Chemometrics and Intelligent Laboratory Systems 274 (2026) 105731

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemometrics

Polynomial model with *Asymmetric Curvature Induced by Excess* for mixture design: optimization of green anthocyanin extraction from *Cenchrus purpureus* using a standardless chromatographic response

Luiz Henrique Cantarino Adriano [a], Patrícia Abranches Geraldo [a], Cecília Pinto Nogueira [b], Juarez Campolina Machado [b], Marcone Augusto Leal de Oliveira [a,c,*]

[a] Grupo de Química Analítica e Quimiometria – GQAQ, Chemistry Department, Institute of Exact Sciences, Federal University of Juiz de Fora (UFJF), Juiz de Fora, MG, 36026-900, Brazil
[b] Embrapa Gado de Leite, Juiz de Fora, MG, 36038-330, Brazil
[c] National Institute of Science and Technology for Bioanalytics – INCTBio, Institute of Chemistry, University of Campinas (UNICAMP), Campinas, SP, 13083-970, Brazil

ARTICLE INFO

Keywords:
Binary mixture modeling
Alternative mixture model
Anthocyanin extraction
Elephant grass (*Cenchrus purpureus*)
Scheffé mixture models

ABSTRACT

Canonical Scheffé mixture models are widely used for binary systems, but their inherent assumption of symmetric component behavior restricts their ability to describe processes in which mixture components fulfill distinct physicochemical roles. This limitation is especially relevant for solvent-based extraction systems involving water and ethanol. In this study, the extraction of anthocyanins from leaves of *Cenchrus purpureus*. using water:ethanol mixtures were investigated, and an alternative modeling framework, named Asymmetric Curvature Induced by Excess (ACIE), was proposed to overcome this conceptual constraint. Experimental and simulated approach were evaluated in this work. Extractions were carried out by static maceration at room temperature using only water and ethanol, in accordance with Green Chemistry principles. The extracts were analyzed by HPLC-DAD, and a semi-quantitative approach was adopted in which the response variable was defined as the sum of the chromatographic peak areas of the two most reproducible anthocyanin signals. Canonical Scheffé linear and quadratic models were unable to adequately represent the system, exhibiting lack of fit and limited predictive capability. In contrast, ACIE, by incorporating asymmetric curvatures for modeling the excess of each component, achieved superior statistical performance and a more realistic representation of the response surface. The model predicted an optimal extraction region between 82 and 92% ethanol, consistent with the distinct roles of ethanol in promoting cell wall disruption and water in anthocyanin solubilization. These results show that ACIE effectively addresses a fundamental limitation of Scheffé-based models relative to asymmetries in binary mixtures and provides a flexible, chemically interpretable framework that may be applicable to other systems characterized by asymmetric component functionality.

1. Introduction

A valuable tool to maximize extraction efficiency and optimize solvent composition is mixture design. Mixture design is a specific class of experimental design in which the independent variables are proportions of different components of a mixture, rather than absolute quantities. Unlike traditional factorial designs, where the factors vary independently, in mixture designs the sum of all components is constrained, making it particularly suitable for studies in which the relative proportions of solvents are critical [1]. The application of mixture design in solvent extraction processes has been widely reported in the literature, with Scheffé canonical models being the most commonly adopted framework [2–6]. This approach allows a systematic and efficient evaluation of how different solvent combinations affect extraction yield and selectivity, enabling the identification of optimal solvent ratios that maximize the recovery of target compounds, mixture design significantly reduces the number of experimental trials compared to traditional optimization methods, saving time and resources and promoting a more sustainable research approach [7,8].

This article is part of a special issue entitled: BR School of Chemometrics published in Chemometrics and Intelligent Laboratory Systems.

* Corresponding author. Grupo de Química Analítica e Quimiometria – GQAQ, Chemistry Department, Institute of Exact Sciences, Federal University of Juiz de Fora (UFJF), 36026-900, Juiz de Fora, MG, Brazil.
E-mail address: marcone.oliveira@ufjf.br (M.A.L. de Oliveira).

https://doi.org/10.1016/j.chemolab.2026.105731
Received 23 December 2025; Received in revised form 26 February 2026; Accepted 16 April 2026
Available online 26 April 2026
0169-7439/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Abbreviations

ACIE	polynomial model asymmetric curvature induced by excess
DAD	diode array detector
HCl	hydrochloric acid
IQR	interquartile range
LOOCV	leave-one-out cross-validation
MAE	mean absolute error
PMCC	polynomial model with central curvature
R^2	coefficient of determination
RMSE	root mean square error
RSD	relative standard deviation
SMC	polynomial model with symmetric marginal curvature

Despite their widespread use, canonical Scheffé mixture models are known to have several theoretical and practical limitations that have encouraged the development of alternative approaches for mixture experiments [9]. Early critiques pointed out that these models struggle to describe simple linear blending behavior, especially in cases where one component mainly acts as a diluent or contributes in an almost additive, inert way [10]. The lack of pure quadratic terms also makes the interpretation of model coefficients less intuitive, since the parameters are linked to pure-component responses rather than clearly describing how changes in individual components affect curvature [11,12]. These issues tend to become more evident in replicated experiments or in systems where the response changes rapidly over limited regions of the simplex, which can lead to an inadequate local description of the response surface [11].

To address these shortcomings, a variety of alternative mixture modeling strategies have been proposed with the goal of easing the rigid structural constraints of the canonical framework. Methods such as generalized blending models [9], reparametrization based on reference mixtures, and partially quadratic formulations [13] offer greater flexibility while still maintaining interpretability and parsimony. Other developments have also explored the use of nonlinear transformations [14] or the selective addition of curvature terms to better capture localized effects and directional changes within the simplex [15,16]. As seen, it can be stated that these efforts reflect an increasing awareness that, in practical applications, particularly when mixture components play distinct functional roles, more flexible formulations can provide a more realistic and informative description than canonical models that enforce uniform, symmetric behavior across the entire composition space.

Anthocyanins are natural pigments belonging to the flavonoid class, widely distributed throughout the plant kingdom. These compounds are responsible for the vibrant colors observed in a variety of plant tissues, providing hues ranging from orange and red to pink, violet, and blue. Their abundance in fruits, flowers, and leaves contributes not only to the visual appeal of plants but also plays essential ecological roles, such as attracting pollinators and seed dispersers [17]. In addition to their aesthetic and ecological functions, anthocyanins have gained significant attention due to their potential health benefits, particularly their antioxidant and anti-inflammatory properties. From a human health perspective, the consumption of anthocyanin-rich foods has been associated with several beneficial effects, including the management of obesity, diabetes, and neurodegenerative diseases. Additionally, anthocyanins may provide cardiovascular protection and exhibit anti-cancer activity [18,19]. These health benefits, combined with the increasing demand for natural and sustainable products, have driven the use of anthocyanins as natural colorants in the food industry, offering an alternative to synthetic dyes [20].

Including solvent proportion, several factors can influence the efficiency of anthocyanin extraction, as widely reported in the literature, including solvent type, extraction time, temperature, pH, and sample preparation methods. Among these variables, pH control is particularly critical, as anthocyanins are significantly more stable under acidic conditions, with maximum stability typically observed at pH values close to two [21]. Maintaining the extraction medium within this acidic range helps preserve the flavylium cation form of anthocyanins, thereby minimizing degradation and structural transformation during extraction. A broad range of solvents has been employed for anthocyanin extraction [22], yet mixtures of ethanol and water are often preferred due to their lower environmental impact compared to pure organic solvents [23–25]. Various extraction techniques have also been investigated, such as ultrasound-assisted extraction and microwave-assisted extraction. Although these methods can enhance extraction efficiency, static maceration remains a valuable approach; despite not achieving the highest yields, it aligns with green chemistry principles, requires minimal to no energy input, and offers greater scalability from laboratory settings to real-world applications due to its simplicity [25,26].

Natural sources such as fruits and leaves are typically used for anthocyanin extraction; however, reliance on these materials presents constraints including high production costs, seasonality, and extensive land use. Consequently, alternative sources are needed to ensure a more cost-effective, sustainable, and year-round supply of these compounds. An important aspect of sustainable anthocyanin production is the selection of renewable, high-yield plant sources. In this context, elephant grass [*Cenchrus purpureus* (Schumach.)], a tropical forage species of African origin, has emerged as a promising candidate. Widely cultivated in tropical regions such as Brazil, it exhibits rapid growth (harvestable within 3-4 months) and high biomass productivity, reaching up to 50 tons of dry matter per hectare per year [27]. The species also shows high photosynthetic efficiency, persistence, rapid regrowth, broad adaptability, and the ability for biological nitrogen fixation [28]. Recent studies have highlighted purple genotypes of elephant grass, whose coloration results from anthocyanin accumulation in leaves and stems. The purple cultivar has been reported to contain approximately 0.76 mg of anthocyanins per 100 g of biomass [29], indicating that elephant grass may serve as a sustainable alternative source of these pigments, avoiding the use of eatable fruit as raw material for extraction.

In this context, a mixture design approach was adopted to investigate a green extraction process for anthocyanins from purple elephant grass using exclusively water and ethanol as solvents and avoiding energy-intensive or auxiliary techniques. Alongside the application of canonical Scheffé mixture models, this study explores alternative modeling strategies aimed at overcoming some of the inherent limitations of the traditional framework, particularly in systems where mixture components may play distinct physicochemical roles. By integrating statistical modeling with chemical reasoning, the proposed approach seeks to evaluate whether increased flexibility in the response surface formulation can improve both model performance and interpretability within the context of solvent-based extraction processes.

2. Materials and methods

2.1. Reagents and solutions

All reagents used for analysis were of analytical grade. Ethanol and methanol was purchased from Neon Comercial (Suzano, SP, Brazil). Formic acid and hydrochloric acid (HCl) were purchased from Vetec (Rio de Janeiro, RJ, Brazil). Water was purified by deionization (Milli-Q system; Millipore, Bedford, MA, USA).

2.2. Plant material

The genotype T44.1 of. was selected for this study due to its elite status within the Embrapa elephant grass breeding program. This genotype was chosen based on its high biomass production and the intense purple coloration observed in its flowers and stems.

Individual plants were established at the Embrapa Gado de Leite headquarters, located in Juiz de Fora, Minas Gerais, Brazil (21°46′57′′S; 43°23′13′′W; 1041 m above sea level). Cultivation followed technical recommendations for elephant grass farming.

A total of 2 kg of leaves were harvested, chopped, and stored fresh in plastic bags at −50 °C until extraction. From this material, 200 g were randomly selected for processing. To increase reproducibility and representativeness of the analysis methods, the selected material was cut into strips using scissors and further reduced in size with an electric food processor.

2.3. Experimental design and statistical modeling

A simplex-centroid mixture design was applied to optimize the extraction of anthocyanins. The two components studied were water acidified with HCl 5%, coded as x_1 and further referred simply as 'water', and ethanol, coded as x_2. The experimental matrix consisted of five binary combinations among the components with proportion ranging from 0% to 100% in increments of 25% in each level. Each experimental condition was tested in triplicate as represented in Table 1.

2.4. Extraction procedure

Approximately 5 g of elephant grass leaf material was weighed and transferred to a suitable vessel. A total of 50 mL of extraction solvent was added, consisting of varying proportions of water acidified with HCl 5% and ethanol, according to the mixture design described in **Section 2.3**. The extraction was performed by static maceration at room temperature (approximately 25 °C) for 24 h, under protection from light. After extraction, the mixtures were filtered through qualitative filter paper and concentrated under reduced pressure using a rotary evaporator at 40 °C to remove ethanol completely.

The ethanol-free extracts were transferred to Falcon tubes or round-bottom flasks, covered with aluminum foil, sealed with Parafilm, and stored at −20 °C until freeze-drying. Lyophilization was conducted for 5 days using an Alpha 1-2 LD Plus freeze-dryer (Christ®), under vacuum at −55 °C. The lyophilized extracts were reconstituted in water acidified with 1% formic acid, and the volume was adjusted to 5 mL in a volumetric flask to ensure standardization. The solutions were centrifuged at 8000 rpm for 3 min and subsequently filtered through a 0.22 μm microfilter before storage at −50 °C for further analysis.

2.5. HPLC-DAD analysis

The extracts were analyzed using a HPLC system equipped with a diode array detector (DAD) (Agilent Technologies, SMC260 Infinity II). Separation was achieved on an Agilent Eclipse Plus C18 column (100 mm × 4.6 mm, 3.5 μm particle size), maintained at 50 °C. The mobile phase consisted of water acidified with 1% formic acid (Solvent A) and methanol (Solvent B).

As proposed by Šimerdová et al. [30], the flow rate was set at 0.8 mL/min with the following gradient program: 0 min: 5% B, 0-4 min: 5-20% B, 4-8 min: 20-25% B, 8-10 min: 25-90% B, 10-10.15 min: 90% B, 10.15-10.30 min: 90-5% B, 10.30-13 min: 5% B.

The injection volume was 15 μL without dilution. The DAD was set to acquire spectra in the range of 190 to 600 nm, with chromatograms monitored at 520 nm, which corresponds to the absorption maximum for anthocyanins. The data were processed using Agilent ChemStation, and the areas of the main anthocyanin peaks were integrated. No external standards were used, and the response variable was expressed in absorbance units (mAU·s).

Table 1
Proportion of water and ethanol defined for the experimental design.

Sample ID	Experimental level	Proportion of water (x_1)	Proportion of ethanol (x_2)
S1	A	1	0
S2	A	1	0
S3	A	1	0
S4	B	0.75	0.25
S5	B	0.75	0.25
S6	B	0.75	0.25
S7	C	0.5	0.5
S8	C	0.5	0.5
S9	C	0.5	0.5
S10	D	0.25	0.75
S11	D	0.25	0.75
S12	D	0.25	0.75
S13	E	0	1
S14	E	0	1
S15	E	0	1

3. Results and discussion

3.1. Chromatographic profile

The samples were analyzed by HPLC without the need for prior dilution. Fig. 1 presents the overlaid chromatograms obtained from these analyses.

The chromatogram reveals five distinct peaks (numbered 1-5), with the major compound eluting at a retention time of approximately 7.497 min. Although four minor compounds were also detected, their chromatographic signals were occasionally distorted due to their low concentrations in the extracts. Even so, the UV-Vis spectra obtained for all five peaks (Fig. 1 of the Supplementary Material) exhibit the characteristic absorption bands around 280 nm and 520 nm that are diagnostic of anthocyanin chromophores, with some spectra additionally showing a band in the 310-340 nm region, commonly associated with acylated anthocyanin derivatives [31]. Focusing on the major peak, evaluation of its UV-Vis spectrum shows a ratio of 0.29 between the absorption bands at 440 nm and 520 nm, which according to Saha et al. [31], together with the wavelength of maximum absorption, is characteristic of a specific anthocyanin. The absence of an absorption band in the 310-340 nm region, typically related to acylation, further indicates that this compound corresponds to a non-acylated anthocyanin. Taken

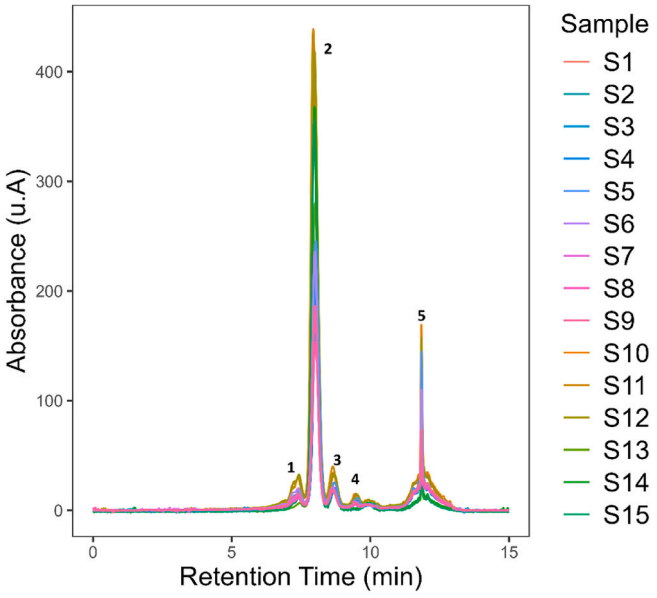

Fig. 1. Overlaid chromatograms corresponding to analyses from a mixture design experiment conducted at five levels with three replicates each. The profiles represent anthocyanin separation monitored at 520 nm, allowing comparison of retention behavior and sign. Each number (1-5) is representative of anthocyanins species detected in the analytical run.

together, these spectral features support its assignment as cyanidin-3-glucoside, what is in agreement with literature reports [32]. Wang et al. [33] recently identified cyanidin-3-glucoside as the dominant anthocyanin in purple elephant grass, where it accounted for approximately 89% of the total anthocyanin content.

3.2. Mixture design analysis

Having established the general anthocyanin profile present in the elephant grass extracts, we proceeded to apply the mixture design approach to develop a regression model capable of estimating the optimal proportion between water and ethanol, aiming to define the most effective extraction condition for the studied material. The first step in this modeling process is to determine the appropriate response variable *y*. Ideally, since the goal of extraction is to maximize anthocyanin yield, the concentration of extracted anthocyanins would serve as a logical choice for *y*. However, this study presents an additional layer of complexity since it is not possible to determine the absolute concentrations of the compounds due to the unavailability of the necessary analytical standards.

Therefore, a semi-quantitative approach was adopted, in which the peak area was used as the best available estimate of analyte concentration in the extracts. For a fixed wavelength, the area under each chromatographic peak is directly proportional to the analyte's concentration. These areas were automatically calculated by the software Agilent ChemStation using the 'Agile 2' integration mode. While this approach does not permit the determination of the exact quantity of each anthocyanin, it allows for reliable comparison of total anthocyanin content between extracts and assessment of the relative proportions of the detected species.

The initial proposal for this optimization study was to use the sum of the areas of all detected anthocyanin peaks. However, upon analyzing the data, natural variability was observed in the measured areas across replicates within the same sample level. These variations are expected and arise from a combination of random errors inherent to every stage of the process, from sample collection to instrumental detection, as well as intrinsic challenges associated with analyzing plant extracts. Each replicate involved unique plant material and was extracted individually under authentic conditions. Additionally, extraction processes themselves have a range of efficiencies that contribute to variability across samples. Therefore, each extract is inherently unique due to differences in plant material and individualized extraction procedures.

There is no specific regulatory standard in the literature for addressing the issue described above. Therefore, a relative standard deviation (RSD) threshold of 15% was adopted as the cutoff for acceptable reproducibility of peak areas. This criterion was based on RDC No. 27/2012 issued by Brazil's National Health Surveillance Agency (ANVISA) and the Bioanalytical Method Validation Guidance published by the U.S. Food and Drug Administration (FDA) in May 2018 [34,35]. While these guidelines are intended for bioanalytical samples and not specifically for plant matrices, their applicability is justified by the comparable complexity of both sample types.

Fig. 2 displays the RSD values for each of the five peaks across the five proposed sample levels. The dashed red line in the graph represents the 15% threshold established earlier. As shown, peaks 1, 4, and 5 exhibit RSD values that exceed this limit in certain replicates, rendering their areas unsuitable for inclusion in the regression model. This elevated variability is likely due to the smaller absolute areas of these peaks, which are more susceptible to fluctuations caused by instrumental noise. Moreover, peaks with smaller areas are more sensitive to small variations relative to the group's mean, given that RSD is a relative, not absolute, measure.

Considering the data presented in Fig. 2, peaks 2 and 3 were the only ones that met the previously established reproducibility criterion. These peaks also correspond to the major constituents in the chromatogram. Therefore, it is considered that using only these two peaks provides a

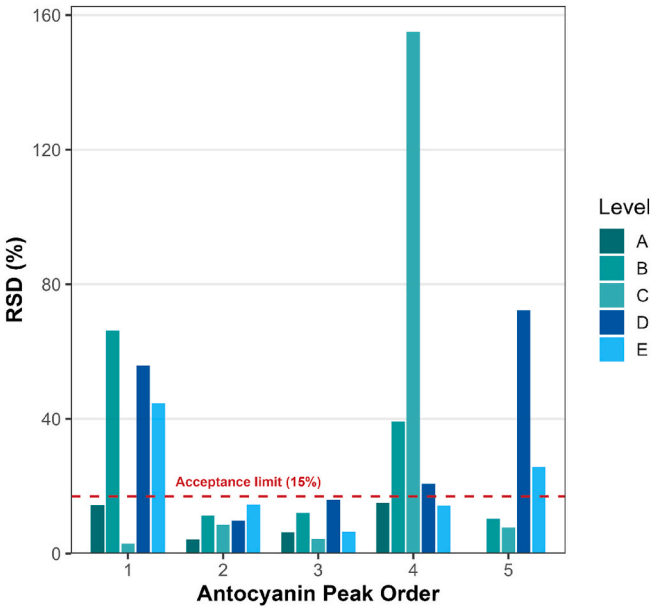

Fig. 2. RSD (%) values for five anthocyanin peaks obtained under five experimental conditions (levels A-E). Each bar represents the variability associated with a given condition for a specific anthocyanin peak. The horizontal dashed red line denotes the acceptance limit of 15% RSD, used as a criterion to evaluate analytical precision. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

reliable representation of the extraction conditions, while avoiding the introduction of unnecessary noise associated with the variability of minor components in subsequent modeling. From this point forward, the response variable *y* in the regression model was defined as the sum of the areas corresponding to peaks 2 and 3 in the chromatograms of each sample. With this definition in place, the regression modeling process could begin. In line with the principle of parsimony, an initial additive linear model was fitted, followed by a quadratic model, in an effort to better understand the behavior of the system under study. The dataset used for constructing these models, as well as those described subsequently, is presented in Table 1 of the Supplementary Material, where x_1 and x_2 represent the experimental variables corresponding to the proportions of acidified water and ethanol, respectively, and *y* denotes the sum of the areas of peaks 2 and 3.

Prior to fitting a parametric regression model, the normality of the response variable was assessed using the Shapiro-Wilk test, which returned a p-value of 0.065. As a result, the null hypothesis was not rejected with a $\alpha = 0.05$, and the response variable was considered to follow a normal distribution. Based on this assumption, two of Scheffé's canonical models the binary linear and quadratic models were evaluated. In this study, however, the most common canonical models proved insufficient to accurately describe the relationship between anthocyanin content and the solvent proportions.

This limitation became evident through lack-of-fit tests, where the linear model yielded a $F_{calculated}$ of 21.802 compared to a $F_{critical}$ value of 8.23, and the quadratic model produced an $F_{calculated}$ of 30.72 against a $F_{critical}$ value of 5.46. The results obtained demonstrate that linear model and quadratic model were both insufficient to capture the nuances of the extraction process. This limitation was particularly evident due models lack of fit and also in the lack of significance of the cross-interaction coefficient in the quadratic model, indicating that a simple linear or even quadratic regression cannot adequately represent the interaction between the solvents in the extractive solution and the anthocyanins present in the plant material. The summary for all models is in Table 2 of the Supplementary Material.

The usual progression for mixtures designs, inside Scheffé's

canonical model framework, would be the special cubic model. However, for binary mixtures the special cubic model collapses to the quadratic form due to the absence of a third component x_3. Without this third dimension, the cubic interaction term cannot be formed, and therefore the model is mathematically identical to the quadratic one. As a result, it fails to account for the additional curvature and interaction complexity observed in the experimental data.

Since Scheffé's framework does not provide a direct way to overcome this limitation for binary systems, an alternative strategy was adopted, based on transforming regular polynomial models into forms that are suitable for modeling mixture systems. To this end, regular polynomial models were first proposed with the inclusion of different forms of quadratic terms. These models were then reparametrized by imposing the mixture constraint, in a manner analogous to the procedure used in Scheffé's canonical models. The resulting formulations are hereafter referred to as the polynomial model with Symmetric Marginal Curvature (SMC), the polynomial model with Central Curvature (PMCC), and the polynomial model with Asymmetric Curvature Induced by Excess (ACIE). For the adjustment of some of these models, an additional experimental point at a 90:10 ethanol:water composition was included. This experiment was performed in triplicate and, for comparison purposes, all models were recalculated considering this addition. The resulting design matrix has full column rank for the five-parameter, no-intercept ACIE model, ensuring that all coefficients are identifiable from the available compositions. Moreover, the inclusion of vertex and asymmetric interior points (including 90:10 ethanol:water) provides sufficient independent information to estimate the higher-order excess terms without undue collinearity. Finally, the design includes replicated runs at each main composition, yielding substantial residual degrees of freedom and pure-error estimation, supporting robust parameter estimation. The original regular polynomial, the additional term, and the resulting mixture polynomial are presented in Table 2. The proofs for all transformations are in Supplementary Material.

For mixture models, quadratic terms such as x_1^2 and x_2^2 are introduced to allow the contribution of each component to vary nonlinearly across the experimental region. By incorporating these terms, the model can represent situations in which changes in the proportion of a given component produce effects on the response that are not strictly proportional. The associated coefficients indicate whether the individual influence of each component intensifies or weakens as its fraction in the mixture increases. Because these terms enter the model in an additive manner, they describe marginal curvature effects linked to each component separately, without explicitly accounting for their combined action. This formulation provides a foundation for capturing deviations from linearity associated with each component, as pursued in SMC.

Building on this idea, additional flexibility is achieved by including asymmetric interaction terms such as $x_1^2x_2$ and $x_1x_2^2$. These terms allow the model to distinguish situations in which the excess of one component affects the response differently than the excess of the other. Specifically, $x_1^2x_2$ emphasizes regions where x_1 dominates the mixture while x_2 remains present in smaller amounts, whereas $x_1x_2^2$ captures the converse situation. By doing so, the model becomes capable of representing asymmetric compositions in which substituting one component for the other does not lead to equivalent responses, a feature that is particularly important when the components play distinct functional roles. Accordingly, these terms are incorporated in ACIE to enable a more nuanced representation of response behavior under asymmetric mixture compositions.

Within the same framework, the term $x_1^2x_2^2$ serves a complementary role by specifically addressing the behavior of the response in the central region of the simplex, where the component proportions are similar. This term reaches its maximum contribution when x_1 and x_2 have comparable values, enabling the model to explicitly control the curvature of the response surface under balanced compositions. A positive coefficient indicates an increase in the response near this central region, while a negative coefficient reflects a decrease, corresponding to a downward curvature. As the mixture moves away from balance toward either extreme, where one of proportions is zero, the influence of this term gradually vanishes. In this way, $x_1^2x_2^2$ complements the asymmetric interaction terms by regulating the response behavior at the center of the simplex, completing the representation of curvature patterns addressed in PMCC and ACIE.

The inclusion of these additional terms expands the descriptive capability of the model when compared with the classical Scheffé polynomials, as it allows the effects of an excess of one component on the response to be explicitly distinguished from those of the other. In Scheffé models, particularly in the canonical quadratic model, deviations from a balanced composition are treated in an essentially symmetric way, meaning that the model responds equivalently to the excess of any component. As a result, mixtures in which one component predominates are represented in a structurally similar manner, regardless of which component is in excess. While this symmetry is appropriate in many situations, it restricts the model's ability to describe asymmetric response surfaces. By incorporating additional terms that break this symmetry, the model becomes capable of capturing variations in the response associated not only with the joint presence of the components, but also with the direction of the imbalance between them, leading to a more flexible and potentially more faithful representation of the effects observed in real mixture systems.

In order to identify the model best suited to describe the studied extraction system, a cross-validation procedure was performed using the experimental data. This statistical method not only assesses how well a model fits the training data but also estimates its capacity to generalize to new, unseen observations. Among the various strategies available, the leave-one-out cross-validation (LOOCV) method was selected. In this approach, 18 models were fitted, each omitting a different experimental point. The predicted response for the omitted point was then compared against its actual value, and model performance was evaluated across all iterations using three metrics: root mean square error (RMSE), mean absolute error (MAE), and the coefficient of determination (R^2). The average results of these metrics are summarized in Fig. 3.

The MAE and RMSE results indicate that the canonical mixture models exhibited the poorest predictive performance, with SMC showing no improvement over the standard quadratic formulation (RMSE = 1112.21, MAE = 926.06). This outcome reflects the limited effectiveness of adding only symmetric marginal quadratic terms, which do not provide additional modeling flexibility beyond that already available in the quadratic Scheffé framework. As a consequence, the inclusion of these terms does not translate into a measurable reduction in prediction error. In contrast, PMCC exhibited a clear improvement in both error metrics (RMSE = 622.71, MAE = 444.55), demonstrating that the inclusion of an additional curvature term yields a more efficient representation of the response surface and leads to a tangible gain in predictive accuracy. The most substantial reduction in prediction error was achieved by ACIE (RMSE = 388.39, MAE = 307.02), indicating that the asymmetric curvature structure introduced by the model is particularly effective in capturing the dominant sources of variability in the data.

Table 2
Regular, curvature, and mixture formulations of the evaluated mixture models.

Model	Regular form	Curvature terms	Mixture form
SMC	$y' = b_0 + b_1x_1^2 + b_2x_2^2 + b_{12}x_1x_2$	$b_1x_1^2, b_2x_2^2$	$y' = b_1^*x_1^2 + b_2^*x_2^2 + b_{12}^*x_1x_2$
PMCC	$y' = b_0 + b_{11}x_1^2 + b_{22}x_2^2 + b_{1122}x_1^2x_2^2$	$b_{11}x_1^2, b_{22}x_2^2, b_{1122}x_1^2x_2^2$	$y' = b_1^*x_1^2 + b_2^*x_2^2 + b_3^{**}x_1x_2 + b_{12}^*x_1^2x_2^2$
ACIE	$y' = b_0 + b_1x_1 + b_2x_2 + b_{112}x_1^2x_2 + b_{122}x_1x_2^2 + b_{1122}x_1^2x_2^2$	$b_{112}x_1^2x_2, b1_{22}x_1x_2^2$	$y' = b_1^*x_1 + b_2^*x_2 + b_{11}x_1^2x_2 + b_{22}x_1x_2^2 + b_{12}x_1^2x_2^2$

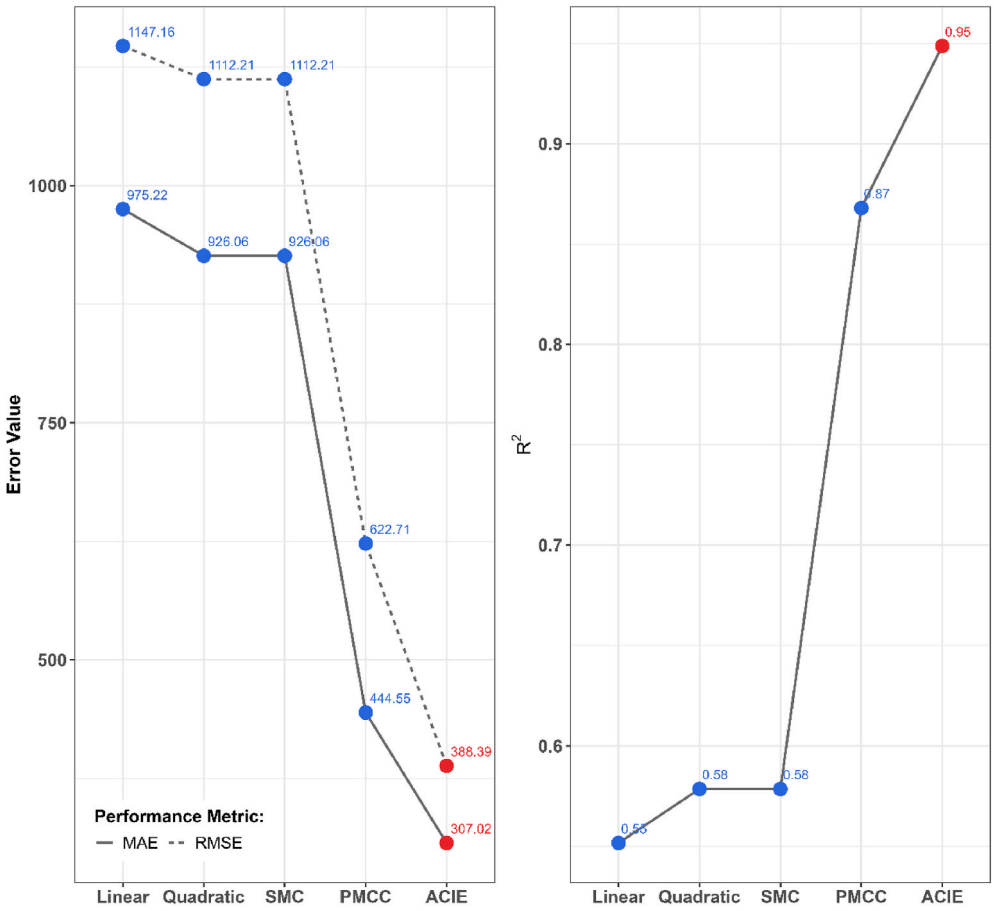

Fig. 3. Comparison of predictive performance among five regression models. The left panel shows MAE (solid line) and RMSE (dashed line), while the right panel presents the R^2. Lower error values and higher R^2 indicate improved model performance, highlighting differences in predictive accuracy across the evaluated models.

This trend is further supported by R^2, with ACIE yielding the R^2 value closest to unity ($R^2 = 0.95$), reinforcing its superior goodness of fit relative to the other models. Importantly, ACIE also passed the lack-of-fit test (p-value = 0.45613), confirming that the improvement in fit is not merely a result of increased model flexibility, but rather reflects a meaningful enhancement in the model's ability to describe the underlying system behavior within the experimental domain. Taken together, these results indicate that while incremental additions of curvature terms may yield limited gains, the inclusion of a more expressive curvature structure, as in ACIE, leads to a consistent and statistically supported improvement in both predictive performance and overall model adequacy. The results for all metrics are available in Table 3 of the Supplementary Material.

To assess the robustness of model comparison with respect to response definition, a sensitivity analysis was performed by refitting all models using an alternative response: the total anthocyanin signal (sum of all detected peaks). Although the inclusion of minor peaks increased response variability and reduced absolute performance metrics, the relative model ranking was preserved, with the ACIE model consistently outperforming the alternative formulations. These results indicate that the superior performance of the ACIE model is robust and does not depend on a specific response definition. The performance metric for all areas is presented in Fig. 2 of the Supplementary Material.

Statistical assumptions of the final model were also thoroughly validated, reinforcing the robustness of the ACIE formulation. The residuals followed a normal distribution, as indicated by the Shapiro-Wilk test (p-value = 0.9968), exhibited homoscedastic behavior according to the Breusch-Pagan test (p-value = 0.2034), and showed no evidence of autocorrelation based on the Durbin-Watson test (p-value = 0.7879). Together, these results indicate that the fundamental assumptions underlying the regression analysis are satisfied, supporting the statistical consistency of the model and its suitability for inference, description, and prediction within the studied extraction system [36].

In terms of complexity control metrics, variance inflation factors (VIF) can be considered as an indicator of potential multicollinearity. However, in mixture modeling, where component proportions are inherently constrained and interdependent, VIF values must be interpreted with caution. As discussed by Montgomery [37], the constrained experimental space in mixture systems naturally induces correlation among regressors, and therefore conventional cutoffs established for unconstrained regression models may not be directly applicable. In Table 4 of the Supplementary Material, we provided those numbers alongside with coefficient significance p-value.

In addition to these diagnostic tests, global goodness-of-fit and predictive performance metrics further support the adequacy of ACIE. The R^2 was found to be very close to the adjusted R^2 ($R^2 = 0.93$), indicating that the inclusion of additional terms does not lead to overfitting and that the explained variance is not artificially inflated by model complexity. This proximity between R^2 and adjusted R^2 suggests that the added curvature terms contribute meaningfully to the model rather than capturing noise. Moreover, the mean absolute percentage error (MAPE = 6.37%) provides complementary evidence of predictive reliability. Values below 10% are generally associated with excellent predictive performance, indicating that the model maintains low relative prediction error across the experimental region [38]. Taken together, these metrics demonstrate that ACIE achieves a balanced trade-off between flexibility and parsimony, offering accurate predictions throughout the simplex while preserving statistical stability and

generalizability.

The final formulation of the ACIE model is presented in Equation (1), while the corresponding response surface predicted by this formulation is shown in Fig. 4. Together, these results provide a comprehensive representation of the system behavior across the entire binary mixture domain, allowing direct visualization of how the predicted response evolves as the solvent composition shifts between water-rich and ethanol-rich regions. The predicted surface is fully consistent with the experimental trends and serves as the basis for interpreting how compositional asymmetries influence extraction performance.

$$y' = (3431,4 \pm 263,8)x_1 + (5376,6 \pm 257,1)x_2 + (24754,1 \pm 4463,5)x_1^2x_2 + (44668,8 \pm 4014,9)x_1x_2^2 - (127803,9 \pm 16016,7)x_1^2x_2^2 \quad \text{(Eq. 1)}$$

The ACIE model coefficients can be interpreted in line with the canonical framework of mixture designs applied to Scheffé systems. The linear coefficients associated with x_1 and x_2 retain their usual chemical interpretation, representing the predicted response at the vertex of the simplex, where one component represent 100% of the mixture. Also, both coefficients are positive, indicating that the anthocyanin content is positively influenced by the presence of either component alone, with neither water nor ethanol individually causing a detrimental effect on extraction. The higher magnitude of the coefficient associated with x_2 suggests that ethanol plays a more prominent role than water in enhancing anthocyanin recovery. Moving to the asymmetric interaction terms, the coefficient of $x_1^2x_2$, which reflects conditions where water is in relative excess, is smaller than that of $x_1x_2^2$, corresponding to ethanol-rich mixtures. This difference indicates that an excess of ethanol affects the response differently from an excess of water and, more importantly, that ethanol-rich conditions are more favorable for increasing the extracted anthocyanin content. Notably, both asymmetric interaction coefficients are positive, meaning that excesses of either component, within their respective regimes, contribute positively to the response. It is not the case for the coefficient associated with the symmetric term $x_1^2x_2^2$, which describes behavior in the central region of the simplex, since it is negative, indicating a reduction in anthocyanin content in balanced mixtures when compared to conditions dominated by one component. These coefficients highlight that anthocyanin extraction is maximized under solvent compositions skewed toward one component, ethanol.

Beyond reproducing the observed response profile, the key strength of ACIE lies in how it achieves this fit: the model is structured to accommodate fundamentally different mixture regimes within a single binary simplex. By incorporating asymmetric curvature terms together with the central curvature term, the model is able to adjust the response independently in three compositionally meaningful regions of the simplex: conditions dominated by excess of water, conditions dominated by excess of ethanol, and conditions where the two components are present in comparable proportions. Rather than enforcing a uniform curvature structure, ACIE allows the response surface to shape locally, adapting to the direction and magnitude of compositional imbalance. This local flexibility is critical in binary systems where different regions of the simplex are governed by different physical and chemical mechanisms rather than forcing a single symmetric response pattern across the entire domain.

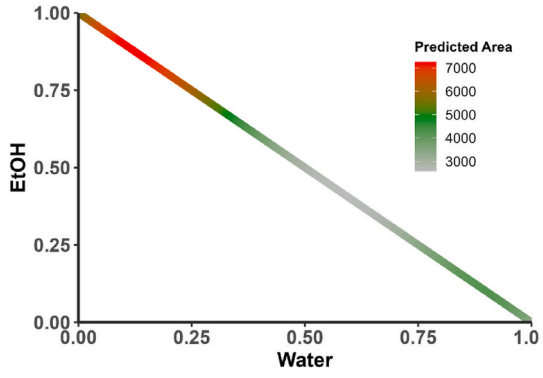

Fig. 4. Predicted response surface generated by the ACIE model, describing the effect of mixture composition on the predicted anthocyanin peak area. The water and ethanol fractions define the mixture composition, while the color scale represents the predicted response in area value. Highest responses are observed toward 82-92% ethanol-rich compositions, indicating the optimum condition for anthocyanins extraction. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

This feature becomes particularly important for the anthocyanin extraction system studied here, in which water and ethanol play fundamentally different functional roles. Water is primarily responsible for solubilizing anthocyanins once they are released, whereas ethanol plays a dominant role in disrupting the plant cell wall, enabling the release of intracellular compounds. Because these roles are not interchangeable, the effect of an excess of ethanol is intrinsically different from the effect of an excess of water. Canonical formulations based on Scheffé's framework impose a strictly symmetric treatment of the mixture components, implicitly assuming that both components contribute to the response in the same way and with the same functional relevance. In the present system, this assumption is not only unrealistic but also detrimental to model performance, as it prevents the model from distinguishing the fundamentally different effects associated with water-rich and ethanol-rich compositions. As a consequence, this enforced symmetry leads to lack of fit in Scheffé-based models, both in their linear and quadratic forms. By contrast, ACIE adopts an explicitly asymmetric formulation that allows each region of the binary simplex to be described according to the specific direction of imbalance, thereby overcoming a central limitation of Scheffé's framework.

The response surface predicted by ACIE, is coded by colors, where the increase of red indicates higher predicted content of anthocyanins at given proportion. Therefore, grey areas in the curve means lower efficiency on the extraction. The obtained result reflects these mechanistic differences in a chemically interpretable way. Extraction efficiency is predicted to decrease near equal water:ethanol proportions, where we suggest that strong solvent-solvent interactions not only reduce the effective affinity of the solvent mixture for anthocyanins but also impair the efficiency of cell-wall disruption. At the other extremes, the model captures additional limitations: at 100% ethanol, the absence of sufficient water limits anthocyanin solubilization, while at 100% water, the lack of ethanol prevents effective disruption of the plant matrix, restricting the release of anthocyanins. ACIE predicts a region of maximum extraction performance in the interval from 82 to 92% ethanol, where ethanol-driven cell-wall disruption is dominant while enough water remains present to maintain solubility. This asymmetric optimum arises naturally from the model structure and aligns with the known physicochemical roles of the two solvents. By capturing these asymmetric, mechanism-driven changes in response across the binary simplex, ACIE goes beyond incremental improvements in fit: it provides a structurally appropriate and practically interpretable framework for binary mixture modeling.

Taken together, these results show that ACIE represents a refinement of existing approaches. By allowing the response to depend not only on composition but also on the direction of compositional imbalance, the model departs from the symmetric assumptions underlying canonical mixture models. This shift constitutes a conceptual advance in the modeling of binary mixtures, particularly for systems in which components play distinct and non-substitutable roles. Although demonstrated here for anthocyanin extraction, the is believed that ACIE framework is broadly applicable to other mixture-driven processes, including solvent-assisted reactions, formulation science, and separation systems. In this sense, ACIE establishes asymmetry as a central and exploitable feature of

binary mixture behavior, rather than a limitation to be overlooked.

3.2.1. Simulation study

Although the predictive performance of the proposed ACIE model was initially assessed using LOOCV, such validation remains inherently internal to the experimental dataset under investigation. While appropriate for small datasets, internal validation alone does not fully address the broader question of how a model may behave when applied to independent systems exhibiting structural characteristics different from those observed experimentally.

To further investigate the generalizability and structural behavior of the proposed model, a structured simulation study was conducted. Simulation provides a controlled framework in which the true data-generating mechanism is known a priori, thereby enabling direct assessment of a model's ability to recover the underlying response surface. Unlike purely empirical validation, simulation allows the systematic control of key features such as symmetry, interaction strength, and local nonlinear effects, while maintaining full control over noise structure and experimental design. In particular, this simulation study was designed to examine whether the advantages observed for the ACIE model in the context of anthocyanin extraction from elephant grass could extend beyond that specific experimental system.

By evaluating model performance across independently generated binary mixture systems with varying structural properties, the analysis begins to address the question of whether the proposed formulation exhibits consistent predictive behavior under alternative response mechanisms. In this sense, the simulation framework serves as an intermediate step between internal validation and broader experimental application, helping to clarify the extent to which the proposed model may be generalizable to other binary mixture system.

To examine whether the performance of the ACIE model observed in the experimental anthocyanin extraction system was specific to that dataset or reflects a more general structural advantage, three independent binary mixture response-generating functions were defined. The purpose of these scenarios was to evaluate whether ACIE maintains consistent predictive behavior when exposed to distinct sources of asymmetry unrelated to the original experimental system. By constructing independent functional forms with controlled structural characteristics, it becomes possible to assess whether the proposed model performs robustly across different asymmetric mechanisms rather than benefiting from a fortuitous match with a single empirical dataset.

The first scenario was based on a quadratic mixture formulation (Equation (2)) and was intentionally selected as a baseline condition, since the quadratic model is correctly specified under this mechanism and is therefore expected to perform optimally. This case serves as a control, ensuring that ACIE does not display artificial superiority when no structural asymmetry is present. The second scenario introduced an additional cubic-type nonlinear component (Equation (3)), generating a moderate degree of asymmetry. In this case, although the quadratic model may still approximate the response reasonably well, it is no longer fully specified, allowing assessment of whether ACIE can accommodate subtle asymmetric deviations. The third scenario incorporated a logistic-type nonlinear modulation (Equation (4)), producing a more pronounced and structurally distinct asymmetry that is inherently more difficult for symmetric quadratic formulations to represent. This final case was designed to evaluate whether ACIE demonstrates a clear performance advantage under stronger asymmetric conditions, consistent with the behavior observed in the experimental application. A representative curvature for each of those models are available in Fig. 3 of the Supplementary Material.

$$f_1(x) = \beta_1 x + \beta_2(1-x) + \beta_{12}x(1-x) \quad \text{(Eq. 2)}$$

$$f_2(x) = f_1(x) + \gamma x(1-x)(x-0.5) \quad \text{(Eq. 3)}$$

$$f_3(x) = f_2(x) + \sigma(k(x-\tau)) \quad \text{(Eq. 4)}$$

Within each of the three data-generating scenarios, a parameter vector ($\boldsymbol{\theta}$) was first randomly sampled to define a specific underlying binary mixture system for that scenario. For the quadratic baseline scenario, the parameter vector was defined as $\boldsymbol{\theta} = [\beta_1, \beta_2, \beta_{12}]$, corresponding to the linear blending coefficients and the quadratic interaction term. For the asymmetric cubic scenario, the parameter vector was defined as $\boldsymbol{\theta} = [\beta_1, \beta_2, \beta_{12}, \gamma]$, where γ controls the magnitude and direction of the additional cubic-type asymmetric contribution. For the asymmetric logistic scenario, the parameter vector was defined as $\boldsymbol{\theta} = [\beta_1, \beta_2, \beta_{12}, \gamma, k, \tau]$, where γ controls the strength of the asymmetric modulation, and k and τ govern the shape and location of the logistic-type transition along the mixture domain.

Given a sampled $\boldsymbol{\theta}$ for a chosen scenario, the noise-free response was evaluated at the mixture compositions defined by the same experimental design adopted in the empirical study, including the additional 90:10 ethanol:water composition and the triplicate structure at each design point. These values define the true underlying response of the system, here denoted as Y_{true}, corresponding to the deterministic output of the generating function without measurement variability. To emulate experimental observations, a random error term (ε) was added to each value of Y_{true} yielding the observed response $Y_{obs} = Y_{true} + \varepsilon$. The error term was assumed to follow a normal distribution with zero mean and constant variance, representing homoscedastic experimental noise. The resulting dataset therefore reproduces the structure of an experimental training dataset in which the measured response arises from the true mixture-dependent signal perturbed by random variability.

The simulated observations Y_{obs} were then used as training data to fit the linear, quadratic and ACIE mixture models using the same mixture proportions as in the experimental study and also in triplicate. After model estimation, predictions were obtained over an independent testing grid spanning the entire binary mixture domain, defined by compositions $x_1 \in [0, 1]$ with uniform increments and $x_2 = 1 - x_1$. These predicted values, denoted Y_{pred}, were compared against the corresponding noise-free responses evaluated at the same grid points. In this manner, predictive performance was quantified by measuring the discrepancy between Y_{pred} and the true underlying response surface Y_{true}. Finally, RMSE, MAE and R^2 were calculated for each model.

The stochastic component introduced in the simulations was assumed to follow a homoscedastic structure across the entire mixture domain. Although heteroscedastic error could be considered in principle, particularly in systems where variability depends on concentration levels, a homoscedastic formulation was adopted here to remain consistent with the assumptions employed in the experimental modeling stage. However, the noise variance was not defined as a fixed absolute value. A fixed error magnitude would disproportionately affect systems with smaller response amplitudes, potentially masking the structural behavior of the generating function and confounding model comparison by introducing variability unrelated to the intrinsic characteristics of the response surface. To avoid this distortion, the noise level was defined proportionally to the magnitude of the underlying signal. Specifically, the error term was defined as $\varepsilon \sim N(\mu = 0, \sigma = 0.10 \times sd(Y_{true}))$, where $sd(Y_{true})$ denotes the standard deviation of the noise-free response evaluated at the training design points. This proportional definition ensures that the signal-to-noise ratio remains comparable across different parameterizations, preventing scenarios in which model discrimination would be dominated by excessive noise rather than structural approximation ability. Although this noise level does not necessarily reproduce the exact magnitude of experimental variability observed in practice, it was considered sufficient to introduce realistic measurement uncertainty while still allowing meaningful comparison of predictive performance among the competing models.

For each scenario, 30 independent parameter vectors $\boldsymbol{\theta}$ were randomly sampled to represent distinct underlying binary mixture systems within the same structural formulation. For each selected parameter set, 100 independent repetitions were performed, differing only in

the realization of the random error term. In each repetition, new error values were generated for all design points, including the triplicate observations, thereby producing 100 independent simulated experimental datasets for the same underlying response surface. Each simulated dataset was subsequently used to fit the competing mixture models, generate predictions over the testing grid, and compute the corresponding performance metrics. This resulted in 100 simulations per $\boldsymbol{\theta}$ set for all three models in every considered scenario. Also, each scenario was evaluated in 30 different conditions. For a model, within a specific, scenario 3000 simulation were performed.

The simulation framework generated a total of 27,000 fitted models across all scenarios, parameter sets, and stochastic repetitions. For each scenario, performance was evaluated within each parameter condition $\boldsymbol{\theta}$, comparing models under identical structural and noise realizations. To summarize the repeated stochastic variability, the median of the 100 repetitions was computed for each condition, resulting in 30 median performance values per model within each scenario. The distribution of these medians for RMSE, MAE, and R^2 is presented in the boxplots shown in Fig. 4 of the Supplementary Material. This approach allows separation of variability due to stochastic noise from variability due to structural differences among parameterizations.

In scenario 1 (quadratic truth), the quadratic model exhibited the best overall performance, as expected given that the true generating function followed a quadratic mixture formulation. The quadratic model achieved a median RMSE of 17.6 (IQR = 7.47), compared to 25.4 (IQR = 13.5) for ACIE. Similarly, the median MAE was 14.7 (IQR = 5.95) for the quadratic model and 21.6 (IQR = 11.7) for ACIE. The median R^2 was 0.998 (IQR = 0.000197) for the quadratic model and 0.996 (IQR = 0.000529) for ACIE. Although the quadratic model was consistently superior in this scenario, the difference was modest, and ACIE demonstrated nearly equivalent predictive accuracy. This behavior is desirable: in a system where the quadratic formulation is correctly specified, it is expected to perform optimally. The fact that ACIE does not artificially outperform the correctly specified model reinforces the internal consistency of the simulation design. At the same time, the comparable performance indicates that ACIE remains stable and applicable even under lower structural complexity, although the quadratic model would naturally be preferred in such a case.

As the structural asymmetry increases, a progressive shift in relative model performance becomes evident. In scenario 2 (asymmetric cubic), where the deviation from symmetry is moderate, the ACIE model clearly outperforms the quadratic formulation. The median RMSE for ACIE was 27.6 (IQR = 22.1), compared to 86.1 (IQR = 55.0) for the quadratic model. Similarly, the median MAE was 23.6 (IQR = 18.7) for ACIE *versus* 76.3 (IQR = 50.2) for the quadratic model, while the median R^2 increased from 0.975 (IQR = 0.112) for the quadratic model to 0.997 (IQR = 0.000861) for ACIE. Although the cubic asymmetry is not extreme, and the quadratic model still retains some capacity to approximate the response surface, its loss of specification leads to both larger absolute errors and substantially greater variability across parameterizations. In contrast, ACIE not only reduces the median prediction error but also exhibits markedly smaller interquartile ranges, indicating greater robustness to structural variation within this moderately asymmetric context.

This contrast becomes even more pronounced in scenario 3 (asymmetric logistic), where asymmetry is stronger and more structurally distinct. In this case, the quadratic model shows a substantial deterioration in performance, with a median RMSE of 121.0 (IQR = 171.0) and MAE of 95.2 (IQR = 144.0), alongside a median R^2 of 0.900 (IQR = 0.376). By comparison, ACIE achieves a markedly lower median RMSE of 62.7 (IQR = 28.2) and MAE of 52.9 (IQR = 22.2), with a higher median R^2 of 0.987 (IQR = 0.0280). Beyond the improvement in central tendency, the reduction in variability is particularly notable: the quadratic model exhibits wide dispersion across parameterizations, whereas ACIE maintains considerably tighter distributions. This behavior indicates that ACIE is structurally more adaptable under pronounced asymmetry, consistently outperforming the canonical quadratic formulation when the true response deviates substantially from symmetric mixture assumptions. The linear model, which performed poorly across all scenarios, is therefore not discussed in detail here, although its results are available in Fig. 4 of the Supplementary Material for completeness.

Because the parameter vectors $\boldsymbol{\theta}$ were randomly sampled within each scenario, the degree of asymmetry was not explicitly controlled. Consequently, even in scenario 3 (asymmetric logistic), some parameterizations may have produced response surfaces that were only moderately asymmetric, which explains why the median performance of the quadratic model does not collapse entirely. In fact, median values alone may suggest that the quadratic formulation remains competitive in certain cases. However, this apparent proximity is largely a consequence of random parameter selection rather than structural adequacy. A closer inspection of the distributions reveals that the stability of the quadratic model under strong asymmetry is markedly poorer: in scenario 3, the R^2 values for the quadratic model exhibit substantial dispersion, with lower tails approaching zero, whereas the ACIE model never produced R^2 values below approximately 0.75. Thus, although central tendency metrics may occasionally appear comparable, the robustness of ACIE under pronounced asymmetry is substantially greater. It is reasonable to expect that, with an increasing number of parameter realizations, the separation between models would become even more evident due to the consistently lower variability and higher structural adaptability of ACIE relative to the quadratic model.

An alternative and more direct way to quantify the predictive gain between models is to evaluate, within each scenario and for each parameter condition, the difference in prediction error between ACIE and the quadratic model. For this purpose, the quantity $\Delta_{RMSE} = RMSE_{ACIE} - RMSE_{quad}$, was computed, so that positive values indicate superior performance of ACIE. This structural behavior becomes clearer when the win-rate analysis and the magnitude of predictive gains are examined jointly. The win-rate results reveal a systematic shift in model superiority across scenarios. In scenario 1 (quadratic truth), the quadratic model outperformed ACIE in all 30 parameter conditions (100%-win rate), confirming that when the data-generating mechanism is truly quadratic, the canonical formulation is correctly specified and performs optimally. In scenario 2 (asymmetric cubic), the balance shifts markedly, with ACIE achieving 26 wins out of 30 conditions (86.7%), while the quadratic model prevailed in only 13.3% of cases. In scenario 3 (asymmetric logistic), the dominance of ACIE becomes even more pronounced, reaching 28 wins out of 30 conditions (93.3%), compared to only 6.67% for the quadratic model. Importantly, this progression reflects not only a change in frequency of superiority but also a clear amplification in the magnitude of predictive differences.

As shown in Fig. 5, scenario 1 is characterized by predominantly negative values, indicating the expected advantage of the quadratic model; in the scenario 2, most Δ_{RMSE} values become positive, demonstrating consistent improvement of ACIE under moderate asymmetry, with moderate gains in magnitude; and in scenario 3, positive Δ_{RMSE} values not only dominate but increase substantially, with several parameterizations exhibiting large predictive gains in favor of ACIE. The increasing dispersion and upward shift of Δ_{RMSE} across scenarios confirm that the predictive advantage of ACIE strengthens as structural asymmetry intensifies.

The simulation study was designed to assess whether the superior performance of the ACIE model observed in the experimental system reflects a transferable structural capability rather than a dataset-specific effect. By exposing the competing models to independently generated mixture systems with distinct sources and intensities of asymmetry, the analysis demonstrated that ACIE behaves consistently across parameterizations and becomes increasingly advantageous as structural deviation from quadratic symmetry intensifies. At the same time, the results

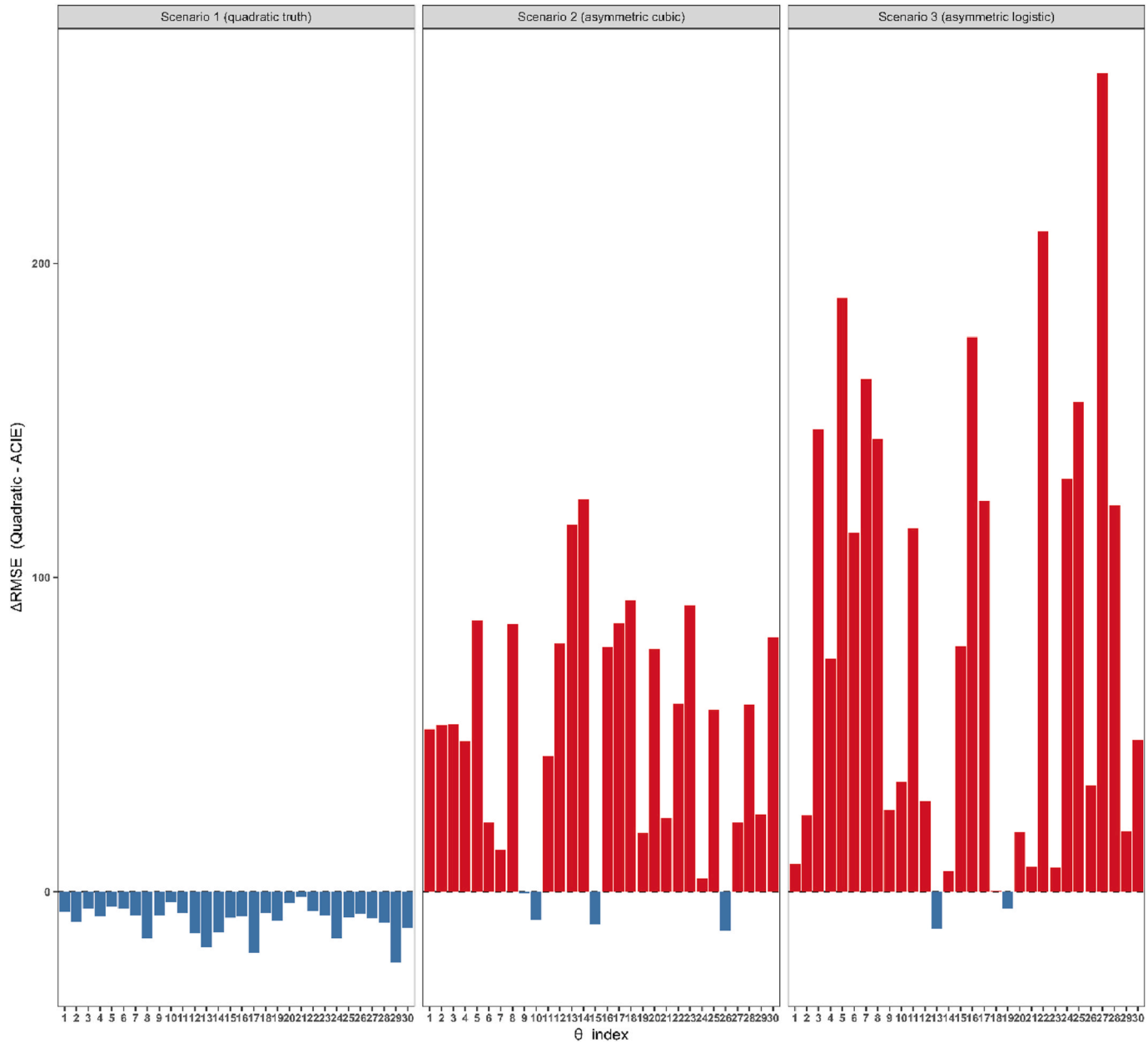

Fig. 5. Distribution of predictive gain between the quadratic and ACIE models across parameter conditions for each simulated scenario. Bars represent the Δ_{RMSE} values computed from the median performance within each parameter set (n = 100 per set). Negative values indicate superior performance of the quadratic model, whereas positive values indicate lower prediction error for ACIE. The three panels correspond to the quadratic (symmetric), asymmetric cubic (moderate asymmetry), and asymmetric logistic (strong asymmetry) scenarios, illustrating how both the frequency and magnitude of predictive differences evolve with increasing structural asymmetry.

confirm that ACIE does not exhibit artificial superiority under correct quadratic specification, thereby reinforcing the internal validity of the comparison framework. Although simulation cannot substitute for independent experimental validation, therefore further application to additional real-world mixture systems remains necessary to fully establish the external generalizability of the model, it provides controlled evidence that the predictive advantage of ACIE is systematic and structurally grounded. These findings support the view that ACIE is particularly well suited and outperform canonical models for binary mixture systems in which asymmetric interaction patterns are expected, while maintaining stable performance under simpler conditions.

Beyond the statistical performance and physicochemical consistency observed, the results of this extraction procedure in this study also demonstrate strong adherence to fundamental principles of Green Chemistry as proposed by Anastas & Warner [39]. The exclusive use of water and ethanol as solvents directly addresses the principle of using safer solvents and auxiliaries, while the absence of energy-intensive assisted techniques aligns with the principle of energy efficiency. In addition, the application of mixture design and experimental design strategies significantly reduced the number of experiments required to characterize the system, contributing to waste minimization and more efficient use of raw materials. From a broader sustainability perspective, these results are also aligned with the United Nations Sustainable Development Goals (American Chemical Society) [40], particularly SDG 12, focused on responsible consumption and production, and SDG 9, which addresses industry, innovation, and infrastructure. This simple, low-energy, and operationally robust approach shows strong potential for industrial adoption, as it facilitates scalability, reduces operational

costs, and lowers barriers to implementation in production lines.

4. Concluding remarks

From a modeling perspective, the results consistently show that the canonical Scheffé models, in both their linear and quadratic forms, were unable to adequately represent the behavior of the system under study. This limitation is not related to the complexity of the phenomenon itself, but rather to a conceptual restriction inherent to these models, the natural imposition of response symmetry, which implicitly assumes that mixture components exert equivalent effects when present in excess. For the extraction system investigated here, in which water and ethanol play distinct physicochemical roles, this assumption proved to be inadequate, leading to lack of fit and reduced predictive capability.

Within this context, the ACIE model represents a significant conceptual advance by explicitly breaking with the symmetry assumption imposed by canonical Scheffé models. By introducing asymmetric curvatures associated with the excess of each component, ACIE allows different regions of the binary simplex to be modeled independently, clearly distinguishing water-rich conditions, ethanol-rich conditions, and intermediate compositions. This structure provides a level of local flexibility that is not available in canonical Scheffé models, enabling the capture of directional variations in the response that more faithfully reflect the physicochemical mechanisms governing the system. As a direct consequence, ACIE exhibited superior statistical performance and a more realistic representation of the response surface, in addition to predicting an extraction optimum between 75 and 90% ethanol. This prediction is fully consistent with the physicochemical roles of the solvents, requiring a high ethanol fraction to promote cell wall disruption while maintaining sufficient water to ensure anthocyanin solubilization. The convergence between the statistical response of the model and the chemical interpretation of the process reinforces the structural adequacy of ACIE for describing the system studied.

We conclude that ACIE constitutes a conceptual evolution relative to canonical Scheffé models by allowing the effects of excess of each component to be treated in an asymmetric and physically coherent manner. This shift in perspective represents a promising contribution in the modeling of binary mixtures, particularly relevant for systems in which components perform distinct and non-interchangeable functions. Although developed and evaluated in the context of binary systems, the proposed framework provides a solid foundation for extensions to multicomponent systems and for future advances in polynomial-based mixture modeling. Beyond its methodological contribution, the application of ACIE in this work enabled the optimization of the extraction system, maximizing the recovery efficiency of anthocyanins, high-value compounds, thereby contributing to the valorization of plant material and aligning the study with principles of Green Chemistry.

CRediT authorship contribution statement

Luiz Henrique Cantarino Adriano: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Patrícia Abranches Geraldo:** Investigation, Writing – original draft, Writing – review & editing. **Cecília Pinto Nogueira:** Investigation. **Juarez Campolina Machado:** Investigation, Methodology, Resources, Writing – review & editing. **Marcone Augusto Leal de Oliveira:** Funding acquisition, Resources, Writing – review & editing.

Declaration of competing interest

The authors have declared no conflict of interest.

Acknowledgements

The authors acknowledge the Coordination of Superior Level Staff Improvement – Brazil (CAPES) – Finance Code 001; Brazilian National Council for Scientific and Technological Development (CNPQ: 445275/2023-5, 406850/2023-2, 303867/2020-5); Funding Authority for Studies and Projects (Finep) (CT-INFRA 01/2013-REF 0633/13); RQ-MG (CEX.RED- 00010-14); National Institute of Science and Technology for Bioanalytic (INCTBio) (FAPESP 2014/50867-3, CNPq 465389/2014-7, 401256/2020-0); State of Minas Gerais Research Support Foundation (FAPEMIG) (APQ-01118-24).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chemolab.2026.105731.

Data availability

Data will be made available on request.

References

[1] J.A. Cornell, Experiments with Mixtures, Wiley, 2002, https://doi.org/10.1002/9781118204221.
[2] V.P. Cavalcanti, S. Aazza, S.K.V. Bertolucci, J.P.M. Rocha, A.D. Coelho, A.J. M. Oliveira, L.C. Mendes, M.M.A. Pereira, L.C. Morais, M.R. Forim, M. Pasqual, J. Dória, Solvent mixture optimization in the extraction of bioactive compounds and antioxidant activities from garlic (Allium sativum L.), Molecules 26 (2021) 6026, https://doi.org/10.3390/MOLECULES26196026.
[3] M.C. DiCiaula, G.C. Lopes, I.S. Scarminio, J.C.P. De Mello, Optimization of solvent mixtures for extraction from bark of schinus terebinthifolius by a statistical mixture-design technique and development of a UV-Vis spectrophotometric method for analysis of total polyphenols in the extract, Quim. Nova 37 (2014) 158–163, https://doi.org/10.1590/S0100-40422014000100026.
[4] A.C. Santos Felix, C.G. Novaes, M. Pires Rocha, G.E. Barreto, B.B. do Nascimento, L. D. Giraldez Alvarez, Mixture design and doehlert matrix for the optimization of the extraction of phenolic compounds from Spondias mombin L apple bagasse agroindustrial residues, Front. Chem. 5 (2018) 305208, https://doi.org/10.3389/FCHEM.2017.00116/BIBTEX.
[5] F. Delaroza, I.S. Scarminio, Mixture design optimization of extraction and mobile phase media for fingerprint analysis of Bauhinia variegata L, J. Separ. Sci. 31 (2008) 1034–1041, https://doi.org/10.1002/JSSC.200700389.
[6] G.L. Scheel, E.D. Pauli, M. Rakocevic, R.E. Bruns, I.S. Scarminio, Environmental stress evaluation of Coffea arabica L. leaves from spectrophotometric fingerprints by PCA and OSC–PLS–DA, Arab. J. Chem. 12 (2019) 4251–4257, https://doi.org/10.1016/J.ARABJC.2016.05.014.
[7] S. Armenta, S. Garrigues, M. de la Guardia, Green analytical chemistry, TrAC, Trends Anal. Chem. 27 (2008) 497–511, https://doi.org/10.1016/J.TRAC.2008.05.003.
[8] M.A. Bezerra, R.E. Santelli, E.P. Oliveira, L.S. Villar, L.A. Escaleira, Response surface methodology (RSM) as a tool for optimization in analytical chemistry, Talanta 76 (2008) 965–977, https://doi.org/10.1016/J.TALANTA.2008.05.019.
[9] L. Brown, A.N. Donev, A.C. Bissett, General blending models for data from mixture experiments, Technometrics 57 (2015) 449–456, https://doi.org/10.1080/00401706.2014.947003.
[10] M.H. Quenouille, Experiments with Mixtures, 1959.
[11] N. Kettaneh-Wold, Analysis of Mixture Data with Partial Least Squares, 1992.
[12] D.R. Cox, A note on polynomial response functions for mixtures. http://biomet.oxfordjournals.org/, 1971.
[13] G.F. Piepel, J.M. Szychowski, J.L. Loeppky, Augmenting Scheffé linear mixture models with squared and/or crossproduct terms, J. Qual. Technol. 34 (2002) 297–314, https://doi.org/10.1080/00224065.2002.11980160.
[14] N.R. Draper, R.C. St John, A mixtures model with inverse terms, Technometrics 19 (1977) 37, https://doi.org/10.2307/1268252.
[15] C. Zhang, H. Peng, D-optimal designs for quadratic mixture canonical polynomials with spline, Stat. Probab. Lett. 82 (2012) 1095–1101, https://doi.org/10.1016/j.spl.2012.02.013.
[16] C. Li, C. Zhang, A-optimal designs for quadratic mixture canonical polynomials with spline, J. Stat. Plann. Inference 207 (2020) 1–9, https://doi.org/10.1016/j.jspi.2019.10.003.
[17] S. Sangeeta, S. Rai, S. Bisht, Mohd Nazim, Chemical properties of anthocyanins sourced from different subtropical fruits, in: Anthocyanins in Subtropical Fruits, CRC Press, New York, 2023, pp. 33–58, https://doi.org/10.1201/9781003242598-2.
[18] E. Pojer, F. Mattivi, D. Johnson, C.S. Stockley, The case for anthocyanin consumption to promote human health: a review, Compr. Rev. Food Sci. Food Saf. 12 (2013) 483–508, https://doi.org/10.1111/1541-4337.12024.
[19] J. He, M.M. Giusti, Anthocyanins: natural colorants with health-promoting properties, Annu. Rev. Food Sci. Technol. 1 (2010) 163–187, https://doi.org/10.1146/annurev.food.080708.100754.
[20] C.T. da Costa, D. Horton, S.A. Margolis, Analysis of anthocyanins in foods by liquid chromatography, liquid chromatography–mass spectrometry and capillary

electrophoresis, J. Chromatogr. A 881 (2000) 403–410, https://doi.org/10.1016/S0021-9673(00)00328-9.
[21] M.Ü. Özgür, E. Çimen, Ultrasound-assisted extraction of anthocyanins from red rose petals and new spectrophotometric methods for the determination of total monomeric anthocyanins, J. AOAC Int. 101 (2018) 967–980, https://doi.org/10.5740/jaoacint.17-0307.
[22] J. Tan, Y. Han, B. Han, X. Qi, X. Cai, S. Ge, H. Xue, Extraction and purification of anthocyanins: a review, J. Agric. Food Res. 8 (2022) 100306, https://doi.org/10.1016/J.JAFR.2022.100306.
[23] C. Capello, U. Fischer, K. Hungerbühler, What is a green solvent? A comprehensive framework for the environmental assessment of solvents, Green Chem. 9 (2007) 927–934, https://doi.org/10.1039/B617536H.
[24] F. Chemat, M.A. Vian, G. Cravotto, Green extraction of natural products: concept and principles, Int. J. Mol. Sci. 13 (2012) 8615–8627, https://doi.org/10.3390/ijms13078615.
[25] L.M. de Souza Mesquita, L.S. Contieri, F.A. e Silva, R.H. Bagini, F.S. Bragagnolo, M. M. Strieder, F.H.B. Sosa, N. Schaeffer, M.G. Freire, S.P.M. Ventura, J.A.P. Coutinho, M.A. Rostagno, Path2Green: introducing 12 green extraction principles and a novel metric for assessing sustainability in biomass valorization, Green Chem. 26 (2024) 10087–10106, https://doi.org/10.1039/d4gc02512a.
[26] J. Azmir, I.S.M. Zaidul, M.M. Rahman, K.M. Sharif, A. Mohamed, F. Sahena, M.H. A. Jahurul, K. Ghafoor, N.A.N. Norulaini, A.K.M. Omar, Techniques for extraction of bioactive compounds from plant materials: a review, J. Food Eng. 117 (2013) 426–436, https://doi.org/10.1016/j.jfoodeng.2013.01.014.
[27] F.M. Ferreira, J.R. Rocha, L.L. Bhering, F.D. Fernandes, F.J. da S. Lédo, J.H. de A. Rangel, M. Kopp, T.M.M. Câmara, V.Q. Ribeiro da Silva, J.C. Machado, Optimal harvest number and genotypic evaluation of total dry biomass, stability, and adaptability of elephant grass clones for bioenergy purposes, Biomass Bioenergy 149 (2021) 106104, https://doi.org/10.1016/j.biombioe.2021.106104.
[28] J.R. Rocha, J.C. Machado, P.C.S. Carneiro, J. da C. Carneiro, M.D.V. Resende, A. Vander Pereira, J.E. de S. Carneiro, Elephant grass ecotypes for bioenergy production via direct combustion of biomass, Ind. Crops Prod. 95 (2017) 27–32, https://doi.org/10.1016/j.indcrop.2016.10.014.
[29] S. Zhou, J. Chen, Y. Lai, G. Yin, P. Chen, K.K. Pennerman, H. Yan, B. Wu, H. Zhang, X. Yi, C. Wang, M. Fu, X. Zhang, L. Huang, X. Ma, Y. Peng, Y. Yan, G. Nie, L. Liu, Integrative analysis of metabolome and transcriptome reveals anthocyanins biosynthesis regulation in grass species Pennisetum purpureum, Ind. Crops Prod. 138 (2019), https://doi.org/10.1016/j.indcrop.2019.111470.
[30] B. Šimerdová, M. Bobríková, I. Lhotská, J. Kaplan, A. Křenová, D. Šatínský, Evaluation of anthocyanin profiles in various blackcurrant cultivars over a three-year period using a fast HPLC-DAD method, Foods 10 (2021) 1745, https://doi.org/10.3390/foods10081745.
[31] S. Saha, J. Singh, A. Paul, R. Sarkar, Z. Khan, K. Banerjee, Anthocyanin profiling using UV-vis spectroscopy and liquid chromatography mass spectrometry, J. AOAC Int. 103 (2021) 23–39, https://doi.org/10.5740/jaoacint.19-0201.
[32] T. Fossen, R. Slimestad, Ø.M. Andersen, Anthocyanins with 4′-glucosidation from red onion, Allium cepa, Phytochemistry 64 (2003) 1367–1374, https://doi.org/10.1016/j.phytochem.2003.08.019.
[33] Z.Y. Wang, P.Y. Lin, C.Y. Hong, K.C.C. Chou, T.J. Lu, Untargeted metabolomics reveals distinct anthocyanin profiles in napier grass (Pennisetum purpureum Schumach.) cultivars, Foods 14 (2025), https://doi.org/10.3390/foods14152582.
[34] ANVISA, Resolução - RDC nº 27, DE 17 DE MAIO DE 2012. Https://Bvsms.Saude.Gov.Br/Bvs/Saudelegis/Anvisa/2012/Rdc0027_17_05_2012.Html, 2012.
[35] FDA, Bioanalytical method validation guidance for industry. Https://Www.Fda.Gov/Regulatory-Information/Search-Fda-Guidance-Documents/Bioanalytical-Method-Validation-Guidance-Industry, 2018.
[36] D.C. Montgomery, Design and Analysis of Experiments, John Wiley & Sons, Inc., 2013.
[37] D.C. Montgomery, S.R. Voth, Multicollinearity and leverage in mixture experiments, J. Qual. Technol. 26 (1994) 96–108, https://doi.org/10.1080/00224065.1994.11979510.
[38] E. Ostertagová, Modelling using polynomial regression, in: Procedia Eng., Elsevier Ltd, 2012, pp. 500–506, https://doi.org/10.1016/j.proeng.2012.09.545.
[39] ECOSChem, Definition and Criteria for Sustainable Chemistry, Created by the Expert Committee on Sustainable Chemistry, 2023. https://global.oup.com/academic/product/green-chemistry-9780198506980. (Accessed 21 December 2025).
[40] American Chemical Society, Chemistry & the UN Sustainable Development Goals - American Chemical Society, (n.d.). https://www.acs.org/green-chemistry-sustainability/education/chemistry-sustainable-development-goals.html (accessed December 21, 2025).