Developing fully transparent, site-level, measurement-based inventories using continuous monitoring data

William Daniels and the
Colorado School of Mines Team

Department of Applied Mathematics and Statistics

COLORADO SCHOOL OF MINES

Developing fully transparent, site-level, measurement-based inventories using continuous monitoring data

Developing fully transparent, site-level, measurement-based inventories using continuous monitoring data

Inventory = sum (equivalent to mean) of emissions over a given time period

Developing fully transparent, site-level, measurement-based inventories using continuous monitoring data

Inventory = sum (equivalent to mean) of emissions over a given time period

Measurement-based = only use measurement data to build the inventory

Developing fully transparent, site-level, measurement-based inventories using continuous monitoring data

Inventory = sum (equivalent to mean) of emissions over a given time period

Measurement-based = only use measurement data to build the inventory

Site-level = only measurements from the specific site used to build the inventory

Developing fully transparent, site-level, measurement-based inventories using continuous monitoring data

Inventory = sum (equivalent to mean) of emissions over a given time period

Measurement-based = only use measurement data to build the inventory

Site-level = only measurements from the specific site used to build the inventory

Fully transparent = all of the methods are open source!

site-level & measurement-based

Site-level = only measurements from the specific site used to build the inventory

Measurement-based = only use measurement data to build the inventory

site-level & measurement-based

Site-level = only measurements from the specific site used to build the inventory

Measurement-based = only use measurement data to build the inventory

Advantages

- No assumptions about similar sites following similar distributions
- No potentially for underestimation to leak through from the inventory

site-level & measurement-based

Site-level = only measurements from the specific site used to build the inventory

Measurement-based = only use measurement data to build the inventory

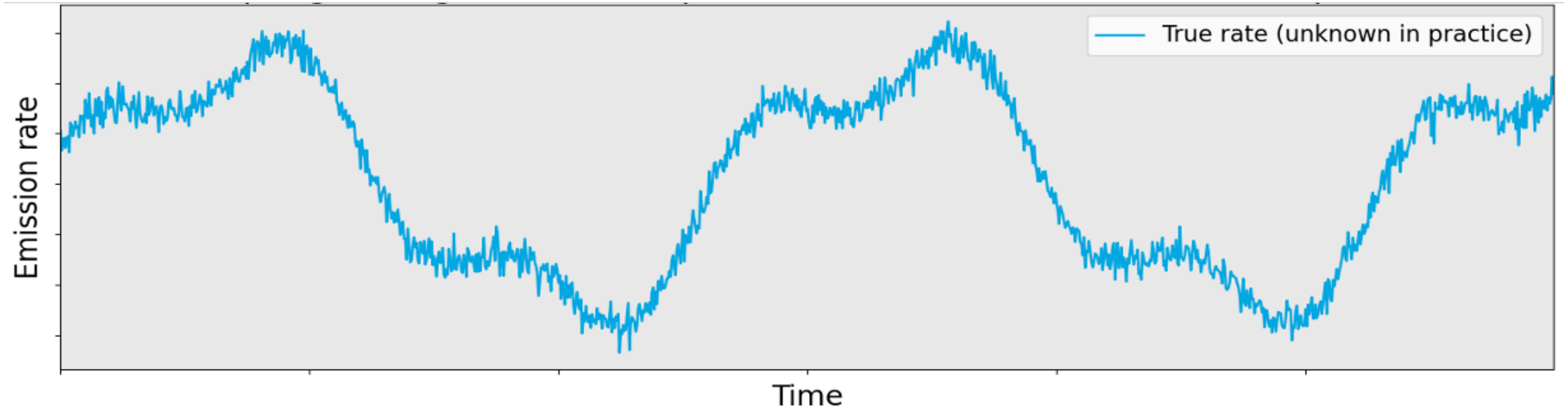
Advantages

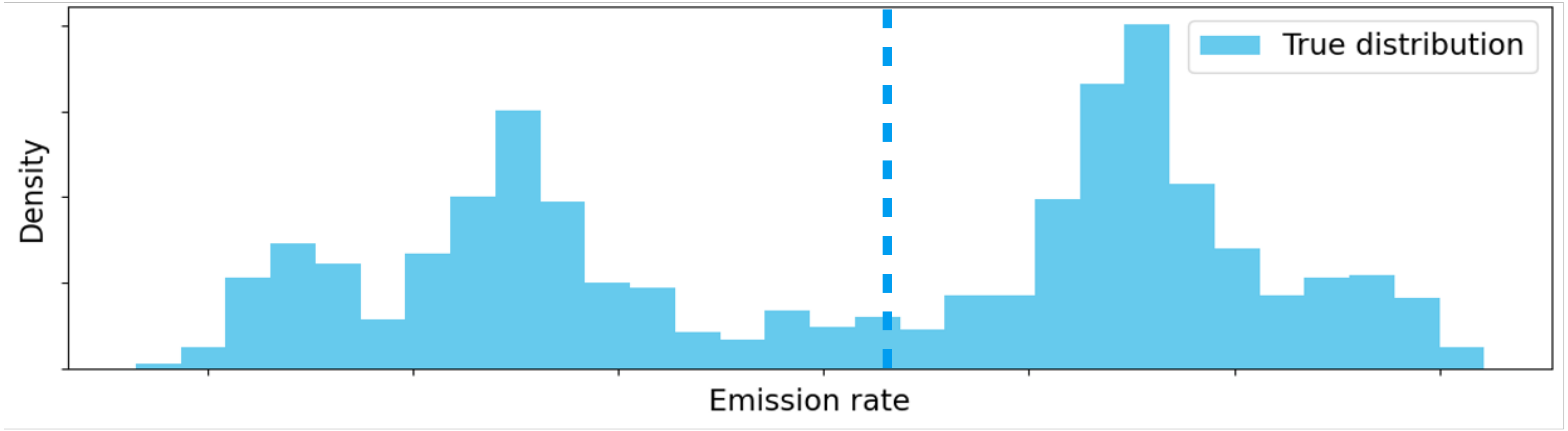
- No assumptions about similar sites following similar distributions
- No potentially for underestimation to leak through from the inventory

Challenges

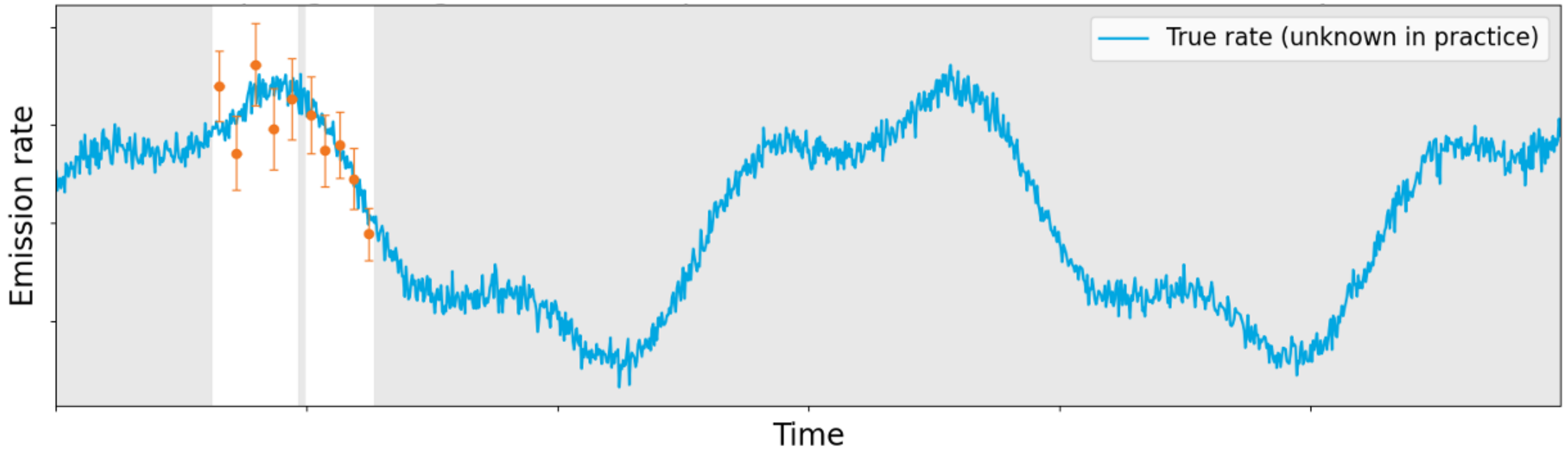
- Source estimate does not necessarily equal root cause
- Need a lot of measurements on each site
- Scaling up requires lots of measurements on lots of sites

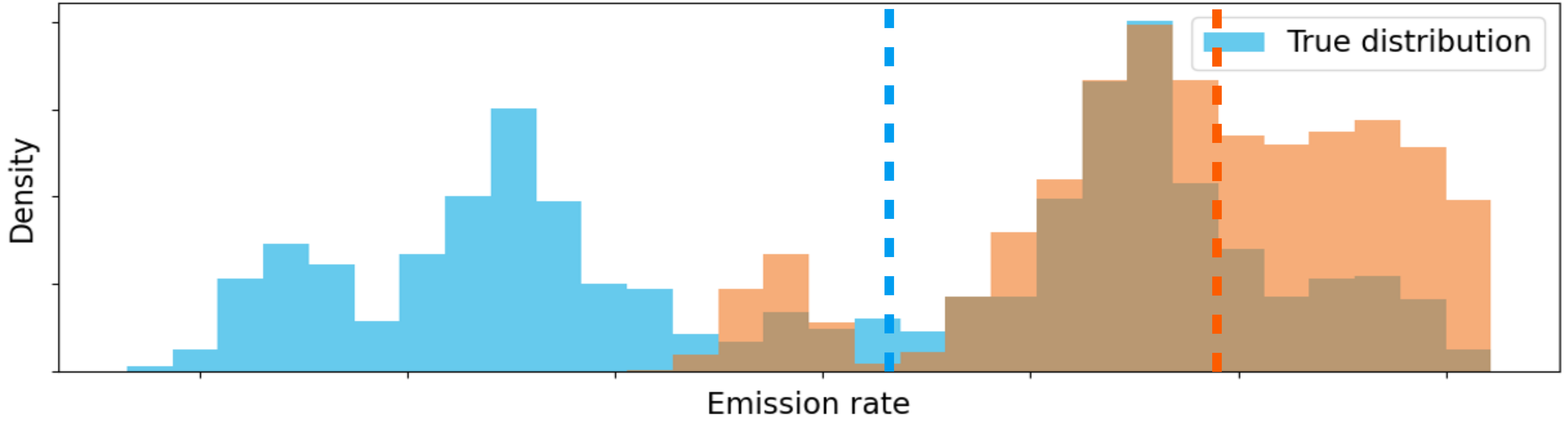
Why do you need so many measurements?



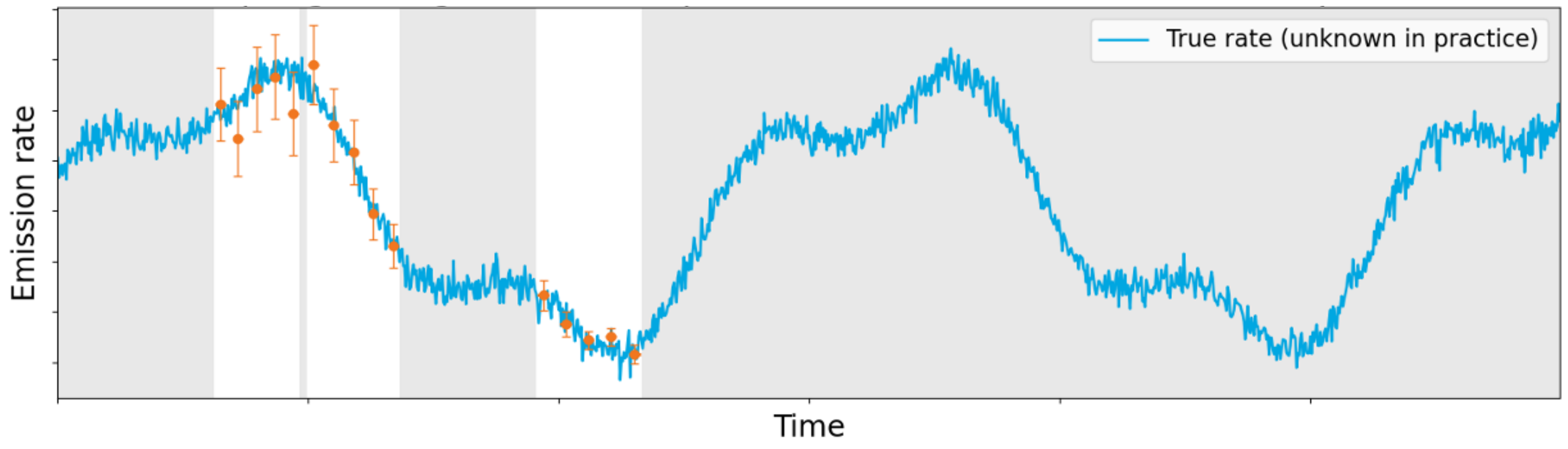


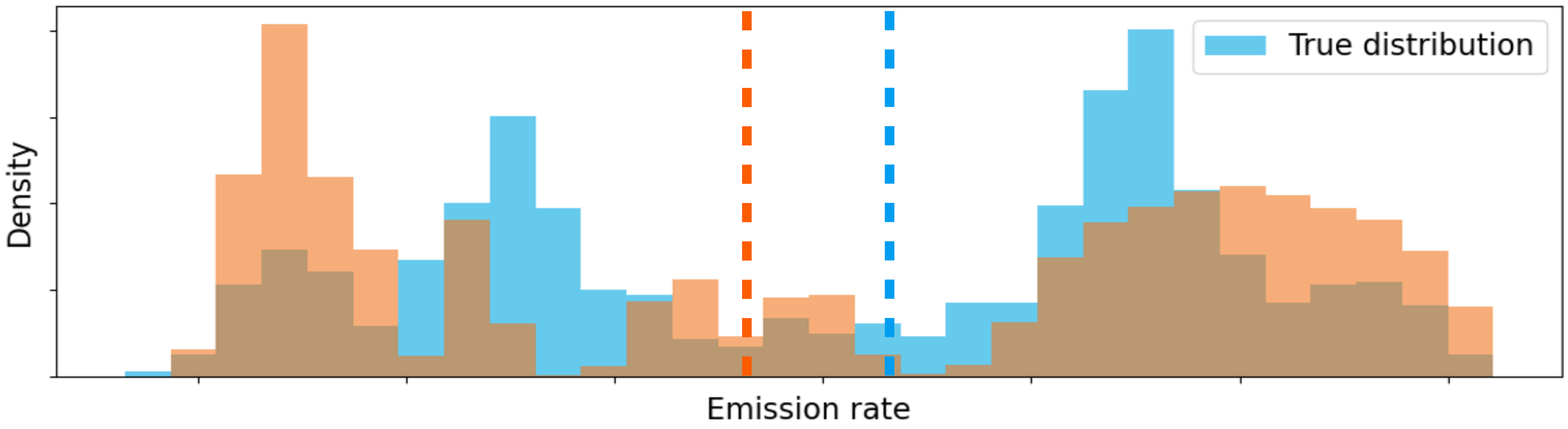
Why do you need so many measurements?



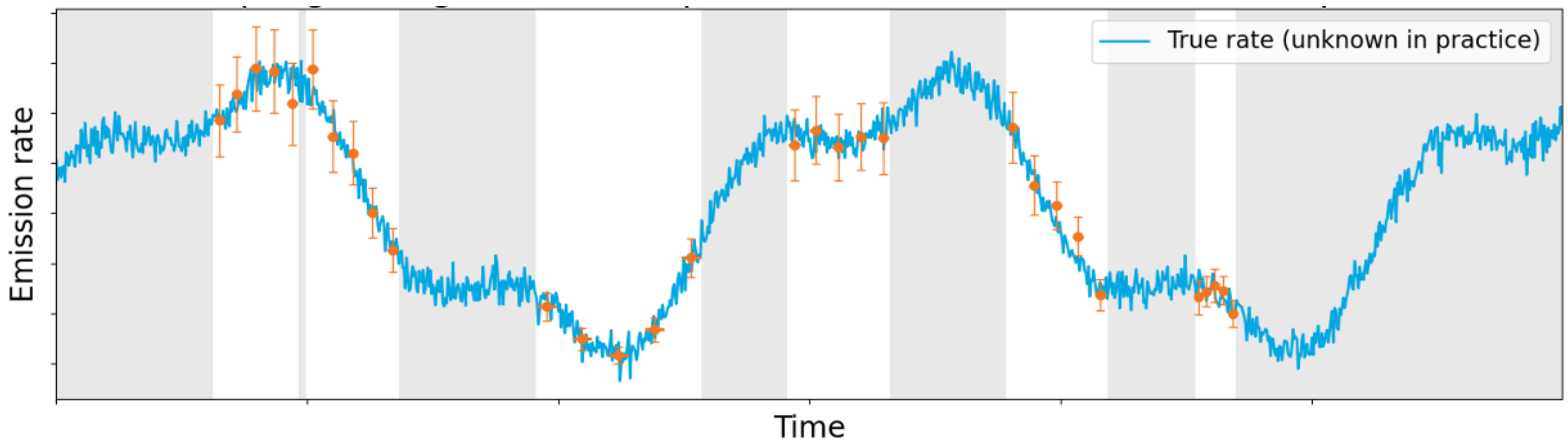


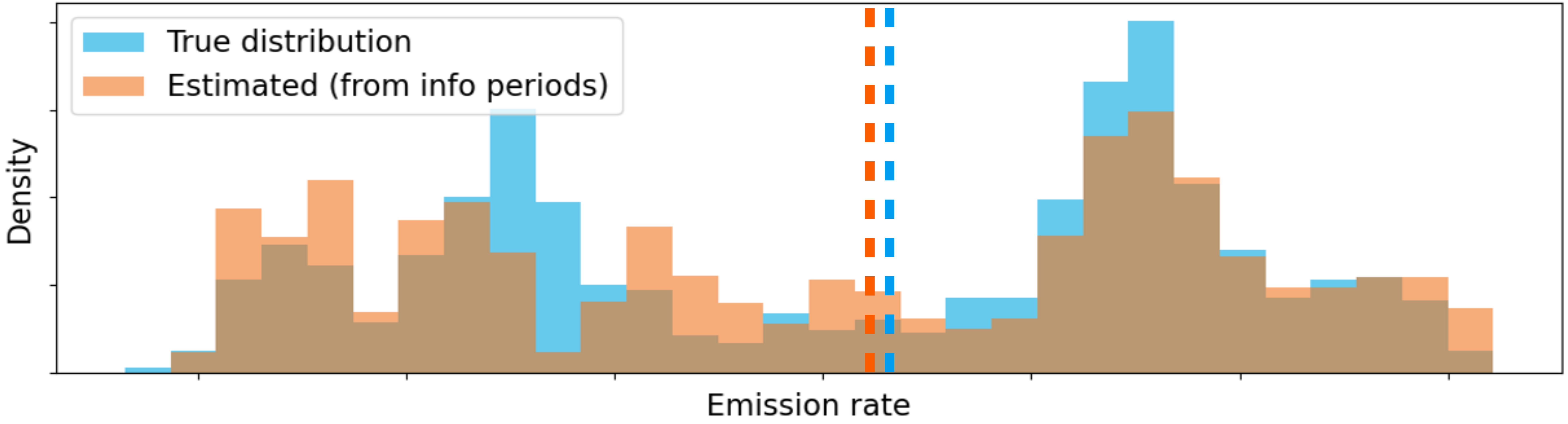
Why do you need so many measurements?



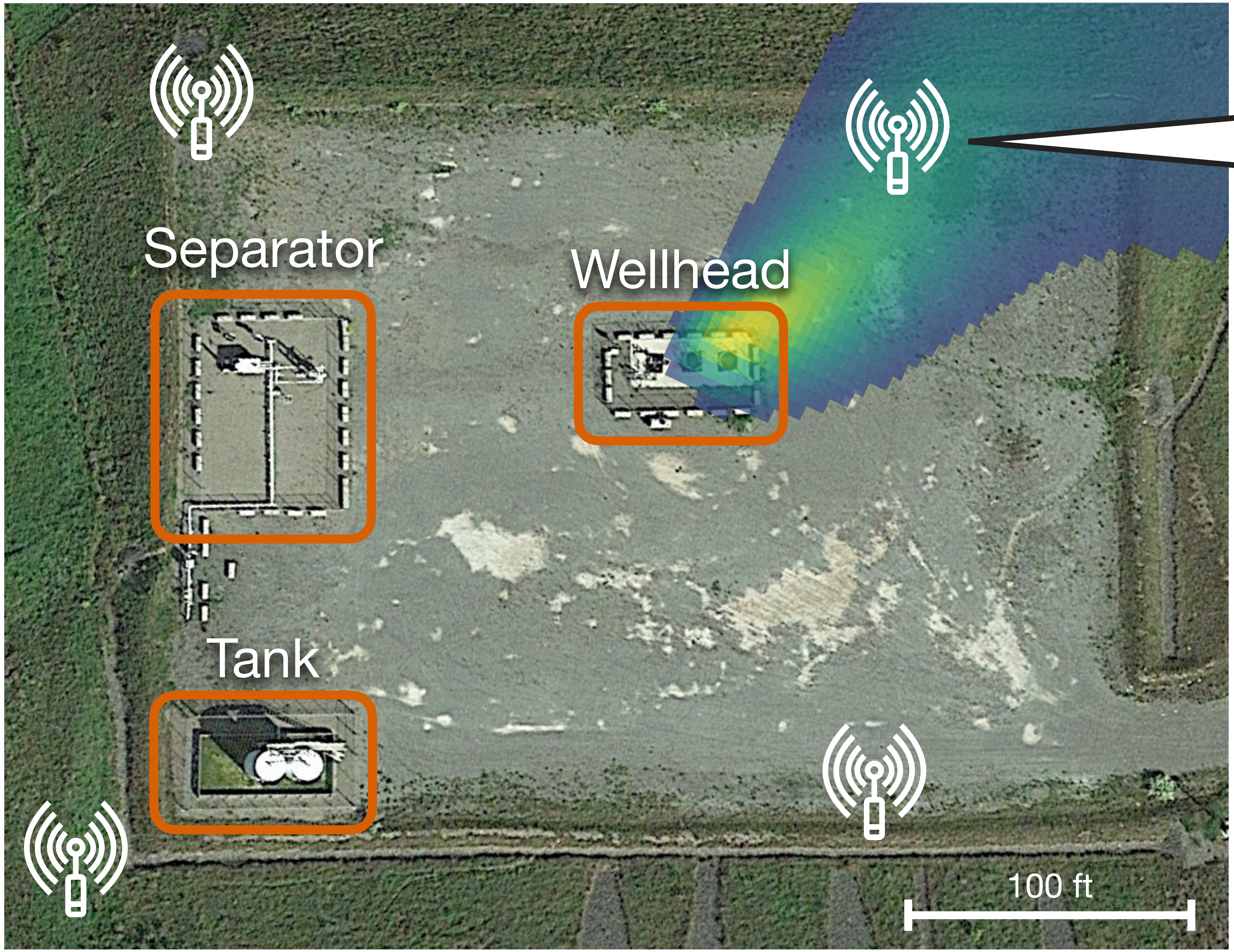


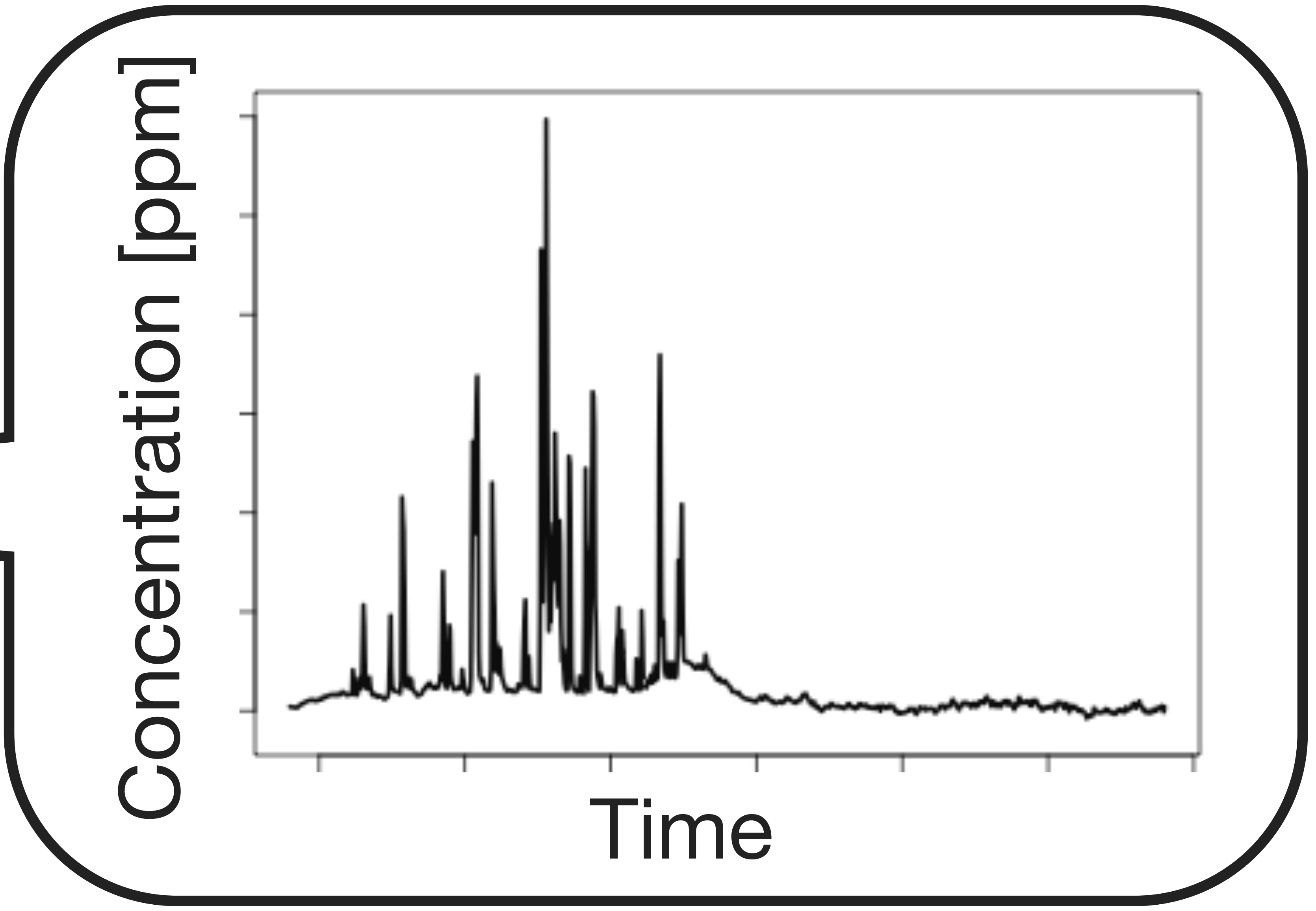
Why do you need so many measurements?





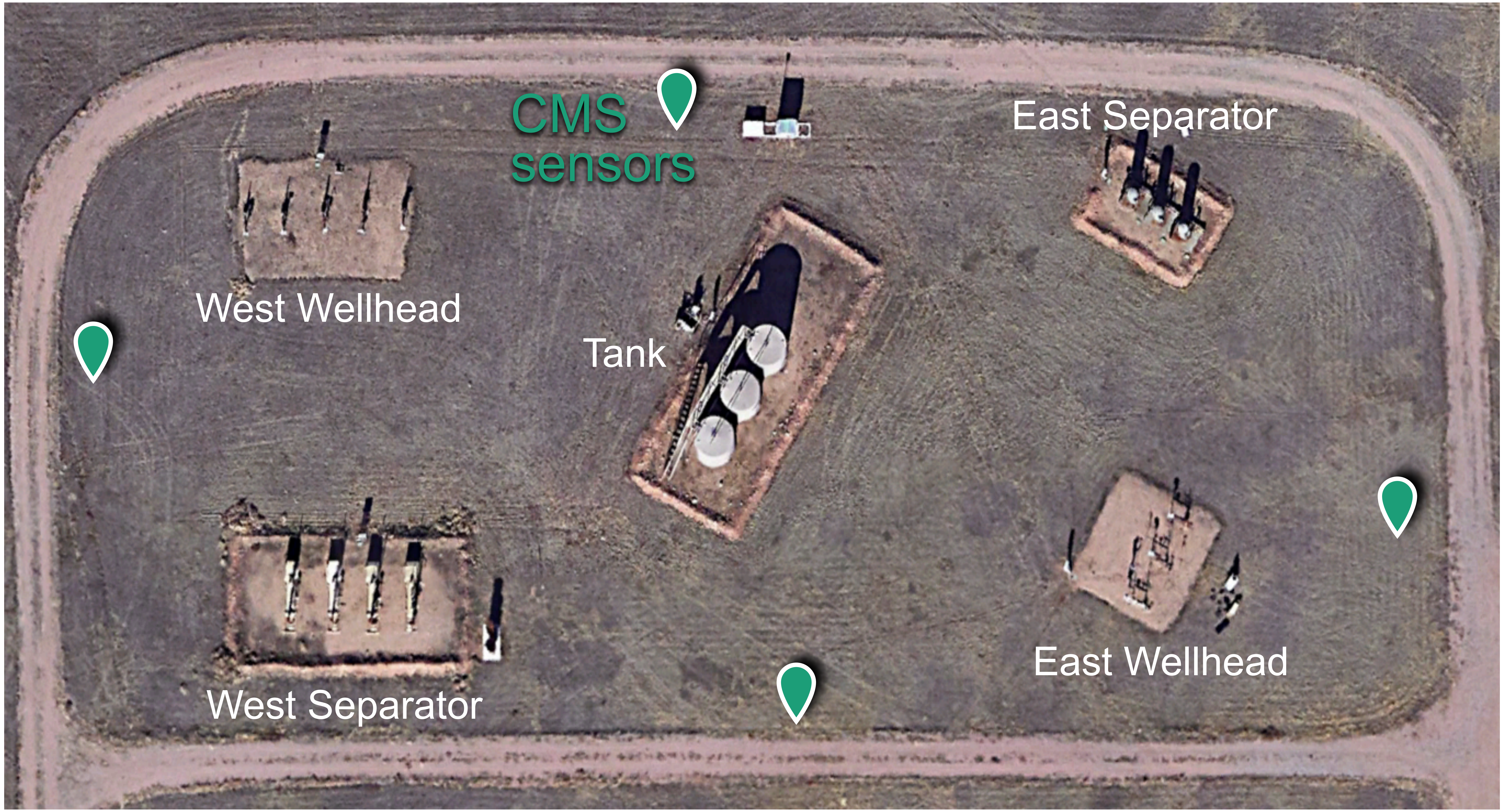
Continuous monitoring 101



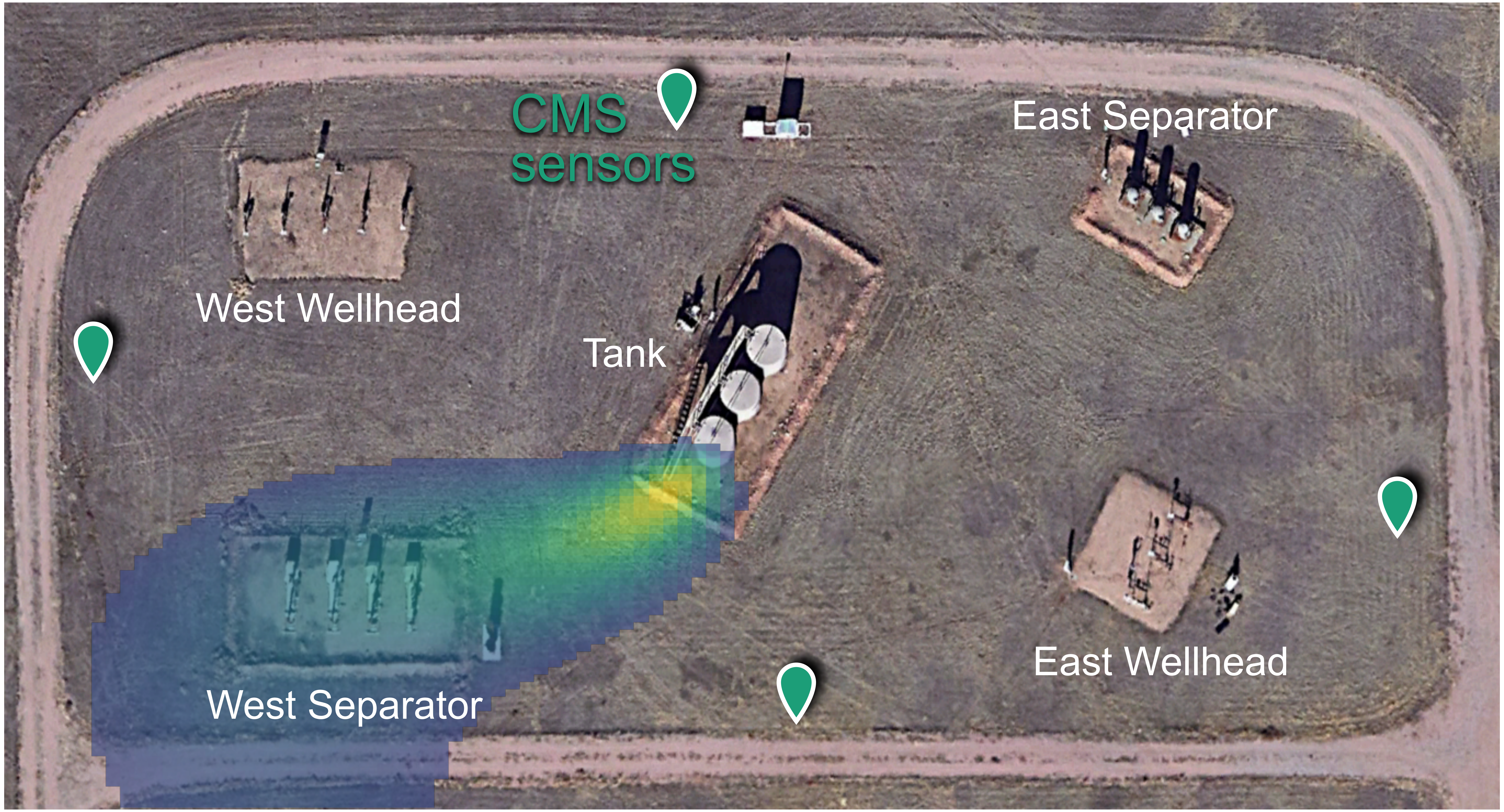


Aerial measurement technology

“Continuous” is rarely truly continuous

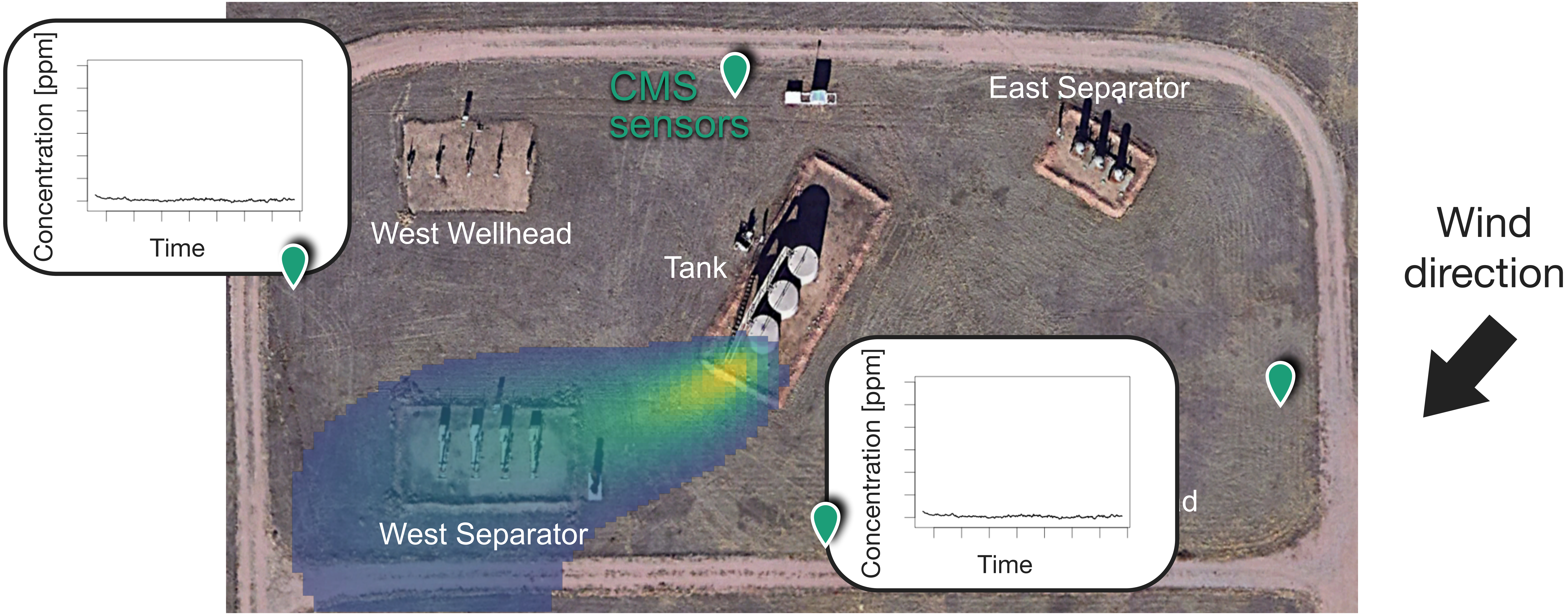


“Continuous” is rarely truly continuous



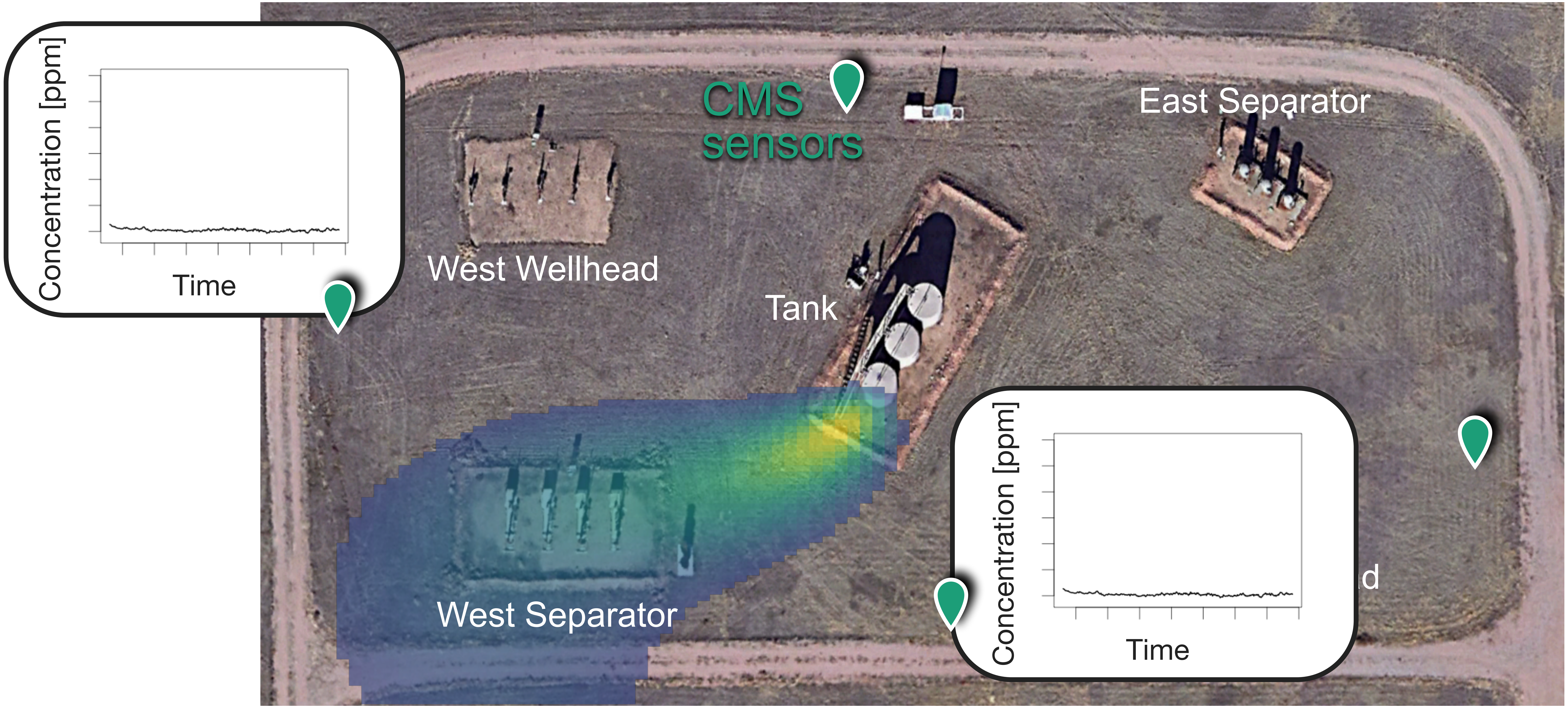
Wind
direction

“Continuous” is rarely truly continuous



No emissions information when the wind blows between sensors

"Continuous" is rarely truly continuous



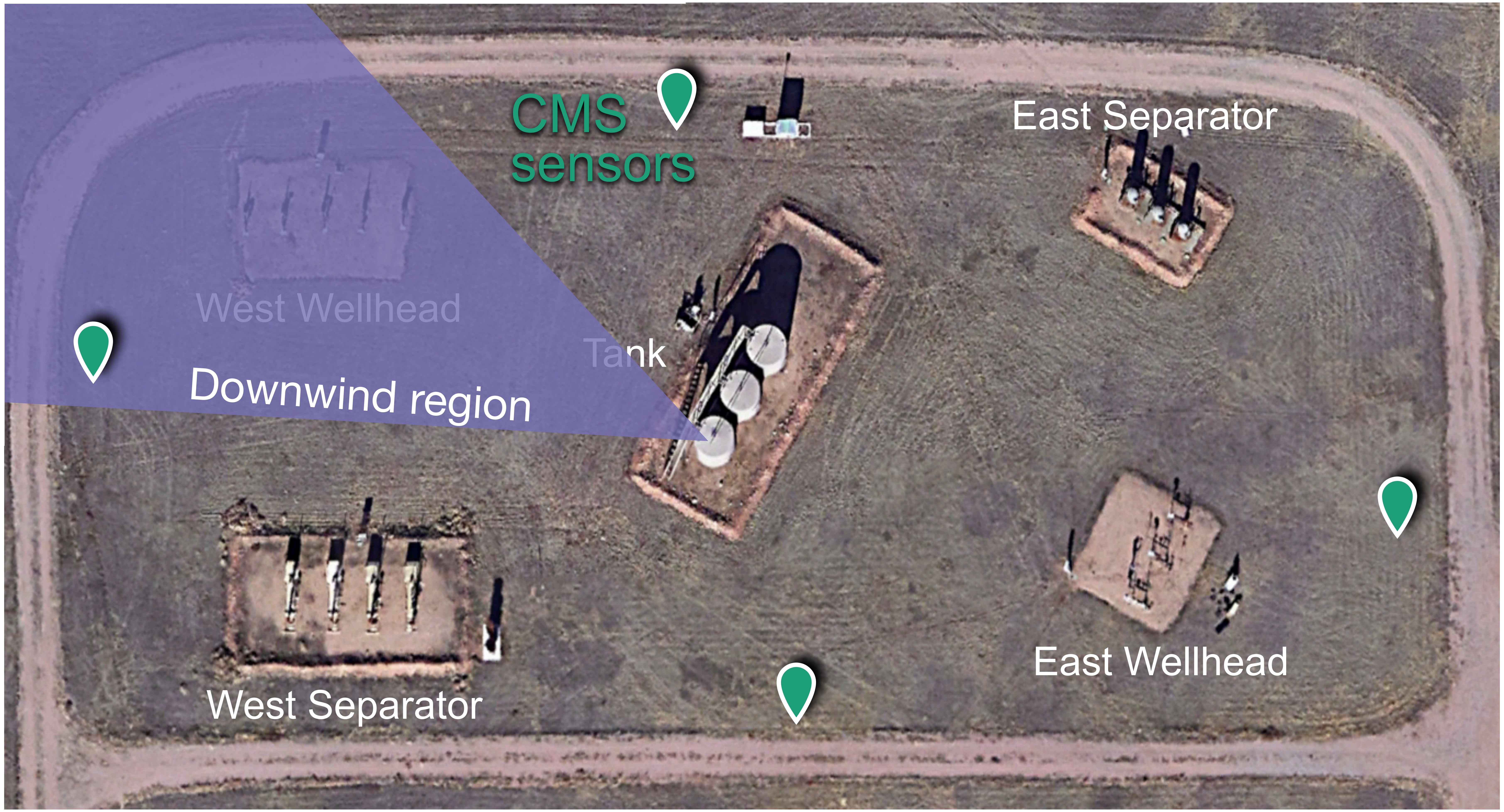
Wind direction

This cannot be interpreted as 0 kg/hr!

However… we can estimate when this happens!

Downwind region **does not** overlap with CMS sensors
= period of "**no information**"

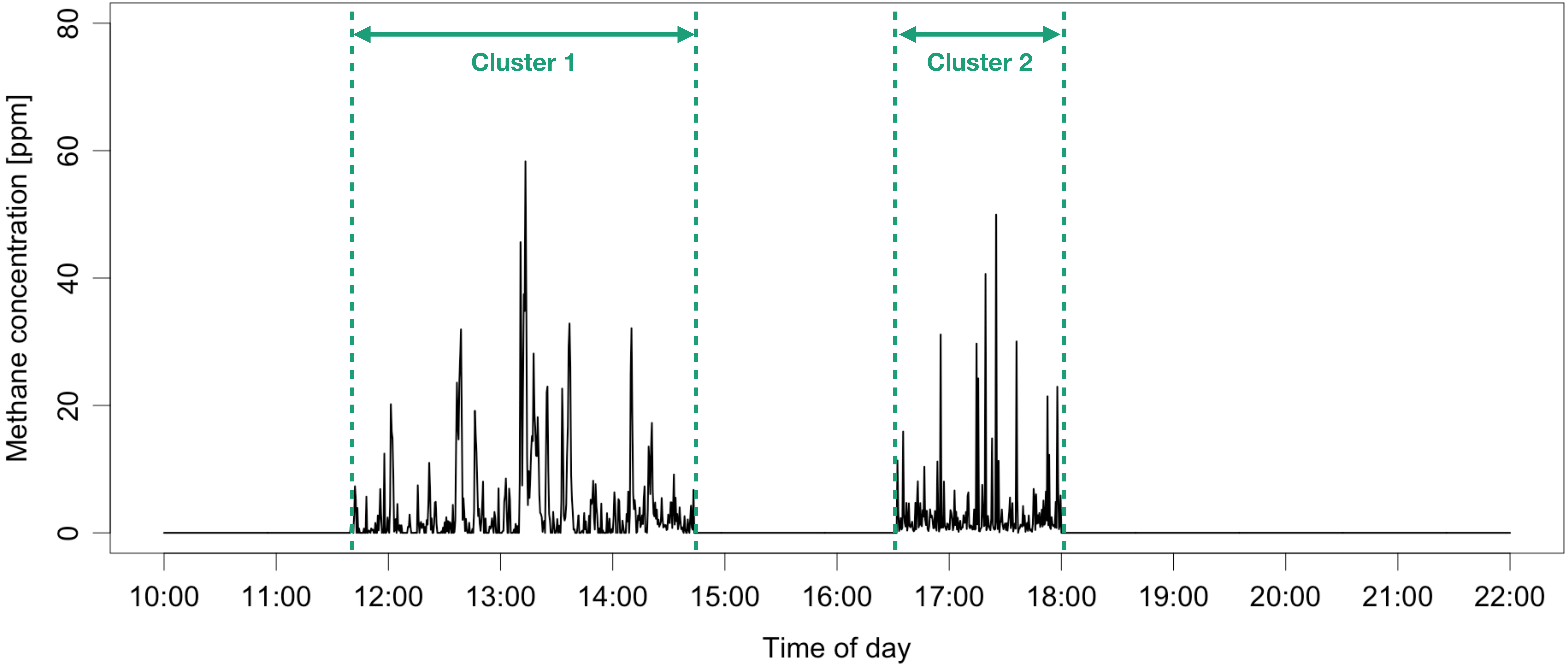
However… we can estimate when this happens!



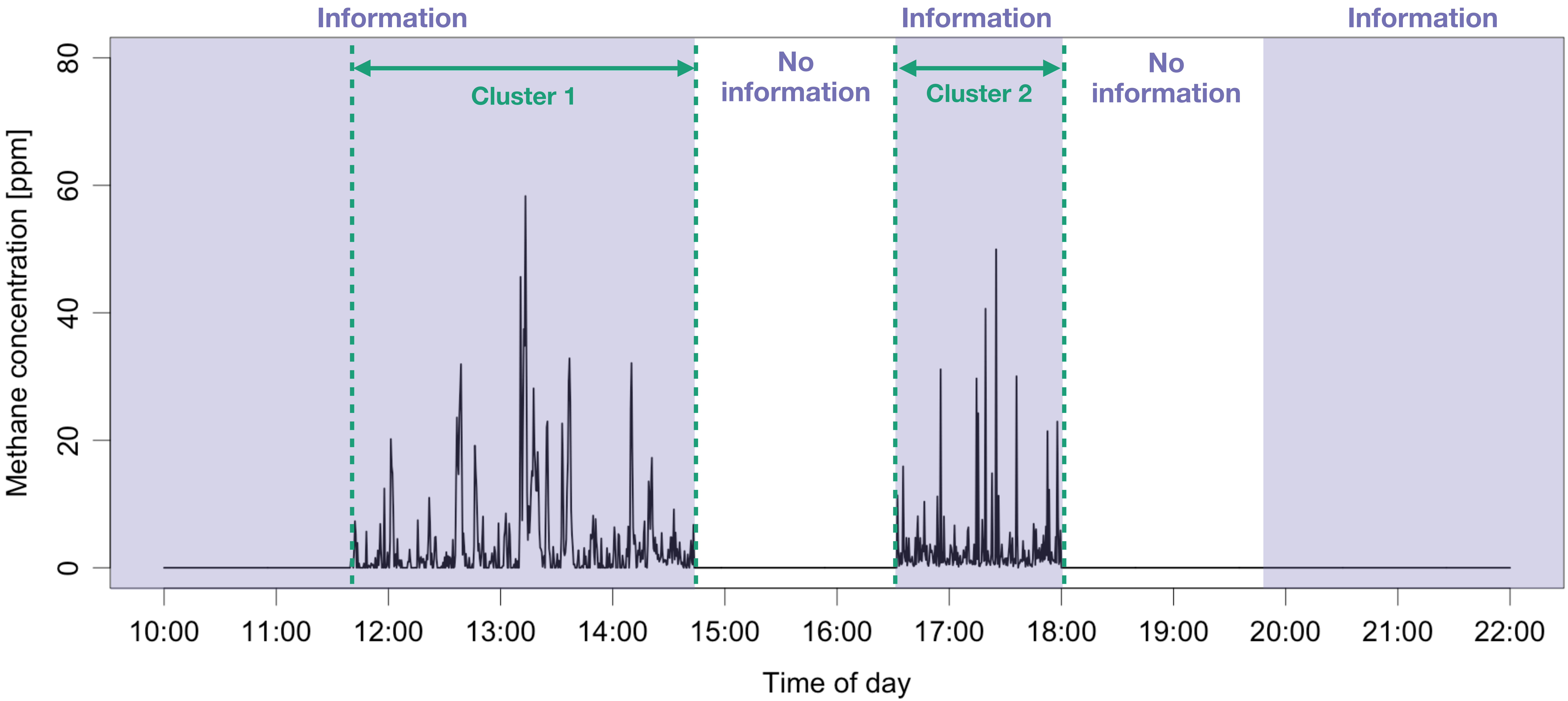
Wind
direction

Downwind region **does** overlap with CMS sensors
= period of "**information**"

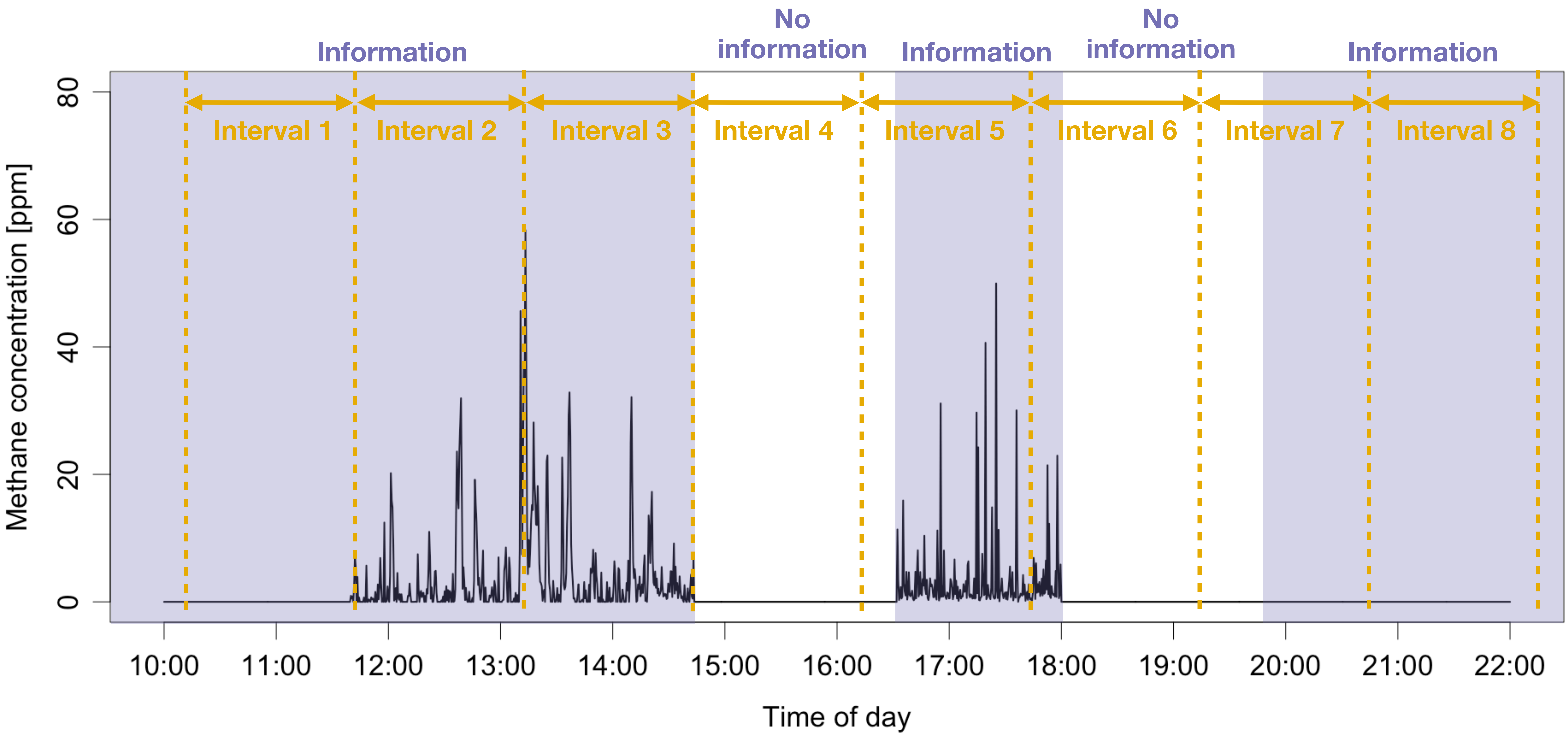
How do periods of information and no information present themselves in the data?



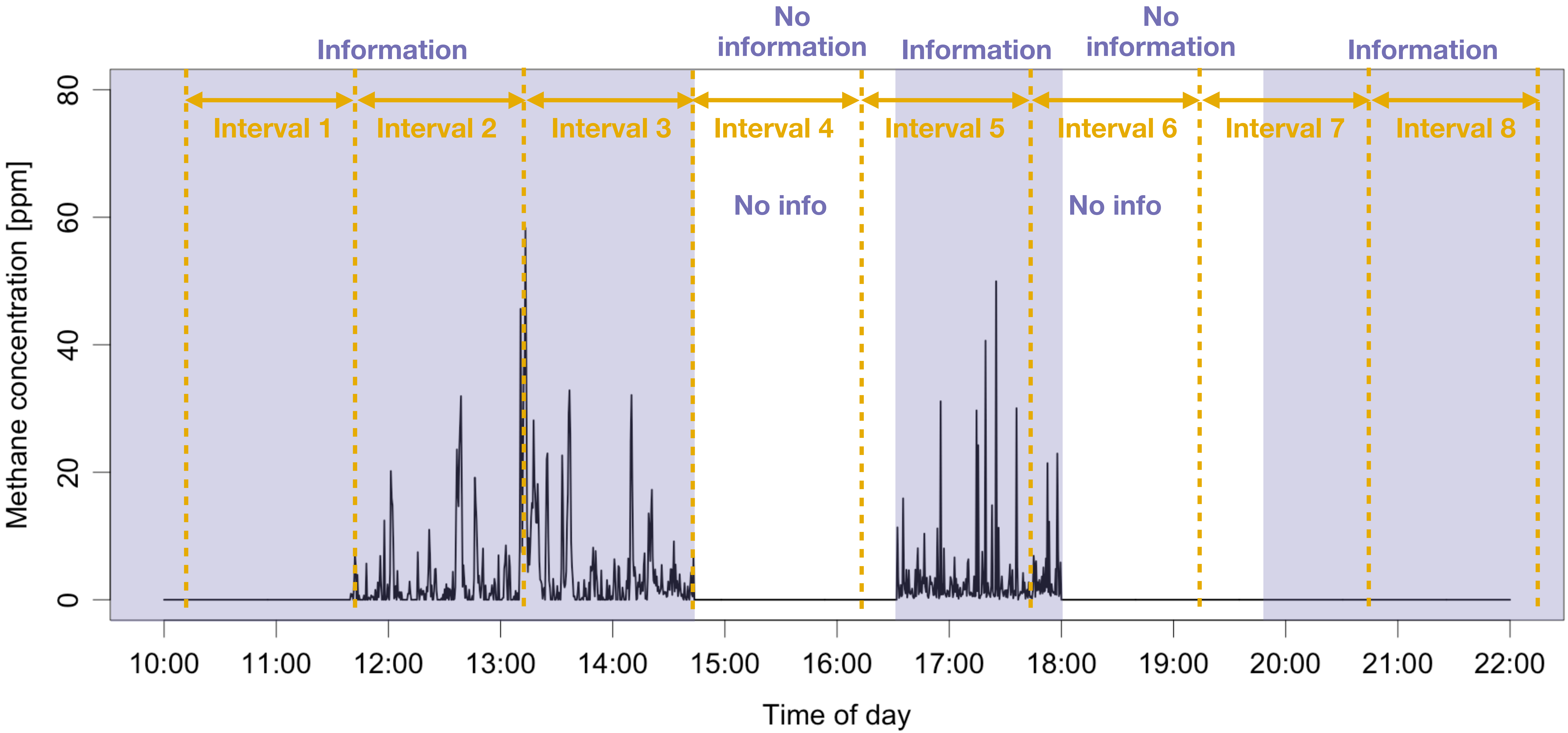
How do periods of information and no information present themselves in the data?



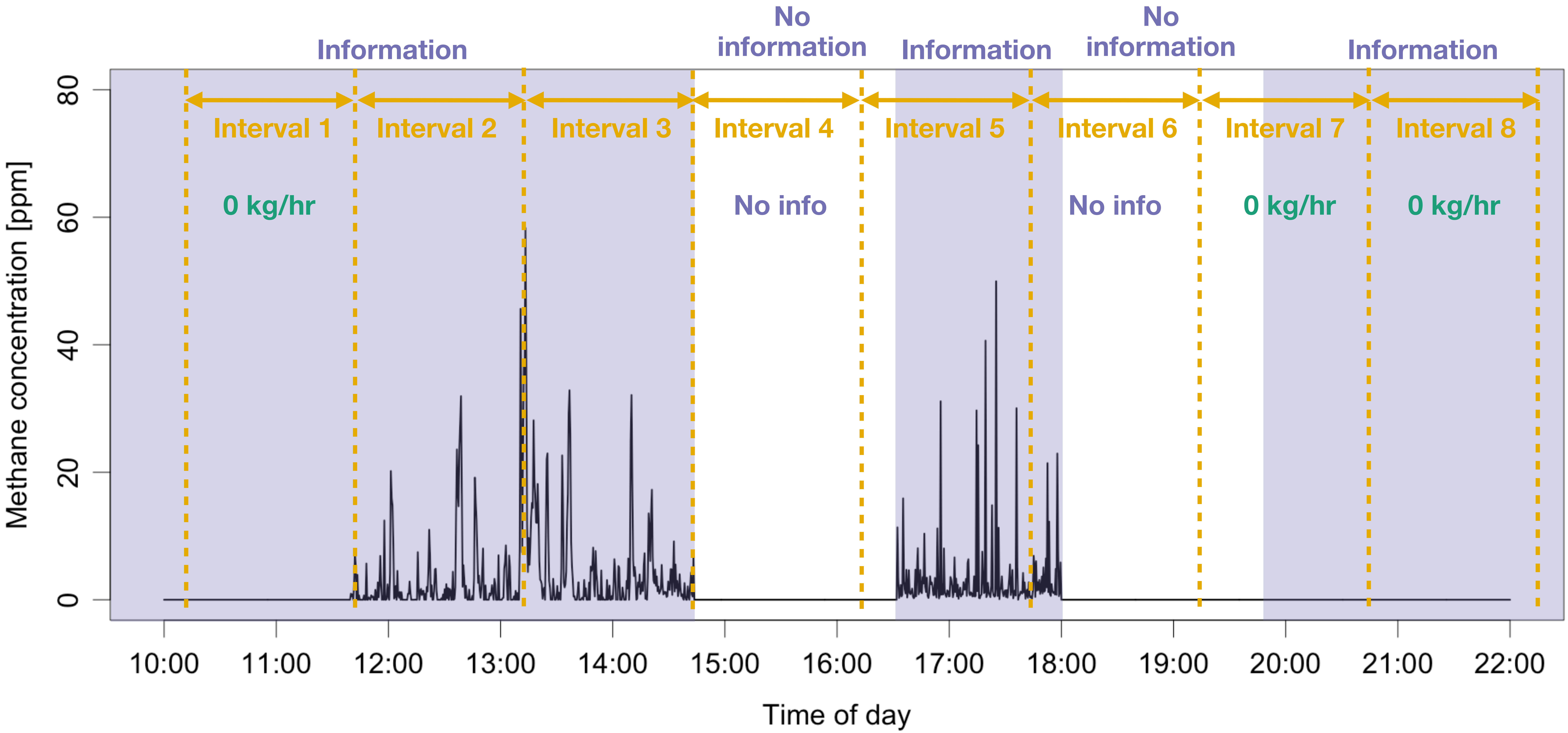
In practice, we estimate emissions on fixed fixed intervals



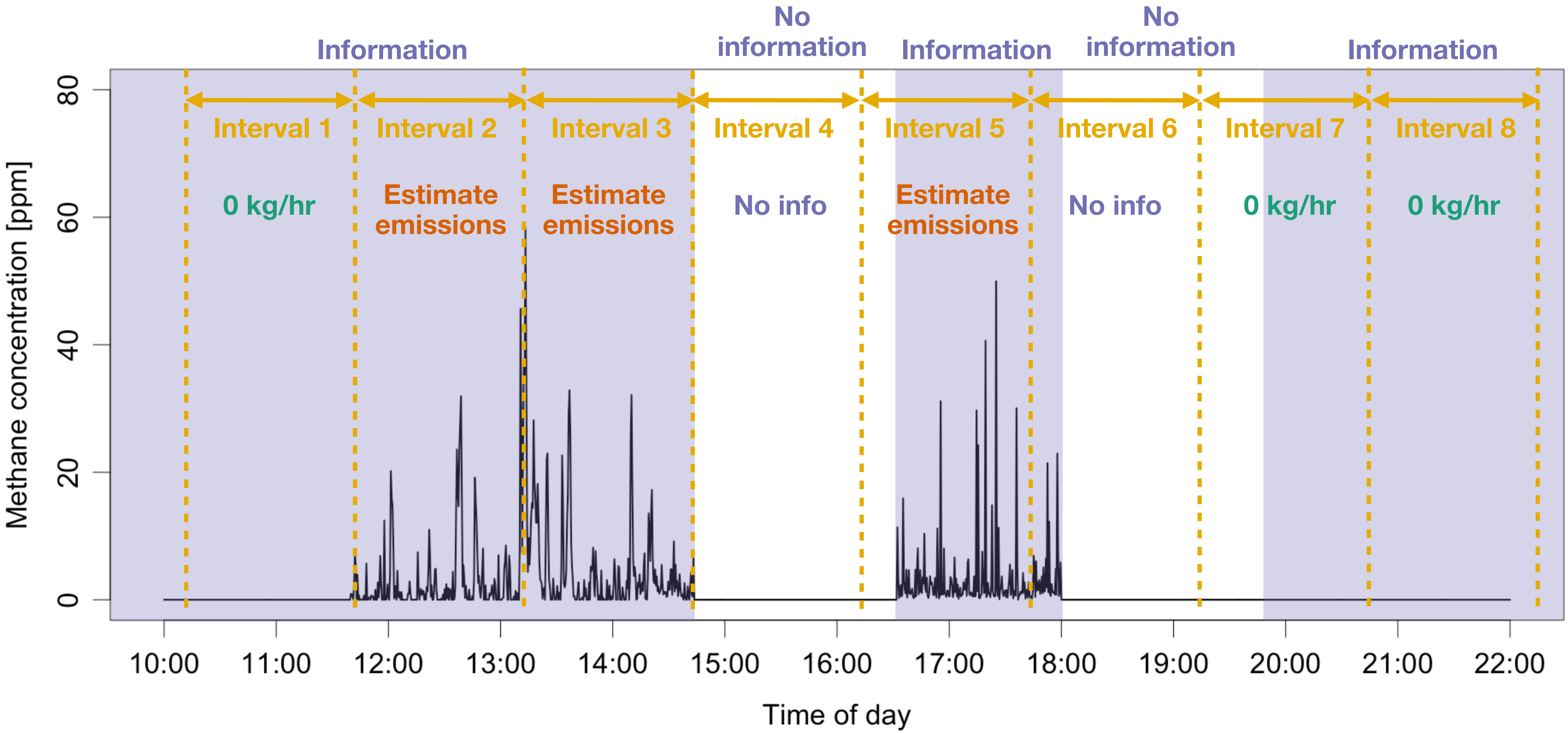
Before building an inventory, we need to identify when an interval has no information,



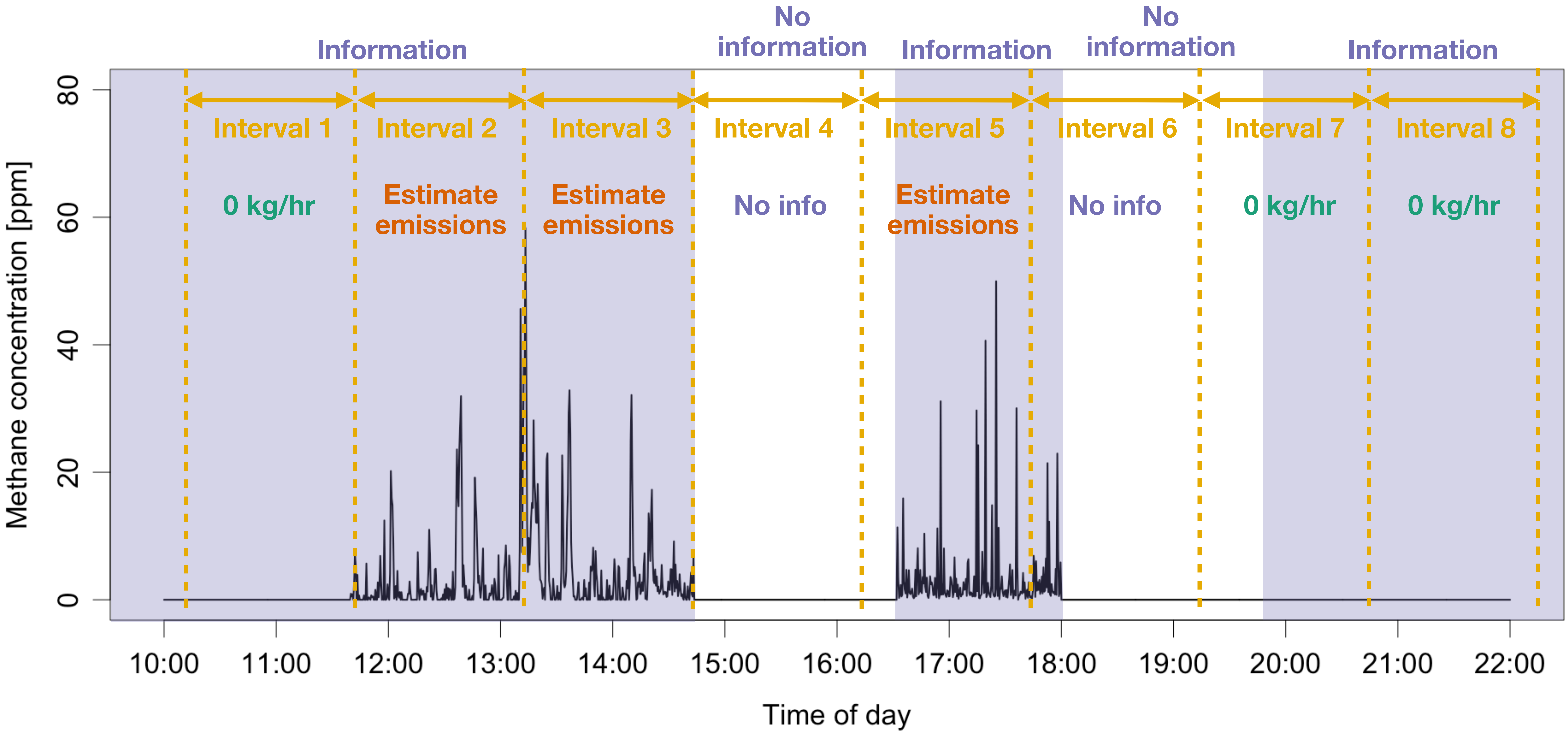
Before building an inventory, we need to identify when an interval has no information, no emissions,

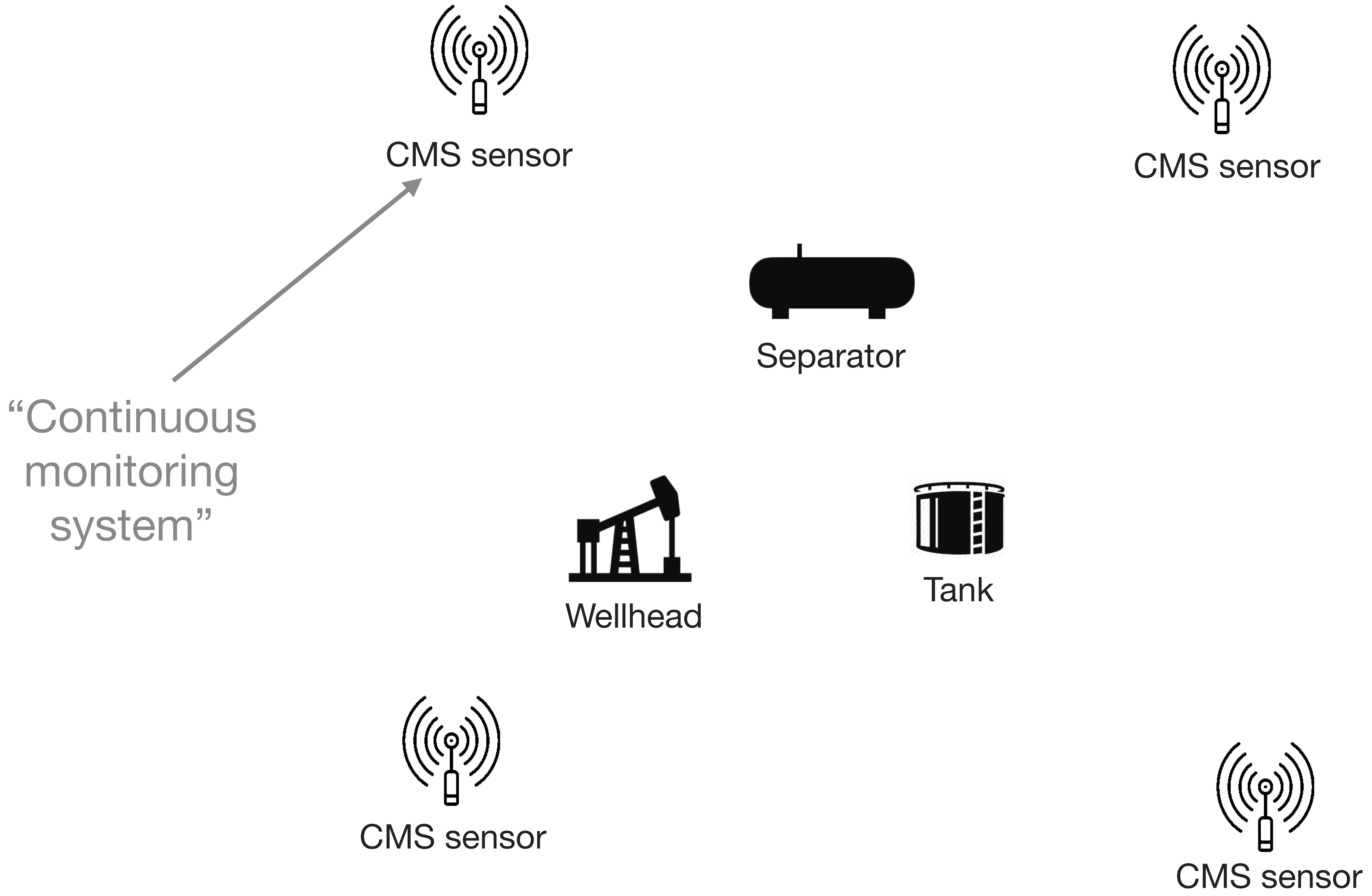


Before building an inventory, we need to identify when an interval has no information, no emissions, or a non-zero emission

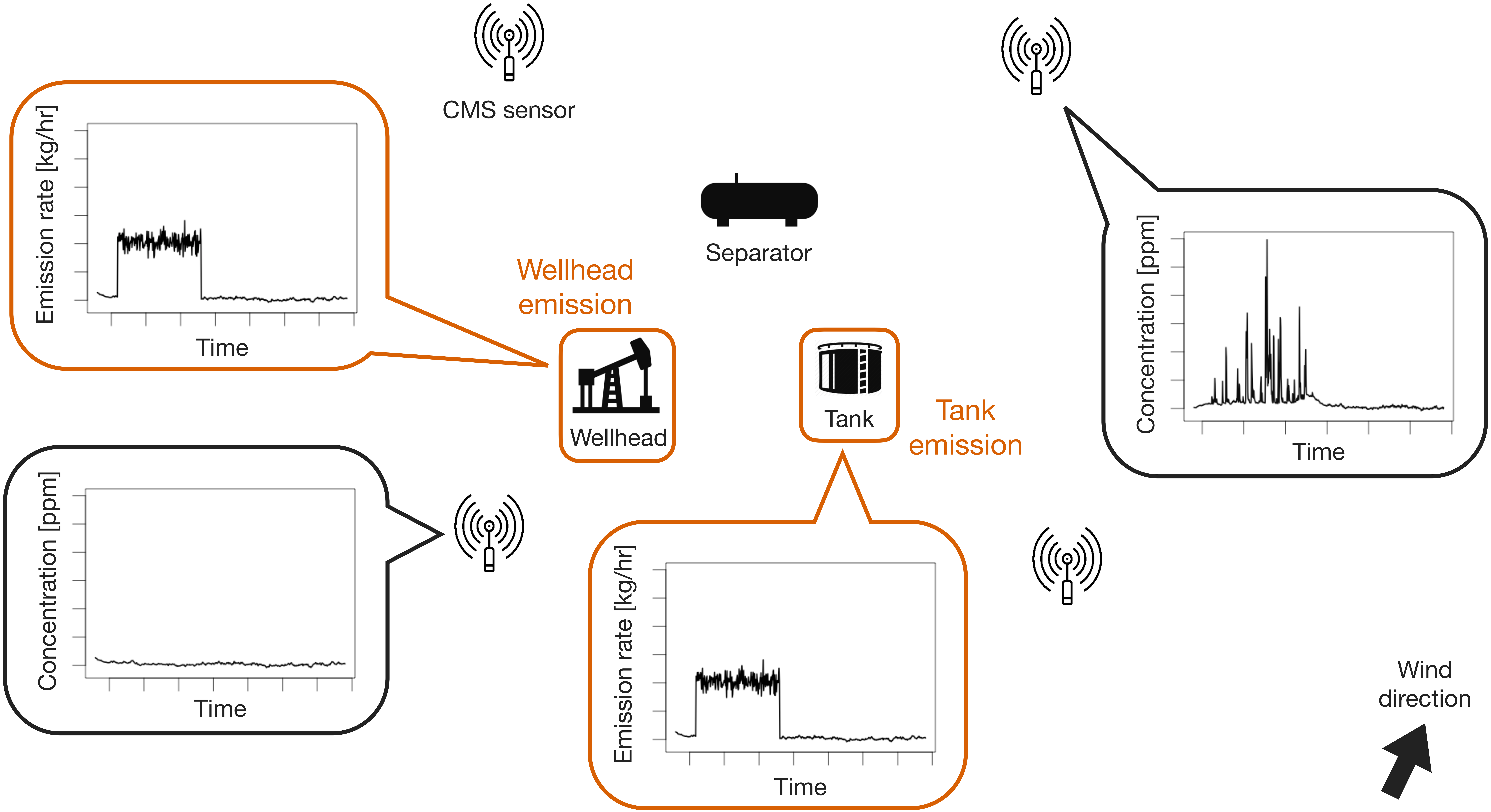


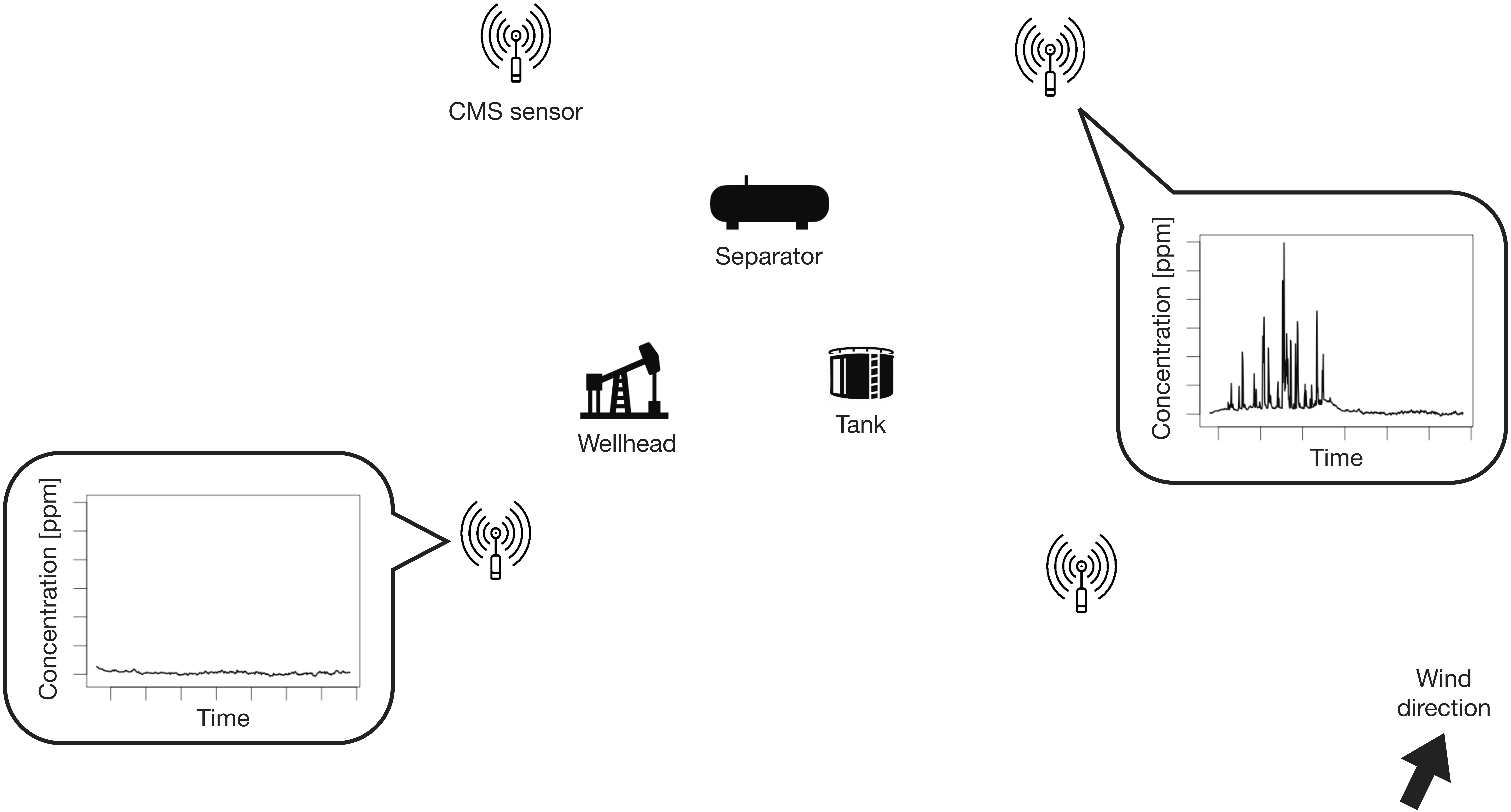
How do you get the inventory? We're almost there…





Estimating multi-source emissions with continuous monitors


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

Multisource detection, localization, and quantification (MDLQ) model

Assume a multiple linear regression model at the data level

n = number of observations
p = number of potential sources

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

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

Concentration observations from CMS sensors

Emission rates for each source

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

Gaussian puff model: mathematical definition

Set up coordinate system so that source is at (0,0,H) and positive x-axis aligns with downwind vector

Total volume of methane contained in puff p

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

Predicted methane concentration at sensor location (x,y,z) and time t from puff p

Exponential decay in concentration in horizontal plane (x, y)

Exponential decay in concentration in vertical dimension (z)

Gaussian puff model: mathematical definition

Set up coordinate system so that source is at (0,0,H) and positive x-axis aligns with downwind vector

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

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

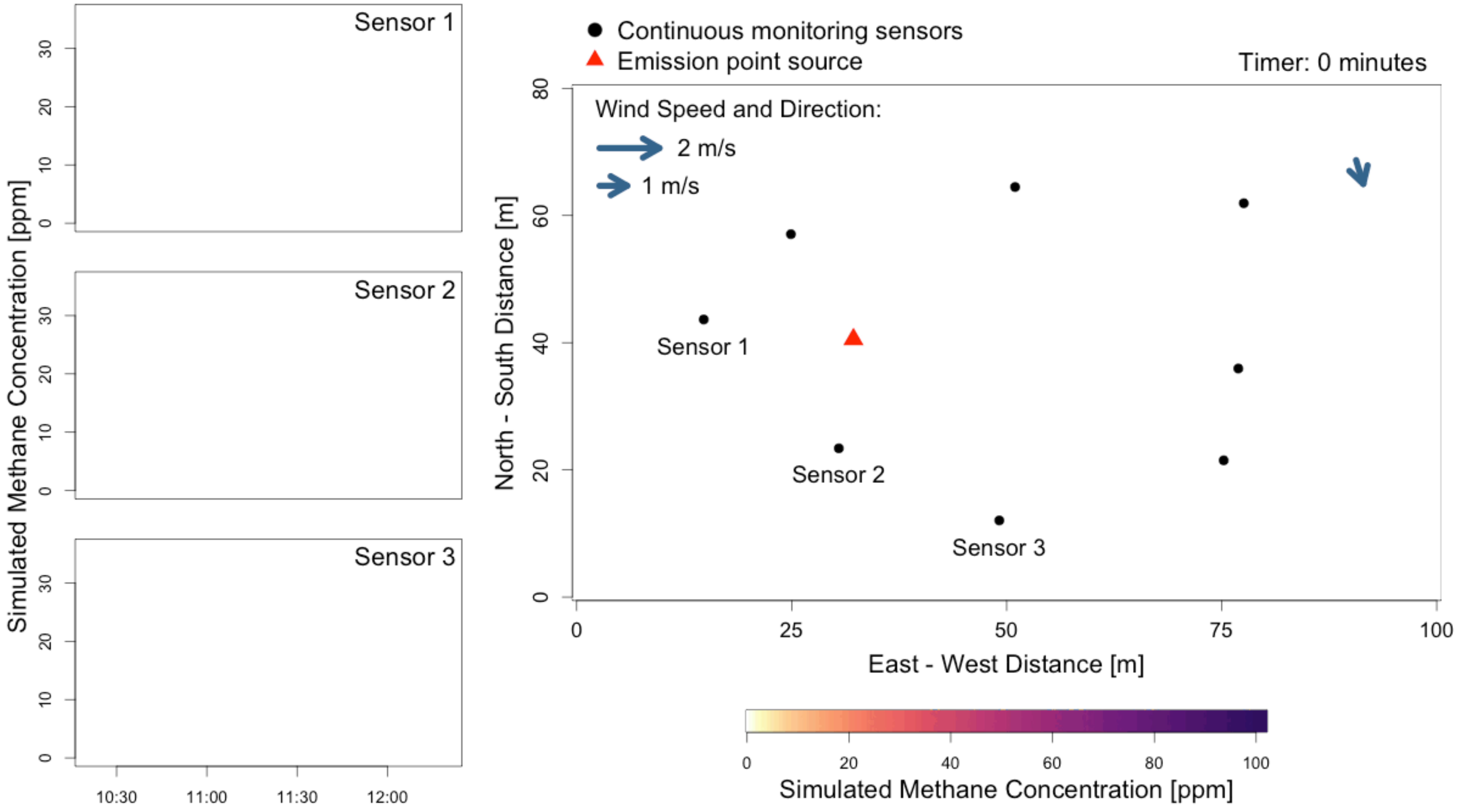
Total volume of methane contained in puff p

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

Predicted methane concentration at sensor location (x,y,z) and time t from puff p

Exponential decay in concentration in horizontal plane (x, y)

Exponential decay in concentration in vertical dimension (z)



Sensor 1
Sensor 2
Sensor 3
Simulated Methane Concentration [ppm]
0
10
20
30
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
Sensor 1
Sensor 2
Sensor 3
North - South Distance [m]
0
20
40
60
80
East - West Distance [m]
0
25
50
75
100
0
20
40
60
80
100
Simulated Methane Concentration [ppm]

Repeat this for all other potential sources!

Multisource detection, localization, and quantification (MDLQ) model

Assume a multiple linear regression model at the data level

n = number of observations
p = number of potential sources

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

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

Concentration observations from CMS sensors

Emission rates for each source

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

Multisource detection, localization, and quantification (MDLQ) model

Assume a multiple linear regression model at the data level

n = number of observations
p = number of potential sources

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

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

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

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

Multisource detection, localization, and quantification (MDLQ) model

Assume a multiple linear regression model at the data level

n = number of observations
p = number of potential sources

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

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

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

Multisource detection, localization, and quantification (MDLQ) model

Assume a multiple linear regression model at the data level

n = number of observations
p = number of potential sources

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

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

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

Multisource detection, localization, and quantification (MDLQ) model

Data-level: $y = X\beta + \epsilon$

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

n = number of observations
p = number of potential sources

The remainder of the hierarchy takes the following form

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

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

"Slab" component is non-negative

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

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

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

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

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

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

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

a_i, b_i, c_i, d_i can contain operator insight

Multisource detection, localization, and quantification (MDLQ) model

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

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

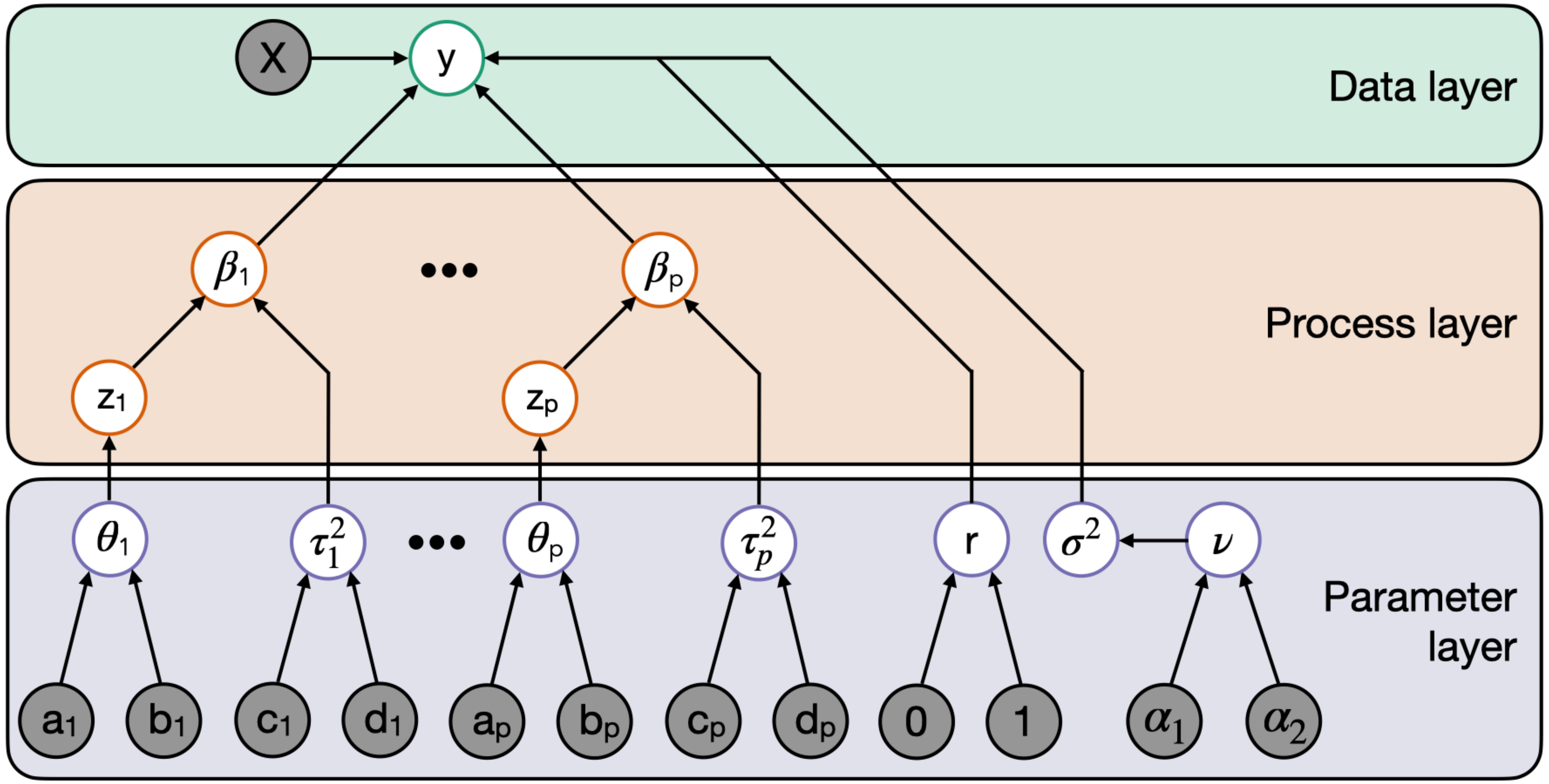
$$\theta_i \sim \mathrm{Beta}(a_i, b_i)$$

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

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

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

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



Sampling from the posterior

We can derive Gibbs updates for all parameters except ν.

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

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

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

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

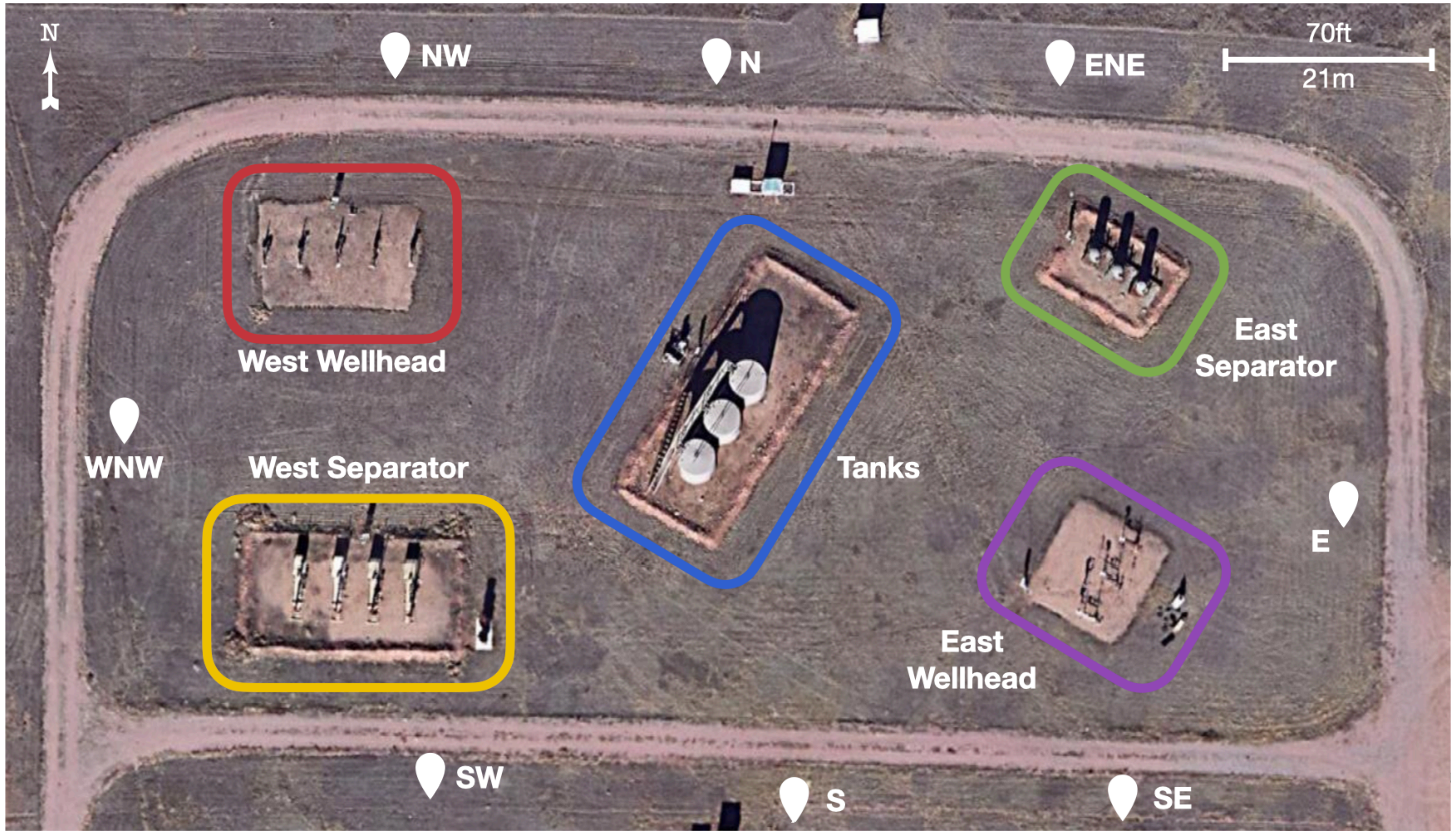
$$\beta_i|\xi \sim \begin{cases} 0 & z_i = 0 \\ \mathcal{N}\left(\left(\frac{X^T R^{-1} X}{\sigma^2}\right)^{-1}\left(\frac{X^T R^{-1} y}{\sigma^2} - \frac{e_i}{\tau_i^2\sigma^2}\right), \left(\frac{X^T R^{-1} X}{\sigma^2}\right)^{-1}\right) & z_i = 1 \end{cases}$$

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

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

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)

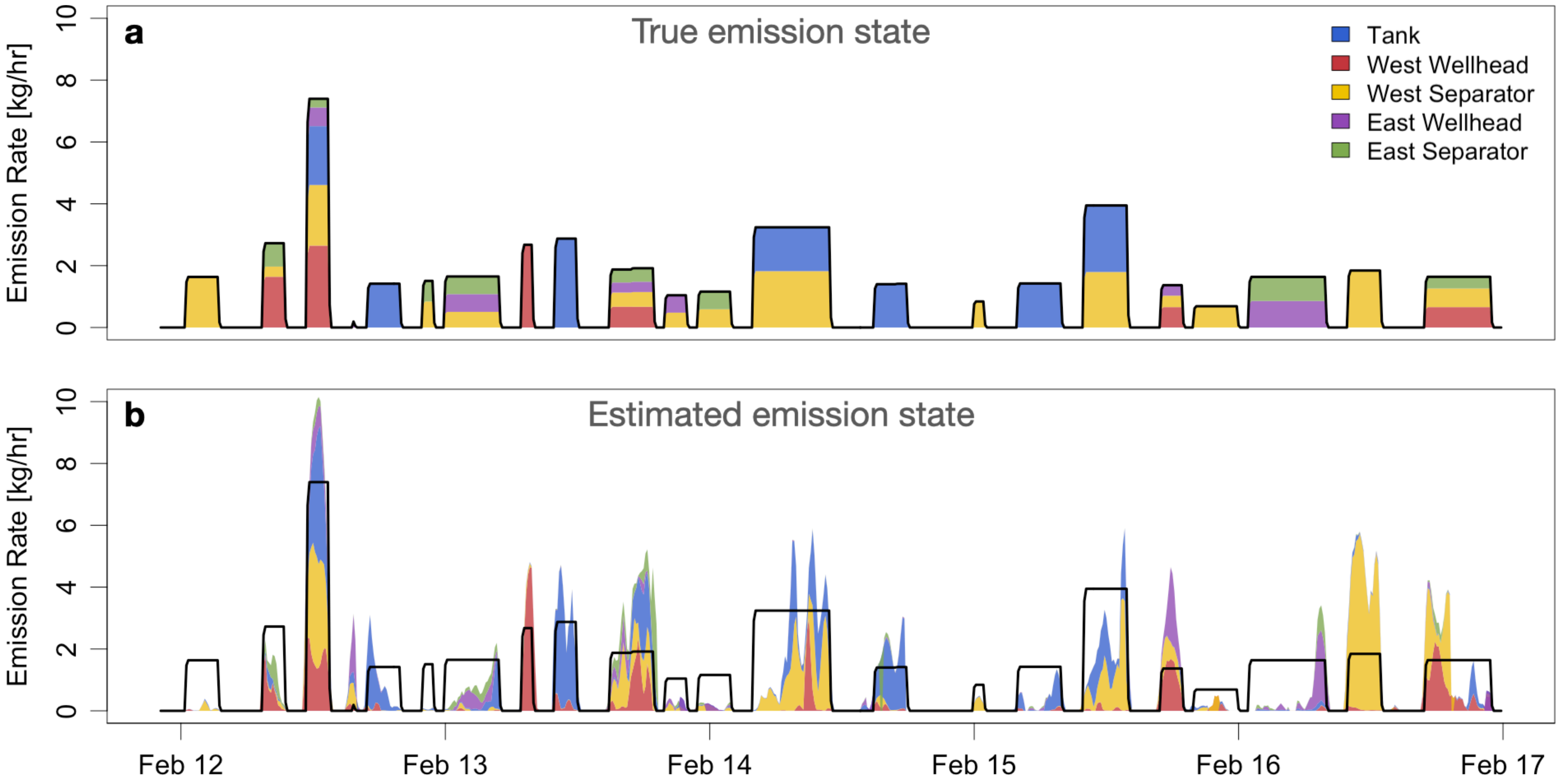
337 controlled releases:

- **99** (29%) single-source
- **238** (71%) multi-source

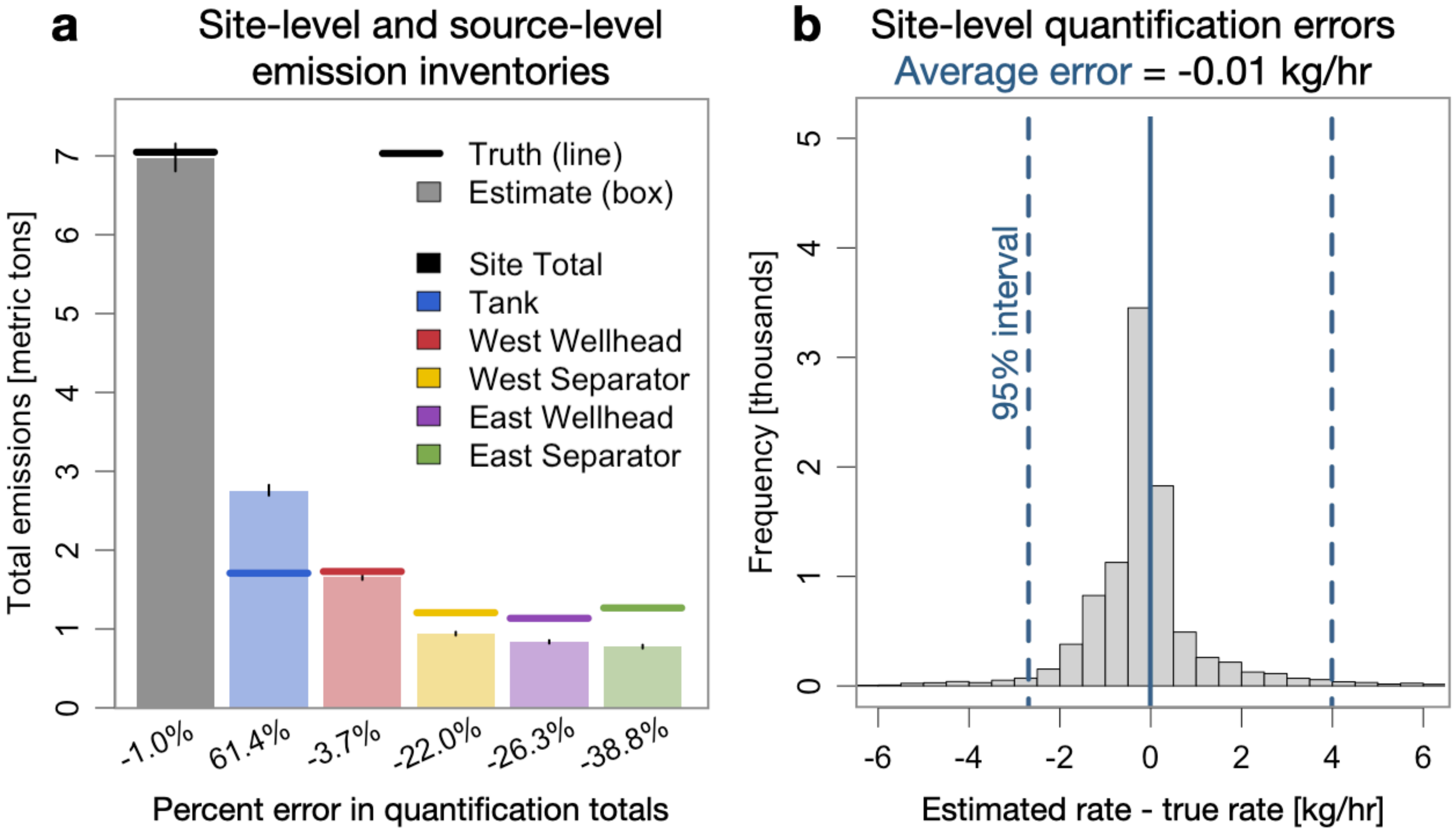
Emission rates range from **0.08** to **7.2** kg/hr

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

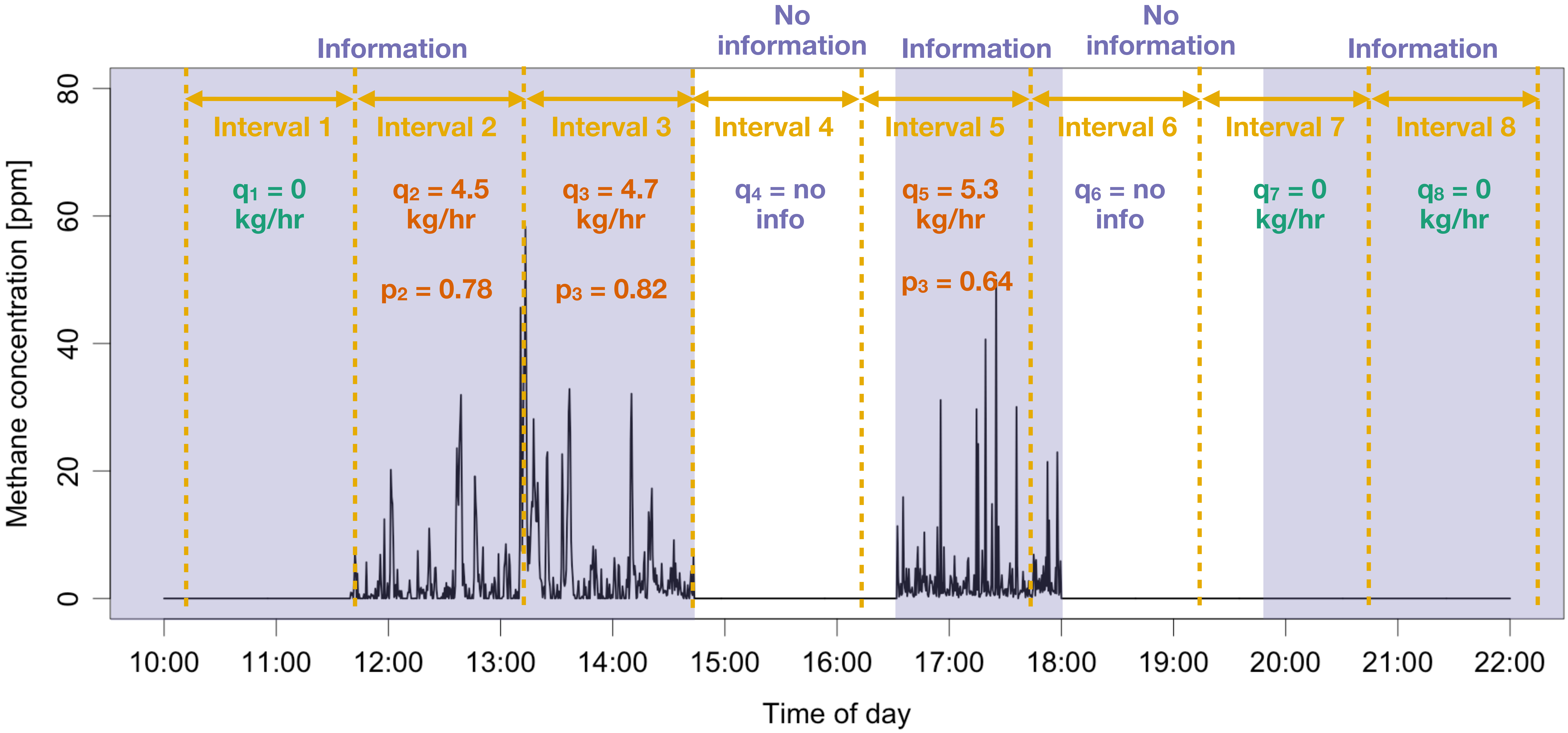
Model evaluation on multi-source controlled release data



Model evaluation on multi-source controlled release data



Need to identify when an interval is no information, no emissions, or a non-zero emission

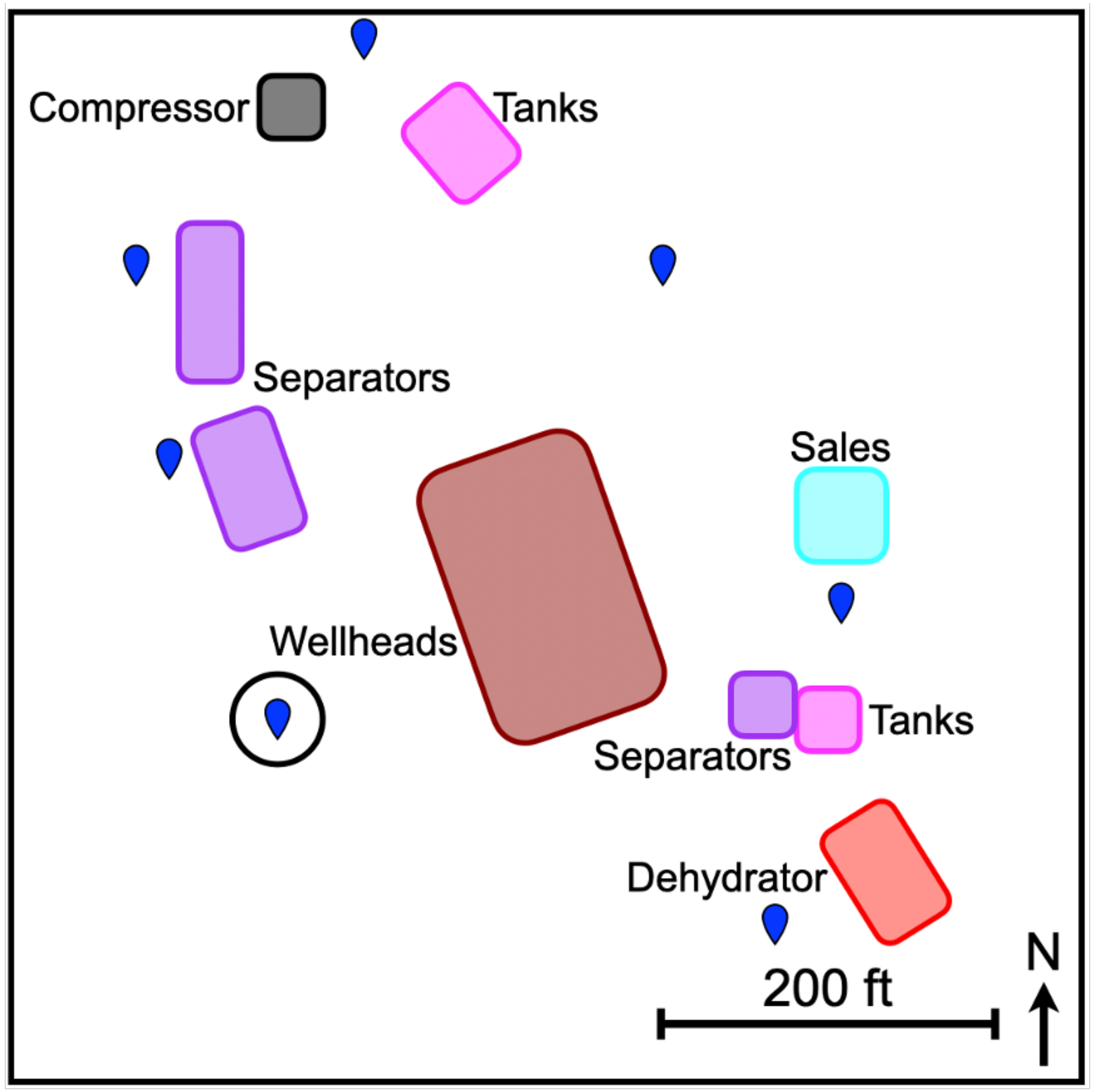


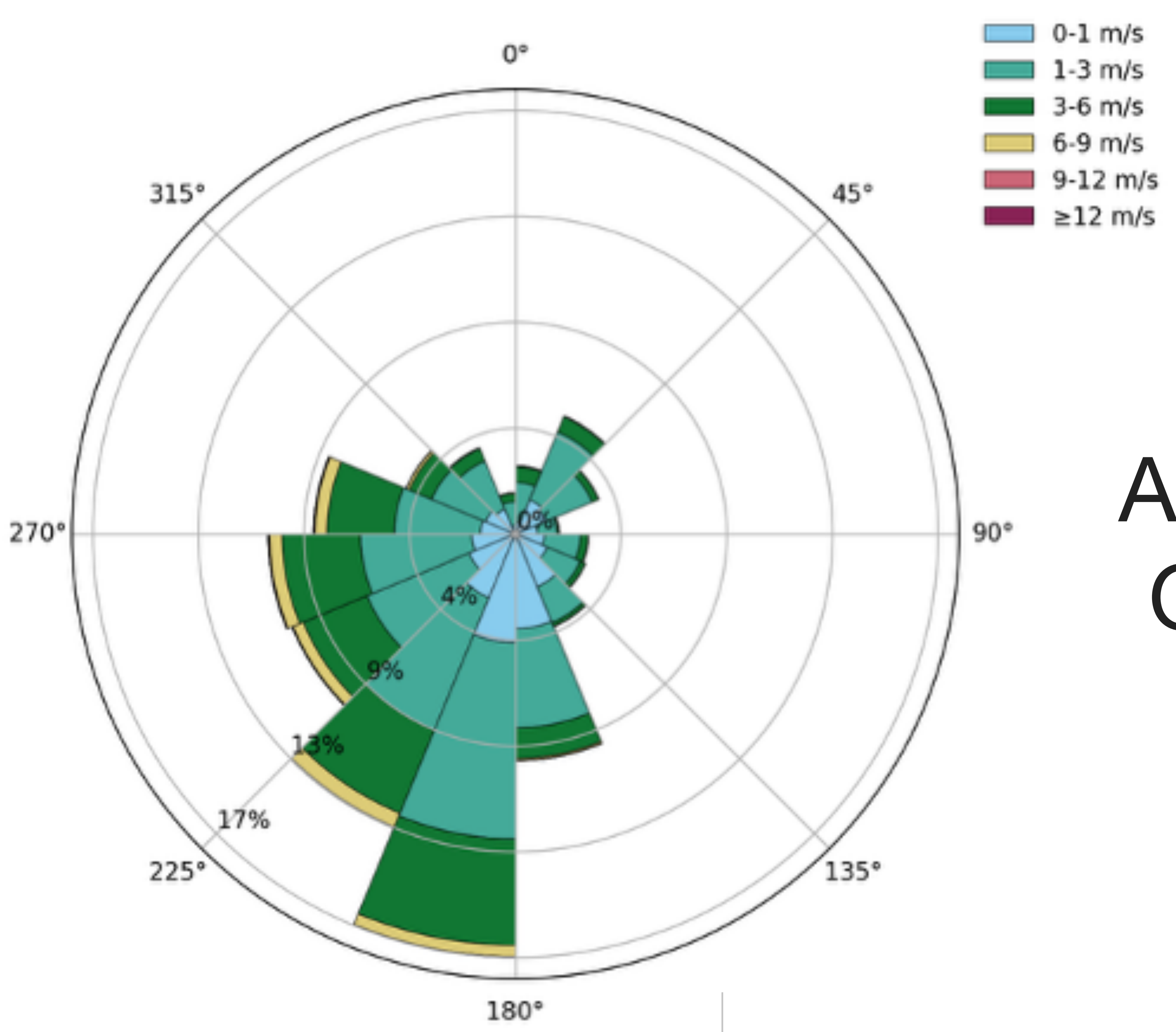
Measurement-based inventory results for the Appalachian Methane Initiative (AMI)

- **We have data from 26 production sites**
  - All are equipped with high-end continuous monitoring point sensors
  - Number of sensors per site varies from 3 to 7

- **57.82 total years of data**
  - Average of 2.22 years per site

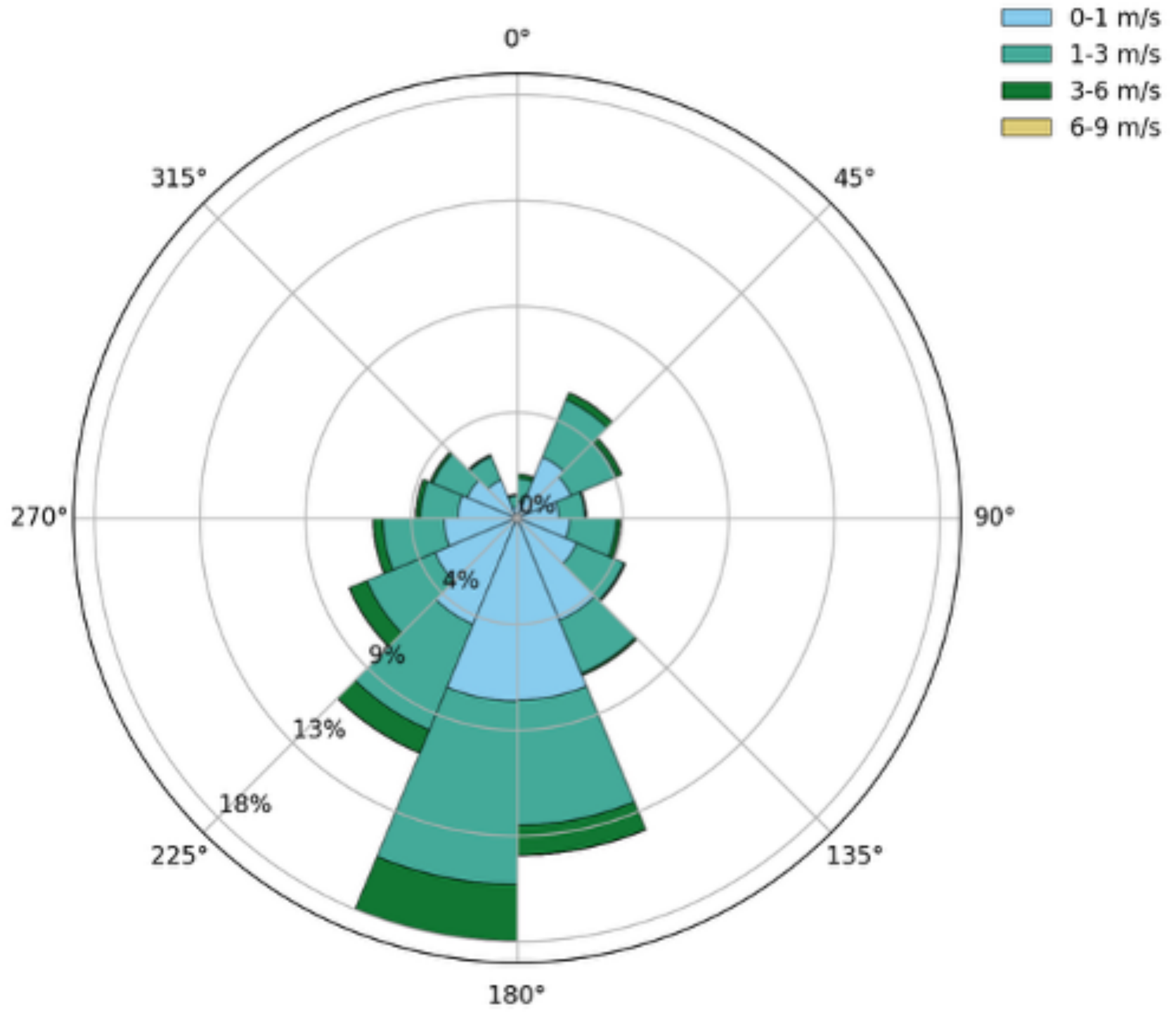
Example: Site 25





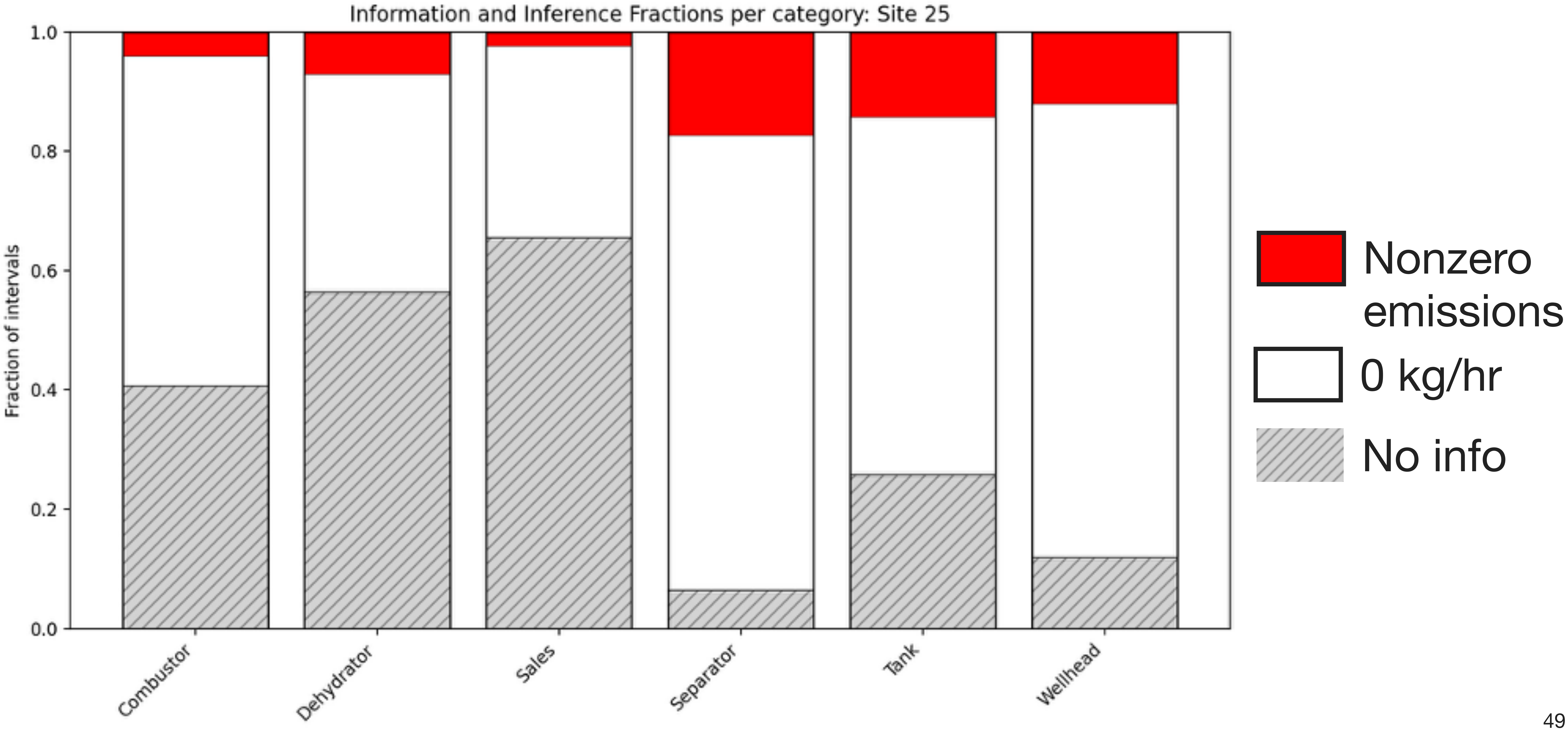
August 2023 to October 2025

September 2025

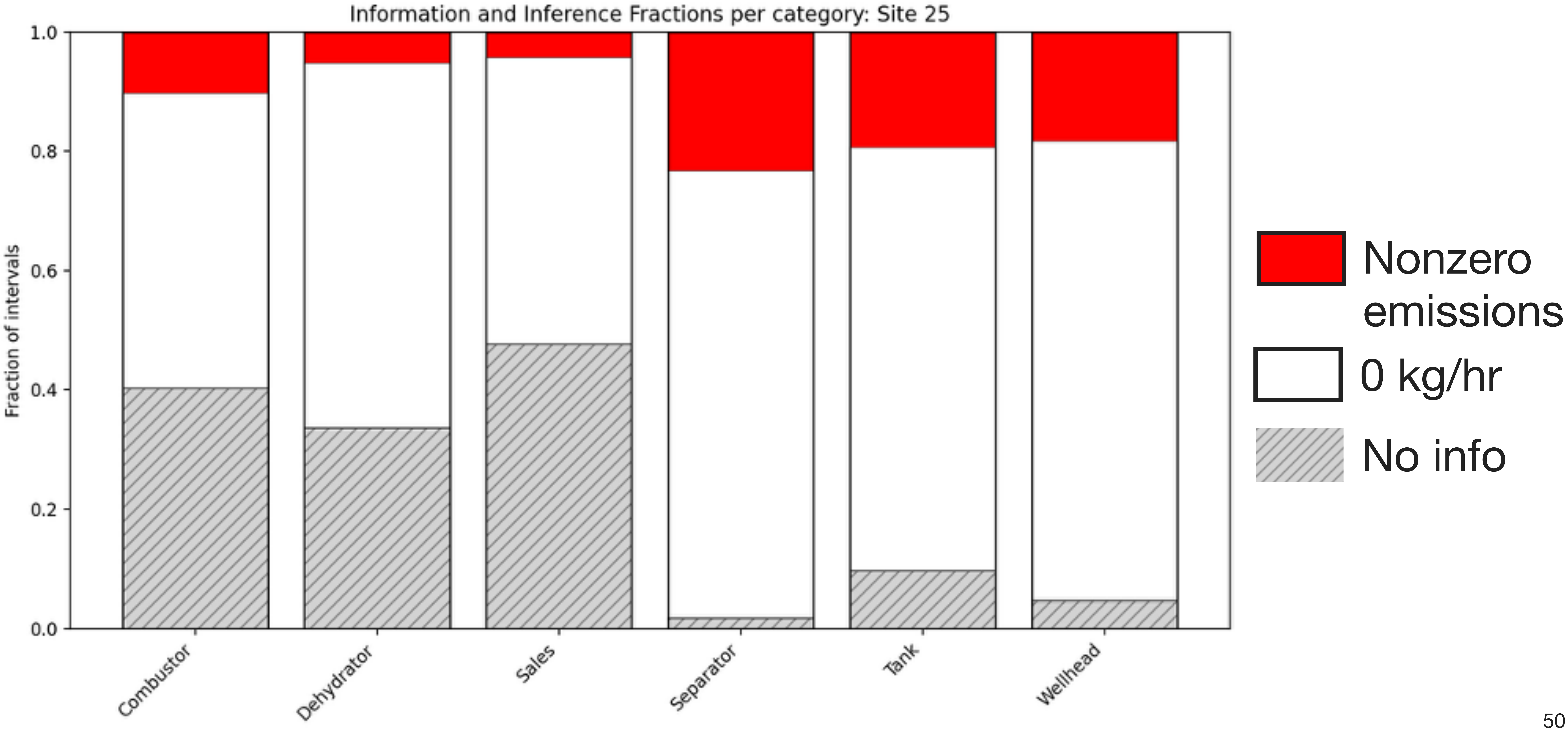


Ratio of no information to information

August 2023
to
October 2025

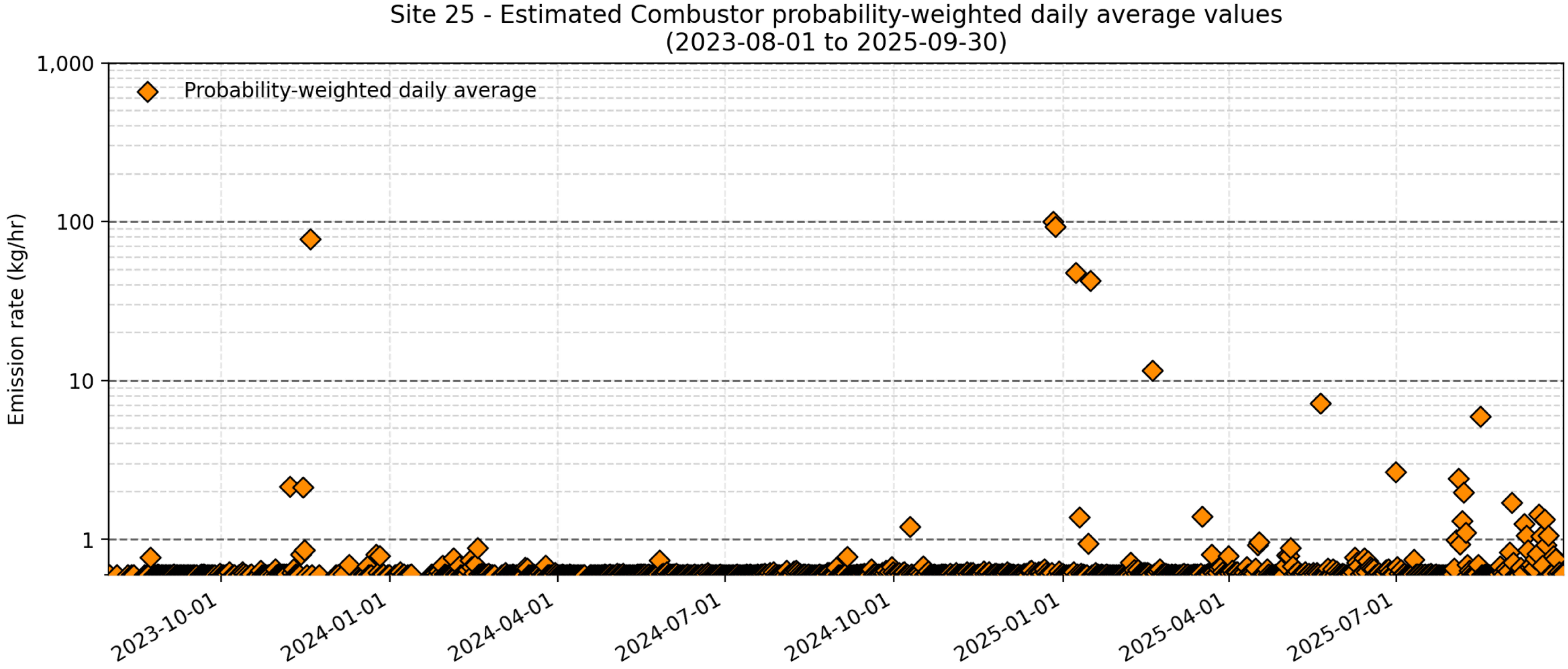


Ratio of no information to information

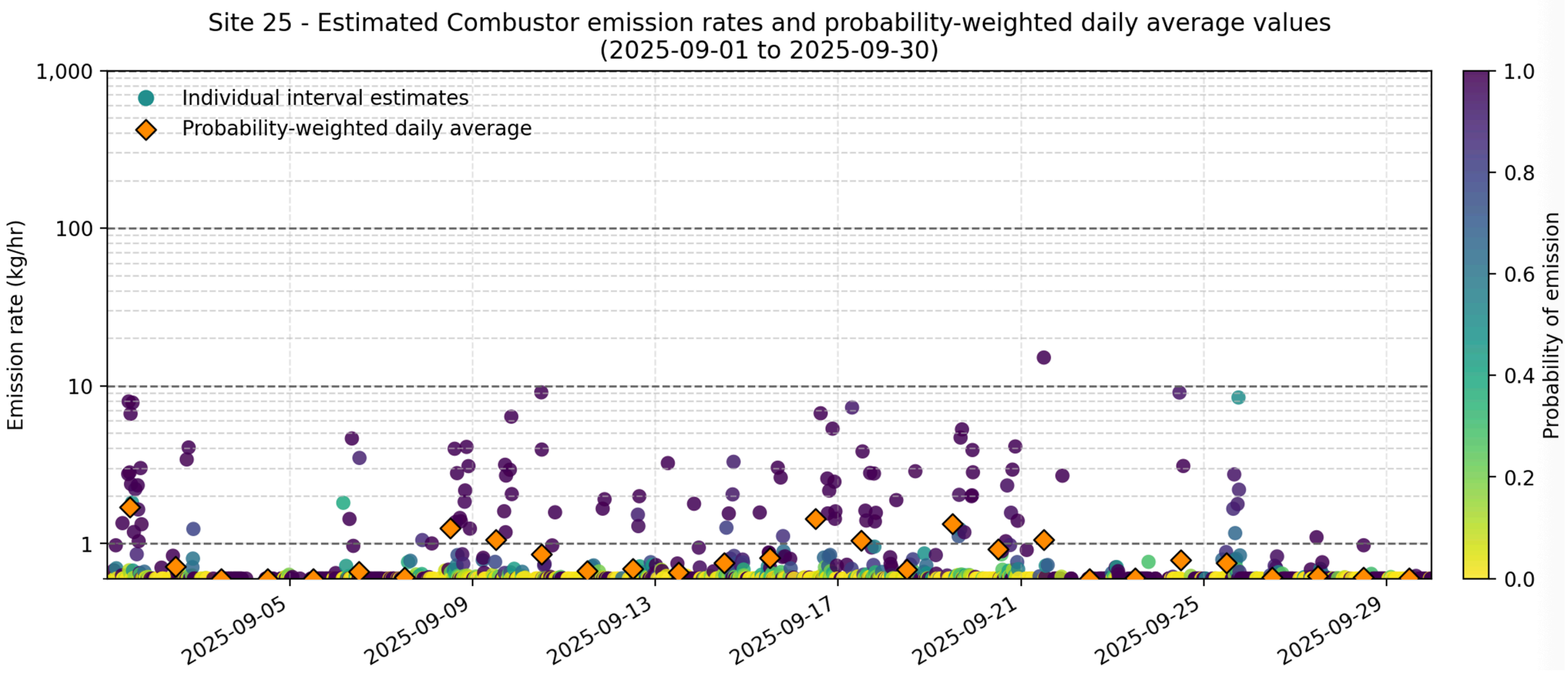


Combustor emission rate estimates over time

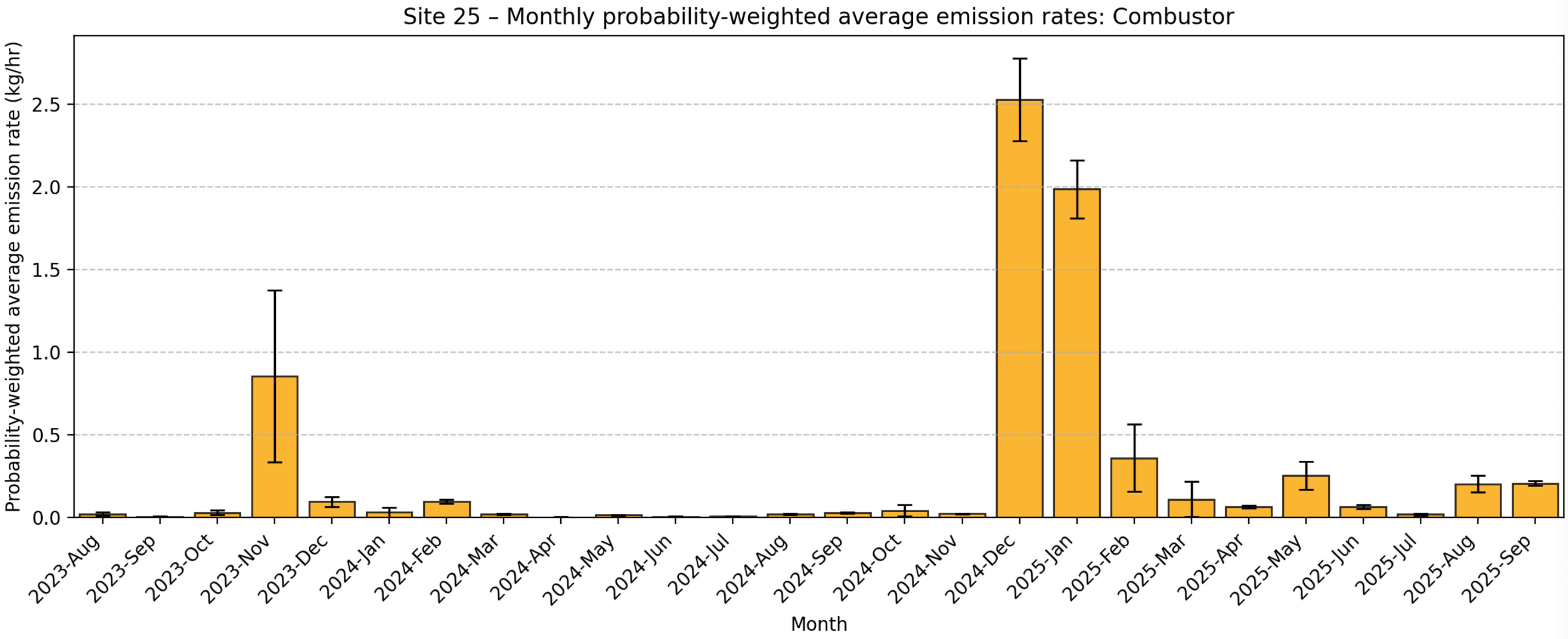
Combustor emission rate estimates over time



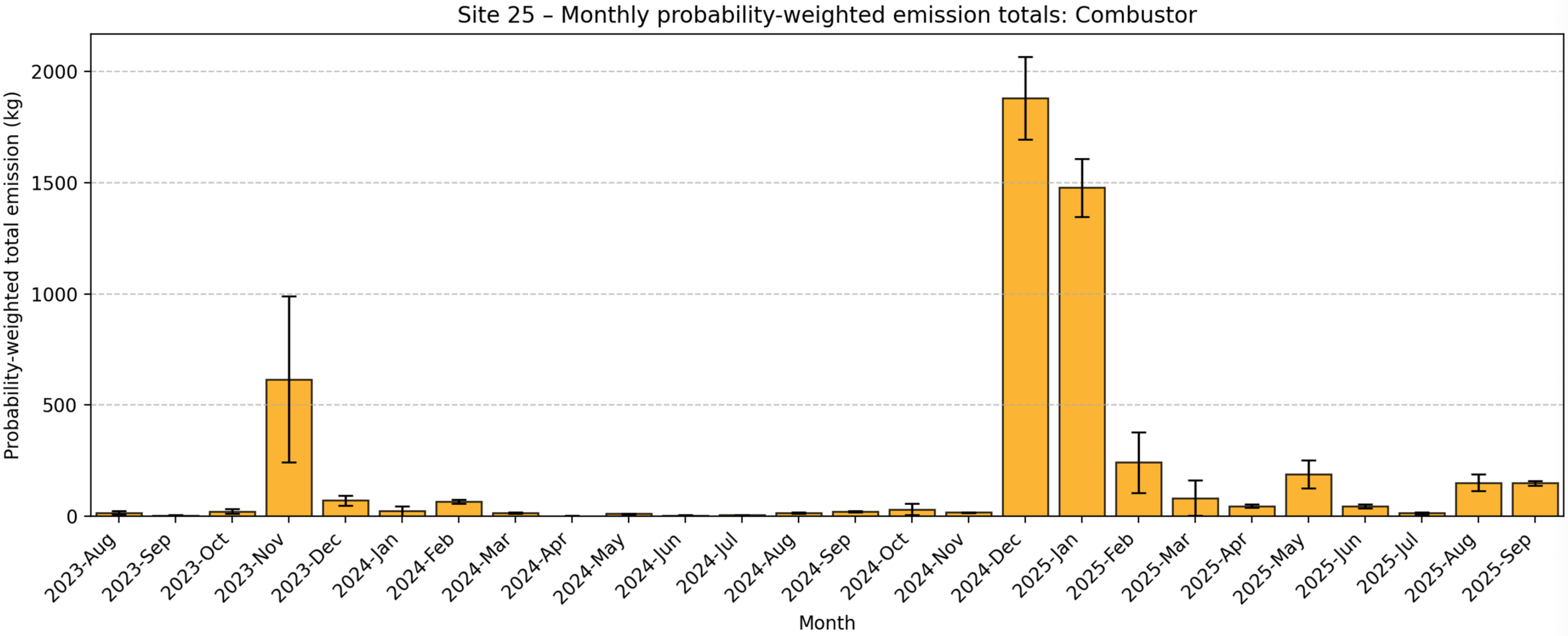
Combustor emission rate estimates over time



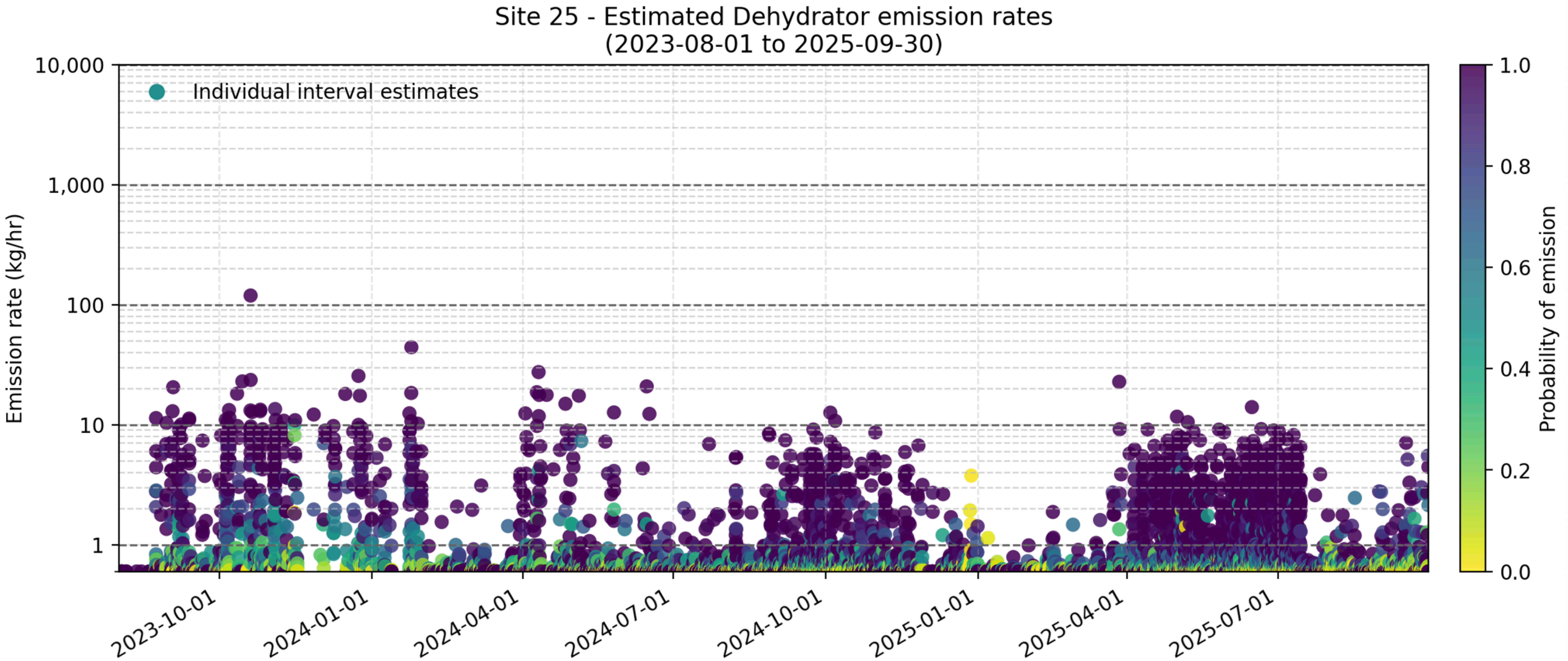
Combustor emission rate estimates over time



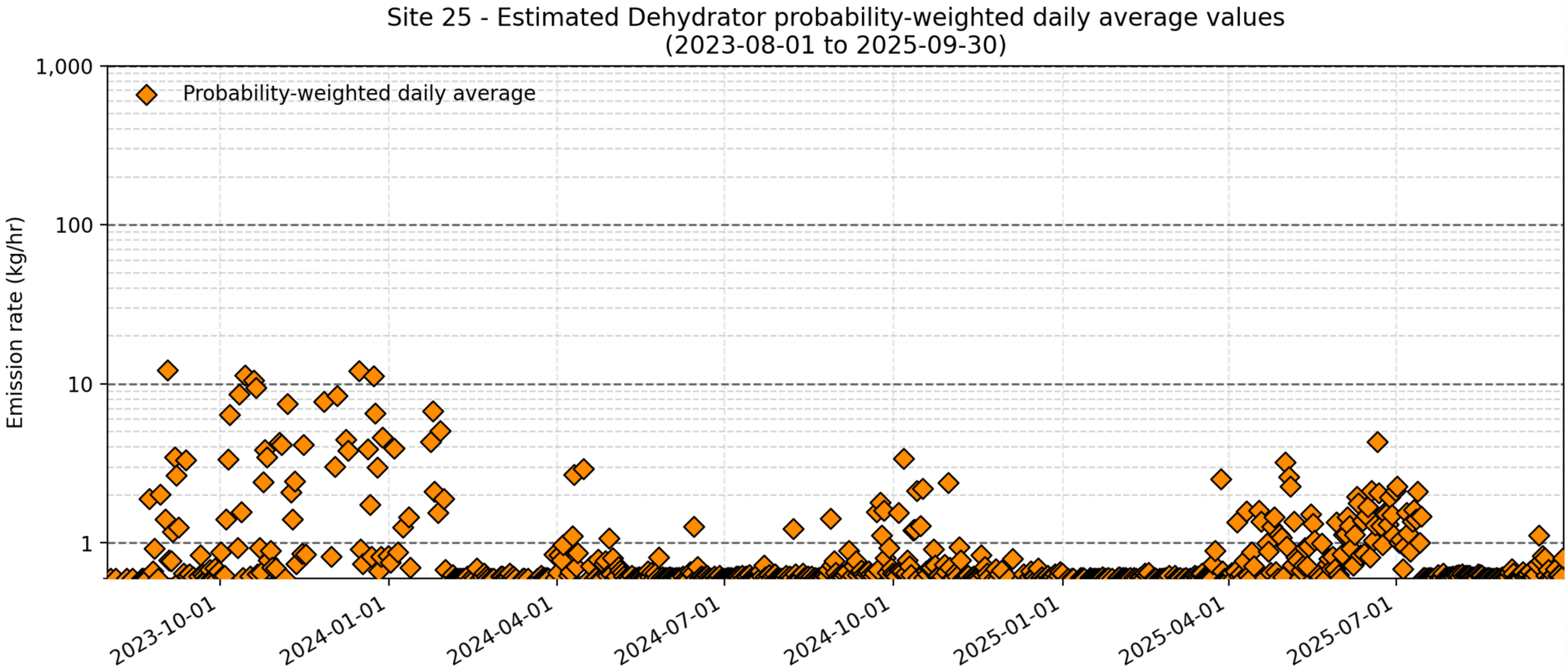
Combustor emission rate estimates over time



Dehydrator emission rate estimates over time

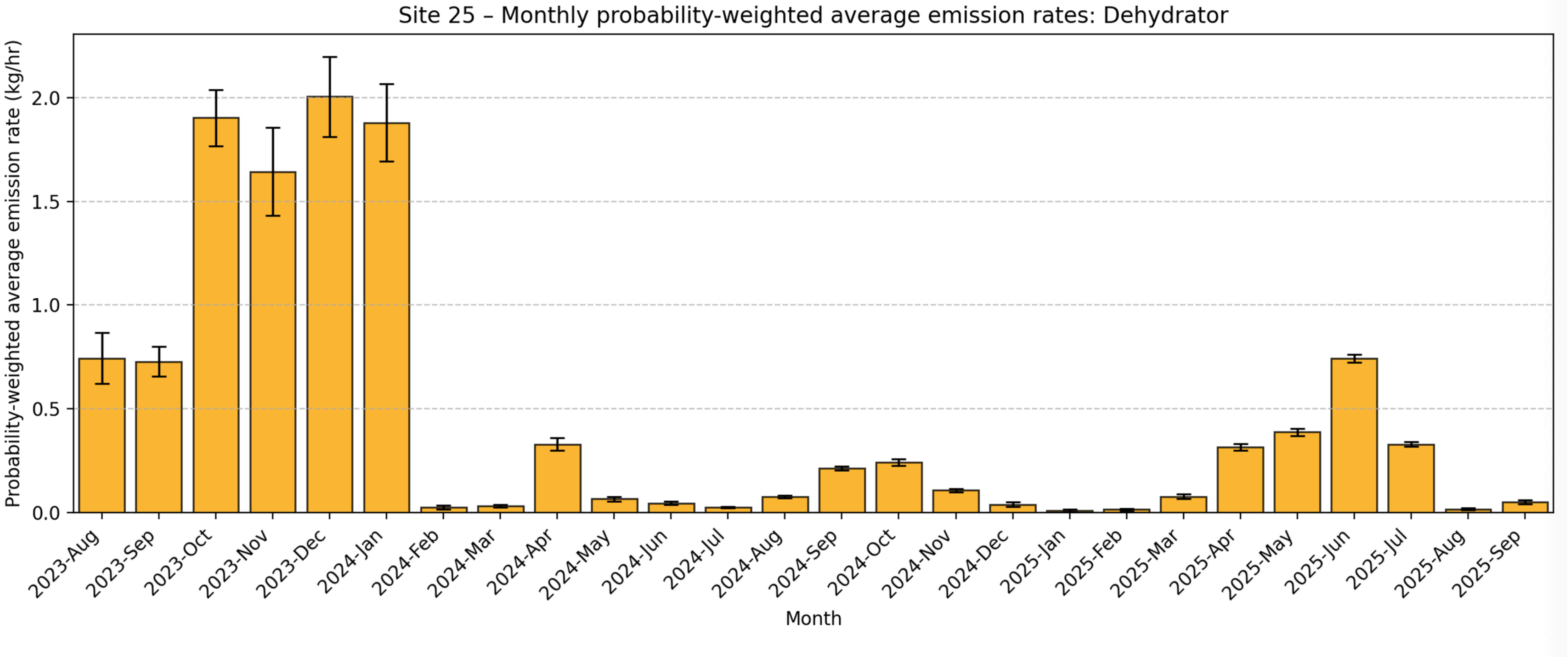


Dehydrator emission rate estimates over time

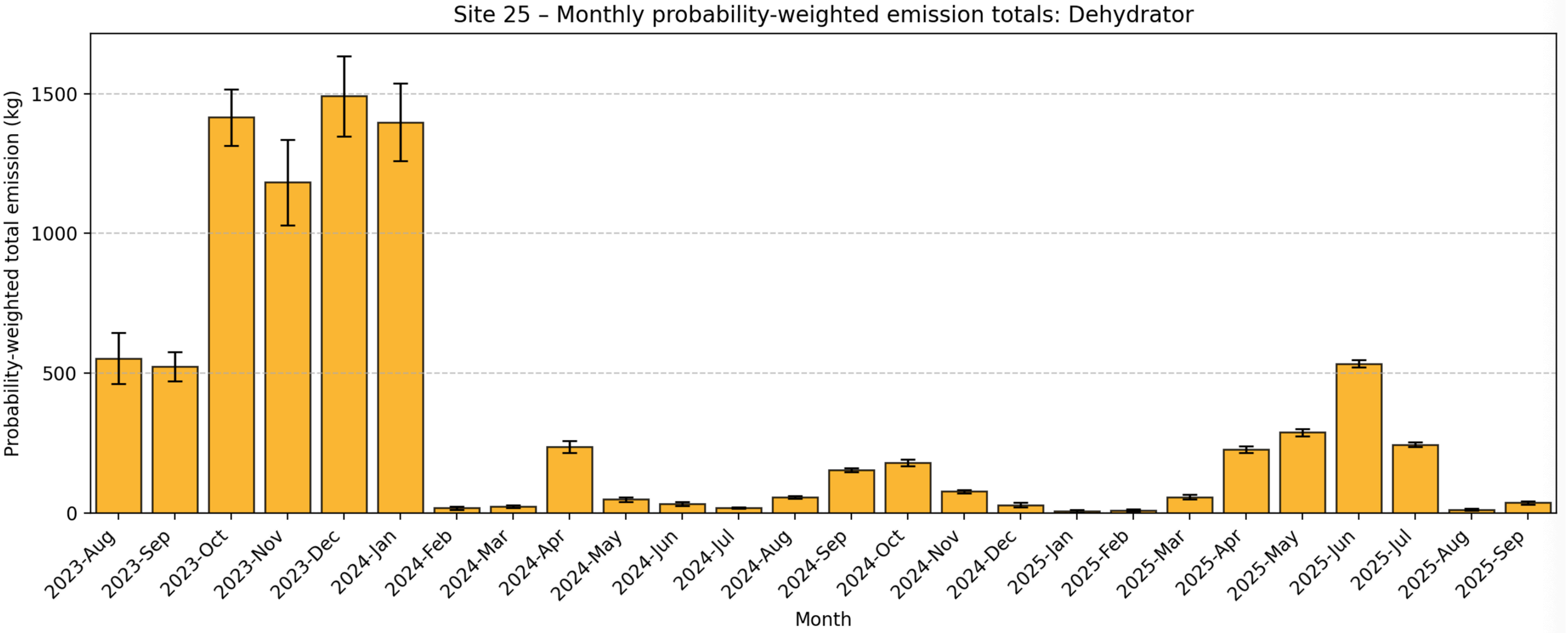


Dehydrator emission rate estimates over time

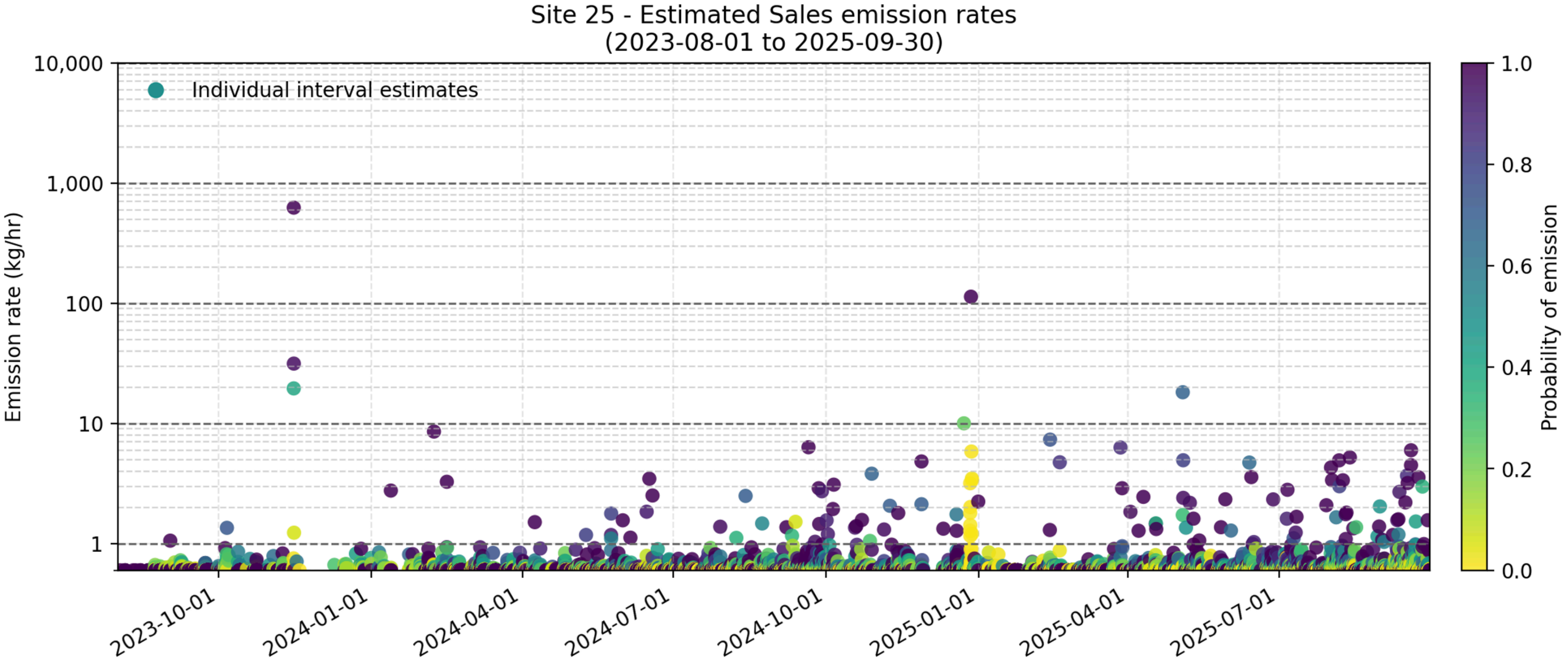
Dehydrator emission rate estimates over time



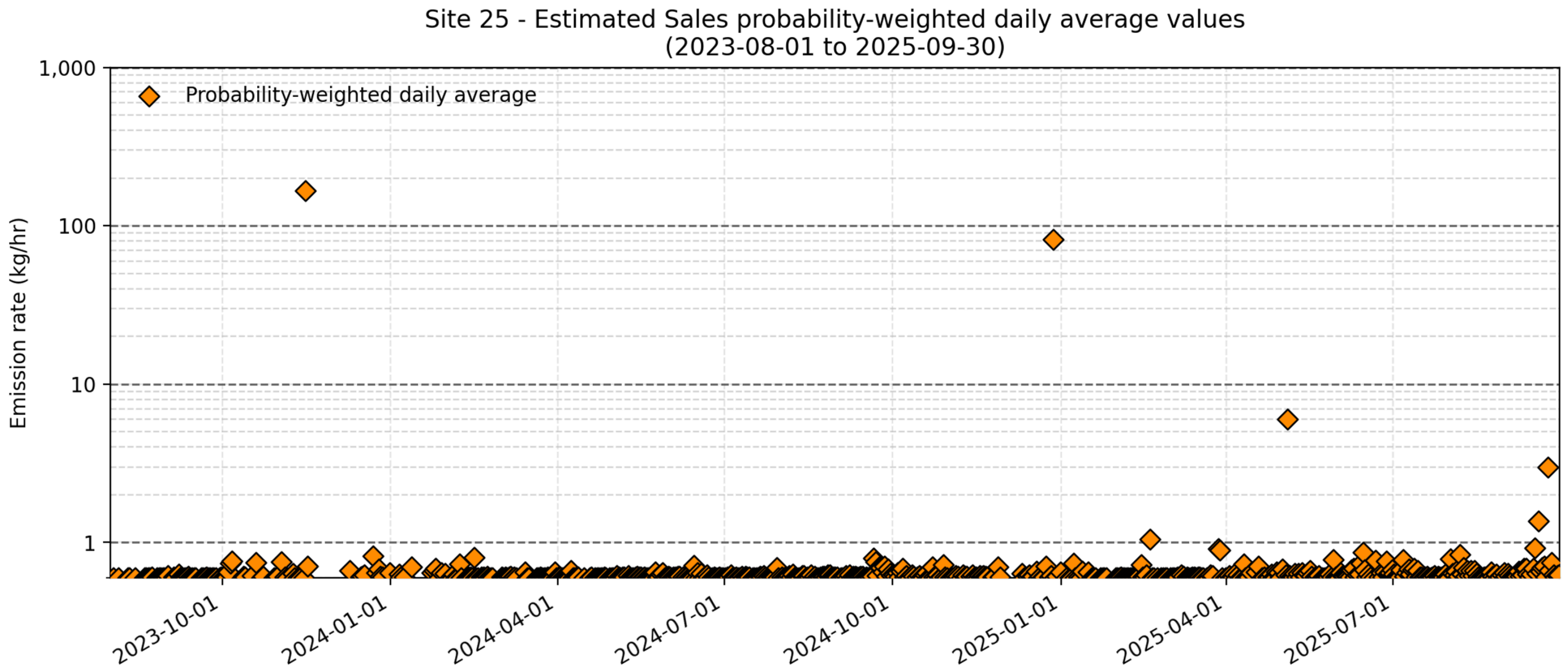
Dehydrator emission rate estimates over time



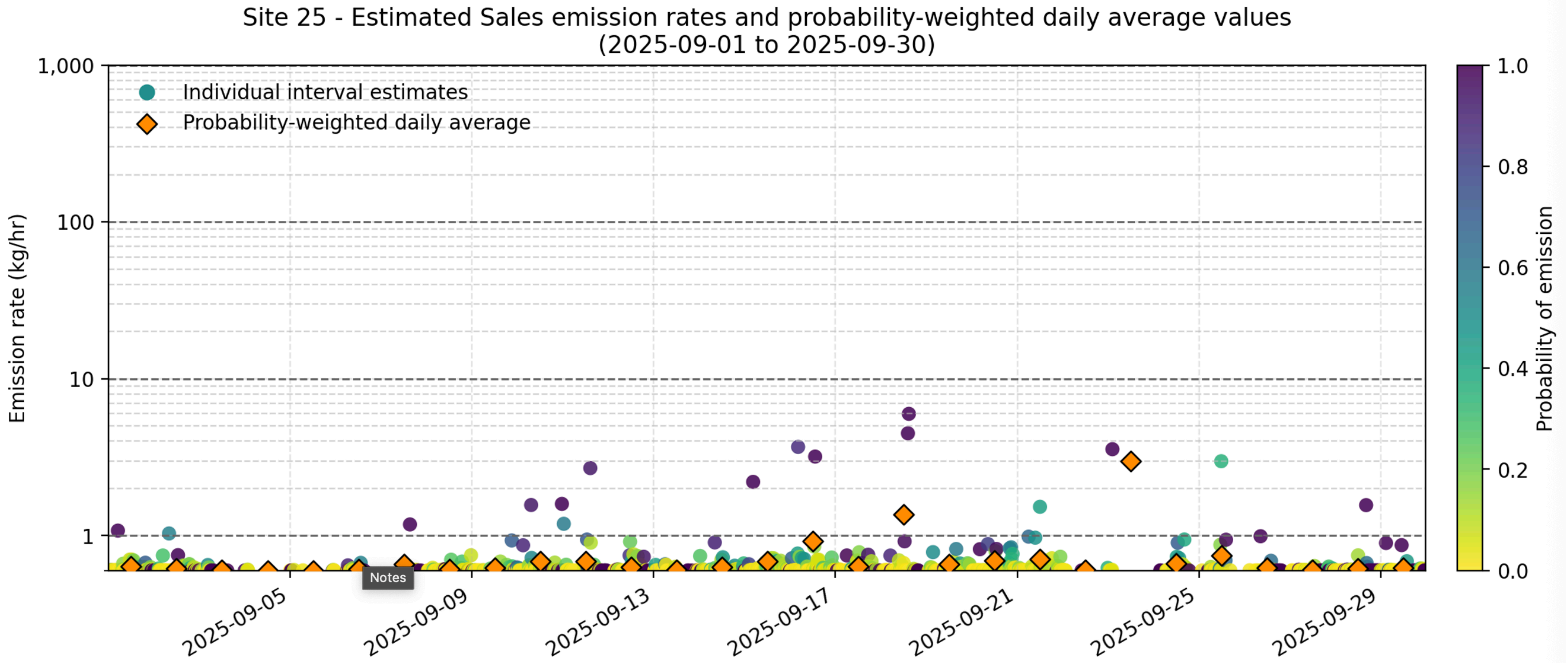
Sales line emission rate estimates over time



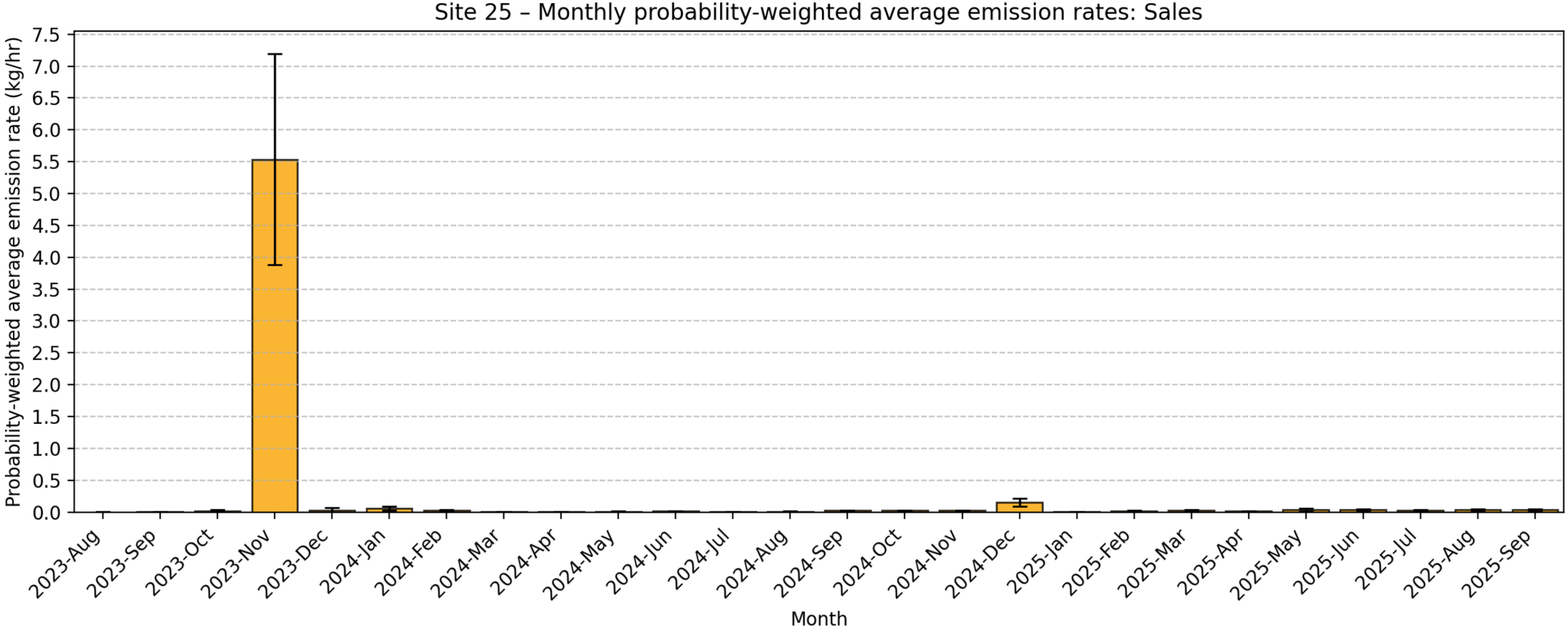
Sales line emission rate estimates over time



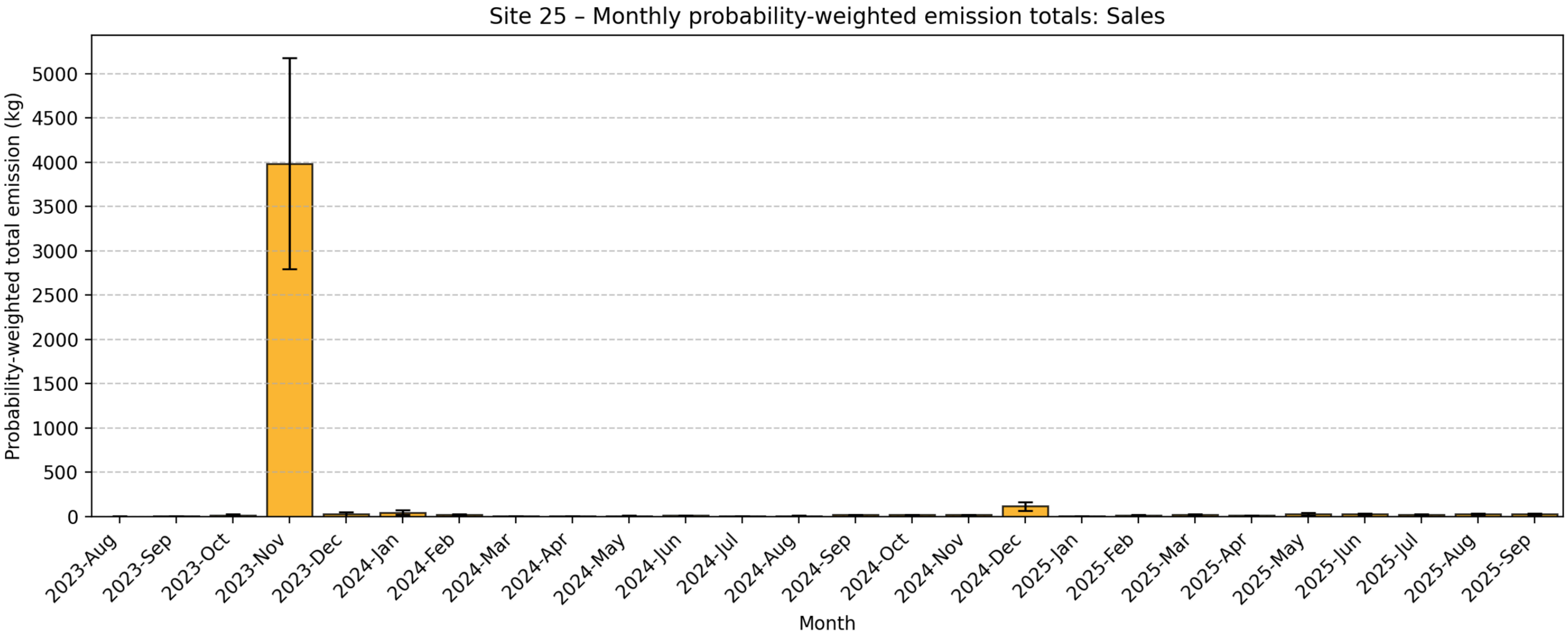
Sales line emission rate estimates over time



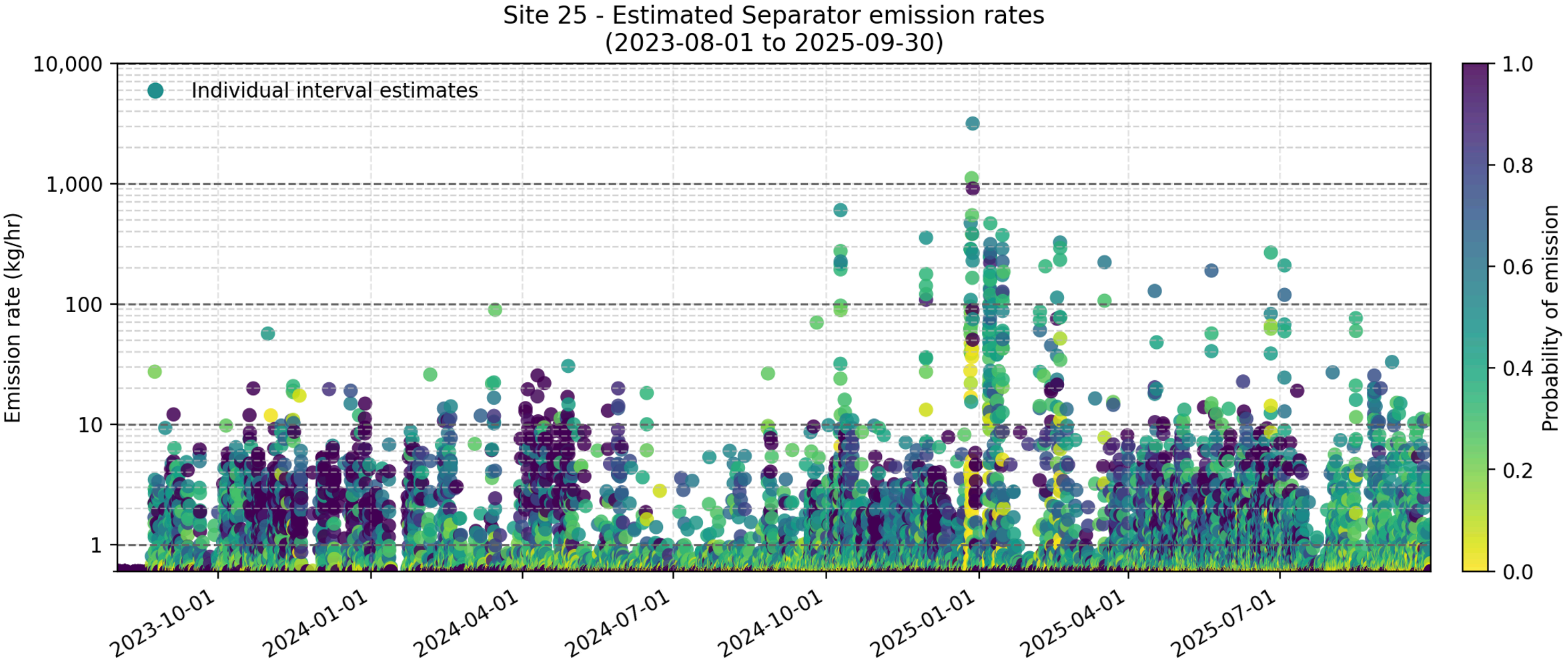
Sales line emission rate estimates over time



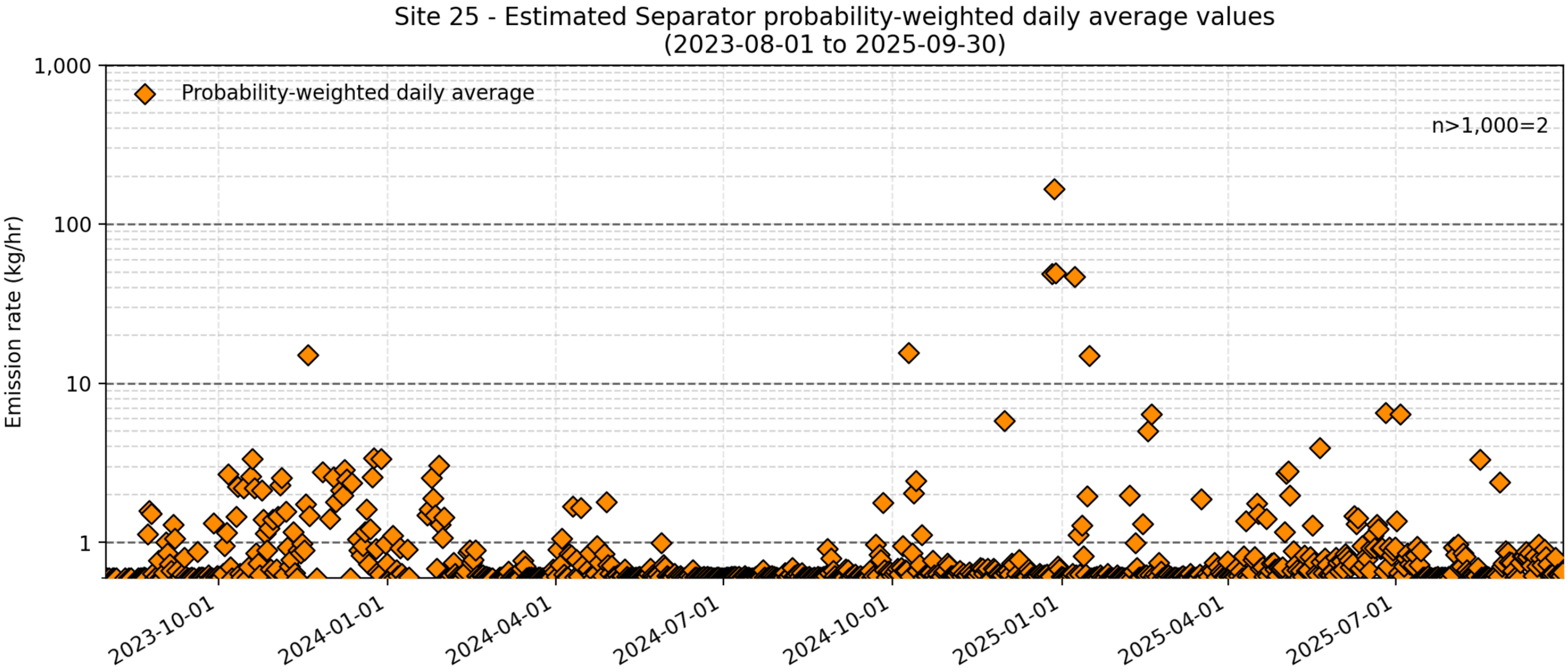
Sales line emission rate estimates over time



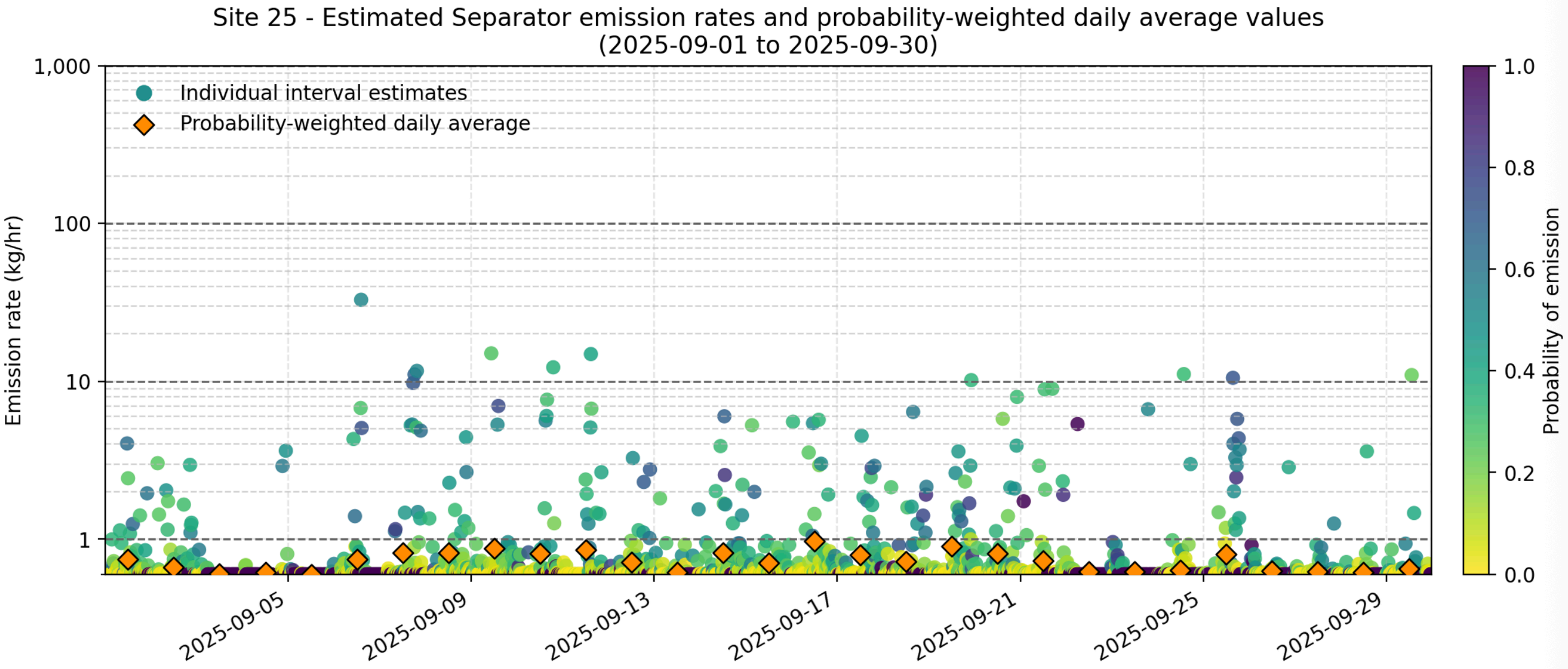
Separator emission rate estimates over time



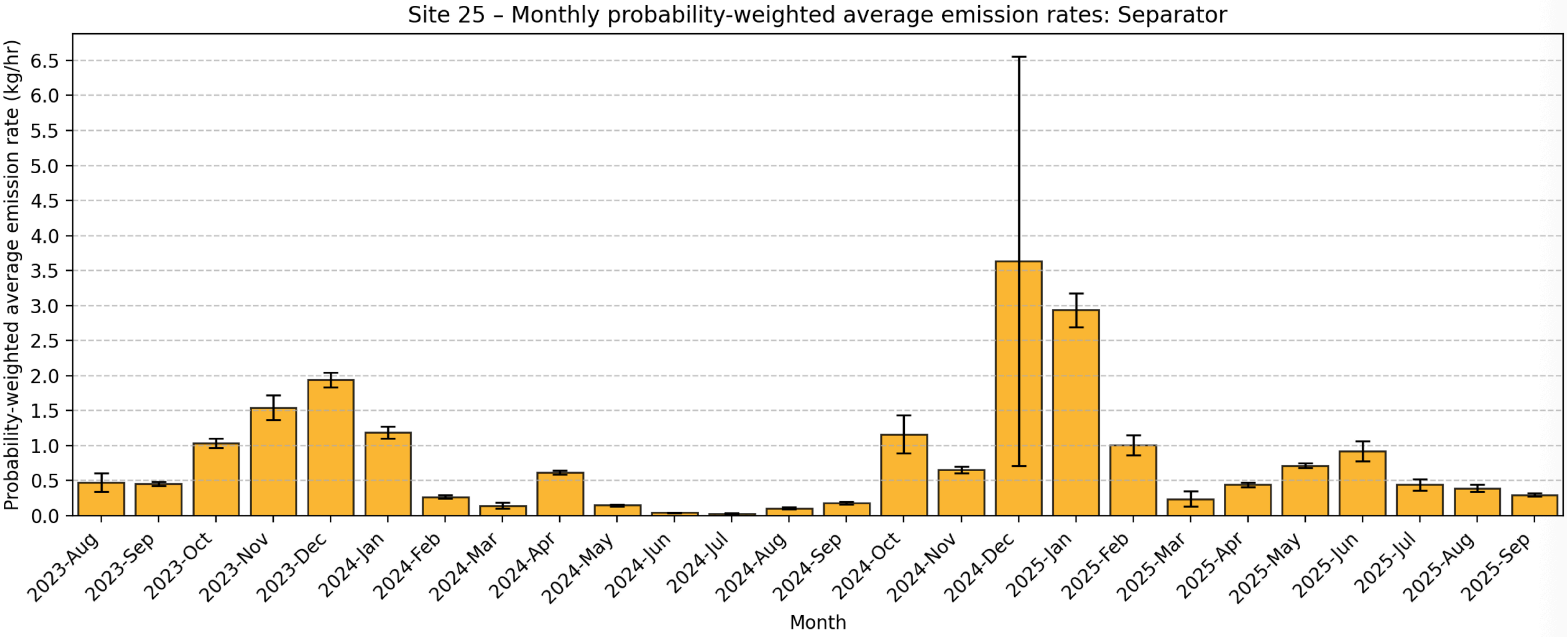
Separator emission rate estimates over time



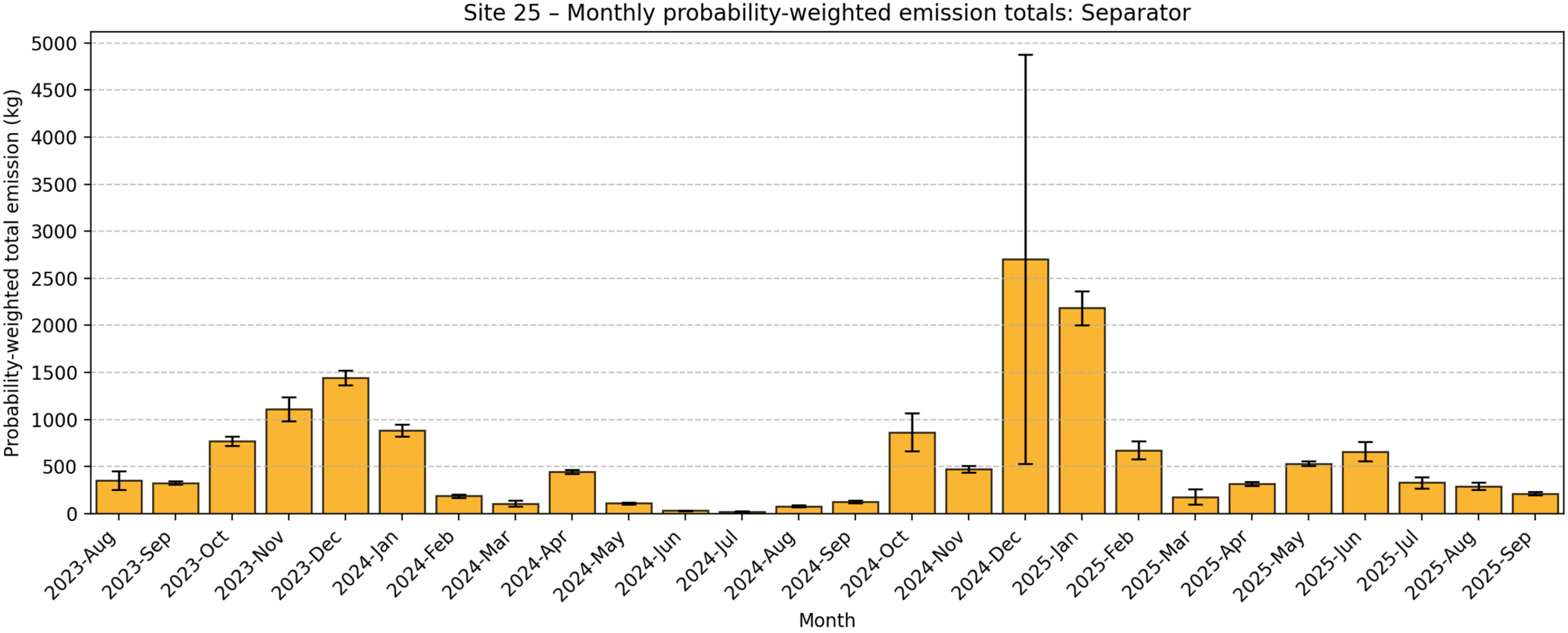
Separator emission rate estimates over time



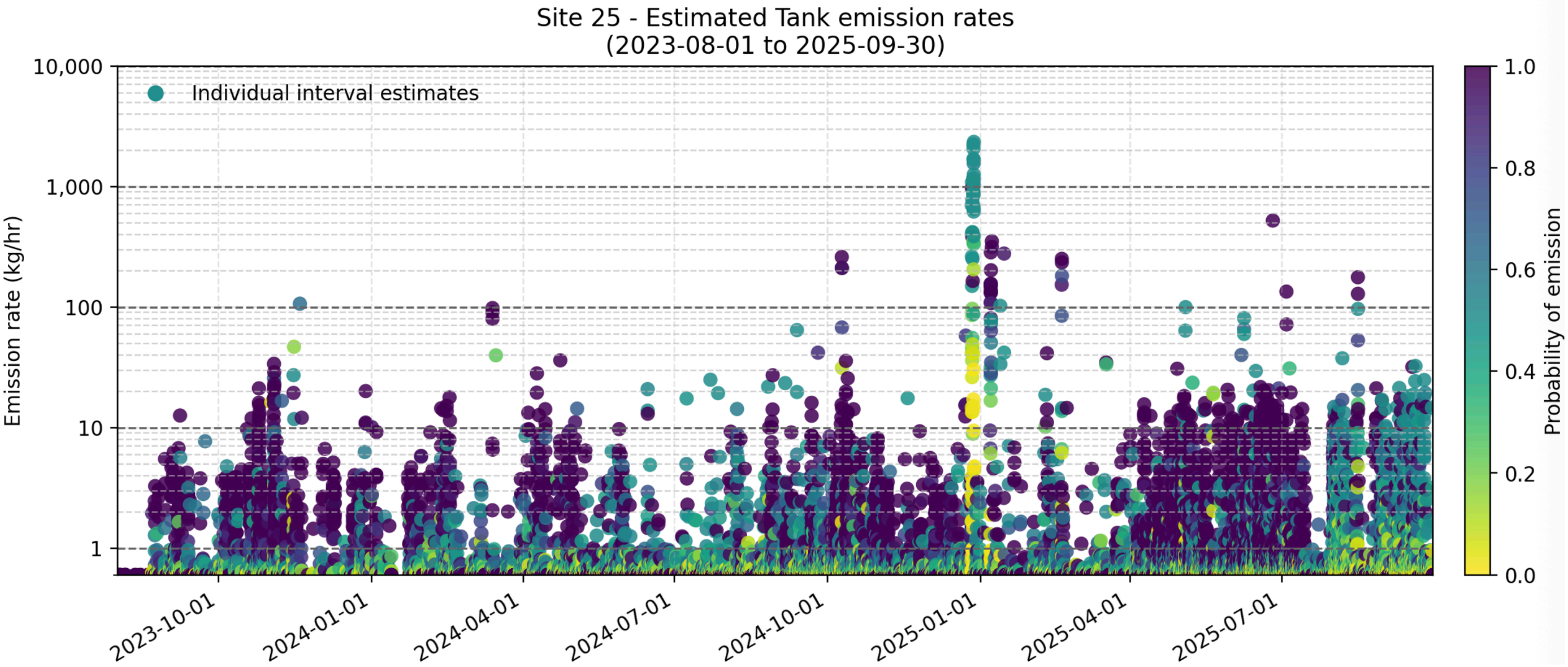
Separator emission rate estimates over time



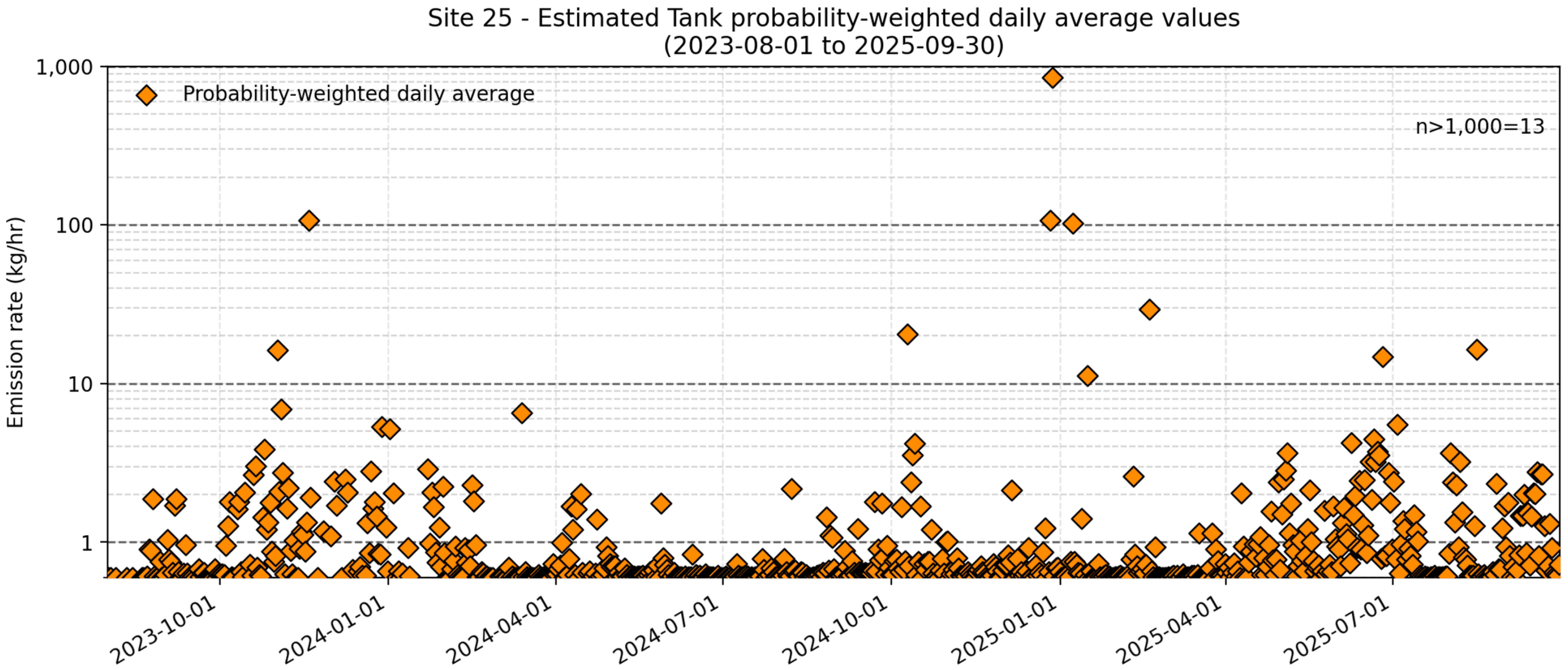
Separator emission rate estimates over time



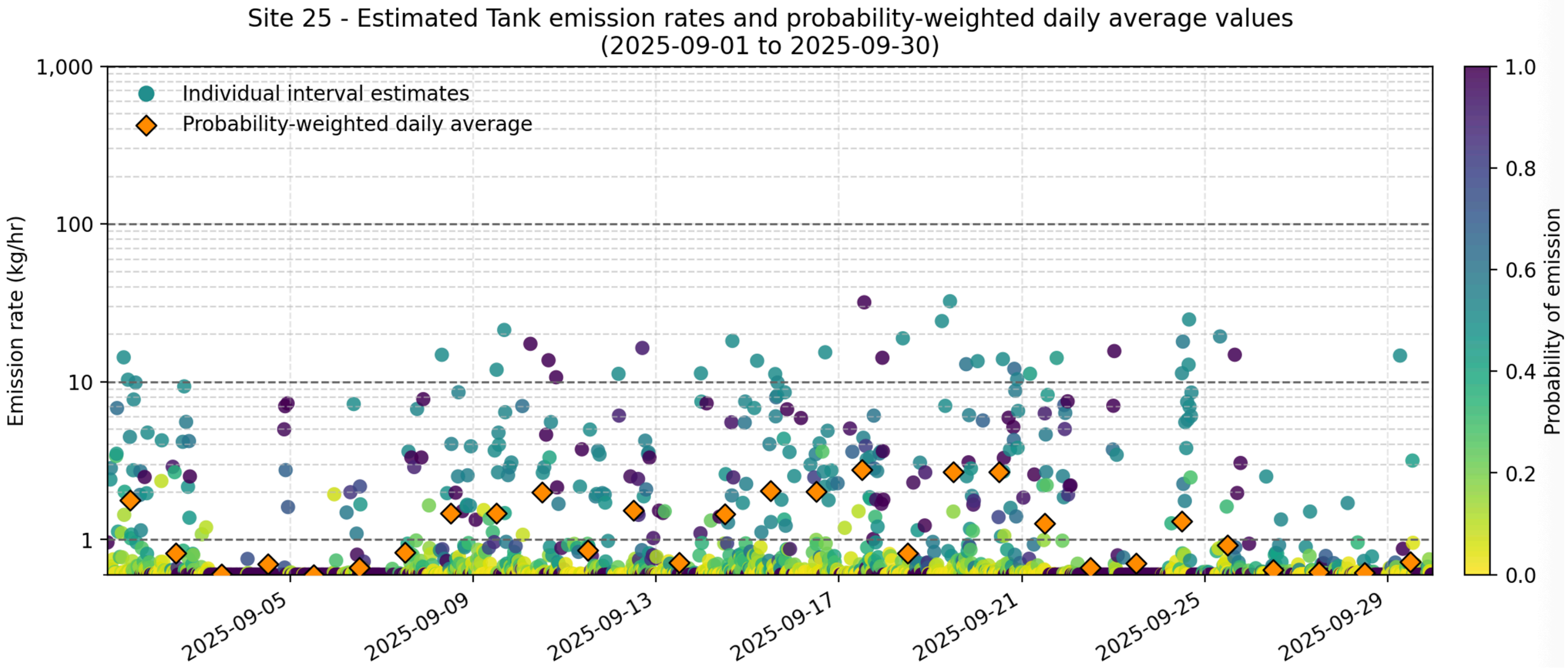
Tank emission rate estimates over time



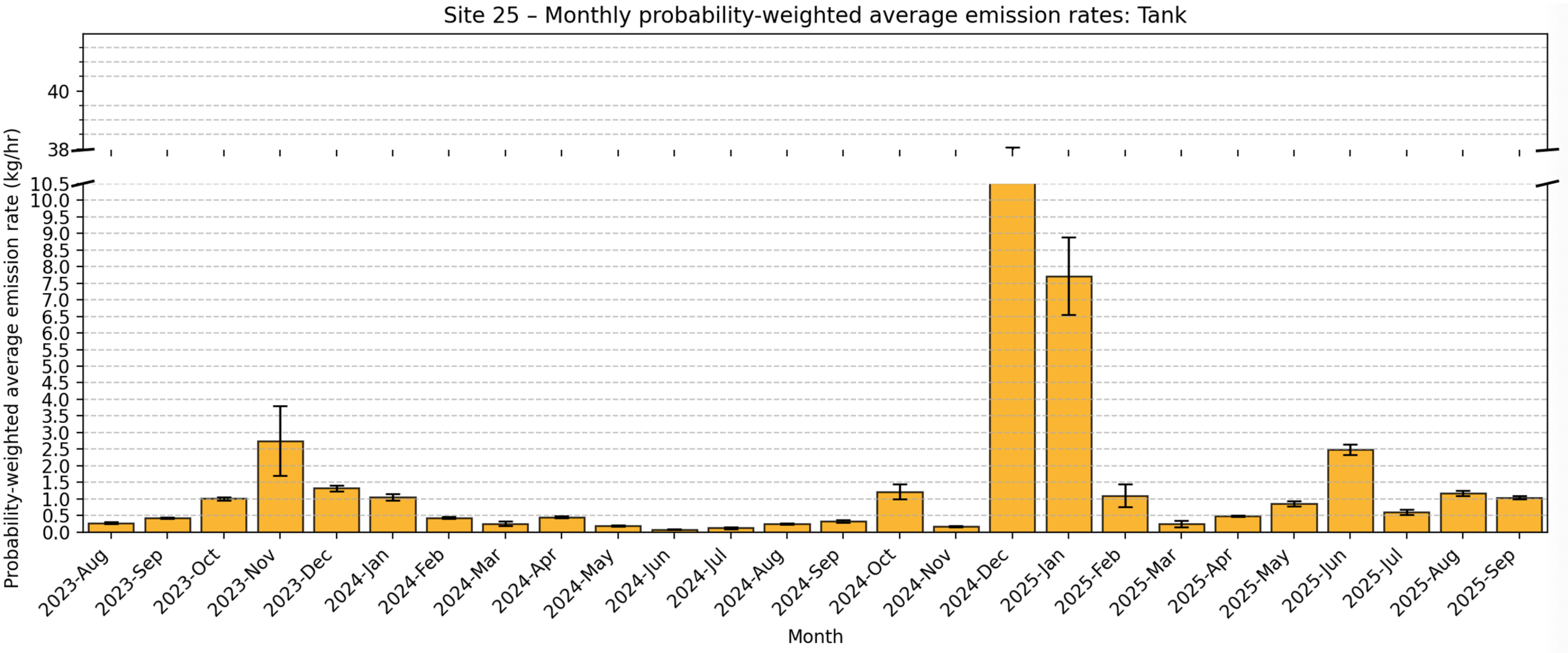
Tank emission rate estimates over time



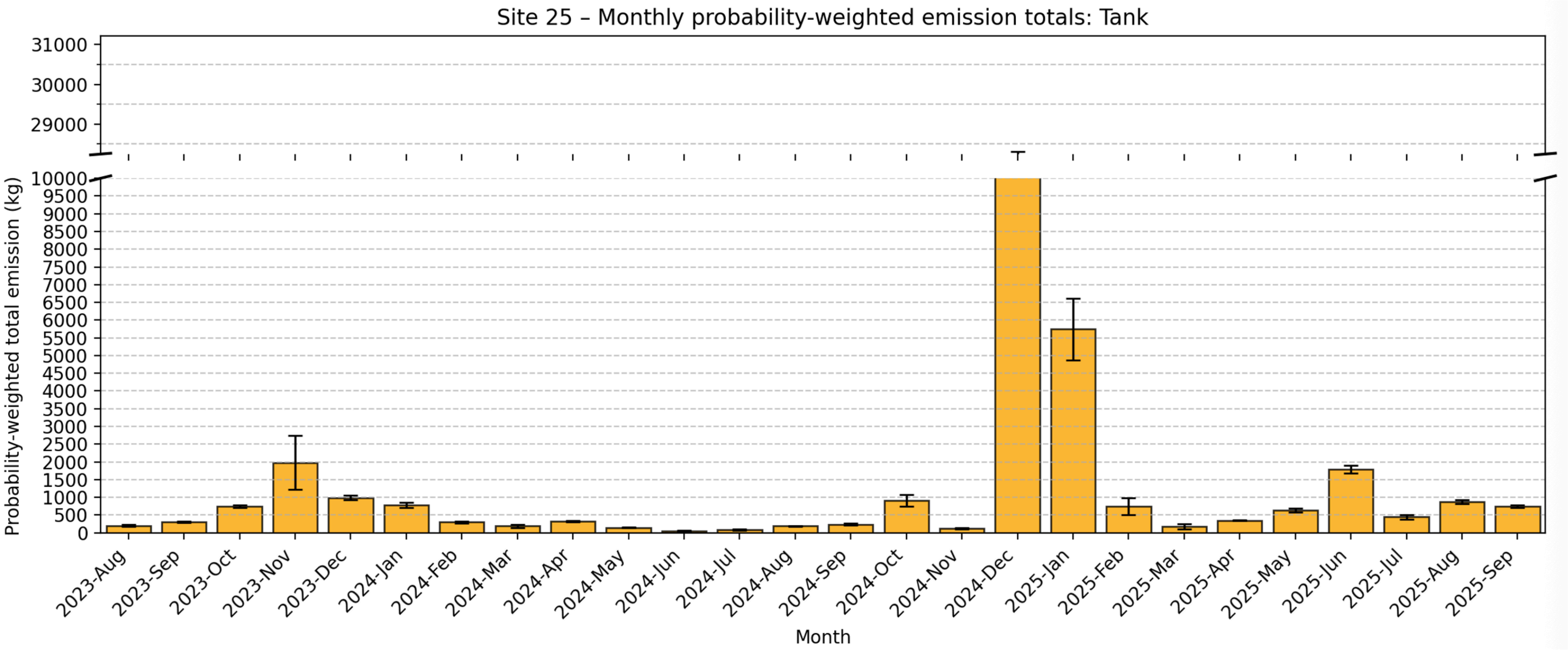
Tank emission rate estimates over time



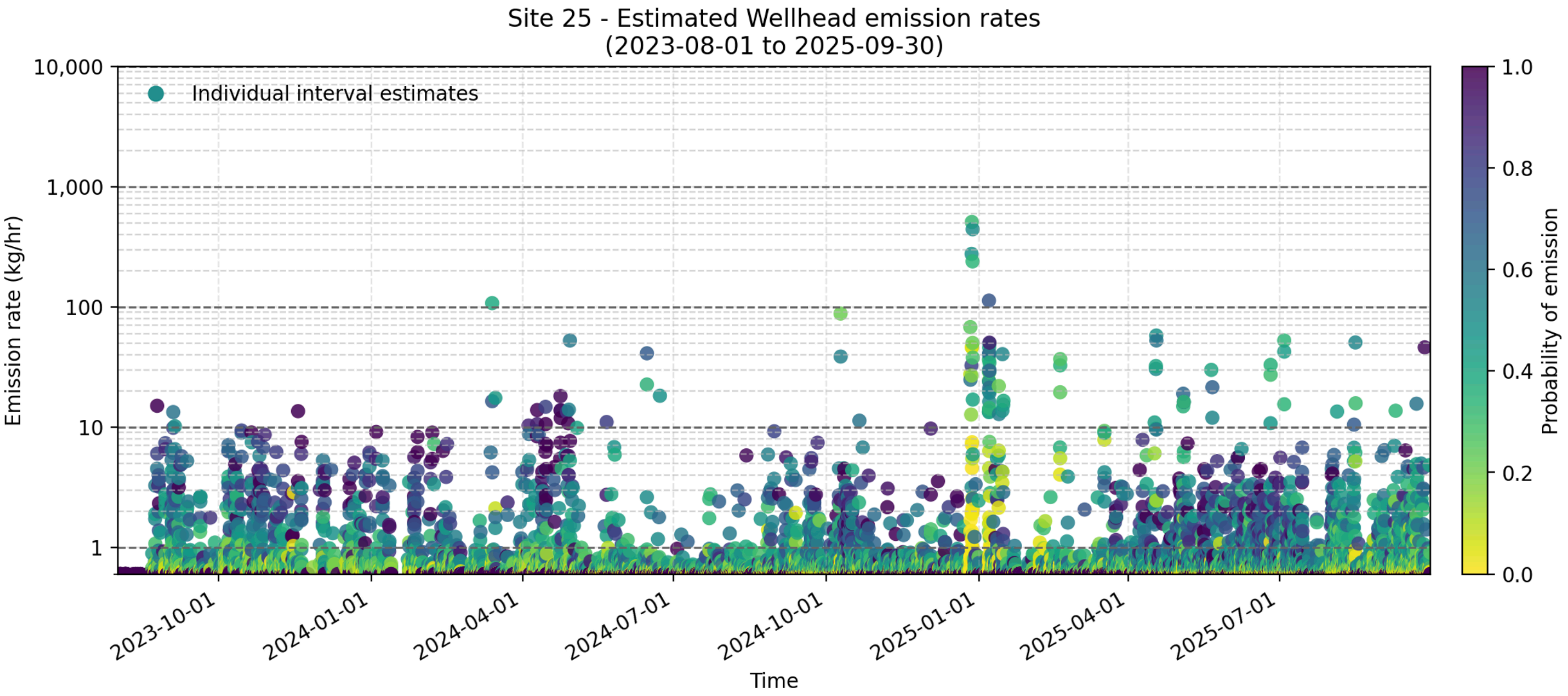
Tank emission rate estimates over time



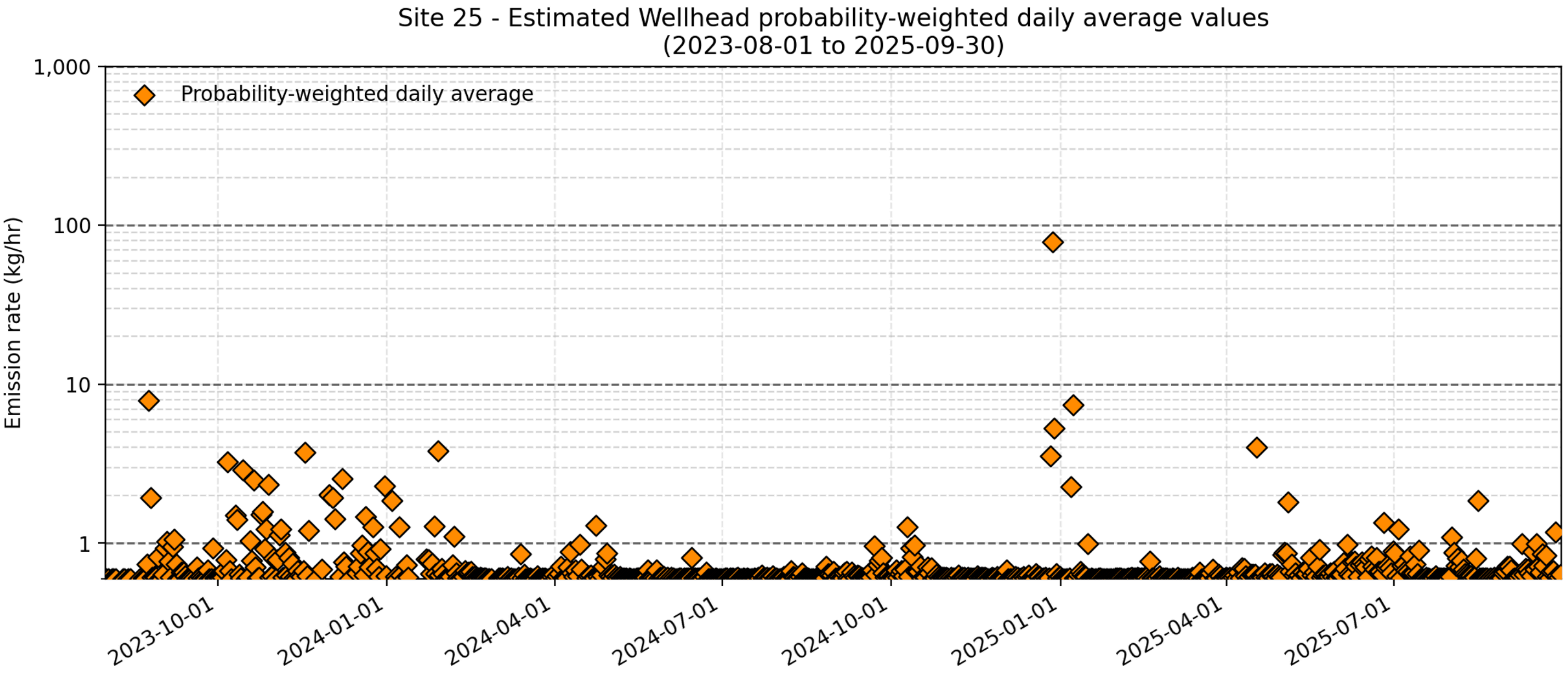
Tank emission rate estimates over time



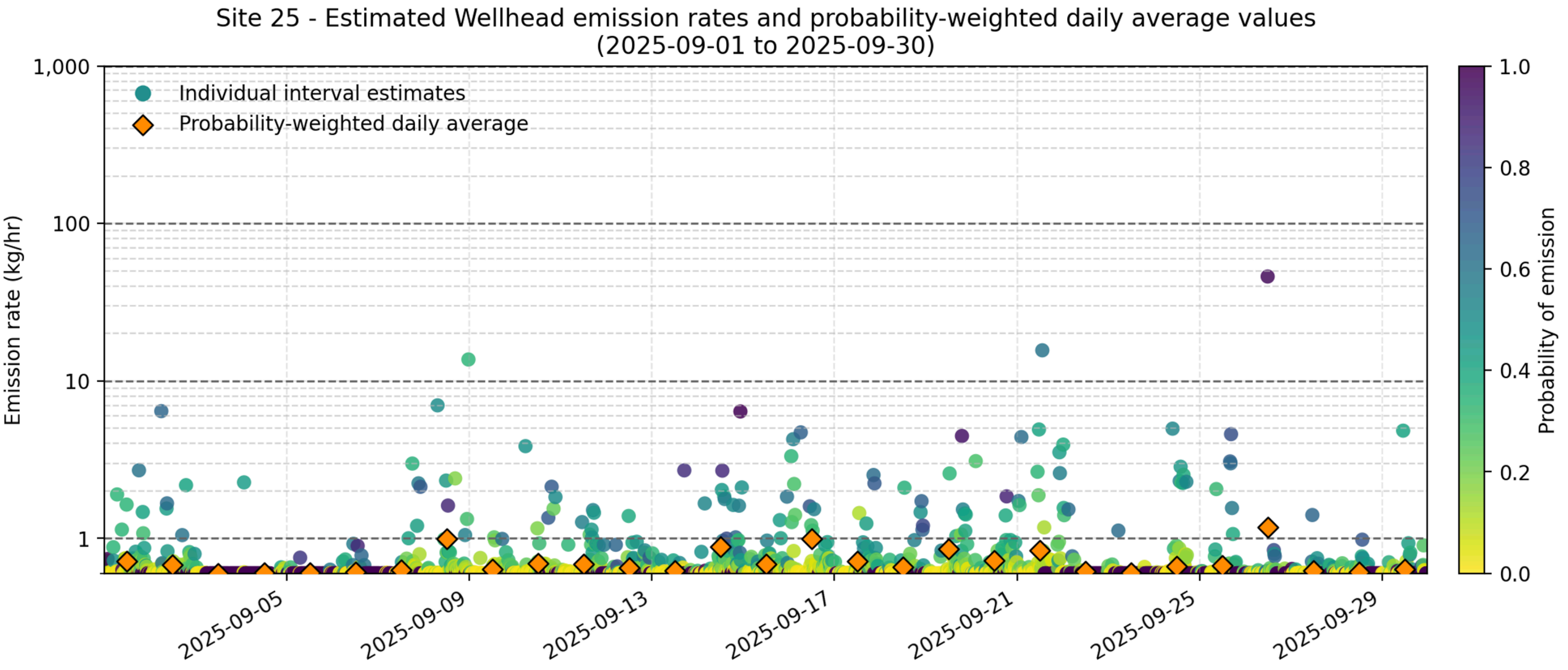
Wellhead emission rate estimates over time



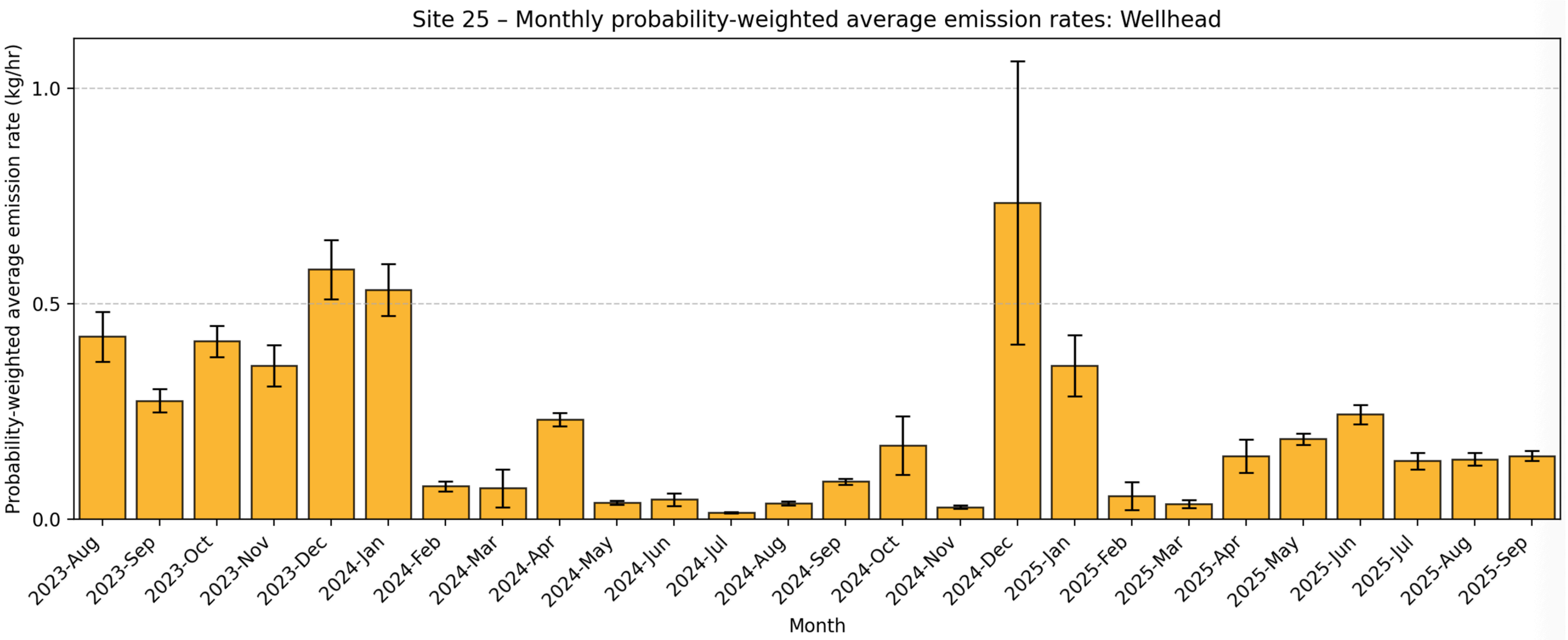
Wellhead emission rate estimates over time



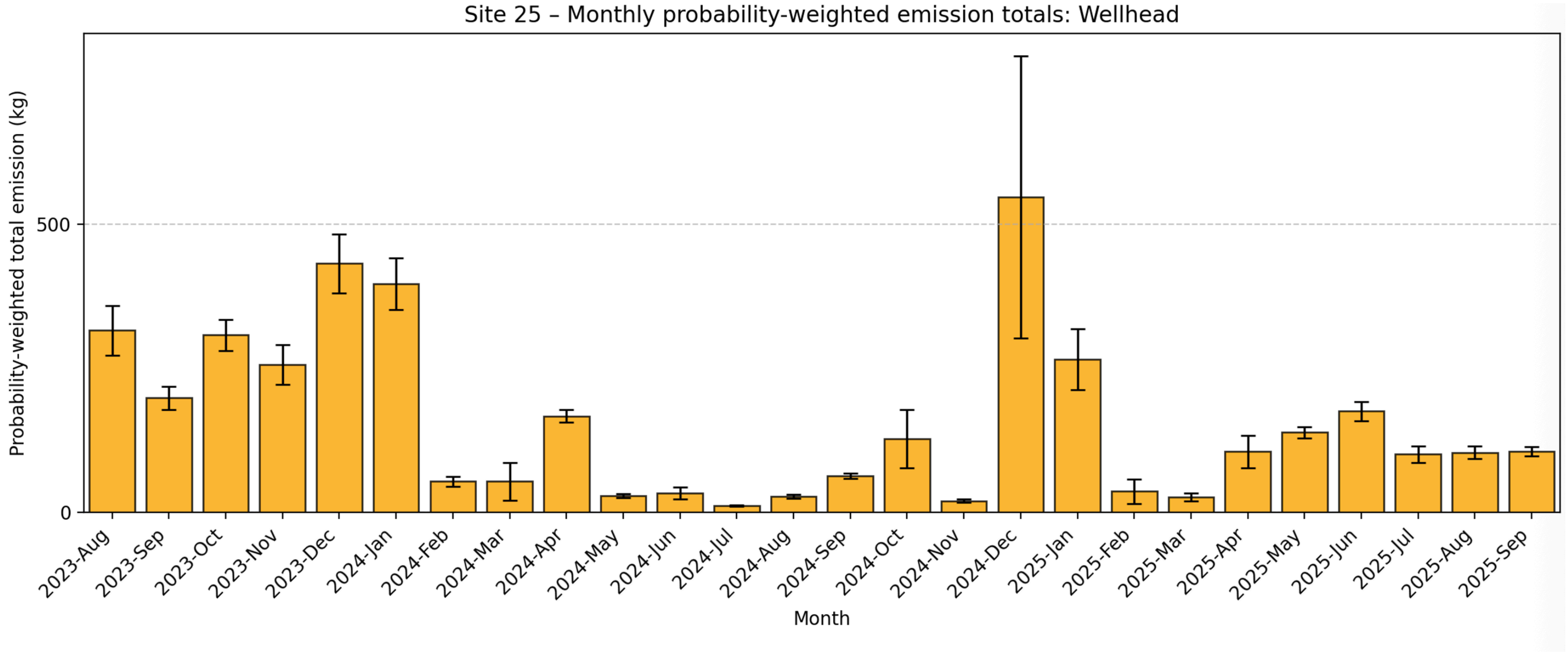
Wellhead emission rate estimates over time



Wellhead emission rate estimates over time

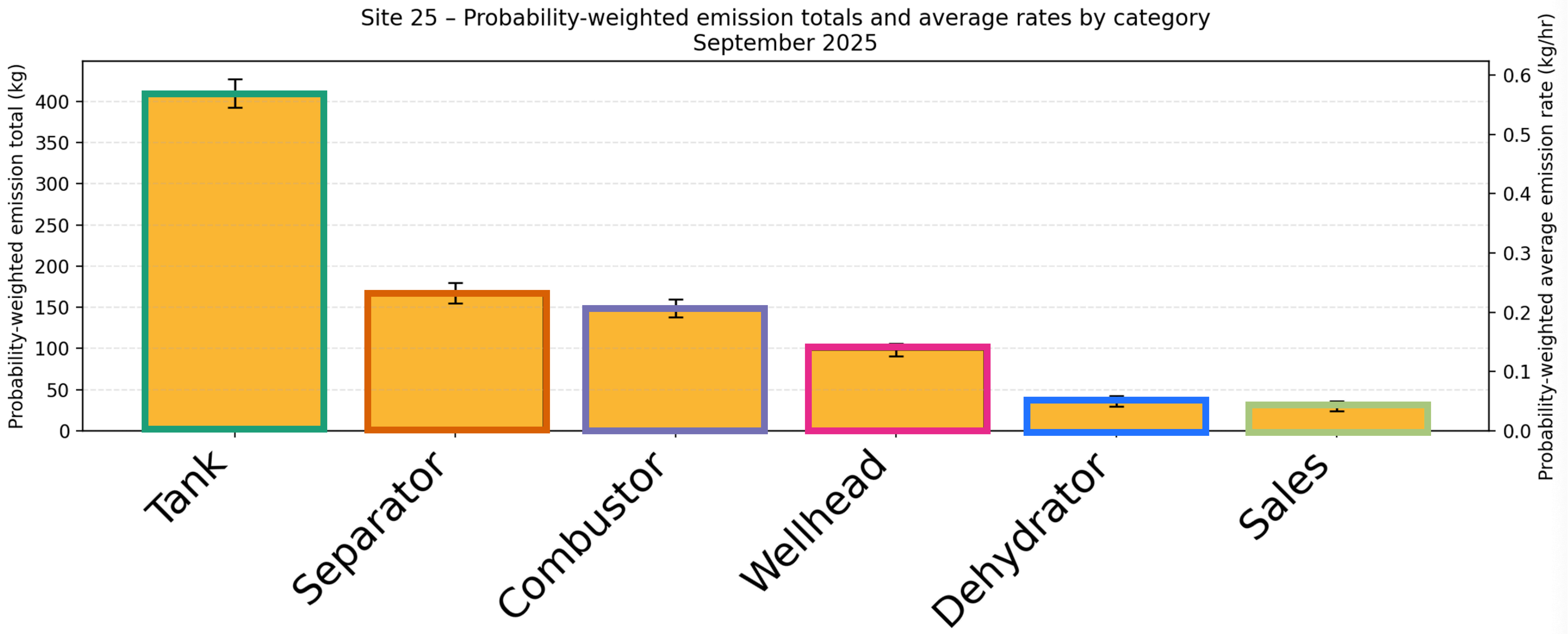


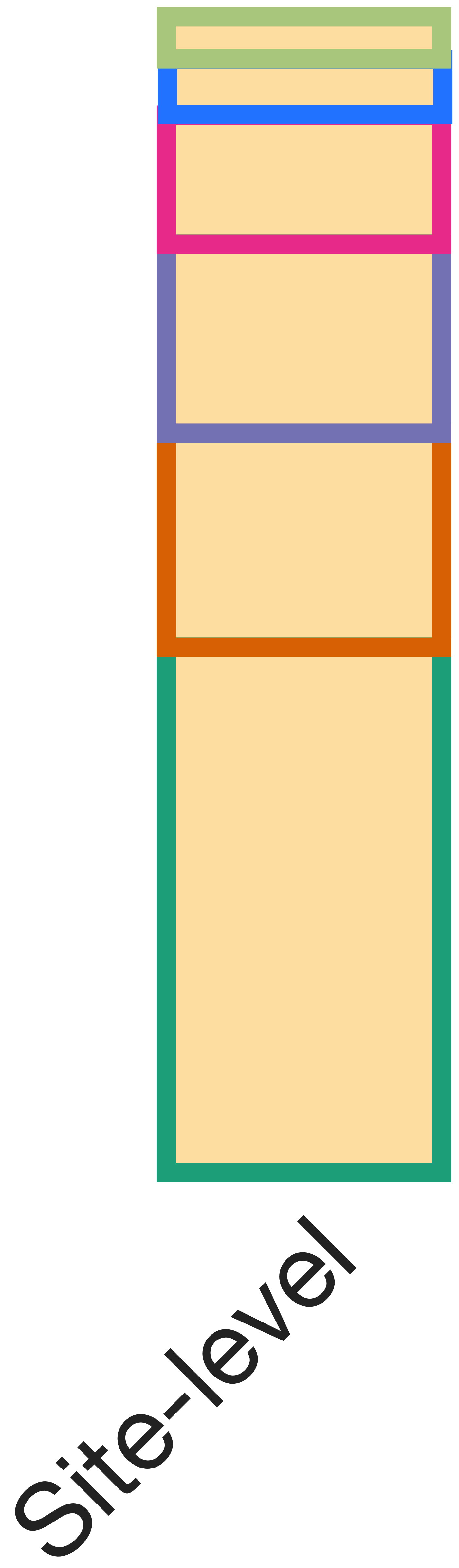
Wellhead emission rate estimates over time



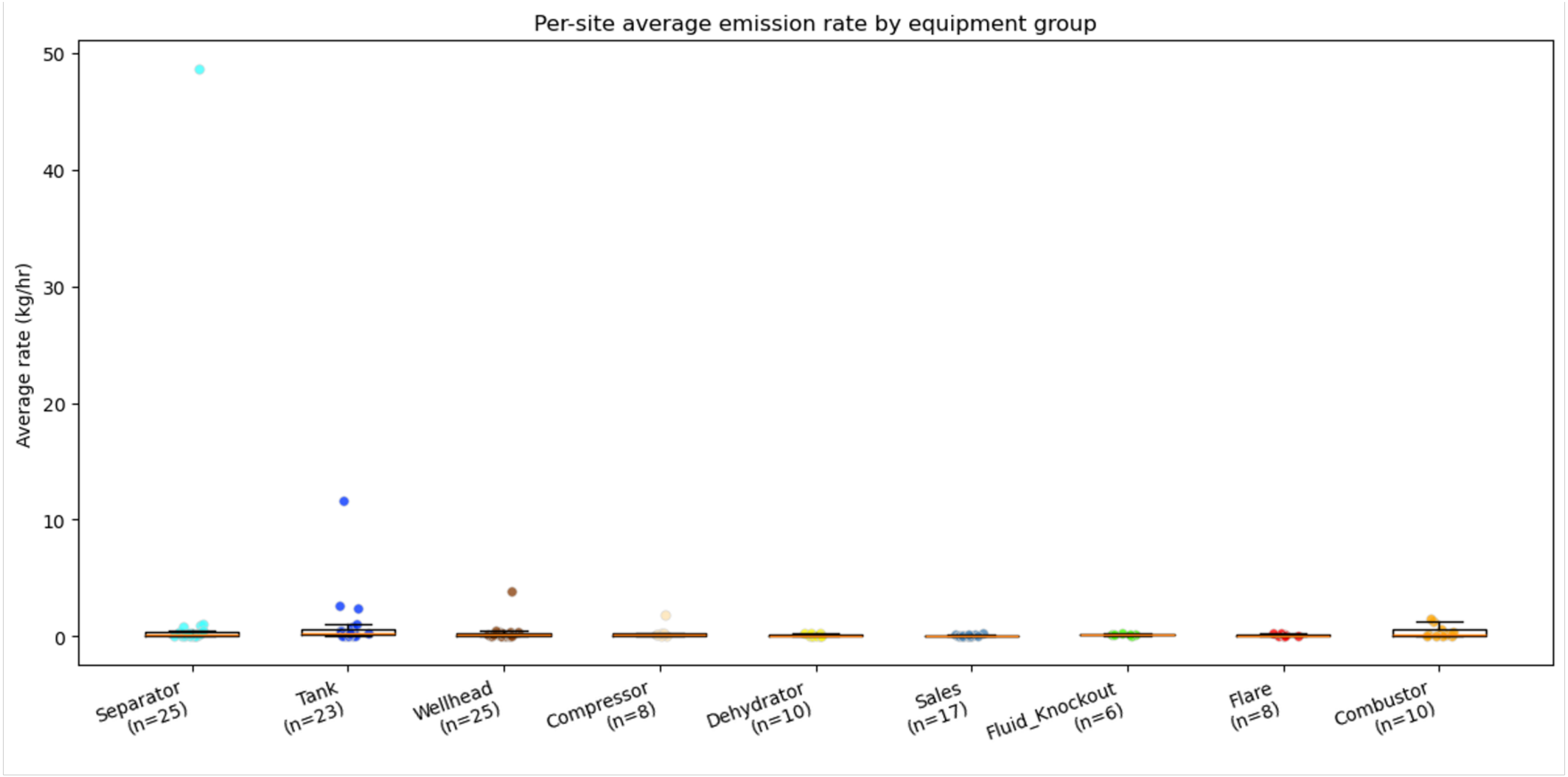
Site 25: September 2025 inventory

890 [830, 951] kg

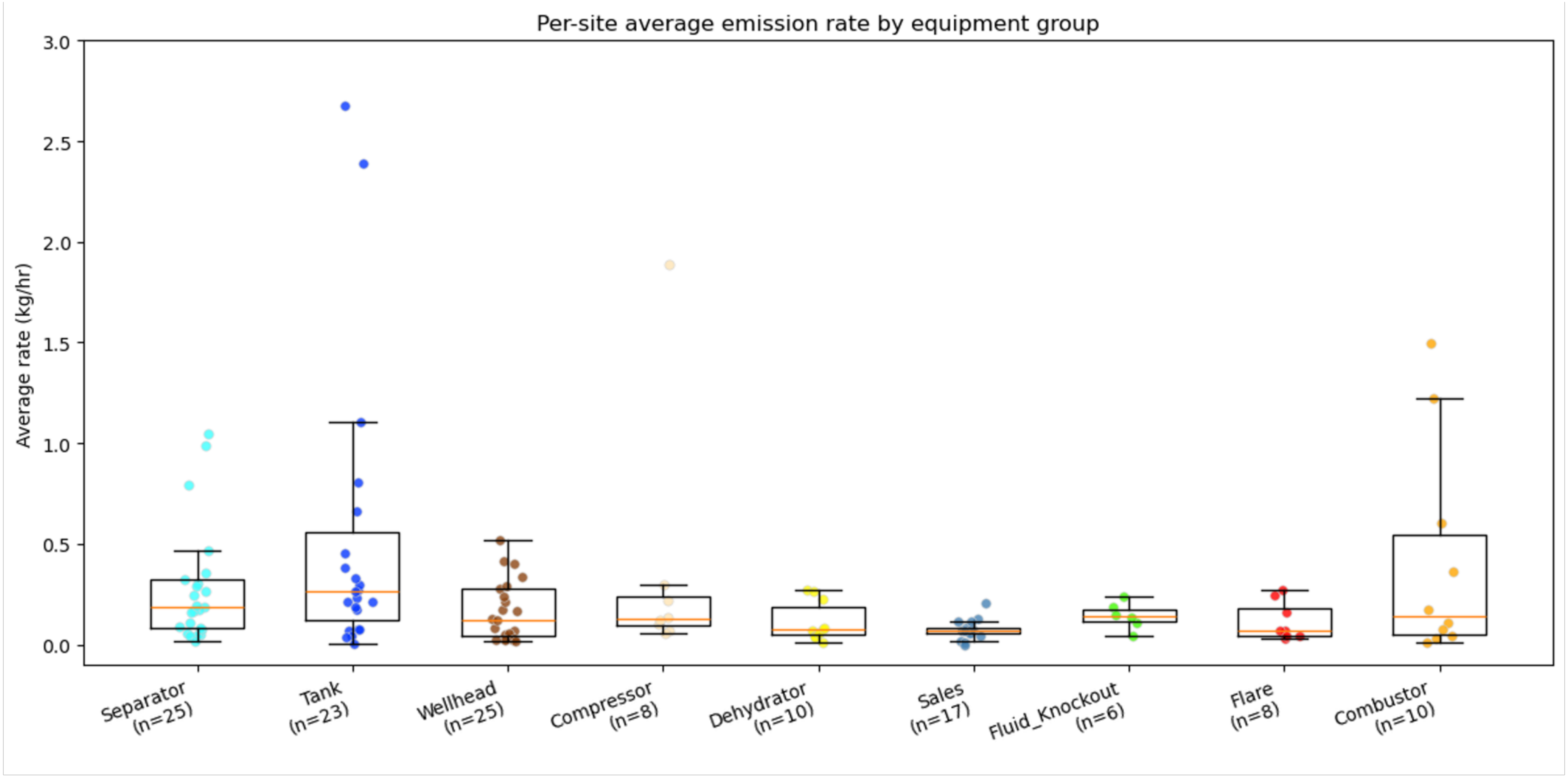




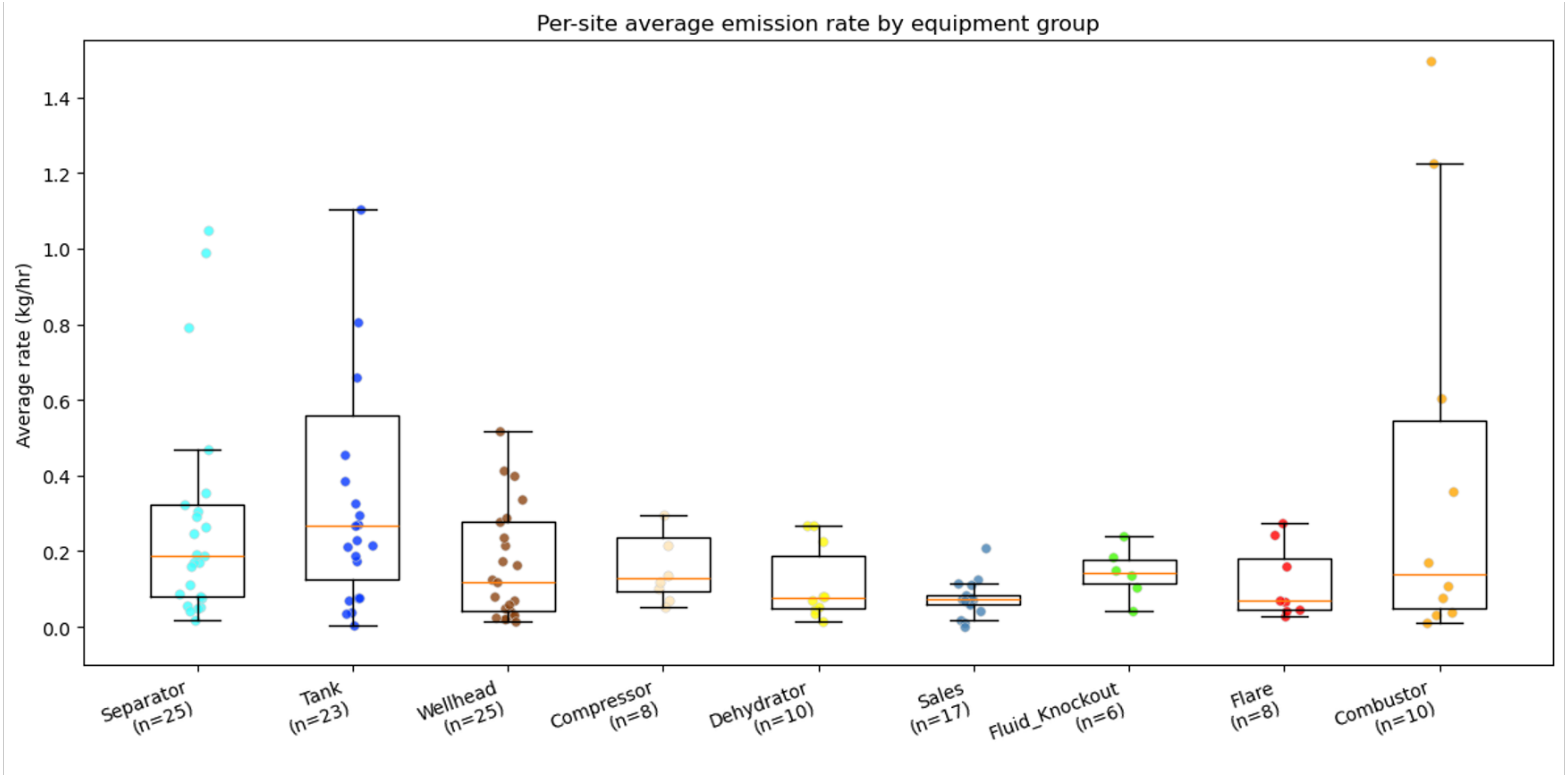
Summary across all 26 sites



Summary across all 26 sites



Summary across all 26 sites



Concluding thought:

- CMS provide enough measurements to create fully measurement-based inventories at the site-level… IF
  - You account for periods of no information
  - You have an unbiased inverse model (or know how to correct for the bias)
- There's a lot of information in the CMS-based emission rates, and we are just getting started analyzing it

Next steps:

- Compare to other inventory methods (UT MII, CSU MAES, GHGRP)
  - We have already done this on 5 sites for the COBE project in Colorado
- Use CMS-based emission rate estimates to inform "prototypical sites" or subsets of sites where distributions are similar
  - E.g. conventional wells

Thank you! Questions?

wdanie16@jhu.edu

COLORADO SCHOOL OF MINES

EEMDL Energy Emissions Modeling and Data Lab

U.S. DEPARTMENT OF ENERGY