ORIGINAL

Comparison of three methods for outliers elimination in regression analysis between the soil saturation extract electrical conductivity and the 1:5 extract electrical conductivity.

Dr.C. Mario González Posada Dacosta, Profesor Titular. [mgonzalezposada@udg.co.cu]
Universidad de Granma. Cuba.

Roberto Alejandro García Reyes. [rgarciar@uho.edu.cu]
Universidad de Holguín. Cuba.

Carlos González Posada Rodríguez. [carlos.damian@crearq.cu]
Empresa de Servicios de Ingeniería y Diseño. Cuba.

Abstract

To express the levels of soil salinity with respect to crop development, the value of the electrical conductivity of the saturation extract (ECe) is generally used although there are a number of drawbacks in its determination, which limit its widespread use. To overcome such problem is possible to make simultaneous measurements of ECe and the 1:5 extract electrical conductivity (EC1:5), easier to measure, for a small number of samples and then with the measured data, to obtain the best regression equation between both variables. The usefulness of this methodology is to estimate ECe data without the need to measure them directly. To ensure the application of the best model, outliers analysis is used to discard possible values that lead to non-optimal parameters of the regression equation to be used in the calibration process. Therefore, this paper analyzes the influence of the outliers elimination method on the quality of the adjustment models, in the ECe vs. EC1:5 linear regression analysis. The data used in this work were obtained from the processing of soil samples taken at sites located in the Cauto River Basin, affected by salinity. It was concluded that the most convenient and simple method for outliers elimination is the one that determines those cases for which its standard residue exceeds double the standard deviation of the set of standard residues of the sample considered in each analysis.

Keywords: outliers; soil salinity; linear regression.

Recibido: 11/02/2019 | **Aceptado**: 11/09/2019

Comparación de tres métodos para la eliminación de outliers en el análisis de regresión entre la conductividad eléctrica del extracto de saturación del suelo y la conductividad eléctrica del extracto 1:5.

Resumen

Para expresar los niveles de salinidad del suelo con respecto al desarrollo de los cultivos, generalmente se utiliza el valor de la conductividad eléctrica del extracto de saturación (ECe), aunque existe un número de problemas en su determinación que limitan su uso generalizado. Para resolver estos problemas es posible realizar mediciones simultáneas de ECe y de la conductividad eléctrica del extracto 1:5 (EC1:5), más fácil de medir, para un número pequeño de muestras y entonces con los datos de las mediciones, obtener la mejor ecuación de regresión entre ambas variables. La utilidad de esta metodología es estimar datos de ECe sin la necesidad de medirlos directamente. Para garantizar la aplicación del mejor modelo, se emplea el análisis de outliers para descartar posibles valores que conduzcan a parámetros no óptimos de la ecuación de regresión a utilizar en el proceso de calibración. Por tanto, este trabajo analiza la influencia del método de eliminación de outliers sobre la calidad de los modelos de ajuste en el análisis de regresión lineal entre ECe y EC1:5. Los datos utilizados en este trabajo fueron obtenidos del procesamiento de muestras de suelo tomadas en puntos ubicados en la Cuenca del Río Cauto, afectados por salinidad. Se concluyó que el método más conveniente y simple para la eliminación es aquel que determina aquellos casos para los cuales su residuo estándar es superior al doble de la desviación estándar del conjunto de residuos estándares de la muestra considerada en cada análisis.

Palabras clave: ouliers; salinidad de suelo; regresión lineal.

Introduction

The salt content of a soil can be estimated in good approximation from a measurement of the electrical conductivity of a saturated soil paste or a more dilute suspension.

To express the levels of soil salinity with respect to crop development, the value of the electrical conductivity of the saturation extract (ECe) is generally used. This procedure consists in preparing a saturated soil paste and then by suction or centrifugation enough amount of extract is obtained to determine its conductivity.

The main advantage of the saturation extract method for evaluating soil salinity is that the saturation percentage is directly related to the different values of the field moisture. Determinations made in many soils indicate that, within a wide range of textures, the saturation percentage is approximately equal to four times the moisture content at 15 atmospheres, which approximates the permanent wilt moisture. The concentration of soluble salts in the saturation extract therefore, tends to be almost half of the concentration of the soil solution in the upper moisture limit and almost a quarter of the concentration that the soil solution would have in the limit lower of moisture. The effect of salts dilution that occurs in soils with fine textures, due to their greater moisture retention capacity, is automatically considered. For this reason, the saturation extract conductivity can be used directly to estimate the effect of soil salinity on plant development (USSL, 1954).

Regardless of the great utility of using the saturation extract electrical conductivity as an index for soil salinity evaluation, there are a number of drawbacks in its determination, which limit its widespread use, especially in routine work linked to studies of diagnosis and monitoring in production areas. The most important are detailed below.

a) Used soil mass: In order to obtain enough amount of extract for measuring its electrical conductivity, depending on the soil texture, it is generally necessary to take 1.5 to 2 kg of soil in each layer corresponding to each sampling point. In soil studies where geostatistical techniques are applied for detailed scale mapping, each sampling can comprise 30 to 100 points, in each of which, samples must be taken at least in four depths. All of the above results are needed to take, to transport and to process a large amount of soil in each study that is carried out, which in many cases, for practical reasons, are impossible.
b) Subjectivity in the saturated paste preparation. The end point of the saturated paste is based on fundamentally subjective criteria (USSL, 1954). In this sense, the analyst's ability and experience is decisive in the quality and repeatability of obtaining the saturation moisture corresponding to each sample. This drawback of course adds another variability component to ECe determinations. The fact that this humidity depends fundamentally on the soil texture, greatly limits the automation of this process and therefore the processing of large numbers of samples in each work session.
c) Obtaining the saturation extract. This step turns out to be the most difficult to execute due to the equipment requirements as well as the cumbersome handling of saturated paste (Slavich and Petterson, 1990; Naddler and Dasberg, 1980).

The afore mentioned drawbacks in relation to the soil saturation extract electrical conductivity measurement, make its use very limited and instead are used other faster and cheaper methods. In our country, the measurement of the electrical conductivity of the 1:5 extract is used and from this value the total soluble salt content (SST,%) is calculated through the expression:

$$SST = 0.297 * EC1:5^{0.87}$$

In fact, the essence of this evaluation is the estimation of the salinity index (SST) from the measurement of another magnitude (EC1:5), a much simpler procedure to perform.

In the case of the saturation extract electrical conductivity it is also possible to use this method, from the measurement of another magnitude that is correlated with the ECe and that is easier to determine. One of the quantities that satisfies such requirements turns out to be the electrical conductivity of the 1:5 extract.

The idea on which the method is based is to make simultaneous measurements of both magnitudes for a small number of samples and then with the measured data, to obtain the best regression equation between both variables. In this step, the rigor in which the data is handled and the analyzes are performed is of the greatest importance, since the accuracy and precision of the ECe estimates will depend on the reliability of the obtained equations. In general cases, the linear regression analysis is used through the least squares adjustment technique. The usefulness of this methodology is that once the procedure described above has been carried out, it is possible to estimate ECe data without the need to measure them directly, which on the one hand, allows the diagnosis of salinity in internationally accepted terms and on the other, facilitates and make this activity cheaper.

To ensure the application of the best model, outliers analysis is used to discard possible values that lead to non-optimal parameters of the regression equation to be used in the calibration process.

The problem of deciding to what extent any particular observation can be considered as belonging to the data sample obtained, as a result of the measurement process, is a difficult issue in which there is still no general consensus for its treatment. In this sense, there are important factors to take into account when deciding which data should be eliminated or not from the sample, among which are the measured magnitude and its variability, the characteristics of the measurement method and the instruments used, the experience of the researcher and the objectives of the analysis to be performed with the obtained data.

The majority of the authors agree to define as outliers those extreme, atypical and infrequent cases of a variable or combination of several variables, which exert a great influence on the statistical calculations of the sample (StatSoft, Inc., 2011). Its treatment is based on the criteria to decide whether or not to eliminate it from the data set and to adjust the methods of statistical analysis to their processing with minimal influence on the results to be obtained.

To reduce the impact in statistical analyzes, caused by outliers, it is first advised to review the collected information, in order to detect errors in its registry. Second, to verify if one of the variables involved is responsible for most outliers considering their elimination. Third, to eliminate those that under some grounded criteria do not fit the population. Finally, if it is not advisable to eliminate those variables responsible for the outliers, then proceed to their transformation.

The existence of outliers and their treatment has great influence when regression analyzes are performed between two or more variables. Taking into account the algorithm of the least squares method, a simple outlier may be able to significantly change the slope of the regression line and therefore the value of the correlation coefficient. In this case, the fundamental tool for the detection of outliers is the analysis of residues, which can be done graphically or analytically. The first route, of course, is applicable to the bivariate case, through the preparation of scatter plots that include the estimated regression line and by inspection of measurements, those cases that are very far from this line, are detected. This procedure has a great subjective content and can only be considered as a preliminary step. From the analytical point of view, several authors reported procedures for the identification of outliers based primarily on the comparison of the residues associated with each value with estimators linked to the degree of dispersion of the data for a given level of significance. Those cases in which the absolute value of its standard residue (SRV) is in a range of two to five times the standard error of the distribution of these errors (σ_{SR}) are generally considered as outliers. The magnitude of the standard residue is calculated through the expression:

$$SRV = \frac{Y_{Obs} - Y_{Est}}{\sqrt{\frac{\sum (Y_{Obs} - Y_{Est})^2}{n-2}}}$$

Other criteria for outliers evaluation are the calculations of Cook's distance and Mahalanobis' distance.

Imagine the independent variables (in a regression equation) as defining a multidimensional space in which each observation can be plotted. In addition, it can be plotted a point representing the means for all independent variables. This "mean point" in the multidimensional space is also called the centroid. The Mahalanobis` distance is the distance of a case from the centroid in the multidimensional space, defined by the correlated independent variables (if the independent variables are uncorrelated, it is the same as the simple Euclidean distance). Thus, this measure provides an indication of whether or not an observation is an outlier with respect to the independent variable values (StatSoft, Inc., 2011).

The Cook´s distance indicates the difference between the computed βi (slope) values and the values one would have obtained, had the respective case been excluded. All distances should be of about equal magnitude; if not, then there is reason to believe that the respective case(s) biased the estimation of the regression coefficients.

During the simple linear regression analysis, for outliers detection, we can proceed to determine those cases in which the absolute value of the standard residue exceeds double the value of the standard error of the distribution of the standard errors. These cases are removed from the database and the analysis is repeated. It may happen that after the previous process, σ_{SR} decreases and there are new cases that satisfy the condition of outliers, so the treatment described above should be repeated until ensuring that there are no outliers in the analyzed sample. This procedure can also be performed based on the calculation of Cook´s and Mahalanobis´ distances. However, for these criteria it may happen that the iterative process is not convergent, in the sense that, after the elimination of each outlier and the repetition of the regression analysis, there are always one or several cases, for which the value of the corresponding distance calculated, differs greatly from those calculated for the rest of the cases. Therefore, its use must be cautious and always analyzing the variation experience dafter each elimination of outliers, by statisticians such as the coefficient of determination, as well as by the slope and intercept of the model obtained and their corresponding estimation errors.

Therefore, this paper analyzes the influence of outliers elimination method on the quality of the adjustment models, in the ECe vs. EC1:5 linear regression analysis.

Population and sample

a) Data used for the analysis: The data used in this work were obtained from the processing of soil samples taken at sites located in the Cauto River Basin, affected by salinity.

   The main study was carried out with 42 samples corresponding to 21 points sampled at depths of 0-20 and 20-40 cm in the Cayo Grande farm, Río Cauto. Table 1 shows the characteristics of the soil samples taken.

Table 1. Characteristics of soil sampling.

Code	Place	Site	Points	Layers	Width
CG	Yara	Cayo Grande	64	4	20 cm
JIG-EG-U18-2	Jiguaní	Emp. Gen. "M.Fajardo". Unidad 18	5	5	20 cm
JUC-LP-B2-1	Jucarito	Lote Piloto. Bloque 2.	6	5	20 cm
JUC-LP-B2-2	Jucarito	Lote Piloto. Bloque 2.	7	4	15 cm
JUC-LP-B3	Jucarito	Lote Piloto. Bloque 3.	8	2	20 cm
MIR	Mir	Sector de Referencia. Monte Alto	6	4	15 cm
RC-G21-B23-C36-1	Río Cauto	Granja 21. Bloque 23. Campo 36.	9	5	20 cm
RC-G21-B23-C36-2	Río Cauto	Granja 21. Bloque 23. Campo 36.	7	6	15 cm
RC-G21-B23-C38	Río Cauto	Granja 21. Bloque 23. Campo 38.	9	4	15 cm

b) Measurement of the electrical conductivity of the saturation extract: The methodology for measurement the saturation extract electrical conductivity followed in this work is explained in detail in the specialized literature (USSL, 1954).

c) Measurement of 1:5 extract electrical conductivity: For measurement 1:5 extract electrical conductivity, the methodology also proposed by the USSL (1954) was followed.

d) Obtaining the calibration equations: Outliers analysis. To study the feasibility of estimating the saturation extract electrical conductivity from data obtained by applying the 1:5 extract electrical conductivity, the linear regression analysis was used (Walpole and Myers, 1989). The coefficients R and R^2 were calculated as well as the slope of the regression line, its intercept and the standard error of the estimate.

Based on the use of the Statistica statistical package (Statsoft, 2011), a comparative study of the most convenient method for identifying outliers was carried out.

In each sample to be analyzed, the impact of outliers elimination on the subsequent adjustment made with the remaining data was evaluated. The coefficient of determination, the standard error of the estimate, the values of the slope and intercept and their corresponding estimation errors as well as the analysis of variance corresponding to each regression were taken as comparative indices. The outliers identification methods were:

- The comparison of standard waste with the standard error of the distribution of these errors,
- Cook's distance analysis and
- Mahalanobis distance analysis.

The methodology consisted of initially carrying out an analysis of residues in the original database, identifying outliers according to the criteria used. The next step was to eliminate from this base, the case to which the highest value of the index used for the identification of the outliers corresponded (standard residue, Mahalanobis` distance or Cook`s distance).

Then, the regression analysis was repeated to the modified database. This process, in the case of the first index used, was repeated until there were no cases that met the elimination condition ($|SRV| > 2\ \sigma_{SR}$). In relation to the other two indices, the process of eliminating outliers was based on the fact that the computer system calculates these distances for each case considered in the regression, so those whose values differed from the corresponding ones were considered as outliers to most cases. For the last two indices, the same number of outliers were eliminated as for the first one, because after each regression analysis performed, there were always cases whose calculated distances differed from those of the rest of them.

Results Analysis

For the analysis of the influence of the outliers on the quality of the regressions, it was used data of saturation extract electrical conductivity (ECe) and 1:5 extract electrical conductivity (EC1:5) corresponding to the sampling of Cayo Grande (CG). Data corresponding to the depths of 0-20 and 20-40 cm in the 21 sampled points were considered, which yielded a total of 42 pairs of data.

Figure 1 shows the simple linear regression line between both variables considering the 42 cases mentioned above. The good quality of the adjustment can be seen firstly by the value of the coefficient of determination and secondly by the relative low dispersion of the experimental points with respect to said regression line. However, it is observed that many of the points fall outside the confidence intervals of the estimate, which suggests a deeper analysis to verify the reliability of the model obtained. The first step was to eliminate one by one, those cases for which, after each regression analysis performed, their corresponding standard residue (SRV) exceeded twice the standard deviation corresponding to the total of said errors.

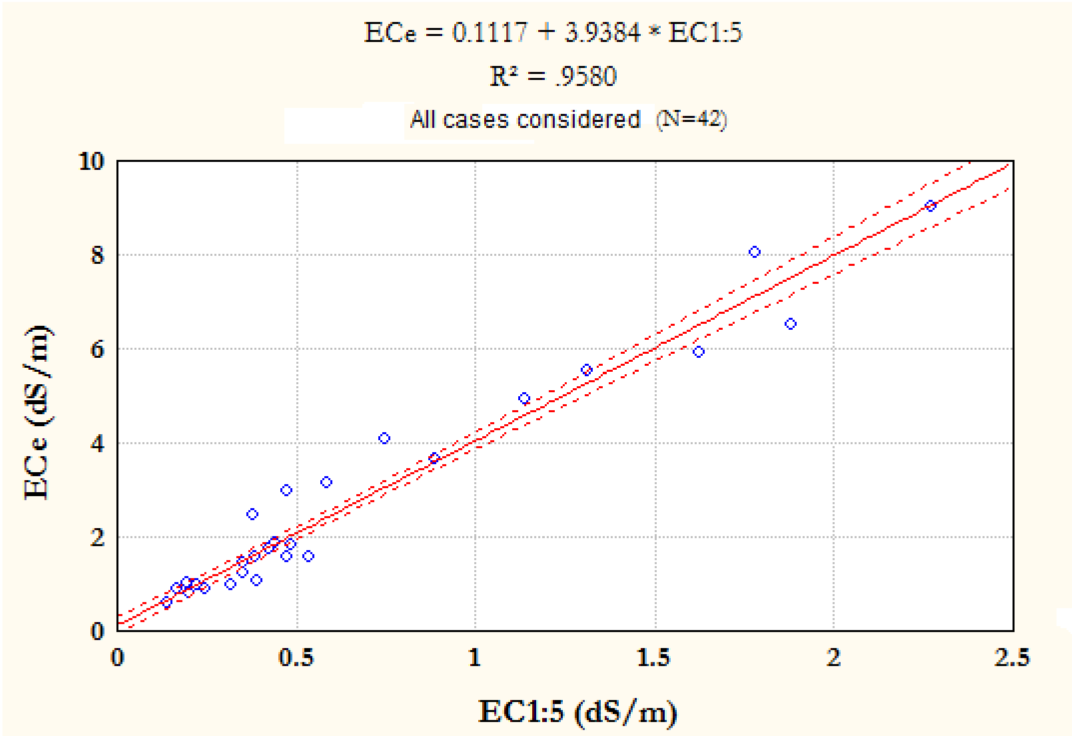

Figure 1.Simple linear regression analysis using all sample data. Cayo Grande (CG).

Obviously this criterion is based on the assumption of the normal distribution of these errors. After the elimination of an outlier, the simple linear regression analysis was carried out again to the remaining data, repeating this process until completing the elimination of 21 outliers, according to this criterion. This procedure was repeated from the original database, but considering as outlier, after each regression, the case to which the greatest value of Cook's distance corresponded. Unlike the previous case, the statistical system used does not use a discriminant criterion or threshold for said distance in relation to the categorization or not of a case as outlier. Therefore, the process of eliminating the outliers assumptions, according to this criterion, was repeated the same number of times (21) as in the previous case. The aforementioned was also carried out using the Mahalanobis` distance as the elimination criterion.

It should be noted that during the elimination procedure, in general for the criteria used, the cases that are eliminated in each one and the order in which said elimination is carried out do not coincide. This is clearly seen in Table 2, where the exclusion sequence followed in each analysis appears. None of the three parameters analyzed shows a tendency in their variation or a correspondence of one with the other.

Table 2.-Sequence of outliers elimination for each method.

Total of cases eliminated	N	Outliers					
		$\|SRV\| > 2\sigma$		Cook`s Dist.		Mahalanobis`Dist.	
		Case	|SRV|	Case	C.D.[1]	Case	M.D.[2]
0	42	23	2.3583	33	0.7668	33	6.8317
1	41	11	2.5704	37	0.4914	37	8.5764
2	40	37	2.6391	15	0.3957	15	9.1919
3	39	2	2.7794	16	1.4275	36	7.6609
4	38	39	2.7726	11	0.2503	32	8.4149
5	37	33	2.4505	23	0.1435	12	6.2227
6	36	17	2.2477	39	0.1457	11	5.9747
7	35	21	2.2284	32	0.2724	39	3.6120
8	34	19	2.3959	36	2.1195	17	3.6662
9	33	15	2.4967	12	1.9024	18	2.9006
10	32	40	2.1481	2	0.5082	42	3.0856
11	31	42	2.3422	9	0.2518	23	3.6265
12	30	38	2.4273	18	0.2512	9	3.0648
13	29	32	2.2060	30	0.3698	30	2.9418
14	28	41	2.1792	8	0.2440	19	2.2380
15	27	20	2.3381	42	0.2176	8	2.6965
16	26	36	2.4276	21	0.2152	2	2.8765
17	25	18	2.1630	19	0.2930	21	3.0623
18	24	3	2.1992	29	0.2445	29	2.7433
19	23	34	2.2053	22	0.2065	20	3.3260
20	22	1	2.0252	28	0.2071	41	3.4704
21	21	ok		17	0.8400	38	3.2122

[1] Cook`S Distance

[2] Mahalanobis` distance

This behavior is logical if we take into account that for its calculation it is necessary to consider the cases that remain in the sample, which are decreasing in each step and in a different way for each method.

Figure 2A shows the variation experienced by the coefficient of determination (R^2) after the elimination of each outlier, according to the three methods used. As can be seen, as cases are eliminated, the regression analysis performed on the remaining data has a different R^2. However, this variation is only optimal for the first case where, at the conclusion of the process, the maximum possible value is practically reached.

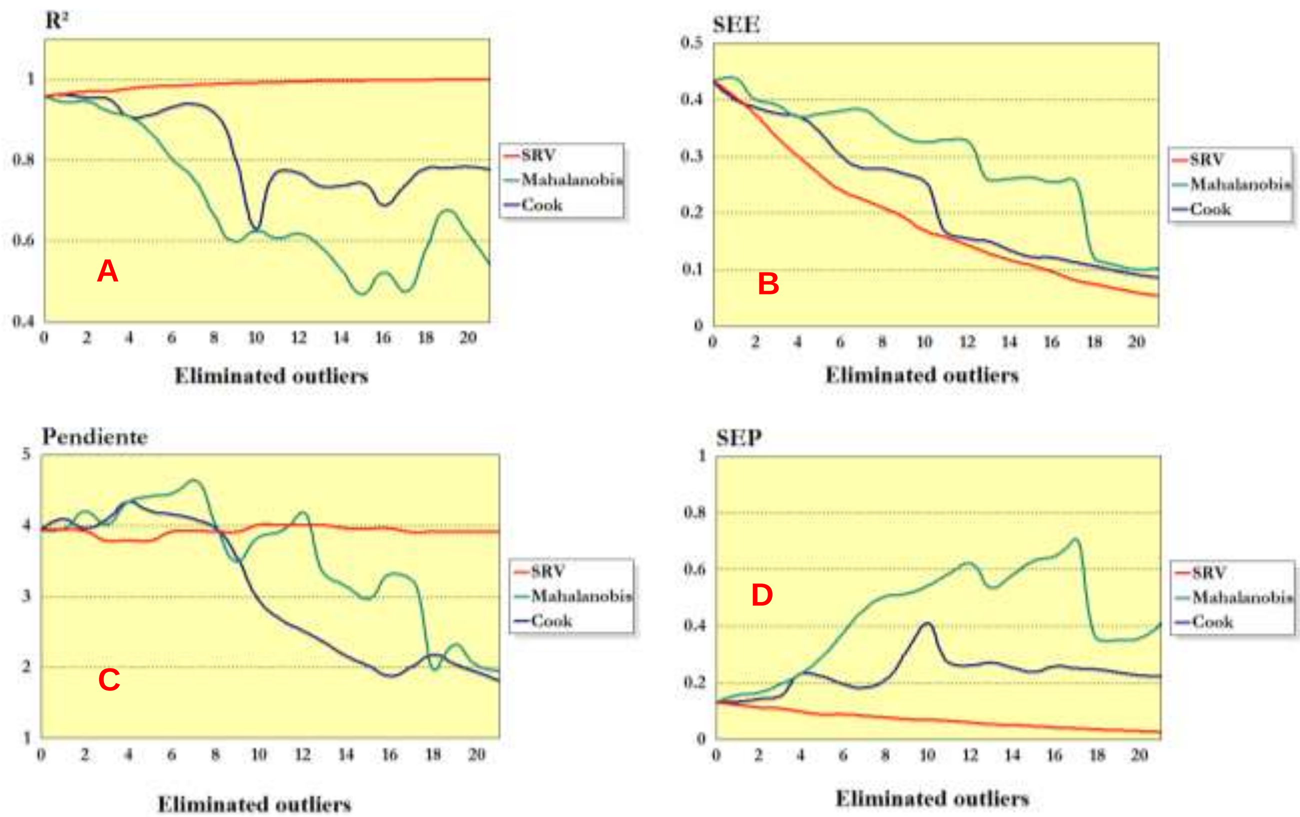

Figure 2.Effect of outliers elimination in the simple linear regression model of each studied method on: A) the coefficient of determination, B) the standard error of the estimation, C) the slope, D) the standard error of the slope.

The other two methods lead to an impoverishment of the quality of adjustment especially when using Mahalanobis distance.

An analysis similar to the previous one is shown in Figure 2B, but considering the standard error of the estimate (SEE).

Again, it can be seen that the elimination of outliers from the values of the standard residues, leads to the greater decrease of this index, which to a large extent also reflects the quality of the adjustment. The behavior of the standard error of the estimate for the other two methods is similar to that of the coefficient of determination.

It should be noted that for the first of the methods discussed, at the end of the elimination process, both statisticians tend to stabilize, experiencing very little variation, which indicates that the optimum in the regression model obtained has been achieved. As will be seen later, this behavior is also observed in the value of the slope and its standard error, reflected in Figures 2C and 2D.

The slope is especially important in the final regression model that is accepted, given the predictive nature of its use. Given the relatively small number of data used in the regression, it is expected that the deletion of one or more cases from the database could lead to significant changes in the value of the slope. This does not happen when the elimination method is the comparison of the standard residues and it is evident that it happens when the distances of Cook or Mahalanobis are considered, for which the successive elimination of cases significantly decreases the value of the slope. In relation to the standard error of the slope, its behavior reflected in Figure 2D, for the comparison of standard errors, is similar to that already analyzed for the standard error of the estimate. However, for the other two criteria the tendency of this parameter is to increase as outliers are eliminated, which is obviously undesirable as discussed above.

Graphically, the results of the analyzed process can be compared through Figures 3A, 3B and 3C. The regression lines obtained after the elimination process of the 21 outliers are presented according to each criterion studied. It is obvious to point out the excellent result shown by the comparison of the standard residues, in comparison with the models that lead to the application of the other two methods of optimization of the regression analysis. For the purpose of comparison, all the graphs were built on the same scale, which allows corroborating what was previously stated about the elimination of those cases with greater value of the independent variable. The manifest superiority of the first method is also observed in relation to the confidence intervals of the regression and although the Cook`s distance method generally offers better results than the Mahalanobis` distance, its application would lead to obtaining a model of linear regression very different from that obtained from the original data set.

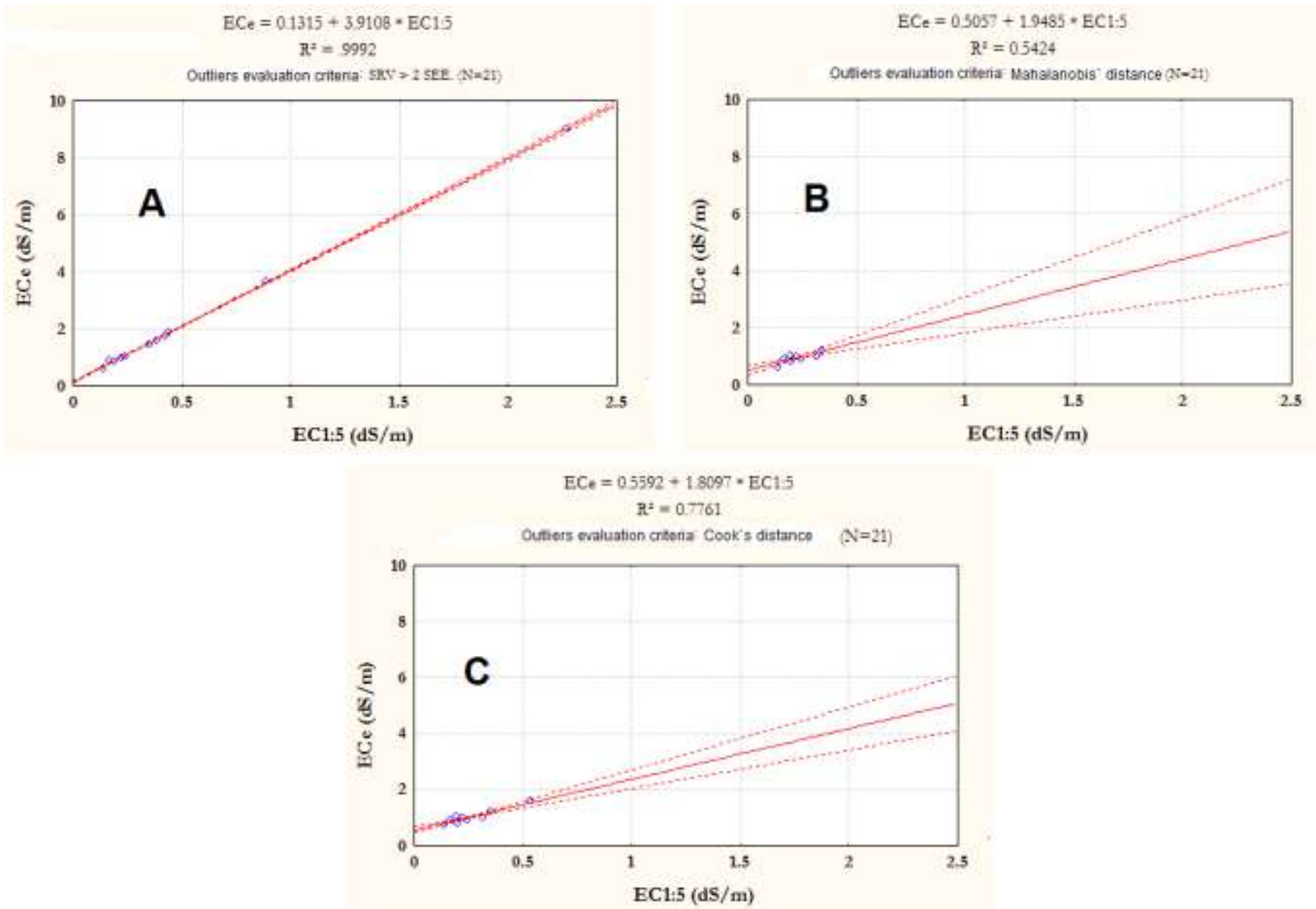

Figura 3. Results of the simple linear regression analysis after the elimination of 21 outliers.

A) |SRV| > 2 σ_{SR} , B) Mahalanobis` distance, C) Cook`s distance.

As a summary, Table 3 shows the results of applying the simple linear regression analysis to the original data set and the samples obtained after the elimination of the outliers considering the three criteria analyzed.

From the study of the statisticians reflected in this table, the inappropriateness of the use of Cook and Mahalanobis distances as criteria for the elimination of outliers in simple linear regression analysis applied to soil salinity indices is appreciated.

Secondly, it is evident that the elimination of those cases that, for a given model, their standard residues are greater than the standard deviation of the set of standard residues, leads to a statistically much more reliable regression model, which is what desired considering the predictive use of said model.

Table 3 Final results of the simple linear regression.

	Original	Method of outliers elimination		
	Data	$\|SRV\|>2 \cdot \sigma SR$	Cook	Mahalanobis
Model: $Y=\beta o+\beta 1 X$				
N	42	21	21	21
R	0.9788	0.9996	0.8810	0.7365
R^2	0.9580	0.9992	0.7761	0.5424
R^2aj	0.9570	0.9991	0.7643	0.5183
SEE	0.4331	0.0543	0.0863	0.1028
Intercept				
βo	0.1117	0.1315	0.5592	0.5057
SE	0.0953	0.0152	0.0584	0.0925
t	1.1723	8.6291	9.5697	5.4645
p	0.2480	0.0000	0.0000	0.0000
Slope				
$\beta 1$	3.9384	3.9108	1.8097	1.9485
SE	0.1303	0.0260	0.2230	0.4106
t	30.2245	150.4983	8.1152	4.7454
p	0.0000	0.0000	0.0000	0.0001
Sum of Squares				
Regression	171.3855	66.7111	0.4910	0.2380
Residual	7.5044	0.0560	0.1416	0.2008
Total	178.8899	66.7671	0.6326	0.4388
Mean Square				
Regression	171.3855	66.7111	0.4910	0.2380
Residual	0.1876	0.0029	0.0075	0.0106
Anova				
F	913.5180	22649.7344	65.8567	22.5191
p-level	0.0000	0.0000	0.0000	0.0001

In addition to the above, it can be verified that the model obtained after the elimination of the 21 outliers according to the standard residues, is based on data with the same range as the original data, which of course, guarantees the quality of the estimate in that same interval.

Conclusions

1. For the detection and elimination of outliers, in order to obtain the best linear regression model for the estimation of the saturation extract electrical conductivity from measurements of the 1:5 extract electrical conductivity, the most convenient and simple method it is the one that determines those cases for which its standard residue exceeds double the standard deviation of the set of standard residues of the sample considered in each analysis.
2. As a result of the study carried out, it is possible to obtain excellent adjustments in the regressions of the variables used, which reaffirms the usefulness of the analyzed method to estimate values of the electrical conductivity of the saturation extract.

Bibliografic references

Adnan, R. &Mohamad, M. N. (2003). Multiple Outliers Detection. Procedures in Linear Regression. *Matematika*, 19, (1), 29–45.

Angiulli,F., Basta S. & Clara Pizzuti.(2006). Distance-Based Detection and Prediction of Outliers.*IEEE Transactions on Knowledge and Data Engineering*, 18, (2).

Cousineau, D. &Chartier, S. (2010). Outliers detection and treatment: a review. *International Journal of Psychological Research*, 3, (1), 58-67.

Jiang, M.F., Tseng S.S. & Su C.M. (2001).Two-phase clustering process for outliers detection. *Pattern Recognition Letters,* 22, 691-700.

Meltem, E.&Ufuk, E. (2016).Outlier detection with Mahalano bissquare distance: incorporating small sample correction factor, *Journal of Applied Statistics*, DOI: 10.1080/02664763.2016.1255313.

Naddler, A. &Dasberg,S. (1980). A comparison of different methods for measuring soil salinity. *Soil Science Society of America Journal*, 44, 725-728.

Slavich, P.G. &Petterson,G.H. (1990).Estimating solution extract salinity from soil paste electrical conductivity: an evaluation of procedures. *Australian Journal of Soil Research*, 28, 517-522.

StatSoft, Inc. (2011). STATISTICA (data analysis software system), version 10.www.statsoft.com.

Tao, W., Qun, L., Bin, C.& Zhonghua, L. (2018). Multiple outliers detection in sparse high-dimensional regression, *Journal of Statistical Computation and Simulation*, 88 (1), 89-107, DOI: 10.1080/00949655.2017.1379521

USSL.(1954). Diagnosis and improvement of saline and alkaline soils. *United States Department of Agriculture. Agricultural Handbook N° 60. Washington, D.C*.

Xuqing, L., Feng, G., Yandong, W.&Zhiguo, Z. (2018).Detecting outliers and influential points: an indirect classical Mahalanobis distance-based method, *Journal of Statistical Computation and Simulation*, DOI: 10.1080/00949655.2018.1448981