Title

Optimizing long-term monitoring of radiation air-dose rates after the Fukushima Daiichi Nuclear Power Plant

Authors

Sun, Dajie
Wainwright, Haruko M
Oroza, Carlos A
et al.

Publication Date

2020-09-01

DOI

10.1016/j.jenvrad.2020.106281

Peer reviewed

Keywords: Fukushima, dose rate, network optimization, machine learning, long-term monitoring

1. Introduction

Since the release of radionuclides in March 2011, air dose rates (i.e., the ambient dose equivalent rate (µSv/h) at 1 m above the ground) near the Fukushima Daiichi Nuclear Power Plant (NPP) have been steadily decreasing (Saito, 2016, 2019). The designated evacuation area has shrunk to 370 km^2, which is less than 3% of the Fukushima Prefecture (Fukushima Prefectural Government, 2017). Currently, radiocesium (^{134}Cs and ^{137}Cs) is the main contaminant of concern in the environment, since it is a major contributor to air-dose rates (Saito, 2016). Many studies have documented reduction of the air dose rates faster than expected from physical decay as a result of both physical and ecological decay (Kinase et al., 2014, 2015, 2017; Saito, 2016, 2019; Wainwright et al., 2018). In addition, other studies have found that the extensive decontamination effort in the region has played a critical role in this recovery process (Yasutaka et al., 2013; Wainwright et al., 2018).

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 eight years, the monitoring program is expected to transition to long-term monitoring beyond 10 years. 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 martial text S1). 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 searching the optima to place 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.

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 study 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.

2. Methodology

Since our methodology is 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 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. 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.

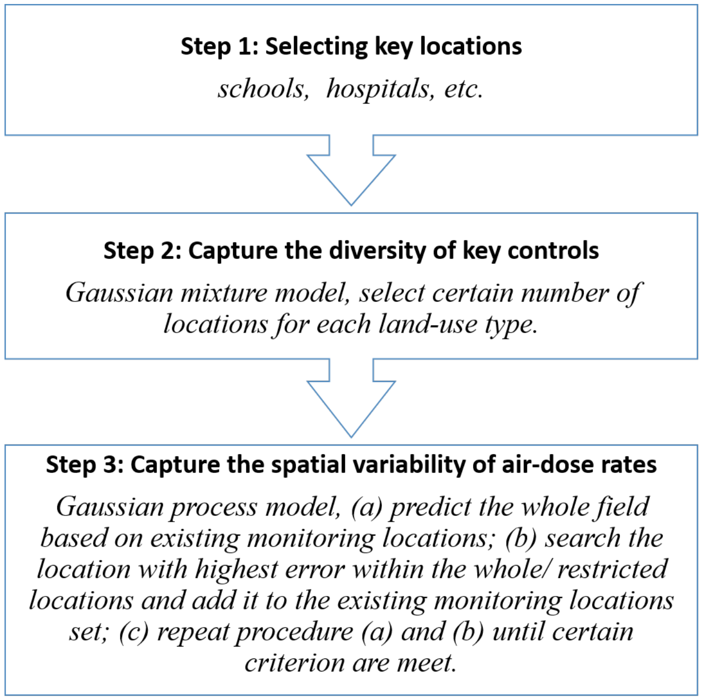

Fig. 1. Flowchart of the optimization Method

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 a 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 $\boldsymbol{x} = [x^{\text{lat}}; x^{\text{lon}}; x^{\text{elevation}}; x^{\text{direction}}; x^{\text{distance}}; x^{\text{landuse}}]$) is a product of a finite number of latent (unobserved) components (i.e., measurements) that follow Gaussian distributions, where x^{lat}, x^{lon}, $x^{\text{elevation}}$, $x^{\text{direction}}$, 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: $N(\boldsymbol{x} \mid \mu, \Sigma)$ 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:

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

where

$$\sum_{m=1}^{M} \pi_m = 1 \tag{2}$$

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. Results and discussion

We demonstrated our methodology 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. 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.

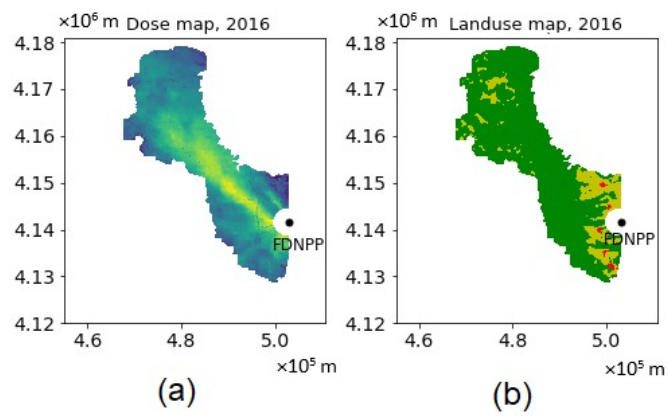

Fig. 2. Input data maps: (a) 2016 integrated air-dose-rate map in log10 microSv/hr, (b) land-cover map. In (b), the green region is forest, the yellow region is cropland, and red region is urban area. The unit of coordinates is meter(m), the black dots in each subplot are the location of Fukushima Daiichi 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 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. 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. 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. 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. 3a, 34 points in Fig. 3b, and 284 points in Fig. 3c.

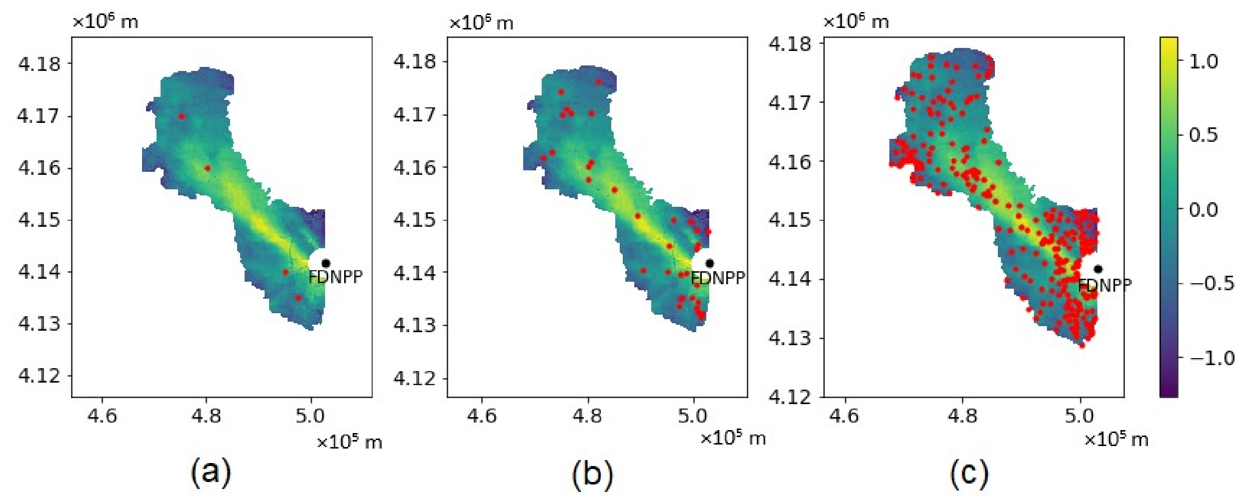

Fig. 3. Proposed monitoring locations over the 2016 integrated map (in log10 microSv/hr) in Case 1 after: (a) Step 1, (b) Step 2 and (c) Step 3. In the figures, the red circles are the monitoring locations.

The overall estimation error (RMSE) is plotted against the number of monitoring locations in Step 3 (Fig. 4). Fig. 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. 4b. Fig. 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.

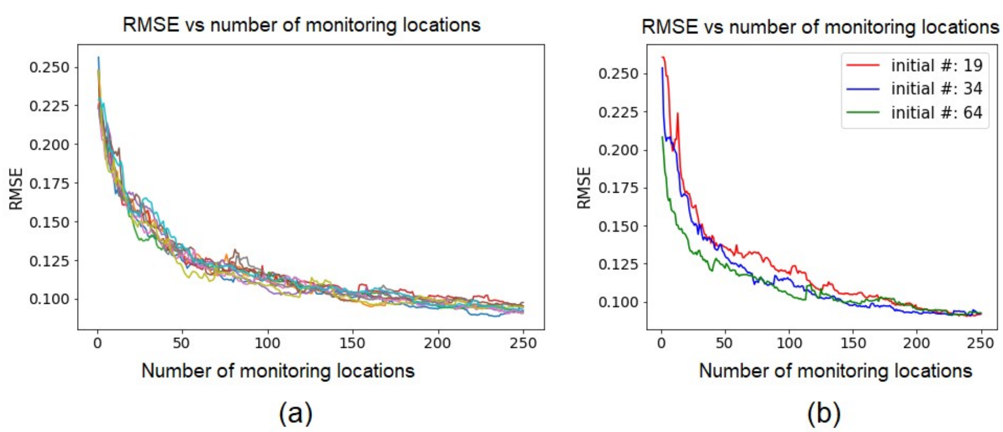

Fig. 4. RMSE vs number of monitoring locations in Step 3 in Case 1: (a) initial monitoring locations number is 34, random sampled top 3% highest estimation error, MC simulated 10 times; (b) random sampled top 3% highest estimation error, with initial monitoring locations number 19, 34, 64.

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.

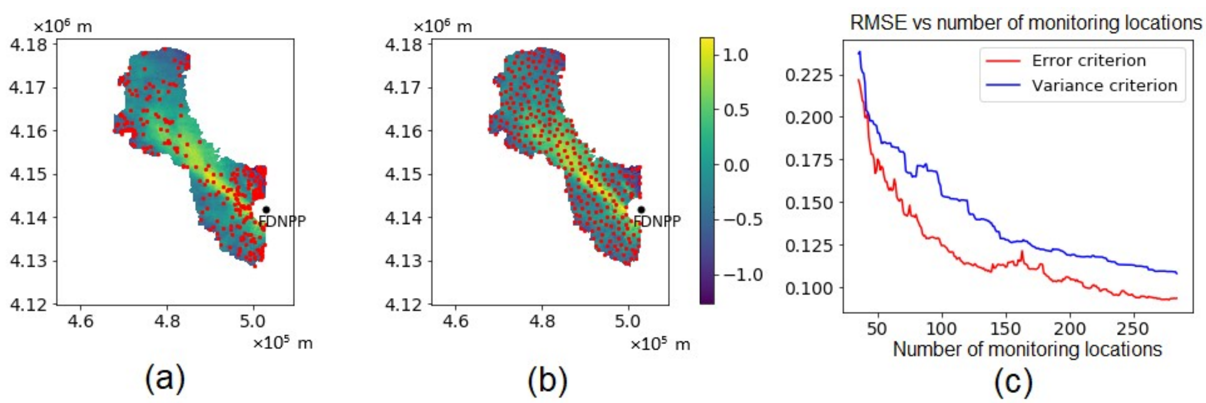

Fig. 5. Monitoring locations configurations by choosing (a) the top 3% of the estimation errors and (b) the top 3% estimation variance and (c) RMSE curves using error criterion vs variance criterion.

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. 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. 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. 5b). In Fig. 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% 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. 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.

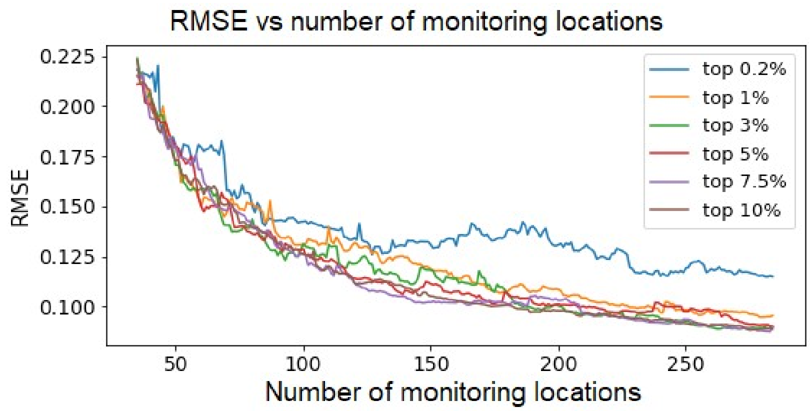

Fig. 6. RMSE as a function of the number of monitoring locations for different parameters within the error-based criterion. In the legend, top 10% means randomly sampling one pixel out of the pixels with top 10% highest error for next sensor, etc.

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. 7 shows the sampling locations at each step for Case 2. Similar to the monitoring configuration without location restriction (Fig. 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 = 4.9×10^5 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.

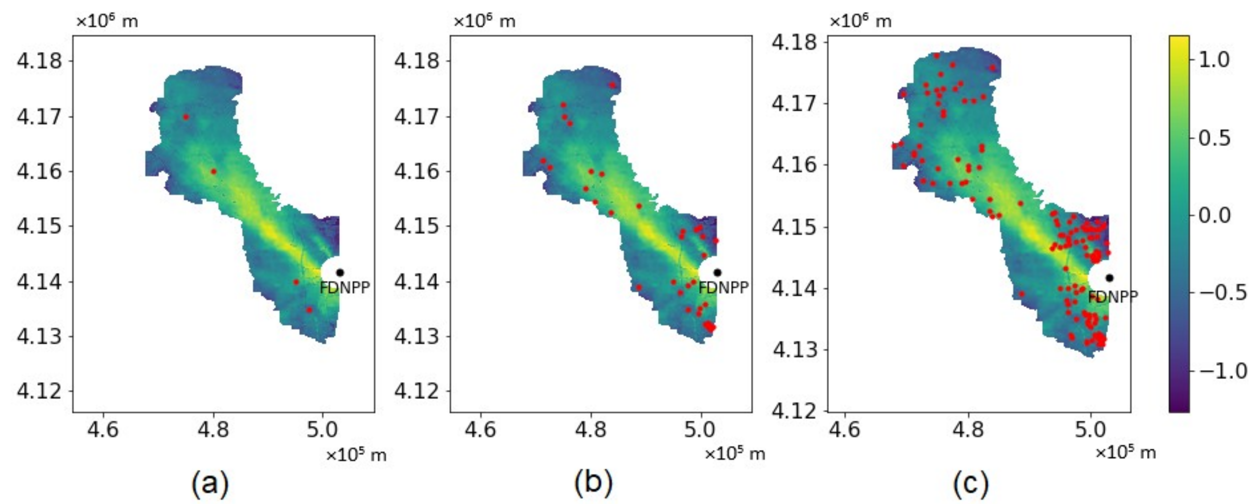

Fig. 7. Proposed sampling locations over the 2016 integrated map (in log10 microSv/hr) in Case 2 after: (a) Step 1, (b) Step 2 and (c) Step 3. In the Fig.s, the red dot are the sampling locations.

Fig. 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. 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. 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. 8b. As consistent with the no-restriction case (Fig. 4b), the number of Step 2 locations do not affect the convergence of RMSE.

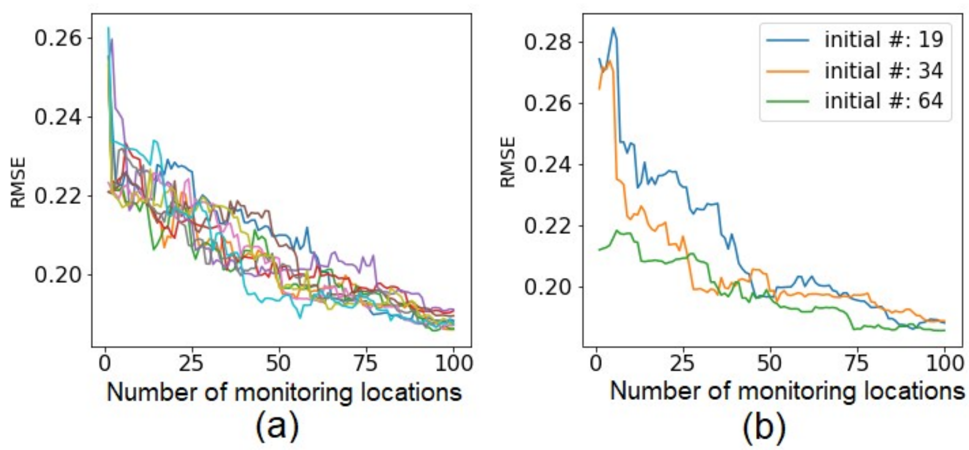

Fig. 8. RMSE vs number of monitoring locations in Step 3 in Case 2: (a) initial locations number is 34, random sampled top 3% highest error, MC simulated 10 times; (b) random sampled top 3% highest error.

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. 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. 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. 9b), there are some locations that are always chosen (red dots in Fig. 9b), while some are less likely to be selected (purple dots in Fig. 9b). These more-selected locations tend to be located within the high heterogeneity region. In addition, the spatial pattern is consistent with Fig. 7c, which is just one instance of the simulation. Fig. 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. 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.

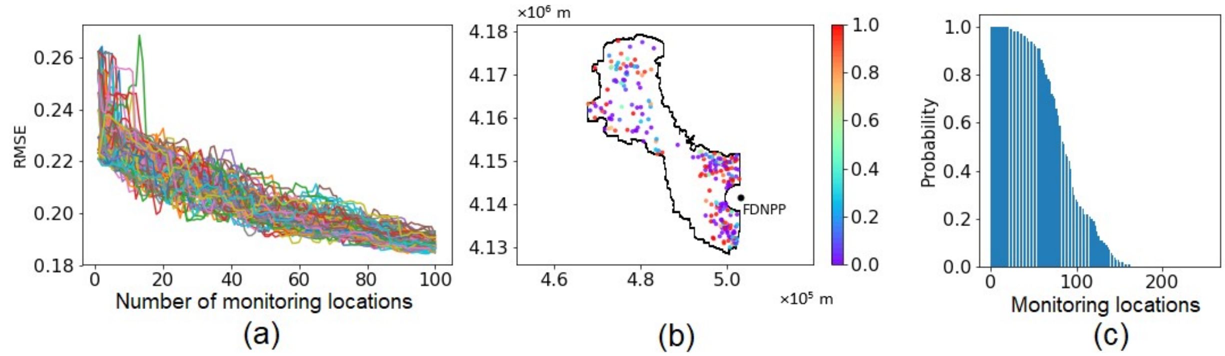

Fig. 9. Results from generating 100 sets of monitoring locations based on the MC simulation: (a) the RMSE curves of the 100 simulations, as a function of monitoring locations, (b) probabilistic map of the monitoring locations among monitoring posts based on 100 simulations, and (c) probability of each location sorted from high (1.0) to low (0.0). In (b), the color of each dot is an indicator of probability.

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.

Acknowledgements

The environmental monitoring data in this study were acquired during the projects commissioned by the Japan Nuclear Regulatory Agency. We thank the people who contributed to collecting the data and compiling them into the JAEA database. Funding for this work was provided by Japan Atomic Energy Agency under Award No. AWD00000626, as part of Work for Others funding from Berkeley Lab, provided by the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. We also thank Dr. Shun'ichi Hisamatsu and two anonymous reviewers for helpful comments to improve our manuscript.

References

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. J. Environ. Radioact. 2018; 192:385-398. https://doi.org/10.1016/j.jenvrad.2018.07.009.

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

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).

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

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

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.

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.

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.

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., 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.

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.

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.

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.

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

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).

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

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, 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, 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.

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., & 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, 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.

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.

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

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.

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.

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.

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

List of Figures

Fig. 1. Flowchart of the optimization Method

Fig. 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 meter(m), the black dots in each subplot are the location of Fukushima Daiichi Nuclear Power Plant(FDNPP).

Fig. 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.

Fig. 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.

Fig. 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

Fig. 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.

Fig. 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 Fig.s, the red dot are the sampling locations.

Fig. 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.

Fig. 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.