DOCTORAL THESIS

THE ASSESSMENT OF THE CROWDSOURCED RADIATION DATA TO PROVIDE RADIOLOGICAL INFORMATION IN FUKUSHIMA PREFECTURE, JAPAN

WIM IKBAL NURSAL

DIVISION OF INTEGRATED ARTS AND SCIENCES
GRADUATE SCHOOL OF INTEGRATED ARTS AND SCIENCES
HIROSHIMA UNIVERSITY
SEPTEMBER 2017

ABSTRACT

The necessary radiological information was not quickly and widely published by Japan Government during the crisis phase of Fukushima Dai-ichi Nuclear Power Plant disaster. This situation encouraged some individuals from various background to organize, design their own detectors, and mobilize lots of layperson to measure radiation in their environment, and to make all the data to be stored and published in the internet. Despite of large measurements data have been collected, it seems that very little recognition from the expert/scientist group has been given to the work that they have done. This study is an attempt to assess the quality of crowdsourced radiation data in Fukushima by examining the agreement of the data with the expert group data and the possibility to extend measurement on the complex landscape such as forests.

A simple linear regression models were made on radiation data from citizen-scientist group and expert group to know the agreement of air dose rate levels and trends in air dose rate reduction between them. We used KURAMA data from seven survey periods to represent expert group data and seven datasets from SAFECAST data to represent citizen scientist group data of which acquisition periods were comparable to the KURAMA survey periods. The *R-squared* of the models showed the citizen-scientist group data correlated well with the corresponding expert data. The slopes of all the regression models, however, indicated that the air dose rate values measured by the citizen-scientist group were about 30 to 60 percent lower than those of the expert group. The air dose rate reduction trend from the crowdsourced data showed a similar decreasing pattern compared to that of the expert group, although the discrepancy in the magnitude of dose reduction between them could be as high as 18 percent. The present of discrepancy in air dose rate values suggest a careful interpretation of radiation information generated solely from crowdsourced radiation data.

Since the crowdsourced radiation data were collected by layperson on the ground, the measurements tend to be done on accessible place such as road side and seldom on difficult to access landscape object such as forests. In this study, we investigate whether extending the target of measurements on the forest using the citizen-science designed detector and UAV (Unmanned Aerial Vehicle) will be possible to provide reliable radiation information as well. We set up one hectare plot on deciduous forest and measured the air dose rate at one meter above the ground using the citizen-science detector. Measurement of air dose was performed on the UAV. The measurement took place while it was moving above forest canopy of the plot. After the data were collected, air dose rate surface model was developed on both datasets. Based on the visual comparison, the air dose surface model measured on the UAV mostly did not show a similar pattern to the surface model of air dose rate measured on the ground. We suspected that it was probably due to the high autocorrelation within the air dose rate data measured on the UAV.

From this study, we learned that while citizen-scientist group with crowdsourcing approach could potentially be alternative source of radiation information and a great partner to the expert groups in radiation data collection. Although there are some limitations in their data and detector, with the agility and openness environment that the citizen science group have, the underlying problems of those limitations may be addressed soon.

ACKNOWLEDGEMENT

Firstly, the author would like to thank to Allah, The Creator of the World and The Ultimate Custodian, because of His Compassion and Mercy that the author received. His blessings would continuously be precipitated with or without this small work. The author would only expect that the outcome of this work in any form would be useful for any interested reader if not contribute to the recovery of nuclear disaster.

In this opportunity, I would to express my great gratitude to Prof. Toshinori Okuda for being relentless and patient to guide the author to complete this work and study in general. The similar appreciation goes to Prof. Toshihiro Yamada who opens a big door for listening and consultation. The author would like to convey warm appreciation to Prof. Tetsuji Okamoto and Prof. Kiriko Sakata for their invaluable help and support that the author received during the period of study in Hiroshima University. The author would express more than thanks to Prof. Kiyoshi Shizuma for his kindness to reply every basic question about radiation that the author asked, to involve the author in river sediment contamination study, and share his vast knowledge on radiation, supervise and suggest the author in measurement, and so forth. Last but not the least, the author is grateful for supports that been given by Phoenix Program of Hiroshima University.

This study and work would not be completed without the patience of my beloved family. I would like to sincerely ask their forgiveness if they ever felt that they were sitting between myself and my academic works. My immense thanks for their kind understanding for the author during the difficult times, although the words would not be able to return those times that belong to them.

LIST OF TABLES

Table 2-1. Public participation in science which can be classified into three categories, based on the scope of participation i.e. contributory, collaborative, and co-created project participations (Bonney et al., 2009) 9

Table 2-2. Approaches and methods for estimating quality measures and indicator (modified from Senaratne et al., 2017)........ 11

Table 3-1. KURAMA dataset and its associated SAFECAST dataset used in this study.. 17

Table 4-1. Statistics of air dose rate of two different ground measurement methods 43

LIST OF FIGURES

Figure 3-1. Correlation diagram of the air dose rate of non-scientist group data (SAFECAST) to scientist group data (KURAMA) which collected in a parallel acquisition time. The figure is composed by seven panels which indicates the survey time of KURAMA system. 19

Figure 3-2. Correlation diagram of air dose rates between the first survey with the subsequent survey of the expert group. 21

Figure 3-3. Correlation Diagram of the Air Dose Rates Between the First Selected Measurement Period to the Subsequent Period of the Citizen Science Group. 23

Figure 3-4. The Number of Observation between Crowdsourced and Expert Approach at Several Land Use Types in Three Different Time Acquisition: (A) 2011, (B) 2012, and (C) 2013. An observation means a measurement conducted at a certain place or a grid, since the whole study area was represented into grid that follows National Grid Square Framework of Japan. Each grid was counted only once. 24

Figure 3-5. The Trend of Air Dose Rate Reduction throughout the Survey Period (2011-2013) which Taken from the Slope of Regression Model Between the First Survey (June 2011) and the Second Survey (December 2011), between the First Survey and the Third Survey (March 2012), the Fourth (August 2012 – October 2012), the Fifth (November 2012 – December 2012), the Sixth (June 2013 – August 2013) and the Seventh (November 2013 – December 2013) 27

Figure 4-1. Study Site in the Iitate Village. The area bounded by green box was area of interest for comparing the air dose measurements from different detectors using UAV. The plot shown by orange grid mesh is a plot in deciduous tree patch. 33

Figure 4-2. The design of research (a) UAV and radiation detectors (b) flight route of UAV 35

Figure 4-3. Auto-correlogram of the Data Measured by GM Counter and Scintillator on the UAV. ACF is the function that shows the level of correlation of the dataset to itself. Lag is the sequence of observation. Figure in the panel (A) shows a significant dependency of every air dose rate value to all measurements which conducted by UAV with GM counter. While a figure in panel (B) shows that an air dose rate values that measured by scintillator on UAV was significantly dependent to the next three observations (air dose rate values). 38

Figure 4-4. The surface model that generated by ordinary kriging interpolation to all survey modes (ground and aerial measurements). Figure shown in the panel (A) is a surface model of air dose rates that measured through ground measurement with stationary approach, panel (B) is surface model of air dose that measured through the UAV with stationary approach, figure in the panel (C) and (D) are air surface model of air dose rates that measured through UAV with GM Counter type detector which flew from North to South and East to West respectively, and panel (E) and (F) are figure of surface model of air dose rate measured through similar method as previous two panels but the detector used was scintillator detector. 39

Figure 4-5. Linear relationship between the UAV data from both detectors its flight direction to the ground data with wandering measurement approach. 41

Figure 4-6. Scatter plot of air dose rate measured at fix points and air dose rate obtained through wandering approach. 42

LIST OF APPENDIX

Appendix 1. The KURAMA data acquired in 2011/12/05 – 2011/12/28 (2nd Survey).....56

Appendix 2. The parallel dataset of crowsourced data to the 2nd survey of KURAMA data 57

Appendix 3. The KURAMA data acquired in 2012/08/20 – 2012/10/12 (the 4th Survey) 58

Appendix 4. The parallel dataset of crowdsourced data to the 4th survey of KURAMA data 59

Appendix 5. The KURAMA data acquired in 2013/11/05 – 2013/12/12 (the 7th Survey) 60

Appendix 6. The parallel dataset of crowdsourced data to the 7th survey of KURAMA data 61

Appendix 7. R code for preprocessing task on crowdsourced radiation data.....................62

TABLE OF CONTENTS

ABSTRACT i

ACKNOWLEDGEMENT iii

LIST OF TABLES iv

LIST OF FIGURES v

LIST OF APPENDIX vii

TABLE OF CONTENTS viii

CHAPTER 1. GENERAL INTRODUCTION 1

1.1. Background 1

  1.1.1. The Need for Radiation Monitoring 1

  1.1.2. Public Involvement in Radiation Monitoring 2

  1.1.3. The Advantage of Public Involvement in Radiation Measurement 3

1.2. Research Questions 4

1.3. The Structure of the Dissertation 5

CHAPTER 2. CROWDSOURCED DATA QUALITY 7

2.1. Crowdsourcing Related Terms and Definition 7

2.2. Crowdsourced Data Quality 10

CHAPTER 3. CITIZEN SCIENCE AND EXPERT-GROUP RADIATION MONITORING DATA COMPARISON 13

3.1. Introduction 13

3.2. Methodology 14

  3.2.1. Citizen Science and Expert Data 14

  3.2.2. Dataset Specification 15

3.2.3. Unit of Observation 15
3.2.4. Data Analysis 18
3.3. Results 20
3.4. Discussion 25
3.4.1. Agreement in Dose Rate Measurement 25
3.4.2. The Reduction of Air Dose Rates 28
3.5. Conclusion 29
CHAPTER 4. SCALING UP CITIZEN SCIENCE DATA BY AERIAL VEHICLE PLATFORM 31
4.1. Introduction 31
4.2. Methodology 32
4.2.1. Study Area and Observation Plot 32
4.2.2. Data Collection Method 32
4.2.3. Data Analysis 36
4.2.4. Result 37
4.3. Discussion 44
4.3.1. Autocorrelation 44
4.3.2. Comparison of Survey Methods in Air Dose Rate Mapping 44
4.4. Conclusion 46
CHAPTER 5. GENERAL CONCLUSION AND RECOMMENDATION 47
5.1. GENERAL CONCLUSION 47
5.2. RECOMMENDATION 48
REFERENCES 50
APPENDIX 55

CHAPTER 1. GENERAL INTRODUCTION

1.1. Background

1.1.1. The Need for Radiation Monitoring

After Fukushima Dai-ichi Nuclear Power Plant (FDNPP) accident followed by the (Yasunari *et al*., 2011) which dominated by forests. Cs-137 has been the responsible pollutant which contributes to long term elevated radiation in Fukushima Prefecture and its neighboring regions due to its huge amount of quantity deposited into the land and has relatively longer half-life (Steinhauser *et al*., 2014).

In Fukushima soil, the mobility of Cs-137 is very low that caused its availability to plants is also low, due to high mineral content that strongly binds the radiocaesium (Nakao *et al*., 2012). Although Cs-137 is less mobile chemically, it is still able to be transported by physical force such as soil erosion. That was why the option to fully decontaminate the forest came into the table of Japan government decision makers, although many studies suggested that it would not be ecologically and economically feasible (such as in Hashimoto *et al*., 2012 and Yasutaka *et al*., 2013). These illustrations would imply that the reduction of air dose and contamination radiocaesium would depend on natural decay and redistribution of the element to the lower plain on which human and their activities are concentrated. Considering that the acceleration of air dose reduction or contamination reduction is not likely for long time, a long term and sustainable radiological monitoring in the whole region would be demanded.

1.1.2. Public Involvement in Radiation Monitoring

Immediately after the accident, Japan government focused on rescuing the survivor and handle the nuclear power crisis. The effort of radiation measurement came after the FDNPP accident and radioactive plume spread widely. The national and systematic radiation mapping project has just begun in June 2011. It has served the purpose of disaster impact assessment, radiation protection formulation, and countermeasures decisions (Saito & Onda, 2015). The measurements were done comprehensively through various methods, platform, and medium, at various scales and extent on various environmental conditions which provide not only a great depth and detail radiological information (such as ground measurement), but also general and extensive radiological information such as (airborne mapping) on various land uses (JAEA, 2014). The whole project was coordinated by MEXT (the ministry of the Education, Sports, Science) which assigned JAEA (the Japanese Atomic Energy Agency) to lead and manage the whole monitoring operation and cooperation (Saito & Onda, 2015).

A while after the explosion of Fukushima Dai-ichi Nuclear Power Plant, when the order for evacuation and necessary screening had been released, radiological information was not quickly and widely published by Japan Government for some reason (Onishi and Fackler, 2011). This attitude with the contrasting Japanese government vows on nuclear safety from previous policy cost the government with the loss of trust from Japanese citizen. A rather similar case was also reported in Three Mile Island accident (Gricar and Baratta, 1983) and indicated in Chernobyl accident (Renn, 1990).

Japanese government is not the only source or institution that providing the radiological information. In the absence of radiological information for the public during the FDNPP crisis, there were many initiatives led by a group of people or an organization to collect, aggregate, and summarize the radiation measurement from various source and publish the information (Brown *et al*., 2016). The data mostly were the air dose rate or ambient air dose rate and measured by various radiation survey meter.

In Fukushima, some of organizations that lead radiation data collection initiatives are persisting still. One of them is SAFECAST (http://blog.safecast.org/about/). As of July 2016, the SAFECAST radiation measurement database exponentially rose to 50 million records data points, which taken from not only in Japan but many places all around the world (SAFECAST, 2016). The measurement is done by volunteers and some SAFECAST staff. The radiation measurement and monitoring has been being conducted to this date. It can be concluded that SAFECAST initiative has been evolved from an immediate response to radiation disaster into long term monitoring over radioactive contaminated or surveillance on elevated radiation.

1.1.3. The Advantage of Public Involvement in Radiation Measurement

It is better to note that expansive measurement by citizen (crowdsourced radiation data collection) to the nuclear disaster affected areas was firstly initiated and conducted in Fukushima, Japan in relation to FDNPP disaster. Prior to FDNPP disaster, there were radiation measurement or monitoring had been done but rather performed in smaller scope or community scale. The example of this is what was done in Pennsylvania, USA after Three Mile Island Nuclear Generating Station No. 2 (TMI-2) accident (Gricar and Baratta, 1983) and Belarus (ex-Soviet Union) after Chernobyl Nuclear Power Plant (CNPP) accident (Lochard, 2007). The latter which was conducted under similar project like ETHOS (a kind of living rehabilitation of the survivor project) has been applied at some villages in Fukushima.

Up to now, there has no study about the advantages of crowdsourcing radiation data implementation for the whole participants that joined the initiative. On the individual basis, we believe that some benefits that obtained by individual through a volunteering activities of both initiatives as such (crowdsourced and rehabilitation program) might be similar for some extent, such as more confident on the measurement result or the measurement result was more acceptable than that measured by the authorities. Hence, that individual or community would be easily to move on with another commitment to alleviate the situation.

For a community radiation monitoring as mentioned above, a successful claim was reported like in Lochard (2007) through a rehabilitation project named ETHOS. Lochard (2015) in his communication about similar project which replicated in Fukushima argued that through a quite intense and direct engagement, the affected residence gained the ability, confident, and dignity to control their lives again. Note that, radiation monitoring was only a part of whole activities within the project, but a crucial one. Despite of the difference on scale of the activities, we believe that radiation monitoring in crowdsourcing and that rehabilitation program shared a common assumption, principles or values, such as engaging the citizen in any risk assessment program that enable them to widely participate (Shirabe *et al.*, 2015), people acceptance on their own measurement rather than the doubtful authorities (Gricar and Baratta, 1983), data transparency, and so forth.

1.2. Research Questions

While the history of radiation data collection initiatives during the crisis caused by FDNPP accident is great and inspiring, the data as a collective effort of lots of people is still need to be assessed before it further used in any decision of crisis or disaster management. The data and the method of data collection are the major points of assessment that addressed by the expert or scientists group toward non-expert group (such as Wiggins *et al.*, 2013, Bordogna *et al.*, 2014, Saito & Onda, 2015, Freitag *et al.*, 2016). On the other hand, the specific assessment on crowdsourcing on radiation data has never been done. To reap maximum benefit from data crowdsourcing, its assessment would be necessary not only for the public as their potential clients and scientists group that might be use their data in the future, but also to a fostering organization of crowdsourcing initiative.

With abundant data on hand which keeps growing in the future, some direct and classical questions addressed to crowdsourced data remain relevant to be asked that become themes of this dissertation:

(1) How true the crowdsourced data represent the fact (radiological situation)?
(2) How to improve the utility of crowdsourced data? Could we extend measurements on places that difficult to access by human such as forests?

1.3. The Structure of the Dissertation

The content of this dissertation follows the research questions mentioned above. Prior to the investigation of these problems, we discussed about the concept of data quality aspect of citizen science data in Chapter 2.

The first question will be addressed in Chapter 3. It will be very difficult to ask the volunteers to measure the air dose at certain location, while we are doing the most accurate measurement at the same time and place. Therefore, in this chapter instead of comparing with the most accurate independent using the advanced equipment available, we propose a method as such that citizen science data is compared with its parallel of scientist group data which acquired relatively in the same way and period, due to all citizen science group data were taken in the past. We used seven datasets scientist group data available and therefore seven equivalent datasets from citizen science group since 2011 until 2013.

In Chapter 4, we reported the use of aerial vehicle for mapping air dose rates above forest canopy using two kinds of detectors, that is GM Pancake type detector (bGeigie Nano) and scintillator type detector (Polimaster PM1703MO-1B). Beside comparing both detectors effectivity in mapping air dose rate on such complex natural object, we explored the nature of the data as a prerequisite for modeling linear relationship between air dose rates and tree (canopy) dimension.

Last but not the least, we described our work on fusioning the citizen science data and expert data for interpolating air dose rate value at no data location to prove that both data are complement to each other. The techniques used was backpropagation method of artificial neural network, a popular method in machine learning. It used two distinct scale datasets, that is citizen science data that measured on the ground and manned aerial vehicle. Several variables used in this method including the distance from surrounding known air dose value points to the no-data points which utilize both detail and coarser scale data, land cover types or the distance to the nearest land cover, slope and aspect of terrain.

The last chapter, we would like to emphasis some points highlighted in these studies followed by recommendation about the use of citizen science data.

CHAPTER 2. CROWDSOURCED DATA QUALITY

2.1. Crowdsourcing Related Terms and Definition

Many researchers perceived crowdsourcing as an approach or method on data acquisition, which is only a part of science production. It is frequently being linked with “citizen science” since crowdsourcing is done mostly by non-professionals (lay person) or what then called “citizen scientists”. In fact, the term “citizen science” was coined or get popularity almost a decade earlier than the term “crowdsourcing”.

The word “citizen science” was coined by Irwin (1995). Instead of formulating the definition by himself, he explored the multiple meanings of citizen science, i.e. science for the citizen and science by the citizen (Irwin, 1995; Pp. xi). The latter seems dominating in most of citizen projects in 21st century (Silvertown, 2009). The term “citizen science” itself seems a bit vague as if science depends upon the science performers. But what do scientists mean with the term is that real scientists professionals or experts working together with public in all or some extent to solve a scientific question (Cohn, 2008).

The other alternative term to citizen science is *crowdsourced science* (Toerpe, 2013) or *crowd science* (Franzoni and Sauermann, 2014) due to the participation came from a significant number of citizens or public, and so forth. The adoption of “crowd” or “crowdsourced” is because of the famous term “crowdsourcing” that coined by Jeff Howe (2006) to remark the Age of Internet that connect people around the world has changed the way how a company obtained a greater potential human labor to support them in running its businesses more efficiently. Toerpe (2013) was also used the term “public participation in science”, that is public action that take part in science method or activities, although it does not mean that any citizen science project would limits the participation. In fact, in many citizen science projects, it is hard to find a project in which the citizen does all the steps of scientific method. Therefore, Rossiter *et al*. (2015) proposed “citizen-assisted science” for more accurate term.

Bonney *et al*. (2009) proposed models of public participation in scientific research in which steps were commonly conducted by the professional scientists in doing their research. The models are illustrated in Table 2-1. Based on the models, they differentiate three categorizes of public participation in scientific research, i.e.: (1) Contributory projects, in which public primary role is collecting data; (2) Collaborative projects, in which the public steps further mainly from collecting samples or data collection to contribute in analysis; (3) Co-created projects is when public and scientist work altogether from the beginning to the end of a research project. The latter is called “extreme citizen science” by Haklay (2013) to give remark of the total participation of the public on a research project.

When it comes to the acquisition of geographic data or information, the term “volunteer geographic information” (VGI) or “crowdsourced geographic information” (CGI) was introduced (Sui *et al*., 2013). Geographic data or information is data or information about the attribute (nature) of a place which has an earth-referenced location. Beside the internet, the advanced of sensor technologies to create representation of the real world as well as other information and communication technology advancement would make individuals who have access to these technologies able to contribute.

Table 2-1. Public participation in science which can be classified into three categories, based on the scope of participation *i.e.* contributory, collaborative, and co-created project participations (Bonney *et al*., 2009)

Steps in Scientific Process	**Steps included in Contributory Projects**	**Steps included in Collaborative Projects**	**Steps included in Co-created Projects**
Choose/define research question			○
Gather information and resources			○
Develop hypotheses (explanations)			○
Design data collection methodologies		△	○
Collect samples and/or record data	○	○	○
Analyze samples		○	○
Analyze data	△	○	○
Interpret analysis result and make conclusion		△	○
Disseminate conclusions/translate results into action	△	△	○
Discuss results and ask new questions			○

○ = Public included in the steps; △ = public sometimes included in the steps

2.2. Crowdsourced Data Quality

Crowdsourced data or citizen-assisted science data quality is not new issue to data scientists. Senaratne *et al*. (2017) found hundreds of articles discussing about the assessment of the data quality, only of crowdsourced geospatial data or what they called as Volunteer Geographic Information (VGI). Especially for geospatial data, ignoring the quality aspect of such data would give undesired result to the data user (Congalton and Green, 1957).

Many scientists put a lot of their interest in crowdsourced data because the volume of the data that potentially could be obtained is so massive. It could be obtained by layperson using a variety of method and equipment. It is no surprising that many cases of crowdsourced data become of what so called big data. Big data is indicated by huge volume, high rate of acquisition speed, high variety (such as types, formats, structures, and so forth) of data. In order to get the maximum benefit from this type of data, scientists need to assessed the quality of the data and compulsorily preprocessed the data before further used in analysis.

There are ways for valuing the quality of crowdsourced geospatial data. Senaratne *et al.* (2017) compiled the list of quality measures formulated by ISO/TC 211 (International Standardized Organization/Technical Committee) with other measures from various researchers. They also enlisted some practical or less analytical quality indicators that enroll as a proxy of quality measure. In addition to those quality expressions, they also made a list of approaches and methods for assessing the quality of the data which can be seen in Table 2-2. For more detail about the measures and indicator, the reader is suggested to peruse the work of Senaratne *et al*. (2017).

Table 2-2. Approaches and methods for estimating quality measures and indicator (modified from Senaratne *et al*., 2017)

Approach	**Methods**
Geographic	Compare with reference data
	Line of sight
	Formal specifications
	Semantic consistency check
	Geometrical analysis
	Intrinsic data check
	Integrity constraints
	Automatic tag recommendation
	Geographic proximity
	Time between observations
	Automatic scale capturing
	Geographic familiarity
Social	Manual inspection
	Manual inspection/annotation
	Manual annotation
	Comparing limitation with previous evaluation
	Linguistic decision making
	Meta-data analysis
	Tokens achieved, peer reviewing
Crowdsourcing	Applying Linus law
Data Mining	Possibilistic truth value
	Cluster analysis
	Latent class analysis
	Correlation statistics
	Automatic detection of outliers
	Regression analysis
	Supervised classification
	Feature classification
	Provenance vocabulary
	Heuristic metrics/fuzzy logic

Senaratne *et al*. (2017) noted that data mining has been used by many researchers to value the quality of crowdsourced data. Data mining works by looking out the pattern from within the data. It can assess the quality of the geospatial data from the general pattern of feature attribute values and use it to solve the problem such as noise by filtering the values (data cleaning). In this way, the geographic data is not necessarily being used in the filtering process.

CHAPTER 3. CITIZEN SCIENCE AND EXPERT-GROUP RADIATION MONITORING DATA COMPARISON

3.1. Introduction

The Fukushima Dai-ichi Nuclear Power Plant accident was a disaster event that notoriously known for governance misconduct by keeping the critical information for the public to know in several days after the accident (New York Times, 2011). During the crisis, the radiation detector was sold out and people had no mean to know the air dose rate of their surroundings (Democracy Now, 2014). Seeing this situation, a group of people tried to initiate large data collection by developing a very simple and handy radiation detector to build by a layman or volunteers, then use the detector for measure radiation, and upload the measurement data (crowd radiation data) to a designated server from which publicly online map showing the air dose situation is generated (SAFECAST, 2016). In the context of disaster management, voluntary mapping potentially be another channel for public to get information about their environment, since there is no one responsible to initiate this movement.

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 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 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 particularly due to 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.

3.2. Methodology

3.2.1. Citizen Science 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).

3.2.2. 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.

3.2.3. Unit of Observation

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.

Table 3-1. KURAMA dataset and its associated SAFECAST dataset used in this study

KURAMA survey	KURAMA's Acquisition Date	Associated 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

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 by a unique code. When there are two or more measurements existed in a grid, the air dose rate values were averaged.

3.2.4. 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.

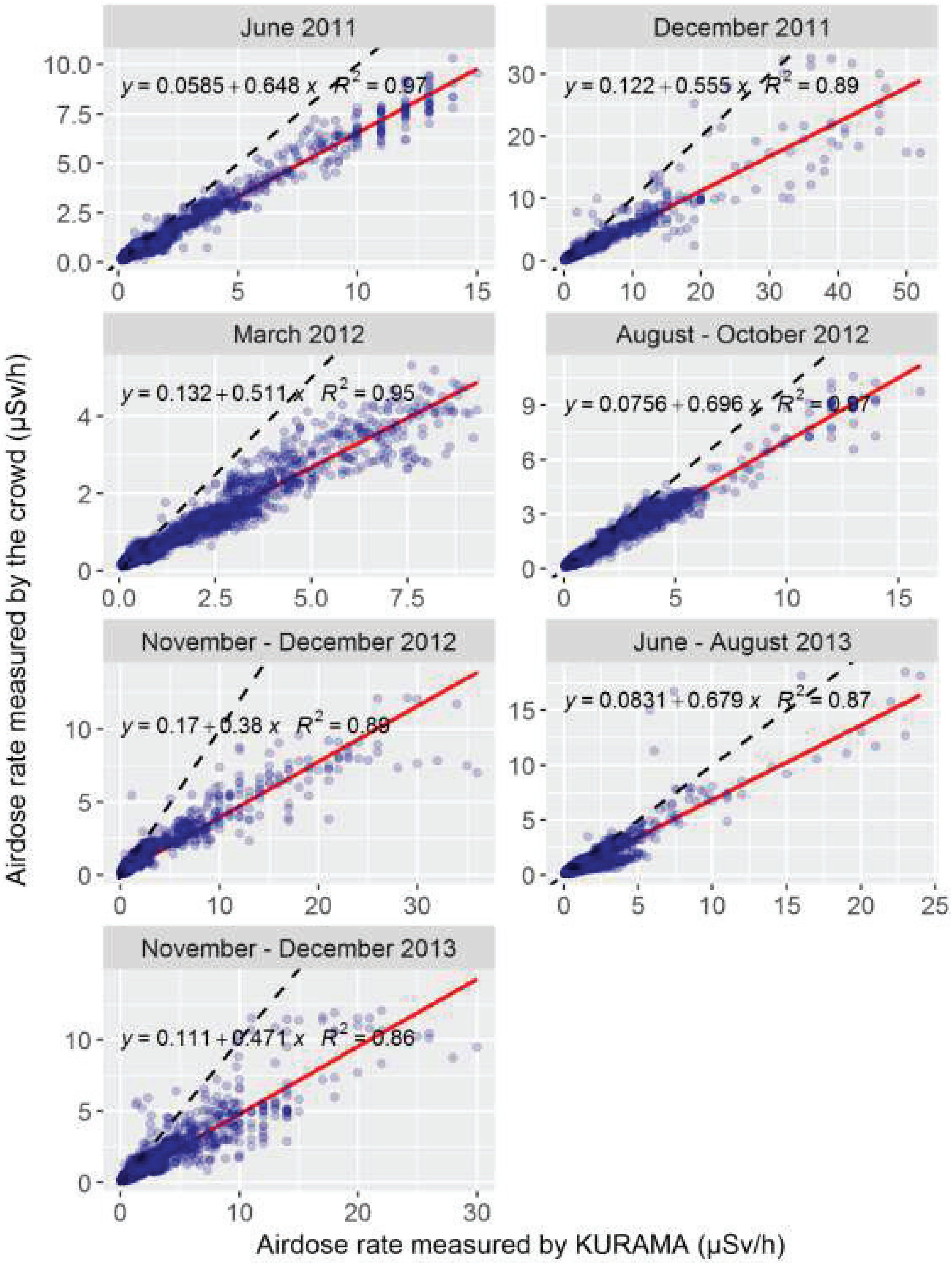


Figure 3-1. Correlation diagram of the air dose rate of non-scientist group data (SAFECAST) to scientist group data (KURAMA) which collected in a parallel acquisition time. The figure is composed by seven panels which indicates the survey time of KURAMA system.

3.3. Results

Figure 3-1 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. It illustrates that in all survey periods, the citizen-scientist group 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 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. Figure 3-2 and Figure 3-3 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 Figure 3-2 (the first row and first column of the figure) of the air dose rate from the second KURAMA survey had decreased by about 32% in comparison to the first KURAMA survey.

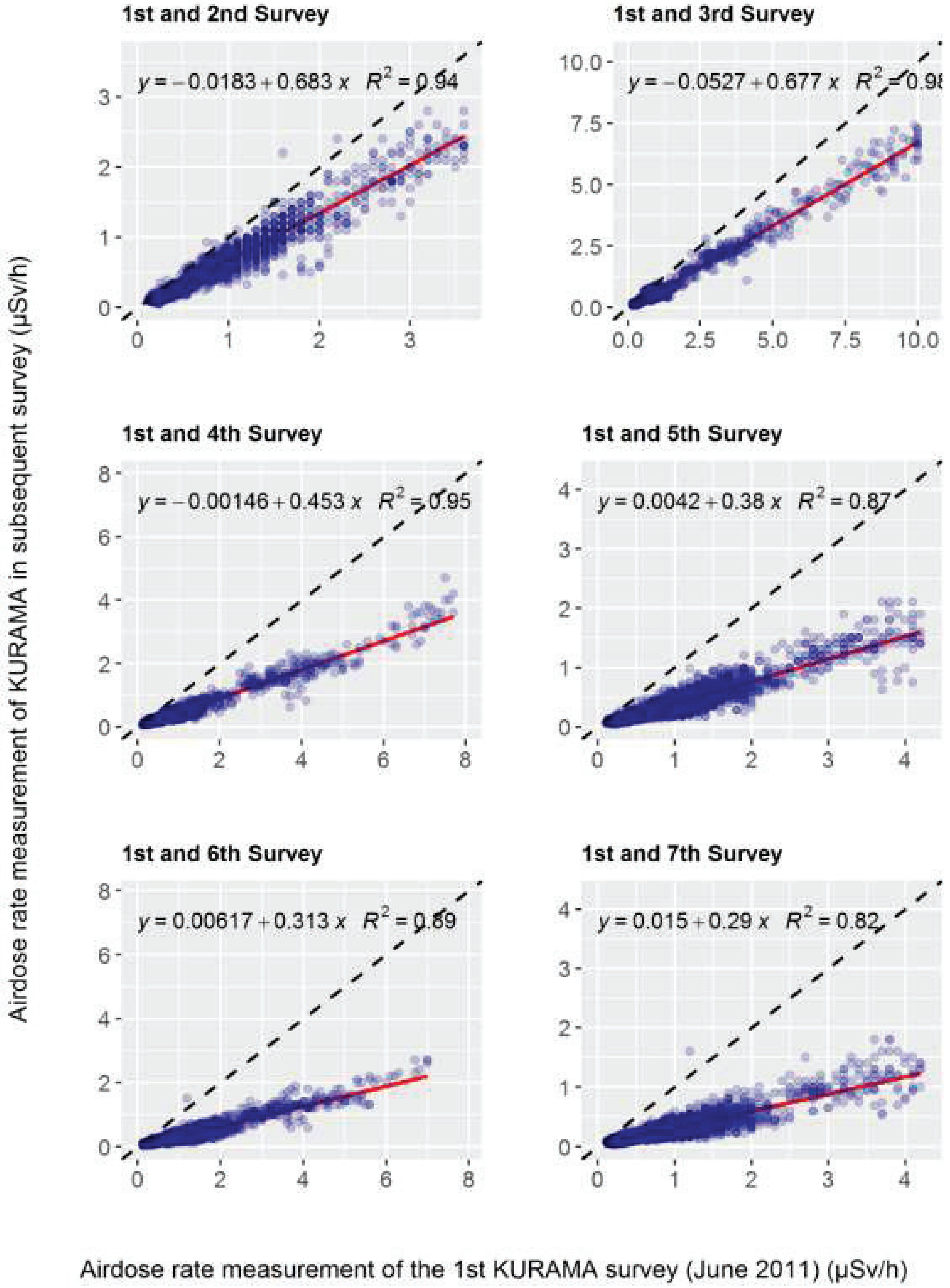


Figure 3-2. Correlation diagram of air dose rates between the first survey with the subsequent survey of the expert group.

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 32%, the fourth by 45%, the fifth by 62%, the sixth by 69%, and the seventh by 71%. During the same period, as illustrated by the first panel of Figure 3-3, the percentage by which the second period of SAFECAST measurements had decreased compared to the first measurement period was 41%. 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 51%, the fourth by 46%, the fifth by 65%, the sixth by 64%, and the seventh by 73%.

Based on Figure 3-4, 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.

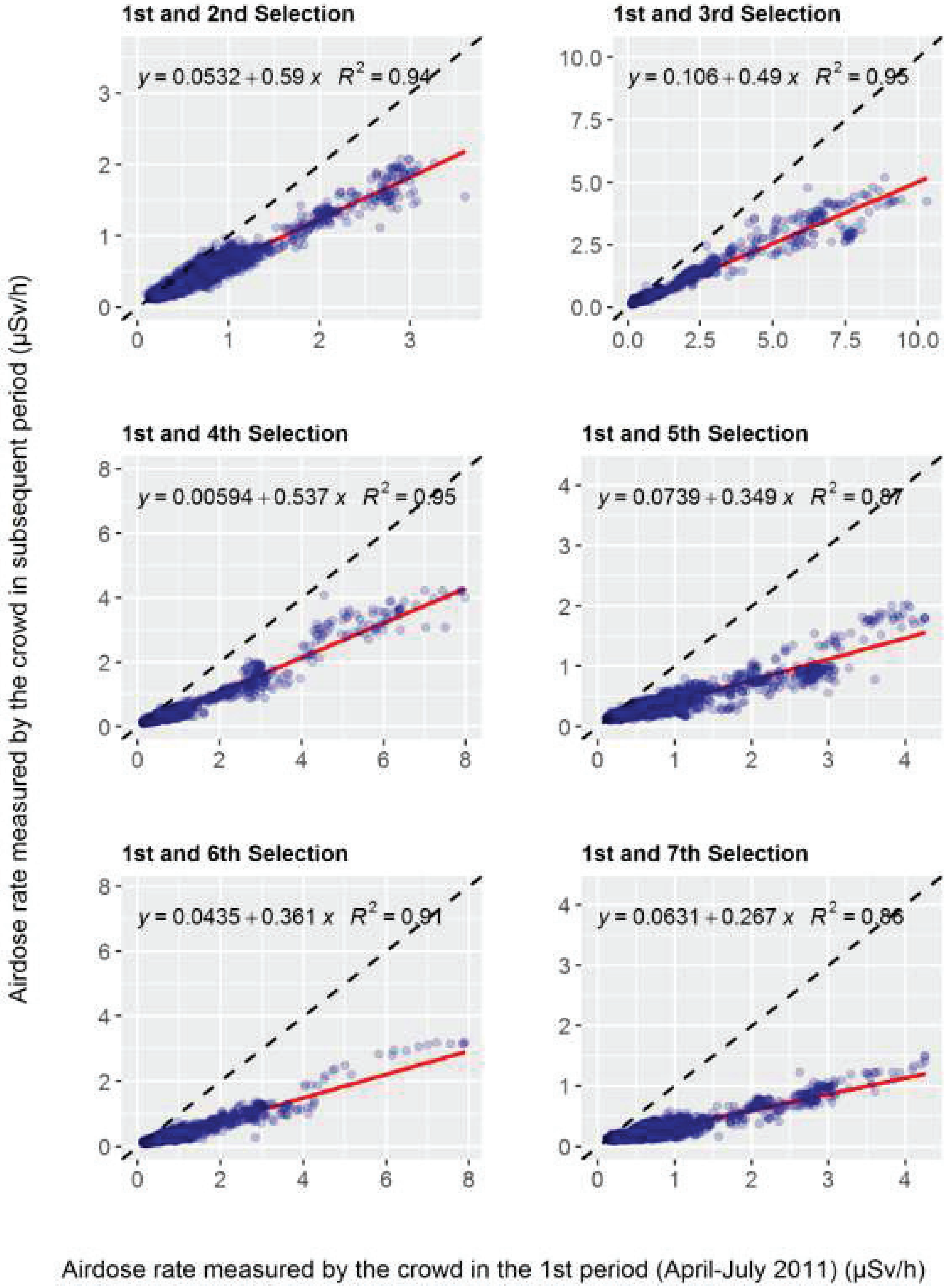


Figure 3-3. Correlation Diagram of the Air Dose Rates Between the First Selected Measurement Period to the Subsequent Period of the Citizen Science Group.

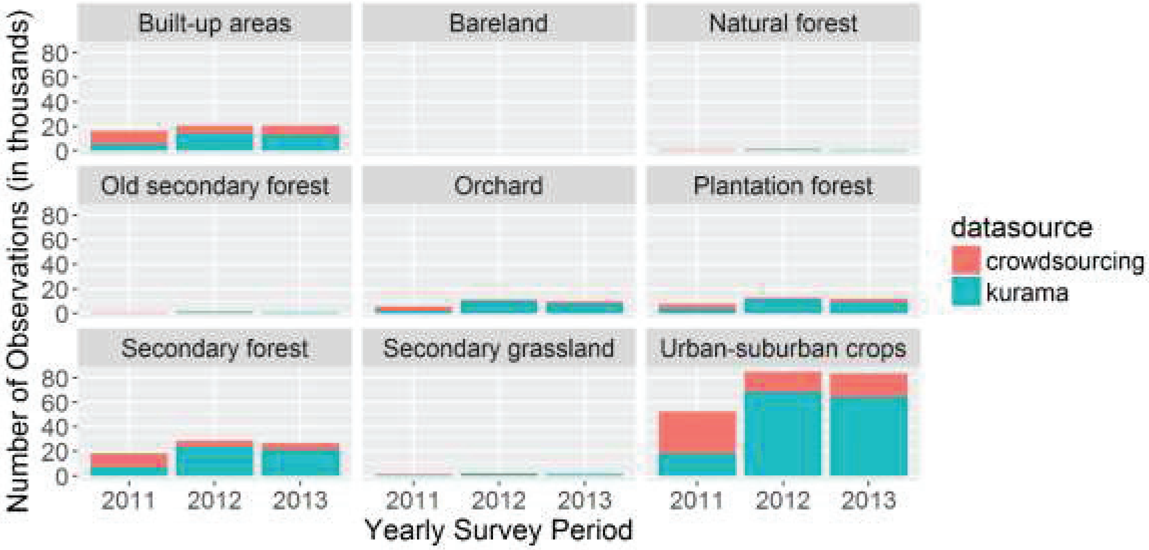


Figure 3-4. The Number of Observation between Crowdsourced and Expert Approach at Several Land Use Types in Three Different Time Acquisition: (A) 2011, (B) 2012, and (C) 2013. An observation means a measurement conducted at a certain place or a grid, since the whole study area was represented into grid that follows National Grid Square Framework of Japan. Each grid was counted only once.

3.4. Discussion

3.4.1. Agreement in Dose Rate Measurement

The large number of observations in the lower range of air dose rates shown in Figure 3-1 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 Figure 3-1 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 30% – 60% 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 from snow that affected measurements were already removed in expert measurements database (JAEA, 2014). Since similar effort has not done yet to citizen science group data, it is likely that snow shielding effect appeared as a lower slope values in the regression lines.

Both citizen science group and expert group data include natural background radiations. The natural background radiations may include radiation from terrestrial sources such as uranium, thorium and radium and extraterrestrial object such as cosmic rays. The detector used by expert group was not designed to detect cosmic rays, therefore cosmic rays might also influence the discrepancy of air dose rates values of both measurement approaches. The intercepts values in Figure 3-1 ranging from 0.06 to 0.17 might reflect the influence of cosmic ray to the citizen science group data. In the highly contaminated areas, the contribution of total background radiation to the air dose rate measurement values is not substantial and so is the cosmic rays. On the other hand, they become significant when the measurement is undertaken in the low to very low contaminated areas or in the normal condition.

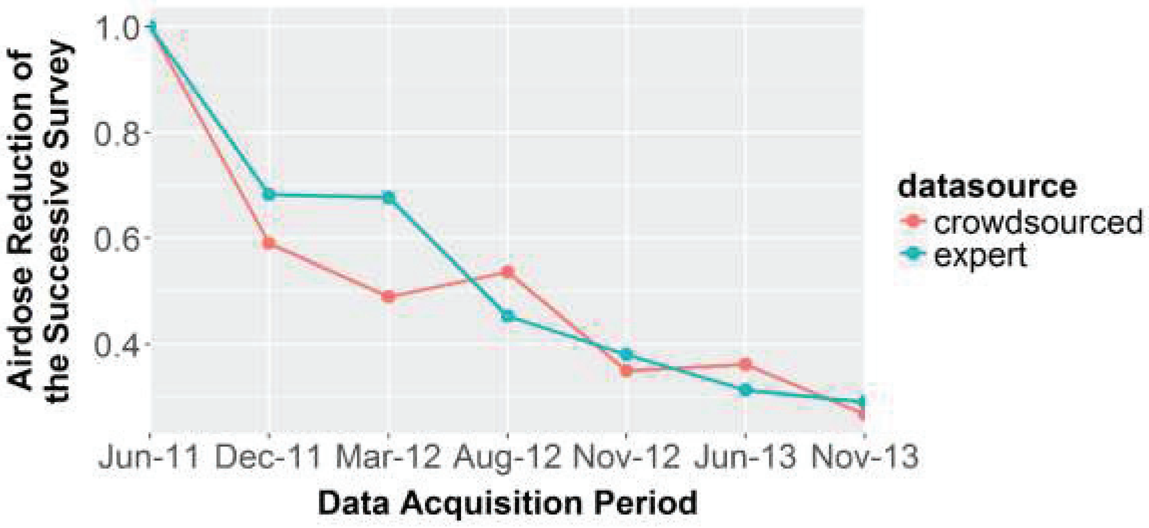


Figure 3-5. The Trend of Air Dose Rate Reduction throughout the Survey Period (2011-2013) which Taken from the Slope of Regression Model Between the First Survey (June 2011) and the Second Survey (December 2011), between the First Survey and the Third Survey (March 2012), the Fourth (August 2012 – October 2012), the Fifth (November 2012 – December 2012), the Sixth (June 2013 – August 2013) and the Seventh (November 2013 – December 2013)

3.4.2. The 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 Figure 3-2 and Figure 3-3, respectively. We took the slopes from Figure 3-2 and Figure 3-3, and used them to see their performance in the course of the survey periods, as illustrated in Figure 3-5. 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 up to 18%. We believe that this difference came about because of detector characteristics and measurement outcomes.

3.5. 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 30% up to 60% 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 18%. 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 of which data are lacking due to government resource limitations or because of national monitoring was ceased already.

CHAPTER 4. SCALING UP CITIZEN SCIENCE DATA BY AERIAL VEHICLE PLATFORM

4.1. Introduction

The citizen science in Japan was remarkably recognized after Fukushima Dai-ichi Nuclear Power Station (FDNPS) by their agility actions in organizing themselves and developing a comprehensive system for radiation monitoring in the environment, feeding and updating the public with the radiological situation in Fukushima. One of their centrality was the design and making of a compact and affordable radiation detector as it was scarce at the time when people needed it most. The detector was designed in such that laypeople with a novice soldering skill would easily to assemble the detector from its electronic parts into a working detector. The current prototype has a much more convenient design since the parts are possible to be assembled without soldering. By the availability and deployment of this detector in significant amount during the crisis, a large volume of radiation data could be collected. Thereafter, the required information generated from those data were fed back to the public while such information was concealed by the government for some time.

Radiation data are still continuously being collected by the volunteers but limited to the accessible places like on the road or of urban and sub-urban areas. The data measured in land uses that remote from the populated places or have natural objects such forests are almost not available. One way to overcome the limitation of human mobility in difficult places like forest is by deploying aerial vehicle which bring along the radiation detector.

The civilian's Unmanned Aerial Vehicle (UAV) is now becoming handier, user-friendlier, affordable for hobbyists. Various UAV of this types have been used for monitoring the contaminated areas by the professional experts. To address the limitation of mobility of citizen science, we conducted a study of scaling up measurement using citizen science detector coupled with UAV on the forest to evaluate whether there was a significant associative pattern between the measurement of airdose rate on the ground to the measurement on the UAV. More detail about data collection is provided in the following section.

4.2. Methodology

4.2.1. Study Area and Observation Plot

The study was conducted in southern part of Iitate village near to its administrative border (Figure 4-1). In a few hundred meter, there was a gateway to the no-stay zone due to high air dose rate and contamination. In general, the study area consisted of some agricultural areas dominated by pasture land and planted trees. Near to site study, there were many a heavy-duty machine that seemed work on decontamination. The area of our study site would be decontaminated soon as informed by the land owner.

Some trees that could be distinguished in this site were deciduous trees, mix deciduous, and conifer tree patches. We investigated about 100 m x 100 m plot on deciduous trees in which the air dose rate on 1 m above the ground was measured. Above the forest canopy of this plot, the air dose rate was measured by using an UAV.

4.2.2. Data Collection Method

4.2.2.1 *Radiation Measurement using UAV*

Two radiation detector types used for collecting air dose data were bGeigie Nano and Polimaster PM1703MO-1B. bGeigie Nano is a compact radiation detector consists of a pan-cake GM counter, flash card for storing the data, GPS and Bluetooth system and device for convenient visualization with the smart phone. These features would enable the integration with the UAV without significant difficulty.

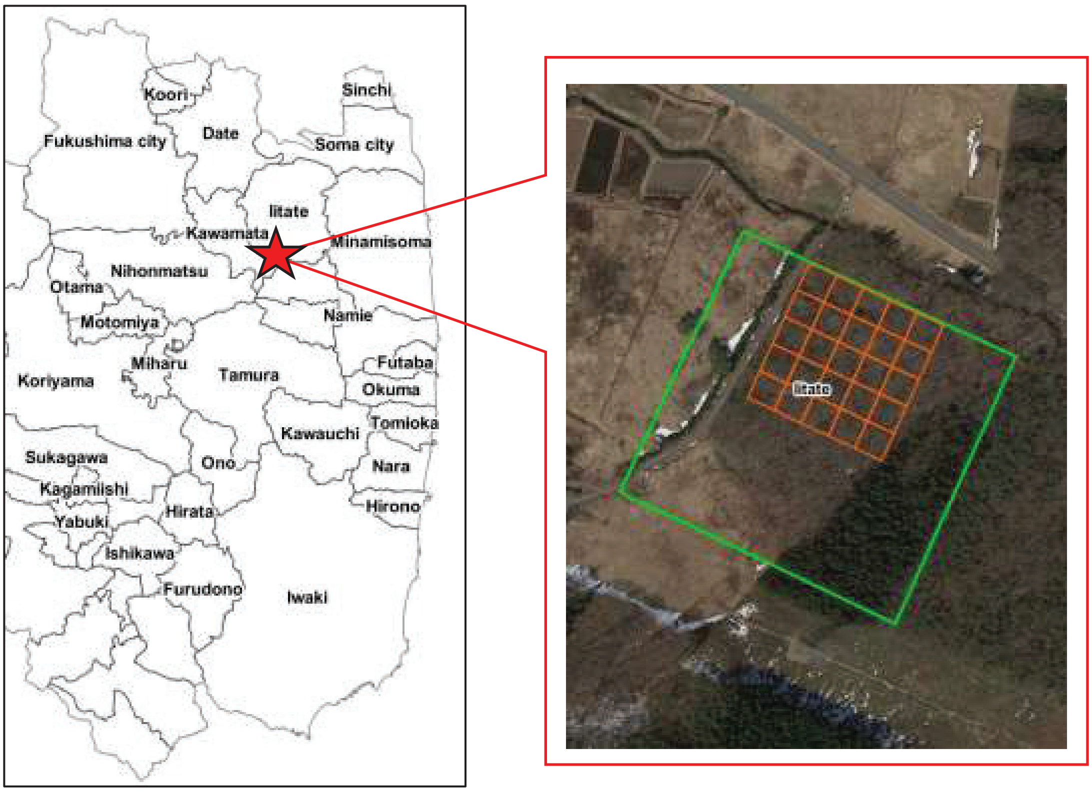


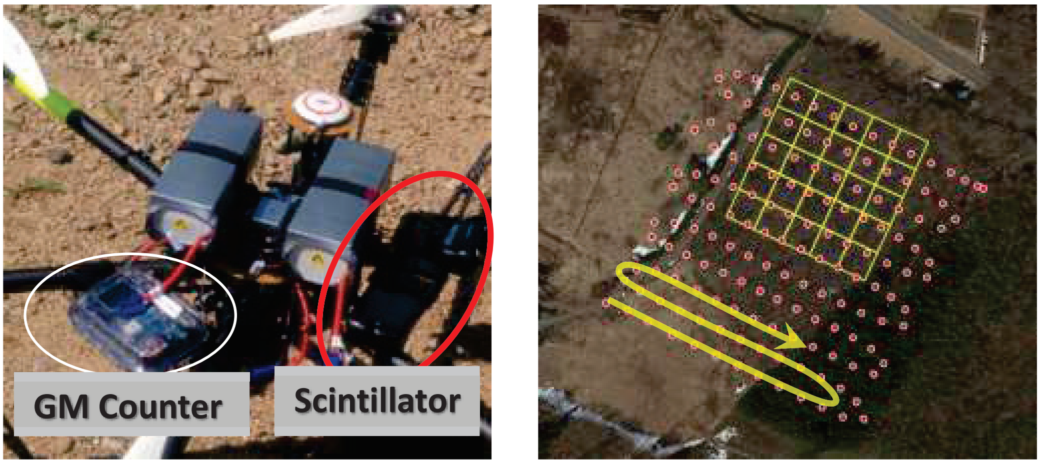
Figure 4-1. Study Site in the Iitate Village. The area bounded by green box was area of interest for comparing the air dose measurements from different detectors using UAV. The plot shown by orange grid mesh is a plot in deciduous tree patch.

Polimaster PM1703MO-1B is a commercial detector which uses a CsI(Tl) material and relatively more sensitive gamma radiation counter than the previous one. Since this detector is not equipped with data logger, the smartphone together with GPS was used as a data logger device. The geographic and air dose data were emitted to the smartphone through a Bluetooth connection and stored to the memory card. These detectors were set together on the UAV and measure the radiation for every flight. The appearance of a prepared UAV before flight was conducted shown by Figure 4-2.

The UAV moved along the comb-shape like paths with the interval distance of about 15 m over the study area. The moving speed of the UAV was 2-3 m/s. The altitude was set up to 40 m above the ground. We cooperated with a private company for handling the UAV to ensure the flight was conducted safely and in a more consistent way. The company was also requested to provide a very accurate elevation data using LIDAR technology and system.

4.2.2.2 *Radiation Measurement on the Ground*

The air dose rate in the deciduous forest at 1 m above the ground was also measured by using bGeigie Nano. The measurement was conducted within the designated plot as shown in Figure 4-1. There were two methods of measurement were done, i.e. stationary and wandering measurements. Stationary measurement was done by staying on the fix position within each subplot for 1 minute and the dose rate value was recorded or written down. Whereas wandering measurement was done by walking and visiting each subplot while the detector system recorded all dose rate values that counted along the track.



(a) (b)

Figure 4-2. The design of research (a) UAV and radiation detectors (b) flight route of UAV

4.2.3. Data Analysis

4.2.3.1 *Autocorrelation*

The data that provided by all radiation detectors are spatial data, i.e. data that have geopositional index relative to the earth surface. By this regard and the fact that all detectors have its own sensitivity to detect radiation which influences the required time to measure for a precise air dose estimation, we suspected that the data hampered by autocorrelation particularly to the data obtained through UAV. We used the autocorrelation function (*ACF*) to detect the existence of autocorrelation in the data from both detectors. For the implementation, we used the available functions or libraries from R Statistical Software's online repository.

4.2.3.2 *Comparison of Survey Methods in Air Dose Rate Mapping*

The original or unprocessed data provided by the detector devices are point type data, means that one air dose rate value will refer to a point on the earth which created for each 5s of bGeigie and 10s of Polimaster detector. Therefore, when the data projected into a planar surface, there will be a gap from one point to another. To grasp the whole radiological situation in the forest, the air dose rates values at the gaps between points will be interpolated. By this way, the spatial pattern of air dose rates would be obtained and it would be relatively easy to compare the air dose rates measurement from different detector and survey modes visually. The method that used for interpolation was ordinary kriging and it was implemented in the R statistical software environment (R Core Team, 2017) with a special statistical module (computing function library) for geostatistics (Pebesma, 2004).

The evaluation of air dose information provided by UAV data was also done using correlation analysis to get a quantitative impression about how the pattern shown by UAV data would relate with the pattern of ground data. The ground data that used in this evaluation was the ground data obtained through wandering measurement.

4.2.4. Result

4.2.4.1 *Autocorrelation*

Figure 4-3 showed that the autocorrelation exists on air dose rates measured by both GM counter and scintillator type detectors on the UAV. It is obvious that the level of autocorrelation on GM counter was higher than the scintillator one by looking at the number of bars that crossing the autocorrelation threshold line. The GM counter data was correlated to each other all the way from the beginning to the end of data recording, although the correlation strength looked decreasing.

The time lag that shown in the figure is related with the time taken for each measurement or recording time. bGeigie was set to record the estimated dose rate for every 5s while Polimaster was every 10s. The time lag of scintillation detector when no significant autocorrelation happened was 3 bars (by neglecting the first bar) means that it will need more than 30s to have an independent observation.

4.2.4.2 *Air Dose Rate Surface Model*

The result of interpolation to all survey data with different methods and detectors can be seen in the Figure 4-4. It can be distinguished visually that the ground measurement has a higher range of air dose rate values than the ones that measured on the UAV. The range of air dose rate values obtained from UAV with GM counter detector was the lowest one. The direction of UAV also affected the measurement as it could be discerned from the difference Figure 4-4 panel C with panel D and panel E with panel F.

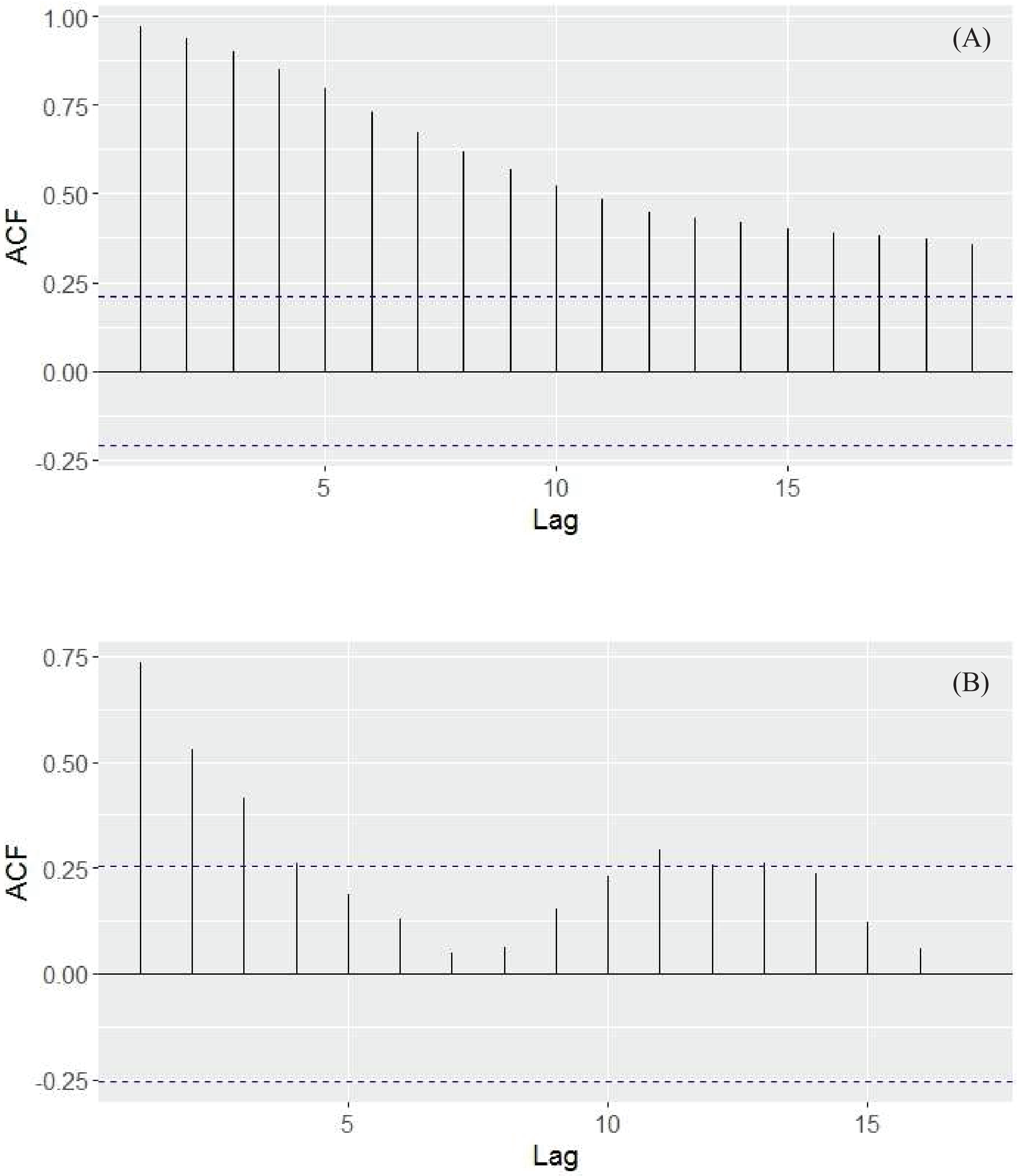


Figure 4-3. Auto-correlogram of the Data Measured by GM Counter and Scintillator on the UAV. ACF is the function that shows the level of correlation of the dataset to itself. Lag is the sequence of observation. Figure in the panel (A) shows a significant dependency of every air dose rate value to all measurements which conducted by UAV with GM counter. While a figure in panel (B) shows that an air dose rate values that measured by scintillator on UAV was significantly dependent to the next three observations (air dose rate values).

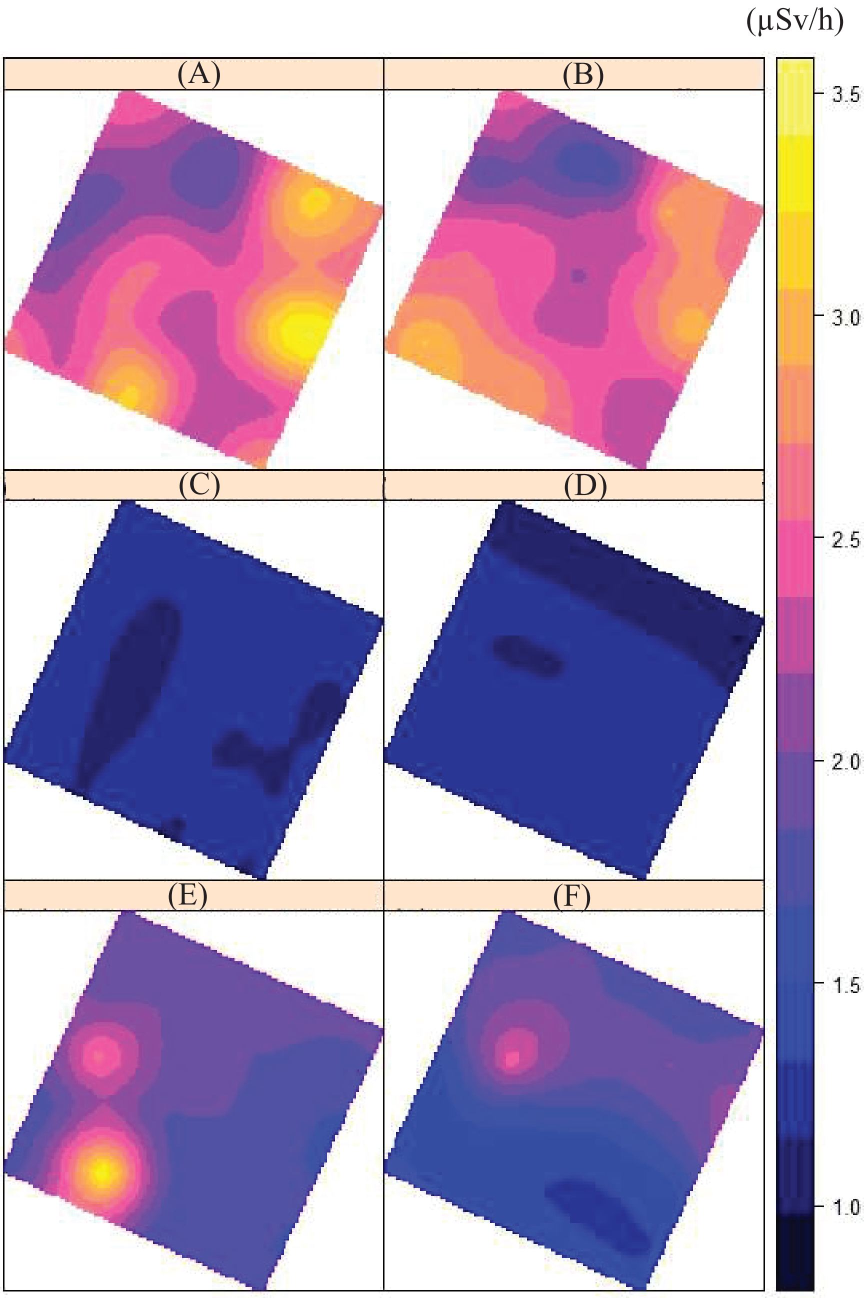


Figure 4-4. The surface model that generated by ordinary kriging interpolation to all survey modes (ground and aerial measurements). Figure shown in the panel (A) is a surface model of air dose rates that measured through ground measurement with stationary approach, panel (B) is surface model of air dose that measured through the UAV with stationary approach, figure in the panel (C) and (D) are air surface model of air dose rates that measured through UAV with GM Counter type detector which flew from North to South and East to West respectively, and panel (E) and (F) are figure of surface model of air dose rate measured through similar method as previous two panels but the detector used was scintillator detector.

Figure 4-5 showed that the linear relationship model between air dose rate measured on UAV and the data measured on the ground under the deciduous trees were very weak and had almost no sensible associated pattern. The variance that can be explained by the models were so small and even some models showed a contrary relationship with a negative regression coefficient. There were a contrast model parameters between the data from different UAV flight direction.

A quite comparable air dose rate surface pattern was shown between one that obtained through measuring for a while at fix position (Figure 4-4A) and one that obtained through measurement while walking throughout the plot (Figure 4-4B). The statistics of 30 random points that distributed over both air dose rate surfaces were close as shown in Table 4-1. From the Figure 4-6 both methods gave a relatively good correlation air dose rate (n=30, $p < 0.001$).

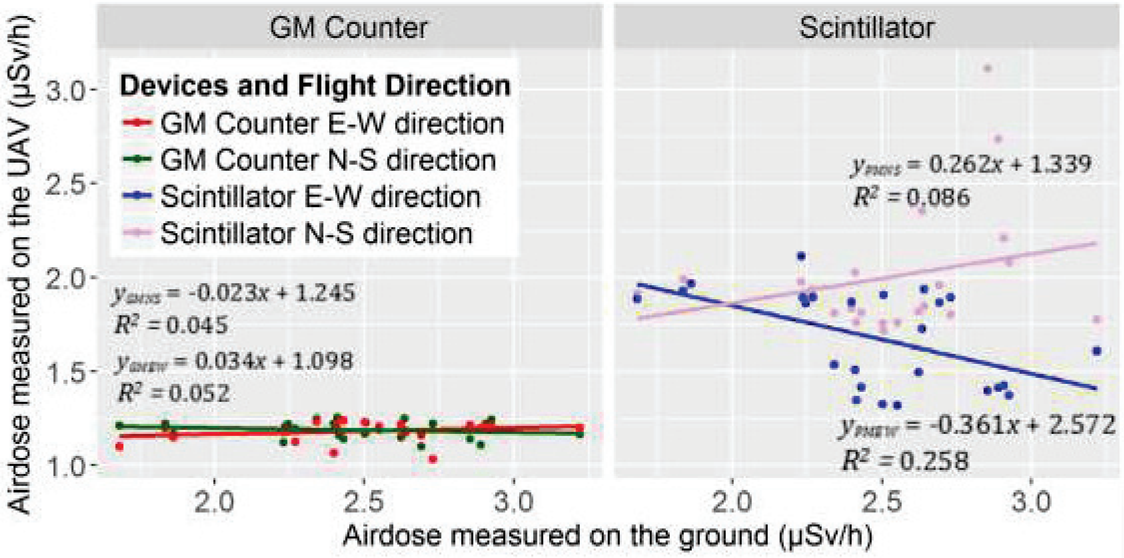


Figure 4-5. Linear relationship between the UAV data from both detectors its flight direction to the ground data with wandering measurement approach.

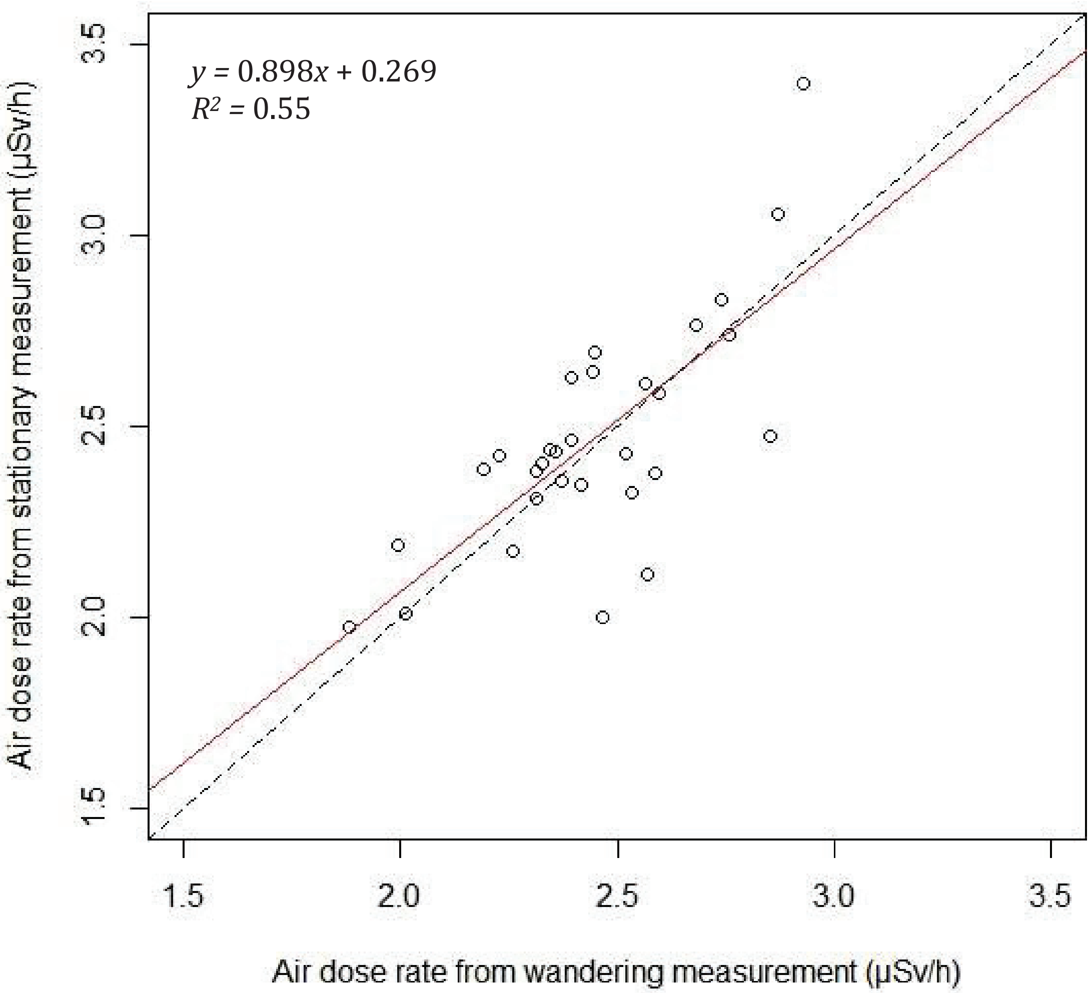


Figure 4-6. Scatter plot of air dose rate measured at fix points and air dose rate obtained through wandering approach.

Table 4-1. Statistics of air dose rate of two different ground measurement methods

Statistical Parameter	Stationary Measurements	Wandering Measurement
Mean	2.465	2.445
Median	2.424	2.430
Standard Deviation	0.305	0.252
Maximum	3.40	2.929
Minimum	1.974	1.884

4.3. Discussion

4.3.1. Autocorrelation

The assumption about the existence of autocorrelation in the UAV data was appeared to be true. As we shortly described earlier, this could be happened when the detectors are moving from one place to another which potentially have different radiation intensity or emission level. While the detector system is still collecting enough counts (detected photon incidence) to estimate air dose, the position of the detector changed to another which gave another incoming count. Therefore, the estimation of air dose rate at certain position was influenced by the counts from previous place. Another possible cause of autocorrelation is that of measuring air dose rate far above the ground, which implies a larger radiation field that could influence the measurement than if measuring the radiation on the ground. Sanada *et al*. (2014) indicated that the radius of radiation field would be equivalent with the height of platform on which the detector was mounted.

4.3.2. Comparison of Survey Methods in Air Dose Rate Mapping

The interpolated air dose rate values that depicted in Figure 4-4 shows that both detectors are not equal in providing radiation information through a mobile mode survey at the altitude using UAV. In general, the air dose rate values measured by bGeigie or GM counter type detector had a lower value at all positions in the observation of interest area. Such phenomenon has been emphasized in another study that comparing the mobile survey data from the same GM counter detector and the data from modified large scintillator (Nursal, 2017). In this study, both detectors were directly exposed to the radiation. Note that this was not the case of GM counter only when the UAV measurement is compared with ground measurement, as scintillation type detector that used in this study also shown the same tendency of low air dose rate values (Figure 4-4E and Figure 4-4F). Isaakson & Rääf (2017) stated that autocorrelation maybe be indicated by a low discrepancy of two areas nearby in the detector than in reality. With this assumption, it is possible that the highest air dose rate values were trimmed out or averaged by the detector with the previous measurement values.

Figure 4-4D – Figure 4-4F shows that the data from both detectors that measured with mobile mode did not follow the pattern of actual air dose rate values on the ground of the plot in deciduous forest patch. It was also confirmed by the Figure 4-5 that showed no significant relationship between measurement on the ground and on the UAV. The figure was further indicated that direction of UAV flight influenced the performance of measurement as well. It is not yet understood why the mobile measurement could be influenced by the direction of the platform. Nevertheless, all those factors mentioned above which argued to be the factors that cause disagreement between ground and UAV measurement would add more complexities in estimating air dose from the altitude using a mobile platform. In several personal communications, Sanada (2015) mentioned that topography and biomass were complex factors that need to be considered in the estimation. Hence, in the meantime, the use of simple and light using a mobile platform such as UAV would not be useful to expand the collection of air dose rate data into the complex land cover like forests.

The wandering method in measuring air dose rate 1 m above the ground seemed to have similar pattern of stationary measurement method in some extent. Some variation cannot be explained by the linear model as depicted in Figure 4-6, that might be due to interpolation effect. Another way to see how similar both method in providing air dose information is by looking the statistical parameters provided in Table 4-1. Based on the information which the table provided, it can be argued that both methods generate a quite similar radiological situation under the deciduous trees. The previous study about different mobility in ground measurement was done by Atarashi-Andoh *et al*. (2015) using a much more sophisticated scintillation detectors. They found that wandering measurement would produce similar information with the stationary method, but the first method got the advantage of more efficient in collecting the data than the latter one. This information would encourage citizen scientists or citizen science organization to collect by or to keep continue using the wandering measurement method.

4.4. Conclusion

Citizen science data was successful in providing radiation data by deploying many dedicated volunteers to measure air dose rate in the environment. As a detector platform, it is normal that a person would avoid the complexity in the landscape. This study shows that extending the radiation data collection using mobile survey or specifically UAV to overcome the complex elements of landscape will remain challenging. It is not only merely because of the predictability of GM counter data on ground air dose in the forest, but also due to the nature of the data itself.

CHAPTER 5. GENERAL CONCLUSION AND RECOMMENDATION

5.1. GENERAL CONCLUSION

Based on the typology of citizen-assisted science classification of Haklay (2013), SAFECAST as a single leading organization in the initiatives can be regarded as an *extreme citizen scientist*. On the other hand, if the initiative was referring not only to the initiator but also the supporting volunteers, the project can be regarded as a contributory citizen-assisted science, that is the contribution from the wide citizen was mainly on collecting and providing the data to the leading organization.

In the aspect of data quality, we have revealed that there was a well the agreement between crowdsourced or non-expert group data and expert group data. The discrepancy of air doe rate values exists where crowdsourced data consistently showed having lower air dose rate values than the expert group data significantly, therefore the release of map based on solely crowdsourced data needs to be accompanied with supplemental information explaining that discrepancy. Extending the capability of data collection to remote area on the complex terrain such as forests using a moving platform and citizen designed detector is not something that would be achieved by the citizen science group in the near time. Our study shows that the data came out from the detector gave a misrepresented air dose on the ground.

This study shows the advantage of public involvement in providing radiation information to the public. Citizen science group which organized within SAFECAST showed an agile response in the crisis of nuclear disaster. Nevertheless, as shown through this study there are some limitation particularly on the data collection which affected to the quality of the data. Since the data and data collection are the integral part of their work, the problems behind the limitations need to be addressed and solved soon.

5.2. RECOMMENDATION

Some recommendations that tenderly be proposed including the area within and beyond this study with supporting arguments are in the following:

1) Although some discrepancy exists between the expert data with crowdsourcing data by the citizen scientists, some extent of agreement between both data also exist. Another aspect of citizen science initiatives to be considered is the level of agility and prolific in providing the radiological information during the crisis. Therefore, we suggest the experts or scientists group to create a creative cooperation with the citizen science group, to produce mutual benefits by tapping a big number of talents (volunteers) and at the same time improve the capacity of the citizen science groups.
2) The area of cooperation between scientist and citizen science group or even with larger groups are considerably numerous. Citizen science groups and experts group perhaps could sit together and discuss how to improve their detectors that able to get more sample rates, how to make the data become more useful for scientists and other citizens, and so forth.
3) The citizen scientists group may expand their campaign to the specific group of people that may collect the data in the complex terrain, such as farmers that persistently live in their contaminated land. We strongly suggest that the cooperation with the specific group would not only a matter of data only, but rather beyond the data collection and the information it generates. In this case, it does not mean that the citizen science group that specialized on technology side does all the job, but rather to cooperate with other

groups that may have another program that beneficial to the farmers or radiation disaster affected people in general.

REFERENCES

Andoh, M., Nakahara, Y., Tsuda, S., Yoshida, T., Matsuda, N., Takahashi, F., Mikami, S., Kinouchi, N., Sato, T., Tanigaki, M., & Takamiya, K. (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.

Atarashi-Andoh, M., Koarashi, J., Takeuchi, E., Tsuduki, K., Nishimura, S. and Matsunaga, T., 2015. Catchment-scale distribution of radiocesium air dose rate in a mountainous deciduous forest and its relation to topography. *Journal of Environmental Radioactivity*, *147*, 1-7.

Bonney, R., Ballard, H., Jordan, R., McCallie, E., Phillips, T., Shirk, J., & Wilderman, C. C. (2009). Public Participation in Scientific Research: Defining the Field and Assessing Its Potential for Informal Science Education. A CAISE Inquiry Group Report. *Online Submission*.

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, pp. 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, pp. 1–29.

Chica-Olmo, J., 2009. Prediction of housing location price by a multivariate spatial method: cokriging. *Journal of Real Estate Research*.

Cohn, J. P. (2008). Citizen science: Can volunteers do real research?. *BioScience*, 58(3), 192-197.

Congalton, R. G., & Green, K. (2008). *Assessing the accuracy of remotely sensed data: principles and practices*. CRC press.

Franzoni, C., & Sauermann, H. (2014). Crowd science: The organization of scientific research in open collaborative projects. *Research Policy*, 43(1), 1-20.

Freitag, A., Meyer, R., & Whiteman, L. (2016). Strategies Employed by Citizen Science Programs to Increase the Credibility of Their Data. *Citizen Science: Theory and Practice*, 1(1), 2. DOI: http://doi.org/10.5334/cstp.6

Gotway, C.A. and Young, L.J., 2002. Combining incompatible spatial data. *Journal of the American Statistical Association*, 97(458), Pp.632-648.

Gricar, B. G., & Baratta, A. J. (1983). Bridging the information gap at Three Mile Island: Radiation monitoring by citizens. *The Journal of Applied Behavioral Science, 19*(1), 35-49.

Haklay, M. (2013). Citizen science and volunteered geographic information: Overview and typology of participation. In *Crowdsourcing geographic knowledge* (pp. 105-122). Springer Netherlands.

Hashimoto, S., Ugawa, S., Nanko, K. and Shichi, K., 2012. The total amounts of radioactively contaminated materials in forests in Fukushima, Japan. *Scientific reports*, 2, 416.

Howe, J. (2006). The rise of crowdsourcing. *Wired magazine*, *14*(6), 1-4.

Irwin, A. (1995). *Citizen science: A study of people, expertise and sustainable development*. Psychology Press.

Isaakson, M and Rääf, C.L. (2017). *Environmental Radioactivity and Emergency Preparedness*. CRC Press.

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

Koarashi, J., Atarashi-Andoh, M., Matsunaga, T., Sato, T., Nagao, S., & Nagai, H. (2012). Factors affecting vertical distribution of Fukushima accident-derived radiocesium in soil under different land-use conditions. *Science of the Total Environment, 431*, 392-401.

Katata, G., Terada, H., Nagai, H. and Chino, M., 2012. Numerical reconstruction of high dose rate zones due to the Fukushima Dai-ichi Nuclear Power Plant accident. *Journal of Environmental Radioactivity*, 111, 2-12.

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

Lochard, J. (2007). Rehabilitation of living conditions in territories contaminated by the Chernobyl accident: the ETHOS Project. *Health physics, 93*(5), 522-526.

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

Nakao, A., Funakawa, S., Tsukada, H., & Kosaki, T. (2012). The fate of caesium-137 in a soil environment controlled by immobilization on clay minerals. *SANSAI : An Environmental Journal for the Global Community, 6*, 17-29.

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

Nursal, W.I., T. Okuda, and T. Yamada. 2017. Potential of Crowdsourcing Approach on Monitoring Radioactivity in Fukushima Prefecture. *Global Environmental Research* (20), 111 – 117.

Onishi, N. & Fackler, M. (2011). Japan Held Nuclear Data, Leaving Evacuees in Peril. *The New York Times*, 8 August 2011, p. A1. Available at http://www.nytimes.com/2011/08/09/ world/asia/09japan.html [Last accessed 12 June 2016].

Pebesma, E.J. (2004). Multivariable geostatistics in S: the gstat package. *Computers & Geosciences*, (30): 683-691.

R Core Team. (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Austria. Vienna, Austria. URL https://www.R-project.org/

Renn, O. (1990). Public responses to the Chernobyl accident. *Journal of Environmental Psychology*, 10(2), 151-167.

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 Onda, Y., 2015. Outline of the national mapping projects implemented after the Fukushima accident. *Journal of Environmental Radioactivity*, 139, 240-249.

Sanada, Y., Sugita, T., Nishizawa, Y., Kondo, A. and Torii, T., 2014. The aerial radiation monitoring in Japan after the Fukushima Daiichi nuclear power plant accident. *Progress in Nuclear Science and Technology*, 4, 76-80.

Schepaschenko, D., See, L., Lesiv, M., McCallum, I., Fritz, S., Salk, C., Moltchanova, E., Perger, C., Shchepashchenko, M., Shvidenko, A. and Kovalevskyi, S., 2015. Development of a global hybrid forest mask through the synergy of remote sensing, crowdsourcing and FAO statistics. *Remote Sensing of Environment*, 162, 208-220.

Senaratne, H., Mobasheri, A., Ali, A. L., Capineri, C., & Haklay, M. (2017). A review of volunteered geographic information quality assessment methods. *International Journal of Geographical Information Science*, *31*(1), 139-167.

Silvertown, J. (2009). A new dawn for citizen science. Trends in ecology & evolution, 24(9), 467-471.

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.

Steinhauser, G., A. Brandl, T. E. Johnson. 2014. Comparison of the Chernobyl and Fukushima nuclear accidents: A review of the environmental impacts. *Science of the Total Environment* (470), 800–817.

Sui, D., Elwood, S., & Goodchild, M. (Eds.). (2012). *Crowdsourcing geographic knowledge: volunteered geographic information (VGI) in theory and practice*. Springer Science & Business Media.

Tanigaki, M., R. Okumura, K. Takamiya, N. Sato,H. Yoshino and H. Yamana (2013) Development of a car-borne ɣ-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.

Toerpe, K. (2013). The rise of citizen science. *The Futurist*, 47(4), 25.

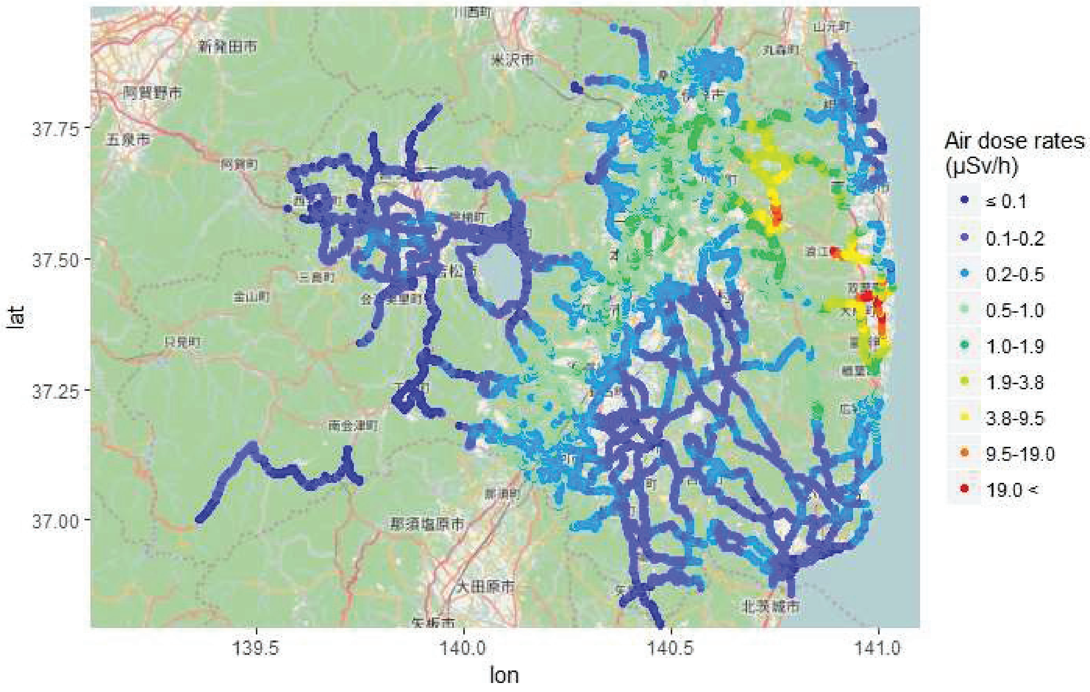
Tsuda, S., T. Yoshida, M. Tsutsumi and K. Saito (2015) Characteristics and verification of a car-borne survey system for dose rates in air: KURAMA-II. *Journal of Environmental Radioactivity*, 139, 260–265.

Wiggins, A., Bonney, R., Graham, E., Henderson, S., Kelling, S., LeBuhn, G., Littauer, R., Lotts, K., Michener, W., Newman, G., Russel, E., Stevenson, R. & Weltzin, J. (2013). Data management guide for public participation in scientific research. DataOne Working Group, 1-41.

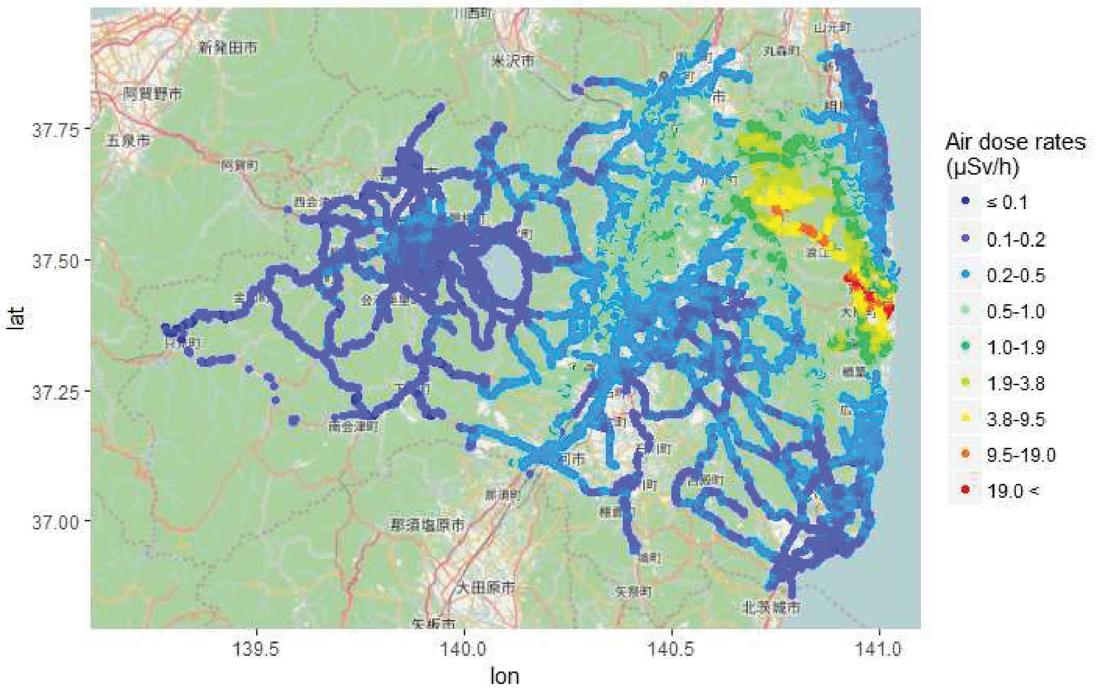
Yasunari, T.J., Stohl, A., Hayano, R.S., Burkhart, J.F., Eckhardt, S. and Yasunari, T., 2011. Cesium-137 deposition and contamination of Japanese soils due to the Fukushima nuclear accident. *Proceedings of the National Academy of Sciences*, 108(49), 19530-19534.

APPENDIX

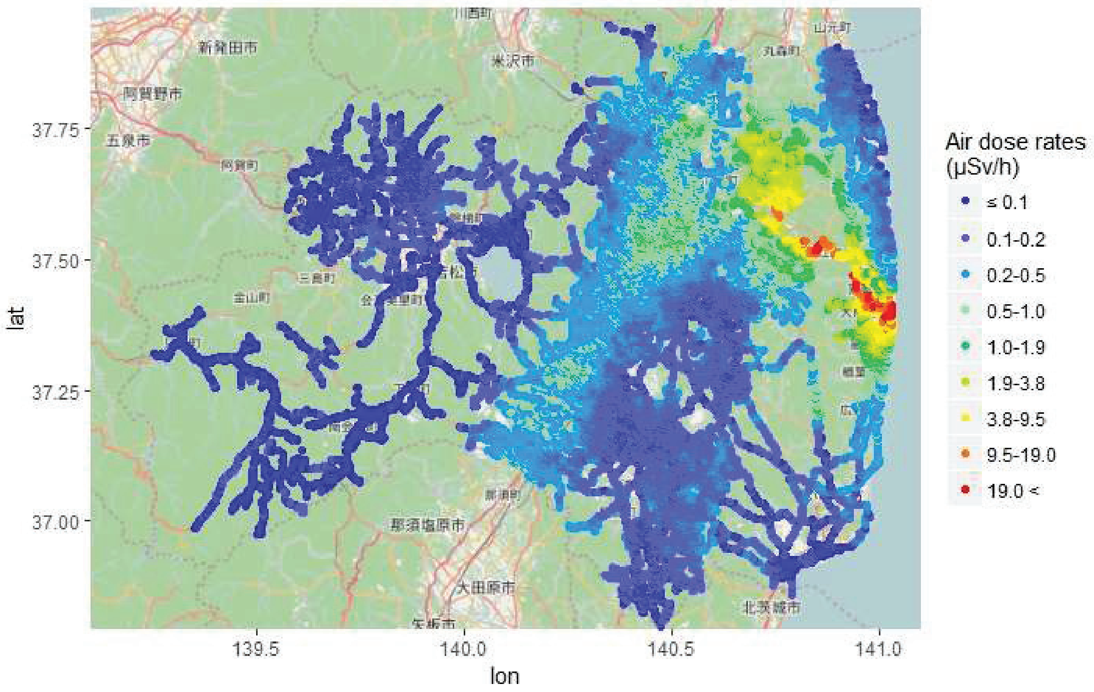
Appendix 1. The KURAMA data acquired in 2011/12/05 – 2011/12/28 (2nd Survey)



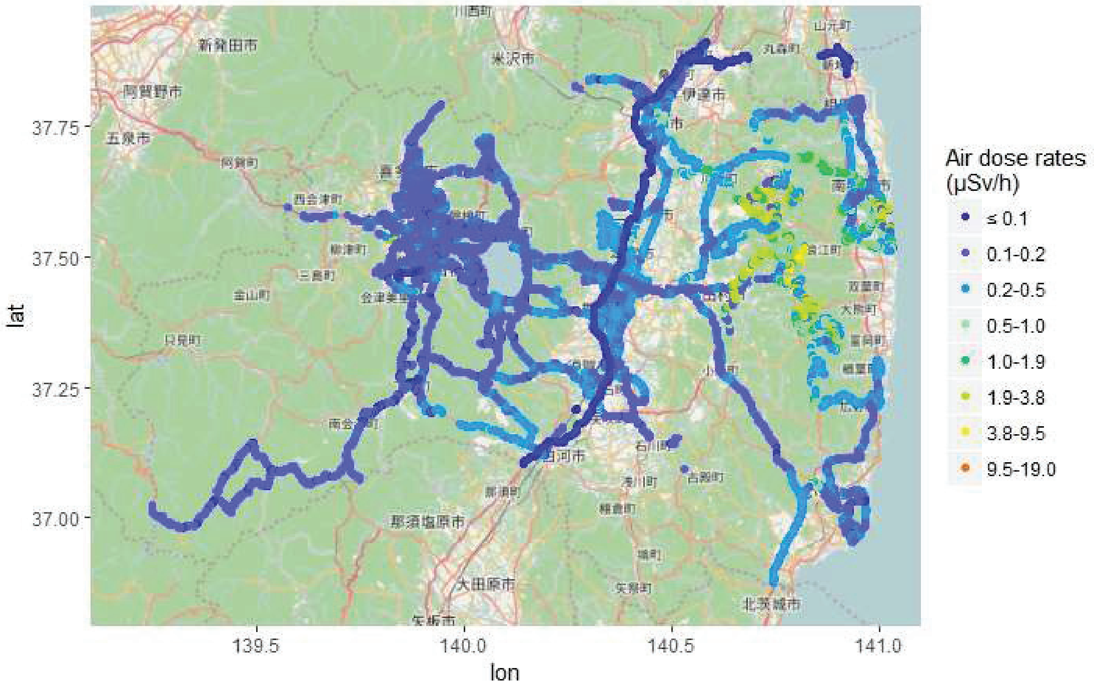
Appendix 2. The parallel dataset of crowsourced data to the 2nd survey of KURAMA data



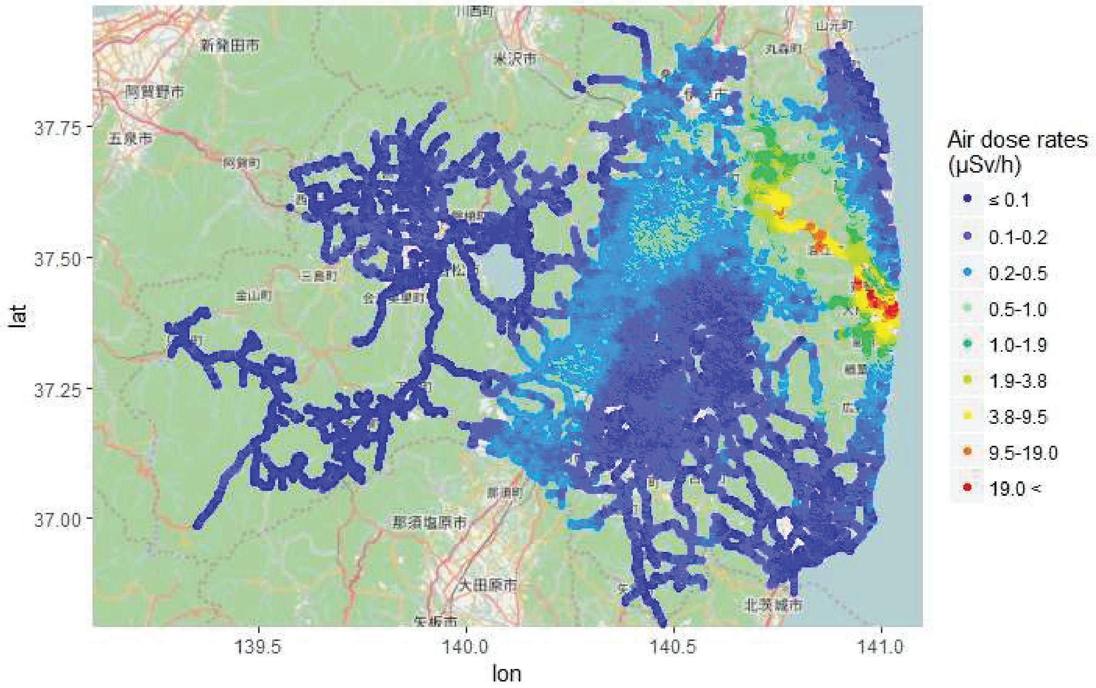
Appendix 3. The KURAMA data acquired in 2012/08/20 – 2012/10/12 (the 4th Survey)



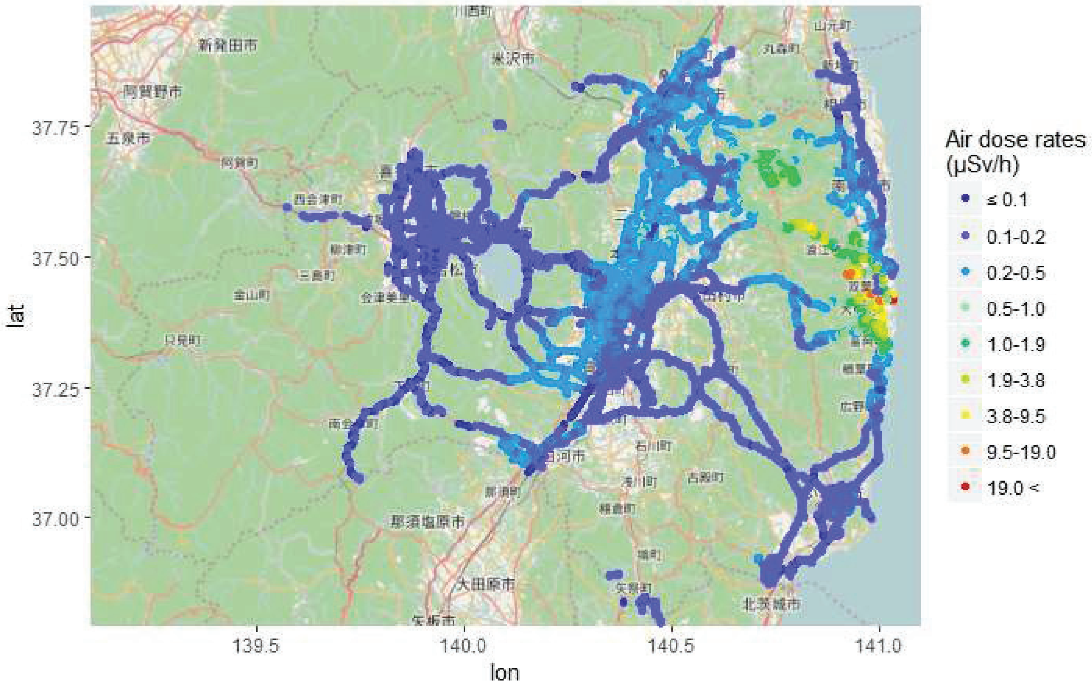
Appendix 4. The parallel dataset of crowdsourced data to the 4th survey of KURAMA data



Appendix 5. The KURAMA data acquired in 2013/11/05 – 2013/12/12 (the 7th Survey)



Appendix 6. The parallel dataset of crowdsourced data to the 7th survey of KURAMA data



Appendix 7. R code for preprocessing task on crowdsourced radiation data

Note about this script:

Consider that the preliminary process on SAFECAST datasets were very important step, we put the code below with many purposes: (1) the reader who interested would find some detail issues within citizen-scientists group radiation data, (2) as a proof that preprocessing task was a tedious and sometimes a must-repeat procedure although the (temporary) result has been generated.

```
rm(list=ls())
options(digits=15)

library(lubridate)
library(sp)
library(rgdal)
library(dplyr)

patokan_hari <- function(date1, date2)
{
  beda_hari <- date2 - date1
  patokan <- date1 + beda_hari/2
  return(patokan)
}
spPromote <- function(dfData, projSystem)
{
  coordinates(dfData) <- ~ Longitude + Latitude
  proj4string(dfData) <- CRS(projSystem)
  return(dfData)
}

spSubset <- function(spData, spClipper)
{
  spData <- spTransform(spData, CRS(proj4string(spClipper)))
  spClipped <- spData[spClipper, ]
  return(spClipped)
}

CRS4326 <- "+init=epsg:4326"
shp <- "ESRI Shapefile"
setwd("/WN/DATA/FUKUSHIMA/AIRDOSE/SAFECAST/")

# SC 2011 ----
load("sc2011-Fukushima-geo2017-02-12 10.52.29.RData")
sc2011Fuku.dup <- sc2011Fuku %>%
                  as.data.frame(.) %>%
                  group_by(Longitude, Latitude) %>%
                  filter(n() > 1) %>%
                  as.data.frame(.) %>%
                  spPromote(., CRS4326)

nrow(sc2011Fuku.dup)
writeOGR(sc2011Fuku.dup,
         dsn=".",
         layer="sc2011-fuku-geo-dup",
         driver=shp)
sc2011Fuku.unq <- sc2011Fuku %>%
                  as.data.frame(.) %>%
                  group_by(Longitude, Latitude) %>%
                  filter(n() == 1)
nrow(sc2011Fuku); nrow(sc2011Fuku.dup); nrow(sc2011Fuku.unq)

# :::: VERY CLOSE POINTS ----
# use distinct
sc2011Fuku.unq <- sc2011Fuku.unq %>%
                  distinct(round(Longitude, digits=4),
                           round(Latitude,digits=4), .keep_all=T)
```

Appendix 7. *Continued*

```
nrow(sc2011Fuku.unq)
head(sc2011Fuku.unq)
names(sc2011Fuku.unq)

sc2011Fuku.unq <- sc2011Fuku.unq[c(1:12)]

# :::: REVISE AIRDOSE ----
sc2011Fuku.unq$airdose <- sc2011Fuku.unq$Value * 0.0029940119760479044

# :::: TRIMMING VERY LOW DOSE ----
summary(sc2011Fuku.unq$airdose)
sc2011Fuku.cln <- subset(sc2011Fuku.unq, airdose >= 0.08)
rm(sc2011Fuku.unq)

# :::: SAVING ----

sc2011Fuku.cln <- spPromote(as.data.frame(sc2011Fuku.cln), CRS4326)
save(sc2011Fuku.cln, file="sc2011-fuku-geo-cln.RData")
writeOGR(sc2011Fuku.cln, dsn=".",
         layer="sc2011-fuku-geo-cln",
         driver=shp)
rm(sc2011Fuku)

# END SC 2011 ----

# SC 2012 ----
load("sc2012-Fukushima-geo2017-02-12 11.02.24.RData")
sc2012Fuku.dup <- sc2012Fuku %>%
                  as.data.frame(.) %>%
                  group_by(Longitude, Latitude) %>%
                  filter(n() > 1) %>%
                  as.data.frame(.) %>%
                  spPromote(., CRS4326)

writeOGR(sc2012Fuku.dup,
         dsn=".",
         layer="sc2012-fuku-geo-dup",
         driver=shp)

sc2012Fuku.unq <- sc2012Fuku %>%
                  as.data.frame(.) %>%
                  group_by(Longitude, Latitude) %>%
                  filter(n() == 1)

nrow(sc2012Fuku); nrow(sc2012Fuku.dup); nrow(sc2012Fuku.unq)
# :::: VERY CLOSE POINTS ----
# use distinct
sc2012Fuku.unq <- sc2012Fuku.unq %>%
                  distinct(round(Longitude, digits=4),
                           round(Latitude,digits=4), .keep_all=T)
nrow(sc2012Fuku.unq)
head(sc2012Fuku.unq)
names(sc2012Fuku.unq)
sc2012Fuku.unq <- sc2012Fuku.unq[c(1:12)]

# :::: REVISE AIRDOSE ----
sc2012Fuku.unq$airdose <- sc2012Fuku.unq$Value * 0.0029940119760479044

# :::: TRIMMING VERY LOW DOSE ----
summary(sc2012Fuku.unq$airdose)
sc2012Fuku.cln <- subset(sc2012Fuku.unq, airdose >= 0.08)
rm(sc2012Fuku.unq)

# :::: SAVING ----

sc2012Fuku.cln <- spPromote(as.data.frame(sc2012Fuku.cln), CRS4326)
save(sc2012Fuku.cln, file="sc2012-fuku-geo-cln.RData")
writeOGR(sc2012Fuku.cln, dsn=".",
         layer="sc2012-fuku-geo-cln",
         driver=shp)
```

Appendix 7. *Continued*

```
rm(sc2012Fuku)
# END SC 2012 ----

# SC 2013 ----
load("sc2013-Fukushima-geo2017-02-12 11.18.08.RData")

sc2013Fuku.dup <- sc2013Fuku %>%
  as.data.frame(.) %>%
  group_by(Longitude, Latitude) %>%
  filter(n() > 1) %>%
  as.data.frame(.) %>%
  spPromote(., CRS4326)

nrow(sc2013Fuku.dup)

writeOGR(sc2013Fuku.dup,
         dsn=".",
         layer="sc2013-fuku-geo-dup",
         driver=shp)

sc2013Fuku.unq <- sc2013Fuku %>%
  as.data.frame(.) %>%
  group_by(Longitude, Latitude) %>%
  filter(n() == 1)

nrow(sc2013Fuku); nrow(sc2013Fuku.dup); nrow(sc2013Fuku.unq)

# :::: VERY CLOSE POINTS ----
# use distinct
sc2013Fuku.unq <- sc2013Fuku.unq %>%
  distinct(round(Longitude, digits=4),
           round(Latitude,digits=4), .keep_all=T)

nrow(sc2013Fuku.unq)
head(sc2013Fuku.unq)
names(sc2013Fuku.unq)

sc2013Fuku.unq <- sc2013Fuku.unq[c(1:12)]

# :::: REVISE AIRDOSE ----
sc2013Fuku.unq$airdose <- sc2013Fuku.unq$Value * 0.0029940119760479044

# :::: TRIMMING VERY LOW DOSE ----
summary(sc2013Fuku.unq$airdose)
sc2013Fuku.cln <- subset(sc2013Fuku.unq, airdose >= 0.08)
rm(sc2013Fuku.unq)

# :::: SAVING ----

sc2013Fuku.cln <- spPromote(as.data.frame(sc2013Fuku.cln), CRS4326)
save(sc2013Fuku.cln, file="sc2013-fuku-geo-cln.RData")
writeOGR(sc2013Fuku.cln, dsn=".",
         layer="sc2013-fuku-geo-cln",
         driver=shp)

rm(sc2013Fuku)

# END SC 2013 ----

# SC 2014 ----
load("sc2014-Fukushima-geo2017-02-12 11.32.36.RData")

sc2014Fuku.dup <- sc2014Fuku %>%
  as.data.frame(.) %>%
  group_by(Longitude, Latitude) %>%
```

Appendix 7. *Continued*

```
filter(n() > 1) %>%
  as.data.frame(.) %>%
  spPromote(., CRS4326)

nrow(sc2014Fuku.dup)

writeOGR(sc2014Fuku.dup,
         dsn=".",
         layer="sc2014-fuku-geo-dup",
         driver=shp)

sc2014Fuku.unq <- sc2014Fuku %>%
  as.data.frame(.) %>%
  group_by(Longitude, Latitude) %>%
  filter(n() == 1)

nrow(sc2014Fuku); nrow(sc2014Fuku.dup); nrow(sc2014Fuku.unq)

# :::: VERY CLOSE POINTS ----
# use distinct
sc2014Fuku.unq <- sc2014Fuku.unq %>%
  distinct(round(Longitude, digits=4),
           round(Latitude,digits=4), .keep_all=T)

nrow(sc2014Fuku.unq)
head(sc2014Fuku.unq)
names(sc2014Fuku.unq)

sc2014Fuku.unq <- sc2014Fuku.unq[c(1:12)]

# :::: REVISE AIRDOSE ----
sc2014Fuku.unq$airdose <- sc2014Fuku.unq$Value * 0.0029940119760479044

# :::: TRIMMING VERY LOW DOSE ----
summary(sc2014Fuku.unq$airdose)
sc2014Fuku.cln <- subset(sc2014Fuku.unq, airdose >= 0.08)
rm(sc2014Fuku.unq)

# :::: SAVING ----

sc2014Fuku.cln <- spPromote(as.data.frame(sc2014Fuku.cln), CRS4326)
save(sc2014Fuku.cln, file="sc2014-fuku-geo-cln.RData")
writeOGR(sc2014Fuku.cln, dsn=".",
         layer="sc2014-fuku-geo-cln",
         driver=shp)

rm(sc2014Fuku)

# END SC 2014 ----

# SC 2015 ?? ----
load("sc2015-Fukushima-geo2017-02-12 11.40.47.RData")
sc2015Fuku.dup <- sc2015Fuku %>%
  as.data.frame(.) %>%
  group_by(Longitude, Latitude) %>%
  filter(n() > 1) %>%
  as.data.frame(.) %>%
  spPromote(., CRS4326)

nrow(sc2015Fuku.dup)

writeOGR(sc2015Fuku.dup,
         dsn=".",
         layer="sc2015-fuku-geo-dup",
         driver=shp)
```

Appendix 7. *Continued*

```
sc2015Fuku.unq <- sc2015Fuku %>%
  as.data.frame(.) %>%
  group_by(Longitude, Latitude) %>%
  filter(n() == 1)

nrow(sc2015Fuku); nrow(sc2015Fuku.dup); nrow(sc2015Fuku.unq)

# :::: VERY CLOSE POINTS ----
# use distinct
sc2015Fuku.unq <- sc2015Fuku.unq %>%
  distinct(round(Longitude, digits=4),
           round(Latitude,digits=4), .keep_all=T)

nrow(sc2015Fuku.unq)
head(sc2015Fuku.unq)
names(sc2015Fuku.unq)

sc2015Fuku.unq <- sc2015Fuku.unq[c(1:12)]

# :::: REVISE AIRDOSE ----
sc2015Fuku.unq$airdose <- sc2015Fuku.unq$Value * 0.0029940119760479044

# :::: TRIMMING VERY LOW DOSE ----
summary(sc2015Fuku.unq$airdose)
sc2015Fuku.cln <- subset(sc2015Fuku.unq, airdose >= 0.08)
rm(sc2015Fuku.unq)

# :::: SAVING ----

sc2015Fuku.cln <- spPromote(as.data.frame(sc2015Fuku.cln), CRS4326)
save(sc2015Fuku.cln, file="sc2015-fuku-geo-cln.RData")
writeOGR(sc2015Fuku.cln, dsn=".",
         layer="sc2015-fuku-geo-cln",
         driver=shp)

rm(sc2015Fuku)
```