PNAS nexus

Main Manuscript for

Predictability of human mobility during the COVID-19 pandemic in the United States

Michal Hajlasz[1], Sen Pei[2]

1. Department of Computer Science, Columbia University, 10027 New York, USA

2. Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, 10032 New York, USA

*Corresponding author: Sen Pei.

Email: sp3449@cumc.columbia.edu

Author Contributions: S.P. conceived the study, M.H. performed the analysis, S.P. and M.H. curated the data, M.H. and S.P. investigated the results. S.P. and M.H. drafted, revised, and reviewed the manuscript.

Competing Interest Statement: All authors declare no competing interests.

Classification: Population Biology; Biophysics and Computational Biology

Keywords: human mobility; predictability; human behavior; COVID-19; epidemics

This PDF file includes:

Main Text
Figures 1 to 6

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/pnasnexus/advance-article/doi/10.1093/pnasnexus/pgae308/7721207 by guest on 29 July 2024

© The Author(s) 2024. Published by Oxford University Press on behalf of National Academy of Sciences. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Human mobility is fundamental to a range of applications including epidemic control, urban planning, and traffic engineering. While laws governing individual movement trajectories and population flows across locations have been extensively studied, the predictability of population-level mobility during the COVID-19 pandemic driven by specific activities such as work, shopping, and recreation remains elusive. Here we analyze mobility data for six place categories at the US county level from February 15th, 2020, to November 23rd, 2021 and measure how the predictability of these mobility metrics changed during the COVID-19 pandemic. We quantify the time-varying predictability in each place category using an information-theoretic metric, permutation entropy. We find disparate predictability patterns across place categories over the course of the pandemic, suggesting differential behavioral changes in human activities perturbed by disease outbreaks. Notably, predictability change in foot traffic to residential locations is mostly in the opposite direction to other mobility categories. Specifically, visits to residences had the highest predictability during stay-at-home orders in March 2020, while visits to other location types had low predictability during this period. This pattern flipped after the lifting of restrictions during summer 2020. We identify four key factors, including weather conditions, population size, COVID-19 case growth, and government policies, and estimate their nonlinear effects on mobility predictability. Our findings provide insights on how people change their behaviors during public health emergencies and may inform improved interventions in future epidemics.

Significance Statement

Human mobility can shape the course of epidemics and may change in response to disease outbreaks. This behavior-epidemic feedback needs to be incorporated into next-generation predictive models for infectious diseases. However, a fundamental question remains: whether such mobility change can be predicted during disease outbreaks? Here we analyze human mobility driven by routine activities at the US county level and measure the predictability using an information-theoretic metric. We find that visits to residences under stay-at-home orders in spring 2020 were more predictable, while visits to other places had a higher predictability during summer 2020. Statistical analysis indicates that weather conditions, population size, COVID-19 case growth, and government policies had large influence on mobility predictability for all six place categories.

Main Text

Introduction

Human mobility is the fabric weaving together our society. Societal operations such as social, economic, and cultural exchanges depend critically on how people move around (1–4). Understanding patterns of human mobility is crucial for a variety of applications spanning epidemic control (5–14), social segregation (15–17), and disaster response (18–21). A plethora of studies have been devoted to discovering laws governing the mobility of individuals and population (1–4, 22–29). For instance, Brockmann et al. discovered scaling laws of human mobility, showing that the distribution of traveling distances decays as a power law (4). Alessandretti et al. found meaningful scales in day-to-day human mobility and showed that aggregating mobility at different scales gives rise to the power-law distribution of traveling distances (1). Schläpfer et al. found that the number of visitors to any location decreases as the inverse square of the product of their visiting frequency and travel distance (2). These studies

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/pnasnexus/advance-article/doi/10.1093/pnasnexus/pgae308/7721207 by guest on 29 July 2024

motivated a large body of research on predicting human mobility at various scales using novel computational methods (30–32).

In the context of infectious diseases, human mobility drives the growth and expansion of pathogens. Prior to the COVID-19 pandemic, there was a rich history of researchers using mobility data to study infectious disease dynamics, spanning diverse pathogens with different transmission modes (e.g., respiratory droplet/airborne (5–7) to vector-borne (8–10)) and across different geographic resolutions (within-city, country, global). However, the use of mobility data in epidemic modeling essentially exploded during the COVID-19 pandemic (33–37), due to previously proprietary mobility data becoming publicly available for use in public health response. For example, Perofsky et al. studied the impacts of human mobility on the citywide transmission dynamics of 18 respiratory viruses in pre- and post-COVID-19 pandemic years using mobile device location data (38); Pullano et al. characterized the spatial connectivity of US counties and explored its implications for geographical disease dynamics using mobile phone-derived mobility data (39); Zhang et al. employed aggregated foot-traffic data and developed a behavior-driven epidemic model to predict neighborhood-level COVID-19 spread in New York City (40). Many epidemic models incorporated human mobility data to predict the progression of epidemics (6, 41–44). However, people may change their behaviors in response to disease outbreaks. It is a consensus that next-generation predictive models for infectious diseases need to incorporate the feedback between epidemic spread and behavioral changes (45, 46). Fundamentally, the success of this approach depends on whether we can accurately predict human mobility during disease outbreaks.

Many previous studies have explored the predictability of human mobility (47–51). Song et al. showed that individual-level mobility trajectories are highly predictable (47). Lu et al. used mobile phone call records to show that the theoretical maximum predictability of human mobility can be approached using Markov Chain based models (50). Chen et al. found that colocation information from both social and non-social sources (i.e., colocation with acquaintances or strangers) contains predictive information of an individual's mobility pattern (48). Liu et al. found that there is a switching phenomenon of spatiotemporal interaction patterns in cities, driven by different human mobility during the active state in the daytime and the sleeping state in the nighttime (52). Residents in larger cities tends to have a shorter sleeping time, which may lower the predictability of human mobility.

Despite these studies, whether the predictability of human mobility persisted during the COVID-19 pandemic remains unclear. Mobility patterns in foot traffic differ across different location types for day-to-day activities such as work, shopping, dining, and recreation, which in turn affects human contact rates and the clustering of contacts. Depending on the relative infection risk in those location types, people may change their mobility in specific place categories, either voluntarily or following government policies implementing non-pharmaceutical interventions (NPIs), such as stay-at-home orders, gathering restrictions, and business closures. For epidemic forecasting informed by mobility data, a fundamental question is to what degree these reactive mobility changes can be predicted. Human mobility data collected during the COVID-19 pandemic offer an opportunity to answer this question.

In this study, we analyzed mobility data measuring foot traffic (i.e., visits) to six place categories (i.e., residences, workplaces, transit stations, parks, groceries and pharmacies, and retail and recreation) in US counties from February 15th, 2020 to November 23rd, 2021. We measured the predictability of mobility changes relative to the baseline prior to the pandemic in the six categories using an information-theoretic metric, permutation entropy (PE). PE is a robust measure for studying the complexity of a time series, introduced by Bandt and Pompe (53). Specifically, PE is robust to observational or dynamical noise in the data, is generally faster than other time series complexity measures, and avoids restrictive parametric model assumptions (54). PE has been successfully applied to various fields such as economics (54) and mechanical

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/pnasnexus/advance-article/doi/10.1093/pnasnexus/pgae308/7721207 by guest on 29 July 2024

engineering (55). In biomedical studies, it has been used to quantify predictability of infectious diseases (56), to differentiate between epileptic and non-epileptic electroencephalographic recordings (57), as well as to typify the complexity of short heart period variability series (58).

Our analysis revealed disparate patterns of predictability across place categories. We further explored the factors associated with mobility predictability using statistical methods, controlling for a range of covariates and spatiotemporal autocorrelation. Interestingly, we found that the predictability of all place categories was impacted by four group of factors, namely, weather conditions, population size, COVID-19 growth, and government policies. The nonlinear effects of these factors on mobility predictability were estimated using a generalized additive model (GAM).

Results

Reactive mobility changes in place categories

We used data from the Google COVID-19 Community Mobility Reports (59) to track changes in foot traffic to six place categories (residential locations, workplaces, transit stations, parks, groceries and pharmacies, and retail and recreation) from February 15th, 2020, to November 23rd, 2021. We focus on the human mobility during the period before the start of the first Omicron wave. The time series measured the daily percent change from baseline in visitation, where the baseline is defined as the median daily visitor numbers from the 5-week period of January 3rd – February 6th, 2020 (59). To remove outliers and weekly oscillations (Fig. S1), we performed a data preprocessing to better track the mobility trend over time (Figs. S2-S3, Supplementary Information). The number of counties analyzed for each day varies, depending on the availability of mobility data. For each category, the minimum and maximum numbers of counties for each day are retail and recreation 570 – 2299; groceries and pharmacies 461 – 2256; parks 69 – 589; transit stations 152 – 937; workplaces 427 – 2101; residences 583 – 1113.

Mobility trends for six place categories in the US are presented in Fig. 1. Mobility to residences peaked during the spring wave in 2020, following stay-at-home orders enforced in many jurisdictions across the US. Subsequently, visitation to residences followed an overall decreasing trend, with another lower peak during the winter of 2020, when COVID-19 surged across the US. Visits to residences can partially represent the time people staying at home (i.e., when people stay at home for a longer time, there will likely be more recorded visitors to residences). During the periods when COVID-19 cases grew rapidly (e.g., spring 2020 and winter 2020), people may stay at home longer following stay-at-home orders or voluntarily to avoid infection risk in other places. As a result, mobility to residences tends to be above the baseline during these periods.

In contrast to residences, visitation to other place categories dropped precipitously following the strict control measures in March 2020, up to 60% for workplaces and transit stations in certain counties. Mobility to workplaces gradually increased throughout late 2020 and 2021 but did not recover to the baseline level by November 2021. Visitation to transit stations resumed to the baseline level in 2021. Groceries and pharmacies were the least impacted places during the COVID-19 pandemic with the maximal mobility reduction around 15% for the median value. The mobility to groceries and pharmacies, as well as retail and recreation, was able to resume baseline-level mobility in the summer of 2021. Park visitation showed a more regular seasonal cycle – more visitation in summers and less visitation in winters. For all place categories except residences, there was an anomalous drop in the mobility data during the summer of 2020, possibly caused by changes in the data calculation method.

Human mobility to different place categories may share similar patterns. To explore this similarity across place categories, we performed a t-Distributed Stochastic Neighbor Embedding (t-SNE) clustering (Materials and Methods) using the mobility time series for each place category in US counties with complete mobility data. For workplaces, residences, retail and recreation, and

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/pnasnexus/advance-article/doi/10.1093/pnasnexus/pgae308/7721207 by guest on 29 July 2024

transit stations, mobility time series from different US counties tend to cluster together (Fig. 2A), suggesting that mobility to each of these place categories shared similar patterns across locations. However, these place categories form distant clusters, indicating that they had distinct mobility patterns. Parks and transit stations had different mobility patterns across US counties, as data points for different counties are distant from each other (Fig. 2A).

To understand the seasonality of mobility, we performed a Fourier analysis to examine the power within mobility time series corresponding to different periods (Materials and Methods) (Fig. 3). From February 15th, 2020 to November 23rd, 2021, mobility to workplaces and residences had a strong weekly pattern with a period of 7 days, suggesting that the COVID-19 pandemic have not changed the weekly rhythm of working. Other place categories also showed a weekly periodicity, but the power was relatively lower. Visits to parks were dominated by a period of 345 days, reflecting an annual seasonality. Mobility to retail and recreation, groceries and pharmacies, and transition stations also had a long period around 300 days, possibly due to that the annual seasonality was disrupted by the COVID-19 pandemic.

Predictability of human mobility

We computed the predictability of human mobility within a time interval using permutation entropy. For human mobility, a higher PE implies a more random time series with less predictability. As a result, we defined the predictability of human mobility as $1 - h_n$, where $h_n \in [0,1]$ is the normalized PE (Materials and Methods). This definition has been used to quantify the predictability of infectious disease time series (56). For each place category and county, we used a sliding time window of 30 days starting from each day to compute daily PE. The PE value for each day quantifies the predictability of mobility over the succeeding 30-day period, with $h_n = 1$ representing complete randomness and $h_u = 0$ representing complete order (i.e., strictly increasing or decreasing). For each day, the minimum and maximum numbers of counties with predictability data are listed below: retail and recreation 569 – 2205; groceries and pharmacies 461 – 2164; parks 69 – 552; transit stations 149 – 879; workplaces 427 – 2047; residences 580 – 1090.

The mobility predictability exhibited differential patterns across place categories (Fig. 4). Most predictability values ranged between 0.4 and 0.8, suggesting a certain level of regularity of mobility in all place categories. Notably, the predictability for residences reached a peak above 0.9 in March 2020, possibly due to synchronized stay-at-home orders effected in many locations in the US. In contrast, the predictability for other five place categories dropped significantly during the same period, indicating highly irregular mobility patterns under the disruption caused by the pandemic. In May 2020, visitation to places other than residences showed an increased predictability, reflecting a more regular increasing trend of mobility following relaxation of control measures. The predictability for residences fell during the same period – the change of predictability generally followed opposite directions for residences and other place categories. After summer 2020, the predictability for all place categories fluctuated but with a smaller magnitude. Relatively, predictability for workplaces and residences had the most volatile fluctuations as mobility in these places was more prone to be impacted by government policies and voluntary behavioral changes. Mobility predictability for parks and groceries and pharmacies were more stable throughout the study period. We performed a t-SNE clustering to examine the similarity between predictability time series for different place categories in US counties (Fig. 2B). Residences and workplaces formed their own clusters, while other place categories clustered together, sharing similar patterns of predictability change.

We visualize the spatial pattern of mobility predictability across US counties using county-level predictability averaged over the study period (Fig. 5). Mobility data for different place categories had varying levels of missingness – a good coverage for retail and recreation and groceries and pharmacies versus a limited coverage for parks and transit stations. Despite the data

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/pnasnexus/advance-article/doi/10.1093/pnasnexus/pgae308/7721207 by guest on 29 July 2024

missingness, we can identify several spatial patterns for mobility predictability. For instance, predictability for residences and workplaces was higher in the west coast and Florida. For retail and recreation and groceries and pharmacies, several states such as California, Arizona, and Florida had a higher predictability than other locations in the middle of the US. Visitation to parks was least predictable in the northeast US. Mobility to transit stations was more predictable in metropolitan areas such as San Francisco, Seattle, Los Angeles, San Diego, etc., where commuters more relied on public transportation systems.

To better compare mobility predictability across the US, we show the distribution of average mobility predictability for each place category in the counties within each of the four US regions defined by the US Census Bureau – Northeast, South, Midwest, and West (Fig. S4). In general, mobility in the West region was more predictability, especially for retail and recreation, groceries and pharmacies, parks, and residences. Within each region, there was a large variation of mobility predictability across counties. Notably, the predictability for visiting parks exhibited the strongest variation across the four US regions among the examined six place categories.

Factors associated with mobility predictability

A myriad of factors can impact human mobility. We employed a generalized additive model (GAM) to estimate the effects of a range of factors on mobility predictability (Materials and Methods). To reduce the temporal autocorrelation in predictability data, we used the *monthly* predictability computed using non-overlapping mobility data for each month separately as response variables. We controlled for a range of covariates at the county level including demographic and socioeconomic factors, disease statistics, commuting behavior, governmental response, and weather conditions (Materials and Methods). Specifically, we included the following covariates at the US county level: (1) log-transformed population size; (2) percentage of Black residents; (3) percentage of Hispanic residents; (4) median household income; (5) monthly new COVID-19 cases; (6) a binary indicator for increasing COVID-19 cases; (7) percentage of residents over 60 years old; (8) percentage of residents over 25 years old without high school diploma; (9) the stringency of government responses to the COVID-19 pandemic, as measured by the Oxford government stringency index; (10) monthly average temperature; (11) monthly precipitation; (12) percentage of residents commuting to work by public transit; (13) percentage of residents working from home. A correlation analysis was performed to ensure that no strong collinearity existed in the covariates (Fig. S5). We used tensor product smooth terms to account for the spatial autocorrelation in the response variables and potential nonlinearities and interactions between some variables. A residual analysis was performed to ensure no spatial and temporal autocorrelations existed in the residuals (Figs. S6-S9).

To quantify the impacts of covariates on mobility predictability, we ranked the covariates using F-values for each place category and selected the top five most impactful factors. Interestingly, this analysis identified four groups of variables that persistently imposed strong impacts on mobility predictability across all place categories (Table S1). These include weather conditions (monthly precipitation and temperature), demographic and socioeconomic factors (population size), local COVID-19 spread (increasing or decreasing cases), and government policy mandates (the Oxford government stringency index).

We visualize the partial effects of several factors on mobility predictability for each place category (Fig. 6, Fig. S6). (1). Weather conditions had profound impacts on human mobility. A higher monthly precipitation was associated with a lower mobility predictability for all place categories, indicating an overall more irregular mobility pattern during months with more rains. Temperature was positively associated with mobility predictability for residences, workplaces, retail and recreation, and transit stations, but had a nonlinear impact on predictability for groceries and pharmacies and parks. (2). Population size was positively associated with mobility predictability except for residences and parks. Generally, mobility in counties with more population tended to

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/pnasnexus/advance-article/doi/10.1093/pnasnexus/pgae308/7721207 by guest on 29 July 2024

be more predictable. (3). Local infection trend also had strong impacts on predictability. An increasing trend of COVID-19 cases was associated with a higher predictability for residences but a lower predictability for other place categories. (4). The impact of government response index was more subtle. Stronger government interventions reduced predictability for mobility in groceries and pharmacies, transit stations, and parks. Intermediate level of government interventions increased predictability for workplaces and retail and recreation and rendered mobility in residences less predictable, showing nonlinear effects for these place categories.

Discussion

Understanding whether human mobility can be predicted during a public health emergency is critical for epidemic modeling and disease control. In this study, using an information-theoretic metric and mobility data from Google COVID-19 Community Mobility Reports, we quantified the predictability of mobility in six place categories at the US county level. We found differential patterns of predictability between residences and other place categories and revealed the spatial variations in mobility predictability. We further used statistical methods to identify common key factors that impacted predictability and quantified their effects for different place categories.

Our analysis found that mobility in the six place categories had different patterns of predictability. Notably, predictability change in residences was mostly in the opposite direction with other places. In addition, government policies had disparate impacts on different places. As shown in Fig. 1, visitation to workplaces and residences was more impacted by NPIs but not so much for groceries and pharmacies and parks. Indeed, essential activities such as getting foods and medicines should be less disrupted by government policies. This finding highlights the need to reduce infection risk in the activities that are not easily altered by administrative orders during future public health emergencies (e.g., improving ventilation and providing food delivery).

The predictability of human mobility varied across the US geographically. While some US states shared similar predictability patterns in certain place categories, the underlying mechanism may be different. For instance, on average, both the west coast and Florida had high predictability for visits to residences. Major cities on the west coast (e.g., San Francisco and Seattle) were generally more vigilant to COVID-19 and have implemented relatively strict NPIs. The high predictability might be due to the synchronized mobility under stay-at-home orders. In contrast, Florida has implemented minimal NPIs, which might also lead to more predictable mobility as people's routine activities were less disrupted. As another example, mobility to workplaces in the west coast was more predictable, possibly because there are several major metropolitan areas in these regions with a large number of commuters, whose mobility to workplaces was strongly modulated by NPIs. Whereas, visiting to workplaces in Florida was more predictable potentially because work-related mobility was less impacted by NPIs. The absence of strong NPIs might explain why Florida stood out with high predictability in several place categories; however, the actual reasons need to be further studied.

After the removal of most governmental interventions, mobility predictability still fluctuated throughout the study period. Such fluctuations were more attributable to voluntary behavior changes in response to infection risk. Understanding the reactive behavior changes in different communities and the underlying mechanisms can help design improved interventions to better incentivize people to change their behaviors to curb disease spread.

Mobility data derived mobile phones have been used to study mobility predictability after natural disasters. For instance, Lu et al. analyzed the predictability of population displacement after the 2010 Haiti earthquake (60, 61). Contrary to the common belief that human mobility becomes less predictable after the disaster, they found the predictability of people's trajectories remained high during the three-month period after the earthquake. Our finding of the relatively high mobility predictability during the COVID-19 pandemic in US counties agrees with this previous result.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/pnasnexus/advance-article/doi/10.1093/pnasnexus/pgae308/7721207 by guest on 29 July 2024

In this study, we used PE to measure the randomness of aggregated mobility time series. Since the aggregated time series were synthesized from a large number of individuals with varying socioeconomic and demographic characteristics, our study has the limitation that the findings cannot reflect the change of predictability among different groups of populations during the COVID-19 pandemic. More granular mobility data with associated socioeconomic and demographic information are needed to support in-depth analyses in future works. Another limitation of the study arises from the Google COVID-19 Community Mobility Reports data. The mobility data used the median value of visitors to different place categories from the 5-week period (January 3rd – February 6th, 2020) as the baseline, which only reflects human mobility during the wintertime. Since the mobility data did not cover the period before 2020, our analysis cannot elucidate the overall mobility change caused by COVID-19 relative to the same dates in previous years.

Real-time forecasting of human mobility has broad applications in disease control and public health. Our findings indicate that human mobility during public health emergencies had a certain level of predictability, providing a theoretical basis for forecasting reactive behavior changes. However, we discovered strong temporal variations in mobility predictability, indicating a pronounced non-stationary nature of the mobility dynamics. This non-stationarity imposes a challenge in operational mobility forecasting in real time. For instance, models trained using historical data may not produce reliable forecasts for the future. Existing predictive models using human mobility data typically assume constant mobility or pre-defined mobility change in the forecast horizon. Those assumptions may generate suboptimal predictions in locations with low predictability of mobility. While our analysis did not aim to find which mobility metrics should be employed in disease forecasting, we identified several factors associated with lower mobility predictability. For example, mobility in locations with small populations and high precipitation was generally less predictable during the COVID-19 pandemic, which may lead to larger uncertainty in disease dynamics and create challenges in real-time forecasting. Further analyses on the choice of mobility metrics and their utilities in disease forecasting across different locations are warranted in future research.

Materials and Methods

Data

We used data from the Google COVID-19 Community Mobility Reports (59), downloaded from Harvard Dataverse (62). This US county-level dataset detailed the daily percent change from baseline in visitation to six different categories - parks, groceries and pharmacies, retail and recreation, transit stations, workplaces, and residences. The baseline is the median value from the 5-week period January 3rd – February 6th, 2020 (59). We limited our analysis to data from February 15th, 2020 to November 23rd, 2021, prior to the wide spread of the Omicron variant. Before running the analysis, we performed preprocessing to remove outliers and deal with data missingness in the raw data. Details are presented in Supplementary Information. Demographic and socioeconomic statistics at the county level were obtained from the American Community Survey data in 2021 (63), downloaded using the R package tidycensus (64). COVID-19 data were downloaded from the Johns Hopkins Coronavirus Resource Center (65). Temperature and precipitation data were compiled from the North America Land Data Assimilation System project (66). COVID-19 government response index was download from the University of Oxford COVID-19 Government Response Tracker (67).

Permutation entropy

The permutation entropy method has three inputs: $\{x_t\}_{t=1,\cdots,T}$ – the time series with length T, τ – the embedding time delay, and n – the embedding dimension or order (53). In our analysis, we

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/pnasnexus/advance-article/doi/10.1093/pnasnexus/pgae308/7721207 by guest on 29 July 2024

fixed $\tau = 1$, as in previous studies using PE (53). The output is a single number, representing the complexity of the time series x. Intuitively, permutation entropy uses the order type of subsequences of the time series to quantify the complexity. As it only considers the relative order of two consecutive numbers (instead of using the exact numerical values), it bypasses the need to arbitrarily define data bins for continuous variables for Shannon entropy and is not impacted by the scale of data. Permutation entropy is also more robust to noises in the time series data.

For a time series $\{x_t\}_{t=1,\cdots,T}$, we examine all $n!$ permutations π of order n which are considered as possible order types of n different numbers. For each π we compute the relative frequency

$$p(\pi) = \frac{\#\{t | t \leq T - n, (x_{t+1}, \cdots, x_{t+n}) \; has \; type \; \pi\}}{T - n + 1}.$$

The permutation entropy of order $n \geq 2$ is defined as

$$H(n) = -\sum p(\pi) \log p(\pi),$$

where the sum runs over all $n!$ permutations π of order n. $H(n)$ ranges from 0 to $\log(n!)$. A strictly increasing or decreasing time series will have permutation entropy of 0 and a completely random time series will have permutation entropy of $\log(n!)$. We can normalize the permutation entropy by dividing it by $\log(n!)$:

$$h_n = H(n)/\log(n!).$$

Here h_n is normalized to ensure the range to be within 0 and 1. In our analysis, we used $1 - h_n$ to define the predictability of mobility.

To compute h_n, we need to define the embedding time delay τ and the order n. A large embedding dimension means that we can detect more distinct ordinal patterns. However, there will be fewer observations of these ordinal patterns, causing our estimates $p(\pi)$ to be less reliable (68). Bandt and Pompe (53) provided 3 heuristics for choosing effective values of τ and n. Namely, $\tau = 1$, $n \leq 7$, and $T > n!$. Staniek and Lehnertz (69) provided empirical evidence for the usefulness of these heuristics in calculating permutation entropy. In our analysis, we set $\tau = 1$. As we used a 30-day time window ($T = 30$), we tested $n = 2,3,4$ and found that $n = 4$ in general led to smallest h_n values. We therefore set $n = 4$ in our analysis to estimate the upper limit of predictability across different orders. For each county in each category, we calculated the predictability of the processed data in sliding 30-day time windows. We used the ordpy.permutation entropy function in Python with an embedding dimension of 4 and embedding time delay of 1. If the 30-day time series had more than 3 missing data points, we replaced the predictability with a NaN value.

Statistical analysis

We performed a t-SNE clustering analysis to examine the similarity between different mobility time series. t-SNE is a nonlinear dimensionality reduction technique for embedding high-dimensional data for visualization in a low-dimensional space of two or three dimensions (70, 71). In the analysis, the mobility data were standardized to have zero mean and unit variance, including only counties with no missing data. The standardized data was then reduced to two dimensions using t-SNE from the sklearn library in Python (72). t-SNE reduces high-dimensional data to two dimensions by constructing probability distributions of pairwise similarities in both the high-dimensional and low-dimensional spaces. It then iteratively adjusts the low-dimensional

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/pnasnexus/advance-article/doi/10.1093/pnasnexus/pgae308/7721207 by guest on 29 July 2024

points to minimize the Kullback-Leibler divergence between these distributions, preserving the local structure and clustering of the original data.

A Fourier analysis was employed to examine the seasonality of human mobility using data from February 15th, 2020 to November 23rd, 2021. For the mobility time series in each county without missing data, we performed a Fourier transform to convert the time series from its original domain to a representation in the frequency domain corresponding to different periods. Before running the analysis, we subtracted the mean from the time series in each county and normalized the Fourier transform by dividing it by the length of the time series. This normalization technique allows for comparison between signals with different sampling rates. We computed the power within the time series corresponding to different periods from 1 day to 365 days with a one-day interval and averaged the power across counties. The periods with locally maximum powers are highlighted in Fig. 3.

We used a GAM model to estimate the effect of factors on mobility predictability. GAM is effective in handling data with spatiotemporal autocorrelations and is computationally efficient. Denote $P_{i,t}$ as the predictability for each place category in county i and month t. The model can be expressed as:

$$\begin{aligned} P_{i,t} = \alpha + te(latitude_i, longitude_i, t) + s(latitude_i, longitude_i, bs = "tp") + \beta_1(log\ population_i) \\ + \beta_2(percent\ Black\ residents_i) + \beta_3(percent\ Hispanic\ residents_i) \\ + \beta_4(median\ household\ income_i) + \beta_5(monthly\ cases_{i,t}) + \beta_6(case\ increasing_{i,t}) \\ + \beta_7(percent\ population\ over\ 60_i) \\ + \beta_8(percent\ residents\ over\ 25\ without\ high\ school\ diploma_i) \\ + \beta_9(Oxford\ government\ response\ index) + \beta_{10}(monthly\ temperature_{i,t}) \\ + \beta_{11}(monthly\ precipitation_{i,t}) \\ + \beta_{12}(percent\ residents\ commuting\ to\ work\ by\ public\ transit_i) \\ + \beta_{13}(percent\ residents\ working\ from\ home_i). \end{aligned}$$

Here, α is the intercept. Case increasing is a binary variable representing the trend of infection in each county. For each month and each county, we performed a linear regression, $y = kx + b$, where x is the day since the beginning of the time series and y is the daily confirmed cases. The linear regression was performed for the daily confirmed cases with each month. Case increasing was defined as 1 if the slope of the regression line, k, is positive and 0 otherwise. Government response index was at the state level, and we used the monthly average in each state to represent control stringency in all counties within that state. Percent residents commuting to work by public transit and percent residents working from home represent commuting behavior of residents in each county. We included them in the GAM to control for the impact of commuting behavior on mobility predictability. We accounted for spatial autocorrelation by including the term $te(latitude_i, longitude_i, t)$. This tensor product smooth term allows the model to account for nonlinear relationships between the response variable and the spatial coordinates, latitude and longitude, calculated to be the geographical center of their respective counties. This tensor product term is a smooth function of several variables where the basis is defined using tensor products of bases for single variables (here the variables are latitude, longitude, and time). An addition term $s(latitude_i, longitude_i, bs = "tp")$ was used, where $bs = "tp"$ represents a thin plate spline that is used primarily for smoothing over two or more dimensions. This term was used to additionally capture spatial patterns that are not necessarily time-dependent to further reduce spatial autocorrelation. The basis dimension for the smooth terms was set to be 3, chosen in order to minimize the Akaike Information Criterion, prevent overfitting in the response curves, and yield reasonable results with the gam.check diagnostics tool. The GAM model was implemented using the mgcv package in R (73).

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/pnasnexus/advance-article/doi/10.1093/pnasnexus/pgae308/7721207 by guest on 29 July 2024

We plotted the correlations between all pairs of covariates to ensure that no strong collinearity existed in the explanatory variables (Fig. S5). Data from all county-month combinations with available predictability data were used in the analysis. To ensure the GAM captured the spatiotemporal autocorrelation in the response variable, we performed a residual analysis. Statistical tests indicated that there were no spatial (Moran's I test) (Fig. S8) and temporal (Durbin-Watson test) autocorrelations in the residuals (Fig. S9). Details on statistical tests are provided in Supplementary Information.

Data and code availability

All data used in this study are publicly available. The Google COVID-19 Community Mobility Reports data were downloaded from Harvard Dataverse. Demographic and socioeconomic statistics were obtained from the American Community Survey. COVID-19 data were downloaded from the Johns Hopkins Coronavirus Resource Center. Temperature and precipitation data were compiled from the North America Land Data Assimilation System project. COVID-19 government response index was download from the University of Oxford COVID-19 Government Response Tracker. All data and custom code supporting the statistical analysis are publicly available at GitHub (https://github.com/cheatingthemichal/USCOVID-Mobility-Predictability).

Acknowledgments

This study was supported by funding from National Science Foundation DMS-2229605, Centers for Disease Control and Prevention U01CK000592 and 75D30122C14289, and Council of State and Territorial Epidemiologists NU38OT00297.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/pnasnexus/advance-article/doi/10.1093/pnasnexus/pgae308/7721207 by guest on 29 July 2024

References

1. L. Alessandretti, U. Aslak, S. Lehmann, The scales of human mobility. *Nature* **587**, 402–407 (2020).
2. M. Schläpfer, *et al.*, The universal visitation law of human mobility. *Nature* **593**, 522–527 (2021).
3. L. M. A. Bettencourt, The Origins of Scaling in Cities. *Science* **340**, 1438–1441 (2013).
4. D. Brockmann, L. Hufnagel, T. Geisel, The scaling laws of human travel. *Nature* **439**, 462–465 (2006).
5. C. Viboud, *et al.*, Synchrony, Waves, and Spatial Hierarchies in the Spread of Influenza. *Science* **312**, 447–451 (2006).
6. D. Brockmann, D. Helbing, The Hidden Geometry of Complex, Network-Driven Contagion Phenomena. *Science* **342**, 1337–1342 (2013).
7. D. Balcan, *et al.*, Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc. Natl. Acad. Sci.* **106**, 21484–21489 (2009).
8. A. Wesolowski, *et al.*, Quantifying the Impact of Human Mobility on Malaria. *Science* **338**, 267–270 (2012).
9. A. Wesolowski, *et al.*, Impact of human mobility on the emergence of dengue epidemics in Pakistan. *Proc. Natl. Acad. Sci.* **112**, 11887–11892 (2015).
10. R. Bomfim, *et al.*, Predicting dengue outbreaks at neighbourhood level using human mobility in urban areas. *J. R. Soc. Interface* **17**, 20200691 (2020).
11. S. Chang, *et al.*, Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* **589**, 82–87 (2021).
12. J. S. Jia, *et al.*, Population flow drives spatio-temporal distribution of COVID-19 in China. *Nature* **582**, 389–394 (2020).
13. R. Li, *et al.*, Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* **368**, 489–493 (2020).
14. S. Pei, T. K. Yamana, S. Kandula, M. Galanti, J. Shaman, Burden and characteristics of COVID-19 in the United States during 2020. *Nature* **598**, 338–341 (2021).
15. H. Nilforoshan, *et al.*, Human mobility networks reveal increased segregation in large cities. *Nature* **624**, 586–592 (2023).
16. E. Moro, D. Calacci, X. Dong, A. Pentland, Mobility patterns are associated with experienced income segregation in large US cities. *Nat. Commun.* **12**, 4633 (2021).
17. A. Amini, K. Kung, C. Kang, S. Sobolevsky, C. Ratti, The impact of social segregation on human mobility in developing and industrialized regions. *EPJ Data Sci.* **3**, 1–20 (2014).
18. B. Hong, B. J. Bonczak, A. Gupta, C. E. Kontokosta, Measuring inequality in community resilience to natural disasters using large-scale mobility data. *Nat. Commun.* **12**, 1870 (2021).
19. H. Deng, *et al.*, High-resolution human mobility data reveal race and wealth disparities in disaster evacuation patterns. *Humanit. Soc. Sci. Commun.* **8**, 1–8 (2021).
20. J. Tang, *et al.*, Resilience patterns of human mobility in response to extreme urban floods. *Natl. Sci. Rev.* **10**, nwad097 (2023).
21. X. Song, Q. Zhang, Y. Sekimoto, R. Shibasaki, Prediction of human emergency behavior and their mobility following large-scale disaster in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '14., (Association for Computing Machinery, 2014), pp. 5–14.
22. F. Simini, M. C. González, A. Maritan, A.-L. Barabási, A universal model for mobility and migration patterns. *Nature* **484**, 96–100 (2012).
23. M. C. González, C. A. Hidalgo, A.-L. Barabási, Understanding individual human mobility patterns. *Nature* **453**, 779–782 (2008).
24. A. Noulas, S. Scellato, R. Lambiotte, M. Pontil, C. Mascolo, A Tale of Many Cities: Universal Patterns in Human Urban Mobility. *PLOS ONE* **7**, e37027 (2012).
25. L. Alessandretti, P. Sapiezynski, V. Sekara, S. Lehmann, A. Baronchelli, Evidence for a conserved quantity in human mobility. *Nat. Hum. Behav.* **2**, 485–491 (2018).
26. H. Barbosa, *et al.*, Human mobility: Models and applications. *Phys. Rep.* **734**, 1–74 (2018).

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/pnasnexus/advance-article/doi/10.1093/pnasnexus/pgae308/7721207 by guest on 29 July 2024

27. C. Song, T. Koren, P. Wang, A.-L. Barabási, Modelling the scaling properties of human mobility. *Nat. Phys.* **6**, 818–823 (2010).
28. X.-Y. Yan, W.-X. Wang, Z.-Y. Gao, Y.-C. Lai, Universal model of individual and population mobility on diverse spatial scales. *Nat. Commun.* **8**, 1639 (2017).
29. R. Li, *et al.*, Simple spatial scaling rules behind complex cities. *Nat. Commun.* **8**, 1841 (2017).
30. G. Savcisens, *et al.*, Using sequences of life-events to predict human lives. *Nat. Comput. Sci.* 1–14 (2023). https://doi.org/10.1038/s43588-023-00573-5.
31. D. Y. Kim, H. Y. Song, Method of predicting human mobility patterns using deep learning. *Neurocomputing* **280**, 56–64 (2018).
32. M. G. Beiró, A. Panisson, M. Tizzoni, C. Cattuto, Predicting human mobility through the assimilation of social media traces into mobility models. *EPJ Data Sci.* **5**, 30 (2016).
33. M. U. G. Kraemer, *et al.*, The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* **368**, 493–497 (2020).
34. M. Chinazzi, *et al.*, The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* **368**, 395–400 (2020).
35. H. S. Badr, *et al.*, Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study. *Lancet Infect. Dis.* **20**, 1247–1254 (2020).
36. S. M. Kissler, *et al.*, Reductions in commuting mobility correlate with geographic differences in SARS-CoV-2 prevalence in New York City. *Nat Commun* **11**, 4674 (2020).
37. S. Pei, S. Kandula, J. Shaman, Differential effects of intervention timing on COVID-19 spread in the United States. *Sci. Adv.* **6**, eabd6370 (2020).
38. A. C. Perofsky, *et al.*, Impacts of human mobility on the citywide transmission dynamics of 18 respiratory viruses in pre- and post-COVID-19 pandemic years. *Nat. Commun.* **15**, 4164 (2024).
39. G. Pullano, L. G. Alvarez-Zuzek, V. Colizza, S. Bansal, Characterizing US spatial connectivity: implications for geographical disease dynamics and metapopulation modeling. [Preprint] (2024). Available at: https://www.medrxiv.org/content/10.1101/2023.11.22.23298916v2 [Accessed 23 May 2024].
40. R. Zhang, *et al.*, Behavior-driven forecasts of neighborhood-level COVID-19 spread in New York City. [Preprint] (2024). Available at: https://www.medrxiv.org/content/10.1101/2024.04.17.24305995v1 [Accessed 23 May 2024].
41. S. Riley, Large-Scale Spatial-Transmission Models of Infectious Disease. *Science* **316**, 1298–1301 (2007).
42. V. Colizza, A. Barrat, M. Barthélemy, A. Vespignani, The role of the airline transportation network in the prediction and predictability of global epidemics. *Proc. Natl. Acad. Sci.* **103**, 2015–2020 (2006).
43. S. Pei, S. Kandula, W. Yang, J. Shaman, Forecasting the spatial transmission of influenza in the United States. *Proc. Natl. Acad. Sci.* **115**, 2752–2757 (2018).
44. S. Pei, J. Shaman, Initial Simulation of SARS-CoV2 Spread and Intervention Effects in the Continental US. *medRxiv* 2020.03.21.20040303 (2020). https://doi.org/10.1101/2020.03.21.20040303.
45. J. Bedson, *et al.*, A review and agenda for integrated disease models including social and behavioural factors. *Nat. Hum. Behav.* **5**, 834–846 (2021).
46. J. J. V. Bavel, *et al.*, Using social and behavioural science to support COVID-19 pandemic response. *Nat. Hum. Behav.* **4**, 460–471 (2020).
47. C. Song, Z. Qu, N. Blumm, A.-L. Barabási, Limits of Predictability in Human Mobility. *Science* (2010). https://doi.org/10.1126/science.1177170.
48. Z. Chen, *et al.*, Contrasting social and non-social sources of predictability in human mobility. *Nat. Commun.* **13**, 1922 (2022).
49. D. do C. Teixeira, J. M. Almeida, A. C. Viana, On estimating the predictability of human mobility: the role of routine. *EPJ Data Sci.* **10**, 49 (2021).
50. X. Lu, E. Wetter, N. Bharti, A. J. Tatem, L. Bengtsson, Approaching the Limit of Predictability in Human Mobility. *Sci. Rep.* **3**, 2923 (2013).

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/pnasnexus/advance-article/doi/10.1093/pnasnexus/pgae308/7721207 by guest on 29 July 2024

51. A. Cuttone, S. Lehmann, M. C. González, Understanding predictability and exploration in human mobility. *EPJ Data Sci.* **7**, 2 (2018).
52. C. Liu, *et al.*, Revealing spatiotemporal interaction patterns behind complex cities. *Chaos Interdiscip. J. Nonlinear Sci.* **32**, 081105 (2022).
53. C. Bandt, B. Pompe, Permutation Entropy: A Natural Complexity Measure for Time Series. *Phys. Rev. Lett.* **88**, 174102 (2002).
54. M. Henry, G. Judge, Permutation Entropy and Information Recovery in Nonlinear Dynamic Economic Time Series. *Econometrics* **7**, 10 (2019).
55. R. Yan, Y. Liu, R. X. Gao, Permutation entropy: A nonlinear statistical measure for status characterization of rotary machines. *Mech. Syst. Signal Process.* **29**, 474–484 (2012).
56. S. V. Scarpino, G. Petri, On the predictability of infectious disease outbreaks. *Nat. Commun.* **10**, 898 (2019).
57. M. Zanin, L. Zunino, O. A. Rosso, D. Papo, Permutation Entropy and Its Main Biomedical and Econophysics Applications: A Review. *Entropy* **14**, 1553–1577 (2012).
58. A. Porta, *et al.*, Entropy, entropy rate, and pattern classification as tools to typify complexity in short heart period variability series. *IEEE Trans. Biomed. Eng.* **48**, 1282–1291 (2001).
59. COVID-19 Community Mobility Report. *COVID-19 Community Mobil. Rep.* Available at: https://www.google.com/covid19/mobility?hl=en [Accessed 9 April 2022].
60. X. Lu, L. Bengtsson, P. Holme, Predictability of population displacement after the 2010 Haiti earthquake. *Proc. Natl. Acad. Sci.* **109**, 11576–11581 (2012).
61. L. Bengtsson, X. Lu, A. Thorson, R. Garfield, J. von Schreeb, Improved Response to Disasters and Outbreaks by Tracking Population Movements with Mobile Phone Network Data: A Post-Earthquake Geospatial Study in Haiti. *PLOS Med.* **8**, e1001083 (2011).
62. China Data Lab, Google Community Mobility Reports with Basemap (US). Harvard Dataverse. https://doi.org/10.7910/DVN/1CLYWS. Deposited 14 November 2022.
63. U. C. Bureau, American Community Survey (ACS). *Census.gov*. Available at: https://www.census.gov/programs-surveys/acs [Accessed 9 April 2022].
64. K. Walker, M. Herman, K. Eberwein, tidycensus: Load US Census Boundary and Attribute Data as "tidyverse" and 'sf'-Ready Data Frames. (2021). Deposited 23 September 2021.
65. COVID-19 United States Cases by County. *Johns Hopkins Coronavirus Resour. Cent.* Available at: https://coronavirus.jhu.edu/us-map [Accessed 17 March 2022].
66. NLDAS: Project Goals | LDAS. Available at: https://ldas.gsfc.nasa.gov/nldas [Accessed 1 January 2024].
67. Oxford Covid-19 Government Response Tracker (OxCGRT). (2022). Deposited 20 May 2022.
68. A. M. Brandmaier, pdc: An R Package for Complexity-Based Clustering of Time Series. *J. Stat. Softw.* **67**, 1–23 (2015).
69. M. Staniek, K. Lehnertz, Parameter selection for permutation entropy measurements. *Int. J. Bifurc. Chaos* **17**, 3729–3733 (2007).
70. G. E. Hinton, S. Roweis, Stochastic Neighbor Embedding in *Advances in Neural Information Processing Systems*, (MIT Press, 2002).
71. L. van der Maaten, G. Hinton, Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
72. F. Pedregosa, *et al.*, Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
73. S. N. Wood, Fast Stable Restricted Maximum Likelihood and Marginal Likelihood Estimation of Semiparametric Generalized Linear Models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **73**, 3–36 (2011).

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/pnasnexus/advance-article/doi/10.1093/pnasnexus/pgae308/7721207 by guest on 29 July 2024

Figures and Tables

Figure 1. Google COVID-19 Community Mobility Reports data. Time series show the county-level percent change in visitation from baseline across 6 categories (retail and recreation, groceries and pharmacies, parks, transit stations, workplaces, and residences) in the US. Solid black lines show the median and shaded areas represent 50% and 90% confidence intervals, respectively. Data were preprocessed to remove outliers and weekly oscillations to better show the mobility trend over time (Supplementary Information). The baseline period is defined as the 5-week period January 3rd – February 6th, 2020.

Figure 2. Clustering of mobility patterns across different place categories. We performed t-SNE clustering using the mobility time series (A) and predictability time series (B) at the county level. Each dot represents a time series (mobility in A and predictability in B) for a place category in a single county from February 15th, 2020 to November 23rd, 2021. Colors are used to distinguish different place categories.

Figure 3. Periodicity of human mobility to different place categories. For each place category, we performed a Fourier analysis on the mobility time series in each county without missing data from February 15th, 2020 to November 23rd, 2021. We computed the power corresponding to different periods from 1 day to 365 days and averaged the power across counties. Periods with strong powers within a neighborhood of periods are highlighted for each place category.

Figure 4. Predictability of mobility in six place categories. For each county, the normalized permutation entropy h_n was computed for a 30-day sliding time window starting from each day. The figures show the daily predictability, $1 - h_n$, for each place categories at the county level. The median (solid black lines), 50%, and 90% confidence intervals (shaded areas) were calculated using all US counties with mobility data. A predictability of zero represents complete randomness and a predictability of one represents complete order (i.e., strictly increasing or decreasing).

Figure 5. Average mobility predictability at the US county level. For each county and place category, we show the predictability averaged over the period from February 15th, 2020 to November 23rd, 2021. Counties without complete predictability data for each place category were shown as blank. Red color means high predictability and blue color means low predictability. The range of predictability is shown in the color bars.

Figure 6. The partial effects of population size, government response, precipitation, temperature, and infection trend on mobility predictability for six place categories. Statistical analysis was performed using a GAM controlling for a variety of covariates at the county level. The solid lines show mean values, and the shaded areas represent 95% CIs. For the binary variable, case increasing, we show the slope coefficients representing the change of predictability when the variable changes from 0 to 1. The bars show the 95% CIs. Partial effects were estimated for the period from February 15th, 2020 to November 23rd, 2021, using all counties with monthly mobility predictability data.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/pnasnexus/advance-article/doi/10.1093/pnasnexus/pgae308/7721207 by guest on 29 July 2024

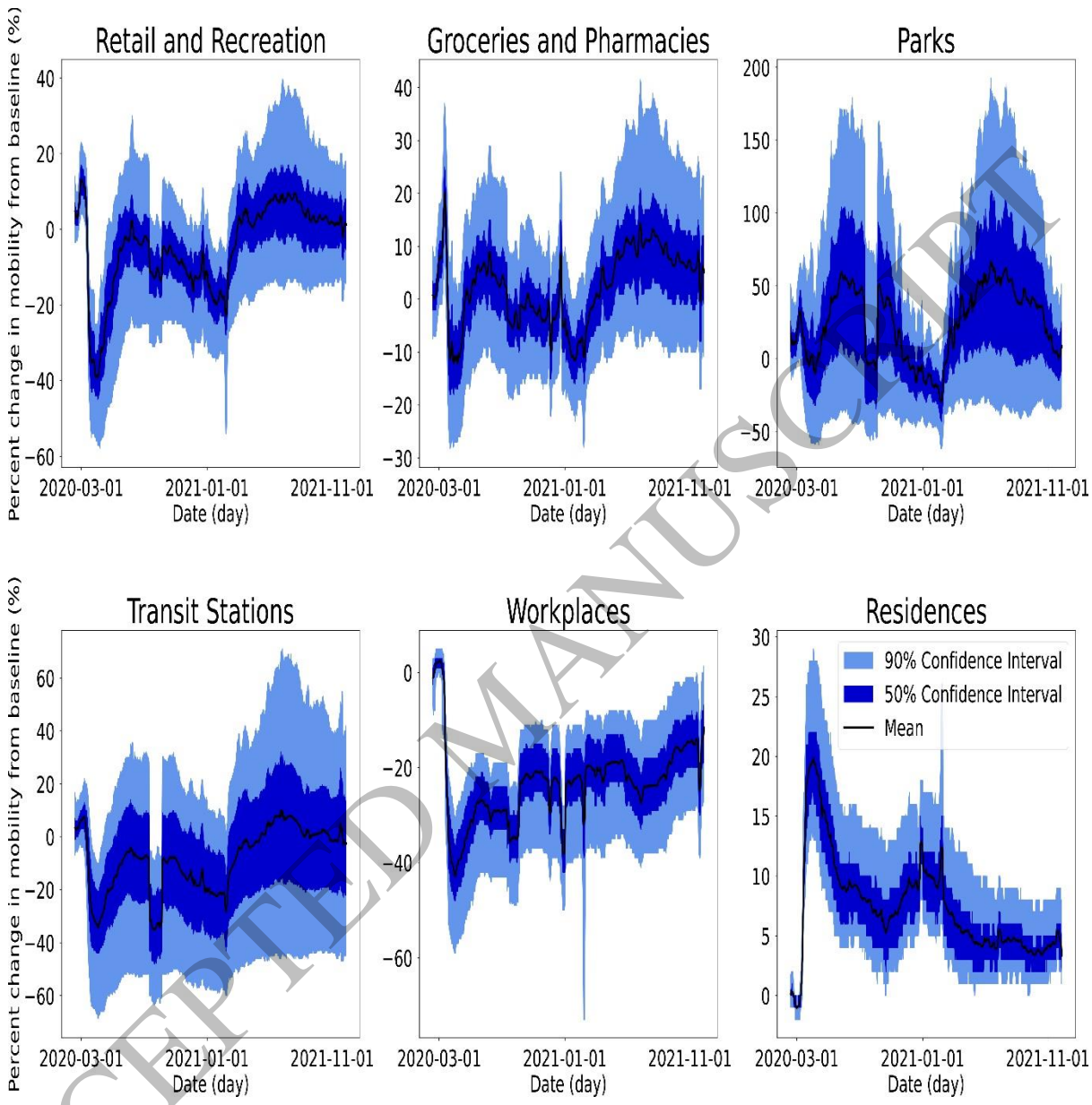

Figure 1
559x355 mm (x DPI)

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/pnasnexus/advance-article/doi/10.1093/pnasnexus/pgae308/7721207 by guest on 29 July 2024

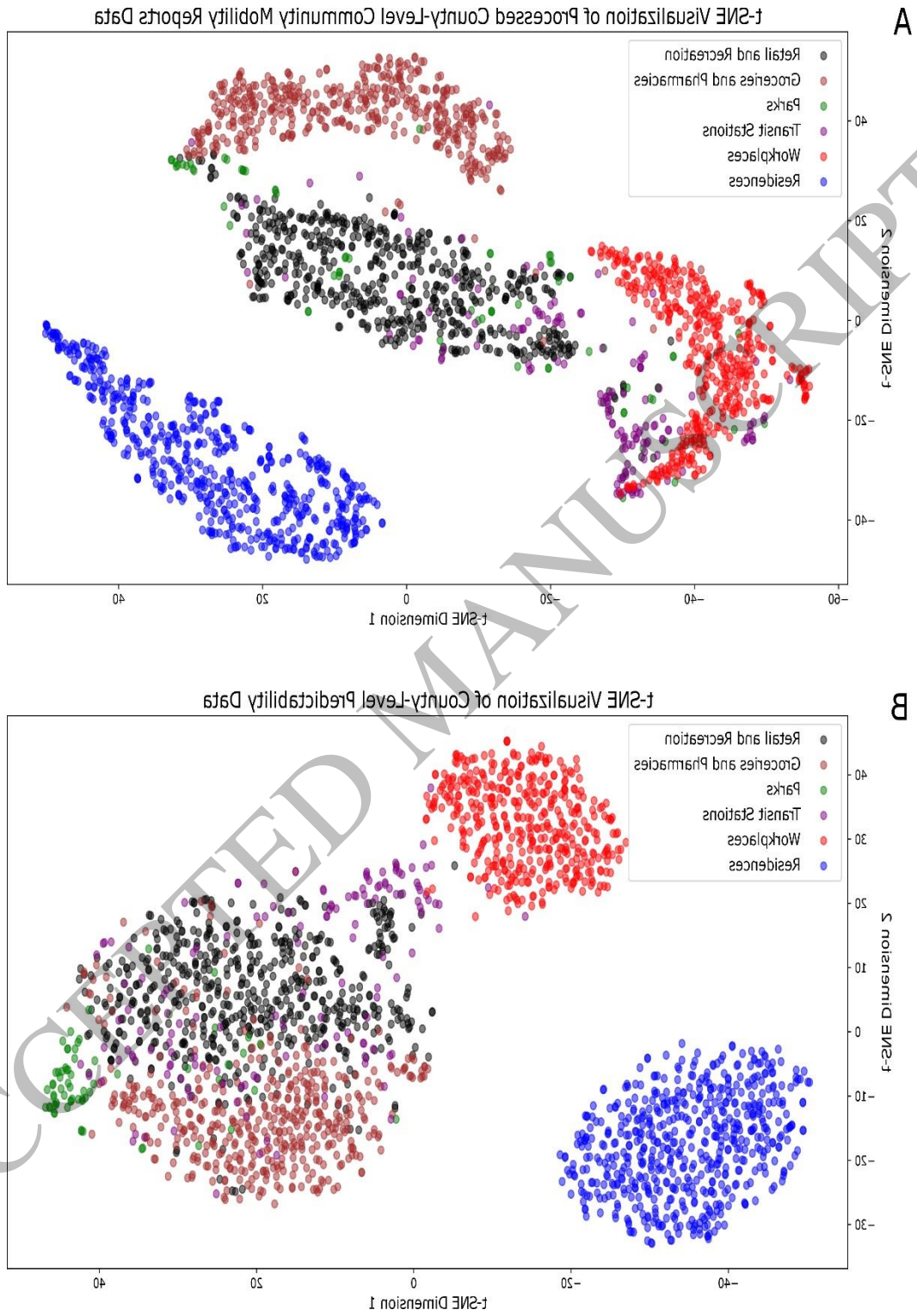

Figure 2
326x348 mm (x DPI)

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/pnasnexus/advance-article/doi/10.1093/pnasnexus/pgae308/7721207 by guest on 29 July 2024

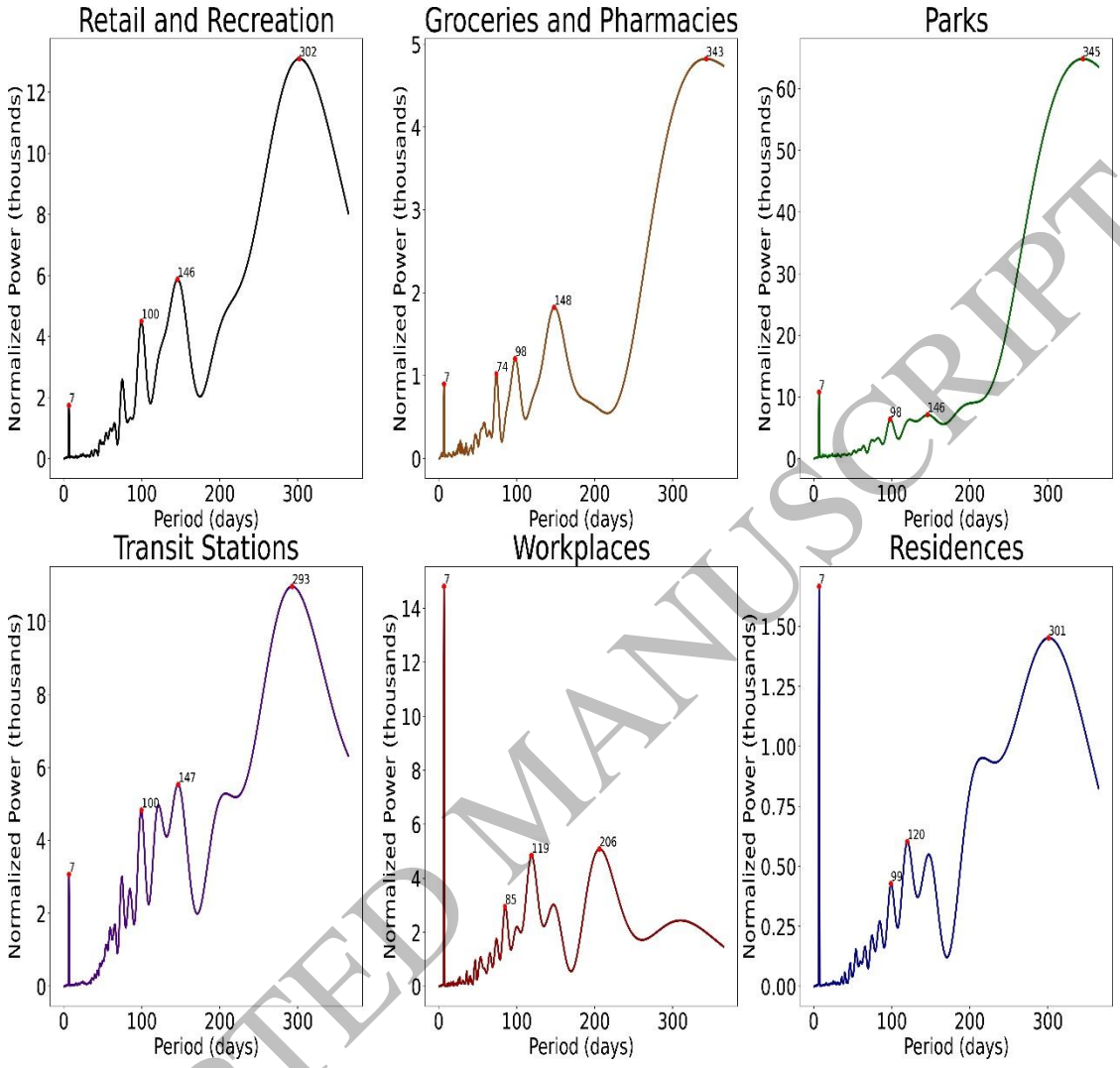

Figure 3
559x376 mm (x DPI)

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/pnasnexus/advance-article/doi/10.1093/pnasnexus/pgae308/7721207 by guest on 29 July 2024

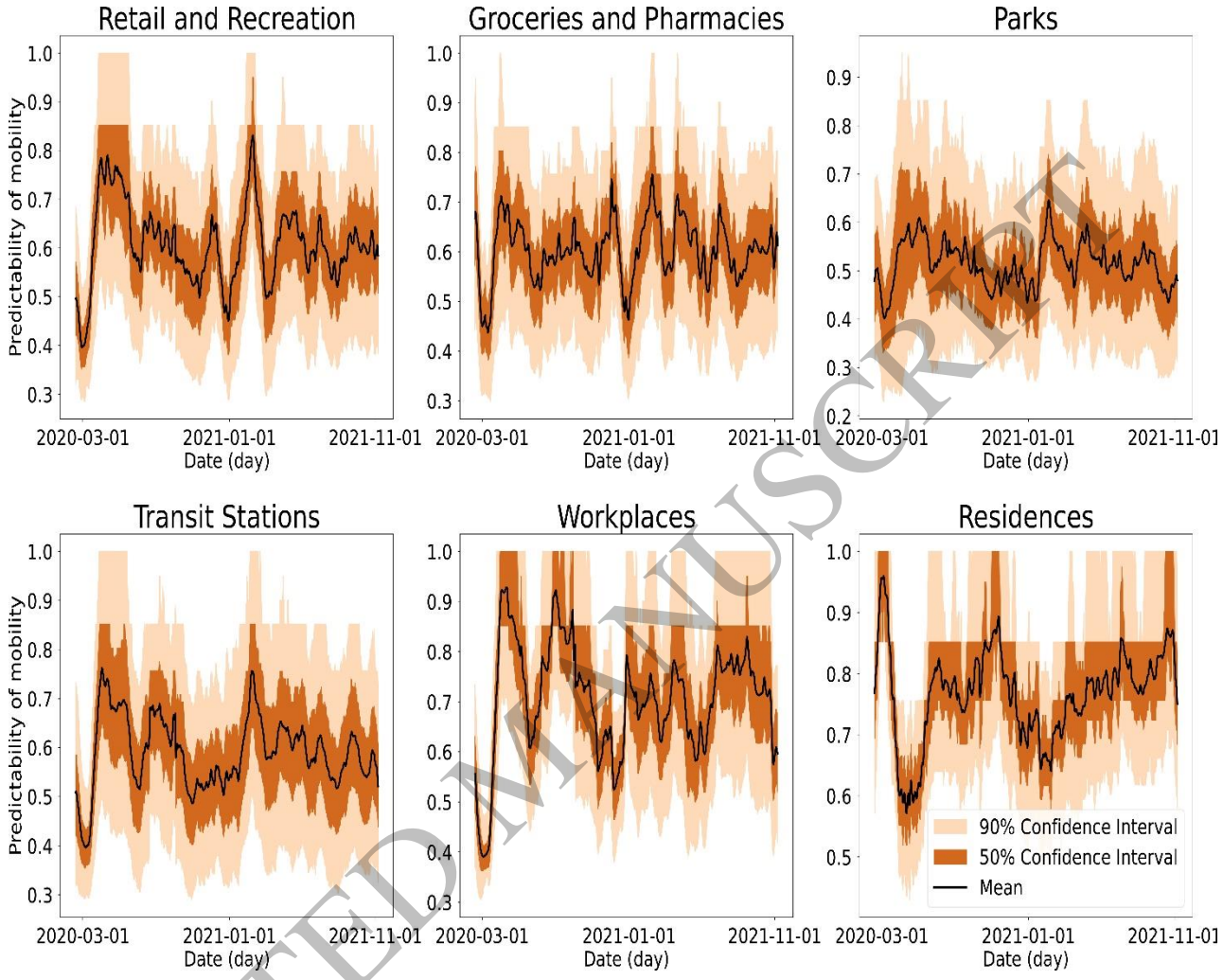

Figure 4
559x348 mm (x DPI)

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/pnasnexus/advance-article/doi/10.1093/pnasnexus/pgae308/7721207 by guest on 29 July 2024

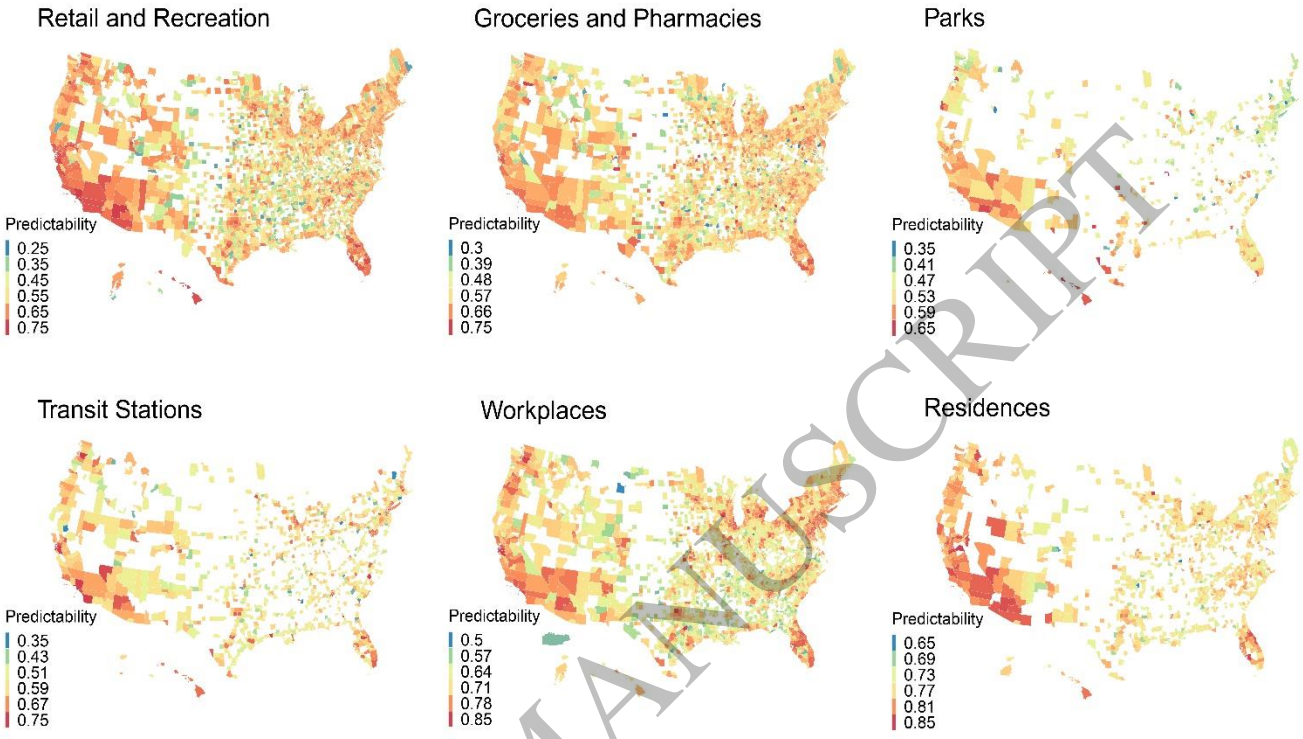

Figure 5
559x314 mm (x DPI)

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/pnasnexus/advance-article/doi/10.1093/pnasnexus/pgae308/7721207 by guest on 29 July 2024

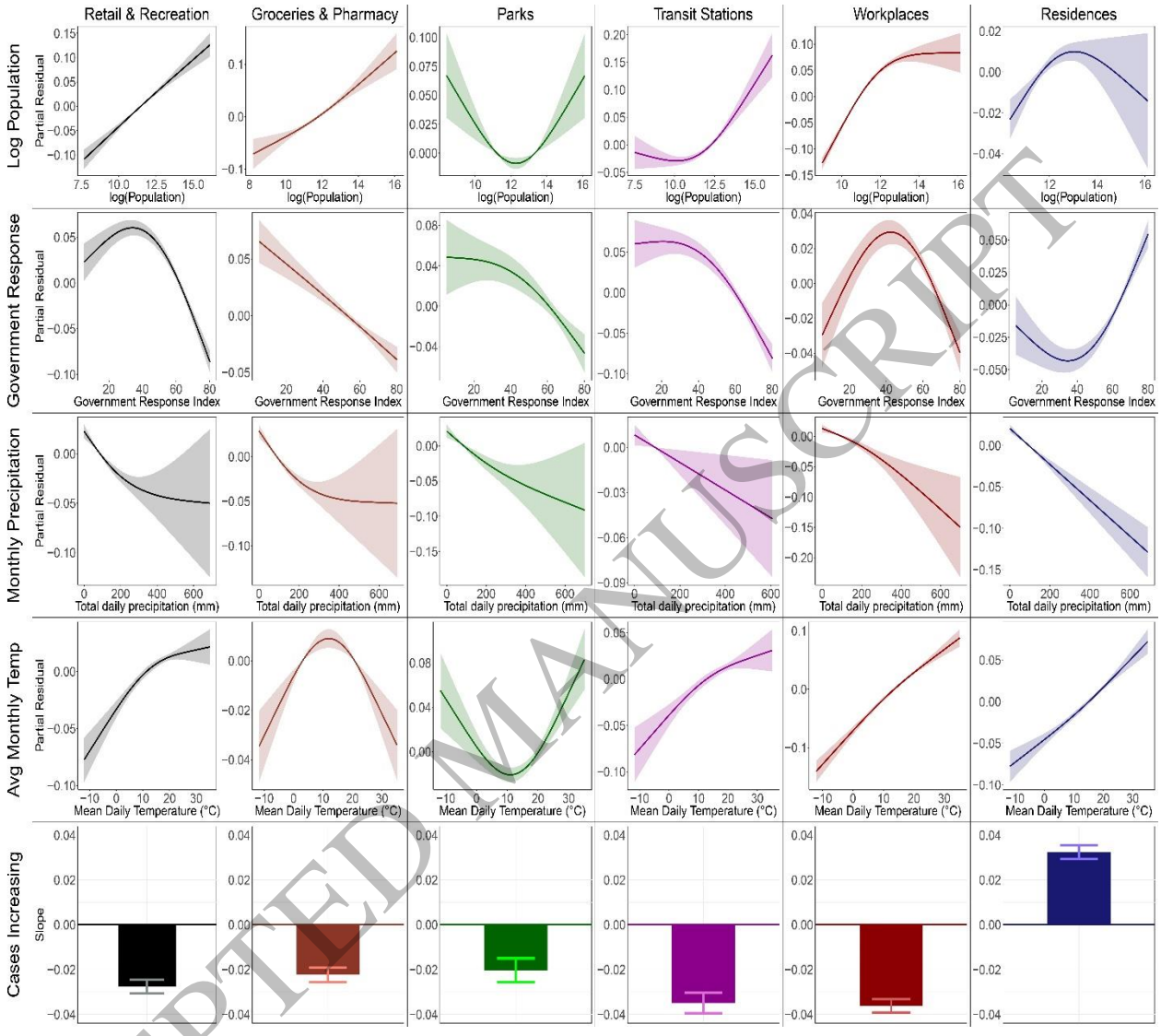

Figure 6
559x419 mm (x DPI)

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/pnasnexus/advance-article/doi/10.1093/pnasnexus/pgae308/7721207 by guest on 29 July 2024