Using Traffic Data to Inform Transmission Dynamics for COVID-19 in Southern California

September 2021

A Research Report from the Pacific Southwest Region University Transportation Center

Sze-chuan Suen

Maged Dessouky

Suyanpeng Zhang

Han Yu

Anthony Nguyen

Daniel J. Epstein Department of Industrial and Systems Engineering, University of Southern California

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. PSR-20-SP95	**2. Government Accession No.** N/A	**3. Recipient's Catalog No.** N/A	
4. Title and Subtitle Using Traffic Data to Inform Transmission Dynamics for COVID19 in Southern California		**5. Report Date** September 2021 **6. Performing Organization Code** N/A	
7. Author(s) Sze-chuan Suen, 0000-0001-9453-5863 Maged Dessouky, 0000-0002-9630-6201 Suyanpeng Zhang, 0000-0002-8957-1293 Han Yu, 0000-0001-5019-805X Anthony Nguyen, 0000-0002-7616-8358		**8. Performing Organization Report No.** PSR-20-SP95	
9. Performing Organization Name and Address METRANS Transportation Center University of Southern California University Park Campus, RGL 216 Los Angeles, CA 90089-0626		**10. Work Unit No.** N/A **11. Contract or Grant No.** USDOT Grant 69A3551747109	
12. Sponsoring Agency Name and Address U.S. Department of Transportation Office of the Assistant Secretary for Research and Technology 1200 New Jersey Avenue, SE, Washington, DC 20590		**13. Type of Report and Period Covered** Final report (August 1, 2020-July 29, 2021) **14. Sponsoring Agency Code** USDOT OST-R	
15. Supplementary Notes N/A			
16. Abstract Understanding COVID patterns for disease control means that we need to incorporate population flow in our modeling, as these trends influence disease transmission. We use ADMS traffic data to build a compartmental disease model of COVID-19. We draw on methodology from infectious disease models, traffic data, and facility location models together in a novel way, and to our knowledge, no prior work has leveraged such detailed traffic information over such a large urban area to inform disease control efforts.			
17. Key Words Infectious disease modeling, compartmental model, traffic flow, origin-destination, COVID-19	**18. Distribution Statement** No restrictions.		
19. Security Classif. (of this report) Unclassified	**20. Security Classif. (of this page)** Unclassified	**21. No. of Pages** 22	**22. Price** N/A

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Contents

About the Pacific Southwest Region University Transportation Center

The Pacific Southwest Region University Transportation Center (UTC) is the Region 9 University Transportation Center funded under the US Department of Transportation's University Transportation Centers Program. Established in 2016, the Pacific Southwest Region UTC (PSR) is led by the University of Southern California and includes seven partners: Long Beach State University; University of California, Davis; University of California, Irvine; University of California, Los Angeles; University of Hawaii; Northern Arizona University; Pima Community College.

The Pacific Southwest Region UTC conducts an integrated, multidisciplinary program of research, education and technology transfer aimed at *improving the mobility of people and goods throughout the region.* Our program is organized around four themes: 1) technology to address transportation problems and improve mobility; 2) improving mobility for vulnerable populations; 3) Improving resilience and protecting the environment; and 4) managing mobility in high growth areas.

U.S. Department of Transportation (USDOT) Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by a grant from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof.

Disclosure

Principal Investigators Sze-chuan Suen and Maged Dessousky conducted this research titled, "Identifying Priority Testing Locations in Southern California for COVID-19 With Transmission Dynamics and Network Data" at the Daniel J. Epstein Department of Industrial and Systems Engineering in the Viterbi School of Engineering at the University of Southern California. The research took place from Aug 1, 2020 to July 30, 2021 and was funded by the Office of the Provost at the University of Southern California through the Zumberge Special Solicitation: Epidemic & Virus Related Research and Development Award in the amount of 85,000. The research was conducted as part of the Pacific Southwest Region University Transportation Center research program.

Acknowledgements

We wish to acknowledge financial support of the Office of the Provost at the University of Southern California through the “Zumberge Special Solicitation: Epidemic & Virus-Related Research and Development Award.”

Abstract

Understanding COVID patterns for disease control means that we need to incorporate population flow in our modeling, as these trends influence disease transmission. We use ADMS traffic data to build a compartmental disease model of COVID-19. We draw on methodology from infectious disease models, traffic data, and facility location models together in a novel way, and to our knowledge, no prior work has leveraged such detailed traffic information over such a large urban area to inform disease control efforts. We find that it is possible to incorporate traffic data into compartmental models. The model estimates that the number of recovered and vaccinated individuals in the population is substantial, indicating that the completely susceptible proportion of the population is likely small – good news if we want to reduce the number of new cases. However, these results are sensitive to model assumptions around traffic flow patterns, infection rates, and other disease parameters, and we leave more rigorous testing around these values to future work.

Research Report

Executive Summary

This project had two objectives. The first was to develop an optimization problem to estimate origin and destinations of traffic flow from vehicular data on arterial highways in LA County. We cleaned and organized traffic flow data, chose nodes and highways for inclusion in the project, and ran data on weekdays and weekends to demonstrate feasibility. The second goal was to formulate a dynamic transmission model of COVID over eight service planning areas using the traffic flow data to estimate population flow/transmission patterns. We incorporated data from the medical literature to inform the model which uses a classic compartmental model structure.

Motivation: There is an urgent need for performing widespread COVID-19 testing to control disease spread, as many infected individuals are asymptomatic or have mild flu-like symptoms, yet are still transmissive. These patients may not seek care and therefore cannot be diagnosed or undergo quarantine, resulting in subsequent fatal downstream infections. Officials have called for increasing testing as a critical step needed to reopen the country.

However, frequent, regular, and complete population testing across the entire US population, or even over a complete metropolitan area, is prohibitively challenging as testing supplies are limited and require trained health staff which could be better put to use in caring for those in need. It is therefore critical to focus testing in high-priority areas, where tests are likely to capture positive cases. While this includes high risk individuals (contacts of positive cases, elderly in nursing homes, etc.), identifying infected individuals more generally as tests become more widely available will provide crucial information on overall disease prevalence and spread to inform future disease control efforts. Synthesizing and using traffic patterns as transportation patterns change will shed light on possible transmission patterns in populated urban areas such as Los Angeles County.

We therefore propose using the USC Archived Data Management System (ADMS)(USC Viterbi School of Engineering Integrated Media System Center & Metrans Transprtation Center USC CSULB, n.d.), which collects and stores traffic data, to create an epidemic model informed by up-to-date origin-destination traffic data. We will use the model to identify which of the service planning regions (SPAs) in Los Angeles (LA) county are at highest risk for unidentified cases and direct testing resources there. This allows our recommendations to incorporate change in transportation patterns due to disease mitigation policies (e.g., social distancing recommendations, etc.).

Research Methodology: To achieve our dual aims, we relied on the Archived Data Management System (ADMS) for information on LA traffic patterns. ADMS collects, archives, and integrates a

variety of transportation datasets from Los Angeles, Orange, San Bernardino, Riverside, and Ventura Counties. ADMS includes access to real-time traffic datasets from i) 9500 highway and arterial loop detectors providing data approximately every 1 minute, and ii) 2500 bus and train GPS location (AVL) data operating throughout Los Angeles County. This data is collected from road sensors that record traffic.

To achieve Aim 1, we used the ADMS data for the 8 service planning areas (SPAs) in LA county to infer dynamic origin-destination patterns using a mathematical optimization approach, as the data was not able to directly tell us where the traffic flow originated nor where its final destination would be. The optimization problem was solved using computational resources provided by USC Viterbi School of Engineering (their HPCC server).

In Aim 2, we used the inferred origin-destination values in a dynamic transmission network model across different areas of LA. To capture COVID-19 disease dynamics, we used a compartment model of disease where populations may flow between different health/treatment states (Susceptible, Exposed, Infected, Unidentified Infected, Identified Infected, Hospitalized, Recovered, Vaccinated, and Dead). The rate of flow between these compartments is determined by a variety of factors, such as the rate of transmission, disease progression, clearance, treatment, death, etc. Compartmental models like these can be captured in a system of differential equations, which allow for tractable analysis and simulation. Mixing between groups from different geographical areas drives the rate of new infections along with the number of susceptible and infected individuals in both areas. This mixing pattern between areas is described by a matrix which was informed by traffic data.

Results: Model calibration and validation showed good model performance. To validate the model, we compared model outcomes on COVID deaths, the number of identified cases (case counts), and hospitalizations, all overall and for each SPA. While we expected, and saw, larger error values for SPA-level outcomes, total error values were within reasonable values.

These results demonstrated that incorporation of traffic data into compartmental models is possible, and that geographically-stratified models may be useful for infectious disease prediction and control. In future work, we plan to further refine this model before using it to generate recommendations on where vaccination and testing centers ought to be located to be most effective.

Introduction

There is an urgent need for performing widespread COVID-19 testing to control disease spread, as many infected individuals are asymptomatic or have mild flu-like symptoms, yet are still transmissive. These patients may not seek care and therefore not be diagnosed or undergo quarantine, resulting in subsequent fatal downstream infections. Officials have called for increasing testing as a critical step needed to reopen the country.

However, frequent, regular, and complete population testing across the entire US population, or even over a complete metropolitan area, is prohibitively challenging as testing supplies are limited and require trained health staff which could be better put to use in caring for those confirmed to be infected. It is therefore critical to focus testing in high-priority areas, where tests are likely to capture positive cases. While this includes high risk individuals (contacts of positive cases, elderly in nursing homes, etc.), identifying infected individuals more generally as tests become more widely available will provide crucial information on overall disease prevalence and spread to inform future disease control efforts. Synthesizing and using traffic patterns as transportation patterns change will shed light on possible transmission patterns in populated urban areas such as Los Angeles County.

We therefore propose using the USC Archived Data Management System (ADMS), which collects and synthesizes traffic data, to create an epidemic model informed by up-to-date origin-destination traffic data. We will use the model to identify which of the service planning regions (SPAs) in Los Angeles (LA) county are at highest risk for unidentified cases and direct testing resources there. This allows our recommendations to incorporate change in transportation patterns due to disease mitigation policies (e.g., social distancing recommendations, etc.).

Methods

Overview

To achieve our dual aims, we rely on the Archived Data Management System (ADMS) for information on LA traffic patterns. ADMS collects, archives, and integrates a variety of transportation datasets from Los Angeles, Orange, San Bernardino, Riverside, and Ventura Counties. ADMS includes access to real-time traffic datasets from i) 9500 highway and arterial loop detectors providing data approximately every 1 minute, and ii) 2500 bus and train GPS location (AVL) data operating throughout Los Angeles County. This data is collected from road sensors that record traffic.

To achieve Aim 1, we use the ADMS data for the 8 service planning areas (SPAs) in LA county (Los Angeles County Department of Public Health, 2020) to infer dynamic origin-destination patterns using a mathematical optimization approach, as the data was not able to directly tell us where the traffic flow originated nor where its final destination would be. The optimization

problem was solved using computational resources provided by USC Viterbi School of Engineering (their HPCC server).

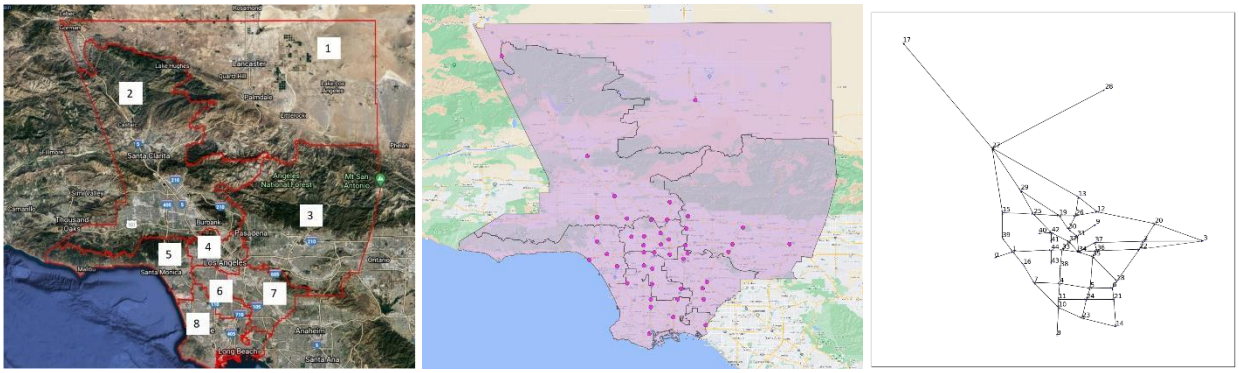
In Aim 2, we used the inferred origin-destination values in a dynamic transmission network model across different areas of LA. To capture COVID-19 disease dynamics, we propose using a compartment model of disease where populations may flow between different health/treatment states (Susceptible, Exposed, Infected, Unidentified Infected, Identified Infected, Hospitalized, Recovered, Vaccinated, and Dead). While there exist other compartment models for COVID-19, these do not focus on LA County, as our work does (Chatterjee et al., 2020; Giordano et al., 2020; Grant, 2020; Ivorra et al., 2020); we expect the county-specific transportation and case counts in LA will be important for future disease prediction. The rate of flow between the compartments in the model is determined by a variety of factors, such as the rate of transmission, disease progression, clearance, treatment, death, etc. Compartmental models like these can be captured in a system of differential equations, which allow for tractable analysis and simulation. Mixing between groups from different geographical areas drives the rate of new infections along with the number of susceptible and infected individuals in both areas. This mixing pattern between areas is described by a matrix which was informed by traffic data.

We describe these steps in more detail in the next sections.

Traffic Flow Analysis

We need to first infer origin-destination patterns from traffic flow information. We do this by identifying a network of nodes over the service planning areas, which are defined by the Los Angeles County Department of Public Health (see Figure 1)(Los Angeles County Department of Public Health, 2020). Nodes are selected intersections of major highways in Los Angeles County.

Figure 1. Service planning areas (left), selected road sensor nodes (middle), and resultant network (right) for analysis.



Data

We used the traffic data from the ADMS database. This data provides traffic information using sensors on highways and local roads. The information includes data on vehicle speed and volume (the number of vehicles that passed by per sensor every 30 seconds).

In order to process the data used for the optimization model, we sum the volume of each sensor on the network link for each 30 second interval. For calculating the speed on the link, we use the average speed from all sensors on that link. Also, in each evaluation time interval, we sum up the volume within that time interval and average speed in the interval. The optimization model also uses traffic information on local roads. We aggregate all volume data on local roads within a small range (1 degree in latitude and longitude direction) of each node and use it as a lower bound constraint on the number of originating and departing demand at that node.

Inference of Origins and Destinations Through Optimization

The main objective for our model is to predict the traffic flow between each origin and destination. The approach is similar to prior work in this area (Ma & Qian, 2018).

To predict each link's traffic flow -- how many cars start from origin to destination, or OD -- we use a non-negative least square model to ensure each link's estimated traffic flow has the smallest least square error compared with the true traffic flow. We formulated a basic optimization model along with a formulation with additional constraints (local traffic flow, symmetricity constraints, etc.), as initial outcomes seemed unlikely and were highly sensitive to assumptions. Adding additional data and constraints therefore seemed to be a reasonable way to arrive at more realistic outcomes. The formulation is presented below.

Basic Optimization model:

The model uses traffic volume and speed data to estimate one day dynamic OD. We first discretize the continuous-time traffic flow into time intervals and assume the average speed remains the same within each time interval. The objective function computes the L2 norm error between the observed link *a* flow $x_a(t)$ on time *t*. All variables with the *hat* notation denotes estimated values. *T* is the time interval set. $q^k_{ij}(t_2)$ represents the estimate traffic flow from origin *i* to destination *j* following path *k* at time t_2. We choose the first three shortest paths between each OD pair. Our problem is therefore:

$$\begin{aligned} \underset{\{\hat{q}^k_{ij}(\cdot)\}_{k,i,j}}{\text{minimize}} \quad & \sum_a \sum_{t \in T} \|x_a(t_1) - \hat{x}_a(t_1)\|^2 \\ \text{subject to} \quad & \hat{q}^k_{ij}(t_2) \geq 0 \quad \forall t_2 \in T \;\; \forall i,j \in \Psi \;\; \forall k \in \Phi_{rs}, \end{aligned}$$

We next turn to the relationship between the link flow $\hat{x}_a(t_1)$ and path flow $\hat{q}^k_{ij}(t_2)$. Define

$$\hat{x}_a(t_1) = \sum_{rs \in \Psi} \sum_{k \in \Phi_{rs}} \rho_{rs}^{ka}(t_1, t_2) \hat{q}_{rs}^k(t_2)$$

Here, Ψ is the set of OD pairs and Φ_{rs} is the set of paths that originated from *r* to *s*. $\rho_{rs}^{ka}(t_1, t_2)$ is the portion of the *k*th path flow departing within time interval t_2 between OD pair *rs* which arrives at link *a* within time interval t_1. As we assumed that the vehicles are spread evenly in time and space, $\rho_{rs}^{ka}(t_1, t_2)$ is calculated by the speed and volume data from each link.

For each traffic pattern, we compute the route choice portions for all OD pairs. Define route choice portion $p_{rs}^k(t_1)$ that it distributes OD demand $q_{rs}(t_1)$ to path flow $q_{rs}^k(t_1)$ using the following equation:

$$q_{rs}^k(t_1) = p_{rs}^k(t_1) q_{rs}(t_1)$$

We use a Logit-based model based on mean travel time for each traffic pattern:

$$p_{rs}^k(t_1) = \frac{\exp\left(-\theta c_{rs}^k(t_1)\right)}{\sum_{k \in \Psi_{rs}} \exp\left(-\theta c_{rs}^k(t_1)\right)}$$

where $c_{rs}^k(t_1)$ represents the mean travel time of path flow *k* in OD *rs* departing at time t_1 for all days. θ is a dispersion factor in the Logit model.

Then the basic optimization model becomes:

$$\begin{aligned} \underset{\{\hat{q}_{ij}^k(\cdot)\}_{k,i,j}}{\text{minimize}} \quad & \sum_a \sum_{t \in T} \left\| x_a(t_1) - \sum_{rs \in \Psi} \sum_{k \in \Phi_{rs}} \rho_{rs}^{ka}(t_1, t_2) p_{rs}^k(t_1) \hat{q}_{rs}(t_1) \right\|^2 \\ \text{subject to} \quad & \hat{q}_{ij}(t_2) \geq 0 \quad \forall t_2 \in T, \end{aligned}$$

Adding local road traffic information to the network: We can use the traffic flow information from local roads to give a lower bound on the total traffic flow that originated from each node at time t_1. Traffic flow $l_m(t_1)$ that originated from node *m*, and traffic flow $d_m(t_1)$ that ends at node *m*. *LB* represents the lower bound on the traffic flow that starts and ends at node *m* on time t_1, which is calculated by local road traffic information, and α is the discount parameter. Here we set $\alpha = 0.8$.

$$l_m(t_1) + d_m(t_1) \geq \alpha \times LB$$

We add these traffic flow constraints into the model by using slack variables.

Adding a constraint around symmetricity: Based on our intuition, the input traffic flow and output traffic flow should not have big difference. Thus, we also introduce a constraint to

ensure this property. The traffic flow f_{ij} from region *i* to region *j* should be similar to the traffic flow f_{ji} from region *j* to region *i* over a day.

$$f_{ij} = f_{ji}$$

Here we define $f_{ij} = \sum_t \sum_{m \in i; n \in j} q_{m,n}(t)$. Similarly, we append this symmetric constraint into our model. We solve the optimization problem to identify OD outcomes for each of our SPAs over each time interval. An example is shown in Figure 2.

Figure 2. Weekday traffic flow 2020.05.18-2020.05.29 (with local road lower bound and symmetricity constraints).

From region	1	2	3	4	5	6	7	8	Sum
1	1672.38	10646.58	17071.54	4802.82	1236.76	845.51	1079.13	1340.76	38695.48
2	4838.59	21167.13	29934.48	9292.96	2280.18	5965.40	2416.65	11281.07	87176.47
3	15297.28	28181.99	31736.14	14865.76	6488.88	12151.00	3131.61	27606.54	139459.19
4	2192.95	9862.55	9408.51	10959.45	8684.06	5107.24	1019.42	2998.53	50232.72
5	1981.73	4646.97	8041.16	12291.89	10663.50	11581.44	2278.57	3376.80	54862.05
6	798.16	2252.44	6982.80	8379.15	2484.85	3477.76	8857.92	6058.13	39291.21
7	1274.24	2992.87	18079.34	2058.00	2538.28	5204.25	12268.14	6263.55	50678.66
8	2794.40	10120.56	12502.05	7197.91	20261.28	8685.91	12031.05	18501.17	92094.33
Sum	30849.74	89871.09	133756.02	69847.94	54637.78	53018.51	43082.49	77426.55	552490.12

This process was repeated for every interval. These outcomes were then used in the dynamic transmission model of disease, which we turn to next.

Dynamic Transmission Model

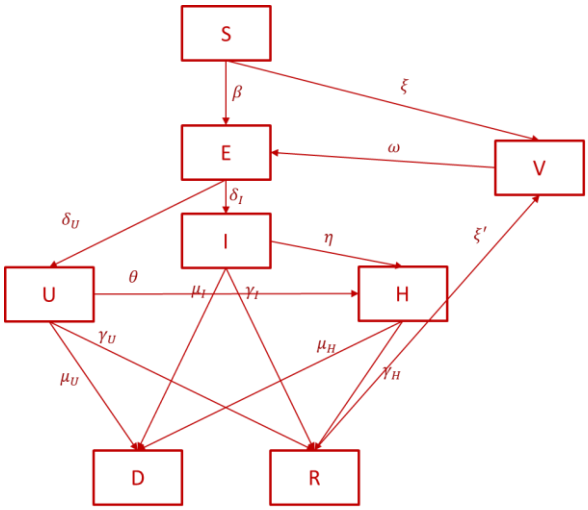
Model Structure

In Aim 2, we used the inferred origin-destination values in a dynamic transmission network model across different areas of LA. To capture COVID-19 disease dynamics, we propose using a compartment model of disease where populations may flow between different health/treatment states (Susceptible, Exposed, Infected, Unidentified Infected, Identified Infected, Hospitalized, Recovered, Vaccinated, and Dead, shown in Figure 3) (Grant, 2020; Kermack et al., 1927). The rate of flow between these compartments is determined by a variety of factors, such as the rate of transmission, disease progression, clearance, treatment, death, etc. Compartmental models like these can be captured in a system of differential equations, which allow for tractable analysis and simulation. Mixing between groups from different geographical areas drives the rate of new infections along with the number of susceptible and

infected individuals in both areas. This mixing pattern between areas is described by a matrix which was be informed by traffic data.

The corresponding ordinary differential equations are:

Figure . Model Schematic.



$$\frac{dS_k}{dt} = -\sum_j \beta_{Ijk}(t) \times S_k(t) \times \frac{I_j(t)}{N_j(t)} - \sum_j \beta_{Ujk}(t) \times S_k(t) \times \frac{U_j(t)}{N_j(t)} - \xi(t) \times S_k(t) \quad (1)$$

$$\frac{dE_k}{dt} = \sum_j \beta_I{}^{jk}(t) \times S_k(t) \times \frac{I_j(t)}{N_j(t)} + \sum_j \beta_{Ujk}(t) \times S_k(t) \times \frac{U_j(t)}{N_j(t)} - (\delta_I(t) + \delta_U(t)) \times E_k(t) + \omega(t) \times V_k(t) \quad (2)$$

$$\frac{dI_k}{dt} = \delta_I(t) \times E_k(t) - (\gamma_I(t) + \mu_I(t) + \eta(t)) \times I_k(t) \quad (3)$$

$$\frac{dU_k}{dt} = \delta_U(t) \times E_k(t) - (\gamma_U(t) + \mu_U(t) + \theta(t)) \times U_k(t) \quad (4)$$

$$\frac{dH_k}{dt} = \eta(t) \times I_k(t) + \theta(t) \times U_k(t) - (\gamma_H(t) + \mu_H(t)) \times H_k(t) \quad (5)$$

$$\frac{dR_k}{dt} = \gamma_I(t) \times I_k(t) + \gamma_U(t) \times U_k(t) + \gamma_H(t) \times H_k(t) - \xi'(t) R_k(t) \quad (6)$$

$$\frac{dV_k(t)}{dt} = \xi(t) \times S_k(t) + \xi'(t) R_k(t) - \omega(t) \times V_k(t) \quad (7)$$

In the next sections, we discuss how we identify values for these equations; after parameterization, we implement these equations in Python (using scipy.integrate.odeint) and simulate disease spread over time.

Time Intervals

To better match empirically observed disease trends, we allow the model parameters to vary over time, which we have split into different intervals, listed in Table 1 below. Values within the same interval remain constant; these constraints reduce overfitting but allow us the flexibility associated with a time-varying model. While the model remains sensitive to these interval

assumptions, we leave further exploration of these choices to future work due to the limited duration of this project.

Table 1. Intervals in disease model

Interval (Days from 3/1 2020)	Months	Description
0-45	March - Mid April	First lock-down
46-90	Mid April - End May	Lock-down & social events
91-135	June - Mid July	Gradual re-opening & looser rules
136-210	Mid July - End Sep	Second wave
211-300	Oct - End Dec	Holiday season
301-360	End Dec - March	Vaccination1
361-390	March - April	Vaccination2
391-480	April - June	Vaccination3
481-516	June - July	Variants

Calibration

We draw model parameters from values in the literature. However, not all values are directly observable (e.g., the rate of becoming transmissive for unidentified cases). We therefore calibrate the model by changing the following calibrated parameters:

Table 2: Calibration Parameters

Parameter	Description
S_I	Multiplier for β_I, where $\beta_I = S_I \times$ traffic flow
S_U	Multiplier for β_U, where $\beta_U = S_U \times$ traffic flow
γ_I	Recovery rate for identified infectious individuals
γ_U	Recovery rate for unidentified infectious individuals
γ_H	Recovery rate for hospitalized individuals
μ_I	Death rate for identified infectious individuals
μ_U	Death rate for unidentified infectious individuals
μ_H	Death rate for hospitalized individuals
δ_I	Rate from exposed to identified infectious
δ_U	Rate from exposed to unidentified infectious
η	Rate from identified infectious to hospitalized
θ	Multiplier on the rate from unidentified infectious to hospitalized, the rate is $\eta * \theta$
ω	Rate from vaccinated to exposed
ξ	Vaccination rate for susceptible group
ξ'	Vaccination rate for recovered group

These values are set such that the simulation output is matches the following calibration targets:

Table 3: Calibration Targets.

Targets	Description
$D_{i,t}$	Cumulative Death number for SPA i at time t
$C_{i,t}$	Cumulative identified Case number for SPA i at time t

$V_{i,t}$	Vaccination number (first dose) for SPA i at time t
H_t	Hospitalization number at time t

The calibration process involves solving the following optimization problem:

$$\min_{\mathbf{p}} \; w_D \sum_i \sum_t \lambda_i(|\hat{D}_{i,t} - D_{i,t}|) + w_C \sum_i \sum_t \lambda_i(|\hat{D}_{i,t} + \hat{H}_{i,t} + \hat{R}_{i,t} + \hat{I}_{i,t} - C_{i,t}|) + w_V \sum_i \sum_t \lambda_i(|\hat{V}_{i,t} - V_{i,t}|) + w_H \sum_t |\hat{H}_t - H_t|$$
$$s.t. \;\; \text{Compartmental Model System Dynamics}$$

where $\widehat{D}_{i,t}$ is the simulated death number for SPA i at time t. $\widehat{H}_{i,t}$ is the simulated hospitalized number for SPA i at time t. $\hat{R}_{i,t}$ is the simulated recovered number for SPA i at time t. $\hat{I}_{i,t}$ is the simulated identified infectious number for SPA i at time t. $\hat{V}_{i,t}$ is the simulated vaccinated number for SPA i at time t. $D_{i,t}$ is the actual death number for SPA i at time t. $C_{i,t}$ is the actual cumulative identified case number for SPA i at time t. $V_{i,t}$ is the actual vaccinated number for SPA i at time t. H_t is the actual hospitalized number for SPA i at time t. All empirical data was drawn from the Los Angeles Department of Public Health COVID dashboard (Los Angeles County Department of Public Health, 2021).

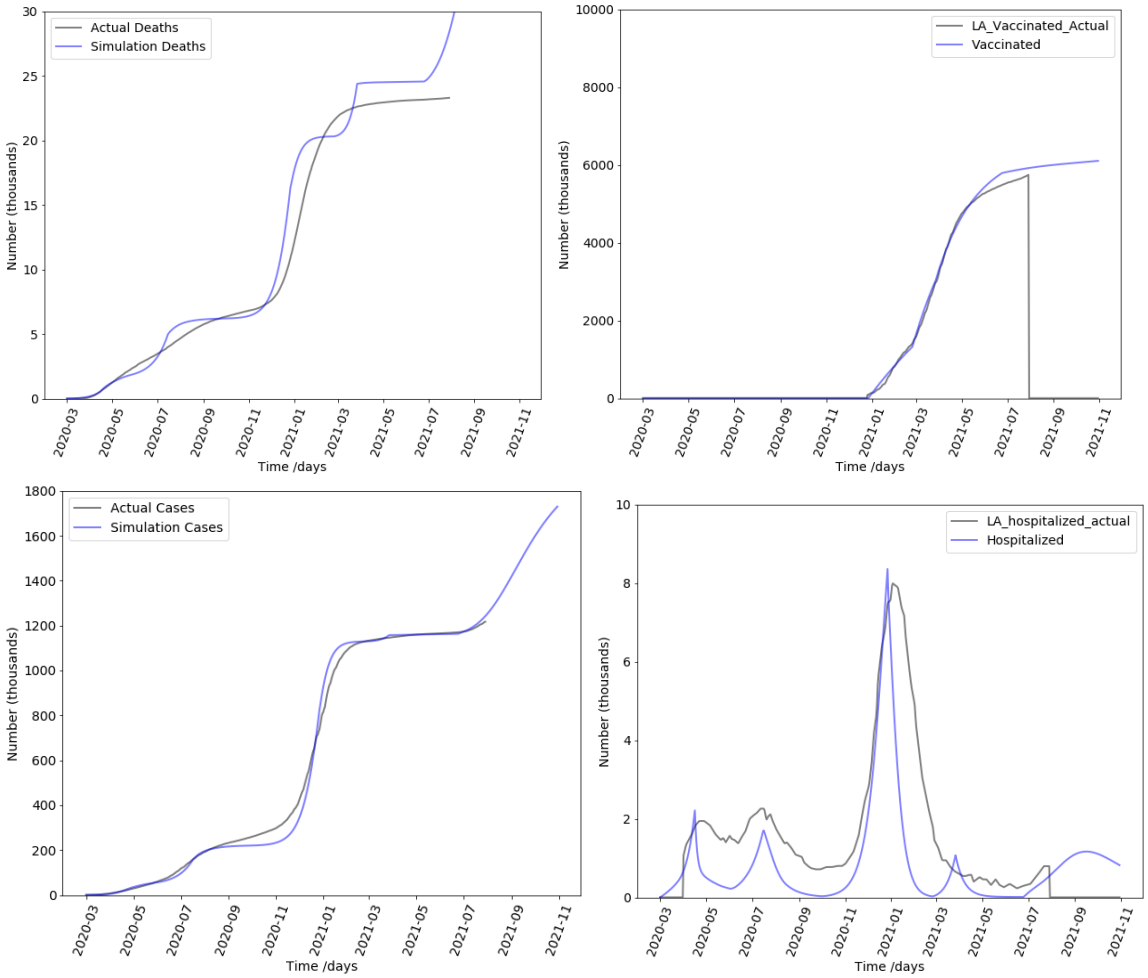
For each of the intervals listed above, we run the ordinary least squares model with this cost function as the objective. We use the calibrated parameter set from the last interval as the initialization point for the current interval. The calibration process is done in Python 3.7.

Results

Dynamic Transmission Model Calibration Outcomes

Calibration outcome values are shown in the Appendix. We show the simulation output compared to empirical data in Figure 4. Panel a through d depicts the deaths, vaccination counts, cases, and hospitalizations, respectively. While there is some noise, we can see that the model outputs seem to capture the general trends seen in the empirical data. While it may be useful to compare outcomes with and without traffic data, this would involve a recalibration of the model, which, due to time constraints, was left for future work.

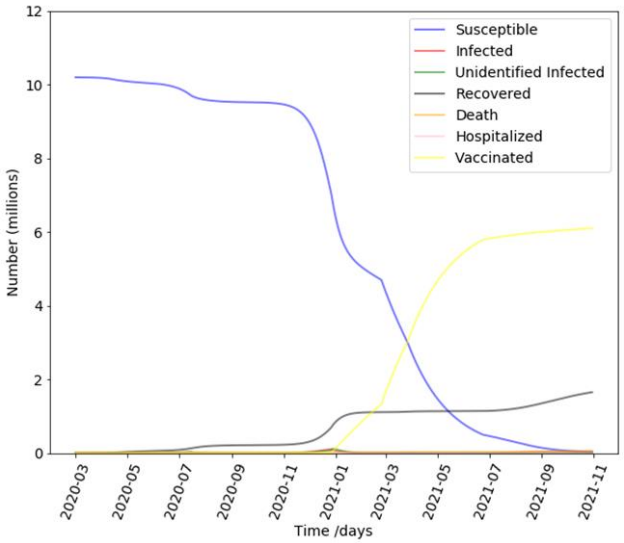
Figure 4. Calibration Results.



Model Results

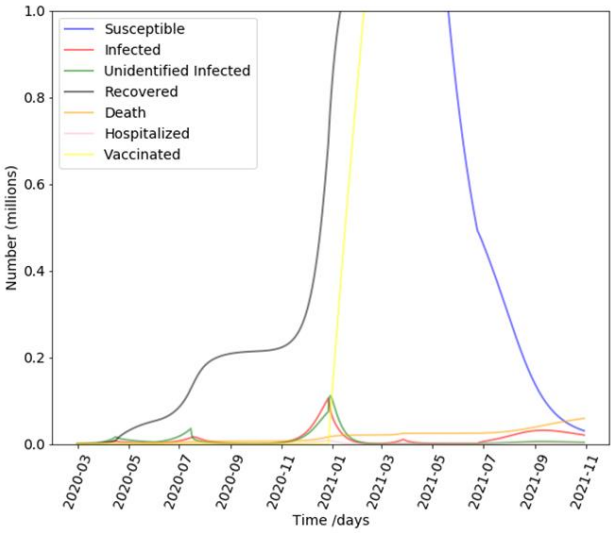
With the model calibrated, we can turn to using the model to perform prediction and simulating counterfactual scenarios. We plot projections for COVID-19 going up to November 2021 in the following Figure 5 and 6, which also provides model outcomes on unobserved values (such as unidentified infected cases).

Figure 5. Model Outcomes.



We additionally provide Figure 6 to better visualize the action between 0 and 1 million in Figure 5.

Figure 6. Model Outcomes (Zoomed In)



These outcomes show that we expect most of the population to be either recovered or vaccinated (and not susceptible) by November 2021, which should greatly reduce the number

of new infections. However, this model does not take into account new variants, which we already know will play an increasingly large role as mutations occur.

Despite this limitation, we believe the model's insights into the ratio of unidentified versus identified cases, effectiveness in vaccination, etc., can still be useful, particularly in explaining trends in the data over the past year. Future projections will heavily depend on policy responses (lockdowns, vaccination booster shots, flu season, new travel, etc.), and therefore continues to be challenging.

Conclusions

This work demonstrates that transmission models can be useful for understanding disease trends, and that such models can incorporate traffic flow patterns. We drew on methodology from infectious disease models, traffic data, and facility location models together in a novel way, and to our knowledge, no prior work has leveraged such detailed traffic information over such a large urban area to inform disease control efforts.

We find that it is possible to incorporate traffic data into compartmental models. The model estimates that the number of recovered and vaccinated individuals in the population is substantial, indicating that the completely susceptible proportion of the population is likely small -- good news if we want to reduce the number of new cases. However, challenges for disease control remain, as variants (which were not included in the model) begin to play a larger role in the number of new cases and deaths.

Directions for Future Research

There remains much work to be done to translate these findings into actionable policy. We ultimately intend for these results to inform the placement of testing sites across the health districts. Achieving this goal would necessitate further stratification of the dynamic transmission model (and traffic data) by health district in addition to service planning area.

This presents additional challenges when calibrating the model, as limited data is available at such fine geographical resolution. Upon calibration, the model would then be used for predicting disease burden in each location. This information could then be used to prioritize districts for testing. Recommendations for the placement of testing center would additionally consider the ease to which individuals could access the site by incorporating road information to and from the selected locations for testing within the prioritized health districts. These recommendations could additionally consider proximity of testing sites to vulnerable populations (nursing homes, dialysis centers, etc.).

References

Chatterjee, K., Chatterjee, K., Kumar, A., & Shankar, S. (2020). Healthcare impact of COVID-19 epidemic in India: A stochastic mathematical model. *Medical Journal Armed Forces India*. https://doi.org/10.1016/j.mjafi.2020.03.022

Giordano, G., Blanchini, F., Bruno, R., Colaneri, P., Di Filippo, A., Di Matteo, A., & Colaneri, M. (2020). Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nature Medicine*, 1–6. https://doi.org/10.1038/s41591-020-0883-7

Grant, A. (2020). Dynamics of COVID-19 epidemics: SEIR models underestimate peak infection rates and overestimate epidemic duration. *MedRxiv*, 2020.04.02.20050674. https://doi.org/10.1101/2020.04.02.20050674

Ivorra, B., Ferrández, M. R., Vela-Pérez, M., & Ramos, A. M. (2020). Mathematical modeling of the spread of the coronavirus disease 2019 (COVID-19) taking into account the undetected infections. The case of China. *Communications in Nonlinear Science and Numerical Simulation*, *88*, 105303. https://doi.org/10.1016/j.cnsns.2020.105303

Kermack, W. O. W. O., McKendrick, A. G. A. G., & Walker, G. T. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A*, *115*(772), 700–721. https://doi.org/10.1098/rspa.1927.0118

Los Angeles County Department of Public Health. (2020). *Service Planning Areas*. http://publichealth.lacounty.gov/chs/SPAMain/ServicePlanningAreas.htm

Los Angeles County Department of Public Health. (2021). *LA County Daily COVID-19 Data*. http://publichealth.lacounty.gov/media/coronavirus/data/

Ma, W., & Qian, Z. (Sean). (2018). Estimating multi-year 24/7 origin-destination demand using high-granular multi-source traffic data. *Transportation Research Part C: Emerging Technologies*, *96*, 96–121. https://doi.org/10.1016/j.trc.2018.09.002

USC Viterbi School of Engineering Integrated Media System Center, & Metrans Transprtation Center USC CSULB. (n.d.). *ADMS App*. Retrieved November 22, 2021, from https://adms.usc.edu/app

Data Management Plan

Products of Research
No data was collected for this study. We used ADMS data and values from the existing literature to parameterize the model.

Data Format and Content
ADMS data was accessed through the online portal.

Data Access and Sharing
Requests to access to the ADMS data should be sent to Genevieve Giuliano <giuliano@price.usc.edu>.

Reuse and Redistribution
ADMS access is restricted. Data access must be requested.

Appendix

Dynamic Transmission Model Calibration Outcomes

Traffic Status	**Period**	S_I	S_U	γ_I	γ_U	γ_H	μ_I	μ_U	μ_H
lockdown	March - Mid April	0.00292	0.00229	0.081	0.085	0.061	0.00091	0.00050	0.01530
lockdown	Mid-April - End May	0.00247	0.00208	0.255	0.234	0.389	0.00163	0.00221	0.01320
reopening	June - Mid July	0.00338	0.00432	0.255	0.235	0.389	0.00209	0.00340	0.01330
2nd wave	Mid July - End Sep	0.00365	0.00331	0.209	0.943	0.357	0.00101	0.00197	0.01250
holidays	Oct - End Dec	0.00984	0.00664	0.210	0.943	0.357	0.00211	0.00258	0.01260
vaccination1	End Dec - March	0.00869	0.00825	0.331	0.952	0.493	0.00093	0.00113	0.01020
vaccination2	March - April	0.01480	0.02380	0.117	0.500	0.106	0.02620	0.00884	0.13200
vaccination3	April - June	0.00120	0.00100	0.100	0.500	0.100	0.00100	0.00100	0.00100

(Calibration outcomes, continued.)

Traffic Status	**Period**	δ_I**_i**	δ_U**_u**	η	θ
lockdown	March - Mid April	0.08370	0.19700	0.02370	0.50000
lockdown	Mid-April - End May	0.07000	0.15900	0.02370	0.50000
reopening	June - Mid July	0.07000	0.15900	0.02370	0.50000
second wave	Mid July - End Sep	0.07000	0.18500	0.02370	0.50000
holidays	Oct - End Dec	0.07110	0.18500	0.02370	0.50000
vaccination1	End Dec - March	0.05760	0.31400	0.02370	0.50000
vaccination2	March - April	0.09990	0.00106	0.03410	0.00326
vaccination3	April - June	0.00262	0.00100	0.00151	0.00100
variants	June - July	0.06600	0.02870	0.00768	0.00107

(Calibration outcomes, continued.)

Traffic Status	**Period**	ξ	ξ' **for Identified**	ξ' **for Unidentified**	ω
lockdown	March - Mid April	N/A	N/A	N/A	N/A
lockdown	Mid-April - End May	N/A	N/A	N/A	N/A
reopening	June - Mid July	N/A	N/A	N/A	N/A
second wave	Mid July - End Sep	N/A	N/A	N/A	N/A
holidays	Oct - End Dec	N/A	N/A	N/A	N/A
vaccination1	End Dec - March	0.00283	0.00890	0.08360	0.00266
vaccination2	March - April	0.01400	0.00274	0.00100	0.00135
vaccination3	April - June	0.02030	0.00586	0.00100	0.00100
variants	June - July	0.00700	0.00188	0.00103	0.03260