



Title	Screening priority pesticides for drinking water quality regulation and monitoring by machine learning : Analysis of factors affecting detectability
Author(s)	Kentaro, Narita; Yoshihiko, Matsui; Taku, Matsushita et al.
Citation	Journal of Environmental Management, 326(Part A), 116738 https://doi.org/10.1016/j.jenvman.2022.116738
Issue Date	2023-01-15
Doc URL	https://hdl.handle.net/2115/93920
Rights	© 2022. This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/
Rights(URL)	https://creativecommons.org/licenses/by-nc-nd/4.0/
Type	journal article
File Information	Journal of Environmental Management 2022 Matsui.pdf



Paper submitted to Journal of Environmental Management

Screening Priority Pesticides for Drinking Water Quality Regulation and Monitoring by Machine Learning: Analysis of Factors Affecting Detectability

Kentaro Narita^a, Yoshihiko Matsui^{b,*}, Taku Matsushita^b, and Nobutaka Shirasaki^b

^a Graduate School of Engineering, Hokkaido University, N13W8, Sapporo 060-8628, Japan

^b Faculty of Engineering, Hokkaido University, N13W8, Sapporo 060-8628, Japan

* Corresponding author: Phone & Fax: +81-11-706-7280, Email: matsui@eng.hokudai.ac.jp

1

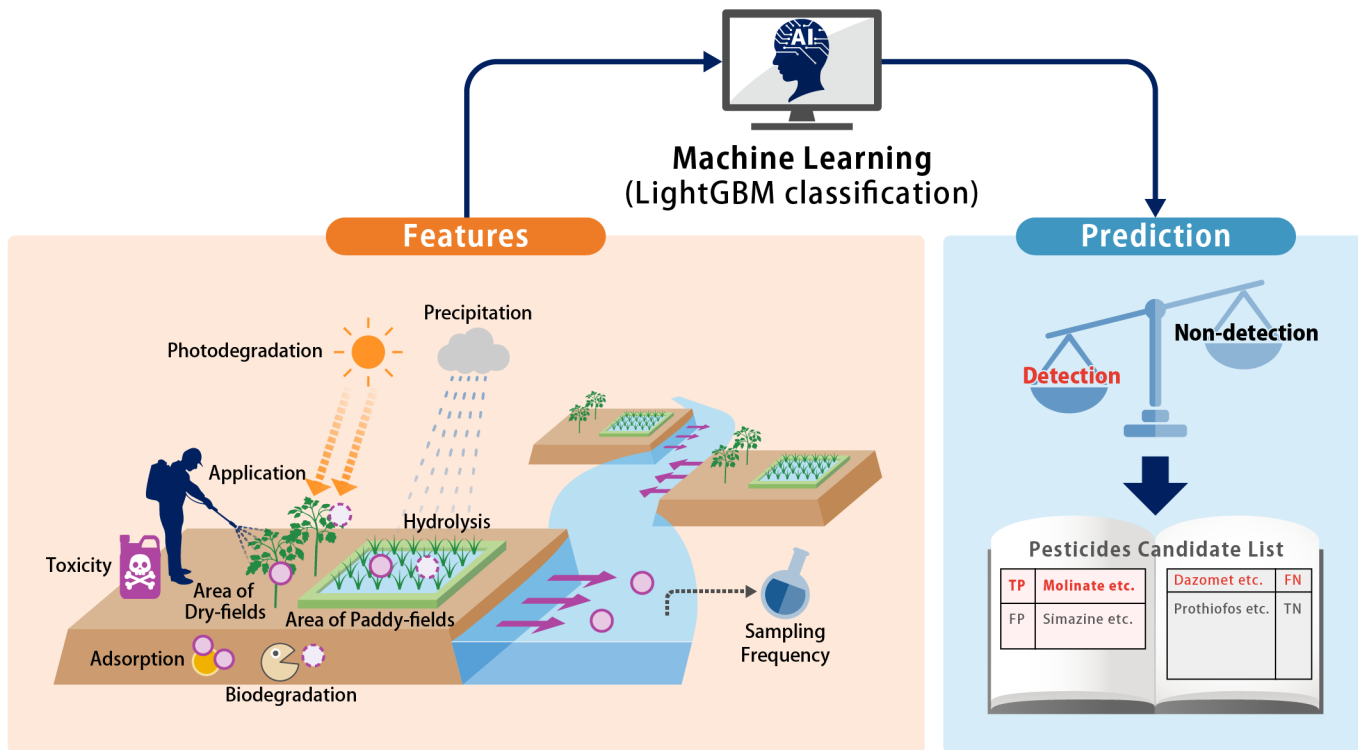
2

3 **Research Highlights**

4

- 5 • Detection/non-detection of pesticides in surface water as a drinking water source
- 6 • LightGBM classification model is superior to other models and score method
- 7 • Key features: frequency of measurements, pesticides to rice paddy, guideline values
- 8 • Existence of threshold/range, e.g. half-life >10-day that increases detectability
- 9 • When detectability is high, other features act synergistically

10



Abstract

Proper selection of new contaminants to be regulated or monitored prior to implementation is an important issue for regulators and water supply utilities. Herein, we constructed and evaluated machine learning models for predicting the detectability (detection/non-detection) of pesticides in surface water as drinking water sources. Classification and regression models were constructed for Random Forest, XGBoost, and LightGBM, respectively; of these, the LightGBM classification model had the highest prediction accuracy. Furthermore, its prediction performance was superior in all aspects of Recall, Precision, and F-measure compared to the detectability index method, which is based on runoff models from previous studies. Regardless of the type of machine learning model, the number of annual measurements, sales quantity of pesticide for rice-paddy field, and water quality guideline values were the most important model features (explanatory variables). Analysis of the impact of the features suggested the presence of a threshold (or range), above which the detectability increased. In addition, if a feature (e.g., quantity of pesticide sales) acted to increase the likelihood of detection beyond a threshold value, other features also synergistically affected detectability. Proportion of false positives and negatives varied depending on the features used. The superiority of the machine learning models is their ability to represent nonlinear and complex relationships between features and pesticide detectability that cannot be represented by existing risk scoring methods.

Keywords: pesticide concentration, guideline, screening, monitoring, prediction, drinking water source.

1. Introduction

Pesticides used in agricultural lands may escape into water systems such as rivers (Jin et al., 2019; Kamata et al., 2020; Panis et al., 2022; Sjerps et al., 2019), making them a chemical group that requires attention in water quality management, particularly for drinking water production. Governments and nongovernmental organizations select certain pesticides and regulate their concentrations in drinking water. The European Union Drinking Water Directive sets the maximum allowable concentration of pesticides in drinking water at 0.5 µg/L in total as well as individually at 0.1 µg/L (EU, 2020). In Japan, pesticides are not set as water quality standard items in the Drinking Water Quality Standards (DWQS), regulatory standards under the Water Supply Law, but are listed in the “Complementary Items to Set the Target for Water Quality Management” (hereafter called the Japanese Drinking Water Quality Guidelines or JDWQG, which are non-enforceable water quality guidelines) to be tested in accordance with the DWQS (MHLWJ, 2022a). They are evaluated using the total pesticide method, which is equivalent to Tier 2 in the chemical mixture exposure assessment of the World Health Organization (WHO, 2017). In the total pesticide method, the detection index (DI) value is defined as the sum of the values obtained by dividing the detection value (DV, detected concentration) of an individual pesticide by the individual target value (guideline value, GV) of that pesticide, and the DI value is supposed to be less than or equal to one (MHLWJ, 2003). In addition, 115 pesticides with a high probability of detection have been selected as a candidate group for examination (pesticide candidate list, formerly called the primary list of pesticides) (MHLWJ, 2022b). It is recommended that water utilities appropriately select pesticides with a high likelihood of detection for testing in water quality examinations by referring to the pesticide candidate list and considering local conditions (MHLWJ, 2003). However, most of these water utilities measure pesticide concentrations only once a year (Kosaka et al., 2013) despite the large seasonal variations in pesticide concentrations (Nakano et al., 2004; Ouyang et al., 2017; Taylor et al., 2021), and they measure all pesticides on the list without selecting pesticides that have a high possibility of detection from the pesticide candidate list. Current pesticide measurement plans therefore are not necessarily based on a sufficient understanding of the

actual presence of pesticides at the water source. On the other hand, the regulator is required to make rational decisions on which pesticides to add and remove from the pesticide candidate list in periodic reviews (Public Welfare Science Council, 2003).

Various prediction methods have been developed to screen pesticides to be regulated as well as those to be examined and monitored, taking into account the toxicity and environmental fate of the pesticides. Some methods are based on actual measurement data (Sang et al., 2022; Tani et al., 2012; Tsaboula et al., 2019b, 2019a), and others predict pesticide concentrations in river water using runoff analysis models (Imaizumi et al., 2018; Matsui et al., 2006; Pullan et al., 2016). However, these methods require a great deal of effort obtain high-frequency measurement data in advance and/or to construct and run detailed simulation models, and thus are therefore not suitable for simple advance prediction of unmeasured pesticide contamination conditions. In this context, several risk scoring methods have been proposed for simple predictive evaluation of the contamination status of unmeasured pesticides; these methods can be used for prioritization of regulation and monitoring targets (Casado-Martinez et al., 2018; Dabrowski et al., 2014; Harmon O'Driscoll et al., 2022; Juraske et al., 2007). The risk scoring method uses available information on degradability, toxicity, and amount used; ranks them by value; gives a score according to the rank; and prioritizes pesticide risks relatively easily based on the total score. However, the ranking system, the setting of scores for each rank, and the calculation method of the overall value are set qualitatively or semi-quantitatively by experts on a case-by-case basis according to the intended use of the method (Vryzas et al., 2020). Furthermore, there are only a limited number of studies that have followed up by comparing the validity of the risk scoring method with actual detection/non-detection and concentrations measured in the environment. Based on the results of pesticide runoff model simulations, we proposed and validated a detectability index to determine the possibility of detecting pesticides in surface water (Narita et al., 2021, 2014). The pesticide candidate list in JDWQG was developed using the detectability index method. On the other hand, the detectability of pesticides may also depend on the frequency of monitoring (Kosaka et al., 2013; Narita et al., 2021). There have been reports of pesticides actually being detected in surface

waters where pesticides had been rarely measured in the past (Asami et al., 2020). Although the detectability index does not include monitoring frequency as an explanatory variable, it may be useful to estimate the detectability of pesticides based on the assumed number of annual pesticide measurements.

The existing risk scoring methods (Casado-Martinez et al., 2018; Dabrowski et al., 2014; Harmon O'Driscoll et al., 2022; Juraske et al., 2007; Yu et al., 2019) and the detectability index method (Narita et al., 2014) are all simple arithmetic methods that calculate a total score or index value using a multiplicative or additive model, but given the development of machine learning methods today, better prediction models can be constructed (Liu et al., 2022). With machine learning, the relationship between objective variables and explanatory variables is constructed as a nonparametric model based on past experience (measured data); for example, machine learning models are being investigated as a method for predicting the concentration of poly- and perfluoroalkyl substances and arsenic in unmeasured locations in the environment (George and Dixit, 2021; Huang et al., 2021; Mohammadi et al., 2022). Studies focusing on pesticides include those that predict the spatial distribution of pesticide concentrations in groundwater or soil (Bedi et al., 2020; X. Wang et al., 2022) and those that predict time-series changes in pesticide concentrations in river water (Yun et al., 2021). However, studies aimed at screening for pesticides with high detection potential, including unmeasured pesticides, have not been conducted despite the high need. In addition, although it is difficult to logically explain the prediction results of machine learning because the model structure is a black box (Molnar, 2022), it is important to ensure they can be explained in order for the model to be widely used by water utilities.

In this study, we did the following with the aim of providing reference information for regulators and water utilities to consider the pesticides to be regulated and measured. 1) We tested several machine learning models for predicting the detectability of pesticides in rivers, selected the best one, and evaluated its effectiveness by comparing its results with those of the detectability index method. 2)

Furthermore, we identified and characterized the main factors associated with the detectability of pesticides.

2. Data and Methods

2.1 Outline of research procedure

A flowchart of the research procedure is shown in Fig. 1. To predict the detectability of pesticides in surface water using a machine learning model, the study was divided into three major phases. First, data for training-validation and testing the model were collected and organized. A classification model and regression model were set up for each of the three modeling algorithms, making a total of six models. These models were trained with training-validation data, and the predictions of pesticide detectability were evaluated using the testing data. The performance of the six models was compared based on several performance metrics, the optimal model was selected, and the explanatory variables that contributed significantly to the prediction results were identified.

2.2 Pesticide concentration and feature value data

2.2.1 Pesticide concentration data

The territory of Japan was divided into 10 regions (Narita et al., 2021, 2014), and the maximum detected concentration of each pesticide in surface water for each region was taken from *Statistics on Water Supply* (JWWA, 2011–2020). Binary variables (detection/non-detection) were used as objective variables in the classification model, and pesticide concentrations were directly used as objective variables in the regression model. This study used data on pesticide concentrations in raw water in the 10-year period from FYs (fiscal years) 2009 to 2018. During this period, the number of pesticides on

the pesticide candidate list ranged from 102 to 123, slightly increasing each year. All pesticides were analyzed by officially authorized and validated methods (See MHLWJ (2022d)). The lower limit of quantification required by the total pesticide method is 1/100 of the individual GV for each pesticide. Thus, for example, even if the GV is 1 µg/L and the measured value is 0.005 µg/L and sufficient after sensitive measurement, the value 0.005 µg/L is not included in the DI because the measured value is less than 0.01 µg/L (1/100 of the GV) (MHLWJ, 2022c). For classification model data, therefore, the pesticide concentration was converted into a binary variable of detection/non-detection, with “detected” being determined when a concentration of more than or equal to 1/100 of the GV was detected and “not detected” when the concentration was less than 1/100 of the GV. Because the required lower limit of the quantification level is uniformly defined according by the GV of each pesticide, the judgment criteria for detection/non-detection do not depend on the particular pesticide measurement method used or laboratory (e.g., water utility) that conducted the test; they are all unified at 1/100 of the GV. Note that the data on pesticides in *Statistics on Water Supply* might contain erroneous records; we used the method of Kosaka et al. to correct any erroneous entries (Kosaka et al., 2017).

The dataset of objective variables used for training, validation, and testing of the model consisted of a total of 10,620 records from 10 years × 10 regions × the number of pesticide types measured per year (101–123). Of these, 8,240 records from 8 of the years were used to train and validate the model. The remaining 2,380 records were used for model testing with reference to previous studies (Jeong et al., 2022; Li et al., 2022). Of the 10,620 records, about 12% (1,303 records) were with “detection”, and 88% (9,317 records) were with “non-detection”. This is considered to be imbalanced for training-validation data, and we interpreted and evaluated the prediction results with that in mind (see section 2.3.2).

2.2.2 Data used for feature values

The following data were used to determine feature values (explanatory variables, Table 1) in the

machine learning models to predict the detectability of pesticides: the amount of pesticide inputs to the watershed, intended usage purpose of pesticides, the percentage of arable land, the degradability of pesticides, the soil adsorption of pesticides, the toxicity of pesticides, the amount of precipitation, and the number of measurements. Although it is desirable to estimate the amount of pesticides actually applied to arable land, it is difficult to determine the amount of pesticides used based on the actual farming operations of a vast number of farmers (MHLWJ, 2020). Therefore, we assumed that the amount of pesticide sold (i.e., the sales quantity) by prefecture in each fiscal year listed in pesticide sales data books (JPPA, 2010–2019) was equal to the amount used in that year, and we aggregated those values for each region and used the aggregate value as the amount of each pesticide used.

Since different crops are cultivated in different regions and pesticide sales differ by the intended use, the use of pesticides may affect prediction results. Therefore, the intended usage purpose of pesticides was added to the features as categorical values. In addition, based on the standard pesticide use for each cropping variety and using the cropped area from the Census of Agriculture and Forestry (MICJ, 2015), pesticide use was divided into the amount used in paddy fields and the amount used in dry fields. Based on the area planted per crop type (MICJ, 2015) and the standard pesticide use per crop type, the pesticide sales quantity was divided into the amount used in paddy fields for rice farming (sales quantity for paddy fields) and the amount used in dry fields (sales quantity for dry fields). Pesticides with low GV_s are therefore more likely to be “detected” even if their measured concentration values are low (Kosaka et al., 2017; Narita et al., 2021). Thus, GV was used as feature value in predicting the detection/non-detection of a pesticide.

The yearly precipitation data by prefecture (MICJ, 2022) were compiled by region. The half-life in soil, half-life in water, and half-life under sunlight were used to denote pesticide degradability, and adsorption was represented by the soil adsorption coefficient normalized by the organic-carbon content (K_{oc}). These values were basically determined from the literature (FSC, 2020; MAFFJ, 2020; MOEJ, 2020; Tomlin, 2006). When different values were given in different sources, the average of the values

was used. When the half-life in soil and K_{oc} were not mentioned in any of the literature, EPISuite (USEPA, 2012) was used to estimate the values. When the value of any of the three types of half-lives was unknown (i.e., missing), the half-life for which a value could be obtained was substituted. The number of measurements for each pesticide was calculated as the sum of the number of measurements in all waterworks in the region from the *Statistics on Water Supply*.

2.3 Methods for predicting pesticide detectability

2.3.1 Machine learning

Of the ensemble machine learning methods that construct highly accurate models by combining multiple decision trees, we used Random Forest (RF) (Breiman, 2001), XGBoost (XGB) (Chen and Guestrin, 2016), and LightGBM (LGB) (Ke et al., 2017). In RF, a number of decision trees (weak learners) are created from randomly extracted data and features using the bootstrap method. The RF classification model integrates the results of each decision tree by taking the majority vote, and the RF regression model outputs the mean value. Examples of its application include the prediction of chemical concentrations in assessments of groundwater (George and Dixit, 2021) and surface water quality (Chen et al., 2020). XGB and LGB are representative methods of the Gradient Boosting Decision Tree, which improve the overall model performance by creating a large number of weak learners in series and repeatedly having the next weak learner correct the errors of the previous weak learner. The main difference between XGB and LGB is the decision tree growth strategy: XGB uses a level-wise (depth-first) strategy, whereas LGB uses a leaf-wise (width-first) strategy. Examples of their application include health risk assessment for soil contamination by chemical substances (M. Wang et al., 2022) and prediction of *E. coli* concentration at beaches (Li et al., 2022).

The above three machine learning models were constructed with two patterns (a classification model in which the objective variable is a binary value of detection/non-detection and a regression model in

which the objective variable is pesticide concentration). The classification model outputs a probability value that takes a value between 0 (non-detection) and 1 (detection), so a threshold value (0.5 or higher for detection in this study) was set for the output, and a binary value of detection/non-detection was used to make a decision. The regression model outputs the pesticide concentration, and in accordance with the total pesticide method, a pesticide concentration of 1/100 of the GV or more was considered as detection and a pesticide concentration of less than 1/100 of the GV was considered as non-detection. The machine learning models were programmed in Python 3.8. RF was implemented from the library schikit-learn (Ver. 1.0.1), and XGB and LGB from the libraries xgboost (Ver. 1.5.0) and lightgbm (Ver. 3.3.1), respectively.

2.3.2 Data handling in model building

In training-validating the models, we used the K-fold cross-validation method to prevent over-training and adopted the model with the highest validation accuracy. The K-fold cross-validation method is a method in which the data are split into K pieces, training is performed on (K-1) pieces, and then validation is performed on the remaining piece. The processes are repeated K times, and K equaled 5 in this study. The hyperparameters of the machine learning models were determined by grid search and preliminary analysis using the Optuna library (Ver. 2.10.0).

In this study, the number of data points for “detection” was smaller than that for “non-detection,” and thus, the objective variable of the machine learning model was imbalanced. It is known that in the case of such imbalanced data, the prediction accuracy of the objective variable with a small number of samples (“detection” in this study) is significantly reduced (Haixiang et al., 2017). Various methods have been proposed to solve this imbalance problem, and we decided to address it by adopting the undersampling method for the classification model and the sampling weight method for the regression model. In this study, the ratio of minority classes (detection) to majority classes (non-detection) was set to 1:1.25 and implemented in the imbalanced-learn library (Ver. 0.8.1). The sampling weight

method is a method of learning by giving weight to a small number of data points (data with a concentration of 1/100 or more of the GV) in the evaluation of the prediction error of the objective variable, thus emphasizing the importance of this group. In this study, the sample weight function and other functions included in each library were used, and the weight was set as the ratio (7.15) of the number of values with concentration $\geq 1/100$ of GV (1,303) to that of the values with a lower concentration (9,317).

2.3.3 Model performance metrics

The prediction results of the models are detection (positive) and non-detection (negative). Prediction results can be classified into four categories according to their truth status: true positive (TP), false positive (FP), true negative (TN), and false negative (FN). “Accuracy,” a representative model performance metric, is defined as the ratio of the number of correct predictions (TP+TN) divided by the total number of predictions (TP+FP+TN+FN). Accuracy is valid when the number of positive and negative classes is approximately equal, but in the case of imbalanced data, it can lead to a misleading result. That is, the model will achieve an overall high level of Accuracy if it is increased only for the dominant class in the dataset. Since the purpose of the model in this study was to accurately predict the “detection” of potentially detectable pesticides, we decided to use indicators that focused on the positive classes (TP, FP, and FN): Recall, Precision, and F-measure. These metrics are defined in equations (1) to (3), respectively.

$$Recall = \frac{TP}{TP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$F\text{-measure} = \frac{2Precision \times Recall}{Precision + Recall} \quad (3)$$

In the case of this study, Recall is extremely important because a model with a low Recall will miss the possibility that a pesticide will be detected. A model with a low Precision will result in many false positives; that is, pesticides that are not likely to be detected will be judged as likely to be detected, which means that the model prediction is not effective for screening pesticide test targets. However, since Recall and Precision sometimes have a trade-off relationship, the F-measure, which is the harmonic mean of the two and is an overall performance metric, was also used.

When evaluating the classification performance of a classification model, ROC (Receiver Operating Characteristic) curves and PR (Precision Recall) curves are sometimes used to check the classification ability by changing the threshold value of a classification model; PR curves are plotted on two axes, Recall and Precision, and are said to be more applicable to imbalanced data than ROC curves (Saito and Rehmsmeier, 2015). Therefore, we used PR curves in the evaluation of the classification models; for quantitative evaluation of the models by PR curves, we used PR-AUC, which represents the area under the curve (AUC) that the PR curve encloses.

2.4 Factor analysis of features using SHAP

We analyzed the sensitivity of each feature to the predictions using SHAP (SHapley Additive exPlanations), an interpretable machine learning method, which has been implemented in the environmental field (Stojić et al., 2019; Wang et al., 2021; Zhang et al., 2022). SHAP is a method to quantify the contribution of multiple features to a single predicted value (a SHAP value), which is a numerical value that quantifies the contribution of a feature as the change in the predicted value when a feature is excluded from the prediction (Lundberg and Lee, 2017). SHAP values were calculated using Tree SHAP (Lundberg et al., 2018) of the shap library (Ver. 0.40.0). A conceptual diagram of the factorial analysis of features using SHAP is shown in Fig. 2. Since 12 types of features are given for each of the 2,380 records used to evaluate the model, $2,380 \times 12$ data points were obtained as SHAP values. The sum of the 12 SHAP values for a single record corresponds to the deviation of the

predicted value for that record (the objective variable, the detection probability value) from the mean of the predicted values for all records. The importance of each feature was interpreted by calculating the contribution of all features in the model as the mean of the absolute SHAP values.

3. Results and Discussion

3.1 Evaluation of machine learning models

3.1.1 Selection of the best model

Six machine learning models (RF, XGB, and LGB each modeled as a classification model [designated with a “c”] and a regression model [designated with an “r”]) were trained by using the training-validation data (detection and feature data from FYs 2009–2016) and constructed to predict pesticide detectability (detection/non-detection in the classification model and pesticide concentration in the regression model). Prediction was performed for each model for FY2017 and FY2018 with testing data to predict the detection/non-detection of pesticides, and the obtained prediction outputs and evaluation data (actual detection results) were combined to calculate a confusion matrix (TP, FP, TN, FN). The prediction models were compared and evaluated using the performance metrics shown in 2.3.3. The best model for predicting pesticide detectability was then selected.

The evaluation results for each model are shown in Table 2. Among the classification models, LGBc had the highest Recall value at 0.81, followed by RFc at 0.79 and XGBc at 0.78, but the difference was small. Precision was the lowest for LGBc (0.36) because it has a trade-off relationship with Recall. The F-measure was highest for RFc (0.52), but the difference between RFc and the other two models was slight. For the regression model, Recall was highest for LGBr (0.85), followed by RFr (0.81) and XGBr (0.62) (the same order as for the classification models). Precision values were XGBr (0.42),

RFr (0.32) and LGBr (0.19), again in reverse order of Recall because of the trade-off relationship. The F-measures were XGBr (0.50), RFr (0.46), and LGBr (0.31). Compared to the classification models, the regression models showed larger differences in their performance metrics values. Because machine learning models are black boxes in terms of model structure, it is difficult to investigate the clear cause of the variation in the prediction performance of the models. However, since the detection data used in this study were imbalanced and biased toward the negative (non-detection) class, one possible cause is that the effect of the imbalanced data could not be completely eliminated, even though measures were taken against it.

The regression model outperformed the classification model for RF and LGB for Recall, but the classification model was higher for XGB. Furthermore, the classification model was higher for Precision and the F-measure. From these results and evaluating the overall effectiveness of the model predictions, the classification model was considered to be superior to the regression model. In other words, when predicting whether or not a pesticide is detected at a certain concentration, it is better to directly predict the detection/non-detection of the pesticide rather than its concentration.

Fig. 3 shows the PR curve and PR-AUC for the prediction results of the classification models; the PR-AUC quantifies the classification performance of the models, and LGBc (0.64) was the best among the three classification models. Based on the above, LGBc, the predictive model that had a high Recall, the highest PR-AUC, and good Precision and F-measure values was considered to be the best prediction model for screening pesticides.

3.1.2 Effectiveness of features and the machine learning models

We assessed the effectiveness of the features for LGBc, which was evaluated as the best machine learning model. Furthermore, its effectiveness was compared with that of the conventional detectability index method (Text S1, SI) (Narita et al., 2014). Here, the machine learning model was

reconstructed using the same features as those used in the detectability index (excluding the number of measurements per year [Nmeas], paddy area fraction [Par], field area fraction [Dar], and use [Use]) in order to compare predictability under identical conditions. The model constructed excluding only Nmeas was designated LGBc_{e1}, and the model constructed excluding Nmeas, Par, Dar, and Use was designated LGBc_{e2}.

In addition, Nmeas was set to be unknown at the time of the prediction (Kosaka et al., 2013; Narita et al., 2021). To determine to what extent Nmeas values affect the detectability of pesticides, the following assumptions were made to obtain the forecast results. Nmeas for the year under testing (FYs 2017–2018) was (1) the same as the number of annual measurements in the previous year, FY 2016, for each pesticide (LGBc_{p1}), and (2) uniformly equal to the average Nmeas (128 measurements/pesticide/year) of all pesticides for the previous fiscal year (FYs 2009–2016, the model training-validation year) (LGBc_{p2}).

Table 3 shows the Recall, Precision, F-measure, and PR-AUC predicted by each LGBc model and the detectability index method. Recall was highest (0.81) for LGBc, which used the correct Nmeas values. LGBc_{p1} had a lower Recall (0.78), and LGBc_{p2} was even lower (0.76). This suggests that the inclusion of Nmeas in the features reduces the number of missed pesticides with high detection potential (false negatives). Furthermore, accuracy is required for the number of annual measurements (a feature value) to ensure prediction accuracy. The Recall of LGBc_{e2} was 0.78, similar to that of LGBc_{e1}, and excluding Par, Dar, and Use did not improve the prediction of pesticide detectability. Compared to LGBc, Recall was lower when Nmeas was omitted from the features in LGBc_{e2}, but there were cases where the Precision, F-measure, and PR-AUC were slightly improved. The number of annual measurements contributed to the improvement of Recall, but its impact on F-measure and PR-AUC was limited.

The prediction results using the detectability index (Narita et al., 2021) were 0.76, 0.28, and 0.41 for Recall, Precision, and the F-measure, respectively, all of which were lower than the prediction

accuracy of LGBc_{e2} (0.78, 0.37, and 0.50, respectively) constructed using the same explanatory feature values as the detectability index method, indicating that the machine learning model was superior in all aspects of prediction accuracy. The detectability index (Narita et al., 2021) is based on a simple multiplicative model representation of the several explanatory variables. Machine learning, on the other hand, is a nonparametric model in which the relationship between the explanatory (feature) and objective variables is not defined a priori (e.g., (Ke et al., 2017)). In machine learning models, complex relationships between multiple feature values and objective variables (detectability), both linear and nonlinear, are incorporated into the model, and the interaction between features can also be reflected, which is evaluated to better reproduce the complex relationships that determine pesticide concentrations that cannot be expressed in conventional additive or multiplicative models. The superiority of the machine learning model was considered to be due to the superiority of the model structure itself over the detectability index method. It could be a better method for predicting the detectability of pesticides in surface water compared to conventional risk scoring approaches, without requiring any special consideration of the pesticide runoff process.

3.2 Interpreting the importance of features using SHAP

The machine learning models based on decision trees used in this study can calculate the importance of features based on the number of times a feature is employed in the training process of the model, but quantitative evaluation of how features influence model outputs is difficult (Strobl et al., 2007). In addition, it is also known that the relative importance of the features decreases when there is multicollinearity among features (Molnar, 2022). To interpret the prediction results of the LGBc model in this study, we calculated SHAP values, and the overall importance of each feature was evaluated by the mean absolute SHAP value: the larger the mean SHAP value, the greater the impact on the model output. As shown in Fig. 4A, Nmeas, Papp, and GV were identified as the top three features; i.e., they had the largest mean SHAP values and made the greatest contributions to the model prediction results. The next most important features were HLS (half-life in soil), Par, and HLWP (half-life in water) for

degradability. The top three features were the same for the RFc and XGBc models, including their rankings (Fig. S1), indicating that these three features are important for prediction, regardless of the model type.

Fig. 4B shows the distribution of SHAP values for each feature along with the size of the feature. Of the top three features, Nmeas and GV have widely distributed SHAP values in both the positive and negative directions, but the relationship between the size of the feature and the positive or negative SHAP value is clear. This suggests that the number of annual measurements and the GVs contribute to either an increase or decrease in detectability, depending on the size of the value, respectively.

In contrast, Papp showed a larger positive spread of SHAP values but a smaller spread in the negative direction. This indicates that when the amounts of pesticides used in paddy fields are small, the decrease in the amount has no significant effect on the decrease in predicted values. Conversely, when the amounts of pesticides are large, the more likely they are to be detected. The distribution of SHAP values for Dapp (pesticide sales quantity for dry fields) showed a similar trend to that of Papp, but its spread was narrower, and the change in SHAP values with the magnitude of the feature value was smaller, indicating that the effect of Dapp was smaller. This is in agreement with the previous finding that pesticides applied to rice paddies are more likely to be detected in river water than pesticides applied to dry fields (Kamata et al., 2020) because they are more likely to be directly discharged into rivers via paddy field drainage channels and other avenues. The effect of the amount of pesticide application was nonlinear in this study, varying with the volume of pesticide application. Although we do not have a physical interpretation of this result, we plan to study it in the future.

3.3 Analysis of the contribution of features

For the four features with the highest mean SHAP values (Nmeas, Papp, GV, and HLS), we visualized the impact of changes in the features on the predictions of the model by using SHAP dependence plots

(Lundberg and Lee, 2017) (Fig. 5). Even when the values of the X-axis features were similar, the Y-axis values (i.e., the SHAP values) varied, indicating the influence of interactions with the other features. The SHAP values for Nmeas are generally in the positive range for pesticides measured more than 100 times/year, indicating that the detectability of pesticides increases with the number of annual pesticide measurements in this range (Fig. 5A1–A3). In fact, previous studies have found that the detection of pesticides is affected by the number of annual measurements (Kosaka et al., 2017; Narita et al., 2021), in agreement with the results of the SHAP analysis. In cases where the amounts of pesticide application are extremely small or the GVs are large, pesticides are either not present in surface water or, if present, have a concentration of less than 1/100 of the GV (the lower limit of quantification). Thus, in these cases, the potential for detection is still considered small, even with a higher number of measurements per year.

The SHAP dependence plots of Papp is shown in Fig. 5B1–B3; the SHAP value of Papp takes a positive value when the pesticide sales quantity is above about 1 tonne/year, but the variance of the SHAP value with respect to Papp was large. This trend was also observed for Dapp (Fig. S2). This indicates that when the amount of pesticide application exceeds a certain amount (about 1 tonne/year in this study), the detectability of the pesticide increases, but its contribution is affected by other characteristics. Although the amount of pesticide use in a watershed has been described as acting linearly or proportionally in conventional quantification methods for pesticide concentration prediction, pesticide detectability, and waterborne exposure risk (Chen et al., 2002; Dabrowski et al., 2014; Narita et al., 2021), the machine learning models indicated that this is not a simple linear or proportional relationship because it is synergistically affected by other characteristics.

The SHAP dependence plots of GV are shown in Fig. 5C1–C3. Small vertical dispersions of SHAP values can be seen, as well as a monotonic decrease with increasing GV, indicating small interactions with other features and a clear, negative relationship with detectability. When focusing on the interaction with Nmeas, there was a tendency for the SHAP value of GV to increase when $GV < 10$

µg/L and N_{meas} was high. Pesticides with small GV_s are set at a lower limit of quantification, and even low concentrations are likely to be detected if they are actually measured. Therefore, an increase in the number of annual measurements will increase the likelihood that water samples will be taken and tested if pesticides are present in the water.

The SHAP dependence plots of HLS are shown in Fig. 5D1–D3; SHAP values were generally in the negative range when HLS was less than about 10 days. Valcke et al. found that pesticides with a half-life in soil of less than 15 days had the lowest score in a score value table for environmental fate (TEKP, Toxicity/Environmental Kinetic Potential) proposed to screen pesticides with high health risks, meaning there was a small exposure risk for those pesticides in the environment (Valcke et al., 2005). Chen et al. proposed an index for the runoff potential of pesticides to surface water (the Surface Water Mobility Index; SWMI), but the index is defined so that the negative impact of the soil half-life on SWMI dominates when the soil half-life is less than 10 days (Chen et al., 2002). In the SHAP dependence plots, the detectability of pesticides decreased when HLS was approximately 10 days or less, showing a trend consistent with these existing risk scoring methods. In the range of HLS exceeding 10 days, the SHAP values dispersed to positive or negative values due to synergetic interactions with other features. There was no clear trend between HLS and SHAP values, and there was no tendency for higher persistence to increase the detection rate. Although pesticide concentrations in groundwater would be expected to be greatly affected by persistence (Sinclair et al., 2006; Yang et al., 2017), the data in this study were pesticide concentrations in river water with relatively short runoff times from the agricultural land to the points of sampling for water quality examination. In this case, persistence above a certain level would not be directly connected to the likelihood of detection of the pesticide.

3.4 Factor analysis on influencing detectability

In the SHAP value table of the prediction results by LGBc, we aggregated the SHAP values for each true/false result to examine the factors affecting those results (Fig. 6). The dominant features for total

SHAP values with respect to TP and FP tended to be the same as the feature importance shown in Fig. 4: Nmeas, Papp, and GV. The percentage contribution of Nmeas to the total SHAP value was 35% for TP and 37% for FP. Because the number of annual measurements contributed almost equally to TP and FP, we considered that the number of annual measurements contributed little to discriminating between TP and FP. This interpretation is in accordance with the fact that the Precision, as defined by TP and FP values, did not decrease by much when the number of annual measurements was removed (as described in section 3.1.2).

Papp and Nmeas contributed equally to the prediction of TN. GV also had a relatively high contribution. Together, the contribution of these three features accounted for about 74% of the total, suggesting that pesticides with low detectability can be explained by Papp, GV, and Nmeas, regardless of pesticide-specific characteristics such as degradability and adsorption. In addition, the trend for FN differed from those of TP, FP, and TN, with GV making a positive contribution to detectability, while the contribution of Nmeas was very small. Degradability (HLWP, HLS, and HLW) accounted for about 76% of the contribution in the negative direction, and Papp also made a small negative contribution. Overall, each of the model features acted on TP, TN, FP, and FN differently.

The pesticides with FN were often pesticides with low pesticide sales quantity for paddy fields, low annual measurement frequency, or high degradability (Table S1). A judgment of undetectable based on these features seems reasonable. For example, pesticides whose pesticide registrations expired before the year in which the pesticide measurements were conducted and that should not have been sold (e.g., pyridafenthion and terbucarb) had zero sales as the feature value, making it reasonable to expect low detectability (they were in fact detected, making them FN). For dazomet, which is highly hydrolyzable, degradation transformation products are also measured in addition to the original pesticide substance during water quality examinations, and the database we used might record the pesticide as detected even when the degradants were detected in some cases (MHLWJ, 2022d). Since this study used degradability as a model feature, implying that high degradability leads to low

detection probability, we considered that pesticides detected as metabolites would likely have low detectability in the model.

Features that could explain the registration expiration and the presence of metabolites were not included in the machine learning models in this study, so it can be said that the models lack features to predict the detectability of these pesticides. The models therefore could be improved if a new database was created after a careful investigation of whether metabolites or the original substance are detected (unfortunately, creating this type of database would be very time consuming). Furthermore, the accuracy of model predictions is affected by the uncertainty of the input data (Dabrowski et al., 2014; Greskowiak et al., 2017), which are feature values (Matsui et al., 2007; Ullucci et al., 2022). In this study, only available data such as pesticide sales and cultivated land area at the regional level were used, and data on pesticide use and cultivated land area upstream in the watershed from the point of measurement of pesticide concentration could not be included as model features, which introduced uncertainty into the model results. In this regard, there is a great potential to improve the accuracy of the predictions by enhancing the features used to predict the likelihood of pesticide detection.

4. Conclusions

Machine learning models for predicting the detectability of pesticides in surface water were developed using statistical data on water supply and published data on pesticide sales and degradability, and their effectiveness was evaluated by comparing their accuracy with an existing prediction method, and factors affecting detectability were analyzed using SHAP. The main conclusions of this study are as follows.

(1) Classification and regression models were constructed for the decision tree-based models RF, XGB, and LGB, and the prediction accuracy of the six models was compared and evaluated. The regression

models outperformed each classification model in Recall, but the Precision and F-measure were higher for the classification models. Overall, the classification models were judged to be superior to the regression model in predicting the detectability of pesticides. Among the classification models, LGBc had a reproducibility of 0.81, a goodness of fit of 0.36, an F-measure of 0.50, and a PR-AUC of 0.64, making it the best model with higher accuracy than RFc and XGBc.

(2) Compared to the prediction using the existing detectability index, which is based on runoff models created in previous studies (Narita et al., 2014), the prediction by LGBce2 was superior in all aspects of Recall, Precision, and F-measure. The superiority of the prediction performance of the machine learning model was considered to be due to not only the larger number of explanatory variables in the model but also the fact that the structure of the model itself is superior in reproducing the complex and nonlinear runoff process of pesticides.

(3) In the SHAP analysis, the number of annual measurements, the pesticide sales quantity for rice-paddy fields, and GV were in the top three features for all classification models (RFc, XGBc, and LGBc) and were important features that explained the prediction results for the models. Pesticide sales quantity for dry fields had a smaller impact than for rice-paddy fields. Each of these features can have a notable effect on increasing the frequency of pesticide detection, but in cases where the likelihood of detection is high (e.g., when the pesticide sales quantity for rice-paddy is high), they can be synergistically influenced by other characteristics.

(4) Recall decreased when the value of the number of annual measurements as a feature was inaccurate or in models where the number of annual measurements was excluded from the model features. The inclusion of the number of annual measurements as a feature helped to reduce FN (missed pesticides with high detection potential). However, the number of annual measurements did not greatly affect Precision, specifically FP. Overall, each of the model features acted on TP, TN, FP, and FN differently.

(5) SHAP dependence plots for the features were created to analyze the effect of changes in the feature values on the predictive values of the model. A positive relationship was found for the predicted value (detectability of pesticides) with the number of annual measurements, pesticide sales quantity for paddy fields, and pesticide sales quantity for dry fields, while a negative relationship was found with the GV. This result also suggests the existence of a threshold value for the pesticide sales, above which the pesticide detectability increases. The detectability of pesticides decreases when HLS is approximately 10 days or less.

(6) Some of the FN results (falsely predicted as undetectable) were reasonable and appeared to be based on inadequate model features, which suggests a room for model improvement.

Overall, we found that machine learning models can be used to evaluate the detectability of pesticides with high accuracy. Machine learning models can represent nonlinear and complex relationships between features and pesticide detection/non-detection that cannot be represented by existing scoring models. The black box nature of machine learning models is a problem when they are used as prediction models for complex and unknown processes, but in this study, we were able to use SHAP to explain to some extent how the values of the model features affect the prediction results. The model results are expected to help users (e.g., regulators and water utilities) make plans about which pesticides to regulate and measure. Since new data are released every year for water supply statistics, it would also be possible to improve the prediction performance of the model by increasing the amount of training data and updating the model periodically.

Acknowledgements

This study was supported by a Health and Labor Sciences Research Grant (19LA1005 and 22LA1007) from the Ministry of Health, Labor, and Welfare, Japan.

References

- Asami, M., Kamata, M., Matsui, Y., 2020. The FY2019 Report of Health and Labor Sciences Research Grant (Research on Health Security Control, 19LA1005): Multi-disciplinary Research on Risk Management of Drinking-water Quality Based on Detection of Pollutants. Ministry of Health, Labor and Welfare of Japan, Tokyo, Japan.
- Bedi, S., Samal, A., Ray, C., Snow, D., 2020. Comparative evaluation of machine learning models for groundwater quality assessment. *Environ. Monit. Assess.* 192, 776. <https://doi.org/10.1007/s10661-020-08695-3>
- Breiman, L., 2001. Random Forests. *Mach. Learn.* 45, 5–32.
- Casado-Martinez, M. del C., Wildi, M., Ferrari, B.J.D., Werner, I., 2018. Prioritization of substances for national ambient monitoring of sediment in Switzerland. *Environ. Sci. Pollut. Res.* 25, 3127–3138. <https://doi.org/10.1007/s11356-017-9082-6>
- Chen, K., Chen, H., Zhou, C., Huang, Y., Qi, X., Shen, R., Liu, F., Zuo, M., Zou, X., Wang, J., Zhang, Y., Chen, D., Chen, X., Deng, Y., Ren, H., 2020. Comparative analysis of surface water quality prediction performance and identification of key water parameters using different machine learning models based on big data. *Water Res.* 171, 115454. <https://doi.org/10.1016/j.watres.2019.115454>
- Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system. *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.* 13-17-Aug, 785–794. <https://doi.org/10.1145/2939672.2939785>
- Chen, W., Hertl, P., Chen, S., Tierney, D., 2002. A pesticide surface water mobility index and its relationship with concentrations in agricultural drainage watersheds. *Environ. Toxicol. Chem.* 21, 298–308. <https://doi.org/10.1002/etc.5620210211>
- Dabrowski, J.M., Shadung, J.M., Wepener, V., 2014. Prioritizing agricultural pesticides used in South Africa based on their environmental mobility and potential human health effects. *Environ. Int.* 62, 31–40. <https://doi.org/10.1016/j.envint.2013.10.001>

616 EU, 2020. Directive (EU) 2020/2184 of the European Parliament and of the Council of 16 December
 617 2020 on the quality of water intended for human consumption.

618 FSC, 2020. (Food Safety Commission of Japan) Risk Assessment Reports of Pesticides [WWW
 619 Document]. URL <https://www.fsc.go.jp/fsciis/evaluationDocument/list?itemCategory=001>
 620 (accessed 6.22.22).

621 George, S., Dixit, A., 2021. A machine learning approach for prioritizing groundwater testing for per-and
 622 polyfluoroalkyl substances (PFAS). *J. Environ. Manage.* 295, 113359.
 623 <https://doi.org/10.1016/j.jenvman.2021.113359>

624 Greskowiak, J., Hamann, E., Burke, V., Massmann, G., 2017. The uncertainty of biodegradation rate
 625 constants of emerging organic compounds in soil and groundwater – A compilation of literature
 626 values for 82 substances. *Water Res.* 126, 122–133. <https://doi.org/10.1016/j.watres.2017.09.017>

627 Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H., Bing, G., 2017. Learning from class-
 628 imbalanced data: Review of methods and applications. *Expert Syst. Appl.* 73, 220–239.
 629 <https://doi.org/10.1016/j.eswa.2016.12.035>

630 Harmon O’Driscoll, J., Siggins, A., Healy, M.G., McGinley, J., Mellander, P.E., Morrison, L., Ryan,
 631 P.C., 2022. A risk ranking of pesticides in Irish drinking water considering chronic health effects.
 632 *Sci. Total Environ.* 829, 154532. <https://doi.org/10.1016/j.scitotenv.2022.154532>

633 Huang, R., Ma, C., Ma, J., Huangfu, X., He, Q., 2021. Machine learning in natural and engineered water
 634 systems. *Water Res.* 205, 117666. <https://doi.org/10.1016/j.watres.2021.117666>

635 Imaizumi, Y., Suzuki, N., Shiraishi, F., Nakajima, D., Serizawa, S., Sakurai, T., Shiraishi, H., 2018.
 636 Development and validation of a simulation method, PeCHREM, for evaluating spatio-temporal
 637 concentration changes of paddy herbicides in rivers. *Environ. Sci. Process. Impacts* 20, 120–132.
 638 <https://doi.org/10.1039/c7em00517b>

639 Jeong, M., Hoon, D., Wook, S., Kim, K., Kim, S., Oa, S., Kim, G., Park, Y., Rae, D., Lee, M., Kang, J.,

640 2022. Identifying the acute toxicity of contaminated sediments using machine learning models.
 641 Environ. Pollut. 312, 120086. <https://doi.org/10.1016/j.envpol.2022.120086>
 642 Jin, X., Liu, Y., Qiao, X., Guo, R., Liu, C., Wang, X., Zhao, X., 2019. Risk assessment of organochlorine
 643 pesticides in drinking water source of the Yangtze River. Ecotoxicol. Environ. Saf. 182, 109390.
 644 <https://doi.org/10.1016/j.ecoenv.2019.109390>
 645 JPPA, 2019. Pesticide directory 2009–2018. Japan Plant Protection Association, Tokyo, Japan.
 646 Juraske, R., Antón, A., Castells, F., Huijbregts, M.A.J., 2007. PestScreen: a screening approach for
 647 scoring and ranking pesticides by their environmental and toxicological concern. Environ. Int. 33,
 648 886–93. <https://doi.org/10.1016/j.envint.2007.04.005>
 649 JWWA, 2020. Statistics on Water Supply 2009–2018, Vol. 92-101. Japan Water Works Association,
 650 Tokyo, Japan.
 651 Kamata, M., Matsui, Y., Asami, M., 2020. National trends in pesticides in drinking water and water
 652 sources in Japan. Sci. Total Environ. 744, 140930. <https://doi.org/10.1016/j.scitotenv.2020.140930>
 653 Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.Y., 2017. LightGBM: A highly
 654 efficient gradient boosting decision tree. Adv. Neural Inf. Process. Syst. 2017-Decem, 3147–3155.
 655 Kosaka, K., Asami, M., Sasaki, M., Matsui, Y., Akiba, M., 2013. Analysis of monitoring programs of
 656 pesticides at water utilities using statistics on water supply. J. Japan Soc. Civ. Eng. Ser. G
 657 (Environmental Res. 69, III_363-III_374. https://doi.org/10.2208/jscejer.69.III_363
 658 Kosaka, K., Asami, M., Sasaki, M., Matsui, Y., Akiba, M., 2017. Relationship Analysis between
 659 Pesticide Monitoring Plan and Their Detection at Waterworks throughout Japan. J. Japan Soc.
 660 Water Environ. 40, 125–133. <https://doi.org/10.2965/jswe.40.125>
 661 Li, L., Qiao, J., Yu, G., Wang, L., Li, H.Y., Liao, C., Zhu, Z., 2022. Interpretable tree-based ensemble
 662 model for predicting beach water quality. Water Res. 211, 118078.
 663 <https://doi.org/10.1016/j.watres.2022.118078>

664 Liu, Y., Yu, J., Sun, H., Li, T., He, X., Lin, Z., 2022. Screening and prioritizing substances in
 665 groundwater in the Beijing–Tianjin–Hebei region of the North China Plain based on exposure and
 666 hazard assessments. *J. Hazard. Mater.* 423, 127142. <https://doi.org/10.1016/j.jhazmat.2021.127142>
 667 Lundberg, S., Lee, S.-I., 2017. A Unified Approach to Interpreting Model Predictions, in: *Advances in*
 668 *Neural Information Processing Systems 30 (NIPS 2017)*. Curran Associates, Inc, pp. 4765–4774.
 669 Lundberg, S.M., Erion, G.G., Lee, S.-I., 2018. Consistent Individualized Feature Attribution for Tree
 670 Ensembles.
 671 MAFFJ, 2020. (Ministry of Agriculture, Forestry and Fisheries, Government of Japan) Pesticide
 672 Assessment Report [WWW Document]. URL https://www.maff.go.jp/j/nouyaku/n_sinsa/index.html
 673 (accessed 6.22.22).
 674 Matsui, Y., Narita, K., Inoue, T., Matsushita, T., 2007. Using precise data sets on farming and pesticide
 675 properties to verify a diffuse pollution hydrological model for predicting pesticide concentration,
 676 *Water Science and Technology*. <https://doi.org/10.2166/wst.2007.437>
 677 Matsui, Y., Narita, K., Inoue, T., Matsushita, T., 2006. Screening level analysis for monitoring pesticide
 678 in river water using a hydrological diffuse pollution model with limited input data, *Water Science*
 679 *and Technology*. <https://doi.org/10.2166/wst.2006.310>
 680 MHLWJ, 2022a. (Ministry of Health, Labor and Welfare of Japan) Drinking water quality standard
 681 [WWW Document]. URL
 682 <https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/topics/bukyoku/kenkou/suido/kijun/kijunchi.html>
 683 #02 (accessed 6.21.22).
 684 MHLWJ, 2022b. (Ministry of Health, Labor and Welfare of Japan) Revision of drinking water quality
 685 standard [WWW Document]. URL
 686 <https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/topics/bukyoku/kenkou/suido/suishitsu/05.html>
 687 (accessed 6.22.22).

688 MHLWJ, 2022c. (Ministry of Health, Labor and Welfare of Japan) Analytical Method for
689 Complementary Items to Set the Target for Water Quality Management [WWW Document]. URL
690 <https://www.mhlw.go.jp/content/10900000/000922150.pdf> (accessed 6.22.22).

691 MHLWJ, 2022d. (Ministry of Health, Labor and Welfare of Japan) Rolling revision of water quality
692 standard [WWW Document]. URL
693 <https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/topics/bukyoku/kenkou/suido/kijun/index.html>
694 (accessed 6.22.22).

695 MHLWJ, 2020. (Ministry of Health, Labor and Welfare of Japan) Shipments of Pesticides by Ingredient
696 by prefecture and region [WWW Document]. URL
697 [https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/topics/bukyoku/kenkou/suido/suishitsu/05_00003](https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/topics/bukyoku/kenkou/suido/suishitsu/05_00003.html)
698 [.html](https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/topics/bukyoku/kenkou/suido/suishitsu/05_00003.html) (accessed 6.22.22).

699 MHLWJ, 2003. (Ministry of Health, Labor and Welfare of Japan) Pesticides in Drinking water quality
700 standard and guidelines [WWW Document]. URL
701 <https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/topics/bukyoku/kenkou/suido/suishitsu/05.html>
702 (accessed 6.22.22).

703 MICJ, 2022. (Ministry of Internal affairs and Communications of Japan) Annual Precipitation by
704 Prefecture (2009-2018) [WWW Document]. URL [https://www.e-stat.go.jp/regional-](https://www.e-stat.go.jp/regional-statistics/ssdsview)
705 [statistics/ssdsview](https://www.e-stat.go.jp/regional-statistics/ssdsview) (accessed 6.22.22).

706 MICJ, 2015. (Ministry of Internal affairs and Communications of Japan) Census of Agriculture and
707 Forestry in Japan [WWW Document]. URL [https://www.e-stat.go.jp/stat-](https://www.e-stat.go.jp/stat-search?page=1&toukei=00500209)
708 [search?page=1&toukei=00500209](https://www.e-stat.go.jp/stat-search?page=1&toukei=00500209) (accessed 6.22.22).

709 MOEJ, 2020. (Ministry of the Environment of Japan) Registration Standard of Pesticide to Prevent Harm
710 to Aquatic Flora and Fauna [WWW Document]. URL [http://www.env.go.jp/water/sui-](http://www.env.go.jp/water/sui-kaitei/kijun.html)
711 [kaitei/kijun.html](http://www.env.go.jp/water/sui-kaitei/kijun.html) (accessed 6.22.22).

712 Mohammadi, M., Naghibi, S.A., Motevalli, A., Hashemi, H., 2022. Human-induced arsenic pollution
 713 modeling in surface waters - An integrated approach using machine learning algorithms and
 714 environmental factors. *J. Environ. Manage.* 305, 114347.
 715 <https://doi.org/10.1016/j.jenvman.2021.114347>
 716 Molnar, C., 2022. Interpretable Machine Learning-A Guide for Making Black Box Models Explainable
 717 [WWW Document]. URL <https://christophm.github.io/interpretable-ml-book/> (accessed 9.22.22).
 718 Nakano, Y., Miyazaki, A., Yoshida, T., Ono, K., Inoue, T., 2004. A study on pesticide runoff from paddy
 719 fields to a river in rural region - 1: Field survey of pesticide runoff in the Kozakura River, Japan.
 720 *Water Res.* 38, 3017–3022. <https://doi.org/10.1016/j.watres.2004.02.013>
 721 Narita, K., Matsui, Y., Iwao, K., Kamata, M., Matsushita, T., Shirasaki, N., 2014. Selecting pesticides for
 722 inclusion in drinking water quality guidelines on the basis of detection probability and ranking.
 723 *Environ. Int.* 63, 114–120. <https://doi.org/10.1016/j.envint.2013.10.019>
 724 Narita, K., Matsui, Y., Matsushita, T., Shirasaki, N., 2021. Selection of priority pesticides in Japanese
 725 drinking water quality regulation: Validity, limitations, and evolution of a risk prediction method.
 726 *Sci. Total Environ.* 751, 141636. <https://doi.org/10.1016/j.scitotenv.2020.141636>
 727 Ouyang, W., Cai, G., Tysklind, M., Yang, W., Hao, F., Liu, H., 2017. Temporal-spatial patterns of three
 728 types of pesticide loadings in a middle-high latitude agricultural watershed. *Water Res.* 122, 377–
 729 386. <https://doi.org/10.1016/j.watres.2017.06.023>
 730 Panis, C., Zanetti Pessôa Candioto, L., Carla Gaboardi, S., Gurzenda, S., Cruz, J., Castro, M., Lemos, B.,
 731 2022. Widespread pesticide contamination of drinking water and impact on cancer risk in Brazil.
 732 *Environ. Int.* 165, 107321. <https://doi.org/10.1016/j.envint.2022.107321>
 733 Public Welfare Science Council, 2003. Revision of Water Quality Standard [WWW Document]. URL
 734 <https://www.mhlw.go.jp/shingi/2003/04/s0428-7.html> (accessed 6.22.22).
 735 Pullan, S.P., Whelan, M.J., Rettino, J., Filby, K., Eyre, S., Holman, I.P., 2016. Development and

736 application of a catchment scale pesticide fate and transport model for use in drinking water risk
 737 assessment. *Sci. Total Environ.* 563–564, 434–447. <https://doi.org/10.1016/j.scitotenv.2016.04.135>
 738 Saito, T., Rehmsmeier, M., 2015. The precision-recall plot is more informative than the ROC plot when
 739 evaluating binary classifiers on imbalanced datasets. *PLoS One* 10, 1–21.
 740 <https://doi.org/10.1371/journal.pone.0118432>
 741 Sang, C., Yu, Z., An, W., Borgen Sørensen, P., Jin, F., Yang, M., 2022. Development of a data driven
 742 model to screen the priority control pesticides in drinking water based on health risk ranking and
 743 contribution rates. *Environ. Int.* 158, 106901. <https://doi.org/10.1016/j.envint.2021.106901>
 744 Sinclair, C.J., Boxall, A.B.A., Parsons, S.A., Thomas, M.R., 2006. Prioritization of Pesticide
 745 Environmental Transformation Products in Drinking Water Supplies. *Environ. Sci. Technol.* 40,
 746 7283–7289. <https://doi.org/10.1021/es0603507>
 747 Sjerps, R.M.A., Kooij, P.J.F., van Loon, A., Van Wezel, A.P., 2019. Occurrence of pesticides in Dutch
 748 drinking water sources. *Chemosphere* 235, 510–518.
 749 <https://doi.org/10.1016/j.chemosphere.2019.06.207>
 750 Stojić, A., Stanić, N., Vuković, G., Stanišić, S., Perišić, M., Šoštarić, A., Lazić, L., 2019. Explainable
 751 extreme gradient boosting tree-based prediction of toluene, ethylbenzene and xylene wet
 752 deposition. *Sci. Total Environ.* 653, 140–147. <https://doi.org/10.1016/j.scitotenv.2018.10.368>
 753 Strobl, C., Boulesteix, A.L., Zeileis, A., Hothorn, T., 2007. Bias in random forest variable importance
 754 measures: Illustrations, sources and a solution. *BMC Bioinformatics* 8.
 755 <https://doi.org/10.1186/1471-2105-8-25>
 756 Tani, K., Matsui, Y., Iwao, K., Kamata, M., Matsushita, T., 2012. Selecting analytical target pesticides in
 757 monitoring: Sensitivity analysis and scoring. *Water Res.* 46, 741–749.
 758 <https://doi.org/10.1016/j.watres.2011.11.036>
 759 Taylor, A.C., Mills, G.A., Gravell, A., Kerwick, M., Fones, G.R., 2021. Passive sampling with suspect

760 screening of polar pesticides and multivariate analysis in river catchments: Informing
 761 environmental risk assessments and designing future monitoring programmes. *Sci. Total Environ.*
 762 787, 147519. <https://doi.org/10.1016/j.scitotenv.2021.147519>

763 Tomlin, C.D.S., 2006. The E-Pesticide Manual. Version 4.0. British Crop Protection Council, Hampshire,
 764 UK.

765 Tsaboula, A., Menexes, G., Papadakis, E.N., Vryzas, Z., Kotopoulou, A., Kintzikoglou, K.,
 766 Papadopoulou-Mourkidou, E., 2019a. Assessment and management of pesticide pollution at a river
 767 basin level part II: Optimization of pesticide monitoring networks on surface aquatic ecosystems by
 768 data analysis methods. *Sci. Total Environ.* 653, 1612–1622.
 769 <https://doi.org/10.1016/j.scitotenv.2018.10.270>

770 Tsaboula, A., Papadakis, E.N., Vryzas, Z., Kotopoulou, A., Kintzikoglou, K., Papadopoulou-Mourkidou,
 771 E., 2019b. Assessment and management of pesticide pollution at a river basin level part I: Aquatic
 772 ecotoxicological quality indices. *Sci. Total Environ.* 653, 1597–1611.
 773 <https://doi.org/10.1016/j.scitotenv.2018.08.240>

774 Ullucci, S., Menaballi, L., Di Giorgi, S., Luini, M., Riva, C., Schlitt, C., Clementi, E., Azimonti, G., 2022.
 775 Pesticides groundwater modelling relies on input data characterised by a high intrinsic variability:
 776 Is the resulting risk for groundwater credible? *Sci. Total Environ.* 839, 156314.
 777 <https://doi.org/10.1016/j.scitotenv.2022.156314>

778 USEPA, 2012. Estimation Program Interface (EPI) Suite [WWW Document].

779 Valcke, M., Chaverri, F., Monge, P., Bravo, V., Mergler, D., Partanen, T., Wesseling, C., 2005. Pesticide
 780 prioritization for a case-control study on childhood leukemia in Costa Rica: A simple stepwise
 781 approach. *Environ. Res.* 97, 335–347. <https://doi.org/10.1016/j.envres.2004.05.011>

782 Vryzas, Z., Ramwell, C., Sans, C., 2020. Pesticide prioritization approaches and limitations in
 783 environmental monitoring studies: From Europe to Latin America and the Caribbean. *Environ. Int.*

143, 105917. <https://doi.org/10.1016/j.envint.2020.105917>

Wang, M., Li, X., Lei, M., Duan, L., Chen, H., 2022. Human health risk identification of petrochemical sites based on extreme gradient boosting. *Ecotoxicol. Environ. Saf.* 233, 113332. <https://doi.org/10.1016/j.ecoenv.2022.113332>

Wang, R., Kim, J.H., Li, M.H., 2021. Predicting stream water quality under different urban development pattern scenarios with an interpretable machine learning approach. *Sci. Total Environ.* 761, 144057. <https://doi.org/10.1016/j.scitotenv.2020.144057>

Wang, X., Yu, D., Ma, L., Lu, X., Song, J., Lei, M., 2022. Using big data searching and machine learning to predict human health risk probability from pesticide site soils in China. *J. Environ. Manage.* 320, 115798. <https://doi.org/10.1016/j.jenvman.2022.115798>

WHO, 2017. Chemical mixtures in source water and drinking-water. World Health Organization, Geneva, Switzerland.

Yang, Y., Lian, X.-Y., Jiang, Y.-H., Xi, B.-D., He, X.-S., 2017. Risk-based prioritization method for the classification of groundwater pesticide pollution from agricultural regions. *Integr. Environ. Assess. Manag.* 13, 1052–1059. <https://doi.org/10.1002/ieam.1950>

Yu, Y., Yu, Z., Xiang, M., Zhou, Z., Hu, G., Zhang, Y., Ma, R., Li, H., 2019. Screening and prioritization of chemical hazards for deriving human health ambient water quality criteria in China. *J. Environ. Manage.* 245, 223–229. <https://doi.org/10.1016/j.jenvman.2019.05.076>

Yun, D., Abbas, A., Jeon, J., Ligaray, M., Baek, S.S., Cho, K.H., 2021. Developing a deep learning model for the simulation of micro-pollutants in a watershed. *J. Clean. Prod.* 300, 126858. <https://doi.org/10.1016/j.jclepro.2021.126858>

Zhang, Z., Huang, J., Duan, S., Huang, Y., Cai, J., Bian, J., 2022. Use of interpretable machine learning to identify the factors influencing the nonlinear linkage between land use and river water quality in the Chesapeake Bay watershed. *Ecol. Indic.* 140, 108977.

808 <https://doi.org/10.1016/j.ecolind.2022.108977>

809

Table 1 Features (explanatory variables) in the machine learning models

	Abbreviation	Data type	Unit
Sales quantity for paddy fields	Papp	Value	tonne/year
Sales quantity for dry fields	Dapp	Value	tonne/year
Half-lives in soil	HLS	Value	Day
Half-lives in water	HLW	Value	Day
Half-lives under sunlight	HLWL	Value	Day
The soil adsorption coefficient normalized by the organic-carbon content	Koc	Value	L/kg
Guideline value	GV	Value	µg/L
Annual precipitation	Preci	Value	km ³ /year
Ratio of paddy area to land area in the region	Par	Value	dimensionless
Ratio of dry area to land area in the region	Dar	Value	dimensionless
Number of annual pesticide measurements	Nmeas	Value	times/year
Applications (herbicide, fungicide, insecticide, or soil fumigant)	Use	Category	—

Table 2 Comparison of prediction performance among the machine learning models. RF, XGB, and LGB denote Random Forest, XGBoost, and LightGBM, respectively. The c denotes the classification model, and r denotes the regression model.

Metrics	Classification model			Regression model		
	RFc	XGBc	LGBc	RFr	XGBr	LGBr
Recall	0.79	0.78	0.81	0.81	0.62	0.85
Precision	0.39	0.37	0.36	0.32	0.42	0.19
F-measure	0.52	0.50	0.50	0.46	0.50	0.31

Table 3 Comparison of the prediction performance among LGBc, two types of modified LGBc models, and the detectability index method.

Metrics	LGBc	LGBc _{p1}	LGBc _{p2}	LGBc _{e1}	LGBc _{e2}	Detectability index
Recall	0.81	0.78	0.76	0.78	0.78	0.76
Precision	0.36	0.36	0.38	0.38	0.37	0.28
F-measure	0.50	0.49	0.50	0.51	0.50	0.41
PR-AUC	0.64	0.61	0.61	0.64	0.66	N/A

LGBc_{p1} was constructed with the Nmeas for the year under testing (FYs 2017–2018) the same as Nmeas in the previous year, 2016, for each pesticide. LGBc_{p2} was constructed with the Nmeas for the year under testing (FYs 2017–2018) uniformly equal to the average Nmeas (128 measurements/pesticide/year) of all pesticides for the previous fiscal year (FYs 2009–2016, the model training-validation year). LGBc_{e1} was constructed excluding only Nmeas; LGBc_{e2} was constructed excluding Nmeas, Par, Dar, and Use.

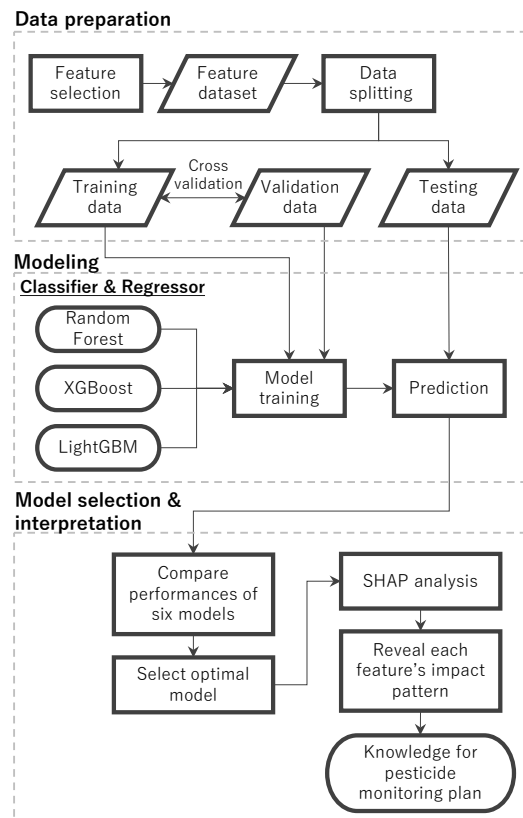


Fig. 1 Flow chart of the research procedure. SHAP: SHapley Additive exPlanations.

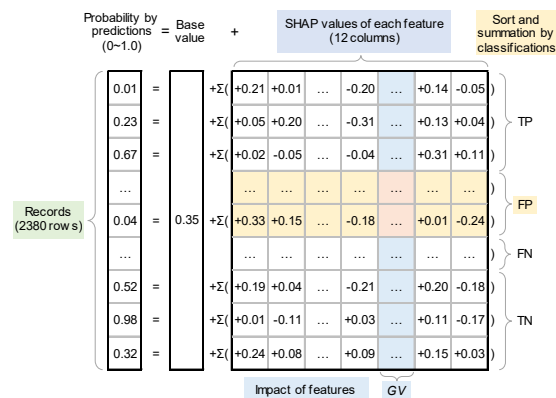


Fig. 2 Concept of factor analysis by the SHAP method

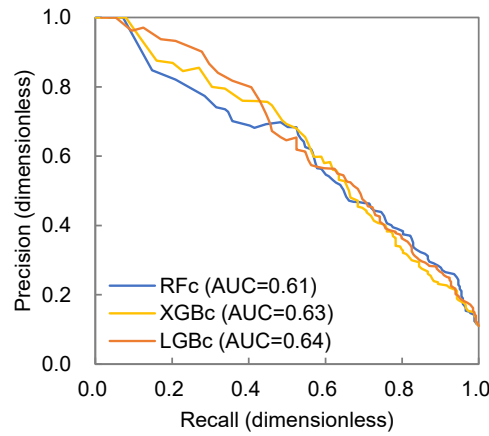


Fig. 3 PR curves of the machine learning models for pesticide detection. RFc, XGBc, and LGBc denote the classification models of Random Forest, XGBoost, and LightGBM, respectively.

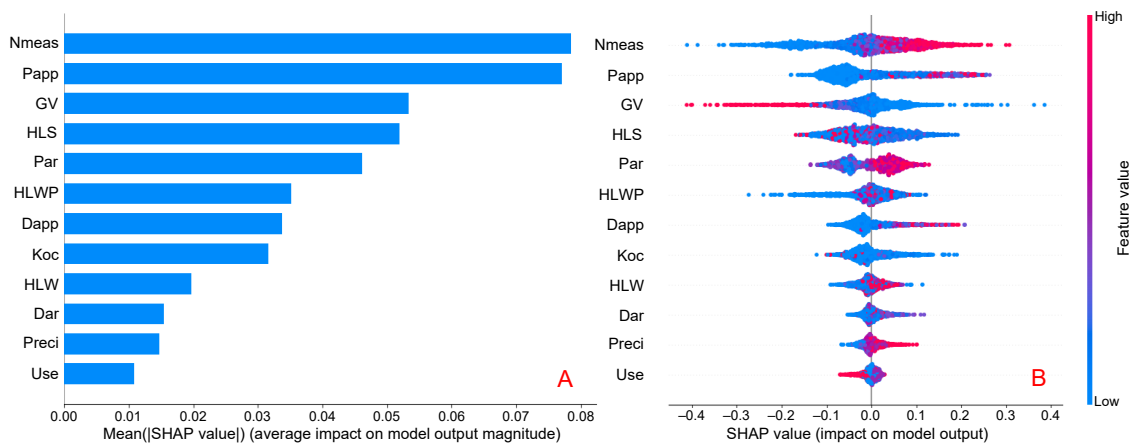


Fig. 4 Importance and contributions of features to the LGBc model prediction: (A) each feature's mean absolute SHAP value over all of the records, and (B) the distribution of SHAP values for all records for each feature. The color of the SHAP value is dependent on the feature value (the greater the feature value, the redder the plot). All features are defined in Table 1.

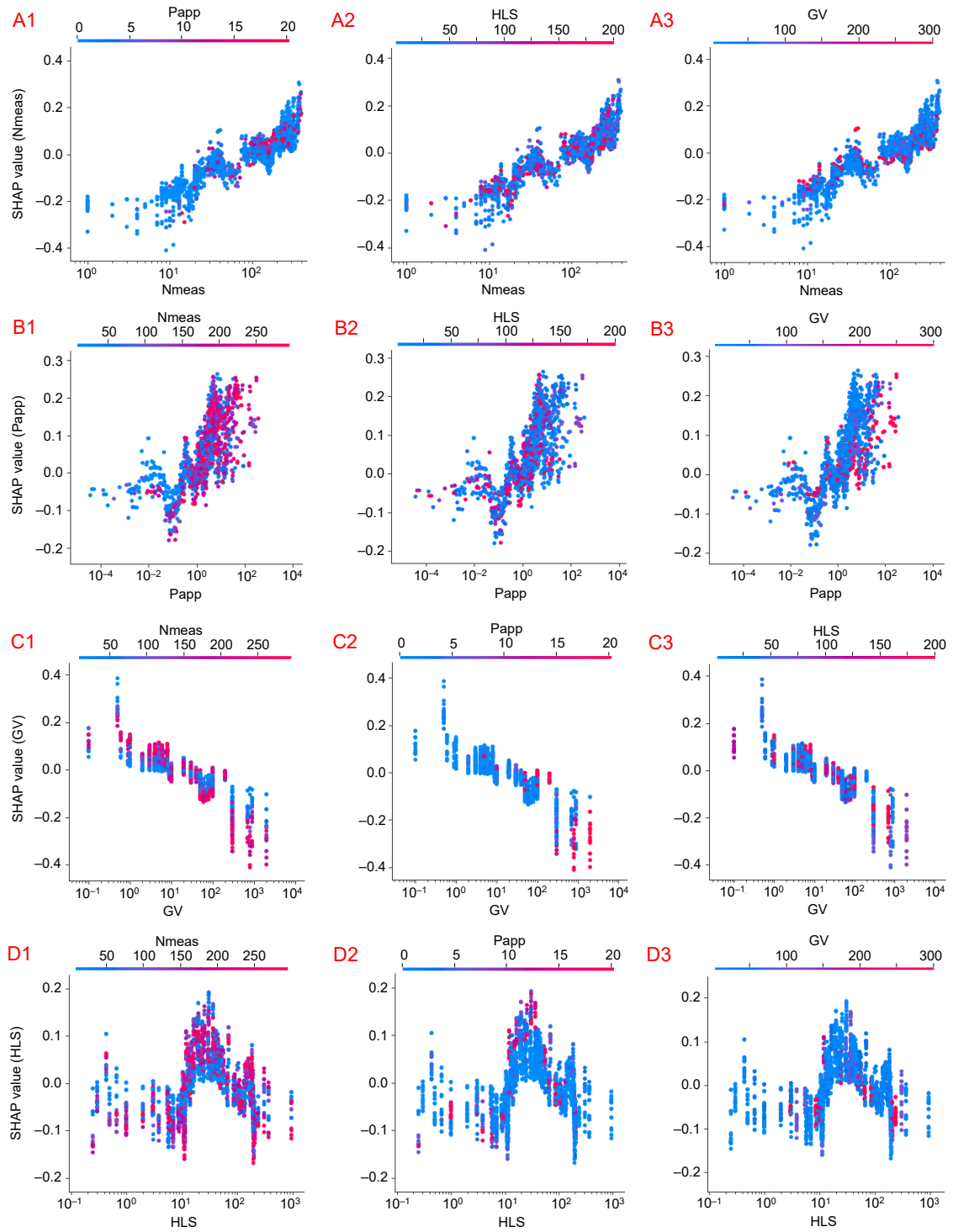


Fig. 5 SHAP dependence plots of the top four features for the probability of pesticide detection. (A1, A2, and A3) SHAP dependence plots of Nmeas; the colors show the magnitude of Papp, GV, and HLS, respectively, to show the influence of each feature on the SHAP value of Nmeas. (B, C and D series) SHAP dependence plots of Papp, GV, and HLS, respectively. In each case, the redder the plot, the greater the feature value.

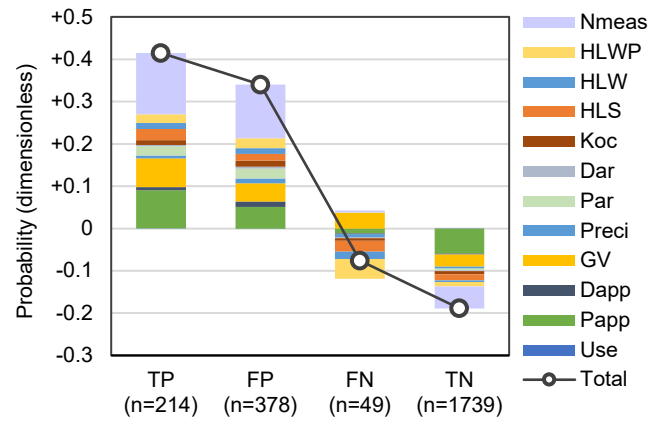


Fig. 6 Total mean SHAP values of the features for TP, FP, FN, and TN (n is the number of records for each feature). The vertical axis represents the prediction probability, which is the sum of the base value and the SHAP value of each feature (see Fig. 2). All features are defined in Table 1.

Supplementary Material

Screening Priority Pesticides for Drinking Water Quality Regulation and Monitoring by Machine Learning: Analysis of Factors Affecting Detectability

Kentaro Narita^a, Yoshihiko Matsui^{b,*}, Taku Matsushita^b, and Nobutaka Shirasaki^b

^a Graduate School of Engineering, Hokkaido University, N13W8, Sapporo 060-8628, Japan

^b Faculty of Engineering, Hokkaido University, N13W8, Sapporo 060-8628, Japan

* Corresponding author: Phone & Fax: +81-11-706-7280, Email: matsui@eng.hokudai.ac.jp

Text S1: Detectability Index

Narita et al. developed a detectability index that evaluates the detectability of each pesticide using pesticide sales quantity by region, GVs, pesticide half-lives, soil adsorption constants, and precipitation as parameters (Narita et al., 2014). The index was applied to pick up unmeasured pesticides that were determined to have high potential for detection. Furthermore, based on their results, the index was validated by investigating whether the pesticides picked up were actually detected or not in newly planned and conducted measurements, evaluating the effectiveness of the detectability index, and improving the index (Narita et al., 2021). In this paper, as a benchmark for evaluating the predictions by the machine learning models, predictions by detectability index were also made, and the validity was compared and evaluated.

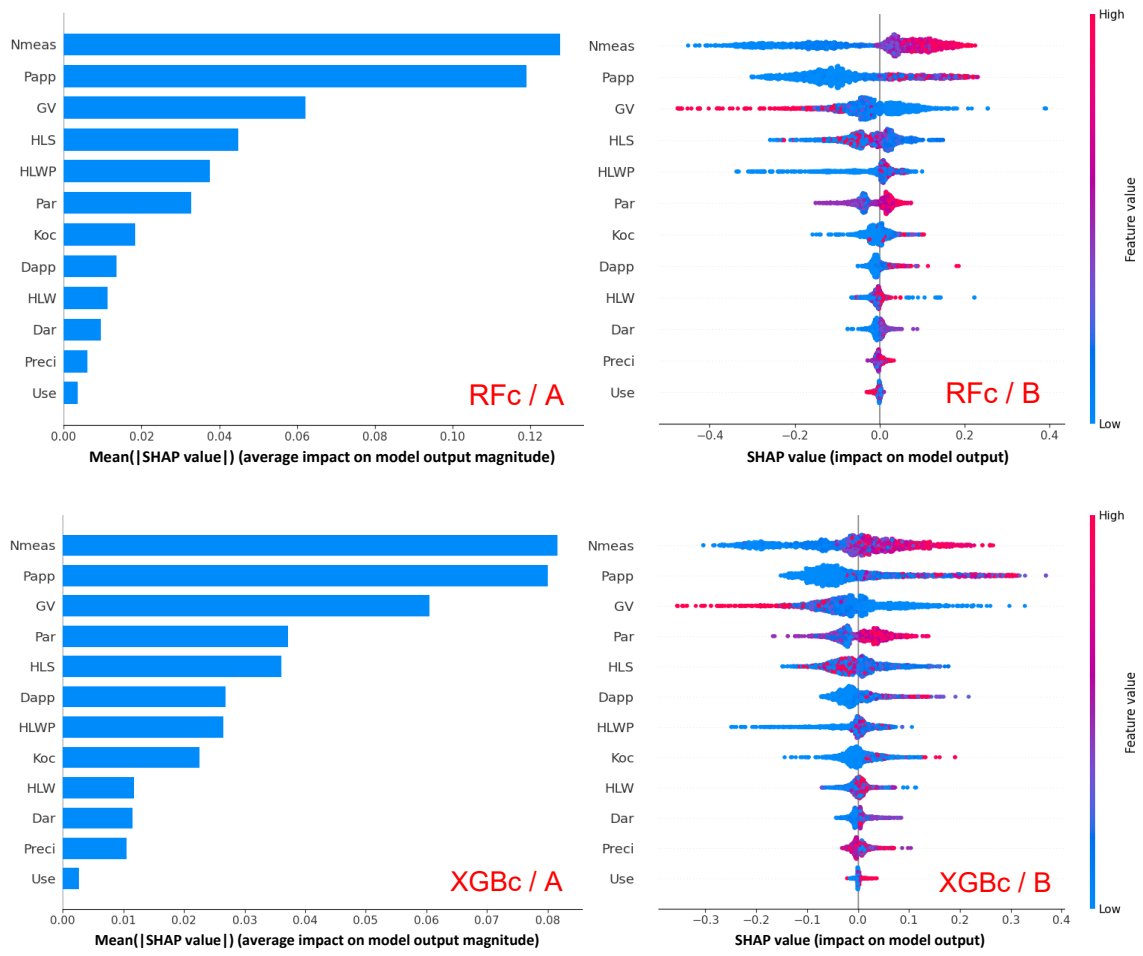


Fig. S1 Importance and contributions of features to the RFc (upper panel) and XGBc (lower panel) model prediction: (A) each feature's mean absolute SHAP values over all records and (B) the distribution of all records' SHAP values, with the SHAP value dots' color dependent on the feature values (the greater the feature value, the redder the plot). All features are defined in Table 1.

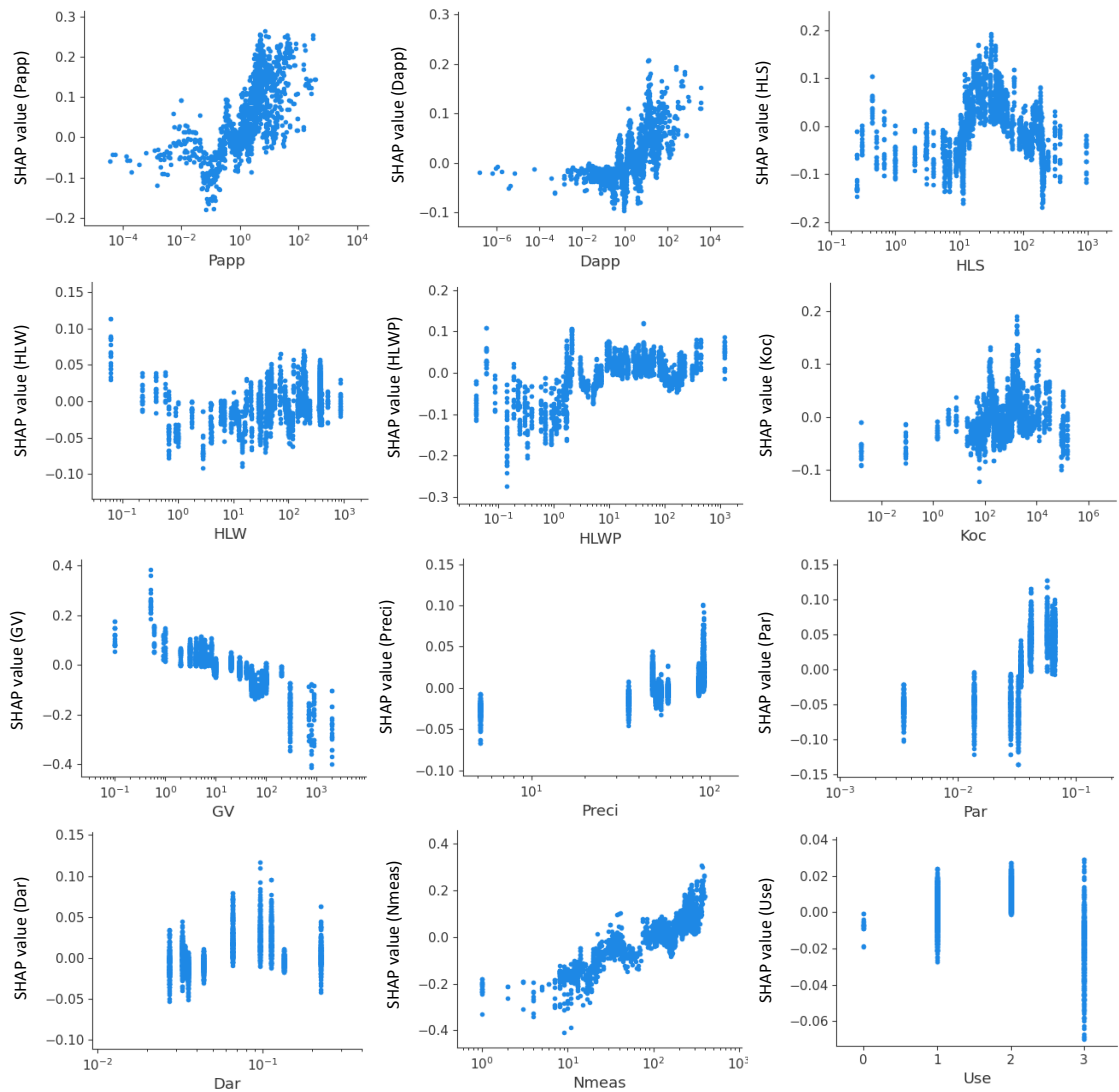


Fig. S2 SHAP dependence plots of all the features for the probability of pesticide detection. All features are defined in Table 1.

Table S1 Records of pesticides determined to be falsely predicted as undetectable (FN).

Fiscal year	Number in the List	Pesticide name	Region	Detection probability value (Range=0–1.0)	Use	Papp (tonne/year)	Dapp (tonne/year)	GV (µg/L)	Preci (km ³ /year)	Par (dimensionless)	Dar (dimensionless)	Koc (L/kg)	HLS (day)	HLW (day)	HLWP (day)	Nmeas (times/year)
2018	112	Methidathion	Kanto	0.495	Insecticide	0.0	9.7	4	52	0.066	0.096	4.8E+2	1.1E+1	4.8E+1	2.2E+1	315
2018	058	Dazomet	Tohoku	0.478	Fungicide	0.0	294.1	10	92	0.057	0.066	3.5E+1	3.0E-1	2.0E+1	8.2E+1	191
2018	089	Buprofezin	Kanto	0.448	Insecticide	1.3	3.6	20	52	0.066	0.096	2.2E+3	1.3E+2	3.8E+2	3.2E+1	232
2018	066	Triclopyr	Kanto	0.439	Herbicide	0.0	4.6	6	52	0.066	0.096	8.5E+1	4.6E+1	4.6E+1	4.6E+1	301
2017	112	Methomyl	Kinki	0.433	Insecticide	0.0	3.9	30	48	0.042	0.033	4.7E+1	6.0E+0	1.6E+2	4.8E+1	276
2018	081	Fenobucarb	Chugoku	0.431	Insecticide	1.7	0.0	30	50	0.034	0.027	2.3E+2	1.0E+1	1.0E+1	1.0E+1	244
2017	109	Fosthiazate	Kanto	0.430	Insecticide	0.0	28.2	3	52	0.066	0.096	5.0E+1	1.3E+1	1.0E+2	1.4E+1	306
2018	093	Prothiofos	Kanto	0.424	Insecticide	0.0	5.0	4	52	0.066	0.096	1.0E+5	4.5E+1	2.3E+2	1.7E+1	194
2017	095	Prothiofos	Kanto	0.422	Insecticide	0.0	5.8	4	52	0.066	0.096	1.0E+5	4.5E+1	2.3E+2	1.7E+1	195
2018	064	Tefuryltrione	Tokai	0.420	Herbicide	5.3	0.0	2	53	0.032	0.044	3.6E+2	1.5E+1	2.1E+1	2.1E+1	82
2017	060	Dazomet	Tokai	0.420	Fungicide	0.0	334.4	10	53	0.032	0.044	3.5E+1	3.0E-1	2.0E+1	8.2E+1	39
2018	040	Chlorpyrifos	Kanto	0.398	Insecticide	0.0	20.0	3	52	0.066	0.096	4.3E+3	1.1E+1	7.2E+1	3.4E+1	375
2018	040	Chlorpyrifos	Kinki	0.372	Insecticide	0.0	3.0	3	48	0.042	0.033	4.3E+3	1.1E+1	7.2E+1	3.4E+1	277
2017	078	Pyridaphenthion	Kanto	0.370	Insecticide	0.0	0.0	2	52	0.066	0.096	3.8E+2	2.1E+1	4.6E+1	2.1E+1	275
2017	037	Glufosinate	Shikoku	0.368	Herbicide	4.9	5.8	20	35	0.028	0.034	2.8E+2	5.7E+0	3.7E+2	1.2E+3	34
2017	042	Cyanazine	Kyushu	0.364	Herbicide	0.0	1.5	4	87	0.040	0.113	1.1E+2	1.4E+1	1.5E+2	2.3E+2	160
2018	003	2,4-PA	Kyushu	0.356	Herbicide	7.7	11.3	20	87	0.040	0.113	5.8E+1	7.0E+0	7.0E+0	7.0E+0	194
2018	049	Dithiocarbamate pesticides	Tokai	0.351	Insecticidal fungicide	0.3	66.0	5	53	0.032	0.044	6.8E+2	1.0E+2	1.0E+2	1.0E+2	80
2018	056	Diazinon	Hokuriku	0.340	Insecticide	0.0	30.5	3	58	0.062	0.035	2.2E+3	1.6E+1	1.6E+1	1.6E+1	159
2017	050	Dithiocarbamate pesticides	Tokai	0.337	Insecticidal fungicide	0.3	58.3	5	53	0.032	0.044	6.8E+2	1.0E+2	1.0E+2	1.0E+2	82
2017	040	Chlorpyrifos	Kinki	0.335	Insecticide	0.0	3.3	3	48	0.042	0.033	4.3E+3	1.1E+1	7.2E+1	3.4E+1	288
2018	058	Dazomet	Tokai	0.308	Fungicide	0.0	302.2	10	53	0.032	0.044	3.5E+1	3.0E-1	2.0E+1	8.2E+1	70
2018	064	Tefuryltrione	Chugoku	0.291	Herbicide	7.2	0.0	2	50	0.034	0.027	3.6E+2	1.5E+1	2.1E+1	2.1E+1	63
2017	066	Tefuryltrione	Chugoku	0.282	Herbicide	5.5	0.0	2	50	0.034	0.027	3.6E+2	1.5E+1	2.1E+1	2.1E+1	67

2018	067	Trichlorphon	Chugoku	0.275	Insecticide	0.0	0.9	5	50	0.034	0.027	7.9E+0	1.5E+2	1.5E+2	1.5E+2	125
2018	037	Glufosinate	Chugoku	0.262	Herbicide	6.5	3.7	20	50	0.034	0.027	2.8E+2	5.7E+0	3.7E+2	1.2E+3	66
2018	042	Cyanazine	Tokai	0.244	Herbicide	0.0	2.4	1	53	0.032	0.044	1.1E+2	1.4E+1	1.5E+2	2.3E+2	104
2018	014	Isoproc carb	Chugoku	0.232	Insecticide	0.0	0.0	10	50	0.034	0.027	3.5E+1	1.2E+1	3.5E+2	4.1E+1	232
2018	051	Cyhalofop butyl	Hokuriku	0.224	Herbicide	6.9	0.0	6	58	0.062	0.035	2.2E+2	2.5E-1	9.7E+1	5.3E+0	120
2017	052	Cyhalofop butyl	Hokuriku	0.224	Herbicide	7.2	0.0	6	58	0.062	0.035	2.2E+2	2.5E-1	9.7E+1	5.3E+0	126
2017	067	Terbucarb	Kanto	0.223	Herbicide	0.0	0.0	20	52	0.066	0.096	5.6E+3	1.5E+2	1.5E+2	1.5E+2	219
2018	048	Disulfoton (Ethylthiometon)	Kanto	0.217	Insecticide	0.0	0.0	4	52	0.066	0.096	2.1E+2	2.0E+2	3.0E+1	1.5E+0	315
2018	048	Disulfoton (Ethylthiometon)	Kinki	0.197	Insecticide	0.0	0.0	4	48	0.042	0.033	2.1E+2	2.0E+2	3.0E+1	1.5E+0	277
2018	088	Butamifos	Kanto	0.194	Herbicide	1.0	11.9	20	52	0.066	0.096	2.1E+3	1.7E+1	3.7E+2	1.4E-1	318
2017	010	Amitraz	Kyushu	0.181	Insecticide	0.0	1.3	6	87	0.040	0.113	8.0E+3	5.0E-1	9.2E-1	5.1E+0	89
2017	066	Tefuryltrione	Tokai	0.171	Herbicide	4.3	0.0	2	53	0.032	0.044	3.6E+2	1.5E+1	2.1E+1	2.1E+1	43
2017	012	Isoxathion	Shikoku	0.168	Insecticide	0.0	0.8	8	35	0.028	0.034	3.1E+4	1.9E+1	4.8E+1	1.9E+0	40
2017	069	Trichlorphon	Shikoku	0.167	Insecticide	0.0	0.7	5	35	0.028	0.034	7.9E+0	1.5E+2	1.5E+2	1.5E+2	38
2018	067	Trichlorphon	Shikoku	0.167	Insecticide	0.0	0.6	5	35	0.028	0.034	7.9E+0	1.5E+2	1.5E+2	1.5E+2	38
2017	062	Thiram	Hokkaido	0.159	Fungicide	0.2	12.9	20	93	0.014	0.225	6.8E+2	1.9E+2	3.7E+2	1.9E+2	24
2018	071	Paraquat	Chugoku	0.143	Herbicide	1.5	1.2	5	50	0.034	0.027	8.8E-2	7.0E+0	3.0E+1	7.0E+0	39
2018	052	Simazine	Shikoku	0.132	Herbicide	0.0	0.1	3	35	0.028	0.034	9.0E+1	4.9E+1	2.8E+1	1.2E+1	36
2018	084	Phenthoate	Kinki	0.132	Insecticide	0.7	2.5	7	48	0.042	0.033	6.4E+2	1.0E+0	1.0E+0	1.0E+0	284
2018	027	Cadusafos	Shikoku	0.123	Insecticide	0.0	0.8	0.6	35	0.028	0.034	2.4E+2	2.5E+1	3.7E+2	1.5E+1	33
2017	073	Paraquat	Chugoku	0.090	Herbicide	1.8	1.3	5	50	0.034	0.027	8.8E-2	7.0E+0	3.0E+1	7.0E+0	43
2018	018	Indanofan	Chugoku	0.074	Herbicide	0.2	0.0	9	50	0.034	0.027	6.3E+2	1.1E+1	1.0E+2	1.2E+1	166
2017	005	MCP	Okinawa	0.027	Herbicide	0.0	0.0	5	5	0.003	0.135	1.6E+2	3.1E+0	1.7E+1	6.3E+0	8
2018	093	Prothiofos	Okinawa	0.010	Insecticide	0.0	2.6	4	5	0.003	0.135	1.0E+5	4.5E+1	2.3E+2	1.7E+1	9
2017	066	Tefuryltrione	Shikoku	0.009	Herbicide	1.9	0.0	2	35	0.028	0.034	3.6E+2	1.5E+1	2.1E+1	2.1E+1	4

References

- Narita, K., Matsui, Y., Iwao, K., Kamata, M., Matsushita, T., Shirasaki, N., 2014. Selecting pesticides for inclusion in drinking water quality guidelines on the basis of detection probability and ranking. *Environ. Int.* 63, 114–120. <https://doi.org/10.1016/j.envint.2013.10.019>
- Narita, K., Matsui, Y., Matsushita, T., Shirasaki, N., 2021. Selection of priority pesticides in Japanese drinking water quality regulation: Validity, limitations, and evolution of a risk prediction method. *Sci. Total Environ.* 751, 141636. <https://doi.org/10.1016/j.scitotenv.2020.141636>