



An innovative multivariate tool for fuel consumption and costs estimation of agricultural operations

Mirko Guerrieri, Marco Fedrizzi, Francesca Antonucci, Federico Pallottino, Giulio Sperandio, Mauro Pagano, Simone Figorilli, Paolo Menesatti and Corrado Costa

Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria (CREA), Unità di ricerca per l'ingegneria agraria. Via della Pascolare 16, 00015 Monterotondo scalo (Roma), Italy.

Abstract

The estimation of operating costs of agricultural and forestry machineries is a key factor in both planning agricultural policies and farm management. Few works have tried to estimate operating costs and the produced models are normally based on deterministic approaches. Conversely, in the statistical model randomness is present and variable states are not described by unique values, but rather by probability distributions. In this study, for the first time, a multivariate statistical model based on Partial Least Squares (PLS) was adopted to predict the fuel consumption and costs of six agricultural operations such as: ploughing, harrowing, fertilization, sowing, weed control and shredding. The prediction was conducted on two steps: first of all few initial selected parameters (time per surface-area unit, maximum engine power, purchase price of the tractor and purchase price of the operating machinery) were used to estimate the fuel consumption; then the predicted fuel consumption together with the initial parameters were used to estimate the operational costs. Since the obtained models were based on an input dataset very heterogeneous, these resulted to be extremely efficient and so generalizable and robust. In details the results show prediction values in the test with r always ≥ 0.91 . Thus, the approach may result extremely useful for both farmers (in terms of economic advantages) and at institutional level (representing an innovative and efficient tool for planning future Rural Development Programmes and the Common Agricultural Policy). In light of these advantages the proposed approach may as well be implemented on a web platform and made available to all the stakeholders.

Additional key words: digital agriculture; precision farming; predictive modelling; machine efficiency; economical assessments.

Abbreviations used: CIOSTA (Commission Internationale de l'Organisation Scientifique du Travail en Agriculture); CTF (Controlled Traffic Farming); LV (Latent Vector); PLS (Partial Least Squares); RMSEC (Root Mean Square Error of Calibration); RPD (Ratio of Standard Error of Performance to Standard Deviation); TAV (Tractor Turn-Around Time); TE (Effective Work Time); TN (Net Time); VIP (Variable Importance in the Projection).

Authors' contributions: Conceived and designed the experiments: MG, MF, GS and MP. Analyzed the data: MG, FA, FP, SF, PM and CC. Wrote the paper: MG, MF, FA, FP, GS and CC.

Citation: Guerrieri, M.; Fedrizzi, M.; Antonucci, F.; Pallottino, F.; Sperandio, G.; Pagano, M.; Figorilli, S.; Menesatti, P.; Costa, C. (2016). An innovative multivariate tool for fuel consumption and costs estimation of agricultural operations. Spanish Journal of Agricultural Research, Volume 14, Issue 4, e0209. <http://dx.doi.org/10.5424/sjar/2016144-9490>.

Received: 17 Feb 2016. **Accepted:** 27 Oct 2016.

Copyright © 2016 INIA. This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial (by-nc) Spain 3.0 Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Funding: Agreement (Protocol MiPAAF DG COSVIR 4963 of 02/03/2012) between Italian Ministry of Agriculture, Food and Forestry Policies (MiPAAF), as part of the project "Mo.Na.Co.", and Consiglio per la Ricerca e la Sperimentazione in Agricoltura – Unità di Ricerca di Agrobiologia e Pedologia (CRA-ABP); Italian Ministry of Agriculture, Food and Forestry Policies (MiPAAF) project AGROENER (D.D. n. 26329).

Competing interests: The authors have declared that no competing interests exist.

Correspondence should be addressed to Corrado Costa: corrado.costa@crea.gov.it

Introduction

Nowadays lowering the resources use with practices not affecting the production rates is becoming more and more crucial due to their scarcity and the increasing competitiveness together with the raising awareness of the pressure agriculture has on the envi-

ronment. An example is given by the introduction of precision farming technologies into conventional farm activities that has provided operators the opportunity to cope with in-field variability and to handle and manage the resources and information efficiently (Fountas *et al.*, 2015). Similarly, choosing agricultural machinery of the appropriate size, and being able to predict

in advance the operational costs, can significantly contribute to optimize the resources use. The estimation of operating costs of agricultural and forestry (Verani *et al.*, 2015) machinery and the definition of economic competitiveness gap (conditionality standards on agricultural farms and short- and medium-term business planning of the agricultural farm activities) are a key factor in planning policies of Rural Development Programs (RDPs) both at national and European level.

Several authors developed different approaches and methods for cost estimation and calculation of mechanized farming operations in relation to their specific needs. Lazzari & Mazzetto (1996) developed the Computed Farm Machinery System model (ComFARMS) to analyse arable farm mechanisation problems from strategic or management standpoints. In this context, the purpose is to enable the user to carry out sensitivity analysis by modifying input off-farm and on-farm data and to perform multi-criteria methods that include subjective and non-numeric information. Moreover, Sogaard & Sørensen (2004) developed a non-linear programming model implemented by using the programming software suite General Algebraic Modelling System (GAMS) to support the process of choosing the optimal level of farm mechanisation in terms of technical capability. It is based on a least-cost concept, which involves all expected fixed and variable costs for a particular farm size and crop plan Camarena *et al.* (2004) developed an integrated approach called MULTIPREDIO created mixed integer linear programming linked to several databases contained in spreadsheets to select agricultural machinery for a multifarm system. Their approach aims to select the optimal machinery set for each farm, which corresponds to the lowest annual mechanisation cost of the multifarm system through the time. Another method has been developed by Gunnarsson & Hansson (2004) to examine different harvesting systems, estimate timeliness costs and to present conclusions on harvesting machinery selection. In the study, the direct machine costs (including both, fixed and variable) were calculated with conventional methods using parameters from the ASAE Standards (ASAE D497.4, 2000). Bochtis *et al.* (2010) developed a targeted approach for the estimation of the operational machinery costs on an annual basis in controlled traffic farming (CTF) systems. Their approach combines four sub-models based on specific algorithmic approaches in order to evaluate the consequences in terms of machinery performance (following different driving directions) when establishing tramlines in a CTF system. The work used specific equation given in the Agricultural Machinery Management Data ASAE Standard (ASAE D497.6, 2009).

Spinelli *et al.* (2011) carried out a study in order to calculate the economic life, the annual use and the residual value of two harvesting machines of forestry products. This was done starting from a large database (coming from Europe and North America) of second-hand machine sale offers containing over 1000 records. The information contained in the study are crucial for machine rate calculation often based on rule-of-thumb assumptions, following the absence of empirical data. In the study of De Toro *et al.* (2012) a model capable to predict the moisture content of wheat using historical weather data, was used to assess the effects of weather on cereal harvesting costs (machine, labour, timeliness, drying). The specific machine costs (*i.e.*, for the combine harvester) were estimated using ASAE standard methods (ASAE D497.5, 2006; ASAE EP496.3, 2006) with the following parameters: depreciation, residual value, annual use and the economic life.

In spite of the numerous elements of randomness illustrated above, traditional estimation of operating costs of agricultural machinery is largely based on deterministic models, which perform the same way for a given set of initial conditions (Abramo *et al.*, 2015). Conversely, in the statistical model with respect to the deterministic ones, randomness is present and variable states are not described by unique values, but rather by probability distributions.

In this study, a Partial Least Squares (PLS) multivariate statistical model was adopted to predict the fuel consumption and costs of six agricultural operations such as: ploughing, harrowing, fertilization, sowing, weed control and shredding. This prediction has been done starting from some selected parameters, which are: soil workability, non-working distance travelled, time per surface-area unit, maximum engine power, purchase price of the tractor and purchase price of the operating machinery in order to optimize crucial agricultural operations and thus increase the farm performance.

Material and methods

Data collection

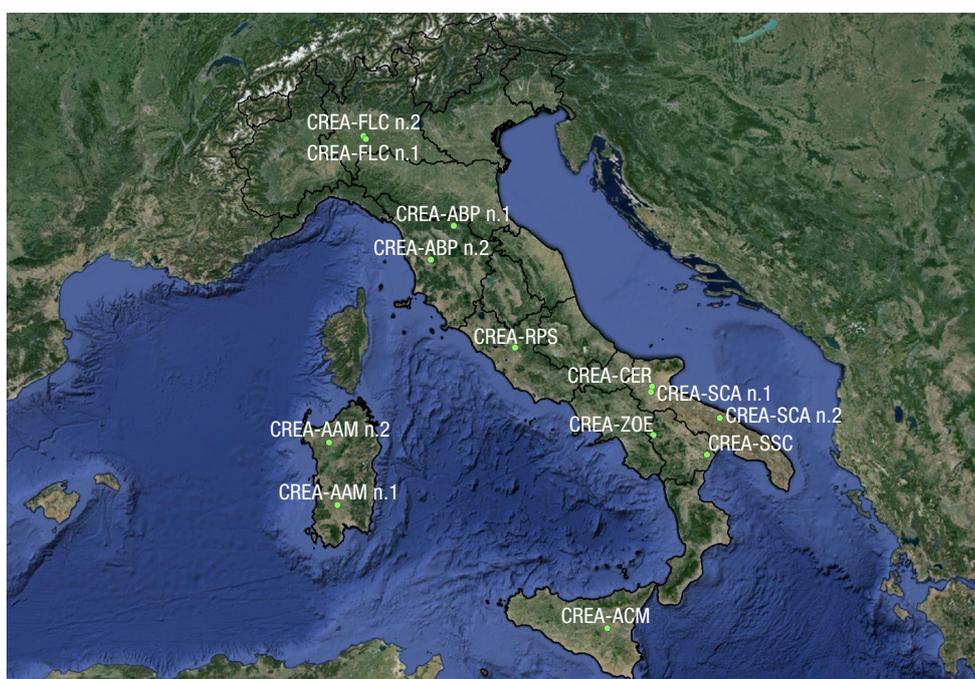
Data were collected from 2011 to 2014 in several experimental fields of CREA located in different parts of the Italian country (Fig. 1). The soils of the various experimental fields have different characteristics regarding slope, texture, shape, surface and crop grown (Table 1).

The different agricultural operations are in details: ploughing (54 observations), harrowing (70 observa-

Table 1. Soil characteristics of experimental fields regarding: number of sites, slope (%), texture (%), shape, surface (m²) and crop grown.

Experimental field	N. sites	Soil slope (%)	Soil texture (%)			Soil shape	Soil surface (m ²)	Crop grown
			Sand	Silt	Clay			
CREA-AAM	1	0	43.2	25.6	31.2	Rectangular	3200	Common wheat grain/Egyptian Clover
CREA-AAM	2	3	—	—	—	Rectangular	1155	Oats herbage
CREA-ABP	1	14.8	5.5	50.1	44.4	Trapezoidal	22320	Sunflower/field bean/common wheat grain
CREA-ABP	2	13	20.0	37.6	42.4	Trapezoidal	18330	Common wheat grain
CREA-ACM	1	0	—	—	—	Rectangular	5000	Durum wheat grain
CREA-CER	1	1	19.1	43.3	37.6	Rectangular	5000	Durum wheat grain
CREA-FLC	1	0	73.0	17.0	10.0	Rectangular	10736	Corn grain/soybean
CREA-FLC	2	0	43.0	39.0	18.0	Trapezoidal	3500	Corn grain
CREA-RPS	1	6	32.9	46.2	20.9	Rectangular	1920	Common wheat grain/field bean
CREA-SCA	1	0	19.5	31.1	49.4	Rectangular	4900	Durum wheat grain
CREA-SCA	2	3	13.2	32.7	54.1	Rectangular	1800	Durum wheat grain
CREA-SSC	1	0	19.0	39.0	42.0	Rectangular	1270	Durum wheat grain
CREA-ZOE	1	3	—	—	—	Trapezoidal	25000	Mix herbage of vetch and oats

—: Unavailable.

**Figure 1.** CREA-AAM n.1, Azienda Podere “Ortigara”, Località Sanluri Scalo (VS); CREA-AAM n.2, Azienda Canu Agro, Siligo (SS); CREA-ABP n.1 Azienda Fagna (Scarperia, FI); CREA-ABP n.2 Azienda Santa Elisabetta, Località Vicarello (Volterra, PI); CREA-ACM, Azienda Libertinia, Catenanuova (EN); CREA-CER, Azienda Manfredini, Località Manfredini, Foggia (FG); CREA-FLC n.1, Azienda Cascina Baroncina, Lodi (LO); CREA-FLC n.2 Azienda Arcagna, Montanaso Lombardo (LO); CREA-RPS, Azienda Tormancina, Loc.Tormancina (Roma); CREA-SCA n.1, Azienda Podere 124, Foggia (FG); CREA-SCA n.2, Azienda Agostinelli, Rutigliano (FG); CREA-SSC, Azienda Campo7, Metaponto (MT); CREA-ZOE, Azienda Bella, Bella (PZ).

tions), fertilization (65 observations), sowing (43 observations), weed control (22 observations) and shredding (19 observations). The agricultural operations were carried out using tractors with different brand and power and operating machineries with different life time, annual machine use, repair and maintenance fac-

tor and accumulated use hours of machine (Fedrizzi *et al.*, 2015). In Table 2, brand and model, engine power of the tractors for different operations are reported.

The work times, at each agricultural operation, were recorded following the recommendations of the Italian Rural Engineering Association (AIGR), which considers

Table 2. Model and engine power of the machines used in the different operations.

Experimental field	Brand and model	Engine power (kW)	Operations
CREA-AAM	Fiat 100 90 DT15	74	Ploughing, harrowing, sowing
CREA-ABP	Same Panther 90	65	Ploughing
	Deutz-Fahr TTV 1145	117	Ploughing
	Fiat 70 CI	54	Ploughing
	Same Kripton 105	77	Harrowing, sowing, shredding
	Same Explorer II 60	44	Fertilization, weed control
CREA-ACM	Agriful 82 C FD L	59	Ploughing
	Goldoni Compact 614.D	45	Fertilization, sowing, weed control
CREA-CER	Fiat Turbo DT 160/90	118	Ploughing
	Fiatagri DT 80-66S	59	Shredding
CREA-FLC	Case 86	63	Ploughing, harrowing
	Fiat 35-66	26	Fertilization
	McCormik	74	Weed control, shredding
	Ford 4610	45	Sowing
CREA-RPS	New Holland TM 135	101	Ploughing
	John Deere 6210	67	Fertilization, sowing, weed control
CREA-SCA	Massey Ferguson 5455	73	Ploughing, harrowing
	Fiat 160/90	118	Ploughing, harrowing
	Deutz-Fahr Agrottron 200	153	Ploughing
	Landini 6500 Mark II	47	Harrowing
	New Holland TN60DA	44	Harrowing
	Fiat 600	44	Fertilization, sowing, weed control
CREA-SSC	New Holland TN80F DT	59	Ploughing, harrowing, fertilization, sowing, weed control
CREA-ZOE	John Deere 6010 AS/1	66	Ploughing, harrowing
	Fiat 540	40	Sowing

the official methodology of the “Commission Internationale de l’Organisation Scientifique du Travail en Agriculture” (CIOSTA) (Manfredi, 1971; Biondi, 1999). In this study, in order to significantly reduce the influence due to the different characteristics of the farms, in terms of type of machinery used, distance of the fields and so on, only some items from the CIOSTA official methodology were considered.

The plots of land of the agricultural holdings examined differ in shape and size, geomorphology, soil composition and geographic location, as well as agronomical and administrative management. In order to significantly reduce the influence induced by this vast variability of agricultural holding characteristics, only several items of the CIOSTA method were taken into account.

Since the small area of experimental fields (about 0.5 ha) for the collection of work times (as reported by CIOSTA), only those related to the effective work time (TE) and to the tractor turn-around time (TAV) (which together represent the net time, TN) were considered. Considering the TN, the hourly operating cost of each tractor and piece of machinery used, was determined by means of specific analytical methods, and successively, the cost *per* surface-area unit was determined for each agricultural operation.

The operating costs, for each agricultural operation, of the tractors and operating machineries, were identi-

fied by considering two main parameters: fixed and variable costs. The former involve the reintegration of the invested capital, the cost of capital using, and the various expenses (insurance, storage and taxes). The variable costs were related to the use of the agricultural machinery and include the expenses incurred for repairs and maintenance, fuel, lubricants and labour. Relatively to the life time and annual use for all tractors, a standard value of 15 years and 1067 hours per year respectively are considered.

The methods proposed in the bibliography are substantially similar in relation to the calculation of the fixed costs, whereas they differ in the formulas and coefficients adopted in calculating the variable costs. As far as this last item is concerned, reference has been made to a specific method (Biondi, 1999). In this study, the method proposed by Biondi (1999) has been updated with precise references to the technical standards that indicate the technical and economic coefficients to be used in the calculations (ASAE D497.4, 2003; ASAE EP496.2, 2003).

The maximum engine power (P; kW), and the purchase price of the tractor (€) were obtained from the publication “Buyers’ guide 2013” (Guida all’acquisto 2013), edited by the Italian magazine “L’informatore Agrario” (<http://www.informatoreagrario.it/ita/riviste/infoagri/131a19/sommario.asp>).

Table 3. Factors used for the fuel consumption calculation in relation to the different operations extracted from Biondi's (1999) methodology.

Processing	Engine power factor in the effective work time (d_{ie}) (%)	Engine power factor in the turning accessory time (d_{lav}) (%)	Specific diesel fuel consumption to the maximum engine power (Sc) (kg/kWh)
Ploughing	85	30	0.25
Fertilizing	40	30	0.25
Weed control	40	30	0.25
Harrowing	70	30	0.25
Sowing	50	30	0.25
Shredding	75	30	0.25

Furthermore, it was also necessary to carry out an economic assessment of all the operating machinery (ploughs, harrows, seeders, fertiliser spreaders, etc.) used in the cultivation activities. The purchase prices of the various machines was determined when possible from the producers price list otherwise contacting specific retailers through personal communications.

The fuel consumption *per* hectare (modified from Biondi, 1999) was calculated considering, first of all, the hourly fuel consumption (Fc_h ; kg/h) per each agricultural operation using the following formula:

$$Fc_h = Sc \cdot P \cdot d \quad [1]$$

where, Sc = specific fuel consumption in kg/kWh; P = maximum engine power in kW; and d = power utilisation factor in %.

Then, d was described considering two different conditions: d_{ie} is the power used during effective operation and d_{lav} is the power used during the turn-around operations and manoeuvres. By this way Eq. [1] becomes:

$$Fc_h = (Sc \cdot P \cdot d_{ie}) + (Sc \cdot P \cdot d_{lav}) \quad [2]$$

To calculate the fuel consumption per hectare (Fc_{ha} ; kg/ha) we modified [2] as follows:

$$Fc_{ha} = T_e \cdot (Sc \cdot P \cdot d_{ie}) + T_i \cdot (Sc \cdot P \cdot d_{lav}) \quad [3]$$

where T_e = effective time consumed during the operation in hours; and T_{lav} = time consumed during the turn-around operations and manoeuvres in hours.

The methodology used for calculating tractors and machineries operational costs, is referred to the one used by Fedrizzi *et al.* (2015).

The values used for the specific fuel consumption (Sc), and the power utilisation factor (d), related to different examined operations are reported in Table 3. The value ranges are derived according to Biondi (1999) methodology. In the development of the exposed calculations, it refers exclusively to the aspects relating

to costs directly connected with the use of machines and equipment excluding the costs of other inputs such as fertilizer, seed and pesticides.

Multivariate modelling

A multivariate modelling approach was adopted to predict six agricultural operations, fuel consumption and costs. A two-step approach was applied.

In the first step, the fuel consumption for each agricultural operation was predicted from the first four variables: time per surface-area unit (h/ha), maximum engine power (kW), purchase price of the tractor (€) and purchase price of the operating machinery (€). Only for the ploughing fuel consumption two additional dummy variables were considered: soil workability (high = 1, low = 0); and minimization of the tractor non-working distance travelled (optimized = 0; not optimized = 1, where not optimized regards outward with ploughing and return without ploughing).

In the second step, the costs for each agricultural operation were predicted from the four above mentioned variables and the fuel consumption predicted as a result of the first step. A PLS regression approach was applied (*e.g.*, Wold *et al.*, 2001; Costa *et al.*, 2013; Infantino *et al.*, 2015; Cutini *et al.*, 2016) on the above mentioned datasets. The PLS is a particular type of multivariate regression which uses a two-block predictive PLS model. The regression analysis objective is achieved by using the equation that minimizes the residual mean square error, or maximizes the coefficient of multiple determination r^2 , which is the most commonly used statistic to measure the forecasting potential of a multiple regression equation. The predictive ability of the model also depends on the number of latent vectors (LV) used. Generally, a good predictive model should have high values of the Pearson correlation coefficient (r) and low values for the root mean square error in calibration (RMSEC). The procedure calculated the ratio of percentage deviation (RPD), which is the ratio of the standard deviation of the measured data to the RMSE

(Williams, 1987). The model chosen was for the number of LV that yielded the highest r , minimum standard error of prevision (SEP) for predicted and known Y-block and maximum RPD. The PLS models were developed starting from a calibration set (training/evaluation set; Forina *et al.*, 2008), consisting of 50% of the sample. The PLS model (cross-validated) was then validated on a set on internal tests consisting of the remaining 50% of the samples. The partitioning was carried out using the sample set partitioning-based on joint X-Y distances (SPXY) algorithm (Harrop Galvao *et al.*, 2005) that takes account of the variability in both X and Y. The independent variables (X-block) datasets were standardised using an autoscale algorithm (*i.e.*, centers columns to zero mean and scales to unit variance). For each PLS model a summary of the relative importance of the independent variables (Y-block) to predict the dependent one is given by variable importance in the projection (VIP) (Febbi *et al.*, 2015; Taiti *et al.*, 2015). The VIP scores estimate the importance of each variable in the PLS-based models and were calculated according to Chong & Jun (2005). The PLS models were developed using a procedure written in the MATLAB 7.1 R14 environment.

The obtained PLS models were applied to a standard sized Italian farm: <10 hectares farm with two tractors with reduced power (118 and 59 kW). This in order to calculate costs per hectare and fuel consumption for the six operations examined.

Results

The performance of PLS models with different LVs in the determination of the fuel consumption for each agricultural operation is summarized in Table 4. The cumulated variance for the X-blocks ranged from

56.9% for weed control to 99.88% for fertilization. For the Y-block the cumulated variances ranged from 5.7% for sowing to 15.85% for fertilization. The RMSEC unbiased values had the same unit range of the original data and were always very low. Considering correlation coefficients in the calibration/validation set, these were always very high ($r > 0.96$); the same result could be observed for test set showing very high performances (> 0.94), except from fertilization (0.91).

Figure 2 shows the scatter plot of the observed versus predicted (biased) values by the PLS models in estimating fuel consumption at different agricultural operations. For both datasets, the biased observations are well distributed along the bisectrix, indicating a good performance in predicting the fuel consumption.

The VIP scores obtained by the PLS regressions to estimate fuel consumption of the different agricultural operations are showed in Table 5. It is possible to observe as, for all the agricultural operations, the time per surface-area unit (h/ha) is the most important variable in predicting fuel consumption. For ploughing, also the soil workability is an important variable in prediction.

The performance of the PLS models with different LVs in the determination of the costs for each agricultural operation, is summarized in Table 6. The cumulated variances for the X-blocks ranged from 96.67% for sowing to 100% for ploughing and shredding. For the Y-block the cumulated variances ranged from 9.82% for weed control to 20.29% for fertilization. The RMSEC unbiased values had the same unit range of the original data, and were very low. Considering the correlation coefficients in the calibration/validation set, these were always very high ($r > 98$); the same result could be observed for test set showing very high performances (> 98).

Figure 3 shows the scatter plot of the observed versus predicted (biased) values by the PLS models in

Table 4. Descriptors and principal results of the Partial Least Squares (PLS) regression model in estimating fuel consumption at different agricultural operations.

PLS descriptors	Agricultural operations					
	Ploughing	Harrowing	Fertilization	Sowing	Weed control	Shredding
Number of samples	54	70	65	43	22	19
Number of variables	6	4	4	4	4	4
Number of LVs ^[1]	3	3	3	3	3	3
% Cumulated variance X-block	89.3	99.85	99.88	99.59	56.9	99.83
% Cumulated variance Y-block	15.28	14.27	15.85	5.77	14.67	15.42
RMSEC ^[2]	46.65	10.29	0.91	4.91	0.99	13.50
Bias	-46.42	-10.26	-0.90	-4.90	-0.99	-13.45
r model ^[3]	0.97	0.98	0.97	0.96	1.00	0.98
r test ^[4]	0.99	0.94	0.97	0.96	0.91	0.96

^[1] LV: latent vector. ^[2] RMSEC: root mean squares error in calibration. ^[3] Model test correlation coefficient. ^[4] Model and test correlation coefficient.

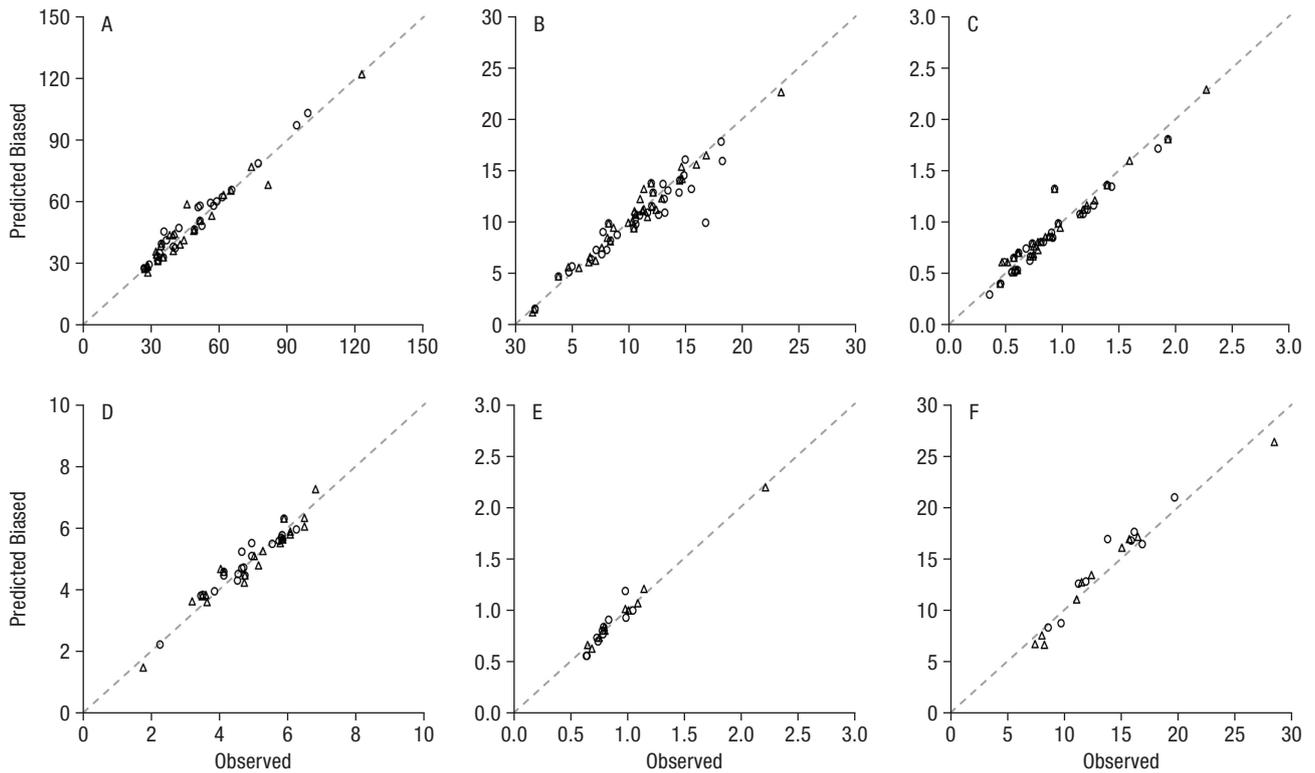


Figure 2. Scatter plot of the observed versus predicted (biased) values by the Partial Least Squares (PLS) models in estimating fuel consumption at different agricultural operations (A, ploughing; B, harrowing; C, fertilization; D, sowing; E, weed control; F, shredding). Triangles indicate the 50% samples used to calibrate/validate the models. Circles indicate 50% samples used to test the model. Dashed line represents the bisectrix (*i.e.*, perfect attribution).

Table 5. Variable importance in the projection (VIP) scores of the partial least squares (PLS) regression models in estimating fuel consumption at different agricultural operations.

VIP	Agricultural operations					
	Ploughing	Harrowing	Fertilization	Sowing	Weed control	Shredding
Soil workability	1.06	–	–	–	–	–
Non-working distance travelled	0.00	–	–	–	–	–
Time per surface-area unit	3.87	3.13	2.92	1.99	2.30	1.50
Maximum engine power	0.44	0.15	0.43	0.71	0.49	0.75
Purchase price of the tractor	0.20	0.08	0.40	0.58	0.51	0.91
Purchase price of the operating machinery	0.44	0.65	0.24	0.73	0.69	0.83

Table 6. Descriptors and principal results of the Partial Least Squares (PLS) regression model in estimating costs at different agricultural operations.

PLS descriptors	Agricultural operations					
	Ploughing	Harrowing	Fertilization	Sowing	Weed control	Shredding
Number of samples	54	70	65	43	22	19
Number of variables	7	5	5	5	5	5
Number of LVs ^[1]	5	3	4	3	4	4
% Cumulated variance X-block	100	99.88	99.91	99.67	99.98	100
% Cumulated variance Y-block	13.88	16.87	20.29	10.3	9.82	9.84
RMSEC ^[2]	219.69	50.21	6.96	39.51	6.83	64.96
Bias	-219.40	-50.02	-6.94	-39.45	-6.83	-64.93
<i>r</i> model ^[3]	0.99	0.98	0.99	0.99	1.00	1.00
<i>r</i> test ^[4]	0.99	0.98	0.99	0.99	0.99	1.00

^[1] LV: latent vector. ^[2] RMSEC: root mean squares error in calibration. ^[3] Model test correlation coefficient. ^[4] Model and test correlation coefficient.

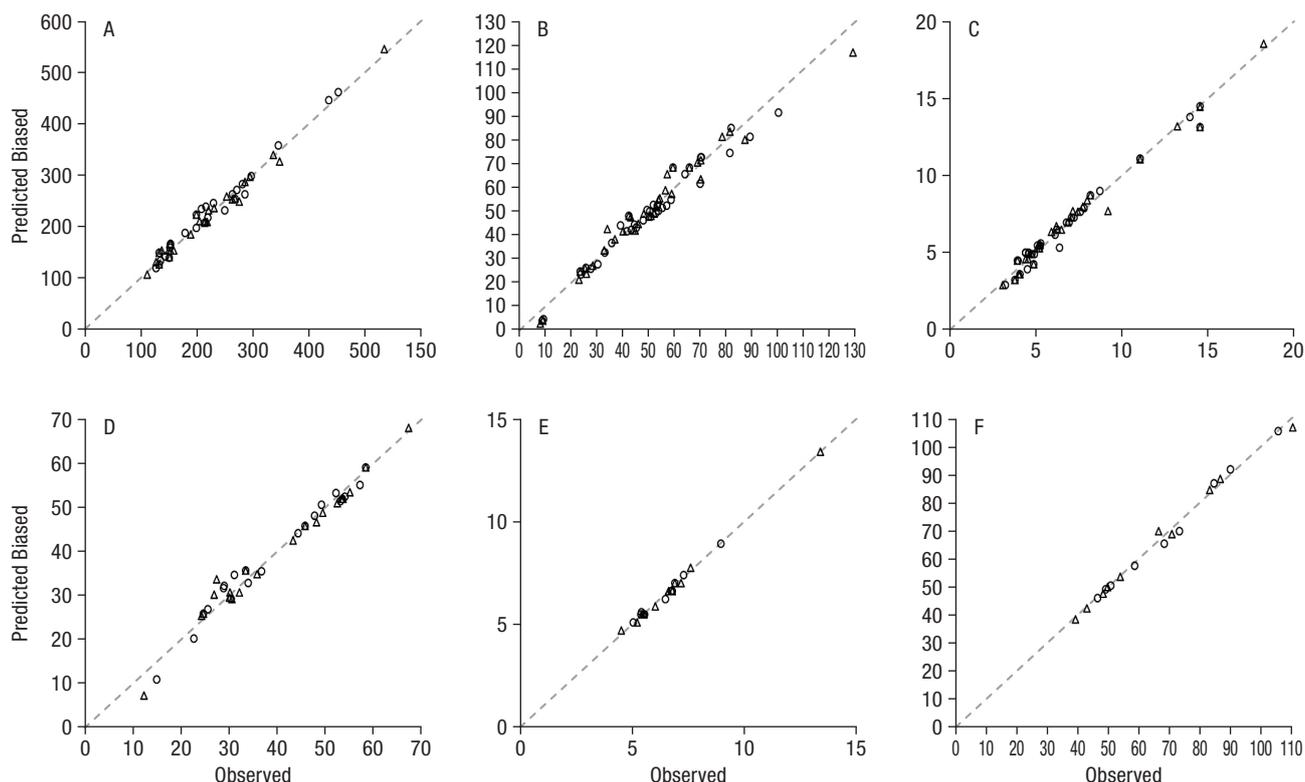


Figure 3. Scatter plot of the observed versus predicted (biased) values by the Partial Least Squares (PLS) models in estimating costs at different agricultural operations (A, ploughing; B, harrowing; C, fertilization; D, sowing; E, weed control; F, shredding). Triangles indicate the 50% samples used to calibrate/validate the models. Circles indicate 50% samples used to test the model. Dashed line represents the bisectrix (*i.e.*, perfect attribution).

Table 7. Variable importance in the projection (VIP) scores of the partial least squares (PLS) regression models in estimating costs at different agricultural operations.

VIP	Agricultural operations					
	Ploughing	Harrowing	Fertilization	Sowing	Weed control	Shredding
Soil workability	0.67	—	—	—	—	—
Non-working distance travelled	0.00	—	—	—	—	—
Time per surface-area unit	2.43	2.18	1.60	1.60	1.95	2.14
Maximum engine power	0.22	0.01	0.96	0.14	0.19	0.30
Purchase price of the tractor	0.21	0.01	0.97	0.11	0.17	0.36
Purchase price of the operating machinery	0.43	0.36	0.08	1.25	0.65	0.48
Fuel consumption	3.03	2.43	1.39	1.90	2.03	1.71

estimating costs at different agricultural operations. For both datasets, the biased observations were well distributed along the bisectrix, indicating a good performance in predicting the costs.

The VIP scores obtained by the PLS regressions to estimate costs of the different agricultural operations are showed in Table 7. It is possible to observe as, for ploughing, harrowing, sowing and weed control the fuel consumption was the most important variable. Equally, the time per surface-area unit (h/ha) was the most important variable for fertilization and shredding. Sowing showed higher values for the purchase price

of the operating machinery, differently from the others (Table 7).

An example of application of the proposed model for a standard sized Italian farm (<10 hectares farm with a reduced mechanization), to calculate costs per hectare and fuel consumption for the six operations examined, is shown in Table 8. The proposed model, differently from the results obtained according to Biondi (1999) methodology, allowed a more simpler calculation of consumption, based on the most simple and expeditious parameters already known by the farmer, as for example the tractor

Table 8. Example of application of the prediction model of fuel consumption and unit costs for the six operations in a standard Italian farm.

Operations	Equipment type	Tractor engine power (kW)	Total work time (h/ha)	Tractor price (€)	Equipment price (€)	Fuel consumption prediction (kg/ha)	Mechanical operation cost (€/ha)
Ploughing	Double reversible plough	118.00	2.70	108307	5162.00	50.91	208.26
Fertilizing	Centrifugal fertilizer spreader	58.82	0.18	54532	500.00	1.13	6.50
Weed control	Sprayer	58.82	0.19	54532	2215.00	1.32	12.92
Harrowing	Disc harrow	58.82	1.39	54532	5060.00	12.91	63.93
Sowing	Pneumatic seed drill	58.82	0.89	54532	17000.00	1.36	53.49
Shredding	Shredder	58.82	1.64	54532	1000.00	1.83	99.44

engine power, total work time, tractor and equipment price.

Discussion

The monitoring of agricultural operations, in particular the costs and fuel consumption of the machinery, is a large important portion for the economical farm balance. In order to obtain an uniform procedure for machinery cost analysis, this study adopt standard models to predict fuel consumption and costs of six agricultural operations, such as ploughing, harrowing, fertilization, sowing, weed control and shredding, on the base of six variables.

For the first time, a two-step approach was applied to predict six agricultural operations fuel consumption and, then, from esteem fuel consumption, the costs in agricultural engineering. The following variables were used to develop a predictive model and thus for the estimation of costs and fuel consumption: time per surface-area unit; maximum engine power; purchase price of the tractor and of the operating machinery; soil workability; and non-working distance travelled. This method on one side uses fewer variables with respect to the ASAE one (ASAE EP496.2, 2003) or other methodologies, and on the other side can be easily utilized by farmers. In fact, the farmer knows the surface-area of their plots, the technical and economic aspects of their tractors and operating machines, and the average time required to perform the various growing operations. Since this method requires as input only this information, its use is greatly simplified if compared to other systems and methodologies more articulated that require numerous other qualified input factors. In this scenario, a great number of authors have developed different modelling approaches for cost and fuel consumption estimation and calculation of mechanized farming operations.

The majority of these methods is based on quantitative agricultural analysis (*e.g.*, income, prices, farm size, efficiency, factor allocation, production, welfare,

etc.) at different levels of scale (Happe *et al.*, 2006). In this study, the analysis is based on a multivariate statistical model, in which randomness is present and variable states are described by probability distributions, with respect to a deterministic one. As reported by Costa *et al.* (2012), a multivariate regression approach is particularly useful when predicting one or more dependent variables from a large set of independent variables, often collinear. In fact, this work adopted a multivariate modelling approach to predict agricultural operations, fuel consumption and costs applying two steps. In the first, the fuel consumption for each agricultural operation was predicted from the above mentioned variables meanwhile in the second, the costs for each agricultural operation were predicted from these variables and the fuel consumption predicted as a result of the first step. All the models constructed for the prediction of fuel consumption resulted to be highly performant.

Considering the correlation coefficients, the r values for test set were > 0.94 , except for fertilization (0.91). Considering also the performances in the test set of the model constructed to predict costs, the r values were all > 0.96 .

Using this approach, it is possible to observe also the importance of the variables in the prediction with the analysis of the VIP scores. For ploughing, harrowing, fertilization, weed control, sowing and shredding, the time per surface-area unit is the most important variable in predicting fuel consumption. For ploughing, also the soil workability is an important variable in prediction. In addition, for ploughing, harrowing, sowing and weed control the fuel consumption is the most important variable in predicting costs. Equally, the time per surface-area unit is the most important variable for fertilization and shredding. Sowing showed higher values for the purchase price of the operating machinery, differently from the others. This is due to the high initial cost of purchase (ranging from €500 to €18,000) and for the oversizing in terms of technical and economic aspects of the operating machineries.

The provisional models applied to a standard Italian farm (<10 hectares farm with a reduced mechanization), showed the possibility to apply such an approach to generate scenarios different purposes, from policy makers to agricultural operators to farm contractors.

In conclusion, the obtained models being based on an input dataset very heterogeneous (in terms of field: shape, dimensions, slope, texture, surface, crop grown etc. and in terms of machines/operators) resulted to be extremely efficient and so generalizable and robust. This represents a crucial characteristic needed to transfer the approach from theory to practice.

The advantages of the proposed predictive model are related to the simplicity for the farmers and policy makers to acquire the necessary information. In fact, to calculate consumption and cost per hectare, is sufficient to know the tractor engine power, total work time, tractor and equipment price. In this way, it is possible to get the desired results, without *a priori* knowledge of a great number of parameters and elements as required by the traditional methodologies.

This approach may results extremely useful for both farmers (in terms of economic advantages, *e.g.* while choosing correct machines dimension and power or operations scheduling and management) and at institutional level representing an innovative and efficient tool in order to define correctly the gasoline reimbursement or planning future Rural Development Programmes and the Common Agricultural Policy. In light of these advantages, the approach may be implemented on a web platform (including specific app) and made available to all the stakeholders.

References

- Abramo G, Costa C, D'Angelo CA, 2015. A multivariate stochastic model to assess research performance. *Scientometrics* 102: 1755-1772. <http://dx.doi.org/10.1007/s11192-014-1474-5>
- ASAE D497.4, 2000. Agricultural machinery management data. In: ASAE Standards, pp: 350-357. Am Soc Agric Biol Eng, St. Joseph, MI, USA.
- ASAE D497.4, 2003. Agricultural machinery management. In: ASAE Standards, pp: 373-380. Am Soc Agric Biol Eng, St. Joseph, MI, USA.
- ASAE D497.5, 2006. Agricultural machinery management data. In: ASAE Standards, pp: 390-398. Am Soc Agric Biol Eng, St. Joseph, MI, USA.
- ASAE D497.6, 2009. Agricultural machinery management data. In: ASAE Standards, pp: 360-367. Am Soc Agric Biol Eng, St. Joseph, MI, USA.
- ASAE EP496.2, 2003. Agricultural machinery management. In: ASAE Standards, pp: 367-372. Am Soc Agric Biol Eng, St. Joseph, MI, USA.
- ASAE EP496.3, 2006. Agricultural machinery management. In: ASAE Standards, pp: 384-390. Am Soc Agric Biol Eng, St. Joseph, MI, USA.
- Biondi P. (Ed.). 1999. *Meccanica agraria: le macchine agricole*. UTET, Torino (Italy). pp: 547-561.
- Bochtis DD, Sørensen CG, Busato P, Hameed IA, Rodias E, Green O, Papadakis G, 2010. Tramline establishment in controlled traffic farming based on operational machinery cost. *Biosyst Eng* 107: 221-231. <http://dx.doi.org/10.1016/j.biosystemseng.2010.08.004>
- Camarena EA, Gracia C, Cabrera Sixto JM, 2004. A mixed integer linear programming machinery selection model for multifarm systems. *Biosyst Eng* 87: 145-154. <http://dx.doi.org/10.1016/j.biosystemseng.2003.10.003>
- Chong IG, Jun CH, 2005. Performance of some variable selection methods when multicollinearity is present. *Chemom Intell Lab Syst* 78: 103-112. <http://dx.doi.org/10.1016/j.chemolab.2004.12.011>
- Costa C, Menesatti P, Spinelli R, 2012. Performance modelling in forest operations through partial least square regression. *Silva Fennica* 46 (2): 241-252. <http://dx.doi.org/10.14214/sf.57>
- Costa C, Antonucci F, Boglione C, Menesatti P, Vandeputte M, Chatain B, 2013. Automated sorting for size, sex and skeletal anomalies of cultured seabass using external shape analysis. *Aquacult Eng* 52: 58-64. <http://dx.doi.org/10.1016/j.aquaeng.2012.09.001>
- Cutini M, Costa C, Bisaglia C, 2016. Development of a simplified method for evaluating agricultural tractor's operator whole body vibration. *J Terramechanics* 63: 23-32. <http://dx.doi.org/10.1016/j.jterra.2015.11.001>
- De Toro A, Gunnarsson C, Lundin G, Jonsson N, 2012. Cereal harvesting - Strategies and costs under variable weather conditions. *Biosyst Eng* 111: 429-439. <http://dx.doi.org/10.1016/j.biosystemseng.2012.01.010>
- Febbi P, Menesatti P, Costa C, Pari L, Cecchini M, 2015. Automated determination of poplar chip size distribution based on combined image and multivariate analyses. *Biomass Bioenerg* 73: 1-10. <http://dx.doi.org/10.1016/j.biombioe.2014.12.001>
- Fedrizzi M, Sperandio G, Guerrieri M, Pagano M, Costa C, Puri D, Fanigliulo R, Bazzoffi P, 2015. Economic competitiveness gap related to the application of the conditionality standards on agricultural farms: evaluation methodology. *Ital J Agron* 10 (s1): 696.
- Forina M, Oliveri P, Lanteri S, Casale M, 2008. Classification techniques, classic and new, for old and new problems. *Chemom Intell Lab Syst* 93: 132-48. <http://dx.doi.org/10.1016/j.chemolab.2008.05.003>
- Fountas S, Sorensen CG, Tsiropoulos Z, Cavalaris C, Liakos V, Gemtos T, 2015. Farm machinery management information system. *Comput Electron Agr* 110: 131-138. <http://dx.doi.org/10.1016/j.compag.2014.11.011>
- Gunnarsson C, Hansson PA, 2004. Optimisation of field machinery for an arable farm converting to organic farming. *Agric Syst* 80: 85-103. <http://dx.doi.org/10.1016/j.agsy.2003.06.005>
- Happe K, Kellermann K, Balmann A, 2006. Agent-based analysis of agricultural policies: an illustration of the

- agricultural policy simulator AgriPoliS, its adaptation and behavior. *Ecol Soc* 111: 49.
- Harrop Galvao RK, Ugulino Araujo MC, Emidio Jose G, Coelho Pontes MJ, Cirino Silva E, Bezerra Saldanha TC, 2005. A method for calibration and validation subset partitioning. *Talanta* 67: 736-740. <http://dx.doi.org/10.1016/j.talanta.2005.03.025>
- Infantino A, Aureli G, Costa C, Taiti C, Antonucci F, Menesatti P, Pallottino F, De Felice S, D'Egidio MG, Mancuso S, 2015. Potential application of PTR-TOFMS for the detection of deoxynivalenol DON in durum wheat. *Food Control* 57: 96-104. <http://dx.doi.org/10.1016/j.foodcont.2015.03.047>
- Lazzari M, Mazzetto F, 1996. A PC model for selecting multicropping farm machinery systems. *Comput Electron Agr* 14: 43-59. [http://dx.doi.org/10.1016/0168-1699\(95\)00036-4](http://dx.doi.org/10.1016/0168-1699(95)00036-4)
- Manfredi E, 1971. Raccomandazione A.I.G.R., IIIª sezione "denominazione, simbolo e unità di misura delle grandezze fondamentali relative all'impiego delle macchine in agricoltura, con particolare riguardo alle colture erbacee". *Rivista di Ingegneria Agraria* 2 (4): 258-260.
- Søgaard HT, Sørensen CG, 2004. A model for optimal selection of machinery sizes within the farm machinery system. *Biosyst Eng* 89: 13-28. <http://dx.doi.org/10.1016/j.biosystemseng.2004.05.004>
- Spinelli R, Magagnotti N, Picchi G, 2011. Annual use, economic life and residual value x of cut to-length harvesting machines. *J Forest Econ* 17: 378-387. <http://dx.doi.org/10.1016/j.jfe.2011.03.003>
- Taiti C, Costa C, Menesatti P, Comparini D, Bazihizina N, Azzarello E, Masi E, Mancuso S, 2015. Class-modeling approach to PTR-TOFMS data: a peppers case study. *J Sci Food Agr* 95: 1757-1763. <http://dx.doi.org/10.1002/jsfa.6761>
- Verani S, Sperandio G, Picchio R, Marchi E, Costa C, 2015. Sustainability assessment of a self-consumption wood-energy chain on small scale for heat generation in Central Italy. *Energies* 8: 5182-5197. <http://dx.doi.org/10.3390/en8065182>
- Williams PC (Ed.), 1987. Variables affecting near-infrared reflectance spectroscopic analysis. In: *Near-infrared technology in the agricultural and food industries*; Williams P, Norris K (Eds.), pp: 43-166. Am Assoc Cereal Chem, St Paul, USA.
- Wold S, Sjostrom M, Eriksson L, 2001. PLS-regression: A basic tool of chemometrics. *Chemom Intell Lab Syst* 58: 109-130. [http://dx.doi.org/10.1016/S0169-7439\(01\)00155-1](http://dx.doi.org/10.1016/S0169-7439(01)00155-1)