Estimating infection-related human mobility networks based on time series data of COVID-19 infection in Japan

Tetsuya Yamadaa, * and Shoi Shi—a, *

a Graduate School of Medicine, The University of Tokyo, 113-0033 Tokyo, Japan
*b Correspondence: tetsuyamada1222@gmail.com; shoishi0322@gmail.com

Abstract

Comprehensive and evidence-based countermeasures against emerging infectious diseases have become increasingly important in recent years. COVID-19 and many other infectious diseases are spread by human movement and contact, but complex transportation networks in 21st century make it difficult to predict disease spread in rapidly changing situations. It is especially challenging to estimate the network of infection transmission in the countries that the traffic and human movement data infrastructure is not yet developed. In this study, we devised a method to estimate the network of transmission of COVID-19 from the time series data of its infection and applied it to determine its spread across areas in Japan. We incorporated the effects of soft lockdowns, such as the declaration of a state of emergency, and changes in the infection network due to government-sponsored travel promotion, and predicted the spread of infection using the Tokyo Olympics as a model. The models used in this study are available online, and our data-driven infection network models are scalable, whether it be at the level of a city, town, country, or continent, and applicable anywhere in the world, as long as the time-series data of infectious per region is available. These estimations of effective distance and the depiction of infectious disease networks based on actual infection data are expected to be useful in devising data-driven countermeasures against emerging infectious diseases worldwide.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Introduction

Since the identification of the coronavirus disease 2019 (COVID-19) in December 2019 [1, 2], over 180 million laboratory-confirmed infections and 4 million deaths have been observed worldwide as of 30 June 2021 [3]. The development of vaccines has certainly reduced the number of new infections worldwide [4, 5], allowing life before COVID-19 to gradually return to normal. However, in countries where vaccination has been delayed, there is a risk that the infection will continue to spread, and the emergence of new mutant virus strains has been observed [6, 7, 8], thus, proper measures against COVID-19 are still required.

In the past year and a half, effective countermeasures against COVID-19 have been established, such as social distancing and wearing of face masks. The effectiveness of these non-pharmaceutical interventions (NPIs) has also been quantitatively evaluated [9, 10, 11, 12], together with the need for appropriate application where necessary depending on the active status of COVID-19. Otherwise, governments have to take the painful and severe measures again, the lockdowns, to prevent the spread of infection. Lockdowns are very effective infection control measures [13, 14, 10]; however, the trade-off between the impact on the economy and on the mental health of individuals cannot be ignored [15, 16, 17, 18]; thus, large-scale lockdown should be avoided as much as possible. As an alternative, selective lockdowns, which prevent the spread of infection by locking down a city when the number of infected individuals increases in that city, or other specific interventions have also been applied with some success in the United Kingdom and in other countries [19, 20, 21, 22, 23].

Since the spread of infection between cities is caused by the movement of people [24, 25, 26], infection does not always spread to cities that neighbor geographically, but spreads between cities where there is a continuous flow of individuals [27, 28]. Therefore, it is useful to create a network for predicting the spread of infection by calculating effective distances based on traffic and mobility of individuals [29, 30]. In fact, this method has been successful in predicting the spread of infection between countries and cities [31]. Conversely, effective distances are easier to estimate in countries where the human movement and traffic data is easy to be obtained and are well developed, although, this is limited to a few countries. Even in the countries or regions where the human movement and traffic data infrastructure is not yet developed, it is required to share a scheme for estimating infection networks based on the minimum necessary information and taking appropriate countermeasures worldwide. In addition, the power of effective distances is that they vary in a time-dependent and environment-dependent manner, as they are influenced by the number of infected people in the mobile population and the number of individuals taking preventive measures, which is difficult to be captured from movement-data-based network. In this context, it is useful to calculate effective distances and estimate the network structure of infection transmission based on the actual number of infected individuals.

In this context, the dynamics of infected individuals in Japan may be useful to estimate effective distances of viral transmission. In the past year and a half, although there has been no explosion of infections similar to the tens of thousands of new infections and thousands of deaths per day observed in the United States and Europe, there has been a constant rate of infections observed in Japan. The inability to impose legally binding restrictions on the movement of people and the opening of commercial facilities has made it difficult to completely control the number of infected people; although, the cooperative nature of the Japanese might prevent the explosion of infections [32, 33, 34]. In other words, in Japan, it is possible to observe a discrepancy between effective distances based on movement data and on the number of infected people, which is difficult in countries or regions that have experienced a rapid increase in infection and repeated lockdowns that completely restrict movement. In addition, Japan has had three peaks of infection, two declarations of the state of emergency (i.e., with moderate restriction on the movement of people), and two stages of travel campaigns, during which the government encouraged domestic travels by subsidizing travel expenses, between 18 March 2020 and 12 March 2021, which have provided abundant data for describing changes in effective distances (Fig. 1). Thus, the unique dynamics of SARS-CoV-2 infection in Japan is a suitable model for calculating the effective distance of the number of infected people, and provides a new lockdown strategy based on an effective real-life-based network.

In this study, by combining the “Susceptible-Exposed-Infectious-Recovered” SEIR model and the diffusion equation [35, 36, 37], we simulated the infection propagation from Tokyo, the capital of Japan, to other prefectures and calculated the effective distance separating them. We then tested the effective distance model envisioning three different declarations of the state of emergency and a travel campaign. A map of Japan based on effective distance was drawn, and this map was used to predict the dynamics of infection associated with the Tokyo 2020 Olympics.
Results

Diffusion of infected population from Tokyo recapitulated the propagation of COVID-19 pandemic in Japan.

To model the propagation of the COVID-19 pandemic in Japan as a diffusion process of the infected population from Tokyo and to estimate effective distances, we first recapitulated a time series using the SEIR model in which the number of infected people in Tokyo depended on the distinct phases of the pandemic, or periods P1, P2, and P3 (Fig. 1). We calibrated the parameters for the number of potentially newly infected individuals in Tokyo during each period using the Markov chain Monte Carlo (MCMC) algorithm, and the summary of the estimated parameters is listed in Supplementary Table 1, which were generally consistent with previous studies on COVID-19 [35, 37]. Models using the estimated parameters accurately recapitulated the temporal dynamics and amplitude of the outbreaks and surges in infections (Supplementary Figure 1a-c). Leveraging these models as sources, we subsequently set up diffusion equations that were calculated as one-dimensional partial differential equations (PDEs). Thus, human mobility outcomes differed and depended on whether the pandemic was on a spreading or shrinking trend; hence, we solved the diffusion equations dividing each period into periods before and after a peak of infection in Tokyo, which generated six periods (i.e., P1a, P1b, P2a, P2b, P3a, and P3b) (Fig. 1). After numerically solving the PDEs, we fit the number of newly confirmed cases in each of the 47 prefectures of Japan to the solutions and estimated the effective distance from Tokyo (Fig. 2, Supplementary Figure 2). The calculated results at the estimated effective distance, particularly during the first part of each period (i.e., P1a, P2a, and P3a), recapitulated the temporal evolution of the number of newly confirmed cases in most prefectures (Fig. 2). Consequently, by considering the propagation of the pandemic as a diffusion process, we estimated the effective distance from Tokyo based on only the time series of newly confirmed cases of COVID-19 in each prefecture.

The effective distance estimated based on the diffusion process from Tokyo revealed the mobility of individuals during the pandemic.

To characterize the estimated effective distance based on the diffusion process from Tokyo, we applied linear regression in which the geographical distances from Tokyo or effective distances based on inter-regional passenger traffic are explanatory variables [31]. As for the effective distance derived from passenger traffic data, we used inter-prefecture passenger volume data obtained in Japan in 2019, built a weighted directed network of mobility probability between prefectures, and estimated it as the shortest path from Tokyo in the graph. The effective distance based on passenger traffic network generally predicted the effective distance based on the diffusion process, whereas the geographical distance did not, indicating that human mobility from Tokyo to other prefectures largely explained the time series of infections (Fig. 3a, Supplementary Figure 3a-c). Notably, in the Kanto region of Japan, which consists of seven prefectures centered around Tokyo, both the distance metrics predicted the effective distance derived from the diffusion equations with high accuracy, implying that, around Tokyo, the dynamics of infections could be largely explained by inter-prefecture travel between Tokyo and other prefectures, which correlated with the geographical distance on a local scale (Fig. 3b, Supplementary Figure 3a). Conversely, in prefectures far away from Tokyo, there would be some factors other than the diffusion of infected individuals from Tokyo that contributed to dynamics of infections. To examine the factors that contributed to the deviations of the effective distance based on the diffusion process from the predicted values in Fig. 3a, we investigated residuals from the predicted values. The residuals of some regions exhibited unequal distributions around the baseline Supplementary Figure 4a,b. For example, residuals of the Kinki regions and the Tohoku regions mostly distributed above and below the baseline, respectively Supplementary Figure 4a,b. These results imply that the deviations from the predicted values attributed to the local, or regional factors other than travels from Tokyo, such as within-prefecture spread or containment of the coronavirus cases or travels from nearby prefectures. Therefore, our findings revealed that the effective distance based on the changes in evolution of infection over time in each prefecture putatively reflected both human mobility and other local effects from nearby prefectures during the pandemic.

Effective distance changes dynamically depended on stages of the pandemic.

Another advantage of the effective distance derived from time series data of infections over that derived from passenger volume data is that human mobility can be estimated dynamically without the need for extensive surveys. Such dynamic estimations based on different stages of the pandemic are particularly important to evaluate effects of specific policies (e.g., “soft lockdown” and travel campaigns) and to define a better strategy. Thus, to investigate changes in human mobility over different stages of the pandemic, we first examined the distribution of the effective distance over the respective periods. Histograms of effective distances revealed that the shapes of the distributions shifted depending on different periods, and particularly during spreading periods.
other plots (Fig. 4a, Supplementary Figure 5a,b), it can be implied that closer effective distances, in particular spreading of the virus and altering their distance from Tokyo. Furthermore, considering the combination with than diffusion from Tokyo, there was a local connectivity between nearby prefectures that contributed to the entire country was conserved over time (Fig. 4b and Supplementary Figure 6b), which indicated that, other infection-related human mobility could be captured during a pandemic. Subsequently, we investigated the changes in the effective distance for each prefecture by region. The general trend of the effective distance across the entire country was conserved over time (Fig. 4b and Supplementary Figure 6b), which indicated that, other than diffusion from Tokyo, there was a local connectivity between nearby prefectures that contributed to the spreading of the virus and altering their distance from Tokyo. Furthermore, considering the combination with other plots (Fig. 4a, Supplementary Figure 5a,b), it can be implied that closer effective distances, in particular like those found in major cities in each region, chiefly contributes to the rapid increase in coronavirus cases against which the government needed to impose restrictions. Consequently, the effective distance estimated from the time series of the newly confirmed cases evolved temporally, which might provide useful information for determining the appropriate mobility level during the pandemic.

A distorted map of Japan, based on effective distances and local interactions, revealed the non-uniform spread of the pandemic across Japan.

To visualize how the pandemic spread across Japan across different periods, we combined information on effective distances and local interactions between prefectures, which we defined as the geographical distance between adjacent prefectures that had large passenger traffic in the 2019 survey, to create a distorted map of Japan. The position (i.e., longitude and latitude) of a capital city in each prefecture was determined as the solution of a mathematical optimization equation using the non-linear least squares calculation [38]. In general, regions to the west of Tokyo, largely the Kinki and Kyushu regions, reduced in size, whereas prefectures around Tokyo were expanded (for example, the Chubu regions were stretched in the north-west direction) in the distorted maps compared to the original one (Fig. 5a-d). These deformations indicated that the coronavirus tended to rapidly spread to the western regions of Japan, while it required additional time towards spread to the north and north-western directions, particularly towards the Tohoku and Chubu regions. Thus, distorting maps based on effective distance and local interactions would be useful to characterize the spread of the coronavirus geographically and to develop specific policies aimed at preventing a large spike in the infectious population.

Quantitative estimation of the scale of the pandemic in each prefecture in different sets of effective distances

Although effective distance allows us to estimate how the virus spreads across Japan, it is still unclear how much the effective distance contributes to determine the scale of the pandemic quantitatively, which would be a useful parameter when developing policies that balance both economic activity and medical resources in each prefecture. To evaluate the effects of the effective distance, we conducted simulations in which only the effective distance was changed from P3. If the effective distance was the same as P1, the total number of infected people would increase to more than twice in peripheral or rural areas, but would not increase to a similar degree in populated prefectures (Fig. 6a). Conversely, if the effective distance was the same as P2, there would be no clear differences between the total number of infected individuals between the central or peripheral prefectures in Japan (Fig. 6a). These results indicated that the effective distance could largely affect the scale of the pandemic, and these effects were heterogeneous between prefectures. Finally, we considered the increase in human mobility that could result from a very large public event such as the 2020 Olympic and the Paralympic Games, which we were to be held in Tokyo and several other prefectures, and estimated the impact on the pandemic (Supplementary Figure 9). Our findings showed that prefectures that held several matches, particularly prefectures that usually did not share strong connectivity with Tokyo, would experience a large increase in the number of patients (Fig. 6b). For example, the number of infected cases in Fukushima prefecture would increase by 80%, while other host prefectures would increase by 40 to 60%. This indicated that the impact of an increase in human mobility due to huge public events were more severe in prefectures that usually had a larger effective distance from Tokyo. Based on these quantitative estimations, we recommend that policymakers should correctly control for the effective distance or distribute medical resources across the whole country. Therefore, effective distance enables us to estimate the scale of the pandemic in a single prefecture.
based on changes human mobility and would be helpful in devising countermeasures against COVID-19 and other infectious diseases over changing situations.

Discussion

In this study, we calculated the effective distance of the spread of COVID-19 in Japan by applying the SEIR model to Tokyo, the capital city of Japan. Despite the presence of social factors that altered human mobility, such as two declarations of the state of emergency and travel campaigns, the spread of COVID-19 was demonstrated mainly by assuming the spread from Tokyo during all phases of the pandemic (Supplementary Figure 1). More specifically, our SEIR model-based predictions revealed that the effective distance from Tokyo to other prefectures changed over each phase, in which the effective distance between Tokyo and Hokkaido was close for P1 and P3, but far for P2, for example (Supplementary Figure 1). Comparing the fitting parameter and effective distance in P1a, P2a and P3a, we can see the effect of government-sponsored travel promotion. In particular, the effective distance between Tokyo and prefectures with many tourist attractions has increased. These findings provided an explanation for the spread of the two outbreaks in Hokkaido and many other prefectures, which could not be explained merely by passenger traffic data alone.

In Fig. 6, we estimated the impact of an huge public event such as the Tokyo Olympics by creating an infection network for the spread of infection, constructing all infection networks based on the SEIR model, and simulating the spread of infection when an outbreak occurs in a city is computationally infeasible. However, once the network structure is drawn, a feasible simulation becomes available. For example, in the case of this study, by assuming the spread of infection from Tokyo, we were able to explain the infection situation in other prefectures with high accuracy (Supplementary Figure 3), suggesting that the infection network in Japan was a simple radial structure centered on Tokyo. Conversely, if the same method is applied to the United States or European countries, where multiple cities or countries, respectively, experienced surges of infection simultaneously [39, 40], it may be possible to obtain a more complex structural organization of the infection network.

By combining the effective distance estimated from infection data with the effective distance based on a transportation network and movement data, we believe that it is possible to establish more effective preventive policies to avoid the spread of infection. It is difficult to quantitatively evaluate the contribution of various modes of transportation, such as airplanes, cars, and trains, to infectious diseases. By comparing the two effective distances, it is possible to determine the means of transportation that contribute to the transmission of the disease and to take selective and effective preventive measures. However, the proposed method has some limitations. Importantly, the assumption of the diffusion process from Tokyo is qualitatively different between periods before and after peaks of infection and does not allow to quantitatively compare effective distances between these periods. Thus, improvements in methodology to the estimate effective distance based on the uniform assumption remain to be addressed in future studies.

Methods

Dataset.

We downloaded the data of the daily number of individuals tested positive in 47 prefectures from the COVID-19 Situation Report in Japan by the Toyo Keizai Online COVID-19 Task Team [41]. The data were originally provided by the Ministry of Health, Labor, and Welfare. Returnees on government charter flights from Wuhan, airport quarantine and passengers, and cases on the Diamond Princess cruise ship were excluded. Due to the data availability from all of the prefectures, we took into consideration the data from 18 March 2020 to 13 March 2021, and divided it into three periods based on the three waves of the pandemic: (P1) from 18 March 2020 to 16 May 2020, (P2) from 17 May 2020 to 3 October 2020, and (P3) from 4 October 2020 to 12 March 2021. Each of these periods was further subdivided into two periods (a and b) depending on the time of the peak of infection. The timeline of the study and the major announcements from the Japanese government are illustrated in Fig. 1.
The survey data on passenger traffic between prefectures in 2019 were from the Ministry of Land, Infrastructure, Transport, and Tourism. The data included the passenger volumes on railways, cars, ships, and airlines. The population data for each prefecture, the distance between prefectures, which is defined by the distance between prefectural capitals, and the GeoPackage of for Japan were downloaded from the Statistics Bureau of Japan, the Geospatial Information Authority in Japan, and Database of Global Administrative Areas (GADM), respectively.


To describe the course of the COVID-19 pandemic in Tokyo, which is the capital and most populated city and likely the epicenter of the pandemic in Japan, we constructed a basic SEIR model, that included the following compartments: Susceptible (S), Exposed (E), Infectious (I), and Recovered (R). The choice of our model was motivated by several previously published studies on the COVID-19 pandemic [30, 35, 37].

The local dynamics of transmission in the model was given by

\[
\dot{S} = -\beta IS \\
\dot{E} = \beta IS - \epsilon E \\
\dot{I} = \epsilon E - \rho I \\
\dot{R} = \rho I.
\]

In the model, the susceptible population (S) becomes exposed to the viral agent upon contact with the infectious population (I) and thus, becomes the exposed population (E). The transmission rate of the virus is expressed as \(\beta\) and \(\beta I\) represents the force of infection. The exposed population becomes an infectious population at the rate \(\epsilon\) (i.e., calculated as the inverse of the latent period), and the infectious population becomes the recovered population (R) at the rate \(\rho\), the or recovery rate.

Parameter estimation using the MCMC algorithm.

Model parameters were estimated using the Bayesian framework by sampling the posterior parameter distribution via an affine-invariant MCMC algorithm using the emcee v3 toolkit [42, 43]. For data reliability, the data used to calibrate the model consisted of the 7-day backward moving average of newly confirmed cases in Tokyo. Considering the rapid increase in the spread of the coronavirus pandemic in Tokyo, we assumed that the count data of newly confirmed cases followed a Gaussian distribution with a mean given by \(\xi I\) and variance given by \(1/\tau\). \(E_0\) and \(I_0\), the number of exposed and infectious individuals on the first day during the period, respectively, were also estimated. For simplicity, the number of recovered individuals, or \(R_0\), was set at zero; thus, the number of susceptible individuals, or \(S_0\), was calculated as \(S_0 = N_t - E_0 - I_0\) where \(N_t\) represented the total population of Tokyo. Consequently, we estimated seven parameters for each period: \(\beta, \epsilon, \rho, \xi, E_0, I_0,\) and \(\tau\). We used the same uniform distributions for all periods as the prior distributions for each parameter, as described in Supplementary Table 1. The statistics of the posterior distributions are also listed in Supplementary Table 1. The convergence of the MCMC chains are judged based on the integrated autocorrelation time in accordance with a previous study [42]. When the number of samples satisfied the inequality given by,

\[ M \geq C\hat{\tau}(M), \]

where \(\hat{\tau}(M)\) was the estimated integrated autocorrelation time, the relative error in the estimation of the posterior distribution was considered sufficiently small.

Model of the spread of the infected population from Tokyo and estimation of effective distance.

The spread of the COVID-19 pandemic in prefectures other than Tokyo was modeled as the diffusion process of infected individuals from Tokyo. We denoted the number of newly confirmed cases as \(u(x,t)\), as follows:

\[
\frac{\partial u}{\partial t} = D \frac{\partial^2 u}{\partial x^2}, \ x \in (0, L), \ t \in (0, T),
\]

where \(D\) was the diffusion coefficient. The parameters used in the simulation are summarized in [TABLE]. For simplicity, the initial conditions for the PDE were expressed as:
\[ u(0, 0) = \xi I_0 \]
\[ u(x, 0) = 0, \ x \in (0, L], \]
which meant that the infectious population was located only in Tokyo in the initial condition. For the boundary conditions, we applied the Dirichlet boundary condition for \( x = 0 \) and the Neumann boundary condition for \( x = L \), or
\[ u(0, t) = \xi I \]
\[ \frac{\partial u}{\partial x} \bigg|_{x=L} = 0. \]

The parameters for the PDE are summarized in Supplementary Table 2. The PDE was replaced by a series of ordinary differential equations (ODEs) by finite difference approximations, which were solved using the SciPy software. The solutions of the PDE, or the diffusion process of the infectious population from Tokyo depending on \( x \) and \( t \) values, are shown in Fig. 2a,b.

To estimate the effective distance from Tokyo at different stages of the pandemic, we considered the diffusion process for six different periods depending on the period P1 to P3 and the peaks of the number of new cases in Tokyo during the periods: P1a, P1b, P2a, P2b, P3a, and P3b. The fitting of the observed data to the simulated data was also conducted for each period separately, and the \( x \), or the effective distance of each prefecture from Tokyo based on the diffusion process, was determined by choosing \( x \) with a minimum root mean square error (RMSE) value. The estimated effective distance for all periods are summarized in Suplementary Table 3. If the effective distance of Tokyo, \( x_T \), estimated from fitting was not equal to zero, the effective distance of all prefectures was adjusted by subtracting \( x_T \) for subsequent analyses.

**Effective distance based on the inter-prefecture network of traffic and mobility data of individuals.**

Since the contagion process is predominantly caused by the movement of people, we can also estimate the effective distance between prefectures based on passenger traffic data. The effective distance derived by this method was used to test the validity of the effective distance derived from the diffusion equation and for its characterization (Fig. 3a,b, Supplementary Figure 3a). We defined the connectivity matrix \( \mathbf{P} \) as follows:

\[ P_{nm} = \frac{F_{nm}}{F_m} \quad (0 \leq P_{nm} \leq 1) \]
\[ F_m = \sum_n F_{nm}, \]
where \( F_{nm} \) represented the inter-prefecture passenger traffic on railways, cars, ships and airlines. Thus, the matrix \( \mathbf{P} \) quantified the fraction of the passenger flux from prefecture \( n \) to \( m \). Given this flux-fraction matrix \( \mathbf{P} \), the effective distance \( d_{nm} \) from prefecture \( n \) to \( m \) is defined as

\[ d_{nm} = (1 - \log P_{nm}) \geq 1. \]

We then constructed a weighted directed graph network consisting of 47 nodes, representing the prefectures in Japan, and edges from prefecture \( n \) to \( m \) with weights of \( d_{nm} \) (i.e., \( 47C_2 = 1081 \) edges in total). Consequently, the effective distance from Tokyo to prefecture \( n \), or \( D_n \), was estimated as the length of the shortest path from Tokyo to the prefecture \( n \) in the graph network.

**Linear regression of the effective distance based on the diffusion process by other distance metrics.**

The effective distance derived from the diffusion equations was predicted by a linear model without intercept, or \( y = ax \), as Tokyo is at the origin. Models were fit to minimize the residual sum of squares. Coefficient of determination (\( R^2 \)) was determined as \( R^2 = 1 - \frac{u}{v} \), where \( u \) is the residual sum of squares and \( v \) is the total sum of squares.
Transformation of a map of Japan based on the effective distance from Tokyo and local connectivity.

We created a transformed map of Japan according to the Mercator projection (i.e., the longitude and latitude corresponding to the $x$ and $y$ coordinates in the Euclidean plane, respectively) depending on the effective distance from Tokyo, which was estimated from the diffusion equation and local interactions between prefectures (Supplementary Figure 7). Local interactions between prefectures were defined as the geographical distance between adjacent prefectures, including those connected by roads. Furthermore, for simplicity, the local connectivity was only considered for adjacent prefectures that had substantial passenger traffic, or 100,000 people per year in at least one direction. The unit of effective distance was converted to that of geographical distance assuming a linear relationship between them locally, as shown in Supplementary Figure 3c,d. In total, 111 links between prefectures (i.e., 46 links that expressed effective distance from Tokyo to other prefectures and 65 links of local interactions) were used to transform a map of Japan to represent the spread of the pandemic.

Distorting a map based on a given proximity can be formulated as a nonlinear least squares problem as follows:

$$\min \sum_{ij \in L} \left( t_{ij} - \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} \right)^2,$$

where $ij$ denoted the links between points $i$ and $j$, and $x$ and $y$ represented the longitude and latitude of a prefecture, respectively. $L$ denoted the set of 111 links between prefectures, and $t_{ij}$ denoted the proximity (i.e., effective distance or local geographical distance). To solve this problem, we adopted the algorithm suggested by Shimizu and Inoue in 2009 [38]. In short, this algorithm considers the norm minimization of bearing changes instead of the norm minimization of variation of coordinates, which is solved by the Levenberg-Marquardt method, which uses bearings between points as the initial condition, and is a natural premise when transforming a map. Next, the nonlinear least squares problem was rewritten as

$$\min \sum_{ij \in L} \left[ \{t_{ij} - (x_j - x_i) \sin \theta'_{ij} - (y_j - y_i) \cos \theta'_{ij}\}^2 
+ \alpha \{(x_j - x_i) \cos \theta'_{ij} - (y_j - y_i) \sin \theta'_{ij}\}^2 \right],$$

where $\theta'_{ij}$ was the approximate bearing of link $ij$ in the transformed map.

When $\alpha = 1$, we obtained:

$$\min \sum_{ij \in L} \left[ \{t_{ij} \sin \theta'_{ij} - (x_j - x_i)\}^2 + \{t_{ij} \cos \theta'_{ij} - (y_j - y_i)\}^2 \right].$$

This meant that we could approximate the solution for the initial nonlinear least squares problem by an iteration of solving two independent linear squares problems. When solving these linear squares problems, we assigned different weights to equations derived from the effective distance and those from local interactions (10:1). The weight was determined so that the distorted maps represented effective distance from Tokyo as well as geographical relationships between nearby prefectures (Supplementary Figure 8a,b). These weights enabled the effective distance from Tokyo to be expressed while retaining geographical information. When $|\theta_{ij} - \theta'_{ij}| < 0.01$, the solution was regarded as reaching convergence.

Simulations to estimate effects of the effective distance on scales of the pandemic.

Period P3 was selected as the base period because we considered that the period reflected most recent human mobility, thus was suitable for estimating an impact of Tokyo 2020 Olympic and Paralympic on the pandemic. The rate of change was calculated as the ratio of estimated total number of infections in simulated distances to the total number of infections in the original distances (i.e., the effective distance during P3). The prefectures in which 2020 Olympic and Paralympic Games were held (i.e., Tokyo, Kanagawa, Saitama, Chiba, Ibaraki, Shizuoka, Fukushima, Miyagi, and Hokkaido) were set three times closer than other prefectures, which modeled increased human mobility due to the games (Supplementary Figure 9).

Data availability

Raw data used in this manuscript are available at https://github.com/tyamadat/seir_pde.
Code availability

All source codes used in this manuscript are open and available at https://github.com/tyamadat/seir_pde.

References


Acknowledgements

We would like to thank M. Katori for helpful discussions. This work was supported by JSPS KAKENHI Grant – Grant-in-Aid for Early-Career Scientists (19K16487 and 21K15136, S.S.), Transformative Research Areas A (20H05894 and 20H05903, S.S.), and Moonshot R and D – MILLENNIA Program Grant (JPMJMS2023-25, S.S.).

Author contributions

T.Y. and S.S. designed the study, conducted the simulations, analyzed the data, and wrote the manuscript.
Fig. 1 | Timeline of the pandemic of coronavirus disease 2019 (COVID-19) in Japan.

The period from 18 March 2020 to 12 March 2021 was considered. The top row indicates the period during which a state of emergency (i.e., "soft lockdown", or mild restrictions on movement, alcohol-serving businesses, and large events) was declared. The second row indicates the period during which travel campaign, or "Go To Travel", during which Japanese government encouraged the citizens to travel domestically by subsidizing travel expenses. The travel campaign had two stages: (1) the first from 22 July 2020 to 30 September 2020, during which Tokyo was excluded from the campaign, and (2) the second from 1 October 2020 to 27 December 2020, during which Tokyo was included and the subsidy was increased from stage 1. The main panel exhibits the 7-day backward moving average of the number of newly confirmed cases in each prefecture, which are highlighted in blue for Tokyo. Depending on the phase of the pandemic, the period was divided into six different periods.
**Fig. 2** | **Effective distance from Tokyo is estimated from diffusion process.**

*a,b.* Solutions of diffusion equations for each period (a: before peaks of infection, b: after peaks of infection). Solutions for \( x \in \{2n \mid 0 \leq n \leq 50, \ n \in \mathbb{Z}\} \) over time are shown. The source of the diffusion process \((x = 0)\) is the solution of the SEIR model, whose parameters were calibrated based on the MCMC algorithm ([Supplementary Figure 1a-c](#)).
Fig. 3 | Effective distance derived from diffusion equations contains both passenger traffic and other local information that contribute to spread of the pandemic.

Relationship between effective distance based on diffusion process from Tokyo and passenger traffic data in 2019 for all prefectures (a) and the prefectures in Kanto region (b). A point represents each prefecture, and the prefectures whose effective distance is larger than 100 were excluded. Linear regression by least mean square and correlation coefficients are shown for each panel.
Effective distance changes dynamically depending on stages of the pandemic.

Fig. 4 | Effective distance changes dynamically depending on stages of the pandemic.

a. Distribution of effective distances in periods P1a, P2a, and P3a. The vertical line in magenta indicates the mean value of the distribution.

b. Effective distance of each prefecture in periods P1a, P2a, and P3a, ordered by "Prefecture Code" defined International Organization for Standardization. Regions in which prefectures belong to are colored differently.
Distorted maps of Japan based on effective distance and local interactions reveal non-uniform spreading of the virus across Japan.

- **Fig. 5** Distorted maps of Japan based on effective distance and local interactions reveal non-uniform spreading of the virus across Japan.
  - **a** Original map of Japan. Each point indicates the capital of each prefecture, colored by region.
  - **b-c** Distorted maps of Japan based on effective distance and local connectivity in periods P1a, P2a, and P3a (Supplementary Figure 7).
Simulations quantitatively estimates the effects of effective distance on a scale of the pandemic in each prefecture. 

\(\textbf{a}\), Estimated increase or decrease of total infection during an equivalent period as P3 when the effective distance from Tokyo is set to the same as P1 (left) or P2 (right) and source (Tokyo) is the same as P3.

\(\textbf{b}\), Estimated increase in infections of infection during an equivalent period as P3 when every prefecture approaches Tokyo because of a large-scale event such as the Olympic. Some prefectures where some games are held are closer to Tokyo than others. The results of the simulated shifts are shown in Supplementary Figure 9.
Supplementary Figure 1 | SEIR model recapitulates time evolution of the number of newly confirmed cases in Tokyo for each period.

a-c. Time evolution of infection derived from the SEIR model whose parameters are fit by the MCMC algorithm.
Supplementary Figure 2 | Diffusion process from Tokyo recapitulates time series data of infections in other prefectures during period P1a.

Cumulative number of newly confirmed cases that were observed (blue) and estimated by the diffusion process from Tokyo (orange) in each prefecture during P1a.
Supplementary Figure 3 | Relationship bewteen effective distance based on diffusion and other distances.

a. Relationships between effective distance based on diffusion process from Tokyo and that on passenger traffic data in 2019 for the whole country (blue + orange) and Kanto region (orange).

b,c. Relationships between effective distance based on diffusion process from Tokyo and Geographical distance from Tokyo for the all prefectures and (a) and the prefectures in Kanto region (b). A point indicates each prefecture. Linear regression by least mean square and correlation coefficient are shown for each panel.
Supplementary Figure 4 | Residuals of the effective distance based on the diffusion process.

Residuals of the effective distance derived from the diffusion process compared to the predicted values based on linear regressions based on effective distance derived from the passenger traffic data (Fig. 3a, Supplementary Figure 5a).
Supplementary Figure 5 | Distribution of effective distance in different periods.

a-b. Effective distance in each prefecture for each period, sorted in ascending order.
Supplementary Figure 6 | Effective distance changes dynamically depending on stages of the pandemic.

a. Distributions of effective distance in periods P1b, P2b, and P3b. The vertical line in magenta indicates the mean value of the distribution. Prefectures whose effective distance was larger than 100 were excluded. b. Effective distance of each prefecture in P1b, P2b, and P3b, ordered by "Prefecture Code" defined International Organization for Standardization. Regions in which prefectures belong to are colored differently.
Supplementary Figure 7 | Local interactions with large passenger volumes.

Links of adjacent prefectures in which more than 100,000 individuals traveled across the border in at least one direction in 2019 are shown.
Distorted maps of Japan based on effective distance in different weight ratio between effective distance and local interactions.

A map of Japan was distorted to represent the effective distance in each period. Distorted maps were determined as solutions for weighted non-linear least squares problems. Various weight ratio between the effective distance and the geographical distances between nearby prefectures which had high passenger volumes was tested (a, b, Fig. 5).
Supplementary Figure 9 | Putative map of Japan during the Tokyo 2020 Olympic used for the simulation in Fig. 6b.

Distorted map based on effective distance during period P3a (left) and modified period P3a (P3’a: Olympic) (right). In period P3’a, the prefecture in which some games will be held, highlighted in the left panel, approached Tokyo more than others.
Supplementary Table 1 | Parameters in the SEIR model and the estimated values.

Seven parameters were estimated using the MCMC algorithm so that the SEIR model represented the time series data of infections in each period. The median of the sampler of each parameter was chosen as a representative parameter.

Supplementary Table 2 | Parameters in the diffusion equations.

The parameters for the diffusion equations were chosen so that the solutions represented the time course of the number of newly confirmed cases in each prefecture.

Supplementary Table 3 | Effective distance in each prefecture in the respective periods.

The effective distance of each prefecture from Tokyo was estimated in each period based on the diffusion process.
<table>
<thead>
<tr>
<th>Epidemiological meaning</th>
<th>Median (P1)</th>
<th>Median (P2)</th>
<th>Median (P3)</th>
<th>Prior</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$ Transmission rate</td>
<td>$3.7e-7$</td>
<td>$1.9e-7$</td>
<td>$4.6e-7$</td>
<td>$U(1e-9, 1e-6)$</td>
</tr>
<tr>
<td>$\epsilon$ $1/\epsilon$ is the latent period of the infection</td>
<td>$9.1e-1$</td>
<td>$1.8e-2$</td>
<td>$4.1e-1$</td>
<td>$U(1e-2, 1)$</td>
</tr>
<tr>
<td>$\rho$ Recovery rate</td>
<td>$4.1$</td>
<td>$2.4e-1$</td>
<td>$5.0$</td>
<td>$U(5e-2, 5)$</td>
</tr>
<tr>
<td>$\xi$ Coefficient for $I$ to express the number of newly confirmed cases</td>
<td>$5.8e-3$</td>
<td>$5.1e-4$</td>
<td>$1.0e-1$</td>
<td>$U(1e-5, 1e-1)$</td>
</tr>
<tr>
<td>$E_0$ Size of $E$ in an initial condition</td>
<td>$9.1e3$</td>
<td>$4.7e3$</td>
<td>$1.0e2$</td>
<td>$U(1e2, 1e4)$</td>
</tr>
<tr>
<td>$I_0$ Size of $I$ in an initial condition</td>
<td>$3.2e3$</td>
<td>$3.8e3$</td>
<td>$3.5e2$</td>
<td>$U(1e2, 1e4)$</td>
</tr>
<tr>
<td>$\tau$ $1/\tau$ is the variance of data noise</td>
<td>$9.6e1$</td>
<td>$1.0e2$</td>
<td>$4.1e-4$</td>
<td>$U(1e-8, 1e2)$</td>
</tr>
<tr>
<td>Meaning</td>
<td>Value</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------</td>
<td>-------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n$</td>
<td>1000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of points in space</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$D$</td>
<td>1000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diffusion coefficient</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$L$</td>
<td>10000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Upper limit of space</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>P1a</td>
<td>P2a</td>
<td>P3a</td>
<td>P1b</td>
</tr>
<tr>
<td>---</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>1. Hokkaido</td>
<td>26</td>
<td>32</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>2. Aomori</td>
<td>41</td>
<td>83</td>
<td>56</td>
<td>41</td>
</tr>
<tr>
<td>3. Iwate</td>
<td>–</td>
<td>100</td>
<td>60</td>
<td>–</td>
</tr>
<tr>
<td>4. Miyagi</td>
<td>34</td>
<td>56</td>
<td>39</td>
<td>32</td>
</tr>
<tr>
<td>5. Akita</td>
<td>45</td>
<td>95</td>
<td>78</td>
<td>50</td>
</tr>
<tr>
<td>6. Yamagata</td>
<td>38</td>
<td>78</td>
<td>62</td>
<td>32</td>
</tr>
<tr>
<td>7. Fukushima</td>
<td>37</td>
<td>80</td>
<td>51</td>
<td>30</td>
</tr>
<tr>
<td>8. Ibaraki</td>
<td>28</td>
<td>51</td>
<td>38</td>
<td>28</td>
</tr>
<tr>
<td>9. Tochigi</td>
<td>38</td>
<td>49</td>
<td>44</td>
<td>35</td>
</tr>
<tr>
<td>10. Gunma</td>
<td>32</td>
<td>61</td>
<td>40</td>
<td>29</td>
</tr>
<tr>
<td>11. Saitama</td>
<td>17</td>
<td>22</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td>12. Chiba</td>
<td>15</td>
<td>29</td>
<td>18</td>
<td>16</td>
</tr>
<tr>
<td>13. Tokyo</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>14. Kanagawa</td>
<td>15</td>
<td>22</td>
<td>8</td>
<td>13</td>
</tr>
<tr>
<td>15. Niigata</td>
<td>38</td>
<td>70</td>
<td>59</td>
<td>31</td>
</tr>
<tr>
<td>16. Toyama</td>
<td>37</td>
<td>70</td>
<td>70</td>
<td>21</td>
</tr>
<tr>
<td>17. Ishikawa</td>
<td>29</td>
<td>62</td>
<td>62</td>
<td>22</td>
</tr>
<tr>
<td>18. Fukui</td>
<td>29</td>
<td>75</td>
<td>74</td>
<td>32</td>
</tr>
<tr>
<td>19. Yamanashi</td>
<td>38</td>
<td>62</td>
<td>59</td>
<td>34</td>
</tr>
<tr>
<td>20. Nagano</td>
<td>39</td>
<td>67</td>
<td>47</td>
<td>30</td>
</tr>
<tr>
<td>21. Gifu</td>
<td>28</td>
<td>50</td>
<td>39</td>
<td>31</td>
</tr>
<tr>
<td>22. Shizuoka</td>
<td>38</td>
<td>50</td>
<td>35</td>
<td>32</td>
</tr>
<tr>
<td>23. Aichi</td>
<td>22</td>
<td>28</td>
<td>13</td>
<td>22</td>
</tr>
<tr>
<td>24. Mie</td>
<td>46</td>
<td>64</td>
<td>49</td>
<td>33</td>
</tr>
<tr>
<td>25. Shiga</td>
<td>37</td>
<td>57</td>
<td>51</td>
<td>29</td>
</tr>
<tr>
<td>26. Kyoto</td>
<td>23</td>
<td>38</td>
<td>32</td>
<td>22</td>
</tr>
<tr>
<td>27. Osaka</td>
<td>11</td>
<td>19</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>28. Hyogo</td>
<td>19</td>
<td>37</td>
<td>19</td>
<td>18</td>
</tr>
<tr>
<td>29. Nara</td>
<td>37</td>
<td>48</td>
<td>41</td>
<td>30</td>
</tr>
<tr>
<td>30. Wakayama</td>
<td>41</td>
<td>55</td>
<td>58</td>
<td>32</td>
</tr>
<tr>
<td>31. Tottori</td>
<td>65</td>
<td>81</td>
<td>78</td>
<td>46</td>
</tr>
<tr>
<td>32. Shimane</td>
<td>51</td>
<td>85</td>
<td>81</td>
<td>36</td>
</tr>
<tr>
<td>33. Okayama</td>
<td>43</td>
<td>61</td>
<td>43</td>
<td>39</td>
</tr>
<tr>
<td>34. Hiroshima</td>
<td>35</td>
<td>50</td>
<td>34</td>
<td>25</td>
</tr>
<tr>
<td>35. Yamaguchi</td>
<td>42</td>
<td>72</td>
<td>59</td>
<td>37</td>
</tr>
<tr>
<td>36. Tokushima</td>
<td>58</td>
<td>73</td>
<td>82</td>
<td>46</td>
</tr>
<tr>
<td>37. Kagawa</td>
<td>47</td>
<td>69</td>
<td>66</td>
<td>37</td>
</tr>
<tr>
<td>38. Ehime</td>
<td>38</td>
<td>61</td>
<td>59</td>
<td>36</td>
</tr>
<tr>
<td>39. Kochi</td>
<td>35</td>
<td>82</td>
<td>56</td>
<td>36</td>
</tr>
<tr>
<td>40. Fukuoka</td>
<td>19</td>
<td>27</td>
<td>28</td>
<td>18</td>
</tr>
<tr>
<td>41. Saga</td>
<td>45</td>
<td>67</td>
<td>65</td>
<td>32</td>
</tr>
<tr>
<td>42. Nagasaki</td>
<td>44</td>
<td>60</td>
<td>58</td>
<td>45</td>
</tr>
<tr>
<td>43. Kumamoto</td>
<td>41</td>
<td>56</td>
<td>43</td>
<td>34</td>
</tr>
<tr>
<td>44. Oita</td>
<td>34</td>
<td>91</td>
<td>54</td>
<td>36</td>
</tr>
<tr>
<td>45. Miyazaki</td>
<td>44</td>
<td>58</td>
<td>55</td>
<td>-</td>
</tr>
<tr>
<td>46. Kagoshima</td>
<td>53</td>
<td>41</td>
<td>53</td>
<td>40</td>
</tr>
<tr>
<td>47. Okinawa</td>
<td>33</td>
<td>50</td>
<td>31</td>
<td>27</td>
</tr>
</tbody>
</table>