ISSN print 0716-0356
ISSN online 2452-5995
ISSN-L 0716-0356

TRILOGÍA

CIENCIA · TECNOLOGÍA · SOCIEDAD

ARTÍCULOS

▪ PRESENTACIÓN

CONFERENCE PAPER

> ARTÍCULO | CONFERENCE PAPER
▪ DESIGN AND IMPLEMENTATION OF A ROBOTIC ARM CONTROLLED BY REINFORCEMENT LEARNING TECHNIQUES

> ARTÍCULO | CONFERENCE PAPER
▪ SMART IRRIGATION THROUGH WATER CONSUMPTION: PREDICTION

> ARTÍCULO | CONFERENCE PAPER
▪ STRUCTURED GRAMMATICAL EVOLUTION FOR MODELING THE MULTI-BAND LIGHT CURVES OF SUPERNOVA.

>COMUNICACIÓN BREVE
▪ JUAN JOSÉ SALDAÑA GONZÁLEZ: ADELANTADO DE LA HISTORIA DE LA CIENCIA EN AMÉRICA.

>DISCURSO
▪ ALEJANDRO *MONO* GONZÁLEZ: LA DESAPARICIÓN DE UNA EXTREMA VISIBILIDAD

UTEM
UNIVERSIDAD TECNOLÓGICA METROPOLITANA
del Estado de Chile

AGOSTO | 2023
Vol. 38 • Nº 49

Conference paper EVIC versión XVII, 2022* **

Nicolás Fierro Flores***

Universidad de Chile Santiago, Chile.

https://orcid.org/0009-0002-3467-2756

Jhon Pilataxi ****

Universidad de Chile Santiago, Chile.

https://orcid.org/0009-0003-6217-9501

Artículo

STRUCTURED GRAMMATICAL EVOLUTION FOR MODELING THE MULTI-BAND LIGHT CURVES OF SUPERNOVA*****

EVOLUCIÓN GRAMATICAL ESTRUCTURADA PARA MODELAR LAS CURVAS DE LUZ MULTIBANDA DE SUPERNOVA

Recibido: 2 de diciembre de 2022 | Aprobado: 24 de abril de 2023 | Versión final: 6 de junio de 2023

Cómo citar este artículo:
Fierro Flores, N. I. y Pilataxi, J. (2023). Structured Grammatical Evolution for Modeling the Multi-band Light Curves of Supernova. Trilogía (Santiago), 38(49), 24-31. Santiago de Chile: Ediciones UTEM.

* EVIC (2022). Escuela de Verano anual en Inteligencia Computacional. IEEE CIS, Utem. Recuperado de: https://evic2022.utem.cl/

** Poster Competition 1 rd . XVII IEEE Latin-American Summer School on Computational Intelligence (EVIC) December 12-14, 2022, Universidad Tecnológica Metropolitana, Chile.

*** This work was supported by the Agencia Nacional de Investigación y Desarrollo (ANID) under Becas/Doctorado Nacional 21211513

**** Departamento de Ingeniera eléctrica, Universidad de Chile. Correo electrónico: nicolas.fierro@ug.uchile.cl

***** Departamento de Ingeniera eléctrica, Universidad de Chile. Correo electrónico: jhon.pilataxi@ug.uchile.cl

ABSTRACT

Supernovas (SNs) have been one of the most studied events in astronomy. However, there are still no models capable of describing this phenomenon in a general and accurate way. These models generally seek to describe a single type of supernova, requires multiple parameter's values and often do not distinguish between the different light bands of the same curve. Structured grammatical evolution allows the generation of a model with data and a given basal structure, which can be designed considering the nature of the problem for which we are looking for a model. In this case, with some mathematical assumptions we can generate a symbolic regression to obtain a model for different types of SNs and for each light band. We can also use this algorithm to fit the parametric model of the supernova and obtain the value of the variables needed to model it.

Keywords: Astroinformatics, Supernova, Genetic Pro- gramming, Structured Grammatical Evolution, Symbolic regression

RESUMEN

Las supernovas (SN) han sido eventos muy estudiados por la astronomía. Sin embargo, aún no existen modelos capaces de describir este fenómeno de manera general y precisa. Estos modelos generalmente buscan describir un solo tipo de supernova, requieren múltiples valores de parámetros y, a menudo, no distinguen entre las diferentes bandas de luz de una misma curva. La evolución gramatical estructurada permite generar un modelo con datos y una estructura base determinada que puede diseñarse teniendo en cuenta la naturaleza del problema para el que buscamos un modelo. En este caso, con algunas suposiciones matemáticas podemos generar una regresión simbólica para obtener un modelo para diferentes tipos de SN y para cada banda de luz. También podemos usar este algoritmo para ajustar el modelo paramétrico de la supernova y obtener el valor de las variables necesarias para modelarlo.

Palabras clave: astroinformática, supernova, programación genética, evolución gramatical estructurada, regresión simbólica

I. INTRODUCTION

There is an interest in describing the supernova phenomenon that increases its luminosity until it reaches a peak and then decreases until it disappears. The Supernova Parametric Model (SPM) attempts to describe the typical behavior of SN light curve [1]. The SPM function is defined as:

$$f_{sne}(t) = f_{early}(t)(1 - g(t)) + f_{late}(t)g(t), \quad (1)$$

$$g(t) = sigmoid(s(t - (\gamma + t_0)), \quad (2)$$

$$f_{early}(t) = A(1 - \beta(t - t_0)/\gamma)/den, \quad (3)$$

$$f_{late}(t) = A(1 - \beta)exp((\gamma + t_0 - t)/den, \quad (4)$$

$$den = 1 + exp(-(t - t_0)/\tau_{rise}), \quad (5)$$

where A affects the brightness scale, t_0 acts as a temporal shift, γ controls the time duration of SN-plateau region, β controls the slope of the SN-plateau region, τ_{rise} controls the required time to reach the maximum brightness, τ_{fall} controls the brightness decay and s is a transition factor. As we can notice, SPM requires the adjustment of multiple variables and the prediction is complex, due to the limited availability of data and difficulties in terms of tracking, since SNs are generally detected some time after the onset. This is why symbolic regression through evolutionary algorithms emerges as an interesting technique to obtain simpler and more specific models for different types of supernovae and the bands of light that compose them.

Symbolic regression (SR) is a type of regression analysis that searches the space of mathematical expressions to find the model that best fits a given dataset, both in terms of accuracy and simplicity.

Evolutionary Algorithms (EA) are computational methods inspired by the principles of natural selection and genetics. Grammatical evolution is an evolutionary algorithm that can evolve complete programs in an arbitrary language using a variable-length binary string. The binary genome determines which production rules in a Backus–Naur form grammar def- inition are used in a genotype-to-phenotype mapping process to a program [2]. Among Evolutionary algorithms, Genetic Programming (GP) has been very popular. Many experiments generate their own language specific for their particular prob- lem. Grammatical Evolution (GE) performs the evolutionary process on variable-length binary strings, unlike other GP algorithms [3]. In Structured Grammatical Evolution (SGE) each gene is linked to a specific non-terminal and is composed by a list of integers. The length of each list is determined by computing the maximum possible number of expansions of the corresponding non-terminal. This structure ensures that when a gene is modified, it does not affect the derivation options of other non-terminals, thus narrowing the number of changes that occur at the phenotypic level. In this work, we utilized SGE to find models that describe six datasets of SNs using two grammars, one of them based on the SPM and the other based on simple math expression.

II. METHOD

We explored two approaches (explicit and simplified gram- mar) to solve Supernova symbolic regression problem. In the first one, we set the SPM structure as the start symbol and evolved a population to optimize the SPM parameters; on the other approach, in simplified grammar, we defined basic structures and operations, and the SGE found the function that best fit to the dataset. In simplified grammar the start symbol is the sum between a number and expression, where an expression is a non-terminal set, that include basic operation [+, −, *, /] and complex operation [*exp*, *sigmoid*], also we defined a free-constant and a variable *t*.

We used six datasets. The first one was generated from SPM using the following parameters: A=0.18, t_0= 33.941, γ= 18.975, β=0.666, τ_{rise} = 13.416 and τ_{fall} = 40.076, those values were taken from [4]. The second dataset was the same as first one but adding Gaussian noise. Additionally, we used four synthetic datasets. The synthetic data represent examples of green and red energy bands of two supernova type.

I. For the Generation of Synthetic Observations we based in [5]. First, a set of optimal SPM parameters $\theta^{(b)*}$ are sampled by using an MCMC posterior distribution estimation from the empirical observations. The MCMC sampling procedure allows us to introduce a moderate diversity of the SPM param- eters when generating a new light-curve. After this, the STW is generated, and the sampled time values are evaluated using the SPM analytical function f_{sne} and a set of optimal SPM parameters $\theta^{(b)*}$. In this way, multiple synthetic observation- fluxes $\hat{\mu}_{i,j}(b)$μˆi,j(b) are generated. Table 1 presents the evolutionary parameter used in all experi- mental runs. The goal was to minimize the error between data and regression function, so we used Root Mean Squared Error (RMSE) as fitness function.

TABLE I

Parameter settings.

Parameter	Value
Number of runs	5
Population Size	500
Number of generations	1000
Crossover probability	0.9
Mutation probability	0.3
Elitism	0.1
Selection type	Tournament

III. RESULTS

Table 2 presents the RMSE (mean and standard deviation of the five independent runs) achieved by both approaches.

TABLE II

Mean and standard deviation of RMSE achieved in five runs. The RMSE value are scaled by 10^{-3}.

Data	Explicit Grammar	Simplified Grammar
SPM	4.43±1.9	6.53±0.77
SPM noise	5.453±0.8	6.27±1.3
Red band Synthetic I	19.11±1.4	21.02±4.1
Green band Synthetic I	12.9±2	17.46±3
Red band Synthetic II	69.32±17	24.74±12
Green band Synthetic II	49.17±16	30.17±2.5

Even though the RMSE achieved with explicit grammar was less than simplifies grammar in the four first used datasets, statistical test proved that there were not statistically significant differences in RSME archived with both grammars. Figures 1 and 2 shows the final function obtained with the proposed grammars in SPM noise dataset. In contrast, in the two-most difficult datasets, red and green band Syntetic II, the functions obtained with simplified grammar were best fit that those obtained with explicit grammar (see Figures 3 and 4). Using SGE with explicit grammar was not enough to find the SPM parameter values, due to math operations between pa- rameters defined in the SPM, and several values combinations gave the same result. Further, the SPM has two parts: early and late, but when we separated the final functions obtained by using the first dataset, both parts were completely different to the original ones (see Figure 5). These results indicate that is possible to find several representations for a dataset, and some of them could be a reduced version of SPM. When we observed the final functions obtained with simplified grammar, we noticed that the majority of functions have the following structure: a number that represent the first observation plus a fraction, where the denominator was similar to the SPM, and the numerator had a sigmoid function that models light curves between the first observation and the peak, so in the Green band Syntetic II, which start with the peak the sigmoid function did not appear in the final functions.

IV. CONCLUSION

SGE was used to generate a model that describes the Supernova, using two different grammars: One that uses the Parametric Model of Supernovae and that allows the adjust- ment of the parameters, and another simpler one composed of sigmoid and exponential operations that generates different models for different cases maintaining a structure that allows the detection of increase, peak and decrease of the flux caused by the phenomenon. For multi-band supernovae SGE using the parametric model achieves good results that are close to the data, while for specific red and green bands the performance is not of quality. On the other hand, SGE with a simpler grammar achieves good performance in both cases, thus allowing to have a model generator that fits different SN types and different bands. The generation of simpler expressions than the parametric model allows a more practical and faster analysis of this phenomenon, as well as allowing to observe and theorize about differences in the models for each type of Supernova. The simplicity of the expressions allows to reduce the time needed for them to be evaluated, which could be useful for massive data processing of different phenomena of this nature.

REFERENCES

V. Villar, E. Berger, G. Miller, R. Chornock, A. Rest, D. Jones, M. Drout,

R. Foley, R. Kirshner, R. Lunnan *et al.*, "Supernova photometric classi- fication pipelines trained on spectroscopically classified supernovae from the pan-starrs1 medium-deep survey," *The Astrophysical Journal*, 8 8 4 , 1, p. 83, 2019.

M. O'Neill and C. Ryan, "Grammatical evolution," *IEEE Transactions on Evolutionary Computation*, 5, 4, pp. 349–358, 2001.

N. Lourenc¸o, F. B. Pereira, and E. Costa, "Sge: a structured representation for grammatical evolution," in *International Conference on Artificial Evolution (Evolution Artificielle)*. Springer, 2015, pp. 136–148.

O´ . A. Pimentel Fuentes, "Clasificación profunda de curvas de luz multi- banda basada en atención," 2022.

O´ . Pimentel, P. A. Estevez, and F. Forster, "Deep attention-based supernovae classification of multi-band light-curves," *arXiv preprint arXiv:2201.08482*, 2022.

V. ANEXO

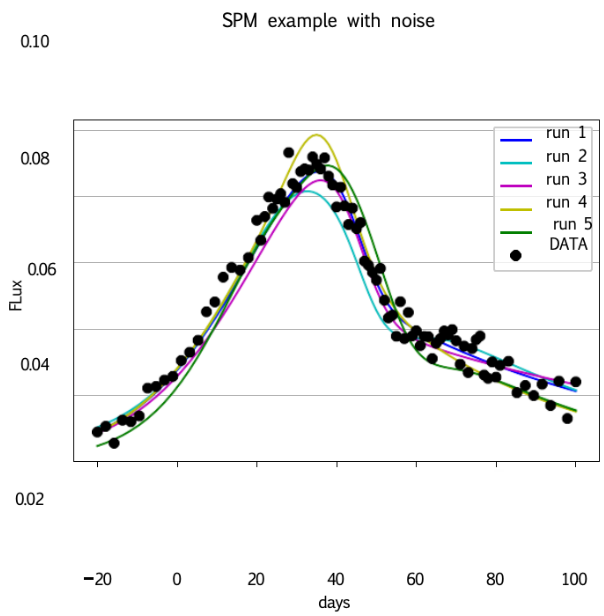

Fig. 1. Functions obtained with explicit grammar in SPM with noise.

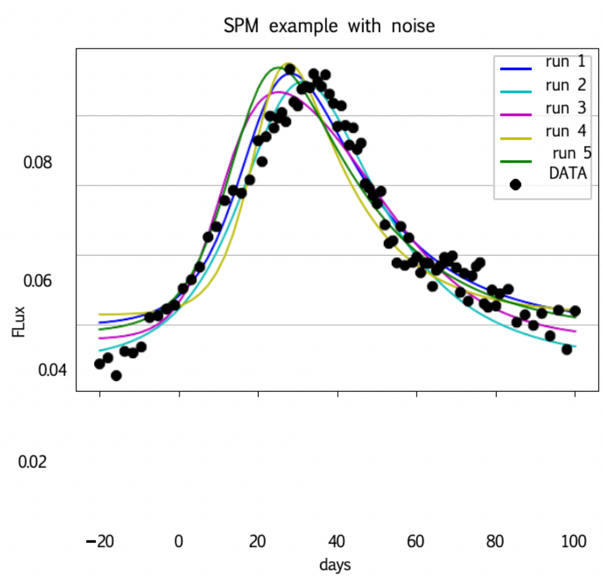

Fig. 2. Functions obtained with simplified grammar in SPM with noise.

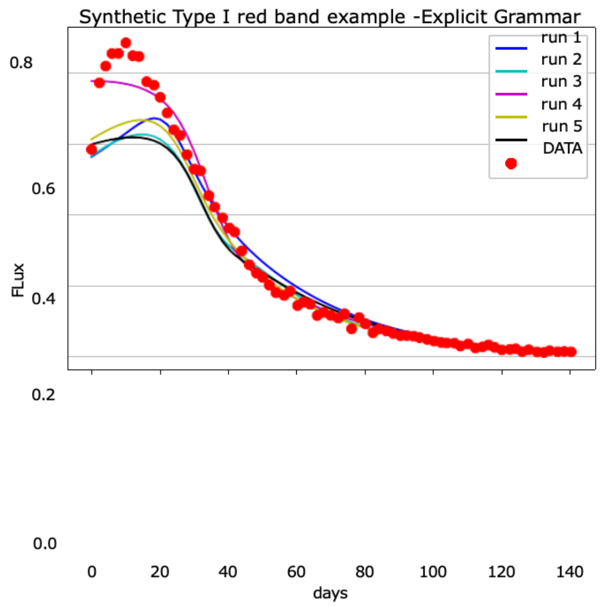

Fig. 3. Functions obtained with explicit grammar in red band synthetic

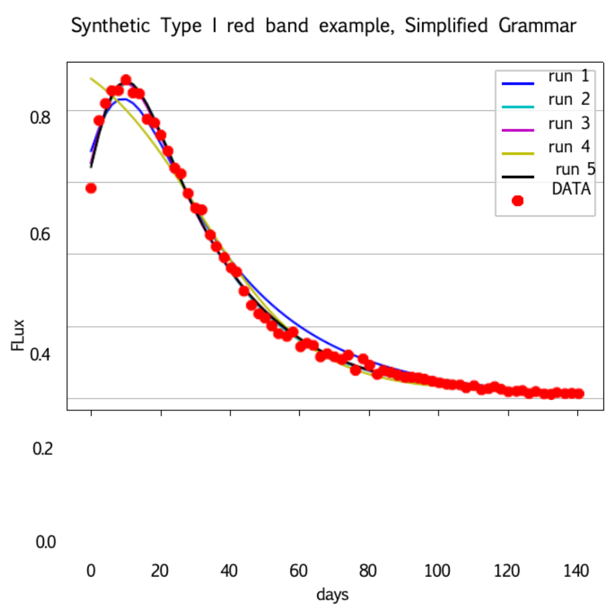

Fig. 4. Functions obtained with simplified grammar in red band synthetic.

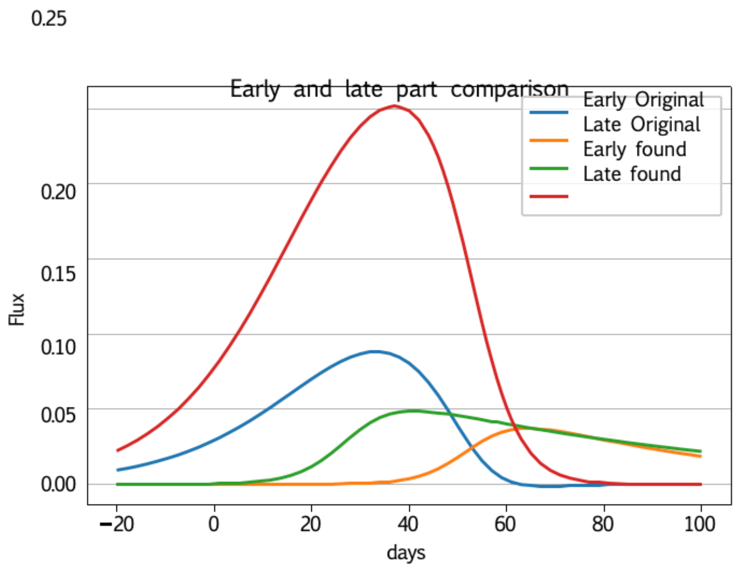

Fig. 5. Early and late comparison of SPM and found function.

UNIVERSIDAD
TECNOLÓGICA
METROPOLITANA
UTEM
del Estado de Chile

TRILOGÍA
CIENCIA · TECNOLOGÍA · SOCIEDAD

EDICIONES UNIVERSIDAD
TECNOLÓGICA METROPOLITANA