Simulation of agricultural logistic processes with k-nearest neighbors algorithm

Sebastian Götz[1], Nina Zimmermann[2], Dirk Engelhardt[2], Heinz Bernhardt[1*]

(1. *TechnischeUniversitätMünchen, Agricultural System Engineering, Am Staudengarten 2, 85354 Freising, Germany;*
2. Steinbeis University, Institute for Management in Logistics, Hanau, Germany.)

Abstract: The topic logistic has become more and more important in German agriculture during the last years. This is caused by a growth of enterprises and machines but also by the enormous extension of the cultivation of renewable resources for the production of energy. To manage these logistical tasks in agriculture in Germany at the moment, different transport systems are preferred. The classical system with tractor and agricultural trailer, transport via truck like it, is typical for the commercial transport of goods and the transport with specialized vehicles which can be classified between both systems. To evaluate these transport processes it is decisive for the farmers to know the key parameters of the single systems like the average fuel consumption (energy) and the average transport speed (time) for their special logistic issue to optimize the input of resources. The aim of the examination is therefore to develop a planning tool for the farmers to evaluate the logistic systems.Within this project examinations have shown that environmental influences like driver, loading, road type, gradient, winding ness, traffic situation, daytime and so on, have an enormous influence on the key parameters energy and time. Current simulation systems for the estimation of the key parameters energy and time have not provided satisfying results for the agricultural sector. Usually they are basing on linear and/or nonlinear equations with a very complex emphasis of the influencing factors and after technical changes on the logistic systems, they cannot be used any longer. Furthermore always only a very small number of influencing factors can be integrated. Therefore in this examination the k-Nearest Neighbors algorithm (kNN) is used. This results in a much more flexible use of different influencing factors with a difference in weight. With the help of test data, the system is learning and creating the kNN algorithm. This can be used for simulation. The advantage of this system of "artificial intelligence" is that the model building can be done in time in the area of the current working point. This makes it possible to integrate even unknown or in their effect not determinable environmental factors. By the training structure and the integration of new test data,the algorithm is much more easily adaptable on new trends. There is also the possibility to train the system optimal on the own conditions with the help of own test data. The k-Nearest Neighbors algorithm which has been determined during the examination, makes it possible, to estimate the key parameters energy and time for the logistic tasks in agriculture with a probability of more than 97%.

Keywords: Agricultural logistic, k-Nearest Neighbors algorithm, energy consumption, Germany.

Citation: Götz,S., N. Zimmermann, D.Engelhardt, and H. Bernhardt. 2015. Simulation of agricultural logistic processes with k-nearest neighbors algorithm. Agric Eng Int: CIGR Journal, Special issue 2015: 18th World Congress of CIGR: 241-245

1 Introduction

Agricultural production happens mostly on the field while processing is done on central places. Therefore transport is a main part of the agricultural production system. The effort for transport can be pointed on the following example (FAO Statistics Division, 2013). The average grain harvest in Germany is about 45.72 Mt. Under German climate conditions 160 to 200 h for threshing can be estimated during July and August(Urbanek, 2010). Therefore 180 h of threshing result in an average harvest of 254.00 t/h which have to be harvested, transported, processed and stored. The transport from the field to the warehouse and then to the processing is a central link between field and processing. Additional to the grain transport which is mentioned in the example, further transports for sugar beets and potatoes are necessary. For livestock breeding the

Received date: 2014-11-25 **Accepted date:** 2015-04-17
Corresponding author: Heinz Bernhardt, Technische Universität München, Agricultural System Engineering, Am Staudengarten 2, 85354 Freising, Germany. Email: heinz.bernhardt@wzw.tum.de.

transport of feed and farm fertilizer has to be regarded. Also the growing production of renewable resources for the energetic and substantial use, rises the number of transports in agriculture.

1.1 Transport simulation

To plan and simulate the transports appropriate models are necessary. The difficulty for correct modelling lay in the determination of the influencing variables. These comprise technical aspects like transport vehicle, engine power, tires and size; road conditions like incline, road surface, curve and road type; social aspects like interaction with other road users, road traffic law, driver etc..

The mathematical description of transport problems is often done by the factoring and description in regression equation. By multiple regressions Hermann (2000) and Weise (2000) describe for example the correlation between the speed of the tractor, their max. designed speed and the mass of the whole transport unit. Herewith a high coefficient of determination of 0.83 is reached (Herrmann, 2000).

The validity of these regression models is limited to the examined ranges of the described factors. For the given example the values of the total mass have to range between 4 600 and 29 610 kg and the maximum speed between 32 and 80 km/h. Due to the fast progression in automotive engineering,these models tend to a fast aging and require an often and time consuming actualization.

1.2 k-Nearest neighbors

A further method of modelling is the k-Nearest Neighbors Method (kNN). It is a non-parametric instance based classification method which is one of the simplest (Altman, 1992) and at the same time oldest (Cover, 1967)methodof mechanical learning.The method allows an estimation of an unknown dependent variable with the help of the similarity of a descriptive variable to those of a reference data set. This procedure allows a high flexibility and actuality of the model by the addition or change of the reference data set, so that an easy adoption into practice is possible (Kostadinov, 2012). The assumption of a certain distribution as it is necessary for regression equation is not applicable for the kNN-estimation.

The fundamental difference between the kNN-estimation and the other methods is that no explicit hypothesis or describing model function hasbeenestablished. The kNN-method is therefore not dependent on an explicit model like the regression methods. The kNN-estimation allows like already explained, the estimation of a target value from a test data set by means of the descriptive independent variable of the target value and their comparison with the descriptive variable of a reference data set, which is also called trainingdata. The first step of the two-stage kNN-estimation lies in the calculation of the nearest neighbor by means of the similarity of the descriptive variable of the reference data set to the descriptive variables of the test data. Therefore the relative nearness of the reference data to the target is calculated by the help of adequate distance measure and afterwards the reference data are arranged in relation of their distance to the target point. That means that from the available training data almost similar data sets are chosen and under the assumption that these data sets are also similar in relation to their target feature used to calculate the target value. The actual calculation of the wanted target value occurs then by the selection of the most frequently value in terms of a majority decision or a weighted average determination of the target value in the training data set. In contrast to a regression model that creates a functional equation over the whole range of a data set the result of a kNN-estimation is only a local approximation which only results in the similarity to the nearest neighbors and therefore changes for another calculation. A standardization of the independent variable is necessary before the distance calculation (Haendel, n.y.; Backhaus, 2008) as the variables have different seizes and units like 3% grade, 60 km/h speed or engine speed of 2100 rpm. For standardization of the data the z-transformation is used as though the relative value is transformed by average value and standard deviation.

2 Material and methods

2.1 Testdata

In an experimental series (Engelhardt, 2002; Bernhardt et. al., 2008;Götzet.al., 2011;Bernhardt et.al., 2012) the test data for the kNN algorithm are created. The fuel consumption and the average speed of different modern agricultural transport systems have been detected. The presented analysis was conducted in Hessen Germany in 2011 and 2012. The test track – as in the whole experimental series – includes cross-town routes and country roads in different states of development, which add up to a length of about 17 kilometres. In this particular analysis, highway sections,have not been covered, because the tractor has no permission for this type of road

Two common tractors with different power spectrums, a Unimog and a Truck have been chosen for the road test. An overview of the towing vehicles technical characteristics is shown in Table 1.

Table 1Keydata of the deployed towing vehicles

	Unit	Tractor 121 kW	Tractor 243 kW	Unimog 210 kW	Truck 310 kW
Nominal output	kW	121	243	210	310
Dead weight	kg	6985	10830	7480	7400
Lenght	m	4.75	5.65	6.11	5.93
Height	m	2.99	3.32	3.49	2.93
Width	m	2.57	2.75	2.50	2.40
Gearing		infinitely variable transmission	infinitely variable transmission	manual transmission	automatic transmission
Ad Blue reservoir	L	0	0	0	25
Maximum design speed	km/h	40	50	80	80

The trailer variants have been chosen carefully in order to match the towing vehicles. Two 18 t 2-axle trailers, a 3-axle trailer and a semi-trailer (with or without dolly-axle) represent the available options (Table 2).

Table2Keyfigures of the deployed trailer variants

		18t 2-axle trailer	3-axle trailer	Semi - trailer	Dolly- axle
Dead weight	kg	4420	5900	8130	2180
Gross vehicle weight	kg	18000	24000	34000	13000
Pay load	kg	13580	18100	25870	
Length	m	7.35	8.20	10.10	
Height	m		1.70	2.00	
Width	m	2.55	2.45	2.50	

For every single vehicle combination the fuel consumption and the position and time data have been recorded. The factors "type of vehicle", "engine power", "engine type", "driver","tyre equipment", "type of street", "acclivity", "landform" and "traffic conditions" have been varied. For the planning of the single tests it was important that the chosen combination is also useful in practice.

The average fuel consumption of the examined trucks for example is about 35.97 L/100km for country lanes and 55.47 L/100km for borough roads. The average speed of the examined tractors is 39 km/h for country lanes and

26.4 km/h for borough roads. The difference between various drivers with the same training background is about 6 %. Also the speed in main-through-roads varies for tractors about 15 % in the course of a day caused by the other road users.

inclination and daily deviations in traffic flow are set as fix data.

As Figure 1 shows there is a good matching of the values estimated by kNN algorithm and the separately measured values on the test track. The variance of the estimated values is in total a little bit lesser than for the measured values.The larger differences in the truck can be explained by the smaller number of measurements and thus the greater integration of the extreme measured values.

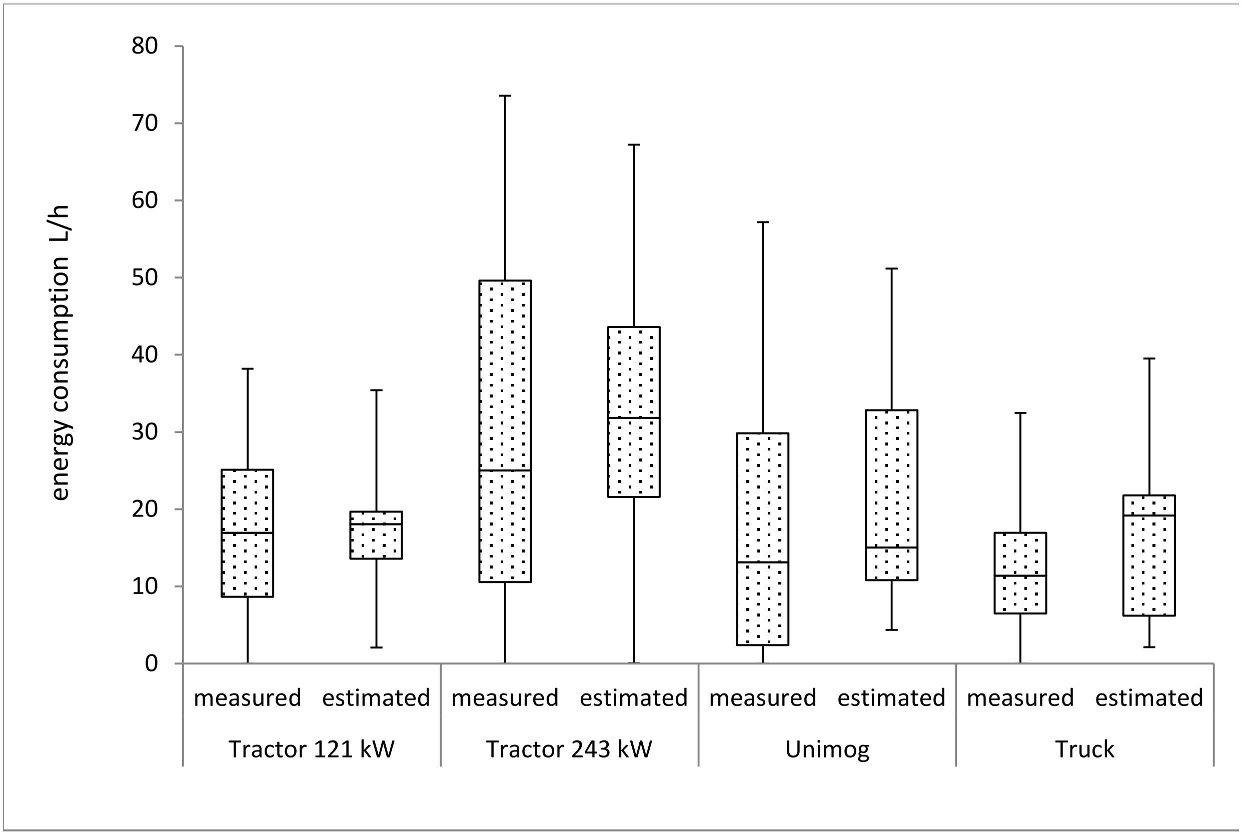

Figure 1 kNNalgorithmenergy consumption.

2.2 kNN results

Based on the determined data the data of a comparable test track are estimated with the help of the kNN algorithm. The kind of vehicle, road type, curves,

3 Discussion and conclusion

It can be shown that the kNN algorithm is a good method to estimate complex logistic problems. These problems are most times indicated by a large amount of effects which have influences on the system. kNN algorithm makes it possible to integrate the influence of possible effects into the calculation for which the concrete impact on the data is not definitely known.

The estimated values for the energy consumption show only a slight deviation to the separately detected

data. Also the variance of the values is with kNN algorithm clearly smaller. The more the used test data for kNN algorithm is adapted to the environment which should be examined,the better are the estimated results. This can be explained by the fact, that not all parameters can be detected and nominated. There are always influences which are hidden in the used data but cannot be clearly named. This is the advantage of the kNN method as it is based on real data, so that it is not necessary; to fix the correlation of the single parameters as it is necessary for a correlation analysis.

With the kNN method there is the possibility to integrate data into an existing estimable model and therefore adapt the estimation on new trends e.g. another engine which leads to changes in the driving characteristics or new transport vehicles with other chassis.

Another advantage of the kNN method is that for the first time in agricultural logistic, so called soft-factors, which result from the interaction with other road users, can be integrated into the estimation of energy consumption and average speed. The number of road user has an enormous influence on the results. Even for slow vehicles like tractor and trailers for which it is assumed that they are only a traffic block,it can be shown, that there is a change in energy consumption and speed caused by the general volume of traffic.

The aim of future examinations is to integrate additional data and to analyze the influence of different fix combinations of variables on the accuracy of the estimation.

[References]

Altman, N. 1992.An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression.*The American Statistician*, 46(3):175–185.

Backhaus, K., B.Erichson,W.Plinke and R. Weiber.2008.*Multivariate Analysemethoden Eine anwendungsorientierte Einführung*.Berlin:Springer.

Bernhardt, H., W.Lixfeld,D.EngelhardtandE.Kolundzija.2008.New Transport and Handling Technology for Optimizing the Logistics Supply Chain during Grain Harvest. *LandtechnikAgricultural Engineering*, 63(2):92-93.

Bernhardt, H.,S.Götz, V.Heizinger,N.Zimmermann andD.Engelhardt.2012. Energy consumption of agricultural transports on influencing factors.*International Conference of Agricultural Engineering CIGR-AgEng 2012*.Available at: http://cigr.ageng2012.org/images/fotosg/tabla_137_C1768.pdf, Valencia / Spain.8.-12. July 2012.

Cover, T.M. andP.E. Hart.1967.Nearest Neighbor Pattern Classification. *IEEE Transactions on Information Theory*, 13(1), 21–27.

Engelhardt, D. 2002. Transportfahrzeuge im Agrarbereich - Lastkraftwagen, Möglichkeiten und Konzepte, Ph. D. Diss. Justus-Liebig-Universität Giessen. Cuvillier Verlag.

FAO Statistics Division. 2013. FAO-ErnteMengen-2005-2010, FAO. http://faostat3.fao.org/home/index.html.

Götz, S., J.Holzer, J.Winkler, D.Engelhardt andH.Bernhardt.2011. Agricultural Logistics – System comparison of transport concepts in grain logistics.*Landtechnik Agricultural Engineering*, 66(5): 381-386.

Haendel, L. 2003.*Clusterverfahren zur datenbasierten Generierung interpretierbarer Regeln unter Verwendung lokaler Entscheidungskriterien*, Dortmund: Uni Dortmund.

Herrmann, A., P.Pickel, andN.Fröba.2000.Transport speeds of tractors.*Landtechnik Agricultural Engineering*,55(5):338–339.

Heyes, D., E.Hipp,A.Zimmermann, X.Römersperger, D.Raudszus, andM.Lienkamp.2012.Systematische Bewertung des Fahrereinflusses auf die Transporteffizienz von Nutzfahrzeugen. B.Sc. Thesis TU München.

Kostadinov, F.,R.Lemm, andO.Thees.2012.Ein Instrument zur Schätzung von Holzernteproduktivitäten mittels kNN-Methode.*Schweizerische Zeitschrift für Forstwesen*, 163, (4):119–129.

Urbanek, M. 2010. Entwicklungstendenzen am Mähdreschermarkt. M. Sc. Thesis Hochschule Neubrandenburg.

Weise, G. 2000.Practical driving performances of agricultural transport vehicles.*Landtechnik Agricultural Engineering*,55(5): 338 – 339.