Bayesian hierarchical model for methane emission source apportionment

William Daniels,

Doug Nycka, Dorit Hammerling

Department of Applied Mathematics and Statistics

Colorado School of Mines

August 5, 2024

Joint Statistical Meetings

Recent regulatory push to measure and mitigate methane emissions!

United States

H. R. 5376 (Inflation Reduction Act)

SEC. 136. (a) The Administrator shall impose and collect a fee from the owner or operator of **each applicable facility** that is required to report methane emissions …

SEC. 136. (g)(2) … calculation of fees under subsection (c) of this section, are based on **empirical data** and accurately reflect the total methane emissions from the applicable facilities.

Recent regulatory push to measure and mitigate methane emissions!

United States

H. R. 5376 (Inflation Reduction Act)

SEC. 136. (a) The Administrator shall impose and collect a fee from the owner or operator of **each applicable facility** that is required to report methane emissions …

alculation of fees
of this section,
al data and
he total methane
pplicable

Amendments adopted by the European Parliament on 9 May 2023 on the proposal for a regulation of the European Parliament

… importers must provide a report with the following information for **each site** from which the import to the Union has taken place …

… information specifying the exporter's, or where relevant, the producer's **direct measurements of site-level methane emissions**, conducted by independent service provider …

European Union

Recent regulatory push to measure and mitigate methane emissions!

United States

H. R. 5376 (Inflation Reduction Act)

SEC. 136. (a) The Administrator shall impose and collect a fee from the owner or operator of **each applicable facility** that is required to report methane emissions …

alculation of fees
 of this section,
al data and
he total methane
pplicable

Amendments adopted by the European Parliament on 9 May 2023 on the proposal for a regulation of the European Parliament

orters must provide a report
the following information for
site from which the import to the
 has taken place …

ormation specifying the
ter's, or where relevant, the
cer's **direct measurements of**
level methane emissions,
conducted by independent service provider …

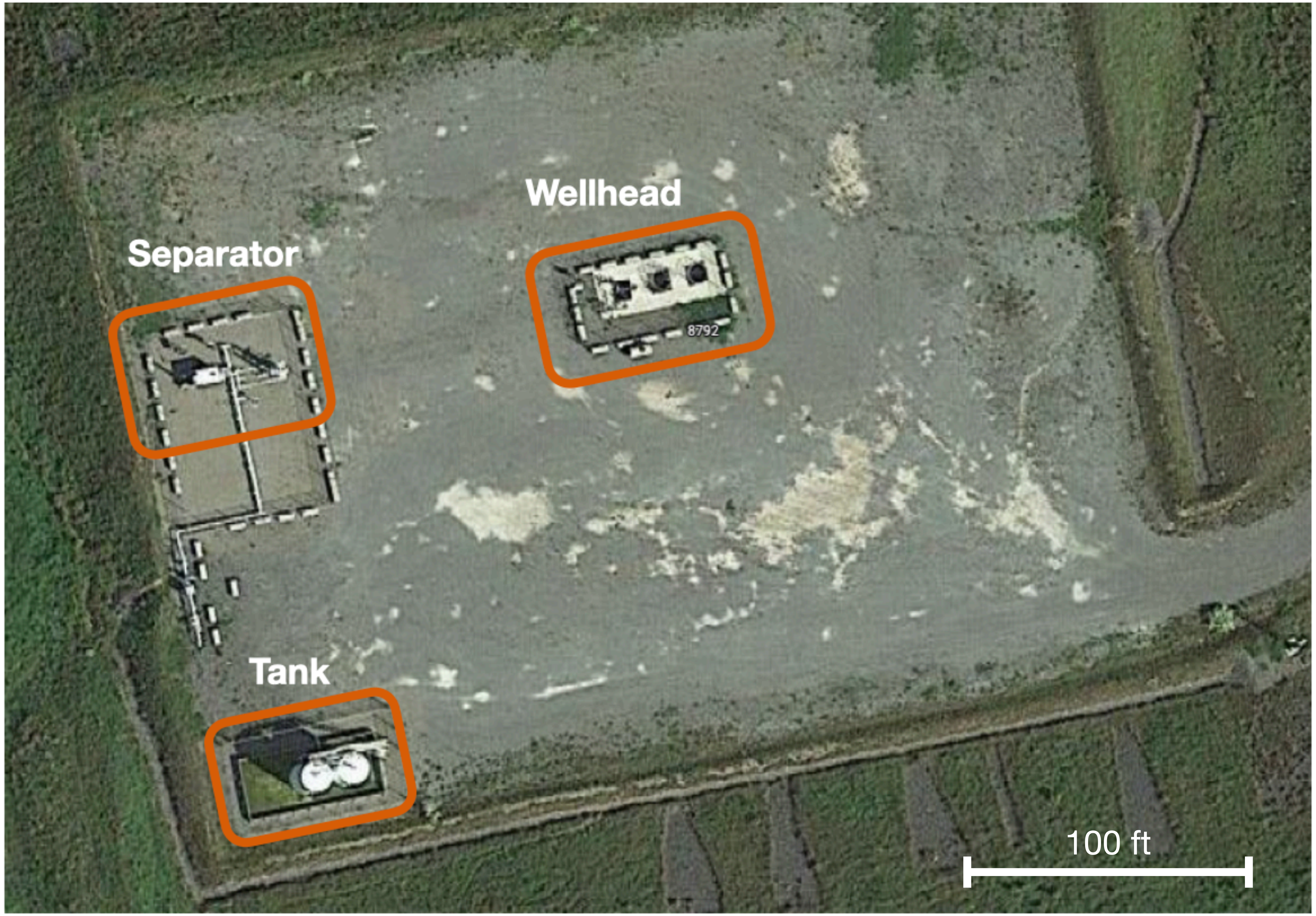
European Union

The Oil & Gas Methane Partnership 2.0 (OGMP 2.0)

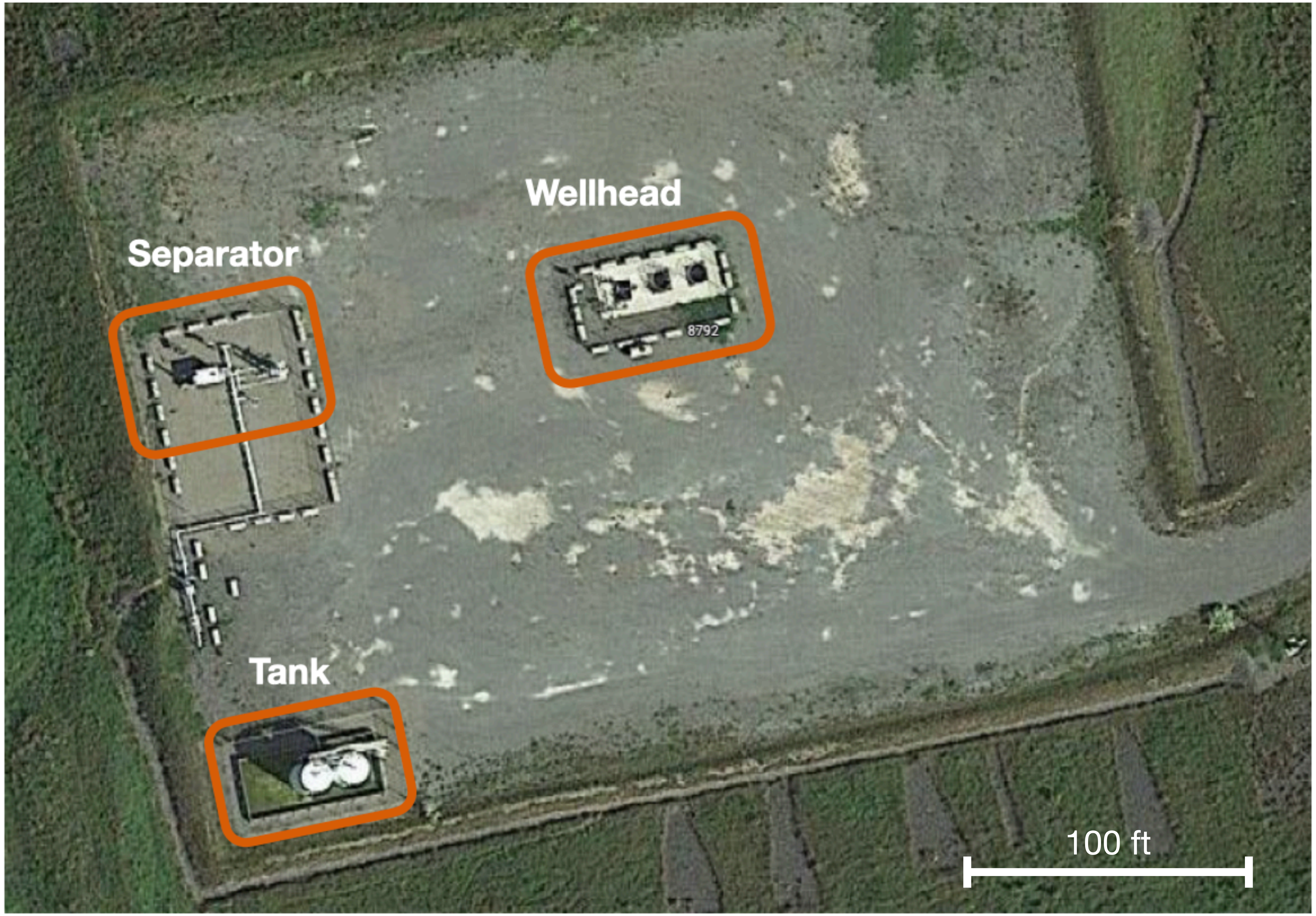
Level 5 — Emissions reported similarly to Level 4, but with the addition of **site-level measurements** (measurements that characterize site-level emissions distribution for a statistically representative population)

Global Initiatives

Example oil and gas site

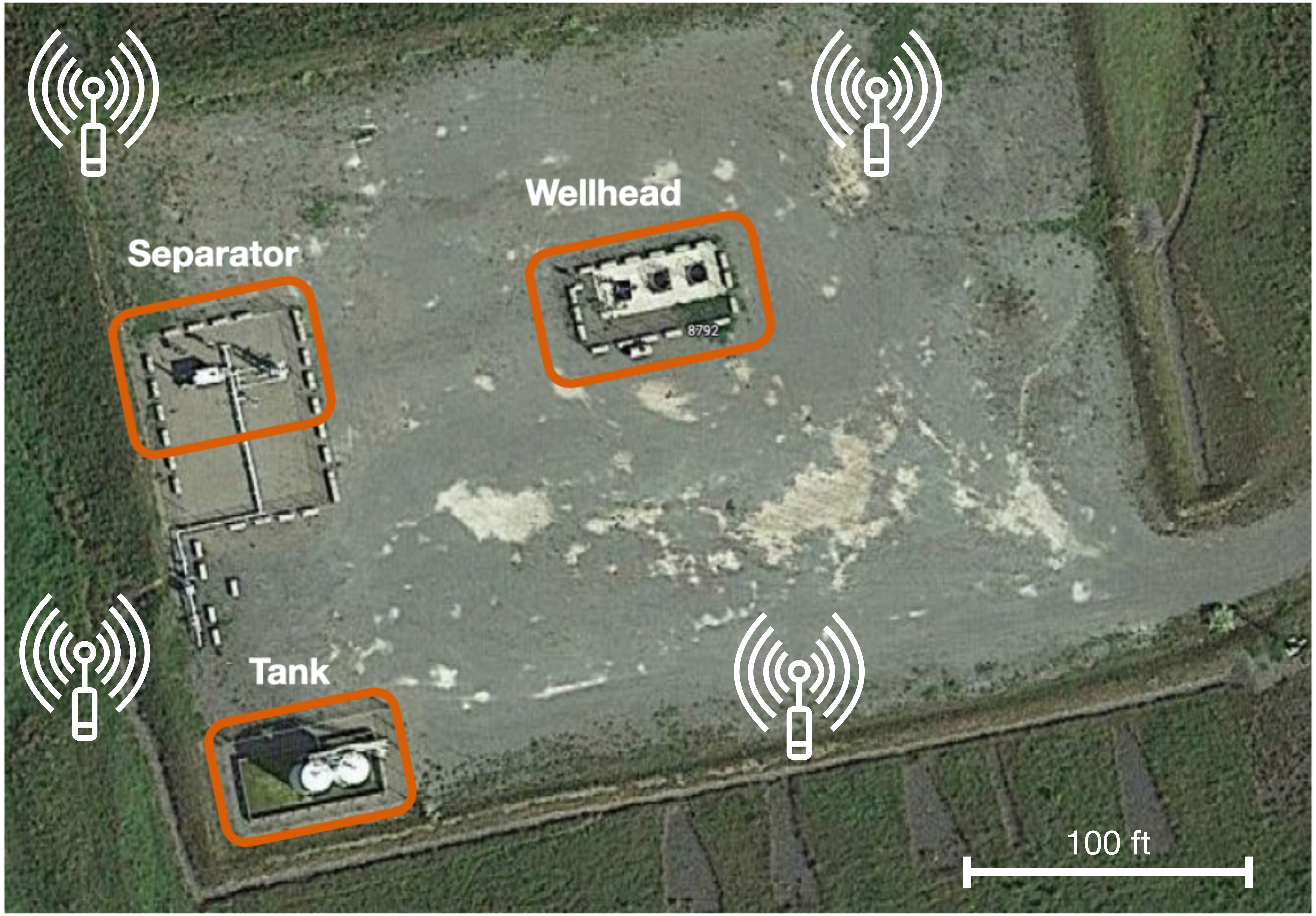


Example oil and gas site



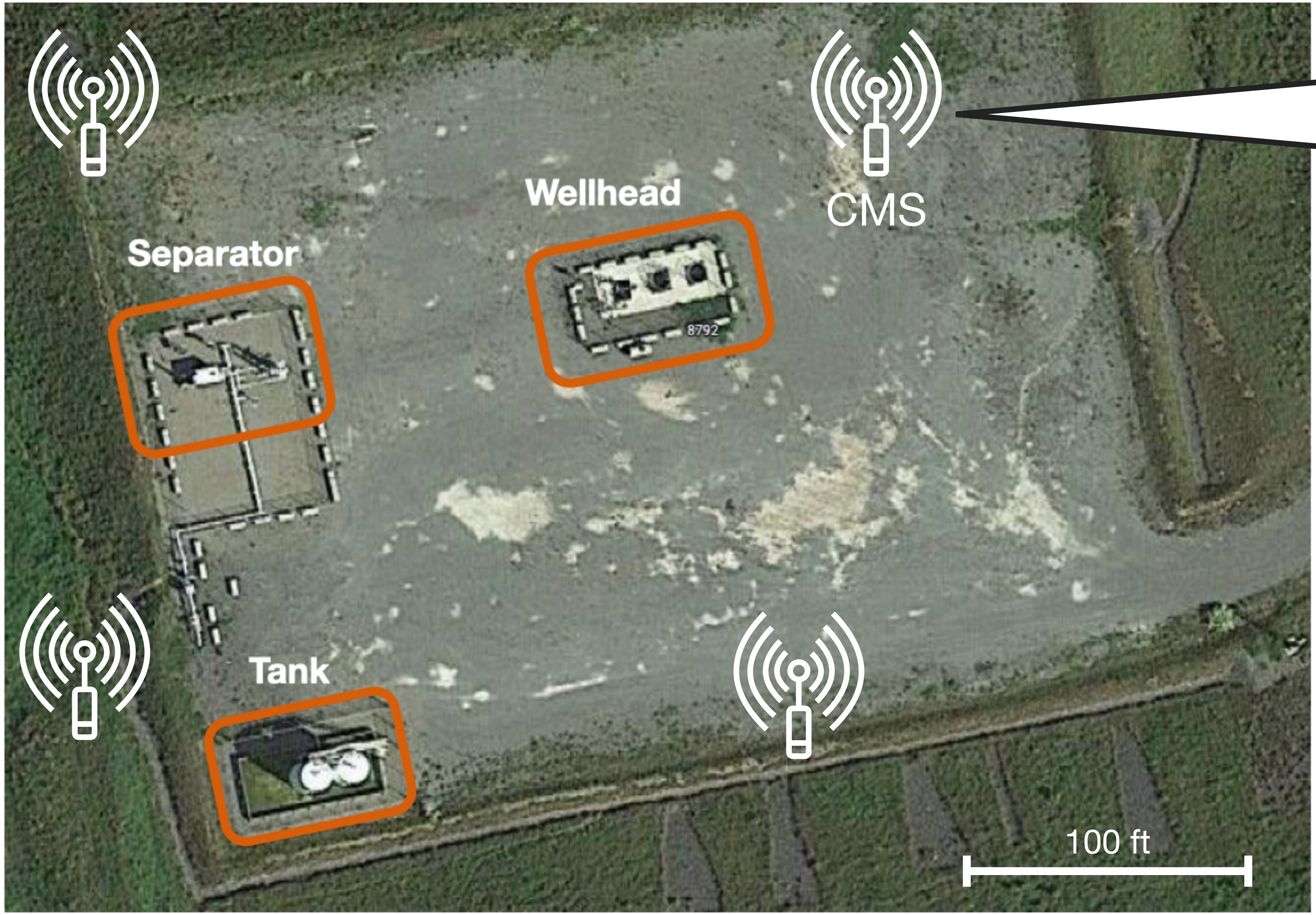
Continuous monitoring system (CMS)

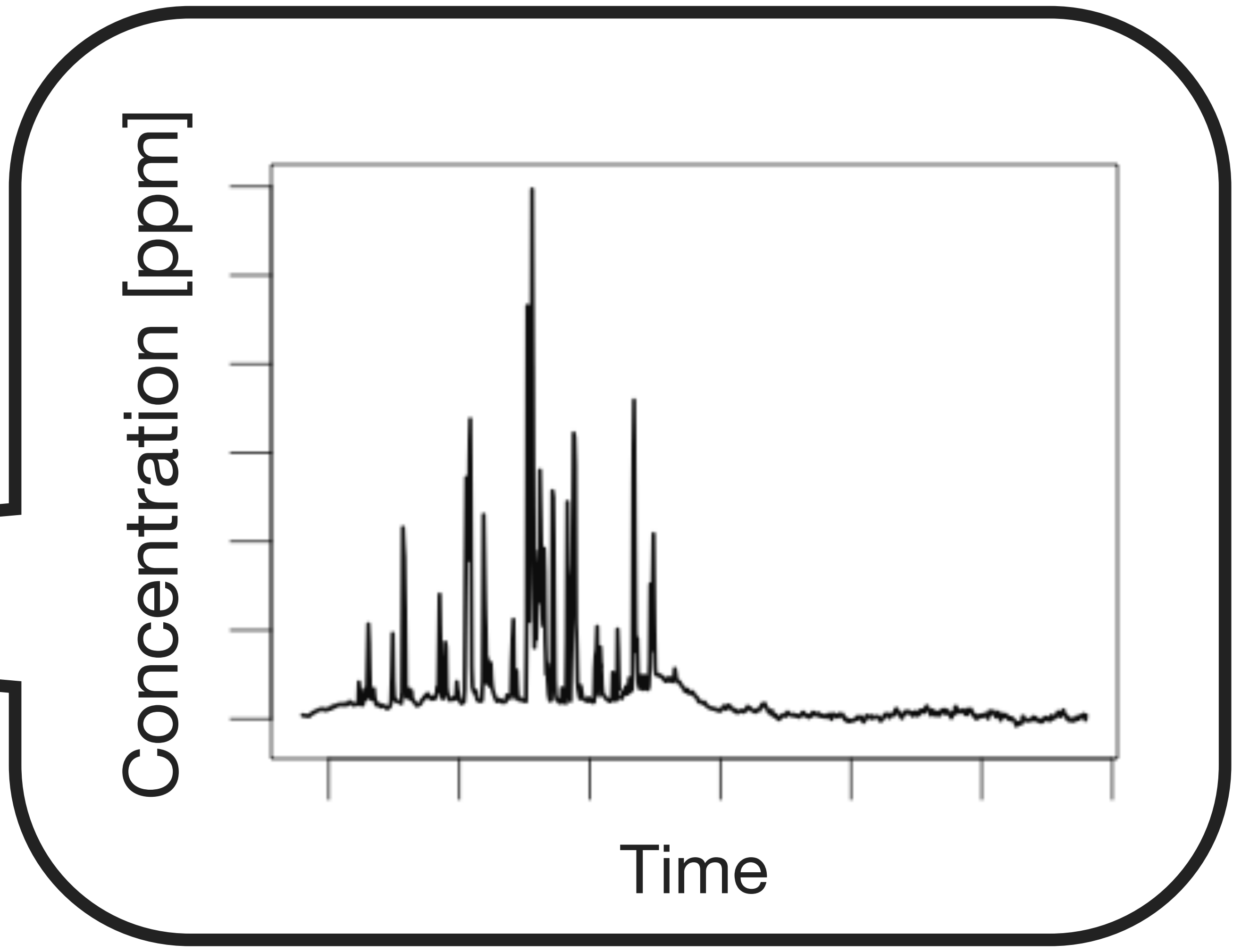
Example oil and gas site



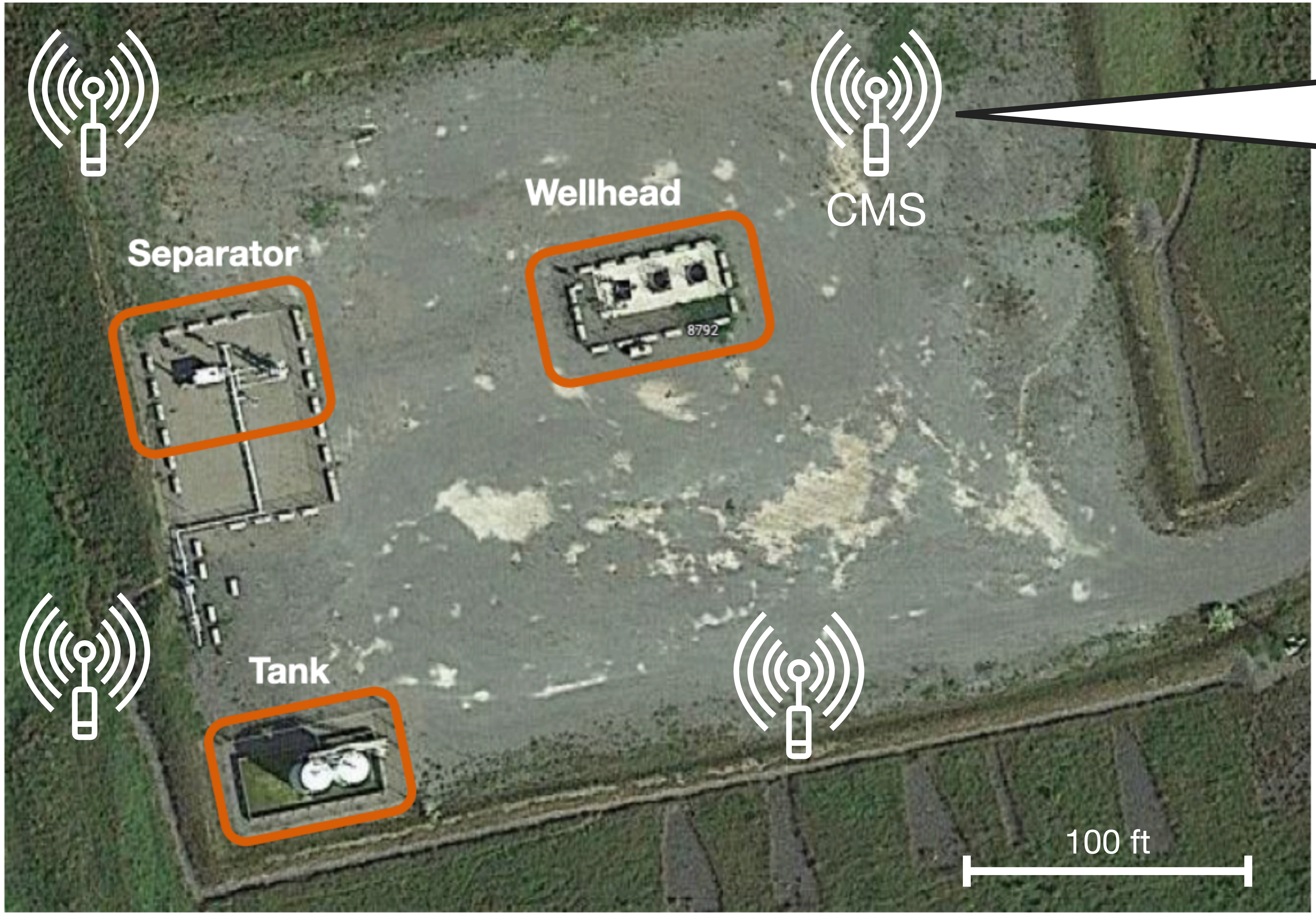
Continuous monitoring system (CMS)

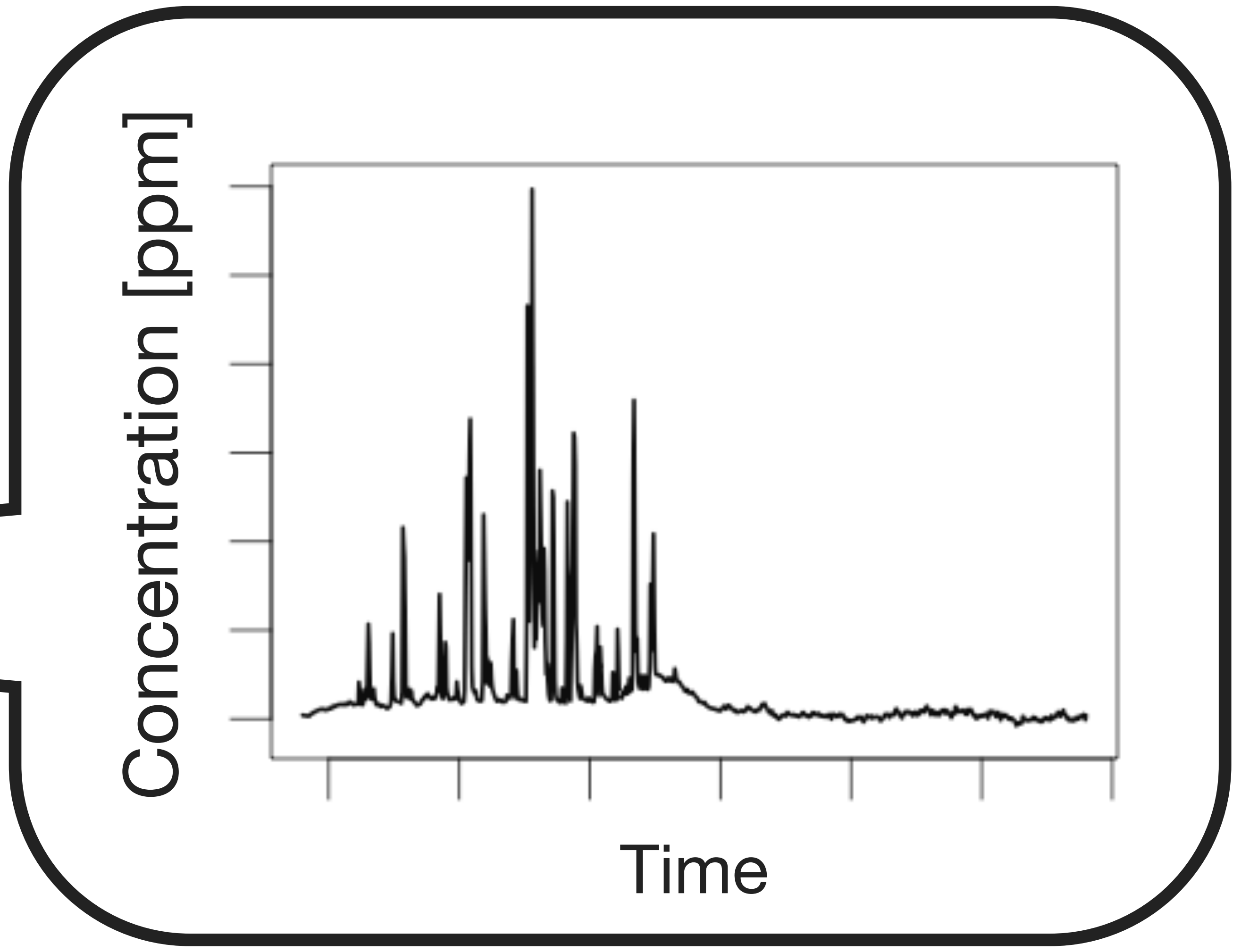
Example oil and gas site





Example oil and gas site





Need an inversion framework to translate raw concentration data into more useful information:

When is a leak happening?

Where is the leak coming from?

How much methane is being emitted?

Model hierarchy

n = number of observations
p = number of potential sources

Assume the standard linear model:

$$y = X\beta + \epsilon$$

$$\epsilon \sim N(0,\sigma^2)$$

$$y \in \mathbb{R}^{n\times 1}, X \in \mathbb{R}^{n\times p}, \beta \in \mathbb{R}^{p\times 1}$$

Concentration observations from CMS sensors

Simulated concentrations from forward model, with each column assuming a different source

Emission rates for each source

This gives us: $y \sim N(X\beta, I\sigma^2)$

Gaussian puff atmospheric dispersion model

Total volume of methane contained in puff p

Total concentration at (x,y,z,t)

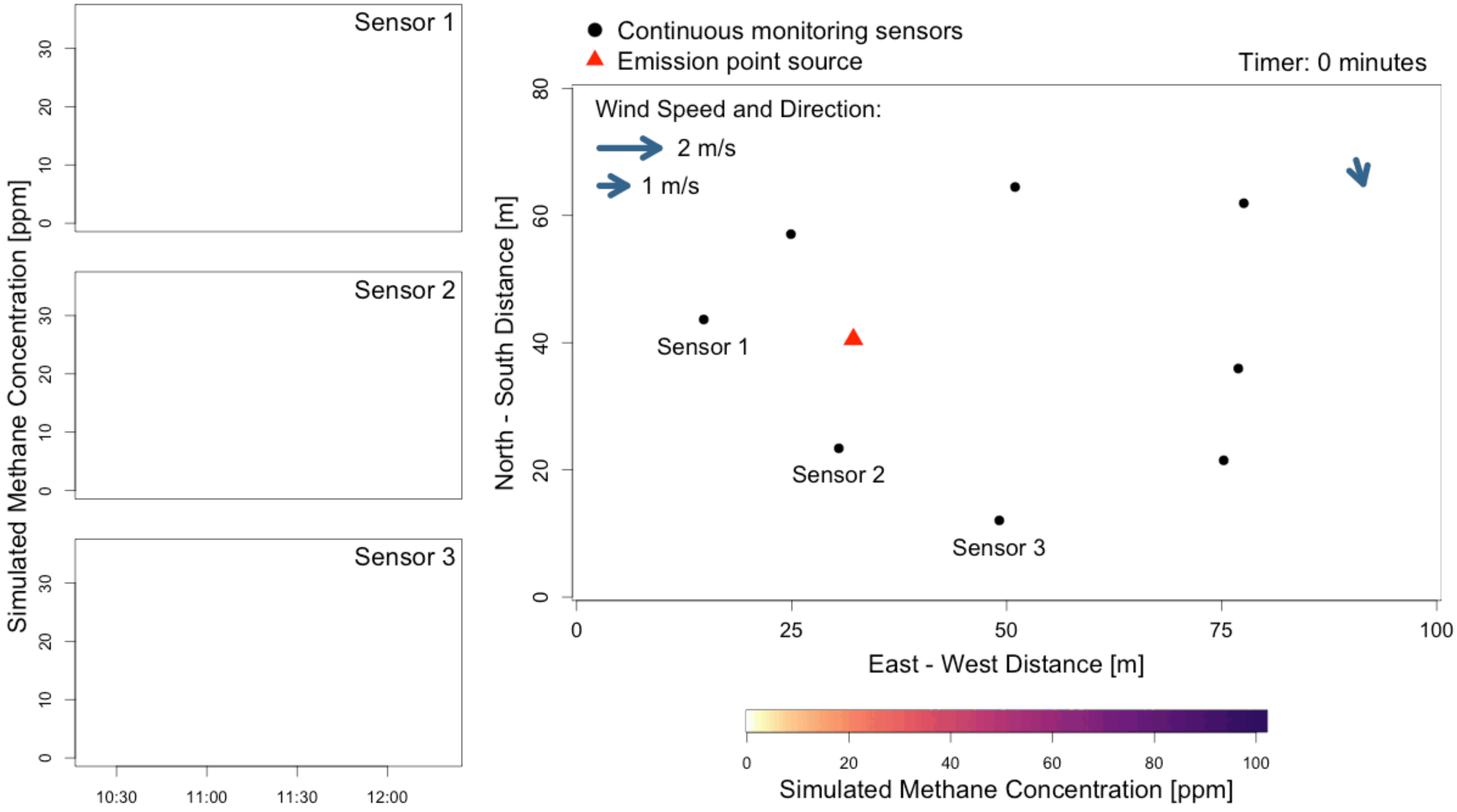
$$c(x, y, z, t) = \sum_{p=1}^{P} c_p(x, y, z, t)$$

$$c_p(x, y, z, t) = \frac{Q}{(2\pi)^{3/2}\sigma_y^2\sigma_z} \exp\left(-\frac{(x-ut)^2+y^2}{2\sigma_y^2}\right)\left[\exp\left(-\frac{(z-H)^2}{2\sigma_z^2}\right)+\exp\left(-\frac{(z+H)^2}{2\sigma_z^2}\right)\right]$$

Concentration contribution of puff p

Decay in puff concentration in horizontal plane (x,y)

Decay in puff concentration in vertical dimension (z)


Sensor 1
Sensor 2
Sensor 3
Simulated Methane Concentration [ppm]
10:30
11:00
11:30
12:00
Continuous monitoring sensors
Emission point source
Timer: 0 minutes
Wind Speed and Direction:
2 m/s
1 m/s
North - South Distance [m]
East - West Distance [m]
Simulated Methane Concentration [ppm]

Repeat this for all other potential sources!

Model hierarchy

n = number of observations
p = number of potential sources

Assume the standard linear model:

$$y = X\beta + \epsilon$$

$$\epsilon \sim N(0,\sigma^2)$$

$$y \in \mathbb{R}^{n\times 1}, X \in \mathbb{R}^{n\times p}, \beta \in \mathbb{R}^{p\times 1}$$

Concentration observations from CMS sensors

Simulated concentrations from forward model, with each column assuming a different source

Emission rates for each source

This gives us: $y \sim N(X\beta, I\sigma^2)$

Model hierarchy

n = number of observations
p = number of potential sources

Assume the standard linear model:

$$y = X\beta + \epsilon$$

$$\epsilon \sim N(0,\sigma^2)$$

Create the following prior structure

Spike-and-slab prior drives estimates to identically zero

Separate probability of emission for each potential emission source

$$\beta_i \sim \begin{cases} 0 & z_i = 0 \\ \mathrm{Exp}(\tau_i^2\sigma^2) & z_i = 1 \end{cases}$$

$$z_i \sim \mathrm{Bernoulli}(\theta_i)$$

$$\theta_i \sim \mathrm{Beta}(a_i, b_i)$$

$$\sigma^2 \sim \text{Inv-Gamma}(\alpha_1, \alpha_2)$$

$$\tau_i^2 \sim \text{Inv-Gamma}(c_i, d_i)$$

Exponential component is non-negative

a_i, b_i, c_i, d_i can contain operator insight

Use a Gibbs sampler to sample from the posterior

Just need to derive all of the necessary conditionals

$$\sigma^2|\xi = \sigma^2|y,\beta \sim \text{Inv-Gamma}\left(\alpha_1 + \frac{n}{2}, \alpha_2 + \frac{(y-X\beta)^T(y-X\beta)}{2}\right)$$

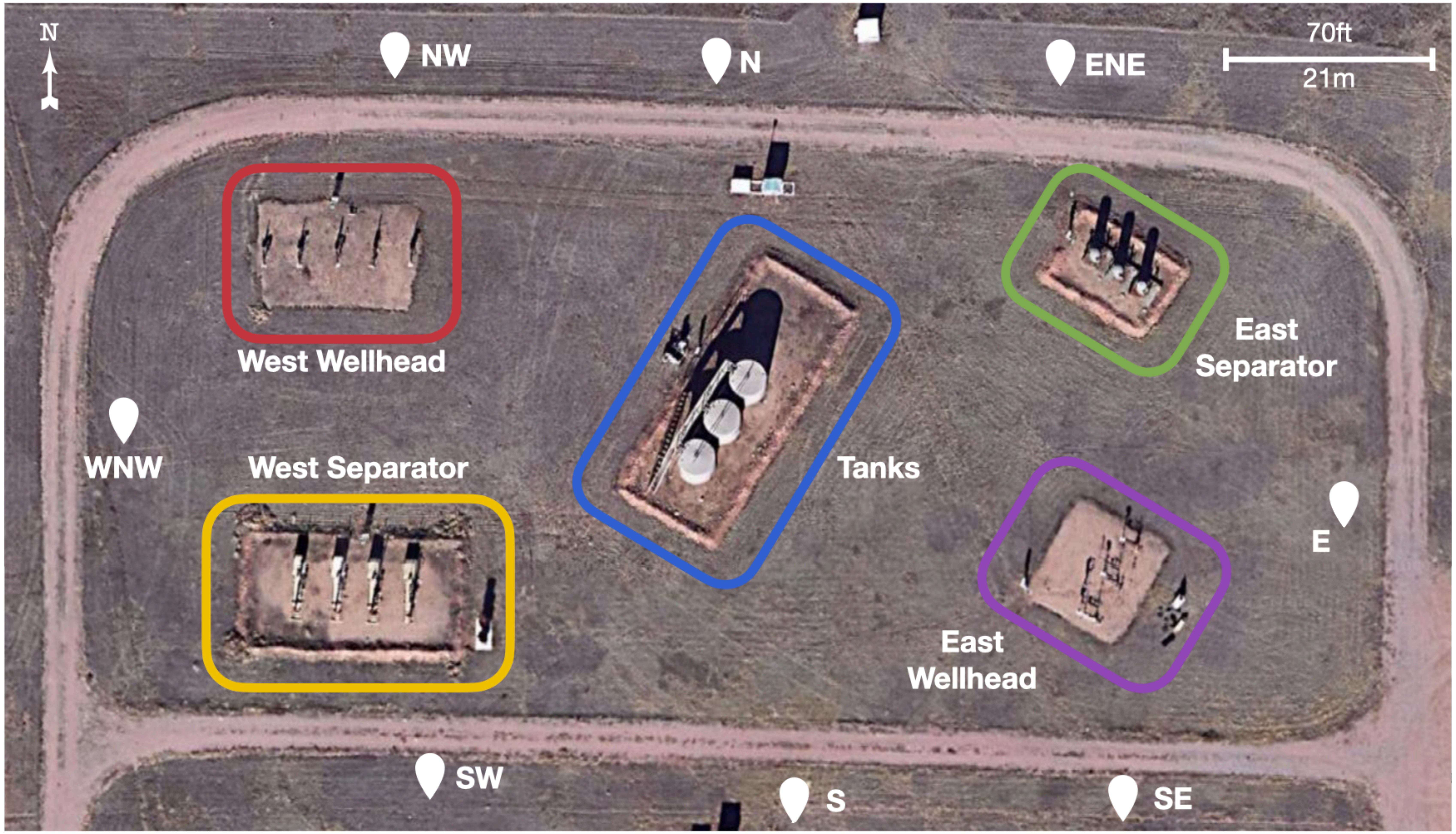
$$\theta_i|\xi = \theta_i|z_i \sim \text{Beta}(z_i + a_i, 1 - z_i + b_i)$$

$$\tau_i^2|\xi = \tau_i^2|\beta_i, z_i \sim \begin{cases} \text{Inv-Gamma}(c_i, d_i) & z_i = 0 \\ \text{Inv-Gamma}\left(1 + c_i, \frac{\beta_i}{\sigma^2} + d_i\right) & z_i = 1 \end{cases}$$

$$\beta_i|\xi = \beta_i|y, \beta_{-i}, \sigma^2, \tau_i^2, z_i \sim \begin{cases} 0 & z_i = 0 \\ \mathcal{N}\left((\frac{X^TX}{\sigma^2})^{-1}(\frac{X^Ty}{\sigma^2} - \frac{e_i}{\tau_i^2\sigma^2}), (\frac{X^TX}{\sigma^2})^{-1}\right) & z_i = 1 \end{cases}$$

$$z_i|\xi = z_i|y, z_{-i}, \beta_{-i}, \sigma^2, \tau^2, \theta \sim \text{Bernoulli}\left(\frac{(1-\theta_i)}{(1-\theta_i) + \frac{\theta_i}{2\tau_i^2\sigma^2}\exp\left(\frac{(x_i^Tw - (1/\tau_i^2))^2}{2\sigma^2 x_i^T x_i}\right)\left(\frac{2\pi\sigma^2}{x_i^T x_i}\right)^{1/2}}\right)$$

Model evaluation on multi-source controlled release data



Methane Emissions Technology Evaluation Center (METEC)

87 multi-source releases

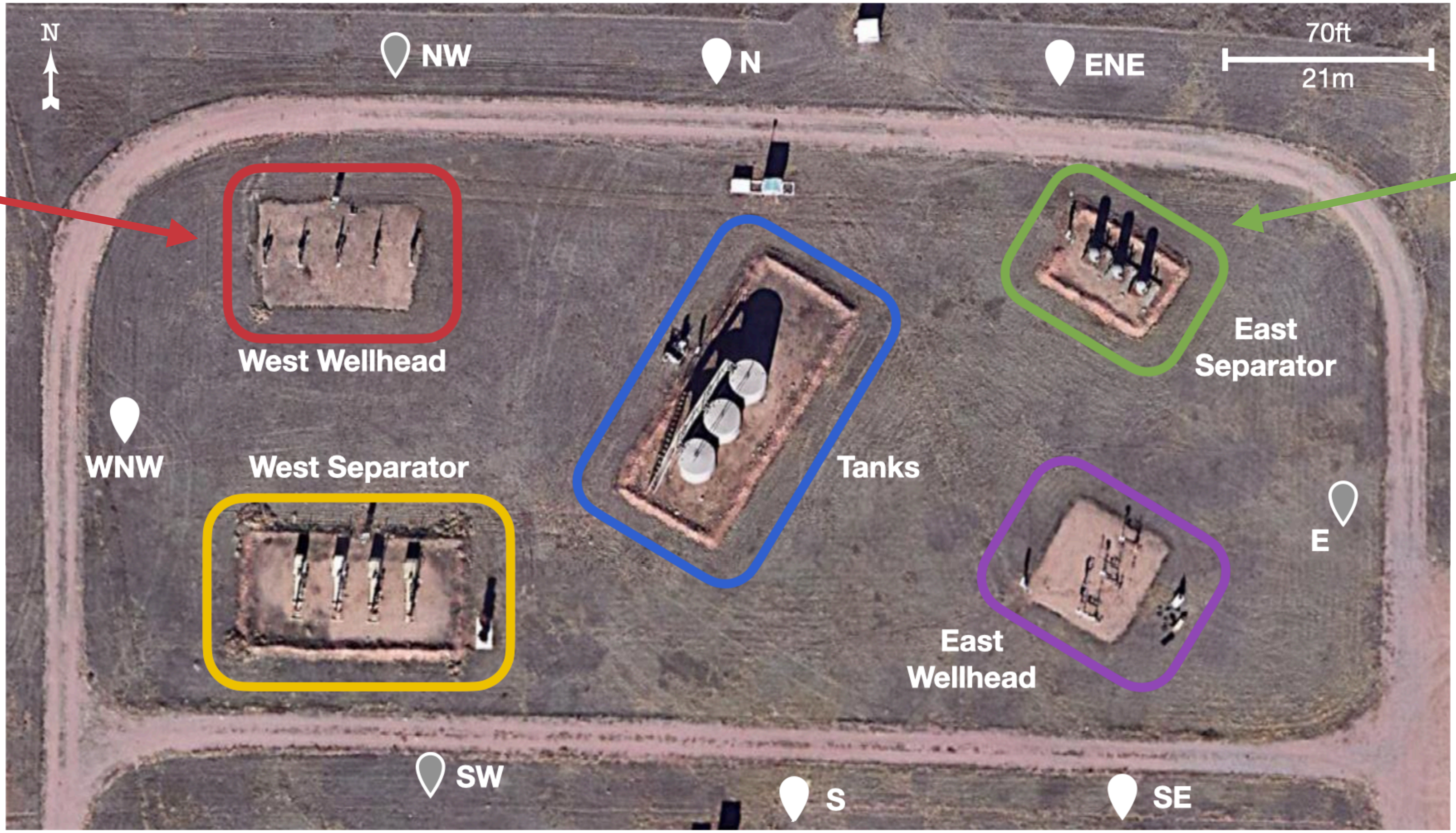
109 single-source releases

196 releases total

Model evaluation on multi-source controlled release data

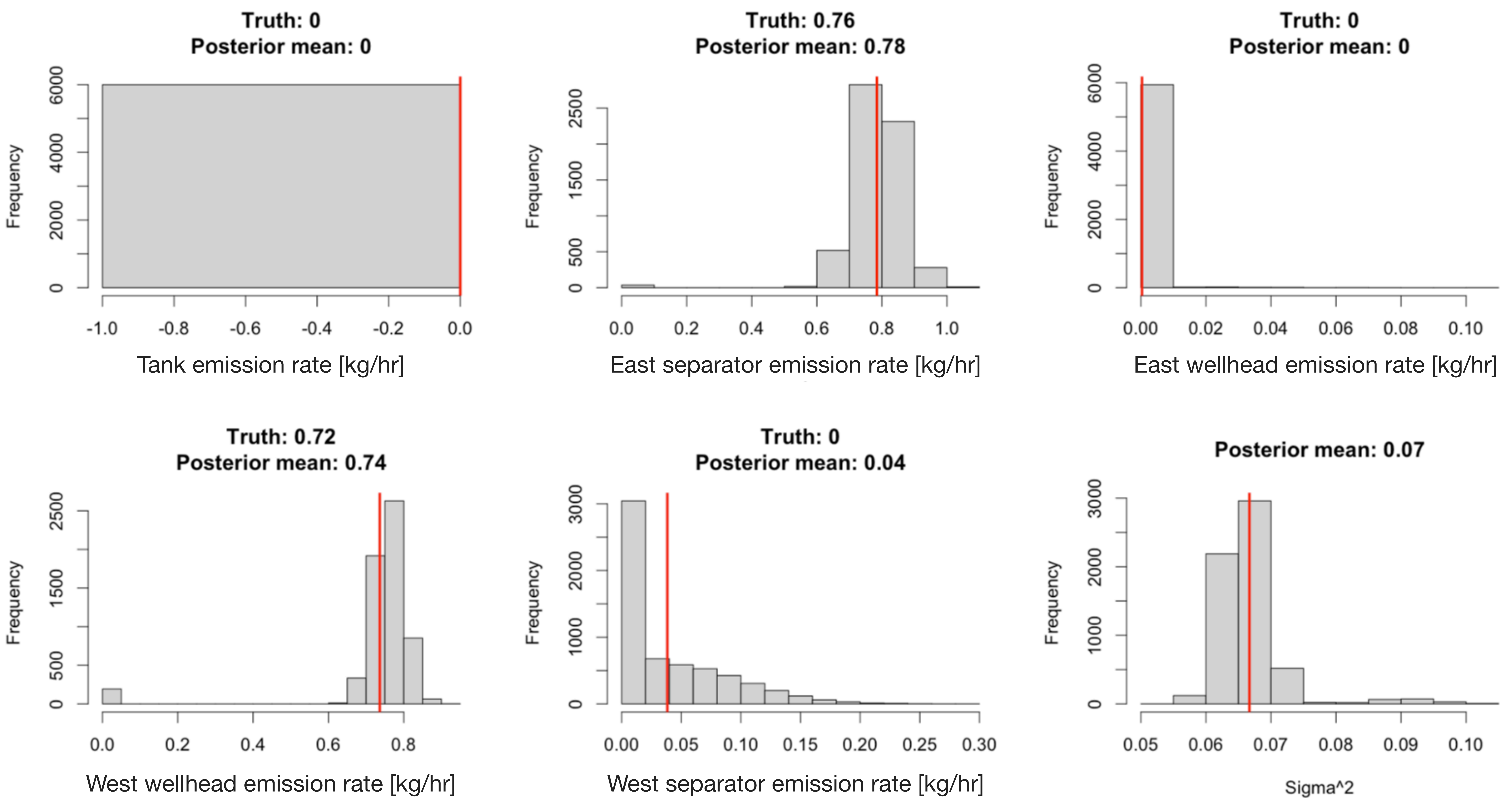
Example two source release

West wellhead emission of 0.72 kg/hr

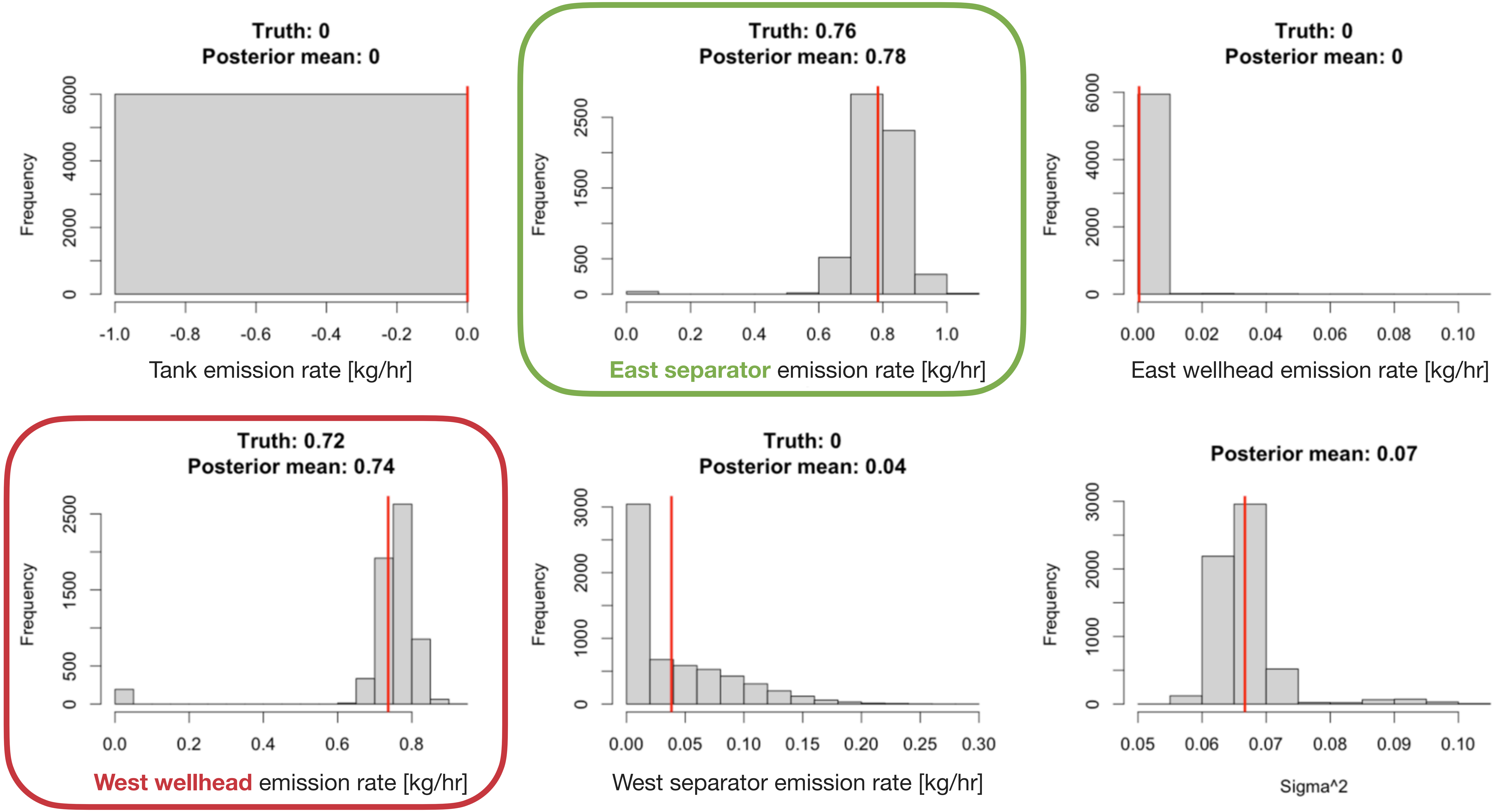


East separator emission of 0.76 kg/hr

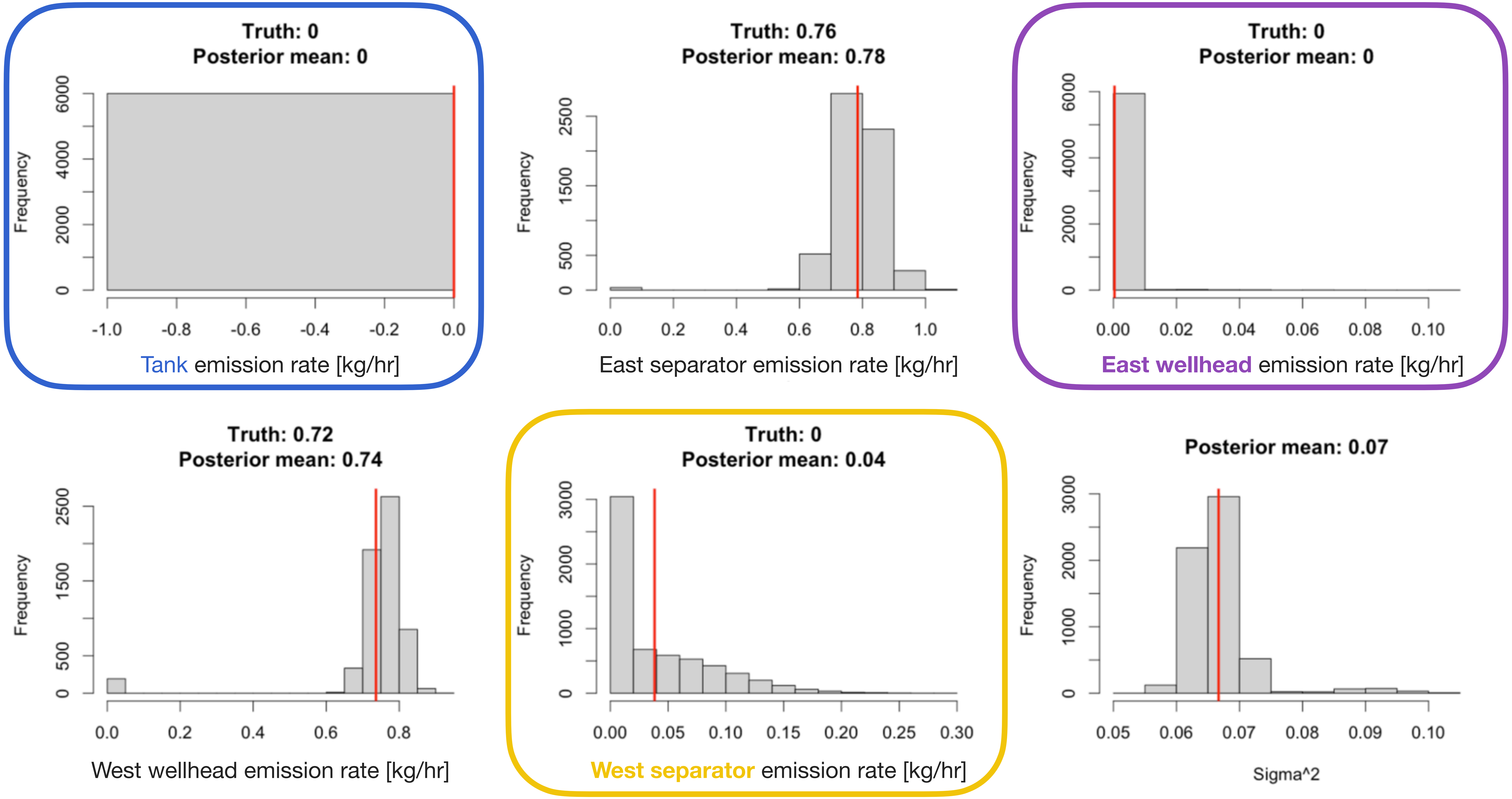
Model evaluation on multi-source controlled release data



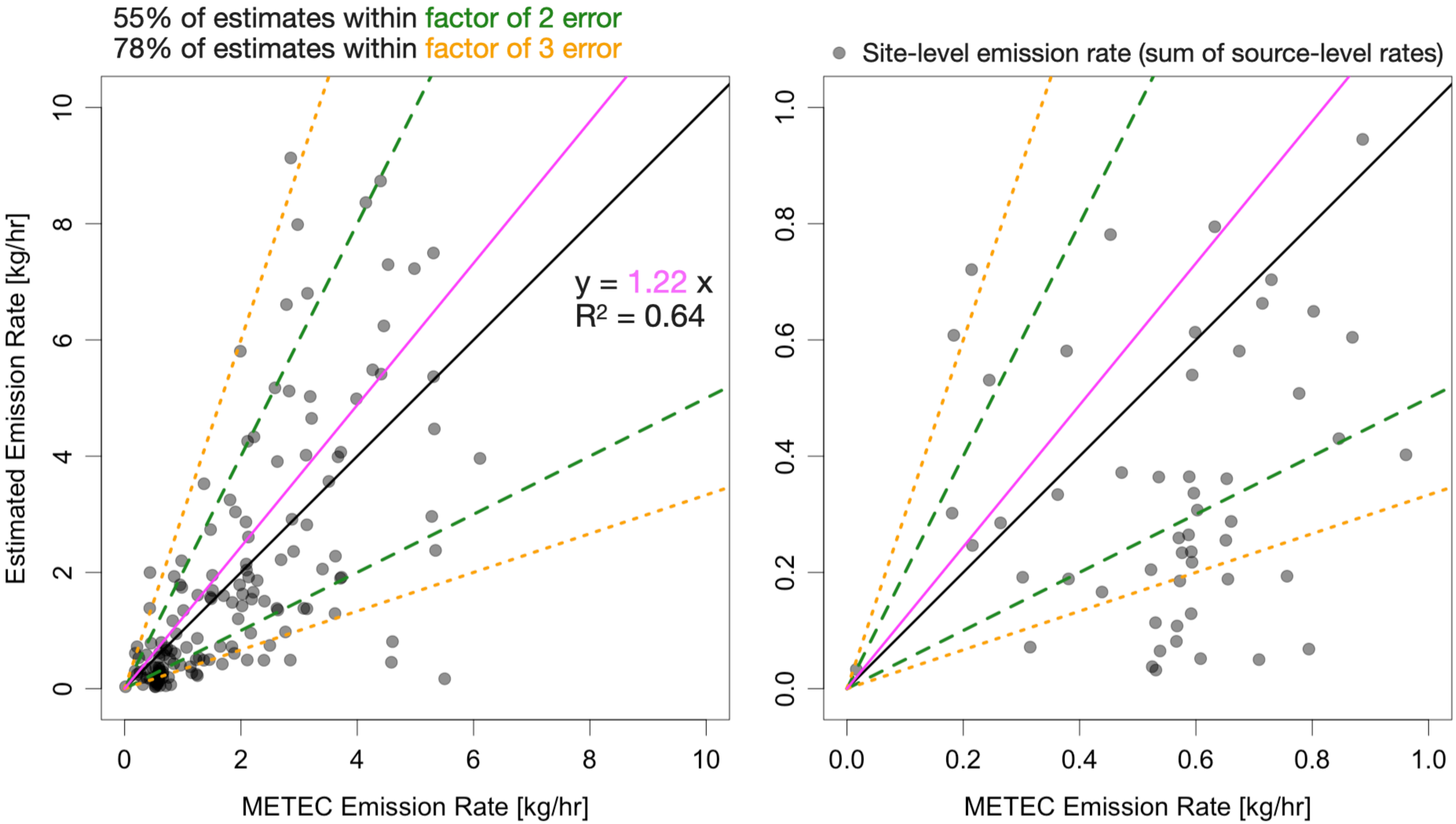
Model evaluation on multi-source controlled release data



Model evaluation on multi-source controlled release data



Model evaluation on multi-source controlled release data



Thank you! Questions?

wdaniels@mines.edu

COLORADO SCHOOL OF MINES

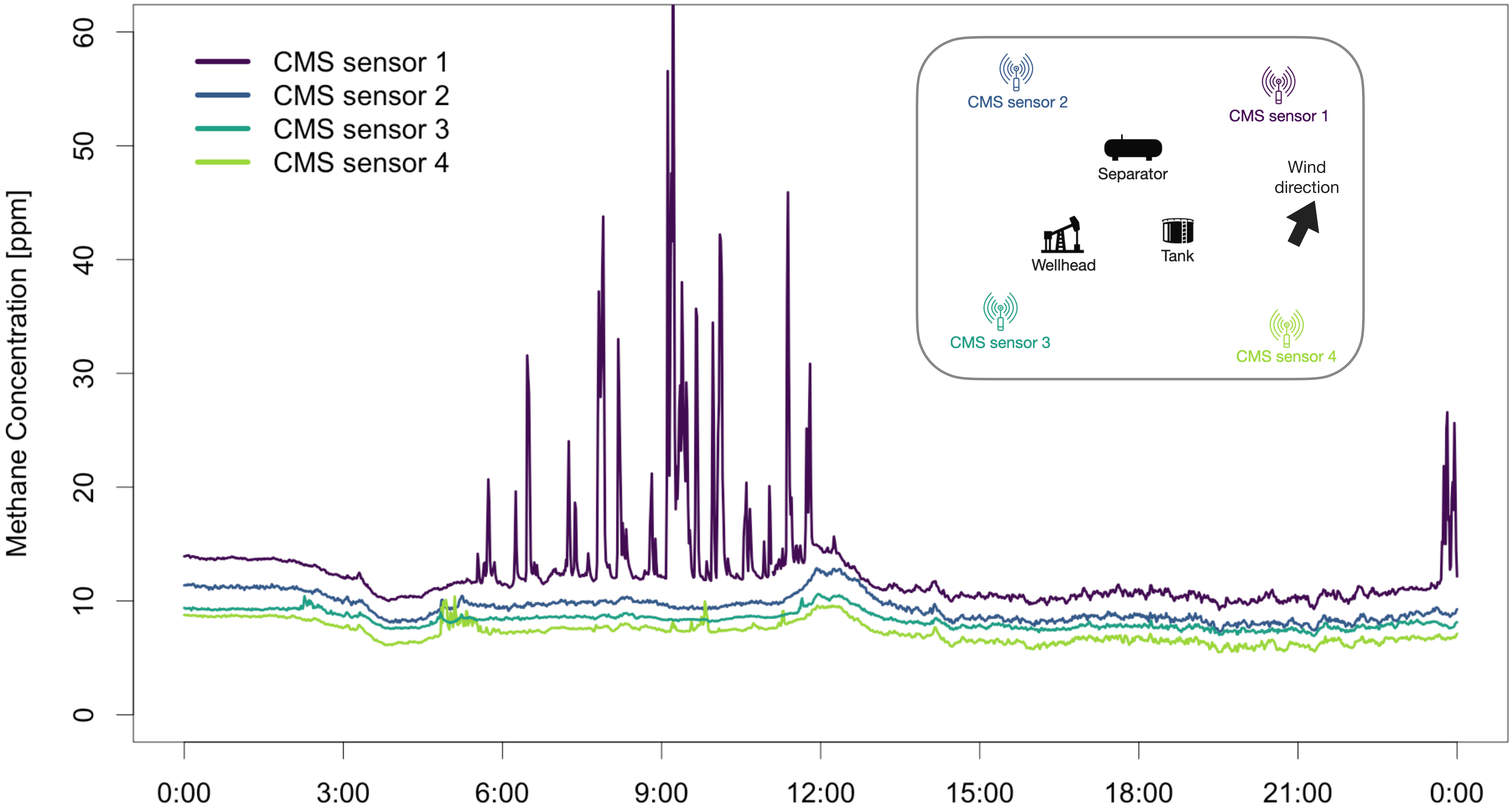
TEXAS
The University of Texas at Austin

COLORADO STATE UNIVERSITY

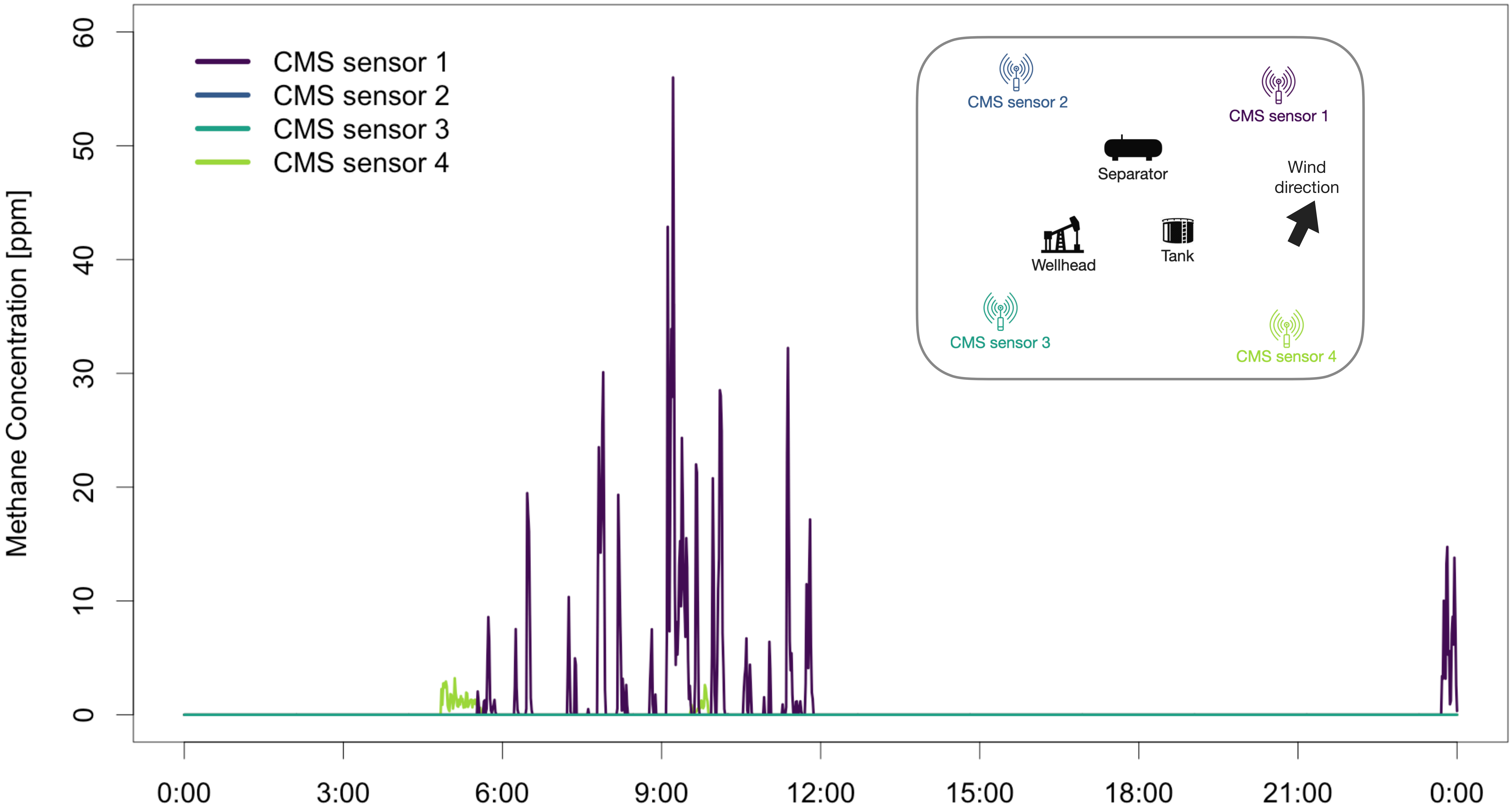
EEMDL
Energy Emissions Modeling and Data Lab

The Payne Institute for Public Policy

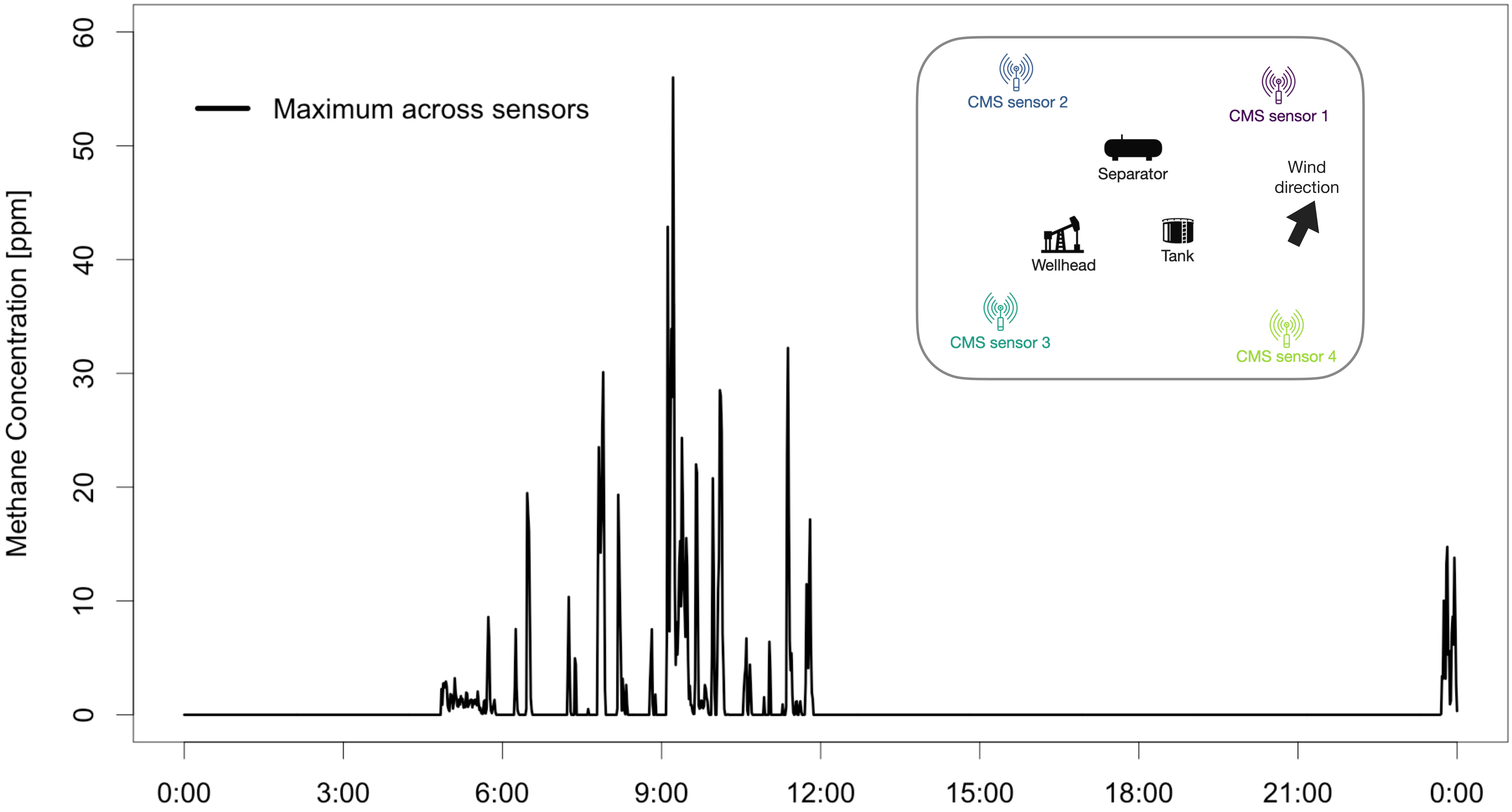
Backup

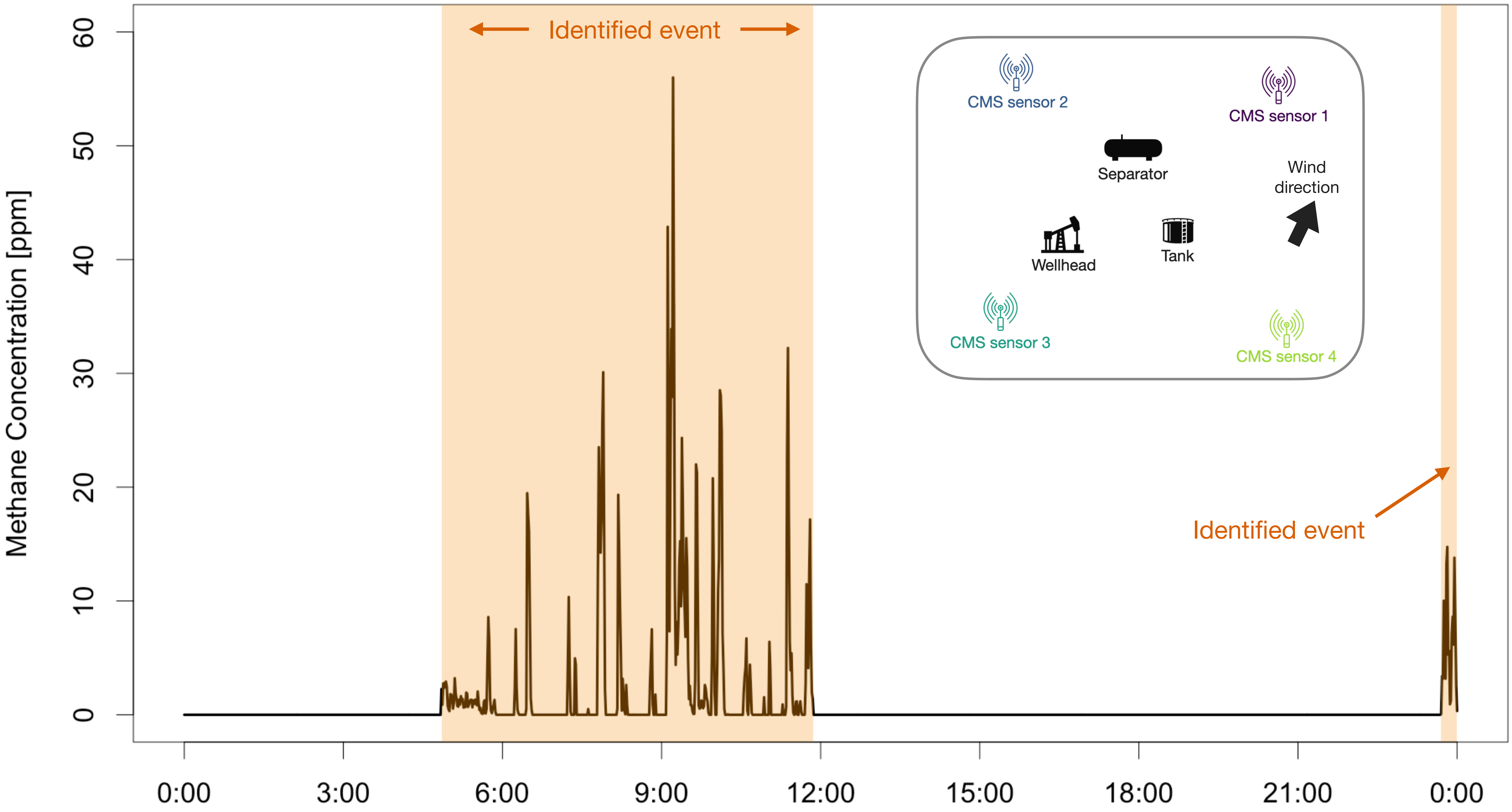


CMS sensor 1
CMS sensor 2
CMS sensor 3
CMS sensor 4
Methane Concentration [ppm]
0
10
20
30
40
50
60
0:00
3:00
6:00
9:00
12:00
15:00
18:00
21:00
0:00
CMS sensor 2
CMS sensor 1
Separator
Wind direction
Wellhead
Tank
CMS sensor 3
CMS sensor 4



Methane Concentration [ppm]
0
10
20
30
40
50
60
0:00
3:00
6:00
9:00
12:00
15:00
18:00
21:00
0:00
CMS sensor 1
CMS sensor 2
CMS sensor 3
CMS sensor 4
CMS sensor 2
CMS sensor 1
Separator
Wind direction
Wellhead
Tank
CMS sensor 3
CMS sensor 4


Maximum across sensors
Methane Concentration [ppm]
0
10
20
30
40
50
60
0:00
3:00
6:00
9:00
12:00
15:00
18:00
21:00
0:00
CMS sensor 2
CMS sensor 1
Separator
Wind direction
Wellhead
Tank
CMS sensor 3
CMS sensor 4


Identified event
Identified event
CMS sensor 2
CMS sensor 1
Separator
Wind direction
Wellhead
Tank
CMS sensor 3
CMS sensor 4
Methane Concentration [ppm]
0
10
20
30
40
50
60
0:00
3:00
6:00
9:00
12:00
15:00
18:00
21:00
0:00

“Wish list” that guides Bayesian hierarchical model development

What we want:

1. Constrain parameters (emission rates) to be non-negative.
   Not likely to be methane sinks on oil and gas sites.
2. Shrink small estimates to identically zero.
   Makes alerting easier.
3. Include operator insight via priors.
   Often well known if a particular source will be leaking given the season, production volume, etc.

Model hierarchy

n = number of observations
p = number of potential sources

Assume the standard linear model:

$$y = X\beta + \epsilon$$

$$\epsilon \sim N(0,\sigma^2)$$

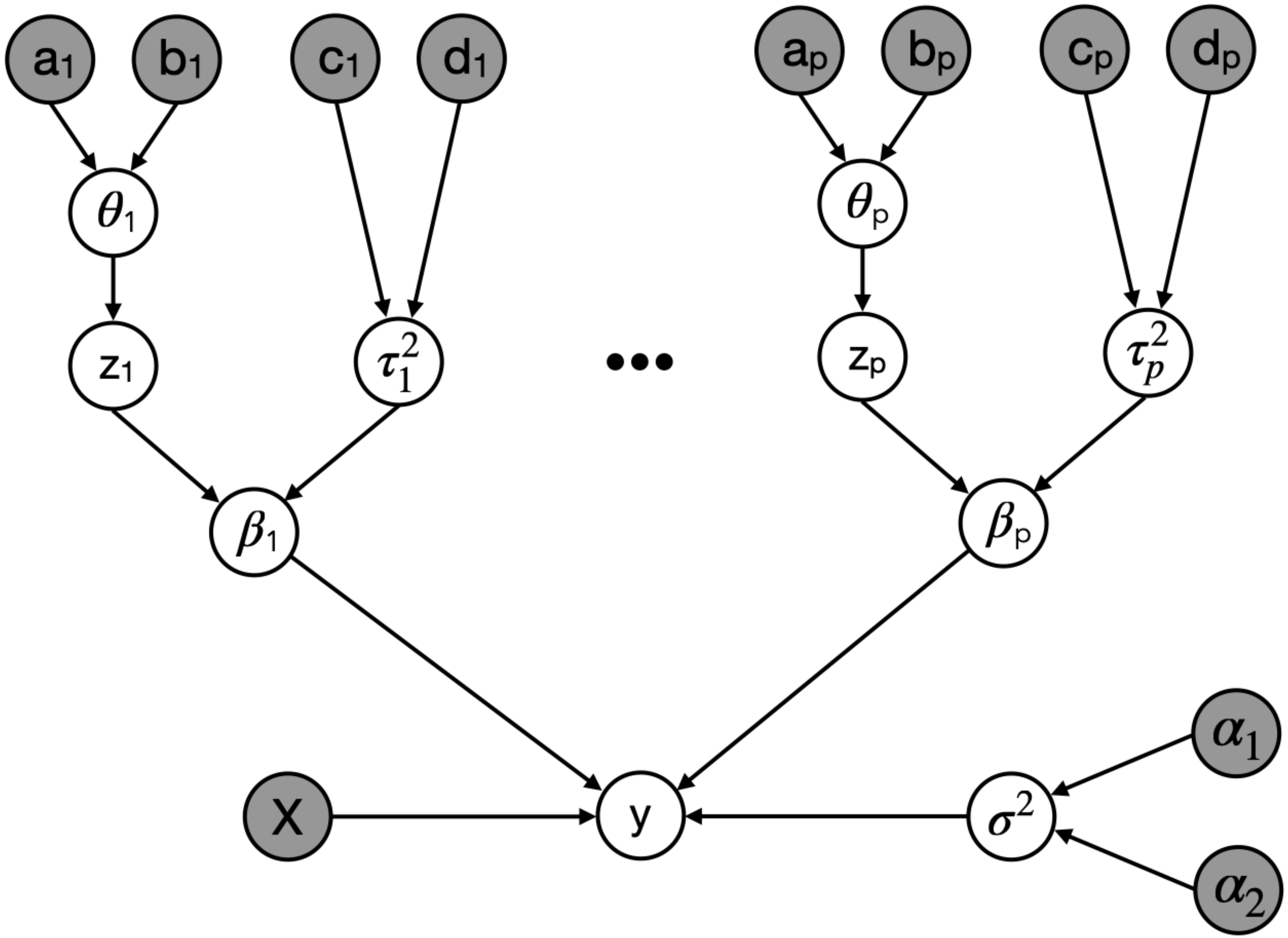
Create the following prior structure

$$\beta_i \sim \begin{cases} 0 & z_i = 0 \\ \mathrm{Exp}(\tau_i^2\sigma^2) & z_i = 1 \end{cases}$$

Achieve spike-and-slab prior using a Dirac delta function

$$f(\beta_i|\tau_i^2, \sigma^2, z_i) = (1 - z_i)\delta_0(\beta_i) + z_i \mathrm{Exp}(\beta_i|\tau_i^2\sigma^2)$$

Model hierarchy



Sampling from the posterior

Let ξ be a vector of all other parameters

$$\xi = \{\beta_1, ..., \beta_p, z_1, ..., z_p, \theta_1, ..., \theta_p, \tau_1^2, ..., \tau_p^2, \sigma^2\}$$

Bayes' theorem gives us a way of getting at the posterior distribution we are interested in

$$p(\xi|y) = \frac{p(y|\xi)p(\xi)}{p(y)} = \frac{p(y|\xi)p(\xi)}{\int p(y|\xi)p(\xi)d\xi}$$

Computing the marginal likelihood is often infeasible, so we can work with proportionality

$$p(\xi|y) \propto p(y|\xi)p(\xi)$$

Metropolis-Hastings can be used to sample this, but can be inefficient in high dimensional space

Sampling from the posterior

Instead, we can use a Gibbs sampler to sample from the posterior:
Sample from the posterior by iteratively sampling from the full conditional for each parameter

The steps below are used to generate the $(c+1)^{th}$ iteration of the cycle

- Step 1: Draw $\xi_1^{(c+1)} \sim p(\xi_1 | \xi_2^{(c)}, \xi_3^{(c)}, ..., \xi_k^{(c)}, y)$
- Step 2: Draw $\xi_2^{(c+1)} \sim p(\xi_2 | \xi_1^{(c+1)}, \xi_3^{(c)}, ..., \xi_k^{(c)}, y)$
- ...
- Step i: Draw $\xi_i^{(c+1)} \sim p(\xi_i | \xi_1^{(c+1)}, \xi_2^{(c+1)}, ..., \xi_{i-1}^{(c+1)}, \xi_{i+1}^{(c)}, ..., \xi_k^{(c)}, y)$
- ...
- Step k: Draw $\xi_k^{(c+1)} \sim p(\xi_k | \xi_1^{(c+1)}, \xi_2^{(c+1)}, ..., \xi_{k-1}^{(c+1)}, y)$

Use a Gibbs sampler to sample from the posterior

Just need to derive all of the necessary conditionals

$$\sigma^2|\xi = \sigma^2|y, \beta \sim \text{Inv-Gamma}\left(\alpha_1 + \frac{n}{2}, \alpha_2 + \frac{(y - X\beta)^T(y - X\beta)}{2}\right)$$

$$\theta_i|\xi = \theta_i|z_i \sim \text{Beta}(z_i + a_i, 1 - z_i + b_i)$$

$$\tau_i^2|\xi = \tau_i^2|\beta_i, z_i \sim \begin{cases} \text{Inv-Gamma}(c_i, d_i) & z_i = 0 \\ \text{Inv-Gamma}\left(1 + c_i, \frac{\beta_i}{\sigma^2} + d_i\right) & z_i = 1 \end{cases}$$

$$\beta_i|\xi = \beta_i|y, \beta_{-i}, \sigma^2, \tau_i^2, z_i \sim \begin{cases} 0 & z_i = 0 \\ \mathcal{N}\left((\frac{X^TX}{\sigma^2})^{-1}(\frac{X^Ty}{\sigma^2} - \frac{e_i}{\tau_i^2\sigma^2}), (\frac{X^TX}{\sigma^2})^{-1}\right) & z_i = 1 \end{cases}$$

$$z_i|\xi = z_i|y, z_{-i}, \beta_{-i}, \sigma^2, \tau^2, \theta \sim \text{Bernoulli}\left(\frac{(1-\theta_i)}{(1-\theta_i) + \frac{\theta_i}{2\tau_i^2\sigma^2}\exp\left(\frac{(x_i^Tw - (1/\tau_i^2))^2}{2\sigma^2 x_i^T x_i}\right)\left(\frac{2\pi\sigma^2}{x_i^Tx_i}\right)^{1/2}}\right)$$

Model evaluation on simulated data: “sanity check”

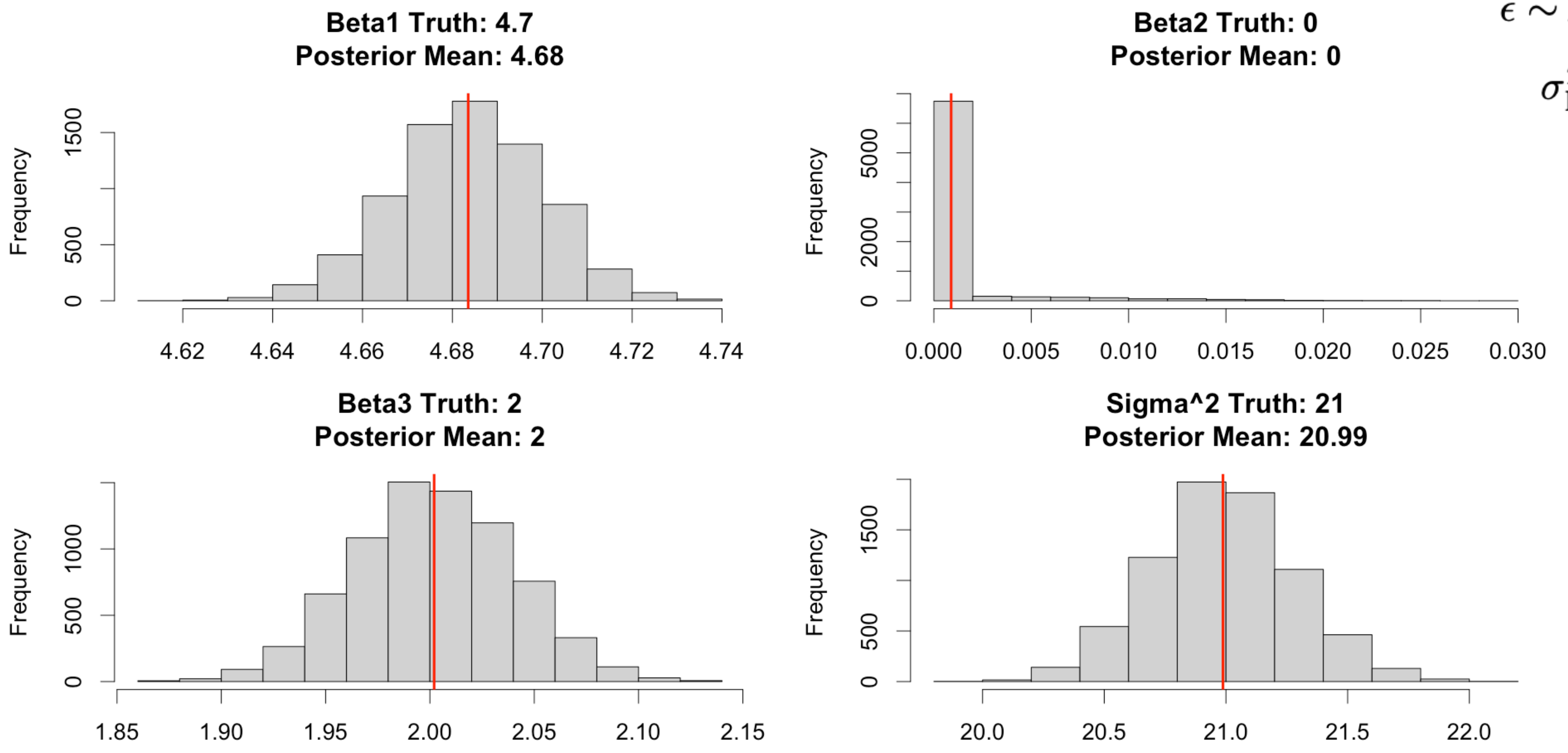
Create fake response data with known parameter values.
Make sure model can retrieve these parameters.

$$y = \beta_{\text{known}} X + \epsilon$$

$$\beta_{\text{known}} = \{4.7, 0, 2\}^T$$

$$\epsilon \sim \mathcal{N}(0, \sigma^2_{\text{known}})$$

$$\sigma^2_{\text{known}} = 21$$



Model evaluation on simulated data: “sanity check”

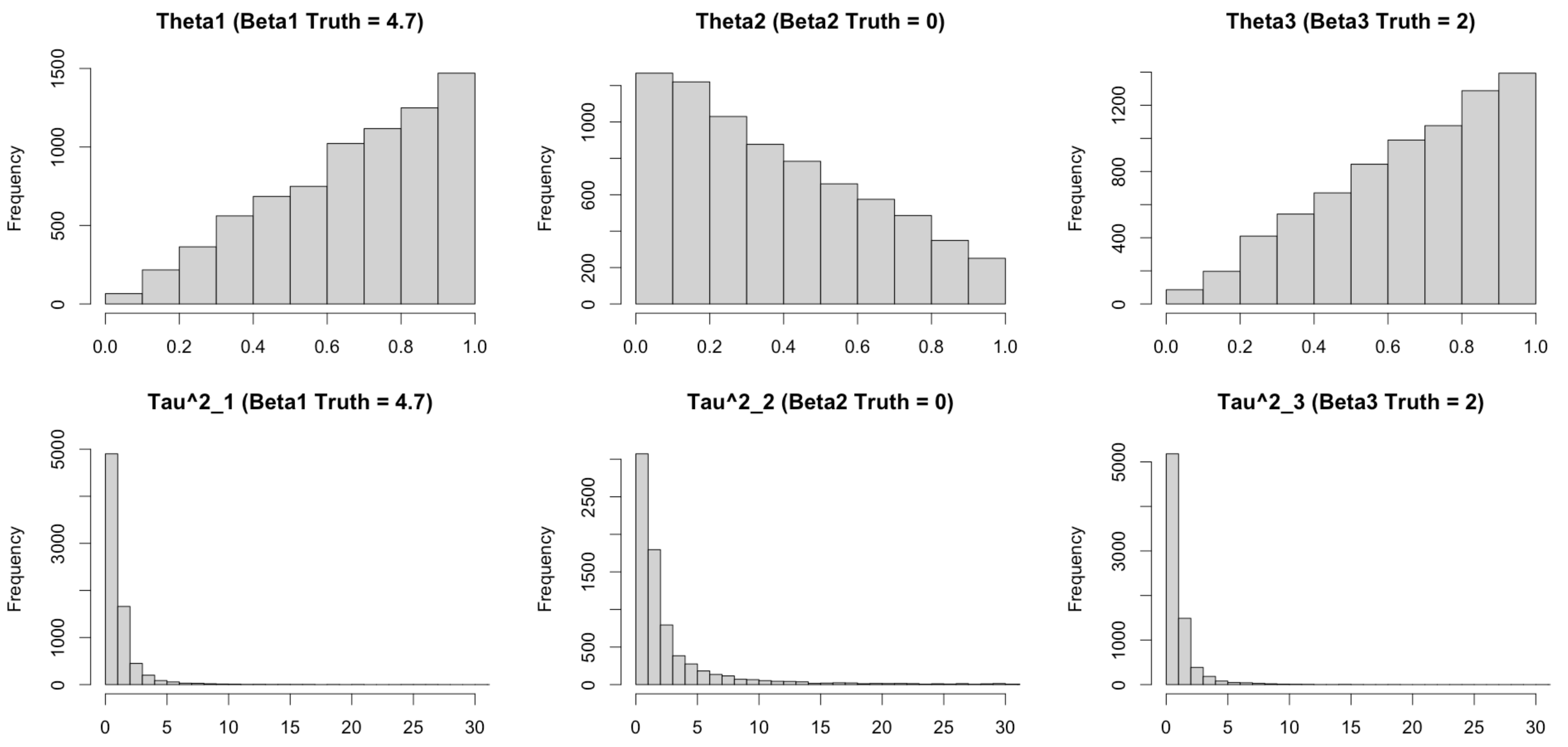
Create fake response data with known parameter values.
Make sure model can retrieve these parameters.

$$y = \beta_{\text{known}} X + \epsilon$$

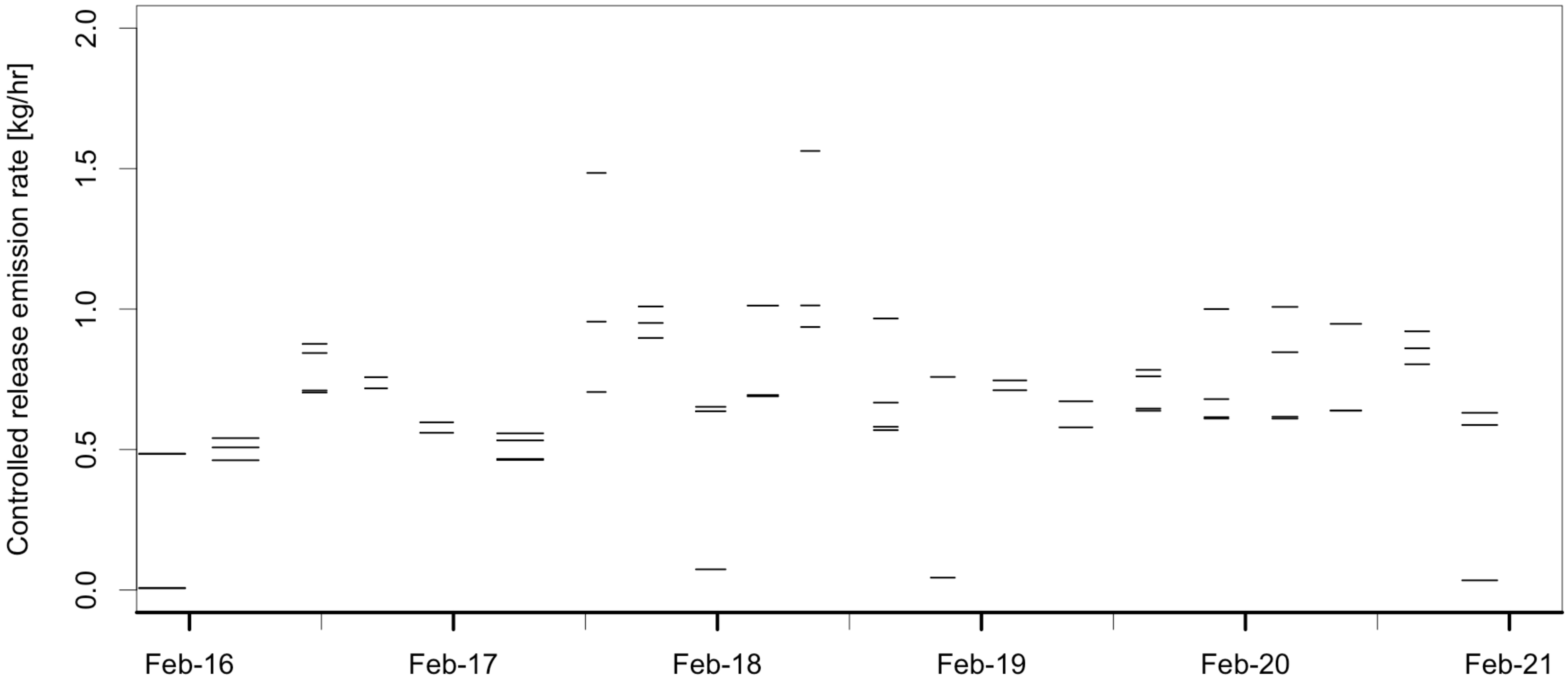
$$\beta_{\text{known}} = \{4.7, 0, 2\}^T$$

$$\epsilon \sim \mathcal{N}(0, \sigma^2_{\text{known}})$$

$$\sigma^2_{\text{known}} = 21$$



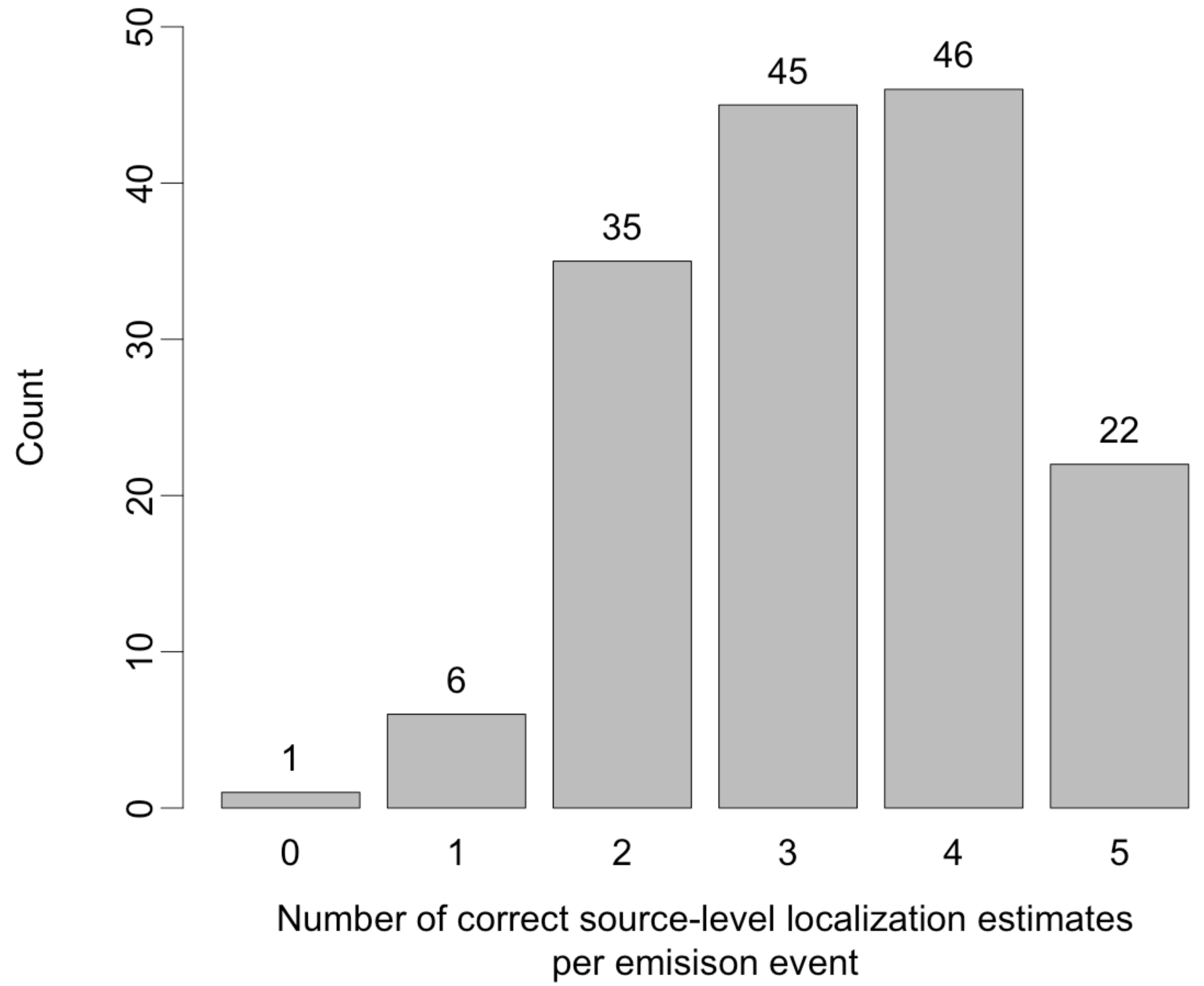
Model evaluation on multi-source controlled release data

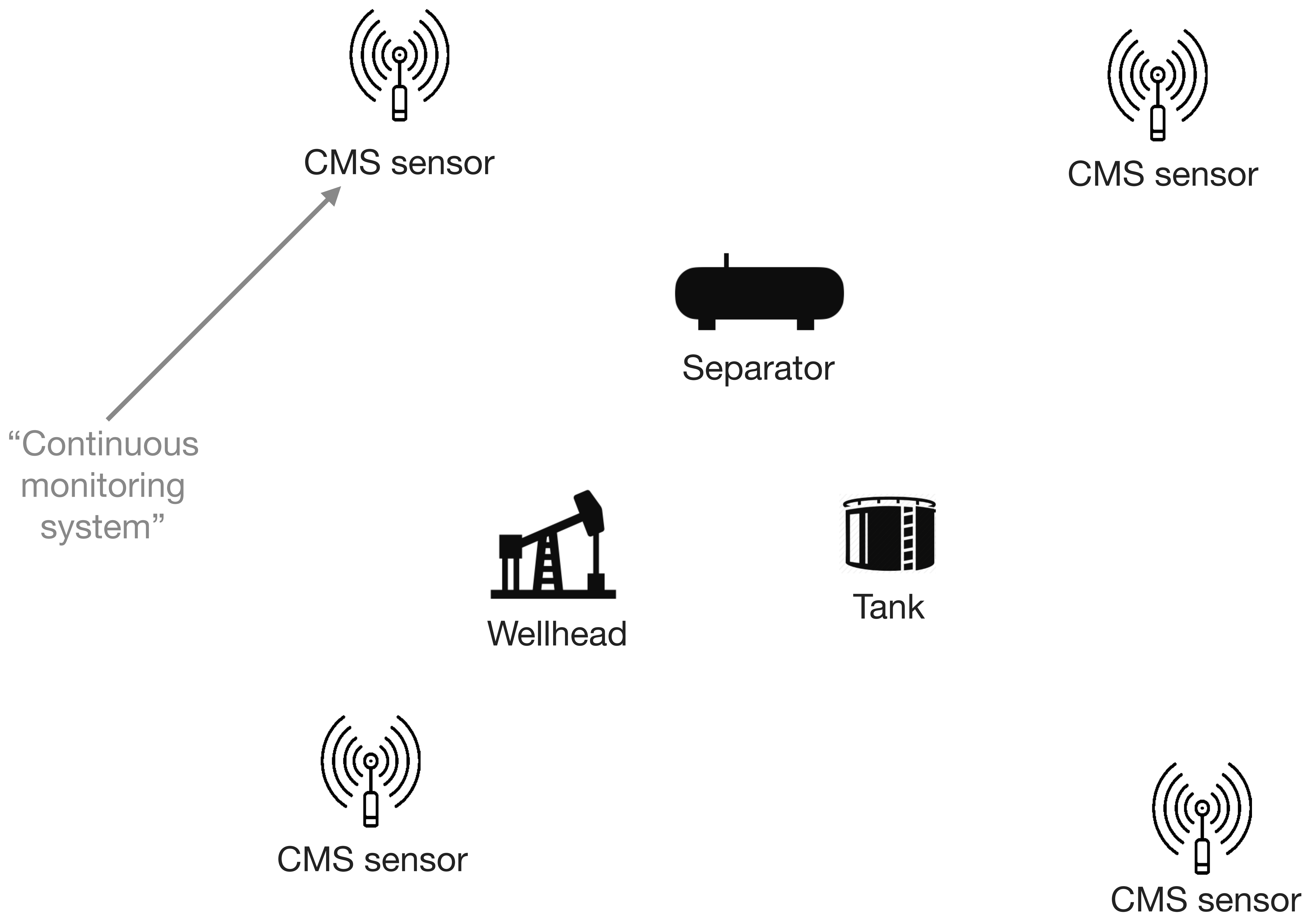


Model evaluation on multi-source controlled release data

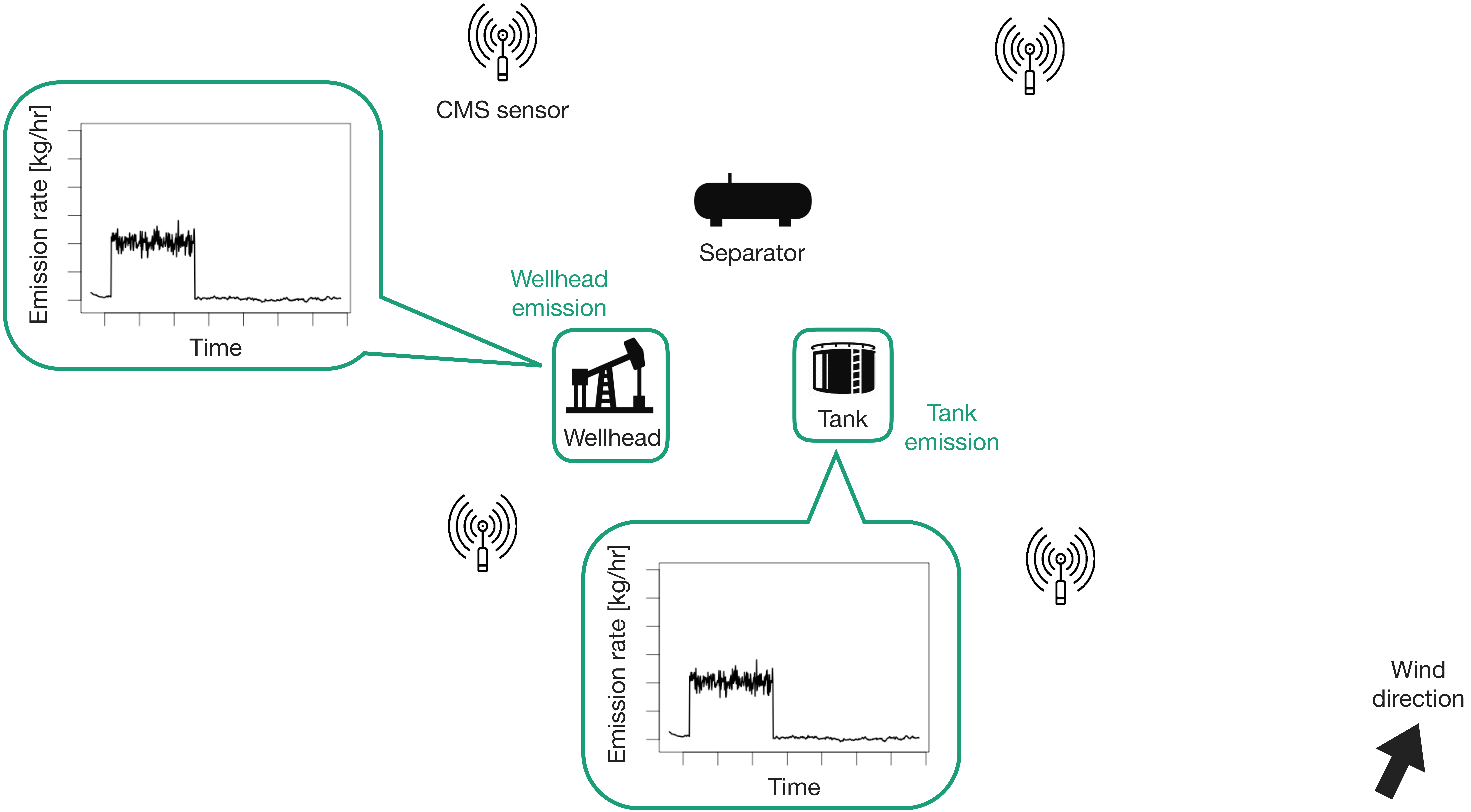
	Tanks	West Wellhead	West Separator	East Wellhead	East Separator
Percent of emission events with correct localization estimate	46%	66%	70%	69%	74%

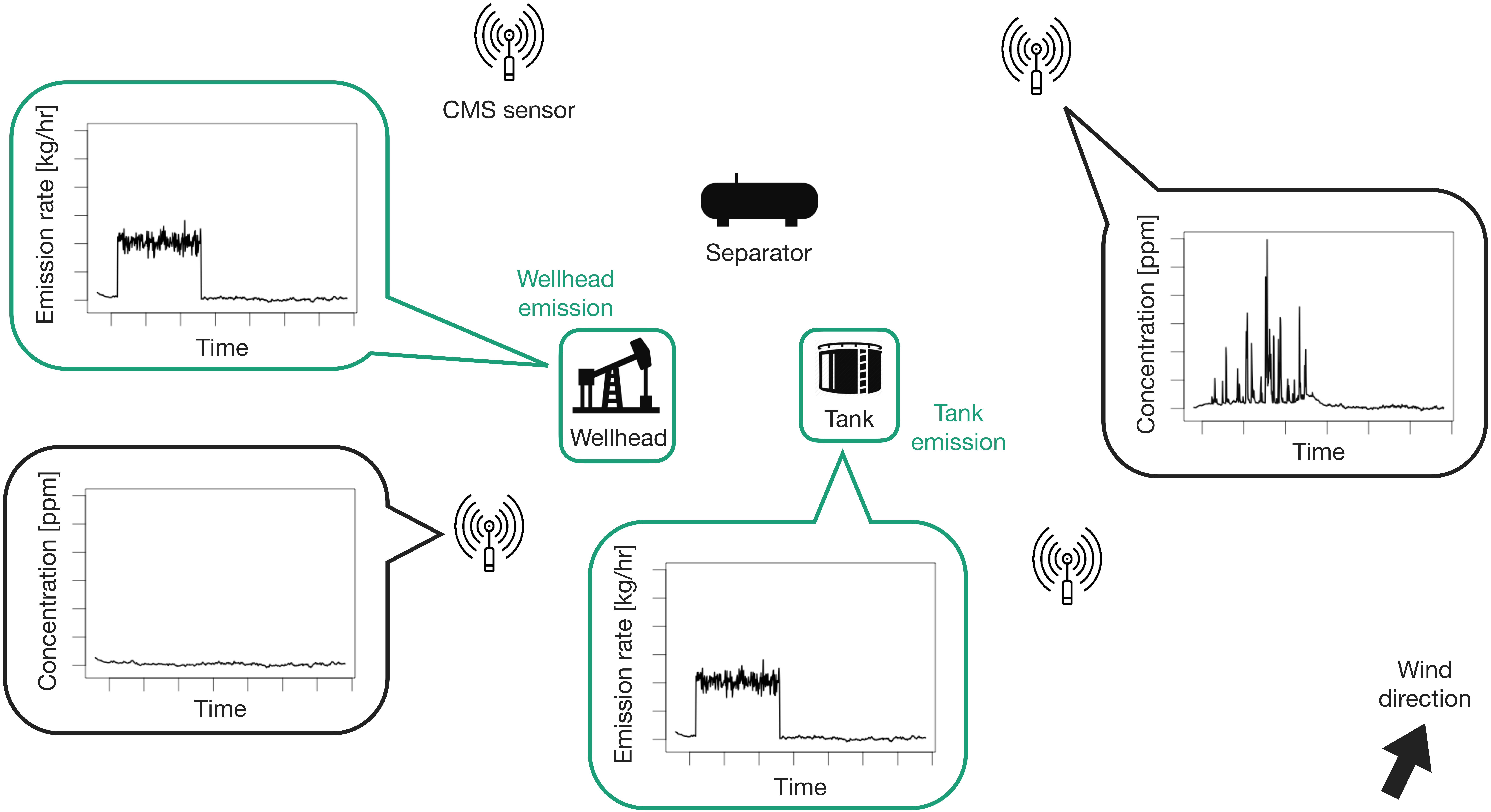
For now, let a localization estimate mean an emission rate estimate > 0.01 kg/hr

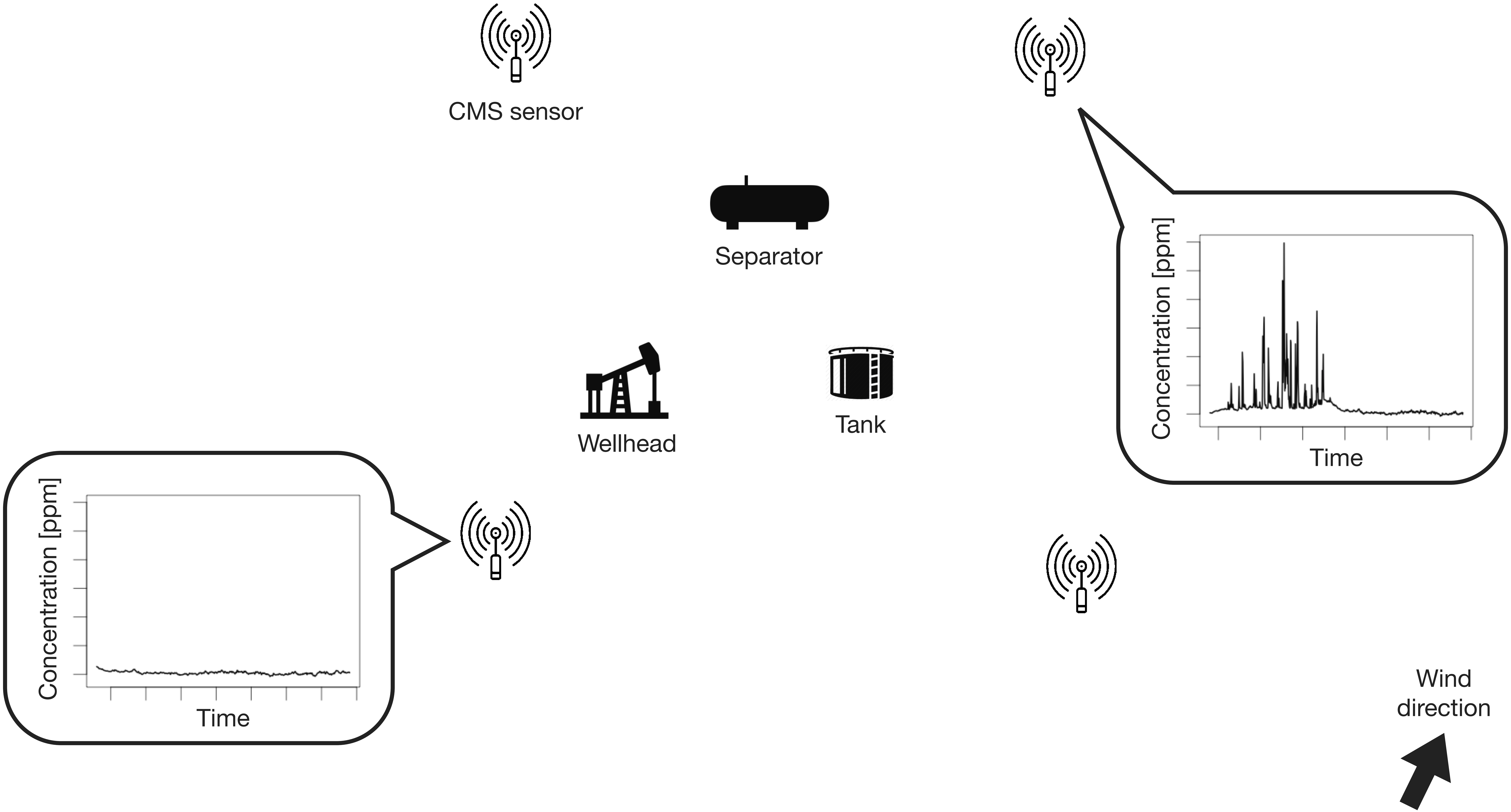




The multi-source continuous monitoring inverse problem


CMS sensor
Emission rate [kg/hr]
Time
Separator
Wellhead emission
Wellhead
Tank
Tank emission
Emission rate [kg/hr]
Time
Wind direction


CMS sensor
Emission rate [kg/hr]
Time
Separator
Wellhead
emission
Wellhead
Tank
Tank
emission
Concentration [ppm]
Time
Concentration [ppm]
Time
Emission rate [kg/hr]
Time
Wind
direction


CMS sensor
Separator
Wellhead
Tank
Concentration [ppm]
Time
Concentration [ppm]
Time
Wind
direction