UC Berkeley

Title

Spatial and temporal prediction of radiation dose rates near Fukushima Daiichi Nuclear Power Plant

Permalink

https://escholarship.org/uc/item/5g29p8h9

Authors

Sun, Dajie
Wainwright, Haruko
Suresh, Ishita
et al.

Publication Date

DOI

Copyright Information

Peer reviewed
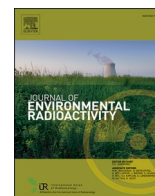
Spatial and temporal prediction of radiation dose rates near Fukushima Daiichi Nuclear Power Plant

Dajie Sun [a,*], Haruko Wainwright [b,c], Ishita Suresh [d], Akiyuki Seki [e], Hiroshi Takemiya [e], Kimiaki Saito [e]

[a] University of California, Berkeley, USA
[b] Lawrence Berkeley National Laboratory, USA
[c] Massachusetts Institute of Technology, USA
[d] Alameda High School/Lawrence Berkeley National Laboratory, USA
[e] Japan Atomic Energy Agency, Tokyo, Japan

ARTICLE INFO

ABSTRACT

In this paper, we have developed a methodology to estimate the spatiotemporal distribution of radiation air dose rates around the Fukushima Daiichi Nuclear Power Plant (FDNPP). In our exploratory data analysis, we found that (1) the temporal evolution of dose rates is composed of a log-linear decay trend and fluctuations of air dose rates that are spatially correlated among adjacent monitoring posts; and (2) the slope of the log-linear environmental decay trend can be represented as a function of the apparent initial dose rates, coordinate position, land-use type, and soil type. From these observations, we first estimated the log-linear decay trend at each location based on these predictors, using the random forest method. We then developed a modified Kalman filter coupled with a Gaussian process model to estimate the dose-rate time series at a given location and time. We applied this method to the Fukushima evacuation zone (as of March 2017), which included 17 monitoring post locations (with monitoring datasets collected between 2014 and 2018) and generated a time series of dose-rate maps. Our results show that this approach allows us to produce accurate spatial and temporal predictions of radiation dose-rate maps using limited spatiotemporal measurements.

1. Introduction

In March 2011, the accident at the Fukushima Daiichi Nuclear Power Plant (FDNPP) resulted in the release of radioactive contaminants to the atmosphere and their deposition in the environment, mostly within the 80 km radius. Radiocesium (^{134}Cs and ^{137}Cs) is considered the main nuclide contributing to the exposure dose in this area (IAEA, 2015; Saito et al., 2015). Multiple agencies have conducted a number of radiation measurements and monitoring campaigns in this region since the accident (e.g., NRA, 2021; Saito and Onda, 2015; 2019; Andoh et al., 2018), which resulted in a large volume of well-archived radiation air-dose-rate data (JAEA, 2021; Seki et al., 2021). These datasets have been analyzed extensively for the purpose of (1) confirming the continuing reduction of contaminant and hazard levels, (2) providing exposure evaluation for the public's return to the evacuation zone, and (3) accumulating the basic datasets for scientific knowledge and future preparation.

In recent years, Wainwright et al. (2017) have developed a data integration method that can integrate multiscale datasets from different types and scales of measurements (airborne, car, and walk surveys), and, from these integrated datasets, generate integrated dose-rate maps. This is based on the Bayesian hierarchical model for combining different datasets and geostatistics for including spatial correlations to constrain the estimation. The integrated dose-rate maps have been able to quantify the temporal changes in air dose rates at the regional scale around the FDNPP, and thereby aid in predicting the spatial distribution of air dose rates in the future (Wainwright et al., 2019). Based on the integrated maps, Sun et al. (2020) have recently proposed an optimization approach for long-term monitoring, selecting a subset of monitoring posts that can capture the spatial heterogeneity of the dose rates more effectively with a reduced number of points.

In parallel, a data-driven environmental decay model has been developed to predict the decay of radiation air dose rates in the environment (Kinase et al., 2014, 2017; Wainwright et al., 2019). Kinase et al. (2017) built a prediction model characterized by ecological
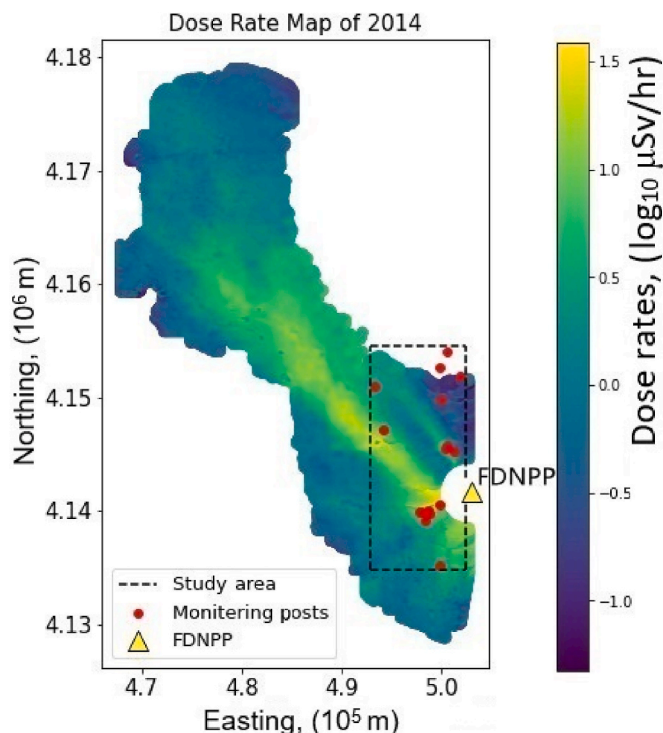

Fig. 1. Diagram of the Study Area. The background is the integrated-dose-rate map of 2014. The color indicates the dose-rate level under logarithm scale based on 10, in log10 μSv/hr. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

half-lives of radiocesium for different land-use types which are parameters in the double exponential formula, and predicted future distribution maps of ambient dose equivalent rates. They also found that dose rates in urban areas were decreasing faster than in cropland or natural forest areas. In parallel, Sanada et al. (2019) evaluated the ecological half-lives of dose-rate reduction in the double exponential formula, based on airborne radiation monitoring. In addition, Andoh et al. (2020) modified the double exponential formula of ecological half-life and developed a two-group model to represent the acceleration of the decrease in the air dose rates observed in the evacuation order area.

There are still challenges in describing the spatially heterogeneous environmental decay over the region. Even though the general tendency toward diminution of air dose rates has been evaluated in the 80 km zone, their trend was found to vary significantly, depending on the location within each land-use type (Saito et al., 2019a, 2019b; Andoh et al., 2018). Although these data-driven models mentioned above can predict the decay trend of the average dose rates within the same land-use category, the detailed spatial-temporal variations have not been considered in previous studies.

The goal of this study is to develop a spatiotemporal data-integration method for creating an integrated radiation air-dose-rate map over space and time. The method is based on the Gaussian process model (GPM), which takes advantage of spatial auto-correlation and covariance models to predict variables in unobserved locations. In order to incorporate the spatial-temporal correlation information of the dose-rate field within the temporal information, we combined GPM and the modified Kalman filter to estimate the spatial-temporal dose-rate time-series at a given time and space.

The Gaussian process model (GPM) is a method of interpolation or prediction, and an extension of geostatistics, assuming random variables with a multivariate normal distribution. By defining the covariance matrix for representing spatial correlations, the GPM can describe the structure of these input variables (such as spatial structure). GPM has a growing popularity in various research fields, particularly in a

geospatial context. For example, Paredes et al. (2021) developed Gaussian process regression models for accurate range and direction estimation of unmanned aerial vehicles. Rong et al. (2019) developed a data-driven nonparametric Bayesian model based on a Gaussian process to describe the lateral motion uncertainty of ship trajectory. For radiation monitoring, Sun et al. (2020) used GPM to interpolate the radiation dose rates and to select monitoring locations to capture the spatial variability across the region.

The Kalman filter (Kalman, 1960) algorithm estimates unknown variables given measurements observed over time. It has been a widely applied concept within many fields, such as time-series analysis in data processing, navigation, and the control of vehicles and spacecraft. Schmidt et al. (2018) presented a Kalman filter-based framework to establish a real-time *in situ* monitoring system for groundwater contamination, based on *in situ* measurable water quality variables, such as specific conductance (SC) and pH.

In our study, we modified the original Kalman filter to accommodate the spatial auto-correlation representing the natural fluctuations and noise. We used this filter to predict the temporal evolution of dose rates for pixels with a real-time monitoring device (to describe the evolution more precisely) or with only a segment of historical data (to fill the gap between segments), and then we applied GPM to interpolate the dose rates for pixels without observations at each time point. This process was repeated for each time point, enabling the construction of a spatial-temporal dose-rate distribution. We demonstrate our approach using the dose-rate datasets from the evacuation zone (as of March 2017).

2. Data description

In this paper, we consider the domain (Fig. 1) corresponding to the integrated dose-rate map of 2014 provided by Wainwright et al. (2017). The size of each grid is 50 m × 50 m, the UTM position of the origin is (460101, 4120001). The dashed box with dimension of 8550 m × 18750 m is the study area, which contains 17 monitoring posts collecting between July 1st, 2017 and Nov 22nd, 2017. We use a section of the time series that has continuous measurements, without snow.

In addition, since the monitoring-post datasets are limited in space, we used the car-survey data (Database for Radioactive Substance Monitoring Data, 2022) from the Tokyo Electric Power Company (TEPCO) to quantify the spatial variability of environmental decay, as well as the correlation between decay rates and initial dose rates. In the car-survey data set, there are 9626 locations, with 13 time-points for each location, starting from March 18, 2014, and ending on April 3, 2018 (see Table S1 in supplementary material for details). The location information of all the car-survey data points can be found in the Results & discussion section (section 4.1). The datasets of soil information, elevation, land-use in the Fukushima region were published by JAEA, and were downloaded from the previous version of Database for Radioactive Substance Monitoring Data (Database for Radioactive Substance Monitoring Data, 2022; Table S2 in supplementary material).

3. Methods

3.1. Environmental decay characterization

In this study, we used the single component decay model with effective half-life in similar ways as Kinase et al. (2014) for describing the environmental decay after March 18th, 2014. Although other papers used the two-component models with ecological half-lives excluding physical decay (Kinase et al., 2017; Sanada et al., 2019; Andoh et al., 2018), the decay can be represented by *effective* half-life based on the log-linear trends after the dose rate reduction according to the ^{134}Cs contribution is reduced (^{134}Cs has a much shorter half-life of 2.06 years compared to ^{137}Cs of 30.2 years). For a single-component exponential decay model, the time series have a linear trend under the logarithm scale, and the slope of this linear trend corresponds to the decay rates,
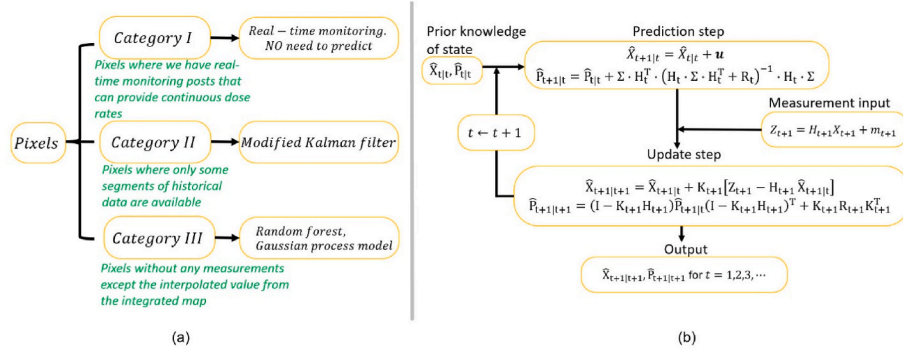
Fig. 2. (a) Pixel's categories and prediction methods for each category; (b) Diagram of modified Kalman filter to predict Category II pixels.

while the intercept of the trend corresponds to the *apparent* initial dose rates at $t = 0$ (i.e., March 11th, 2011), extrapolated from the time after which the single-component exponential decay is defined (i.e., March 18th, 2014). For convenience, we use the terms "slope" and "intercept" in this study, which have the same meaning as "decay rates" and the "initial dose rates", respectively.

We used the random forest method to estimate the slope at unobserved locations. Random Forest is a machine learning method for classification or regression that operates by constructing a large number of decision trees from bootstrapped subsampled data at training time and outputting the mode of the classes (classification) or mean/average prediction (regression) of the individual trees (Tin Kam, 1995). The random forest method works well with correlated predictors similar to ridge regressions (Hastie et al., 2001). We used cross-validation to choose the hyper-parameters.

We investigated the relationship between the slope and intercept, including the background radiation. We decomposed the dose rates into an exponential part and a background part, as shown in Eq. (1):

$$x_t = \log\left(x_0 e^{-\lambda_{eff} t} + h\right) \tag{1}$$

where x_t is the log-dose rate at time t, x_0 is the apparent initial dose rate, λ_{eff} is the effective decay rate, and h is the natural background. The absolute value of the slope in the logarithm scale is computed as Eq. (2):

$$\left|\frac{d}{dt}x_t\right| = \frac{\lambda_{eff} x_0 e^{-\lambda_{eff} t}}{x_0 e^{-\lambda_{eff} t} + h} = \lambda_{eff}\frac{1}{1 + e^{\lambda_{eff} t} h/x_0} \approx \begin{cases} \lambda_{eff}, & e^{\lambda_{eff} t} h/x_0 \approx 0. \quad (a) \\ 0, & e^{\lambda_{eff} t} h/x_0 \gg 1. \quad (b) \end{cases} \tag{2}$$

Specifically, when the dose-rate level is much higher than the background, the influence of the natural background can be ignored, and the slope represents the effective decay rate, as shown in Case (a) of Eq. (2); when the dose rate is not high enough, the existence of the background will pull the slope toward zero, as shown in Case (b) of Eq. (2).

3.2. Spatiotemporal estimation

Our study domain was discretized into the same grid (with grid size 50 m × 50 m) as the integrated map. As mentioned in Section 3.1, we assume that the dose-rate time-series for pixels under logarithm scale can be modeled by a linear trend plus a fluctuation term. At time t, the dose-rate vector for n pixels can be defined as $X_t = \{x_{t,1}, .., x_{t,n}\}$:

$$X_t = \boldsymbol{u}\, t + \boldsymbol{b} + \boldsymbol{\varepsilon}_t \tag{3}$$

where X_t is the n-element vector representing dose rates under logarithm scale, $\boldsymbol{u}$ is a n-element vector that indicates the spatially variable slope, $\boldsymbol{b}$ is the n-element intercept vector, and $\boldsymbol{\varepsilon}_t$ is the n-element vector representing the fluctuation term from the trend. We assume that $\boldsymbol{\varepsilon}_t$ is normally distributed and independent between different time steps, but they are spatially correlated among different pixels at the same time step:

$$cov(\varepsilon_t, \varepsilon_s) = \begin{cases} \Sigma_t, & when\ t = s \\ 0, & otherwise \end{cases} \tag{4}$$

where t and s are time step indices, Σ_t is a $n \times n$ matrix indicating that the fluctuation parts are spatially correlated.

3.2.1. Time-series estimation

All the pixels can be classified into three categories: (I) pixels where we have real-time monitoring posts that can provide continuous dose rates; (II) pixels where only some segments of historical data are available, such the locations where monitoring posts are removed; and (III) pixels without any measurements except the interpolated value from the integrated map. These categories are shown in Fig. 2a. We use n_1, n_2, n_3 to denote the number of pixels in each category, and use vector Z_t (with k_t elements) to represent the observation vector at time step t. In this study, we selected monitoring-post data from JAEA as the observation vector Z_t to illustrate our method. The dose rates for the Category I pixels do not need to be estimated, because they were defined by the real-time monitoring posts. A general spatial-temporal prediction can be decomposed into two steps to estimate the dose rates for the Category II and Category III pixels:

3.2.1.1. Step One: Estimate the air dose rates in the Category II pixels for one time step.
We define a dose-rate vector for the pixels in Categories I and II at time step t as $X_t = (x_{1,t}, x_{2,t}, \cdots, x_{(n_1+n_2),t})^T$, where $x_{i,t}$ is the dose rate for pixel i at time step t and (n_1+n_2) equals the total number of pixels in Categories I and II here. We define the $(n_1+n_2) \times (n_1+n_2)$ covariance matrix of X_t as P_t. The decay model of Eq. (3) can be rewritten as

$$X_{t+1} = X_t + \boldsymbol{u} + w_t \tag{5}$$

where $\boldsymbol{u}$ is a vector of (n_1+n_2) elements, representing the slopes of the linear trend of all the pixels in Categories I and II. $\boldsymbol{u}$ can be fitted from the historical data. w_t is defined as

$$w_t = \varepsilon_t - \varepsilon_{t+1} + \boldsymbol{u} \tag{6}$$

The $(n_1+n_2) \times (n_1+n_2)$ covariance matrix between w_s and w_t is defined as Q_t,

$$Q_t = cov(w_s, w_t) = \begin{cases} 2\Sigma_t, & if\ t = s \\ -\Sigma_t, & if\ |t - s| = 1 \\ 0, & otherwise \end{cases} \tag{7}$$

where matrix Σ_t is the covariance of the noise ε among different pixels with dimension $(n_1+n_2) \times (n_1+n_2)$ at time step t.

For every time step, we will have k_t-elements measurement/observation vector Z_t,
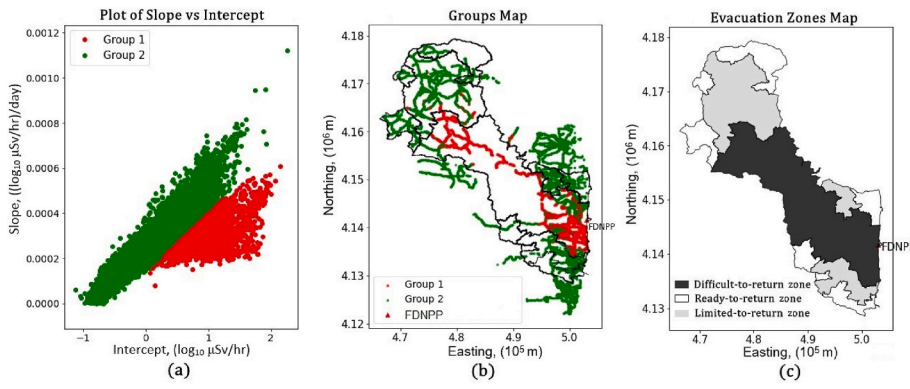
$$Z_t = H_t X_t + m_t \tag{8}$$

Fig. 3. (a) Plot of the slope as a function of the intercept; (b) the two groups in the physical co-ordinates; (c) the evacuation zone map. In (a), Group 1 (red points): data points that are off the linear trend of Slope ~ Intercept; Group 2(green points): data points which maintain a linear trend between slop and intercept. The location map of each group in (a) is shown in (b), and the black solid curve in (b) has the same zone boundaries as (c). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

where

$$cov(m_t, m_s) = \begin{cases} R_t \ if \ t = s \\ 0, \quad otherwise \end{cases} \tag{9}$$

R_t is the covariance matrix of additive measurement noise with dimension of $k_t \times k_t$. H_t is the $k_t \times (n_1 + n_2)$ observation matrix, determined by the data model (i.e., how each type of measurement, airborne, car, and monitoring post, was represented by the dose rates of all pixels and integrated within our method), and m_t is the additive measurement noise (k_t elements vector) whose covariance is R_t.

In order to estimate the system state determined by Eqs. (5)–(9), we developed a modified Kalman filter that consists of iterations of two steps: a prediction step and an update step, as shown in Fig. 2b. The prediction step gives the predicted values in the next time step based on the temporal evolution, while the update step integrates this prediction and data values. (A detailed derivation can be found in the supplementary material S3; only the result is presented here.)

(a) Prediction step:

The prediction step computes the system status at time step $t+1$ based on the predicted system status at time step t.

$$\widehat{X}_{t+1|t} = \widehat{X}_{t|t} + \boldsymbol{u} \tag{10}$$

$$\widehat{P}_{t+1|t} = \widehat{P}_{t|t} + \Sigma_t H_t^T (H_t \Sigma_t H_t^T + R_t)^{-1} H_t \Sigma_t \tag{11}$$

where $\widehat{X}_{t+1|t}$ and $\widehat{P}_{t+1|t}$ represent the prediction of the dose-rate vector and covariance matrix for the time step $(t+1)$ conditioned on the dose rate of the previous time step t.

(b) Update step:

The update step serves to refine the estimation of system status at time step $t+1$ based on the observation of Z_{t+1}.

$$\widehat{X}_{t+1|t+1} = \widehat{X}_{t+1|t} + K_{t+1} [Z_{t+1} - H_{t+1}\widehat{X}_{t+1|t}] \tag{12}$$

$$\widehat{P}_{t+1|t+1} = (I - K_{t+1}H_{t+1})\widehat{P}_{t+1|t}(I - K_{t+1}H_{t+1})^T + K_{t+1}R_{t+1}K_{t+1}^T \tag{13}$$

where

$$K_{t+1} = \widehat{P}_{t+1|t}H_{t+1}^T [H_t\widehat{P}_{t+1|t}H_{t+1}^T + R_{t+1}]^{-1} \tag{14}$$

is a $(n_1 + n_2) \times k_{t+1}$ matrix. The subscripts of $\widehat{P}_{t+1|t+1}$, $\widehat{X}_{t+1|t+1}$ mean the prediction of P, X at time step $t+1$ conditioned on Z_{t+1}.

3.2.1.2. Step Two: Predict dose-rate time-series for the Category III pixels for the same time step as Step One. We added subscript '3' for each variable in Eq. (5) for Category III pixels:

$$X_{3,t} = \boldsymbol{u}_3 t + \boldsymbol{b}_3 + \boldsymbol{\varepsilon}_{3,t} \tag{15}$$

where $\boldsymbol{u}_3 t + \boldsymbol{b}_3$ serves as the linear trend of the dose-rate time-series and $\boldsymbol{\varepsilon}_{3,t}$ consists of the fluctuation part. Here, the dimension of $X_{3,t}$, $\boldsymbol{u}_3$, $\boldsymbol{b}_3$ and $\boldsymbol{\varepsilon}_{3,t}$ are n_3.

In order to predict the dose-rate times for the Category III pixels that have no historical data, the following components are required: (1) the initial values (dose rates at $t = 0$, which is $\mathbf{b}$ in Eq. (15)), (2) the slope of the linear trend ($\boldsymbol{u}$ in Eq. (15)), and (3) the fluctuation part ($\boldsymbol{\varepsilon}_t$ in Eq. (15)). This prediction step consists of three sub-steps to compute the three components:

(1) The apparent initial values of the time series can be selected from the integrated-dose-rate map of 2014 developed by Wainwright et al. (2017, 2019).
(2) The slope of the linear trend can be predicated using the random forest method with the initial values, land-use type, soil type (Table S2), Easting/Northing position (in the UTM system) as predictor variables.
(3) Fluctuations around the trends can be predicted by the dose rates conditioned on the Category I pixels and the predicted values for the Category II pixels at the same time step, using the Gaussian process model (GPM). GPM assumes that the dose-rate values at unmeasured locations follow a multivariate normal distribution, and the unobserved variables can be predicted by conditioning distribution on the observed variables.

Specifically, similar to Eq. (15), where we used the n_3-element vector $\varepsilon_{3,t}$ to denote the fluctuation vector for the pixels from Category III at time step t, we use the $(n_1 + n_2)$-element vector $\varepsilon_{1,t}$ to denote the fluctuation vector for the pixels from Categories I and II at the same time step. Then the best prediction of $\varepsilon_{3,t}$ conditioning on $\varepsilon_{1,t}$ is:

$$\varepsilon_{3,t|1} = \mu_3 + \Sigma_{31}\Sigma_{11}^{-1}(\varepsilon_{1,t} - \mu_1) \tag{16}$$

$$\Sigma_{3|1} = \Sigma_{33} - \Sigma_{31}\Sigma_{11}^{-1}\Sigma_{13} \tag{17}$$

In the above, $\varepsilon_{1,t}$ is a $(n_1 + n_2)$-element vector, $\varepsilon_{3,t|1}$ is a n_3-element vector, μ_1, μ_3 are prior mean vectors, which are 0 here, and covariance matrix was defined as $cov\left(\begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{3,t} \end{pmatrix}, \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{3,t} \end{pmatrix}\right) = \begin{pmatrix} \Sigma_{11} & \Sigma_{13} \\ \Sigma_{31} & \Sigma_{33} \end{pmatrix}$. The dimension of Σ_{33} is $n_3 \times n_3$, the dimension of Σ_{11} is $(n_1 + n_2) \times (n_1 + n_2)$, and the dimension of $\Sigma_{3|1}$ is $n_3 \times n_3$.

By repeating the use of the modified Kalman filter (Eq. (10)-(14)) and GPM (Eq. (16) and (17)), the time series for all the pixels (Category II, Category III) can be reconstructed. Alternatively, we can repeat Step One for all the time steps and then repeat Step Two for all the time steps.
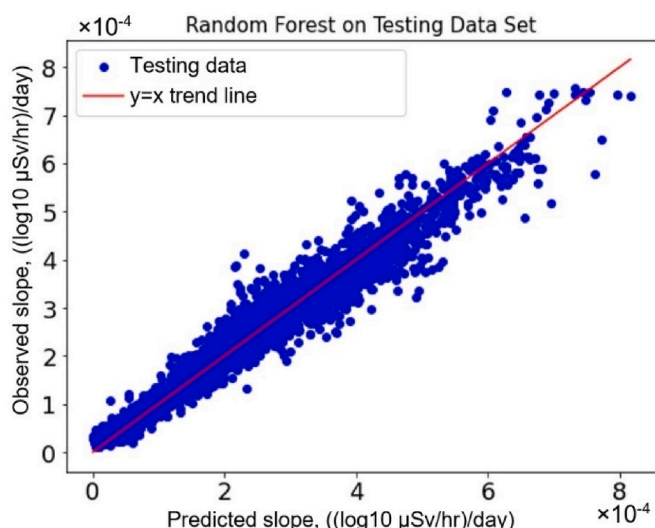
Fig. 4. Slope prediction using Random Forest vs observed value on testing data set. In the Random Forest process, indicators are intercept, land-use type, Easting/Northing position and soil type. Data are from TEPCO car-survey dataset since 2014-01-01.
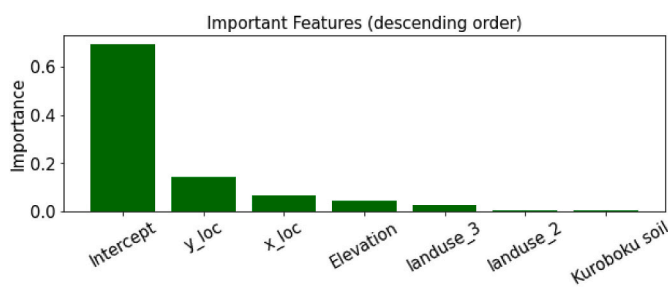

Fig. 5. Feature importance as given by Random Forest. The intercept is the term we used in this study equivalent to initial dose rates; the x_loc, y_loc are the Easting/Northing position in the UTM coordinate system; landuse_3 is land-use type 3, which stands for forest; and landuse_2 means cropland; Kuroboku soil is a soil type.

4. Results & discussion

4.1. Spatial heterogeneity of environmental decay

Based on the car-survey data, we first explored the relationship between the slope (i.e., effective decay rates) and the intercept (i.e., the apparent initial dose rates) in Fig. 3a. We have found the bifurcation or the two groups in this relationship. We then plotted the map of these data points and found that these two groups largely correspond to the different spatial zones (Fig. 3b): (1) the difficult-to-return (or high-risk) zone, and (2) the ready-to-return zone, and limited-to-return zone (Fig. 3c). The difficult-to-return zones have low effective decay rates (i. e., slower reduction), even though the initial dose rates are high. This finding is possibly a result of human activities within the ready-to-return zones that mobilized the soil particles and radiocesium, accelerating the decay. In addition, the difficult-to-return zone is dominated by forests (83% of the area), where the environmental decay is reported to be slower (Kinase et al., 2014; Andoh et al., 2018; Saito et al., 2019a).

The Random Forest regression result is presented in Fig. 4, with the predictive performance of R^2 equal to 0.93, when we applied the regression model to the testing dataset that was not a part of the fitting process. The random forest method also provides a parameter-importance ranking in the regression model, shown in Fig. 5. We found that the apparent initial dose rate is the most influential factor in determining the decay rates (i.e., the slope). Minor contributions are detected from land-use type, spatial coordinates, and soil.

4.2. Spatiotemporal estimation results

4.2.1. Category II pixels estimation result

We selected representative pixels one for the high dose-rate region (pixel 103547) and the other for the low dose-rate region (pixel 103570) to evaluate our method. These two locations are from the testing set (not included in the estimation). The time-series prediction for the two Category II pixels is shown in Fig. 6. The Pearson correlation coefficient R is 0.91 in Fig. 6a, while the Pearson correlation coefficient R is 0.25 in Fig. 6b. Most of the measurements in Fig. 6b still fall into the confidence interval. The predicted dose rates capture the fluctuation part in addition to the log-linear decay trend of dose-rate time-series. In general, the locations with the higher initial dose rate have smaller uncertainty (Fig. 6a), while the ones with the lower initial dose rate have larger uncertainty (Fig. 6b). After looking into the locations of these pixels, we found that the uncertainty is mostly determined by the distance to the closest conditioning pixel.

4.2.2. Category III pixels estimation result

The dose-rate time-series at the Category III pixels are predicted by combining the environmental decay rates (i.e., slope) from the Random Forest and the spatiotemporal fluctuations predicted by GPM (Fig. 7). We test the performance by excluding the data at the two locations (chose to represent low-dose rates and high-dose-rates pixels) during the estimation process as the testing data. Similar to the estimation of the Category II pixels, the Pearson correlation coefficient is higher for the location with the high initial dose rate (Fig. 7b) than the one with the
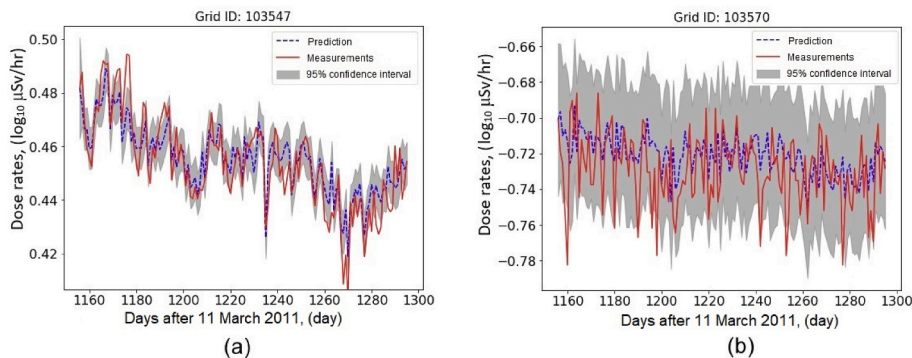

Fig. 6. Comparison of Category II pixels, (a) for pixel 103547; (b) for pixel 103570. In the plots, the red line is for the measured values, the blue dashed line for the predicted values, and the grey band for the 95% confidence interval, some measurements lie beyond the 95% confidence interval due to randomness. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)
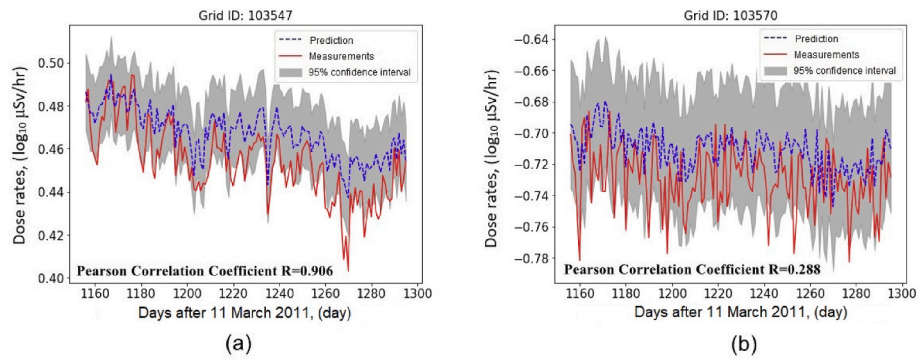
Fig. 7. Prediction vs true value for Category III: (a) for pixel 103570 (representing the low-dose-rate pixels); (b) for pixel 103547 (representing the high-dose-rate pixels). In the plots, red line: Measurements; blue dashed line: Prediction; grey band: 95% confidence interval. The measurement lines were excluded in the estimation. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)
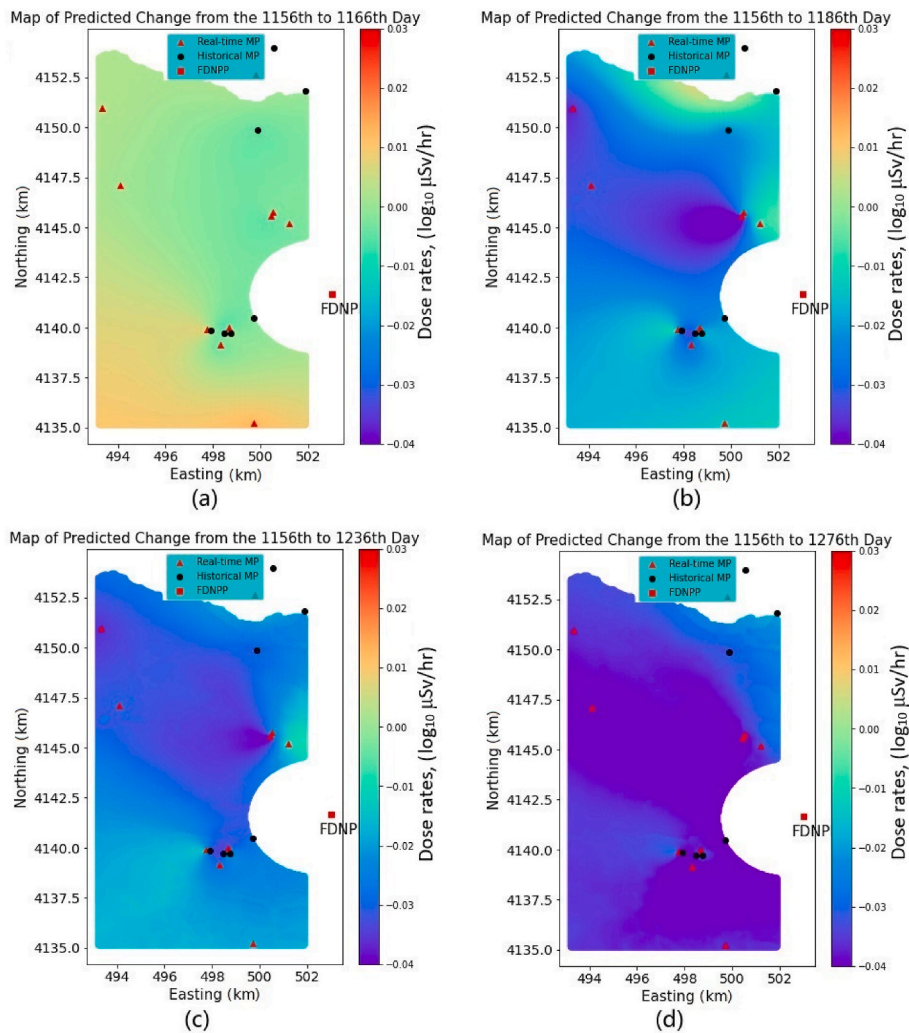

Fig. 8. Predicted dose-rate change for the study area in the log-scale: (a) to (d) the dose rate decrease from the 1156th day to the 1166th, 1186th, 1236th, and 1276th days after the accident. In the color bar, negative values indicate a decrease. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

low initial dose rate (Fig. 7a). In Fig. 7b, the fluctuation is well captured, such as the lower dose rates around Days 1200 and 1250.

Fig. 8 provides the predicted spatiotemporal evolution of the dose rates, which is the change from the apparent initial dose rate on the first day of this time-series (1156th day after the accident) to 10 days, 30 days, 80 days, and 120 days. The color of each pixel represents the

change in dose rates in the log scale, which is heterogeneous across the space. The blue/violet region expands from Fig. 8a–d as the dose rates decrease across the region; this trend is more obvious in the northwest direction, which is assumed to be the high-dose-rate zone. This observation agrees with the conclusion that we draw from Fig. 7—that higher dose rates will decrease at higher rates. In the southwest region, the dose

rates initially increase slightly representing by the red color (Fig. 8a), which is also shown in the data time-series plots between 1156th and 1166th day in Fig. 7. This increase is considered to be the natural fluctuation component, which could be attributed to soil moisture, cosmic rays or the deposition of radon decay daughters during precipitation (Bogena et al., 2015; Barbosa et al., 2018; Mercier et al., 2009).

5. Conclusions

In this work, we investigated the heterogeneity of environmental air-dose-rate decay near Fukushima area based on the random forest method, and developed a methodology by which to estimate the spatial-temporal distribution of the dose rates, using a combination of algorithms based on the Gaussian process models and the modified Kalman filter. We found that a dose-rate time-series can be modeled as a linear decay trend with random fluctuations around the trend, that the linear trend is primarily dependent on the initial dose rate (after 2014), and that the fluctuation of radiation dose rates is spatially correlated among adjacent monitoring posts. We then estimated the environmental decay rate (i.e., the slope of log-linear trend), using the random forest method as a function of the initial dose rate, land-use type, spatial coordinates, and soil as predictors. In addition, we then demonstrated that the modified Kalman filter can be applied to the dose-rate prediction/integration of the pixels with a partial or fully historical data set (Category II pixels), and the fluctuations of the time series from the linear trend at the Category III pixels can be predicted using the Gaussian process model. In summary, our method can successfully estimate air dose rates continuously over time at a given location without monitoring posts. This methodology can be used as a general methodology for improving the existing optimization method (Sun et al., 2020), as well as for dose-rate exposure evaluation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jenvrad.2022.106946.

References

Andoh, M., Mikami, S., Tsuda, S., Yoshida, T., Matsuda, N., Saito, K., 2018. 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. https://doi.org/10.1016/j.jenvrad.2018.07.009.

Andoh, M., Sasaki, M., Saito, K., 2020. Evaluation of decreasing trend in air dose rate and ecological half-life within an 80 km range from Fukushima Dai-ichi Nuclear Power Plant, using car-borne survey data measured by KURAMA systems up to 2018. J. Nucl. Sci. Technol. 57 (12), 1319–1330. https://doi.org/10.1080/00223131.2020.1789008.

Barbosa, S., Huisman, J.A., Azevedo, E.B., 2018. Meteorological and soil surface effects in gamma radiation time series-Implications for assessment of earthquake precursors. J. Environ. Radioact. 195, 72–78. https://doi.org/10.1016/j.jenvrad.2018.09.022.

Bogena, H.R., Huisman, J.A., Güntner, A., Hübner, C., Kusche, J., Jonard, F., et al., 2015. Emerging methods for noninvasive sensing of soil moisture dynamics from field to catchment scale: a review. Wiley Interdisciplinary Reviews: Water 2 (6), 635–647. https://doi.org/10.1002/wat2.1097.

Database for Radioactive Substance Monitoring Data. (n.d.). Retrieved on April 26, 2022, from https://emdb.jaea.go.jp/emdb_old/en/.

Hastie, T., Tibshirani, R., Friedman, J.H., 2001. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, Berlin. https://doi.org/10.1007/978-0-387-84858-7.

Tin Kam, Ho, 1995. Random decision forests (PDF). Proceedings of the 3rd International Conference on Document Analysis and Recognition 278–282. Montreal, QC, 14–16 August 1995, Archived from the original (PDF) on 17 April 2016. Retrieved 26 Dec 2020.

International Atomic Energy Agency, 2015. The Fukushima Daiichi Accident, ISBN 978-92-0-107015-9.

Kalman, R.E., 1960. A New Approach to Linear Filtering and Prediction Problems. https://doi.org/10.1115/1.3662552.

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. Protect. Dosim. 160 (4), 318–321. https://doi.org/10.1093/rpd/ncu014.

Kinase, S., Takahashi, T., Saito, K., 2017. Long-term predictions of ambient dose equivalent rates after the Fukushima Daiichi nuclear power plant accident. J. Nucl. Sci. Technol. 54 (12), 1345–1354. https://doi.org/10.1080/00223131.2017.1365659.

Mercier, J.F., Tracy, B.L., d'Amours, R., Chagnon, F., Hoffman, I., Korpach, E.P., et al., 2009. Increased environmental gamma-ray dose rate during precipitation: a strong correlation with contributing air mass. J. Environ. Radioact. 100 (7), 527–533. https://doi.org/10.1016/j.jenvrad.2009.03.002.

NRA (Nuclear Regulation Authority, 2021. Comprehensive Radiation Monitoring Plan. https://radioactivity.nsr.nsr.go.jp/en/contents/16000/15099/24/274_20210401.pdf.

Paredes, J.A., Álvarez, F.J., Hansard, M., Rajab, K.Z., 2021. A Gaussian process model for UAV localization using millimetre wave radar. Expert Syst. Appl., 115563 https://doi.org/10.1016/j.eswa.2021.115563.

Rong, H., Teixeira, A.P., Soares, C.G., 2019. Ship trajectory uncertainty prediction based on a Gaussian process model. Ocean Eng. 182, 499–511. https://doi.org/10.1016/j.oceaneng.2019.04.024.

Saito, K., Onda, Y., 2015. Outline of the national mapping projects implemented after the Fukushima accident. J. Environ. Radioact. 139, 240–249. https://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., Saegusa, J., Seki, A., Takemiya, H., Shibata, T., 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, 308–319. https://doi.org/10.1016/j.jenvrad.2014.02.014.

Saito, K., Mikami, S., Andoh, M., Matsuda, N., Kinase, S., Tsuda, S., et al., 2019a. 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. 210, 105878 https://doi.org/10.1016/j.jenvrad.2018.12.020.

Saito, K., Mikami, S., Andoh, M., Matsuda, N., Kinase, S., Tsuda, S., et al., 2019b. Temporal change in radiological environments on land after the Fukushima Daiichi nuclear power plant accident. Journal of Radiation Protection and Research 44 (4), 128–148. https://doi.org/10.14407/jrpr.2019.44.4.128.

Sanada, Y., Urabe, Y., Sasaki, M., Ochi, K., Torii, T., 2019. Evaluation of ecological half-life of dose rate based on airborne radiation monitoring following the Fukushima Dai-ichi nuclear power plant accident. J. Environ. Radioact. 210, 105816 https://doi.org/10.1016/j.jenvrad.2018.07.016.

Schmidt, F., Wainwright, H.M., Faybishenko, B., Denham, M., Eddy-Dilek, C., 2018. In situ monitoring of groundwater contamination using the Kalman filter. Environ. Sci. Technol. 52 (13), 7418–7425. https://pubs.acs.org/doi/abs/10.1021/acs.est.8b00017.

Seki, A., Saito, K., Takemiya, H., 2021. Current status of the environmental monitoring database on the accident at Fukushima Daiichi nuclear power plant. J. Radiol. Prot. 41, S89–S98. https://doi.org/10.1088/1361-6498/abfbc1, 2021.

Sun, D., Wainwright, H.M., Oroza, C.A., Seki, A., Mikami, S., Takemiya, H., Saito, K., 2020. Optimizing long-term monitoring of radiation air-dose rates after the Fukushima Daiichi nuclear power plant. J. Environ. Radioact. 220, 106281 https://doi.org/10.1016/j.jenvrad.2020.106281.

Wainwright, H.M., Seki, A., Chen, J., Saito, K., 2017. A multiscale Bayesian data integration approach for mapping air dose rates around the Fukushima Daiichi Nuclear Power Plant. J. Environ. Radioact. 167, 62–69. https://doi.org/10.1016/j.jenvrad.2016.11.033.

Wainwright, H.M., Seki, A., Mikami, S., Saito, K., 2019. Characterizing regional-scale temporal evolution of air dose rates after the Fukushima Daiichi Nuclear Power Plant accident. J. Environ. Radioact. 210, 105808 https://doi.org/10.1016/j.jenvrad.2018.09.006.