



**MOMENTO OPTIMO PARA LA COSECHA DE LA MICROALGA *Ankistrodesmus gracilis*  
EN RELACIÓN CON SU CANTIDAD Y CALIDAD**

**MARCO ANTONIO IMUES FIGUEROA**

**UNIVERSIDAD NACIONAL DE COLOMBIA  
FACULTAD DE CIENCIAS  
ESPECIALIZACIÓN EN ESTADÍSTICA  
PASTO, COLOMBIA  
2008**

**MOMENTO OPTIMO PARA LA COSECHA DE LA MICROALGA *Ankistrodesmus gracilis*  
EN RELACIÓN CON SU CANTIDAD Y CALIDAD**

**MARCO ANTONIO IMUES FIGUEROA**

**Trabajo de Grado presentado como requisito parcial para optar al título de  
Especialista en Estadística**

**Asesora  
MARTHA PATRICIA BOHORQUEZ CASTAÑEDA**

**UNIVERSIDAD NACIONAL DE COLOMBIA  
FACULTAD DE CIENCIAS  
ESPECIALIZACIÓN EN ESTADÍSTICA  
PASTO, COLOMBIA  
2008**

## CONTENIDO

	<b>Pág.</b>
RESUMEN	6
SUMMARY	7
INTRODUCCIÓN	8
1. METODOLOGÍA	10
2. MARCO TEÓRICO	11
2.1 AVANCES EN EL CULTIVO DE ALGAS	11
2.2 MODELOS DE ECUACIONES SIMULTANEAS	13
2.2.1 Autocorrelación	13
2.2.2 Identificación de ecuaciones	17
2.2.3 Estimación por el método de mínimos cuadrado en dos etapas	18
3. PRESENTACIÓN A ANÁLISIS DE LOS RESULTADOS	20
3.1 ANÁLISIS EXPLORATORIO	20
3.2 ANÁLISIS DE CORRELACIÓN	24
3.3 ANÁLISIS PRELIMINAR DE REGRESIÓN LINEAL MÚLTIPLE	25
3.3.1 Modelo para la variable Proteína	25
3.3.2 Modelo para la variable Energía	26
3.3.3 Modelo para la variable Extracto etéreo (EE)	27
3.4 ANÁLISIS DE AUTOCORRELACIÓN	28
3.5 IDENTIFICACIÓN DE ECUACIONES Y VARIABLES INSTRUMENTALES	31
3.6 MÉTODO DE MÍNIMOS CUADRADO EN DOS ETAPAS (MC2E)	33
3.7 DETERMINACIÓN DEL MOMENTO ÓPTIMO DE COSECHA	39
4. CONCLUSIONES	40
REFERENCIAS BIBLIOGRÁFICAS	41

## LISTA DE TABLAS

	<b>Pág.</b>
Tabla 1. Estadísticas descriptivas para las variables en estudio, en período de 15 días.	20
Tabla 2. Estadísticas descriptivas para las variables en estudio, en período de 13 días.	22
Tabla 3. Correlación múltiple entre las diferentes variables.	24
Tabla 4. Matriz de correlaciones parciales.	25
Tabla 5. Análisis de regresión lineal preliminar de la primera ecuación ( $Y_{1t} = \text{Proteína}$ ).	26
Tabla 6. Análisis de regresión lineal preliminar de la segunda ecuación ( $Y_{2t} = \text{Energía}$ ).	27
Tabla 7. Análisis de regresión lineal preliminar de la tercera ecuación ( $Y_{3t} = \text{EE}$ ).	28
Tabla 8. Prueba de autocorrelación mediante el estadístico $d$ de Durbin-Watson.	29
Tabla 9. Matriz de coeficientes para la identificación de ecuaciones mediante la condición de rango.	32
Tabla 10. Análisis de regresión para la estimación de variables endógenas.	34
Tabla 11. Análisis de regresión para la estimación de la primera ecuación del sistema.	35
Tabla 12. Análisis de regresión para la estimación de la segunda ecuación del sistema.	37
Tabla 13. Análisis de regresión para la estimación de la tercera ecuación del sistema.	38
Tabla 14. Valores predichos para el punto óptimo $Y_{it}$ particular.	39

## LISTA DE FIGURAS

		<b>Pág.</b>
Figura 1.	Diagrama de dispersión para las variables incluidas en el estudio, con datos en un período de 15 días.	21
Figura 2.	Diagrama de dispersión para las variables incluidas en el estudio, con datos en un período de 13 días.	23
Figura 3.	Diagramas de la función de autocorrelación simple y parcial de las variables endógenas.	30

## RESUMEN

El presente trabajo se desarrolló con el fin de determinar el momento óptimo para la cosecha de la microalga *Ankistrodesmus gracilis* cultivada en laboratorio, de manera que se maximice el crecimiento poblacional y la disponibilidad de nutrientes y se minimice el contenido de fibra, utilizando métodos de ecuaciones simultáneas. Para esto se aprovechó la base de datos suministrada por Coral (2006), de la que fue tomada la información de Contenido de proteína ( $Y_{1t}$ ), Concentración de energía ( $Y_{2t}$ ) y Contenido de extracto etéreo ( $Y_{3t}$ ) como variables endógenas y Densidad poblacional ( $X_{1t}$ ), Contenido de clorofila a ( $X_{2t}$ ) y Contenido de fibra ( $X_{3t}$ ) como variables exógenas, además de los días de cultivo ( $T$ ), en un período de 13 días. Se corrieron modelos de regresión de cada una de las variables endógenas sobre todas las demás variables, con el propósito de tener una guía para la formulación del sistema preliminar de ecuaciones simultáneas, constituido por tres ecuaciones de tipo endógeno. Dado la existencia visible de una serie de tiempo, se aplicaron pruebas de Durbin-Watson para detectar autocorrelación serial de las variables endógenas con sus respectivos rezagos suponiendo  $H_1 = \rho > 0$ , encontrando que la variable Proteína corresponde a un proceso AR(1), la variable Energía a un proceso AR(2) y el Extracto etéreo a un proceso AR(1), corroborado por las funciones de autocorrelación simple y parcial; en consecuencia, se incluyeron en los modelos las respectivas variables rezagadas ( $Y_{1t-1}$ ,  $Y_{2t-2}$ ,  $Y_{3t-1}$ ). Mediante la verificación de la condición de orden y la condición de rango, se pudo establecer que el sistema está sobreidentificado, por tal motivo, es posible aplicar el método de mínimos cuadrado en dos etapas (MC2E), considerando las variables rezagadas como variables instrumentales (VI) para los modelos de ecuaciones reducidas. En la primera etapa de MC2E se obtuvieron estimadores de los parámetros estructurales mediante el método de mínimos cuadrados ordinarios (MCO), con base en las ecuaciones en la forma reducida y mediante éstos obtener los correspondientes valores ajustados para las variables endógenas ( $\hat{Y}_{it}^*$ ). En la segunda etapa de MC2E se reemplazaron las variables endógenas del segundo miembro por  $\hat{Y}_{it}^*$  y se estimaron las ecuaciones mediante MCO, con lo cual se obtienen estimaciones insesgadas y consistentes. Los análisis anteriores permiten consolidar el sistema de ecuaciones simultáneas, de la siguiente manera:

$$\begin{aligned}\hat{Y}_{1t} &= -0,186801 + 0,00188877X_{1t} + 0,0000167509X_{2t} - 0,015422X_{3t} + 0,769953Y_{1t-1} - 0,00000175684\hat{Y}_{2t}^* - 0,000775469T \\ \hat{Y}_{2t} &= -23656,8 + 104,531X_{1t} + 1,72874X_{2t} + 268,214X_{3t} - 0,356208Y_{2t-2} - 4370,53\hat{Y}_{1t}^* - 17481\hat{Y}_{3t}^* + 119,746T \\ \hat{Y}_{3t} &= -0,960238 + 0,00705133X_{1t} + 0,0000887397X_{2t} + 0,370497Y_{3t-1} - 0,193376\hat{Y}_{1t}^* - 0,0000118743\hat{Y}_{2t}^* - 0,0227127T\end{aligned}$$

Con base en los modelos estimados previamente, para el sistema de ecuaciones múltiples, se realizó la predicción para cada uno de los puntos  $Y_{it}$  particular, a partir de los cuales fue posible determinar el momento óptimo de cosecha de la microalga a los 11 días de cultivo.

**Palabras clave:** microalga, *Ankistrodesmus gracilis*, ecuaciones simultaneas, mínimos cuadrado ordinarios, mínimos cuadrado en dos etapas.

## SUMMARY

This work was developed in order to determine the optimal time to harvest the microalgae *Ankistrodesmus gracilis* grown in the laboratory, so as to maximize the population growth and nutrient availability and minimize the fiber content, using simultaneous equations. For this was used the database provided by Coral (2006), which was taken from the information content of protein ( $Y_{1t}$ ) concentration of gross energy ( $Y_{2t}$ ) and ether extract content ( $Y_{3t}$ ) as endogenous variables and population density ( $X_{1t}$ ), chlorophyll *a* content ( $X_{2t}$ ) and fiber content ( $X_{3t}$ ) as exogenous variables, in addition to the days of cultivation ( $T$ ), in a period of 13 days. It was rushed regression models for each of the endogenous variables on all other variables, with the goal of having a guide for her preliminary design of the system of simultaneous equations, consisting of three equations type endogenous type. Given the visible presence of a series of time were applied Durbin-Watson test to detect serial autocorrelation of the endogenous variables with their respective setbacks assuming  $H_1 = \rho > 0$ , finding that the protein variable is a process AR(1), a gross energy variable Energy is a process AR(2) and the ethereal extract a process AR(1), supported by the functions of simple and partial autocorrelation and consequently, were included in the models behind the respective variables ( $Y_{1t-1}$ ,  $Y_{2t-2}$ ,  $Y_{3t-1}$ ). By verifying the order condition and the rank condition, it could be established that the system is sobre-identified, therefore, it is possible to apply the method of minimal square in two stages (MS2S), considering the lagging variables as instrumental variables (IV) for the model of reduced equations. In the first phase of MS2S were obtained estimators of the structural parameters using the method of ordinary least squares (OLS), based on the equations in the short form and get them through the appropriate settings for the endogenous variables ( $\hat{Y}_{it}^*$ ). In the second stage of MS2S were replaced endogenous variables by  $\hat{Y}_{it}^*$  on the second member and was estimated by the equations through OLS, thus obtained are unbiased and consistent. The previous analysis allows to consolidate the system of simultaneous equations, as follows:

$$\begin{aligned}\hat{Y}_{1t} &= -0,186801 + 0,00188877X_{1t} + 0,0000167509X_{2t} - 0,015422X_{3t} + 0,769953Y_{1t-1} - 0,00000175684\hat{Y}_{2t}^* - 0,000775469T \\ \hat{Y}_{2t} &= -23656,8 + 104,531X_{1t} + 1,72874X_{2t} + 268,214X_{3t} - 0,356208Y_{2t-2} - 4370,53\hat{Y}_{1t}^* - 17481\hat{Y}_{3t}^* + 119,746T \\ \hat{Y}_{3t} &= -0,960238 + 0,00705133X_{1t} + 0,0000887397X_{2t} + 0,370497Y_{3t-1} - 0,193376\hat{Y}_{1t}^* - 0,0000118743\hat{Y}_{2t}^* - 0,0227127T\end{aligned}$$

Based on the models previously estimated, for the system of multiple equations, the prediction was made for each point  $Y_{it}$  particular, from which it was possible to determine the optimal time to harvest at 11 days of microalgae culture.

**Key words:** microalga, *Ankistrodesmus gracilis*, simultaneous equations, ordinary least squares, minimum square in two stages.

## INTRODUCCIÓN

El incremento de la producción acuícola ha llevado a la necesidad de cultivar nuevas especies con mayor potencial de crecimiento, lo cual implica la masiva producción de alimento vivo, especialmente de microalgas, de vital importancia para el cultivo de moluscos, crustáceos, peces y diversas especies de zooplancton.

El alimento preferido por las especies de peces en su estado larvario, está constituido por alimento vivo (plancton), tanto fito (microalgas) como zooplancton, siendo el primero de gran importancia por cuanto es la base de la cadena trófica, bien sea como alimento directo de las larvas o como alimento de muchas especies de zooplancton, que más tarde servirán de alimento de otras especies ícticas. Por lo tanto, el desarrollo y supervivencia de las larvas depende de la presencia de organismos que conforman el fitoplancton y el zooplancton, debido a su composición nutricional así como por su tamaño y forma que permiten su fácil captura e ingestión.

El cultivo de microalgas en gran escala, data del año 1949 en Israel, Japón y Estados Unidos y otros países en todo el mundo, utilizando y estudiando diferentes especies de algas. Estos métodos, implementados en laboratorios controlados, hoy en día se desarrollan eficientemente, con tecnologías de punta y equipos avanzados para el control del ambiente.

Algunos autores, como Sipaúba y Rocha<sup>1</sup>, encontraron que la microalga *Ankistrodesmus gracilis* produce buenos resultados como alimento para zooplancton y larvas de peces, puesto que tiene un crecimiento exponencial dentro de un período de 8 a 15 días de cultivo y por su alto valor nutritivo.

Las técnicas actuales recomiendan el momento más adecuado para cosechar las algas, al finalizar la fase exponencial, dado que en ese punto se tiene la máxima concentración celular, lo que indica que se tiene mayor cantidad de alimento para los organismos que se desea alimentar. En la especie de microalga en mención, ésta comúnmente se realiza entre los 12 y los 13 días de cultivo.

Sin embargo, la importancia de este alimento está basada en su valor nutritivo, esto es, el contenido de algunos componentes químicos como proteína, energía bruta y extracto etéreo, los cuales deberían tenerse en cuenta para determinar el momento del aprovechamiento del cultivo algal, dada su disponibilidad en el material vegetal.

Según Fábregas<sup>2</sup>, el valor nutricional del alga varía en función de las condiciones de cultivo, de la composición del medio, la fase de crecimiento, el tamaño de la célula y la digestibilidad. Por su parte Sipaúba<sup>3</sup> sostiene que los niveles de proteína pueden ser mejorados al cosecharse en la fase

---

<sup>1</sup> SIPAUBA, L. H. e ROCHA, O. 1993. Cultivo em larga escala de organismos planctônicos para alimentação de larvas e alevinos de peixes: I-algas clorofíceas. Em: Biotemas. Jaboticabal, SP, Brasil: FCAVJ-UNESP, Vol. 6, No. 1.

<sup>2</sup> FABREGAS, J. 1986. Marine microalgae as a potential source of minerals in fish diets. In: Ecol. Brasil, Vol. 7, No. 51.

<sup>3</sup> SIPAUBA, L. H. 1994. Limnología aplicada à Aqüicultura. Jaboticabal, SP, Brasil: FUNEP. Boletim Técnico do CAUNESP No. 1.

exponencial, puesto que es donde hay mayor aprovechamiento de nutrientes y mayor proceso de fotosíntesis.

De los anteriores planteamientos se deduce el máximo volumen de población no es suficiente criterio para definir el momento óptimo de cosecha, ya que el valor nutritivo (considerando todos los nutrientes, excepto la fibra) puede que no sea el indicado dado el incremento de la lignificación natural de las paredes celulares, por lo que se debe buscar que esta variables se encuentren en un equilibrio razonable, esto es, máxima población, máxima disponibilidad de nutrientes y mínimo contenido de fibra. Desde este punto de vista, es necesario optimizar el momento más adecuado de cosecha mediante un sistema de ecuaciones simultáneas, donde se tiene un grupo de variables endógenas (nutrientes) y otro grupo de variables exógenas.

El presente trabajo pretende determinar el momento óptimo para la cosecha de la microalga *Ankistrodesmus gracilis* cultivada en laboratorio, de manera que se maximice el crecimiento poblacional y la disponibilidad de nutrientes y se minimice el contenido de fibra, utilizando métodos de ecuaciones simultáneas.

## 1. METODOLOGÍA

Se utilizó la base de datos suministrada por Coral<sup>4</sup>, que contempla información de las variables: Densidad poblacional de las algas, Concentración de Proteína, Contenido de Energía bruta, Concentración de Extracto etéreo (EE), Contenido de Clorofila *a*, Contenido de Fibra, Días de cultivo.

La información fue recolectada durante todo el ciclo biológico, en un período de 15 días, tomando tres muestras diarias en dos recipientes de cultivo de 2,5 litros, en cuyas muestras se midieron las variables antes mencionadas.

Teniendo en cuenta que todas las variables, a excepción de la fibra, crecen hasta el día 13 y posteriormente decrecen, lo cual distorsiona la relación entre variables, haciendo que la correlación no sea significativa en la mayoría de los casos, se tomaron los datos únicamente hasta finalizar la fase exponencial, cuando se tiene el máximo crecimiento poblacional.

Utilizando la metodología propuesta por Juan, Kizys y Manzanedo<sup>5</sup> y Gujarati<sup>6</sup>, para la construcción de modelos de ecuaciones simultáneas, se formularon tres ecuaciones, cada una de las cuales representa un modelo de regresión para las variables endógenas sobre las exógenas:

- **Variables endógenas.** Corresponden a las variables que describen las características propias de las algas, que tienen que ver con la concentración de nutrientes presentes.

$Y_{1t}$  = Contenido de proteína.

$Y_{2t}$  = Concentración de energía.

$Y_{3t}$  = Contenido de extracto etéreo.

- **Variables exógenas.** Aquí se han incluido las variables independientes, que tienen influencia sobre la concentración de nutrientes en las algas.

$X_{1t}$  = Densidad poblacional.

$X_{2t}$  = Contenido de clorofila *a*.

$X_{3t}$  = Contenido de fibra.

$T$  = Días de cultivo

---

<sup>4</sup> CORAL, L. C. 2006. Estudio del crecimiento y desarrollo de la microalga *Ankistrodesmus gracilis* (Bold & Wynne, 1985) (chlorophyta) con diferentes medios de cultivo en laboratorio. Trabajo de grado (Ingeniero en Producción Acuícola). Pasto, Colombia: Universidad de Nariño, Facultad de Ciencias Pecuarias, Departamento de Recursos Hidrobiológicos. 117 p.

<sup>5</sup> JUAN, A; KIZYS, R. y MANZANEDO, L. M. 2004. Modelos multiecuacionales: Modelos de ecuaciones simultáneas. Cataluña: Universidad Oberta de Cataluña, Proyecto e-Math. Disponible en Internet, URL: <http://www.uoc.edu/in3/emath/docs/MMultiecuacionales.pdf>.

<sup>6</sup> GUJARATI, D. N. 2006. Principios de econometría. Madrid: McGraw-Hill.

## 2. MARCO TEÓRICO

### 2.1 AVANCES EN EL CULTIVO DE ALGAS

Según Bold y Wynne<sup>7</sup>, *Ankistrodesmus gracilis* es un alga unicelular perteneciente a la clase Chlorophyta, orden Chlorococcales, familia Oocystaceae, género *Ankistrodesmus* y especie *gracilis*.

Para Asselbom y Parodi<sup>8</sup>, esta especie presenta células de 12,5 a 17,5 µm de largo y de 2,5 a 5,0 µm de ancho, con forma de media luna y con los extremos ligeramente aguzados. El cloroplasto es parietal y presenta tilacoides paralelamente dispuestos, separados a veces por gránulos de almidón o interrumpidos por la presencia de un pirenoide, penetrado sólo en su región periférica por los tilacoides. El núcleo ocupa una posición central en la célula y posee una pronunciada escotadura donde se aloja el dictiosoma. El citoplasma se caracteriza por su homogeneidad y por la presencia de abundantes ribosomas libres.

Por su parte, Sipaúba y Rocha<sup>9</sup> hallaron que la microalga *Ankistrodesmus gracilis* ha dado buenos resultados como alimento para zooplancton y larvas de peces, dado su crecimiento acelerado y su alto valor nutritivo, además, encontraron que en la alimentación de zooplancton aumentaba la tasa de reproducción, en comparación con otras 16 especies de algas.

En cuanto al cultivo en masa, Lesmos<sup>10</sup> afirma que esta microalga tiene buenos resultados con volúmenes menores a 250 litros, produciendo grandes concentraciones celulares en corto período de tiempo.

Tal como afirma Alvarez<sup>11</sup>, las plantas utilizan la energía lumínica, a través de la fotosíntesis, para transformarla en energía química. La fotosíntesis de las algas tiene básicamente el mismo proceso que ocurre en las plantas superiores, en donde la luz es absorbida por varios pigmentos fotosintéticos, especialmente la clorofila *a*, que puede absorber hasta 695 nm de longitud de onda. La fijación fotosintética de la clorofila *a* es responsable de la generación primaria de los compuestos orgánicos: carbohidratos, grasas, proteínas, que conjuntamente con la cantidad de carbono total o energía fijada, forman la producción primaria bruta.

Las algas unicelulares se reproducen por simple división celular en bipartición repetida, para formar nuevos individuos de iguales características a la progenitora, proceso también conocido como fisión binaria, para incrementar el crecimiento de la población<sup>12</sup>.

---

<sup>7</sup> BOLD, H. C. y WYNNE, M. J. 1985. Introduction to the algae: structure and reproduction. New Jersey: Prentice Hall. 586 p.

<sup>8</sup> ASSELBOM, V. M. y PARODI, E. R. 2003. Ficología. En: XXIX Jornadas Argentinas de Botánica & Reunión anual de la Sociedad Botánica de Chile.

<sup>9</sup> SIPAUBA, L. H. e ROCHA, O., Op. Cit.

<sup>10</sup> LESMOS, A. M. 2001. Cultura em larga escala da microalga *Ankistrodesmus gracilis* (CHLOROPHYCEAE) e do microcrustáceo *Diaphanosoma birgei* em laboratório. Tese (Doctor em Aqüicultura). Jaboticabal, SP, Brasil: CAUNESP.

<sup>11</sup> ALVAREZ, H. 1994. Aspectos biológicos generales. Folleto de algas. Guayaquil, Ecuador: Escuela Superior Politécnica del Litoral.

<sup>12</sup> Ibid.

Sipaúba y Rocha<sup>13</sup> mencionan los principales factores físicos y químicos del agua, para el cultivo en condiciones controladas: luz, temperatura, aireación, nutrientes mayores (NPK) y menores, dióxido carbónico, pH, dureza y alcalinidad. Estos factores deben ser controlados plenamente en los cultivos que se realizan en laboratorio.

Según Alvarez<sup>14</sup>, el crecimiento de las algas puede darse en varias fases:

- **Fase de ajuste.** Es la etapa de adaptación que sufren las algas a las nuevas condiciones de cultivo. Esta fase es corta y no hay incremento neto de la población.
- **Fase exponencial.** Aquí las células se duplican sucesivamente en intervalos iguales de tiempo.
- **Fase de retardo o declinación.** El tiempo requerido para duplicar la población aumenta, reduciéndose la tasa de crecimiento.
- **Fase estacionaria.** Es muy corta y no hay un incremento neto de la población, ya que la tasa de crecimiento es igual a la tasa de mortalidad.
- **Fase de muerte.** Aquí la tasa de mortalidad supera la tasa de multiplicación celular.

En cuanto a su valor nutritivo, la energía bruta es uno de los nutrientes más importantes en la evaluación de los alimentos, dado que es la fuente de toda la actividad orgánica. Está constituida por la energía suministrada por la proteína, los lípidos y los carbohidratos. Su presencia en las algas aumenta a medida que las células crecen y su concentración en la población será mayor a medida que este se multiplica. El extracto etéreo comprende las grasas que contienen las células algales y son una fuente importante de energía. Marshal<sup>15</sup> asegura que generalmente se da una disminución de proteína por célula cuando hay un aumento del contenido de extracto etéreo.

La fibra refleja la fracción de materiales derivados de la pared celular que aumenta a medida que la planta crece o envejece<sup>16</sup> y las proteínas microalgales brindan pobre digestibilidad o disponibilidad si el contenido de fibra aumenta, es decir, cuando las plantas maduran<sup>17</sup>. La fibra constituye el residuo orgánico insoluble, generalmente considerada como el carbohidrato no disponible en una dieta, cuyo contenido empieza a incrementarse a mitad de la fase exponencial<sup>18</sup>. Los contenidos de clorofila a también varían según la edad de la planta y algunos de estos nutrientes, de vital importancia en la alimentación de las larvas de peces, pueden disminuir a medida que el contenido de fibra aumenta.

Coral<sup>19</sup> (2006) encontró que la concentración absoluta de las células algales ( $Y$ ) muestran un crecimiento de forma exponencial, con respecto al tiempo de cultivo ( $X$ ), representado por el modelo  $\hat{y} = 51,877e^{0,183X}$ , con  $R^2 = 0,7342$ .

---

<sup>13</sup> SIPAUBA, L. H. e ROCHA, O., Op. Cit.

<sup>14</sup> ALVAREZ, H., Op. Cit.

<sup>15</sup> MARSHALL, D. 1987. Biología de las algas: Enfoque fisiológico. México: Limusa.

<sup>16</sup> MORRIS, H. 1999. Composición bioquímica y evaluación de la calidad proteica de la biomasa autotrófica de *Chlorella vulgaris*. En: Revista cubana de alimentación y nutrición. Santiago de Cuba: Centro de Investigaciones de Energía Solar. Vol. 13, No. 1.

<sup>17</sup> QUINTANA, M. M. 1999. Composición bioquímica y evaluación de la calidad proteica: Suplementos dietéticos. En: Revista cubana de alimentación y nutrición. Santiago de Cuba: Centro de Investigaciones de Energía Solar. Vol. 13, No. 2.

<sup>18</sup> TACON, A. G. 1987. The nutrition and feeding of farmed fish and shrimp. A training manual II: Nutrient sources and composition. Brasilia: FAO. 74 p.

<sup>19</sup> CORAL, L. C., Op. Cit.

Este mismo autor, al comparar dos tipos de fertilizantes, uno comercial (CHU12) y otro orgánico (extracto de macrófitas mas NPK) pudo determinar que los contenidos nutricionales de la *Ankistrodesmus gracilis* no presentaron diferencias estadísticas significativas, al igual que el crecimiento poblacional y el contenido de clorofila a, estableciendo que la utilización del fertilizante orgánico tiene menores costos.

## 2.2 MODELOS DE ECUACIONES SIMULTÁNEAS

De acuerdo con Gujarati<sup>20</sup>, los modelos de regresión lineal clásicos (MRLC), donde se tiene una única ecuación, la variable dependiente  $Y$  está expresada en función de una o más variables explicativas  $X_i$ . Sin embargo, hay situaciones en las que no se puede sostener una relación unidireccional entre  $Y$  y las  $X$ ; es muy posible que las  $X$  no sólo afecten a  $Y$ , sino que  $Y$  también afecte a una o más  $Y$ , con una relación bilateral de retroalimentación, en cuya situación los estimadores de mínimos cuadrados ordinarios MCO serían sesgado e inconsistentes. En estos casos se hace necesario plantear más de una ecuación, estructurando los denominados **modelos de regresión con ecuaciones simultáneas**.

Juan, Kizys y Manzanedo<sup>21</sup> recomiendan aplicar estos sistemas de ecuaciones, cada una de las cuales representa un modelo de regresión lineal, cuando la variable dependiente en una ecuación actúa también como variable explicativa en otra ecuación, en modelos con problemas de endogenidad; entonces se estará ante un modelo de ecuaciones simultaneas o modelo multiecuacional. Para solventar los problemas de sesgo e inconsistencia de los estimadores, recomiendan calcular antes la forma reducida del sistema, en la cual cada una de las variables endógenas del sistema es expresada únicamente como función lineal de las variables exógenas del modelo, para calcular luego estimadores mínimos cuadrados indirectos MCI, que dan lugar a estimadores menos sesgado pero consistentes.

En el modelo de regresión lineal clásico (MRLC), los estimadores MCO son los mejores estimadores lineales insesgados, pero en problemas de simultaneidad éstos no suelen serlo, dado que uno de los supuestos más fuertes del MRLC, exige que el término de error aleatorio  $u_t$  y las variables explicativas no deben estar correlacionados. Debido a la correlación existente entre  $Y_t$  y  $u_t$  en los modelos de ecuaciones simultáneas, indicando un problema de autocorrelación, los estimadores son sesgados (en muestras pequeñas) e inconsistentes (en muestras grandes)<sup>22</sup>.

El método alternativo, conocido con el nombre de mínimos cuadrados en dos etapas (MC2E), que son estimadores insesgados y consistentes en estos casos, es decir, a medida que el tamaño de la muestra aumenta indefinidamente, estos estimadores convergen a los verdaderos valores de los parámetros de la población; sin embargo, en muestras pequeñas, o finitas, es posible que los estimadores MC2E sean sesgados, pero son consistentes, al contrario de los MCO que son sesgados e inconsistentes.

**2.2.1 Autocorrelación.** Novales<sup>23</sup> explica que en algunos casos, especialmente cuando los datos están inmersos en una serie de tiempo, la matriz de correlaciones tiene estructura no escalar, esto es, algunos de los elementos fuera de la diagonal son diferentes de cero, situación que proviene del hecho de que el término de error del modelo tiene relación consigo mismo a través del tiempo, producido por existencia de ciclos de tendencia, variables omitidas, relaciones lineales o por relaciones dinámicas. En este caso:

---

<sup>20</sup> GUJARATI, D. N. 2006. Principios de econometría. Madrid: McGraw-Hill.

<sup>21</sup> JUAN, A; KIZYS, R. y MANZANEDO, L. M., Op. Cit.

<sup>22</sup> GUJARATI, D. N., Op. Cit.

<sup>23</sup> NOVALES, A. 1993. Econometría. 2a ed. Madrid: McGraw-Hill.

$$y_t = X^t \beta + u_t, \quad t = 1, 2, \dots, T$$

$$u_t = \rho u_{t-1} + \varepsilon_t$$

Según Gujarati<sup>24</sup>, la autocorrelación se puede definir como “la correlación entre miembros de observaciones ordenados en el tiempo o en el espacio”. Entonces, existe autocorrelación cuando el término de error de un modelo está correlacionado consigo mismo a través del tiempo, es decir  $E(u_t u_s) \neq 0$ . Es decir, el valor esperado del producto de los dos términos de error distintos  $u_t$  y  $u_s$  es diferente de cero.

De acuerdo con este autor, los ciclos de tendencia hacen referencia a la existencia de períodos con autocorrelaciones positivas alternados con autocorrelaciones negativas, es decir, ciclos de tendencias crecientes y decrecientes o rachas de valores altos y bajos.

La omisión de variables también genera autocorrelación, ya que variables explicativas de mucha importancia, que pueden resultar determinantes en la predicción y que se encuentran correlacionadas con  $u_t$ , no son incluidas en el modelo y se dejan dentro de  $u_t$ , por lo cual el factor de perturbación resulta autocorrelacionado.

Las relaciones no lineales son una consecuencia de los ciclos de tendencia, lo cual implica un tipo de relación diferente a la lineal entre  $u_t$  e  $y$ , con rachas de valores positivos alternados con negativos, resultando en problemas de autocorrelación.

Gujarati<sup>25</sup> plantea las consecuencias más importantes de la autocorrelación, las siguientes:

- a. Los estimadores de mínimos cuadrados siguen siendo lineales e insesgados.
- b. Los estimadores MCO ya no son los mejores estimadores lineales insesgados, puesto que no son eficientes, es decir, no tienen varianza mínima.
- c. Las varianzas estimadas de los estimadores MCO están sesgadas, ya que en ocasiones las fórmulas para su cálculo subestiman las auténticas varianzas y errores estándar, inflando los valores de  $t$ .
- d. Como consecuencia de lo anterior, los test habituales de  $t$  y de  $F$  no suelen ser confiables.
- e. La fórmula habitual para calcular la varianza del error (CME) es un estimador sesgado de  $\sigma^2$  y, en algunos casos, es probable que la subestime.
- f. Por consiguiente, el  $R^2$  calculado convencionalmente puede ser una medida poco fiable del verdadero.
- g. Las varianzas calculadas convencionalmente y los errores estándar de previsión también pueden ser ineficientes.

En el caso de estudio en el presente proyecto, es posible que haya una relación temporal o retardo entre  $Y$  y las  $X$  o  $Y_{t-1}$ , como puede darse en la ecuación:

$$Y_{1t} = \beta_{10} + \beta_{11} X_{1t} + \beta_{12} X_{2t} + \gamma_{23} Y_{1t-1} + \gamma_{21} Y_{2t-2} + T + u_{1t}$$

---

<sup>24</sup> GUJARATI, D. N., Op. Cit.

<sup>25</sup> Ibid.

Aquí  $Y_{1t}$  (contenido de proteína) en el momento  $t$  depende de  $X_{1t}$  (densidad poblacional) en el momento  $t$ ,  $X_{1t}$  (contenido de clorofila a) en el momento  $t$ , de  $Y_{1t-1}$  en el momento  $t-1$ , de  $Y_{2t-2}$  (concentración de energía bruta) en el momento  $t-2$ .

Como muestra este modelo, debido a los retardos  $Y_{1t-1}$  y  $Y_{2t-2}$ , la relación entre el contenido de proteína con la concentración de energía bruta no es contemporánea. Según Gujarati<sup>26</sup>, a esto modelos se les denomina **modelos dinámicos**, es decir, implican cambios a lo largo del tiempo, debido a que el efecto de una variación unitaria del valor de las variables explicativas retardadas se deja sentir durante cierto número de períodos. Más técnicamente los denomina **modelos de retardos distribuidos**, por cuanto el efecto se distribuye a lo largo de varios períodos.

Este mismo autor explica que su estimación tiene algunos problemas prácticos, como los siguientes:

- Hay necesidad de determinar cuántos valores retardados de las variables explicativas se debe introducir.
- Si se introducen demasiados retardos, es posible que los grados de libertad disminuyen significativamente, puesto que por cada coeficiente estimado se pierde un grado de libertad, y a medida que éstos se reducen la inferencia estadística es menos fiable.
- Inclusive con muestras grandes, se pueden encontrar problemas de multicolinealidad. Además, los coeficientes de los términos sucesivamente retardados pueden alternar de signo, dificultando la interpretación de algunos de ellos.

Una forma ingeniosa de reducir el número de términos retardados en los modelos de retardos distribuidos, consiste en adoptar el planteamiento utilizado por los **modelos de Koyck de expectativas adaptativas y de ajustes parciales**, el cual se podría reducir a la siguiente expresión:

$$Y_{1t} = C_{10} + C_{12}X_{1t} + C_{13}Y_{1t-1} + T + \nu_{1t}$$

Donde  $\nu$  es el término de error. Este modelo se puede denominar **modelo autorregresivo**, porque el valor retardado de la variable dependiente aparece como variable explicativa. Ahora se deben estimar sólo cuatro incógnitas, con un gran ahorro en grado de libertad, dado que todos los términos retardados de la regresión son sustituidos por un único valor retardado de  $Y$ .

Sin embargo, con esto se han creado algunos problemas:

- Puesto que  $Y_{1t}$  es aleatoria,  $Y_{1t-1}$  también es aleatoria, por lo cual es necesario asegurarse que el término de error  $\nu_{1t}$  y la variable retardada  $Y_{1t-1}$  no están correlacionadas, de lo contrario los estimadores MCO no sólo están sesgado sino que también son inconsistentes, y aún si no están correlacionados, éstos son sesgados en muestras pequeñas, aunque en muestras grandes (asintóticas) tienden a ser consistentes.
- Si  $\nu_{1t}$  tiene una correlación serial de orden uno, los estimadores MCO son sesgados, pero además son inconsistentes y los procedimientos tradicionales de contrastación de  $t$  y  $F$  son inválidos.

Por otra parte, algunas veces, los modelos de regresión con datos de series temporales ofrecen resultados espurios, en el sentido de que los resultados parecen, a primera vista, buenos, pero una investigación más detenida revela que resultan sospechosos. En estos casos, un  $R^2 > d$  (Durbin-

---

<sup>26</sup> Ibid.

Watson) es una regla intuitiva para sospechar que la regresión estimada parece un problema de regresión espuria, es decir, es posible que no exista realmente una relación significativa, fenómeno que sucede cuando se hace una regresión de una serie temporal no estacionaria sobre otra serie temporal no estacionaria<sup>27</sup>.

De acuerdo con Wooldrige<sup>28</sup>, un proceso  $y_t = \rho_1 y_{t-1} + e_t$ ,  $t = 1, 2, \dots$ , donde el punto de partida es  $y_0 (t = 0)$  y  $\{e_t : t = 1, 2, \dots\}$  es una sucesión i.i.d. con media cero y varianza  $\sigma^2$ ; además se supone que  $y_0$  y  $e_t$  son independientes y que  $E(y_0) = 0$ , se le denomina **proceso autorregresivo de orden uno AR(1)**. Un modelo AR(1) es esencialmente importante en el análisis de regresión múltiple con datos de series de tiempo. Tal como sostiene Madala<sup>29</sup>, considerando un modelo de ecuaciones simultáneas, éstas reciben el nombre de estructurales si tienen la siguiente forma:

$$\begin{aligned} Y &= \alpha_1 + \beta_1 Z + \gamma_1 X + u_1 \\ Z &= \alpha_2 + \beta_2 Z + \gamma_2 W + u_2 \end{aligned}$$

En este caso,  $Y$  y  $Z$  son variables endógenas, mientras que  $X$  y  $W$  son las variables exógenas. Dado que las variables exógenas son independientes de los términos de error  $u_1$  y  $u_2$  y satisfacen los requisitos usuales para la estimación de mínimos cuadrados ordinarios, es posible estimar las regresiones de  $Y$  y  $Z$  por medio de MCO, aunque no es posible estimar estas ecuaciones por dicho método.

Si se resuelve el sistema de ecuaciones para  $Y$  y  $Z$  en términos de  $X$  y  $W$  se obtiene:

$$\begin{aligned} Y &= \frac{\alpha_1 \beta_2 - \alpha_2 \beta_1}{\beta_2 - \beta_1} + \frac{\gamma_1 \beta_2}{\beta_2 - \beta_1} X - \frac{\gamma_2 \beta_1}{\beta_2 - \beta_1} W + u_1 \\ Z &= \frac{\alpha_1 - \alpha_2}{\beta_2 - \beta_1} + \frac{\gamma_1}{\beta_2 - \beta_1} X - \frac{\gamma_2}{\beta_2 - \beta_1} W + u_2 \end{aligned}$$

Estas se conocen como ecuaciones en la forma reducida. Entonces es posible escribirlas de la siguiente manera:

$$\begin{aligned} Y &= \pi_1 + \pi_2 X + \pi_3 W + u_1 \\ Z &= \pi_4 + \pi_5 X + \pi_6 W + u_2 \end{aligned}$$

Los  $\pi$ ' se conocen como parámetros de la forma reducida. Entonces las estimaciones de estas ecuaciones por mínimos cuadrados ordinarios proporcionan estimadores consistentes para los parámetros de la forma reducida.

Con base en lo anterior, a partir de las ecuaciones  $Y$  y  $Z$  sobre  $X$  y  $W$  es posible recuperar los parámetros de las ecuaciones originales, mediante el método de mínimos cuadrados indirectos (MCI).

<sup>27</sup> Ibid.

<sup>28</sup> WOOLDRIGE, J. M. 2001. Introducción a la Econometría. Trad. Por Francisco Javier Dávila. México: Thomson.

<sup>29</sup> MADDALA, G. S. 1996. Introducción a la Econometría. 2ª ed. Prentice Hall.

Sin embargo, antes de decidir si se aplica este método, es necesario resolver el problema de identificación.

**2.2.2 Identificación de ecuaciones.** Según Juan, Kizys y Manzanedo<sup>30</sup>, el problema de identificación hace referencia a la posibilidad o no de estimar los parámetros estructurales de un modelo de ecuaciones simultáneas a partir de los parámetros de la forma reducida asociada, los cuales sí se podían estimar mediante MCO.

En este sentido:

- Se dice que una ecuación está **no identificada** cuando no se tiene suficiente información para estimar los parámetros de la forma estructural de la ecuación.
- Se dice que una ecuación está **sobreidentificada** cuando haya más de una combinación posible de valores estimados para los parámetros de la forma estructural.
- Finalmente, una ecuación estará **exactamente identificada** cuando sólo sea posible obtener una única estimación de los parámetros estructurales.
- Dado un modelo multiecuacional en forma estructural, se dice que es un sistema exactamente identificado cuando todas sus ecuaciones lo sean. En tal caso, será posible obtener, de forma unívoca, estimaciones para los parámetros estructurales a partir de las estimaciones de la forma reducida.

Maddala<sup>31</sup> explica que estos sistemas lineales disponen de una sencilla regla de conteo que se conoce como condición de orden para la identificación, y es la siguiente:

Sean:  $g$  = número de variables endógenas en el sistema.

$k$  = total de variables (endógenas y exógenas) que faltan en la ecuación bajo consideración.

Entonces:

- Si  $k = g - 1$ , entonces la ecuación está exactamente identificada.
- Si  $k > g - 1$ , entonces la ecuación está sobreidentificada.
- Si  $k < g - 1$ , entonces la ecuación está subidentificada.

Aunque la condición de orden es sólo una condición necesaria y no suficiente (el hecho de que se cumpla  $k \geq g - 1$  no implica de por sí que la ecuación esté identificada), en la gran mayoría de los casos proporciona la respuesta correcta al problema de la identificación sin necesidad de recurrir a la condición de rango.

Dado que la condición de orden no siempre es necesaria y suficiente y puede resultar complicada cuando se trata de más de dos ecuaciones, es necesario utilizar una forma más sistemática para hacerlo, utilizando la condición de rango, la cual consiste en determinar si la ecuación considerada puede obtenerse como una combinación lineal de las demás ecuaciones, conformando una matriz donde el rango sea  $(g - 1)$ . Sin embargo, este método establece si la ecuación está identificada o no; con base en la condición de orden es posible saber si se identifica exactamente o se sobreidentifica. Además, la condición de rango puede proporcionar ecuaciones no estimables, es decir que sus resultados no

---

<sup>30</sup> JUAN, A; KIZYS, R. y MANZANEDO, L. M., Op. Cit.

<sup>31</sup> MADDALA, G. S., Op. Cit.

pueden tener interpretación dentro del problema, por lo tanto es conveniente verificar con la condición de orden.

**2.2.3 Estimación por el método de mínimos cuadrados en dos etapas.** Tal como afirma Maddala<sup>32</sup>, después de decidir sobre el problema de identificación, es necesario abordar la estimación de los modelos de ecuaciones simultáneas. La utilización de otros métodos, como MCI, sin embargo, requieren que las ecuaciones estén exactamente identificadas y, además, puede resultar muy complicado si existen varias ecuaciones, razón por la cual no se usa con frecuencia y es necesario considerar otros métodos de aplicación más generalizada.

A este respecto, Gujarati<sup>33</sup> afirma que si es posible encontrar una variable instrumental de  $Y$  tal que, aunque se parece a  $Y$ , no está correlacionada con  $u_t$ , se puede utilizar directamente el método de MCO para estimar los parámetros de la ecuación en referencia. Después de encontrar esa variable instrumental se puede aplicar el método de mínimos cuadrados en dos etapas (MC2E), el cual implica dos aplicaciones sucesivas de MCO, utilizando las variables  $y$  como regresoras y no como instrumentales.

Juan, Kizys y Manzanedo<sup>34</sup> dicen que el método MC2E permite obtener estimadores consistentes para los parámetros estructurales en el caso de ecuaciones sobreidentificadas o exactamente identificadas. El método MC2E consiste en:

- **Primera etapa:** Para cada variable endógena explicativa de la ecuación en la forma reducida, hallar la ecuación de regresión de ésta sobre todas las variables exógenas del sistema, mediante MCO y obtener los valores predichos.
- **Segunda etapa:** Con las ecuaciones de regresión obtenidas, hallar los valores estimados  $\hat{y}$  para cada variable endógena, y realizar la regresión de la variable endógena dependiente sobre las variables explicativas usando dichos valores estimados (en lugar de los valores observados), mediante MCO.

De acuerdo con Maddala<sup>35</sup>, los estimadores no cambian cuando se reemplaza  $y = bY + cZ + u$  por  $\hat{y}$ , entonces se obtiene la ecuación normal que se puede escribir como:

$$y = \hat{y} + v$$

Donde  $v$  no se correlaciona con  $Z$ , puesto que  $\hat{y}$  es función lineal de  $Z$ . De allí que en la segunda etapa del método MC2E, es posible reemplazar todas las variables endógenas en la ecuación por sus valores predichos a partir de las formas reducidas y después estimar la ecuación mediante MCO.

Entonces, los errores estándar estimados en la segunda etapa serán diferentes debido a que la variable dependiente es  $\hat{y}$  en lugar de  $y$ . Sin embargo los errores estándar son incorrectos, por lo tanto no importa si las variables endógenas de un segundo miembro o todas ellas se reemplazan por  $\hat{y}$  en la segunda etapa del método.

---

<sup>32</sup> Ibid

<sup>33</sup> GUJARATI, D. N., Op. Cit.

<sup>34</sup> JUAN, A; KIZYS, R. y MANZANEDO, L. M., Op. Cit.

<sup>35</sup> MADDALA, G. S., Op. Cit.

A fin de obtener estimadores consistentes de los parámetros, cuando  $Y$  y  $u$  están correlacionadas, se necesita información adicional, proveniente de alguna variable que satisface ciertas propiedades, la cual se denomina **variable instrumental** de  $Y$ <sup>36</sup>.

Dado un modelo de regresión con una ecuación estructural, donde se sospecha autocorrelación en alguna de las variables explicativas, o porque esa variable explicativa resulta ser endógena en otra ecuación del modelo de ecuaciones simultáneas, se hace necesario buscar una variable instrumental de la variable explicativa endógena, la cual no se correlaciona con  $u$ . Las variables instrumentales pueden generarse a partir de retardados de las mismas variables endógenas, con el fin de poder obtener una ecuación exactamente identificada, en cuyo caso los estimadores MCO son insesgados y consistentes, con lo cual se estaría aplicando un método de mínimos cuadrados en dos etapas (MC2E).

---

<sup>36</sup> WOOLDRIGE, J. M., Op. Cit.

### 3. PRESENTACIÓN Y ANÁLISIS DE LOS RESULTADOS

#### 3.1 ANÁLISIS EXPLORATORIO

En la Tabla 1 se presentan las principales medidas de estadística descriptiva, para las diferentes variables incluidas en el estudio. Este análisis inicial se efectuó tomando los datos originales, durante todo el período biológico, el cual tiene duración de 15 días, que equivale a 30 datos, puesto que el ensayo se desarrolló con una repetición. Aquí se puede resaltar una variabilidad baja (representada por el coeficiente de variación) en las variables Proteína, Energía y Extracto Etéreo, media en Clorofila y alta en Fibra y Población. Esto indica que las variables endógenas tienen baja dispersión, no así las variables exógenas.

**Tabla 1. Estadísticas descriptivas para las variables en estudio, en período de 15 días.**

Estadístico	Proteína	Energía	EE	Población	Clorofila	Fibra
Media	43,9337	5150,5428	26,0530	311,1643	1259,8520	7,2803
Error típico	0,3339	57,2418	0,1614	31,7735	38,9550	0,4976
Mediana	43,8450	5132,3427	25,8700	387,9900	1233,8500	7,1000
Desviación estándar	1,8288	313,5260	0,8839	174,0308	213,3652	2,7254
Varianza	3,3445	98298,5414	0,7813	30286,7029	45524,7004	7,4280
Coef. de variación	0,0416	0,0609	0,0339	0,5593	0,1694	0,3744
Curtosis	-0,5426	-0,7866	-0,8461	-1,3220	-0,2658	-1,0533
Asimetría	-0,3131	0,3261	0,1376	-0,5701	0,5588	0,1635
Rango	6,3400	1124,1539	3,4100	503,2200	755,5000	9,5700
Mínimo	40,3400	4565,2790	24,2800	21,8000	918,5000	3,2600
Máximo	46,6800	5689,4329	27,6900	525,0200	1674,0000	12,8300

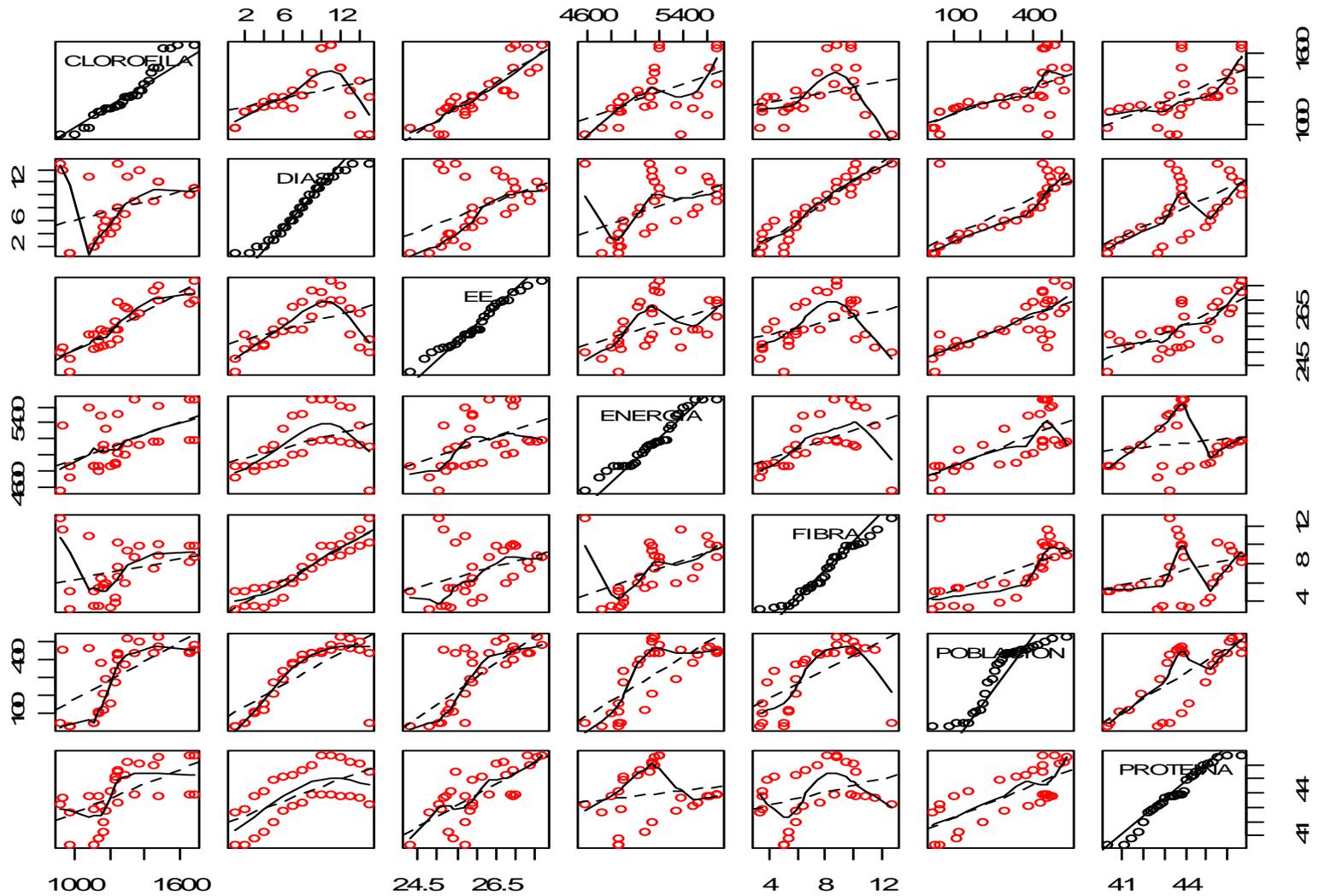
Por otra parte, la totalidad de las variables tienen un valor de asimetría cercano a cero y tienen tendencia a una distribución platicúrtica, con valores negativos en este coeficiente. Esta característica es importante, por cuanto las predicciones mediante modelos en los cuales se asume un supuesto de normalidad, como es el caso de este estudio, las predicciones serán significativas.

La Figura 1 representa la distribución de los datos, en una matriz de dispersión, donde se incluye la totalidad de los datos, durante 15 días de estudio. Un análisis detallado de estos diagramas permite asegurar que todas las variables tienen una fase de incremento exponencial, la cual aumenta hasta el día 12 o 13, de allí se presenta una fase estacionaria y luego otra de descenso. La única variable que tiene un comportamiento un tanto diferente es la Fibra, ya que a partir de la mitad de la fase exponencial se presenta un fuerte aumento hasta el final del período, tal como afirma Tacon<sup>37</sup> y su incremento influye en la disminución de la disponibilidad de los nutrientes<sup>38</sup>. Como consecuencia, la Clorofila, la Energía, el EE y la Proteína disminuyen más temprano, es decir, tan pronto como el incremento de la Fibra es significativo.

<sup>37</sup> TACON, A. G., Op. Cit.

<sup>38</sup> QUINTANA, M. M., Op. Cit.

Figura 1. Diagrama de dispersión para las variables incluidas en el estudio, con datos en un período de 15 días.



Los anteriores argumentos llevaron a la necesidad de incluir en el estudio los datos correspondientes al período comprendido entre los días 1 y 13, es decir, hasta el punto de máximo crecimiento poblacional, que representa el momento que actualmente se considera adecuado para la cosecha, puesto que se tiene la máxima cantidad de biomasa y por lo tanto el mayor volumen de alimento. Por otra parte, si se tienen en cuenta más allá del día 13, cuando la mayoría de las variables presentan decrecimiento, es posible que las relaciones de asociación y dependencia entre las diferentes variables, no sean significativas por no ajustarse a una relación lineal, además, el comportamiento más allá de ese punto no es de interés en la producción de plancton.

La Tabla 2 muestra las medidas de estadística descriptiva utilizando un período de 13 días de crecimiento del alga, que en total está representado por 26 datos. De manera similar, la variabilidad es baja en las variables endógenas (Proteína, Extracto etéreo y Energía) y en las exógenas (Clorofila, Fibra, Población) es alta.

**Tabla 2. Estadísticas descriptivas para las variables en estudio, en período de 13 días.**

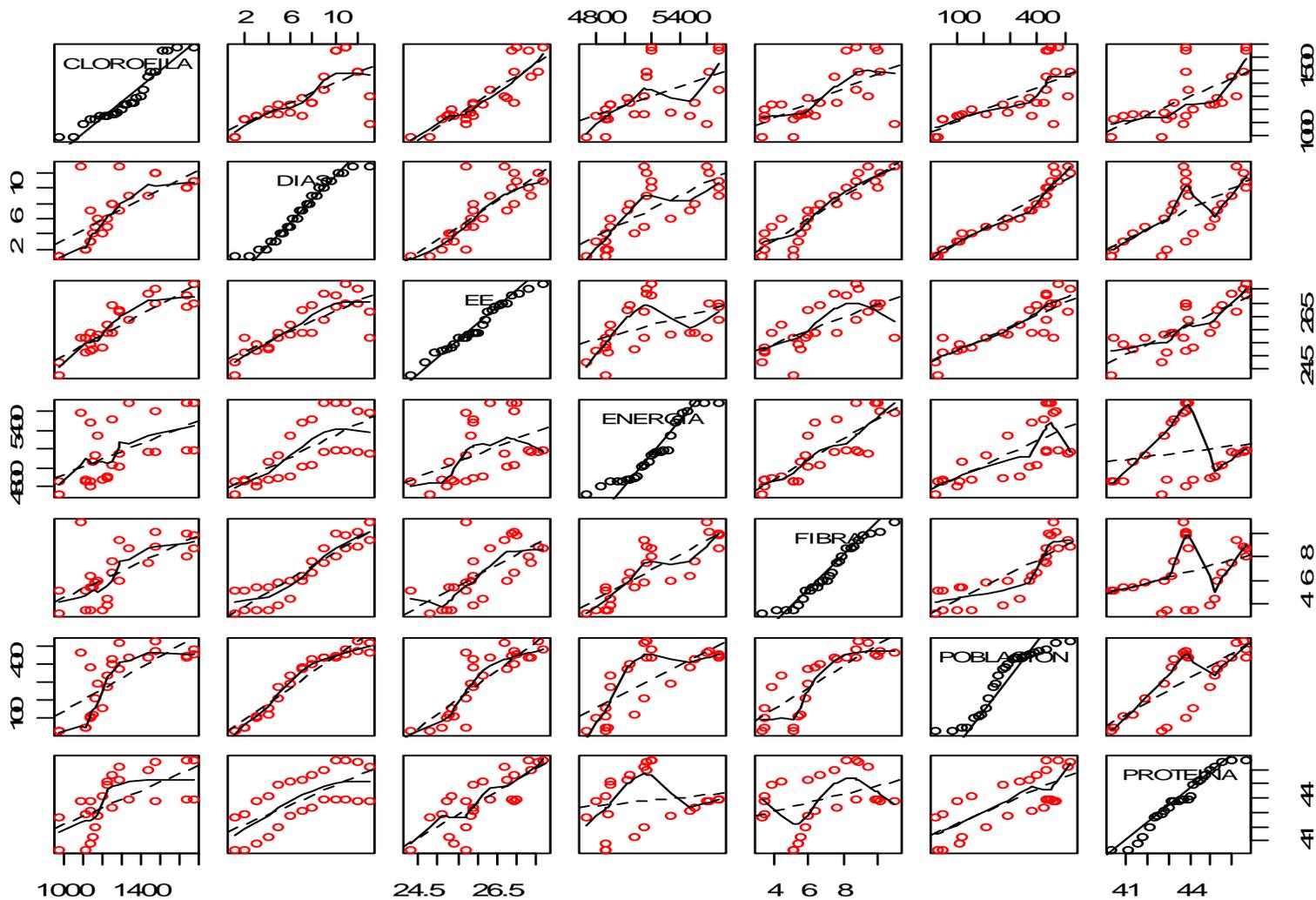
<b>Estadística</b>	<b>Proteína</b>	<b>Energía</b>	<b>EE</b>	<b>Población</b>	<b>Clorofila</b>	<b>Fibra</b>
Media	43,8312	5166,3672	26,1308	304,3762	1282,3115	6,6850
Error típico	0,3704	61,3678	0,1760	33,7102	40,5745	0,4659
Mediana	43,8450	5140,8963	25,8800	373,4600	1233,8500	6,1750
Desviación estándar	1,8888	312,9155	0,8974	171,8888	206,8903	2,3757
Varianza	3,5674	97916,1341	0,8054	29545,7621	42803,5923	5,6437
Coef. de variación	0,0431	0,0606	0,0343	0,5647	0,1613	0,3554
Curtosis	-0,6425	-1,0861	-0,7825	-1,3548	-0,4007	-1,1937
Asimetría	-0,2407	0,4869	-0,0166	-0,4954	0,7288	0,1693
Rango	6,3400	973,4635	3,4100	503,2200	697,5000	7,7000
Mínimo	40,3400	4715,9694	24,2800	21,8000	976,5000	3,2600
Máximo	46,6800	5689,4329	27,6900	525,0200	1674,0000	10,9600

De forma similar, el valor del coeficiente de asimetría es cercano a cero, en todas las variables, y la distribución es levemente platicúrtica, por cuanto los valores del coeficiente son negativos y cercanos a uno, con las consecuencias benéficas para los procesos posteriores, que se mencionaron anteriormente.

Como se observa en la Figura 2, cuando se utilizan los datos para un período de 13 días, las variables se ajustan más a una línea recta, lo cual garantiza mayor ajuste en la relación lineal entre variables. Como ya se ha dicho, las únicas variables con comportamiento diferente son la Fibra que se incrementa notoriamente desde la mitad de la fase exponencial, además de la Proteína y la Energía, que empiezan a disminuir su disponibilidad desde el día 10; posiblemente son las que más sufren la influencia por el incremento de la Fibra.

Debido al comportamiento de las diferentes variables, según se explicó anteriormente, en adelante se realizarán todos los análisis y procedimientos, utilizando los datos obtenidos durante un período de cultivo de 13 días.

Figura 2. Diagrama de dispersión para las variables incluidas en el estudio, con datos en un período de 13 días.



### 3.2 ANÁLISIS DE CORRELACIÓN

El análisis de correlación múltiple, que se presenta en la Tabla 3, permite concluir que los mayores valores del coeficiente de Pearson corresponden a las asociaciones de todas las variables con Días, es decir, están asociadas con el tiempo de cultivo; de igual forma muestra valores altos de  $r$  la asociación entre Extracto etéreo y Clorofila, Población y Clorofila, Población y EE, Proteína y EE, Fibra y Energía, Población y Energía, Población y Fibra, Proteína y Población; con un valor menor de  $r$ , que la cual se considera una relación fuerte por ser significativa al 95% de confianza, ( $p < 0,05$ ), se encuentran los pares de variables Energía y Clorofila, Fibra y Clorofila, Proteína y Clorofila, Fibra y Energía.

**Tabla 3. Correlación múltiple entre las diferentes variables.**

	Clorofila	Días	EE	Energía	Fibra	Población	Proteína
<b>Clorofila</b>	r	0,716	0,879	0,5388	0,6618	0,7235	0,637
	n	26	26	26	26	26	26
	p-valor	0	0	0,0045	0,0002	0	0,0005
<b>Días</b>	0,716	r	0,8133	0,7344	0,9161	0,9606	0,6801
	26	n	26	26	26	26	26
	0	p-valor	0	0	0	0	0,0001
<b>EE</b>	0,879	0,8133	r	0,4692	0,6808	0,8523	0,7841
	26	26	n	26	26	26	26
	0	0	p-valor	0,0156	0,0001	0	0
<b>Energía</b>	0,5388	0,7344	0,4692	r	0,8374	0,7225	0,1659
	26	26	26	n	26	26	26
	0,0045	0	0,0156	p-valor	0	0	0,4181
<b>Fibra</b>	0,6618	0,9161	0,6808	0,8374	r	0,8376	0,3796
	26	26	26	26	n	26	26
	0,0002	0	0,0001	0	p-valor	0	0,0558
<b>Población</b>	0,7235	0,9606	0,8523	0,7225	0,8376	r	0,7408
	26	26	26	26	26	n	26
	0	0	0	0	0	p-valor	0
<b>Proteína</b>	0,637	0,6801	0,7841	0,1659	0,3796	0,7408	R
	26	26	26	26	26	26	N
	0,0005	0,0001	0	0,4181	0,0558	0	p-valor

No se encontró correlación lineal significativa entre la Proteína con la variable Energía, aunque la correlación lineal con la Fibra es significativa al 95% de confianza, dado su valor de probabilidad ( $p=0,0558$ ).

Sin embargo, cuando se analiza las Correlaciones Parciales (Tabla 4), en todos los casos resultó ser significativa, positiva entre EE y Clorofila, Fibra y Días, Población y Energía, Energía y Clorofila, Proteína y Días, Población y EE, y negativa entre Energía y EE, Proteína y Energía, Proteína y Fibra. Esto indica que su relación depende en gran medida de la acción conjunta entre las diferentes variables.

**Tabla 4. Matriz de correlaciones parciales.**

	<b>Clorofila</b>	<b>Días</b>	<b>EE</b>	<b>Energía</b>	<b>Fibra</b>	<b>Población</b>	<b>Proteína</b>
Clorofila		-0,1454	0,7680	0,5798	0,2400	-0,4911	0,4658
Días	-0,1454		-0,1020	-0,0309	0,8571	0,4058	0,5411
EE	0,7680	-0,1020		-0,5524	0,0795	0,5820	-0,1240
Energía	0,5798	-0,0309	-0,5524		0,0543	0,7096	-0,5577
Fibra	0,2400	0,8571	0,0795	0,0543		-0,1141	-0,6599
Población	-0,4911	0,4058	0,5820	0,7096	-0,1141		0,4152
Proteína	0,4658	0,5411	-0,1240	-0,5577	-0,6599	0,4152	

### 3.3 ANÁLISIS PRELIMINAR DE REGRESIÓN LINEAL MÚLTIPLE

Antes de iniciar la estimación de los modelos, se realizaron pruebas de normalidad, mediante el estadístico de Shapiro-Wilks. Al probar la hipótesis nula  $H_0$ : la muestra proviene de una población con distribución normal, no existieron elementos para que sea aceptada ( $p < 0,05$ ) en las variables Días de cultivo y Población. Para solucionar este problema, los valores utilizados corresponden a transformaciones mediante raíz cuadrada en las variables Población ( $X_{1t}$ ) y Días de cultivo ( $T$ ); en las variables Proteína ( $Y_{1t}$  y  $Y_{1t-1}$ ), EE ( $Y_{3t}$  y  $Y_{3t-1}$ ) y Fibra ( $X_{3t}$ ), se realizaron transformaciones mediante logaritmo natural, con el fin de solucionar problemas de heterocedasticidad, detectados en los análisis de residuos al realizar la estimación de los modelos.

**3.3.1 Modelo para la variable Proteína.** Se corrió un análisis de regresión para el modelo Proteína sobre las demás variables:

$$\text{Proteína} = \text{Población} + \text{Clorofila} + \text{Fibra} + \text{Energía} + \text{EE} + \text{Días}$$

En la Tabla 5 se muestra el análisis de varianza para el modelo de la primera ecuación, donde la variable regresora es la Proteína; se puede deducir que el modelo es significativo al 99% de confiabilidad ( $p < 0,01$ ), es decir, existe dependencia lineal de la Proteína con relación a las demás variables. Allí también se establece un  $R^2$  alto y la posible presencia de autocorrelación, dado el valor del estadístico de Durbin-Watson.

Con base en los anteriores argumentos, con relación al análisis de regresión y el criterio Cp de Mallow, se decidió excluir la variable EE y se planteó el primer modelo del sistema multicuacional, de la siguiente manera:

$$Y_{1t} = \beta_{10} + \beta_{11}X_{1t} + \beta_{12}X_{2t} + \beta_{13}X_{3t} + \gamma_{12}Y_{2t} + T + u_{1t}$$

La no significancia del EE se puede deber a que el EE es una fuente importante de energía, producida por los lípidos<sup>39</sup>, pero no la única, puesto que la Energía está producida por diversos componentes como las proteínas, los carbohidratos y los lípidos<sup>40</sup>, por lo tanto ésta última puede ser más importante,

<sup>39</sup> MARSHALL, D., Op. Cit.

<sup>40</sup> FABREGAS, J., Op. Cit.

y mostrar mayor influencia sobre el contenido de proteína.

**Tabla 5. Análisis de regresión lineal preliminar de la primera ecuación ( $Y_{1t}$  = Proteína).**

Análisis		Valor F	p-valor	R <sup>2</sup>	R <sup>2</sup> ajustado	Durbin Watson	p-valor
ANOVA		38,5000	0,0000	93,9030	91,4642	1,40792	0,0037
<b>Análisis de regresión</b>							
Parámetro	Estimador	Valor T	p-valor	CM	Valor F	p-valor	
	Constante	-0,684611	-1,645	0,1208			
$X_{1t}$	Población (A)	0,00323049	1,11587	0,2820	0,0128849	103,83	0,0000
$X_{2t}$	Clorofila (B)	0,00003271	0,956077	0,3542	0,00112097	9,03	0,0089
$X_{3t}$	Fibra (C)	-0,102976	-3,7862	0,0018	0,00832111	67,05	0,0000
$Y_{2t}$	Energía (D)	-0,00005175	-2,16289	0,0471	0,00492674	39,70	0,0000
$Y_{3t}$	EE (E)	0,323547	0,904742	0,3799	0,000040232	0,32	0,5775
$T$	Días (F)	0,0688168	3,32992	0,0046	0,00137607	11,09	0,0046
<b>Selección del mejor modelo</b>							
MSE	R <sup>2</sup>	R <sup>2</sup> ajustado	Cp Mallow	Variables			
0,000122693	93,5703	91,561	5,81856	ABCDF			

**3.3.2 Modelo para la variable Energía.** A continuación se corrió el análisis de regresión para el modelo:

$$\text{Energía} = \text{Población} + \text{Clorofila} + \text{Fibra} + \text{Proteína} + \text{EE} + \text{Días}$$

Tal como se indica en la Tabla 6, el modelo propuesto es significativo, lo cual quiere decir que existe relación lineal entre las variables regresoras y la Energía como variable respuesta. Esta relación de dependencia explica el 91,2306% de la variabilidad, según el  $R^2$  y el 87,7229% según el  $R^2$  ajustado. Parece existir autocorrelación, dado el valor del estadístico de Durban Watson.

El análisis de regresión lineal (estadístico T) muestra que las variables Fibra y Días no tienen un efecto significativo, individualmente, sobre la Energía. La mayor influencia está ocasionada por el EE y la Proteína, acorde con la afirmación de Fabregas<sup>41</sup>.

Teniendo en cuenta lo explicado por este autor, según el cual, la presencia de proteína en las algas aumenta a medida que las células crecen, su concentración en la población será mayor a medida que ésta se multiplica y disminuye a medida que aumenta la lignificación de las paredes celulares, además de los análisis de regresión y criterio Cp de Mallow (Tabla 6), se decidió incluir las variables Clorofila y Días de cultivo en el modelo, definido de la siguiente manera:

$$Y_{2t} = \beta_{20} + \beta_{21}X_{1t} + \beta_{22}X_{2t} + \beta_{23}X_{3t} + \gamma_{21}Y_{1t} + \gamma_{23}Y_{3t} + T + u_{2t}$$

<sup>41</sup> Ibid.

**Tabla 6. Análisis de regresión lineal preliminar de la segunda ecuación ( $Y_{2t}$  = Energía).**

Análisis	Valor F	p-valor	R <sup>2</sup>	R <sup>2</sup> ajustado	Durbin Watson	p-valor
ANOVA	26,0100	0,0000	91,2306	87,7229	1,6644	0,0232
<b>Análisis de regresión</b>						
Parámetro	Estimador	Valor T	p-valor	CM	Valor F	p-valor
Constante	-10807,7	-3,35754	0,0043			
$X_{1t}$ Población (A)	80,5972	4,17492	0,0008	732179,0	66,46	0,0000
$X_{2t}$ Clorofila (B)	0,895092	3,75531	0,0019	2012,99	0,18	0,6751
$X_{3t}$ Fibra (C)	199,374	0,562192	0,5823	461526,0	41,90	0,0000
$Y_{1t}$ Proteína (D)	-4593,79	-2,16289	0,0471	413493,0	37,54	0,0000
$Y_{3t}$ EE (E)	-7670,02	-2,70335	0,0163	109731,0	9,96	0,0065
$T$ Días (F)	-27,5395	-0,107289	0,9160	126,806	0,01	0,9160
<b>Selección del mejor modelo</b>						
MSE	R <sup>2</sup>	R <sup>2</sup> ajustado	Cp Mallow	Variables		
10335,6	91,2239	88,4813	5,01151	ACDEF		

**3.3.3 Modelo para la variable Extracto etéreo (EE).** El análisis de regresión para la tercera ecuación, donde se tiene como variable endógena el EE, para el modelo:

$$EE = \text{Población} + \text{Clorofila} + \text{Fibra} + \text{Proteína} + \text{Energía} + \text{Días}$$

El análisis de varianza para el modelo revela que existe relación lineal ( $p < 0,01$ ) entre las variables regresoras propuestas y la variable respuesta, que en este caso es el Extracto etéreo (Tabla 7). El modelo representa el 94,9507% de la variabilidad total, según el  $R^2$  y el 92,9309% según el  $R^2$  ajustado y parece que el EE no está autocorrelacionado ( $p > 0,05$ ), de acuerdo con el estadístico de Durban-Watson.

En concordancia con lo expresado por Alvarez<sup>42</sup>, quien considera que el EE se incrementa cuando se tiene mayor actividad clorofílica, va aumentando a medida que la población se reproduce y por ser una fuente de energía, al aumentar este nutriente se ve incrementado el EE. Con base en los resultados del análisis de regresión y criterio Cp de Mallow (Tabla 7), se planteó la tercera ecuación del sistema, así:

$$Y_{3t} = \beta_{30} + \beta_{31}X_{1t} + \beta_{32}X_{2t} + \gamma_{31}Y_{1t} + \gamma_{32}Y_{2t} + T + u_{3t}$$

<sup>42</sup> ALVAREZ, H., Op. Cit.

**Tabla 7. Análisis de regresión lineal preliminar de la tercera ecuación ( $Y_{3t} = EE$ ).**

Análisis	Valor F	p-valor	R <sup>2</sup>	R <sup>2</sup> ajustado	Durbin Watson	p-valor
ANOVA	47,0100	0,0000	94,9507	92,9309	2,0751	0,0743
Análisis de regresión						
Parámetro	Estimador	Valor T	p-valor	CM	Valor F	p-valor
Constante	-0,926226	-4,42216	0,0005			
$X_{1t}$ Población (A)	0,0052537	3,22926	0,0056	0,011494100	187,37	0,0000
$X_{2t}$ Clorofila (B)	0,0000749072	4,83794	0,0002	0,003847630	62,72	0,0000
$X_{3t}$ Fibra (C)	0,0574523	2,58214	0,0208	0,000141127	2,30	0,1501
$Y_{1t}$ Proteína (D)	0,159935	0,904742	0,3799	0,000792129	12,91	0,0027
$Y_{2t}$ Energía (E)	-0,000042712	-2,70335	0,0163	0,000777311	12,67	0,0029
$T$ Días (F)	-0,0343708	-2,02381	0,0612	0,000251259	4,10	0,0612
Selección del mejor modelo						
MSE	R <sup>2</sup>	R <sup>2</sup> ajustado	Cp Mallow	Variables		
0,0000606495	94,6751	93,0111	5,81856	ABDEF		

Con base en los análisis previos, el sistema de ecuaciones simultáneas se ha estructurado como sigue:

$$\begin{aligned}
 Y_{1t} &= \beta_{10} + \beta_{11}X_{1t} + \beta_{12}X_{2t} + \beta_{13}X_{3t} + \gamma_{12}Y_{2t} + T + u_{1t} \\
 Y_{2t} &= \beta_{20} + \beta_{21}X_{1t} + \beta_{22}X_{2t} + \beta_{23}X_{3t} + \gamma_{21}Y_{1t} + \gamma_{23}Y_{3t} + T + u_{2t} \\
 Y_{3t} &= \beta_{30} + \beta_{31}X_{1t} + \beta_{32}X_{2t} + \gamma_{31}Y_{1t} + \gamma_{32}Y_{2t} + T + u_{3t}
 \end{aligned}$$

Donde:

$\beta$  = Parámetros de las variables exógenas.

$\gamma$  = Parámetros de las variables endógenas.

$Y_{1t}$  = Contenido de proteína.

$Y_{2t}$  = Concentración de energía.

$Y_{3t}$  = Contenido de extracto etéreo.

$X_{1t}$  = Densidad poblacional.

$X_{2t}$  = Contenido de clorofila a.

$X_{3t}$  = Contenido de fibra.

$T$  = Días de cultivo.

$u_{t}$  = Error asociado a la  $t$ -ésima observación.

### 3.4 ANÁLISIS DE AUTOCORRELACIÓN

Dado la existencia visible de una serie de tiempo, con el fin de determinar la autocorrelación serial de las variables endógenas, se realizó la prueba de Durbin-Watson entre cada una de las variables y su primer rezago, considerando un proceso autorregresivo de orden uno AR(1), es decir para constatar si  $Corr(u_t, u_{t-1}) = 0$ . Los resultados se muestran en la Tabla 8, suponiendo una prueba de hipótesis con  $H_1 = \rho > 0$ , donde se observa la existencia de autocorrelación en la variable Proteína ( $p < 0,05$ ) y la variable EE ( $p < 0,1$ ), no así en la variable Extracto etéreo ( $p > 0,1$ ).

**Tabla 8. Prueba de autocorrelación mediante el estadístico  $d$  de Durbin-Watson.**

Variable Dependiente $Y_t$	Variable independiente $Y_{t-1}$	$R^2$	$R^2$ ajustado	Valor $d$ Durbin-Watson	p-valor
Proteína	Proteína (rezagada)	0,9689	0,9675	1,4401	0,04927
Energía	Energía (rezagada)	0,9284	0,9251	1,4728	0,30420
Extracto etéreo	Extracto etéreo (rezagada)	0,7020	0,6885	1,8673	0,06440

En la Figura 3 se presentan los diagramas de la función de autocorrelación simple y la función de autocorrelación parcial, para las variables exógenas. A este respecto, Peña<sup>43</sup> sostiene que determinar el orden de un proceso autorregresivo a partir de su función de autocorrelación simple es difícil, ya que ésta es una mezcla de crecimientos exponenciales y sinusoidales, que se amortiguan al avanzar el retardo y no presentan rasgos fácilmente identificables para determinar el orden del proceso, por lo cual se debe apoyar en la función de autocorrelación parcial, puesto que existe un efecto directo de  $y_{t-2}$  sobre  $y_t$ .

En el caso de la Energía, en donde se obtuvo un estadístico de Durbin-Watson no significativo para el proceso AR(1), cuando se observa la función de autocorrelación simple parece seguir un proceso AR(2), con el primer parámetro positivo ( $\Phi_1 > 0$ ) y el segundo negativo ( $\Phi_2 < 0$ ), corroborado por la función de autocorrelación parcial. Considerando este caso, al calcular el estadístico de Durbin-Watson, para la prueba de autocorrelación entre  $Y_{2t}$  sobre  $Y_{2t-2}$  su valor es 1,4182, significativo el 94% de confiabilidad ( $p = 0,5294$ )

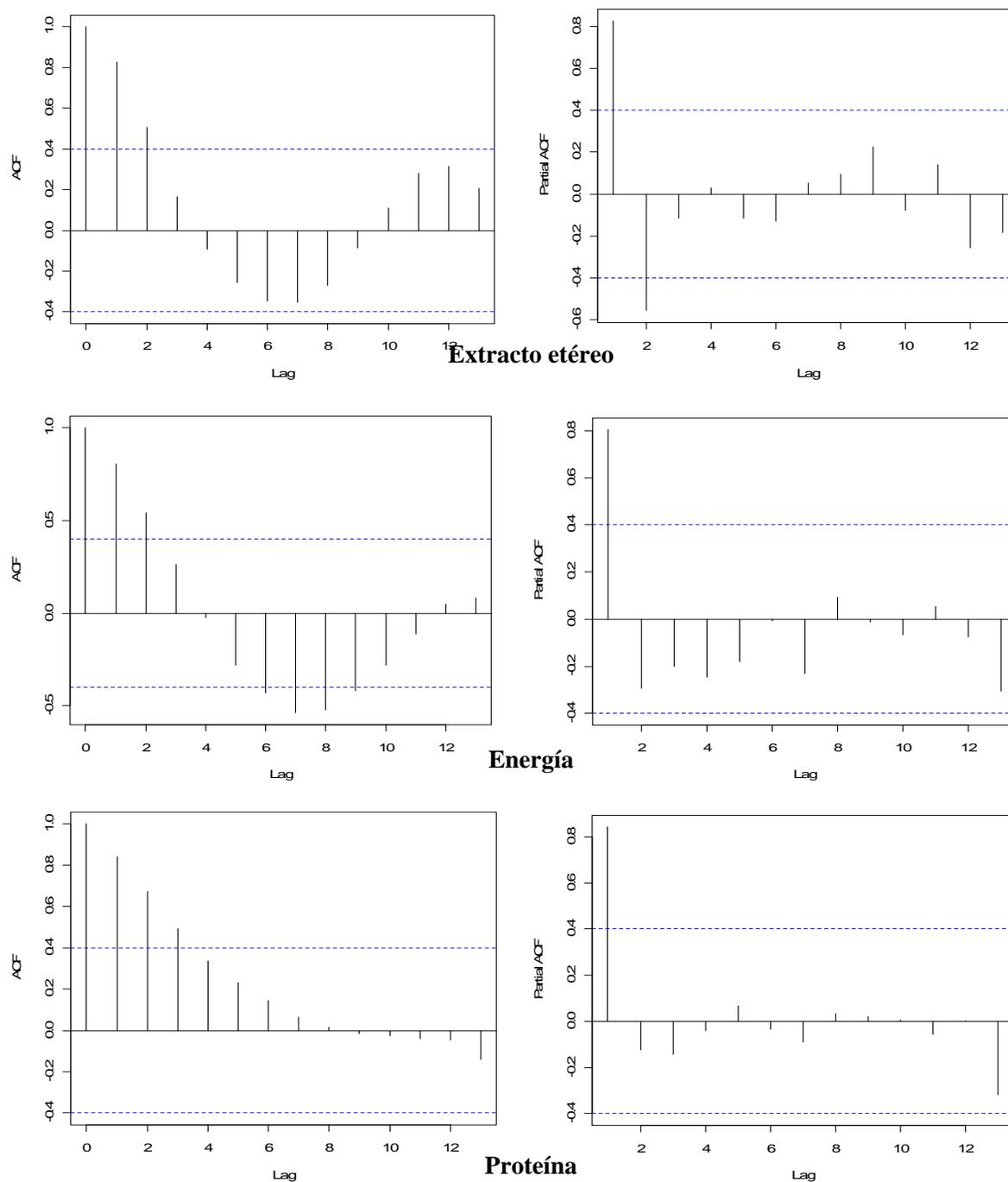
Para el EE, la función de autocorrelación simple parece indicar un proceso AR(1), con  $\Phi_1 > 0$  y  $\Phi_2 < 0$ , y el análisis de la función de autocorrelación parcial lleva a concluir que puede tratarse de un proceso AR(1), con  $\Phi_1 > 0$ . Esto permite aclarar la duda dejada cuando se analizó la autocorrelación mediante el estadístico de Durbin-Watson.

El análisis de los diagramas correspondientes a las funciones de autocorrelación simple y parcial para la variable Proteína, permiten visualizar más claramente un proceso AR(1), con  $\Phi_1 > 0$ .

Los argumentos planteados anteriormente llevan a pensar en la necesidad de incluir algunas variables rezagadas en el modelo, así:  $Y_{1t-1}$  (Proteína con rezago de un período),  $Y_{2t-2}$  (Energía con rezago de dos períodos) y  $Y_{3t-1}$  (Extracto etéreo con rezago de un período).

<sup>43</sup> PEÑA, D. 2001. Análisis de series temporales. Bogotá: Alianza Editorial. 590 p.

**Figura 3. Diagramas de la función de autocorrelación simple y parcial de las variables endógenas.**



De esta manera, considerando que las variables rezagadas son las mismas variables endógenas retardadas en un período de tiempo, se propone el sistema multiecuacional con los modelos ajustados, de la siguiente manera:

$$\begin{aligned}
Y_{1t} &= \beta_{10} + \beta_{11}X_{1t} + \beta_{12}X_{2t} + \beta_{13}X_{3t} + \gamma_{11}Y_{1t-1} + \gamma_{12}Y_{2t} + T + u_{1t} \\
Y_{2t} &= \beta_{20} + \beta_{21}X_{1t} + \beta_{22}X_{2t} + \beta_{23}X_{3t} + \gamma_{21}Y_{1t} + \gamma_{22}Y_{2t-2} + \gamma_{23}Y_{3t} + T + u_{2t} \\
Y_{3t} &= \beta_{30} + \beta_{31}X_{1t} + \beta_{32}X_{2t} + \gamma_{31}Y_{1t} + \gamma_{32}Y_{2t} + \gamma_{33}Y_{3t-1} + T + u_{3t}
\end{aligned}$$

Donde:

$\beta$  = Parámetros de las variables exógenas

$\gamma$  = Parámetros de las variables endógenas.

$Y_{1t}$  = Contenido de proteína.

$Y_{1t-1}$  = Contenido de proteína, con rezago de un período.

$Y_{2t}$  = Concentración de energía.

$Y_{2t-2}$  = Concentración de energía, con rezago de dos períodos.

$Y_{3t}$  = Contenido de extracto etéreo.

$Y_{3t-1}$  = Contenido de extracto etéreo, con rezago de un período.

$X_{1t}$  = Densidad poblacional.

$X_{2t}$  = Contenido de clorofila a.

$X_{3t}$  = Contenido de fibra.

$T$  = Días de cultivo.

$u_t$  = Error asociado a la  $t$ -ésima observación.

Sin embargo, como se dijo antes, todas las variables incluidas en este trabajo tienen un comportamiento de tendencia y, de acuerdo con lo expuesto por Wooldrige<sup>44</sup>, la tendencia temporal se considera exógena, por lo tanto el modelo debe indicar los parámetros de las variables rezagadas con  $\beta$ , reformulando el sistema así:

$$\begin{aligned}
Y_{1t} &= \beta_{10} + \beta_{11}X_{1t} + \beta_{12}X_{2t} + \beta_{13}X_{3t} + \beta_{14}Y_{1t-1} + \gamma_{12}Y_{2t} + T + u_{1t} \\
Y_{2t} &= \beta_{20} + \beta_{21}X_{1t} + \beta_{22}X_{2t} + \beta_{23}X_{3t} + \beta_{24}Y_{2t-2} + \gamma_{21}Y_{1t} + \gamma_{23}Y_{3t} + T + u_{2t} \\
Y_{3t} &= \beta_{30} + \beta_{31}X_{1t} + \beta_{32}X_{2t} + \beta_{34}Y_{3t-1} + \gamma_{31}Y_{1t} + \gamma_{32}Y_{2t} + T + u_{3t}
\end{aligned}$$

### 3.5 IDENTIFICACIÓN DE ECUACIONES Y VARIABLES INSTRUMENTALES

Dado el sistema de ecuaciones previamente definido en la sección anterior, donde  $E(u_{jt}) = 0$  y  $u_{jt}$  no se correlacionan con  $X_{1t}$ ,  $X_{2t}$ ,  $X_{3t}$  ni  $T$ , tal como se determinó al realizar el análisis de correlaciones parciales (Tabla 4). Según ha explicado Wooldrige<sup>45</sup>, la tendencia temporal es exógena, en consecuencia  $u_{jt}$  tampoco se correlaciona con  $Y_{1t-1}$ ,  $Y_{2t-2}$  ni  $Y_{3t-1}$ . Las variables  $Y_{1t}$ ,  $Y_{2t}$  y  $Y_{3t}$  son variables explicativas endógenas, por lo tanto tal vez se correlacionen con  $u_{jt}$ .

Par definir el problema de identificación de las ecuaciones, mediante la condición de orden y la condición de rango, siguiendo el procedimiento recomendado por Maddala<sup>46</sup>, se planteó la matriz que se muestra en la Tabla 9.

<sup>44</sup> WOOLDRIGE, J. M., Op. Cit.

<sup>45</sup> Ibid.

<sup>46</sup> MADDALA, G. S., Op. Cit.

En este sistema, el número de variables endógenas menos uno es 2; por condición de orden todas las ecuaciones están sobreidentificadas, dado que el número de variables endógenas y exógenas excluidas es mayor al número total de variables endógenas menos uno.

**Tabla 9. Matriz de coeficientes para la identificación de ecuaciones mediante la condición de rango.**

Ecuación	$Y_{1t}$	$Y_{2t}$	$Y_{3t}$	$Y_{1t-1}$	$Y_{2t-2}$	$Y_{3t-1}$	$X_{1t}$	$X_{2t}$	$X_{3t}$	$T$
1	0	1	0	1	0	0	1	1	1	1
2	1	0	1	0	1	0	1	1	1	1
3	1	1	0	0	0	1	1	1	0	1

La primera ecuación del sistema presenta dos restricciones de exclusión ( $Y_{2t-2}$  y  $Y_{3t-1}$ ) que serán utilizadas como variables instrumentales para  $Y_{2t}$ , por lo tanto la ecuación satisface la condición de orden, con dos restricciones sobreidentificadas, puesto que se tiene una sola variable endógena en el lado derecho de la ecuación y se dispone de dos variables instrumentales para su estimación. Para verificar la condición de rango, al eliminar la primera fila y dejar las columnas correspondientes a las variables faltantes, se obtiene la siguiente matriz:

1	1	1	0
1	0	0	1

Aquí se puede encontrar dos filas y dos columnas cualesquiera ( $g - 1 = 2$ ) cuyos elementos no son todos cero, por lo tanto, según la condición de rango la ecuación está identificada.

La segunda ecuación también satisface la condición de orden, dado que tiene dos restricciones de exclusión ( $Y_{1t-1}$  y  $Y_{3t-1}$ ) que se pueden utilizar como variables instrumentales para la estimación de  $Y_{1t}$  y  $Y_{3t}$ , variables endógenas incluidas en el lado derecho del modelo; aquí sólo hay una restricción sobreidentificada. Con el fin de verificar la condición de rango, al eliminar la segunda fila y dejar las columnas correspondientes a las variables faltantes, la matriz resultante es:

1	1	0
1	0	1

De nuevo es posible encontrar dos filas y dos columnas cuyos elementos no son todos cero, por lo tanto, según la condición de rango la segunda ecuación está identificada.

De manera similar, en la tercera ecuación se observan tres restricciones de exclusión ( $X_{3t}$ ,  $Y_{1t-1}$  y  $Y_{2t-2}$ ) como variables instrumentales para estimar una variable endógena ( $Y_{2t}$ ) incluida en el lado derecho de la ecuación, por cuyo motivo satisface la condición de orden con dos restricciones sobreidentificadas. En este caso, al verificar la condición de rango, si se elimina la tercera fila y se dejan las columnas correspondientes a las variables faltantes, la matriz resultante es:

0	1	0	1
1	0	1	1

De manera similar, es posible encontrar dos filas y dos columnas cuyos elementos no son todos cero, por lo tanto, según la condición de rango la tercera ecuación está identificada.

De acuerdo con Wooldrige<sup>47</sup>, salvo si hay una falla evidente de identificación, la ecuación que satisface la condición de rango se considera identificada y por lo tanto el método es necesario y suficiente, sin embargo la condición de orden es necesaria para establecer el grado de identificabilidad, pero no es suficiente. Para esto es indispensable establecer como condición que  $T \geq 0$ , dado que no puede existir un momento de tiempo negativo; de otro lado se podría exigir todos los  $\beta_{ij} \neq 0$  en una misma ecuación, para asegurar la identificación mediante la condición de orden, pero esto puede darse por hecho, ya que es prácticamente imposible que todos los  $\beta_{ij} = 0$ , en cuyo caso se podrían presentar ecuaciones no identificadas; como todas las ecuaciones del sistema están sobreidentificadas, con suficientes restricciones de exclusión, sería poco probable que se presentara un caso de no identificación.

### 3.6 MÉTODO DE MÍNIMOS CUADRADOS EN DOS ETAPAS (MC2E)

Ahora se buscará la estimación consistente de los parámetros estructurales a través del método de MC2E. Dado que las ecuaciones de la forma reducida, tal como se han presentado en este trabajo, son predictores óptimos, incluso pueden ser funciones de esperanza condicional, se pueden obtener estimadores consistentes, a partir de ecuaciones estructurales identificadas, estén o no sobreidentificadas, que podrían ser estimadores insesgados, mediante la regresión lineal de  $Y_{it}$  sobre las variables  $X_{it}$  y las rezagadas  $Y_{1t-1}$ ,  $Y_{2t-2}$  y  $Y_{3t-1}$  (que se consideran exógenas), cuyas regresiones se denotan de la siguiente manera:

$$\begin{aligned}\hat{Y}_{1t}^* &= \hat{\pi}_0 + \hat{\pi}_1 X_{1t} + \hat{\pi}_2 X_{2t} + \hat{\pi}_3 X_{3t} + \hat{\pi}_4 Y_{1t-1} + \hat{\pi}_5 Y_{2t-2} + \hat{\pi}_6 Y_{3t-1} \\ \hat{Y}_{2t}^* &= \hat{\pi}_7 + \hat{\pi}_8 X_{1t} + \hat{\pi}_9 X_{2t} + \hat{\pi}_{10} X_{3t} + \hat{\pi}_{11} Y_{1t-1} + \hat{\pi}_{12} Y_{2t-2} + \hat{\pi}_{13} Y_{3t-1} \\ Y_{3t}^* &= \hat{\pi}_{14} + \hat{\pi}_{15} X_{1t} + \hat{\pi}_{16} X_{2t} + \hat{\pi}_{17} X_{3t} + \hat{\pi}_{18} Y_{1t-1} + \hat{\pi}_{19} Y_{2t-2} + \hat{\pi}_{20} Y_{3t-1}\end{aligned}$$

La **primera etapa** consiste en obtener estimadores mediante MCO, para las ecuaciones de la forma reducida planteadas previamente y, a partir de éstos, obtener los correspondientes valores ajustados<sup>48</sup>. Mediante este procedimiento se busca eliminar las correlaciones que probablemente existan entre  $Y_{it}$  y  $u_{it}$ , en cuyo caso  $\hat{Y}_{it}^*$  es una estimación del valor medio de  $Y$  condicional sobre las variables regresoras.

La estimación mediante MCO para la variable  $Y_{it}^*$ , produjo los siguientes modelos:

$$\begin{aligned}\hat{Y}_{1t}^* &= -0,142112 + 0,000998772X_{1t} + 0,0000132628X_{2t} - 0,0105273X_{3t} + 0,849833Y_{1t-1} - 0,000013524Y_{2t-2} - 0,0228623Y_{3t-1} \\ \hat{Y}_{2t}^* &= -5462,9 + 37,38X_{1t} + 0,205352X_{2t} + 329,908X_{3t} - 2049,39Y_{1t-1} + 0,468658Y_{2t-2} - 4885,35Y_{3t-1} \\ \hat{Y}_{3t}^* &= -0,92194 + 0,00376819X_{1t} + 0,0000816511X_{2t} + 0,00343195X_{3t} - 0,0631164Y_{1t-1} - 0,0000402846Y_{2t-2} + 0,303424Y_{3t-1}\end{aligned}$$

Los análisis de varianza para los modelos estimados indican que se debe rechazar  $H_0: \beta = 0$ , puesto que  $p < 0,01$ , en todos los casos, por lo tanto se concluye que existe relación de dependencia lineal

<sup>47</sup> WOOLDRIGE, J. M., Op. Cit.

<sup>48</sup> GUJARATI, D. N., Op. Cit.

entre las variables regresoras y la variable dependiente de cada modelo, con confiabilidad del 99%. En la Tabla 10 se muestran los datos correspondientes al análisis de regresión para los modelos estimados mediante MCO en la primera etapa.

Aquí se puede observar, como una particularidad, que los valores de  $R^2$  y  $R^2$  ajustado son altos, lo cual, según los expresado por Maddala<sup>49</sup>, indica que los valores estimados de las variables endógenas están muy cercanos a sus valores observados y, por tanto, es menos probable que éstas últimas estén correlacionadas con las perturbaciones estocásticas en las ecuaciones estructurales originales.

**Tabla 10. Análisis de regresión para la estimación de variables endógenas.**

Parámetro	Proteína ( $\hat{Y}_{1t}$ )		Energía ( $\hat{Y}_{2t}$ )		EE ( $\hat{Y}_{3t}$ )	
	Valor $t$	p-valor	Valor $t$	p-valor	Valor $t$	p-valor
Constante	-1,21742	0,2423	-4,09184	0,0010	-6,17524	0,0000
Población ( $X_{1t}$ )	0,944599	0,3598	3,09104	0,0075	2,72657	0,0156
Clorofila ( $X_{2t}$ )	1,3586	0,1944	1,83925	0,0858	6,39915	0,0000
Fibra ( $X_{3t}$ )	-0,622593	0,5429	1,70594	0,1086	0,155285	0,8787
Proteína AR(1) ( $Y_{1t-1}$ )	7,87541	0,0000	-1,66054	0,1176	-0,447491	0,6609
Energía AR(2) ( $Y_{2t-2}$ )	-1,48837	0,1574	4,50967	0,0004	1,51272	0,1511
EE AR(1) ( $Y_{3t-1}$ )	-0,14898	0,8836	-2,78347	0,0139	-3,39192	0,0040
$R^2$	98,2336		96,2563		94,9443	
$R^2$ ajustado	97,5271		94,7588		92,9220	
Error estándar	0,00599608		68,5777		0,00783727	
Durbin-Watson	2,13327	0,0656	1,81065	0,0497	2,15485	0,0590

La **segunda etapa** consiste en reemplazar las variables endógenas del segundo miembro por  $\hat{Y}_{it}^*$  y estimar las ecuaciones mediante MCO. Por lo tanto el método de MC2E permite reemplazar todas las variables predictoras endógenas en la ecuación por sus valores predichos a partir de las formas reducidas. Ahora, las ecuaciones sobreidentificadas de Proteína, Energía y EE pueden escribirse como:

$$\begin{aligned}
 Y_{1t} &= \beta_{10} + \beta_{11}X_{1t} + \beta_{12}X_{2t} + \beta_{13}X_{3t} + \beta_{14}Y_{1t-1} + \gamma_{12}\hat{Y}_{2t}^* + T + u_{1t}^* \\
 Y_{2t} &= \beta_{20} + \beta_{21}X_{1t} + \beta_{22}X_{2t} + \beta_{23}X_{3t} + \beta_{24}Y_{2t-2} + \gamma_{21}\hat{Y}_{1t}^* + \gamma_{23}\hat{Y}_{3t}^* + T + u_{2t}^* \\
 Y_{3t} &= \beta_{30} + \beta_{31}X_{1t} + \beta_{32}X_{2t} + \beta_{34}Y_{3t-1} + \gamma_{31}\hat{Y}_{1t}^* + \gamma_{32}\hat{Y}_{2t}^* + T + u_{3t}^*
 \end{aligned}$$

Donde:

$$\begin{aligned}
 u_{1t}^* &= u_{1t} + \gamma_{12}\hat{u}_{2t}^* \\
 u_{2t}^* &= u_{2t} + \gamma_{21}\hat{u}_{1t}^* + \gamma_{23}\hat{u}_{3t}^* \\
 u_{3t}^* &= u_{3t} + \gamma_{31}\hat{u}_{1t}^* + \gamma_{32}\hat{u}_{2t}^*
 \end{aligned}$$

<sup>49</sup> MADDALA, G. S., Op. Cit.

Ahora, en la segunda etapa, se vuelve a aplicar MCO a las ecuaciones así transformadas. Como característica importante, las estimaciones obtenidas son consistentes, es decir, convergen hacia los verdaderos valores, a medida que el tamaño de la muestra aumenta indefinidamente.

El modelo de regresión de  $Y_{1t}$  (**Proteína**) sobre las variables incluidas en la primera ecuación, representa una relación lineal, ya que se debe rechazar la hipótesis nula ( $p < 0,01$ ), cuyo modelo estimado es:

$$\hat{Y}_{1t} = -0,186801 + 0,00188877X_{1t} + 0,0000167509X_{2t} - 0,015422X_{3t} + 0,769953Y_{1t-1} - 0,00000175684\hat{Y}_{2t}^* - 0,000775469T$$

Donde:

$\hat{Y}_{1t}$  = Contenido de proteína estimada, transformada mediante logaritmo natural.

$X_{1t}$  = Población, transformada mediante raíz cuadrada.

$X_{2t}$  = Contenido de clorofila a.

$X_{3t}$  = Contenido de fibra transformada mediante logaritmo natural.

$Y_{1t-1}$  = Contenido de proteína, transformada mediante logaritmo natural, con rezago de un período.

$\hat{Y}_{2t}^*$  = Valores estimados de la concentración de energía mediante regresión con el modelo reducido, en la primera etapa de MC2E.

$T$  = Días de cultivo, valores transformados mediante raíz cuadrada.

Este modelo muestra un  $R^2 = 98,1574\%$  y  $R^2$  ajustado =  $97,4203\%$ , valores que indican un buen ajuste del modelo para  $Y_{1t}$ . De igual manera se calcula un error estándar del modelo bajo (0,00612418) y estadístico de Durbin-Watson no significativo ( $d = 2,08389$ ;  $p = 0,0665$ ). Las estimaciones correspondientes al análisis de regresión se indican en la Tabla 11.

**Tabla 11. Análisis de regresión para la estimación de la primera ecuación del sistema.**

Parámetro	Estadístico T	p-valor	Estadístico F	p-valor	ANOVA	
					Cociente F	p-valor
Constante	-1,00407	0,3313			133,18	0,0000
$X_{1t}$ Población (A)	1,35292	0,1961	343,55	0,0000		
$X_{2t}$ Clorofila (B)	1,28856	0,2171	29,89	0,0001		
$X_{3t}$ Fibra (C)	-0,835132	0,4167	221,86	0,0000		
$Y_{1t-1}$ Proteína (D)	4,40111	0,0005	202,16	0,0000		
$\hat{Y}_{2t}$ Energía (E)	-1,03763	0,3159	1,59	0,2259		
$T$ Días (F)	-0,0482981	0,9621	0,00	0,9621		
<b>Selección del mejor modelo</b>						
<b>MSE</b>	<b>R<sup>2</sup></b>	<b>R<sup>2</sup> ajustado</b>	<b>Cp Mallow</b>	<b>Variables</b>		
0,0000394521	97,9325	97,2864	6,83039	ABCDEF		

El análisis de varianza (ANOVA) indica que existe relación lineal entre las variables regresoras y la variable dependiente ( $p < 0,01$ ). El análisis de regresión para la significancia individual de las variables

regresoras (Estadístico T) muestra que la única variable que tiene efecto significativo en el modelo es la Proteína cuando se incluye con rezago de un período (AR1). Sin embargo, al realizar el análisis según el orden de ajuste (Estadístico F), es decir, la inclusión de las variables condicionadas a la presencia de las demás en el modelo, se encuentra que no tienen influencia significativa sobre el modelo las variables Energía y Días de cultivo.

Con base en este análisis, se corrieron modelos excluyendo las variables no significativas, y todos ellos mostraron valores de los estadísticos  $R^2$  y  $R^2$  ajustado menores, cuadrado medio del error y error estándar de estimación mayores, por lo cual se decidió adoptar el modelo completo, por ser el de mejor comportamiento según el criterio de Cp de Mallow, además de las razones expuestas por Sipaúba<sup>50</sup> y Marshal<sup>51</sup>, quienes afirman que generalmente se da una disminución de proteína por célula cuando la planta envejece y hay un aumento del contenido de extracto etéreo, fuente fundamental de energía.

El modelo de regresión de  $Y_{2t}$  (**Energía**) sobre las variables incluidas en la segunda ecuación, representa una relación lineal, ya que hay evidencia suficiente para rechazar la hipótesis nula  $H_0: \beta = 0$  ( $p < 0,01$ ), cuyo modelo estimado es:

$$\hat{Y}_{2t} = -23656,8 + 104,531X_{1t} + 1,72874X_{2t} + 268,214X_{3t} - 0,356208Y_{2t-2} - 4370,53\hat{Y}_{1t}^* - 17481\hat{Y}_{3t}^* + 119,746T$$

Donde:

$\hat{Y}_{2t}$  = Concentración de energía.

$X_{1t}$  = Densidad poblacional, transformada mediante raíz cuadrada.

$X_{2t}$  = Contenido de clorofila a.

$X_{3t}$  = Contenido de fibra, transformada mediante logaritmo natural.

$Y_{2t-2}$  = Concentración de energía, con rezago de dos períodos.

$\hat{Y}_{1t}^*$  = Valores estimados del contenido de proteína de acuerdo con la ecuación reducida de la primera etapa de MC2E.

$\hat{Y}_{3t}^*$  = Valores estimados del contenido de EE de acuerdo con la ecuación reducida de la primera etapa de MC2E.

$T$  = Días de cultivo, valores transformados mediante raíz cuadrada.

El modelo que representa la segunda ecuación tiene un  $R^2 = 96,3255\%$  y un  $R^2$  ajustado =  $94,4882\%$ , lo cual indica que hay un buen ajuste para el contenido de Energía, con un error estándar del modelo bajo (70,3253) y estadístico de Durbin-Watson significativo ( $d = 1,80941$ ;  $p=0,0308$ ).

Al efectuar el análisis de regresión (Tabla 12), cuando se estudia el efecto de las variables regresoras sobre el modelo, en forma individual (Estadístico T), es posible establecer que la Fibra y la Energía (AR2), al igual que los Días de cultivo, no son significativas. Pero cuando el análisis se hace por orden de ajuste, esto es, la inclusión condicionada a que las demás variables se encuentren en el modelo (Estadístico F), las variables cuyo efecto no es significativo en el modelo son la Clorofila y los Días de cultivo. Si bien el aporte condicional de la Clorofila no es significativo, su importancia individual si lo es y esto justifica su presencia en el modelo. En forma contraria, la Fibra y la Energía (AR2) no

<sup>50</sup> SIPAUBA, L. H., Op. Cit.

<sup>51</sup> MARSHALL, D., Op. Cit.

muestran un aporte significativo en forma individual, pero cuando se estudia su presencia condicional, su efecto sobre la Energía muestra significancia estadística.

**Tabla 12. Análisis de regresión para la estimación de la segunda ecuación del sistema.**

Parámetro	Estadístico T	p-valor	Estadístico F	p-valor	ANOVA	
					Cociente F	p-valor
Constante	-3,0402	0,0088			52,43	0,0000
$X_{1t}$ Población (A)	5,16956	0,0001	148,04	0,0000		
$X_{2t}$ Clorofila (B)	3,16105	0,0069	0,41	0,5338		
$X_{3t}$ Fibra (C)	0,98227	0,3426	93,32	0,0000		
$Y_{2t-2}$ Energía (D)	-0,984392	0,3416	82,89	0,0000		
$\hat{Y}_{1t}$ Proteína (E)	-2,3241	0,0357	34,32	0,0000		
$\hat{Y}_{3t}$ EE (F)	-2,7915	0,0144	7,76	0,0146		
$T$ Días (G)	0,509809	0,6181	0,26	0,6181		
Selección del mejor modelo						
MSE	$R^2$	$R^2$ ajustado	Cp Mallow	Variables		
4832,57	95,8966	94,6143	5,63416	ABCDEFG		

Teniendo en cuenta que el criterio de Cp de Mallow propone un modelo que incluye todas las variables, el cual tiene los mejores valores de  $R^2$ ,  $R^2$  ajustado y MSE; además, los argumentos de Sipaúba<sup>52</sup>, quien asegura que la energía bruta está constituida por la energía suministrada por la proteína, los lípidos y los carbohidratos y su presencia en las algas aumenta a medida que las células crecen y su concentración en la población será mayor a medida que este se multiplica, llevaron a adoptar el modelo completo propuesto inicialmente.

El modelo de regresión de  $Y_{3t}$  (EE) sobre las variables incluidas en la tercera ecuación, representa una relación de dependencia lineal, ya que no hay evidencia suficiente para aceptar la hipótesis nula ( $p < 0$ ), cuyo modelo estimado es:

$$\hat{Y}_{3t} = -0,960238 + 0,00705133X_{1t} + 0,0000887397X_{2t} + 0,370497Y_{3t-1} - 0,193376\hat{Y}_{1t}^* - 0,0000118743\hat{Y}_{2t}^* - 0,0227127T$$

Donde:

$\hat{Y}_{3t}$  = Contenido de Extracto etéreo, valores transformados mediante logaritmo natural.

$X_{1t}$  = Densidad poblacional, valores transformados mediante raíz cuadrada.

$X_{2t}$  = Contenido de clorofila a.

$Y_{3t-1}$  = Contenido de extracto etéreo, valores transformados mediante logaritmo natural, con rezago de un período.

$\hat{Y}_{1t}^*$  = Valores estimados de la concentración de Proteína, mediante la ecuación reducida, en la primera etapa de MC2E.

$\hat{Y}_{2t}^*$  = Valores estimados de la concentración de Energía, mediante la ecuación reducida, en la primera etapa de MC2E.

$T$  = Días de cultivo, valores transformados mediante raíz cuadrada.

<sup>52</sup> SIPAUBA, L. H., Op. Cit.

La tercera ecuación presenta un  $R^2 = 94,7625\%$  y un  $R^2 = 92,6675\%$ , lo cual indica que el modelo tiene una representatividad alta sobre la variabilidad del contenido de EE en las algas, con un error estándar del modelo (0,00797691) bastante bajo y estadístico de Durbin-Watson significativo ( $d = 2,20861$ ;  $p=0,0398$ ).

De la Tabla 13 se deduce que el modelo estimado con la tercera ecuación es significativo, esto es, la ANOVA permite concluir que existe relación lineal ( $p < 0,01$ ) entre las variables regresoras incluidas y el EE. Este análisis de regresión, cuando se estudia el efecto de las variables incluidas individualmente en el modelo (Estadístico T), no resulta significativo ( $p < 0,05$ ) en el EE (AR1), la Proteína, la Energía y los Días de cultivo; pero si el análisis se hace mediante un ANOVA adicional para las variables en orden de ajuste (Estadístico F), sólo la variable que hace referencia a los Días de cultivo no presenta un efecto significativo ( $p < 0,05$ ) sobre el EE. Es posible que suceda esto, en todas las variables, dado el comportamiento de los nutrientes a través del tiempo, los que aumentan o disminuyen según el efecto conjunto de las demás variables; de todas maneras, se está tratando una serie de tiempo dinámica que está supeditada al crecimiento de la planta en un período de cultivo, como manifiesta Alvarez<sup>53</sup> y se ha demostrado a través del presente estudio.

**Tabla 13. Análisis de regresión para la estimación de la tercera ecuación del sistema.**

Parámetro	Estadístico T	p-valor	Estadístico F	p-valor	ANOVA	
					Cociente F	p-valor
Constante	-1,82562	0,0879			45,2300	0,0000
$X_{1t}$ Población (A)	3,0268	0,0085	180,64	0,0000		
$X_{2t}$ Clorofila (B)	3,57166	0,0028	60,47	0,0000		
$Y_{3t-1}$ EE (C)	1,13675	0,2735	11,22	0,0044		
$\hat{Y}_{1t}$ Proteína (D)	-0,69386	0,4984	7,72	0,0141		
$\hat{Y}_{2t}$ Energía (E)	-0,928279	0,3680	10,57	0,0054		
$T$ Días (F)	-0,88276	0,3913	0,78	0,3913		
<b>Selección del mejor modelo</b>						
<b>MSE</b>	<b>R<sup>2</sup></b>	<b>R<sup>2</sup> ajustado</b>	<b>Cp Mallow</b>	<b>Variables</b>		
0,0000627533	94,4904	92,7687	5,77927	ABCDEF		

Por otra parte, es necesario tener en cuenta que el Extracto etéreo comprende las grasas que contienen las células algales y son una fuente importante de energía<sup>54</sup>, contenidos que varían según la fase de crecimiento<sup>55</sup>. Con base en estos argumentos y atendiendo a la prueba para selección del mejor modelo, según el criterio Cp de Mallow, se incluyeron todas las variables propuestas inicialmente.

Los análisis anteriores permiten consolidar el sistema de ecuaciones simultáneas, de la siguiente manera:

$$\hat{Y}_{1t} = -0,186801 + 0,00188877X_{1t} + 0,0000167509X_{2t} - 0,015422X_{3t} + 0,769953Y_{1t-1} - 0,00000175684\hat{Y}_{2t}^* - 0,000775469T$$

$$\hat{Y}_{2t} = -23656,8 + 104,531X_{1t} + 1,72874X_{2t} + 268,214X_{3t} - 0,356208Y_{2t-2} - 4370,53\hat{Y}_{1t}^* - 17481\hat{Y}_{3t}^* + 119,746T$$

$$\hat{Y}_{3t} = -0,960238 + 0,00705133X_{1t} + 0,0000887397X_{2t} + 0,370497Y_{3t-1} - 0,193376\hat{Y}_{1t}^* - 0,0000118743\hat{Y}_{2t}^* - 0,0227127T$$

<sup>53</sup> ALVAREZ, H. Op. Cit.

<sup>54</sup> SIPAUBA, L. H., Op. Cit.

<sup>55</sup> ALVAREZ, H. Op. Cit.

En este sistema se puede ver claramente el efecto positivo de la Población y el contenido de Clorofila sobre las variables endógenas (Proteína, Energía y EE); el contenido de Fibra tiene un efecto negativo sobre la concentración de Proteína, lo cual es coherente con lo expresado por Asselbom y Parodi<sup>56</sup>, referido a la disminución de la primera cuando la segunda aumenta. Los rezagos (AR1) de las variables Proteína y EE tienen un efecto positivo, indicando que su concentración aumenta a medida que pasa el tiempo, pero en la Energía (AR2) es negativa, con un proceso contrario. La Energía tiene influencia negativa sobre la Proteína; la Energía y el EE tienen efecto negativo sobre la Energía y la Proteína y la Energía muestran un efecto negativo sobre el contenido de EE, mostrando que la interacción entre las variables endógenas siempre será inversamente proporcional.

### 3.7 DETERMINACIÓN DEL MOMENTO ÓPTIMO DE COSECHA

Con base en los modelos estimados previamente, para el sistema de ecuaciones múltiples, se realizó la predicción para cada uno de los puntos  $Y_{it}$  particular, a partir de los cuales fue posible determinar el momento óptimo de cosecha de la microalga.

**Tabla 13. Valores predichos para el punto óptimo  $Y_{it}$  particular.**

Variable	Valor estimado
Días de cultivo	11
Clorofila (pg/célula)	0,1674
Fibra (%)	8,77
Población (células $\times 10^5$ /ml)	475,57
Extracto etéreo (%)	27,5816
Contenido de Proteína (%)	46,7142
Energía bruta (kcal/kg)	5222,49

De acuerdo con esta estimación, el momento óptimo de cultivo en el cuál se debe realizar la cosecha es a los 11 días, punto en el cual es posible tener la máxima concentración de nutrientes, especialmente Proteína (46,7142%) y Energía bruta (5222,49 kcal/kg) y el menor contenido de fibra (8,77%). Si bien no corresponde a la máxima concentración poblacional, es uno de los valores más altos, que maximiza las variables endógenas. Esto contrasta con los argumentos de Alvarez<sup>57</sup>, Sipaúba y Rocha<sup>58</sup> y Coral<sup>59</sup> quienes realizaron la cosecha a los 13 días, es decir, cuando termina la fase exponencial, argumentando que es el momento en el cual se tiene la mayor concentración poblacional.

<sup>56</sup> ASSELBOM, V. M. y PARODI, E. R., Op. Cit.

<sup>57</sup> ALVAREZ, H., Op. Cit.

<sup>58</sup> SIPAUBA, L. H. e ROCHA, O., Op. Cit.

<sup>59</sup> CORAL, L. C., Op. Cit.

#### 4. CONCLUSIONES

- Todas las variables muestran una fase de incremento exponencial, la cual aumenta hasta el día 12 o 13, de allí se presenta una fase estacionaria y luego otra de descenso, a excepción de la Fibra, la que inicia un fuerte incremento a partir de la mitad de la fase exponencial, hasta el final del período.
- Las asociaciones más fuertes corresponden a la relación de todas las variables con Días de cultivo, entre Extracto etéreo y Clorofila, Población y Clorofila, Población y EE, Proteína y EE, Fibra y Energía, Población y Energía, Población y Fibra, Proteína y Población.
- El análisis de Correlaciones Parciales evidencia asociación fuerte y positiva entre EE y Clorofila, Fibra y Días, Población y Energía.
- Los modelos de ecuaciones simultáneas para describir el cultivo del alga *Ankistrodesmus gracilis* deben incluir variables regresoras como rezagos de las endógenas. En el caso del presente trabajo, la inclusión de las variables endógenas Proteína y EE con rezago de un período y la Energía con rezago de dos períodos, resultaron tener un efecto significativo en los modelos.
- Por condición de orden, todas las ecuaciones del sistema están sobre identificadas, lo cual es corroborado por la condición de rango como una condición necesaria y suficiente.
- Los modelos estimados del sistema de ecuaciones simultáneas, mediante el método de mínimos cuadrados en dos etapas, para el cultivo de la microalga *Ankistrodesmus gracilis*, permiten observar las interacciones entre variables para optimizar el momento de cosecha, de acuerdo con su valor nutricional, que no es posible con los métodos actuales.
- No existe claridad sobre la significancia del efecto del tiempo de cultivo sobre las variables endógenas, quizá por que su comportamiento corresponde a la influencia combinada de todas las variables en estudio. Su inclusión se mantiene dado el tipo de sistema de ecuaciones que se encuentra inmerso en una serie de tiempo.
- El procedimiento permite ver claramente el efecto negativo del contenido de Fibra sobre los nutrientes, especialmente sobre la Proteína, además, la interacción entre las variables endógenas siempre será inversamente proporcional.
- Es importante realizar estudios similares, utilizando el método MC2E, en el cultivo de especies de microalgas colombianas, incluyendo muestras grandes, con datos repetidos.
- El día 11 mostró ser el momento óptimo para la cosecha de la microalga *Ankistrodesmus gracilis* de manera que se maximiza la concentración de nutrientes, especialmente Proteína (46,7142%) y Energía bruta (5222,49 kcal/kg) y minimiza el contenido de fibra (8,77%).

## REFERENCIAS BIBLIOGRÁFICAS

- ALVAREZ, H. 1994. Aspectos biológicos generales. Folleto de algas. Guayaquil, Ecuador: Escuela Superior Politécnica del Litoral. 9 p.
- ASSELBOM, V. M. y PARODI, E. R. 2003. Ficología. En: XXIX Jornadas Argentinas de Botánica & Reunión anual de la Sociedad Botánica de Chile. 38 p.
- BOLD, H. C. y WYNNE, M. J. 1985. Introduction to the algae: structure and reproduction. New Jersey: Prentice Hall. 586 p.
- CORAL, L. C. 2006. Estudio del crecimiento y desarrollo de la microalga *Ankistrodesmus gracilis* (Bold & Wynne, 1985) (chlorophyta) con diferentes medios de cultivo en laboratorio. Trabajo de grado (Ingeniero en Producción Acuícola). Pasto, Colombia: Universidad de Nariño, Facultad de Ciencias Pecuarias, Departamento de Recursos Hidrobiológicos. 117 p.
- FABREGAS, J. 1986. Marine microalgae as a potential source of minerals in fish diets. In: Ecol. Brasil, Vol. 7, No. 51. p. 237-245.
- GUJARATI, D. N. 2006. Principios de econometría. Madrid: McGraw-Hill. 546 p.
- JUAN, A; KIZYS, R. y MANZANEDO, L. M. 2004. Modelos multiecuacionales: Modelos de ecuaciones simultáneas. Cataluña: Universidad Oberta de Cataluña, Proyecto e-Math. 18 p. Disponible en Internet, URL: <http://www.uoc.edu/in3/emath/docs/MMultiecuacionales.pdf>.
- LESMOS, A. M. 2001. Cultura em larga escala da microalga *Ankistrodesmus gracilis* (CHLOROPHYCEAE) e do microcrustaceo *Diaphanosoma birgei* em laboratorio. Tese (Doctor em Aqüicultura). Jaboticabal, SP, Brasil: CAUNESP. 82 p.
- MADDALA, G. S. 1996. Introducción a la Econometría. 2ª ed. Prentice Hall.
- MARSHALL, D. 1987. Biología de las algas: Enfoque fisiológico. México: Limusa. 236 p.
- MORRIS, H. 1999. Composición bioquímica y evaluación de la calidad proteica de la biomasa autotrófica de *Chlorella vulgaris*. En: Revista cubana de alimentación y nutrición. Santiago de Cuba: Centro de Investigaciones de Energía Solar. Vol. 13, No. 1.
- NOVALES, A. 1993. Econometría. 2a ed. Madrid: McGraw-Hill. 676 p.
- PEÑA, D. 2001. Análisis de series temporales. Bogotá: Alianza Editorial. 590 p.
- QUINTANA, M. M. 1999. Composición bioquímica y evaluación de la calidad proteica: Suplementos dietéticos. En: Revista cubana de alimentación y nutrición. Santiago de Cuba: Centro de Investigaciones de Energía Solar. Vol. 13, No. 2.
- SIPAUBA, L. H. 1994. Limnología aplicada à Aqüicultura. Jaboticabal, SP, Brasil: FUNEP. Boletim Técnico do CAUNESP No. 1. 71 p.

SIPAUBA, L. H. e ROCHA, O. 1993. Cultivo em larga escala de organismos planctônicos para alimentação de larvas e alevinos de peixes: I-algas clorofíceas. Em: Biotemas. Jaboticabal, SP, Brasil: FCAVJ-UNESP, Vol. 6, No. 1. p. 96-106.

TACON, A. G. 1987. The nutrition and feeding of farmed fish and shrimp. A training manual II: Nutrient sources and composition. Brasilia: FAO. 74 p.

WOOLDRIGE, J. M. 2001. Introducción a la Econometría. Trad. Por Francisco Javier Dávila. México: Thomson. 816 p.