Real-time people movement estimation in large disasters from several kinds of mobile phone data

Yoshihide Sekimoto
University of Tokyo
Tokyo, Japan
sekimoto@iis.u-tokyo.ac.jp

Akihito Sudo
University of Tokyo
Tokyo, Japan
sudoa@iis.u-tokyo.ac.jp

Takehiro Kashiyama
University of Tokyo
Tokyo, Japan
ksym@iis.u-tokyo.ac.jp

Toshikazu Seto
University of Tokyo
Tokyo, Japan
tosseto@csis.u-tokyo.ac.jp

Hideki Hayashi
Hitachi Ltd.
Tokyo, Japan
hideki.hayashi.xu@hitachi.com

Akinori Asahara
Hitachi Ltd.
Tokyo, Japan
akinori.asahara.bq@hitachi.com

Hiroki Ishizuka
KDDI Corporation
Tokyo, Japan
hk-ishizuka@kddi.com

Satoshi Nishiyama
KDDI Corporation
Tokyo, Japan
sa-nishiyama@kddi.com

Ubicomp/ISWC'16 Adjunct , September 12-16, 2016, Heidelberg, Germany
ACM 978-1-4503-4462-3/16/09.
http://dx.doi.org/10.1145/2968219.2968421

Abstract
Recently, an understanding of mass movement in urban areas immediately after large disasters, such as the Great East Japan Earthquake (GEJE), has been needed. In particular, mobile phone data is available as time-varying data. However, much more detailed movement that is based on network flow instead of aggregated data is needed for appropriate rescue on a real-time basis. Hence, our research aims to estimate real-time human movement during large disasters from several kinds of mobile phone data. In this paper, we simulate the movement of people in the Tokyo metropolitan area in a large disaster situation and obtain several kinds of fragmentary movement observation data from mobile phones. Our approach is to use data assimilation techniques combining with simulation of population movement and observation data. The experimental results confirm that the improvement in accuracy depends on the observation data quality using sensitivity analysis and data processing speed to satisfy each condition for real-time estimation.

Author Keywords
People mass movement, disaster activity, mobile phone data, data assimilation

ACM Classification Keywords
Multiagent systems, Spatial databases, GIS, Engineering, Sociology, Real time, Human safety

Introduction
Recently, the understanding of the mass movement of people in urban areas immediately after large disasters, such as in the Great East Japan Earthquake (GEJE), has become needed. In particular, mobile phone data is available as time-varying data (e.g., every hour) so that aggregated population data, even on a commercial basis (e.g., as "Mobile spatial statistics," available from NTT Docomo and "Congestion statistics" from Zenrin Datacom).

However, much more detailed movement that is based on network flow instead of aggregated data is needed for appropriate rescue on a real-time basis for different locations.

Hence, our research aims to estimate real-time population movement in large disasters from several kinds of mobile phone data. We first review related work in various fields in the next section. We then present our approach. In this paper, we simulate population movement in the Tokyo metropolitan area in a large disaster situation to obtain some fragmentary human movement observation data such as aggregated mesh-based population data from Call Detail Record (CDR) data from telecommunication companies and fire spread data from some mobile phone applications. Our approach is to use data assimilation techniques combined with people movement simulation and observation data. In the final section, we present our experimental evaluation results with respect to estimation accuracy, sensitivity analysis, and data processing speed to satisfy the conditions for real-time estimation.

Related work
Traffic condition estimation via cellular networks
Demissie et al. [4] executed real-time traffic volume estimation using cellular networks. Their achievement of real time estimation is notable. However, their target was road traffic status, which does not include walking or stationary people. Iqbal et al. [6] developed origin-destination matrices from mobile phone data. Their data includes user ID, which means they utilized non-anonymized data. Song et al. [10] estimates the behavior of people on a daily scale several days after an earthquake has occurred. Sevtsuk et al. [9] analyzed daily urban population mobility patterns using aggregated mobile phone data. A number of studies related to this field have been conducted. However, to our best knowledge, there is no previous study for the real-time accurate prediction of population distribution or flow via anonymized mobile phone data, which is applicable to various situations such as the day of a massive earthquake.

Nonlinear filtering in traffic fields
Mihaylova et al. [7] estimated freeway traffic volume using particle filters, where observations are collected by video camera. Cheng et al. [3] used mobile phone data to estimate highway traffic volume, also using a particle filter. In addition, Bayen and his colleagues [13,14] have done a number of studies that have applied data assimilation methods to traffic engineering. They employed a standard nonlinear method, and we believe this is the first study applying specialized particle filters to a problem in the GIS field.

our approach

Assumed situation

In this subsection, we explain the current disaster management system of national and local governments when a large-scale disaster occurs.

The national government and local governments (municipalities) set up their emergency headquarters within 30 min after a large-scale disaster occurs. The national government then examines the situation to allocate resources such as relief supplies and rescue teams for prefectures. The local governments also examine the situation to ensure the safety and security of residents and request support from neighboring prefectures. However, it is difficult to examine the situation because it takes a long time to understand its nature in a disaster. In the GEJE, prefectures hit by the disaster were almost unable to determine the number of deaths, missing persons, seriously injured persons, and persons with minor injuries. In this situation, local governments receive information about the earthquake intensity via TV and radio broadcasts, and government employees investigate their local neighborhood. Therefore, it is difficult to collect information in a disaster situation.

To solve this problem, we propose a system for estimating the movement of people and disaster situation from several kinds of mobile phone data.

First, the system starts to collect mobile phone CDR data when it detects that a large-scale disaster has occurred. At the same time, it starts to collect disaster information from mobile phones owned by some people. Fifteen minutes after the disaster occurs, it displays fragmentary disaster information collected from mobile phones. Thirty minutes after the disaster occurs, it displays the estimated movement of people and disaster status. After that, it updates the estimated results every 30 minutes.

As a result, local governments can then examine the situation to ensure the safety and security of residents based on fragmentary disaster information 15 min after the disaster occurs. After 30 min, they start to set up their emergency headquarters and request support from neighboring prefectures based on the estimated results of the movement of people and the disaster situation. At the same time, the national government establishes the Headquarters for Major Disaster Management and consults with the local government to allocate resources such as relief supplies and rescue teams to the prefectures based on the estimated results.

Creation of assumed situation

In the case of the Tokyo metropolitan area, the Japanese National Government has assumed some disaster scenarios depending on the date and time of a large earthquake. For example, if it happens in winter, collapsed buildings will tend to burn in one direction because of strong winds and dryness. The distribution patterns of people immediately after an earthquake are quite different depending on the time, e.g., nighttime, working hours, or morning/evening rush hour. Based on these points, initial population distribution, fire occurrence points, wind speed, and direction are generated according to the date and time (Figure 1).

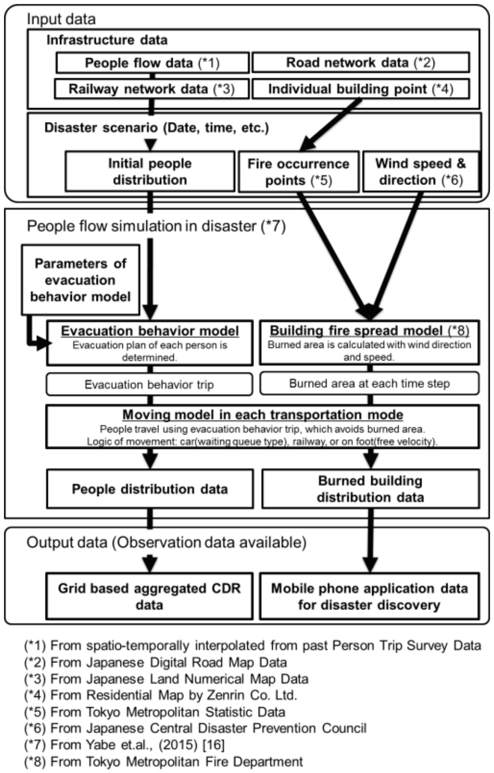

Figure 1: Creation of assumed disaster situations in the Tokyo metropolitan area

The initial population distribution at the input disaster time consists of population flow data that is spatio-temporally interpolated using past traffic statistics data collected in 2008 (See [8] for details of the interpolation method). Next, fire occurrence points are assumed based on statistical fire probability data published by Tokyo Metropolitan Fire Department. Fire occurrence points are stochastically allocated on building point data published by ZENRIN Co., Ltd.

To estimate the movement of people in this research, we consider an evacuation behavior model and building fire spread model. The building fire spread model is based on the model proposed in 1997 by the Tokyo Metropolitan Fire Department [12]. The evacuation behavior model is based on [15], in which activity start time and destination are stochastically determined depending on elapsed time. After determining these parameters, travel routes to escape the burning areas are determined. Moving spatio-temporal positions are decided depending on each transportation mode such as car, railway, or on foot.

Finally, we acquire fragmentary spatio-temporal observation data for the movement of people and burning buildings using a mobile phone application for disaster discovery. These two kinds of observation data are described in detail in the next subsections.

Grid based population data from aggregated CDR data
Our approach utilizes two kinds of data obtained from mobile phones. For the first type, we utilize aggregated mesh-based population data from the CDR data of a mobile telecommunication company. These CDR data are automatically and routinely recorded without the user's intention. In fact, every time a user makes or receives a phone call, short message service (SMS), multimedia messaging service (MMS), or data communication regarding the interaction and location of the user (to locate a base station for communication), this data is logged as CDR for billing purposes. Even if the user does not operate the mobile phone intentionally, some applications on the mobile phone periodically send "keep-alive" messages [1,2] every few minutes. Using these data, we are able to determine an individual user's location every several minutes using triangulation over the round trip delays of each base station to which the mobile phone of the user is connected. The provider of this data, as a mobile operator, has obtained a clear opt-in permission for the use of CDRs from each user for this research.

In this work, we generated an aggregated mesh-based population for privacy issues. The spatio-temporal resolution of each mesh is 500 m^2 and 5 min When a large-scale disaster occurs, some base stations will immediately stop service because of damages caused by the disaster. Further, other base stations will gradually stop because of power supply loss within three to six hours after the disaster occurs. Therefore, the population data from the CDR will also decrease over time. We created the observation population data of our experiment to consider the shutdown rate of base stations using an actual base station spatial distribution according to the disaster level.

Fire spread observation from mobile phone applications
The other type of data used from mobile phones is information on the sightings of building fires from individuals. This information is referred to as disaster information. We assume that this information is collected by a disaster response application on

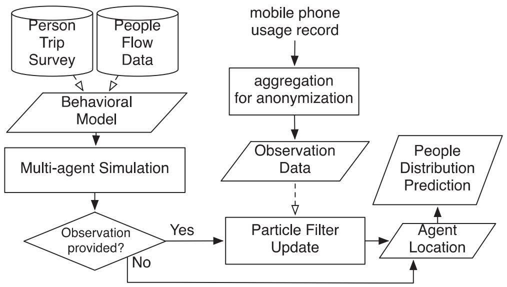

Figure 2: Population movement estimation algorithm from mobile phone mesh data

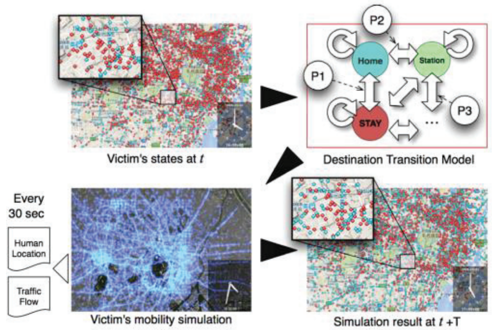

Figure 3: Multi-agent simulator

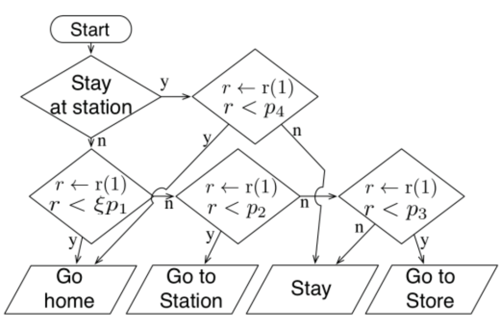

Figure 4: Behavior model for a disaster situation

smartphones. In this application, a person sends disaster information and his/her position using a simple and easy to use operation. In this research, the disaster information is created from population movement data, fire spread data, and data about the areas in which mobile phones cannot be used after a large-scale earthquake occurs. The fire spread and service loss data is transformed into time-series grid data. We assume that it is individuals in the fire spread areas but not in the service-loss areas that send the disaster information.

Estimation Methodology

Population distribution is estimated under the framework of data assimilation by formulating it using a state space model (See [11] for details). The state variables of the state space model are person location parameters and the behavior model, which is described later. In the system model, a multi-agent simulation based on the behavior model was used to describe the temporal development of state variables of person location, and Gaussian noise was added only to the parameters of the behavior model. The number of people in each grid cell, determined using Global Positioning System (GPS) or CDR data, was used as observation data while filtering was performed using a particle filter after creating the observation model based on the sampling and sensing processes (Figure 2,3,4).

System integration for real-time estimation

Using the estimation methodology described in the previous section, we implemented a system that integrates the simulation, grid based aggregated CDR, and fire spread observation data, assuming a large disaster. Figure 5 shows the overall timeline of data processing for this estimation system, which consists of population flow estimation (in blue), mobile phone CDR aggregation (brown), fire spread observation (red), and visualization (light green). In this example, the earthquake occurs at 18:00. Before that, the system provides normal population movement from human movement flow data. Mobile phone aggregated CDR data are uploaded every 5 min and used for population movement assimilation. Simultaneously, the fire spread observation data from smartphone applications is updated every 30 min and used to search for similar fire spread patterns in a database. These results are used to determine escape activities during the population movement assimilation.

Search speed of fire spread patterns

In the real-time data estimation timeline, the estimation of fire spread distribution is efficiently processed within the given time. We use a spatio-temporal similarity search method [5] to estimate fire spread distribution. This method searches a database that stores many simulated disaster scenario results represented by time-series grid data to find similar scenarios.

Figure 6 shows how the spatio-temporal similarity search estimates the fire spread distribution. First, we create a scenario dataset of fire spread data before a large-scale earthquake occurs. We input building information, wind direction, wind speed, and a fire point and output the scenario data. The fire spread scenario data is represented by time-series grid data that includes a time interval, grid geometry, and the number of burning buildings. This scenario dataset is stored in a database. After a large-scale earthquake occurs, the spatio-temporal similarity search method searches the scenarios of fire spread data for those that are similar to the disaster information obtained from the mobile phones. This disaster information includes the time, position, fire-flag (which indicates a fire burning at that time instance), and position. Here, the spatio-temporal search method extracts scenarios in which the grid data spatially and temporally intersects with the observed data at more than *k* points. This extracted scenario dataset is combined (by set union) with the observed to estimate the real fire spread distribution. The number of burning buildings in the grid data from multiple scenarios is the maximum number of burning buildings. Figure 7 shows an example of an estimated fire spread distribution.

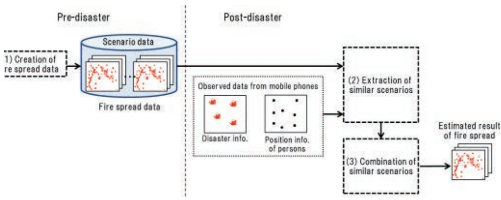

Figure 6. Estimation of fire spread using spatio-temporal similarity search

Figure 7. Example of estimated results of fire spread distribution ((c) Map tiles by Stamen Design, under CC BY 3.0. Data by OpenStreetMap, under CC BY SA)

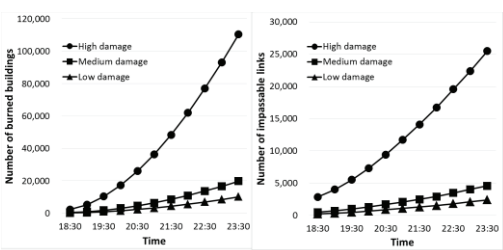

Figure 8. Time transition of damaged building and road by fire spread simulation

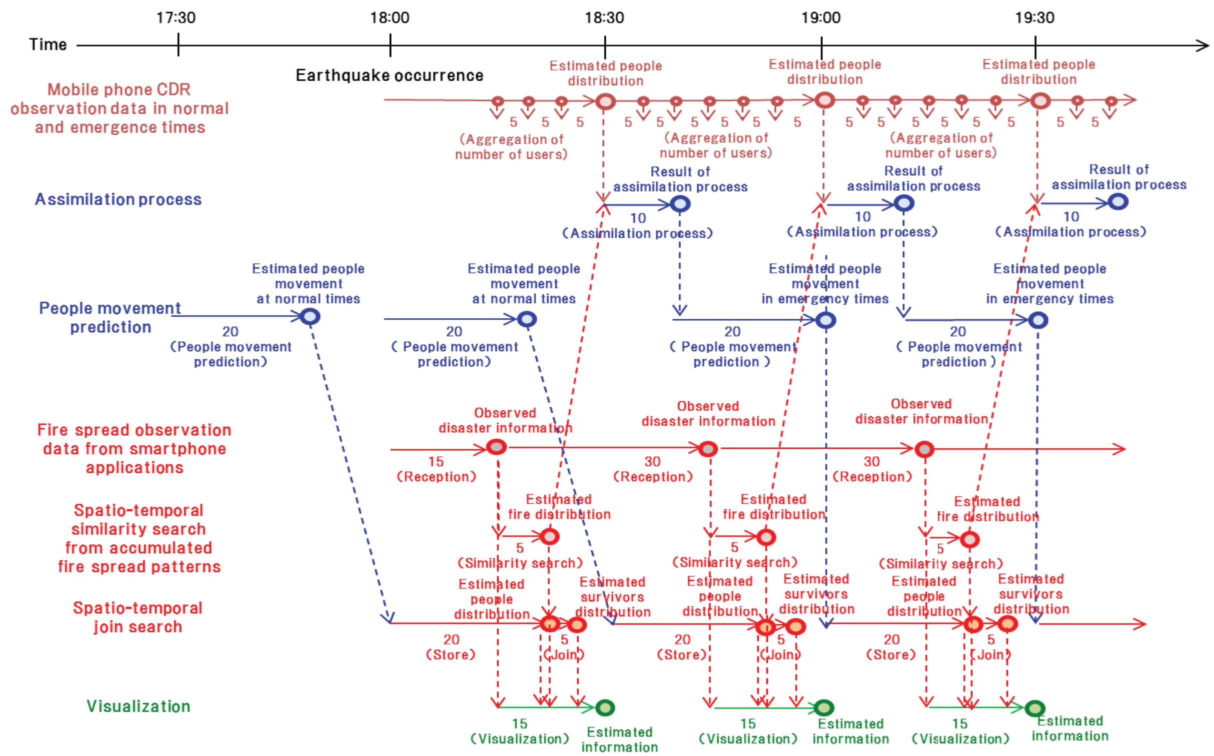

Figure 5: Timeline of data processing for real-time estimation

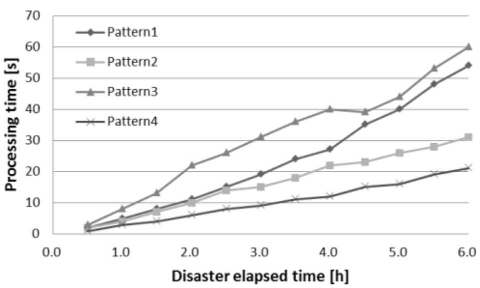

Figure 9. Processing time of fire spread distribution estimation.

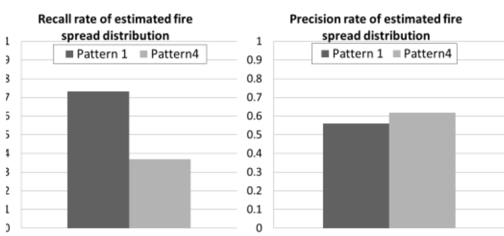

Figure 10. Recall and precision of fire spread distribution estimation.

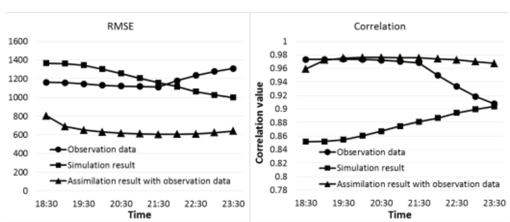

Figure 11. Time transition of estimation accuracy

Experiments

Experiment conditions

Table 1 shows the parameters of the four disaster patterns used in these experiments, which differ in terms of damage scale (or number of fire points, where 900 points indicates high damage and 150 points indicates low damage), disaster occurrence time, and number of disaster observations. In our simulation model, the number of burned buildings and impassable road links both increase depending on the disaster scale, as shown in Figure 8.

#	Damage scale	Number of fire points	Disaster occurrenc e time	Number of disaster observatio ns per 6 h
1	High	900	18:00	100,000
2	Low	150	18:00	100,000
3	High	900	13:00	100,000
4	High	900	18:00	8,000

Table 1. Disaster patterns used in the experiments

Estimation of fire spread distribution

In this section, we present the processing time needed to estimate the fire spread distribution. Additionally, we evaluate the estimation accuracy.

In this experiment, we created about 6,400 fire spread scenarios. The time ranges over 24 h and the spatial range consists of Tokyo and the surrounding areas. For the grid data, the time interval is 10 min, and the grid cells are about 100 m^2. The total number of time-series grid data items in these scenarios is about 15M items. The estimated results are also represented by time-series grid data, whose time interval and grid geometry are the same as the those of the scenarios. Given the data process flow for real-time estimation, we require the time to estimate the fire spread distribution to be less than 5 min.

Figure 9 shows the fire spread distribution estimation time for all disaster patterns. This results show that the estimation of fire spread distribution over all disaster patterns is at most 50 s, which satisfies the performance requirement.

Figure 10 shows the recall and precision of the estimated fire spread distribution for 23 wards in central Tokyo, six hours after a large-scale earthquake occurs. We believe that the number of disaster observations mainly affects the accuracy of the estimated fire spread distribution. Therefore, we compared the accuracy of estimated fire spread in Patterns 1 and 4. The left graph shows the recall of estimated fire spread distribution, which is defined by the ratio of the correctly estimated burning grid points to the number of burning grid points in the true data. The right graph shows the precision of the estimated fire spread distribution. This precision is the ratio of the number of correctly estimated burning grid points to the number of estimated burning grid points.

The results shows that the estimated result of Pattern 1 achieves higher recall than that of Pattern 4. This is because the disaster information in Pattern 1 is distributed widely and the estimated fire spread distribution is hence wider. However, Pattern 1 leads to lower precision than Pattern 4. This result shows that the number of imprecise burning grid points in Pattern

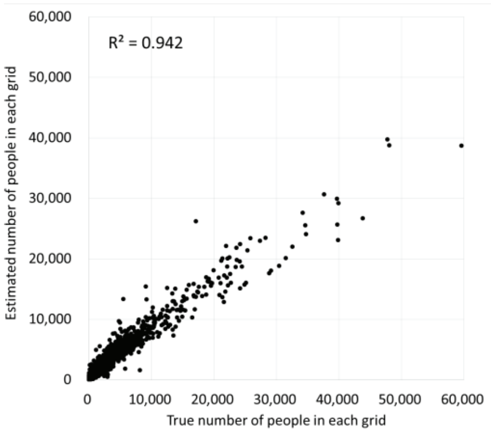

Figure 12. Correlation between true and estimated number of people

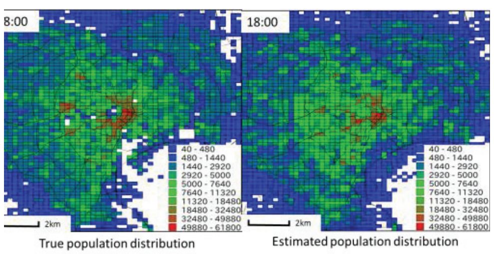

Figure 13. Comparison of true and estimated population distribution

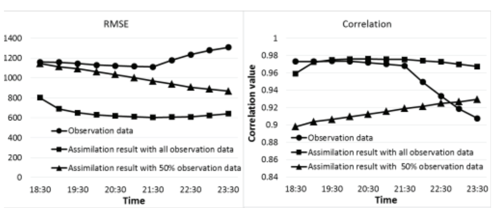

Figure 14. Accuracy with respect to observation data loss

1 increases. It is a future task to reduce the number of imprecise burning grid points.

Estimation of population distribution

Here, we explain the estimation results based on our proposed assimilation approach. Figure. 11 shows the estimation accuracy in terms of root mean squared error (RMSE) and correlation between the estimated and true number of people in each cell over time. We can easily see that the assimilated results are much better in terms of RMSE and correlation. In particular, Figure 12 compares the true and estimated number of people in all grids at 19:00 and shows good correlation, as $r^2 = 0.942$. In addition, Figure 13 shows the distribution results. The estimated distribution is a little smoother than the true distribution.

Sensitivity analysis based on observation data

In addition, we analyzed how the sensitivity depends on observation data. Figure 14 illustrates the difference in accuracy depending on the volume of observations used in the assimilation methods. We can see a clear difference in estimation accuracy given the observation data.

As shown in Figure 15, in contrast, the difference in penetration rate for the fire observation application does not affect the estimation accuracy. This seems to be because a high penetration of this application helps search for better fire expansion pattern candidates, but better candidates did not directly reflect on the accuracy of the population distribution in this study.

We then investigated whether the difference depends on the disaster scale. Consequently, we can see that a small-scale disaster provided better accuracy in Figure 16. This seems to be because of the changes were smaller in scale.

Finally, the processing time of the population distribution estimation is shown in Figure 17. This figure shows that the processing time increases with time because the number of people who are moving increases because of the evacuation. However, this processing time still remains within the 30 min of each assimilation step.

Conclusions

In this paper, we estimated real-time population movement in a large disaster from several kinds of mobile phone data using data assimilation techniques. In the experiments, we showed that the sensitivity of the results depends on the observation data quality and quantity in the Tokyo metropolitan area. Moreover, the processing time was well within the 30 min assimilation step duration. However, some problems still remain. The accuracy of the estimated network-based population movement volume is still not sufficient, although we estimate population movement. Additional improvement will be needed using network-based volume observation data.

Acknowledgements

A part of this work is supported by a consignment to research and develop techniques for the use and application of real-time information on the G-space platform from the Ministry of Internal Affairs and Communications, Japan. The authors greatly appreciate this kind and helpful support.

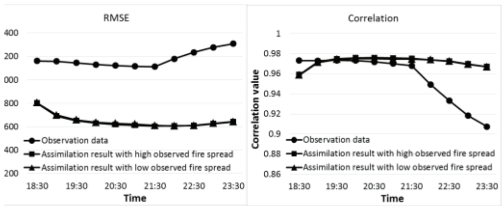

Figure 15. Accuracy with respect to penetration rate for fire observation application

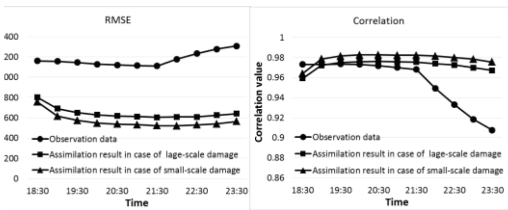

Figure 16. Accuracy comparison depending on disaster scale

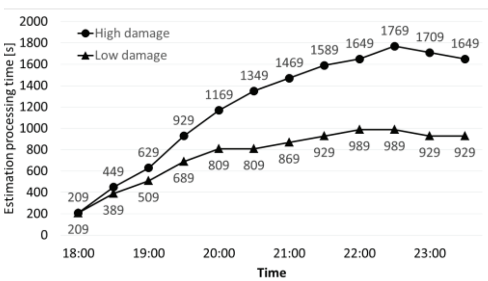

Figure 17. Processing time of estimation over 30 min

REFERENCES

1. Andrius Aucinas, Narseo Vallina-Rodriguez, Yan Grunenberger, Vijay Erramilli, Konstantina Papagiannaki, Jon Crowcroft, and David Wetherall. 2013. Staying online while mobile: The hidden costs. In *Proceedings of the Ninth ACM Conference on Emerging Networking Experiments and Technologies, CoNEXT'13*, 315–320, ACM, New York, USA.
2. Matthias Böhmer, Brent Hecht, Johannes Schöning, Antonio Krüger, and Gernot Baue. 2011. Falling asleep with Angry Birds, Facebook and Kindle: A large scale study on mobile application usage. In *Proce edings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services, MobileHCI'11*, 47–56, ACM, New York, USA.
3. Peng Cheng, Zhijun Qiu, and Bin Ran. 2006. Particle filter based traffic state estimation using cell phone network data. In *IEEE Conference on Intelligent Transportation Systems (ITSC2006)*, 1047–1052.
4. Merkebe G. Demissie, G. H. de Almeida Correia, and Carlos Bento. 2013. Intelligent road traffic status detection system through cellular networks handover information: An exploratory study. *Transportation Research Part C: Emerging Technologies*, 32, 76–88.
5. Hideki Hayashi, Akinori Asahara, Natsuko Sugaya, Yuichi Ogawa, and Hitoshi Tomita. 2015. Spatio-temporal similarity search method for disaster estimation. In *2nd Workshop on Advances in Software and Hardware for Big Data to Knowledge Discovery (ASH), in Conjunction with 2015 IEEE Conference on Big Data Conference*, 2462–2469.
6. M. Shahadat Iqbal, Charisma F. Choudhury, Pu Wang, and Marta C. Gonzalez. 2014. Development of origin–destination matrices using mobile phone call data, *Transportation Research Part C: Emerging Technologies*, 40, 63–74.
7. Lyudmila Mihaylova and Rene Boel. 2004. A particle filter for freeway traffic estimation, In *43rd IEEE Conference* on *Decision and Control (CDC2004)*, 2, 2106–2111.
8. Yoshihide Sekimoto, Ryosuke Shibasaki, Hiroshi Kanasugi, Tomotaka Usui, and Yasunobu Shimazaki. 2011. PFLOW: Reconstruction of people flow recycling large-scale social survey data. *IEEE Pervasive Computing*, 10, 4, 27–35.
9. Andres Sevtsuk and Carlo Ratti. 2010. Does urban mobility have a daily routine? Learning from the aggregate data of mobile networks, *Journal of Urban Technology*, 17, 1, 41–60.
10. Xuan Song, Quanshi Zhang, Yoshihide Sekimoto, and Ryosuke Shibasaki. 2014. 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*, 5–14.
11. Akihito Sudo, Takehiro Kashiyama, Takahiro Yabe, Hiroshi Kanasugi, and Yoshihide Sekimoto. 2016. Human mobility estimation following massive disaster using filtering approach. *Journal of Disaster Research*.11, 2, 217–224.
12. Tokyo Metropolitan Fire Department. 1997. Understanding and countermeasure of fire occurrence and spread by a big earthquake.
13. Daniel B. Work, Olli-Pekka Tossavainen, Quinn Jacobson, and Alexandre M. Bayen. 2009. Lagrangian sensing: Traffic estimation with mobile devices, In *American Control Conference (ACC '09)*, 1536–1543. IEEE.
14. Daniel B. Work, Olli-Pekka Tossavainen, Sebastien Blandin, Alexandre M. Bayen, Tochukwu Iwuchukwu, and Kenneth Tracton. 2008. An ensemble Kalman filtering approach to highway traffic estimation using GPSs enabled mobile devices. In *47th IEEE Conference on Decision and Control (CDC 2008)*, 5062–5068.
15. Takahiro Yabe, Yoshihide Sekimoto, Hiroshi Kanasugi, and Takehiro Kashiyama. 2015. Making real-time predictions of people's irregular movement in a metropolitan scale under disaster situations. *International Conference on Computers in Urban Planning and Urban Management (CUPUM2015)*.