

A statistical framework for reconstructing epidemic curves and modelling wastewater



Institute for
Pandemics



Emily Somerset
PhD student
University of Toronto

Wastewater-based reproduction rates for epidemic curve reconstruction. *Submitted work. Joint with:*

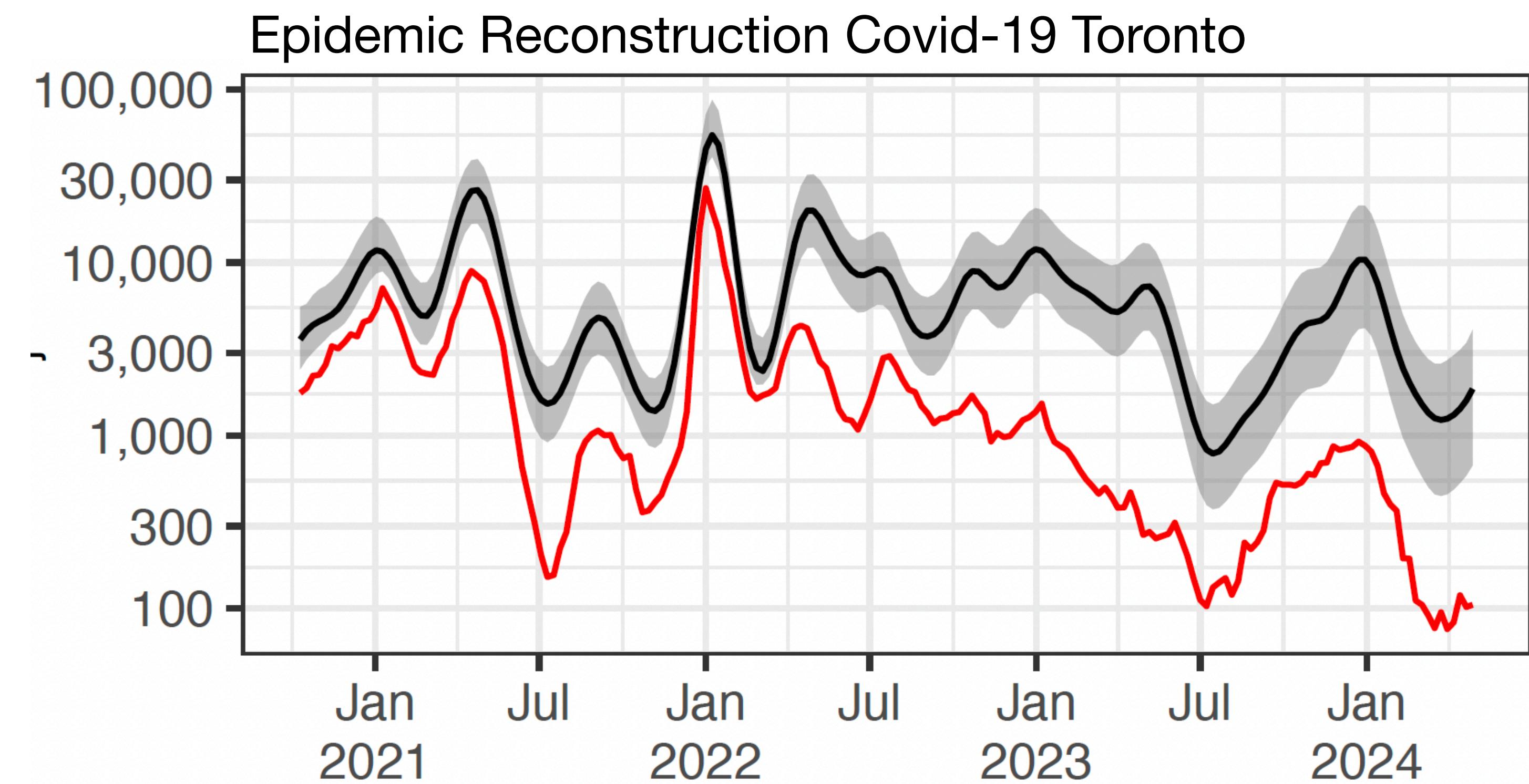


Justin Slater
Assistant Professor
University of Guelph



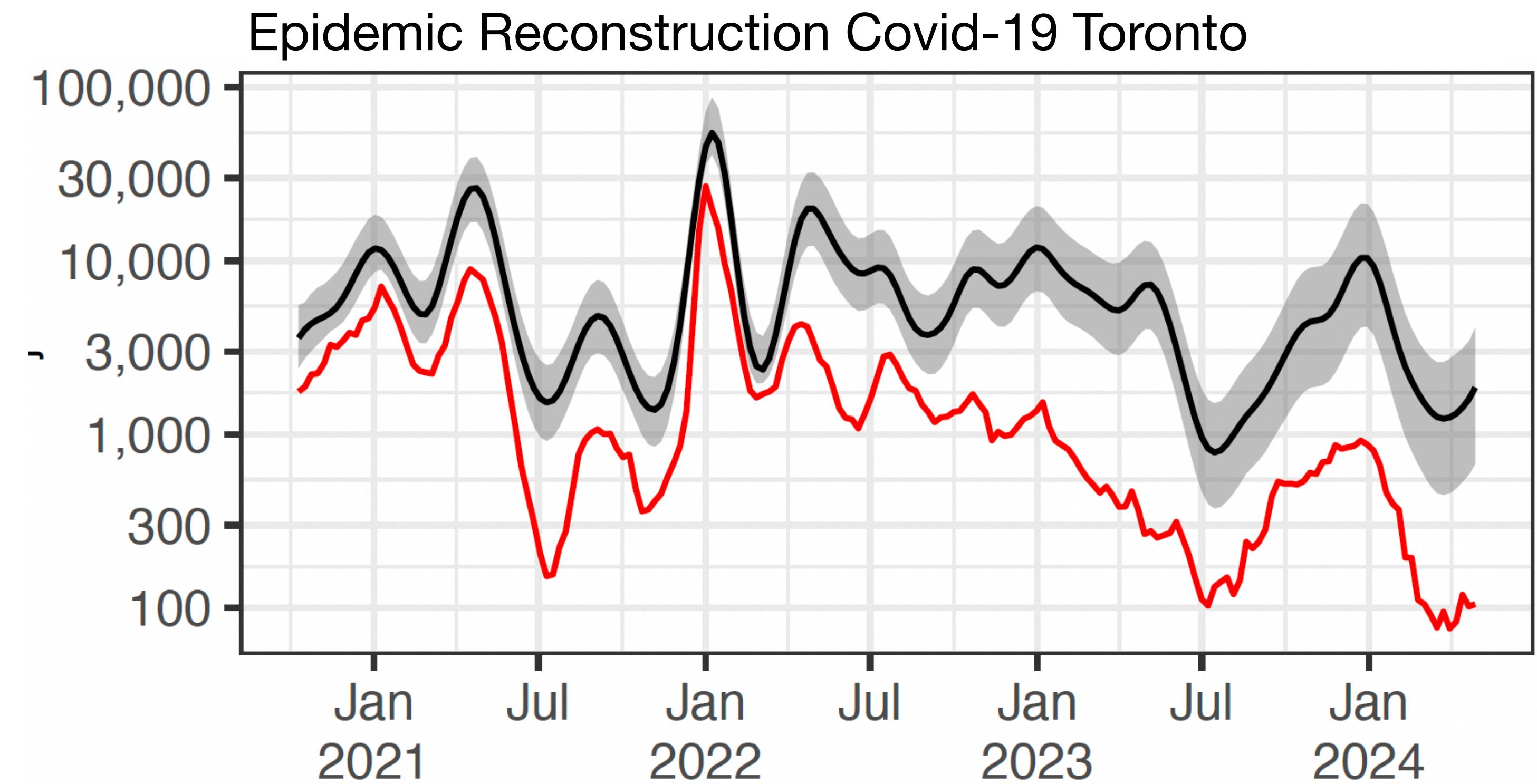
Patrick Brown
Associate Professor
University of Toronto

Goal: Estimate the number of people with an infectious disease.



Goal: Estimate the number of people with an infectious disease.

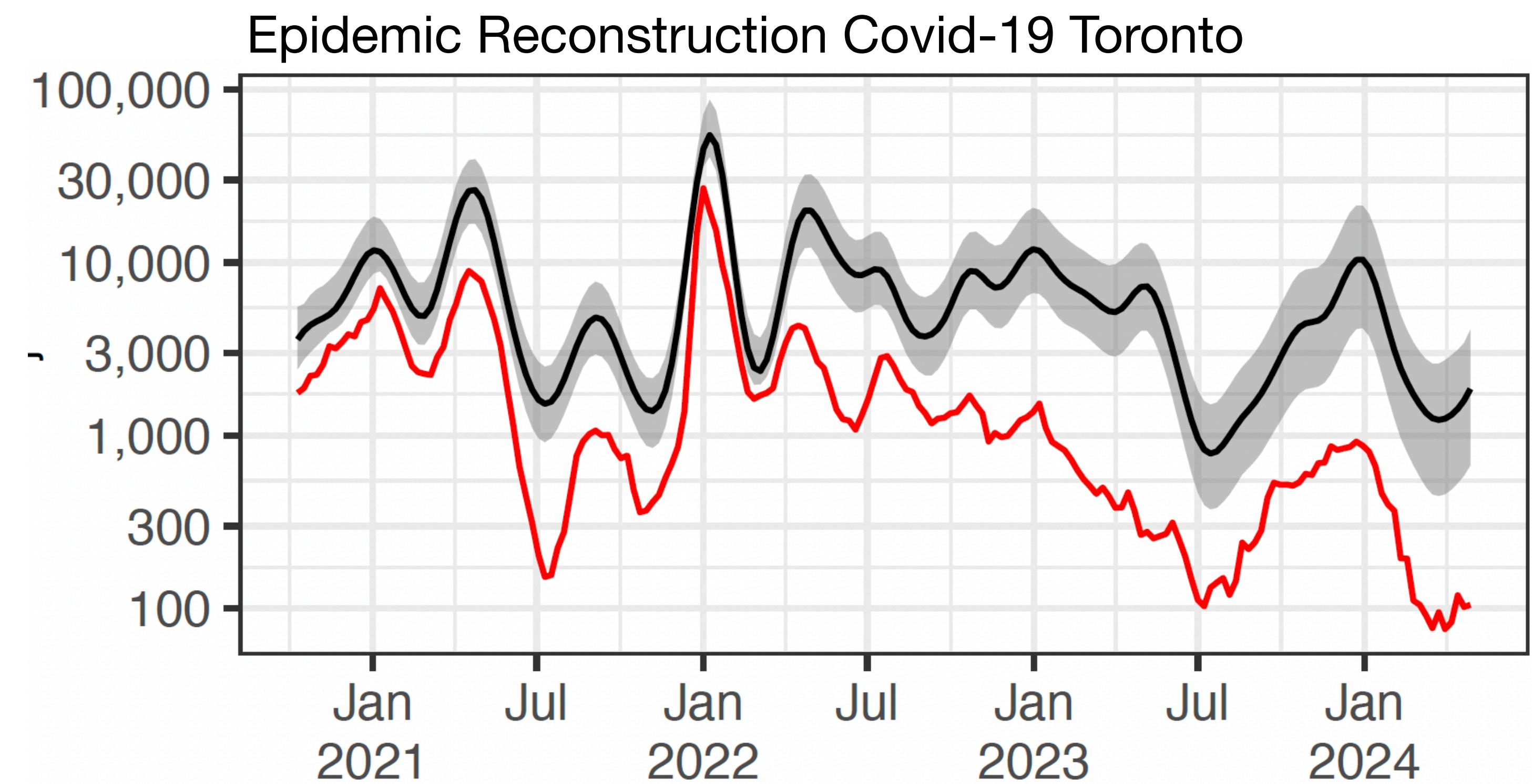
Y_t = observed cases at t



Goal: Estimate the number of people with an infectious disease.

Y_t = observed cases at t

I_t = true number of infections



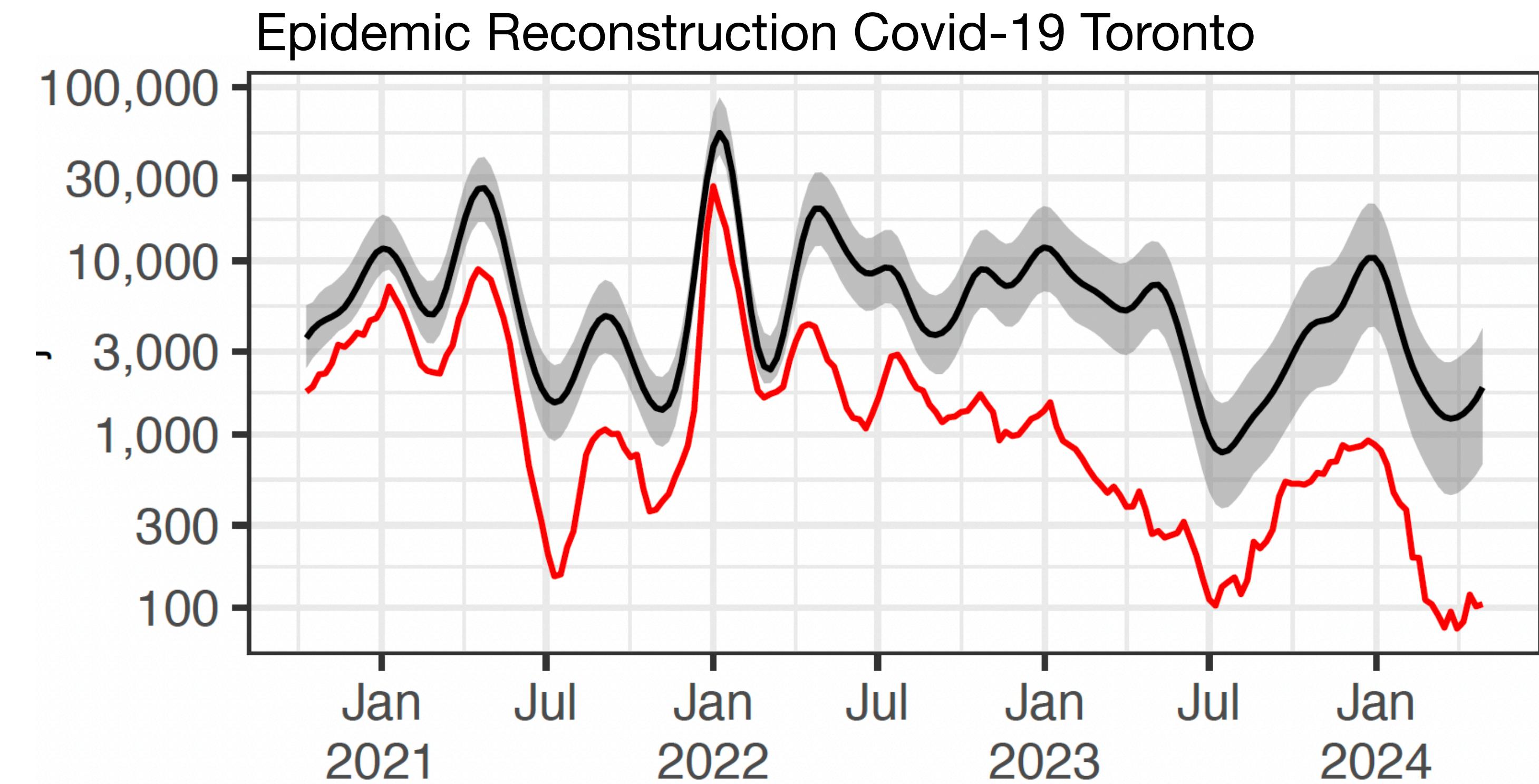
Goal: Estimate the number of people with an infectious disease.

Y_t = observed cases at t

I_t = true number of infections

θ = parameter vector with real-world meanings, e.g

- Reproduction rates
- Reporting fractions



Want an infectious disease model like:

$$p(I_t, \theta | Y_t, \text{other data})$$

Commonly used infectious disease model [1]: Poisson autoregressive model

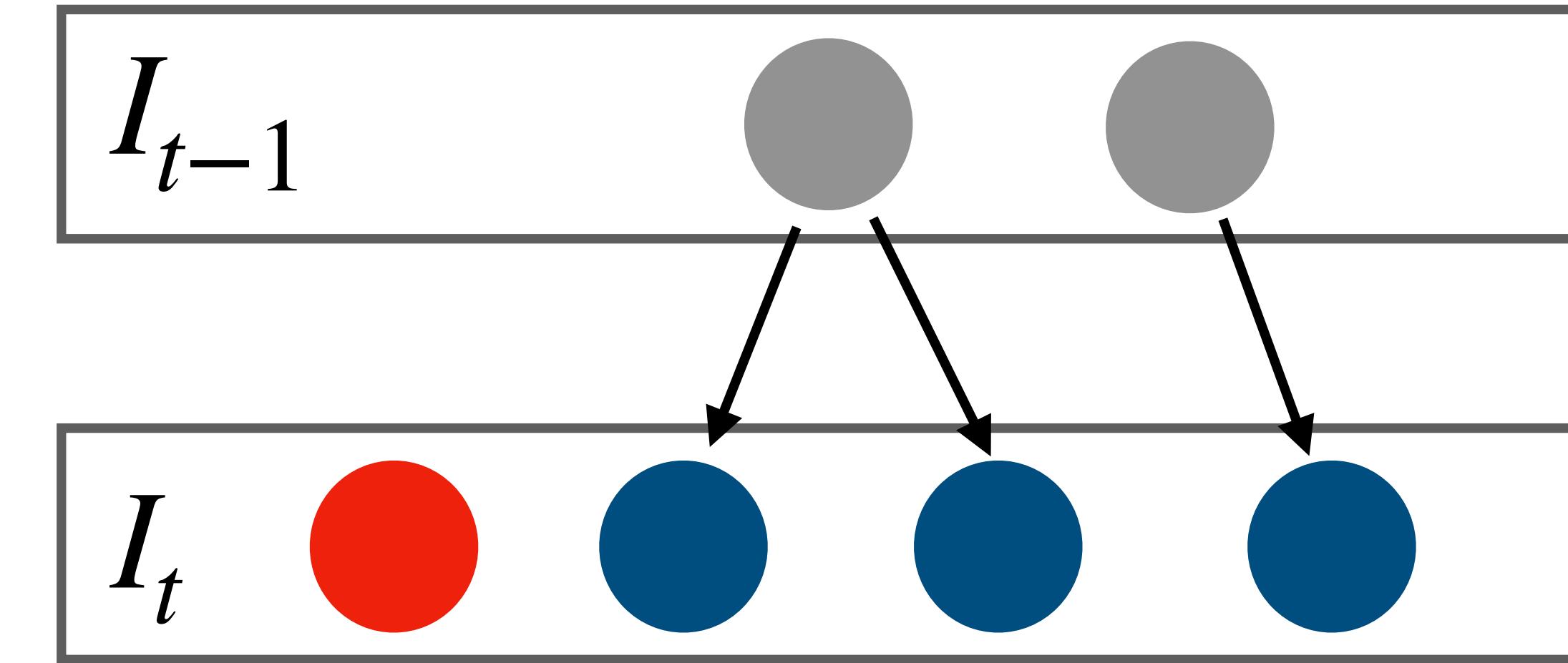
$$I_t | I_{t-1} \sim \text{Pois}(\lambda_t)$$

$$\lambda_t = \nu + \alpha I_{t-1}$$

Commonly used infectious disease model [1]: Poisson autoregressive model

$$I_t | I_{t-1} \sim \text{Pois}(\lambda_t)$$

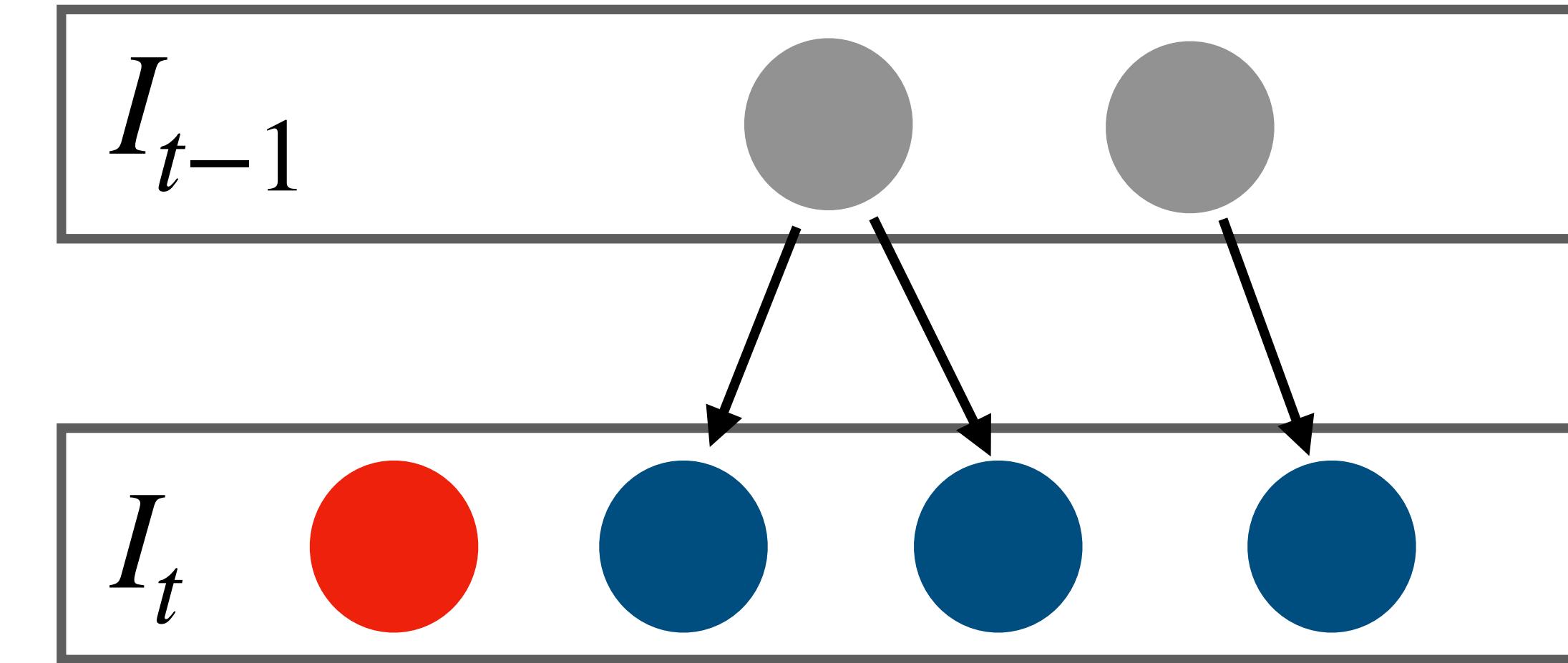
$$\lambda_t = \nu + \alpha I_{t-1}$$



Commonly used infectious disease model [1]: Poisson autoregressive model

$$I_t | I_{t-1} \sim \text{Pois}(\lambda_t)$$

$$\lambda_t = \nu + \underline{\alpha I_{t-1}}$$

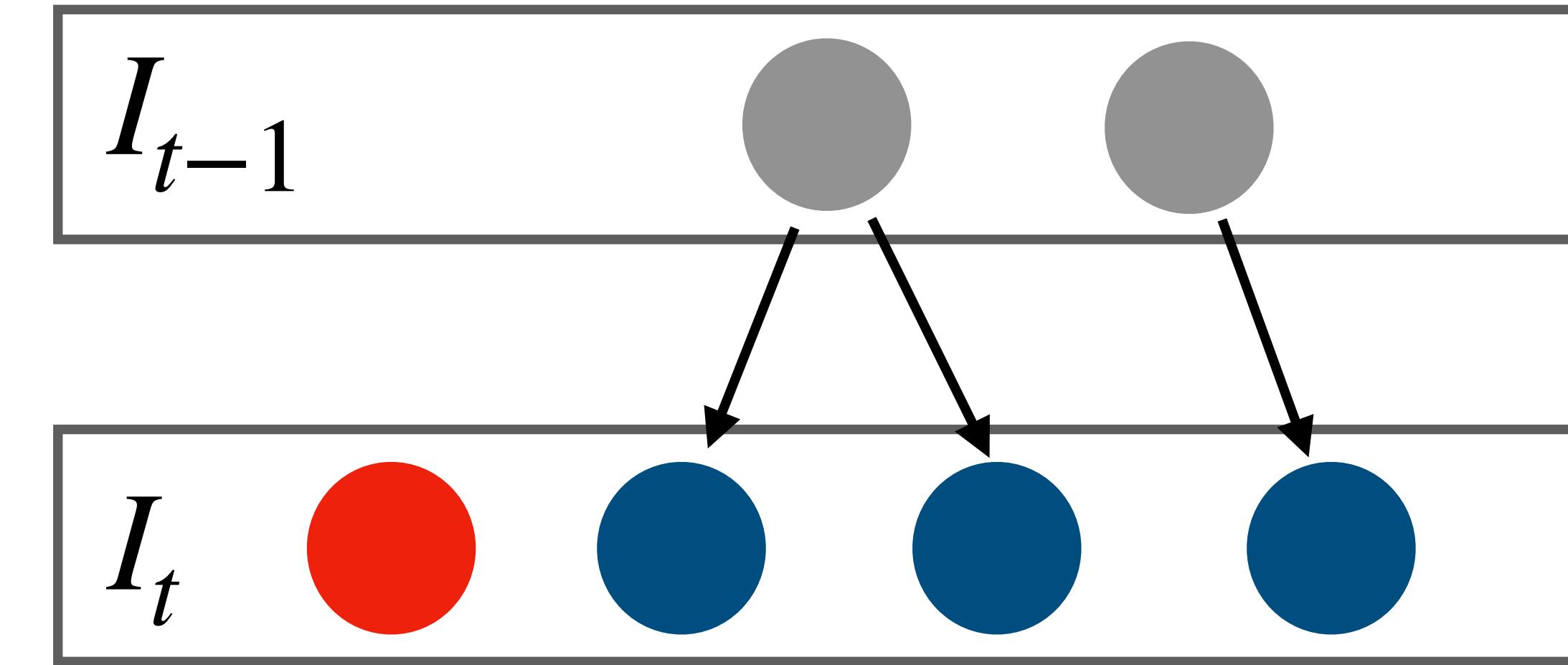


α is the **Expected number of offspring** - Reproduction rate

Commonly used infectious disease model [1]: Poisson autoregressive model

$$I_t | I_{t-1} \sim \text{Pois}(\lambda_t)$$

$$\lambda_t = \nu + \alpha I_{t-1}$$



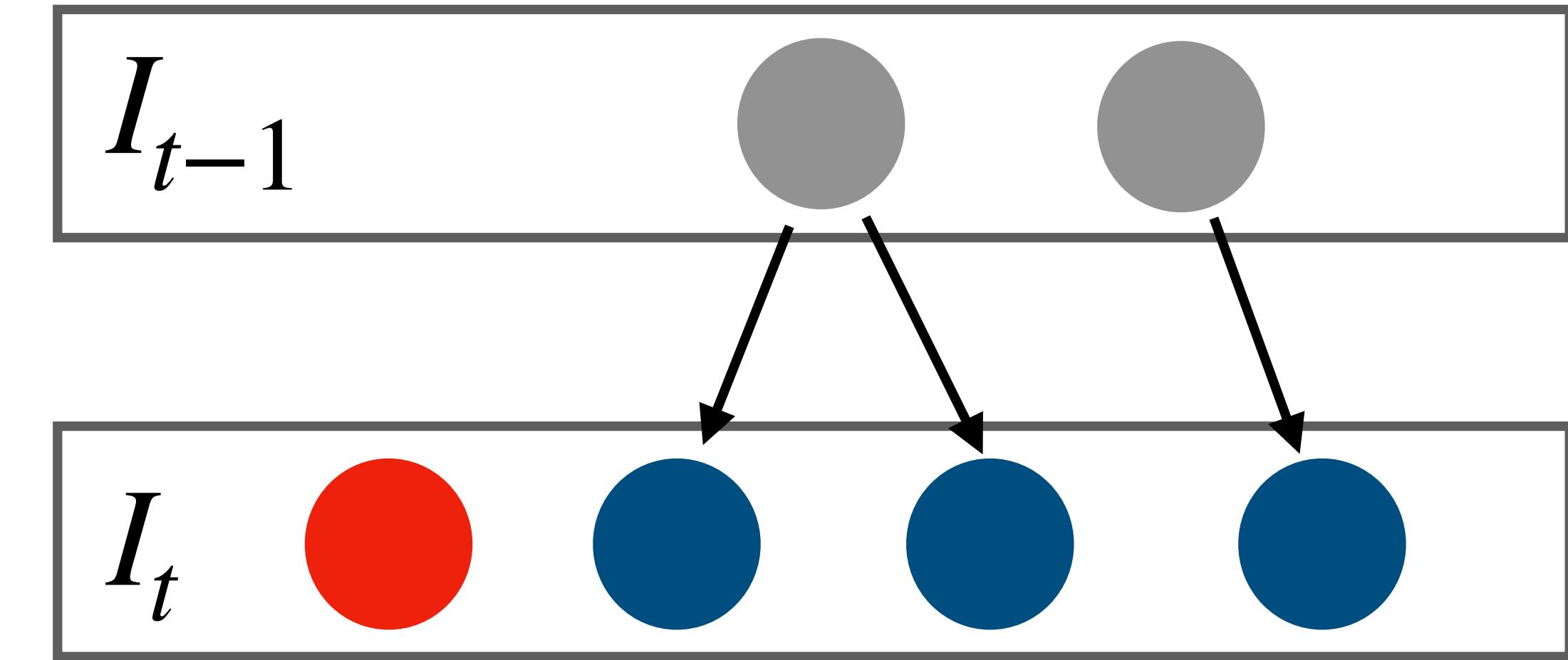
α is the **Expected number of offspring** - Reproduction rate

ν is the expected number of cases not explained by I_{t-1}

Commonly used infectious disease model [1]: Poisson autoregressive model

$$I_t | I_{t-1} \sim \text{Pois}(\lambda_t)$$

$$\lambda_t = \nu + \alpha I_{t-1}$$



α is the **Expected number of offspring** - Reproduction rate

ν is the expected number of cases not explained by I_{t-1}

Old cases are **the only cause** of new cases → mechanistic model

$$I_t | I_{t-1} \sim \text{Pois}(\lambda_t)$$

$$\lambda_t = \nu + \alpha I_{t-1}$$

$$I_t | I_{t-1} \sim \text{Pois}(\lambda_t)$$

$$\lambda_t = \nu + \alpha I_{t-1}$$

- We don't observe the true number of infectious I_t

$$I_t | I_{t-1} \sim \text{Pois}(\lambda_t)$$

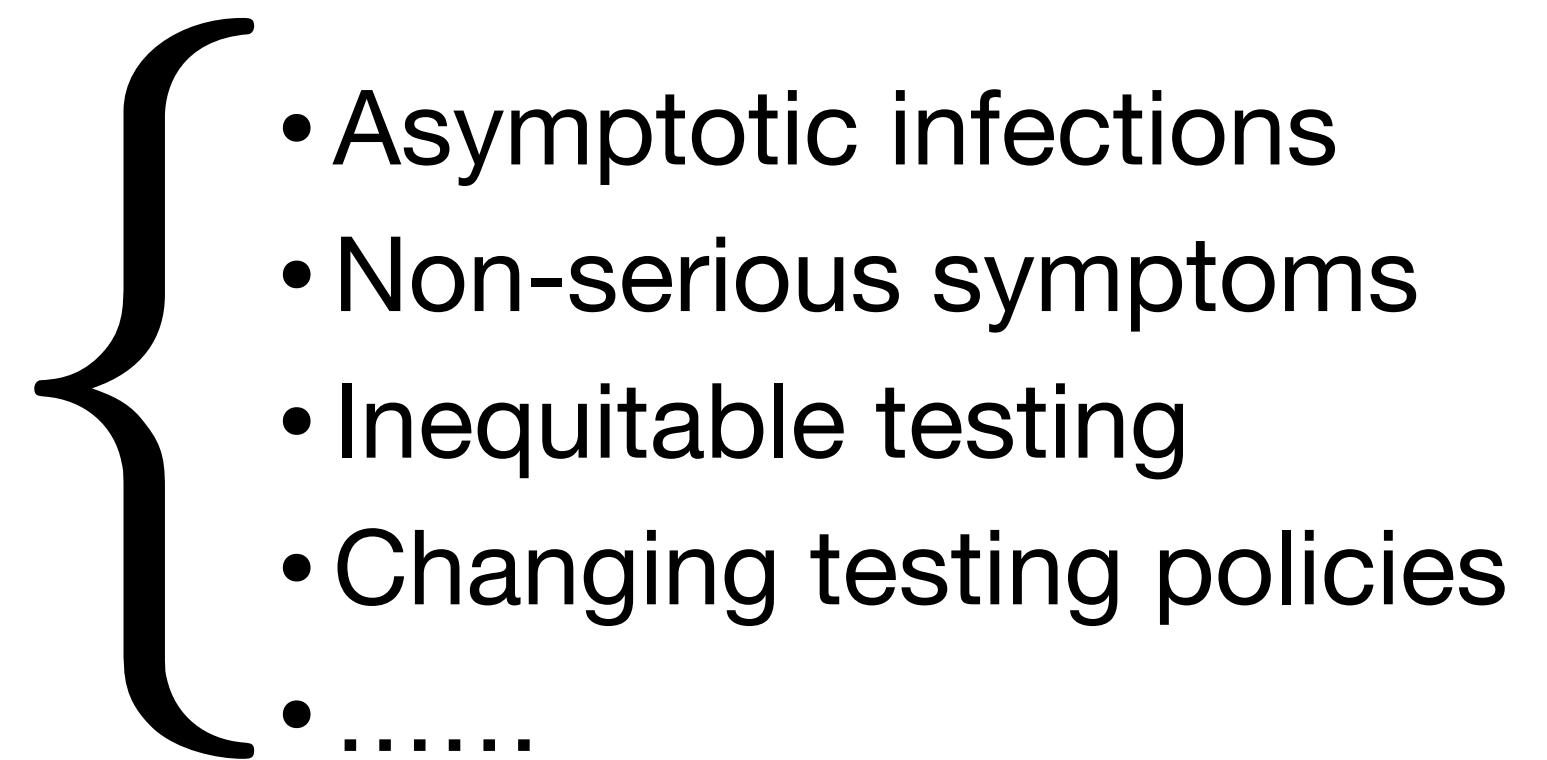
$$\lambda_t = \nu + \alpha I_{t-1}$$

- We don't observe the true number of infectious I_t
- We observe an **under-reported** infection count Y_t

$$I_t | I_{t-1} \sim \text{Pois}(\lambda_t)$$

$$\lambda_t = \nu + \alpha I_{t-1}$$

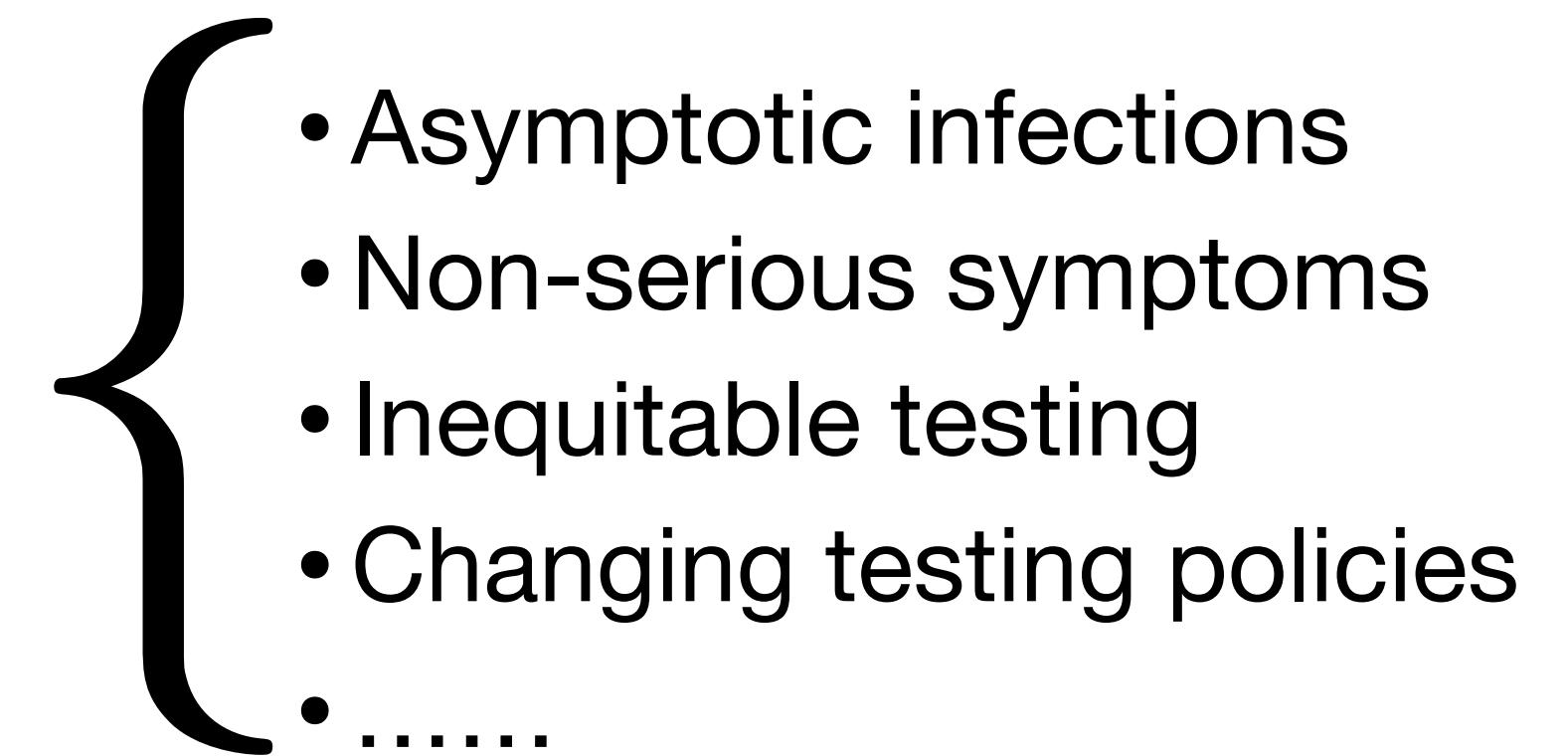
- We don't observe the true number of infectious I_t
- We observe an **under-reported** infection count Y_t



$$I_t | I_{t-1} \sim \text{Pois}(\lambda_t)$$

$$\lambda_t = \nu + \alpha I_{t-1}$$

- We don't observe the true number of infectious I_t
- We observe an **under-reported** infection count Y_t
- If we fit:



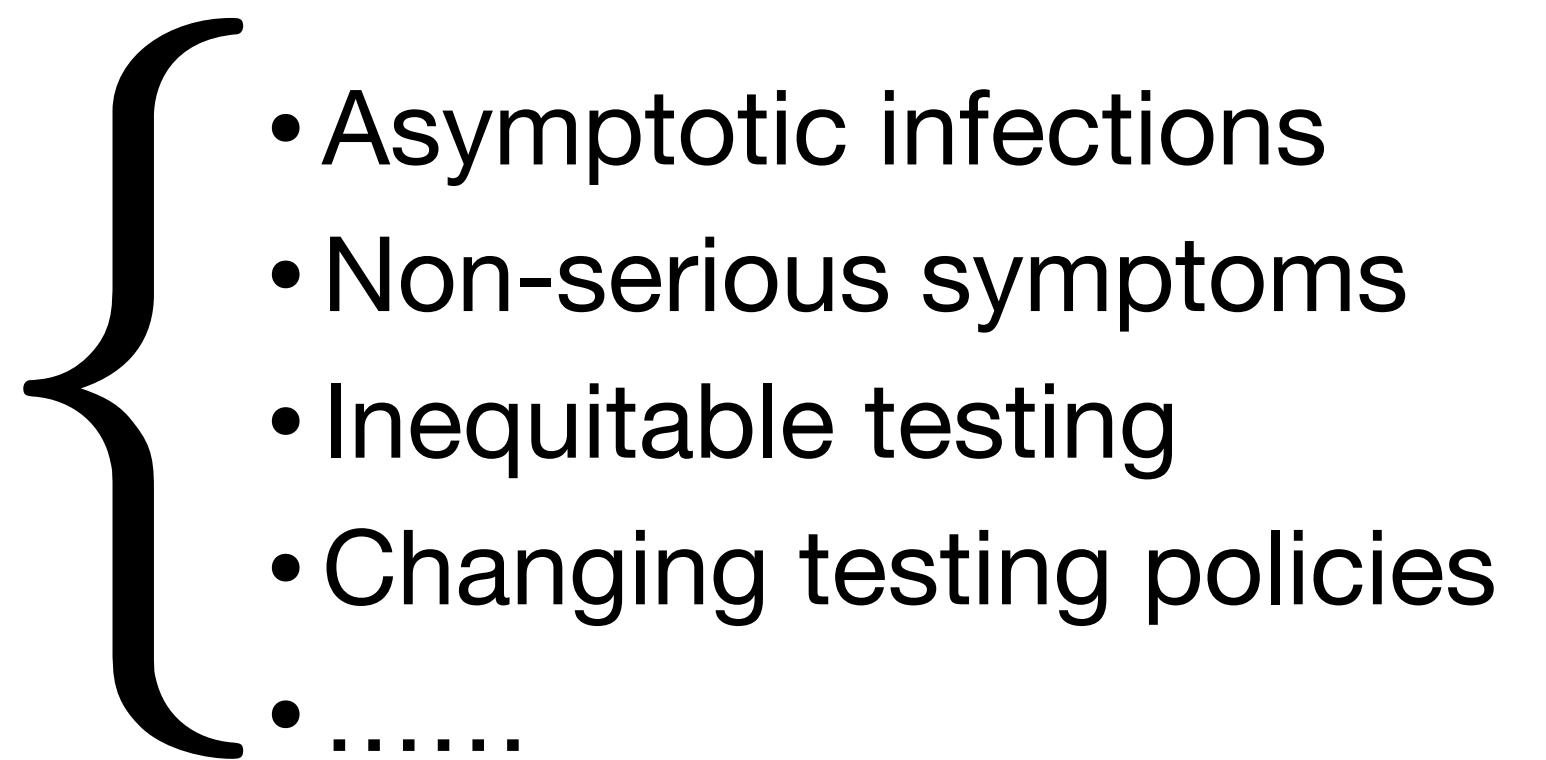
$$Y_t | Y_{t-1} \sim \text{Pois}(\lambda_t)$$

$$\lambda_t = \nu + \alpha Y_{t-1}$$

$$I_t | I_{t-1} \sim \text{Pois}(\lambda_t)$$

$$\lambda_t = \nu + \alpha I_{t-1}$$

- We don't observe the true number of infectious I_t
- We observe an **under-reported** infection count Y_t
- If we fit:



$$Y_t | Y_{t-1} \sim \text{Pois}(\lambda_t)$$

$$\lambda_t = \nu + \alpha Y_{t-1}$$

- Biased inference of ν and α

Model the under-reporting mechanism!

$$Y_t | I_t \sim \text{Bin}(I_t, \pi)$$

**Binomial
thinning**

$$I_t | I_{t-1} \sim \text{Pois}(\lambda_t)$$

$$\lambda_t = \nu + \alpha I_{t-1}$$

π = Reporting fraction

I_t = true infection counts at time t

Integer-valued state-space model

In practice, we also want time-varying reporting probability and reproduction rates:

$$Y_t | I_t \sim \text{Bin}(I_t, \pi_t)$$

$$I_t | I_{t-1} \sim \text{Pois}(\lambda_t, \lambda_t)$$

$$\lambda_t = \nu + \alpha_t I_{t-1}$$

$$\text{logit}(\pi_t) \sim \text{RW}(1)$$

π_t = Reporting probability at time t

α_t = reproduction rate at time t

In practice, we also want time-varying reporting probability and reproduction rates:

$$Y_t | I_t \sim \text{Bin}(I_t, \pi_t)$$

$$I_t | I_{t-1} \sim \text{Pois}(\lambda_t, \lambda_t)$$

$$\lambda_t = \nu + \alpha_t I_{t-1}$$

$$\text{logit}(\pi_t) \sim \text{RW}(1)$$

π_t = Reporting probability at time t

α_t = reproduction rate at time t

Leads to an unidentifiable model

In practice, we also want time-varying reporting probability and reproduction rates:

$$Y_t | I_t \sim \text{Bin}(I_t, \pi_t)$$

$$I_t | I_{t-1} \sim \text{Pois}(\lambda_t, \lambda_t)$$

$$\lambda_t = \nu + \alpha_t I_{t-1}$$

$$\text{logit}(\pi_t) \sim \text{RW}(1)$$

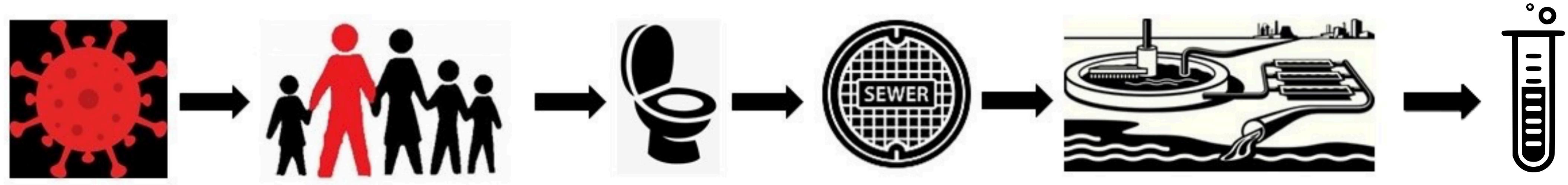
π_t = Reporting probability at time t

α_t = reproduction rate at time t

A solution: Integrate additional information from wastewater data!

Leads to an unidentifiable model

Wastewater Surveillance:



Credit: Government of Newfoundland Labrador Canada [2].

- Pathogens being monitored through wastewater [3]
 - SARS-CoV-2
 - Influenza
 - Respiratory syncytial virus
 - Human Metapneumovirus
 - Parainfluenza
 - Norovirus
 - Rotavirus
 - Monkey pox
 - Hepatitis A
 - Candida auris
 - EDV68
 - Human Adenovirus Group F



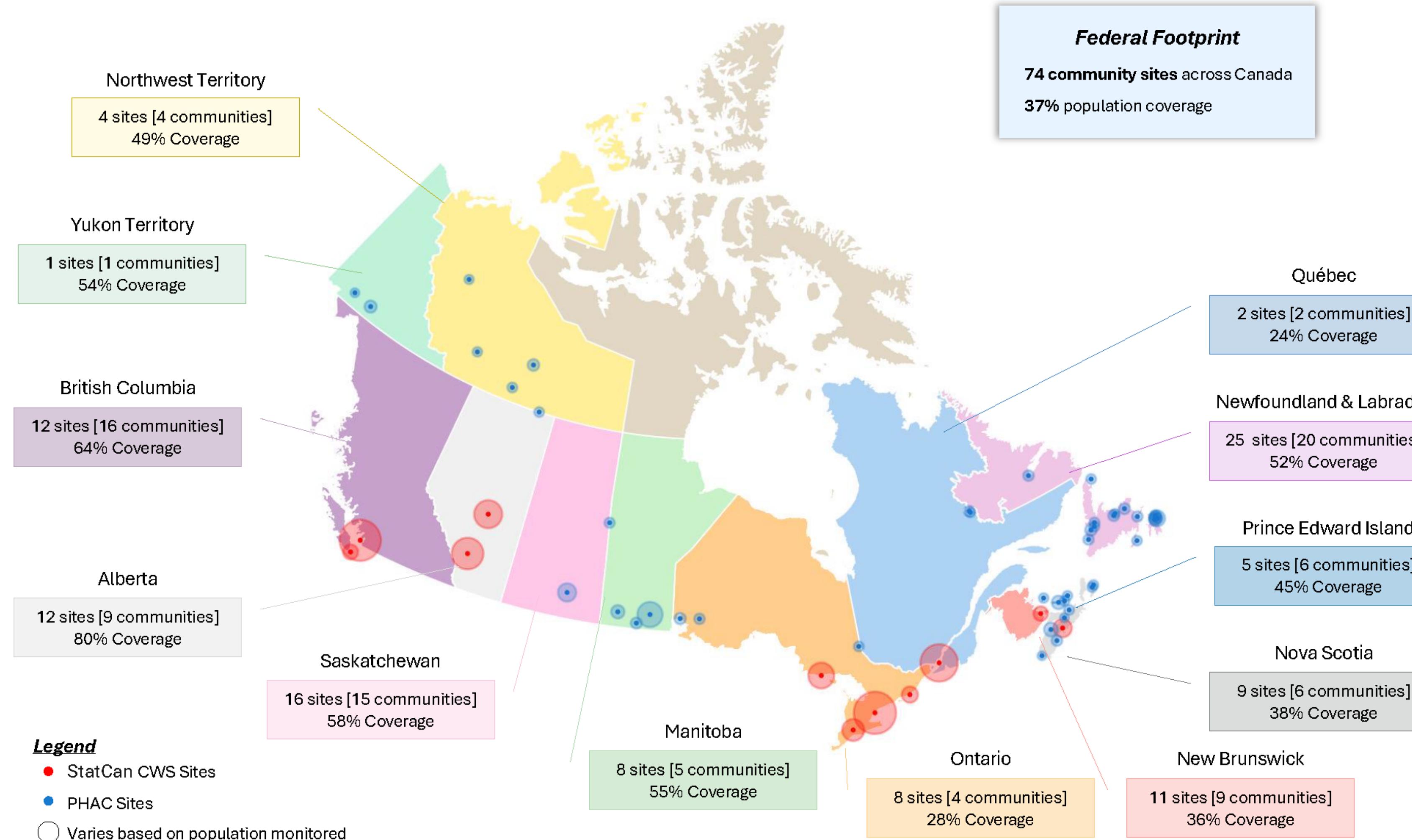
Public Health
Agency of Canada

Agence de la santé
publique du Canada

Canada

Current Federal, Provincial and Territorial Wastewater Monitoring Networks

Covers ~51% of the population connected to sewer systems across Canada

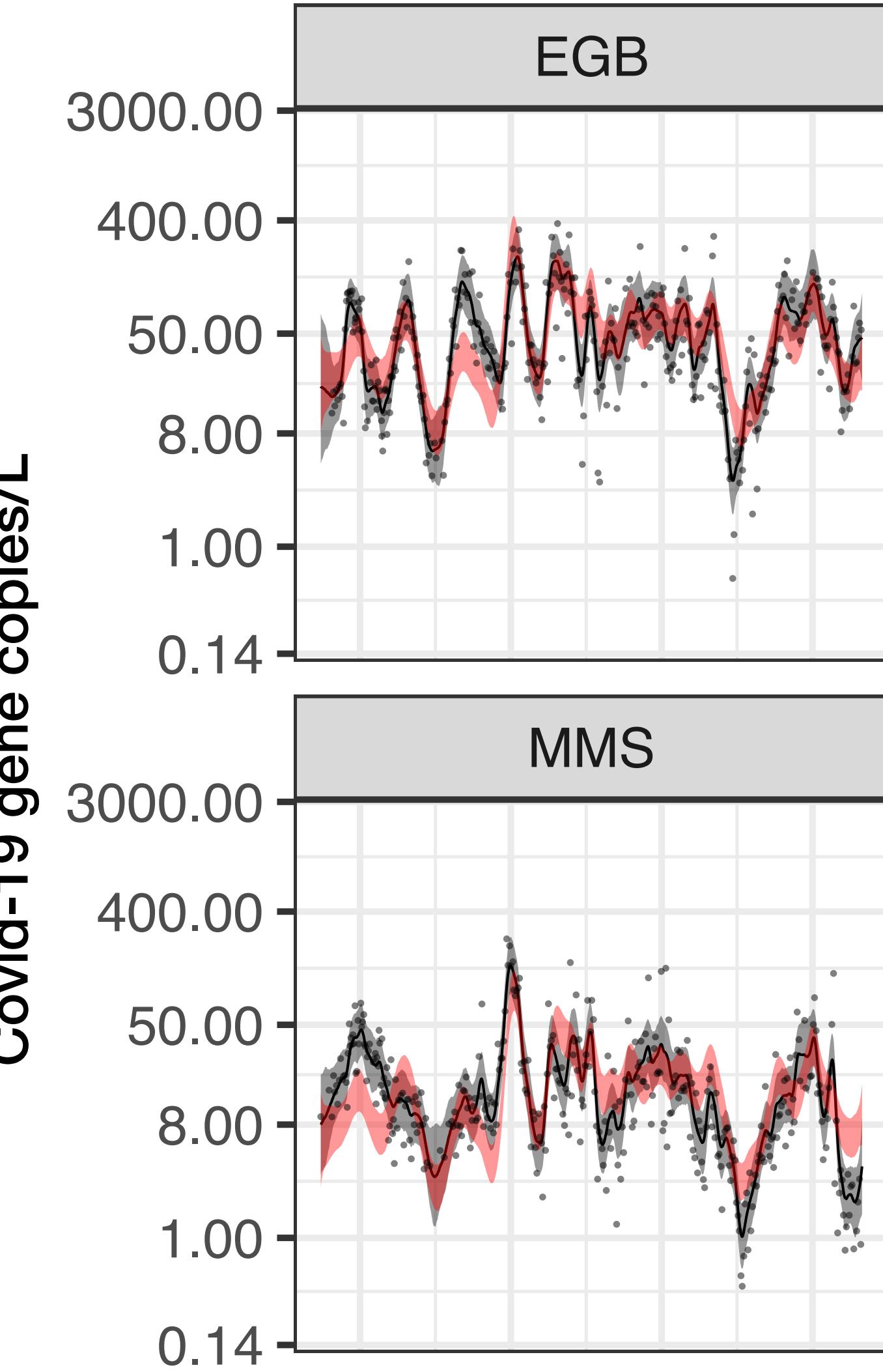
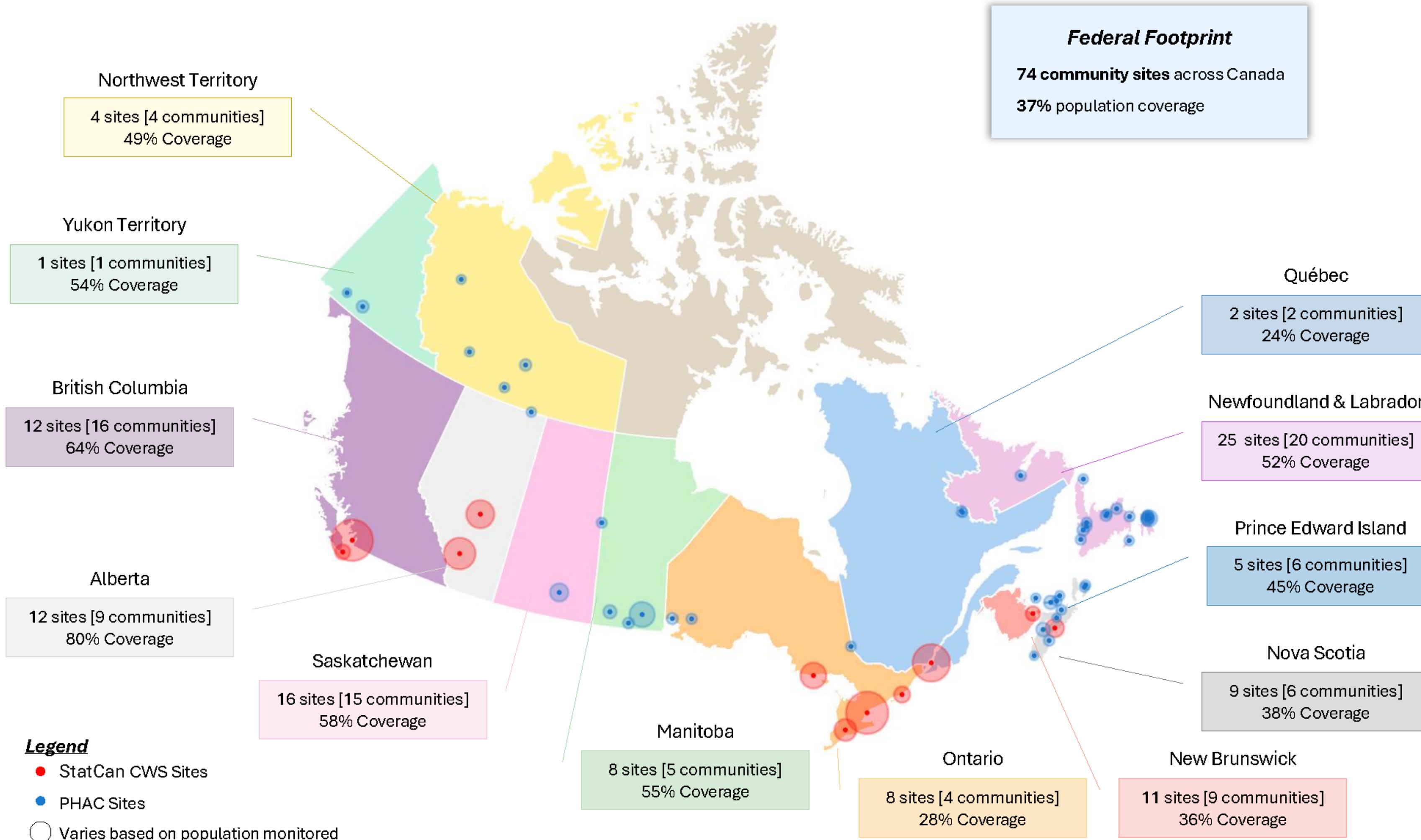




Canada

Current Federal, Provincial and Territorial Wastewater Monitoring Networks

Covers ~51% of the population connected to sewer systems across Canada

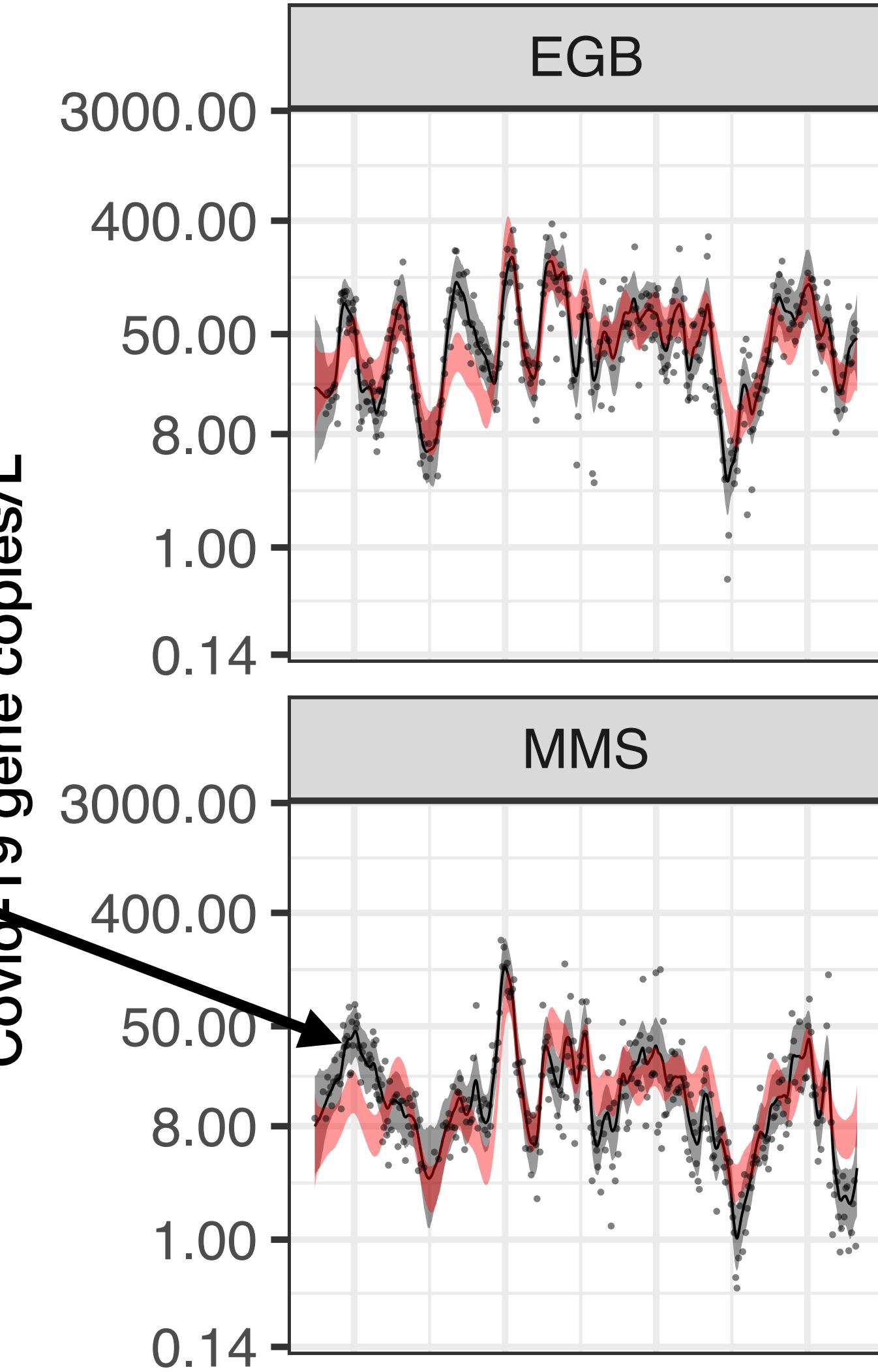




Canada

Current Federal, Provincial and Territorial Wastewater Monitoring Networks

Covers ~51% of the population connected to sewer systems across Canada

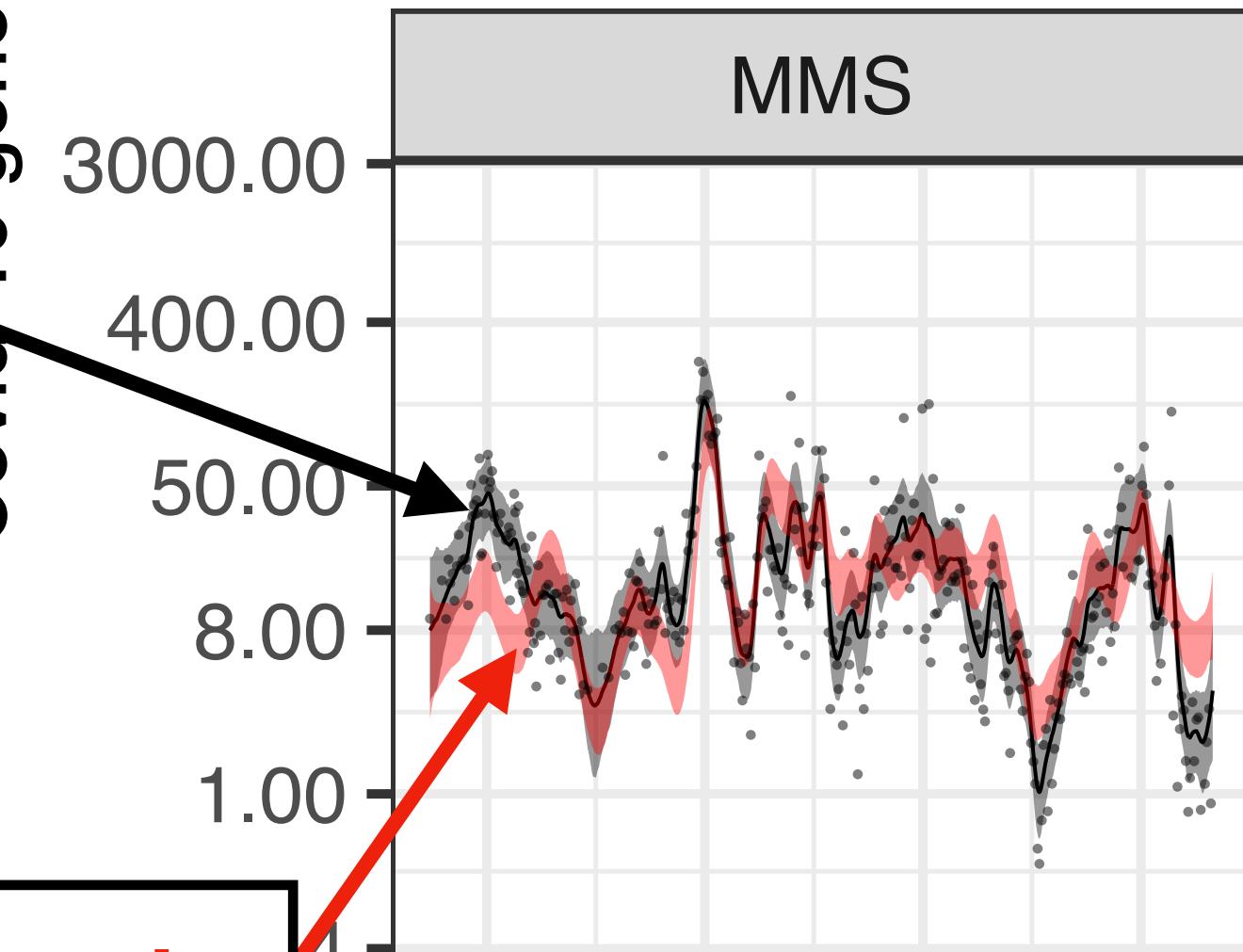
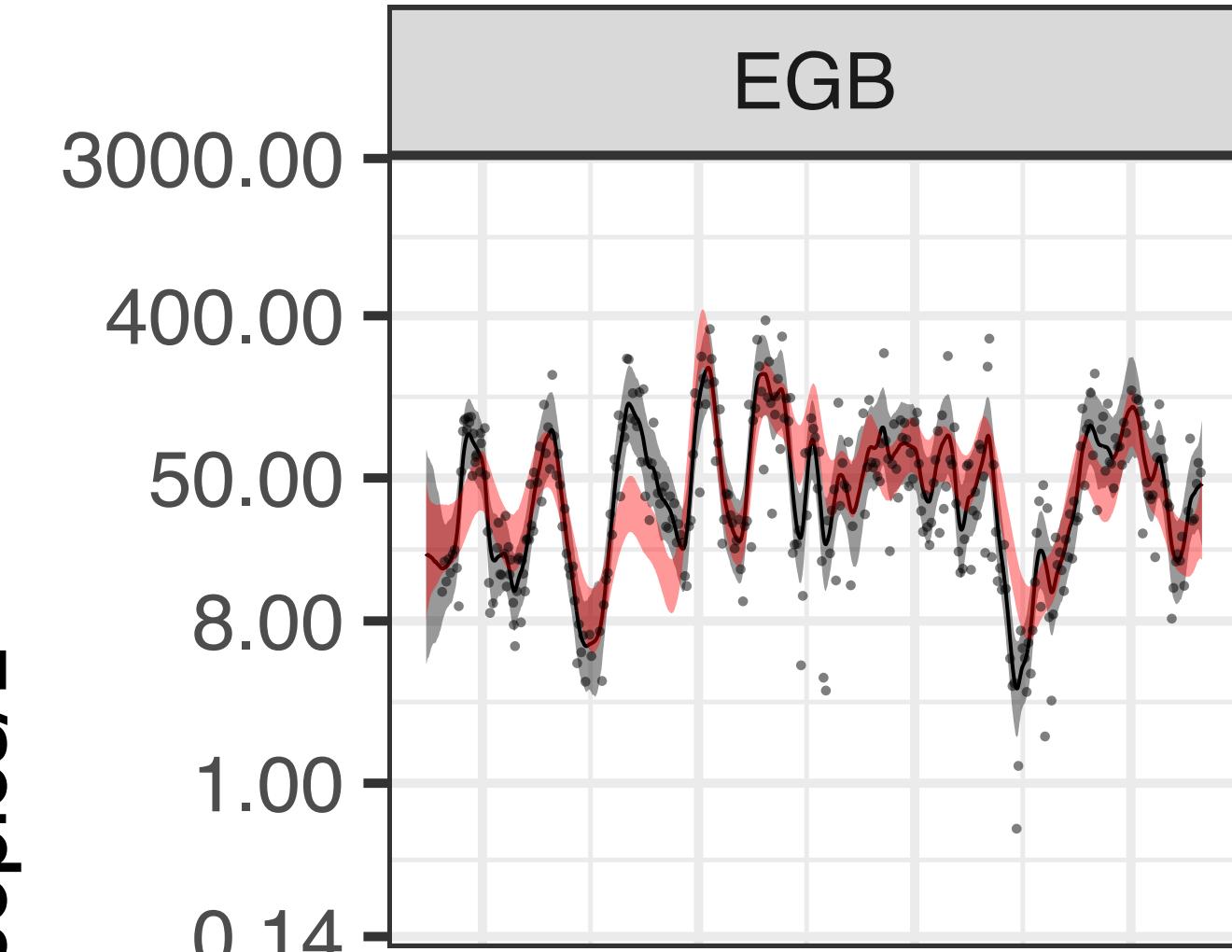




Canada

Current Federal, Provincial and Territorial Wastewater Monitoring Networks

Covers ~51% of the population connected to sewer systems across Canada

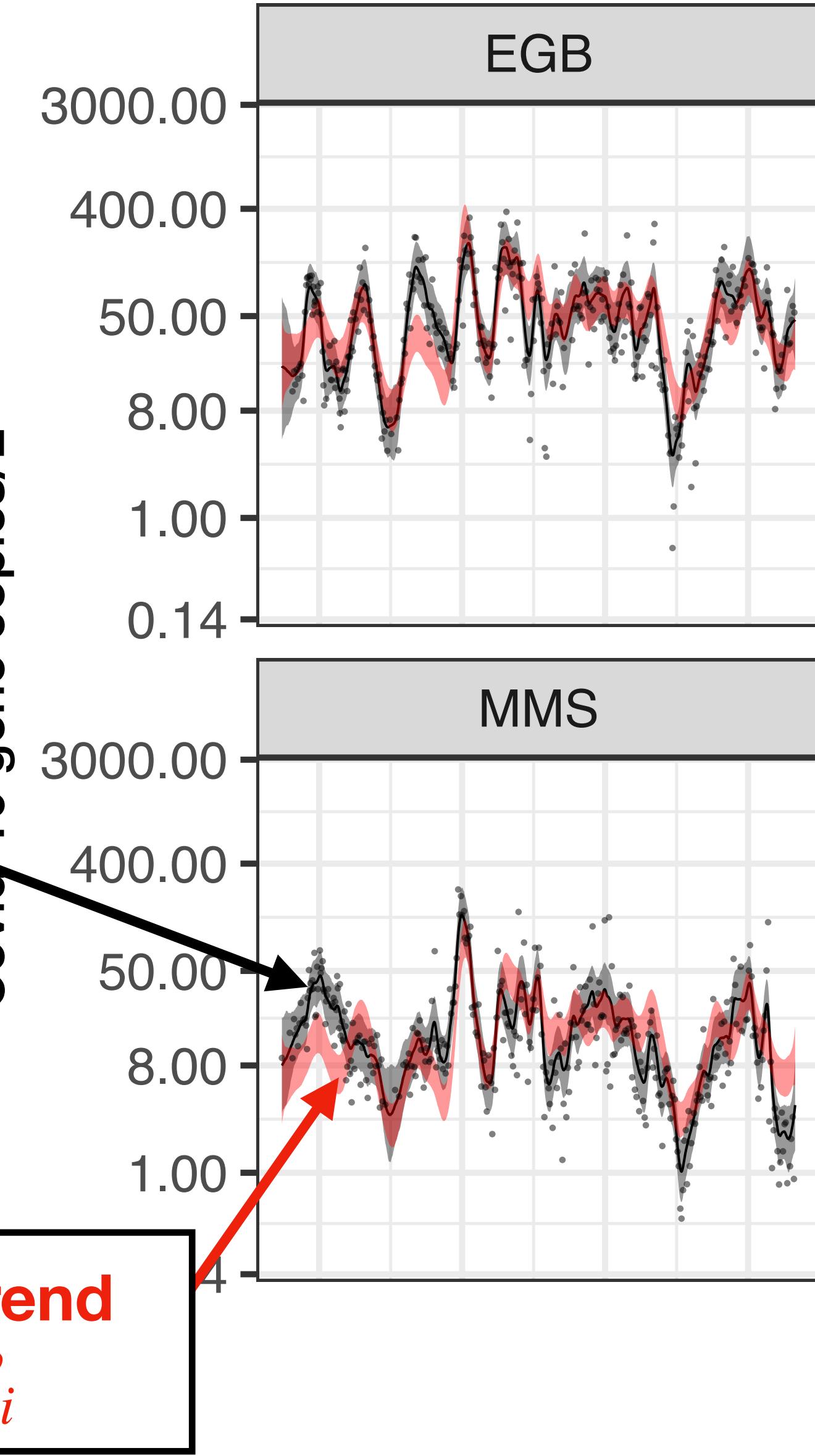
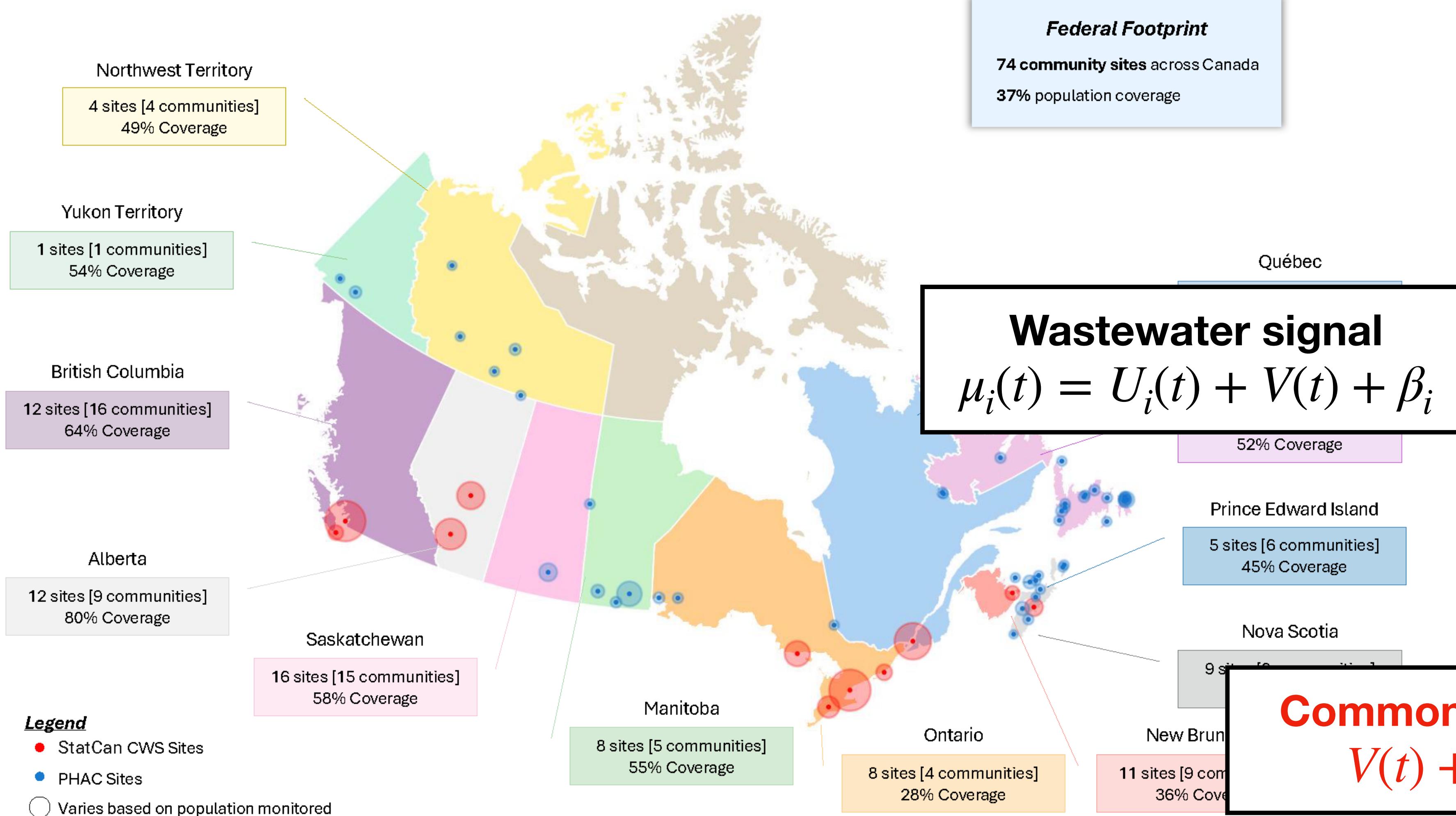




Canada

Current Federal, Provincial and Territorial Wastewater Monitoring Networks

Covers ~51% of the population connected to sewer systems across Canada





- In Somerset & Brown (2024) we used **differentiable Gaussian Processes (GPs)** to model differentiable wastewater signals
 - A. $V(t) \sim \text{IWP}_3(\sigma)$, inference on $V(t), V'(t), V''(t)$
 - B. $U_i(t) \sim \text{Matérn}(1.5, \phi, \tau)$, inference on $U_i(t), U'_i(t)$

*Journal of the Royal Statistical Society Series C:
Applied Statistics*, 2024, **00**, 1–16
<https://doi.org/10.1093/rsssc/qlae073>



Original Article

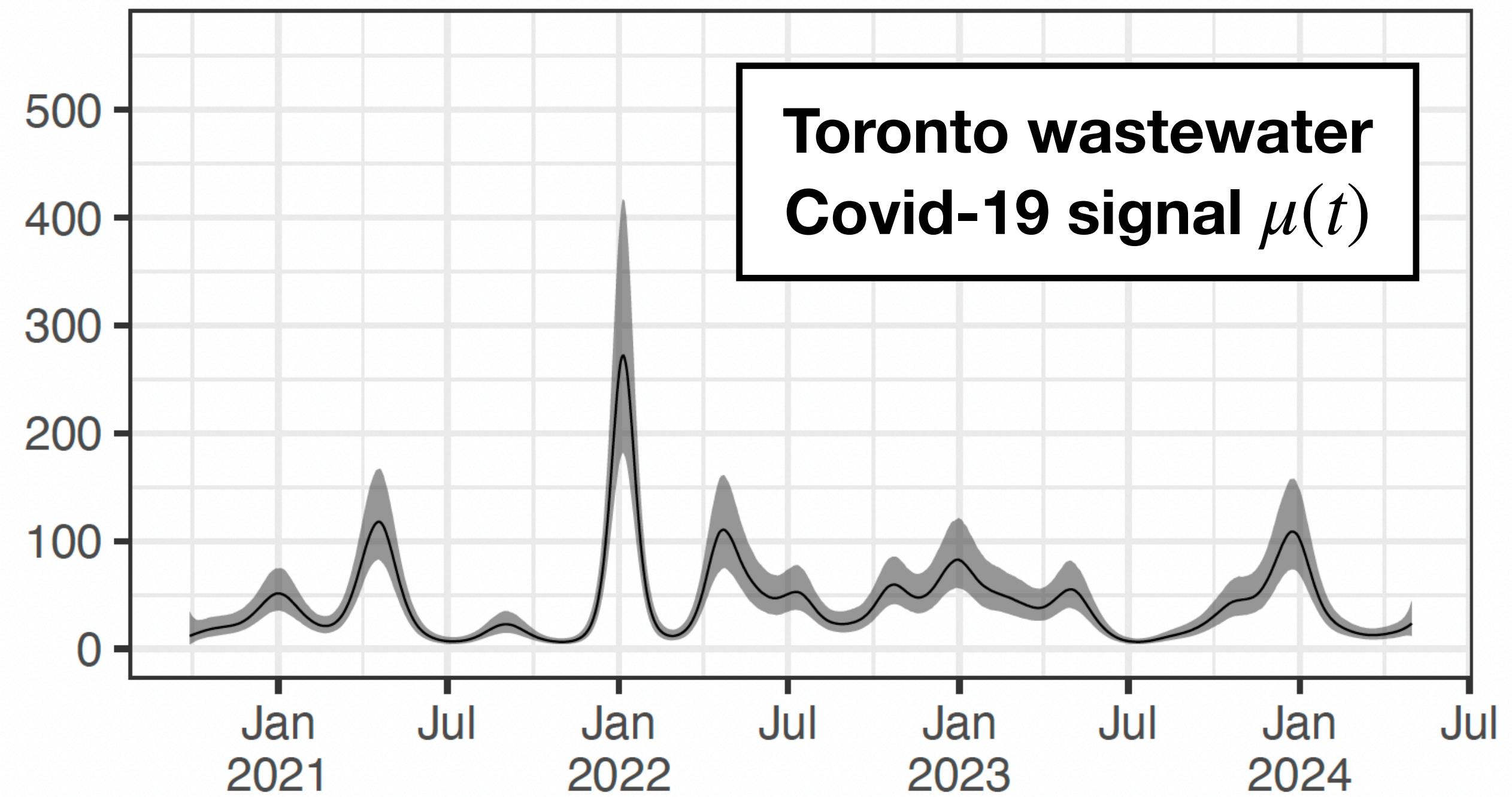
Wastewater surveillance using differentiable Gaussian processes

Emily Somerset¹ and Patrick E. Brown^{1,2}

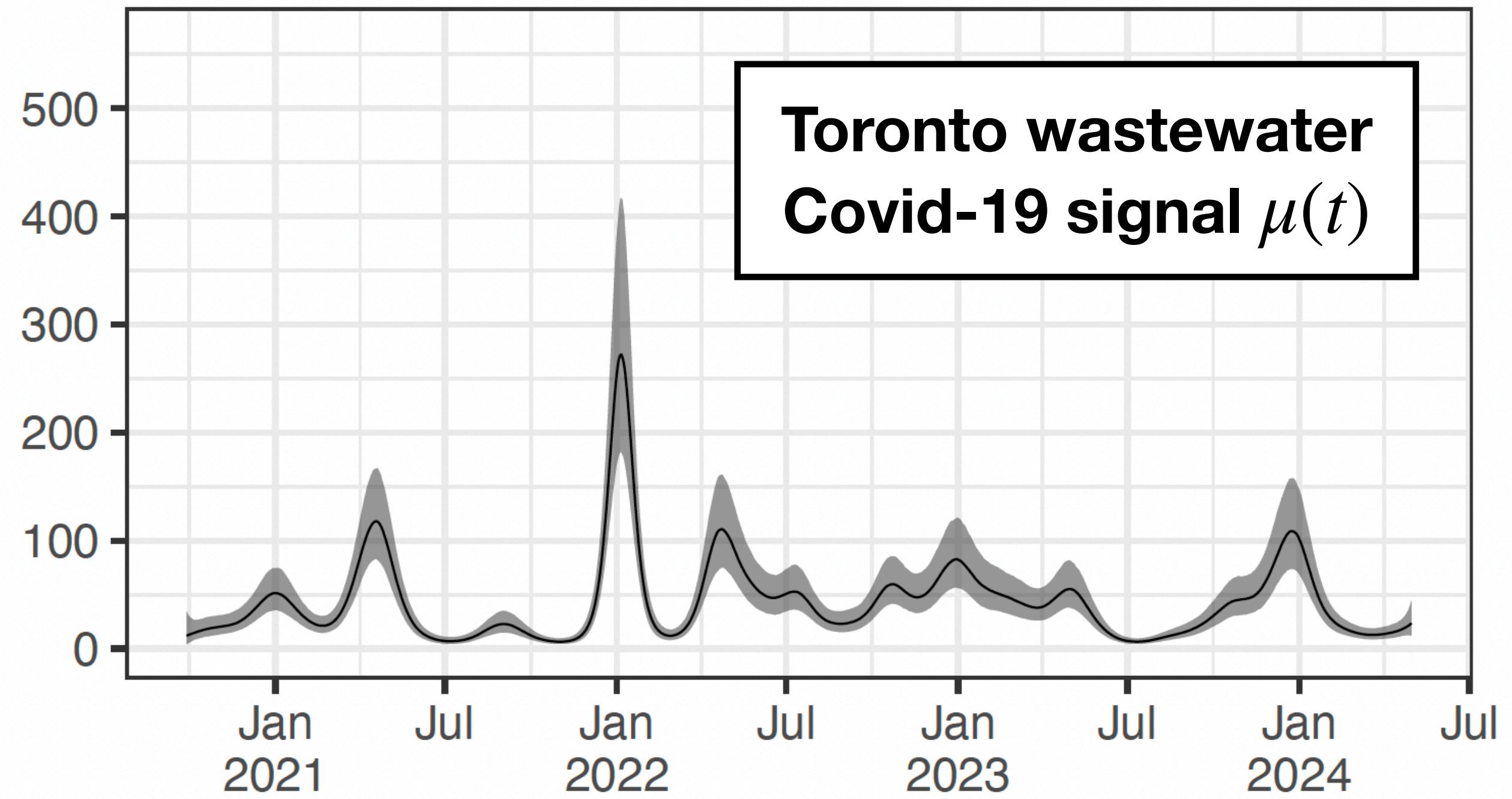
¹Department of Statistical Sciences, University of Toronto, Toronto, Canada

²Centre for Global Health Research, Unity Health, Toronto, Canada

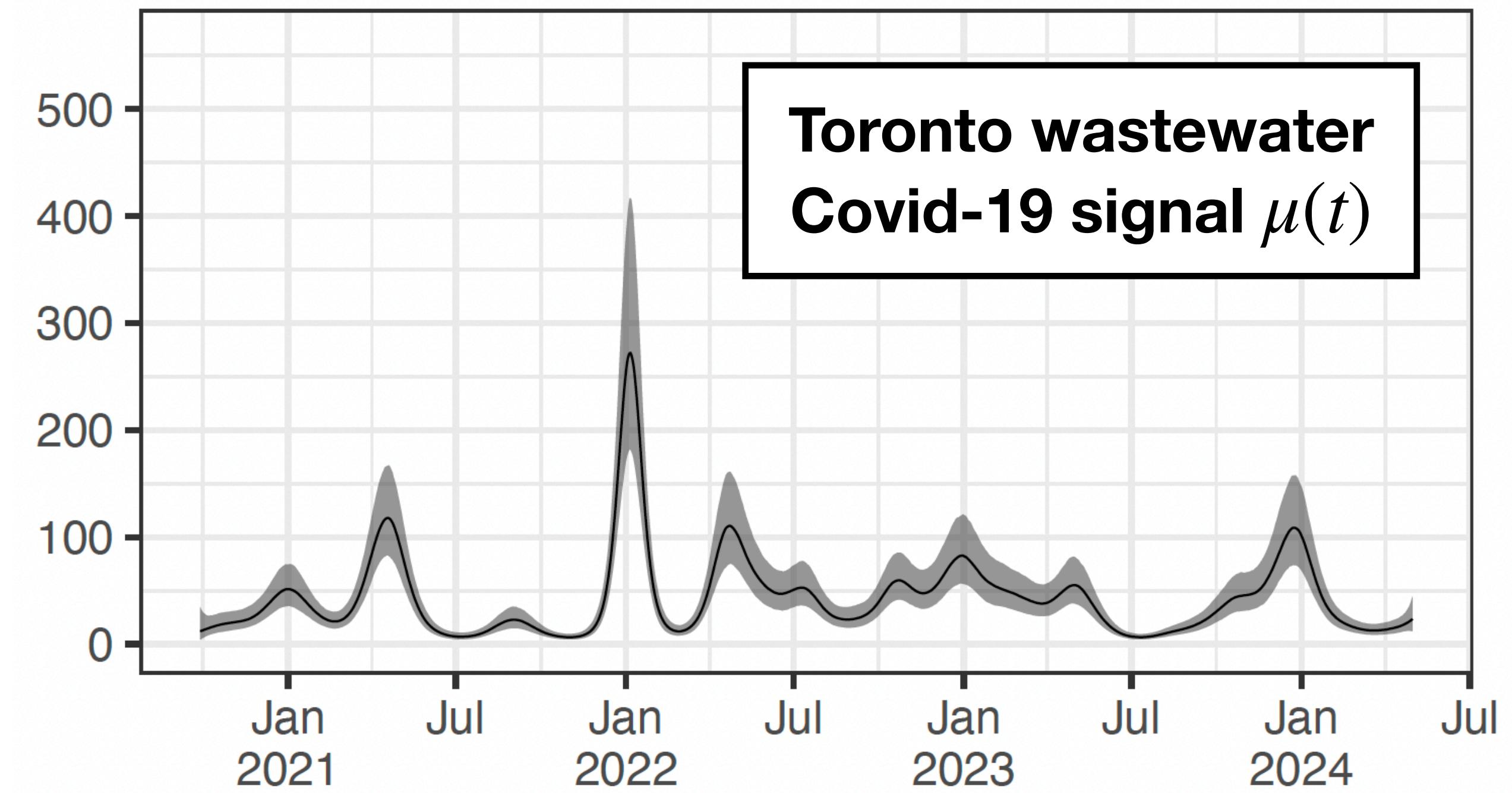
Address for correspondence: Emily Somerset, Department of Statistical Sciences, University of Toronto, 700 University Avenue, 9th Floor, Toronto, ON, Canada M5G 1Z5. Email: emily.somerset@mail.utoronto.ca



- $\mu(t)$ virus shed by humans at time t

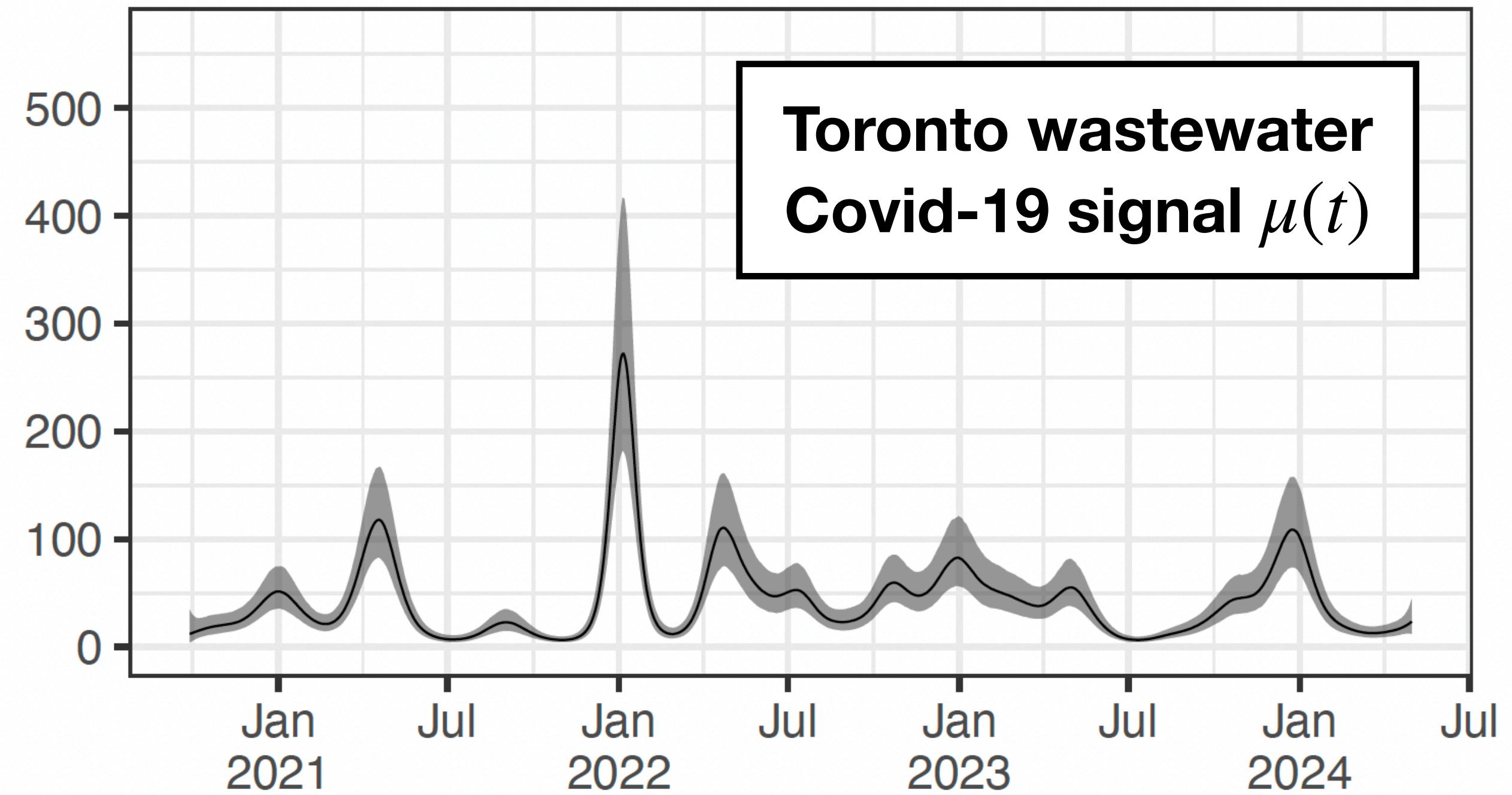


- $\mu(t)$ virus shed by humans at time t
- Roughly, if the number of people infected with COVID-19 doubles, we expect:



- $\mu(t)$ virus shed by humans at time t
- Roughly, if the number of people infected with COVID-19 doubles, we expect:

$$\frac{\mu(t+h)}{\mu(t)} \approx 2$$

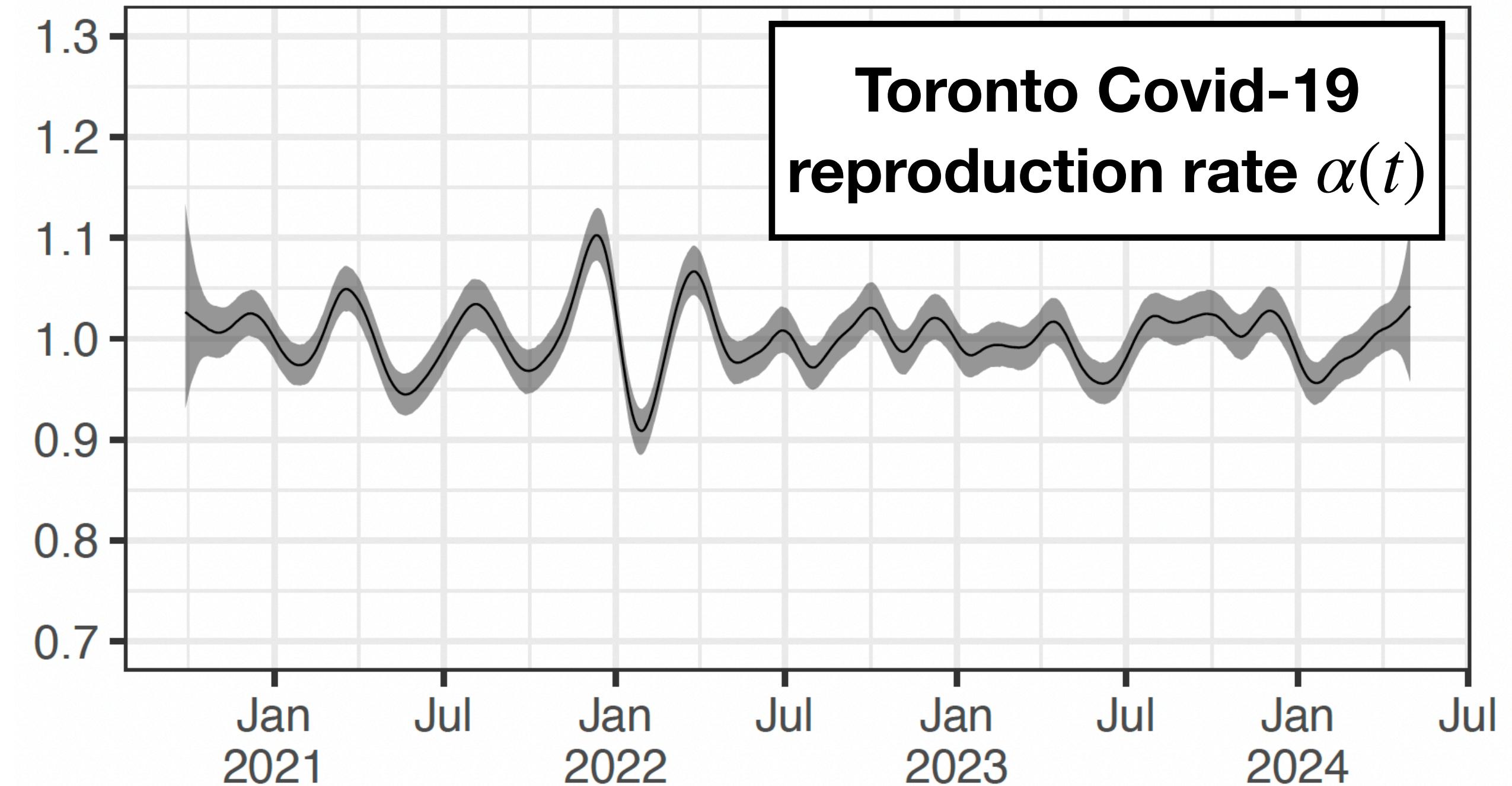


- $\mu(t)$ virus shed by humans at time t
- Roughly, if the number of people infected with COVID-19 doubles, we expect:

$$\frac{\mu(t+h)}{\mu(t)} \approx 2$$

- Leads to:

$$\alpha(t) = \lim_{h \rightarrow 0} \left(\frac{\mu(t+h)}{\mu(t)} \right)^{1/h} = \exp \left(\frac{\mu'(t)}{\mu(t)} \right)$$



Inference on *differentiable* $\mu(t)$ through our wastewater model

The joint wastewater and case count model:

$$\left. \begin{array}{l} Y_t | I_t \sim \text{Bin}(I_t, \pi_t) \\ I_t | I_{t-1} \sim \text{Pois}(\lambda_t) \\ \lambda_t = \nu + \alpha_t X_{t-1} \end{array} \right\} \begin{array}{|c|} \hline \text{Infection model} \\ \hline \end{array}$$

Wastewater model $\left\{ G_{it} \sim \text{Waterwater model}(\alpha_t, \vec{\gamma}_{ww}) \right.$

...

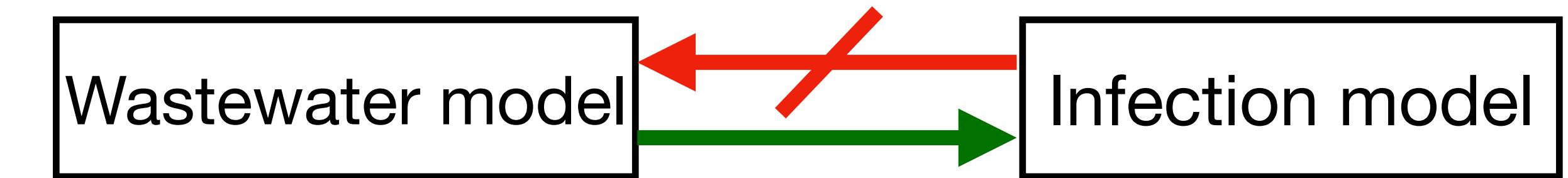
The joint wastewater and case count model:

$$\left. \begin{aligned} Y_t | I_t &\sim \text{Bin}(I_t, \pi_t) \\ I_t | I_{t-1} &\sim \text{Pois}(\lambda_t) \\ \lambda_t &= \nu + \alpha_t X_{t-1} \end{aligned} \right\} \quad \begin{array}{|bbox=665 405 835 465} \text{Infection model} \end{array}$$

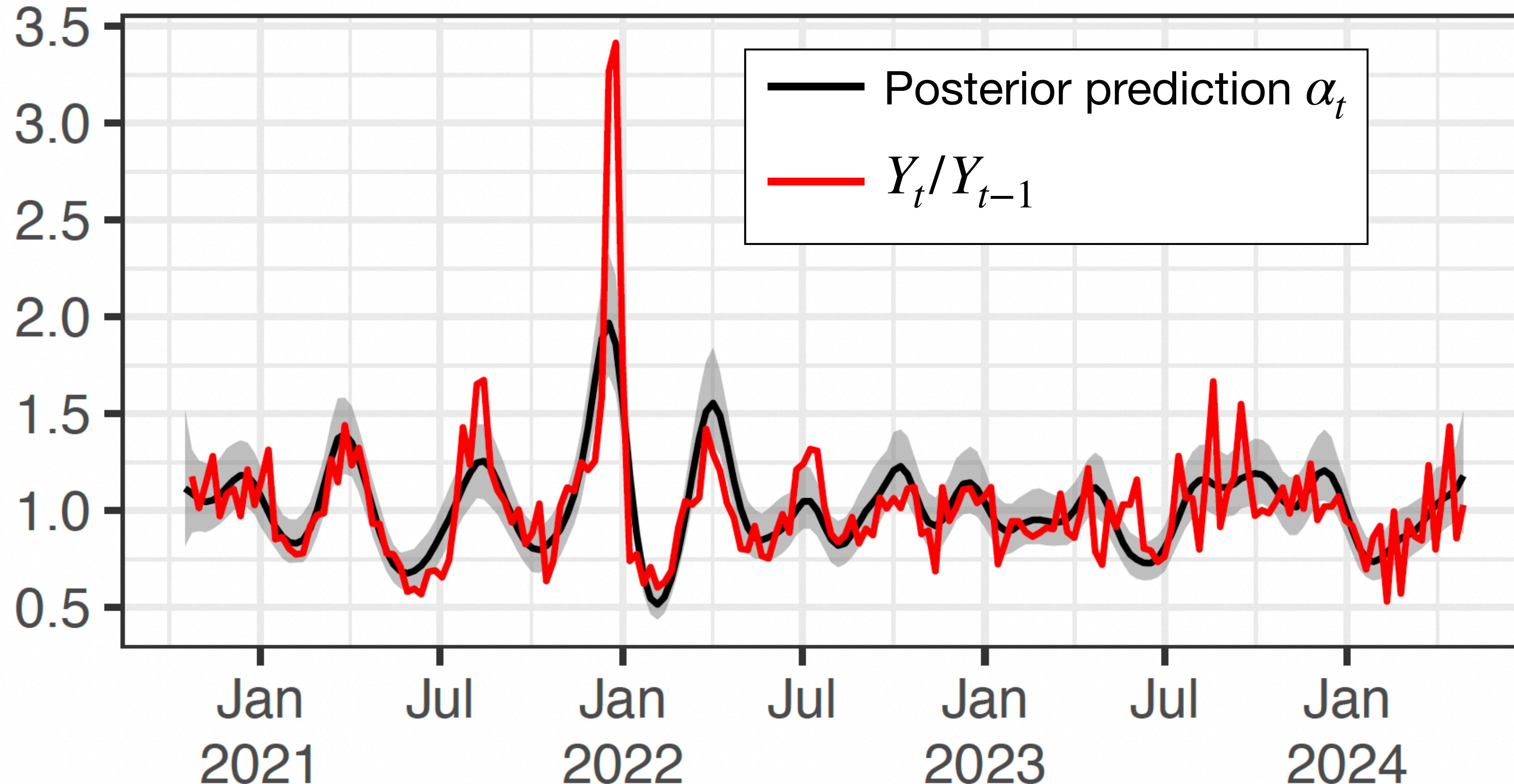
Wastewater model $\left\{ G_{it} \sim \text{Waterwater model}(\alpha_t, \vec{\gamma}_{ww}) \right.$

...

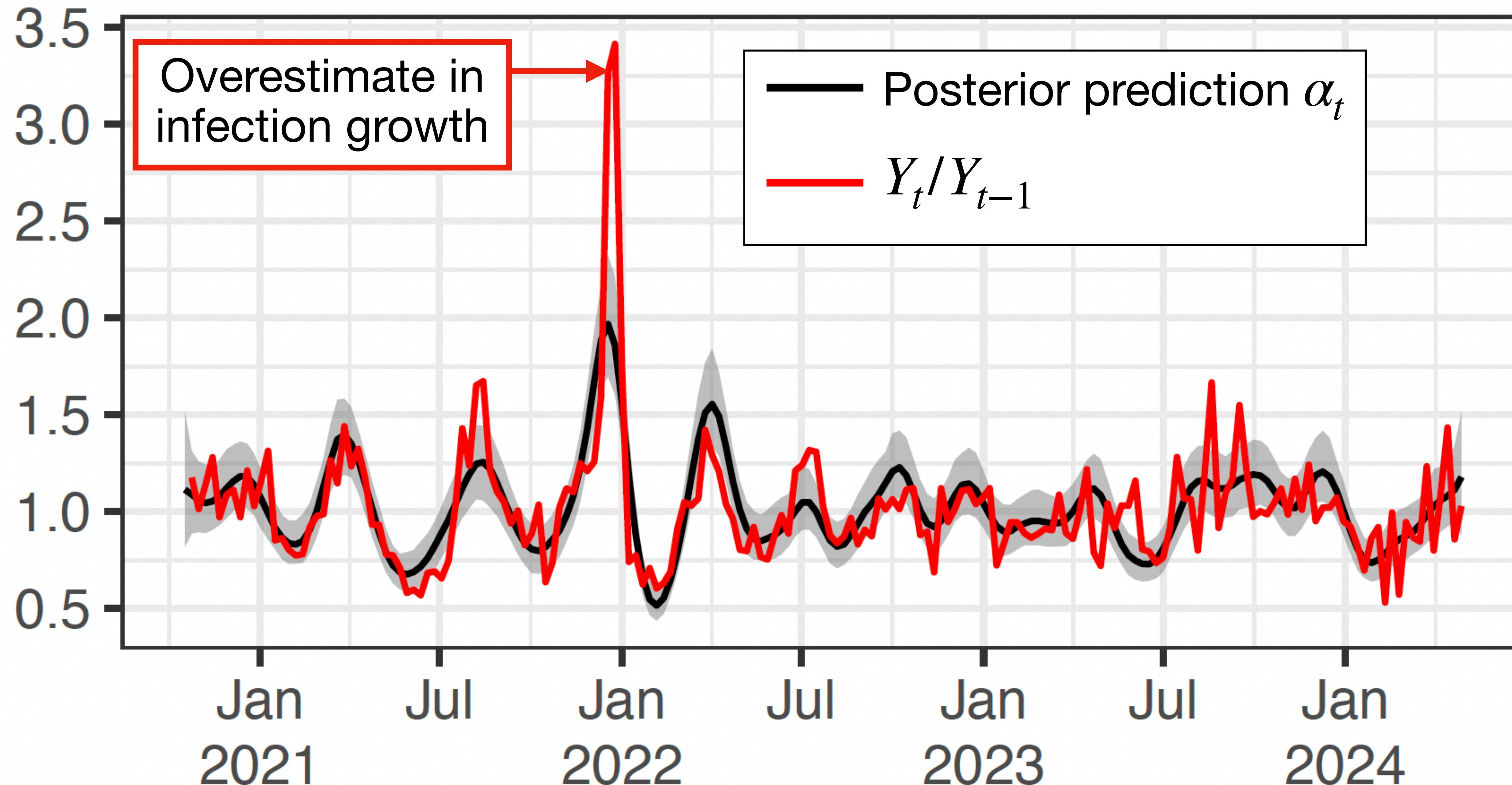
Fit this model using **modular inference [5]** to remove feedback from infection model into wastewater.

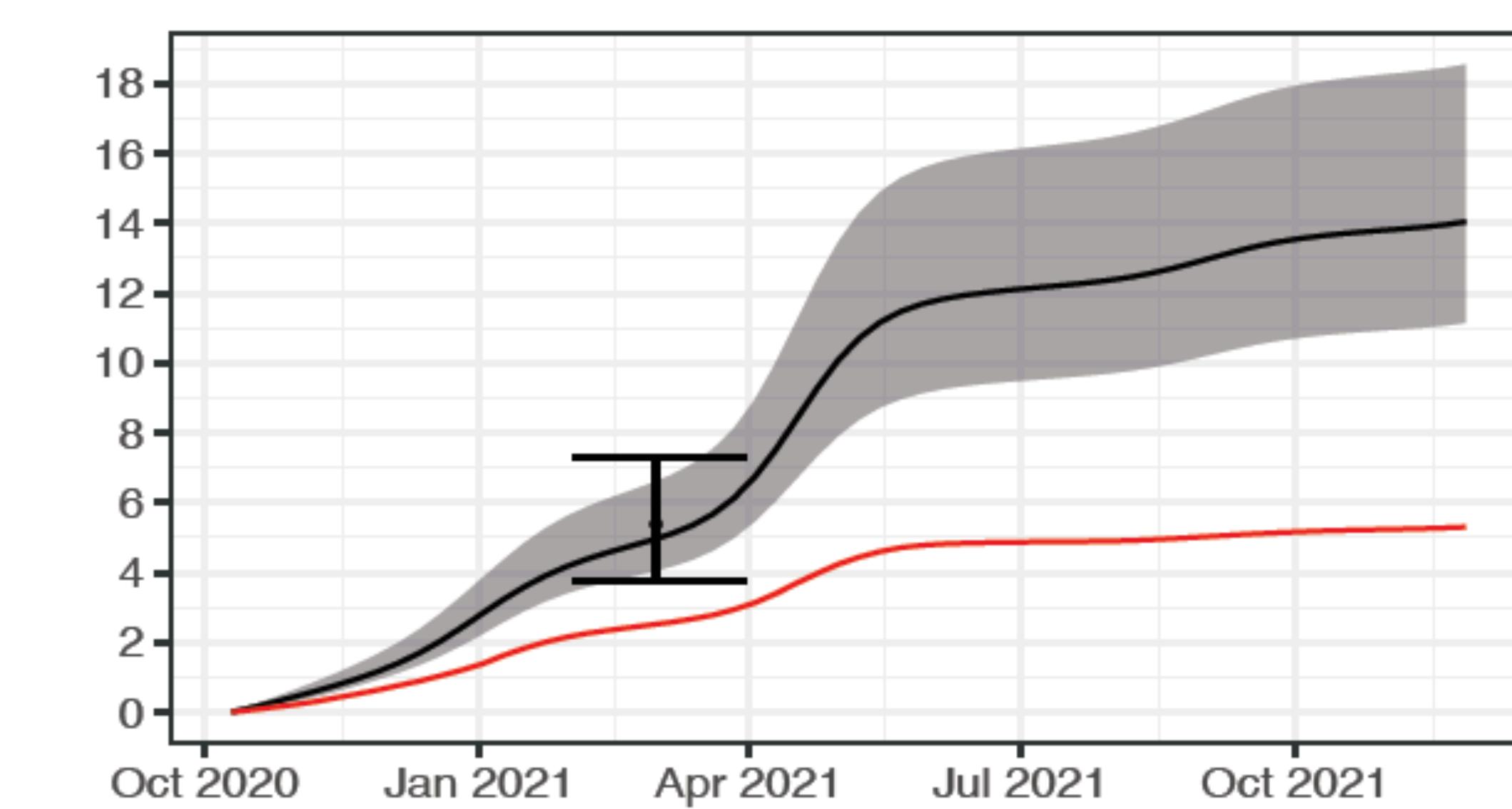
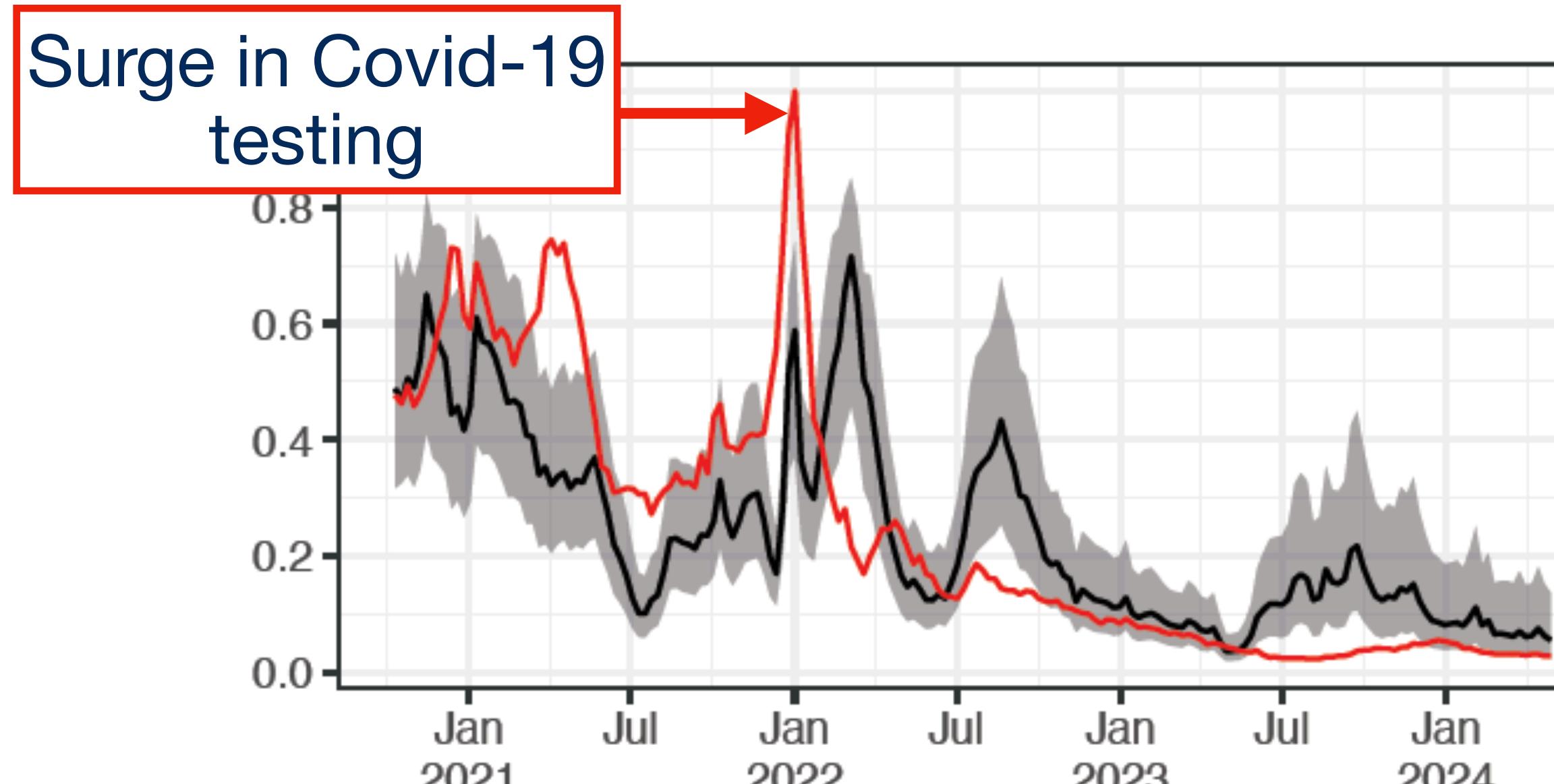
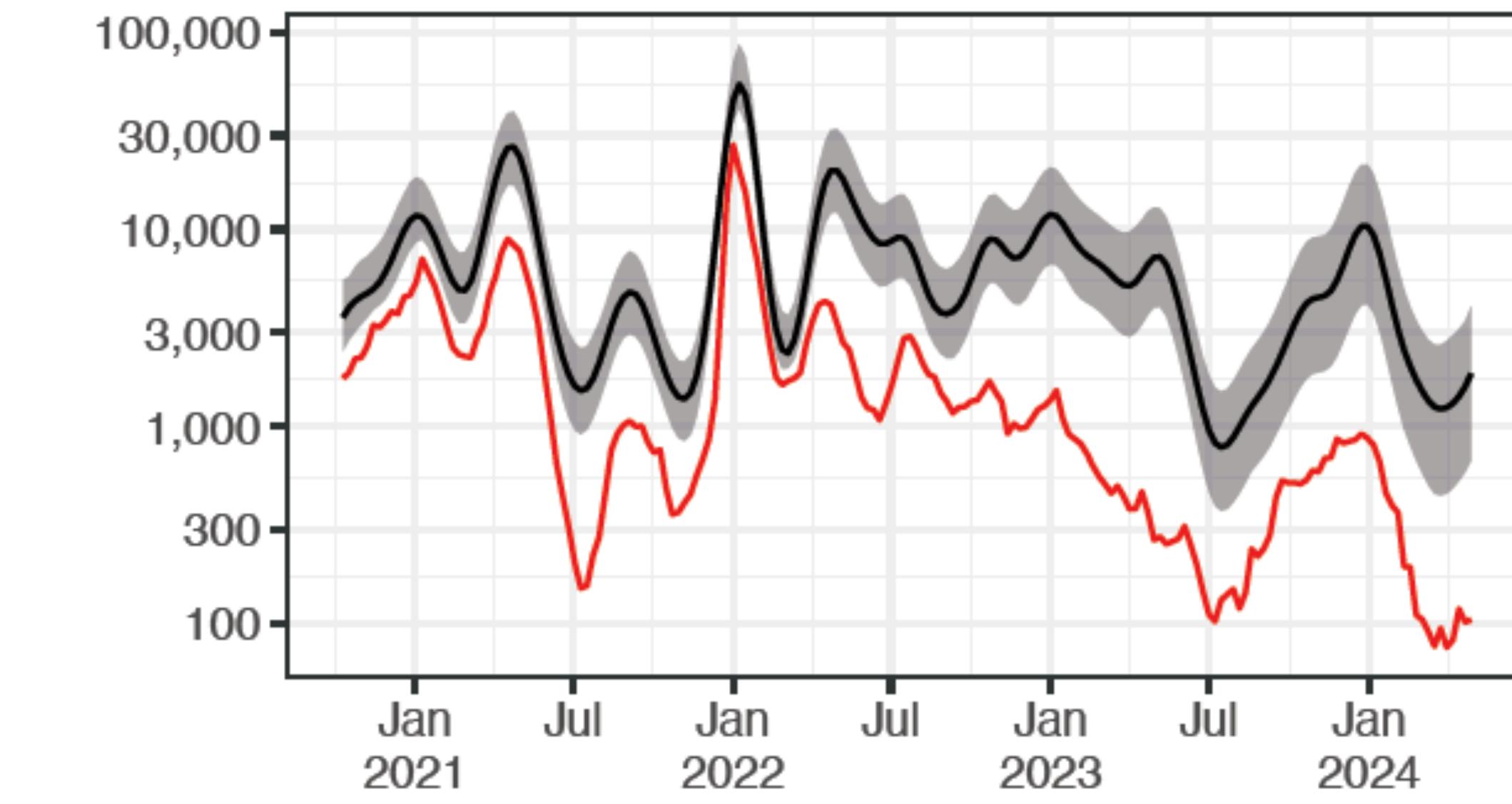
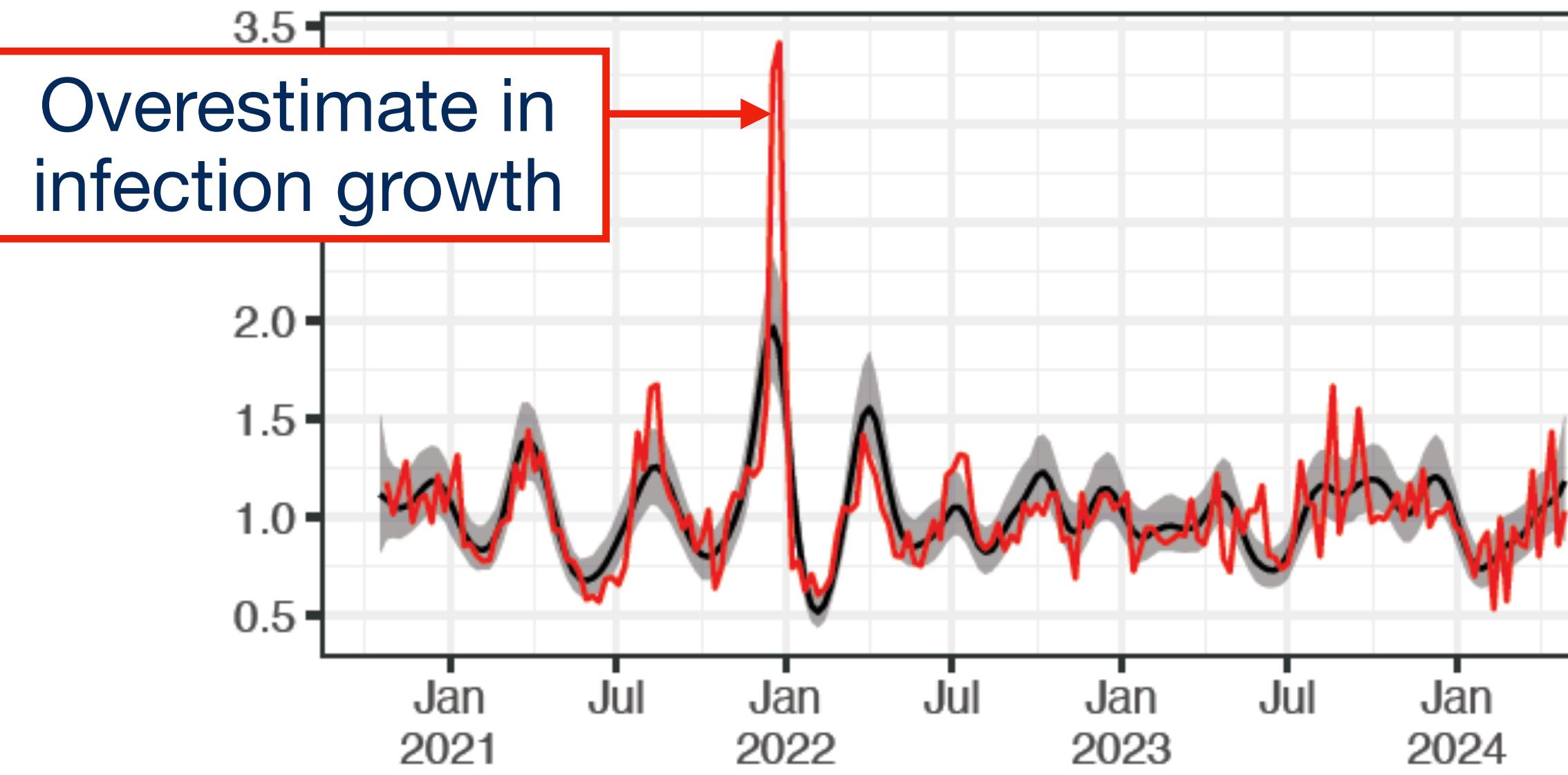


Results: Wastewater-based reproduction rates



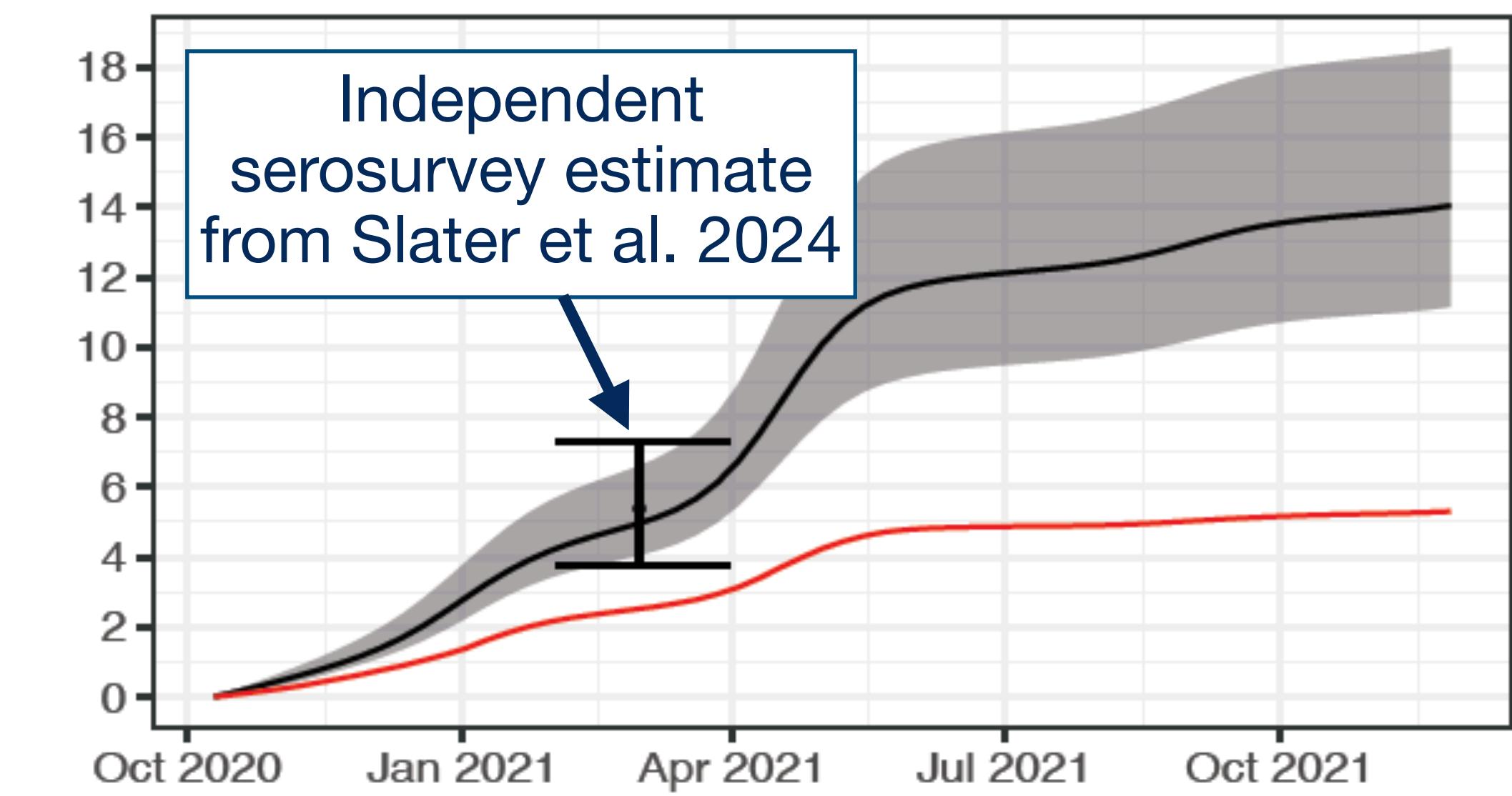
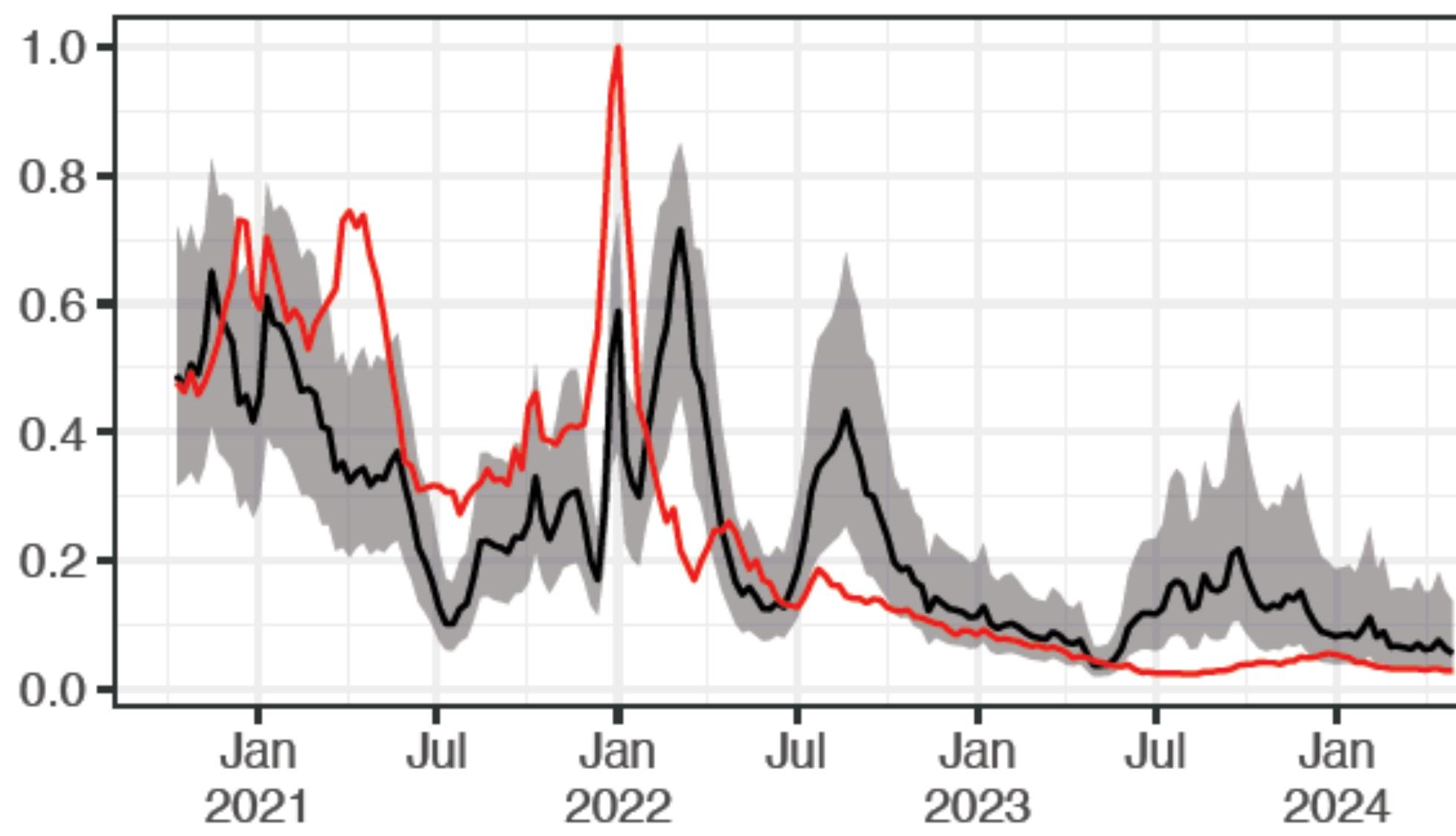
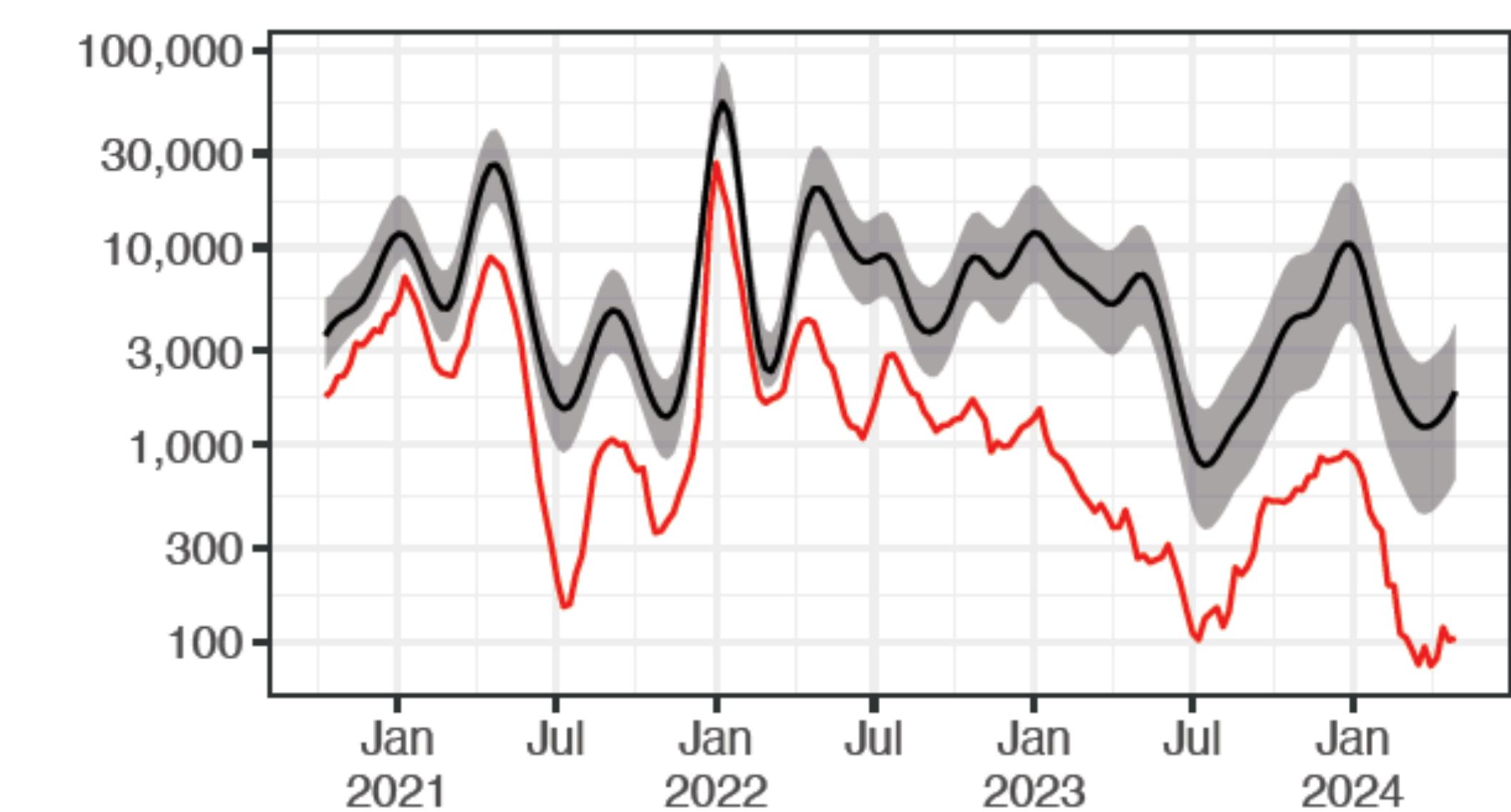
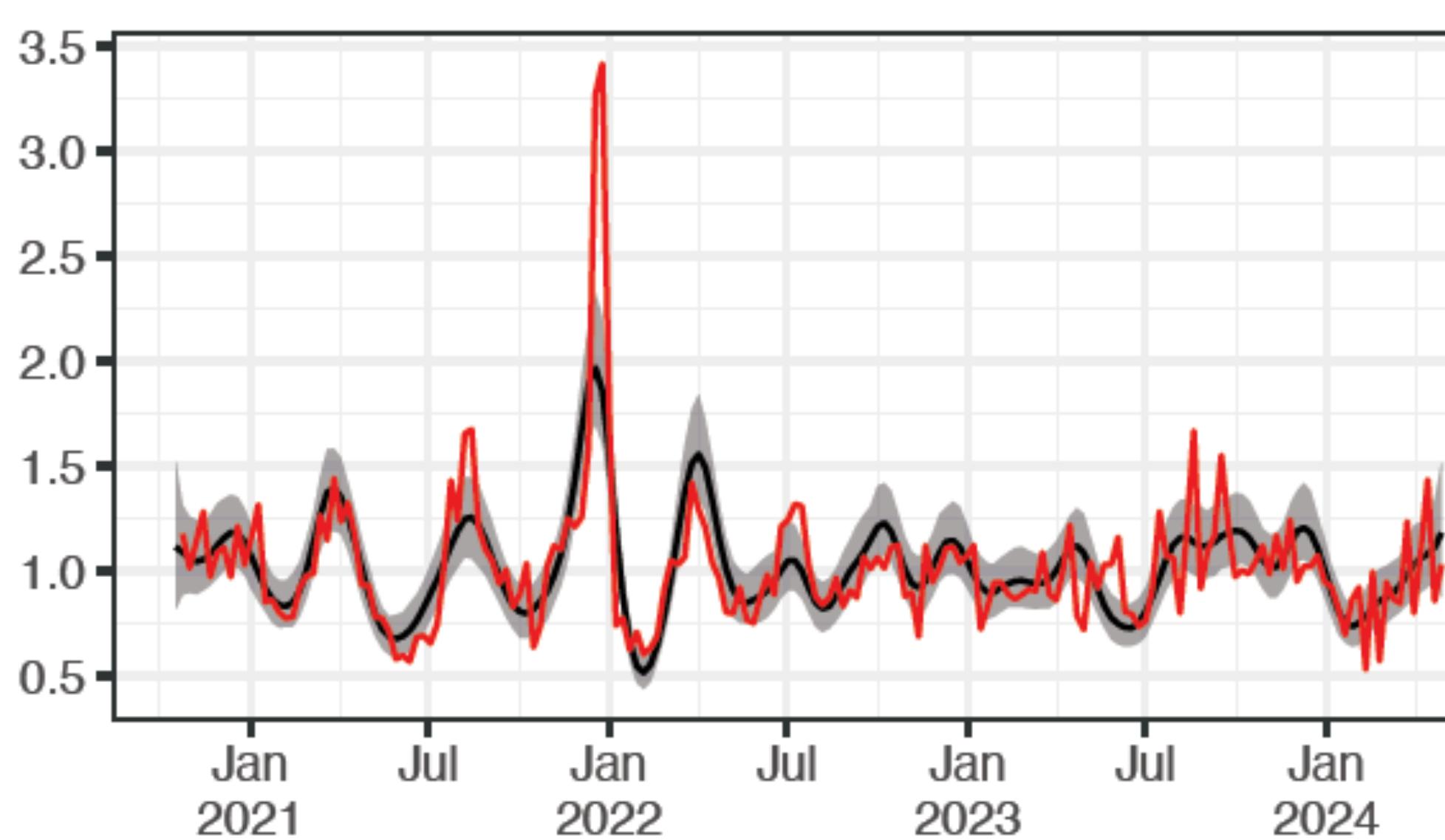
Results: Wastewater-based reproduction rates





(c) Reporting probability, π_j

(d) Pre-Omicron cumulative incidence, $C_j\%$





UNIVERSITY OF TORONTO
FACULTY OF ARTS & SCIENCE

Department of Statistical Sciences

Key takeaways:

Key takeaways:

- When modelling infectious disease data, account for the under-reporting mechanism

Key takeaways:

- When modelling infectious disease data, account for the under-reporting mechanism
- Good epidemic curve reconstructions involve multiple data sources

Key takeaways:

- When modelling infectious disease data, account for the under-reporting mechanism
- Good epidemic curve reconstructions involve multiple data sources
- Using differentiable processes when modelling wastewater data is very useful

Key takeaways:

- When modelling infectious disease data, account for the under-reporting mechanism
- Good epidemic curve reconstructions involve multiple data sources
- Using differentiable processes when modelling wastewater data is very useful
- Introduced wastewater-based reproduction rates

Key takeaways:

- When modelling infectious disease data, account for the under-reporting mechanism
- Good epidemic curve reconstructions involve multiple data sources
- Using differentiable processes when modelling wastewater data is very useful
- Introduced wastewater-based reproduction rates
- Incorporated them into infectious disease models

Key takeaways:

- When modelling infectious disease data, account for the under-reporting mechanism
- Good epidemic curve reconstructions involve multiple data sources
- Using differentiable processes when modelling wastewater data is very useful
- Introduced wastewater-based reproduction rates
- Incorporated them into infectious disease models
- Serosurveys, if available, can be used to corroborate results

References:

1. Quick, C., Dey, R., & Lin, X. (2021). Regression models for understanding COVID-19 epidemic dynamics with incomplete data. *Journal of the American Statistical Association*, 116(536), 1561-1577.
2. ArcGIS Dashboards. "COVID-19 Dashboard." Accessed May 29, 2024. <https://www.arcgis.com/apps/dashboards/aa7f07e518b9461497b3ccc740ae4bb6>.
3. Public Health Agency of Canada. "Covid-19 Wastewater Surveillance Dashboard." Canada.ca. Accessed May 29, 2024. <https://health-infobase.canada.ca/covid-19/wastewater/>.
4. Somerset, Emily, and Patrick E. Brown. "Wastewater surveillance using differentiable Gaussian processes." *Journal of the Royal Statistical Society Series C: Applied Statistics* (2024): qlae073.
5. Martyn Plummer. "Cuts in Bayesian graphical models". In: *Statistics and Computing* 25(2015), pp. 37–43.

References:

1. Quick, C., Dey, R., & Lin, X. (2021). Regression models for understanding COVID-19 epidemic dynamics with incomplete data. *Journal of the American Statistical Association*, 116(536), 1561-1577.
2. ArcGIS Dashboards. "COVID-19 Dashboard." Accessed May 29, 2024. <https://www.arcgis.com/apps/dashboards/aa7f07e518b9461497b3ccc740ae4bb6>.
3. Public Health Agency of Canada. "Covid-19 Wastewater Surveillance Dashboard." Canada.ca. Accessed May 29, 2024. <https://health-infobase.canada.ca/covid-19/wastewater/>.
4. Somerset, Emily, and Patrick E. Brown. "Wastewater surveillance using differentiable Gaussian processes." *Journal of the Royal Statistical Society Series C: Applied Statistics* (2024): qlae073.
5. Martyn Plummer. "Cuts in Bayesian graphical models". In: *Statistics and Computing* 25(2015), pp. 37–43.