Title

Long-term Radiation Monitoring Strategies after Nuclear Power Plant Accidents

Author

Sun, Dajie

Publication Date

2021

Peer reviewed|Thesis/dissertation

Long-term Radiation Monitoring Strategies after Nuclear Power Plant Accidents

By

Dajie Sun

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering - Nuclear Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Haruko Murakami Wainwright, Co-Chair
Professor Jasmina Vujic, Co-Chair
Professor Karl van Bibber
Professor Mark Hansen

Summer 2021

To Liting (Melissa) Wang

Abstract

Long-term Radiation Monitoring Strategies after Nuclear Power Plant Accidents

by

Dajie Sun

Doctor of Philosophy in Nuclear Engineering

University of California, Berkeley

Professor Haruko Murakami Wainwright, Chair

Since the Fukushima Daiichi Nuclear Power Plant (FDNPP) accident in 2011, radiation measurements and monitoring have been conducted continuously. Radiation air dose rate datasets have been archived extensively in this area. There are several different types of measurements: fixed-point measurements, walk surveys, car surveys, airborne surveys, and monitoring posts. They have different spatial coverage, footprints, and uncertainty. Currently, the monitoring program is expected to transition to long-term monitoring after ten years' monitoring. The challenge of long-term monitoring is to build a cost effective and sustainable strategy for minimizing the cost associated with the number of monitoring locations or sampling, while maximizing the ability to meet the objectives of long-term monitoring.

This study aims to develop the long-term radiation monitoring strategies after the FDNPP accident. In this dissertation, we tackle three key challenges: (1) multiscale spatial data integration, (2) monitoring optimization, and (3) spatiotemporal data integration. First, we developed an efficient algorithm for integrating the multiscale data sets; the algorithm is based on Kriging to estimate the dose rates for unobserved locations. Secondly, we developed a strategy and an algorithm to optimize the monitoring post placement and their number. This strategy is designed in order to reduce the number of sensors while capturing spatial heterogeneity. The algorithm is based on Gaussian process model to capture and estimate the heterogeneity of air-dose rates across the domain. Lastly, we built a Kalman-filter based algorithm combined with Gaussian Process Model to predict the spatial–temporal distribution of radiation dose rates. We expect that these methods will have valuable contributions for the long-term monitoring in the Fukushima region, but also for the preparation for the future nuclear accidents.

Table of contents

List of figures

List of tables

Acknowledgments

First and foremost, I want to express my special thanks to my advisor Haruko Murakami Wainwright. It was my honor to be her first Ph.D. student. For the past 4 years, she has been teaching me consciously and unconsciously, she is a very kindhearted person and excellent researcher. I really appreciated her contribution of her time, ideas, and funding which enable me, who was constantly battling with English language, to finish my Ph.D. and lead a wonderful life in Berkeley.

I also want to thank Professor Karl Van Bibber, who was the Chair of the Department of Nuclear Engineering and helped me find an advisor to continue my study in Berkeley. Without his help, my life might have been different.

Meanwhile, I am grateful to Professor Jasmina Vujic and Professor Mark Hansen, who attended my qualifying exam and gave much important feedbacks. They are also members of my thesis committee. I also want to thank Dr. Carlos Oroza, Dr. Kimiaki Saito, Dr. Satoshi Mikami and Dr. Akiyuki Seki, who were co-authors of my publication and contributed to the third chapter of this thesis. Thanks for their time, dedication, and efforts.

My life in Berkeley was so enjoyable, in part due to so many friends I meet here, such as William Mason who taught me shooting, hiking and fishing, Al Kenany, who shared with me a lot of valuable life experiences. Nicolas Min helped me tremendously on looking for jobs and offered kind assistance on revising my job resume. They became important part of my life in Berkeley. My life was also enriched by wonderful cohort members in my research group, such as Dinara Ermakova, who told me about her wonderful experience of the places she has traveled.

Lastly, I would like to thank to my parent, for all their constant care, love and encouragement. Thank you to my girlfriend Liting Wang who gave me her support, encouragement, and love for all three years since we first met, and this is especially important during the final stages of this Ph.D. journey.

In this thesis, I used environmental monitoring data collected during the projects commissioned by the Japan Nuclear Regulatory Agency. I want to thank the people who contributed to collecting the data and compiling them into the JAEA database.

Thank you for all the people mentioned above, and numerous unnamed love, kindness and support along this precious academic journey. You all have been extremely helpful and I sincerely and greatly appreciate you all.

Dajie Sun,

UC Berkeley

July, 2021

1 Background

1.1 Fukushima Accident

The Fukushima Daiichi Nuclear Power Plant (FDNPP) is located on the east coast of Japan in the towns of Ōkuma and Futaba in the Fukushima Prefecture—and next to the Pacific Ocean (as indicated in Fig. 1.1). On 11 March 2011, the Great East Japan Earthquake, magnitude 9.0, hit the east coast of Japan, causing considerable damage to that region and creating a large tsunami that destroyed the backup generators of the FDNPP. The combined effect of earthquake and tsunami led to the loss of off-site and on-site electrical power. Then the reactor core lost its cooling function, the water level dropped, and hydrogen was generated due to cladding-water reaction at high temperature. The hydrogen leaked out from the containment and stayed inside the building in which the containment was located. Gradually, the hydrogen accumulated and eventually exploded and destroyed the building. Such process led to the discharge of radioactive nuclides into the environment (TEPCO, 2013). The entire sequence of events is shown in Fig. 1.2.

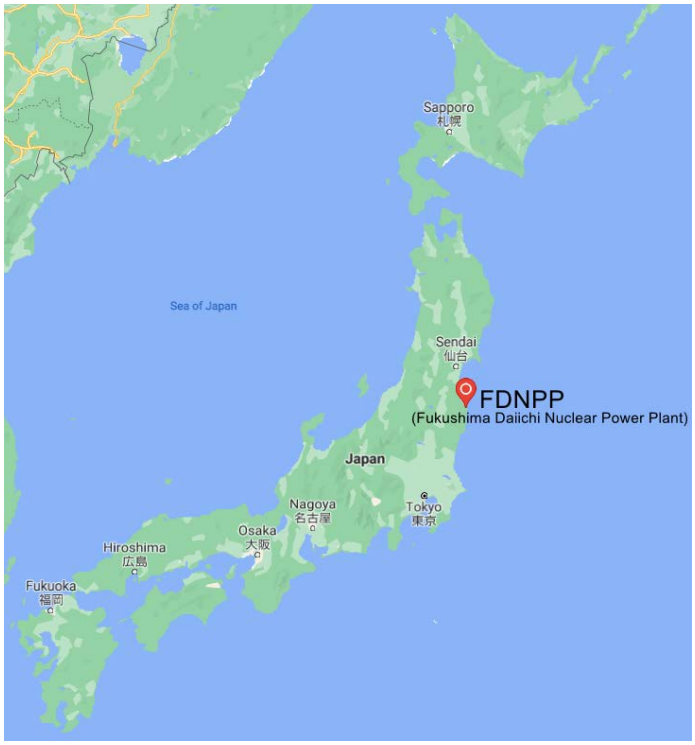

Fig. 1.1 A map of the Fukushima Daiichi Nuclear Power Plant

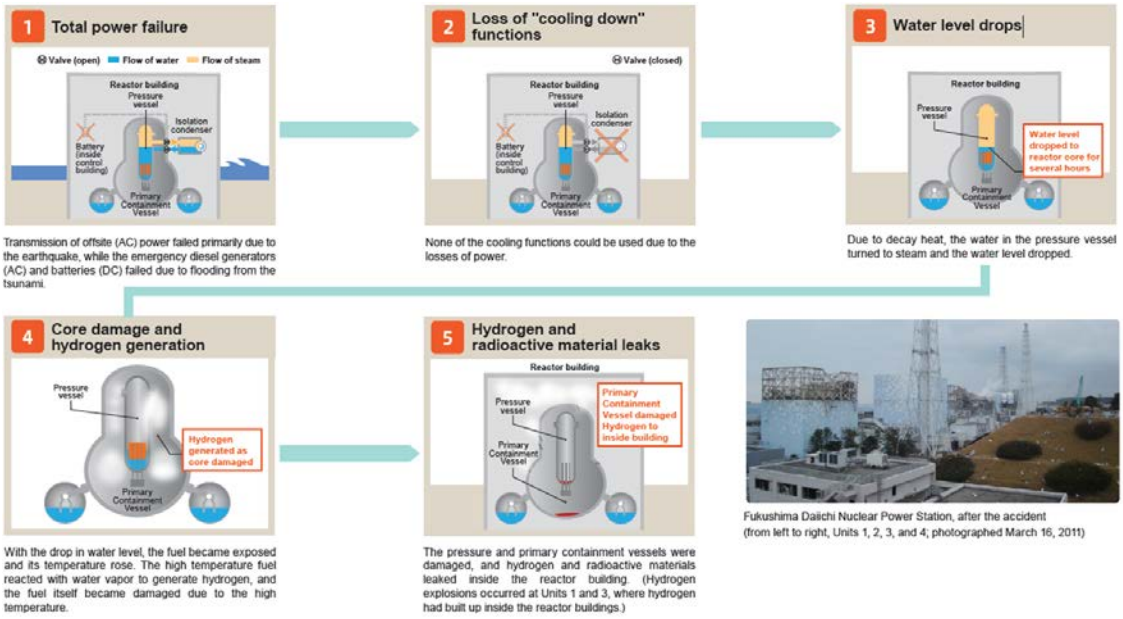

Fig. 1.2 General Outline of the Development of the FDNPP Accident on March 11, 2011 (TEPCO, 2013)

Table 1.1 provides an estimation of radionuclide releases into the atmosphere from the FDNPP accident, published by the Nuclear Industry Safety Agency (NISA, 2011), based on burn-up analysis. It was estimated that the radiocesium (mainly ^{137}Cs and ^{134}Cs) released into the atmosphere by FDNPP (about 3.3×10^{16} Bq) was one order of magnitude smaller than that released by the Chernobyl nuclear accident of 1986 (Saito et al., 2015). Several long-life radionuclides, especially radiocesium, would remain in the environment for decades and pose a risk of exposure to the public.

Table 1.1 Estimated radioactivity releases into the atmosphere in the FDNPP accident given by the Nuclear Industry Safety Association (NISA, 2011)

Radionuclide	Halflife	Released radioactivity (Bq)			
		Reactor 1	Reactor 2	Reactor 3	Total
^{134}Cs	2.1 y	7.1 X 10^{14}	1.6 X 10^{16}	8.2 X 10^{14}	1.8 X 10^{16}
^{137}Cs	30 y	5.9 X 10^{14}	1.4 X 10^{16}	7.1 X 10^{14}	1.5 X 10^{16}
^{89}Sr	50.5 d	8.2 X 10^{13}	6.8 X 10^{14}	1.2 X 10^{15}	2.0 X 10^{15}
^{90}Sr	29.1 y	6.1 X 10^{12}	4.8 X 10^{13}	8.5 X 10^{13}	1.4 X 10^{14}
^{140}Ba	12.7 d	1.3 X 10^{14}	1.1 X 10^{15}	1.9 X 10^{15}	3.2 X 10^{15}
^{127m}Te	109 d	2.5 X 10^{14}	7.7 X 10^{14}	6.9 X 10^{13}	1.1 X 10^{15}
^{129m}Te	33.6 d	7.2 X 10^{14}	2.4 X 10^{15}	2.1 X 10^{14}	3.3 X 10^{15}
^{131m}Te	30 h	2.2 X 10^{15}	2.3 X 10^{15}	4.5 X 10^{14}	5.0 X 10^{15}
^{132}Te	78.2 h	2.5 X 10^{16}	5.7 X 10^{16}	6.4 X 10^{15}	8.8 X 10^{16}
^{103}Ru	39.3 d	2.5 X 10^{09}	1.8 X 10^{09}	3.2 X 10^{09}	7.5 X 10^{09}
^{106}Ru	368.2 d	7.4 X 10^{08}	5.1 X 10^{08}	8.9 X 10^{08}	2.1 X 10^{09}
^{95}Zr	64 d	4.6 X 10^{11}	1.6 X 10^{13}	2.2 X 10^{11}	1.7 X 10^{13}
^{141}Ce	32.5 d	4.6 X 10^{11}	1.7 X 10^{13}	2.2 X 10^{11}	1.8 X 10^{13}
^{144}Ce	284.3 d	3.1 X 10^{11}	1.1 X 10^{13}	1.4 X 10^{11}	1.1 X 10^{13}
^{239}Np	2.4 d	3.7 X 10^{12}	7.1 X 10^{13}	1.4 X 10^{12}	7.6 X 10^{13}
^{238}Pu	87.7 y	5.8 X 10^{08}	1.8 X 10^{10}	2.5 X 10^{08}	1.9 X 10^{10}
^{239}Pu	24,065 y	8.6 X 10^{07}	3.1 X 10^{09}	4.0 X 10^{07}	3.2 X 10^{09}
^{240}Pu	6537 y	8.8 X 10^{07}	3.0 X 10^{09}	4.0 X 10^{07}	3.2 X 10^{09}
^{241}Pu	14.4 y	3.5 X 10^{10}	1.2 X 10^{12}	1.6 X 10^{10}	1.2 X 10^{12}
^{91}Y	58.5 d	3.1 X 10^{11}	2.7 X 10^{12}	4.4 X 10^{11}	3.4 X 10^{12}
^{143}Pr	13.6 d	3.6 X 10^{11}	3.2 X 10^{12}	5.2 X 10^{11}	4.1 X 10^{12}
^{147}Nd	11 d	1.5 X 10^{11}	1.3 X 10^{12}	2.2 X 10^{11}	1.6 X 10^{12}

^{242}Cm	162.8 d	1.1 X 10^{10}	7.7 X 10^{10}	1.4 X 10^{10}	1.0 X 10^{11}
^{131}I	8 d	1.2 X 10^{16}	1.4 X 10^{17}	7.0 X 10^{15}	1.6 X 10^{17}
^{132}I	2.3 h	1.3 X 10^{13}	6.7 X 10^{06}	3.7 X 10^{10}	1.3 X 10^{13}
^{133}I	20.8 h	1.2 X 10^{16}	2.6 X 10^{16}	4.2 X 10^{15}	4.2 X 10^{16}
^{135}I	6.6 h	2.0 X 10^{15}	7.4 X 10^{13}	1.9 X 10^{14}	2.3 X 10^{15}
^{127}Sb	3.9 d	1.7 X 10^{15}	4.2 X 10^{15}	4.5 X 10^{14}	6.4 X 10^{15}
^{129}Sb	4.3 h	1.4 X 10^{14}	5.6 X 10^{10}	2.3 X 10^{12}	1.4 X 10^{14}
^{99}Mo	66 h	2.6 X 10^{09}	1.2 X 10^{09}	2.9 X 10^{09}	6.7 X 10^{09}

1.2 Monitoring Program

In order to protect the public from radiation exposure and take the right countermeasures, Japanese government branches and research institutions needed accurate and detailed information about the distribution of the deposited radiocesium and resultant air-dose rates (i.e., the ambient dose equivalent rate (μSv/h) at 1 m above the ground)**.** Three months after the FDNPP accident, the Japan Atomic Energy Agency (JAEA), in cooperation with other organizations, conducted a mapping project in order to develop a comprehensive elucidation of the contamination conditions (Saito et al., 2019). Meanwhile, aerial radiation monitoring (ARM), a separate project that preceded the mapping project, had been conducted by the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT) and the U.S. Department of Energy (DOE). These types of radiation measurements have been conducted more than 40 times since the accident (extending from the immediate surrounding area of the FDNPP to the entire country of Japan), obtained the distribution of ambient dose-rate and deposition of radioactive cesium in Japan during this time (Sanada et al., 2014, 2015). The periods of each campaign are summarized in Fig. 1.3 (Saito et al., 2019), whereas details of each type of measurements are explained in Section 1.3. These methods listed in *Fig. 1.3* as well as monitoring posts were also used by the national and local government to perform continuous environmental monitoring. Furthermore, several nonprofit organizations have performed independent environmental monitoring and have accumulated large amounts of radiation data; for example, Nursal et al. (2016) made comparisons of data agreeability from different organizations.

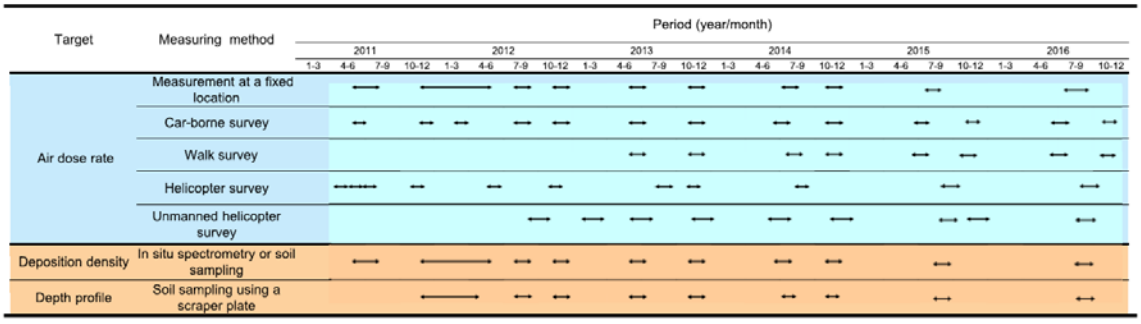

Fig. 1.3 Time table of Periods of large-scale environmental monitoring conducted in national projects (Saito et al., 2019).

After ten years of monitoring, massive data sets were accumulated and the monitoring program is expected to transition from conducting short-term monitoring to long-term

monitoring. With long-term monitoring, the economic cost of such a program needs to be considered along with its effectiveness in meeting its scientific objectives. This cost consideration brings up the topic of optimization. While there are many optimization algorithms by which to optimize radiation monitoring networks, what makes our study unique is that we are building an optimization algorithm to improve radiation monitoring based on spatially extensive datasets and spatial information. Details will be given in Chapter 3.

1.3 Different Type of Measurements

There are four major types of measurements for air dose rates, as indicated in Fig. 1.3: (1) measurements at fixed locations (referred to as "fixed-point measurements" in some publications), (2) car-borne surveys (shortened to "car surveys" in some publications), (3) walk surveys, and (4) helicopter surveys (including unmanned helicopter surveys). In some publications (such as in Wainwright et al., 2017), helicopter surveys and unmanned helicopter surveys are often merged as "air surveys." Besides the four major types mentioned above, monitoring posts were also used by National and local governments for continuous environmental monitoring. Saito et al. (2019) presented a detailed summary of the different character of each type of measurement. Here, only a brief description is provided.

Measurement at a fixed location is the most accurate method among the four. It uses a NaI(Tl) detector and takes measurements at undisturbed fields, such as natural forests, where neither human disturbance nor high flooding are expected. Fixed-point measurement has the highest accuracy, while its spatial coverage is the smallest among the four.

Walk surveys are the second most accurate method among the four. For the period we studied, these were accomplished by a person carrying a Kyoto University RAdiation MApping system (KURAMA-II) walking around a 1×1 km^2 square area for as long a distance as possible. The KURAMA-II mobile survey system, developed by Kyoto University, was updated from the KURAMA-I. The measurement result was automatically transferred through a cellular phone network (Tanigaki et al., 2015). Continuously measured air dose rates were averaged over a 20×20 m^2 area to reduce the statistical fluctuations, so these walk surveys had a larger coverage area than the fixed point measurements.

Car-borne surveys were conducted by a moving car equipped with the KURAMA-II system along the road. In order to reduce statistical fluctuations, continuously measured air-dose rates were averaged over a 100×100 m^2 area; thus, the car-borne surveys had a larger coverage area than the walk-survey measurements and were less accurate because it was averaged over the 100×100 m^2 area. Furthermore, car-borne surveys were restricted to areas with roads and could not cover entire targeted areas.

Helicopter surveys involved attaching a large NaI(Tl) detector with high detection efficiency placed inside or under a helicopter. The helicopter's standard fly height was 300 meters; it had a larger coverage area compared to car-borne surveys; Takahashi et al. (2017) presented a coverage model for helicopter surveys. Wainwright et al. (2017, 2018) found that, in the case of the Fukushima accident, the values in airborne survey data were found to be systematically slightly higher than those measured on land. Unmanned helicopter surveys used a LaBr detector and were performed within a 5 km radius of the FDNPP site, where it was difficult to enter by land. The standard fly height is 80-100 m, which was lower than the manned helicopter surveys, and this led to a higher precision and smaller coverage area (Sanada et al., 2015). A comparison summary is listed in Table 1.2 (Saito et al., 2019)

Table 1.2 Evaluation of each measurement methods employed in large-scale environmental monitoring after FDNPP accident. A higher number indicates a higher rank in the evaluation.

Measurement Method	**Evaluated score**		**Noteworthy**
	Data accuracy	Coverage	
Fixed point measurement	5	1	Give reference value
Car-borne survey	4	2	Related to living environment
Walk survey	3	3	Enormous amount of data
Air-borne survey (unmanned helicopter)	2	4	Supplement helicopter survey
Air-borne survey (helicopter)	1	5	Cover whole area

As shown in Table 1.2, different types of measurement have different spatial coverages and precisions (or uncertainties), and calibration is also needed for some types. In fact, for this period (the accident and its aftermath), massive data sets from these types of measurements have been accumulated, and these different data sets (and the useful information from these data sets) have been integrated to generate the most accurate and comprehensive dose-rate-distribution maps possible. Chapter 2 will discuss the existing algorithm for integrating different types of data sets and the way to extend the existing algorithm.

1.4 Time Trend of the Dose Rates for Recent Years

There are three main causes for the decrease in area dose rates:

(1) Physical decay of radiocesium

Radiocecium has two main isotopes:^{134}Cs has a shorter half-life (of 2.1 years), and ^{137}Cs is 30.2 years, the ratio of their initial released radioacitivity is approximately 6:5 as shown by Table 1.1. As a consequence, ^{134}Cs contributed more to the air dose rate reduction than ^{137}Cs initially. However, as time went on, the contribution of ^{134}Cs was reduced after certain points in time. Saito et al. (2019) presented a diagram of the evolution of their contribution, as shown in Fig. 1.4.

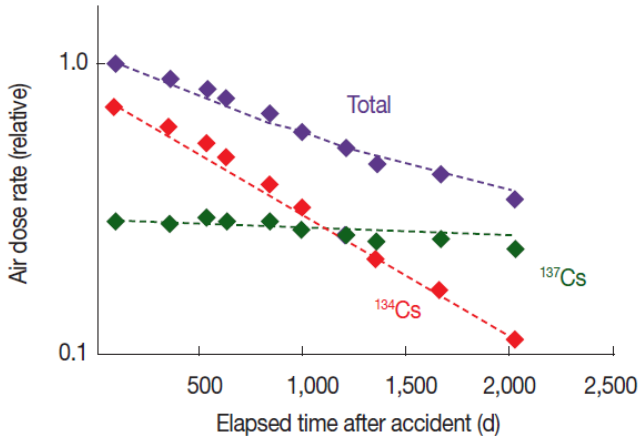

Fig. 1.4 Temporal changes in air dose rates as result of decrease in radiocesium deposition density

As shown in Fig. 1.4, the contribution of ^{137}Cs surpassed ^{134}Cs in around 1100 days after the accident, and both of them go down approximately linearly and continuously with time under logarithm scale.

(2) Migration of radiumcesium in the soil

The soil could shield the gamma ray emitted from the radiocesium underground. In other words, the dose rates would decrease as a result of the vertical penetration of the radiocesium into the soil (Saito et al., 2018). At the same time, radiocesium is known to migrate laterally by sediment transport. Even though the horizontal migration of radiocesium in some undisturbed area (such as natural forests) is very small, it can be significant in certain circumstances, such as urban areas and cultivated fields. In these areas, the horizontal migration would accelerate the decrease in air dose rates. Generally speaking, the migration in undisturbed areas, such as natural forests, would be slow, but in places with human activity, it would be accelerated. For example, it was found that radiocesium deposited on artificial structures (such as paved roads and houses) was washed off in a short time in the early stages after deposition, and this resulted in a rapid air-dose-rate reduction. In conclusion, due to the influence of migration, the rates of decrease for different land-use types are different.

(3) Human decontamination activities.

Decontamination involves removing radioactive materials from areas of human activity in order to promptly decrease their impact on human health and the environment. After the accident, decontamination was performed extensively over wide areas. Evrard et al.(2019) gave a detailed summary and evaluation of the decontamination plan. According to the result posted by the Ministry of Environment (Environmental remediation subsection), as a result of decontamination, air dose rates were reduced by a factor of more than two in residential areas and farmland, by measuring and comparing the air dose rates above decontaminated locations before and after the decontamination, as shown in Fig. 1.5. Some studies have found that the extensive decontamination effort in the region played a critical role in this recovery process (Yasutaka et al., 2013; Wainwright et al., 2018).

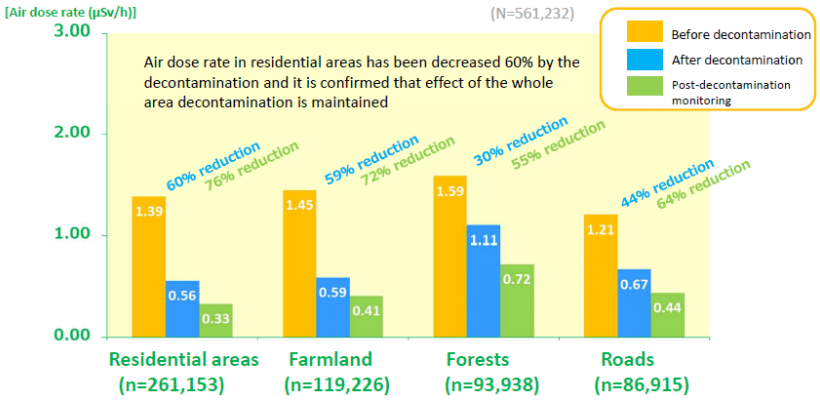

Fig. 1.5 Effect of Decontamination for different land-use types. The chart shows the air dose rate average in each category (Ministry of Environment. Environmental remediation., 2021).

In Fig. 1.5, the reduction of dose rates in different land-use areas are different. In residential areas with extensive human activity, the dose rates were reduced by 60% percent due to decontamination, while in the forests, where there are fewer human activities, the reduction factor can be 30%.

Combined with factors mentioned above, the contribution of each factor to the overall dose rate reduction was summarized by Saito et al. (2019), as shown in Fig. 1.6.

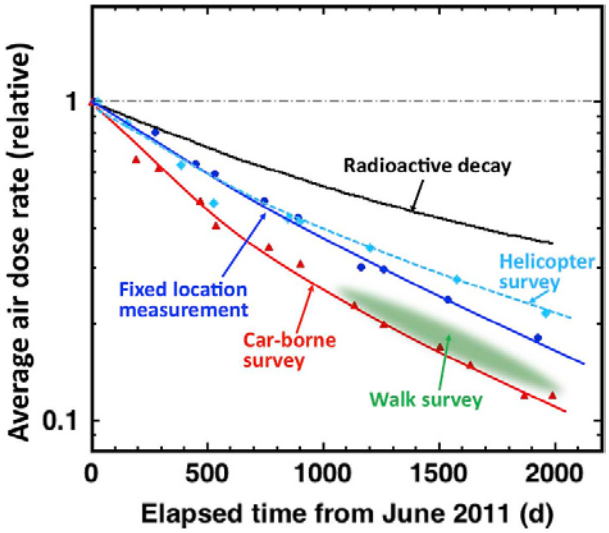

Fig. 1.6 Temporal changes in average air dose rates evaluated by using data from car-borne surveys, fixed-location measurements in undisturbed fields, helicopter surveys, and walk surveys. Air dose rates were normalized to those in June 2011.

In Fig. 1.6, the reason air dose rates as measured by helicopter surveys have decreased more slowly than those derived from other measurements is because helicopter surveys detect gamma rays over wide areas that include forests with little human activity, while car surveys and walk surveys only take measurements in places that have human activity, which accelerated the rate of decrease.

The literature reviews (in Chapter 4) have suggested that even though there are many data-driven models developed already to predict the decay trend of the average dose rates for the same land-use category, the reality is that different locations will decay at different speeds, even within the same land-use category. As of now, there is no mathematical model for predicting dose-rate decay for certain locations within a certain land-use category, soil type, vegetation, etc. In Chapter 4, we fill that gap and build a model that would predict decay rates for specific locations and various conditions.

1.5 Scope of the Dissertation

The above sections have presented the background of the Fukushima Daiichi Nuclear Power Plant Accident and revealed the existing challenges in developing a monitoring strategy after the accident. They include multiscale data integration, sensor network optimization, and dose-rate time-series prediction. The goal of the study presented in this thesis is to develop and apply models to solve these challenges. This study adds new perspectives for people to evaluate the existing monitoring program and provide possible ways to improve it.

This study is unique with respect to the extent of its analysis. Never before has such wide-ranging and continuous data sets been collected and readily presented to the public for this type of study. Most previous studies in this area have borrowed or extended research techniques from the Chernobyl accident or nuclear test monitoring after the

Second World War. In addition, this study applies statistical methods, especially machine learning techniques, which enables us to extract actionable information from a large volume of data sets.

In the following chapters, I will address specifically the issues I brought up at the beginning of this section as follows: Chapter 2: Multiscale Data Integration, extending the existing algorithm of multiscale datasets integration developed by Haruko Wainright, and solving its computation efficiency issue; Chapter 3: Sensor Network Optimization, developing a monitoring strategy end with optimized sensor network; Chapter 4: Spatial-temporal Distribution Prediction, building an approach to predict the radiation level spatially and temporally; and Chapter 5: Conclusion, summarizing the work done in this thesis.

2 Multi-scale data integration

2.1 Introduction

As mentioned in Chapter 1, monitoring of air dose rates (i.e., ambient dose equivalent rates) in the area around the FDNPP have been performed continuously and massive measurement data has been collected since the accident (e.g., Saito and Onda, 2015; Mikami et al., 2015). Generally speaking, there are various techniques applied in the air dose rates measurements, such as walk surveys (a person carrying portable monitoring systems), fixed-point measurement, car surveys, and airborne surveys. Monitoring posts were also widely used in places with a lot of people, such as schools, hospitals, urban area, etc. As discussed in section 1.3, different types of measurement have different spatial coverage and precision (see Table 1.2 in Chapter 1).

Different measurement devices generate different types of data sets. As different types of data sets have different levels of accuracy and different support scale (i.e., support volume, resolution), there are discrepancies among these different types of surveys in terms of measured air dose rates, even they were collected at around the same time and same locations. Different data sets will also have different data sample density and spatial coverage. For example, car surveys are only possible along road, but with high spatial density; Walk surveys are often limited to selected neighborhoods because of the time and labor they require; Airborne surveys were initially conducted in restricted areas where human beings are not able to enter due to high radiation level.

For different characteristics mentioned above, there is a need to develop an approach to integrate different types of measurements, and to provide an integrated map of air dose rates by taking into account the characteristics.

This chapter is an extension of an existing integration method based on Bayesian Geostatistical approach, developed by Wainwright et al. (2017). We first summarize this existing approach in section 2.2 and 2.3 and then built a new approach in section 2.4 and 2.5.

2.2 Exploratory Data Analytis

In Wainwright et. al (2017), only walk surveys, car surveys and airbone surveys were considered. The comparison of different types of measurements in different type of land use are shown in Fig. 2.1 and Fig. 2.2. Fig. 2.1 and Fig. 2.2 not only showed the correlation between them, but also provided a way to calibrate the car survey and air survey data sets (the walk surveys were set as a standard due to their higher accuracy).

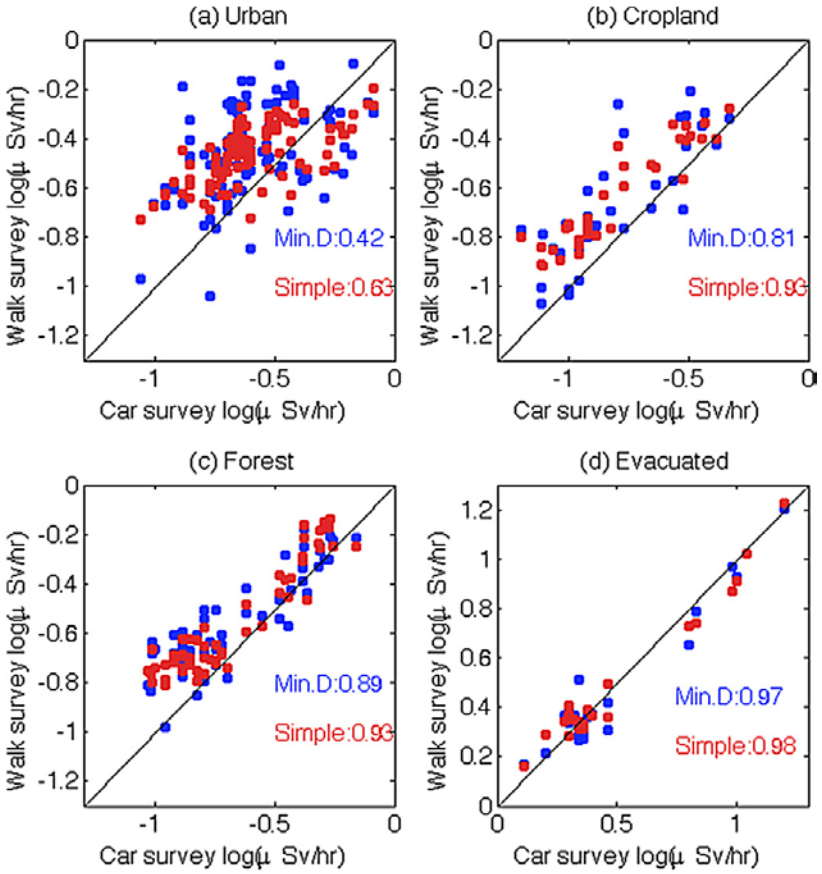

Fig. 2.1 Comparison between the walk survey and car survey data: (a) urban, (b) cropland, (c) forest and (d) evacuated areas. The blue points are the co-located points identified by the minimum distance method. The red points are the average of the walk survey points using the simple average method. The correlation coefficients are attached in each plot. Graphic used with permission from Wainwright 2017.

In Fig. 2.1, the blue points are using the minimum distance to find the co-located points for comparison while the red points are using a simple average method. Obviously, the blue points have a larger scattering distribution than the red points. It is also obvious that the car survey data tends to underestimating the walk survey data in urban, cropland and forest. That is probably because the human activity and car activities removed the radiocesium in the middle of the roads.

Similar to Fig. 2.1, Fig. 2.2 gives the comparison of air surveys and walk surveys. The blue points in Fig. 2.2 mean that the co-locations are selected by minimum distance method while the red points by weighted average which average out the small-scale variability in the airborne survey data. It is also obvious that the airborne surveys tend to overestimate the air dose rates, that is probably because a helicopter flying 300 meters above the ground is able to collect more gamma rays from a much larger spatial angle than other surveys (Saito et al., 2019). Using the weighted average method (the red points

in Fig. 2.2) significantly improves the correlation between airborne data and walk survey data.

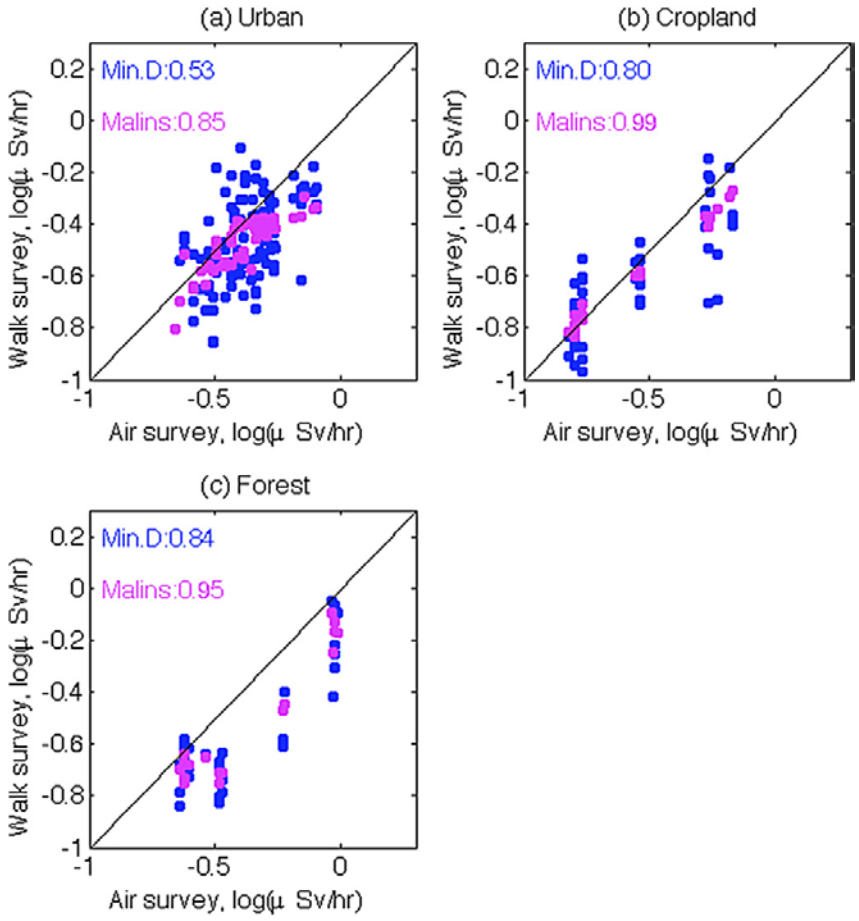

Fig. 2.2 Comparison between the walk survey and air survey data in: (a) urban, (b) cropland, and (c) forest. The blue points are the co-located points identified by the minimum distance. The magenta points are the average of the walk survey points using the weights computed by the radiation transport simulation. Graphic used with permission from Wainwright 2017.

In summary, Fig. 2.1 and Fig. 2.2 not only show the correlation between different kind of measurement but also present a way to calibrate different measurement data (assuming the walk surveys are the most accurate and can be used as a reference in calibration). However, for measurements that are taken in locations that are far apart, calibration method doesn't work in this context as different types of measurements at different locations are not comparable. An integration method must be developed to incorporate all these type of measurements.

2.3 Bayesian Hierarchical Approach

In Wainwright et al. (2017) paper, the Bayesian hierarchical approach was developed to integrate the three type of data sets.

We denote the air dose rate at i_{th} pixel by y_i, where $i = 1, \cdots, n$. The three datasets(walk survey, car survey, airborne survey) were denoted by three vectors, representing the airborne survey data z_A (each data point is represented by $z_{A,j}$, where $j = 1, \cdots, m_A$), car survey data z_C (each data point is represented by $z_{C,j}$, where $j = 1, \cdots, m_C$), and walk survey data z_W (each data point is represented by $z_{W,j}$, where $j = 1, \cdots, m_W$). Based on Bayes's rule:

$$p(y|z_A, z_C, z_W) \propto p(z_A|y)p(z_C|y)p(y|z_W) \tag{2.1}$$

Due to the correlation described in Fig. 2.1 and Fig. 2.2, airborne survey data z_A and car survey z_C can be represented as

$$\begin{cases} z_A = Ay + \varepsilon_A \\ z_C = Cy + \varepsilon_C \end{cases} \tag{2.2}$$

where A is a $m_A\ by\ n$ matrix and C is a $m_C\ by\ n$ matrix that includes the correlation parameters and weights (obtained from the Monte Carlo transport code developed by Malins et at. (2015, 2016)), ε_A and ε_C are all uncertainty vectors with diagonal covariance matrix. Furthermore, as y is a multivariate normal distribution, z_A and z_C also follow

$$\begin{cases} z_A \sim MVN(Ay, D_A) \\ z_C \sim MVN(Cy, D_C) \end{cases} \tag{2.3}$$

where D_A is $m_A\ by\ m_A$ diagonal covariance matrix of ε_A, and D_A is $m_C\ by\ m_C$ diagonal covariance matrix of ε_C.

we use the walk survey data as conditional points to constrain the distribution of y as $p(y|z_W)$. we assume

$$y|z_W \sim MVN(\mu_C, \Sigma_C) \tag{2.4}$$

where μ_C, Σ_C are parameters estimated from section 2.3.

So, we have the air dose rate vector y follows

$$y \sim MVN(Q^{-1}g, Q^{-1}) \tag{2.5}$$

where

$$\begin{cases} Q = \Sigma_c^{-1} + A^T D_A^{-1} A + C^T D_C^{-1} C \\ g = \Sigma_c^{-1}\mu_C + A^T D_A^{-1} z_A + C^T D_C^{-1} z_C \end{cases} \tag{2.6}$$

In the computation, y is a vector representing the dose rate at all the vertices in the grid

over the study area(or whole field). So, the means of y(which is $Q^{-1}g$) is the integrated dose rate map, and $Q^{-1}$is the estimation variance.

All the above equations ((2.1) to (2.6)) consist the Bayesian Hierarchical Approach in Wainwright's original paper (Wainwright et al., 2017).

2.4 New Approach.

Bayesian Hierarchical Approach does provide a way to integrate different types of datasets, but the result (Equation (2.5) and (2.6)) involves a large matrix inversion, which is very computation expensive, and requires huge memory to store the $n \times n$ matrix Q.

In this chapter we developed a new method for data integration, which was much more computationally efficient and proved to equal the result given by the original method.
Let us define a new variable:

$$y^* = \begin{pmatrix} z_A \\ z_C \\ y_1 \\ \vdots \\ y_n \end{pmatrix} \tag{2.7}$$

$p(y|z_W) \sim MVN(\mu_C, \Sigma_C)$ as the prior distribution of y, then, the covariance of $y^*$is

$$\mathrm{cov}(y^*, y^*) = \begin{pmatrix} A\Sigma_c A^T + D_A & A\Sigma_c C^T & A\Sigma_c \\ C\Sigma_c A^T & C\Sigma_c C^T + D_C & C\Sigma_c \\ \Sigma_c A^T & \Sigma_c C^T & \Sigma_c \end{pmatrix} \tag{2.8}$$

Using the simple kriging method, the conditional variance and mean of y based on z_A and z_C are

$$\begin{cases} \mu_{y|(z_A,z_C)} = \mu_c + (\Sigma_c A^T, \quad \Sigma_c C^T) \begin{pmatrix} A\Sigma_c A^T + D_A & A\Sigma_c C^T \\ C\Sigma_c A^T & C\Sigma_c C^T + D_C \end{pmatrix}^{-1} \begin{pmatrix} z_A - A\mu_c \\ z_C - C\mu_c \end{pmatrix} \\ \Sigma_{y|(z_A,z_C)} = \Sigma_c - (\Sigma_c A^T, \quad \Sigma_c C^T) \begin{pmatrix} A\Sigma_c A^T + D_A & A\Sigma_c C^T \\ C\Sigma_c A^T & C\Sigma_c C^T + D_C \end{pmatrix}^{-1} \begin{pmatrix} A\Sigma_c \\ C\Sigma_c \end{pmatrix} \end{cases} \tag{2.9}$$

The variance of each pixel can even be simplified as

$$\begin{pmatrix} \sigma_{y1}^2 \\ \vdots \\ \sigma_{yn}^2 \end{pmatrix} = diag\left(\Sigma_{y|(Z_A,Z_C)}\right) \\ = diag(\Sigma_c) - \text{colSums}\left(\Sigma_{\text{acy}}(.*)\{\ (\Sigma_{\text{ac}})^{-1}\Sigma_{\text{acy}}\}\right) \tag{2.10}$$

Where $\Sigma_{\text{ac}} = \begin{pmatrix} A\Sigma_c A^T + D_A & A\Sigma_c C^T \\ C\Sigma_c A^T & C\Sigma_c C^T + D_C \end{pmatrix}$, and $\Sigma_{\text{acy}} = \begin{pmatrix} A\Sigma_c \\ C\Sigma_c \end{pmatrix}$, the operator $(.*)$ means multiplying element by element.

In the solution above, the $\begin{pmatrix} A\Sigma_c A^T + D_A & A\Sigma_c C^T \\ C\Sigma_c A^T & C\Sigma_c C^T + D_C \end{pmatrix}$ is an $(m_A + m_C) \times (m_A + m_C)$ (where m_A is the number of air surveys and m_C is the number of car borne surveys), as $(m_A + m_C)$ is much smaller than n, so the result above (Equation (2.9)) is much more computational efficient compared with the result given by Bayesian hierarchical approach (Equation (2.5)~(2.6)). A detailed analysis will be made in next section.

It was proven that, even though the computation complexity of Equation (2.5) and (2.6) and Equation (2.9) are different, the two sets of equations are actually equal, the detail proof can be found in Appendix I.

2.5 Comparison

Computation complexity

In the original result, Σ_C is the covariance matrix of the whole field with a dimension of n×n, where n is the total number of pixels in the field and could be tens of thousands or hundreds of thousands (in the real dataset, $n \approx 20\ 000$).

The issue is that calculating the inverse of the large matrix Σ_C and Q (both dimension are $n \times n$) is a very difficult due to complexity and memory constraint. In software R, the *solve.default()* function solves the system equation $QZ = I$ to find $Z = Q^{-1}$. The result of *help(solve)* command indicates that it called a Lapack routine DGESV (LAPACK 3.10.0, 2021) which uses the LU decomposition. Specifically, there are two steps to get the inverse of the large matrix Q:

(1) Using LU decomposition to transform $QZ = I$ to $UZ = I^*$ (where I^* is no longer a diagonal matrix), the complexity is $\frac{n^3}{3}$ for LU and $\frac{n^3}{2}$ for I^*, which is $\frac{5n^3}{6}$

(2) Solving for Z from $UZ = I^*$, the complexity is $\frac{n^3}{2}$.

So, the complexity for matrix inverse of a $n \times n$ matrix Q is $\frac{5n^3}{6} + \frac{n^3}{2} = \frac{4n^3}{3}$.

Ignore the lower order term, the total complexity of Equation is

$$\frac{4n^3}{3} + n \times n_a^2 + n \times n_c^2 + \frac{4n^3}{3} + \frac{n^3}{3} + n \times n_a + n \times n_c + \frac{n^3}{3} = \frac{10}{3}n^3 \qquad (2.11)$$

Which is $O(n^3)$.

On the other hand, for the improved algorithm, to get $(\Sigma_{\text{ac}})^{-1}\Sigma_{\text{acy}}$, we can use Cholesky decomposition:

$$\Sigma_{\text{ac}} = U^T U \qquad (2.12)$$

$$(\Sigma_{\text{ac}})^{-1}\Sigma_{\text{acy}} = backSolve\left(U, forwardSolve\left(U^T, \Sigma_{\text{acy}}\right)\right) \qquad (2.13)$$

Σ_{ac} is a $n_{ac} \times n$ dimensional matrix, where n_{ac} number of airborne and car survey data, and n is the number of pixels of the whole field to predict. Usually, $n_{ac} \ll n$.

The complexity of Equation (2.12) is $\frac{n_{ac}^3}{6}$; The complexity of Equation (2.13) is $\frac{n_{ac}(n_{ac}+1)}{2}n = \frac{1}{2}nn_{ac}^2$; The complexity of Equation (2.10) is $n \times n_{ac}$. So, the complexity in total is

$$\frac{n_{ac}^3}{6} + \frac{1}{2}nn_{ac}^2 + n \times n_{ac} = \frac{1}{2}nn_{ac}^2 \qquad (2.14)$$

In Equation (2.14), we ignored the lower order terms.

Compare Equation (2.14) with Equation (2.11), we can see that the complexity reduced by

$$\frac{\frac{10}{3}n^3}{\frac{1}{2}nn_{ac}^2} = 6.67\left(\frac{n}{n_{ac}}\right)^2 \qquad (2.15)$$

In our data, $\frac{n}{n_{ac}} \approx 10$, so the prediction expressed by our new approach can be 600~700 times faster.

Furthermore, once we put matrix Σ_{ac} into the memory, Equation (2.10) and (2.13) can be calculated in parallel. So the least memory required is the size of Σ_{ac}, which is only 1% the size of Σ_{y} in Equation (2.5).

However, there is one thing about how to calculate Σ_{ac}. Define

$$z_{ac} = \begin{pmatrix} z_a \\ z_c \end{pmatrix} = P \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = Py \tag{2.16}$$

$$\Sigma_{ac} = cov(z_{ac}, z_{ac}) = P\ cov(y, y)P^T = P\Sigma_y P^T \tag{2.17}$$

The complexity of matrix form of Equation (2.17) is

$$n_{ac} \times n^2 + n \times n_{ac}^2 = n^2 n_{ac} \tag{2.18}$$

If we want to bypass large matrix Σ_y due to memory restraint, we could calculate the elements of Σ_{ac}, and the complexity will be:

$$n_{ac}^2 \times (2n + n_{correlated}^2) = 2n \times n_{ac}^2 \tag{2.19}$$

$n_{correlated}$ is the number of pixels around pixel z_{ac} that need to be averaged in the airborne/car survey model, $n_{correlated}$ is also the non-zero elements in each row of matrix A in the first equation of (2.6).

The seeming complexity in Equation (2.19) is smaller than Equation (2.17). But when we calculate the each element of Σ_{ac}, we need to select the correlated pixels around (the non-zero elements in each row of matrix *A*), and build the covariance matrixes one by one. As there are nugget effect and 3 different types of land-use in our dataset, we need piles of for loops and if else loops in Python, and this makes the code hundreds of times slower than the matrix calculation in Equation (2.17). However, this can be done in parallel, and completely removed the memory restraint. If we can implement this method in Fortran or C, it should be much faster.

Another possible improvement for the future work is about Equation (2.16)~ (2.17): we encounter large matrix Σ_y when calculating the auto-covariance of z_{ac}. For a stationary Gaussian field, the auto-covariance of z_{ac} is only determined by the distances within each pair of pixels. So, there are two possible solutions: (1) make a table of the auto-covariance as a function of the distance, and just interpolate from the table to calculate each element of the auto-covariance matrix of z_{ac}; (2) decompose z_{ac} into weighted average of the pixels around z_{ac} (i.e. $z_{ac} = P'y'$, here P' is a small matrix and y' is a small vector) beforehand and each time, just use re-calculate the location of y' to calculate its auto-covariance $\Sigma_{y'}$(which is a small matrix), and then use $P'\Sigma_{y'}P'^T$ instead of $P\Sigma_y P^T$ to calculate the auto-covariance matrix of z_{ac}, this way will bypass the big matrix Σ_y. The two possible improvement will be tried in the future work.

3 Sensor network long-term optimization strategy

3.1 Background

After the accident, monitoring programs such as mapping project, ARM, etc. have been conducted continuously in Fukushima area. It has been found that the air dose rates (i.e., the ambient dose equivalent rate (μSv/h) at 1 m above the ground) has been readily decreasing since the accident (Saito, 2016, 2019), and the designated evacuation area has shrunk to 370 km^2, which is less than 3% of the Fukushima Prefecture (Fukushima Prefectural Government, 2017). By now, the monitoring program is expected to transition to long-term monitoring after 10 years. Since the extensive data accumulation has led to an understanding of contaminant distributions and mobility (Eddy-Dilek et al., 2014), the long-term monitoring has different objectives from the monitoring shortly after the accident for remedial activities.

Since the release event, radiation measurements and monitoring have been conducted continuously in this region. Monitoring has played a critical role in protecting the public, guiding decontamination efforts, and planning the return of evacuated residents. Radiation measurements have been carried out using various techniques and platforms. In addition to the conventional monitoring posts, new monitoring posts have been installed at more than 3,500 locations in the region, providing continuous, real-time air dose rates. To quantify the temporal changes in air dose rates, fixed-point measurements and soil sampling of undisturbed land have been done once or twice per year to provide the most accurate measurements of radiation dose rates (Mikami et al., 2015, 2019). In parallel, walk surveys (Andoh et al., 2018a), car surveys (Andoh et al., 2015, 2018b), and airborne surveys (Sanada et al., 2014, 2018) have been performed over the region once or twice a year to characterize the spatial distribution of radiation dose rates (Saito and Onda, 2015). The air dose rates are found to be significantly correlated with Cs-137 concentrations in soil (NRA, 2011a; Onda et al., 2015; Masoudi et al., 2019), so that they are considered as proxies of soil contamination in the region.

After 10 years, the monitoring program is expected to transition to long-term monitoring. The objectives of long-term monitoring are often different from monitoring during remedial activities, since such monitoring starts after extensive data accumulation has led to an understanding of contaminant distributions and mobility (Eddy-Dilek et al., 2014). The main long-term monitoring objectives are to (1) confirm the continuing reduction of contaminant and hazard levels, (2) provide assurance for the public, and (3) accumulate basic datasets for scientific knowledge and future preparation. At the same time, long-term monitoring is critical for detecting changes or anomalies in contaminant mobility (if they occur), or for detecting any unexpected processes or events. At the former nuclear weapon sites in the U.S.A. for example, monitoring activities have been continuing for more than 30 years, providing critical data and assurance for the local communities near the sites (Schmidt et al., 2018). This is particularly important for radiologically

contaminated sites where the environmental and health impacts are often exaggerated and false information can have a significant socioeconomic impact (Sawano et al., 2019).

The challenge of long-term monitoring is to build a cost effective and sustainable strategy by minimizing the cost associated with the number of monitoring locations or sampling, while maximizing the ability to meet the objectives listed above. In contrast to the monitoring activities during remediation, long-term monitoring has to be carefully planned, considering cost, spatial coverage, and the priorities of local communities and governments. Although there are a variety of factors to prioritize monitoring locations such as population density and socioeconomic and psychological factors, science-based methods could support or augment such prioritization. In particular, we may develop an optimization strategy for the radiation monitoring network—specifically by providing a logical way to determine the number and locations of different monitoring platforms.

Monitoring network optimization has been widely studied and applied in many fields, such as air-pollution monitoring, water-quality monitoring, snow-thickness measurements, and soil-pollution measurements. As a result of reviewing literature from 1978 to 2019, there have been many approaches that are developed for monitoring network optimization, such as spatial simulated annealing (SSA), genetic algorithms (GA), ant colony optimization (ACO), particle swarm optimization (PSO), the entropy-based Bayesian method, information theory, and surrogate-based optimization combined with random forests or kriging method. (More details regarding these algorithms and related literatures can be found in the supplementary material text S1 of Sun et al., 2020). In most of these approaches, optimization is done in two steps. The first step involves making predictions to create a map of contamination, using contaminant transport models, historical data, or the Kriging method. The second step involves optimally placing sensors based on objective functions; there are multiple algorithms available such as GA, ACO, PSO, and GA.

There have been several approaches proposed to optimize radiation monitoring networks (Melles et al., 2008; Heuvelink et al., 2010). Melles et al. (2008) developed an algorithm to optimize the air dose rate monitoring network of point measurement, by minimizing the average kriging standard deviation to find the optimal monitoring station locations. The approach by Heuvelink et al. (2010) is based on spatial simulated annealing to optimize the measurement of radionuclide concentrations spatially based on mobile measuring devices or sensors, by minimizing the expected weighted sum of false-positive and false-negative detection areas (the definition of positive and negative were in Heuvelink et al. (2010)).

Recently, environmental monitoring has been evolving to deploy airborne platforms and technologies, including drone and airborne measurements, that allow spatially extensive characterization and mapping (e.g., Wainwright et al., 2017). In particular, airborne radiation monitoring technologies have been advanced significantly in the past decade (Sanada et al., 2014; Sanada and Torii, 2015; Vetter et al., 2019). Working with multiple radiation survey datasets, Wainwright et al. (2017; 2018) has developed a multiscale

data-integration methodology – based on Bayesian hierarchical models and geostatistics – which has enabled the integration of datasets from these three kinds of surveys with different spatial coverage and footprints, as well as the creation of integrated maps of air dose rates over the region. Taking advantage of such airborne measurements, Oroza et al. (2016) proposed a novel machine-learning-based approach that optimizes the sensor-network configuration to capture the heterogeneous distribution of snow depths. There are now opportunities to improve the radiation monitoring based on spatially extensive datasets and spatial information.

The objective of this chapter is to develop a general methodology for optimizing regional-scale radiation monitoring, by extending the methodology developed by Oroza et al. (2016) for radiation monitoring. Specifically, for Fukushima, the focus is on either reducing the number of existing monitoring posts while keeping the high-priority locations (such as at schools and public facilities) and capturing spatial heterogeneity, or placing walk/car survey locations at minimum-but-sufficient locations. For simplicity, we assume in this study that the monitoring cost is proportional to the number of monitoring locations. In parallel, we aim to generalize this concept for any network applied to existing or potential contamination events. In principle, we assume that radiation monitoring networks are required to capture (1) the spatial heterogeneity of radiation dose rates; (2) key locations such as hospitals, schools, and public facilities; and (3) key features such as different land uses, terrains, and other factors that are known to control radionuclide mobility.

Our methodology is versatile: we can use the same approach to reduce the number of measurements from the existing points, as well as to establish new measurement locations, with some constraints such as accessibility (e.g., roads and public lands). Compared to the previous studies on radiation monitoring optimization, our unique contribution is that we use the spatially distributed radiation air dose rate map during the optimization rather than simple interpolation of point measurements. We demonstrate this methodology with a limited number of datasets at limited spatial scale, using an integrated radiation-dose-rate map created by Wainwright et al. (2017) as the true distribution of the air-dose rates.

3.2 Optimization Methodology

Since our methodology has been applied here for long-term monitoring, we assume that there has been an accumulation of datasets to aid in identifying the spatial distribution of air dose rates and in understanding of their changes. Specifically, in the Fukushima region, the air dose rates have been mapped extensively. Soon after the accident, the air dose rates indicated different decreasing tendency depending on the locations, since the mobile portion of radioceasium migrated at different speeds depending on, for example, surface land-cover types and human activities (Kinase et al., 2014; Saito et al., 2019). In

this analysis, we considered the geographical range, and also included currently known factors (i.e., land-cover type) that influence the radiocesium movement. In recent years after the migratory radiocesium has migrated, many studies have reported the spatially uniform reduction of dose rates over the region, except for a steeper decrease in the decontaminated region (Wainwright et al., 2018). This is because cesium is strongly bound to soil particles, and its mobility is quite limited in the environment. Therefore, we may assume that the current dose-rate map can be used to plan future monitoring activities. We use the current integrated map of air dose rates as a reference map to select monitoring locations (Wainwright et al., 2018). The steps of our methodology are shown as Fig. 3.1. Details of each step is discussed later.

In the following sections, we use the term "monitoring locations" or "monitoring points" to represent the locations for monitoring posts, survey data points, or dose-rate measurements. This is equivalent to "sensor locations" in Oroza et al. (2016) and other literature.

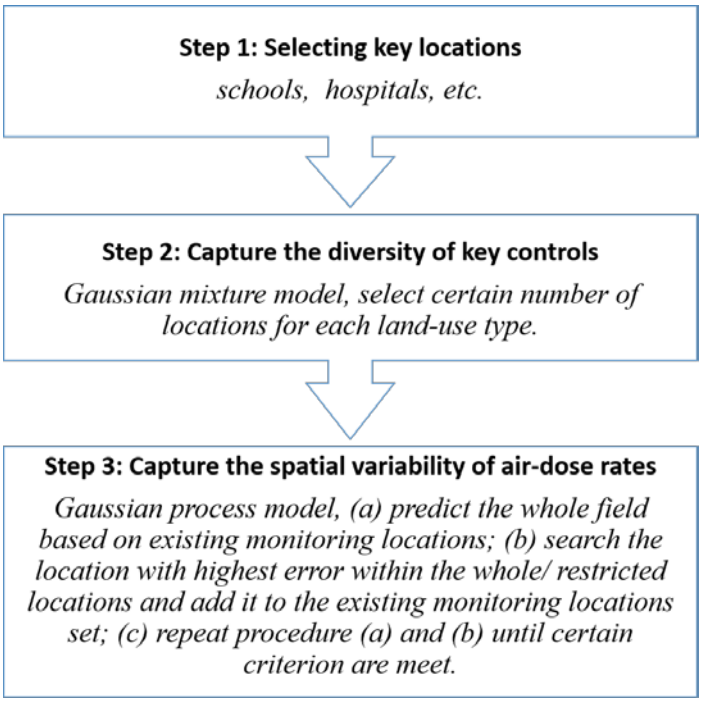

Fig. 3.1 Flowchart of the optimization Method

Step 1. Key locations

In the first step, we place monitoring points at key locations or pre-determined locations such as compliance points, schools, or hospitals. Although their number and locations can be negotiable, it is often the case that there are a set of locations required for monitoring, based on regulations or public need.

Step 2. Capture the diversity of key controls

There are key environmental controls that are known to affect the reduction of air-dose rates or the heterogeneity of the air-dose rates, such as land-cover types (Saito et al., 2019). To capture such effects more effectively, we may want to distribute monitoring points at the most representative locations of different parameters or features, such as elevation, distance/direction from the source, or spatial extent (latitude/longitude). This allows us to diversify the monitoring locations across different environmental variables, which is particularly important for scientific research and understanding, as well as for finding any additional or unexpected effects in the future. Thus, after establishing key locations in Step 1, in Step 2 we add more monitoring locations to capture key features.

Following Oroza et al. (2016), we use the Gaussian mixture model (GMM) to determine the monitoring locations so as to identify the most representative locations. A GMM assumes that a feature space (e.g., the combined $x = [x^{lat}; x^{lon}; x^{elevation}; x^{direction}; x^{distance}; x^{landuse}]$) is a product of a finite number of latent (unobserved) components (i.e., measurements) that follow Gaussian distributions, where x^{lat}, x^{lon}, $x^{elevation}$, $x^{direction}$, $x^{distance}$ and $x^{landuse}$ are the raster datasets for latitude, longitude, elevation, direction from the plant, distance from the plant and land use type, respectively. The purpose of using a GMM here is to find the representative values in feature space, (i.e. the center points of clusters) rather than to quantify the parameter uncertainty. The monitoring network's ability to observe each point in the feature space is represented using a multivariate normal distribution: $N(x \mid \mu, \Sigma)$ where μ and Σ are the mean and covariance, respectively. This is the parametric expression for each component of the mixture. The mean of the normal distribution is selected to be the measurement location in the feature space as a representative location. Multiple Gaussian distributions (multiple measurement locations) are combined and weighted with mixing parameters π_m from an ensemble of M mixture elements:

$$p(x) = \sum_{m=1}^{M} \pi_m N(x \mid \mu_m, \Sigma_m) \tag{3.1}$$

where $p(x)$ is the probability density at x, and

$$\sum_{m=1}^{M} \pi_m = 1 \tag{3.2}$$

We use the expectation maximization (EM) algorithm to place the Step 2 sets of monitoring locations (McLachlan and Peel, 2004; Pedregosa et al., 2011). The EM algorithm is an iterative process in which the algorithm identifies the most likely parameter estimates for the mixture of multivariate normal distributions to represent the data. Within this algorithm, we use a spherical covariance function to update the model weights, covariance, and means with each iteration. Once the maximization step no longer increases the log-likelihood, the

process terminates, and the optimized monitoring locations have been found. We then perform a nearest neighbor search through the full feature space (i.e., not subsampled) to find the physical location that most closely matches the features of each mean estimate.
The previous studies in this region (e.g., Saito et al., 2019 and Kinase et al. 2014) have shown that the land-cover type is known to influence the environmental decay of the air dose rates. Since GMM does not include categorical variables, we assign a fixed number of monitoring locations in each land-cover type and distribute them according to the other numerical features within each land-cover type. The feature matrices for each subregion are extracted and scaled before the GMM is fit in each region.

Step 3. Capture the spatial variability of air-dose rates

In this step, a Gaussian process model (GPM) is used to add monitoring locations to capture the spatial variability across the region, following Oroza et al. (2016). A Gaussian process model is based on spatial auto-correlation and covariance models, which are equivalent to the geostatistical model used in Wainwright et al. (2017). Although Oroza et al. (2016) included the dependency of the target variables on environmental variables such as elevation, we use only the spatial correlations, since the spatial distribution of the radiation dose rates are largely governed by the plume path and initial deposition—although there are also some minor effects caused by environmental controls such as elevation, land use, and other parameters which can be expended to depend on needs. We assume an exponential covariance model, the parameters of which are simultaneously estimated. We assumed the same parameters for the domain without considering the land cover types, which is different from Wainwright et al. (2017).

We add one monitoring location at a time, sequentially based on the estimation result. With each iteration, the air dose-rate map is estimated using GPM, conditioned on the current locations. The values at the monitoring locations are taken from the reference map, which in this case is the integrated dose-rate map developed by Wainwright et al. (2017; 2018). The difference between the estimated and reference map is quantified by the absolute error at each pixel. A new monitoring location is placed at a randomly selected pixel within the top three percent of the absolute error. We note that such randomness is necessary to avoid the effect of outliers, since the maximum error is often affected by such outliers. At each iteration, we compute the Root Mean Square Error (RMSE) over all the pixels that do not have monitoring locations. RMSE is used as a summary statistic to quantify the overall estimation error of this map. This step is repeated until the RMSE converges, the desired number of monitoring locations are placed, or the RMSE falls lower than the required threshold. We consider that the convergence-based criteria could be most appropriate, since it is often difficult to define the number of monitoring locations based on the absolute RMSE values. We may define the minimum-but-sufficient number of monitoring locations based on the convergence of RMSE, such that RMSE with the reduced number of monitoring locations is within a certain range (i.e., a few percent) from the one of the existing locations.

The use of the estimation error is different from Oroza et al. (2016) or other studies (Araki et al., 2015; Masoudi et al., 2019; Zhuang et al., 2011), who placed monitoring locations

based on the estimation variance. The estimation variance (or often called kriging variance) is calculated based on the interpolation of point measurements without using the actual values in the reference map. In our case, the reference map – i.e., the integrated map of air dose rates – is available over the region (Wainwright et al., 2016), and it is known that the relative spatial distribution of the air dose rates does not change over time significantly. We hypothesize that, using the estimation error (as the difference between the reference map and the interpolated map), we can maximize the use of information currently available and we can further improve the monitoring network compared to using the estimation variance. We evaluate the impact of the difference between using the estimation error and variance in a synthetic scenario.

We have implemented our algorithms using the Scikit-learn package in PYTHON (Pedregosa et al., 2011). We have made multiple improvements in the algorithms compared to Oroza et al. (2016), such as restricting monitoring locations (for example, representing the availability of power, and the accessibility of locations and existing monitoring locations).

3.3 Result and Discussion

We demonstrated our methodology by using the datasets in the designated evacuation area (as of March 2017). We used the 2016 integrated map created in Wainwright et al. (2018), along with other spatially extensive data, including elevation, land-cover type, and distance and direction from the NPP (Fig. 3.2). The pixel size was 50 m by 50 m. We used the high-resolution land-use and land-cover map of Japan (version 14.02) created by the Japan Aerospace Exploration Agency (Takahashi et al., 2013). In this demonstration, we focused on the methodology development, aiming to test our algorithm performance. We created a hypothetical set of priority locations to be used in Step 1.

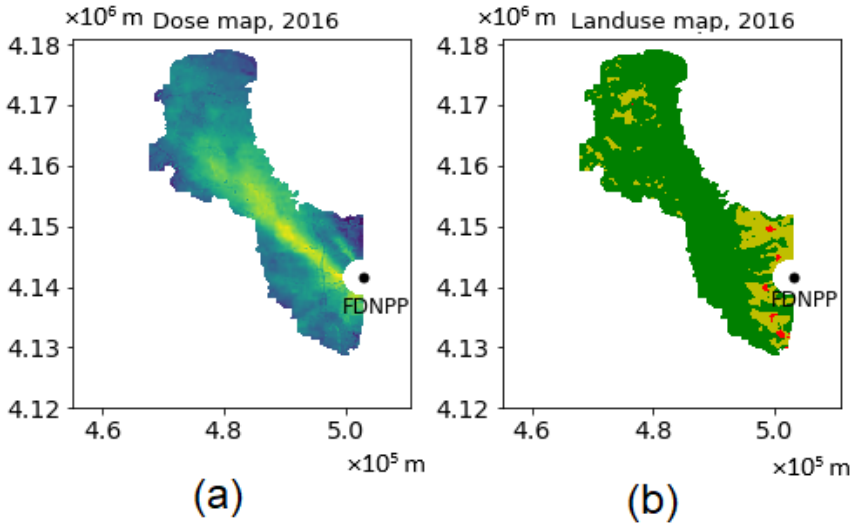

Fig. 3.2 Input data maps: (a) 2016 integrated air-dose-rate map in log10 microSv/hr, (b) land-cover map. In (b), the green region is forest, the yellow region is cropland, and red region is urban area. The unit of coordinates is meters (m), the black dots in each subplot are the location of Fukushima Daiichi Nuclear Power Plant (FDNPP).

To represent different uses, we considered two cases: (1) across the domain without any location restrictions, (2) at the limited locations selected in advance. In Case 1, we considered all the pixels that are candidate locations for monitoring. Case 1 was used mainly to demonstrate the algorithms and to explore the effect of parameters within the optimization algorithms. Case 2 mimicked the situation in which the goal would be to reduce the number of existing monitoring locations, or the restricted locations along the roads or accessible locations.

Case 1: Placement without location constraints

Fig. 3.3 shows the monitoring locations at each step for Case 1. As mentioned above, the Step 1 locations are hypothetical for the demonstration purpose. We assume that the four Step-1 locations are the prioritized locations that are fixed a priori (Fig. 3.3a). The monitoring points are added to diversify various environmental properties in Step 2, so that the monitoring locations are distributed widely throughout the area (Fig. 3.3b). We assume ten locations in each land-cover type, so that 30 points are placed in total. The points are distributed over the domain to cover the range of dose rates and space. In Step 3, the algorithm adds 250 points to capture the heterogeneity in the dose rates, so that it places monitoring locations in-between the Step 1 and Step 2 points (Fig. 3.3c), as well as in the region where the spatial heterogeneity is high and the dose rate changes more rapidly in a short distance (e.g., the region near the power plant). There are four points in Fig. 3.3a, 34 points in Fig. 3.3b, and 284 points in Fig. 3.3c.

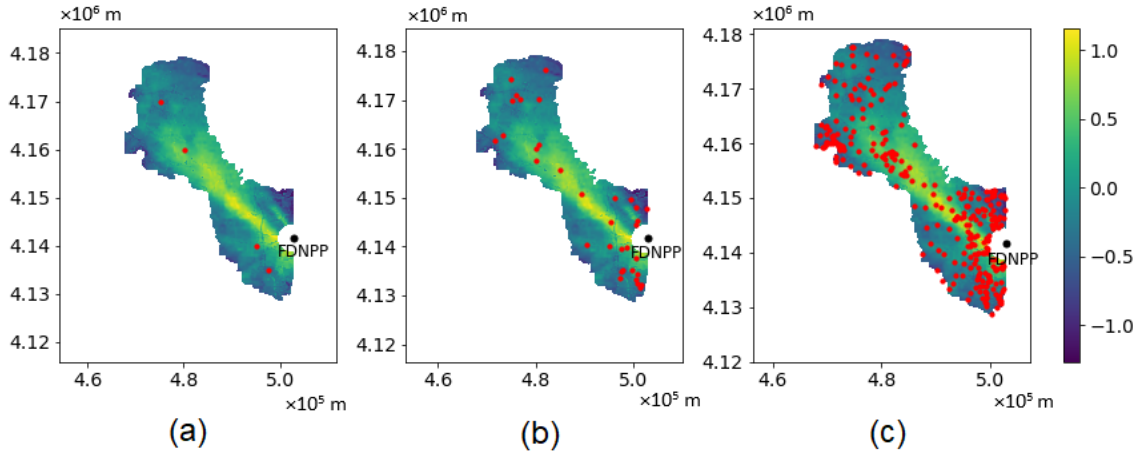

Fig. 3.3 Proposed monitoring locations over the 2016 integrated map (in log10 microSv/hr) in Case 1 after: (a) Step 1, (b) Step 2 and (c) Step 3. In the figures, the red circles are the monitoring locations.

The overall estimation error (RMSE) is plotted against the number of monitoring locations in Step 3 (Fig. 3.4). Fig. 3.4a examines the effect of the randomness, since the point at each iteration is selected randomly within the pixels that have the top 3% estimation errors. RMSE decreases rapidly at the beginning and converges to a certain value. This is because once there are enough monitoring locations to capture the heterogeneity, additional locations have a diminishing effect. In addition, such RMSE convergence is attributed possibly to random errors in the dose-rate measurements or spatially uncorrelated variability in the dose-rate distribution. All the curves are fairly similar, suggesting that the randomness effect is quite minimal within the optimization algorithm.

In addition, we compare several numbers for the Step-2 monitoring locations; five, 10, and 20 in each land-cover type (i.e., the initial number in Step 3 is 19, 34, and 64, respectively), as shown in Fig. 3.4b. Fig. 3.4b illustrates that when the number of monitoring locations is high in Step 2, the initial RMSE is low, but it converges to the same value. The number of initial monitoring locations does not have a significant impact on the final distribution and RMSE, or on the ability of the monitoring network to capture the heterogeneity of the dose rates.

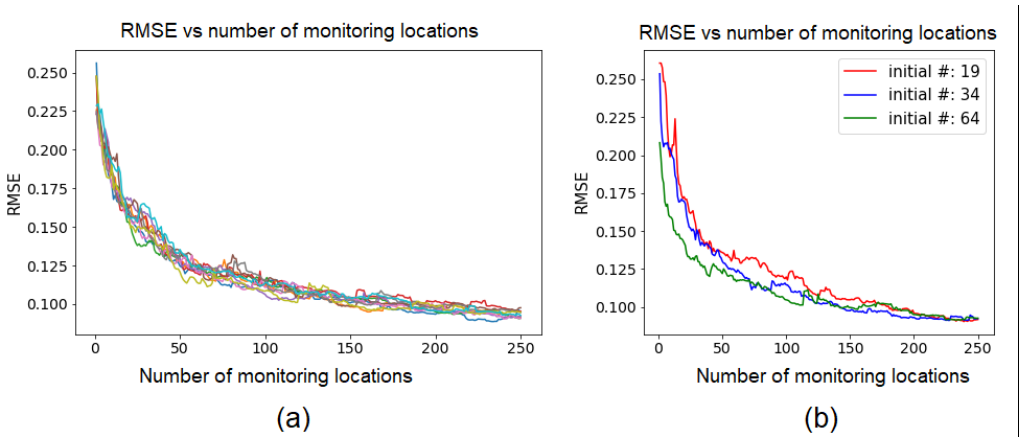

Fig. 3.4 RMSE vs number of monitoring locations in Step 3 in Case 1: (a) initial monitoring locations number is 34, random sampled top 3% highest estimation error, MC simulated 10 times; (b) random sampled top 3% highest estimation error, with initial monitoring locations number 19, 34, 64.

In addition, we investigated the effect of the selection criteria to select the next monitoring location in Step 3. The original algorithm in Oroza et al. (2016) selected the next location based on the estimation variance from GPM—i.e., choosing one location among the top 3% variance pixels or the largest variance pixel. We proposed an alternative for choosing the next location based on the estimation error computed as the difference between the reference and interpolated maps in Step 3.

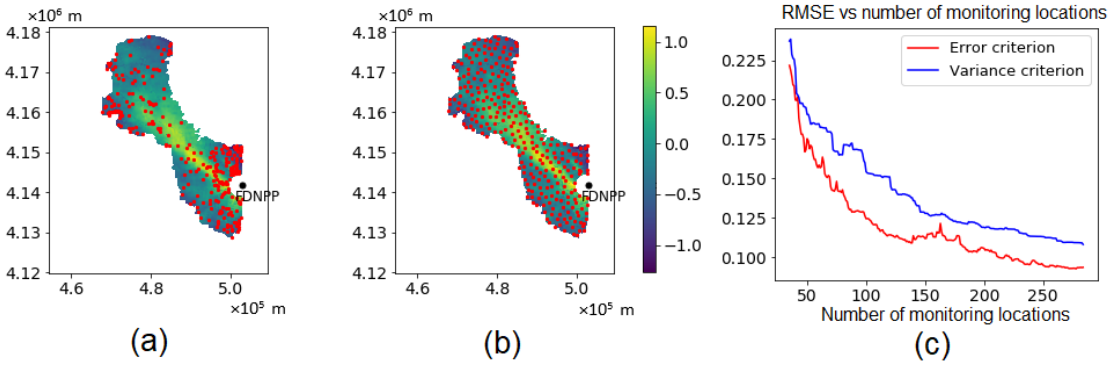

Fig. 3.5 Monitoring locations configurations by choosing (a) the top 3% of the estimation errors and (b) the top 3% estimation variance and (c) RMSE curves using error criterion vs variance criterion.

The two criteria make a large difference in terms of the RMSE and spatial configuration of monitoring locations. When estimation error is used as the criterion (Fig. 3.5a), there are many clusters in the map. The clusters tend to be located where the radiation dose rate is more heterogeneous over a short distance. In the region where the spatial heterogeneity is high, the interpolation becomes high, and more monitoring locations are needed to capture the spatial heterogeneity. On the other hand, when estimation variance criterion is used (Fig. 3.5b), the monitoring locations are more uniformly distributed over the

domain. As a property of GPM, the highest predicted variance is the middle points among neighboring sensors. Therefore, this variance-based criterion tends to choose locations in the middle of an existing network, which ultimately results in a uniform sensor network (Fig. 3.5b). In Fig. 3.5c, the estimation error-based criterion yields a more rapid decrease in RMSE than the variance-based criterion, as well as a smaller RMSE when the RMSE is converged. This result suggests that the estimation error-based criterion can add points more effectively where the heterogeneity is large, and can capture the heterogeneity with fewer numbers of monitoring locations.

In our algorithm, we randomly selected one location among the top 3% with largest error instead of choosing the largest one to reduce the influence of outliers. However, the choice of 3% seems rather arbitrary, and therefore this parameter has to be evaluated. We consider that such random selection can effectively attenuate the effect of outliers, although such a selection scheme could also reduce the prediction power, since the algorithm could choose the pixels with lower estimation error—there is an apparent trade-off. To evaluate what is the best sampling scope for our algorithm, we tested different percentages: 0.2%, 1%, 3%, 5%, 7.5%, 10%, and compared the reduction of RMSE as a function of the number of monitoring locations. Fig. 3.6 shows that the reduction is the most effective between 3% and 7.5%. The RMSE is higher for the smallest percentage (0.2%) due to the outlier effects, and also for the largest percentage (10%) due to the fact that the large estimation-error pixels are missed. This confirms the presence of the trade-offs, and the parameters have to be optimized for each case.

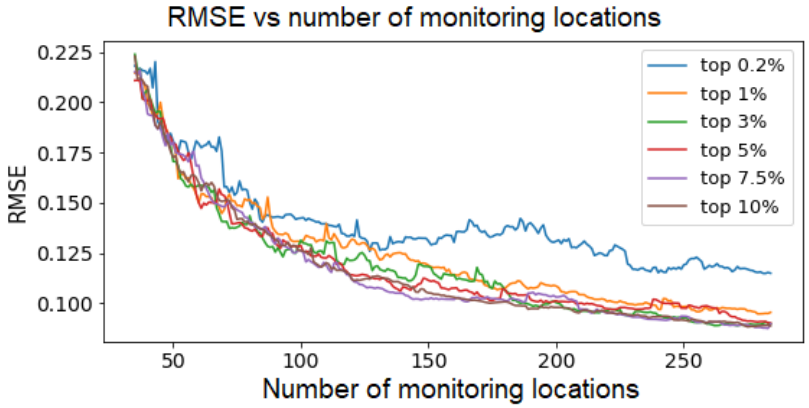

Fig. 3.6 RMSE as a function of the number of monitoring locations for different parameters within the error-based criterion. In the legend, top 10% means randomly sampling one pixel out of the pixels with top 10% highest error for next sensor, etc.

Case 2: Placement with the location restriction

In Case 2, we demonstrated the monitoring network optimization with location restriction. We used actual monitoring post locations except for Step 1. In Step 2, we added 10 locations for each land-use type. In Step 3, we selected 100 out of the 255 existing monitoring locations. Fig. 3.7 shows the sampling locations at each step for Case 2. Similar to the monitoring configuration without location restriction (Fig. 3.3c), the monitoring locations are concentrated in the region where the spatial heterogeneity is high. The difference is that there is a missing region around Easting = 4.9 x 105 m, where there are no existing monitoring locations. This difference may suggest that locations that are currently missing but are needed to capture the regional-scale heterogeneity of radiation dose rates.

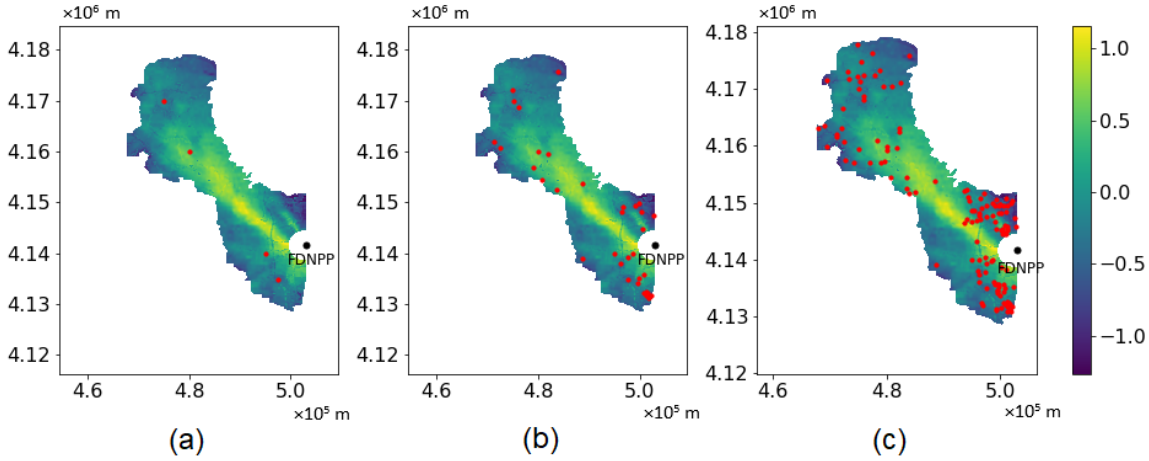

Fig. 3.7 Proposed sampling locations over the 2016 integrated map (in log10 microSv/hr) in Case 2 after: (a) Step 1, (b) Step 2 and (c) Step 3. In the Figures, the red dots are the sampling locations.

Fig. 3.8 shows the effect of the randomness within the algorithm and the number of Step 2 locations, when the locations are restricted to the existing monitoring locations. In Fig. 3.8a, after repeating the simulations ten times, the RMSE curves are plotted against the number of monitoring locations. The RMSE decreases with fluctuation at the beginning and converges to a certain value. The converged value is higher than the no-restriction case in Fig. 3.4a, and the RMSE converges slowly compared to the no-restriction case, since the number of pixels that can be chosen is much smaller. The existing monitoring locations are not necessarily capturing the spatial heterogeneity of contamination. In Fig. 8b, we compare several numbers of Step-2 sampling locations: five, ten, and twenty in each land-cover type (i.e., the initial number in Step 3 is 19, 34, and 64, respectively), as shown in Fig. 3.8b. As consistent with the no-restriction case (Fig. 3.4b), the number of Step 2 locations do not affect the convergence of RMSE.

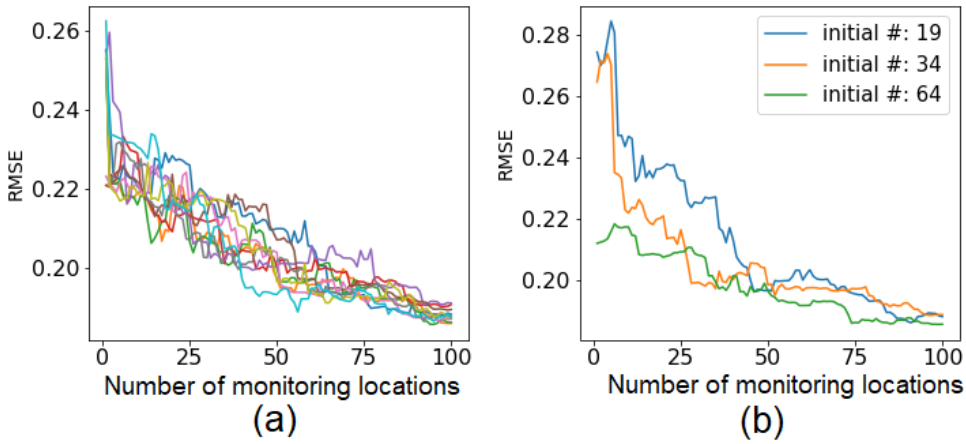

Fig. 3.8 RMSE vs number of monitoring locations in Step 3 in Case 2: (a) initial locations number is 34, random sampled top 3% highest error, MC simulated 10 times; (b) random sampled top 3% highest error.

Since our algorithm has a random selection (e.g., within top 3% of largest errors) within each iteration, there could be randomness in the final monitoring locations. There is a concern that random simulations may yield totally different design networks. We need to evaluate how this randomness affects the monitoring locations. We created a probabilistic map—the probability of each location to be chosen as a monitoring location — to represent the randomness within the algorithm. Using the Monte Carlo simulation, we created the 100 sets of monitoring locations that are equally likely (Fig. 3.9). The probabilities are computed by the frequency of being selected in the Monte Carlo simulations. Within each set, 100 locations were selected out of 255 pre-selected locations, since the RMSE appears to converge around 100 locations.

Fig. 3.9a shows that RMSE generally decreases as a function of the number of monitoring locations and converges to a similar value. In the probability-based monitoring network (Fig. 3.9b), there are some locations that are always chosen (red dots in Fig. 3.9b), while some are less likely to be selected (purple dots in Fig. 3.9b). These more-selected locations tend to be located within the high heterogeneity region. In addition, the spatial pattern is consistent with Fig. 3.7c, which is just one instance of the simulation. Fig. 3.9c shows the probability of being selected for each location sorted from high (1.0) to low (0.0). For example, there are 28 locations (from 0 to 27, around 11 percent out of total) that are 100% (always) selected, while 78 locations (from 177 to 254, around 30 percent out of total) are never selected during the 100 simulations. The slope of the distribution in Fig. 3.9c reflects the ambiguity of our algorithm, i.e., steeper means less randomness. The steep curve results suggest that the randomness might not affect the monitoring location significantly, and the algorithm can identify both the locations that are highly important, as well as the locations that have a negligible impact on the ability to capture spatial heterogeneity.

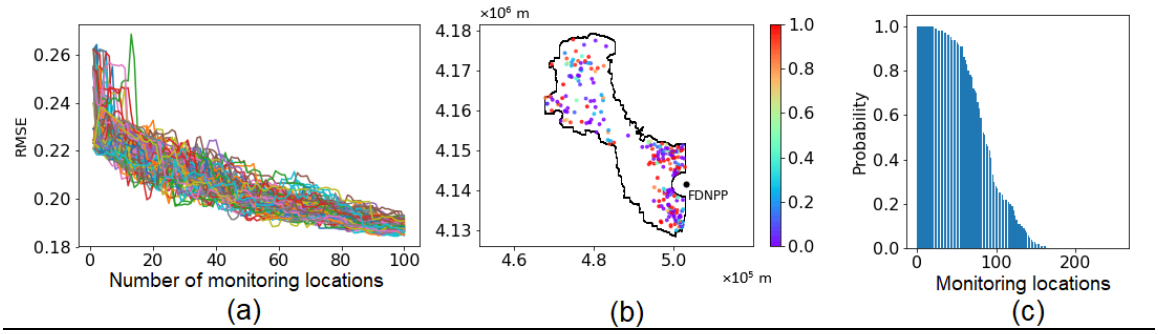

Fig. 3.9 Results from generating 100 sets of monitoring locations based on the MC simulation: (a) the RMSE curves of the 100 simulations, as a function of monitoring locations, (b) probabilistic map of the monitoring locations among monitoring posts based on 100 simulations, and (c) probability of each location sorted from high (1.0) to low (0.0). In (b), the color of each dot is an indicator of probability.

3.4 Conclusion

In this work, we have developed a methodology for optimizing monitoring locations of air dose rates at the regional scale. This methodology can be used as a general methodology either for reducing the number of existing monitoring locations (such as monitoring posts), or for optimally placing mobile measurements, such as car or walk surveys. Three steps are taken in order to determine monitoring locations in a systematic manner: (1) prioritizing the critical locations, such as schools or regulatory requirement locations, (2) diversifying locations across the key environmental controls that are known to influence contaminant mobility and distributions based on a Gaussian mixture model, and (3) capturing the heterogeneity of air dose rates across the domain based on a Gaussian process model. We use the integrated dose-rate map from Wainwright et al. (2017; 2018) as the reference map and distribute the sampling in such a way as to capture the heterogeneity of the reference map.

Our results have shown that this approach enables us to add or subtract monitoring locations in a systematic manner such that the heterogeneity of air dose rates is captured by the minimal number of monitoring locations. We acknowledge that our algorithm does not include socioeconomic factors that influence overall exposure dose to the public. The population density or traffic volume (along each road) can be additional spatial layers that are readily available and can be included (such as Sun et al., 2019). The algorithm can accommodate other factors such as agricultural information or key facilities. At the same time, capturing the overall spatial distribution of air dose rates is important for risk assessments or decontamination planning. In fact, many people in this region enter the non-populated forested area for edible wild plants or for forestry (Miura, 2016). We consider that our algorithm in this paper is the first step of monitoring optimization by capturing the spatial heterogeneity; we can add other information and their priority weights according to the user's needs.

In addition, we acknowledge that this algorithm would not provide additional protection or remediation methods. However, having an accurate map of contamination allows people to avoid highly contaminated areas or to concentrate decontamination resources to appropriate areas. In addition, long-term monitoring is important to provide the correct information about the stability of the contaminant distribution, and the reduction of radiation level to the people in the other regions. Improving air dose rate mapping with the limited number of monitoring locations, hence, contributes significantly to protecting public health as well as to supporting the local economy.

4 Spatial-temporal dose rate distribution Prediction

4.1 Introduction

During the Fukushima accident, radioactive nuclides were released into the environment and deposited around the area of FDNPP; mostly within the 80-km radius. Radiocesium is considered the main nuclide that contributes to the exposure dose (IAEA, 2015). Since the accident, multiple agencies have performing radiation measurements and monitoring in this region (Saito et al., 2018; Kinase et al., 2017; Andoh et al., 2018), which resulted in a large volume of well-archived radiation air dose rate data (JAEA, 2021). These datasets have been analyzed extensively for the purpose of (1) confirming system stability and the continuing reduction of contaminant and hazard levels, (2) providing exposure evaluation for the public's return to the evacuation zone, (3) accumulating the basic datasets for scientific knowledge and future preparation. It has been found that the air dose rates near FDNPP decreased to approximately 10% of its initial level by November 2018 (Saito, 2016, 2019; Andoh et al., 2020). As time goes on, the monitoring program is expected to transition from short-term monitoring to long-term monitoring. Different with short-term monitoring, the challenge of long-term monitoring is building a cost effective and sustainable strategy by minimizing the cost associated with the number of monitoring locations or sampling, while maximizing the ability to meet the objectives of monitoring, Sun et al., (2020) developed a sensor network optimization method to accommodate the transition to long-term monitoring.

However, Sun's work was based on the assumption that the dose rates distribution map is stabilized without significant variation. Even though the general tendency of air dose rates and deposition densities in the 80 km zone is downward, their decreasing speed were found to vary significantly according to location (Saito et al., 2018, 2019; Andoh et al., 2018). Specifically, dose rates in the urban area decrease faster that in the area of cropland or nature forest. Saito et al. (2018) presented a detailed summary and analysis with respect to these variabilities across the region.

In parallel, a data-driven environmental decay model has been developed to predict the decay of radiation air dose rates in the environment (Wainwright et al., 2019). Kinase et al. (2017) built a prediction model characterized by ecological half-lives of radioactive cesium for different land-use type, and predicted distribution maps of ambient dose equivalent rates for the next 30 years. Sanada et al. (2018) evaluated the ecological half-life of dose rate, which is a parameter in the double exponential formula, based on airborne radiation monitoring and use this function to predict the air dose rate on average. Andoh et al. (2020) modified the double exponential formula of ecological half-life and developed a two-group model to model the acceleration of the decrease in the air dose rates observed in the evacuation order area.

Although these data-driven models mentioned above can predict the decay trend of the average dose rates within the same land-use category, spatial-temporal correlation factors

were not considered within previous work. To incorporate this spatial heterogeneity factor and enhance the practicability and efficiency of our optimization method, it is necessary to integrate the spatial-temporal factor during optimization.

In geostatistics, Gaussian process model (GPM) is a method of interpolation for which the interpolated values are modeled by a Gaussian process. Oroza et al. (2016) used the GPM to model the snow depth around the network. Following Oroza et al. (2016), Sun et al. (2010) used GPM to find locations to add monitoring locations to capture the spatial variability across the region. Kalman filter is an algorithm that uses a series of measurements observed over time to filter out the noise and other inaccuracies and obtain a more accurate estimation of unknown variables than those based on a single measurement alone. Kalman filter is a widely applied concept in many fields, such as time series analysis in data processing, navigation, and control of vehicles, spacecraft, etc. Schmidt et al. (2018) presented a Kalman filter-based framework to establish a real-time in situ monitoring system for groundwater contamination based on in situ measurable water quality variables, such as specific conductance (SC) and pH. In this paper, GPM is applied to predict unobserved location conditioning on observed locations while Kalman filter is to fill the gap of observed locations using observation of other observed locations.

The goal of this study is to develop a spatiotemporal data integration method for creating an integrated radiation air dose rate map over space and time, by extending the Bayesian hierarchical method developed by Wainwright et al (2017; 2018). The method is based on GPM which takes advantage of spatial auto-correlation and covariance models to predict variables in unobserved locations. In order to incorporate the spatial correlation information of the dose rates field with the temporal information, we combined GPM and Kalman filter to predict the spatial-temporal dose rate time-series at a given time and location. Specifically, Kalman filter was used to predict the temporal evolution of dose rates for pixels with real-time monitoring device (to make them more precise) or only segment of historical data (to fill the gap between segments), and then GPM was applied to interpolate the dose rates for pixels without observations at each time point. This process was repeated for each time point and the spatial-temporal dose rates distribution could be constructed. Through this way, our algorithm utilizes the spatial correlation information in prediction, which make it different from the models mentioned at the beginning of this paragraph. We demonstrate our approach using the dose rate datasets from the evacuation zone (as of March 2017) between July 1st 2017 and Nov 22nd 2017.

4.2 Data Description

In this paper, the original dataset is the monitoring poster dataset acquired by JAEA. Although JAEA accumulated a large number of datasets of air dose rates since the accident, only a subset of datasets was used in this study to demonstrate our approach. Fig. 4.1 shows the integrated dose rates map of 2014 provided by Wainwright et al. (2017), with the color indicates the dose rates level under logarithm scale based on 10. The size of each grid is 50m×50m, the UTM position of the origin is (460101, 4120001). The dashed box with dimension of 8550m ×18750m is the study area where contains 17

monitoring posts with data collected between July 1st 2017 and Nov 22nd 2017. In addition, the car survey data from TEPCO was also used to verify the correlation between slope and intercept. In the car survey data set we used, there are 9626 locations, with 29 time-points for each location, starting from Aug 2nd, 2011 and end on April 3rd 2018.

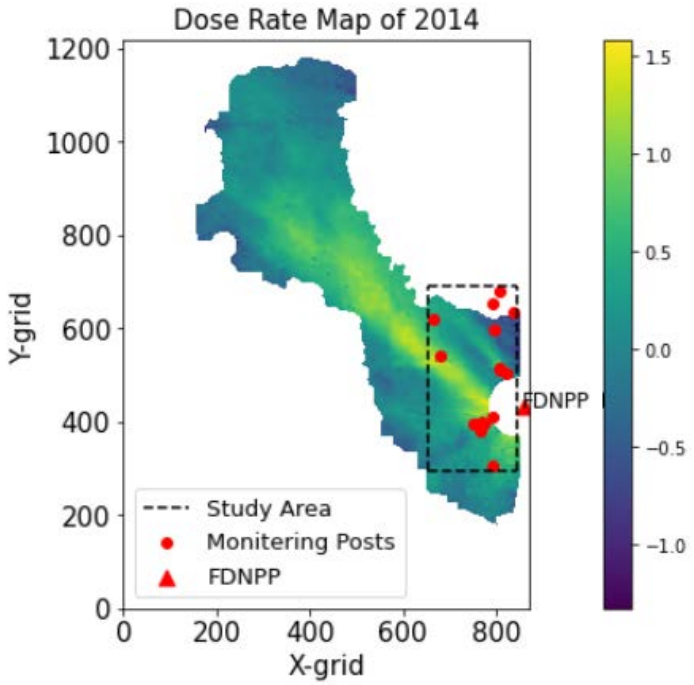

Fig. 4.1 Diagram of the Study Area, the background is the integrated dose rates map of 2014.

4.3 Methods and Results

4.3.1 Environmental Decay Model:

In our study, the area was discretized into the same grid (with grid size 50m×50m) as the integrated map. All the pixels can be classified into three categories: (1) pixels that we have real-time monitoring posts which can provide continuous dose rates; (2) pixels at which only some segments of historical data are available, such as fixed points measurement, car survey, airborne survey; and (3) pixels without any measurements except the interpolated value from the Integrated Map. Assume that there are n pixels in Category I&II and m pixels in Category III. Since the total number of pixels is $(n + m)$, the average dose rates within each pixel could be indicated as a vector $x(t)$ at time t, vector $x(t)$ has $(n + m)$ elements. We assume that the dose rates time-series for pixels under logarithm scale can be modeled by a linear trend plus a fluctuation term. For a single time step t_i, the dose rates can be represented as

$$\boldsymbol{x}(t_i) = \boldsymbol{u} \cdot t_i + \boldsymbol{b} + \boldsymbol{\varepsilon_i} \tag{4.1}$$

Although other papers used the two-component models (Kinase et al., 2017; Sanada et al. 2018;), the decay can be represented by log-linear trends after the ^{134}Cs contribution is reduced as ^{134}Cs has a much shorter half-life (2.06 years) than ^{137}Cs (30.2 years). As a result, compared with two-component models, the linear model makes the algorithm simpler without losing accuracy.

In Equation (4.1), $\boldsymbol{x}(t_i)$**,** $\boldsymbol{u}$**,** $\boldsymbol{b}$ and $\boldsymbol{\varepsilon_i}$ are all vectors of $n + m$ elements. We consider discrete time $t_i (i = 0, 1, \ldots)$ which is a scale with subscript i denotes the i_{th} time step, $\boldsymbol{x}(t_i)$ denotes the dose rates at t_i, $\boldsymbol{\varepsilon}_i$ is the fluctuation at t_i, and $\boldsymbol{u}$**,** $\boldsymbol{b}$ are the slope and intercept of the linear trend. In our study, we assume that $\boldsymbol{\varepsilon_i}$ is normally distributed and independent between different time steps, but that its components are spatially correlated among different pixels at the same time step:

$$cov(\varepsilon_i, \varepsilon_j) = \begin{cases} \Sigma, & when\ i = j \\ 0, & otherwise \end{cases} \tag{4.2}$$

where $\boldsymbol{\varepsilon_i}$, $\boldsymbol{\varepsilon_j}$ are all vectors of $n + m$ elements and Σ is a $(n + m) \times (n + m)$ matrix which indicates the fluctuation parts are space correlated.

The decay rate $\boldsymbol{u}$ which is determined by the slope of the linear trend (or decay rate) can change over locations. The decay rate is known to be correlated with initial dose rates, landuse type, spatial position and other factors. For Category I&II pixels, $\boldsymbol{u}$ can be fitted from the measurements. For Category III (unobserved locations), we used random forest (trained on Category I and II pixels) to predict $\boldsymbol{u}$. Random Forest is a machine learning method for classification or regression that operate by constructing a multitude of decision trees at training time and output the mode of the classes (classification) or mean/average prediction (regression) of the individual trees (Tin Kam, 1995).

4.3.2 Time-series estimation:

Step one: predict the temporal dose rates in the pixels in Category II for a time step

In this step (predicting the temporal dose rates for Category II pixels,) we found four possible solutions: Gaussian Process Model, Linear Regression, Kalman Filter and our own improved filter. Here, we introduced the details of each solution and a comparison from the viewpoint of mathematics was made in Appendix II.

(1). Gaussian Process Model (GPM):

As the dose rates from different locations are spatial correlated, the GPM is an option to interpolate the dose rates for each unmeasured location at each time step, just as we did in Chapter 3. Specifically, for each time step, we collect available measurements and (I) estimate the spatial parameters (such as variance and length scale), and then (II) use the estimated spatial parameters to construct the covariance matrix and apply the kriging method for interpolation.

However, there are many assumptions in the GPM which can't be strictly followed. For example, we assume the covariance is a function only of the distance without considering the fact that the covariance for two pairs of locations may be different even the two pairs have the same distance within. We also assume that the nugget effect to be a very small value based on our experience (such as 0 or 10^{-8}, etc.). These assumptions will deteriorate our model's prediction accuracy.

(2). Linear Regression

Based on the understanding of GPM, some improvements can be done for this problem. In this problem, we have some historical measurements of the time-series for some locations, and we can extract the covariance information from the historical data rather than estimating from the data at each time step.

Assume we have the covariance matrix as $\mathrm{cov}\left(\begin{pmatrix}\mathrm{X}_1\\ \mathrm{X}_2\end{pmatrix},\begin{pmatrix}\mathrm{X}_1\\ \mathrm{X}_2\end{pmatrix}\right)=\Sigma=\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}$, where $\mathrm{X}_1=\begin{pmatrix}\mathrm{x}_1\\ \vdots\\ \mathrm{x}_\mathrm{k}\end{pmatrix}$ is vector of k available measurements, and $\mathrm{X}_2=\begin{pmatrix}\mathrm{x}_{\mathrm{k}+1}\\ \vdots\\ \mathrm{x}_{\mathrm{k}+\mathrm{l}}\end{pmatrix}$ is vector of l unmeasured location to be estimated, then the prediction of X_2 conditioning on X_1 is given by

$$\mathrm{X}_{2|1}=\mu_2+\Sigma_{21}\Sigma_{11}^{-1}(\mathrm{X}_1-\mu_1) \tag{4.3}$$

The above equation implies that $\mathrm{X}_{2|1}$ is a linear combination of elements of X_1:

$$\mathrm{X}_{2|1} = \begin{pmatrix} \beta_{0,1} & \beta_{1,1} & \cdots & \beta_{k,1} \\ \vdots & \vdots & \vdots & \vdots \\ \beta_{0,l} & \beta_{1,l} & \cdots & \beta_{k,l} \end{pmatrix} \begin{pmatrix} 1 \\ x_1 \\ \vdots \\ x_k \end{pmatrix} \tag{4.4}$$

Or, for a specific (i_{th}) location:

$$x_{i|1,\cdots k} = \beta_{0,i} + \beta_{1,i}x_1 + \cdots + \beta_{k,i}x_k \tag{4.5}$$

Based on equation (4.5), we can do a linear regression from the historical data set to predict $\begin{pmatrix} \mathrm{x}_{\mathrm{k+1}} \\ \vdots \\ \mathrm{x}_{\mathrm{k+l}} \end{pmatrix}$ from $\begin{pmatrix} \mathrm{x}_1 \\ \vdots \\ \mathrm{x}_{\mathrm{k}} \end{pmatrix}$.
The result of this linear regression method is shown as Fig. 4.2

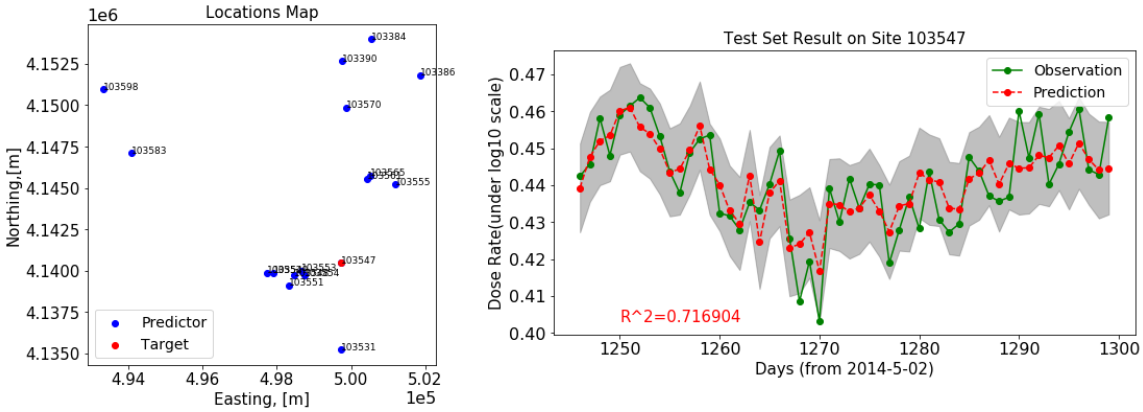

Fig. 4.2 (a) Locations of monitoring posts used for estimating the fluctuation, the dose rate at the "target" post was estimated; (b) The Prediction result using linear regression for the category II pixels.

To illustrate how linear regression works, we use site 103547 as the target to estimate, and other sites as predictors. In Fig. 4.2, plot (a) show the location map of each site, and (b) show the prediction vs. observation, the grey band is the 95% confidence interval.

(3). Kalman Filter

The Linear Regression method doesn't require the covariance function and nugget effect assumption, so it is easy to be implemented. However, it requires all the predictors to be available at each time step: if one predictor is missing, the prediction can't be made. Due to this reason, we turned to Kalman filter which can be more general to handle the cases that some predictors are missing.

At t_{th} time step, assume we have observation vector Z_t, whose dimension is k_t, and k_t can be different for different time step. We define a dose rate vector at time step t as

$X_t = \left(x_{1,t}, x_{2,t}, \cdots, x_{n,t}\right)^T$, where $x_{i,t}$ is the dose rate for pixel i at time step t and n is the number of pixels in category I, and define the $n \times n$ covariance matrix of X_t as P_t. Then, $\hat{X}_{t+1|t}$ and $\hat{P}_{t+1|t}$ represent the prediction of dose rate and covariance matrix for the time step $(t+1)$ conditioned on the dose rate at the time previous step t.

Based on the decay model of Eq.(4.1), the vector form can be written as

$$X_{t+1} = X_t + \boldsymbol{u} + w_t \tag{4.6}$$

and

$$w_t = X_{t+1} - X_t + \boldsymbol{u} \tag{4.7}$$

And the $n \times n$ covariance matrix Q_t between w_s and w_t is

$$Q_t = cov(w_s, w_t) = \begin{cases} cov(X_{t+1} - X_t)\ if\ t = s \\ 0, \quad otherwise \end{cases} \tag{4.8}$$

For every time step, we will have k_t-elements measurement/observation vector Z_t,

$$Z_t = H_t X_t + m_t \tag{4.9}$$

where

$$cov(m_t, m_s) = \begin{cases} R_t\ if\ t = s \\ 0, \quad otherwise \end{cases} \tag{4.10}$$

and $\boldsymbol{u}$ is a vector of *n* elements, representing the slopes of the linear trend of all the pixels in Category II, it can be fitted from the historical data; the $n \times n$ covariance matrix Σ here is the covariance of the noise ε among different pixels with dimension $n \times n$, and R_t is the covariance matrix of additive measurement noise with dimension of $k_t \times k_t$. H_t is the $k_t \times n$ observation matrix, is determined by the data model, i.e., how each type of measurement (airborne, car, monitoring post) was represented by the dose rates of all pixels and integrated in our method, m_t is the additive measurement noise (k_t elements vector) whose covariance is R_t.

Equation. (4.6) to (4.10) match a standard Kalman Filter (Kalman, 1960) model which is made up of two steps:

Predict step:

$$\hat{X}_{t+1|t} = \hat{X}_{t|t} + \boldsymbol{u} \tag{4.11}$$

$$\hat{\mathrm{P}}_{\mathrm{t+1|t}} = \hat{\mathrm{P}}_{\mathrm{t|t}} + \mathrm{Q_t} \tag{4.12}$$

Update step:

$$\hat{\mathrm{X}}_{\mathrm{t+1|t+1}} = \hat{\mathrm{X}}_{\mathrm{t+1|t}} + \mathrm{K_{t+1}}\left[\mathrm{Z_{t+1}} - \mathrm{H_{t+1}}\,\hat{\mathrm{X}}_{\mathrm{t+1|t}}\right] \tag{4.13}$$

$$\widehat{\mathrm{P}}_{t+1|t+1} = (\mathrm{I} - \mathrm{K}_{t+1}\mathrm{H}_{t+1})\widehat{\mathrm{P}}_{t+1|t}(\mathrm{I} - \mathrm{K}_{t+1}\mathrm{H}_{t+1})^{\mathrm{T}} + \mathrm{K}_{t+1}\mathrm{R}_{t+1}\mathrm{K}_{t+1}^{\mathrm{T}} \quad (4.14)$$

where,

$$\mathrm{K}_{t+1} = \widehat{\mathrm{P}}_{t+1|t}\mathrm{H}_{t+1}^{\mathrm{T}}\left[\mathrm{H}_{t}\widehat{\mathrm{P}}_{t+1|t}\mathrm{H}_{t+1}^{\mathrm{T}} + \mathrm{R}_{t+1}\right]^{-1} \quad (4.15)$$

is a $n \times k_{t+1}$ matrix called Kalman Gain. The subscripts of $\widehat{\mathrm{P}}_{t+1|t}$, $\widehat{\mathrm{X}}_{t+1|t}$ mean the prediction of P, X at time step $t+1$ conditioned on time step t.

The result of Kalman filter (Equation (4.11) to (4.15)) can be seen at Fig. 4.3

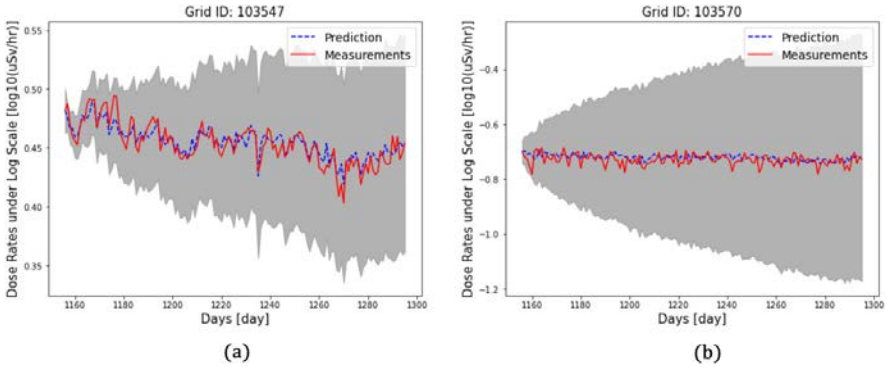

Fig. 4.3. Comparison of Category II pixels, (a) for pixel 103547; (b) for pixel 103570; grey band is 95% confidence interval.

The time-series prediction for the Category II pixels is shown in Fig. 4.3. These two locations in Fig. 4.3 are from the testing set, which is randomly selected 40% of the data set. The predicted and measured values are compared at the two pixels (pixel 103547 starting from the higher dose rate and pixel 103570 starting from low dose rates). There are two characteristics with respect to the 95% confidence interval in Fig. 4.3. First, the width of the confidence interval grows larger with time t, this is inconsistent with our understanding of Kalman filter that gives a non-increasing confidence interval, such as Figure.4 in Schmidt et al.(2018). However, after carefully investigation, we found out the increasing width of confidence interval in Fig. 4.3 was due to the weak constraint of Equation (4.9). In Equation (4.9), we set the observation matrix as $H_t = \left(I_{k_t \times k_t}, 0_{k_t \times (n-k_t)}\right)$, the constraint applied to the $(k_t + 1)_{th}$ to the n_{th} pixels were fulfilled by the $n \times n$ covariance matrix Σ. This constraint reduced the uncertainty induced by Equation (4.6) to an order of $\sqrt{t}$. As a result, the width of the confidence interval actually increases following shape of a square root function. Furthermore, although the confidence interval seems "weird", the prediction of X_t seems reasonable. Actually, the prediction of X_t when $R_t = 0$ is same as the prediction given by GPM. All these math proofs can be found in Appendix II. Secondly, the 95% confidence intervals are different in order between Fig. 4.3a and Fig. 4.3b. In Fig. 4.3a, the 95% confidence interval at $t = 1300$ day has a width of around 0.2, which includes most of the measured values. The Pearson correlation coefficients R is 0.91. In Fig. 4.3b, the 95% confidence interval at

$t = 1300$ day is 0.8 in width, and Pearson correlation coefficients R is 0.25 while most measurements still fall into the confidence interval. We can find that Fig. 4.3a that started from the higher dose rates has smaller uncertainty while Fig. 4.3b that started from low dose rate has larger uncertainty. There are many reasons that may cause larger uncertainty, the closest distance to the conditioning pixels, the malfunction of the measurement device, ration of signal to background etc. After looking into the locations of these pixels, we believe that the uncertainty is mostly determined by the distance to the closest conditioning pixel.

(4). Improved Filter

Even though the prediction of X_t in Kalman filter is reasonable, the predicted uncertainty is of little use as it is growing larger with time t. We investigated into this phenomenon and found out that the assumption of Kalman filter (Equation (4.8)) was not strictly followed in our problem. In our problem,

$$Q_t = cov(w_s, w_t) = \begin{cases} 2\Sigma, & if\ \mathrm{t} = \mathrm{s} \\ -\Sigma, & if\ |t-s| = 1 \\ 0, & otherwise \end{cases} \tag{4.16}$$

where $\Sigma = cov(\varepsilon_s, \varepsilon_t)$. Based on this new Q_t, our improved filter will be (math details of this improved filter can be found in Appendix III)

Predict step:

$$\hat{X}_{t+1|t} = \hat{X}_{t|t} + \boldsymbol{u} \tag{4.17}$$

$$\widehat{\mathrm{P}}_{\mathrm{t+1|t}} = \widehat{\mathrm{P}}_{\mathrm{t|t}} + \Sigma \cdot \mathrm{H}_{\mathrm{t}}^{\mathrm{T}} \cdot (\mathrm{H}_{\mathrm{t}} \cdot \Sigma \cdot \mathrm{H}_{\mathrm{t}}^{\mathrm{T}} + \mathrm{R}_{\mathrm{t}})^{-1} \cdot \mathrm{H}_{\mathrm{t}} \cdot \Sigma \tag{4.18}$$

Update step:

$$\widehat{\mathrm{X}}_{\mathrm{t+1|t+1}} = \widehat{\mathrm{X}}_{\mathrm{t+1|t}} + \mathrm{K}_{\mathrm{t+1}}\left[\mathrm{Z}_{\mathrm{t+1}} - \mathrm{H}_{\mathrm{t+1}}\widehat{\mathrm{X}}_{\mathrm{t+1|t}}\right] \tag{4.19}$$

$$\widehat{\mathrm{P}}_{\mathrm{t+1|t+1}} = (\mathrm{I} - \mathrm{K}_{\mathrm{t+1}}\mathrm{H}_{\mathrm{t+1}})\widehat{\mathrm{P}}_{\mathrm{t+1|t}}(\mathrm{I} - \mathrm{K}_{\mathrm{t+1}}\mathrm{H}_{\mathrm{t+1}})^{\mathrm{T}} + \mathrm{K}_{\mathrm{t+1}}\mathrm{R}_{\mathrm{t+1}}\mathrm{K}_{\mathrm{t+1}}^{\mathrm{T}} \tag{4.20}$$

where,

$$\mathrm{K}_{\mathrm{t+1}} = \widehat{\mathrm{P}}_{\mathrm{t+1|t}}\mathrm{H}_{\mathrm{t+1}}^{\mathrm{T}}\left[\mathrm{H}_{\mathrm{t}}\widehat{\mathrm{P}}_{\mathrm{t+1|t}}\mathrm{H}_{\mathrm{t+1}}^{\mathrm{T}} + \mathrm{R}_{\mathrm{t+1}}\right]^{-1} \tag{4.21}$$

Equation (4.17) to (4.21) constitute the new filter we developed in this study. Compared with Kalman filter (Equation (4.11) to (4.15)), only $\widehat{\mathrm{P}}_{\mathrm{t+1|t}}$ in (4.18) and (4.12) are different, other equations are the same.

The result of our improved filter is illustrated in Fig. 4.4,

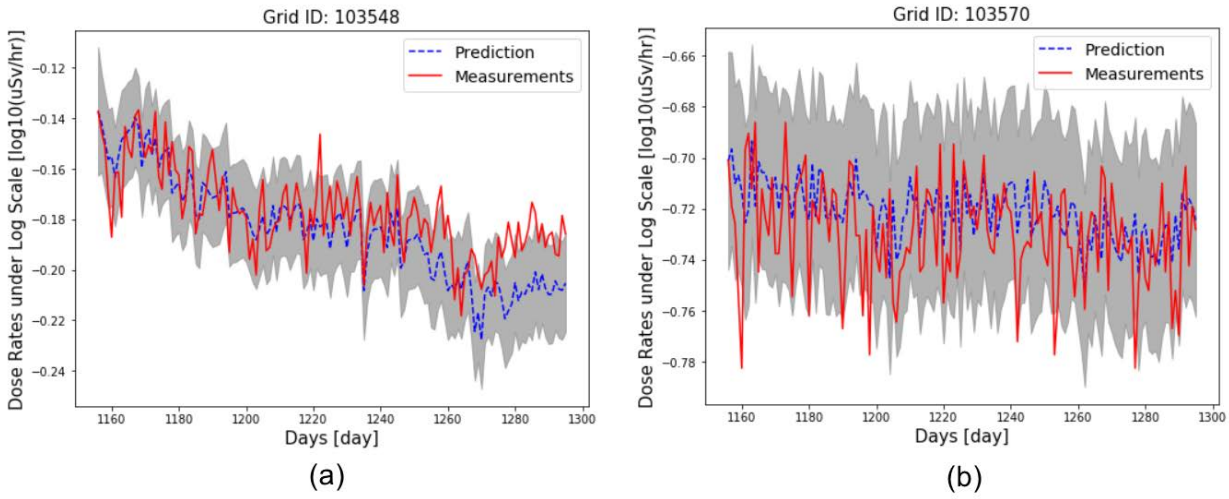

Fig. 4.4 Comparison of Category II pixels, (a) for pixel 103548; (b) for pixel 103570; grey band is 95% confidence interval.

Compared with Fig. 4.3, we can see that Fig. 4.4 presents more reasonable confidence interval while their prediction of X_t are same (mathematical explanation of this phenomenon can be found in Appendix II).

As a summary, four possible solutions were discussed in the above, and a comparison between the Kalman filter and our own improved filter are made in Appendix II. In the end, we chose our own improved filter as the final solution for Step one.

Step two: Predict dose rate time-series for pixels in Category III for the same time step as Step one.

Based on Equation (4.1), to predict the dose rates times for pixels in Category III which have no historical data, three components are need: (1) initial values (dose rates at $t = 0$, which is **b** in Equation (4.1)), (2) the slope of the linear trend ($\boldsymbol{u}$ in Equation (4.1)), (3) the fluctuation part ($\boldsymbol{\varepsilon_i}$ in Equation (4.1)). So, this prediction step can make up of three sub-steps to compute the three components:

(a). Initial values of the time series can be selected from the Integrated Dose Rates map of 2014, which was already developed by Wainwright et al. (2017, 2018).

(b). Slope of the linear trend can be predicated using Random Forest with initial values, land-use type, x, y coordinate position as predictor variables.

(c). At the same time step, fluctuations around the trends can be predicted by dose rates conditioned on the predicted values at the Category I&II pixels, using Gaussian Process Model(GPM). GPM assumes that the dose rate values at unmeasured locations follows a multivariate normal distribution, and the unobserved variables can be predicted by conditioning distribution on the observed variables.

Specifically, assume n-elements vector F_1denotes the fluctuation vector for the pixels from Category I and II, the m-elements vector F_2 denotes the fluctuation vector for the pixels from Category III at the same time step, then the best prediction of F_2 conditioning on F_1 is

$$\mathrm{F}_{2|1} = \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathrm{F}_1 - \mu_1) \tag{4.22}$$

$$\Sigma_{2|1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \tag{4.23}$$

In the above, F_1 is a n-elements vector, $F_{2|1}$is a m-elements vector, μ_1, μ_2 are prior mean vectors which are 0 here, and Σ_{ij} are covariance matrix between (F_i, F_j). The dimension of Σ_{22} is $m \times m$, the dimension of Σ_{11} is $n \times n$, and the dimension of $\Sigma_{2|1}$ is $m \times m$.

Step three: repeat Step one and Step two for all the time points.

Through the repeating of Step one and two for all the time step, the time series for all the pixels (Category II and Category III) can be reconstructed. Alternatively, we can repeat step one for all the time steps and then repeat step two for all the time steps. Both ways will give the same result.

4.3.3 Prediction of Trend Slope

In order to estimate the dose rate time-series of pixels in category III, both the fluctuation part and the slope of the linear trend need to be estimated. The fluctuation part could be estimated through GPM at each time step (will be discussed in section 4.3.5). In this section, we investigated how to predict the slope of the linear trend of dose rate time-series. For Category I and II pixels, slope of time-series trend can be obtained from linear fitting of the measurements. We found that the slope of time-series trend (i.e., slope) is correlated with the initial dose rates (i.e., intercept) as shown in Fig. 4.5a. A random forest regression was applied to predict the trend (i.e., slope) as a function of the intercept, land-use type, spatial coordinates for category III pixels (with no measurements). Since the monitoring post datasets are limited in space, we used the car survey datasets to characterize the environmental decay trend.

The regression result is presented in Fig. 4.5b, with the predictive performance of R^2 equal to 0.934. The random forest method also provides a parameter importance ranking in the regression model. We found that the initial dose rate is the most influential factor, determining the decay rate (i.e., the slope). The minor contribution is detected from the land-use type, and spatial coordinates.

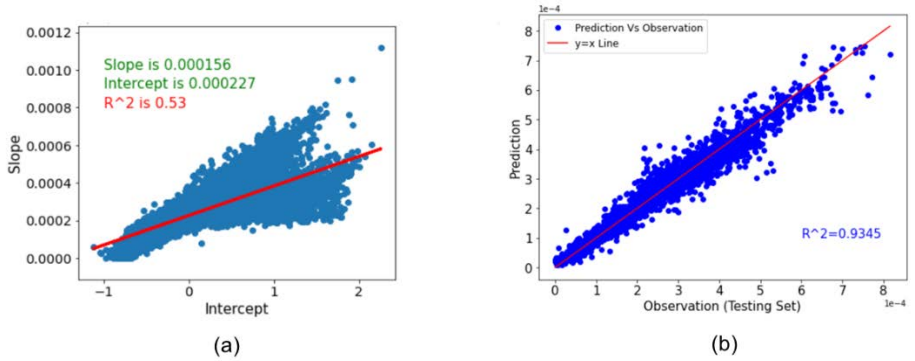

Fig. 4.5. Slope prediction result: (a) Slope vs Intercept (initial dose rates); (b) Slope prediction using Random Forest vs true value on testing data set, in the Random Forest process, indicators are Intercept, land-use type, x,y coordinate position. Data are from TEPCO carborne dataset since 2014-01-01.

In addition, we further investigated the two branches that appear in the correlations between the intercept and slope in Fig. 4.5a. We manually separate the data points into the two groups: the red cluster (lower part) and blue cluster (upper part) as shown in Fig. 4.6a.

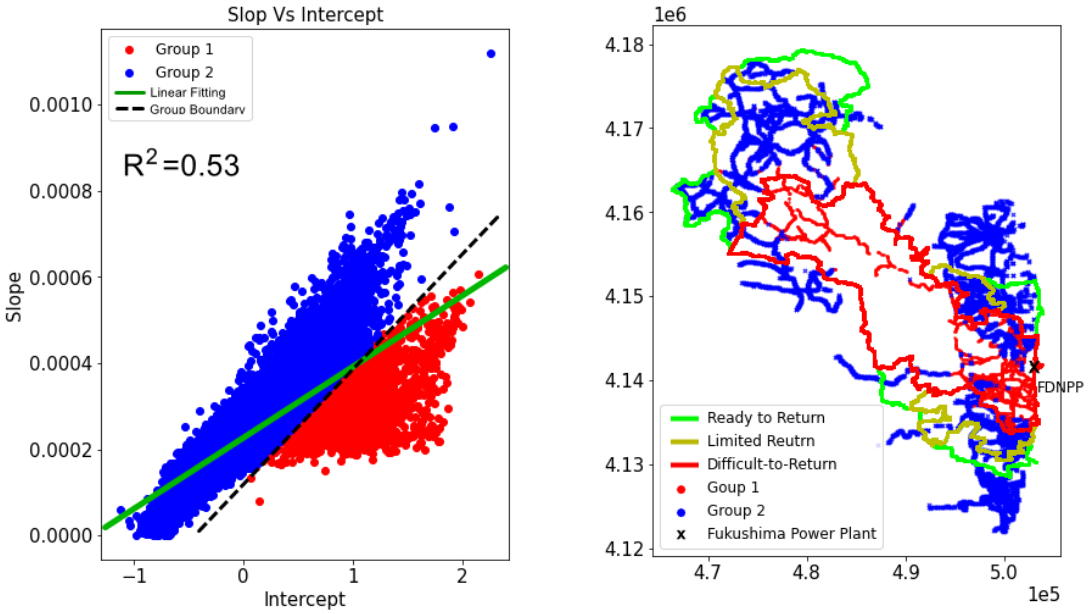

Fig. 4.6 Slope and intercept correlation and relative distribution: (a) the slope vs intercept plot, (b) the spatial distribution of two clusters in (a). In (b), the region closed by lime curves is the evacuation zone that is ready to lift, the region closed by yellow curves is evacuation zone that still has restriction to return, the region closed by the red curve is the zone of high risk that is not allowed to return.

We found that these two branches correspond the different zones in the region (see Fig. 4.6b); the Difficult-to-return (or high-risk) zone and Ready/Limited-to-return zone. The Difficult-to-return zones have low environmental decay rates (i.e., smaller slopes), even though the high initial dose rates are high. It is possibly because the human activities

within the Ready-to-return zones mobilized the soil particles and radiocesium, accelerating the decay (Saito et al., 2018). In addition, the Difficult-to-return zone is dominated by forests, where the environmental decay is reported to be slower (Kinase et al., 2014; Saito et al. 2018). As a result, after we included the xy-location, landuse type as indicator variables in the random forest model, the two branches in Fig. 4.5a merged and we got a good agreement in Fig. 4.5b.

In order to study the mechanism behind the slope ~ intercept relationship, we went through many possible explanations. Firstly, we investigated the influence of the natural background. Specifically, when the dose rates level is much higher than the natural background, the influence of the natural background can be ignored, then the slope of the linear trend equals the ecological half-lives; when the dose rates is not high enough, the existence of the background will pull the slope toward zero. As this result, this possible explanation predicts that high dose rates have higher slope than the lower dose rates due to the influence of natural background. In order to check this, once again, we divided the data into three groups: (1) data points that are off the linear trend of Slope ~ Intercept (indicated by red points), (2) dose rates in the last day (2580^{th} day) that are higher than 10 times of the natural background and not in group 1 (indicated by yellow points), (3) dose rates in the last day that are lower than 10 times of the natural background and not in group 1 (indicated by green points). The three groups are plotted in Fig. 4.7a,b:

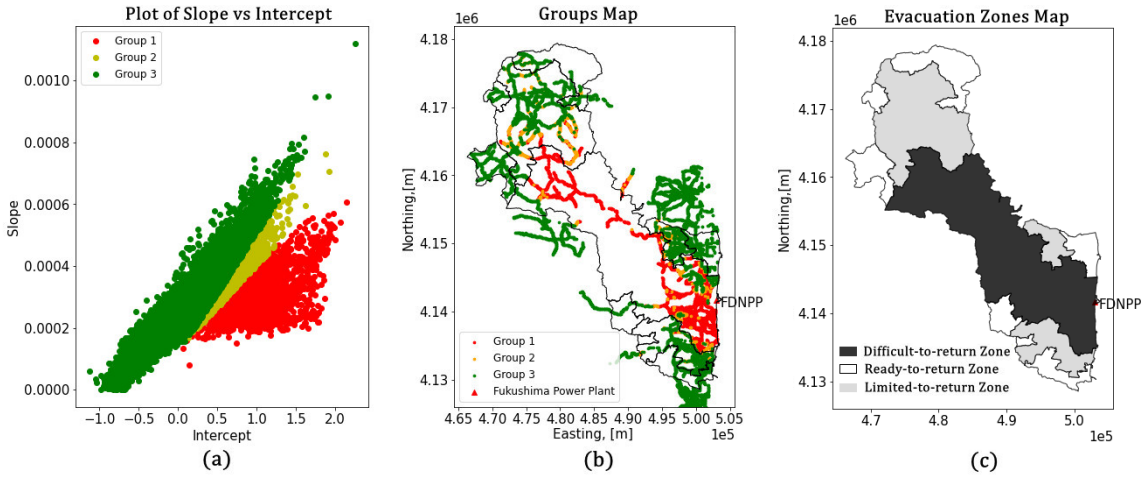

Fig. 4.7 (a) Plot of slope vs. intercept; (b) Location map for the three groups; (c) Evacuation Zones map.

In Fig. 4.7a, there are more green points than yellow points, which means most of the place in the Ready/Limited-to-return Zones has lower dose rates than 10 times the natural background. For these yellow points whose dose rates are far more than 10 times of the nature background, their slopes are still varying greatly, which means that the influence of natural background can't tell us the whole story, there probably be other factors that influence the slope.

As for the second possible explanation -- soil influence, we plotted the soil information of the Ready/Limited-to-return Zones and Difficult-to-return Zone in Fig. 4.8. As shown in Fig. 4.8, the two type of Zones has similar composition of each soil type. For

comparison, we also plotted the slope information of the two different Zones, as shown in Fig. 4.9. Obviously, the slope of Kuroboku soil (which is the most comman soil type as shown in Fig. 4.8) in the Difficult-to-return Zone is slight smaller than the counterpart in the Ready/Limited-to-return Zones as indicated in Fig. 4.9. Furthermore, we drew the box plot of the intercept (which is equivalent to the initial dose rates under logarithm scale) for each soil type, as indicated by Fig. 4.10. In Fig. 4.10, we could find that the initial dose rates are higher in the Difficult-to-return Zone than the Ready/Limited-to-return Zones for all the soil types.

Furthermore, we did the one-way ANOVA to check the means of slope/intercept in the two groups (Difficult-to-return Zone Vs Ready/Limited-to-return Zones) equal or not. A one-way ANOVA is a test used to determine whether or not there is a statistically significant difference between the means of two or more independent groups. If the overall p-value from the ANOVA test is less than some significance level (such as $\alpha = 0.05$), then we have sufficient evidence to say that at least one of the means of the groups is different from the others. The result of slope and intercept test is as Table 4.1:

Table 4.1 One-way ANOVA test of slope/intercept for two groups

Variable to test	Statistic of ANOVA	p-value
Slope	84.43	7.0×10^{-20}
Intercept	1222.74	3.47×10^{-227}

Therefore, differences in time-series slopes of dose rates under logarithm scale probably caused by difference in soil type. Specifically, immature soil is a type of soil that lacks a well-developed profile, usually because it has not had enough time for one to develop by normal soil-forming processes, it is characterized by slight weathering of the mineral material and generally small amounts of organic matter in the profile (Duchaufour P., 1982). Kuroboku soils (black soils rich in humus content) are found on terraces, hills, and gentle slopes throughout Japan, soil organic matter also strongly affects the liquid limit and plastic limit of Andisols (Aragaki et al., 1987). It is reasonable to infer that the high level dose rates radioactive nuclides migrate more easily in the immature soil than the Kuroboku soil.

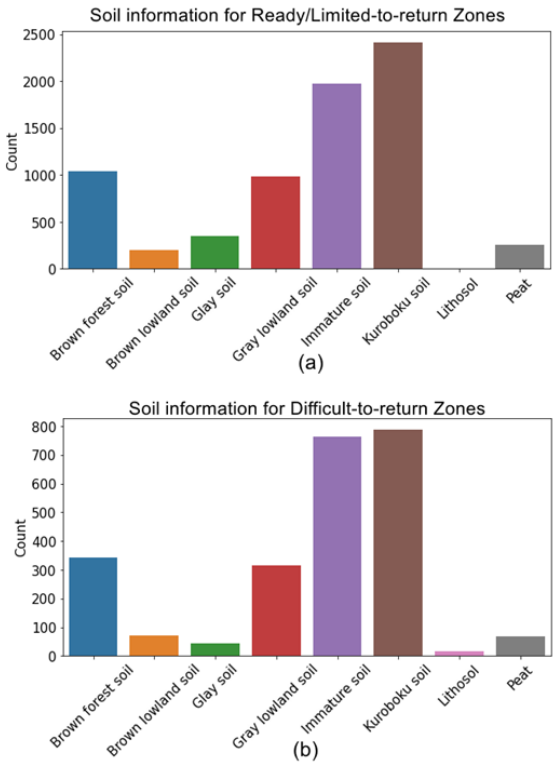

Fig. 4.8 Soil type information for (a) Ready/Limited-to-return Zones, (b) Difficult-to-return Zone.

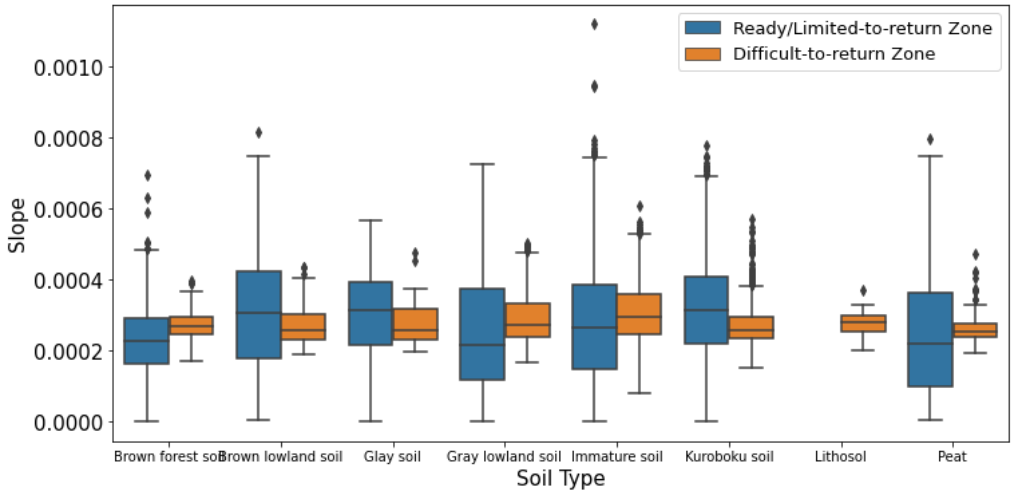

Fig. 4.9 Box plot of slope for Ready/Limited-to-return Zones vs Difficult-to-return Zone.

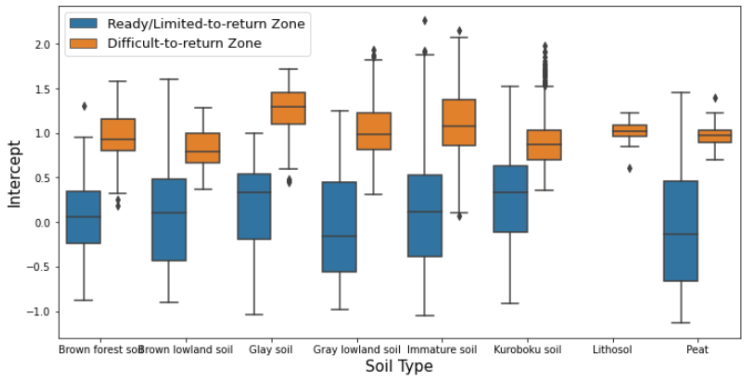

Fig. 4.10 Box plot of intercept (or initial dose rates) for vs soil type for Ready/limited-to-return zone and difficult-to-return zone.

4.3.4 Dose Rate Time-Series Prediction

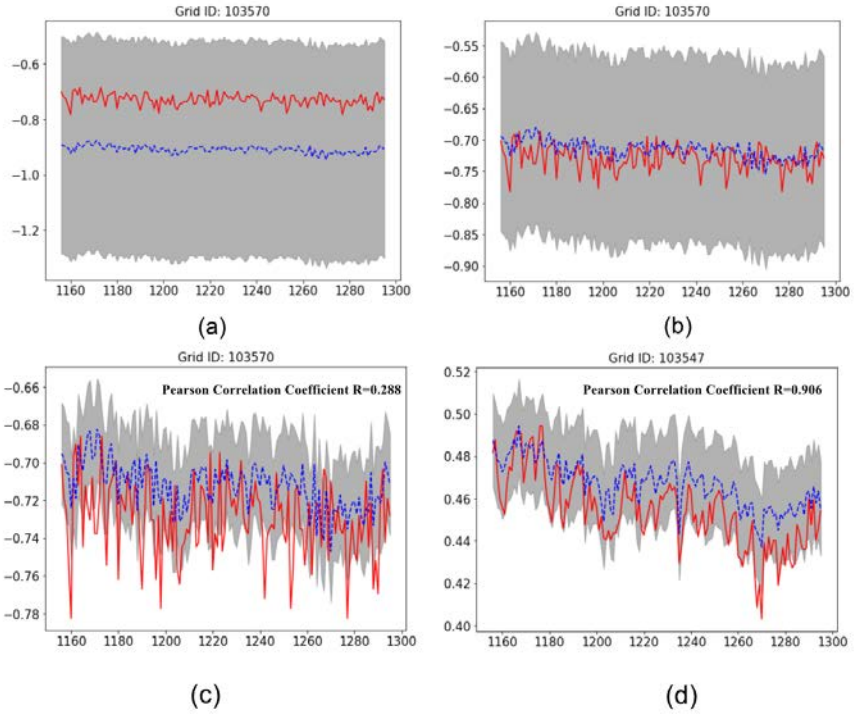

Fig. 4.11 Prediction vs True value for Category II: (a) for pixel 103570 (representing the low dose rates pixels) with initial values borrowed from the integrated map; (b) for pixel 103570 with adjusted initial values from true value, with variance estimated by GPM Python package; (c) for pixel 103570 with adjusted the initial values from true value, with variance estimated by the mean of the variances of the time-series of the 17 locations;(d) for pixel 103547 (representing the high dose rates pixels) with adjusted the initial values from true value. In the plots, Red line: Measurements; blue dashed line: Prediction; grey band: 95% confidence interval.

The dose-rate time-series at the Category III pixels are predicted by combining the environmental decay rates (i.e., slope) predicted by the random forest and the spatiotemporal fluctuations predicted by GPM, results are shown as Fig. 4.11. In Fig. 4.11a, there is obvious deviation between the prediction and the measurement, that discrepancy is due to discrepancy between the initial value extracted from the integrated map and the true initial value. After initial value been corrected by the true value, a new comparison is presented in Fig. 4.11b. Meanwhile, the confidence interval is exaggerated in Fig. 4.11a and Fig. 4.11b, that is due to the over estimation of variance by the GPM package in Python (Gaussian Process. scikit-learn 0.24.2., 2011). In this study, we only have 17 locations, which led to large uncertainty of estimated variance. Because the GMP package doesn't provide too much flexibility for users, we wrote our own code of Kriging method that using the mean of the variances of the time-series of the 17 locations instead of the overestimated variance by the GPM package, and got the result as shown in Fig. 4.11c and Fig. 4.11d. Comparing Fig. 4.11c and Fig. 4.11d, we can see that Fig. 4.11d presents a higher Pearson correlation coefficient for higher dose rates, which means our algorithm can predict better for higher dose rates pixels, and this conclusion could also be found in Fig. 4.5.

Fig. 4.12 provides the predicted spatiotemporal evolution of the dose rates, which is the decrease from the initial dose rate at 10 days, 30 days, 80 days and 120 days after the first day of this time-series (which is 1156th day after the accident). The color of each pixel represents the decrease of dose rates in the log scale, which is heterogeneous across the space. The blue/violet region is actually expanding from Fig. 4.12a to Fig. 4.12d, and this trend is more obvious north west direction which is supposed to be the high dose rates zone, it agrees with the conclusion we draw from Fig. 4.11 that higher dose rates will decrease at higher rates.

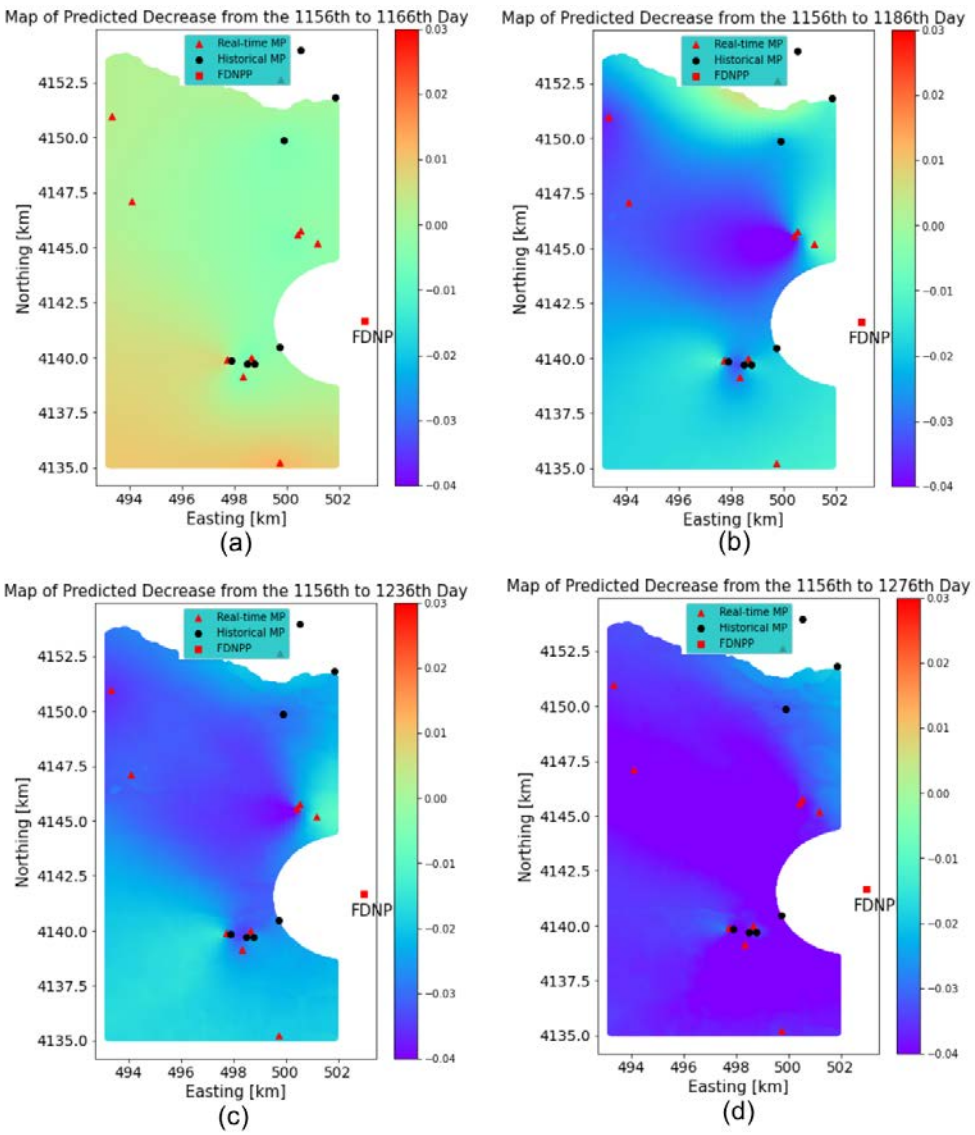

Fig. 4.12 Predicted Dose Rates Decrease for the Study Area: (a) to (d) the dose rate decrease from 1156th day to the 1166th,1186th,1236th and 1276th days after the accident. In the color bar, negative value means decrease.

4.3.5 Further Discussion

The result in Section 4.3.4 is promising as it provided a way to estimate the radio dose rates in the future for any location. It could also provide support for policy makers to evaluate the dose exposure at any location within the evacuation zone and decide when to lift the evacuation order. Furthermore, the predicted map can merge with the integrated map if they are available at the same time point, which will make the dose rates distribution information continuous and more accurate. The algorithm to merge prediction map with integrated map is provided as

$$\begin{cases} u' = \dfrac{u_1\sigma_2^2 + u_2\sigma_1^2}{\sigma_2^2 + \sigma_1^2} \\ \sigma'^2 = \dfrac{\sigma_2^2\sigma_1^2}{\sigma_2^2 + \sigma_1^2} \end{cases} \tag{4.24}$$

where u_1, σ_1 are predicted value and stand deviation (or uncertainty) from this study, and u_2, σ_2 are value and stand deviation (or uncertainty) from the integrated map.

Meanwhile, Equation (4.1) can be extended beyond linear trend model. In this study $\boldsymbol{u}$ in Equation (4.6) and (4.11) is a constant during each iteration as Equation (4.1) is a linear trend model. If the trend in Equation (4.1) is not linear, then u in Equation (4.6) and (4.11) can be obtained by the difference the trend and is not a constant in general, this extends our model's versatility.

As the last point, the observation matrix H can be extended to incorporate multiscale measurement, which make our model can be extended to integrated different type of measurement. i.e.,

$$Z_i = \sum_{j=1}^{n} H_{i,j} X_j + m_i \tag{4.25}$$

For different type of measurement, $H_{i,j}$ is different and was constructed in the same way as in Equation (2.2).

4.4 Conclusions

In this chapter, we have verified the spatial correlation of the dose rates time-series and developed the methodology to predict spatial-temporal distribution of the dose rates using a combination of algorithms, based on the Gaussian process models and our own improved filter. We found that: (1) dose rate time-series for a single pixel can be modeled as a linear decay trend with random fluctuations around the trend, the linear trend is primarily dependent on the initial dose rate (in 2014), and the fluctuation of radiation dose rates are spatially correlated among adjacent monitoring posts, (2) the our own improved filter can be applied for the dose rates prediction/integration in the pixels with part or fully historical data set (pixels in Category I&II), (3) the environmental decay rate (i.e., the slope of log-linear trend) can be predicted using the random forest method as a function of the initial dose rate, land-use type, and the spatial coordinates as predictors, (4) the fluctuations of the time series from the linear trend at the Category III pixels can be predicted using the Gaussian Process Model. As a summary, our method can successfully estimate the air dose rates continuously over time at the location without monitoring posts. As it is able to estimate the dose rates for any location at any time, this methodology can be used as a general methodology either for improving the existing optimization method or dose rates exposure evaluation for the purpose of planning the

return of evacuated residents.

5 Conclusion

Radiation monitoring has been continuously performed for over ten years since the Fukushima Daiichi Nuclear Power Plant accident, which resulted in a large volume of datasets. Analyzing huge-volume data sets, and optimizing the existing monitoring network are critical for and scheduling the return of evacuated people. The research studies presented in this thesis solved three challenges that emerged from radiation monitoring after the FDNPP accident.

First, we extended the existing multiscale-dataset-integration approach to improve computational efficiency, including parallelization. We simplified the matrix expression of the result and improved the computing efficiency, making it hundreds of times faster and enabling it to be run on a PC. Furthermore, the new approach can easily be implemented simply by slightly revising the existing computation code. We not only saved computational resources in dealing with this problem, but also provided a possible solution in the future for similar problems.

Then, we developed a sensor network optimization method for optimizing monitoring locations of air-dose rates by reaching its local optima, making it feasible to use fewer sensors (approximately half the original number) in monitoring the region without losing much spatial information. Since long-term monitoring plays an important role in confirming the stability of the contaminant distribution, and in the reduction of radiation levels to the public, this optimization method makes a significant contribution to protecting public health and to supporting the local community. Furthermore, this method could be extended and generalized to other contamination problems that involve reducing the number of monitoring locations or optimally placing mobile measurements.

Finally, we further studied the spatial-temporal heterogeneity of environmental decay across the regions that had not been analyzed in previous studies. Our model can provide a dose-rate time-series estimation for a specific location at given time, something which was not available in previous studies. This is critical for exposure evaluations, essential for the public's safe return to the evacuation zone. We also identified the key factors controlling environmental decay rates and investigated a possible explanation for the dependence of decay rates on soil types and other factors. Such a perspective could be very helpful in understanding ecological decay mechanisms.

Overall, this dissertation provided efficient solutions to resolve the three challenges related to long-term radiation monitoring strategies after nuclear power plant accidents, and it could provide useful suggestions to decision-makers working on long-term monitoring programs after nuclear power plant accidents.

Supplemental Information

References

Andoh M, Mikami S, Tsuda S, Yoshida T, Matsuda N, Saito K. (2018b) Decreasing trend of ambient dose equivalent rates over a wide area in eastern Japan until 2016 evaluated by car-borne surveys using KURAMA Systems. J. Environ. Radioact. 2018; 192:385-398. https://doi.org/10.1016/j.jenvrad.2018.07.009.

Andoh M, Nakahara Y, Tsuda S, Yoshida T, Matsuda N, Takahashi F, Mikami S, Kinouchi N, Sato T, Tanigaki M, Takamiya K, Sato N, Okumura R, Uchihori Y, Saito K. Measurement of air dose rates in wide area around the Fukushima Daiichi nuclear power plant through a series of car-borne surveys. J. Environ. Radioact. 2015; 139:266-280. https://doi.org/10.1016/j.jenvrad.2014.05.014.

Andoh M, Yamamoto H, Kanno T, Saito K. (2018a). Measurement of ambient dose equivalent rates by walk survey around Fukushima Dai-ichi Nuclear Power Plant using KURAMA-II until 2016. J. Environ. Radioact. 190-191:111-121. https://doi.org/10.1016/j.jenvrad.2018.04.025.

Andoh, M., Mikami, S., Tsuda, S., Yoshida, T., Matsuda, N., & Saito, K. (2018b). Decreasing trend of ambient dose equivalent rates over a wide area in eastern Japan until 2016 evaluated by car-borne surveys using KURAMA systems. Journal of environmental radioactivity.

Andoh, M., Sasaki, M., & Saito, K. (2020). Evaluation of decreasing trend in air dose rate and ecological half-life within an 80 km range from Fukushima Dai-ichi Nuclear Power Plant, using car-borne survey data measured by KURAMA systems up to 2018. Journal of Nuclear Science and Technology, 57(12), 1319-1330.

Araki, S., Iwahashi, K., Shimadera, H., Yamamoto, K., & Kondo, A. (2015). Optimization of air monitoring networks using chemical transport model and search algorithm. Atmospheric Environment, 122, 22-30. https://doi.org/10.1016/j.atmosenv.2015.09.030

Diggle, P. J., Tawn, J. A., & Moyeed, R. A. (1998). Model‐based geostatistics. Journal of the Royal Statistical Society: Series C (Applied Statistics), 47(3), 299-350.

Eddy-Dilek, C.; Millings, M. R.; Looney, B. B.; and Denham, M. E. Innovative Strategy for Long Term Monitoring of Metal and Radionuclide Plumes 2014. Savannah River Site (SRS), Aiken, SC (United States); Retrieved from https://www.osti.gov/scitech/biblio/1122792(accessed 1/12/2020).

Evacuation Orders and Restricted Areas. http://fukushimaontheglobe.com/the-earthquake-and-the-nuclear-accident/evacuation-orders-and-restricted-areas#current(retrieved on Jan 23, 2021)

Evrard, O., Laceby, J. P., & Nakao, A. (2019). Effectiveness of landscape decontamination following the Fukushima nuclear accident: a review. Soil, 5(2), 333-350.

Fukushima Prefectural Government, http://www.pref.fukushima.lg.jp/site/portal-english/en03-08.html, retrieved on December 19th, 2017.

Gaussian Process. scikit-learn 0.24.2. (2011). https://scikit-learn.org/stable/modules/gaussian_process.html.

Heuvelink, G. B., Jiang, Z., De Bruin, S., & Twenhöfel, C. J. (2010). Optimization of mobile radioactivity monitoring networks. International Journal of Geographical Information Science, 24(3), 365-382. https://doi.org/10.1080/13658810802646687

Ho, Tin Kam (1995). Random Decision Forests (PDF). Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, 14–16 August 1995. pp. 278–282. Archived from the original (PDF) on 17 April 2016. Retrieved 26 Dec 2020

International Atomic Energy Agency. The Fukushima Daiichi Accident. ISBN:978-92-0-107015-9. 2015.

Kinase, S., Sato, S., Sakamoto, R., Yamamoto, H., & Saito, K. (2015). Changes in ambient dose equivalent rates around roads at Kawamata town after the Fukushima accident. Radiation Protection Dosimetry, 167(1-3), 340-343. https://doi.org/10.1093/rpd/ncv275.

Kinase, S., Takahashi, T., & Saito, K. (2017). Long-term predictions of ambient dose equivalent rates after the Fukushima Daiichi nuclear power plant accident. Journal of Nuclear Science and Technology, 1-10. https://doi.org/10.1080/00223131.2017.1365659

Kinase, S., Takahashi, T., Sato, S., Sakamoto, R., & Saito, K. (2014). Development of prediction models for radioactive caesium distribution within the 80-km radius of the Fukushima Daiichi nuclear power plant. Radiation Protection Dosimetry, 160(4), 318-321. https://doi.org/10.1093/rpd/ncu014.

Kunii, N., Fujimura, M. S., Komasa, Y., Kitamura, A., Sato, H., Takatsuji, T., ... & Kimura, S. (2018). The Knowledge and Awareness for Radiocesium Food Monitoring after the Fukushima Daiichi Nuclear Accident in Nihonmatsu City, Fukushima Prefecture. International journal of environmental research and public health, 15(10), 2289.

LAPACK 3.10.0. (n.d.). LAPACK. Retrieved August 1, 2021, from http://www.netlib.org/lapack/explore-html/d7/d3b/group__double_g_esolve_ga5ee879032a8365897c3ba91e3dc8d512.html

Malins, A., Kurikami, H., Nakama, S., Saito, T., Okumura, M., Machida, M., & Kitamura, A. (2016). Evaluation of ambient dose equivalent rates influenced by vertical

and horizontal distribution of radioactive cesium in soil in Fukushima Prefecture. Journal of environmental radioactivity, 151, 38-49.

Malins, A., Okumura, M., Machida, M., Takemiya, H., and Saito, K. (2015) Fields of View for Environmental Radioactivity. Proceedings of the 2015 International Symposium on Radiological Issues for Fukushima's Revitalized Future (available at arXiv:1509.09125).

Masoudi, P., Le Coz, M., Cazala, C., & Saito, K. (2019). Spatial properties of soil analyses and airborne measurements for reconnaissance of soil contamination by 137Cs after Fukushima nuclear accident in 2011. Journal of environmental radioactivity, 202, 74-84. https://doi.org/10.1016/j.jenvrad.2018.11.014

McLachlan, G., and D. Peel (2004), Finite Mixture Models, John Wiley, N. Y.

Melles, S. J., Heuvelink, G. B., Twenhöfel, C. J., & Stöhlker, U. (2008, June). Sampling optimization trade-offs for long-term monitoring of gamma dose rates. In International Conference on Computational Science and Its Applications (pp. 444-458). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-69839-5_33

Mikami S, Tanaka H, Matsuda H, Sato S, Hoshide Y, Okuda N, Suzuki T, Sakamoto R, Andoh M, Saito K. The deposition densities of radiocesium and the air dose rates in undisturbed fields around the Fukushima Dai-ichi nuclear power plant; their temporal changes for five years after the accident. J. Environ. Radioact. 2019; 139:320-343. https://doi.org/10.1016/j.jenvrad.2019.03.017

Mikami, S., Maeyama, T., Hoshide, Y., Sakamoto, R., Sato, S., Okuda, N., ... & Fujiwara, M. (2015). Spatial distributions of radionuclides deposited onto ground soil around the Fukushima Dai-ichi Nuclear Power Plant and their temporal change until December 2012. Journal of Environmental Radioactivity, 139, 320-343. https://doi.org/10.1016/j.jenvrad.2014.09.010.

Ministry of Environment. Environmental remediation. http://josen.env.go.jp/en/decontamination/ (retrieved on Jan 29, 2021)

Miura, S. (2016). The effects of radioactive contamination on the forestry industry and commercial mushroom-log production in Fukushima, Japan. In Agricultural implications of the Fukushima nuclear accident (pp. 145-160). Springer, Tokyo.

NISA (Nuclear Industry Safety Agency), 2011. http://www.meti.go.jp/press/2011/10/

20111020001/20111020001.pdf (accessed 30.01.14.) (in Japanese).

Nuclear Regulation Authority of Japan (NRA), 2011. Revision of the airborne monitoring result based on comparison of the distribution map of radioactivity (map of radiocesium concentration in soil) by MEXT. http://radioactivity.nsr.go.jp/ja/contents/5000/4931/view.html, Accessed date: 1 April 2018 (in Japanese).

Oroza, C. A., Zheng, Z., Glaser, S. D., Tuia, D., & Bales, R. C. (2016). Optimizing embedded sensor network design for catchment‐scale snow‐depth estimation using LiDAR and machine learning. Water Resources Research, 52(10), 8174-8189. https://doi.org/10.1002/2016WR018896.

Pedregosa, F., et al. (2011), Scikit-learn: Machine learning in Python. Journal of machine learning research, 12, 2825–2830.

Saito K, Mikami S, Andoh M, Matsuda N, Kinase S, Tsuda S, Yoshida T, Sato T, Seki A, Yamamoto H, Sanada Y, Wainwright-Murakami H, Takemiya H. (2019). Summary of temporal changes in air dose rates and radionuclide deposition densities in the 80 km zone over five years after the Fukushima Nuclear Power Plant accident. Journal of environmental radioactivity. https://doi.org/10.1016/j.jenvrad.2018.12.020.

Saito K, Yamamoto H, Mikami S, Andoh M, Matsuda N, Kinase S, Tsuda S, Yoshida T, Matsumoto S, Sato T, Seki A, Takemiya H. (2016) Radiological conditions in the environment around the Fukushima Daiichi nuclear power plant site. Global Environ. Res. 20, 15–22.

Saito, K., & Onda, Y. (2015). Outline of the national mapping projects implemented after the Fukushima accident. Journal of environmental radioactivity 139, 240-249. https://doi.org/10.1016/j.jenvrad.2014.10.009.

Saito, K., Mikami, S., Andoh, M., Matsuda, N., Kinase, S., Tsuda, S., ... & Yoshimura, K. (2019). Temporal Change in Radiological Environments on Land after the Fukushima Daiichi Nuclear Power Plant Accident. Journal of Radiation Protection and Research, 44(4), 128-148.

Saito, K., Yamamoto, H., Mikami, S., Andoh, M., Matsuda, N., Kinase, S., ... & Takemiya, H. (2016). Radiological conditions in the environment around the Fukushima Daiichi Nuclear Power Plant site. Global Environ. Res, 20, 15-22.

Sanada Y, Sugira T, Nishizawa Y, Kondo A, Torii T. (2014). The aerial radiation monitoring in Japan after the Fukushima Daiichi nuclear power plant accident. J. Prog. In Nucl. Sci. Technol. 4:76-80. https://doi.org/10.15669/pnst.4.76.

Sanada Y, Urabe Y, Sasaki M, Ochi K, Torii T. (2018). Evaluation of ecological half-life of dose rate based on airborne radiation monitoring following the Fukushima Daiichi nuclear plant accident. J. Environ. Radioact. 192:417–425. https://doi.org/10.1016/j.jenvrad.2018.09.014.

Sanada, Y., & Torii, T. (2015). Aerial radiation monitoring around the Fukushima Dai-ichi nuclear power plant using an unmanned helicopter. Journal of environmental radioactivity, 139, 294-299.

Sanada, Y., Sugita, T., Nishizawa, Y., Kondo, A., & Torii, T. (2014). The aerial radiation monitoring in Japan after the Fukushima Daiichi nuclear power plant accident. Prog. Nucl. Sci. Technol, 4(7).

Sanada, Y., Urabe, Y., Sasaki, M., Ochi, K., & Torii, T. (2019). Evaluation of ecological half-life of dose rate based on airborne radiation monitoring following the Fukushima Dai-ichi nuclear power plant accident. Journal of environmental radioactivity, 210, 105816.

Sawano, T., Ozaki, A., Hori, A., & Tsubokura, M. (2019). Combating 'fake news' and social stigma after the Fukushima Daiichi Nuclear Power Plant incident—the importance of accurate longitudinal clinical data. QJM: An International Journal of Medicine. https://doi.org/10.1093/qjmed/hcz049.

Schmidt, F., Wainwright, H. M., Faybishenko, B., Denham, M., & Eddy-Dilek, C. (2018). In situ monitoring of groundwater contamination using the Kalman filter. Environmental Science & Technology, 52(13), 7418-7425.

Schmidt, F., Wainwright, H. M., Faybishenko, B., Denham, M., & Eddy-Dilek, C. (2018). In situ monitoring of groundwater contamination using the Kalman filter. Environmental science & technology, 52(13), 7418-7425.

Sun, C., Li, V. O., Lam, J. C., & Leslie, I. (2019). Optimal Citizen-Centric Sensor Placement for Air Quality Monitoring: A Case Study of City of Cambridge, the United Kingdom. IEEE Access, 7, 47390-47400. https://doi.org/10.1109/ACCESS.2019.2909111

Sun, D., Wainwright, H. M., Oroza, C. A., Seki, A., Mikami, S., Takemiya, H., & Saito, K. (2020). Optimizing long-term monitoring of radiation air-dose rates after the Fukushima Daiichi Nuclear Power Plant. Journal of environmental radioactivity, 220, 106281.

Takahashi et al., "JAXA High Resolution Land-Use and Land-Cover Map of Japan," 2013 IEEE International Geoscience and Remote Sensing Symposium - IGARSS, Melbourne, VIC, 2013, pp. 2384-2387. https://doi.org/10.1109/IGARSS.2013.6723299.

Takahashi, Y., Fan, Q., Suga, H., Tanaka, K., Sakaguchi, A., Takeichi, Y., ... & Kanivets, V. V. (2017). Comparison of solid-water partitions of radiocesium in river waters in fukushima and chernobyl areas. Scientific reports, 7(1), 1-11.

Tanigaki, M., Okumura, R., Takamiya, K., Sato, N., Yoshino, H., Yoshinaga, H., ... & Yamana, H. (2015). Development of KURAMA-II and its operation in Fukushima. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 781, 57-64.

Tanigaki, W. I., Okuda, T., & YAMADA, T. (2016). Potential of crowdsourcing approach on monitoring radioactivity in Fukushima prefecture. Global Environmental Research, 20, 111-8.

Tealab, A. (2018). Time series forecasting using artificial neural networks methodologies: A systematic review. Future Computing and Informatics Journal, 3(2), 334-340.

The Development of and Lessons from the Fukushima Daiichi Nuclear Accident, TEPCO, 2013. https://www.tepco.co.jp/en/decommision/accident/images/outline01.pdf (retrieved on Jan 23, 2021)

Vetter, K., Barnowski, R., Cates, J. W., Haefner, A., Joshi, T. H., Pavlovsky, R., & Quiter, B. J. (2019). Advances in nuclear radiation sensing: Enabling 3-D gamma-ray vision. Sensors, 19(11), 2541.

Wainwright et al. (2017), White paper: Field and Virtual Testbeds for Cost-effective Sustainable Remediation, Enhanced Attenuation and Long-Term Monitoring, online access: https://aodocs.altirnao.com/drive/redirect/lbl.gov/QaWDxMgs30LaSFFBRy

Wainwright, H. M., Seki, A., Chen, J., & Saito, K. (2017). A multiscale Bayesian data integration approach for mapping air dose rates around the Fukushima Daiichi Nuclear Power Plant. Journal of Environmental Radioactivity, 167, 62-69. https://doi.org/10.1016/j.jenvrad.2016.11.033.

Wainwright, H. M., Seki, A., Mikami, S., & Saito, K. (2018). Characterizing regional-scale temporal evolution of air dose rates after the Fukushima Daiichi Nuclear Power Plant accident. Journal of Environmental Radioactivity, 189, 213-220. https://doi.org/10.1016/j.jenvrad.2018.09.006.

Wainwright, H. M., Seki, A., Mikami, S., & Saito, K. (2019). Characterizing regional-scale temporal evolution of air dose rates after the Fukushima Daiichi Nuclear Power Plant accident. Journal of environmental radioactivity, 210, 105808.

Yasutaka, T., Iwasaki, Y., Hashimoto, S., Naito, W., Ono, K., Kishimoto, A., ... & Nakanishi, J. (2013). A GIS-based evaluation of the effect of decontamination on effective doses due to long-term external exposures in Fukushima. Chemosphere, 93(6), 1222-1229. https://doi.org/10.1016/j.chemosphere.2013.06.083.

Zhuang, X., & Liu, R. (2011, June). The optimization of regional air quality monitoring network based on spatial analysis. In 2011 19th International Conference on Geoinformatics (pp. 1-6). IEEE. https://doi.org/10.1109/GeoInformatics.2011.5980772

Appendix I

This appendix is to show that the result got by Bayesian hierarchical approach (section 2.3) equals the result of our new approach (section 2.4). which is to prove

$$\begin{cases} Q^{-1}g = \mu_c + (\Sigma_c A^T, \quad \Sigma_c C^T)\begin{pmatrix} A\Sigma_c A^T + D_A & A\Sigma_c C^T \\ C\Sigma_c A^T & C\Sigma_c C^T + D_C \end{pmatrix}^{-1}\begin{pmatrix} z_A - A\mu_c \\ z_C - C\mu_c \end{pmatrix} \\ Q^{-1} = \Sigma_c - (\Sigma_c A^T, \quad \Sigma_c C^T)\begin{pmatrix} A\Sigma_c A^T + D_A & A\Sigma_c C^T \\ C\Sigma_c A^T & C\Sigma_c C^T + D_C \end{pmatrix}^{-1}\begin{pmatrix} A\Sigma_c \\ C\Sigma_c \end{pmatrix} \end{cases} \quad \text{(A.1)}$$

where

$$\begin{cases} Q = \Sigma_c^{-1} + A^T D_A^{-1} A + C^T D_C^{-1} C \\ g = \Sigma_c^{-1}\mu_C + A^T D_A^{-1} z_A + C^T D_C^{-1} z_C \end{cases}$$

As the first step, we first show that

$$Q^{-1} = \Sigma_c - (\Sigma_c A^T, \quad \Sigma_c C^T)\begin{pmatrix} A\Sigma_c A^T + D_A & A\Sigma_c C^T \\ C\Sigma_c A^T & C\Sigma_c C^T + D_C \end{pmatrix}^{-1}\begin{pmatrix} A\Sigma_c \\ C\Sigma_c \end{pmatrix} \quad \text{(A.2)}$$

which is also to prove

$$(\Sigma_c^{-1} + A^T D_A^{-1} A + C^T D_C^{-1} C)^{-1} = \Sigma_c - (\Sigma_c A^T, \quad \Sigma_c C^T)\begin{pmatrix} A\Sigma_c A^T + D_A & A\Sigma_c C^T \\ C\Sigma_c A^T & C\Sigma_c C^T + D_C \end{pmatrix}^{-1}\begin{pmatrix} A\Sigma_c \\ C\Sigma_c \end{pmatrix} \quad \text{(A.3)}$$

The left side can be simplified as

$$(\Sigma_c^{-1} + A^T D_A^{-1} A + C^T D_C^{-1} C)^{-1} = \left((I + A^T D_A^{-1} A\Sigma_c + C^T D_C^{-1} C\Sigma_c)\Sigma_c^{-1}\right)^{-1} = \Sigma_c (I + A^T D_A^{-1} A\Sigma_c + C^T D_C^{-1} C\Sigma_c)^{-1}$$

As Σ_c is invertible, it can be canceled the right sides, we have

$$(I + A^T D_A^{-1} A\Sigma_c + C^T D_C^{-1} C\Sigma_c)^{-1} = I - (A^T, \quad C^T)\begin{pmatrix} A\Sigma_c A^T + D_A & A\Sigma_c C^T \\ C\Sigma_c A^T & C\Sigma_c C^T + D_C \end{pmatrix}^{-1}\begin{pmatrix} A\Sigma_c \\ C\Sigma_c \end{pmatrix} \quad \text{(A.4)}$$

Both side multiplying by $(I + A^T D_A^{-1} A\Sigma_c + C^T D_C^{-1} C\Sigma_c)$, we have

$$I = (I + A^T D_A^{-1} A\Sigma_c + C^T D_C^{-1} C\Sigma_c)\left[I - (A^T, \quad C^T)\begin{pmatrix} A\Sigma_c A^T + D_A & A\Sigma_c C^T \\ C\Sigma_c A^T & C\Sigma_c C^T + D_C \end{pmatrix}^{-1}\begin{pmatrix} A\Sigma_c \\ C\Sigma_c \end{pmatrix}\right] \quad \text{(A.5)}$$

After expansion, Equation (A.5) can be simplified as

$$A^T D_A^{-1} A + C^T D_C^{-1} C = (I + A^T D_A^{-1} A\Sigma_c + C^T D_C^{-1} C\Sigma_c)(A^T, \quad C^T)\begin{pmatrix} A\Sigma_c A^T + D_A & A\Sigma_c C^T \\ C\Sigma_c A^T & C\Sigma_c C^T + D_C \end{pmatrix}^{-1}\begin{pmatrix} A \\ C \end{pmatrix} \quad \text{(A.6)}$$

which is

$$(A^T D_A^{-1}, \quad C^T D_C^{-1}) \begin{pmatrix} A \\ C \end{pmatrix} = (I + A^T D_A^{-1} A\Sigma_c + C^T D_C^{-1} C\Sigma_c)(A^T, \quad C^T) \begin{pmatrix} A\Sigma_c A^T + D_A & A\Sigma_c C^T \\ C\Sigma_c A^T & C\Sigma_c C^T + D_C \end{pmatrix}^{-1} \begin{pmatrix} A \\ C \end{pmatrix} \quad \text{(A.7)}$$

As $\begin{pmatrix} A \\ C \end{pmatrix}$ has more columns than rows, the only solution for Equation (A.7) is

$$(A^T D_A^{-1}, \quad C^T D_C^{-1}) = (I + A^T D_A^{-1} A\Sigma_c + C^T D_C^{-1} C\Sigma_c)(A^T, \quad C^T) \begin{pmatrix} A\Sigma_c A^T + D_A & A\Sigma_c C^T \\ C\Sigma_c A^T & C\Sigma_c C^T + D_C \end{pmatrix}^{-1} \quad \text{(A.8)}$$

Multiplied by $\begin{pmatrix} A\Sigma_c A^T + D_A & A\Sigma_c C^T \\ C\Sigma_c A^T & C\Sigma_c C^T + D_C \end{pmatrix}$ on both sides of (A.8), we have

$$(A^T D_A^{-1}, \quad C^T D_C^{-1}) \begin{pmatrix} A\Sigma_c A^T + D_A & A\Sigma_c C^T \\ C\Sigma_c A^T & C\Sigma_c C^T + D_C \end{pmatrix} = (I + A^T D_A^{-1} A\Sigma_c + C^T D_C^{-1} C\Sigma_c)(A^T, \quad C^T) \quad \text{(A.9)}$$

The left side of Equation (A.9) becomes a row vector:

$$\begin{aligned} &\left(A^T D_A^{-1}(A\Sigma_c A^T + D_A) + C^T D_C^{-1} C\Sigma_c A^T, \right. \\ &\left. A^T D_A^{-1} A\Sigma_c C^T + C^T D_C^{-1}(C\Sigma_c C^T + D_C)\right) \\ &= (A^T + A^T D_A^{-1} A\Sigma_c A^T + C^T D_C^{-1} C\Sigma_c A^T, C^T + C^T D_C^{-1} C\Sigma_c C^T \\ &+ \ A^T D_A^{-1} A\Sigma_c C^T) \end{aligned} \quad \text{(A.10)}$$

The right side of Equation (A.9) become another row vector

$$\begin{aligned} &(I + A^T D_A^{-1} A\Sigma_c + C^T D_C^{-1} C\Sigma_c)(A^T, \quad C^T) \\ &= (A^T + A^T D_A^{-1} A\Sigma_c A^T + C^T D_C^{-1} C\Sigma_c A^T, C^T \\ &+ C^T D_C^{-1} C\Sigma_c C^T + \ A^T D_A^{-1} A\Sigma_c C^T) \end{aligned} \quad \text{(A.11)}$$

Compare the right side of Equation (A.10) and (A.11), they are same. So, we have proved Equation (A.9) which is equation to prove Equation (A.1)

Appendix II

The appendix provides the Comparison of standard Kalman filter and our own improved filter.

Problem description:

We have 17 locations (n=17), the dose rates at time step t is $X_t = \begin{pmatrix} x_0 \\ \vdots \\ x_{16} \end{pmatrix}$, we want to predict $x_{10}, \cdots, x_{16}$ conditioning on $x_0, \cdots, x_9$, using standard Kalman Filter, set

$$Z_t = H_t X_t + v_t \tag{A.12}$$

Where $H_t = \left(I_{p\times p}, 0_{p\times(n-p)}\right) = \begin{pmatrix} 1 & & & 0 & \cdots & 0 \\ & \ddots & & \vdots & \ddots & \vdots \\ & & 1 & 0 & \cdots & 0 \end{pmatrix}$, and $p = 10, n = 17$ here.

v_t is the measurement errors, and $cov(v_t, v_s) = R_t \cdot 1(t == s)$. In our problem, $R_t = 0_{p\times p}$, but for genelization, we keep R_t here.

(1). Standard Kalman Filter

$$X_{t+1} = FX_t + u + w_t \tag{A.13}$$

and $cov(w_t, w_s) = \mathrm{Q_t} \cdot 1(t == s)$

$$Z_t = H_t X_t + v_t \tag{A.14}$$

$cov(v_t, v_s) = R_t \cdot 1(t == s)$. In our problem, $R_t = 0$.

It can be solved by two steps:

(a). Predict Step:

$$\hat{X}_{t+1|t} = \hat{X}_{t|t} + u \tag{A.15}$$

$$\hat{P}_{t+1|t} = \hat{P}_{t|t} + \mathrm{Q_t} \tag{A.16}$$

(b). Update Step

$$\hat{X}_{t+1|t+1} = \hat{X}_{t+1|t} + K_{t+1}\left[Z_{t+1} - H_{t+1}\,\hat{X}_{t+1|t}\right] \tag{A.17}$$

$$P_{t+1|t+1} = (I - K_{t+1}H_{t+1})P_{t+1|t}(I - K_{t+1}H_{t+1})^T + K_{t+1}R_{t+1}K_{t+1}^T \quad \text{(A.18)}$$

Where the gain matrix is given by

$$K_{t+1} = P_{t+1|t}H_{t+1}^T\left[H_tP_{t+1|t}H_{t+1}^T + R_{t+1}\right]^{-1} \quad \text{(A.19)}$$

Write $P_{t+1|t} = \begin{pmatrix} P_{p\times p} & P_{p\times(n-p)} \\ P_{(n-p)\times p} & P_{(n-p)\times(n-p)} \end{pmatrix}$ in blocks form. Then,

$$P_{t+1|t}H_{t+1}^T = \begin{pmatrix} P_{p\times p} & P_{p\times(n-p)} \\ P_{(n-p)\times p} & P_{(n-p)\times(n-p)} \end{pmatrix}\begin{pmatrix} I_{p\times p} \\ 0_{(n-p)\times p} \end{pmatrix} = \begin{pmatrix} P_{p\times p} \\ P_{(n-p)\times p} \end{pmatrix} \quad \text{(A.20)}$$

and $H_tP_{t+1|t}H_{t+1}^T = P_{p\times p}$, then the Kalman Gain becomes

$$K_{t+1} = \begin{pmatrix} P_{p\times p} \\ P_{(n-p)\times p} \end{pmatrix}\left(P_{p\times p} + R_t\right)^{-1} \quad \text{(A.21)}$$

And

$$\begin{aligned} I - K_{t+1}H_{t+1} &= I - \begin{pmatrix} P_{p\times p} \\ P_{(n-p)\times p} \end{pmatrix}\left(P_{p\times p} + R_t\right)^{-1}\left(I_{p\times p}, 0_{p\times(n-p)}\right) \\ &= \begin{pmatrix} R_t\left(P_{p\times p} + R_t\right)^{-1} & 0 \\ -P_{(n-p)\times p}\left(P_{p\times p} + R_t\right)^{-1} & I_{(n-p)\times(n-p)} \end{pmatrix} \end{aligned} \quad \text{(A.22)}$$

So, plug above into $P_{t+1|t+1}$, (A.22) can be re-written as

$$(I - K_{t+1}H_{t+1})P_{t+1|t}(I - K_{t+1}H_{t+1})^T = \begin{pmatrix} B_{p\times p} & B_{p\times(n-p)} \\ B_{(n-p)\times p} & B_{(n-p)\times(n-p)} \end{pmatrix} \quad \text{(A.23)}$$

Here, $B_{p\times p} = R_t\left(P_{p\times p} + R_t\right)^{-1}P_{p\times p}\left(P_{p\times p} + R_t\right)^{-1}R_t$,

$B_{p\times(n-p)} = R_t\left(P_{p\times p} + R_t\right)^{-1}R_t\left(P_{p\times p} + R_t\right)^{-1}P_{p\times(n-p)}$, and $B_{(n-p)\times p} = \left(B_{p\times(n-p)}\right)^T$.

We really care about $B_{(n-p)\times(n-p)}$ because its diagonal elements is the predicted uncertainty for $(x_{10}, \cdots, x_{16})$.

$$
\begin{aligned}
B_{(n-p)\times(n-p)} = & \underbrace{P_{(n-p)\times p}\left(P_{p\times p}+R_t\right)^{-1}P_{p\times p}\left(P_{p\times p}+R_t\right)^{-1}P_{p\times(n-p)}}_{(1)} \\
& - \underbrace{P_{(n-p)\times p}\left(P_{p\times p}+R_t\right)^{-1}P_{p\times(n-p)}}_{(2)} \\
& - \underbrace{P_{(n-p)\times p}\left(P_{p\times p}+R_t\right)^{-1}P_{p\times(n-p)}}_{(3)} + \underbrace{P_{(n-p)\times(n-p)}}_{(4)} \\
& = \underbrace{P_{(n-p)\times(n-p)} - P_{(n-p)\times p}\left(P_{p\times p}+R_t\right)^{-1}P_{p\times(n-p)}}_{(I)} \\
& - \underbrace{P_{(n-p)\times p}\left(P_{p\times p}+R_t\right)^{-1}R_t\left(P_{p\times p}+R_t\right)^{-1}P_{p\times(n-p)}}_{(II)}
\end{aligned}
\tag{A.24}
$$

We can see that the first term (green, term (I)) of (A.24) is kriging variance (with nugget effect), and the second term(blue, term(II)) is not defined, and will be 0 when $R_t = 0$.

So, when $R_t = 0$, the predicted variance equals kriging variance.

Special case: $\mathbf{R_t = 0}$

When $R_t = 0$, $B_{p\times p} = 0$, $B_{p\times(n-p)} = 0$, $K_{t+1}R_{t+1}K_{t+1}^T = 0$, and then

$$
P_{t+1|t+1} = \begin{pmatrix} 0_{p\times p} & 0_{p\times(n-p)} \\ 0_{(n-p)\times p} & P_{(n-p)\times(n-p)} - P_{(n-p)\times p}\left(P_{p\times p}\right)^{-1}P_{p\times(n-p)} \end{pmatrix} \tag{A.25}
$$

Then, due to

$$
\hat{P}_{t+1|t} = \hat{P}_{t|t} + \mathrm{Q_t} \tag{A.26}
$$

We have

$$
\hat{P}_{t+1|t} = \begin{pmatrix} Q_{p\times p} & Q_{p\times(n-p)} \\ Q_{(n-p)\times p} & Q_{(n-p)\times(n-p)} + \left(P_{(n-p)\times(n-p)} - P_{(n-p)\times p}\left(P_{p\times p}\right)^{-1}P_{p\times(n-p)}\right) \end{pmatrix} \tag{A.27}
$$

Then,

$$
\begin{aligned}
\hat{P}^{t+1|t+1}_{(n-p)\times(n-p)} &= \hat{P}^{t+1|t}_{(n-p)\times(n-p)} - \hat{P}^{t+1|t}{}_{(n-p)\times p}\left(\hat{P}^{t+1|t}_{p\times p}\right)^{-1}\hat{P}^{t+1|t}_{p\times(n-p)} \\
&= \Big(Q_{(n-p)\times(n-p)} \\
&\quad + \left(P_{(n-p)\times(n-p)} - P_{(n-p)\times p}\left(P_{p\times p}\right)^{-1}P_{p\times(n-p)}\right)\Big) \\
&\quad - Q_{(n-p)\times p}\left(Q_{p\times p}\right)^{-1}Q_{p\times(n-p)} \\
&= \left(P_{(n-p)\times(n-p)} - P_{(n-p)\times p}\left(P_{p\times p}\right)^{-1}P_{p\times(n-p)}\right) \\
&\quad + \left(Q_{(n-p)\times(n-p)} - Q_{(n-p)\times p}\left(Q_{p\times p}\right)^{-1}Q_{p\times(n-p)}\right) \\
&= \hat{P}^{t|t}_{(n-p)\times(n-p)} \\
&\quad + \left(Q_{(n-p)\times(n-p)} - Q_{(n-p)\times p}\left(Q_{p\times p}\right)^{-1}Q_{p\times(n-p)}\right)
\end{aligned}
\tag{A.28}
$$

So, each iteration of $\hat{P}^{t+1|t+1}_{(n-p)\times(n-p)}$, it will add a term $\left(Q_{(n-p)\times(n-p)} - Q_{(n-p)\times p}\left(Q_{p\times p}\right)^{-1}Q_{p\times(n-p)}\right)$,

Which is to say

$$
\begin{aligned}
\hat{P}^{t+1|t+1}_{(n-p)\times(n-p)} &= constant + (t+1) \\
&\quad \times \left(Q_{(n-p)\times(n-p)} - Q_{(n-p)\times p}\left(Q_{p\times p}\right)^{-1}Q_{p\times(n-p)}\right)
\end{aligned}
\tag{A.29}
$$

The constant is determined by the initial state of the covariance, which is to say,

$$
uncertainty \propto \sqrt{t + constant} \tag{A.30}
$$

So we have a plot like

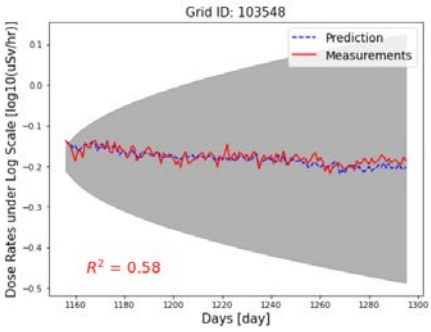

Fig. A.1 The plot of prediction (X_t) with uncertainty width ($\mu - 2\sigma,\ \mu + 2\sigma$), where μ is the mean of X_t over the plot range.

The prediction of X_t:

$$\hat{X}_{t+1|t} = \hat{X}_{t|t} + u \tag{A.31}$$

$$\hat{X}_{t+1|t+1} = \hat{X}_{t+1|t} + K_{t+1}\left[Z_{t+1} - H_{t+1}\,\hat{X}_{t+1|t}\right] \tag{A.32}$$

Then, $\left[Z_{t+1} - H_{t+1}\,\hat{X}_{t+1|t}\right] = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}_{t+1} - \begin{pmatrix} \hat{x}_1 \\ \vdots \\ \hat{x}_p \end{pmatrix}_{t+1|t}$ (with hat is prediction, without hat is observation)and $K_{t+1} = \begin{pmatrix} P_{p\times p} \\ P_{(n-p)\times p} \end{pmatrix}\left(P_{p\times p} + R_t\right)^{-1} = \begin{pmatrix} I_{p\times p} \\ P_{(n-p)\times p}P_{p\times p}^{-1} \end{pmatrix}$ ($here\ R_t = 0$)

Then,

$$\hat{X}_{t+1|t+1} = \begin{pmatrix} \hat{x}_1 \\ \vdots \\ \hat{x}_p \\ \vdots \\ \hat{x}_n \end{pmatrix}_{t+1|t} + \begin{pmatrix} I_{p\times p} \\ P_{(n-p)\times p}P_{p\times p}^{-1} \end{pmatrix}\left(\begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}_{t+1} - \begin{pmatrix} \hat{x}_1 \\ \vdots \\ \hat{x}_p \end{pmatrix}_{t+1|t}\right)$$

$$= \begin{pmatrix} x_1 \\ \vdots \\ x_p \\ \hat{x}_{t+1|t}[(p+1):n] - P_{(n-p)\times p}P_{p\times p}^{-1}\left(x[(p+1):n] - \hat{x}_{t+1|t}[(p+1):n]\right) \end{pmatrix} \tag{A.33}$$

where we define $\hat{x}_{t+1|t}[(p+1):n] = \begin{pmatrix} \hat{x}_{p+1} \\ \vdots \\ \hat{x}_n \end{pmatrix}_{t+1|t}$

The red term in (A.33) is exactly the same as the kriging prediction. So, we predicted that the standard Kalman filter will give exactly the same prediction as kriging method **(it is easy to verify that the standard kalman filter is still kriging even when $\boldsymbol{R_t \neq 0}$**, but the covariance matrix used are different)

(2). Our Own Improved Filter

The only difference is the update step:

$$\hat{P}_{t+1|t} = \hat{P}_{t|t} + \Sigma \cdot H_t^T \cdot (H_t \cdot \Sigma \cdot \mathrm{H_t^T} + R_t)^{-1} \cdot H_t \cdot \Sigma \tag{A.34}$$

Where $2\Sigma = \mathrm{Q_t}$,

We calculated that

$$\begin{aligned}
&\Sigma \cdot H_t^T \cdot (H_t \cdot \Sigma \cdot \mathrm{H_t^T} + R_t)^{-1} \cdot H_t \cdot \Sigma \\
&= \begin{pmatrix} \Sigma_{\mathrm{p\times p}}\left(\Sigma_{\mathrm{p\times p}} + R_t\right)^{-1}\Sigma_{\mathrm{p\times p}} & \Sigma_{\mathrm{p\times p}}\left(\Sigma_{\mathrm{p\times p}} + R_t\right)^{-1}\Sigma_{\mathrm{p\times (n-p)}} \\ \Sigma_{\mathrm{(n-p)\times p}}\left(\Sigma_{\mathrm{p\times p}} + R_t\right)^{-1}\Sigma_{\mathrm{p\times p}} & \Sigma_{\mathrm{(n-p)\times p}}\left(\Sigma_{\mathrm{p\times p}} + R_t\right)^{-1}\Sigma_{\mathrm{p\times (n-p)}} \end{pmatrix} \\
&= \begin{pmatrix} \Sigma_{\mathrm{p\times p}} & \Sigma_{\mathrm{p\times (n-p)}} \\ \Sigma_{\mathrm{(n-p)\times p}} & \Sigma_{\mathrm{(n-p)\times p}}\Sigma_{\mathrm{p\times p}}^{-1}\Sigma_{\mathrm{p\times (n-p)}} \end{pmatrix} (when\ R_t = 0)
\end{aligned} \tag{A.35}$$

So, when $R_t = 0$,

$$\hat{P}_{t+1|t} = \hat{P}_{t|t} + \begin{pmatrix} \Sigma_{\mathrm{p\times p}} & \Sigma_{\mathrm{p\times (n-p)}} \\ \Sigma_{\mathrm{(n-p)\times p}} & \Sigma_{\mathrm{(n-p)\times p}}\Sigma_{\mathrm{p\times p}}^{-1}\Sigma_{\mathrm{p\times (n-p)}} \end{pmatrix} \tag{A.36}$$

Similar to equation (A.29),

$$\begin{aligned}
\hat{P}_{(n-p)\times(n-p)}^{t+1|t+1} &= constant + (t+1) \\
&\times \left(Q_{(n-p)\times(n-p)} - Q_{(n-p)\times p}\left(Q_{p\times p}\right)^{-1}Q_{p\times(n-p)}\right)
\end{aligned} \tag{A.37}$$

Where Q is replaced by $\begin{pmatrix} \Sigma_{\mathrm{p\times p}} & \Sigma_{\mathrm{p\times (n-p)}} \\ \Sigma_{\mathrm{(n-p)\times p}} & \Sigma_{\mathrm{(n-p)\times p}}\Sigma_{\mathrm{p\times p}}^{-1}\Sigma_{\mathrm{p\times (n-p)}} \end{pmatrix}$, we have

$$\begin{aligned}
&\hat{P}_{(n-p)\times(n-p)}^{t+1|t+1} = constant + (t+1) \\
&\times \left(\Sigma_{\mathrm{(n-p)\times p}}\Sigma_{\mathrm{p\times p}}^{-1}\Sigma_{\mathrm{p\times (n-p)}} - \Sigma_{\mathrm{(n-p)\times p}}\Sigma_{\mathrm{p\times p}}^{-1}\Sigma_{\mathrm{p\times (n-p)}}\right) \\
&= constatnt + (t+1) \times 0 = constant
\end{aligned} \tag{A.38}$$

(it is also easy to verify that the covariance $\hat{P}_{(n-p)\times(n-p)}^{t+1|t+1}$ is constant when $R_t \neq 0$)

Equation (A.38) explained why our filter will give a constant covariance(uncertainty).

For prediction of X_t:

As we have previous calculated in Equation (A.33):

$$\hat{X}_{t+1|t+1} = \begin{pmatrix} x_1 \\ \vdots \\ x_p \\ \hat{x}_{t+1|t}[(p+1):n] - P_{(n-p)\times p}P_{p\times p}^{-1}\left(x[(p+1):n] - \hat{x}_{t+1|t}[(p+1):n]\right) \end{pmatrix}$$

Which only evolves $P_{(n-p)\times p}$ *and* $P_{p\times p}^{-1}$ (**upper left and off-diagonal terms of matrix P**). We also found that

$$P_{t+1|t+1} = \begin{pmatrix} 0_{p\times p} & 0_{p\times(n-p)} \\ 0_{(n-p)\times p} & P_{(n-p)\times(n-p)} - P_{(n-p)\times p}\left(P_{p\times p}\right)^{-1} P_{p\times(n-p)} \end{pmatrix} \quad \text{(A.39)}$$

So, the difference of our filter with standard Kalman filter is that, Kalman filter adds Q_t to $P_{t|t}$(or $P_{t+1|t+1}$) in each time step, while our filter adds

$$\begin{pmatrix} \Sigma_{\mathrm{p\times p}} & \Sigma_{\mathrm{p\times(n-p)}} \\ \Sigma_{\mathrm{(n-p)\times p}} & \Sigma_{\mathrm{(n-p)\times p}}\Sigma_{\mathrm{p\times p}}^{-1}\Sigma_{\mathrm{p\times(n-p)}} \end{pmatrix}$$
$$= \begin{pmatrix} \frac{1}{2}\mathrm{Q_{p\times p}} & \frac{1}{2}\mathrm{Q_{p\times(n-p)}} \\ \frac{1}{2}\mathrm{Q_{(n-p)\times p}}\mathrm{Q_{p\times p}^{-1}}\,\mathrm{Q_{p\times(n-p)}} & \Sigma_{\mathrm{(n-p)\times p}}\Sigma_{\mathrm{p\times p}}^{-1}\Sigma_{\mathrm{p\times(n-p)}} \end{pmatrix} \quad \text{(A.40)}$$

to $P_{t|t}$ in each time step. The lower right term(green term) is different with lower right of Q_t, but it doesn't matter because the kriging prediction of X_t only determined by the upper left term and two off-diagonal terms of covariance matrix. In a word, our filter will give the same prediction of X_t as standard Kalman filter.

Conclusion:

1. When $R_t = 0$, the standard Kalman filter will give the same prediction as kriging method for $x_{10}, \cdots, x_{16}$. When $R_t \neq 0$, the standard kalman filter is still Kriging method but difference covariance matrix predicted and used, so the prediction will be different with kriging.
2. When $R_t = 0$, the covariance given by standard Kalman filter will increase linearly with step t. the uncertainty is $\delta \propto \sqrt{t + constant}$
3. whether $R_t = 0$ or not, Our Filter will give a constant covariance matrix (uncertainty, confident interval width)
4. When $R_t = 0$, the prediction of X_t given by standard Kalman fitler, our Filter and kriging method, are the same.

Appendix III

This appendix provides the mathematical support for our new filter for dose rates temporal prediction.

1. Problem and Assumptions

Let X_t to be the dose rates vector of all the pixels at time step t, and $X_t = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, where n is the total number of pixels.

$$X_t = T_t + \varepsilon_t \tag{A.41}$$

Where, we assume $T_t = u \cdot t + b$ is a linear trend , ε_t is the random noise and

$$cov(X_t, X_s) = cov(\varepsilon_t, \varepsilon_s) = \begin{cases} \Sigma, & if\ t = s \\ 0, & otherwise \end{cases} \tag{A.42}$$

from equation (A.41), we have

$$X_{t+1} - X_t = (T_{t+1} + \varepsilon_{t+1}) - (T_t - \varepsilon_t) = (T_{t+1} - T_t) + \varepsilon_{t+1} - \varepsilon_t = u + \varepsilon_{t+1} - \varepsilon_t$$

or,

$$X_{t+1} = FX_t + u + w_t \tag{A.43}$$

where u is the average dose rate reduction per day; in this problem, $F = 1$; (F was put here just to be consistent with a standard Kalman filter)

$$w_t = \varepsilon_{t+1} - \varepsilon_t \tag{A.44}$$

It is easy to prove that

$$cov(w_s, w_t) = cov(\varepsilon_{s+1} - \varepsilon_s, \varepsilon_{t+1} - \varepsilon_t) = \begin{cases} 2\Sigma, & if\ \mathrm{t = s} \\ -\Sigma, & if\ |t - s| = 1 \\ 0, & otherwise \end{cases} \tag{A.45}$$

(Equation (A.45) above is different with Equation (A.43) in Ian Reid's lecture notes)

and

$$cov(X_t, w_s) = cov(\varepsilon_t, \varepsilon_{s+1} - \varepsilon_s) = \begin{cases} \Sigma, & if\ \mathrm{t = s + 1} \\ -\Sigma, & if\ \mathrm{t = s} \\ 0, & otherwise \end{cases} \tag{A.46}$$

the measurement vector Z_t represents k measurements

$$Z_t = H_t X_t + v_t \tag{A.47}$$

H_t is $k \times n$ matrix, where

$$H_t(i,:) = \begin{cases} (0,\cdots,1,\cdots 0), & for\ piont\ measurement\ of\ pixel\ j \\ \left(0,\cdots,\frac{1}{m},\cdots,\frac{1}{m}\cdots 0\right), & for\ airborne\ or\ car\ survey\ measurement \end{cases} \tag{A.48}$$

$$Cov(v_t, v_s) = \begin{cases} R_t, & t = s \\ 0, & otherwise \end{cases} \tag{A.49}$$

$$Cov(w_t, v_s) = 0, \quad for\ all\ t, s \tag{A.50}$$

Where R_t is a $k \times k$ diagonal matrix.

In Equation (A.48), if the i_{th} elements of Z_t is a point measurement, such as fixed-points measurement, monitoring post, walk survey, and that measurement locates within the j_{th} pixel., then $H_t(i,l) = \begin{cases} 1, & when\ l = j \\ 0, & otherwise \end{cases}$; if the i_{th} elements of Z_t is car or airborne survey, it can be represented as a weighted average of pixels around, i.e., for $j_{th}^{(1)}, j_{th}^{(2)}, \cdots, j_{th}^{(m)}$ pixel with weight $w_1, \cdots w_m$, then $H_t(i,l) = \begin{cases} w_t, & when\ l = j^{(t)}, for\ \ t = 1,2,\cdots m \\ 0, & otherwise \end{cases}$.

The initial system state has a known mean $X_{0|0}$ and covariance matrix $P_{0|0}$

$$\begin{cases} X_{0|0} = E[X_0] \\ P_{0|0} = E[(X_{0|0} - X_0)(X_{0|0} - X_0)^T] \end{cases} \tag{A.51}$$

Given the above assumptions the task is to determine, given a set of observations $Z_1, \cdots Z_{t+1}$, the estimation filter that at the $(t+1)_{th}$instance in time generates an optimal estimate of the state X_{t+1}, which we denote by $\hat{X}_{t+1}$, that minimizes the expectation of the squared-error loss function

$$E\left[\left\|X_{t+1} - \hat{X}_{t+1}\right\|^2\right] = E\left[(X_{t+1} - \hat{X}_{t+1})(X_{t+1} - \hat{X}_{t+1})^T\right] \tag{A.52}$$

2. Derivation

The prediction of state X_{t+1} based on the observations up to time t, $Z_1, \cdots Z_t$ namely $\hat{X}_{t+1}|Z^t$, where $Z^t = [Z_1, \cdots Z_t]$.

$$\hat{X}_{t+1|t} = E[X_{t+1}|Z_1, Z_2, \cdots Z_k] = E[X_{t+1}|Z^t] \tag{A.53}$$

Then the Predicted state is given by

$$\begin{aligned} \hat{X}_{t+1|t} = E[X_{t+1}|Z^t] &= E[F_t X_t + G_t u_t + w_t|Z^t] \\ &= F_t \hat{X}_{t|t} + G_t u_t + E(w_t|Z^t) \end{aligned} \tag{A.54}$$

$$\begin{aligned} P_{t+1|t} &= E\left[\left(X_{t+1} - \hat{X}_{t+1|t}\right)\left(X_{t+1} - \hat{X}_{t+1|t}\right)^T|Z^t\right] \\ &= E\left[\left(F_t X_t + w_t - F_t \hat{X}_{t|t}\right.\right. \\ &\quad \left.\left. - E(w_t|Z^t)\right)\left(F_t X_t + w_t - F_t \hat{X}_{t|t} - E(w_t|Z^t)\right)^T|Z^t\right] \\ &= E\left[\left(F_t\left(X_t - \hat{X}_{t|t}\right)\right.\right. \\ &\quad \left.\left. + \left(w_t - E(w_t|Z^t)\right)\right)\left(F_t\left(X_t - \hat{X}_{t|t}\right)\right.\right. \\ &\quad \left.\left. + \left(w_t - E(w_t|Z^t)\right)\right)^T|Z^t\right] \\ &= F_t P_{t|t} F_t^T + \underbrace{E\left[\left(w_t - E(w_t|Z^t)\right)\left(w_t - E(w_t|Z^t)\right)^T\middle|Z^t\right]}_{(1)} \\ &\quad + \underbrace{2E\left[F_t\left(X_t - \hat{X}_{t|t}\right)\left(w_t - E(w_t|Z^t)\right)^T|Z^t\right]}_{(2)} \end{aligned} \tag{A.55}$$

Item (1) in Equation (A.55) can be simplified as

$$\begin{aligned} &E\left[\left(w_t - E(w_t|Z^t)\right)\left(w_t - E(w_t|Z^t)\right)^T\middle|Z^t\right] \\ &= E[\|\varepsilon_{t+1} - \varepsilon_t - E[\varepsilon_{t+1} - \varepsilon_t||Z^t]\|^2|Z^t] \\ &= E[\|\varepsilon_{t+1} - \varepsilon_t - E[-\varepsilon_t||Z^t]\|^2|Z^t] \\ &= E[\|\varepsilon_{t+1}\|^2|Z^t] + E[\|\varepsilon_t - E[\varepsilon_t||Z^t]\|^2|Z^t] \\ &+ E[\varepsilon_{t+1}(\varepsilon_t - E[\varepsilon_t||Z^t])^T|Z^t] \\ &= E[\|\varepsilon_{t+1}\|^2] + E[\|\varepsilon_t - E[\varepsilon_t||Z^t]\|^2|Z^t] + 0 \\ &= \Sigma + \Sigma - \Sigma \cdot H_t^T \cdot (H_t \cdot \Sigma \cdot \mathrm{H_t^T} + R_t)^{-1} \cdot H_t \cdot \Sigma \\ &= 2\Sigma - \Sigma \cdot H_t^T \cdot (H_t \cdot \Sigma \cdot \mathrm{H_t^T} + R_t)^{-1} \cdot H_t \cdot \Sigma \end{aligned} \tag{A.56}$$

Note, $E[\|\varepsilon_t - E[\varepsilon_t||Z^t]\|^2|Z^t]$ in equation (A.56) is calculated by

$$\begin{aligned} E[\|\varepsilon_t - E[\varepsilon_t||Z^t]\|^2|Z^t] &= E[\|X_t - E[X_t||Z^t]\|^2|Z^t] = \Sigma_{X_t|Z^t} = \Sigma_{X_t|Z_t} \\ &= \Sigma_{XX} - \Sigma_{XZ}\Sigma_{ZZ}^{-1}\Sigma_{ZX} \\ &= \Sigma - \Sigma \cdot H_t^T \cdot (H_t \cdot \Sigma \cdot \mathrm{H_t^T} + R_t)^{-1} \cdot H_t \cdot \Sigma \end{aligned} \tag{A.57}$$

, which is also called the Kriging variance.

Item (2) in Equation (A.55) can be simplified as

$$\begin{aligned}
&2E\left[F_t\left(X_t-\hat{X}_{t|t}\right)\left(w_t-E(w_t|Z^t)\right)^T|Z^t\right]\\
&=2F_t\cdot E\left[\left(X_t-\hat{X}_{t|t}\right)\left(w_t-E(w_t|Z^t)\right)^T|Z^t\right]\\
&=2F_t\cdot E\left[\left(X_t-\hat{X}_{t|t}\right)\left(\varepsilon_{t+1}-\varepsilon_t-E(\varepsilon_{t+1}-\varepsilon_t|Z^t)\right)^T|Z^t\right]\\
&=-2F_t\cdot E\left[\left(X_t-\hat{X}_{t|t}\right)\left(\varepsilon_t-E(\varepsilon_t|Z^t)\right)^T|Z^t\right]=\\
&=-2F_t\cdot E\left[\left(X_t-\hat{X}_{t|t}\right)\left((T_t+\varepsilon_t)-E(T_t+\varepsilon_t|Z^t)\right)^T|Z^t\right]\\
&==-2F_t\cdot E\left[\left(X_t-\hat{X}_{t|t}\right)\left(X_t-E(X_t|Z^t)\right)^T|Z^t\right]\\
&=-2F_t\cdot E\left[\left(X_t-\hat{X}_{t|t}\right)\left(X_t-\hat{X}_{t|t}\right)^T|Z^t\right]=-2F\cdot\Sigma_{X_t|Z_t}\\
&=\ -2\mathrm{F}\cdot\Sigma+2\mathrm{F}\cdot\Sigma\cdot H_t^T\cdot(H_t\cdot\Sigma\cdot\mathrm{H_t^T}+R_t)^{-1}\cdot H_t\cdot\Sigma
\end{aligned}\tag{A.58}$$

Plug Equation (A.58), (A.56) into Equation (A.55), we have

$$\begin{aligned}
P_{t+1|t}=F_tP_{t|t}F_t^T+2(1-F)\Sigma+(2\mathrm{F}-1)\Sigma\cdot H_t^T\cdot(H_t\cdot\Sigma\cdot\mathrm{H_t^T}+R_t)^{-1}\\
\cdot H_t\cdot\Sigma
\end{aligned}\tag{A.59}$$

In our problem, $F=1$, Equation (A.59) can be even simplified as

$$P_{t+1|t}=F_tP_{t|t}F_t^T+\Sigma\cdot H_t^T\cdot(H_t\cdot\Sigma\cdot\mathrm{H_t^T}+R_t)^{-1}\cdot H_t\cdot\Sigma\tag{A.60}$$

Then, we assume that the estimation of $\hat{X}_{t+1|t+1}$ is a linear weighted sum of K_{t+1} and K'_{t+1},

$$\hat{X}_{t+1|t+1}=K'_{t+1}\hat{X}_{t+1|t}+K_{t+1}Z_{t+1}\tag{A.61}$$

Where K'_{t+1} and K_{t+1} are weighting or gain matrices (of different sizes). Our problem now is reduced to finding the K'_{t+1} and K_{t+1} that minimize the conditional mean squared estimation error where of course the estimation error is given by:

$$\tilde{X}_{t+1|t+1}=\hat{X}_{t+1|t+1}-X_{k+1}\tag{A.62}$$

3. The Unbiased Condition

Four the filter to be unbiased, we require that $E\left(\hat{X}_{t+1|t+1}\right)=E(X_{k+1})$

$$\begin{aligned} E\left[\hat{X}_{t+1|t+1}\right] &= E\left[K'_{t+1}\hat{X}_{t+1|t} + K_{t+1}H_{t+1}X_{t+1} + K_{t+1}v_{t+1}\right] \\ &= K'_{t+1}E\left[\hat{X}_{t+1|t}\right] + K_{t+1}H_{t+1}E[X_{t+1}] + K_{t+1}E[v_{t+1}] \\ &= K'_{t+1}E\left[\hat{X}_{t+1|t}\right] + K_{t+1}H_{t+1}E[X_{t+1}] \end{aligned} \tag{A.63}$$

The prediction is unbiased because

$$\begin{aligned} E\left[\hat{X}_{t+1|t}\right] &= E[F_tX_t + G_tu_t + w_t|Z^t] \\ &= F_tE\left[X_{t|t}\right] + G_tu_t + E[E[w_t|Z^t]] = F_tE\left[X_{t|t}\right] + G_tu_t \\ &= E[X_{t+1}] \end{aligned} \tag{A.64}$$

$(note\ E[E[w_t|Z^t]] = 0)$

Combine Equation (A.63) and (A.64),

$$E\left[\hat{X}_{t+1|t+1}\right] = (K'_{t+1} + K_{t+1}H_{t+1})E[X_{t+1}] \tag{A.65}$$

The condition that $\hat{X}_{t+1|t+1}$ be unbiased reduce the requirement to

$$K'_{t+1} + K_{t+1}H_{t+1} = I\ or\ K'_{t+1} = I - K_{t+1}H_{t+1} \tag{A.66}$$

Then, Equation (A.61) can be rewritten as

$$\begin{aligned} \hat{X}_{t+1|t+1} &= (I - K_{t+1}H_{t+1})\hat{X}_{t+1|t} + K_{t+1}Z_{t+1} \\ &= \hat{X}_{t+1|t} + K_{t+1}(Z_{t+1} - H_{t+1})\hat{X}_{t+1|t} \end{aligned} \tag{A.67}$$

K_{t+1} above is known as Kalman Gain.

4. The Error Covariance

$$\begin{aligned} X_{t+1} - \hat{X}_{t+1|t+1} &= X_{t+1} - \left((I - K_{t+1}H_{t+1})\hat{X}_{t+1|t} + K_{t+1}Z_{t+1}\right) \\ &= X_{t+1} \\ &\quad - \left((I - K_{t+1}H_{t+1})\hat{X}_{t+1|t} + K_{t+1}(H_{t+1}X_{t+1} + v_{t+1})\right) \\ &= (I - K_{t+1}H_{t+1})\left(X_{t+1} - \hat{X}_{t+1|t}\right) - K_{t+1}v_{t+1} \end{aligned} \tag{A.68}$$

So, the updated covariance matrix is

$$
\begin{aligned}
P_{t+1|t+1} &= E\left[\left(X_{t+1} - \hat{X}_{t+1|t+1}\right)\left(X_{t+1} - \hat{X}_{t+1|t+1}\right)^T\right] \\
&= E\left[\left\|(I - K_{t+1}H_{t+1})\left(X_{t+1} - \hat{X}_{t+1|t}\right) - K_{t+1}v_{t+1}\right\|^2\right] \\
&= (I - K_{t+1}H_{t+1})E\left[\left\|X_{t+1} - \hat{X}_{t+1|t}\right\|^2\right](I - K_{t+1}H_{t+1})^T \\
&+ K_{t+1}E[\|v_{t+1}\|^2]K_{t+1}^T \\
&- 2(I - K_{t+1}H_{t+1})E\left[\left(X_{t+1} - \hat{X}_{t+1|t}\right)v_{t+1}\right]K_{t+1}^T
\end{aligned} \tag{A.69}
$$

and with

$$
\begin{aligned}
E[\|v_{t+1}\|^2] &= R_{t+1} \\
E\left[\left\|X_{t+1} - \hat{X}_{t+1|t}\right\|^2\right] &= P_{t+1|t} \\
E\left[\left(X_{t+1} - \hat{X}_{t+1|t}\right)v_{t+1}\right] &= 0
\end{aligned}
$$

We obtain

$$
P_{t+1|t+1} = (I - K_{t+1}H_{t+1})P_{t+1|t}(I - K_{t+1}H_{t+1})^T + K_{t+1}R_{t+1}K_{t+1}^T \tag{A.70}
$$

5. Choose the Kalman Gain

$$
\begin{aligned}
L &= \min_{K_{t+1}} E\left[\tilde{X}_{t+1|t+1}^T \tilde{X}_{t+1|t+1} | Z^{t+1}\right] \\
&= \min_{K_{t+1}} trace\left(E\left[\tilde{X}_{t+1|t+1}\tilde{X}_{t+1|t+1}^T | Z^{t+1}\right]\right) \\
&= \min_{K_{t+1}} trace\left(P_{t+1|t+1}\right)
\end{aligned} \tag{A.71}
$$

For any matrix A and a symmetric matrix B,

$$
\frac{\partial}{\partial A} trace(ABA^T) = 2AB \tag{A.72}
$$

So,

$$
\frac{\partial L}{\partial K_{t+1}} = -2(I - K_{t+1}H_{t+1})P_{t+1|t}H_{t+1}^T + 2K_{t+1}R_{t+1} = 0 \tag{A.73}
$$

Re-arranging gives an equation for the gain matrix

$$
K_{t+1} = P_{t+1|t}H_{t+1}^T\left(H_{t+1}P_{t+1|t}H_{t+1}^T + R_{t+1}\right)^{-1} \tag{A.74}
$$

6. Summary for Key Equations of Each Step

Predict Step:

$$\hat{X}_{t+1|t} = \hat{X}_{t|t} + u \tag{A.75}$$

$$\hat{P}_{t+1|t} = \hat{P}_{t|t} + \Sigma \cdot H_t^T \cdot (H_t \cdot \Sigma \cdot \mathrm{H_t^T} + R_t)^{-1} \cdot H_t \cdot \Sigma \tag{A.76}$$

Update Step

$$\hat{X}_{t+1|t+1} = \hat{X}_{t+1|t} + K_{t+1}\left[Z_{t+1} - H_{t+1}\,\hat{X}_{t+1|t}\right] \tag{A.77}$$

$$P_{t+1|t+1} = (I - K_{t+1}H_{t+1})P_{t+1|t}(I - K_{t+1}H_{t+1})^T + K_{t+1}R_{t+1}K_{t+1}^T \tag{A.78}$$

Where the gain matrix is given by

$$K_{t+1} = P_{t+1|t}H_{t+1}^T\left[H_t P_{t+1|t} H_{t+1}^T + R_{t+1}\right]^{-1} \tag{A.79}$$