Transactions on Computational and Applied Mathematics (2024)
Clausius Scientific Press, Canada
DOI: 10.23977/tracam.2024.040106
ISSN 2616-1826 Vol. 4 Num. 1

Application of GAMLSS Model in Analysis of Composition of Glass Relics in China

Ruisheng Zhang, Xinhong Liu*, Yanqi Zhang, Xuanyi Jin

Beijing Institute of Petrochemical Technology, Beijing, 102617, China
**Corresponding author: liuxinhong@bipt.edu.cn*

Keywords: GAMLSS, SEP3 distribution, Normal distribution

Abstract: Based on a systematic analysis of the data characteristics of unearthed glass relics, a GAMLSS model was established for the main components, which proved that introducing explanatory variables into the prediction of location and scale parameters can more effectively explain the heterogeneity and skewness characteristics of each main component. The correlation between glass type, decoration, surface weathering, and color was explored. The presence or absence of weathering components on the surface was predicted. The empirical results show that the GAMLSS model can not only provide a basis for the prediction of various components of glass relics, but also make the classification of glass more stable and reasonable. This provides a new reference for studying the origin of glass in China.

1. Introduction

Glass has a long history of development and is one of the earliest artificial materials invented by humans. Due to the ancient Chinese people's love for glass, glass, as a treasure of ancient society, developed and flourished, and the related craftsmanship and technology of glass also developed rapidly. Glass products were not only important materials in ancient China's foreign trade, but also reflected the highly developed craftsmanship and technology of ancient China[1-3]. The appearance of glass in our country occurred later than the discovery of glass in the world. The ancients absorbed foreign technology and made local materials, so local glass products have a similar appearance to foreign ones, but their chemical composition is different. Through chemical composition analysis of unearthed glass relics, not only has a certain understanding of the source of ancient Chinese glass products been obtained, but also valuable information has been provided for the study of the development history of ancient Chinese glass[4].

Based on data from a cultural relic glass unearthed in China, this article finds that the degree of surface weathering of glass is significantly correlated with the type of glass. In addition, patterns and types are also correlated under surface weathering factors. The SEP3 distribution and normal distribution were used to fit the main components, and a GAMLSS model was established between the various components and related factors of glass. The rationality of the model was tested using the software package gamlss[5-6] in R software, and the composition and related relationships of the main components were obtained. This provides new reference materials for studying the origin of glass in China.

2. GAMLSS model

Rigby and Stasinopoulos detailed the model in 2005- a generalized additive model based on position, scale, and shape parameters (GAMLSS). This model is more flexible and complex than the multiple regression model. The response variables are no longer limited to exponential distribution families, and the system part not only considers the relationship between positional parameters and explanatory variables, but also incorporates the relationship between scale parameters, shape parameters, and explanatory variables. These improvements enable the GAMLSS model to better explain data issues.

Assuming the probability density function of the response variable is $f\left(y|\boldsymbol{\theta}\right)$, Among them, $\boldsymbol{\theta}$ is a vector $\boldsymbol{\theta}=\left(\theta_1,\theta_2,\cdots,\theta_p\right)^T$ composed of p parameters. Under given conditions $\boldsymbol{\theta}=\boldsymbol{\theta}^i=\left(\theta_{i1},\theta_{i2},\cdots,\theta_{ip}\right)$, $y_i\left(i=1,2,\cdots,n\right)$ are independent of each other, and their probability density function is $f\left(y_i|\boldsymbol{\theta}=\boldsymbol{\theta}^i\right)$. The GAMLSS package written by Stasinopoulos and Rigby in 2007 contains over 60 different distributions. This article adopts the Skew Exponential Power type 3 distribution (SEP3) and an unbiased normal distribution that can describe unbiased data. The probability density function of the SEP3 distribution is

$$f_1\left(y\right)=\frac{c}{\sigma}\left\{\exp\left[-\frac{1}{2}|vz|^{\tau}\right]I(y<\mu)+\exp\left[-\frac{1}{2}|\frac{z}{v}|^{\tau}\right]I(y\geq\mu)\right\},\quad -\infty<y<\infty$$

Among them, the position parameter and scale parameter are $-\infty<\mu<\infty$, $\sigma>0$, $v>0$, $T>0$ $Z=\frac{y-\mu}{\sigma}$, $c=v\tau/[(1+v^2)2^{1/\tau}\Gamma(1/\tau)]$ and I are indicative functions.

The expectation of the SEP3 distribution is $\mathrm{E}\left(Y\right)=\mu+\sigma E(Z)$, the variance is $Var\left(Y\right)=\sigma^2V(Z)$, while $\mathrm{E}\left(Z\right)=2^{1/\tau}\Gamma(2/\tau)(v-1/v)/\Gamma(1/\tau)$, $\mathrm{E}\left(Z^2\right)=2^{2/\tau}\Gamma(3/\tau)(v^3+1/v^3)/[\Gamma(1/\tau)(v+1/v)]$.

The systematic part of the GAMLSS model[5] can establish a regression model between $\boldsymbol{\theta}^i$ and explanatory variables. $\mathbf{y}^T=\left(y_1,y_2,\cdots,y_n\right)$ is a vector composed of observed values of response variables, where $g_k\left(\cdot\right)\left(k=1,2,\cdots,p\right)$ is a known monotonic connection function. The form of connecting regression models is:

$$g_k\left(\boldsymbol{\theta}_k\right)=\boldsymbol{\eta}_k=\mathbf{X}_k\boldsymbol{\beta}_k+\sum_{j=1}^{J_k}\mathbf{Z}_{jk}\boldsymbol{\gamma}_{jk} \tag{1}$$

At this point, both $\boldsymbol{\theta}_k$ and $\boldsymbol{\eta}_k$ are n-dimensional vectors, $\mathbf{X}_k$ is a design matrix of a known $n\times J_k'$, $\mathbf{Z}_{jk}$ is a matrix of $n\times q_{jk}$, $\boldsymbol{\beta}_k^T=\left(\beta_{1k},\beta_{2k},\cdots,\beta_{J_k'k}\right)$ is a J_k'-dimensional parameter vector, and $\boldsymbol{\gamma}_{jk}$ is a random vector with a q_{jk}-ary normal distribution. Equation (1) is called a generalized additive model based on position, scale, and shape parameters (GAMLSS). The GAMLSS model has many simplified forms, such as when $k=1,2,\cdots,p$, $J_k=0$, it can be simplified as:

$$g_k\left(\boldsymbol{\theta}_k\right)=\boldsymbol{\eta}_k=\mathbf{X}_k\boldsymbol{\beta}_k \tag{2}$$

This article adopts simplified equation (2). For a general distribution, the first two parameters generally represent position and scale, represented by μ and σ respectively.

The estimation of parameters is carried out using empirical Bayesian methods. These algorithms are relatively mature and can be implemented using the R software package gamlss. The GAMLSS model has been widely applied in many fields in recent years[7-8].

3. Empirical analysis

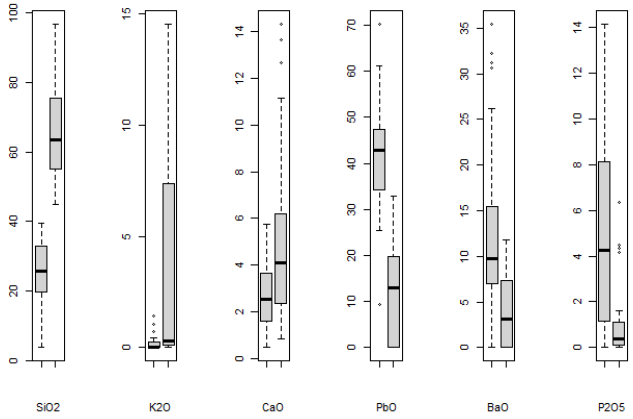

Figure 1: Box plot of the proportion of six main components

In question C of the 2022 National Mathematical Modeling Competition, there is a batch of relevant data on ancient Chinese glass products[9]. Archaeologists have classified these cultural relics into two types based on their chemical composition and other detection methods: high potassium glass and lead barium glass. There are a total of 58 cultural relics and their classification information, including patterns, colors, surface weathering, and the proportion of corresponding main components for each cultural relic. The cumulative sum of the proportions of each component should be 100%, but due to detection methods and other reasons, the cumulative sum of the proportions of its components may not be 100%. This article considers 67 data points between 85% and 105% of the cumulative composition ratio as valid data.

It can be clearly seen from Figure 1 that there is a significant difference between the main components of the weathering point and the main components of the no weathering point. The presence or absence of weathering has a certain impact on each component.

3.1 Data processing and analysis

The surface weathering of glass is likely related to the type, color, texture, environment, a small amount of data, etc. Fisher's exact test[10] is used for classification data, with a significance level of 5%. The P-values of the correlation measurement between glass surface weathering, decoration, and color are 0.098 and 0.384, both greater than 0.05, indicating poor correlation and accepting the original hypothesis; The P-value between the degree of glass surface weathering and glass type is 0.035 and less than 0.05, rejecting the original hypothesis, indicating a significant correlation between the degree of glass surface weathering and glass type. That is, there is a good correlation between the degree of glass weathering and glass type. Detailed data analysis is shown in Table 1. In addition, the patterns and types were subjected to Cochran Mantel Haenszel tests under surface

weathering conditions[11], and the two were also conditionally correlated.

Table 1: The relationship between various factors and glass surface weathering factors

Various factors	Fisher's exact test P-value	Relevant degree
ornamentation	0.098	Unrelated
colour	0.384	Unrelated
type	0.035	correlation

Table 2: Normal distribution test for each major component

	SiO_2	K_2O	CaO	PbO	BaO	P_2O_5
skewness	0.179	2.066	1.015	0.257	1.622	1.445
kurtosis	2.153	5.700	3.110	2.031	5.447	4.199
Shapiro test	0.963	0.526	0.879	0.929	0.806	0.764

From Table 2, it can be seen that all types of cultural relics have obvious skewness and flat peaks, and the Shapiro test statistic values are significant. From the Shapiro test, it can be seen that each component does not follow a normal distribution.

3.2 Establishment of model

GAMLSS is an effective mathematical method for dealing with the interdependence between multiple variables in no normal distributions. By using this method and statistical software R, a GAMLSS model is established for each component and factor under two different distribution assumptions, and the rationality and coefficient of the model are tested for significance.

GAMLSS model 1with normal distribution:

$$f(y_i) \sim N(\mu_i, \sigma), \ \mu_i = \mathbf{x}_i \boldsymbol{\beta}_1, \ \sigma \text{ is a constant;}$$

GAMLSS model 2 for SEP3 Distribution:

$$f(y_i) \sim SEP3(\mu_i, \sigma_i, \nu, \tau), \ \mu_i = \mathbf{x}_i \boldsymbol{\beta}_2, \ \sigma_i = \mathbf{x}_i \boldsymbol{\alpha}, \ \nu, \tau \text{ are constants;}$$

Among them, $\mathbf{x}_i \left(i = 0,1,\cdots,4\right)$ is a design matrix composed of intercept and decoration X_1 (there are three types of decoration, namely A, B, and C. Decoration A is the basic class), type X_2 (there are two types, namely high potassium glass and lead barium glass, and high potassium glass is the basic class), color X_3 (there are eight colors, namely blue-green, light blue, purple, deep blue, light green, green, dark green, and black, with blue-green being the basic class), and weathering sampling point X_4 (there are two types, either weathered sampling points or no weathered sampling points, and the basic class is weathered sampling points). $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are coefficient vectors for positional parameters, and $\boldsymbol{\alpha}$ is the coefficient vector of the scale parameter.

3.3 Analysis of model results

Using the gamlss package in R, parameter estimates of mathematical models for the relationship between SiO_2 composition and decoration, type, and (no) weathering point were obtained for two different models. The estimation results are shown in Table 3. The C coefficient of pattern in model 1 and model 2 is not significant, and the estimated values of other parameters are significant and not

zero. The estimated values of the parameters in model 2 are τ =0.789 and ν =59.237, both of which are significantly non-zero, and the connection functions are logarithmic.

The standard for evaluating which distribution fits well is AIC=-2 logarithmic likelihood function value+2 * number of parameters. It can be seen that the smaller the AIC value, the better the fitting. The AIC values of the normal distribution and SEP3 distribution are 483.312 and 465.961, respectively. Obviously, the GAMLSS model established by the SEP3 distribution fits better. In addition, the residual analysis of SiO_2 content in model 1 and model 2 is shown in Figure 2. It can be seen that the randomized residuals of the SEP3 distribution fluctuate within the range of [-2, 2], and there is no obvious trend. The distribution shape of the residuals approximates the standard normal distribution, and the QQ plot of the residuals is almost a 45 degree straight line. This indicates that the SEP3 distribution has a good fitting effect on SiO_2 content data with heteroscedasticity.

Table 3: Parameter estimation results of two models for SiO_2 composition

	Model 1		Model 2	
	Estimated value of $\boldsymbol{\beta}_1$	Estimated value of $\log(\sigma)$	Estimated value of $\boldsymbol{\beta}_2$	Estimated value of $\boldsymbol{\alpha}$
Intercept term	36.76	8.152	42.376	2.577
Decoration (B)	25.113		18.370	-1.688
Decoration (C)	-2.599		-0.238	0.336
Type (lead barium)	-9.444		-15.077	
Weathering free sampling points	32.091		34.406	
AIC	483.312		465.961	

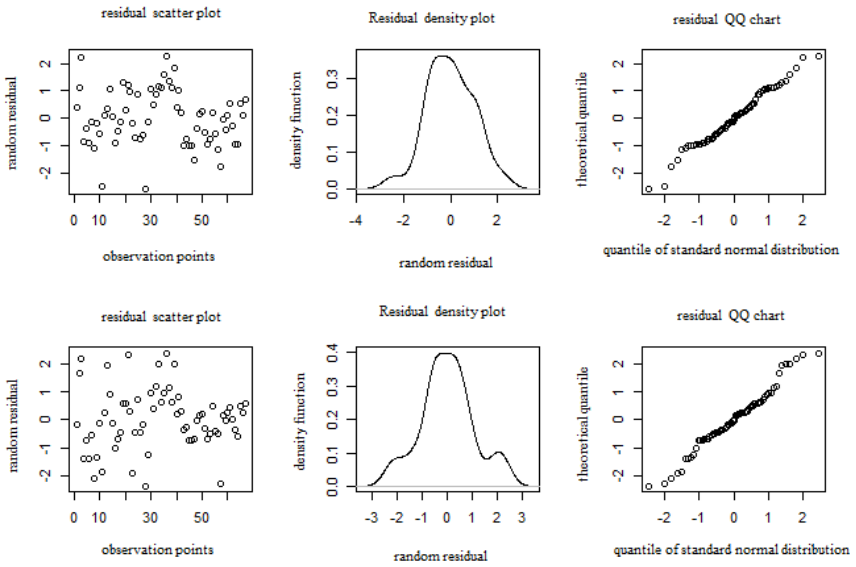

Figure 2: Residual distribution diagram of SiO_2 model 1 and model 2

The mathematical model of the relationship between the component CaO in cultural relics and patterns, types, and (no) weathering points under two different models is as follows:

Linear regression model 3 with normal distribution:

$$f(y_i) \sim N(\mu_i, \sigma), \quad \mu_i = \mathbf{x}_i \boldsymbol{\beta}_1, \quad \sigma \text{ is a constant;}$$

GAMLSS model 4 with normal distribution:

$$f(y_i) \sim N(\mu_i, \sigma_i), \mu_i = \mathbf{x}_i \boldsymbol{\beta}_2, \sigma_i = \mathbf{x}_i \boldsymbol{\alpha};$$

Among them, $\mathbf{x}_i (i = 0,1,\cdots,4)$ is consistent with the design matrix in the previous SiO_2 composition model. $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are the coefficient vectors of the position parameter, and $\boldsymbol{\alpha}$ is the coefficient vector of the scale parameter. The coefficient of pattern in model 3 and model 4 is similar to model 1 and model 2. The AIC values of the two models with normal distribution are 278.290 and 259.625, respectively. Obviously, the GAMLSS model established with normal distribution fits better.

4. Comparison of component estimation for glass artifacts under two different models

In order to compare the estimation performance of the main components of glass artifacts under two different models, this article conducted a detailed analysis of the absolute error of the fitting values of the main components of glass artifacts unearthed in China. After comparison, the absolute errors of each major component under the fitting of model 2 are smaller than the fitting values under model 1, indicating that the GAMLSS model used in this article is effective in solving data with skewness and heteroscedasticity characteristics.

In addition, we also found that the proportion of some components of various types of glass changed before and after weathering, and model 2 can also make good predictions on this issue. The detailed results are shown in Table 4.

Table 4: Predicted values of the main components of each type before and after *weathering*

	SiO_2	K_2O	CaO	PbO	BaO	P_2O_5
Before high potassium weathering	69.191	11.258	5.107	0.387	0.857	1.321
After high potassium weathering	38.296	11.258	6.717	22.215	0.857	5.507
Before lead and barium weathering	58.852	0.220	1.360	20.387	10.950	1.321
After weathering of lead and barium compounds	27.957	0.220	2.970	42.215	10.950	5.507

5. Conclusion

Based on the data of various components of glass artifacts, correlation analysis was conducted on factors such as glass type, decoration, color, and weathering using R software. We confirmed a clear correlation between glass type, decoration, and weathering. The data of the main components of glass relics have skewness and heteroscedasticity characteristics. This article uses SEP3 distribution and normal distribution to fit the main components of glass relics, and establishes a GAMLSS model between each main component and glass type, decoration, color, and weathering, which is an innovative application. This model can be applied to predict the proportion of various components in cultural relics, and plays a certain reference role in the identification of glass cultural relics types.

Acknowledgements

The authors gratefully acknowledge the financial support from the 2023 Higher Education Science Research Planning Project of the China Association of Higher Education - Practice and

Research on University Mathematics Curriculum Based on Project-Based Learning (23SX0411).

References

[1] Cui Jianfeng, Wu Xiaohong, Tan Yuanhui, et al. Composition analysis of ancient glassware unearthed from Chu tombs during the Warring States period in the Yuanshui River Basin of Hunan Province [J]. Journal of Ceramics, 2009, 37 (11): 1909-1913+1918.
[2] Gan Fuxi, Zhao Hongxia, Li Qinghui, et al. Scientific analysis and research on unearthed Warring States period glass products in Hubei Province [J]. Jianghan Archaeology, 2010 (02): 108-116+151+0.
[3] Huang Xiaojuan, Yan Jing, Wang Hui. Scientific analysis and research on silicate beads unearthed from M4 in Majiayuan Warring States Cemetery, Gansu [J]. Spectroscopy and Spectral Analysis, 2015, 35 (10): 2895-2900.
[4] Hu Zhizhong, Li Pei, Jiang Luman, et al. Analysis of Composition and Source of Ancient Glass Materials LA-ICP-MS [J]. Rock and Mineral Testing, 2020, 39 (04): 505-514.
[5] Rigby, R. A., Stasinopoulos, D. M. Generalized Additive Models for Location, Scale and Shape (with Discussion) [J]. Applied Statistics, 2005, 54(3): 507-554.
[6] Stasinopoulos, M., Rigby, B. Generalized Additive Models for Location Scale and Shape (GAMLSS) in R [J]. Journal of Statistical Software, 2007, 23(7): 1-46.
[7] Liu Xinhong, Feng Yuan, Mi Haijie. The Application of GAMLSS Model in Car Insurance Pricing [J]. Mathematical Practice and Understanding, 2017, 47 (11): 1-8.
[8] Mo Shuhong, Li Chenxing, Xing Hua, et al. Research on Annual Runoff of Xiaoli River Basin Based on GAMLSS Model [J]. Journal of Applied Fundamentals and Engineering Science, 2022, 30 (01): 40-49.
[9] Question C of the 2022 National College Student Mathematical Modeling Competition [EB/OL]. The official website of the National College Student Mathematical Modeling Competition, http://www.mcm.edu.cn 2022-9-15/ 2023-3-24.
[10] Fisher, R. A. The logic of inductive inference [J]. Journal of the Royal Statistical Society Series A, 1935, 98, 39–54.
[11] Alan Agresti. Categorical data analysis (second edition) [M].2002, New York: Wiley.