npj | science of food

Article

Published in partnership with Beijing Technology and Business University & International Union of Food Science and Technology

https://doi.org/10.1038/s41538-026-00891-8

Modeling the relationship between sensory attributes and liking using generalized additive mixed models: a case study of plant-based meat alternatives

Check for updates

Sol Kim [1,6], Ju-Eun Nho [2,6], Hyoung Il Son [3,4,5], Min-A Kim [2] & Soo-Jung Kim [1,5]

This study compared two analytical frameworks, partial least squares regression (PLSR) and a generalized additive mixed model (GAMM), to analyze the relationship between sensory attributes and consumer liking in plant-based meat alternatives. Consumer liking was measured using a nine-point hedonic scale, and sensory intensities were obtained using the rate-all-that-apply method. Model performance was assessed using a nested cross-validation framework with multiple validation structures. Across all validation structures, the ordinal GAMM showed higher predictive performance than PLSR, accounting for ordinal ratings and nonlinear relationships. Both approaches identified consistent key sensory drivers of liking, including smoked flavor, meaty flavor, and beany odor. GAMM also revealed curvilinear relationships with optimal intensity ranges and interactions between attributes. Overall, PLSR efficiently summarizes attribute importance under near-linear conditions, whereas GAMM captures nonlinear dynamics and interactions. The complementary use of both approaches enables a more comprehensive understanding of sensory drivers of consumer acceptance.

Developing consumer-oriented food products is a crucial strategy for market success[1]. This perspective encompasses more than a simple assessment of product acceptability; it emphasizes identifying consumer-relevant product characteristics and systematically incorporating these insights into product development, which requires an analytical understanding of the perceptual determinants that affect consumer preferences. Accordingly, in sensory and consumer science, the relationship between sensory attributes and consumer liking must be understood, and the sensory factors that drive liking must be determined[2]. Once sensory–hedonic relationships are empirically established, they offer a basis for constructing predictive models that estimate how targeted modifications of sensory attributes affect consumer acceptance. Such modeling frameworks are useful in product optimization and quality improvement, where systematic adjustments in formulation or processing conditions are necessary to enhance consumer-perceived quality. By quantifying and predicting the associations between sensory attributes and liking, these models guide refinement decisions and contribute to more efficient, evidence-based optimization processes[3].

Partial least squares regression (PLSR) has been widely employed as a representative multivariate analytical method to identify the relationship between sensory attributes and consumer liking. As a technique that addresses multicollinearity among predictor variables, PLSR enables the prediction of consumer liking (Y) from sensory attributes matrices (X). Owing to these methodological strengths, PLSR has been applied across diverse product categories, including milk, cheese, tea, and baked products, to elucidate how specific sensory attributes contribute to consumer liking[4–8]. In addition to modeling the association between X and Y, PLSR identifies predictor variables that best explain the variation in consumer liking[9]. Collectively, these

[1]Department of Integrative Food, Bioscience and Biotechnology, Chonnam National University, 77 Yongbong-ro, Buk-gu, Gwangju, Republic of Korea. [2]Department of Food and Nutrition, Chonnam National University, 77 Yongbong-ro, Buk-gu, Gwangju, Republic of Korea. [3]Department of Convergence Biosystems Engineering, Chonnam National University, 77 Yongbong-ro, Buk-gu, Gwangju, Republic of Korea. [4]Interdisciplinary Program in IT-Bio Convergence System, Chonnam National University, 77 Yongbong-ro, Buk-gu, Gwangju, Republic of Korea. [5]Research Center for Biological Cybernetics, Chonnam National University, 77 Yongbong-ro, Buk-gu, Gwangju, Republic of Korea. [6]These authors contributed equally: Sol Kim, Ju-Eun Nho. e-mail: makim@jnu.ac.kr; bioksj@jnu.ac.kr

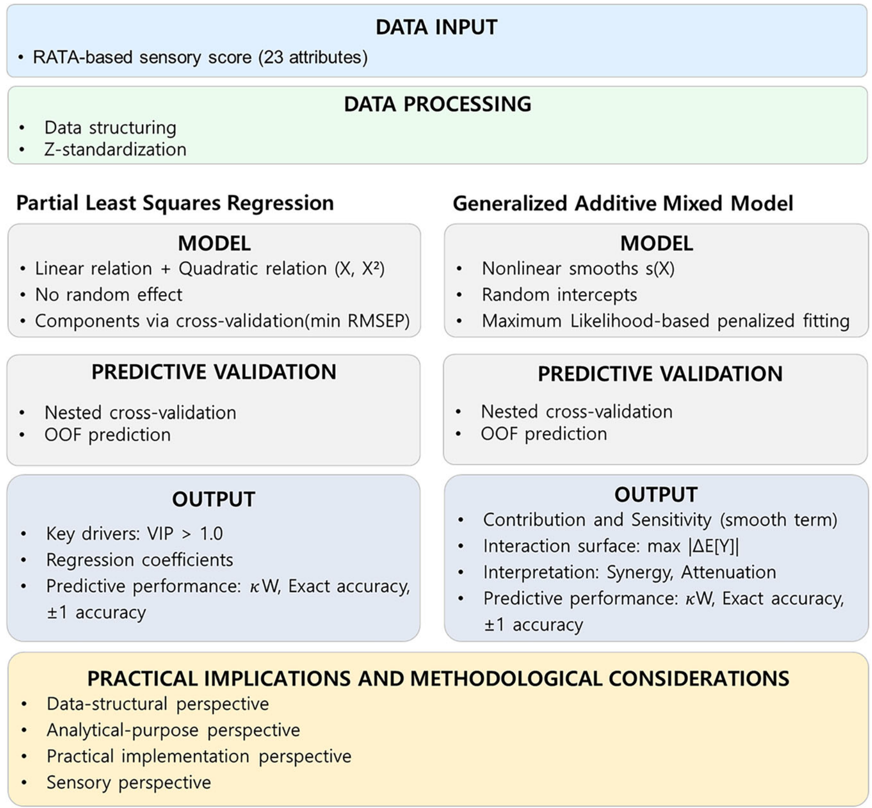

Fig. 1 | Analytical framework for modeling consumer liking from sensory attributes using partial least squares regression (PLSR) and the generalized additive mixed model (GAMM). The workflow summarizes the analysis of sensory attribute data and consumer liking scores using PLSR to evaluate linear multivariate relationships and GAMM to model potentially nonlinear attribute–liking associations.

applications demonstrate the utility of PLSR in identifying critical drivers of liking across diverse product matrices.

Although PLSR has been widely applied to model the relationship between sensory attributes and consumer liking, it assumes linear proportionality between predictors and responses. However, in practice, sensory responses often exhibit nonlinear behavior (e.g., saturation effects or inverted-U relationships), limiting the ability of linear models to capture such complex associations accurately[10]. Moreover, PLSR is extremely sensitive to noise and outliers and cannot adequately account for interconsumer variability or correlations from repeated measurements[11]. When data are dominated by nonlinear effects or high consumer heterogeneity, the interpretability and predictive reliability of PLSR become restricted. Although nonlinear transformations of predictor variables can partially mitigate these problems, such approaches often increase the computational complexity and complicate algorithmic implementation[12].

Beyond these model-related constraints, consumer-liking data are typically collected using a nine-point hedonic scale. From a theoretical perspective, responses on this scale represent ordinal data, as the numerical categories do not guarantee equal psychological spacing between adjacent labels[13]. Despite this approach, consumer researchers generally treat these discrete category ratings as if they were points on a continuum and were interval data[14]. This practice facilitates parametric statistical methods (e.g., analysis of variance and linear regression). However, applying parametric models to data that deviate from interval-scale properties can cause distortions in parameter estimates or reduced estimation efficiency, particularly when respondents disproportionately select extreme categories or the assumed scale continuity is violated[15]. Many sensory studies have analyzed each sensory attribute independently, causing fragmented interpretations and increased computational burden[15]. Together, these methodological and structural limitations indicate the need for a unified and flexible modeling framework that can jointly address the nonlinear and hierarchical nature of sensory–consumer data.

To address these limitations, the generalized additive mixed model (GAMM) is proposed as an alternative framework that accommodates nonlinear relationships while incorporating random effects to account for consumer- and product-level variability. The GAMM integrates smooth functions for fixed effects with random effects, enabling flexible estimation of nonlinear trajectories for individual sensory attributes[16,17]. This approach can capture multidimensional response patterns and interconsumer heterogeneity that conventional linear models may fail to detect[18]. GAMMs have been applied across diverse scientific domains to model nonlinear and hierarchical data structures, including behavioral adaptation in animal science[19], crop productivity in agriculture[20], body mass trajectories in epidemiology[18], and autocorrelated time-series linguistic data in language sciences[21]. However, their application in sensory and consumer research remains limited.

To address this gap, this study proposes a GAMM-based analytical framework for evaluating linear and nonlinear relationships between sensory attributes and consumer liking. This framework is applied to rate-all-that-apply (RATA) sensory data from plant-based meat alternatives (PMAs) (Fig. 1). PMAs have emerged as an alternative in response to growing concerns regarding the environmental and health impacts of conventional meat consumption, and the global market is projected to expand from USD 16.7 billion in 2024 to USD 100.3 billion by 2033[22]. However, consumer acceptance remains limited due to sensory discrepancies between plant-based products and conventional meat, particularly in taste, aroma, texture, and appearance[23,24]. Therefore, systematic sensory analysis is necessary to identify key sensory drivers of liking and to inform product optimization. The objective of this study is to evaluate whether GAMM can serve as a framework for modeling hierarchical and nonlinear sensory–consumer response patterns in product optimization.

Results

Cross-validated predictive performance

The empirical distribution of overall liking scores was first examined to characterize the response structure. As shown in Fig. S1, the ratings span the full nine-point scale but exhibit a non-uniform distribution with multiple peaks, specifically at scores 3 and 6, rather than forming a

Table 1 | Predictive performance of partial least squares regression (PLSR), ordinal (generalized additive mixed model) GAMM, and ordinal logistic regression (OLR) evaluated under random, product-grouped (LOPO), and panelist-grouped (LOPaO) validation schemes

Model		Metrics		
		Weighted Kappa (κW)	Exact accuracy	Accuracy ± 1
PLSR	Random	0.739	0.240	0.632
	LOPO	0.715	0.241	0.601
	LOPaO	0.728	0.218	0.608
Ordinal GAMM	Random	0.829	0.397	0.730
	LOPO	0.759	0.300	0.601
	LOPaO	0.781	0.369	0.667
OLR	Random	0.802	0.394	0.694
	LOPO	0.763	0.309	0.667
	LOPaO	0.751	0.333	0.649

continuous Gaussian pattern. The concentration of responses at discrete score levels, combined with sparse observations in certain categories, reflects the ordinal nature of hedonic judgments. Such a distribution supports the use of an ordinal modeling approach over continuous Gaussian regression, which assumes normality and equal intervals between response categories.

To model the relationship between sensory attributes and hedonic liking, two modeling approaches were applied: (1) a GAMM with an ordinal categorical response and (2) a PLSR model that treats liking as a continuous variable. Although the two models differ in the structure of the response variable (ordinal vs. continuous), their performance was evaluated using common quantitative metrics to enable comparison under a consistent predictive framework.

Model evaluation was conducted using out-of-fold predictions obtained from a nested cross-validation procedure. Specifically, the outer fold was used to assess model generalization performance, while the inner fold was used to select model hyperparameters. The validation framework consisted of three schemes: random tenfold, product-grouped eightfold (LOPO), and panelist-grouped tenfold (LOPaO).

This multi-validation framework was designed to evaluate not only the general predictive performance of the models but also their robustness under different generalization scenarios, such as predicting preferences for new products or new consumers. The model comparison results are summarized in Table 1.

Overall, ordinal GAMM demonstrated higher predictive performance than PLSR across all three validation structures. In the random tenfold cross-validation, ordinal GAMM achieved $\kappa W = 0.829$, exact accuracy = 0.397, and accuracy within ±1 = 0.730, whereas PLSR showed $\kappa W = 0.739$, exact accuracy = 0.240, and accuracy within ±1 = 0.632. Similar patterns were observed in the product-grouped and panelist-grouped validation schemes. Under the product-grouped validation, ordinal GAMM achieved $\kappa W = 0.759$, compared with $\kappa W = 0.715$ for PLSR. Likewise, under the panelist-grouped validation, ordinal GAMM yielded $\kappa W = 0.781$, whereas PLSR showed $\kappa W = 0.728$.

To further evaluate the influence of model structure on predictive performance, an ordinal logistic regression (OLR) model with a linear predictor was introduced as a baseline model. The OLR model was evaluated using the same nested cross-validation framework, including random tenfold, product-grouped eightfold (LOPO), and panelist-grouped tenfold (LOPaO) validation schemes. The results indicated that ordinal GAMM showed higher predictive agreement than OLR under the random and panelist-grouped validation structures. For example, in the random tenfold validation, OLR achieved $\kappa W = 0.802$, whereas ordinal GAMM achieved $\kappa W = 0.829$. In contrast, under the product-grouped eightfold (LOPO) validation, OLR slightly outperformed ordinal GAMM in terms of predictive agreement (Table 1).

Considering that RATA intensity scores are measured on a 1–5 ordinal scale, an additional analysis was conducted to examine the influence of predictor representation. Specifically, a model in which RATA intensity variables were treated as ordered factors rather than continuous predictors was additionally fitted. However, under this specification, model estimation became unstable across cross-validation folds and did not yield reliable predictive performance. Therefore, in the main analysis, RATA intensity variables were retained as Z-standardized continuous predictors.

Evaluation of model adequacy and interpretability

In Figure S2, diagnostic checks were performed using residual plots, including histograms, residual quantile–quantile (Q–Q) plots, residual–linear predictor plots, and observed-fitted plots.

The residual Q–Q plot displays modest deviations from the theoretical normal quantile line, including in the central quantile region. The residual histogram similarly demonstrated no evident skewness or multimodality, supporting the adequacy of the residual structure. In the residual–linear predictor plot, the residuals were dispersed randomly around zero; however, a mild S-shaped curvature emerged in the lower range of the linear predictor, and residual clustering attributable to the stepwise cut-point structure was observed in the midrange. Heteroscedasticity refers to a systematic increase or decrease in the variance of residuals as a function of the predictor and is typically determined in residual–fitted plots via a cone-shaped widening or narrowing of the residual spread.

This study evaluates heteroscedasticity using the residual–linear predictor plot. The analysis revealed no systematic changes in residual variance across the full range of the linear predictor. Even in regions associated with large predictor values, the residuals did not display directional spreading or increased dispersion; instead, they remained stably centered around zero.

In the observed-fitted plot, the response variable has a nine-point ordinal structure; therefore, the plot displays horizontal step-like bands rather than the continuous 45° diagonal pattern expected for continuous outcomes. In this analysis, higher liking scores are associated with larger fitted latent values, indicating that the model appropriately captures the underlying latent preference structure.

Concurvity diagnostics revealed low pairwise concurvity values among smooth terms (range: 0.00–0.38, median: 0.03; Fig. S3). To determine whether each sensory attribute exhibits a linear or nonlinear relationship to consumer liking, the edf of the corresponding smooth terms was examined. In Fig. 2, edf values range from 1.00 to 2.38, indicating a modest variation in the degree of nonlinearity across attributes. Most attributes exhibited nearly linear or only weakly nonlinear effects ($edf \approx 1.0$–1.5), whereas a limited subset of attributes showed more pronounced curvature. In particular, Cooked_soybean_F ($edf = 2.38$), Crumbly_M ($edf = 2.34$), and Hardness_M ($edf = 2.21$), Oily_M ($edf = 2.18$), Glossy_A ($edf = 2.14$), and Smoked_F ($edf = 2.07$) displayed moderate nonlinear tendencies. All k-index values exceeded 0.97 ($p > 0.05$), confirming that the basis dimensions were adequate to capture the functional form of each smooth term.

To elucidate these nonlinear tendencies, partial-effect functions were visualized for the attributes with the highest edf values (Fig. 3). Distinct nonlinear patterns were observed across several attributes. For example, Cooked_soybean_F, Crumbly_M, and Hardness_M presented inverted-U-shaped patterns, in which liking increased at moderate intensities but declined beyond certain thresholds. Oily_M displayed a monotonic decrease across the range, indicating that excessive oiliness may adversely influence liking. Glossy_A showed a shallow nonlinear pattern with a gradual upward turn at higher intensities, suggesting a weaker but non-negligible deviation from linearity. Conversely, Smoked_F presented a monotonic, linear increase in liking, indicating a consistently positive contribution of smoky flavor to the overall hedonic responses. Notably, the mouthfeel-related attributes (e.g., Hardness_M and Crumbly_M) predominantly presented inverted-U-shaped patterns.

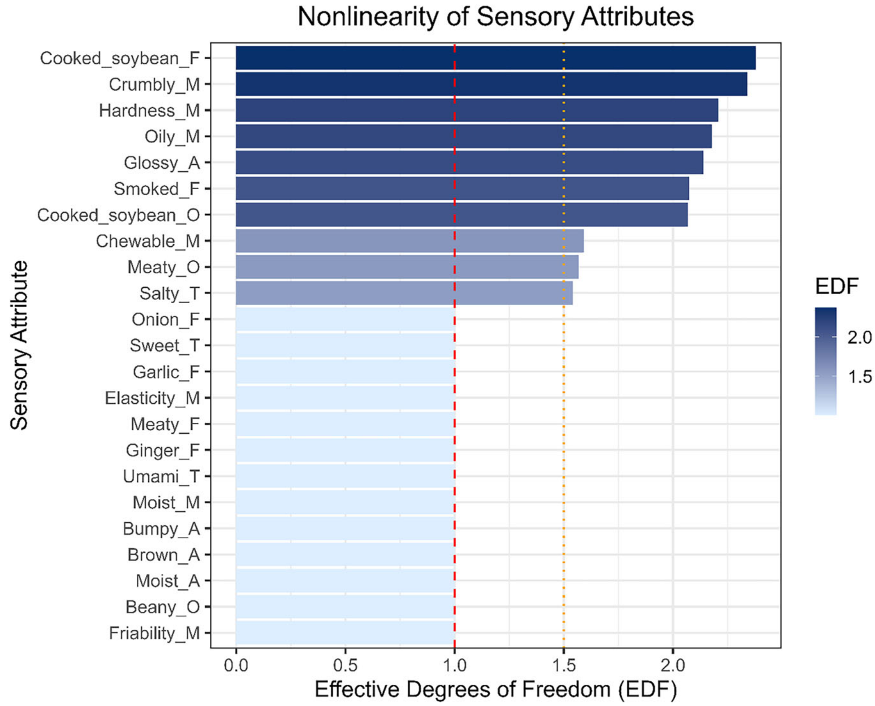

Fig. 2 | Effective degrees of freedom (edf) of sensory attributes in the ordinal generalized additive mixed model (GAMM). Bar color intensity indicates the degree of nonlinearity (darker blue = stronger curvature). The dashed red line (*edf* = 1) denotes the threshold for linear effects.

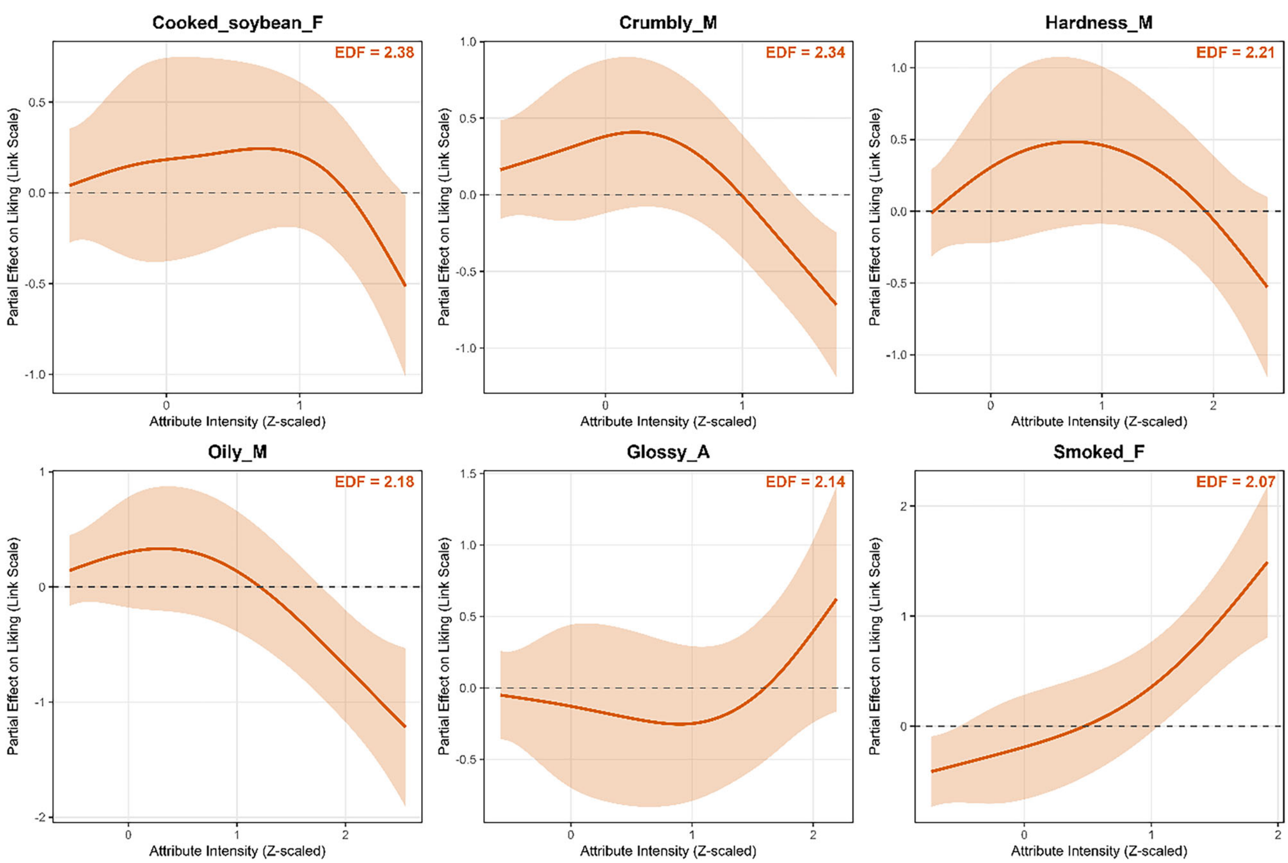

Fig. 3 | Partial-effect functions of sensory attributes exhibiting nonlinear relationships (edf >1). Shaded areas denote 95% confidence intervals, indicating estimation uncertainty and the stability of the curve.

Quantitative identification of sensory drivers of consumer liking

The crucial sensory drivers of consumer liking were examined using two analytical approaches, PLSR and GAMM. In the PLSR framework, the importance of predictor variables was evaluated using the VIP index. The first-order PLSR model ($n_{comp} = 4$) identified Smoked_F, Meaty_F, Crumbly_M, Oily_M, Glossy_A, Meaty_O, Umami_T, Moist_A, Moist_M, and Beany_O as important predictors (VIP > 1), indicating that these attributes substantially contributed to the variation in liking scores (Fig. 4a).

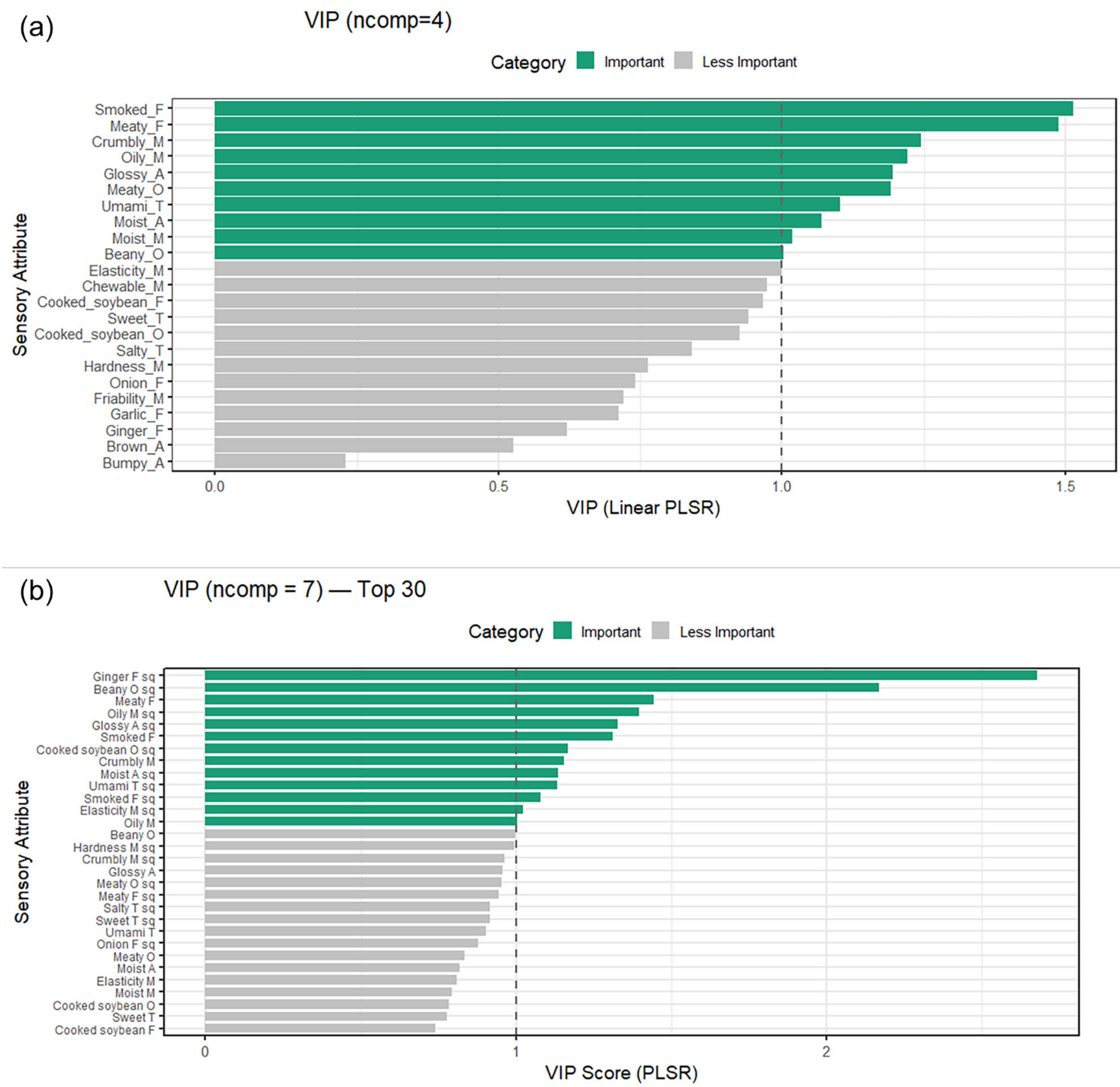

Fig. 4 | Variable importance in projection (VIP) scores of sensory attributes derived from partial least squares regression (PLSR) models. a Linear PLSR model with four components ($n_{comp} = 4$). **b** Quadratic-extended PLSR model including squared terms ($n_{comp} = 7$). Attributes with VIP >1.0 (green) were identified as critical drivers of consumer liking. The suffix "_sq" denotes the squared (quadratic) term of the corresponding sensory attribute.

Among these high-VIP predictors, Crumbly_M, Moist_A, and Beany_O exhibited negative regression coefficients, indicating a detrimental effect on liking, whereas the remaining attributes had positive effects. To account for potential nonlinear effects, the extended PLSR model incorporating squared terms was examined ($n_{comp} = 7$). This model partially confirmed the findings from the first-order model while identifying Ginger_F_sq, Beany_O_sq, and Oily_M_sq as influential predictors (Fig. 4b). The prominence of these squared terms indicates that several sensory attributes display nonlinear relationships (e.g., inverted-U or plateauing) with liking, suggesting that an additive linear structure alone cannot completely capture their effects. This extended PLSR framework offers evidence of the curvilinear contributions to consumer liking.

In parallel, the GAMM framework quantifies the partial contribution of each sensory attribute to the predicted liking score using two complementary metrics: the SD of each partial-effect curve, reflecting the overall variability in contribution, and the peak-to-peak range, representing the sensitivity of liking response to changes in attribute intensity (Fig. 5a, b). Both metrics consistently identified Smoked_F, Meaty_F, Oily_M, Beany_O, Meaty_O, and Crumbly_M as the most influential drivers of liking. These findings reveal strong agreement with the critical attributes identified using the PLSR models, demonstrating agreement between the two analytical frameworks in determining the sensory factors most strongly associated with consumer liking.

Bivariate interaction surfaces of sensory attributes

Figure S4 summarizes the top 30 bivariate sensory–attribute interactions ranked by the maximum absolute nonlinear deviation on the observed hedonic scale (max|ΔE[Y]|) among the 253 tested attribute combinations. The nonadditive interaction effect was defined as the difference between the predicted liking values from the GAMM including the interaction term and those from the corresponding additive model containing only the main effects of each attribute ($\Delta E[Y] = \hat{Y}_{int} - \hat{Y}_{add}$).

Interaction candidates were screened using a penalized interaction framework (select = TRUE). Among the 253 tested attribute combinations,

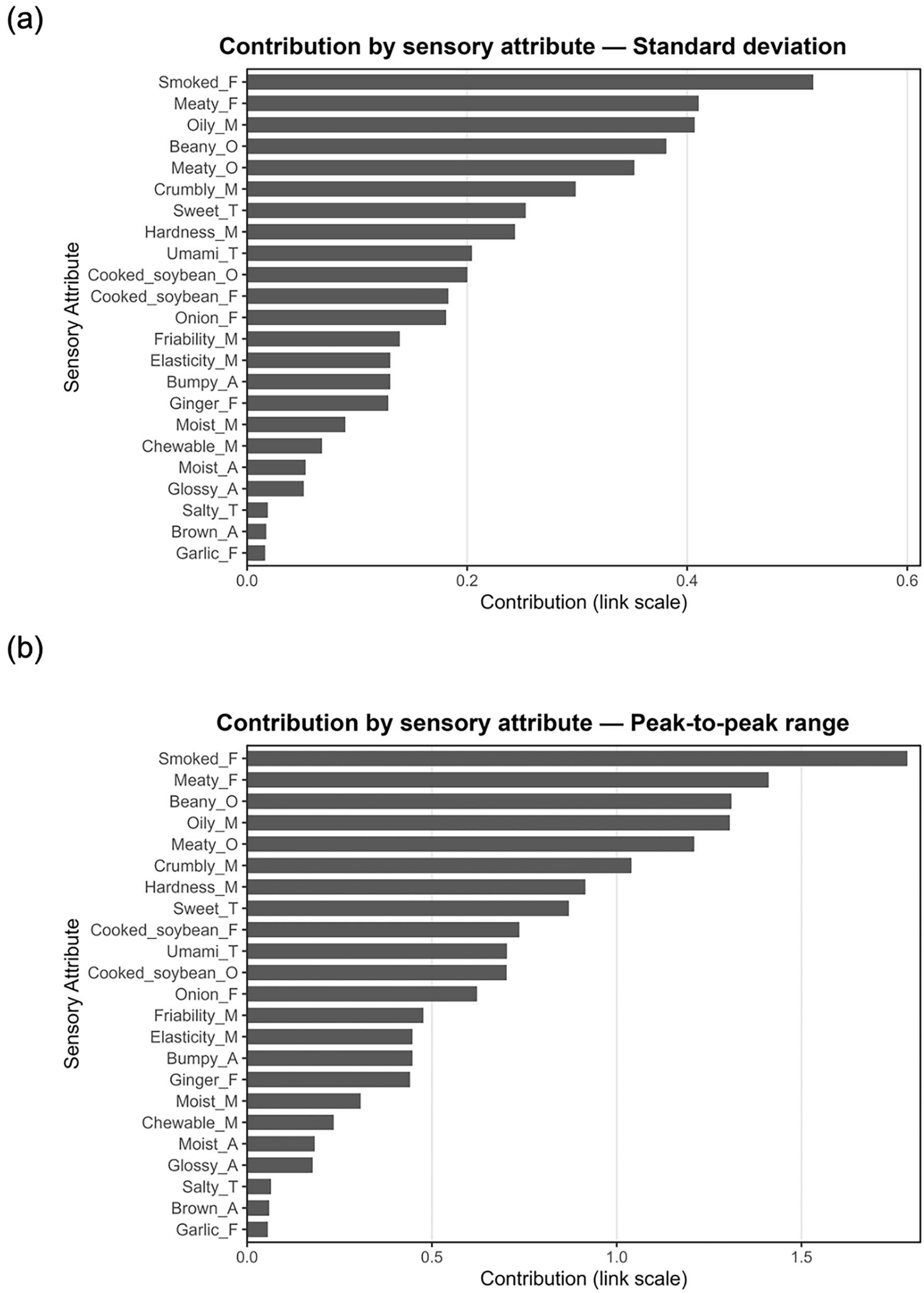

Fig. 5 | Relative contribution of each sensory attribute to predicted consumer liking derived from the ordinal generalized additive mixed model (GAMM). **a** Standard deviation of the partial-effect curves, reflecting the overall variance explained. **b** Peak-to-peak range of the partial-effect curves, representing the sensitivity of liking to changes in each attribute. Attributes are ranked in descending order of contribution within each criterion.

46 pairs satisfied both the Benjamini–Hochberg adjusted p value criterion ($p < 0.05$) and the $\Delta AIC \geq 2$ threshold, and were therefore considered statistically supported interactions. To ensure that the evaluation reflected the data-supported sensory region, effect sizes were calculated within the 5–95% quantile range of each attribute. Observed-scale effect metrics included the maximum absolute deviation (max $|\Delta E[Y]|$), the mean absolute deviation (mean $|\Delta E[Y]|$), and the proportion of the sensory space exceeding the 0.5-point just-noticeable difference (JND) threshold.

Across the screened interactions, the magnitude of nonlinear deviations ranged from ~0.2 to 2.11 units on the observed hedonic scale, with the largest deviation observed for the Meaty_F × Cooked_soybean_O pair (max $|\Delta E[Y]| = 2.11$). Several of the higher-ranked interactions exceeded the 0.5-point JND threshold within localized regions of the sensory space, indicating that departures from additive predictions may reach perceptible levels under specific combinations of sensory intensities.

Figure 6 presents the two-dimensional interaction surfaces for the six highest-ranked pairs selected based on max $|\Delta E[Y]|$. All surfaces were fitted using the penalized GAMM interaction term (select = TRUE) and evaluated within the 5–95% quantile region of the sensory space. The complexity of each interaction smooth was assessed using the edf, which ranged approximately from 2.8 to 6.5, indicating moderate nonlinear structure rather than highly irregular surfaces.

Several characteristic patterns emerged from these interaction surfaces. First, relatively strong nonadditive deviations were observed in combinations involving meaty and cooked-soybean-related flavor attributes. In particular, the Meaty_F × Cooked_soybean_O interaction exhibited a comparatively large effect size ($edf = 5.89$, $\Delta AIC = 6.42$, BH-adjusted $p \approx 0$), with deviations from the additive prediction increasing in regions where both attributes intensified simultaneously. Similarly, the Glossy_A × Meaty_F interaction showed substantial model improvement ($edf = 5.92$, $\Delta AIC = 10.78$, BH-adjusted $p \approx 0$), suggesting a nonlinear relationship between visual appearance and perceived meaty flavor.

Second, attributes associated with cooked-soybean flavor appeared repeatedly among the higher-ranked interaction pairs. For example, the Glossy_A × Cooked_soybean_O combination exhibited detectable nonlinear deviations ($edf = 6.47$, $\Delta AIC = 3.15$, BH-adjusted $p \approx 0$), indicating that visual glossiness and cooked-soybean aroma jointly influence hedonic responses within specific regions of the sensory space.

Third, interactions between taste and texture attributes were also observed. The Umami_T × Moist_M pair showed nonlinear variation in predicted liking, suggesting that the influence of umami intensity on liking depends on the perceived level of moisture. In contrast, the Sweet_T × Beany_O interaction exhibited a relatively simpler structure. In addition, the Beany_O × Meaty_O pair indicated nonlinear relationships among aroma-related attributes. Overall, the estimated interaction surfaces did not exhibit globally oscillating or highly wavy patterns across the sensory space. Instead, nonlinear deviations were localized within restricted regions of attribute intensity, forming smooth gradient-like structures with localized curvature. This pattern suggests context-dependent interaction effects rather than widespread nonlinear fluctuations driven by model overfitting.

To further assess the uncertainty of the estimated practical effect sizes, a panelist-cluster bootstrap analysis was conducted for a subset of high-ranked interaction pairs identified during the screening stage. For each pair, the maximum absolute deviation on the observed scale ($\max|\Delta E[Y]|$) was recalculated across bootstrap resamples to obtain 95% confidence intervals. The bootstrap results indicated that effect magnitude varied across interaction pairs, with several high-ranked interactions maintaining substantial deviations on the observed hedonic scale. For example, the Meaty_F × Cooked_soybean_O interaction showed a point estimate of max $|\Delta E[Y]| = 2.11$ with a bootstrap 95% confidence interval of ~0.91–2.77, indicating a relatively stable deviation from additive predictions. Similarly, the Glossy_A × Meaty_F interaction exhibited a max $|\Delta E[Y]|$ of 1.11 with a bootstrap 95% confidence interval of ~0.13–2.18, suggesting a moderate nonlinear interaction effect, although the relatively wide confidence interval indicates greater uncertainty in the effect magnitude compared with the top-ranked interaction pair. Other interactions exhibited smaller or more uncertain effect magnitudes, suggesting that the practical relevance of interaction effects may vary depending on the specific attribute combination.

Discussion

The primary aim of this study was not to determine the superiority of any analytical model but to explore how different statistical approaches can function complementarily in sensory–consumer research. This work aims to illustrate how each approach offers distinct insight, depending on the structural data characteristics and intended analytical objectives. To this end, two modeling frameworks were compared: the GAMM and PLSR. Although these models differ considerably in their underlying assumptions, flexibility, and interpretability, this study presents a quantitative and qualitative discussion of the unique advantages and limitations of each framework in practical data analysis and application contexts.

The comparison of modeling approaches showed that ordinal GAMM generally achieved higher predictive performance than PLSR across all validation structures. This finding underscores the importance of accounting for the ordinal nature of hedonic liking when modeling consumer preference data. In sensory and consumer research, hedonic ratings are often treated as continuous variables; however, they are inherently ordinal measurements representing ordered preference categories rather than interval-scale values. By modeling the response variable within an ordinal framework, GAMM preserves the ordered structure of the hedonic scale and avoids assuming equal distances between adjacent categories. Explicitly incorporating this ordinal structure may therefore provide more reliable predictions of consumer liking than models that treat the response as continuous.

Further insight into the role of model structure was obtained from the comparison between ordinal GAMM and the baseline OLR model. Although both models retain the ordinal nature of the response variable, OLR assumes a strictly linear relationship between predictors and the latent preference scale, whereas GAMM extends the ordinal framework by incorporating additive smooth functions that allow more flexible modeling of sensory–hedonic relationships. Diagnostic results indicated that most sensory attributes exhibited approximately linear relationships with liking ($edf \approx 1$), while a smaller subset showed mild nonlinear tendencies (maximum $edf \approx 2.38$), suggesting the presence of limited curvature or threshold-like effects for certain attributes. These structural differences were also reflected in the cross-validation results. Under random validation and panelist-grouped validation, ordinal GAMM showed higher predictive agreement than OLR, suggesting that nonlinear flexibility can improve prediction when the training data sufficiently cover the sensory response space. In contrast, under product-grouped validation (LOPO), OLR slightly outperformed ordinal GAMM, possibly reflecting the stricter generalization scenario of predicting liking for entirely unseen products, where simpler linear models may generalize more robustly than more complex nonlinear models.

When selecting and validating statistical models, goodness-of-fit must be assessed and the underlying model assumptions must be satisfied. A properly fitted model should adequately explain the observed data while minimizing the risk of erroneous inference. In this context, residual analysis plays a central role in identifying potential deficiencies in model fitting, as well as detecting outliers or influential observations, and is therefore an indispensable component of regression diagnostics[25].

Model diagnostics indicate that the fitted ordinal GAMM adequately represents the sensory–consumer data (Fig. S2). In cumulative ordinal models, Pearson and deviance residuals do not follow a normal reference distribution even when the model is correctly specified[26]. Consequently, Q–Q plots in this setting are primarily used to detect severe model misspecification rather than to assess residual normality. The Q–Q plot observed in this study exhibits residual patterns consistent with those widely reported in ordinal regression diagnostics[27].

Specifically, the subtle departures in the central quantile region and the gentle S-shaped form correspond to the expected distributional characteristics of cumulative ordinal models rather than to anomalous behavior. Importantly, the absence of pathological residual patterns—such as extreme curvature, multimodal distributions, or heavy-tailed deviations—indicates that no substantial lack of fit occurred in the ordinal GAMM. Such residual structures are characteristic of cumulative link–based ordinal models and are not typically indicative of model misspecification[27].

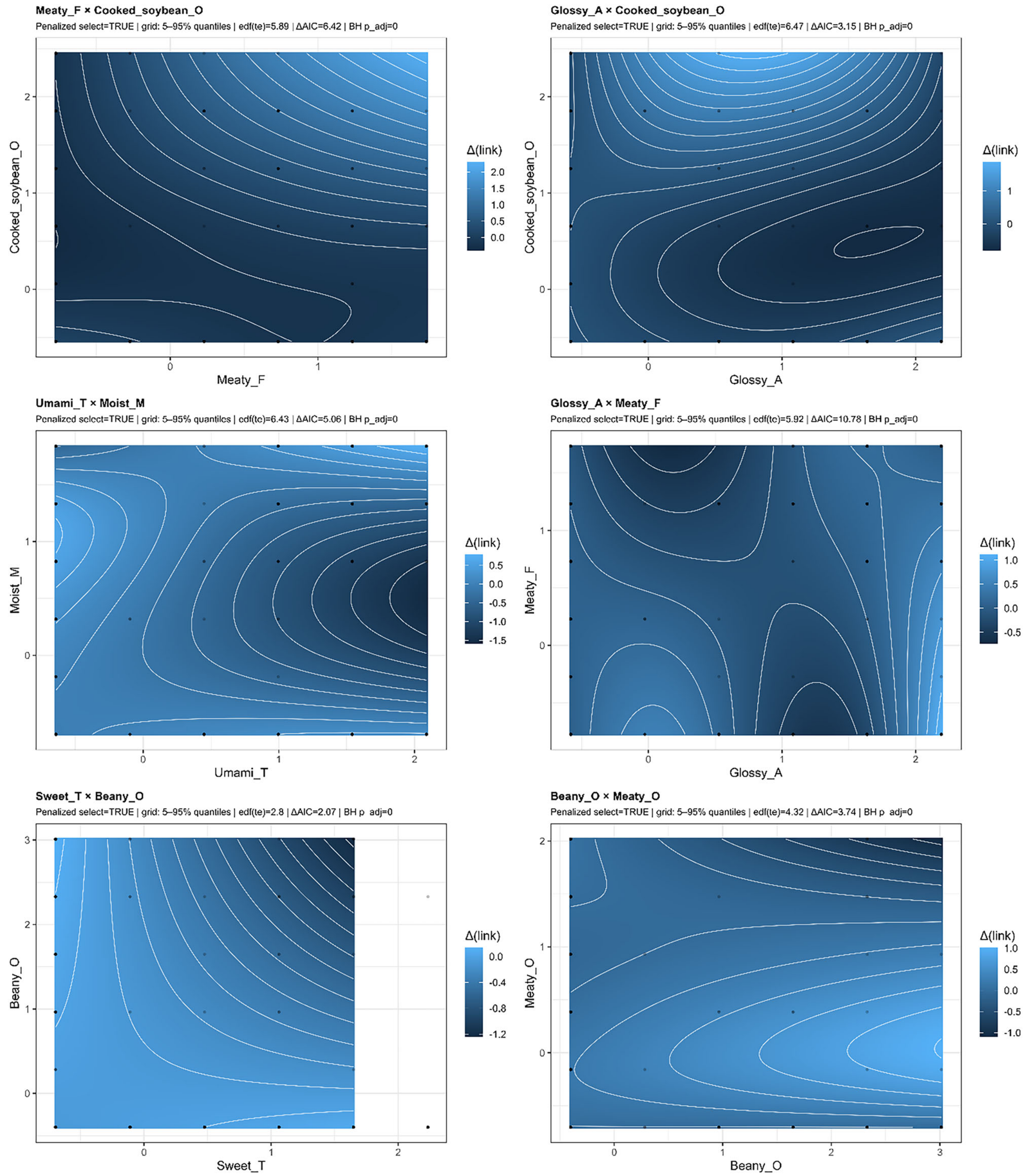

Fig. 6 | Two-dimensional GAMM interaction surfaces illustrating nonadditive sensory effects for the six highest-ranked attribute pairs. The panels show interaction surfaces for Meaty_F × Cooked_soybean_O, Glossy_A × Cooked_soybean_O, Umami_T × Moist_M, Glossy_A × Meaty_F, Sweet_T × Beany_O, and Beany_O × Meaty_O, arranged from top left to bottom right. The x- and y-axes represent the two sensory attributes in each pair, and the color scale indicates the estimated nonadditive effect on the model link scale, expressed as Δ(link). Points indicate the observed sample locations used for model estimation.

Heteroscedastic variance, when present, can distort standard error estimation and lead to inaccurate test statistics and confidence intervals[28,29]. In cumulative link models, failure to account for heteroscedasticity may further bias the estimation of location parameters[30]. However, in the present analysis, no systematic changes in residual variance were observed across the range of the linear predictor, providing no evidence of heteroscedasticity in the fitted ordinal GAMM.

In addition, the observed-fitted patterns arise naturally from the discrete nature of ordinal responses. Accordingly, the primary diagnostic criterion in this context is whether the fitted values exhibit a directionally consistent relationship with the ordered response categories, rather than conformity to a continuous reference pattern.

Concurvity analysis further demonstrated that the smooth terms were sufficiently independent. The highest observed concurvity value (0.38, Glossy_A × Moist_A) reflects a plausible perceptual association between attributes rather than statistical redundancy (Fig. S3). Overall, these results confirm that multicollinearity among smooth terms was limited, allowing stable estimation of individual smooth effects.

To facilitate interpretation of the nonlinear relationships, partial-effect functions were examined. A partial-effect function represents a univariate response curve describing how the predicted liking score changes as the intensity of a single sensory attribute varies, while holding all other predictors constant. These visualizations serve as essential interpretive tools within the GAMM framework, as they reveal both the direction of change in liking (increasing or decreasing) and the functional shape of the response. Importantly, response features such as optimal midrange intensities, plateau effects, or adverse responses at excessively high levels cannot be adequately captured by linear regression approaches (Fig. 3).

From a sensory perspective, prior literature has consistently identified texture as a critical determinant of quality in PMAs[31], and the present findings reinforce this view. In particular, the inverted-U-shaped patterns observed for certain attributes, such as Hardness_M, in the GAMM analysis are consistent with findings from prior research on food formulation[32,33]. These nonlinear response functions indicate that consumer acceptance of texture does not follow a simple monotonic pattern but instead depends on achieving attribute-specific optimal intensity ranges. Deviations toward insufficient or excessive intensity frequently result in reduced liking, consistent with well-documented sensory challenges associated with PMAs. Consequently, the inverted-U-shaped relationships identified for several mouthfeel-related attributes within the GAMM framework directly reflect this sensitivity to textural imbalances and provide indirect support for the plausibility of the nonlinear patterns identified within the GAMM framework.

Collectively, these findings demonstrate that the proposed GAMM framework is sufficiently flexible to detect nonlinear effects when they are present, while also confirming that linear relationships remain predominant for many sensory attributes.

From a data-structural perspective, the PLSR framework assumes a continuous response structure[34], treating hedonic ratings as quantitative variables with equal intervals between adjacent categories[35]. Such an interval-scale assumption is commonly adopted in sensory and consumer research when applying linear models, such as PLSR, PCA, or LMM[36–38].

In practice, however, consumer liking data are ordinal rating data composed of discrete ordered categories. Applying a normal linear model to such responses, therefore, represents an approximation in which categorical ratings are treated as continuous variables. This simplification can be acceptable when the number of response categories is sufficiently large, and the response distribution is approximately symmetric[35].

However, the empirical distribution of liking scores in the present dataset suggests that this approximation may be limited. As shown in Fig. S1, the ratings span the full nine-point scale but are unevenly distributed across discrete categories rather than forming a continuous Gaussian pattern. Such a distribution reflects the categorical nature of hedonic judgments and highlights the limitations of treating liking scores as continuous variables. Under these conditions, modeling the response within an ordinal framework provides a more appropriate representation of consumer preference data.

In contrast, the GAMM models ordinal response structures while allowing nonequidistant intervals. By incorporating smoothing functions to capture nonlinear relationships and random effects to account for consumer heterogeneity, GAMM interprets hedonic data as ordinal judgments rather than continuous scores[39,40].

Accordingly, the comparison between PLSR and GAMM should not be viewed solely as a competition in predictive accuracy but rather as an assessment of data–model congruence (i.e., the extent to which the statistical assumptions of each model align with the structural properties of the data). As summarized in Table 1, although both regression frameworks achieved reasonable predictive accuracy, GAMM showed greater structural alignment with ordinal, nonequidistant nine-point responses, capturing nonlinear and categorical consumer behaviors more effectively.

Although PLSR remains advantageous due to its simplicity, computational efficiency, and straightforward implementation (especially in routine or industrial settings requiring rapid calibration), its linear formulation limits its ability to capture the nonlinear, nonequidistant nature of hedonic ratings[41]. In contrast, GAMM flexibly models complex response surfaces and cross-modal interactions, offering deeper insight into the perceptual mechanisms underlying consumer liking. Thus, the distinction between the two frameworks is not between prediction and explanation, but between analytical depth and structural fidelity, with respect to how well the model assumptions align with the true properties of the sensory–consumer data.

From an implementation standpoint, PLSR also benefits from its broad availability in standard statistical software. The method can be rapidly fitted, interpreted, and reproduced, making it useful for iterative product development or industrial settings where fast, explainable modeling is necessary. In practice, PLSR is particularly suitable for rapid screening in the initial stages of new product development or when the number of sensory attributes is small and linear relationships are reasonably expected. In addition, its VIP-based variable importance measures provide intuitive, easily communicable summaries, facilitating practical decision-making and communication with nonspecialist stakeholders.

In contrast, GAMM involves more complex parameter tuning and greater computational demands, especially when modeling nonlinear interactions and random structures. However, this complexity enables GAMM to accommodate ordinal responses, nonlinear patterns, and hierarchical data structures flexibly, delivering a more faithful representation of the characteristics of sensory–consumer datasets. Accordingly, GAMM is more appropriate when hedonic responses are measured on ordinal scales with a limited number of categories, when nonlinear patterns (e.g., saturation, thresholds, or inverted-U functions) are anticipated, or when cross-modal interactions between sensory attributes must be explored. Moreover, GAMM explicitly accounts for consumer heterogeneity via random effects, offering more detailed insight into the perceptual mechanisms underlying consumer responses. Thus, the choice between these two analytical frameworks reflects a practical balance between accessibility and adaptability: PLSR offers ease of implementation, whereas GAMM delivers structural sophistication and interpretive precision.

From a sensory perspective, individual sensory modalities do not operate as entirely independent modules, as cross-modal interactions may occur whereby one sensory modality can influence the response of another[42]. Accordingly, food perception reflects the integrative processing of multiple sensory attributes, which are more likely to influence overall acceptability through interactive mechanisms rather than operating in a strictly additive manner. In this context, a principal strength of GAMM lies in its capacity to explicitly capture nonlinear and interaction effects among sensory attributes. Whereas PLSR summarizes overall trends under the assumption that each attribute contributes linearly and independently, GAMM constructs continuous response surfaces through smoothing functions, thereby enabling more flexible modeling of how combinations of attributes influence consumer liking.

In the present study, this characteristic was reflected in several cross-modal interactions identified in the interaction analysis. For example, combinations involving meaty and cooked-soybean related attributes (e.g., Meaty_F × Cooked_soybean_O*)* showed relatively strong nonlinear deviations from additive predictions. In addition, interactions involving appearance and flavor attributes (e.g., Glossy_A × Meaty_F) and taste–texture relationships (e.g., Umami_T × Moist_M) illustrated that the influence of individual sensory attributes on liking may depend on the intensity of other attributes. These results suggest that consumer liking cannot always be represented as a simple linear aggregation of individual attributes but may emerge within a multivariate context of interacting sensory dimensions. Notably, the attributes involved in these interactions, such as meaty flavor and cooked-soybean odor, have also been identified as key drivers of liking in previous studies on plant-based meat alternatives[43–45]. These findings provide insight into how individually important sensory attributes may jointly influence consumer liking within a broader sensory configuration.

Nevertheless, model selection should not be determined solely by the interpretive depth or analytical complexity of a given framework. In practical analytical settings, the structural characteristics of the dataset and the configuration of the sample should also be considered. When the number of product samples is limited, linear dimension-reduction approaches such as PLSR may provide more stable predictive performance by summarizing the dominant variance structure of the data. In contrast, when the response variable is measured on an ordinal scale, when nonlinear relationships between predictors and liking responses are present (e.g., edf >1), or when consumer-level hierarchical variability must be modeled, GAMM may provide a more appropriate analytical framework that better reflects the underlying data structure. From this perspective, PLSR may serve as a practical tool for rapid exploratory screening of key sensory drivers, whereas GAMM may be more suitable for the detailed interpretation of nonlinear effects, hierarchical structure, and interactions among sensory attributes.

Several limitations of the present study should be acknowledged. First, the number of product samples included in the analysis was relatively small (N = 8). Because nonlinear mixed models estimate both smoothing parameters and variance components simultaneously, the stability of these estimates may be influenced by the composition of the product set when the number of products is limited. Although the nested cross-validation framework suggested reasonable predictive stability, future studies incorporating a larger and more diverse product space would be beneficial to further evaluate the robustness and reproducibility of the proposed modeling approach.

Second, the present analysis was conducted within a single food category. Sensory attribute structures and consumer preference mechanisms may vary across food systems; therefore, the extent to which the observed modeling behavior generalizes to other product categories remains uncertain. Additional studies involving different sensory contexts would help confirm the broader applicability of the findings.

Third, the measurement characteristics of the RATA methodology warrant consideration. Previous studies comparing RATA with conventional intensity scaling methods have generally reported broadly comparable findings[46,47], providing empirical support for the validity of RATA-based sensory measurements. Nevertheless, unlike conventional intensity scales, RATA employs a binary rating procedure in which respondents first indicate the presence of an attribute before rating its intensity. Because intensity ratings are conditional on prior attribute selection, this two-step structure may introduce measurement variability. Specifically, because intensity ratings are only recorded for selected attributes, lower-intensity regions of the sensory space may be underrepresented, potentially destabilizing smooth estimates in those regions and shifting the apparent location of nonlinear features such as inverted-U peaks toward the center of the observed data range.

In the present study, RATA intensity scores were treated as observed predictors in the GAMM framework, and—as in most regression analyses based on self-reported sensory data—some degree of measurement error may therefore be present. Such an error may attenuate estimated effect magnitudes or influence the apparent location and strength of nonlinear relationships captured by spline-based smooth terms.

Future research may address this limitation by integrating nonlinear mixed modeling approaches with latent variable frameworks that explicitly account for measurement error. For example, combining GAMM with structural equation modeling (SEM) could allow the separation of latent sensory constructs from observed RATA responses while retaining the capacity to model nonlinear and hierarchical relationships among sensory attributes and consumer liking.

Finally, although the modeling framework provided useful predictive insights, the present study did not include independent experimental validation of the key sensory relationships identified by the models. Because the nonlinear patterns and interaction effects were inferred from observational sensory data, their causal interpretation remains limited. Future research should therefore aim to experimentally verify the identified attribute relationships through controlled product manipulations. Such validation would help assess the reliability and reproducibility of the proposed modeling approach and support its practical application in product quality optimization.

Methods

Sensory evaluation

For the case study, this work uses a dataset from a previous study[48], in which 200 consumers evaluated seven PMAs compared with a pork-based *tteok-galbi* (short-rib patties). Participants were recruited in South Korea through online postings targeting a university community and nearby residents. Individuals reporting allergies to relevant ingredients (e.g., soy, wheat, pork, or eggs) or oral conditions potentially compromising sensory evaluation were excluded from participation. Participants were randomly assigned to one of two sensory assessment methods: check-all-that-apply (CATA; $N = 100$) or rate-all-that-apply (RATA; $N = 100$). Demographic characteristics were as follows: CATA group (male = 25, female = 75; mean age = 24.20 ± 8.27 years); RATA group (male = 27, female = 73; mean age = 23.90 ± 7.99 years).

Sample evaluation followed a monadic sequential design, in which each sample was assessed independently. For each sample, participants first rated overall liking using a nine-point hedonic scale (1 = extremely dislike to 9 = extremely like), followed by sensory assessment using their assigned method. Both methods employed the same 23 sensory attributes covering five dimensions: appearance (four attributes), odor (three attributes), taste (three attributes), flavor (six attributes), and mouthfeel (seven attributes). These descriptors were selected based on previous research investigating sensory characteristics of meat and plant-based patties[49–52]. Representative examples include "glossy appearance" and "moist appearance." In the CATA procedure, participants indicated all perceived attributes. In the RATA procedure, participants indicated the presence of each attribute and, if perceived, rated its intensity on a five-point scale (1 = very weak to 5 = very strong). As the panel consisted of untrained consumers, no formal training on descriptor definitions was provided; however, procedural instructions were delivered to ensure comprehension of the rating task. To minimize primacy bias, the presentation order of sensory terms was randomized across participants[53].

For the present analysis, only the RATA dataset was employed because intensity ratings were deemed more informative for predictive modeling. The intensity scores for the sensory attributes were used as predictor variables, whereas overall liking was designated as the dependent variable. Further details of the experimental design have been reported previously[48].

Liking modeling using PLSR

In the PLSR analysis, the independent variables comprised 23 sensory intensity attributes evaluated via the RATA procedure, encompassing five sensory dimensions: appearance, odor, taste, flavor, and mouthfeel. Each observation corresponded to an individual consumer–sample combination (100 consumers × 8 samples = 800 observations). All sensory attributes were z-standardized (mean = 0, SD = 1) to eliminate scale differences between variables.

Although RATA intensity scores are ordinal by definition, they were treated as standardized continuous predictors because the scale approximates interval-level measurement and allows stable estimation of regression-based sensory models. To evaluate the potential impact of predictor treatment, a sensitivity analysis was additionally conducted by modeling RATA attributes as ordered factors.

The dependent variable was the individual overall liking score for each observation, measured on a nine-point hedonic scale (1 = extremely dislike to 9 = extremely like). Squared terms of all predictors were included in the model to capture potential nonlinear relationships between sensory attributes and consumer liking[41]. This approach enabled the evaluation of linear and curvature effects in a primarily linear modeling framework. The PLSR method was conducted using the *pls* package in R, and the optimal number of latent components was determined within each outer training set using five-fold inner cross-validation by selecting the number of components that minimized the mean prediction error (RMSE).

The relative importance of each sensory attribute was quantified using the variable importance in projection (VIP) score, where attributes exhibiting VIP > 1.0 are considered the most influential drivers of liking[54]. Additionally, regression coefficients were assessed to interpret the direction of influence (positive or negative) for each attribute. When quadratic terms were included, curvature visualizations based on the fitted coefficients were applied to illustrate the potential nonlinear response patterns in the sensory–liking associations.

Liking modeling using GAMM

Model specification and fitting. An ordinal GAMM framework was implemented using the *mgcv* package in R (v 4.3.1) to investigate the linear and nonlinear associations between sensory attributes and consumer overall liking. The dataset structure, predictor variables, and preprocessing procedures are identical to those in the PLSR analysis.

The GAMM framework extends the generalized linear model by incorporating smooth functions that flexibly model nonlinear response patterns within a cumulative ordinal modeling framework, while accounting for hierarchical random variability at the consumer and product levels[55,56]. The generalized model formulation is expressed in Eq. (1):

$$g\left(\Pr\left(Y_i \le c|X_i\right)\right) = \theta_c + \left[\sum_{j=1}^{m} s_j\left(X_{ji}\right) + b_{consumer(i)} + b_{product(i)}\right],\ c = 1, \ldots, C-1, \tag{1}$$

where $\Pr\left(Y_i \le c|X_i\right)$ denotes the cumulative probability of the liking score Y_i, which is less than or equal to category c (for $C = 9$ ordinal categories).

In the mgcv::ocat family, the ordinal response is modeled using a latent variable formulation in which category probabilities are determined by threshold parameters ($\boldsymbol{\theta_c}$) separating adjacent ordinal levels. The smooth function $s_j(\cdot)$ represents the nonlinear effect of the j-th sensory attribute, while $b_{\mathrm{consumer}(i)}$ and $b_{\mathrm{product}(i)}$ represent random intercepts capturing interconsumer and interproduct variability, respectively. Random effects were implemented in *mgcv* using penalized regression spline terms with a random-effect basis ("re"), corresponding to random intercepts for repeated observations at the consumer and product levels. For model identifiability, the first cut-point parameter is fixed internally in the OCAT formulation.

Thin-plate regression splines were employed as the smoothing basis. For interaction screening and model comparison across alternative fixed-effect structures (e.g., interaction models), smoothing parameters were initially estimated using maximum likelihood (ML) within the penalized likelihood framework of *mgcv*, which enables direct comparison of model fit via ΔAIC. For final model estimation and interpretation of smooth effects and interaction effects, the selected models were subsequently refitted using restricted maximum likelihood (REML) to obtain more stable smoothing parameter estimates.

To prevent overfitting and suppress excessive wiggliness in interaction surfaces, penalized smoothing with automatic term selection (select = TRUE) was applied within the mgcv framework. The complexity of tensor-product smooth terms was evaluated using the effective degrees of freedom (edf), where edf = 1 indicates a linear relationship and edf >1 indicates increasing nonlinearity in the predictor–response relationship.

Prediction and category mapping

For each observation, the fitted ordinal GAMM yields estimated probabilities for each of the nine hedonic categories based on the cumulative probability structure of the model. In practice, category-level probabilities were derived from the fitted cumulative model, which internally converts cumulative probabilities into category-specific probabilities for each ordinal level. These probabilities reflect the model-based likelihood that a given observation belongs to each ordinal level of the 9-point liking scale.

The final predicted hedonic category was determined by selecting the ordinal level with the highest estimated probability among the nine categories. These predicted categories were subsequently used to compute cross-validated weighted kappa and exact accuracy.

Model adequacy and interpretability analysis

Model adequacy was assessed using randomized quantile (Dunn–Smyth) residual diagnostics, a concurvity evaluation, and a visual inspection of the smooth-term structures on the link scale. In addition, the edf associated with each smooth term was assessed to determine the degree of nonlinearity in the predictor–response relationships.

The relative contribution of each sensory predictor was quantified using two complementary indices: (i) the standard deviation (SD), representing the overall variance contribution to the linear predictor, and (ii) the peak-to-peak amplitude, reflecting the magnitude of the response sensitivity to attribute intensity changes.

These GAMM-based interpretations were compared with the VIP scores obtained using the PLSR model to evaluate the consistency and complementarity between the PLSR and GAMM perspectives on the sensory drivers of liking. Figure 1 summarizes the overall analytical workflow, including the PLSR and subsequent GAMM modeling procedures.

Interaction surface estimation

To explore potential synergistic or antagonistic interactions between critical sensory attributes, two-dimensional tensor-product smooth functions, $te(X_i, X_j)$, were fitted for data-driven significant attribute pairs identified using the ΔAIC comparison and Benjamini–Hochberg false discovery rate (FDR) correction.

In the *mgcv* framework, $te(\cdot)$ represents a tensor product of smooth basis functions that flexibly model nonlinear interaction surfaces between predictors with potentially different scales. To mitigate the risk of overfitting and suppress excessive wiggliness in the estimated surfaces, penalized smoothing with automatic term selection (select = TRUE) was applied during model estimation. This shrinkage mechanism allows non-informative smooth components to be penalized toward zero, thereby reducing the likelihood of detecting spurious interaction patterns. Interaction effects were visualized only within the data-supported region defined by the 5–95% quantiles of the predictor distributions in order to avoid extrapolation outside the observed data range.

These interaction surfaces characterize how combinations of sensory attribute intensities jointly influence the predicted liking response in the fitted ordinal GAMM. To quantify the practical magnitude of interaction effects on the observed hedonic scale, predicted category probabilities from the ordinal GAMM were used to compute the expected liking score, E[Y]. Differences between additive and interaction models were visualized as Δ-surfaces, representing deviations from additive expectations and enabling the interpretation of cross-modal sensory relationships (e.g., flavor–texture or taste–appearance interactions) that are not captured by conventional linear models. Positive values of ΔE[Y] indicate synergistic departures from additive predictions, whereas negative values indicate antagonistic departures.

For each interaction pair satisfying the screening criteria, the expected liking difference between the additive and interaction models (ΔE[Y]) was calculated across a trimmed predictor grid. In addition to the maximum and mean absolute interaction effects, the proportion of the predictor space in which the expected liking difference exceeded 0.5 hedonic units (defined here as a practical just-noticeable difference, JND, on the observed 9-point scale) was quantified, allowing statistically significant but perceptually negligible interactions to be distinguished[57].

Uncertainty in the observed-scale interaction effect sizes was evaluated using a panelist-cluster bootstrap procedure applied to the top-ranked interaction pairs. In each bootstrap iteration, panelists were resampled with replacement while preserving the hierarchical data structure. The interaction models were refitted for each bootstrap sample, and the maximum absolute deviation on the observed hedonic scale (max $|\Delta E[Y]|$) was recalculated. The resulting bootstrap distribution was used to derive 95% confidence intervals for the interaction effect size.

Liking modeling using OLR

An OLR model with linear predictors was additionally fitted as a linear ordinal baseline to isolate the contribution of nonlinear smooth components in the ordinal GAMM framework. The OLR model was fitted using the same predictor variables and nested cross-validation framework described above, and was implemented using the *polr* function in the MASS package in R.

Cross-validated predictive performance

Predictive performance was assessed using a nested cross-validation framework in which the outer loop consisted of ten folds and the inner loop consisted of five folds, thereby ensuring an unbiased evaluation of generalization performance and preventing information leakage during hyperparameter tuning. Within each outer fold, models were trained exclusively on the corresponding training subset, and out-of-fold (OOF) predictions were generated for the held-out test subset. All model selection and tuning procedures were conducted strictly within the training data. Final performance metrics were calculated using the aggregated OOF predictions across all folds.

To evaluate model generalizability under different data-partitioning scenarios, three validation schemes were implemented: random 10-fold cross-validation, product-grouped leave-one-product-out cross-validation (LOPO), and panelist-grouped leave-one-panelist-out cross-validation (LOPaO).

Weighted kappa (κ_W)

Predictive agreement between observed and predicted hedonic categories was evaluated using the weighted kappa statistic (κ_W) with quadratic weighting. For both modeling frameworks, predicted outputs were mapped to the corresponding nine-point ordinal categories (1–9) to obtain final classifications, as defined in Eq. (2):

$$\kappa_W = \frac{\sum_{i,j} w_{ij} O_{ij} - \sum_{i,j} w_{ij} E_{ij}}{1 - \sum_{i,j} w_{ij} E_{ij}}, \tag{2}$$

where O_{ij} and E_{ij} denote the observed and expected proportions of ratings in category pair (i,j), respectively, and w_{ij} represents the weight assigned to the disagreement between categories i and j. The weight w_{ij} is typically defined using a quadratic weighting scheme:

$$w_{ij} = 1 - \left(\frac{i-j}{k-1}\right)^2, \tag{3}$$

where k denotes the total ordinal categories ($k = 9$). The cross-validated κ_W (quadratic), computed from OOF predictions, was used as the primary metric for model comparison.

Exact accuracy

Exact accuracy was defined as the proportion of observations for which the predicted category exactly matched the observed category:

$$\text{Accuracy} = \frac{1}{n}\sum_{i=1}^{n} 1\left(\widehat{y}_i = y_i\right) \tag{4}$$

where y_i denotes the observed hedonic category, $\widehat{y}_i$ the predicted category, n the total number of observations, and $1(\cdot)$ the indicator function.

Accuracy within ±1 category

Accuracy within ±1 category was additionally calculated to reflect practically acceptable deviations in ordinal predictions. This metric was defined as the proportion of observations for which the absolute difference between the observed and predicted hedonic categories was less than or equal to one:

$$\text{Accuracy}_{\pm 1} = \frac{1}{n}\sum_{i=1}^{n} 1\left(y_i - \widehat{y}_i \le 1\right) \tag{5}$$

where y_i denotes the observed hedonic category, $\widehat{\mathbf{y}}_\mathbf{i}$ the predicted category, n the total number of observations, and $1(\cdot)$ the indicator function. This metric accounts for near-miss predictions in ordinal scales, where a one-category deviation may represent a minor discrepancy rather than substantive disagreement. Similar "accuracy within n" measures have been proposed for evaluating ordinal classification performance[58]. All accuracy metrics, including exact accuracy and ±1 accuracy, were computed from aggregated OOF predictions obtained under the nested cross-validation framework.

Data availability

The authors declare that all data supporting the findings of this study are included in the article. The code used in this study is not publicly available at the time of publication, as it is part of ongoing research, but is available from the corresponding author upon reasonable request.

Received: 22 December 2025; Accepted: 12 May 2026;
Published online: 23 May 2026

References

1. Stewart-Knox, B. & Mitchell, P. What separates the winners from the losers in new food product development? *Trends Food Sci. Technol.* **14**, 58–64 (2003).
2. Van Kleef, E., Van Trijp, H. C. & Luning, P. Internal versus external preference analysis: An exploratory study on end-user evaluation. *Food Qual. Prefer.* **17**, 387–399 (2006).
3. Bi, J. A review of statistical methods for determination of relative importance of correlated predictors and identification of drivers of consumer liking. *J. Sens. Stud.* **27**, 87–101 (2012).
4. Zhang, X. et al. Sensory profile and Beijing youth preference of seven cheese varieties. *Food Qual. Prefer.* **22**, 101–109 (2011).
5. Yang, J.-E. & Lee, J. Consumer perception and liking, and sensory characteristics of blended teas. *Food Sci. Biotechnol.* **29**, 63–74 (2020).
6. Su, Y. et al. Sensory description and consumer hedonic perception of ultra-high temperature (UHT) milk. *Foods* **11**, 1350 (2022).
7. Ju, S. et al. Development of nano soy milk through sensory attributes and consumer acceptability. *Foods* **10**, 3014 (2021).
8. Frauenhoffer, S., Ola, E., Jervis, S., Gauch, S. E. & Seo, H. S. Comparison analyses of consumer acceptance, evoked emotions, and purchase intent between gluten-containing versus gluten-free chocolate chip cookie products. *J. Food Sci.* **90**, e70276 (2025).
9. Cheng, Q. & Sun, D.-W. Application of PLSR in correlating physical and chemical properties of pork ham with different cooling methods. *Meat Sci.* **70**, 691–698 (2005).
10. Endrizzi, I. et al. Relationships between intensity and liking for chemosensory stimuli in food models: a large-scale consumer segmentation. *Foods* **11**, 5 (2021).
11. Zou, W. et al. Targeted prediction of sensory preference for fermented pomegranate juice based on machine learning. *LWT* **201**, 116260 (2024).
12. Krishnamurthy, R., Srivastava, A., Paton, J., Bell, G. & Levy, D. Prediction of consumer liking from trained sensory panel information: evaluation of neural networks. *Food Qual. Prefer.* **18**, 275–285 (2007).
13. Moskowitz, H. R. Magnitude estimation: notes on what, how, when, and why to use it. *J. Food Qual.* **1**, 195–227 (1977).
14. Lim, J. Hedonic scaling: a review of methods and theory. *Food Qual. prefer.* **22**, 733–747 (2011).
15. Alves, J. C. R., Palma, G. R. & de Lara, I. A. R. Unified beta regression model with random effects for the analysis of sensory attributes. Preprint at arXiv:2504.05996 (2025).

16. Urdapilleta, I., TIJUS, C. A. & Nicklaus, S. Sensory evaluation based on verbal judgments. *J. Sens. Stud.* **14**, 79–95 (1999).
17. Cabrera, M. & Taylor, G. Modelling spatio-temporal data of dengue fever using generalized additive mixed models. *Spat. Spatiotemporal Epidemiol.* **28**, 1–13 (2019).
18. Ding, M. et al. Delineation of body mass index trajectory predicting lowest risk of mortality in US men using generalized additive mixed model. *Ann. Epidemiol.* **26**, 698–703. e692 (2016).
19. Ferrari, C. et al. Application of a generalized additive mixed model in time series study of dairy cow behavior under hot summer conditions. *J. Dairy Sci.* **108**, 1554–1572 (2025).
20. Mare, Y., Zewotir, T. & Belay, D. B. Additive modeling of zonal level crop production in Ethiopia. *Sci. Rep.* **15**, 12391 (2025).
21. Baayen, R. H., van Rij, J., De Cat, C. & Wood, S. *Mixed-Effects Regression Models In Linguistics* (Springer, 2018).
22. IMARC Group. Plant-based meat market size, share, trends and forecast by product type, source, meat type, distribution channel, and region, 2025–2033. IMARC Group (2025).
23. Barker, S. & McSweeney, M. B. Sensory characterization of yellow pea and ground chicken hybrid meat burgers using static and dynamic methodologies. *J. Food Sci.* **87**, 5390–5401 (2022).
24. Elzerman, J. E., van Boekel, M. A. J. S. & Luning, P. A. Exploring meat substitutes: consumer experiences and contextual factors. *Br. Food J.* **115**, 700–710 (2013).
25. Araripe, P. P. & de Lara, I. A. R. Ordinal data and residual analysis: review and application. *Braz. J. Biometrics* **41**, 287–310 (2023).
26. Liu, D. & Zhang, H. Residuals and diagnostics for ordinal regression models: a surrogate approach. *J. Am. Stat. Assoc.* **113**, 845–854 (2018).
27. McCullagh, P. Regression models for ordinal data. *J. R. Stat. Soc. B* **42**, 109–127 (1980).
28. White, H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* **48**, 817–838 (1980).
29. Allison, P. D. Comparing logit and probit coefficients across groups. *Sociol. Methods Res.* **28**, 186–208 (1999).
30. Tutz, G. & Berger, M. Tree-structured scale effects in binary and ordinal regression. *Stat. Comput.* **31**, 17 (2021).
31. Appiani, M., Cattaneo, C. & Laureati, M. Sensory properties and consumer acceptance of plant-based meat, dairy, fish and eggs analogs: a systematic review. *Front. Sustain. Food Syst.* **7**, 1268068 (2023).
32. Palanisamy, M., Töpfl, S., Aganovic, K. & Berger, R. G. Influence of iota carrageenan addition on the properties of soya protein meat analogues. *LWT* **87**, 546–552 (2018).
33. Majzoobi, M., Talebanfar, S., Eskandari, M. H. & Farahnaky, A. Improving the quality of meat-free sausages using κ-carrageenan, konjac mannan and xanthan gum. *Int. J. Food Sci. Technol.* **52**, 1269–1275 (2017).
34. Vicente-Gonzalez, L., Frutos-Bernal, E. & Vicente-Villardon, J. L. Partial least squares regression for binary data. *Mathematics* **13**, 458 (2025).
35. Christensen, R. H. B. & Brockhoff, P. B. Analysis of sensory ratings data with cumulative link models. *J. Soc. Fr. Statistique* **154**, 58–79 (2013).
36. Plaehn, D. & Lundahl, D. S. An L-PLS preference cluster analysis on French consumer hedonics to fresh tomatoes. *Food Qual. Prefer.* **17**, 243–256 (2006).
37. Ghosh, D. & Chattopadhyay, P. Application of principal component analysis (PCA) as a sensory assessment tool for fermented food products. *J. Food Sci. Technol.* **49**, 328–334 (2012).
38. Armitage, R. M., Iatridi, V., Vi, C. T. & Yeomans, M. R. Phenotypic differences in taste hedonics: The effects of sweet liking. *Food Qual. Prefer.* **107**, 104845 (2023).
39. Sóskuthy, M. Generalised additive mixed models for dynamic analysis in linguistics: A practical introduction. Preprint at arXiv:1703.05339 (2017).
40. Gertheiss, J., Scheipl, F., Lauer, T. & Ehrhardt, H. Statistical inference for ordinal predictors in generalized additive models with application to bronchopulmonary dysplasia. *BMC Res. Notes* **15**, 112 (2022).
41. de Souza, V. R., Popper, R., Plamenov, V., Wojnicz, P. & Martinez, J. Traditional preference mapping and computational machine learning techniques: a comparative study of approaches to guide product development. *Food Qual. Prefer.* **120**, 105251 (2024).
42. Driver, J. & Spence, C. Multisensory perception: beyond modularity and convergence. *Curr. Biol.* **10**, R731–R735 (2000).
43. Ettinger, L. et al. Consumer perception and acceptability of plant-based alternatives to chicken. *Foods* **11**, 2271 (2022).
44. Petrat-Melin, B. & Dam, S. Textural and consumer-aided characterisation and acceptability of a hybrid meat and plant-based burger patty. *Foods* **12**, 2246 (2023).
45. Asgar, M. A., Fazilah, A., Huda, N., Bhat, R. & Karim, A. A. Nonmeat protein alternatives as meat extenders and meat analogs. *Compr. Rev. Food Sci. Food Saf.* **9**, 513–529 (2010).
46. Oppermann, A. K. L., de Graaf, C., Scholten, E., Stieger, M. & Piqueras-Fiszman, B. Comparison of rate-all-that-apply (RATA) and descriptive sensory analysis (DA) of model double emulsions with subtle perceptual differences. *Food Qual. Prefer.* **56**, 55–68 (2017).
47. Jeong, Y., Kwak, H. S., Lim, M., Kim, Y. J. & Lee, Y. Validation of optimization methods for sensory characteristics using rate-all-that-apply and intensity scales: a case study of apple juice. *Foods* **13**, 2853 (2024).
48. Nho, J.-E., Shin, J.-W. & Kim, M.-A. Sensory drivers of liking for plant-based meat alternatives: insights from check-all-that-apply with ideal profile integration. *Food Sci. Biotechnol.* **34**, 3579–3588 (2025).
49. Adhikari, K. et al. Development of a lexicon for beef flavor in intact muscle. *J. Sens. Stud.* **26**, 413–420 (2011).
50. Chun, S., Chambers, E. I. V. & Chambers, D. H. Effects of shiitake (*Lentinus edodes* P.) mushroom powder and sodium tripolyphosphate on texture and flavor of pork patties. *Foods* **9**, 611 (2020).
51. Chung, L. & Chung, S. J. Cross-cultural comparisons among the sensory characteristics of fermented soybean using Korean and Japanese descriptive analysis panels. *J. Food Sci.* **72**, S676–S688 (2007).
52. Maughan, C., Tansawat, R., Cornforth, D., Ward, R. & Martini, S. Development of a beef flavor lexicon and its application to compare the flavor profile and consumer acceptance of rib steaks from grass or grain-fed cattle. *Meat Sci.* **90**, 116–121 (2012).
53. Ares, G. et al. Visual attention by consumers to check-all-that-apply questions: insights to support methodological development. *Food Qual. Prefer.* **32**, 210–220 (2014).
54. Wold, S., Sjöström, M. & Eriksson, L. PLS-regression: a basic tool of chemometrics. *Chemometrics Intell. Lab. Syst.* **58**, 109–130 (2001).
55. Shadish, W. R., Zuur, A. F. & Sullivan, K. J. Using generalized additive (mixed) models to analyze single case designs. *J. Sch. Psychol.* **52**, 149–178 (2014).
56. Hastie, T. J. in *Statistical models in S* (eds Chambers, J. M. & Hastie, T. J.) Ch. 7 (CRC Press, 2017).
57. Gacula, M. Jr. & Rutenbeck, S. Sample size in consumer test and descriptive analysis. *J. Sens. Stud.* **21**, 129–145 (2006).
58. Gaudette, L. & Japkowicz, N. Evaluation methods for ordinal classification. In *Canadian Conference on Artificial Intelligence* 207–210 (Springer, 2009).

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) (2023R1A2C1006540) funded by the Ministry of Science and ICT (MSIT). This work was also supported by the NRF (2022R1C1C1012074 and RS-2025-00521107) funded by MSIT. In addition, this work was supported by the Regional Innovation System & Education (RISE) through the Gwangju

RISE Center, funded by the Ministry of Education (MOE) and the Gwangju Metropolitan Government, Republic of Korea (2025-RISE-05-011).

Author contributions
Sol Kim: Writing–original draft, conceptualization, formal analysis, methodology, software, validation, and visualization. Ju-Eun Nho: Writing–original draft, conceptualization, data curation, investigation, methodology, and resources. Hyoung Il Son: Writing–review & editing, project administration, and supervision. Min-A Kim: Writing–review & editing, conceptualization, project administration, resources, and supervision. Soo-Jung Kim: Writing–review & editing, conceptualization, project administration, and supervision.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41538-026-00891-8.

Correspondence and requests for materials should be addressed to Min-A Kim or Soo-Jung Kim.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.