

## Research Article

# Prediction of Frost Occurrences Using Statistical Modeling Approaches

Hyojin Lee,<sup>1</sup> Jong A. Chun,<sup>1</sup> Hyun-Hee Han,<sup>2</sup> and Sung Kim<sup>3</sup>

<sup>1</sup>APEC Climate Center, Climate Research Department, 12 Centum 7-ro, Haeundae-gu, Busan 48058, Republic of Korea

<sup>2</sup>Department of Horticultural Crop Research, National Institute of Horticultural and Herbal Science, 100 Nongsaeongmyeong-ro, Iseo-myeon, Wanju-gun, Jeollabuk-do 55365, Republic of Korea

<sup>3</sup>Republic of Korea Air Force Weather Wing, Gyeryong-si, Chungcheongnam-do 32809, Republic of Korea

Correspondence should be addressed to Jong A. Chun; [jachun@apcc21.org](mailto:jachun@apcc21.org)

Received 5 February 2016; Revised 24 March 2016; Accepted 29 March 2016

Academic Editor: Herminia García Mozo

Copyright © 2016 Hyojin Lee et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We developed the frost prediction models in spring in Korea using logistic regression and decision tree techniques. Hit Rate (HR), Probability of Detection (POD), and False Alarm Rate (FAR) from both models were calculated and compared. Threshold values for the logistic regression models were selected to maximize HR and minimize FAR for each station, and the split for the decision tree models was stopped when change in entropy was relatively small. Average HR values were 0.92 and 0.91 for logistic regression and decision tree techniques, respectively, average POD values were 0.78 and 0.80 for logistic regression and decision tree techniques, respectively, and average FAR values were 0.22 and 0.28 for logistic regression and decision tree techniques, respectively. The average numbers of selected explanatory variables were 5.7 and 2.3 for logistic regression and decision tree techniques, respectively. Fewer explanatory variables can be more appropriate for operational activities to provide a timely warning for the prevention of the frost damages to agricultural crops. We concluded that the decision tree model can be more useful for the timely warning system. It is recommended that the models should be improved to reflect local topological features.

## 1. Introduction

It is widely known that many perennial crops such as fruit tree in South Korea are prone to be damaged from late-spring frost events. Dehydration resulting from the extracellular ice formation leads to permanent tissue damage of crops through a freeze event from a frost event [1, 2]. Frost events can be divided into two categories: radiation frosts and advective frosts [3]. The former tends to occur at the meteorological characteristics of clear skies, no wind, and a low dewpoint temperature. The latter typically occurs under the meteorological conditions of cloudy skies, moderate to strong winds, no temperature inversion, and low humidity.

There have been many studies on meteorological conditions at frost events [4, 5]. Kwon et al. [4] analyzed the following eight meteorological variables at each station in South Korea when frost events occurred from 1973 to 2007: minimum temperature (denoted as TMIN), grass minimum

temperature (denoted as GMINT), dewpoint temperature (denoted as Dewpoint), and wind speed (denoted as Wind) on frost occurrence days, mean relative humidity (denoted as RH<sub>mean</sub>), minimum relative humidity (denoted as RH<sub>min</sub>), and cloud amount (denoted as Cloud) on one day before the frost occurrence days, and difference between maximum temperature on one day before the frost occurrence days and minimum temperature on the frost occurrence days (denoted as T<sub>diff</sub>). These meteorological variables have been used to estimate the frost probability. For example, Floor [6] used wind speed, total cloud amount, minimum temperature, and grass minimum temperature for the estimation of frost events at Eelde (Netherlands).

Frost warning systems have been developed based on meteorological variables. Three levels of frost warnings are currently issued by The National Weather Service (NWS) based on air temperatures and wind speeds [7]. The criteria for the three levels (frost warning, frost/free warning, and

frost warning) are  $0^{\circ}\text{C}$  and  $16\text{ km h}^{-1}$ , for air temperature and wind speed, respectively. Chevalier et al. [3] developed a web-based fuzzy expert system for frost event warnings based on predicted air and dewpoint temperatures and observed current wind speeds. An expert system was used to predict the frost occurrence on roads and bridges [8]. The prediction systems used observed maximum and minimum temperatures from the previous day and estimates of air temperature, dewpoint temperature, cloud cover, precipitation, and average wind speed.

Frost occurrences can be described as a binary variable since we can divide days into two: days when frost occurs and when frost does not occur. The logistic regression and decision tree techniques can be used to predict frost occurrences. Those techniques have been used for various fields including agronomy [9, 10], meteorology [11], and medicine [12]. However, to the best of our knowledge, few studies have been conducted on frost prediction using those techniques. These techniques are much simpler than fuzzy logic techniques. These simple methods can be more useful for operational activities to prevent frost damages to agricultural crops. The objective of this study was to develop the prediction models of frost occurrences using the logistic regression and decision techniques. We identified the most selected meteorological variables by the two techniques and compared the frost prediction models from both techniques.

## 2. Materials and Methods

**2.1. Study Site and Data Collection.** For this study, six stations (Chuncheon: 101, Suwon: 119, Seosan: 129, Cheongju: 131, Gwangju: 156, and Jinju: 192) were selected as described in Figure 1. The numbers after each station are the station numbers named by the Korea Meteorological Administration (KMA). Based on a study by Kwon et al. [4], the eight meteorological variables were selected for this study. The frost occurrences and eight meteorological variables from the years 1973 to 2014 were collected from the KMA. For the developed system for frost predictions or warnings, two [7] to seven [8] meteorological variables were used. For this study, spring seasons defined as March to May were focused, especially since late frost in spring seasons frequently damaged overall crop growth seriously.

The meteorological characteristics and their statistics at the frost occurrence days at the six stations from 1973 to 2014 are summarized in Table 1. The Seosan station with 909 days found that the most frost events occurred, while the smallest was observed at the Gwangju station with 510 days. Frost in spring occurred when minimum temperature is approximately  $-2^{\circ}\text{C}$  at most of stations except for Gwangju station ( $-0.2^{\circ}\text{C}$ ). Frost events at Chuncheon, Suwon, Seosan, and Cheongju stations were observed when the range of grass minimum temperature was between  $-6.1$  and  $-7.4^{\circ}\text{C}$ , while frost events at Gwangju and Jinju stations occurred when the range of grass minimum temperature was between  $-4.1$  and  $-4.9^{\circ}\text{C}$ . For wind speed, the range when frost occurred was  $1.6$  to  $2.3\text{ m s}^{-1}$ .

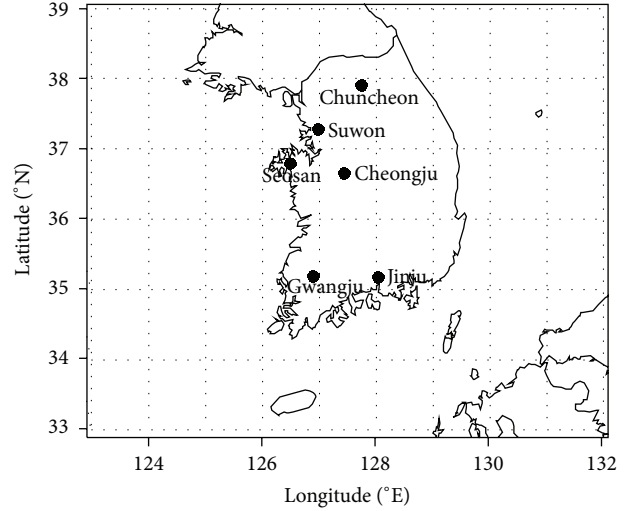


FIGURE 1: Location map of study sites used to analyze meteorological characteristics at frost occurrences.

**2.2. Logistic Regression.** The eight meteorological variables which were identified as the most influential meteorological variables on frost events by Kwon et al. [4] were used for the explanatory variables of the prediction models. Frost occurrences ( $Y$  of (2)) can be defined as binary variable (0 for no frost event and 1 for frost event) in logistic regression modeling. For frost occurrence probability  $\pi(x_i)$ , the logistic regression equation can be given as

$$\text{logit} [\pi(x_i)] = \exp(\alpha_0 + \beta'x_i), \quad (1)$$

where  $\alpha_0$  is the intercept of the linear term and  $\beta'$  is the coefficients corresponding to explanatory variable  $x_i$ . The probability where frost events occur can be defined as

$$P(Y = 1) = \frac{\exp(\alpha_0 + \beta'x_i)}{1 + \exp(\alpha_0 + \beta'x_i)}. \quad (2)$$

On the other hand, the probability where frost events do not occur can be calculated as

$$P(Y = 0) = 1 - P(Y = 1). \quad (3)$$

The Hosmer-Lemeshow significance test [13] with a significance level of 0.05 was used as a criterion to select explanatory variables. A backward elimination method was used for the selection of the explanatory variables. First, the model was fitted with all those eight meteorological variables, and the explanatory variable with the highest  $p$  value greater than the selection criterion (i.e., insignificant at the significance level of 0.05) was removed. Second, the model was fitted with the rest of the explanatory variables and we eliminated the explanatory variable whose  $p$  value was highest among insignificant explanatory variables at the significance level of 0.05. This elimination was repeated until all explanatory variables were significant at the significance level of 0.05.

TABLE 1: Meteorological characteristics at frost occurrences at the study stations in spring.

Station	Statistic	TMIN	GMINT <sup>†</sup> °C	Dewpoint	$T_{diff}$ <sup>‡</sup>	RH <sub>mean</sub> <sup>§</sup> %	RH <sub>min</sub> <sup>¶</sup>	Wind m s <sup>-1</sup>	Cloud %
Chuncheon	<i>N</i>	896	896	896	895	895	895	896	895
	Min.	-16.2	-23	-21.6	3.7	25.1	8.0	0.2	0.0
	Max.	11.8	4.7	10.9	43.2	97.8	96.0	6.0	100.0
	Mean	-2.6	-7.4	-3.7	13.9	60.9	31.5	1.6	36.9
	Std.	2.6	3.0	3.6	4.3	11.9	12.9	0.8	28.7
Suwon	<i>N</i>	792	792	792	792	792	792	792	792
	Min.	-11.3	-15.7	-19.8	4.8	26.0	11.0	0.2	0.0
	Max.	6.1	2.1	6.6	36	97.3	95.0	5.5	100.0
	Mean	-1.4	-6.7	-2.0	12.0	64.8	36.1	1.7	33.6
	Std.	2.6	3.2	3.5	3.9	12.0	12.5	0.7	29.2
Seosan	<i>N</i>	909	909	909	909	909	909	909	909
	Min.	-9.8	-14.8	-15.5	3.2	30.4	11.0	0.6	0.0
	Max.	5.7	0.3	6.9	37.7	97.8	91.0	7.0	100.0
	Mean	-2.0	-6.1	-2.0	12.5	67.5	37.1	2.3	35.8
	Std.	2.5	2.8	3.5	4.0	11.8	13.7	1.0	28.3
Cheongju	<i>N</i>	683	683	683	683	683	683	683	683
	Min.	-12	-17.4	-13.1	3.8	28.8	10.0	0.5	0.0
	Max.	5.4	0.8	5.6	38.8	96.5	93.0	5.3	100.0
	Mean	-1.6	-6.9	-3.0	13.3	60.7	31.9	1.9	35.8
	Std.	2.5	3.0	3.4	4.3	11.1	12.6	0.8	28.5
Gwangju	<i>N</i>	510	510	510	509	509	509	510	509
	Min.	-10.1	-12.6	-12.2	2.9	32.3	6.0	0.5	0.0
	Max.	8.6	5.3	9.7	36.2	92.3	81.0	7.0	100.0
	Mean	-0.4	-4.1	-2.3	12.1	61.4	33.7	2.1	35.8
	Std.	2.2	2.4	3.2	4.1	11.4	12.8	0.8	27.8
Jinju	<i>N</i>	537	537	537	537	537	537	537	537
	Min.	-10.4	-14.1	-16.5	5.3	25.0	8.0	0.4	0.0
	Max.	5.1	2.4	6.9	42.7	90.8	82.0	5.6	100.0
	Mean	-2.2	-4.9	-2.8	15.9	56.5	26.6	1.7	29.9
	Std.	2.2	2.6	3.9	4.1	13.1	12.3	0.7	27.8

<sup>†</sup> Grass minimum temperature.

<sup>‡</sup> Difference between maximum (one day before frost occurrence day) and minimum (on frost occurrence day) temperatures.

<sup>§</sup> Mean relative humidity.

<sup>¶</sup> Minimum relative humidity.

2.3. *Decision Tree.* Frost occurrence can be described as binary feature (i.e., “Yes-No” response). For this study, a binary decision tree model was used for the frost prediction model. For the split, entropy was used for a separation criterion. Parent’s node entropy ( $E$ ) was calculated by (4) as follows:

$$E = -\sum_{j=1}^2 \frac{n_{+j}}{n_{++}} \log\left(\frac{n_{+j}}{n_{++}}\right), \quad (4)$$

$$E_L = -\sum_{j=1}^2 \frac{n_{1j}}{n_{1+}} \log\left(\frac{n_{1j}}{n_{1+}}\right), \quad (5)$$

$$E_R = -\sum_{j=1}^2 \frac{n_{2j}}{n_{2+}} \log\left(\frac{n_{2j}}{n_{2+}}\right), \quad (6)$$

$$E_D = E - \frac{n_{1+}}{n_{++}} E_L - \frac{n_{2+}}{n_{++}} E_R, \quad (7)$$

TABLE 2: Two-dimensional ( $2 \times 2$ ) contingency table.

Observation	Forecast		Total
	No frost	Frost occurrence	
No frost	$D$ (negative correction)	$B$ (false alarm)	$B + D$
Frost occurrence	$C$ (miss)	$A$ (hit)	$A + C$
Total	$C + D$	$A + B$	$A + B + C + D$

where  $n_{++}$  is the number of parent's node data and  $n_{ij}$  is the number of data in  $i$ th node and  $j$  class.  $E$  is the entropy of parent's node,  $E_L$  is the entropy of the left branch,  $E_R$  is the entropy of the right branch, and  $E_D$  is the entropy difference between decision node and its branches.

A decision tree model consists of decision nodes and leaf nodes. Each decision node has exactly two branches in a binary tree. The topmost decision node (i.e., root node) in a tree was selected as an explanatory variable (i.e., the best predictor variable) with the maximum entropy. The frost occurrences were split to maximize the difference of  $E_L$  and  $E_R$ . The entropies of two branches were then calculated using (5) and (6). The explanatory variable with the largest  $E_D$  by (7) was selected as a decision node. To avoid overfitting problem in building decision trees, we stopped the split where change in  $E_D$  was relatively small.

**2.4. Logistic Regression versus Decision Tree.** Logistic regression and decision tree techniques were used to develop prediction models for frost occurrence events in spring, and their predictabilities were compared. The results predicted from these developed models were summarized in the  $2 \times 2$  contingency table (i.e., Table 2). Using these tables, we calculated Hit Rate (HR), Probability of Detection (POD), and False Alarm Rate (FAR) as the skill scores for the predictability of the models and these skill scores from both models were compared. These skill scores can be calculated by

$$\begin{aligned} \text{HR} &= \frac{A + D}{A + B + C + D}, \\ \text{POD} &= \frac{A}{A + C}, \\ \text{FAR} &= \frac{B}{A + B}, \end{aligned} \quad (8)$$

where  $A$ ,  $B$ ,  $C$ , and  $D$  are defined in Table 2.

The probability of frost occurrences using (2) can be calculated. When the probability is greater than a threshold value (default = 0.5), the model will predict that a frost event occurs. In contrast, the model will predict that a frost event does not occur with the probability smaller than a threshold value.  $A$ ,  $B$ ,  $C$ , and  $D$  in (8) can be changed by adjusting the threshold value. Subsequently, the three skill scores (HR, POD, and FAR) can be changed depending on

the threshold value. We selected threshold values for the logistic regression models to maximize HR and POD and minimize FAR for each station. The Leave-One-Out-Cross Validation technique was used to estimate the performance of the developed models using both techniques.

Since frost events are very local and too many meteorological variables may not be always available at all local sites, limited number of explanatory variables might be more suitable for the operational purpose of the frost occurrence prediction models. We compared the developed models using both logistic regression and decision tree techniques and proposed a better technique for the operational use based on their performance and the number of selected explanatory variables.

### 3. Results and Discussion

**3.1. Logistic Regression.** The prediction models for the frost occurrence events were developed using the logistic regression techniques. As shown in Table 3, the threshold values which were determined to maximize HR and POD and to minimize FAR for the logistic regression model varied with the range of 0.42 (Jinju) to 0.49 (Suwon). The HR, POD, and FAR values resulting from these threshold values were approximately 0.9, 0.8, and 0.2, respectively (Table 4). Three variables, TMIN, Dewpoint, and Wind, were selected at all six stations, while the variables,  $\text{RH}_{\text{mean}}$  and cloud, were used at five stations except for the Suwon and Cheongju stations, respectively. Particularly, TMIN and Wind among the three variables have been commonly used for the predictions of frost events [3, 6–8]. Seven out of eight meteorological variables except for  $T_{\text{diff}}$  were selected for the prediction models using the logistic regression technique. These seven variables have been frequently found in previous studies [3, 6, 8].  $T_{\text{diff}}$  in those studies have not been used for the frost predictions or warnings. GMINT was selected at four stations and  $\text{RH}_{\text{min}}$  was the least selected explanatory variable. These results implied that minimum and grass minimum temperatures are negatively correlated with the probability of the frost occurrence. On the contrary, Dewpoint and  $\text{RH}_{\text{mean}}$  are positively correlated with that. Intercepts of each equation (Table 3) at the six stations varied with the range of  $-2.554$  (Cheongju) and  $-0.19$  (Suwon). The highest coefficient of TMIN was  $-0.262$  at Gwangju and the lowest one  $-0.92$  at Jinju. Although these two stations are located in the same latitude, odds ratio of TMIN at Gwangju is approximately twice as high as that at Jinju. While the coefficients of TMIN largely varied over the six stations, the variations of the coefficients of Wind and Dewpoint were somewhat small over the six stations. Average coefficients of Wind and Dewpoint were about 0.20 and  $-0.42$ , respectively.

Figure 2 depicts the observed and predicted number of daily frost events in spring during the study period. The results from the Leave-One-Out Cross Validation method are also described in Figure 2. The models overestimated the frost events in early March at all stations, while slightly underestimated results were shown in April at all stations.

TABLE 3: Logistic regression models and threshold values for stations.

Station	$\alpha_0 + \beta'x$ in (2)	Threshold
Chuncheon	$-0.384 - 0.761\text{TMIN} + 0.150\text{Dewpoint} + 0.019\text{RH}_{\text{mean}} - 0.372\text{Wind} - 0.104\text{Cloud}$	0.43
Suwon	$-0.190 - 0.484\text{TMIN} - 0.364\text{GMINT} + 0.264\text{Dewpoint} - 0.461\text{Wind} - 0.094\text{Cloud}$	0.49
Seosan	$-1.44 - 0.498\text{TMIN} - 0.389\text{GMINT} + 0.186\text{Dewpoint} + 0.022\text{RH}_{\text{mean}} - 0.381\text{Wind} - 0.107\text{Cloud}$	0.44
Cheongju	$-2.554 - 0.416\text{TMIN} - 0.501\text{GMINT} + 0.292\text{Dewpoint} + 0.034\text{RH}_{\text{mean}} - 0.026\text{RH}_{\text{min}} - 0.343\text{Wind}$	0.48
Gwangju	$-1.525 - 0.262\text{TMIN} - 0.539\text{GMINT} + 0.153\text{Dewpoint} + 0.037\text{RH}_{\text{mean}} - 0.529\text{Wind} - 0.225\text{Cloud}$	0.45
Jinju	$-1.499 - 0.920\text{TMIN} + 0.152\text{Dewpoint} + 0.047\text{RH}_{\text{mean}} - 0.029\text{RH}_{\text{min}} - 0.439\text{Wind} - 0.0175\text{Cloud}$	0.42

TABLE 4: Hit Rate (HR), Probability of Detection (POD), and False Alarm Rate (FAR) by the logistic regression and decision tree models.

Station	Logistic regression			Decision tree		
	HR	POD	FAR	HR	POD	FAR
Chuncheon	0.899	0.767	0.190	0.900	0.813	0.259
Suwon	0.913	0.798	0.231	0.903	0.808	0.311
Seosan	0.918	0.812	0.154	0.907	0.866	0.285
Cheongju	0.927	0.816	0.242	0.920	0.731	0.139
Gwangju	0.929	0.747	0.303	0.916	0.761	0.469
Jinju	0.935	0.760	0.216	0.940	0.804	0.246

3.2. *Decision Tree.* Figure 3 displays the results using the decision tree technique. For example, the tree in Figure 3(a) represents that the tree model predicts that frost will occur when  $\text{TMIN} < -0.35^\circ\text{C}$  and  $\text{Wind} > 2.45 \text{ m s}^{-1}$ . Like TMIN in Figure 3(a), the same meteorological variable might be used for the separation criterion in a tree. The terminal nodes (i.e., denoted as “N” in Figure 3) which were classified into nonfrost occurrences had different probability of frost occurrences. For example, the tree model in Figure 3(a) will predict no frost as long as  $\text{TMIN} \geq -0.35^\circ\text{C}$ . However, the probability of no frost can be different whether the interval which consists of  $-0.35$  and  $1.05^\circ\text{C}$  includes minimum temperature. For this case (i.e., Figure 3(a)), the probability of no frost was 0.98 when  $\text{TMIN} \geq 1.05^\circ\text{C}$  and that was 0.61 when  $-0.35^\circ\text{C} \leq \text{TMIN} < 1.05^\circ\text{C}$ . The average of Wind was  $3.1 \text{ m s}^{-1}$ . This value is slightly lower than the criterion of wind speed ( $4.4 \text{ m s}^{-1}$ ) used by Perry [7] and slightly higher than that ( $2 \text{ m s}^{-1}$ ) of Floor [6] and is between the two criteria proposed by Chevalier et al. [3]. They proposed three categories based on two wind speeds (less than  $2.2 \text{ m s}^{-1}$ , between  $2.2 \text{ m s}^{-1}$  and  $4.4 \text{ m s}^{-1}$ , and greater than  $4.4 \text{ m s}^{-1}$ ).

The two to three variables were used in the decision tree model for frost prediction at study stations. For example, TMIN and Wind were selected at Chuncheon station, and GMINT, Dewpoint, and Wind were selected at Cheongju station. Unlike the result of the logistic regression technique, TMIN was selected at only three stations, while GMINT was selected at the same stations as the results of the logistic regression technique (i.e., Suwon, Seosan, Cheongju,

and Gwangju stations). Overall, four explanatory variables (TMIN, GMINT, Wind, and Dewpoint) out of the eight meteorological variables were selected at the study stations with the decision tree technique. Wind was used at five stations for the decision tree technique. This implies that wind speed is the most important separation criterion for the construction of tree.

The predicted frost events using the decision tree technique were compared with the observed frost events for the study period (Figure 4). Unlike the results using the logistic regression technique, there was no apparent pattern in early March. However, the model using the decision tree technique overestimated frost events at the Cheongju station in March and underestimated those in April.

3.3. *Logistic Regression versus Decision Tree.* The average numbers of selected explanatory variables were 5.7 and 2.3 for logistic regression and decision tree models, respectively. The number of selected variables for logistic regression models is slightly smaller than the number of variables (7) used by Takle [8]. While that for decision tree models is slightly higher than that (two variables) by Perry [7], that is, slightly smaller than those by Floor [6] and Chevalier et al. [3]. Four and three meteorological variables were used by Floor [6] and Chevalier et al. [3], respectively. TMIN, Wind, and Dewpoint were selected at all six stations from the logistic regression technique, while Wind (five stations) was the most selected variable from the decision tree technique,



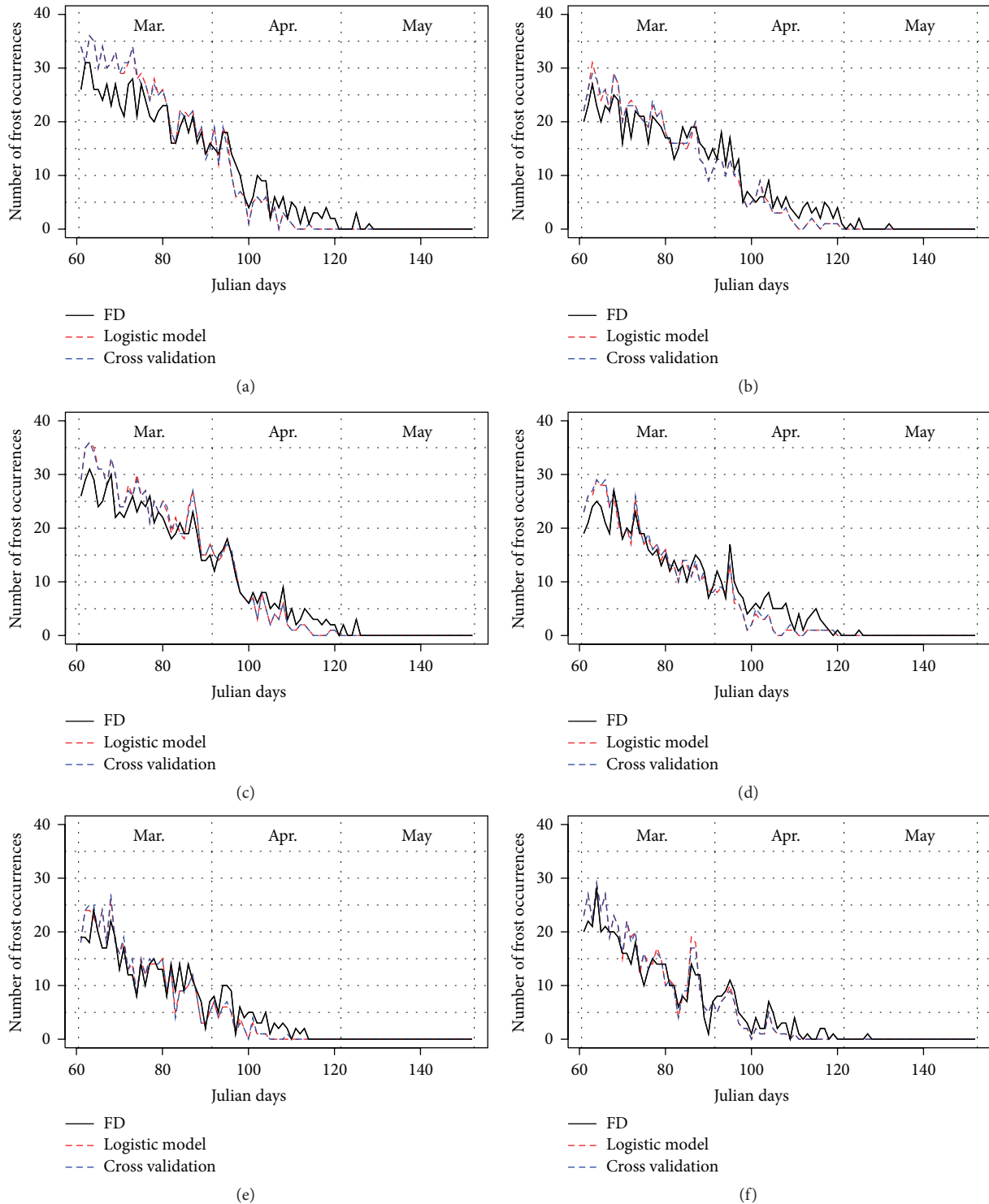


FIGURE 2: Observed and predicted frost in spring using the logistic regression models: (a) Chuncheon (101), (b) Suwon (119), (c) Seosan (129), (d) Cheongju (131), (e) Gwangju (156), and (f) Jinju (192).

followed by GMINT (four stations) and TMIN (three stations). Particularly, GMINT were most selected (at the same stations: Suwon, Seosan, Cheongju, and Gwangju) from both techniques. This result seems consistent with the result where

Kwon et al. [4] reported that GMINT was more influential than TMIN.

Skill scores calculated using the  $2 \times 2$  contingency table for the two different models are summarized in Table 4.

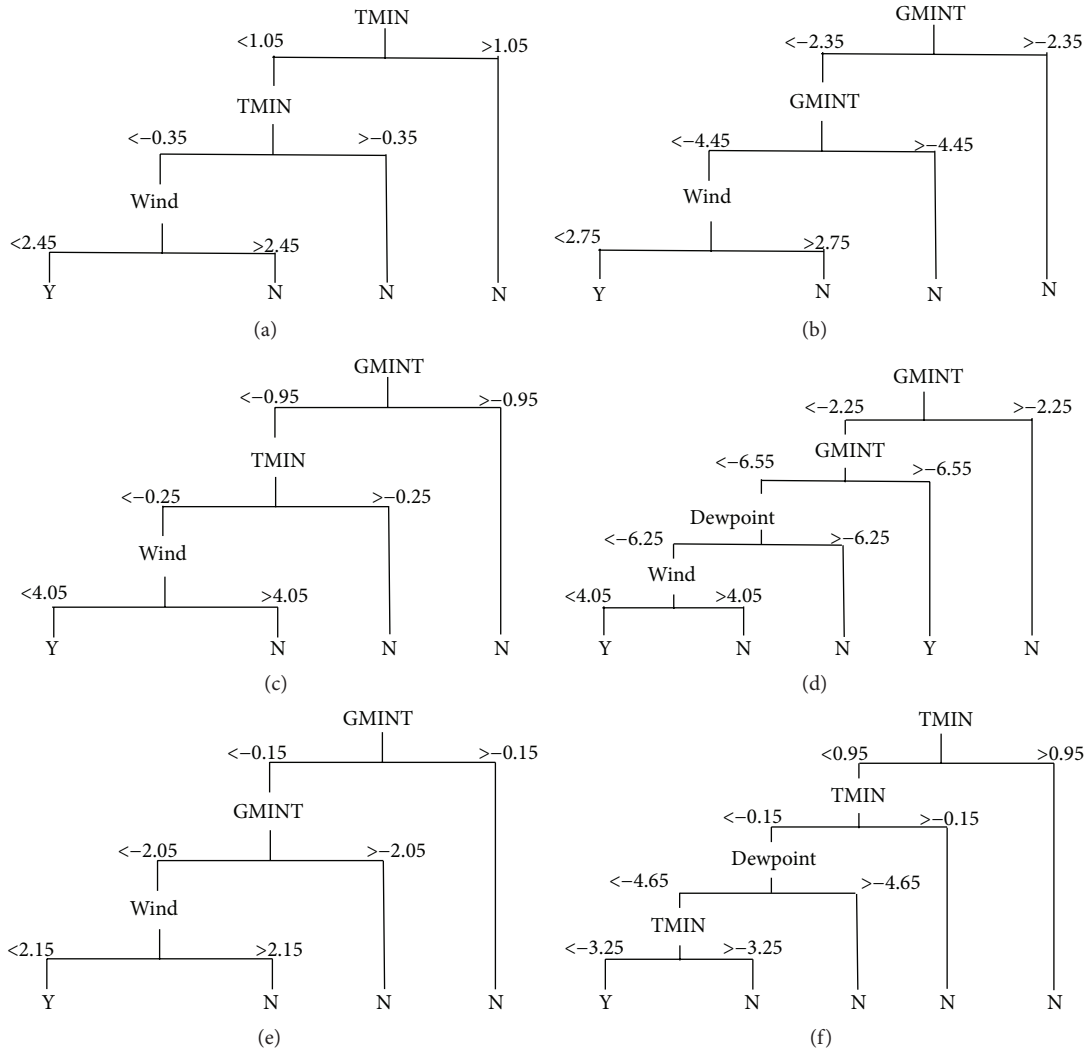


FIGURE 3: Determined number of nodes using the decision tree models: (a) Chuncheon (101), (b) Suwon (119), (c) Seosan (129), (d) Cheongju (131), (e) Gwangju (156), and (f) Jinju (192). GMINT: grass minimum temperature, TMIN: minimum temperature, and Wind: wind speed.

The HR and POD values for the logistic regression models varied with the range between 0.899 (Chuncheon) and 0.935 (Jinju) and with the range 0.767 (Chuncheon) to 0.816 (Cheongju), respectively (Table 4). The HR values for the decision tree models varied with a similar range to that for the logistic regression models. The POD values for the decision tree models were slightly higher than those for the logistic regression models, while the FAR values for the decision tree models were slightly higher than those for the logistic regression models. Unfortunately, those skill scores did not show consistent results from the two techniques. However, the POD values would be more appropriate skill score for the case where frost events do not occur very often, since a main purpose of the frost prediction model would be a timely warning for the prevention of the frost damages. Furthermore, fewer explanatory variables were selected for the decision tree models, suggesting that these models can be more appropriate for operational activities. These results imply that the decision tree model for the frost

occurrence event prediction can be more useful to provide a timely warning for the prevention of the frost damages to agricultural farms. It is concluded that this proposed technique may be useful to better support farmers by providing adequate strategies to reduce frost damages through a timely warning.

#### 4. Conclusions

To develop prediction models for frost events in the six stations in the Korean peninsula, we used the logistic regression and decision tree techniques using eight meteorological variables. Although three skill scores (HR, POD, and FAR) resulting from the two techniques were not consistent, the decision tree models were selected for the potential operational activities since fewer explanatory variables were used in the models and higher POD values than those for the logistic regression models. It is concluded that the prediction models of frost occurrence events in spring can be used

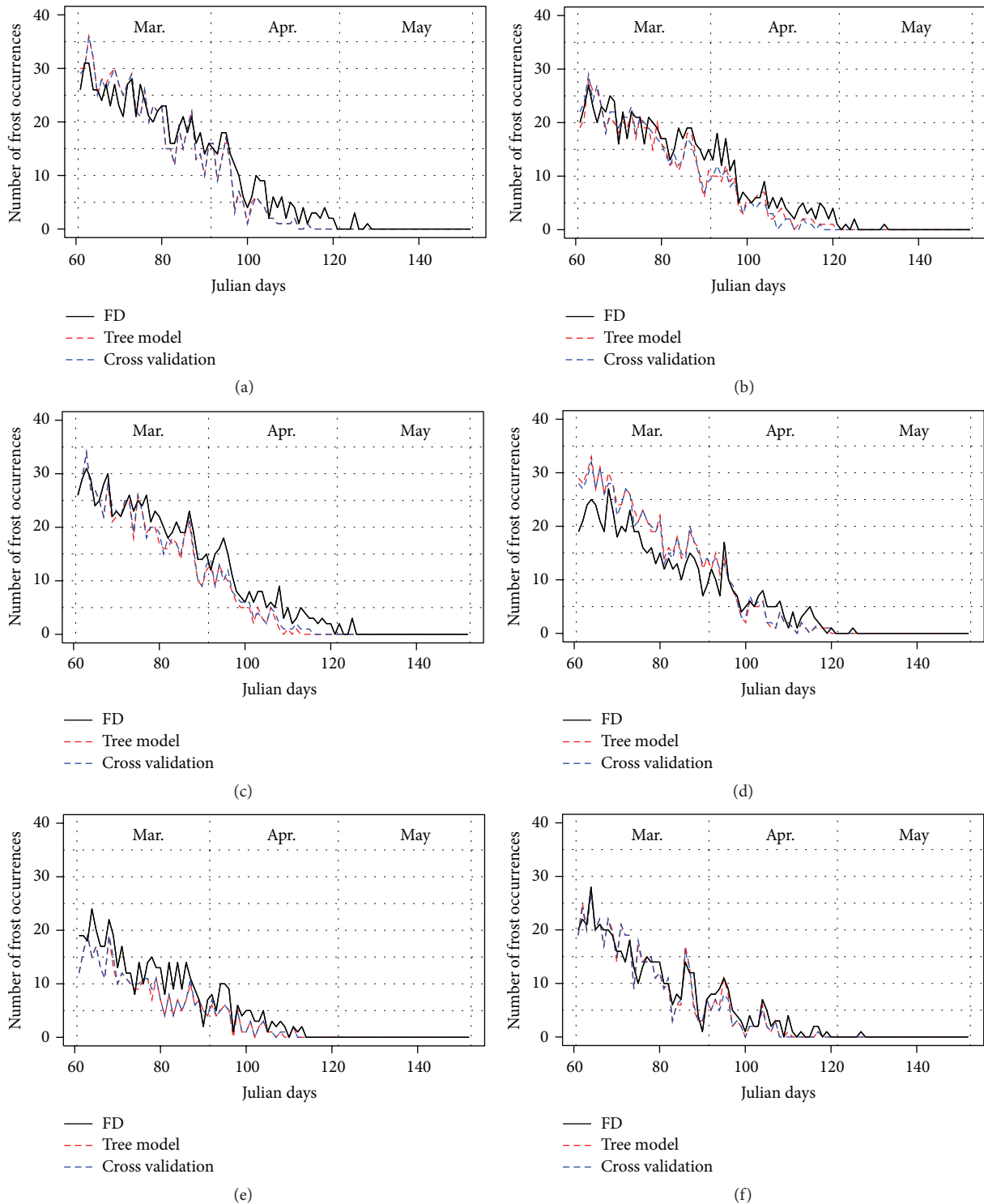


FIGURE 4: Observed and predicted frost in spring using the decision tree models: (a) Chuncheon (101), (b) Suwon (119), (c) Seosan (129), (d) Cheongju (131), (e) Gwangju (156), and (f) Jinju (192).

to prevent frost damages to agricultural crops by providing timely interventions for frost damages to agricultural farms and facilities.

For the forecast of frost events, predicted meteorological variables should be used. The predictability of frost

occurrences might be different according to forecast lead times. However, this study does not address this question. It is recommended that this question be addressed in a further study. In addition, since the frost events are very local and are affected by local topological characteristics, it is



recommended that the models should be improved to reflect these local topological characteristics.

## Competing Interests

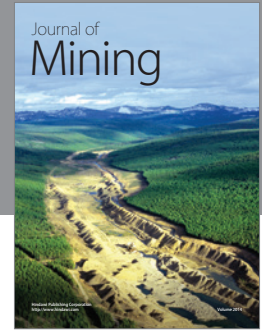
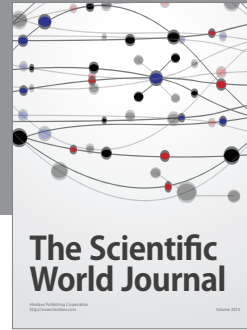
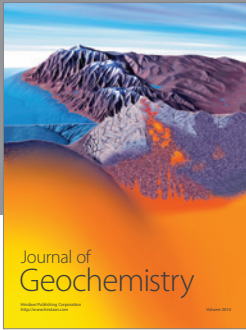
The authors declare that they have no competing interests.

## Acknowledgments

This research was supported by the APEC Climate Center. Also, the authors would like to acknowledge the Korea Meteorological Administration, which is responsible for the meteorological dataset used for this research.

## References

- [1] R. L. Snyder and J. P. de Melo-Abreu, *Frost Protection: Fundamentals, Practice and Economics*, Food and Agriculture Organization of the United Nations, Rome, Italy, 2005.
- [2] J. Rodrigo, "Spring frosts in deciduous fruit trees—morphological damage and flower hardiness," *Scientia Horticulturae*, vol. 85, no. 3, pp. 155–173, 2000.
- [3] R. F. Chevalier, G. Hoogenboom, R. W. McClendon, and J. O. Paz, "A web-based fuzzy expert system for frost warnings in horticultural crops," *Environmental Modelling and Software*, vol. 35, pp. 84–91, 2012.
- [4] Y.-A. Kwon, H.-S. Lee, W.-T. Kwon, and K.-O. Boo, "The weather characteristics of frost occurrence days for protecting crops against frost damage," *Journal of the Korean Geographical Society*, vol. 43, no. 6, pp. 824–842, 2008 (Korean).
- [5] K.-H. Hwang, J.-T. Lee, J.-I. Yun, S.-O. Hur, and K.-M. Shim, "Characteristics of nocturnal cooling at a pear orchard in frost-prone area," *Korean Journal of Agricultural and Forest Meteorology*, vol. 3, no. 4, pp. 206–214, 2001 (Korean).
- [6] C. Floor, "Estimating grass minimum temperature and probability of ground frost at Eelde (Netherlands)," *Meteorological Magazine*, vol. 11, no. 1409, pp. 261–265, 1989.
- [7] K. B. Perry, "Frost/freeze protection for horticultural crops," Leaflet no. 705-A, North Carolina Cooperative Extension Service, 1994.
- [8] E. S. Takle, "Bridge and roadway frost: occurrence and prediction by use of an expert system," *Journal of Applied Meteorology*, vol. 29, pp. e727–e734, 1990.
- [9] P. Battilani, A. Pietri, C. Barbano, A. Scandolara, T. Bertuzzi, and A. Marocco, "Logistic regression modeling of cropping systems to predict fumonisin contamination in maize," *Journal of Agricultural and Food Chemistry*, vol. 56, no. 21, pp. 10433–10438, 2008.
- [10] L. Paton, M. C. Troffaes, N. Boatman, M. Hussein, and A. Hart, "Multinomial logistic regression on Markov chains for crop rotation modelling," in *Information Processing and Management of Uncertainty in Knowledge-Based Systems: 15th International Conference, IPMU 2014, Montpellier, France, July 15–19, 2014, Proceedings, Part III*, vol. 444 of *Communications in Computer and Information Science*, pp. 476–485, Springer, Berlin, Germany, 2014.
- [11] A. Geetha and G. M. Nasira, "Data mining for meteorological applications: decision trees for modeling rainfall prediction," in *Proceedings of the 5th IEEE International Conference on Computational Intelligence and Computing Research (ICIC '14)*, pp. 1–4, IEEE, Coimbatore, India, December 2014.
- [12] W. J. Long, J. L. Griffith, H. P. Selker, and R. B. D'Agostino, "A comparison of logistic regression to decision-tree induction in a medical domain," *Computers and Biomedical Research*, vol. 26, no. 1, pp. 74–97, 1993.
- [13] D. W. Hosmer and S. Lemeshow, *Applied Logistic Regression*, John Wiley & Sons, New York, NY, USA, 1989.



**Hindawi**

Submit your manuscripts at  
<http://www.hindawi.com>

