OPEN ACCESS Freely available online
Journal of Forest Research: Open Access
Research Article

Determination of Climate Predictor Variables Effecting on Annual Cone Harvest and Seed Yield of Korean pine (*Pinus koraiensis Siebold & Zucc.*) Seed Orchards

Kim YY[1*], Ku JJ[1], Kim JH[1], Lim HI[1], Han J[2]

[1]*Division of Forest Bioinformation, National Institute of Forest Science, 39, Onjeong-ro, Gwonseon-gu, Suwon-si, Gyeonggi-do, Republic of Korea;* [2]*Department of Seed and Seedling Management, National Forest Seed and Variety Center, 72, Suhoeri-ro, Suanbo-myeon, Chungju-si, Chungcheongbuk-do, Republic of Korea*

ABSTRACT

Background: Korean pine (*Pinus koraiensis* Siebold & Zucc.) is a native pine species of Korea. The relationship between climate conditions and seed production of Korean pine has not been clearly revealed yet compared to other pines. This study was conducted to identify climate variables that are significantly related to cone and seed production in three seed orchards of Korean pine.

Methods: Regression model for cone harvest and seed yield was built using climate elements such as monthly mean temperature, monthly total precipitation, and monthly hours of sunshine as predictor variables.

Results and Discussion: The seed yield model had higher predictive precision than the cone harvest. For the seed yield model, seven climate variables were associated with seven major phenological periods in the three-year reproductive cycle of the pines: long-shoot bud (LSB) bursting, LSB development, pollen and cone bud dormancy (in two years before the seeding year), flowering, pollination, seed cone dormancy (in one year before the seeding year) and cone and seed maturation (in the seeding year). Particularly, precipitation during LSB bursting two years before the seeding year was a major climate variable limiting seed yield the most. The remaining variables associated with other phenological periods were the minor ones significantly affecting the seed yield.

Conclusion: The model appears to be meaningful clearly showing which climate variables are associated with the seed production of Korean pine and also to what extent they affect it. Further studies on a more advanced predictive model are needed, based on this study.

Keywords: Korean pine; Seed production; Seed orchards; Predictive model; Climate; Phenology; Reproductive cycle

INTRODUCTION

Korean pine (*Pinus koraiensis* Siebold & Zucc.) is one of the native pines in Korea. It usually grows in high-altitude natural forests over 1,000 meters, but also grows relatively well in plantation forests of 300 to 800 meters. The total forest area of the species is 230,282 ha, which is 5.3% of the total forest area of Korea [1]. Clonal seed orchards of the species had been established in three regions across the country, with a total area of 90 ha. Over the past 10 years (2009-2018), 11.4 tons of seeds were yielded annually from the seed orchards and used to create 1,024 ha of afforested forests per year [1].

For successful afforestation, a sufficient quantity of seeds and seedlings must be prepared in advance. However, it is not easy to collect and supply the required amount of seeds at the right time in line with the actual process of afforestation, because the seed yield is not constant across years due to environmental influences. Hence, predicting the optimal timing and environmental factors for maximizing seed yield has long been a major concern [2]. Many studies have been conducted in conifer species, such as examining the phenology and reproductive cycle [3-6], the effect of tree age, growth, density [7-9], and modeling for temporal variation of the seed yield [10-14].

Most pines are known to have a three-year reproductive cycle. The cycle is a sequential process of initiation of reproductive bud primordia (two years before the seeding year), flowering and pollination (one year before), and fertilization and seed

Correspondence to: Kim Yong-Yul, Division of Forest Bioinformation, National Institute of Forest Science, 39, Onjeong-ro, Gwonseon-gu, Suwon-si, Gyeonggi-do, Republic of Korea, Tel: +82312901150; E-mail: yykim@korea.kr

Received: September 17, 2020, **Accepted:** October 05, 2020, **Published:** October 12, 2020

Citation: Kim YY, Ku JJ, Kim JH, Lim HI, Han J (2020) Determination of Climate Predictor Variables Effecting on Annual Cone Harvest and Seed Yield of Korean pine (*Pinus koraiensis Siebold & Zucc.*) Seed Orchards. Fores Res. 9:239. doi: 10. 35248/2168-9776.20.9.239

development (the seeding year) [15]. Many researchers have examined which climate factors in the reproductive cycle have a significant effect on flowering and the consequent seed production of western white pine (*Pinus monticola*) [16,17], Norway spruce (*Picea abies*), Scots pine (*Pinus sylvestris*) [2,18], and white spruce (*Picea glauca*) [10]. The common result of these studies is that pines flower abundantly in weather conditions where the first summer (one year before the flowering year) is cool and the second (the flowering year) is warm; further, the most significant climate predictors are temperature in May (positive correlation) and summer rainfall (negative correlation) [2]. With respect to the Korean pine, only three studies have reported climate factors related to the number and weight of cones per tree [19-21]. The results of the studies differ from each other, so assumably, the significant common climate factors have not yet been identified. Further relevant studies are needed to confirm this.

The aim of this study was to develop a predictive model for the cone harvest and seed yield of Korean pine seed orchards using weather data from 12 years (2008-2019). Data included monthly mean temperature, monthly total precipitation, and monthly hours of sunshine. Subsequently, sets of climate predictor variables that significantly account for annual variation in cone harvest and seed yield were selected based on a stepwise regression analysis. Following this, the selected variables were examined whether they were associated with phenological reproductive processes in the three-year cycle of pines, so to determine whether they could be the predictive climate factor for cone and seed production of Korean pine seed orchards.

MATERIALS AND METHODS

Seed orchards description

The seed orchards of Korean pine are located in three regions: Chuncheon-city (37° 52.5′ N, 127° 36.5′ E, 439 m in elevation) and Gangneung-city (36° 35′ N, 128° 49′ E, 889 m) in Gangwon Province and Chungju-City (36° 52′ N, 127° 59′ E, 444 m) in Chungbuk province. They are all clonal seed orchards, which comprise grafts of selected genotypes. The total area spans 90 ha; its details are given in Table 1.

Data collection and preparation

The data for cone harvest and seed yield over the past 12 years (2008-2019) were obtained from the National Forest Seed Variety Center. They are the annual total quantity (kg) of the harvested cones and purified sound seeds in each seed orchard. However, the distributions of the data highly were skewed, and had serious departures from normality in the Shapiro-Wilk test (W=0.76, p<0.0001) [22]. The raw data were transformed to natural logarithms to make them much less skewed and more symmetric [23]. The log-transformed data were used as response variables in a subsequent regression analysis.

Fourteen years (2006-2019) of meteorological data for three climate factors, namely, monthly mean temperature (MMTp), monthly total precipitation (MTPr), and monthly hours of sunshine (MHSs), were obtained from website of the Korea Meteorological Administration (https://data.kma.go.kr/cmmn/main.do). They are the recorded data from the weather stations closest to each of the three seed orchards. A total of 102 predictor variables were created and used for a subsequent regression analysis. They included 30 climate variables for 10 months (from January to October) of a given seeding year (t0), 36 for 12 months (from January to December) of one year before the given seeding year (t1), and 36 for 12 months (from January to December) of two years before the given seeding year (t2). The names of the 102 variables were written in the following manner: the year (t0, t1, t2), the climate factor (MMTp, MTPr, MHSs), and the number of months (1 to 12) were arrayed sequentially in the name. For example, the name "t0_MMTp_5" denotes the climate predictor variable for the "monthly mean temperature" of "May" in the seeding year "t0".

A new data set was prepared and used for a subsequent regression analysis, in which the response variables (the log-transformed cone and seed production) in a given year in a given seed orchard corresponded to the 102 predictor variables (climate variables) in the same year in the same seed orchard.

Data analysis

All statistical analyses were performed using SAS JMP® Pro (https://www.jmp.com). Distributions for the time series of the cone harvest, seed yield, log-transformed cone harvest and log-transformed seed yield were analyzed. Additionally, time series plots were prepared and analyzed to determine whether represent a trend and seasonality. Pearson's simple and partial correlation coefficients between the log-transformed data and 102 climate variables were calculated [22,24].

The following model was used for the stepwise regressing analysis:

$$LN(Y_i) = b_0 + \sum_{j=1}^{102} b_j X_{ij} + \varepsilon_{i,}$$

where Y_i, $i=1,\ldots,36$, denotes the *i*-th observation of the cone harvest or seed yield, b_0 is an intercept, X_{ij} is the respective value of the *j*-th climate predictor variable X_j, and b_j, $j=1,\ldots,102$, denotes the linear coefficient. The error term ε_i reflects factors that are not accounted for in the model.

A stepwise regression method was used to select candidate

Table 1: Details of three seed orchards of Korean pine.

Name of seed orchard	Region	Location	Mean elevation (m)	Establishment periods (year)	Number of trees	Area (ha)	Trees/ha
CJ	Chungju-City, Chungbuk Province	36° 52′ N, 127° 59′ E	444	1970-1974	3,410	20.5	166.3
CC	Chuncheon-City, Gangwon Province	37° 53′ N, 127° 37′ E	439	1978-1995	6,695	38.0	176.2
GN	Gangneung-City, Gangwon Province	36° 35′ N, 128° 49′ E	889	1970-1972	1,887	31.5	59.9
Total	-	-	-	1970-1995	11,992	90.0	133.2

regressions of the best climate variables for the log-transformed cone harvest and seed yield, in which a threshold value of 0.05 was used to specify whether a given variable can be entered into the model or whether the variable can remain in the model [24]. The plot for the residuals using predicted values from the selected regression was examined to test whether the residuals were random. The Durbin-Watson test was performed to examine autocorrelation between consecutive residuals. The variance inflation factor (VIF) for each predictor variable was checked to assess multicollinearity of the regression [24,25].

Examination of associations between predictor variables and phenological reproductive processes

Associations between the predictor variables of the selected regression models and phenological processes during the three-year reproductive cycle of the pines were examined as follows; a time-series table of phenological processes during the three-year reproductive cycle of western white pine was prepared by reference to the table presented in the paper of Eis [17]. Then, the predictor variables were assigned to the corresponding periods in the three-year reproductive cycle to examine which variables were associated with the phenological processes of which periods, and also the associations were compared with those of relevant studies in western white pine [16,17] and Korean pine [19].

RESULTS

Annual cone harvest and seed yield

The annual mean of cone harvest, seed yield, log-transformed cone harvest (LN (cone)) and seed yield (LN (seed)) between 2008-2019 was 25,803 kg, 3.999 kg, 9.7, and 8.0, respectively (Table 2). It appeared that Chungju and Gangneung seed orchards annually harvested more cone than Chuncheon, but had a lower seed yield. However, there was no significant difference among seed orchards in the annual mean of cone harvest (F=0.6387, p=0.5348), seed yield (1.3315, 0.2779), LN (cone) (0.0423, 0.9586), and LN (seed) (2.3553, 0.1106).

The 12-year time series plot for the mean value of the cone harvest, seed yield, LN (cone), and LN (seed) are shown in Figure 1. The values did not show a distinct increasing or decreasing trend. Moreover, the results from the time-series regression analysis showed that all slope of the fitting regression line were close to zero. This indicates that the time series had no trend component. Conversely, at first glance, the time series plot seemed likely to have a seasonal cycle with a three-year interval between peaks. However, the value of the autocorrelation function of the time series was less than 0.33, indicating that there is no seasonality. Therefore, no further analysis was conducted in this study, because 12 years of time series data on seed production were found to have no significant trend or seasonal components.

Climate variable correlations

The pair-wise correlation matrix between 17 variables (2 response variables and 15 climate predictor variables) is present in Table 3. In the simple correlation, 10 climate variables were significantly related to the LN (cone) and LN (seed), in which five variables were significantly correlated with the LN (cone), and seven with the LN (seed), and two significantly correlated with both. In the partial correlation, however, five variables were significantly related to LN (cone) and LN (seed), in which two were significantly related to LN (cone), and three to LN (seed). The two variables, t2_MTPr_2 and t2_MTPr_10, were significantly related to the LN (cone) and LN (seed) in both correlations. Each of the 15 variables had the same sign of correlation coefficient to both the LN (cone) and LN (seed);

Table 2: Descriptive statistics for cone harvest and seed yield, and their logarithms.

Source	Overall mean					Regional mean		
	Mean	Std.Dv	Std.Err	CV	p < W	GN	CC	CJ
Cone	25,803.4	26,968.8	4,625.1	104.5	< 0.0001	27,825.2	18,946.5	31,605.7
LN (Cone)	9.7	0.9	0.2	9.7	0.8542	9.8	9.7	9.7
Seed	3,998.8	2,720.6	453.4	68.0	0.0088	3,529.6	5,033.9	3,432.8
LN (Seed)	8.0	0.8	0.1	10.3	0.0802	7.8	8.4	7.8

GN: Gangneung; CC: Chuncheon; CJ: Chungju; Std.Dv: Standard Deviation; Std.Err: Standard Error; CV: Coefficient of Variation; p < W: p-value for Shapiro-Wilk W test.

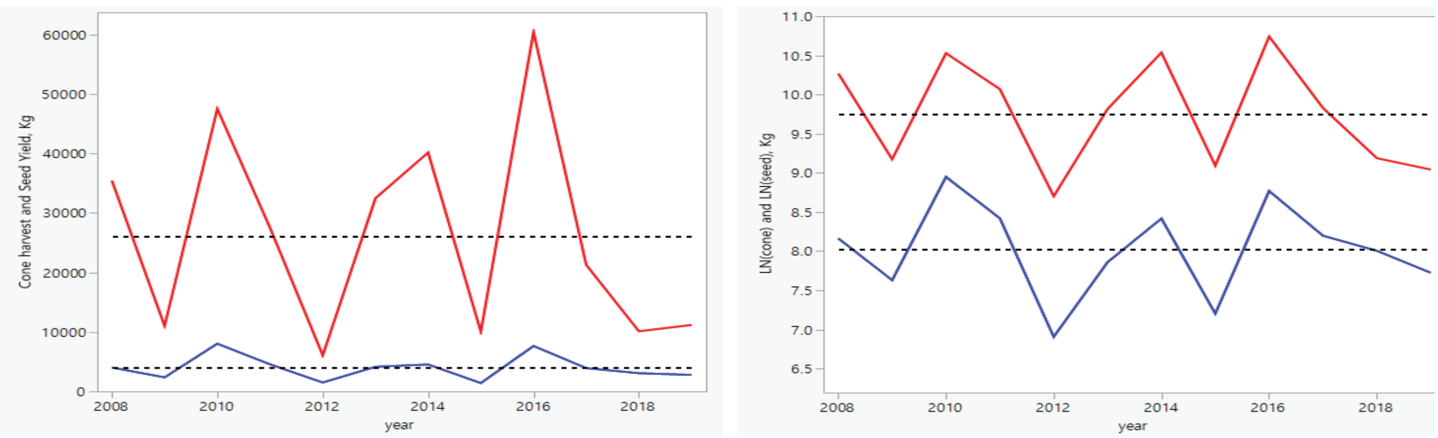

Figure 1: Time series plot for mean of cone harvest, seed yield (left), LN (cone) and LN (seed) (right). Red line: Cone harvest and LN (cone); Blue line: Seed yield and LN (seed); Black dotted line: Mean value.

Table 3: Pearson's simple and partial correlations between 17 variables.

			Year t2								Year t1					Year t0		
Variables			LN (seed)	MTPr_2	MMTp_4	MHSs_6	MTPr_7	MTPr_9	MTPr_10	MHSs_10	MTPr_5	MTPr_6	MHSs_6	MTPr_10	MTPr_11	MTPr_2	MTPr_10	MHSs_10
		LN (cone)	0.4039 (0.0694)	-0.2946 (0.1949)	0.1689 (0.4642)	-0.2237 (0.3297)	-0.0071 (0.9757)	0.0812 (0.7264)	**0.4877 (0.0249)**	**0.5454 (0.0106)**	-0.0530 (0.8194)	0.1517 (0.5114)	0.0698 (0.7637)	0.0592 (0.7989)	0.0689 (0.7666)	0.1533 (0.5071)	-0.3093 (0.1725)	0.3957 (0.0758)
	LN (seed)	**0.8156 (0.0000)**	LN (seed)	**-0.6349 (0.0020)**	0.1857 (0.4203)	0.0525 (0.8211)	**0.4599 (0.0359)**	0.1088 (0.6388)	0.1440 (0.5334)	-0.1113 (0.6310)	**-0.4655 (0.0335)**	0.1007 (0.6640)	0.4100 (0.0649)	-0.1514 (0.5123)	0.3901 (0.0804)	0.3541 (0.1153)	-0.3877 (0.0825)	-0.0665 (0.7745)
Year t2	MTPr_2	**-0.4702 (0.0050)**	**-0.6031 (0.0001)**	t2_ MTPr_2	-0.3079 (0.0677)	-0.0157 (0.9462)	0.4142 (0.0620)	0.0698 (0.7638)	**0.4803 (0.0275)**	0.2094 (0.3623)	**-0.6171 (0.0029)**	0.2845 (0.2113)	0.3847 (0.0850)	-0.1711 (0.4584)	**0.5236 (0.0148)**	0.3647 (0.1041)	**-0.5982 (0.0042)**	0.3081 (0.1742)
	MMTp_4	0.1871 (0.2894)	**0.4210 (0.0106)**	0.1943 (0.3988)	t2_ MMTp_4	0.3450 (0.1256)	-0.2354 (0.3044)	**-0.4448 (0.0433)**	-0.4136 (0.0624)	-0.3269 (0.1481)	-0.0030 (0.9897)	-0.0058 (0.9801)	-0.1067 (0.6453)	-0.3029 (0.1820)	0.0556 (0.8109)	-0.1899 (0.4097)	0.2454 (0.2836)	-0.3917 (0.0790)
	MHSs_6	**-0.3854 (0.0244)**	-0.2484 (0.1440)	-0.0666 (0.6997)	0.2659 (0.1170)	t2_ MHSs_6	0.1258 (0.5869)	-0.0718 (0.6771)	-0.1556 (0.3648)	0.1193 (0.6067)	0.1190 (0.6074)	0.1345 (0.5611)	**0.4927 (0.0233)**	**0.6273 (0.0023)**	**-0.4360 (0.0482)**	-0.1894 (0.4110)	0.1068 (0.6450)	0.1352 (0.5590)
	MTPr_7	0.0368 (0.8362)	0.1713 (0.3178)	-0.0144 (0.9337)	-0.0025 (0.9884)	-0.0683) (0.6921)	t2_ MTPr_7	**-0.3602) (0.0309)**	0.1342 (0.4351)	-0.2785 (0.1000)	0.1027 (0.5513)	-0.3014 (0.0740)	-0.1250 (0.4676)	-0.1135 (0.5099)	-0.3269 (0.0516)	-0.0203 (0.9064)	0.2173 (0.3440)	0.0225 (0.8962)
	MTPr_9	-0.3037 (0.0808)	**-0.3987 (0.0160)**	0.2378 (0.1625)	**-0.4650 (0.0043)**	0.3021 (0.1831)	**-0.4377 (0.0472)**	t2_ MTPr_9	-0.2682 (0.2399)	-0.3498 (0.1201)	-0.1460 (0.5276)	0.2895 (0.2031)	**-0.4859 (0.0255)**	**-0.4944 (0.0227)**	0.1572 (0.4963)	-0.2582 (0.2585)	0.3229 (0.1534)	0.0435 (0.8514)
	MTPr_10	**0.3437 (0.0466)**	0.2589 (0.1273)	0.1931 (0.2592)	-0.0084 (0.9610)	0.1151 (0.6193)	-0.1145 (0.6211)	-0.2036 (0.2337)	t2_ MTPr_10	**-0.4396 (0.0461)**	0.0824 (0.7224)	-0.2815 (0.2163)	-0.1028 (0.6575)	-0.1768 (0.4432)	0.0347 (0.8812)	**-0.5151 (0.0169)**	**0.5048 (0.0196)**	-0.3318 (0.1417)
	MHSs_10	**0.3661 (0.0332)**	0.1129 (0.5120)	-0.1087 (0.5281)	-0.0382 (0.8250)	-0.0497 (0.7734)	-0.2885 (0.2047)	-0.2592 (0.1268)	-0.0038 (0.9827)	t2_ MHSs_10	-0.0419 (0.8570)	-0.1604 (0.4874)	-0.1352 (0.5591)	-0.0306 (0.8951)	0.0284 (0.9027)	-0.2278 (0.3207)	0.3805 (0.0889)	-0.2052 (0.3723)
Year t1	MTPr_5	-0.2381 (0.1751)	-0.1839 (0.2831)	-0.2663 (0.1164)	0.0583 (0.7355)	0.2906 (0.0855)	0.3453 (0.1253)	0.0390 (0.8212)	-0.2643 (0.1193)	-0.2332 (0.1710)	t1_ MTPr_5	**0.3430 (0.0406)**	-0.0072 (0.9667)	-0.1804 (0.4340)	0.3484 (0.1217)	-0.2919 (0.0840)	-0.2644 (0.2467)	-0.0034 (0.9845)
	MTPr_6	-0.3059 (0.0785)	**-0.3714 (0.0257)**	0.2801 (0.0980)	-0.2023 (0.2368)	0.1129 (0.5121)	-0.1933 (0.4013)	**0.6505 (0.0000)**	-0.2456 (0.1488)	-0.2623 (0.1222)	0.4069 (0.0672)	t1_ MTPr_6	-0.0481 (0.8360)	-0.0701 (0.7628)	0.1844 (0.4235)	-0.2370 (0.3010)	0.0240 (0.9178)	0.0524 (0.8216)
	MHSs_6	0.1921 (0.2764)	**0.3506 (0.0360)**	-0.1442 (0.4015)	**0.5219 (0.0011)**	**0.4488 (0.0060)**	**-0.4429 (0.0444)**	**-0.4550 (0.0053)**	0.2802 (0.0979)	0.1507 (0.3802)	0.1982 (0.3892)	-0.2148 (0.2083)	t1_ MHSs_6	-0.2774 (0.2234)	-0.0208 (0.9288)	-0.1552 (0.5017)	**0.4418 (0.0449)**	-0.0853 (0.7133)
	MTPr_10	-0.1308 (0.4609)	-0.1839 (0.2829)	-0.0470 (0.7854)	-0.0316 (0.8550)	**0.4386 (0.0075)**	-0.1542 (0.5045)	-0.2759 (0.1034)	-0.2118 (0.2150)	0.2560 (0.1317)	0.0777 (0.6525)	-0.0846 (0.6238)	0.0611 (0.7234)	t1_ MTPr_10	**0.5001 (0.0210)**	-0.0734 (0.6704)	-0.1122 (0.6282)	0.1020 (0.5537)
	MTPr_11	0.2717 (0.1202)	0.1791 (0.2958)	0.2502 (0.1410)	-0.0193 (0.9109)	**-0.3607 (0.0307)**	-0.2466 (0.2813)	0.1318 (0.4435)	0.2778 (0.1009)	0.1630 (0.3421)	-0.1373 (0.4245)	0.2388 (0.1608)	0.0660 (0.7023)	0.0223 (0.8972)	t1_ MTPr_11	0.1830 (0.2855)	0.3528 (0.1167)	-0.1709 (0.3191)
Year t0	MTPr_2	**0.4073 (0.0168)**	**0.4567 (0.0051)**	-0.2034 (0.2342)	0.1311 (0.4461)	**-0.3605 (0.0308)**	-0.2347 (0.3059)	-0.2873 (0.0893)	-0.0420 (0.8077)	0.2201 (0.1972)	0.1018 (0.6607)	**-0.3869 (0.0198)**	0.1286 (0.4549)	0.0040 (0.9864)	-0.0798 (0.7309)	t0_ MTPr_2	**0.5220 (0.0152)**	-0.0659 (0.7765)
	MTPr_10	-0.0421 (0.8133)	-0.0313 (0.8564)	-0.1062 (0.5375)	0.1098 (0.5240)	0.3014 (0.0740)	-0.3036 (0.0718)	0.0639 (0.7112)	0.1613 (0.3472)	0.2436 (0.1523)	-0.0711 (0.6801)	-0.0592 (0.7316)	**0.4470 (0.0063)**	0.0262 (0.8795)	0.1275 (0.4587)	0.2076 (0.2244)	t0_ MTPr_10	0.0765 (0.6576)
	MHSs_10	-0.2659 (0.1285)	**-0.4668 (0.0041)**	0.3244 (0.0536)	**-0.5348 (0.0008)**	0.1239 (0.4715)	0.0936 (0.6864)	**0.4073 (0.0137)**	-0.2031 (0.2347)	-0.0547 (0.7514)	0.0186 (0.9361)	0.2937 (0.0821)	-0.2966 (0.0790)	0.0849 (0.7145)	-0.2327 (0.3100)	-0.1691 (0.3240)	0.3249 (0.1507)	t0_ MHSs_10

The values of the simple correlation coefficient are listed in the diagonal below, while those of the partial correlation in the above diagonal.
Values in parentheses are p-values. MTPr: Monthly Total Precipitation; MHSs: Monthly Sunshine Hours; MMTp: Monthly Mean Temperature.
The bold numbers are the coefficient and probability of the significantly correlated pairs at p = 0.05.

that is the same direction of the relationship, even though the coefficient value was not the same. Among the 15 variables, one was MMTp, 10 were MTPr, and the remaining four were MHSs.

Regression model for LN (Cone) and LN (Seed)

Two models for LN (cone) and LN (seed) were selected from the stepwise regression analysis as follows:

LN (cone)=10.0375 + (-0.0212) × (t2_MTPr_2) + (-0.0086) × (t2_MHSs_6) + (0.0077) × (t2_MTPr_10) + (0.0089) × (t2_MHSs_10),

LN (seed)=7.3969 + (-0.0278) × (t2_MTPr_2) + (-0.0009) × (t2_MTPr_7) + (-0.0059) × (t1_MTPr_5) + (0.0083) × (t1_MHSs_6) + (-0.0036) × (t1_MTPr_10) + (0.0091) × (t1_MTPr_11) + (-0.0049) × (t0_MTPr_10).

The model for LN (cone) with four climate predictor variables (t2_MTPr_2, t2_MHSs_6, t2_MTPr_10, and t2_MHSs_10) was highly significant at a level of p=0.05 for the F-ratio (F=12.2, $p \leq 0.0001$), accounting for 62.6% of the total variation in the LN (cone) (R^2=0.6263, RMSE=0.6165) (Table 4). The plot for the observed values by the predicted values from the model parameter estimates showed that the model fit the observed values relatively well, even though there were some deviations between the values (Figure 2). The plot for the residuals by the predicted values showed that the residuals were randomly scattered around zero with no obvious pattern. The Shapiro-Wilk W test for normality of the model was not significant at a level of p=0.05 (W=0.9756, p=0.6310). All the VIF values for four variables were much smaller than 5 (Table 5), indicating that the model for LN (cone) has no significant multicollinearity with four predictor variables. The two variables (t2_MTPr_2 and t2_MHSs_6) were negative for the LN (cone), while the other two (t2_MTPr_10, and t2_MHSs_10) were positive (Table 5). Based on the partial regression coefficient and the absolute value of the standardized coefficient, t2_MTPr_2 was the most important predictor variable (0.2211, 0.5397), and the relative importance of the remaining three variables was similar (0.0877-0.1964, 0.2979-0.3862).

In the model for the LN (seed), seven climate predictor variables (t2_MTPr_2, t2_MTPr_7, t1_MTPr_5, t1_MHSs_6, t1_MTPr_10, t1_MTPr_11 and t0_MTPr_10) were included, and the model was highly significant at a level of p=0.05 (F=21.4, $p \leq 0.0001$), accounting for 80.3% of the total variation in LN (seed) (R^2=0.8425, RMSE=0.3673) (Table 4). The plot for the observed values by the predicted values from the model parameter estimates showed that the model fitted well with the observed values, even though there were a few deviations between the observed and the predicted values (Figure 3). The plot for the residual by the predicted value showed that the residuals of the model were randomly scattered around zero with no obvious pattern (Figure 3). The Shapiro-Wilk W-test for normality of the model was not significant at a level of p=0.05 (W=0.9641, p=0.2865). The four climate predictor variables (t2_MTPr_2, t1_MTPr_5, t1_MTPr_10, and t0_MTPr_10) were negatively related to LN (seed), while the other three variables (t2_MTPr_7, t1_MHSs_6 and t1_MHPr_11) were positive (Table 6). All VIF values for the seven variables were much smaller than 5, indicating that the model for LN (seed) had no significant multicollinearity with seven variables (Table 6). Based on the partial regression coefficient and the absolute value of the standardized coefficient, t2_MTPr_2 was the most important predictor variable in the model (0.3637, 0.7884), and the relative importance of the remaining six variables was similar (0.0306-0.0583, 0.1860-0.4326).

The regression results can be summarized as follows: first, approximately 63% of the total variation in LN (cone) can be explained by the fitting model consisting of four climate predictor variables of two years before the seeding year. A higher cone harvest would be expected if there was less precipitation in February, longer sunshine in June and October, and more precipitation in October. Second, approximately 84% of the total variation in LN (seed) can be accounted for by the fitting model composed of seven climate predictor variables (two of two years before the seeding year, four of one year before the seeding year, one of the seeding year). A higher seed yield would be expected if there was less precipitation in February of the two years before, May and October of one year before, and October of the seeding year. Additionally, a higher seed yield would be expected if more precipitation occurred in July of two years before and November of one year before, and if longer sunshine occurred in June of one year before. Third, February precipitation for two years before the seeding year is the climate predictor variable that most affects both cone harvest and seed yield. It would be expected that the lower the February precipitation, the higher the cone harvest and seed yield.

Associations of predictor variables with phenological reproductive processes

Ten climate predictor variables of the regression models for the LN (cone) and LN (seed) were assigned to five phenological phases in the three-year reproductive cycle (Table 7). All four variables associated with LN (cone) were assigned only to the first phenological phase.

Table 4: Goodness-of-fit of regression models for logarithms of cone harvest and seed yield.

Source	R^2	Adj-R^2	RMSE	Mean	AICc	BIC	***F*** statistics		Durbin-Watson test	
							F-ratio	p > ***F***	DW	p < DW
LN (cone)	0.6263	0.5748	0.6165	9.7	73.3	79.3	12.2	<.0001	1.89	0.3361
LN (seed)	0.8425	0.8031	0.3673	8.0	45.9	53.2	21.4	<.0001	1.88	0.3679

R^2: Coefficient of determination for the model; Adj-R^2: Adjusted R^2; RMSE: Root Mean Square Error; AICc: Akaike Information Criterion; BIC: Bayesian Information Criterion; p < DW: p-value for the Durbin Watson test.

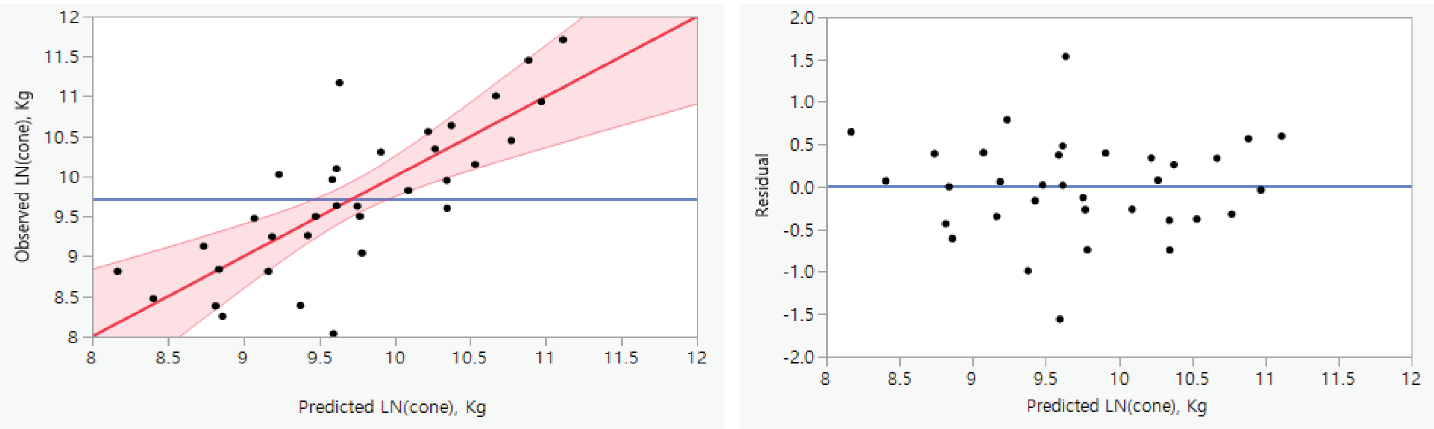

Figure 2: Plot for observed values of LN (cone) (left) and for residuals (right) by predicted values. Blue line denotes mean value.

Table 5: Parameter estimates of model for cone harvest (LN (cone)).

Source	Estimates	Standard Error	95% lower limit	95% upper limit	Partial R^2	Standardized coefficient	T-test		VIF
							t-ratio	Prob > |t|	
Intercept	10.0375	0.9594	8.0752	11.9997	-	-	10.46	<.0001	-
t2_MTPr_2	-0.0212	0.0046	-0.0306	-0.0119	0.2211	-0.5397	-4.64	<.0001	1.05
t2_MHSs_6	-0.0086	0.0029	-0.0146	-0.0026	0.1211	-0.3392	-2.94	0.0065	1.04
t2_MTPr_10	0.0077	0.0023	0.0029	0.0124	0.1964	0.3862	3.29	0.0026	1.07
t2_MHSs_10	0.0089	0.0034	0.0019	0.0159	0.0877	0.2979	2.61	0.0142	1.01

VIF: Variation Inflation Factor.

Conversely, the seven variables related to LN (seed) were assigned to three phenological phases and one interphase; two of the seven variables were assigned to the first phenological phase, two variables to both the second phase and interphase, and one variable to the third phase. The assignment pattern of the climate variables for LN (cone) is similar to that of Rehfeldt et al. [16], but the pattern for LN (seed) is similar to that of Eis [17] and Joo et al. [19].

It is noteworthy in this study that many climate variables (five out of 10 variables) were assigned to the period of the first phenological phase where the processes of the LSB development and the initiation of pollen and cone bud take place. Further, in particular, the assigned variables explained a large proportion of total variation in each of the LN (cone) and LN (seed) (62.6% and 39.7%, respectively). Rehefeldt et al. [16] and Eis [17] also showed similar results, where four of six variables and eight of 16 variables were assigned to the same phase in each study. According to Owens [15], the LSB is a different type of vegetative bud of pine species compared with that of other conifers. It contains all types of buds, such as pollen and seed cone bud, short-shoot bud, and lateral branch bud, which is found on new elongated shoots in the following year. These facts indicate that the normal formation and development of LSB two years before the seeding year may be a prerequisite for determining the amount of flowering, cone harvest and seed yield. Consequently, it is thought that the six climate variables assigned to the first phenological phase of the LSB development and initiation of pollen and seed cone bud in this study may be good predictors for cone harvest and seed yield of Korean pine.

Particularly noteworthy is the February precipitation among the five climate variables allocated to the first phenological phase, which is the most influential variable that accounted for 22.1% and 36.3% of the total variation of LN (cone) and LN (seed), respectively, and is the limiting factor that negatively affected both of them the most. February in Korea is the period between late winter and early spring, so the temperature is low. Moreover, the average temperature in February in the three seed orchards over the past 12 years was -2.12°C (± 2.51). Therefore, it can be inferred that the higher the precipitation at this time, the lower the temperature, which would inhibit or delay the LSB bursting and development, and eventually lead to a reduction in flowering and consequently cone and seed production.

Five of seven variables related to LN (seed) were assigned to the following phenological periods; namely, two variables were assigned to the period of flowering and pollination in one year before the seeding year, two to the period of seed cone dormancy in the same year, and one to the period of the cone and seed maturation in the seeding year. These three groups of variables assigned to each period can be ranked in the following order based on the proportion of total variation in LN (seed) explained by them: May precipitation and Jun sunshine hours (9.1%) in one year before the seeding year, October and November precipitation (7.0%) in the same year, and October precipitation (5.8%) in the seeding year. The assignment of the five variables to the phenological periods in this study seems to be slightly similar to that of Eis [17] rather than that of Joo et al. [19], except for October precipitation in the seeding year. In addition, this study and Eis [17] showed that the variables related to the LN (seed) and cone number were not assigned to the period

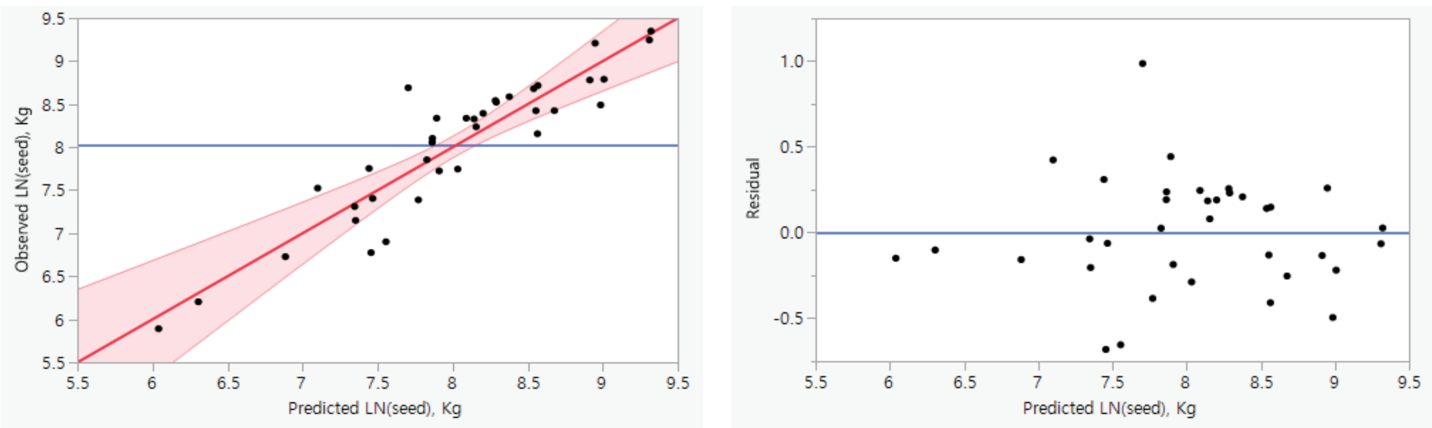

Figure 3: Plot for observed values of LN (seed) (left) and for residuals (right) by predicted values. Blue line denotes mean value.

Table 6: Parameter estimates of model for seed yield (LN (seed)).

Source	Estimates	Standard Error	95% lower limit	95% upper limit	Partial R^2	Standardized coefficient	T-test		VIF
							t-ratio	Prob > |t|	
Intercept	7.3969	0.4451	6.4851	8.3087	-	-	16.62	<.0001	-
t2_MTPr_2	-0.0278	0.0029	-0.0337	-0.0219	0.3637	-0.7884	-9.67	<.0001	1.18
t2_MTPr_7	0.0009	0.0003	0.0004	0.0015	0.0336	0.2817	3.38	0.0022	1.24
t1_MTPr_5	-0.0059	0.0013	-0.0085	-0.0033	0.0376	-0.3656	-4.64	<.0001	1.10
t1_MHSs_6	0.0083	0.0019	0.0044	0.0121	0.0533	0.3744	4.43	0.0001	1.27
t1_MTPr_10	-0.0036	0.0015	-0.0066	-0.0006	0.0306	-0.1860	-2.45	0.0208	1.02
t1_MTPr_11	0.0091	0.0017	0.0055	0.0127	0.0398	0.4326	5.24	<.0001	1.21
t0_MTPr_10	-0.0049	0.0016	-0.0081	-0.0017	0.0583	-0.2731	-3.11	0.0043	1.37

VIF: Variation Inflation Factor

Table 7: Associations of climate predictor variables with phenological processes of three-year reproductive cycle.

Phenology of western white pine (Owens, 2004)	Year and month	Korean Pine				Western white pine	
		This study		Joo et al. [19]		Rehfeldt et al. [16]	Eis [17]
		LN (cone)	LN (seed)	Cone weight	Seed weight	Cone number	Cone number
Long-shoot bud (LSB) development	t2_Feb	PR (-)①	PR (-)①	TP (+)			
	t2_Jun	SS (-)③				Dry: PR (-)	Sunny: SS (+) Warm: TP (+), Dry
	t2_Jul		PR (+)⑤			Dry: PR (-)	Sunny: SS (+) Warm, Dry
Pollen and cone bud initiation	t2_Aug					Wet: PR (+)	Warm: TP (+)
	t2_Sep					Wet: PR (+)	Warm: TP (+) Wet: PR (+)
Pollen and cone bud dormancy	t2_Oct	SS(+)④, PR(+)②					Warm: TP (+)
	t1_Jan				TP (-)		
	t1_Mar						Warm: TP (+)
Flowering	t1_Apr						Warm: Min TP (-) Max TP (+)
	t1_May		PR (-)④				Warm: TP (+), Sunny
Pollination	t1_Jun		SS (+)③		SS (+)		
Magagametophyte development	t1_Jul						Wet: PR (+)
	t1_Aug			TP (-)	TP (-)		
Seed cone dormancy	t1_Sep						Warm: TP (+)
	t1_Oct		PR (-)⑦				Warm: TP (+)
	t1_Nov		PR (+)②				Warm: TP (+)

PR: Precipitation; SS: Sunshine; TP: Mean Temperature; Min and Max TP: Minimum and Maximum Temperature. T_0: Seeding Year; T_1: One Year Before; T_2: Two Years Before.
The (+) and (-) means the “positive” and “negative” association with the cone and seed yield. The circled superscript number indicates the rank of the relative influence of the predictor variables in the regression models.

of fertilization in Jun and July, the seeding year. Conversely studies by Rehfeldt et al. [16] and Joo et al. [19] showed that dry weather conditions and July temperature were correlated with the cone number and the weight of cone and seed. The comparison between the studies indicates that there were no common climate variables assigned to the phenology period of one year before the seeding year and the seeding year. Further studies are needed to ascertain whether the difference between the studies is caused by different statistical methods used in each study or by different local weather conditions in each study site, to determine the more generalized climate variables related to seed yield of Korean pine [15-26].

DISCUSSION

Given that the value of R^2 and RMSE of the regression model and the association of predictor variables of the model with the phenological phases, the model for LN (seed) seems to be statistically more reliable in explaining the total variation of LN (seed) and also practically more applicable to determine the optimal timing of seed production rather than the model for LN (cone). However, the model for LN (seed) in this study has the weakness of fewer observations (36 data) compared to the number of predictor variables, resulting in the possibility of over-fitting of the model. In general, it has been suggested that there should be at least 10 observations per variable in a regression analysis [25], which means that at least 70 observations of seed yield would be needed in the model for LN (seed) with seven climate variables in this study. Therefore, a Least Absolute Shrinkage and Selection Operator analysis [26] was additionally conducted to reduce the number of predictor variables of the model, where the number of variables included in the shrunk new model was set to four appropriate variables for the 36 observations in this study [26]. The result of the additional analysis showed that the new model comprised four variables (t2_MTPr_2, t1_MTPr_5, t1_MHSs_6, t0_MTPr_10) with a R^2 of 0.5217 and a square root of the mean squared prediction error of 0.5643, became lower in predictive precision than that of the previous one (0.8425, 0.3239). All considered, it is difficult to determine which of the two models would be better in explaining the variation in seed yield in the seed orchards of Korean pine, and it will only be possible after additional data are obtained and analyzed.

It seems that the best way would be to obtain more observations within approximately five years and then build a new model. The number of observations should be more than ten times the number of variables included in the new model. However, within five years, it is impossible to obtain more than the minimum number of observations only from the three seed orchards of Korean pine. Instead, it is desirable to obtain data from many forests with density and age similar to that of the seed orchards in multiple regions with different weather conditions in Korea and adjacent countries, if possible. Subsequently, a new model should be re-built using recently developed modeling methods. In such case, the results of this study are expected to be used as preliminary data, namely as training or comparative dataset for future research.

CONCLUSION

In this study, we built regression models that could significantly explain large proportion of total variation in cone harvest and seed yield in three seed orchards of Korean pine over the past 12 years (2008-2019) using climate predictor variables. The results showed that the model for seed yield would be better than that of cone harvest in predicting the quantity of seed production and making mid-term plans for seed collection. The model for seed yield indicated that seed yield was significantly related to seven climate variables (temperature and sunshine hours) assigned to seven phenological periods for the LSB bursting, LSB development, pollen and cone bud dormancy (in two years before the seeding year), flowering, pollination, seed cone dormancy (in one year before the seeding year), and cone and seed maturation (in the seeding year). Particularly, precipitation during the period of the LSB bursting two years before the seeding year was a major limiting climate variable that had the greatest effect on seed yield. The remaining variables assigned to other phenological periods were the minor variables that significantly and independently affected the seed yield.

The model built in the study seems to be meaningful in that it clearly present which climate variables are associated with the seed production of Korean pine and also to what extent they affect it, despite the possibility of over-fitting due to relatively many variables compared to the number of observations. It is hoped that further studies on a more advanced predictive model will be carried out based on these results.

ACKNOWLEDGEMENT

We greatly appreciate SAS JMP Korea for allowing us to use the JMP Pro software for statistical analysis of this study. We also thank the anonymous reviewers for their useful comments on this manuscript.

REFERENCES

1. Korea Forest Service. Statistical Year Book of Forestry. 2019.
2. Pukkala T, Hokkanen T, Nikkanen T. Prediction models for the annual seed crop of Norway spruce and Scots pine in Finland. Silva Fenn. 2010;44(4):629-642.
3. Owston PW. The shoot apex in eastern white pine: its structure, seasonal development, and variation within the crown. Can J Bot. 1968;47(7):1181-1188.
4. Owens JN, Molder M. Development of long-shoot terminal buds of western white pine (*Pinus monticola*). Can J Bot. 1977;55:1308-1321.
5. Owens JN, Molder M. Seed-cone differentiation and sexual reproduction in western white pine (*Pinus monticola*). Can J Bot. 1977;55(20):2574-2590.
6. Philipson JJ. Predicting cone crop potential in conifers by assessment of developing cone buds and cones. Forestry. 1997;70(1):87-96.
7. Ordonez JL, Retana J, Espelta JM. Effects of tree size, crown damage and tree location on post-fire survival and cone production of *Pinus nigra* trees. For Ecol Manag. 2005;206(1-3):109-117.
8. Ayari A, Khouja ML. Ecophysiological variables influencing Aleppo pine seed and cone production: a review. Tree Physiol. 2014;34(4):426-437.
9. Nguyen TT, Tai DT, Zhang P, Razaq M, Shen HL. Effect of thinning intensity of tree growth and temporal variation of seed and cone production in a *Pinus koraiensis* plantation. J For Res. 2019;30(3):835-845.
10. Krebs CJ, LaMontagne JM, Kenney AJ, Boutin S. Climatic determinants of white spruce cone crops in the boreal forest of southwestern Yukon. Bot. 2012;90(2):113-119.
11. Bisi F, Von Hardenberg J, Bertolino S, Wauters LA, Imperio S, Preatoni DG, et al. Current and future conifer seed production in the Alps: testing weather factors as cues behind masting. Eur J For Res. 2016;135(4):743-754.
12. Mencuccini M, Piussi P, Sulli AZ. Thirty years of seed production in a subalpine Norway spruce forests: Patterns of temporal and spatial variation. For Ecol Manag. 1995;76:109-125.
13. Mutke S, Gordo J, Gil L. Variability of Mediterranean stone pine cone production: yield loss as response to climate change. Agric For Meteorol. 2005;132:263-272.
14. Nussbaumer A, Waldner P, Apuhtin V, Aytar F. Impact of weather cues and resource dynamics on masting occurrence in the main forest tree species in Europe. For Ecol Manag. 2018;429:336-350.
15. Owens JN. The reproductive biology of western white pine. Forest Genetic Councils of British Columbia. Extension Note No. 2004: 04.
16. Rehfeldt GE, Stage AR, Bingham RT. Strobili development in western white pine: periodicity, prediction, and association with weather. For Sci. 1971;17:454-461.
17. Eis S. Association of western pine cone crops with weather variables. Can J For Res. 1975;6(1):6-12.
18. Nikkanen T, Ruotsalainen S. Variation in flowering abundance and its impact on the genetic diversity of the seed crop in a Norway spruce seed orchard. Silva Fenn. 2000;34(3):205-222.
19. Joo YT, Chon SK, Chung DJ. The effect of some meteorological factors on the seed characteristics in Korean white pine (*Pinus koraiensis* S. et Z.): The weight of cone and seed per cone. Korean J Agric Forest Meteorol. 1999;1:20-29.
20. Kim IH, Shin MY, Kim YC, Chon SK. Evaluation of reproductive growth in a mature stand of Korean pine under simulated climatic condition. Korean J Agric Forest Meteorol. 2001;3:185-98.
21. Shin MY, Jang YS, Han SU, Kim YC. Effects of local climatic conditions on the yearly cone production in progeny test stands of Korean white pine. Korean J Agric Forest Meteorol. 2002;4(3):141-150.
22. Rushing H, Karl A, Wisnowski J. Design and analysis of experiments by Douglas Montgomery: A Supplement for Using JMP○RR. Cary, NC: SAS Institute Inc. 2013.
23. Feng C, Wang H, Lu N, Tu XM. Log transformation: Application and interpretation in biomedical research. Statistics in Medicine. 2002;32:230-239.
24. Field A, Miles J, Field Z. Discovering Statistics Using R. Newbury Park, CA: SAGE Publications Ltd. 2012; p: 992.
25. Troutt MD. Regression, 10k rule of thumb for. Encyclopedia of Statistical Sciences. New York: Wiley, USA. 2006.
26. Klimberg R, McCullough P. Fundamentals of Predictive Analytics with JMP. (2nd edn). Cary, NC: SAS Institute Inc., USA, 2016.