Development of statistical and machine learning models to predict the occurrence of radiation fog in Japan

Momoko Negishi | Hiroyuki Kusaka

1Graduate school of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan
2Center for Computational Sciences, University of Tsukuba, Tsukuba, Japan

Abstract
This study develops statistical and machine learning models based on discriminant analysis, logistic regression analysis, and a support vector machine (SVM) to predict the occurrence of radiation fog in Japan. The selection of a suitable set of explanatory variables for the models was made using the Akaike information criterion (AIC). The accuracies of the three models were measured and compared. To determine the optimum combination of explanatory variables, temperature, humidity, wind speed, precipitation, sunshine, and visibility data were considered. Based on the root mean square error (RMSE) and AIC values, the best combination of variables was found to include: the presence of precipitation, mean wind speed during the night, minimum temperature during the night, the amount of temperature cooling during the night, the minimum humidity during the previous day, and visibility at 18:00. A comparison of the predictive accuracies of the three models using the selected variable combination showed that the discriminant model produced a critical success index (CSI, or threat score) of 22.5, while the logistic regression model and the SVM model both produced better CSI results—with scores of 33.8 and 38.2, respectively.

KEYWORDS
AIC, discriminant model, logistic regression model, radiation fog, SVM

1 INTRODUCTION

In Japan, traffic accidents often occur due to visibility problems caused by fog (e.g., Yamamoto, 2000; Ohashi & Suido, 2021). Notably, basins in Japan have been found to have more fog days than plains; that is, basins are more prone to radiation fog than plains (Akimoto & Kusaka, 2010). Radiation fog is more likely to cause serious accidents. Indeed, there have been cases of multiple vehicle collisions due to radiation fog in the Aizu Basin (Yamamoto & Oyamada, 2000). To prevent or reduce the occurrence of such accidents, it is necessary to understand the factors that produce radiation fog and to develop an ability to reliably predict its occurrence.

Radiation fog in Japan, which is the subject of this study, is fog that tends to occur in terrain such as valleys and basins; its generation is known to be influenced by radiative cooling at night and the formation of cold lakes created by cold airflow. Radiation fog has been observed at fixed points on the ground and over an area using satellites in many parts of the world (e.g., Bendix, 2002; Cuxart & Jimenez, 2012; Lee et al., 1997; Meyer & Lala,
1990; Teshiba et al., 2004; Underwood et al., 2004; van Schalkwyk & Dyson, 2013). By observational and numerical studies, the conditions for radiation fog generation have been clarified (e.g., Bergot, 2012; Brown & Roach, 1976; Gerber, 1981; Koračin et al., 2001; Liu et al., 2011; Mason, 1982; Meyer & Lala, 1990; Price et al., 2018; Roach et al., 1976; Smith et al., 2018). Researchers have also been reproduced and predicted radiation fog using conceptual one-dimensional boundary-layer models, high-resolution three-dimensional numerical weather prediction (NWP) models, multi-model mesoscale ensemble prediction system, or large eddy simulation models (e.g., Bott et al., 1990; Duynkerke, 1991; Guedalia & Bergot, 1994; Meyer & Rao, 1999; Nakanishi, 2000; Bott & Trautmann, 2002; Pagowski et al., 2004; Bergot et al., 2007; Zhou & Ferrier, 2008; Zhou & Du, 2010; Van der Velde et al., 2010; Bergot et al., 2015; Cui et al., 2019; Ito et al., 2019; Gu et al., 2019; Smith et al., 2020). Recent studies show that sub-km high-resolution NWP models and large eddy simulation (LES) models can simulate and predict radiation fog better. However, the accuracy has a high sensitivity to surface properties (Smith et al., 2020).

Machine learning is increasingly being used for radiation fog prediction as well as other fogs (e.g., Durán-Rosal et al., 2018), as an example, Fabbian and de Dear (2007) used a neural network to successfully predict radiation fog at the Canberra airport. More recently, in K-C Miao et al. (2020), they used several different methods to make predictions of radiation fogs and compared the accuracy of each method. Radiation fog prediction methods using machine learning have been developed.

Prediction of radiation fog has also been conducted in Japan. Yamada et al. (1999) developed a statistical model and used it to predict the radiation fog that occurs at New Chitose Airport. Prediction of radiation fog occurring in basins has also been previously studied (Ohashi et al., 2012). Ohashi used a discriminant equation to predict radiation fog at night in the Miyoshi Basin.

The purpose of this study is to develop a predictive model for radiative basin fog using statistical and machine learning methods.

Two aspects of this study make it particularly unique. Firstly, the study uses an effective approach to select the explanatory variables for the model, an important step in ensuring the accuracy of the model’s predictions. Secondly, multiple models are developed and compared for accuracy.

2 | METHOD

2.1 | Study area and target time

The target area for predicting the occurrence of basin fog is the Chichibu Basin in Saitama Prefecture (Figure 1a,b). There are two reasons for this selection: Firstly, there is the JMA’s Chichibu Special District Meteorological Observatory (hereinafter referred to as the Chichibu observatory) in this area (Figure 1b), and the observations are called Automated Meteorological Data Acquisition System (AMeDAS) data. More than 20 years of AMeDAS observation data, including visibility data, are available and easily provide sufficient training data for our models. Secondly, the Chichibu Basin is the closest area to Tokyo where ‘sea of clouds’ occur. In recent years, ‘sea of clouds’ has become a tourist attraction among the Japanese, and it is possible that accurate fog prediction can be used to support and enhance such tourist activity in the Chichibu area and elsewhere.

The target time for fog prediction is the night-time from 0100 to 0800 Japan Standard Time (JST) during autumn (20 September–31 December). This is because, as is known well, fog occurs most frequently in the autumn (Figure 2). When fog appears at night, traffic accidents are more likely to occur. Thus, predicting the occurrence of fog at night is crucial to determine whether roads should be closed, or vehicle speeds should be restricted. It is also assumed that the prediction of fog will be made at 1800 JST on the previous day.

2.2 | Data

The near-surface air temperature, relative humidity, wind speed, wind direction, precipitation, pressure, and sunshine duration are measured automatically and hourly at the Chichibu observatory. Visibility is also measured automatically and hourly using a transmissometer. The accuracy of the instruments is tested regularly. The JMA’s certification standards are 0.1°C for air temperature, 1% for relative humidity, 0.1 m/s for wind speed, and 0.5 mm for precipitation (JMA, 2021).

The observation data used in this study included visibility, air temperature, relative humidity, wind speed, and precipitation. The data for 1998–2016 are used as the training data to determine the parameters in the prediction models (i.e., statistical and machine learning models), and the remaining data for 2017–2019 are used as test data.

In our study, if visibility is reported to be less than 1 km at least once from 0100 to 0800 JST, the day is treated as a ‘fog occurrence day’, under the assumption that fog has appeared. By screening rainfall cases, we distinguish poor visibility caused by radiative fog and heavy rainfall. Although we do not distinguish between advection fog and basin fog, we determine that the fall night-time fog in this area is basin fog based on wind direction, wind speed, and weather maps.
2.3 Difficulty in numerical weather prediction

Before developing the models, it is important to check NWP data could be effectively used to predict fog. The NWP data used in this study are MSM-GPV data, which are output from the prediction system operated by the JMA. The JMA used a non-hydrostatic mesoscale meteorological model called ‘NHM’ for regional-scale weather prediction in the operational prediction system until February 2017 (e.g., Saito et al., 2006, 2007). Subsequently, the JMA changed the model to a newly developed non-hydrostatic meteorological model called ‘asuca’ (JMA, 2019). The locations of the AMeDAS and MSM-GPV points are shown in Figure 1b. The distance between the two points was approximately 1.5 km. The elevations of the AMeDAS and MSM grid points were 232 m and approximately 320 m, respectively. The difference in the elevation was 88 m.

We compared the relative humidity of the MSM-GPV data with that of the AMeDAS observation data. As Figure 3 shows, the GPV predictions are far different from the AMeDAS observations and almost never approach 100% humidity. Thus, the predicted values of humidity by GPV alone are insufficient to effectively forecast the presence of fog. Indeed, we attempted to reproduce the fog occurrence using the GPV data and following estimation equation used in the previous studies (e.g., Doran et al., 1999; Gu et al., 2019), but found it difficult to do that (Figure 4). This supports that humidity of GPV data should not be used for prediction as it is.

\[
\text{VIS}_{\text{RUC}} = 10^3 \times 60 \times e^{\left[\frac{-2.5 \times (\text{RH} - 15)}{80}\right]} \quad \text{(Rapid Update Cycle [RUC] method)}
\]

\[
\text{VIS}_{\text{fsl}} = 1609 \times 6000 \times \frac{T - T_d}{\text{RH}^{1.75}}
\]
The daily minimum temperature, temperature drop, and wind speed of the MSM-GPV data were also compared with those of the AMeDAS data (Table 1). Herein, the temperature drop represents the difference between the temperature at 1800 JST on the previous day and the minimum temperature appeared between 0000 JST and 0600 JST on the same day. Table 1 summarizes the mean error of each element. The mean error values were 18% for maximum humidity, $1.5^\circ$C for temperature drop (Temp drop), $1.5^\circ$C for minimum temperature, and 0.4 m/s for wind speed. It was found that both temperature and wind speed produce a much smaller error than humidity. The accuracy for the minimum temperature is particularly good. If no bias correction or guidance is applied, the accuracy of the maximum humidity predictions is relatively poor, which means that the predictive power of water vapour content is poorer than the other variables.

### 2.4 Prediction method used in this study

From the above, it seems that direct prediction of fog using the JMA’s NWP model (GPV data) is difficult. There are two other possible methods at least to overcome this problem:

1. Using bias-corrected relative humidity of GPV data.
2. Using statistical methods or machine learning.

In this study, we adopted method (2). It should be noted that we are not rejecting method (1), but rather have chosen method (2) as our focus.

### 2.5 Statistical and machine learning models used in the study

Our goal is to predict two categories of events, namely the presence or absence of fog, using the statistical and machine learning models. The presence or absence of fog was determined based on the visibility values. The visibility values were used by numerically assigning a visibility value of 1 as ‘fog’, when the visibility was $\leq 1$ km, and 0 as ‘no fog’, when the visibility was $>1$ km.

For the prediction models, we selected discriminant analysis, logistic regression analysis, and SVM, which are typical methods for classifying events into two categories. Along with the predictions of our three models, benchmark predictions were made using the discriminant equation developed in Ohashi et al. (2012). In this study, we refer to Ohashi’s discriminant equation as discriminant formula (O); we refer to our newly developed discriminant formula as discriminant formula (N). The Ohashi’s method is described in Appendix A. The general nature of each of three methods used in the present study is described below.

#### 2.5.1 Linear discriminant analysis

In linear discriminant analysis (LDA), the data elements are divided into two groups using a linear discriminant function, which produces a division like that shown in Figure 5a. The general form of the discriminant function is

$$Z = a_0 + a_1x_1 + a_2x_2 + \ldots + a_nx_n$$

The discrimination results produced by a linear discriminant function are easy to judge, as group membership is indicated by the sign of $Z$. However, since the data are divided into two groups by a straight line, groups with complicated boundaries cannot be fully separated. This method has been used in a prior study of basin fog prediction in Japan (Ohashi et al., 2012).
2.5.2 | Logistic regression analysis

Logistic regression analysis discriminates by logit curves and is suitable for cases where the target variable is qualitative, and the explanatory variables are quantitative. The logit model produces

$$P = \frac{e^{(a_0 + a_1x_1 + a_2x_2 + \ldots + a_nx_n)}}{1 + e^{(a_0 + a_1x_1 + a_2x_2 + \ldots + a_nx_n)}}$$

where $P$ is the predicted value (a value between 0 and 1), $x_i$ are the explanatory variables (in this case, the input data for the weather elements), and $a_i$ represents the explanatory variable coefficients, with an $a_0$ constant.

The objective variable in our study is the presence or absence of fog, expressed as 1 or 0: a day with fog is given a value of 1; a day without fog is given a value of 0.

When values for the explanatory variables are input, the prediction result $p$ is treated as a probability. The user can set a boundary value for deciding whether the target event occurs (in our case, the appearance of fog) based on the prediction result (for example, a $p$ value greater than 0.4 might be treated as an indicator that fog will occur).

2.5.3 | Support vector machine (SVM)

A support vector machine is a type of machine learning approach commonly used for discrimination with supervised data. Here, the discrimination targets are grouped using machine learning (Figure 5c). In our model, the presence or absence of fog is represented by 1 and 0.

Since a non-linear separator can be used as the boundary line for grouping, even groups with complicated boundaries can be finely divided. In the development stage, it is necessary to tune the method to ensure suitability. While this method is easy to use, it outputs only the results of the discrimination, not the details.

2.6 | Which meteorological elements should be used for explanatory variables in the prediction models?

In statistical modelling and machine learning, the choice of explanatory variables is highly important. Table 2 lists the candidate variables that could be produced from the available data. In selecting the candidates, we focused on several key points. Firstly, for the variables related to water vapour, we decided to use the observed values of the day rather than the predicted values, for the reasons cited above (i.e., because of the poor GPV accuracy that results when there is no bias correction or guidance). For variables that were found to have relatively good GPV accuracy, such as temperature and average wind speed, the next day’s predicted values should be used in the statistical and machine learning models. However, observations are used for all variables instead of GPV data in the present study. This is because the purpose of the present study is model development and model intercomparison. Namely, the predictions performed in the present study are so-called perfect model experiments using observations, which can be considered reasonable in terms of basic research.

In this study, we check correlations between the meteorological elements, and then select the explanatory variables of our models using the Akaike information criterion (AIC).

To choose which meteorological elements to be used as the explanatory variables in our predictive equations, we firstly checked their correlations. We decided not to use elements with a correlation of 0.4 or higher at the same time. In addition, we calculated the variance inflation factor (VIF) to check for multicollinearity. Those elements that produced a VIF value of 10 or more were not used together. This criterion is common and is used in several previous studies (e.g., Chatterjee & Hadi, 2012; Li et al., 2016).

We next used the AIC, combining meteorological elements that were admissible according to the above check.
In general, the AIC is used to find better combinations of explanatory variables that improve the relative accuracy of a predictive model (e.g., Sato et al., 2020; Schatz & Kucharik, 2014). It considers the trade-off between model fit and the total number of variables (elements, parameters) included in the model; the smaller the AIC value, the more effective, in relative terms, is the combination of variables being considered. The AIC value is calculated as

\[
AIC = -2 \times (\text{maximum log likelihood} + \text{number of free parameters})
\]

The logistic regression analysis model was used as the prototype model, and each AIC was checked by fitting a combination of meteorological elements. AIC checks were performed for all possible combinations of the meteorological elements, and the total number of combinations was 511. Table 3 shows the best combination of meteorological elements according to the AIC results. As indicated, the lowest AIC (1123.6) was for the six-element combination: precipitation, mean night-time wind speed, night-time minimum temperature, night-time temperature drop, previous day’s minimum relative humidity, and visibility at 1800 JST.

The time for each meteorological element used in the models was from 0000 to 2400 JST on the previous day for precipitation and from 0000 to 0600 JST on the same day for mean night-time wind speed, minimum night-time temperature, and night-time temperature drop. The time of 1800 JST for visibility, which is close to the time of sunset, was selected because our objective was to predict fog occurrence at night. The other reason is that the night before would be an appropriate timing for forecasting and usefulness in the case of actual operations.

### TABLE 2
Candidate meteorological elements and their data content details

| Data format | Data source          | Period               |
|-------------|----------------------|----------------------|
| Precipitation | 0, 1                | Amount of precipitation | Previous day |
| Daylight hours | Numerical value | Daylight hours | Previous day |
| Average night-time wind speed | Numerical value | Wind speed | Night-time(0:00–6:00) |
| Lowest night-time temp. | Numerical value | Temperature | Night-time(0:00–6:00) |
| Night-time temp. drop | Numerical value | Temperature | Night-time(0:00–6:00) |
| Daily minimum humidity | Numerical value | Humidity | Previous day |
| Amount of water vapour at 18:00 | Numerical value | Temp., Humidity | 18:00 on the previous day |
| Dew point temp. at 18:00 | Numerical value | Dew point temp. | 18:00 on the previous day |
| Visibility at 18:00 | Numerical value | Visibility | 18:00 on the previous day |

### TABLE 3
The best combination of the explanatory variables chosen according to the AIC

| Selected | |
|----------|---|
| Precipitation | ○ |
| Daylight hours | ○ |
| Average night-time wind speed | ○ |
| Lowest night-time temp. | ○ |
| Night-time temp. drop | ○ |
| Daily minimum humidity | ○ |
| Amount of water vapour at 18:00 | ○ |
| Dew point temp. at 18:00 | ○ |
| Visibility at 18:00 | ○ |
| AIC | 1123.6 |

## 3 RESULT

### 3.1 Prediction results

As indicated earlier, the six variables selected for inclusion in our predictive models were precipitation, mean night-time wind speed, night-time minimum temperature, night-time temperature drop, previous day’s minimum humidity, and visibility at 18:00.

The results from the prediction equations developed using each of the three methods, as well as the benchmark predictions of discriminant function (O), are shown in Table 4. To compare results, the number of a false positive and the number of misses were counted. Furthermore, since fog, the target of these predictions, is a phenomenon with a low daily occurrence rate, the ‘critical success index’ (CSI) is used rather than the accuracy rate as a measure of performance. CSI is calculated as

\[
CSI = \frac{FO}{N - XX}
\]
where \( FO = \) predicted and observed, \( N = \) total number of predictions, \( XX = \) no prediction and no observation.

The results of the CSI calculations are summarized in Table 4 and Figure 6. Details of the prediction results are described below.

### 3.1.1 | Discriminant formula (O)

Of the four methods, discriminant formula (O) produced the fewest misses, but a false positive was the most. The CSI was 19.3.

### 3.1.2 | Discriminant formula (N)

The number of misses increased but the number of strikes decreased compared with discriminant formula (O). The CSI was 22.5, which is better than discriminant formula (O).

### 3.1.3 | Logistic regression analysis

The number of misses here is somewhat higher than for the discriminant formula, but the number of strikes is much lower. The CSI was 33.8, which is better than discriminant formula (N).

### 3.1.4 | SVM

The number of missed days is high—more than half of the total fog days—but the number of strikes has almost disappeared. The CSI was 38.2, which is the highest among all the methods; thus, it can be said that accuracy was good.

Based on the CSI comparisons, SVM performed best, followed in descending order by the logistic regression model and discriminant equation (N). All three methods were found to be better than the benchmark discriminant formula (O).

Although this study evaluated the increase in CSI score, the results were mainly dominated by false positives (false alarms). When using prediction results to make decisions regarding stopping traffic at night, it is also important to reduce false negative (misses). False negatives of discriminant, logistic regression analysis, and SVM are respectively 0.03, 0.04, and 0.06.

### 3.2 | Fog events unpredicted by the two statistical models

The prediction results of discriminant equation (N) are shown in Figure 7. Here, \( Z = 0 \) serves as the boundary between predicted fog days (which are on the minus side of 0) and predicted non-fog days (on the plus side of 0). Actual fog days are shown as white bars; actual non-fog days are shown as black bars. According to Figure 7, there are almost no cases where the prediction is far off. However, there are many errors in the vicinity of \(-1 < Z < 1\) (i.e., near the \( Z = 0 \) boundary).

The results of the logistic regression model are shown in Figure 8. In this case, \( p > 0.2 \) was considered to be indicative of a fog day. As shown, most of the non-fog days (black circles) in the sample are in the \( p < 0.2 \) region. On the other hand, there are several cases where the value of \( p \) is much larger than 0.2 even though the day is actually a non-fog day.
**DISCUSSION**

**4.1 Comparison with a standard model developed by Ohashi et al. (2012)**

In contrast to the previous study by Ohashi et al. (2012), which we used for a benchmark, we were able to develop a discriminant equation that was suitable for the Chichibu Basin (one that increased the CSI by 3.5). One possible reason for this is the difference in the explanatory variables used to develop the two equations. In the previous study, only three explanatory variables were included: maximum possible cooling, humidity, and wind speed. On the other hand, six variables were used in our study: night-time minimum temperature, night-time temperature drop, night-time mean wind speed, previous day's precipitation, previous day's minimum humidity, and visibility (at 18:00). The difference in the results of the previous and newly proposed discriminant equations suggests that the three explanatory variables of maximum possible cooling, humidity, and wind speed are not sufficient to discriminate between fog and non-fog days. More generally, this means that the selection of appropriate explanatory variables is important for improving the prediction accuracy.

It should also be noted that the maximum possible cooling used in Ohashi et al. (2012) requires many variables and is complicated to derive, whereas the meteorological factors used in this study are simple to produce. This is a small point, but it represents an improvement.

**4.2 Handling of ‘mist’ data**

One concern is the impact of ‘mist’ days on our modeling. We checked the data for all days in the period under study and found that many of the days were ‘mist’ days with visibility between 1 km and 10 km, accounting for nearly half of the total data (Figure 9). In some years, this was more than 50% of the total number of days. It is possible that these ‘mist’ data interfere with the identification of fog. We leave this consideration to future study.

**4.3 Experiments at other locations**

To test the performance of SVM (over the conventional method of discriminant analysis), we conducted prediction experiments and comparative analysis using variable selection by AIC, SVM, and discriminant analysis models at other sites.

The target area was the Shinjo Basin (latitude 38.7613, longitude 140.3056), which is located on the Sea of Japan in northern Japan. The Shinjo Basin was selected for the following two reasons. Firstly, the Shinjo Basin is well known for its frequent foggy nights in fall season, which caused severe traffic accidents. Secondly, Chikazawa and Wada (2004) have conducted studies for predicting fog at night using the method of discriminant analysis.

The Shinjo Basin is surrounded by mountains at an elevation of approximately 1000 m. There is a regional observation station of JMA in Shinjo Basin, where...
temperature, humidity, wind direction, wind speed, precipitation, and visibility are observed hourly (Figure 10). The predicted results from the discriminant analysis, logistic regression analysis, and SVM models are shown in Figure 11. The CSI of the predicted results from the discriminant analysis, logistic regression analysis, and SVM were 54.2, 59.8, and 64.1, respectively. The results were similar to those of Chichibu as shown in Figure 6. Therefore, we can conclude that predicting basin fog at night in autumn using the AIC variable selection plus SVM model has better accuracy than the conventional method using discriminant analysis in both Chichibu and Shinjo Basin.

5 | CONCLUSIONS

This study aims at predicting radiation fog occurred in the Chichibu Basin in Japan using models based on three different estimation methods. The data used in the models were obtained from AMeDAS dataset.

The AIC was used to select a suitable set of explanatory variables for the models, which are important for predicting radiation fog. The explanatory variables selected were precipitation, mean night-time wind speed, night-time minimum temperature, night-time temperature drop, minimum previous day humidity, and visibility at 1800 JST.

The prediction accuracies of the discriminant, logistic regression, and SVM models, all of which used the same set of input variables, were compared. Based on a comparison of CSI values, SVM produced the best results, with a CSI of 38.2, followed by the logistic regression model (33.8) and the linear discriminant model (22.5). The proposed model was found to be more effective and better than the discriminant equation used in the previous studies. Good performance of the model is also shown in another location (Shinjo Basin).

The predictions performed in this study were perfect model experiments using observations, which can be considered reasonable in terms of basic research. For practical applications, it is necessary to verify the model using forecast data rather than observations, which would be considered in future studies. This study focussed on the presence or absence of fog and did not address the timing of fog generation and disappearance. It would be better if we could develop a model for predicting the timing of fog occurrence and disappearance with more benefits. As this requires higher level of prediction, we would consider this a subject for the future study.

ACKNOWLEDGEMENTS

This research was partly supported by the Multi-disciplinary Cooperative Research Program in the Center for Computational Sciences, University of Tsukuba, and by JSPS KAKENHI Grant Number JP19H01155. We would like to express our sincere gratitude to all of the people involved in this study. Figures 1 and 10 have been
made using the GSI maps by the Geospatial Information Authority of Japan.

AUTHOR CONTRIBUTIONS
Momoko Negishi: Data curation (lead); formal analysis (lead); methodology (equal); visualization (lead); writing – original draft (equal); writing – review and editing (equal). Hiroyuki Kusaka: Conceptualization (equal); funding acquisition (lead); methodology (equal); project administration (lead); resources (lead); supervision (lead); writing – original draft (equal); writing – review and editing (equal).

ORCID
Hiroyuki Kusaka https://orcid.org/0000-0002-0326-7179

REFERENCES
Akimoto, Y. & Kusaka, H. (2010) A climatological study of fog in Japan based on event data. Atmospheric Research, 151, 200–211.
Bendix, J. (2002) A satellite-based climatology of fog and low-level stratus in Germany and adjacent areas. Atmospheric Research, 64, 3–18.
Bergot, T. (2012) Small-scale structure of radiation fog: a large-eddy simulation study. Quarterly Journal of the Royal Meteorological Society, 139, 1099–1112.
Bergot, T., Escobarb, J. & Massona, V. (2015) Effect of small-scale surface heterogeneities and buildings on radiation fog: large-eddy simulation study at Paris-Charles de Gaulle airport. Quarterly Journal of the Royal Meteorological Society, 141, 285–298.
Bergot, T., Terradellas, E., Cuxart, J., Mira, A., Liechti, O., Mueller, M. et al. (2007) Intercomparison of single-column numerical models for the prediction of radiation fog. Journal of Applied Meteorology and Climatology, 46, 504–521.
Bott, A., Sievers, U. & Zhulkowski, W. (1990) A radiation fog model with a detailed treatment of the interaction between radiative transfer and fog microphysics. Journal of the Atmospheric Sciences, 47, 2153–2166.
Bott, A. & Trautmann, T. (2002) PAFOG - a new efficient forecast model of radiation fog and low-level stratiform clouds. Atmospheric Research, 64, 191–203.
Brown, R. & Roach, W.T. (1976) The physics of radiation fog: I - a field study. Quarterly Journal of the Royal Meteorological Society, 102, 335–354.
Chatterjee, S. & Hadi, A.A. (2012) Regression analysis by example, 5th edition. New Jersey, USA: John Wiley & Sons, Inc.
Chikazawa, M. & Wada, K. (2004) Prediction of radiation fog by discriminant analysis in Shinjo Basin Yamagata prefecture (in Japanese). Journal of Meteorological Research, 56, 71–81.
Cui, C., Bao, Y., Yuan, C., Li, Z. & Zong, C. (2019) Comparison of the performances between the WRF and WRF-LES models in radiation fog – a case study. Atmospheric Research, 226, 76–86.
Cuxart, J. & Jimenez, M.A. (2012) Deep radiation fog in a wide Closed Valley: study by numerical modeling and remote sensing. Pure and Applied Geophysics, 169, 911–926.
Doran, J. A., Roohr, P. J., Beberwyk, D. J., Brooks, G. R., Gayno, G. A., Williams, R. T., Lewis, J. M., & Lefevre, R. J. (1999) The MM5 at the Air Force Weather Agency-New products to support military operations. The 8th Conference on Aviation, Range, and Aerospace Meteorology. NOAA/NWS, Dallas, Texas. Durán-Rosal, A.M., Fernández, J.C., Casanova-Mateo, C., Sanz-Justo, J., Salcedo-Sanz, S. & Hervás-Martínez, C. (2018) Efficient fog prediction with multi-objective evolutionary neural networks. Applied Soft Computing, 70, 347–358.
Duynkerke, P.G. (1991) Radiation fog: a comparison of model simulation with detailed observations. Monthly Weather Review, 119, 324–341.
Fabbian, D. & de Dear, R.J. (2007) Application of artificial neural network forecasts to predict fog at Canberra international airport. Weather and Forecasting, 22, 372–381.
Gerber, H.E. (1981) Microstructure of a radiation fog. Journal of the Atmospheric Sciences, 38, 454–458.
Gu, Y., Kusaka, H., Doan, Q.-V. & Tan, J. (2019) Impacts of urban expansion on fog types in Shanghai, China: numerical experiments by WRF model. Atmospheric Research, 220, 57–74.
Guedalia, D. & Bergot, T. (1994) Numerical forecasting of radiation fog. Part II: a comparison of model simulation with Severe observed fog events. Monthly Weather Review, 122, 1231–1246.
Ito, J., Nagoshi, T. & Nino, H. (2019) A numerical study of ‘Hijikawa-Arashi’: a thermally driven gap wind visualized by nocturnal fog. Journal of Applied Meteorology and Climatology, 58, 1293–1307.
Japan Meteorological Agency. (2019). Outline of the operational numerical weather prediction at the Japan meteorological agency. Appendix to WMO Technical Progress Report on the Global Data-Processing and Forecasting System and Numerical Weather Prediction. 229 pp., https://www.jma.go.jp/jma-eng/jma-center/nwp/outline2019-nwp/index.htm.
Japan Meteorological Agency. (2021), private communication to JMA weather counselling office. 26 November 2021.
Koračin, D., Lewis, J., Thompson, W.T., Dorman, C.E. & Businger, J.A. (2001) Transition of stratus into fog along the California coast: observations and modeling. Journal of the Atmospheric Sciences, 58, 1714–1731.
Lee, T., Turk, J. & Richardson, K. (1997) Stratus and fog products using GOES-8/9 3.9-mm data. Weather and Forecasting, 12, 664–677.
Li, Z., Wang, K., Zhou, C. & Wang, L. (2016) Short communication modelling the true monthly mean temperature from continuous measurements over global land. International Journal of Climatology, 36, 2103–2110.
Liu, D., Yang, J., Niu, S. & Li, Z. (2011) On the evolution and structure of a radiation fog event in Nanjing. Advances in Atmospheric Sciences, 28, 223–237.
Mason, J. (1982) The physics of radiation fog. Journal of the Meteorological Society of Japan, 60, 486–499.
Meyer, M.B. & Lala, G.G. (1990) Climatological aspects of radiation fog occurrence at Albany, New York. Journal of Climate, 3, 577–586.
Meyer, W.D. & Rao, G.V. (1999) Radiation fog prediction using a simple numerical model. Pure and Applied Geophysics, 155, 57–80.
Miao, K., Han, T., Yao, Y., Lu, H., Chen, P., Wang, B. et al. (2020) Application of LSTM for short term fog forecasting based on meteorological elements. Neurocomputing, 408, 285–291.
Nakanishi, M. (2000) Large-Eddy simulation of radiation fog. Boundary-Layer Meteorology, 94, 461–493.
Ohashi, Y., Iwashita, M. & Kusamoto, M. (2012) Possibility of predicting basin fog by using the maximum possible cooling amount and surveillance camera image. Journal of Agricultural Meteorology, 68, 97–106.
Ohashi, Y. & Suido, M. (2021) Numerical simulations of upslope fog observed at Beppu Bay in Oita prefecture, Japan. *Meteorological Applications*, 28, e2003.

Pagowski, M., Gultepe, I. & King, P. (2004) Analysis and simulation of an extremely dense fog event in southern Ontario. *Journal of Applied Meteorology*, 43, 3–16.

Price, J.D., Lane, S., Boutle, I.A., Smith, D.K.E., Bergot, T., Lac, C. et al. (2018) A field and modelling study to improve our understanding and forecasting of radiation fog. *Bulletin of the American Meteorological Society*, 99, 2061–2077.

Roach, W.T., Brown, R., Caughey, S.J., Garland, J.A. & Readings, C. J. (1976) The physics of radiation fog: I - a field study. *Quarterly Journal of the Royal Meteorological Society*, 102, 313–333.

Saito, K., Fujita, T., Yamada, Y., Ishida, J., Kumagai, Y., Aranami, K. et al. (2006) The operational JMA nonhydrostatic mesoscale model. *Monthly Weather Review*, 134, 1266–1298.

Saito, K., Ishida, J., Aranami, K., Hara, T., Segawa, T., Narita, M. et al. (2007) Nonhydrostatic atmospheric models and operational development at JMA. *Journal of the Meteorological Society of Japan*, 85B, 271–304.

Sato, T., Kusaka, H. & Hino, H. (2020) Quantitative assessment of the contribution of meteorological variables to the prediction of the number of heat stroke patients for Tokyo. *SOLA*, 16, 104–108.

Schatz, J. & Kucharik, C.J. (2014) Seasonality of the urban heat Island effect in Madison, Wisconsin. *Journal of Applied Meteorology and Climatology*, 53, 2371–2386.

Smith, D.K.E., Renfrew, I.A., Dorling, S.R., Price, J.D. & Boutle, I. A. (2020) Sub-km scale numerical weather prediction model simulations of radiation fog. *Quarterly Journal of the Royal Meteorological Society*, 147, 746–763.

Smith, D.K.E., Renfrew, I.A., Price, J.D. & Dorling, S.R. (2018) Numerical modelling of the evolution of the boundary layer during a radiation fog event. *Weather*, 73, 310–316.

Teshiba, M., Hashiguchi, H., Uematsu, A., Tanaka, H., Ohmori, Y. & Fukao, S. (2004) Fog observations with a millimeter-wave scanning radar at Miyoshi basin, Japan. *Earth, Planets and Space*, 56, 259–268.

Underwood, S.J., Ellrod, G.P. & Kuhnert, A.L. (2004) A multiple-case analysis of nocturnal radiation-fog development in the Central Valley of California utilizing the GOES night-time fog product. *Journal of Applied Meteorology*, 43, 297–311.

Van der Velde, I.R., Steeneveld, G.J., Wichers Schreur, B.G.J. & Holtslag, A.A.M. (2010) Modeling and forecasting the onset and duration of severe radiation fog under frost conditions. *Monthly Weather Review*, 138, 4237–4253.

van Schalkwyk, L. & Dyson, L. (2013) Climatological characteristics of fog at Cape Town international airport. *Weather and Forecasting*, 28, 631–646.

Yamada, T., Mizuno, H. & Yamamoto, A. (1999) Winter fog forecast for the new Chitose airport:forecast skill using a forecast equation for the onset time of radiation fog (in Japanese). *Journal of Meteorological Research*, 51, 59–66.

Yamamoto, A. (2000) A climatology of disasters by low visibility with dense fog in Japan (in Japanese). *Journal of Japan Society for Natural Disaster Science*, 19, 99–110.

Yamamoto, A. & Oyamada, K. (2000) A local dense fog in Aizu Basin, Japan - a case study on 1 December 1998 - (in Japanese). *Journal of Meteorological Research*, 52, 17–30.

Zhou, B. & Du, J. (2010) Fog prediction from a multimodel mesoscale ensemble prediction system. *Weather and Forecasting*, 25, 303–322.

Zhou, B. & Ferrier, B.S. (2008) Asymptotic analysis of equilibrium in radiation fog. *Journal of Applied Meteorology and Climatology*, 47, 1704–1722.

How to cite this article: Negishi, M., & Kusaka, H. (2022). Development of statistical and machine learning models to predict the occurrence of radiation fog in Japan. *Meteorological Applications*, 29(2), e2048. https://doi.org/10.1002/met.2048

**APPENDIX A**

Ohashi’s method

Fog predictions were performed for the Miyoshi Basin (latitude 34.8026, longitude 132.8573) in Hiroshima Prefecture. Discriminant equation was used as the statistical model. The meteorological elements used were temperature, humidity, and wind speed, and all these meteorological elements used to build the statistical model were obtained from observations. In other words, all the tests were perfect model experiments using the observations. The presence or absence of fog was determined using the images captured by the cameras installed in the basin. The temperature, humidity, and wind speed were obtained using meteorological instruments installed by the authors. The period of observation and prediction by discriminant analysis was from 16 October to 12 December 2009. Additional information can be found in Ohashi et al. (2012).