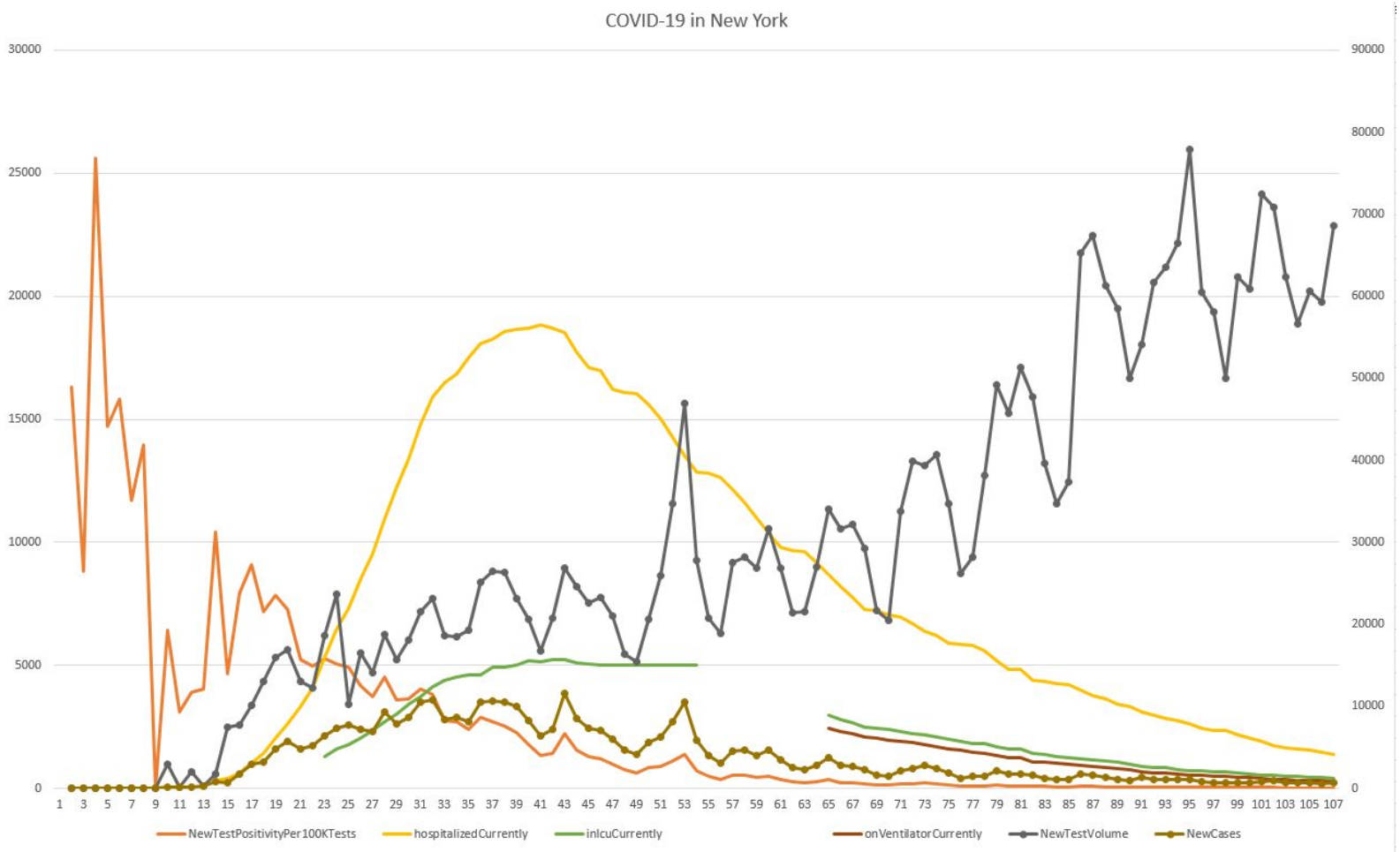
Particle Filtering & PMCMC for Integrating Wastewater Surveillance & Health System Data: Motivations, Mathematical & Algorithmic Structure and Wastewater Data Essentials

Nathaniel Osgood
Computational Epidemiology & Public Health Informatics Laboratory
University of Saskatchewan

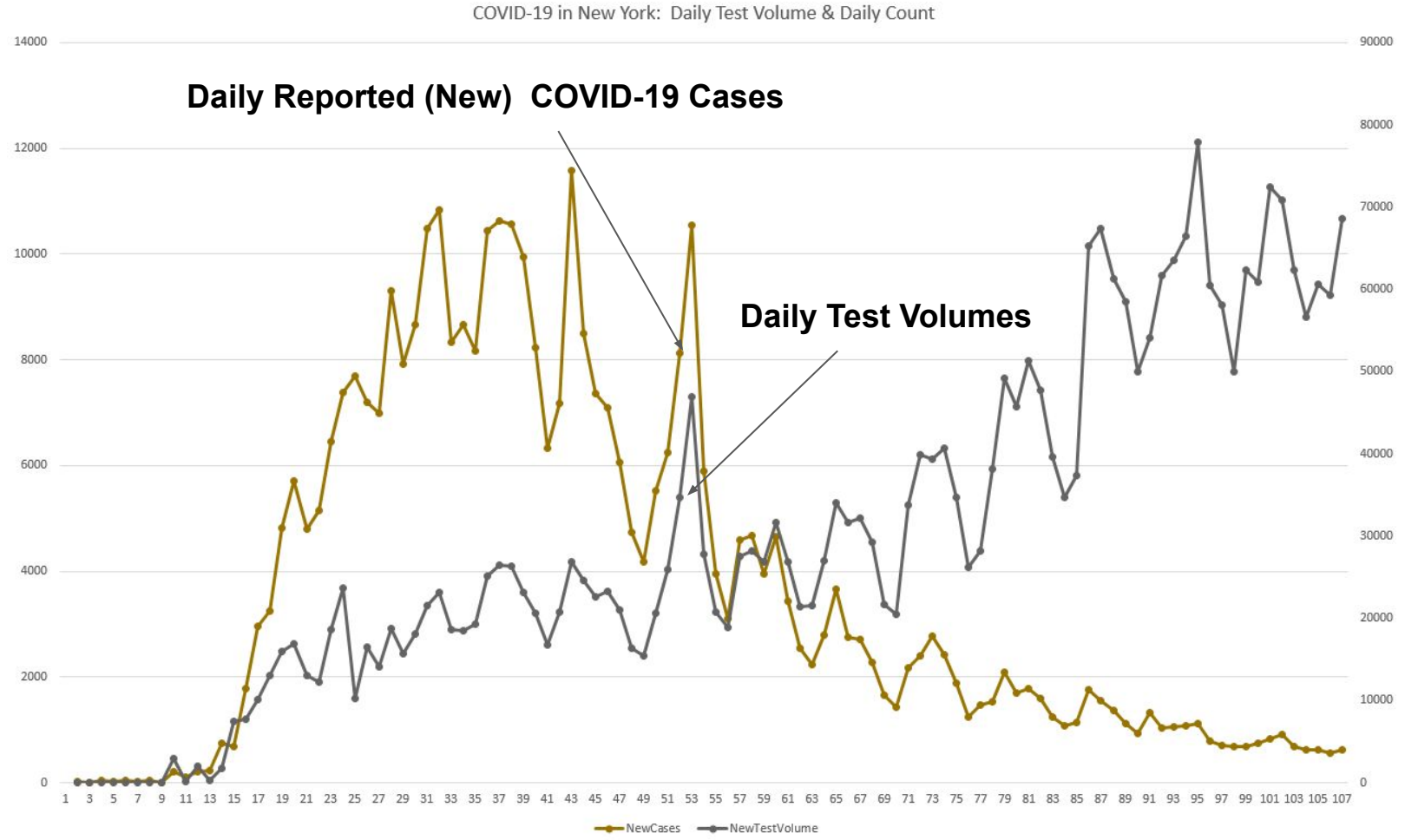
Agenda

- Motivation
- Particle Filtering with dynamic models
- Wastewater essentials
- Key elements of wastewater-specific model formulation
  - Likelihood
  - Forms of nexus between model state and wastewater concentration
- [As time allows] PMCMC
- Conclusions

Making Sense of Interconnected Time Series: NY



Intertwining of Testing & Case Counts



Wastewater Epidemiology

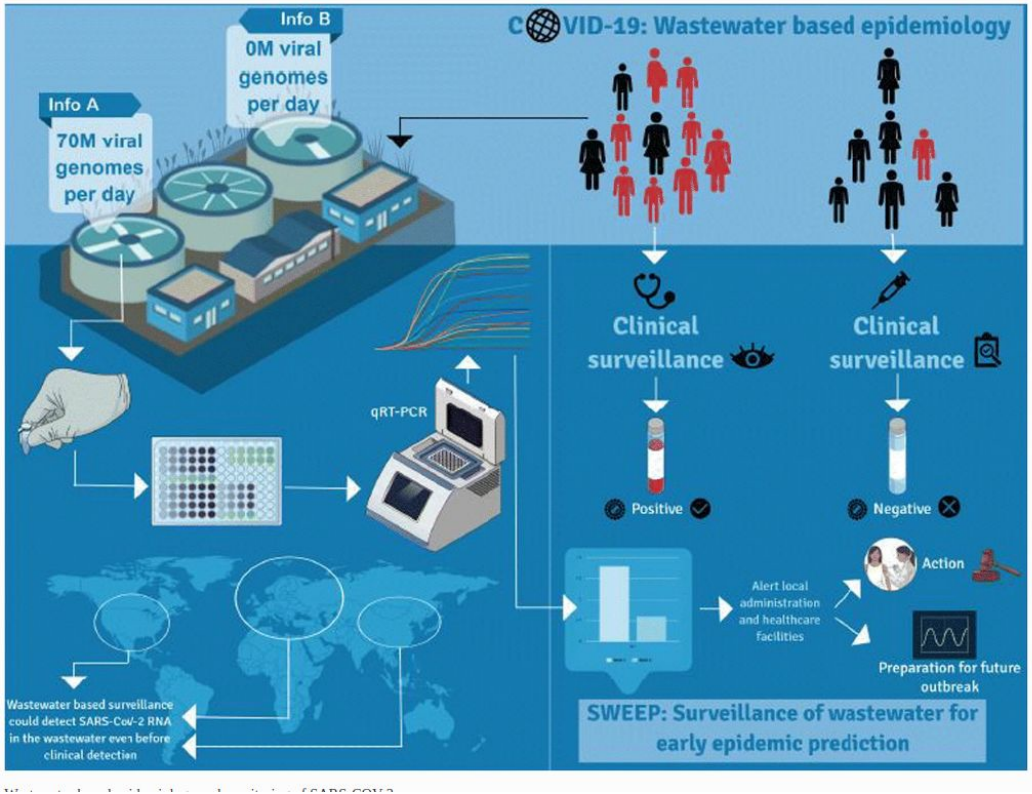
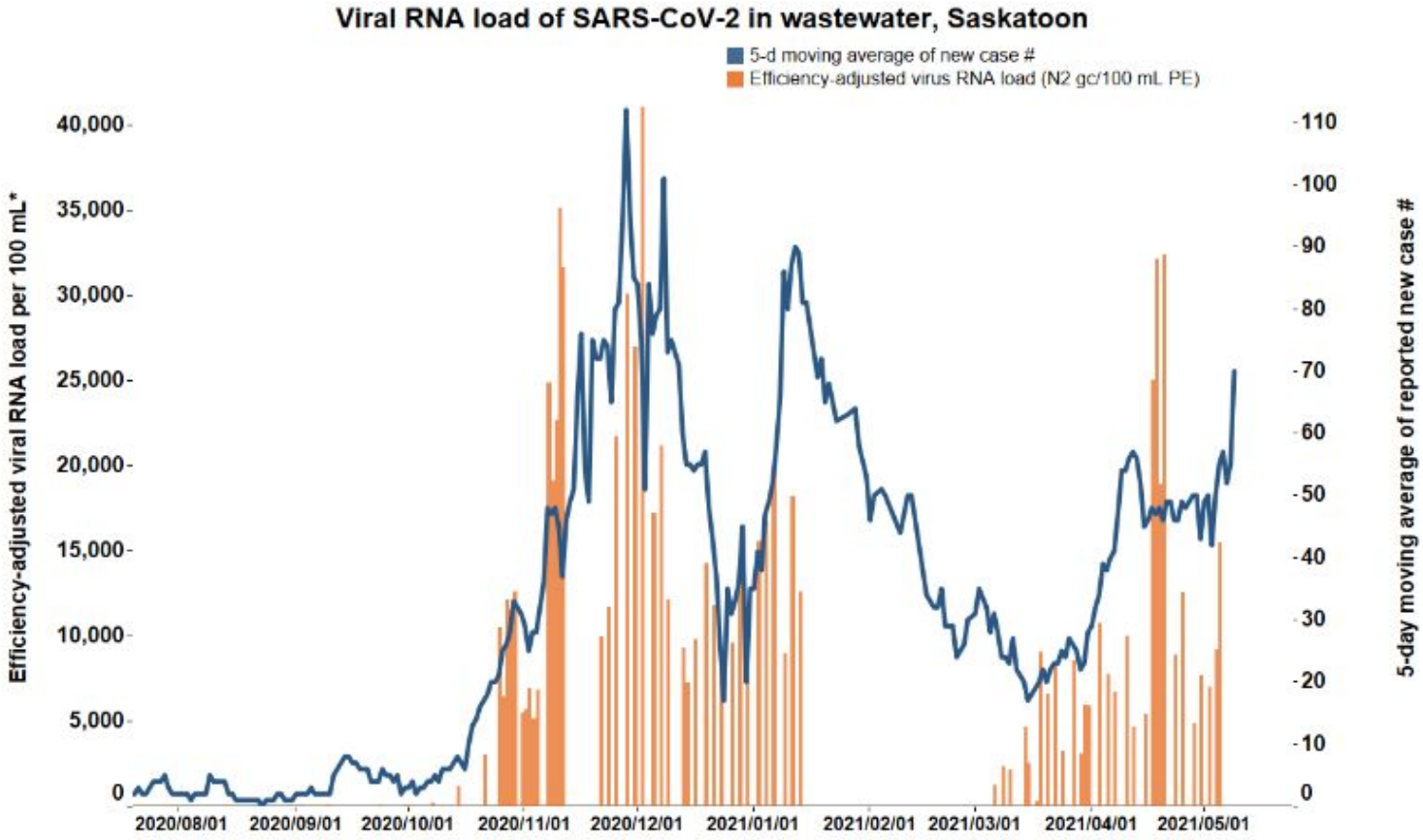
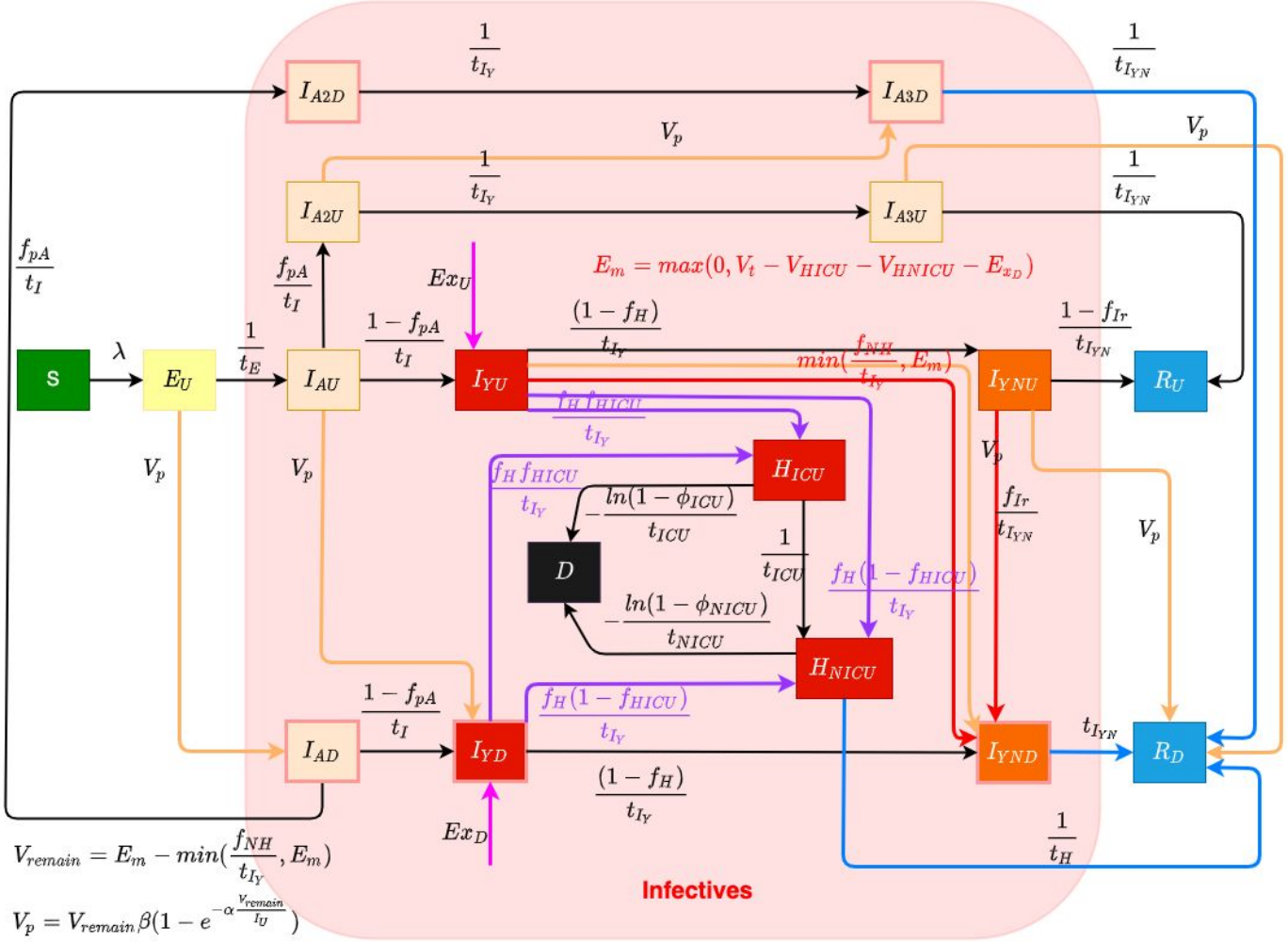


Image source: Kumar, M., Mohapatra, S., Mazumder, P. et al. Making Waves Perspectives of Modelling and Monitoring of SARS-CoV-2 in Aquatic Environment for COVID-19 Pandemic. Curr Pollution Rep 6, 468–479 (2020). https://doi.org/10.1007/s40726-020-00161-5

Wastewater Data



Source: University of Saskatchewan Toxicology Centre via COVID-19 Early Indicators, https://water.usask.ca/covid-19/#MeasuringVirusIndicatorsinWastewaterasanEarlyWarningofCOVID19Outbreaks



Example COVID-19 Dynamic Model

Associated ODEs

Stocks:

$$\frac{d\mathbf{S}}{dt} = -\lambda S$$

$$\frac{d\mathbf{E_U}}{dt} = \lambda S - \frac{E_U}{t_E} - V_p\frac{E_U}{I_U}$$

$$\frac{d\mathbf{E_D}}{dt} = V_p\frac{E_U}{I_U} - \frac{E_D}{t_E}$$

$$\frac{d\mathbf{I_{AU}}}{dt} = \frac{E_U}{t_E} - \frac{I_{AU}}{t_I} - V_p\frac{I_{AU}}{I_U}$$

$$\frac{\mathbf{I_{AD}}}{dt} = \frac{E_D}{t_E} + V_p\frac{I_{AU}}{I_U} - \frac{I_{AD}}{t_I}$$

$$\frac{d\mathbf{I_{A2U}}}{dt} = f_{pA}\frac{I_{AU}}{t_I} - \frac{I_{A2U}}{t_{I_Y}} - V_p\frac{I_{A2U}}{I_U}$$

$$\frac{d\mathbf{I_{A2D}}}{dt} = f_{pA}\frac{I_{AD}}{t_I} + V_p\frac{I_{A2U}}{I_U} - \frac{I_{A2D}}{t_{I_Y}}$$

$$\frac{d\mathbf{I_{A3U}}}{dt} = \frac{I_{A2U}}{t_{I_Y}} - V_p\frac{I_{A3U}}{I_U} - \frac{I_{A3U}}{t_{I_{YN}}}$$

$$\frac{d\mathbf{I_{A3D}}}{dt} = \frac{I_{A2D}}{t_{I_Y}} + V_p\frac{I_{A3U}}{I_U} - \frac{I_{A3D}}{t_{I_{YN}}}$$

$$\frac{d\mathbf{I_{YU}}}{dt} = Ex_D\frac{1-f_S}{f_S} + (1.0-f_{pA})\frac{I_{AU}}{t_I} - \frac{I_{YU}}{t_{I_Y}} - V_p\frac{I_{YU}}{I_U} - min(\frac{I_{YU}f_{NH}}{t_{I_Y}}, E_m)$$

$$\frac{d\mathbf{I_{YD}}}{dt} = Ex_D + (1-f_{pA})\frac{I_{AD}}{t_I} + V_p\frac{I_{YU}}{I_U} - \frac{I_{YD}}{t_{I_Y}}$$

$$\frac{d\mathbf{H_{ICU}}}{dt} = I_{YU}\frac{f_H f_{HICU}}{t_{I_Y}} + I_{YD}\frac{f_H f_{HICU}}{t_{I_Y}} - \frac{H_{ICU}}{t_{ICU}} - (H_{ICU}\frac{-ln(1-\phi_{ICU})}{t_{ICU}})$$

$$\frac{d\mathbf{H_{NICU}}}{dt} = I_{YU}\frac{f_H(1-f_{HICU})}{t_{I_Y}} + I_{YD}\frac{f_H(1-f_{HICU})}{t_{I_Y}} + \frac{H_{ICU}}{t_{ICU}} - (H_{NICU}\frac{-ln(1-\phi_{NICU})}{t_{NICU}}) - \frac{H_{NICU}}{t_H}$$

$$\frac{d\mathbf{I_{YNU}}}{dt} = I_{YU}\frac{1-f_H}{t_{I_Y}} - \frac{I_{YNU}}{t_{I_{YN}}} - V_p\frac{I_{YNU}}{I_U}$$

$$\frac{d\mathbf{I_{YND}}}{dt} = f_{I_r}\frac{I_{YNU}}{t_{I_{YN}}} + V_p\frac{I_{YNU}}{I_U} + \frac{I_{YD}}{t_{I_Y}} + min(\frac{I_{YU}f_{NH}}{t_{I_Y}}, E_m) - \frac{I_{YND}}{t_{I_{YN}}}$$

$$\frac{d\mathbf{R_U}}{dt} = (1-f_{I_r})\frac{I_{YNU}}{t_{I_{YN}}} + \frac{I_{A3U}}{t_{I_{YN}}}$$

$$\frac{d\mathbf{R_D}}{dt} = \frac{I_{YND}}{t_{I_{YN}}} + \frac{H_{ICU}}{t_H} + \frac{I_{A3D}}{t_{I_{YN}}}$$

$$\frac{d\mathbf{D}}{dt} = (H_{NICU}\frac{-ln(1-\phi_{NICU})}{t_{NICU}}) + (H_{ICU}\frac{-ln(1-\phi_{ICU})}{t_{ICU}})$$

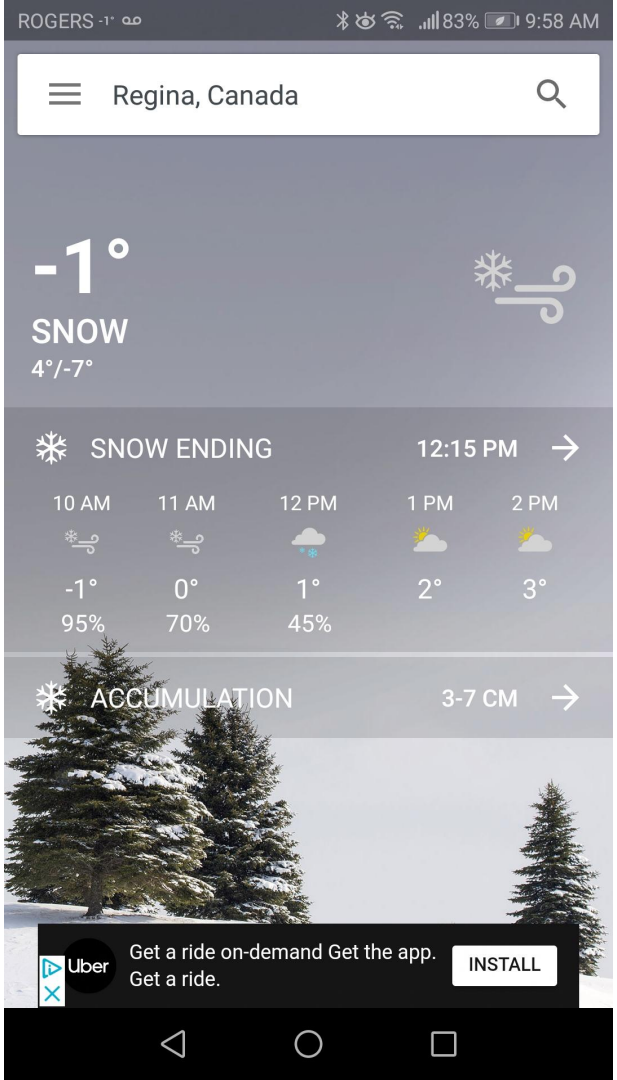
Dynamic Modeling & Empirical Situation

- Models can offer great utility in
  - Evaluating intervention tradeoffs
  - Understanding drivers
  - Anticipating evolution of situation
  - Understanding long-term behaviour
  - Assessing sensitivity to model assumptions
- Many uses are greatly strengthened when start model in state representative of the current situation
- Stronger stakeholder confidence & understanding of model results given alignment with empirical data

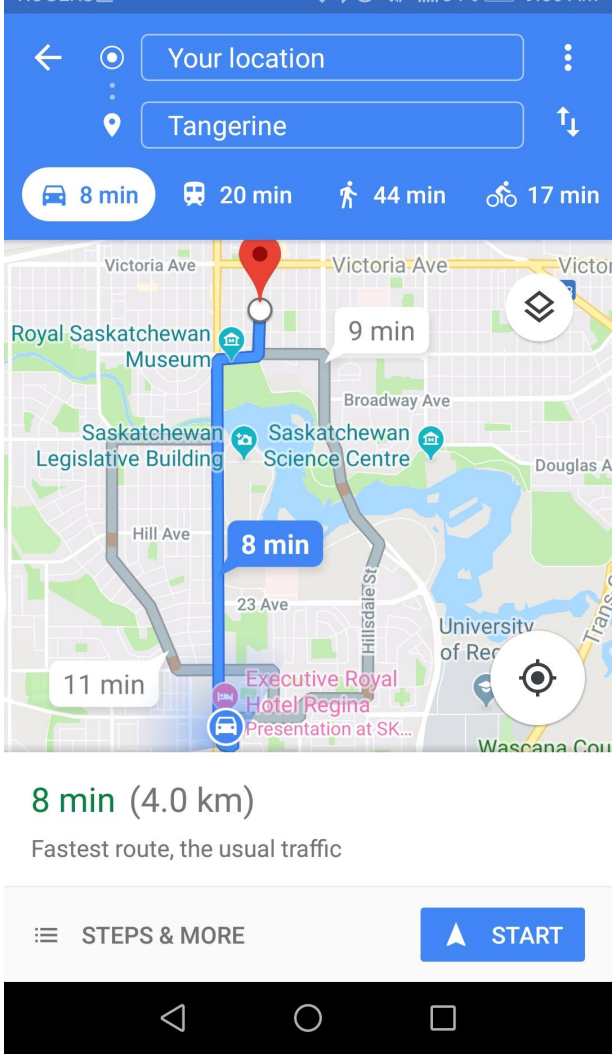
Reflection

- Models as approximations: Unassisted, *even the most sophisticated models* eventually diverge from empirical situation as time passes
  - Divergence between model state & empirical state
- Some relevant challenges: **Stochastics,** exogenous changes, approximations, omissions, heterogeneity ...
- Divergence can strongly affect perception of drivers, intervention tradeoffs, ability to project

As with weather models, communicable disease models offer greater utility when updated to reflect the latest evidence & use that to anticipate future state



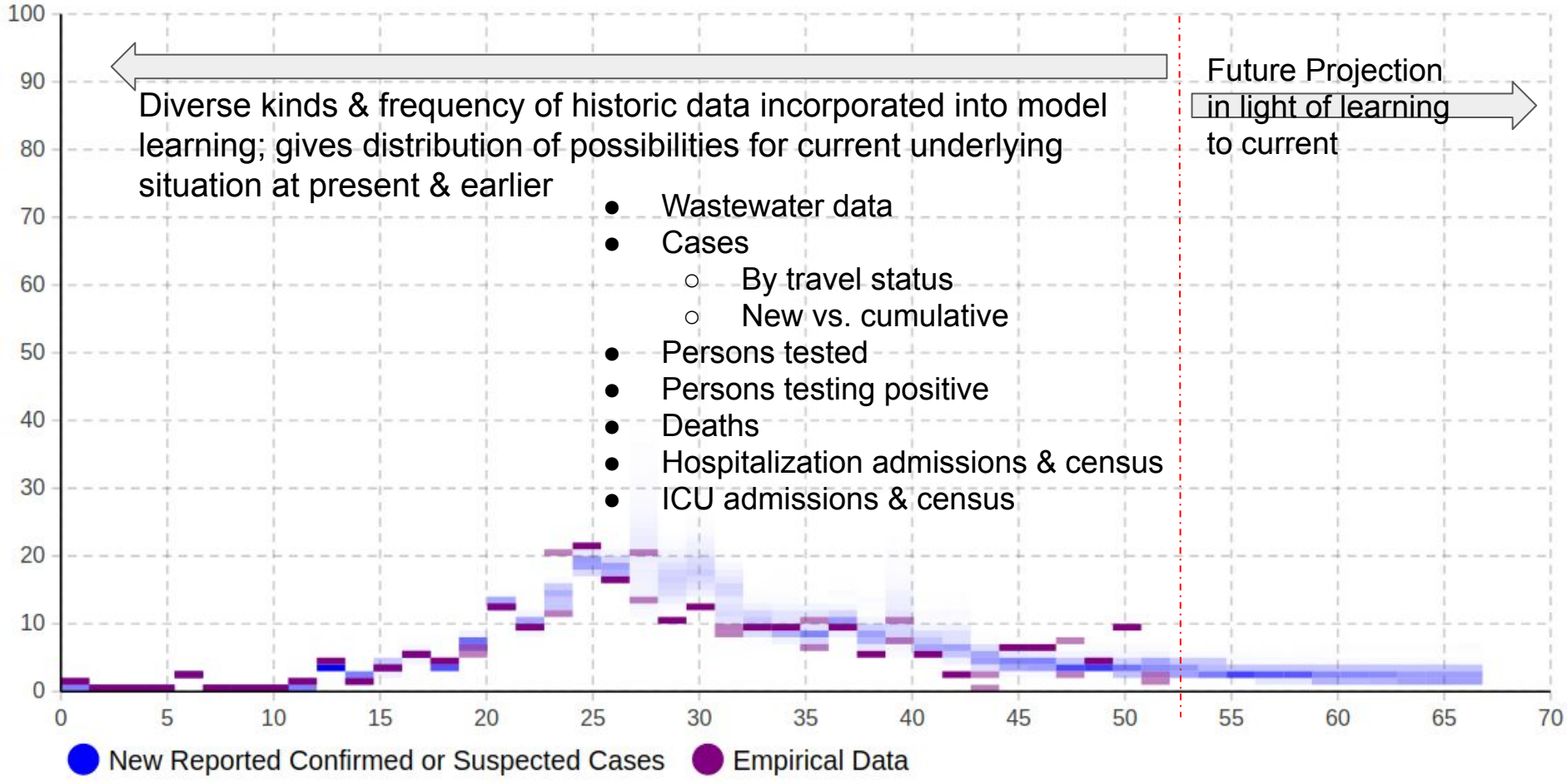
Like a GPS, we'd like recommendations tradeoffs b/t ways to reach our goals based on where we are in fact now -- not where we expected to be now



Agenda

- Motivation
- Particle Filtering with dynamic models
- Wastewater essentials
- Key elements of wastewater-specific model formulation
  - Likelihood
  - Forms of nexus between model state and wastewater concentration
- [As time allows] PMCMC
- Conclusions

Adaptive Planning -- Observing Unfolding Evidence



Bayesian Machine Learning & Dynamic Models

- **MCMC:** Sample from $p_M(\theta|y_{1:T})$: posteriors of ***deterministic*** dynamic model static parameters, latent states, scenario results, and incremental scenario gains.
- **Particle Filtering/SMC:** Sample from $p_{\theta,M}(x_{1:T}|y_{1:T})$: posteriors of ***stochastic*** dynamic model latent states stochastically evolving parameters, scenario results, and incremental scenario gains.
- **Particle MCMC (PMCMC):** Sample from $p_M(\theta,x_{1:T}|y_{1:T})$: posteriors of ***stochastic*** dynamic model latent states, stochastically evolving parameters, scenario results, and incremental scenario gains *and static parameters*.

Key Facts About how PF Works

The simulation model …

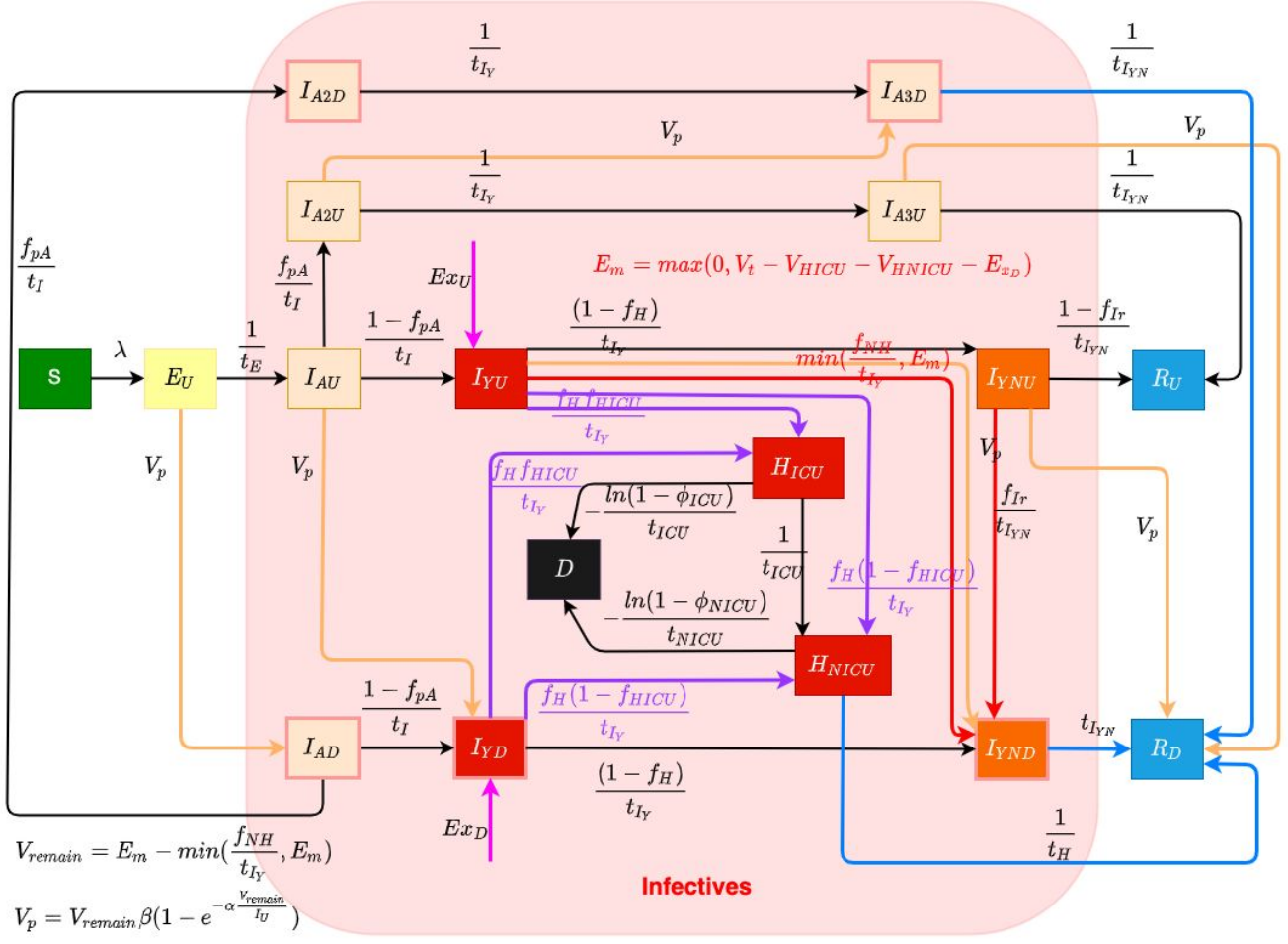
- Includes ***stochastic processes***
- Runs typically during prediction step between observation points
- Entire state "corrected" to align w/empirical data after observation

Is performed ***recursively***

- Rather than re-estimate the state over all time points *de novo* when new data comes in, the estimate when new data comes in depends on earlier estimates

Samples from the state (and trajectory) distribution

- **Each sample represented by "particle"** & contains a **complete hypothesis about state**
- Particles reflect "competing hypotheses" as to value of the current state
- Exploits **importance sampling**: distribution is sampled by associating samples from proposal dist (particles) w/**weights**
- There is a "**survival of the fittest**" of particles (hypotheses)



At any one time, each particle would hold a specific value for each state variable & is characterized by state vector

$$[S, E_U, E_D, I_{AU}, I_{AD}, I_{A2U}, I_{A2D}, I_{A3U}, I_{A3D}, I_{AU}, I_{AD}, I_{YU}, I_{YD}, I_{YNU}, I_{YND}, H_{ICU}, H_{NICU}, R_U, R_D, D, C_\beta, \alpha, f_{NH}, f_H]^T$$

How to Perform PF on Agg. Model in a Nutshell

Start with stochastic System Dynamics/Compartmental/ODE model

Subscript model by 100s to 1000s of particles

Each particle has its own full copy of model state (anything that could differ b/t realizations)

Sample from initial model state from prior distribution; set weights uniformly to 1

(Prediction phase): **Between observations**

All particles evolve according to standard model dynamics (just perform integration of each particle's state until next observation; all particles survive)

Particle weights remain invariant

(Update phase): **At observation points** (daily for SK data considered here): For each particle, multiply particle weight by likelihood of observing the empirical observation vector, given particle state

Resampling/"Survival of the fittest": If effective sample size is too low (too much disparity in weights) following observation, particles are resampled according to their weights, and weight is reset to 1

Particles with high weights reproduce; with low weights disappear

Trajectories can be sampled by maintaining *ancestry matrix holding lineages*

Resampling

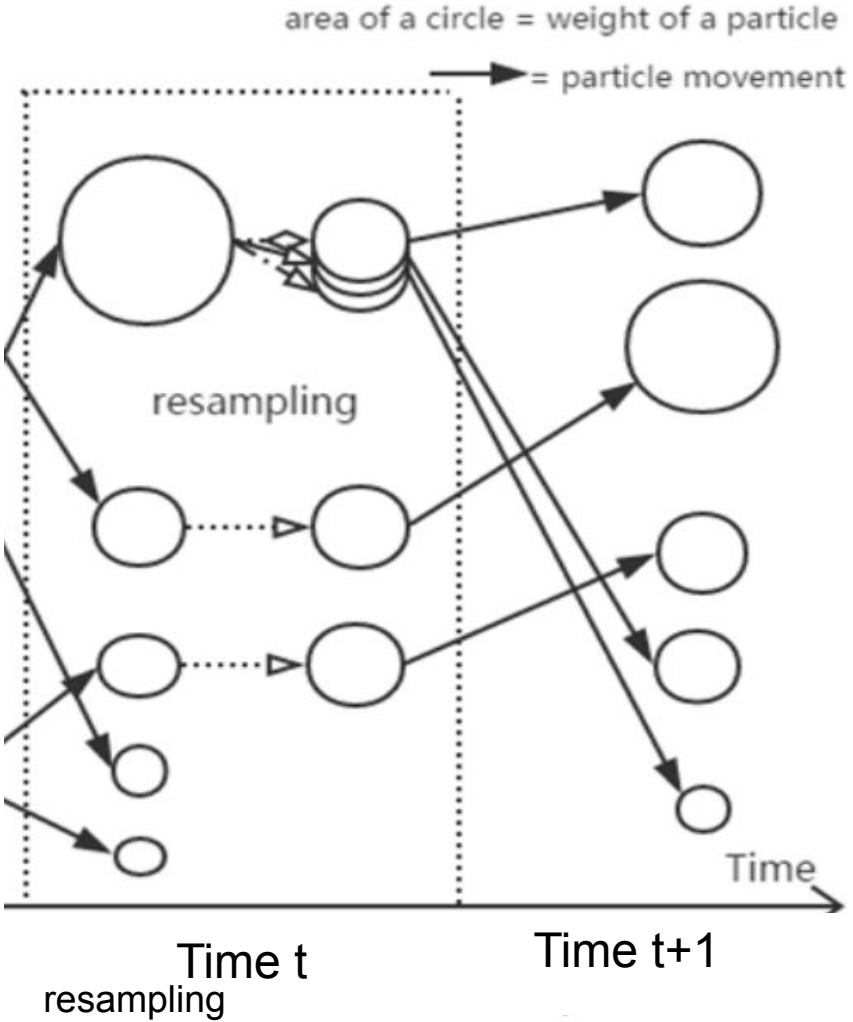


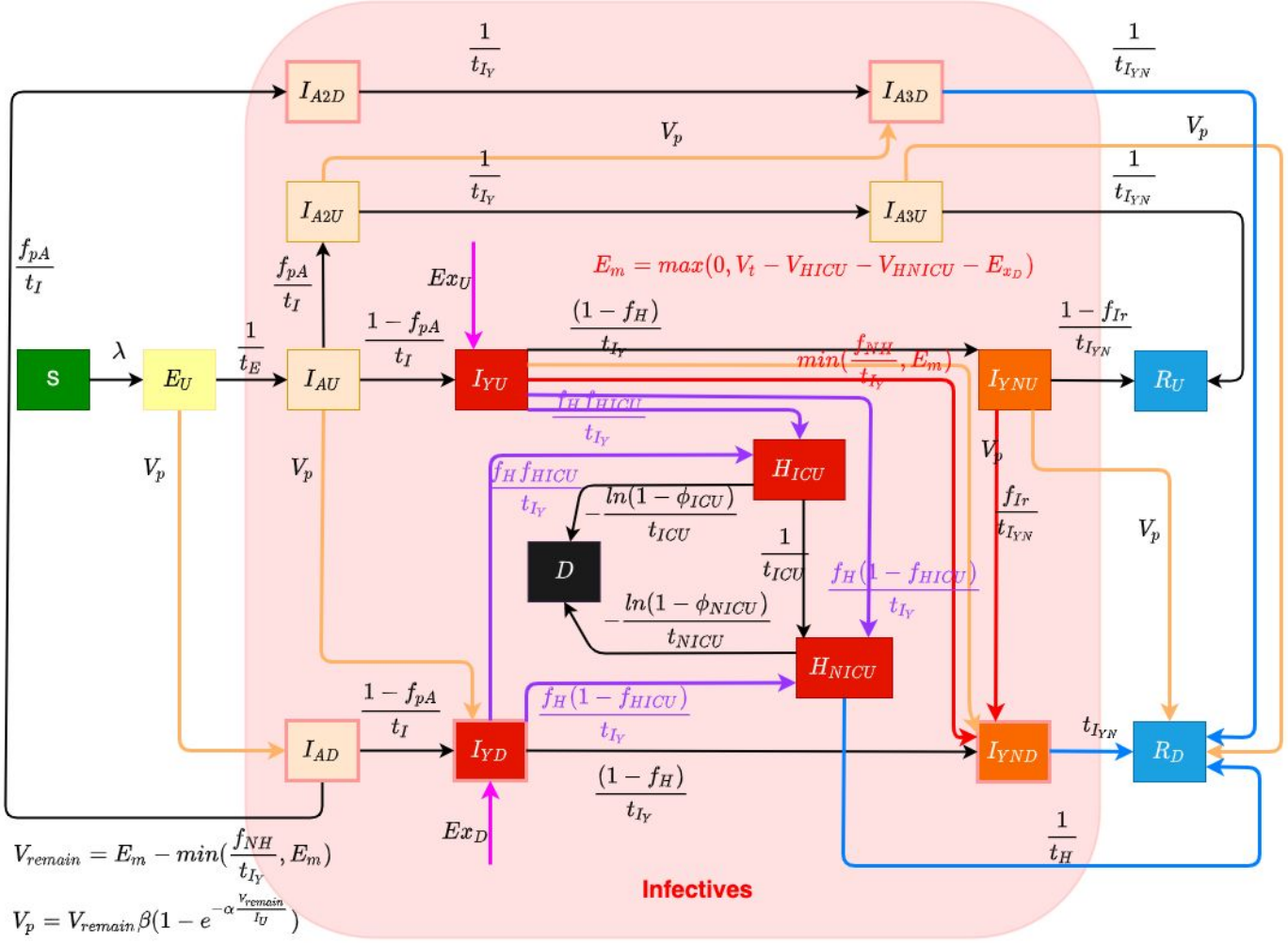
Image adapted from Xiaoyan Li M.Sc. Thesis

Particle Filtering Requirements for Compartmental Model

- This approach can be readily applied to broad classes of ODE models
- Each type of data used must have corresponding element(s) in the dynamic model to which that data can be compared in a likelihood function, e.g.,
  - ICU/Non-ICU Hospital census: One or more corresponding state variable each
  - Reported cases: Diagnosis flows/transitions
  - ICU/Non-ICU hospital admissions: Hospitalization flows/transitions
  - Deaths: Hospitalization flows/transitions
  - Hospital admissions: Hospitalization flows/transitions
  - **Wastewater concentrations: One or more state variables characterizing shedding population**

Example: CEPHIL's COVID Compartmental Model

- Separate treatment of travel- from non-travel diagnosed cases
- Data: case, test volume & positives, hospital and ICU census & admissions, deaths
- Test data as central to mediating active flows
- Hospitalization as a function of underlying later-stage infected cases -- not of reported case counts
- Parameter uncertainty with respect to key parameters
  - Contact
  - Hospitalization
  - Non-hospitalized case parameters
  - Testing parameter
- This model is applied daily in Saskatchewan, but also for all other Canadian provinces (for PHAC) and weekly for First Nations Reserves (for FNIHB)



The mathematical structure of the COVID-19 dynamic model employed in particle filtering

State Vector Associated with Each Particle

$$[S, E_U, E_D, I_{AU}, I_{AD}, I_{A2U}, I_{A2D}, I_{A3U}, I_{A3D}, I_{AU}, I_{AD}, I_{YU}, I_{YD}, I_{YNU}, I_{YND}, H_{ICU}, H_{NICU}, R_U, R_D, D, C_\beta, \alpha, f_{NH}, f_H]^T$$

Regular state variables of the System Dynamics/ODE/ Compartmental model

"Dynamic Parameters" whose (potentially transformed) values are engaged in random walks over time

Stochastically Evolving State Variables

The following variables exhibit **random walks** of variable, transformed to lie within a bounded range

- *c:* Contacts Per Day (cβ traditionally)
- f_H: Fraction of underlying infections that require hospitalization
- f_{NH}: Fraction of Incidence that is reported as cases
- α*:* Efficiency of active testing

Likelihood Functions Explored

Likelihood functions $p(y_t^M | x_t^N)$ give the likelihood of the empirical datum, given the particle state

Common distributions

- Binomial
- Negative Binomial/Pascal
- Poisson
- Normal
- Lognormal

Form of Likelihood Function 2: Dealing with Multiple Datasets

- Form when all data is present:

$$\mathcal{L} = \mathcal{L}_{NewReportedEndogenousCases} \times \mathcal{L}_{CumulativeReportedEndogenousCases} \times \mathcal{L}_{CumulativeICUAdmissioncases} \times \mathcal{L}_{CumulativeNICUAdmissioncases} \times \mathcal{L}_{ICUCensusCases} \times \mathcal{L}_{NICUCensusCases} \times \mathcal{L}_{CumulativeDeathCases} \times \mathcal{L}_{ViralConcentration}$$

- This model update step occurs *daily* for SK
- When any type of empirical data above is absent, that likelihood term is omitted from the calculation of the overall likelihood $\mathcal{L}$
- Wastewater is incorporated in the final term for those episodic days in which it is available

Form of Likelihood Function 2: Dealing with Multiple Datasets

Two examples of (negative binomial) sub-likelihood functions:

$$\frac{i_{et}}{i_{et}+r_{er}}$$

1. The sub-likelihood function of only considering new reported COVID cases:

$$\mathcal{L}_{NewReportedEndogenousCases} = \binom{y_{et}+r_{er}-1}{r_{er}-1} {p_{er}}^{r_{er}} (1-p_{er})^{y_{et}}$$

$$\frac{H_{ICU}}{H_{ICU}+r_{ICU}}$$

2. The sub-likelihood function of only considering new reported COVID cases:

$$\mathcal{L}_{ICUCensusCases} = \binom{y_{ICU}+r_{ICU}-1}{r_{ICU}-1} {p_{ICU}}^{r_{ICU}} (1-p_{ICU})^{y_{ICU}}$$

Recall: **How to Perform PF on Agg. Model in a Nutshell**

Start with stochastic System Dynamics/Compartmental/ODE model

Subscript model by 1000s or more particles (COVID-19 provincial models: 150,000 particles is standard)

Each particle has its own **full copy of model state** (anything that could differ b/t realizations)

Sample from initial model state from prior distribution; set weights uniformly to 1

(Prediction phase): **Between observations**

All particles evolve according to standard model dynamics (just perform integration of each particle's state until next observation; all particles survive)

Particle weights remain invariant

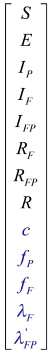
(Update phase): **At observation points**: For each particle, multiply particle weight by likelihood of observing the empirical observation vector given that particle's state

Resampling/"Survival of the fittest": If effective sample size is too low (too much disparity in weights) following observation, particles are resampled according to their weights, and weight is reset to 1

Particles with high weights reproduce; those with low weights tend to disappear

Trajectories can be sampled by maintaining *ancestry matrix holding lineages*

Example Scenario For day t



Particles

	P1	P2	P3	…
Prior Weight	0.2	0.17	0.01	
Particle-Specific hypothesis about State	902 31 42 220 421 121 157 106 0.02 0.1 0.5 0.2 0.5	495 22 52 523 221 272 351 64 0.05 0.1 0.05 0.2 0.5	992 5 2 1 0 0 0 0 0.001 0.5 0.5 0.2 0.5	
Particle calculated: Endo. case presentations	12	15	2	
Hospital census	20	2	0.5	
Likelihoods of Observing Empirical data	0.22	0.5	0.001	
	0.75	0.15	0.03	
Composite Likelihood	0.165	0.075	0.00003	
Updated unnormalized weights	0.033	0.01275	3E-07	
Updated weights	0.066 902 31 42	0.0255 495 22 52	6E-07 992 5 2	

From model flows, taking into account test volumes

Empirical COVID-19 observations y_t
For day t:

Endogen. presenting cases	10
Hospital Census	15

Weight Normalization

- **Resampling could occur here**

Recall: **How to Perform PF on Agg. Model in a Nutshell**

Start with stochastic System Dynamics/Compartmental/ODE model

Subscript model by 1000s or more particles (COVID-19 provincial models: 150,000 particles is standard)

Each particle has its own **full copy of model state** (anything that could differ b/t realizations)

Sample from initial model state from prior distribution; set weights uniformly to 1

(Prediction phase): **Between observations**

All particles evolve according to standard model dynamics (just perform integration of each particle's state until next observation; all particles survive)

Particle weights remain invariant

(Update phase): **At observation points**: For each particle, multiply particle weight by likelihood of observing the empirical observation vector given that particle's state

Resampling/"Survival of the fittest": If effective sample size is too low (too much disparity in weights) following observation, particles are resampled according to their weights, and weight is reset to 1

Particles with high weights reproduce; those with low weights tend to disappear

Trajectories can be sampled by maintaining *ancestry matrix holding lineages*

Resampling

Triggered after weight updates & normalization
if effective sample size lies below a certain threshold
(or a certain fraction of the total number of particles)

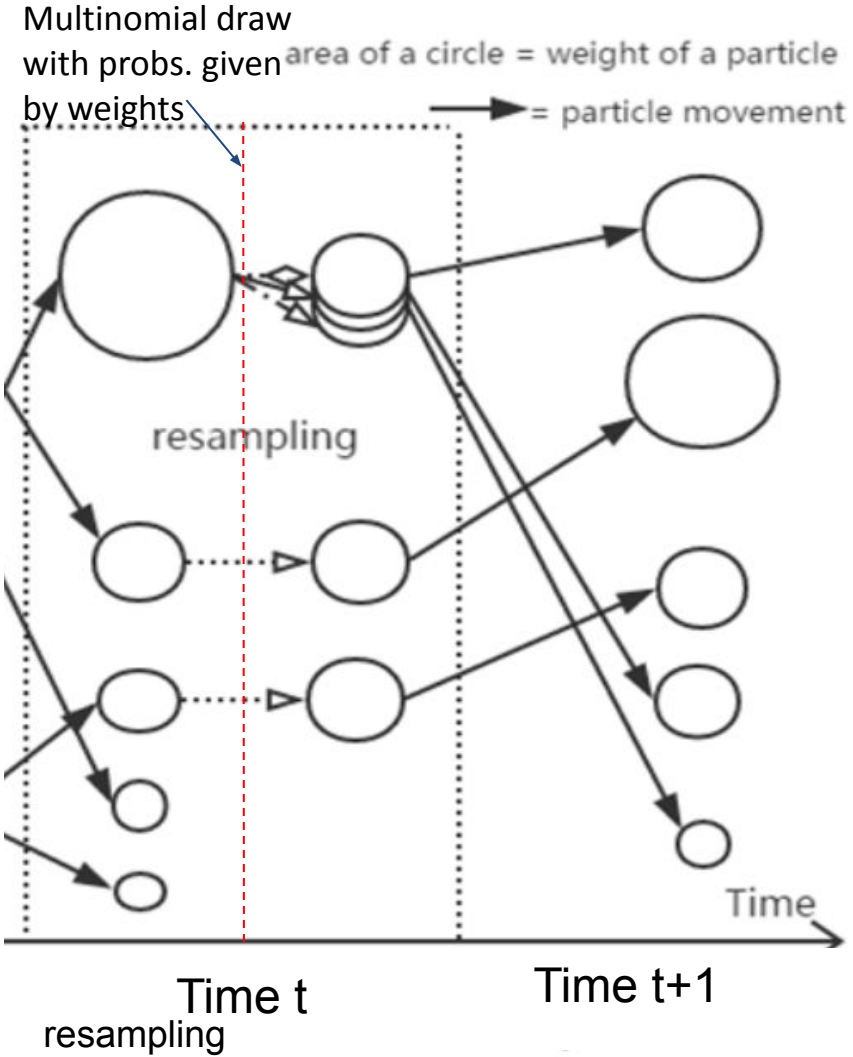


Image adapted from Xiaoyan Li M.Sc. Thesis

Agenda

- Motivation
- Particle Filtering with dynamic models
- Wastewater essentials
- Key elements of wastewater-specific model formulation
  - Likelihood
  - Forms of nexus between model state and wastewater concentration
- [As time allows] PMCMC
- Conclusions

Wastewater Surveillance Essentials

- Living and working locations produce wastewater from plumbing (sinks, dishwashers, toilets, showers, bathtubs, washing machines, industrial processes, ...)
- Canadian municipalities with populations in the 1000s typically have centralized wastewater infrastructures (Cf. leach fields, septic tanks, etc.)
- Wastewater sampling is performed upstream of many levels of treatment
- Even central municipal wastewater systems differ strongly in size, design
    - Stormwater systems can be either combined with wastewater systems (e.g., in older cities) or separate
    - Wastewater takes time to flow to treatment processes
- Wastewater surveillance for some communicable diseases (including COVID-19) makes heavy use of fecal indicators
- Fecal shedding can be highly variable across the natural history of infection
- Active sampling seek to sample effluent streams or accumulations
- Passive sampling can sample from plumbing pipes within homes

Wastewater Tradeoffs

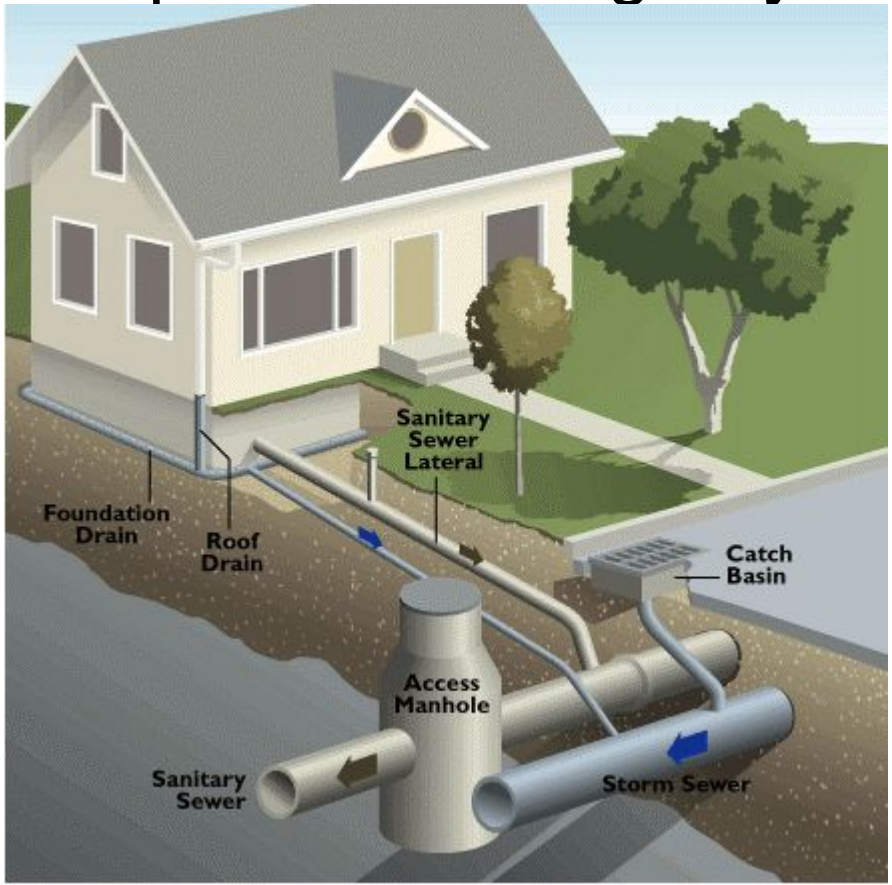
Advantages

- Low actual & perceived privacy intrusion
- Broad to narrow geographic coverage
- Independence of care-seeking behaviour
- Less logistical effort than individual testing
- Reaching those who would not opt in to testing
- For some populations, safer than testing
- Versatility: Capacity to examine syndemics, behavioural dynamics, cortisol levels, substance & pharmaceutical use, antimicrobial resistance, etc.

Disadvantages

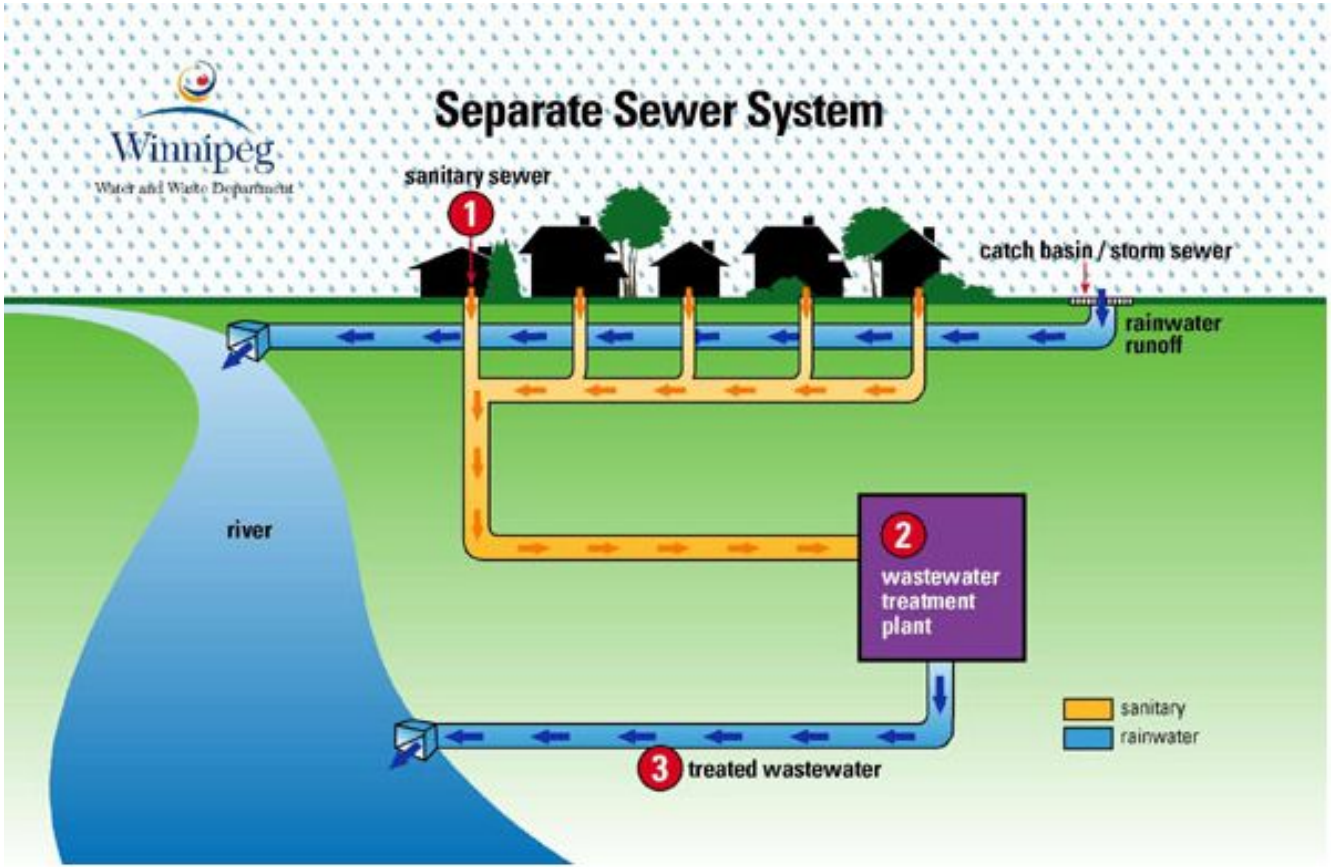
- Don’t identify ill individuals
- Uncertainties on shedding profiles for conventional lineage & variants
- In some infrastructures, difficulty of securing reliable signal
- Signal interpretation

Separated Sewage Systems



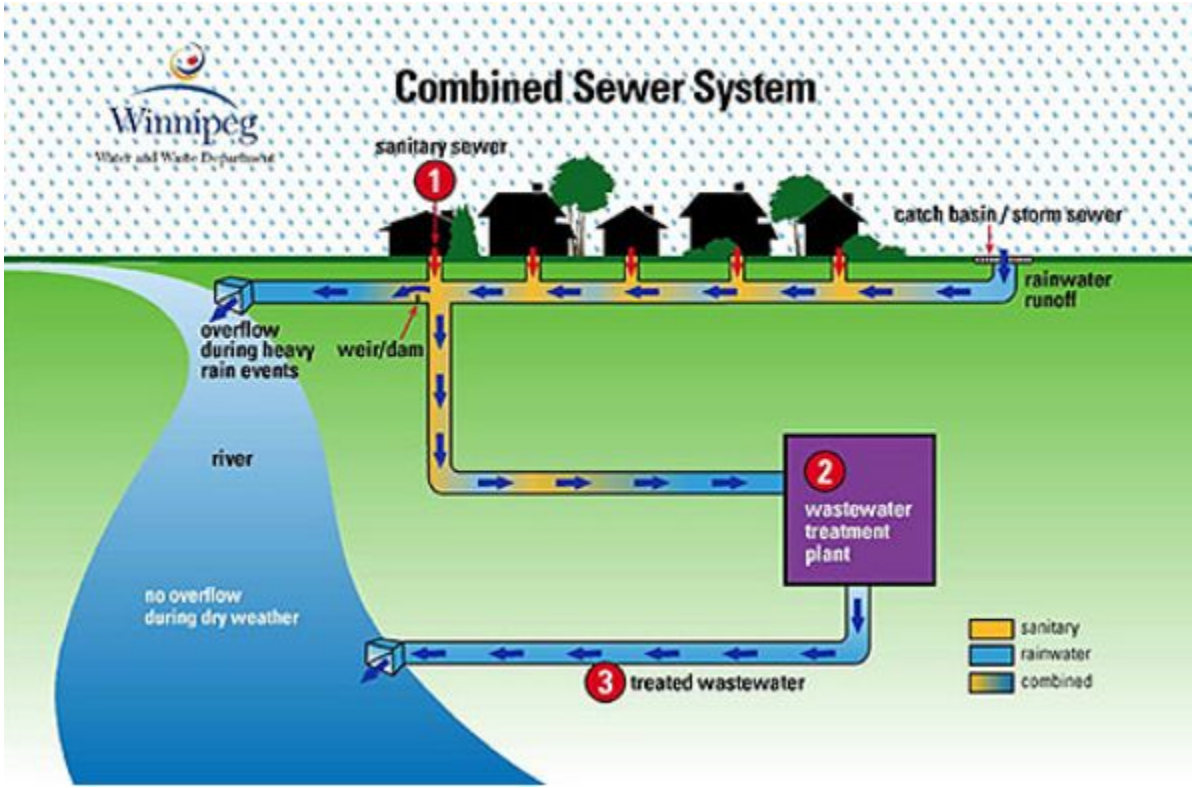
Source: City of Noblesville, IN. *Sewer Collection System*. https://www.cityofnoblesville.org/egov/images/1477437946_34112_o.png

Separate Wastewater & Stormwater Systems



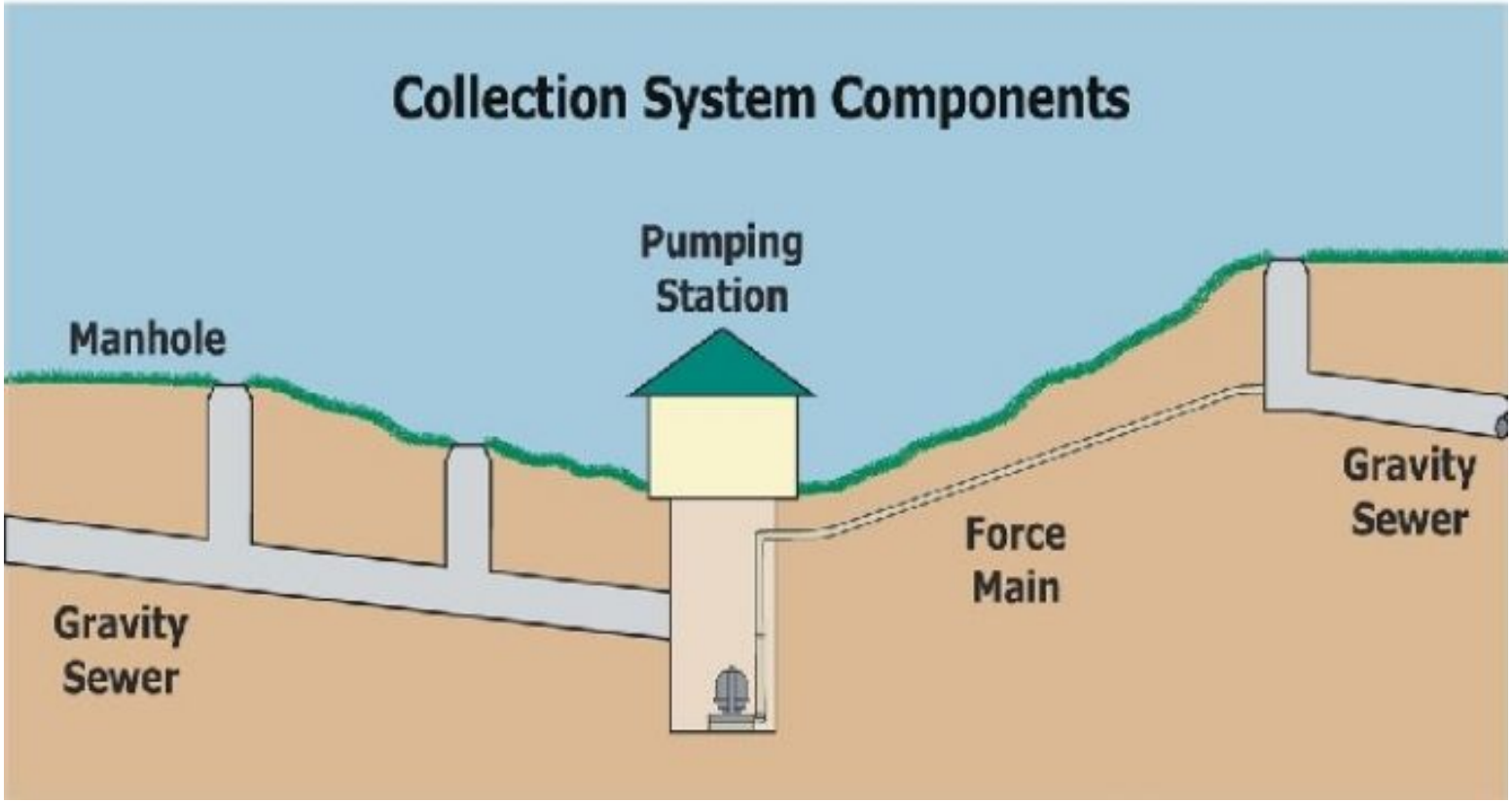
Source: Winnipeg *Water and Waste Department*. https://www.winnipeg.ca/waterandwaste/images/sewage/separate.jpg

Combined Wastewater & Stormwater System



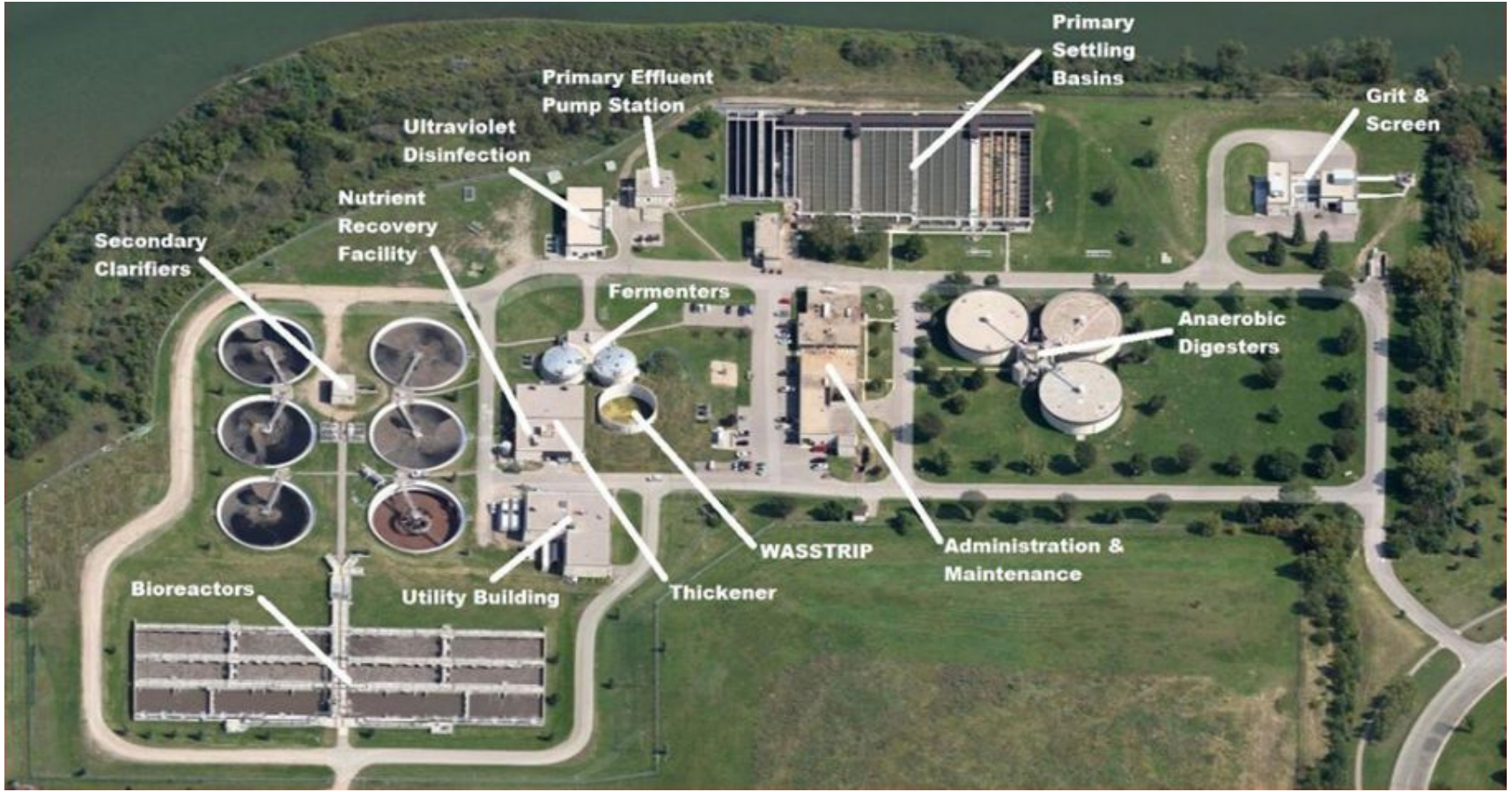
Source: Winnipeg *Water and Waste Department*. https://www.winnipeg.ca/waterandwaste/images/sewage/combined.jpg

Wastewater Pump Stations



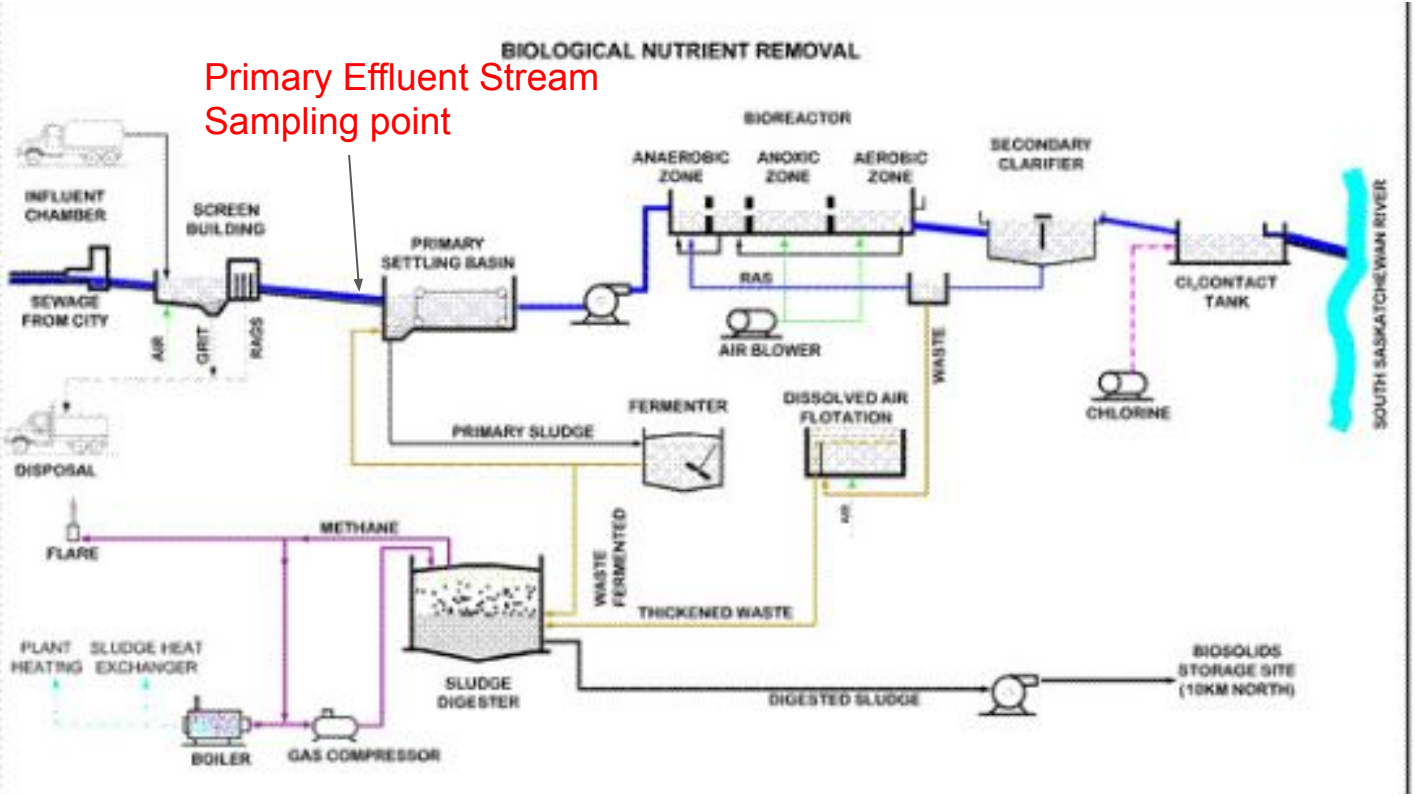
Pennsylvania. Dept. Env. Protection, Office of Water Management. Module 1 Introduction to Wastewater Treatment Plant Operator https://slidetodoc.com/presentation_image_h/852709aaa90086b4f126aa67a8756d4c/image-32.jpg

Example Wastewater Treatment Plant



Source:https://www.saskatoon.ca/sites/default/files/images/wastewater_treatment_plant_labelled_areas_for_web_page_-_2014.jpg

Sampling Points



Source: https://www.saskatoon.ca/sites/default/files/images/biological_nutrient_removal.gif

Multiple Forms of Wastewater Assays

Quantitation

- This approach directly supporting estimation of a key latent variable: The size of the SARS-CoV-2 shedding population
- While valuable, lower sensitivity
- Currently addressed with Bayesian methods

Detection

- Higher sensitivity
- Valuable for new variants, appearance of low prevalence variant in vulnerable communities)
- Can be conducted in parallel with broader SARS-CoV-2 quantitation for same region
- Also addressable with Bayesian methods

Key Elements of Wastewater Context Important for Modeling: Infrastructure

- Is there a municipal wastewater treatment plan?
- Is the stormwater sewer system combined with wastewater sewer system?
- What are the lengths of time from plumbing leaves the source to the wastewater plant? Does this differ across the city?
- Does the wastewater data accumulate in certain locations?
- Is sampling conducted from an effluent stream, or an accumulation?
- How well is the wastewater protected from external temperatures, water & salts ?
- Does the catchment area involve any acute care centres?
- What is the shedding profile for the natural course of infection?
- What is the sampling regime during the day?
- Is the goal to sample at the domicile, facility-specific level, neighbourhood level, regions within a municipality, or municipal level?

Key Elements of Wastewater Context Important for Modeling: Sampling & Beyond

- Active sampling: How frequently/after what flow volumes are the samples taken (e.g., do they account for diurnal variation?)
- Passive sampling: Over what period of time are the samples accumulating?
- How quickly are the samples refrigerated? Transported to the lab?
- Are samples normalized in the laboratory?
- Is the wastewater using solid phase or liquid phase samples?
- What are the goals of sampling: Detection or quantitation?
- What normalization procedures are followed? What comparator concentrations are referenced?

Geographic Resolution

- The geographic resolution of sampling affects
  - Logistics & cost
  - Actionability of results
  - Sampling method (e.g., passive vs. active)
  - Intervention options
  - Alerting options
  - Choice of quantitation vs. detection

Agenda

- Motivation
- Particle Filtering with dynamic models
- Wastewater essentials
- Key elements of wastewater-specific model formulation
  - Likelihood
  - Forms of nexus between model state and wastewater concentration
- [As time allows] PMCMC
- Conclusions

Wastewater Data Examined Here: Infrastructure Context

- All wastewater & health data considered here is specifically from Saskatoon
- Saskatoon has almost entirely separate wastewater & stormwater infrastructures
  - Daily flow is not affected by precipitation, first flush or snowmelt
  - Minimal risk of pH&salinty changes
  - Small amounts of infrastructural coupling could apply during extreme weather events not experienced during the sampling timeframe
- Wastewater data is from municipal-wide primary effluent wastewater (~80ML/Day)
- To normalize with respect to within-day effects with respect to effluent flow, autosampling is conducted throughout day from a moving incoming effluent stream
- U of S Toxicology Centre estimates a transit time from homes to the wastewater plant primary effluent of below 12 hours & possibly half that ⇒ less degradation
- Wastewater transition to the wastewater treatment plant is made by sewer several meters below ground, which preserve temperature in the 12-18C range

Wastewater Data Examined Here: Sampling Chronology

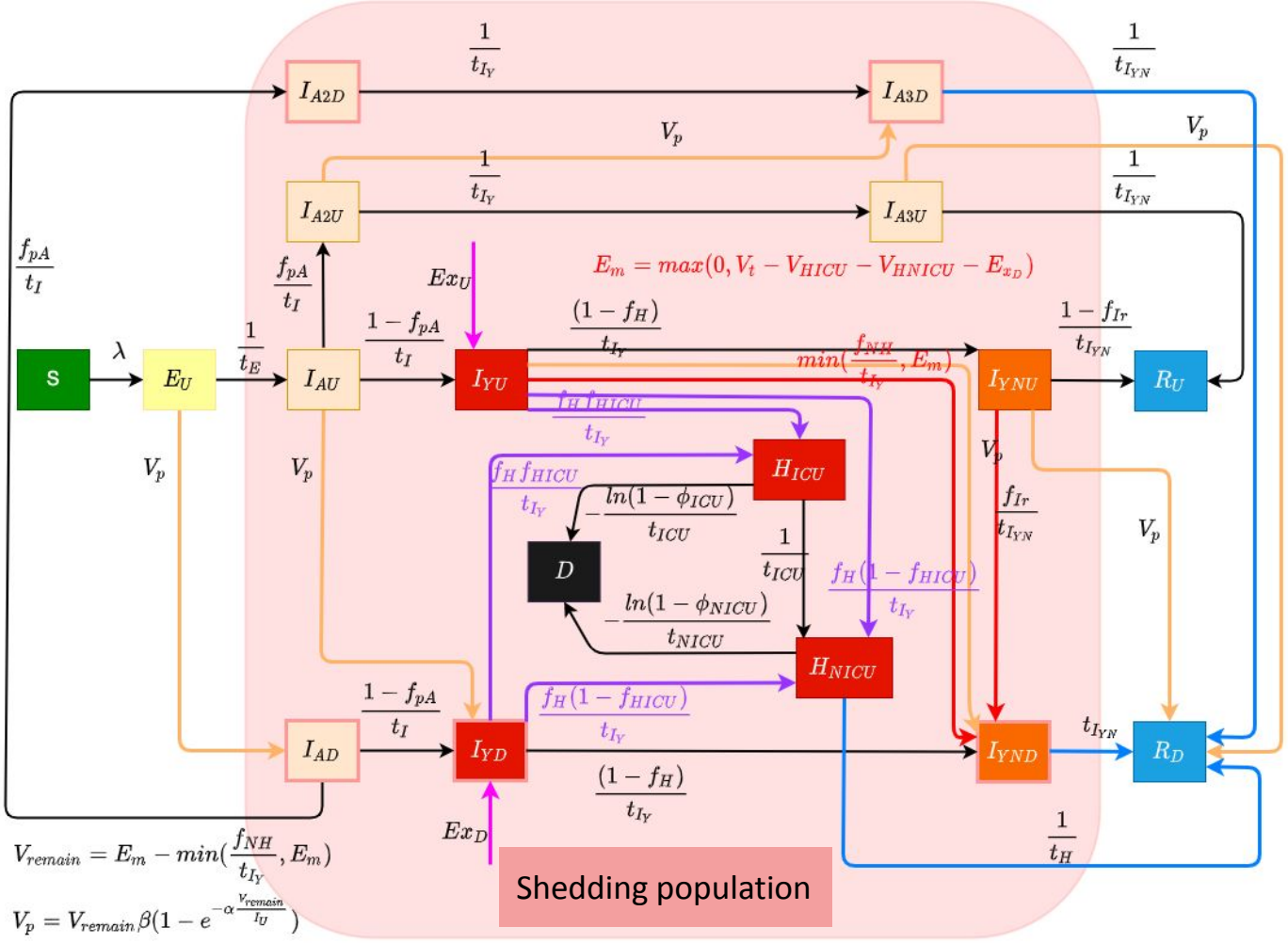
- 58 total wastewater samples were reported
- Sampling was conducted
  - Singleton July 2020 sample
  - Sept. 2020 - Jan. 2021: Sampling concentrated on and after October 21, Regular 3 days/week reporting for November 11, 2020 onwards
  - Sampling hiatus Jan 14-March 5, 2021 (inclusive) due to lack of funding
  - Regular 3 days sampling resumed March 6, 2021

Wastewater Data Examined Here: Sampled Storage & Laboratory Context

- Wastewater plant freezes samples & stores throughout week
- Samples are shipped to Toxicology centre at end of sampling week
- Laboratory samples from a given week are analyzed in a block
- Tox. Centre provides efficiency-adjusted viral concentration estimates (N2 copies/100 mL)
- Results are reported following normalization with respect to efficiency of RNA extraction, qPCR (judged via a reference surrogate armed virus) and recovery ratio

Key Parameter: Ratio Between Reported Normalized Viral Concentration & Shedding Population (γ)

- Based on feedback from laboratory partners, we assumed a linear scaling relationship between shedding population & reported viral concentrations
- The scaling term is γ
- We sought to bound possible values for coefficient γ based on estimates of shedding population and empirical concentration values
- Wastewater infrastructure will affect the extent of other model structures required



The mathematical structure of the COVID-19 dynamic model employed in particle filtering

Form of Likelihood Function 2: Dealing with Multiple Datasets

Two examples of the sub-likelihood function:

1. The sub-likelihood function only considering new reported COVID ca $\frac{i_{et}}{i_{et}+r_{er}}$

$$\mathcal{L}_{NewReportedEndogenousCases} = \binom{y_{et}+r_{er}-1}{r_{er}-1} {p_{er}}^{r_{er}} (1-p_{er})^{y_{et}}$$

2. The sub-likelihood function only considering wastewater data (when available)

$$\mathcal{L}_{ViralConcentration} = \binom{y_{VC}+r_{VC}-1}{r_{VC}-1} {p_{VC}}^{r_{VC}} (1-p_{VC})^{y_{VC}}$$ where

$$p_{VC} = \frac{x_{VC}}{x_{VC}+r_{VC}}$$

$$x_{VC} = \gamma(I_{AU} + I_{AD} + I_{A2U} + I_{A2D} + I_{A3U} + I_{A3D} + I_{YU} + I_{YD} + I_{YNU} + I_{YND})$$

Model Structural Elements to Consider for Wastewater

- Delayed relationship between shedding population & WW concentration
- Accumulations: Basing WW concentration on smoothing of shedding population over multiple days
- Region-specific delays & dynamics
- Invariant vs. time-varying linear relationship between shedding population & viral concentration: e.g., based on precipitation, occurrence of snowmelt

Uses of the Particle Filtered Model

- **Population tomography:** Providing a consensus portrait (via a joint distribution) of the situation now and in the past
- **Projection/Forecasting**: Projection forward from **now** with model dynamics and “status quo” or diffusive assumptions concerning active testing, contact patterns, etc.
- **Backcasting**: Historical reconstruction based on earlier & later data
- **“What if” questions**: Evaluation of intervention portfolios or other “what if” scenarios from today

Uses of the Particle Filtered Model

- **Population tomography:** Providing a consensus portrait (via a joint distribution) of the situation now and in the past
- **Projection/Forecasting**: Projection forward from **now** with model dynamics and "status quo" or diffusive assumptions concerning active testing, contact patterns, etc.
- **Backcasting**: Historical reconstruction based on earlier & later data
- **"What if" questions**: Evaluation of intervention portfolios or other "what if" scenarios from today

Result: Population Tomography

Uses of the Particle Filtered Model

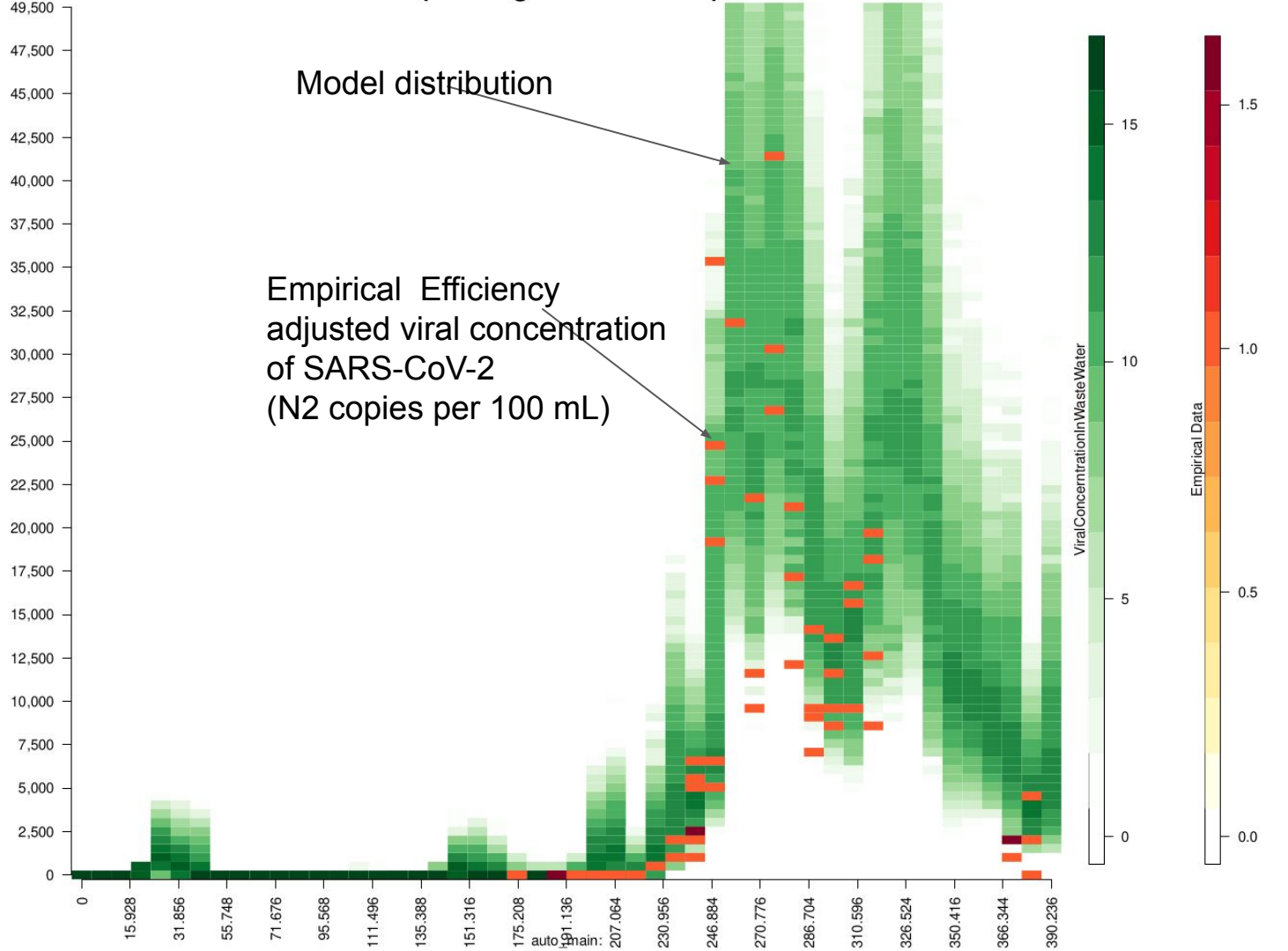
- **Population tomography:** Providing a consensus portrait of the situation now and in the past
- **Projection/Forecasting**: Projection forward from **now** with model dynamics and "status quo" or diffusive assumptions concerning active testing, contact patterns, etc. Requires
  - Specification of active test volumes
  - Treatment of uncertainty
- **Backcasting**: Historical reconstruction based on earlier & later data
- **"What if" questions**: Evaluation of intervention portfolios or other "what if" scenarios from today

Findings

Summary

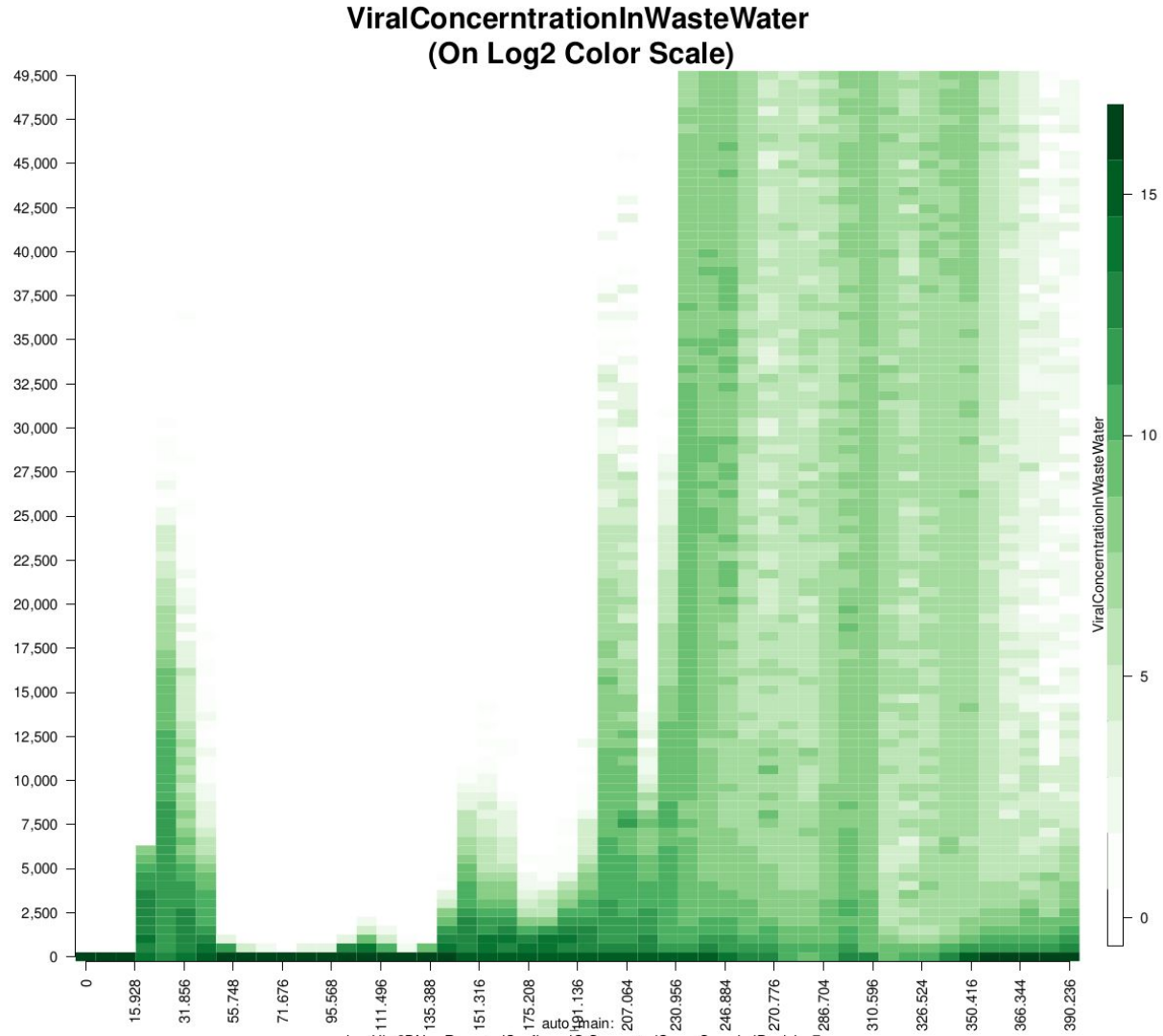
- Inclusion of even modest, episodic WW viral concentration data in PF allows
  - Effective PF estimation of latent state accounting for health system & WW data
  - Altered (likely sharpened) model estimates of health-system related data
- Results clearly suggest that WW data may support better matching against health system-derived empirical data
- Large particle complements (~150K particles) are recommended for consistency in results
- Given efficiency adjusted concentration data and separate WW & stormwater sewers, it appears adequate to characterizing a temporally invariant relationship between wastewater viral concentration & count of infectives
- Simplicity of this relationship emphasizes the attractiveness the PMCMC

Discrepancy from Wastewater Data: Considering Wastewater in Likelihood with Constant coefficient γ (Method B)



Discrepancy from Wastewater Data: **Not** Considering Wastewater in Likelihood

This distribution is not informed by WW data, and is at some distance from actual wastewater data seen in prev. slide.



Impact of Considering Wastewater on Saskatoon-Specific Estimates
150K Particle Results

Model Scenario	Effective Reproductive Number (ERN, R*)	Estimate of Count of Undiagnosed Infectives	Count of Daily New Infections	Force of infection
No wastewater data considered in likelihood	0.95 (0.15, 3.03)	996 (595, 1878)	81.9 (8.9, 431.6)	0.00028 (3e-05, 0.00147)
Wastewater data with constant coefficient relating concentration & infectives	1.4 (0.28, 3.61)	1248 (612, 2426)	152 (18, 671)	0.00051 (5.9e-05, 0.0022)

Evaluation Options Explored

- Posterior discrepancy -- compare posterior predicted and empirical observables
- Temporal cross-validation -- Projection (Train in one period of time, test evolution on on another)

Evaluation Options Explored

- Posterior discrepancy -- compare posterior predicted and empirical observables
- Temporal cross-validation -- Projection (Train in one period of time, test evolution on on another)

Assessing Posterior Discrepancy Between Model Predictions & Empirical Health System Data

- We sought to assess how the presence of wastewater data affects particle filtering performance, through comparison of model results with
  - Saskatoon-specific Empirical Health System Data
  - Wastewater data (where possible)
- 3 Sets of Scenarios were conducted
  - **Scenario Set 1**: Assessing impact of no use of wastewater data vs. two different ways of integrating wastewater data via handling of scaling coefficient γ
    - No use of wastewater data: 1 realization
    - Constant coefficient γ: 2 realizations
    - Time varying coefficient γ: 2 realizations
    - 75K Particles
  - **Scenario Set 2**: Assessing impact of no use of wastewater data vs. constant scaling coefficient γ
    - 2 realizations each
    - 75K Particles
  - **Scenario Set 3**: Assessing impact of no use of wastewater data vs. constant scaling coefficient γ
    - 1 realizations each
    - 150K Particles
- It is of key importance to realize that posterior discrepancy is calculated over entire time period of health system data, but WW data is only available for a small fraction of that time

Posterior Discrepancy from All Data
40K Particles, Time horizon Feb 23, 2020 - Mar 26, 2021

Model Scenario within Scenario Set 1	**Discrepancy from Health System Data over entire model time horizon**	**Discrepancy from both Health System & Wastewater Data over entire model time horizon**
No wastewater data considered in likelihood	3.73	N/A (No discrepancy from WW data assessed)
Wastewater data with temporally varying coefficient relating concentration & infectives	3.70	3.82
Wastewater data with constant coefficient relating concentration & infectives	3.74	3.82

Posterior Discrepancy from Empirical Health System Data

75K & 150K Particles; Time horizon Feb 23, 2020 - Mar 31, 2021

Model Scenario	**75K Particles 2 Realizations Each Scenario Set 2**	**150K Particles 1 Realization Each Scenario Set 3**
No wastewater data considered in likelihood	3.713	3.722
Wastewater data with constant coefficient relating concentration & infectives	3.710	3.694

Impact of Wastewater Data Frequency & Smoothing on **Posterior** Discrepancy

Model Scenario within Scenario Set 1	**Discrepancy from Health System Data over entire model time horizon**
No wastewater data considered in likelihood	
Wastewater data sampled 1 day per week	3.79
Wastewater data sampled 3 days per week	3.85
Wastewater data sampled 3 days per week & averaged	3.78

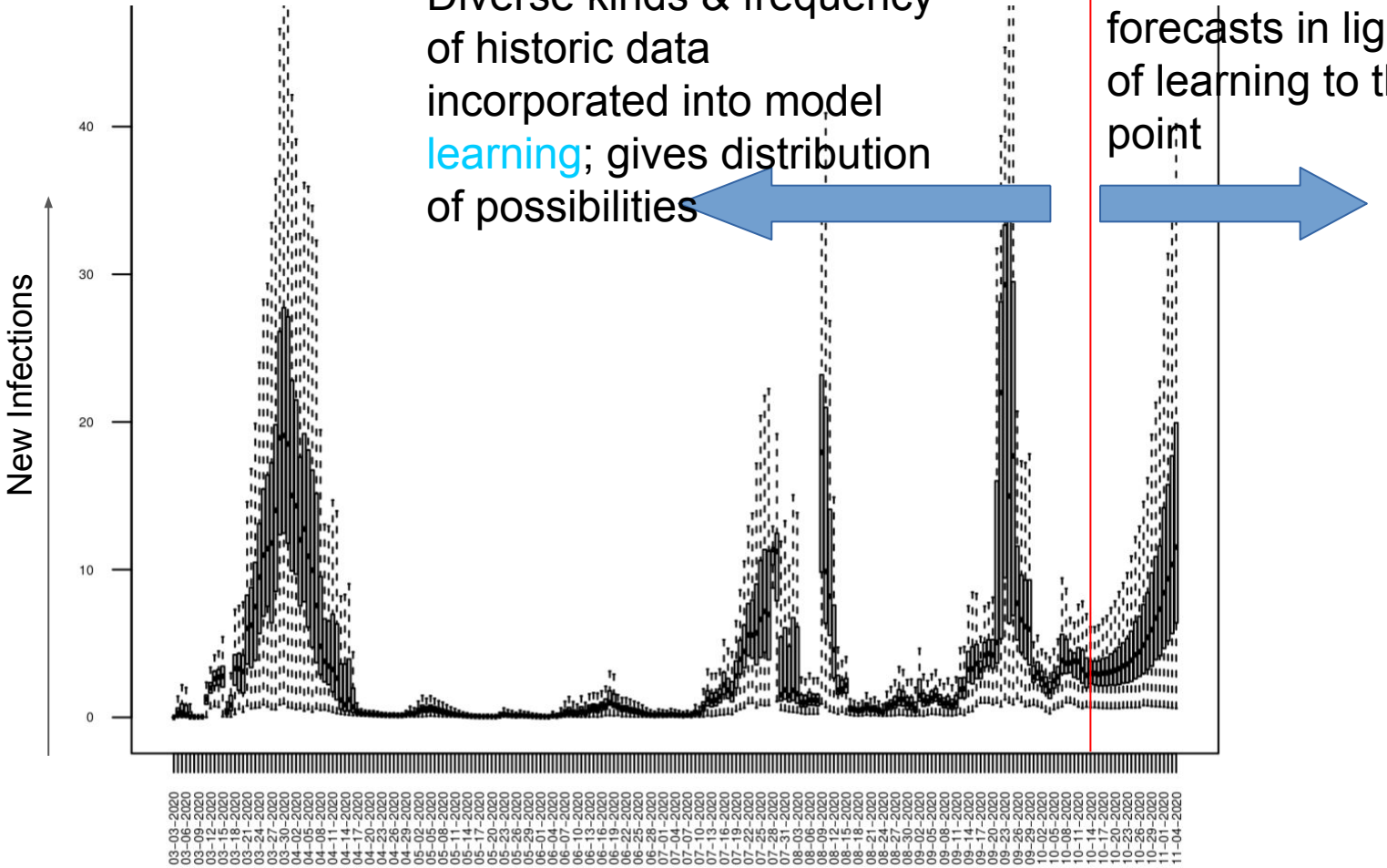
Evaluation Options Explored

- Posterior discrepancy -- compare posterior predicted and empirical observables
- Projection (Train in one period of time, test evolution on on another)

Projection Discrepancies

Diverse kinds & frequency of historic data incorporated into model learning; gives distribution of possibilities

Model forecasts in light of learning to this point



Projection Discrepancy from Empirical Health System Data

150K Particles; health system data time horizon Feb 23, 2020 - Mar 24, 2021; projection period March 25, 2021 - April 8, 2021

Model Scenario	**2 Realizations of 150,000 particles for each base scenario, with each realization being extended with 10 projection realizations utilizing 1000 particles**
No wastewater data considered in likelihood	1.663
Wastewater data with constant coefficient relating concentration & infectives	1.596

Mann-Whitney one-way U-Test
exact p-value: 0.04296

The consideration of wastewater data in the likelihood employed during particle filtering yields a statistically significant reduction in **projection discrepancy** of the particle filtered estimates of current state and the empirical data during the projection time period.

Agenda

- Motivation
- Particle Filtering with dynamic models
- Wastewater essentials
- Key elements of wastewater-specific model formulation
  - Likelihood
  - Forms of nexus between model state and wastewater concentration
- [As time allows] PMCMC
- Conclusions

PMCMC-Based Analytics for Wastewater

Bayesian Machine Learning & Dynamic Models

- **MCMC:** Sample from $p_M(\theta|y_{1:T})$: posteriors of ***deterministic*** dynamic model static parameters, latent states, scenario results, and incremental scenario gains.
- **Particle Filtering/SMC:** Sample from $p_{\theta,M}(x_{1:T}|y_{1:T})$: posteriors of ***stochastic*** dynamic model latent states stochastically evolving parameters, scenario results, and incremental scenario gains.
- **Particle MCMC (PMCMC):** Sample from $p_M(\theta,x_{1:T}|y_{1:T})$: posteriors of ***stochastic*** dynamic model latent states, stochastically evolving parameters, scenario results, and incremental scenario gains *and static parameters*.

PMCMC Basics

- Supports estimating (via sampling from) **joint** distributions of
  - Parameters (static and evolving)
  - System state over time
- High computational expense
  - 10,000+ MCMC iteration (burn-in period required)
  - Each MCMC iteration requires running particle filtering to sample *trajectory* of latent states
  - Each particle in particle filtering must run model
  - Often seek to run multiple MCMC walkers (chains)
  - Silver lining: Highly parallelizable
- MCMC is actually a *family* of algorithms

Family of Algorithms

- Presented here: Particle Marginal PMCMC
- Particle Gibbs PMCMC
  - more controversial
- Particle independent PMCMC

See

Andrieu C., Doucet A. *Particle Markov chain Monte Carlo methods.* J. R. Statist. Soc. B (2010) 72, Part 3, pp. 269–342

Basic Approach to Sampling from $p(\theta, x_{1:T} | y_{1:T})$

Samples jointly
From parameters
& latent state

- Find initial parameters θ by looping to find non-zero posterior value
- MCMC (Metropolis-Hastings) iteration, sampling parameters via MH-algorithm, by considering candidates θ*
  - **Particle filter**, sampling from **trajectories conditional on candidate θ***
    - This particle filtering assumes current candidate parameter value θ* as parameter values
    - Every observation triggers resampling following weight update
    - **Entire trajectory** is sampled, according to **final weight**
      - This is part of sample if θ* accepted
      - An ancestry matrix is maintained
    - Weights across successive observations across all trajectories support calculating the **posterior for** the **candidate parameter value θ* (NB: Can be sensitive to particle count!)**
  - For each sampled value, emit sampled values (new or existing)
    - If accepted candidate θ*, emit θ* and sampled trajectory of latent state
    - If did not accept candidate, state in current place, and (re-)emit those (pre-existing) parameter values **and** the latent state (previously) sampled from them

Particle Filtering with ABMs

- Guidelines for effective particle filtering with ABMs have yet to be elucidated
- Given high nominal (& likely moderately high intrinsic) dimensionality of state space, non-sparse coverage requires high # of particles
- Exceptionally weighty computational resource demand
  - High dimensionality => High number of particles
  - Per-ensemble high because each particle is associated with a …
  - Complete model state representation High memory need
  - ABM: Large populations & inter-agent interactions High computational burden
- Our lines of research
  - Case studies of tradeoffs
  - Creation of ABM-specialized platforms
  - Large-scale parallel implementation (GPU, future: FPGA)
  - Distributed computation

Particle Filtering & PMCMC Advantages

- Ready use PF of with only sparse wastewater data
- Ease of incorporating different variants
- Capacity to apply at diverse levels (facilities, neighbourhood, municipal, SAG, etc.)
- Option to make use of cheaper, passive sampling over periods of time
- Ability to project forward
- Pontential to take advantage of both
    - Lower-sensitivity quantitation
    - Higher-sensitivity (e.g., variant specific) dichotomous presence/absence data
        - Can have stochastic arrivals that are systematically selected against if absent, but otherwise selected for

Agenda

- Motivation
- Particle Filtering with dynamic models
- Wastewater essentials
- Key elements of wastewater-specific model formulation
  - Likelihood
  - Forms of nexus between model state and wastewater concentration
- [As time allows] PMCMC
- Conclusions

Summary

- By combining ML & dynamic modeling, particle filtering & PMCMC provide a convenient, versatile and straightforward means of incorporating wastewater data into transmission model estimates of epidemiological & health service demand state
- **PF use of WW data allows model output to better match empirical data**
- Beyond providing distribution-based estimates of underlying latent state, this method provides the capacity to project state forward & conduct “what if” analyses
- **PF use of wastewater data improves projection accuracy**
- Wastewater data can be incorporated in a straightforward fashion by relating the model estimate of the shedding population to wastewater concentrations
- The approach scales to handle episodic and presence/absence wastewater data
- Results were aided by local infrastructural characteristics, incl. separate wastewater & stormwater sewers, rapid transit times, autosampling of primary effluent streams
- This approach can be readily adapted to other dynamic models&municipal contexts
- Some municipal infrastructures may require model structure elaboration
- In this case, effective application of particle filtering requires 100,000s of particles & substantial computational effort for reliable results at a local & regional level

Acknowledgements

- Doctoral student Xiaoyan Li provided essential leadership in integrating wastewater data into particle filtering, evaluating & refining formulation
- Postdoctoral fellow Dr. Jeremy Eng integrated wastewater data into the PMCMC model and evaluated the results
- Dr. John Giesy, Dr. Yuwei Xie, Dr. Markus Brinkmann and Dr. Kerry McPhedran of the U Sask. Toxicology Centre provided the wastewater used, key background understanding concerning wastewater analysis, lab processes, wastewater infrastructure, shedding, and related areas
- The Saskatchewan Health Authority & Saskatchewan Ministry of Health made possible and realized provision of health system data, funded the development of the model, shaped its requirements, the infrastructure (special thanks to Dr. J. Basran & Jennifer Zerff)
- PHAC for the wastewater contract that accelerated this work, helped provide trainee support, as well as PHAC & FNIHB for supporting wastewater infras.
- NDO wants to extend special gratitude for SYK for inspiring this work