agronomy

MDPI

Article

Parameterization and Calibration of Wild Blueberry Machine Learning Models to Predict Fruit-Set in the Northeast China Bog Blueberry Agroecosystem

Hongchun Qu [1,2,*], Rui Xiang [1], Efrem Yohannes Obsie [2], Dianwen Wei [3] and Francis Drummond [4,5]

1 Institute of Ecological Safety and College of Automation, Chongqing University of Posts and Telecommunications, Chongqing 400065, China; xiangrui154356@gmail.com
2 College of Computer Science, Chongqing University of Posts and Telecommunications, Chongqing 400065, China; l201910008@stu.cqupt.edu.cn
3 Institute of Natural Resources and Ecology, Heilongjiang Academy of Sciences, Harbin 150040, China; wdwzrs@163.com
4 School of Biology and Ecology, University of Maine, Orono, ME 04469, USA; fdrummond@maine.edu
5 Cooperative Extension, University of Maine, 5722 Deering Hall, Orono, ME 04469, USA
* Correspondence: hcchyu@gmail.com

Abstract: Data deficiency prevents the development of reliable machine learning models for many agroecosystems, especially those characterized by a dearth of knowledge derived from field data. However, other similar agroecosystems with extensive data resources can be of use. We propose a new predictive modeling approach based upon the concept of *transfer learning* to solve the problem of data deficiency in predicting productivity of agroecosystems, where productivity is a nonlinear function of various interacting biotic and abiotic factors. We describe the process of building metamodels (machine learning models built and trained on simulation data) from simulations built for one agroecosystem (US wild blueberry) as the source domain, where the data resource is abundant. Metamodels are evaluated and the best metamodel representing the system dynamics is selected. The best metamodel is re-parameterized and calibrated to another agroecosystem (Northeast China bog blueberry) as the target domain where field collected data are lacking. Experimental results showed that our metamodel developed for wild blueberry achieved 78% accuracy in fruit-set prediction for bog blueberry. To demonstrate its usefulness, we applied this calibrated metamodel to investigate the response of bog blueberry to various weather conditions. We found that an 8% reduction in fruit-set of bog blueberry is likely to happen if weather becomes warmer and wetter as predicted by climate models. In addition, southern and eastern production regions will suffer more severe fruit-set decline than the other growing regions. Predictions also suggest that increasing commercially available honeybee densities to 18 bees/m^2/min, or bumble bee densities to 0.6 bees/m^2/min, is a viable way to compensate for the predicted 8% climate induced fruit-set decline in the future.

Keywords: berry crop; fruit-set prediction; machine learning; transfer learning; agroecosystems

Citation: Qu, H.; Xiang, R.; Obsie, E.Y.; Wei, D.; Drummond, F. Parameterization and Calibration of Wild Blueberry Machine Learning Models to Predict Fruit-Set in the Northeast China Bog Blueberry Agroecosystem. *Agronomy* **2021**, *11*, 1736. https://doi.org/10.3390/agronomy11091736

Academic Editor: Andrea Peruzzi

Received: 26 July 2021
Accepted: 26 August 2021
Published: 29 August 2021

1. Introduction

Bog blueberry (*Vaccinium uliginosum* L.), is a Eurasian and North American flowering plant species in the genus *Vaccinium* within the heath family [1,2]. The commercial value of bog blueberries comes from their antioxidant content, which is the highest of 40 common fruits and vegetables [3]. Bog blueberries are widely distributed and managed in the highlands of Northern China [4] as a major agroecosystem supporting the local rural economy. As of 2019, farmers in three provinces in Northeast China manage 3.5 million hectares (15% of the nationwide blueberry coverage) of bog blueberry with an estimated annual yield of 0.6 million metric tons (less than 5% of the national blueberry yield), according to the Heilongjiang Bureau of Agriculture and Forestry [5]. Increasing yield, fruit quality, and long-term economic stability in bog blueberry production requires better

understanding of the fundamental ecological processes of cross-pollination as a result of various interacting factors. The efficiency of several wild blueberry species fruit production, bog blueberry included, depend heavily on cross-pollination by bees [6–10] and weather factors such as temperature and precipitation [11,12]. For example, weather factors not only directly affect phenology and physiology of blueberry plants [13], but also directly affect the activities of bees during pollination, such as foraging duration during bloom. If we can understand the relationships between fruit-set (a proxy to yield) and various interacting factors, from which reliable fruit-set prediction can be made, then more effective management efforts can be initiated accordingly to improve productivity [14]. However, this is not an easy task since fruit-set is usually a highly nonlinear function of plant genetic and morphological traits, bee species composition and density, soil fertility, weather conditions, as well as other spatial and temporal biotic and abiotic factors [15].

To capture these complex relationships, computer simulations and statistical models are often relied upon by investigators [14–18]. Computer simulation models, particularly agent-based simulation models are popular in agricultural research because they can express detailed causalities of interacting ecological processes that are fundamental for nonlinear behavior prediction [19]. However, these approaches need considerable input of knowledge exchange between experts and modelers, and also require extensive computer power to provide meaningful results if the dimensionality of the parameter space is high. Conventional statistical models can fit mathematical relations between variables based on collected data, which are critical for inferring relationships. Unfortunately, statistical models suffer the limitations of representing emerging behaviors and spatial heterogeneities, which usually cannot be ignored in extrapolating nonlinear dynamics in ecosystems. In recent years, machine learning methods, such as Random Forest, Gradient boosted models, K-means, and other techniques, have gained much attention as useful predictive modeling tools for fruit-set and yield estimation [20,21]. This approach takes advantage of algorithms to learn hidden complex patterns in huge datasets without too much human intervention and bias, but still achieve remarkable prediction accuracy [22]. However, machine learning models require large amounts of data for training and testing before reliable results can be achieved [23]. This is a limiting factor in the analysis and prediction of many systems. To overcome this drawback, the marriage of machine learning algorithms with computer simulations, i.e., metamodeling (machine learning models built and trained on simulation data to overcome the obstacle of data limitation), has become an emergent solution for situations where sufficient data cannot be collected for mimicking the highly nonlinear dynamics in agroecosystems [24].

Empirical research in China on bog blueberry is quite limited due to the undeveloped agricultural status of the far Northeast Highlands. Even though large-scale ecological recovery research has been conducted in these areas, systematic field observations on bog blueberry's fruit production in association with bee density, composition and weather conditions are limited [25]. This prevents the building of predictive models that require comprehensive datasets for formulating the structure of the models, and then the subsequent training, and validating necessary for model evaluation. To the best of our knowledge, currently no predictive model has been found worldwide for bog blueberry fruit-set prediction except for computer simulations, but there have been statistical models and metamodels built for the similar blueberry species, *V. angustifolium* [15]. This sibling species called wild blueberry or lowbush blueberry, is intensively managed as a commercial fruit crop in the northeastern United States and the Canadian Maritimes [26]. Blueberry biologists and agronomists have found that bog blueberry shares similar traits to wild blueberry (its sibling species) in ploidy level (tetraploid), low genetic diversity, growth form (rhizomatous shrub being clonal), pollination and reproduction system, as well as sharing suitable climatic conditions [27]. The pioneering research on wild blueberry has encouraged us to investigate the possibility of calibrating the predictive models built and validated on wild blueberry for the prediction of bog blueberry fruit-set.

Our primary goal in this research, therefore, was to test the hypothesis that predictive models successfully used for wild blueberry can be adapted to bog blueberry fruit-set prediction with adjusted parameter estimation and validation on a relatively small dataset. Although the two species are very similar, we still are not clear about whether a non-reciprocal pollen compatibility relationship across different genotypes exists for bog blueberry and what the proportion of plants that are self-compatible is in this species [28]. These relationships are critical for estimating the pollination efficiency and consequently predicting fruit-set [15]. We hypothesized that the pollination simulation model built for wild blueberry captures the main structure of pollination dynamics in bog blueberry. This hypothesis is based upon the similar population reproductive structure that is shared by the two blueberry species [15,28,29]. Both species are primarily heterogamous, meaning that outcrossing is essential, but they also share autogamy (a proportion of self-compatible individuals within populations). Based upon this similar reproductive and genetic structure, the metamodels based upon the wild blueberry simulation model should be able to be used to identify bog blueberry specific patterns in pollination from the limited bog blueberry data that we had available for modeling. Once we successfully transferred the metamodel into bog blueberry fruit-set prediction, we applied it as a tool to evaluate the response of Northeast China's bog blueberry agroecosystem to weather factors, which have been regarded as the most important abiotic variables affecting blueberry fruit-set (a proxy to yield) [11,12]. In addition, we also investigated the viable management strategies that could likely offset the negative effects of climate change on bog blueberry productivity.

2. Materials and Methods

2.1. Overview

Our study combines field observation and computer modeling (Figure 1) to produce reliable predictions of bog blueberry's fruit-set and utilize them to predict the productivity of Northeast China's bog blueberry agroecosystem in response to various weather conditions and predicted climate change. A simulated dataset previously generated by the wild blueberry pollination simulation model [15] was employed to develop several machine learning algorithms. We use the terms "machine learning algorithm" and "metamodel" interchangeably in this paper throughout the several steps of our analytical approach: feature selection, model formulation and selection. Once the best metamodel with the highest prediction accuracy was selected, its parameters were then estimated by being calibrated to the bog blueberry fruit-set data collected from 6 fields in Northeast China in 2015 and 2016. Then the calibrated metamodel was validated against a second independent bog blueberry fruit-set data collected at the same 6 fields from 2017 to 2019, which had not been used for metamodel construction. After validation, the metamodel was used for three experiments to predict bog blueberry fruit-set in Northeast China under several weather conditions. The geographical pattern of fruit-set change, as well as, potential management strategies to mitigate negative effects of climate change were also investigated as model applications.

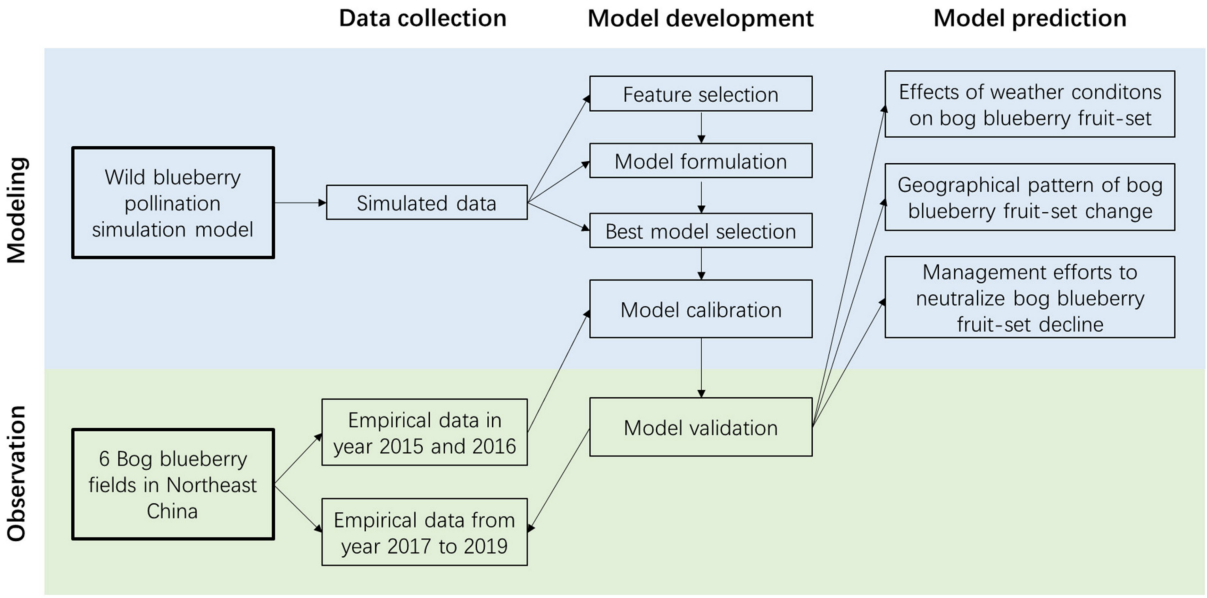


Figure 1. Flow diagram of the study's stages and tasks.

2.1.1. Study Area

We predicted the fruit-set of bog blueberries in 151 growth regions in three provinces: Heilongjiang, Jilin and Inner Mongolia in Northeast China. These regions range from 116.49° E to 132.53° E, and from 40.86° N to 52.58° N, as shown in Figure 2. Each black dot represents the geographical coordinates of the meteorological station in each of the 151 production areas. Yellow triangles are locations of the 6 bog blueberry fields where fruit-set data was used for metamodel calibration (2015–2016) and validation (2017–2019). These sites were Mohe (MH), Tahe (TH) Jiamusi (JMS) and Yichun (YC) in Heilongjiang province, Genhe (GHS) in Inner Mogolia Autonomous Region and Changbai Mountain area (CBS) in Jilin province.

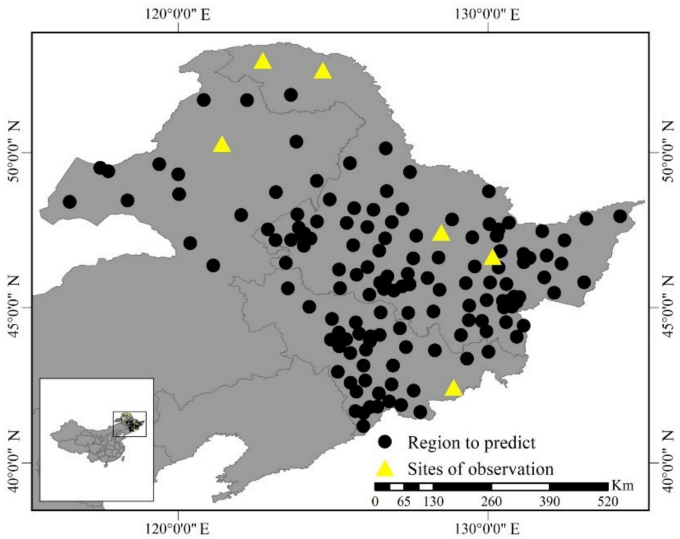


Figure 2. Study area in Northeast China for bog blueberry fruit-set prediction.

2.1.2. Datasets and Simulation Data

Data used for metamodel development was downloaded from a publicly accessible wild blueberry fruit-set dataset [30], see Mendeley Data: https://data.mendeley.com/datasets/p5hvjzsvn8/1, accessed on 25 August 2021, which was generated by the Wild Blueberry Pollination Simulation Model [15]. The dataset contains 777 records of simulated

wild blueberry fruit-set, each of which is an average of 100 replications of simulation experiments. These simulations were conducted by varying the following parameters: the average clone size (CS) in square meters within a wild blueberry field; pollinator density (common taxa) in bees per square meter per minute for Honeybee (HB), Bumble bee (BB), Andrena (AD) and Osmia (OS); the highest point (MaxUTR), the lowest point (MinUTR) and the average (AvUTR) of the upper range of the daily air temperature along the blueberry production season; the highest point (MaxLTR), the lowest point (MinLTR), the average (AvLTR) of the lower range of the daily air temperature during the blueberry production season; the total number of rainy days (RD, or precipitation are used interchangeably in this paper) during the bloom season when the daily precipitation is higher than 2.5 cm; and the average number of days that rain (AvRD) during the bloom season. Table 1 summarizes the range of parameter values that the simulations covered. These ranges represent typical wild blueberry spatial traits, bee pollinator density and four possible climatic trends around current weather conditions, which were Warm and Dry, Warm and Wet, Cool and Dry, Cool and Wet, respectively. In other words, the simulation data contain patterns that reveal the relationships between fruit-set and the features of plant, pollinator, and weather conditions.

Table 1. Summary of simulated wild blueberry fruit-set data [30].

Parameter (Feature)	Number of Records	Unit	Range	Mean
Clone size (CS)	777	m^2	10~40	18.768
Honeybee (HB)	777	bees/m^2/min	0~18.43	0.417
Bumble bee (BB)	777	bees/m^2/min	0~0.585	0.282
Andrena (AD)	777	bees/m^2/min	0~0.75	0.469
Osmia (OS)	777	bees/m^2/min	0~0.75	0.562
MaxOfUpperTRange (MaxUTR)	777	°C	20.9~34.8	27.9
MinOfUpperTRange (MinUTR)	777	°C	3.9~14	9.8
AverageOfUpperTRange (AvUTR)	777	°C	14.6~26.1	20.4
MaxOfLowerTRange (MaxLTR)	777	°C	10.1~20.1	15.2
MinOfLowerTRange (MinLTR)	777	°C	−4.3~0.6	−1.8
AverageOfLowerTRange (AvLTR)	777	°C	5.1~13.3	9.2
RainDays (RD)	777	day	1~34	18.309
AverageRainDays (AvRD)	777	day	0.06~0.56	0.32

2.1.3. Empirical Data

From 2015 to 2019, researchers at the Institute of Natural Resources and Ecology (Heilongjiang Academy of Sciences, China) recorded bog blueberry fruit-set at the clone level (Figure 3) in six locations (see Section 2.1 Study area) [25]. Fruit-set was calculated as the total number of viable fruits in one clone (genet or genetically unique plant) at harvest divided by the number of flower buds before bloom in the same clone. Due to uncertainties of the long duration of these field observations, sample sizes in each year and at each field site varied from 58 to 300. Unfortunately, no plant spatial traits or bee visitation rates had been recorded together with the fruit-set observations. In addition to fruit-set, historical weather data including precipitation and air temperature from 2015 to 2019 were collected from the China National Meteorological Center (http://data.cma.cn, accessed on 25 August 2021) at the geographical location of each field site. The six field observations on bog blueberry fruit-set are included in Appendix A.

(**A**) (**B**)

Figure 3. A bog blueberry field site (**A**) in Amul, Mohe (MH) and one of the sampled clones (**B**). Photos were taken by Dr. Dianwen Wei.

2.1.4. Feature Selection

In the process of developing a metamodel for fruit-set prediction, there may be features that are irrelevant or of minor importance. The contribution of some of the types of features can be determined to be minimal as compared to the most important features obtained as the result of feature selection. The total 13 features which were obtained from the publicly available dataset [30] of the Wild Blueberry Simulation Model have a different unit and value range and therefore, we normalized them into a common scale (e.g., between 0 and 1) without distorting differences in the ranges of values. We then used five machine learning methods that have their own built-in feature selection function: Decision Tree (DT), Random Forest (RF), Gradient Boost Decision Tree (GBDT), AdaBoost, and Extreme Gradient Boosting (XGBoost). All five methods were employed to select the subset of features having the highest coefficient of determination (R^2). Statistically, the coefficient of determination (R^2) explains the proportion of variance in the predicted result that is explained by the features. We iteratively tested the coefficient of determination of the five machine learning methods on all subsets of features, in which the number of features in a subset increased from 1 to 13 [31]. It was observed that the XGBoost method achieved the highest R^2 value (0.929) among the five machine learning algorithms when the number of features reached eight (8) (Figure 4). The eight (8) feature subset selected by XGBoost were CS, HB, BB, AD, OS, MaxUTR, MinUTR, and RD. These eight (8) features were used for developing the metamodels.

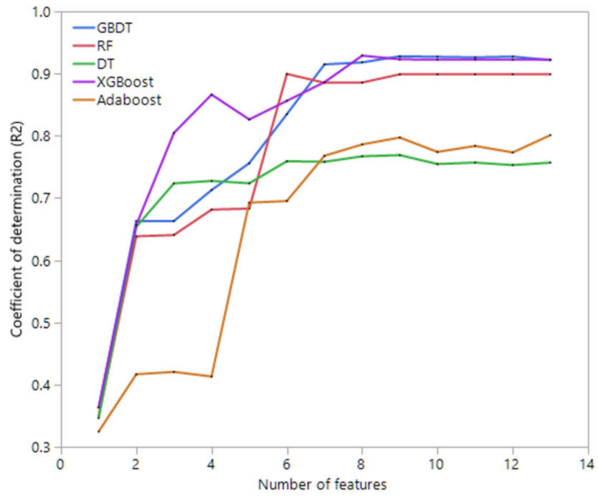


Figure 4. The number of features and their associated coefficient of determination (R^2) for the five machine learning algorithms DT, RF, GBDT, AdaBoost, and XGBoost.

2.2. Model Development

The Python 3.6 software development environment was employed to develop nine metamodels from the simulated wild blueberry dataset [30], in which wild blueberry fruit-set was the dependent (response) variable, and the eight (8) selected features (see Section 2.4) were used as independent (explanatory) variables.

2.2.1. Multiple Linear Regression (MLR)

MLR is a statistical modeling technique that involves predicting a numeric value given multiple independent variables. MLR models have been used extensively in the agricultural research field [32] to develop predictive models that assume a linear relationship between more than one explanatory variable and a response variable by fitting an additive linear equation to observed data. In this study, the goal of using multiple linear regression is to model the linear relationship between the eight (8) selected features and wild blueberry fruit-set, which is given by the general form as in Equation (1):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon \quad (1)$$

where Y is wild blueberry fruit-set in the ith sample, i.e., the dependent variable; X_k corresponds to the independent variables (CS, HB, BB, AD, OS, MaxUTR, MinUTR, and RD); β_0 is the equation intercept; β_k is the corresponding linear coefficient of the k^{th} independent variable; and ε is random error. The training process is responsible for determining the most likely coefficients for the MLR model from learning patterns in the dataset (see Section 2.3).

2.2.2. Support Vector Regression (SVR)

The Support Vector Machine (SVM) algorithm was first introduced by Vladimir Vapnik and his colleagues [33] and is a class of supervised machine learning algorithms for classification problems. Later on, Drucker et al. [34] proposed Support vector regression based on the concept of Vapnik's support vectors. The main purpose of SVM regression is to represent complex relationships through nonlinear mapping. Residual error is minimized by adding a hyperplane and maximizing the distance between predicted and observed values. Initially, the variables are modeled from the original space to a high-dimensional feature space by a kernel function, where the kernel functions can be (linear, polynomial, Gaussian, etc.) and depend on the relationship between the explanatory and response variables. Then, a linear model is built in the derived feature space to minimize error, where it becomes linearly separable [35]. The main hyperparameter fine-tuning of the SVM model includes a kernel function, cost parameter(C), gamma (γ) and the impact of regularization. In this study, the performance of different kernel functions was compared, and finally a Gaussian kernel function was selected. This significantly reduces the risk of overfitting and improves generalization.

2.2.3. K-Nearest Neighbor (K-NN)

Initially, K-NN was used only for classification. However, over the past few decades, this model has also been used for both classification and regression modeling. The K-NN algorithm is a nonlinear instance-based machine learning method [36] which is based on the distance of the predictor variables to the nearest training group known to the model [37]. When K-NN is applied to solve regression problems, the value of the response is calculated as a weighted sum of the responses of all the k neighbors, in which the weight is inversely proportional to the distance from the input record [38]. The neighbors' distance can be calculated by Euclidean, Manhattan, and Minkowski distance formulas. However, in this

study, the Minkowski distance formula was chosen by comparing the performance of different distance measurement metrics, given by Equation (2):

$$\left(\sum_{i=1}^{k} \left| X_i - Y_i \right|^q \right)^{\frac{1}{q}} \quad (2)$$

where k is the number of nearest neighbors, x_i and y_i are the distance between two points, and q is a real value between 1 and 2.

2.2.4. Decision Trees (DT)

Decision Trees (DTs) are non-parametric algorithms that fall under the category of supervised machine learning algorithms and can handle large and complex datasets effectively without a complex parameter structure [39]. Decision tree regression assesses the features of an object and trains a model in the tree structure to predict data to produce meaningful results in the future. In this study, we used the C4.5 algorithm [40] to train the DT model for predicting wild blueberry fruit-set.

2.2.5. Random Forest (RF)

Random Forest (RF) first introduced by Breiman [41], is one of the most widely employed ensemble of machine learning methods which create multiple regression trees that are generated by a large set of decision trees for computing regression models [42]. When compared to using a single decision tree which often creates an unstable model, RF makes predictions by combining predictions from several decision trees using Bootstrap aggregation or a Bagging technique [43]. Bootstrapping involves random sampling of data with replacement and has been proved to effectively reduce and control variance (overfitting) of a predictive model. Random forest training includes training for each decision tree on a randomly selected subset of features and data. Then, the final prediction result is obtained by majority vote or averaging the outputs from each of the sub-models. This can significantly improve the predictability of the RF model in terms of accuracy and generalizability. With respect to handling noisy data, Random forests are more robust. In addition, Random Forests perform well at capturing tabular data with numeric features and maintaining nonlinear interactions between the response variable and the predictors [44].

2.2.6. Adaptive Boosting (AdaBoost)

AdaBoost regression is a machine learning meta-algorithm proposed by [45] that begins with fitting a regressor on the training dataset and then adds additional replications of the original regressor on the same dataset, but with performance improvement based on error information collected from its predecessors [46]. In this way, a final model based upon a strong learning basis can be expected. To improve performance, the algorithm can be used in conjunction with many other learning algorithms.

2.2.7. Gradient Boosting Decision Tree (GBDT)

Jerome Friedman [47] in 1999 first introduced GBDT. This method is a popular machine learning algorithm that optimizes the predicted value of a model in a series of steps during the learning process. With the aim of minimizing the loss function, every single iteration of the decision tree involves adjusting the values of the coefficients, weights, or biases employed on each of the input variables used to predict the target value. There are four main benefits to applying GBDT for predictive model development: (1) feature selection is inherently performed during the learning process; (2) not prone to use of collinear or identical features; (3) models are relatively easy to interpret, and (4) it is easy to specify different loss functions. Hence, in order to solve predictive problems in both classification and regression domains, GBDT is among the most widely used methods used in machine learning.

2.2.8. Artificial Neural Networks (ANN)

Artificial Neural Networks represents a network model for emulating a biological neural system which consists of input, hidden, and output layers [48]. The input layer directly receives the data, whereas the output layer produces the required output. The layers between the input and output layers are hidden layers where the intermediate computation takes place. The process of creating a neural network begins with the perceptron. The perceptron receives the inputs which correspond to the independent variables, and the neurons in the hidden layer multiplies the inputs by weights based upon multiple passes of the data. The results are then transferred to the output layer via an activation function (such as relu, tanh, sigmoid) [49,50]. In this study, we adopted a three-layer perceptron ANN having the activation function "ReLu" (because of its computational efficiency) [31] in the hidden layers, which uses Adam as the gradient descent method. The activation function used in our study is the non-linear rectified linear unit (ReLu) function, which has output 0 for input less than 0 and raw output otherwise. It is mathematically expressed in Equation (3).

$$R(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases} \tag{3}$$

2.2.9. eXtreme Gradient Boosting (XGBoost)

In recent years, many agricultural researchers have been using XGBoost to predict crop yields [30]. The eXtreme Gradient Boosting (XGBoost) method is an ensemble machine learning algorithm based on the principles of the gradient boosting decision tree method. With this method boosted trees are efficiently constructed and can automatically operate parallel computation 10 times faster than gradient boosting [51]. It supports supervised machine learning algorithms, including classification, regression and ranking. For better performance, XGBoost provides three additional features compared to GBDT. First, the weights of each new tree can be scaled down by given constant leaves which reduces the influence of a single tree on the final score. Second, introduction of a regularized loss function avoids the problem of overfitting. Third, the Taylor expansion method is used to approximate the loss function thereby speeding up the process of optimization [52]. The predicted values of the tree ensemble model are computed using the formulae in Equations (4) and (5).

$$\begin{gathered} \hat{y}_i = \varnothing\,(x_i) = \sum_{k=1}^{n} k = f_k(x_i), f_k \in F\# \\ F = f(x) = w_{q_{(x)}}\left(q\ : R^m \to T, w \in R^t\right) \end{gathered} \tag{4}$$

Equation (4) represents the regression tree space where x_i is the input, $\hat{y}_i$ is the output, q the tree structure, and the number of leaves in the tree are identified as T. Each f_k matches the independent tree structure q and the weight w, where w_i represents the score of the i^{th} leaf. This predicted value can be evaluated by:

$$L(\varnothing) = \sum_i l(\hat{y}_i\,, y_i) + \sum_k \Omega(f_k) \tag{5}$$

$$where, \quad \Omega(f) = y^T + \frac{1}{z}\lambda w^2$$

L is the loss function which is the difference between the predicted value $\hat{y}_i$ and the response value y_i. The entity Ω represents the complexity of the model and is a regularization term that has the function of smoothing the weight to avoid overfitting.

2.3. Model Evaluation and Selection

We employed a widely used statistical approach, cross-validation, to address the problem of overfitting, in which we built models using 5-fold cross validation methodology. The five folds were divided into training (four-folds) and testing (one-fold) sets. The cross-validation was repeated in n rounds or trials. In each round the wild blueberry simulation dataset was partitioned into a complementary subset. The model estimation was conducted on one subset (the training set), while the validation was conducted on the other subset (the validation or test set). The wild blueberry simulation dataset was partitioned differently in different rounds, and the validation results were averaged over the n rounds as an estimation of the model's prediction accuracy.

In this study, to examine the prediction accuracy of the nine metamodels for wild blueberry fruit-set predictions, four different statistical evaluation metrics were employed. First, the coefficient of determination (R^2) was used, which is defined as the proportion of the variance in the response variable that is explained by the independent variables. Second, we used the mean absolute error (MAE), which is defined as the absolute mean difference between the i^{th} actual fruit-set y_i and the i^{th} predicted fruit-set $\hat{y}_i$. Third, we used the mean average percentage error (MAPE), defined as the prediction accuracy as an average of ratios, each of which is the i^{th} prediction error $y_i - \hat{y}_i$ divided by the i^{th} actual value y_i. Fourth, the root mean squared error (RMSE) was used and is defined as a measure of the quadratic mean difference between predicted and actual fruit-set. The mean absolute error (MAE), mean average percentage error (MAPE) and root mean squared error (RMSE) were computed using the Equations (6)–(8); respectively:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{6}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{7}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{8}$$

where n is the number of samples, y_i is the observed fruit-set and $\hat{y}_i$ is the model-predicted fruit-set. The coefficient of determination (R^2) was computed according to [53] and is given by Equation (9):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y}_i)^2} \tag{9}$$

where y_i , $\hat{y}_i$ and n are as defined above, and $\overline{y}_i$ is the mean observed wild blueberry fruit-set. The metamodel with the lowest prediction error and the highest R^2 was selected as the best metamodel.

2.4. Model Calibration and Validation

When the best metamodel was selected as the candidate for wild blueberry fruit-set prediction, it needed to be further calibrated to fit the reproductive dynamics of bog blueberry before specific predictions could be made for bog blueberry. The best metamodel trained on the wild blueberry simulation dataset was regarded as the best representation of the relationships between fruit-set and the eight variables by learning the basic pattern of pollination dynamics produced by the Wild Blueberry Pollination Simulation Model. We hypothesized that the selected metamodel has the potential to describe a similar pattern of pollination dynamics in the bog blueberry agroecosystem [30]. This hypothesis was tested in two steps. Step one was based upon exploring and estimating the best parameter set of the best metamodel for bog blueberry. Step two was based upon comparing the predictions with bog blueberry fruit-set not used previously in order to examine the goodness of fit.

The best metamodel of wild blueberry prediction had eight parameters (or features, including clone size, bee taxon-specific densities and weather) that determine fruit-set. However, the prediction only needs weather data from bog blueberry fields as input. Other parameters such as clone size, densities of different bee taxa are unknown and need to be predetermined before fruit-set prediction can be made for the bog blueberry agroecosystems. Therefore, to calibrate the best metamodel to bog blueberry, it was necessary to find optimal values for the subset of parameters that (1) minimized the difference between the predicted bog blueberry fruit-set and the observed of bog blueberry fruit-set on six bog blueberry fields in years 2015 and 2016; and simultaneously (2) maximized the total coefficient of determination (R^2, see Equation (9)) of these parameters that explain the variance of the model. The first objective can be mathematically formulated as:

$$J_1(\theta) = \sum_{i=1}^{M} \sum_{j=1}^{N} e_{i,j}(\theta) \tag{10}$$

where $e_{i,j}(\theta)$ is a metric function that evaluates the prediction error, which was specified as the RMSE in equation 8; M and N are the number of fields (i.e., (6)) and the number of years (i.e., (2)) being predicted, respectively; θ denotes the following subset of parameters selected for optimization from the metamodel: Clone size (CS), density of Honeybees (HB), Bumble bees (BB), Andrena bees (AD) and Osmia bees (OS).

The second objective can be formulated as:

$$J_2(\theta) = \sum_{i=1}^{M} \sum_{j=1}^{N} \frac{1}{R^2_{i,j}(\theta)} \tag{11}$$

where $R^2_{i,j}(\theta)$ is the coefficient of determination (R^2) of θ that explains the variance from the metamodel, which is defined in Equation (9). Finally, we looked for the set of values for the 5 parameters θ that optimize both objectives, which can be formulated as a multi-objective problem:

$$\begin{matrix} \min J(\theta) \\ \theta \in \Re^5 \end{matrix} = [J_1(\theta), J_2(\theta)] \in \Re^2 \tag{12}$$

A multi-objective optimization method [54] was employed to search the best parameter set for the metamodel that is expected to achieve the highest prediction accuracy and interpretability for bog blueberry fruit-set. To minimize the random effect, we replicated the optimization algorithm 100 times, each of which was considered to have reached convergence when both of the objectives did not make further improvement in values smaller than 10^{-6} upon successive iterations.

Once the best values of the five parameters (except for the three weather relevant parameters) were found for the metamodel, its predictions were validated on the mean fruit-set observed in the same six bog blueberry fields in the years 2017–2019. Our conceptual approach was that if the validation results were acceptable, it indicated that the metamodel had successfully learned the difference in pollination patterns between wild blueberry and bog blueberry. The metamodel could then be regarded as successfully calibrated to bog blueberries and would be ready for making realistic field predictions.

2.5. Prediction Applications

We conducted three prediction experiments to evaluate how different weather conditions expected in the near future (5~10 years) might affect bog blueberry fruit-set in northeastern China. Specifically, we used the calibrated metamodel, i.e., XGBoost, to make predictions for bog blueberry fruit-set in the 151 growth regions with several combinations of temperature and precipitation variation while keeping other parameters unchanged. Then the variation in predicted fruit-set of these regions were analyzed to determine if a geographical pattern existed. Finally, if specific weather conditions caused a decline in fruit-set, several management efforts which were available to be adjusted in the metamodel,

such as increasing commercial honeybee or bumble bee densities, were tested for potential mitigation of the fruit-set decline under the stress of weather factors.

2.5.1. Bog Blueberry Fruit-Set in Various Weather Conditions during Bloom

According to a recent climate research report, the most likely climate change direction in Northeast China in the near future (5~10 years) would be higher air temperatures with more rainy days during the bog blueberry production season. These conditions may pose a threat to bog blueberry's fruit-set. The primary goal of this experiment was to estimate bog blueberry's fruit-set under higher temperature and precipitation conditions. However, the opposite weather conditions were also considered in this experiment to elucidate the full effects of a highly variable fluctuation in weather on bog blueberry crop production. Specifically, we conducted a full factorial prediction experiment, in which the two factors (i.e., MaxUTR and RD) were systematically varied on seven levels (decreased or increased by 5%, 15% and 25% with respect to the baseline scenario indicating the recent weather conditions between 2017 and 2019). Therefore, there were $7 \times 7 = 49$ combinations of predictions made. The predicted fruit-set of the 151 bog blueberry growth regions in Northeast China were plotted and projected onto a GIS map with ArcGIS [55], from which the variation of predicted bog blueberry fruit-set can be observed along a gradient of weather change.

2.5.2. Geographical Pattern of Bog Blueberry Fruit-Set Change

We were particularly interested in whether the change in weather during bloom evenly or irregularly affects bog blueberry fruit-set across the 151 growth regions in Northeast China. If the effects were irregular, we expected to observe geographical patterns in response. For example, are specific parts of the growing regions more likely affected by weather change than others? It is important to evaluate the potential for bog blueberry production along a geographical gradient under a changing climate and provide insight into management and conservation. Specifically, we used the predictions generated in the previous experiment to analyze the changes in correlation between fruit-set and the longitudes and latitudes of the predicted regions when the air temperature positively or negatively 5%, 15% and 25% deviated from the current climate baseline (2017–2019). We also conducted a similar analysis for scenarios in which the number of rainy days during bloom were changed in the same proportional pattern.

2.5.3. Bee Density Enhancement for Compensating Bog Blueberry Fruit-Set Decline

Some weather conditions may negatively affect pollinator foraging activities, which could cause a decline in bog blueberry fruit-set. We were especially interested in whether some factors that are relevant to bog blueberry production can be manipulated to increase fruit-set. Hopefully these efforts can compensate or mitigate a decline in fruit-set caused by inappropriate weather conditions. Of the three groups of eight features (parameters of the XGBoost model), weather (RD, MaxUTR and MinUTR), plant spatial traits (CS) and bee density (HB, BB, OS and AD); only bee density is subject to human intervention. Commercial honeybee and bumble bees are available in many regions in Northeast China. Therefore, it is feasible to increase the density of honeybees or bumble bees without reducing the density of wild bees for the purpose of enhancing the pollination service in a field. By exploring the parameter space of the XGBoost model, we aimed at finding the optimal density for honeybees or bumble bees that can independently compensate for fruit-set decline. Specifically, we made two sets of predictions with the metamodel to fit the relationships between bee density and bog blueberry fruit-set while keeping other factors unchanged in accordance with the weather conditions where fruit-set was negatively affected. Then we quantitatively estimated the percentage of extra honeybee or bumble bee investment necessary to mitigate the negative effect of climate change.

3. Results

3.1. Metamodel Formulation and Selection

Nine metamodels were constructed and their performances in terms of prediction ability were evaluated on the simulated wild blueberry pollination dataset with a five-fold cross-validation method. The three best metamodels that had the lowest MAE, MAPE, and RMSE and the highest R^2 were: RF (Random Forest), GBDT (Gradient Boosting Decision Tree), and XGBoost (eXtreme Gradient Boosting) (Table 2). The remaining six metamodels were not considered further in future metamodeling steps. Even though RF and GBDT had almost the same level of prediction error as XGBoost, we selected XGBoost as the best metamodel and used it for bog blueberry fruit-set prediction because it had a higher R^2 (0.932) than RF (0.922) and GBDT (0.927). The XGBoost metamodel had the lowest unexplained variance in wild blueberry fruit-set among the three-best performed metamodels. The highest R^2 achieved by XGBoost has given the best model generalization capability in dealing with unexpected patterns in validation data not used to construct the metamodels. The scatter plots (Figure 5) illustrate how accurate the fruit-set in the validation data can be predicted by different metamodels trained in the training set of the wild blueberry pollination dataset.

Table 2. The model's performance on the simulated wild blueberry pollination dataset.

Model	Evaluation Metrics			
	MAE	**MAPE**	**RMSE**	**R^2**
SVM	0.044	9.025%	0.053	0.553
MLR	0.034	7.282%	0.042	0.721
AdaBoost	0.030	6.398%	0.038	0.773
DT	0.013	3.093%	0.03	0.845
ANN	0.022	4.720%	0.029	0.861
KNN	0.014	3.152%	0.024	0.912
RF	0.011 [1]	2.520%	0.022	0.922
GBDT	0.011 [1]	2.483% [1]	0.021 [1]	0.927
XGBoost	0.011 [1]	2.489%	0.021 [1]	0.932 [2]

[1] bold denotes the best overall lowest value for MAE, MAPE and RMSE; [2] bold denotes the best overall highest value for the R^2.

In addition to model prediction performance, feature importance in the three best metamodels were also evaluated to understand the influence of potential predictive factors on blueberry fruit-set. The importance of features was estimated by the coefficient statistics between each feature and fruit-set and then standardized to 100% as a relative importance measure. Results showed that CS (clone size), RD (Raining days), and MaxUTR (the highest point of the upper range of the daily air temperature along the blueberry product season) were the three most important factors influencing blueberry fruit-set across the three best metamodels (Figure 6). It was not surprising that CS was the most important factor affecting blueberry fruit-set due to the outcrossing nature of blueberry (i.e., differential outcrossing success depending upon clone genetics of sire and recipient [28]) and bee foraging behavior, which is confirmed by previous empirical [6] and simulation research results [15]. However, both wild blueberry and bog blueberry are wild species that are not planted, so farmers have no control over this feature. The two weather parameters, precipitation and air temperature showed the next highest influence on blueberry fruit-set.

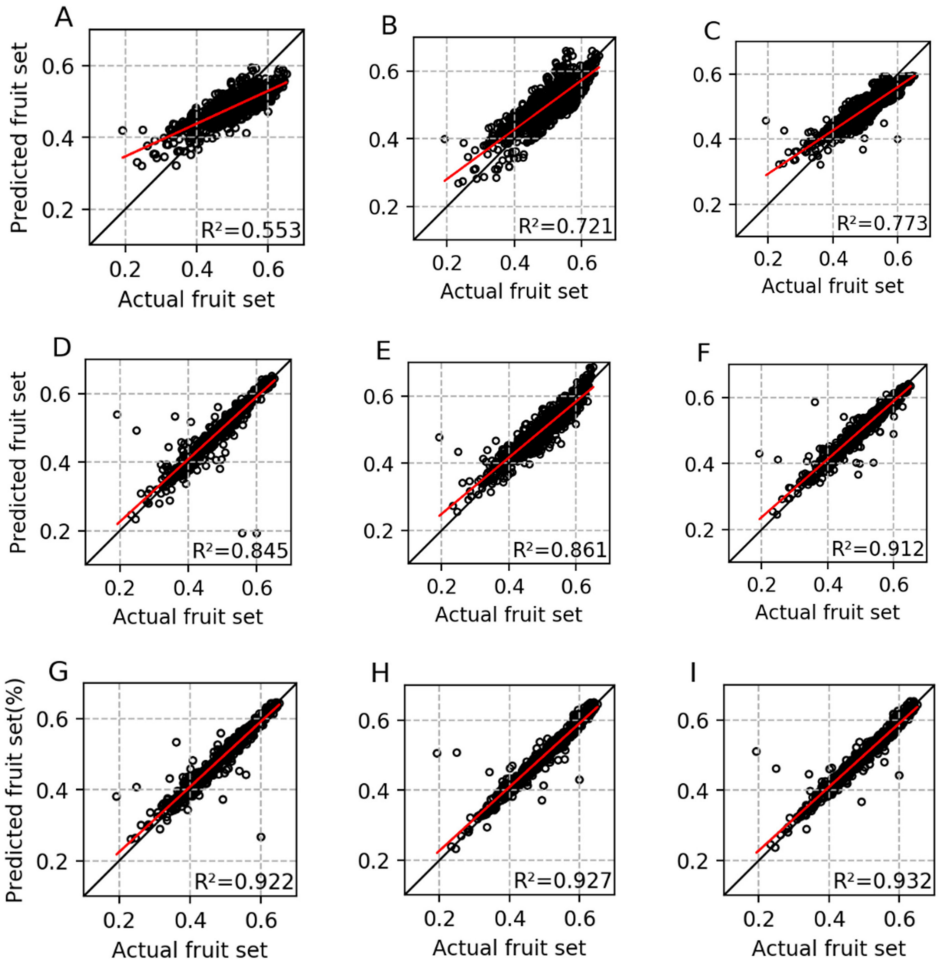


Figure 5. Scatter plots of actual fruit-set versus predicted fruit-set based on the wild blueberry pollination simulation validation dataset for MLR (**A**), SVR (**B**), AdaBoost (**C**), DT (**D**), ANN (**E**), KNN (**F**), RF (**G**), GBDT (**H**) and XGBoost (**I**). Black lines are 1:1 slopes and the red lines are least squares regression lines.

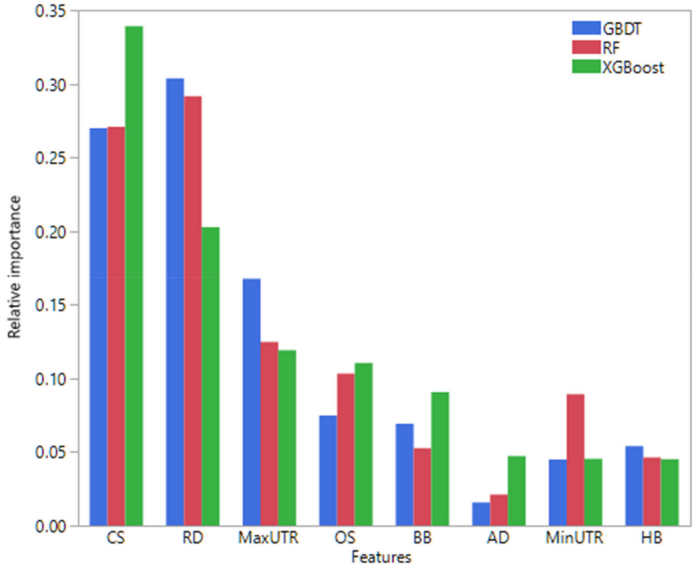


Figure 6. Relative importance of features affecting results of three different metamodels (GBDT, RF, and XGBoost). Clone size, the number of rainy days, and maximum upper temperature ranges were the three most important features across metamodels.

3.2. Metamodel Calibration and Validation

The XGBoost metamodel outperformed the remaining eight metamodels in previous training and evaluation steps on the wild blueberry pollination dataset. It was therefore selected as the candidate model to predict bog blueberry fruit-set since it had learned the basic dynamics of pollination and fruit-set for blueberry species. The XGBoost model was then calibrated to the empirical bog blueberry fruit-set dataset collected at the six field sites (CBS, GHS, JMS, MH, TH and YC) in years 2015 and 2016. The calibration process is a multi-objective optimization [55] that simultaneously minimizes prediction error and maximizes interpretability by finely tuning the parameters. Figure 7A,C,E,G,I shows the predicted versus the observed bog blueberry fruit-set for the six fields in years 2015 and 2016 without calibration, i.e., the XGBoost metamodel trained on wild blueberry data were directly used in bog blueberry fruit-set prediction. After calibration, the average accuracy of bog blueberry fruit-set prediction increased from 0.658 to 0.975, as shown in Figure 7B,D,F,H,J, in which the prediction accuracy is defined as the coefficient of determination (R^2) of the linear regression between predicted and observed fruit-set. The visualizations in Figure 7 also shows that specific ranges of parameters, such as clone size and bee density, are also the optimal settings for bog blueberry. Table 3 presents the means and 95% confidence intervals of the optimal parameter set for the XGBoost model that had been calibrated to bog blueberry observations. The optimal parameter set obtained on the bog blueberry dataset showed a slight difference in contrast with the original version trained and tested on the wild blueberry dataset. Clone size shrunk from 18.78 to 15.32; honeybee density increased from 0.417 to 1.026; bumble bee density decreased from 0.282 to 0.232, etc. These variations of model parameters likely reflect the difference of plant physiology, phenology, and genetics in vegetative and reproductive growth between wild blueberry and bog blueberry.

Table 3. The optimal parameter set of the best metamodel found by the multi-objective optimization algorithm [54] when XGBoost was calibrated to the observed fruit-set of the six bog blueberry field sites (CBS, GHS, JMS, MH, TH, and YC) in the years 2015 and 2016.

Parameter	Unit	Mean	95% CI (Upper)	95% CI (Lower)
Clone size (CS)	m^2	15.32	15.945	14.694
Honeybee (HB)	bees/m^2/min	1.026	1.068	0.984
Bumble bee (BB)	bees/m^2/min	0.232	0.241	0.223
Andrena bee (AD)	bees/m^2/min	0.435	0.453	0.417
Osmia bee (OS)	bees/m^2/min	0.637	0.663	0.611

The calibrated XGBoost model with the optimal parameter set was validated on bog blueberry fruit-set data collected at the six field sites (CBS, GHS, JMS, MH, TH and YC), but in different years between 2017 and 2019. Table 4 shows that 14 out of 18 bog blueberry fruit-set predictions made by XGBoost fell into the 95% confidence interval of the field observations, which achieved an overall 78% prediction accuracy. If we consider the 4 missed predictions (Wilcoxon Signed Rank test, $p < 0.05$) further, the worst case (CBS in year 2018) had only a 4% prediction error between the lower bound of the 95% confidence interval (0.504) and the XGBoost prediction (0.482); while the best case (MH in year 2018) had less than 1% prediction error between the lower bound of the 95% confidence interval (0.509) and the XGBoost prediction (0.504). This evidence suggests a successful model validation if one takes into account the general acceptance rate of model prediction [56]. Therefore, the XGBoost metamodel can be adapted and trusted in bog blueberry fruit-set predictions.

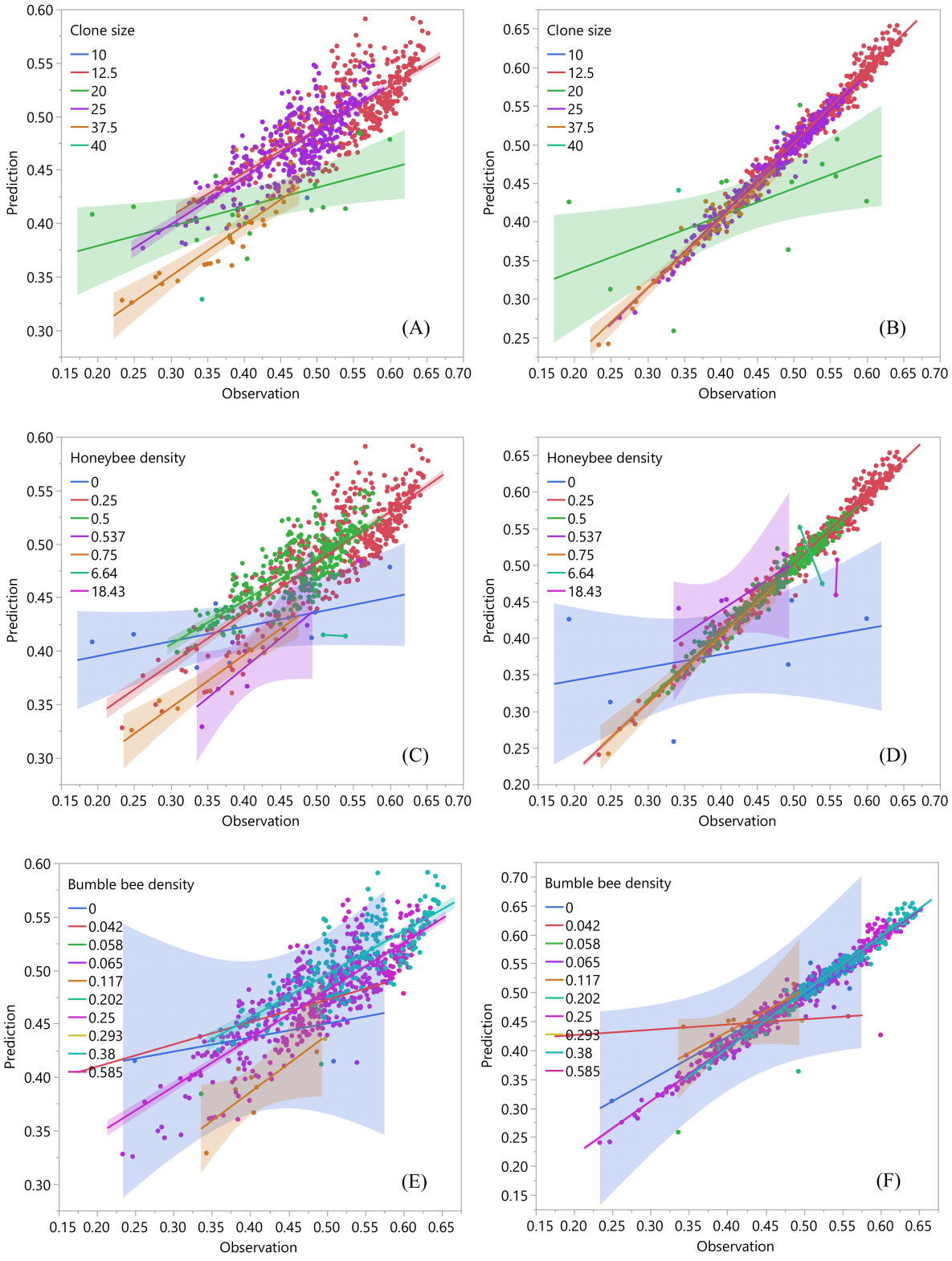


Figure 7. *Cont.*

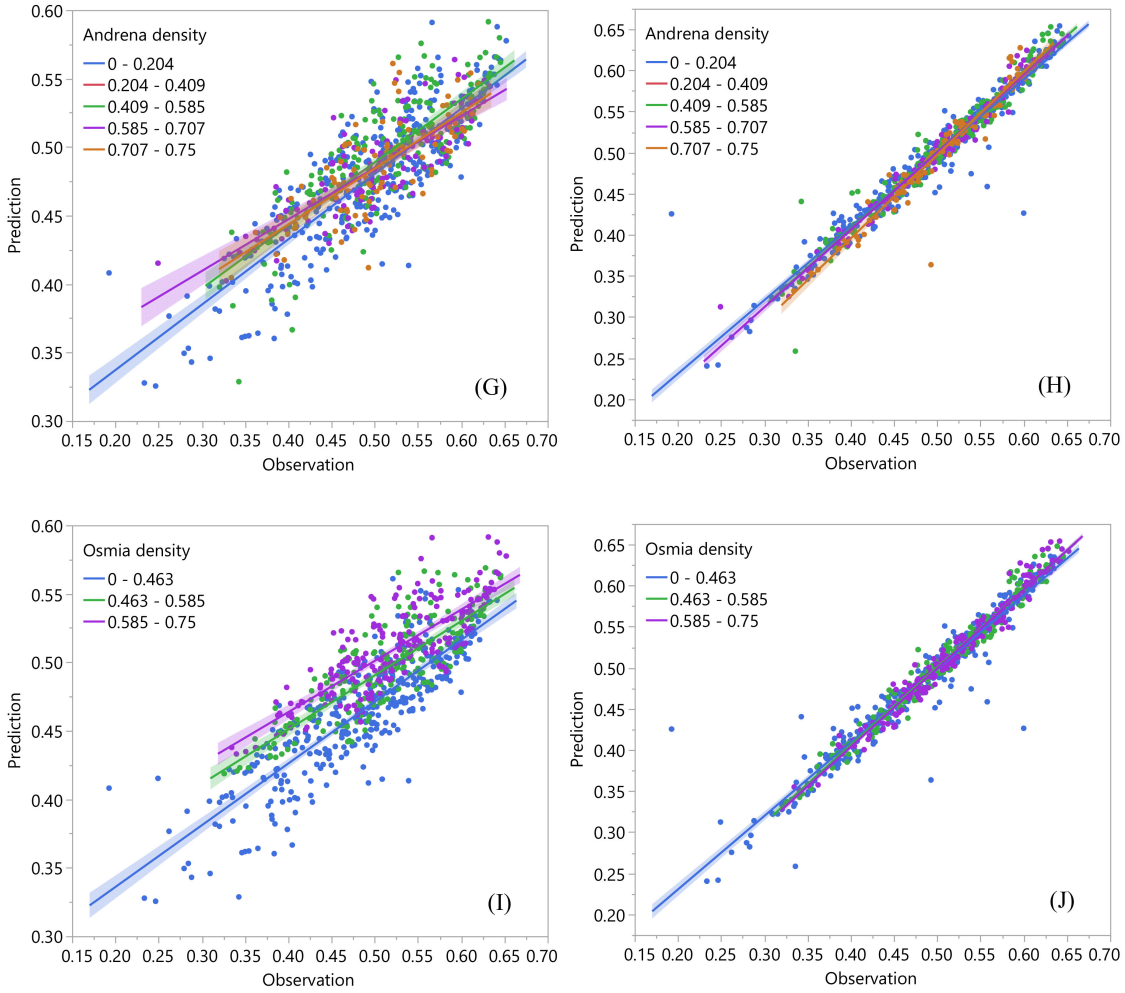


Figure 7. Accuracy of bog blueberry fruit-set prediction before (**A**,**C**,**E**,**G**,**I**) and after (**B**,**D**,**F**,**H**,**J**) parameter optimizations were conducted. The prediction accuracy is defined as the coefficient of determination (R^2) of the linear regression between predictions and observations in fruit-set.

Table 4. Validation results of the calibrated metamodel (XGBoost). The model predictions were compared with field observations in six sites (CBS, GHS, JMS, MH, TH, and YC) in the years between 2017–2019, using the Wilcoxon Signed-Rank test. Bold and italic numbers indicate statistically significant differences ($\alpha = 0.05$), which means that the metamodel failed to predict bog blueberry fruit-set in a specific field site and year.

Field Site	Year of Observation	Field Fruit-Set Mean	Field Fruit-Set 95% CI (Upper)	Field Fruit-Set 95% CI (Lower)	Predicted Fruit-Set	df	*p*-Value
CBS	2017	0.494	0.519	0.468	0.498	163	0.724
CBS	2018	0.520	0.537	0.504	0.482	287	***<0.001***
CBS	2019	0.560	0.576	0.543	0.570	272	0.220
GHS	2017	0.452	0.491	0.413	0.466	67	0.475
GHS	2018	0.554	0.592	0.516	0.576	58	0.257
GHS	2019	0.520	0.549	0.491	0.504	145	0.283
JMS	2017	0.456	0.487	0.429	0.458	111	0.872
JMS	2018	0.567	0.587	0.548	0.576	205	0.382
JMS	2019	0.495	0.533	0.458	0.481	75	0.446
MH	2017	0.462	0.504	0.420	0.466	71	0.829
MH	2018	0.534	0.559	0.509	0.504	128	***0.020***
MH	2019	0.545	0.570	0.521	0.524	102	0.089
TH	2017	0.471	0.502	0.440	0.461	125	0.525
TH	2018	0.541	0.567	0.514	0.546	102	0.697
TH	2019	0.509	0.537	0.481	0.564	122	***<0.001***
YC	2017	0.582	0.603	0.560	0.544	113	***<0.001***
YC	2018	0.564	0.592	0.536	0.556	99	0.564
YC	2019	0.487	0.519	0.455	0.510	95	0.154

3.3. Bog Blueberry Fruit-Set Predictions

The first experiment revealed that under the current weather conditions (baseline scenario (T + R)) in which the variation of air temperature and precipitation was zero, the 151 bog blueberry growth regions had a mean fruit-set of 0.403 and a median of 0.374. The highest fruit-set prediction recorded was 0.521 in Oroqin county (50.6° N, 123.73° E). The lowest fruit-set prediction was 0.373 and recorded in multiple regions including Horqin county (46.36° N, 121.13° E). As expected, higher temperature significantly drove the fruit-set decline in most regions (ANOVA, F = 6.360, df = 603, $p < 0.001$). The rise of air temperature alone by 5, 15 and 25% caused 3.5% (mean = 0.389), 4.2% (0.386) and 4.2% (0.386) decline in bog blueberry fruit-set in the 151 regions (Figure 8A). The lowest fruit-set recorded was 0.317 in Sunwu county (49.25° N, 127.2° E) in both scenarios where air temperature was raised by 15 and 25%. Surprisingly, fruit-set in currently warmer weather did not drop by an equivalent percentage as the increase in air temperature. We also did not find further fruit-set decline when the air temperature in these regions was already 15% higher than current weather conditions (Tukey-Kramer HSD test, $p = 0.866$). This indicates a nonlinear relationship between bog blueberry fruit-set and possible warmer weather conditions. The opposite direction in weather change, i.e., when the air temperature in these regions declined by 5, 15 and 25%, promoted bog blueberry fruit-set by 7.7% (0.434), 21.8% (0.491) and 18.9% (0.479) in the 151 regions (ANOVA, F = 200.723, df = 603, $p < 0.001$, Figure 8A).

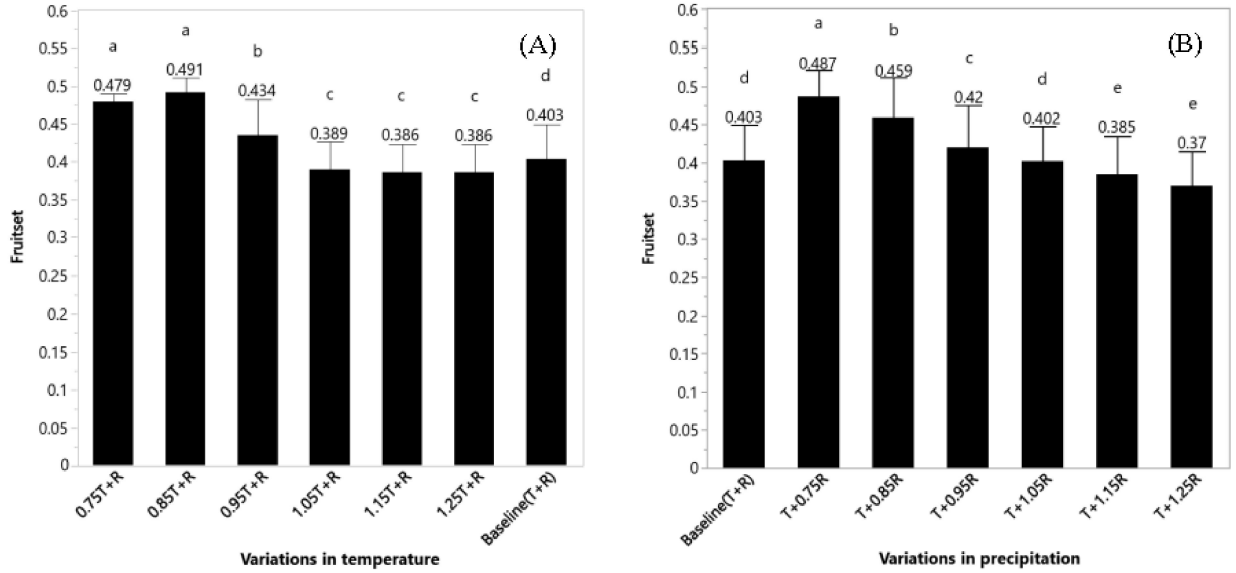


Figure 8. Effects of change in air temperature (**A**) or precipitation (**B**) on bog blueberry's fruit-set in the 151 growing regions in Northeast China. Error bars are one standard deviation of the mean. Different letters on bars indicate significant differences ($\alpha = 0.05$) between climate change scenarios.

If the weather became dryer and the air temperature remained unchanged, the average fruit-sets of the 151 bog blueberry growth regions were increased by 4.2, 13.9, and 20.8% when the number of rainy days decreased by 5, 15, and 25% (Figure 8B), which were significantly higher than those fruit-sets associated with the baseline weather conditions (Turkey-Kramer HSD, $p < 0.001$). The highest fruit-set recorded was 0.528 in Manchuria county (49.35° N, 117.19° E). On the contrary, if the weather was more wet, the average fruit-sets in the 151 bog blueberry growing regions showed 0.2, 4.4, and 8.2% decline in contrast with the baseline scenario when the number of rainy days increased by 5, 15, and 25% (Figure 8B). It appears that only a 5% precipitation increase in the future will not (Turkey-Kramer HSD, $p = 0.999$) cause a decline in bog blueberry fruit-set. The lowest fruit-set in wetter weather conditions was 0.317 in 28 counties.

Overall, increased air temperature and/or precipitation posed a larger magnitude of threat to bog blueberry fruit-set than that the benefit of increased fruit-set brought

about by an opposite weather direction of decreased air temperature or precipitation (see Appendix A, Figure A1). The nonsymmetrical effects of temperature or precipitation variation indicated that the most likely trend of future weather may cause severe damage to bog blueberries as we initially expected. Most bog blueberry production regions in Northeast China will suffer fruit-set declines if the weather continues to warm or become more wet, a model prediction of future climate change.

Predictions also provided evidence of interactions between air temperature and precipitation (ANOVA, df = 36, F = 15.821, $p < 0.001$). The highest fruit-set was 0.562 in Manchuria county (49.35° N, 117.19° E) under the cooler and dryer weather conditions (air temperature was 15% lower, and precipitation was 25% lower than the baseline situation). This record was also the highest fruit-set value in our predictions for all possible combinations of changes in temperature and precipitation. It suggests that cooler and dryer weather might be ideal for bog blueberry reproduction and harvestable yields. The lowest fruit-set was recorded as 0.317 in more than 60 counties under the warmer and wetter weather scenario (air temperature and precipitation were both increased 25% over baseline), which is much worse than the negative effect caused by increased air temperature or precipitation alone. This is a warning signal to future bog blueberry production in Northeast China, again as warmer more rainy springs are to be expected. Bog blueberry fruit-set predictions in the 151 regions under different weather variation scenarios are geographically shown in Figures 9 and 10, in which fruit-set of counties were categorized into three levels (less than 0.4, between 0.4 and 0.5, higher than 0.5).

The second experiment found that overall, bog blueberry fruit-set is positively correlated with latitude of the production regions. This is true for both field observations and metamodel predictions. Northern regions had higher fruit-set than southern ones. But no apparent trend was detected in this relationship when air temperatures varied while precipitation was static (see Appendix A, Table A1, $p = 0.595$). This finding suggests that warmer or cooler weather in the near future is not likely to change the geographical pattern of the locations of the highest bog blueberry productivity in Northeast China. Surprisingly, a strong trend in the correlations between latitude and the change in fruit-set was detected when precipitation was varied (Table A2, R2 = 0.833, $p < 0.001$). The higher the precipitation level, the stronger the correlation between latitude and fruit-set. The trend indicated that if the level of precipitation goes up in the future, the northern bog blueberry production regions (such as Mohe, Tahe, and Oroqin counties) will have an advantage in achieving high bog blueberry fruit-set compared to the southern regions (such as Changbaishan, Shulan and Tonghua counties). However, if weather becomes dryer, the southern regions will benefit more from optimal weather growing conditions and the difference in fruit-set levels between northern and southern regions will shrink.

The decline in bog blueberry fruit-set due to climate induced increases in air temperature and rainy days can be estimated at 8%, from 0.403 down to 0.37 (Figure 8B). According to the third prediction experiment, the 8% decline in fruit-set can be compensated by intentionally increasing honeybee density from the current 1.0 bees/m^2/min to 18 bees/m^2/min (Figure 11A). If all other factors are fixed and only honeybee density is increased, the 8% of bog blueberry fruit-set that has to be compensated for requires much higher honeybee density than a scenario where all factors are fully interactive and able to be manipulated. However, due to the recognized high efficiency of bumble bees as pollinators for blueberry, increasing commercial bumble bee density alone from the current 0.232 bees/m^2/minute to 0.568 bees/m^2/minute (Figure 11B) can enhance bog blueberry fruit-set by 8%. One should note that even though the more efficient bumble bee can offset climate induced fruit-set decline, a bottleneck exists since our predictions show that it is almost impossible to have higher bog blueberry fruit-set than 0.60 by only introducing bumble bees (bumble bee density higher than 0.65 bees/m^2/minute) because predicted proportion fruit set increases at a decreasing rate with increasing bumble bee forager density.

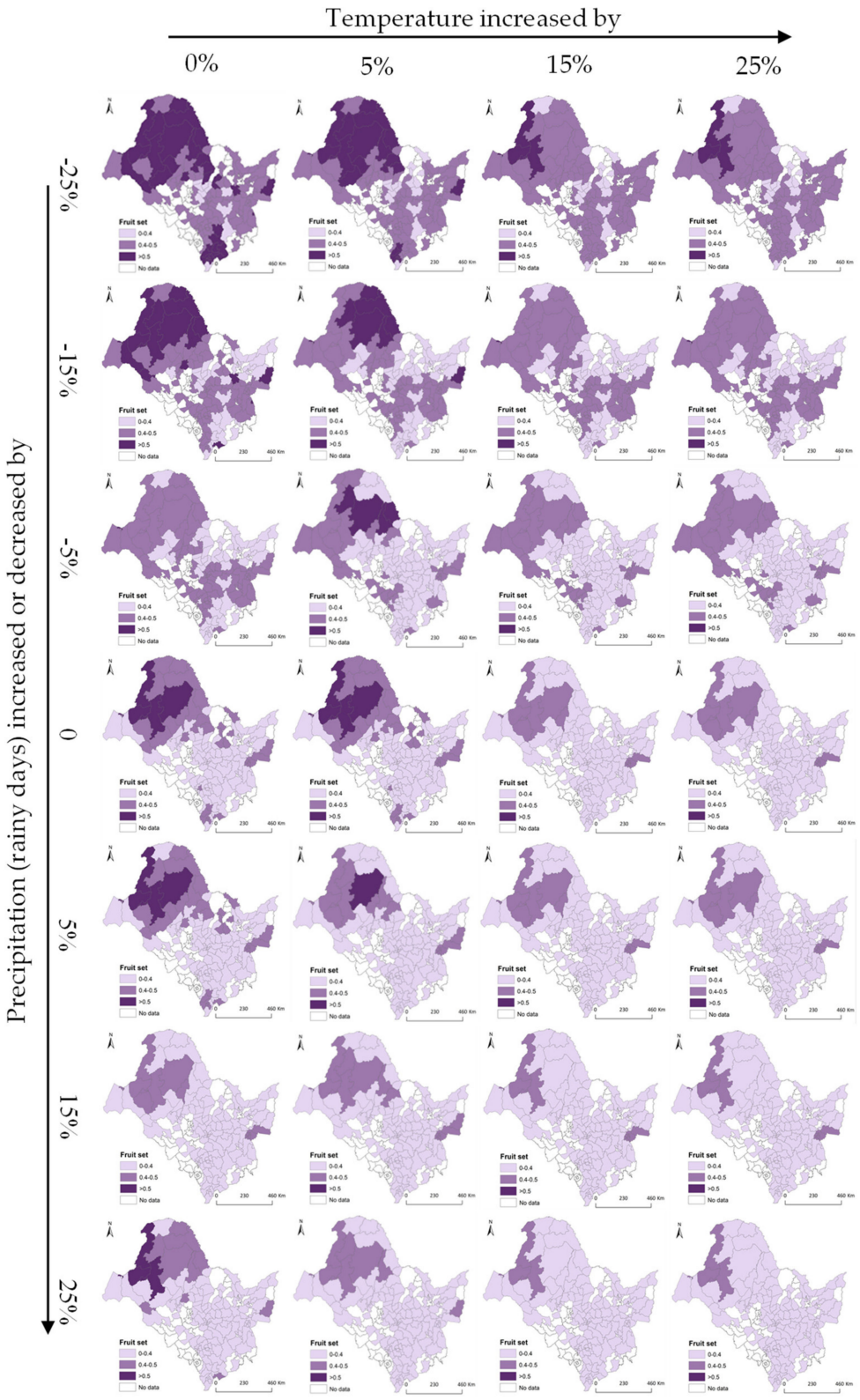


Figure 9. Fruit-set predictions in 151 bog blueberry production regions in Northeast China with increased temperatures and precipitation.

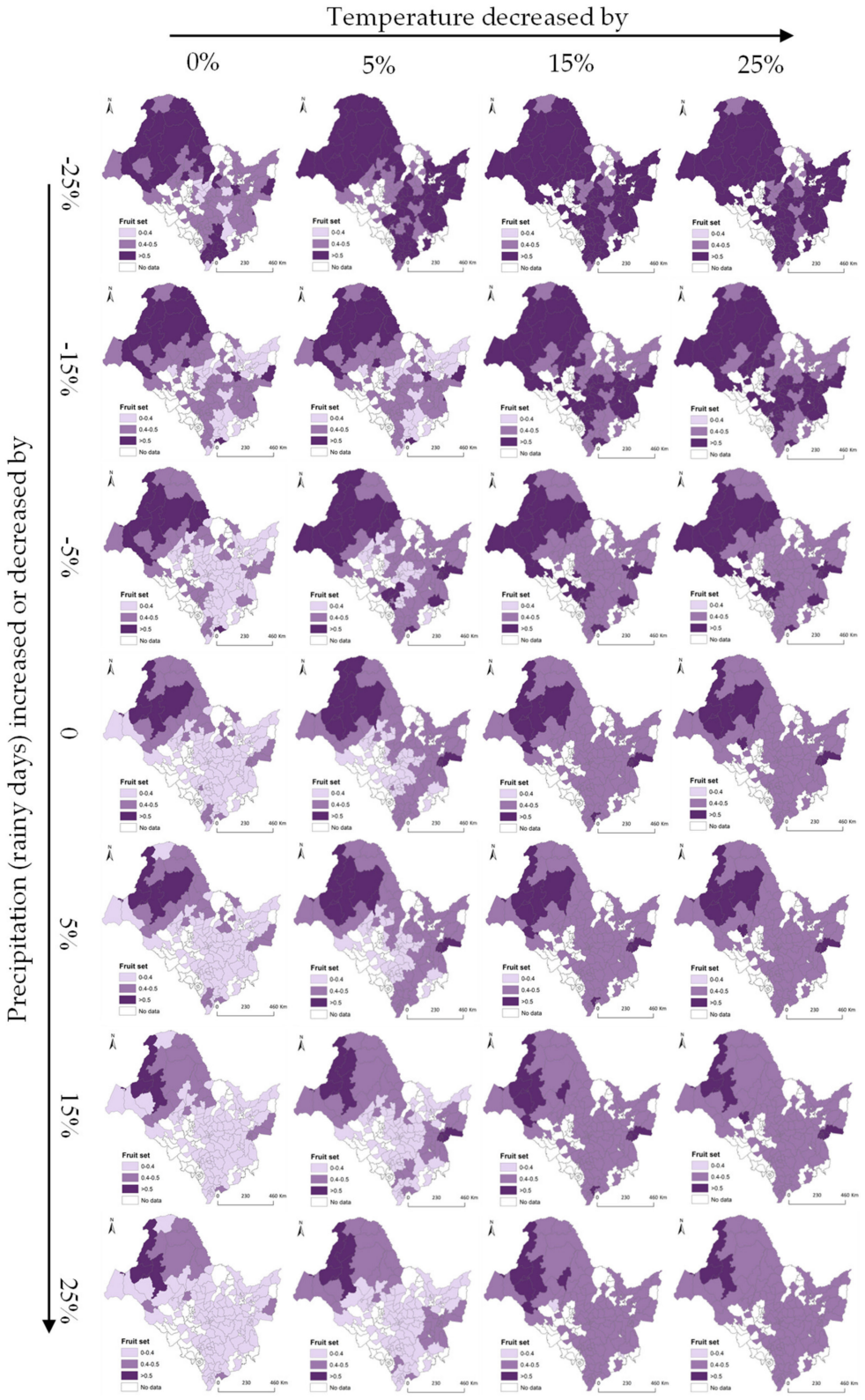


Figure 10. Fruit-set predictions in 151 bog blueberry growing regions in Northeast China with decreased temperatures and precipitation.

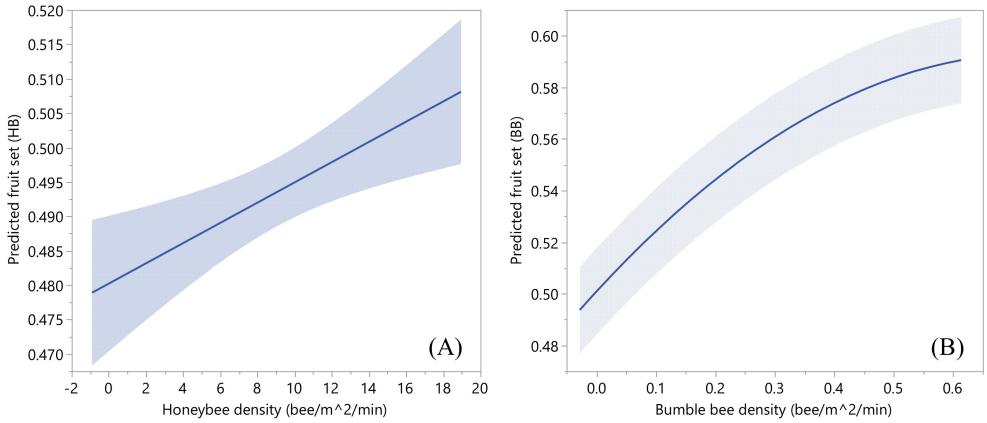


Figure 11. Predicted bog blueberry proportion fruit-set based on the density of commercial honeybee density (**A**) and bumble bee density (**B**) using the XGBoost model parameterized for recent weather conditions while keeping other parameters at default values. Blue-grey areas around the trend lines represent a 95% confidence interval envelope.

4. Discussion

4.1. Modelling Approach and Interpretation

This research contributed to our scientific knowledge with a novel modeling approach to predicting the fruit-set of Northeast China's bog blueberry based upon a very limited amount of data for model development. As a major agroecosystem in Northeast China; bog blueberry's yield, fruit quality and economic stability are important to the local rural economy. Machine learning models that capture the complex nonlinear relationships between fruit-set and various interacting biotic and abiotic factors are essential for developing adaptive bog blueberry management strategies. However, due to the lack of systematic field experiments and observations on bog blueberries in Northeast China, it was unlikely that in the next few years agricultural scientists could directly develop and train reliable machine learning models for prediction purposes. We therefore utilized previously developed and validated computer simulation models based on US wild blueberry production as a reliable data generator to feed nine machine learning models. The best machine model with the highest prediction performance from the previous step was parameterized and calibrated on a small bog blueberry data set (less than 1500 samples). After calibration, the metamodel was used for bog blueberry fruit-set prediction. This approach allows machine learning models to be adapted to new prediction tasks even though the available data for retraining is limited. The theoretical foundation for this approach is well-known and referred to as *transfer learning* [57]. The approach provides a general machine learning framework where a model trained (aka knowledge gained) in one domain of interest can be applied (aka transferred) to a new but similar domain where data is usually limited. In recent years, transfer learning has been widely used in deep learning models for crop disease detection [58–60], but it is not common for non-deep learning models [61], which may need more fine-tuning with multi-objective optimization techniques [54]. However, transfer learning would work with any type of machine learning algorithms where the base features are the same [62] with different magnitudes of interpretability [63]. In our research, the best machine learning model trained on the wild blueberry simulation dataset was regarded as the best representation of the relationships between fruit-set and the eight predictor variables by learning the basic pattern of pollination dynamics of multiple blueberry species. Although bog blueberry may have diverse genotypes (e.g., the proportion of self-compatible genotypes, [64]) that could cause different pollination dynamics, these differences can still be explained by a metamodel with re-parameterization on a small amount of bog blueberry data. The difference of decomposed feature-level variance before and after metamodel calibration (the visualizations in Figure 7) might shed light on this point.

Unlike the deep neural network-based transfer learning approach [65], in which only the very few last layers are replaced and retrained while front layers are kept as storage of common knowledge between the source and the target domain [57], our approach uses a similar concept of knowledge storage, but with a different way of retraining [62]. In our approach, the basic patterns of pollination dynamics are captured and stored in the machine learning models (metamodels) trained on vast amounts of simulation data. However, we did not split these metamodels, instead, the difference in pollination dynamics between different blueberry species was offset by adjusting a partial parameter set. Specifically, clone size (CS) and the density of different bee species (OS, BB, AD, and HB) were chosen for adjusting with the help of a multi-objective optimization method. In doing so, the best value of these features in the metamodels maximize the likelihood of predicting bog blueberry fruit-set. The merits of this technique are threefold. First, directly adjusting parameters (not for model manipulation) instead of the structure of a metamodel for retraining is helpful in some scenarios where the features learned are not multiscale. Splitting the model for retraining (e.g., like deep neural networks do) may not work. Second, directly adjusting parameters is straightforward and can be done very quickly, because searching an optimal set by means of a strong multi-objective optimization procedure has much lower computational complexity than training a deep neural network until it converges [54]. Third, since the number of parameters in our approach (four) is much less than the number of hyperparameters in a deep neural network (usually thousands), the retraining process requires much less data [63], which implies that our approach is more robust and universal than many transfer learning situations where data insufficiency is common. However, we acknowledge that our method, like all metamodeling approaches based on simulation data, is restricted by how effectively the simulation model can be exploited. If the simulation experimental design is not representative of the system or the number of experiments conducted is not adequate, information loss is expected and missed patterns might bias prediction results.

4.2. Practical Implications of Model Prediction

Our metamodel predictions showed that warmer, wetter weather is detrimental to optimal fruit set. As a measure of potential yield, a decline in fruit set has direct implications on profitability and sustainability of bog blueberry production in Northeast China. The current geographic distribution of bog blueberry is the cool, temperate (circumpolar) regions of the northern hemisphere [66]. This current climate envelope of cool, dry weather suggests that bog blueberry will not be tolerant to higher temperature and increased precipitation [67]. Our metamodel prediction experiments focused on reproduction and fruit set. Based upon wild blueberry, some of the mechanisms of decreased fruit set due to climate warming and increased precipitation are: reduced activity of bees during rainy days, reduction of the longevity of stigma viability during warm springs, faster rate of progression through the bloom period by the flowering clones (genetically unique individual plants), asynchrony between the emergence of wild bees and the early blooming clones, and increased blossom loss due to infection by *Monilinia vaccinii-corymbosi* (Reade) (the causal fungal organism of mummy berry disease) [15,26,68]. Parkinson and Mulder [2] also support our hypothesis that bog blueberry fruit set will be affected in a similar way that wild blueberry fruit set is by warming. In Alaska, USA, they found that bog blueberry exhibited a decline in fruit set in warmer lowland sites relative to cooler upland sites.

However, the loss in bog blueberry productivity due to climate warming and increased precipitation may be even greater than the losses that we modeled. Increased precipitation can result in other plant diseases due to fungi that reduce photosynthetic area and overall vigor of the plant and may make plants susceptible to other pathogens [26]. Graae et al. [69] support our hypothesis that climate change may increase fungal disease in bog blueberry as it does in wild blueberry. They found increased seed infection by fungi under warmer temperatures. Also in wild blueberry, Tasnim et al. [13] found that already occurring climate change resulting in warming and increased potential evapotranspiration is having

negative effects on wild blueberry photosynthetic rates and overall plant growth rates. We expect that these same responses will be observed in the sibling species, bog blueberry. One of the few experiments with bog blueberry under changing weather conditions showed that nutrient uptake under warming declines in bog blueberry [70]. Therefore, the predicted decline in bog blueberry fruit set due to climate change may be only one component of a more complex suite of detrimental responses as a result of future climate change.

5. Conclusions

In our research, a new predictive modeling approach was introduced based upon the concept of transfer learning to solve the problem of data deficiency in predicting productivity of agroecosystems. Our paper delineates the process of building machine learning models (metamodels) from simulations built for one agroecosystem, where data resources are abundant; evaluating and selecting the best metamodel representing the dynamics of the system; and re-parameterizing and calibrating the metamodel to another agroecosystem where crop species-specific data is scare. To test our proposal, we used the pollination dynamics of the wild (lowbush) blueberry agroecosystem in the US as the source domain and the bog blueberry agroecosystem in Northeast China as the target domain. We have shown that knowledge gained in one crop system by modeling and simulation can be used for other crop systems sharing similar traits, which could save considerable investment in field observation and data collection [71]. Another contribution of our work is that it provides a feasible path for packaging knowledge and making it available across the scientific and grower community, which is often a neglected aspect of sustainable solution development [72]. In addition to the methodological exploration, we also applied our approach to predicting the response of productivity (i.e., fruit-set, as a proxy to yield) of Northeast China's bog blueberry agroecosystem to different weather conditions, which are regarded to be the most important abiotic factor affecting bog blueberry yield. Our application showed that in the most likely climate change scenario (i.e., higher temperatures with more precipitation), a decrease in fruit-set of Northeast China bog blueberries, the worst-case scenario being an 8% reduction was expected. The geographical pattern also revealed that the southern and eastern production regions will suffer more severe fruit-set decline than the rest of the bog blueberry growing regions. We suggest that while controlling blueberry clone size and wild bee populations is unlikely to be successful, increasing commercially available honeybee density from the current 1 bee/m^2/minute to 18 bees/m^2/minute, or commercially available bumble bee density from the current 0.2 bee/m^2/minute to 0.6 bees/m^2/minute, might be a viable way to compensate for the predicted 8% climate induced fruit-set decline in the near future.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/agronomy11091736/s1.

Author Contributions: H.Q.: Conceptualization, Methodology, Formal analysis, Writing—Original draft preparation, Project administration, Funding acquisition; R.X.: Investigation, Validation, Software, Visualization; E.Y.O.: Software, Writing—Original draft preparation; D.W.: Investigation, Data Curation; F.D.: Conceptualization, Supervision, Resources, Writing—Review & Editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China grant number [61871061] and the National Key Research and Development Program of China Grant number [2016QY01W0200].

Institutional Review Board Statement: Not applicable.

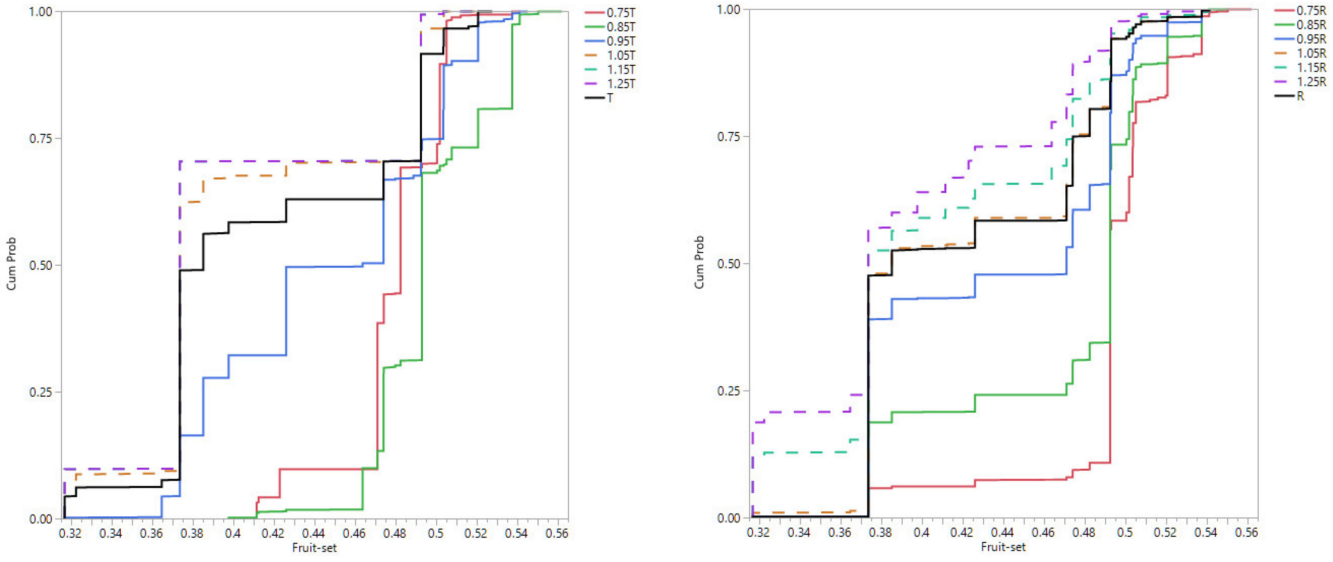
Informed Consent Statement: Not applicable.

Data Availability Statement: Data is contained within the Supplementary Material.

Acknowledgments: We thank the three anonymous reviewers for their careful reading of our manuscript and their insightful comments, which help us to improve the manuscript. We are also grateful to several field workers and graduate students for data collection and preprocessing.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A



(**A**) (**B**)

Figure A1. Cumulative probability of predicted fruit-set in the 151 bog blueberry production regions in Northeast China under potential future temperature (**A**) or precipitation (**B**) change. T and R denote the baseline scenario where no temperature (T) or precipitation (R) changes were applied to current conditions.

Table A1. Correlations between latitude/longitude and bog blueberry fruit-set in the 151 regions in Northeast China with changes of **air** temperature from −25% to 25% (Pearson's product-moment correlation test) when the number of rainy days was kept the same as the baseline scenario. Bold and italic numbers indicate statistically significant correlations ($\alpha = 0.05$).

	0.75T + R	**0.85T + R**	**0.95T + R**	**Baseline (T + R)**	**1.05T + R**	**1.15T + R**	**1.25T + R**
Latitude	0.211 (***p = 0.009***)	0.046 ($p = 0.579$)	0.233 (***p = 0.004***)	0.344 (***p <0.001***)	0.348 (***p <0.001***)	0.245 (***p = 0.003***)	0.245 (***p = 0.003***)
Longitude	−0.244 (***p = 0.003***)	0.005 ($p = 0.947$)	−0.101 ($p = 0.219$)	−0.274 (***p = 0.001***)	−0.217 (***p = 0.007***)	−0.190 (***p = 0.020***)	−0.190 (***p = 0.020***)

Table A2. Correlations between latitude/longitude and bog blueberry fruit-set in the 151 regions in Northeast China with changes in precipitation from −25% to 25% (Pearson's product-moment correlation test) when air temperature was kept the same as the baseline scenario. Bold and italic numbers indicate statistically significant correlations ($\alpha = 0.05$).

	T + 0.75R	**T + 0.85R**	**T + 0.95R**	**Baseline (T + R)**	**T + 1.05R**	**T + 1.15R**	**T + 1.25R**
Latitude	0.089 ($p = 0.279$)	0.139 ($p = 0.088$)	0.193 (***p = 0.018***)	0.344 (***p < 0.001***)	0.324 (***p < 0.001***)	0.303 (***p < 0.001***)	0.374 (***p < 0.001***)
Longitude	−0.103 ($p = 0.208$)	−0.393 (***p < 0.001***)	−0.376 (***p < 0.001***)	−0.274 (***p = 0.001***)	−0.263 (***p < 0.001***)	−0.260 (***p < 0.001***)	−0.488 (***p < 0.001***)

References

1. Holloway, P.S. Managing Wild Bog Blueberry, Lingonberry, Cloudberry and Crowberry Stands in Alaska. Scholarworks. 2006. Available online: http://hdl.handle.net/11122/2828 (accessed on 25 August 2021).
2. Parkinson, L.V.; Mulder, C.P. Patterns of pollen and resource limitation of fruit production in *Vaccinium uliginosum* and *V. vitis-idaea* in Interior Alaska. *PLoS ONE* **2020**, *15*, e0224056. [CrossRef]
3. Su, S.; Wang, L.; Wu, J.; Li, B.; Wang, W.; Wang, L. Chemical compositions and functions of *Vaccinium uliginosum*. *Chin. J. Bot.* **2016**, *51*, 691.
4. Li, Y.; Yu, H. The current status and future of the blueberry industry in China. *Acta Hortic.* **2009**, *810*, 445–456.

5. Jiafeng, J.; Jiguang, W.; Hong, Y.; Shan'an, H. The developing blueberry industry in China. In *Modern Fruit Industry*; IntechOpen: London, UK, 2019.
6. Aras, P.; De Oliveira, D.; Savoie, L. Effect of a honey bee (Hymenoptera: Apidae) gradient on the pollination and yield of lowbush blueberry. *J. Econ. Entomol.* **1996**, *89*, 1080–1083. [CrossRef]
7. Asare, E.; Hoshide, A.K.; Drummond, F.A.; Criner, G.K.; Chen, X. Economic risk of bee pollination in Maine wild blueberry, *Vaccinium angustifolium*. *J. Econ. Entomol.* **2017**, *110*, 1980–1992. [CrossRef]
8. Bushmann, S.L.; Drummond, F.A. Analysis of Pollination Services Provided by Wild and Managed Bees (Apoidea) in Wild Blueberry (*Vaccinium angustifolium* Aiton) Production in Maine, USA, with a Literature Review. *J. Agron.* **2020**, *10*, 1413. [CrossRef]
9. Drummond, F.A. Behavior of bees associated with the wild blueberry agro-ecosystem in the USA. *Int. J. Entomol. Nematol.* **2016**, *2*, 21–26.
10. Javorek, S.K.; Mackenzie, K.E.; Vander Kloet, S. Comparative pollination effectiveness among bees (Hymenoptera: Apoidea) on lowbush blueberry (Ericaceae: *Vaccinium angustifolium*). *Ann. Entomol. Soc. Am.* **2002**, *95*, 345–351. [CrossRef]
11. Urbanowicz, C.; Virginia, R.A.; Irwin, R.E. Pollen limitation and reproduction of three plant species across a temperature gradient in western Greenland. *Arct. Antarct. Alp. Res.* **2018**, *50*, S100022. [CrossRef]
12. White, S.N.; Boyd, N.S.; Van Acker, R.C. Growing degree-day models for predicting lowbush blueberry (*Vaccinium angustifolium* Ait.) ramet emergence, tip dieback, and flowering in Nova Scotia, Canada. *HortScience* **2012**, *47*, 1014–1021. [CrossRef]
13. Tasnim, R.; Drummond, F.; Zhang, Y.-J. Climate Change Patterns of Wild Blueberry Fields in Downeast, Maine over the Past 40 Years. *Water* **2021**, *13*, 594. [CrossRef]
14. Russell, J.C.; Lecomte, V.; Dumont, Y.; Le Corre, M. Intraguild predation and mesopredator release effect on long-lived prey. *Ecol. Model.* **2009**, *220*, 1098–1104. [CrossRef]
15. Qu, H.; Drummond, F. Simulation-based modeling of wild blueberry pollination. *Comput. Electron. Agric.* **2018**, *144*, 94–101. [CrossRef]
16. Eaton, L.J.; Nams, V.O. Honey bee stocking numbers and wild blueberry production in Nova Scotia. *Can. J. Plant Sci.* **2012**, *92*, 1305–1310. [CrossRef]
17. Kirk, A.K.; Isaacs, R. Predicting flower phenology and viability of highbush blueberry. *HortScience* **2012**, *47*, 1291–1296. [CrossRef]
18. Yarborough, D.E. Factors Contributing to the Increase in Productivity in the Wild Blueberry Industry. *Small Fruits Rev.* **2004**, *3*, 33–43. [CrossRef]
19. Qu, H.; Seifan, T.; Tielbörger, K.; Seifan, M. A spatially explicit agent-based simulation platform for investigating effects of shared pollination service on ecological communities. *Simul. Model Pract. Theory* **2013**, *37*, 107–124. [CrossRef]
20. Koirala, A.; Walsh, K.B.; Wang, Z.; McCarthy, C. Deep learning–Method overview and review of use for fruit detection and yield estimation. *Comput. Electron. Agric.* **2019**, *162*, 219–234. [CrossRef]
21. Sirsat, M.S.; Mendes-Moreira, J.; Ferreira, C.; Cunha, M. Machine Learning predictive model of grapevine yield based on agroclimatic patterns. *Eng. Agric. Environ. Food* **2019**, *12*, 443–450. [CrossRef]
22. Chlingaryan, A.; Sukkarieh, S.; Whelan, B. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Comput. Electron. Agric.* **2018**, *151*, 61–69. [CrossRef]
23. Zhou, Z.-H. Learnware: On the future of machine learning. *Front. Comput. Sci.* **2016**, *10*, 589–590. [CrossRef]
24. Shahhosseini, M.; Martinez-Feria, R.A.; Hu, G.; Archontoulis, S.V. Maize yield and nitrate loss prediction with machine learning algorithms. *Environ. Res. Lett.* **2019**, *14*, 124026. [CrossRef]
25. Sun, B. Territory and Natural Resources Study. *Territ. Nat. Resour. Stud.* **2019**, *1*, 83–87.
26. Yarborough, D.; Drummond, F.; Annis, S.; D'Appollonio, J. In Maine wild blueberry systems analysis. In Proceedings of the XI International Vaccinium Symposium 1180, Orlando, FL, USA, 10–14 April 2016; pp. 151–160.
27. Yarborough, D. In Improving Northern bilberry (Vaccinium uliginosum) production. In Proceedings of the X International Symposium on Vaccinium and Other Superfruits 1017, Maastricht, The Netherlands, 17–22 June 2012; pp. 223–229.
28. Bell, D.J.; Rowland, L.J.; Stommel, J.; Drummond, F.A. Yield variation among clones of lowbush blueberry as a function of genetic similarity and self-compatibility. *J. Am. Soc. Hortic.* **2010**, *135*, 259–270. [CrossRef]
29. Alsos, I.G.; Engelskjøn, T.; Brochmann, C. Conservation genetics and population history of *Betula nana*, *Vaccinium uliginosum*, and *Campanula rotundifolia* in the arctic archipelago of Svalbard. *Arct. Antarct. Alp.* **2002**, *34*, 408–418. [CrossRef]
30. Obsie, E.Y.; Qu, H.; Drummond, F. Wild blueberry yield prediction using a combination of computer simulation and machine learning algorithms. *Comput. Electron. Agric.* **2020**, *178*, 105778. [CrossRef]
31. Harteveld, D.O.; Grant, M.R.; Pscheidt, J.W.; Peever, T.L. Predicting Ascospore release of *Monilinia vaccinii-corymbosi* of blueberry with machine learning. *Phytopathology* **2017**, *107*, 1364–1371. [CrossRef] [PubMed]
32. Abdel-Sattar, M.; Aboukarima, A.M.; Alnahdi, B.M. Application of artificial neural network and support vector regression in predicting mass of berry fruits (*Ziziphus mauritiana* Lamk.) based on fruit axial dimensions. *PLoS ONE* **2021**, *16*, e0245228. [CrossRef]
33. Cortes, C. Support-vector network. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
34. Drucker, H.; Burges, C.J.; Kaufman, L.; Smola, A.; Vapnik, V. Support vector regression machines. *Adv. Neural Inf. Process. Syst.* **1997**, *9*, 155–161.
35. Smola, A.; Schölkopf, B. A tutorial on support vector regression. *Stat. Comput.* **2004**, *14*, 199–222. [CrossRef]
36. Aha, D.W.; Kibler, D.; Albert, M.K. Instance-based learning algorithms. *Mach. Learn.* **1991**, *6*, 37–66. [CrossRef]

37. Appelhans, T.; Mwangomo, E.; Hardy, D.R.; Hemp, A.; Nauss, T. Evaluating machine learning approaches for the interpolation of monthly air temperature at Mt. Kilimanjaro, Tanzania. *Spat. Stat.* **2015**, *14*, 91–113. [CrossRef]
38. Chen, G.H.; Shah, D. *Explaining the Success of Nearest Neighbor Methods in Prediction*; Now Publishers: Boston, MA, USA, 2018.
39. Song, Y.; Lu, Y. Decision tree methods: Applications for classification and prediction. *Shanghai Arch. Psychiatr.* **2015**, *27*, 130–135.
40. Quinlan, J.R.J. Ross Quinlan_C4. 5_ Programs for Machine Learning. pdf. *Morgan Kaufmann* **1993**, *5*, 302.
41. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
42. Cao, J.; Zhang, Z.; Tao, F.; Zhang, L.; Luo, Y.; Han, J.; Li, Z. Identifying the contributions of multi-source data for winter wheat yield prediction in China. *Remote Sens.* **2020**, *12*, 750. [CrossRef]
43. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [CrossRef]
44. Konduri, V.S.; Vandal, T.J.; Ganguly, S.; Ganguly, A.R. Data science for weather impacts on crop yield. *Front. Sustain. Food Syst.* **2020**, *4*, 52. [CrossRef]
45. Freund, Y.; Schapire, R.E. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.* **1997**, *55*, 119–139. [CrossRef]
46. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
47. Friedman, J. Greedy function approximation: A gradient boosting machine. *Ann. Statist.* **2001**, *29*, 1189–1232. [CrossRef]
48. Kaul, M.; Hill, R.L.; Walthall, C. Artificial neural networks for corn and soybean yield prediction. *Agric. Syst.* **2005**, *85*, 1–18. [CrossRef]
49. Brown, M.E.; Lary, D.J.; Vrieling, A.; Stathakis, D.; Mussa, H. Neural networks as a tool for constructing continuous NDVI time series from AVHRR and MODIS. *Int. J. Remote Sens.* **2008**, *29*, 7141–7158. [CrossRef]
50. Hu, Q.; Weng, X. Estimating impervious surfaces from medium spatial resolution imagery using the self-organizing map and multi-layer perceptron neural networks. *Remote Sens.* **2009**, *113*, 2089–2102. [CrossRef]
51. Chen, T.; He, T.; Benesty, M.; Khotilovich, V.; Tang, Y.; Cho, H. Xgboost: Extreme gradient boosting. Available online: https://cran.r-project.org/web/packages/xgboost/index.html (accessed on 25 August 2021).
52. Romeiko, X.X.; Guo, Z.; Pang, Y.; Lee, E.K.; Zhang, X. Comparing Machine Learning Approaches for Predicting Spatially Explicit Life Cycle Global Warming and Eutrophication Impacts from Corn Production. *Sustainability* **2020**, *12*, 1481. [CrossRef]
53. Montgomery, D.C.; Peck, E.A.; Vining, G.G. *Introduction to Linear Regression Analysis*; John Wiley & Sons: Hoboken, NJ, USA, 2021.
54. Qu, H.; Yin, L.; Tang, X. An automatic clustering method using multi-objective genetic algorithm with gene rearrangement and cluster merging. *Appl. Soft Comput.* **2021**, *99*, 106929. [CrossRef]
55. Scott, L.; Janikas, M. Spatial statistics in ArcGIS. In *Handbook of Applied Spatial Analysis*; Springer: Berlin, Germany, 2010.
56. Puntel, L.A.; Sawyer, J.E.; Barker, D.W.; Thorburn, P.J.; Castellano, M.J.; Moore, K.J.; VanLoocke, A.; Heaton, E.A.; Archontoulis, S.V. A systems modeling approach to forecast corn economic optimum nitrogen rate. *Front. Plant Sci.* **2018**, *9*, 436. [CrossRef] [PubMed]
57. Pan, S.J.; Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **2009**, *22*, 1345–1359. [CrossRef]
58. Coulibaly, S.; Kamsu-Foguem, B.; Kamissoko, D.; Traore, D. Deep neural networks with transfer learning in millet crop images. *Comput. Ind.* **2019**, *108*, 115–120. [CrossRef]
59. Rangarajan Aravind, K.; Raja, P. Automated disease classification in (Selected) agricultural crops using transfer learning. *Automatika* **2020**, *61*, 260–272. [CrossRef]
60. Sun, X.; Wei, J. In Identification of maize disease based on transfer learning. *J. Phys. Conf. Ser.* **2020**, *1437*, 012080. [CrossRef]
61. Mendes, A.; Togelius, J.; Coelho, L.d.S. Multi-Stage Transfer Learning with an Application to Selection Process. *arXiv* **2020**, arXiv:2006.01276.
62. Lee, C.-K.; Lu, C.; Yu, Y.; Sun, Q.; Hsieh, C.-Y.; Zhang, S.; Liu, Q.; Shi, L. Transfer learning with graph neural networks for optoelectronic properties of conjugated oligomers. *J. Chem. Phys.* **2021**, *154*, 024906. [CrossRef]
63. Storm, H.; Baylis, K.; Heckelei, T. Machine learning in agricultural and applied economics. *Eur. Rev. Agric. Econ.* **2020**, *47*, 849–892. [CrossRef]
64. Drummond, F.A.; Rowland, L.J. The ecology of autogamy in wild blueberry (*Vaccinium angustifolium* Aiton): Does the early clone get the bee? *J. Agron.* **2020**, *10*, 1153.
65. Moon, T.; Son, J.E. Knowledge transfer for adapting pre-trained deep neural models to predict different greenhouse environments based on a low quantity of data. *Comput. Electron. Agric.* **2021**, *185*, 106136. [CrossRef]
66. Jacquemart, A.-L. Biological flora of the British Isles, no. 193. *Vaccinium uliginosum* L. *J. Ecol.* **1996**, *84*, 771–785. [CrossRef]
67. Kong, W.-S.; Kim, K.; Lee, S.; Park, H.; Cho, S.-H. Distribution of high mountain plants and species vulnerability against climate change. *J. Environ. Impact Assess.* **2014**, *23*, 119–136. [CrossRef]
68. Drummond, F. Reproductive biology of wild blueberry (*Vaccinium angustifolium* Aiton). *Agriculture* **2019**, *9*, 69. [CrossRef]
69. Graae, B.J.; Alsos, I.G.; Ejrnaes, R. The impact of temperature regimes on development, dormancy breaking and germination of dwarf shrub seeds from arctic, alpine and boreal sites. *Plant Ecol.* **2008**, *198*, 275–284. [CrossRef]
70. Aerts, R.; Cornelissen, J.; Van Logtestijn, R.; Callaghan, T. Climate change has only a minor impact on nutrient resorption parameters in a high-latitude peatland. *Oecologia* **2007**, *151*, 132–139. [CrossRef] [PubMed]

71. Jones, J.W.; Antle, J.M.; Basso, B.; Boote, K.J.; Conant, R.T.; Foster, I.; Godfray, H.C.J.; Herrero, M.; Howitt, R.E.; Janssen, S. Toward a new generation of agricultural system data, models, and knowledge products: State of agricultural systems science. *Agric. Syst.* **2017**, *155*, 269–288. [CrossRef] [PubMed]
72. McNunn, G.; Heaton, E.; Archontoulis, S.; Licht, M.; VanLoocke, A. Using a crop modeling framework for precision cost-benefit analysis of variable seeding and nitrogen application rates. *Front. Sustain. Food Syst.* **2019**, *3*, 108. [CrossRef]