Smart Agricultural Technology 11 (2025) 101043

Contents lists available at ScienceDirect

Smart Agricultural Technology

journal homepage: www.journals.elsevier.com/smart-agricultural-technology

Predicting tilling and seeding operation times in grain production: A comparison of machine learning and mechanistic models

Luca Scheurer *, Tobias Zimpel, Joerg Leukel

Department of Information Systems and Digital Technologies, University of Hohenheim, Stuttgart, Germany

ARTICLE INFO

Keywords:
Artificial intelligence
Crop production
Decision-making
Field operations
Predictive analytics
Process time
Cost-effective
Scheduling

ABSTRACT

Field operations management in grain production requires accurate and timely predictions of operation times for machine tasks. While machine learning (ML) is being adopted more widely in operations management, little is known about its ability to predict tilling and seeding operation times. The aim of this study was to evaluate the prediction performance of ML models for these operation times by using readily available tractor and operations data rather than dynamic environmental data. We collected data between March 2022 and August 2023 from 70 grain fields in the southwest of Germany, including variables such as tractor speed, engine speed, fuel consumption, and field geometry. Operation times exhibited high variability (coefficient of variation [CV] = 0.88). Nine ML algorithms and two conventional mechanistic models proposed by the American Society of Agricultural and Biological Engineers (ASAE EP496.3) were evaluated in a temporal external validation. Random forest (RF) models outperformed all other models, achieving a normalized root mean square error (NRMSE) of 0.215 and a coefficient of determination (R^2) of 0.910. Compared to a conventional mechanistic model, the RF model reduced the mean absolute error (MAE) by 37.8 %, and enhanced the R^2 by 0.107. The study results highlight the potential of our approach to predict tilling and seeding operation times in grain production without increasing the effort for data collection, offering an accessible and cost-effective solution for resource-constrained grain farming systems that experience data shortages.

1. Introduction

Efficient grain production depends on the precise scheduling of machinery to carry out essential field operations, such as tilling and seeding, within specific agronomic time windows. Timely alignment of tilling and seeding with growth periods and weather conditions is important, as it has been shown to improve planting conditions and promote favorable crop development [9,13,33]. Conversely, delays or inefficiencies in machinery deployment for tilling and seeding can impare crop growth, which then leads to reduced yield potential and increased operational costs [12]. The scheduling of machines across fields with varying characteristics underscores the signifcance of organized and adaptive scheduling practices [8,20,32]. Accurate predictions of tilling and seeding operation times enable better planning, resource allocation, and optimization of machinery deployment to meet those time-sensitive demands [9,23,31].

Previous studies on operation times prediction in grain production primarily focused on deterministic models, also referred to as mechanistic models, which estimate operation times based on field size, machinery specification, and task type [11,19,27]. These models employ predefined mathematical equations and logical rules grounded in physical relationships and engineering principles, and thus offer transparent and interpretable results that facilitate systematic comparisons across different field conditions and machinery types. While these models can provide general estimates, they generally fail to account for the complex relationships within field operations. Factors such as field shape, terrain variability, and obstacles can influence operation times by reducing maneuvering efficiency, while variations in soil conditions and operator proficiency further compromise operational performance [6,9]. Consequently, mechanistic models typically assume ideal or average conditions and do not readily adapt to the fluctuations introduced by real-world factors. Therefore, there is a need for more flexible predictive approaches that can capture these fluctuating variables and enable more accurate predictions of operation times.

Machine learning (ML) is increasingly being used to predict times and durations in grain production, such as harvesting time, growth stages, and machine scheduling, where field configurations, machine performance, and operation-specific variables play an important role [3,

---

* Corresponding author.
E-mail addresses: luca.scheurer@uni-hohenheim.de (L. Scheurer), tobias.zimpel@uni-hohenheim.de (T. Zimpel), joerg.leukel@uni-hohenheim.de (J. Leukel).

https://doi.org/10.1016/j.atech.2025.101043
Received 1 April 2025; Received in revised form 9 May 2025; Accepted 21 May 2025

8,35]. These operation-specific factors, which are challenging to account for using deterministic models, are better handled by ML algorithms that learn the intricate relationships between variables and the target outcome based on historical data. From an operational perspective, ML models serve as decision-support tools that integrate historical field data to identify and quantify subtle patterns and variations in operational performance, thereby enabling more refined and accurate predictions to support scheduling and resource allocation.

Recent studies suggest that ML-based approaches can significantly enhance operational efficiency in agricultural machinery management. For instance, Bettucci et al. [7] demonstrated that using a Random Forest classifier enables accurate classification of machinery operational states during ploughing, thereby converting complex data streams into actionable insights for improved decision-making. Similarly, Fedrizzi et al. [17] developed an artificial neural network (ANN) model to estimate net operation time per unit area across fields of various shapes. Their approach, which considered factors such as working speed and the number of turnings, achieved a coefficient of determination (R^2) of 0.63. In another study, Seyyedhasani and Dvorak [29] reformulated field work scheduling as a vehicle routing problem. Their method predicted operating times with a mean absolute percentage (MAPE) error between 2 % and 6 % depending on routing.

Although previous studies have highlighted the potential of ML models in agricultural operations, to our knowledge, their usefulness has not yet been examined for tilling and seeding operations, and it remains unclear whether ML models are more accurate than deterministic models in this context. An important challenge are the additional costs for acquiring dynamic environmental data, such as weather data from external weather stations and soil conditions through manual sampling, and integrating data from various systems using different protocols and data formats. This challenge can be circumvented by exploring how readily available data from existing machinery and systems can be combined to maximize prediction performance. This approach would also avoid costly soil sampling and analysis. Furthermore, developers must still select appropriate ML algorithms and assess the contributions of various features to optimize model performance.

Against this backdrop, the objectives of our research are to (1) develop ML models for tilling and seeding operation times prediction in grain farming based on historical field data and machinery data, (2) compare the performance of these ML models to that of domain-specific mechanistic models [1], and (3) evaluate the impact of field characteristics, machinery types, and operator factors on prediction performance.

2. Materials and methods

2.1. Study area

The study used data from tillage and seeding operations performed on 70 grain fields in the southwest of Germany. Sixty-two fields were located in the district of St. Wendel in Saarland and managed by a commercial farm. The remaining eight fields were located in Baden-Württemberg and used by the University of Hohenheim for field trials. Field sizes ranged from 0.17 to 15.75 ha, with an average size of 2.48 ha (SD = 2.64 ha). Thiry-three fields (47 %) had at least five corners, with the shapes ranging from rectangles to more complex forms up to 16 corners. Sixteen of the Wendel fields (19 %) were partially obstructed by trees, utility poles, or similar obstacles.

2.2. Experimental design

We employed a repeated-measures design to assess how the performance of different ML models compares to two domain-specific mechanistic models. We examined nine different ML algorithms, namely: ANN, Elastic Regression (ELR), eXtreme Gradient Boosting (XGB), Lasso Regression (LAR), Multiple Linear Regression (MLR), Partial Least Squares (PLS), Random Forest (RF), Ridge Regression (RIR), and Support Vector Machine (SVM). Models trained using these algorithms were compared both among themselves and against the mechanistic models. The mechanistic models were derived from a standard by the American Society of Agricultural and Biological Engineers. The standard ASAE EP496.3 is a reference framework for mechanistic modeling in agricultural machinery management, which provides standardized parameters that allow consistent and systematic comparisons with ML-based prediction models [1]. One of the mechanistic models, fixed tractor speed (FTS), was derived directly from the standard, whereas the second model, moving average tractor speed (MAS), was adapted from the original version by incorporating the moving average tractor speed.

2.3. Machine learning process

Fig. 1 shows the flowchart of the ML process. The process began with data collection on seeding and tilling operations, followed by data preprocessing to prepare the dataset for training. The dataset was then used to train nine ML models, each corresponding to a different experimental condition and algorithm tested, and the baseline mechanistic models. Finally, the models' performance was evaluated using a temporal external validation. The validation ensures robust model generalization by preserving the natural temporal order of observations, which is crucial for non-stationary time series, where cross-validation may yield misleading performance estimates [4,14].

2.3.1. Data collection

The data collection for tilling and seeding operations on summer and winter wheat was carried out across two consecutive growing seasons, spanning from March 2022 to August 2023. In total, data for 259 operations were collected, including 118 tilling and 127 seeding operations on the Wendel fields and 14 seeding operations on the Hohenheim fields.

Tilling and seeding operations were conducted using a fleet of tractors, including the Claas Arion 650, Fendt 314 Vario, Fendt 716 Vario, Fendt 720 Vario, and Fendt 724 Vario, in conjunction with seed drills (Kverneland E Drill and Horsch Express 3 KR) and the Lemken Karat 9 cultivator. Each tractor was equipped with a real-time telemetry module, which continuously captured performance data for the machinery in use, including fuel consumption, engine rotation speed, and tractor speed. Furthermore, the telemetry module recorded positional data, including GPS-derived coordinates and elevation. Due to privacy concerns expressed by the participating farms, operator identifiers were not collected and therefore could not be considered as a feature.

The data were collected at a frequency of one reading per second and automatically uploaded to a network provided by EXA Computing GmbH (Hamm, Germany) for real-time processing. This data processing facilitated the detection of specific operational states, including active fieldwork, field turning, idling, preparation, and road transport. The processed data were exported as individual datasets for each operation and distributed through a cloud platform.

2.3.2. Data preprocessing

We transformed the collected data into features from which prediction models can be trained. The data preprocessing included data cleaning, data decomposition, and feature engineering.

During *data cleaning* we verified the completeness and validity of the data, adjusting where necessary. Our analysis was focused solely on the plausibility of the data; hence, no statistical outlier analysis was performed. Although no values were missing, we identified nine operations that had untrustworthy GPS positions or consisted of only a single activity state. To verify GPS data, we conducted a visual inspection using Google Maps satellite imagery and identified untrustworthy positions based on unusual jumps or locations in unlikely areas such as highways or built-up zones. Most anomalies occurred at the Hohenheim fields, possibly due to interference from nearby buildings. The presence of only

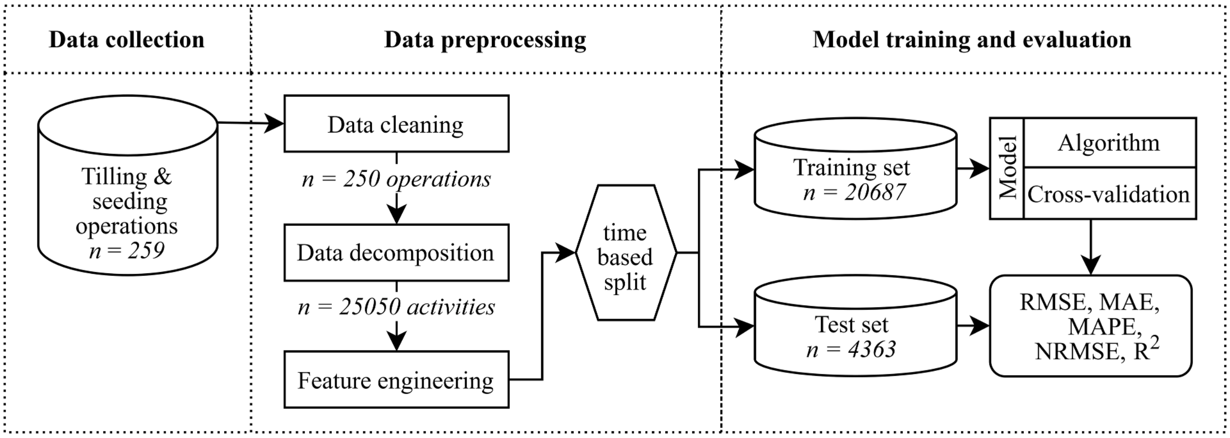


Fig. 1. Flow chart of the machine learning process.

one activity state implied that the machine was merely engaged in a single activity, such as working or turning. Such observations likely resulted from technical issues or the use of unconventional cultivators or seed drills. In total, we excluded nine operations (3.5 % of the initial dataset) resulting in a cleaned dataset comprising 250 operations.

We performed a *data decomposition* by dividing each total operation time into discrete activities such as working and turning times to increase data granularity. The cumulative time of these activities comprised the total operation time, such as tilling and seeding. Each operation was segmented into working and turning phases, generating individual samples for each activity while excluding waiting times. Waiting times were omitted due to their independence from key determinants (e.g., machinery performance and field size), as they often stem from external factors like equipment malfunctions or adjustments. While this exclusion improves model focus, the predicted operation times reflect the controllable and performance-relevant portions of field work, and may slightly underestimate total field time in practical use, particularly in cases involving prolonged interruptions or irregular delays. Since tilling and seeding are sequences of working and turning activities, each field operation encompasses multiple instances of these activities. This decomposition enabled us to disregard field-specific determinants (e.g., field entrances, shapes, obstacles) because each sample in the dataset represented a single activity within the operation, inherently capturing these spatial determinants in the aggregated activity sum. This approach also allowed us to determine net working time, thereby enabling the assessment of field efficiency [17,23]. The data decomposition yielded 25050 distinct activities derived from the set of 250 tilling and seeding operations.

Table 1
Constructed features.

Type	#	Feature	Definition
Continuous	1	Length (m)	Driven length of activity in meters
	2	Elevation (m)	Elevation gained while driving of activity in meters
	3	Straight-line distance (m)	Shortest straight-line distance between the starting and ending point of activity in meters
	4	Tractor speed (m*s^{-1})	Mean tractor speed of comparable tasks in m*s^{-1}
	5	Engine speed (1000 rpm)	Mean engine speed of comparable task in 1000 rotations per minute
	6	Fuel consumption (l*h^{-1})	Mean fuel consumption of comparable tasks in l*h^{-1}
	7	Week (n)	Week of the year in which the task was conducted
Categorical	8	Task type	Tilling or seeding task
	9	Activity state	Working or turning
	10	Field ID	Field identifier based on characters and numbers
	11	Tractor-machine combination	Tractor and machine combination identifier based on characters and numbers
	12	First use of tractor-machine combination on the field	True if a tractor-machine combination was not used so far for an operation on that field, otherwise false
	13	First use of tractor-machine-combination	True if a tractor-machine combination was not used so far for any operation, otherwise false

In *feature engineering*, we defined seven continuous and six categorical features based on available data for operations management. Tables 1 shows all 13 features used. *Length (m)* and *Elevation (m)* provide information regarding the dimensions, obstructions, and configuration of portions of a field. *Straight-line distance (m)* in contrast with *Length (m),* addressed the radius and diverse turning styles. *Tractor speed (m/s), Engine speed (1000* rpm*)*, and *Fuel consumption (l/h)* were interdependent parameters that collectively characterized the machine's power usage and output [10]. These activity-based features were calculated by averaging the activity values of the preceding field operations of the same type. Without such data, the mean value of all similar activities within the dataset was used as a proxy for farmers' experience. Temporal dependencies were captured in the feature *Week (n)*. Data on soil condition were not included in model training to ensure the development of ML models that operate independently of costly sensors or soil sampling. Collecting such data would have required additional resources, making large-scale implementation more challenging. Alternative methods, such as weather-based estimations or existing sensor data, were evaluated but found to be either impractical or not sufficiently accurate. Instead, models were developed under the assumption that farmers generally perform operations under comparable, workable conditions for each tillage or seeding task. This approach enabled a cost-effective data collection strategy while maintaining the models' applicability in agricultural settings.

As categorical features, we constructed *Task type* and *Activity state* to distinguish between tilling or seeding operations and working or turning activities. We created a field identifier *Field-ID* to cover all field differences such as shape, field entry points, and obstacles summarily. *Machine-field-combination* provided information about the specific tractor and drilling or tilling machine utilized. Additionally, we tracked the first deployment of each tractor-machine combination for each field (*First use of machine-field combination on the field*) and overall (*First use of tractor-machine combination*). Finally, we applied standard scaling to normalize all continuous variables and one-hot encoding to convert all categorical variables into binary format. The latter transformation increased the number of features from 13 to 86. The feature engineering for the 250 tasks resulted in a dataset of 25050 samples (13132 samples for working activities and 11918 samples for turning activities).

In the last step, the dataset was split time-based into a training set ($n = 20687$; 82.6 %) and a test set ($n = 4363$; 17.4 %). The training set includes data from the nine months preceding August 1, 2023, while the test set consists of data from the following one month.

2.3.3. *Model training*

We defined each ML model as a combination of algorithm and its corresponding hyperparameter configuration trained on a fixed dataset [24]. Specifically, we employed nine algorithms available in *Scikit-learn v1.2.2* [28]: ANN, ELR, XGB, LAR, MLR, PLS, RF, RIR, and SVM. We focused on standard algorithms commonly used and well-understood by most developers, prioritizing algorithms that are straightforward to implement and require less development effort compared to deep learning (DL) models. The number of features in our dataset is relatively low, making DL models less appropriate for this study. To optimize the performance of each algorithm, we specified ranges for key hyperparameters, focusing on those deemed most influential for performance. Table 2 shows the hyperparameters used for each algorithm. Computational times for model training were a few minutes only, using standard ML hardware installed on a local machine.

Hyperparameter optimization was performed using grid search with ten-fold cross-validation, applying the techniques of Bergstra and Bengio [5] and Kohavi [25]. MLR and PLS were used without grid search as the hyperparameters provided are optimized for run-time efficiency and are not expected to affect performance. We assessed each configuration by mean absolute error (MAE), chosen for its clarity and direct interpretability in measuring prediction errors [2,21]. While grid search tends to be less efficient than random search [5], we constrained the hyperparameter ranges to reduce computation time.

To compare these ML models against domain-specific mechanistic models, we defined two baseline models. The baseline models were calculated by multiplying the activity length by a speed factor, which is either the moving average speed or a fixed speed within the typical ranges for tilling and seeding operations as defined by the American Society of Agricultural and Biological Engineers [1]. Table 3 shows the formulas used.

2.3.4. *Model evaluation*

Prediction models were evaluated on an external test set ($n = 4363$) that included new observations not used in the model training and chronologically succeeded the training data. To assess the performance of the models, we used the following standard metrics [15,34]: MAE, root mean squared error (RMSE), R^2, MAPE, and normalized RMSE by mean (NRMSE). Table 4 provides the formulas for these metrics. For the final model evaluation on the test set, the ML model with the lowest MAE and the baseline model with the lower MAE were selected.

Table 2
Algorithms and hyperparameter ranges for the training of nine ML models.

#	Algorithm	Hyperparameter	Range (first value is the default value)
ML1	ANN	Activation	[identify, logistic, tanh, relu]
		Alpha	[0.00001, 0.0001, 0.001]
		Solver	[LBFGS, SGD, Adam]
ML2	ELR	Alpha	[0.1, 1.0, 10.0]
ML3	XGB	C	[1.0, 0.1, 10]
		Epsilon	[0.1, 0.01, 1.0]
		Kernel	[RBF, Linear, Poly, Sigmoid]
ML4	LAR	Alpha	[0.1, 1.0, 10.0]
ML5	MLR	-	-
ML6	PLS	-	-
ML7	RF	Bootstrap	[True, False]
		Criterion	[squared error, absolute error]
		Estimators	[100, 10, 10,000]
ML8	RIR	Alpha	[0.1, 1.0, 10.0]
ML9	SVM	C	[1.0, 0.1, 10]
		Epsilon	[0.1, 0.01, 1.0]
		Kernel	[RBF, Linear, Poly, Sigmoid]

Note. ANN = Artificial Neural Networks, ELR = Elastic Regression, XGB = eXtreme Gradient Boosting, LAR = Lasso Regression, MLR = Multiple Linear Regression, PLS = Partial Least Squares, RF = Random Forest, RIR = Ridge Regression, SVM = Support Vector Machine.

Table 3
Baseline models.

#	Baseline	Definition
B1	Moving average speed	Length × Average tractor speed in similar previous operations
B2	Fixed tractor speed	For seeding operations: Length × 2.2 m/s For tilling operations: Length × 3.1 m/s

3. Results

3.1. *Statistics of observed variables*

Table 5 provides descriptive statistics for the continuous variables. The target variable, *activity time*, ranged from 2 to 317 s, with an average value of 34.76 s, which aligns with expectations for the study fields. The coefficient of variation (CV), indicating variability, was high for the target variable (0.88), length (1.08), straight-line distance (1.32), and fuel consumption (0.50). Overall, the fields were highly heterogeneous in geometry.

Fig. 2 shows the correlation matrix for the continuous variables to identify monotonic relationships (Spearman rank correlation analysis). Activity time exhibited a very strong positive correlation with length ($r_s = 0.92$) and a strong positive correlation with straight-line distance ($r_s = 0.70$). Among the independent variables, notable correlations included fuel rate and tractor speed ($r_s = 0.82$) as well as length and straight-line distance ($r_s = 0.85$), both of which were anticipated. Elevation showed no correlations, while week showed a single weak negative correlation. In contrast, all other variables exhibited multiple correlations, with at least one reaching moderate strength. It is important to note that most of the used ML algorithms can also learn non-monotonic relationships [22]. Consequently, a lack of correlation does not imply that a feature has little relevance for prediction.

Table 6 presents descriptive statistics for the observed categorical variables. The majority of operations were categorized as seeding operations (56.41 %). Of the total field operation time, 53.42 % was spent in the working activity state. The distributions of categories for field ID and tractor-machine combination aligned with expectations. The first use of a tractor-machine combination accounted for 4.00 % of the observations, while the first use for individual fields comprised 41.80 %.

3.2. *Prediction performance using cross-validation (months 1 to 9)*

Table 7 summarizes the cross-validation averaged results using data collected in months 1 to 9. The baseline models MAS (B1) and FTS (B2)

Table 4
Performance metrics used in model training and evaluation.

Metric	Definition	Formula
MAE	Mean absolute error	$\frac{1}{n}\sum_{i=1}^{n}\lvert y_i - \hat{y}_i\rvert$
RMSE	Root mean square error	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$
R^2	Coefficient of determination	$1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$
MAPE	Mean absolute percentage error	$\frac{1}{n}\sum_{i=1}^{n}\left\lvert\frac{y_i - \hat{y}_i}{y_i}\right\rvert \times 100$
NRMSE	Normalized root mean square error (NRMSE divided by mean observed value)	$\frac{RMSE}{\bar{y}}$

Note. n = number of observations. y = observed value. $\hat{y}$ = predicted value. $\bar{y}$ = arithmetic mean of observed values.

Table 5
Descriptive statistics of continuous variables (n = 25050).

Variable	Mean	SD	Min	Max	CV
Length (m)	79.55	85.65	0.12	816.31	1.08
Elevation (m)	−0.01	7.28	−109.20	89.80	n/a
Straight-line distance (m)	68.28	90.38	0.03	786.47	1.32
Tractor speed ($m*s^{-1}$)	2.14	0.53	0.83	3.37	0.25
Engine speed (1000 rpm)	1393.49	148.97	930.06	1706.33	0.11
Fuel consumption ($l*h^{-1}$)	16.77	8.43	5.75	35.43	0.50
Week (n)	30.58	8.95	10.00	41.00	0.29
Activity time (s)	34.76	30.70	2.00	317.00	0.88

Note. CV = Coefficient of variation. n/a = not applicable due to distribution parameters.

showed differences in performance. MAS (B1) had an R^2 of 0.841, with an MAE of 7.7 s, MAPE of 31.0 %, and RMSE of 11.0 s. FTS (B2) performed worse across all metrics, with a R^2 of 0.794, MAE of 9.0 s, MAPE of 26.8 %, and RMSE of 12.6 s.

Among the ML models, all but one algorithm achieved an R^2 between around 0.90 or 0.91, and they differed primarily in RMSE and MAPE. XGB (ML3) and RF (ML7) achieved the best performance with an R^2 of 0.910 and an MAE of 5.4 s. These models outperformed all other algorithms regarding MAE, MAPE, and RMSE. However, RF slightly outperformed XGB in both RMSE and MAPE, with values of 8.1 s and 23.8 %, compared to 8.4 s and 24.5 % for XGB. Other models, such as ANN (ML1) and SVM (ML9), showed similar performance with MAE values ranging from 5.8 to 6.1 s and RMSE values of 8.6 and 8.7 s. In contrast, models like PLS (ML6) showed a larger performance gap (R^2 = 0.842, MAPE = 40.6 %, and RMSE = 10.7 s).

3.3. *Prediction performance in temporal external validation (month 10)*

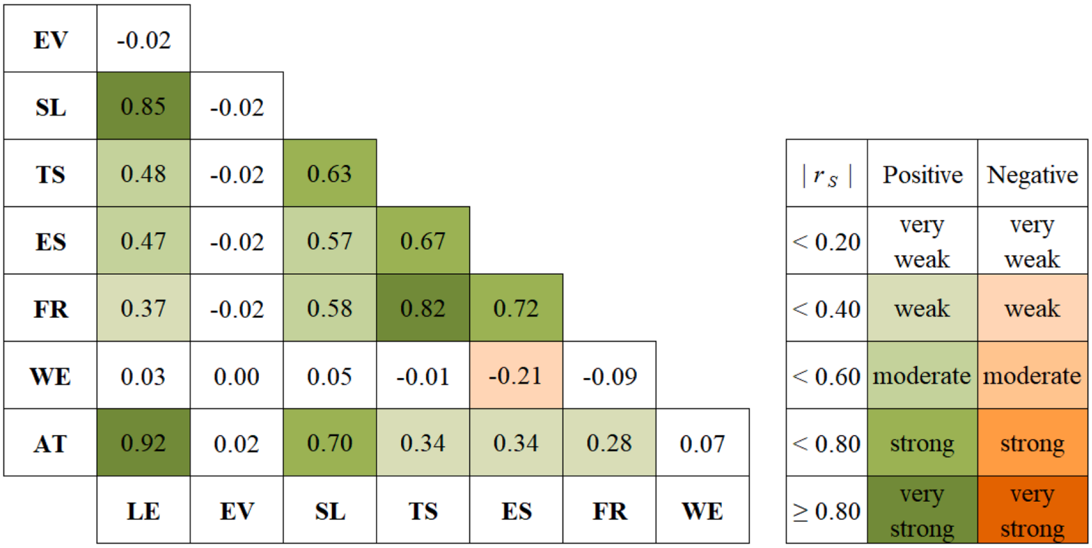
Table 8 shows the performance of the RF model and the baseline model MAS (B1), which were selected based on their performance in the cross-validation. The MAS model achieved an R^2 of 0.803, MAPE of 26.1 %, and NRMSE of 0.318 %. The RF model showed higher performance than the MAS model across all metrics (R^2 = 0.910, MAPE = 17.4 %, and NRMSE =0.215). Scatter plots of observed and predicted activity time for each model are shown in Fig. 3.

To gain further insights into the role of specific features, we analyzed the relative importance of features as indicated by the Gini index for RF models. The following results were obtained for the test set. Length had the highest Gini index (91 %), followed by straight-line distance (4 %), cumulative field ID (2 %), and elevation (1 %).

4. Discussion

This study evaluated the performance of nine ML models in predicting tilling and seeding operation times, using data from small fields in Southwest Germany. We found that ML models, particular RF, were more accurate than two domain-specific mechanistic models. ML models demonstrated their potential as cost-effective solutions for field operations management by using readily available input data such as tractor speed, engine speed, fuel consumption, and field geometry.

Concerning the comparison of mechanistic models to ML models, all tested ML algorithms outperformed the deterministic models in the cross-validation. The ML models demonstrated R^2 values ranging from 0.842 to 0.910, with a reduced MAE of 18.9 % to 36.4 % compared to the best baseline model (MAS: R^2 = 0.841; MAE = 7.7 s). This difference in performance can be attributed to the ability of ML algorithms to better capture the high variability inherent in tilling and seeding operation



EV	-0.02						
SL	0.85	-0.02					
TS	0.48	-0.02	0.63				
ES	0.47	-0.02	0.57	0.67			
FR	0.37	-0.02	0.58	0.82	0.72		
WE	0.03	0.00	0.05	-0.01	-0.21	-0.09	
AT	0.92	0.02	0.70	0.34	0.34	0.28	0.07
	LE	**EV**	**SL**	**TS**	**ES**	**FR**	**WE**

$\lvert r_S \rvert$	Positive	Negative
< 0.20	very weak	very weak
< 0.40	weak	weak
< 0.60	moderate	moderate
< 0.80	strong	strong
≥ 0.80	very strong	very strong

Fig. 2. Correlation matrix (n = 25050, Spearman rank correlation coefficients). AT = activity time; ES = engine speed; EV = elevation; FR = fuel rate; LE = length; SL = straight-line distance; TS = tractor speed; WE = week.

Table 6
Descriptive statistics of categorical variables (n = 25050).

Variable	Categories	Mode	Mode Percentage	Minority	Minority Percentage
Task type	2	Seeding	56.41 %	Tilling	43.59 %
Activity state	2	Working	53.42 %	Turning	47.58 %
Field ID	76	Wendel-F28	6.57 %	Hoh-F4	0.02 %
Tractor-machine combination	4	Fendt314	54.01 %	Claas660	0.03 %
First use of tractor-machine combination on field	2	No	58.20 %	Yes	41.80 %
First use of tractor-machine combination	2	No	96.00 %	Yes	4.00 %

Table 7
Performance of two baseline and nine ML prediction models using cross-validation and data from months 1 to 9 ($n = 20687$).

#	Algorithm	MAE (s)	RMSE (s)	R^2	MAPE (%)	NRMSE
B1	MAS	7.7	11.0	0.841	31.0	0.321
B2	FTS	9.0	12.6	0.794	26.8	0.364
ML1	ANN	6.1	8.6	0.895	29.6	0.253
ML2	ELR	6.1	8.6	0.897	32.7	0.252
ML3	XGB	5.4	8.4	0.909	19.0	0.245
ML4	LAR	5.7	8.2	0.907	27.3	0.240
ML5	MLR	5.6	8.2	0.905	27.5	0.240
ML6	PLS	8.1	10.7	0.842	40.6	0.312
ML7	RF	5.4	8.1	0.910	18.9	0.238
ML8	RIR	6.2	8.6	0.896	29.3	0.252
ML9	SVM	5.8	8.7	0.902	25.4	0.254

Note. ANN = Artificial Neural Networks, ELR = Elastic Regression, FTS = Fixed tractor speed, LAR = Lasso Regression, MAS = Moving average speed, MLR = Multiple Linear Regression, PLS = Partial Least Squares, RF = Random Forest, RIR = Ridge Regression, SVM = Support Vector Machine, XGB = eXtreme Gradient Boosting.

Table 8
Performance of the selected prediction models on data from month 10 ($n = 4363$).

#	Algorithm	MAE (s)	RMSE (s)	R^2	MAPE (%)	NRMSE
B1	MAS	8.2	11.1	0.803	26.1	0.318
ML7	RF	5.1	7.5	0.910	17.4	0.215

Note. MAS = Moving average speed, RF = Random Forest.

data, which deterministic models often struggle to account for. Mechanistic models rely on predefined, simplified relationships and assumptions, limiting their capacity to represent complex and nonlinear relationships and interactions between factors such as field characteristics, machinery types, and operator behavior. This assumption is further supported by the feature importance analysis, in which straight-line distance and cumulative field ID ranked second and third in terms of Gini index, respectively. These features, which reflect aspects of operator behavior and field variability, highlight factors that mechanistic models typically overlook or cannot adequately represent. Some caution is warranted when interpreting Gini indices, as they may not fully capture the contributions of individual variables; future studies could benefit from employing explainable AI techniques to provide more comprehensive insights.

Regarding the comparison of ML models, the results showed that the RF model demonstrated the best prediction performance in the cross-validation ($R^2 = 0.910$), followed by XGB, which achieved the similar performance ($R^2 = 0.909$). Both RF and XGB outperformed all other ML models, demonstrating that specifically ensemble learners are effective for predicting tilling and seeding operation times and managing the complexity of agricultural operations. RF maintained its prediction performance in the temporal external validation, where it achieved an R^2 of 0.910 and had a reduced MAE by 37.8 %, compared to the baseline model MAS. These results confirm RF's ability to generalize to later data, and suitability for field tilling and seeding operation time prediction.

Our study results on prediction models for tilling and seeding operation times in grain production have several implications for using such models in practice. First, our findings demonstrate the ability of ML models to better predict operation times than conventional mechanistic models. This ability enables greater alignment of tilling and seeding operations with agronomic windows, improving operational planning and machinery scheduling, and enhancing overall crop productivity.

Second, the results highlight that ML models enable higher prediction performance for operation times prediction without requiring additional data collection, making the approach accessible. This aspect is particularly relevant for grain farming systems, in which resources and access to advanced data collection tools may be limited. The approach enables adoption without requiring upfront investments in specialized sensors and retrieval of dynamic enviromental data, such as weather data from external stations and soil conditions through in-field sampling. Additionally, the models' ability to learn from historical data, commonly recorded by modern field machinery and practices, facilitate deployment with minimal setup. This capability enhances operational efficiency and improves scheduling [18,30]. The cost-effectiveness of these models is also enhanced by their scalability with new data, as they can be applied without requiring extensive resource investment. The scalability makes ML prediction models a promising option for resource-constrained farming systems aiming to modernize operations [18,26].

Third, these models offer value for operational planning by generating outputs that can be integrated into decision support systems. One practical use case involves predicting longer operation times for newly managed fields, allowing farm managers to proactively allocate

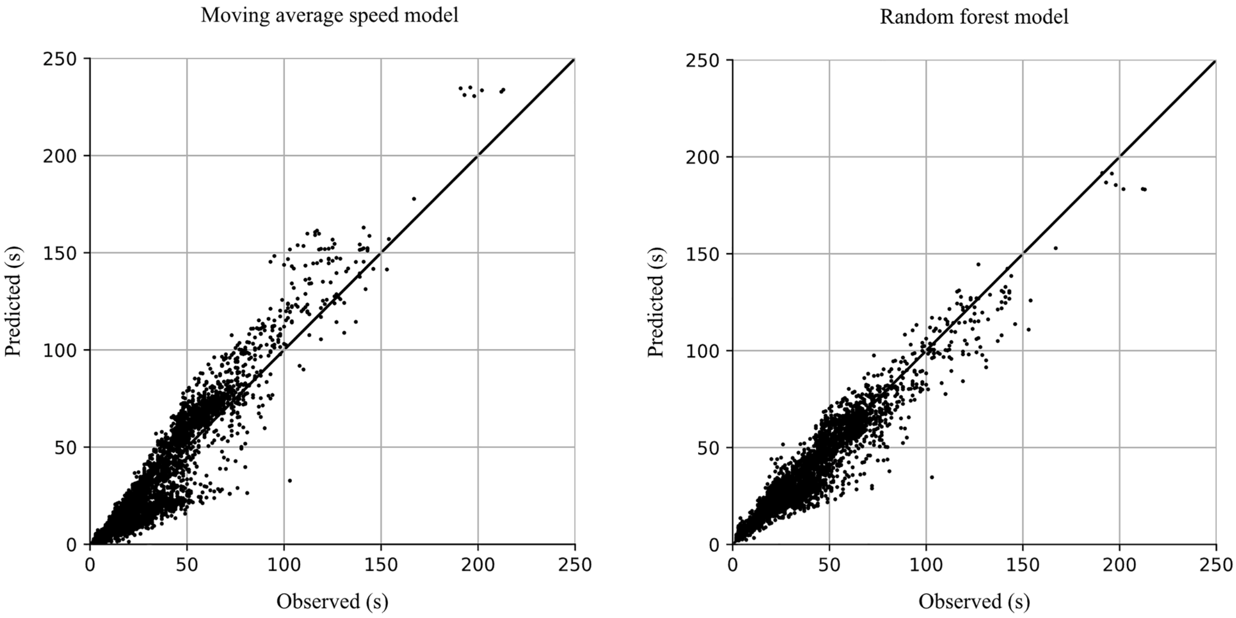


Fig. 3. Observed versus predicted activity times for the selected models MAS (Moving Average Speed, left) and Random Forest (right) on the test set ($n = 4363$).

additional resources or adjust schedules to accommodate the increased effort. Conversely, when predicted operation times are significantly shorter than expected, freed-up resources can be reassigned to other tasks, increasing overall farm efficiency.

Fourth, this study emphasizes challenges in developing prediction models for operation times that are highly variable, as it was the case in our dataset. This variability is primarily driven by great diversity in field geometry as well as operator skills and other environmental influences. However, ML models were effective in capturing complex relationships between input variables and operation times, leading to more accurate predictions compared to conventional models. Although the results demonstrate the effectiveness of ML models for seeding and tilling in grain production, their generalizability to other operations, such as harvesting, spraying, and fertilizing, remains uncertain. The operation times of these tasks may be contingent upon different variables, which would require further validation to assess the models' applicability across a broader range of field activities and crops. Furthermore, the models relied on a small set of input variables, excluding factors like weather conditions, soil parameters, and individual operator behavior, which could impact operation times, and thus could be examined in future studies. The models requires further evaluation to assess their applicability to regions beyond those studied. Overall, although this study highlights the potential of ML models for predicting tilling and seeding operation times, further research is essential to enhance their generalizability and improve their predictive performance in diverse grain farming systems.

5. Conclusion

This study examined how ML models can be employed for predicting tilling and seeding operation times in grain production. We evaluated and compared nine ML models using different algorithms, all of which outperformed conventional mechanistic models in the temporal external validation. Ensemble learners, especially RF, showed the highest prediction performance. The improvements in performance were substantial, and no additional effort was required to collect data other than operational data that was already available and recorded. These characteristics make ML models a cost-effective approach for supporting field operations and machinery scheduling. Overall, the findings suggest that ML models, particularly RF, offer a practical, accessible tool for improving operational decision-making without significant additional costs for data collection.

Funding

This work was supported by the Federal Ministry of Food and Agriculture (BMEL), Germany, under grant 28DK109A20. Publishing fees were supported by the Funding Programme Open Access Publishing of the University of Hohenheim.

CRediT authorship contribution statement

Luca Scheurer: Writing – original draft, Software, Methodology, Investigation, Conceptualization. **Tobias Zimpel:** Writing – review & editing, Software, Investigation. **Joerg Leukel:** Writing – review & editing, Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethics Statement

Not applicable: This manuscript does not include human or animal research.

Acknowledgment

The authors would like to thank the project partners EXA Computing GmbH and the Petershof in Marpingen for their tireless data collection and provision of the dataset, especially Niklas Rauber.

Data availability

The authors do not have permission to share data.

References

[1] ASABE, Agricultural Machinery Management, ASABE, 2020. https://elibrary.asabe.org/azdez.asp?jid=2&aid=36431&cid=s2000&t=2.
[2] P.L. Bartlett, S. Boucheron, G. Lugosi, Model selection and error estimation, Mach. Learn. 48 (1/3) (2002) 85–113, https://doi.org/10.1023/A:1013999503812.
[3] A. Bechar, C. Vigneault, Agricultural robots for field operations. Part 2: operations and systems, Biosyst. Eng. 153 (2017) 110–128, https://doi.org/10.1016/j.biosystemseng.2016.11.004.
[4] C. Bergmeir, J.M. Benítez, On the use of cross-validation for time series predictor evaluation, Inf. Sci. (Ny.) 191 (2012) 192–213, https://doi.org/10.1016/j.ins.2011.12.028.
[5] J. Bergstra, Y. Bengio, Random search for hyper-parameter optimization, J. Mach. Learn. Res. 13 (10) (2012) 281–305. https://www.jmlr.org/papers/v13/bergstra12a.html.
[6] H. Bernhardt, J. Bartenschlager, M. Stettmer, M. Mederle, Turning Structures and Driving Strategies of Different fields. In 2018 ASABE Annual International Meeting, 2018, American Society of Agricultural and Biological Engineers, 2018, https://doi.org/10.13031/aim.201800501.
[7] F. Bettucci, M. Sozzi, M. Benetti, L. Sartori, A data-driven approach to agricultural machinery working states analysis during ploughing operations, Smart Agric. Technol. 8 (2024) 100511, https://doi.org/10.1016/j.atech.2024.100511.
[8] D.D. Bochtis, C.G Sørensen, P. Busato, Advances in agricultural machinery management: a review, Biosyst. Eng. 126 (2014) 69–81, https://doi.org/10.1016/j.biosystemseng.2014.07.012.
[9] D.D. Bochtis, C.G Sørensen, O. Green, D. Moshou, J. Olesen, Effect of controlled traffic on field efficiency, Biosyst. Eng. 106 (1) (2010) 14–25, https://doi.org/10.1016/j.biosystemseng.2009.10.009.
[10] J. Boysen, L. Zender, A. Stein, Modeling the soil-machine response of secondary tillage: a deep learning approach, Smart Agric. Technol. 6 (2023) 100363, https://doi.org/10.1016/j.atech.2023.100363.
[11] D.R. Buckmaster, J.W. Hilton, Computerized cycle analysis of harvest, transport, and unload systems, Comput. Electron. Agric. 47 (2) (2005) 137–147, https://doi.org/10.1016/j.compag.2004.11.015.
[12] N.E. Caicedo Solano, G.A. García Llinás, J.R Montoya-Torres, Operational model for minimizing costs in agricultural production systems, Comput. Electron. Agric. 197 (2022) 106932, https://doi.org/10.1016/j.compag.2022.106932.
[13] P.M. Carr, R.D. Horsley, W.W. Poland, Tillage and seeding rate effects on wheat cultivars: I. Grain production, Crop. Sci. 43 (1) (2003) 202–209, https://doi.org/10.2135/cropsci2003.2020.
[14] V. Cerqueira, L. Torgo, I. Mozetič, Evaluating time series forecasting models: an empirical study on performance estimation methods, Mach. Learn. 109 (11) (2020) 1997–2028, https://doi.org/10.1007/s10994-020-05910-7.
[15] T. Chai, R.R. Draxler, Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature, Geosci. Model. Dev. 7 (3) (2014) 1247–1250, https://doi.org/10.5194/gmd-7-1247-2014.
[16] M. Fedrizzi, F. Antonucci, G. Sperandio, S. Figorilli, F. Pallottino, C. Costa, An artificial neural network model to predict the effective work time of different agricultural field shapes, Span. J. Agric. Res. 17 (1) (2019) e0201, https://doi.org/10.5424/sjar/2019171-13366.
[17] R. Gautron, O.-.A. Maillard, P. Preux, M. Corbeels, R. Sabbadin, Reinforcement learning for crop management support: review, prospects and challenges, Comput. Electron. Agric. 200 (2022) 107182, https://doi.org/10.1016/j.compag.2022.107182.
[18] S. Gupta, H. Rikhtehgar Berenji, M. Shukla, N.N. Murthy, Opportunities in farming research from an operations management perspective, Prod. Oper. Manage 32 (6) (2023) 1577–1596, https://doi.org/10.1111/poms.13967.
[19] P. He, J. Li, X. Wang, Wheat harvest schedule model for agricultural machinery cooperatives considering fragmental farmlands, Comput. Electron. Agric. 145 (2018) 226–234, https://doi.org/10.1016/j.compag.2017.12.042.
[20] M. Hubl, Modeling an agricultural process coordination problem to enhance efficiency and resilience with methods of artificial intelligence, Modellier. 2022 Satell. Event. (2022) 6–17, https://doi.org/10.18420/modellierung2022ws-003. Gesellschaft für Informatik e.V.
[21] R.J. Hyndman, A.B. Koehler, Another look at measures of forecast accuracy, Int. J. Forecast. 22 (4) (2006) 679–688, https://doi.org/10.1016/j.ijforecast.2006.03.001.
[22] G. James, D. Witten, T. Hastie, R. Tibshirani, An Introduction to Statistical Learning, Springer US, 2021, https://doi.org/10.1007/978-1-0716-1418-1.

[23] A. Janulevičius, E. Šarauskis, A. Čiplienė, A. Juostas, Estimation of farm tractor performance as a function of time efficiency during ploughing in fields of different sizes, Biosyst. Eng. 179 (2019) 80–93, https://doi.org/10.1016/j.biosystemseng.2019.01.004.

[24] M. Kearns, Y. Mansour, A.Y. Ng, D. Ron, An experimental and theoretical comparison of model selection methods, Mach. Learn. 27 (1) (1997) 7–50, https://doi.org/10.1023/A:1007344726582.

[25] R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, in: IJCAI'95, Proceedings of the 14th International Joint Conference on Artificial Intelligence 2, Morgan Kaufmann Publishers Inc, 1995, pp. 1137–1143. Volume.

[26] A.A. Mana, A. Allouhi, A. Hamrani, S. Rehman, I. el Jamaoui, K. Jayachandran, Sustainable AI-based production agriculture: exploring AI applications and implications in agricultural practices, Smart Agric. Technol. 7 (2024) 100416, https://doi.org/10.1016/j.atech.2024.100416.

[27] A. Orfanou, P. Busato, D.D. Bochtis, G. Edwards, D. Pavlou, C.G. Sørensen, R. Berruto, Scheduling for machinery fleets in biomass multiple-field operations, Comput. Electron. Agric. 94 (2013) 12–19, https://doi.org/10.1016/j.compag.2013.03.002.

[28] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, et al., Scikit-learn: machine learning in Python, J. Mach. Learn. Res. 21 (9) (2009) 1263–1284.

[29] H. Seyyedhasani, J.S. Dvorak, Reducing field work time using fleet routing optimization, Biosyst. Eng. 169 (2018) 1–10, https://doi.org/10.1016/j.biosystemseng.2018.01.006.

[30] R. Sharma, S.S. Kamble, A. Gunasekaran, V. Kumar, A. Kumar, A systematic literature review on machine learning applications for sustainable agriculture supply chain performance, Comput. Oper. Res. 119 (2020) 104926, https://doi.org/10.1016/j.cor.2020.104926.

[31] C.G. Sørensen, V. Nielsen, Operational analyses and model comparison of machinery systems for reduced tillage, Biosyst. Eng. 92 (2) (2005) 143–155, https://doi.org/10.1016/j.biosystemseng.2005.06.014.

[32] M. Spekken, S.de Bruin, Optimized routing on agricultural fields by minimizing maneuvering and servicing time, Precis. Agric. 14 (2) (2013) 224–244, https://doi.org/10.1007/s11119-012-9290-5.

[33] R. Tsimba, G.O. Edmeades, J.P. Millner, P.D. Kemp, The effect of planting date on maize grain yields and yield components, Field. Crops. Res. 150 (2013) 135–144, https://doi.org/10.1016/j.fcr.2013.05.028.

[34] C.J. Willmott, K. Matsuura, Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance, Clim. Res. 30 (1) (2005) 79–82. http://www.jstor.org/stable/24869236.

[35] Y. Yue, J.-.H. Li, L.-.F. Fan, L.-.L. Zhang, P.-.F. Zhao, Q. Zhou, N. Wang, Z.-.Y. Wang, L. Huang, X.-.H. Dong, Prediction of maize growth stages based on deep learning, Comput. Electron. Agric. 172 (2020) 105351, https://doi.org/10.1016/j.compag.2020.105351.