Potential of Crowdsourcing Approach on Monitoring Radioactivity in Fukushima Prefecture

Wim Ikbal NURSAL*, Toshinori OKUDA and Toshihiro YAMADA

Graduate School of Integrated Arts and Sciences, Hiroshima University,
1-7-1 Higashi-Hiroshima, Hiroshima 739-8521, Japan
**e-mail: wimnursal@hiroshima-u.ac.jp*

Abstract

Crowdsourced data and professional scientists' data were compared for agreement regarding air dose rate levels and trends in air dose rate reduction to assess the value of public initiatives in radiation data collection during nuclear crisis response. This study used seven KURAMA datasets from seven survey periods to represent expert group data. To represent non-expert group data, we used seven datasets from SAFECAST's database the collection period of which was comparable to the KURAMA survey periods. A simple linear regression model was separately applied to a pair of combination datasets from different sources and also to a pair of first survey period datasets and subsequent datasets from the same source. The R-squared of the models showed the non-scientist group data correlating well with the corresponding expert data. The slopes of all the regression models, however, indicated that the air dose rate values measured by non-expert group were about 40 to 70 percent lower than those of the expert group. The air dose rate reduction trend from the crowd data showed a similar decreasing pattern compared to that of the expert group, although the discrepancy in the magnitude of dose reduction between them was as high as 14 percent. The discrepancy in air dose rate values suggest a careful interpretation of radiation information generated solely from crowdsourced data. Nevertheless, given the strong linear relationship of crowd data with scientist groups' data, the superior number of data points during a crisis, and the flexibility and agility of self-organization, we argue that the public could be a great partner to scientist groups in radiation data collection.

Key words: contamination, crowdsourcing, Fukushima, monitoring, radioactivity, SAFECAST

1. Introduction

Crowdsourcing of radiation data or voluntary data collection by lay people is a new approach to data collection on radiation in Japan. SAFECAST, an international Civil Society Organization (CSO) for citizen science and the environment, led such an initiative in response to a lack of radiation information available to the public soon after the Fukushima Daiichi Nuclear Power Plant accident (Fukushima accident; SAFECAST, 2011). SAFECAST still continues to collect data and develop a methodology for radiation data collection. It has gained interest and participation from people around the world. To date, SAFECAST's database has risen exponentially to 50 million records (data points) per July 2016 (SAFECAST, 2016).

The data collected by non-experts or non-scientist groups has been undermined by other groups skeptical of the validity of the data collection methodology (Bordogna *et al*., 2014). In our view, there is no perfect method for anything, including data collection. On the other hand, data quality assessment of any source is necessary before the data can be used for scientific or policy-making purposes. From this stance, we would like to assess how much the non-expert and expert data agree with each other by narrowing the geographic focus to Fukushima Prefecture area as the place most affected by the Fukushima disaster.

2. Materials and Methods

2.1 Crowdsourced and Expert Data

Crowdsourced data mostly consist of radiation measurements on the ground using a unique device carried by a moving vehicle such as a car (Brown *et al*., 2016). The device, called a "bGeigie," is a radiation detector integrated with electronics designed by SAFECAST for collecting necessary information, including geographic coordinates, dates and times, and storing this information in a flash memory card. Based on its specifications, the bGeigie uses a pancake-type Geiger-Muller detector (SAFECAST, 2013), widely known as a "GM counter."

Expert data such as the Nuclear Regulatory Authority

(NRA) or Japan Atomic Energy Agency's (JAEA) database include much thematic information. Not only do they include air dose rate measurements but also radioactive concentrations in many media (such as soil, fresh and marine water, the atmosphere and food). From this database, we selected air dose rates measured by the NRA through a car-borne survey known as KURAMA. We chose these data because the data collection methodology was quite similar to that of the citizen science group, that is, the use of a car carrying a radiation detector for measuring radiation. KURAMA has uses a NaI(Tl) scintillator and recently a CsI(Tl) based scintillator to detect and measure gamma radiation. Tsuda *et al*. (2015) provided a thorough investigation on the air dose rate and energy characteristics of this detector. The technical development of the KURAMA system was described by Tanigaki *et al*. (2013) and Tanigaki *et al*. (2015).

2.2 Dataset Specification

This study uses all KURAMA data available from 2011 to 2013, comprising seven datasets from seven surveys (JAEA, 2014). The data collection effort started and ended on a particular date and took around a week to two months to complete. These datasets as well as other thematic data can be freely accessed through http://emdb.jaea.go.jp/emdb/en/.

Since the non-expert group collected the data on an irregular basis, SAFECAST's database was divided into seven datasets which measured close to or within a KURAMA survey period. Each SAFECAST dataset holds an accumulation of three months (90 days) of measurements. Theoretically, the air dose rates of Cs-137 do not significantly decrease within six months. Table 1 introduces the selected datasets from both data sources used in this study.

2.3 Unit of Analysis

The non-scientist group and scientist group did not necessarily measure radiation exactly in the same location. Therefore it is impossible to compare the air dose rates from the two groups at any point in the study area. To solve this spatial problem, we assume that when both measurements were conducted at a distance of less than 100 meters from each other, the ambient doses measured did not differ significantly. This assumption adopts the opinion of Andoh *et al*. (2015) who argued that 90% of an air dose rate measured at a specified location comes from a radius of 60 meters from the contaminated area. The idea was implemented by representing the study area as a matrix of 100 meter square grids. The index of grids follows the National Standard Grid Square Framework (Ministry of Internal Affairs and Communication, 1996). Each grid was assigned a unique code. In each grid where two or more measurements existed, the air dose rate values were averaged.

2.4 Data Analysis

The easiest way to know how expert and crowdsourced approaches compare in measuring radiation would be by relating their data to each other, since the datasets are both about air radiation doses in the open environment. For our comparison we adopted a simple linear regression analysis, which is a widely used, very useful, straightforward statistical tool.

We performed linear regression analyses on two dataset combinations: (1) between datasets from different data sources, the acquisition periods of which are comparable, and (2) between datasets from the same source that had different survey periods. In the latter analysis, a radiation dose reduction rate from one survey period to the next could be assumed. The degree of decline in air dose rates may be a good way of comparing both methods in viewing the dynamics of radioactivity in the study area.

Finally, we further examined the number of observation from both approaches across different kinds of land cover. We used the seventh vegetation survey data from the Geospatial Information Authority of Japan to provide information about land cover in Fukushima Prefecture.

3. Results

Figure 1 presents the pattern we used for comparing air dose rate measurements by the citizen science group and by national expert group at the same place. Each panel in the figure shows a significant number of observations concentrated in the lower range of air dose rates and fewer observations in higher range values. Similarly, the variation in air doses becomes broader as air dose values increase.

Figure 1 illustrates that in all survey periods, the non-expert data correlate quite well with the expert data. On the other hand, the actual air dose rate values from non-expert measurements seem to be lower than those of the

Table 1 The KURAMA dataset and its analogous SAFECAST dataset used in this study.

KURAMA survey	KURAMA's Acquisition Dates	Analogous SAFECAST dataset
1st	2011/06/06 – 2011/06/13	2011/05/26 – 2011/07/25
2nd	2011/12/05 – 2011/12/28	2011/11/02 – 2012/01/31
3rd	2012/03/13 – 2012/03/30	2012/03/21 – 2012/05/05
4th	2012/08/20 – 2012/10/12	2012/08/01 – 2012/10/30
5th	2012/11/05 – 2012/12/10	2012/10/09 – 2013/01/07
6th	2013/06/12 – 2013/08/08	2013/05/12 – 2013/09/08
7th	2013/11/05 – 2013/12/12	2013/10/09 – 2014/01/07

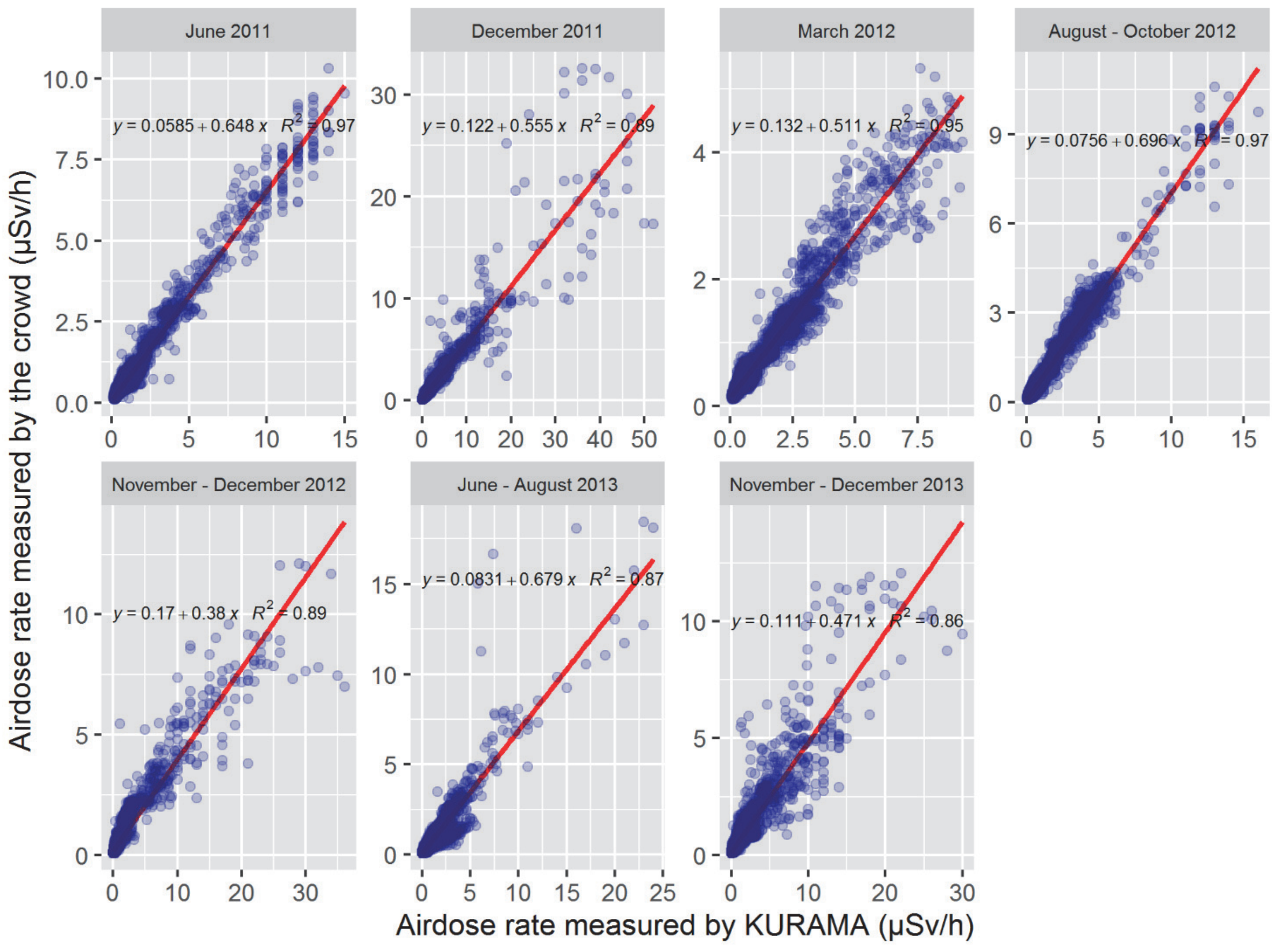

Fig. 1 Correlation diagrams of non-scientist group air dose rate data from (SAFECAST) to scientist group data (KURAMA), collected during parallel acquisition times. The figure is organized into seven panels according to the survey times of the KURAMA system.

professional group measurements in all observation periods. Represented by the slope of the figure, the discrepancy level between the crowdsourcing and professional approached a factor of 0.65 in the first survey period (June 2011), followed by 0.56 in the second survey period (December 2011), 0.51 in the third period (March 2012) 0.70 in the fourth period (August – October 2012), 0.38 in the fifth period (November – December 2012), 0.68 in the sixth period (June – August 2013), and 0.48 in the seventh survey period (November – December 2013).

To discover how similar the non-expert and expert methods were in depicting air dose rate trends, the data of the same group in the first survey period were paired with those of a subsequent survey period. Figures 2 and 3 present the estimated slopes, showing that both expert and non-expert data demonstrated a continuous decreasing trend from the first survey period towards the latest. The first panel of Fig. 2 (the first row and first column of the figure) of the air dose rate from the second KURAMA survey had decreased by about 25% in comparison to the first KURAMA survey. The subsequent panels of the same figure show that air dose rate in the third survey measurement by the expert group had decreased by about 36%, the fourth by 46%, the fifth by 56%, the sixth by 60%, and the seventh by 61%. During the same period, as illustrated by the first panel of Fig. 3, the percentage by which the second period of SAFECAST measurements had decreased compared to the first measurement period was 34% . Correspondingly, as shown in subsequent panels of the figure, the air dose rate in the third measurement period by the citizen science group had decreased by 50%, the fourth by 47%, the fifth by 64%, the sixth by 65%, and the seventh by 72%.

Based on Fig. 4, in the early measurement period of 2011, the citizen science group collected more radiation data than the national authority. The number of observations, however, fell off in the following years. Both radiation measurement approaches showed relatively significant numbers of observations in urban and suburban environments but a lack of observations in forested areas throughout the survey periods.

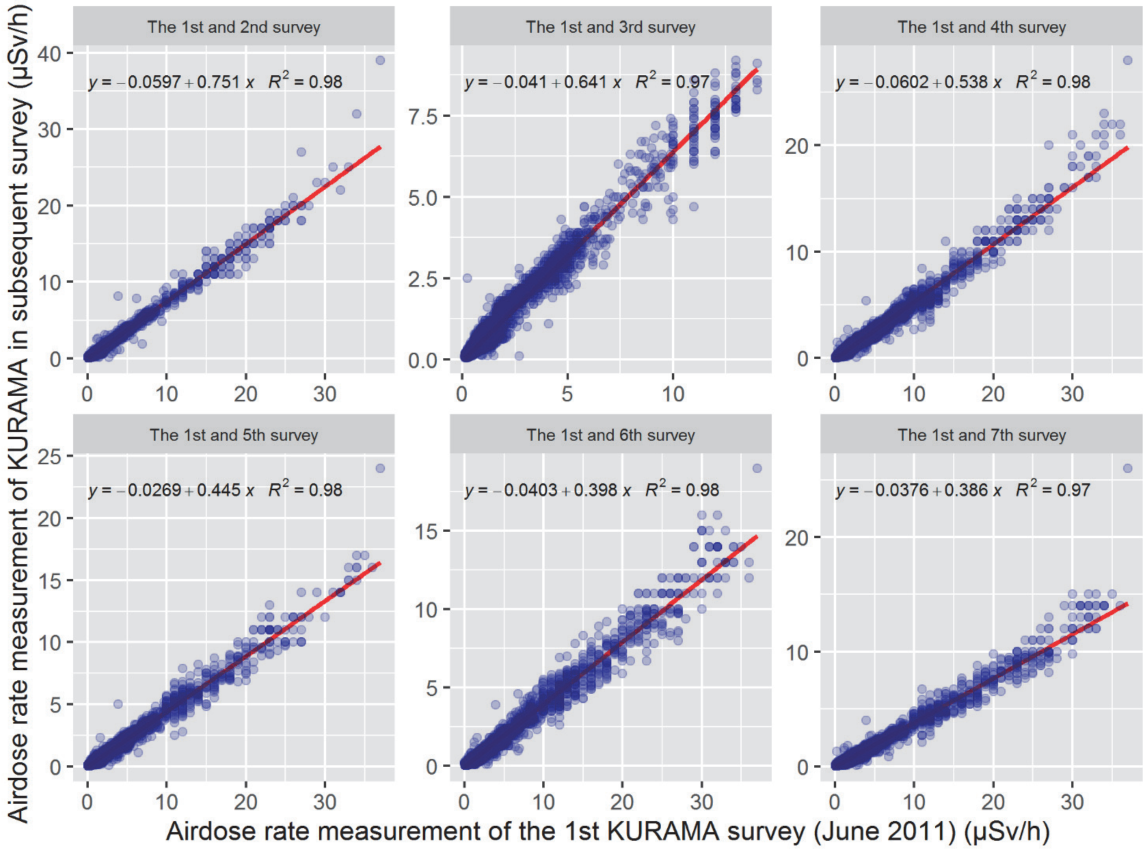

Fig. 2 Correlation diagram of air dose rates between the first and the subsequent surveys of the expert group.

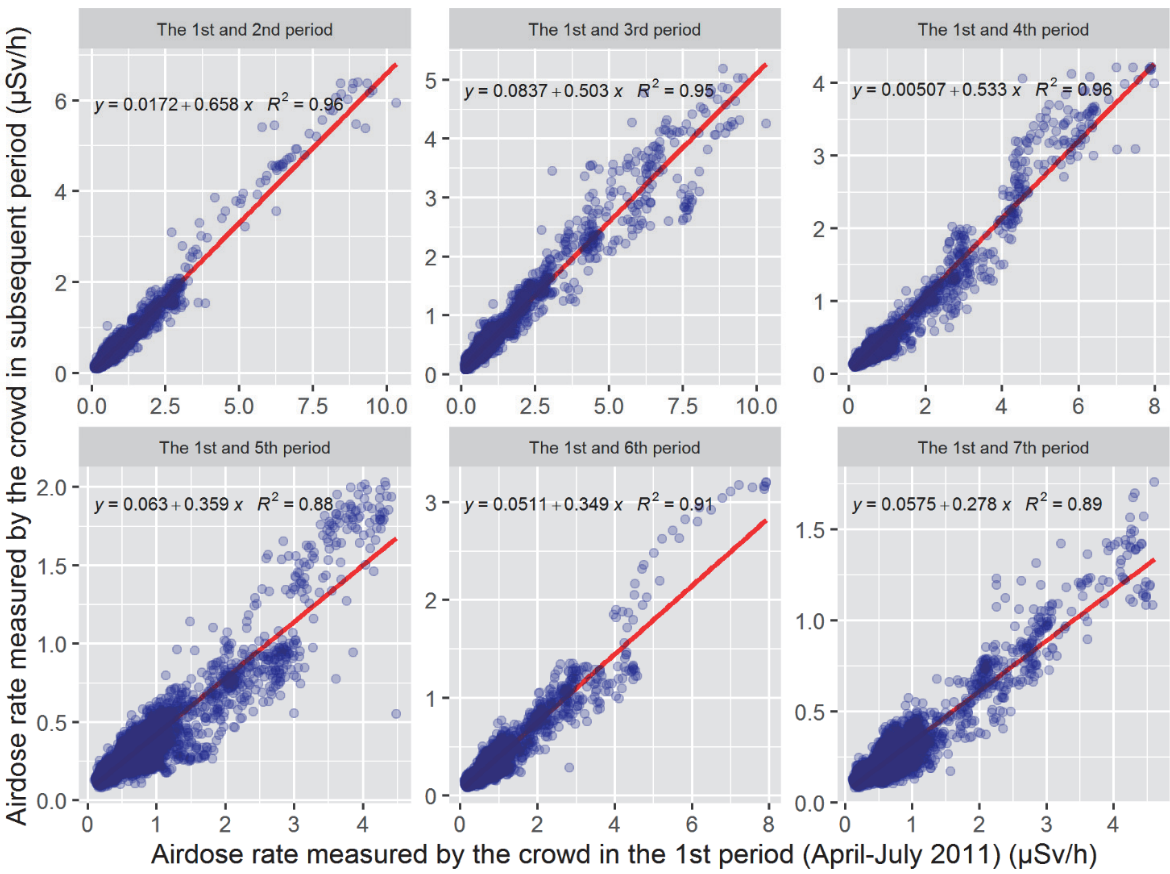

Fig. 3 Correlation diagrams of the air dose rates between the first selected measurement period and subsequent periods of the citizen science group.

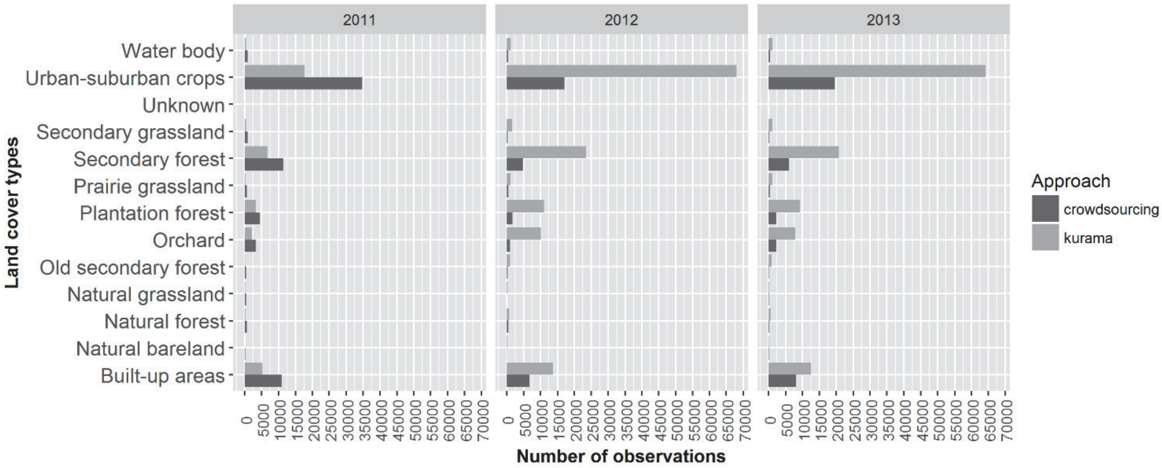

Fig. 4 Numbers of observations provided by crowdsourcing and the expert survey across several land cover types in Fukushima Prefecture after the Fukushima accident.

4. Discussion

4.1 Agreement in Dose Rate Measurements

The large number of observations in the lower range of air dose rates shown in Fig. 1 was probably due to the large extent of the contaminated area with such a range of air dose rates and partly because access to the highly contaminated part of Fukushima Prefecture was restricted. Meanwhile, the high variation in the high air dose rate regions might be due to the detector types used. Knoll (2010) stated that application of this kind of detector is less useful at high counting rates because of a well-known dead time phenomenon that necessitates application of a dead time correction. Any bGeigie instrument utilizes a dead time compensation formula in its counting system (SAFECAST, 2014). Another possible cause of high variation in the upper range of air dose rate measurements is a seasonal factor, together with the detector sensitivity factor mentioned earlier. It is quite clear from the figure that the magnitude of variation was relatively larger in the survey periods of December 2011, March 2012, November – December 2012, and November – December 2013. During these periods, there could have been some amount of snow cover in parts of Fukushima Prefecture when measurements were undertaken. Tanigaki *et al*. (2015) found that air dose readings by the KURAMA-II system were greatly affected by heavy snow occurrence. The sixth panel of the June – August survey period, when measurements were conducted in summer, is an exception, but there is a chance of outliers also affecting the variation.

The coefficient correlation in Fig. 1 signifies that the non-expert data can be estimated from the expert group data. Although both data are well correlated, the non-expert measurements of air dose rate values of can be 40% – 70% lower than the values of the expert group measurements. We suspect that at least two factors that might contribute to the discrepancy. First, it might be due to the how the detector is mounted on the car. The bGeigie is usually set on either side of a car window (supported by a belt strap locked to the hand grip inside the car). It thus faces either to the left or the right side of the car. Because of this placement, some number of photons coming from behind the detector might be blocked by the body of the car. Also, due to the physical design of the detector, such that a thick steel case covers its back side, the direction the detector is facing would affect the number of photons coming into the detector's window. Second, as we recognized previously, seasonal conditions might also influence the response of GM counters. The slopes of the regression lines shown in the panels associated with winter or early spring measurements were relatively smaller than the slopes in other panels. The shielding effect of snow that had affected the measurements had already been compensated for in the expert measurements database (JAEA, 2014). Since similar efforts have not made yet with the citizen science group data, it is likely that the snow-shielding effect have shown up as a lower slope values in the regression lines.

Both the citizen science group and expert group data include natural background radiation. Natural background radiation may include radiation from terrestrial sources such as uranium, thorium and radium, and extraterrestrial phenomena such as cosmic rays. The detector used by the expert group was not designed to detect cosmic rays, therefore cosmic rays may also have influenced the discrepancy between the air dose rate values of the two measurement approaches. The intercept values in Fig. 1, ranging from 0.06 to 0.17, might reflect the influence of cosmic rays on the citizen science group data. In highly contaminated areas, the contribution of total background radiation to air dose rate mesurement values would not be substantial and neither would that of cosmic rays. On the other hand, they become significant when measurements are undertaken in low to very low contaminated areas or under normal conditions.

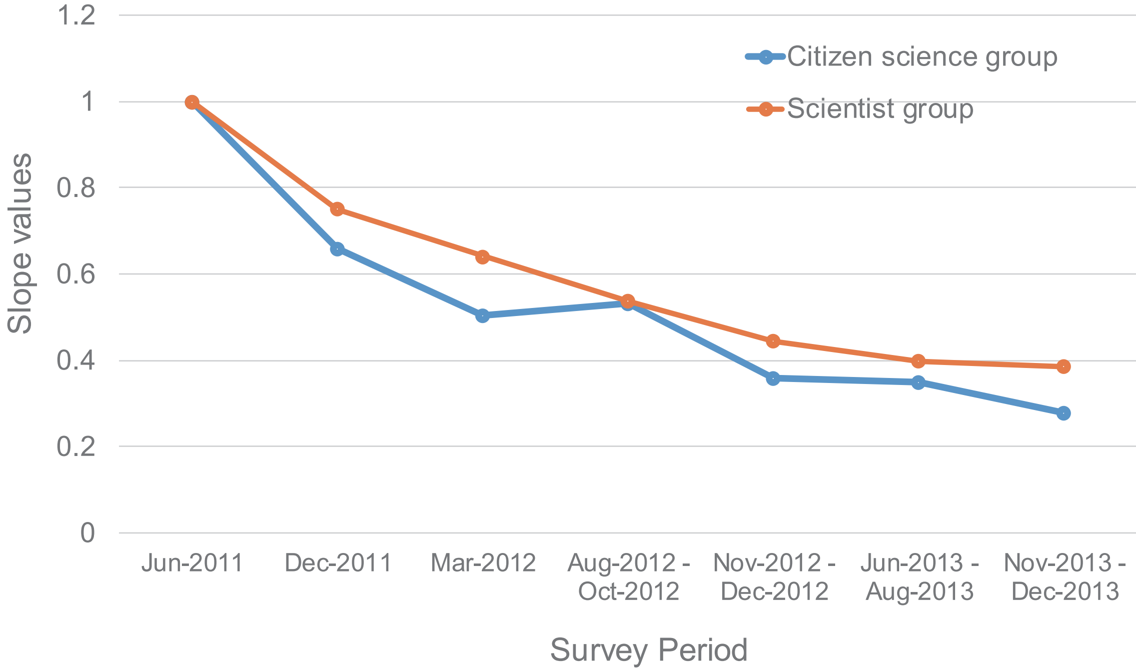

Fig. 5 Air dose rate reduction trends throughout the survey period (2011-2013), taken from the slope of regression model between the first survey (June 2011) and second survey (December 2011), third survey (March 2012), fourth survey (August 2012 – October 2012), fifth survey (November 2012 – December 2012), sixth survey (June 2013 – August 2013) and seventh survey (November 2013 – December 2013).

4.2 Reduction of Air Dose Rates

The non-expert data show a decreasing trend in air dose rates, and so do the expert group data, as discerned from the slopes of the regression lines in Figs. 2 and 3, respectively. We took the slopes from Figs. 2 and 3, and used them to see their performance in the course of the survey periods, as illustrated in Fig. 5. It clearly shows that the slopes from both non-professional scientists and expert groups are going in the same direction, which has started leveling since the last period of the survey. The difference in the air dose rate reduction between the two methods based on their regression slopes ranges from 0.5% to 14%. We believe that this difference came about as a consequence of detector characteristics and measurement outcomes.

4.3 Radiation Measurements in Forested Areas

It is reasonable for any measurement which depends primarily on human or so-called human-centric sensor types (Srivastava *et al*., 2012) to have limited movement throughout a wide landscape, particularly a landscape that has a complex physical form or condition, such as various forms of land cover, terrain or road access as in Fukushima Prefecture. What we can learn from the citizen science approach is that the response of the public toward provision of radiation information in the early period of the crisis was very positive given that many observations were done in urban and suburban environments (Fig. 4).

Improvement for more extensive monitoring of forested land is imperative in the long term, since radioactive matter is highly concentrated in the forests (Hashimoto *et al*., 2012). Improvement of the crowdsourcing approach may be accomplished by segmenting the volunteers (Rossiter *et al*., 2015) or developing a tasking system (Boulos *et al*., 2011) prioritizing areas that have not yet been visited or need to be revisited. SAFECAST has been developing a fixed sensor network installed in fixed locations and positions as an alternative system for monitoring the environment with less effort from human operators. This new instrument will continuously update the air dose rate in near real time.

5. Conclusion

This study extends the discourse about the quality of information that could be acquired by citizen participation in science. We investigated the radiation data of SAFECAST and data managed by the NRA and JAEA for agreement on air dose rate values, reduction in air dose rates from 2011 to 2013, and the number of measurements across several land cover types. We presented evidence that the air dose rate values from crowdsourced radiation measurements are well correlated with scientist group measurements. The real air dose rate values from the citizen science groups, however, were lower than those of the expert groups, ranging from 40% up to 70% lower. We also assessed trends in air dose rates indicated by the slope of the linear regression model between two datasets of different survey periods but from the same source. The result showed that the trend of air dose reduction generated from citizen science group data followed the same direction of the trend provided from scientist group data. The magnitude of air dose rate reduction of the citizen group data toward the expert group data is lower than that of the expert group, the discrepancy between which can be as high as 14%. We discussed some factors that might cause such discrepancies in measurement values, which are mainly associated with GM counter characteristics and sensitivity. We provided evidence that a crowdsourcing approach to radiation data is responsive to a crisis. Especially for urban and suburban areas, a crowdsourcing approach could potentially be relied on to provide radiation information after a nuclear accident.

Given the significant discrepancy in air dose rate values, we would like to suggest that the radiation information provided by citizen science, especially from measurements with GM counters would need supplemental and comparative material with a brief explanation on the existing discrepancies. Regarding the utilization of citizen science data, we would emphasize the importance of preprocessing or pre-analysis stages including data selection and conversion, before further using them for generating information. Since the air dose values show discrepancies with a seasonal pattern, data selection based on period of measurement is crucial. The selection of datasets based on the detector type from the SAFECAST database is important as well because the database may contain numerous measurements using a variety of detectors. Each detector has unique conversion factors to other measurement units.

We argue that, since both data sources have quite a good linear relationship, the use of citizen science through the SAFECAST database to provide radiation information is worth consideration. It is particularly beneficial in areas lacking data due to government resource limitations or because national monitoring is no longer required there.

Acknowledgement

We would like to express our sincere gratitude to the reviewers for their constructive comments. We are in debt to and sincerely thank certain prominent SAFECAST members: Pieter Franken, Nick Dolezal and Joe Moross for helping us understand and use their data. We also want to express our gratitude toward the Japan Atomic Energy Agency (JAEA) for providing public access to their data. We also thank Prof. Kiyoshi Shizuma from Hiroshima University for valuable knowledge and input to improve this study. The Hiroshima University Phoenix Leader Education Program (Hiroshima Initiative) for the Renaissance from Radiation Disaster funded by the Ministry of Education, Culture, Sports, Science, and Technology financially supported this study.

References

Andoh, M., Y.Nakahara, S. Tsuda, T. Yoshida, N. Matsuda, F.Takahashi, S. Mikami, N. Kinouchi, T. Sato, M. Tanigaki and K. Takamiya (2015) Measurement of air dose rates over a wide area around the Fukushima Dai-ichi Nuclear Power Plant through a series of car-borne surveys. *Journal of Environmental Radioactivity*, 139: 266–280.

Bordogna, G., P. Carrara, L. Criscuolo, M. Pepe and A. Rampini (2016) On predicting and improving the quality of Volunteer Geographic Information projects. *International Journal of Digital Earth*, 9: 134–155. DOI: <http://dx.doi.org/10.1080/17538947.2014.976774>

Brown, A., P. Franken, S. Bonner, N. Dolezal and J. Moross (2016) Safecast: successful citizen-science for radiation measurement and communication after Fukushima. *Journal of Radiological Protection*, 36: S82–S101.

Boulos, M.N.K., B. Resch, D.N. Crowley, J.G. Breslin, G. Sohn, R. Burtner, W.A. Pike, E. Jezierski and K.Y.S. Chuang (2011) Crowdsourcing, citizen sensing and sensor web technologies for public and environmental health surveillance and crisis management: trends, OGC standards and application examples. *International Journal of Health Geographics*, 10: 1–29.

Hashimoto, S., S. Ugawa, K. Nanko and K. Shichi (2012) The total amounts of radioactively contaminated materials in forests in Fukushima, Japan. *Scientific Reports*, 2: 416 – 420.

Japan Atomic Energy Agency (JAEA) (2014) Database for Radioactive Substance Monitoring Data. <http://emdb.jaea.go.jp/emdb/en/>

Knoll, G.F. (2010) *Radiation Detection and Measurement*. John Wiley & Sons. New Jersey.

Ministry of Internal Affairs and Communication (MIAC) (1996) Method of Demarcation for Grid Square. <http://www.stat.go.jp/english/data/mesh/05.htm>

Neuman, R. (2015) Driving greater awareness of the Fukushima disaster's health impacts with cloud analytics. <https://customers.microsoft.com/en-US/story/driving-greater-awareness-of-the-fukushima-disasters-h>

Rossiter, D.G., Liu, J., Carlisle, S. & Zhu, A.X. (2015) Can citizen science assist digital soil mapping?. *Geoderma*, 259: 71–80.

SAFECAST (2011) About SAFECAST. <http://blog.safecast.org/ about >

SAFECAST (2013) bGeigieNano Features_ specifications. <https://github.com/Safecast/ bGeigieNanoKit /wiki/ bGeigieNano-Features_specifications>

SAFECAST (2014) Nano Operation Manual.

SAFECAST (2016) 50 Million Data Points. <http://blog.safecast.org/2016/07/50-million-data-points>

Saito, K. and Y. Onda (2015) Outline of the national mapping projects implemented after the Fukushima accident. *Journal of Environmental Radioactivity*, 139: 240 – 249.

Srivastava, M., T. Abdelzaher and B. Szymanski (2012) Human-centric sensing. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 370: 176 – 197.

Tanigaki, M., R. Okumura, K. Takamiya, N. Sato,H. Yoshino and H. Yamana (2013) Development of a car-borne γ-ray survey system, KURAMA. *Nuclear Instruments and Methods in Physics Research A*, 726: 162–168.

Tanigaki, M., R. Okumura, K. Takamiya, N. Sato, H. Yoshino, H. Yamana, H. Yoshinaga and Y. Kobayashi (2015) Development of KURAMA-II and its operation in Fukushima. *Nuclear Instruments and Methods in Physics Research A*, 781: 57–64.

Tsuda, S., T. Yoshida, M. Tsutsumi and K. Saito (2015) Characteristics and verification of a car-borne survey system for dose rates in air: KURAMA-II. *Journal of Environmental Radioactivity*, 139: 260 – 265.

(Received 20 December 2016, Accepted 6 March 2017)