Development of Yield Prediction System Based on Real-time Agricultural meteorological Information

Haedong Lee *, Aekyung Moon*

** ETRI, 218 Gajeong-ro, Yuseong-gu, 305-700, Korea*

haenamu@etri.re.kr, akmoon@etri.re.kr

***Abstract*—This paper contains about the research and the building of an effective agricultural yield forecasting system based on real-time monthly weather. It is difficult to predict the agricultural crop production because of the abnormal weather that happens every year and rapid regional climate change due to global warming. The development of agricultural yield forecasting system that leverages real-time weather information is urgently required. In this research, we cover how to process the number of weather data(monthly, daily) and how to configure the prediction system. We establish a non-parametric statistical model on the basis of 33 years of agricultural weather information. According to the implemented model, we predict final production using the monthly weather information. This paper contains the results of the simulation.**

***Keywords*—Prediction, Meteorology, Agriculture, Yield**

I. INTRODUCTION

The agricultural observation induces farmers to adjust their agricultural production and shipment. And it contributes the government's supply policy and price stabilization measures. To do business in the country, such as agriculture observations, there have been a lot of research and system building. However, due to global warming and abnormal weather, agricultural forecast is becoming more complex and more difficult. We can recognize the gravity of this situation from the cabbage wave repeated every year. Agricultural production of next year is affected from crop degree of preceding year, price, consumption patterns, imported agricultural products. It takes into consideration with the impact of a combination of many different factors. The prediction of agricultural products can be called divine notch.

At domestic, several government-related institutes performed many studies and government-funded projects in order to support reliable prediction service. In order to reliably predict the quality we need a lot of statistical data. However, in the past, many data were not stacked because of the lack of computing processing power and data storage technology. According to recent advances in technology, large amounts of distributed storage technologies, increased computing power have been surged. Unlike in the past, big data technology as recently spotlighted has been greatly improved.

This paper includes the research necessary to constructing reliable prediction and information system through convergence between ICT technology and agricultural observation technology. This allows the existing agricultural observation project and system to be more effective and efficient. This is a good fusion case of ICT and agricultural practices that will contribute to increasing national competitiveness.

II. RELATED WORKS

There are several previous studies about building crop yield model according to meteorological factors.

At [2], probability estimation method was applied. The panel was constructed using regional domestic data. It uses two-way method in consideration of the cross-sectional effect and time-series. As a result, it was developed the model that predicts the crops yield per unit in consideration of weather factors. However, diseases and pests and weather disasters were not taken into account because of the lack of experimental statistics. In addition, it used the average data of the last five years for the diurnal temperature range, sunshine hours, high/low temperature because KMA(Korea meteorological Administration)[3] did not include those as long-term forecast items. There is the limitation that it is qualitative estimates rather than quantitative estimates

At [4], the research was conducted to analyse fluctuating vegetable price and patterns and factors. This study aimed to identify time series fluctuation features including trends and seasonality of each vegetable item, and to investigate the factors of fluctuating supply and demand. Fluctuating vegetable prices were analysed by means of a method of using fluctuation indicators and a method of estimating and examining time series models. APC and CV were used as indicators for showing fluctuating prices. ARIMA and GARCH were applied as time series model[4].

At recent, meteorological factors has been highlighted in relation with climate change. Thus many studies were conducted crop forecasting taking into account with meteorological variables.

In this study, we developed 'yield per crop prediction model' taking into account various weather factors, as well as the characteristics of the specific crops. We use cumulative weather of the year and reflect the latest weather collected in real-time.

III. PREDICTION

In this paper, we estimate the yearly yield of the apple. Cultivation area of the apple, as a fruit tree is composed of the area of young tree and of mature tree. Total year production output is calculated as the multiplication of total cultivated area and yield per unit area. We emphasize on estimating the yield of the crops utilizing weather information.

A. *Explanation variables*

We used daily meteorological raw data of the 35 weather stations that collected for 33 years and generated nationwide weather information: processing the raw data and made the monthly average. Those are average monthly temperature, monthly hours of sunshine, average monthly rainfall as shown in the table 1. The following table shows the used meteorological data as the inputs of modeling and in addition, we generate the secondary meteorological data based on this information. We try to reflect the characteristics of apples with these into the model. Those are daily temperature difference, the degree of high temperature, the degree of rainfall and the sunscald.

TABLE 1. USED METEOROLOGY DATA IN KOREA

		unit	year	month	period
apple	yield	weight	○		1980~2012
Average Weather Data (Korea)	Maximum Temp	℃		○	1980~2012
	Minimum Temp	℃		○	1980~2012
	Average Temp	℃		○	1980~2012
	Rain Amount	mm		○	1980~2012
	Hours of Sunshine	Hr		○	1980~2012

The sugar content of the apple is mainly influenced by temperature difference of the day and the night. Temperature decreases at night forces the plant to stop breathing and to accumulate nutrients in the apple fruit. And then fruit size and sugar content increases. Temperature difference from March to October period was applied.

When the temperature is above 30°C, carbohydrate consumption by respiration becomes more than carbohydrate synthesis. As a result, it prevents accumulation of material. Formation of fruit buds becomes too poor and fruit enlargement becomes worse. So we add the degree of high temperature as the secondary variable.

(a) Sunburn (b) Sunscald (c) Delayed Sunscald

Figure 1. Damaged Apple Photos[5]

Unlike other crops, the sunscald may occur in apple. Sunscald occurs when tomatoes, peppers, or apples are exposed to the direct rays of the sun during hot weather. The damaged areas are vulnerable to attack by insects, fungi, and bacteria[5]. At Korea, when the temperature is 32 ° C or more from July to August, this phenomenon happens. We add sunscald-related variable into the model of July to August.

Excessive rainfall in the growing period negatively affects the production of apples. We compute the degree of rainfall from precipitation data of rainy season from May to September and use it as a new variable.

B. *Apple Production Model*

In this study, kernel smoothing[6], one of the non-parametric regression methods were used

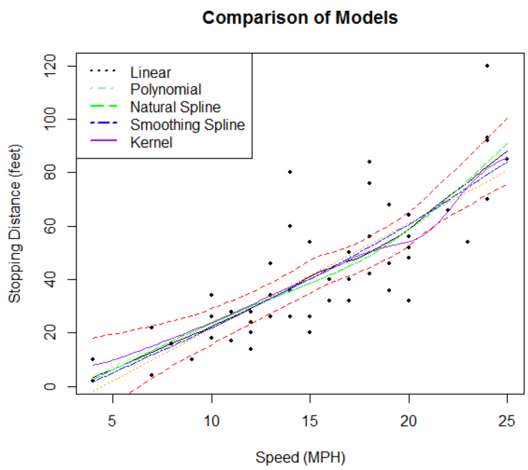

Figure 2. Comparison of Models [7]

Non-parametric method, as shown in the figure above, is the method used when it is difficult to find the linearity of the data. The crop production forecast model was developed taking into account the specific characteristics of crop and the various weather factors. Based on current monthly weather information as well as cumulative weather, we predicted the yearly production for an apple. We use explanation variables(as shown at table 2): yearly production, monthly average maximum temperature, monthly average temperature, monthly average minimum temperature, monthly average minimum temperature, Average monthly precipitation, daily temperature difference, change degreee in precipitation, degree of high temperature, degree of sunscald, degree of rainfall and constructed non-parametric model for apple and predicted total annual apple production at every month.

The following formula shows apple production model. The dependent variable *PY* means domestic annual production of apple. *avg_temp_tot*, *shine_temp_tot*, *rain_amt_tot* are the basic explanatory variables. *weather$diff*, *weather$deghigh*, *weather$deghigh2*, *weather$degrain*' are the introduced explanatory variables for the characteristics of crop.

```
PY = f
{ (  avg_temp_tot, shine_temp_tot, rain_amt_tot ),
  (  weather$diff, weather$deghigh, weather$deghigh2,
          weather$deghigh3,weather$degrain  )
}
```

Figure 3. Apple Production Model

The *weather$diff* means a diurnal temperature range and we use data from March to October at apple prediction model. The *weather$deghigh* is the degree of high temperature that means the difference between average of total maximum temperature and maximum temperature of each month and we use them from May to June. The *weather$deghigh2* is the degree of sunscald that give a negative impact on growth of apple and we use them from July to August. The *weather$degrain* is the degree of rainfall which means difference between total average rainfall and average rainfall of each month and we use them from May to September. all weather data of past months are reflected on the prediction model. For example, when we predict the annual production at March, the model uses average temperature, sunshine of March as well as those from January to February as explanatory variables. So when we predict it at December, all meteorological data of January to December is reflected.

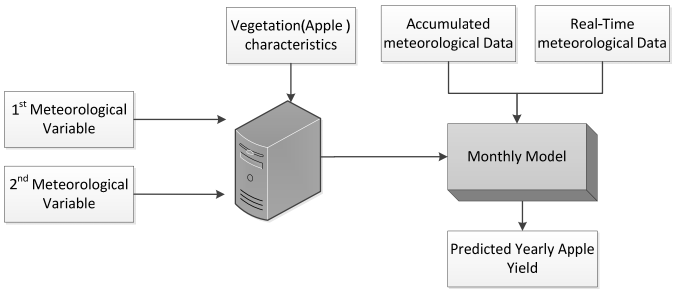

Figure 4. Prediction System Data Flow Diagram

We use R, one of the statistical package to make the model. R provides a variety of statistics functions and *npreg* function contained in library(*np*) is used. *npreg* computes a kernel regression estimate of a one dimensional dependent variable on p-variate explanatory data, given a set of evaluation points, training points (consisting of explanatory data and dependent data), and a bandwidth specification using the method of Li and Racine[1].

C. System Result

The simulation was run on a Lenovo E31 workstation(intel xeon processor 3.4GHz, 8GB DDR). The greater the number of the monthly weather, the more the calculation time exponentially increases. But, after the model is completed, this running time is not important.

TABLE 2. R CODE RUNNING TIME FOR MONTHLY MODELING

Jan	Feb	March	April	May	June
0.0075	0.08 91	0.3668	0.9819	2.5377	2.4479
July	Aug	Sep	Oct	Nov	Dec
8.4280	8.7203	11.6377	18.5103	18.8139	24.7184

(unit : minute)

December model's MAPE is 5.713087e-12 and R-squared is 1. So December model's forecast is high predictive. In summary, January/February model's MAPE is low. Other models from March to December apple production forecast have high values as R-squared value.

TABLE 3. APPLE STATISTICS MODEL ACCURACY

Model sort	Only average weather		Crop characteristics	
Accuracy	MAPE	R-squared	MAPE	R-squared
result_apple_1	0.013988	0.206275	0.013988	0.206275
result_apple_2	0.011706	0.030230	0.013988	0.206275
result_apple_3	0.011223	0.112620	2.81E-08	1
result_apple_4	0.015831	0.090724	9.26E-05	0.999887
result_apple_5	0.009942	0.264819	0.000201	0.998786
result_apple_6	0.009282	0.408570	3.58E-08	1
result_apple_7	0.004792	0.842816	2.50E-06	1.000000
result_apple_8	0.007775	0.566404	4.74E-10	1
result_apple_9	0.004540	0.833044	2.16E-12	1
result_apple_10	0.006191	0.777461	4.57E-10	1
result_apple_11	0.005337	0.771296	1.38E-15	1
result_apple_12	0.011761	0.204779	5.71E-12	1

IV. CONCLUSIONS

This paper includes the research necessary to constructing reliable prediction and information system through convergence between ICT technology and agricultural observation technology. In this research, we cover how to process the number of weather data and how to configure the prediction system. We establish a non-parametric statistical model. We use the meteorological data as the inputs of modelling and generate the secondary meteorological data for reflecting the characteristics of apples. At the future, we will consider more factors and improve the system prediction ability.

REFERENCES

[1] Qi Li and Jeff Racine, "*Predictor Relevance And Extramarital Affairs*," JOURNAL OF APPLIED ECONOMETRICS., 2004.

[2] Sukho Han, Byounghoon Lee, Misung Park, Junho Seung, Hyunseok Yang, Sungchul Shin, "*A Study of Building Crop Yield Forecasting Model considering Meteorological elements*," Korea Rural Economic Institute, Policy Research Report, 2011.

[3] Korea meteorological Administration, www.kma.go.kr.

[4] Yong-Sun Lee, Jong-Jin Kim, S74u-Jeong Noh, "*Analysis of Vegetable Price Fluctuations*," Korea Rural Economic Institute, Policy Research Report, 2012.

[5] "*Market Diseases of Apples, Pears, and Quinces*," Washington State University, http://postharvest.tfrec.wsu.edu.

[6] M. P. Wand & M. C. Jones, "*Kernel Smoothing*," Monographs on Statistics and Applied Probability, Chapman & Hall, 1995.

[7] "*That's Smooth*," http://statistical-research.com/thats-smooth/, October 10, 2013

Haedong Lee received the BS and MS degrees in computer science for Kyungpook National University, Korea, in 1999 and 2001, respectively. In 2001, he joined Electronics and Telecommunication Research Institute, Korea. He is a senior researcher in the Embedded System Research Team. His research interests include convergence of agriculture and ICT, plant image processing, and controlled horticulture.

Aekyung Moon received the MS and PhD degrees in computer engineering from Yeungnam University, Korea, in 1992 and 2000, respectively. In 2000, she joined Electronics and Telecommunication Research Institute, Korea. Currently, she is a senior researcher in the Embedded System Research Team. Her research interests include distributed systems, contextaware middleware, sensor networks, and recommendation systems.