Sampling variability under extreme skewness:

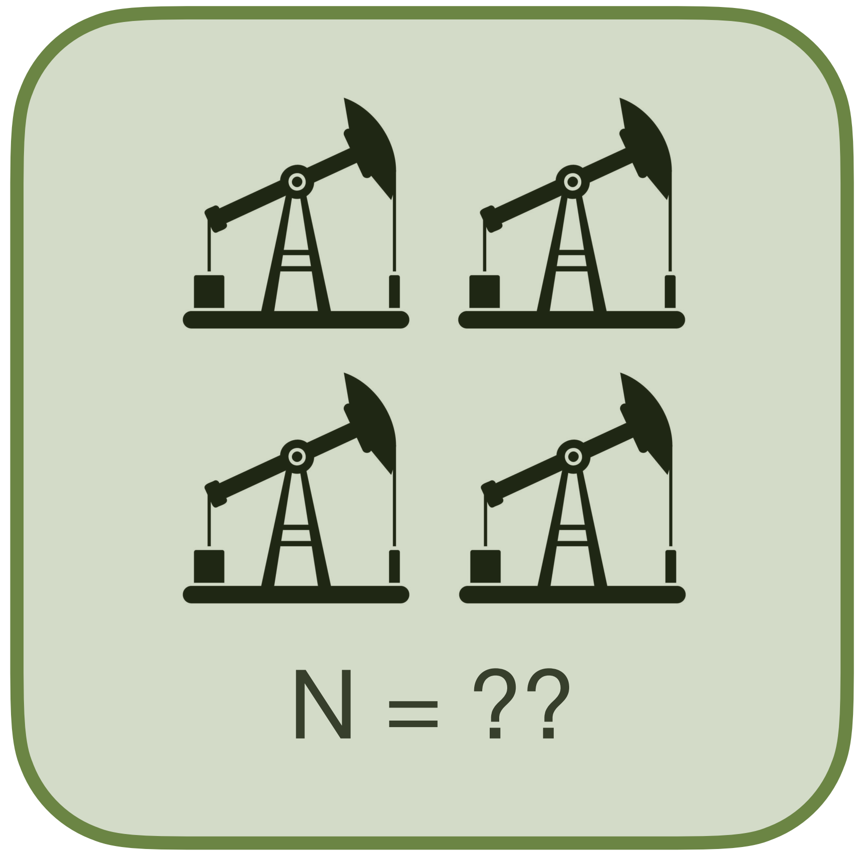
Sample size guidance for future methane measurement campaigns

William Daniels
Dept. of Environmental Health and Engineering
Johns Hopkins University

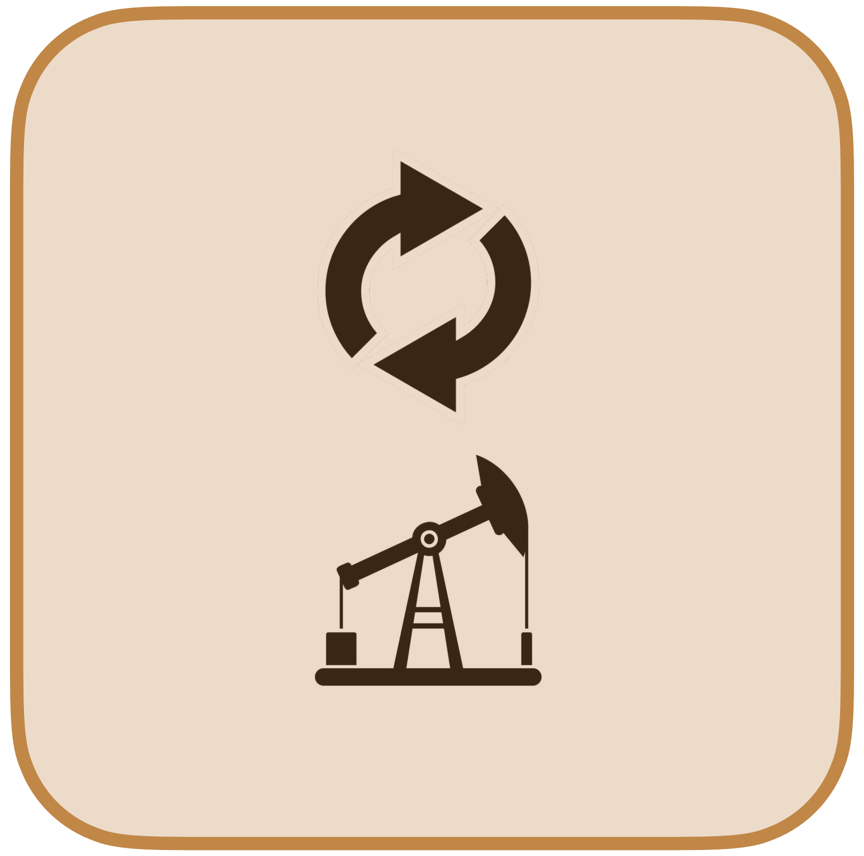
Michael Basanese, Dorit Hammerling
Dept. of Applied Mathematics and Statistics
Colorado School of Mines

Lots to consider when planning a measurement campaign

Budget



How many sites to measure



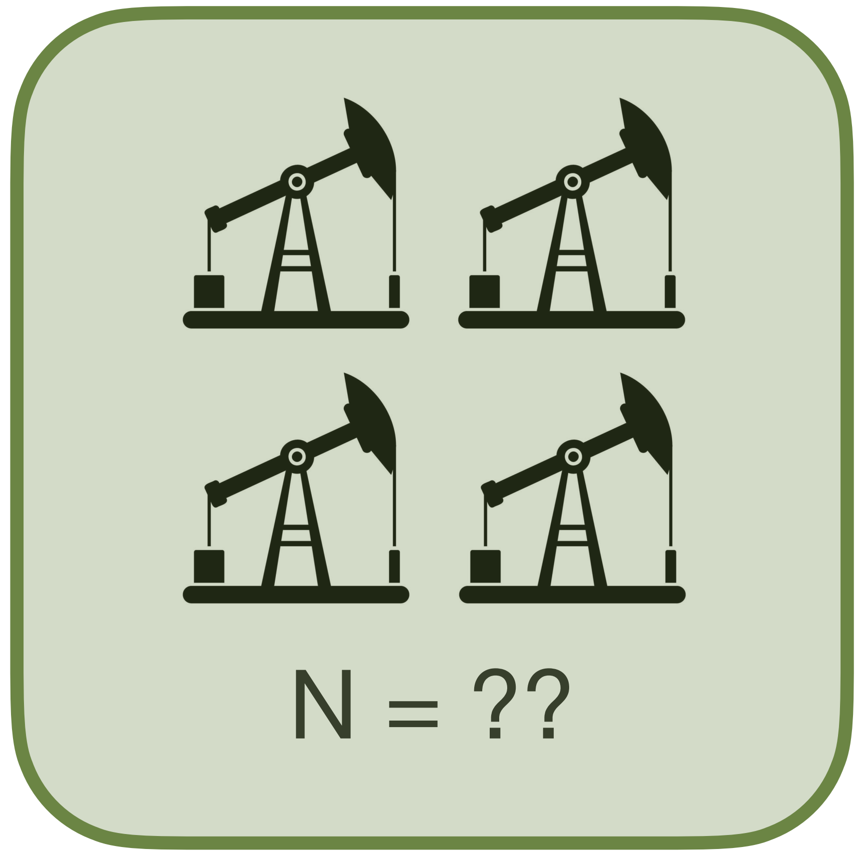
Site repeats vs. no repeats

Stratification across site types

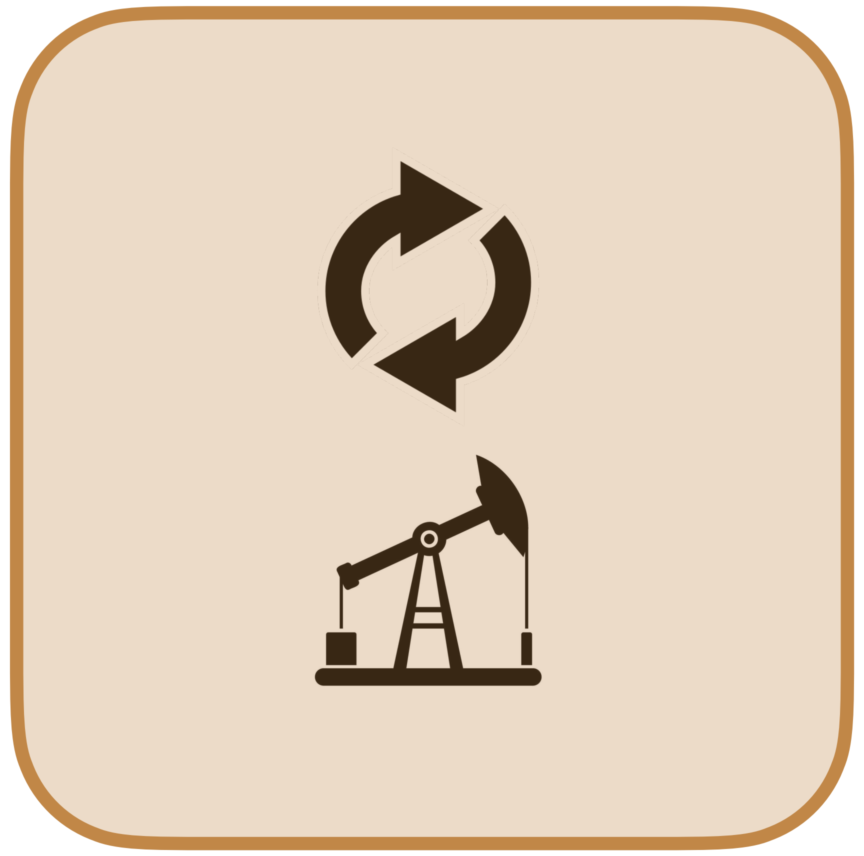
•••

Lots to consider when planning a measurement campaign

Budget



How many sites to measure



Site repeats vs. no repeats

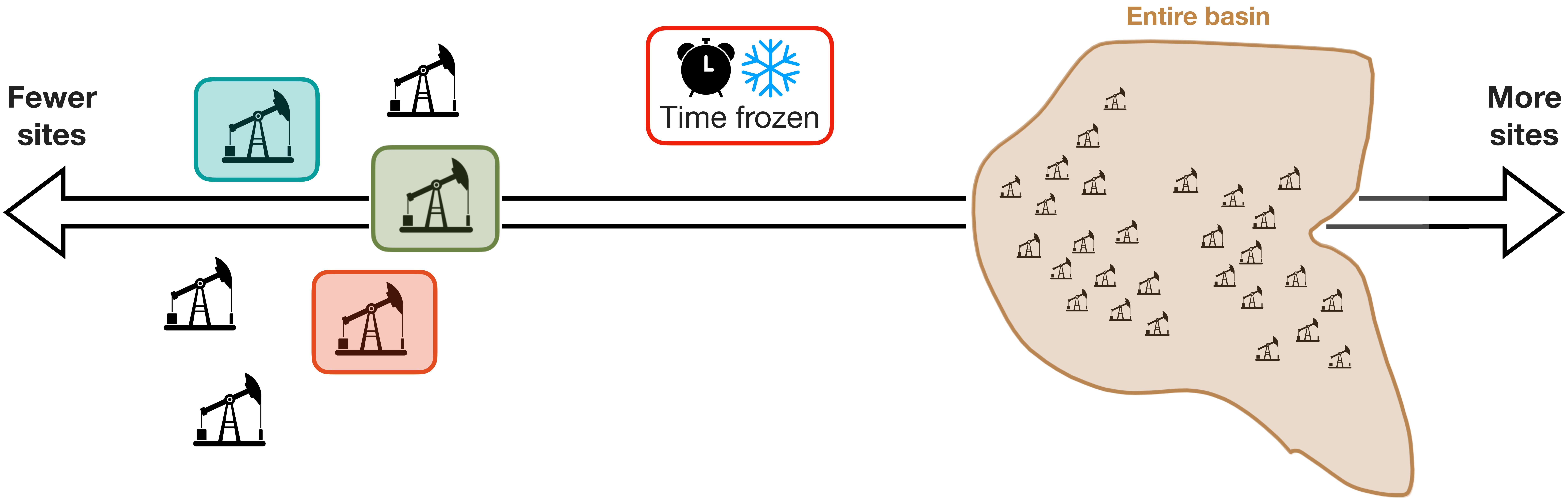
Stratification across site types

•••

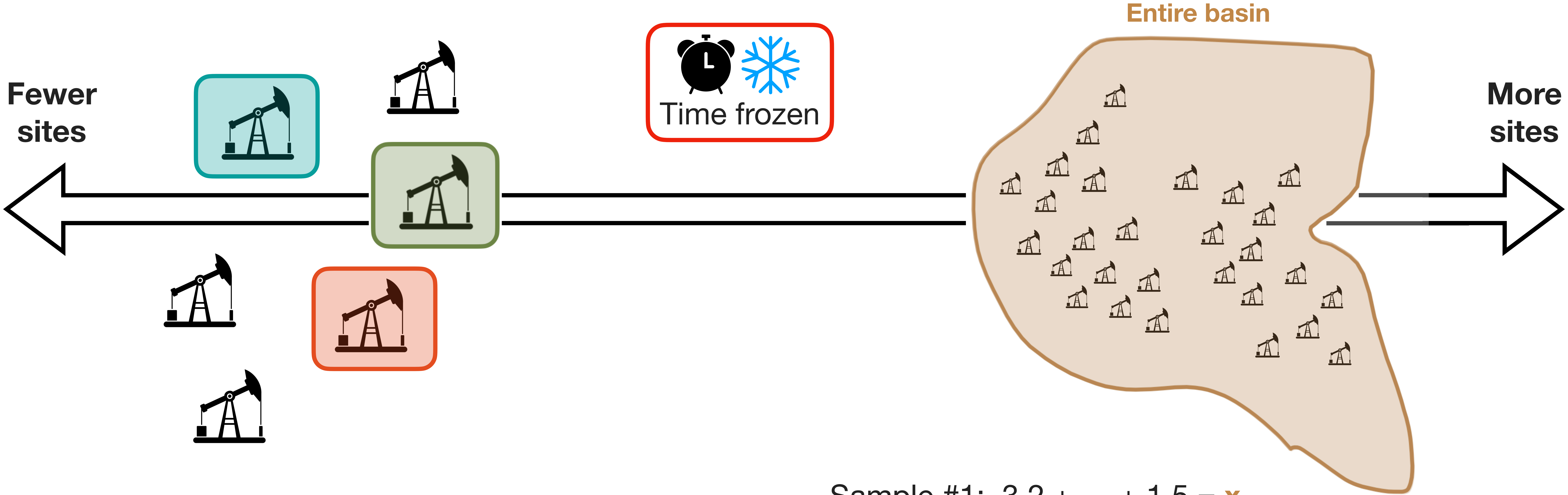
We focus on one piece of the puzzle…

… and approach it through the lens of statistical variability

What is sampling variability?



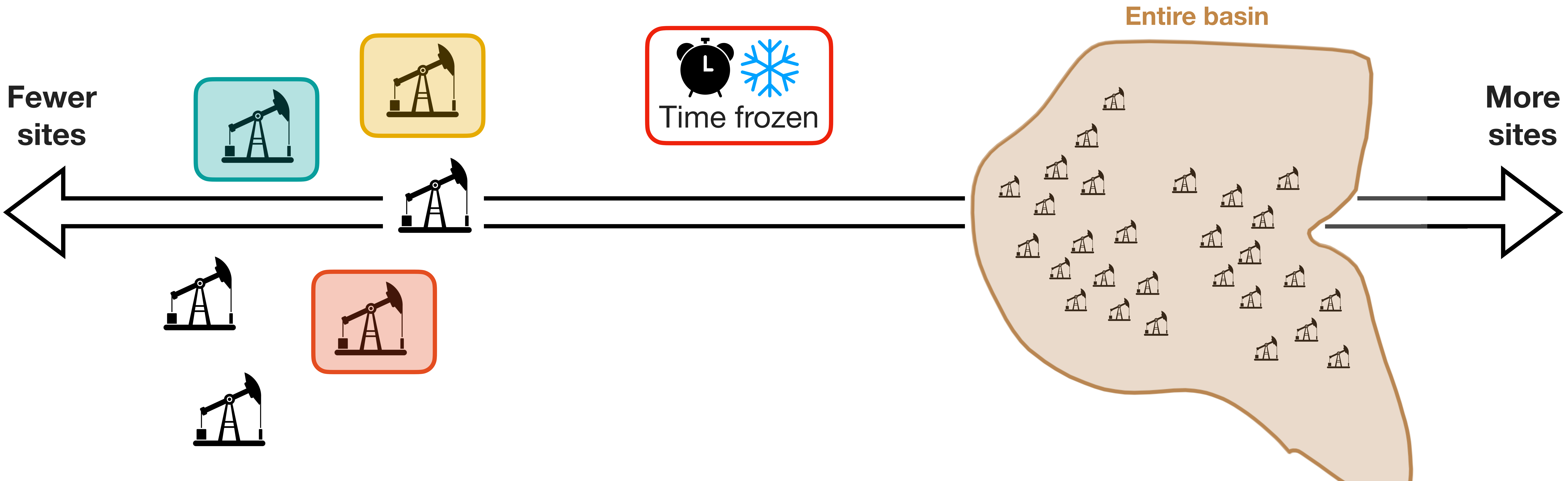
What is sampling variability?



Sample #1: **3** + **0** + **1.5** = 4.5

Sample #1: 3.2 + … + 1.5 = **x**

What is sampling variability?



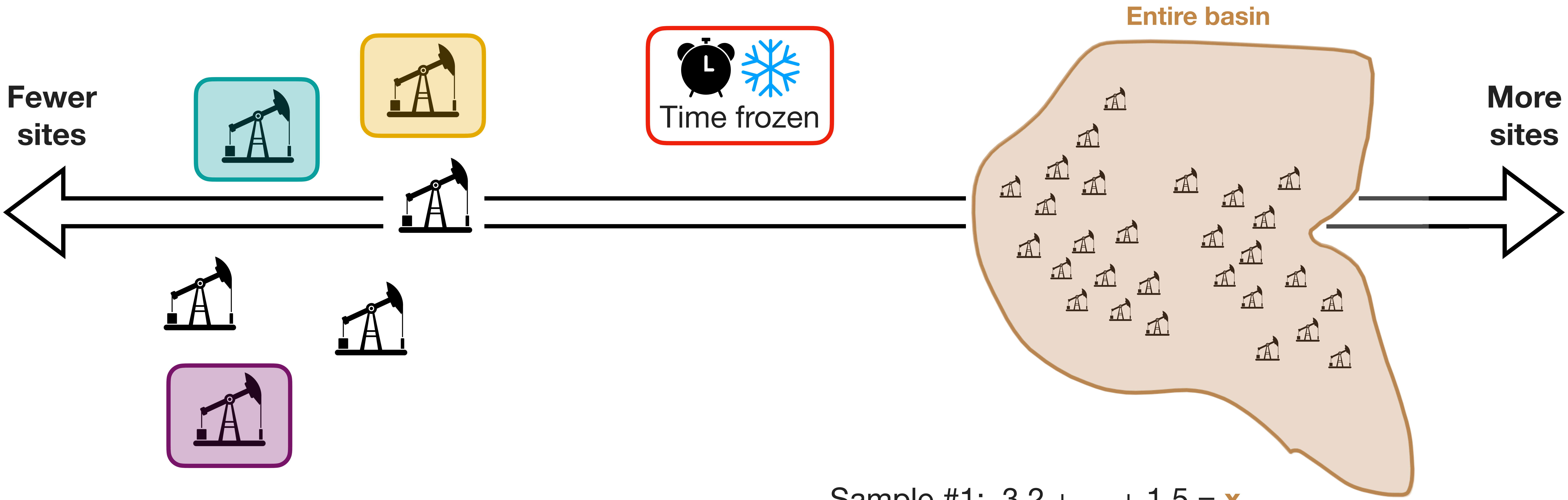
Sample #1: 3 + 0 + 1.5 = 4.5

Sample #2: 3 + 1.5 + 0.4 = 4.9

Sample #1: 3.2 + … + 1.5 = x

Sample #2: 3.2 + … + 1.5 = x

What is sampling variability?



Sample #1: 3 + 0 + 1.5 = 4.5

Sample #2: 3 + 1.5 + 0.4 = 4.9

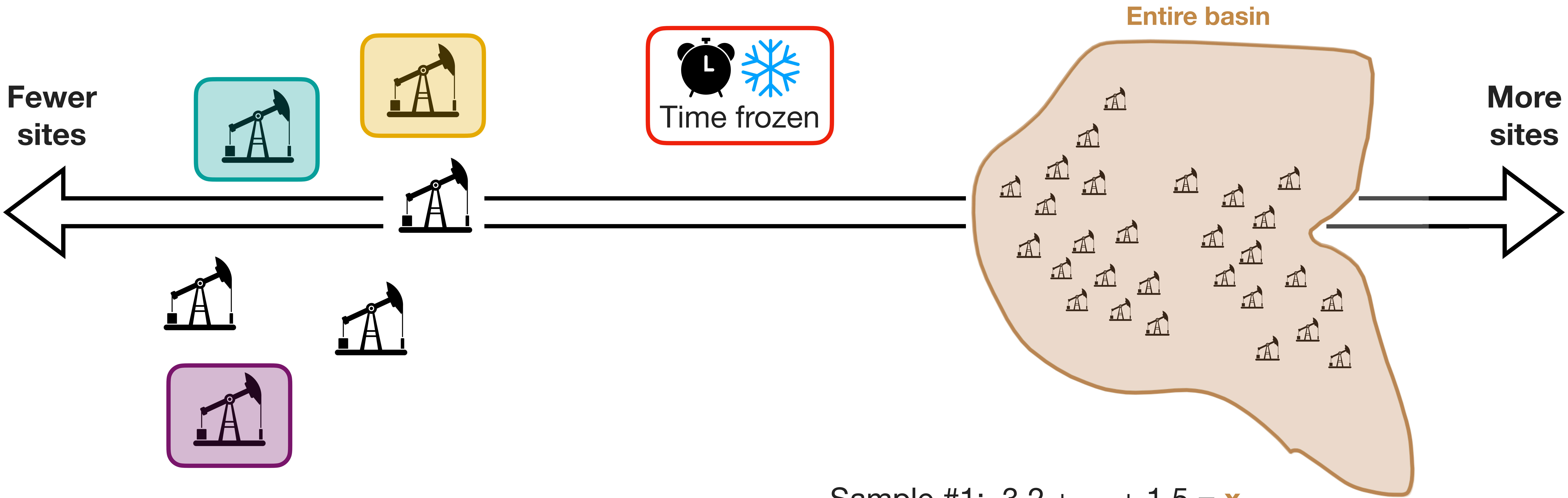
Sample #2: 3 + 0.4 + 243 = 246.4

Sample #1: 3.2 + … + 1.5 = x

Sample #2: 3.2 + … + 1.5 = x

Sample #3: 3.2 + … + 1.5 = x

What is sampling variability?



Sample #1: 3 + 0 + 1.5 = 4.5

Sample #2: 3 + 1.5 + 0.4 = 4.9

Sample #2: 3 + 0.4 + 243 = 246.4

⋮

Sampling variability:
55x difference in the sum between samples

Sample #1: 3.2 + … + 1.5 = x

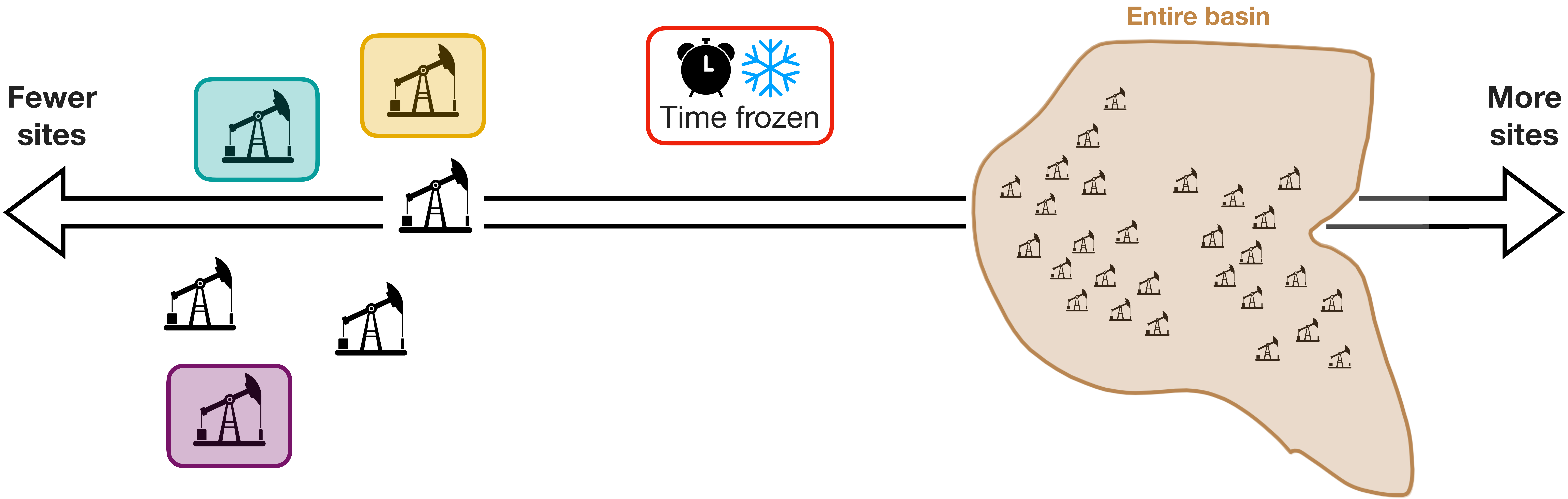
Sample #2: 3.2 + … + 1.5 = x

Sample #3: 3.2 + … + 1.5 = x

⋮

Sampling variability:
1x difference in the sum between samples

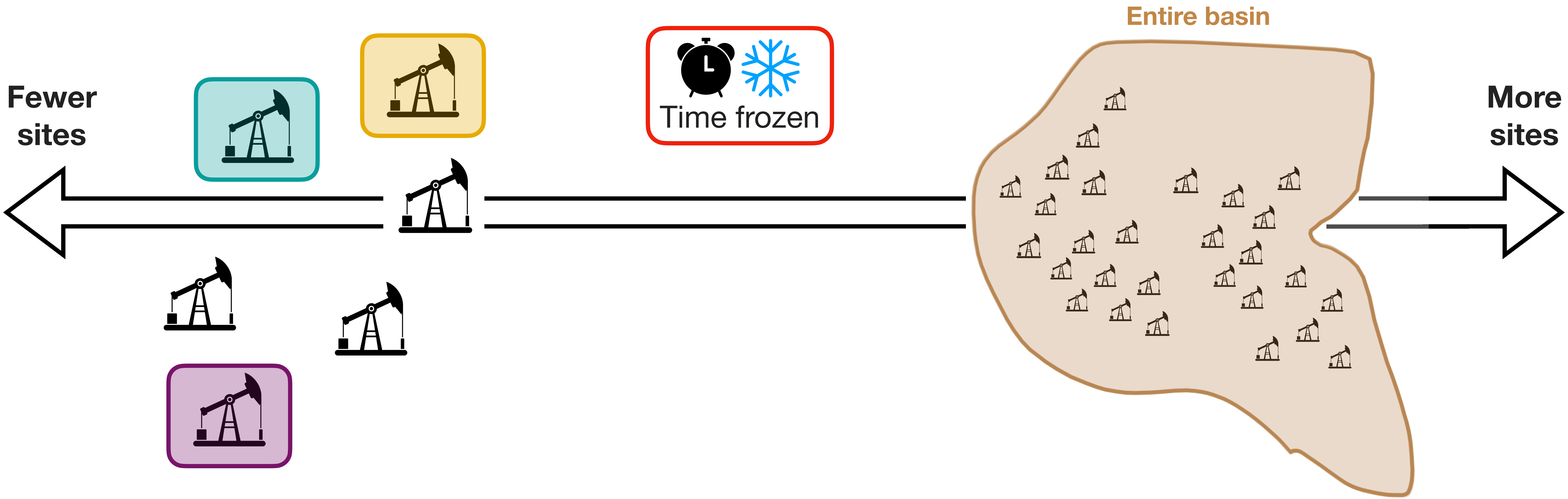
(exactly the same, by definition)



Sampling variability:
55x difference in the sum between samples

What are we trying to do?

Sampling variability:
1x difference in the sum between samples

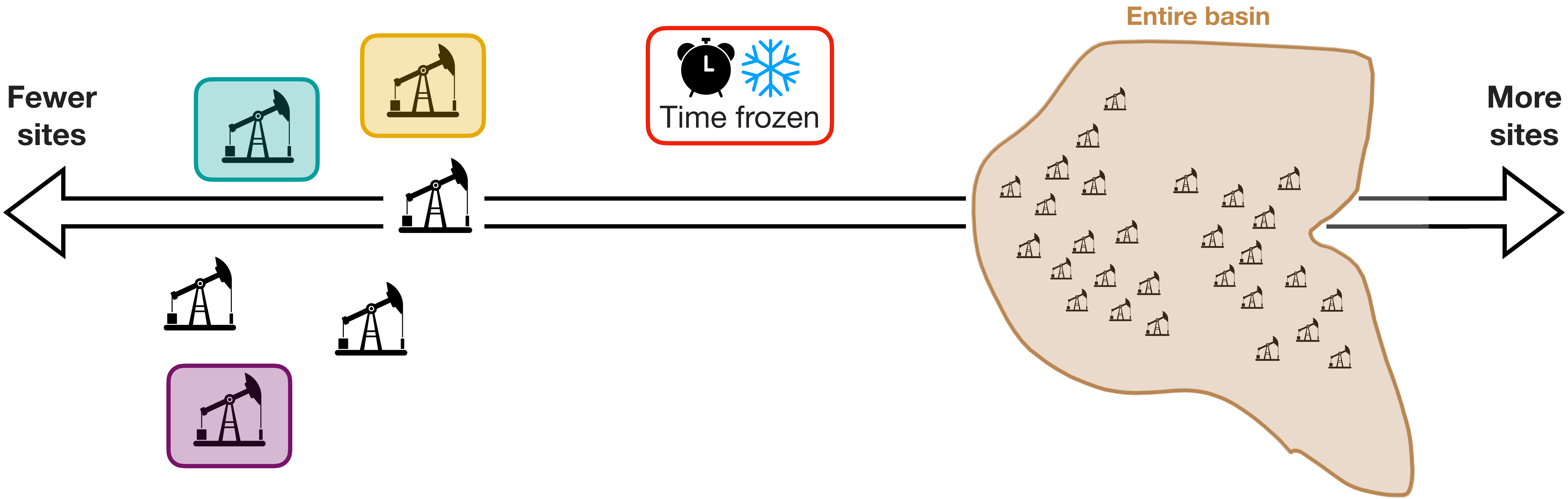


Sampling variability:
55x difference in the sum between samples

What are we trying to do?

Sampling variability:
1x difference in the sum between samples

How quickly does the 55x converge to 1x as you measure more sites?



Sampling variability:
55x difference in the sum between samples

What are we trying to do?

Sampling variability:
1x difference in the sum between samples

How quickly does the 55x converge to 1x as you measure more sites?

How many sites do you need to measure to limit sampling variability to a given threshold?

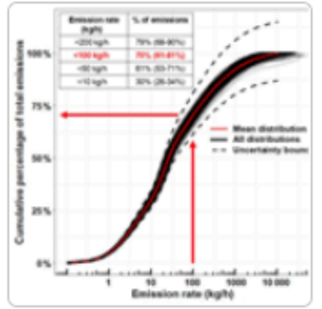
This would be an easy problem if we knew everything in advance..

This would be an easy problem if we knew everything in advance..

.. but we don't. Instead, turn to emission rate distributions from previous campaigns.

Research article | 04 Feb 2025

Small emission sources in aggregate disproportionately account for a large majority of total methane emissions from the US oil and gas sector



James P. Williams, Mark Omara, Anthony Himmelberger, Daniel Zavala-Araiza, Katlyn MacKay, Joshua Benmergui, Maryann Sargent, Steven C. Wofsy, Steven P. Hamburg, and Ritesh Gautam

Williams et al. (2025)

- Tracer-based, inverse Gaussian, OTM 33, Hi-Flow
- Relatively low detection thresholds: ~0.1 kg/hr

Colorado Ongoing Basin Emissions (COBE) Updated Final Report

Jenna A. Brown[1], Michael Moy[1], Arthur Santos[1], Ethan Rimelman[1], Winrose Mollel[1], Olga Khaliukova[2], Callan Okenberg[2], William S. Daniels[2], Dorit M. Hammerling[2], Daniel Zimmerle[1], and Anna L. Hodshire[1]

[1]Energy Institute, Methane Emissions Technology Evaluation Center, Colorado State University, Fort Collins, Colorado, USA
[2]Dept. of Applied Mathematics and Statistics, Colorado School of Mines, Golden, Colorado, USA

November 20, 2025

Brown et al. (2025)

- Bridger GML, GHGSat, and Insight M
- Varying detection thresholds: ~3-50 kg/hr

Article | Published: 13 March 2024

US oil and gas system emissions from nearly one million aerial site measurements

Evan D. Sherwin, Jeffrey S. Rutherford, Zhan Zhang, Yuanlei Chen, Erin B. Wetherley, Petr V. Yakovlev, Elena S. F. Berman, Brian B. Jones, Daniel H. Cusworth, Andrew K. Thorpe, Alana K. Ayasse, Riley M. Duren & Adam R. Brandt

Nature **627**, 328–334 (2024) | Cite this article

Sherwin et al. (2024)

- Carbon Mapper (AVIRIS-NG) and Insight M
- Higher detection thresholds: ~6-90 kg/hr

Environmental Science & Technology > Vol 57/Issue 33 > Article

Cite Share Jump to Expand

Open Access

ENERGY AND CLIMATE | August 10, 2023

Extension of Methane Emission Rate Distribution for Permian Basin Oil and Gas Production Infrastructure by Aerial LiDAR

William M. Kunkel*, Asa E. Carre-Burritt, Grant S. Aivazian, Nicholas C. Snow, Jacob T. Harris, Tagert S. Mueller, Peter A. Roos, and Michael J. Thorpe*

Kunkel et al. (2023)

- Bridger GML
- Relatively low detection threshold: ~3 kg/hr

This would be an easy problem if we knew everything in advance..

.. but we don't. Instead, turn to emission rate distributions from previous campaigns.

This would be an easy problem if we knew everything in advance..

.. but we don't. Instead, turn to emission rate distributions from previous campaigns.

Generate sample size guidance from each reference distribution

based on three error metrics for the sample mean

For each reference distribution:

1. Let μ represent the "true" distribution mean
2. Sample 1,000 times without replacement at sample sizes ranging from 1 to the length of the distribution
3. For each sample size:
    a. Let $\bar{\boldsymbol{x}}^* = \{\bar{x}_1^*, \ldots, \bar{x}_{1000}^*\}$ represent the 1,000 sample means
    b. Compute the percent error in the sample means as $\boldsymbol{\epsilon} = 100 \times (\bar{\boldsymbol{x}}^* - \mu)/\mu$
    c. Compute error metrics
        i. $\mathrm{Q}_{50} = \text{median}(\boldsymbol{\epsilon})$
        ii. $\epsilon_{\text{max}} = \max(\boldsymbol{\epsilon})$
        iii. $\mathbb{P}_{10} = \dfrac{\text{number of } |\boldsymbol{\epsilon}| < 10}{\text{length of } \boldsymbol{\epsilon}}$
4. Sample size guidance: pick the smallest sample that keeps the error metrics within a given threshold

Generate sample size guidance from each reference distribution

based on three error metrics for the sample mean

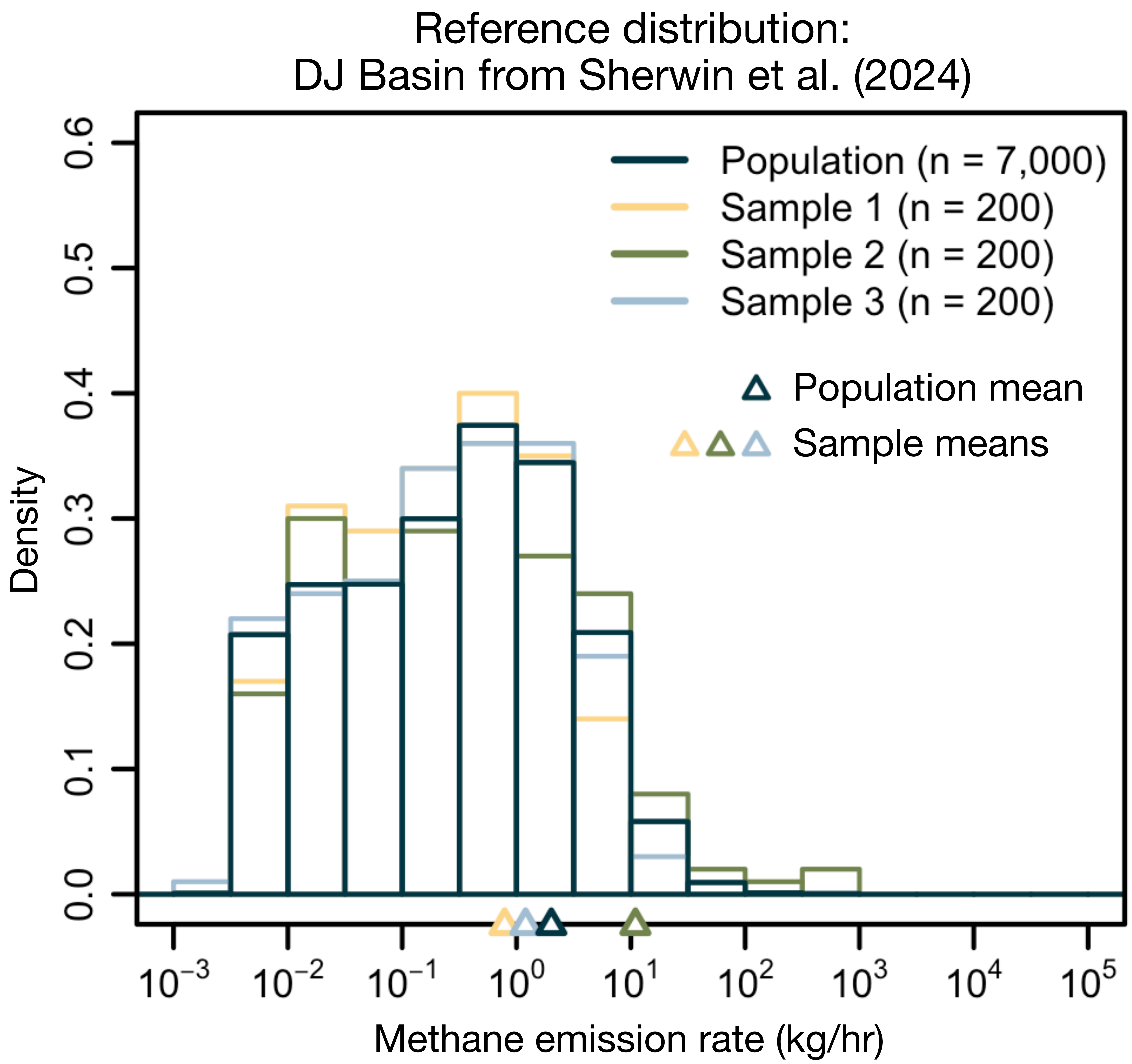
For each reference distribution:

1. Let μ represent the "true" distribution mean
2. Sample 1,000 times without replacement at sample sizes ranging from 1 to the length of the distribution
3. For each sample size:
    a. Let $\bar{\boldsymbol{x}}^* = \{\bar{x}_1^*, \ldots, \bar{x}_{1000}^*\}$ represent the 1,000 sample means
    b. Compute the percent error in the sample means as $\boldsymbol{\epsilon} = 100 \times (\bar{\boldsymbol{x}}^* - \mu)/\mu$
    c. Compute error metrics
        i. $\mathrm{Q}_{50} = \mathrm{median}(\boldsymbol{\epsilon})$
        ii. $\epsilon_{\mathrm{max}} = \mathrm{max}(\boldsymbol{\epsilon})$
        iii. $\mathbb{P}_{10} = \dfrac{\text{number of } |\boldsymbol{\epsilon}| < 10}{\text{length of } \boldsymbol{\epsilon}}$
4. Sample size guidance: pick the smallest sample that keeps the error metrics within a given threshold

Why base sample size guidance on the sample mean?

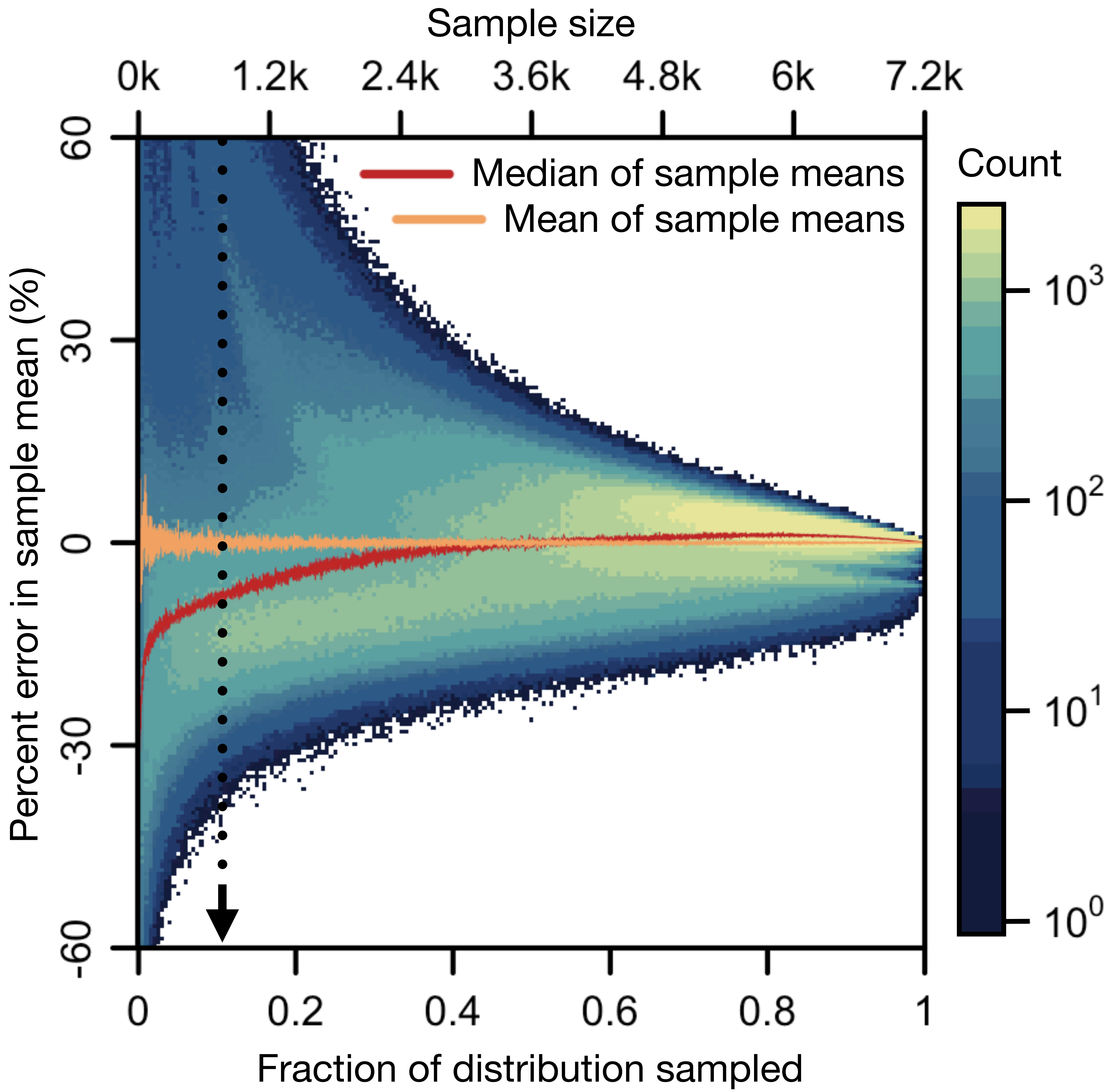
Why sample without replacement?

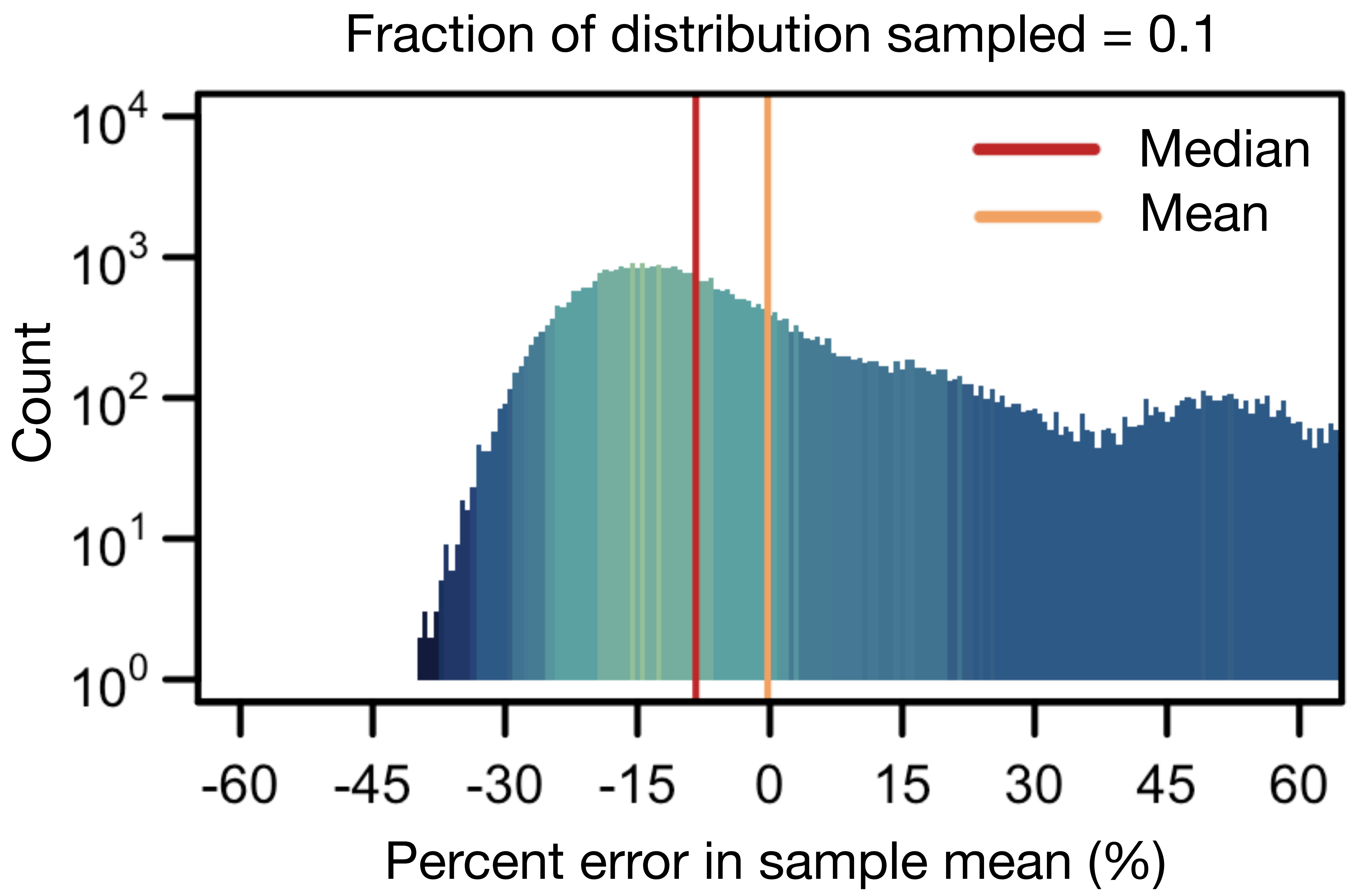
Extreme skew can cause large sampling variability



Extreme skew can cause large sampling variability

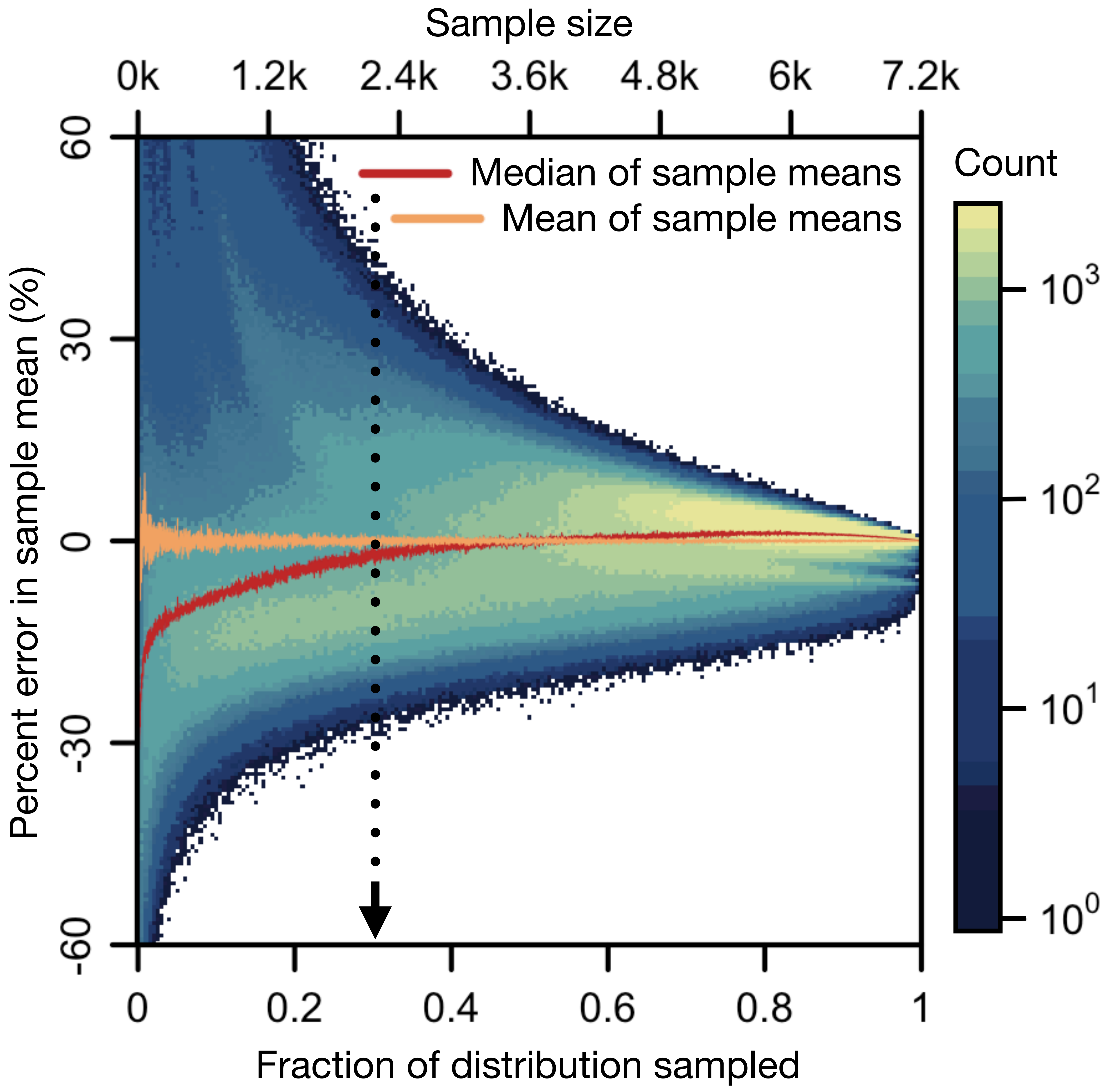
Extreme skew can cause large sampling variability

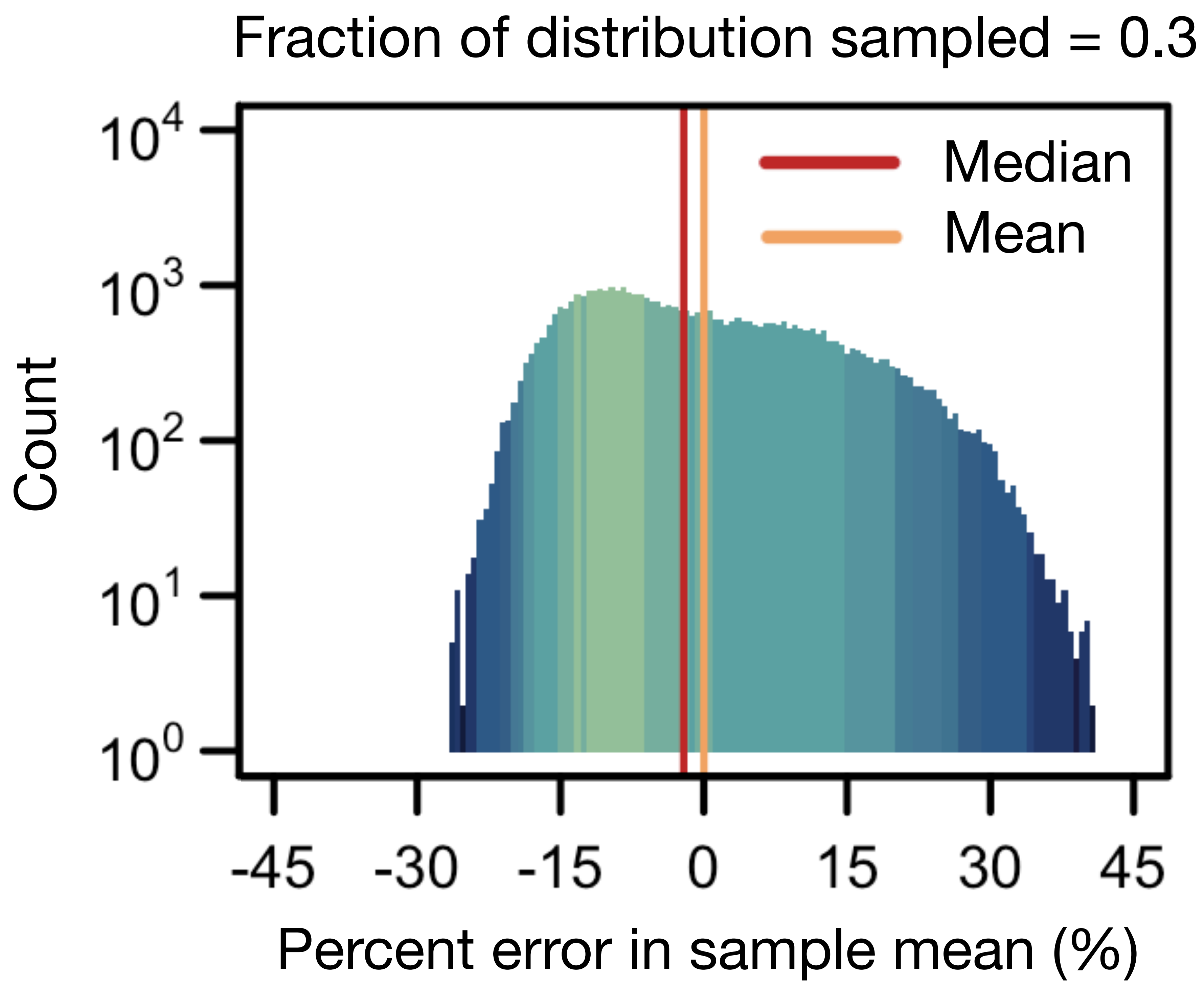




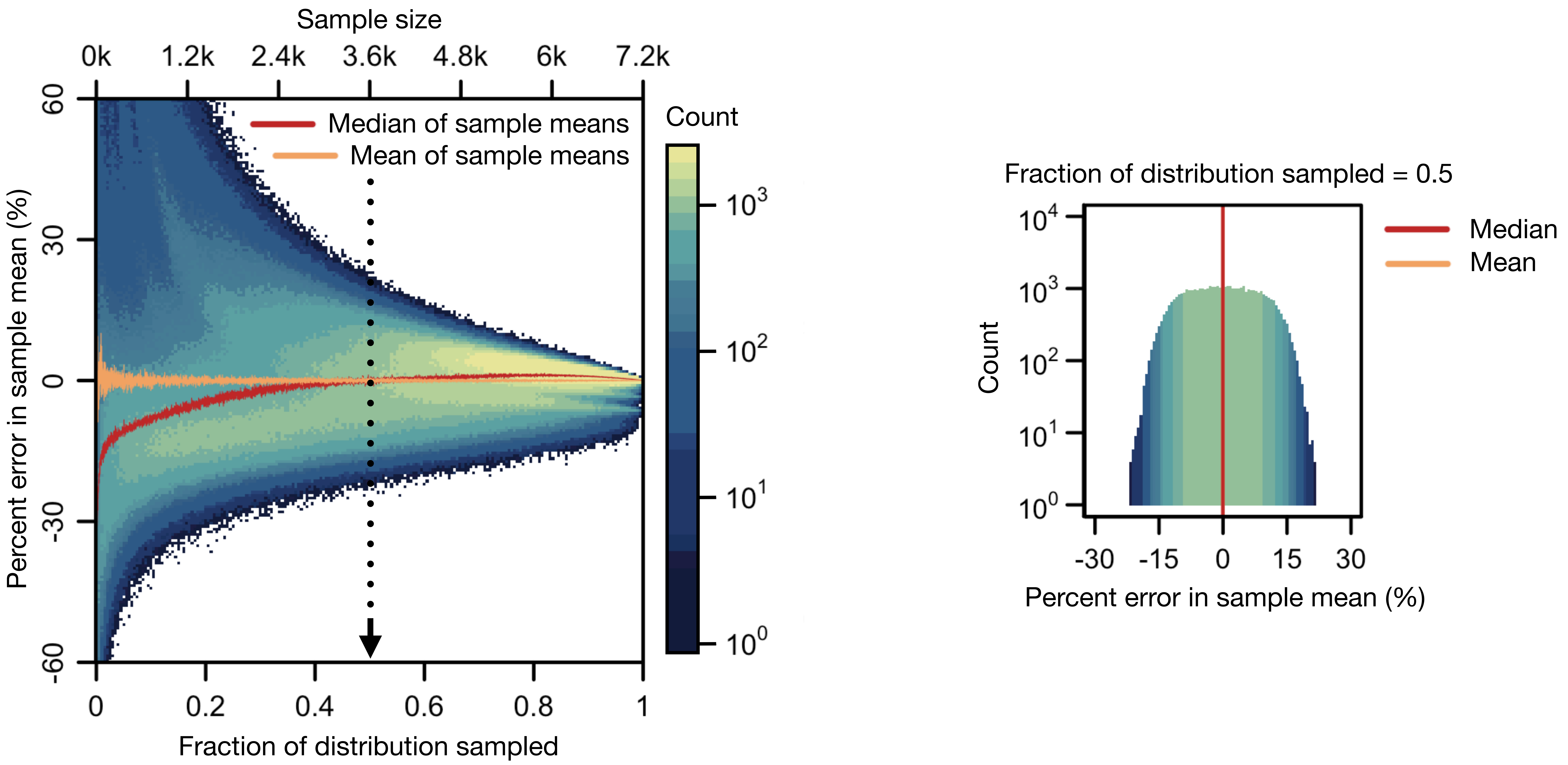
At small sample sizes, most samples underestimate, but some vastly overestimate!

Extreme skew can cause large sampling variability

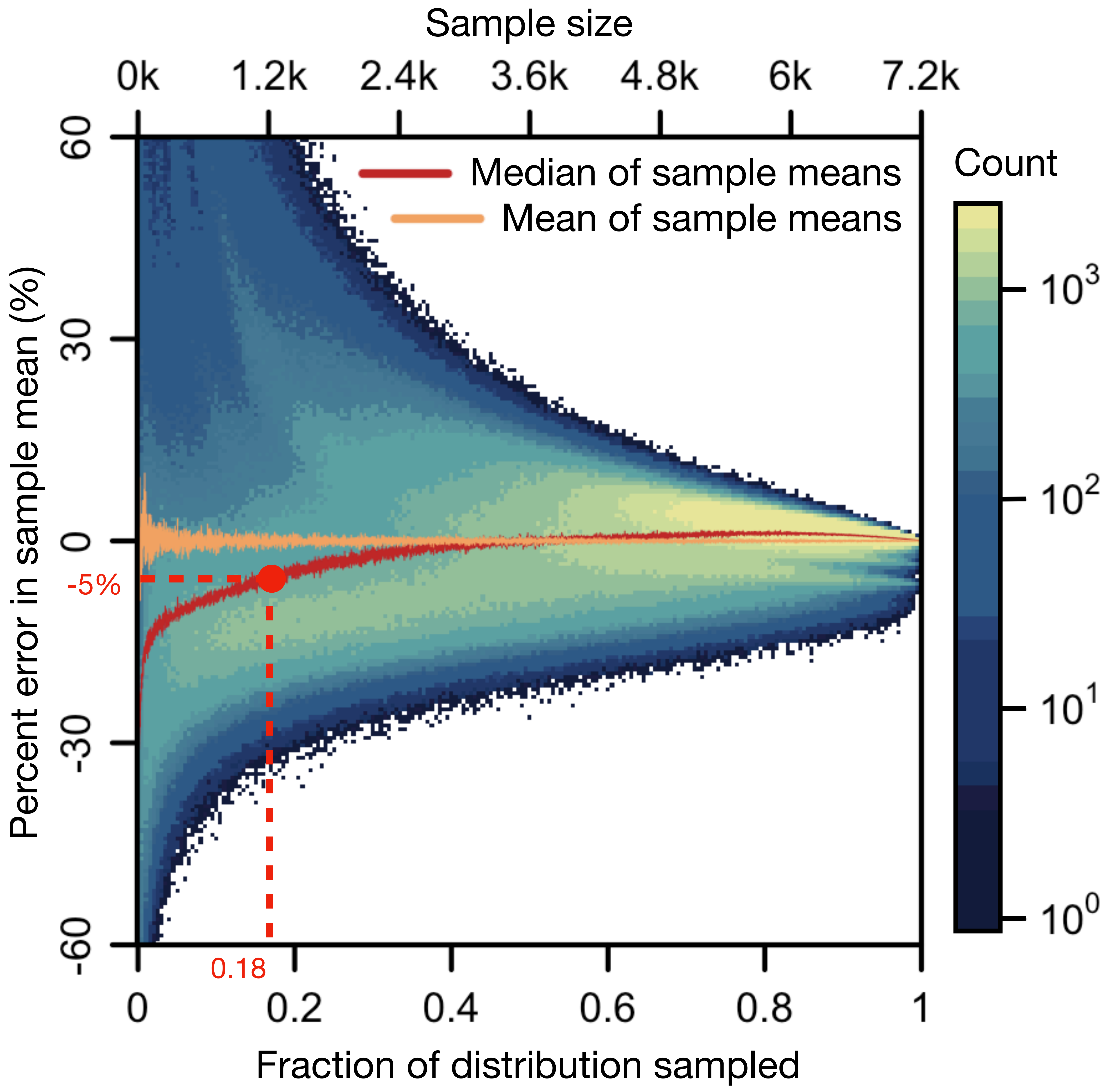




Extreme skew can cause large sampling variability



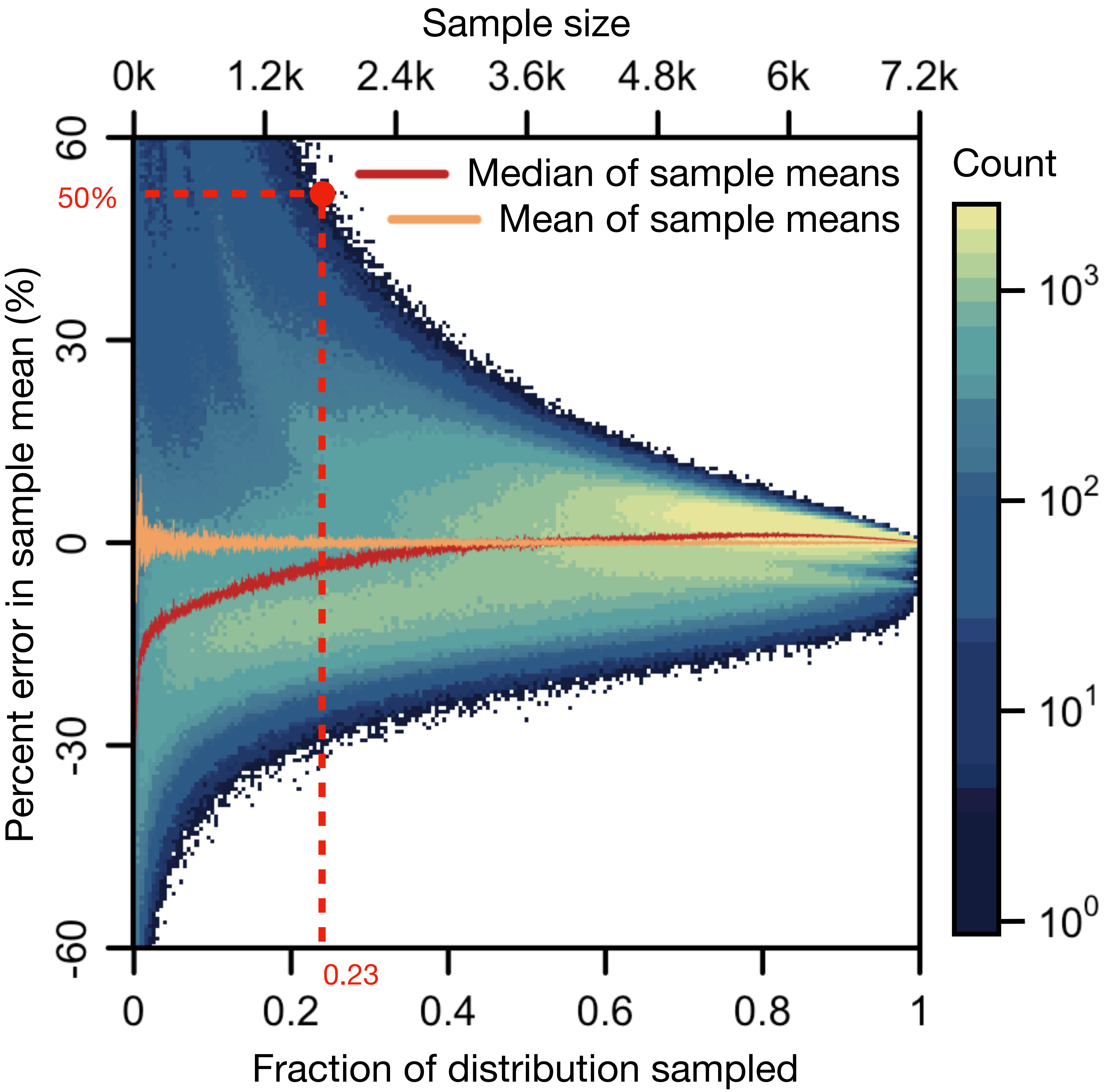
How to generate sample size guidance based on error metrics



If I want: $Q_{50} = \text{median}(\epsilon) > -5\%$

Then I need to sample at least: 18% of sites

How to generate sample size guidance based on error metrics



If I want: $\epsilon_{\text{max}} = \max(\epsilon) < 50\%$

Then I need to sample at least: 23% of sites

Sample size guidance based on maximum error

Sample size guidance based on maximum error

Very large samples (>50% of the basin) may be required to accurately characterize average emissions

Sample size guidance based on maximum error

Very large samples (>50% of the basin) may be required to accurately characterize average emissions **if super-emitters are both large and rare.**

Sample size guidance based on maximum error

For mitigation: more measurements to basins with many super-emitters.

For inventory development: more measurements to basins with few super-emitters.

Sample size guidance based on maximum error

More work is needed to definitively characterize the distribution of methane emission rates across sites in US oil and gas basins.

Implications and take-aways

Very large samples (>50% of the basin) may be required to accurately characterize average emissions if super-emitters are both large and rare.

Differences in emission characteristics between basins necessitates basin-specific sampling strategies.

For mitigation: more measurements to basins with many super-emitters.
For inventory development: more measurements to basins with few super-emitters.

More work is needed to definitively characterize the distribution of methane emission rates across sites in US oil and gas basins.

This framework can be extended to temporal variability, but we need better estimates of emission rate distributions over time to do this well. Depending on the scale you care about, this would help determine the number of basin-level aerial measurement campaigns to do per year or the number of continuous point sensors to put around a given facility.

Thank you!
Questions?

wdanie16@jh.edu

JOHNS HOPKINS UNIVERSITY

COLORADO SCHOOL OF MINES

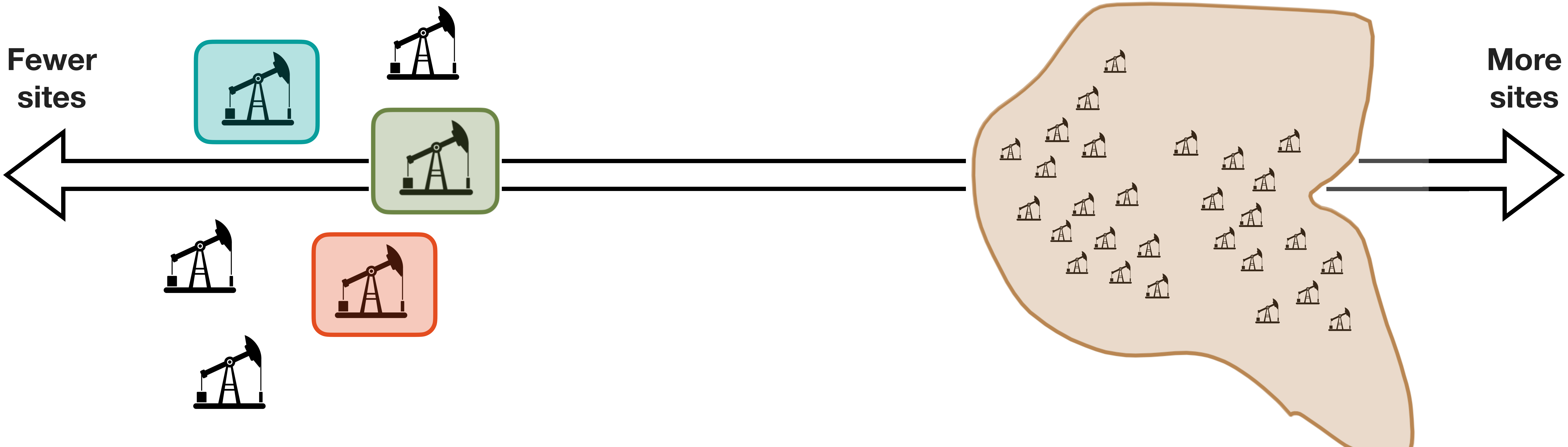
EEMDL Energy Emissions Modeling and Data Lab

CAN CH4

Daniels and Hammerling (2026). In revision, Communications Earth and Environment.

Webtool

What is temporal variability?



Revisit #1: **3** + **0** + **1.5** = 4.5

Revisit #2: **2** + **0.5** + **1.2** = 3.7

Revisit #3: **2.3** + **0** + **74** = 76.3

⋮

Temporal variability:
17x difference in the sum between revisits

Revisit #1: 3.2 + … + 1.5 = **x**

Revisit #2: 2 + … + 203.5 = **y**

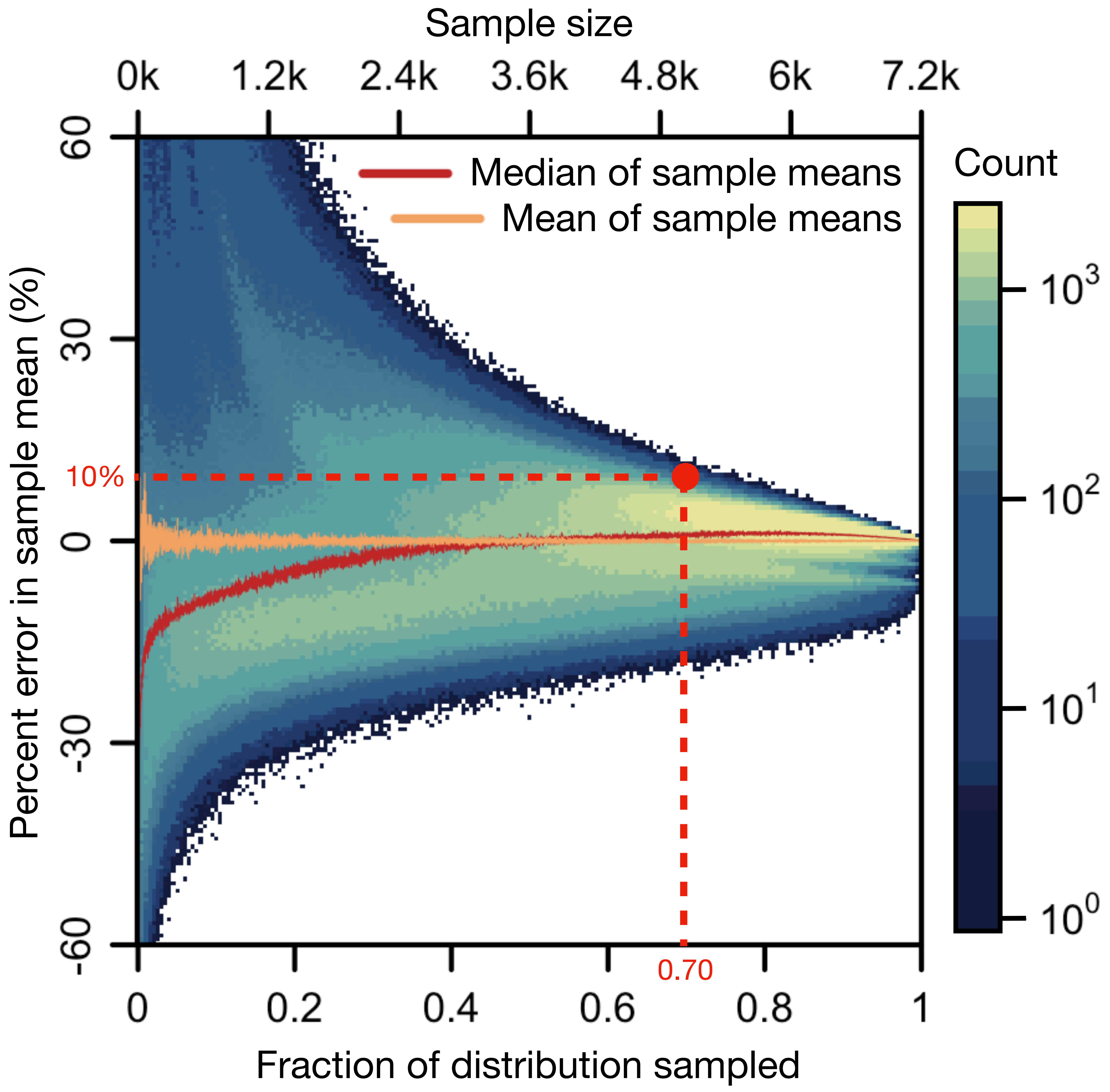
Revisit #3: 2.5 + … + 1.7 = **z**

⋮

Temporal variability:
??x difference in the sum between revisits

(but definitely smaller than 17x!)

How to generate sample size guidance based on error metrics



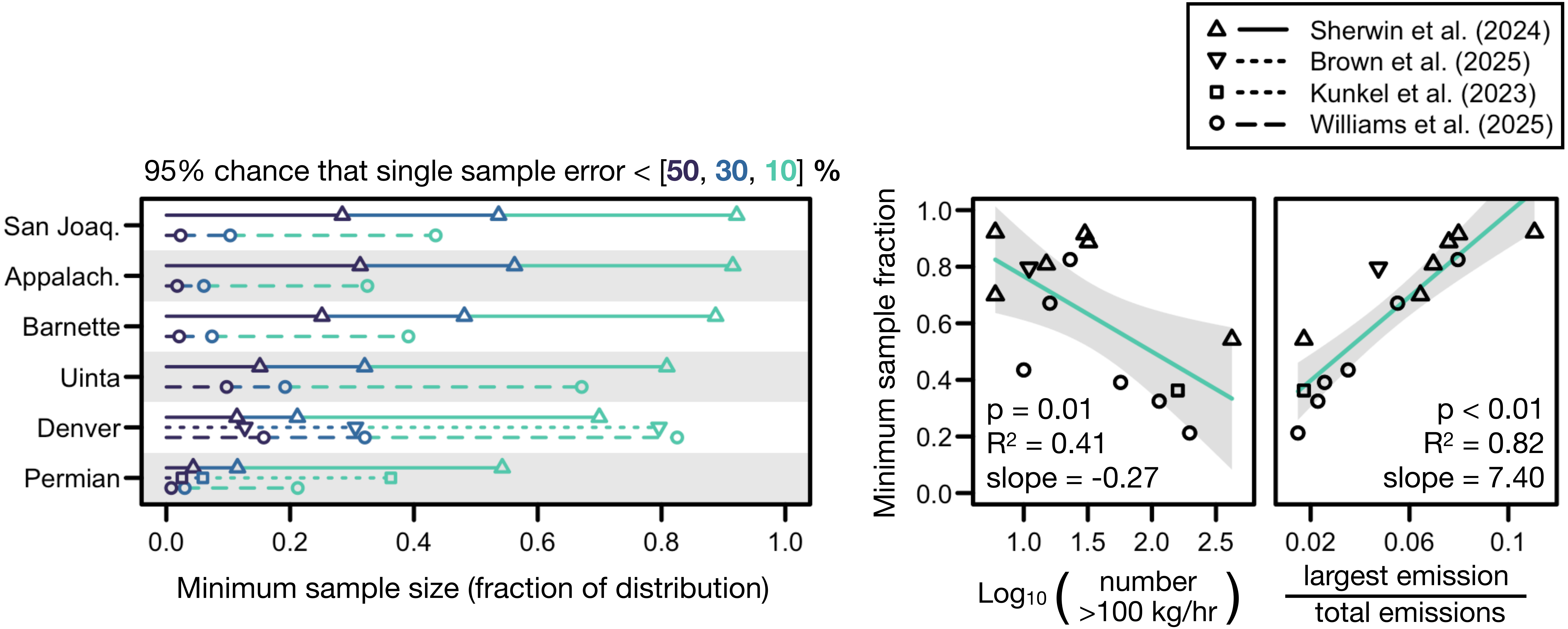
"I want to be 95% sure that the next sample I take has error less than 10%"

If I want: $\mathbb{P}_{10} = \dfrac{\text{number of } |\boldsymbol{\epsilon}| < 10}{\text{length of } \boldsymbol{\epsilon}} < 95\%$

Then I need to sample at least: 70% of sites

Sample size guidance based on median error

Sample size guidance based on single sample error



Median error > [-5, -3, -1] %
Denver
San Joaq.
Appalach.
Uinta
Barnette
Permian
Sherwin et al. (2024)
Brown et al. (2025)
Kunkel et al. (2023)
Williams et al. (2025)
slope = -0.03
R² = 0.15
p = 0.17
slope = -1.0e-5
R² = 0.05
p = 0.43
Maximum error < [100, 50, 25] %
slope = 0.01
R² = 0.01
p = 0.81
slope = 2.4e-5
R² = 0.06
p = 0.38
95% chance that single sample error < [50, 30, 10] %
slope = 0.01
R² < 0.01
p = 0.85
slope = 2.2e-5
R² = 0.04
p = 0.52
Minimum sample size (fraction of distribution)
Mean
SD

Median error > [-5, -3, -1] %
Denver
San Joaq.
Appalach.
Uinta
Barnette
Permian
Sherwin et al. (2024)
Brown et al. (2025)
Kunkel et al. (2023)
Williams et al. (2025)
slope = -0.04
R² = 0.01
p = 0.72
slope = -0.03
R² = 0.03
p = 0.57
Maximum error < [100, 50, 25] %
slope = -0.57
R² = 0.45
p = 0.01
slope = -0.29
R² = 0.52
p < 0.01
95% chance that single sample error < [50, 30, 10] %
slope = -0.70
R² = 0.45
p = 0.01
slope = -0.36
R² = 0.53
p < 0.01
Minimum sample size (fraction of distribution)
Log10(skew)
Log10(kurtosis)