A Bayesian Analysis of the Radioactive Releases of Fukushima

Ryota Tomioka
Department of Mathematical Informatics
University of Tokyo

Morten Mørup
Section for Cognitive Systems
Technical University of Denmark

Abstract

The Fukushima Daiichi disaster 11 March, 2011 is considered the largest nuclear accident since the 1986 Chernobyl disaster and has been rated at level 7 on the International Nuclear Event Scale. As different radioactive materials have different effects to human body, it is important to know the types of nuclides and their levels of concentration from the recorded mixture of radiations to take necessary measures. We presently formulate a Bayesian generative model for the data available on radioactive releases from the Fukushima Daiichi disaster across Japan. From the sparsely sampled measurements the model infers what nuclides are present as well as their concentration levels. An important property of the proposed model is that it admits unique recovery of the parameters. On synthetic data we demonstrate that our model is able to infer the underlying components and on data from the Fukushima Daiichi plant we establish that the model is able to account for the data. We further demonstrate how the model extends to include all the available measurements recorded throughout Japan. The model can be considered a first attempt to apply Bayesian learning unsupervised in order to give a more detailed account also of the latent structure present in the data of the Fukushima Daiichi disaster.

1 Introduction

Following the Tohoku earthquake and the 15 meter tsunami wave on 11 March, 2011 the Fukushima Daiichi nuclear disaster was caused by a series of equipment failures. The disaster is considered the largest nuclear accident since the 1986 Chernobyl disaster and has been rated at level 7 on the International Nuclear Event Scale. Radioactive materials were released from the Fukushima containment vessels as a result of accidental or uncontrolled events as well as deliberate venting to reduce gaseous pressure [14]. Very quickly people living in Japan became worried about the various health risks that may result from exposure to excessive amounts of radioactivity. As a result, the Tokyo Electronic Power Company (TEPCO) and authorities started to release radioactive dose measurements across Japan, these measurements are ongoing.

Unfortunately, these measurements do not directly answer what nuclides the measured radioactivity is caused by as the measurements are formed from *mixture of radiations* from various types of radionuclides (unstable atoms that emit radiation), such as, iodine-131 (^{131}I) and cesium-134/137 (^{134}Cs/^{137}Cs) [8, 6, 5]. Since different radionuclides have different effects to human body, it is important to know *which type of radionuclides are located where in what concentrations* in order to take necessary measures.

The standard approach for doing this is either localized in space or time. The in-situ gamma-ray spectrometry [10] can be useful in separating different radionuclides, but requires expensive equipment and long measurement time; therefore it is spatially localized. On the other hand, an airplane or a helicopter can be used to measure the radioactive dose rate of a large area. However, instantaneous dose rates do not directly give concentrations of different types of radionuclides as the measured dose rate is a mixture of various radionuclides. Recently, Yasunari et al. [15] analyzed the radioactivity levels in deposition combined with a atmospheric transport model called FLEXPART. Although this is a promising direction, their focus was to reveal the concentration of a single nuclide ^{137}Cs.

Figure 2 shows some of the measurements from the Fukushima Daiichi nuclear plant for the one month period following the earthquake. From a machine learn-

Appearing in Proceedings of the 15^{th} International Conference on Artificial Intelligence and Statistics (AISTATS) 2012, La Palma, Canary Islands. Volume XX of JMLR: W&CP XX.

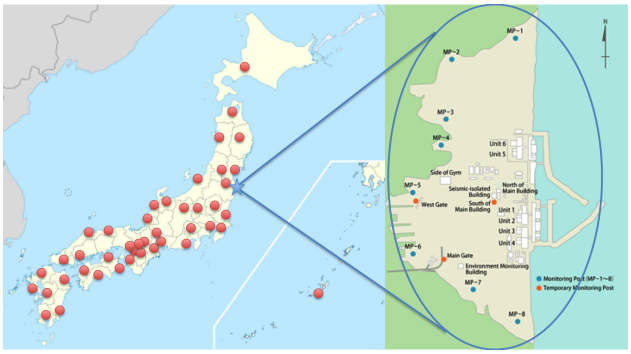


Figure 1: A map showing the 48 regions that we consider in our modeling. The first region consists of nine measurement points at the Fukushima Daiichi nuclear power plant (blue circle). The other 47 regions each consist of one measurement point located at the prefectural capital cities.

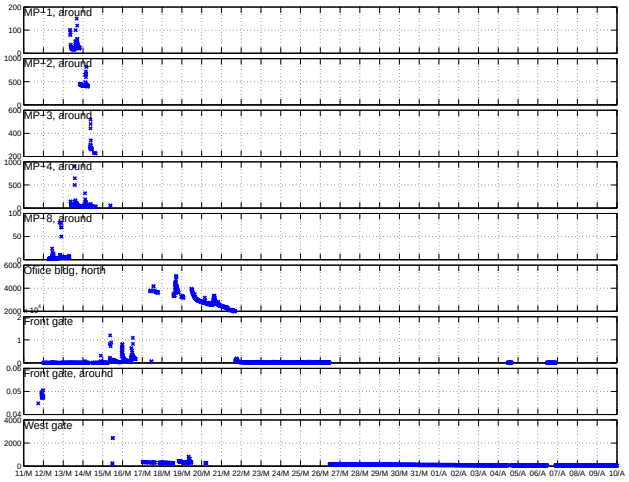


Figure 2: Raw data from the Fukushima Daiichi nuclear plant. The time period from March 11th to April 10th is plotted.

ing point of view, this is a massively missing de-mixing (i.e. blind source separation) problem.

Our goal is to provide a tool that combines both the spatial and temporal dimensions of the data and estimates the types of radionuclides as well as their spatio-temporal distribution. In order to overcome the massively missing measurements, we employ natural but strong modeling assumptions. Our basic assumption is that the radiations from different radionuclides contribute additively to the measured dose rate, and each nuclide has its own decay curve with a fixed but unknown decay constant. As we show, these assumptions are in theory sufficient to make the identification of nuclides and their concentrations possible. Furthermore the proposed model also takes local and global structure of the measurement into account. On the local side, we define groups of nearby measurement points, in which the radionuclide concentrations are assumed to be shared. On the other hand, only the set of decay constants for different radionuclides are shared across groups, which allows us to estimate region specific nuclide concentrations and aggregate samples from different regions to improve the estimation of decay constants.

We hope our modeling framework will turn useful in order to *predict the level of radioactivity between measurements as well as estimate which radionuclides are present in what concentrations.* This will hopefully improve our understanding of the extent of the Fukushima Daiichi disaster and its impact on both nature and humans in the exposed environments.

2 Methods

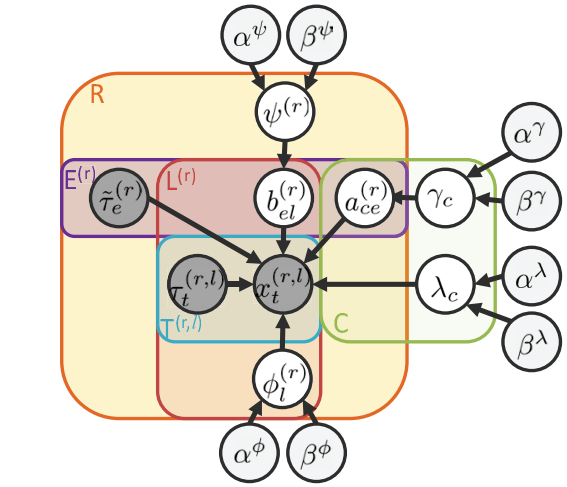
We will consider a total of R different regions indexed by r. Within each region there are $L^{(r)}$ measurement locations indexed by l and a total of $E^{(r)}$ radioactive incidence events indexed by e each occurring at time $\tilde{\tau}_e^{(r)}$. We will assume the events in all regions are known a priori or can be directly inferred from the observed data defined by a significant increase in the radioactive level between samples. We will consider a total of C radioactive nuclides indexed by c. The Fukushima data is sampled both inhomogeneous and sparsely across regions and locations. For each location and region there is a total of $T^{(r,l)}$ observed measurements over time indexed by t such that the tth observation of the radiation level in region r at location l denoted $x_t^{(r,l)}$ occurs at time $\tau_t^{(r,l)}$. We will let $\mathcal{N}_+$ denote the normal distribution truncated to the non-negative orthant and $\exp_+(-\tau) = \begin{cases} \exp(-\tau) & \text{if } \tau \geq 0 \\ 0 & \text{otherwise} \end{cases}$ be the exponential function truncated to zero such that radioactive events can not influence the radioactivity level back in time.

2.1 A generative model of the Fukushima data

For each region r we are provided with the following measurements $\{\boldsymbol{x}^{(r,1)}, \boldsymbol{x}^{(r,2)}, \dots \boldsymbol{x}^{(r,L^{(r)})}\}$, $\{\boldsymbol{\tau}^{(r,1)}, \boldsymbol{\tau}^{(r,2)}, \dots \boldsymbol{\tau}^{(r,L^{(r)})}\}$ as well as the time of the radioactive incidence events $\tilde{\boldsymbol{\tau}}^{(r)}$. For the measured radioactivity level $x_t^{(r,l)}$, in region r at location l at

Draw nuclide decay constants	λ_c	$\sim Gamma(\alpha^\lambda, \beta^\lambda)$
Draw variance of nuclide concentrations	γ_c	$\sim InvGamma(\alpha^\gamma, \beta^\gamma)$
Draw region and event specific nuclide concentrations	$a_{ce}^{(r)}$	$\sim \mathcal{N}_+(0, \gamma_c)$
Draw region specific variance of concentrations	$\psi^{(r)}$	$\sim InvGamma(\alpha^\psi, \beta^\psi)$
Draw region, location and event specific concentrations	$b_{el}^{(r)}$	$\sim \mathcal{N}_+(0, \psi^{(r)})$
Define the average radioactive activity level as	$\mu_t^{(r,l)}$	$= \sum_{e,c} a_{ce}^{(r)} b_{el}^{(r)} \exp_+(-\lambda_c(\tau_t^{(r,l)} - \tilde{\tau}_e^{(r)}))$
Draw region and location specific noise level	$\phi_l^{(r)}$	$\sim InvGamma(\alpha^\phi, \beta^\phi)$
Draw the observed radioactive level	$x_t^{(r,l)}$	$\sim \mathcal{N}(\mu_t^{(r,l)}, \phi_l^{(r)})$

Table 1: Generative process for the Fukushima Daiichi data.



$$x_t^{(r,l)} \approx \mu_t^{(r,l)} = \sum_{e,c} a_{ce}^{(r)} b_{el}^{(r)} \exp_+(-\lambda_c(\tau_t^{(r,l)} - \tilde{\tau}_e^{(r)}))$$

Figure 3: Illustration of the generative model for the Fukushima data. Light gray circles denote predefined priors, and dark shaded circles the observed data.

time $\tau_t^{(r,l)}$ we define the generative process given in Table 1. A graphical model of the generative process is given in Figure 3 and examples of data generated according to the process is illustrated in Figure 4.

We are particularly interested in establishing what nuclides are in the data as well as their concentration level. We therefore treat the decay constants λ_c as stochastic variables in order to infer which nuclides account for the data unsupervised. We further specify in the generative model the variance of each nuclide concentration γ_c separately. Inference on the prior variances typically results in a sparse solution, i.e., most prior variances tend to zero. This process, known as automatic relevance determination [1], enables us to effectively estimate the number of nuclides. Notice also, that across regions the prior over nuclide concentrations and decay constants are shared, hence, these

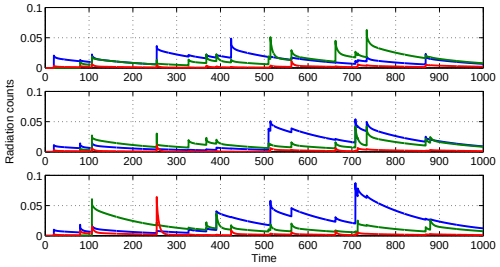


Figure 4: Three random draws from the generative model with $R = 1$, $L = 3$, and $C = 5$. The time of events $\tilde{\tau}_e$ are sampled from a poisson point process with the average waiting time 50.

parameters are supported by all the available data.

2.2 Model limitations

Let $z_{lec}^{(r)}$ denote the concentration of the radioactive nuclide c at event e and location l in region r. The model proposed assumes the following decomposition $z_{lec}^{(r)} = b_{el}^{(r)} a_{ce}^{(r)}$. That is, the relative concentrations of each nuclide for each event is the same within a region given by $a_{ce}^{(r)}$ but can differ in proportion across locations by $b_{el}^{(r)}$. This assumption significantly reduces the number of model parameters from $L^{(r)}E^{(r)}C$ to $E^{(r)}(C + L^{(r)})$.

The data is based on count data which can be assumed to be distributed according to a Poisson distribution. For computational convenience we use the Normal distribution instead as likelihood function for the data. We note however that for sufficiently large values of count occurrences ($x_t^{(r,l)} > 10^3$) the Poisson distribution can well be approximated by a Normal distribution.

In the generative model the timing of each radioactive incidence events $\tilde{\tau}_e^{(r)}$ is assumed to be known a priori as these events are available from official records and can also be well estimated from the data, i.e. if $x_{t-1}^{(r,l)} < x_t^{(r,l)}$ we will create the event e given by $\tilde{\tau}_e^{(r)} = \tau_t^{(r,l)}$.

Defining events from data has however its limitations. The actual timing of the event can occur anywhere between $\tau_{t-1}^{(r,l)}$ and $\tau_t^{(r,l)}$ but from a modeling point of view a change in the timing of the event from $\tilde{\tau}_e^{(r)}$ to $\hat{\tau}_e^{(r)} \in (\tau_{t-1}^{(r,l)}, \tau_t^{(r,l)})$ can be absorbed by an equivalent change in the concentration level of the nuclides $a_{ce}^{(r)}$ to $\hat{a}_{ce}^{(r)}$ such that $a_{ce}^{(r)} \exp_+(-\lambda_c(\tau_t^{(r,l)} - \tilde{\tau}_e^{(r)})) = \hat{a}_{ce}^{(r)} \exp_+(-\lambda_c(\tau_t^{(r,l)} - \hat{\tau}_e^{(r)}))$. Furthermore, if multiple events occur between $\tau_{t-1}^{(r,l)}$ and $\tau_t^{(r,l)}$ these events can similarly be collapsed into a single event that combines the contributions to the radioactivity of the multiple events. Thus, the timing of the events and number of events detected is limited to the temporal resolution of the data. However, as events are shared across locations events only need to be detected within one of the measured locations within a region.

The model does not account for potential correlations across regions for instance caused by weather patterns or regions relative proximity to each other. It is however very challenging to share information across regions as the correlation structure is non-stationary, i.e. weather patterns can change and the spatial propagation can be nuclide specific and non-linear.

Finally, nuclide decay constants for which λ_c is very small (i.e. in the order of years) can be difficult to detect as this requires samples that are separated far in time. Furthermore, nuclides that have similar decay constants can be difficult to separate. The model is also unable to determine how the nuclides are bound to the measurement environment and there can be issues of measurements being inconsistently recorded. This is not accounted for in the model beyond the region and location specific noise variance parameter $\phi_l^{(r)}$.

2.3 Inference

Let $\boldsymbol{\theta} = \{\alpha^\lambda, \beta^\lambda, \alpha^\psi, \beta^\psi, \alpha^\phi, \beta^\phi\}$ denote fixed hyperparameters. Let $\mathcal{X}$ denote the complete available data. From the generative model we can define the following joint distribution

$$
\begin{aligned}
P(\mathcal{X}, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\psi}^{(1:R)}, \boldsymbol{\phi}^{(1:R)}, \boldsymbol{A}^{(1:R)}, \boldsymbol{B}^{(1:R)} | \boldsymbol{\theta}) = \\
\Bigg(\prod_r \Bigg[\prod_{l,t} p(x_t^{(r,l)} | \mu_t^{(r,l)}, \phi_l^{(r)}) \Bigg] \cdot \Bigg[\prod_{l,e} p(b_{el}^{(r)} | 0, \psi_e^{(r)}) \Bigg] \\
\cdot \Bigg[\prod_{e,c} p(a_{ce}^{(r)} | 0, \gamma_c) \Bigg] \cdot \Bigg[\prod_l p(\phi_l^{(r)} | \alpha^\phi, \beta^\phi) \Bigg] \\
\cdot p(\psi^{(r)} | \alpha^\psi, \beta^\psi) \Bigg) \cdot \Bigg[\prod_c p(\lambda_c | \alpha^\lambda, \beta^\lambda) \cdot p(\gamma_c | \alpha^\alpha, \beta^\alpha) \Bigg].
\end{aligned}
$$

We will use Markov chain Monte Carlo (MCMC) to estimate the model parameters from the above joint posterior. We will exploit conjugacy of the Gamma and the truncated normal distribution in order to derive Gibbs updates for γ_c, $\psi^{(r)}$, $\phi_l^{(r)}$, $a_{ec}^{(r)}$, $b_{le}^{(r)}$. These posterior distributions are given by

$$
\begin{aligned}
\gamma_c &\sim InvGamma(\alpha^\gamma + \sum_{r,l} \frac{E^{(r)}}{2}, \beta^\gamma + \sum_{r,e} \frac{(a_{ce}^{(r)})^2}{2}), \\
\psi^{(r)} &\sim InvGamma(\alpha^\psi + \frac{E^{(r)} L^{(r)}}{2}, \beta^\psi + \frac{|\boldsymbol{B}^{(r)}|_F^2}{2}), \\
\phi_l^{(r)} &\sim InvGamma(\alpha^\phi + \frac{T^{(r,l)}}{2}, \beta^\phi + \sum_t \frac{(x_t^{(r,l)} - \mu_t^{(r,l)})^2}{2}), \\
\boldsymbol{a}_e^{(r)} &\sim \mathcal{N}_+(\mu^{\boldsymbol{a}_e^{(r)}}, \boldsymbol{\Sigma}^{\boldsymbol{a}_e^{(r)}}), \\
\boldsymbol{b}_l^{(r)} &\sim \mathcal{N}_+(\mu^{\boldsymbol{b}_l^{(r)}}, \boldsymbol{\Sigma}^{\boldsymbol{b}_l^{(r)}}), \quad \text{where} \\
\boldsymbol{\Sigma}^{\boldsymbol{a}_e^{(r)}} &= \left(\left(\sum_l \tfrac{1}{\phi_l^{(r)}} \boldsymbol{P}^{(r,e,l)\top} \boldsymbol{P}^{(r,e,l)} \right) + \mathrm{diag}(\boldsymbol{\gamma}_c^{-1}) \right)^{-1}, \\
\boldsymbol{\mu}^{\boldsymbol{a}_e^{(r)}} &= \boldsymbol{\Sigma}^{\boldsymbol{a}_e^{(r)}} (\sum_l \boldsymbol{P}^{(r,e,l)} \boldsymbol{\xi}^{(r,e,l)} / \phi_l^{(r)}), \\
\boldsymbol{\Sigma}^{\boldsymbol{b}_l^{(r)}} &= \left(\frac{1}{\phi_l^{(r)}} \boldsymbol{R}^{(r,l)\top} \boldsymbol{R}^{(r,l)} + (1/\psi^{(r)}) \boldsymbol{I} \right)^{-1}, \\
\boldsymbol{\mu}^{\boldsymbol{b}_l^{(r)}} &= \boldsymbol{\Sigma}^{\boldsymbol{b}_l^{(r)}} (\boldsymbol{R}^{(r,l)\top} \boldsymbol{x}_l^{(r)} / \phi_l^{(r)}).
\end{aligned}
$$

In the expressions above the $T^{(r,l)}$ dimensional vector $\boldsymbol{\xi}^{(r,l)}$, the $T^{(r,l)} \times C$ matrix $\boldsymbol{P}^{(r,e,l)}$, and the $T^{(r,l)} \times E^{(r)}$ matrix $\boldsymbol{R}^{(r,l)}$ are defined as follows:

$$
\begin{aligned}
\boldsymbol{\xi}^{(r,e,l)} &= \Big(x_t^{(r)} - \sum_{e' \neq e} \sum_c a_{ce'}^{(r)} b_{e'l}^{(r)} \mathrm{e}_+^{-\lambda_c(\tau_t^{(r,l)} - \tilde{\tau}_{e'}^{(r)})} \Big)_t, \\
\boldsymbol{P}^{(r,e,l)} &= \Big(b_{el}^{(r)} \exp_+(-\lambda_c(\tau_t^{(r,l)} - \tilde{\tau}_e^{(r)})) \Big)_{t,c}, \\
\boldsymbol{R}^{(r,l)} &= \Big(\sum_c a_{ce}^{(r)} \exp_+(-\lambda_c(\tau_t^{(r,l)} - \tilde{\tau}_e^{(r)})) \Big)_{t,e}.
\end{aligned}
$$

In order to sample from $\mathcal{N}_+$ we used the slice-sampler proposed in [12].

Since λ_c is constrained to be non-negative we will re-parametrize $\lambda_c = \exp(\upsilon_c)$ and sample υ_c by a random-walk using a normal distribution as proposal distribution with mean given by the current value of υ_c and variance defined by the inverse Hessian of the re-parametrized negative log-likelihood function, i.e. $Q(\upsilon_c^* | \upsilon) \sim \mathcal{N}(\upsilon_c, (-\frac{\partial^2 \log P(\upsilon_c | \boldsymbol{\theta}_{\setminus \upsilon_c})}{\partial \upsilon_c^2})^{-1})$. For the log-likelihood function and the Hessian of the re-parametrized problem we have

$$
\begin{aligned}
-\log P(\upsilon_c | \boldsymbol{\Theta}_{\setminus \upsilon_c}) =& (\alpha^\lambda - 1)\upsilon_c - \exp(\upsilon_c)/\beta^\lambda \\
& - \sum_r \frac{\sum_{t,l} (x_{tl}^{(r)} - \mu_{tl}^{(r)})^2}{2\phi_l^{(r)}} + \text{const.}, \\
-\frac{\partial^2 \log P(\upsilon_c | \boldsymbol{\Theta}_{\setminus \upsilon_c})}{\partial \upsilon_c^2} =& \exp(\upsilon_c)/\beta^\lambda \\
& - \Bigg[\sum_r \sum_{lt} (\sum_e q_{tec}^{(r,l)} \lambda_c (\tau_t^{(r,l)} - \tilde{\tau}_e^{(r)})
\end{aligned}
$$

$$\cdot(1+\lambda_c(\tau_t^{(r,l)}-\tilde{\tau}_e^{(r)})))(x_{lt}^{(r)}-\mu_{lt}^{(r)})^2/\phi_l^{(r)}\Bigg]$$

$$+\left[\sum_r\sum_{lte}(q_{tec}^{(r,l)}\lambda_c(\tau_t^{(r,l)}-\tilde{\tau}_e^{(r)}))^2/\phi_l^{(r)}\right],$$

where $\boldsymbol{\Theta}_{\setminus v_c}$ is used to denote all parameters of the model except v_c and $q_{tec}^{(r,l)}=a_{ce}^{(r)}b_{el}^{(r)}\exp_+(-\lambda_c(\tau_t^{(r,l)}-\tilde{\tau}_e^{(r)}))$. The accept rate is given by the Metropolis-Hastings ratio $\min\{1,\frac{P(v_c^*|\boldsymbol{\theta}_{\setminus v_c^*})Q(v_c|v_c^*)}{P(v_c|\boldsymbol{\theta}_{\setminus v_c})Q(v_c^*|v_c)}\}$.

Accounting for background radiation: We include a location and region dependent bias term denoting the level of background radiation. This is trivially implemented by introducing an additional (background) event ($e=0$) with $\tilde{\tau}_0^{(r)}=0$ and additional (background) component ($c=0$) defined by $\lambda_0=0$ such that $a_{0e}^{(r)}=\begin{cases}1 & \text{if } e=0\\ 0 & \text{otherwise}\end{cases}$. The background radiation level is estimated by sampling the parameter $b_{l0}^{(r)}$ in the Gibbs update of $\boldsymbol{b}_l^{(r)}$.

2.4 Parameter Identifiability

We will in this section derive two important properties of the proposed model; the components are unique up to permutation and scaling and the initial relative nuclide concentrations can be estimated from $\boldsymbol{A}^{(r)}$.

2.4.1 Identifiability of the model

The estimated value of the radiation level given by $\mu_t^{(r,l)}=\sum_{e,c}a_{ce}^{(r)}b_{el}^{(r)}\exp_+(-\lambda_c(\tau_t^{(r,l)}-\tilde{\tau}_e^{(r)}))$ can be considered a non-linear factor analytic type decomposition into the nuclide concentration matrix $\boldsymbol{A}^{(r)}$, location specific concentrations $\boldsymbol{B}^{(r)}$ and a third order tensor of time courses with elements given by $v_{tec}=\exp_+(-\lambda_c(\tau_t^{(r,l)}-\tilde{\tau}_e^{(r)}))$.

Now, for simplicity omitting the index r for a region, we can write the reconstructed data for the lth location as a simple linear system as follows: $\boldsymbol{\mu}^{(l)}=\sum_{c=1}^C\boldsymbol{V}_c\,\mathrm{diag}(\boldsymbol{a}_c)\boldsymbol{b}^{(l)}$. If we define an alternative solution given by an invertible matrix $\boldsymbol{M}$ we have $\boldsymbol{\mu}^{(l)}=\sum_{c=1}^C\tilde{\boldsymbol{V}}_c\,\mathrm{diag}(\tilde{\boldsymbol{a}}_c)\tilde{\boldsymbol{b}}^{(l)}$, where

$$\sum_{c=1}^C\tilde{\boldsymbol{V}}_c\,\mathrm{diag}(\tilde{\boldsymbol{a}}_c)=\sum_{c=1}^C\boldsymbol{V}_c\,\mathrm{diag}(\boldsymbol{a}_c)\boldsymbol{M},\ \tilde{\boldsymbol{b}}^{(l)}=\boldsymbol{M}^{-1}\boldsymbol{b}^{(l)}.$$

Since there is no restriction on $\tilde{\boldsymbol{b}}^{(l)}$, the above second equation is automatically satisfied for any $\boldsymbol{M}$. However, we can see that in the first equation, we have ET equations but only EC degrees of freedom for $\tilde{\boldsymbol{A}}$, C degrees of freedom for $\tilde{\lambda}_c$ ($c=1,\ldots,C$), and E^2 degrees of freedom for $\boldsymbol{M}$. Therefore, if $ET>EC+C+E^2$, it is unlikely that the solution of the nonlinear equation exists.

2.5 Estimating the nuclide concentrations from $\boldsymbol{A}^{(r)}$

Consider a single event at $\tilde{\tau}=0$ such that the observed radiation dose x_t is composed of C nuclides decaying at different decay constants λ_c ($c=1,\ldots,C$). Denote the concentration of the cth radionuclide at time t as $z_c(\tau_t)$. The observed radiation dose x_t can be written as $x_t=\sum_c\chi_c z_c(\tau_t)$, where χ_c is a nuclide specific constant that can be found in physics literature.

Note that although our model predicts as $x_t=\sum_c a_c e^{-\lambda_c\tau_t}$, ignoring the indices for region and location, the latent nuclide concentration $z_c(\tau_t)$ does *not* simply decay as $z_c(\tau)=z_c(0)e^{-\lambda_c\tau}$, in which case we would have $a_c=\chi_c z_c(0)$. This is because some radionuclides decay into other nuclides.

However, if we know additionally which nuclides decay into which other nuclides, we can recover the initial nuclide concentration $z_c(0)$ from the estimated "effective" nuclide concentration $a_1,\ldots,a_C$. This is possible because the decay process can be described by the linear differential equation $\dot{\boldsymbol{z}}(\tau)=\boldsymbol{M}\boldsymbol{z}(\tau)$, where $\boldsymbol{z}(\tau)=(z_1(\tau),\ldots,z_c(\tau))^\top$ and $\boldsymbol{M}$ is an upper triangular matrix that describes the decay chain. For example if the jth nuclide decays into the ith nuclide ($i<j$) at rate λ, the jth column vector of $\boldsymbol{M}$ has λ and $-\lambda$ in the ith and the jth element, respectively. The matrix $\boldsymbol{M}$ is upper triangular because the decay chain is a DAG. By assuming that the decay constants are distinct, the characteristic polynomial of $\boldsymbol{M}$ has C distinct roots and $\boldsymbol{M}$ can be diagonalized by a regular matrix $\boldsymbol{P}$. In this case, the solution of the above differential equation can be written as follows:

$$\boldsymbol{z}(\tau)=\boldsymbol{P}\,\mathrm{diag}\left(e^{-\lambda_1\tau},\ldots,e^{-\lambda_C\tau}\right)\boldsymbol{P}^{-1}\boldsymbol{z}(0).$$

Thus the estimated coefficient a_c can be written as $a_c=\tilde{\chi}_c\tilde{z}_c(0)$, where $\tilde{\boldsymbol{\chi}}=\boldsymbol{P}^\top\boldsymbol{\chi}$ and $\tilde{\boldsymbol{z}}(0)=\boldsymbol{P}^{-1}\boldsymbol{z}(0)$. Since $\boldsymbol{\chi}$ and $\boldsymbol{P}$ are known, we can recover the initial nuclide concentration $\boldsymbol{z}(0)$ from $a_1,\ldots,a_C$.

3 Results

In this section we first demonstrate that our sampling method described in Section 2 can recover global radiation dose from largely missing observations on synthetic data. Next, we apply our method to the survey meter recordings from Fukushima Daiichi nuclear power plant. Finally we apply our method to combine the data from the Fukushima Daiichi power plant and regional measurements across Japan.

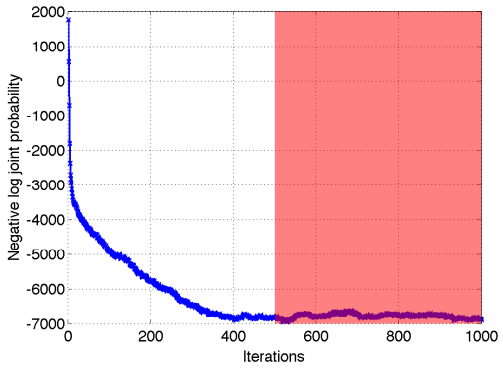


Figure 5: The negative log joint probability function is plotted against the number of MCMC iterations. The shaded area from 500 to 1000 iterations are used to compute the posterior mean for the missing elements.

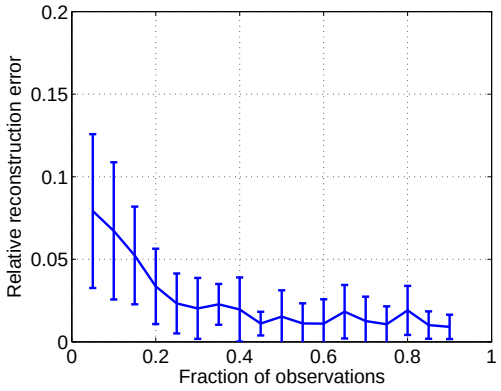


Figure 6: The relative reconstruction error (1) is plotted against the fraction of observed elements.

3.1 Synthetic data

We generate synthetic hourly data corresponding to a period of one month with five locations ($L = 5$)[1]. We assume three underlying radionuclides having fixed decay constants corresponding to half lives 8.8, 0.88, and 0.088 days.

We split the generated data into training and test sets by randomly holding out certain fraction of data points as missing. We run the sampler described in the previous section only with the training set and report the reconstruction error on the test set.

The following error metric is used to measure the mean squared error of the posterior prediction relative to the signal power:

$$\text{relerr} = \frac{\sum_l \frac{1}{T^{(l,\text{test})}} \sum_{t \in \mathcal{T}^{(l,\text{test})}} (x_t^{(l)} - \langle \mu_t^{(l)} \rangle)^2}{\sum_l \frac{1}{T^{(l)}} \sum_t (x_t^{(l)})^2}, \quad (1)$$

where $\mathcal{T}^{(l,\text{test})}$ is the set of indices corresponding to the missing data and $T^{(l,\text{test})} = |\mathcal{T}^{(l,\text{test})}|$ is the number of missing data for the lth location. The posterior average $\langle \mu_t^{(l)} \rangle$ is computed by sampling for a certain number of iterations after the burn-in period.

Figure 5 shows the time course of the negative log joint probability (i.e., $-\log(P(\mathcal{X}, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\psi}^{(1:R)}, \boldsymbol{\phi}^{(1:R)}, \boldsymbol{A}^{(1:R)}, \boldsymbol{B}^{(1:R)} | \boldsymbol{\theta}))$) against the sampling steps. From the plot, we can see that the negative log probability decreases until 400 iterations and become more flat after that point. Therefore, we used the samples from 500 to 1000 iterations to compute the posterior mean prediction.

[1] In the simulation we only have a single region, i.e. $R = 1$ and the r index is therefore ignored.

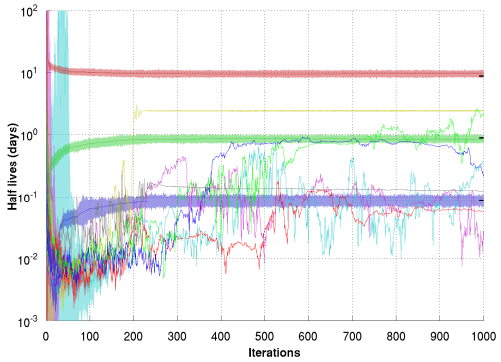


Figure 7: The half lives of the radionuclides estimated by the proposed sampler on the synthetic data. The sampler assumed ten nuclide candidates ($C = 10$), while the data was generated with three nuclides with half lives 8.8, 0.88, and 0.088 (days) indicated by the short black lines in the right side of the plot.

Figure 7 shows the half lives of the estimated radionuclides against the number of iterations. The decay constant λ_c is converted into the half life by $t_{\text{half}} = \log(2)/\lambda_c$, where log is the natural logarithm. The solid curve shows the half lives against the number of iterations, and the shaded area around each curve indicates the relative degree of support of the corresponding radionuclide from the data (given by the standard deviation of the estimated nuclide concentration variance $\sqrt{\gamma_c}$.) We can see that after 200 iterations, three nuclides that have half lives close to the truth are constantly supported (wide area around the curves). The not so well supported components are heavily fluctuating as they are poorly supported by the data and have very little influence on the model prediction.

Figure 8 visualizes the prediction results for a partic-

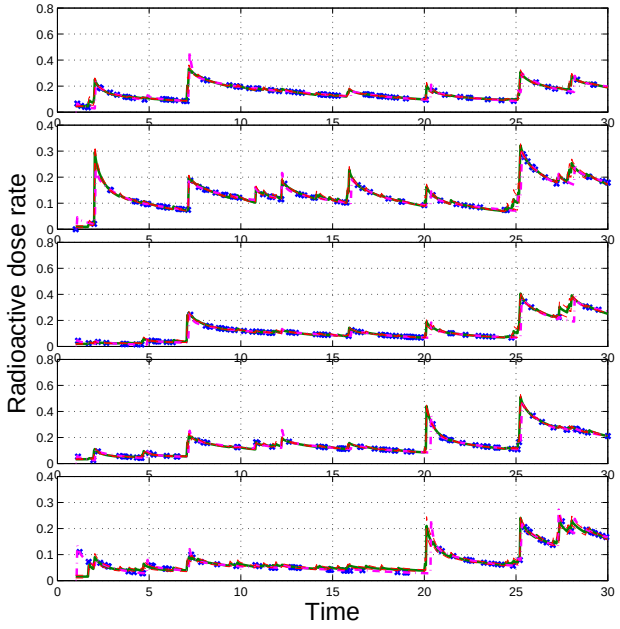


Figure 8: Prediction result on a synthetic data set. The dashed magenta curves show the true radioactivity. The blue crosses indicate the observed data points. The solid green curves show the mean posterior prediction, and the red dashed curves show the one standard deviation around the mean.

ular data. Although we only use 10% of the data for training (90% missing), the prediction results agree well with the true curves.

3.2 Fukushima Daiichi data set

The Fukushima Daiichi data set is originally released from Tokyo Electronic Power Company (TEPCO) [13] and collected in a machine readable format by a community effort.[4] The original data set contains 24 unique location tags. After removing locations with less than 10 samples, we analyze dose rate measurements from nine locations, namely "MP-1, around", "MP-2, around", "MP-3, around", "MP-4, around", MP-8, around", "Office building, north", "Front gate", "Front gate, around", "West gate". These locations are indicated in the blue circle in Figure 1.

Figure 9 shows the half lives of the estimated radionuclides against the number of iterations for a 10 component analysis of the Fukushima Daiichi data. The decay constant λ_c is converted into the half life by $t_{\text{half}} = \log(2)/\lambda_c$. The solid curve shows the half lives against the number of iterations, and the shaded area around each curve indicates again the relative degree of support of the corresponding radionuclide from the data. We can see that after about 300 iterations 3 prominent components emerge; two prominent quickly decaying components that could pertain to ^{131}Te, ^{132m}I, ^{132}I, or ^{135}I, and a relatively slow components that could correspond to ^{133}Xe or ^{131}I.

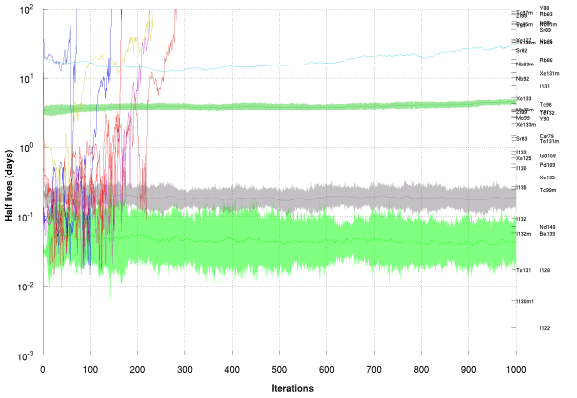


Figure 9: The half lives of the radionuclides estimated by the proposed model for the Fukushima Daiichi data set.

Figure 10 visualizes the prediction results by the estimated model. While the measured data is well accounted for by the model we see that the model attempts to predict regions that are not supported by measurements by aggregating information across the locations. The standard deviations for the predictions are indicated by the red dashed lines.

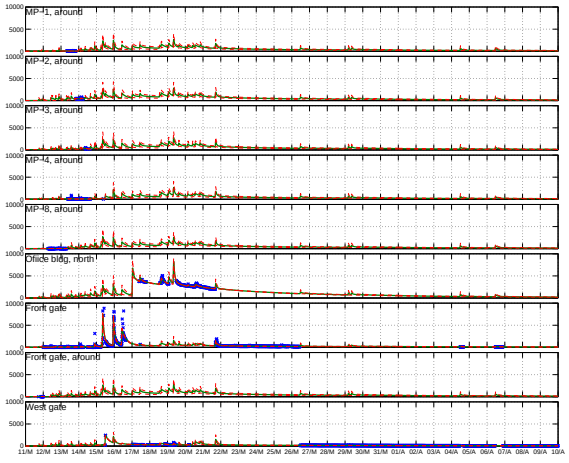


Figure 10: Prediction result on the Fukushima Daiichi data set. The green solid curve shows the posterior mean prediction. The red dashed curves show the one standard deviation around the mean.

3.3 Joint analysis of Fukushima Daiichi and MEXT data

The MEXT data set is originally released from Ministry of Education, Culture, Sports, Science, and Technology (MEXT), Japan [7]. We use the CSV format data maintained by Haruhiko Okumura [9]. The data set contains hourly radiation dose from 47 prefectural capital cities across Japan. Most cities use NaI scintillation detectors, but some cities transportable monitoring posts. In addition, each measuring station is located at a different height from the ground. In that sense, the data set is rather noisy and heterogeneous. We assume 1 μGy/h = 1 μSv/h. For this analysis we assume that there are in total 48 regions. The first region contains the nine locations we analyzed in the previous subsection from the Fukushima Daiichi plant. The 2nd to the 48th regions correspond to the 47 prefectural capital cities (see also Figure 1) and contains a single measurement site.

Figure 11 visualize the predictive mean dose rates for Utsunomiya city in Tochigi prefecture and Shinjuku-ward in Tokyo prefecture for the range of March 11th to April 10th. We can see that while the radioactivity from the Utsunomiya city mainly caused by the single event on March 15th, that for Tokyo is caused by events on March 15th, March 16th, 21st, and 22nd. This is natural because Tokyo is at least 200 km away from the Fukushima power plant, whereas Utsunomiya city is much closer to the plant. The predictions for the Fukushima Daiichi plant is very similar to Figure 10 and therefore not shown.

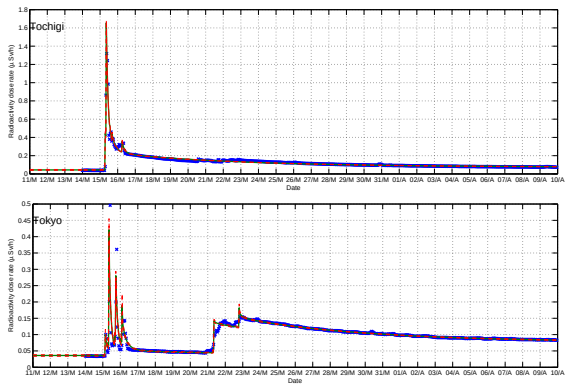


Figure 11: The predictive mean dose rate (green curve) for Utsunomiya city in Tochigi prefecture (top panel) and for Shinjuku-ward in Tokyo prefecture (bottom panel) are plotted in the range of March 11th to April 10th. The blue crosses are measurements.

4 Discussion

We have derived a generative model for the measurements of radioactive releases of the Fukushima Daiichi disaster. From very sparsely sampled measurements the model can estimate the presence and concentration levels of the underlying nuclides. In particular, our parametrizations is identifiable and in theory enable recovery of the initial relative concentration levels of the nuclides.

The Bayesian formulation naturally admits to impose additional structure into the problem. In the future we envision that information such as weather patterns can be used to inform about spatio-temporal correlation that can be modelled for instance by imposing Gaussian Process priors [11] on $\boldsymbol{B}^{(1:R)}$. We further envision that events can be inferred in the model by introducing birth-death processes that are able to propose and remove events during sampling based on reversible-jump MCMC [2, 3].

The model we have derived can be considered a first attempt at aggregating the sparse data available on the Fukushima Daiichi disaster, and we believe as more data will be available model estimation will improve. In particular, the model is not currently able to well infer slow decaying nuclides such as ^{134}Cs/^{137}Cs as these decay constants are poorly determined from the short period of time data has been recorded compared to the half-lives of these nuclides that are in the order of years. The model is further challenged by the very sparsely sampled data considered, thus in the future we will try to include data from other sources.

It will take long before the full extent of the Fukushima Daiichi disaster will be understood but there is no doubt that Bayesian methods for inferring latent structure in the sparse information available is a powerful tool in order to improve our understanding of the disaster. In this effort this can be considered a perhaps naïve but tractable first step.

Acknowledgements. We would like to thank Mikkel N. Schmidt, Ole Winther, Tue Herlau, and Haruhiko Okumura for their helpful discussions and comments. This work was supported in part by MEXT KAKENHI 22700138, 23240019, and NTT Communication Science Laboratories.

References

[1] Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, 1 edition, 2007.

[2] Peter J. Green. Reversible jump Markov chain Monte Carlo computation and Bayesian model

determination. *Biometrica*, 82(4):711–732, 1995.

[3] Peter J. Green and David I. Hastie. Reversible jump MCMC. 2009.

[4] Ryo Ichimiya. A summary page of radiation monitoring data and graphs related to the nuclear power plant accident in Fukushima, Japan. `https://sites.google.com/site/radmonitor311/top_english`.

[5] A. Ioannidou, S. Manenti, L. Gini, and F. Groppi. Fukushima fallout of ^{131}I, ^{137}Cs, ^{134}Cs at Milano, Italy. In *Proceedings of the 7th international conference on natural computation (ICNC)*, pages 2082–2085, 2011.

[6] M. Manolopoulou, E. Vagena, S. Stoulos, A. Ioannidou, and C. Papastefanou. Radioiodine and radiocesium in Thessaloniki, Northern Greece due to the Fukushima nuclear accident. *J Environ Radioact*, 102(8):796–7, 2011.

[7] Ministry of Education, Culture, Sports, Science, and Technology (MEXT). Reading of environmental radioactivity level by prefecture. `http://radioactivity.mext.go.jp/en/`.

[8] Masson O. Tracking of Airborne Radionuclides from the Damaged Fukushima Dai-Ichi Nuclear Reactors by European Networks. *Environmental Science & Technology*, 45(18):7670–7677, 2011.

[9] Haruhiko Okumura. `http://oku.edu.mie-u.ac.jp/~okumura/stat/data/`.

[10] International Commission on Radiation Units and Measurements. Gamma-ray spectrometry in the environment. *ICRU report 53*, 1994.

[11] Carl Edward Rasmussen. Gaussian processes for machine learning. MIT Press, 2006.

[12] Mikkel Schmidt. Linearly constrained bayesian matrix factorization for blind source separation. In Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams, and A. Culotta, editors, *Advances in Neural Information Processing Systems 22*, pages 1624–1632. 2009.

[13] Tokyo Electronic Power Company (TEPCO). Radiation dose measured in the Fukushima Daiichi Nuclear Power Station. `http://www.tepco.co.jp/nu/fukushima-np/f1/index-e.html`.

[14] Wikipedia. Fukushima Daiichi nuclear disaster. `http://en.wikipedia.org/wiki/Fukushima_Daiichi_nuclear_disaster`.

[15] T.J. Yasunari, A. Stohl, R.S. Hayano, J.F. Burkhart, S. Eckhardt, and T. Yasunari. Cesium-137 deposition and contamination of Japanese soils due to the Fukushima nuclear accident. *Proceedings of the National Academy of Sciences*, 108(49):19530–19534, 2011.