Kobe University Repository : Kernel

PDF issue: 2025-12-05

An Easily Installed Method of the Estimation of Soybean Yield Based on Meteorological Environments with Regression Analysis

Nitta, Akihiro ; Chonan, Yuya ; Hayashi, Satoshi ; Nakamura, Takuji ; Tsuji, Hiroyuki ; Murakami, Noriyuki ; Nishide, Ryo ; Ohkawa, Takenao …

(Citation)
Engineering Proceedings,9(1):26

(Issue Date)
2021-11-30

(Resource Type)
conference paper

(Version)
Version of Record

(Rights)
© 2021 by the authors. Licensee MDPI, Basel, Switzerland.
Creative Commons Attribution 4.0 International license

(URL)
https://hdl.handle.net/20.500.14094/0100481878

engineering proceedings

Proceeding Paper

An Easily Installed Method of the Estimation of Soybean Yield Based on Meteorological Environments with Regression Analysis †

Akihiro Nitta [1,*], Yuya Chonan [2], Satoshi Hayashi [2], Takuji Nakamura [2], Hiroyuki Tsuji [2], Noriyuki Murakami [2], Ryo Nishide [3], Takenao Ohkawa [1] and Seiichi Ozawa [4]

[1] Graduate School of System Informatics, Kobe University, Kobe 657-0013, Japan; ohkawa@kobe-u.ac.jp
[2] Hokkaido Agricultural Research Center, NARO, Sapporo 062-0045, Japan; y.chonan@affrc.go.jp (Y.C.); hayashis@affrc.go.jp (S.H.); takuwan@affrc.go.jp (T.N.); tuzihiro@affrc.go.jp (H.T.); noriyuki@affrc.go.jp (N.M.)
[3] The Center for Data Science Education and Research, Shiga University, Hikone 522-8522, Japan; ryo-nishide@biwako.shiga-u.ac.jp
[4] Graduate School of Engineering, Kobe University, Kobe 657-08501, Japan; ozawasei@kobe-u.ac.jp
* Correspondence: Aki_0811@cs25.scitec.kobe-u.ac.jp; Tel.: +81-78-803-6333

† Presented at the 13th EFITA International Conference, online, 25–26 May 2021.

Abstract: A simple method for estimating soybean yield under ideal environments in Japan is proposed. Several models that simulate soybean yield have been proposed in other countries; however, direct adaptation to Japanese species is difficult in terms of climatic and regional characteristics. In addition, they often require variety-specific information or various environmental information, which is sometimes hard to simulate. Therefore, we attempted to create a simple simulation model with meteorological data as the main input to the model. The proposed model ignores the features that need setting for each cultivated field and is composed of a statistical model instead of a physiological analysis for the sake of brevity. Although the prediction accuracy of the model needs to be improved, we can use it as a decision support system for soybean cultivation because it requires only location information and can be easily introduced by many farmers.

Keywords: regression analysis; data mining; smart farming

check for updates

Citation: Nitta, A.; Chonan, Y.; Hayashi, S.; Nakamura, T.; Tsuji, H.; Murakami, N.; Nishide, R.; Ohkawa, T.; Ozawa, S. An Easily Installed Method of the Estimation of Soybean Yield Based on Meteorological Environments with Regression Analysis. *Eng. Proc.* **2021**, *9*, 26. https://doi.org/10.3390/engproc2021009026

Academic Editors: Dimitrios Aidonis and Aristotelis Christos Tagarakis

Published: 30 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In recent years, due to the decrease in number and aging of farmers, there is a need to improve and renovate agricultural work in Japan. Accurate predictions of the growth of crops lead to appropriate cultivation management and improve the production yield. Crop growth tends to be affected by the surrounding environment. In particular, the yield is the result cumulatively affected by the growing environment over the whole cultivation period [1]. Collecting and recording various detailed data such as plant heights or leaves are indexes which may pave the way to generate a highly accurate model. However, the application of such models is limited only to fields with special equipment or requiring additional work to collect these detailed data. Therefore, predicting growth using information available without special measurement equipment or additional work is important. Meteorological conditions can be obtained from climate satellites and are the dominant environmental factors in cultivation [2].

Soybean is a principal crop in Japan and more sensitive to climatic conditions than rice. To obtain high yields, cultivating the appropriate species in a suitable season is important. Applying growth models is an effective way to improve plans. In other countries, work on developing models for soybean has been conducted, but these models are difficult to apply in Japan due to differences in the varieties and the climate. In this respect, our approach enables regional flexibility by readjusting models. In this study, we aim to quantitatively evaluate the cultivation techniques of soybean farmers in Japan as a part of cultivation management support. We estimate the climatic productivity [1] through a growth model using meteorological data as inputs. By comparing the climatic productivity with the actual yield, we can evaluate the conditions of our field. In this study, the growth model consists of two sub-models. One is the prediction of the developmental progress, and the other is the prediction of seedling formation. By integrating the results of the analysis in each sub-model, climatic productivity is finally determined. The proposed growth model in this study requires only the longitude and latitude for obtaining meteorological data, which makes it easy to install.

2. Materials and Methods

Crop growth is often described as a system that integrates several sub-processes such as developmental progress or dry matter production. In this study, the growth model has two sub-models: the vegetative process, which describes the growth of the plant body; and the maturation process, which describes the formation of the seedlings.

2.1. Vegetative Process

The growth of soybean is the cumulative outcome of quantitative changes in the plant body due to the growing environment. The rate of this quantitative change is the developmental Rate (DVR), and its integration is the developmental index (DVI). In detail, DVI has a relationship with DVR, as detailed in Equation (1), which is called the DVR model [3]. To express the continuous vegetative process of soybeans, Table S1 shows the correspondence of DVI to the growth stage of soybeans defined by Fehr et al. [4].

$$\mathrm{DVR}_i = \sum_{j=1}^{i} \mathrm{DVR}_j \tag{1}$$

Temperature and day length have a significant effect on the developmental phenomena of crops [1]. Temperature and day length can be used to calculate DVR, as shown in Equation (2). Equation (2) is based on the model for rice [5], which divides the periods before and after the flowering stage due to differences in the reactivity of soybean.

$$\mathrm{DVR}_i = \begin{cases} \frac{1}{G_v} \cdot \frac{1-\exp\left(B_L\left(L_j-L_c\right)\right)}{1-\exp\left(-A_T\left(T_j-T_k\right)\right)} & \text{for } 0 \leq \mathrm{DVI}_i < 1 \\ \frac{\alpha}{G_v} \cdot \frac{1-\exp\left(B_L\left(L_j-L_c\right)\right)}{1-\exp\left(-A_T\left(T_j-T_k\right)\right)} + \varepsilon & \text{for } 1 \leq \mathrm{DVI}_i \leq 2 \end{cases} \tag{2}$$

A_T and B_L are coefficients representing the sensitivity to temperature and day length, respectively. T_i and L_i are the respective temperature and day length on day i. T_k is the temperature at which the DVR halves under a certain day length, L_c is the critical day length for growth to proceed, and G_v is the minimum number of days required for the VE-R2 stage. δ is a fixed DVR value that is added over days, and α is the ratio of the effect of weather to the effect of δ. Some parameters need to be optimized. In this study, we used Bayesian optimization [6] with each parameter as an input variable and the total number of growing days as an objective variable.

2.2. Reproductive Process

We describe the estimation of climatic productivity as a prediction of the maturation process. The climatic productivity is calculated in three steps: division of the growing period, characterization of the meteorological data, and construction of a regression model. First, DVI divides the whole cultivation period into three periods: $0 \leq \mathrm{DVI} < 1.0$, $1.0 \leq \mathrm{DVI} < 1.2$ and $1.2 \leq \mathrm{DVI} \leq 2.0$. Soybeans respond differently to the environment at specific growth stages, such as around the flowering stage, and we chose this partitioning method based on previous studies [7]. Then, the following six meteorological elements are obtained as features for each divided period: average temperature, average solar radiation, average precipitation, percentage of days with precipitation, average wind speed, and maximum wind speed. We examined a wide variety of meteorological elements as features. The elements mentioned above were selected after determining multicollinearity by calculating variance inflation factors (VIFs) [8]. There are six types of meteorological elements acquired for three sections; therefore, the total number of input features is eighteen. Finally, these features are used to construct a regression model to obtain the climatic productivity. To suppress overlearning and to reduce dimensionality, LASSO regression is applied as a method to construct regression model. Moreover, we also examined the prediction accuracy in comparison with the conventional method without dividing the growing period. Climatic productivity was estimated in this study; therefore, the fields were assumed to be under ideal conditions, without any growth-inhibiting factors such as insect damage or disease, and with sufficiently fertile soil.

3. Results

The dataset used for the analysis was the cultivation records of 293 fields collected from different locations in Japan. The fields in this dataset show high yield potential, making them ideal for the estimation of climatic productivity. Therefore, these fields were kept free from factors inhibiting growth such as insect damage or disease, and the soil conditions sufficiently were fertile. The weather information was obtained from NARO's Agro-Meteorological Grid Square Data System [9,10], which was developed and operated by the National Agriculture and Food Research Organization (NARO).

In terms of the sub-model for predicting the vegetative process, the RMSE between the estimated and measured growing days was 3.91 days for the interval $0 \leq \text{DVI} < 1$ and 5.45 days for the interval $1 \leq \text{DVI} \leq 2$. This result indicates that the sub-model can sufficiently estimate the growth process of soybean. Furthermore, the results imply the capability for evaluating the cultivation environment based on the growth stage in the sub-model for predicting the maturation process. Focusing on individual plots, heavy rain caused growth inhibition in some fields, which resulted in large errors. The lack of accounting for precipitation in the sub-model led to a large error. Regarding the sub-model for predicting the maturation process, the RMSE between the estimated and measured values was 67.2 g m^{-2}, and the RMSE without dividing the growing period was 76.9 g m^{-2}. There was a significant difference in the results between the divided and undivided growing periods. Thus, dividing the growing period by growth stages is beneficial to evaluate the effects of climatic factors.

Figure S1 is a scatter plot of the estimated and measured values for the applied growth model. Figure S1 shows that the higher (or lower) the yield is, the larger the error. There is normality in the yield of this dataset, with values biased around the mean. The tendency of the regressor to especially reduce the error of data near the mean probably accounts for this result. For this, accurate estimation of off-center values is important. On the other hand, the results were inadequate because the expected error was less than 50 g m^{-2}. The limitation of the input information for the convenience of operation caused this issue. Furthermore, we believe that the analysis ignored the differences among species, which is the reason for the small amount of data in this study, also contributing to the issue. In the future, we will improve the input features and investigate the analysis method, while guaranteeing the simplicity of the operation.

4. Conclusions

In this study, we modeled the growth of soybean, taking into account the effect of meteorological factors, and estimated the climatic productivity. This model requires no special measurement equipment or additional work because we anticipated a practical operating situation. The information necessary for the introduction of the model is meteorological data, which can be obtained if the longitude, latitude, and the dates of the growing period are available. Hence, the proposed model has high installation applicability. The model has two sub-models, one for the vegetative process and the other for the maturation process.

This model accurately predicts the growth of the plant body and estimates the formation of the seedlings. In predicting the vegetative process, the model succeeded in achieving sufficiently plausible results. Although the estimation of the maturation process needs continuous improvement based on the simplicity of operation, the climatic productivity assesses the productivity by comparison with the actual yield.

Supplementary Materials: The following are available online at https://github.com/Aki0811/EFITA2021_ID130, Table S1: Stage of Soybean Development and Corresponding DVI; Figure S1: Scatter plot of estimated and measured yields.

Author Contributions: Conceptualization, T.O.; methodology, A.N.; software, A.N.; validation, A.N., Y.C., S.H., T.N., H.T. and N.M.; formal analysis, A.N.; investigation, A.N., Y.C., S.H., T.N., H.T. and N.M.; resources, Y.C., S.H., T.N., H.T., N.M. and S.O.; data curation, A.N.; writing—original draft preparation, A.N.; writing—review and editing, Y.C., S.H., T.N., H.T., N.M., R.N., T.O. and S.O.; visualization, A.N.; supervision, T.O.; project administration, T.O.; funding acquisition, T.N. and T.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partly funded by Ministry of Agriculture, Forestry and Fisheries: Development of diagnostic methods and countermeasure techniques for overcoming high-yield inhibitory factors.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy.

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

References

1. Horie, T. A Model for Evaluating Climatic Productivity and Water Balance of Irrigated Rice and Its Application. *Southeast Asian Stud.* **1987**, *25*, 62–74.
2. Garner, W.W.; Allard, H.A. Effect of the relative length of day and night and other factors of the environment on growth and reproduction in plant. *J. Agric. Res.* **1920**, *18*, 553–606. [CrossRef]
3. Wit, C.T.; Brouwer, R.; Penning, F.W.T. The simulation of photosynthetic systems. In Proceedings of the International Biological Program/Production Process, Technical Meeting, Trebon, Czech Republic, 14–21 September 1969; pp. 44–70.
4. Fehr, W.R.; Caviness, C.R. "Stage of soybean development", Agriculture and Home Economic Experiment Station. *Spec. Rep.* **1977**, *80*, 1–12.
5. Horie, T.; Nakagawa, H. Modelling and Prediction of Developmental Process in Rice: 1. Structure and method of parameter estimation of a model for simulating developmental process toward heading. *Jpn. J. Crop. Sci.* **1990**, *59*, 687–695. [CrossRef]
6. Shahriari, B.; Swersky, K.; Wang, Z.; Adams, R.P.; Freitas, N.D. Taking the Human out of the Loop: A Review of Bayesian Optimization. *Proc. IEEE* **2016**, *104*, 148–175. [CrossRef]
7. Nakano, S.; Kumagai, E.; Shimada, S.; Sameshima, R.; Ohno, H.; Homma, K.; Shiraiwa, T. Modeling of Phenological Development Stage and Impact of Elevated Air Temperature on the Phenological Development of Soybean Cultivars in Japan. *Jpn. J. Crop Sci.* **2015**, *84*, 408–417. [CrossRef]
8. Marquardt, D.W. Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation. *Technometrics* **1970**, *12*, 605–607. [CrossRef]
9. Ohno, H.; Sasaki, K.; Ohara, G.; Nakazono, K. Development of grid square air temperature and precipitation data compiled from observed, forecasted, and climatic normal data. *Clim. Biosph.* **2016**, *16*, 71–79. [CrossRef]
10. Kominami, Y.; Sasaki, K.; Ohno, H. *User's Manual for the Agro-Meteorological Grid Square Data*, 4th ed.; NARO: Sapporo, Japan, 2019; p. 67.