University of Thi-Qar Journal of agricultural research
Thi-Qar Journal of agricultural research
ISSN Onlin:2708-9347, ISSN Print: 2708-9339 Volume 15, Issue 1 (2026) PP 243-250
https://jam.utq.edu.iq/index.php/main https://doi.org/10.54174/2e0f0q75

Operation-Level Prediction of Tractor Productivity and Fuel Consumption in Vineyard Operations

[1]Rawaz Jalal Hama Ali, [2]Fawzy Faidhullah Khurshid, [3]Adnan Fattah Khalid [4]Shaee Adeeb Ghareeb

[1,2,4]Department of Biotechnology and Crop Science, College of Agricultural Engineering Sciences, University of Sulaimani, Kurdistan-Iraq

[3]Department of Manufacturing and Industrial Engineering, Faculty of Engineering, Koya University

E-mail: Rawaz.hamaali@univsul.edu.iq
E-mail:Fawzy.khurshid@univsul.edu.iq
E-mail:adnan.fattah@koyauniversity.org
E-mail: Shaee.gharib@univsul.edu.iq

Abstract

Precise prediction of tractor operational output and fuel use is important for machinery management and sustainable agricultural operations. This study evaluated whether operation-level CAN-Bus and GNSS summaries can support tractor-agnostic prediction of effective field capacity and fuel consumption per unit area in vineyard operations. An open vineyard tractor telemetry dataset was used, with two targets: HectaresPerH and LitersPerHa. Three predictive approaches were compared: an activity-mean baseline, Ridge regression, and Histogram-based Gradient Boosting Regression. Model performance was evaluated using leave-one-tractor-out validation with nested hyperparameter tuning, allowing the assessment of transferability to unseen tractors. For HectaresPerH, the activity-mean baseline, Ridge, and HGBR produced very similar pooled performance, with the baseline slightly strongest by RMSE and R2. This indicates that tractor productivity at the operation-summary level was largely captured by activity context. For LitersPerHa, both machine-learning models improved over the baseline, and HGBR achieved the best pooled performance, although the improvement was modest. Wilcoxon signed-rank tests on fold-wise RMSE did not show statistically significant differences between the main model comparisons. Activity-level error analysis showed that predictive difficulty varied across vineyard tasks. The findings indicate that CAN-derived operation summaries contain useful predictive signal, but simple operational baselines remain important. The main contribution of the study is a realistic unseen-tractor evaluation framework.

Keyword: CAN-Bus telemetry; field capacity; fuel consumption; tractor performance; vineyard operations

I. INTRODUCTION

Improving tractor productivity and fuel consumption is a central objective in agricultural engineering because machinery performance affects timeliness, operating cost, energy use, and farm-level resource management. Effective field capacity and fuel consumption per unit area are widely used indicators because they connect machine behaviour with field output and resource efficiency (ASABE, 2011; Kolator, 2021). Modern tractors increasingly generate CAN-Bus and GNSS data that describe engine state, motion, location, and operation context. These data allow real-world evaluation of tractor use beyond controlled test conditions (Mattetti et al., 2021; Angelucci and Mattetti, 2024; Götz et al., 2025).

University of Thi-Qar Journal of agricultural research
Thi-Qar Journal of agricultural research
ISSN Onlin:2708-9347, ISSN Print: 2708-9339 Volume 15, Issue 1 (2026) PP 243-250
https://jam.utq.edu.iq/index.php/main https://doi.org/10.54174/2e0f0q75

Vineyard mechanization is a relevant case for telemetry-based analysis because operations are repeated but heterogeneous. Tasks such as crop protection, tillage, ridging, mulching, and harvesting differ in speed, load state, maneuvering, and fuel demand. Recent vineyard research has shown that tractor performance, fuel consumption, and operational efficiency vary by activity, confirming that task type is part of the machinery-performance structure rather than background information (Alemanno et al., 2025; Rossi and Alemanno, 2024).

Machine-learning models are increasingly used for agricultural machinery prediction, but many studies emphasize within-dataset accuracy rather than transferability to unseen machines. This distinction is important because random record-level splitting can overestimate performance when observations from the same tractor appear in both training and testing sets. For structured data, validation should respect group structure when the practical use case is prediction for unseen groups (Roberts et al., 2017). When hyperparameters are tuned, nested validation is also recommended to reduce optimistic bias in error estimation (Varma and Simon, 2006).

The objective of this study was to evaluate whether operation-level CAN-Bus and GNSS summaries can support tractor-agnostic prediction of tractor productivity and fuel consumption in vineyard operations. Ridge regression and Histogram-based Gradient Boosting Regression were compared with an activity-mean baseline under leave-one-tractor-out validation with nested tuning. The contribution of this paper is a realistic evaluation of unseen-tractor prediction using tractor telemetry.

II. MATERIALS AND METHODS

A. Dataset and targets

This study used the Tractor Performances in Vineyard Operations dataset. The dataset contains operation-level summaries derived from vineyard tractor telemetry obtained from "Tractor performances in vineyard operations" by Rossi and Alemanno (2024) 26 vineyard fields in Tuscany, Italy collected from CAN-Bus and GNSS sources. Each record represents a summarized operation rather than a high-frequency time-series segment. The analysis focused on two outcomes: HectaresPerH, interpreted as effective field capacity, and LitersPerHa, interpreted as fuel consumption per unit area. These outcomes are directly relevant to tractor productivity and fuel-use assessment in agricultural machinery management (ASABE, 2011; Kolator, 2021).

B. Predictors and leakage control

Predictors were drawn from operation-level descriptors related to engine state, motion, slope, temperature, PTO behaviour, and activity context. The tractor identifier was used only as a grouping variable for validation and was not used as a predictor. Variables that directly represented or could reconstruct the target variables were excluded to reduce target leakage risk. This handling is summarized in TABLE (1).

TABLE (1): PREDICTOR HANDLING AND LEAKAGE-CONTROL SUMMARY

Variable group	Model role	Reason
Activity	Included	Operational context strongly affects vineyard tractor behaviour
Wheel speed, engine speed, torque, PTO, temperature, slope	Included	Telemetry and operation descriptors available before target calculation
Tractor identifier	Excluded from predictors; used for grouping	Prevents memorization of tractor-specific patterns
HectaresPerH and LitersPerHa	Targets only	Response variables
Area, time, total fuel, or directly target-derived variables	Excluded when target-adjacent	Potential target-construction or proxy leakage risk

ISSN Onlin:2708-9347, ISSN Print: 2708-9339 Volume 15, Issue 1 (2026) PP 243-250
https://jam.utq.edu.iq/index.php/main https://doi.org/10.54174/2e0f0q75

C. Models, validation, and metrics

Three approaches were evaluated: an activity-mean baseline, Ridge regression, and Histogram-based Gradient Boosting Regression. Ridge regression was selected because regularized linear models are appropriate for multicollinear predictors (Hoerl and Kennard, 1970). HGBR was selected as a non-linear tree-based model capable of capturing interactions and threshold-like responses (Friedman, 2001). The baseline was included to determine whether machine learning improved on a simple practical reference predictor.

The outer evaluation used leave-one-tractor-out validation. In each fold, all observations from one tractor were held out for testing, while the remaining tractors were used for training and tuning. Hyperparameters were tuned only within the training tractors using an inner validation procedure, giving a nested design. Performance was summarized using mean absolute error, root mean square error, and coefficient of determination. Fold-wise RMSE values were compared using the Wilcoxon signed-rank test (Wilcoxon, 1945; Roberts et al., 2017; Varma and Simon, 2006).

D. Interpretation and activity-level analysis

Model behaviour was examined using predicted-versus-observed plots, predictor correlation analysis, permutation importance, and activity-level error summaries. Because tractor telemetry variables are physically and statistically linked, interpretation was treated as descriptive rather than causal. This caution is consistent with the known limitations of marginal and importance-based interpretation under feature dependence (Apley and Zhu, 2020; Molnar et al., 2023).

III. RESULTS AND DISCUSSION

A. Overall predictive performance

TABLE (2) summarizes the pooled predictive performance. For HectaresPerH, all three approaches performed similarly. The activity-mean baseline achieved the lowest RMSE and highest R2, while Ridge had a slightly lower MAE. This indicates that machine-learning models added little value beyond activity context for productivity prediction at the operation-summary level.

TABLE (2): POOLED PERFORMANCE METRICS UNDER LEAVE-ONE-TRACTOR-OUT VALIDATION

Target	Model	MAE	RMSE	R2
HectaresPerH	BaselineActivityMean	0.476	0.686	0.562
HectaresPerH	Ridge	0.472	0.693	0.554
HectaresPerH	HGBR	0.477	0.694	0.552
LitersPerHa	BaselineActivityMean	2.657	3.401	0.344
LitersPerHa	Ridge	2.571	3.304	0.381
LitersPerHa	HGBR	2.529	3.268	0.394

For LitersPerHa, both machine-learning models improved over the activity-mean baseline. HGBR achieved the best pooled RMSE and R2, followed closely by Ridge. The improvement was modest, but it suggests that fuel consumption per unit area depends on more complex relationships among motion, load state, and operation context than productivity alone. Similar complexity in tractor fuel use has been reported in machinery performance studies (Pitla et al., 2016; Kolator, 2021).

B. Fold-wise model comparison

The Wilcoxon signed-rank tests in TABLE (3) showed that none of the main model comparisons were statistically significant. Therefore, the numerical advantage of the activity baseline for HectaresPerH and the numerical advantage of HGBR for LitersPerHa should not be interpreted as strong model superiority across tractors.

University of Thi-Qar Journal of agricultural research

Thi-Qar Journal of agricultural research

ISSN Onlin:2708-9347, ISSN Print: 2708-9339 Volume 15, Issue 1 (2026) PP 243-250

https://jam.utq.edu.iq/index.php/main https://doi.org/10.54174/2e0f0q75

TABLE (3): WILCOXON SIGNED-RANK TESTS FOR FOLD-WISE RMSE COMPARISONS

Target	Comparison	Statistic	p-value
HectaresPerH	Ridge vs HGBR	8	0.6875
HectaresPerH	Ridge vs Baseline	9	0.8438
LitersPerHa	Ridge vs HGBR	6	0.4375
LitersPerHa	Ridge vs Baseline	6	0.4375

C. Predicted-versus-observed behaviour

The predicted-versus-observed plots in Fig. 1 show that the models captured meaningful structure in both targets, but residual variability remained substantial. This visual pattern is consistent with the moderate R2 values in TABLE (2). The scatter also indicates reduced precision at the extremes, which is expected when operation-level summaries are used instead of detailed time-series data.

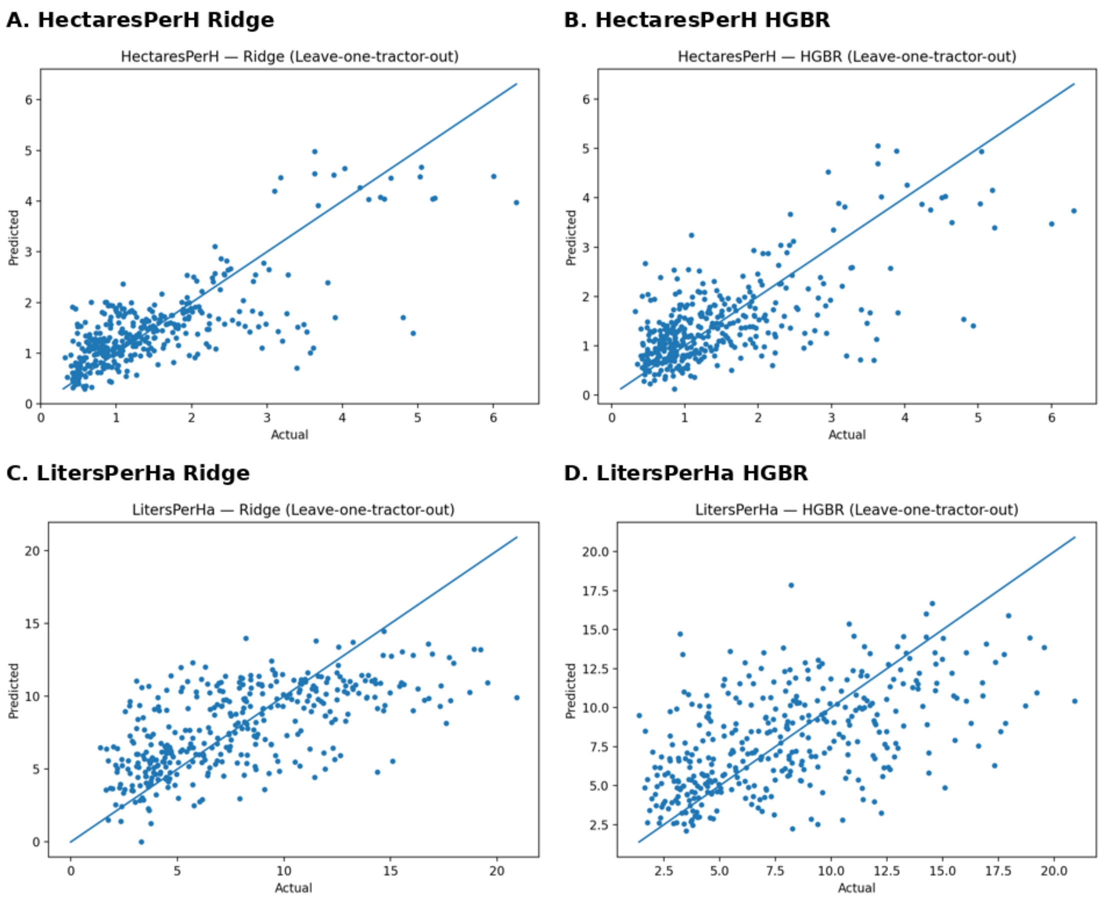

Fig. 1. Predicted-versus-observed values for HectaresPerH and LitersPerHa under leave-one-tractor-out validation.

University of Thi-Qar Journal of agricultural research
Thi-Qar Journal of agricultural research
ISSN Onlin:2708-9347, ISSN Print: 2708-9339 Volume 15, Issue 1 (2026) PP 243-250
https://jam.utq.edu.iq/index.php/main https://doi.org/10.54174/2e0f0q75

D. Activity-level errors

Activity-level errors showed that prediction difficulty varied across vineyard operations. TABLE (4) reports examples of high- and low-error activities. Harvest and crop protection with multiple rows were among the more difficult activities, while leaf removal and crop protection with single rows showed lower errors. These patterns support the interpretation that a single global model may average across distinct operational regimes.

TABLE (4): REPRESENTATIVE HIGH- AND LOW-ERROR ACTIVITIES BY TARGET

Target	Error group	Activity	MAE Ridge	MAE HGBR
HectaresPerH	High error	Crop protection MR	0.874	0.852
HectaresPerH	High error	Harvest	0.840	0.754
HectaresPerH	Low error	Leaf removal	0.135	0.245
HectaresPerH	Low error	Unridging	0.255	0.352
LitersPerHa	High error	Harvest	4.820	4.599
LitersPerHa	High error	Shoot removal	3.506	3.301
LitersPerHa	Low error	Crop protection SR	1.212	1.145
LitersPerHa	Low error	Mulching	1.972	1.928

E. Predictor correlation and interpretation

The predictor correlation matrix in Fig. 2 shows substantial dependence among several telemetry variables, especially engine and powertrain indicators. This is mechanically reasonable, because power, torque, engine speed, and thermal variables are not independent in tractor operation. Therefore, individual importance rankings should be interpreted as overlapping signals within correlated predictor groups rather than as independent causal effects.

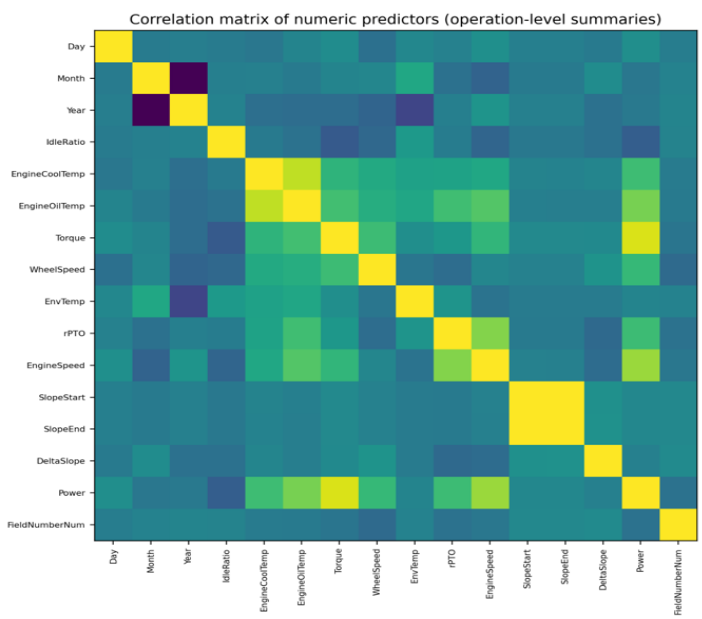

Fig. 2. Correlation matrix of numeric predictors used in the operation-level telemetry summaries.

University of Thi-Qar Journal of agricultural research

Thi-Qar Journal of agricultural research

ISSN Onlin:2708-9347, ISSN Print: 2708-9339 Volume 15, Issue 1 (2026) PP 243-250

https://jam.utq.edu.iq/index.php/main https://doi.org/10.54174/2e0f0q75

Permutation importance results in Fig. 3 showed that powertrain, motion, slope, and activity-related variables contributed to prediction. However, because predictors were correlated, these outputs were used as diagnostic summaries only. This interpretation is consistent with Apley and Zhu (2020) and Molnar et al. (2023), who emphasized that marginal and permutation-based interpretation can be misleading when feature dependence is strong.

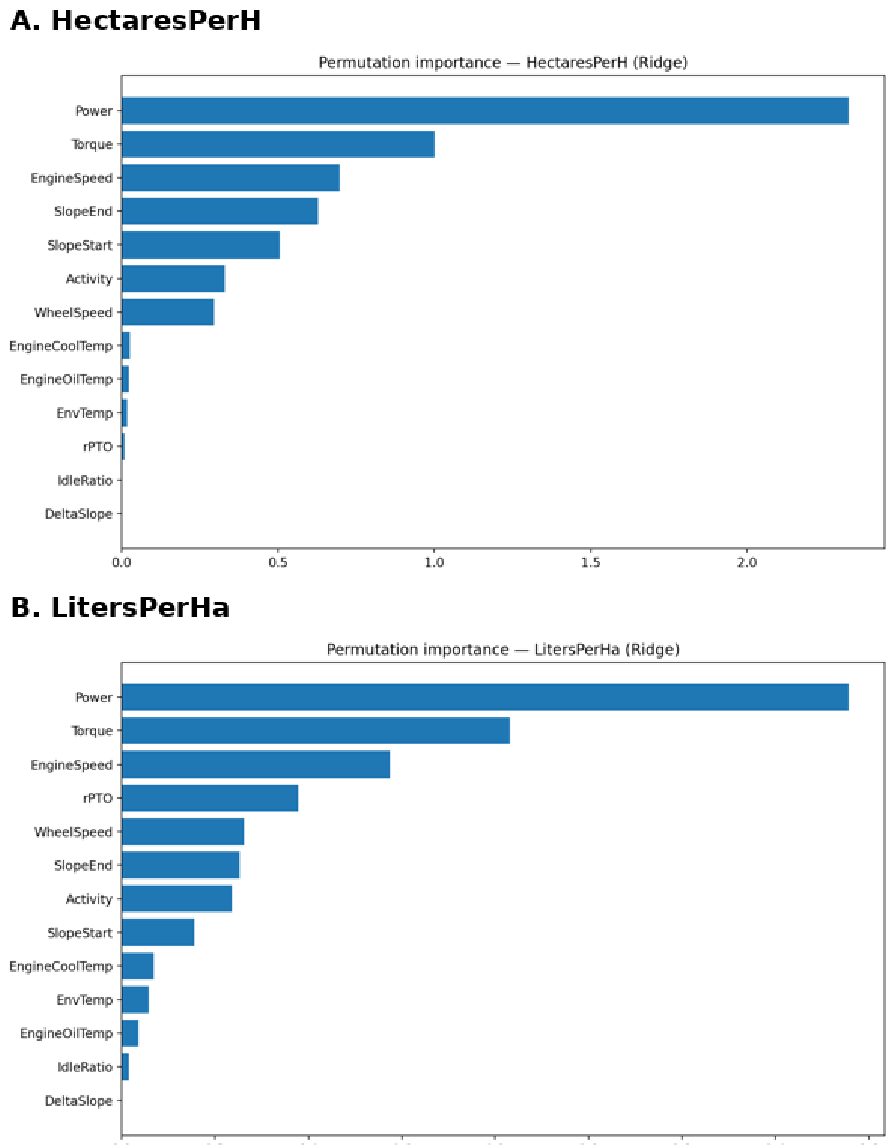

Fig. 3. Permutation importance summaries for Ridge models predicting HectaresPerH and LitersPerHa.

Overall, the findings show that operation-level CAN-Bus and GNSS summaries contain useful information for tractor performance analysis, but the value of machine learning depends on the target. Productivity was captured almost as well by a

ISSN Onlin:2708-9347, ISSN Print: 2708-9339 Volume 15, Issue 1 (2026) PP 243-250

https://jam.utq.edu.iq/index.php/main https://doi.org/10.54174/2e0f0q75

simple activity baseline, while fuel consumption per unit area showed modest improvement from machine learning. The strongest contribution of the study is therefore methodological: it provides an unseen-tractor validation framework that gives a stricter and more realistic estimate of predictive utility than record-level random splitting.

IV. CONCLUSIONS

This study evaluated operation-level prediction of tractor productivity and fuel consumption in vineyard operations using CAN-Bus and GNSS telemetry summaries. Under leave-one-tractor-out validation with nested tuning, the activity-mean baseline performed as well as or slightly better than Ridge and HGBR for HectaresPerH, indicating that activity context captured much of the productivity signal. For LitersPerHa, Ridge and HGBR improved over the baseline, with HGBR achieving the best pooled performance, although the improvement was modest and statistically resonable. The results show that tractor telemetry can support practical machinery-performance benchmarking, but model claims must remain conservative under unseen-tractor evaluation. Future work should expand external validation across more tractors and seasons, compare activity-specific models, and use correlation-aware interpretation methods.

V. REFERENCES

Alemanno, R., Rossi, P., Monarca, D. and Bencini, A. (2025). Evaluation of tractor performance, efficiency and fuel consumption in vineyard activities. Scientific Reports, 15, 8416. https://doi.org/10.1038/s41598-025-93526-z

Angelucci, L. and Mattetti, M. (2024). The development of reference working cycles for agricultural tractors. Biosystems Engineering, 242, 29-37. https://doi.org/10.1016/j.biosystemseng.2024.04.004

Apley, D.W. and Zhu, J. (2020). Visualizing the effects of predictor variables in black box supervised learning models. Journal of the Royal Statistical Society: Series B, 82(4), 1059-1086. https://doi.org/10.1111/rssb.12377

ASABE. (2011). ASABE EP496.3: Agricultural machinery management data. St. Joseph, Michigan: American Society of Agricultural and Biological Engineers.

Friedman, J.H. (2001). Greedy function approximation: A gradient boosting machine. The Annals of Statistics, 29(5), 1189-1232. https://doi.org/10.1214/aos/1013203451.

Götz, K., Kusuma, A., Dörfler, A. and Lienkamp, M. (2025). Agricultural load cycles: Tractor mission profiles from recorded GNSS and CAN bus data. Data in Brief, 60, 111494. https://doi.org/10.1016/j.dib.2025.111494

Hoerl, A.E. and Kennard, R.W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1), 55-67. https://doi.org/10.1080/00401706.1970.10488634.

Kolator, B.A. (2021). Modeling of tractor fuel consumption. Energies, 14(8), 2300. https://doi.org/10.3390/en14082300

Mattetti, M., Maraldi, M., Lenzini, N., Fiorati, S., Sereni, E. and Molari, G. (2021). Outlining the mission profile of agricultural tractors through CAN-BUS data analytics. Computers and Electronics in Agriculture, 184, 106078. https://doi.org/10.1016/j.compag.2021.106078

ISSN Onlin:2708-9347, ISSN Print: 2708-9339 Volume 15, Issue 1 (2026) PP 243-250
https://jam.utq.edu.iq/index.php/main https://doi.org/10.54174/2e0f0q75

Molnar, C., Freiesleben, T., König, G., Casalicchio, G., Wright, M.N. and Bischl, B. (2023). Relating the partial dependence plot and permutation feature importance to the data generating process. In: Explainable Artificial Intelligence. Communications in Computer and Information Science, 1901, pp. 456-479. https://doi.org/10.1007/978-3-031-44064-9_24

Pitla, S.K., Luck, J.D., Werner, J., Lin, N. and Shearer, S.A. (2016). In-field fuel use and load states of agricultural field machinery. Computers and Electronics in Agriculture, 121, 290-300. https://doi.org/10.1016/j.compag.2015.12.023

Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J.J., Schröder, B., Thuiller, W., Warton, D.I., Wintle, B.A., Hartig, F. and Dormann, C.F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography, 40, 913-929. https://doi.org/10.1111/ecog.02881

Rossi, P. and Alemanno, R. (2024). Tractor performances in vineyard operations [Data set]. Mendeley Data. https://doi.org/10.17632/y3h5xzkjc6.1

Varma, S. and Simon, R. (2006). Bias in error estimation when using cross-validation for model selection. BMC Bioinformatics, 7, 91. https://doi.org/10.1186/1471-2105-7-91

Wilcoxon, F. (1945). Individual comparisons by ranking methods. Biometrics Bulletin, 1(6), 80-83. https://doi.org/10.2307/3001968.