1° Congreso Argentino de Estadística

SELECTION OF A SUITABLE PROBABILITY MODEL FOR THE ANALYSIS OF BIOCHEMICAL DATA FROM SOIL

MARIA SOL ROSSI[1], RODOLFO MON[1], CARLOS IRURTIA[1]

[1]*De los Reseros y Nicolás Repetto S/N, Hurlingham, Instituto de Suelos / INTA*

rossi.maria@inta.gob.ar

ABSTRACT

Compacted soil in Argentine and other countries causes serious loss of agricultural production. In previous studies we used biological indicators of soil quality to describe the soil in different systems of agricultural management. In this study, we studied the microbial activity from soil based on enzyme activities measurements at agricultural soil located in the Northeast of the Province of Buenos Aires. This followed an experimental design consisting of three blocks completely randomized with three treatments: decompacted soil (DECOMP), compacted soil (COMP) and non-cultivated soil *(*CONTROL). The size of the experimental unit was 4.200 m. The number of variable continuous quantitative biochemical measures was 10. Soil samples were taken with 15-days frequent biochemical analyses during the period June 2010 to April 2013. Multivariate time series analysis was used to study the behavior of soil enzyme activities. A time series algorithm was used to study the biochemical profile of the soil subjected to three cycles of agriculture with the ts library in the R environment under Debian GNU/Linux system. The exact likelihood was computed via a state-space representation of the ARIMA process, and the innovations and their variance found by a Kalman filter. We demonstrate that the biochemical characterization of the soil is related to the state of its compaction using the ARIMA model. These data lead to the idea that the interrelation of biochemical variables from soil could be used as an accessor of the degree of current compaction of soil. Finally the effect of mechanical decompaction of soil on biochemical variables for 3 years under no-tillage as a management system was tested.

Key Words: *soil enzyme activities. multivariates time series analysis. soil decompaction.*

INTRODUCTION

Compacted soil in Argentine and other countries causes serious loss of agricultural production. In previous studies as Rossi et al (2008) and Rossi et al (2010) we used biological indicators of soil quality to describe the soil in different systems of agricultural management. The degree of soil compaction and the soil microbiome were studied by numerous authors such as Hartmann *et al.* (2014), Barik *et al.* (2014). In this paper we investigate the microbial activity of soil on the basis of enzymatic profiles over time in contrasting situations of soil state. For this we use time series analysis using a set of statistical techniques that allows describing and predicting the behavior of a time series and modeling the stochastic process from which they come in order to make predictions. The ARIMA model family (Autoregressive Integrated Moving Average model and extensions) according to Box and Jenkins (1970), Brockwell *et al.* (2002), Hyndman and Khandakar (2008) is widely used and presents good results for near prediction horizons in time of time series that present seasonal behaviors. This paper introduces the ARIMA family of models, completely analyzes a series and analyzes aspects of stationarity, seasonality and diagnosis to build the model that best fits it. The time series is analyzed with the statistical programming environment and language R (2008).

METHODS

In this study, we studied the microbial activity from soil based on enzyme activities measurements at agricultural soil located in the Northeast of the Province of Buenos Aires. This followed an experimental design consisting of three blocks completely randomized with three groups: decompacted soil (DECOMP), compacted soil (COMP) and non-cultivated soil *(*CONTROL). The size of the experimental unit was 4.200 m. The number of variable continuous quantitative biochemical measures was 10. Soil samples were taken with 15-days frequent biochemical analyses during the period June 2010 to April 2013 according to Alef and Nannipieri (1995). Multivariate time series analysis was used to study the behavior of soil enzyme activities according to Wold (1938) and Wiener (1949). A time series algorithm was used to study the biochemical profile of the soil subjected to three cycles of agriculture with the ts library in the R (2015) environment under Debian GNU/Linux system. The exact likelihood was computed via a state-space representation of the ARIMA process, and the innovations and their variance found by a Kalman filter. Determinations of infiltration with a simple ring, resistance to penetration with a shock penetrometer and bulk density from 0 to 10 cm and from 10 to 20 cm deep were made according to Pilatti and de Orellana (2000).

The R language commands used to analyze the time series stored in the class object are presented below.

```
####### 1. IDENTIFICATION
# TIME CHART
plot(tsdata, xlab = "", ylab = "PROCEDIMIENTOS INICIADOS", main = "GRÁFICO TEMPORAL")
# CHART ADF PACF
mx=12*2
par(mfrow=c(1,2))
tsdata %>% acf(lag.max=mx, xaxt="n", main = TeX("$y_t$"), xlab = "", ylab = "ACF")
axis(1, at=0:mx/12, labels=0:mx)
tsdata %>% pacf(lag.max=mx, xaxt="n", main = "", xlab = "", ylab = "PACF")
axis(1, at=0:mx/12, labels=0:mx)
# CLASSICAL DECOMPOSITION
tsdata %>% decompose(type = "multiplicative") %>% plot()
# SEASONAL CHART
seasonplot(tsdata, col = rainbow(7), year.labels=FALSE, year.labels.left=TRUE,
ylab = "PROCEDURES", xlab = "MES", main = "SEASONAL TEMPORARY GRAPH")
# ADF TEST
adfTest(tsdata) # RESULT: NO STATIONARY
# SUGGESTED NUMBER OF DIFFERENCES (SIMPLE OR SEASONAL, RESPECTIVELY)
ndiffs(tsdata) # RESULT d = 1
nsdiffs(tsdata) # RESULT D = 1
#BOX-COX TRANSFORMATION
lambda <- BoxCox.lambda(tsdata) #SUGGEST lambda = 0.318
tsdata.transformed <- tsdata %>%
BoxCox(lambda)
# STATIONARY SERIES
stationary <- tsdata.transformed %>% diff(differences = 1) %>% diff(differences = 1,
lag = 12)
stationary %>% plot(xlab ="", ylab = TeX("$\\nabla_{12}(\\nabla y_t^{(\\lambda)})$"),
main = "TEMPORARY GRAPH")
par(mfrow=c(1,2))
stationary %>% acf(lag.max=mx, xaxt="n", xlab = "", ylab = "ACF",
main = TeX(" $\\nabla_{12}(\\nabla y_t^{(\\lambda)})$"))
axis(1, at=0:mx/12, labels=0:mx)
stationary %>% pacf(lag.max=mx, xaxt="n", main = "", xlab = "", ylab = "PACF")
axis(1, at=0:mx/12, labels=0:mx)
####### 2. AJUST
# AFC AND PACF PLOT SUGGEST ARIMA(2,1,2)(0,1,1)[12]
t1 <- tsdata.transformed %>% Arima(order = c(2,1,2), seasonal = list(order = c(0,1,1),
period=12))
####### 3. DIAGNOSIS
# RESIDUALS ARE STATIONARY
plot(t1$residuals, xlab = "", ylab = "", main = "RESIDUALS")
mx = 12
par(mfrow = c(2,1))
plot(t1$residuals, xlab = "", ylab = "RESIDUALS", main = "RESIDUALS")
```

```
t1$residuals %>% acf(lag.max=mx, xaxt="n", xlab = "", ylab = "ADF", main = "")
axis(1, at=0:mx/12, labels=0:mx)
#  RESIDUALS ARE INDEPENDENTS
t1$residuals %>% Box.test(type = "Ljung-Box")
tsdiag(t1)
#RESIDUES FOLLOW A NORMAL DISTRIBUTION
par(mfrow=c(1,2))
qqnorm(t1$residuals, main = "GRÁFICO QQ", xlab = "THEORETICAL QUANTILES", ylab =
"SAMPLE QUANTILES")
qqline(t1$residuals, col = "red")
hist(t1$residuals, main = "HISTOGRAM OF RESIDUALS", xlab = "RESIDUALS", ylab =
"FRÉQUENCE")
t1$residuals %>% jarque.bera.test() #RESULT: p-valor 0.9348
####### 4. PREDICTION
horizonte <- 12*2
pred <- t1 %>% forecast(h = horizonte, level = c(95))
# UNDO TRANSFORM
pred$mean <- pred$mean %>% InvBoxCox(lambda)
pred$lower <- pred$lower %>% InvBoxCox(lambda)
pred$upper <- pred$upper %>% InvBoxCox(lambda)
pred$x <- pred$x %>% InvBoxCox(lambda)
pred$fitted <- pred$fitted %>% InvBoxCox(lambda)
pred$residuals <- pred$residuals %>% InvBoxCox(lambda)
#MEASURES OF PREDICTION
accuracy(pred)
# PREDICTION
pred %>% plot(shaded = FALSE, xlab = "AÑOS", ylab = "PROCEDURES",
main = TeX("$SARIMA(2,1,2)(0,1,1)_{12}$"))
lines(pred$fitted, col = "red")
legend("topleft", legend=c("SERIE", "PREDICTION", "CONFIDENCE INTERVAL AT 95%",
"RESULTS"), col=c("black", "blue", "black", "red"), lty=c(1,1,2,1), lwd = 2,
cex = 0.6)
### AUTOMATIC MODELING
t1 <- tsdata %>% auto.arima()
horizont <- 12*2
pred <- t1 %>% forecast(h = horizont, level = c(95))
```

RESULTS

As we know the stochastic process $\{Y_t\} = \{Y_t \mid t \in Z\}$ is stationary if its statistical properties do not depend on the period of time in which it is observed.
In the first place, a preliminary analysis was carried out where the classic decomposition of the time series was carried out in the form

$$Y_t = S_t \times T_t \times R_t$$

It was observed that the mean and the variance were not constant, therefore, the Box-Cox transformation was applied to stabilize the variance and the differentiations were made to convert the constant mean.

Then a statistical model was developed that describes the stochastic process. Figures 1, 2, 3 and 4 show that the time series has a seasonal behavior that appears as two peaks in July and September. This suggests that a seasonal ARIMA model would fit the data well.

SARIMA(2, 1, 1) × (0, 1, 1)12

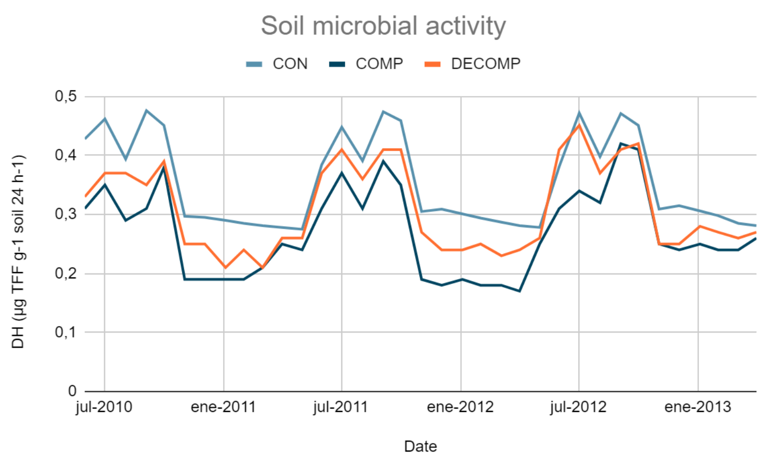

Figure 1: Time series of soil dehydrogenase activity every 15 days of each year

In the first instance, we can observe that the time series are not stationary since their respective means and variances are not constant. This interpretation from the graph was corroborated in the calculations, therefore the time series had to be transformed.

The microbial activity of the soil measured on the basis of the dehydrogenase activity was higher in the undisturbed soil (CON) with respect to the activity in the soil subjected to mechanical loosening (DECOMP) and with respect to the soil without mechanical loosening treatment (COMP) $p < 0.05$. A trend of increased microbial activity over time was also observed in the mechanical decompaction treatment (DECOMP) with respect to the undisturbed soil (CON).

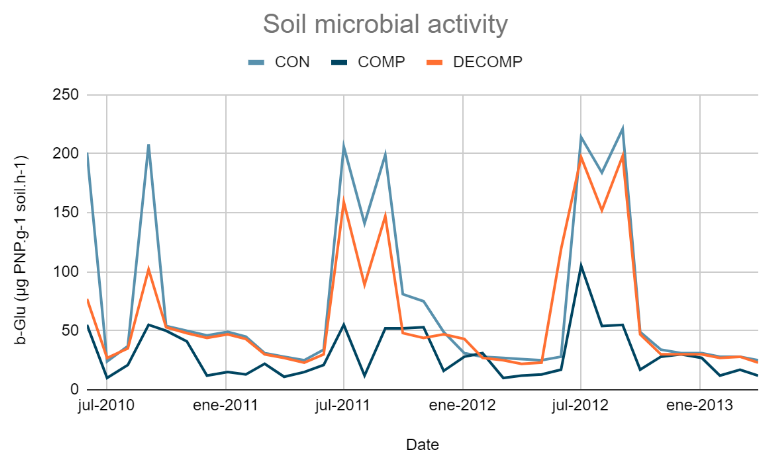

Figure 2: Time series of soil ß glucosidase activity every 15 days of each year

In these measurements we can see that the time series are not stationary since their respective means and variances are not constant. This interpretation from the graph was corroborated in the calculations, therefore the time series had to be transformed.

The microbial activity of the soil measured on the basis of beta glucosidase activity was higher in the undisturbed soil (CON) with respect to the activity in the soil subjected to mechanical loosening (DECOMP) and with respect to the soil without mechanical loosening treatment (COMP) $p < 0.05$. It was also observed that the microbial activity increased over time in the mechanical decompaction treatment (DECOMP) with respect to the undisturbed soil (CON) $p < 0.05$. On the other hand, the microbial activity of the compacted soil (COMP) and of the undisturbed soil (CON) did not present variations except for those explained by seasonality during the entire period studied.

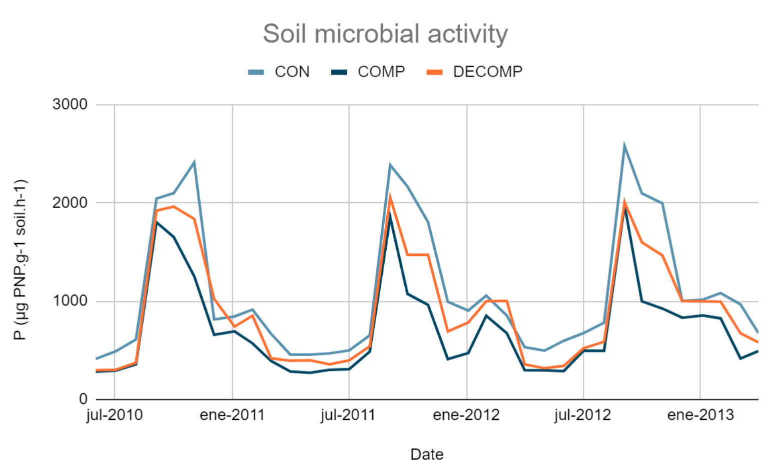

Figure 3: Time series of soil phosphatase activity every 15 days of each year

In the first instance, we can observe that the time series are not stationary since their respective means and variances are not constant. This interpretation from the graph was corroborated in the calculations, therefore the time series had to be transformed.

The microbial activity of the soil measured on the basis of phosphatase activity was higher in the undisturbed soil (CON) presented a statistically significant difference $p<0.05$ with respect to the other groups when measuring the activity in the soil subjected to mechanical decompaction (DECOMP) and with respect to the soil without mechanical decompaction treatment (COMP). However, it was observed that the microbial activity of the compacted soil (COMP), of the soil under mechanical decompaction (DECOMP) and of the undisturbed soil (CON) did not present variations except for those explained by seasonality during the entire period studied. Therefore, the microbial activity based on the activity of the phosphatase enzyme did not result in an indicator of change compared to the state of soil compaction under the conditions studied.

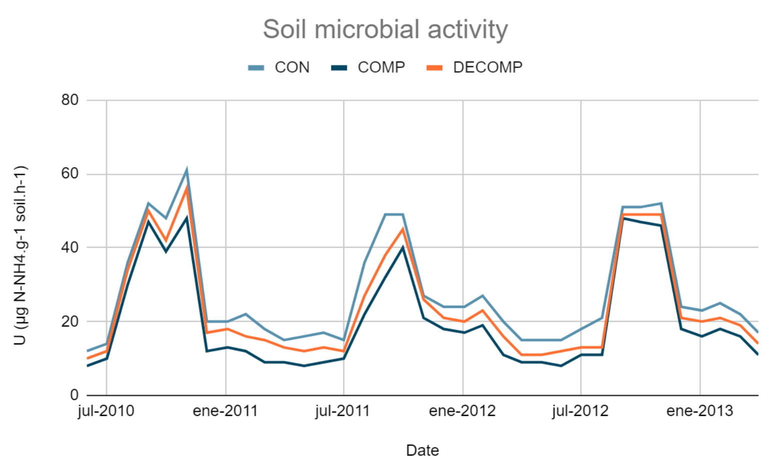

Figure 4: Time series of soil urease activity every 15 days of each year

In the first instance, we can observe that the time series are not stationary since their respective means and variances are not constant. This interpretation from the graph was corroborated in the calculations, therefore the time series had to be transformed.

The microbial activity of the soil measured on the basis of urease activity presented a higher trend in the undisturbed soil (CON) with respect to the activity in the soil subjected to mechanical loosening (DECOMP) and with respect to the soil without mechanical loosening treatment. (COMP). However, it was observed that the microbial activity of the compacted soil (COMP), of the soil under mechanical decompaction (DECOMP) and of the undisturbed soil (CON) did not present variations except for those explained by seasonality during the entire period studied. Therefore, the microbial activity based on the activity of the phosphatase enzyme did not result in an indicator of change compared to the state of soil compaction under the conditions studied.

The stationary transformed time series differentiated from the original series showed a stationary behavior in the preliminary analysis of the soil biochemical data, so the ARIMA model was applied. The seasonal ARIMA model with the lowest Akaike information criterion using R.The predictions and their realized 95% confidence interval. The forecasts have the increasing trend and seasonal behavior that one might expect when examining older data. We demonstrate that the biochemical characterization of the soil is related to the state of its compaction using the ARIMA model.

The lot under study is located in a cartographic unit that presents three series of soil, Delgado series 40%; Santa Isabel Series 40% and Teodelina Series 20%. The infiltration curves show very low infiltration rates with rapid value declines. Ten minutes after the start of the measurement, the values drop to very low values close to the base value.

The critical apparent density data was calculated according to Pillatti and de Orellana (2000). In the layer from 0 to 10 cm, the average value corresponding to the three series is 1.382 and for the layer from 10 to 20 cm it is 1.354, which indicates that up to 20 cm the soil of the lot under study presents Da values above the critical values. Considering the critical value 100%, only one value is below 90%, a value from which radical growth is considered critical.

On the other hand, if the maximum apparent density data is taken as that cited by Ferreras *et al.* (2007), obtained by means of the proctor test, the maximum reference value is somewhat higher than that estimated by Pillatti and de Orellana (2000). Therefore, the relative bulk density values decrease. The quotient between the bulk density measured in the field and that measured by the Proctor test establishes the Relative Compaction (RC), a value that allows comparison between different Da values.

If the CR values are very high, close to 90%, it means that the soil is very close to the maximum compaction it can admit, with a severe decrease in the largest pores, affecting crop growth (Ferreras *et al.* 2007).

We work with the family of ARIMA models for the analysis of time series based on the concept of stochastic process and understanding a time series as a realization of it. Next, the fundamental concepts for the analysis of time series were worked on, the family of ARIMA models was built and the data set was analyzed.

CONCLUSION

The model of the ARIMA family that was adjusted to the series following the established criteria was chosen. Once the model was estimated, the requirements were verified in the diagnostic stage and the desired results were obtained. Also, the predictions provided by the model were found together with a graph of the same in which it can be seen that the series continues with a growing trend and with its characteristic seasonal behavior.

Finally, the modeling of the series was used in the R language library that follows an algorithm that aims to achieve the model that contributes the least AIC.

These data lead to the idea that the interrelation of biochemical variables from soil could be used as an accessor of the degree of current compaction of soil. Finally the effect of mechanical decompaction of soil on biochemical variables for 3 years under no-tillage as a management system was tested. The soil presents values of apparent density, resistance to penetration and infiltration, characteristic of compacted soils. It is estimated that the productive potential of this soil is compromised by physical compaction up to 30 cm deep. A mechanical decompaction up to 25 - 30 cm deep will improve the physical conditions of the soil in the layer of the profile with the highest root activity. In this way, infiltration will be improved and the soil will be explored by the roots more easily. These changes in soil properties are expected to be reflected in increased yields, although this is not always the case as yield is shaped by many other variables.

This work was support by PE AEAI-271171. Diagnóstico de la compactación del suelo en siembra directa y técnicas para la descompactación y su control. INTA-Instituto de Suelos.

REFERENCES

Alef, K. y P. Nannipieri. 1995. Methods in applied soil microbiology and biochemistry. Academic Press Harcourt Brace and Company Publishers Great Britain. https://www.elsevier.com/books/methods-in-applied-soil-microbiology-and-biochemistry/alef/978-0-12-513840-6

Barik, K., Aksakal, E. L., Islam, K. R., Sari, S., & Angin, I. (2014). Spatial variability in soil compaction properties associated with field traffic operations. Catena, 120, 122-133. https://doi.org/10.1016/j.catena.2014.04.013

Brockwell, P. J., Davis, R. J., Rose, C., Richard A. Davis, P. J. B., Smith, M. D., Calder, M. V., & Springer (Firm). (2002). Introduction to Time Series and Forecasting. https://www.academia.edu/42933730/Introduction_to_Time_Series_and_Forecasting_Third_Edition

Box, G. y Jenkins, G. (1970) Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco. https://www.scirp.org/(S(i43dyn45teexjx455qlt3d2q))/reference/referencespapers.aspx?referenceid=2087370

Ferreras, L., Magra, G., Besson, P., Kovalevski, E. y F. García. (2007). Indicadores de calidad física en suelos de la región pampeana norte de Argentina bajo siembra directa. Revista de la asociación Argentina de la Ciencia del Suelo 25(2): 159-172. http://www.scielo.org.ar/scielo.php?pid=S1850-20672007000200007&script=sci_arttext

Hartmann, M., Niklaus, P. A., Zimmermann, S., Schmutz, S., Kremer, J., Abarenkov, K., & Frey, B. (2014). Resistance and resilience of the forest soil microbiome to logging-associated compaction. The ISME journal, 8(1), 226-244. https://www.nature.com/articles/ismej2013141

Hyndman, R. J., Khandakar, Y. (2008). Automatic Time Series Forecasting: The forecast Package for R. Journal of Statistical Software, 27(3), 1–22. https://doi.org/10.18637/jss.v027.i03

Pilatti M.A. y J. A. de Orellana (2000).The Ideal soil. Critical values of an a Ideal Soil for Mollisols in the North of the Pampean Region, Argentina.J. Sustainable Agric.17:89 - 111.

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org

Rossi, M.S., Casas, R.R., Michelena, R.O., García, M. & Pérez, B.A. (2008). Indicadores bioquímicos y microbiológicos para la descripción del estado de salud del suelo. En Libro de Resúmenes del XXI Congreso Argentino de la Ciencia del Suelo. 1 Ed – San Luis: Asociación Argentina de la Ciencia del Suelo, 400 pp. AACS - Asociación Argentina Ciencia del Suelo. Enlace al Trabajo expandido

Rossi M.S., Michelena R. O., Casas, R. R. 2010. Looking for Biological Indicators of Soil using Hierarchical Clustering. XXII Congreso Argentino de la Ciencia del Suelo. Rosario, Santa Fe. AACS - Asociación Argentina Ciencia del Suelo Enlace al Trabajo expandido

Wiener, N. (1949). Extrapolation, Interpolation, and Smoothing of Stationary Time Series. With Engineering Applications. The Technology Press of the Massachusetts Institute of Technology, Cambridge.
https://mitpress.mit.edu/9780262730051/extrapolation-interpolation-and-smoothing-of-stationary-time-series/

Wold, H. O. A. (1938). A Study in the Analysis of Stationary Time Series. Almqvist and Wiksell.
https://www.scirp.org/(S(351jmbntvnsjt1aadkozje))/reference/referencespapers.aspx?referenceid=1936239