26th International Congress on Modelling and Simulation, Adelaide, SA, Australia, 30 Nov–4 Dec 2025
mssanz.org.au/modsim2025

Predicting lettuce biomass under six climate zones using machine learning in a digital twin system

Shu Liang [a,b,c], **Mst Irin Parvin** [a,b], **Nguyen Van Duc Long** [a,b], **Marc Escribà-Gelonch** [d], **Volker Hessel** [a,b,c]

[a] School of Chemical Engineering, The University of Adelaide, Adelaide 5005, Australia, [b] ARC Centre of Excellence Plants for Space (P4S), The University of Adelaide, Australia, [c] Andy Thomas Centre for Space Resources (ATCSR), The University of Adelaide, Adelaide 5005, Australia, [d] Higher Polytechnic Engineering School, University of Lleida, Lleida, Spain
Email: shu.liang@adelaide.edu.au

Abstract: This study simulated the (extreme) environments of six climate zones in a growth chamber, expanding the application boundaries of digital twins in plant growth modeling, especially in coping with future climate change ("Climate 2050") and space plant cultivation, providing new perspectives and technical paths. The study used three machine learning models, linear regression, random forest and boosted decision tree, to predict the biomass of lettuce under 4 scenarios. The random forest model performed best in all scenarios, with a coefficient of determination R^2 close to 0.99 and a mean absolute percentage error (MAPE) of 6%. By introducing correction factors, the prediction accuracy of the model was further improved for the scenarios of no watering, fertilization and soil-sand mixture. Combining real-time climate input with machine learning predictions, this study demonstrated the potential for the widespread application of the digital twin framework in future climate-resilient agriculture and space plant cultivation.

Keywords: *Digital twin, Machine learning, Space agriculture*

1. INTRODUCTION

As global climate change intensifies, agricultural production is facing unprecedented environmental challenges, especially under multiple pressures such as extreme temperatures, frequent droughts and soil degradation, the yield and quality of crops are becoming increasingly unstable. The Intergovernmental Panel on Climate Change (IPCC) reports that the global mean temperature has increased by approximately 0.8-1.3°C relative to the pre-industrial baseline of 1850–2019. However, to more accurately simulate and predict the operating status of future agricultural systems, digital twin (DT) technology is gradually becoming a key tool to promote the development of smart agriculture. This technology can provide high-timeliness and high-precision decision support for agricultural management by connecting real physical systems with virtual models in real time. At the same time, with the development of artificial intelligence, embedding machine learning models into digital twin systems has become an important way to improve crop growth prediction capabilities.

2. METHODOLOGY AND RESULTS

This study collected real-time meteorological data from six cities in typical climate zones around the world (Lleida, Adelaide, Paris, San Luis, Singapore and Fairbanks), and replicated them in controlled growth chambers in turn, simulating the climate of a new city every week, achieving a "virtual switch" of the environment, and reproducing the extreme growth environment under the future "Climate 2050". We used three machine learning algorithms, including linear regression, random forest and gradient boosting tree, in this system to build and compare biomass prediction models under different scenarios.

We designed 4 experimental scenarios (normal conditions, no watering, fertilization, and soil-sand mixture) to predict the growth performance of lettuce and compared the prediction effects of three machine learning models: linear regression, random forest, and boosted decision tree. Figure 1 shows that linear regression significantly overestimates the biomass under normal and fertilization conditions, indicating that its underfitting problem is serious; while random forest performs most stably in each scenario, and the predicted value is closest to the measured value, with R^2 reaching 0.984 and MAPE of 6.3%. Weekly biomass variation was modest across the 4 climate zones, with the exception of a significantly faster rate of biomass growth during the spring in arctic Fairbanks.

In addition, we improve the model predictions for the no watering, fertilization, and soil-sand mixture scenarios by adding correction factors. The optimal factor was selected by calculating MAPE and relative error. The results showed that the average absolute deviation MAPE was around 6% under all scenarios, indicating that the model prediction deviation was small and generally reliable. In different scenarios, the correction factor was selected to avoid the model overestimating or underestimating plant growth. By comparison, the model predictions are best when the correction factors are 0.12, 1.6, and 0.7. respectively, Table 1 (Liang et al., 2025).

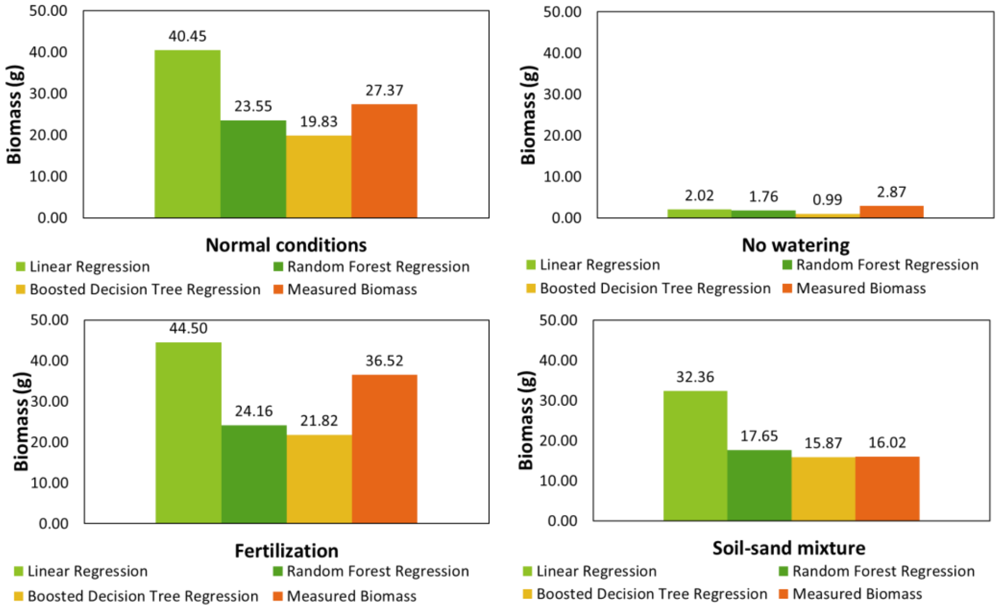

Figure 1 Comparison between the predicted and measured values of the three algorithms under different scenarios. (Source: Liang et al., 2025)

Table 1 Model calibration results under different scenarios using correction factors to improve prediction accuracy

Scenario	No.	Correction factor	MAPE/%	Predicted Value	Relative Error/%
No watering	1	0.10	6.11%	2.24	21.95
	2	0.12	6.36%	2.68	6.62
	3	0.14	6.17%	3.09	7.67
Fertilisation	2	1.5	6.03%	33.44	8.43
	2	1.6	6.05%	35.75	2.10
	3	1.7	6.07%	38.76	6.13
Soil-sand mixture	1	0.7	6.03%	15.44	3.62
	2	0.8	6.11%	17.88	11.61
	3	0.9	5.99%	19.77	23.41

3. CONCLUSIONS AND OUTLOOK

This study introduced the "Climate 2050" scenario and simulated the extreme environments of six climate zones in the growth chamber, expanding the application boundaries of digital twins in plant growth modeling, especially in coping with future climate change and space plant cultivation, providing new perspectives and technical paths. The system accurately predicted lettuce biomass ($R^2 \approx 0.99$, MAPE ≈ 6%) under multiple scenarios. This study not only improves the prediction capability of biomass by machine learning in digital twin systems but also lays the foundation for the image-based IoT digital twin loop that we subsequently developed. The system collects lettuce images through a camera, combines image binarization, CNN 3D deconvolution model, and multi-agent generation system (MAGS) to achieve real-time recognition and intelligent decision-making control of plant growth status. Although this research focuses on biomass prediction driven by climate variables, related image recognition methods will be further integrated in subsequent studies to achieve more comprehensive plant growth modelling.

ACKNOWLEDGEMENTS

The authors would like to thank the Australian Research Council Centre of Excellence ‘Plants for Space’ (P4S) (grant number: CE230100015), and the members of the Department of Agricultural and Forestry Sciences and Engineering of the University of Lleida for their financial, logistical and material support.

REFERENCES

Yuan, X., Li, S., Chen, J., Yu, H., Yang, T., Wang, C., ... & Ao, X. (2024). Impacts of global climate change on agricultural production: a comprehensive review. *Agronomy*, *14*(7), 1360.

Liang, S., Nguyen, V. D. L., Parvin, M. I., Escribà-Gelonch, M., Lantin, S., & Hessel, V. (2025). Digital Twin for Lettuce Growth in Six Climate Zones (‘Climate 2050’). *Computers and Electronics in Agriculture.* Submitted for publication.

