Detecting high-emitting methane sources in oil/gas fields using satellite observations

Daniel H. Cusworth[1,2], Daniel J. Jacob[1,2], Jian-Xiong Sheng[2], Joshua Benmergui[2], Alexander J. Turner[3], Jeremy Brandman[4], Laurent White[4], and Cynthia A. Randles[4]

[1]Department of Earth and Planetary Sciences, Harvard University, Cambridge, 02138, USA
[2]School of Engineering and Applied Sciences, Harvard University, Cambridge, 02138, USA
[3]College of Chemistry/Department of Earth and Planetary Sciences, University of California, Berkeley, CA, USA
[4]ExxonMobil Research and Engineering Company, Annandale, NJ, USA

Correspondence to: Daniel Cusworth (dcusworth@fas.harvard.edu)

Abstract. Methane emissions from oil/gas fields originate from a large number of relatively small and densely clustered point sources. A small fraction of high-mode emitters can make a large contribution to the total methane emission. Here we conduct observation system simulation experiments (OSSEs) to examine the potential of recently launched or planned satellites to detect and locate these high-mode emitters through measurements of atmospheric methane columns. We simulate atmospheric methane over a generic oil/gas field (20-500 production sites of different size categories in a 50×50 km^2 domain) for a 1-week period using the WRF-STILT meteorological model with 1.3×1.3 km^2 horizontal resolution. The simulations consider many random realizations for the occurrence and distribution of high-mode emitters in the field by sampling bi-modal probability density functions (pdfs) of emissions from individual sites. The atmospheric methane fields for each realization are observed virtually with different satellite and surface observing configurations. Column methane enhancements observed from satellites are small relative to instrument precision, even for high-mode emitters, so an inverse analysis is necessary. We compare L_1 and L_2 regularizations and show that L_1 regularization effectively provides sparse solutions for a bi-modally distributed variable and enables the retrieval of high-mode emitters. We find that the recently launched TROPOMI instrument (low Earth orbit, 7×7 km^2 nadir pixels, daily return time) and the planned GeoCARB instrument (geostationary orbit, 2.7×3.0 km^2 pixels, 2× or 4×/day return time) are successful (>80% detection rate, <20% false alarm rate) at locating high-emitting sources for fields of 20-50 emitters within the 50×50 km^2 domain as long as skies are clear. They are unsuccessful for denser fields. GeoCARB does not benefit significantly from more frequent observations (4×/day vs. 2×/day) because of temporal error correlation in the inversion, unless under partly cloudy conditions where more frequent observation increases the probability of clear sky. It becomes marginally successful when allowing a 5-km error tolerance for localization. A next-generation geostationary satellite instrument with 1.3×1.3 km^2 pixels, hourly return time, and 1 ppb precision can successfully detect and locate the high-mode emitters for a dense field with up to 500 sites in the 50×50 km^2 domain. The capabilities of TROPOMI and GeoCARB can be usefully augmented with a surface air observation network of 5-20 sites, and in turn the satellite instruments increase the detection capability that can be achieved from the surface sites alone.

1 Introduction

Anthropogenic methane emissions from oil/gas fields originate from a large number of relatively small and densely clustered point sources (Allen et al., 2013). For example, the Barnett Shale in Texas has over 20000 well pads spread over a 300×300 km^2 domain, contributing 40% of total oil/gas emissions from the region (Lyon et al., 2015). It has been estimated that 7% of the wells contribute 50% of total well emissions (Rella et al., 2015; Zavala-Araiza et al., 2015). Identifying such high-emitting wells is of both economic and environmental interest. We present here observing system simulation experiments (OSSEs) to examine the potential of using satellite observations of atmospheric methane for this purpose.

Satellites measure backscattered solar radiation in the shortwave infrared (SWIR) from which atmospheric columns of methane can be retrieved with near uniform sensitivity down to the surface under clear-sky conditions (Jacob et al., 2016). The satellite record for SWIR methane began with the SCIAMACHY instrument (2003-2012; Frankenberg et al., 2005), which provided coarse resolution measurements (30 x 60 km^2 in nadir). The currently operating GOSAT instrument (2009-; Kuze et al., 2016) has finer resolution (10-km diameter pixels) but sparse coverage (individual pixels 250 km apart). The TROPOMI instrument launched in October 2017 provides complete daily coverage at 7×7 km^2 nadir resolution (Hu et al., 2018). The geostationary GeoCARB instrument, to be launched in the early 2020s, is currently planned to provide 2.7×3 km^2 pixel resolution with a return time that may range from one to four times per day (Polonsky et al., 2014; O'Brien et al., 2016). Other geostationary methane satellite missions have been proposed with various combinations of more frequent coverage, finer pixel resolution, and higher instrument precision (Fishman et al., 2012; Butz et al., 2015; Xi et al., 2015; Propp et al., 2017).

A number of studies have examined the value of satellite observations for quantifying methane sources. Inverse analyses of SCIAMACHY and GOSAT data have focused on quantifying emissions at ~100 km regional scales (Bergamaschi et al., 2013; Wecht et al., 2014a; Alexe et al., 2015; Turner et al., 2015). OSSEs have shown the potential for TROPOMI and GeoCARB to effectively constrain emissions at the 25-100 km scale without the multiyear averaging required by SCIAMACHY and GOSAT (Wecht et al., 2014b; Sheng et al., 2018a). Other OSSEs have examined the potential for satellites to quantify large point sources from plume observations (Buchwitz et al., 2013; Rayner et al., 2014; Varon et al., 2018). A recent study by Turner et al. (2018) evaluated the capability of TROPOMI and GeoCARB to quantify emissions in the Barnett Shale down to the kilometer scale for a 1-week observing period. They found that GeoCARB should have some capability for constant sources over a 1-week period but not for transient sources. Hase et al. (2017) simulated surface and aircraft pseudo-observations over North America and used them to constrain North American emissions at $1° \times 1°$ resolution. They found that sparse optimization better constrained local methane hotspots than the standard Bayesian approach.

Here we target a different problem. Given a dense population of production sites (wells) in an oil/gas field, can satellites localize high-mode emitters to enable corrective action? In this problem, quantifying emissions is not as important as identification of the high-mode emitters. The location of the individual point sources is known, but their mode of emission (normal low-mode or high-mode) is unknown. Once a well starts emitting in the high mode, it continues doing so until corrective action is taken. Satellites offer an attractive monitoring approach for identifying high-mode emitters but their

capability may be limited by return frequency, cloud cover, pixel resolution, error in the atmospheric transport model needed to relate the plume to the location of emission, and limitations in the inverse method for identifying sparse high-mode sources. Here we will evaluate the potential of different satellite observing configurations and inverse methods to address this problem with application to TROPOMI, GeoCARB, and finer-resolution geostationary data. We will also examine whether the information from satellites can be usefully complemented with a supporting network of surface observations.

2 Observing System Simulation Experiment

We consider a hypothetical oil/gas field of dimension 50×50 km^2 with 20, 50, 100, or 500 randomly placed production sites (wells), corresponding to site densities of 0.008 km^{-2}, 0.02 km^{-2}, 0.04 km^{-2}, and 0.2 km^{-2}, respectively. The latter case corresponds to the average site density in the Barnett Shale. We create a large ensemble of emission scenarios in each case where different random subsets of sites of different production size categories (small: 10-100 million cubic feet per day (Mcf/d) where 1 Mcf/d = 0.028 Mm^3/d; medium: 100-1000 Mcf/d; large: 1000+ Mcf/d) are in the high-emission mode, and we simulate the resulting atmospheric methane concentration fields with the WRF meteorological model at 1.3×1.3 km^2 resolution. We then sample this pseudo-atmosphere with different satellite and surface observing configurations and apply different inverse methods to detect the high emitters. Detection success is evaluated for each observing configuration and inverse method using statistics for the ensemble of emission scenarios. We describe in this section the different elements of the OSSE.

2.1 Constructing an ensemble of emission fields

Production sites within the 50×50 km^2 domain are randomly placed on the 1.3×1.3 km^2 WRF model grid, with at most one site per grid cell. Emission statistics for the sites are based on observations from the Barnett Shale Coordinated Campaign (Lyon et al. 2015). For each scenario we randomly assign a production size category to each site with 23% of the sites as small, 62% as medium, and 15% as large (Rella et al., 2015). We then assign an emission rate for each site by randomly sampling the bi-modal probability density functions (pdfs) describing low-mode emissions and high-mode emissions for each size category (Lan et al., 2015; Rella et al., 2015; Yacovitch et al., 2015). We assume no other sources in the domain.

Figure 1 shows the pdfs of methane emissions for each production site size category. We flag production sites to be in the high-emission mode if they exceed an emission threshold of 40 kg h^{-1} (axis break in Figure 1), which corresponds on average to 5% of all the sites. High-mode emissions from small facilities are much lower, centered around 24 kg h^{-1}, and would be difficult to distinguish from the normal (low) emission mode. Thus we do not attempt to detect them as high-mode emitters.

Figure 2 shows a sample realization of the oil/gas field with 24 small production sites, 67 medium sites, and 9 large sites (100 total) within the 50×50 km^2 domain. In this realization there are five sites in the high-emission mode. We generate

500 emission scenarios in the same fashion as Figure 2 by randomly assigning size categories for each site (small, medium, large) and randomly sampling the emission pdfs from Figure 1.

2.2 Constructing pseudo-observations of atmospheric methane

We use the meteorological simulation previously generated by Turner et al. (2018) for a 1-week period (October 19-25, 2013) in the Barnett Shale. This simulation applied the Weather Research and Forecasting Model (WRF; Skamarock et al., 2008) at 1.3 km horizontal resolution to drive the Stochastic Time-Inverted Lagrangian Transport (STILT) model (Nehrkorn et al., 2010). STILT is a receptor-oriented Lagrangian particle dispersion model that defines the source footprints for individual atmospheric observations. Turner et al. (2018) applied it to generate 1.3×1.3 km² hourly footprints for any daytime surface or atmospheric column observation in a 70×70 km² domain. Footprints for each column were obtained by releasing and tracking back in time 100 particles from vertical levels centered at 28 m, 97 m, 190 m, 300 m above ground, and 8 additional levels up to 14 km altitude spaced evenly on a pressure grid. The column footprints were weighted with a typical near-uniform SWIR averaging kernel for satellite observations (Worden et al., 2015). Surface observations are taken in the lowest model layer (centered at 28 m above ground) and the corresponding footprints are obtained by releasing and tracking back in time 100 particles at the observation location and time. We use the ensemble of footprints generated by Turner et al. (2018) and add to it hourly footprints for surface observations at night. The 70×70 km² observing domain encompasses our 50×50 km² oil/gas field plus 10 km outside the boundaries (Figure 2) to account for plume transport.

The 70×70 km² archive of WRF-STILT footprints allows us to immediately compute the time-dependent methane concentration field associated with any emission scenario. Figure 3 shows a sample footprint, expressing the sensitivity of atmospheric concentrations at a given location and time i to the emission field upwind. Column footprints are about an order of magnitude smaller than surface footprints because of the dilution effect of the column measurement. Taking the footprints to represent the true atmospheric transport relating emissions to atmospheric concentrations for that location and time, we can combine them with any realization of our emission field (Section 2.1) to generate the true time-dependent methane concentrations in the domain to be sampled by the instruments.

Satellite observations of methane column concentrations are conventionally expressed in unit of dry column mean mixing ratio (ppb), which is the ratio of the vertical column density of methane to the vertical column density of dry air (Jacob et al., 2016). The footprint for location and time i is mathematically represented as $\mathbf{h}_i = (\partial y_i / \partial \mathbf{x})^T$ (units ppb μmol⁻¹ m² s) where y_i is the methane concentration (ppb) for that location and time, and $\mathbf{x}$ (μmol m⁻² s⁻¹) is a vector of dimension n describing the emission field for the n emitters in the domain. We select the footprint information that corresponds to the locations of the n emitters so that $\mathbf{h}_i$ is also a vector of n dimension. The true atmospheric concentration can be immediately constructed for any emission field $\mathbf{x}$ as $y_i = \mathbf{h}_i \bullet \mathbf{x} + b$, where $\bullet$ denotes the scalar product and b is a background assumed here to be constant.

A given methane observing configuration makes m observations of the domain over the 1-week simulation period. The true methane concentrations for that observation ensemble can be assembled as an m-dimensional vector $\mathbf{y_{true}} = \mathbf{Hx}$ where

$\mathbf{H} = \partial\mathbf{y_{true}}/\partial\mathbf{x}$ is the $m \times n$ Jacobian matrix of footprints with rows $\mathbf{h}_i^T$. The pseudo-observations are then generated as $\mathbf{y} = \mathbf{y}_{true}$ + $\sigma\boldsymbol{\varepsilon}$ where σ is the instrument precision (one standard deviation) and the vector $\boldsymbol{\varepsilon}$ is a random realization of Gaussian noise with mean value of zero and standard deviation of unity for each vector element. SWIR instruments may also suffer from systematic errors but we do not account for those here in the absence of information. The largest source of systematic error on our scale would likely be the inhomogeneity in surface reflectivity (Pfister et al., 2005).

The mean daytime 10 m horizontal wind speed inside the observing domain during the simulated week is 5.4 m/s. Stronger winds could further dilute plumes within an observing domain, making the ability for satellite detection of emitters more difficult; on the other hand, the model transport error is less for stronger winds (Varon et al., 2018).

2.3 Satellite and surface observing configurations

Table 1 describes the different satellite observing configurations evaluated in this work including TROPOMI, GeoCARB with 2 or 4 return times per day, and an aspirational next-generation geostationary instrument with 1.3×1.3 km^2 pixel resolution, 1 ppb precision, and hourly return frequency between 8 and 17 local time (LT). Successful methane retrievals from satellites require clear sky. The probability of clear sky in a partly cloudy domain depends greatly on pixel size (Remer et al., 2012). Results for a partly cloudy condition would depend on the particular cloud configuration and would be difficult to generalize. Here we assume clear-sky conditions to avoid this complication, but the detection probability for high-mode emitters should then be viewed as an upper limit. In particular, it should be recognized that no detection from satellite is possible for a cloudy domain.

We also wish to determine the benefit of a well-positioned surface air monitoring network for supplementing the satellite observations. Assume that we have M fixed monitoring instruments to deploy measuring surface air methane concentrations *in situ*. We want to place them in a configuration that maximizes the information that they would provide, assuming an isotropic wind for generality. A trivial solution would be to place an instrument at each production site, in which case the monitoring problem would be fully solved, but this solution may not be practical for a large number of production sites. Given a known spatial distribution of emitters (the locations of the production sites), we use the k-means spatial clustering approach (Hartigan and Wong, 1979) to select monitoring site locations minimizing the distances to emitter locations. Figure 2 shows the selected locations for five surface monitoring sites. We assume that these sites report hourly data with 1 ppb precision and that the background concentration in surface air is constant, consistent with the assumption made for satellite observations. A variable background would complicate the problem but could be retrieved as part of the inversion (Wecht et al., 2014b).

An important consideration in the interpretation of satellite observations is that methane column enhancements from individual point sources are typically small relative to instrument precision, even in the high-emitting mode (Jacob et al., 2016; Varon et al., 2018). Figure 4 shows the pixel-resolved distribution of atmospheric methane column enhancements above the background for a single pass of the different satellite instruments sampling the emission field of Figure 2. The enhancements

are less than 1 ppb even for 1.3×1.3 km² pixels and are weaker at coarser pixel resolution. This is less than the single-scene precision of the satellite instruments (Table 1). Successful detection of high-mode emitters thus requires the sampling of many pixels, across the plume and/or through repeated sampling, to reduce the noise. This is less of an issue for surface air measurements, where methane enhancements are an order of magnitude higher (Figure 3). On the other hand, surface

5   monitoring sites are spatially sparse. For both satellite and surface air observations, a formal inverse analysis of the ensemble of atmospheric observations accounting for plume transport is required for detection of the high-mode emitters.

2.4 Inverse methods

Given a set of observations **y** and Jacobian matrix **H**, we need an inverse method to determine the best solution $\hat{\mathbf{x}}$ of the

10   emission field **x** at predetermined locations. The inversion should be able to detect the small fraction of sources in the high-emitting mode, with detection being more important than quantification. This is known as a sparse-solution problem, where most elements of the emission field **x** are very small (for which an optimized value of zero would be acceptable), and a few of the elements are relatively large. We use regularized least squares regression (Hansen, 2010), also known as Tikhonov Regularization, where the solution is found by minimizing the cost function $J(\mathbf{x})$,

$$J(\mathbf{x}) = (\mathbf{Hx} - \mathbf{y})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{Hx} - \mathbf{y}) + \lambda\|\mathbf{x}\|_L^p \quad (1)$$

Here the first term on the right hand side represents the ordinary least-squares cost function, such that the solution would minimize the residuals between the prediction **Hx** and the observations weighted by the observational error covariance matrix

20   **R**. The second term represents an adjustable parameter λ and the L-norm of **x**, which is a measure of the magnitude of the vector **x** defined as the following:

$$\|x\|_L = \sqrt[L]{\Sigma_{k=1}^n |x_k|^L} \quad (2)$$

Adding this second term in the cost function penalizes the total magnitude of **x** in the solution, which reduces overfitting to

25   noise and regularizes the solution. When $L = 1$ and $p = 1$, this is known as L_1-regularization or the least absolute shrinkage and selection operator (LASSO; Tibshirani, 1996), and Equation 1 takes the form:

$$J(\mathbf{x}) = (\mathbf{Hx} - \mathbf{y})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{Hx} - \mathbf{y}) + \lambda\sum_{k=1}^n |x_k| \quad (3)$$

When $L = 2$ and $p = 2$, Equation 1 takes the form known as L_2-regularization or Ridge Regression (Evgeniou et al., 2000):

30

$$J(\mathbf{x}) = (\mathbf{Hx} - \mathbf{y})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{Hx} - \mathbf{y}) + \lambda\mathbf{x}^T\mathbf{x} \quad (4)$$

Equation 4 is equivalent to the standard Bayesian optimization (Rodgers, 2000) assuming Gaussian distributions, a prior emission estimate of zero, and uniform prior error variance of λ^{-1}.

The observational error covariance matrix $\mathbf{R} = (r_{ij})$ accounts for both instrument and model transport errors. Representation errors are negligible due to the model grid resolution being finer or the same resolution as the instrument pixels (Turner et al., 2018). The diagonal terms add the corresponding error variances in quadrature:

$$r_{ii} = \sigma_I^2 + \sigma_M^2 \quad (5a)$$

where σ_I is the instrument error standard deviation as given by the precision in Table 1, and σ_M is the model transport error standard deviation previously estimated to be 4 ppb for methane columns (Turner et al., 2018). Given the order of magnitude difference in sensitivity between satellite columns and surface measurements (Figure 3), we assume σ_M to be 40 ppb for surface measurements. Off-diagonal terms account for model transport error correlation between different observations. Following Turner et al. (2018), we assume a temporal error correlation length scale (τ) of 2 hours and a spatial error correlation length scale (ℓ) of 40 km:

$$r_{ij} = \sigma_M^2 \exp\left\{-\frac{d}{\ell}\right\} \exp\left\{-\frac{t}{\tau}\right\} \quad \text{for } i \neq j \quad (5b)$$

where d and t are the distance and elapsed time between observations y_i and y_j.

Additional model transport error correlation applies when combining satellite and surface air observations in the inversion, since the footprints can be similar (Figure 3). To quantify this error correlation, we use the work of Sheng et al. (2018b), who jointly compared column (TCCON) and surface air (NOAA) measurements of methane at Lamont, Oklahoma with GEOS-Chem transport model simulations. By correlating the model-observation differences for coincident column (i) and surface air (j) observations we find a model transport error correlation coefficient cor(i, j) = 0.65 that we apply to the corresponding off-diagonal terms:

$$r_{ij} = \text{cor}(i,j)\, \sigma_{Mi}\sigma_{Mj}\exp\left\{-\frac{d}{\ell}\right\} \exp\left\{-\frac{t}{\tau}\right\} \quad (5c)$$

Inverse solutions derived using L_1-regularization produce sparser solutions than the L_2 counterpart (Tibshirani, 1996), which is desirable for our application and has previously been shown to produce good results for constraining methane hotspots (Hase et al. 2017). Here we will perform both L_1 and L_2 inversions and compare the results. Minimization of $J(\mathbf{x})$ in Equations 3 and 4 to obtain the solution $\hat{\mathbf{x}}$ corresponding to $dJ/d\mathbf{x} = 0$ is done numerically using coordinate gradient descent (Friedman et al., 2009). The regularization parameter λ is chosen so that the mismatch between model and observations is small, but not so small that the solution $\hat{\mathbf{x}}$ is overfit to random noise, which would occur when $\lambda = 0$. We use the process of 5-fold cross-

validation to select an optimal λ value (Arlot and Celisse, 2010). This process randomly samples **H** and **y** into a training and validation set. Minimization of J is done on the training set using an array of λ values. The process is repeated five times, and the value of λ that on average minimizes the residual error on the validation set is retained.

Figure 5 shows the distribution $\hat{x}$ from a single realization of emissions, GeoCARB 4×/day pseudo-observations, and both L_1 and L_2 regularization. In this simulation, L_1 regularization enables the retrieval of high-mode emitters while L_2 regularization is more restrictive in allowing excursions from the low-mode mean.

2.5 Detection of high-emission modes

Success in the detection of high-mode emitters from the distribution of $\hat{x}$ can be determined by comparison to the actual occurrence and location of these emitters as defined in Section 2.1 and illustrated in Figure 2. In a real-world application we would not know the actual pdfs of emissions (Figure 1), so we need to diagnose the occurrence of high-mode emitters on the basis of anomalies in the distribution of $\hat{x}$. We define high-mode elements as being more than S standard deviations from the mean of the $\hat{x}$ distribution, where S is varied in the 1.65-2.5 range to examine the associated sensitivity. Using anomaly detection on $\hat{x}$ instead of a fixed threshold (e.g., 40 kg h^{-1}) allows for generalization to other emission fields where the mean normal and high modes may be different than the Barnett Shale. Figure 5 shows thresholds for classifying high-mode emitters using anomaly detection and a fixed value of 40 kg h^{-1}. The L_1 threshold is larger than the L_2 threshold, but smaller than the 40 kg h^{-1}. Had the fixed threshold been used, some high-mode emitters (relative to $\hat{x}$) would have not been classified as such.

The detection of high-mode emitters by the inversion is graded into four categories: 1) true positives (TP), or the inversion correctly identifying the locations of the high-mode emitters, 2) true negatives (TN), or the inversion correctly identifying the locations of the low-mode emitters, 3) false positives (FP), or the inversion signaling a high-mode emitter when in reality the emitter is in the low mode, and 4) false negatives (FN), or the inversion signaling a low-mode emitter when in reality the emitter is in the high mode.

We compile these grades into three overall performance metrics (Brasseur and Jacob, 2017). The probability of detection (POD) is defined as the ratio of true positives to true positives plus false negatives:

$$POD = \frac{\Sigma\,TP}{\Sigma\,TP + \Sigma\,FN}$$

This metric measures the ability to detect high-mode emitters. The false alarm ratio (FAR) is defined as the ratio of false positives to false positives plus true positives:

$$FAR = \frac{\Sigma\,FP}{\Sigma\,TP + \Sigma\,FP}$$

This metric measures the reliability of high-mode emission occurrences detected by the inversion.

A perfect observing system would have a POD of one and a FAR of zero. Here we define a successful observing system as achieving a POD of 0.8 (80%) and a FAR of 0.2 (20%). These criteria, although somewhat arbitrary, allow us to succinctly summarize the success of each observing configuration.

5   We combine the POD and FAR metrics into one overall performance metric called the Equitable Threat Score (ETS; Wang, 2014):

$$ETS = \frac{\Sigma\,TP - \alpha}{\Sigma\,TP + \Sigma\,FP + \Sigma\,FN - \alpha}$$

where α is the number of TP predictions that are expected by chance:

$$\alpha = \frac{(\Sigma\,TP + \Sigma\,FP)(\Sigma\,TP + \Sigma\,FN)}{\Sigma\,TP + \Sigma\,FP + \Sigma\,FN + \Sigma\,TN} = \frac{1}{N}\frac{\Sigma\,FP}{FAR}\frac{\Sigma\,TP}{POD}$$

and $N = \Sigma\,TP + \Sigma\,FP + \Sigma\,FN + \Sigma\,TN$. The ETS measures how well the high-mode emitters detected by the observing system correspond to the actual occurrences, beyond what could be achieved by chance. A perfect observing system has an ETS of one, and a system performing worse than chance would have a negative ETS. An observing system with POD of 0.8 and FAR of 0.2 has an ETS of 0.65 for a field where 5% of emitters are in the high mode. We take this as our ETS criterion

15   for successful detection.

3 Results and discussion

3.1 Performance of different satellite and surface observing systems

We begin by testing the ability of each satellite configuration of Table 1 to detect high-mode emitters from fields of 20 to 500

20   randomly scattered production sites within the 50×50 km^2 domain. For a given number of sites, we conduct each test for 500 different realizations of the emission field assigning randomly each production site to a size category (small, medium, large) and sampling randomly the pdfs of Figure 1. Emitter locations are fixed across all 500 realizations. Figure 6 shows the POD, FAR, and ETS results for a field of 100 emitters and compares the results of L_1 and L_2 regularizations. The values represent the mean results for the ensemble of 500 realizations, and the error bars represent the range of results when the high-mode

25   detection threshold S is varied from 1.65 to 2.5. We find that L_1 regularization provides better predictions for all cases. This is especially the case for the next-generation satellite, where L_1 regularization produces a POD of 0.85 with a near-perfect FAR of 0.04. L_2-regularization is more conducive to spreading emissions across a broader array of state vector elements. The better performance of L_1-regularization is also observed for other site densities (not shown). We use L_1 regularization in what follows.

Figure 6 also compares the performances of the satellite observing systems to those of an ensemble of 5-20 optimally

30   placed (k-means) surface sites. We find that the surface observing system performs comparably to GeoCARB. We explore combining satellite and surface observations into a single prediction in Section 3.3.

The results from Figure 6 show that TROPOMI and GeoCARB are unsuccessful in locating high-mode emitters for a field of 100 production sites (0.04 sites km^{-2}). We examine the sensitivity of this result to site density. Figure 7 compares the detection results for fields of 20, 50, 100, and 500 production sites within the 50×50 km^2 domain. For a field of only 20 emitters, TROPOMI is successful and GeoCARB produces near perfect results. For a field of 50 emitters, TROPOMI is no longer successful, but GeoCARB is still marginally successful due to finer pixel resolution and higher instrument precision. We find in general that GeoCARB gains little by sampling four times a day (4×/day) vs. 2×/day. This is due to the temporal model error correlation between successive GeoCARB observations. Accounting for cloud cover would show more benefit from 4×/day observations, since a higher frequency of observations allows a greater chance of sampling clear-sky conditions, although the benefit depends on the cloud persistence time scale (Sheng et al., 2018a).

The ability of a satellite observing configuration to localize high-mode emitters thus depends not only on repeat time, resolution, precision, and cloud cover, but also on the density of emitters within a field. For the high-density fields of 100 and 500 production sites considered here (0.04 sites km^{-2} and 0.2 sites km^{-2}), we find that only the next-generation satellite instrument is successful. Actual fields can be even denser but we are limited in our investigation by the 1.3 × 1.3 km^2 resolution of the WRF simulation. Detecting individual high-mode emitters in denser fields would require geostationary satellite observations with sub-km pixels but this is beyond the scope of current proposals.

3.2 Spatial tolerance in detection of high-mode emitters

The results from Figure 7 are somewhat pessimistic regarding the ability of near-future satellite observations (TROPOMI and GeoCARB) to detect the locations of high-mode emitters in dense emission fields. It may be acceptable to relax the localization criterion. If the observing system detects a false positive that is sufficiently close to the actual location of a high-mode emitter, then the detection may still have some value. In our OSSE setup, localization is effectively limited by the 1.3 × 1.3 km^2 grid resolution of the WRF simulation. To examine the sensitivity to localization, we repeated the analysis allowing for 3-5 km tolerance of false predictions. Figure 8 shows the results for a field of 100 emitters. We find that spatial tolerance significantly improves the performance of GeoCARB but still falls short of our success criterion. The FAR decreases below 0.2 for 3 km tolerance and below 0.1 for 5 km tolerance, but the POD only improves to 0.7 and thus the ETS remains below 0.65.

3.3 Combining satellite and surface observations

We saw in Section 3.1 that only the next-generation satellite instrument can successfully detect high-mode emitters when the site density is high. Here we examine if a combination of satellite and surface observations can improve detection, i.e., if TROPOMI and GeoCARB could benefit from an *in situ* supporting surface network and vice versa. This is addressed with a joint inversion of the satellite and surface observations, taking into account the error correlation between the two as described in Section 2.4.

Figure 9 shows the results for a field of 100 emitters. The already successful next-generation instrument shows no benefit from added surface sites. On the other hand, the surface sites provide greatly added value to TROPOMI and GeoCARB. Adding 10-20 surface sites enables near-successful detection of the high-mode emitters. At the same time, TROPOMI and GeoCARB data add significantly to the performance of a surface observing system alone by providing observations with more spatial coverage. We find that TROPOMI and GeoCARB perform similarly when added to surface sites, and that their main benefit is to decrease the FAR. Accounting for clouds would show more benefit for GeoCARB because the finer pixels allow for more frequent clear-sky observations (Sheng et al., 2018a).

4 Conclusions

We performed observing system simulation experiments (OSSEs) to test the ability of near-future satellite instruments measuring atmospheric methane (TROPOMI, GeoCARB, next-generation geostationary) to detect high-mode point source emitters among a dense field of individual point sources, alone or supported by a surface monitoring network. We focused on the practical problem of detecting high-mode emitters in an oil/gas production field with a high density of wells. Remote detection from satellites, combined with operator knowledge, could supplement on-site leak detection and repair (LDAR) programs to identify and fix unexpected high emitters. Our results in these meteorological conditions can be summarized usefully in terms of answers to questions that a field manager might have:

Can I rely on satellite data alone to detect high-mode emitters among the production sites in my oil/gas field?

We find that TROPOMI and GeoCARB can detect high-mode emitters as long as the density of point sources is relatively small (20 sites within our 50×50 km^2 domain, or a density of 0.008 km^{-2}) and skies are clear. GeoCARB shows little difference in success rate (Equitable Threat Score (ETS) > 0.65) for 2 or 4 overpasses per day. GeoCARB is marginally successful for 50 sites (0.02 km^{-2}) but fails for 100 sites (0.04 km^{-2}). A next-generation geostationary satellite instrument with ~1-km pixel resolution and hourly return time would deliver precise detection in dense fields up to 500 sites (0.2 km^{-2}). Allowing for a 5-km spatial error tolerance for localization, we find that GeoCARB comes close to successful detection in a field of 100 sites.

How should I analyze the satellite observations to detect high-mode emitters?

Detection of high-mode emitters from satellite observations is not a simple matter of flagging hot spots because the methane column enhancements are typically small compared to instrument precision, even for high-mode emitters. Repeated clear-sky observation combined with inverse analysis using an atmospheric transport model is needed. We find that an inversion with L_1 regularization produces better results than L_2-regularization. This is expected since the L_1 regularization method is designed to recover sparse signals.

Can I usefully supplement satellite information with surface monitoring?

Both TROPOMI and GeoCARB add significantly to the information provided by a surface monitoring network of 5-20 sites within the 50×50 km^2 domain, and conversely the addition of a surface network enhances significantly the information that can be retrieved from TROPOMI and GeoCARB. The combination of these satellite instruments with the surface monitors can deliver successful detection of high-mode emitters through a joint inversion. Adding surface sites provides no benefit to the

5   next-generation geostationary instrument, which can successfully detect high-mode emitters on its own as long as skies are clear.

Data Availability. The WRF-STILT model is available for download at https://uataq.github.io/stilt/. A worked through example of the high-mode detection observing system simulation experiment (OSSE) described in this paper is available in

10   the Supplementary Information for this paper.

Acknowledgments. This work was supported by the ExxonMobil Research and Engineering Company, the US Department of Energy (DOE) Advanced Research Projects Agency – Energy (ARPA-E), and the NASA Earth Science Division. This research used the Savio computational cluster resource provided by the Berkeley Research Computing program at the

15   University of California, Berkeley (supported by the UC Berkeley Chancellor, Vice Chancellor for Research, and Chief Information Officer). This research also used resources from the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

References

Alexe, M., Bergamaschi, P., Segers, A., Detmers, R., Butz, A., Hasekamp, O., Guerlet, S., Parker, R., Boesch, H., Frankenberg,

20   C., Scheepmaker, R. A., Dlugokencky, E., Sweeney, C., Wofsy, S. C., and Kort, E. A.: Inverse modelling of CH$_4$ emissions for 2010-2011 using different satellite retrieval products from GOSAT and SCIAMACHY, Atmospheric Chemistry and Physics, 15, 113–133, doi:10.5194/acp-15-113-2015, URL https://www.atmos-chem-phys. net/15/113/2015/, 2015.

Allen, D. T., Torres, V. M., Thomas, J., Sullivan, D. W., Harrison, M., Hendler, A., Herndon, S. C., Kolb, C. E., Fraser, M. P., Hill, A. D., Lamb, B. K., Miskimins, J., Sawyer, R. F., and Seinfeld, J. H.: Measurements of methane emissions at natural

25   gas production sites in the United States, Proceedings of the National Academy of Sciences, 110, 17 768–17 773, doi:10.1073/ pnas.1304880110, URL http://www.pnas.org/content/110/44/17768, 2013.

Arlot, S. and Celisse, A.: A survey of cross-validation procedures for model selection, Statistics surveys, 4, 40–79, 2010.

Bergamaschi, P., Houweling, S., Segers, A., Krol, M., Frankenberg, C., Scheepmaker, R. A., Dlugokencky, E., Wofsy, S. C., Kort, E. A., Sweeney, C., Schuck, T., Brenninkmeijer, C., Chen, H., Beck, V., and Gerbig, C.: Atmospheric CH4 in the first

30   decade of the 21st century: Inverse modeling analysis using SCIAMACHY satellite retrievals and NOAA surface

measurements, Journal of Geophysical Research: Atmospheres, 118, 7350–7369, doi:10.1002/jgrd.50480, URL http://dx.doi.org/10.1002/jgrd.50480, 2013.

Bousserez, N., Henze, D. K., Rooney, B., Perkins, A., Wecht, K. J., Turner, A. J., Natraj, V., and Worden, J. R.: Constraints on methane emissions in North America from future geostationary remote-sensing measurements, Atmospheric Chemistry and Physics, 16, 6175–6190, doi:10.5194/acp-16-6175-2016, URL https://www.atmos-chem-phys.net/16/6175/2016/, 2016.

Brasseur, G. P. and Jacob, D. J.: Modeling of Atmospheric Chemistry, Cambridge University Press, 2017.

Buchwitz, M., Reuter, M., Bovensmann, H., Pillai, D., Heymann, J., Schneising, O., Rozanov, V., Krings, T., Burrows, J. P., Boesch, H., Gerbig, C., Meijer, Y., and Lo¨scher, A.: Carbon Monitoring Satellite (CarbonSat): assessment of atmospheric CO_2 and CH_4 retrieval errors by error parameterization, Atmospheric Measurement Techniques, 6, 3477–3500, doi:10.5194/ amt-6-3477-2013, URL https://www.atmos-meas-tech.net/6/3477/2013/, 2013.

Butz, A., Orphal, J., Checa-Garcia, R., Friedl-Vallon, F., von Clarmann, T., Bovensmann, H., Hasekamp, O., Landgraf, J., Knigge, T., Weise, D., Sqalli-Houssini, O., and Kemper, D.: Geostationary Emission Explorer for Europe (G3E): mission concept and initial performance assessment, Atmospheric Measurement Techniques, 8, 4719–4734, doi:10.5194/ amt-8-4719-2015, URL https://www.atmos-meas-tech.net/8/4719/2015/, 2015.

Fishman, J., Iraci, L. T., Al-Saadi, J., Chance, K., Chavez, F., Chin, M., Coble, P., Davis, C., DiGiacomo, P. M., Edwards, D., Eldering, A., Goes, J., Herman, J., Hu, C., Jacob, D. J., Jordan, C., Kawa, S. R., Key, R., Liu, X., Lohrenz, S., Mannino, A., Natraj, V., Neil, D., Neu, J., Newchurch, M., Pickering, K., Salisbury, J., Sosik, H., Subramaniam, A., Tzortziou, M., Wang, J., and Wang, M.: The United States' Next Generation of Atmospheric Composition and Coastal Ecosystem Measurements: NASA's Geostationary Coastal and Air Pollution Events (GEO-CAPE) Mission, Bulletin of the American Meteorological Society, 93, 1547–1566, doi:10.1175/BAMS-D-11-00201.1, URL https://doi.org/10.1175/BAMS-D-11-00201.1, 2012.

Frankenberg, C., Meirink, J. F., van Weele, M., Platt, U., and Wagner, T.: Assessing Methane Emissions from Global Space-Borne Observations, Science, 308, 1010–1014, doi:10. 1126/science.1106644, URL http://science.sciencemag.org/content/308/5724/1010, 2005.

Friedman, J., Hastie, T., and Tibshirani, R.: Regularization paths for generalized linear models via coordinate descent, Journal of statistical software, 33, 1, 2010.

Hartigan, J. A. and Wong, M. A.: Algorithm AS 136: A k-means clustering algorithm, Journal of the Royal Statistical Society. Series C (Applied Statistics), 28, 100–108, 1979.

Hansen, P.C.: Discrete inverse problems: insight and algorithms. Vol. 7. Siam, 2010.

Hase, N., Miller, S. M., Maass, P., Notholt, J., Palm, M., and Warneke, T.: Atmospheric inverse modeling via sparse reconstruction, Geoscientific Model Development, 10, 3695–3713, URL http://search.proquest.com.ezp-prod1.hul.harvard.edu/ docview/1949064962?accountid=11311, copyright - Copyright Copernicus GmbH 2017; Last updated - 2018-01-29, 2017.

5 Hu, H., Jochen, L., Rob, D., Tobias, B., de Brugh Joost, A., Ilse, A., André´, B., and Otto, H.: Toward Global Mapping of Methane With TROPOMI: First Results and Intersatellite Comparison to GOSAT, Geophysical Research Letters, 0, doi:10.1002/2018GL077259, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2018GL077259, 2018.

Jacob, D. J., Turner, A. J., Maasakkers, J. D., Sheng, J., Sun, K., Liu, X., Chance, K., Aben, I., McKeever, J., and Frankenberg, C.: Satellite observations of atmospheric methane and their value for quantifying methane emissions,
10 Atmospheric Chemistry and Physics, 16, 14 371–14 396, doi:10.5194/acp-16-14371-2016, URL https://www.atmos-chem-phys.net/16/14371/2016/, 2016.

Kuze, A., Suto, H., Shiomi, K., Kawakami, S., Tanaka, M., Ueda, Y., Deguchi, A., Yoshida, J., Yamamoto, Y., Kataoka, F., Taylor, T. E., and Buijs, H. L.: Update on GOSAT TANSO-FTS performance, operations, and data products after more than 6 years in space, Atmospheric Measurement Techniques, 9, 2445–2461, doi:10.5194/amt-9-2445-2016, URL https://www.
15 atmos-meas-tech.net/9/2445/2016/, 2016.

Lan, X., Talbot, R., Laine, P., and Torres, A.: Characterizing Fugitive Methane Emissions in the Barnett Shale Area Using a Mobile Laboratory, Environmental Science & Technology, 49, 8139–8146, doi:10.1021/es5063055, URL http://dx.doi.org/10.1021/es5063055, pMID: 26148552, 2015.

Lyon, D. R., Zavala-Araiza, D., Alvarez, R. A., Harriss, R., Palacios, V., Lan, X., Talbot, R., Lavoie, T., Shepson, P.,
20 Yacovitch, T. I., Herndon, S. C., Marchese, A. J., Zimmerle, D., Robinson, A. L., and Hamburg, S. P.: Constructing a Spatially Resolved Methane Emission Inventory for the Barnett Shale Region, Environmental Science & Technology, 49, 8147–8157, doi:10.1021/es506359c, URL http://dx.doi.org/10.1021/es506359c, pMID: 26148553, 2015.

Nehrkorn, T., Eluszkiewicz, J., Wofsy, S. C., Lin, J. C., Gerbig, C., Longo, M., and Freitas, S.: Coupled weather research and forecasting–stochastic time-inverted lagrangian transport (WRF–STILT) model, Meteorology and Atmospheric Physics, 107,
25 51–64, doi:10.1007/ s00703-010-0068-x, URL https://doi.org/10.1007/s00703-010-0068-x, 2010.

O'Brien, D. M., Polonsky, I. N., Utembe, S. R., and Rayner, P. J.: Potential of a geostationary geoCARB mission to estimate surface emissions of CO 2, CH 4 and CO in a polluted urban environment: case study Shanghai, Atmospheric Measurement Techniques, 9, 4633, 2016.

Pfister, G., Gille, J. C., Ziskin, D., Francis, G., Edwards, D.P., Deeter, M.N., and Abbott, E.: Effects of a spectral surface
30 reflectance on measurements of backscattered solar radiation: Application to the MOPITT methane retrieval, Journal of Atmospheric and Oceanic Technology 22, 5, 566-574, 2005.

Polonsky, I. N., O'Brien, D. M., Kumer, J. B., O'Dell, C. W., and the geoCARB Team: Performance of a geostationary mission, geoCARB, to measure CO_2, CH_4 and CO column-averaged concentrations, Atmospheric Measurement Techniques, 7, 959–981, doi:10. 5194/amt-7-959-2014, URL https://www.atmos-meas-tech.net/7/959/2014/, 2014.

Propp, A. M., Benmergui, J. S., Turner, A. J., and Wofsy, S. C.: MethaneSat: Detecting Methane Emissions in the Barnett Shale Region, AGU Fall Meeting Abstracts, 2017.

Rayner, P. J., Utembe, S. R., and Crowell, S.: Constraining regional greenhouse gas emissions using geostationary concentration measurements: a theoretical study, Atmospheric Measurement Techniques, 7, 3285–3293, doi:10.5194/amt-7-3285-2014, URL https://www. atmos-meas-tech.net/7/3285/2014/, 2014.

Rella, C. W., Tsai, T. R., Botkin, C. G., Crosson, E. R., and Steele, D.: Measuring Emissions from Oil and Natural Gas Well Pads Using the Mobile Flux Plane Technique, Environmental Science & Technology, 49, 4742–4748, doi:10.1021/acs.est.5b00099, pMID: 25806837, 2015.

Remer, L. A., Mattoo, S., Levy, R. C., Heidinger, A., Pierce, R. B., and Chin, M.: Retrieving aerosol in a cloudy environment: aerosol product availability as a function of spatial resolution, Atmos. Meas. Tech., 5, 1823-1840, https://doi.org/10.5194/amt-5-1823-2012, 2012.

Rodgers, C. D.: Inverse methods for atmospheric sounding: theory and practice, vol. 2, World scientific, 2000.

Sheng, J.-X., Jacob, D. J., Maasakkers, J. D., Zhang, Y., and Sulprizio, M. P.: Comparative analysis of low-Earth orbit (TROPOMI) and geostationary (GeoCARB, GEO-CAPE) satellite instruments for constraining methane emissions on fine regional scales: application to the Southeast US, Atmospheric Measurement Techniques Discussions, 2018, 1–15, doi:10.5194/amt-2018-121, URL https://www.atmos-meas-tech-discuss.net/ amt-2018-121/, 2018a.

Sheng, J.-X., Jacob, D. J., Turner, A. J., Maasakkers, J. D., Sulprizio, M. P., Bloom, A. A., Andrews, A. E., and Wunch, D.: High-resolution inversion of methane emissions in the Southeast US using SEAC4RS aircraft observations of atmospheric methane: anthropogenic and wetland sources, Atmospheric Chemistry and Physics, 18, 6483–6491, doi:10.5194/ acp-18-6483-2018, URL https://www.atmos-chem-phys.net/18/6483/2018/, 2018b.

Skamarock, W., Klemp, J., Dudhia, J., Gill, D., Barker, D., Dudha, M., Huang, X., Wang, W., and Powers, Y.: A description of the advanced research WRF Ver. 30, National Center for Atmospheric Research US, 2008.

Tibshirani, R.: Regression shrinkage and selection via the lasso, Journal of the Royal Statistical Society. Series B (Methodological), pp. 267–288, 1996.

Turner, A. J., Jacob, D. J., Wecht, K. J., Maasakkers, J. D., Lundgren, E., Andrews, A. E., Biraud, S. C., Boesch, H., Bowman, K. W., Deutscher, N. M., Dubey, M. K., Griffith, D. W. T., Hase, F., Kuze, A., Notholt, J., Ohyama, H., Parker, R., Payne, V. H., Sussmann, R., Sweeney, C., Velazco, V. A., Warneke, T., Wennberg, P. O., and Wunch, D.: Estimating global and

North American methane emissions with high spatial resolution using GOSAT satellite data, Atmospheric Chemistry and Physics, 15, 7049–7069, doi:10.5194/acp-15-7049-2015, URL https://www.atmos-chem-phys.net/15/7049/2015/, 2015.

Turner, A. J., Jacob, D. J., Benmergui, J., Brandman, J., White, L., and Randles, C. A.: Assessing the capability of different satellite observing configurations to resolve the distribution of methane emissions at kilometer scales, Atmospheric Chemistry and Physics Discussions, 2018, 1–23, doi:10.5194/acp-2018-164, URL https://www.atmos-chem-phys-discuss.net/acp-2018-164/, 2018.

Varon, D. J., Jacob, D. J., McKeever, J., Jervis, D., Durak, B. O. A., Xia, Y., and Huang, Y.: Quantifying methane point sources from fine-scale (GHGSat) satellite observations of atmospheric methane plumes, Atmospheric Measurement Techniques Discussions, 2018, 1–25, doi:10.5194/amt-2018-171, URL https://www.atmos-meas-tech-discuss.net/ amt-2018-171/, 2018.

Wang, C.-C.: On the Calculation and Correction of Equitable Threat Score for Model Quantitative Precipitation Forecasts for Small Verification Areas: The Example of Taiwan, Weather and Forecasting, 29, 788–798, doi:10.1175/WAF-D-13-00087.1, URL https://doi.org/10.1175/WAF-D-13-00087.1, 2014.

Wecht, K. J., Jacob, D., Frankenberg, C., Jiang, Z., and Blake, D.: Mapping of North American methane emissions with high spatial resolution by inversion of SCIAMACHY satellite data, Journal of Geophysical Research: Atmospheres, 119, 7741–7756, doi:10.1002/ 2014JD021551, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/ 2014JD021551, 2014a.

Wecht, K. J., Jacob, D. J., Sulprizio, M. P., Santoni, G. W., Wofsy, S. C., Parker, R., Bosch, H., and Worden, J.: Spatially resolving methane emissions in California: constraints from the CalNex aircraft campaign and from present (GOSAT, TES) and future (TROPOMI, geostationary) satellite observations, Atmospheric Chemistry and Physics, 14, 8173–8184, doi:10.5194/ acp-14-8173-2014, URL https://www.atmos-chem-phys.net/14/8173/2014/, 2014b.

Worden, J. R., Turner, A. J., Bloom, A., Kulawik, S. S., Liu, J., Lee, M., Weidner, R., Bowman, K., Frankenberg, C., Parker, R., and Payne, V. H.: Quantifying lower tropospheric methane concentrations using GOSAT near-IR and TES thermal IR measurements, Atmospheric Measurement Techniques, 8, 3433–3445, doi:10.5194/amt-8-3433-2015, URL https://www.atmos-meas-tech.net/8/3433/2015/, 2015.

Xi, X., Natraj, V., Shia, R. L., Luo, M., Zhang, Q., Newman, S., Sander, S. P., and Yung, Y. L.: Simulated retrievals for the remote sensing of CO_2, CH_4, CO, and H_2O from geostationary orbit, Atmospheric Measurement Techniques, 8, 4817–4830, doi:10.5194/amt-8-4817-2015, URL https://www.atmos-meas-tech.net/8/4817/2015/, 2015.

Yacovitch, T. I., Herndon, S. C., Petron, G., Kofler, J., Lyon, D., Zahniser, M. S., and Kolb, C. E.: Mobile Laboratory Observations of Methane Emissions in the Barnett Shale Region, Environmental Science & Technology, 49, 7889–7895, doi:10.1021/es506352j, URL http:// dx.doi.org/10.1021/es506352j, pMID: 25751617, 2015.

Zavala-Araiza, D., Lyon, D., Alvarez, R. A., Palacios, V., Harriss, R., Lan, X., Talbot, R., and Hamburg, S. P.: Toward a Functional Definition of Methane Super-Emitters: Application to Natural Gas Production Sites, Environmental Science & Technology, 49, 8167–8174, doi:10.1021/acs.est.5b00133, URL http://dx.doi.org/10.1021/acs.est. 5b00133, pMID: 26148555, 2015.

Table 1. Observing configurations considered in this work.

Instrument	Observation Frequency	Pixel size (km^2)	Precision[a] (ppb)	Number of observations[b]
Satellites				
TROPOMI	daily[c]	7.0 x 7.0	11[d]	567
GeoCARB 2×/day	2× daily[e]	2.7 x 3.0	4.0[f]	7700
GeoCARB 4×/day	4× daily[g]	2.7 x 3.0	4.0	15400
Next generation[h]	hourly[i]	1.3 x 1.3	1.0	164500
Surface sites[j]	hourly[k]	point	1.0	840 – 3360[l]

[a]Dry column mean mixing ratio for the satellite observations, local mixing ratio for the surface observations
[b]One week of clear-sky conditions in the 70×70 km^2 domain
[c]13 local solar time (LT)
[d]Butz et al. (2012)
[e]12 and 16 LT
[f]O'Brien et al. (2016)
[g]10, 12, 14, and 16 LT
[h]Aspirational instrument combining the characteristics of instruments currently at the proposal stage (Fishman et al., 2012; Butz et al., 2015; Xi et al., 2015)
[i]Between 8 and 17 LT
[j] *in situ* measurements of surface air concentrations
[k]Day and night
[l]For 5 to 20 surface sites

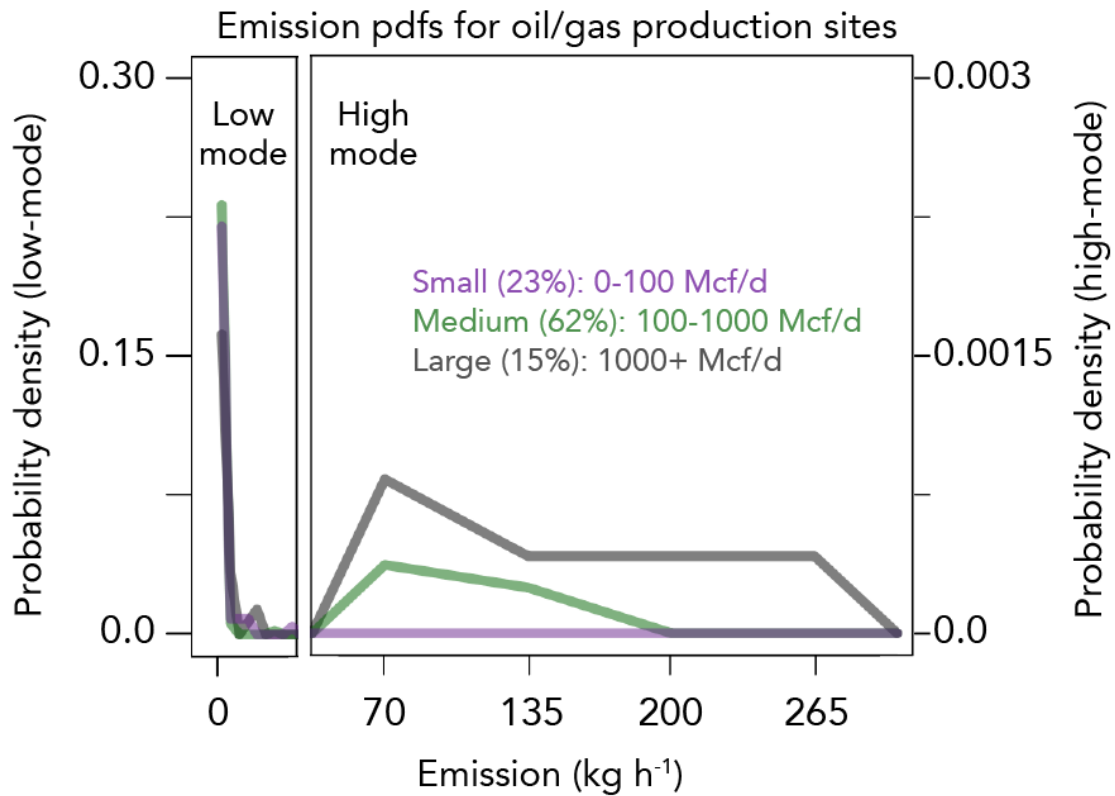
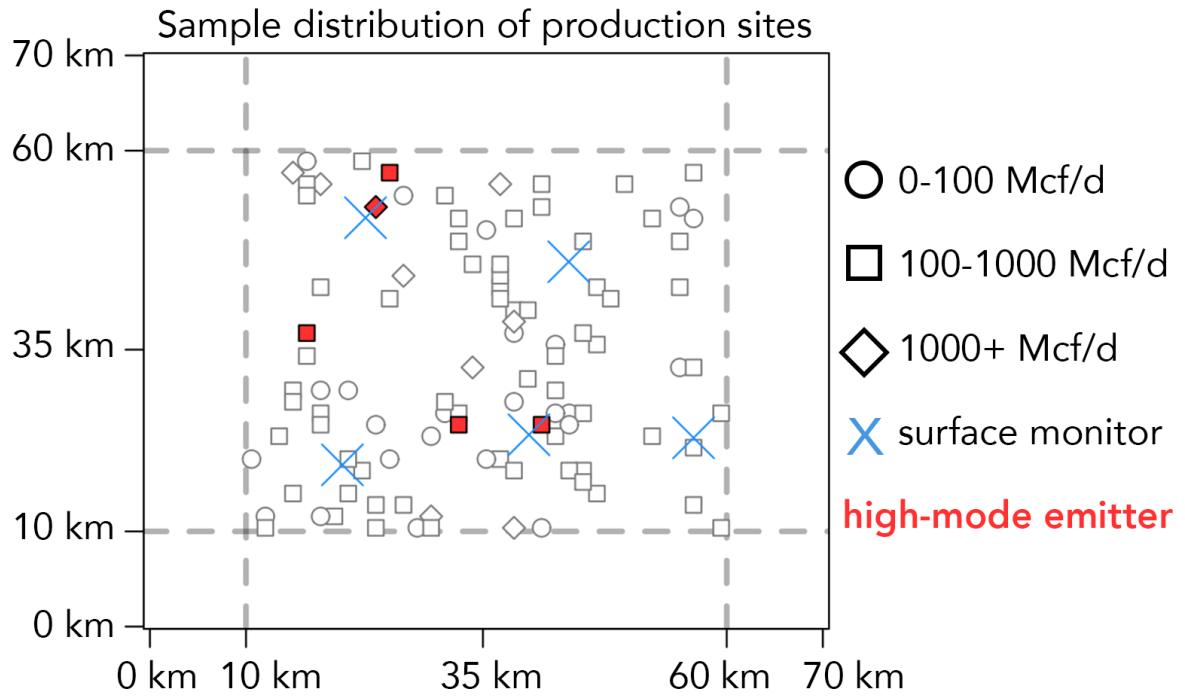



Figure 1. Probability density functions (pdfs) of emissions for oil/gas production sites of different production size categories (small, medium, and large), taken from Barnett Shale observations (Lan et al., 2015; Rella et al., 2015; Yacovitch et al., 2015). Note the difference in y-axis scales between the left (low-mode) and right (high-mode) panels. The axis break at 40 kg h^{-1} represents the threshold for flagging an emitter as high.

5 **Figure 2**. Sample realization of emissions from a hypothetical oil/gas production field with 100 production sites of different production size categories (symbols) within a 50×50 km^2 domain (dashed line). Different production size categories are shown with symbols. Red shading indicates high-mode emitters. Blue symbols mark the locations of five surface air monitoring sites placed according to the k-means algorithm.

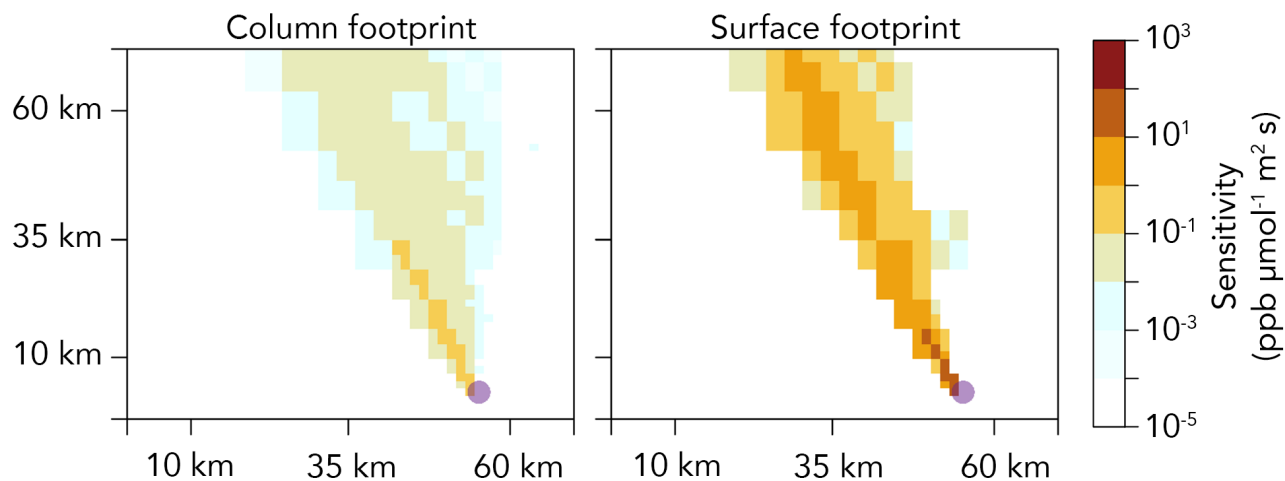



Figure 3: Sample sensitivities of observed atmospheric concentrations (column and surface) to surface emissions upwind, defining the emission footprint for that observation. Values are shown here for a particular observation point (purple dot) and time (October 19, 2013 at 09 LT). Concentrations are in mixing ratio units of ppb (dry column mean mixing ratio for the column) and emissions are in units of $\mu mol \ m^{-2} \ s^{-1}$.

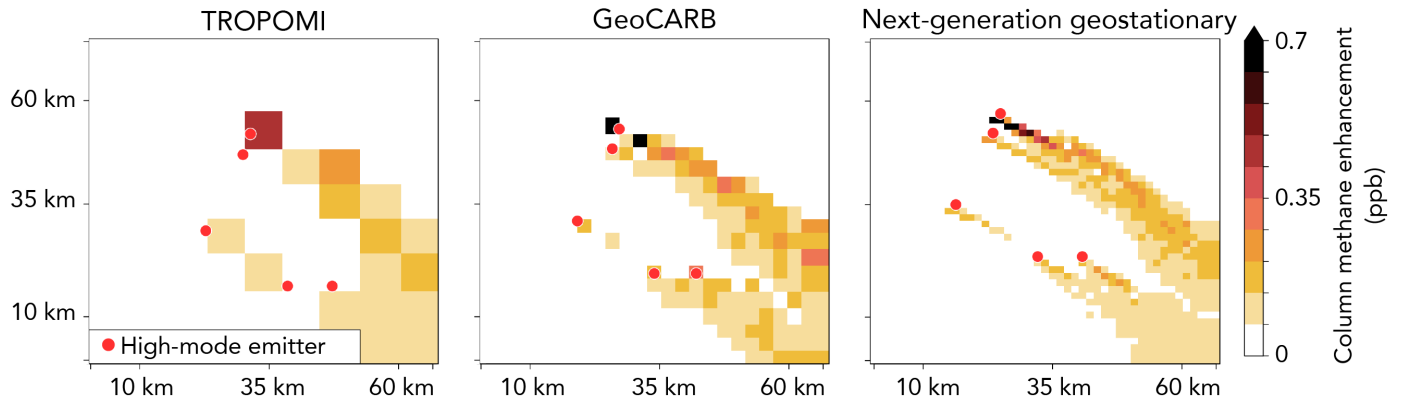



Figure 4: Simulated noiseless methane column enhancement for sampling by single overpasses of TROPOMI, GeoCARB, and a next-generation high-resolution geostationary satellite (Table 1). Emission field is that of Figure 2. The locations of the five high-mode emitters in that field are indicated. Values are for 22 October 2013 at 13 LT.

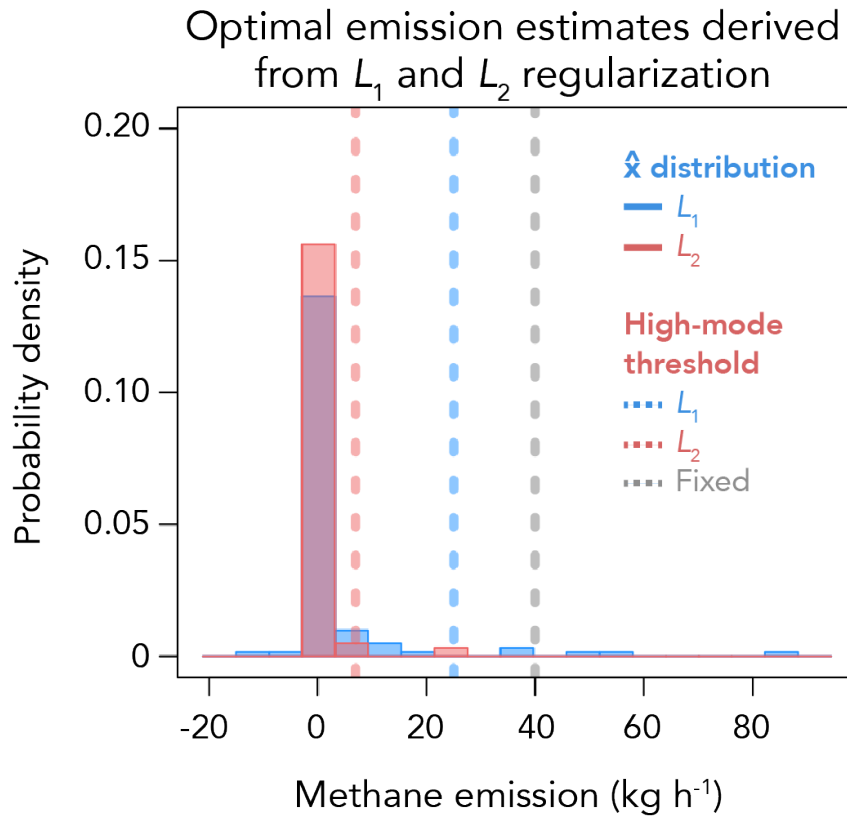



Figure 5: An example distribution of the optimal emission estimate $\hat{x}$ for a realization of the emission inventory (100 sites), GeoCARB 4×/day pseudo-observations, and L_1 or L_2 regularization. Dashed lines represent the thresholds to classify an emitter as high-mode, determined either from the distribution $\hat{x}$ or from a fixed prior value (here 40 kg h^{-1}).

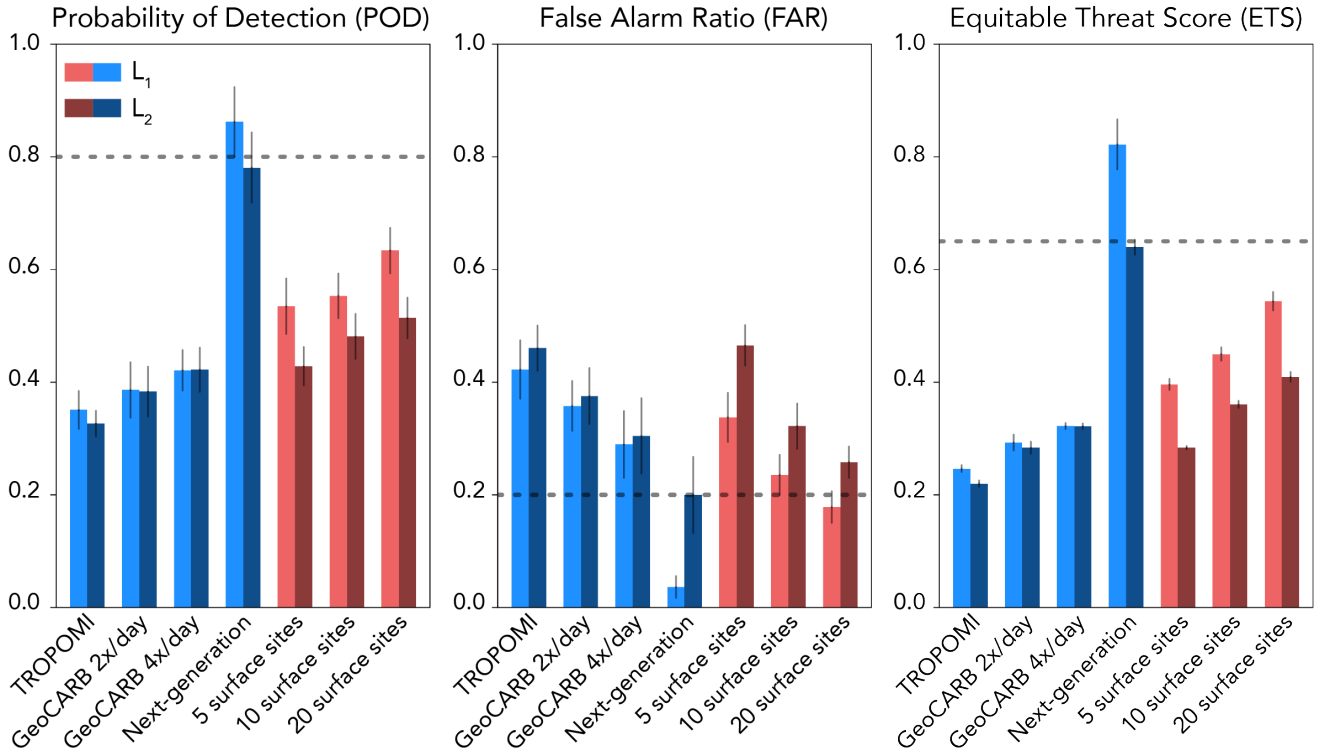



Figure 6. Probability of detection (POD), false alarm ratio (FAR), and Equitable Threat Score (ETS) of high-mode emitters for each satellite and surface observing configuration. Each bar represents the mean of 500 observing system simulation experiments (OSSEs), where 100 production sites in a 50×50 km^2 domain were used to construct 500 random realizations of an emission field including different subsets of high-mode emitters. For each observing configuration, the left bar (lighter color) shows results for the inversion with L_1 regularization, and the right bar (darker color) is for the L_2-regularization. The dashed lines represent the POD, FAR, and ETS criteria for successful observing systems. Here and in following figures, the vertical lines measure the sensitivity to the choice of threshold for diagnosing high-mode emitters in the inversion.

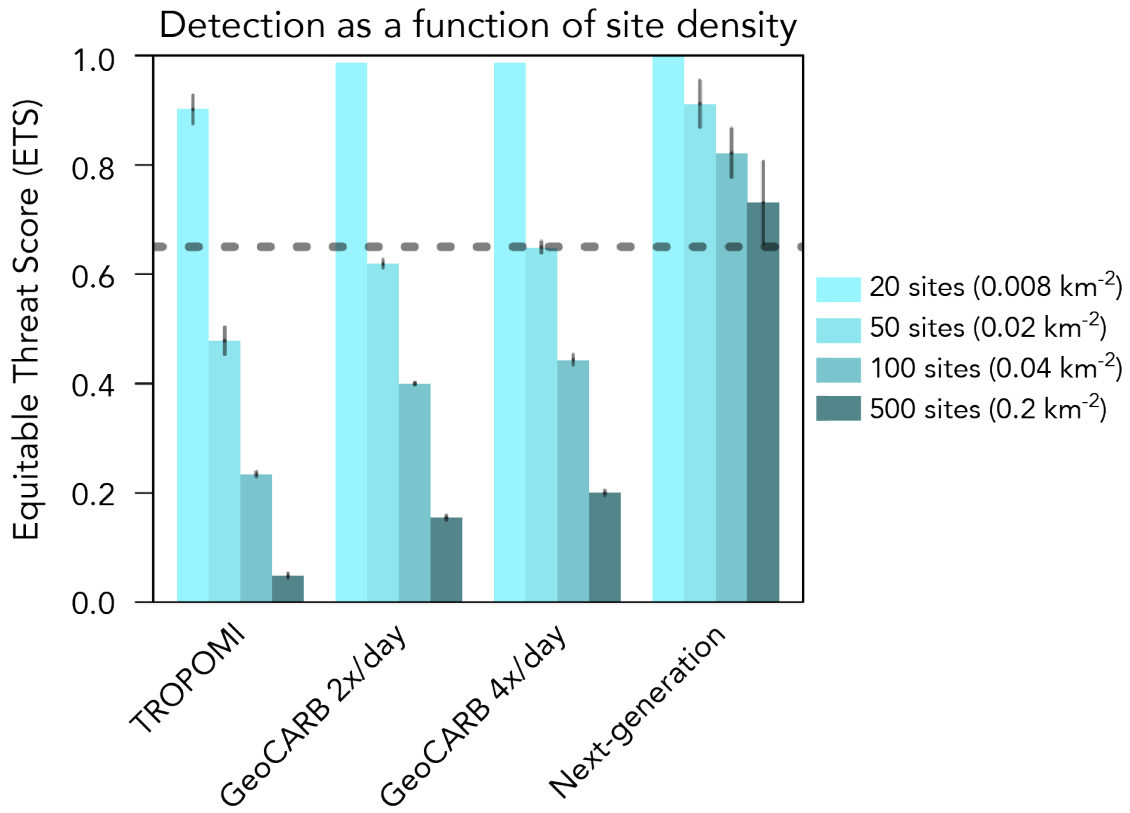



Figure 7. Equitable Threat Score (ETS) for each satellite observing configuration, varying the density of production sites (20-500 sites in 50×50 km^2 domain). Results are from the L_1 inversion. The dashed line represents the ETS criterion for successful observation.

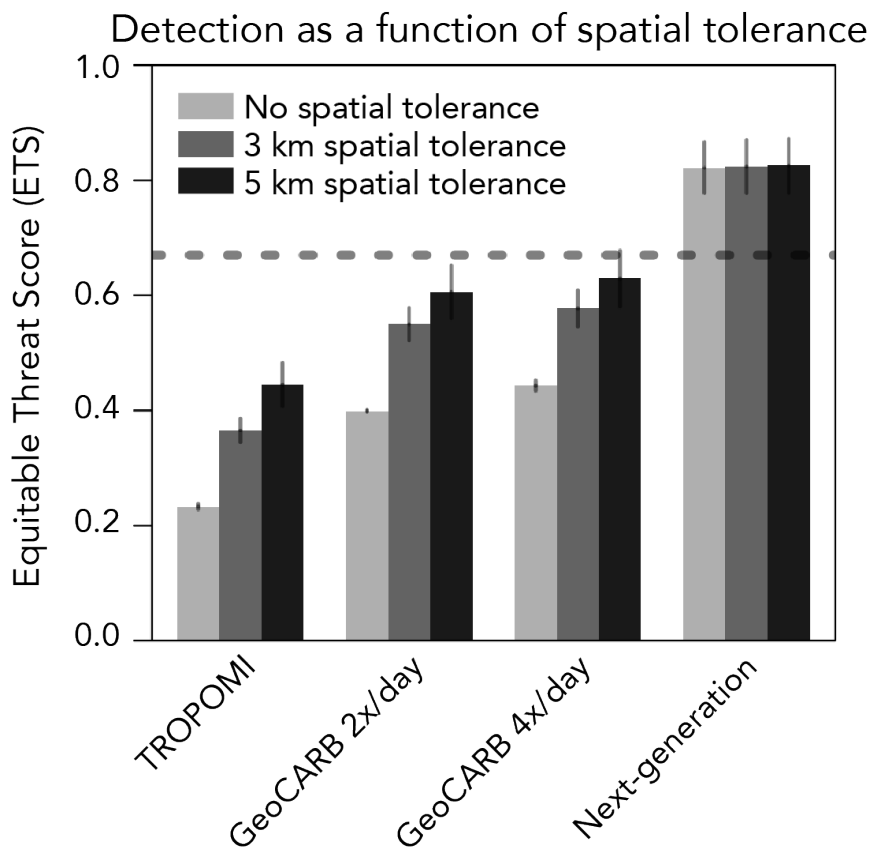



Figure 8: Effect of introducing spatial tolerance in the detection of high-mode emitters. Spatial tolerance is the radius within which a high-mode emitter must be located in order for a prediction to be called true positive (TP). The results are for an emission field with 100 production sites in the 50×50 km^2 domain. Only results from the L_1 inversion method are shown. The dashed line represents the ETS success criterion.

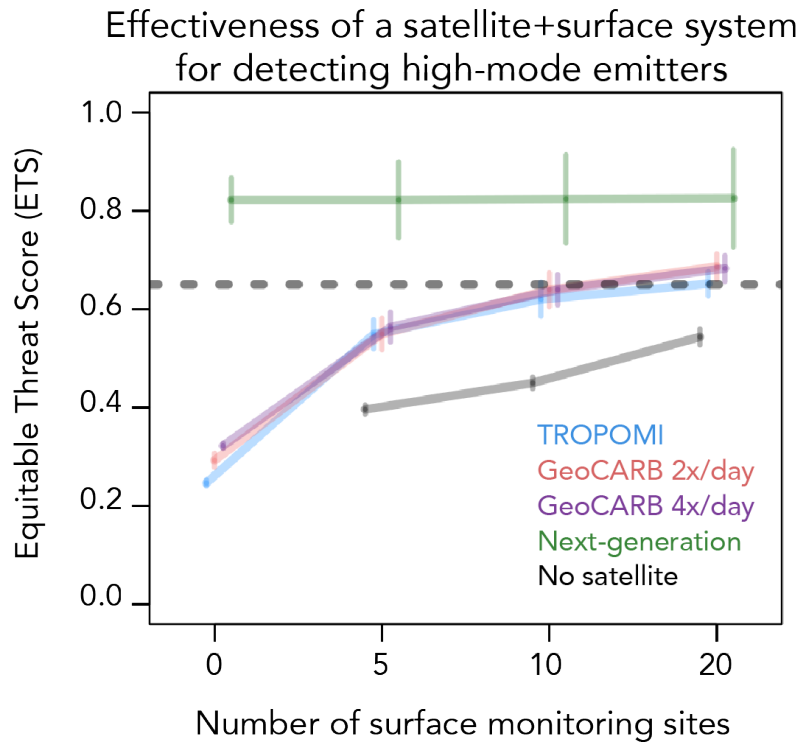



Figure 9: Effectiveness of a combined satellite and surface observing system for detecting high-mode emitters in an oil/gas field of 100 emitters over a 50×50 km^2 domain, as determined from joint inversion of the observations. The dashed line represents the ETS success criterion.