



ISSN: 2319-9490

RESEARCH ARTICLE

HOW REGRESSIVE OBJECTIVE REGRESSION METHODOLOGY FINDS INFORMATION BEYOND A WHITE NOISE

^{1,*}Ricardo Osés Rodríguez, ²Jaime Wilfrido Aldaz Cárdenas, ³Rigoberto Fimia Duarte,
²Jorge Jagger Segura Ochoa, ⁴Iosbel Burgos Alemán, ⁵Claudia Osés Llanes,
⁶Nancy Guadalupe Aldaz Cárdenas, ⁶Jenny Janeth Segura Ochoa, ²Danilo Fabián Yáñez Silva,
²Víctor Danilo Montero Silva, ²Fredy Rodrigo Guillín Nuñez and ²Edison Rivelino Ramón Curay

¹Provincial Meteorological Center (CMP), Marta Abreu prolongation # 59 upstairs Juan Bruno Zayas corner, Villa Clara CUBA

²Career of Veterinarian Medicine, Faculty of Agricultural Science, Natural and Environmental Resources, Bolívar State University, Bolívar Province, Ecuador

³Faculty of Health Care Technology, University of Medical Sciences, Villa Clara, Cuba

⁴Provincial Hospital "Arnaldo Milian Castro", Santa Clara City, Cuba

⁵Center of Health, Epidemiology and Microbiology, Santa Clara Cuba

⁶"Canine and Feline" Veterinarian Clinic, Guaranda, Bolívar Province, Ecuador

Received 03rd April, 2018; Accepted 17th May, 2018; Published 30th June, 2018

ABSTRACT

The objective of this paper is to determine how methodology ROR can give information when having a series, whose autocorrelation diagrams are a white noise. Deceased by car accidents variable was used for Cuba in the period 2006- 2011. The future projection of the data series can be obtained by using modeling ROR, which provides a new and important way, promising for the series that behave as a white noise, because this gives new information for the series and its behavior. Deceased tendency is to diminish at about 29 persons a year; deceased persons depend on quantity of deceased two years backwards. As the model is perfect and errors with zero value are obtained, Cristóbal Numbers are introduced, which could create new statistics and prediction models. This study was done using the statistical package of Social Sciences (SPSS) Version 13.

Key words: Cuba, Methodology ROR, prediction models, prognosis, Regression, white noise.

Copyright © 2018, Ricardo Osés Rodríguez *et al.* This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Citation: Ricardo Osés Rodríguez, Jaime Wilfrido Aldaz Cárdenas, Rigoberto Fimia Duarte *et al.*, 2018. "How regressive objective regression methodology finds information beyond a white noise" *International Journal of Current Research in Life Sciences*, 7, (06), 2257-2261.

INTRODUCTION

The information in a data series is important when projecting the future, but what happens when the modeling series is a white noise? And there is no information in previous steps allowable to model for the future. In this article methodology ROR is treated and how important information is obtained by using it to project the future behavior of the series. Methodology ROR (Regressive methodology¹) is carried out in several steps, which are explained in this article; however it is necessary to detail in this work since mathematic point of view. In this methodology a curve adjustment is made using the square minimum method, which is explained below. It is frequently necessary to represent by a functional relation, data given as a group of points X -Y. For example, an experiment has been carried out and the points have been obtained X – Y

represented in a graphic of Y against X. As these points are going to be used for computer calculus, several problems are confronted. There are experimental errors in the values of Y. Variations due to experimental errors must be softened somehow.

1. The value of Y is desirable to know, corresponding to some value of X, which is found between two experimental values of X.
2. It would be desirable – in fact it can be the main purpose of the calculus – extra polar, that is to determine Y value, corresponding to X value outside the rank of experimental values of this variable.

All these considerations lead to the necessity of a functional relation between X and Y in equation form and this must be simple. The question is to determine a curve to approximate data with adequate precision. The first question presented is this: How is it going to be decided if a given curve is a good "adjustment" to data?

**Corresponding author:* Ricardo Osés Rodríguez,
Provincial Meteorological Center (CMP), Marta Abreu prolongation # 59
upstairs Juan Bruno Zayas corner, Villa Clara CUBA.

This discussion will be simpler if a new term is defined. The deviation in a given point is the difference between Y experimental value and Y calculated value from a functional relation. The question to adjust a curve to data can be formulated: What condition can be put to the deviations to achieve an adequate curve?

An attractive possibility is to ask for the sum of deviations as small as possible. If a prima is used to identify 2 Y values calculated beginning from a searched functional relation, this means to ask for

$$\sum_{i=1}^N (Y_i - Y_i')$$

be a minimum, in which N is the number of data points but the attractive of this possibility disappears when considering the case simply to adjust a line at two points. This difficulty could be avoided when specifying absolute values, that is to require minimizing

$$\sum_{i=1}^N |Y_i - Y_i'|$$

To find a minimum value, it cannot be derived because the absolute value function does not have a derivate in its minimum. It could be asked for the maximum error to be a minimum, which is Chebyshev approximation, but this leads to a complicated iterative process to determine the functional relation. Thus, the square minimums criteria arrives, in which it is asked for a minimum value of

$$\sum_{i=1}^N (Y_i - Y_i')^2$$

As it can be seen, this expression can be differentiated to determine its minimum. It leads to lineal equations, which in many cases are of practical interest, and from the beginning are easy to solve. Finally, there are statistical considerations that suggest the square minimums criteria is a good criterion, besides the computer facilities. Thus, the function of approximation for methodology ROR is written as follows:

$$Y_i' = c_1 * \delta_1(x_i) + c_2 * \delta_2(x_i) + c_3 * \text{NoC}(x_i).$$

Where:

$$\delta_1(x_i) = \begin{cases} 0 & \text{if } X_i = 2n \\ 1 & \text{if } X_i = 2n+1. \end{cases} \quad n=0,1,\dots,N$$

$$\delta_2(x_i) = \begin{cases} 1 & \text{if } X_i = 2n \\ 0 & \text{if } X_i = 2n+1. \end{cases} \quad n=0,1,\dots,N$$

$$\text{NoC}(x_i) = x_i, \quad x_i = 0,1,\dots,N.$$

The objective is to determine C1, C2 and C3, to minimize

$$S = \sum_{i=1}^N (Y_i - Y_i')^2$$

$$S = \sum_{i=1}^N (Y_i - c_1 * \delta_1(x_i) - c_2 * \delta_2(x_i) - c_3 * \text{NoC}(x_i))^2$$

As it is known when minimizing S considered as function of C1, the partial derivate of S is equal to zero with respect to C1, the result is:

$$(\partial S / \partial c_1) = (-2) \sum_{i=1}^N (Y_i - c_1 * \delta_1(x_i) - c_2 * \delta_2(x_i) - c_3 * \text{NoC}(x_i)) = 0.$$

Making equal to zero and readequating, it is obtained:

$$c_1 * \sum_{i=1}^N \delta_1(x_i) + c_2 * \sum_{i=1}^N \delta_2(x_i) + c_3 * \sum_{i=1}^N \text{NoC}(x_i) = \sum_{i=1}^N Y_i.$$

$$c_1 * \left(\sum_{i=1}^N \delta_1(x_i) \right) + c_2 * \left(\sum_{i=1}^N \delta_2(x_i) \right) + c_3 * \left(\sum_{i=1}^N \text{NoC}(x_i) \right) = \sum_{i=1}^N Y_i.$$

Deriving S with respect to C2 and then with respect to C3 and making each of the results equal to zero, two more equations are obtained in the unknown C1, C2, C3, the three simultaneous equations in these three incognitos are named normal equations to adjust an equation to the set of data, then:

$$c_1 * \left(\sum_{i=1}^N \delta_1(x_i) \right) + c_2 * \left(\sum_{i=1}^N \delta_2(x_i) \right) + c_3 * \left(\sum_{i=1}^N \text{NoC}(x_i) \right) = \sum_{i=1}^N Y_i.$$

$$c_1 * \left(\sum_{i=1}^N \delta_1(x_i) \right) + c_2 * \left(\sum_{i=1}^N \delta_2(x_i) \right) + c_3 * \left(\sum_{i=1}^N \text{NoC}(x_i) \right) = \sum_{i=1}^N Y_i.$$

$$c_1 * \left(\sum_{i=1}^N \delta_1(x_i) \right) + c_2 * \left(\sum_{i=1}^N \delta_2(x_i) \right) + c_3 * \left(\sum_{i=1}^N \text{NoC}(x_i) \right) = \sum_{i=1}^N Y_i.$$

To find a “better” function for data; it is only necessary to carry out the sums needed and solve the three equations system, this combination explains a great quantity of variance Y_i , then it is obtained:

$$Y_i' = c_1 * \delta_1(x_i) + c_2 * \delta_2(x_i) + c_3 * \text{NoC}(x_i).$$

The errors left are $e_i = (Y_i - Y_i')$ then the cross correlation is calculated of e_i with $Y_{i-n}(x_i)$ in the following formula:

$$\text{Corr}(e_i, Y_{i-n}(x_i)) = \frac{\text{Cov}(e_i, Y_{i-n}(x_i))}{[\text{Var}(e_i) * \text{Var}(Y_{i-n}(x_i))]^{1/2}} \text{ and the maximum value of that}$$

function is selected.

Function that is the corresponding peak named t, then variable Y_t is calculated and the system is resolved this time with variable Y_t ,

$$(\partial e_i / \partial c_i) = (\partial S_2 / \partial c_i) = 0 \text{ this time with the function:}$$

$$S_2 = (Y_i - c_1 * \delta_1(x_i) - c_2 * \delta_2(x_i) - c_3 * \text{NoC}(x_i) - c_4 * Y_t(x_i))^2,$$

Then an error e_2 is left, which is cross correlated with $g_{i-k}(x_i)$ as hexogen variable similar to e_i obtaining a new peak in t for variable $g_{i-k}(x_i)$, and the system is resolved again (t can be of different order to the one calculated for the function $Y_{i-n}(x_i)$).

Coefficients^{a,b}

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	DI	-981.340	308.741	-.1381	-3.179	.003
	DS	-795.908	304.288	-.1120	-2.616	.013
	NoC	7.166	3.007	.374	2.383	.023
	Lag3angiostot	.880	.190	.719	4.630	.000
	Lag3XY1	33.632	12.277	1.626	2.739	.010

a. Dependent Variable: Angiostotal

b. Linear Regression through the Origin

Table 1. Correlograms of Deceased variable
Autocorrelations

Series: Deceased

Lag	Autocorrelation	Std.Error ^a	Box-Ljung Statistic		
			Value	df	Sig. ^b
1	.309	.323	.918	1	.338
2	.114	.289	1.074	2	.584
3	-.212	.250	1.791	3	.617
4	-.300	.204	3.949	4	.413

a. The underlying process assumed is independent (white noise).

b. Based on the asymptotic chi-square approximation.

This time $(\partial_{ci}/\partial c_i) = (\partial S_3/\partial c_i) = 0$ in such way that

$$S_3 = (Y_i - c_1 * \delta_1(x_i) - c_2 * \delta_2(x_i) - c_3 * \text{NoC}(x_i) - c_4 * Y_t(x_i) - c_5 * g_t(x_i))^2,$$

An error e_4 is finally obtained, which must have media zero and variance 1 and the process is stopped obtaining the highest quantity of variance as possible, in this approximation data of the same function $Y_{i-n}(x_i)$ and hexogen data of the function $g_{i-k}(x_i)$. This methodology has been used in the model for angiostrongilosis³ variable, where it was obtained the following model of function:

Where $DS = \delta_1(x_i)$ and $DI = \delta_2(x_i)$ NoC = NoC(x_i), is the tendency and $\text{Lag3angiostot} = Y_{i-n}(x_i)$ is the regressive angiostrongilosis in three bimonthly periods ($t=3$) and lag3XY1 is the hexogen variable Mean Temperature in Yabú station ($g_{i-k}(x_i)$) regressive in three bimonthly periods where t is equal to 3, the same as for angiostrongilosis. The objective of this paper is to determine how methodology ROR can give information when having a series, whose auto correlograms are a white noise.

MATERIALS AND METHODS

For this work the quantity of death by car accident was used, in Cuba from 2006 to 2011, besides the ROR methodology was used as (Osés, R. and Grau, 2011), the methodology of Box and Jenkins (Box *et al.*, 1994) was not used since this has limitations as (Osés, R. and Grau, 2011), we used SPSS Version 13. The Regressive method was used for the analysis (Osés, R. and Grau, 2011).

This methodology is also used in the forecast of great intensity earthquakes in Cuba (Osés *et al.*, 2012a), besides it was implemented in mosquitoes control in (Fimia *et al.*, 2012) and these results were used in the study of climatic change applied to animal health in Villa Clara, Cuba (Osés *et al.*, 2012), the mathematical modeling was applied to Malaria (Fimia *et al.*, 2012). The ROR methodology is widespread to meteorology, for example in the modeling of cold front and the impact of sun spots as in (Osés *et al.*, 2012b), The ROR methodology is also applied to long term prediction of larval density of anopheles mosquitoes in (Osés *et al.*, 2012), besides in (Osés *et al.*, 2014) was done a long term-prognostic (1 year in advance) to obtain daily forecast of meteorological variables in Sancti Spiritus, Cuba. The Regressive methodology opens a high range of applications to the modeling of any time series of data. Data for the model obtained for the deceased were taken from State Statistics Office (Oficina Estatal de Estadísticas).

RESULTS AND DISCUSSION

Below it is shown the total correlograms and the partial autocorrelations of Deceased variable (Table 1 and Table 2). As it can be seen a white noise is presented. Then, there is no information in previous steps in the series allowable to predict the future; however the methodology ROR¹ is applied to the deceased variable obtaining the following model. The model is perfect, it explains the 100 % of variance, and the model has not errors (Table 3). In Table 4 the model parameters can be shown. The tendency NoC of the deceased is to diminish at about 29 persons, the deceased depend on the quantity of deceased two years backwards (Lag2deceased). As the standard errors are zero in Table 4,

Table 2. Partial Autocorrelogram of Deceased series**Partial Autocorrelations**

Series: Deceased

Lag	Partial Autocorrelation	Std. Error
1	.309	.408
2	.021	.408
3	-.279	.408
4	-.192	.408

Table 3. Summary of the model for the quantity of deceased**Model Summary^{c, d}**

Model	R	R Square ^a	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	1.000 ^b	1.000	.	.	1.998

- a. For regression through the origin (the no-intercept model), R Square measures the proportion of the variability in the dependent variable about the origin explained by regression. This CANNOT be compared to R Square for models which include an intercept.
- b. Predictors: Lag2Deceased, DI, NoC, DS
- c. Dependent Variable: Deceased
- d. Linear Regression through the Origin

Table 4. Model Parameters of deceased at National level**Coefficients^{a, b}**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	DS	826.886	.000	.801	.	.
	DI	823.773	.000	.798	.	.
	NoC	-29.250	.000	-.186	.	.
	Lag2Deceased	.045	.000	.049	.	.

- a. Dependent Variable: Deceased
- b. Linear Regression through the Origin

Table 5. Results of the model
Unstandardized Predicted Value
Unstandardized Residual
Case Summaries^a

	Year	Deceased	Unstandardized Predicted Value	Unstandardized Residual
1	2006.00	855.00	.	.
2	2007.00	775.00	.	.
3	2008.00	778.00	778.00000	.00000
4	2009.00	742.00	742.00000	.00000
5	2010.00	716.00	716.00000	.00000
6	2011.00	682.00	682.00000	.00000
7	2012.00	.	654.68182	.
8	2013.00	.	620.77273	.
Total N	8	6	6	4

- a. Limited to first 100 cases.

the statistical value T of Student (t) and its meaning cannot be obtained. As it is known: $t_1 = B_1 / S_{b_1}$ (according to (Sánchez and Torres, 1986), where B, are the coefficient of the regression and S_{b_1} are the standard errors and t_1 is the student t test parameters'; that is why to calculate t of each coefficient is impossible, because a division by zero must have been developed; for this it is suggested to use the calculus form according to (Osés, 2007), in which it is presented how to operate Cristosols numbers (numbers divided by zero). Another work that can be consulted in relation to this, it can be seen in (Osés, 2012) in which are presented numbers seen as a ring and as a group. Through these numbers new statistics and prediction models can be obtained. The prognosis for 2012 and 2013 is the following (Table 5).

Conclusion

Through modeling ROR, information has been obtained for the future projection of data series, which provides a new and important way, promising for the series that behave as a white noise, because this gives new information for the series and its behavior. Deceased tendency is to diminish at about 29 persons; deceased persons depend on quantity of deceased two years backwards. Through Cristosols Numbers new statistics and prediction models can be obtained.

REFERENCES

- Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. and 1994. Time Series Analysis Forecasting and Control, 3rd edition. Prentice –Hall Inc., New Jersey.
- Fimia R.; Osés, R. & Otero M. and 2012. La malaria. Modelación matemática en el análisis entomológico. Evolución de la malaria, modelación matemática y boletines de Vigilancia Epidemiológica en Cuba. Editorial Académica Española (eae. ISBN: 978-3-659-02519-8) for Marca registrada de Academic Publishing GmbH & Co. KG Academia, Germany.
- Fimia, R.; Osés, R., Otero, M., Diéguez, L., Cepero, O., González, R., Silveira, E. and Corona, E. and 2012. El control de mosquitos (Diptera: Culicidae) utilizando métodos biomatemáticos en la Provincia de Villa Clara. REDVET Rev. electrón. vet. [http://www.veterinaria.org/revistas/redvetVolumen 13 Nº 3](http://www.veterinaria.org/revistas/redvetVolumen%2013%20Nº%203) - <http://www.veterinaria.org/revistas/redvet/n03031.html>.
- McCracken, D. and Dorn, S. and 1971. Métodos numéricos y programación FORTRAN, Edición Revolucionaria. Instituto Cubano del Libro, 1971.
- Oficina Estatal de Estadísticas (www.one.cu)
- Osés R. and 2007. División por cero, un intento de resolver sistemas incompatibles, interpretación filosófica. Centro Meteorológico Provincial de Villa Clara, Cuba [inédito].
- Osés R., Fimia, R.; Silveira, E., Hernández, W., Saura, G., Pedraza A. and Gonzalez, R. and 2012. Modelación matemática hasta el año 2020 de la densidad larvaria anofelínica de mosquitos (Diptera Culicidae) en Caibarién, Provincia de Villa Clara Cuba. REDVET Rev. electrón. vet. [http://www.veterinaria.org/revistas/redvetVolumen 13 Nº 3](http://www.veterinaria.org/revistas/redvetVolumen%2013%20Nº%203) - <http://www.veterinaria.org/revistas/redvet/n030312.html>.
- Osés, R. and 2012. Los Cristosoles. Grupos y Anillos. Centro Meteorológico Provincial de Villa Clara, Cuba [inédito].
- Osés, R. and Fimia, R. and 2012. Modelos de Pronóstico de Angiostrongilosis y Fasciolosis. Editorial Académica Española. ISBN: 978-3-8484-5355-9.
- Osés, R. and Grau R. 2011. Modelación regresiva (ROR), versus modelación ARIMA, usando variables dicotómicas en mutaciones del VIH. Universidad Central Marta Abreu de las Villas, 25 de Febrero. Editorial Feijóo. ISBN: 978-959-250-652-7.
- Osés, R., Fimia, R.; Saura, G., Pedraza, A., Ruiz, N. and Socarrás, J. and 2014. Long Term Forecast of Meteorological Variables in Sancti Spiritus. CUBA. Applied Ecology and Environmental Sciences, 2 (1): 37-42. ISSN (Print): 2328-3912, ISSN (Online): 2328-3920. DOI: 10.12691/aees-2-1-6.
- Osés, R., Saura, G. and Pedraza, A. and 2012a. Modelación matemática ROR aplicada al pronóstico de terremotos de gran intensidad en Cuba, REDVET Rev. electrón. vet. <http://www.veterinaria.org/revistas/redvet>. Volumen 13, (05B) – <http://www.veterinaria.org/revistas/redvet/n050512B.html>.
- Osés, R., Saura, G. and Pedraza, A. and 2012b. Modelación de la cantidad de frentes fríos en Cuba, impacto de manchas solares. REDVET Rev. electrón. vet. Volumen 13 Nº 05B - <http://www.veterinaria.org/revistas/redvet/n050512>
- Osés, R.; Saura, G.; Otero, M.; Pedraza, A.; Socarrás, J. and Ruiz, N. and 2012. Cambio climático e impacto en la salud animal en la Provincia de Villa Clara, CUBA. REDVET Rev. electrón. vet. [http://www.veterinaria.org/revistas/redvetVolumen 13 Nº 05B](http://www.veterinaria.org/revistas/redvetVolumen%2013%20Nº%2005B) -<http://www.veterinaria.org/revistas/redvet/n050512B.html> ,
- Sánchez, R. and Torres, J.A. and 1986. Estadística Elemental. Editorial Pueblo y Educación, p 249, 1986.
