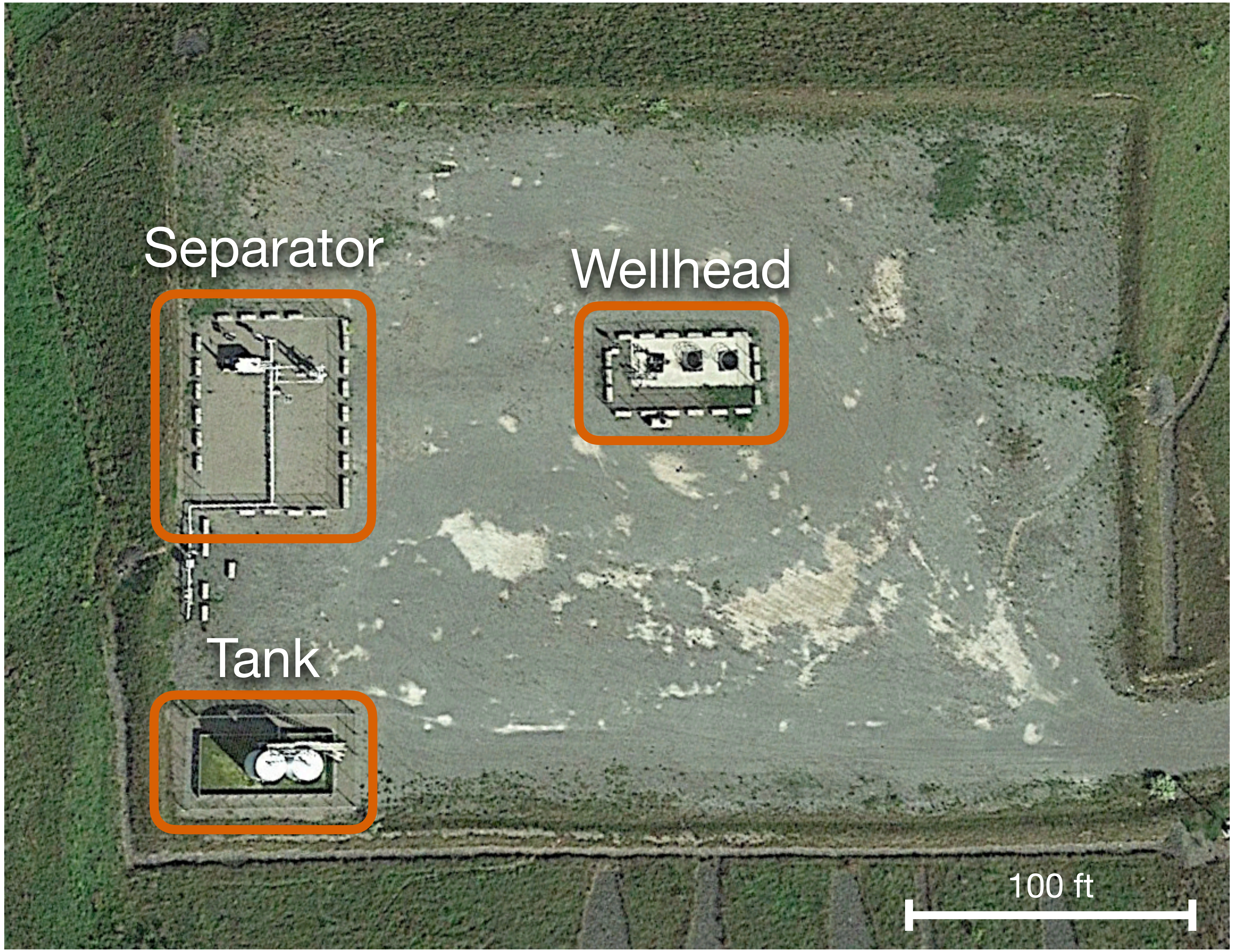
CMS Series #3:
Modeling multisource methane emissions on oil and gas sites

William Daniels

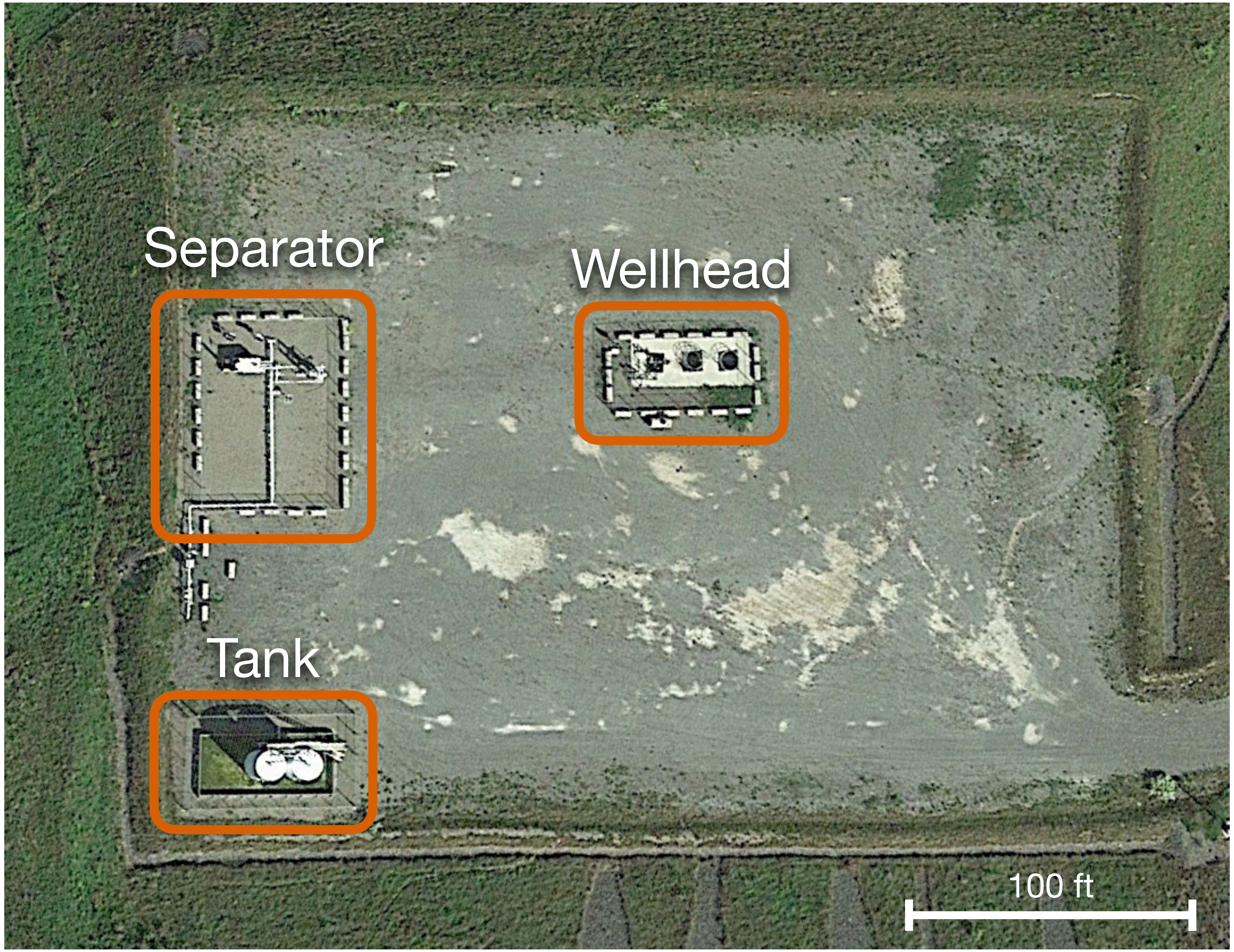
Department of Applied Mathematics and Statistics

COLORADO SCHOOL OF MINES

Example production oil and gas site

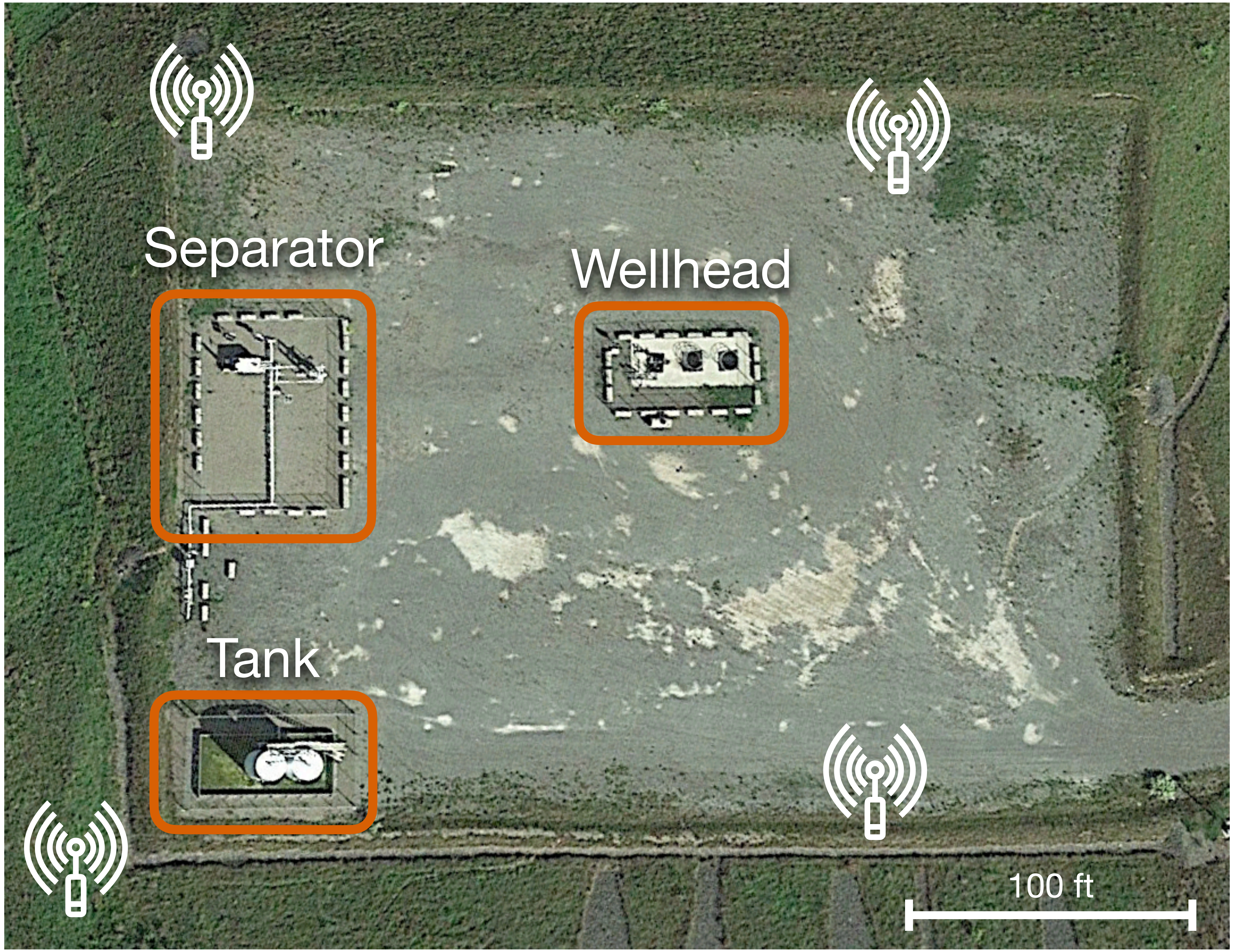


Example production oil and gas site



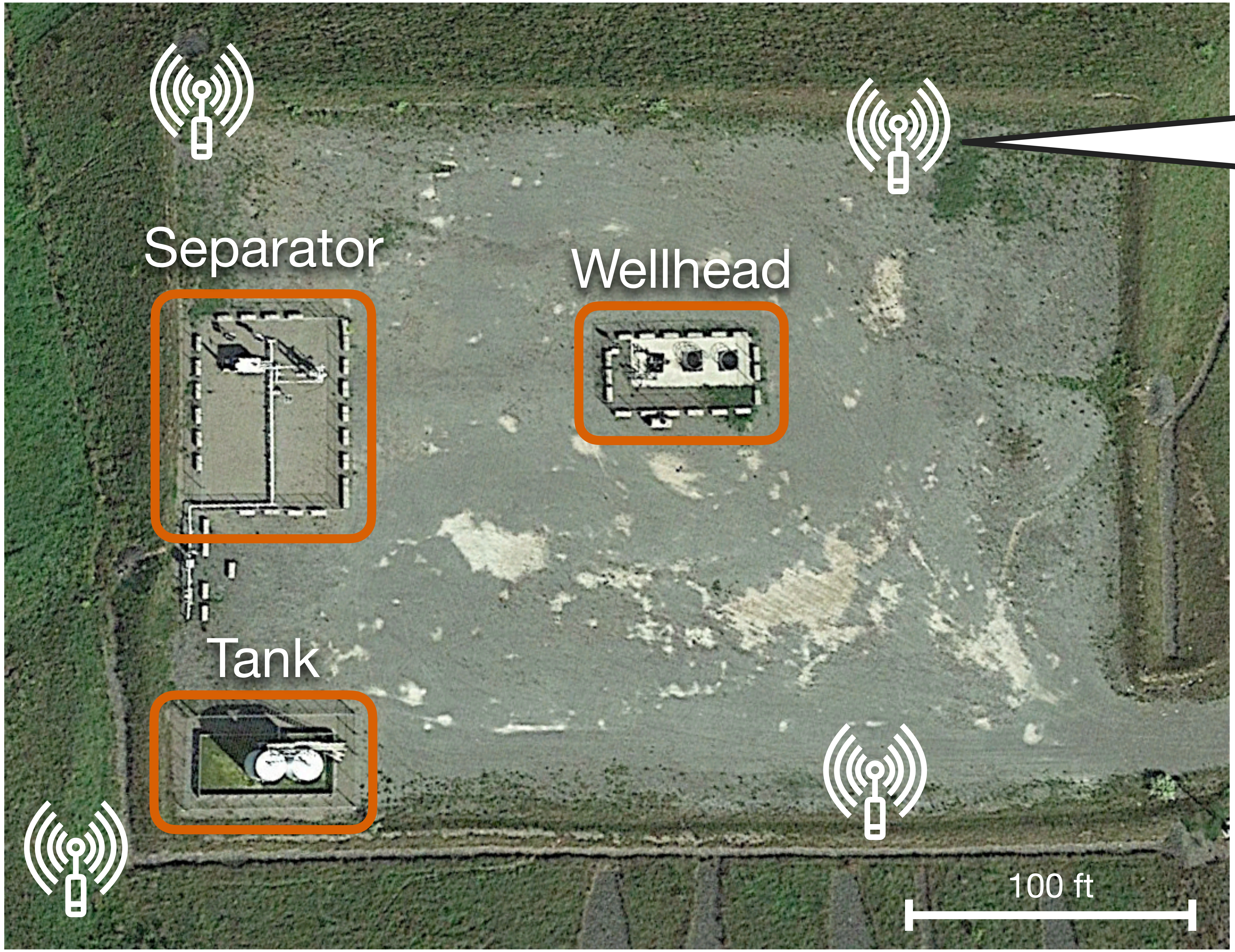
Continuous monitoring system (CMS)

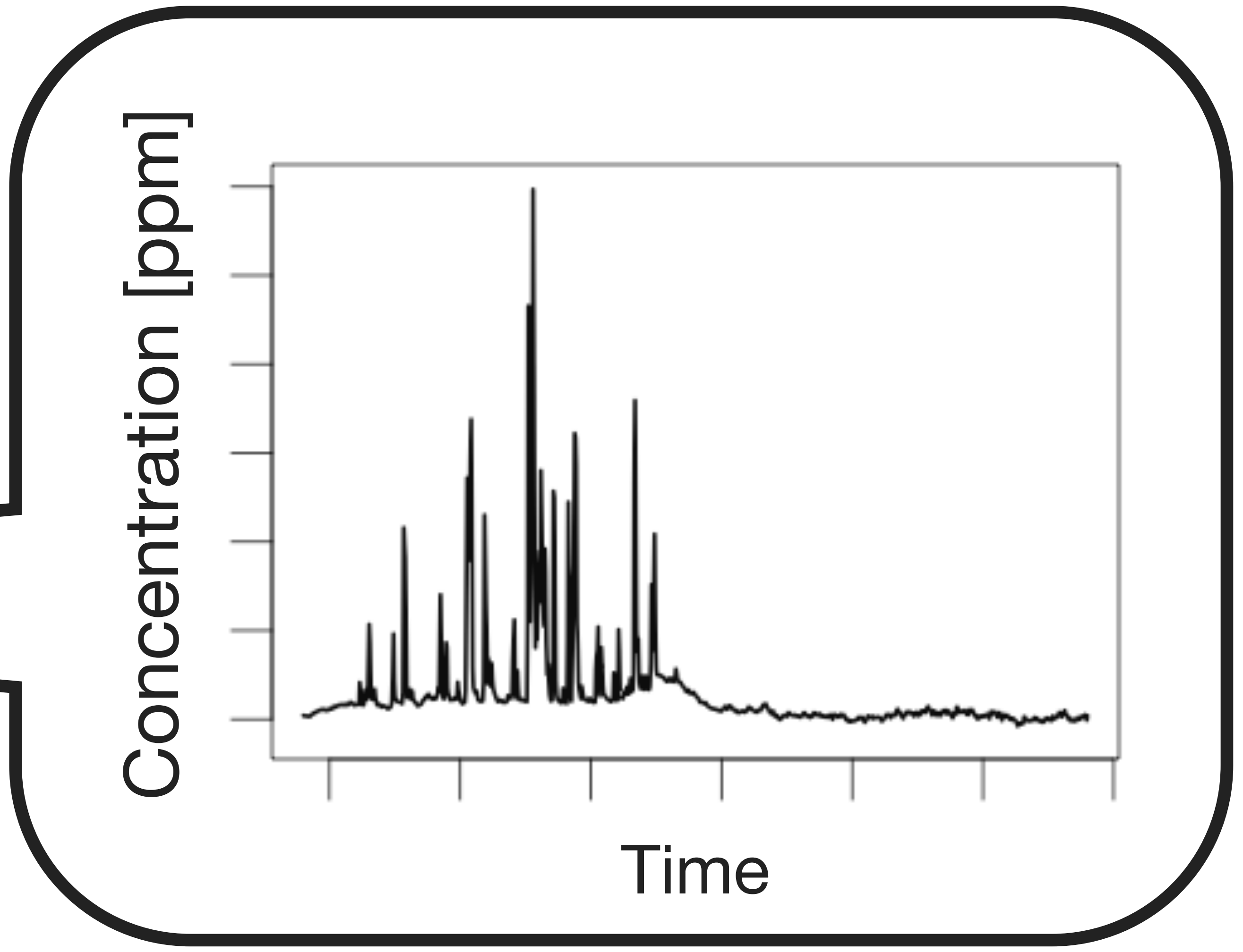
Example production oil and gas site



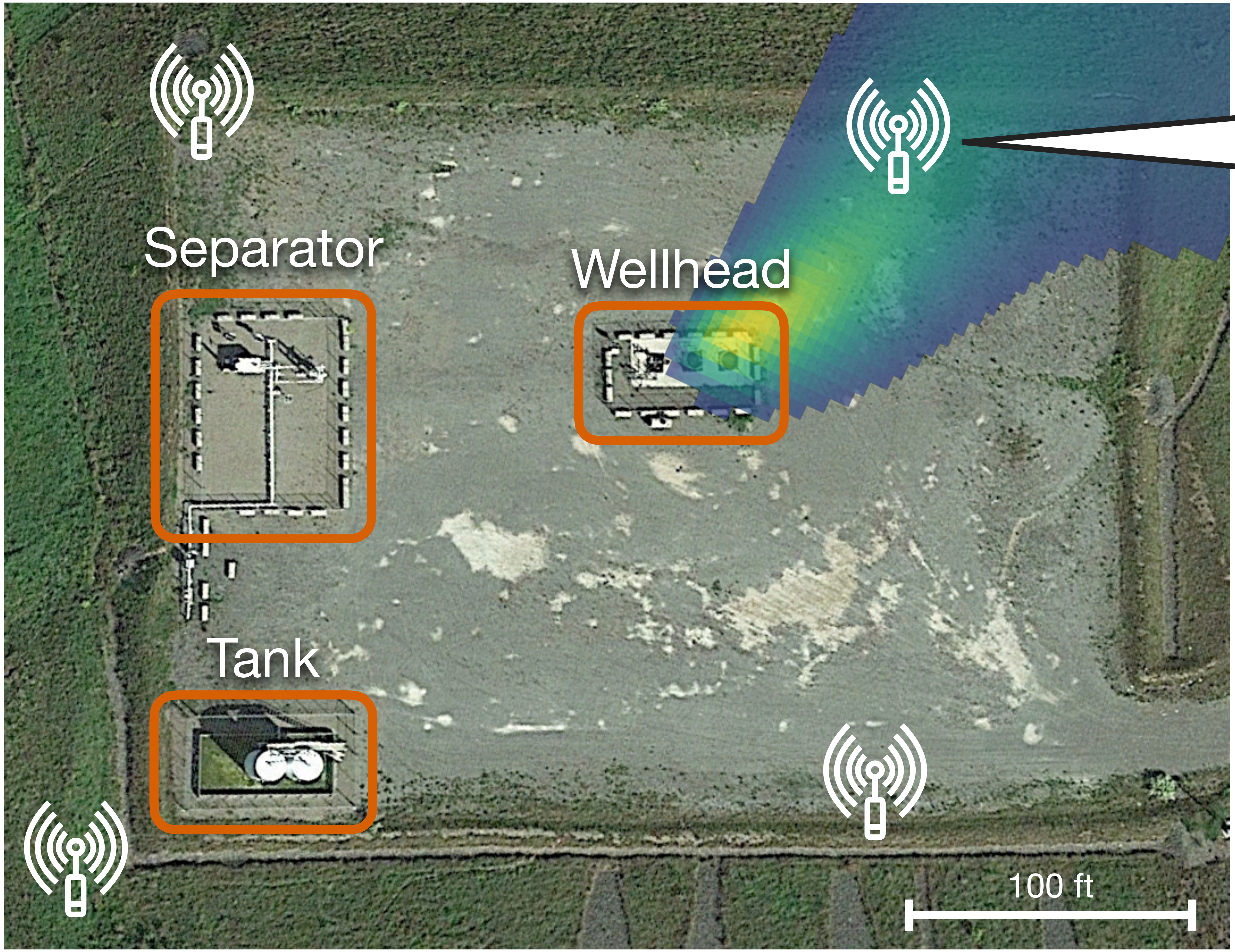
Continuous monitoring system (CMS)

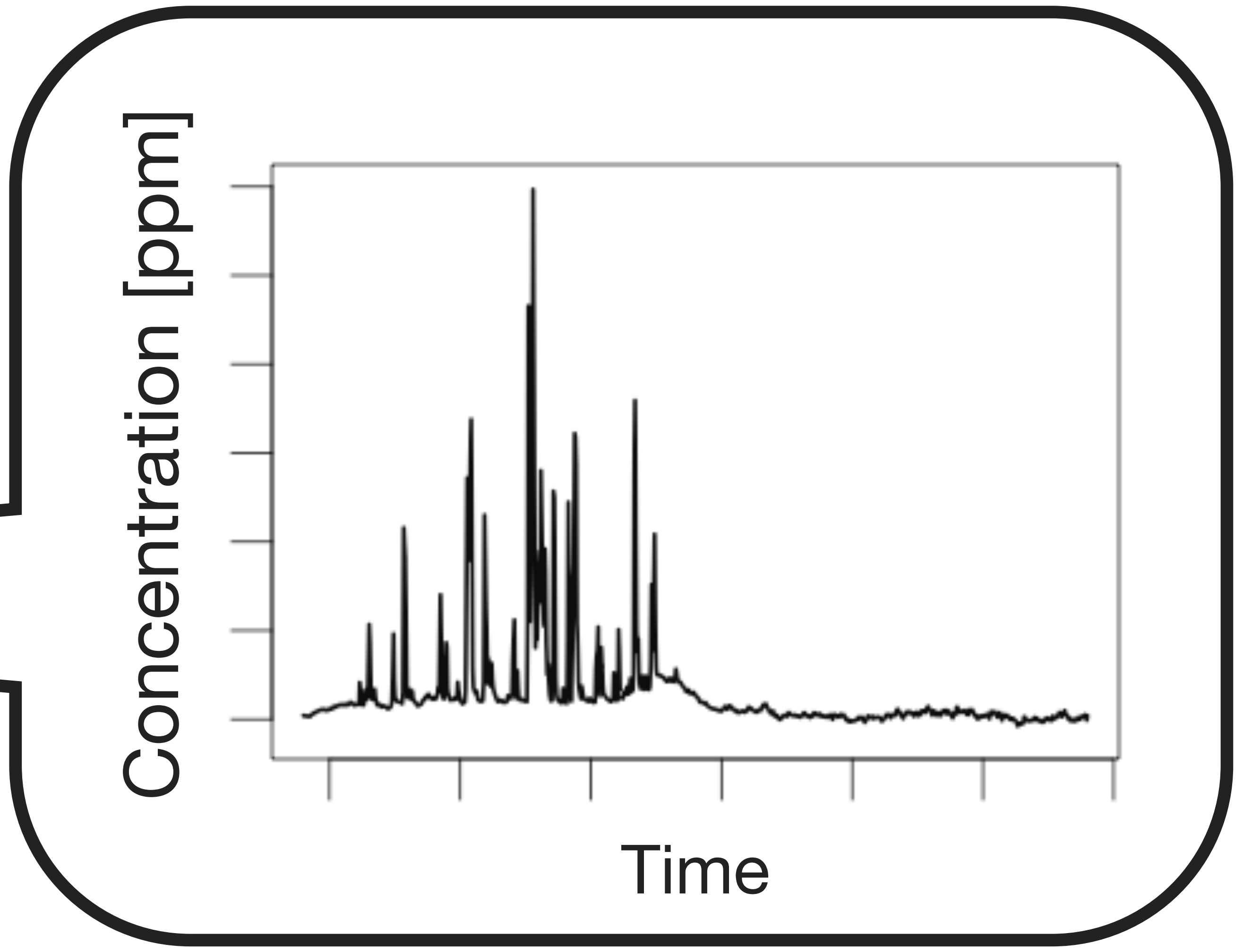
Example production oil and gas site





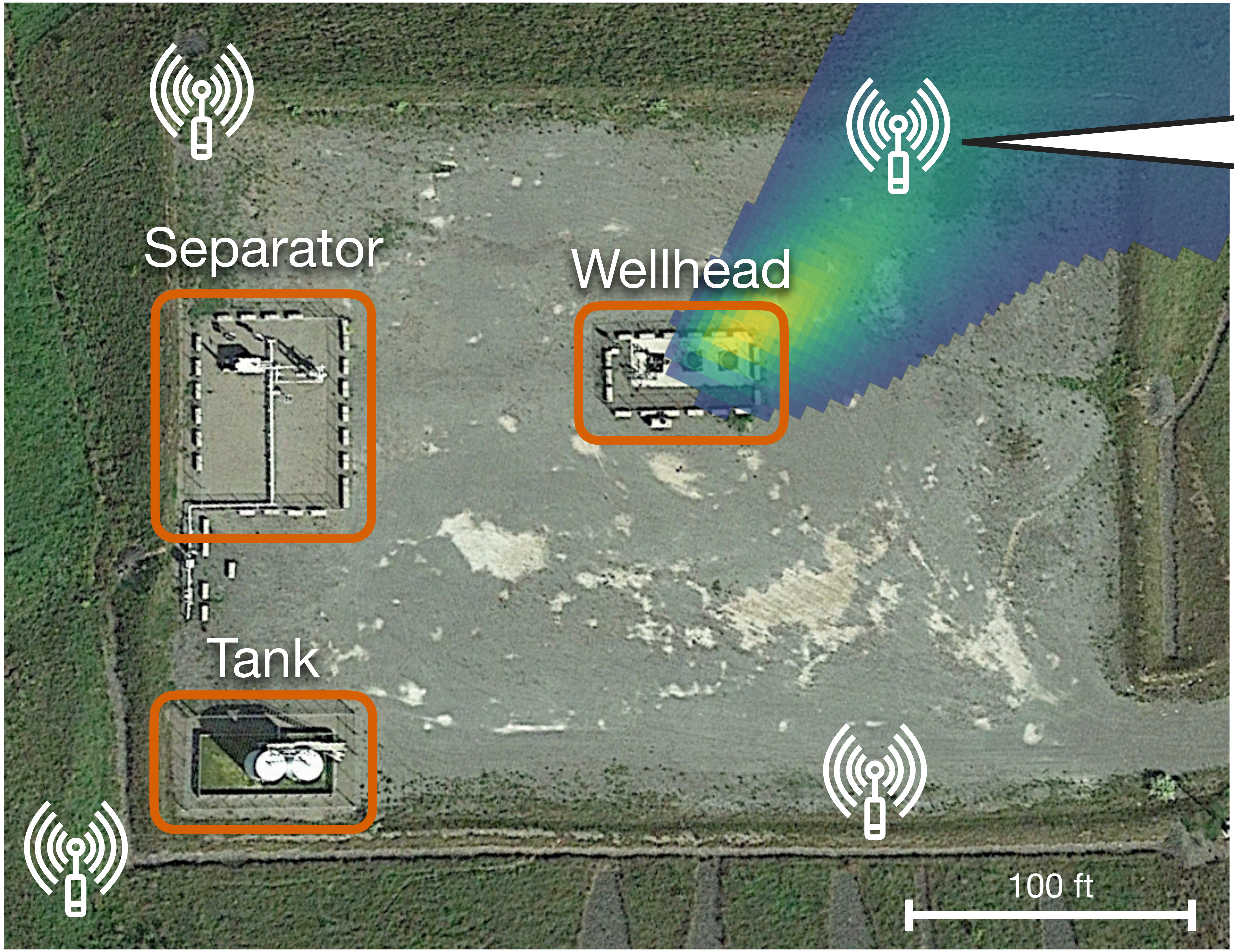
Example production oil and gas site

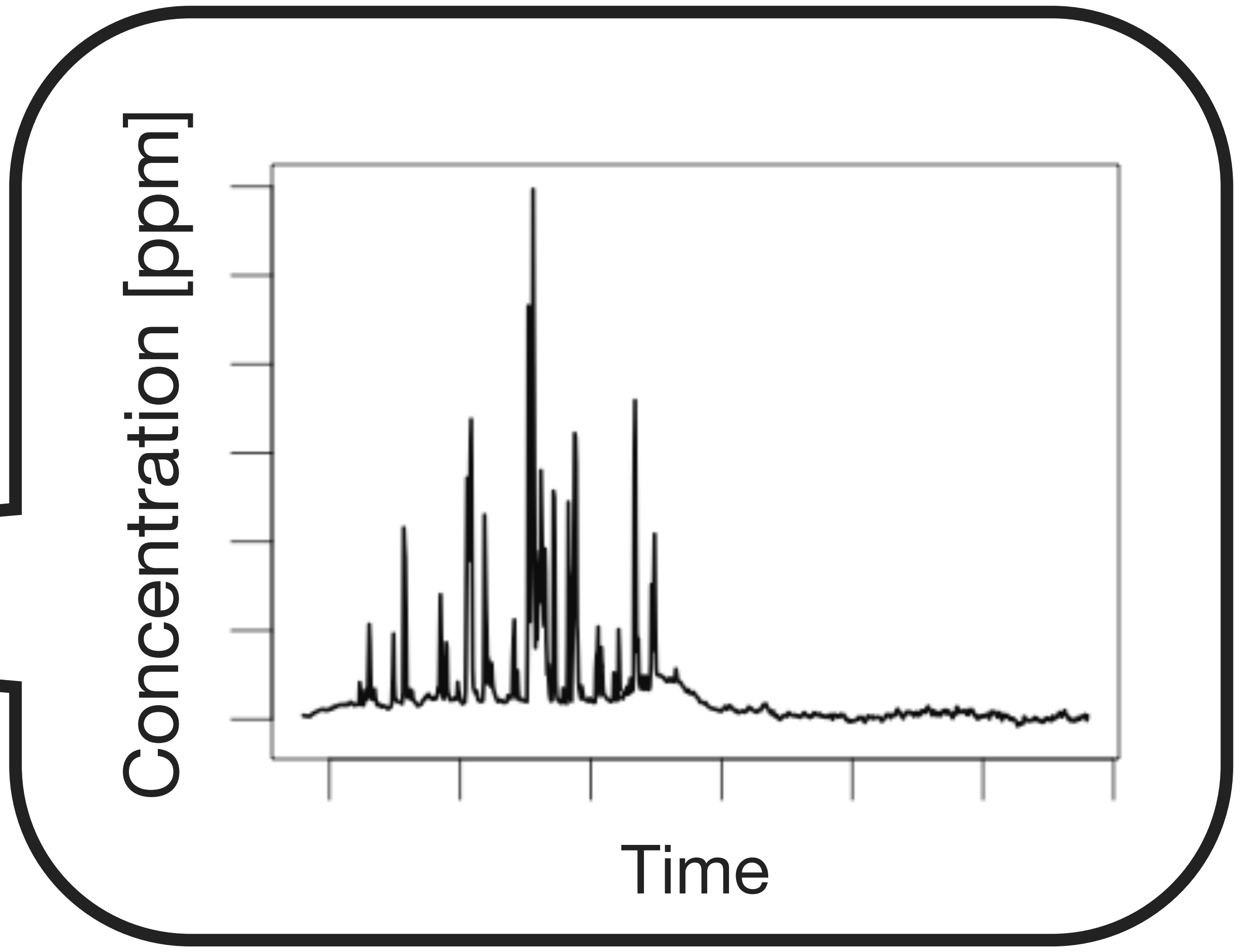




Aerial measurement technology

Example production oil and gas site



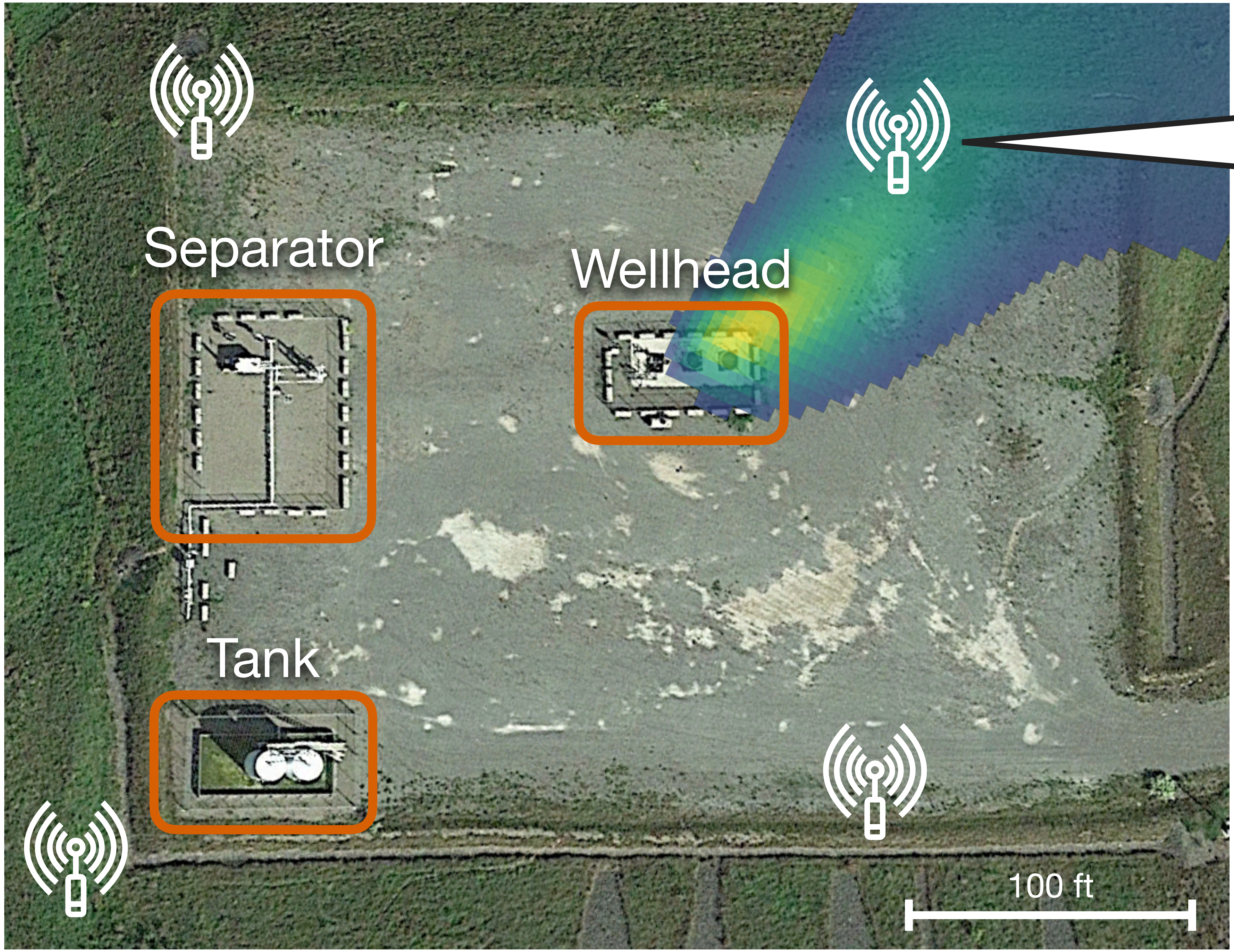


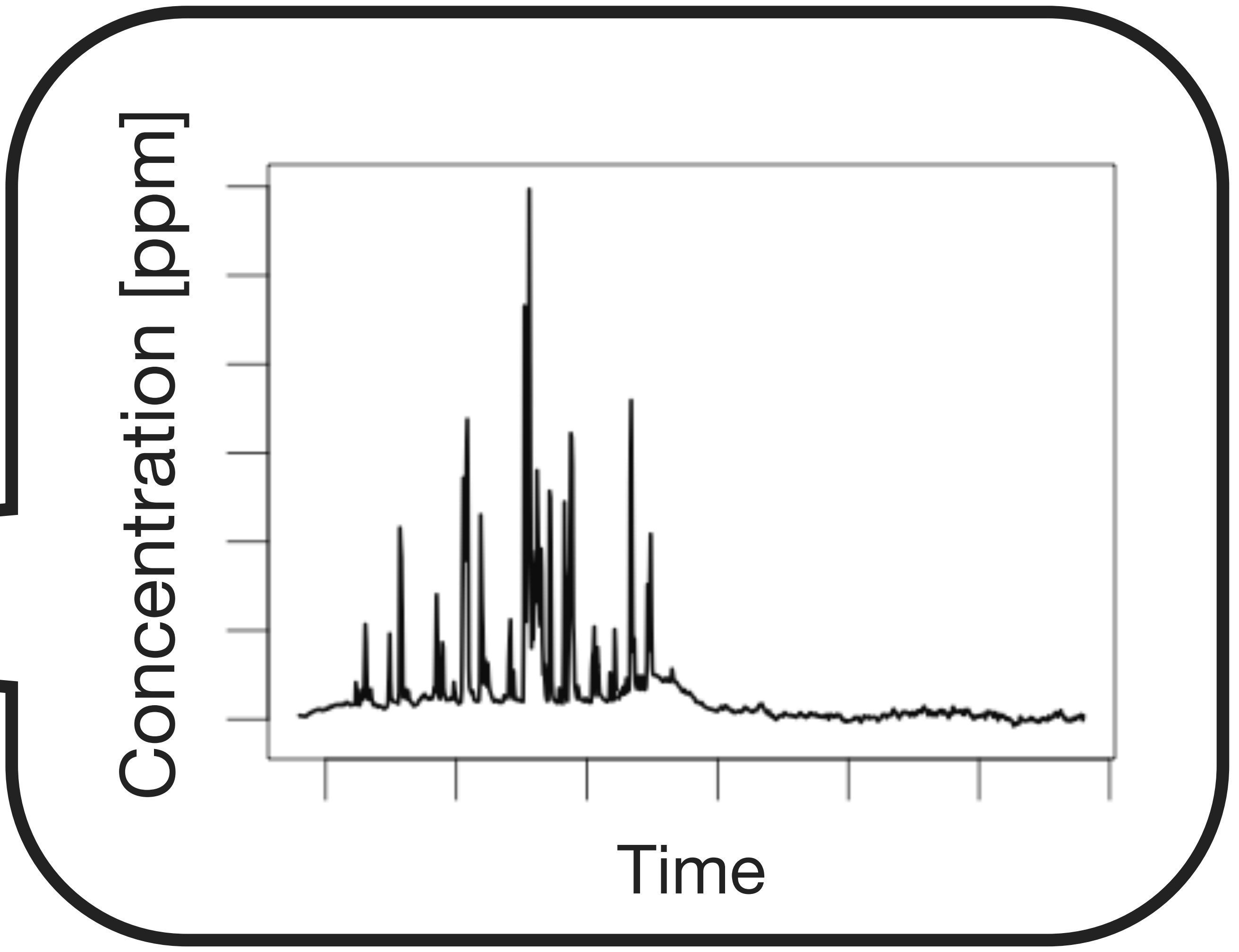
Aerial measurement technology

Bottom-up inventory estimate =

1 wellhead	x	wellhead emission factor	+
1 separator	x	separator emission factor	+
1 tank	x	tank emission factor	

Example production oil and gas site





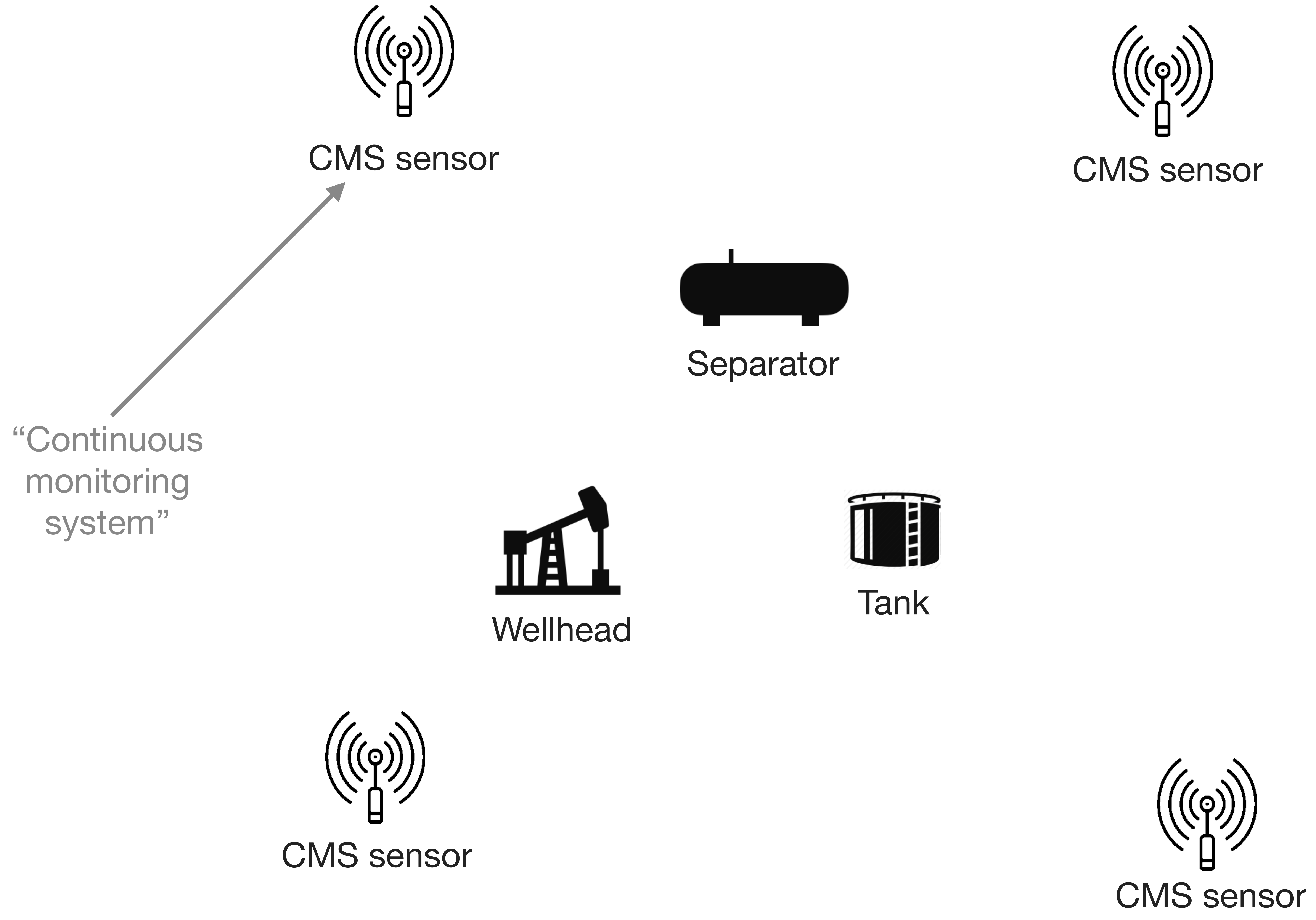
Event detection:
When is an emission happening?

Localization:
Where is the emission coming from?

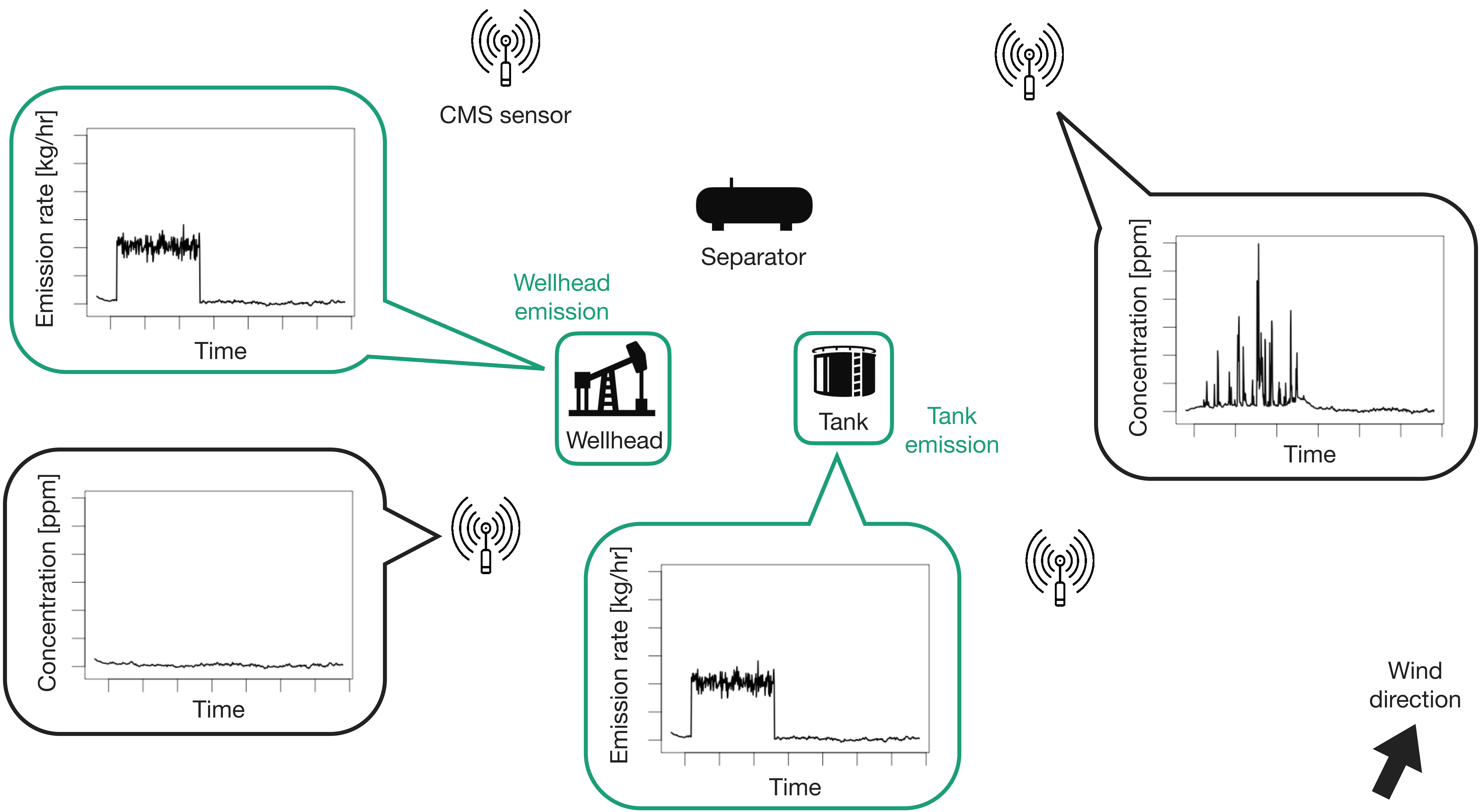
Quantification:
How much is being emitted?

Chapter 4:

Multi-source emission detection, localization, and quantification



The multi-source continuous monitoring inverse problem


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

Model hierarchy

n = number of observations
p = number of potential sources

Assume a multiple linear regression model at the data level

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

$$y \equiv \{y_1, \ldots, y_n\}, \beta \equiv \{\beta_1, \ldots, \beta_p\}, X \in \mathbb{R}^{n \times p}$$

Concentration observations from CMS sensors

Emission rates for each source

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

Model hierarchy

n = number of observations
p = number of potential sources

Assume a multiple linear regression model at the data level

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

$$y \equiv \{y_1, \ldots, y_n\}, \beta \equiv \{\beta_1, \ldots, \beta_p\}, X \in \mathbb{R}^{n \times p}$$

Assume that the errors $\epsilon \equiv \{\epsilon_1, \ldots, \epsilon_n\}$ are are identically distributed, Gaussian, and autocorrelated such that

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

Model hierarchy

n = number of observations
p = number of potential sources

Assume a multiple linear regression model at the data level

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

$$y \equiv \{y_1, \ldots, y_n\}, \beta \equiv \{\beta_1, \ldots, \beta_p\}, X \in \mathbb{R}^{n \times p}$$

Assume that the errors $\epsilon \equiv \{\epsilon_1, \ldots, \epsilon_n\}$ are are identically distributed, Gaussian, and autocorrelated such that

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

Let the errors follow an AR(1) process such that

$$\epsilon_i = r\epsilon_{i-1} + w$$

Autocorrelation coefficient

Gaussian white noise

Model hierarchy

n = number of observations
p = number of potential sources

Assume a multiple linear regression model at the data level

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

$$y \equiv \{y_1, \ldots, y_n\}, \beta \equiv \{\beta_1, \ldots, \beta_p\}, X \in \mathbb{R}^{n \times p}$$

Assume that the errors $\epsilon \equiv \{\epsilon_1, \ldots, \epsilon_n\}$ are are identically distributed, Gaussian, and autocorrelated such that

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

Let the errors follow an AR(1) process such that

$$\epsilon_i = r\epsilon_{i-1} + w$$

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

Model hierarchy

n = number of observations
p = number of potential sources

Given an AR(1) process for ϵ, the correlation matrix is

$$R = \begin{bmatrix} 1 & r & r^2 & \dots & r^{n-1} \\ r & 1 & r & \dots & \vdots \\ r^2 & r & 1 & \dots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r^{n-1} & \dots & \dots & \dots & 1 \end{bmatrix}$$

Model hierarchy

n = number of observations
p = number of potential sources

Given an AR(1) process for ϵ, the correlation matrix is

$$R = \begin{bmatrix} 1 & r & r^2 & \dots & r^{n-1} \\ r & 1 & r & \dots & \vdots \\ r^2 & r & 1 & \dots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r^{n-1} & \dots & \dots & \dots & 1 \end{bmatrix}$$

which has closed form expressions for the inverse and determinant:

$$R^{-1} = \frac{1}{(1-r^2)} \begin{bmatrix} 1 & -r & 0 & \dots & 0 \\ -r & 1+r^2 & -r & \dots & \vdots \\ 0 & -r & 1+r^2 & \dots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & \dots & 1 \end{bmatrix} \quad \text{and} \quad |R| = (1-r^2)^{n-1}$$

Model hierarchy

n = number of observations
p = number of potential sources

Data-level: $y = X\beta + \epsilon$
$\epsilon \sim N(0,\sigma^2 R)$

The remainder of the hierarchy takes the following form

Spike-and-slab prior allows samples to be identically zero

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

"Slab" component is non-negative

$$z_i \sim \mathrm{Bernoulli}(\theta_i)$$
$$\theta_i \sim \mathrm{Beta}(a_i, b_i)$$
$$\tau_i^2 \sim \text{Inv-Gamma}(c_i, d_i)$$
$$\sigma^2 \sim \text{Inv-Gamma}(\nu/2,\ \nu/2)$$
$$\nu \sim \text{Inv-Gamma}(\alpha_1, \alpha_2)$$
$$r \sim \mathrm{Uniform}(0, 1)$$

Proportion of samples where $z_i = 1$ gives posterior probability that source i is emitting

a_i, b_i, c_i, d_i can contain operator insight

Model hierarchy

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

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

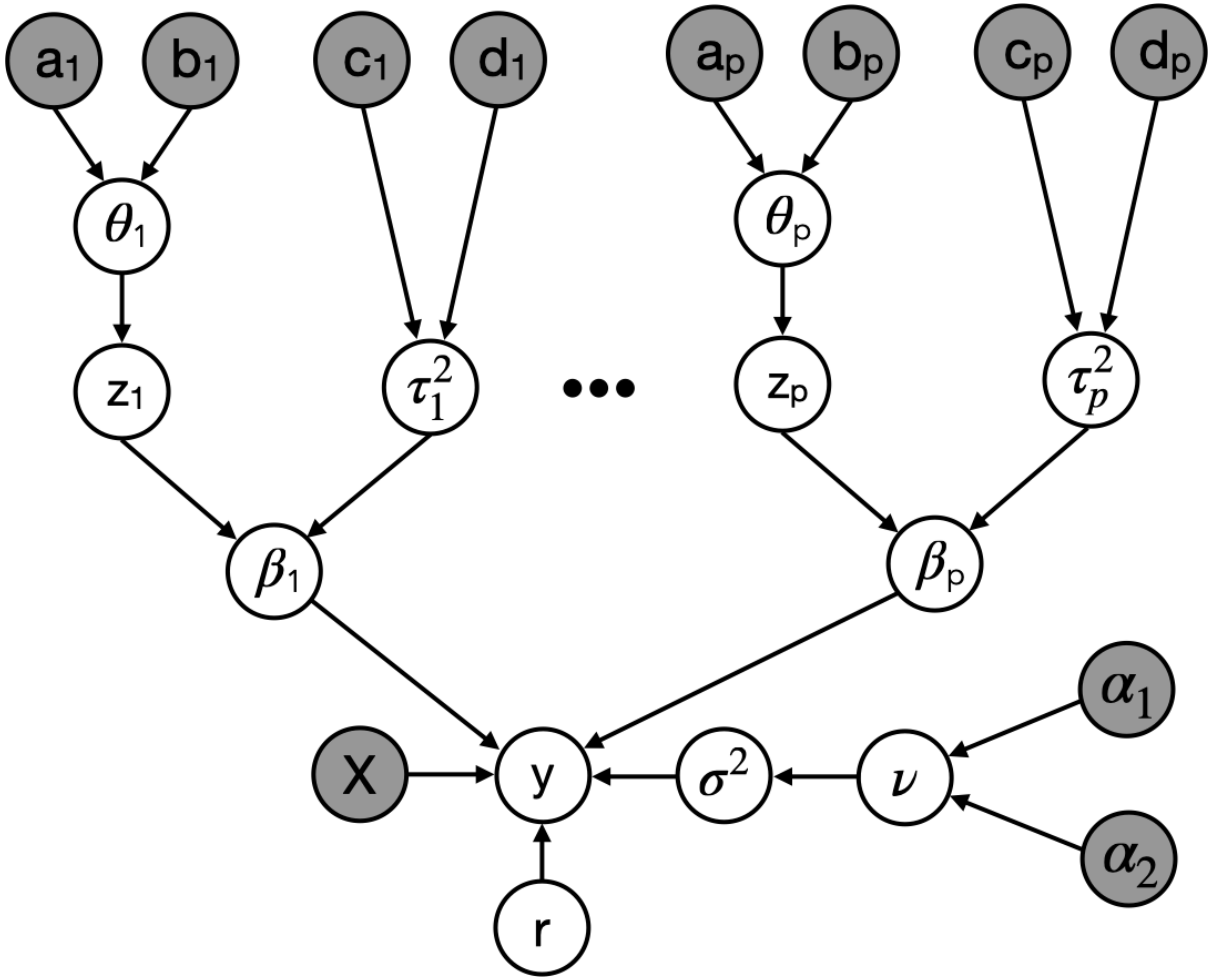
$$\theta_i \sim \mathrm{Beta}(a_i, b_i)$$

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

$$\sigma^2 \sim \text{Inv-Gamma}(\nu/2,\ \nu/2)$$

$$\nu \sim \text{Inv-Gamma}(\alpha_1, \alpha_2)$$

$$r \sim \mathrm{Uniform}(0, 1)$$



Sampling from the posterior

We can derive Gibbs updates for all parameters except ν.

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

$$\sigma^2|\xi \sim \text{Inv-Gamma}\left(\frac{\nu}{2} + \frac{n}{2}, \frac{\nu}{2} + \frac{1}{2}(y - X\beta)^T R^{-1}(y - X\beta)\right)$$

$$r|\xi \sim \begin{cases} \mathcal{N}(X\beta, \sigma^2 R) & 0 < r < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\tau_i^2|\xi \sim \text{Inv-Gamma}\left(z_i + c_i, \frac{\beta_i}{\sigma^2} + d_i\right)$$

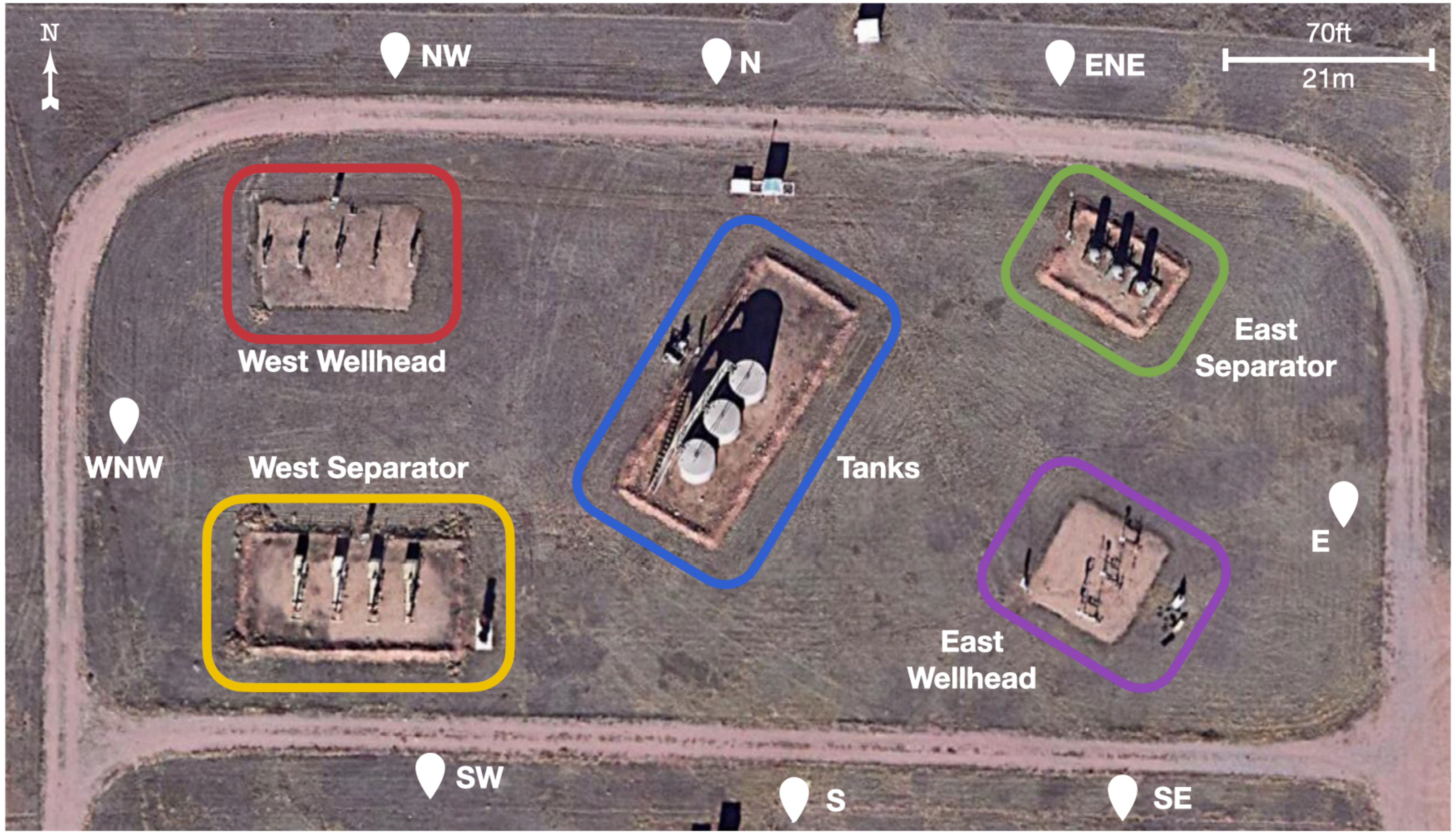
$$\beta_i|\xi \sim \begin{cases} 0 & z_i = 0 \\ \mathcal{N}\left(\left(\frac{X^T R^{-1} X}{\sigma^2}\right)^{-1}\left(\frac{X^T R^{-1} y}{\sigma^2} - \frac{e_i}{\tau_i^2 \sigma^2}\right), \left(\frac{X^T R^{-1} X}{\sigma^2}\right)^{-1}\right) & z_i = 1 \end{cases}$$

$$z_i|\xi \sim \text{Bernoulli}\left(1 - \frac{1 - \theta_i}{(1 - \theta_i) + \theta_i\left(\frac{1}{\tau_i^2 \sigma^2}\right)\exp\left(\frac{\left(\sum_{j=1}^n (w_j X_{j,i}^* + w_j^* X_{j,i}) - \frac{2}{\tau_i^2}\right)^2}{4\sigma^2 \sum_{j=1}^n X_{j,i} X_{j,i}^*}\right)\left(\frac{2\sigma^2 \pi}{\sum_{j=1}^n X_{j,i} X_{j,i}^*}\right)^{1/2}\left(\frac{1}{2}\right)}\right)$$

$$\nu|\xi \sim ? \quad \text{(Use a Metropolis–Hastings step)}$$

Iterative samples from each full conditional gives you samples from the joint posterior!

Model evaluation on multi-source controlled release data



Methane Emissions Technology Evaluation Center (METEC)

337 controlled releases:
- **99** (29%) single-source
- **238** (71%) multi-source

Emission rates range from **0.08** to **7.2** kg/hr

Emission durations range from **0.5** to **8** hours

Simulation study

Vary the degree of autocorrelation

For each controlled release, replace actual concentration observations with

$$\tilde{y} = X\beta_T + \tilde{\epsilon}$$

where β_T are the true emission rates and are errors that follow an AR(1) process.

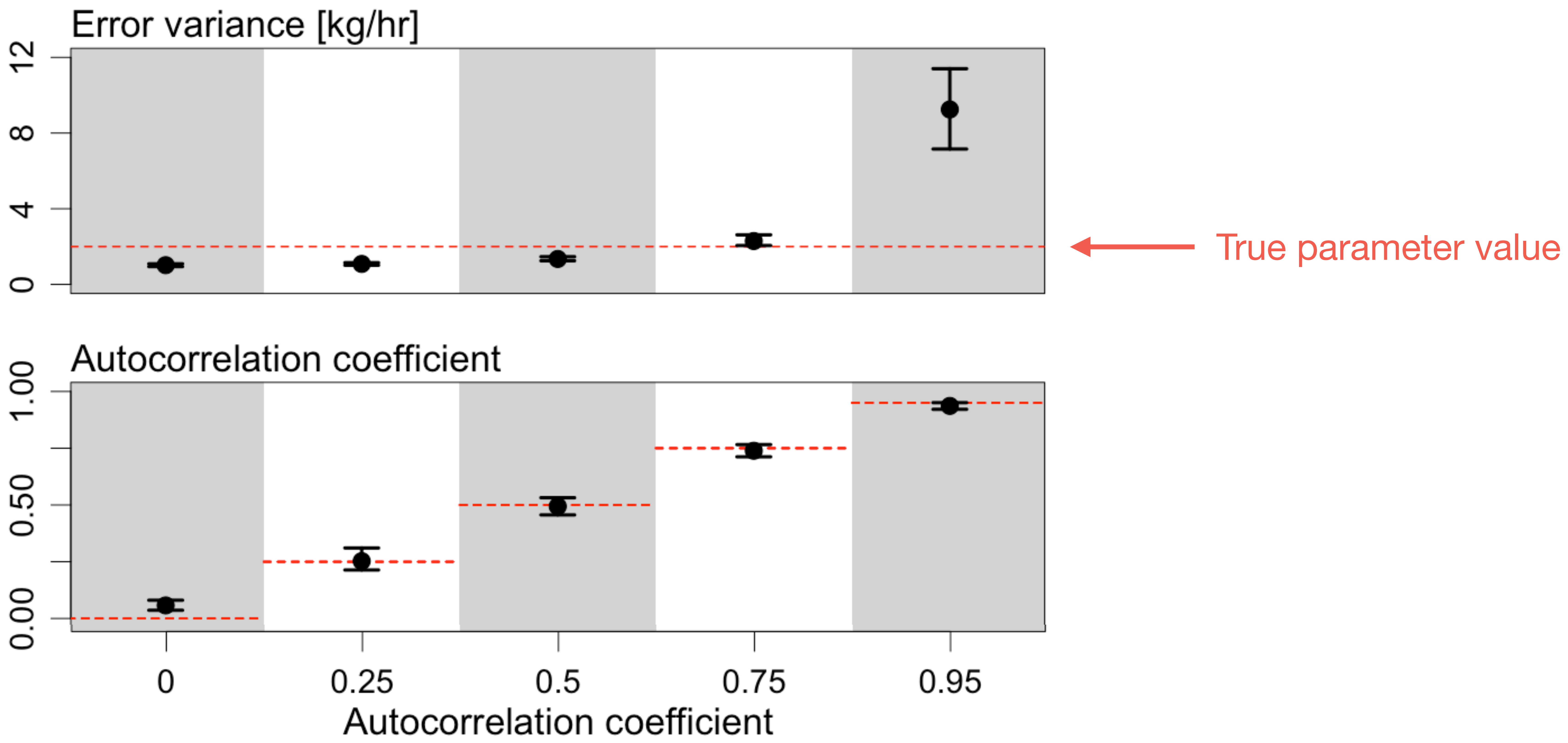
Simulation study

Vary the degree of autocorrelation

For each controlled release, replace actual concentration observations with

$$\tilde{y} = X\beta_T + \tilde{\epsilon}$$

where β_T are the true emission rates and are errors that follow an AR(1) process.



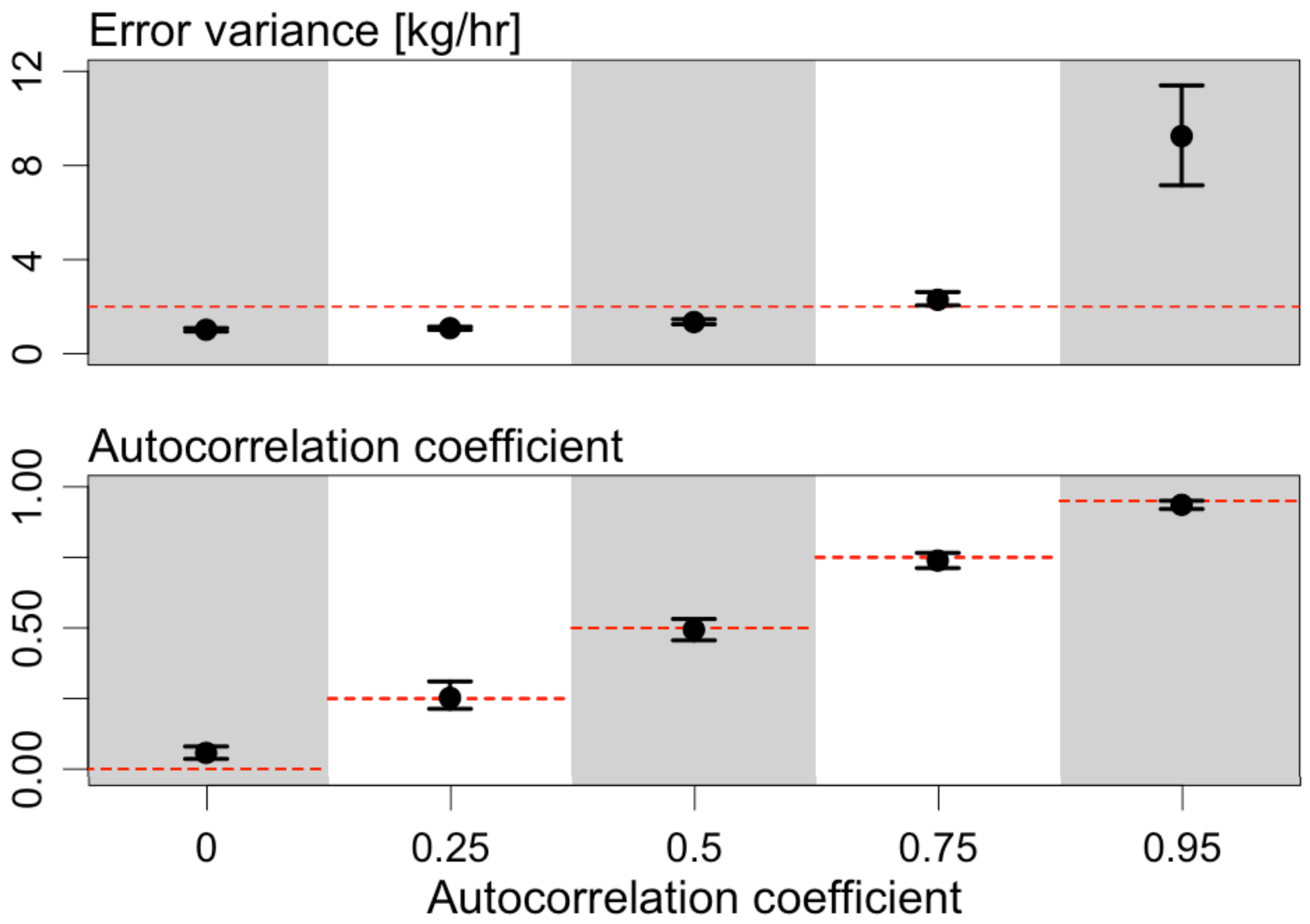
Simulation study

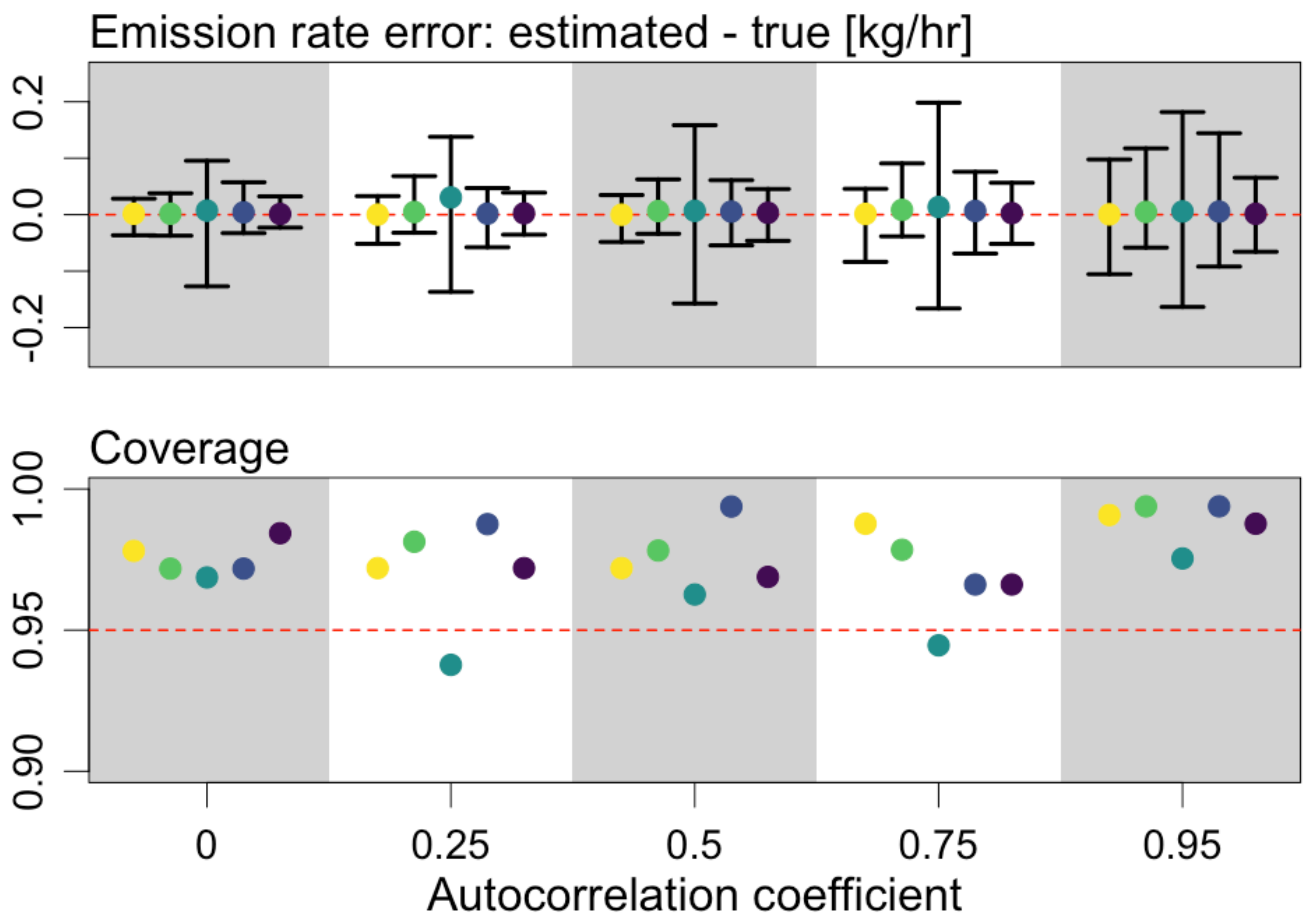
Vary the degree of autocorrelation

For each controlled release, replace actual concentration observations with

$$\tilde{y} = X\beta_T + \tilde{\epsilon}$$

where β_T are the true emission rates and are errors that follow an AR(1) process.





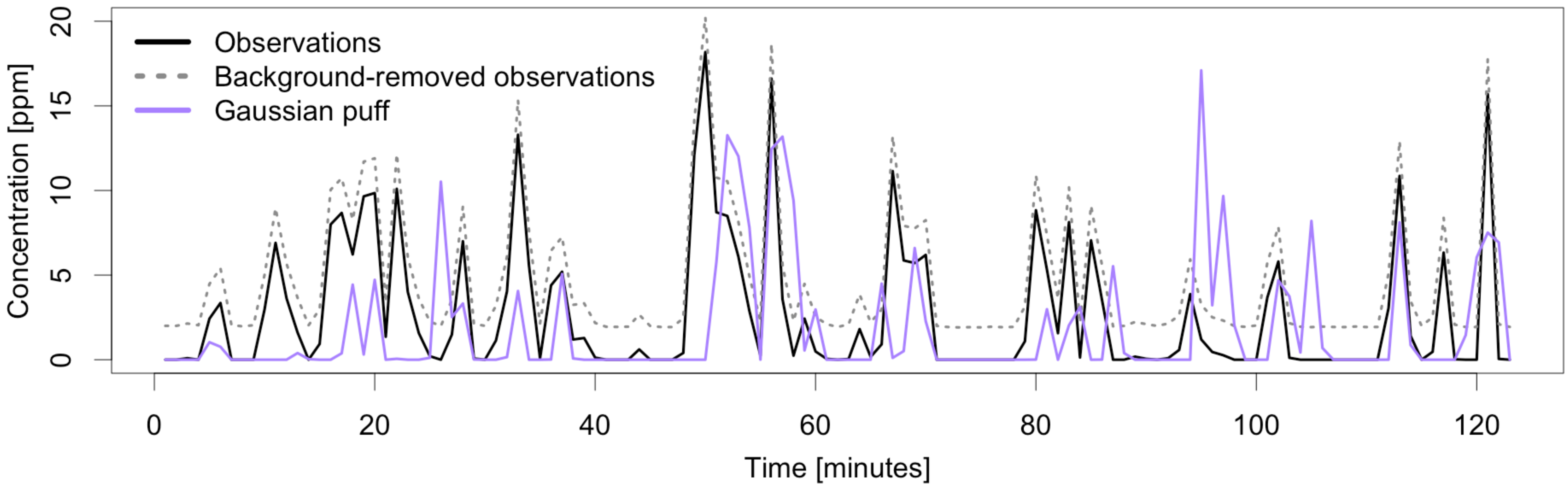
Simulation study

Vary the degree of spike misalignment

For each controlled release, replace actual concentration observations with

$$\tilde{y} = X\beta_T + \tilde{\epsilon}$$

but move a given percent of the spikes in the fake observations to a different time during the release.



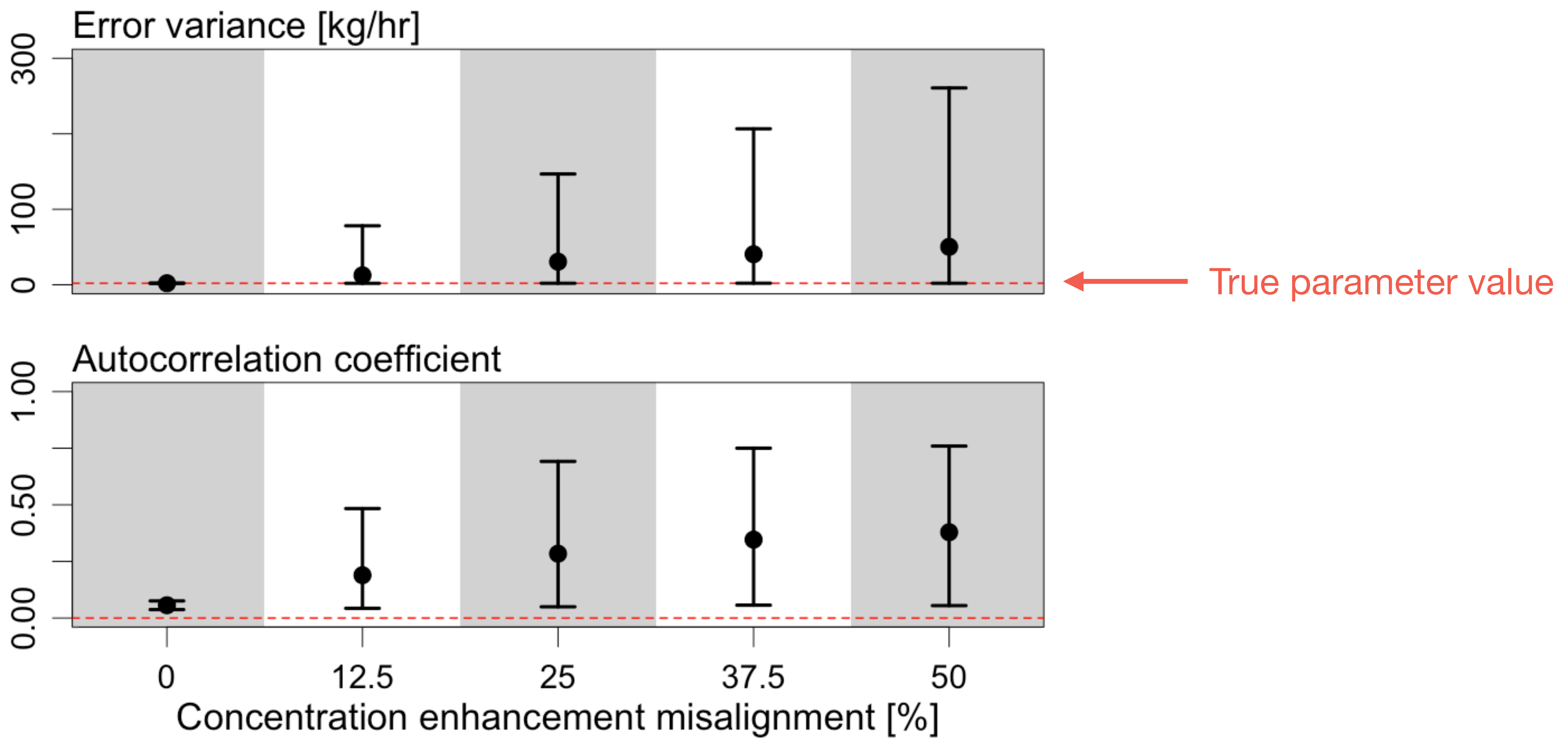
Simulation study

Vary the degree of spike misalignment

For each controlled release, replace actual concentration observations with

$$\tilde{y} = X\beta_T + \tilde{\epsilon}$$

but move a given percent of the spikes in the fake observations to a different time during the release.



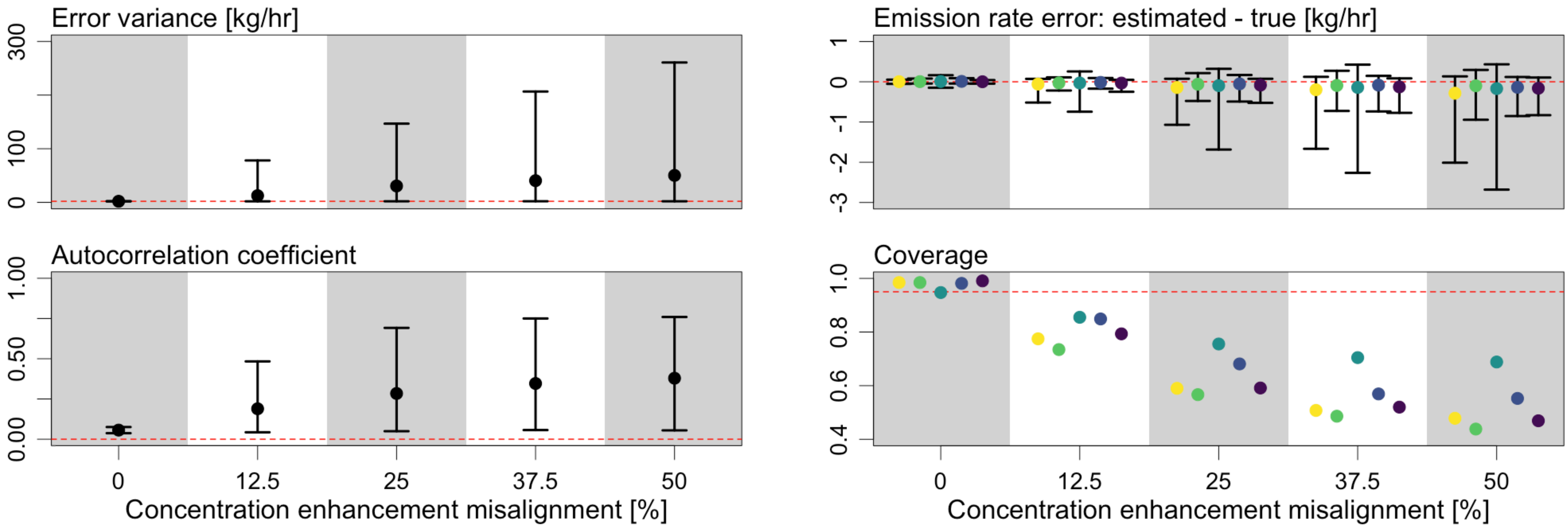
Simulation study

Vary the degree of spike misalignment

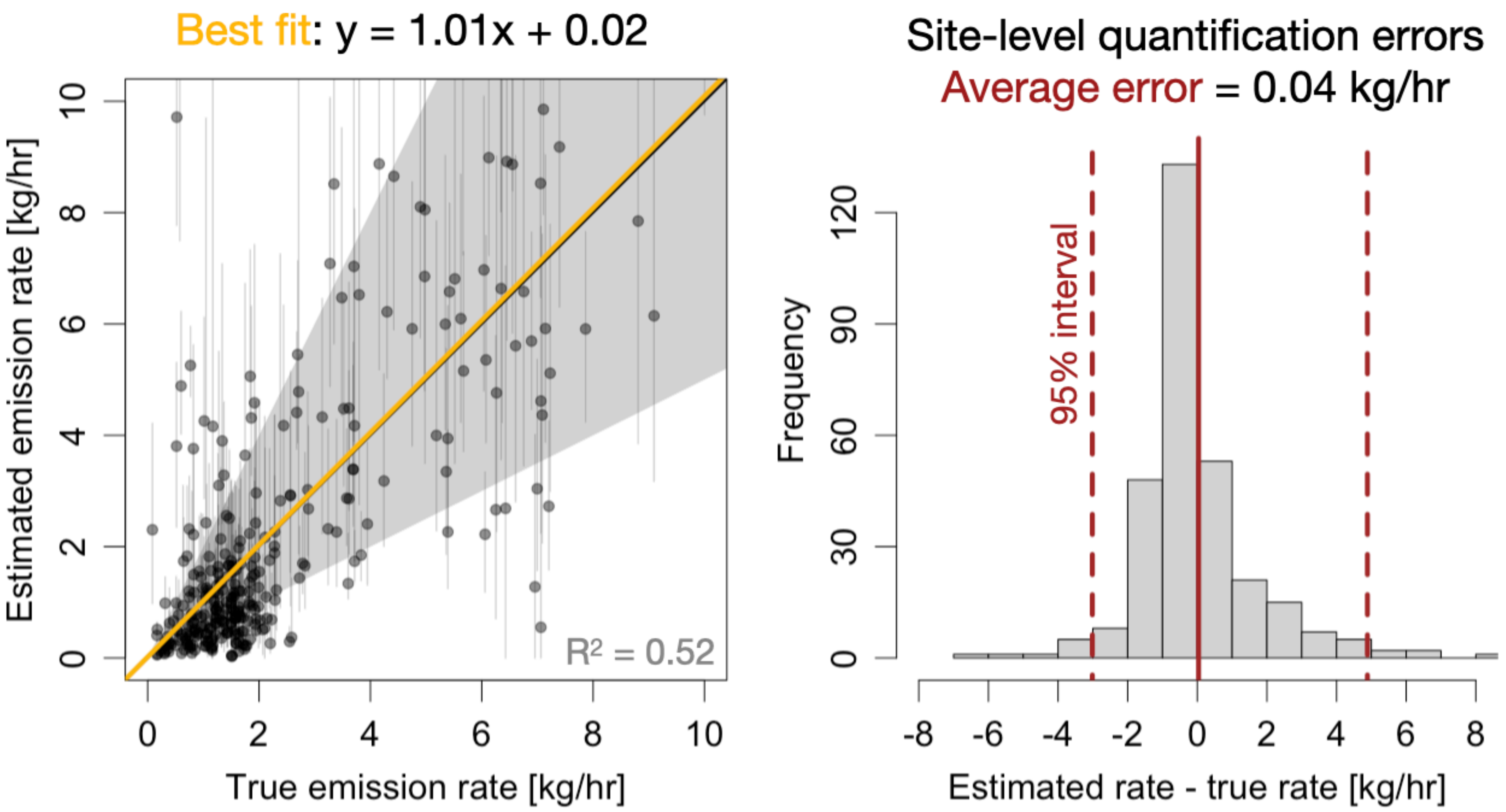
For each controlled release, replace actual concentration observations with

$$\tilde{y} = X\beta_T + \tilde{\epsilon}$$

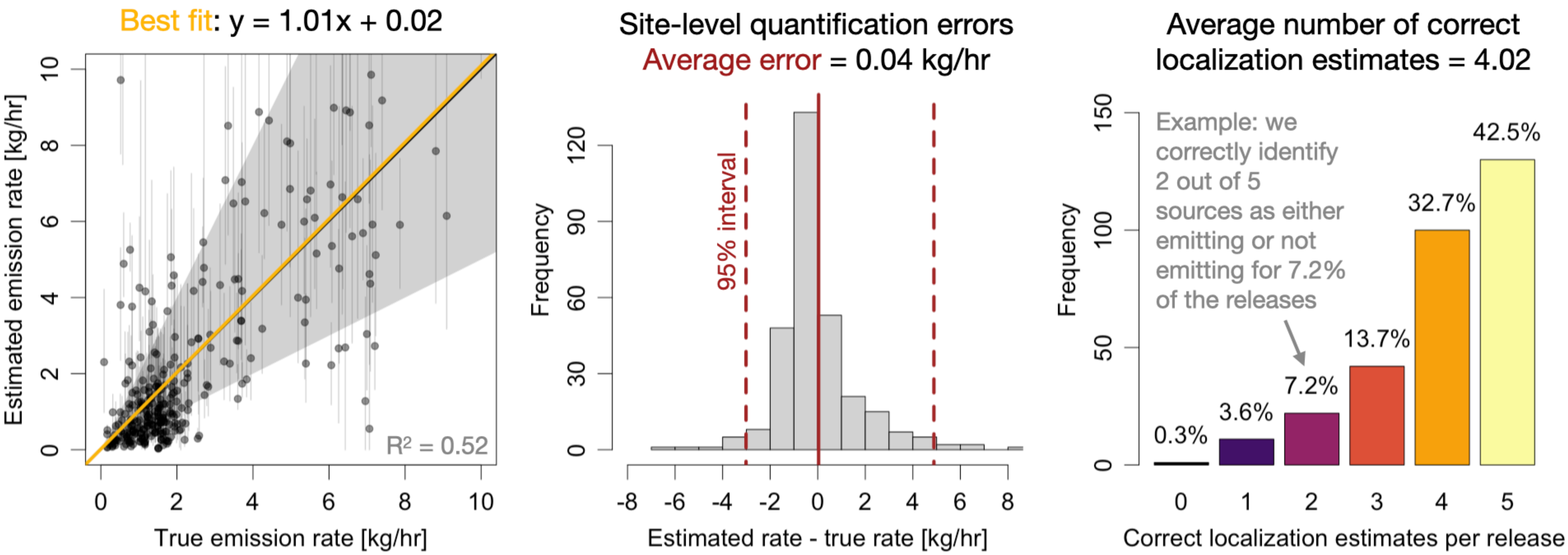
but move a given percent of the spikes in the fake observations to a different time during the release.



Model evaluation on multi-source controlled release data



Model evaluation on multi-source controlled release data



CMS Series #3:

Multi-source emission detection, localization, and quantification

A Bayesian hierarchical model for methane emission source apportionment.

William Daniels, Douglas Nychka, Dorit Hammerling.
Annals of Applied Statistics, submitted, (2025).

Thank you!

COLORADO SCHOOL OF MINES

EEMDL
Energy Emissions Modeling and Data Lab

U.S. DEPARTMENT OF ENERGY