A Multiscale Bayesian Data Integration Approach for Mapping Air Dose Rates around the Fukushima Daiichi Nuclear Power Plant – 17166

Haruko Wainwright*, Akiyuki Seki**, Jinsong Chen*, Kimiaki Saito**
* Lawrence Berkeley National Laboratory
** Japan Atomic Energy Agency

ABSTRACT

This paper presents a multiscale data integration method to estimate the spatial distribution of air dose rates in the regional scale around the Fukushima Daiichi Nuclear Power Plant. We integrate various types of datasets, such as ground-based walk and car surveys, and airborne surveys, all of which have different scales, resolutions, spatial coverage, and accuracy. This method is based on geostatistics to represent spatial heterogeneous structures, and also on Bayesian hierarchical models to integrate multiscale, multitype datasets in a consistent manner. The Bayesian method allows us to quantify the uncertainty in the estimates, and to provide the confidence intervals that are critical for robust decision-making. Although this approach is primarily data-driven, it has great flexibility to include mechanistic models for representing radiation transport or other complex correlations. We demonstrate our approach using three types of datasets collected at the same time over Fukushima City, Japan: (1) coarse-resolution airborne surveys covering the entire area, (2) car surveys along major roads, and (3) walk surveys in multiple neighborhoods. Results show that the method can successfully integrate three types of datasets in a consistent manner, and create an integrated map (including the confidence intervals) of air dose rates over the domain in high resolution.

INTRODUCTION

Radiation measurements and monitoring in the region around the Fukushima Daiichi NPP have been performed continuously since the accident [1,2]. Since the accident, radiation measurements have been conducted using various techniques and platforms such as walk, car and airborne surveys. Such mapping efforts are essential to protect the public, guide decontamination efforts, estimate the amount of decontamination waste, and plan the return of evacuated residents.

With many data survey types available, however, it has become clear that there are discrepancies among them in terms of measured dose-rate values, even collected at the same time and same locations. Most discrepancies are due to the fact that each type of data has a different level of accuracy and a different support scale (i.e., support volume, resolution). For example, airborne surveys measure average dose rates over a much larger area (typically a several-hundred-meter radius) than ground-based measurements (~several tens of meters). Such averaging becomes particularly problematic, because detailed surveys with handheld monitors (i.e., walk surveys) often find that soil contamination and radiation dose rates are both highly heterogeneous, having many hotspots [3]. Physics-based radiation transport

modeling by Malins et al. [4] also showed that the uncertainty in the above-ground dose-rate measurements is associated with the horizontal heterogeneity of cesium distribution in soil rather than the vertical distribution.

In addition to the resolution and support volumes, the different data types also have different spatial sampling density and different spatial coverage. Car survey data are, for example, limited to the locations along roads, even though their data provide relatively high-resolution dose rate, both in space and in time [5]. Walk surveys are further limited in spatial coverage and often clustered in several neighborhoods, since it takes time and physical labor for a person to walk around with a device, even though the walk survey could represent the health risk of an average person walking in the neighborhood. The airborne survey has the large spatial coverage at the regional scale, although the resolution is low due to the spatial averaging effect [6].

Recently, Wainwright et al. [7] developed a Bayesian hierarchical modeling approach to integrate multiscale datasets (i.e., car, walk and airborne surveys), and also to estimate the spatial distribution of air dose rates at 1 m above the ground surface in high resolution across the regional scale. Bayesian approaches have been used in environmental science as a flexible and expandable framework to integrate multiscale datasets [8]. Wainwright et al. [7] demonstrated how this method could be used to integrate low-resolution and high-uncertainty airborne and car survey data with high-resolution (but sparse) walk survey data in a consistent manner.

This study aims to apply their methodology to the area with a larger spatial extent and higher dose rates within the evacuation zones. We used the datasets collected in Fukushima Prefecture, Japan, in November 2014 by Japan Atomic Energy Agency (JAEA). Such integrated and more resolved maps are important for assessing the return of the residents and their potential exposure dose.

METHODS

A Bayesian hierarchical model consists of a series of statistical sub-models mainly in two categories: data models and process models. The process models—in this context—describe the spatial pattern (or map) of dose rates within the domain, representing the spatial trend and heterogeneity of contamination. The data models connect this pattern and the actual data, given measurement errors. These data models can represent, for example, a direct ground-based measurement or a function of the pattern—for example, spatial averaging over a certain area for a low-resolution airborne dataset. The overall model— a series of statistical submodels—is flexible and expandable, able to include complex correlations (such as correlations with land use, soil texture, or topography) or various observations. Once all the submodels are developed, we can estimate the parameters, as well as the radiation map and its confidence interval, using sampling methods or optimization methods.

To develop an integrated map, we denote the radiation dose rate at i-th pixel by y_i,

where $i = 1, \ldots, n$. We also denote three datasets by three vectors, representing the airborne survey data $\boldsymbol{z}_A$ (each data point is represented by $z_{A,j}$, where $j = 1,\ldots, m_A$), car survey data $\boldsymbol{z}_C$ (each data point is represented by $z_{C,j}$, where $j = 1,\ldots, m_C$), and walk survey data (each data point is represented by $z_{W,j}$, where $j = 1,\ldots, m_W$). The goal is to estimate the posterior distribution of the radiation dose-rate map $\boldsymbol{y}$ (i.e., the vector representing the radiation dose rates at all the pixels) conditioned on three datasets ($\boldsymbol{z}_A$, $\boldsymbol{z}_C$ and $\mathbf{z}_W$), written as $p(\mathbf{y} \mid \mathbf{z}_A, \mathbf{z}_C, \boldsymbol{z}_W)$. By applying Bayes' rule, we can rewrite this posterior distribution as:

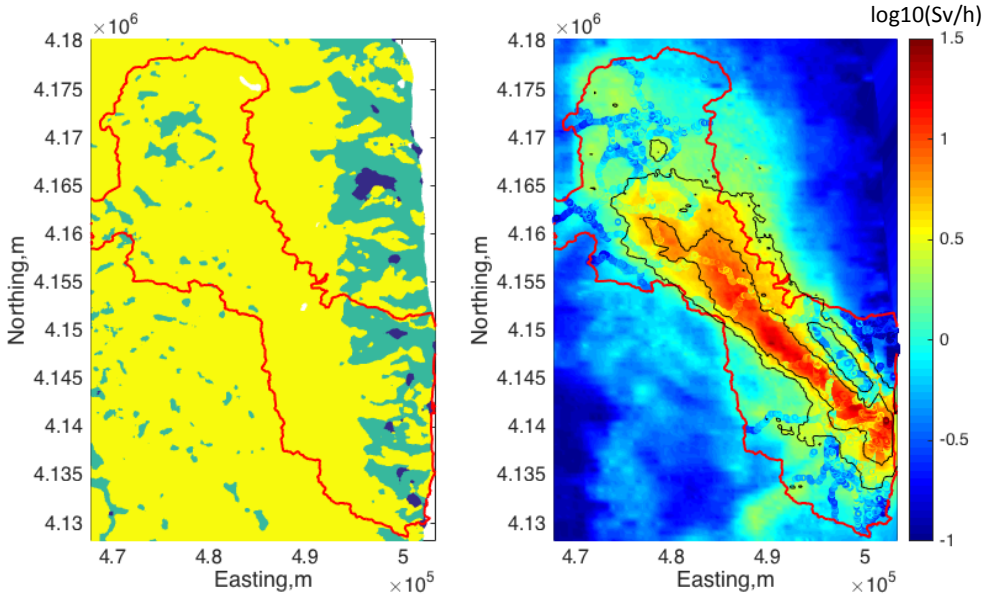
$$p(\boldsymbol{y} \mid \boldsymbol{z}_A, \boldsymbol{z}_C, \boldsymbol{z}_W) \propto p(\boldsymbol{z}_A \mid \boldsymbol{y})\, p(\boldsymbol{z}_C \mid \boldsymbol{y}) p(\boldsymbol{y} \mid \boldsymbol{z}_W) \quad (1)$$

We assume that the datasets are conditionally independent of each other, given the air dose rate distribution.

The detail descriptions of mathematical formulation are available in Wainwright et al. [7]. The first distributions $p(\boldsymbol{z}_A \mid \boldsymbol{y})$ and $p(\mathbf{z}_C \mid \boldsymbol{y})$ represent the data models to connect the low-resolution data (i.e., airborne and car survey data) and air dose rate distribution at each pixel ($\boldsymbol{y}$). The spatial average functions are included in these conditional distributions. We selected simple averaging for representing car surveys, and weighted averaging for airborne surveys, the weights of which were computed based on the radiation transport simulations [4]. The third distribution $p(\boldsymbol{y} \mid \mathbf{z}_W)$ represents the process model (i.e., geostatistical model [9]) to describe the spatial pattern given the measured dose rates in the walk surveys. We also assume that the parameters in the data and process models are estimated and well constrained through EDA, and hence they are fixed during this Bayesian estimation.

RESULTS

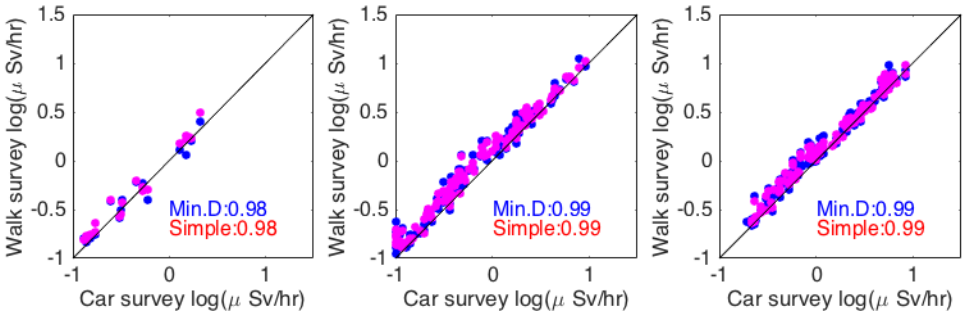
In this study, we integrated the three types of datasets over the evacuation zone (726km^2) collected by JAEA in Fall 2014. In addition, we used the high-resolution land-use and land-cover map of Japan (version 14.02) created by Japan Aerospace Exploration Agency [10]. The domain is mostly covered by a forest (Figure 1a), although there are some urban areas (i.e., towns and villages) and cropland near the Pacific Ocean. By overlaying the car, walk and air survey data (Figure 1b), we see that the car and walk survey data show smaller-scale variability than the airborne data, and that the air survey data overestimates the air dose rate. Compared to the datasets used in Wainwright et al. [7], the dose rate is generally higher in the evacuation zone. In addition, the evacuation zone has a larger spatial coverage of forested areas, and less human activities.



(a) (b)

Figure 1. (a) Land cover types (blue = urban, green = cropland and yellow = forest) and (b) original datasets (the car and walk survey data over the airborne data). In (b), the car and walk survey can be seen as lines and dots of lower-dose values compared to the surrounding air survey data. In (a) and (b), the red polygons are the evacuation zone extent. In (b), the black contour lines are the threshold of 20mSv/yr and 50mSv/yr.

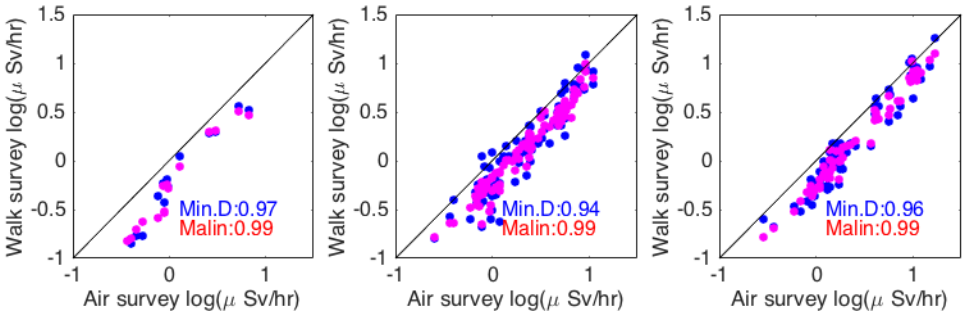
The comparison between the car and walk surveys (Figure 2) shows that the car survey underestimates the walk survey data, which is consistent with Wainwright et al. [7]. The correlation coefficients are high (>0.98) in all the land-use types. Simple averaging improves the correlation coefficients only slightly, which is different from Wainwright et al. [7]. This would be due to the fact that there are less human activities on the roads.



(a) (b) (c)

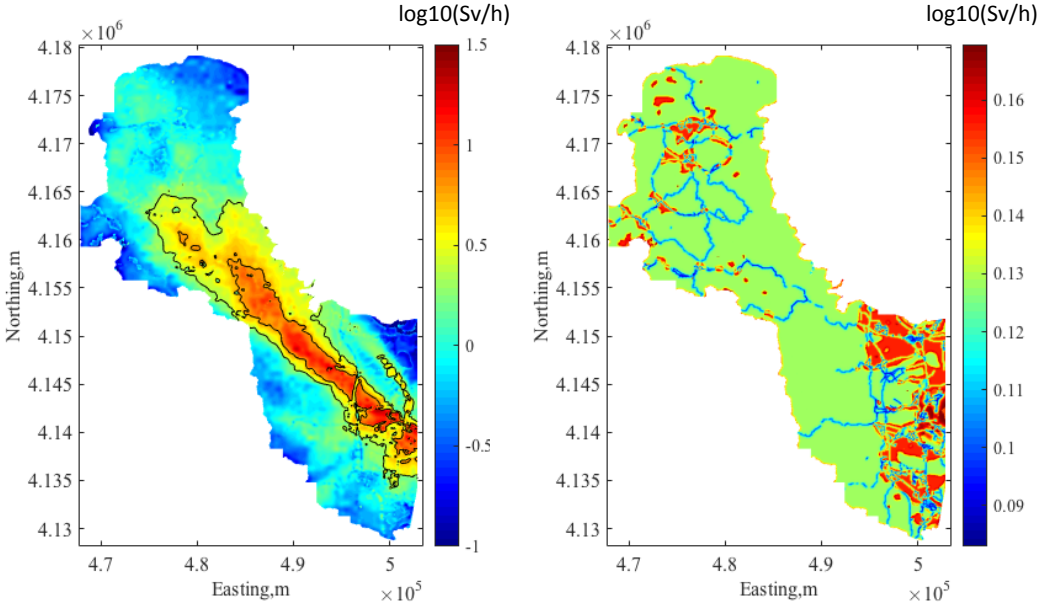
Figure 2. Comparison between the car and walk survey data at co-located points in the evacuation zone (Fall 2014): (a) urban, (b) cropland, and (c) forest. The blue circles are the co-located points identified by the minimum distance. The pink circles are the averaged data (simple averaging). The numbers in each plot is the correlation coefficients based on the minimum distance (Min.D) and simple averaging (Simple).

The comparison between the air and walk surveys (Figure 3) shows that the air survey overestimates the dose rate compared to the walk survey data. The blue circles (co-located points identified by the minimum distance) show larger scattering compared to the walk-car comparison, possibly due to the large spatial heterogeneity within the airborne survey footprints. The averaging improves the correlation to 0.99 in all land-cover types.



(a) (b) (c)

Figure 3. Comparison between the air and walk survey data at co-located points in the evacuation zone (Fall 2014): (a) urban, (b) cropland, and (c) forest. The blue circles are the co-located points identified by the minimum distance. The pink circles are the averaged data, , using the weighted average based on the radiation transport simulations by Malins et al. [4]. The numbers in each plot is the correlation coefficients based on the minimum distance (Min.D) and weighted averaging (Malins)



(a) (b)

Figure 4. (a) Integrated radiation dose-rate map within the evacuation zone, and (b) the map of standard deviation, representing the uncertainty in the estimation. In (a), the black contour lines are the threshold of 20mSv/yr and 50mSv/yr.

The estimated map (mean field) from the data integration in Figure 4a shows more detailed and finer-resolution heterogeneity than the original airborne data (Figure1b), although the general trend is very similar. The systematic bias (or shift) in the airborne data was also corrected. The area of above 20mSv/yr is 218 km^2 in the integrated map, which is smaller than the one in the original airborne survey (305 km^2). Correcting such overestimation would be important, since 20mSv/yr is often used as the threshold value for the policy decisions. The map of standard deviation (Figure 4b) represent the estimation uncertainty. The urban areas have increased standard deviation, since the variability of air dose rates is high. The areas near the roads have smaller uncertainty, since the air dose rates are more constrained by the car and walk surveys.

Figure 5 shows the validation result to evaluate the performance of the data integration and the dose-rate estimation. One hundred of the walk survey data were excluded from the estimation in each land-use type, and they were used for validation purposes. Without the data integration, the airborne data at co-located points (blue dots) exhibit larger scatters and a systematic bias compared to the walk survey data. After the data integration, the predicted values (based on our approach and the three datasets) are tightly distributed around the one-to-one line and are mostly included in the 99% confidence interval. Figure 5 shows that this method successfully estimates the fine-resolution dose-rate map based on the spatially sparse walk and car survey data and airborne data. Having such a confidence interval would be useful for practical applications, such as estimating the range of the potential health effects or evaluating the return scenarios for the evacuated zones.

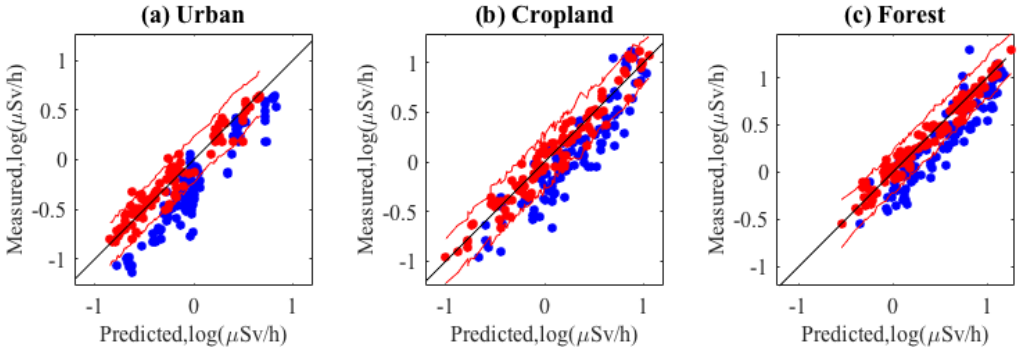


Figure 5. Comparison between the predicted and measured air dose rates (log-transformed) at the walk-survey data locations not used for the estimation (Fall 2014 data). The red dots represent the predicted values based on the data integration method; the blue dots are the co-located airborne data without using the integration. The black line is the one-to-one line; the red lines are the 95% confidence intervals.

CONCLUSION

In this study, we showed the second demonstration of out Bayesian data integration approach for the Fukushima evacuation zones with high air dose rates. The Bayesian hierarchical model developed by Wainwrithe et al. [7] was effective to

integrate multiscale, multitype dose-rate measurements, and also to create the high-resolution air dose rate over the large spatial extent. The estimated map captured more detailed spatial heterogeneity than the regional airborne survey data. In addition, the integrated map reflects the physical understanding of the radionuclide distribution found in the datasets; for example, the urban area has higher uncertainty and larger confidence intervals, particularly in the areas without walk survey datasets. Having the confidence intervals of possible air dose rates would help estimate health risks or plan the return to the evacuation zones in more robust manner in a more robust manner. In addition, the integrated map would be a critical initial condition for predicting the future air dose rates in the Fukushima region.

REFERENCES

1. Saito, K., & Onda, Y., Outline of the national mapping projects implemented after the Fukushima accident. Journal of Environmental Radioactivity, 139(C), 240–249. doi:10.1016/j.jenvrad.2014.10.009, 2015
2. Mikami, S., Maeyama, T., Hoshide, Y., Sakamoto, R., Sato, S., Okuda, N., et al., The air dose rate around the Fukushima Dai-ichi Nuclear Power Plant: its spatial characteristics and temporal changes until December 2012. Journal of Environmental Radioactivity, 139(C), 250–259. doi:10.1016/j.jenvrad.2014.08.020, 2015
3. Japan Atomic Energy Agency, Establishing the methodology to understand the long-term impact of radionuclides released during the Fukushima Daiichi Nuclear Power Plant accident, JAEA Report, http://fukushima.jaea.go.jp/initiatives/cat03/entry05.html, (in Japanese), 2012
4. Malins, A., Kurikami, H., Nakama, S., Saito, T., Okumura, M., Machida, M., & Kitamura, A., 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, 2016
5. Tsuda, S., Yoshida, T., Tsutsumi, M., & Saito, K., Characteristics and verification of a car-borne survey system for dose rates in air: KURAMA-II. Journal of Environmental Radioactivity, 139(C), 260–265. doi:10.1016/j.jenvrad.2014.02.028, 2015.
6. Torii, T., Y. Sanada, T. Sugita, A. Kondo, Y. Shikaze and Y. Urabe, Investigation of Radionuclide Distribution Using Aircraft for Surrounding Environmental Survey from Fukushima Daiichi Nuclear Power Plant, JAEA-Technology 2012-036, (in Japanese), 2012.
7. Wainwright, H.M., A. Seki, J.Chen and K. Saito, A multiscale Bayesian data integration approach for mapping radionuclide contamination in the regional scale, Journal of Environmental Radioactivity, 2016.
8. Wikle, C.K., R.F. Milliff, D. Nychka and L. M. Berliner, Spatiotemporal hierarchical Bayesian modeling: tropical ocean surface winds, J. Am. Stat. Assoc., 96(454), 382-397, 2001.
9. Diggle, P. and Ribeiro, P. J., Model-based geostatistics. Springer Science & Business Media, 2007.
10. Takahashi, M., Nasahara, K. N., Tadono, T., Watanabe, T., Dotsu, M., Sugimura, T., & Tomiyama, N., JAXA High Resolution Land-Use and Land-

Cover Map of Japan. In Geoscience and Remote Sensing Symposium (IGARSS), 2013 IEEE International (pp. 2384-2387). IEEE, 2013.

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. We also thank Marilyn Saarni for English editing. 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.