Improvement of training data for dose rate distribution using an artificial neural network

Miyuki Sasaki\textsuperscript{1*}, Yukihisa Sanada\textsuperscript{1}

\textsuperscript{1} Sector of Fukushima Research and Development, Japan Atomic Energy Agency

\textsuperscript{*}sasaki.miyuki@jaea.go.jp

Abstract. This study presents the evaluation results of the validity of the visualization map of the ambient dose rate at 1 m above the ground level using an artificial neural network. The dose rate map created using the artificial neural network-based method is found to reproduce ground-based survey results better than conventional methods. Suggested to improve the validity of the airborne radiation survey visualization, applying the color data obtained using a photogrammetry system is a new experience.

Keywords: Machine learning, Artificial neural networks, Airborne radiation monitoring, Fukushima Daiichi Nuclear Power Plant accident

1. Introduction

The airborne radiation survey (ARS) was conducted in various areas after the Fukushima Daiichi Nuclear Power Plant (FDNPP) accident in 2011 [1]. The ARS can quickly perform measurements in a wide range, but the distribution resolution of the ambient dose rate (air dose rate) by the ARS is rough compared with a ground-based survey because the detector position is far from the radiation source (ground). The inverse problem analysis and machine learning methods have been proposed to realize a more accurate visualization of the ARS data [2][3][4]. In the Fukushima environmental field, providing an improved position resolution of the dose rate distribution map is helpful in estimating the dose exposure of residents and in making decisions as regards lifting the evacuation zone.

In this study, we visualize the air dose rate distribution at 1 m above the ground level (agl) based on artificial neural networks (ANNs). The basic methodology of the ANNs for converting the ARS data is performed based on a previous study [4]. To improve the practical map, topography, and photographic color data excluded in the previous studies are added to the input variables. The training data of the ARS and the ground-based surveys around the FDNPP are obtained for the ANN construction. The ANN is constructed to convert the ARS data to an air dose rate of 1 m agl. We also evaluate the effect of each input variable on the conversion at the constructed ANN. This study shows the evaluation results of the validity of the visualization map of the ambient dose rate at 1 m agl using an ANN.
2. Method

2.1. Data acquisition

The ARS data were obtained using an unmanned helicopter (FAZER-R-G2, Yamaha Motor Co., Ltd., Shizuoka, Japan). The unmanned helicopter used approximately 50–80 m, 5–10 m s\(^{-1}\), and 50–150 m agl as the flight line space, flight speed, and flight altitude, respectively. The radiation detector consisted of a LaBr\(_3\) (Ce) scintillation detector (38 mm φ × 38 mm H × three detectors) and detected gamma rays in the 50–2800 keV energy range (Japan Radiation Engineering Co., Ltd., Ibaraki, Japan). The detector stored the gamma ray spectrum and the Global Navigation Satellite System (GNSS) location information every second. The digital elevation model (DEM) of the Geospatial Information Authority of Japan [5] and the three-dimensional (3D) orthophoto data (digital surface model [DSM]) from the photo data obtained by the unmanned helicopter were used as the topography information. Orthophoto data were used for Red-Green-Blue color mode (RGB) of the photographic color data. The air dose rate at 1 m agl (ground-based data) determined by a walk survey as the ground-based survey data was acquired together with the ARS data. The walk survey detector used was KURAMA-II (CsI: 13 mm × 13 mm × 20 mm) [6]. KURAMA-II saves the air dose rate and the GNSS location information every 3 s and converts the obtained gamma ray pulse to the air dose rate using the \(G(E)\) function. Figure 1 shows the \(G(E)\) function used in KURAMA-II.

![Figure 1: G(E) function of KURAMA-II](image)

2.2. Training dataset

The gamma ray spectrum data by the ARS, topographical information, and photo color data obtained using an unmanned helicopter were used as the input variables for the ANN construction. The air dose rates at 1 m agl measured using a walk survey meter were the objective variables. Training datasets (input and objective variables) were obtained around the FDNPP. Table 1 presents information on the input and objective variables. The ANN was constructed by varying the combination of the input variables. All data were averaged with a 10 m × 10 m mesh. The training data set was a total of 83495 sets.
Table 1: Information on the input and objective variables

| Data type   | Parameter | Unit      | Description                                                                 |
|-------------|-----------|-----------|-----------------------------------------------------------------------------|
| Input       | AGL       | m         | Absolute altitude in measurement (Flight altitude) — (Surface altitude)    |
|             | TC        | cps       | Count rate of 50–2800 keV                                                  |
|             | E50       | cps       | Count rate of 50–450 keV                                                   |
|             | E450      | cps       | Count rate of 450–900 keV                                                  |
|             | E900      | cps       | Count rate of 900–1400 keV                                                 |
|             | E1400     | cps       | Count rate of 1400–2800 keV                                                |
|             | DSM       | m         | Tree and building height in measurement (Digital Surface Model) — (Digital Elevation Model) |
|             | R         | -         | Aerial photo color data: Red (0–255)                                       |
|             | G         | -         | Aerial photo color data: Green (0–255)                                     |
|             | B         | -         | Aerial photo color data: Blue (0–255)                                      |
| Objective   | Ground-based data | µSv h⁻¹ | Air dose rate of 1 m above the ground level |

2.3. Network construction

A cascade correlation was used to construct the ANN [7]. It is a supervised learning algorithm that constructs a feedforward network. Learning begins with a minimal network consisting of only an input layer and an output layer. New hidden layers are added in each stage to minimize the overall network error. The ANN used in this study is a network that solves regression problems. The ANN structure is simple. All activation functions in the network are tanh functions. The output function is a sigmoid function. The objective function is cross-entropy with an added ridge regression. The adaptive subgradient method was used to update the weights and the biases [8]. NeuralWorks Predict (NeuralWare, Carnegie, USA) was used to construct and learn the ANN. The ANN was constructed with a pattern using the following input variables:
- Pattern-1. AGL and TC
- Pattern-2. AGL, TC, and DSM
- Pattern-3. AGL, TC, and RGB
- Pattern-4. AGL and 4C
- Pattern-5. AGL, 4C, and DSM
- Pattern-6. AGL, 4C, DSM, and RGB

2.4. Conventional method

The conversion validity was compared with the conversion result of the ANN constructed in each pattern and the conversion result by the flat sauce model (FSM), which was conventionally used for the ARS value conversion [9]. The FSM is a conventional conversion method that calculates the air dose rate at 1 m agl using the measurement altitude (Table 1: AGL) and the radiation count rate (Table 1: TC) in the sky. It calculates the radiation count rate by assuming that the count rate exponentially decays according to the measured altitude (distance to the ground). The FSM conversion formula is expressed as follows through (1):

\[ Y_{\text{FSM}_i} = (TC_i - BG) \cdot CD \cdot \exp[AF(AGL_i - H_{std})] \]  

where, \( Y_{\text{FSM}_i} \) is the converted value by the FSM in mesh \( i \) (µSv h⁻¹: air dose rate at 1 m agl); \( TC \) is the gamma ray counting rate (cps); \( BG \) is the detector background (cps); \( CD \) is
the air dose rate conversion coefficient for converting the gamma ray counting rate to the air
dose rate ($\mu$Sv h$^{-1}$ cps$^{-1}$); $AF$ is the altitude correction coefficient obtained by applying the
relationship between altitude and gamma ray count rate to an exponential function (m$^{-1}$);
$AGL$ is the absolute altitude in the measurement (m); and $H_{std}$ is the converted altitude (=1
m). The converted data was averaged with a 10 m $\times$ 10 m mesh. The number of mesh $i$ of
FSM is the same as the number of test datasets.

2.5. Evaluation parameter
The air dose rate was visualized by the ANN using test data other than the training data used
for the ANN construction. The test data set was a total 16264 set, and this data was 10 m $\times$
10 m mesh. The accuracy of the ANN conversion value was evaluated by comparing the
ground-based values in the same area. The root mean square error (RMSE) was used as the
index of the conversion accuracy. The $RMSE$ is expressed as follows through (2):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(Y_i - G_i)^2}{N}}.$$  \hspace{1cm} (2)

where, $N$ is the number of data; $Y_i$ is the converted value (output value) of the ARS data in
mesh $i$ ($\mu$Sv h$^{-1}$: air dose rate at 1 m agl); and $G_i$ is the ground-based survey value in mesh $i$
($\mu$Sv h$^{-1}$). When the RMSE approaches 0, it means that the converted value is close to the
ground-based survey value.

In addition, a relative deviation ($RD$) histogram calculated using the ground-based
measurement and converted values was created, and the accuracy was evaluated. The $RD$ is
expressed as follows using (3):

$$RD_i = \frac{(Y_i - G_i)}{G_i}.$$  \hspace{1cm} (3)

When the RD approaches 0, it means that the converted value is close to the ground-based
survey value. When RD is a positive value, it means that the converted value is converted
higher than the ground-based survey value.

2.6. Evaluation conditions of the input variables
The permutation importance was calculated to evaluate the effect of each input variable on
the ANN output [10]. It was computed as an RMSE comprising the ANN output value
obtained from the randomly created input variable and the ANN output value when the
specified input variable was randomly changed to prevent the data bias of the training data set
from affecting the evaluation of the permutation importance. The calculation was performed
using the ANN of Pattern-6. The calculation flow of the permutation importance is expressed
as follows:

1) A randomly created data of input variables (20000 sets) is input to the constructed ANN,
and gets the output values of ANN.
2) Randomly change the input variable of the dataset in 1) that calculates the permutation
importance, and get the output values of ANN.
3) Calculate the $RMSE$ using the output values obtained in 1) and 2).

A continuous input variable was input to the ANN constructed to evaluate the conversion
tendency. The tendency of the output value obtained from the ANN was also evaluated. The
conversion tendency was evaluated in the following patterns:

1) comparison of the output tendencies of Pattern-1 and the FSM conversion method;
2) comparison of the output tendency of Pattern-2 and the shielding tendency according to the height of the trees in the forest obtained in the field experiment; and
3) comparison of the conversion tendencies for each field color in Pattern-3.

The shielding tendency of the tree height in 2) was calculated from the data obtained from unmanned helicopters and the ground-based data in multiple forests. The calculation method detail has been described in a previous study [3]. These field experiment data were not used to construct the ANN.

3. Results and discussions

3.1. Validity of the air dose rate map

Figure 2(a) shows the $RD$ histograms comparing the ARS measurement data converted to the air dose rate at 1 m agl using each method and the ground-based data. Figure 2(b) presents the $RMSE$ defined as the error between each conversion result and the $RD$ median in Fig. 2(a). The ANN results showed $RD$ median approaches 0 when compared to the FSM. The $RD$ histograms of each conversion result of the evaluation pattern of the ANN seemed to be no different when different input variables were used. The ANN conversion result showed a decrease in the $RMSE$ compared to the FSM. However, no big difference was found in the $RMSE$ of each conversion result of the ANN. The results in Fig. 2 indicate that the ANN is more accurate than the FSM.

![Figure 2: $RD$ and $RMSE$ of the ARS conversion and ground measurement values: (a) histograms of $RD$ and (b) table of the $RMSE$ and $RD$ median values](image)

Figure 3 shows the results of converting the ARS measurement data (part of the test data) to an air dose rate at 1 m agl using the ANN and the FSM. Figure 3(a) presents the orthophoto data taken by an unmanned helicopter and the results of a walk survey data. Figure 3(b) depicts the conversion result by the FSM, while Fig.3(c)–(f) illustrate those by the ANN. The ANN constructed with Pattern-1 was used in (c); that with Pattern-4 was used in (d); that with Pattern-5 was employed in (e); and that with Pattern-6 was used in (f). The ANN conversion for the maps in Fig. 3 was multiplied by 1.6 to make the same $RD$ median of the FSM. Evaluating the difference in the distribution of each conversion is easy under
similar RD median and air dose rate scale. When the ANNs were multiplied by 1.6, the histogram for each ANN in Fig. 2 showed a median and a shape similar to those in the FSM. The conversion result of the ANN with a few input variables (Fig. 3(c)) exhibited a similar distribution to the FSM. The ANN distribution representation was more detailed when the input variables were increased. The difference was particularly remarkable in Fig. 3(f), with RGB as the input variable. The air dose rate in Fig. 3(f) was converted to a lower value on the roads and railroad tracks.

The results in Fig. 2–3 indicate that the ARS data converted by the ANN more accurately visualized the air dose rate at 1 m agl compared to those converted by the FSM. The ANN was more accurate than the RMSE of the FSM conversion value when compared with the RMSE of the ANN in Pattern-1. Even though the same input variable was used, the accuracy improved because the conversion parameters were optimized by the training.

![Figure 3: Air dose rate maps converted by each method](image)
(a) orthophoto data and results of a walk survey data, (b) the conversion result by the FSM, (c) the conversion result by the ANN: Pattern-1, (d) ANN: Pattern-4, (e) ANN: Pattern-5 and (f) ANN: Pattern-6. The mesh size of the air dose distribution is 10 m × 10 m.

### 3.2. Evaluation of input variables

Figure 4 shows the result of the permutation importance of each input variable in the ANN of Pattern-6. The input variable with the highest permutation importance was E450. The next largest input variable was AGL. The validity of the air dose rate map strongly depended on the flight altitude of an unmanned helicopter and the count rate of the radiocesium gamma ray
that reaches the detector. Such an evaluation helps us understand the important parameter.

3.3. Characteristics of the training data pattern

Figure 5 shows the comparison results of the 3D correlation of the three parameters of conversion factor, total count rate, and AGL of the ANN in Pattern-1 and the FSM conversion result to reveal the characteristics of the training data pattern. The conversion factor (CF) on the vertical axis in Fig. 5 was obtained by dividing the converted air dose rate from the TC. The CF of the FSM changed with the AGL (Fig. 5(a)) because the AF was unified, regardless of the TC (Eq. (1)). In contrast, the CF changed with the AGL and the TC (Fig. 5(b)), showing that the ANN analysis automatically corrected the AF by training data. The difference in the conversion tendency caused by the gamma ray energy was included in the network by ANN training. The main radionuclide around the FDNPP, for which the training data were obtained, was radioactive cesium. Natural radionuclides also existed in the field. The count rates of E900 and E1400 on the high-energy side, including the effects from the natural radionuclides around the FDNPP, were almost constant at all locations. In other words, the low TC area was largely influenced by the natural radionuclides (higher energy band), while the high TC area was largely influenced by radioactive cesium (lower energy band). Figure 5 shows that the dose rate trends were converted to high when the TC was low and converted to low when the TC was high. The objective variable (ground-based value) of the training data was the value obtained by converting the gamma ray pulse to the air dose rate using the $G(E)$ function. The ground-based values were converted to lower dose rates in the lower energy bands and higher dose rates in the higher energy bands. This tendency and that shown in Fig. 5 matched. In addition, the FSM did not include the difference in attenuation caused by the gamma ray energy in the conversion calculation. Theoretically, the contribution of the radiation derived from the natural nuclides, which have a different energy from radiocesium, is not negligible at the low-level TC. The ANN was more accurate than the FSM because it considers these effects by training.
Second, Figure 6 shows the ratio of the converted air dose rate and the measured ground-based data in Pattern-2. The vertical axis of the ANN denotes the value obtained by dividing the ANN conversion value when DSM–DEM = 0 by the ANN-converted value. Figure 6 illustrates the shielding factor of the forest trees obtained in the actual field experiment. The vertical axis of the measurement results of the forest shielding effect depicts the value obtained by dividing the value of the ARS data converted to 1 m agl using the FSM by the air dose rate measured at 1 m agl (ground-based values). The two graphs in Fig. 6 show the decay according to the exponential function, which demonstrated similar trends. The ANN analysis automatically represented the forest shielding factor by training data.

Lastly, Figure 7 shows the $CF$ for each color area in the ANN of Pattern-3. The average RGB color value in each area (i.e., forest, paddy field, building, and road) was used as the input variable for conversion. The RGB average values of all training data were used as the input variables. The vertical axis in Fig. 7 depicts the value ($CF$) obtained by dividing the converted value using the ANN by $TC$. The graph colors in Fig. 7 were reproduced with the average value of RGB in each area. “Forest” was converted slightly higher, while “Road” was converted slightly lower compared to the other areas. In other words, the ANN analysis automatically corrected the location factor using the training data. Theoretically, the air dose rate on the road can easily be decreased compared with that in the surroundings because radiocesium on the road is well known to be easily washed off [11]. In Fig. 3, the dose rates of the roads and railroad tracks were low when converted by the ANN including RGB as an input variable. In Fig. 7, “Road” was converted to a lower dose rate compared to the other areas. The roads and the railroad tracks were decontaminated around the FDNPP, and the training data were taken after the decontamination work was started. The inclusion of RGB to the input variable was considered to have added the color feature information on the roads and railroad tracks to the conversion calculation, consequently converting these areas to a lower dose rate. The dose rate of “Forest” in Fig. 7 was higher than that of the other areas. Training was considered to have added the shielding effect of the forest to the “Forest” color area.

When the input variable of the ANN was increased, the air dose rate distribution in Fig. 3 changed, but the $RMSE$ in Fig. 2(b) did not improve because the objective variable was limited. To obtain the objective variables, the walk survey was conducted in places where people could walk (e.g., sidewalks, rice field footpaths, and playgrounds). In the future, the
optimum evaluation method must be considered.

Figure 6: Comparison of the forest shielding effect and the ANN conversion tendency of Pattern-2

\[ y = \exp(-0.0127x) \]

\[ y = \exp(-0.0152x) \]

Figure 7: Conversion tendency of each area by the ANN of Pattern-3

4. Conclusions

In this study, we proposed a methodology for visualizing air dose rate distributions using ANNs. Adding the terrain and the photo color data as the input variables slightly improved the accuracy of the ANN conversion. The ANN that trained the ARS measurement data around the FDNPP learned the characteristics of the difference in the nuclide and decontamination statuses around the FDNPP.

Future studies are expected to focus on the improvement of the training data quality using additional measurement and simulation data in difficult approach areas, such as forests and farmlands. In addition, new conversion factors will be acquired from the conversion trends obtained through learning.
References

[1] Nuclear Regulation Authority in Japan: Monitoring information of environmental radioactivity level, https://radioactivity.nsr.go.jp/en/ (Accessed at 4 October, 2021).

[2] M. Sasaki, A. Ishizaki, Y. Sanada: Development of analysis method for airborne radiation monitoring using the inverse problem solutions, Prog. Nucl. Sci. Tech., 6 (2019), 63–67.

[3] M. Sasaki, Y. Sanada, A. Yamamoto: Application of the forest shielding factor to the maximum-likelihood expectation maximization method for airborne radiation monitoring, Radiat. Protect. Dosim., 184 (2019), 400-404.

[4] M. Sasaki, Y. Sanada, E. W. Katengeza, A. Yamamoto: New method for visualizing the dose rate distribution around the Fukushima Daiichi Nuclear Power Plant using artificial neural networks, Scientific Reports, 11 (2021), 1857.

[5] GSI (Geospatial information authority of Japan): Maps & Geospatial Information, https://www.gsi.go.jp/ENGLISH/page_e30031.html (Accessed at 4 October, 2021).

[6] M. Tanigaki, R. Okumura, K. Takamiya, N. Sato, H. Yoshino, H. Yoshinaga, Y. Kobayashi, A. Uehara, H. Yamana: Development of KURAMA-II and its operation in Fukushima. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 781 (2015), 57-64.

[7] S. E. Fahlman, C. Lebiere: The Cascade-Correlation Learning Architecture. Advances in Neural Information Processing Systems 2, (1990), 524-532.

[8] J. Duchi, E. Hazan, Y. Singer: Adaptive Subgradient Methods for Online Learning and Stochastic Optimization. The Journal of Machine Learning Research. 12 (2011), 2121-2159.

[9] Y. Sanada, T. Torii: Aerial radiation monitoring around the Fukushima Dai-ichi nuclear power plant using an unmanned helicopter, J. Environ. Radioact., 139 (2015), 294–299.

[10] A. Altmann, L. Toloşi, O. Sander, T. Lengauer: Permutation importance: a corrected feature importance measure, Bioinformatics, 26:10 (2010), 1340-1347.

[11] M. Ando, Y. Nakahara, S. Tsuda, T. Yoshida, N. Matsuda, F. Takahashi, S. Mikami, N. Kinouchi, T. Sato, M. Tanigaki, K. Takamiya, N. Sato, R. Okumura, Y. Uchihori, K. Saito: Measurement of air dose rates over a wide area around the Fukushima Dai-ichi Nuclear Power Plant through a series of car-borne surveys, Journal of Environmental Radioactivity, 139 (2015), 266-280.