



HAL
open science

Integrating Predictive Process Monitoring Techniques in Smart Agriculture

Simona Fioretto, Dino Ienco, Roberto Interdonato, Elio Masciari

► **To cite this version:**

Simona Fioretto, Dino Ienco, Roberto Interdonato, Elio Masciari. Integrating Predictive Process Monitoring Techniques in Smart Agriculture. ISMIS 2024 - 27th International symposium on methodologies for intelligent systems, Jun 2024, Poitiers, France. pp.306-313, 10.1007/978-3-031-62700-2_27. hal-04619543

HAL Id: hal-04619543

<https://hal.science/hal-04619543v1>

Submitted on 12 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Integrating Predictive Process Monitoring Techniques in Smart Agriculture

Simona Fioretto³[0009-0006-8700-8188], Dino Ienco²[0000-0002-8736-3132],
Roberto Interdonato¹[0000-0002-0536-6277], and Elio
Masciari³[0000-0002-1778-5321]

¹ CIRAD, INRIA, Univ. Montpellier, France roberto.interdonato@cirad.fr

² INRAE, INRIA, Univ. Montpellier, France dino.ienco@inrae.fr

³ University of Naples Federico II, Naples 08544, IT
{simona.fioretto,elio.masciari}@unina.it

Abstract. Problems related to the environment are increasingly commonly known and consequently also technology is adapting to find suitable solutions. The ancestral technique of crop rotation was identified as a solution to address the problems related to pollution due to intensive food production (i.e. using fertilizers and pesticides). To ensure that this technique can actually improve food production, it is necessary to understand how modern technologies can support it; in particular the analysis of crop rotation can support farmers in decision making process and the optimization of farm management practices. The aim of this paper is to investigate how predictive process monitoring techniques can enhance crop rotation strategies by leveraging Agriculture 4.0 through real-time monitoring, resulting in more accurate and adaptive strategies. It is a position paper that proposes research questions for further study, which may help to develop the research area.

Keywords: Predictive Process Monitoring · Crop Rotation · Machine Learning.

1 Introduction

The expected increase of the world population by about one third by 2050 leads directly to an increase in the demand for food [18]. Fertilisers and pesticides are employed to meet this ever-increasing demand, by preventing diseases that could cause a drastic reduction in production outputs. However, this new system of food production, without technological interventions or containment measures, will lead to an expected increase in pollution of 50-92% [18], causing environmental problems and not meeting the growing demand for food [4].

In this scenario, crop rotation can help in reducing the use of pesticides and fertilizers, meeting the needs of sustainable agriculture. Crop rotation is an ancestral agricultural technique that involves the successive cultivation of different crops in a specified order in the same field [1]. It allocates fields to different crops and selects the appropriate crop type [20]. Therefore, predictions

about crop rotation sequences could support the farming system management in targeting production strategies. However, forecasting a suitable cropping plan is difficult due to the co-existence of many factors, such as weather, soil type, diseases, and pests, which strongly impact crop growth [7].

However, [20] suggests the use of detailed spatio-temporal data to improve the prediction of crop rotation sequences. Many proposed approaches leverage both spatial and temporal data, either alone or in combination. By modelling the crop sequence as a process, and integrating multi-dimensional and time-series data using event logs, it is possible to apply Predictive Process Monitoring (PPM) [2] to predict the crop rotation sequence. PPM is a subset of Process Mining (PM) techniques that use event log data to predict process behaviour [15,14].

This aim of this position paper is to present existing and emerging agricultural approaches, to identify related issues and to raise research questions. The structure of the paper is the following: in Section 2 we provide an overview of the current state of the art. In Section 3, after identifying the gaps that require further investigation, the research questions are discussed along with related comments. Finally, in Section 4, we present concluding remarks.

2 State of the art

The aim of this section is to introduce Process Mining, Prediction Techniques, and Predictive Process Monitoring methods used in agriculture.

2.1 Process Mining techniques

Process Mining (PM) is a research field that uses event logs to gain insights into business processes. PM includes *process discovery* (generating a model from an event log), *conformance checking* (comparing a process model with an event log of the same process) and *process enhancement* (improving a process model by using information from recorded processes in the event log) [21].

In [11] the authors utilise PM techniques to investigate user behaviour when employing decision support systems (DSS). The study specifically targeted farmers who used DSS for cultivar selection. The system assists farmers in selecting cultivars based on their characteristics, which can be used to identify potential features to add, when redesigning the system.

In [12] the authors combine PM and simulation-based technology to redesign processes. They use a bottom-up approach to identify processes and their performance from process data. Then, simulation on future performances of the process is employed to support the redesign phase. Finally, the approach is assessed on a web-based DSS in a Dutch agriculture organization, gaining insight into damages caused by parasites and the related financial consequences. Conclusions report that the redesign phase can benefit from potential performance gains.

The approaches analysed so far are summarised in Table 1. Therefore, it is clear that PM in the agricultural field is mainly used to identify the processes underlying the DSS used by farmers, using process discovery techniques.

Study	Approach	Task
[11]	Process Discovery	User Behavior
[12]	Process Discovery and Simulation	BP Redesign

Table 1. Process Mining methods

2.2 Prediction techniques

The proposal [9] predicts the crop yield at the end of the cropping season on ten crops. They merge 2 data-sources, one for agricultural production data, and one for weather information data. Then, they compare regression methods, such as Multiple Linear Regression, M5-Prime Regression Trees (RT), Perceptron Multilayer Neural Networks, Support Vector Regression, and k-Nearest Neighbor, finding the best attribute subset for each method using a complete algorithm. The techniques were then evaluated using Root Mean Squared Error (RMSE), Root Relative Squared Error (RRSE), Correlation Factor (R), and Mean Absolute Error (MAE) metrics, showing M5 as the best tool.

The authors [17] predict crop yield under soil salinity effects, using vegetation indices and Stepwise Linear Regression (SLR) derived from satellite images. Firstly, they create crop pattern maps of the area and extract field types using object-based classification. Secondly, they predict crop yield and map soil salinity. Finally, they investigate the correlation between yield loss and soil salinity.

The authors [10] predict crop yield before sowing, considering rainfall and soil characteristics. First, they use historical rainfall data to forecast future rainfall quantities using ARIMA and Exponential Sequencing. Since rainfall data are seasonal, a time-series analysis approach takes into account the trend. Second, they use Recurrent Neural Network (RNN) to learn crop yield patterns based on soil properties. Finally, the RNN uses the previous predictions on rainfall and soil type to predict crop yield.

The study [19] proposes various data mining techniques for predicting crop yield in the agricultural domain based on rainfall data. After selecting the relevant features, the authors focus on using regression techniques to develop a predictor model. The dataset used contains information on crop type, season, year, crop coverage area, and production.

The authors in [16] propose a framework. First, they combine LoRa Wireless Sensor Networks (WSN) with short and medium area coverage to collect and organise data in time-series. Then, they apply ARIMA prediction model for predicting future complications in production and propose a notification system to support corrective actions and achieve a better crop yield. Finally, they do 3 experiments to analyze the proposed framework.

The proposal [8] considers pre-crop values and crop rotation matrices. The latter contains factors having a positive or negative effect on the next crop, mediated through the soil. First, they train an XGBoost Regressor to predict crop rotation matrices using Normalized Difference Vegetation Index (NDVI) around harvesting time. They used satellite image data combined with weather, soil, and land-use data to identify yield-enhancing crop combinations. Results

show that the proposal can support farmers in adapting their crop rotation to local requirements and potential crops. However, its limitation lies in the use of the NDVI indicator for yield potential, as it measures plant land cover and also evaluates weeds.

The authors in [13] propose a two objective approach. First, they find the environmental factors that influence crop productivity and suitability. Then, they use six different tree-based ensemble learning models namely, Random Forest (RF), Gradient Boosting, AdaBoost, XGBoost, LightGBM, and CatBoost to predict crop suitability and productivity. The input data include environmental conditions, such as soil and climate measurements, and their impact on crop growth. The evaluation of the models on accuracy, precision, recall and F1-score, show that XGBoost and LightGBM outperform AdaBoost and RF.

The analysed approaches are summarised in Table 2. They are mainly focused on crop yield prediction, mostly using techniques for analysing and predicting time series data. Some studies also integrate Machine Learning (ML) and Artificial Neural Networks (ANN), either alone or in conjunction with ARIMA models.

Study	Approach	Prediction Task
[9]	Regression Methods	Crop Yield
[17]	SLR	Crop Yield
[10]	ARIMA Exponential Sequencing RNN	Crop Yield Patterns
[19]	Regression Methods	Crop Yield
[16]	ARIMA	Deviations
[8]	XGBoost	Crop Rotation Matrices
[13]	Tree-based Ensemble Learning	Crop Suitability Crop Productivity

Table 2. Prediction approaches

2.3 Predictive monitoring approaches

Predictive Process Monitoring (PPM) is a subset of PM techniques, which leverages event log data to predict the future behaviour of a process in execution by using prediction algorithms [2]. In fact, event log data can integrate multi-dimensional and time-series data, combining various factors affecting the crop growth.

In [5] the authors based on crop history data, use PM techniques and Markov Chains principles to predict and visualize the crop that will be planted in the following year ($n+1$) for each field in the same geographic location. The methodology involved preparing the dataset and constructing Directly Follow Graphs (DFGs) to depict crop relationships through PM. Subsequently, a prediction

model was developed using the weights derived from DFGs to generate an associated adjacency matrix. Finally, the transition matrix was estimated. Finally, the model performs predictions using the transition matrix, the SVP dataset and a coding dictionary. The evaluation of the model and presentation of the results are done using DFGs.

In [22] the authors model the crop growth as a process. More in details, they obtain crop harvest predictions by integrating static and dynamic information on crop growth into a single data input, i.e. the event log. First, they define the process model of crop growing process. Then, they use replay to merge various input data into one event log. Then, they extract prefixes from the obtained event log and group them into buckets, encoded as feature vectors with an encoding mechanism chosen according to agronomic knowledge. Finally they train statistical models for making explainable and transparent predictions. The experimentation on a real-life data case study shows promising results.

The authors in [6] leverage Long Short Time Memory (LSTM) networks to predict the most probable crop rotation scenario in a field. This is because crop rotation is a sequence-based problem. The paper employs the Seq2Seq architecture on LSTM cells auto-encoder to address the problem of predicting the most probable sequence of crops grown in a field, not only in year $(n+1)$, but also in year $(n+x)$. The LSTM predicts both past and future scenarios. To propose plausible results, these predictions are compared with predictions from a probability-conditioned model. The results of the two models are integrated.

PPM approaches analysed so far in the context of agriculture are summarised in Table 3. Only a few approaches related to PPM are reported in the literature, addressing 3 different tasks. In addition, they also focus on different approaches, the first one relies on the application of PM to show the results based on relations between activities with the use of DGF, and the second and the third ones use SL methods and ANN respectively to make predictions.

Study	Approach	Prediction Task
[5]	Markov Chains DGF	Crop Rotation
[22]	SL	Crop Harvest
[6]	ANN (LSTM)	Crop Sequence

Table 3. Predictive Monitoring Approaches

The above approaches using PPM or PM techniques rely on the transformation of collected data into event log. Therefore, the application of PM based approaches can raise limitations related to data collection and transformation. As a final point, the prediction approaches presented in Section 2.2 and Section 2.3 rely on datasets collected from multiple sources, depending on the prediction task and the availability of data in the field.

3 Research Challenges

Current research suggests that Agriculture 4.0, also known as the 4th agricultural revolution, is gaining momentum with promising results, also supported by digital transformation achieved by the introduction of new technologies. The literature review in Section 2 clearly shows an increasing tendency towards the application of PM techniques, Prediction Techniques, and PPM approaches.

Although not widely explored, PPM approaches could improve agricultural forecasting by integrating time series data with multi-dimensional data, thus enabling online forecasting. Furthermore, PPM facilitates what-if scenario analysis and the decision-making process. However, as previously mentioned, the application of PPM requires both historical and online event log data from multiple sources to be integrated. This integration is necessary to identify the features that influence the crop prediction task and to obtain more accurate predictions.

Research Questions Current approaches in the literature raise questions about the potential benefits of the application of PPM techniques in precision agriculture. These observations trigger and motivate our following research questions:

RQ₁: Which prediction task should be faced? (e.g. if the goal is to achieve the maximum production, the prediction of the crop sequence is obviously more influential than the prediction of crop harvest)

RQ₂: Which approach should be used? (e.g., based on the prediction task and on the available data, what approach should be used in the field of predictive process monitoring? Are reported approaches showing better results than others? Is the rationale of the prediction process important?)

Remarks This paper presents findings that open up fascinating possibilities for future research. Section 2 provides an overview of the current state of the art, indicating that PPM techniques have the potential to enhance crop predictions by integrating time-series data with multi-dimensional data. However, limited research studies suggest that further investigation is necessary. Since PPM predicts the future outcome of a process in execution [3], prediction results may assist farmers in decision-making processes, e.g. by re-adapting the decisions on crops based on possible predicted deviations.

The discussion of the above research questions could bring to important results, having a strong impact on the implementation of PPM in agriculture:

- First, by identifying the prediction task it is possible to detect potential future deviations (for instance a different crop production level than expected before it happens), provide results and predict probable outcomes of the crops (i.e., the crop yield, the crop harvest influencing the income), or to recommend solutions (i.e., predict the crop sequence or crop rotation);
- Second, by identifying the possible prediction models to be used, it is possible to conduct a comparative analysis of the models in order to choose the one which fits the objectives of the research (i.e., the one respecting the computational requirements and the interpretability of the model)

4 Conclusion

In this paper, the issue of food increasing demand is addressed, including concerns regarding environmental impacts. Among the available solutions, crop rotation is widely identified as an ancestral technique that can address the aforementioned problem. Along with the use of advanced technologies, it can effectively support farmers in field management. Therefore, predictive process monitoring is identified as an emerging approach currently used to address this problem. After providing a brief overview of the proposed approaches, we identified the issues related to the application of predictive monitoring in the research field. These issues are mainly related to the collection of input data. This, in turn, raises the research question to be addressed as future work, which is mostly attributable to identifying the predictive task to be addressed and the prediction model to be assessed.

References

1. Castellazzi, M., Wood, G., Burgess, P., Morris, J., Conrad, K., Perry, J.: A systematic representation of crop rotations. *Agricultural Systems* **97**(1), 26–33 (2008). <https://doi.org/https://doi.org/10.1016/j.agsy.2007.10.006>, <https://www.sciencedirect.com/science/article/pii/S0308521X07001096>
2. Di Francescomarino, C., Ghidini, C.: Predictive Process Monitoring, pp. 320–346. Springer International Publishing, Cham (2022). https://doi.org/10.1007/978-3-031-08848-3_10
3. Di Francescomarino, C., Ghidini, C., Maggi, F.M., Milani, F.: Predictive process monitoring methods: Which one suits me best? In: International conference on business process management. pp. 462–479. Springer (2018)
4. Dias, T., Dukes, A.E., Antunes, P.M.: Accounting for soil biotic effects on soil health and crop productivity in the design of crop rotations. *Journal of the science of food and agriculture* **95** **3**, 447–54 (2015), <https://api.semanticscholar.org/CorpusID:7890471>
5. Dupuis, A., Dadouchi, C., Agard, B.: Predicting crop rotations using process mining techniques and markov principals. *Computers and Electronics in Agriculture* **194**, 106686 (2022)
6. Dupuis, A., Dadouchi, C., Agard, B.: Methodology for multi-temporal prediction of crop rotations using recurrent neural networks. *Smart Agricultural Technology* **4**, 100152 (2023)
7. Dury, J., Schaller, N., García, F., Reynaud, A., Bergez, J.E.: Models to support cropping plan and crop rotation decisions. a review. *Agronomy for Sustainable Development* **32**, 567–580 (2012), <https://api.semanticscholar.org/CorpusID:16687797>
8. Fenz, S., Neubauer, T., Heurix, J., Friedel, J.K., Wohlmuth, M.L.: Ai-and data-driven pre-crop values and crop rotation matrices. *European Journal of Agronomy* **150**, 126949 (2023)
9. González Sánchez, A., Frausto Solís, J., Ojeda Bustamante, W., et al.: Predictive ability of machine learning methods for massive crop yield prediction (2014)

10. Kulkarni, S., Mandal, S.N., Sharma, G.S., Mundada, M.R., Meeradevi: Predictive analysis to improve crop yield using a neural network model. In: 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI). pp. 74–79 (2018). <https://doi.org/10.1109/ICACCI.2018.8554851>
11. Mărușter, L., Faber, N.R., Jorna, R.J., van Haren, R.J.: A process mining approach to analyse user behaviour. In: International Conference on Web Information Systems and Technologies. vol. 2, pp. 208–214. SCITEPRESS (2008)
12. Mărușter, L., Van Beest, N.R.: Redesigning business processes: a methodology based on simulation and process mining techniques. *Knowledge and Information Systems* **21**, 267–297 (2009)
13. Nti, I.K., Zaman, A., Nyarko-Boateng, O., Adekoya, A.F., Keyeremeh, F.: A predictive analytics model for crop suitability and productivity with tree-based ensemble learning. *Decision Analytics Journal* **8**, 100311 (2023). <https://doi.org/https://doi.org/10.1016/j.dajour.2023.100311>, <https://www.sciencedirect.com/science/article/pii/S2772662223001510>
14. Pasquadibisceglie, V., Appice, A., Castellano, G., van der Aalst, W.M.P.: PROMISE: coupling predictive process mining to process discovery. *Inf. Sci.* **606**, 250–271 (2022). <https://doi.org/10.1016/J.INS.2022.05.052>, <https://doi.org/10.1016/j.ins.2022.05.052>
15. Pravičević, S., Appice, A., Malerba, D.: Process mining to forecast the future of running cases. In: Appice, A., Ceci, M., Loglisci, C., Manco, G., Masciari, E., Ras, Z.W. (eds.) *New Frontiers in Mining Complex Patterns - Second International Workshop, NFMCP 2013, Held in Conjunction with ECML-PKDD 2013, Prague, Czech Republic, September 27, 2013, Revised Selected Papers. Lecture Notes in Computer Science*, vol. 8399, pp. 67–81. Springer (2013). https://doi.org/10.1007/978-3-319-08407-7_5, https://doi.org/10.1007/978-3-319-08407-7_5
16. dos Santos, U.J.L., Pessin, G., da Costa, C.A., da Rosa Righi, R.: Agriprediction: A proactive internet of things model to anticipate problems and improve production in agricultural crops. *Computers and electronics in agriculture* **161**, 202–213 (2019)
17. Satir, O., Berberoglu, S.: Crop yield prediction under soil salinity using satellite derived vegetation indices. *Field crops research* **192**, 134–143 (2016)
18. Springmann, M., Clark, M., Mason-D’Croz, D., Wiebe, K., Bodirsky, B.L., Las-saletta, L., De Vries, W., Vermeulen, S.J., Herrero, M., Carlson, K.M., et al.: Options for keeping the food system within environmental limits. *Nature* **562**(7728), 519–525 (2018)
19. Surya, P., Aroquiaraaj, I.L., et al.: Crop yield prediction in agriculture using data mining predictive analytic techniques. *International Journal of Research and Analytical Reviews* **5**(4), 783–787 (2018)
20. Upcott, E.V., Henrys, P.A., Redhead, J.W., Jarvis, S.G., Pywell, R.F.: A new approach to characterising and predicting crop rotations using national-scale annual crop maps. *Science of the Total Environment* **860**, 160471 (2023)
21. Van Der Aalst, W.: Process mining: Overview and opportunities. *ACM Transactions on Management Information Systems (TMIS)* **3**(2), 1–17 (2012)
22. Yang, J., Ouyang, C., Dik, G., Corry, P., ter Hofstede, A.H.: Crop harvest forecast via agronomy-informed process modelling and predictive monitoring. In: International Conference on Advanced Information Systems Engineering. pp. 201–217. Springer (2022)