

Modelling COVID-19 in the North American region with a metapopulation network and Kalman filter

Matteo Perini^{1*}, Teresa K. Yamana¹, Marta Galanti¹, Jiyeon Suh¹, Roselyn Kaondera-Shava¹, Jeffrey Shaman^{1,2}

¹*Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, New York, United States of America*

²*Columbia Climate School, Columbia University, New York, New York, United States of America*

* *Corresponding author*

Email: matteo.perini@columbia.edu

Summary

Background

Metapopulation models provide platforms for understanding infectious disease dynamics and predicting clinical outcomes across interconnected populations, particularly for large epidemics and pandemics like COVID-19.

Methods

We developed a novel metapopulation model for simulating respiratory virus transmission in the North America region, specifically for the 96 states, provinces, and territories of Canada, Mexico and the United States. The model is informed by COVID-19 case data, which are assimilated using the Ensemble Adjustment Kalman filter (EAKF), a Bayesian inference algorithm, and commuting and mobility data, which are used to build and adjust the network and movement across locations on a daily basis.

Findings

This model-inference system provides estimates of transmission dynamics, infection rates, and ascertainment rates for each of the 96 locations from January 2020 to March 2021. The results highlight differences in disease dynamics and ascertainment among the three countries.

Interpretation

The metapopulation structure enables rapid simulation at large scale, and the data assimilation method makes the system responsive to changes in system dynamics. This model can serve as a versatile platform for modeling other infectious diseases across the North American region.

Funding

US Centers for Disease Control and Prevention Contract 75D30122C14289; US NIH Grant AI163023.

Introduction

36 Mathematical models have been used to simulate infectious diseases outcomes, infer transmission
37 dynamics, and predict future disease burden. These tools can inform public health strategies by testing
38 control methods and identifying effective interventions¹. During the SARS-CoV-2 pandemic,
39 unprecedented data availability enabled application of mathematical models in many locations and at
40 different geographical scales worldwide²⁻⁵. Metapopulation modeling approaches provide an efficient
41 framework for simulating and evaluating the spatiotemporal progression of infectious disease over
42 large geographic areas. In this model form, populations are aggregated within typically fixed
43 geographic units (e.g. provinces, cities), which allows resolution of spatial disease patterns without the
44 computational expense and micro-behavioral assumptions required for agent-based models. At large
45 geographical scales, agent-based models require high-performance computing (HPC) clusters. For
46 example, Bhattacharya et al. developed a platform to enable the real-time execution of an agent-based
47 COVID-19 model for the United States on more than ten thousand CPU cores⁶.

48
49 Many metapopulation models have been developed to describe the dynamics of infectious diseases at
50 different geographic scales^{7,8}; however, a key challenge in model development lies in accurately
51 determining the movement patterns of individuals among subpopulations. Some metapopulation
52 models use fixed or arbitrary sized geographical areas (cells) as subpopulations, which can then be
53 aggregated to match the resolution of available case, census and movement data^{9,10}. One example of a
54 multi-national system is the GLEAM (Global Epidemic And Mobility) platform¹⁰⁻¹², which estimates
55 the flux of individuals among arbitrary subpopulations centered around major transportation hubs
56 (usually airports) and uses commuting and air travel data¹³.

57
58 In 2020, the US reported the highest number of COVID-19 cases and deaths globally, with the first
59 case identified in Washington state on January 20th and three pandemic waves manifesting during the
60 year¹⁴. Canada reported its first case in Toronto on January 25th and experienced a decline in cases
61 during the summer followed by a resurgence in the fall. Mexico had a similar epidemiological history,
62 but the first wave developed later during the summer¹⁵. The literature currently lacks a comprehensive
63 COVID-19 model for the North American region at continental scale. Here, we present a
64 metapopulation susceptible-exposed-infectious-recovered (SEIR) for COVID-19 for the majority of
65 the North American region (i.e. Canada, United States and Mexico) from the beginning of the
66 pandemic to widespread availability and distribution COVID-19 vaccines. The model is coupled with
67 a data assimilation algorithm and is informed by daily COVID-19 cases and a commuting network
68 that is adjusted by daily mobility trends. Unlike the GLEAM platform, in this model the flux of
69 individuals among subpopulations of the metapopulation model is calculated using the daily work
70 commuting patterns coupled with random movements among the geographical locations, as in
71 previous work⁴.

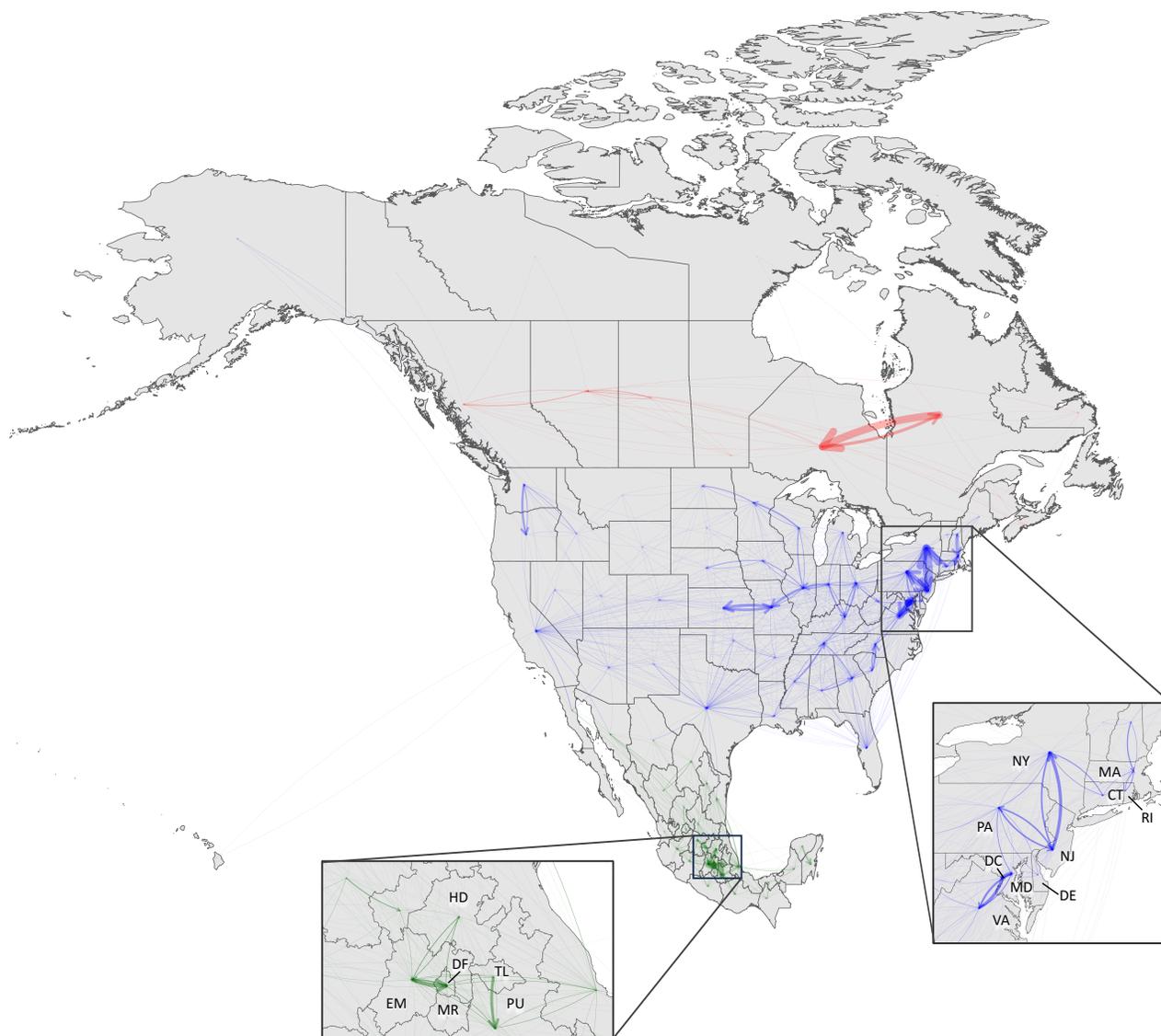
72
73 The model developed here provides temporal estimates of transmission and ascertainment rates for
74 COVID-19 for the North American Region, filling a gap in the existing literature regarding models
75 for this region. The flux patterns and the continental-level metapopulation structure used in this study

76 helps investigation of disease dynamics of COVID-19 at a larger scale, for the majority of the North
78 American region, and can reveal dynamics that are discernible only at such a broad scale. Lastly, the
work commuting matrix structure developed for this study can be paired to different compartmental
80 models and disease surveillance data to study the dynamics of other infectious diseases such as
influenza.

82 **Methods**

We developed a metapopulation susceptible-exposed-infectious-recovered (SEIR) model for the
84 North American region. Specifically, the model represents the 96 first-level administrative divisions
of Canada (10 provinces and 3 territories), United States (50 states and 1 federal district) and Mexico
86 (31 states and 1 autonomous city) represented in **Figure 1**, with a total population of 483 million,
based on census data¹⁶⁻¹⁸. Mixing is simulated as two types of movement: daily commuting and

88 random movement (i.e. all the daily movement across divisions due to reasons other than work commuting).



90 **Figure 1 Commuting to work matrix in most of the North American region.** The map shows the 10
92 provinces and 3 territories of Canada, the 50 states and 1 federal district of United States and the 31 states
94 and 1 autonomous city of Mexico that have been used in the model. The arrows represent the flux of
individuals commuting daily to work to another location. Arrow size represents the number of commuters,
and color represents the country of origin: red for Canada, blue for United States, green for Mexico.

Daily commuting matrix

96 Daily commuting among locations were retrieved and derived from four national datasets: i) Canadian
2016 census (Statistics Canada) Commuting Flow from Geography of Residence to Geography of
98 Work¹⁹; ii) Canada Frontier Counts (Statistics Canada): Number of vehicles travelling between Canada

100 and the United States²⁰; iii) 2011-2015 5-Year American Community Survey (ACS) Commuting Flows
101 (United States Census Bureau)²¹; and iv) Mexican Intercensal Survey 2015 (National Institute of
102 Statistics and Geography, INEGI)¹⁸. Information from these datasets were processed and combined
103 according to the methods described in **Supplementary Note 1** to obtain the commuting work matrix
104 represented in **Figure 1**. The matrix contains the number of people that commute daily to work in
105 another location (states for Mexico and US, provinces or territories for Canada). To account for
106 control measures and closures enacted during the estimation period, the commuting work matrix was
107 scaled based on daily mobility activity, derived from the “change in workplace visitors” trends from
108 Google Community Mobility Reports²² (see **Supplementary Figure 1**). To account for movement
109 between locations for purposes other than work commuting, the model includes daily random
110 movement among locations; this movement was set to be proportional to the average number of
111 working commuters among each location pair (see **Table 1**).

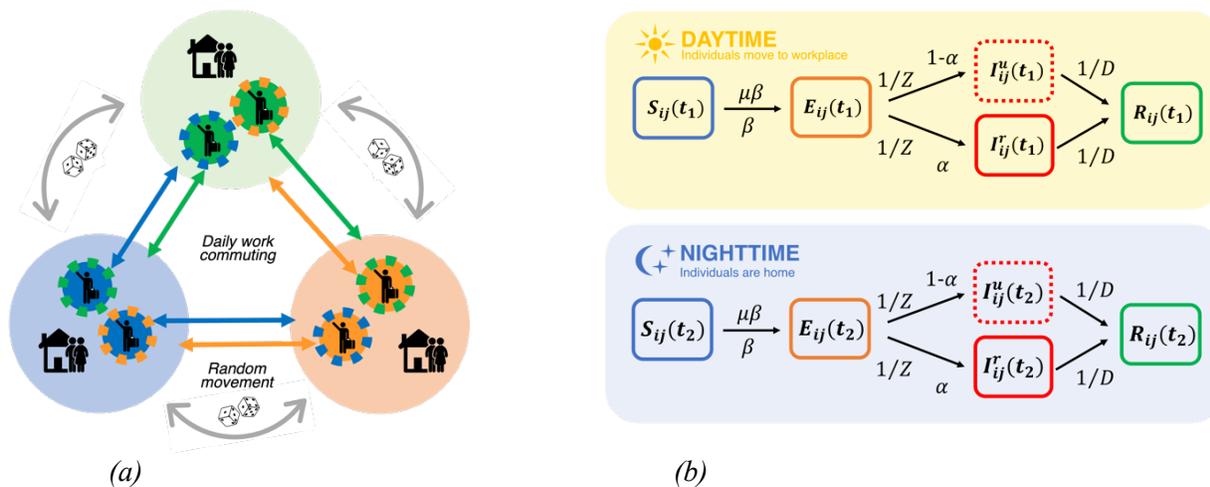
112 **Transmission model**

113 The metapopulation model resolves daytime and nighttime mixing differences, depicting diurnal
114 changes in contact among subpopulations. Transmission occurs as a discrete Markov process during
115 both day and nighttime, following the structure of previous studies in the US⁴. The transmission
116 dynamics are described by the equations in **Supplementary Note 2 (eqs. S4-S13)**. In these equations,
117 S_{ij} , E_{ij} , I_{ij}^r , I_{ij}^u , and N_{ij} represent the susceptible, exposed, reported infectious, unreported infectious,
118 and total population in the subpopulation commuting from location j to location i ($i \leftarrow j$).
119 Additionally, we assume that no individuals enter or leave the model, and that there is no loss of
120 immunity following primary infection, given the relatively short simulation time period. For the same
121 reason, we can compute the R_{ij} (Removed population) as $R_{ij} = \sum_1^t I_{ij}^r + \sum_1^t I_{ij}^u$ which is the
122 cumulative sum of daily new reported and unreported cases, or the sum of the individuals that exited
123 the two infectious compartments.

124 The parameters of the model are: β , the transmission rate of reported infections; μ , the relative
125 transmissibility of unreported infections; Z , the average latency period (from infection to
126 contagiousness); D , as the average duration of contagiousness; α , the fraction of documented
127 infections (ascertainment rate); and θ , a multiplicative factor adjusting random movement. A distinct
128 transmission rate, $\mu\beta$, is defined for undocumented infections: we assume that these individuals show
129 little to no symptoms during infection and are less contagious than documented infections⁵. Each
130 equation is integrated using a Poisson process to capture the stochastic nature of transmission
131 dynamics. In total, the model consists of 3,268 metapopulations. To reduce the dimension of the
132 system being estimated and to improve identifiability, we fixed the values of the parameters related to
133 disease progression (Z , D , and μ) and the multiplicative factor for random movement (θ) based on
134 previously published findings^{4,14}. Local transmission rates β_t^l and ascertainment rates α_t^l are estimated
135 for each of the 96 locations, l at time t . The hyperparameter R_t^l , the time-varying reproductive
136 number, was derived using the next-generation matrix approach^{4,23}:

$$137 \quad (1) \quad R_t^l = \beta_t^l D [\alpha_t^l + (1 - \alpha_t^l) \mu]$$

- 140 R_t^l Time-varying reproductive number for location l at time t
- D Duration of contagiousness
- 142 β_t^l Transmission rate of location l at time t
- α_t^l ascertainment rate of location l at time t
- μ relative transmissibility of unreported infections



144 (a)

146 **Figure 2. Metapopulation structure and compartmental model:** (a) Daily work commuting - during the

148 daytime, some individuals commute from their home to their workplace in another location and mix with

150 the populations present there. During the nighttime, those commuters return home and mix with other

152 residents who live in the same location. Random movement - individuals may travel among locations for

reasons other than work. These random visitors circulate among subpopulations following a Markov

process, causing a population exchange in all locations (b) I_{ij}^r reported infected and I_{ij}^u unreported infected

from location from location j to location i ($i \leftarrow j$); t_1 daytime duration; t_2 nighttime duration; β

transmission rate; α ascertainment rate; μ relative transmissibility of unreported cases; Z latency period;

D duration of contagiousness.

154 **COVID-19 cases and data assimilation**

156 The model is informed by daily COVID-19 confirmed case data retrieved from COVID-19 Open

158 Data — Google Health²⁴ starting from January 20th 2020 to March 31st 2021 for the 96 first-level

160 administrative divisions of Canada, United States and Mexico. This time frame was chosen to capture

the first three COVID-19 waves¹⁴, before the emergence of the Delta and subsequent variants and

162 more widespread uptake of COVID-19 vaccines. A 7-day moving average was applied to the daily

cases data to smooth out daily fluctuations in reporting. The smoothing also mitigates the impact of

reporting delays or inconsistencies in case data, providing a more reliable indicator of overall disease

dynamics.

164 The estimation of state variables and parameters is carried out by data assimilation implemented using

166 the Ensemble Adjustment Kalman Filter (EAKF) algorithm²⁵, as in previous studies^{26,27}. Kalman filters

use Bayes' rule to update state variables and parameters. Normality is assumed for the prior

168 distribution and the likelihood so that the posterior distribution can be characterized by the mean and

170 the covariance. In this model, an ensemble of 300 simulations is integrated to generate a prior
172 distribution of parameters and state variables, including estimation of the observed state variable. At
174 each observation time point, the model is halted, and the ensemble and observation are used to
176 calculate the Kalman gain, which is used to update the observed state variable. The unobserved state
178 variables and parameters are then updated in proportion to the same Kalman gain. Finally, the
posterior estimates are used as priors and model integration through time continues to the next
observation. The formulas used to calculate the Kalman gain for the observed and unobserved
variables and parameters are presented in the **Supplementary Note 3 (eqs, S14-S15)**. To properly
balance the influence of the observational data in the assimilation process, it is crucial to estimate its
error, which is typically unknown. Here, we estimated the Observational Error Variance of the
observational data as shown in **eq. (2)**, similarly to prior works^{5,28,29}:

$$180 \quad (2) \quad OEV_t^l = \max \left(5, \frac{(O_t^l)^2}{100} \right)$$

OEV_t^l Observational Error Variance of location l at time t

182 O_t^l average cases in location l in the week before time t

184 Repeated filter adjustments tend to decrease model ensemble variance reducing the impact of new
observations in subsequent estimates. This may lead to divergence, in which the filter ceases adjusting
186 the model state²⁵. To avoid divergence, at each timestep we applied a multiplicative factor (1.01) to
inflate the prior ensemble of the observed variable (daily reported infected I^r) and the estimated
188 parameters (ascertainment rate α and transmission rate β). Additionally, we reinitialized the values of
the estimated parameters for a fraction (2%) of the ensemble members every 7 days. This
190 reinitialization enables the system to periodically readjust estimations whenever the ensemble variance
begins to shrink, thus preventing divergence.

192 Further details on the SEIR-EAKF assimilation process are available in the **Supplementary Note 3**.

194 **Model initialization**

196 Each ensemble member was initialized with a set of parameter and state variable estimates to resemble
the epidemiological conditions at the beginning of the SARS-CoV-2 pandemic in 2020. The parameter
198 values related to disease progression (Z , D and μ) were drawn from distributions with ranges reported
in **Table 1**, in accordance with Li et al⁵. These parameters are assumed to remain fixed over time and
200 to have consistent values across all locations, representing inherent biological characteristics of the
disease. The random movement factor θ is also constant over time, with a random value drawn from
202 a uniform distribution between 0 and 0.2 assigned to each location (see **Table 1**). This value represents
the relative volume of random movement compared to commuting, where $\theta = 0.15$ indicates that
204 the number of random visitors is 15% of the average number of commuters between two locations.
Exposed (E), reported infectious (I^r) and unreported infectious (I^u) individuals were initialized with
206 random draws from a uniform distribution that ranged from 1 to 9 in each subpopulation.

208 The three countries exhibited substantive differences in healthcare and testing capacity during the
210 pandemic. Mexico had particularly low testing rates^{30,31}, suggesting that only severe cases were assayed
212 due to limited availability of test kits. During 2020, the United States experienced the highest number
214 of cases, as well as relatively high nation-wide ascertainment rates, as shown in prior inference
216 studies^{4,14}. For Canada, the estimate of national testing rates or the ascertainment rate is lacking;
218 however, Ontario, the most populous province in Canada, accounting for ~37% of the population,
initially faced challenges in ramping up its testing capacity, but its centralized resource strategy was
able to increase test capacity during 2020³². Given these differences among the three countries, we
assigned country-specific initial ranges for the ascertainment rate α . For the US we used SARS-CoV-
2 seroprevalence estimates and cumulative reported cases to estimate the value of the ascertainment
rate. Specifically we used infection-induced seroprevalence from blood samples collected in US during
July 2020 (3.5%)³³ to derive an initial estimate of α in the US, this formula is shown in **eq. (3)**:

$$220 \quad (3) \quad \alpha_{US}^0 = \text{median}\left(\frac{\sum \text{cases}_j}{IIS_{US} \times \text{pop}_j}\right)$$

α_{US}^0 Initial ascertainment rate for the US

222 $\sum \text{cases}_j$ Cumulative reported cases of state j up to July 2020

IIS_{US} Infection-induced seroprevalence in the United States during July 2020 (3.5%)³³

224 pop_j Population of state j

226 Based on this formula, we set the initial prior distribution mean to 0.25 for α in US.

228 Infection-induced seroprevalence was not available for Canada and Mexico during the time frame of
interest. Instead, considering the very low testing rate for Mexico^{30,31} we set the initial prior distribution
230 mean to 0.08, corresponding to a ~70% decrease from the US. Canada faced early issues establishing
testing facilities³², so we set the initial prior distribution mean to 0.12 for α , between the values of US
232 and Mexico (~50% decrease compared to the US).

234 The initial values of β were drawn for each location from the normal distribution with mean $\mu = 1.93$
and standard deviation $\sigma = 0.75$ similarly to Li et al.⁵(see **Table 1**).

236 The SEIR-EAKF model-inference system can potentially estimate parameters and state variables
238 values that violate physicality. This could happen, for example, if a state variable or estimated
parameter is adjusted by the filter to a value below zero. To avoid this issue, we applied constraints to
240 both the state variables and parameters, subsequently re-assigning values that fell outside specified
bounds. Specifically, ensemble members associated with any state variable that exhibited values less
242 than or equal to zero were assigned values from the previous day. The Susceptible population S_{ij} is
not directly adjusted by the EAKF. Instead, it is computed using the total population count and the
244 EAKF-adjusted state variables as $S_{ij} = N - E_{ij} - R_{ij}$. This guarantees population mass balance and is
feasible under the assumption of non-reinfection among individuals. Additionally, it prevents the

246 variable S_{ij} from exceeding the total population, which aligns with the assumption of a constant
 248 population.

The lower bound of α was set to 0.025 corresponding to 5 reported cases every 200 infections. For
 250 Canada and US, this bound increased linearly by 0.5% at each day, or: $\alpha_{t+1}^{min} = \alpha_t^{min} * 1.005$. This
 252 increase was imposed to reflect the efforts of the two countries to increase the detection capabilities
 254 of local and national health systems^{32,34-36}. Conversely, Mexico implemented a sentinel surveillance
 256 system³⁷ in which only the hospitalized cases and 10% of the mild cases were tested³⁸. Focusing testing
 258 efforts on symptomatic individuals is a cost-effective strategy, but it also greatly increases the
 260 proportion of unreported cases, resulting in lower ascertainment rates. For this reason, the lower α
 bound of Mexico did not increase over time in our model. The upper bound for α was fixed for all
 locations at 0.6 (i.e. 60 cases per 100 infections). We allowed the transmission rate β to span a broad
 range of values, with the lower bound set to 0.2 and the upper bound set to 4. The descriptions, prior
 ranges and bounds for all the model parameters are shown in **Table 1**.

	Description	Initial prior range distribution	bounds	Parameter type
Z	Latency period	$U(2, 5)$ days	Fixed	Fixed
D	Duration of contagiousness	$U(2, 5)$ days		
μ	Relative transmissibility of unreported infections	$U(0.2, 0.45)$		
θ	Random movement factor	$U(0.0, 0.2)$		
β	Transmission rate	$N(1.93, 0.75)$	[0.2, 4]	Estimated
α	Ascertainment rate - Canada	$U(0.08, 0.17)$	[0.025, 0.6]*	
	Ascertainment rate – United States	$U(0.13, 0.37)$		
	Ascertainment rate - Mexico	$U(0.04, 0.12)$		

* Mexico lower bound is fixed to 0.025 over time; Canada and US lower bound increases by 0.5% daily: $\alpha_{t+1}^{min} = \alpha_t^{min} * 1.005$

262 **Table 1:** parameters description, initial prior distribution, ranges, and parameter type

264 System identifiability

To assess the identifiability of the model, we first tested our framework using synthetically generated
 266 datasets. Specifically, we generated a suite of synthetic outbreaks using the model (**eqs. S4-S13**) in free
 268 simulation, each with arbitrarily assigned values of the epidemiological parameters and initial
 conditions for state variables. We then ran the full model-inference system 100 times assimilating the
 270 daily new reported cases time series generated by each of these free simulations to test the system's
 272 ability to accurately estimate state variable and parameter values. As in practice with actual data, we
 fixed the parameters related to disease progression (Z , D , and μ) and the multiplicative factor for
 random movement (θ), while the ascertainment rate α and transmission rate β were estimated for
 each location. The initial prior distribution and range of the parameters were as reported in **Table 1**.

274

Results

276 Daily commuting matrix

278 The daily commuting matrix depicted in **Figure 1** was obtained by combining and processing the
279 national datasets reported in the Method section (full description in **Supplementary Note 1**); it is
280 summarized in **Table 2**. Out of a combined population of 483 million, the Canadian population
281 accounts for the 7.3% (35.1 million), the US for the 67.2% (324.3 million) and Mexico for the 25.6%
282 (123.3 million). Around 8.9 million people, or 1.84% of the represented North American population,
283 commutes daily to another state/province/territory or country to work. Inter-country commuting
284 accounts for just the 1.12% (99,369) of the 8.9 million commuters. Finally, the percentage of internal
285 commuters for each country is similar to the population percentage relative to the total population of
286 the three countries.

Residence	Workplace	Commuters	% over Total Commuters
Canada	Canada	527,300	5.93
Canada	USA	1,962	0.02
USA	Canada	4,605	0.05
USA	USA	5,334,072	60.02
USA	Mexico	12,892	0.15
Mexico	USA	79,910	0.90
Mexico	Mexico	2,925,820	32.92
		Total Commuters	% over total Population
		8,890,144	1.84

288 **Table 2:** Number of commuters by country of residence and workplace. Individuals with residence and
289 workplace in the same country are commuting between states/provinces/territories within that country. The
290 last column shows the percentage over the total commuters. The last row shows the total number of cross-
291 border (state, province, territory, or country) commuters and its percentage over the total population of the
292 model.

294 System Identifiability

295 To verify the convergence of the estimated parameters (ascertainment rate α and transmission rate β)
296 to the synthetic truth values created in free simulation, we plotted the model-generated data points
297 over the boxplot distributions of the 3000 estimated parameter values (300 ensemble members x 100
298 ensemble simulations) at the end of each outbreak among the locations (see **Supplementary Figure**
299 **2**). Most of the true values fall in the interquartile of the distribution of the estimated values (68% for
300 α and 84% for β) while all the true values falls in the 95% CI of the distribution. This demonstrates
301 the ability of the system to estimate local time-varying α and β values.

302

Simulation with Case Data

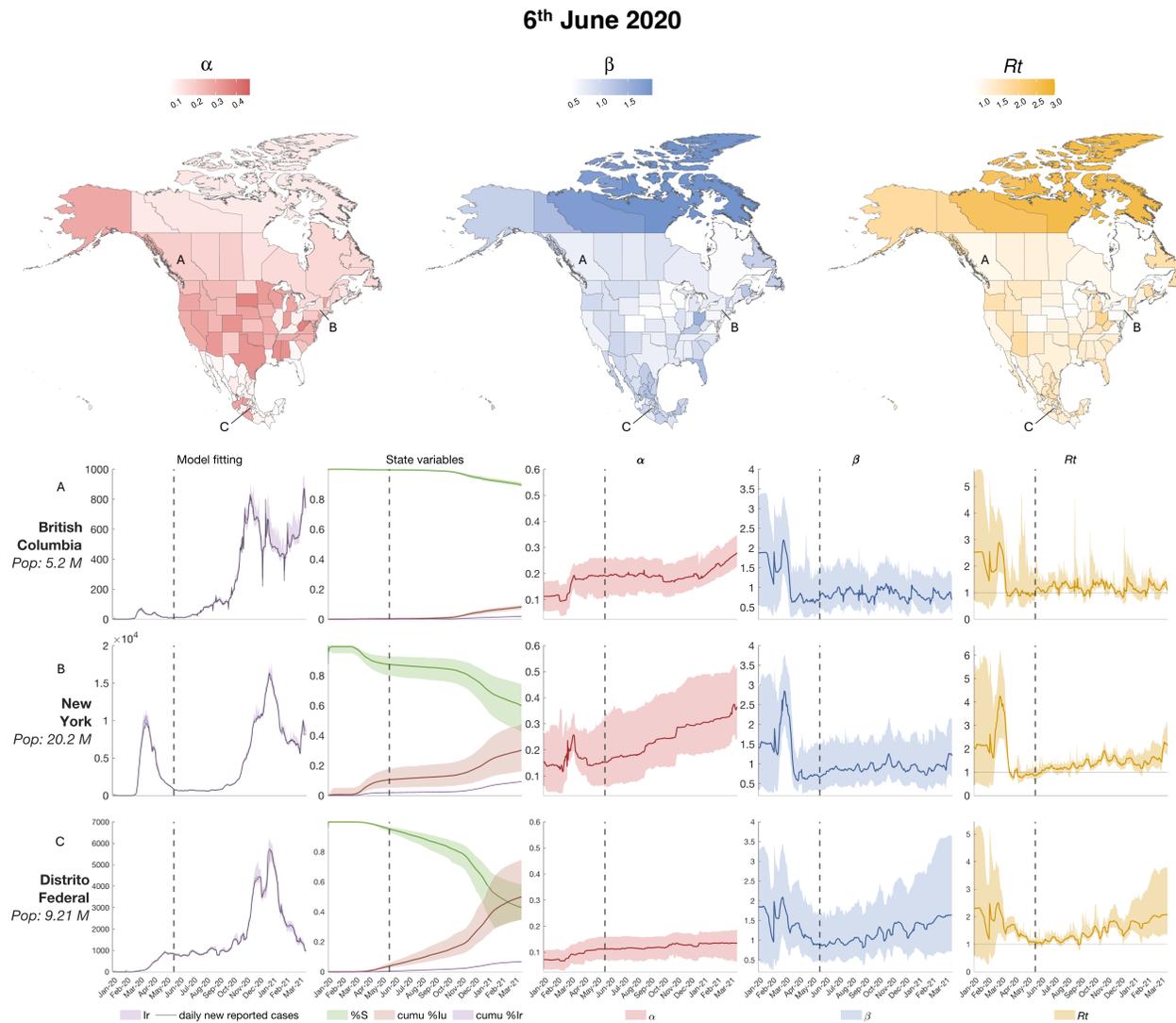
304 The model-inference system was run with the real case data of Canada, United States and Mexico
starting from January 20th 2020 to March 31st 2021. During this time frame, the three countries

306 experienced asynchronous outbreaks as depicted in **Supplementary Figure 1**. To explore disease
 308 dynamics over time and location, we selected three timepoints when cases were declining in most of
 the locations of the North American region, roughly corresponding to the end of the three pandemic
 310 waves experienced in United States during 2020¹⁴. The first wave began in January 2020 and lasted
 through the spring; we selected June 6, 2020 as the first timepoint. **Figure 3** shows the estimated
 312 values of the parameters (ascertainment rate α and transmission rate β) and hyperparameters (basic
 reproductive number R_t) in all 96 locations of the study region at this time point. In addition, the
 lower panels of **Figure 3-5** show the time progression of model fitting (modeled daily new reported
 314 cases and observed daily cases), select state variables (S , cumulative I^r , cumulative I^u) and parameter
 and hyperparameter estimates (α , β , R_t) for a total of nine selected locations. These locations were
 316 selected among the 96 to represent the epidemiological progression in different geographical areas of
 the North American region, focusing on some of the most populous and epidemiologically relevant
 318 locations of the three countries. In **Figure 3** the three selected locations indicated in the maps are
 British Columbia (Canada), New York (United States) and Distrito Federal (Mexico City, Mexico).
 320 The sub-national parameter values for each of the three countries were aggregated to the national level
 using a population-weighted average. On June 6, 2020, the population-weighted average
 322 ascertainment rate α of the mean estimates in US states (0.26) was 1.5 times higher than in Canada
 (0.17) and 2.7 times higher than in Mexico (0.10). The population-weighted average transmission rate
 324 β of the mean estimates for Mexican states was 1.03, which is 1.4 times higher than in Canada (0.72)
 and the US (0.72). The population-weighted average of the time-varying reproductive number R_t of
 326 the mean estimates was 0.99 for Canadian provinces and territories, 1.11 for the United States, and
 1.22 for Mexican states. These values are also reported in **Table 3**.

328

Location	Jun. 6 th 2020	Sep. 7 th 2020	Mar. 15 th 2021
Ascertainment rate α			
Canada	0.17	0.18	0.26
United States	0.26	0.31	0.36
Mexico	0.10	0.11	0.13
Transmission rate β			
Canada	0.72	0.95	0.84
United States	0.72	0.64	0.85
Mexico	1.03	0.86	0.85
Time-varying basic reproduction number R_t			
Canada	0.99	1.32	1.31
United States	1.11	1.04	1.49
Mexico	1.22	1.04	1.07

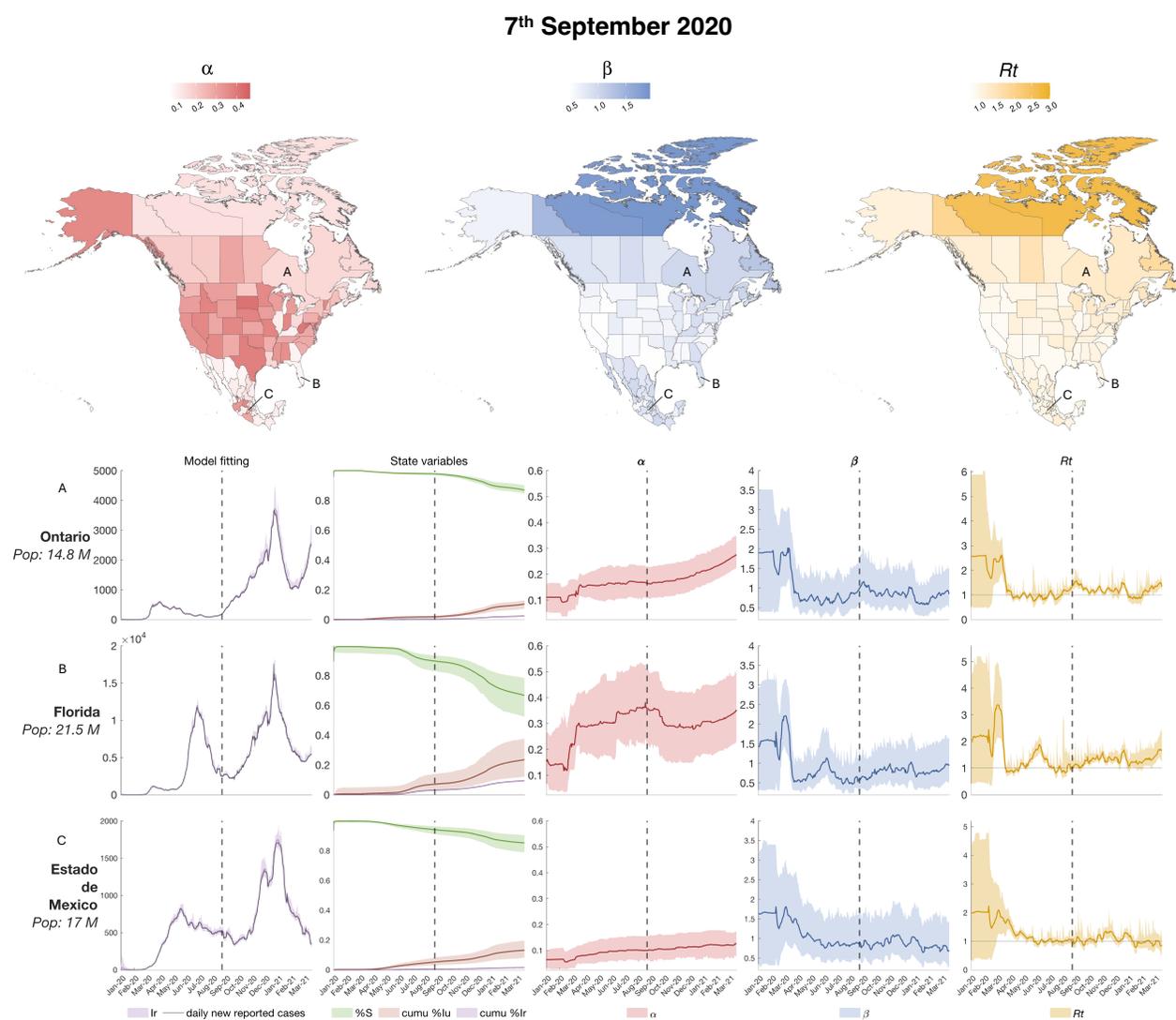
330 **Table 3:** Population-weighted average of the mean estimates values of the parameters (ascertainment rate
 332 α and transmission rate β) and hyperparameters (time-varying basic reproductive number R_t) for the three
 countries at three selected timepoints. The intensity of cell colors in the table corresponds to their values:
 higher values are represented by more intense colors.



334 **Figure 3. Estimated state variables and parameters on June 6, 2020:** the three maps on the top panel show
 336 the value of the of the parameters (ascertainment rate α and transmission rate β) and the hyperparameter
 338 (basic reproductive number R_t) for all 96 locations on June 6, 2020. The bottom panels show the model
 340 fitting (i.e. the estimated observed variable over the 7-day smoothed daily new reported cases), three state
 342 variables illustrating disease progression (the susceptible population, cumulative reported infectious and
 cumulative unreported infectious), and the parameters α , β and R_t for 3 selected locations: British
 Columbia (Canada), New York (US) and Distrito Federal (Mexico City, Mexico). The color shaded areas
 represent the 95% credible interval from the 300-member ensemble. The dotted vertical lines indicate the
 timepoint of reference for the maps.

344 During summer 2020, some locations, particularly in the US, experienced a second wave consisting of
 346 a resurgence of cases, with a decline at the beginning of fall. **Figure 4** shows the values of the state
 variables and parameters on September 7, 2020. The maps show the parameter estimates across all

348 locations, while the temporal dynamics of state variable and parameter estimates are shown for three
 350 of the nine selected locations: Ontario (Canada), Florida (US) and Estado de México (Mexico). On
 352 September 7, 2020, the population-weighted average ascertainment rate α of the mean estimates in
 354 the United States was 0.31, which is 1.74 times higher than in Canada (0.17) and 2.8 times higher than
 356 in Mexico (0.11). The population-weighted average transmission rate β of the mean estimates in
 Canadian provinces and territories was 0.95, which is 1.5 times higher than in the US (0.64) and 1.1
 times higher than in Mexican (0.86). The population-weighted average of the time-varying
 reproductive number R_t of the mean estimates was 1.32 for Canadian provinces and territories, 1.26
 times higher than in the United States (1.04) and Mexican states (1.04). These values are also reported
 in **Table 3**.

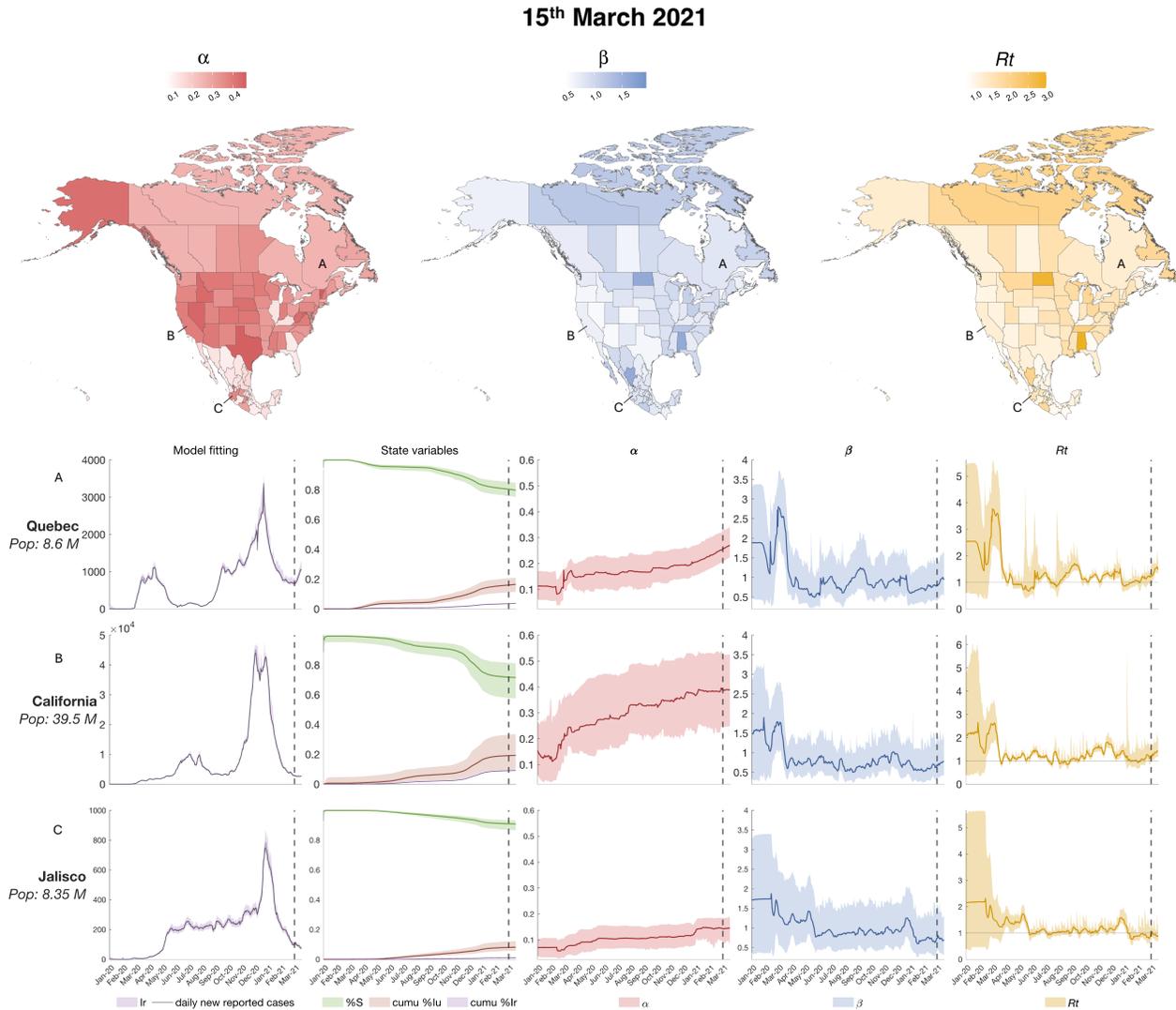


358 **Figure 4. Estimated state variables and parameters on September 7, 2020:** the three maps on the top
 360 panel show the value of the of the parameters (ascertainment rate α and transmission rate β) and the
 hyperparameter (basic reproductive number R_t) for all 96 locations on September 7, 2020. The bottom

362 *panels show the model fitting (i.e. the estimated observed variable over the 7-day smoothed daily new*
364 *reported cases), three state variables illustrating disease progression (the susceptible population,*
366 *cumulative reported infectious and cumulative unreported infectious), and the parameters α , β and R_t for*
3 selected locations: Ontario (Canada), Florida (US) and Estado de México (Mexico). The color shaded
areas represent the 95% credible interval from the 300-member ensemble. The dotted vertical lines indicate
the timepoint of reference for the maps.

368 Most locations experienced a more severe outbreak during the autumn-winter wave of 2020/2021,
before the widespread availability of the COVID-19 vaccines. We plot the estimated parameters and
370 hyperparameters on March 15, 2021 in **Figure 5**. The state variable and parameter estimate time series
are also shown for three of the nine selected locations: Quebec (Canada), California (United States),
372 and Jalisco (Mexico). On March 15, 2021, the population-weighted average ascertainment rate α of
the mean estimates in the United States was 0.36, 1.4 times higher than Canada (0.26) and 2.7 times
374 more than Mexico (0.13). The population-weighted average transmission rate β of the mean estimates
was 0.83 in Canada and 0.85 in US and Mexico. The population-weighted average of the time-varying
376 reproductive number R_t was 1.49 for the United States, 1.14 times higher than Canadian provinces
and territories (1.31) and 1.39 times higher than Mexican states (1.04). These values are also reported
378 in **Table 3**. All the estimated mean values of the parameters (ascertainment rate α and transmission
rate β) and hyperparameters (basic reproductive number R_t) for all the location on these three
380 timepoints are reported in **Supplementary Table 3**.

382



384 **Figure 5. Estimated state variables and parameters on March 15, 2021:** the three maps on the top
 386 panel show the value of the of the parameters (ascertainment rate α and transmission rate β) and the
 388 hyperparameter (basic reproductive number R_t) for all 96 locations on March 15, 2021. The bottom panels
 390 show the model fitting (i.e. the estimated observed variable over the 7-day smoothed daily new reported
 392 cases), three state variables illustrating disease progression (the susceptible population, cumulative
 reported infectious and cumulative unreported infectious), and the parameters α , β and R_t for 3 selected
 locations: Quebec (Canada), California (United States), and Jalisco (Mexico). The color shaded areas
 represent the 95% credible interval from the 300-member ensemble. The dotted vertical lines indicate the
 timepoint of reference for the maps.

394 Discussion

396 In this study we developed a SEIR metapopulation model structure and combined it with a data
 assimilation algorithm (EAKF) to reproduce COVID-19 outbreak dynamics and estimate important

398 epidemiological parameters across most of the North American region during the first three pandemic
waves¹⁴. The model has demonstrated identifiability in estimating system state variables, as well as the
400 ascertainment rates, α , and transmission rates, β , for the 96 first-level administrative divisions of
Canada, United States, and Mexico. The metapopulation structure provided computational efficiency
402 in comparison to agent based models that require HPC (high-performance computing) and cluster
computing approaches to run at large scale⁶. The transmission module of the model structure relies
404 on the daily work commuting patterns across states, provinces, and territories for disease spreading,
with the capability to adjust the daily commuting matrix using Google Mobility Report trends data²²
to account for travel restrictions implemented during the pandemic.

406
As depicted for selected locations in **Figure 3-5** and reported for all the locations in **Supplementary**
408 **Table 3**, the SEIR-EAKF system estimated large disparities in the ascertainment rates α among the
three countries^{4,14,30-32}. Moreover, it estimated a gradual increase in the ascertainment rates α
410 throughout the three major COVID-19 outbreaks during 2020 and at the beginning of 2021 in each
country. Specifically, the US showed substantially higher α values in almost all states across the three
412 waves: at the last time point (March 15, 2021) the majority of states had reached an ascertainment rate
of $\sim 40\%$. Canada and Mexico showed smaller ascertainment rates than the US at the end of the
414 estimation: most of the Canadian provinces and territories had an ascertainment rate of $\sim 26\%$, while
Mexican states had less than 20%. Note that in Canada and the United States, the increase in α is
416 partially influenced by the increasing lower imposed during inference, whereas Mexico has a fixed α
lower bound over time.

418
Transmission rates β remained generally low during the estimation period for the majority of
420 locations. In the United States, the estimated β values in the majority of locations appeared to exhibit
minimal variation throughout the estimation period, while in Mexico transmission rates generally
422 decreased gradually over time. A notable exception was Distrito Federal (Mexico City), the very
densely populated capital of Mexico. The relatively high β value at the end of the estimation in March
424 2021 (1.6) was not unexpected in this location. High population density allows for more contact
opportunities, increasing the rate of transmission and the basic reproduction number as show in
426 previous analysis⁴. A few additional exceptions to the national trends are discussed in the
Supplementary Note 4.

428
The hyperparameter R_t (time-varying basic reproductive number) is proportional to β and α , as
430 reported in **eq. (1)**. The majority of the mean estimated R_t values were above the epidemic threshold
of $R_t = 1$ at all the three selected time points and all locations had mean $R_t > 1$ at least once among
432 the three selected time point. Moreover, the lower estimated mean values for R_t never dropped below
0.75, or below 0.86 at the last time point. These values suggest a sustained epidemic in the North
434 American region.

436 Overall, the three countries demonstrated distinct and quite isolated epidemiological histories. This is
438 supported by the results presented in **Figures 3-5** and **Supplementary Table 3** where the trends
440 estimated for the parameter and hyperparameter values can be clustered by country, highlighting the
442 national and interconnected epidemiological developments they have undergone. This phenomenon
444 arises from the commuting network structure depicted in **Figure 1**, where few arrows traverse national
446 borders, illustrating three major networks (the countries) with limited interconnection.

448 The model developed for this work uses a new commuting matrix, based on national census data and
450 surveys, to model worker flows in the American region (Canada, US, and Mexico). It assumes most
452 movements are daily work commutes and accounts for additional movements with a random factor
454 proportional to the commuting flux. Moreover, the system couples the compartmental model (SEIR)
456 with the Ensemble Adjustment Kalman Filter (EAKF), adjusting the system state variables and
458 parameters daily on the basis of the case data. This approach differs from other models used to
460 estimate infectious diseases on a multi-national and/or continental scale, such as the GLEAM
462 platform¹⁰⁻¹². Those platforms estimate the movement of individuals among various arbitrary
464 subpopulations located around major transportation hubs and leverages commuting and air travel
466 data. In contrast, our approach utilizes state, province, and territory boundaries to define locations
468 and is not informed by air travel data. Instead of adjusting the estimation, the GLEAM platform uses
470 case data and other datasets to conduct a calibration run aimed at identifying the optimal set of
472 parameters that best fit the real data in its estimations³⁹.

474 The role of asymptomatic individuals has been shown to be central for the spread of viruses like
476 SARS-CoV-2 and influenza, with most transmissions occurring with exposure times exceeding an
478 hour⁴⁰. Therefore, it is reasonable to center a model system around work-related commuting;
480 infectious and asymptomatic individuals share the same space with coworkers for several hours,
482 increasing their probability of infection; subsequently, each worker returns home to their resident
484 locations, further increasing the probability of spreading disease to their families. Conversely, the
486 GLEAM platform focuses on global disease spread, showing that the global spatiotemporal patterns
488 of disease spreading are mainly determined by the airline network⁴¹. Additionally, the inclusion of air
490 travel data enables GLEAM to capture global phenomena such as the external introduction or
492 reintroduction of the virus during the estimation process. In contrast, the SEIR-EAKF model
494 developed in this work is more sensitive to local changes in trends, as it leverages case data to adjust
496 estimations. The importance of modeling each introduction event is particularly relevant at the
498 beginning of an epidemic. This significance decreases once the epidemic in a region is primarily driven
500 by internal transmission dynamics. Our model accommodates the external introduction of new
502 infections through stochastic integration instead of relying on international flight data. However, the
504 inclusion of global air travel data can improve system estimation, particularly in the initial weeks
506 following the introduction of virus and for epidemics less prevalent than COVID-19. These
508 differences make the two modeling approaches particularly effective at capturing different aspects and
510 phases of epidemics. Therefore, combining these and other approaches into a single model or multi-
512 model ensemble could improve the capacity to estimate and predict disease parameters.

478 The development of multinational dynamical models entails more time and effort compared to
480 localized models because it requires the reconciliation of heterogeneous data sources. For example,
482 the three countries in this study conducted independent census surveys that needed to be carefully
484 interpreted and meticulously merged to ensure the resulting multinational contact network was
486 homogeneous. The great advantage of using realistic contact networks that encompass multiple
488 countries is the ability to infer simultaneously across broader regions (e.g. North America), rather than
comparing results from separate local inference systems, thus highlighting the geographic spread of
the disease across borders. Furthermore, the inference system implemented in this study can serve as
a platform for modeling other respiratory infectious diseases, such as influenza, by pairing the North
American commuting network developed here with other mathematical models.

Robust models for estimating disease parameters not only help to understand disease dynamics and
assess responses at different locations, but they also can be adapted for use during epidemics to
generate forecasts that can help health authorities in developing and implementing more informed
policies. Using spatially resolved dynamical models applied at scale, it is possible to compare how
various factors such as public health policies, population density, and mobility patterns affect disease
spread and control in different locations. This comparative inference helps identify the most effective
strategies and conditions for controlling epidemics, providing valuable insights for tailoring public
health interventions to specific regions or populations. Additionally, this model can be implemented
to support monitoring systems and counterfactual simulation, enhancing public health preparedness
and response by enabling data-driven and location-specific strategies that can directly improve
epidemic control of the region.

500

Contributors

502 Data curation: MP. Formal Analysis: MP, TKY, MG. Investigation: MP. Software: MP, TKY, MG, JS
RKS. Methodology: MP, JS, RKS. Writing – original draft: MP, JS, RK. Validation: TKY, MG.
504 Supervision: TKY, MG, JS. Writing – review & editing: TKY, MG, JS. Funding acquisition: JS.
Conceptualization: JS. Project administration: JS.

506 Data sharing statement

The model scripts and the input data are publicly available on GitHub
508 https://github.com/MatteoPS/NA_SEIR-EAKF

Declaration of interest

510 JS and Columbia University disclose partial ownership of SK Analytics. JS discloses consulting for
BNI. All other authors declare no competing interests.

512

514 **References**

- 516 1. Pagel C, Yates CA. Role of mathematical modelling in future pandemic response policy. *BMJ* [Internet]. 2022 Sep 15 [cited 2024 Mar 18];e070615. Available from: <https://www.bmj.com/lookup/doi/10.1136/bmj-2022-070615>
- 518 2. Zhou X, Ma X, Gao S, Ma Y, Gao J, Jiang H, et al. Measuring the worldwide spread of COVID-
520 19 using a comprehensive modeling method. *BMC Med Inform Decis Mak* [Internet]. 2023
522 Sep 15 [cited 2024 Mar 18];21(S9):384. Available from:
<https://bmcmidinformedecismak.biomedcentral.com/articles/10.1186/s12911-023-02213-4>
- 524 3. Zhang W, Xie R, Dong X, Li J, Peng P, Dr Santibanez Gonzalez E. SEIR-FMi: A coronavirus
526 disease epidemiological model based on intra-city movement, inter-city movement and medical
528 resource investment. *Comput Biol Med* [Internet]. 2022 Oct [cited 2024 Mar 18];149:106046.
Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0010482522007636>
- 526 4. Sen Pei, Kandula S, Shaman J. Differential effects of intervention timing on COVID-19 spread in
528 the United States. *Sci Adv* [Internet]. 2020 Dec 4 [cited 2023 Dec 22];6(49):eabd6370.
Available from: <https://www.science.org/doi/10.1126/sciadv.abd6370>
- 530 5. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, et al. Substantial undocumented infection
532 facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* [Internet]. 2020
May [cited 2023 Dec 22];368(6490):489–93. Available from:
<https://www.science.org/doi/10.1126/science.abb3221>
- 534 6. Bhattacharya P, Machi D, Chen J, Hoops S, Lewis B, Mortveit H, et al. Novel multi-cluster
536 workflow system to support real-time HPC-enabled epidemic science: Investigating the impact
of vaccine acceptance on COVID-19 spread. *J Parallel Distrib Comput* [Internet]. 2024 Sep 1
[cited 2024 Jun 3];191:104899. Available from:
<https://www.sciencedirect.com/science/article/pii/S0743731524000637>
- 538 7. Kong L, Duan M, Shi J, Hong J, Chang Z, Zhang Z. Compartmental structures used in modeling
540 COVID-19: a scoping review. *Infect Dis Poverty* [Internet]. 2022 Jun 21 [cited 2024 Jan
23];11(1):72. Available from: <https://idpjournal.biomedcentral.com/articles/10.1186/s40249-022-01001-y>
- 542 8. Wang P, Zheng X, Liu H. Simulation and forecasting models of COVID-19 taking into account
544 spatio-temporal dynamic characteristics: A review. *Front Public Health* [Internet]. 2022 Oct 18
[cited 2024 Jan 23];10:1033432. Available from:
<https://www.frontiersin.org/articles/10.3389/fpubh.2022.1033432/full>
- 546 9. Venkatramanan S, Sadilek A, Fadikar A, Barrett CL, Biggerstaff M, Chen J, et al. Forecasting
548 influenza activity using machine-learned mobility map. *Nat Commun* [Internet]. 2021 Feb 9
[cited 2024 Jun 4];12(1):726. Available from: <https://www.nature.com/articles/s41467-021-21018-5>
- 550 10. Balcan D, Gonçalves B, Hu H, Ramasco JJ, Colizza V, Vespignani A. Modeling the spatial
spread of infectious diseases: The GLObal Epidemic and Mobility computational model. *J*

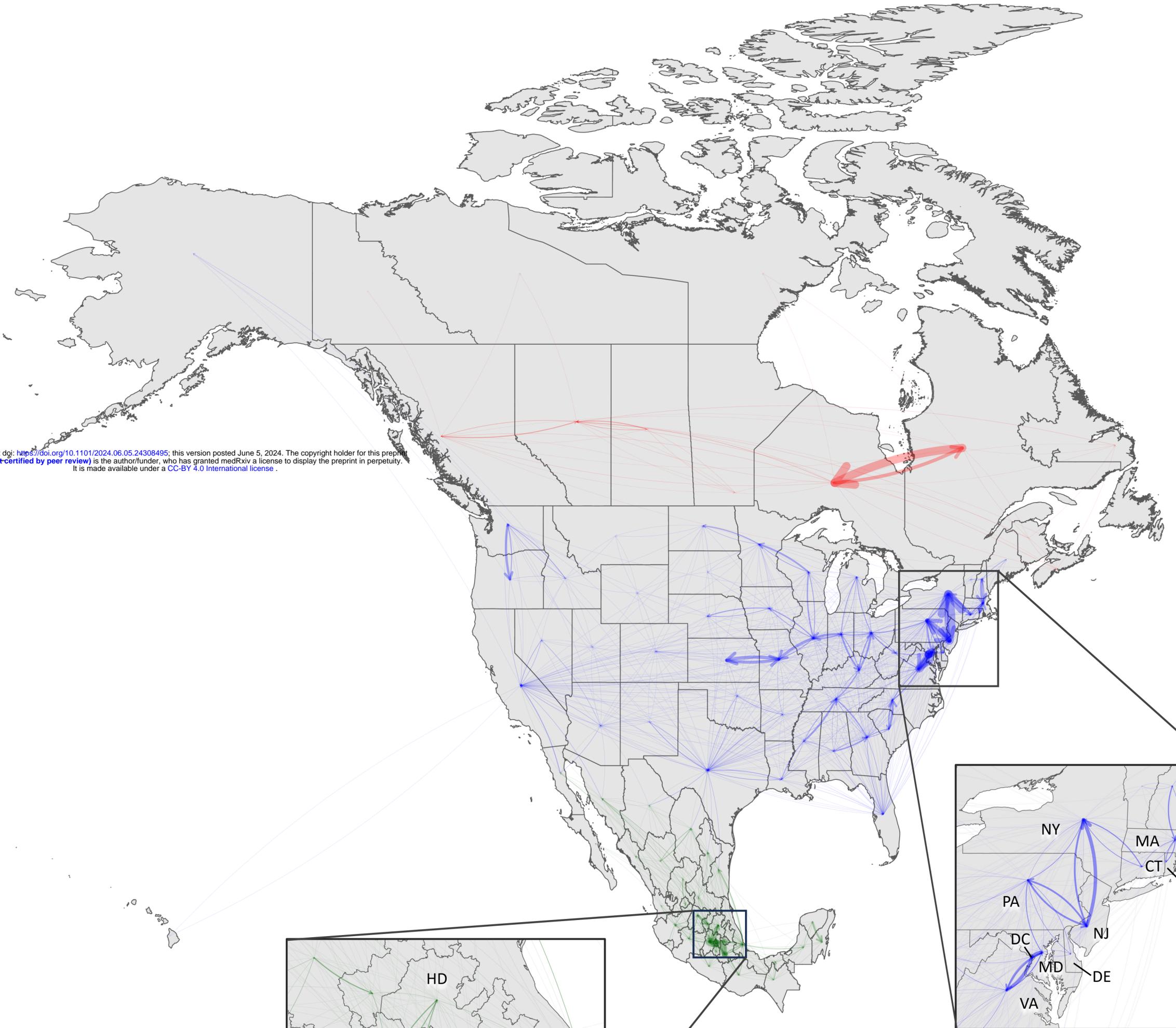
- 552 Comput Sci [Internet]. 2010 Aug [cited 2024 Jan 22];1(3):132–45. Available from:
553 <https://linkinghub.elsevier.com/retrieve/pii/S1877750310000438>
- 554 11. Pastore-Piontti A, Zhang Q, Gomes MFC, Rossi L, Poletto C, Colizza V, et al. Real-Time
555 Assessment of the International Spreading Risk Associated with the 2014 West African Ebola
556 Outbreak. In: Chowell G, Hyman JM, editors. Mathematical and Statistical Modeling for
557 Emerging and Re-emerging Infectious Diseases [Internet]. Cham: Springer International
558 Publishing; 2016 [cited 2024 Jan 23]. p. 39–56. Available from:
http://link.springer.com/10.1007/978-3-319-40413-4_4
- 560 12. Zhang Q, Sun K, Chinazzi M, Pastore Y Piontti A, Dean NE, Rojas DP, et al. Spread of Zika
561 virus in the Americas. Proc Natl Acad Sci [Internet]. 2017 May 30 [cited 2024 Jan 23];114(22).
562 Available from: <https://pnas.org/doi/full/10.1073/pnas.1620161114>
- 564 13. Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, et al. The effect of travel
565 restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science
566 [Internet]. 2020 Apr 24 [cited 2024 Jan 22];368(6489):395–400. Available from:
<https://www.science.org/doi/10.1126/science.aba9757>
- 568 14. Pei S, Yamana TK, Kandula S, Galanti M, Shaman J. Burden and characteristics of COVID-19
569 in the United States during 2020. Nature [Internet]. 2021 Oct 14 [cited 2024 Jan
570 19];598(7880):338–41. Available from: <https://www.nature.com/articles/s41586-021-03914-4>
- 572 15. CSSEGISandData. CSSEGISandData/COVID-19 [Internet]. COVID-19 Data Repository by
573 the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. 2024
574 [cited 2024 Jan 19]. Available from: <https://github.com/CSSEGISandData/COVID-19>
- 576 16. Statistics Canada. Population and Dwelling Count Highlight Tables, 2016 Census – Canada,
577 provinces and territories [Internet]. [cited 2024 May 13]. Available from:
[https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/hlt-fst/pd-
578 pl/Table.cfm?Lang=Eng&T=101&S=50&O=A](https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/hlt-fst/pd-pl/Table.cfm?Lang=Eng&T=101&S=50&O=A)
- 578 17. Bureau UC. Population [Internet]. Census.gov. [cited 2024 May 13]. Available from:
<https://www.census.gov/topics/population.html>
- 580 18. INEGI. Intercensal Survey 2015 [Internet]. [cited 2023 Dec 22]. Available from:
<https://en.www.inegi.org.mx/programas/intercensal/2015/#Microdatos>
- 582 19. Government of Canada SC. Commuting Flow from Geography of Residence to Geography of
583 Work - Census Metropolitan Areas and Census Agglomerations: Main Mode of Commuting
584 (10), Commuting Duration (6) and Sex (3) for the Employed Labour Force Aged 15 Years and
585 Over Having a Usual Place of Work, in Private Households, 2016 Census - 25% Sample Data
586 [Internet]. 2017 [cited 2024 Apr 3]. Available from: [https://www12.statcan.gc.ca/census-
587 recensement/2016/dp-pd/dt-td/Rp-
588 eng.cfm?TABID=4&LANG=E&A=R&APATH=3&DETAIL=0&DIM=0&FL=A&FREE=
589 0&GC=0&GL=-
590 1&GID=1354564&GK=0&GRP=1&O=D&PID=111333&PRID=10&PTYPE=109445&S=
0&SHOWALL=0&SUB=0&Temporal=2017&THEME=125&VID=0&VNAMEE=&VNA
MEF=%20\(2017\)&D1=0&D2=0&D3=0&D4=0&D5=0&D6=0](https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/dt-td/Rp-eng.cfm?TABID=4&LANG=E&A=R&APATH=3&DETAIL=0&DIM=0&FL=A&FREE=0&GC=0&GL=-1&GID=1354564&GK=0&GRP=1&O=D&PID=111333&PRID=10&PTYPE=109445&S=0&SHOWALL=0&SUB=0&Temporal=2017&THEME=125&VID=0&VNAMEE=&VNAMEF=%20(2017)&D1=0&D2=0&D3=0&D4=0&D5=0&D6=0)

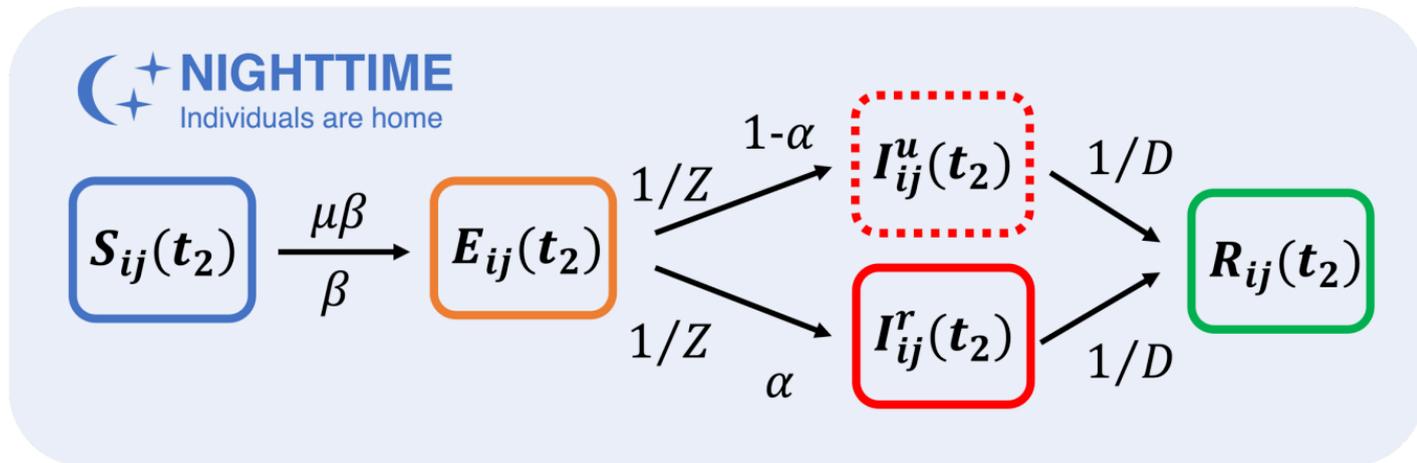
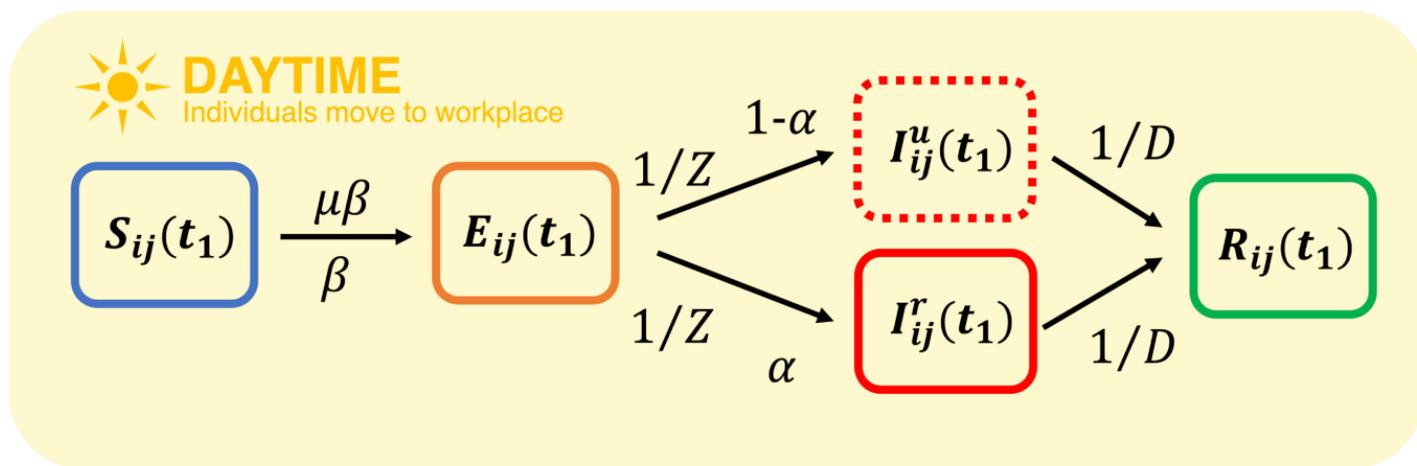
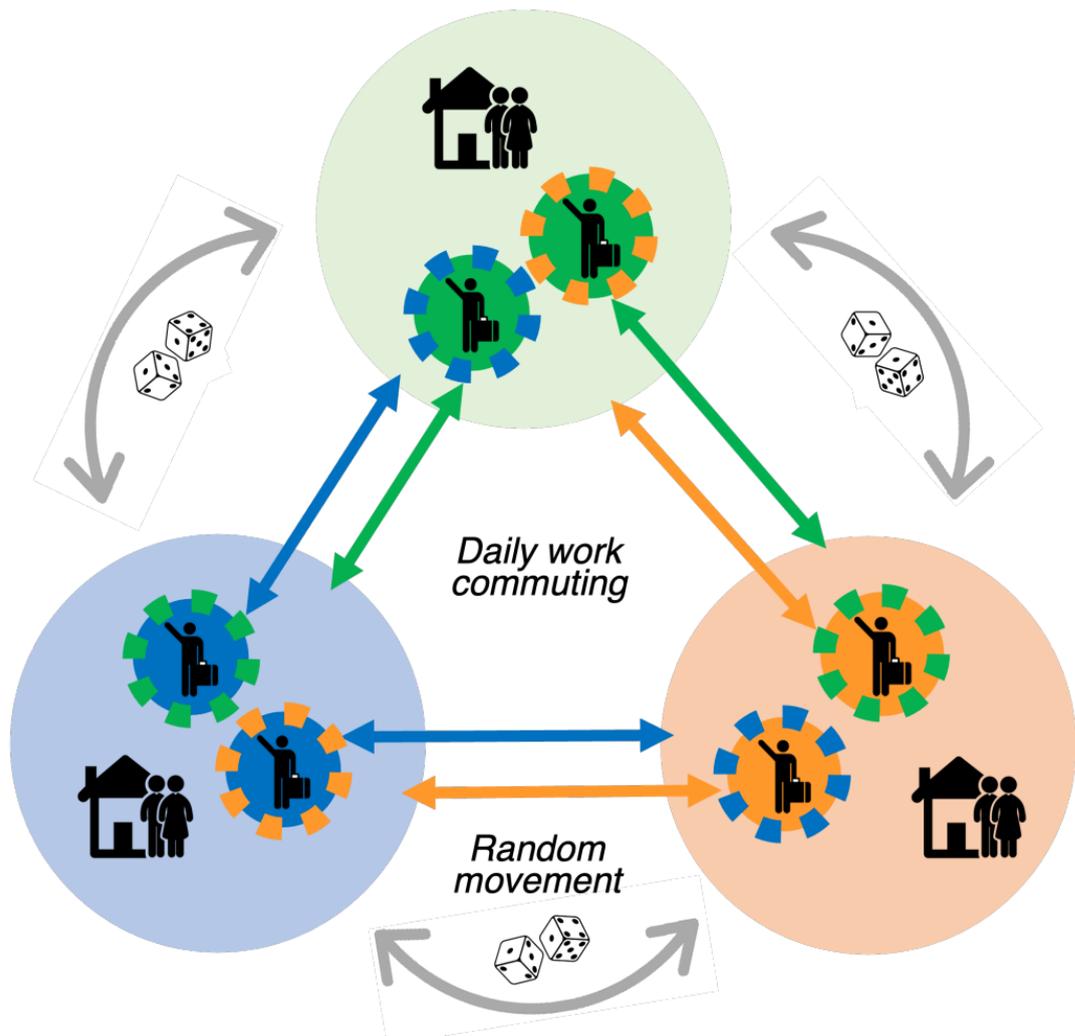
- 592 20. Statistics Canada. Number of vehicles travelling between Canada and the United States
[Internet]. Government of Canada; [cited 2023 Dec 22]. Available from:
594 <https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=2410000201>
21. US Census Bureau. 2011-2015 5-Year ACS Commuting Flows [Internet]. [cited 2023 Dec 22].
596 Available from: <https://www.census.gov/data/tables/2015/demo/metro-micro/commuting-flows-2015.html>
- 598 22. COVID-19 Community Mobility Reports [Internet]. [cited 2023 Dec 22]. Available from:
<https://www.google.com/covid19/mobility/>
- 600 23. Diekmann O, Heesterbeek JAP, Roberts MG. The construction of next-generation matrices for
compartmental epidemic models. *J R Soc Interface* [Internet]. 2010 Jun 6 [cited 2024 Jan
602 23];7(47):873–85. Available from:
<https://royalsocietypublishing.org/doi/10.1098/rsif.2009.0386>
- 604 24. COVID-19 Open Data — Google Health [Internet]. [cited 2023 Dec 22]. Available from:
<https://health.google.com/covid-19/open-data/data-sources>
- 606 25. Anderson JL. An Ensemble Adjustment Kalman Filter for Data Assimilation. *Mon Weather Rev*
[Internet]. 2001 Dec [cited 2023 Dec 22];129(12):2884–903. Available from:
608 [http://journals.ametsoc.org/doi/10.1175/1520-0493\(2001\)129<2884:AEAKFF>2.0.CO;2](http://journals.ametsoc.org/doi/10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2)
- 610 26. Shaman J, Karspeck A, Yang W, Tamerius J, Lipsitch M. Real-time influenza forecasts during
the 2012–2013 season. *Nat Commun* [Internet]. 2013 Dec 3 [cited 2024 Jan 9];4(1):2837.
Available from: <https://www.nature.com/articles/ncomms3837>
- 612 27. DeFelice NB, Little E, Campbell SR, Shaman J. Ensemble forecast of human West Nile virus
cases and mosquito infection rates. *Nat Commun* [Internet]. 2017 Feb 24 [cited 2024 Jan
614 9];8(1):14592. Available from: <https://www.nature.com/articles/ncomms14592>
- 616 28. Yang W, Karspeck A, Shaman J. Comparison of Filtering Methods for the Modeling and
Retrospective Forecasting of Influenza Epidemics. Ferguson NM, editor. *PLoS Comput Biol*
[Internet]. 2014 Apr 24 [cited 2024 Jan 23];10(4):e1003583. Available from:
618 <https://dx.plos.org/10.1371/journal.pcbi.1003583>
- 620 29. Pei S, Kandula S, Yang W, Shaman J. Forecasting the spatial transmission of influenza in the
United States. *Proc Natl Acad Sci* [Internet]. 2018 Mar 13 [cited 2024 Jan 23];115(11):2752–7.
Available from: <https://pnas.org/doi/full/10.1073/pnas.1708856115>
- 622 30. Dahal S, Banda JM, Bento AI, Mizumoto K, Chowell G. Characterizing all-cause excess
mortality patterns during COVID-19 pandemic in Mexico. *BMC Infect Dis* [Internet]. 2021
624 Dec [cited 2024 Mar 20];21(1):432. Available from:
<https://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-021-06122-7>
- 626 31. Hasell J, Mathieu E, Beltekian D, Macdonald B, Giattino C, Ortiz-Ospina E, et al. A cross-
country database of COVID-19 testing. *Sci Data* [Internet]. 2020 Oct 8 [cited 2024 Mar
628 20];7(1):345. Available from: <https://www.nature.com/articles/s41597-020-00688-8>

- 630 32. Lee MHY, Xu G, Cheng F, Khalid AF. Testing surge capacity—A Canadian COVID-19
experience, Ontario’s surge capacity for the first wave. *Health Policy* [Internet]. 2021 Oct [cited
632 2024 Mar 26];125(10):1291–6. Available from:
<https://linkinghub.elsevier.com/retrieve/pii/S0168851021002025>
- 634 33. Jones JM, Stone M, Sulaeman H, Fink RV, Dave H, Levy ME, et al. Estimated US Infection-
and Vaccine-Induced SARS-CoV-2 Seroprevalence Based on Blood Donations, July 2020-May
636 2021. *JAMA* [Internet]. 2021 Oct 12 [cited 2024 Jan 12];326(14):1400. Available from:
<https://jamanetwork.com/journals/jama/fullarticle/2784013>
- 638 34. The COVID Tracking Project [Internet]. The COVID Tracking Project. [cited 2024 Feb 26].
Available from: <https://covidtracking.com/data/national>
- 640 35. Canada PHA of. COVID-19 daily epidemiology update: Testing and variants [Internet]. aem.
2020 [cited 2024 Feb 26]. Available from: <https://www.canada.ca/en.html>
- 642 36. Mathieu E, Ritchie H, Rodés-Guirao L, Appel C, Giattino C, Hasell J, et al. Coronavirus
Pandemic (COVID-19). *Our World Data* [Internet]. 2020 Mar 5 [cited 2024 Feb 26]; Available
from: <https://ourworldindata.org/coronavirus-testing>
- 644 37. Dirección General de Epidemiología. Lineamiento estandarizado para la vigilancia
646 epidemiológica y por laboratorio de enfermedad por 2019-nCoV. 2020 [Internet]. Available
from: [https://www.gob.mx/salud/documentos/lineamiento-estandarizado-para-la-vigilancia-
epidemiologica-y-por-laboratorio-de-la-enfermedad-respiratoria-viral](https://www.gob.mx/salud/documentos/lineamiento-estandarizado-para-la-vigilancia-epidemiologica-y-por-laboratorio-de-la-enfermedad-respiratoria-viral)
- 648 38. De La Cruz-Hernández SI. Another Vision of the Situation of the COVID-19 Pandemic in
Mexico During 2020. *Disaster Med Public Health Prep* [Internet]. 2022 Dec [cited 2024 Apr
650 23];16(6):2296–8. Available from:
652 [https://www.cambridge.org/core/product/identifier/S1935789321003402/type/journal_articl
e](https://www.cambridge.org/core/product/identifier/S1935789321003402/type/journal_article)
- 654 39. Davis JT, Chinazzi M, Perra N, Mu K, Pastore y Piontti A, Ajelli M, et al. Cryptic transmission
of SARS-CoV-2 and the first COVID-19 wave. *Nature* [Internet]. 2021 Dec [cited 2024 May
1];600(7887):127–32. Available from: <https://www.nature.com/articles/s41586-021-04130-w>
- 656 40. Ferretti L, Wymant C, Petrie J, Tsallis D, Kendall M, Ledda A, et al. Digital measurement of
SARS-CoV-2 transmission risk from 7 million contacts. *Nature* [Internet]. 2024 Feb [cited
658 2024 Mar 29];626(7997):145–50. Available from: [https://www.nature.com/articles/s41586-
023-06952-2](https://www.nature.com/articles/s41586-023-06952-2)
- 660 41. Balcan D, Colizza V, Gonçalves B, Hu H, Ramasco JJ, Vespignani A. Multiscale mobility
662 networks and the spatial spreading of infectious diseases. *Proc Natl Acad Sci* [Internet]. 2009
Dec 22 [cited 2024 Mar 29];106(51):21484–9. Available from:
<https://www.pnas.org/doi/10.1073/pnas.0906910106>

664

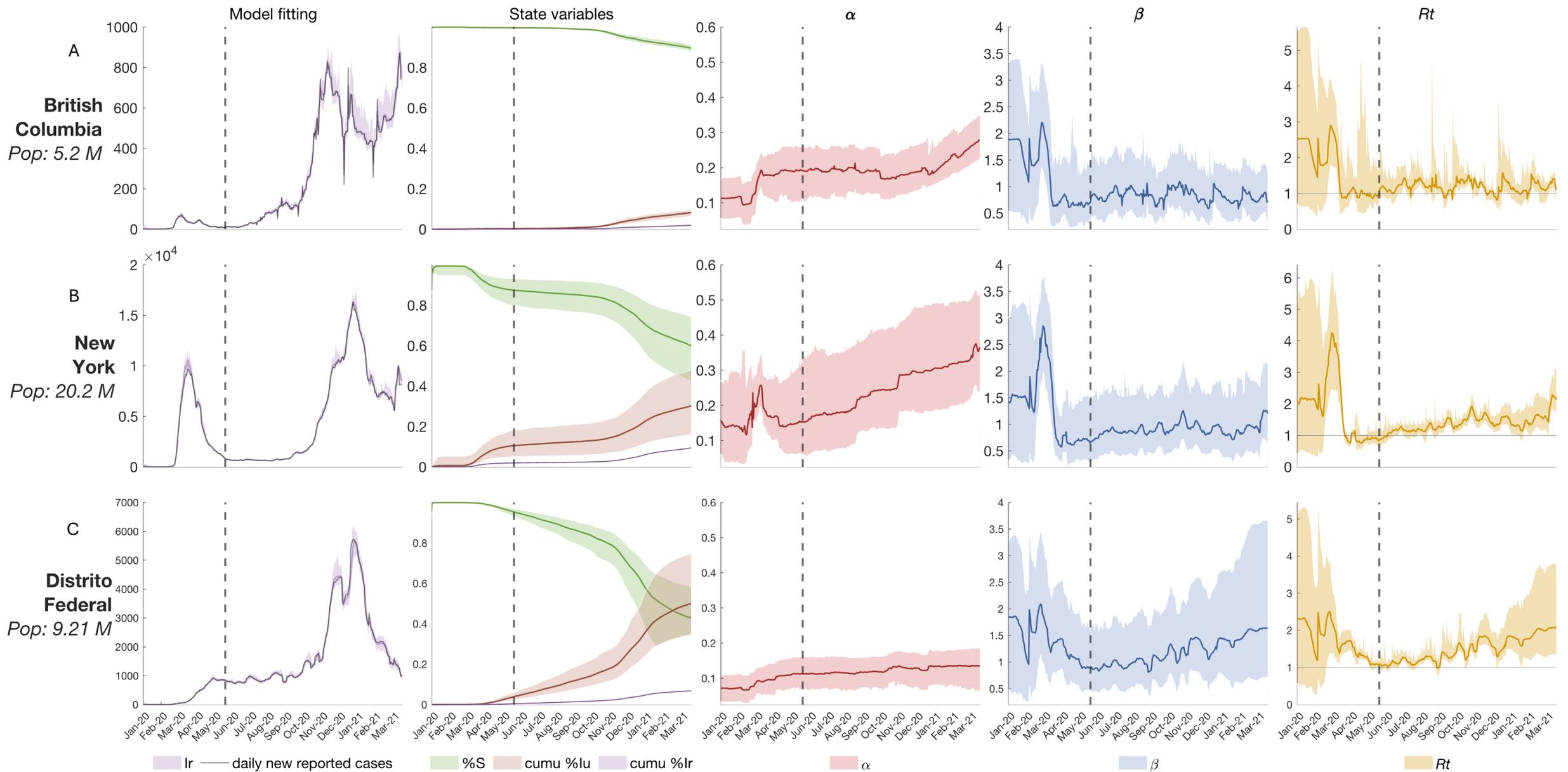
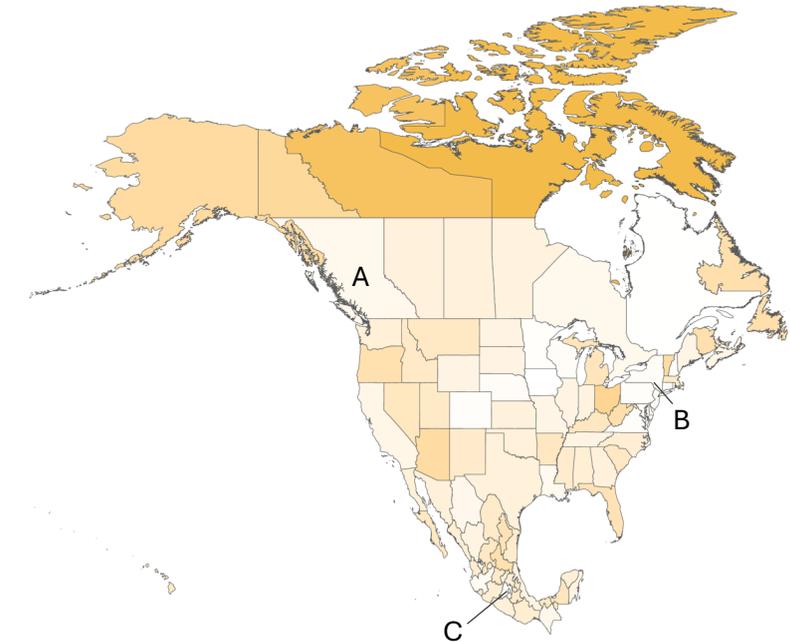
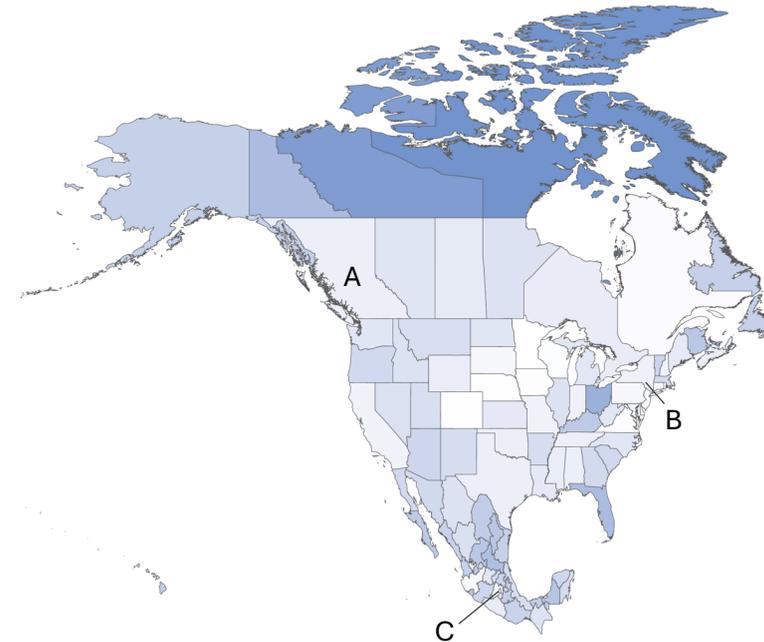
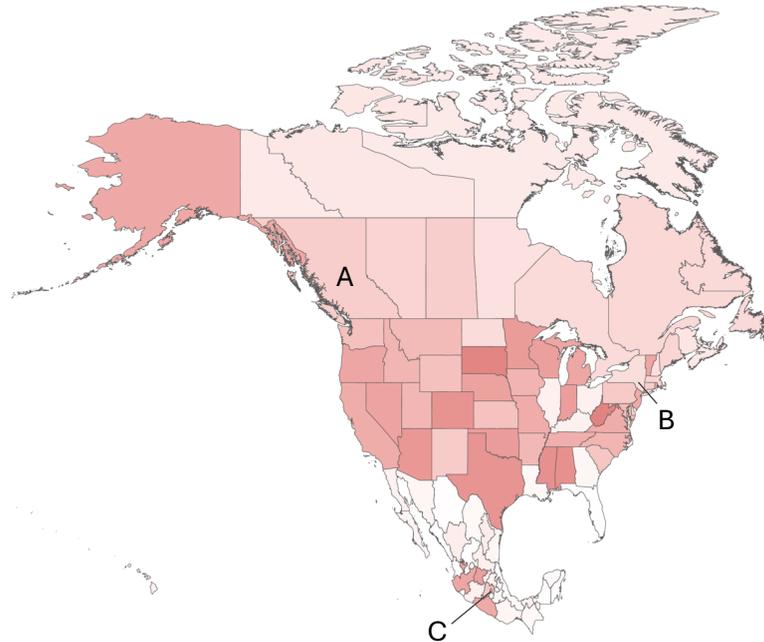
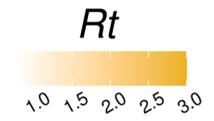
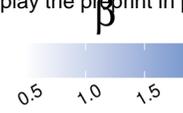
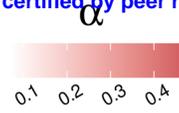
medRxiv preprint doi: <https://doi.org/10.1101/2024.06.05.24308495>; this version posted June 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).





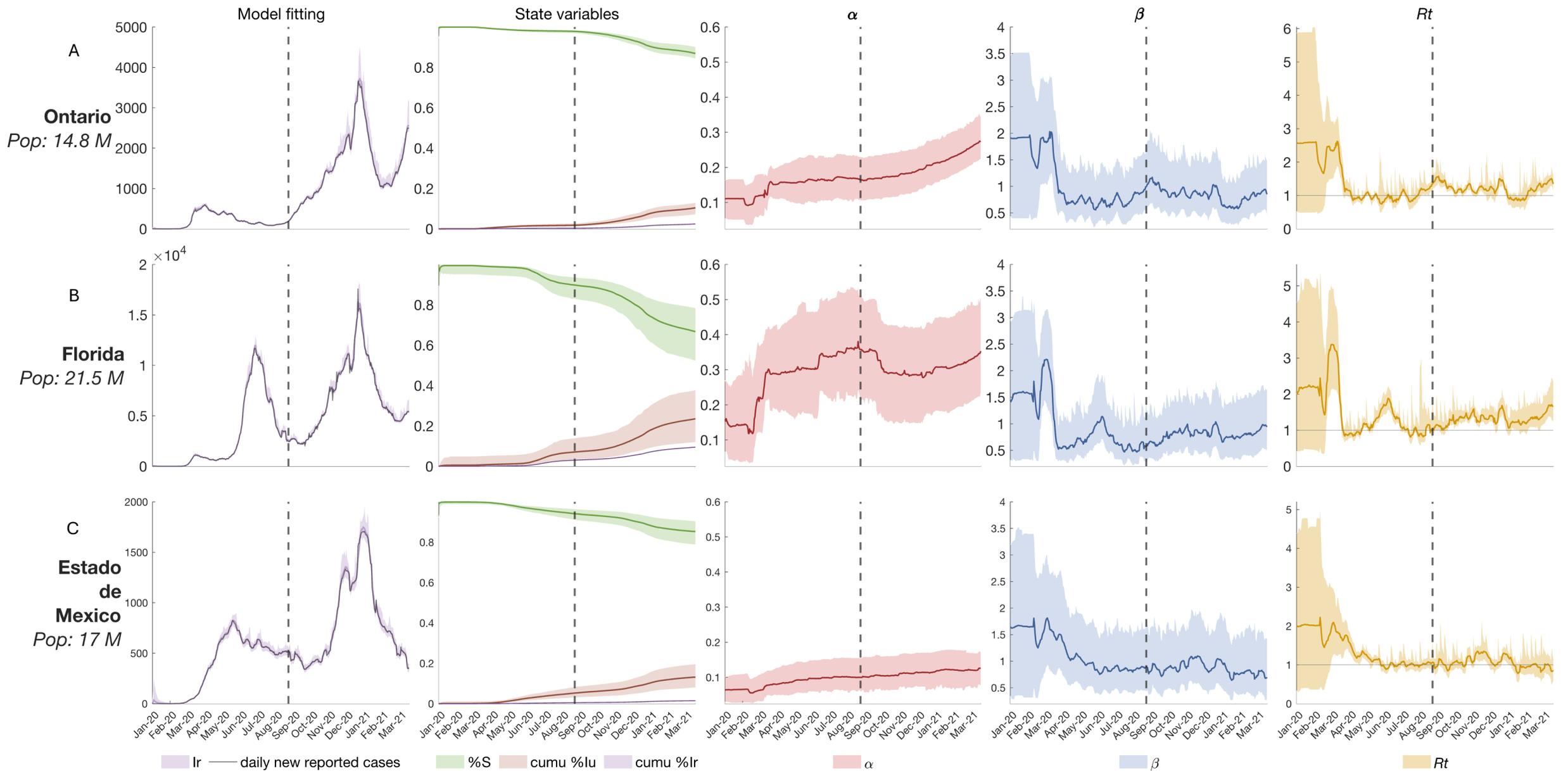
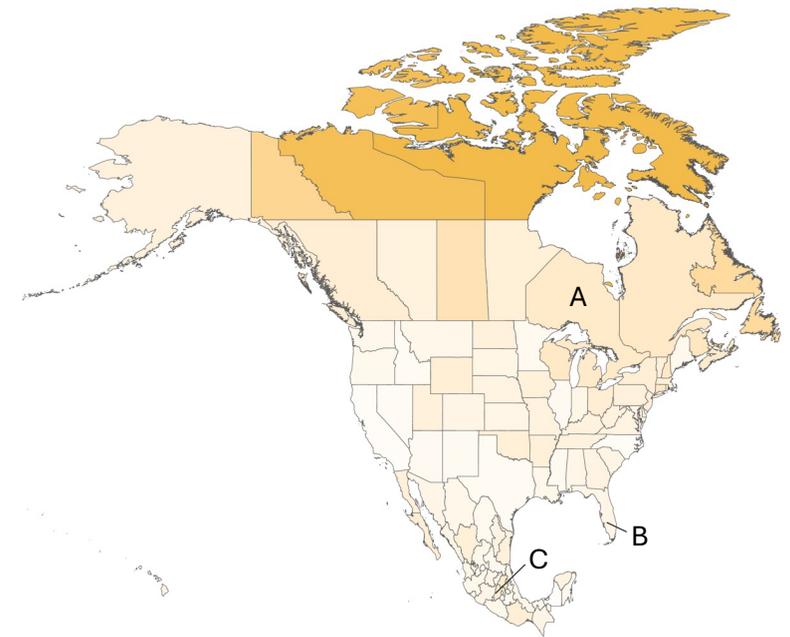
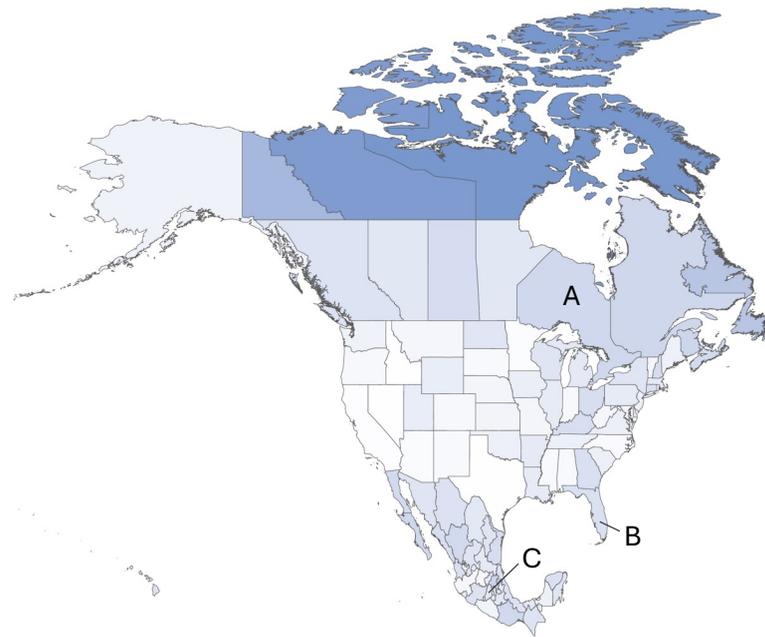
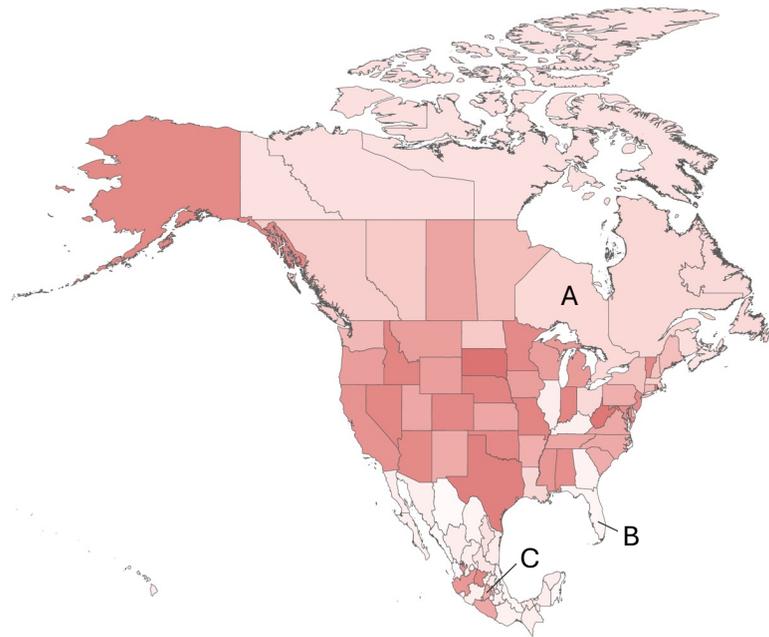
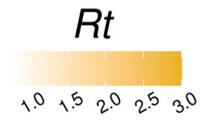
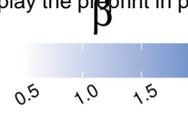
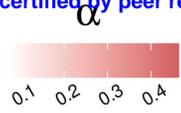
6th June 2020

medRxiv preprint doi: <https://doi.org/10.1101/2024.06.05.24308495>; this version posted June 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).



7th September 2020

medRxiv preprint doi: <https://doi.org/10.1101/2024.06.05.24308495>; this version posted June 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).



15th March 2021

medRxiv preprint doi: <https://doi.org/10.1101/2024.06.05.24308495>; this version posted June 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

