Title

A multiscale Bayesian data integration approach for mapping air dose rates around the Fukushima Daiichi Nuclear Power Plant

Authors

Wainwright, Haruko M

Seki, Akiyuki

Chen, Jinsong

et al.

Publication Date

2017-02-01

DOI

10.1016/j.jenvrad.2016.11.033

Peer reviewed

Journal of Environmental Radioactivity xxx (2016) 1–8

Contents lists available at ScienceDirect

Journal of Environmental Radioactivity

journal homepage: www.elsevier.com/locate/jenvrad

A multiscale Bayesian data integration approach for mapping air dose rates around the Fukushima Daiichi Nuclear Power Plant

Haruko M. Wainwright [a, *], Akiyuki Seki [b], Jinsong Chen [a], Kimiaki Saito [c]

[a] Earth Sciences Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, MS 74R-316C, Berkeley, CA 94720-8126, USA
[b] Japan Atomic Energy Agency, Center for Computational Science & E-system, 178-4-4 Wakashiba, Kashiwa, Chiba, 227-0871, Japan
[c] Japan Atomic Energy Agency, Fukushima Environmental Safety Center, 2-2-2 Uchisawai-cho, Chiyoda, Tokyo, 100-0011, Japan

ARTICLE INFO

Article history:
Received 24 May 2016
Received in revised form
27 November 2016
Accepted 28 November 2016
Available online xxx

Keywords:
Mapping of air dose rates
Bayesian hierarchical models
Fukushima Daiichi Nuclear Power Plant accident
Geostatistics
Multi-scale
Data integration

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, multi-type 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 in 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 and create an integrated map (including the confidence intervals) of air dose rates over the domain in high resolution. Moreover, this study provides us with various insights into the characteristics of each dataset, as well as radiocaesium distribution. In particular, the urban areas show high heterogeneity in the contaminant distribution due to human activities as well as large discrepancy among different surveys due to such heterogeneity.

1. Introduction

The accident at the Fukushima Daiichi Nuclear Power Plant (NPP) after the Great East Japan Earthquake resulted in the release of radioactive contaminants to the atmosphere and environment in March 2011. The radioactive contaminants were subsequently deposited on soil and plants via wet and dry precipitation (Tanaka, 2012). The soils in Fukushima and neighboring prefectures are contaminated with deposition of multiple radionuclides, such as ^{131}I, ^{129m}Te, ^{110m}Ag, ^{134}Cs and ^{137}Cs. The radiocaesium contamination is currently considered to be the most serious long-term health hazard due to its activity released and half-life.

Measurements and monitoring of air dose rates (i.e., ambient dose equivalent rates) in the region around the Fukushima Daiichi NPP have been performed continuously since the accident (e.g., Saito and Onda, 2015; Mikami et al., 2015). The measurements have been conducted using various techniques such as walk surveys using portable monitoring systems, car surveys, and airborne surveys. Soil samples also have been collected to assess the extent of contamination in the terrestrial environment (Saito et al., 2015). Such mapping efforts are essential to protect the public, to guide decontamination efforts, and to plan the return of evacuated residents.

In addition to traditional fixed-location monitoring posts or handheld monitors, advanced measurement techniques have been developed, and are currently used routinely in the area. Among them, airborne monitoring or surveys have been extensively used to map the air dose rates in the regional scale (e.g., 100 km radius) with resolution of several hundred meters. In the airborne surveys, a radiation detector is mounted on helicopters, which then fly over the target area (Torii et al., 2012). Calibration methods have been

* Corresponding author.
E-mail addresses: hmwainwright@lbl.gov (H.M. Wainwright), seki.akiyuki@jaea.go.jp (A. Seki), jchen@lbl.gov (J. Chen), saito.kimiaki@jaea.go.jp (K. Saito).

http://dx.doi.org/10.1016/j.jenvrad.2016.11.033
0265-931X/

developed to compute air dose rates at 1 m above the ground surface by considering attenuation in the air and various factors. Airborne data have provided vital information to identify the extent of large-scale contaminant distribution as well as to make policy decisions for radiation protection purposes. In addition, a GPS-aided mobile radiation monitoring system mounted on a car or motorcycle—the Kyoto University RAdiation MApping system (KURAMA, Tanigaki et al., 2015) —has been used extensively to characterize the distribution of air dose rates along roads in real time (Andoh et al., 2015). The KURAMA system has a significant advantage for real-time monitoring, with the help of online data streaming and cloud storage system (Tsuda et al., 2015). The KURAMA system has been rigorously tested and verified to be acceptable for mass monitoring of air dose rates (Tsuda et al., 2015).

With many data survey types available, it has become clear that there are discrepancies among them in terms of measured air dose rates, even collected at around the same time and same locations. This is mainly because 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 air dose rates over a much larger area (typically a radius of several hundred meters) than ground-based measurements (~several tens of meters). Such averaging becomes particularly problematic, because soil contamination and air dose rates are both highly heterogeneous, having many hotspots (JAEA, 2012). Physics-based radiation transport modeling by Malins et al. (2016) also showed that uncertainty in the above-ground air dose rates is associated with the horizontal heterogeneity of radio-caesium 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-density data points along the roads (Tsuda et al., 2015). 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. The airborne survey has the large spatial coverage at the regional scale, although the resolution is low due to the averaging, missing many hotspots and detailed heterogeneity.

To reconcile such discrepancies, 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 and uncertainty of each type of measurement. In environmental science, monitoring and spatiotemporal mapping of various properties—such as CO_2 concentration, wind velocity or reactive transport properties in subsurface—have been the focus of extensive research in the past decades. Although many traditional datasets are point-scale and sparse in time and space, recent advanced datasets can cover large areas, such as remote sensing in atmospheric/terrestrial sciences and geophysical techniques in subsurface sciences. 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 (e.g., Wikle et al., 2001; Zhou and Michalak, 2009; Wainwright et al., 2014, 2015).

Particularly, the Bayesian hierarchical approach has been proposed as a flexible and expandable framework to integrate multiscale datasets (Wikle et al., 2001; Wainwright et al., 2014, 2015). A Bayesian hierarchical model typically 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 air dose rates within the domain, representing the spatial trend and heterogeneity of contamination. The data models connect this pattern with actual data, given measurement errors. These data models can represent, for example, a direct ground-based measurement or a function of the pattern such as spatial averaging over a certain area for a low-resolution airborne dataset. The overall model— a series of statistical sub-models—is flexible and expandable so that it can include complex correlations or various types of observations. Once all the sub-models are developed, we can estimate the map of air dose rates and its confidence interval, using sampling-based or optimization-based methods. One of the main advantages is that this method can quantify the possible estimation errors and provide confidence intervals of the estimated air dose rates at any given location.

In addition, geostatistics have been developed to characterize the spatial heterogeneity of environmental properties and to interpolate those properties based on sparse measurements (Deutsch and Journel, 1998; Diggle and Ribeiro, 2007). Geostatistics are based on the spatial autocorrelation that determines the heterogeneity structure based on available datasets; i.e., how the property value changes over space. Diggle et al. (1998) has applied a model-based geostatistical approach to characterize the radionuclide concentrations (^{137}Cs) and to estimate the environmental decay rate on the Rongelap Island. In addition, geostatistical models are often used as process models within Bayesian approaches or Bayesian hierarchical models to integrate multiscale multi-type datasets (e.g., geophysics and core data, or remote sensing and ground-based measurements), and to estimate spatially heterogeneous properties (e.g., Chen et al., 2001, 2006; Sassen et al., 2012; Wainwright et al., 2014, 2015).

In this study, we develop a Bayesian geostatistical approach to integrate multiscale datasets (i.e., car, walk and airborne surveys) and to estimate the spatial distribution of air dose rates at 1 m above the ground surface in high resolution across the regional scale. Since we consider that the walk surveys represent the exposure dose of average individuals walking on the streets, we estimate the distribution of air dose rates equivalent to the walk survey data. In addition, we aim to gain a significant insight into the characteristics of each survey, as well as the spatial heterogeneity and trend of air dose rates. We demonstrate our approach using the datasets collected in Fukushima Prefecture, Japan, in November 2013 by Japan Atomic Energy Agency (JAEA).

2. Data description

In this paper, we integrated three types of datasets of air dose rates collected in Fukushima prefecture of Japan, including walk, car and airborne surveys. Although JAEA accumulated a vast amount of datasets of air dose rates since the accident, we used a subset of datasets to demonstrate our approach. Fig. 1 shows the domain of interest in this study, which is the northwestern region of the Fukushima Daiichi NPP approximately within the 80-km radius. We used the datasets collected around the same time in November 2013. We assumed that the effect of radiocaesium decay was negligible among these three surveys. All the datasets were publically available and downloaded from the data management system developed by JAEA (Seki et al., 2016).

We used the processed and converted airborne data equivalent to air dose rates at 1 m above the ground surface. The detailed procedures of data acquisition and calibration can be found in Torii et al. (2012). The actual flight altitude was approximately 300 m above the ground. The publicly available airborne data are interpolated on the 250 m-resolution grid. As shown in Fig. 1a, the airborne data capture the variability of air dose rates at the regional scale, and show the highly contaminated area extending in the

Please cite this article in press as: Wainwright, H.M., et al., A multiscale Bayesian data integration approach for mapping air dose rates around the Fukushima Daiichi Nuclear Power Plant, Journal of Environmental Radioactivity (2016), http://dx.doi.org/10.1016/j.jenvrad.2016.11.033

northwest direction from the NPP location, which was the prevailing wind direction at the time of the accident. The airborne data are weighted average over the area around each data point, since they represent the total gamma radiation emitted from a large area on the ground. Torii et al. (2012) estimated that the zone of influence would be within a circle in the same radius as the flight latitude. Alternatively, the radiation transport simulation described by Malins et al. (2015, 2016) can be used to estimate the relative contribution of gamma ray emitted from the ground to the air dose rate measured by a detector at any given height as a function of source radius. In the simulation, it was assumed that Cs-137 in the ground distributes exponentially with a relaxation mass depth of 1 g/cm^2. The modeling results (Fig. 2) show, for example, that, at 5 m above the ground, the 190-m radius represents 95% of total gamma ray counts (assuming Cs-137). At 300 m above the ground, the 95-percent radius increases to 547 m.

The car survey data were collected using the KURAMA system mounted on a car. Tsuda et al. (2015) described this system, including GPS and data streaming to a cloud data storage in detail. Air dose rates (i.e., ambient dose equivalent rates) and coordinates were measured every several seconds, while the car was moving. As shown in Fig. 1b, the data points are distributed along major roads with high sampling density. Since the car survey detects gamma ray emitted along roads on a moving platform, the measured air dose rate is considered to be an average of an extended area. JAEA's data management system stores the data averaged in each 100-m mesh grid (Seki et al., 2016), mainly to obtain sufficient statistical accuracy in the results.

The walk survey data were collected also using the KURAMA system. In the walk survey, a person carries the system and walks around each neighborhood so that the walk survey data tend to be distributed in clusters (Fig. 1b). The walk survey data were collected from mostly publicly accessible areas, such as pedestrian paths along roads. The walk survey data were processed and averaged in each 20-m mesh grid in the database (Seki et al., 2016).

In addition, we used the high-resolution land-use and land-cover map of Japan (version 14.02) created by Japan Aerospace Exploration Agency (downloaded from http://www.eorc.jaxa.jp/ALOS/lulc/lulc_jindex.htm; in Japanese). The map is based on the 50-m resolution multispectral data from the Advanced Visible and Near Infrared Radiometer type 2 (AVNIR-2) on the Advanced Land Observation Satellite (Takahashi et al., 2013). The land is classified into nine types: water, urban, paddy, crop, grass, deciduous forest, evergreen forest, bare land, and snow and ice. In this study, due to the data availability, we re-grouped these nine types into three groups: urban, crop (paddy, crop, grass) and forest (deciduous forest, evergreen forest, bare land). To reduce the noise, we smoothed the landscape type by selecting the dominant land cover type within 300 m of each pixel.

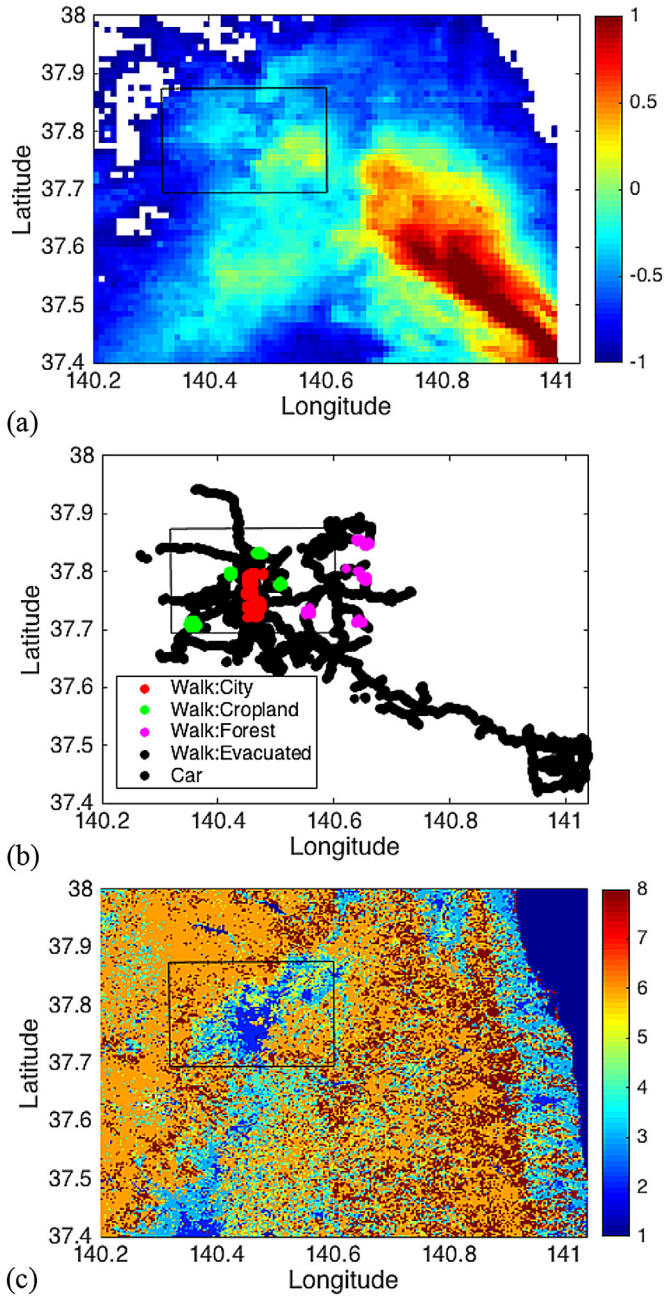

Fig. 1. (a) Air dose rate map from an airborne survey (in log10, μSv/h), (b) car and walk survey locations used in this study, and (c) the land use map based on the AVNIR-2 satellite. The black rectangles are the vicinity of Fukushima City used for the estimation. In (b), the walk survey locations were categorized into different land use types.

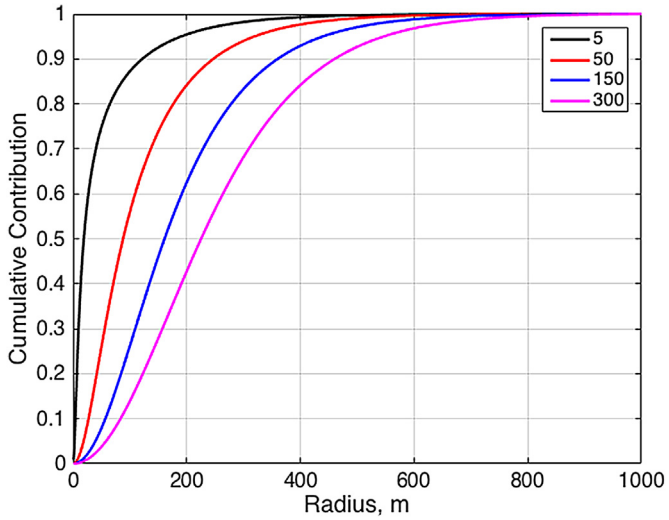

Fig. 2. Cumulative contribution of gamma radiation from deposited radiocaesium sources to the air dose rate as a function of source radius at the different altitude of a detector: 5, 50, 150 and 300 m above the ground.

Please cite this article in press as: Wainwright, H.M., et al., A multiscale Bayesian data integration approach for mapping air dose rates around the Fukushima Daiichi Nuclear Power Plant, Journal of Environmental Radioactivity (2016), http://dx.doi.org/10.1016/j.jenvrad.2016.11.033

3. Methodology

3.1. Exploratory data analysis

Before building a statistical model for the Bayesian estimation, we performed an exploratory data analysis (EDA) to understand the characteristics of these datasets and to quantify the spatial heterogeneity of air dose rates. We first identified the correlations among the walk, car and airborne survey data within or near Fukushima City. We investigated the dependency of such correlations on land-cover types (i.e., urban, crop, forest), since several studies have reported the effect of land-cover types on the air dose rates in the environment (Kinase et al., 2014; Mikami et al., 2015). In addition, we added the datasets from the high-dose and evacuated area (Futaba Machi). The main focus was to establish the correlations between the low-resolution and high-uncertainty datasets (i.e., car and airborne survey) and the high-resolution walk survey data. The Pearson product-moment correlation coefficient was computed for each pair of datasets.

Such a correlation analysis is typically done by extracting the data points of two or more types that are located very close to each other (i.e., co-located data points). To represent different surveys in our study, however, we need to consider a different zone of influence (or spatial averaging) depending on the survey type. We tested several averaging methods to represent the low-resolution car and airborne survey data: (1) simple averaging within the zone of influence, (2) inverse-distance weighted (IDW) averaging within the zone of influence, and (3) weighted averaging with the weights computed from radiation transport modeling (Malins et al., 2015, 2016). The zone of influence for simple and IDW averaging is the circle with the radius of 100 m for the car survey data, and the one of the 300 m for the airborne survey data, respectively.

In addition, we performed a geostatistical analysis to investigate the spatial variability and spatial correlation structure of air dose rates in each land-cover type. The underlying assumption in geostatistics is that the values at two points are more likely to be similar when they are closer (e.g., Deutsch and Journel, 1998; Diggle and Ribeiro, 2007; Dwivedi et al., 2016). The measure of similarity is defined by the Pearson's correlation coefficient as a function of distance; the spatial correlation decays to zero as the distance increases. Alternatively, we may define the measure of difference, referred to as *variogram*, since the two locations are more likely to have different values when they are farther apart. In addition, even when the two points are at the same location, the values may be uncertain due to measurement errors or small-scale variability. This spatially uncorrelated portion of variability is called *nugget*. The geostatistical (or variogram) analysis intends to estimate (1) magnitude of variability (or spatial heterogeneity) as variance, (2) fraction of spatially uncorrelated variability within the total variability as nugget ratio (i.e., the ratio between the total variance and nugget variance), (3) spatial correlation range (i.e., the range at which the spatial correlation becomes zero or approximately zero), and (4) variogram model (i.e., the function to describe the decay of spatial correlations) such as exponential and spherical models. We used the standard variogram approach in the GeoR package of the statistical software R (Ribeiro and Diggle, 2001). The more detailed descriptions and mathematical formulation in geostatistics can be found in Diggle and Ribeiro (2007).

3.2. Data integration and estimation

To develop an integrated map, we denote the air dose rate at i-th pixel by y_i, where $i = 1, \ldots, n$. Similarly, we denote the 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 probability distribution of the air dose rate map $\boldsymbol{y}$ (i.e., the vector representing the air dose rates at all the pixels) conditioned on the three datasets ($\boldsymbol{z}_A$, $\boldsymbol{z}_C$ and $\boldsymbol{z}_W$), written as $p(\boldsymbol{y} \mid \boldsymbol{z}_A, \boldsymbol{z}_C, \boldsymbol{z}_W)$. By applying Bayes' rule, we can rewrite this posterior distribution as:

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

where "$\propto$" denotes "is proportional to". We assume that the datasets are conditionally independent of each other, given the distribution of air dose rates. The first two conditional distributions $p(\boldsymbol{z}_A \mid \boldsymbol{y})$ and $p(\boldsymbol{z}_C \mid \boldsymbol{y})$ represent the data models to connect the distribution of the air dose rates ($\boldsymbol{y}$) with the low-resolution data (i.e., airborne or car survey data). The last distribution $p(\boldsymbol{y} \mid \boldsymbol{z}_W)$ represents the process model to describe the spatial pattern given the measured air 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 (Section 3.1), and hence they are considered as known parameters during this Bayesian estimation. Each sub-model (i.e., conditional distribution) in Equation (1) is described in the following.

The first conditional distribution $p(\boldsymbol{z}_A \mid \boldsymbol{y})$ in Equation (1) is the data model representing the airborne data as a function of the air dose rate map $\boldsymbol{y}$. We assume a spatial weighted averaging function:

$$z_{A,j} = f_A\left(\sum w_{A,i,j} y_i\right) + \varepsilon_{A,j}, \quad (2)$$

where $w_{A,i,j}$ is the weight determined by the distance between i-th pixel and j-th airborne data point, $\varepsilon_{A,j}$ is the error associated with each airborne data point, and $f_A(\bullet)$ is a function representing the correlation between the airborne data and air dose rate distribution (such as a linear function from the linear correlation analysis). We computed the weight using the Monte Carlo radiation transport code developed by Malins et al. (2015, 2016) according to the flight altitude of 300 m. In this study, we assume a linear model such that $\boldsymbol{z}_A = A\boldsymbol{y} + \boldsymbol{\varepsilon}_A$, where the m_A-by-n matrix A includes the parameters and also the averaging weights, and $\boldsymbol{\varepsilon}_A$ is the error vector for the airborne data; $\boldsymbol{\varepsilon}_A = \{\varepsilon_{A,j},\ j = 1, \ldots, m_A\}$. We assume that each component $\varepsilon_{A,j}$ follows an independent normal distribution with zero-mean, and that the error variance σ_A can be determined from the correlation analysis between two datasets (Section 3.1). Consequently, we can define the m_A-by-m_A data-error covariance matrix D_A; each of the diagonal components is σ_A. The airborne data vector follows a multivariate Gaussian distribution with the mean $A\boldsymbol{y}$ and the covariance matrix D_A, represented by $\boldsymbol{z}_A \sim MVN(A\boldsymbol{y}, D_A)$.

Similarly, we can define the car survey data as an averaged function of the air dose rate map $\boldsymbol{y}$ as:

$$z_{C,j} = f_C\left(\sum w_{C,i,j} y_i\right) + \varepsilon_{C,j}. \quad (3)$$

where $w_{C,i,j}$ is the weight determined by the distance between i-th pixel and j-th data point, $\varepsilon_{C,j}$ is an error associated with each data point, and $f_C(\bullet)$ is a function representing the correlation between the car survey data and air dose rate distribution. Again, we assume a linear model $\boldsymbol{z}_C = C\boldsymbol{y} + \boldsymbol{\varepsilon}_C$, where the m_C-by-n matrix C includes the correlation parameters and weights. We also define the m_C-by-m_C data-error covariance matrix D_C, with the diagonal components of error variance σ_C. The car survey data vector thus follows a multivariate Gaussian distribution $\boldsymbol{z}_C \sim MVN(C\boldsymbol{y}, D_C)$.

As a process model, we assume that $\boldsymbol{y}$ is a multivariate Gaussian field described by geostatistical parameters. Since the walk survey

Please cite this article in press as: Wainwright, H.M., et al., A multiscale Bayesian data integration approach for mapping air dose rates around the Fukushima Daiichi Nuclear Power Plant, Journal of Environmental Radioactivity (2016), http://dx.doi.org/10.1016/j.jenvrad.2016.11.033

data represent the target air dose rates to be estimated, we use the walk survey data as conditional points to constrain the distribution of $\boldsymbol{y}$ as $p(\boldsymbol{y}|\boldsymbol{z}_W)$. If a walk survey data value is available at a pixel, we assume that the air dose rate at the location is known. We determine the geostatistical parameters from EDA described in Section 3.1. The conditional distribution is the multivariate distribution with the mean μ_c and covariance Σ_c based on the geostatistical parameters and conditioning points $\boldsymbol{z}_W$: $\boldsymbol{y} \sim MVN(\mu_c, \Sigma_c)$ (Diggle and Ribeiro, 2007).

Since three conditional distributions in Equation (1) are multivariate Gaussian, we can derive an analytical form of this posterior distribution as a multivariate normal distribution with mean $Q^{-1}\boldsymbol{g}$ and covariance Q^{-1}, where

$$Q = \Sigma_c^{-1} + A^T D_A^{-1} A + C^T D_C^{-1} C \quad (4)$$

$$\boldsymbol{g} = \mu_c + A^T D_A^{-1} z_A + C^T D_C^{-1} z_C. \quad (5)$$

Each term in Equations (4) and (5) is already defined above. We may directly sample $\boldsymbol{y}$ from $\boldsymbol{y} \sim MVN(Q^{-1}\boldsymbol{g}, Q^{-1})$ for the Monte Carlo simulations, or estimate the mean (or expected) air dose rate map by $Q^{-1}\boldsymbol{g}$. The estimation variance at each pixel is the diagonal component of Q^{-1}, which represents the uncertainty in the estimation results. We implemented the algorithm in a FORTRAN code, which is released with an open-source license.

4. Results and discussion

4.1. Exploratory data analysis

Fig. 3 shows the comparison between the walk and car survey data. Finding co-located points using the minimum distance (blue circles) causes large scatter, especially in the urban area (Fig. 3a). Taking a simple average (red circles) reduces the scatter and improves the correlation, particularly in the urban area. In all the land-cover types, the correlation was statistically significant with p-values less than 10^{-3}. In Table 1, the correlation coefficients are computed based on the different averaging methods. The simple average provides the best correlation coefficient. The effect of averaging is most significant in the urban area, possibly because human activities and complex terrain created large heterogeneity in air dose rates (equivalent to the walk survey data). In addition, the walk survey data included data from small side streets, which resulted in the discrepancy to the car survey data collected on main streets. On the other hand, there was little improvement by averaging in the forest, possibly because the heterogeneity is smaller and the walk and car survey points tended to be along the same roads.

The car survey data tend to underestimate the walk survey data at many data points in Fig. 3. This is possibly because radiocaesium tends to be removed and re-distributed by human and car activities in the middle of the roads (Andersson et al., 2002). This is also true in the cropland (Fig. 3b) and forest (Fig. 3c) near Fukushima City. In the evacuated area (Fig. 3d), however, this is not the case, and the correlation coefficient is significantly higher (0.99). This is probably because there is no human activity to re-distribute dirt on the streets in the evacuated area.

Similarly, the airborne and walk survey data are compared in Fig. 4 and Table 2. Since there are no airborne data available from the evacuated area (due to the immediate vicinity of the power plant), we compared the data only in the three land-cover types. The correlation is statistically significant in all the land-cover types (p-values are less than 10^{-3}). Fig. 4 also shows that the airborne survey data tend to overestimate the air dose rates, possibly because the calibration was done only at several locations with an undisturbed flat surface having relatively higher air dose rates.

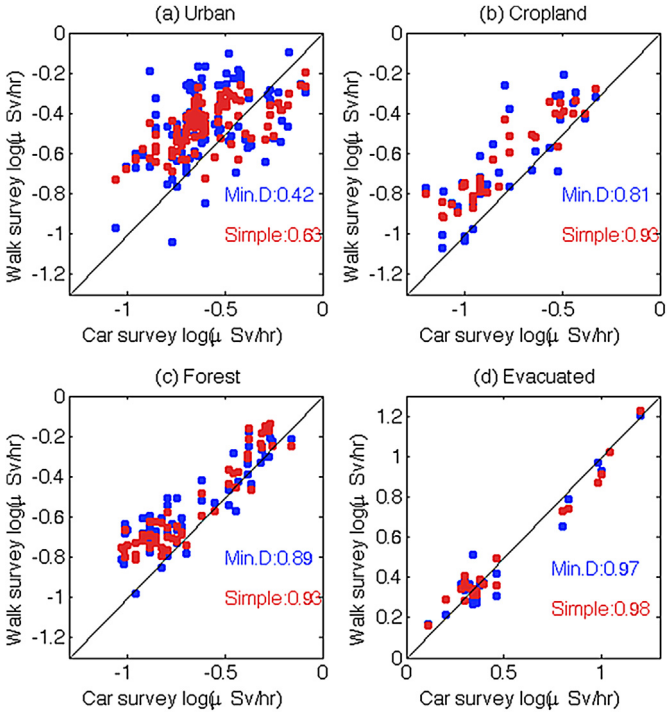

Fig. 3. Comparison between the car and walk survey data: (a) urban, (b) cropland, (c) forest and (d) evacuated areas. The blue circles are the co-located points identified by the minimum distance. The red circles are the average of the walk survey points using the simple average. In each plot, the correlation coefficients are shown.

Table 1
Correlation coefficients between the car and walk survey data. In addition to finding co-located points by the minimum distance (Minimum), we compared several averaging methods: (1) simple averaging within the zone of influence (Simple), (2) inverse-distance weighted (IDW) averaging within the zone of influence, and (3) weighted averaging with the weights computed from radiation transport modeling (Malins).

	Urban	Cropland	Forest	Evacuated
Minimum	0.42	0.81	0.89	0.97
Simple	0.63	0.93	0.93	0.98
IDW	0.62	0.91	0.93	0.99
Malins	0.56	0.88	0.92	0.98

Similar to Fig. 3, selecting the co-located data points by the minimum distance (blue circles) shows larger variability, 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 (red circles), the correlation improves significantly in all land-cover types, particularly in the urban area (Fig. 4a). Such a large improvement (similar to the car/walk comparison in Fig. 3) in the urban area suggests that the air dose rate distribution in the urban area is much more variable and heterogeneous within several hundred meters than the cropland or forest.

Fig. 5 shows the variogram analysis result in each land-cover type near Fukushima City. Among the three land-cover types, the urban area has higher variability than other land cover types, which is consistent with the observations in Figs. 3 and 4. The urban area also has the highest nugget variance (i.e., the variance at distance 0), suggesting the large variability within a small scale, and the presence of hot or cold spots. This is consistent with the fact that

Please cite this article in press as: Wainwright, H.M., et al., A multiscale Bayesian data integration approach for mapping air dose rates around the Fukushima Daiichi Nuclear Power Plant, Journal of Environmental Radioactivity (2016), http://dx.doi.org/10.1016/j.jenvrad.2016.11.033

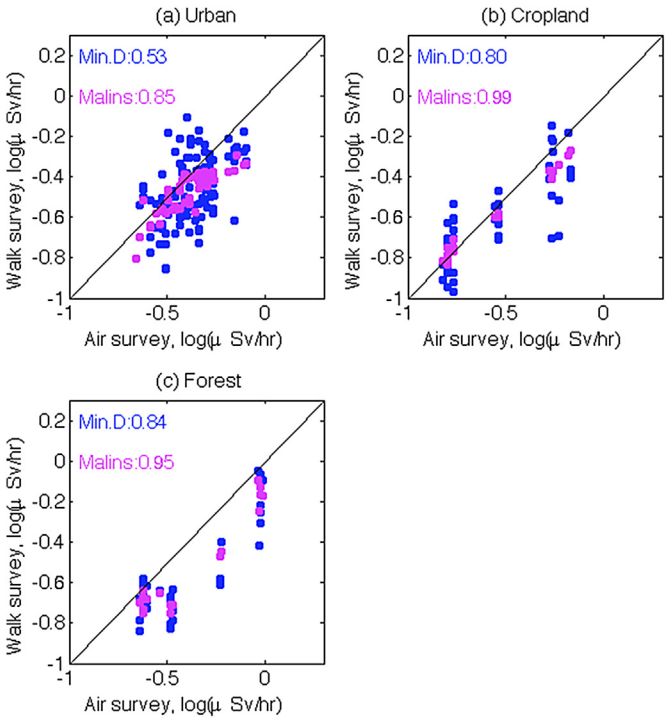

Fig. 4. Comparison between the air and walk survey data in: (a) urban, (b) cropland, and (c) forest. The blue circles are the co-located points identified by the minimum distance. The magenta circles are the average of the walk survey points using the weights computed by the radiation transport simulation.

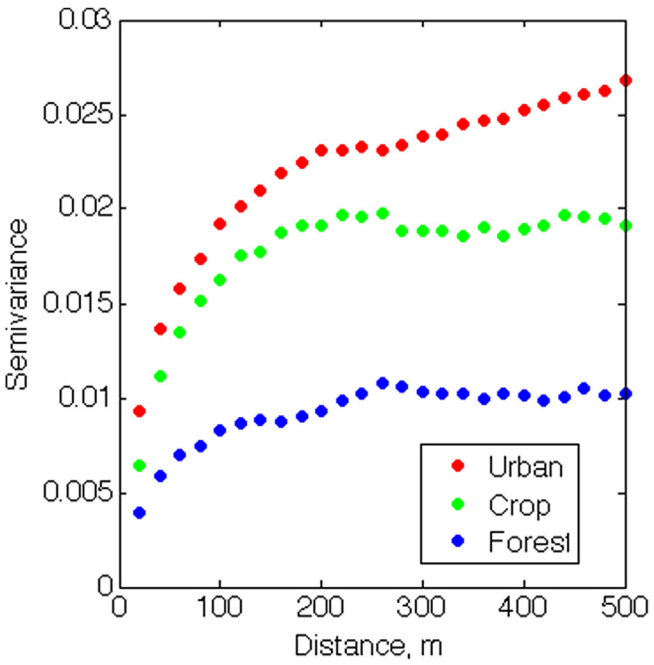

Fig. 5. Empirical variograms for three land-use types.

Table 2
Correlation coefficients between the airborne and walk survey data. In addition to finding co-located points by the minimum distance (Minimum), we compared several averaging methods: (1) simple averaging within the zone of influence (Simple), (2) inverse-distance weighted (IDW) averaging within the zone of influence, and (3) weighted averaging with the weights computed from radiation transport modeling (Malins).

	Urban	Cropland	Forest
Minimum	0.53	0.80	0.84
Simple	0.84	0.99	0.95
IDW	0.81	0.98	0.93
Malins	0.85	0.99	0.95

the urban area has more complex microtopography (trenches, buildings, sidewalks) as well as more human activities that drive radionuclide re-distribution. The forest, on the other hand, has the small variability and long range, suggesting that the air dose rates are less variable, and also change more smoothly over space.

4.2. Data integration and estimation

In this study, we demonstrated the data integration and estimation in the vicinity of Fukushima city (Fig. 6a) to create the integrated map of air dose rates in high resolution (50 m by 50 m) over the domain of 25 km by 20 km (i.e., 500 by 400 pixels). The domain includes the major land cover types, such as urban, cropland and forest (Fig. 6b). In the airborne survey data (Fig. 6a), the south-eastern portion of the domain has higher air dose rates, possibly because the area lies along the initial plume direction and is also a forested area where the temporal reduction of air dose rates is slow. By overlaying the car, walk and air survey data (Fig. 6a), we see that the car and walk survey data show smaller-scale variability than the airborne data, and that the airborne data overestimates the air dose rates.

The estimated map (mean field) from the data integration in Fig. 7a shows more detailed and finer-resolution heterogeneity than the original airborne data (Fig. 6a), although the general trend is very similar due to the large spatial coverage of the airborne data. The systematic bias (or shift) in the airborne data was also corrected. There is a slight shift at the boundaries of the different land-cover types, since the correlation between the walk and airborne survey data is different in each land-cover type. As shown in Fig. 6b, the standard deviation is smaller in the vicinity the car and walk survey data points, where the estimation is constrained by those data and the spatial correlation. The standard deviation is larger in the urban area in Fukushima City, representing the larger uncertainty and variability of air dose rates in the urban area, which is consistent with the data analysis results.

Fig. 8 shows the validation result to evaluate the performance of the data integration and the estimation. One hundred points of the walk survey data (randomly selected) are excluded from the estimation, and used for the 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 lines and are mostly included in the 95% confidence interval. Particularly, the significant bias in the airborne data is corrected in the forested area (Fig. 8c). Based on these results, we may argue that this method has successfully created the fine-resolution integrated map of air dose rates based on the spatially sparse walk and car survey data, and spatially extensive airborne survey data.

5. Conclusion

In this paper, we developed a Bayesian geostatistical method for integrating multiscale, multi-type measurements of air dose rates,

Please cite this article in press as: Wainwright, H.M., et al., A multiscale Bayesian data integration approach for mapping air dose rates around the Fukushima Daiichi Nuclear Power Plant, Journal of Environmental Radioactivity (2016), http://dx.doi.org/10.1016/j.jenvrad.2016.11.033

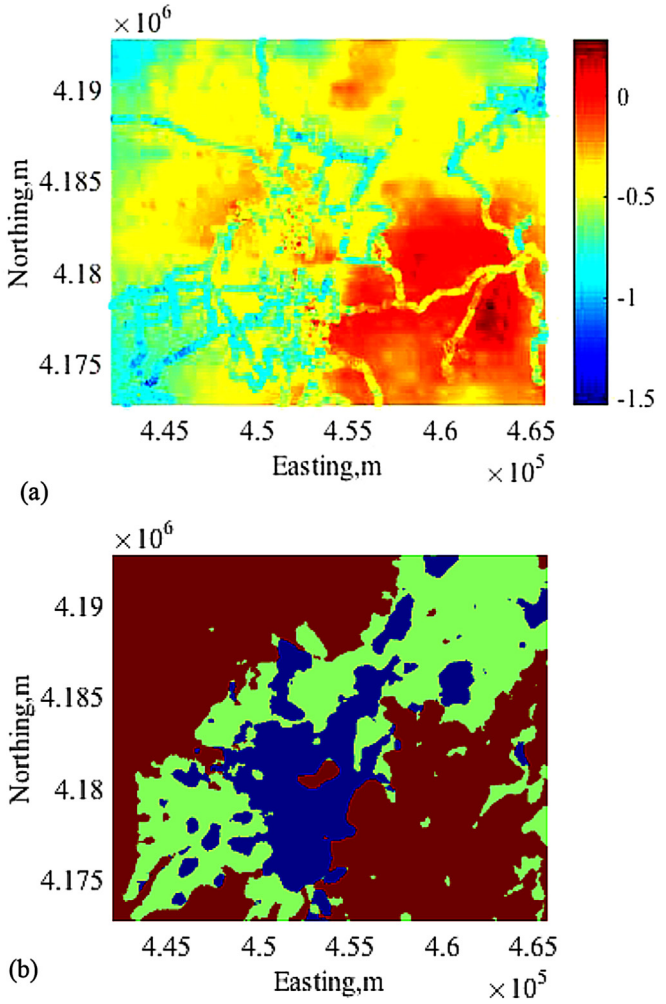

Fig. 6. (a) Walk, car and airborne survey data plotted together over Fukushima City in log10(μSv/h), and (b) the land-cover type (blue is urban, green is cropland and red is forest). In (a), the walk survey and car survey locations can be seen as the lines of blue to orange, since the values are lower than the airborne data.

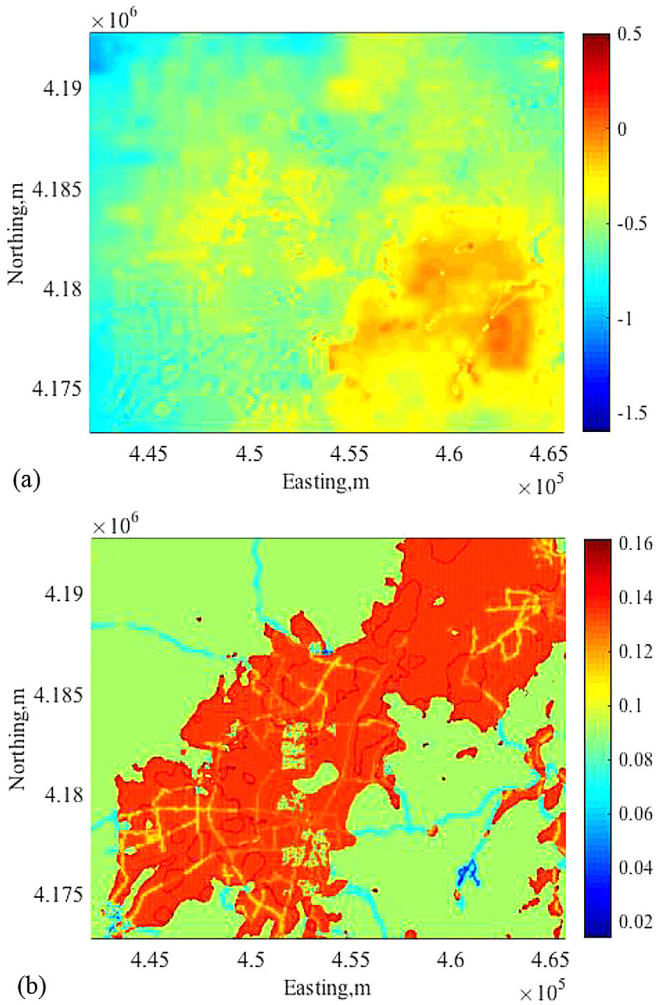

Fig. 7. (a) Integrated dose-rate map (or mean field) based on the developed data integration method, and (b) the estimated standard deviation. In all the plots, the data values are log-transformed.

and 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 method was also able to provide the confidence intervals of the estimated map, representing the uncertainty associated with the data and spatial heterogeneity. Although the current demonstration results were limited in a particular area and time, the results shown here suggested that the effective combination of ground-based data and airborne data could provide detailed and integrated maps of air dose rates in the regional scale around the Fukushima Daiichi NPP. In addition, the statistical analyses provided various insights into both the characteristics of each dataset and the spatial heterogeneity of contamination.

The exploratory data analysis showed that the car and airborne survey data could be represented by the spatial average of more detailed air dose rate distribution from the walk survey data. The spatial average had the highest impact in the urban area, where the co-located air dose rate had the biggest discrepancy between the walk and air surveys as well as between the walk and car surveys. This would be because the urban area had a large heterogeneity in air dose rates and radiocaesium distribution induced by human activities and complex terrains. In the forest and cropland area, on the other hand, the car and airborne survey data were more consistent with the walk survey data.

The geostatistical analysis also revealed that the spatial heterogeneity of air dose rates depended on different land-cover types. The air dose rates were more variable in the urban area than in the cropland and forest areas. More detailed surveys would be required in the urban area to accurately characterize the contamination, while sparse measurements would be sufficient for interpolation in the cropland and forested areas.

The developed estimation method was very powerful to take into account such data correlations and spatial characteristics found in the data analysis for integrating multiscale data and for estimating the distribution of air dose rates over a large area. In addition, we used the weighted averaging factors derived from physics-based radiation transport models to accurately represent the airborne data. The estimated map not only captured more detailed spatial heterogeneity than the original airborne survey data, but also reflected the physical understanding of the radiocaesium distribution found in the datasets; for example, the urban area had higher uncertainty and larger confidence intervals, particularly in the areas without walk survey datasets. Having the confidence intervals would help estimate the range of exposure dose or plan the return to the evacuation zones in a more robust manner.

Please cite this article in press as: Wainwright, H.M., et al., A multiscale Bayesian data integration approach for mapping air dose rates around the Fukushima Daiichi Nuclear Power Plant, Journal of Environmental Radioactivity (2016), http://dx.doi.org/10.1016/j.jenvrad.2016.11.033

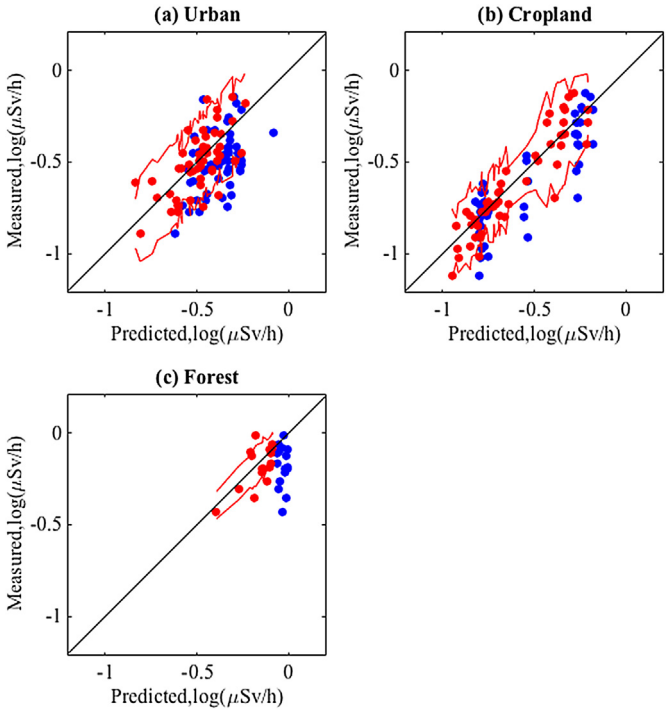

Fig. 8. Comparison between the predicted and measured air dose rates (log-transformed) at the walk-survey data locations not used for the estimation. The red dots represent the predicted values based on the data integration; the blue dots are the co-located airborne data before the integration. The black line is the one-to-one line; the red lines are the 95% confidence intervals.

In future work, we plan to improve the estimation approach, for example, by including other information such as topography, and by including the correlation between the air dose rate and actual caesium concentrations in soil. In addition, we will carry out spatiotemporal integration—by integrating spatially sparse but continuous-time monitoring data, and temporally sparse but spatially extensive data, such as airborne data—to provide a detailed map of air dose rates and radionuclide contamination at regional scale at any given location and time, including the associated confidence intervals.

Acknowledgement

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.

References

Andersson, K.G., Roed, J., Fogh, C.L., 2002. Weathering of radiocaesium contamination on urban streets, walls and roofs. J. Environ. Radioact. 62 (1), 49–60.

Andoh, M., Nakahara, Y., Tsuda, S., Yoshida, T., Matsuda, N., Takahashi, F., Mikami, S., Kinouchi, N., Sato, T., Tanigaki, M., Takamiya, K., 2015. Measurement of air dose rates over a wide area around the Fukushima Dai-ichi Nuclear Power Plant through a series of car-borne surveys. J. Environ. Radioact. 139, 266–280.

Chen, J., Hubbard, S.S., Rubin, Y., 2001. Estimating hydraulic conductivity at the South Oyster Site from geophysical tomographic data using Bayesian techniques based on the normal linear regression model. Water Resour. Res. 37, 1603–1613.

Chen, J., Hubbard, S.S., Peterson, J.E., Williams, K., Fienen, M., Jardine, P., Watson, D., 2006. Development of a joint hydrogeophysical inversion approach and application to a contaminated fractured aquifer. Water Resour. Res. 42 (6).

Deutsch, C.V., Journel, A.G., 1998. GSLib. Geostatistical Software Library and User's Guide, p. 369.

Diggle, P., Ribeiro, P.J., 2007. Model-based Geostatistics. Springer Science & Business Media.

Diggle, P.J., Tawn, J.A., Moyeed, R.A., 1998. Model-based geostatistics. J. R. Stat. Soc. Ser. C Appl. Statistics) 47 (3), 299–350.

Dwivedi, D., Dafflon, B., Arora, B., Wainwright, H.M., Finsterle, S., November 1 2016. Spatial analysis and geostatistical techniques. In: Singh, V.P. (Ed.), Handbook of Applied Hydrology, second ed. McGraw-Hill Education. ISBN-10 0071835091; ISBN-13: 978–0071835091, 1440 pages. https://www.amazon.com/Handbook-Applied-Hydrology-Second-Vijay/dp/0071835091/.

Japan Atomic Energy Agency, 2012. 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).

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. Radiat. Prot. Dosim. 160 (4), 318–321.

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).

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. J. Environ. Radioact. 151, 38–49.

Mikami, S., Maeyama, T., Hoshide, Y., Sakamoto, R., Sato, S., Okuda, N., Sato, T., Takemiya, H., Saito, K., 2015. The air dose rate around the Fukushima Dai-ichi Nuclear Power Plant: its spatial characteristics and temporal changes until December 2012. J. Environ. Radioact. 139 (C), 250–259. http://dx.doi.org/10.1016/j.jenvrad.2014.08.020.

Ribeiro, P.J., Diggle, P.J., 2001. geoR: a package for geostatistical analysis, 15-18. ISSN R-NEWS 1 (2), 1609–3631.

Saito, K., Onda, Y., 2015. Outline of the national mapping projects implemented after the Fukushima accident. J. Environ. Radioact. 139 (C), 240–249. http://dx.doi.org/10.1016/j.jenvrad.2014.10.009.

Saito, K., Tanihata, I., Fujiwara, M., Saito, T., Shimoura, S., Otsuka, T., Onda, Y., Hoshi, M., Ikeuchi, Y., Takahashi, F., Kinouchi, N., 2015. 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. Radioact. 139 (C), 308–319. http://dx.doi.org/10.1016/j.jenvrad.2014.02.014.

Sassen, D.S., Hubbard, S.S., Bea, S.A., Chen, J., Spycher, N., Denham, M.E., 2012. Reactive facies: an approach for parameterizing field scale reactive transport models using geophysical methods. Water Resour. Res. 48, W10526.

Seki, A., Saito, K., Takemiya, H., 2016. Providing the Radioactive Substance Monitoring Data on the Fukushima Nuclear Accident through Database System submitted for publication.

Takahashi, M., Nasahara, K.N., Tadono, T., Watanabe, T., Dotsu, M., Sugimura, T., Tomiyama, N., 2013, July. JAXA High Resolution Land-use and Land-cover Map of Japan. In: Geoscience and Remote Sensing Symposium (IGARSS), 2013 IEEE International. IEEE, pp. 2384–2387.

Tanaka, S., 2012. Accident at the Fukushima dai-ichi nuclear power stations of TEPCO —outline & lessons learned. Proc. Jpn. Acad. Ser. B Phys. Biol. Sci. 88 (9), 471–484.

Tanigaki, M., Okumura, R., Takamiya, K., Sato, N., Yoshino, H., Yoshinaga, H., Kobayashi, Y., Uehara, A., Yamana, H., 2015. Development of KURAMA-II and its operation in Fukushima. Nuclear Instruments and methods in physics research section a: accelerators, spectrometers. Detect. Assoc. Equip. 781, 57–64.

Torii, T., Sanada, Y., Sugita, T., Kondo, A., Shikaze, Y., Urabe, Y., 2012. Investigation of Radionuclide Distribution Using Aircraft for Surrounding Environmental Survey from Fukushima Daiichi Nuclear Power Plant. JAEA-Technology, pp. 2012–2036 (in Japanese) 2012.

Tsuda, S., Yoshida, T., Tsutsumi, M., Saito, K., 2015. Characteristics and verification of a car-borne survey system for dose rates in air: KURAMA-II. J. Environ. Radioact. 139 (C), 260–265. http://dx.doi.org/10.1016/j.jenvrad.2014.02.028.

Wainwright, H.M., Chen, J., Sassen, D.S., Hubbard, S.S., 2014. Bayesian hierarchical approach and geophysical data sets for estimation of reactive facies over plume scales. Water Resour. Res. 50 (6), 4564–4584.

Wainwright, H.M., Orozco, A.F., Bücker, M., Dafflon, B., Chen, J., Hubbard, S.S., Williams, K.H., 2015. Hierarchical Bayesian method for mapping biogeochemical hot spots using induced polarization imaging. Water Resour. Res. 52, 533–551. http://dx.doi.org/10.1002/2015WR017763.

Wikle, C.K., Milliff, R.F., Nychka, D., Berliner, L.M., 2001. Spatiotemporal hierarchical bayesian modeling tropical ocean surface winds. J. Am. Stat. Assoc. 96 (454), 382–397.

Zhou, Y., Michalak, A.M., 2009. Characterizing attribute distributions in water sediments by geostatistical downscaling. Environ. Sci. Technol. 2009 43 (24), 9267–9273.

Please cite this article in press as: Wainwright, H.M., et al., A multiscale Bayesian data integration approach for mapping air dose rates around the Fukushima Daiichi Nuclear Power Plant, Journal of Environmental Radioactivity (2016), http://dx.doi.org/10.1016/j.jenvrad.2016.11.033