Simultaneous equation method for estimating the optimal moment for harvesting microalgae

[1]Marco A. Imués-Figueroa, [1]Gustavo A. Torres-Valencia, [2]Jennifer Gómez-Hernández

[1] Department of Hydrobiological Resources, Faculty of Animal Sciences, University of Nariño, Pasto, Colombia; [2] Program of Aquaculture Production Engineering, Faculty of Animal Sciences, University of Nariño, Pasto, Colombia. Corresponding author: M. A. Imués-Figueroa, marcoi@udenar.edu.co

Abstract. Current techniques recommend harvesting microalgae at the end of the exponential phase when the maximum cellular concentration is higher; in the culture of *Ankistrodesmus gracilis*, the mean times are between 12 and 13 days. However, this is not sufficient criteria to define the optimal harvest time, since the nutritional value may not be as indicated, especially in intensive cultures using photobioreactors. The present work applies simultaneous equations method to determine the optimal moment for the harvest of *A. gracilis* grown in a helical tubular photobioreactor. It maximizes population growth and nutrient availability and minimizes fiber content. Daily samples were collected for 15 days to determine: protein content (Y_{1t}), energy concentration (Y_{2t}) and ethereal 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). Regression models of each of the endogenous variables dependent on the exogenous variables were run, to have a guide for formulating the preliminary system of simultaneous equations, consisting of three endogenous-type equations. A time series is visible, and a serial autocorrelation of the endogenous variables with lag was detected in the variables protein (AR1), energy (AR2), ethereal extract (AR1), and lagging variables (Y_{1t-1}, Y_{2t-2}, Y_{3t-1}). Models, whose instrumental variables are fully identified: regression parameters were estimated using the two-stage least squares method (2SLSM). Based on these models estimated for the multiple equation system, the prediction was made for each of the particular Y_{it} points. It was possible to determine the optimal time for harvesting the microalgae after 11 days of cultivation, optimizing the time of harvest in terms of quantity (population density) and quality (nutrients).

Key Words: exponential growth phase, regression model, two-stage least squares method (2SLSM).

Introduction. The intensification of aquaculture production using new techniques and the cultivation of species with the highest growth potential has led to the need for massive production of live food, especially microalgae. It is a feed of vital importance for the cultivation of mollusks, crustaceans, fish, and various zooplankton species, constituting the preferred food in the early stages of growth, and as primary producers, the base of food chain (Prieto et al 2006). Therefore, the development and survival of the larvae depend on the presence of organisms that make up phytoplankton and zooplankton, due to their nutritional composition, as well as their size and shape, which allow their easy capture and digestion (Sipaúba-Tavares & Rocha 2003).

Sipaúba-Tavares and Rocha (2003) found that the *Ankistrodesmus gracilis* microalgae can be used on a large scale for zooplankton food and fish larvae since it has exponential growth within a period of 8 to 15 days of culture. It has a high nutritional value and excellent results for larvae growth (Sipaúba-Tavares et al 2009), considering its resistance to adverse conditions and its nutritional quality (Sipaúba-Tavares et al 2017).

Commercially, microalgae are cultivated in ponds, whose size varies according to the dimensions of the raceways, between 1 and 5 m^2; they are autotrophic organisms, and they are growing outdoors, where they are diluted for maximum productivity between 0.5 - 0.7 g.L^{-1}, requiring intensive use of energy for the harvest (Travieso et al 2001).

Nowadays, photobioreactors (PHBR) are closed systems with different designs, have been widespread among massive microalgae culture systems, which maximize the growth of the microalgae (Contreras et al 2003). PHBR are characterized by the regulation of most of the population growth parameters (Huarachi-Olivera et al 2015), which allow control of cultivation conditions (Moheimani et al 2012), with higher biomass yield (Wang et al 2013) in axenic monocultures (Ruiz 2011). Studies carried out by Matabanchoy et al (2020) indicate that the helical tubular photobioreactor (HTPHBR) type is the most suitable for the cultivation of *Chlorella* sp., and possibly for other species.

Current techniques recommend the most appropriate time to harvest the algae at the end of the exponential phase (Bold & Wynne 1985; Votolina et al 2000). At this point, maximum cell concentration is obtained, with the highest amount for the organisms for feed. *A. gracilis* is commonly harvested between 12 and 13 days of cultivation (Sipaúba-Tavares & Pereira 2008).

However, the importance of microalgae as food lies in its nutritional value, where raw energy is one of the most important nutrients, as it is the source of all cellular activity, consisting of the energy supplied by proteins, lipids and carbohydrates and their presence in algae increases as cells grow and their concentration in the population will be higher as it multiplies (Marshall 1987). Ethereal extract (EE) comprises the fats that the algal cells contain and are an important source of energy; a decrease in protein availability generally occurs when EE increases (Neumann et al 2018).

On the other hand, fiber reflects a fraction of materials derived from the cell wall that increases as the plant grows or ages (Morris et al 1999) and the microalgal proteins provide poor digestibility or availability if fiber content increases, when plants are mature (Quintana et al 1999). Fiber constitutes the insoluble organic residue, generally considered as carbohydrate not available in a diet, the content of which begins to increase in the middle of the exponential phase (McDermid et al 2005). The chlorophyll contents also vary according to the age of the plant and some of these nutrients, it is of vital importance in the feeding of fish larvae, which can decrease as fiber content increases.

In this way, nutritional value of the algae varies depending on culture conditions, composition of the medium, growth phase, size of the cell, and digestibility (Fabregas & Herreros 1986). Protein levels can be improved by harvesting in the exponential phase since this is where there is greater use of nutrients and a greater process of photosynthesis (Sipaúba- Tavares & Rocha 2003).

Therefore, maximum density is not sufficient criteria for harvesting, since the nutritional value may not be as indicated, given the increase in the natural lignification of the cell walls. So, it should be sought that these variables are found in a reasonable balance to obtain maximum population, maximum nutrient content, and minimum fiber content.

In situations like these, where it is necessary to establish causal or dependency relationships with the intervention of multiple variables that present bilateral feedback, models of simultaneous equations can be an optimization alternative (Gujarati 2006). In a system of various equations, each one represents a model linear regression, whose dependent variable in one equation also acts as an explanatory variable in another. However, this causes endogeneity problems, in addition to being involved in a time series with a serial autocorrelation (Novales 1993), which give rise to biased and inconsistent estimators when applying predictive models. Consequently, there is a need to solve these problems by estimating the regression parameters using the two-stage least square method (2SLSM) (Londoño 2005).

The present work proposes the use of the simultaneous equations method to predict the most suitable moment for the harvest of the *Ankistrodesmus gracilis* microalgae grown in an HTPHBR, to maximize population growth and the availability of nutrients while minimizing fiber content, optimizing the balance between quantity and quality.

Material and Method

Location, laboratory, and equipment. The cultivation of the microalgae was carried out at the Aquaculture Laboratory of the University of Nariño, located to the north-west of the city of San Juan de Pasto (Colombia), between the geographical coordinates 1° 12' 52.48" north, and 77° 16' 41.22" west, at the height of 2,486 meters above sea level and an average ambient temperature of 13.8°C (Imués-Figueroa et al 2018). The research took place during the second semester of 2019.

An HTPHBR of 45 L volume was used, built-in crystal hose of one inch in diameter and 56 m in total length, wound on electro-welded mesh with a diameter of 0.4 m, and useful height of 1.53 m, powered by an airlift pump. The light was provided by six 22-watt, 2,400 lumen white-light LEDT8 lamps, conditioned with the necessary equipment to maintain the stability of temperature, air, and CO_2 supply, volume, flow, and speed of the fluid.

Biological material, cultivation protocol and laboratory analysis. The *A. gracilis* strain comes from the Aquaculture Laboratories of the University of Nariño, where it is kept in pure form. An inoculum of this strain was carried out and raised to a volume of 3 L and to be sown in the HTPHBR with 45 liters of previously treated water.

During the cultivation, Guillard's F/2 (Nasler) nutrient was supplied in a ratio of one milliliter for each liter of the fluid, cultivating by applying the protocol developed in the Laboratory, adapted from Food and Agriculture Organization (Helm & Bourne 2006).

The cultivation was carried out for 15 days, in three different periods. Daily, one-liter samples were taken to obtain the necessary biomass to tri-measure the variables: population density, protein, ether extract, energy, fiber, and chlorophyll-a; the volume was replaced immediately. Bromatological analyzes were carried out in the Bromatology Laboratories of the University of Nariño.

Statistical analysis. Variables were classified into endogenous and exogenous, adopting the following criteria:

Endogenous variables that describe the characteristics of algae, which have to do with the concentration of nutrients present:
Y_{1t} = protein content (percentage).
Y_{2t} = gross energy ($kcal.kg^{-1}$).
Y_{3t} = ether extract (percentage).

Exogenous variables are independent variables that influence the concentration of nutrients in algae:
X_{1t} = population density ($cell.ml^{-1}$).
X_{2t} = chlorophyll-a content ($pg.L^{-1}$).
X_{3t} = fiber content (percentage).
T = cultivation days

With the data obtained, corresponding to the study variables, exploratory analyzes were performed with measures of central tendency and dispersion, as well as statistical graphs, to determine the data´s behavior. Linear correlation analyzes were carried out to identify the variables that will be included in the regression models since the existence of dependency implies the existence of association (Solarte et al 2009). With this, the theoretical models were formulated, and the parameters were estimated, following the procedures proposed by Hausman (1983) and Gujarati (2006). Statistical analyzes were performed using Statgraphics Centurion XVI software.

Results and Discussion

Exploratory analysis. A detailed analysis of the diagrams of Figure 1-A, in 15 cultivation days, show an exponential increase phase until the 12th or 13th day in all variables, then a stationary phase and a decrease phase. Fiber behavior is different, since it increases from the middle of the exponential phase to end of period, in concordance with McDermid et al (2005); this increase in fiber influences the decrease in the availability of various nutrients (Quintana et al 1999); therefore, chlorophyll-a, energy, EE and protein decrease from day 10, are suffering the effects of the significant increase in fiber.

These arguments led to include in the study the data corresponding to the period up to the point of maximum population growth (Figure 1-B), in which it is the harvest point with the maximum biomass amount. The decrease phase may cause the association and dependency relationships between the different variables to be nonsignificant, as they do not adjust to a linear relationship, distorting the dependency relationship. Furthermore, behavior beyond that point is not of interest in plankton production.

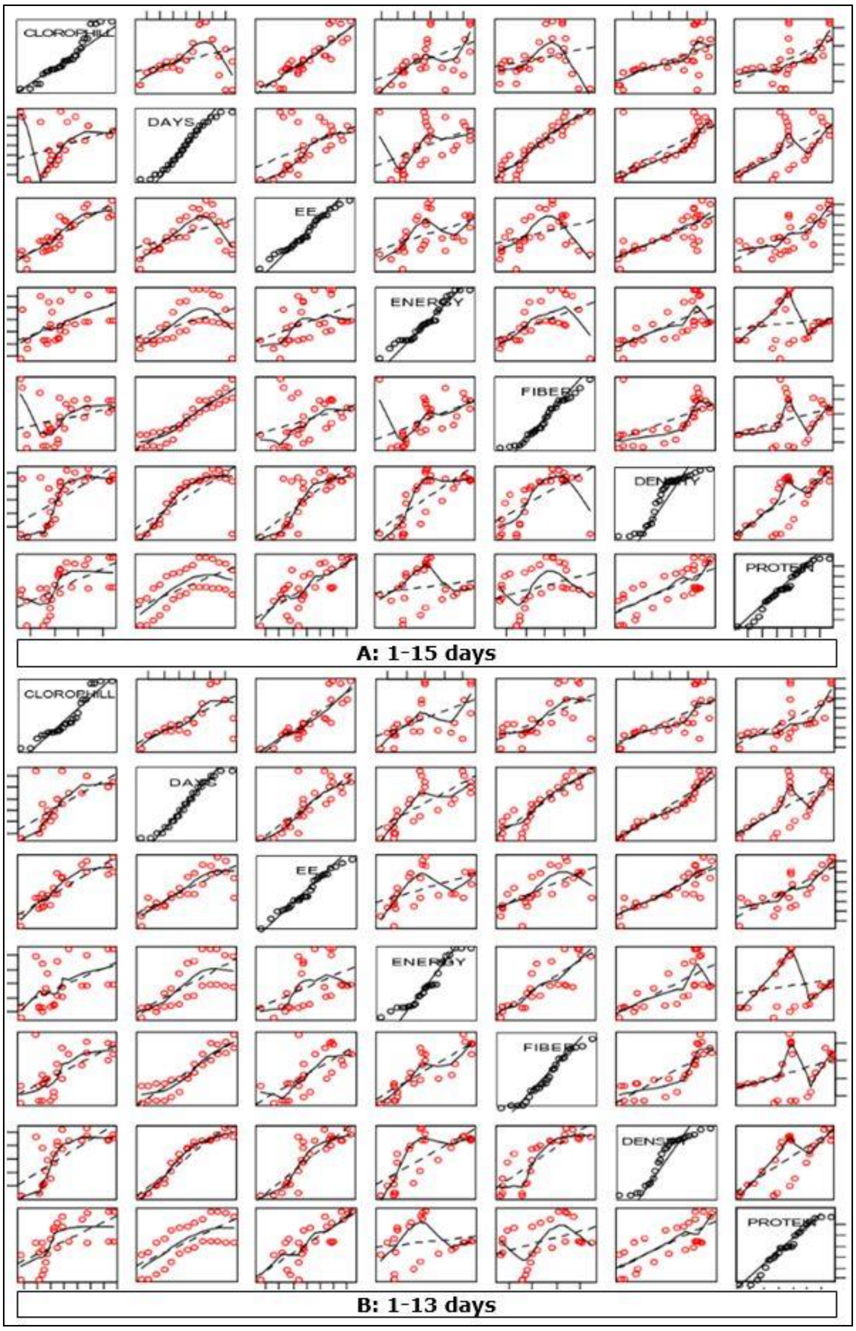

Figure 1. Graphic analysis of the study variables, of *Ankistrodesmus gracilis* microalgae cultured in periods of 15 (A) and 13 (B) days.

In 13 days of cultivation, it is observed that the variables adjust more to a straight line, which guarantees a more significant adjustment in the linear relationship between variables. As already it is mentioned, the only variables with different behavior are fiber, which increases markedly from the middle of the exponential phase; in addition, protein and energy decrease in availability from day 10; possibly they suffer the most from the negative influence of the increase in fiber.

The respective statistics, for 13 days of algae growth, with a total of 117 data per variable (Table 1), show low variability in endogenous variables (protein, ethereal extract, and energy) and high in exogenous ones (chlorophyll-a, fiber, density).

Table 1

Descriptive statistics for the study variables, in 13 days period of *Ankistrodesmus gracilis*

Statistics	*Protein* %	*Energy* *kcal*	*EE* %	*Cells density* $\times 10^5$	*Chlorophil-a* *pg*	*Fiber* %
Mean	43.8312	5166.3672	26.1308	504.3762	1282.3115	9.6850
Typical error	0.3704	61.3678	0.1760	43.7102	40.5745	0.5659
Median	43.8450	5140.8963	25.8800	573.4600	1233.8500	9.1750
Standard deviation	1.8888	312.9155	0.8974	271.8888	206.8903	3.3757
Coeff. of variation	0.0431	0.0606	0.0343	0.5391	0.1613	0.3486
Kurtosis	-0.6425	-1.0861	-0.7825	-1.3548	-0.4007	-1.1937
Asymmetry	-0.2407	0.4869	-0.0166	-0.4954	0.7288	0.1693

The asymmetry values are close to zero in all the variables; the distribution is slightly platykurtic, with the consequent approximation to a normal distribution, beneficial for the later statistical processes.

An analysis of partial correlations (Montgomery & Runger 2003) indicates the existence of a linear association relationship between all the variables, by presenting a Pearson correlation coefficient between medium and high, with a high significance ($p<0.05$). Those indicate that their relationship depends on the joint action between the different variables, which guarantees the viability of multiple regression models.

Preliminary multiple linear regression analysis. Before estimating the models, tests were performed for the statistical assumptions of normality, homocedasticity, and independence. For the first assumption, the hypothesis was denied for the variables of days and density, in which case, there was a need to transform the data using a square root. The homoscedasticity hypothesis was denied for protein, EE and fiber, and the transformation of these variables must be carried out using a natural logarithm. In the case of independence´s assumption, it is not fulfilled because it is a time series, and its solution is the reason for this study.

The theoretical models with which the regression analysis was performed are the following:

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

Where: Y_{1t} (protein content) at time t depends on the value of β_{10} (intercept), X_{1t} (population density), X_{2t} (chlorophyll-a content), X_{3t} (fiber content), Y_{2t} (gross energy concentration), Y_{3t} (fiber content), T (cultivation days) and error (u_{1t}), at time t.

Similarly, Y_{2t} (gross energy concentration) at time t depends on the value of β_{20} (intercept), X_{1t} (population density), X_{2t} (chlorophyll-a content), X_{3t} (fiber content), Y_{1t} (protein content), Y_{3t} (ethereal extract), T (cultivation days) and error (u_{2t}), at time t.

In the case of Y_{3t} (ethereal extract) at time t, it depends on the value of β_{30} (intercept), X_{1t} (population density), X_{2t} (content of chlorophyll-a), X_{3t} (fiber content), Y_{1t} (protein content), Y_{2t} (gross energy), T (cultivation days) and error (u_{3t}), at time t. Table

2 shows the analysis of variance for the models, corresponding to each of the regressive variables.

Table 2

Estimation of linear regression models of the simultaneous equations system (α = 0.05)

	Model 1: Y_{1t} (Protein)			*Model 2: Y_{2t} (Energy)*			*Model 3: Y_{3t} (EE)*		
Analysis	*F p-value*	R^2	*Durbin-Watson p-value*	*F p-value*	R^2	*Durbin-Watson p-value*	*F p-value*	R^2	*Durbin-Watson p-value*
ANOVA	0.0000	93.903	0.0037	0.0000	91.2306	0.0232	0.0000	94.9507	0.0443
Regression analysis									
Variable	$\hat{\beta}$	*T p-value*	*F p-value*	$\hat{\beta}$	*T p-value*	*F p-value*	$\hat{\beta}$	*T p-value*	*F p-value*
Intercept	-0.684611	0.1208		-10807.7	0.0043		-0.926226	0.0005	
X_{1t} Density (A)	0.00323049	0.2820	0.0000	80.5972	0.0008	0.0000	0.0052537	0.0056	0.0000
X_{2t} Chlorophil-a (B)	0.00003271	0.3542	0.0089	0.895092	0.0019	0.6751	0.0000749072	0.0002	0.0000
X_{3t} Fiber (C)	-0.102976	0.0018	0.0000	199.374	0.5823	0.0000	0.0574523	0.0208	0.1501
Y_{1t} Protein (D)				-4593.79	0.0471	0.0000	0.159935	0.3799	0.0027
Y_{2t} Energy (E)	-0.00005175	0.0471	0.0000				-0.000042712	0.0163	0.0029
Y_{3t} EE (F)	0.323547	0.3799	0.5775	-7670.02	0.0163	0.0065			
T Days (G)	0.0688168	0.0046	0.0046	-27.5395	0.9160	0.9160	-0.0343708	0.0612	0.0612
Selection of the best model									
	R^2	CP-Mallow	Var.	R^2	CP-Mallow	Var.	R^2	*CP-Mallow*	*Var.*
	93.5703	5.81856	ABCEG	91.224	5.01151	ACDFG	94.6751	5.81856	ABDEG

It is evident that all the estimated models are significant as a whole ($p<0.05$) and show a high coefficient of determination, greater than 91%, although with a non-significant hypothesis of independence of error (Durbin-Watson) ($p<0.05$), since it is a temporal series, with autocorrelated data.

In Model 1, although the EE value is high, it is not significant ($p>0.05$), which may be due to lipids are an important source of energy (Marshal 1987). But it is not the only one since energy is produced by various components such as proteins, carbohydrates, and lipids (Fabregas & Herreros 1986). Therefore, it can be more important, with a greater influence on protein content.

In Model 2, the linear regression analysis (T statistic) shows that the fiber and days variables do not have a significant individual effect ($p>0.05$) on energy. The most significant influence is caused by EE and protein, according to the statements of Fabregas and Herreros (1986). They state that the presence of protein in algae increases as cells grow and its concentration in the population will be higher than as it multiplies and decreases as the cell walls' lignification increases.

In the case of Model 3, EE seems not to be influenced by fiber; however, in concordance with it expressed by Gómez (2007). He considers that EE increases when greater chlorophyll activity increases as the population reproduces, and because it is an energy source, as this nutrient increase, the EE is increased.

Based on the previous analyzes, considering the coefficient of determination and the CP-Mallow statistic, the Simultaneous Equation System was structured as follows:

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

Where: β = parameters of exogenous variables; γ = parameters of endogenous variables; Y_{1t} = protein content; Y_{2t} = concentration of energy; Y_{3t} = ethereal extract content; X_{1t} = population density; X_{2t} = chlorophyll-a content; X_{3t} = fiber content; T = cultivation days; $u_{.t}$ = error associated with the t-th observation.

Autocorrelation analysis. Given the existence of a time series, to determine the serial autocorrelation of the endogenous variables Durbin-Watson test was performed between the variables and their first lag considering an autoregressive process of order one AR(1), to verify if Corr($u_t u_{t-1}$)=0. The results are shown in Table 3, assuming H_1: $\rho>0$, where the existence of autocorrelation in the protein variable is observed, as in EE ($p<0.1$), not for energy ($p>0.1$).

Table 3

Test of autocorrelation by *d* Durbin-Watson statistic

Dependent variable Y_t	*Independent variable Y_{t-1}*	R^2	R^2 *adjusted*	*d-value Durbin-Watson*	*p-value*
Protein	Protein (lagged)	0.9689	0.9675	1.4401	0.04927
Energy	Energy (lagged)	0.9284	0.9251	1.4728	0.30420
Ethereal extract	Ethereal extract (lagged)	0.7020	0.6885	1.8673	0.06440

Concerning the autocorrelation functions, for exogenous variables, according to Peña and Marín (2016), determining the order of an autoregressive process based on its simple autocorrelation function presents difficulties. Since this is a mixture of exponential growths and sinusoidal, which are cushioned as the delay progresses and do not have readily identifiable features to determine the order of the process, so it must be supported by the partial autocorrelation function, since there is a direct effect of y_{t-2} on y_t.

Thus, in the case of the energy variable, with a Durbin-Watson statistic not significant for the AR(1) process, when the simple autocorrelation function is observed, an AR(2) process appears to follow, with the first positive parameter ($\Phi_1 > 0$) and the second negative ($\Phi_1 < 0$), corroborated by the partial autocorrelation function, for which the Durbin-Watson statistic (1.4182) for AR(2) was estimated, assuming autocorrelation Y_{2t} over Y_{2t-2} considered significant (p=0.05294). For EE variable, the autocorrelation function seems to indicate an AR(1) process, with $\Phi_1>0$ and $\Phi_1<0$, and the analysis of partial autocorrelation function leads to the conclusion that it may be an AR(1) process with $\Phi_1>0$, clarifying the doubt of autocorrelation using the Durbin-Watson statistic. In the protein variable, an AR(1) process is visualized, with $\Phi_1>0$.

The arguments raised above lead to the need to include some lagged variables in the model, like this: Y_{1t-1} (protein with a lag of one period), Y_{2t-2} (energy with a lag of two periods) and Y_{3t-1} (ethereal extract with a lag of one period). In this way and following Mallow's CP recommendation for the best model, considering that there is a small decrease in R^2, not significant, a multi-equation system with adjusted models is proposed. In addition, all the variables included have a trending behavior and, according to what was exposed by Wooldrige (2010), the temporal trend is considered exogenous. Therefore, the model must indicate the parameters of the variables lagging with β, in the following system:

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

Where: β = parameters of exogenous variables; γ = parameters of the endogenous variables; Y_{1t} = protein content; Y_{1t-1} = protein content, with a lag of one period; Y_{2t} = concentration of energy; Y_{2t-2} = concentration of energy, with a lag of two periods; Y_{3t} = ethereal extract content; Y_{3t-1} = ethereal extract content, with a lag of one period; X_{1t} = population density; X_{2t} = chlorophyll-a content; X_{3t} = fiber content; T = days of cultivation; $u_{.t}$ = error associated with the t-th observation.

As shown in these models, due to the Y_{1t-1} and Y_{2t-2} delays, the relationship between protein content and gross energy concentration is not contemporary, becoming dynamic models (Gujarati 2006). They imply changes over time because the effect of a unit variation in the value of the delayed explanatory variables is felt for several periods (distributed delays), with the effect distributed over several periods.

Identification of equations and instrumental variables. Given the system of previously defined equations, where $E(u_{jt}) = 0$ and u_{jt} do not correlate with X_{1t}, X_{2t}, X_{3t} or T, determined in the analysis of partial correlations, as explained by Wooldrige (2010), the temporal trend is exogenous, consequently u_{jt} also does not correlate with Y_{1t-1}, Y_{2t-2} or Y_{3t-1}. Since, Y_{1t}, Y_{2t}, and Y_{3t} are endogenous explanatory variables, they may be correlated with u_{jt}. A problem of identification of the equations arises, through the order and rank conditions, following the procedure recommended by Maddala (1996). With which shows that, by order condition, all the equations are overidentified, since the number of excluded endogenous and exogenous variables is greater than the total number of endogenous variables minus one.

Based on this condition, the first system equation (protein) presents two exclusion restrictions (Y_{2t-2} and Y_{3t-1}) that will be used as instrumental variables for Y_{2t}. Therefore, the equation satisfies the order condition, with two overidentified restrictions, since there is a single endogenous variable on the right side of the equation, and two instrumental variables are available for estimation.

The second equation (energy) also satisfies the order condition, since it has two exclusion restrictions (Y_{1t-1} and Y_{3t-1}) that can be used as instrumental variables for the estimation of Y_{1t} and Y_{3t}, endogenous variables included in the side right of the model, clarifying that there is a single overidentified restriction.

Similarly, in the third equation (EE), three exclusion restrictions (X_{3t}, Y_{1t-1}, and Y_{2t-2}) are observed as instrumental variables to estimate an endogenous variable (Y_{2t}) included on the right side of the equation, for which reason satisfies the order condition with two overidentified constraints.

According to Wooldrige (2010), unless there is an obvious identification failure, equation that satisfies range condition is considered identified, and therefore the method is necessary and sufficient. However, the order condition is necessary to establish the degree of identifiability, but it is not sufficient. For this, it is essential to establish a condition $T \geq 0$, since there cannot be a negative time; on the other hand, all $\beta_{ij} \neq 0$ could be required in the same equation, to ensure identification using the order condition; but this can be taken for granted since it is practically impossible for all $\beta_{ij} = 0$, in which case they may present unidentified equations. Since all equations in the system are overidentified, with sufficient exclusion restrictions, a case of non-identification would be unlikely to occur.

Two-stage least squares method (2SLSM). The consistent estimation of the structural parameters will now be sought through the 2LSLM method, recommended by Gujarati (2006) for these cases. Since the reduced form equations, as presented above, are optimal predictors; they can even be functions of conditional hope. Consistent estimators can be obtained, from identified structural equations, whether or not they are overidentified, which could be estimators unbiased, through the linear regression of Y_{it} on the variables X_{it} and the laggards Y_{1t-1}, Y_{2t-2} and Y_{3t-1} (which are considered exogenous).

According to this method, the first stage consists of obtaining estimators using ordinary least squares (OLS), for the equations of the reduced form previously proposed and, from these, obtaining the corresponding adjusted values, which seeks to eliminate the correlations that they probably exist between Y_{it} and u_{it}. With this, the following estimated models were obtained:

$$\hat{Y}_{1t} = -0.142112 + 0.000998772X_{1t} + 0.0000132628X_{2t} - 0.0105273X_{3t} + 0.849833Y_{1t-1} - 0.000013524Y_{2t-2} - 0.0228623Y_{3t-1}$$

$$\hat{Y}_{2t} = -5462.9 + 37.38X_{1t} + 0.205352X_{2t} + 329.908X_{3t} - 2049.39Y_{1t-1} + 0.468658Y_{2t-2} - 4885.35Y_{3t-1}$$

$$\hat{Y}_{3t} = -0.92194 + 0.00376819X_{1t} + 0.000081651X_{2t} + 0.00343195X_{3t} - 0.0631164Y_{1t-1} - 0.0000402846Y_{2t-2} + 0.303424Y_{3t-1}$$

ANOVA for the estimated models indicates that Ho: $\beta = 0$ should be rejected, since $p<0.01$, in all cases, therefore, it is concluded that there is a linear dependency relationship between return variables and dependent variable, of each model.

Furthermore, R^2 and R^2-adjusted values are high, which indicates that estimated values of endogenous variables are very close to their observed values. Therefore, the latter is less likely to be correlated with stochastic disturbances in the original structural equations (Maddala 1996).

The second stage consists of replacing the endogenous variables of the second member with the estimated variables and estimating the equations by OLS. Therefore, the 2SLSM method allows all the endogenous predictor variables in the equation to be replaced by their predicted values from the reduced forms, and the OLS method is reapplied, which estimates are consistent as the sample size increases indefinitely (Montgomery & Runger 2003). The final models estimated with this methodology are the following:

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

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

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

Where: $\hat{Y}_{1t}^{*}$ = estimated protein content, transformed by natural logarithm; $\hat{Y}_{2t}^{*}$ = concentration of energy; $\hat{Y}_{3t}^{*}$= ethereal extract content, values transformed by natural logarithm; X_{1t} = population density, transformed by square root; X_{2t} = chlorophyll-a content; X_{3t} = fiber content, transformed by natural logarithm; Y_{1t-1} = protein content, transformed by natural logarithm, with a lag of one period; Y_{2t-2} = concentration of energy, with a lag of two periods; Y_{3t-1} = EE content, with a lag of one period; $\hat{Y}_{1t}$ = estimated values of protein content according to the reduced equation of the first stage of 2SLSM; $\hat{Y}_{2t}$ = estimated values of energy concentration by regression with the reduced model, in the first stage of 2SLSM; $\hat{Y}_{3t}$ = estimated values of protein content according to the reduced equation of the first stage of 2SLSM; T = days of cultivation; values transformed by a square root.

From the analysis of the system of actions in global form, it is possible to see the positive effect of the population and the content of chlorophyll-a on the endogenous variables (protein, energy, and EE); fiber content has a negative effect on protein concentration, which is consistent with what was expressed by Peña and Marín (2016), referring to the decrease in the former when the latter increases. The lags (AR1) of the variables protein and EE have a positive effect, indicating that their concentration increases as time passes. In the case of energy (AR2), it has a negative influence on protein; energy and EE have a negative effect on energy and protein, and energy show a negative effect on EE content, indicating that the interaction between endogenous variables will always be inversely proportional.

The model for $\hat{Y}_{1t}^{*}$ (protein) shows a high coefficient of determination (R^2 = 98.1574%) and a low standard error (0.00612418), indicating a good fit for Y_{1t}. Furthermore, non-significant Durbin-Watson statistic ($p>0.05$) ensures the absence of self-correlation of errors, therefore independence.

ANOVA indicates that there is a linear relationship between the return variables and the dependent variable ($p<0.01$). The regression analysis for the individual significance of the return variables (T-Statistic) shows that the only variable that has a significant effect on the model is protein when it is included with a lag of one period. However, in the analysis according to the adjustment order (F-Statistic), which contains the variables conditioned to the presence of others in the model, it is found that the variables energy and days of cultivation have no significant influence on the model.

Despite this, the complete model was adopted, for having an equally high determination coefficient (97.9395%) and for being the one with the best performance according to the CP-Mallow criterion. In addition, they are considering the reasons given by Sipaúba-Tavares (1994) and Marshal (1987). They affirm that there is generally a decrease in protein in cell when the plant ages, and there is an increase in the content of ethereal extract, a fundamental source of energy.

The second regression model for $\hat{Y}^*_{2t}$ (energy) represents a linear relationship since there is sufficient evidence to reject the null hypothesis Ho: $\beta = 0$ ($p<0.01$), with a good fit when presenting a coefficient of high determination (96.3255%) and low standard error (70.3253). When performing the regression analysis of the individual effect of the return variables, it is possible to establish that fiber, energy (AR2) and days of cultivation are not significant ($p>0.05$); however, in the analysis in order of adjustment, chlorophyll-a and days of culture are without significant effect ($p>0.05$).

Although the conditional contribution of chlorophyll-a is not significant, its individual importance it is, and this justifies its presence in the model. On contrary, fiber and energy (AR2) do not show a significant contribution individually, but when the conditional presence is studied, its effect on energy shows statistical significance.

Taking into account CP-Mallow criterion, a model is proposed that includes all the variables, which has the best values of R^2 (95.8966%) and MSE (4832.57), supported by Sipaúba-Tavares (1994), who ensures that the gross energy is made up of the energy supplied by the protein, lipids and carbohydrates and its presence in the algae increases as the cells grow and its concentration in the population will be higher as it multiplies, explanations which led to the adoption of the complete model initially proposed.

The third estimated model $\hat{Y}^*_{3t}$ (EE) represents a significant linear dependency relationship ($p<0.05$), with a good fit ($R^2 = 94.7625\%$) and a low error (0.0000627533). Although EE(AR1), protein, energy, and days of culture do not seem significant ($p>0.05$) when the individual effect is analyzed, the joint adjustment order indicates that only days of culture may not have a significant effect on the EE. It is possible that this happens in all the variables, given the behavior of the nutrients over time, which increase or decrease according to the joint effect of the other variables. In any case, a dynamic time series is being treated that is subject to the growth of the plant in a growing period, as stated by Gómez (2007) and has been demonstrated through the present study.

On the other hand, it is necessary to consider that the ethereal extract comprises the fats that the algal cells contain and are an important source of energy (Sipaúba-Tavares et al 2017). These contents vary according to the growth phase (Gómez 2007). Based on these arguments and based on the test for selecting best model, according to CP-Mallow criterion (5.77927), all the variables initially proposed were included.

Optimal harvest time. Based on the previously estimated models, for the multiple equation system, prediction was made for each of the particular Y_{it} points, from which it was possible to determine the optimal harvest time for the microalgae (Table 4).

Table 4

Forecasted values for the particular optimal point Y_{it}, for *Ankistrodesmus gracilis* culture in helical tubular photobioreactor

Variable	*Estimated value*	*Initial value*
Cultivation days	11	13
Chlorophyll-a (pg.cell^{-1})	0.1674	0.1282
Fiber (%)	8.77	9.6850
Density (cells×10^5 ml^{-1})	475.57	504.3762
Ethereal extract (%)	27.5816	26.1308
Protein content (%)	46.7142	43.8312
Gross energy (kcal.kg^{-1})	5222.49	5166.3672

According to this estimate, the optimal time of cultivation at which the harvest should be carried out has been estimated at 11 days. At that point, it is possible to have the maximum concentration of nutrients in protein (46.7142%), gross energy (5222.49

kcal/kg), and ethereal extract (27.5816%) and the lowest fiber content (8.77%). Although the population density is lower (475.57 cells$\times10^5$ ml^{-1}), it is possible to have an adequate biomass. This contrasts with the arguments of Gómez (2007), Sipaúba-Tavares and Rocha (2003) and Sipaúba-Tavares et al (2009), who carried out the harvest at 13 days that is when the exponential phase ends, arguing that it is the moment in which it has the highest population concentration. The decrease in two days results in the best quality of the microalgae as food, reducing production costs; in addition, to optimizing the quantity and the supply time.

Conclusions. All the variables show a phase of exponential increase, which increases until day 12 or 13, from there a stationary phase appears and then another one of descent, except for the fiber, which begins a strong increase from the middle of the exponential phase until the end of the period.

The models of simultaneous equations to describe the cultivation of the algae *Ankistrodesmus gracilis* should include the endogenous return variables protein and EE with one period lag and energy with a two-period lag, with a significant effect on the models.

By order condition, all the equations of the system are overidentified, which is corroborated by the rank condition as a necessary and sufficient condition.

The estimated models of the simultaneous equation system, using the 2SLSM, for the cultivation of the microalgae *Ankistrodesmus gracilis*, allow optimizing the harvest time, according with its nutritional value, which is not possible with current methods.

There is no clarity about the significance of the effect of cultivation time on endogenous variables, perhaps because its behavior corresponds to the combined influence of all variables under study. Its inclusion is maintained given the type of the equation system that is immersed in a temporary series.

The procedure allows us to see the negative effect of the fiber content on the nutrients, especially on the protein; additionally, the interaction between the endogenous variables will always be inversely proportional.

It is important to carry out similar studies, using the 2SLS method, in the cultivation of marine and freshwater microalgae species from other regions of the world, including large samples, with repeated data and other quality variables, such as lipid profiles.

Day 11 showed to be the optimal moment for the harvest of the microalgae *Ankistrodesmus gracilis* so that maximizes the concentration of nutrients, especially protein (46.7142%) and gross energy (5222.49 kcal.kg^{-1}) and minimizes the fiber content (8.77%), with an adequate population density (475.57 cellls.ml^{-1}).

Acknowledgements. To the Vice Chancellor of Research and Social Interaction of the University of Nariño (Colombia), for funding this research.

Conflict of interest. The authors declare that there is no conflict of interest.

References

Bold H. C., Wynne M. J., 1985 Introduction to the algae: structure and reproduction. Prentince Hall, New Jersey, 586 p. ISBN:9780134777467.

Contreras C., Peña J. M., Flores L. B., Cañizares R. O., 2003 [Advances in the conceptual design of photobioreactors for the cultivation of microalgae]. Interciencia; 28(8):450-456. ISSN 0378-1844 [in Spanish].

Fabregas J., Herrero C., 1986 Marine microalgae as a potential source of minerals in fish diets. Aquculture, 51(3-4):237-243.

Gómez L., 2007 [Microalgae: ecological and biotechnological aspects]. Revista Cubana de Química XIX (2):3-20. ISSN: 0258-5995 [in Spanish].

Gujarati D. N., 2006 [Principles of econometrics, 3rd edition]. McGraw-Hill, Madrid, 546 p. ISBN:84-481-4632-8 [in Spanish].

Hausman J., 1983 Chapter 7 Specification and estimation of simultaneous equation models. Handbook of Econometrics, 1:391-448.
Helm M., Bourne N., 2006 [Bivalves hatchery culture: a practical manual. Fishing Technical Document 471]. FAO, Roma, 184 p. ISBN: 92-5-304224-4 [in Spanish].
Huarachi-Olivera R., Yapo-Pari U., Dueñas-Gonza A., González-Juárez R., Condori-Huamanga J., Pacheco-Salazar D. G., Soto-Flores J., 2015 [Adaptability of *Spirulina* (Arthrospira) *platensis* (Cyanophyta) in conical tubular photobioreactor under ambient conditions]. Idesia, 33(1):103-112 [in Spanish].
Imués-Figueroa M. A., Torres-Valencia G., Sanguino-Ortiz W., Chapman F., 2018 [Simple growth rate (SIG) of the rotifer *Brachionus calyciflorus* with different food types]. Investig.pecu., 5(1):17-23 [in Spanish].
Londoño W., 2005 [Multiple equation models, var models and cointegration]. Universidad Eafit, Departamento de Ciencias Básicas, Maestría en Matemáticas Aplicadas, Medellín, Colombia, 114 p [in Spanish].
Maddala G. S., 1996 [Introduction to econometrics, 2nd ed]. Prentice Hall, México, 715 p. ISBN: 968-880-697-8 [in Spanish].
Marshall D., 1987 [Algae biology: physiological approach]. Limusa, México, 236 p. ISBN: 9681821092 [in Spanish].
McDermid K. J., Stuercke B., Haleakala O. J., 2005 Total dietary fiber content in Hawaiian marine algae. Botánica Marina, 48(5):437-440.
Moheimani N., Borowitzka M., Isdepsky A., Sing S., 2012 Standard methods for measuring growth of algae and their composition. In: Developments in applied phycology. (DAPH, volume 5), Borowitzka M. & Cheronet A. (Eds). Springer Science, Australia, pp. 265-284.
Montgomery D. C., Runger G. C., 2003 [Probability and statistics applied to engineering, 2nd edition]. Limusa, México, 937 p. ISBN: 968-18-5915-4 [in Spanish].
Morris H., Quintana M., Almarales A., Hernández L., 1999 [Biochemical composition and evaluation of the protein quality of the autotrophic biomass of *Chlorella vulgaris*]. Rev. cuba. aliment. nutr., 13(2):123-128 [in Spanish].
Neumann U., Derwenskus F., Gille A., Louis S., Schmid-Staiger U., Briviba K., Bischoff S., 2018 Bioavailability and safety of nutrients from the microalgae *Chlorella vulgaris*, *Nannochloropsis oceanica* and *Phaeodactylum tricornutum* in C57BL/6 Mice. Nutrients, 10(8):965.
Novales A., 1993 [Econometrics, 2nd edition]. McGraw-Hill, Madrid, 676 p. ISBN:84-481-0128-6 [in Spanish].
Peña E., Marín H., 2016 [Applied phycology. Seaweed characteristics, uses and cultivation]. Universidad del Valle, Cali, Colombia, 168 p. ISBN 9789587652833 [in Spanish].
Prieto M. J., Mogollón J., Castro A. L., Sierra L. A., 2005 [Effect of the medium and culture conditions on the productivity of three marine diatoms with aquaculture potential]. Revista MVZ Córdoba, 10(1):544-554 [in Spanish].
Quintana M., Hernández L., Morris H., Fernández M., 1999 [Content of some vitamins in cultures of microalgae *Chlorella* sp.]. Cuban Journal of Food and Nutrition, 13(1):9-13 [in Spanish].
Ruiz A., 2011 [Start-up of a microalgae culture for the removal of nutrients from urban wastewater previously treated anaerobically]. Tesis de grado Máster Universitario en Ingeniería Hidráulica y Medio Ambiente. Universidad Politécnica de Valencia, Valencia, España, 96 p [in Spanish].
Sipaúba-Tavares L. H., Ibarra L. C., Fioresi T. B., 2009 *Ankistrodesmus gracilis* (Reisch) Korsikov (Chlorophyta) laboratory cultured in CHU_{12} and macrophyte with NPK media. B. Inst. Pesca, 35(1):111-118.
Sipaúba-Tavares L. H., Lusser-Segali A. M. D., Berchielli-Morais F. A., Scardoeli-Truzzi, B., 2017 Development of low-cost culture media for *Ankistrodesmus gracilis* based on inorganic fertilizer and macrophyte. Acta Limnol. Bras., 29:e5. https://doi.org/10.1590/s2179-975x3916.

Sipaúba-Tavares L. H., Pereira A. M. L., 2008 Large scale laboratory cultures of *Ankistrodesmus gracilis* (Reisch) Korsikov (Chlorophyta) and *Diaphanosoma biergei* Korinek, 1981 (Cladocera). Braz. J. Biol., 68(4):875-883.
Sipaúba-Tavares L., 1994 [Limnology applied to aquaculture]. FUNEP. Boletim Técnico do CAUNESP No. 1, Jaboticabal, SP, Brasil, 71 p [in Portuguese].
Sipaúba–Tavares L., Rocha O., 2003 [Production of plankton (Phytoplankton and Zooplankton) to feed aquatic organisms]. RiMa, São Carlos, Brasil, 106 p. ISBN-10:8586552178 [in Portuguese].
Solarte C., García H., Imués M. A., 2009 [Biostatistics: applications in animal production and health]. Universidad de Nariño, Pasto, Colombia, 134 p. ISBN: 958-9479-39-1 [in Spanish].
Travieso L., Hall D. O., Rao K. K., Benítez F., Sánchez E., Borja R., 2001 A helical tubular photobioreactor producing *Spirulina* in a semicontinuous mode. International Biodeterioration & Biodegradation, 47(3):151–155.
Voltolina D., Nieves M., Piña P., 2000 [Quality of microalgae for aquaculture]. En: Avances en Nutrición Acuícola IV. Memorias del IV Simposium Internacional de Nutrición Acuícola. Noviembre 15-18, 1998. Civera-Cerecedo R., Pérez-Estrada C. J., Ricque-Marie D., Cruz-Suárez L. E. (Eds.). La Paz, B.C.S., México, pp. 28-32 [in Spanish].
Wang J., Han D., Sommerfeld M., Lu C., Hu Q., 2013 Effect of initial biomass density on growth and astaxanthin production of *Haematococcus pluvialis* in an outdoor photobioreactor. Journal of Applied Phycology 25:253–260.
Wooldrige J. M., 2010 [Introduction to econometrics: a modern approach, 4th edition in Spanish]. Traducido por Ma. del Carmen Enriqueta Henao Roa, Érika M. Jasso y Hernan D´Borneville. Cengage Learning, México, 865 p. ISBN-13: 978-607-481-312-8. ISBN-10: 607-481-312-4 [in Spanish].

Received: 24 August 2021. Accepted: 14 December 2021. Published online: 27 August 2022.
Authors:
Marco Antonio Imués-Figueroa, Department of Hydrobiological Resources, Faculty of Animal Sciences, University of Nariño, Pasto, Colombia, e-mail: marcoi@udenar.edu.co
Gustavo Adolfo Torres-Valencia, Department of Hydrobiological Resources, Faculty of Animal Sciences, University of Nariño, Pasto, Colombia, e-mail: gustavotorresvalencia@gmail.com
Jennifer Gómez-Hernández, Program of Aquaculture Production Engineering, Faculty of Animal Sciences, University of Nariño, Pasto, Colombia, e-mail: jenifergomezhernandez@gmail.com

How to cite this article:
Imués-Figueroa M. A., Torres-Valencia G. A., Gómez-Hernández J., 2022 Simultaneous equation method for estimating the optimal moment for harvesting microalgae. AACL Bioflux 15(4):2158-2170.