Skoltech
Skolkovo Institute of Science and Technology

Jury Member Report – Doctor of Philosophy thesis.

Name of Candidate: Dmitry Shadrin

PhD Program: Computational and Data Science and Engineering

Title of Thesis: Data-driven modeling of plant growth dynamics in controlled environments

Supervisor: Professor Maxim Fedorov

Name of the Reviewer: Nikolay V. Brilliantov

I confirm the absence of any conflict of interest (Alternatively, Reviewer can formulate a possible conflict)	**Signature:** **Date: 19-09-2020**

The purpose of this report is to obtain an independent review from the members of PhD defense Jury before the thesis defense. The members of PhD defense Jury are asked to submit signed copy of the report at least 30 days prior the thesis defense. The Reviewers are asked to bring a copy of the completed report to the thesis defense and to discuss the contents of each report with each other before the thesis defense.

If the reviewers have any queries about the thesis which they wish to raise in advance, please contact the Chair of the Jury.

Reviewer's Report

The thesis by Dmitry Shadrin covers a wide range of problems of the data science application addressed at the enhancement of precision agriculture. Namely, the focus of the research is on the development and implementation of data-driven approaches for robust and accurate assessment and modeling of plant growth dynamics in controlled environments. The topic of the dissertation is high relevant for scientific community, making impact in this research direction could help in better understanding of such complex natural system as plants using data-driven approaches. The actual content of the dissertation fully corresponds to its topic. The overall quality of the dissertation is high, it is well structured and meets all the standards of PhD thesis writing. The exposition of the material is rather clear.

The author succeeded in solving the main problem of plant growth assessment and modeling discussed in his dissertation by the implementation of the state-of-the art AI and modeling technologies. As usual for the AI technology, he widely used Convolutional Neural Networks for performing computer vision tasks such as segmentation and Recurrent Neural Networks for building the “black-box” models that allow prediction of the plant growth dynamics. The results obtained by the implementation of AI are novel and have a considerable potential for practical application.

The thesis traditionally starts with introduction and literature review of the background. In the Introduction, the author addresses the main topics of the thesis, objectives and own contribution. A valuable part of the Introduction is the quantitative analyses of the specific publication activity on the topic, which demonstrates its growing importance. In this chapter he also performs a comprehensive review on the artificial growth systems, sensor systems in precision agriculture, imaging technologies for plant phenotype, deep learning methods for growth dynamics prediction, hybrid models for growth dynamics prediction, approaches for seeds germination monitoring, computer vision methods for diseases detection. The reviewed topics are in logical correspondence with the following conducted research.

The Second Chapter of the thesis is devoted to the “bottom-up” modeling. Here, the author performs numerical simulations and shows the advantages and limitations of the differential equations based modeling of the plant growth dynamics. In this Chapter it was clearly shown that results of “bottom-up” modeling of plant growth is highly sensitive to parameters used in the differential-equation and lots of assumptions should be done even for the simplest models. In other words, although the pure “bottom-up” modeling is good for perform theoretical investigations, it is not very suitable for practical applications. Therefore, the author changes the approach and apply the hybrid and data-driven methods.

The Third Chapter focuses on the hybrid methods. The main idea exploited in this Chapter is to extract automatically the features that describe growth dynamics, using data-driven methods and then to adapt and apply Dynamic Mode Decomposition (DMD), Kalman filtering or 2D/3D imaging coupled with non-linear least-square for growth dynamics prediction. The author shows the possibility and benefits in application of abovementioned methods for modeling the plant growth dynamics. Noteworthy, the application of these methods to the plant growth dynamics is novel. For instance, while the DMD has been never applied the for plant growth modeling, the thesis demonstrates that it possesses a high application potential. It was suggested to use the features derived from the set of differential equations as a state vector and then to use the DMD approach to perform system identification. The extended Kalman filter approach has been applied for the plant growth dynamics prediction, supplemented by the computer vision (CV) techniques for the feature extraction (e.g. the leaves area). This method showed a high computational efficiency, without scarifying an accuracy, as compared to the standard non-linear least squares. Merging 2D/3D approaches allowed the plant biomass prediction. It should be noticed that the author has elaborated all the proposed by him methods on the own experimental data. He creates several experimental setups for growing plants in artificial controlled conditions. These setups included comprehensive sensor systems allowing to collect unique datasets that consist of data that describe environmental parameters and plant growth dynamics (sequences of images). He shared all obtained datasets with the research community for testing their own developments.

The Fourth Chapter addresses the data-driven enhancement for plant growth modeling in controlled environments. In this Chapter the machine learning (ML) techniques are widely used. In particular, such techniques, as the recurrent neural networks and convolutional neural networks for the creation of "black-box" models. Convolutional neural networks (CNN) for segmentation of plants allow the leaves area calculation. Additionally, CNNs are used for the instantaneous segmentation allowing a tracking of an each leaf the growth. Recurrent neural networks (RNN) are used as the basis for the prediction of growth dynamics. One of the important features of Dmitry's research is that he implemented end-to-end solutions. This opens wide possibilities for a direct industrial application of obtained in the thesis results. The practical usefulness has been proved in several experiments. In the first one, he successfully developed and tested the CV system equipped with RNN algorithm for growth dynamics prediction. The accuracy and robustness of the predictions were rather high. This proved the concept of the possibility to model the plant growth dynamics using only data as an input. The second one was the industrial experiment in the greenhouse where developed the CV and ML system allowed to assess and predict plant growth dynamics in accurate and automatic way. Moreover, the possibility of the biomass prediction at the industrial scale has been demonstrated. Hence the author research bridges the laboratory investigations and industrial applications. In particular, he proposed and created systems of automatic assessment of the seed germination rate; the modern methods of CV are widely used here.

The sustainability of plant growth is a very important problem and the author conducted experiments for obtaining Hyper-spectral data on diseases development on different stages. Here he proposes his own approach for finding optimal wavebands for detection diseases on the early stages. These results also find applications in the industrial systems for early disease detection. All the above methods are highly relevant in current state-of-the art research in precision agriculture.

The last part of the thesis is aimed at the assessment of the parameters that influence the plant growth dynamics using machine learning techniques. Current approaches are not able to accurately assess the set of factors that have dramatical effects on growth dynamics. The first problem that is solved in the thesis is the soil phytotoxicity assessment and derivation of the main factors that influence the phytotoxicity and biological response. It was shown that using ML techniques such as NNs and SVR it is possible to do the accurate predictions of these effects. Another problem that regularly appears is modelling of a spatial distribution of environmental parameters. It the thesis it was proposed to use the advanced Gaussian process regression technique with automatic kernel structure selection. This is a recently appeared technique and its application for solving the particular problem is novel. It was shown that the proposed approach outperformed existing standard techniques. This, in turn, allow to enhance the quality, robustness and automatization of the modeling of spatial distribution of environmental parameters. Overall, this study gives the possibility to include missing important parameters into existing models and simulators making them more effective in plant growth modeling.

The material of the thesis is based on 12 research papers, published exclusively in Q1/Q2 journals or in recognizable conferences mainly in IEEE organization.

To conclude, I highly esteem the presented work, which combines both theoretical and experimental components, where the most modern methods have been applied and developed. I also wish to stress that the author succeeded to produce very practical results, based at the most fundamental basis. A considerable part of the thesis results may be directly deployed without further elaboration.

In spite of my very positive attitude, I would like to mention some drawbacks of the thesis, which mainly refer to the exposition of the material:

1) Is it justified to apply Verhulst model to the leaves growth? What I mean is that this model assumes some maximal amount of the growing substance, while I doubt that there are any limitations (except for the seasonal changes) for the leaves growth. The author should comment on this.
2) Integrating Eq. (3.2) it is assumed that the parameters are constant. At the same time in Figs. 3.11 and 3.12 these parameters are shown as functions of time. It should be stressed in the text (or the figure captures) that these quantities are not the true one, but just a current estimates. In the present version it is not clearly stated.
3) Eq. (3.18) does not make much sense. Please re-write this piece of the text the exposition of the material is very awkward here.
4) Are experiments in Chapter 4 the same as in the previous Chapters? Please be more specific.
5) There are many undefined variables and awkward expressions:
    - For instance, h (width?) and C (nutrients concentration?), (page 54) are not defined,
    - “v” is mixed with “u” in page 54
    - du is “dimensionless unit”(page 56) – does not make sense
    - incomplete figure capture for Fig. 3.9
    - wrong sentence after Eq. (3.2)

- Eq. (3.19) is to be better explained. The expression "water potential" is very ambiguous -- is the chemical potential meant? "n" and "w" there are not explained -- neither in the text nor in the Table 3.3 below.

Provisional Recommendation

x *recommend that the candidate should defend the thesis by means of a formal thesis defense*

☐ ~~*I recommend that the candidate should defend the thesis by means of a formal thesis defense only after appropriate changes would be introduced in candidate's thesis according to the recommendations of the present report*~~

☐ ~~*The thesis is not acceptable and I recommend that the candidate be exempt from the formal thesis defense*~~