

A Multiscale Bayesian Data Integration Approach for Mapping Radionuclide Contamination

Haruko Murakami Wainwright, Masahiko Okumura
and Kimiaki Saito

Abstract This chapter presents a multiscale data fusion method to estimate the spatial distribution of radiation dose rates at regional scale around the Fukushima Daiichi nuclear power plant. We integrate various types of radiation measurements, such as ground-based hand-held monitors, car-borne surveys, and airborne surveys, all of which have different 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. Although this approach is primarily data-driven, it has great flexibility, enabling it to include mechanistic models for representing radiation transport or other complex processes and correlations. As a first demonstration, we show a simple case study in which we integrate two datasets over Fukushima City, Japan: (1) coarse-resolution airborne survey data covering the entire city and (2) high-resolution ground-based car-borne data along major roads. Results show that the method can successfully integrate two datasets in a consistent manner and generate an integrated map of air dose rates over the domain in high resolution. A further advantage of this method is that it can quantify estimation errors and estimate confidence intervals, which are necessary for modeling and for robust policy planning. In addition, evaluating correlations among different datasets provides us with various insights into the characteristics of each dataset, as well as radionuclide transport and distribution. The resulted maps have started being used by local governments to plan the residents' return, and they are expected to be used for additional policy decisions in the future such as decontamination planning.

Keywords Radiation dose rate mapping · Bayesian hierarchical models · Geostatistics

H.M. Wainwright (✉) · K. Saito
Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA
e-mail: hmwainwright@lbl.gov

M. Okumura
Japan Atomic Energy Agency, 2-2-2 Uchisawai-Cho, Chiyoda, Tokyo 100-0011, Japan

1 Introduction

Radiation measurements and monitoring in the region around the Fukushima Daiichi nuclear power plant (NPP) have been performed continuously since the accident [1, 2]. Such mapping is essential for protecting the public, guiding decontamination efforts, estimating the amount of decontamination waste, and also in planning the return of evacuated residents. Radiation measurements have been conducted using various techniques such as portable hand-held monitors, car-borne surveys, and airborne surveys. Soil samples have been collected to assess the extent of contamination in the terrestrial environment [3].

Despite such large-scale and continuous efforts, there are still significant challenges in mapping the radiation dose rates and radionuclide contamination. Detailed ground-based measurements have revealed that the radiation dose rates and contamination are both quite heterogeneous, often with many hotspots [4]. Although many datasets are becoming available, it has been difficult to integrate those datasets, since each type of data has a different level of accuracy and represents a different support scale (i.e., spatial coverage and resolution). For example, although ground-based car-borne data provide high-resolution air dose rates, car-borne data are limited to the locations along roads [5]. Airborne surveys have been extensively used to map dose rates in the regional spatial coverage (e.g., 100 km radius) [6]; such data are, however, known to exhibit some discrepancies with co-located ground-based measurements. These discrepancies result mainly from the differences in support volume, since airborne measurements represent the average dose rate over a much larger area (typically a several-hundred-meter radius) than ground-based measurements (\sim several tens of meters).

In environmental science, monitoring and spatial-temporal mapping of various properties—such as CO_2 concentration, wind velocity or reactive transport properties in subsurface—have been the focus of extensive research. Although many traditional datasets have been sparse in time and space, more recently available datasets can cover large areas, such as remote sensing data in atmospheric/terrestrial sciences and data from geophysical techniques in subsurface science. Such datasets, however, are known to have some discrepancy with traditional point measurements, because they tend to have a larger support volume (or lower resolution), such that each pixel represents the average of heterogeneous properties in the vicinity. Various approaches have been proposed to integrate remote-sensing or geophysical datasets with traditional point measurements [7–9]. Many of them are based on geostatistics, a powerful tool for characterizing spatial heterogeneity (or correlation) structure based on available datasets [8–10]. In addition, a Bayesian framework is often used to integrate different datasets consistently and also to quantify the uncertainty associated with the estimated maps [7, 9].

In this study, we develop a Bayesian data-integration approach to estimate the spatial distribution of air dose rates and radionuclide contamination in high resolution across the regional scale (several kilometers to several tens of kilometers). We integrate various radiation measurements, with particular focus on airborne and

ground-based measurements. Geostatistical approaches are used to identify spatial correlations and represent small-scale heterogeneity that is not resolved in the coarse-resolution airborne data. We employ a Bayesian hierarchical model for integrating the low-resolution airborne data and sparse ground-based data. We demonstrate our approach using the airborne and car-borne datasets collected in Fukushima City, Japan, in 2012. This integration aims to provide a more resolved and integrated dose-rate map, and also quantify the uncertainty associated with the map for modeling and for policy planning. This approach is now being used to map the radiation dose rate in the Fukushima region, particularly in the evacuation zone [11, 12].

2 Methodology

Our approach is based on a Bayesian hierarchical model [7, 8], which typically consists of three types of statistical submodels: (1) data models, (2) process models, and (3) prior models. The process models describe the spatial pattern (or map) of dose rates within the domain, given the parameters defined by the prior models. Geostatistical models are often used as process models based on the spatial heterogeneity structure identified by available datasets. 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 prior models determine the distributions or ranges of the parameters based on pre-existing information. 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 fully quantify the uncertainty, here we will use the Markov-chain Monte-Carlo method. When the domain size and the number of pixels are large, optimization-based methods will be used to obtain the mean estimate and its asymptotic confidence intervals.

In this chapter, we show one simple example; the integration of airborne and car-borne measurements. We assume that each point of the airborne measurements is the weighted average of radiations from radionuclides distributed over the ground. To develop an integrated map, we denote the radiation dose rate at i -th pixel by y_i , where $i = 1, \dots, n$. We also denote two vectors, representing the airborne data \mathbf{z}_A (each data point is represented by $z_{A,j}$, where $j = 1, \dots, m_A$) and car-borne data \mathbf{z}_V (each data point is represented by $z_{V,j}$, where $j = 1, \dots, m_V$). The goal is to estimate the posterior distribution of the radiation dose-rate map \mathbf{y} (i.e., the vector representing the radiation dose rate at all the pixels) conditioned on two datasets (\mathbf{z}_A and \mathbf{z}_V), written as $p(\mathbf{y}|\mathbf{z}_A, \mathbf{z}_V)$. By applying Bayes' rule, we can rewrite this posterior distribution as:

$$p(\mathbf{y}|\mathbf{z}_A, \mathbf{z}_V) \propto p(\mathbf{z}_A | \mathbf{y}) p(\mathbf{y} | \mathbf{z}_V). \quad (1)$$

The first conditional distribution $p(\mathbf{z}_A | \mathbf{y})$ is a data model representing the airborne data as a function of the dose-rate distribution \mathbf{y} . We assume a spatial weighted averaging function of the dose rate map:

$$z_{A,j} = \sum_{i \in C_j} w_{i,j} y_i + \varepsilon_j. \quad (2)$$

where C_j represents the pixels within the spatial averaging range, $w_{i,j}$ is the weight determined by the distance between i -th pixel and j -th airborne data point, and ε_j is an error associated with each data point. We assume an inverse square distance function for the weights, $w_{i,j}$. The weight can be computed by radiation transport equations [11].

The range is typically considered to be equal to the flight height, following Torii et al. [6]. We assume that the error ε_j includes not only measurement errors associated with hardware (such as instrument noises) but also the uncertainty associated with other factors (such as the variable height of buildings, small-scale spatial variability). In this example, we assume that ε_j follows an independent normal distribution with zero-mean and the error variance σ_A can be determined from the correlation analysis between two datasets.

As a process model, we assume that \mathbf{y} is a multivariate Gaussian field described by geostatistical parameters. For simplicity, we assume that the ground-based measurements have insignificant errors compared to the airborne measurements, and hence we use the ground measurements as conditional points to constrain the distribution of \mathbf{y} as $p(\mathbf{y}|\mathbf{z}_V)$. Since two conditional distributions are both multivariate Gaussian, we can derive an analytical form of this posterior distribution as a multivariate normal distribution with mean $\mathbf{Q}^{-1}\mathbf{g}$ and covariance \mathbf{Q}^{-1} , where $\mathbf{Q} = \Sigma_c^{-1} + \mathbf{A}^T \mathbf{D}^{-1} \mathbf{A}$ and $\boldsymbol{\mu} = \boldsymbol{\mu}_c + \mathbf{A}^T \mathbf{D}^{-1} \mathbf{z}_A$ [7]. In \mathbf{Q} and \mathbf{g} , $\boldsymbol{\mu}_c$ and Σ_c are the conditional mean and covariance given the ground-based data (\mathbf{z}_V) and geostatistical parameters. \mathbf{D} is the data-error covariance matrix; each of the diagonal components is σ_A . The matrix \mathbf{A} is n -by- m_A sparse matrices, where $A_{ij} = w_{ij}$ if i -th pixel is within the range C_j ; otherwise A_{ij} is 0. We may directly sample \mathbf{y} or estimate the mean (or expected) dose map by $\mathbf{Q}^{-1}\mathbf{g}$. Although the example shown here is quite simple, we may add more complexity, such as, for example, physics-based radiation transport models (instead of weighted averaging) to represent the airborne data or data-derived correlations between dose rates and land use or topography [2].

3 Demonstration

In this example, we applied the developed method to the air-dose-rate data collected in Fukushima, Japan. We used the ground-based car-borne data from the second car-borne survey (December 5–28, 2011; <http://radioactivity.nsr.go.jp/en/contents/>

5000/4688/24/255_0321_ek.pdf), as well as the airborne data from the fourth airborne monitoring survey (October 22–November 5, 2011; http://radioactivity.nsr.go.jp/en/contents/4000/3179/24/1270_1216.pdf). We assume that the effect of radiocesium decay is negligible between the two surveys. The airborne data were processed and converted to the values equivalent to the dose rate one meter above the ground surface.

We first compare the co-located data values of the car-borne datasets to the airborne datasets. Direct comparison (Fig. 1a) shows significant scatters in the higher dose region, although the datasets are clearly correlated (the correlation coefficient is 0.78). The car-borne data have larger variability (larger variance), suggesting that small-scale variability is averaged out in the airborne data. When we take into account the weighted spatial average for the airborne data (Fig. 1b), the correlation improves significantly (the correlation coefficient is 0.84). We would note that the airborne data values are systematically higher than the car-borne ones, with and without spatial averaging. There are several possible reasons for such a shift: (1) there could be calibration issues in the airborne data, and (2) the center of roads (where the car-borne data are collected) is known to have lower contamination than the side of the roads or undisturbed land. For demonstration purposes, we assume that the car-borne data are still accurate representation of the radiation map in this example.

Using the correlation between the airborne and car-borne datasets that we found in Fig. 1, we determined the error variance σ_A as well as the shift factor. In addition, geostatistical parameters were estimated based on the variogram analysis of the car-borne datasets, representing the spatial correlation structure of small-scale heterogeneity. Figure 2a shows the airborne survey data in the part of Fukushima; the eastern portion of the domain has higher dose rates, possibly because the area

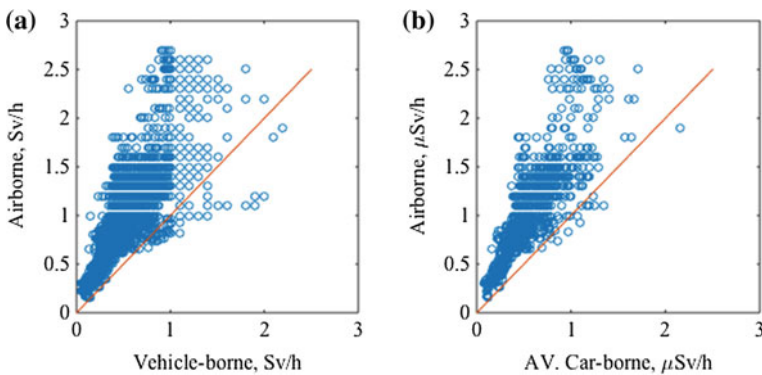


Fig. 1 Comparison between the car-borne data and airborne data: **a** direct comparison of data values and **b** including spatial averaging for the airborne data

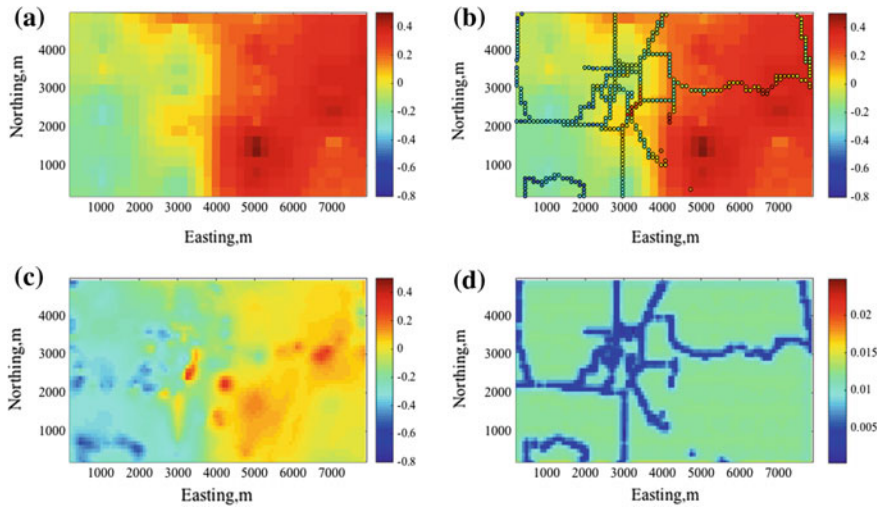
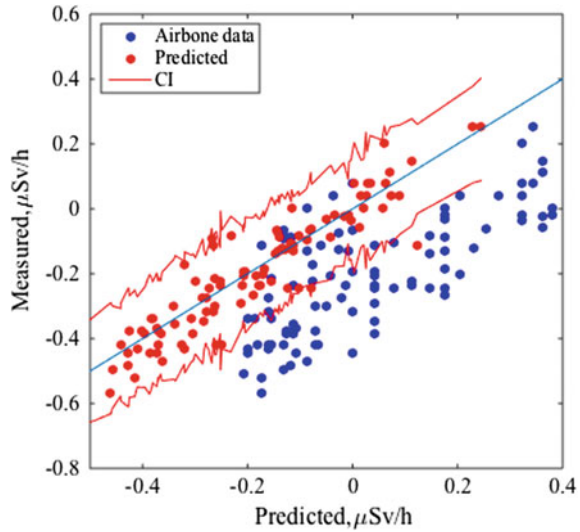


Fig. 2 **a** Airborne dose-rate data over Fukushima City (December 2011), **b** car-borne data (colored circles) over the airborne data (colored map), **c** the estimated integrated dose-rate map (mean field) based on the developed data integration, and **d** the estimation variance. In all the plots, the data values are log-transformed

lies along the initial plume direction and is also a forested area. By overlaying the car-borne and airborne data (Fig. 2b), we see that the car-borne data show smaller-scale variability than the airborne data, and that the airborne data overestimates the air dose rate. The estimated map (mean field) from the data integration in Fig. 2c shows more detailed and finer-resolution heterogeneity than the original airborne data (Fig. 2a). The systematic shift in airborne data was also corrected. As shown in Fig. 2d, the estimation variance is smaller near the car-borne data points, since the model includes spatial correlation.

Figure 3 shows the validation result to evaluate the performance of the data integration and the dose-rate estimation. One hundred of the car-borne data are excluded from the estimation, and used for validation purposes. Without the data integration, the airborne data (blue dots) have large scatters and a systematic shift compared to the car-borne measured data. After the data integration, the predicted values (based on the both airborne and car-borne data at other locations) are tightly distributed around the one-to-one line and are mostly included in the 95% confidence interval. Figure 3 shows that this method successfully estimates the fine-resolution dose-rate map based on the spatially sparse car-borne data and coarse-resolution airborne data. Having such a confidence interval would be useful for practical applications, such as estimating the range of the potential health effects or estimating the decontamination waste volume.

Fig. 3 Comparison between the predicted and measured air dose rates (log-transformed) at the car-borne data locations not used for the estimation. The *red dots* represent the predicted values based on the data integration method; the *blue dots* are the airborne data before the integration. The *blue line* is the one-to-one line; the *red lines* are the 95% confidence intervals



4 Summary and Future Work

In this chapter, we described a multiscale hierarchical Bayesian method for integrating multiscale, multitype dose-rate measurements. As an example, we illustrated how this method could be used to integrate coarse-resolution airborne data and fine-resolution (but sparse) car-borne data in a consistent manner, with the estimation uncertainty quantified. Although the current example model is still simple, results have suggested that the effective combination of ground-based data and airborne data could provide detailed and integrated maps of radiation air dose rates at regional scale around the Fukushima Daiichi NPP. In addition, this method could quantify estimation errors or confidence intervals, representing the uncertainty associated with the integrated maps. We also showed that statistical analyses could provide various insights into both the characteristics of each dataset and the spatial trend of contamination, which would be useful for predicting future radiation levels at the regional scale.

Further improvement was made to improve the estimation approach by including other information, such as the correlations between dose rates and land use and/or topography [11]. Physics-based radiation transport models are used to replace the spatial averaging function to accurately represent airborne data [11]. In the future, spatiotemporal integration—by integrating spatially sparse but continuous-time monitoring data and temporally sparse but spatially extensive data, such as airborne data—will be carried out to provide a detailed map of the air dose rate and radionuclide contamination at regional scale, at any given location and time, including their confidence interval.

These integrated maps of radiation dose rates has started being used in the local governments and agencies to plan the return of residents [12]. These maps can be used to provide additional important measures with the estimates of upper and lower bounds such as the expected cost and waste amount from decontamination. Having the confidence intervals or the upper/lower bounds of those measures will be helpful for planning the decontamination in more robust manner and also preparing for the worst-case scenarios. In addition, the detailed map and its confidence interval are critical to inform the public of the contamination level as well as the progress of decontamination for better decision-making.

References

1. K. Saito, Y. Onda. Outline of the national mapping projects implemented after the Fukushima accident. *J. Environ. Radioactivity* **139**(C), 240–249 (2015). doi:10.1016/j.jenvrad.2014.10.009
2. S. Mikami, T. Maeyama, Y. Hoshide, R. Sakamoto, S. Sato, N. Okuda et al. The air dose rate around the Fukushima Dai-ichi Nuclear Power Plant: its spatial characteristics and temporal changes until December 2012. *J. Environ. Radioactivity* **139**(C), 250–259. (2015). doi:10.1016/j.jenvrad.2014.08.020
3. K. Saito, I. Tanihata, M. Fujiwara, T. Saito, S. Shimoura, T. Otsuka et al, Detailed deposition density maps constructed by large-scale soil sampling for gamma-ray emitting radioactive nuclides from the Fukushima Dai-ichi nuclear power plant accident. *J. Environ. Radioactivity* **139**(C), 308–319. (2015). doi:10.1016/j.jenvrad.2014.02.014
4. 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 (2012), <http://fukushima.jaea.go.jp/initiatives/cat03/entry05.html>. (in Japanese)
5. S. Tsuda, T. Yoshida, M. Tsutsumi, K. Saito, Characteristics and verification of a car-borne survey system for dose rates in air: KURAMA-II. *J. Environ. Radioactivity*, **139**(C), 260–265 (2015). doi:10.1016/j.jenvrad.2014.02.028
6. T. Torii, Y. Sanada, T. Sugita, A. Kondo, Y. Shikaze, Y. Urabe, Investigation of radionuclide distribution using aircraft for surrounding environmental survey from Fukushima Daiichi nuclear power plant. JAEA-Technology 2012-036, (2012) (in Japanese)
7. C.K. Wikle, R.F. Milliff, D. Nychka, L.M. Berliner, Spatiotemporal hierarchical Bayesian modeling: tropical ocean surface winds. *J. Am. Stat. Assoc.* **96**(454), 382–397 (2001)
8. H.M. Wainwright, D. Sassen, J. Chen, S.S. Hubbard, Bayesian hierarchical approach for estimation of reactive facies over plume-scales using geophysical datasets. *Water Resour. Res.* (2014)
9. Y. Zhou, A.M. Michalak, Characterizing attribute distributions in water sediments by geostatistical downscaling. *Environ. Sci. Technol.* **43**(24), 9267–9273 (2009)
10. P. Diggle, P.J. Ribeiro, *Model-Based Geostatistics*. Springer Science & Business Media (2007)
11. H.M. Wainwright, J. Chen, A. Seki, K. Saito, A multiscale Bayesian data integration approach for mapping radionuclide contamination in the regional scale. *J Environ Radioact.* **167**, 62–69 (2017a)
12. H.M. Wainwright, A. Seki, J. Chen, K. Saito, A multiscale Bayesian data integration approach for mapping air dose rates around the Fukushima Daiichi Nuclear Power Plant. Proceeding of WM2017 Conference, March 5–9, 2017, Phoenix, Arisona, USA (2017b)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

