OXFORD

Development of a prediction model for ambient dose equivalent rate distribution based on ecological half-life profiles using LASSO regression and KURAMA data

Yoshiaki Shikaze[1,*] , **Kimiaki Saito**[2], **Naoki Tanimura**[3], **Kazuya Yoshimura**[4] ,
Liu Xudong[2], **Masahiko Machida**[2]

[1]Nuclear Emergency Assistance and Training Center, Nuclear Safety and Emergency Preparedness Institute, Japan Atomic Energy Agency, 178-4-4 Wakashiba, Kashiwa, Chiba 227-0871, Japan
[2]Center for Computational Science & e-Systems, Japan Atomic Energy Agency, 178-4-4 Wakashiba, Kashiwa, Chiba 227-0871, Japan
[3]Science Solution Division, Mizuho Research & Technologies, Ltd., 2-3 Kanda-Nishikicho, Chiyoda-ku, Tokyo 101-8443, Japan
[4]Fukushima Research and Engineering Institute, Japan Atomic Energy Agency, 45-169 Sukakeba, Kaibama, Haramachi-ku, Minamisoma, Fukushima 975-0036, Japan

*Corresponding author. Japan Atomic Energy Agency, 178-4-4 Wakashiba, Kashiwa, Chiba 227-0871, Japan.
E-mail: shikaze.yoshiaki@jaea.go.jp

Abstract

The two-component model, comprising a fast-decay and a slow-decay component, has been widely used to approximate the decreasing trends of air dose rates in contaminated areas surrounding major nuclear accident sites. However, its adequacy is yet to be thoroughly validated. This study analyzed extensive car-borne survey data collected from 2011 to 2016 after the Fukushima Daiichi Nuclear Power Plant accident using the least absolute shrinkage and selection operator regression with a high-degree-of-freedom model. This analysis aimed to evaluate the adequacy of the two-component model and investigate the profiles of ecological half-lives. The results demonstrate that the two-component model can approximate the decreasing trend of air dose rates in the Fukushima area well in most cases. However, in ~20% of the cases, the one-component model provided a better fit. The fast-decay component in the two-component model exhibited a sharp ecological half-life peak below 1 y, with a frequency distribution peaking at 0.3–0.4 y. In contrast, the slow-decay component displayed a broader half-life peak in approximately half of the cases, with a frequency distribution spanning several years to over 50 y. The reduction speed of air dose rates was fastest in urban areas, followed by paddy fields, croplands, deciduous forests, and evergreen forests. The reduction speed decreased as the initial air dose rate increased, a trend explained by the weight assigned to the fast-decay component rather than the value of its ecological half-life. Future predictions of air dose rate distributions were made using a prediction model formula that incorporated the average ecological half-life profiles calculated for each land-use and initial air dose rate category. Prediction accuracy was verified through comparison with integrated map data, which merge air dose rate datasets obtained using different monitoring methods and represent the most currently reliable source. The predicted values tended to decay faster overall than the integrated map data, with an average deviation within 10% over the six-year period. This discrepancy arises because the car-borne survey data were collected on paved roads, where air dose rates decrease rapidly due to the quick washout of radiocesium. The differences between the prediction model values and integrated map values were larger for forests than for farmlands (paddy fields and croplands) and urban areas. This suggests that the reduction in air dose rates is slower in pure forests, where the car-borne surveys were rarely conducted.

Introduction

The Fukushima Daiichi Nuclear Power Plant accident (Fukushima accident) in 2011 caused the widespread deposition of large amounts of anthropogenic radionuclides in Japan, considerably increasing environmental radiation levels [1]. Over time, although radiation levels in certain affected areas have declined, some other regions are expected to remain highly radioactive for prolonged periods owing to the long half-life of ^{137}Cs. Understanding these radiation level trends is crucial for implementing effective measures aimed at radiation protection, environmental remediation, social recovery, and future accident preparedness.

After the Fukushima accident, various organizations conducted extensive environmental monitoring, accumulating substantial data [2–5]. In particular, monitoring activities initiated by the Ministry of Education, Culture, Sports, Science, and Technology and later managed by the Nuclear Regulatory Authority provided systematic and reliable data [2]. These data are publicly accessible through a designated database [6, 7].

Multiple studies have analyzed the decline in ambient dose equivalent rates (air dose rates) using ecological half-lives [8–16], which reflect the rate of decrease in air dose rates caused by radiocesium migration, excluding physical decay contributions. Among these, many studies have used a two-component model for ecological half-life determination, assuming that the air dose rate decreases in the form of two exponential functions representing short and long ecological half-lives.

Assessments following the Fukushima accident support the effectiveness of the two-component model as it fits monitoring data well. Most studies in this context have estimated the short half-life to be below 1 y, while the long half-life has been reported to range from several to several tens of years, as summarized by Saito *et al.* [17]. However, the suitability of the two-component model as the best-fitting model for observed data remains uncertain, warranting further investigations into the optimal number of components for fitting the time series of monitoring data recorded after the accident and how ecological half-lives vary with land use and dose rates.

In a previous study, the least absolute shrinkage and selection operator (LASSO) regression analysis [18, 19] was performed using a model with multiple exponential functions representing different ecological half-lives. This analysis aimed to identify the optimal number of components for modeling the observed air dose rate trend. In general regression analysis on observed data with uncertainties, numerous parameters can lead to overfitting, where the regression curve closely follows data fluctuations. LASSO regression analysis mitigates this issue by adding an L1 regulation term to the loss function, enabling the determination of both the optimal number of basis functions and their parameters.

Furthermore, future predictions of air dose rate distributions can be made using ecological half-life profiles derived from the weights of each decay component identified through LASSO analysis, tailored to specific land use conditions and initial air dose rate ranges.

Recently, integrated air dose rate maps were created by combining datasets obtained from various monitoring methods, including airborne, car-borne, walk surveys, and fixed-location measurements [20]. These maps provide the most reliable radiation-level data across extensive areas and time spans. This integration method was designed to estimate air dose rate distributions equivalent to those obtained from walk surveys. Data integration was conducted using a geostatistical Bayesian method applied to extensive monitoring data systematically collected from the entire Fukushima region, and accuracy validation was performed using multiple approaches. These maps are expected to be effectively utilized for providing initial data of predictions and validating prediction performance.

The primary goal of this study was to determine the optimal number of components in the trend model for approximating temporal changes in air dose rates observed after the Fukushima accident and to characterize ecological half-lives. To this end, LASSO regression analysis was applied to extensive car-borne survey data collected after the accident [12, 21]. The secondary goal was to develop a prediction model using ecological half-life profiles derived from LASSO analysis and evaluate its accuracy by comparing the predicted values of 2022, calculated based on integrated map data of 2016, with actual integrated map data of 2022. Here, pixels with initial air dose rates of 0.2 μSv/h or less in June 2011 were excluded as their radiation levels were comparable to background air dose rates (BGs) determined from the car-borne survey data [22], making it difficult to analyze the decreasing trend due to radiocesium accurately.

Materials and methods
Air dose rate data
The air dose rate data used for the LASSO analysis had to meet several key criteria. In particular, measurements were required to have started during the early phase after the accident and to have been conducted regularly at specific intervals to accurately capture the temporal trend. Additionally, the measurements were required to cover wide areas with diverse conditions, including varying land uses, geological settings [23–25], and air dose rate levels, yielding sufficient air dose rate data suitable for statistical analysis. Ensuring the reliability and accuracy of the measurements was also essential.

Given these requirements, data from the car-borne surveys conducted as part of the Fukushima mapping project were selected [12, 21]. This project adopted four methods to obtain air dose rate measurements, with car-borne surveys generating extensive data. The first car-borne survey was initiated in June 2011 and subsequently conducted twice a year. These surveys covered extensive areas in eastern Japan [12, 21] using the Kyoto University RAdiation MApping (KURAMA) system [26, 27].

Table 1. Details of the car-borne survey data used for the least absolute shrinkage and selection operator (LASSO) analysis. Pixels containing data from >11 out of the 13 campaigns were selected for inclusion in the analysis.

Campaign no.	Duration (day/month/year)	Elapsed time after the accident (y)	Number of pixels
1	6/6/2011–13/6/2011	0.24	45 273
2	5/12/2011–28/12/2011	0.76	36 355
3	13/3/2012–30/3/2012	1.02	38 741
4	20/8/2012–12/10/2012	1.51	132 380
5	5/11/2012–10/12/2012	1.69	89 385
6	12/6/2013–8/8/2013	2.32	129 666
7	5/11/2013–12/12/2013	2.69	119 522
8	23/6/2014–8/8/2014	3.34	102 473
9	4/11/2014–5/12/2014	3.68	102 881
10	29/6/2015–4/8/2015	4.34	100 235
11	2/11/2015–18/12/2015	4.70	89 849
12	27/6/2016–5/8/2016	5.34	100 854
13	31/10/2016–16/12/2016	5.69	86 255

The measurement performance of the system was evaluated using a standard irradiation facility to assess gamma energy, dose rate, and directional responses [28]. This evaluation confirmed that the performance met the Japan Industrial Standard (JIS) Z4333, the authorized standard for commercially available survey meters in Japan (JIS, 2006). Andoh *et al.* estimated the overall uncertainty of the car-borne measurements using the KURAMA-II system to be 15% [21]. During the mapping project, car-borne survey data were averaged over a 100×100 m^2 area to minimize statistical fluctuations. Consequently, temporal changes in the average air dose rate were analyzed within each 100×100 m^2 pixel.

The analysis used car-borne survey data from Fukushima Prefecture collected during 13 campaigns, extending from June 2011 to December 2016. This time frame was chosen to align with those of previous studies [29–31] for comparability under similar conditions. Table 1 provides an overview of the 13 survey campaigns. For consistency, the middle day of each campaign period was designated as the measurement day for all collected data. When determining the measurement dates for each campaign period as described above, the uncertainty of the measurement dates is within half of each campaign period, i.e. at most 1 month, which is deemed negligible compared to other uncertainties. The time elapsed after the accident was calculated from noon on 15 March 2011. The following analysis was performed for pixels containing data from >11 of the 13 campaigns. In other words, a pixel was included in the analysis if it had no more than one missing data point across the 13 campaigns. A total of 21 341 data points met this criterion.

If a missing datum was encountered in the series of air dose rates, it was complemented using the random forest method [32, 33]. This technique estimates missing data either by leveraging information from other locations or by analyzing data from different time points at the same location. In this case, the former approach was deemed more suitable. Consequently, we utilized a random forest model trained on reference data from multiple locations to estimate the missing values, thereby generating a complete dataset free of gaps.

Least absolute shrinkage and selection operator analysis

In the LASSO analysis, the decreasing trend of observed air dose rates at each pixel was fitted using a linear combination model comprising an arbitrary number of exponential functions with different ecological half-lives. The optimal parameters were determined through this approach. The following equation was used for the fitting:

$$D(t) = c_0 + \frac{k \exp(-\lambda_{134}t) + \exp(-\lambda_{137}t)}{k+1}$$
$$\times \sum_{i=1}^{N_e} c_i \exp(-\lambda_i t) \qquad (1)$$

where the symbols have the following meanings:
$D(t)$: air dose rate at time t (μSv/h),
k: ratio of air dose rates contributed by ^{134}Cs and ^{137}Cs at $t = 0$,
λ_{134} and λ_{137}: radioactive decay constants of ^{134}Cs and ^{137}Cs (y^{-1}), respectively,
N_e: number of assumed basis functions,
$\exp(-\lambda_i t)$: basis function,

λ_i: attenuation constant related to each ecological half-life (y^{-1}),

c_i: coefficient of each basis function (μSv/h).

In this study, N_e was set to 50, meaning that 50 exponential functions with different attenuation constants were employed. Previous studies [29–31] have shown that the distribution of ecological half-lives is primarily composed of fast half-lives (<1 y) and slow half-lives (5–20 y). Following these studies, we set N_e to 50 as the number of basis functions sufficient to depict the distribution when half-lives range from 0.1 to 50 y. Here, c_0 represents the background radiation level at each location, and c_i was constrained to be nonnegative (nonnegative LASSO). The i of c_i is an integer from 1 to 50 ($=N_e$). The ecological half-life $t_{1/2}$ is related to the attenuation constant λ through the equation $\ln2/t_{1/2} = \lambda$.

The optimal combination of c_i coefficients was determined by minimizing the following objective function:

$$f(\mathbf{C}) = \frac{1}{N_t} \sum_{s=1}^{N_t} (D_m(t_s) - D(t_s))^2 + \alpha \|\mathbf{C}\|_1 \qquad (2)$$

where the symbols have the following meanings:
C: vector representation of $c_1, \ldots, c_{Ne}$,
N_t: number of sequential data,
$D_m(t_s)$: air dose rate predicted by the model at t_s,
$D(t_s)$: observed air dose rate at t_s,
α: L1 norm coefficient.

Here, N_t was set to 13 to match the number of car-borne survey campaigns. The L1 norm term added to the objective function prevents the selection of excessive exponential functions. The results of the LASSO analysis depend on the value of α: A larger α tends to select a smaller number of exponential functions. In this study, the α value was determined objectively using a cross-validation method.

The L1 norm coefficient, called the hyperparameter α in LASSO, can be optimized using cross-validation. According to the validation scheme, the input data are divided into k-fold groups, and one of these groups is used for testing, while the remaining $k-1$ groups are used for training. LASSO regression is then applied to the training data, yielding $k-1$ LASSO solutions for each α. The error is evaluated on the test data to optimize α [34]. For this study, k was set to five.

The shortest and longest ecological half-lives were assumed to be 0.1 and 50 y, respectively, based on the results of previous studies. Subsequently, the ecological attenuation coefficients (λ_i) were chosen to evenly represent air dose rate attenuation trends across these half-lives (Fig. 1). Table 2 lists the ecological half-lives corresponding to the λ_i values. Data from 0.1 to 50 y in Table 2 was set using an empirical formula to ensure

exponential growth, as in previous studies [29–31]. A set of c_i values was determined through LASSO regression analysis for each of the 21 341 targeted pixels, classified based on land-use categories according to Version 14.2 of the Japan Aerospace Exploration Agency's (JAXA's) data [35]: water, urban, paddy, crop, grass, deciduous forest, evergreen forest, and bare surface.

Prediction method

The prediction was performed using formula (1) by extending the parameter t into the future. The coefficients c_i, defined in formula (1) as the weights of the basis functions (exponential functions with different ecological half-lives) for each pixel were normalized to create an ecological half-life profile for prediction. The normalized c_i values ($i = 1$–50) of all measurement points were classified according to five land-use categories (paddy fields, croplands, urban areas, deciduous forests, and evergreen forests) and four ranges of air dose rates obtained from the initial car-borne survey (0.2–0.5, 0.5–1.0, 1.0–1.9, and >1.9 μSv/h), resulting in 20 conditions. For each condition, the average normalized c_i values ($i = 1$–50) were calculated to serve as the ecological half-life profile for prediction.

The decreasing trend of air dose rates is well known to vary according to land-use category [9, 10, 12, 21]. Additionally, previous studies have noted that these trends depend on the evacuation order classifications of the given areas, likely due to variations in human activities such as decontamination. In this study, we used the aforementioned initial air dose rate categories instead of evacuation order classifications. This approach avoids the arbitrariness of selecting evacuation order classifications, which change over time. Because evacuation orders are determined based on air dose rate levels, the initial air dose rates must be related to the evacuation orders and serve as more objective indicators of the decreasing trends. However, in order to ensure statistical accuracy (Table 3), the category of '>1.9 μSv/h', which is divided into five categories in the air dose rate distribution maps in the previous report [12], was analyzed as a single category.

The data used for prediction were selected based on the variation between the values calculated using parameters from LASSO regression analysis and the actual values measured during the car-borne survey. To calculate the average normalized c_i values ($i = 1$–50) for each condition, $\sim$10% of the LASSO results, which showed significant deviations, were excluded.

To examine the variations, instead of using the residual sum of squares directly, we calculated the sum of the residual squares over 13 periods, normalized by the calculated value for each period (referred to as nRSS).
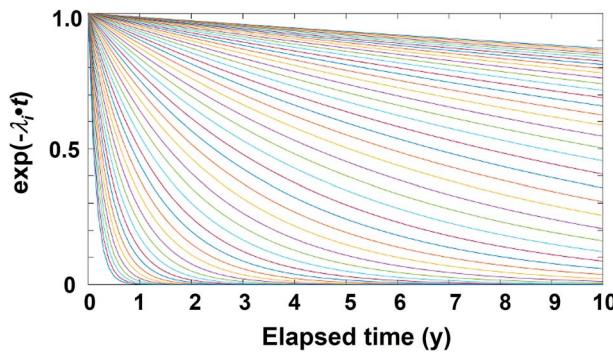
Figure 1. Tendency of air dose rate reduction for different ecological half-lives assumed in the basis functions.

Table 2. Ecological half-lives assumed in the basis functions for the LASSO analysis. The 50 half-lives, ranging from 0.1 to 50 y, corresponded to their respective ecological attenuation coefficients (λ_i) to evenly represent air dose rate attenuation trends.

No.	Ecological half-life (y)									
	1	2	3	4	5	6	7	8	9	10
0	0.10	0.12	0.14	0.16	0.18	0.22	0.25	0.29	0.34	0.40
10	0.46	0.54	0.63	0.73	0.85	0.99	1.15	1.34	1.56	1.81
20	2.11	2.44	2.83	3.28	3.80	4.39	5.07	5.84	6.71	7.70
30	8.82	10.1	11.5	13.0	14.7	16.5	18.5	20.7	23.0	25.4
40	27.9	30.4	33.0	35.6	38.2	40.8	43.3	45.6	47.9	50.0

Table 3. Number of pixels classified by land use and initial dose rate.

Land use	Number of pixels					Total
	Initial air dose rate (μSv/h)					
	0–0.2	0.2–0.5	0.5–1.0	1.0–1.9	>1.9	
Paddy	865	2188	1712	670	300	5735
Crops	507	1487	1001	658	338	3991
Deciduous forests	428	2051	1076	552	411	4518
Evergreen forests	265	914	346	168	232	1925
Urban	582	1860	1673	885	172	5172
Total	2647	8500	5808	2933	1453	21 341

This normalization ensured that smaller decayed values in later periods did not disproportionately influence the results. Data with an nRSS of ≤0.6 (90% of the total) were selected, and the average normalized c_i values ($i = 1$–50) were calculated to create the ecological half-life profile for each condition.

Using the average normalized c_i values ($i = 1$ to 50) representing the ecological half-life profiles for each condition, the temporal variation of air dose rates at each pixel, starting from 15 March 2011, can be calculated with formula (1). The BG was assumed to be 0.050 μSv/h, based on previous studies [36, 37].

In this study, the initial air dose rate for prediction was set using the integrated map data described earlier. For predicting values beyond 2016, the following formula (3) was applied, taking the air dose rate from the 2016 integrated map dataset as the initial value. The predicted value for year T is given by

$$\text{Predicted value for } T = [ADR_{int}(2016) - BG]$$
$$\times [DR_{2011}(T)/DR_{2011}(2016)]$$
$$+ BG \quad (3)$$

where the symbols have the following meanings:
$ADR_{int}(2016)$: air dose rate obtained from the integrated map in 2016 and is used as the initial value in this prediction calculation for the years after 2016,
BG: air dose rate attributed to background radiation,
$DR_{2011}(T)$: decay rate from 2011 in the year T (>2016),

which is the ratio of 'air dose rate subtracted by background radiation' in the year T to that in 2011, $DR_{2011}(2016)$: decay rate from 2011 in 2016, which is the ratio of 'air dose rate subtracted by background radiation' in 2016 to that in 2011.

$[ADR_{int}(2016) - BG]$ is the decaying component. $DR_{2011}(T)$ is the ratio of (air dose rate—BG) in T to that in 2011, where the year T is after 2016. By multiplying the decaying component by the fraction $[DR_{2011}(T)/DR_{2011}(2016)]$, the decaying component in the year T is obtained. Finally, by adding BG, air dose rate in the year T is obtained. The air dose rate at each 50 m × 50 m pixel in the integrated map was predicted using this formula.

The predicted map was created by combining the results for different land-use categories and initial air dose rates. Land use was classified into paddy fields, croplands, urban areas, deciduous forests, and evergreen forests for the predictive mapping. For areas not classified under these categories, the ecological half-life profile of paddy fields was used. During the prediction, the land-use category of each pixel was determined using JAXA's Ver. 21.11 data [38], which represent the land-use conditions for the prediction period.

Results and discussion

A set of c_i values was determined through LASSO regression analysis for each of the 21 341 targeted pixels, classified based on land-use categories according to Version 14.2 of the JAXA's data [39]: water, urban, paddy, crop, deciduous forests, evergreen forests, and bare surface. Our analysis specifically focused on the five dominant land-use categories—urban, paddy, crop, deciduous forests, and evergreen forests—because each contained a sufficient number of pixels for statistical analysis.

Additionally, only pixels with initial air dose rates >0.2 μSv/h in June 2011 were included. Table 3 presents the number of pixels in each of the five dominant categories. After filtering base on initial dose rates, the total number of analyzed pixels was 18 694.

Figure 2 illustrates examples of the decreasing trend in air dose rates, calculated using Equation (1) with the determined parameters and observed air dose rates. The results indicate that the LASSO analysis successfully reconstructed the decreasing trends.

Model component features

The number of components required to adequately represent the decreasing trend of air dose rates was determined based on the LASSO analysis results. The following rule was established to count the number of components: if nonzero c_i values were determined sequentially, they were recognized as a single component. For instance, if nonzero values were identified for c_{11}, c_{12}, and c_{13}, the series of these coefficients was counted as one component. In this analysis, 50 discrete ecological half-lives were formulated. If an actual ecological half-life existed between two formulated half-lives, it was expected that c_i values would be determined for the two exponential functions corresponding to the half-lives on either side of the actual value. Furthermore, radiocesium migration in the real environment is complex, and some radiocesium groups with similar migration characteristics are challenging to analyze separately, leading to their classification within the same component.

Figure 3 illustrates examples of the determined c_i values at four different pixels, all of which were identified as two-component models based on the established rule. The fast-decay component consistently exhibited a sharp peak below 1 y, while the slow-decay component displayed a variety of shapes and positions. These variations will be discussed later.

The number of components required to adequately represent the trend was three or fewer in all cases. Table 4 lists the proportions of the component counts for the selected models across the five land-use categories, showing negligible differences based on land use. The proportion of three-component models was <1% in all land-use categories. Approximately 80% of the cases employed the two-component model, while 20% used the one-component model. These results indicate that the two-component model, which has been widely used, provides an appropriate approximation for representing radiation level trends in most cases.

Next, we analyzed how many exponential functions constituted each model component. Figures 4 and 5 present statistics on the number of exponential functions constituting the two-component and one-component models, respectively. The fast-decay component of the two-component model showed a sharp ecological half-life peak. In over 90% of the cases, this component was represented by one or two exponential functions, and in ∼98% of the cases, it was represented by one to three functions. Thus, the ecological half-life of the fast-decay component was typically represented by a single value or a narrow half-width peak.

The slow-decay component of the two-component model exhibited different features. Approximately 50% of the cases used a single exponential function, while the other 50% employed multiple functions, distributed relatively evenly up to 20 functions. The former cases included those in which the ecological half-life was determined to be 50 y. In these cases,
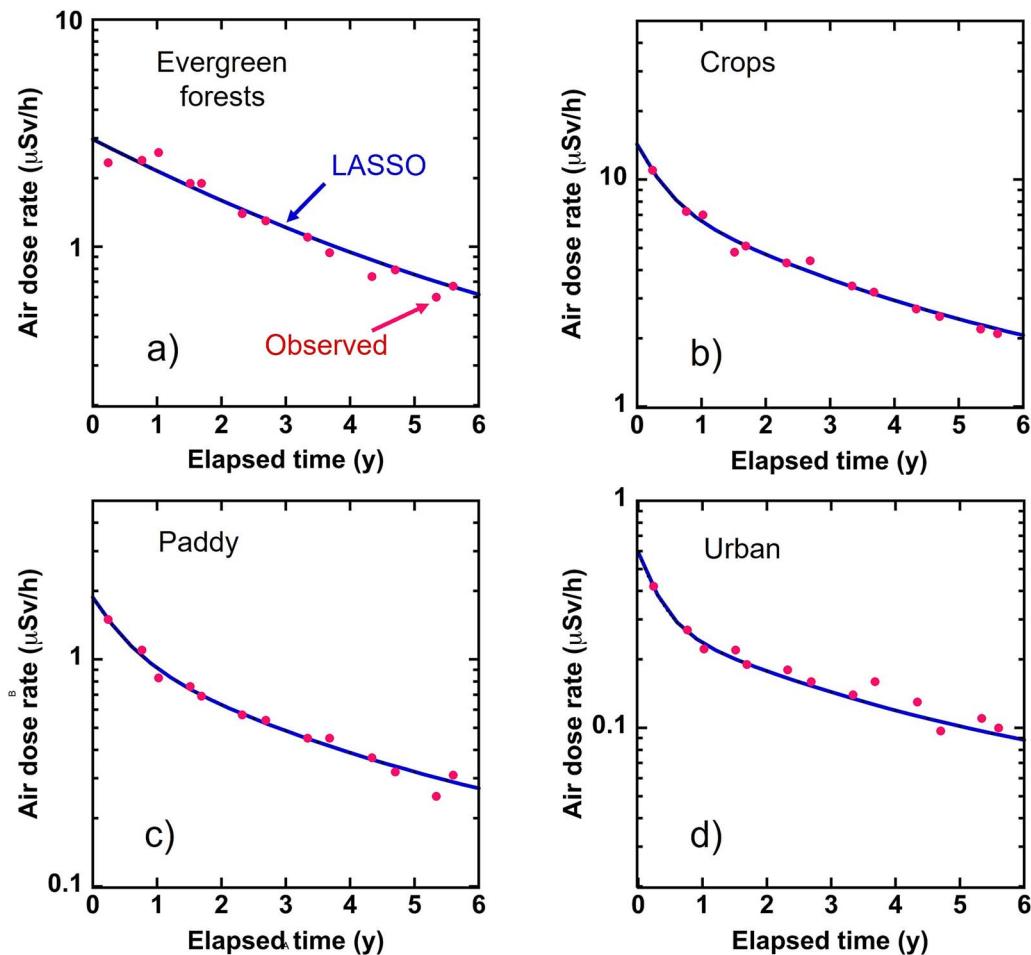
Figure 2. Comparison of LASSO analysis results with observed air dose rates.

Table 4. Proportion of selected component count models.

Land use	Proportion		
	Number of components		
	1	2	3
Paddy	0.21	0.78	0.006
Crops	0.18	0.81	0.007
Deciduous forests	0.17	0.82	0.006
Evergreen forests	0.24	0.75	0.004
Urban	0.19	0.81	0.005
Total	0.19	0.8	0.006

the actual half-life likely had a range exceeding 50 y. Therefore, most ecological half-lives of the slow-decay component had some degree of width.

For the one-component model, two or three exponential functions are the most likely, with 80% of this model comprising fewer than seven basis functions and the remaining model comprising up to 25 functions. The half-lives of the one-component model exceed 1 y, corresponding to the slow-decay component of the two-component model.
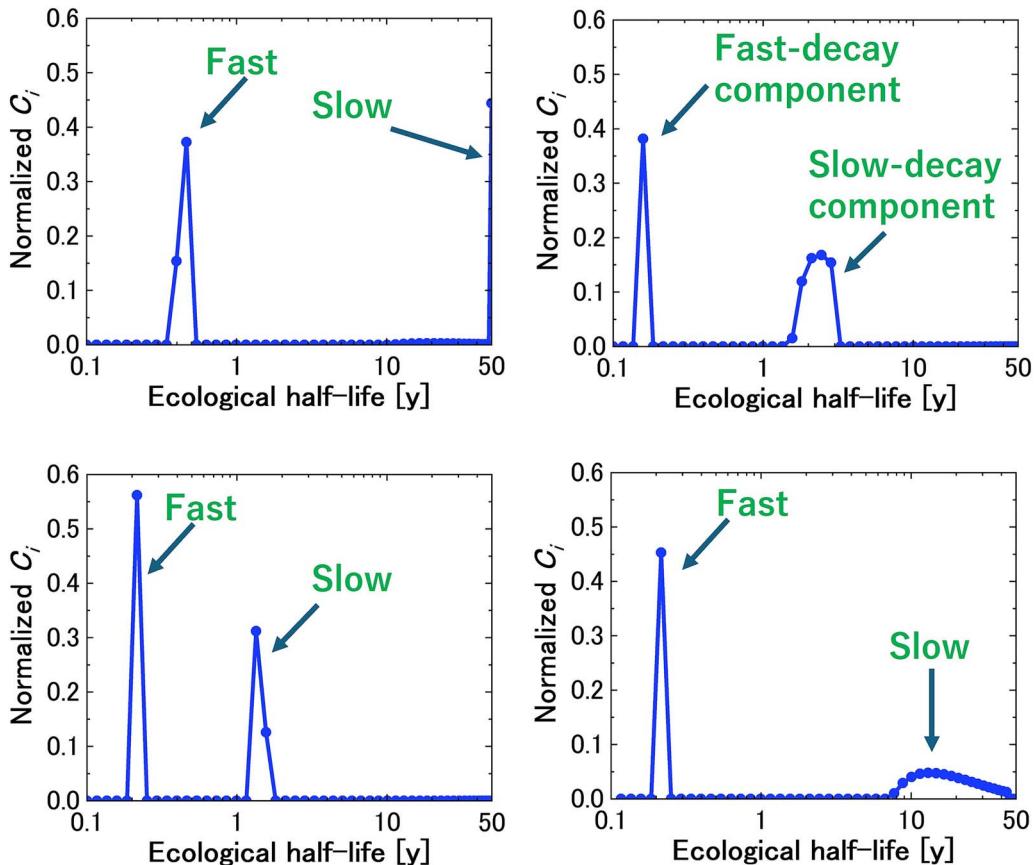
Figure 3.   Examples of the determined c_i values at four different pixels of paddy fields.
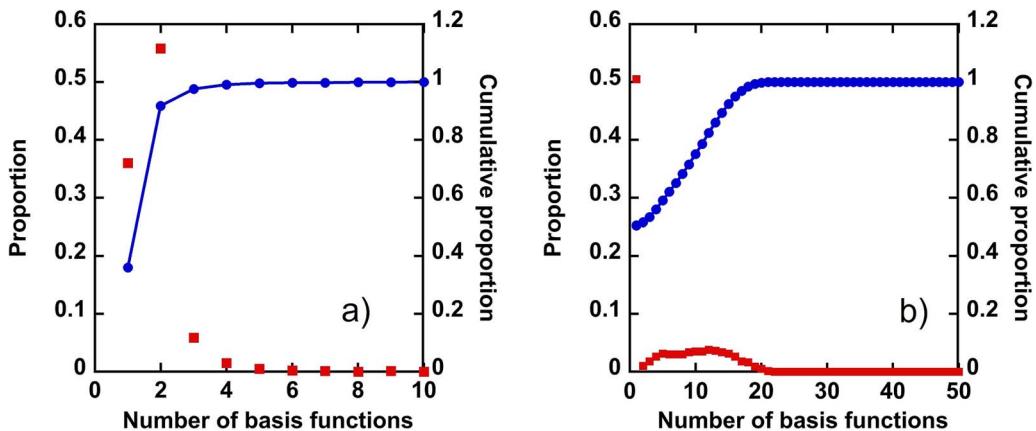
Figure 4.   Distributions of the basis function counts for (a) the fast-decay and (b) slow-decay components of the two-component model. The left vertical axis represents proportion, shown as filled squares in the figure, and the right vertical axis represents cumulative proportion, shown as filled circles in the figure.

The reasons for the aforementioned specific features are not yet fully understood and represent topics for future research. However, it is speculated that the

5-year period following the accident was sufficient to determine the exact half-life of the fast-decay component, which is generally <1 y. Conversely, this period
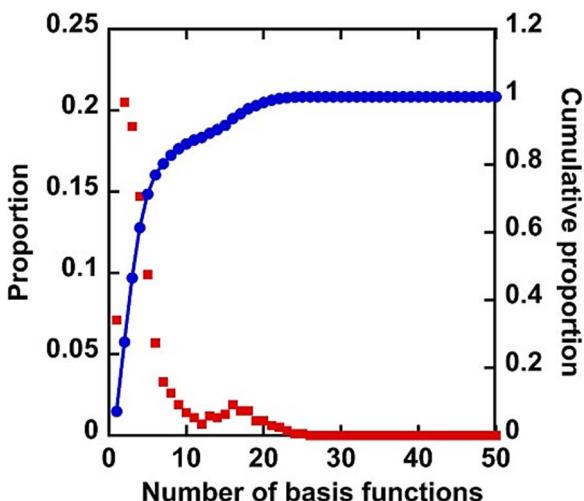
Figure 5. Distributions of the basis function counts for the one-component model. The left vertical axis represents proportion, shown as filled squares in the figure, and the right vertical axis represents cumulative proportion, shown as filled circles in the figure.

may have been excessively short to determine the half-life of the slow-decay component, which spans several decades.

Average profile of ecological half-lives

The characteristics of the average ecological half-life profiles were analyzed by normalizing the coefficients c_1–c_{50} to their sum as their absolute values vary with the radiation level. This normalization was intended to evaluate the relative relationships among the coefficients. Each coefficient was then averaged across all pixels within the same land-use category, where the initial air dose rates exceeded 0.2 μSv/h.

Figure 6 illustrates the average profiles of normalized coefficients for basis functions with different ecological half-lives across the five land-use categories. Regardless of land use, the profiles display a distinct peak below a half-life of 1 y, corresponding to the fast-decay component. Additionally, the profile reveals a broad, low hump spanning half-lives from 1 to >10 years, corresponding to the slow-decay component. The coefficient c_{50}, representing a half-life of 50 y, maintains a specific value. The average normalized coefficient values were summed for half-life ranges of <1 and >1 y (Table 5).

The difference in the summed coefficient values reflects variations in the trend of air dose rate reduction. Figure 7 illustrates the decreasing trend of air dose rates during the 2 years following the accident for different land-use categories. These fundamental features of land-use dependency are consistent with those found in previous studies on car-borne survey

data [10, 12]. The decrease in air dose rates occurs in the following order: urban, crop, paddy, deciduous forests, and evergreen forests. This order aligns with that for the summed coefficient values for ecological half-lives below 1 y. Therefore, the speed of air dose rate reduction in the early stage is associated with the weight (summed coefficients) assigned to exponential functions with ecological half-lives below 1 y (Table 5), rather than the average ecological half-life values shown in Table 6.

The mobility of radiocesium in the environment explains the observed differences in dose rate reduction. Radiocesium deposited on artificial surfaces, such as paved roads, paved parking lots, and building walls and roofs, is quickly removed or washed off after deposition [39], resulting in a rapid air dose rate reduction. In contrast, radiocesium deposited in forests moves slowly within the forest system [40, 41], with minimal discharge from the system [42]. Consequently, dose rate reduction in forests is generally slower, with evergreen forests exhibiting slower reduction rates compared to deciduous forests, consistent with direct observations of air dose rates in forests [43].

Since the car-borne surveys were conducted on paved roads, where deposited radiocesium is removed relatively quickly, the survey data tend to show a faster reduction trend compared to measurements on undisturbed fields. However, gamma rays emitted from the surrounding environment outside the roads are also detected during the car-borne surveys. Consequently, the car-borne survey data reflect the air dose rate reduction tendencies associated with land use.

To investigate the dependence of air dose rate reduction on initial dose rates, the ecological half-life profiles were classified into four initial air dose rate ranges in June 2011: (i) 0.2–0.5 μSv/h, (ii) 0.5–1.0 μSv/h, (iii) 1.0–1.9 μSv/h, and (iv) >1.9 μSv/h.

Figure 8 illustrates the average profiles of ecological half-lives for three representative land-use categories—evergreen forests, paddy, and urban—classified by these four initial dose rate ranges. These categories were selected to avoid over-plotting and potential confusion from excessive data. The ecological half-life profile data [c_i ($i = 1$–50)] for each condition of land use and initial air dose rate category (20 conditions) required for the prediction model equation are listed in Table 7 in Appendix I. The weights assigned to the fast-decay component decrease as the initial air dose rate increases, while the weights for the slow-decay components correspondingly increase. For high dose rate ranges, particularly above 1.9 μSv/h, a distinct peak is observed in the slow-decay component around ecological half-lives of several years.

Figure 9 illustrates the decreasing trend of air dose rates classified by initial dose rate ranges. The reduction
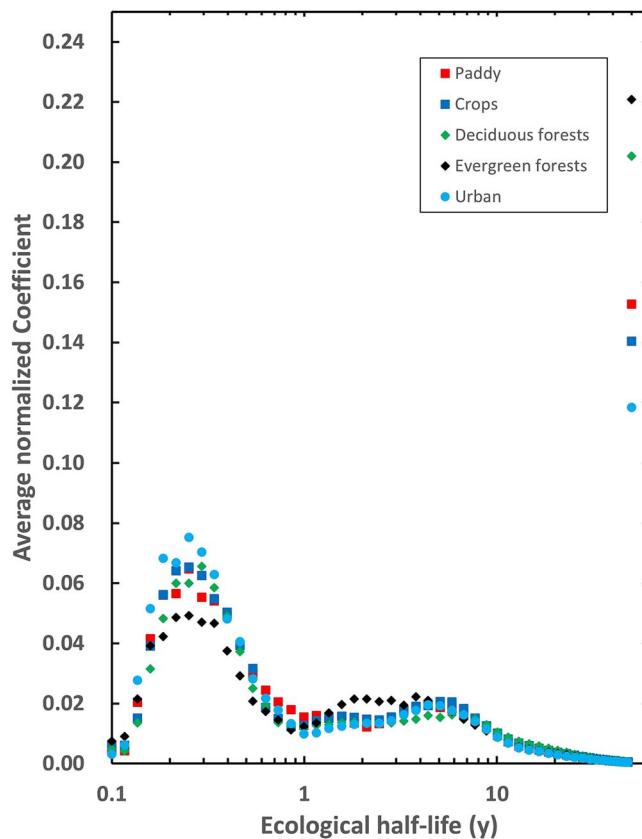
Figure 6. Profile of average ecological half-lives corresponding to the five land-use categories for pixels with initial air dose rates exceeding 0.2 μSv/h. The coefficient c_{50}, representing a half-life of 50 y, maintains a specific value.

Table 5. Sum of the average normalized coefficients of the basis functions for periods of <1 and >1 y.

Dose rate range (μSv/h)	Sum of normalized coefficients									
	Paddy		Crop		Deciduous forests		Evergreen forests		Urban	
	<1 y	>1 y	<1 y	>1 y	<1 y	>1 y	<1 y	>1 y	<1 y	>1 y
0.2–0.5	0.60	0.40	0.62	0.38	0.58	0.42	0.52	0.48	0.64	0.36
0.5–1.0	0.54	0.46	0.54	0.46	0.49	0.51	0.46	0.54	0.57	0.43
1.0–1.9	0.53	0.47	0.52	0.48	0.46	0.54	0.37	0.63	0.61	0.39
>1.9	0.33	0.67	0.32	0.68	0.31	0.69	0.27	0.73	0.36	0.64

rate slows as the initial air dose rate increases, with a noticeably slower decrease observed above 1.9 μSv/h. This trend can be explained by the weight proportions of the fast- and slow-decay components shown in Table 5. A larger proportion of the fast-decay component results in a faster dose rate reduction. However, as the initial dose rate increases, the weight proportion of the fast-decay component decreases, leading to a slower reduction in air dose rates. In low-radiation areas, human activities continued even after the accident, whereas in highly contaminated areas, resident activities were restricted. Human activities, including both official and voluntary decontamination efforts, tend to accelerate air dose rate reductions [17], explain-
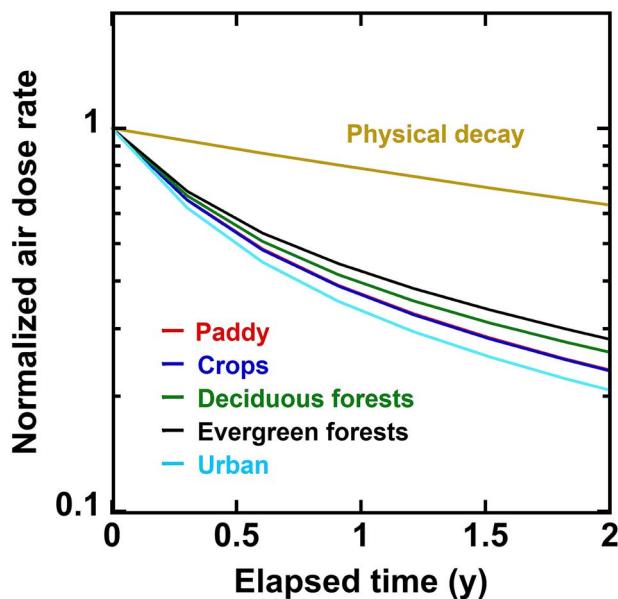
Figure 7. Variations in the decreasing tendency of air dose rates across different land-use categories.

Table 6. Average ecological half-lives for periods below 1 y and exceeding 1 y.

Dose rate range (μSv/h)	Average ecological half-lives (y)									
	Paddy		Crops		Deciduous forests		Evergreen forests		Urban	
	<1 y	1–50 y	<1 y	1–50 y	<1 y	1–50 y	<1 y	1–50 y	<1 y	1–50 y
0.2–0.5	0.32	7.4	0.34	6.7	0.33	7.2	0.35	6.6	0.30	7.0
0.5–1.0	0.40	7.1	0.34	7.4	0.35	8.9	0.35	6.9	0.36	6.7
1.0–1.9	0.42	6.9	0.39	8.0	0.38	8.4	0.32	7.1	0.36	6.6
>1.9	0.32	5.1	0.26	4.8	0.29	6.7	0.26	5.4	0.33	5.0

ing the connection between the initial air dose rate and reduction speed.

The average ecological half-lives for the fast-decay and slow-decay components remain relatively consistent (Table 6). The ecological half-life of 50 y was excluded from the calculation of the average half-life for the slow-decay component because the corresponding exponential function represented an ecological half-life of 50 y and a broader range exceeding 50 y. For the fast-decay component, most average half-lives range between 0.3 and 0.4 y, regardless of initial dose rate or land use. The average half-life for the slow-decay component decreases slightly as the initial dose rate increases, but most values fall between 5 and 8 y.

Future prediction

The ecological half-life profiles classified by conditions for the predictive model, obtained using the method described in the Prediction Method section, are shown in Fig. 8.

The predicted map for 2022 within the 80 km zone was compared with the integrated map for the same year, as shown in Fig. 10. The predicted map (left side of Fig. 10) and the integrated map (right side of Fig. 10) appear reasonably consistent.

A numerical comparison was performed to evaluate the accuracy of the prediction model against the integrated map. Data from each land-use category with initial air dose rates above 0.2 μSv/h were used for this comparison. This threshold was chosen because attenuation trends cannot be reliably evaluated by LASSO analysis for air dose rates near the background level, which reduces the accuracy of the prediction calculation. Additionally, the importance of prediction accuracy decreases for locations with low radiation levels.
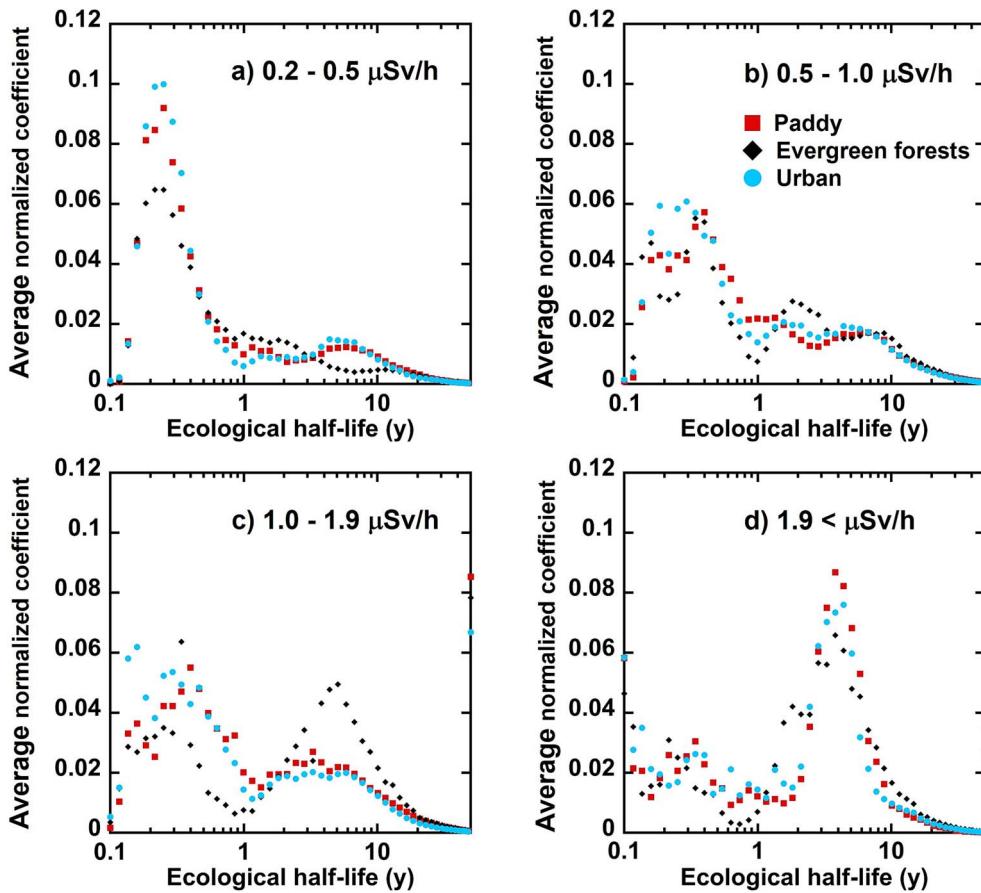
Figure 8. Profiles of average ecological half-lives classified by the initial dose rate.

Figure 11 shows 2D plots with the integrated map values on the vertical axis and the predicted values on the horizontal axis. To ensure accuracy, data within 3 km from 1F were excluded, as air dose rate data in this zone were limited and the integrated map data within this area are considered less accurate.

As depicted in Fig. 11, the data were fitted to a function $y = Ax$, and the proportional constant (A) and coefficient of determination (R^2) were obtained, as presented. The proportional constant obtained is one way of expressing the overall ratio. When comparing the two groups based on this ratio, the results show that the value of the overall prediction model is ~10%–40% smaller than the value of the overall integrated map. This means that the air dose rate value predicted for 2022 starting from 2016 is smaller than the air dose rate value of the integrated map for 2022 with which it should be compared, meaning that the decay over 6 years is faster. Prediction performance is better for paddy fields, croplands, and urban areas, while it is relatively less accurate for deciduous and evergreen

forests. The reasons for these findings will be discussed later.

The R^2 value is ~0.9 or higher, demonstrating a strong correlation between the predicted and integrated map values. For reference, dashed lines representing double and half proportional values are included, showing that most data points fall within this range, or within a factor of two.

To assess the percentage difference between the predicted values and the integrated map values, histograms of the relative differences ([prediction model] – [integrated map])/[integrated map]) are shown for each land use category in Fig. 12a–e. The mean relative difference ranged from −0.018 to −0.062, confirming that the predicted air dose rates are smaller than the integrated air dose rates. However, the degree of relative difference differs from that for the proportion constant A in Fig. 11. The average relative difference remains within 10% for all land-use categories, likely because the proportional constant is more strongly influenced by data from higher air dose rate regions, where prediction performance is poorer.
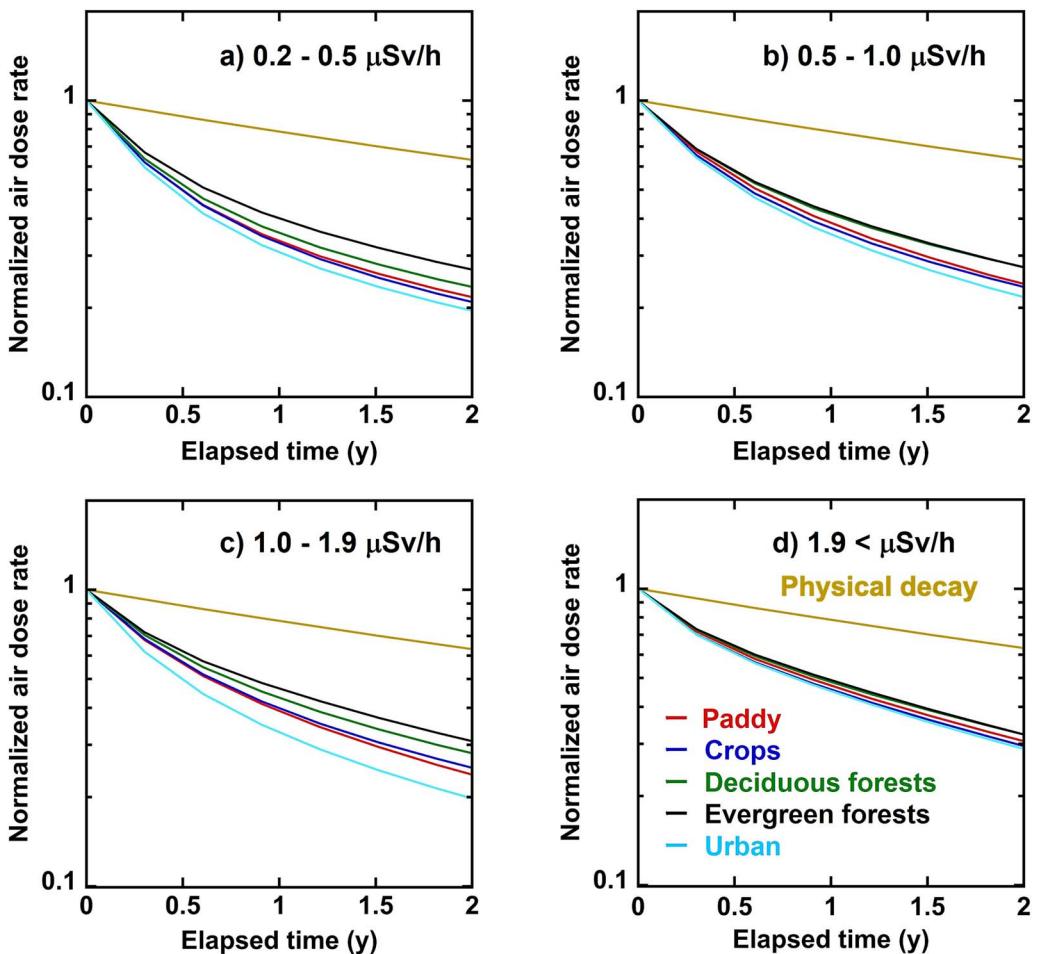
Figure 9. Variations in the reduction trends of air dose rates based on initial dose rates.
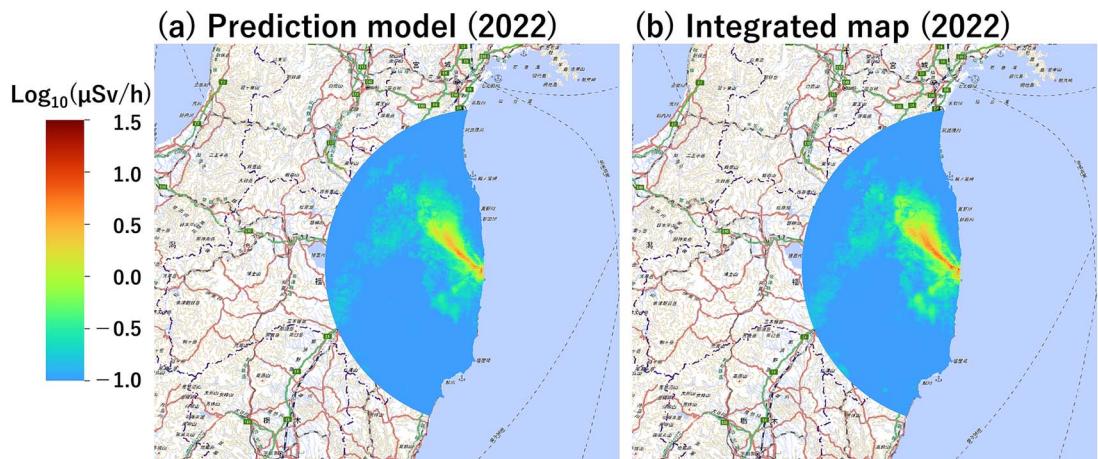

Figure 10. Air dose rate distribution map within 80 km from 1F; (a) prediction model for 2022 based on the integrated map for 2016 and (b) integrated map for 2022. These maps are based on the Basic map (standard map) published by Geospatial Information Authority of Japan.
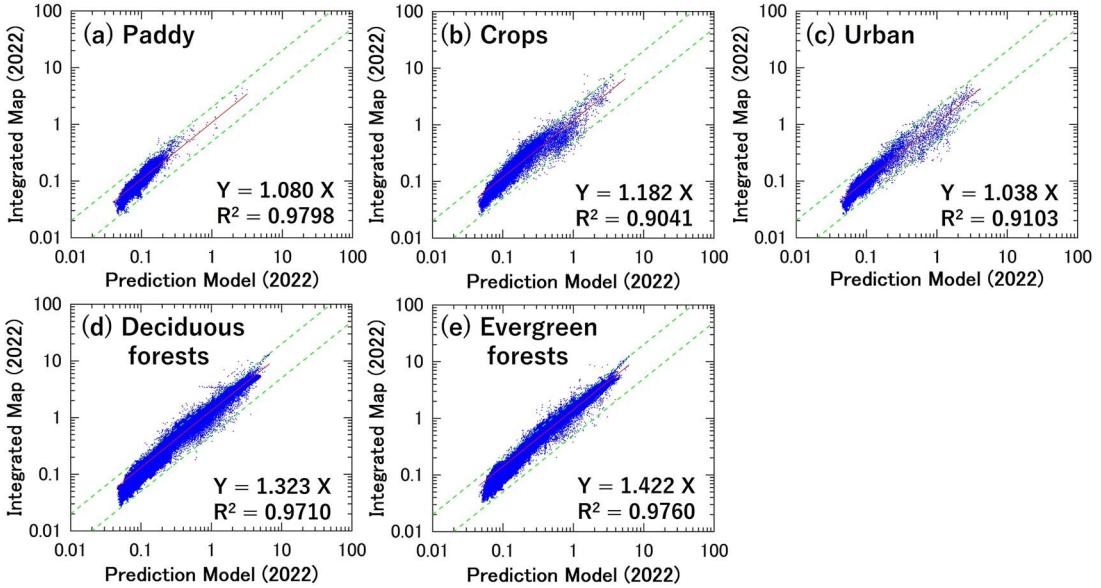
Figure 11. Two-dimensional plots comparing integrated and predicted air dose rates in 2022 for: (a) paddy fields, (b) croplands, (c) urban areas, (d) deciduous forests, and (e) evergreen forests.
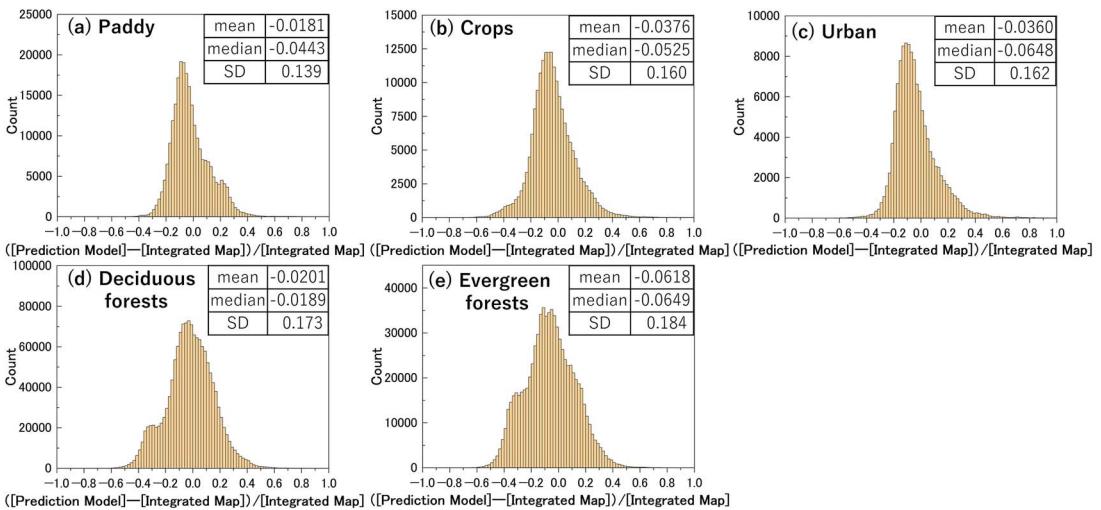
Figure 12. Histograms of the relative difference (([prediction model] − [integrated map])/[integrated map]): (a) paddy fields, (b) croplands, (c) urban areas, (d) deciduous forests, and (e) evergreen forests.

The standard deviation (SD) ranges from 0.139 to 0.184, with most data falling within the range of −0.5 to 2, or within a factor of two. This percentage deviation analysis also indicates that the prediction performance is better for paddy fields, croplands, and urban areas, while it is poor for forests.

In Appendix II, for Figs 11 and 12, four figures for the initial air dose rate categories for each land use are also shown in Figs 15 and 16, respectively.

To evaluate the variation in predicted values for each year, the air dose rates from the prediction calculations (2017–22, starting from the integrated map value in 2016) were analyzed after normalization by the integrated map value in 2016. The average values, as well as the 10th, 50th, and 90th percentile ratios for each land use with initial air dose rates above 0.2 μSv/h, were statistically calculated and are presented in Table 3. The temporal variations in these statistical values according to land use are shown in Fig. 13a–e. For deciduous
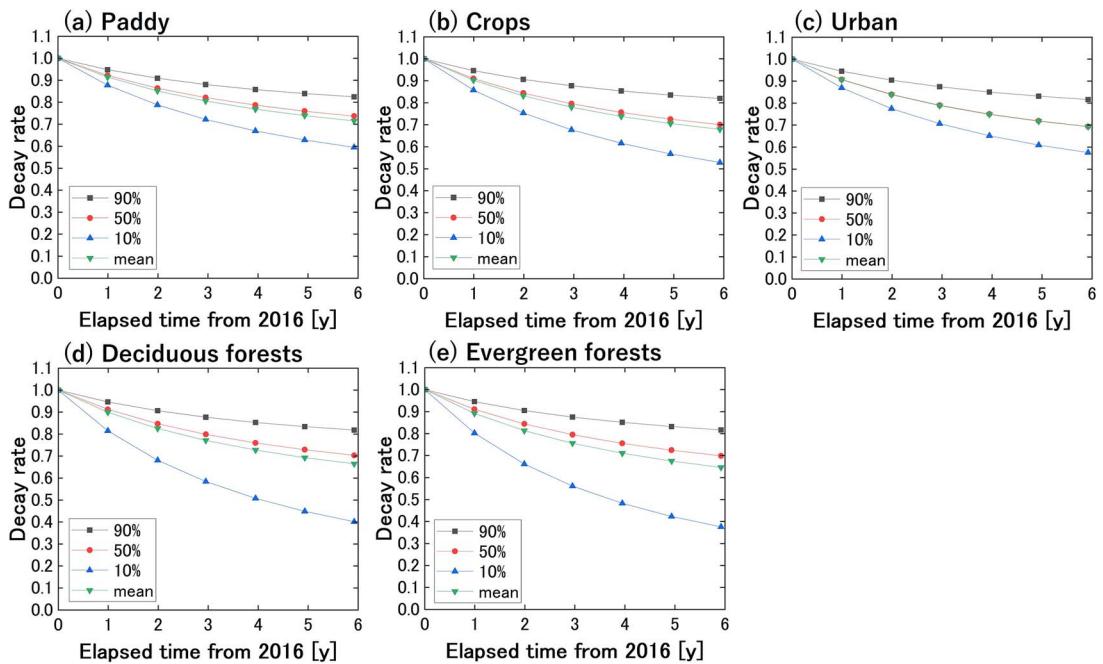
Figure 13. Temporal variations in decay rate (based on 2016) for: (a) paddy fields, (b) croplands, (c) urban areas, (d) deciduous forests, and (e) evergreen forests.

and evergreen forests, the 10th percentile values are particularly low.

A comparison with the integrated map confirms that the model can predict air dose rates up to 2022 with an average deviation of a few tens of percent based on regression analysis and within 10% based on percentage deviation analysis. Overall, the prediction model tends to produce lower values than the integrated map. This discrepancy arises because the model is based on car-borne survey data collected on paved roads, where air dose rates decrease more rapidly. The differences from the integrated map values are larger for forests than for farmlands (paddy and crops) and urban areas. This trend is reasonable, as the reduction in air dose rates is slower in pure forests. Air dose rates measured on roads in forested areas during car-borne surveys differ from those measured in undisturbed forests. In fact, it has been reported that the integrated map was improved by incorporating air dose rates measured in pure forest areas [20].

Figure 17 in Appendix II shows the temporal variation in the decay rate (based on 2016) of predictions, which is largely dependent on the ratio of air dose rates to background radiation at all measurement points under each condition. Since the same tendency is observed for all land uses, four figures for each initial air dose rate category for paddy fields are shown as an example.

Figure 14a–c show air dose rate distribution maps for 2030, 2035, and 2040, calculated using the integrated map for 2022 as the initial condition. The predictions indicate a significant overall decrease in air dose rates. The future air dose rate distribution maps show that contamination will decrease substantially, particularly in the area near 1F and the northwest region, where initial contamination levels were high.

Conclusions

LASSO regression analysis applied to extensive car-borne survey data from Fukushima clarified the features of air dose rate reduction trends following a large-scale nuclear accident. The widely used two-component model adequately approximated the decreasing trends of air dose rates in most cases. However, in some instances, the one-component model provided a better fit. The fast-decay component of the two-component model exhibited a sharp peak in ecological half-life, with the frequency distribution peaking at 0.3–0.4 y. In contrast, approximately half of the slow-decay component displayed a broad half-life peak, represented by multiple exponential functions. The half-life of the slow-decay component varied widely, from several years to several decades. The speed of air dose rate reduction in the years immediately following the accident was strongly influenced by the weight assigned
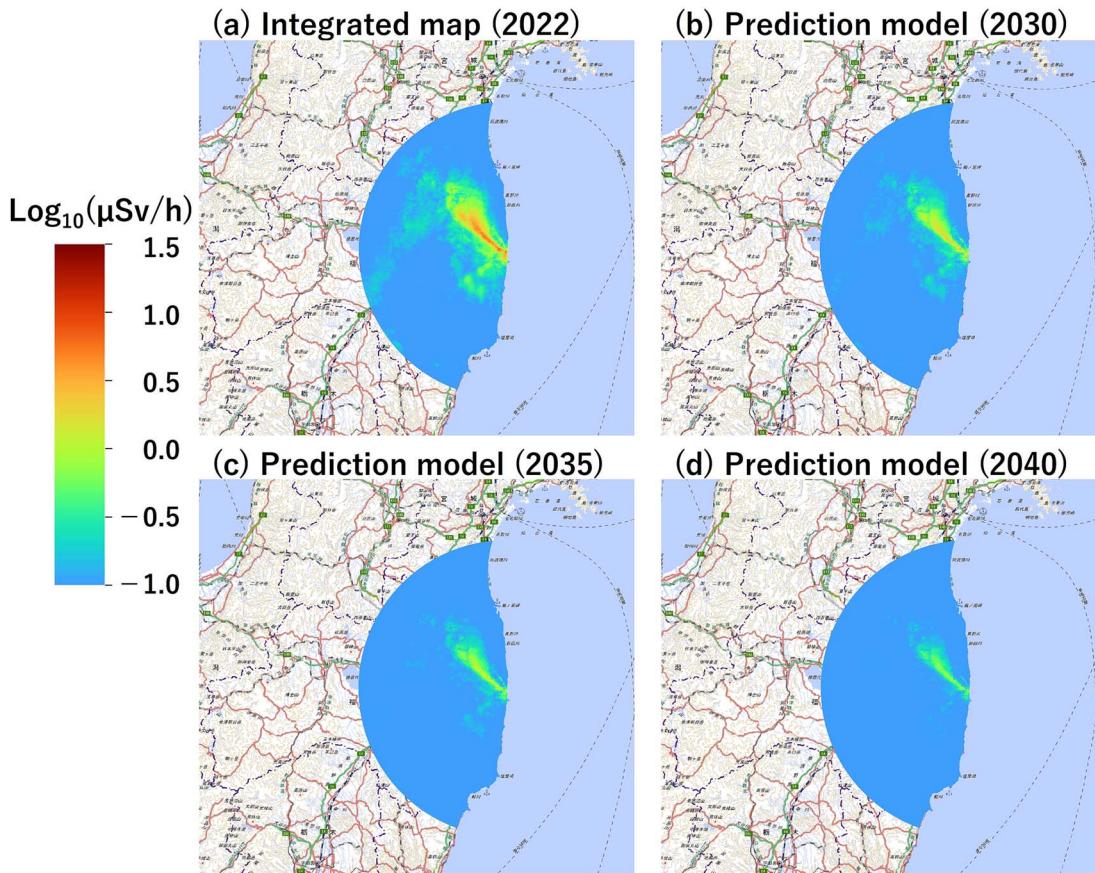
Figure 14. Integrated map (2022) and future predicted maps (2030, 2035, and 2040) of air dose rate distribution within an 80 km radius of 1F: (a) integrated map for 2022, (b) prediction model for 2030, (c) prediction model for 2035, and (d) prediction model for 2040. These maps are based on the Basic map (standard map) published by Geospatial Information Authority of Japan.

to the fast-decay component of the two-component model, rather than the absolute value of its half-life. Air dose rates decreased fastest in urban areas, followed by paddy fields, croplands, deciduous forests, and finally, evergreen forests. This trend can be explained by the migration properties of radiocesium in the environment. Additionally, as initial air dose rates increased, the reduction speed decreased. The findings of this study provide essential insights for modeling the temporal changes in contamination conditions. Future predictions were performed using a model formula that incorporates average ecological half-life profiles calculated for each condition of land use category and initial air dose rate range. Verification of the prediction accuracy through comparison with integrated map data confirmed that the predicted values decayed faster than the integrated map values, with an average deviation within 10% based on the percentage deviation analysis. The variance of the data was generally within a factor of two. Furthermore, the predicted values reflected

the characteristics of the measurement method and land use.

Acknowledgements

Supplementary data

Supplementary data are available at *Radiation Protection Dosimetry online*.

Conflict of interest

None declared.

Funding

References

1. IAEA (International Atomic Energy Agency). *The Fukushima Daiichi Accident*, IAEA, Austria, 2015; ISBN:978-92-0-107015-9.

2. Saito K, Onda Y. Outline of the national mapping projects implemented after the Fukushima accident. *J Environ Radioact* 2015;**139**:240–9. https://doi.org/10.1016/j.jenvrad.2014.10.009.

3. Sanada Y, Sugita T, Nishizawa Y. *et al.* The aerial radiation monitoring in Japan after the Fukushima Daiichi nuclear power plant accident. *J Prog In Nucl Sci Technol* 2014;**4**: 76–80.

4. Fukushima Prefecture. *Fukushima Prefecture Environmental Radiation Monitoring-Mesh Investigation*, 2024; https://www.pref.fukushima.lg.jp/site/portal/ps-monitaring-mesh.html, Accessed data: 6 March 2024.

5. MAFF (Ministry of Agriculture Forestry and Fisheries of Japan). *Monitoring results of ambient dose rate in forest of Fukushima Prefecture (in Japanese)*, 2014; https://www.rinya.maff.go.jp/kanto/press/kikaku/140716.html, Accessed on March 2024.

6. Japan Atomic Energy Agency, *Database for Radioactive Substance Monitoring Data*, 2024; https://emdb.jaea.go.jp/emdb/, Accessed date: 6 March 2024.

7. Seki A, Saito K, Takemiya H. Current status of the environmental monitoring database on the accident at Fukushima Daiichi nuclear power plant. *J Radiol Prot* 2021;**41**:S89–98. https://doi.org/10.1088/1361-6498/abfbc1.

8. Gale HL, Humphreys DLO, Fisher EMR. Weathering of caesium-137 in soil. *Nature* 1964;**201**:257–61. https://doi.org/10.1038/201257a0.

9. Kinase S, Sato S, Takahashi T. *et al.* Ecological half-life of radioactive caesium within the 80 km radius of the Fukushima Daiichi nuclear power plant. *IRPA2014 Abstract Book* 2014;163–6.

10. Kinase S, Takahashi T, Saito K. Long-term prediction of ambient dose equivalent rates after the Fukushima Daiichi nuclear power plant accident. *J Nucl Sci Technol* 2017;**54**: 1345–54. https://doi.org/10.1080/00223131.2017.1365659.

11. Andoh M, Yamamoto H, Kanno T. *et al.* Measurement of air dose rates by walk survey around the Fukushima Daiichi nuclear power plant using KURAMA-II until 2016. *J Environ Radioact* 2018;**190-191**:111–21. https://doi.org/10.1016/j.jenvrad.2018.04.025.

12. Andoh M, Mikami S, Tsuda S. *et al.* Decreasing trend of ambient dose equivalent rates over a wide area in eastern Japan until 2016 evaluated by car-borne surveys using KURAMA systems. *J Environ Radioact* 2018;**192**:385–98. https://doi.org/10.1016/j.jenvrad.2018.07.009.

13. Sanada Y, Urabe Y, Sasaki M. *et al.* Evaluation of ecological half-life of dose rate based on airborne radiation monitoring following the Fukushima Daiichi nuclear plant accident. *J Environ Radioact* 2018;**192**:417–25. https://doi.org/10.1016/j.jenvrad.2018.07.016.

14. Wainwright HM, Seki A, Mikami S. *et al.* Characterizing regional-scale temporal evolution of air dose rates after the Fukushima Daiichi nuclear power plant accident. *J Environ Radioact* 2018;**189**:213–20. https://doi.org/10.1016/j.jenvrad.2018.04.006.

15. UNSCEAR (United Nations Scientific Committee on the Effects of Atomic Radiation). *Levels and effects of radiation exposure due to the nuclear accident after the 2011 great east-Japan earthquake and tsunami. UNSCEAR 2013 Report, Volume I, Scientific Annex A*, New York: United Nations, 2014.

16. UNSCEAR (United Nations Scientific Committee on the Effects of Atomic Radiation). *Levels and effects of radiation exposure due to the accident at the Fukushima Daiichi Nuclear Power Station: implications of information published since the UNSCEAR 2013 Report. UNSCEAR 2020 Report, Scientific Annex B*, New York: United Nations, 2022.

17. Saito K, Mikami S, Andoh M. *et al.* Summary of temporal changes in air dose rates and radionuclide deposition densities in the 80 km zone over five years after the Fukushima nuclear power plant accident. *J Environ Radioact* 2019;**210**:105878. https://doi.org/10.1016/j.jenvrad.2018.12.020.

18. Tibshirani R. Regression shrinkage and selection via the lasso: a retrospective. *J R Stat Soc Series B* 2011;**73**:273–82. https://doi.org/10.1111/j.1467-9868.2011.00771.x.

19. Muthukrishnan R, Rohini R. LASSO: A feature selection technique in predictive modeling for machine learning. *2016 IEEE International Conference on Advances in Computer Applications (ICACA)*, New York: IEEE, 2016, 18–20.

20. Sakura K, Kurikami H, Wainwright H. *et al.* Integrated radiation air dose rate maps over the 80 km radius of the Fukushima Daiichi Nuclear Power Plant and the entire Fukushima Prefecture during 2011–2022. *J Environ Radioact* 2024;**280**:107554. https://doi.org/10.1016/j.jenvrad.2024.107554.

21. Andoh M, Nakahara Y, Tsuda S. *et al.* Measurement of air dose rates in wide area around the Fukushima Daiichi nuclear power plant through a series of car-borne surveys. *J Environ Radioact* 2015;**139**:266–80. https://doi.org/10.1016/j.jenvrad.2014.05.014.

22. Andoh M, Matsuda N, Saito K. Evaluation of ambient dose equivalent rates owing to natural radioactive nuclides in eastern Japan by car-borne surveys using KURAMA-II(in Japanese). *Trans At Energy Soc Jpn* 2017;**16**:63–80.

23. Ministry of the Environment. Review of the behavior of radioactive cesium in soil (in Japanese). In: *Environmental Restoration Review Committee (15th meeting)*, 2015; 2-2:4-5, https://www.env.go.jp/jishin/rmp/conf/15.html, Accessed on 29 May 2025.

24. Imai N, Okai T. *Natural Radiation Map: Addendum to the "Geochemical Map of Japan" (in Japanese)*. Geological Survey of Japan, National Institute of Advanced Industrial Science and Technology, Ibaraki, 2014.

25. Minato S. Distribution of terrestrial gamma-ray dose rates in Japan (in Japanese). *J Earth Sci* 2006;**115**: 87–95.

26. Tanigaki M, Okumura R, Takamiya K. *et al.* Development of a car-borne g-ray survey system, KURAMA. *Nucl Instr Meth Phys Res Sec A* 2013;**726**:162–8. https://doi.org/10.1016/j.nima.2013.05.059.

27. Tanigaki M, Okumura R, Takamiya K. *et al.* Development of KURAMA-II and its operation in Fukushima.

Nucl Instrum Meth Phys Res 2015;**781**:57–64. https://doi.org/10.1016/j.nima.2015.01.086.

28. Tsuda S, Yoshida T, Tsutsumi M. *et al.* Characteristics and verification of a car-borne survey system for dose rates in air: KURAMA-II. *J Environ Radioact* 2015;**139**:260–5. https://doi.org/10.1016/j.jenvrad.2014.02.028.

29. Liu X, Machida M, Saito K. *et al. Investigation on distribution of radioactive substances in Fukushima (4) Study on the ecological half-lives of ambient dose rates using car-borne survey data with fused LASSO algorithm, Fall Meeting, Atomic Energy Society of Japan*. Toyama, Japan, 2019.

30. Saito K, Machida M, Liu X. *et al. Investigation on distribution of radioactive substances in Fukushima (16); Development of air dose rate temporal change model employing machine learning, Fall Meeting, Atomic Energy Society of Japan*. Japan: Hitachi, 2022.

31. Liu X, Machida M, Saito K. *et al. A Study of KURAMA measurement data by sparse modeling, International Topical Workshop on Fukushima Decommissioning Research (FDR 2019)*. Naraha, Japan, 2019.

32. Breiman L. Random forests. *Mach Learn* 2001;**45**:5–32. https://doi.org/10.1023/A:1010933404324.

33. Stekhoven DJ, Bühlmann P. MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics* 2012;**28**:112–8. https://doi.org/10.1093/bioinformatics/btr597.

34. Shi W, Machida M, Yamada S. *et al.* Inverse estimation scheme of radioactive source distributions inside building rooms based on monitoring air dose rates using LASSO: theory and demonstration. *Prog Nucl Energy* 2023;**162**:104792. https://doi.org/10.1016/j.pnucene.2023.104792.

35. JAXA (Japan Aerospace Exploration Agency). *High-Resolution Land-Use and Land-Cover Map. Ver 14.2*, 2014; (provided by JAEA's HP) https://emdb.jaea.go.jp/emdb_old/en/portals/2030101000/, Accessed date: 31 October 2024.

36. Mikami S, Tanaka H, Okuda N. *et al.* In-situ spectrometry of terrestrial gamma rays using portable germanium detectors in area of 80 km radius around the Fukushima Daiichi nuclear power plant(in Japanese). *Trans At Energy Soc Jpn* 2021;**20**:159–78. https://doi.org/10.3327/taesj.J20.010.

37. Saito K, Endo A. III characteristics of environmental gamma rays and the exposure doses. (in Japanese). *Radioisotopes* 2014;**63**:585–602. https://doi.org/10.3769/radioisotopes.63.585.

38. JAXA (Japan Aerospace Exploration Agency). *High-Resolution Land-Use and Land-Cover Map. Ver 21.11*, 2021; (provided by JAEA's HP) https://www.eorc.jaxa.jp/ALOS/en/lulc/data/index.htm, Accessed date: 31 October 2024.

39. Yoshimura K, Saito K, Fujiwara K. Distribution of 137Cs on components in urban area four years after the Fukushima Dai-ichi Nuclear Power Plant accident. *J Environ Radioact* 2017;**178-179**:48–54. https://doi.org/10.1016/j.jenvrad.2017.07.021.

40. Kato K, Onda Y, Saidin ZH. *et al.* Six-year monitoring study of radiocesium transfer in forest environments following the Fukushima nuclear power plant accident. *J Environ Radioact* 2018;**210**:105817. https://doi.org/10.1016/j.jenvrad.2018.09.015.

41. Takahashi J, Onda Y, Hihara D. *et al.* Six-year monitoring of the vertical distribution of radiocesium in three forest soils after the Fukushima Dai-ichi nuclear power plant accident. *J Environ Radioact* 2018;**210**:105811. https://doi.org/10.1016/j.jenvrad.2018.09.009.

42. Funaki H, Yoshimura K, Sakuma K. *et al.* Evaluation of particulate 137Cs discharge from a mountainous forested catchment using reservoir sediments and sinking particles. *J Environ Radioact* 2019;**210**:105814. https://doi.org/10.1016/j.jenvrad.2018.09.012.

43. Kato H, Onda Y, Yamaguchi T. Temporal changes of the ambient dose rate in the forest environments of Fukushima prefecture following the Fukushima reactor accident. *J Environ Radioact* 2018;**93-94**:20–6.