ISSN: 2319-6505

Available Online at http://journalijcar.org

International Journal of Current Advanced Research
Vol 4, Issue 2, pp 17-21, February 2015

International Journal of Current Advanced Research

ISSN: 2319 - 6475

RESEARCH ARTICLE

HOW MUCH METHODOLOGY ROR EXPLAINS THE RAIN ERRORS IN CAIBARIÉN, CUBA

Osés Rodríguez, Ricardo[1], Burgos Alemán, Iosbel [2], Osés Llanes, Claudia[3], Otero Martin Meylin[1], C. Fimia Duarte, Rigoberto[4] and Cruz Camacho, Lisvette[4]

[1]Meteorological Provincial Center, Villa Clara, Cuba
[2]Provincial Hospital Arnaldo Milián Castro, Santa Clara, Cuba
[3]Center of Health, Epidemiology and Microbiology, Santa Clara, Cuba
[4]Faculty of Health Care Technology, Santa Clara, Cuba

ARTICLE INFO

Article History:

Received5th, January, 2015
Received in revised form15th, January, 2015
Accepted12th, February, 2015
Published online 28th, February, 2015

Key words:
Cuba, Methodology ROR, Rain, Regression, White Noise

ABSTRACT

The objective of this work is to determine how much information the methodology ROR can give when having a series whose selfcorrelograms are a white noise. It was used the variable of the monthly fallen rain in Caibarién, Cuba, in the period 1977-2014. By using modeling ROR, it has been obtained information for the future projection of data series of errors by modeling ARIMA, as this type of modeling opens an important and promising way for the series that behave as white noise, by giving new information for the series and its behavior. The model explains the 8.7 % of variance, with errors of 44.6 mm. The errors tendency is the increase around 0.004 mm, although it is not statistically significant, the errors depend on the errors 1 month behind. All the work was carried out with the help of the statistical package of Social Sciences (SPSS) Version 13.

INTRODUCTION

The information in a data series is important when projecting the future, but: what happens when the modeling series is a white noise?, therefore there is no information in previous steps allowing to model towards the future. This article deals about methodology ROR and how through its use, important information can be obtained to project the future behavior of the series.

The Methodology **ROR** (Regression Objective Regressive[1]) is carried out in several steps which are explained in this article; however it is necessary to detail them since the mathematical point of view.

In this methodology, it is done an adjustment of curves using the method of square minimums, which are explained as follows.

It is frequently necessary to represent by means of a functional relation data that have been given as a group of points X - Y. For instance, an experiment has been made and the points X – Y have been obtained in a graphic of Y against X. As these points are going to be used for computer calculus, several problems have to be faced:

1. There are experimental errors in Y values. It should be softened somehow the variations due to experimental errors.
2. The value of Y is wanted corresponding to some value of X which is found between two experimental values of X.
3. It would be desirable– in fact it can be the main purpose of the calculus– to transfer, that is to determine the value of Y corresponding to a value of X out of the range 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 hopefully simple. The question is then to determine a curve to approximate data with enough precision. The first question is this: How to decide if a curve given is a good "adjustment" to data?

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

A possibility that could be attractive is to ask for the deviations sums can be as small as possible. If a prime is used to identify the values[2] of Y calculated beginnning from the wanted 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 simple case of adjusting a line at two points. This difficulty could be avoided by specifying absolute values, that is requiring to minimize.

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

But it cannot be derived to find a minimum value because the function absolute value has no derivate in its minimum. It can be thought to ask for the maximum error is a minimum, which is Chebyshev approximation, but this leads to an iterative process complicated to determine the functional relation.
This leads to the criterion of the square minimums, in which is inquired a minimum value of:

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

As it can be observed, this expression can be different to determine its minimum and leads to equations that are linear in many cases of practical interest, and from the beginning, are easy to solve. Finally, there are statistical considerations that suggest the square minimums are a good criterion, besides its computer facility.

Thus it is written the function of approximation for methodology ROR as follows:

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

Where:

$$\delta_1 (xi) = \begin{cases} 0 & \text{if } Xi = 2n \\ & n = 0,1\ldots N \\ 1 & \text{if } Xi = 2n + 1. \end{cases}$$

$$\delta_2 (xi) = \begin{cases} 1 & \text{if } Xi = 2n \\ & n = 0,1\ldots N \\ 0 & \text{if } Xi = 2n + 1. \end{cases}$$

$\mathbf{NoC}(xi) = xi$, $xi = 0,1,\ldots, N.$

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

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

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

It is known that to minimize S considered as function of C1 is equals to zero the partial derivate of S with respect to C1, the result is:

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

Making equal to zero and readjusting, it is obtained:

$$c_1 * (\sum_{i=1}^{N} \delta_1 (xi)) + c_2 * (\sum_{i=1}^{N} \delta_2 (xi)) + c_3 * (\sum_{i=1}^{N} \mathbf{NoC}(xi)) = \sum_{i=1}^{N} Yi.$$

$$c_1 * (\;)_{N}^{0} + c_2 * (\;)_{0}^{N} + c_3 * (\sum_{i=1}^{N} (xi)) = \sum_{i=1}^{N} Yi.$$

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

$$c_1 * (\;)_{N}^{0} + c_2 * (\;)_{0}^{N} + c_3 * (\sum_{i=1}^{N} (xi)) = \sum_{i=1}^{N} Yi.$$

$$c_1 * (\;)_{N}^{0} + c_2 * (\;)_{0}^{N} + c_3 * (\sum_{i=1}^{N} (xi)) = \sum_{i=1}^{N} Yi.$$

$$c_1 * (\;)_{N}^{0} + c_2 * (\;)_{0}^{N} + c_3 * (\sum_{i=1}^{N} (xi)) = \sum_{i=1}^{N} Yi.$$

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

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

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

$$Corr(ei, Y_{i-n}(xi)) = \frac{Cov(ei, Y_{i-n}(xi))}{[Var(ei) * Var(Y_{i-n}(xi))]^{1/2}}$$

and the maximum value is selected of this function that is the corresponding peak named t, it is calculated then the variable Y_t and the system is resolved again, this time with the variable Yt,

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

$$S_2 = (Yi - c_1 * \delta_1 (xi) - c_2 * \delta_2 (xi) - c_3 * \mathbf{NoC}(xi) - c_4 * Yt(xi))^2 ,$$

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

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

$$S_3 = (Yi - c_1 * \delta_1 (xi) - c_2 * \delta_2 (xi) - c_3 * \mathbf{NoC}(xi) - c_4 * Yt(xi) - c_5 * g_t (xi))^2 ,$$

At the end it is obtained an error e_4 which should 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 are used $Y_{i-n}(xi)$ and hexogen data of the function $g_{i-k}(xi)$.

This methodology has been used in the model for variable angiostrongilosis[3], where the following model of function was obtained:

Coefficients[a,b]

Mode		Unstandardized Coefficients B	Unstandardized Coefficients Std. Err	Standardized Coefficients Beta	t	Sig.
1	DI	-981.34	308.74	-1.381	-3.179	.003
	DS	-795.90	304.28	-1.120	-2.616	.013
	NoC	7.166	3.007	.374	2.383	.023
	Lag3angios	.880	.190	.719	4.630	.000
	Lag3XY1	33.632	12.271	1.626	2.739	.010

a. Dependent Variable: Angiostotal

b. Linear Regression through the Origin

Where DS= $\delta_1(xi)$ and DI = $\delta_2(xi)$ NoC= **NoC**(xi), is the tendency and Lag3angiostot = $Y_{i-n}(xi)$ is the regressive angiostrongilosis in three bimonthly (t=3) and lag3XY1 is the hexogen variable Mean Temperature in Yabú station ($g_{i-k}(xi)$ regressive in three bimonthly where t is equal to 3, the same for angiostrongilosis.

The objective of this work is to determine how much variance provides methodology ROR when modeling the errors of the variable modeled rain by ARIMA that is when there is a series whose selfcorrelograms are a white noise.

MATERIALS AND METHODS

It is used the variable monthly fallen rain in Caibarién (figure 1), Cuba, in the period 1977-2014, and the methodology ROR [1] was also used. To model the error series of rain, Box and Jenkins methodology was used [4]. It was worked with the statistical package spss version 13.

For the analysis of errors the regressive method was used [1].

This methodology is also used for the prognosis of high intensity earthquake in Cuba [5], besides it was implemented in mosquitoes control [6], and the results were used in the study of Climatic Change applied to animal health in Villa Clara, Cuba [7], the mathematical modeling was applied to malaria [8]. The methodology ROR is greatly spread in Meteorology, for example in the modeling of cold fronts and the impact of sun spots [9]. The methodology ROR is also applied for predictions of anopheles mosquito larval density [10]; moreover it was done a long term prognosis (one year) of meteorological variables in Santi Spíritus, Cuba [11].

The regressive methodology opens a wide range of applications for the modeling of any series data of time.

The data for the modeling obtained for the rain were taken from Provincial Meteorological Center [12].

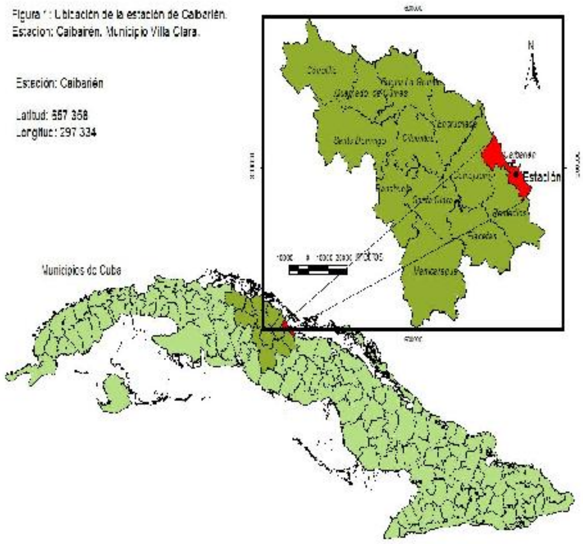

Figure 1 Location of Caibarién station. Villa Clara, Cuba.

RESULTS AND DISCUSSION

Subsequently, they are shown the total correlograms and partial auto correlations of errors of the variable rain in Caibarién (Table 1 and Table 2), this model was calculated with modeling ARIMA. As it can be appreciated there is a presence of a white noise.

Table 1 Correlograms of variable Rain in Caibarién modeled by ARIMA

Auto correlations

Series: Unstandardized Residual

Lag	Autocorrelation	Std.Error[a]	Box-Ljung Statistic Value	Box-Ljung Statistic df	Box-Ljung Statistic Sig.[b]
1	.000	.028	.000	1	.987
2	.006	.028	.050	2	.975
3	.042	.028	2.416	3	.491
4	-.011	.028	2.586	4	.629
5	-.028	.027	3.627	5	.604
6	-.003	.027	3.639	6	.725
7	-.020	.027	4.184	7	.758
8	-.018	.027	4.630	8	.796
9	.023	.027	5.341	9	.804
10	.020	.027	5.855	10	.827
11	-.021	.027	6.460	11	.841
12	.066	.027	12.314	12	.421
13	-.029	.027	13.446	13	.414
14	.030	.027	14.663	14	.402
15	.043	.027	17.089	15	.314
16	-.044	.027	19.635	16	.237

A. The underlying process assumed is independence (whitenoise).
B. Based on the asymptotic chi-square approximation.

Thus, there is no information in previous steps in the series allowing to predict towards the future, however methodology ROR[1] is applied to the errors of rain obtaining the following model.

The model explains the 8.7 % of variance; the model has errors of 44.6 mm (Table 3).

Table 2 Partial self correlograms of the rain in Caibarién modeled with ARIMA

Partial Autocorrelations

Series: Unstandardized Residual

Lag	Partial Autocorrelation	Std.Error
1	.000	.028
2	.006	.028
3	.042	.028
4	-.011	.028
5	-.029	.028
6	-.005	.028
7	-.019	.028
8	-.016	.028
9	.023	.028
10	.021	.028
11	-.021	.028
12	.063	.028
13	-.031	.028
14	.033	.028
15	.038	.028
16	-.041	.028

Table 3 Summary of the model for the series of errors of rain in Caibarién (xc7)

Model Summary[c,d]

Model	R	R Square[a]	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.087[b]	.008	.005	44.59382603	2.000

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: Lag1Error, DS, DI, NoC

c. Dependent Variable: Error for xc7 from ARIMA, MOD_1 NOCON

d. Linear Regression through the Origin

In Table 4 it is appreciated that the analysis of variance is significant with an F of Fisher of 2.5 significant to the 95 %.

Table 4 Analysis of variance of errors modeling

ANOVA[c,d]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	19870.256	4	4967.564	2.498	.041[a]
	Residual	2609055	1312	1988.609		
	Total	2628926[b]	1316			

a. Predictors: Lag1Error, DS, DI, NoC

b. This total sum of squares is not corrected for the constant because the constant is zero for regression through the origin.

c. Dependent Variable: Error for xc7 from ARIMA, MOD_1 NOCON

d. Linear Regression through the Origin

In Table 5 it can be appreciated the parameters of the model. The tendency NoC of errors is to increase in 0.004 mm, the errors depend on the errors one month behind (**Lag1Error**). Although DI and the tendency NoC are not significant, they are kept in the model because they provide variance to the model.

Table 5 Parameters of the model of errors of rain in Caibarién.

Coefficients[a,b]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	DS	-6.362	2.846	-.101	-2.235	.026
	DI	-2.819	2.848	-.045	-.990	.322
	NoC	.004	.003	.067	1.165	.244
	Lag1Error	.053	.028	.053	1.936	.053

a. Dependent Variable: Error for xc7 from ARIMA, MOD_1 NOCON

b. Linear Regression through the Origin

CONCLUSIONS

Through modeling ROR, it can be obtained information for the future projection of data series. This type of modeling opens an important and promising way for the series that behave as a white noise, providing new information for the series and its behavior. The model explains the 8.7 % of variance, with errors of 44.6 mm.

The tendency of errors is to increase around 0.004 mm, although it is not statistically significant, the errors depend on the errors one month behind.

Bibliography

1. Box, G. E. P. and Jenkins, G. M., and Reinsel, G. C., 1994: Time Series Analysis Forecasting and Control, 3rd edition. Prentice –Hall Inc., New Jersey.
2. Centro Meteorológico Provincial Villa Clara Departamento de Clima http://meteoweb.vcl.insmet.cu/meteoweb/
3. FIMIA R, OSÉS R, OTERO M, *ET AL*, (2012). La malaria. Modelación matemática en el análisis entomoepidemiológico. Evolución de la malaria, modelación matemática y boletines de Vigilancia Epidemiológica en Cuba. Editorial Académica Española, (ISBN 978-3-659-02519-8) for Marca registratada de Academic Publishing GmbH & Co. KGAcademia.
4. FIMIA, R., OSÉS, R., OTERO, M., DIEGUEZ, L., CEPERO, O., GONZALEZ, R., SILVEIRA, E., CORONA, E., (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/redvet Volumen 13 Nº 3 -

http://www.veterinaria.org/revistas/redvet/n03031.html.

5. McCracken, D; Dorn, S. (1971). Métodos numéricos y programación FORTRAN, Edición Revolucionaria. Instituto Cubano del Libro, 1971.
6. OSES R, FIMIA R, SAURA G, PEDRAZA A, RUIZ N, SOCARRAS J, (2014).Long Term Forecast of Meteorological Variables in Sancti Spiritus. CUBA..*Applied Ecology and Environmental Sciences*. Volume 2 Number 1 February 2014.ISSN (Print): 2328-3912, ISSN (Online): 2328-3920. 2014, 2(1), 37-42 DOI: 10.12691/aees-2-1-6.
7. OSES R., FIMIA R,, SILVEIRA, E ,HERNANDEZ W, SAURA G., PEDRAZA A,GONZALEZ R (2012). Modelación matemática hasta el año 2020 de la densidad larvaria anofelínica de mosquitos (Diptera Culicidae) en Caibarien Provincia de Villa Clara Cuba .REDVET Rev. electrón. vet. http://www.veterinaria.org/revistas/redvet Volumen 13 Nº 3 - http://www.veterinaria.org/revistas/redvet/n030312.html.
8. OSES, R. SAURA, G., PEDRAZA, A. a (2012). Modelación matemática ROR aplicada al pronóstico de terremotos de gran intensidad en Cuba,REDVET Rev. electrón. vet. http://www.veterinaria.org/revistas/redvet.Volumen13Nº05B – http://www.veterinaria.org/revistas/redvet/n050512B.html.
9. OSES, R. SAURA, G., PEDRAZA, A. b (2012). Modelación de la cantidad de frentes fríos en Cuba, impacto de manchas solares. (2012) REDVET Rev. electrón. vet. Volumen 13 Nº 05B - http://www.veterinaria.org/revistas/redvet/n050512
10. OSÉS, R., SAURA, G., OTERO, M.,PEDRAZA, A, SOCARRAZ, J. RUIZ, N. Cambio climático e impacto en la salud animal en la Provincia de Villa Clara.CUBA.(2012) Volumen 13 Nº 05B - http://www.veterinaria.org/revistas/redvet/n050512B.html ,
11. Oses, R.; 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.
12. Oses, R; Fimia, R (2012) Modelos de Pronóstico de Angiostrongilosis y Fasiolósis. Editorial Académica Española,ISBN:978-3-8484-5355-9,.

****