Statistical tool to estimate and optimize the intensity of the dependence between the parameters of a dynamic system

CARDEI PETRU
Computerized engineering
INMA
6, Ion Ionescu de la Brad Blvd., Sector 1, Bucharest
ROMANIA

Abstract: - The article presents a method of investigating the intensity of the connection between the parameters measured in experiments for different dynamic processes. The method is based on a function constructed from the correlation between the rows of records, called variables. The result is addressed to all specialists who try to obtain relationships between the parameters of a process, starting from the results of experimental research. It is considered that the relationships sought between the variables of the researched process do not have a theoretical foundation or do not have a corresponding one. The tool for assessing the intensity of the connection between the main parameter and the secondary parameters of a dynamic process is tested on soil tillage processes in agriculture. A hierarchy of the intensity of the connection between the main parameter and the secondary parameters or combinations thereof is provided. Using these hierarchies, researchers can create regressions based on the priorities offered by these hierarchies.

Key-Words: - dynamic, process, parametric, statistics, estimators, intensity, connection

Received: August 9, 2021. Revised: March 27, 2022. Accepted: April 24, 2022. Published: June 3, 2022.

1 Introduction

The fundamental problem from which the research whose results are given in this article starts is an old one, over a century old. It consists in trying to predict and minimize the soil tillage draft force generated by agricultural machinery intended for soil work. The literature is very rich in this field. We have given the main references in this field of research in [1], [2], and [3]. Prediction is one of the most important goals of science, which is generally aimed in any field of science. The purely theoretical forecast, although not impossible, is rare and the accuracy is debatable. In general, the experiment is the one that provides answers for the construction of models capable of predictions and optimizations, and the whole experiment is the one that can validate the model thus constructed. The experimental program is decisive in the generation of the mathematical model and its exploitation, [11-12]. Over time, the original problem has taken on enormous dimensions, covering virtually the entire scientific field.

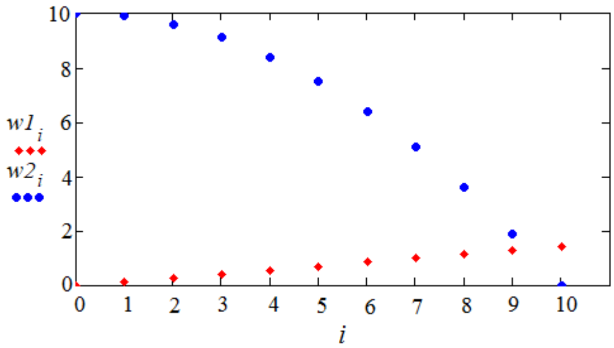
The experimental method is integrated into a more general field of research, in which the results presented in this article have their place: System identification. The field of systems identification uses statistical methods to construct mathematical models of dynamical systems, [13]. The field of systems identification also includes the optimal design of experiments, which aims to generate efficient information, able to provide mathematical models appropriate to achieve the proposed goals. One of the most common approaches to identifying systems is to start experimenting and then try to determine the mathematical relationships between the parameters of the system, without going into details about what is going on inside the system. This approach to the problem is called identifying the black box system, [14].

A dynamic mathematical model in this context is a mathematical description of the dynamic behaviour of a system or process, either in terms of time or frequency. Examples include in this field are physical processes, such as the movement of a falling body under the influence of gravity, the working process of an agricultural machine, the processes of interaction of plants with the soil, etc. One of the many possible applications of system

identification is the problem of control systems. For example, it is the basis for modern data-driven control systems, in which the concepts of system identification are integrated into the design of the controller and lay the groundwork for formal evidence of controller optimization, [14].
From all that is described regarding the identification of systems in the paragraph above, this article provides a method and a statistical tool for assessing the intensity of the connection between the parameters of a dynamic system or a dynamic process. The statistical tool consists of a function based on the correlation coefficient between two variables (series of experimental records of numerical values of some process parameters) which have one or more parameters, in relation to which the optimal mathematical model of the relationships will be calculated. These results are compared in order to rank the intensities of the dependencies between the process parameters. The method based on this function elaborates the strategy of using the constructed function to evaluate the intensity of the connection between the variables (parameters) of the experimental research process and the formation of a hierarchy capable of suggesting to researchers how to form regressions capable of predictions and / or optimizations.
The proposed method and tool are simple and easy to understand alternatives to the sophisticated statistical methods used in the field of systems identification: descriptive statistics, regressions, machine learning, hypothesis testing, etc. [15-16]. The method can be extended using other estimators of descriptive statistics, to achieve optimal statistical models of dynamic processes. *This article is an advanced version of the preprint (not reviewed and unpublished) archived in Research Gate networking*, [17].

2 Problem Formulation

It is considered several experimental data strings, called variables, each with N components (characterizing, for example, a dynamic process). In general, is supposed that there are no theoretically or empirically established relationships between the variables considered. In the current research, in such situations, mathematical statistics are used to obtain a measure of the relationship between various variables (correlation coefficient) and characterization of dependence through various types of regressions.
Through experimentation, data series (variables) are obtained for which the intensity of the connection and the form of the dependence are investigated. The intensity of the link between two experimental variables can be conveniently estimated using the correlation coefficient between the two variables, according to, for example, [4]. Because the main function with which we will operate in this research works using the operator of the correlation coefficient provided by [5], in writing the function we will keep the name of that operator. Therefore, is considering, for example, two experimental variables: $w1 = \{w1_i, i = 1,\ldots,N\}$, $w2 = \{w2_i, i = 1,\ldots,N\}$, $N = 10$, data numerically and graphically in fig. 1.



$$w1 = \{0.0,0.143,0.286,0.428,0.571,0.714,0.6857,0.999,1.142,1.285,1.428\}$$
$$w2 = \{10.0,9.9,9.6,9.1,8.4,7.5,6.4,5.1,3.6,1.9,0.0\}$$

Fig. 1. Two one-dimensional variables with ten components each, numerical values and graphical representation.

The basic function is defined through formula (1):

$$\Lambda: V \times V \times \mathbb{R} \to \mathbb{R},$$
$$\Lambda(v1, v2, x) = corr(v1, v2^x) \qquad (1)$$

where V is a set of vector variables, and $\mathbb{R}$ is the set of real numbers, the domain of definition of the function Λ being the Cartesian product of V with itself and with $\mathbb{R}$. By $v2^x$ we mean the variable obtained from the variable $v2$ by raising the power x of each component (which models the hypothetical form of the dependency between the two variables) , obviously if the operation is allowed. For the correlation coefficient function we used the abbreviated notation:

$$corr: V \times V \to \mathbb{R},$$
$$corr(v1, v2) = \frac{cvar(v1, v2)}{stdev(v1) \cdot stdev(v2)} \qquad (2)$$

where the covariance function is defined by the next formula:

$$cvar: V \times V \to \mathbb{R},$$
$$cvar(v1, v2) = \frac{1}{m \cdot n} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \left[(v1_i - \overline{v1}) \cdot \cdot (v2_j - \overline{v2}) \right] \quad (3)$$

$\overline{v1}$ and $\overline{v2}$ being the average values of the vectors $v1$, and $v2$, and the function:

$$stdev: V \to \mathbb{R},$$
$$stdev(v) = \sqrt{\frac{1}{N} \cdot \sum_{i=0}^{N-1} (v_i - \bar{v})^2} \quad (4)$$

is the standard mean deviation of the variable v. For two fixed variables (for example those in fig. 1), the function Λ becomes a function of a real variable. The variation of this function is difficult to study theoretically, and for our immediate purposes, it is easier to study this function numerically. Please note that when researchers look for exponential dependency laws, in general, the exponents are not very large, so we will represent the function $\Lambda(x)$ over a reasonable length of time for this application. By selection (scanning several intervals for x), we found the interval between 1.5 and 2.5, the interval in which the function Λ has a minimum equal to -1 for the value $x = 2$ of the argument.

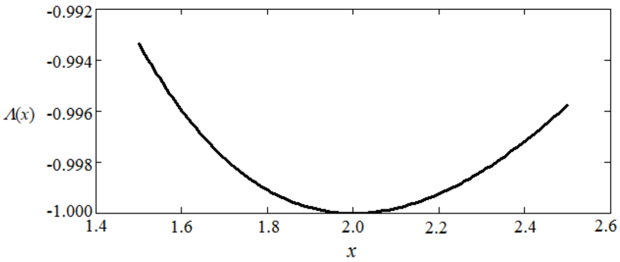


Fig. 2 Dependence of the function Λ, on x, for x between 1.5 and 2.5.

As an immediate interpretation, the function Λ indicates an inverse (in the sense that if $w2$ increases, then $w1$ decreases) dependence between the variables $w1$ and $w2$, of the form:

$$w2 = f(w1^2) \quad (5)$$

If we specify that we chose the two variables as the temporal and the spatial variable of a vertical fall of a point body, with zero initial velocity:

$$h(t) = h_0 - \frac{g}{2} t^2 \quad (6)$$

where h is the current height of the body, h_0 is the height at the initial moment, $t = 0$, $g = 9.81 \text{m} / \text{s}^2$ is the local gravitational acceleration, and t is the time, then the indication of the function Λ, is not a surprise but seems more like a promise to obtain a tool to complete statistical investigations performed on experimental data. At a theoretical level, there remain, for the function Λ and, in general for the research method described, some problems that need to be solved:

D1) the theoretical study of the variation of the function Λ must be tried (even if it will not be possible to solve until the end by analytical methods), to find out if there is always an extreme point, or if there are cases in which it has several points of extreme.

D2) if there are cases in which the function Λ has several extreme points, hypotheses must be advanced on the choice of a physically acceptable variant for the relationship whose model is being tried.

D3) If the function Λ does not show extreme values at physically acceptable intervals, can it be concluded that the variables are independent?

D4) The construction of generalized functions of type (1), such as:

$$\Lambda: V \times V \times \mathbb{R} \to \mathbb{R},$$
$$\Lambda\Big(v1, u\big(v1, u(v2, x)\big)\Big) = corr(v1, v2^x) \quad (7)$$

where $u(v2, x)$ is a function defined on $V \times \mathbb{R}$, and can be inspired by additional information on the phenomenon studied or by the experience of the statistical operator. These problems remain for specialized theoretical studies. In the next chapter, we will focus on some experimental results from the activity of researchers in the field of agricultural tillage machines.

3 Results and comments

In this chapter, the proposed research method will be applied in order to estimate the intensity of the soil tillage draft force connection for a machine designed for soil tillage, with a series of parameters of the process interaction with the soil, in experiments, as well as with their combinations. For this purpose, the experimental data published in [6], obtained for a plough with two mouldboards, will be used. The data is voluminous, it is in the public domain and that is why we do not reproduce it. The authors [6] vary the humidity, depth and working speed and measure the soil tillage draft force. The behaviour of function (1) will be investigated, for variable $v1$ is the soil tillage draft force, and variable $v2$ is specified in table 1, first column. For convenience in writing, denote by h, the soil moisture, by a working depth, respectively by v, the working speed, and by $\boldsymbol{R}$, the soil tillage draft force.

Table 1 Evaluations of the connection between the soil tillage draft force and other parameters of the soil processing with the plough with two mouldboards from [6].

Variable $v2$	x extreme	Extreme of correlation coefficient
h	-0.00003	0.124
a	3.547	0.871
v	1.228	0.423
a^2	1.774	0.871
av	1.547	0.928
av^2	1.237	0.829
$(av)^2$	0.774	0.928

The results in Table 1 show that the parameters with the greatest influence on the tensile strength, $\boldsymbol{R}$, are the combinations av and $(av)^2$, for which the correlation coefficient takes values greater than 0.9. The working depth is dominant in these products. This working depth, a, at the first or second power, gives a maximum correlation coefficient of 0.871, while the working speed, v, at the first power reaches a maximum correlation coefficient with the value 0.423, all with the soil tillage draft force, $\boldsymbol{R}$. These results motivate the consideration of the multiplicative combinations between depth and working speed, in all the calculation formulas proposed for the soil tillage draft force, listed in [1], [2] and [3]. All calculation formulas for soil tillage draft force, examined in [1], [2] and [3] have integer exponents for working depth, working speed or combinations thereof. If taking into account the results of the investigation method presented in this paper, the exponents would become numbers that are no longer integers, then, in order to save the dimensional correctness of the formulas, it is proposed that these combinations be replaced by dimensionless reports, which can be raised to any real powers without creating dimensional problems for formulas. A correctly dimensional solution which uses exponents that are not integers is suggested in [7]. Results somewhat similar to those in Table 1, are obtained by processing other experiments. For example, for the experimental results in [8], the estimates in Table 2 are obtained.

Table 2 Evaluations of the connection between the soil tillage draft force and other parameters of the soil processing with the plough with two mouldboards from [8].

Variable $v2$	x extreme	Extreme of correlation coefficient
a	0.00007	0.933
v	0.415	0.339
a^2	0.000035	0.933
av	0.941	0.824
av^2	1.339	0.666
$(av)^2$	0.471	0.824

The experimental results given in [9], for a system of physical simulation of the working process of some chisel working body, are described in table 3.

Table 3 Estimates of the relationship between soil tillage draft and other parameters of soil processing with the plough with two mouldboards from [9].

Variable $v2$	x extreme	Extreme of correlation coefficient
a	1.039	0.689
v	1.659	0.038
a^2	0.519	0.689
av	0.512	0.424
av^2	9.581	0.332
$(av)^2$	0.256	0.424
Rake angle	1.048	0.631
Cut angle	-0.00001487	0.111
$a^2 \cdot$sin(rake angle)	1.25	0.971
$av \cdot$sin(rake angle)	0.982	0.727

The experiments whose data are published in [10] bring in the set of measured data some parameters rarely considered: soil density, resistance to soil cone index, soil moisture, cohesion and shear stress of the soil. An analysis similar to the above of these records leads to the results in Table 4.

Table 4 Estimates of the relationship between soil tillage draft force and other parameters of soil processing with the plough with two mouldboards from [10].

Variable *v2*	*x* extreme	Extreme of correlation coefficient
a (working depth)	1.088	0.659
Soil density	18.307	0.812
Resistance to vertical penetration of the soil	1.808	0.842
Soil moisture	91.387	-0.387
Soil cohesion	0.0000001	0.871
Ground shear stress	4.495	0.925

The results obtained using the research method described are supported by the function of investigating the intensity of the link between the process parameters, (1), allowing the extraction of useful observations for the use of this technique or research methods.

First of all, it is noted that the working depth is the process parameter most intensely correlated with the soil tillage draft force (or which has the greatest influence on the soil tillage draft force). It is also observed that at the working speeds characteristic of soil tillage works, the influence of the working speed parameter on the soil tillage draft force is small in relation to the influence of other parameters working depth, the tilt angle of the working body, shear stress and soil cone index. It can be suspected that at high speeds, exceeding a certain threshold whose value depends on the mechanical and structural characteristics of the soil, the influence of the working speed on the soil tillage draft force becomes higher.

It is also possible to observe the usefulness of finding the intensity of the connections between the process parameters, using intermediate functions: in table 3 it is observed that the soil tillage draft force depends on the angle of inclination, but more intensely on its sine.

4 Conclusion

This article proposes a tool for researching the intensity of the connection between the parameters of a process, studied experimentally. The final goal is to obtain sufficiently intense parametric combinations related to the main parameter of the process so that an acceptable regression formula can be proposed. The regression formula will be physically adjusted so that the requirement that the arguments of the transcendent functions are dimensionless is met.

The proposed research method and the function which is the main tool have been tested and provide the intensity of the link between the main parameter of the tillage process in agriculture (soil tillage draft force) and the other process parameters, some input, others command and control (depth, width and working speed, in particular).

The proposed investigation method produces a hierarchy of the influence of parameters or parametric combinations on the main parameter of soil processing, the soil tillage draft force. This ranking is then used to create possible regressions for predicting the soil tillage draft force.

Although it is tested on the process of tillage in agriculture, the proposed tool can be used similarly for other processes in agriculture or other fields in which experimental research is carried out to find relationships between characteristic parameters, relationships that cannot be substantiated theoretically.

The tool can be further refined until optimal regressions are obtained, at least in the first stage, obviously only if there will be requests.

References:

[1] Cardei P., Muraru S., Muraru C., Sfiru R., Relative ordering tests for draft force models in soil tillage, *ERDEV Proceedings conference, Jelgava*, 2021.

[2] Cardei P., Muraru S., Muraru C., Sfiru R., Validation and ordering test for an formula of soil tillage draft force, extended to the dependence on soil moisture, *preprint*, DOI: 10.13140/RG.2.2.26955.54569, 2019.

[3] Cardei P., Condruz P., Sfiru R., Muraru C., Test for physical laws of draft force generated in tillage operations, *ERDEV Conference Proceedings*, 2020.

[4] Peck R., Olsen C., Devore J., Introduction in Statistics & Data Analysis, 3rd Edition, *Thomson Brooks/Cole.*, 2008.

[5] MathSoft, Inc., Mathcad 2000, Rference Manual, *MathSoft, Inc., Cambridge, MA 02142 USA*, 1999.

[6] Ranjbar I., Rashidi M., Najjarzadeh I., Niazkhani A., Niyazadeh M., Modeling of Mouldboard Plow Draft Force Based on Tillage Depth and Operation Speed, *Middle-East Journal of Scientific Research* , 17(7): 891-897, 2013.

[7] Moeenifar A., Mousavi-Seyedi S.,R., Kalantari D., Influence of tillage depth, penetration angle and forward speed on the soil/thin-blade interaction force, *Agric Eng Int: CIGR Journal*, vol. 16, no.1, p. 69-74 , 2014.

[8] Al-Suhaibani S.A., Ghaly A.E., Effect of Plowing Depth of Tillage and Forward Speed on the Performance of Medium Size Chisel Plow Operating in a Sandy Soil, *American Journal of Agricultural and Biological Sciences*, 5 (3), 2010, 247-255.

[9] Fechete-Tutunaru L.V.(2019) Gaspar F., Gyorgy Z., Soil-tool interaction of a simple tillage tool in sand, *E3S Web of Conference* s 85, 0.8007, 2019.

[10] Abich S.O., Gitau A.N., Nyaanga D.M., Prediction of subsoiling draft requirements of silt loam soils from empirical data: neural network versus multiple linear regression, *Journal of Engineering in Agriculture and Environment*, vol. 7, No. 2, 2021, 61:72.

[11] Tamura R., Takei Y., Imai Sh., Nakahara M., Shibata S., Nakashiki T., Experimental design for the highly accurate prediction of material properties using descriptors obtained by measurement, *Science and Technology of Advanced Materials: Methods*, Taylor & Francis Online, 2021.

[12] Dolatsara H.A., Chen Y-J., Leonard R.D., Megahed F. M., Jones-Farmer L.A., Explaining Predictive Model Performance: An Experimental Study of Data Preparation and Model Choice, *Big Data*, 2021.

[13] Soderstrom T., Stoica P., *System identification*, New York: Prentice Hall, 1989.

[14] https://en.wikipedia.org/wiki/System_identification#cite_note-1

[15] Lund Research Ltd., *Descriptive an Inferential Statistics*, 2014.

[16] https://en.wikipedia.org/wiki/Statistics

[17] Cardei P., A tool for assensing the intensity of the connection between the parameters of a dynamic system, preprint DOI: 10.13140/RG.2.2.33666.76487, 2022.