Prediction of Yield Fluctuations Caused by Extreme Weather in Greenhouse Tomato Cultivation

Reo Kamiyama[1], Ken-ichi Minamino[1]

[1]*Graduate School of Software and Information Science, Iwate Prefectural University, 152-52 Sugo, Takizawa-shi, Iwate, 0200693, Japan*

Abstract

In recent years, the effects of climate change have become more pronounced, and extreme weather events have significantly affected the agricultural sector. In this study, we aimed to analyze the impact of record-high temperatures between June to August of 2023 and 2024 on the yield of greenhouse tomato cultivation in Iwate Prefecture, Japan. Random Forest regression model was used to analyze the yield prior to the occurrence of exceptional fluctuations and to forecast such fluctuations based on the deviations from actual yields. The results revealed that the method was effective in forecasting the signs of yield fluctuation. In the future, this method can be applied to risk prediction and pest control measures in greenhouse environments.

Keywords

Smart Agriculture, Tomato Cultivation, Machine Learning, Yield Fluctuations, Extreme Weather

1. Introduction

Between 2023 and 2024, both the frequency and severity of extreme weather events increased, driven by the ongoing progression of global warming. In agriculture, production management based on empirical rules (such as reliance on conventional weather patterns) is no longer sufficient to cope with this situation, making it increasingly difficult to ensure stable yields [1]. Even under seemingly controlled environments, such as greenhouses, changes in external temperature and solar radiation affect the internal climate, which result in yield fluctuations [2].

In Japan, crop growth is increasingly affected by several factors, such as heat stress during summer, insufficient sunlight caused by prolonged rainfall during the rainy season, and the emergence of new pests and diseases. Exceptional fluctuations in yield and quality have been reported, even in greenhouse tomato cultivation, where environmental conditions are typically controlled [3, 4]. In recent years, yield prediction models that are developed using environmental and yield data collected before the frequent occurrence of extreme weather events have become less accurate. During the summer of 2024, prolonged heatwaves made it difficult to control the internal temperature of greenhouses, leading to serious quality issues, such as poor fruit set and decreased sugar content. Furthermore, in recent years, complex and unprecedented factors have caused abnormal plant development and unanticipated reductions in yield; these factors include

The 7th International Symposium on Advanced Technologies and Applications in the Internet of Things (ATAIT 2025), September 10-11, 2025, Kusatsu, Shiga, Japan

s231w006@s.iwate-pu.ac.jp (R. Kamiyama); minamino@iwate-pu.ac.jp (K. Minamino)

the simultaneous outbreak of new pests (e.g., tobacco whitefly) and diseases (e.g., powdery mildew and late blight) that have adapted to warming climates [5, 6]. These fluctuations not only destabilize the income of producers but also pose significant challenges in securing a stable food supply at the regional and national levels. Existing studies have analyzed yield fluctuations using statistical analyses such as multiple regression and machine learning models such as regression [7, 8, 9]. However, few studies have been conducted to predict abnormal yield fluctuations caused by global warming.

In this study, we aimed to investigate the nature of exceptional yield fluctuations and develop a prediction method. Using environmental and yield data collected from IoT sensors at the Iwate Wakae Farm, Inc [10] (Morioka City, Iwate Prefecture, Japan), we aimed to develop a yield prediction model based on conditions observed before the frequent occurrence of extreme weather events. We then quantitatively aimed to evaluate the discrepancies between the predicted and actual yields and identify early signs and timing of anomalies, with the goal of developing a method to detect yield fluctuations in advance. Iwate Wakae Farm, Inc is a large-scale facility equipped with a Venlo-type greenhouse that employs a long-term multiphase cultivation system. The greenhouse covered an area of 2,160 m^2, and a hydroponic system using rockwool as the growing medium has been employed in the greenhouse. The facility is equipped with several environmental sensors, including solar radiation, temperature, humidity, and CO_2, allowing for high-precision monitoring of the cultivation environment. The comprehensive data collection could significantly enhance the reliability of the analyses conducted in this study.

Figure 1: Iwate Wakae Farm, Inc.

2. Yield prediction model before the frequent occurrence of extreme weather events (Normal model)

2.1. Explanatory and Objective Variables

The primary explanatory variables used in this study were solar radiation and average daily temperature. Over the past four weeks, solar radiation had accumulated, which is closely associated with photosynthetic activity and fruit enlargement. Previous studies have demonstrated that this time span is highly correlated with the progression of growth stages from fruit set to maturation (Figure 2) [7]. We aimed to capture the distribution of the daily average temperatures rather than relying on simple mean values. Therefore, we introduced a histogram-based representation (hereafter referred to as the H function), which divided the daily average temperatures over the past eight weeks into 1 °C bins and used the relative frequency of each bin as a feature variable (Figure 3) [8]. Weekly cumulative yield for the following week was set as the objective variable (Figure 4). Daily yields were estimated using a three-week moving average to account for practical conditions, such as scheduled harvest breaks (e.g., once a week), which lead to periodic gaps in yield records. This smoothing approach facilitated the extraction of underlying yield trends. Previous studies have confirmed that the influence of harvest breaks is typically confined to period spanning one week before and one week after the break [9].

Cumulative period (weeks)	1	2	3	4	5	6	7	8	9
Correlation coefficient	0.46	0.57	0.65	0.66	0.65	0.66	0.67	0.66	0.63
Frequency		Increase		Max			Small Fluctuations		

Figure 2: Correlation between cumulative solar radiation and weekly yield.

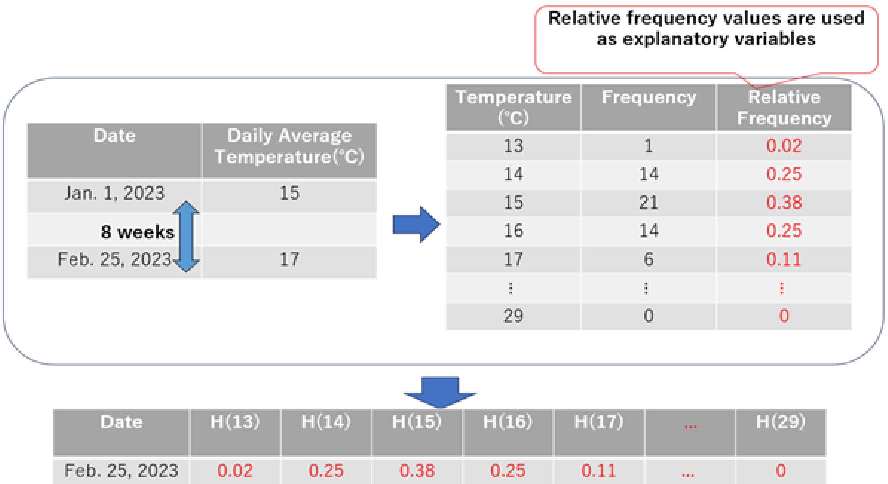

Figure 3: Example output of H function on Feb 25, 2023.

2.2. Model Building and Evaluation

The yield prediction model was constructed using a random forest regression algorithm, which is well-suited for capturing nonlinear relationships and offers strong robustness against external disturbances [11]. A normal model (model18–22) was trained on data obtained between December 2018 and August 2022, the period before the frequent occurrence of extreme weather events. For comparison, a model with extended period (model18–23) was also developed using data up to August 2023, including the potentially abnormal year 2023. The test datasets consisted of data from the year following each training period: December 2022 to August 2023 for model18–22,

Harvest date	Yield（kg）	Weekly cumulative yield for the following week（kg）
Jan. 11, 2023	71.1	418.88
Jan. 12, 2023	66.2	389.22
Jan. 13, 2023	68.4	362.62
Jan. 14, 2023	74.76	329.32
Jan. 15, 2023	-	329.32
Jan. 16, 2023	89.9	289.44
Jan. 17, 2023	71.2	258.74
Jan. 18, 2023	48.42	233.08

Figure 4: Objective Variable (weekly cumulative yield).

and December 2023 to August 2024 for model18–23. Model performance was evaluated using the coefficient of determination (R^2) and root mean square error (RMSE).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{1}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{2}$$

Where n is the number of observations, y_i is the i-th observed value, $\hat{y}_i$ is the corresponding predicted value, and $\bar{y}$ is the mean of the observed values. Table 1 presents the training results of the developed prediction models. Significant differences in the coefficient of determination and RMSE were observed between model18-22 and model18-23. This indicates that in recent years, when weather patterns have become increasingly non-stationary due to extreme events, accurate yield prediction models must be constructed by selecting long-term training datasets that comprehensively capture exceptional weather conditions and the resulting growth abnormalities [12]. However, the accumulation of extensive datasets during transitional periods remains challenging. Therefore, this study proposed a method for continuously monitoring deviations from the predictions generated by a normal model. This approach could enable the early detection of anomaly signs and associated risks [12].

Table 1
Model Performance Metrics

	R^2	RMSE
model_{18-22}	0.86	247.2(Kg)
model_{18-23}	0.14	479.3(Kg)

3. Analysis of exceptional yield fluctuations

Figures 5 and 6 depict the trends in the observed and predicted yields for 2023 and 2024, respectively. These results revealed that two periods (February–April 2023 and April–July 2024) were characterized by notably large yield. The yield between April–July 2024 was primarily influenced by the outbreak of pests, such as the tobacco whitefly, as confirmed by reports and observational records from farmers. In this section, we focus on analyzing the factors contributing to the yield fluctuations between February and April 2023. Furthermore, we predict and analyze seasonal outbreaks of pests that could cause exceptional yield fluctuations. Interviews with growers revealed that the decline in yield observed around February 2023 (Figure 5) was primarily attributed to reduced solar radiation. To verify this, we compared the accumulated solar radiation over the four months immediately preceding this period, particularly from December to January, with that of other years (Figure 7). Notably, yield between 2019–2020 were excluded owing to missing data. The analysis showed that the accumulated solar radiation from November to January in the 2022–2023 and 2023–2024 seasons was approximately 60–90 MJ/m² lower than the average from 2018 to 2022 (492.5 MJ/m²), indicating a clear downward trend. Such a reduction in solar radiation likely had a significant impact on fruit set and enlargement, resulting in a marked decrease in yield. These findings suggest that continuous monitoring of deviations in solar radiation from typical levels can provide valuable insights for predicting yield fluctuations at least one to two months in advance.

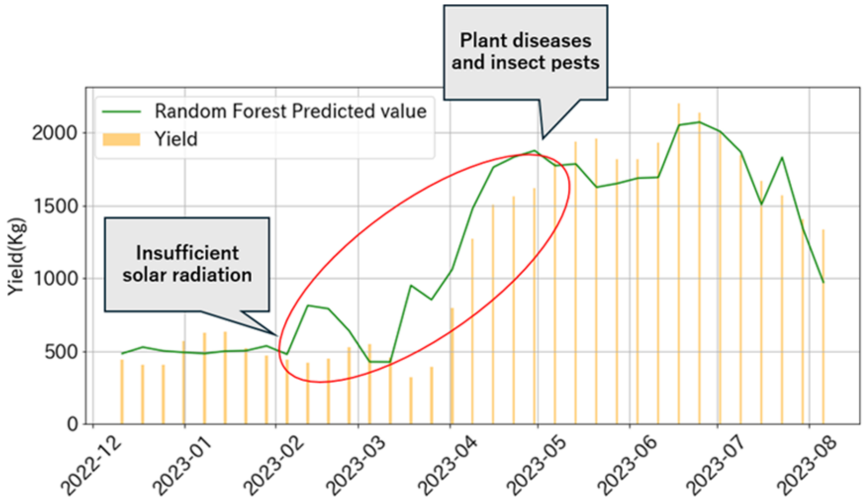

Figure 5: 2023 Yield Fluctuation Analysis

Outbreaks of pest infestations have traditionally been limited to greenhouse whiteflies [13]. This pest is relatively cold-tolerant, easy to manage, and generally does not cause severe damage. However, in recent years, a new pest species, the tobacco whitefly [14] has emerged; this species is highly tolerant to high temperatures and the damage it causes has been rapidly increasing. Tobacco whiteflies are difficult to control using conventional pest management strategies and become particularly active in warmer environments, leading to significant effects on yield

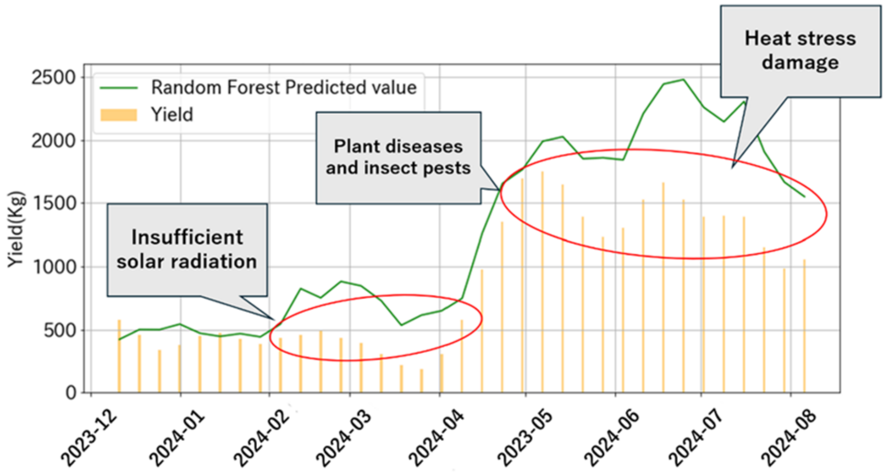

Figure 6: 2024 Yield Fluctuation Analysis

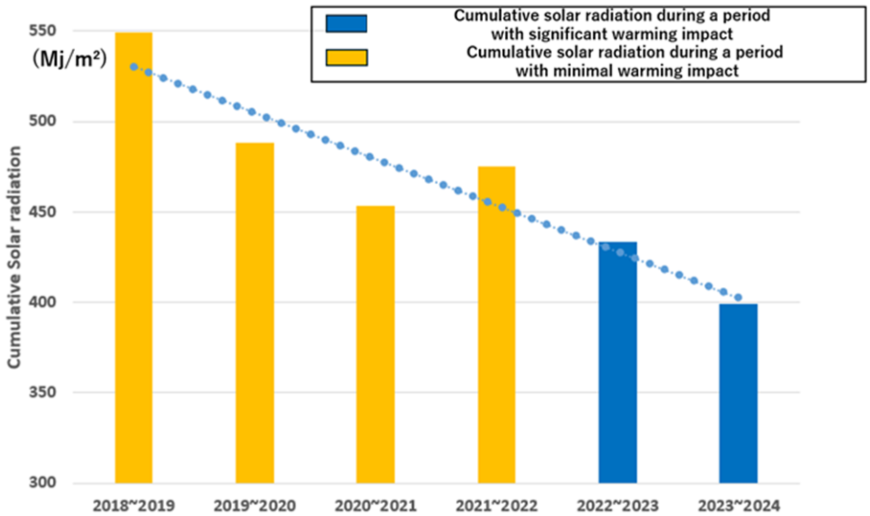

Figure 7: Changes in accumulated solar radiation from November to January in recent years

during hot periods. Additionally, disease damage caused by pathogens, such as sooty mold and powdery mildew [15, 16], has also been reported. The simultaneous occurrence of pests and diseases imposes substantial stress on plant growth, leading to significant yield reduction. Whitefly infestations in the facility reached their first peaks on May 11, 2023, and April 13, 2024 for 2022–2023 and 2023–2024 periods, respectively (Figure 8). Using a multiple regression model [17] combined with the effective accumulated temperature [18], we developed a prediction model and determined that the period from egg to larval stage occurred around May 1, 2023, for the former peak and around April 1, 2024, for the latter. They were used as the explanatory variables: temperature, humidity, the number of pests in the previous week, Leaf Area Index,

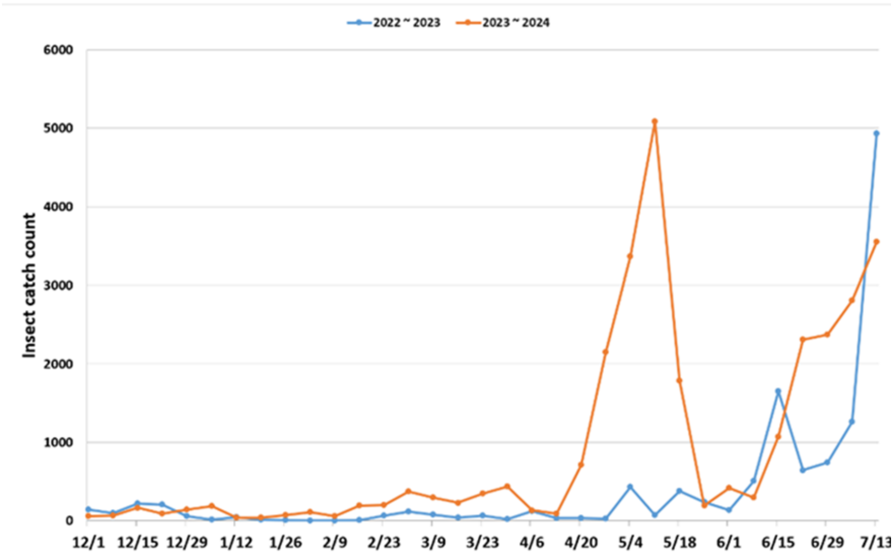

Figure 8: Number of whiteflies caught in the west greenhouse

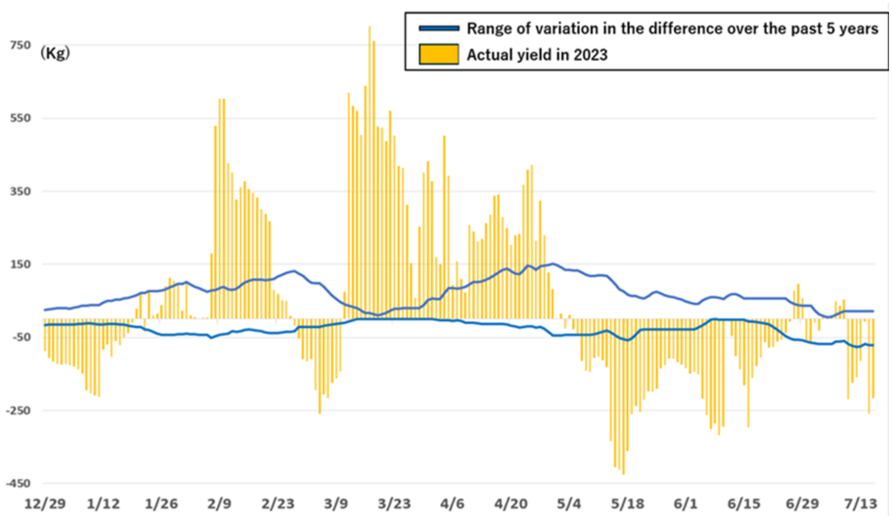

Figure 9: Difference between the predicted value obtained using the normal model and actual yield

and yield. The infestation in 2023 is primarily caused by the greenhouse whitefly, which is effectively controlled by conventional pest management methods. In contrast, the 2024 outbreak was mainly due to the tobacco whitefly, which resisted conventional control measures, leading to continuous infestation spanning from the first peak to the second peak. Figure 9 shows a graph of the differences between the predicted value obtained using the normal model and actual yield in 2022-2023. The blue lines in Figure 9 represent the range of differences for the past five years from 2018 to 2022 (maximum: +151.4 kg, minimum: -76.5 kg). In other words, results that exceed or fall below these blue lines indicate irregular yield fluctuations compared with typical years. Because the prediction model forecasts one week in advance, it enables proactive responses. Moreover, combining this early warning with predictions of the underlying causes of the exceptional yield fluctuations described earlier, rapid and effective countermeasures have become possible.

4. Summary

In this study, we developed a yield prediction model for facility-grown tomatoes and analyzed yield fluctuations by examining the differences between the predicted and actual yields. We also explored the potential to predict the early signs of such variability. The analysis revealed that irregular yield fluctuations were mainly caused by insufficient solar radiation, pests, and diseases. Future work will focus on developing a model that can forecast the likelihood of exceptional yield fluctuations and estimate the extent of yield reduction in advance, thereby enabling effective countermeasures at the production site.

Acknowledgments

We would like to express our deepest gratitude to Iwate Wakae Farm, Inc. for their generous cooperation in this research. They willingly provided us with valuable in-greenhouse environmental and yield data and kindly accommodated our numerous requests for meetings and data collection. Their support was essential for the successful completion of this study.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] Ministry of Agriculture, Forestry and Fisheries (MAFF), Efforts to address global warming in the agriculture, forestry and fisheries sector, https://www.maff.go.jp/j/kanbo/kankyo/seisaku/attach/pdf/index-87.pdf, 2024. Accessed: 2025-08-25.

[2] Japan Meteorological Agency (JMA), Weather summary for summer 2024 (june–august), https://www.data.jma.go.jp/cpd/longfcst/seasonal/202408/202408s.html, 2024. Accessed: 2025-08-25.

[3] K. Nishimura, H. Okamoto, Current status and challenges of climate change adaptation information platform, Journal of Agricultural Meteorology 75 (2019) 1–8.

[4] Y. Takahashi, T. Sakata, Analysis of the relationship between meteorological factors and growth abnormalities in tomato cultivation, Environment Control in Biology 29 (2020) 105–112.

[5] K. Nakajima, Occurrence and impact of the tobacco whitefly in protected horticulture, Journal of the Japanese Society for Horticultural Science (2023).

[6] E. Hashimoto, Y. Honda, Pest and disease management and its effectiveness in tomato cultivation during high-temperature periods, Agriculture and Environment 45 (2021) 33–40.
[7] K. Oibayashi, K. Minamino, Improving agricultural management using a yield prediction model for greenhouse tomato cultivation, in: Proceedings of the 85th National Convention of the Information Processing Society of Japan, 2023, pp. 523–524.
[8] T. Endo, S. Warisawa, T. Yamamoto, I. Yamada, Study on prediction method (calculation model) of harvest date and yield in low-truss dense planting of tomato, in: Proceedings of the 2017 Annual Meeting of the Japanese Society of Agricultural, Biological, and Environmental Engineering, volume A03, 2017, pp. 14–15.
[9] K. Hisekida, H. Nishina, Study on improving productivity in large-scale tomato greenhouses: Harvest prediction of tomatoes based on cumulative solar radiation, Japanese Society of Agricultural, Biological and Environmental Engineers and Scientists 19 (2007) 11–18.
[10] Iwate wakae farm, inc, https://www.facebook.com/wakaefarm/?locale=ja_JP, 2025. Accessed: 2025-08-25.
[11] L. Breiman, Random forests, Machine learning 45 (2001) 5–32.
[12] R. Kamiyama, K.-i. Minamino, Analysis of exceptional yield variations caused by extreme weather in greenhouse tomato cultivation, 2025. Graduate School of Software and information Science, Iwate Prefectural University.
[13] National Research and Development Agency, National Institute for Environmental Studies, Invasive species of japan, https://www.nies.go.jp/biodiversity/invasive/DB/detail/60330e.html, 2025. Accessed: 2025-08-25.
[14] National Research and Development Agency, National Institute for Environmental Studies, Invasive species of japan, https://www.nies.go.jp/biodiversity/invasive/DB/detail/60320e.html, 2025. Accessed: 2025-08-25.
[15] Institute of Vegetable and Floriculture Science, National Agriculture and Food Research organization, Sooty mold information, https://www.naro.go.jp/laboratory/nivfs/kakibyo/plant_search/ha/botan/post_567.html, 2025. Accessed: 2025-08-25.
[16] Institute of Vegetable and Floriculture Science, National Agriculture and Food Research organization, Powdery mildew information, https://www.naro.go.jp/laboratory/nivfs/kakibyo/plant_search/ha/hanamizuki/post_132.html, 2025. Accessed: 2025-08-25.
[17] C. R. Rao, Linear statistical inference and its applications, 2nd ed., John Wiley & Sons, 1973.
[18] Y. Hirano, R. Kamiyama, K. Minamino, Development of a whitefly occurrence prediction model using machine learning in greenhouse tomato cultivation, in: Proceedings of the 87th The National Convention of the Information Processing Society of Japan, 6ZL-03, 2025.