Journal of the Korean Society of Grassland and Forage Science 36(3) : 223~236, 2016 ISSN 2287-5824 (Print)
http://dx.doi.org/10.5333/KGFS.2016.36.3.223 ISSN 2287-5832 (Online)

Research Article

Models for Estimating Yield of Italian Ryegrass in South Areas of Korean Peninsula and Jeju Island

Jing Lun Peng, Moon Ju Kim, Byong Wan Kim and Kyung Il Sung*
Department of Feed Science and Technology, College of Animal Life Sciences, Kangwon National University, Chuncheon, 24341, Republic of Korea

ABSTRACT

The objective of this study was to construct Italian ryegrass (IRG) dry matter yield (DMY) estimation models in South Korea based on climatic data by locations. Obviously, the climatic environment of Jeju Island has great differences with Korean Peninsula. Meanwhile, many data points were from Jeju Island in the prepared data set. Statistically significant differences in both DMY values and climatic variables were observed between south areas of Korean Peninsula and Jeju Island. Therefore, the estimation models were constructed separately for south areas of Korean Peninsula and Jeju Island separately. For south areas of Korean Peninsula, a data set with a sample size of 933 during 26 years was used. Four optimal climatic variables were selected through a stepwise approach of multiple regression analysis with DMY as the response variable. Subsequently, via general linear model, the final model including the selected four climatic variables and cultivated locations as dummy variables was constructed. The model could explain 37.7% of the variations in DMY of IRG in south areas of Korean Peninsula. For Jeju Island, a data set containing 130 data points during 17 years were used in the modeling construction via the stepwise approach of multiple regression analysis. The model constructed in this research could explain 51.0% of the variations in DMY of IRG. For the two models, homoscedasticity and the assumption that the mean of the residuals were equal to zero were satisfied. Meanwhile, the fitness of both models was good based on most scatters of predicted DMY values fell within the 95% confidence interval.
(**Key words :** Italian ryegrass, Yield estimation model, Korean Peninsula, Jeju Island)

Ⅰ. INTRODUCTION

Along with the development of the economy and society, the requirement of high quality livestock products is continuously growing (Tilman et al., 2002; Thornton, 2010). Meanwhile, the high quality forage supplement for ruminant animals is highly related to animal physiology conditions and the safety and quality of livestock products (Robinson et al., 2004). Therefore, high quality livestock production is greatly based on the production of high quality forages.

For the reason of being limited of small area of suitable land for agriculture, the forage production scale in South Korea is not big (Lee and Muller, 2012). Korea imports lots of forages for livestock industry, this costs lots of money. Meanwhile, long way transportation of forages may result in quality decreasements of forage (Choi et al., 2014). Therefore, improvement of domestic production of forages crops and grasses becomes very necessary for both reducing costs and producing high quality forages for Korean livestock industry.

Italian ryegrass (*Lolium multiflorum* Lam., IRG) has high feed value and is the representative winter forage crop in South Korea; meanwhile, IRG silage is popular at the south areas of Republic of Korea (Sung et al., 2012). To produce high quality forages, a forecasting system for yield of IRG is very necessary for production management and planning annual imports of forages (Mkhabela et al., 2011).

Furthermore, as the climate faces more and more unstable fluctuations, the global warming is becoming a new challenge in agriculture and animal industries. Plants, especially forage crops, may face more stress in relation with environmental changes such as cold stress and heat stress (Hatfield et al., 2011). Ignoring climatic stresses will lead to negative effects on forage production and subsequently the quality

* Corresponding author : Kyung Il Sung, Department of Feed Science and Technology, College of Animal Life Sciences, Kangwon National University, Chuncheon, 24341, South Korea, Tel: +82-33-250-8635, Fax: +82-33-242-4540, E-mail: kisung@kangwon.ac.kr

and safety of animal products such as meat and milk. Meanwhile, climate change in South Korea was more significant than world average level (Chung et al., 2007; Hong et al., 2014). Therefore, a forage crop yield prediction tool considering climatic conditions becomes very necessary in South Korea.

Crop yield prediction research has a long history, since 1960s, scientists in United States and Netherland had started to construct several crop growth models to predict the crop yields (Oteng-Darko et al., 2013). Based on the development of computer science and optical observation technologies such as remote sensing satellite and infrared spectrometry, yield estimation technologies based on the crop growth models or optical observation technologies have been developed tremendously (Rauff and Bello, 2015). However, most researches were focusing on food crops and cash crops, and lots of measures and variables in these models resulted in difficulties to modify them in South Korea. Furthermore, few researches were paid attention to forage crops and grasses, especially in South Korea. Therefore, statistical method such as regression analysis was considered as the proper way to construct forage crops yield prediction models based on the actual situation in South Korea. Therefore, this research was conducted to construct the yield forecasting models of IRG based on climatic data via statistical methods.

Ⅱ. MATERIALS AND METHODS

1. Data collection and preparation

The IRG data set used in this research was collected from the results of the adaptability test of imported varieties of grasses and forage crops operated by National Agricultural Cooperative Federation, the reports on joint research projects for new plant variety development operated by Rural Development Administration, research papers in Journal of the Korean Society of Grassland and Forage Science, research reports about livestock experiments operated by Korean National Livestock Research Institute, and Korean crop (farm) survey reports during the 27 years from 1986 to 2013. The sample size of the raw data was 1107 with 125 forage cultivators. Repeated records and undependable records (32) were eliminated and 1075 data points were kept in the final IRG data set.

Raw meteorological data including daily mean temperature, daily maximum temperature, daily minimum temperature, daily precipitation, and sunshine duration was collected from website of meteorological administration based on the cultivated locations in IRG data set. Meteorological data from the nearest meteorological administration was used when some IRG cultivated locations had no meteorological office. Afterwards, temperature, sunshine, and rainfall related variables including autumnal growing days (AGD, day), autumnal accumulated temperature (AAT, ℃), spring growing days (SGD, day), spring accumulated temperature (SAT, ℃), period to accumulated temperature 150 (PAT150, day), period to accumulated temperature 100 (PAT100, day), spring sunshine time (SST, hour), spring rainfall (SRF, mm), spring rainfall days (SRD, day), highest temperature in January (HTJ, ℃), mean temperature in January (MTJ, ℃), and lowest temperature in January (LTJ, ℃) were prepared. The detailed explanations of the generated climatic variables were presented in Table 1.

Finally, the IRG data set and the data set containing generated climatic variables were combined into the final data set used for statistical analyses. Data points with missing values (n=32) were eliminated, and the data set was divided into two data sets based on cultivated locations, therefore the data set of south areas of Korean Peninsula with a sample size of 940 and the data set of Jeju Island with a sample size of 135 were prepared.

For data set of south areas of Korean Peninsula, the outliers (n=7) were deleted after detection via box-plots under the normality assumption. Therefore, a final data set (n=933) during 26 years from 1986 to 2013 (data records in 1987 were only in Jeju Island) with dry matter yield (DMY) values of IRG, 20 cultivated locations, and climatic variables was generated and used in the following analyses. For the data set of Jeju Island, 5 outliers were eliminated and the final data set during 17 years from 1993 to 2013 (no data records in 1996, 1999, 2000, and 2008) contained 130 data points. The cultivated locations and sample size in each location including Jeju Island and the other 22 cultivated locations in the south areas of Korean Peninsula were showed in Fig. 1.

Table 1. The description of generated climatic variables

Climatic Variables	Description
AGD (autumnal growing days, day)	the number of growing days from the sowing date to the day on which the mean daily temperature is above 0℃ in autumn
AAT (autumnal accumulated temperature, °C)	the accumulated temperature from the sowing date to the day on which the mean daily temperature is above 0℃ in autumn
SGD (spring growing days, day)	the number of growing days from the day on which the mean daily temperature is above 0℃ in the next spring to the harvest day
SAT (spring accumulated temperature, ℃)	the accumulated temperature from 1 January to the harvest day in the next spring
PAT150 (period to accumulated temperature 150, day)	the number of days from 1 January to the day on which the accumulated temperature reaches 150℃
PAT100 (period to accumulated temperature 100, day)	the number of days from 1 January to the day on which the accumulated temperature reaches 100℃
SST (spring sunshine time, hour)	the accumulated sunshine hours from the day on which the mean daily temperature is above 5℃ in the next spring to the harvest day
SRF (spring rainfall, mm)	the total spring rainfall from 1 January until the harvest day
SRD (spring rainfall days, day)	the number of days with rainfall from 1 January until the harvest day
HTJ (highest temperature in January, ℃)	the mean of the maximum daily temperature in the coldest month (January in South Korea)
MTJ (mean temperature in January, ℃)	the mean of the mean daily temperature in the coldest month (January in South Korea)
LTJ (lowest temperature in January, ℃)	the mean of the minimum daily temperature in the coldest month (January in South Korea)

2. Statistical Analyses

1) Independent samples t-test

Independent samples t-test was used for the confirmation of statistical differences of the values (DMY and climatic variables) between south areas of Korean Peninsula and Jeju Island. A p value level of 0.05 was considered as the standard significance level in this research. The independent samples t-test is used to test the two population means are equal or not when these two populations are separately independent and identically distributed (Armitage et al., 2008). When assuming the variances of the two samples were equal, the test statistics (t value, pooled variance, and degree of freedom) calculated by the equation as follows:

$$t=\frac{\overline{x_1}-\overline{x_2}}{\sqrt{s^2_{pooled}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$$s^2_{pooled}=\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$$

$$df = n_1+n_1-2$$

where $\overline{x_1}$ and $\overline{x_2}$ are the means, s^2_{pooled} is the pooled variance of samples, n_1 and n_2 are the sample sizes and t is a Student t quantile with n_1+n_2-2 degrees of freedom (StatsDirect Limited, 2016).

When assuming the variances of the two samples were unequal, the test statistics (are calculated by the equation as follows:

$$d=\frac{\overline{x_1}-\overline{x_2}}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$$

$$df=\frac{\left[\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right]^2}{\frac{(s_1^2/n_1)^2}{n_1-1}+\frac{(s_2^2/n_2)^2}{n_2-1}}$$

where $\overline{x_1}$ and $\overline{x_2}$ are the means, s^2_{pooled} is the pooled variance of samples, n_1 and n_2 are the sample sizes, and d is the test statistic used as a Student t quantile with df freedom (StatsDirect Limited, 2016).

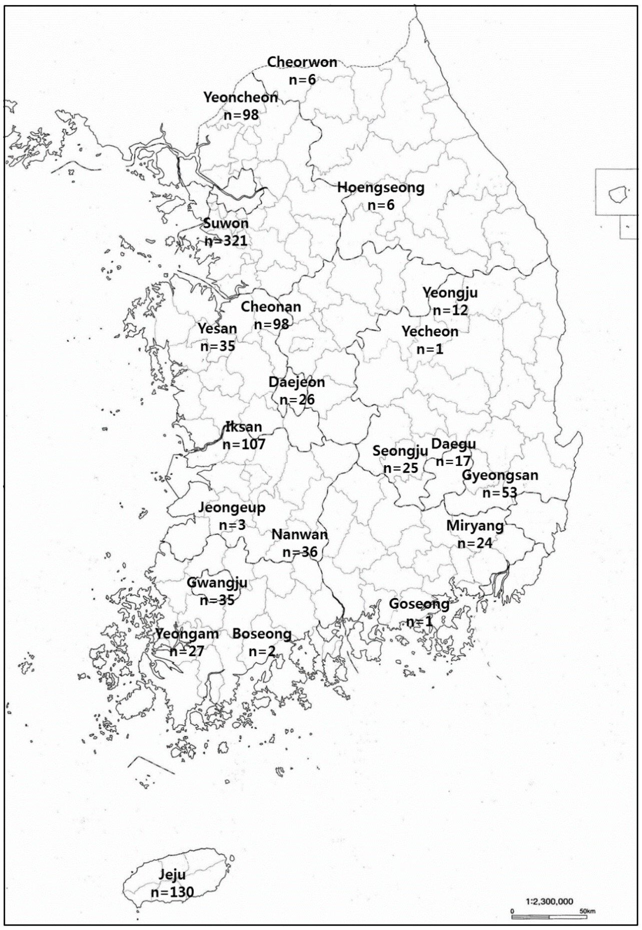

Fig. 1. Map with sample size of the cultivated locations in the final data set.

2) Multiple regression analysis

Multiple regression analysis is used to assess the relationship between continuous independent variables (at least two) and the dependent variable (Mardia et al., 1979). The equation of multiple regression model is as follows:

$$Y_{n\times1} = X_{n\times(p+1)}\beta_{(p+1)\times1} + \varepsilon_{n\times1},\ \varepsilon\sim \text{i.i.d. N } (0,\ \delta^2)$$

Where Y, X, β, ε are the vector of the response variable, the vector of explanatory variables, the matrix of coefficients of explanatory variables, and the vector of residual, respectively. ε is independent and identically distributed.

Multicollinearity is the phenomenon in which two or more explanatory variables in a multiple regression analysis are highly correlated (Farrar and Glauber, 1967). The correlation coefficients were calculated through correlation analysis among explanatory variables. Referring to the variance inflation factor (VIF) of the explanatory variables, the variables with multicollinearity could be detected.

3) General linear model

General linear model was used for constructing models including continuous climatic variables and dummy variables (Mardia et al., 1979). The equation of general linear model is as follows:

$$Y_{n\times1} = X_{n\times(p+1)}\beta_{(p+1)\times1} + Z_{n\times c}\gamma_{c\times1} + \varepsilon_{n\times1},\ \varepsilon\sim \text{i.i.d. N } (0,\ \delta^2)$$

Where Y is response variable, X is explanatory variables. β is coefficient of explanatory variable, Z is dummy variable (c is the dimension of categories), γ is coefficient of dummy variables and ε is (residual) error. ε is independent and identically distributed under homogeneity of variance. Dummy variable is the indicator that takes 0 or 1, where $2^q \le c$ (q is the dimension of indicators).

4) Model evaluation

Residual diagnostic was used to check the fitness of the model. Standardized residuals of the final model were calculated. Then, Probability-Probability plot (P-P plot), scatter plot of standardized residuals against predicted values, and the plot of 95% confidence interval were prepared to check the fitness of the final models to the data sets used in this research.

5) Analysis software

Microsoft Excel 2010 (Microsoft Corp 2010) was used to prepare the data sets and SPSS 21.0 (IBM Corp 2012) was used to perform all the statistical analyses in this study. Meanwhile, only main effects of the explanatory variables were investigated for better simplicity and predictability of the yield prediction models.

Ⅲ. RESULTS AND DISCUSSION

1. Independent sample t-test results for DMY and climatic factors between cultivated locations in south areas of Korean Peninsula and Jeju Island

The results of independent sample t-test for DMY of IRG between cultivated locations in south areas of Korean

Peninsula and Jeju Island were shown in Table 2. The homoscedasticity assumption was not satisfied ($p < 0.05$) based on the results of F test. This may be caused by the difference of sample size between cultivated locations in south areas of Korean Peninsula and Jeju Island (933 vs. 130). The DMY of cultivated locations in south areas of Korean Peninsula was about 4,000 kg/ha significantly less than that of Jeju Island ($p < 0.01$). These results indicated that the difference of DMY between cultivated locations in south areas of Korean Peninsula and Jeju Island was not simply from the sampling error, but from effects of other factors such as climate conditions.

The difference of DMY between cultivated locations in south areas of Korean Peninsula and Jeju Island was due to the effects of soil, climate, and cultivation technologies (Chung, 2007). Meanwhile, climatic factors were considered as the most important one (Hatfield et al., 2011). Temperature, humidity, rainfall, and sunshine were considered as the measurable climatic factors which affected the difference of DMY. In this research, temperature, rainfall, and sunshine related variables were considered.

The results of independent sample t-test for climatic variables of IRG between cultivated locations in south areas of Korean Peninsula and Jeju Island were shown in Table 3. The results of test for homoscedasticity (F test) showed that AAT, LTJ, PAT150, SRF, and SRD did not satisfy the homoscedasticity assumption ($p<0.05$), in contrast, AGD, MTJ, HTJ, SGD, SAT, PAT100, and SST satisfied the homoscedasticity assumption ($p>0.05$). Furthermore, all the climatic variables except PAT100 and PAT150 were significantly higher in Jeju Island ($p<0.05$). In contrast, PAT100 and PAT150 were significantly higher ($p<0.05$) in cultivated locations in south areas of Korean Peninsula.

Based on the results above, it is obvious that both the DMY values and climatic variables were significantly different ($p < 0.05$) between cultivated locations in south areas of Korean Peninsula and Jeju Island. Therefore, in the following analyses, the yield estimation model of IRG will be constructed separately.

2. Model construction for IRG in cultivated locations in south areas of Korean Peninsula

The descriptive statistics, normality, and multicollinearity diagnostics for all the variables in the data set of south areas of Korean Peninsula were presented in Table 4. The mean of DMY was 9343.97 kg/ha and the first quartile and the third quartile were 7042.00 kg/ha and 11376.00 kg/ha, respectively. The mean and median were similar (9343.97 kg/ha vs. 9220.00 kg/ha) and the differences between mean and the first and third quartile were similar (2301.97 kg/ha vs. 2032.03 kg/ha). Therefore, it was judged that the response variable DMY was symmetrically distributed. Meanwhile, other variables were also symmetrically distributed based on the same reasons.

MTJ, SGD, SAT, PAT100, and PAT150 were considered to have multicollinearity problems based on the results of VIF when it is bigger than 10 (Allison, 1999). The reason for multicollinearity of SGD and SAT might be many temperature related variables was included as explanatory variables, and for the multicollinearity between PAT100 and PAT150, the reason might be they were both time related variables and shared some overlapping information. Furthermore, the reason of the presence of multicollinearity problem of MTJ could be explained by three January temperature related variables were included together.

To solve the multicollinearity, correlation analysis including all explanatory variables was performed and the correlation matrix was presented in Table 5. The strong correlations (correlation coefficient > 0.7) were observed between AGD

Table 2. The results of independent sample t-test for DMY[1] of IRG[2] between cultivated locations in south areas of Korean Peninsula and Jeju Island

	F[3]		t[4]			Mean difference (kg/ha)
	Statistics	p-value	Statistics	df	p-value	
Equal variance	31.57	.00	-10.98	1073.00	.00	-4041.40
Not equal variance			-8.39	153.05	.00	

[1] DMY: dry matter yield. [2] IRG: Italian ryegrass.
[3] F: test for homoscedasticity. [4] t: test for mean comparison.

Table 3. The results of independent sample t-test for climatic variables of IRG[1)] between cultivated locations in south areas of Korean Peninsula and Jeju Island

		F[3)]		t[4)]			Mean difference
		Statistics	p-value	Statistics	df	p-value	
AGD[2)] (day)	Equal variance	0.23	0.63	-3.77	1073.00	0.00	-4.51
	Not equal variance			-3.37	163.65	0.00	
AAT (℃)	Equal variance	15.65	0.00	-14.85	1073.00	0.00	-310.45
	Not equal variance			-11.50	153.76	0.00	
LTJ (℃)	Equal variance	27.02	0.00	-25.78	1073.00	0.00	-8.37
	Not equal variance			-32.84	216.14	0.00	
MTJ (℃)	Equal variance	0.91	0.34	-30.25	1073.00	0.00	-6.77
	Not equal variance			-31.61	180.13	0.00	
HTJ (℃)	Equal variance	0.11	0.74	-23.05	1073.00	0.00	-5.69
	Not equal variance			-22.45	171.84	0.00	
SGD (day)	Equal variance	0.26	0.61	-13.08	1073.00	0.00	-23.30
	Not equal variance			-12.82	172.49	0.00	
SAT (℃)	Equal variance	1.38	0.24	-10.25	1073.00	0.00	-316.43
	Not equal variance			-10.75	180.65	0.00	
PAT100 (day)	Equal variance	1.39	0.24	34.12	1073.00	0.00	42.99
	Not equal variance			29.37	160.76	0.00	
PAT150 (day)	Equal variance	4.38	0.04	35.94	1073.00	0.00	42.82
	Not equal variance			29.51	157.40	0.00	
SRF (mm)	Equal variance	3.74	0.05	-7.01	1073.00	0.00	-70.62
	Not equal variance			-6.69	169.68	0.00	
SRD (day)	Equal variance	6.69	0.01	-10.71	1073.00	0.00	-9.39
	Not equal variance			-9.76	165.39	0.00	
SST (hr)	Equal variance	2.71	0.10	-6.94	1073.00	0.00	-113.46
	Not equal variance			-7.42	183.11	0.00	

1) IRG: Italian ryegrass.
2) AGD, autumnal growing days; AAT, autumnal accumulated temperature; LTJ, lowest temperature in January; MTJ, mean temperature in January; HTJ, highest temperature in January; SGD, spring rainfall days; SAT, spring accumulated temperature; PAT100, period to accumulated temperature 100; PAT150, period to accumulated temperature 150; SST, spring sunshine time; SRF, spring rainfall; SRD, spring rainfall days.
3) F: test for homoscedasticity.
4) t: test for mean comparison.

and AAT, LTJ and MTJ, SGD and SAT, PAT100 and PAT150, MTJ and PAT100, MTJ and PAT150, SST and SGD, SST and SAT. Due to the multicollinearity, the variables LTJ, SGD and PAT100 were eliminated. Meanwhile, for the reason that HTJ was not significantly correlated with DMY, it was also eliminated. Therefore, the rest variables included AGD, AAT, MTJ, SAT, PAT150, SST, SRF, and SRD were used in multiple regression analysis.

As showed in Table 6, the optimal climatic variables were selected using the stepwise approach of multiple regression analysis. In this model, the VIF of all the explanatory variables (SAT, AGD, SRF, and AAT) were less than 2 which means it could be concluded that there is no multicollinearity.

Furthermore, the effects of explanatory variables could be recognized by checking the changing degrees of plus-minus signs and magnitudes of Pearson’s correlation coefficients,

Table 4. Descriptive statistics, normality, and multicollinearity diagnostics for all the variables in data set of south areas of Korean Peninsula

	Mean	Median	SE[2)]	Quartile		VIF[3)]
				1st	3rd	
DMY[1)]	9343.97	9220.00	117.67	7042.00	11376.00	
AGD	83.39	85.00	0.42	77.00	91.00	7.40
AAT	843.35	851.30	7.01	725.50	1000.50	8.00
LTJ	-8.50	-8.10	0.12	-10.90	-6.20	4.99
MTJ	-1.93	-1.93	0.08	-2.97	-0.61	11.19
HTJ	4.64	4.70	0.09	3.40	6.00	2.79
SGD	108.26	104.00	0.63	94.00	120.00	23.02
SAT	997.63	925.70	11.08	733.90	1170.90	22.03
PAT100	65.03	67.00	0.43	58.00	76.00	28.61
PAT150	74.75	77.00	0.41	68.00	84.00	27.62
SST	503.79	462.90	5.86	377.60	587.10	3.35
SRF	267.57	258.60	3.56	194.80	337.70	2.31
SRD	37.69	37.00	0.31	31.00	44.00	2.40

1) DMY, dry matter yield; AGD, autumnal growing days; AAT, autumnal accumulated temperature; LTJ, lowest temperature in January; MTJ, mean temperature in January; HTJ, highest temperature in January; SGD, spring rainfall days; SAT, spring accumulated temperature; PAT100, period to accumulated temperature 100; PAT150, period to accumulated temperature 150; SST, spring sunshine time; SRF, spring rainfall; SRD, spring rainfall days.

2) SE: standard error.

3) VIF: variance inflation factor.

Table 5. Correlation matrix including all the variables in data set of south areas of Korean Peninsula

	DMY	AGD	AAT	LTJ	MTJ	HTJ	SGD	SAT	PAT100	PAT150	SST	SRF	SRD
DMY[1)]	1	.185**	.123**	.011	.116**	.040	.527**	.499**	-.232**	-.248**	.357**	.452**	.340**
AGD		1	.917**	.280**	.275**	.040	.095**	.010	-.112**	-.111**	.050	.034	.006
AAT			1	.328**	.282**	.077*	.107**	.017	-.061	-.048	.069*	-.011	.021
LTJ				1	.827**	.398**	.213**	.063	-.477**	-.464**	.185**	-.113**	-.243**
MTJ					1	.692**	.355**	.148**	-.765**	-.755**	.197**	.048	-.046
HTJ						1	.270**	.112**	-.633**	-.618**	.074*	.077*	.205**
SGD							1	.945**	-.399**	-.395**	.728**	.637**	.518**
SAT								1	-.182**	-.186**	.764**	.625**	.540**
PAT100									1	.978**	-.138**	-.245**	-.149**
PAT150										1	-.184**	-.247**	-.148**
SST											1	.337**	.202**
SRF												1	.607**
SRD													1

* P <0.05, ** P < 0.01.

1) DMY, dry matter yield; AGD, autumnal growing days; AAT, autumnal accumulated temperature; LTJ, lowest temperature in January; MTJ, mean temperature in January; HTJ, highest temperature in January; SGD, spring rainfall days; SAT, spring accumulated temperature; PAT100, period to accumulated temperature 100; PAT150, period to accumulated temperature 150; SST, spring sunshine time; SRF, spring rainfall; SRD, spring rainfall days.

Table 6. Results of stepwise multiple regression analysis to detect the optimal climatic variables in the data set of south areas of Korean Peninsula (R square = .324**, adjusted R square = .321**)

	Coefficient		Standard coefficient	t[3]	p-value	VIF[4]	Correlation coefficient		
	B	SE[2]					Pearson's	Partial	Part
Constant	2564.368	1022.373		2.508	0.012				
SAT[1]	3.744	0.368	0.353	10.187	0.000	1.644	0.499	0.317	0.275
AGD	45.576	7.664	0.162	5.947	0.000	1.013	0.185	0.192	0.161
SRF	6.518	1.159	0.197	5.624	0.000	1.690	0.452	0.182	0.152
AAT	-33.450	8.115	-0.116	-4.122	0.000	1.079	-0.248	-0.134	-0.111

1) SAT, spring accumulated temperature; AGD, autumnal growing days; SRF, spring rainfall; AAT, autumnal accumulated temperature.
2) SE: standard error.
3) t:: Student t quantile for testing the significance of variables.
4) VIF: variance inflation factor.

partial correlation coefficients, and part correlation coefficients of explanatory variables (Cohen et al., 2013). AGD almost had no overlapping effect with other variables based on the result that the magnitudes of correlation coefficients had no big differences; this means AGD was independent with other variables. Therefore, the effect of AGD could be interpreted as DMY will increase 45.576 when AGD increases 1 unit. Meanwhile, other variables (SAT, SRF, and AAT) might have overlapping effects with each other based on their correlation coefficients, and the effects of them could not be interpreted in the same way because they were not independent with each other.

Pearson's correlation coefficient could be used to investigate the effect sizes of explanatory variables on the response variable. SAT had the biggest Pearson's correlation coefficient which was 0.499, the Pearson's correlation coefficients of the rest variables were shown as descending in the sequence of SRF, AAT, and AGD. This means the variables related to next spring had strong effects on the DMY of IRG in this data set.

The final yield estimation model was constructed by adding cultivated locations as dummy variables with the selected climatic variables (SAT, AGD, SRF, and AAT) via general linear model. The result was showed in Table 7. The adjusted R square was 32.0% ($p < 0.01$). The adjusted R square of each variable was instead by partial eta squared to express the effect size because dummy variables were included in model. The IRG yield estimation model based on climatic data by locations in South areas of Korean Peninsula was as follows:

$$\text{DMY} = 131.924\text{AGD} - 5.349\ \text{AAT} + 3.703\text{SAT} + 7.305\text{SRF} - 3162.076 + \text{Location}$$

For model of a specific location, the coefficients of the location should be inserted in the model in the Location item in the equation. For example, the IRG yield estimation model of Suwon would be $\text{DMY} = 131.924\text{AGD} - 5.349\ \text{AAT} + 3.703\text{SAT} + 7.305\text{SRF} - 2594.22$ after inputting the Location constant value 567.857.

By comparing the results of Table 6 with 7, it was found that AGD and AAT had obvious changes after adding the location variable. This might mean that AGD and AAT were not independent with cultivated locations, which means AGD and AAT could reflect the different characteristics of the cultivated locations. Meanwhile, it was shown that there were no big changes in the regression coefficients of SAT and SRF. This was thought to be that farmers in different cultivated locations would adjust the seeding and harvest dates in spring to ensure enough accumulated temperature and water supplement for the growth and yield production of IRG. SAT had the biggest partial eta squared, the partial eta squared of the rest variables were shown as descending in the sequence of AGD, SRF and AAT. This might indicate that accumulated temperature in spring has the biggest effect on yield production and subsequently rainfall in spring also had an important influence.

Fig. 2 presented the plot of residual for assessing the goodness-of-fit in this model. Here, in the P-P plot of regression standardized residuals, the points arranged on the line almost well showed that the normality assumption of

Table 7. Results of general linear model for IRG yield estimation in south areas of Korean Peninsula (R square = .337**, adjusted R square = .320**)

Parameter	Coefficient	SE[2)]	t[3)]	p-value	partial eta squared
Constant	-3162.076	1648.548	-1.918	0.055	0.004
AGD[1)]	131.924	22.162	5.953	0.000	0.037
AAT	-5.349	1.390	-3.849	0.000	0.016
SAT	3.703	0.452	8.194	0.000	0.068
SRF	7.305	1.323	5.522	0.000	0.032
[Location = Gyeongsan]	-41.161	1384.929	-0.030	0.976	0.000
[Location = Goseong]	3936.470	3359.348	1.172	0.242	0.001
[Location = Gwangju]	-864.090	1424.817	-0.606	0.544	0.000
[Location = Namwon]	-193.218	1416.677	-0.136	0.892	0.000
[Location = Daegu]	1843.588	1565.656	1.178	0.239	0.002
[Location = Daejeon]	-421.197	1431.581	-0.294	0.769	0.000
[Location = Miryang]	-703.658	1436.114	-0.490	0.624	0.000
[Location = Boseong]	-2897.180	2542.658	-1.139	0.255	0.001
[Location = Seongju]	193.418	1425.302	0.136	0.892	0.000
[Location = Suwon]	567.857	1339.137	0.424	0.672	0.000
[Location = Yeoncheon]	1653.458	1334.046	1.239	0.216	0.002
[Location = Yeongam]	-298.356	1523.431	-0.196	0.845	0.000
[Location = Yeongju]	1123.827	1578.060	0.712	0.477	0.001
[Location = Yesan]	598.584	1391.665	0.430	0.667	0.000
[Location = Yecheon]	-1698.347	3369.608	-0.504	0.614	0.000
[Location = Iksan]	815.162	1329.150	0.613	0.540	0.000
[Location = Jeongeup]	830.642	2225.808	0.373	0.709	0.000
[Location = Cheonan]	89.580	1350.240	0.066	0.947	0.000
[Location = Cheorwon]	-2445.800	1803.071	-1.356	0.175	0.002
[Location = Hoengseong]	0[a]				

1) AGD, autumnal growing days; AAT, autumnal accumulated, SAT, spring accumulated temperature; temperature; SRF, spring rainfall.
2) SE : standard error.
3) t : Student t quantile for testing the significance of variables.

the residual is satisfied. Furthermore, in the scatter plot of standardized residuals against the predicted values of DMY, the points scattering without a particular pattern means that the homoscedasticity and the assumption that the mean of the residuals is equal to zero was well fulfilled.

The figure of 95% confidence interval was presented in Fig. 3, most scatters of the predicted values of DMY were in the 95% confidence interval indicated that though the adjusted R squared of this model is not big, the fitness of the model is good and acceptable.

Here, the reasons of low adjusted R squared of this model might be that other crop growth related variables were not included in this study. Additionally, researches on interaction terms of the explanatory variables were also applied. Though the model fitness turned a little better, the interpretation of the interaction terms was difficult. Therefore, in the final model, the interaction terms were not included.

3. Yield estimation model construction for IRG in Jeju Island

The Descriptive statistics, normality, and multicollinearity

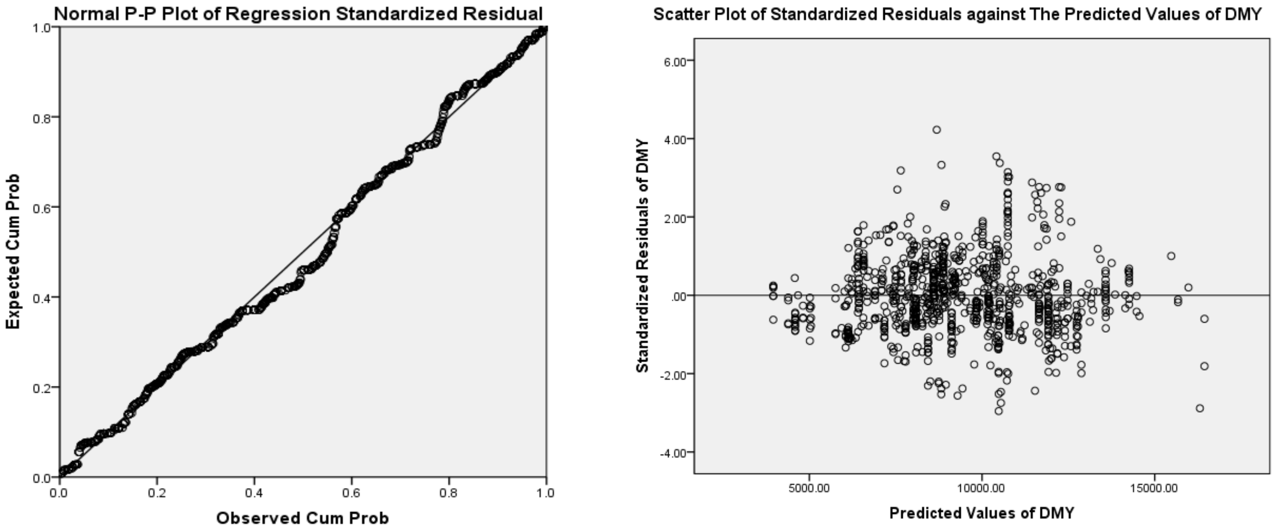

Fig. 2. Results of residual diagnostics of model for Italian ryegrass yield estimation in south areas of Korean Peninsula (left: normal P-P plot of regression standardized residual, right: scatter plot of standardized residuals against predicted values of dry matter yield from the model).

diagnostics for all the explanatory variables in Jeju Island were presented in Table 8. The mean of DMY was 12910.88 kg/ha and the first quartile and the third quartile were 9207.00 kg/ha and 15930.00 kg/ha, respectively. The mean and median were similar (12910.88 kg/ha vs. 13209.00 kg/ha) and the differences between mean and the first and third quartile were not dramatically different (3703.88 kg/ha vs. 3019.12 kg/ha). Therefore, DMY could be judged as

Table 8. Descriptive statistics, normality, and multicollinearity diagnostics for all the variables in data set in Jeju Island

	Mean	Median	SE[2)]	Quartile		VIF[3)]
				1st	3rd	
DMY[1)]	12910.88	13209.00	406.81	9207.00	15930.00	
AGD	87.82	87.00	1.32	80.00	99.00	66.15
AAT	1152.07	1119.55	27.09	975.60	1411.90	49.72
LTJ	-0.17	0.20	0.23	-0.70	1.40	30.78
MTJ	4.79	5.51	0.21	4.94	6.02	292.15
HTJ	10.28	10.30	0.25	9.40	11.40	4.39
SGD	131.06	128.00	1.74	123.00	143.00	234.26
SAT	1303.38	1203.70	27.34	1130.40	1483.30	234.94
PAT100	22.23	14.00	1.45	13.00	19.00	379.89
PAT150	32.11	27.00	1.45	24.00	30.00	269.99
SST	616.00	576.70	14.51	512.20	671.80	17.10
SRF	337.63	282.90	10.31	262.70	424.30	10.32
SRD	46.85	43.00	0.94	38.00	53.00	14.97

1) DMY, dry matter yield; AGD, autumnal growing days; AAT, autumnal accumulated temperature; LTJ, lowest temperature in January; MTJ, mean temperature in January; HTJ, highest temperature in January; SGD, spring rainfall days; SAT, spring accumulated temperature; PAT100, period to accumulated temperature 100; PAT150, period to accumulated temperature 150; SST, spring sunshine time; SRF, spring rainfall; SRD, spring rainfall days.

2) SE: standard error.

3) VIF: variance inflation factor.

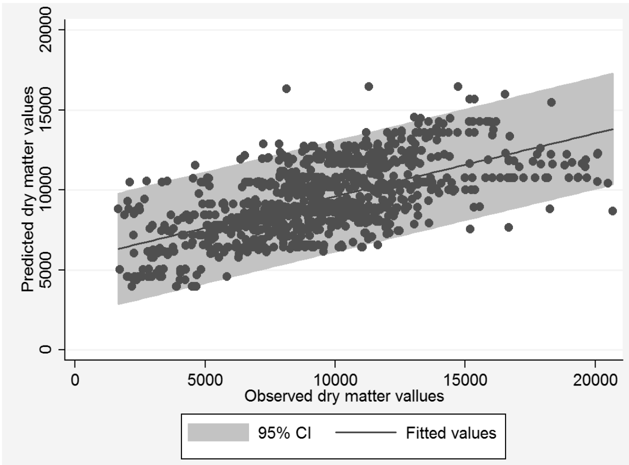

Fig. 3. Scatter plot including mean regression line and 95% confidence interval for observed dry matter yield and predicted dry matter yield from the model for Italian ryegrass yield estimation in south areas of Korean Peninsula.

symmetrically distributed and other variables were also symmetrically distributed based on the same reason.

All the variables except HTJ were considered to have multicollinearity problems based on the VIF values were bigger than 10. To solve the multicollinearity, correlation analysis was performed and the correlation matrix was presented in Table 9. Strong correlation relationships (correlation coefficient > 0.7) were observed between AGD and AAT, LTJ and MTJ, SGD and SAT, PAT100 and PAT150, MTJ and PAT100, MTJ and PAT150, LTJ and PAT100, LTJ and PAT150, SST and SRD. Due to the multicollinearity, the variables AGD, LTJ, SGD, PAT100, PAT150, and SRD were eliminated. Meanwhile, for the reason that HTJ was not significantly correlated with DMY, it was also deleted. However, though in this data set, MTJ was not significantly correlated with DMY, it was an important climatic variable which could affect the vernalisation of IRG, so it was kept. Therefore, the rest variables included AAT, MTJ, SAT, SRF, and SST were used in multiple regression analysis.

As showed in Table 10, the optimal IRG yield prediction model considering the selected climatic variables in Jeju Island was generated via the stepwise approach of multiple regression analysis. The equation of the model was as follows:

$$DMY = -8.044AAT + 18.640SST - 7.542SAT + 9.610SRF + 17282.191$$

The adjusted R square of this model was 51.0% ($p < 0.01$). In this model, the VIF values of all the explanatory

Table 9. Correlation matrix including all the variables of the data set in Jeju Island

	DMY	AGD	AAT	LTJ	MTJ	HTJ	SGD	SAT	PAT100	PAT150	SRF	SRD	SST
DMY[1)]	1	-.579**	-.603**	-.216*	-.014	-.022	.193*	.315**	.063	.054	.459**	.236**	.402**
AGD		1	.979**	.173*	.045	.005	-.371**	-.466**	-.174*	-.099	-.291**	-.423**	-.518**
AAT			1	.246**	.101	.022	-.356**	-.437**	-.198*	-.134	-.292**	-.388**	-.510**
LTJ				1	.868**	.437**	.379**	.286**	-.812**	-.818**	.010	.226**	.249**
MTJ					1	.648**	.327**	.296**	-.953**	-.971**	.113	.369**	.389**
HTJ						1	.153	.157	-.583**	-.634**	-.139	.423**	.486**
SGD							1	.969**	-.352**	-.334**	.658**	.408**	.705**
SAT								1	-.281**	-.274**	.722**	.445**	.746**
PAT100									1	.988**	-.177*	-.332**	-.324**
PAT150										1	-.109	-.383**	-.338**
SRF											1	0.015	.429**
SRD												1	.702**
SST													1

* P <0.05, ** P < 0.01.

1) DMY, dry matter yield; AGD, autumnal growing days; AAT, autumnal accumulated temperature; LTJ, lowest temperature in January; MTJ, mean temperature in January; HTJ, highest temperature in January; SGD, spring rainfall days; SAT, spring accumulated temperature; PAT100, period to accumulated temperature 100; PAT150, period to accumulated temperature 150; SST, spring sunshine time; SRF, spring rainfall; SRD, spring rainfall days.

Table 10. Results of multiple regression model for IRG yield estimation in Jeju Island (R square = .526**, adjusted R square = .510**)

	Coefficient		Standard	t[3]	p-value	VIF[4]	Correlation coefficient		
	B	SE[2]	coefficient				Pearson	Partial	Part
Constant	17282.191	2389.706		7.232	0.000				
AAT[1]	-8.044	1.064	-0.536	-7.558	0.000	1.324	-0.603	-0.560	-0.466
SST	18.640	2.882	0.665	6.467	0.000	2.785	0.459	0.501	0.398
SAT	-7.542	1.714	-0.507	-4.400	0.000	3.498	0.315	-0.366	-0.271
SRF	9.610	3.230	0.243	2.975	0.004	1.764	0.236	0.257	0.183

[1] AAT, autumnal accumulated temperature; SST, spring sunshine time; SAT, spring accumulated temperature; SRF, spring rainfall.
[2] SE: standard error.
[3] t: Student t quantile for testing the significance of variables.
[4] VIF: variance inflation factor.

variables (AAT, SST, SAT, and SRF) were less than 4 which means it could be concluded that there is no multicollinearity. Furthermore, AAT, SST, and SRF had no overlapping effect with other variables based on the results that the magnitudes of correlation coefficients (Pearson's, Partial, and Part) were similar. Therefore, the effects of AAT, SST, and SRF could be interpreted in the way that DMY will increase by the values of regression coefficient when these variables increase 1 unit. Meanwhile, SAT might have overlapping effects and couldn't be explained in the same way.

Fig. 4 presented the results of residual diagnostics for assessing the goodness-of-fit of this model. Here, the points in the P-P plot exhibiting no specific pattern around the 45-degree line represent a good fit. Furthermore, the points were scattered well without a particular pattern on the figure of standard prediction and standard residual, this means that the assumptions of homoscedasticity and the mean of the residuals are equal to zero were satisfied. As presented in Fig. 5, the fitness of the model was confirmed based on most scatters of the predicted DMY values fell in the 95% confidence interval. Similar with the model of south areas of Korean Peninsula, the interaction terms were not included in the final model to for the same reason.

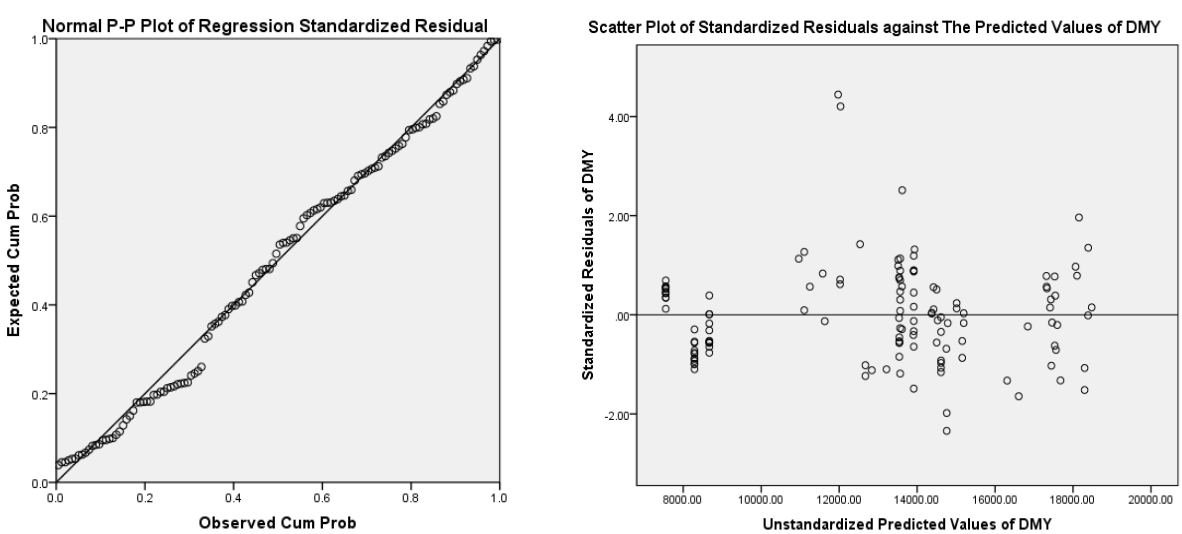

Fig. 4. Results of residual diagnostics of model for Italian ryegrass yield estimation in Jeju Island (left: normal P-P plot of regression standardized residual, right: scatter plot of standardized residuals against predicted values of dry matter yield from the model).

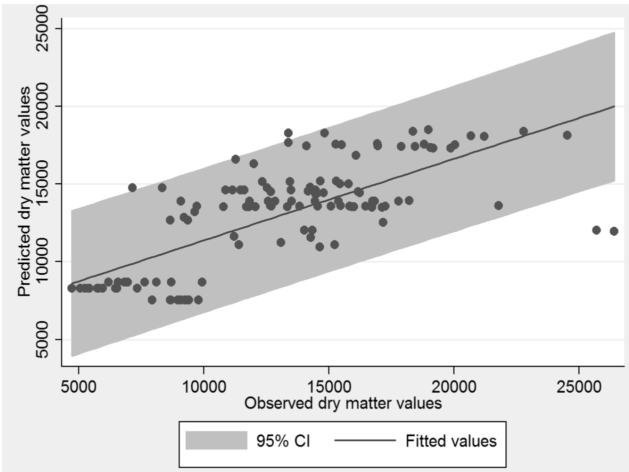

Fig. 5. Scatter plot including mean regression line and 95% confidence interval for observed dry matter yield and predicted dry matter yield from the model for Italian ryegrass yield estimation in Jeju Island.

Ⅳ. CONCLUSION

From the study results, we have demonstrated the differences of DMY and climatic variables between the cultivated locations in south areas of Korean Peninsula and Jeju Island, the yield estimation models of IRG were constructed separately. The results showed that the adjusted R square of IRG yield estimation models of south areas of Korean Peninsula and Jeju Island were 32.0% and 51.0%, respectively. The fitness of both the two models was acceptable based on the results of model diagnostics. The adjusted R square was bigger in the model of Jeju Island suggested that, in the further research, construct the IRG yield estimation model for each location based on improvement of sample size of each cultivated location could decrease the disturbance of spatial variances and increase the precision of modeling. Meanwhile, collection and adding more yield related variables such as soils, cultivars, and cultivation management technologies could help to increase the adjusted R square of models.

Ⅴ. ACKNOWLEDGEMENT

This study was supported by a grant from the Bio-industry Technology Development Program (313020-04), Ministry of Agriculture, Food, and Rural Affairs, Republic of Korea and also with the support of the “Cooperative Research Program for Agriculture Science and Technology Development (Project No. PJ01028303),” Rural Development Administration, Republic of Korea.

Ⅵ. REFERRENCES

Allison, P.D. 1999. Multiple regression: A primer. Pine Forge Press. Newbury Park. pp. 141-143.

Armitage, P., Berry, G. and Matthews, J.N.S. 2008. Statistical methods in medical research. 4th ed, John Wiley & Sons, Malden, USA. pp.83-162.

Choi, S.W., Lee, C.S., Park, C.H., Kim, D.H., Park, S.K., Kim, B.G. and Moon, S.H. 2014. Prediction of nutrient composition and in-vitro dry matter digestibility of corn kernel using near infrared reflectance spectroscopy. Journal of the Korean Society of Grassland and Forage Science. 34(4):277-282.

Chung, Y.S., Yoon, M.B. and Kim, H.S. 2004. On climate variations and changes observed in South Korea. Climatic Change. 66(1-2): 151-161.

Chung, C.H. 2007. Vegetation response to climate change on Jeju Island, South Korea, during the last deglaciation based on pollen record. Geosciences Journal. 11(2):147-155.

Cohen, J., Cohen, P., West, S.G. and Aiken, L.S. 2013. Applied multiple regression/correlation analysis for the behavioral sciences. Routledge. New York. pp 310-316.

Excel (2010) Microsoft Excel 2010. Microsoft Corp., Redmond, WA, USA.

Farrar, D.E. and Glauber, R.R. 1967. Multicollinearity in regression analysis: the problem revisited. The Review of Economic and Statistics. 49(1):92-107.

Hatfield, J.L., Boote, K.J., Kimball, B.A., Ziska, L.H., Izaurralde, R.C., Ort, D., Thomson, A.M. and Wolfe, D. 2011. Climate impacts on agriculture: implications for crop production. Agronomy Journal. 103(2):351-370.

Hong, J.H., Kim, S.K., Kim, D.G., Hong, S.C., Lee, J.B., Moon, K.J., Cha, J.S., Hong, J.S., Ma, Y.I., Kim, S.Y., Jung, H.C. and Choi, Y.E. 2014. Korean Climate Change Assessment Report 2014. Korean National Institute of Environment Research. Incheon. pp.1-45.

Lee, S. and Muller, A.R. 2012. South Korean external strategy qualms: Analysis of Korean overseas agricultural investment within the global food system. Proceedings of Global Land Grabbing II conference. Ithaca. USA. pp. 1-31.

Mardia, K.V., Kent, J.T. and Bibby, J.M. 1979. Multivariate Analysis (Probability and Mathematical Statistics). Academic Press, London-New York-Toronto-Sydney-San Francisco. pp. 1-518.

Mkhabela, M.S., Bullock, P., Raj, S., Wang, S. and Yang, Y. 2011. Crop yield forecasting on the Canadian Prairies using MODIS NDVI data. Agricultural and Forest Meteorology. 151(3): 385-393.

Oteng-Darko, P., Yeboah, S., Addy, S.N.T., Amponsah, S. and Danquah, E.O. 2013. Crop modeling: a tool for agricultural research－a review. Journal of Agriculture Research and Development. 2(1): 1-6.

Rauff, K.O. and Bello, R. 2015. A review of crop growth simulation models as tools for agricultural meteorology. Agricultural Sciences. 6(9):1098.

Robinson, J. 2004. Pasture Perfect: The Far-Reaching Benefits of Choosing Meat, Eggs, and Dairy Products from Grass-Fed Animals. Vashon Island Press. Washington. pp. 1-152.

SPSS (2012) IBM SPSS Statistics 21.0. IBM Corp., Somers, New York. USA.

StatsDirect Limited. Unpaired (Two Sample) t Test. Available from: http://www.statsdirect.com/help/default.htm#parametric_methods/unpaired_t.htm%3FTocPath%3DParametric%2520methods%7C_____3. Accessed Aug. 13, 2016.

Sung, K.I., Nejad, J.G., Song, Y.H., Kim, S.Y., Lee, B.H. and Kim, W.H. 2012. A comparison of feeding whole crop barley mixed with Italian ryegrass silage versus tall fescue hay for Holstein growing cattle. Proceedings of the XVI International Silage Conference. Finland. pp. 508-509.

Thornton, P.K. 2010. Livestock production: recent trends, future prospects. Philosophical Transactions of the Royal Society B: Biological Sciences, 365(1554), 2853-2867.

Tilman, D., Cassman, K.G., Matson, P.A., Naylor, R. and Polasky, S. 2002. Agricultural sustainability and intensive production practices. Nature. 418(6898):671-677.

(Received August 11, 2016 / Revised September 5, 2016 / Accepted September 6, 2016)