Black Sea Journal of Statistics
Doi: 10.5281/zenodo.18234708

Open Access Journal
e-ISSN: 3108-4265

Research Article
Volume 1 - Issue 2: 31-35 / December 2025

CURVE MODELLING PROCEDURE IN SAS PACKAGE PROGRAM

Mustafa ŞAHİN[1]*, İsmail GÖK[1], Esra YAVUZ[2]

[1]Kahramanmaraş Sütçü İmam University, Faculty of Agriculture, Departman of Agricultural Biotechnology, 46100, Kahramanmaraş, Türkiye
[2]Şırnak University, Cizre Vocational School, Department of Accounting and Tax, 73200, Cizre, Şırnak, Türkiye

Abstract: Growth, lactation, and yield curves are widely used in animal husbandry. In addition to pedigree records, they are widely used as a supporting tool in animal breeding. Curve models based on the herd average provide an indication of the herd's overall structure, while individual models provide insight into individual development. In this context, they are used as a supporting tool, particularly in the selection and culling stages. Although curve models are widely used, the number of programs available in this field is quite limited. Among these limited programs, the SAS statistical package stands out, as it offers sufficient user intervention and is widely used. This study will cover the basic steps of the curve modeling procedure (data definition, data input, model, curve, and curve feature definition). For this purpose, the curve modeling procedure in the SAS package will be discussed in detail using a numerical example related to animal husbandry, ultimately providing an important resource for researchers working in this field.

Keywords: SAS, Curve, Procedure

***Corresponding author:** Kahramanmaraş Sütçü İmam University, Faculty of Agriculture, Departman of Agricultural Biotechnology, 46100, Kahramanmaraş, Türkiye
E mail: ms66@ksu.edu.tr (M. ŞAHİN)
Mustafa ŞAHİN https://orcid.org/0000-0003-3622-4543
İsmail GÖK https://orcid.org/0000-0002-0759-1187
Esra YAVUZ https://orcid.org/0000-0002-5589-297X

Received: November 03, 2025
Accepted: November 25, 2025
Published: December 15, 2025
Cite as: Şahin, M., Gök, İ., & Yavuz, E. (2025). Curve modelling procedure in SAS package program. *Black Sea Journal of Statistics, 1*(2), 31–35.

1. Introduction

Curve modeling is a modeling technique that fundamentally enables the mathematical representation of geometric shapes and processes. Widely used in computer-aided design, engineering, and graphics applications, this method has also gained significant ground in the modeling of biological systems and the digitalization of agricultural activities in recent years. As the livestock sector embraces data-driven decision-making processes with the digital transformation, mathematical methods such as curve modeling have become increasingly functional in this field.

In animal husbandry, curve modeling can be used for various purposes, such as generating animal growth curves, analyzing milk yield over time, and examining the relationships between feed intake and weight gain. In this context, parametric curve models (e.g., Gompertz, Logistic, and Richards curves) are widely preferred to mathematically describe the biological development processes of animals. Furthermore, these models offer an important tool for making predictions to increase production efficiency, optimizing feeding strategies, and supporting economic decision-making processes (Bayazıt at al., 2022).

SAS software, widely used in curve modeling, is a powerful statistical analysis platform. In particular, procedures such as PROC NLIN, PROC NLMIXED, PROC GAM, PROC TRANSREG, and PROC SGPLOT can address a wide variety of curve modeling needs, from nonlinear regression analyses to generalized add-on models. Using these SAS procedures, users can define parametric or nonparametric curves, evaluate model fit, and make predictions.

This study will introduce the basic principles of curve modeling in the SAS environment and conduct curve modeling on a sample model using SAS procedures.

2. Materials and Methods

2.1. Materials

Data obtained from the forage legume clover using the in vitro gas production technique were used in the study. Gas production values (ml/1 g DM) were obtained at 3, 6, 12, 24, 48, 72, and 96 hours using (Makkar and Blümmel, 1997). The in vitro gas production method and are given in Table 1. The SAS statistical package program was used in the study (SAS, 1999; Goodwin at al., 2018).

Table 1. Gas production values of clover legume forage plant (ml/1 g DM)

Time	Gas Production
3	118.26
6	180.83
12	241.08
24	304.13
48	356.83
72	386.85
96	397.59

2.2. Methods

In this study, a numerical example from the field of animal nutrition is used to illustrate the curve modeling procedure in the SAS statistical package. The Orskov model (Orskov and McDonald, 1979; dos Santos Cabral at al., 2019; Panik, 2010; Panik, 2013), one of the first equations developed for estimating feed digestibility, is used in the modeling of this numerical example. This model is given in the following equation (1).

$$Y = a + b\,(1 - e^{-ct}) \quad (1)$$

In the equation,

a: represents the initial amount of gas or the amount of digestion,

b: represents the slowly produced gas or the amount of digestion,

c: represents the gas production rate or the rate of digestion,

t: represents the incubation time, and

Y: represents the gas production at the t-th time. Here, "a+b" represents the total amount of gas produced throughout digestion (Orskov and McDonald, 1979; Özkan and Kamalak, 2023).

Implementing any model in the SAS statistical package requires various steps. These include variable definition, data input, statistical model definition, and graphical definition (Beyazıt at al., 2022).

2.2.1. Variable definition

This first stage consists of defining the data set name, data type, data input order, data input method, data precalculation, and data label definition. Variable definition is outlined below.

Data ...;
İnput ... , ... , ... , ...;
Title "………………";
IF/Then/Else;
Drop / Keep;
Rename;
Label;
Cards; or Datalines;

Data: This command is used to name the data set.

Input: This command assigns data to variables. At least one space must be created between variable names. Variables are of two types: numeric and non-numeric. Non-numeric variables must be prefixed with a "$" sign. Variables without this character will be considered numeric variables. Placing the @@ symbol at the end of the input command line indicates that data input is made side by side according to the variable definition.

Title: This command determines the title of the SAS output report.

IF/Then/Else: This command sequence is used to define conditional operations before analysis.

Drop/Keep: This command is used to determine which variables in the data set will be kept or discarded.

Rename: This command is used to change the previously specified variable name.

Label: This command is used to define descriptive labels for variables.

Cards or datalines: This command indicates that the variable definition phase is complete and that data input will begin.

Data input:

This stage consists of input data in accordance with the definitions made in the input variable definition stage. Data input must be carried out in accordance with the variable definition stage. In other words, the data input stage is entirely dependent on the variable definition stage (Korosteleva, 2018).

Statistical model definition:

This stage defines the equations to be used in curve modeling. Symbols such as =, *, /, −, +, and ^ are commonly used in defining equations. Correct model definition is crucial for obtaining statistically significant curves. Furthermore, obtaining accurate outputs such as model parameter estimates and error terms depends on the correct definition of the model. The main headings are briefly outlined below.

Proc nlin; Starts the nonlinear regression procedure. In other words, if the relationship between the dependent and independent variables is nonlinear, that is, if the equation cannot be expressed using linear regression, the "nlin" procedure expresses these relationships using nonlinear mathematical models and performs parameter estimation.

Title "...."; Adds a title to the output.

Parameters a=... b=...; Provides initial estimates for the model's parameters. Appropriate values must be provided for parameter estimation and are mandatory.

Model y = a + b*...; The model's mathematical equation is defined here.

Output out=new p=Model1 r=e; The results are saved to the "new" data set. The estimated "Y" values are placed in "Model1", and the error terms of the estimated model are stored in "e1".

Run; Enables the execution of the procedure described above. Runs the procedure.

Graph (curve and curve properties) definition:

This is the stage in which the curve graph for the estimated model is drawn during the statistical model definition phase. In this stage, the graphical features related to the graph, such as line shape, color, size, axis labels, and positions, are defined. Briefly, it consists of the main headings listed below.

legend1; This is the section where settings related to the legend box within the chart are made. The relevant settings are provided under the following headings.

label = none ;
value = (j = ... "Actual data" j = ... "Estimated");
mode = protect ; t

position=(right inside middle) ; General graphic settings such as font, symbol, text color, and weight are made in this section. Related settings are provided under the following headings.

cborder=... ;

cshadow=…;
across=…;
shape=symbol(…,….);;
goptions ftext='…..' htext=…. gunit=pct ctext=….. csymbol=; In this section, how the graphic outputs will be drawn (visual settings such as font, size, color) are defined.
symbol1 i=…c=…. v=…; This section defines the symbols and lines for the actual data.
symbol2 i=…c=…v=…l=…; This section defines the symbols and lines for the model prediction curve.
axis1 and axis2 label=(angle=… rotate=… '……..') minor=….; In this section, the necessary definitions for vertical and horizontal axis lines are made.
Proc gplot data=new; Used to display the relationship between variables as a line and scatter plot. For the model defined below, the subcodes that plot the line and scatter plots on top of each other are provided.
plot y*t=1 model1*t=2 / overlay
legend=legend1
vaxis=axis1
haxis=axis2
frame
cframe=white;
run; quit;

3. Results and Discussion

The curve modeling steps and relevant codes of the SAS statistical package program for the gas production values of the legume forage plant in Table 1 are given in Table 2.

Table 2. Curve modeling stages and codes

VD	data; input t y; cards
DI	3 118.26 6 180.83 12 241.08 24 304.13 48 356.83 72 386.85 96 397.59 ;

SMD	proc nlin;
	title "orskov";
	parameters a=82.5 b=305 c=0.05;
	model y=a+b*(1-2.7182**(-c*t));
	output out=new p=MODEL1 r=e1; run;

DF	proc print; run;
	legend1 label=none value=
	(j=left "Gas Production" j=left "Orskov model")
	mode=protect position=(right inside middle)
	cborder=white cshadow=white across=1
	shape=symbol(6,2.5);
	position=center value=(justify=center);
	goptions ftext='Arial' htext=2.5 gunit=pct
	ctext=black csymbol=star;
	symbol1 i=none c=black v=star;
	symbol2 i=spline c=red v=none l=1;
	axis1 label=(angle=90 rotate=0 'Trifolium Gas Production Values') minor=none;
	axis2 label=('Measurement Times')
	minor=none;
	proc gplot ;plot y*t=1 MODEL1*t=2; /frame
	cframe=white legend=legend1
	vaxis=axis1 haxis=axis2 overlay ; run;

VD: variable definition, DI: data input, SMD: statistical model definition, DF: defining graphs (curves and curve properties).

When the program codes given in Table 2 were run in the SAS statistical package program, the model variance analysis results (Table 3), model parameter estimates (Table 4) and the Orskov gas production curve of the data set (Figure 1) were obtained as follows. The coefficient estimates obtained through iteration using the Gauss-Newton method (Stroup at al., 2009) are given in Table 5.

Table 3. Model variance analysis results

Source	DF	SS	MS	F	Pr > F
Model	2	68560.5	34280.2	234.46	<.0001
Error	4	584.8	146.2		
Corrected Total	6	69145.3			

SS: Sum of Squares, MS: Mean Square.

Table 4. Model parameter estimates

P	E	AE	ACL
a	82.6964	15.9795	38.3303-127.1
b	308.0	15.4298	265.1-350.8
c	0.0556	0.00770	0.0342-0.0769

P= parameter, E= estimate, AE= approximate standard eror, ACL= approximate 95 % confidence limits.

Table 5. Iterative phase for coefficient estimates (Gauss-Newton)

Iter	a	b	c	SS
0	82.5000	305.0	0.0500	1155.8
1	84.5951	306.6	0.0544	588.2
2	83.0422	307.8	0.0553	584.9
3	82.7581	308.0	0.0555	584.8
4	82.7074	308.0	0.0556	584.8
5	82.6983	308.0	0.0556	584.8
6	82.6967	308.0	0.0556	584.8
7	82.6964	308.0	0.0556	584.8

SS= Sum of Squares, IP= iterative phase.

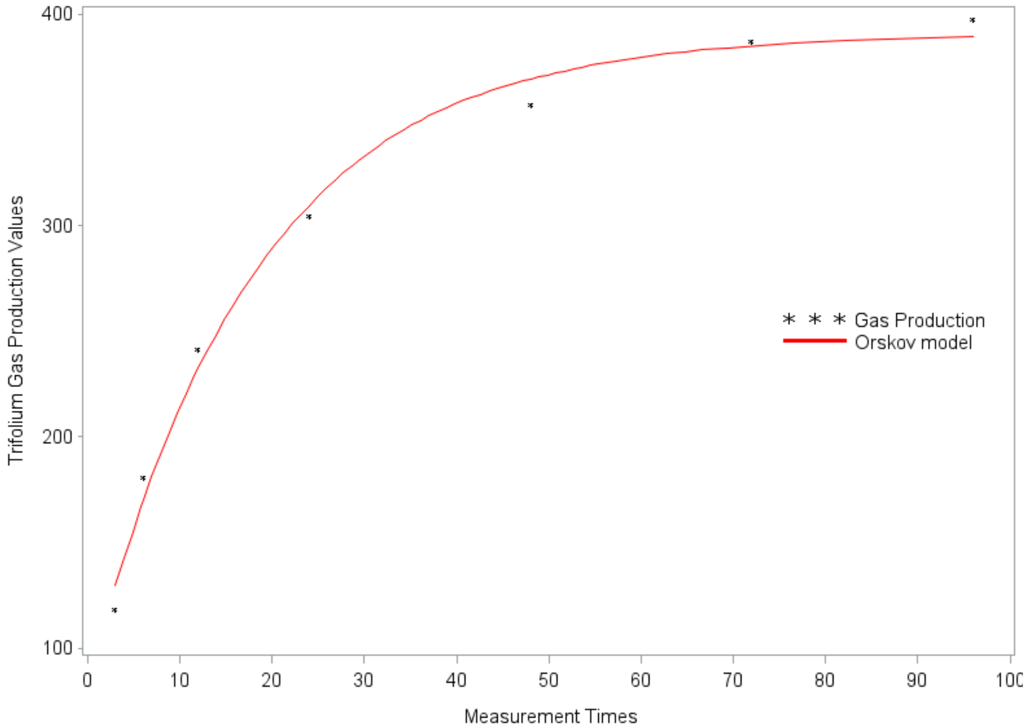

Figure 1. Orskov gas production curve of a forage legume.

When the variance analysis results given in Table 3 are examined, it is seen that the model is found to be statistically significant (P<0.0001). When the results given in Table 4 are examined, it is seen that the amount of gas obtained initially is 82.6964, the amount of gas obtained slowly is 308.0, and the gas production rate or digestion rate is 0.0556. If the initial iteration values for the amount of gas obtained initially, the amount of gas obtained slowly, and the gas production rate are given accurately, the t coefficient estimates can be obtained in the 7th iteration, as seen in Table 5. This clearly demonstrates the importance of the initial iteration values. When Figure 1 is examined, it is seen that the observation values are distributed around the curve obtained with the Orskov model. The coefficient of determination calculated from Table 3 is 0.991 (R^2=1-(584.8/69145.3)). This value is an indication that the created model can represent the point distribution at a high level.

4. Conclusion

In this study, curve modeling analyses based on the Orskov and McDonald model were successfully performed using SAS software. The findings demonstrated that the model is a powerful tool for understanding biological processes and provides a high degree of fit to the data.

The basis for this success lies not only in the selection of the correct model but also in the accurate, careful, and deliberate execution of SAS coding. Especially in nonlinear regression analyses, the correct definition of the model form, the appropriate entry of initial values, the clear definition of parameter constraints, and the correct interpretation of the outputs are crucial. Thanks to the flexible structure offered by SAS, these processes could be carried out systematically and reliable results were achieved.

Accuracy and care are essential in statistical analyses performed with SAS, as coding errors or minor inaccuracies in model definition can completely invalidate the results. Properly written SAS codes not only ensured the accuracy of statistical calculations but also supported the scientific validity of the obtained results.

In conclusion, this study demonstrated that accurate SAS coding is a fundamental cornerstone of the scientific modeling process. Precision in coding is a factor that directly affects the success of the model and plays a critical role in ensuring the reliability and repeatability of the results obtained.

Author Contributions

The percentages of all authors' contributions are presented below. All authors reviewed and approved the final version of the manuscript.

	M.Ş.	İ.G.	E.Y.
C	35	35	30
D	35	35	30
S	35	35	30
DCP	35	35	30
DAI	35	35	30
L	35	35	30
W	35	35	30
CR	35	35	30
SR	35	35	30
PM	35	35	30
FA	35	35	30

C=Concept, D= design, S= supervision, DCP= data collection and/or processing, DAI= data analysis and/or interpretation, L= literature search, W= writing, CR= critical review, SR= submission and revision, PM= project management, FA= funding acquisition.

Conflict of Interest

The authors declare that there is no conflict of interest.

Ethical Consideration

Ethics committee approval was not required for this study because of there was no study on animals or humans.

Acknowledgements

This study has been published in the proceeding book of the IX. International Congress on Domestic Animal Breeding, Genetics and Husbandry-2025 (ICABGEH-25), Szeged, Hungary, August 12-14, 2020.

References

Bayazıt, M., Şahin, M., & Tolun, T. (2022). Modeling of in vitro gas production. *Black Sea Journal of Agriculture, 5*(3), 260–264. https://doi.org/10.47115/bsagriculture.1091017

dos Santos Cabral, Í., Azevêdo, J. A. G., Pina, D. dos S., Pereira, L. G. R., Fernandes, H. J., Almeida, F. M., Souza, L. L., de Lima, R. F., & Cirne, L. G. A. (2019). Evaluation of models utilized in in vitro gas production from tropical feedstuffs. *Semina: Ciências Agrárias, 40*(1), 443–456. https://doi.org/10.5433/1679-0359.2019v40n1p443

Goodwin, D. K., Ramsey, A. F., & Chvosta, J. (2018). *Applied econometrics with SAS: Modeling demand, supply, and risk*. SAS Press.

Korosteleva, O. (2018). *Advanced regression models with SAS and R*. Chapman & Hall.

Makkar, H. P. S., & Blümmel, M. (1997). In vitro rumen apparent and true digestibilities of tannin-rich forages. *Animal Feed Science and Technology, 67*(3–4), 245–251.

Orskov, E. R., & McDonald, I. (1979). The estimation of protein degradability in the rumen from incubation measurements weighted according to the rate of passage. *Journal of Agricultural Science, 92*(2), 499–503.

Özkan, Ç. Ö., & Kamalak, A. (2023). İn-vitro gaz üretim tekniğinde kullanılan matematik modeller ve karşılaştırma yöntemleri. İçinde *Tarımsal araştırmalarda istatistiksel analizler* (ss. 1–7). Holistence Publications.

Panik, M. J. (2010). *Regression modeling: Methods, theory, and computation with SAS*. Chapman & Hall.

Panik, M. J. (2013). *Growth curve modeling: Theory and applications*. John Wiley & Sons.

SAS Institute Inc. (1999). *The SAS system for Windows. Release 8.0.1*. SAS Institute Inc.

Stroup, W. W., Milliken, G. A., Claassen, E. A., & Wolfinger, R. D. (2009). *Generalized linear mixed models: Modern concepts, methods and applications*. SAS Press.