engineering proceedings

MDPI

Proceeding Paper

Growth Curves Modelling and Its Application †

Ana García-Burgos [1,*,‡], Beatriz González-Alzaga [2], María José Giménez-Asensio [2], Marina Lacasaña [2], Nuria Rico-Castro [1,‡] and Desirée Romero-Molina [1,‡]

1 Department of Statistics and Operational Research, University of Granada, 18071 Granada, Spain; nrico@ugr.es (N.R.-C.); deromero@ugr.es (D.R.-M.)
2 Andalusian School of Public Health, 18011 Granada, Spain; beatriz.gonzalez.easp@juntadeandalucia.es (B.G.-A.); mariajosesases@hotmail.com (M.J.G.-A.); marina.lacasana.easp@juntadeandalucia.es (M.L.)
* Correspondence: agburgos@ugr.es

† Presented at the 9th International Conference on Time Series and Forecasting, Gran Canaria, Spain, 12–14 July 2023.
‡ These authors contributed equally to this work.

Abstract: In this article, we compare two ways of modelling measures of fetal growth. The goal is to impute the missing information for certain ultrasound measurements that are observed at different times and with different numbers of observations. To analyze the effect that other variables have, such as environmental exposure to certain substances or diet, on fetal growth based on these data, we need to handle the information measured at the same instant of time for all the individuals under study, preferably in three time windows of pregnancy (first trimester, week 12; second trimester, week 20; third trimester, week 34). For this, data at these chosen times, in case they are not available, must be imputed from the available information using an appropriate statistical model. One option is to use a linear model, specifically a generalized least squares model that is fitted to the features shown in the data. The other option is to use diffusion processes, estimating their parameters based on the available information. In both options, missing data can be estimated with the unconditional fitted model, conditional on the previous available measurement, or conditional to the closest measurement.

Keywords: growth curves; diffusion processes; linear models

Citation: García-Burgos, A.; González-Alzaga, B.; Giménez-Asensio, M.J.; Lacasaña, M.; Rico-Castro, N.; Romero-Molina, D. Growth Curves Modelling and Its Application. *Eng. Proc.* **2023**, *39*, 66. https://doi.org/10.3390/engproc2023039066

Academic Editors: Ignacio Rojas, Hector Pomares, Luis Javier Herrera, Fernando Rojas and Olga Valenzuela

Published: 6 July 2023

Growth Curves Modelling and Its Application

The aim of this work is to compare different methods for statistical modelling growth curves. We compare two different methodologies in the study of a dataset from the GENEIDA (Genetics, Early life environmental Exposures and Infant Development in Andalusia) https://www.easp.es/web/geneida/ (accessed on 27 June 2023) project. This project details a cohort born in 2014, made up of 800 mother–chid pairs. They are followed up during pregnancy, birth and childhood. One of the objectives of the project is to understand how diet and exposure to environmental substances of the pregnant mothers affect fetal growth. To this end, we have some ultrasound measurements performed during pregnancy, which are as follows:

- Biparietal diameter (BPD): distance in millimetres between both parietal bones of the baby's head.
- Abdominal circumference (AC): distance in millimetres around the abdomen.
- Head circumference (HC): distance in millimetres around the head measured above the eyebrows and ears.
- Femur length (FL): length in millimetres of baby's femur.
- Estimated fetal weight (EFW): we estimate the fetal weight using Hadlock's formula (Hadlock et al. [1]).

The data obtained in this study have the following characteristics: the echographic information is measured at different instants of time and the number of ultrasounds is different for each mother. We can see these two characteristics in Figure 1, since the three individuals represented have seven, four and five ultrasound measurements, respectively, at different gestational ages.

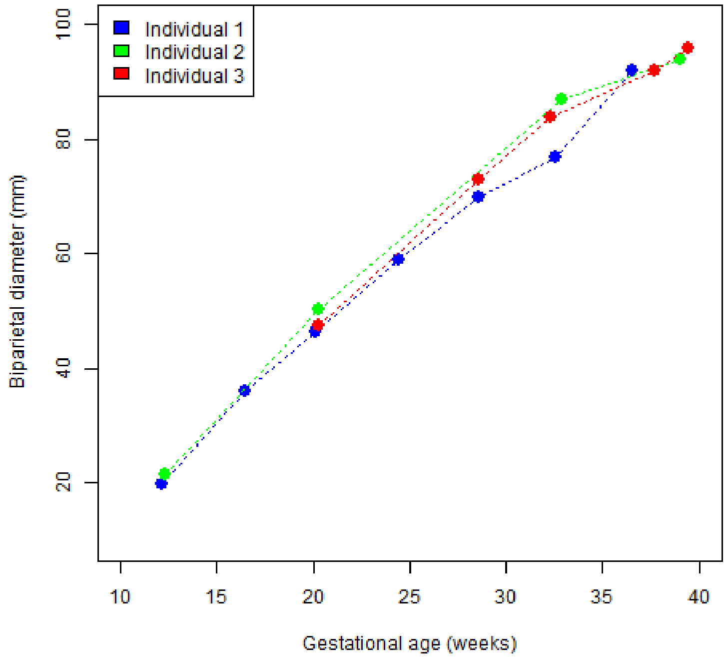


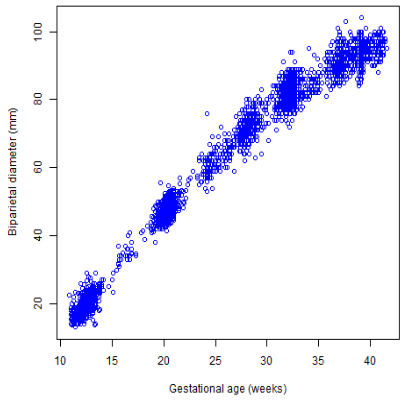
Figure 1. Biparietal diameter of three individuals at different gestational ages.

However, in order to analyze the effect of certain variables on fetal growth based on these data, we need them to be measured at the same instant of time for all individuals. To do this, these data must be imputed based on the available information, preferably using an appropriate statistical model. There are different methodologies to find such a model that fits a set of observations of a quantitative variable that evolves, presenting a growth throughout the time. The curves that model this type of data are called growth curves and their use homogenizes the information from the ultrasound measurements.

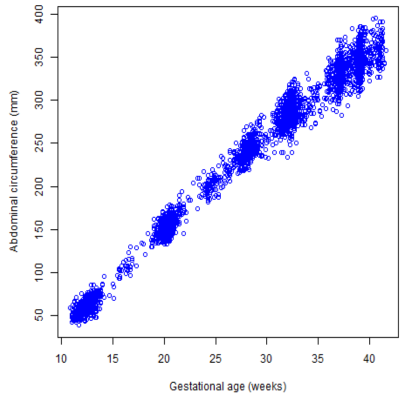
In this work, we propose a comparison of two different methodologies in growth modelling to impute the ultrasound measurements at the desired instant of time: linear models and diffusion processes.

Firstly, we approach the problem using linear models. This consists of considering that the observations are a function of a variable, time, and there is a linear relationship that relates them. Following Iñiguez et al. [2], we created models to predict the five fetal measurements at 12, 20 and 34 weeks of gestation. Initially, we tried to use the generalized linear model, but some of the hypotheses failed. In particular, these two factors stood out: heteroskedasticity of the residuals and autocorrelation. Therefore, the generalized least squares model has been used to obtain the predictions, since this model is less restrictive in terms of assumptions (Kariya and Kurata [3]).

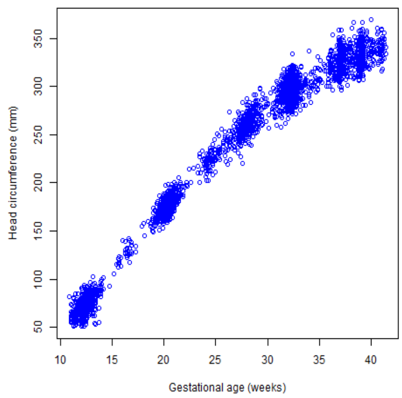
Secondly we solve the problem through diffusion processes. In this case, the observed variable $X(t)$ evolves over time t and at each instant there is a probability distribution for $X(t)$ that depends both on time and on the values observed at previous times $X(0), X(1), \ldots, X(t-1)$. Diffusion processes are useful for modelling time-dependent variables that increase, usually with an exponential or sigmoidal trend (Baudoin [4]). We choose the type of process based on the characteristics of the observed sample paths. In this case, as the data have a sigmoidal or exponential trend, as we can see in Figure 2, we set the mixed Gompertz–lognormal process. In particular cases, it includes the lognormal process, associated with an exponential curve, and the Gompertz-type process, related to a sigmoidal curve (Romero et al. [5]).



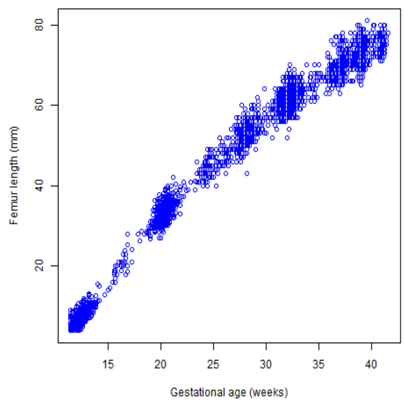
(**a**) Biparietal diameter.



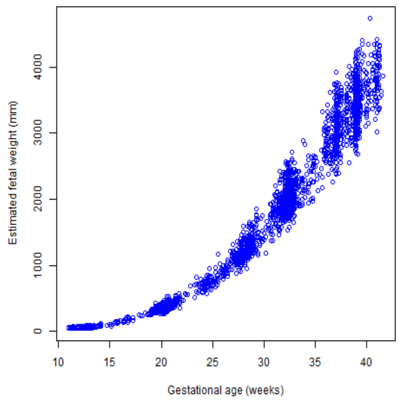
(**b**) Abdominal circumference.



(**c**) Head circumference.



(**d**) Femur lenght.



(**e**) Estimated fetal weight.

Figure 2. Observed sample paths.

Before fitting the growth curves, we carried out an exhaustive clean up of the data by studying the outliers and eliminating defective data. In the case of linear models, we fit a model for each measure with their respective confounders to later review the influential data and recalculate the model. In the case of diffusion processes, we performed a weighted cluster analysis to group the data depending on the result of the analysis and we fit a process for each cluster in each measurement. Finally, we obtained the data of the measurement in the desired gestational age using an unconditional model, one conditioned to the previous data and another conditioned to the closest available data.

After making the adjustments using linear models and diffusion processes, we compared the results of the two methodologies to find out which best imputes the data. To do this, we used different measures to study the error made in the data imputations (Shcherbakov et al. [6]).

Author Contributions: Conceptualization, A.G.-B., N.R.-C. and D.R.-M.; methodology, A.G.-B., N.R.-C. and D.R.-M.; software, A.G.-B., N.R.-C. and D.R.-M.; validation, N.R.-C. and D.R.-M.; formal analysis, A.G.-B., N.R.-C. and D.R.-M.; investigation, A.G.-B., B.G.-A., M.L., N.R.-C. and D.R.-M.; resources, B.G.-A., M.J.G.-A. and M.L.; data curation, B.G.-A., M.J.G.-A. and M.L.; writing—original draft preparation, A.G.-B., N.R.-C. and D.R.-M.; writing—review and editing, A.G.-B., N.R.-C. and D.R.-M.; visualization, A.G.-B.; supervision, N.R.-C. and D.R.-M.; project administration, M.L.; funding acquisition, M.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and the protocol was approved by the Hospital Ethics Committee, the Ethics Committee of "Consejería de Salud y Familias, Junta de Andalucía" (PI-0405-2014) and "Consejería de Igualdad, Salud y Políticas Sociales, Junta de Andalucía" (PI13/01559). We follow the standards described in Andalusian and Spanish laws of personal data protection and biomedical research for the treatment of information and biological samples of human origin.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data available on request due to restrictions eg privacy or ethical. The data presented in this study are available on request from the corresponding author. The data are not publicly available due to the informed consent, obtained from all participants or the legally responsible before they participated in the study, does not establish the transfer to third parties or make it public.

Acknowledgments: FQM147-Análisis estadístico de datos multivariantes y procesos estocásticos.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

GENEIDA	Genetics, Early life environmental Exposures and Infant Development in Andalusia
BPD	Biparietal Diameter
AC	Abdominal circumference
HC	Head circumference
FL	Femur lenght
EFW	Estimated fetal weight

References

1. Hadlock, F.P.; Harrist, R.B.; Sharman, R.S.; Deter, R.L.; Park, S.K. Estimation of fetal weight with the use of head, body, and femur measurements—A prospective study. *Am. J. Obstet. Gynecol.* **1985**, *151*, 333–337. [CrossRef] [PubMed]
2. Iñiguez, C.; Esplugues, A.; Sunyer, J.; Basterrechea, M.; Fernández-Somoano, A.; Costa, O.; Estarlich, M.; Aguilera, I.; Lertxundi, A.; Tardón, A.; et al. INMA Project.: Prenatal Exposure to NO_2 and Ultrasound Measures of Fetal Growth in the Spanish INMA Cohort. *Environ. Health Perspect.* **2016**, *124*, 235–242. [CrossRef] [PubMed]
3. Kariya, T.; Kurata, H. *Generalized Least Squares*; Wiley Series in Probability and Statistics; John Wiley & Sons: Hoboken, NJ, USA, 2004.
4. Baudoin, F. *Diffusion Processes and Stochastic Calculus*; European Mathematical Society/American Mathematical Society: West Lafayette, IN, USA, 2014.
5. Romero, D.; Rico, N.; Arenas, M. A new diffusion process to epidemic data. *Lect. Notes Comput. Sci.* **2013**, *8111*, 69–76.
6. Shcherbakov, M.V.; Brebels, A.; Shcherbakova, N.L.; Tyukov, A.; Janovsky, T.; Kamaev, V.A. A Survey of Forecast Error Measures. *World Appl. Sci. J.* **2013**, *24*, 171–176.