Analyzing Population Flows Using Mobile Spatial Statistics and Contributions to Society During the COVID-19 Pandemic

Kenji Shinoda
Group Leader
Area Marketing Division
DOCOMO InsightMarketing, INC

Hiroto Akatsuka
System Design & Development Group
X-Tech Development Department
NTT DOCOMO, INC.

Toshihiro Suzuki, Ph. D.
General Manager
Area Marketing Division
DOCOMO InsightMarketing, INC

Masayuki Terada
Senior Manager / Security Principal
X-Tech Development Department
NTT DOCOMO, INC.

1. Introduction

It has been approximately two years since the novel corona virus (COVID-19) began to spread, drastically changing the patterns of our lifestyles and activities. As variants more infectious than before raged around the world, the citizens of Japan, together with government and private enterprise mounted their response. Although the rates of serious cases are decreasing as vaccination rates increase, it is still not clear how infection rates could change in the future.

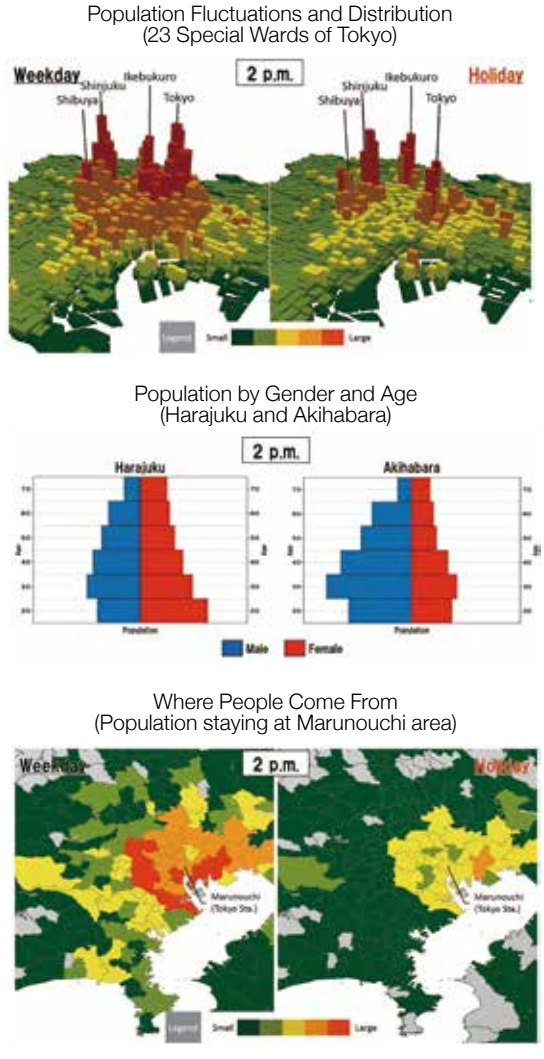
Understanding movement in the population and adopting measures such as the "Three Cs" to reduce spread were important social efforts to limit expansion of COVID-19. As one tool to address such issues, Big Data techniques utilizing location data from mobile devices have attracted much attention. Typical examples of this are the population maps utilizing Mobile Spatial Statistics (MSS). This service provides real-time knowledge of fluctuations in population throughout Japan, and by comparing with results from the previous year; it can provide an understanding of changes in population flows. As an initial use of this service, the Cabinet Secretariat and public organizations in Tokyo and other regions are using MSS to propose policies to counter the spread of infection and to measure the effects of such measures. It has also been used in research to elucidate the relationship between spread of infection and population flows, contributing to society daily.

Section 2 of this article gives an overview of MSS, and Section 3 shows the important role it plays in society, in measures taken to counter the spread of infection. Section 4 describes how it is being used actively to research population flows related to COVID-19, together with some concrete examples. Finally, Section 5 describes how MSS employs differential privacy to ensure robust privacy protection.

2. Overview of MSS

MSS provides statistical data of the population in all areas of Japan, hourly, 24-hours-a-day and 365-days-a-year, based on data from phones and other mobile devices operating on NTT DOCOMO's network (Figure 1). The statistics are computed

Figure 1: Overview of MSS



using a very large sample encompassing the approximately 82 million devices of Japan residents and 12 million belonging to foreigners visiting Japan, generating results that are highly-reliable statistically. For Japan residents, population estimates by attributes such as sex, age-bracket and area of residence can also be computed based on mobile phone subscriber information. Statistics are available geographically on a 500 m mesh covering the entire country*, in age brackets down to five-year increments, and by residential area down to villages. In each base-station area, the mobile devices being served are reported periodically, and based on aggregate numbers of mobile devices; estimates are made taking the market-share of NTT DOCOMO mobile phones into consideration.

One of the most important characteristics of MSS is that data is maintained over a long period of time. Data has been accumulated since services for residents of Japan began in October, 2013, so comparisons with past data can be done. A second characteristic is that data is very fresh. The real-time version of domestic population statistics can provide data as recent as one hour old, which is near-real-time, so it was used to understand population flows during recent measures taken for COVID-19, which changed greatly from day-to-day.

MSS are generated in three processing steps. The first step is "de-identification," which converts mobile phone data to a form from which individuals cannot be identified. The second step is "aggregation," which computes total numbers of phones according to various attributes and then increases the values according to the market-share of NTT DOCOMO to create results that also consider the population that are not NTT DOCOMO users, and the third step is "anonymization," which removes values for areas that have very few people. Thus, personal information and privacy are protected by using statistical processing so that individuals cannot be identified. A conference on "Development of Society and Industry using MSS," was held by experts and guidelines for public and industry use were published, and since the service began in 2013, it has been used in many fields, both government and private, while also ensuring protection of privacy.

For the real-time domestic population distribution statistics currently being used the most for COVID-19 related population flow analysis, "differential privacy" has been used, which is discussed in Section 5.

3. Infectious disease countermeasures using MSS

As mentioned in Section 1, the strong threat of COVID-19 resulted in a social demand to restrict the movement of people. Then, in May 2020, MSS began publishing data on a web site in the form of population maps. The population maps showed heat maps of population in real time on a 500 m mesh covering the entire country. These could be used for various purposes such as checking conditions in congested areas, requesting residents not to leave their areas, or verifying the effects of measures taken (Figure 2). MSS were also used on the Cabinet Secretariat's "Novel Corona Virus Infection Countermeasures" dedicated web site[1]. The site was established to understand population flows for avoiding the Three Cs, and to measure the effects of infection countermeasures. When it was first published, it provided a way to check the ratios of population increase or decrease relative to the population at a reference point in time, the day before or before the state of emergency was declared. It provided comparisons of populations at around 3 pm in representative locations in each prefecture, such as around the Sapporo train station, in Shinjuku, and the Shibuya Center shopping area. As of May 2022, it displays graphs of one-year population increases and decreases for every day, using the peak values for each time period in all of 2019 as the basis (1.0), and for major locations and entertainment areas in every prefecture (Figure 3).

MSS are provided to the national government as well as the Tokyo metropolitan government and other regional public organizations. For example, in Tokyo the state of congestion in major commercial areas such as Shibuya, Shinjuku and Ginza

Figure 2: Population map

Figure 3: Graph of population rates of fluctuation in major locations and entertainment districts, posted on the Cabinet Secretariat web page

* Some areas are covered with a 125 m mesh, and others are covered with a 250 m mesh.

was represented with pictograms and used to warn and encourage citizens to avoid the Three Cs[2].

Previously, MSS has been widely used for purposes such as disaster-prevention planning, community building efforts, and tourism promotion, but in this case it was used broadly for COVID-19 as an index of levels of congestion in public areas, to avoid the Three Cs, thereby contributing to society with measures to reduce movement of people and spread of infection.

4. Population-flow research using MSS

MSS are generated for events like COVID-19 that have had a major effect on society, and have been extremely useful for understanding macro-level changes in population flows. There has been much research using this data, and it has elucidated the relationships between COVID-19, related policies, and population flows, with resulting contributions to society.

Measures to suppress movement of the population, especially during the first declaration of a state of emergency in 2020, had a huge effect on people's movements, and MSS showed this clearly. ARIMURA et al. used MSS to analyze changes in population density in Sapporo after the first declaration of a state of emergency in Japan. Their results showed that during that period, the number of people staying home increased, those travelling to the center of town decreased, and this resulted in decreases of up to 90% in the population density in normally-congested areas[3].

Analysis using MSS has also been used to show behavioral characteristics of Japanese citizens. HARA et al. showed that during COVID-19, people were refrained from taking vacations or traveling between prefectures, even without strong government regulation, resulting in reduced population density in areas that are usually crowded. They also suggested that, based on this, the return of population flows was gradual after the state of emergency was rescinded [4]. KAJITANI et al. also conducted analysis in eight prefectures where infection rates increased, showing a correlation between rates of mobility and infection risk. They also combined this with business data, showing a correlation between potential for contact in each business type and the effective reproduction number [5].

There has also been research comparing the effects of the multiple declarations of a state of emergency. TSUBOI et al. analyzed population flows under two declarations of emergency, using MSS to show that suppression of population flow was smaller during the second declaration than it was during the first declaration[6].

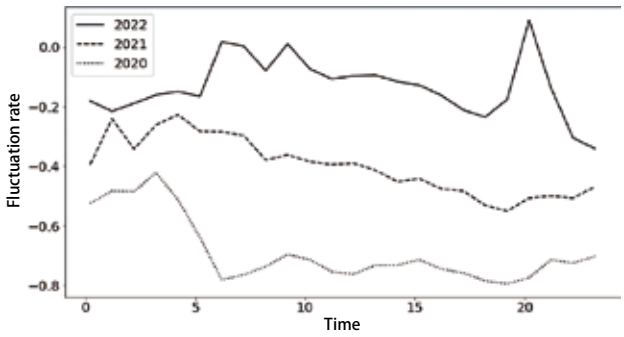
Note that statistical data such as MSS, based on base station positioning, is superior to point-based population-flow data such as GPS, for understanding macro-level behavior [7]. With GPS, the sample only includes users of the application, producing biases by region and other attributes, but MSS is able to obtain data from base stations deployed throughout Japan and makes estimates using NTT DOCOMO's large-scale sample data, so it can produce more-reliable population statistics data.

Finally, MSS was used to observe people's activities during Golden Week (GW) in 2022, compared with prior years. For the five-day period starting April 29, which is Showa Day, the population in the area surrounding Haneda Airport was compared for 2019 (before COVID-19) and the following years (during COVID-19). The conditions for the computations are shown in the table. The populations for each year were represented by peak population values in each hour of the day during the five days of each GW, and rates of increase or decrease were evaluated relative to the peak values at each time in 2019. Results of the computations are shown in Fig. 4. The solid line is for 2022, the dashed line is for 2021, and the dotted line is for 2020, showing rates of increase or decrease each year. 2020 was during the first declaration of a state of emergency, with daytime populations greatly reduced by up to almost 80%. 2021 was also under a state of emergency, but it was the third such declaration, and reductions were not as great as in 2020. Compared with before COVID-19, daytime population was down almost 40%, indicating that people were restraining themselves from taking vacations. Then, during GW in 2022, with widespread vaccinations and removal of restrictions on movement, daytime populations were only reduced by at most approximately 20%. For the time periods around 8 am

Table: Computation conditions

Item	Description
Reference year	2019
Comparison years	2020, 2021, 2022
Time period	Five days starting Apr. 29 in each year
Area covered	A 500 m mesh surrounding Haneda Airport Totals: 53392652, 53392653, 53392662, 53392663

Figure 4: Fluctuation rate trends by comparing peak values



and 8 pm, when many people go out or arrive home, the rates had almost completely recovered. Using MSS in this way, to show statistically how movement of people changes from one hour to the next, is an extremely powerful technique.

5. MSS and protecting privacy

NTT DOCOMO considers protection of user privacy to be of the utmost importance, and has published guidelines to be observed in the production and provision of MSS. MSS are strictly population statistical data and do not provide any information that can be related to individual people. For this reason, information that can identify individuals, such as phone numbers, are not used. Information is also summarized in creating statistics, such as by converting birthdays into age brackets, and addresses into cities and towns. However, summarizing information and creating statistics cannot guarantee that customer privacy is 100% protected. For example, if an attacker knows that there is only one 15-year-old female in a given region, there is potential to discover that female's activity from the real-time version of the domestic population distribution statistics. In this way, there is a danger that existing privacy protection techniques may not cover cases such as when an attacker has access to background information, or a new method of attack is discovered.

To address such issues, NTT DOCOMO has focused attention on a technique called differential privacy[8] which was proposed in 2006. Differential privacy is a technology that is being implemented by Google and Apple, and is also reported to have been used for the 2020 USA national census[9]. In simple terms, a small amount of noise is added to statistical data to satisfy differential privacy criteria. An important feature of differential privacy, in contrast to conventional privacy protection criteria such as the k-anonymity[10], is that the safety of privacy can be guaranteed mathematically. Safety from an attacker with any background information or a new method of attack is given mathematically, which may provide a way to avoid the cat-and-mouse game with respect to privacy protection.

NTT DOCOMO is continuing to study ways to apply differential privacy to large-scale geo-spatial data[11], and using them for privacy protection in the real-time version of domestic population-distribution statistics. This enables us to provide statistical information with levels of both safety and usefulness that are difficult to achieve with simple statistics or conventional privacy protection technologies.

6. Conclusion

This article has described examples of analyzing population flows and their contribution to society, particularly during the COVID-19 pandemic, using MSS, which estimates populations based on mobile phone and other mobile device data from base-stations. Although the situation is improving with increasing vaccination rates, the effects of COVID-19 on society in the future are still not clear, and statistical data on population remains an important tool for refining countermeasures. We hope that MSS will continue to play a role in society, increasing safety for residents of Japan and helping with development of society and industry.

References

[1] Cabinet Secretariat, COVID-19 Information and Resources, https://corona.go.jp/dashboard/
[2] Tokyo Office of the Governor for Policy Planning, https://www.seisakukikaku.metro.tokyo.lg.jp/cross-efforts/corona/people-flow-analysis.html
[3] Mikiharu Arimura, Tran Vinh HA, Kota Okumura, Takumi Asada: Changes in urban mobility in Sapporo city, Japan due to the Covid-19 emergency declarations, Transportation Research Interdisciplinary Perspectives, Volume 7, https://doi.org/10.1016/j.trip.2020.100212(2020).
[4] Yusuke Hara, Hiromichi Yamaguchi: Japanese travel behavior trends and change under COVID-19 state-of- emergency declaration: Nationwide observation by mobile phone location data, Transportation Research Interdisciplinary Perspectives, Volume 9, https:/ doi.org/10.1016/j.trip.2020.100288 (2021).
[5] Yoshio Kajitani, Michinori Hatayama: Explaining the Effective Reproduction Number on the COVID-19 by Mobility and Enterprise Statistics: Evidences during the First Wave in Japan and Implications to the Mobility Reduction Policy, PLoS ONE 16(3), https:/ doi.org/10.1371/journal.pone.0247186 (2020).
[6] Kazufumi Tsuboi, Naoya Fujiwara, Ryo Itoh: Influence of trip distance and population density on intra-city mobility patterns in Tokyo during COVID-19 pandemic, ArXiv, https://doi.org/10.48550/arXiv. 2201.01398 (2022)
[7] Kenta Kikuchi, Takuya Maruyama, "Analysis of conditions escaping disaster during COVID-19: Case of Reiwa 2 Heavy rain in Hitoyoshi City, Kumamoto Prefecture," JSCE Proceedings, Vol. 64, ROMBUNNO.01-02 (2021).
[8] Dwork, C.: Differential Privacy, Proc. 33rd intl. conf. Automata, Languages and Programming-Volume PartII (Bugliesi, M., Preneel, B., Sassone, V. and Wegener, I., eds.), Lecture Notes in Computer Science, Vol. 4052, Springer, pp. 1-12(2006).
[9] U.S. CENSUS BUREAU: Disclosure Avoidance for the 2020 Census: An Introduction, 2021.
[10] Sweeney, L.: k-anonimity: a model for protecting privacy, Intl. J. Uncertainty, Fuzziness and Knowledge- Based Systems, Vol. 10, No. 05, pp. 557-570(2002).
[11] Masayuki Terada, Ryohei Suzuki, Takayasu Yamaguchi, Sadayuki Hongo: "On Publishing Large Tabular Data with Differential Privacy," IPSJ Journal, Vol. 56, No. 9, pp. 1801-1816 (2015).

Cover Art

Naruto Whirlpool, Awa Province, from Famous Views of the 60 Provinces

Utagawa Hiroshige (1797-1858)

Source: National Diet Library, NDL Image Bank (https://rnavi.ndl.go.jp/imagebank/)