Supplementary Information

Estimating the Impact of Human Mobility Restrictions on COVID-19 Transmission in Lagos, Nigeria: A Hybrid Simulation Framework Integrating Mobility and Epidemiological Data

Kailun Liu[1,+], Xin Wu[1,+], Meagan C. Fitzpatrick[2], Weiyu Luo[1], Olanrewaju Lawal[3], Natalia Blanco[2], Kristen Stafford[2], Ruohan Li[1], Lele Zhang[1], Alash'le Abimiku[2,4] & Chenfeng Xiong[1,*]

[1]Department of Civil and Environmental Engineering, Villanova University, Villanova, PA, United States
[2]School of Medicine, University of Maryland Baltimore, Baltimore, Maryland, United States of America
[3]Department of Geography and Environmental Management, University of Port Harcourt, Port Harcourt, Nigeria
[4]Institute of Human Virology Nigeria, Abuja, Nigeria

[*]corresponding.chenfeng.xiong@villanova.edu
[+]These authors contributed equally to this work

List of Supplementary Items

Supplementary Table 2 Distance-bin distribution of trips by LGA from Janunary to Febuary, 2020).

Supplementary Table 3 Distance-bin distribution of trips by LGA from Match to April 2020).

Supplementary Table 4 Relative percentage change in distance-bin shares between Mat-Apr and the Jan–Feb periods.

Supplementary Table 5 Constant and calibrated model parameters for SEAPIR model and Bayesian optimization.

Supplementary Note 1| Study area and population data

We discretize Lagos State using a geohash-6 spatial resolution (approximately 1.22 km × 0.61 km), yielding a total of 5,367 cells across the state. Geohash is a hierarchical geocode system that encodes geographic locations into short alphanumeric strings for spatial indexing and aggregation[1]. To estimate the spatial distribution of population at the geohash-6 level, we utilize the WorldPop 2020 gridded population dataset and spatially align it with the geohash-6 grid. WorldPop provides up-to-date, high-resolution population estimates derived from satellite imagery and census-based projections, calibrated to match the 2020 estimates of the United Nations Population Division (UNPD)[2]. According to this dataset, the total estimated population of Lagos State in 2020 is 12,765,635.

Population counts for each geohash-6 cell in the geohash-6 set $\mathscr{G}$ are computed using an area-weighted allocation approach, in which population values from overlapping WorldPop raster cells are proportionally assigned to a geohash cell according to the fraction of raster cell area intersecting that geohash polygon. Specifically, the population of geohash cell g is calculated as

$$Z_g = \sum_{r \in \mathscr{R}} Z_r \cdot \frac{\text{Area}_{r \cap g}}{\text{Area}_r}, \qquad \forall g \in \mathscr{G}, \tag{1}$$

where Z_r denotes the population of WorldPop raster cell $r \in \mathscr{R}$, $\text{Area}_{r\cap g}$ is the area of overlap between raster cell r and geohash cell g, and Area_r is the total area of raster cell r. Geohash-6 cells that are fully located on water bodies are assigned zero population values. We further aggregate the geohash-6 population to the 20 Local Government Areas (LGAs), denoted by the set $\mathscr{L}$, using a majority-area assignment rule: each geohash-6 cell is assigned to the LGA that contains the largest proportion of its area, and its population is aggregated accordingly. The population of LGA l is then given by

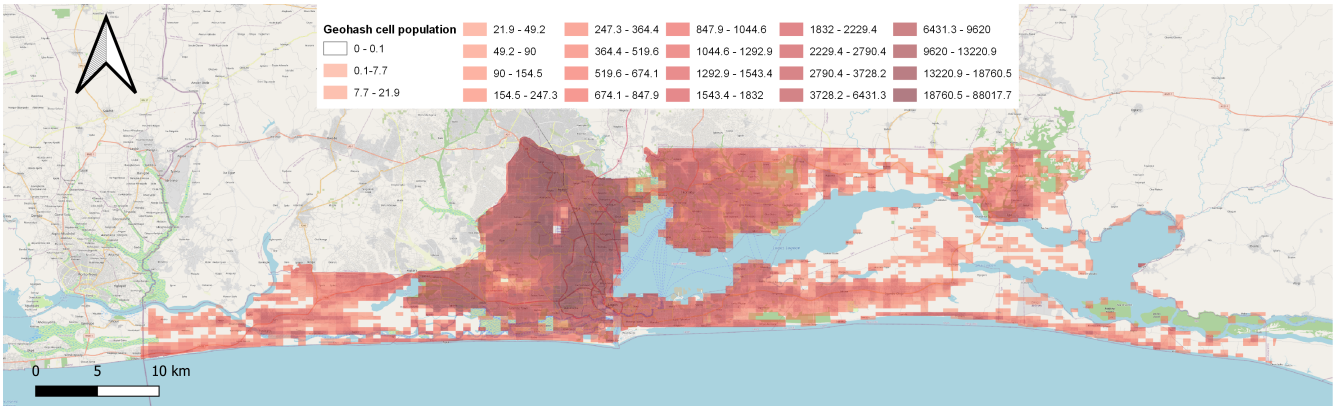
$$Z_l = \sum_{g \in \mathscr{G}_l} Z_g + \phi_l, \qquad \forall l \in \mathscr{L}, \tag{2}$$

where $\mathscr{G}_l \subset \mathscr{G}$ is the set of geohash-6 cells assigned to LGA l. ϕ_l is a correction factor used to ensure consistent population aggregation when geohash-6 cells span multiple LGAs and are allocated according to area overlap, thereby avoiding overcounting and under-counting.

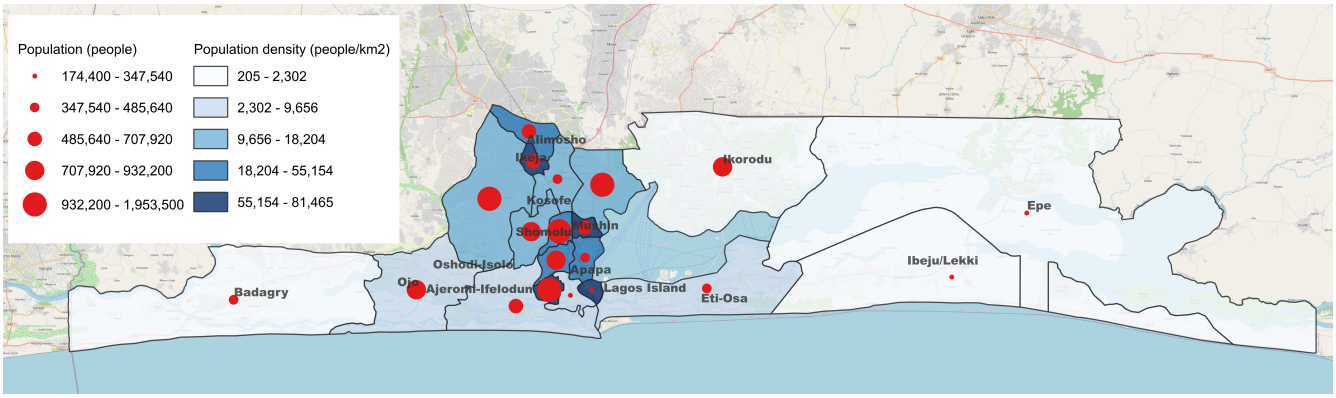
The population assigned to each geohash-6 cell is used to control the total number of synthetic agents initiate from that geohash cell in the simulation, while the aggregated LGA-level population estimates serve as inputs to the gravity model for estimating cross-LGA mobility flows across the state (see details in **Supplementary Note 9**, *LBS-based trip-level gravity model*). Supplementary Fig. 1 a illustrates the spatial distribution of population at the geohash-6 level. The heat map highlights the concentration of residents within the Lagos metropolitan area. Supplementary Fig. 1 b presents the estimated population distribution across the 20 LGAs after aggregating the geohash-6 population to the LGA polygons.

Supplementary Note 2 | Stay-put share analysis

To evaluate whether the LBS data used in this study provide a representative characterization of population-level mobility dynamics, we compared LBS-derived within-cell stay-put shares with the Meta Movement Range baseline[3] across 20 LGAs in Lagos State in Supplementary Fig.2. In Supplementary Table 1, Pearson correlation (PCR) and Spearman correlation (SCR) coefficients and mean L1 are computed. Across the 20 LGAs, the correlation coefficients span a wide range, from 0.28 to 0.90. The median PCR and SCR are approximately 0.74 and 0.62, respectively. Notably, 17 out of 20 LGAs exhibit PCR coefficients exceeding 0.6, indicating a strong and consistent temporal alignment between the two independent mobility data sources. Higher correlations are generally observed in central and high-density LGAs, while lower correlations occur primarily in peripheral areas with lower population density and reduced data coverage, such as Badagry and Epe. The consistency between PCR and SCR results further reflects robust temporal alignment in stay-put behaviors of the two datasets. We also report the ratio of within-cell stay-put shares between the two datasets in the table. Across LGAs, the ratio typically falls in the range of approximately 40%–70% which is consistent with differences in spatial aggregation, as Meta mobility is measured on 0.6 m × 0.6 m tiles, whereas the LBS data are aggregated on coarser geohash-6 cells (approximately 1.22 km × 0.61 km), which cover roughly twice the area. This also indicates that LBS-derived mobility achieves penetration levels comparable in magnitude to Meta mobility once normalized by spatial unit.



(a)



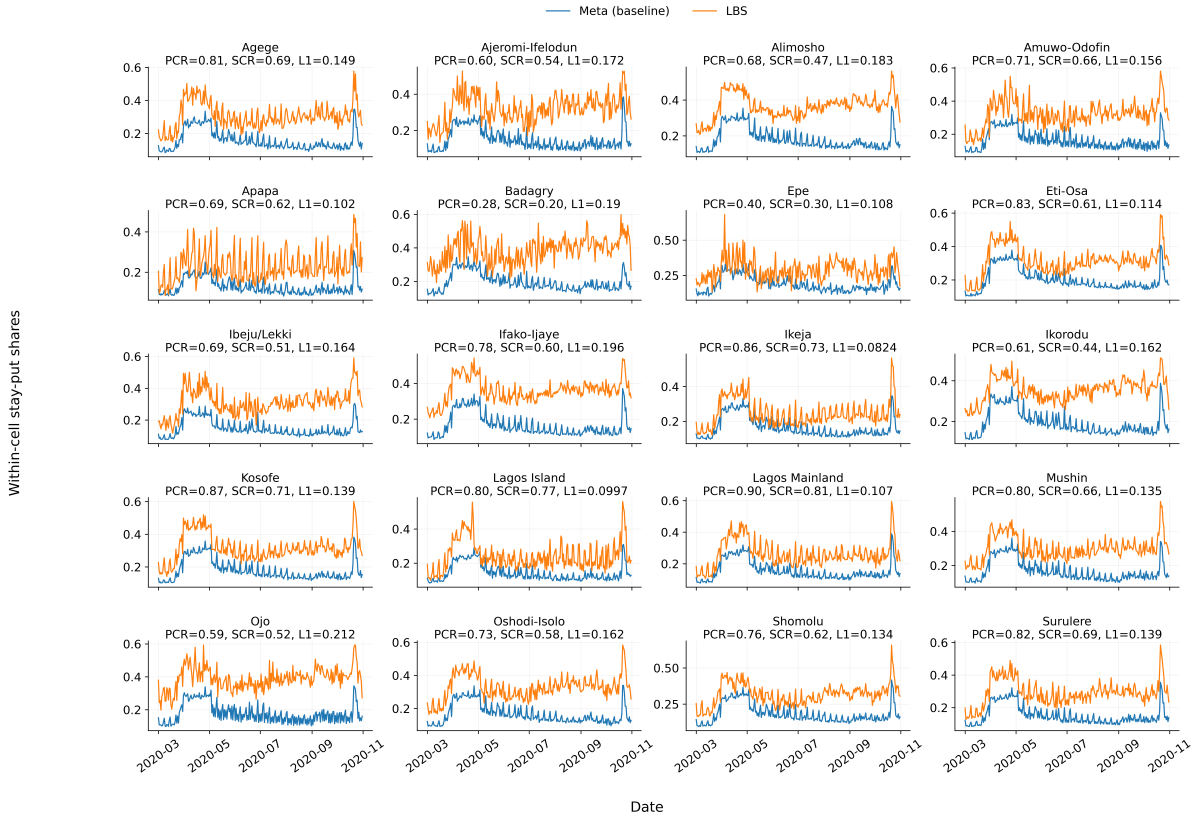
(b)

Supplementary Figure 1. Population spatial distribution in Lagos, Nigeria. a, Population distribution on Geohash-6 cells in Lagos, Nigeria, colored by assigned population derived from WorldPop 2020 (unit: people)[2] **b**, Population (unit: people) and its density (unit: people/km^2) distribution on LGAs in Lagos, Nigeria.

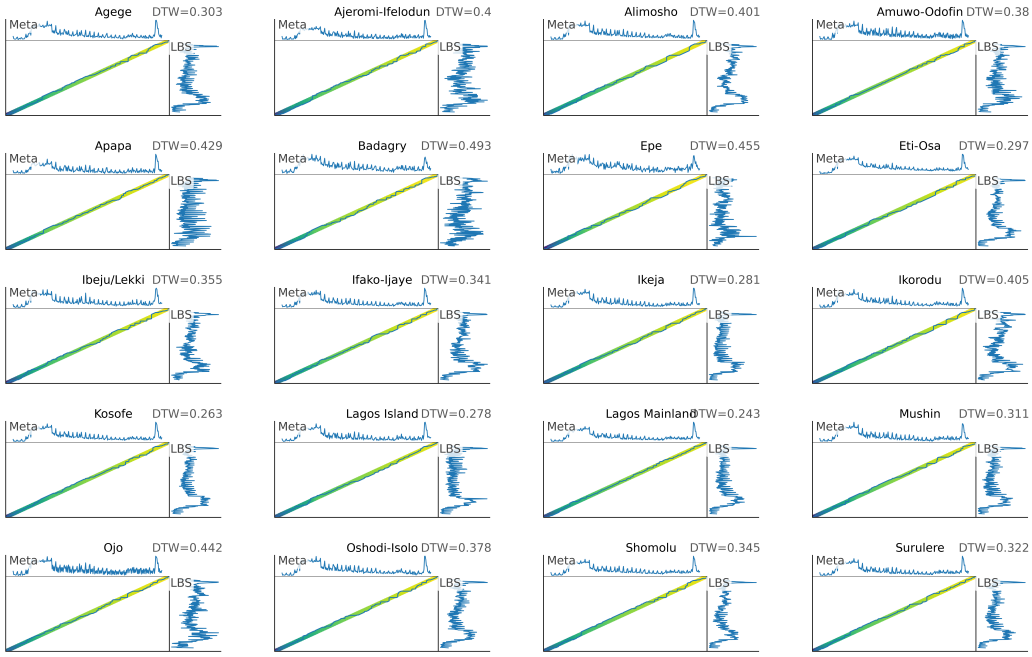
While correlation-based metrics (PCR and SCR) and mean L1 distance assess temporal agreement under pointwise alignment, they do not account for potential local temporal shifts between the two datasets. To further evaluate the robustness of temporal consistency, we conducted a dynamic time warping (DTW) analysis of within-cell stay-put share time series. The 7-day rolling window is used in the analysis and both Meta and LBS time series were subsequently standardized using z-score normalization within each LGA to remove scale differences. In Supplementary Fig. 3. The DTW distance which represents the expected absolute deviation between the two standardized time series along the optimal alignment path, measured in units of standard deviation are calculated using dynamic programming algorithm for each LGA. For example, a DTW distance of 0.43 indicates that, after temporal alignment, the two series differ on average by approximately 0.43 standard deviations. Across LGAs, DTW distances are consistently small, typically below 0.5, indicating that after allowing for local temporal misalignment, the two datasets differ on average by less than half a standard deviation. Together with the PCR and SCR coefficients and the small mean L1 distances, the consistently low DTW distances provide converging evidence that LBS data could capture the stay-put dynamics across LGAs.

Supplementary Note 3 | Cell visits analysis

We apply the same normalization to the LBS trips in the trip roster data by computing relative changes in the number of geohash-6 cells visited per individual from Mar. 1 to Oct. 31, 2020. Supplementary Fig. 4 compares baseline-normalized mobility trajectories averaged across LGAs. Relative to the Meta Movement Range data, LBS-derived mobility exhibits smaller variance and less extreme declines. This difference reflects structural distinctions between the two representations: (i) coarser spatial aggregation at the geohash-6 level smooths visit-count fluctuations compared with Bing tile level 16, and (ii) the NextGen NHTS–aligned, trip-based reconstruction emphasizes salient travel while filtering short-range or incidental



Supplementary Figure 2. Comparison of within-cell stay-put shares derived from Meta baseline and LBS data across LGAs in Lagos State. Time series of within-cell stay-put shares estimated from the Meta Movement Range baseline (blue) and LBS data (orange) are shown for 20 LGAs in Lagos State between March and October 2020. Each panel corresponds to one LGA and reports the Pearson correlation (PCR) coefficient, Spearman rank correlation (SCR) coefficient, and mean L1 distance between the two series. Here, the mean L1 distance quantifies the average absolute deviation between the two time series over time.



Supplementary Figure 3. DTW-based validation of temporal consistency between Meta baseline and LBS stay-put shares across Lagos LGAs. DTW comparisons of within-cell stay-put share time series derived from the Meta Movement Range baseline and LBS data are shown for 20 LGAs in Lagos State. Both time series were smoothed using a 7-day rolling window and standardized using z-score normalization within each LGA to remove scale differences. For each LGA, the top and right subpanels display the original Meta and LBS time series, respectively, while the central panel illustrates the DTW alignment path between the two standardized series. Reported DTW distances quantify the expected absolute deviation along the optimal alignment path and are measured in units of standard deviation.

Supplementary Table 1. Cross-validation of LBS-derived mobility against Meta Movement Range data across Lagos State LGAs (March 1–October 31, 2020). Pearson and Spearman correlation coefficients quantify temporal agreement between the two independent mobility datasets. Stay-put shares are defined as a ratio of active user proportions measured within Bing tiles (Facebook data) or within geohash-6 cells (Our proposed LBS data)

	Pearson	Spearman	Ratio of within-cell stay-put shares (Facebook tile / LBS geohash)	Pop. Density (persons/km^2)
Agege	0.81	0.69	57%	55,804
Ajeromi/Ifelodun	0.60	0.54	50%	81,465
Alimosho	0.68	0.47	52%	10,646
Amuwo Odofin	0.71	0.66	50%	4,264
Apapa	0.69	0.62	56%	8,171
Badagary	0.28	0.20	49%	790
Epe	0.40	0.30	67%	205
Eti-Osa	0.83	0.61	59%	2,402
Ibeju/Lekki	0.69	0.52	58%	419
Ifako/Ijaye	0.78	0.60	46%	19,658
Ikeja	0.86	0.73	66%	11,113
Ikorodu	0.61	0.44	55%	1,900
Kosofe	0.87	0.72	57%	16,849
LagosIsland	0.80	0.77	59%	62,381
Mainland	0.90	0.81	62%	23,964
Mushin	0.80	0.66	61%	54,991
Ojo	0.59	0.52	40%	5,222
Oshodi/Isolo	0.73	0.58	52%	17,235
Shomolu	0.76	0.62	56%	58,000
Surulere	0.82	0.70	62%	37,201
Median	0.74	0.62	57%	N/A
Mean	0.71	0.59	56%	N/A

Notes: Population density is calculated using population estimates from WorldPop report and[4], together with LGA area information[5].

movements.

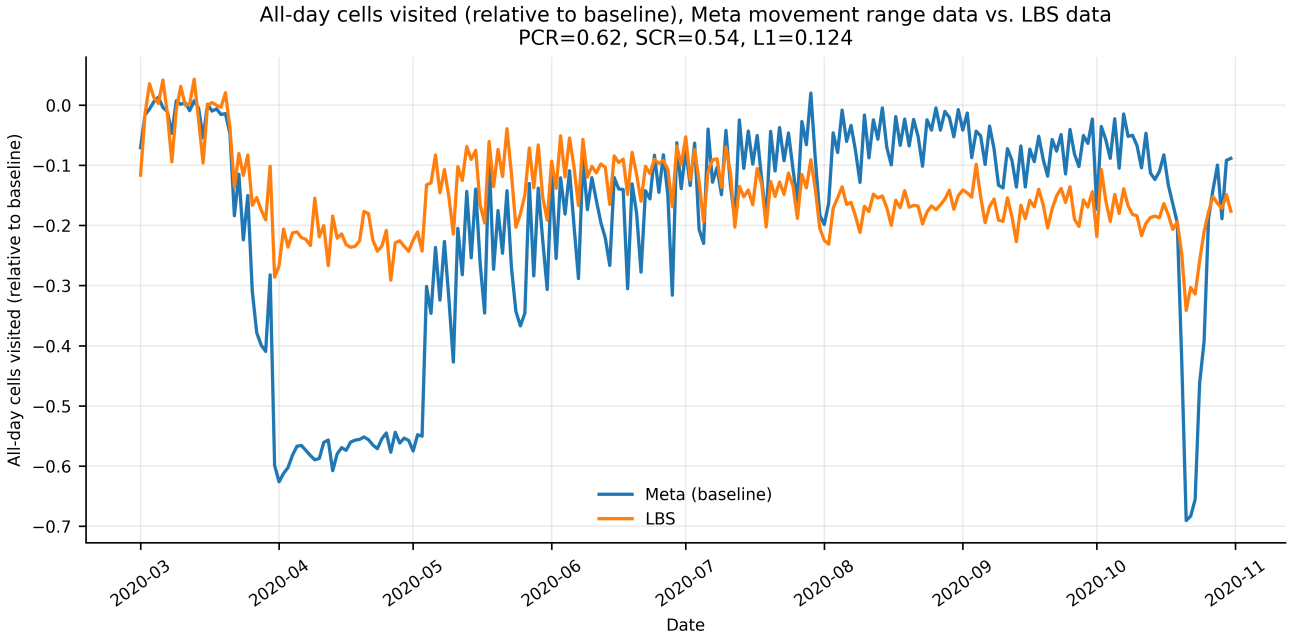
Despite these differences, the two datasets display consistent co-varying temporal patterns (PCR = 0.62; SCR = 0.54), capturing the sharp mobility reductions during early lockdown and the gradual recovery thereafter. The mean L1 distance remains small, indicating close alignment during pre-lockdown and reopening phases. Supplementary Fig. 5 further presents LGA-level comparisons of baseline-normalized changes in all-day cells visited derived from the Meta Movement Range Maps and the LBS trip roster. PCR values exceed 0.5 in 12 of the 20 LGAs. Compared with within-cell stay-put shares, visit-based mobility indicators show weaker agreement across datasets, as reflected by generally lower PCR and SCR values. This outcome is expected, as visit counts are more sensitive to changes in trip length and spatial resolution.

Supplementary Note 4 | Cross-LGA tours across LGAs in Lagos State

After identify the tour templates, it is important to characterize cross LGA mobility patterns. We compute the daily proportion of cross-LGA tours across all tours using LBS-derived trip roster dataset. To fully capture the temporal variations, we use the proportions **before** the tours are pooled to two-week windows to generate daily representative tour templates. These proportions are used to determine the relative shares of three tour types: (i) within-LGA tours, (ii) cross-LGA tours, and (iii) within-geohash tours. These ratios serve as key parameters in the tour-flow estimation process (see **Supplementary Note 9** , *Tour-based flow estimation*, for details). Supplementary Fig. 6 shows the time series for cross-LGA tour proportion from January to October 2020.

Supplementary Note 5 | Cross-LGA travel demand estimation and temporal variation.

Cross-LGA origin–destination (OD) travel demand was estimated using a gravity model. The model is calibrated on a weekly basis, and the resulting cross-LGA OD demand is fitted to LBS-derived trip observations. Supplementary Fig. 7 presents the resulting OD demand between LGAs after fitting to LGA-level population and trip rates (see **Supplementary Note 9**, *LBS-based trip-level gravity model*, for details). This procedure captures the structural patterns of mobility flows between LGAs and provides population-level constraints for cross-resolution mobility reconstruction. Supplementary Fig. 8 further presents gravity-model-estimated OD volumes aggregated by month from January to October 2020, highlighting persistent inter-LGA connectivity alongside NPI-induced temporal mobility reductions, particularly during April.



Supplementary Figure 4. Baseline-normalized mobility change based on visited spatial units (February baseline). Daily time series of mobility change normalized to a February day-of-week baseline, averaged across all LGAs in Lagos State, comparing Meta Movement Range data (Bing tiles) and LBS-derived mobility (geohash-6 cells) from March to October 2020. The PCR, SCR, and mean L1 distance between the two series are reported.

The estimated cross-LGA OD flows are used as constraints to estimate tour flows for each representative tour template. It is important to note that representative tour templates are derived by pooling individual observed tours within multiple rolling two-week time windows over the simulation period. Geohash-6–level population constraints and the proportions of different tour types are jointly embedded to estimate daily tour flows for each representative tour template (see **Supplementary Note 9**, *Tour-based flow estimation*, for details). Supplementary Fig. 9 demonstrates strong consistency between gravity-model–estimated cross-LGA OD demand and the estimated tour flows across all months.

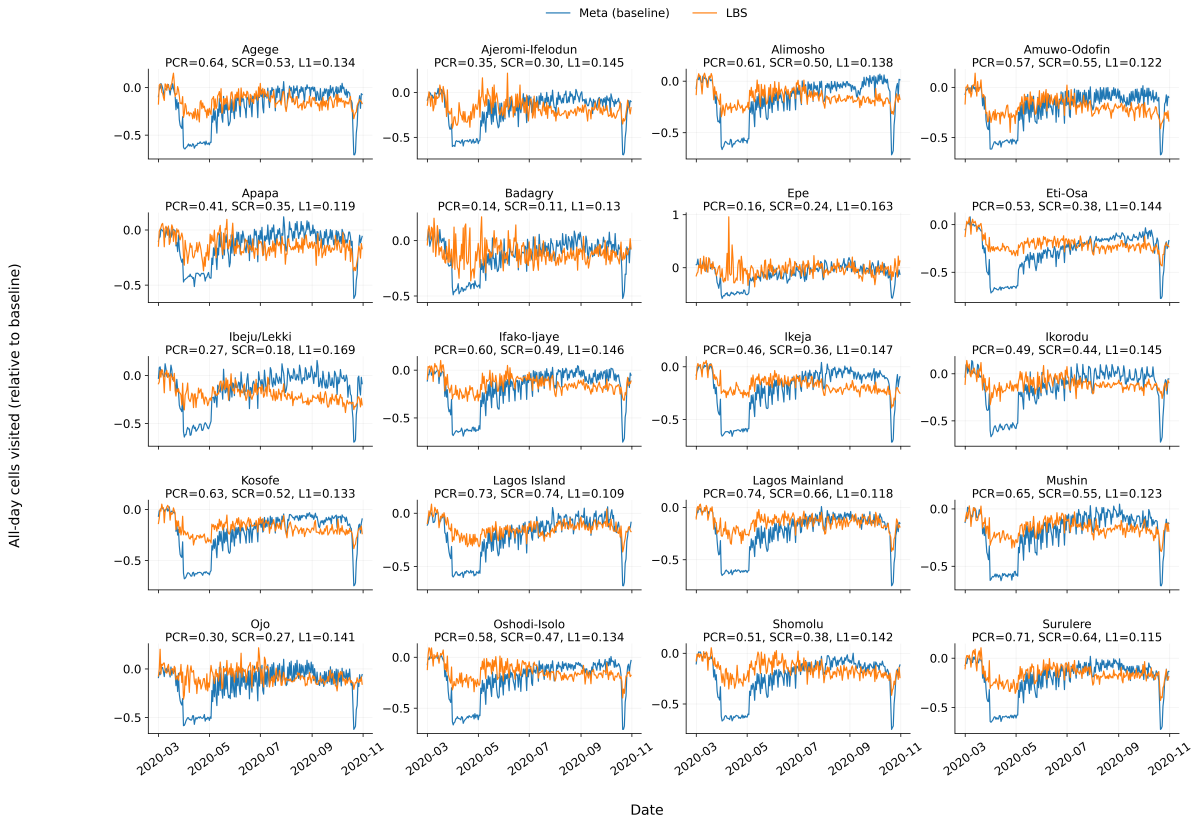
Supplementary Note 6| Agent generation

Based on the estimated tour flows, we generate a corresponding population of agents originating from each geohash-6 cell. Supplementary Figure 10 shows the daily numbers of generated cross-LGA and within-LGA agents. The time series reveals a pronounced reduction in cross-LGA mobility during the lockdown period, followed by a gradual resumption after the onset of Phase 1 reopening, reflecting the implementation and subsequent relaxation of mobility restrictions.

Tour flows can be further aggregated to geohash-6–level OD flows. Supplementary Tables 2, 3, and 4 summarize the distance-bin composition of geohash-6–level trips for each LGA by comparing the Jan–Feb baseline with the March–April lockdown period. Across LGAs, short-distance trips (within 5 km) dominate overall mobility, accounting for approximately 60–70% of trips, consistent with gravity-model distance-decay patterns. The relative differences between the Jan–Feb and March–April distributions provide quantitative evidence of mobility contraction and a strengthened distance-decay effect during the early lockdown phase, indicating a systematic shift toward shorter-distance travel. This pattern is also consistent with the temporal trends observed in Supplementary Figure 10.

Supplementary Note 7| Epidemic modeling and Bayesian optimization

Table 5 summarizes the epidemiological parameters used in the SEAPIR compartmental model, distinguishing between constant parameters and calibrated parameters in this study. The fixed parameters describe intrinsic disease progression processes, including the mean latent period (Λ), the probability of asymptomatic infection (ρ), the mean presymptomatic duration (α), and the mean infectious period (γ_1, γ_2), which are adopted from prior epidemiological studies. These parameters are assumed to be biologically stable and are therefore fixed across all simulation scenarios.In contrast, other parameters are calibrated



Supplementary Figure 5. LGA-level comparison of baseline-normalized changes in all-day cells visited. Time series of baseline-normalized changes in all-day cells visited are shown for 20 LGAs in Lagos State, comparing the Meta Movement Range Maps (blue) and LBS trip roster data (orange) from March to October 2020. Values represent relative deviations from a February day-of-week–specific baseline. Each panel reports the PCR, SCR, and mean L1 distance between the two series.

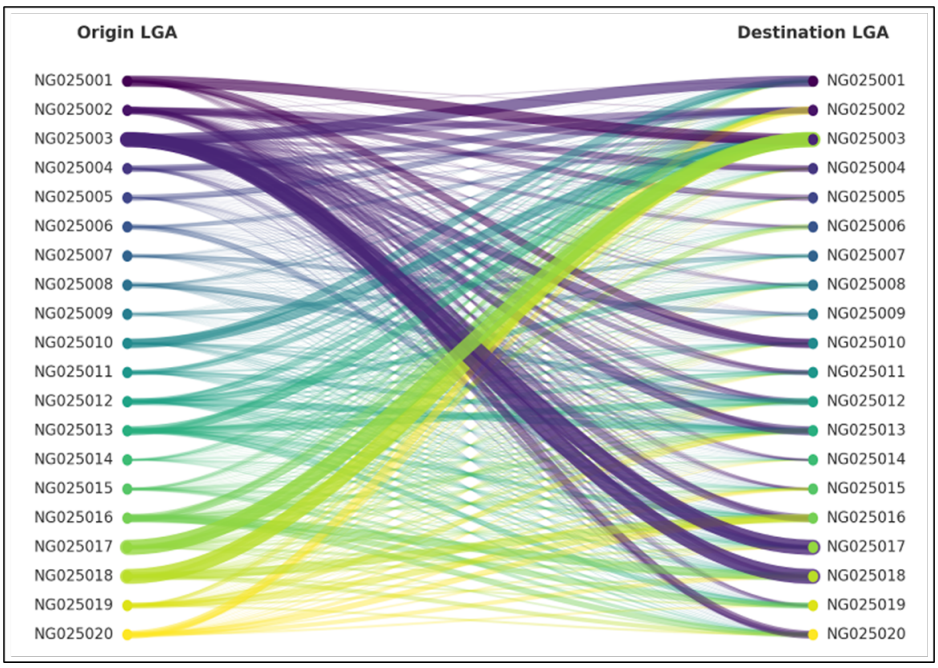
Supplementary Table 2. Distance-bin distribution of trips by LGA from Janunary to Febuary, 2020). The baseline pattern exhibits a clear distance-decay structure consistent with gravity-based mobility, with short-distance trips dominating across all LGAs.

LGA	<5 km	5.1–10 km	10.1–15 km	15.1–20 km	20.1–25 km	>25 km
Agege	60.4	15.5	7.3	6.5	5.9	4.5
Ajeromi-Ifelodun	55.1	16.5	9.8	9.3	5.1	4.4
Alimosho	70.7	11.8	7.4	4.3	2.4	3.3
Amuwo-Odofin	71.8	12.2	6.1	4.9	1.6	3.3
Apapa	67.9	10.1	6.4	5.6	5.6	4.3
Badagry	38.3	8.0	4.5	4.1	3.1	42.1
Epe	39.9	6.2	1.6	4.6	3.2	44.6
Eti-Osa	60.9	11.4	7.5	6.7	5.9	7.5
Ibeju/Lekki	43.3	6.5	3.4	6.2	3.8	36.8
Ifako-Ijaye	72.3	8.6	4.4	4.7	4.3	5.9
Ikeja	73.8	12.9	4.7	3.5	2.5	2.6
Ikorodu	58.0	8.2	5.1	7.5	7.7	13.5
Kosofe	61.8	14.1	12.8	5.5	2.2	3.6
Lagos Island	46.5	21.7	8.8	6.4	10.4	6.2
Lagos Mainland	62.4	17.9	7.1	6.2	2.7	3.7
Mushin	65.5	15.2	11.3	3.9	1.6	2.7
Ojo	60.2	7.1	6.8	8.7	7.1	10.0
Oshodi-Isolo	72.8	15.6	6.8	1.7	1.1	1.9
Shomolu	55.9	16.6	16.3	5.5	2.1	3.7
Surulere	70.2	12.3	7.9	4.9	2.3	2.3
Average	60.4	12.4	7.3	5.5	4.0	10.3

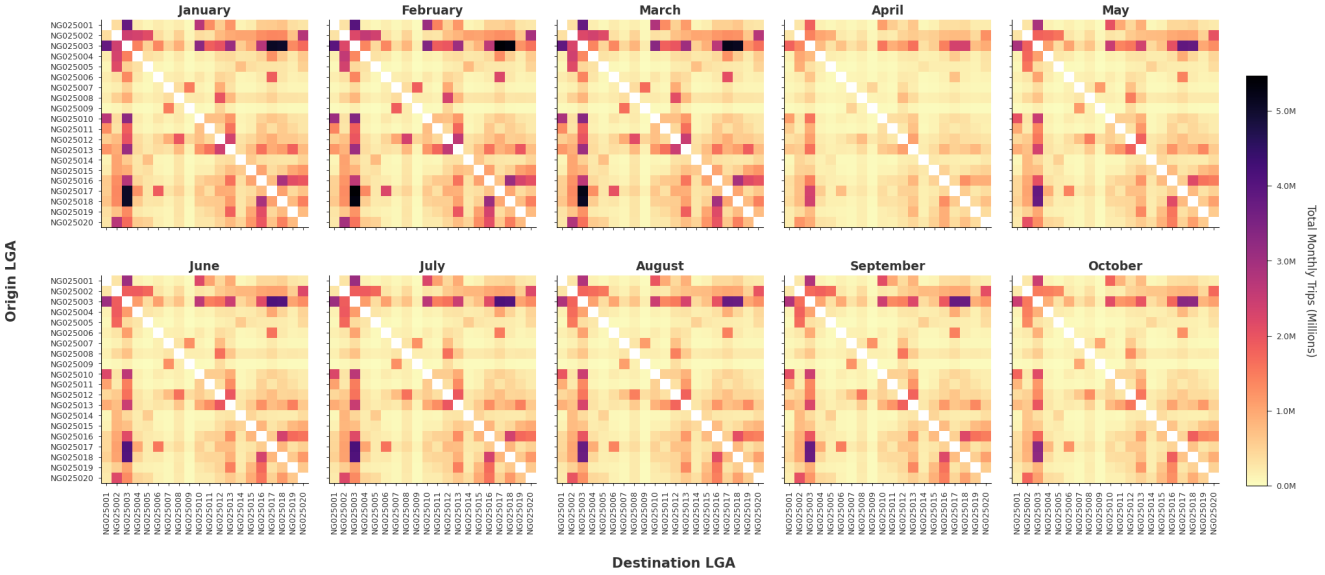
Supplementary Figure 6. Temporal evolution of cross-LGA tour proportions across LGAs in Lagos State Across LGAs, cross-LGA tour proportions exhibit pronounced temporal variation, with substantial declines during lockdown periods and heterogeneous recovery trajectories following phased reopening.

Supplementary Table 3. Distance-bin distribution of trips by LGA from Match to April 2020). Compared with the Jan–Feb baseline, short-distance mobility intensifies while long-distance travel is suppressed, reflecting mobility contraction under non-pharmaceutical interventions.

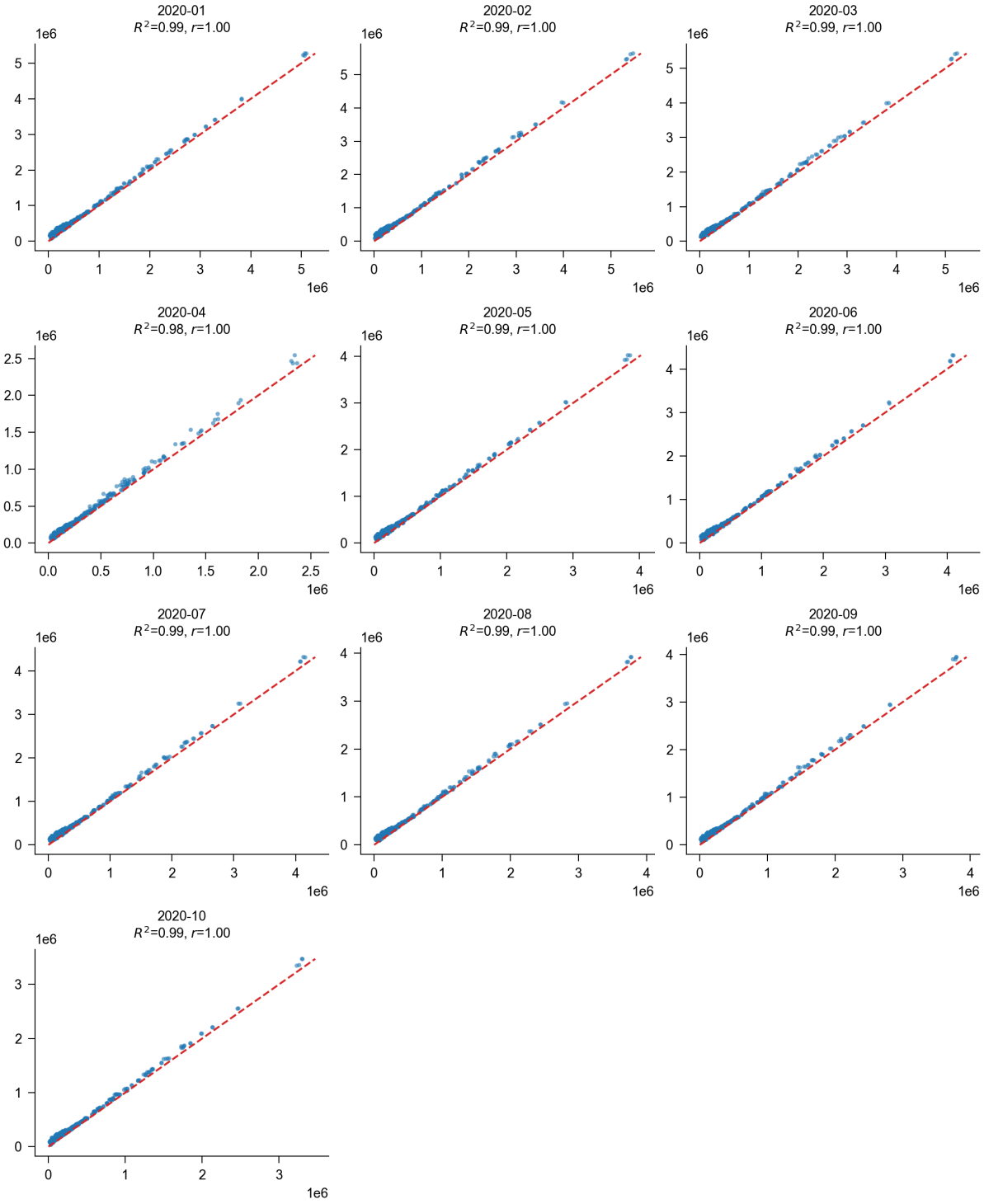
LGA	<5 km	5.1–10 km	10.1–15 km	15.1–20 km	20.1–25 km	>25 km
Agege	67.2	11.7	7.1	5.3	4.9	3.8
Ajeromi-Ifelodun	59.5	13.8	8.1	9.0	4.3	5.2
Alimosho	74.5	10.2	6.9	3.2	2.0	3.3
Amuwo-Odofin	70.8	11.8	6.4	5.4	1.6	3.9
Apapa	66.1	10.0	7.1	7.0	4.5	5.3
Badagry	37.9	6.0	4.3	3.7	3.6	44.3
Epe	43.5	5.9	1.9	3.2	2.3	43.2
Eti-Osa	65.3	9.5	6.7	5.6	5.4	7.5
Ibeju/Lekki	47.0	6.4	3.4	4.9	3.2	35.2
Ifako-Ijaye	75.3	6.6	4.1	4.0	4.5	5.6
Ikeja	76.5	10.9	4.4	3.6	2.0	2.6
Ikorodu	63.3	7.4	4.2	5.9	7.0	12.3
Kosofe	68.5	10.8	10.5	4.9	2.0	3.2
Lagos Island	47.2	20.8	8.1	6.5	11.0	6.5
Lagos Mainland	66.3	14.2	7.0	5.8	3.1	3.5
Mushin	70.6	12.7	9.3	3.3	1.8	2.2
Ojo	63.4	6.2	6.0	6.9	8.2	9.3
Oshodi-Isolo	77.8	12.2	5.9	1.5	1.0	1.7
Shomolu	62.4	13.5	13.8	4.7	2.7	3.0
Surulere	72.3	10.8	7.8	4.1	2.7	2.3
Average	63.8	10.6	6.6	4.9	3.9	10.2



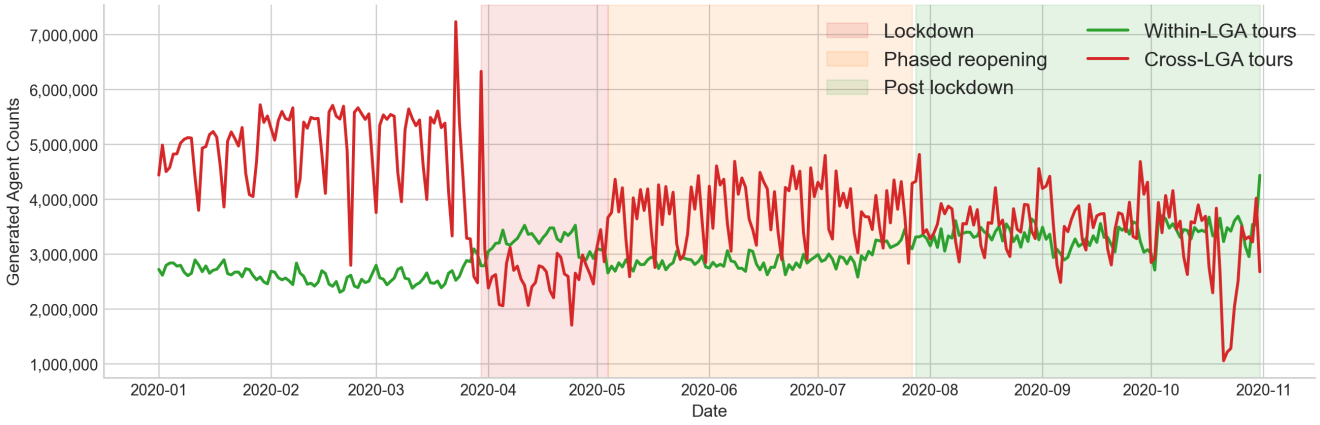
Supplementary Figure 7. Aggregate cross-LGA OD travel flows in Lagos State estimated by the gravity model. The figure visualizes total inter-LGA travel volumes aggregated over the study period, where ribbons connect origin LGAs (left) to destination LGAs (right). Ribbon width represents the relative magnitude of estimated OD flows after iterative proportional fitting. Origin and destination LGAs are indexed using official Nigerian LGA codes.



Supplementary Figure 8. Gravity-model-estimated origin–destination (OD) travel volumes between LGAs in Lagos State, aggregated by month from January to October 2020. Each panel shows the normalized OD matrix for a given month, where rows correspond to origin LGAs and columns correspond to destination LGAs. Color intensity indicates the relative magnitude of cross-LGA travel demand estimated by the gravity model after iterative proportional fitting.Origin and destination LGAs are indexed using official Nigerian LGA codes.



Supplementary Figure 9. Consistency between gravity-model-estimated cross-LGA OD demand and estimated tour flows. Each panel compares monthly aggregate OD volumes obtained from the gravity model (horizontal axis) with those estimated by aggregating geohash-level tour flows (vertical axis) from January to October 2020. The dashed line indicates the 1:1 correspondence. High coefficients of determination (R^2) across all months demonstrate strong cross-resolution consistency between aggregate OD constraints and tour-level mobility flows.



Supplementary Figure 10. Temporal evolution of generated agents and their tour types across different lockdown phases.

Supplementary Table 4. Relative percentage change in distance-bin shares between Mat-Apr and the Jan–Feb periods. Across most LGAs, short-distance trips increase while medium- and long-distance trips decline, highlighting the strengthened distance-decay effect under lockdown consistent with gravity-based mobility contraction.

LGA	<5 km	5.1–10 km	10.1–15 km	15.1–20 km	20.1–25 km	>25 km
Agege	11.3	-24.6	-3.4	-18.1	-16.8	-14.2
Ajeromi-Ifelodun	8.1	-15.9	-16.8	-3.1	-14.7	18.4
Alimosho	5.4	-13.8	-7.8	-25.3	-17.9	-2.2
Amuwo-Odofin	-1.4	-3.1	4.7	10.3	0.2	17.7
Apapa	-2.7	-0.9	11.4	24.1	-20.1	22.1
Badagry	-0.9	-24.3	-3.7	-8.5	17.7	5.4
Epe	9.2	-4.9	20.2	-29.5	-29.3	-3.2
Eti-Osa	7.2	-17.0	-10.8	-16.3	-8.5	-0.8
Ibeju/Lekki	8.6	-2.4	-1.3	-21.2	-15.8	-4.3
Ifako-Ijaye	4.2	-23.4	-6.9	-14.9	4.4	-4.3
Ikeja	3.7	-15.4	-6.9	3.9	-19.1	-2.5
Ikorodu	9.2	-8.8	-17.6	-21.7	-10.2	-9.5
Kosofe	10.9	-22.8	-18.3	-11.5	-7.7	-9.7
Lagos Island	1.5	-4.3	-8.7	1.5	6.3	4.1
Lagos Mainland	6.2	-20.3	-1.2	-5.6	14.5	-5.3
Mushin	7.8	-16.0	-17.3	-13.7	13.5	-16.5
Ojo	5.3	-13.4	-11.8	-21.4	15.7	-6.7
Oshodi-Isolo	6.8	-22.1	-14.5	-12.7	-13.0	-8.7
Shomolu	11.7	-18.6	-15.4	-14.1	27.1	-19.7
Surulere	2.9	-12.2	-1.1	-16.0	16.6	-0.8
Average	5.7	-14.2	-6.4	-10.7	-2.8	-2.0

using Bayesian optimization (BO). Specifically, the transmission coefficient for presymptomatic individuals (β_3) and the daily number of imported exposures (ω) are jointly estimated to match observed epidemic signals under seroprevalence constraints. Transmission coefficients for symptomatic (β_1) and asymptomatic (β_2) individuals are specified as fixed proportions of β_3, reflecting relative infectiousness reported in the literature.

Supplementary Table 5. Constant and calibrated model parameters for SEAPIR model and Bayerisan optimization.

Parameter	Value	Description	Sources
Constant Parameters			
Λ	5.5 days	Mean latent period	Xin et al.[6]
ρ	0.308	Probability exposed becomes asymptomatic	Nishiura et al.[7]
α	2.3 days	Mean presymptomatic duration	He et al.[8]
γ_1, γ_2	3.5 days	Mean infectious duration (Asymp./Symp.)	Li et al.[9]
Calibrated Parameters			
β_3	Calibrated variable	Transmission coefficient (Presymptomatic)	This study
β_1	$0.89\,\beta_3$	Transmission coefficient (Symptomatic)	Vilches et al.[10]
β_2	$0.26\,\beta_3$	Transmission coefficient (Asymptomatic)	Vilches et al.[10]
ω	Calibrated variable	Daily imported exposures (Feb 24–Mar 30)	This study

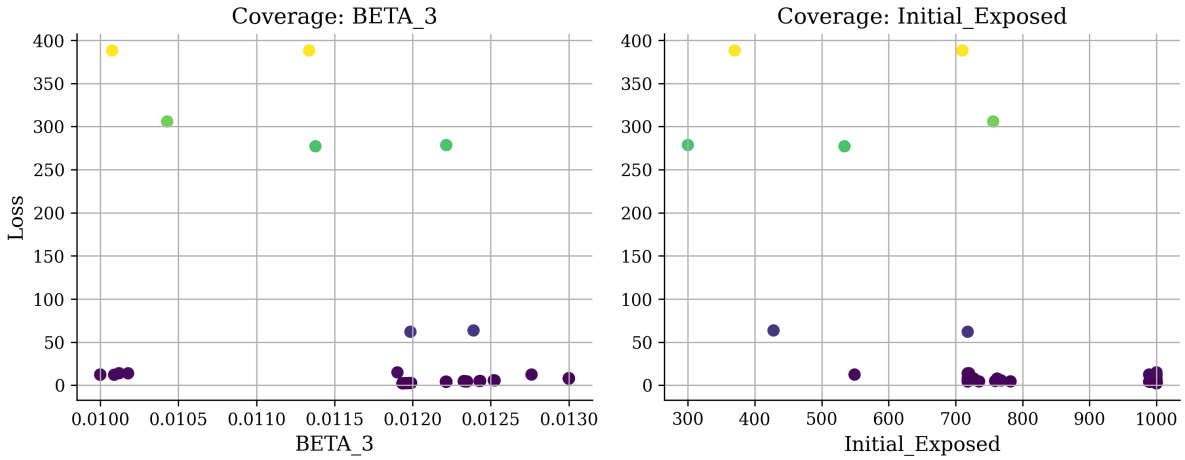
Fig. 11 summarizes the BO procedure and convergence diagnostics used to calibrate key epidemiological parameters. After BO, the estimated symptomatic transmission coefficient is $\beta_1 = 0.0106$, the asymptomatic transmission coefficient is $\beta_2 = 0.0031$, and the presymptomatic transmission coefficient is $\beta_3 = 0.0119$. On average, the model estimates that approximately 1,000 exposed individuals are imported via air travel per day. Imported exposures are assumed to occur only on days **without** air travel restrictions. When calibrating the average number of imported exposures per day, we exclude the period from the onset of lockdown on March 30, 2020, through the end of Phase 3 reopening on July 27, 2020, during which no domestic or international flights arrived in Lagos. Accordingly, imported exposed individuals are set to zero for this interval. This assumption is corroborated by official reports documenting the suspension of air travel during this period[11].

Supplementary Note 8| Scenario configuration

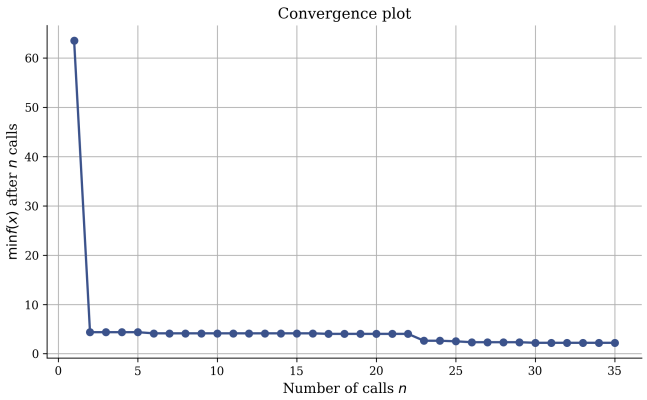
All counterfactual mobility scenarios are generated using the observed mobility patterns as the baseline reference. The observed scenario is segmented into four consecutive demand **regimes** according to policy phases: (i) **pre-lockdown**, (ii) **lockdown**, (iii) **phased reopening**, and (iv) **post-lockdown**. Each regime is characterized by distinct mobility intensity and temporal structure. For the observed scenario, daily agents are directly generated from LBS-based tour flows. For counterfactual scenarios, mobility demand is resampled from empirical distributions observed within specific policy regimes while preserving day-of-week effects. Specifically, for each regime, weekday-specific demand distributions (e.g., Monday, Tuesday, Sunday) are constructed, and counterfactual daily mobility is generated by random sampling from the corresponding distributions, as shown in Fig. 12

1. In the *No Lockdown* scenario, the pre-lockdown demand regime is used as the sole reference for the entire study period. Counterfactual pre-lockdown demand throughout the simulation window is generated by sampling from the weekday-specific distributions estimated during the pre-lockdown period, thereby removing the impact of all lockdown and reopening interventions.

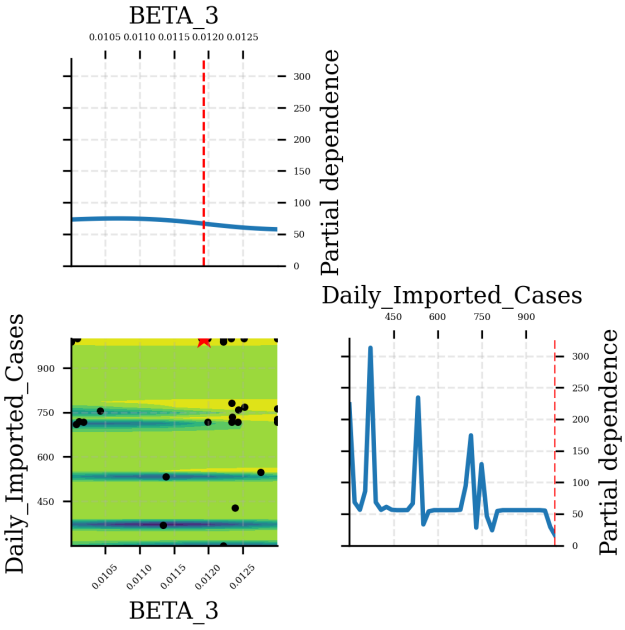
2. In the *Early Lockdown* scenario, the duration of the lockdown regime is kept identical to that in the observed scenario, but the entire lockdown window is shifted earlier in time. Accordingly, counterfactual lockdown demand is generated during this shifted period by sampling from weekday-specific mobility distributions observed during the original lockdown regime. The phased reopening period is shifted earlier in the same manner, with counterfactual phased reopening demand generated using the corresponding observed reopening distributions. Finally, counterfactual post-lockdown demand is generated analogously from the observed post-lockdown regime and applied through the end of the simulation period.



(a) Parameter convergence during Bayesian optimization



(b) Convergence of the Bayesian optimization objective across iterations



(c) Objective landscape of the composite loss across the calibration space

Supplementary Figure 11. Bayesian calibration and convergence diagnostics of epidemiological parameters. a, Parameter convergence during Bayesian optimization for the transmission rate (β_3) and the imported exposed population, showing sampled points across the calibration space. **b,** Convergence of the composite loss function across 35 iterations, indicating rapid stabilization after an initial exploration phase. **c,** Objective landscape of the composite loss function with respect to β_3 and daily imported exposures, illustrating the existence of a well-defined minimum. The optimization identifies a stable parameter set ($\beta = 0.01193$, daily imported exposures $\approx 1{,}000$), which is used for the reported epidemic trajectories.

3. In the *Delayed-lockdown* scenario, the lockdown regime is shifted later in time by the same duration. The intervening period created by the delay is filled using weekday-specific demand distributions from the pre-lockdown regime, while the lockdown, phased reopening, and post-lockdown regimes are generated using their corresponding observed demand distributions.

Supplementary Note 9| Technical details

The simulation period spans from January 1 to October 31 and is indexed by $d = 1, 2, \ldots, D$. The set of simulation days is denoted by $\mathscr{D}$, which is partitioned into two additional collections of temporal units: a set of non-overlapping bi-weekly windows $b \in \mathscr{B}$ and a set of weekly windows $w \in \mathscr{W}$. Different modules are calibrated at different temporal resolutions in accordance with their functional roles and data requirements. Although the adopted temporal resolutions vary across modules, they share a common two-step structure:

Window-level processing: model-specific processing (e.g., pooling, or calibration) is first performed within an appropriate temporal window (weekly or bi-weekly) to ensure statistical adequacy;

Daily temporal construction: daily-level construction is then carried out to recover fine-grained mobility dynamics.

Preparing trip roster data

This module starts from transforming raw LBS sightings into observed OD trips, referred to as *trip roster* data, with imputed home locations using a processing pipeline aligned with NextGen NHTS methodologies[12]. This step converts irregular and device-specific location pings into structured OD trip records that serve as the basis for subsequent tour reconstruction.

Because many devices do not generate complete daily tours due to missing, intermittent, or unevenly sampled observations, the inferred OD trips may be fragmented. To address this issue, intermediate trips are inserted to connect temporally adjacent OD trips based on spatial proximity and temporal continuity, thereby reconstructing complete daily tour chains. Each constructed daily tour starts from the inferred home geohash-6 cell, visits one or more intermediate geohash-6 cells, and ultimately returns to the original home location. To construct complete daily tours from fragmented LBS-derived trips, the following rules are applied:

1. **Home location determination.** Home locations for individual devices are inferred at the geohash-6 level as the location with the highest cumulative dwell time during nighttime hours (9:00 PM–6:00 AM), provided that the device is observed on at least seven distinct nights within a given month.

2. **Intermediate trip insertion.** Intermediate trips are inserted to reconstruct complete daily tours based on the temporal ordering of LBS-recorded trip origins and destinations. Specifically, if a trip ends at geohash x_1 at time t_1 and the subsequent trip starts at geohash x_2 at time t_2, an intermediate trip is added from x_1 at t_1 to x_2 at t_2 to ensure continuity of the trip chain ($t_2 > t_1$).

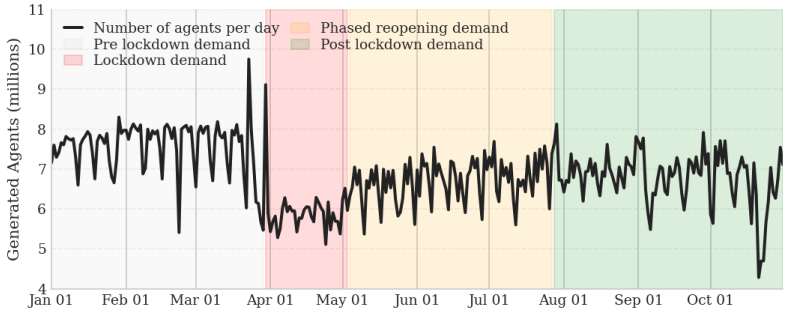
3. **Short time-gap trip merging.** If the time gap between the destination time of a trip (t_1) and the origin time of the subsequent trip (t_2) is smaller than a predefined threshold (approximately 24 minutes), the two successive trips are merged into a single trip. This rule reflects the assumption that extremely short stays are unlikely to generate meaningful activity participation or infection exposure.

4. **Overnight tour handling and daily time definition.** Trips spanning overnight periods are simplified to split into two separate daily tours to ensure that agents return to their home locations within each simulated day. Specifically, a return-to-home trip is appended to the end of the first day, and a departure-from-home trip is added at the beginning of the following day. A simulation day is defined from 3:00 AM to 2:59 AM of the subsequent calendar day to avoid artificial tour fragmentation caused by late-night and early-morning activities.

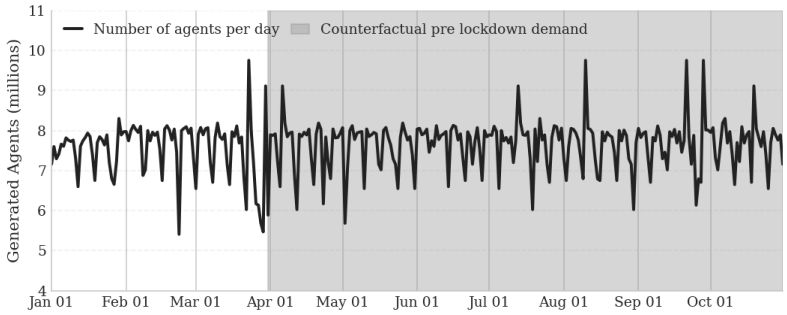
LBS-based tour template generator

Let the constructed daily tours in trip roster data for geohash-6 cell g on day d be denoted by $\overline{\mathscr{T}}_{d,g}$. Representative tour templates for each geohash-6 cell $g \in \mathscr{G}$ are constructed by pooling constructed daily tours within pre-specified bi-weekly time windows $b \in \mathscr{B}$. Specifically, for each bi-weekly window b and geohash-6 cell g, the tour template set is defined as

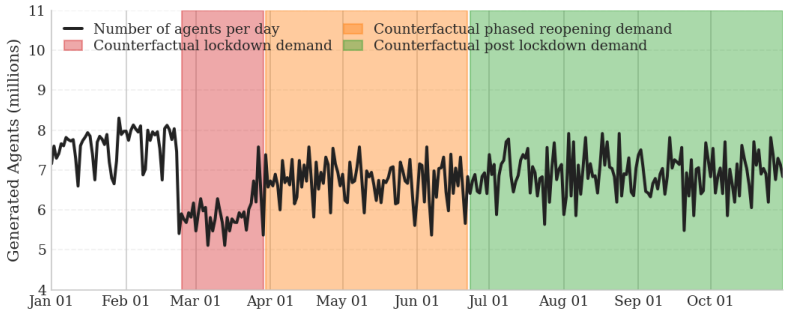
$$\mathscr{T}_{b,g} = \bigcup_{d \in b} \overline{\mathscr{T}}_{d,g}, \qquad \forall b \in \mathscr{B},\ g \in \mathscr{G}. \tag{3}$$



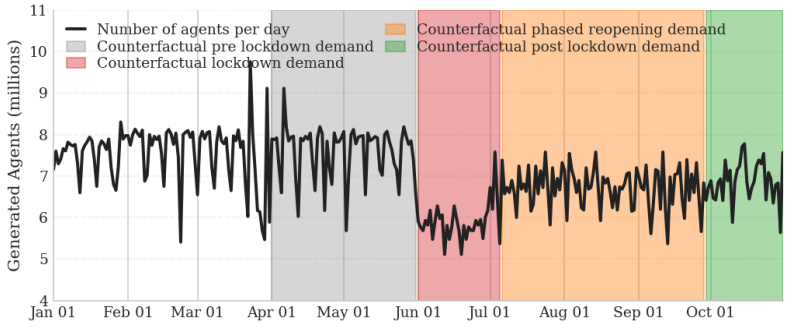
(a) Observed baseline



(b) No lockdown counterfactual



(c) Early lockdown starting Feb 24



(d) Delayed lockdown starting Jun 1

Supplementary Figure 12. Daily generated agents under observed and counterfactual mobility scenarios. Panels show the daily number of generated agents under (a) the observed baseline scenario, (b) the *No Lockdown* counterfactual, (c) the *Early Lockdown* scenario with the lockdown starting on February 24, and (d) the *Delayed Lockdown* scenario with the lockdown starting on June 1. Shaded backgrounds indicate different mobility demand regimes, including pre-lockdown, lockdown, phased reopening, and post-lockdown periods. In counterfactual scenarios, daily mobility demand is generated using regime-specific counterfactual demand distributions resampled from the corresponding observed policy phases while preserving day-of-week effects.

Here, $d \in b$ denotes day d belonging to bi-weekly window b. The complete set of tour templates across all bi-weekly windows in $\mathscr{B}$ and geohash-6 cells in $\mathscr{G}$ is then given by

$$\mathscr{T} = \bigcup_{b \in \mathscr{B}} \bigcup_{g \in \mathscr{G}} \mathscr{T}_{b,g}. \tag{4}$$

The daily tour template set for geohash-6 cell g on day d is defined as

$$\mathscr{T}_{d,g} = \mathscr{T}_{b(d),g}, \qquad \forall d \in \mathscr{D}. \tag{5}$$

Here, $b : \mathscr{D} \to \mathscr{B}$ denotes the mapping that assigns each day d to its corresponding bi-weekly window $b(d)$.

The use of bi-weekly windows reflects a deliberate trade-off between temporal responsiveness and statistical adequacy. Shorter windows improve sensitivity to rapid behavioral changes but may suffer from data sparsity and sampling fluctuations in LBS observations, whereas longer windows provide a larger and more statistically representative set of tour templates at the cost of reduced temporal resolution. The adopted window length balances these competing considerations, ensuring sufficient behavioral diversity while preserving dynamics that are critical for epidemic monitoring and simulation.

LBS-based trip-level gravity model

Trip-level spatial interaction is modeled using a gravity formulation, in which OD flows depend on origin population, destination population, and inter-zonal distance. The gravity model is calibrated at a weekly resolution and subsequently disaggregated to daily mobility flows.

Weekly gravity model calibration. To estimate OD flows across LGAs in Lagos, we develop a macroscopic travel demand model between LGAs calibrated by observed OD trip roster data derived from raw LBS sightings (prior to the insertion of intermediate trips in Module I). For each **weekly** window $w \in \mathscr{W}$, observed OD trips are aggregated across all days within the week and used to calibrate a gravity model of the form

$$\widehat{q}_{l,l'}(w) = K_w \frac{Z_l^{\mu_{1,w}} Z_{l'}^{\mu_{2,w}}}{d_{l,l'}^{\phi_w}}, \qquad l, l' \in \mathscr{L},\ w \in \mathscr{W}. \tag{6}$$

Here, $F_{l,l'}(w)$ denotes the total number of trips between origin LGA l and destination LGA l' during week w; P_l and $P_{l'}$ represent the populations of the origin and destination LGAs, respectively; and $d_{l,l'}$ denotes the centroid-to-centroid distance between LGAs. The parameters $\mu_{1,w}$, $\mu_{2,w}$, and ϕ_w, together with the constant K_w, are estimated via log-linear regression using weekly aggregated LBS OD trip volumes.

The resulting weekly OD trip estimates $\widehat{q}_{l,l'}(w)$ are further adjusted using an iterative proportional fitting (IPF) procedure to enforce a *doubly constrained gravity model*. IPF is a standard approach in transportation modeling for rescaling preliminary OD matrices such that both origin-side production totals and destination-side attraction totals are satisfied[13]. Specifically, IPF iteratively rescales $\widehat{q}_{l,l'}(w)$ to obtain adjusted flows $q_{l,l'}(w)$ that satisfy

$$\sum_{l' \in \mathscr{L}} q_{l,l'}(w) = O_l(w), \qquad \sum_{l \in \mathscr{L}} q_{l,l'}(w) = D_{l'}(w), \quad \forall w \in \mathscr{W}. \tag{7}$$

Here, $O_l(w)$ denotes the expected total trip production of origin LGA l in week w, and $D_{l'}(w)$ denotes the expected total trip attraction of destination LGA l' in the same week. In this study, weekly productions and attractions are assumed to be symmetric and proportional to population size, such that

$$O_l(w) = D_l(w) = Z_l \cdot r_{l,w} \cdot n_w \cdot \eta_{l,w}, \qquad \forall l \in \mathscr{L}, w \in \mathscr{W} \tag{8}$$

where Z_l is the population of LGA l, $r_{l,w}$ denotes the average number of trips per person per day in week w, n_w is the number of days in week w (i.e., 7 days), and $\eta_{l,w}$ denotes the proportion of cross-LGA trips. It should be noted that the proportion of cross-LGA trips is conceptually distinct from the proportion of cross-LGA tours mentioned in **Supplementary Note 4**. Tours originating from other LGAs may include trips whose starting points lie within the currently considered LGA.

Specifically, both the weekly trip rate $r_{l,w}$ and the cross-LGA trip proportion $\eta_{l,w}$ are empirically derived from the **LBS-based trip roster data** as reference quantities. Specifically, the weekly trip rate $r_{l,w}$ is computed as the average daily trip rate per active device within week w:

$$r_{l,w} = \frac{1}{n_w} \sum_{d \in w} \frac{N_{\text{trip}}(d,l)}{N_{\text{device}}(d,l)}, \qquad \forall w \in \mathscr{W}, l \in \mathscr{L} \tag{9}$$

where $N_{\text{trip}}(d,l)$ denotes the total number of observed trips on day d from LGA l, $N_{\text{device}}(d,l)$ denotes the number of active LBS devices on that day from the LGA, and n_w is the number of days in week w. $N_{\text{device}}(d,l)$ denotes the number of observed tours originating from LGA l on day d in the trip roster dataset, **prior to** pooling into bi-weekly time windows to generate daily representative tour templates. The cross-LGA trip proportion $\eta_{l,w}$ is defined as the average daily share of trips whose origin and destination belong to different LGAs:

$$\eta_{l,w} = \frac{1}{n_w} \sum_{d \in w} \frac{N^C_{\text{trip}}(d,l)}{N_{\text{trip}}(d,l)}, \qquad \forall w \in \mathscr{W}, l \in \mathscr{L} \tag{10}$$

where $N^C_{\text{trip}}(d,l)$ represents the number of cross-LGA trips observed on day d from LGA l. Here, $d \in w$ denotes day d belonging to weekly window w.

Daily OD flow construction. After weekly gravity model calibration and IPF adjustment, a weekly cross-LGA OD matrix $q_{l,l'}(w)$ is obtained for each week $w \in \mathscr{W}$. To recover fine-grained temporal mobility dynamics, the weekly OD flows are further disaggregated to the daily level using the within-week trip distribution observed in the LBS-based trip roster data. Specifically, for each day d belonging to week w, the within-week trip share is defined as

$$s_{d,l}(w) = \frac{N_{\text{trip}}(d,l)}{\sum_{d \in w} N_{\text{trip}}(d,l)}, \qquad \sum_{d \in w} s_{d,l}(w) = 1, \quad \forall d \in w \in \mathscr{W}, l \in \mathscr{L} \tag{11}$$

where $N_{\text{trip}}(d)$ denotes the total number of observed trips in trip roster data on day d in the trip roster dataset. Let $w(d) \in \mathscr{W}$ denote the week containing day d. The daily OD flow between LGAs l and l' on day d is then constructed via proportional allocation:

$$q_{l,l'}(d) = s_{d,l}\big(w(d)\big)\, q_{l,l'}\big(w(d)\big), \qquad \forall l, l' \in \mathscr{L},\ d \in \mathscr{D}. \tag{12}$$

Here, daily OD flows are derived from the temporal distribution of all observed trips, rather than only cross-LGA trips, to avoid noise and instability caused by sparse cross-LGA observations.

Tour-based flow estimation

Now, we use hierarchical flow representation used in the layered computational graph (CG) framework in the field of travel demand modeling to estimate tour flows.[14–17]. For each day $d \in \mathscr{D}$ and home geohash $g \in \mathscr{G}$, a daily tour-template set $\mathscr{T}_{d,g}$ is provided, which we partition into three mutually exclusive categories:

$$\mathscr{T}_{d,g} = \mathscr{T}^G_{d,g} \cup \mathscr{T}^L_{d,g} \cup \mathscr{T}^C_{d,g}, \qquad \mathscr{T}^G_{d,g} \cap \mathscr{T}^L_{d,g} = \mathscr{T}^G_{d,g} \cap \mathscr{T}^C_{d,g} = \mathscr{T}^L_{d,g} \cap \mathscr{T}^C_{d,g} = \varnothing. \tag{13}$$

Here, $\mathscr{T}^G_{d,g}$ denotes *within-geohash* tours, $\mathscr{T}^L_{d,g}$ denotes *within-LGA but cross-geohash* tours, and $\mathscr{T}^C_{d,g}$ denotes *cross-LGA* tours.

Variables. Based on the partition, we define the following layered variables

$X^G_{d,g} \geq 0$ denotes the total daily *within-geohash* trip outflow produced by residents of geohash g on day d.

$X^L_{d,g} \geq 0$ denotes the total daily *within-LGA but cross-geohash* trip outflow produced by residents of geohash g on day d.

$X^C_{d,g} \geq 0$ denotes the total daily *cross-LGA* trip outflow produced by residents of geohash g on day d.

$X^G_{d,g,\tau} \geq 0$ denotes the flow assigned to tour template $\tau \in \mathscr{T}^G_{d,g}$ on day d for residents of geohash g.

$X^L_{d,g,\tau} \geq 0$ denotes the flow assigned to tour template $\tau \in \mathscr{T}^L_{d,g}$ on day d for residents of geohash g.

$X^C_{d,g,\tau} \geq 0$ denotes the flow assigned to tour template $\tau \in \mathscr{T}^C_{d,g}$ on day d for residents of geohash g.

$p^G_{d,g,\tau} \in [0,1]$ denotes the *proportion* of the within-geohash outflow $X^G_{d,g}$ allocated to tour $\tau \in \mathscr{T}^G_{d,g}$ on day d.

$p^L_{d,g,\tau} \in [0,1]$ denotes the *proportion* of the within-LGA but cross-geohash outflow $X^L_{d,g}$ allocated to tour $\tau \in \mathscr{T}^L_{d,g}$ on day d.

$p^C_{d,g,\tau} \in [0,1]$ denotes the *proportion* of the cross-LGA outflow $X^C_{d,g}$ allocated to tour $\tau \in \mathscr{T}^C_{d,g}$ on day d.

Parameters Parameters $\overline{p}^G_{d,g}$, $\overline{p}^L_{d,g}$, and $\overline{p}^C_{d,g} \in [0,1]$ are empirically estimated from the LBS-based trip roster data. They represent the observed proportions of *within-geohash*, *within-LGA but cross-geohash*, and *cross-LGA* tour types, respectively, among all trips produced by residents of geohash cell g on day d. Specifically, these proportions are computed as

$$\overline{p}^G_{d,g} = \frac{N^G_{\text{device}}(d,g)}{N_{\text{device}}(d,g)}, \qquad \overline{p}^L_{d,g} = \frac{N^L_{\text{device}}(d,g)}{N_{\text{device}}(d,g)}, \qquad \overline{p}^C_{d,g} = \frac{N^C_{\text{device}}(d,g)}{N_{\text{device}}(d,g)}, \qquad \forall d \in \mathscr{D}, g \in \mathscr{G} \tag{14}$$

where $N_{\text{device}}(d,g)$ denotes the total number of active LBS devices from geohash cell g on day d, and $N^G_{\text{device}}(d,g)$, $N^L_{\text{device}}(d,g)$, and $N^C_{\text{device}}(d,g)$ denote the numbers of devices whose observed trips are classified as within-geohash, within-LGA but cross-geohash, and cross-LGA trips, respectively. It is noted that these proportions are calculated using the original trip roster, before tours are pooled into bi-weekly time windows.

To couple cross-LGA template flows with the cross-LGA OD matrix, we define$\delta^{l,l'}_{d,g,\tau} \in \{0,1\}$ is a known incidence parameter for $\tau \in \mathscr{T}^C_{d,g}$: $\delta^{l,l'}_{d,g,\tau} = 1$ if tour τ corresponds to an cross-LGA OD pair from LGA l to LGA l' on day d, and 0 otherwise.

Layered CG constraints. The hierarchical flow representation is enforced through the following constraints.
(C1) Three-type tours split consistency (within-geohash / within-LGA / cross-LGA):

$$X^G_{d,g} = \overline{p}^G_{d,g} Z_g, \qquad X^L_{d,g} = \overline{p}^L_{d,g} Z_g, \qquad X^C_{d,g} = \overline{p}^C_{d,g} Z_g, \qquad \forall d \in \mathscr{D},\ g \in \mathscr{G}. \tag{15}$$

(C2) Fow conservation constraints:

$$X^G_{d,g,\tau} = p^G_{d,g,\tau} X^G_{d,g,\tau}, \qquad X^L_{d,g,\tau} = p^L_{d,g,\tau} X^L_{d,g,\tau}, \qquad X^C_{d,g,\tau} = p^C_{d,g,\tau} X^C_{d,g,\tau}, \qquad \forall d \in \mathscr{D},\ g \in \mathscr{G}. \tag{16}$$

(C3) Cross-LGA consistency with the macroscopic OD layer:

$$q_{l,l'}(d) = \sum_{g \in \mathscr{G}} \sum_{\tau \in \mathscr{T}^C_{d,g}} \sum_{k \in \{G,L,C\}} \delta^{l,l'}_{d,g,\tau} X^k_{d,g,\tau}, \qquad \forall d \in \mathscr{D},\ \forall (l,l') \in \mathscr{L} \times \mathscr{L},\ l \neq l'. \tag{17}$$

To obtain the least biased allocation of flows over templates subject to the hierarchical constraints, we adopt a maximum-entropy formulation. Equivalently, we minimize the negative entropy of template flows:

$$\max_{\{p^k_{d,g,\tau}\}} \mathscr{H} = \sum_{d \in \mathscr{D}} \sum_{g \in \mathscr{G}} \sum_{k \in \{G,L,C\}} X^k_{d,g} \left(- \sum_{\tau \in \mathscr{T}^k_{d,g}} p^k_{d,g,\tau} \ln p^k_{d,g,\tau} \right). \tag{18}$$

The resulting optimization problem is a convex program with linear equality constraints. It can be solved either via its dual formulation using interior-point methods, or by directly optimizing the primal objective through gradient-based backpropagation, treating the layered computational graph as a differentiable model[14]. For tour categories without additional structural constraints, the maximum-entropy objective yields uniform allocations across admissible templates. Importantly, the proposed framework is flexible and can readily accommodate additional behavioral assumptions or structural constraints.

According to the layered CG structure, the the overall template share and estimated tour flow can be calculated as follows:

$$p_{d,g,\tau} = \overline{p}^k_{d,g} p^k_{d,g,\tau}, \qquad \forall \tau \in \mathscr{T}^k_{d,g}; \forall d \in \mathscr{D}, g \in \mathscr{G}, k \in \{G,L,C\}. \tag{19}$$

$$X_{d,g,\tau} = Z_g p_{d,g,\tau}, \qquad \forall \tau \in \mathscr{T}_{d,g}; \forall d \in \mathscr{D}, g \in \mathscr{G}. \tag{20}$$

Based on the estimated daily tour-level flows, geohash-level OD flows, daily trip production from the original geohash cell $O_g(d)$, and daily visit counts at each cell $V_{g'}(d)$ can be further derived as follows:

$$O_g(d) = \sum_{\tau \in \mathscr{T}^C_{d,g} \cup \mathscr{T}^L_{d,g}} X_{d,g,\tau}, \qquad \forall d \in \mathscr{D},\ g \in \mathscr{G}, \tag{21}$$

$$q_{g,g'}(d) = \sum_{g'' \in \mathscr{G}} \sum_{\tau \in \mathscr{T}^C_{d,g''} \cup \mathscr{T}^L_{d,g''}} \delta^{g,g'}_{d,g'',\tau} X_{d,g'',\tau}, \qquad \forall d \in \mathscr{D},\ g' \neq g''; g,g' \in \mathscr{G}, \tag{22}$$

$$V_{g'}(d) = \sum_{g \in \mathscr{G}} q_{g,g'}(d), \qquad \forall d \in \mathscr{D},\ g' \in \mathscr{G}. \tag{23}$$

Epidemic simulation

We modeled disease progression using a SEAPIR framework (Susceptible–Exposed–Asymptomatic–Presymptomatic–Infectious–Recovered). The simulation operates on a hybrid timescale: while agent mobility and exposure risks are evaluated dynamically based on daily trajectories, biological state transitions occur discretely at the day boundary.

SEAPIR compartmental dynamics. For the susceptible (S), exposed (E), asymptomatic (A), presymptomatic (P), infectious (I), and recovered (R) compartments, we have the following difference equations

$$S_g(d+1) - S_g^M(d) = -c_H S_g^M(d) \frac{\beta_1 I_g(d) + \beta_2 A_g(d) + \beta_3 P_g(d)}{Z_g}, \qquad \forall d \in \mathscr{D},\ g \in \mathscr{G}. \tag{24}$$

$$E_g(d+1) - E_g^M(d) = c_H S_g^M(d) \frac{\beta_1 I_g(d) + \beta_2 A_g(d) + \beta_3 P_g(d)}{Z_g} - \frac{E_g^M(d)}{\Lambda}, \qquad \forall d \in \mathscr{D},\ g \in \mathscr{G}. \tag{25}$$

$$P_g(d+1) - P_g(d) = (1-\rho)\frac{E_g^M(d)}{\Lambda} - \frac{P_g(d)}{\alpha}, \qquad \forall d \in \mathscr{D},\ g \in \mathscr{G}. \tag{26}$$

$$A_g(d+1) - A_g(d) = \rho \frac{E_g^M(d)}{\Lambda} - \frac{A_g(d)}{\gamma_1}, \qquad \forall d \in \mathscr{D},\ g \in \mathscr{G}. \tag{27}$$

$$I_g(d+1) - I_g(d) = \frac{P_g(d)}{\alpha} - \frac{I_g(d)}{\gamma_2}, \qquad \forall d \in \mathscr{D},\ g \in \mathscr{G}. \tag{28}$$

$$R_g(d+1) - R_g(d) = \frac{A_g(d)}{\gamma_1} + \frac{I_g(d)}{\gamma_2}, \qquad \forall d \in \mathscr{D},\ g \in \mathscr{G}. \tag{29}$$

where $S_g(d), E_g(d), A_g(d), P_g(d), I_g(d), R_g(d)$ denote the compartment counts for geohash g in day d. The parameters $\beta_1, \beta_2, \beta_3$ represent transmission coefficients for symptomatic, asymptomatic, and presymptomatic individuals, respectively. c_H is the household contact rate; ρ is the probability of asymptomatic progression; Λ is the mean latent period; α is the mean presymptomatic duration; and γ_1, γ_2 denote the mean infectious periods for asymptomatic and symptomatic (infected) cases. To further illustrate the dynamics among the compartments, Figure 13 presents the flow diagram of the SEAPIR model.

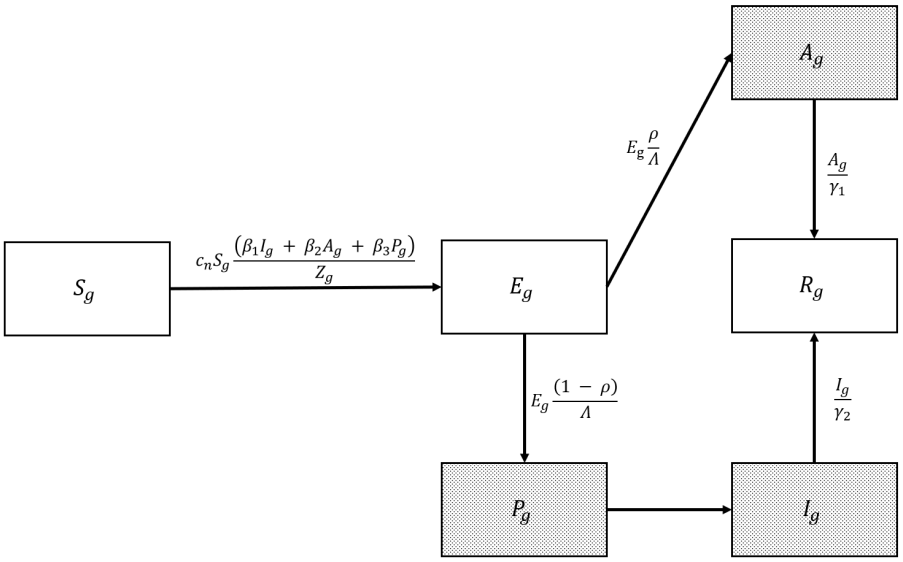
The above compartmental dynamics are formulated in discrete time with a clear separation between *mobility-induced mixing* and *epidemic progression*. The superscript $(\cdot)^M$ denotes epidemiological states *after daytime mobility mixing* but *before* disease progression. Mobility is assumed to affect only the transition from the susceptible to the exposed compartment, as population movement modifies contact opportunities but does not directly alter latent, presymptomatic, asymptomatic, infected, or recovered states. Accordingly, mobility-adjusted states appear only in the first two equations through $S_g^M(d)$ and $E_g^M(d)$, while all remaining compartments follow standard disease progression dynamics. Specifically, the reduction from $S_g^M(d)$ to $S_g(d+1)$ is entirely driven by infection incidence at home, and newly generated exposures during the mobility-adjusted phase enter the exposed compartment and subsequently progress at rate $1/\Lambda$. All other transitions are governed by epidemiological parameters $(\rho, \alpha, \gamma_1, \gamma_2)$ and are unaffected by mobility.

Mobility-driven exposure. We integrated the compartmental framework with the agent-based mobility model through a daily interaction cycle. Let $\mathscr{E}(\tau) \subseteq \mathscr{G} \times \mathscr{G}$ denote the geohash-to-geohash trips contained in tour τ. For each edge $(q, q') \in \mathscr{E}(\tau)$, the destination geohash q' is treated as a daytime mixing location where infection may occur. Here, mobility is assumed to affect only the transition from susceptible to exposed. Accordingly, the mobility-adjusted states satisfy

$$S_g^M(d) - S_g(d) = -S_g(d) \sum_{\tau \in \mathscr{T}_{d,g}} \sum_{(q,q') \in \mathscr{E}(\tau)} p_{d,g,\tau} \pi_{d,g,\tau,g'}. \tag{30}$$

$$E_g^M(d) - E_g(d) = S_g(d) \sum_{\tau \in \mathscr{T}_{d,g}} \sum_{(q,q') \in \mathscr{E}(\tau)} p_{d,g,\tau} \pi_{d,g,\tau,g'}. \tag{31}$$

The above two equations enforce conservation of population between S and E during the daytime mobility-mixing stage. By using the layered CG constraints mentioned before. According to the layered CG structure, we can further estimate daily **geohash-level** trip production flows and visit counts from I, A, and P compartments, i.e., $O_g^I(d), O_g^A(d), O_g^P(d)$ and $V_g^I(d), V_g^A(d), V_g^P(d)$:



Supplementary Figure 13. Flow diagram of the SEAPIR compartmental model, with infectious compartments highlighted

$$X^I_{d,g,\tau} = I_g(d)\, p_{d,g,\tau}, \qquad \forall \tau \in \mathscr{T}_{d,g}; \forall d \in \mathscr{D}, g \in \mathscr{G}. \tag{32}$$

$$X^A_{d,g,\tau} = A_g(d)\, p_{d,g,\tau}, \qquad \forall \tau \in \mathscr{T}_{d,g}; \forall d \in \mathscr{D}, g \in \mathscr{G}. \tag{33}$$

$$X^P_{d,g,\tau} = P_g(d)\, p_{d,g,\tau}, \qquad \forall \tau \in \mathscr{T}_{d,g}; \forall d \in \mathscr{D}, g \in \mathscr{G}. \tag{34}$$

$$O^I_g(d) = \sum_{\tau \in \mathscr{T}^C_{d,g} \cup \mathscr{T}^L_{d,g}} X^I_{d,g,\tau}, \qquad \forall d \in \mathscr{D},\ g \in \mathscr{G}, \tag{35}$$

$$O^A_g(d) = \sum_{\tau \in \mathscr{T}^C_{d,g} \cup \mathscr{T}^L_{d,g}} X^A_{d,g,\tau}, \qquad \forall d \in \mathscr{D},\ g \in \mathscr{G}, \tag{36}$$

$$O^P_g(d) = \sum_{\tau \in \mathscr{T}^C_{d,g} \cup \mathscr{T}^L_{d,g}} X^P_{d,g,\tau}, \qquad \forall d \in \mathscr{D},\ g \in \mathscr{G}, \tag{37}$$

$$q^I_{g,g'}(d) = \sum_{g'' \in \mathscr{G}} \sum_{\tau \in \mathscr{T}^C_{d,g''} \cup \mathscr{T}^L_{d,g''}} \delta^{g,g'}_{d,g'',\tau} X^I_{d,g'',\tau}, \qquad \forall d \in \mathscr{D},\ g' \neq g''; g, g' \in \mathscr{G}, \tag{38}$$

$$q^A_{g,g'}(d) = \sum_{g'' \in \mathscr{G}} \sum_{\tau \in \mathscr{T}^C_{d,g''} \cup \mathscr{T}^L_{d,g''}} \delta^{g,g'}_{d,g'',\tau} X^A_{d,g'',\tau}, \qquad \forall d \in \mathscr{D},\ g' \neq g''; g, g' \in \mathscr{G}, \tag{39}$$

$$q^P_{g,g'}(d) = \sum_{g'' \in \mathscr{G}} \sum_{\tau \in \mathscr{T}^C_{d,g''} \cup \mathscr{T}^L_{d,g''}} \delta^{g,g'}_{d,g'',\tau} X^P_{d,g'',\tau}, \qquad \forall d \in \mathscr{D},\ g' \neq g''; g, g' \in \mathscr{G}, \tag{40}$$

$$V^I_{g'}(d) = \sum_{g \in \mathscr{G}} q^I_{g,g'}(d), \qquad \forall d \in \mathscr{D},\ g' \in \mathscr{G}. \tag{41}$$

$$V^A_{g'}(d) = \sum_{g \in \mathscr{G}} q^A_{g,g'}(d), \qquad \forall d \in \mathscr{D},\ g' \in \mathscr{G}. \tag{42}$$

$$V^P_{g'}(d) = \sum_{g \in \mathscr{G}} q^P_{g,g'}(d), \qquad \forall d \in \mathscr{D},\ g' \in \mathscr{G}. \tag{43}$$

Based on the above geohash-level outflow quantities and visit counts, a simplified exposure probability is formulated to capture mobility-driven interactions. The probability is defined as a function of the effective infectious population present in a geohash during the day. It follows an assumption, under which exposure risk increases with the number of infectious, asymptomatic,

and presymptomatic individuals encountered through mobility-driven contacts.

The formulation accounts for visit counts to each geohash cell while netting out individuals who leave the cell during the day, such that the effective population reflects daytime mixing rather than static residential counts. To maintain tractability, interactions are simplified by assuming that the order of visits within a day does not affect exposure risk and that outbound travelers do not contribute to within-cell contact opportunities after departure. For a susceptible individual whose home geohash is g and whose tour τ visits destination geohash g', exposed probability at geohash g' is modeled as

$$\pi_{d,g,\tau,g'} = 1 - \exp\left[-c_N \cdot \frac{\beta_1 I_{g'}(d) + V^I_{g'}(d) - O^I_{g'}(d) + \beta_2 A_{g'}(d) + V^A_{g'}(d) - O^A_{g'}(d) + \beta_3 P_{g'}(d) + V^P_{g'}(d) - O^P_{g'}(d)}{Z_{g'} + V_{g'}(d) - O_{g'}(d)}\right] \quad (44)$$

The formulation captures how individuals visiting geohash cell g' may interact with individuals from other geohash cells as a result of mobility-driven population mixing.

Bayesian optimization

To estimate key parameters governing transmission and importation, BO is employed to calibrate the epidemic simulation under seroprevalence constraints. A Gaussian process surrogate model is used to approximate this mapping based on a limited number of simulation runs, capturing both the expected simulation outcomes and associated uncertainty across the parameter space. An acquisition function then guides the sequential selection of parameter candidates, balancing exploration of uncertain regions and exploitation of parameter settings that better reproduce observed epidemic constraints.

Epidemiological parameters We calibrate two epidemiological parameters within bounded search intervals:

1. The average daily importation volume of exposed individuals, denoted as γ;
2. the presymptomatic transmission coefficient, denoted as β_3.

Given that the initial seeding condition is fixed, the importation volume ι_{imp} becomes the primary driver of early epidemic growth, jointly with β_3. To mitigate parameter non-identifiability, the symptomatic and asymptomatic transmission coefficients are constrained to fixed proportions of β_3, with $\beta_1 = 0.89\beta_3$ and $\beta_2 = 0.26\beta_3$, respectively. This reduction preserves the empirically established hierarchy of infectiousness reported in the literature.

The epidemic simulation is initialized on February 1, 2020, to capture the earliest potential silent transmission chains. To standardize the epidemic starting point, the system is seeded with a single exposed individual placed in the highest-density geohash cell. Subsequent exogenous introductions are modeled as an average daily inflow of exposed individuals at rate γ, beginning on February 1 and terminating upon the implementation of the interstate travel ban (lockdown). Thus during the lockdown, we restrict $\gamma = 0$. The average imported exposure estimation is reintroduced on August 1, 2020, corresponding to the resumption of international flights and the end of the phased reopening period. This specific timing is corroborated by international air travel trends derived from T-100 market data, which exhibit a distinct recovery trajectory starting in August. To capture the spatial heterogeneity of risk, these imported exposures are distributed probabilistically to cells ranked by population density, concentrating introductions in major transport and commercial hubs[18,19].

Loss functions We calibrate the parameters β_3 and γ by minimizing the following loss function,

$$J(\beta_3, \gamma) = w_{\text{peak}} J_{\text{peak}}(\beta_3, \gamma) + w_{\text{cum}} J_{\text{cum}}(\beta_3, \gamma) + w_{\text{act}} J_{\text{act}}(\beta_3, \gamma), \quad (45)$$

where $w_{\text{peak}}, w_{\text{cum}}$, and w_{act} are nonnegative weights.

The first term in the objective function encodes the peak-timing principle. Although the magnitude of officially reported case counts is subject to substantial under-ascertainment, the timing of the downturn in reported cases ($t_{\text{target}} =$ July 31, 2020) is treated as a conservative upper bound on the true epidemic peak. Accordingly, we define an asymmetric loss that penalizes only late simulated peaks. Let t_{off} denote the official peak time inferred from reported cases, and let $t_{\text{sim}}(\beta_3, \gamma)$ denote the simulated peak time of daily new exposed individuals (as an infection proxy) under parameters (β_3, γ). The peak-timing loss is defined as

$$J_{\text{peak}}(\beta_3, \gamma) = \lambda \left[\frac{t_{\text{sim}}(\beta_3, \gamma) - t_{\text{off}}}{\tau}\right]_+^2 + \left[\frac{t_{\text{sim}}(\beta_3, \gamma) - t_{\text{off}}}{\tau}\right]_+, \quad (46)$$

where $[x]_+ = \max(x, 0)$, τ is a temporal scaling constant (e.g., $\tau = 7$ days), and $\lambda = 2.0$ controls the penalty strength. The linear term ensures that any delay beyond the official peak incurs a nonzero penalty, while the quadratic term increasingly penalizes larger deviations, reflecting growing implausibility of substantially delayed epidemic peaks.

The second term enforces consistency between the simulated cumulative number of infections and the seroprevalence-based estimate of population-level exposure. Let $C_{\text{sim}}(\beta_3, \gamma)$ denote the cumulative number of infected individuals (including

recovered and currently infectious) at the survey time, and let C_{target} be the corresponding empirical mean inferred from the seroprevalence survey. The cumulative-prevalence loss is defined as

$$J_{\text{cum}}(\beta_3, \gamma) = \left| \frac{C_{\text{sim}}(\beta_3, \gamma) - C_{\text{target}}}{\sigma_{\text{sero}}} \right|, \tag{47}$$

where σ_{sero} denotes the standard deviation derived from the reported 95% confidence interval of the seroprevalence estimate[20].

The third term constrains the simulated active prevalence to be consistent with the seroprevalence survey at the observation date. Let $I_{\text{sim}}(\beta_3, \gamma)$ denote the number of actively infectious individuals predicted by the model at the survey time, and let P_{target} be the corresponding empirical estimate. The active-prevalence loss is defined as

$$J_{\text{act}}(\beta_3, \gamma) = \left| \frac{I_{\text{sim}}(\beta_3, \gamma) - P_{\text{target}}}{\sigma_{\text{active}}} \right|, \tag{48}$$

where σ_{active} is also the standard deviation derived from the reported confidence interval of active prevalence[20].

Together, the three loss terms jointly constrain epidemic timing, cumulative exposure, and active prevalence. We then solve the resulting calibration problem using BO, which iteratively runs the simulation and updates a surrogate model of $J(\beta_3, \gamma)$ to propose new parameter candidates, until convergence criteria are met.

References

1. Geohash.es – geohash tools. https://www.geohash.es/ (2025). Accessed: 2025-12-18.
2. Bondarenko, M., Kerr, D., Sorichetta, A., Tatem, A. & WorldPop. Census/projection-disaggregated gridded population datasets, adjusted to match the corresponding UNPD 2020 estimates, for 51 countries across sub-saharan africa using building footprints, DOI: 10.5258/SOTON/WP00683 (2020).
3. Humanitarian Data Exchange (HDX). Movement range maps. https://data.humdata.org/dataset/movement-range-maps (2025). Accessed: 2025-12-28.
4. Bobrovitz, N. *et al.* Global seroprevalence of SARS-CoV-2 antibodies: A systematic review and meta-analysis. *PLOS ONE* **16**, e0252617, DOI: 10.1371/journal.pone.0252617 (2021). Publisher: Public Library of Science.
5. City Population. Lagos state: Administrative division and population statistics (2024). Accessed: 2025-01-10.
6. Xin, H. *et al.* Estimating the latent period of coronavirus disease 2019 (COVID-19). *Clin. Infect. Dis.* **74**, 1678–1681, DOI: 10.1093/cid/ciab746 (2022).
7. Nishiura, H. *et al.* Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19). *Int. J. Infect. Dis.* **94**, 154–155, DOI: 10.1016/j.ijid.2020.03.020 (2020).
8. He, X. *et al.* Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat. Medicine* **26**, 672–675, DOI: 10.1038/s41591-020-0869-5 (2020).
9. Li, R. *et al.* Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* **368**, 489–493, DOI: 10.1126/science.abb3221 (2020).
10. Vilches, T. N. *et al.* Estimating COVID-19 infections, hospitalizations, and deaths following the US vaccination campaigns during the pandemic. *JAMA Netw. Open* **5**, e2142725, DOI: 10.1001/jamanetworkopen.2021.42725 (2022).
11. Ibrahim, R. L., Ajide, K. B. & Olatunde Julius, O. Easing of lockdown measures in nigeria: Implications for the healthcare system. *Heal. Policy Technol.* **9**, 399–404, DOI: 10.1016/j.hlpt.2020.09.004 (2020).
12. Zhang, L. *et al.* Next generation national household travel survey national origin destination data passenger origin-destination data methodology documentation. *Fed. Highw. Adm. Scholar]* (2021).
13. de Dios Ortúzar, J. & Willumsen, L. G. *Modelling transport* (John wiley & sons, 2024).
14. Wu, X., Guo, J., Xian, K. & Zhou, X. Hierarchical travel demand estimation using multiple data sources: A forward and backward propagation algorithmic framework on a layered computational graph. *Transp. Res. Part C: Emerg. Technol.* **96**, 321–346 (2018).
15. Ma, W., Pi, X. & Qian, S. Estimating multi-class dynamic origin-destination demand through a forward-backward algorithm on computational graphs. *Transp. Res. Part C: Emerg. Technol.* **119**, 102747 (2020).
16. Wu, X. *et al.* Simultaneous estimation of induced, diverted, and ex-post demand for railway passengers: an interpretable machine learning framework based on constrained computational graphs. *Transp. Res. Part E: Logist. Transp. Rev.* **202**, 104283 (2025).

17. Zhou, X. S. *et al.* Flow-through tensors: A unified computational graph architecture for multi-layer transportation network optimization. *Artif. Intell. for Transp.* **1**, 100006 (2025).

18. Balcan, D. *et al.* Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc. Natl. Acad. Sci.* **106**, 21484–21489, DOI: 10.1073/pnas.0906910106 (2009). Publisher: Proceedings of the National Academy of Sciences.

19. Chinazzi, M. *et al.* The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* **368**, 395–400, DOI: 10.1126/science.aba9757 (2020).

20. Audu, R. A. *et al.* Seroprevalence of SARS-CoV-2 in four states of nigeria in october 2020: A population-based household survey. *PLOS global public health* **2**, e0000363, DOI: 10.1371/journal.pgph.0000363 (2022).