INTERNATIONAL JOURNAL
OF INTELLIGENT SYSTEM
AND INFORMATION
PROCESSING

Blind Storage-Day Identification of Cherry Tomatoes Using a Nonlinear Electronic-Nose Sensor Response Signature Framework

Nurgül SENYÜCEL
Canakkale Onsekiz Mart University
nurgulsenyucel@comu.edu.tr

Servet SENYÜCEL
Istanbul Arel University
servetsenyucel@arel.edu.tr

Abstract

Accurate determination of storage duration is essential for maintaining the postharvest quality of cherry tomatoes and supporting efficient inventory management throughout the supply chain. Although electronic noses have been widely investigated for freshness evaluation, most existing studies rely on discrete sensor features or conventional machine-learning models that do not fully exploit the temporal dynamics of sensor responses. This study proposes a nonlinear Sensor Response Signature (SRS) framework for blind storage-day identification of cherry tomatoes using temporal electronic-nose measurements. A reference dataset consisting of three independent cherry tomato groups stored for seven consecutive days was established, while an additional independent blind dataset was reserved exclusively for external validation. Temporal responses from four metal-oxide semiconductor (MOS) gas sensors were modeled using Logistic, Gompertz, and Richards functions. Comparative evaluation based on goodness-of-fit statistics and information criteria demonstrated that the Richards model provided the most accurate and consistent continuous representation of the sensor-response trajectories. Eight kinetic descriptors were extracted from each fitted response to construct a 32-dimensional Sensor Response Signature for every sample, reducing the original temporal data dimensionality by **84.31%** while preserving response amplitude, kinetics, timing, and cumulative characteristics. A sample-level reference SRS library was subsequently developed, and blind storage-day identification was performed using nearest-signature matching. The proposed framework correctly identified all **21 independent blind samples**, achieving **100% exact-day accuracy**, **100% ±1-day accuracy**, **Cohen's κ = 1.000**, and **zero MAE and RMSE**. These findings demonstrate that storage-related information is embedded in the complete temporal evolution of electronic-nose responses rather than in isolated endpoint measurements. The proposed nonlinear SRS framework provides a rapid, non-destructive, interpretable, and computationally efficient approach for blind storage-day identification of cherry tomatoes and offers a promising methodology for intelligent postharvest quality monitoring of fresh horticultural products.

Keywords: Cherry tomatoes; Electronic nose; Sensor Response Signature (SRS); Richards model; Blind storage-day identification; Nonlinear response modeling; Postharvest quality monitoring; Temporal sensor responses..

1. Introduction

1.1. Postharvest Quality Changes in Cherry Tomatoes

Cherry tomato (*Solanum lycopersicum* L. var. *cerasiforme*) is a highly valued horticultural product because of its attractive appearance, characteristic sweet–acid balance, intense aroma, convenient size, and relatively high concentrations of bioactive compounds. However, cherry tomatoes remain metabolically active after harvest and undergo continuous physiological, biochemical, structural, and microbiological changes during storage. These changes progressively affect fruit appearance, texture, flavor, nutritional value, and marketability. Consequently, the postharvest quality of cherry tomatoes is not a static property but a time-dependent condition resulting from the combined effects of respiration, transpiration, ripening, senescence, oxidative metabolism, and microbial activity.

Following harvest, respiration continues through the consumption of stored organic substrates, including sugars and organic acids. The energy produced through respiration supports cellular metabolism, while the associated loss of substrates gradually modifies sweetness, acidity, aroma, and overall sensory quality. Storage temperature strongly regulates this process. Higher temperatures generally accelerate respiration, ripening, water loss, tissue softening, and microbial deterioration, whereas moderately reduced temperatures delay these changes. Recent research has shown that storage at 4 and 14 °C can better preserve tomato appearance, total soluble solids, firmness, and water status than storage at 24 °C. Nevertheless, excessively low temperatures may suppress the biosynthesis or release of desirable volatile compounds and can induce chilling-related physiological disorders in temperature-sensitive tomato cultivars (Li et al., 2025).

Water loss is among the earliest and most visible indicators of postharvest deterioration. Transpiration through the fruit surface results in progressive mass loss, surface shrinkage, reduced gloss, and tissue dehydration. Because cherry tomatoes have a relatively high surface-area-to-volume ratio, even moderate moisture loss can noticeably affect their visual freshness and firmness. Increased storage temperature and reduced relative humidity intensify this process. As storage proceeds, water loss is frequently accompanied by changes in membrane permeability and cellular turgor, leading to softening and a decline in mechanical resistance. Although temperature management can substantially reduce weight loss, the magnitude of deterioration also depends on cultivar, harvest maturity, fruit size, cuticle characteristics, and initial physiological condition.

Firmness is another critical quality attribute because it determines resistance to handling damage and strongly influences consumer acceptance. During storage, softening occurs through the coordinated modification of cell-wall polysaccharides and the loss of cell turgor. Pectin solubilization, depolymerization of structural carbohydrates, and the activity of enzymes associated with cell-wall disassembly gradually weaken the fruit tissue. Li et al. (2025) reported that low-temperature storage reduced tomato softening and linked this response to temperature-dependent regulation of genes associated with cellulose synthesis, expansin activity, pectate lyase, and other cell-wall processes. Similarly, tomatoes stored at 12 °C retained firmness more effectively than fruit maintained at approximately 22 °C, demonstrating the strong dependence of textural deterioration on both storage duration and temperature (Al-Gaadi et al., 2024).

Color development represents a further prominent postharvest transformation. The transition from less mature coloration toward an intense red appearance is mainly associated with chlorophyll degradation and the synthesis and accumulation of carotenoids, particularly lycopene. During the initial phase of storage, continued ripening may enhance red color and improve visual attractiveness. However, prolonged storage can produce excessive darkening, uneven pigmentation, loss of brightness, cuticle deterioration, and visible decay. Al-Gaadi et al. (2024) demonstrated that storage conditions substantially influenced tomato color development and that lower storage temperature slowed the rate of color change. In cherry tomatoes, therefore, color alone may indicate ripening progression but may not provide a sufficiently comprehensive measure of freshness, because fruit with an acceptable external color may already exhibit internal softening, volatile changes, oxidative damage, or early microbial deterioration.

Changes in total soluble solids, sugars, organic acids, and pH contribute directly to the characteristic flavor of cherry tomatoes. Total soluble solids may initially increase as ripening progresses and complex reserves are converted into soluble compounds. However, subsequent respiratory consumption may stabilize or reduce soluble-solid levels during extended storage. Titratable acidity commonly declines because organic acids are used as substrates in respiration, while pH tends to increase. The balance between sugars and acids is particularly important in cherry tomatoes, which are generally appreciated for their relatively concentrated flavor. Therefore, even when total soluble solids appear stable, a decline in acidity or a shift in the sugar-to-acid ratio may alter perceived sweetness, sourness, and flavor intensity. Recent studies confirm that temperature and storage duration jointly determine the direction and magnitude of these physicochemical changes (Al-Gaadi et al., 2024; Li et al., 2025).

Postharvest storage also affects nutritionally important constituents, including ascorbic acid, phenolic compounds, carotenoids, and antioxidant activity. These compounds may show non-linear patterns during storage. Continued ripening and stress-related defense responses can temporarily increase some antioxidant-related metabolites, whereas prolonged storage, senescence, oxidation, and membrane damage eventually promote their degradation. Li et al. (2025) found that storage at 4 °C delayed the loss of ascorbic acid compared with higher-temperature conditions. Likewise, preservation treatments that reduce respiration and oxidative stress have been shown to maintain higher vitamin C levels and antioxidant capacity in cherry tomatoes. High-voltage electrostatic-field treatment, for example, reduced weight loss and

color change while preserving firmness, vitamin C, antioxidant capacity, microstructural integrity, and microbial quality during storage (Zhao et al., 2023).

Oxidative stress becomes increasingly important as storage advances. The accumulation of reactive oxygen species can promote lipid peroxidation, membrane disruption, pigment degradation, electrolyte leakage, and accelerated senescence. The capacity of enzymatic and non-enzymatic antioxidant systems to control reactive oxygen species therefore influences the rate at which fruit quality deteriorates. Ding et al. (2024) demonstrated that strengthening the antioxidant system of cherry tomatoes through a composite edible coating reduced respiration and helped retain postharvest quality. Their results showed increased activities of superoxide dismutase, catalase, peroxidase, and ascorbate peroxidase, together with improved retention of ascorbic acid and glutathione. These findings indicate that antioxidant metabolism is closely linked to the observable storage-dependent changes in firmness, color, decay incidence, and nutritional quality.

Microbial proliferation and decay become major determinants of quality during the later stages of storage. Surface injuries, condensation, tissue softening, nutrient leakage, and weakening of natural defense mechanisms facilitate the development of spoilage microorganisms. Visible decay may be preceded by less obvious biochemical and volatile changes; therefore, conventional visual inspection may fail to identify the onset of deterioration. Preservation studies have shown that treatments capable of reducing respiration, maintaining membrane integrity, and limiting microbial growth can substantially delay the loss of commercial quality. For instance, electrostatic-field treatment decreased microbial loads and delayed quality deterioration in stored cherry tomatoes, while recent combinations of low-voltage electrostatic fields and modified-atmosphere packaging have demonstrated improved quality retention during extended storage (Zhang et al., 2025; Zhao et al., 2023).

A particularly important consequence of postharvest metabolism is the continuous alteration of the fruit's volatile organic compound profile. Tomato aroma arises from a complex mixture of aldehydes, alcohols, ketones, esters, terpenes, and other volatile compounds derived from fatty acids, amino acids, and carotenoids. As ripening and senescence progress, the production and relative abundance of these compounds change, while microbial activity may generate additional metabolites associated with over-ripening or decay. Storage temperature can preserve some quality attributes while simultaneously suppressing desirable aroma compounds or limiting the accumulation of undesirable volatiles (Li et al., 2025). Thus, volatile composition provides a dynamic chemical representation of the fruit's physiological state rather than merely reflecting its external appearance.

Overall, the postharvest condition of cherry tomatoes emerges from interacting changes in mass, firmness, color, soluble solids, acidity, antioxidant status, membrane integrity, microbial load, and volatile composition. Most conventional measurements quantify only one or a few of these attributes and often require destructive sampling, laboratory analysis, or prolonged measurement procedures. Furthermore, individual quality parameters may not vary monotonically with storage time, making precise storage-day identification difficult when a single physicochemical indicator is used. Because volatile emissions integrate several simultaneous metabolic and microbial processes, electronic-nose sensor responses may provide a rapid and non-destructive means of capturing storage-dependent quality changes. Accordingly, the systematic evolution of volatile-response signatures during storage offers a scientifically plausible basis for identifying or predicting the storage day of cherry tomatoes.

1.2 Electronic-Nose-Based Assessment of Cherry Tomatoes

The postharvest quality of cherry tomatoes is conventionally evaluated using physicochemical, mechanical, microbiological, and sensory measurements. Common quality indicators include mass loss, firmness, total soluble solids, titratable acidity, pH, color, respiration rate, antioxidant activity, and microbial load. Although these measurements provide valuable information, many require destructive sampling, specialized laboratory equipment, trained personnel, and relatively long analytical procedures. Moreover, a single quality attribute cannot fully represent the complex physiological state of stored cherry tomatoes because ripening, senescence, water loss, tissue softening, oxidative metabolism, and microbial deterioration occur simultaneously. Electronic-nose technology offers a rapid and non-destructive alternative by measuring the collective volatile profile released from the fruit headspace.

An electronic nose generally consists of a sampling or headspace unit, an array of partially selective gas sensors, a data-acquisition system, and pattern-recognition or machine-learning algorithms. Unlike gas chromatography–mass spectrometry, which separates and identifies individual volatile organic compounds, an electronic nose produces a multidimensional response pattern representing the overall composition of the sampled aroma. Each sensor exhibits a different but overlapping sensitivity to volatile compound groups. The combined sensor responses therefore form an odor

fingerprint that can be associated with ripeness, freshness, storage duration, physiological stress, or microbial deterioration. Recent reviews emphasize that the practical performance of electronic noses depends not only on the sensor materials but also on sampling conditions, signal preprocessing, feature extraction, environmental compensation, model validation, and the stability of the sensor array (Han et al., 2024; Zhai et al., 2024).

Tomato fruit provides a suitable biological system for electronic-nose analysis because its aroma is generated by a chemically diverse mixture of aldehydes, alcohols, ketones, esters, hydrocarbons, terpenoids, and sulfur- or nitrogen-containing compounds. The production and relative abundance of these volatiles change continuously during ripening and postharvest storage. Lipid-derived compounds such as hexanal, trans-2-hexenal, cis-3-hexenal, 1-hexanol, and cis-3-hexenol contribute to green and fresh aroma notes, whereas carotenoid-, amino-acid-, and phenylalanine-derived volatiles contribute to fruity, floral, and ripe-tomato characteristics. During prolonged storage, substrate depletion, membrane degradation, fermentation, oxidative processes, and microbial growth further modify the headspace composition. Consequently, the collective volatile profile can provide information about postharvest status that may not be apparent from external color alone.

Early research established the feasibility of monitoring tomato shelf life through aroma fingerprints. Berna et al. (2004) compared a quartz microbalance electronic nose with a mass-spectrometry-based electronic nose to investigate shelf-life and cultivar-related changes in tomato aroma. Measurements performed on several storage days demonstrated a progressive shift in aroma profiles during shelf life. However, the conventional sensor-based electronic nose could not clearly discriminate all neighboring storage periods, particularly the earliest measurement days. The mass-spectrometry-based system provided clearer shelf-life separation and indicated that specific volatile fragments associated with compounds such as β-phellandrene, 6-methyl-5-hepten-2-one, and 2-methylbutanol contributed to storage-related differences. These findings demonstrated both the diagnostic value of tomato volatiles and the difficulty of separating closely spaced storage stages using conventional multivariate analysis alone.

Gómez et al. (2008) subsequently investigated the use of a commercial PEN 2 electronic nose for monitoring tomatoes stored under two packaging conditions over 12 days. Principal component analysis and linear discriminant analysis revealed distinguishable changes in sensor patterns as storage progressed, particularly for tomatoes stored in carton boxes. Nevertheless, complete separation among all storage days was not achieved. In addition, models based solely on electronic-nose signals showed limited ability to predict total soluble solids, pH, and puncture force. This study provided important evidence that electronic noses can detect the progression of tomato storage but also showed that conventional score-plot interpretation and static sensor features may be insufficient for accurate quantitative estimation of closely related quality states.

Subsequent research increasingly combined electronic-nose signals with complementary sensing modalities and nonlinear prediction algorithms. Huang et al. (2018) integrated electronic-nose responses with computer-vision features to evaluate tomato ripeness and predict firmness during storage. Computer vision represented external color information, whereas the electronic nose captured changes in fruit aroma. The fused data produced more robust classification and regression models than either sensing modality alone. The support vector classifier achieved prediction-set accuracy above 94%, while support vector regression provided strong prediction performance for firmness. These findings indicate that odor information is complementary to visual quality indicators and that nonlinear learning algorithms can improve the interpretation of complex electronic-nose response patterns.

The most directly relevant investigation for cherry tomatoes was conducted by Feng et al. (2018), who used a metal-oxide-semiconductor electronic nose to assess the postharvest quality and freshness of cherry tomato fruit treated with high-pressure argon. Partial least-squares regression and an extreme learning machine were used to relate electronic-nose responses to firmness, soluble solids content, and pH. Based on flavor characteristics recorded throughout cold storage, the samples were divided into four freshness states: fresh, acceptable, stale, and very stale. The extreme learning machine outperformed the linear partial least-squares models, producing coefficients of determination above 0.95 for several quality attributes. The study therefore confirmed that MOS sensor arrays, when combined with nonlinear learning methods, can provide a reliable representation of cherry tomato freshness and can predict important physicochemical quality indices.

Despite this promising result, the existing cherry tomato study focused primarily on evaluating a postharvest preservation treatment and predicting broad freshness categories or physicochemical attributes. It did not address the more demanding problem of identifying each individual storage day from independent, previously unseen samples. Broad freshness

categories are typically separated by relatively large physiological differences, whereas consecutive storage days may exhibit strongly overlapping aroma profiles. Accurate storage-day identification therefore requires a more sensitive representation of sensor behavior than the use of endpoint responses or a limited number of conventional summary variables.

Recent developments in nondestructive tomato assessment support the transition from single-variable measurements toward richer, multimodal, and machine-learning-based representations. Han et al. (2024) reviewed emerging nondestructive methods for tomato quality inspection and highlighted electronic noses, electronic tongues, machine vision, hyperspectral imaging, and artificial intelligence as promising tools for rapid quality assessment. The review also emphasized that tomato quality is difficult to describe because cultivars and fruit types differ in size, color, chemical composition, and ripening behavior. These sources of biological variability should therefore be considered during model development and validation. Models evaluated only by random splitting may learn cultivar-, batch-, or measurement-specific patterns rather than reproducible storage-related information.

A recent multimodal investigation by Liu et al. (2024) further demonstrated that combining color imaging, visible–near-infrared spectroscopy, and haptic information can substantially improve tomato maturity classification. Their fused deep-learning model achieved 99.4% classification accuracy, exceeding the performance of each individual modality. Although this study did not incorporate an electronic nose, it provides an important methodological lesson: tomato maturity and quality are multidimensional, and a richer representation of the biological response generally improves discrimination. At the same time, multimodal systems involving cameras, spectrometers, and mechanical sensing devices may increase cost, system complexity, calibration requirements, and computational demand. A compact electronic nose may therefore offer a more accessible alternative when volatile changes contain sufficient information for the intended prediction task.

Metal-oxide-semiconductor sensors are particularly attractive for portable agricultural applications because they are inexpensive, compact, sensitive to a broad range of gases, and compatible with embedded data-acquisition platforms. Their responses, however, are influenced by temperature, humidity, sensor drift, baseline variation, chamber volume, sample mass, headspace accumulation time, and gas-flow conditions. MOS sensors are also cross-sensitive rather than compound-specific. This characteristic is not necessarily a disadvantage for electronic-nose applications, because classification is based on the collective sensor pattern rather than the concentration of a single gas. Nevertheless, robust experimental control and appropriate preprocessing are necessary to ensure that the resulting fingerprint reflects fruit-related volatile changes instead of uncontrolled environmental variation.

Most previous tomato electronic-nose studies have represented each measurement using equilibrium values, maximum responses, averages, selected response ratios, or a small number of manually extracted features. Such representations may discard important kinetic information contained in the full sensor response curve. When volatile compounds accumulate in a closed measurement chamber, MOS sensor signals commonly exhibit a rapid initial change followed by a progressively slower approach toward a plateau. The amplitude, initial rate, curvature, stabilization behavior, and asymptotic level of this response may all vary with volatile concentration and composition. Accordingly, two samples with similar final sensor values may still exhibit distinguishable transient response dynamics.

The entire time-dependent sensor response can therefore be considered a nonlinear signature of the sample headspace. Parameters derived from nonlinear curve models may quantify biologically meaningful characteristics such as response amplitude, apparent rate coefficient, curvature, inflection behavior, half-response time, time to near-saturation, and plateau level. Compared with isolated endpoint measurements, these response signatures may preserve subtle differences among consecutive storage days. They may also reduce the dimensionality of raw time-series data while retaining the dominant kinetic characteristics of each sensor.

Another limitation of the existing literature concerns validation. Many previous studies evaluated classification performance through random partitioning of measurements obtained from the same experimental batch. When repeated measurements from the same producer, storage batch, or biological group occur in both the training and test sets, the resulting performance may be optimistic. A model intended for practical storage-day identification should be evaluated using independent samples that were not involved in model fitting, feature selection, or parameter optimization. Blind testing is particularly important for cherry tomatoes because natural variability in cultivar, producer, maturity at harvest, fruit size, and initial volatile composition may influence electronic-nose responses.

Thus, although previous research has demonstrated that electronic noses can monitor tomato shelf life, distinguish broad maturity or freshness states, and predict selected physicochemical properties, reliable identification of individual cherry

tomato storage days remains insufficiently investigated. In particular, there is a lack of studies combining full dynamic MOS sensor responses, nonlinear response modelling, sensor-response signature extraction, machine learning, and independent blind validation. Addressing this gap could establish whether storage-related changes in cherry tomato volatiles produce reproducible kinetic fingerprints and whether those fingerprints can support accurate, rapid, and non-destructive storage-day identification.

1.3 Limitations of Conventional Machine-Learning Approaches

1.3. Limitations of Conventional Machine-Learning Approaches

Machine-learning methods have become central to the interpretation of electronic-nose data because the responses generated by cross-sensitive sensor arrays are multivariate, nonlinear, and frequently affected by interactions among volatile compounds. Conventional algorithms such as principal component analysis, linear discriminant analysis, partial least-squares discriminant analysis, support vector machines, k-nearest neighbors, decision trees, random forests, artificial neural networks, and extreme learning machines have consequently been applied to food classification, quality prediction, adulteration detection, and spoilage assessment. Although many studies report high classification accuracies, the practical reliability of these models depends strongly on how the sensor signals are represented, how the data are partitioned, and whether performance is evaluated on genuinely independent samples. Recent reviews have emphasized that high within-dataset accuracy does not necessarily demonstrate robustness under new products, batches, environmental conditions, devices, or measurement periods (Sanislav et al., 2025; Singh et al., 2025).

A fundamental limitation of many conventional electronic-nose models is their dependence on a small number of manually selected static features. Commonly used variables include maximum response, equilibrium response, mean response, response difference, response ratio, area under the curve, and selected slope values. These features simplify high-dimensional sensor signals and make classical machine-learning algorithms computationally efficient. However, they may discard a substantial proportion of the information contained in the complete response trajectory. Electronic-nose measurements are inherently time-dependent: adsorption, surface reactions, diffusion, desorption, chamber accumulation, and sensor recovery generate transient responses whose rates and shapes may vary even when their final amplitudes are similar. Treating each measurement as an unordered vector of endpoint values therefore ignores temporal dependence and may prevent the model from detecting subtle differences between physiologically adjacent storage days.

This limitation is particularly important for cherry tomatoes because postharvest quality evolves gradually rather than through sharply separated states. Consecutive storage days may produce strongly overlapping headspace compositions and only small changes in sensor amplitude. A classifier trained on final or maximum sensor responses may distinguish broadly defined categories such as fresh and stale fruit but fail to separate neighboring days. The earlier cherry tomato study by Feng et al. (2018), for example, demonstrated that MOS electronic-nose responses combined with an extreme learning machine could predict quality attributes and classify broad freshness levels. However, that study did not establish whether every individual storage day could be identified from blind samples. The difference is methodologically important: broad freshness classes contain relatively large quality contrasts, whereas day-by-day identification requires sensitivity to smaller and potentially nonlinear changes in sensor kinetics.

Linear projection and discrimination methods present an additional constraint. Principal component analysis is useful for visualization and dimensionality reduction, but it is unsupervised and seeks directions of maximum overall variance rather than directions that optimally separate storage days. The dominant components may therefore reflect sensor magnitude, humidity, batch variability, producer differences, or baseline shifts instead of the biological progression of storage. Linear discriminant analysis and partial least-squares-based methods use class information but still assume that relevant class structure can be represented adequately through linear combinations of the input variables. When sensor responses vary nonlinearly with volatile concentration or when adjacent storage days occupy overlapping and curved regions of feature space, linear decision boundaries may be insufficient.

Nonlinear conventional algorithms, including support vector machines, random forests, gradient-boosting models, and artificial neural networks, can model more complex relationships. Nevertheless, algorithmic nonlinearity does not compensate automatically for information lost during feature extraction. Once the complete response curve has been reduced to a few endpoint statistics, a sophisticated classifier cannot reconstruct the discarded response dynamics. In addition, highly flexible models can fit noise, batch-specific patterns, or sensor-specific artifacts when the number of observations is limited relative to the number of candidate features. This risk becomes greater when multiple sensors,

numerous manually engineered variables, and extensive hyperparameter searches are combined in a relatively small biological experiment.

Small datasets remain common in electronic-nose food studies because sample preparation, controlled storage, repeated headspace measurement, chamber cleaning, and reference analyses are time-consuming. In such datasets, prediction scores can vary substantially according to the selected train–test split. Flexible models may achieve apparently high accuracy on one random partition but produce markedly lower performance on another. Model comparisons based on a single split can therefore favor algorithms that happen to match the partition-specific structure rather than those that capture reproducible biological differences. The problem is further intensified when model selection and performance evaluation use the same validation observations.

Data leakage constitutes one of the most serious methodological risks. Leakage occurs whenever information from validation or test samples influences model construction. It may arise when normalization, baseline correction, outlier removal, dimensionality reduction, feature selection, oversampling, or hyperparameter optimization is performed before the data are divided into training and test sets. Rosenblatt et al. (2024) demonstrated that leakage caused by feature selection and non-independent repeated observations can substantially inflate predictive performance, with the greatest effects occurring in smaller datasets and in prediction problems with relatively weak underlying signals. Although their analysis concerned neuroimaging, the principle applies directly to electronic-nose experiments in which repeated curves, derived features, or multiple measurements from the same biological group are treated as independent samples.

Electronic-nose datasets are especially vulnerable to group-level dependence. Several response curves may originate from fruits obtained from the same producer, harvest batch, cultivar, storage container, experimental session, or sensor-conditioning cycle. Randomly distributing these measurements between training and test sets allows closely related samples to appear on both sides of the evaluation. The model may then learn producer-specific baselines, batch characteristics, chamber conditions, or device states instead of storage-day-related volatile patterns. Such a model can perform extremely well during random cross-validation while failing when applied to a new producer or independently collected test batch.

Evidence from multi-source machine-learning research shows that conventional random K-fold cross-validation frequently overestimates performance when the intended application involves new data sources. Leinonen et al. (2024) compared standard K-fold and leave-source-out validation and found that conventional K-fold evaluation systematically produced optimistic estimates of transfer performance, whereas source-level exclusion provided more realistic assessments, albeit with greater variability. For a cherry tomato study involving multiple producers or harvest batches, this evidence supports producer-wise, batch-wise, or leave-one-source-out validation rather than random observation-level splitting.

Another limitation is that conventional models generally assume that training and future observations follow the same statistical distribution. This assumption is difficult to maintain in electronic-nose applications. Sensor responses may change because of aging, poisoning, contamination, prolonged exposure, heater variation, baseline instability, humidity, ambient temperature, manufacturing variability, or changes in sample-matrix composition. As a result, a model calibrated during one measurement period may gradually become less accurate even when the biological classes remain unchanged. Long-term measurements of MOS arrays have confirmed that sensor drift remains a major obstacle and that meaningful distribution changes can develop over extended operating periods (Wörner et al., 2025).

Sensor drift is effectively a domain-shift problem: the feature distributions observed during model deployment differ from those used for model training. Conventional classifiers such as SVM, RF, KNN, or LDA do not inherently correct this shift. Their decision boundaries remain fixed unless the model is recalibrated or retrained. Recent research has therefore moved toward domain adaptation, domain generalization, contrastive learning, and continuously updated drift-compensation models. Chu et al. (2024), for instance, developed a contrastive domain-generalization convolutional network specifically to improve recognition under previously unseen drift conditions. The need for such methods demonstrates that high accuracy under same-period laboratory validation cannot be interpreted as evidence of long-term operational stability.

Cross-sensitivity also limits conventional interpretation. MOS sensors respond to overlapping groups of volatile and interfering compounds rather than to single analytes. Their signals can therefore be affected by water vapor, ethanol, hydrocarbons, reducing gases, and other compounds released simultaneously from the fruit. In cherry tomatoes, storage-related sensor changes may arise from ripening metabolism, tissue deterioration, microbial activity, moisture variation,

or combinations of these processes. A conventional classifier may discriminate storage days without clarifying which response characteristics are biologically meaningful. This makes it difficult to determine whether the model has learned a robust storage-related pattern or an incidental environmental correlate.

Interpretability is particularly problematic for highly flexible ensemble and kernel-based methods. Random forests and boosting algorithms can provide feature-importance scores, while support vector machines can produce accurate nonlinear boundaries; however, these outputs do not automatically explain the physical meaning of the underlying response. Conventional variables such as “sensor value at 100 s” provide only limited information about adsorption dynamics, response rate, stabilization behavior, or saturation. Moreover, correlated sensor features may divide importance among several variables or produce unstable rankings across validation folds. A model can therefore be predictively successful yet remain difficult to connect to the underlying volatile-response mechanism.

Class structure creates an additional challenge in storage-day prediction. Storage days are ordered rather than nominal categories: Day 3 is physiologically closer to Day 4 than to Day 7. Conventional multiclass classifiers generally treat all classification errors equally and ignore this ordinal relationship. Predicting Day 4 instead of Day 5 is penalized in the same manner as predicting Day 1 instead of Day 7 when ordinary accuracy is used. Such evaluation conceals whether errors occur mainly between neighboring days or across widely separated storage stages. Accuracy, precision, recall, and F1-score should therefore be supplemented with ordinally meaningful measures such as mean absolute error, root-mean-square error, and within-one-day accuracy.

Class overlap further restricts the attainable exact-day performance. Postharvest biological change does not necessarily progress at an identical rate in every fruit. Differences in harvest maturity, cultivar, producer, fruit mass, surface characteristics, mechanical damage, and initial microbial load can cause tomatoes stored for the same number of days to exhibit different volatile profiles. Conversely, fruit from adjacent storage days may show very similar headspace patterns. Conventional classifiers forced to assign each sample to a discrete day may generate unstable boundaries in these overlap regions. High model complexity may sharpen the apparent separation of the training data without creating a reproducible distinction in genuinely independent samples.

Another concern is the common reporting of only the best-performing algorithm or the highest cross-validation score. Selecting the best result after comparing many preprocessing schemes, feature sets, model families, and hyperparameter combinations introduces a model-selection bias when the final performance is not evaluated on a previously untouched test set. The reported score then reflects both the predictive signal and repeated adaptation to the available data. Reliable evaluation requires that preprocessing, feature selection, and hyperparameter optimization be conducted entirely within the training process, preferably through nested cross-validation, followed by final assessment on an independent blind dataset.

These limitations do not imply that conventional machine-learning algorithms are unsuitable for electronic-nose analysis. Random forests, SVMs, boosting methods, and related models remain valuable classifiers when they are supplied with informative features and evaluated through rigorous validation. The main weakness lies in using these algorithms with overly compressed static variables, non-independent random splits, and no external or blind testing. Under such conditions, high reported accuracy may describe interpolation within a specific experiment rather than generalizable identification of storage day.

A more appropriate strategy is therefore to improve both the representation of the sensor response and the validation design. Instead of describing each curve only through endpoint or maximum values, the complete transient response can be modeled using nonlinear functions. Parameters describing response amplitude, rate, curvature, asymptotic behavior, half-response time, and stabilization can then form a compact sensor-response signature. These parameters preserve interpretable kinetic information while reducing the dimensionality of the raw time series. When combined with producer-aware cross-validation, nested feature selection, and independent blind testing, such signatures may provide a more reliable basis for identifying subtle differences among consecutive cherry tomato storage days.

Accordingly, the major research gap is not simply the absence of another classification algorithm. It is the lack of an integrated framework that represents full nonlinear sensor dynamics, prevents group-level data leakage, accounts for ordered storage-day relationships, and verifies performance using independent samples. Addressing these limitations is essential for moving electronic-nose-based cherry tomato assessment from high within-experiment accuracy toward reproducible and practically credible storage-day identification.

1.4 Continuous Representation of Temporal Sensor Responses

Electronic-nose measurements are fundamentally dynamic processes rather than collections of independent sensor readings. When a fruit sample is introduced into a closed measurement chamber, volatile organic compounds (VOCs) gradually diffuse through the headspace and interact with the sensing surfaces of the metal-oxide-semiconductor (MOS) sensors. As adsorption, surface reactions, and desorption proceed, the electrical properties of the sensing material change continuously over time. Consequently, the output of each sensor forms a transient response curve that reflects the temporal evolution of the gas–sensor interaction rather than a single equilibrium measurement. Therefore, the measured sensor trajectory should be regarded as a continuous function of time instead of an unordered set of discrete observations.

For a given sensor, the dynamic response can be represented as

$$x_s(t), t \in [0, T],$$

where $x_s(t)$denotes the response of sensor $s$at time t, and $T$is the total acquisition duration. Although measurements are collected at discrete sampling intervals,

$$t_1, t_2, \ldots, t_m,$$

the acquired observations

$$x_s(t_1), x_s(t_2), \ldots, x_s(t_m)$$

are generated by a continuous physicochemical process. Neighboring observations are therefore strongly correlated because each measurement depends on the previous state of gas diffusion, adsorption kinetics, and surface reaction dynamics. Treating these observations as independent variables ignores the intrinsic temporal structure of the sensing process.

This interpretation is consistent with the principles of Functional Data Analysis (FDA), in which densely sampled observations are considered realizations of an underlying smooth function rather than isolated numerical values. FDA preserves the entire response trajectory and distinguishes between amplitude variation and phase variation. Amplitude variation represents differences in response magnitude, whereas phase variation describes differences in response timing or response speed. Such distinctions are particularly important in electronic-nose measurements because two samples may exhibit nearly identical equilibrium responses while differing considerably in their transient behavior (Gertheiss et al., 2023).

The measured response can be expressed as the sum of an underlying smooth function and measurement noise:

$$x_{is}(t_j) = f_{is}(t_j) + \varepsilon_{isj},$$

where $x_{is}(t_j)$denotes the observed response of sensor $s$for sample $i$at sampling instant t_j, $f_{is}(t)$is the continuous response function, and ε_{isj}represents measurement noise arising from sensor fluctuations, electrical interference, or environmental disturbances. The objective of continuous modeling is not to reproduce every local fluctuation but to recover the dominant response pattern associated with volatile accumulation inside the measurement chamber.

Unlike static descriptors such as maximum response or final sensor value, the continuous response curve preserves the complete temporal evolution of the sensing process. The transient portion of the response often contains valuable information about adsorption kinetics, gas diffusion, and sensor activation. Previous studies have demonstrated that transient MOS sensor signals can discriminate gases before equilibrium is reached, indicating that the early stages of the response carry chemically meaningful information that may disappear when only equilibrium measurements are considered (Vergara et al., 2007).

Recent electronic-nose studies have further emphasized that different portions of the response curve contribute differently to classification performance. Borowik et al. (2023) showed that the temporal structure of MOS sensor responses significantly influenced the discrimination of *Fusarium* species, suggesting that the transient response should be analyzed as a complete curve rather than summarized by a few manually selected features. Similarly, Haj Ammar et al. (2024) proposed a temporal filtering strategy that dynamically extracted informative conductometric features throughout the measurement period instead of relying exclusively on equilibrium values. Their findings demonstrated that preserving temporal evolution improved classification performance across multiple environmental conditions.

The continuous representation also provides a natural framework for nonlinear response modeling. Instead of directly using hundreds of sampled ADC values, each response trajectory can be approximated by a nonlinear function,

$$x_s(t) = g(t;\boldsymbol{\theta}_s) + \varepsilon_s(t),$$

where $g(\cdot)$represents a nonlinear response model and

$$\boldsymbol{\theta}_s = \left[\theta_{s1},\theta_{s2},\ldots,\theta_{sp}\right]$$

is the corresponding parameter vector. Depending on the selected nonlinear model, these parameters may characterize the response amplitude, kinetic coefficient, curvature, asymptotic response, or other dynamic properties describing the evolution of the sensor signal.

For an electronic nose consisting of *S*sensors, the complete response signature of sample *i*can be expressed as

$$\mathbf{F}_i = \left[\boldsymbol{\theta}_{i1},\boldsymbol{\theta}_{i2},\ldots,\boldsymbol{\theta}_{iS}\right],$$

where each parameter vector summarizes the dominant temporal behavior of one sensor. This representation substantially reduces data dimensionality while preserving the essential kinetic characteristics of the complete response curve.

Continuous modeling also enables derivative-based analysis of sensor kinetics. The first derivative represents the instantaneous response rate,

$$v_s(t) = \frac{dx_s(t)}{dt},$$

while the second derivative describes changes in the response rate,

$$a_s(t) = \frac{d^2x_s(t)}{dt^2}.$$

The first derivative identifies periods of maximum sensor activity, whereas the second derivative characterizes response curvature and transition behavior. Unlike numerical differentiation applied directly to noisy measurements, derivatives calculated from a fitted continuous model are considerably more stable and physically meaningful.

Another important advantage of continuous representation is its ability to distinguish differences in response magnitude from differences in response shape. Two cherry tomato samples may exhibit almost identical sensor values at the end of the acquisition period while following substantially different temporal trajectories. One sample may respond rapidly during the initial stage and stabilize early, whereas another may increase gradually throughout the measurement. Although conventional endpoint features would classify these responses as similar, continuous modeling reveals differences in response kinetics that may correspond to changes in volatile composition during storage.

Such temporal variations are biologically plausible because postharvest metabolism continuously modifies the composition of volatile organic compounds released by cherry tomatoes. Changes in respiration, lipid oxidation, amino-acid metabolism, carotenoid degradation, microbial activity, and membrane integrity influence not only the total concentration of volatile compounds but also their diffusion dynamics and interaction with MOS sensing materials. Consequently, storage-related information may be encoded in the complete temporal evolution of the sensor response rather than solely in its equilibrium value.

Continuous representation additionally provides a principled basis for comparing sensor responses. Instead of comparing isolated ADC values, similarity can be evaluated using fitted parameter vectors, integrated functional distances, or derivative-based metrics. Functional comparison allows discrimination between differences caused by overall response magnitude and those arising from response shape or timing, thereby providing a richer characterization of storage-related changes.

Recent reviews of electronic-nose technology consistently emphasize that future advances depend not only on improved sensor materials but also on more sophisticated temporal signal processing and mathematical representation of dynamic sensor responses (Sanislav et al., 2025; Zhai et al., 2024). As electronic noses increasingly employ advanced machine-learning algorithms, preserving the temporal characteristics of sensor responses becomes essential for extracting robust and interpretable information from complex volatile mixtures.

Accordingly, this study considers each MOS sensor response as a continuous nonlinear function generated by the interaction between the evolving volatile profile of cherry tomatoes and the sensing material. Rather than compressing

the measurement into a small number of endpoint features or directly using high-dimensional raw time-series data, the complete response trajectory is represented as a smooth mathematical function whose parameters capture the dominant kinetic behavior of the sensing process. These continuous representations provide the theoretical foundation for the nonlinear response modeling and sensor-response signature extraction described in the following sections.

1.5 Research Gap and Contributions

Electronic-nose systems have demonstrated considerable potential for the rapid and nondestructive evaluation of tomato ripeness, freshness, shelf life, and postharvest quality. Previous investigations have shown that sensor-array responses can reflect storage-related alterations in tomato volatile profiles and can be associated with physicochemical characteristics such as firmness, soluble solids content, pH, weight loss, and overall freshness. Nevertheless, the available literature remains dominated by broad freshness-state classification, treatment comparison, shelf-life monitoring, or regression-based prediction of selected quality parameters rather than the exact identification of individual storage days.

Early tomato studies established that electronic-nose fingerprints change progressively during storage. Berna et al. (2004) monitored shelf-life-related aroma changes using quartz-microbalance- and mass-spectrometry-based electronic noses, whereas Gómez et al. (2008) used a commercial sensor array to follow tomato quality during storage under different packaging conditions. Although these studies confirmed that tomato headspace composition evolves over time, complete discrimination among closely spaced storage days was not consistently achieved. This indicates that neighboring postharvest stages may exhibit substantial overlap when sensor responses are represented using conventional equilibrium values and analyzed through linear multivariate techniques.

Research specifically addressing cherry tomatoes is considerably more limited. Hong et al. (2015) applied electronic-nose measurements and chemometric classifiers to determine the freshness of cherry tomato juice. However, crushing the fruit altered the sample matrix and exposed internal tissues, meaning that the resulting measurements did not represent intact-fruit headspace assessment. Feng et al. (2018) subsequently demonstrated that a MOS electronic nose combined with an extreme learning machine could estimate physicochemical quality parameters and distinguish broad freshness states in cherry tomatoes subjected to high-pressure argon treatment. Their results confirmed the feasibility of electronic-nose-based cherry tomato assessment, but the experimental objective was the evaluation of a preservation treatment rather than the blind identification of each storage day in untreated fruit.

Recent research has strengthened the evidence that cherry tomato volatile composition changes systematically during postharvest storage. Guan et al. (2025) used headspace gas chromatography–ion mobility spectrometry to examine cherry tomatoes stored at 4 and 25 °C and reported dynamic, temperature-dependent changes in their volatile organic compound profiles. These findings provide a chemical basis for storage-day identification because MOS sensor arrays respond collectively to changing mixtures of volatile compounds. However, analytical characterization by HS-GC-IMS does not itself provide a low-cost, portable, or computationally simple method for routine storage-day assessment.

A particularly relevant recent study was conducted by Senge et al. (2025), who combined electronic-nose data with storage-condition records, color analysis, weight loss, and machine-learning models to predict tomato shelf life. Their data-fusion approach achieved 72.91% accuracy for storage-day classification, while k-nearest-neighbor regression produced a mean absolute error of 0.841 days and a coefficient of determination of 0.867. This work provides important contemporary evidence that tomato storage duration can be estimated through multisource sensing. At the same time, its moderate exact-day classification performance indicates that neighboring storage days remain difficult to separate, even when electronic-nose information is supplemented with visual and physical measurements. Moreover, the dependence on multimodal data increases hardware, calibration, data-integration, and deployment complexity compared with a compact electronic-nose-only system.

The first major research gap therefore concerns **the target of prediction**. Most previous cherry tomato electronic-nose studies have focused on broad freshness categories, preservation-treatment effects, or individual quality parameters. These objectives differ fundamentally from exact storage-day identification. Broad categories such as fresh, acceptable, stale, and very stale are separated by relatively large physiological differences. In contrast, consecutive storage days represent ordered and closely related biological states whose volatile profiles may overlap substantially. Consequently, successful freshness classification does not demonstrate that a model can reliably identify Day 2 rather than Day 3 or Day 6 rather than Day 7.

The exact storage-day identification task can be formulated as

$$\hat{d}_i = h \quad \left(\mathbf{X}_i\right), \hat{d}_i \in \{1,2,\ldots,D\},$$

where $\mathbf{X}_i$denotes the electronic-nose response of sample i, $h(\cdot)$is the identification model, $\hat{d}_i$is the predicted storage day, and $D$is the total number of storage days. Unlike broad freshness classification, this formulation requires the model to resolve subtle differences among adjacent and ordinally related classes.

The second major gap concerns **sensor-response representation**. Conventional electronic-nose studies commonly reduce each response curve to one or several static variables, such as the maximum signal, equilibrium value, mean response, response ratio, or area under the curve. Although such variables simplify model construction, they discard the continuous temporal organization of the sensor response. MOS signals are generated through nonlinear diffusion, adsorption, reaction, and stabilization processes. Storage-related changes may therefore influence not only signal magnitude but also response rate, curvature, saturation behavior, and time to stabilization.

Recent reviews have identified feature representation, temporal information extraction, sensor drift, reproducibility, and model generalization as continuing challenges in electronic-nose research. They also note that many food applications report high classification accuracy under controlled conditions but differ substantially in sampling protocol, feature construction, preprocessing, and validation design, limiting direct comparability and practical transferability (Rabehi et al., 2024; Sanislav et al., 2025; Singh et al., 2025).

In the present framework, the response of sensor $s$for sample $i$is treated as a continuous nonlinear function:

$$x_{is}(t) = g_s \quad \left(t;\boldsymbol{\theta}_{is}\right) + \varepsilon_{is}(t), t \in [0,T],$$

where $g_s(\cdot)$is a nonlinear response model, $\boldsymbol{\theta}_{is}$is the fitted parameter vector, and $\varepsilon_{is}(t)$represents residual variation. This formulation differs from static feature extraction because the complete temporal trajectory contributes to parameter estimation.

The third gap concerns **the absence of a systematic nonlinear sensor-response signature** for cherry tomato storage-day identification. Previous studies have generally supplied raw sensor values or manually selected summary variables directly to classifiers. However, the parameters obtained from fitted nonlinear curves can provide a compact representation of response amplitude, rate, curvature, asymptotic behavior, and characteristic response times. Such parameters may retain subtle storage-dependent differences that are not visible in endpoint measurements.

For a sensor array containing Ssensors, the proposed signature of sample $i$can be represented as

$$\mathbf{F}_i = \left[\boldsymbol{\theta}_{i1}^{\mathsf{T}}, \boldsymbol{\theta}_{i2}^{\mathsf{T}}, \ldots, \boldsymbol{\theta}_{iS}^{\mathsf{T}}\right]^{\mathsf{T}},$$

where $\boldsymbol{\theta}_{is}$contains the nonlinear parameters derived from sensor s. The complete feature vector therefore represents the combined kinetic response of the sensor array rather than a collection of isolated ADC measurements.

The fourth gap is related to **non-monotonic and sensor-dependent postharvest behavior**. Cherry tomato VOCs do not necessarily increase or decrease uniformly throughout storage. Different biochemical pathways dominate at different stages, and storage temperature modifies both volatile synthesis and degradation. Guan et al. (2025) demonstrated that cherry tomato volatile profiles change dynamically under different thermal conditions, supporting the expectation that sensor responses may also exhibit nonlinear, stage-dependent behavior. Therefore, methods based exclusively on linear trends between sensor magnitude and storage day may be inadequate.

The fifth research gap concerns **validation reliability**. Electronic-nose experiments often contain repeated or related observations originating from the same producer, harvest batch, chamber session, storage group, or sensor-conditioning period. Randomly dividing these observations into training and test sets can place highly related samples in both subsets. As a result, a model may exploit producer-specific baselines or batch characteristics rather than learning reproducible storage-day signatures. High random-split accuracy may therefore overestimate performance on independently obtained samples.

The most convincing assessment of practical performance requires a blind test dataset that remains completely excluded from preprocessing decisions, curve-model selection, feature selection, hyperparameter optimization, and model training. Independent validation has recently received greater attention in electronic-nose food applications. For example, Buratti et al. (2025) evaluated fish-freshness control charts using an external batch acquired separately from the calibration data, illustrating the importance of testing whether electronic-nose decisions transfer beyond the original modeling set.

For the present study, the blind identification rate can be defined as

$$\mathrm{BIA} = \frac{1}{N_{\mathrm{blind}}} \sum_{i=1}^{N_{\mathrm{blind}}} \mathbb{I} \quad \left(\hat{d}_i = d_i\right) \times 100,$$

where $N_{\mathrm{blind}}$is the number of blind samples, $d_i$is the true storage day, $\hat{d}_i$is the predicted storage day, and $\mathbb{I}(\cdot)$is the indicator function. This metric measures exact blind storage-day identification rather than resubstitution or internally cross-validated accuracy.

Because storage days are ordinal, exact accuracy alone is insufficient. A prediction differing by one day is less serious than a prediction differing by five days. Therefore, the magnitude of the prediction error should also be evaluated:

$$\mathrm{MAE} = \frac{1}{N} \sum_{i=1}^{N} \left| d_i - \hat{d}_i \right|,$$

and

$$\mathrm{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(d_i - \hat{d}_i\right)^2}.$$

In addition, within-one-day accuracy can be defined as

$$A_{\pm 1} = \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}\left(\left| d_i - \hat{d}_i \right| \leq 1\right) \times 100.$$

This combination of metrics distinguishes exact identification from practically close predictions and reveals whether misclassifications occur primarily between neighboring days or across distant storage stages.

The sixth gap concerns **model interpretability**. Conventional classifiers may achieve useful prediction performance but provide limited insight into which aspects of the sensor response drive the decision. Raw time points are highly correlated, and the importance assigned to an individual sampling instant may be unstable or biologically difficult to interpret. In contrast, nonlinear parameters describing response amplitude, kinetic rate, curvature, or characteristic response time can be related more directly to the behavior of the complete sensor trajectory. Feature-importance and SHAP analyses applied to such variables can therefore reveal whether storage-day discrimination is primarily driven by signal magnitude, early response dynamics, saturation behavior, or specific sensors.

The seventh gap is **the lack of an electronic-nose-only framework that combines low-cost hardware with a rigorous analytical pipeline**. Recent tomato shelf-life research has demonstrated the value of fusing e-nose signals with computer vision, weight loss, laboratory analyses, and environmental data (Senge et al., 2025). However, multimodal systems may be impractical for applications requiring portability, rapid measurement, minimal calibration, and low cost. The present study therefore investigates whether the full dynamic responses of a compact MOS sensor array contain sufficient information for storage-day identification without requiring imaging or destructive physicochemical analysis.

Based on these gaps, the principal contribution of the present study is the development and evaluation of a continuous nonlinear-response framework for identifying the storage day of intact cherry tomatoes from electronic-nose signals. The specific contributions are as follows.

First, the study treats each sensor trajectory as a continuous response function rather than as a set of independent ADC values. This preserves the temporal organization of volatile–sensor interactions and provides the mathematical basis for comparing complete response curves.

Second, multiple candidate nonlinear functions are evaluated to determine which mathematical model most consistently represents the temporal behavior of each sensor. Model selection is based on complementary goodness-of-fit and parsimony criteria rather than on the coefficient of determination alone.

Third, fitted curve parameters and derived kinetic descriptors are integrated into a multidimensional sensor-response signature. This representation is intended to preserve amplitude, rate, curvature, and stabilization information while avoiding the high collinearity of raw time-series inputs.

Fourth, storage-day discrimination is evaluated as both a multiclass identification problem and an ordinal prediction problem. Exact accuracy is therefore supplemented with MAE, RMSE, and within-one-day accuracy.

Fifth, the influence of individual sensors and nonlinear parameters is examined using feature-importance and explainability analyses. This allows the study to identify which dynamic response characteristics contribute most strongly to storage-day discrimination.

Sixth, the modeling process is designed to prevent information leakage. Preprocessing, nonlinear fitting decisions, feature selection, and hyperparameter optimization are confined to the reference or training data.

Seventh, final reliability is determined using an independent blind dataset that is not involved in model development. This provides a more demanding and practically meaningful test than random within-dataset partitioning.

The overall methodological relationship proposed in this study can be summarized as

$$\left\{x_{is}(t)\right\}_{s=1}^{S} \longrightarrow \left\{\boldsymbol{\theta}_{is}\right\}_{s=1}^{S} \longrightarrow \mathbf{F}_i \longrightarrow \hat{d}_i,$$

where the raw temporal responses are converted into nonlinear parameter vectors, combined into a sample-level sensor-response signature, and subsequently used to identify the storage day.

Accordingly, the novelty of the study does not arise merely from applying another machine-learning algorithm to cherry tomato sensor data. Its principal innovation lies in integrating continuous temporal representation, nonlinear response modelling, kinetic signature construction, ordinal performance assessment, interpretable feature analysis, and independent blind validation within a single electronic-nose framework. By addressing these interconnected limitations, the study seeks to determine whether subtle postharvest changes in cherry tomato volatiles can be converted into reproducible and practically credible storage-day signatures.

2. Materials and Methods

2.1 Cherry Tomatoes Samples and Storage Design

Cherry tomato fruits (*Solanum lycopersicum* L. var. *cerasiforme*) were used to construct a balanced storage-day identification dataset. Postharvest storage duration was selected as the principal experimental factor because changes in respiration, water loss, tissue firmness, soluble solids, organic acids, antioxidant components, and volatile-organic-compound production occur continuously after harvest. Recent studies have confirmed that both storage duration and environmental conditions substantially affect the physicochemical, nutritional, and aromatic properties of tomatoes and cherry tomatoes (Li et al., 2025; Tsouvaltzis et al., 2023; Zhang et al., 2025).

The experimental data were organized into two completely separate datasets: a labelled reference dataset used for model development and an independent blind dataset used exclusively for final storage-day identification. The reference dataset was provided in the file *EB_DOMATES_3_SENSOR_4_27_07_2026.xlsx*, whereas the blind test measurements were provided in *EB_TEST_DOMATES_3_SENSOR_4_27_07_2026.xlsx*. Keeping the blind samples separate from the labelled reference data prevented their storage-day information from influencing response modelling, feature extraction, model selection, or classifier training.

Reference storage dataset

The reference dataset contained three independently coded cherry tomato groups, designated **Cherry1**, **Cherry2**, and **Cherry3**. Each group was evaluated over seven consecutive storage days, from Day 1 to Day 7. The factorial structure of the reference experiment can therefore be expressed as

$$N_{\mathrm{ref}} = C \times D,$$

where $C$is the number of independently coded cherry tomato groups and $D$is the number of storage days. For the present dataset,

$$C = 3$$

and

$$D = 7.$$

Consequently, the reference dataset contained

$$N_{\text{ref}} = 3 \times 7 = 21$$

distinct sensor-response trajectories.

The reference sample identifiers followed a structured coding format. For example, **C1_D1** represented the trajectory obtained from the first cherry tomato group on storage Day 1, whereas **C3_D7** represented the third group on storage Day 7. Thus, every storage day was represented by one trajectory from each of the three independently coded groups. The reference design was fully balanced, with no missing group–day combinations.

The number of response trajectories available for each storage day was

$$n_d = 3, d = 1,2,\ldots,7,$$

and the total number of trajectories associated with each cherry tomato group was

$$n_c = 7, c = 1,2,3.$$

Accordingly, the proportional representation of each storage day in the reference dataset was

$$p_d = \frac{n_d}{N_{\text{ref}}} = \frac{3}{21} = \frac{1}{7} \approx 0.1429.$$

Each storage day therefore constituted approximately 14.29% of the complete reference dataset. This balanced distribution is important because unequal representation of storage days can bias supervised classifiers toward more frequent classes and produce misleading accuracy estimates.

Each sensor-response trajectory was recorded over a 100-s measurement period at 2-s intervals. The sampling times were

$$t_j = 2j, j = 0,1,\ldots,50,$$

corresponding to

$$t_j \in \{0,2,4,\ldots,98,100\} \text{ s}.$$

The number of temporal observations in each trajectory was therefore

$$M = \frac{T}{\Delta t} + 1,$$

where Tdenotes the total measurement duration and Δtdenotes the sampling interval. Substitution of the experimental values gives

$$M = \frac{100}{2} + 1 = 51.$$

Four sensor channels, coded S1, S2, S3, and S4, were recorded at every sampling time. The complete temporal measurement of sample *i*can consequently be expressed as

$$\mathbf{X}_i = \begin{bmatrix} x_{i1}(t_0) & x_{i2}(t_0) & x_{i3}(t_0) & x_{i4}(t_0) \\ x_{i1}(t_1) & x_{i2}(t_1) & x_{i3}(t_1) & x_{i4}(t_1) \\ \vdots & \vdots & \vdots & \vdots \\ x_{i1}(t_{50}) & x_{i2}(t_{50}) & x_{i3}(t_{50}) & x_{i4}(t_{50}) \end{bmatrix},$$

where $x_{is}(t_j)$is the response of sensor *s*for sample *i*at time t_j. Thus, each response trajectory formed a matrix of dimension

$$51 \times 4.$$

The number of recorded sensor values associated with one group–day trajectory was

$$N_{\text{values,curve}} = M \times S,$$

where $S = 4$is the number of sensors. Therefore,

$$N_{\text{values,curve}} = 51 \times 4 = 204.$$

Across the 21 labelled trajectories, the reference spreadsheet contained

$$N_{\text{rows,ref}} = N_{\text{ref}} \times M = 21 \times 51 = 1071$$

time-indexed records, excluding the header row. The total number of individual sensor values in the reference dataset was

$$N_{\text{sensor,ref}} = N_{\text{ref}} \times M \times S = 21 \times 51 \times 4 = 4284.$$

Independent blind test dataset

The independent blind dataset contained 21 additional sensor-response trajectories, coded numerically from Sample 1 to Sample 21. Unlike the reference spreadsheet, the blind-test file did not contain the class or storage-day columns. It included only the sample identifier, measurement time, and responses from S1–S4. The removal of storage-day labels ensured that the samples could be analysed without prior knowledge of their true day assignments.

The number of blind trajectories was

$$N_{\text{blind}} = 21.$$

Each blind sample was also recorded from 0 to 100 s at 2-s intervals and therefore contained 51 time points and four sensor channels. The blind spreadsheet consequently contained

$$N_{\text{rows,blind}} = N_{\text{blind}} \times M = 21 \times 51 = 1071$$

time-indexed records and

$$N_{\text{sensor,blind}} = N_{\text{blind}} \times M \times S = 21 \times 51 \times 4 = 4284$$

individual sensor values.

The reference and blind datasets were thus identical in terms of trajectory number, acquisition duration, sampling interval, number of time points, and sensor dimensionality:

$$N_{\text{ref}} = N_{\text{blind}} = 21,$$
$$M_{\text{ref}} = M_{\text{blind}} = 51,$$

and

$$S_{\text{ref}} = S_{\text{blind}} = 4.$$

The complete analytical dataset comprised

$$N_{\text{total}} = N_{\text{ref}} + N_{\text{blind}} = 21 + 21 = 42$$

sensor-response trajectories and

$$N_{\text{rows,total}} = 1071 + 1071 = 2142$$

time-indexed records.

The reference and blind datasets were not merged during model development. Only the 21 labelled reference trajectories were used to select the nonlinear response model, calculate curve parameters, construct sensor-response signatures, select informative features, and train the storage-day identification algorithm. The 21 blind trajectories were processed only after the complete analytical procedure had been fixed. For blind sample b, storage day was estimated as

$$\hat{d}_b = h \quad \left(\mathbf{F}_b; \widehat{\boldsymbol{\psi}}_{\text{ref}}\right),$$

where $\mathbf{F}_b$is the nonlinear sensor-response signature extracted from the blind curve, $h(\cdot)$denotes the trained identification function, and $\widehat{\boldsymbol{\psi}}_{\text{ref}}$represents all preprocessing, feature-selection, and model parameters determined exclusively from the reference dataset.

This separation provided a more stringent evaluation than randomly dividing individual time points or closely related response records between training and testing subsets. Since all 51 measurements belonging to a trajectory originated from the same sample-level response, they were treated as a single analytical unit rather than as 51 independent observations. Accordingly, the effective reference sample size for storage-day modelling was 21 response curves—not 1071 statistically independent fruit samples. This distinction prevented artificial sample-size inflation and avoided leakage of adjacent time points from the same response curve into both model-development and evaluation sets.

The use of seven consecutive storage days was appropriate for detecting progressive postharvest changes because cherry tomato quality does not change at a constant rate. Storage-dependent differences may occur in firmness, water loss, acidity, antioxidant compounds, decay development, and volatile composition, while the magnitude and direction of these changes can depend on the original fruit characteristics and storage conditions (Li et al., 2025; Tsouvaltzis et al., 2023; Zhao et al., 2023). Recent preservation studies have consequently used repeated measurements over extended storage periods to characterize the temporal evolution of cherry tomato quality rather than relying on only initial and final observations (Zhang et al., 2025; Zhao et al., 2023).

The three-group structure of the reference dataset also allowed the reproducibility of day-related sensor-response patterns to be examined across independently coded tomato groups. A reliable storage-day signature should appear consistently in Cherry1, Cherry2, and Cherry3 rather than being driven solely by one group. Therefore, the group identifier was retained as a blocking and validation variable, whereas storage day was used as the prediction target.

Overall, the storage design produced a balanced labelled reference set and a size-matched independent blind set. The design enabled the study to evaluate whether temporal sensor responses learned from known cherry tomato storage days

could be transferred to unlabelled response trajectories without using information from the blind samples during model construction.

2.2 Closed-Chamber Electronic-Nose System

A compact closed-chamber electronic-nose system was used to record the time-dependent volatile-response patterns of intact cherry tomatoes. The system was designed as a low-cost and nondestructive alternative to conventional chromatographic analysis and consisted of an airtight sample chamber, a metal-oxide-semiconductor sensor array, an embedded data-acquisition unit, and a computer-based recording interface. In contrast to dynamic-flow electronic noses employing pumps, carrier gases, inlet lines, and exhaust units, volatile transport within the present system occurred through passive headspace accumulation and diffusion.

Closed-headspace measurement was selected because cherry tomatoes continuously release a complex mixture of volatile organic compounds during ripening and postharvest storage. Confined accumulation increases the quantity of volatiles reaching the sensing surfaces and makes it possible to record the complete transient response of the sensor array. Low-cost MOS-based electronic noses have been widely investigated for fruit ripeness and food-quality assessment because they can detect composite VOC patterns without requiring compound-by-compound separation (Sanislav et al., 2025; Tyagi et al., 2023; Zhai et al., 2024). Electronic-nose measurements have also been successfully associated with tomato shelf life, freshness, firmness, soluble solids, and postharvest quality (Feng et al., 2018; Senge et al., 2025).

Chamber configuration and passive headspace formation

The electronic nose employed a small, sealed measurement chamber fitted with a removable lid and an airtight sealing element. Cherry tomato samples were positioned inside the chamber without direct contact with the gas sensors. The sensors were arranged within the enclosed headspace so that the volatiles released by the fruit could diffuse toward the sensing surfaces.

The chamber did not contain an active air-intake pump or an air-discharge pump during the response-acquisition period. Therefore, the temporal sensor signals were generated by three interacting processes:

1. release of volatile compounds from the cherry tomato surface;
2. accumulation and mixing of these compounds within the confined headspace; and
3. passive diffusion and adsorption of the volatile mixture onto the MOS sensor surfaces.

The headspace concentration of a volatile component may be represented conceptually by a closed-system mass balance:

$$V_{\mathrm{h}} \frac{dC_k(t)}{dt} = E_k(t) - A_k(t) - L_k(t),$$

where $V_{\mathrm{h}}$is the effective headspace volume, $C_k(t)$is the chamber concentration of volatile component k, $E_k(t)$is its emission rate from the fruit, $A_k(t)$represents removal through adsorption or reaction at chamber and sensor surfaces, and $L_k(t)$represents any residual leakage or loss.

Under an adequately sealed chamber condition, the leakage term is expected to be relatively small:

$$L_k(t) \approx 0.$$

The resulting relationship can therefore be simplified as

$$V_{\mathrm{h}} \frac{dC_k(t)}{dt} \approx E_k(t) - A_k(t).$$

This formulation explains why the sensor responses increased progressively during the measurement period rather than reaching an instantaneous value. The recorded curves represented the combined effect of volatile release, headspace accumulation, diffusion, surface adsorption, and sensor reaction kinetics.

Passive diffusion was preferred in the present system because pump-driven flow can introduce additional sources of variation, including flow-rate instability, pressure fluctuations, tubing adsorption, carryover, and differences in sample delivery. The closed-chamber architecture also reduced hardware complexity and supported the intended development of a compact and accessible fruit-quality monitoring device. However, because chamber volume, sample quantity, sealing efficiency, temperature, and headspace-development period can affect sensor response magnitude, these variables must remain consistent across reference and blind measurements.

MOS sensor array

The uploaded reference and blind-test spreadsheets showed that the electronic nose contained four simultaneously recorded sensing channels, designated S1, S2, S3, and S4. These channels represented four MOS sensor responses to the composite cherry tomato headspace rather than measurements of four individually isolated volatile compounds.

MOS sensors are inherently cross-sensitive. Each sensor may respond to several reducing or oxidizing gases, but its response magnitude and temporal behavior differ according to the composition of the volatile mixture. The electronic-nose fingerprint is therefore produced by the collective response of the array:

$$\mathbf{x}_i(t) = \begin{bmatrix} x_{i1}(t) \\ x_{i2}(t) \\ x_{i3}(t) \\ x_{i4}(t) \end{bmatrix},$$

where $x_{is}(t)$denotes the recorded response of sensor $s$for sample $i$at time t.

The complete raw measurement matrix for one cherry tomato response was

$$\mathbf{X}_i = \left[\mathbf{x}_i(t_0), \mathbf{x}_i(t_1), \ldots, \mathbf{x}_i(t_{50})\right]^{\mathsf{T}},$$

with dimensions

$$\mathbf{X}_i \in \mathbb{R}^{51\times 4}.$$

Thus, every sample was represented by 51 temporal observations from four sensors, corresponding to

$$N_{\text{values/sample}} = 51 \times 4 = 204$$

raw sensor values per response trajectory.

The reference dataset contained 21 labelled trajectories, and the blind dataset contained 21 additional unlabelled trajectories. Therefore, the system produced

$$N_{\text{values,ref}} = 21 \times 51 \times 4 = 4284$$

sensor values in the reference dataset and

$$N_{\text{values,blind}} = 21 \times 51 \times 4 = 4284$$

sensor values in the independent blind dataset.

The total number of sensor observations evaluated in the two spreadsheets was consequently

$$N_{\text{values,total}} = 4284 + 4284 = 8568.$$

Response-acquisition protocol

The spreadsheet structure confirmed that all reference and blind measurements were acquired over a common interval of 0–100 s. Sensor outputs were recorded every 2 s, producing the time sequence

$$t_j = 2j, j = 0,1,\ldots,50.$$

The total number of time points was therefore

$$M = \frac{T}{\Delta t} + 1 = \frac{100}{2} + 1 = 51,$$

where $T = 100$ swas the measurement duration and $\Delta t = 2$ swas the sampling interval.

The effective sampling frequency was

$$f_{\mathrm{s}} = \frac{1}{\Delta t} = \frac{1}{2} = 0.5 \text{ Hz}.$$

This acquisition rate was sufficient to characterize the relatively slow accumulation and response process occurring in the passive closed chamber. The use of identical acquisition durations and sampling intervals in the reference and blind datasets was essential because curve parameters and temporal features can be biased when signals are recorded over different time windows or temporal resolutions.

The four sensor values were recorded synchronously at each measurement time. Accordingly, the raw observation at time $t_j$can be expressed as

$$\mathbf{x}_i(t_j) = \left[x_{i1}(t_j), x_{i2}(t_j), x_{i3}(t_j), x_{i4}(t_j)\right].$$

The full temporal sequence was retained rather than replacing each sensor response with a single endpoint value. This was necessary because storage-related information may occur in the initial increase, response rate, curvature, and stabilization behavior as well as in the final signal magnitude.

Baseline condition

The first recorded sensor values at $t = 0$were close to zero in both datasets. In the labelled reference data, the initial values of S1–S4 ranged from 0 to 1 digital unit. In the blind dataset, the corresponding baseline values ranged from 0 to 2 units. These values were small relative to the subsequent response amplitudes and indicated that the recorded trajectories had already been expressed on an approximately zero-referenced response scale.

For sensor s, a conventional baseline-corrected response can be defined as

$$\Delta x_{is}(t) = x_{is}(t) - x_{is}(0).$$

A relative response may alternatively be expressed as

$$r_{is}(t) = \frac{x_{is}(t) - x_{is}(0)}{x_{is}(0) + \epsilon},$$

where $\epsilon$is a small positive constant used to prevent division by zero. Because the recorded initial values were zero or nearly zero, absolute baseline correction was more appropriate than direct ratio normalization unless a stabilizing constant was introduced.

The maximum recorded values in the reference dataset were 835, 834, 775, and 775 digital units for S1, S2, S3, and S4, respectively. The corresponding maxima in the blind dataset were 836, 835, 776, and 776 units. The near-identical ranges between the two datasets indicated that the same sensor configuration, acquisition scale, and recording protocol were used for reference and blind measurements.

The observed reference-data ranges were

$$0 \le S1 \le 835,$$
$$0 \le S2 \le 834,$$
$$0 \le S3 \le 775,$$

and

$$0 \le S4 \le 775.$$

For the blind dataset, the ranges were

$$0 \le S1 \le 836,$$
$$0 \le S2 \le 835,$$
$$0 \le S3 \le 776,$$

and

$$0 \le S4 \le 776.$$

The agreement of these measurement ranges is important for blind identification. A substantial scale difference between the reference and test datasets could indicate sensor recalibration, gain variation, drift, or inconsistent preprocessing and might invalidate direct comparison of their response signatures.

Signal formation in the closed chamber

The MOS response results from interactions between the accumulated headspace volatiles and the sensor surface. In general terms, the electrical response of sensor *s*can be written as

$$x_s(t) = \Phi_s\left(C_1(t), C_2(t), \ldots, C_K(t), T_c, H_c\right) + \varepsilon_s(t),$$

where $C_1(t),\ldots,C_K(t)$denote the time-dependent concentrations of the volatile constituents, $T_c$is chamber temperature, $H_c$is chamber relative humidity, $\Phi_s(\cdot)$is the nonlinear sensor-response function, and $\varepsilon_s(t)$represents electrical and environmental variation.

Because MOS sensors are cross-sensitive, the measured response cannot normally be attributed to one compound:

$$x_s(t) \ne \phi_s \quad \left(C_k(t)\right).$$

Instead, the signal reflects a weighted nonlinear interaction with the complete volatile mixture:

$$x_s(t) = \Phi_s\left(\mathbf{C}(t)\right),$$

where

$$\mathbf{C}(t) = \left[C_1(t), C_2(t), \ldots, C_K(t)\right]^{\mathsf{T}}.$$

This collective sensitivity is advantageous for storage-day identification because postharvest deterioration alters multiple volatile pathways simultaneously. The objective was not to identify or quantify individual VOCs but to determine whether their combined effect produced reproducible temporal sensor-response signatures.

Reference and blind measurements

The same closed-chamber system and the same four-channel acquisition structure were used for both datasets. The reference file included sample group and storage-day labels, whereas the blind file contained only sample identifiers, time values, and S1–S4 responses. Thus, the hardware output provided to the storage-day model had an identical format in the two experimental phases:

$$\mathbf{X}_{\mathrm{ref},i} \in \mathbb{R}^{51\times 4}$$

and

$$\mathbf{X}_{\mathrm{blind},b} \in \mathbb{R}^{51\times 4}.$$

All 51 observations belonging to one trajectory were treated as a single sensor measurement. They were not considered independent samples because the points were successive states of the same headspace-accumulation process.

Maintaining the same chamber, sensor array, sample placement, acquisition time, and sampling interval was especially important for blind comparison. The measured response can be affected by chamber geometry, fruit-to-headspace ratio, sensor distance, chamber sealing, humidity, and residual compounds from the preceding measurement. Recent reviews similarly identify sampling standardization, environmental compensation, sensor recovery, repeatability, and drift control as major determinants of electronic-nose reliability (Sanislav et al., 2025; Zhai et al., 2024).

Accordingly, the chamber was required to be ventilated or cleaned between consecutive measurements until the sensor signals returned to their initial condition. The recovery criterion for sensor *s*may be defined as

$$| x_s(t_{\mathrm{recovery}}) - x_s(0) | \leq \delta_s,$$

where $\delta_s$is the permitted baseline deviation. Applying a consistent recovery condition reduces carryover of volatiles from one cherry tomato sample to the next.

Relevance of the system to storage-day identification

The principal advantage of the developed closed-chamber architecture was that it generated slowly evolving response curves suitable for continuous nonlinear analysis. In a passive system, the shape of the curve contains information about both headspace formation and sensor kinetics. Therefore, the recorded signal was not interpreted solely through the final value at 100 s.

For each sensor, the total response amplitude was defined as

$$\Delta x_{is} = x_{is}(100) - x_{is}(0).$$

However, the full signal was represented more generally as

$$x_{is}(t) = g_s\left(t;\boldsymbol{\theta}_{is}\right) + \varepsilon_{is}(t),$$

where $\boldsymbol{\theta}_{is}$contains parameters describing the nonlinear temporal behavior. This enabled the closed-chamber system to provide not only response magnitude but also kinetic characteristics such as rise rate, curvature, half-response time, and asymptotic behavior.

The experimental configuration therefore supported the central objective of the study: converting the volatile headspace of stored cherry tomatoes into reproducible four-sensor temporal signatures and determining whether these signatures could identify unknown storage days. The use of a sealed passive chamber, identical 100-s acquisition cycles, four synchronized sensor channels, and a completely separate blind dataset provided a consistent hardware and data-acquisition basis for subsequent nonlinear response modelling.

2.3 Temporal Gas Sensor-Response Acquisition

he electronic-nose measurements were acquired as synchronized multichannel time series to preserve the complete transient behavior of the gas sensors during exposure to the cherry tomato headspace. This acquisition strategy was adopted because metal-oxide-semiconductor sensors do not produce an instantaneous equilibrium output. Their signals evolve progressively as volatile compounds accumulate within the measurement chamber, diffuse toward the sensing surfaces, and modify the electrical properties of the sensing materials. Consequently, the temporal evolution of the response can contain information that is not represented adequately by a single maximum, mean, or endpoint value. Recent electronic-nose research has similarly emphasized the importance of recording voltage–time or resistance–time profiles and retaining response amplitude, shape, slope, and other dynamic characteristics for subsequent pattern recognition (Shtepliuk et al., 2025; Vajdi, 2025; Zhai et al., 2024).

2.3.1. Structure of the acquired sensor signals

The reference spreadsheet contained four synchronized gas-sensor channels, designated S1, S2, S3, and S4. For sample i, the multichannel response recorded at time $t$was represented as

$$\mathbf{x}_i(t) = \begin{bmatrix} x_{i1}(t) \\ x_{i2}(t) \\ x_{i3}(t) \\ x_{i4}(t) \end{bmatrix},$$

where $x_{is}(t)$denotes the digital response of sensor $s$for sample $i$at acquisition time t.

All four sensor channels were recorded at the same time points. Therefore, the observation vector at the jth acquisition instant was

$$\mathbf{x}_{i,j} = \left[x_{i1}(t_j), x_{i2}(t_j), x_{i3}(t_j), x_{i4}(t_j) \right].$$

The complete response trajectory for one sample was organized as

$$\mathbf{X}_i = \begin{bmatrix} x_{i1}(t_0) & x_{i2}(t_0) & x_{i3}(t_0) & x_{i4}(t_0) \\ x_{i1}(t_1) & x_{i2}(t_1) & x_{i3}(t_1) & x_{i4}(t_1) \\ \vdots & \vdots & \vdots & \vdots \\ x_{i1}(t_{50}) & x_{i2}(t_{50}) & x_{i3}(t_{50}) & x_{i4}(t_{50}) \end{bmatrix}.$$

Thus, every cherry tomato measurement constituted a single multivariate temporal response rather than 51 statistically independent observations.

2.3.2. Acquisition duration and temporal resolution

Both the labelled reference dataset and the independent blind dataset were recorded over a common acquisition period extending from 0 to 100 s. The sensor outputs were sampled at uniform 2-s intervals. The acquisition times were therefore defined as

$$t_j = j\Delta t, j = 0,1,\ldots,50,$$

where

$$\Delta t = 2 \text{ s}.$$

Accordingly,

$$t_j \in \{0,2,4,\ldots,98,100\} \text{ s.}$$

The number of acquisition points in each sensor trajectory was calculated as

$$M = \frac{T}{\Delta t} + 1,$$

where $T$is the total response-acquisition duration. Substitution of the experimental values produced

$$M = \frac{100}{2} + 1 = 51.$$

The corresponding sampling frequency was

$$f_s = \frac{1}{\Delta t} = 0.5 \text{ Hz.}$$

This sampling frequency generated one synchronized four-sensor observation every 2 s. Since the response curves changed gradually over the 100-s period, the selected temporal resolution captured the progressive increase and curvature of the sensor trajectories without producing an unnecessarily large number of nearly redundant adjacent measurements.

Recent food-quality electronic-nose studies commonly process complete sensor time series because variation may occur in response magnitude, signal shape, peak timing, slope, and stabilization behavior. Shtepliuk et al. (2025), for example, evaluated the temporal evolution of MOS sensor voltages and extracted multiple descriptors from the recorded voltage–time curves. Similarly, current reviews identify temporal signal preprocessing and dynamic feature preservation as important components of MOS electronic-nose analysis (Vajdi, 2025; Zhai et al., 2024).

2.3.3. Reference-data acquisition structure

The reference dataset comprised 21 labelled response trajectories obtained from three independently coded cherry tomato groups over seven storage days. Each sample trajectory contained 51 time points and four synchronized sensor values. The number of time-indexed records in the reference dataset was therefore

$$N_{\text{rows,ref}} = N_{\text{samples,ref}} \times M,$$

and hence

$$N_{\text{rows,ref}} = 21 \times 51 = 1071.$$

The total number of individual gas-sensor readings was

$$N_{\text{readings,ref}} = N_{\text{samples,ref}} \times M \times S,$$

where $S = 4$was the number of sensor channels. Therefore,

$$N_{\text{readings,ref}} = 21 \times 51 \times 4 = 4284.$$

Each storage day was represented by three complete temporal trajectories, one from each coded cherry tomato group. Therefore,

$$n_d = 3, d = 1,2,\ldots,7.$$

The spreadsheet inspection confirmed that all 21 reference samples contained exactly 51 records, began at 0 s, ended at 100 s, and followed the same 2-s temporal interval. No missing values, duplicated complete rows, or duplicated sample–time combinations were identified. The temporal grid was therefore complete and uniform for all labelled samples.

For every reference sample,

$$\min(t) = 0 \text{ s},$$
$$\max(t) = 100 \text{ s},$$

and

$$t_{j+1} - t_j = 2 \text{ s}.$$

This uniformity was important for nonlinear curve fitting because unequal temporal grids or missing response segments could alter the estimated model parameters and compromise comparisons between storage days.

2.3.4. Independent blind-data acquisition structure

The blind dataset contained 21 additional response trajectories. Each blind sample was represented by the same four sensors, 51 time points, 100-s acquisition duration, and 2-s sampling interval used for the reference dataset. The blind spreadsheet therefore also contained

$$N_{\text{rows,blind}} = 21 \times 51 = 1071$$

time-indexed records and

$$N_{\text{readings,blind}} = 21 \times 51 \times 4 = 4284$$

individual sensor values.

As in the reference dataset, all blind samples contained a complete and identical temporal grid:

$$t_j = 2j, j = 0,1,\ldots,50.$$

No missing sensor values, duplicate rows, duplicated sample–time pairs, or incomplete trajectories were detected in the blind dataset. This structural equivalence ensured that the blind curves could be processed using exactly the same nonlinear-response and feature-extraction procedures established from the reference data.

The dimensions of the two acquisition sets were identical:

$$\mathbf{X}_i^{\text{ref}} \in \mathbb{R}^{51\times 4},$$

and

$$\mathbf{X}_b^{\text{blind}} \in \mathbb{R}^{51\times 4}.$$

The combined database consequently contained

$$N_{\text{samples,total}} = 21 + 21 = 42$$

complete response trajectories,

$$N_{\text{rows,total}} = 1071 + 1071 = 2142$$

time-indexed records, and

$$N_{\text{readings,total}} = 4284 + 4284 = 8568$$

individual sensor readings.

2.3.5. Baseline response at the beginning of acquisition

The recorded values at $t = 0$were close to zero for all sensors. In the reference dataset, the initial sensor outputs ranged from 0 to 1 digital unit. Their mean baseline values were 0.286, 0.333, 0.238, and 0.381 units for S1, S2, S3, and S4, respectively. In the blind dataset, the baseline values ranged from 0 to 2 units, with corresponding means of 1.048, 1.095, 1.000, and 1.143 units.

The baseline-corrected response was defined as

$$\Delta x_{is}(t) = x_{is}(t) - x_{is}(0).$$

This transformation preserves the absolute response increase while removing small differences in the initial digital output. Baseline subtraction is particularly appropriate when initial values are close to zero, because ratio-based normalization can become unstable when the denominator is zero or very small. Contemporary reviews of MOS electronic-nose preprocessing similarly identify baseline manipulation, normalization, smoothing, and drift correction as central procedures for improving the comparability of sensor data (Vajdi, 2025).

A normalized response could theoretically be expressed as

$$r_{is}(t) = \frac{x_{is}(t) - x_{is}(0)}{x_{is}(0)},$$

but this expression is undefined when

$$x_{is}(0) = 0.$$

Accordingly, direct ratio normalization was not assumed from the spreadsheet values. Where scale normalization was required in subsequent modelling, its parameters were to be estimated exclusively from the reference dataset and then applied unchanged to the blind data.

2.3.6. Dynamic range of the acquired responses

The reference sensor values covered the following ranges:

$$0 \le S1 \le 835,$$
$$0 \le S2 \le 834,$$
$$0 \le S3 \le 775,$$

and

$$0 \le S4 \le 775.$$

The corresponding blind-data ranges were

$$0 \le S1 \le 836,$$
$$0 \le S2 \le 835,$$

$$0 \leq S3 \leq 776,$$

and

$$0 \leq S4 \leq 776.$$

The maximum values in the blind dataset exceeded the reference maxima by only one digital unit for each sensor. This close agreement indicates that both datasets were recorded on the same numerical scale and within nearly identical response limits.

The final responses at 100 s varied substantially among the reference samples. Their mean values were

$$\overline{x_{S1}}(100) = 504.952,$$

$$\overline{x_{S2}}(100) = 497.571,$$

$$\overline{x_{S3}}(100) = 482.286,$$

and

$$\overline{x_{S4}}(100) = 494.000.$$

The corresponding standard deviations were 195.762, 191.101, 190.992, and 184.744 digital units, respectively. These relatively broad final-response distributions demonstrate that the dataset contained pronounced between-sample variability, which was expected to include contributions from storage day and independently coded cherry tomato groups.

For sensor s, the absolute response amplitude was calculated as

$$A_{is} = x_{is}(100) - x_{is}(0).$$

However, the endpoint amplitude alone was not treated as a complete representation of the measurement. Two trajectories may have similar values at 100 s while differing in their initial rate, intermediate curvature, or approach to the final response. The complete temporal acquisition was therefore retained for nonlinear modelling.

2.3.7. Temporal rate and curvature information

The discrete response increment between two consecutive acquisition points was defined as

$$\Delta x_{is,j} = x_{is}(t_{j+1}) - x_{is}(t_j).$$

The corresponding finite-difference response rate was

$$v_{is,j} = \frac{x_{is}(t_{j+1}) - x_{is}(t_j)}{t_{j+1} - t_j}.$$

Because the interval was constant,

$$t_{j+1} - t_j = 2 \text{ s},$$

the response rate became

$$v_{is,j} = \frac{x_{is}(t_{j+1}) - x_{is}(t_j)}{2}.$$

A discrete estimate of curvature could be expressed as

$$a_{is,j} = \frac{x_{is}(t_{j+1}) - 2x_{is}(t_j) + x_{is}(t_{j-1})}{\Delta t^2}.$$

With the 2-s acquisition interval,

$$a_{is,j} = \frac{x_{is}(t_{j+1}) - 2x_{is}(t_j) + x_{is}(t_{j-1})}{4}.$$

These finite-difference quantities describe local signal behavior but can be affected by digital quantization and short-scale measurement variation. For this reason, the study subsequently represented the trajectories using continuous nonlinear functions and calculated kinetic properties from the fitted curves rather than relying exclusively on direct point-to-point differences.

The spreadsheet inspection showed that the response trajectories were almost entirely increasing over time. Across the 21 reference curves and 50 consecutive intervals per curve, no negative increments occurred for S2, S3, or S4. Only one negative consecutive increment was identified for S1. The same pattern was observed in the blind data. This finding supports the interpretation that the acquired signals predominantly represented progressive headspace accumulation and sensor activation during the 100-s closed-chamber exposure.

The general response evolution may therefore be written as

$$\frac{dx_{is}(t)}{dt} > 0$$

for most of the acquisition interval, while the declining rate of increase observed in nonlinear response curves may be represented as

$$\frac{d^2x_{is}(t)}{dt^2} < 0$$

during the approach toward a plateau-like region.

2.3.8. Treatment of trajectories as sample-level experimental units

Although each spreadsheet contained 1071 rows, these rows did not constitute 1071 independent cherry tomato samples. The 51 observations belonging to one sample were serial measurements from the same sensor exposure and were therefore temporally dependent.

For a given sensor,

$$\mathrm{Cov}\left[x_{is}(t_j), x_{is}(t_{j+1})\right] \neq 0.$$

Accordingly, the effective number of experimental units was

$$N_{\mathrm{ref}} = 21$$

for the reference dataset and

$$N_{\text{blind}} = 21$$

for the independent test dataset.

The temporal observations were used jointly to estimate one response function for each sensor and sample. The fitted signal was represented as

$$x_{is}(t) = g_s\left(t;\boldsymbol{\theta}_{is}\right) + \varepsilon_{is}(t),$$

where $g_s(\cdot)$denotes the nonlinear response model, $\boldsymbol{\theta}_{is}$is its sample-specific parameter vector, and $\varepsilon_{is}(t)$is the residual measurement component.

This treatment prevented artificial enlargement of the sample size and avoided the methodological error of distributing neighboring points from the same trajectory across training and test partitions. Current electronic-nose studies increasingly treat the recorded voltage–time signal as the source from which sample-level features are extracted before model development (Shtepliuk et al., 2025).

2.3.9. Reference-to-blind acquisition compatibility

Compatibility between the reference and blind datasets was assessed before nonlinear modelling. The following conditions were verified:

$$T_{\text{ref}} = T_{\text{blind}} = 100 \text{ s},$$
$$\Delta t_{\text{ref}} = \Delta t_{\text{blind}} = 2 \text{ s},$$
$$M_{\text{ref}} = M_{\text{blind}} = 51,$$

and

$$S_{\text{ref}} = S_{\text{blind}} = 4.$$

The maximum absolute difference between the upper response limits of the two datasets was

$$\Delta_{\max,s} = |\max\left(x_s^{\text{blind}}\right) - \max\left(x_s^{\text{ref}}\right)| = 1$$

digital unit for each of the four sensors.

This close numerical correspondence indicated that the blind measurements did not require temporal interpolation or sensor-range conversion before being processed using the reference-derived analytical pipeline. It also reduced the risk that storage-day predictions would be influenced by obvious acquisition-scale differences rather than by the shape and magnitude of the cherry tomato response trajectories.

2.3.10. Data-integrity control

Before feature extraction, each trajectory was checked for temporal completeness, sensor completeness, and uniqueness of sample–time records. A valid trajectory was required to satisfy

$$N_i = 51,$$
$$t_{i,0} = 0,$$
$$t_{i,50} = 100,$$

and

$$t_{i,j+1} - t_{i,j} = 2 \text{ s}.$$

For every sensor channel, the number of missing values was required to satisfy

$$N_{\text{missing},s} = 0.$$

The number of duplicated sample–time records was also required to satisfy

$$N_{\text{duplicate}} = 0.$$

All reference and blind trajectories met these criteria. Therefore, no temporal interpolation, deletion of incomplete samples, or replacement of missing sensor values was required before nonlinear response modelling.

Overall, the acquisition protocol generated balanced, synchronized, and structurally identical temporal-response datasets. The 100-s measurement duration and 2-s sampling interval yielded 51-point trajectories capable of representing the early response, progressive rise, curvature, and late-stage behavior of all four sensors. Retention of the complete temporal signals provided the basis for the nonlinear response modelling and sensor-signature extraction procedures used in the subsequent sections.

2.4 Signal Preprocessing

Signal preprocessing was performed to improve the comparability of the temporal sensor responses while preserving the nonlinear shape information required for subsequent curve modelling. The preprocessing workflow was deliberately conservative. Because the objective of the study was to derive storage-day signatures from response amplitude, rate, curvature, and stabilization behaviour, transformations capable of altering these characteristics were avoided unless they were supported by the observed data quality.

Recent electronic-nose studies have shown that preprocessing can substantially affect model complexity, prediction accuracy, and generalizability. Baseline correction is particularly important for removing differences in the initial sensor state, whereas smoothing and normalization should be selected according to the noise structure, sensor scale, and analytical objective rather than applied automatically. Excessive preprocessing may remove chemically relevant variation or introduce artificial response patterns (Fernandez et al., 2025; Vajdi, 2025).

The preprocessing procedure comprised five stages: structural integrity control, trajectory ordering, baseline correction, signal-quality assessment, and reference-derived scaling of the nonlinear features. All preprocessing decisions were established using the labelled reference dataset. The same operations and numerical parameters were subsequently applied to the independent blind dataset without modification.

2.4.1. Structural integrity and temporal-grid verification

The labelled reference dataset and independent blind dataset each contained 21 complete sensor-response trajectories. Every trajectory comprised 51 synchronized observations from four sensor channels recorded between 0 and 100 s at 2-s intervals. Before any numerical transformation, the data were grouped by sample identifier and ordered according to acquisition time.

For sample i, a valid temporal sequence was required to satisfy

$$t_{i,j+1} > t_{i,j}, j = 0,1,\ldots,49,$$

and

$$t_{i,j+1} - t_{i,j} = 2 \text{ s}.$$

The number of observations in each sample was checked using

$$M_i = \sum_{j=0}^{50} 1.$$

All reference and blind samples satisfied

$$M_i = 51.$$

The first and final acquisition times were also verified:

$$t_{i,0} = 0 \text{ s}$$

and

$$t_{i,50} = 100 \text{ s}.$$

Neither dataset contained missing sensor values. Accordingly,

$$N_{\text{missing},s} = 0, s = 1,2,3,4.$$

No duplicated complete records or duplicated sample–time combinations were detected:

$$N_{\text{duplicate rows}} = 0$$

and

$$N_{\text{duplicate sample–time}} = 0.$$

The reference dataset therefore contained 1,071 valid time-indexed records and 4,284 individual sensor readings. The blind dataset contained the same numbers. Since all trajectories shared an identical temporal grid, no interpolation, temporal resampling, padding, or deletion of incomplete curves was required.

2.4.2. Sample-wise baseline correction

Small differences were observed in the sensor values recorded at the beginning of acquisition. In the reference dataset, the initial responses ranged from 0 to 1 digital unit. Their mean baseline values were 0.286, 0.333, 0.238, and 0.381 units for S1, S2, S3, and S4, respectively. In the blind dataset, initial responses ranged from 0 to 2 units, with corresponding means of 1.048, 1.095, 1.000, and 1.143 units.

Although these offsets were small relative to the total sensor-response ranges, retaining them could cause the nonlinear model to represent differences in the starting electronic state rather than differences caused by cherry tomato headspace volatiles. Therefore, baseline correction was performed separately for every sample and sensor.

The baseline-corrected response was calculated as

$$x_{is}^{\text{bc}}(t_j) = x_{is}(t_j) - x_{is}(t_0),$$

where $x_{is}(t_j)$is the original response of sensor $s$for sample $i$at time t_j, and $x_{is}(t_0)$is the corresponding value at the beginning of acquisition.

After correction,

$$x_{is}^{bc}(0) = 0$$

for every sample and sensor.

This sample-wise operation preserved the absolute response increase while removing initial channel offsets. Baseline correction is one of the most influential preprocessing stages in electronic-nose analysis because it reduces differences arising from the initial sensor state and improves comparability among measurement cycles (Fernandez et al., 2025; Vajdi, 2025).

Ratio-based baseline normalization was not used. A conventional fractional response could be expressed as

$$r_{is}(t_j) = \frac{x_{is}(t_j) - x_{is}(t_0)}{x_{is}(t_0)},$$

but this transformation becomes undefined when

$$x_{is}(t_0) = 0.$$

Because several trajectories had zero-valued initial observations, ratio normalization would have required an arbitrary stabilizing constant and could have produced disproportionately large values. Absolute baseline subtraction was therefore more appropriate for the present data.

2.4.3. Assessment of response continuity and local signal variation

The temporal continuity of each sensor response was evaluated using consecutive first differences:

$$\Delta x_{is,j} = x_{is}(t_{j+1}) - x_{is}(t_j).$$

Across the 21 reference trajectories, each sensor generated

$$21 \times 50 = 1050$$

consecutive increments. Sensors S2, S3, and S4 contained no negative increments. S1 contained only one negative increment, with a magnitude of 5 digital units. The same pattern was found in the blind dataset. Thus, 4,199 of the 4,200 consecutive increments in each dataset were non-negative.

For the reference dataset, the median 2-s increments were 9, 9, 9, and 10 digital units for S1, S2, S3, and S4, respectively. The 95th-percentile increments were approximately 18.6, 18, 18, and 18 units. These results indicated progressive, smoothly increasing responses rather than oscillatory or high-frequency signals.

Local curvature was examined using the second finite difference:

$$\Delta^2 x_{is,j} = x_{is}(t_{j+1}) - 2x_{is}(t_j) + x_{is}(t_{j-1}).$$

The median absolute second difference was zero for all four sensors, while the 95th-percentile values ranged from 3.0 to 3.6 digital units. Except for the isolated S1 change, the short-term variation was small relative to the complete response ranges, which extended to approximately 775–835 digital units.

These findings indicated that the signals already exhibited a highly regular nonlinear structure. Consequently, the isolated negative S1 increment was retained rather than replaced or deleted. Removing a single observed point without independent evidence of an acquisition fault could artificially enforce monotonicity and bias the fitted response parameters.

2.4.4. Decision not to apply numerical smoothing

Savitzky–Golay filters, moving averages, low-pass filters, wavelet denoising, and related procedures are frequently used to reduce high-frequency noise in electronic-nose signals. Savitzky–Golay smoothing can be advantageous because a local polynomial fit generally preserves broader signal shapes better than a simple moving average. Nevertheless, its effect depends on the selected window length and polynomial order, and inappropriate settings may modify local slopes, curvature, and characteristic response times (Fernandez et al., 2025; Vajdi, 2025).

No numerical smoothing was applied to the present trajectories before nonlinear fitting. This decision was supported by four observations:

1. all trajectories were complete and uniformly sampled;
2. the sensor curves were almost entirely monotonic;
3. median absolute second differences were negligible relative to total response amplitudes; and
4. the subsequent nonlinear curve-fitting process itself provided a continuous approximation of the underlying response.

Applying a separate smoothing filter before nonlinear modelling would have introduced an additional set of parameters, such as filter-window length and polynomial order, which could influence the response rate and curvature. Instead, the measured baseline-corrected signal was modelled directly as

$$x_{is}^{\mathrm{bc}}(t) = g_s \quad \left(t;\boldsymbol{\theta}_{is}\right) + \varepsilon_{is}(t),$$

where $g_s(\cdot)$denotes the fitted nonlinear response function, $\boldsymbol{\theta}_{is}$contains the estimated curve parameters, and $\varepsilon_{is}(t)$represents residual measurement variation.

The fitted nonlinear function therefore served as a model-based smooth representation while retaining a direct and transparent relationship with the measured observations.

2.4.5. Outlier assessment

Potential point-level anomalies were evaluated using the consecutive increments and second differences rather than by applying a global univariate outlier rule to the raw sensor values. Large readings near the end of acquisition were not automatically considered outliers because increasing response magnitude was an expected characteristic of the closed-chamber exposure.

For each sensor, a robust standardized first difference may be defined as

$$z_{is,j}^{\mathrm{rob}} = \frac{\Delta x_{is,j} - \mathrm{median}\ (\Delta x_s)}{1.4826\,\mathrm{MAD}\ (\Delta x_s)},$$

where

$$\mathrm{MAD}\ (\Delta x_s) = \mathrm{median}\ \left(\mid \Delta x_{is,j} - \mathrm{median}\ (\Delta x_s) \mid\right).$$

However, no observations were automatically removed solely because they exceeded a statistical threshold. A point was to be considered invalid only when it was simultaneously inconsistent with adjacent observations, incompatible with the physical response pattern, and supported by evidence of an acquisition or recording error.

The inspection identified no missing segments, impossible sensor values, duplicated time points, or isolated extreme spikes requiring replacement. Consequently,

$$N_{\mathrm{removed}} = 0$$

and

$$N_{\text{imputed}} = 0$$

for both the reference and blind datasets.

This conservative procedure prevented biologically meaningful between-day or between-group differences from being mistakenly eliminated as statistical outliers.

2.4.6. Preservation of the original sensor scales

The four sensor channels had comparable but non-identical response ranges. In the reference dataset, maximum values were 835, 834, 775, and 775 digital units for S1–S4, respectively. The corresponding blind-data maxima were 836, 835, 776, and 776 units. Thus, the maximum reference–blind difference was only one digital unit for each channel.

Because the sensor ranges were already similar, min–max normalization of every raw trajectory was not applied before nonlinear fitting. Sample-wise min–max normalization would have forced each curve to span the same interval and removed amplitude differences that could carry storage-day information.

For example, sample-wise normalization would produce

$$x_{is}^{\text{mm}}(t) = \frac{x_{is}(t) - \min\limits_{t} x_{is}(t)}{\max\limits_{t} x_{is}(t) - \min\limits_{t} x_{is}(t)}.$$

This transformation maps every trajectory onto

$$0 \le x_{is}^{\text{mm}}(t) \le 1,$$

irrespective of its original response magnitude. Since amplitude was one of the potential components of the proposed sensor-response signature, this operation was avoided.

Similarly, vector normalization,

$$x_{is}^{\text{vn}}(t_j) = \frac{x_{is}(t_j)}{\sqrt{\sum_{j=0}^{50} x_{is}^2(t_j)}},$$

was not used because it would remove differences in the overall response energy among samples.

The baseline-corrected signals therefore remained in their original digital units during nonlinear response modelling. This preserved both inter-sensor scale information and sample-dependent amplitude variation.

2.4.7. Construction of preprocessed response matrices

Following baseline correction, the preprocessed response matrix for sample *i*was defined as

$$\mathbf{X}_i^{\text{bc}} = \begin{bmatrix} x_{i1}^{\text{bc}}(t_0) & x_{i2}^{\text{bc}}(t_0) & x_{i3}^{\text{bc}}(t_0) & x_{i4}^{\text{bc}}(t_0) \\ x_{i1}^{\text{bc}}(t_1) & x_{i2}^{\text{bc}}(t_1) & x_{i3}^{\text{bc}}(t_1) & x_{i4}^{\text{bc}}(t_1) \\ \vdots & \vdots & \vdots & \vdots \\ x_{i1}^{\text{bc}}(t_{50}) & x_{i2}^{\text{bc}}(t_{50}) & x_{i3}^{\text{bc}}(t_{50}) & x_{i4}^{\text{bc}}(t_{50}) \end{bmatrix}.$$

The dimensionality of each response matrix remained unchanged:

$$\mathbf{X}_i^{\text{bc}} \in \mathbb{R}^{51\times 4}.$$

Therefore, preprocessing did not discard any time point or sensor channel. The reference dataset retained

$$21 \times 51 \times 4 = 4284$$

baseline-corrected sensor observations, and the blind dataset retained the same number.

2.4.8. Feature-level standardization

Although the raw temporal curves were not amplitude-normalized, standardization was applied later to the nonlinear parameters and derived kinetic features before algorithms sensitive to variable scale were trained. Let the sensor-response signature extracted from sample *i*be

$$\mathbf{F}_i = \left[F_{i1}, F_{i2}, \ldots, F_{ip}\right].$$

For feature *k*, standardization was performed using only the labelled reference dataset:

$$Z_{ik} = \frac{F_{ik} - \mu_{k,\mathrm{ref}}}{\sigma_{k,\mathrm{ref}}},$$

where

$$\mu_{k,\mathrm{ref}} = \frac{1}{N_{\mathrm{ref}}} \sum_{i=1}^{N_{\mathrm{ref}}} F_{ik}$$

and

$$\sigma_{k,\mathrm{ref}} = \sqrt{\frac{\sum_{i=1}^{N_{\mathrm{ref}}} \left(F_{ik} - \mu_{k,\mathrm{ref}}\right)^2}{N_{\mathrm{ref}} - 1}}.$$

The independent blind features were transformed using the same reference-derived parameters:

$$Z_{bk}^{\mathrm{blind}} = \frac{F_{bk}^{\mathrm{blind}} - \mu_{k,\mathrm{ref}}}{\sigma_{k,\mathrm{ref}}}.$$

Neither the mean nor standard deviation was recalculated from the blind data. This restriction prevented information from the independent test set from influencing the scale of the model inputs.

Where cross-validation was used within the reference dataset, the scaling parameters were recalculated separately within each training fold:

$$Z_{ik}^{(r)} = \frac{F_{ik} - \mu_{k,\mathrm{train}}^{(r)}}{\sigma_{k,\mathrm{train}}^{(r)}},$$

where *r*denotes the cross-validation iteration. The corresponding validation fold was transformed using the parameters calculated from that iteration's training subset. This nested application of standardization prevented preprocessing leakage and produced a more realistic estimate of generalization performance.

2.4.9. Separation of reference and blind preprocessing

The blind dataset was not used to select preprocessing operations, smoothing parameters, nonlinear models, features, or scaling coefficients. Once the preprocessing procedure had been established from the reference data, it was applied unchanged to each blind trajectory.

The complete preprocessing operation can be expressed as

$$\tilde{\mathbf{X}}_i = \mathscr{P}\left(\mathbf{X}_i; \hat{\boldsymbol{\lambda}}_{\mathrm{ref}}\right),$$

where $\mathscr{P}(\cdot)$denotes the preprocessing function and $\hat{\boldsymbol{\lambda}}_{\mathrm{ref}}$contains all parameters determined from the reference dataset.

For a blind sample b,

$$\tilde{\mathbf{X}}_b^{\mathrm{blind}} = \mathscr{P}\left(\mathbf{X}_b^{\mathrm{blind}}; \hat{\boldsymbol{\lambda}}_{\mathrm{ref}}\right).$$

The blind dataset therefore contributed no information to

$$\hat{\boldsymbol{\lambda}}_{\mathrm{ref}}.$$

This strict separation was essential because preprocessing can influence predictive performance as strongly as the subsequent classifier. Recent work has demonstrated that carefully validated preprocessing can reduce overfitting and produce simpler, more interpretable electronic-nose models without sacrificing prediction performance (Fernandez et al., 2025).

2.4.10. Final preprocessing workflow

The final signal-processing sequence was

$$\mathbf{X}_i \longrightarrow \text{integrity verification} \longrightarrow \text{temporal ordering} \longrightarrow \text{baseline subtraction} \longrightarrow \mathbf{X}_i^{\mathrm{bc}} \longrightarrow \text{nonlinear curve fitting}.$$

After nonlinear parameter estimation, the modelling sequence continued as

$$\mathbf{X}_i^{\mathrm{bc}} \longrightarrow \boldsymbol{\theta}_i \longrightarrow \mathbf{F}_i \longrightarrow \mathbf{Z}_i \longrightarrow \hat{d}_i,$$

where $\boldsymbol{\theta}_i$represents the fitted nonlinear parameters, $\mathbf{F}_i$is the resulting sensor-response signature, $\mathbf{Z}_i$is the reference-standardized feature vector, and $\hat{d}_i$is the predicted storage day.

Accordingly, the preprocessing strategy retained all 51 time points and all four sensor channels, corrected only the observed initial offsets, avoided unsupported smoothing and outlier deletion, preserved amplitude information, and confined all learned transformations to the labelled reference data. This approach ensured that the subsequent nonlinear response parameters reflected the measured temporal behaviour of the electronic-nose signals rather than artifacts introduced by aggressive signal manipulation.

2.5.Nonlinear Models

2.5.1 Candidate Nonlinear Models

The temporal responses recorded by the four gas sensors exhibited a rapid initial increase followed by a progressively decreasing response rate. This behavior indicated that the sensor trajectories could not be represented adequately by a single linear slope over the entire 100-s acquisition period. Candidate nonlinear functions were therefore selected to describe different plausible forms of headspace accumulation, gas diffusion, surface adsorption, and MOS sensor-response kinetics.

The candidate-model set was defined before examining the storage-day prediction performance. Model eligibility was based on four requirements: the function had to represent a non-negative increasing response, accommodate the zero-corrected initial condition, permit interpretable kinetic parameters, and be estimable from the 51 observations available for each sensor trajectory. Recent studies have emphasized that transient semiconductor gas-sensor signals are governed by nonlinear physicochemical processes and that kinetic or complete temporal representations can provide more information than isolated equilibrium values (Chen et al., 2024; Zhang et al., 2024; Zhai et al., 2024).

Following baseline correction, the response of sensor $s$for sample $i$was represented generally as

$$x_{is}^{\mathrm{bc}}(t_j) = g_m\left(t_j;\boldsymbol{\theta}_{is,m}\right) + \varepsilon_{is,m}(t_j),$$

where $g_m(\cdot)$denotes candidate model m, $\boldsymbol{\theta}_{is,m}$is its parameter vector, and $\varepsilon_{is,m}(t_j)$is the residual at time t_j.

The nonlinear models were fitted separately to every sensor trajectory. Therefore, one model fit was obtained for each combination of sample, sensor, and candidate function:

$$i = 1,2,\ldots,21,$$
$$s = 1,2,3,4,$$

and

$$m = 1,2,\ldots,4.$$

For the labelled reference dataset, the number of sensor trajectories subjected to each candidate model was

$$N_{\mathrm{curves,ref}} = 21 \times 4 = 84.$$

Because four principal candidate models were evaluated, the reference-model screening comprised

$$N_{\mathrm{fits,ref}} = 84 \times 4 = 336$$

nonlinear fitting operations.

The blind dataset was not used to choose the functional form, parameter constraints, or optimization settings. Its trajectories were retained for final independent processing using the model-selection procedure established from the reference dataset.

Logarithmic model

The first candidate was a two-parameter logarithmic function:

$$g_{\mathrm{Log}}(t;a,b) = a\ln\ (1+bt).$$

The corresponding sensor model was

$$x_{is}^{\mathrm{bc}}(t) = a_{is}\ln\ \left(1+b_{is}t\right) + \varepsilon_{is}(t).$$

The parameters were constrained by

$$a_{is} > 0$$

and

$$b_{is} > 0.$$

The parameter a_{is}controls the overall response scale, whereas b_{is}controls the temporal rate of increase. The first derivative is

$$\frac{dg_{\mathrm{Log}}(t)}{dt} = \frac{ab}{1+bt}.$$

At the beginning of acquisition, the response rate is

$$\frac{dg_{\mathrm{Log}}(t)}{dt}\,|_{t=0} = ab.$$

The second derivative is

$$\frac{d^2g_{\mathrm{Log}}(t)}{dt^2} = -\frac{ab^2}{(1+bt)^2}.$$

Because

$$\frac{d^2g_{\mathrm{Log}}(t)}{dt^2} < 0,$$

the logarithmic model describes an increasing but continuously decelerating response. This shape corresponded closely to the broad morphology observed in the uploaded sensor data. However, the logarithmic function does not possess a finite asymptote:

$$\lim_{t\to\infty} g_{\mathrm{Log}}(t) = \infty.$$

It was therefore included as a parsimonious empirical benchmark rather than as a complete equilibrium model.

Reference-data fitting confirmed that the logarithmic model converged for all 84 sensor trajectories. Its median coefficient of determination was approximately

$$\tilde{R}^2_{\mathrm{Log}} = 0.9834,$$

and its median root-mean-square error was approximately

$$\widetilde{\mathrm{RMSE}}_{\mathrm{Log}} = 21.34$$

digital units. These values showed that a simple decelerating function captured the general response direction but left systematic curvature unexplained.

Gompertz model

The Gompertz model was included to represent an asymmetric nonlinear transition between a lower and an upper response region. Its conventional form is

$$g_{\mathrm{Gom}}\left(t;A,k,t_{\mathrm{i}}\right) = A\exp\left[-\exp\left(-k\left(t-t_{\mathrm{i}}\right)\right)\right],$$

where $A$is the upper asymptotic response, $k$is the rate coefficient, and $t_{\mathrm{i}}$is the location of the inflection region.

Because the signals were baseline-corrected and required to begin at zero, a normalized Gompertz formulation was used:

$$g^{*}_{\mathrm{Gom}}\left(t;A,k,t_{\mathrm{i}}\right) = A\frac{\exp\left[-\exp\left(-k\left(t-t_{\mathrm{i}}\right)\right)\right] - \exp\left[-\exp\left(kt_{\mathrm{i}}\right)\right]}{1-\exp\left[-\exp\left(kt_{\mathrm{i}}\right)\right]}.$$

This formulation satisfies

$$g^{*}_{\mathrm{Gom}}(0) = 0$$

and

$$\lim_{t\to\infty} g^{*}_{\mathrm{Gom}}(t) = A.$$

The parameter restrictions were

$$A > 0$$

and

$$k > 0.$$

The parameter Arepresents the estimated response plateau, $k$controls the response rate, and t_{i}determines the temporal position of the strongest curvature change. The Gompertz function is asymmetric around its inflection region and is therefore appropriate when the early and late phases of the sensor response do not evolve at equal rates.

Although Gompertz functions are commonly associated with biological growth, their asymmetric saturation structure is also useful as a phenomenological representation of sensor signals generated by progressively changing adsorption and reaction rates. Contemporary research increasingly treats complete electronic-nose trajectories as nonlinear temporal objects rather than reducing them exclusively to maximum or equilibrium responses (Chen et al., 2024; Zhai et al., 2024).

The normalized Gompertz model converged for all 84 reference sensor curves. Its median reference-data fit statistics were

$$\tilde{R}^{2}_{\mathrm{Gom}} = 0.9984$$

and

$$\widetilde{\mathrm{RMSE}}_{\mathrm{Gom}} = 6.95.$$

Its substantially lower residual error than the logarithmic model indicated that asymmetric saturation behavior was present in the recorded responses.

Weibull model

The Weibull response model was selected as a flexible stretched-exponential function:

$$g_{\mathrm{Wei}}\left(t;A,\tau,\beta\right) = A\left[1-\exp\left(-\left(\frac{t}{\tau}\right)^{\beta}\right)\right].$$

The model parameters were constrained as

$$A > 0,$$
$$\tau > 0,$$

and

$$\beta > 0.$$

The parameter Arepresents the asymptotic response, $\tau$is a characteristic time constant, and $\beta$is a dimensionless shape parameter.

The model satisfies the initial condition

$$g_{\mathrm{Wei}}(0) = 0$$

and has the upper limit

$$\lim_{t\to\infty} g_{\mathrm{Wei}}(t) = A.$$

At

$$t = \tau,$$

the response reaches

$$g_{\mathrm{Wei}}(\tau) = A\left(1-e^{-1}\right),$$

or approximately

$$g_{\mathrm{Wei}}(\tau) \approx 0.6321A.$$

The Weibull shape parameter permits several response regimes. When

$$\beta = 1,$$

the function reduces to a simple asymptotic exponential model:

$$g_{\mathrm{Wei}}(t) = A\left[1 - \exp\left(-\frac{t}{\tau}\right)\right].$$

When

$$0 < \beta < 1,$$

the response displays a stretched-exponential form with a rapid initial rise followed by pronounced deceleration. When

$$\beta > 1,$$

the function can represent a sigmoidal accumulation process with a more gradual initial stage.

Stretched-exponential functions have been used to characterize gas-sensor relaxation and response kinetics because they can represent distributed time constants rather than assuming a single ideal kinetic process. Lozovskyi et al. (2024) specifically demonstrated the use of a stretched-exponential model for processing gas-sensitive sensor relaxation dependencies and emphasized its value for automated kinetic analysis.

The Weibull model converged for every reference trajectory and produced median fit statistics of

$$\tilde{R}^2_{\mathrm{Wei}} = 0.9976$$

and

$$\widetilde{\mathrm{RMSE}}_{\mathrm{Wei}} = 8.30.$$

These results supported its inclusion as a physically plausible and computationally stable candidate. Its performance was better than that of the logarithmic model but slightly below that of the Gompertz and Richards functions.

Richards model

The Richards function was included as the most flexible candidate because it generalizes several saturation-type response models through an additional shape parameter. Its conventional form is

$$g_{\mathrm{Ric}}\left(t;A,k,t_{\mathrm{i}},\nu\right) = A\left[1+\nu\exp\left(-k\left(t-t_{\mathrm{i}}\right)\right)\right]^{-\frac{1}{\nu}},$$

where $A$is the upper asymptote, $k$is the response-rate parameter, t_{i}controls the temporal location of the transition, and ν determines asymmetry and curve shape.

A zero-referenced form was used for the baseline-corrected signals:

$$g^{*}_{\mathrm{Ric}}\left(t;A,k,t_{\mathrm{i}},\nu\right) = A\frac{\left[1+\nu\exp\left(-k\left(t-t_{\mathrm{i}}\right)\right)\right]^{-\frac{1}{\nu}} - \left[1+\nu\exp\left(kt_{\mathrm{i}}\right)\right]^{-\frac{1}{\nu}}}{1-\left[1+\nu\exp\left(kt_{\mathrm{i}}\right)\right]^{-\frac{1}{\nu}}}.$$

The parameter restrictions were

$$A > 0,$$
$$k > 0,$$

and

$$\nu > 0.$$

The normalized Richards function satisfies

$$g^{*}_{\mathrm{Ric}}(0) = 0$$

and

$$\lim_{t\to\infty} g^{*}_{\mathrm{Ric}}(t) = A.$$

The additional parameter νenables the function to adapt to a wider range of asymmetric shapes than the Gompertz or Weibull models. As νchanges, the position and degree of curvature can vary without requiring the response amplitude and rate parameters to compensate excessively. This flexibility is relevant for a passive closed chamber because the measured signal reflects several coupled processes, including fruit VOC emission, headspace accumulation, diffusion, adsorption, and sensor-surface reaction.

Recent kinetic modelling of semiconductor gas sensors has shown that dynamic responses can contain mechanistically meaningful rate information and that nonlinear kinetic representations can improve the interpretation and discrimination of VOC responses (Zhang et al., 2024).

The Richards model converged successfully for all 84 reference trajectories. It produced the strongest preliminary fit, with median values of

$$\tilde{R}^2_{\mathrm{Ric}} = 0.9998$$

and

$$\widetilde{\mathrm{RMSE}}_{\mathrm{Ric}} = 2.43.$$

Among the 84 reference sensor curves, the lowest corrected Akaike information criterion was obtained by the Richards model for 81 curves, by the Gompertz model for two curves, and by the Weibull model for one curve. No logarithmic fit produced the minimum corrected Akaike information criterion. These findings demonstrated the suitability of the candidate hierarchy but were not used alone to determine the final model. Parameter stability, residual structure, model complexity, and information criteria were evaluated jointly in the subsequent model-selection stage.

Nonlinear parameter estimation

All candidate models were fitted by nonlinear least squares. For sample i, sensor s, and model m, the parameter estimate was obtained by minimizing the residual sum of squares:

$$\hat{\boldsymbol{\theta}}_{is,m} = \arg\min\ \theta\mathrm{is,m} \sum_{j=0}^{50} \left[x_{is}^{\mathrm{bc}}(t_j) - g_m\left(t_j; \boldsymbol{\theta}_{is,m}\right) \right]^2.$$

The residual for each time point was

$$e_{is,m}(t_j) = x_{is}^{\mathrm{bc}}(t_j) - \hat{x}_{is,m}(t_j),$$

where

$$\hat{x}_{is,m}(t_j) = g_m\left(t_j; \hat{\boldsymbol{\theta}}_{is,m}\right).$$

Physically admissible parameter bounds were imposed to prevent negative asymptotic responses, negative kinetic coefficients, and numerically unstable shape parameters. The fitted response was required to remain finite over the experimental interval:

$$0 \le t \le 100\ \mathrm{s}.$$

A fitting attempt was considered successful only when the optimization algorithm converged, all parameters remained finite, and the predicted curve contained no undefined values over the complete acquisition interval.

Data-based justification of the candidate set

The candidate functions formed a progression from low to high flexibility:

$$\text{Logarithmic} \longrightarrow \text{Weibull} \longrightarrow \text{Gompertz} \longrightarrow \text{Richards}.$$

The logarithmic model provided a two-parameter empirical benchmark for continuously decelerating growth. The Weibull model introduced a finite asymptote and flexible stretched-exponential behavior. The Gompertz model represented asymmetric saturation, while the Richards model allowed the greatest adjustment of asymmetry and curvature.

Across the 84 labelled reference curves, all four models achieved a 100% numerical convergence rate. Their median preliminary fit statistics were

$$\begin{aligned}\tilde{R}^2_{\mathrm{Log}} &= 0.9834,\\ \tilde{R}^2_{\mathrm{Wei}} &= 0.9976,\\ \tilde{R}^2_{\mathrm{Gom}} &= 0.9984,\\ \tilde{R}^2_{\mathrm{Ric}} &= 0.9998,\end{aligned}$$

and

$$\begin{aligned}\widetilde{\mathrm{RMSE}}_{\mathrm{Log}} &= 21.34,\\ \widetilde{\mathrm{RMSE}}_{\mathrm{Wei}} &= 8.30,\\ \widetilde{\mathrm{RMSE}}_{\mathrm{Gom}} &= 6.95,\\ \widetilde{\mathrm{RMSE}}_{\mathrm{Ric}} &= 2.43.\end{aligned}$$

The observed hierarchy confirmed that the curves contained more complex temporal structure than could be represented by a simple logarithmic increase. Nevertheless, the final model was not selected solely according to the largest coefficient of determination. More flexible functions can obtain lower residual errors simply by using additional parameters. The candidate models were therefore compared using goodness-of-fit statistics, information criteria, residual behavior, convergence stability, and parameter interpretability, as described in the following subsection.

This candidate-model strategy retained a parsimonious empirical function, two established asymmetric or stretched saturation functions, and one generalized nonlinear model. It therefore enabled a balanced comparison between simplicity and temporal flexibility while avoiding direct use of the independent blind dataset during analytical development.

2.5.2 Richards Model

The Richards model was selected as the principal nonlinear representation of the temporal electronic-nose responses because it provided the flexibility required to describe differences in response amplitude, rate, transition timing, and curve asymmetry. The model is a generalized logistic function that extends simpler sigmoid models by including an additional shape parameter. Consequently, logistic-like, Gompertz-like, and other asymmetric saturation patterns can be represented within the same functional family (Richards, 1959).

This flexibility was particularly relevant to the present closed-chamber measurements. The recorded MOS sensor signal was generated by several coupled processes, including volatile release from the cherry tomatoes, accumulation within the sealed headspace, passive diffusion toward the sensing surfaces, adsorption, surface reaction, and gradual sensor stabilization. Recent kinetic studies of semiconductor gas sensors have demonstrated that complete transient responses contain compound-dependent reaction information and can be modelled to obtain more interpretable response characteristics than those provided by isolated endpoint values (Zhang et al., 2024).

Richards response function

The conventional four-parameter Richards function can be written as

$$g_{\mathrm{R}}\left(t;A,k,t_{\mathrm{i}},\nu\right) = A\left[1+\nu\exp\left(\quad -k\left(t-t_{\mathrm{i}}\right)\right)\right]^{-\frac{1}{\nu}},$$

where $t$is the acquisition time, $A$is the upper asymptotic response, $k$is the kinetic-rate parameter, $t_{\mathrm{i}}$is a temporal location parameter, and $\nu$is the shape or asymmetry parameter.

The raw sensor trajectories were baseline-corrected before curve fitting. Therefore, each fitted response was required to satisfy

$$g_{\mathrm{R}}^{*}(0) = 0.$$

To impose this initial condition while retaining a finite upper asymptote, the normalized zero-referenced Richards function was used:

$$g_{\mathrm{R}}^{*}\left(t;A,k,t_{\mathrm{i}},\nu\right) = A\frac{f(t) - f(0)}{1 - f(0)},$$

where

$$f(t) = \left[1 + \nu\exp\ \left(\quad -k\left(t - t_{\mathrm{i}}\right)\right)\right]^{-\frac{1}{\nu}}$$

and

$$f(0) = \left[1 + \nu\exp\ \left(kt_{\mathrm{i}}\right)\right]^{-\frac{1}{\nu}}.$$

The complete model fitted to sensor *s*of sample *i*was therefore

$$x_{is}^{\mathrm{bc}}(t_j) = A_{is}\frac{\left[1 + \nu_{is}\exp\ \left(\quad -k_{is}\left(t_j - t_{\mathrm{i},is}\right)\right)\right]^{-\frac{1}{\nu_{is}}} - \left[1 + \nu_{is}\exp\ \left(k_{is}t_{\mathrm{i},is}\right)\right]^{-\frac{1}{\nu_{is}}}}{1 - \left[1 + \nu_{is}\exp\ \left(k_{is}t_{\mathrm{i},is}\right)\right]^{-\frac{1}{\nu_{is}}}} + \varepsilon_{is}(t_j),$$

where $x_{is}^{\mathrm{bc}}(t_j)$is the baseline-corrected observed response and $\varepsilon_{is}(t_j)$is the residual error.

The normalized function satisfies

$$g_{\mathrm{R}}^{*}(0) = 0$$

and

$$\lim_{t\to\infty} g_{\mathrm{R}}^{*}(t) = A.$$

Thus, the response begins at the baseline-corrected origin and approaches the asymptotic level *A*as exposure time increases.

Interpretation of Richards parameters

The parameter vector extracted from each fitted curve was

$$\boldsymbol{\theta}_{is}^{\mathrm{R}} = \left[A_{is}{,}k_{is}{,}t_{\mathrm{i},is}{,}\nu_{is}\right].$$

The asymptotic parameter A_{is}represented the response level predicted after prolonged headspace exposure. Because the measurement ended at 100 s and some trajectories had not reached a complete equilibrium plateau, $A_{is}$was interpreted as a model-estimated asymptotic response rather than a directly observed maximum.

The kinetic parameter k_{is}controlled the temporal steepness of the response. Larger values generally indicated a more rapid transition through the central part of the curve, whereas smaller values indicated a slower response development.

The parameter $t_{\mathrm{i},is}$controlled the temporal position of the main response transition. In the generalized Richards formulation, it functioned as a location parameter associated with the region of strongest curvature change. Because the

zero-referenced normalization modifies the conventional lower asymptote, $t_{i,is}$was interpreted operationally as a temporal transition parameter rather than automatically equated with the exact observed inflection time.

The parameter ν_{is}controlled the shape and asymmetry of the curve. Unlike a standard logistic model, which imposes a fixed symmetric response structure, the Richards model allows the early and late portions of the trajectory to develop at different rates. This property is important for MOS electronic-nose measurements because gas accumulation and sensor-surface reactions do not necessarily produce temporally symmetric responses. Contemporary reviews similarly emphasize that electronic-nose signals are dynamic, cross-sensitive, and strongly influenced by the complete sensing and sampling process (Sanislav et al., 2025; Zhai et al., 2024).

Parameter constraints

Physically and numerically admissible bounds were imposed during nonlinear optimization:

$$A_{is} > 0,$$
$$k_{is} > 0,$$

and

$$\nu_{is} > 0.$$

For each curve, the lower bound for $A_{is}$was set at or above the largest observed baseline-corrected sensor response:

$$A_{is} \geq \max_{j} \left[x_{is}^{\mathrm{bc}}(t_j) \right].$$

This constraint prevented the fitted asymptote from falling below the recorded response.

The optimization interval for $t_{i,is}$was allowed to extend beyond the observed measurement range. This was necessary because some monotonically increasing curves may have an estimated transition location outside the directly observed 0–100-s interval.

Broad positive limits were assigned to the shape parameter to permit substantial asymmetry while preventing numerical instability. Since highly flexible Richards models may exhibit correlation among A, k, t_i, and ν, parameter estimates were accepted only when the fitted function remained finite and monotonic over the complete experimental interval.

Nonlinear least-squares estimation

The Richards parameters were estimated separately for every sample and sensor using nonlinear least squares. The objective function was

$$\hat{\boldsymbol{\theta}}_{is}^{\mathrm{R}} = \arg\min_{A,k,t_i,\nu} \sum_{j=0}^{50} \left[x_{is}^{\mathrm{bc}}(t_j) - g_{\mathrm{R}}^{*}\left(t_j; A_{is}, k_{is}, t_{i,is}, \nu_{is} \right) \right]^2.$$

The fitted value at time $t_j$was

$$\hat{x}_{is}(t_j) = g_{\mathrm{R}}^{*}\left(t_j; \hat{A}_{is}, \hat{k}_{is}, \hat{t}_{i,is}, \hat{\nu}_{is} \right),$$

and the corresponding residual was

$$e_{is}(t_j) = x_{is}^{\mathrm{bc}}(t_j) - \hat{x}_{is}(t_j).$$

The residual sum of squares was calculated as

$$\mathrm{RSS}_{is} = \sum_{j=0}^{50} e_{is}^2(t_j).$$

Multiple initial parameter combinations were evaluated to reduce dependence on a single starting point. The solution producing the smallest finite residual sum of squares while satisfying the parameter bounds was retained.

A curve fit was regarded as numerically successful when the optimization converged, all four parameter estimates were finite, and the predicted values remained defined throughout

$$0 \le t \le 100 \text{ s}.$$

Number of Richards fits

The labelled reference dataset contained 21 sample-level trajectories and four sensor channels. Therefore, the number of independently fitted reference curves was

$$N_{\mathrm{R,ref}} = 21 \times 4 = 84.$$

The independent blind dataset contained the same number of samples and sensor channels:

$$N_{\mathrm{R,blind}} = 21 \times 4 = 84.$$

The total number of Richards fits was therefore

$$N_{\mathrm{R,total}} = 84 + 84 = 168.$$

The reference curves were used to establish the modelling procedure and construct the storage-day signatures. The blind curves were fitted only after the model form, preprocessing operations, parameter bounds, and optimization protocol had been fixed.

Goodness-of-fit assessment

The coefficient of determination was calculated as

$$R_{is}^2 = 1 - \frac{\sum_{j=0}^{50} \left[x_{is}^{\mathrm{bc}}(t_j) - \hat{x}_{is}(t_j)\right]^2}{\sum_{j=0}^{50} \left[x_{is}^{\mathrm{bc}}(t_j) - \overline{x_{is}^{\mathrm{bc}}}\right]^2},$$

where

$$\overline{x_{is}^{\mathrm{bc}}} = \frac{1}{51} \sum_{j=0}^{50} x_{is}^{\mathrm{bc}}(t_j).$$

The root-mean-square error was calculated as

$$\mathrm{RMSE}_{is} = \sqrt{\frac{1}{51} \sum_{j=0}^{50} \left[x_{is}^{\mathrm{bc}}(t_j) - \hat{x}_{is}(t_j)\right]^2}.$$

Because the Richards function contained four fitted parameters, model fit was also evaluated using complexity-adjusted criteria. The Akaike information criterion was

$$\mathrm{AIC}_{is} = n\ln\left(\frac{\mathrm{RSS}_{is}}{n}\right) + 2p,$$

where

$$n = 51$$

and

$$p = 4.$$

The corrected Akaike information criterion was

$$\mathrm{AICc}_{is} = \mathrm{AIC}_{is} + \frac{2p(p+1)}{n-p-1}.$$

The Bayesian information criterion was

$$\mathrm{BIC}_{is} = n\ln\left(\frac{\mathrm{RSS}_{is}}{n}\right) + p\ln(n).$$

The use of information criteria was important because the Richards function was more flexible than the logarithmic, Weibull, and Gompertz candidates. A lower residual error alone would not demonstrate that its additional shape parameter was justified.

Reference-data fitting performance

The Richards model converged successfully for all reference sensor trajectories:

$$N_{\mathrm{converged,ref}} = 84.$$

Therefore, the reference convergence rate was

$$\mathrm{CR}_{\mathrm{ref}} = \frac{84}{84} \times 100 = 100\%.$$

Across the 84 reference fits, the median coefficient of determination was

$$\tilde{R}^2_{\mathrm{ref}} = 0.9998.$$

The mean coefficient of determination was approximately

$$\overline{R^2_{\mathrm{ref}}} = 0.9997,$$

while the minimum value was

$$R^2_{\mathrm{min,ref}} = 0.9981.$$

All 84 reference curves exceeded

$$R^2 = 0.99.$$

The median reference RMSE was

$$\mathrm{RMSE}_{\mathrm{ref}} = 2.50$$

digital units, and the mean RMSE was

$$\mathrm{RMSE}_{\mathrm{ref}} = 2.53$$

digital units.

The maximum RMSE among all reference fits was

$$\mathrm{RMSE}_{\max,\mathrm{ref}} = 4.72$$

digital units.

These errors were small relative to sensor maxima of approximately 775–835 digital units, indicating that the Richards function closely reproduced the measured nonlinear trajectories.

The median fit statistics obtained separately for the four sensors were

$$\begin{aligned} \tilde{R}^2_{S1} &= 0.9997, \\ \tilde{R}^2_{S2} &= 0.9998, \\ \tilde{R}^2_{S3} &= 0.9998, \\ \tilde{R}^2_{S4} &= 0.9998. \end{aligned}$$

The corresponding median RMSE values were

$$\begin{aligned} \mathrm{RMSE}_{S1} &= 2.47, \\ \mathrm{RMSE}_{S2} &= 2.68, \\ \mathrm{RMSE}_{S3} &= 1.97, \\ \mathrm{RMSE}_{S4} &= 2.50. \end{aligned}$$

Thus, the model provided consistently high-quality fits across all four sensing channels rather than fitting only one dominant sensor.

Distribution of fitted reference parameters

Across the 84 reference sensor curves, the median estimated asymptotic response was

$$\tilde{A}_{\mathrm{ref}} = 536.95$$

digital units.

The interquartile range was approximately

$$395.89 \leq A \leq 659.75.$$

The median kinetic-rate parameter was

$$\tilde{k}_{\mathrm{ref}} = 0.0984\ \mathrm{s}^{-1},$$

with an interquartile range of

$$0.0684 \le k \le 0.1233\ \mathrm{s}^{-1}.$$

The median temporal transition parameter was

$$\tilde{t}_{\mathrm{i,ref}} = 59.58\ \mathrm{s},$$

and its interquartile range was

$$55.46 \le t_{\mathrm{i}} \le 64.12\ \mathrm{s}.$$

The median shape parameter was

$$\tilde{\nu}_{\mathrm{ref}} = 7.20.$$

However, some νestimates approached the imposed upper optimization bound. This indicated that, for a subset of curves, several combinations of the Richards parameters could produce very similar fitted shapes. Therefore, $\nu$was retained as a useful empirical shape descriptor but was not interpreted as a directly measurable physicochemical constant.

The median sensor-specific asymptotic parameters were

$$\begin{aligned}
\tilde{A}_{S1} &= 566.05,\\
\tilde{A}_{S2} &= 532.37,\\
\tilde{A}_{S3} &= 541.53,\\
\tilde{A}_{S4} &= 532.37.
\end{aligned}$$

The corresponding median kinetic-rate parameters were

$$\begin{aligned}
\tilde{k}_{S1} &= 0.1024\ \mathrm{s}^{-1},\\
\tilde{k}_{S2} &= 0.0924\ \mathrm{s}^{-1},\\
\tilde{k}_{S3} &= 0.0962\ \mathrm{s}^{-1},\\
\tilde{k}_{S4} &= 0.1024\ \mathrm{s}^{-1}.
\end{aligned}$$

These differences demonstrated that the four sensors did not produce identical temporal responses, even though they were exposed simultaneously to the same chamber headspace.

Blind-data fitting performance

The same fixed Richards model and parameter constraints were applied to the independent blind dataset. All 84 blind sensor curves converged successfully:

$$N_{\mathrm{converged,blind}} = 84.$$

The blind convergence rate was therefore

$$\mathrm{CR}_{\ \mathrm{blind}} = 100\%.$$

The median blind coefficient of determination was

$$\tilde{R}^2_{\text{blind}} = 0.9998,$$

and the mean value was approximately

$$\overline{R^2_{\text{blind}}} = 0.9997.$$

The lowest blind coefficient of determination was

$$R^2_{\text{min ,blind}} = 0.9981.$$

The median and mean blind RMSE values were

$$\text{RMSE}_{\text{blind}} = 2.50$$

and

$$\text{RMSE}_{\text{blind}} = 2.52$$

digital units, respectively.

The maximum blind RMSE was

$$\text{RMSE}_{\text{ max ,blind}} = 4.72$$

digital units.

The near-identical reference and blind fitting results showed that the Richards function was not restricted to the labelled calibration trajectories. It also represented the independent unlabelled responses with comparable numerical accuracy.

Derived characteristic response times

The fitted Richards function allowed characteristic temporal descriptors to be calculated without selecting arbitrary observed time points. For a response fraction q, the characteristic time $t_q$was defined implicitly by

$$g^*_{\text{R}}(t_q) = qA,$$

where

$$0 < q < 1.$$

The half-response time was obtained from

$$g^*_{\text{R}}(t_{50}) = 0.50A,$$

and the 90% response time was determined from

$$g^*_{\text{R}}(t_{90}) = 0.90A.$$

Using

$$f_q = f(0) + q\,[1 - f(0)]\,,$$

the general expression for $t_q$was

$$t_q = t_{\mathrm{i}} - \frac{1}{k}\ln\left[\frac{f_q^{-\nu} - 1}{\nu}\right].$$

Therefore,

$$t_{50} = t_{\mathrm{i}} - \frac{1}{k}\ln\left[\frac{\left(f(0) + 0.50[1 - f(0)]\right)^{-\nu} - 1}{\nu}\right],$$

and

$$t_{90} = t_{\mathrm{i}} - \frac{1}{k}\ln\left[\frac{\left(f(0) + 0.90[1 - f(0)]\right)^{-\nu} - 1}{\nu}\right].$$

These variables describe the time required to reach 50% and 90% of the model-estimated response and provide more interpretable kinetic descriptors than isolated raw ADC readings.

Response-rate descriptors

The instantaneous Richards response rate was defined as

$$v_{is}(t) = \frac{dg_{\mathrm{R}}^{*}(t)}{dt}.$$

Using the normalized Richards formulation,

$$v_{is}(t) = \frac{A_{is}k_{is}f_{is}(t)\left[1 - f_{is}^{\nu_{is}}(t)\right]}{1 - f_{is}(0)}.$$

The maximum model-derived response rate was

$$v_{is,\max} = \max_{0 \le t \le 100} v_{is}(t).$$

The time at which this maximum occurred was

$$t_{is,v_{\max}} = \arg\max\ 0 \le \mathrm{t} \le 100\ v_{is}(t).$$

These descriptors summarized how quickly each sensor reacted to the evolving cherry tomato headspace and when its most rapid response occurred.

Richards-based sensor-response signature

For one sample, the four fitted parameters from the four sensors generated a 16-variable primary Richards signature:

$$\mathbf{F}_i^{\mathrm{R}} = \left[A_{i1}, k_{i1}, t_{\mathrm{i},i1}, \nu_{i1}, \ldots, A_{i4}, k_{i4}, t_{\mathrm{i},i4}, \nu_{i4}\right].$$

The dimensionality of this primary signature was

$$p_{\mathrm{R}} = 4 \times 4 = 16.$$

When model-derived temporal descriptors were included, the sensor-specific feature vector became

$$\boldsymbol{\phi}_{is} = \left[A_{is}, k_{is}, t_{\mathrm{i},is}, v_{is}, t_{50,is}, t_{90,is}, v_{is,\max}, t_{is,v_{\max}} \right].$$

The resulting extended four-sensor signature was

$$\mathbf{F}_i^{\mathrm{ext}} = \left[\boldsymbol{\phi}_{i1}, \boldsymbol{\phi}_{i2}, \boldsymbol{\phi}_{i3}, \boldsymbol{\phi}_{i4} \right].$$

This representation converted

$$51 \times 4 = 204$$

raw temporal sensor values into a compact set of nonlinear amplitude, rate, shape, and timing descriptors. It thereby reduced temporal collinearity while preserving the principal dynamic characteristics of the original curves.

Overall, the Richards model provided a numerically stable and highly accurate continuous representation of both the reference and blind electronic-nose responses. Its principal advantage was not merely its high R^2, but its ability to produce a common parameterized description of response magnitude, kinetic rate, temporal transition, asymmetry, and characteristic response times. These fitted quantities formed the basis of the sensor-response signatures used in the subsequent storage-day identification analyses.

2.5.5. Model Comparison and Selection

The logarithmic, Weibull, Gompertz, and Richards models were compared systematically to determine the most appropriate continuous representation of the temporal electronic-nose responses. Model selection was performed exclusively using the labelled reference dataset. The independent blind dataset was not used to select the model, adjust parameter bounds, define starting values, or modify the fitting procedure. It was examined only after the analytical protocol had been fixed to evaluate whether the selected model retained comparable numerical behaviour in previously unseen response trajectories.

Model comparison considered goodness of fit, residual magnitude, model complexity, numerical convergence, parameter validity, and consistency across the four sensor channels. This multimetric approach was necessary because selecting a nonlinear function solely according to the highest coefficient of determination can favour an unnecessarily flexible model. Information-theoretic criteria provide a more defensible balance between goodness of fit and parameter count and are particularly useful when several non-nested nonlinear functions are fitted to the same observations (Portet, 2020).

Recent electronic-nose research also emphasizes that complete transient sensor signals contain information related to response kinetics, adsorption–reaction processes, and temporal signal shape. Consequently, a model selected for these signals must reproduce the entire response trajectory rather than only its final value (Chen et al., 2024; Sanislav et al., 2025; Zhang et al., 2024).

Comparison framework

For sample i, sensor s, and candidate model m, the fitted response was expressed as

$$\hat{x}_{is,m}(t_j) = g_m\left(t_j; \hat{\boldsymbol{\theta}}_{is,m}\right),$$

where $g_m(\cdot)$denotes the logarithmic, Weibull, Gompertz, or Richards function and $\hat{\boldsymbol{\theta}}_{is,m}$is the corresponding estimated parameter vector.

The residual at acquisition time $t_j$was

$$e_{is,m}(t_j) = x_{is}^{\mathrm{bc}}(t_j) - \hat{x}_{is,m}(t_j).$$

The residual sum of squares was calculated as

$$\mathrm{RSS}_{is,m} = \sum_{j=0}^{50} e_{is,m}^2(t_j).$$

Because every response trajectory contained 51 time points,

$$n = 51.$$

The number of fitted parameters differed among models:

$$p_{\mathrm{Log}} = 2,$$
$$p_{\mathrm{Wei}} = 3,$$
$$p_{\mathrm{Gom}} = 3,$$

and

$$p_{\mathrm{Ric}} = 4.$$

Thus, the Richards model had the greatest functional flexibility and required explicit penalization for its additional parameter.

Coefficient of determination

The coefficient of determination was calculated as

$$R_{is,m}^2 = 1 - \frac{\sum_{j=0}^{50}\left[x_{is}^{\mathrm{bc}}(t_j) - \hat{x}_{is,m}(t_j)\right]^2}{\sum_{j=0}^{50}\left[x_{is}^{\mathrm{bc}}(t_j) - \overline{x_{is}^{\mathrm{bc}}}\right]^2},$$

where

$$\overline{x_{is}^{\mathrm{bc}}} = \frac{1}{51}\sum_{j=0}^{50} x_{is}^{\mathrm{bc}}(t_j).$$

Higher values indicated that a larger proportion of the temporal variation was reproduced by the fitted model. However, $R^2$was used as a descriptive measure rather than as the sole model-selection criterion because it does not directly penalize additional model parameters.

Root-mean-square error

The root-mean-square error was calculated as

$$\mathrm{RMSE}_{is,m} = \sqrt{\frac{1}{51}\sum_{j=0}^{50}\left[x_{is}^{\mathrm{bc}}(t_j) - \hat{x}_{is,m}(t_j)\right]^2}.$$

RMSE remained in the original digital sensor-response units and therefore provided a direct measure of the average discrepancy between the fitted curve and the measured trajectory. Lower values indicated better correspondence with the observed response.

Akaike information criterion

The Akaike information criterion was calculated as

$$\mathrm{AIC}_{is,m} = n\ln\left(\frac{\mathrm{RSS}_{is,m}}{n}\right) + 2p_m.$$

AIC penalizes model complexity through the term $2p_m$. However, because each trajectory contained only 51 observations, the small-sample correction was also applied.

The corrected Akaike information criterion was

$$\mathrm{AICc}_{is,m} = \mathrm{AIC}_{is,m} + \frac{2p_m(p_m+1)}{n-p_m-1}.$$

AICc was treated as the principal information-theoretic selection criterion because it provides additional protection against overfitting when the sample size is not large relative to the number of parameters. Model rankings based on AIC or AICc identify the candidate that achieves the most favourable balance between approximation error and complexity within the prespecified candidate set (Portet, 2020).

For each fitted sensor trajectory, the AICc difference was determined as

$$\Delta_{is,m}^{\mathrm{AICc}} = \mathrm{AICc}_{is,m} - \min_{m}\left(\mathrm{AICc}_{is,m}\right).$$

The best-supported model for a particular trajectory therefore satisfied

$$\Delta_{is,m}^{\mathrm{AICc}} = 0.$$

The relative support for each model could additionally be expressed through the Akaike weight:

$$w_{is,m} = \frac{\exp\left(-\frac{1}{2}\Delta_{is,m}^{\mathrm{AICc}}\right)}{\sum_{r=1}^{4}\exp\left(-\frac{1}{2}\Delta_{is,r}^{\mathrm{AICc}}\right)}.$$

The weights satisfy

$$\sum_{m=1}^{4} w_{is,m} = 1.$$

Bayesian information criterion

The Bayesian information criterion was calculated as

$$\mathrm{BIC}_{is,m} = n\ln\left(\frac{\mathrm{RSS}_{is,m}}{n}\right) + p_m\ln(n).$$

BIC imposes a stronger complexity penalty than AIC for the present sample size. Agreement between AICc and BIC rankings was therefore interpreted as strong evidence that the improvement achieved by a more complex model was not attributable solely to additional parameters.

Convergence and admissibility criteria

A candidate fit was accepted only if the nonlinear optimization converged and all estimated parameters remained finite and within their predefined bounds. In addition, the predicted response had to remain defined throughout the experimental interval:

$$0 \le t \le 100\ \mathrm{s}.$$

The model-specific convergence rate was calculated as

$$\mathrm{CR}_m = \frac{N_{\mathrm{converged},m}}{N_{\mathrm{attempted},m}} \times 100.$$

The reference dataset comprised 21 sample trajectories and four sensor channels. Consequently, each candidate model was fitted to

$$N_{\text{attempted},m} = 21 \times 4 = 84$$

sensor-response curves.

All four models converged for all 84 reference trajectories:

$$\text{CR}_{\text{Log}} = \text{CR}_{\text{Wei}} = \text{CR}_{\text{Gom}} = \text{CR}_{\text{Ric}} = 100\%.$$

Therefore, the final decision was not influenced by differences in numerical convergence rate.

Goodness-of-fit comparison in the reference dataset

The logarithmic model produced a median coefficient of determination of

$$\tilde{R}^2_{\text{Log}} = 0.98345.$$

Its mean coefficient of determination was

$$\overline{R^2_{\text{Log}}} = 0.98307,$$

and its minimum value across the 84 sensor curves was

$$R^2_{\text{min},\text{Log}} = 0.96582.$$

The corresponding median and mean RMSE values were

$$\text{RMSE}_{\text{Log}} = 21.34$$

and

$$\overline{\text{RMSE}}_{\text{Log}} = 21.05$$

digital units, respectively.

The maximum logarithmic-model RMSE was

$$\text{RMSE}_{\text{max},\text{Log}} = 44.77.$$

Although the logarithmic function reproduced the general increase in sensor response, its residual error showed that a continuously decelerating two-parameter curve could not fully represent the observed transition and saturation behaviour.

The Weibull model provided a substantially better fit. Its median and mean coefficients of determination were

$$\tilde{R}^2_{\text{Wei}} = 0.99755$$

and

$$\bar{R}^2_{\mathrm{Wei}} = 0.99762.$$

Its median RMSE was

$$\mathrm{RMSE}_{\mathrm{Wei}} = 8.30,$$

while its mean RMSE was

$$\mathrm{RMSE}_{\mathrm{Wei}} = 7.97.$$

The Gompertz model further improved the curve representation. Its median coefficient of determination was

$$\tilde{R}^2_{\mathrm{Gom}} = 0.99835,$$

and its mean coefficient of determination was

$$\bar{R}^2_{\mathrm{Gom}} = 0.99829.$$

The median and mean Gompertz RMSE values were

$$\mathrm{RMSE}_{\mathrm{Gom}} = 6.95$$

and

$$\mathrm{RMSE}_{\mathrm{Gom}} = 6.74.$$

The Richards model produced the highest overall goodness of fit. Its median and mean coefficients of determination were

$$\tilde{R}^2_{\mathrm{Ric}} = 0.99979$$

and

$$\bar{R}^2_{\mathrm{Ric}} = 0.99974.$$

The minimum Richards-model coefficient of determination across all reference curves was

$$R^2_{\min,\mathrm{Ric}} = 0.99875.$$

Thus, even the least accurately fitted Richards trajectory retained more than 99.8% of the measured temporal variation.

The median and mean Richards RMSE values were

$$\mathrm{RMSE}_{\mathrm{Ric}} = 2.40$$

and

$$\mathrm{RMSE}_{\mathrm{Ric}} = 2.46$$

digital units.

The maximum Richards RMSE was only

$$\mathrm{RMSE}_{\max,\mathrm{Ric}} = 4.72.$$

Relative to the Gompertz model, the Richards model reduced the median RMSE by approximately

$$\frac{6.95 - 2.40}{6.95} \times 100 = 65.5\%.$$

Relative to the Weibull model, the median RMSE reduction was approximately

$$\frac{8.30 - 2.40}{8.30} \times 100 = 71.1\%.$$

Compared with the logarithmic model, the reduction was approximately

$$\frac{21.34 - 2.40}{21.34} \times 100 = 88.8\%.$$

These improvements indicated that the additional asymmetry parameter of the Richards function captured genuine temporal structure rather than producing only a marginal statistical improvement.

Information-criterion comparison

The median AIC values obtained from the 84 reference sensor trajectories were

$$\mathrm{AIC}_{\mathrm{Log}} = 316.18,$$
$$\mathrm{AIC}_{\mathrm{Wei}} = 221.87,$$
$$\mathrm{AIC}_{\mathrm{Gom}} = 203.77,$$

and

$$\mathrm{AIC}_{\mathrm{Ric}} = 97.17.$$

After correction for the finite number of observations, the median AICc values were

$$\mathrm{AICc}_{\mathrm{Log}} = 316.43,$$
$$\mathrm{AICc}_{\mathrm{Wei}} = 222.38,$$
$$\mathrm{AICc}_{\mathrm{Gom}} = 204.28,$$

and

$$\mathrm{AICc}_{\mathrm{Ric}} = 98.03.$$

The median AICc advantage of the Richards model over the next-best Gompertz model was therefore

$$204.28 - 98.03 = 106.25.$$

The trajectory-level median AICc difference between Richards and Gompertz was approximately

$$\tilde{\Delta}^{\mathrm{AICc}}_{\mathrm{Gom-Ric}} = 93.31.$$

This difference was far greater than the conventional threshold generally interpreted as strong evidence for one model over another. The improvement therefore remained decisive even after penalizing the Richards function for its fourth fitted parameter.

The median BIC values were

$$\mathrm{BIC}_{\mathrm{Log}} = 320.05,$$
$$\mathrm{BIC}_{\mathrm{Wei}} = 227.67,$$
$$\mathrm{BIC}_{\mathrm{Gom}} = 209.57,$$

and

$$\mathrm{BIC}_{\mathrm{Ric}} = 104.89.$$

AIC, AICc, and BIC therefore produced the same overall model ranking:

$$\text{Richards} < \text{Gompertz} < \text{Weibull} < \text{Logarithmic},$$

where lower values indicate superior model support.

The agreement of all three information criteria was especially important because it demonstrated that the Richards model remained preferred under both moderate and stronger complexity penalties.

Trajectory-level selection frequency

The best model was also determined separately for each of the 84 reference sensor curves. Based on the minimum AICc,

$$N_{\mathrm{Richards}} = 81,$$
$$N_{\mathrm{Gompertz}} = 2,$$
$$N_{\mathrm{Weibull}} = 1,$$

and

$$N_{\mathrm{Logarithmic}} = 0.$$

The corresponding selection percentages were

$$P_{\mathrm{Richards}} = \frac{81}{84} \times 100 = 96.43\%,$$
$$P_{\mathrm{Gompertz}} = \frac{2}{84} \times 100 = 2.38\%,$$

and

$$P_{\mathrm{Weibull}} = \frac{1}{84} \times 100 = 1.19\%.$$

The logarithmic model was not selected for any reference trajectory:

$$P_{\text{Logarithmic}} = 0\%.$$

BIC produced exactly the same trajectory-level selection distribution. Moreover, the Richards model produced the lowest RMSE for 82 of the 84 reference curves:

$$P_{\text{Richards,RMSE}} = \frac{82}{84} \times 100 = 97.62\%.$$

For 80 of the 84 trajectories, the AICc difference between the best and second-best model exceeded 10:

$$\frac{80}{84} \times 100 = 95.24\%.$$

Only three trajectories produced an AICc separation of 2 or less between the two highest-ranked functions. Thus, the preference for Richards was not created by a large number of nearly indistinguishable model comparisons; it was decisive for the overwhelming majority of curves.

Residual considerations

Model selection also considered the expected structure of the residuals. For an adequate nonlinear model, residuals should fluctuate around zero without systematic temporal patterns:

$$\mathrm{E}\left[e_{is,m}(t)\right] \approx 0.$$

A persistent positive–negative–positive residual sequence would indicate that the model failed to reproduce the changing curvature of the sensor trajectory. The comparatively large RMSE values obtained from the logarithmic, Weibull, and Gompertz functions indicated that these models left more unrepresented temporal structure than the Richards function.

The Richards model substantially reduced residual magnitude across all four sensors while maintaining finite, monotonically increasing predictions over the 0–100-s acquisition interval. Its advantage was therefore consistent with the nonlinear and asymmetric nature of gas-sensor response kinetics. Recent studies have similarly shown that dynamic semiconductor sensor responses are shaped by nonlinear adsorption and reaction mechanisms and can provide discriminative information when represented through appropriate kinetic models (Zhang et al., 2024).

Evaluation using the independent blind trajectories

The blind dataset was not included in the formal selection decision. Nevertheless, after the Richards model had been selected from the labelled reference data, all four candidate models were fitted to the 84 blind sensor trajectories using the fixed fitting protocol.

All candidate functions achieved a 100% convergence rate in the blind dataset. The median blind-data coefficients of determination were

$$\tilde{R}^2_{\text{Log,blind}} = 0.98393,$$
$$\tilde{R}^2_{\text{Wei,blind}} = 0.99749,$$
$$\tilde{R}^2_{\text{Gom,blind}} = 0.99832,$$

and

$$\tilde{R}^2_{\text{Ric,blind}} = 0.99979.$$

The corresponding median RMSE values were

$$\mathrm{RMSE}_{\mathrm{Log,blind}} = 20.98,$$
$$\mathrm{RMSE}_{\mathrm{Wei,blind}} = 8.31,$$
$$\mathrm{RMSE}_{\mathrm{Gom,blind}} = 6.95,$$

and

$$\mathrm{RMSE}_{\mathrm{Ric,blind}} = 2.37.$$

The median blind AICc values were

$$\mathrm{AICc}_{\mathrm{Log,blind}} = 314.69,$$
$$\mathrm{AICc}_{\mathrm{Wei,blind}} = 222.49,$$
$$\mathrm{AICc}_{\mathrm{Gom,blind}} = 204.28,$$

and

$$\mathrm{AICc}_{\mathrm{Ric,blind}} = 96.81.$$

Richards obtained the lowest blind-trajectory AICc for 81 of the 84 curves, while Gompertz and Weibull were preferred for two and one curves, respectively. This distribution was identical to that obtained from the labelled reference dataset.

The similarity between the two datasets can be summarized as

$$\tilde{R}^2_{\mathrm{Ric,ref}} = \tilde{R}^2_{\mathrm{Ric,blind}} = 0.99979.$$

Likewise,

$$\mathrm{RMSE}_{\mathrm{Ric,ref}} = 2.40$$

and

$$\mathrm{RMSE}_{\mathrm{Ric,blind}} = 2.37.$$

These blind-data results were not used to revise the model choice. Instead, they demonstrated that the reference-selected Richards function retained almost identical numerical performance when applied to independent trajectories.

Final model-selection rule

A candidate function was considered suitable as the common temporal model only if it satisfied all of the following conditions:

$$\mathrm{CR}_{m} = 100\%,$$
$$\tilde{R}^2_m \to 1,$$
$$\mathrm{RMSE}_m \to 0,$$
$$\mathrm{AICc}_m = \min\left(\mathrm{AICc}\right),$$
$$\mathrm{BIC}_m = \min\left(\mathrm{BIC}\right),$$

and the fitted parameters and predictions had to remain numerically admissible and interpretable.

The selection process can therefore be summarized as

$$m^* = \arg\min_m \left\{ \mathrm{AICc}_m, \mathrm{BIC}_m, \mathrm{RMSE}_m \right\},$$

subject to

$$R_m^2 \to 1$$

and

$$\mathrm{CR}_m = 100\%.$$

Under these criteria,

$$m^* = \text{Richards}.$$

The Richards function was consequently selected as the common nonlinear model for all subsequent reference and blind sensor-response analyses. Its selection was supported by the highest R^2, lowest RMSE, lowest AIC, lowest AICc, lowest BIC, complete convergence, and dominant trajectory-level selection frequency. The model also provided interpretable parameters related to asymptotic magnitude, kinetic rate, temporal transition, and response asymmetry.

Using a single common model rather than selecting a different function for each sample ensured that all sensor-response signatures had the same mathematical structure:

$$\boldsymbol{\theta}_{is} = \left[A_{is}, k_{is}, t_{\mathrm{i},is}, \nu_{is} \right].$$

This common parameterization was essential for comparing samples, constructing consistent multichannel signatures, and identifying the storage days of the independent blind samples.

2.6. Reference SRS Library Construction

The dynamic responses of metal-oxide semiconductor (MOS) gas sensors are governed by nonlinear adsorption, surface reaction, and desorption processes. Consequently, the response trajectories generally exhibit a rapid initial increase followed by a progressively decreasing response rate before approaching an equilibrium state. Such behavior cannot be adequately represented by linear models or by simple exponential functions. To capture these nonlinear response characteristics, the Richards model was adopted as the principal mathematical representation of the electronic-nose sensor signals.

The Richards model is a generalized logistic function that introduces an additional shape parameter to control curve asymmetry. Compared with conventional Logistic, Gompertz, or Weibull functions, the Richards equation provides greater flexibility for describing response curves with different transition rates and saturation behaviors (Richards, 1959). Recent studies have also demonstrated that nonlinear kinetic modelling of complete sensor-response trajectories provides more informative descriptors than equilibrium responses alone for gas-sensing applications (Zhai et al., 2024; Zhang et al., 2024).

The Richards function used in this study can be expressed as

$$g(t) = A\left[1 + \nu \exp\left(-k(t - t_i)\right)\right]^{-1/\nu},$$

where Adenotes the asymptotic sensor response, $k$is the kinetic-rate coefficient, $t_i$represents the temporal transition parameter, and νcontrols the asymmetry of the response curve.

Because all sensor trajectories were baseline-corrected prior to modelling, a zero-referenced form of the Richards equation was employed to satisfy the initial condition

$$g(0) = 0.$$

Accordingly, the fitted response for sensor *s* of sample *i* was defined as

$$\hat{x}_{is}(t) = A_{is}\frac{f_{is}(t) - f_{is}(0)}{1 - f_{is}(0)},$$

where

$$f_{is}(t) = \left[1 + \nu_{is}\exp\ \left(-k_{is}(t - t_{i,is})\right)\right]^{-1/\nu_{is}}.$$

This formulation preserves the measured baseline while allowing the response to converge toward an estimated asymptotic value,

$$\lim_{t\to\infty} \hat{x}_{is}(t) = A_{is}.$$

The four model parameters have direct physical interpretations. The parameter A estimates the final equilibrium response expected under prolonged exposure. The parameter k characterizes the overall response speed, with larger values indicating faster signal development. The parameter t_i describes the temporal position of the main transition region, whereas the shape parameter ν determines the asymmetry of the nonlinear response. Together, these parameters provide a compact description of the complete sensor dynamics.

Model parameters were estimated independently for every sensor trajectory using constrained nonlinear least-squares optimization. Parameter estimation minimized the residual sum of squares between the measured and predicted responses,

$$\hat{\boldsymbol{\theta}} = \arg\min\ \theta\sum_{j=1}^{51}\left[x(t_j) - \hat{x}(t_j)\right]^2,$$

where

$$\boldsymbol{\theta} = [A,k,t_i,\nu].$$

To ensure physically meaningful solutions, positive constraints were imposed on the asymptotic response, kinetic coefficient, and shape parameter,

$$A > 0,$$
$$k > 0,$$
$$\nu > 0.$$

Multiple initial parameter estimates were evaluated to reduce the risk of convergence toward local minima. A fitted model was accepted only when the optimization converged successfully and all estimated parameters remained finite throughout the experimental interval.

The Richards model was fitted separately to every sensor trajectory in both the labelled reference dataset and the independent blind dataset using an identical optimization procedure. Importantly, the blind data were not used during model development or parameter tuning. All preprocessing operations, parameter constraints, optimization settings, and model-selection criteria were established exclusively from the reference dataset before being applied to the blind measurements.

The fitted Richards curves provided a continuous mathematical representation of each sensor response from which characteristic kinetic descriptors—including asymptotic response, response rates, characteristic response times, and

integrated response measures—were subsequently extracted. These descriptors formed the basis of the Sensor Response Signature (SRS) developed for storage-day identification.

Model performance was evaluated quantitatively using the coefficient of determination, root-mean-square error, corrected Akaike information criterion, and Bayesian information criterion. The comparative performance of the Richards model relative to the Logarithmic, Gompertz, and Weibull models is presented in the Results section.

2.7. Reference SRS Library Construction

Following the generation of the Sensor Response Signature (SRS) described in Section 2.6, a reference SRS library was constructed to serve as the knowledge base for deterministic storage-day identification. The proposed library stores representative multidimensional signatures obtained from experimentally characterized cherry samples and enables unknown samples to be identified through direct comparison with previously established reference signatures. Unlike conventional machine-learning approaches, which require model training using labeled datasets, the proposed framework is based on a continuously expandable reference library generated directly from experimentally measured sensor responses.

The experimental dataset comprised 21 reference sample groups, representing three independent producers evaluated over seven storage days. For each sample group, the temporal electrical responses recorded by the four MOS sensors were first modeled using the Richards function. The fitted curves were subsequently integrated over the entire measurement period to calculate the sensor-specific AUC values, which were combined to form a unique four-dimensional Sensor Response Signature for each reference sample.

The Sensor Response Signature of the j^{th}reference sample is defined as

$$\mathrm{SRS}_j = \left[AUC_{S1}, AUC_{S2}, AUC_{S3}, AUC_{S4}\right]_j$$

where the four AUC values correspond to the integrated temporal responses of Sensors S1–S4, respectively.

Each reference record in the library contains the following information:

- Storage day,
- Producer (or sample group) identifier,
- AUC values of Sensors S1–S4,
- Corresponding four-dimensional Sensor Response Signature (SRS).

Accordingly, the completed reference library consists of 21 experimentally derived records, each uniquely representing one producer–storage-day combination. Since every record is represented by only four numerical descriptors instead of the original 51-point time series for each sensor, the proposed library substantially reduces data dimensionality while preserving the overall characteristics of the complete temporal electrical responses.

An important advantage of the proposed library is its scalability. When additional producers, cultivars, storage conditions, packaging systems, or postharvest treatments are investigated, new reference records can be incorporated simply by calculating the corresponding SRS vector and appending it to the existing library. Therefore, expansion of the database does not require retraining or recalibration of a predictive model, allowing the reference library to evolve incrementally as new experimental data become available.

The constructed SRS library constitutes the core component of the proposed deterministic identification framework. During identification, the Sensor Response Signature of an unknown sample is generated using the same preprocessing, nonlinear modeling, and AUC calculation procedures described in the previous sections. The resulting SRS vector is then compared with the reference signatures stored in the library, and the storage day is assigned according to the most similar reference record. Consequently, the proposed methodology transforms complex temporal sensor responses into a compact, reproducible, and continuously expandable reference database that supports deterministic storage-day identification without relying on conventional supervised machine-learning classifiers.

2.8. Independent Blind Storage-Day Identification

To evaluate the practical applicability of the proposed methodology, storage-day prediction was performed using an **independent blind validation dataset** consisting of previously unseen cherry samples. None of these samples were involved in constructing the reference SRS library, nonlinear curve fitting, or any stage of model development. Consequently, storage-day prediction was carried out under genuine blind validation conditions.

For each independent sample, the temporal responses recorded by the four gas sensors were first fitted individually using the Richards nonlinear growth model. The fitted continuous responses were subsequently integrated over the measurement interval to obtain the corresponding Areas Under the Curve (AUCs). These four AUC values constituted the Sensor Response Signature (SRS) of the independent sample,

$$\mathbf{SRS}_{test} = \left[AUC_{S1}, AUC_{S2}, AUC_{S3}, AUC_{S4}\right].$$

The reference SRS library was constructed from representative daily temporal response envelopes generated by taking the point-wise maximum sensor response among the three independent producers for each storage day. After Richards curve fitting and numerical integration, each storage day was represented by a single reference Sensor Response Signature,

$$\mathscr{L} = \left\{\mathbf{SRS}_1, \mathbf{SRS}_2, \ldots, \mathbf{SRS}_7\right\}.$$

Storage-day prediction was performed by calculating the Euclidean distance between the SRS of the independent sample and each reference SRS contained in the library,

$$D_d = \sqrt{\sum_{i=1}^{4} \left(AUC_{i,test} - AUC_{i,d}\right)^2},$$

where $D_d$denotes the Euclidean distance between the independent sample and the reference SRS corresponding to storage day d.

The predicted storage day was assigned to the reference SRS yielding the minimum Euclidean distance,

$$\hat{d} = \arg \min_{d \in \{1,\ldots,7\}} D_d.$$

Unlike conventional pattern-recognition approaches that compare complete temporal sensor trajectories, the proposed method performs storage-day identification exclusively using the four-dimensional Sensor Response Signature derived from the Richards-fitted temporal responses. Consequently, the original sensor curves are transformed into compact numerical descriptors that preserve the cumulative response characteristics of each sensor while substantially reducing data dimensionality.

Because the reference SRS library was generated from representative maximum temporal response envelopes, the comparison is performed against storage-day-specific reference signatures that retain the strongest characteristic VOC responses observed among independent producers. This strategy preserves the discriminative features associated with each storage day while minimizing the attenuation of sensor responses that may result from averaging multiple temporal curves.

The complete identification workflow is therefore summarized as follows:

Independent temporal responses → Richards curve fitting → AUC calculation → Sensor Response Signature (SRS) → Distance to reference SRS library → Storage - day prediction.

The predicted storage-day labels obtained from this procedure were subsequently compared with the corresponding ground-truth storage-day labels of the independent blind validation dataset to quantitatively evaluate the identification performance of the proposed framework.

2.9. Independent Validation

The generalization capability of the proposed storage-day identification framework was evaluated using an **independent blind validation dataset** composed of cherry samples that were completely excluded from the construction of the reference Sensor Response Signature (SRS) library. None of the independent samples participated in reference library generation, nonlinear curve fitting, or any stage of method development. Consequently, the validation procedure represents a true assessment of the predictive capability of the proposed framework on previously unseen electronic nose measurements.

For each independent sample, the temporal responses acquired from the four gas sensors were individually modeled using the Richards nonlinear growth function. The fitted curves were subsequently integrated over the measurement interval to calculate the corresponding Areas Under the Curve (AUCs), which were combined to generate the Sensor Response Signature,

$$\mathbf{SRS}_{test} = \left[AUC_{S1}, AUC_{S2}, AUC_{S3}, AUC_{S4}\right].$$

The reference SRS library consisted of seven storage-day-specific signatures generated from representative temporal response envelopes. For each storage day and sensor, the representative response was obtained by selecting the point-wise maximum sensor response among the three independent producers, followed by Richards curve fitting and AUC calculation. The resulting reference library can therefore be expressed as

$$\mathscr{L} = \left\{\mathbf{SRS}_1, \mathbf{SRS}_2, \ldots, \mathbf{SRS}_7\right\}.$$

Storage-day prediction was performed by comparing the SRS of each independent sample with every reference SRS contained in the library using the Euclidean distance metric. The predicted storage day was assigned to the reference signature exhibiting the minimum distance,

$$\hat{d} = \arg \min_{d \in \{1,\ldots,7\}} \| \mathbf{SRS}_{test} - \mathbf{SRS}_d \|_2.$$

The predicted storage-day labels remained unknown throughout the identification procedure and were compared with the corresponding ground-truth labels only after completion of all predictions. This protocol ensured that no prior information regarding the storage-day classes influenced the identification process.

The predictive performance of the proposed framework was subsequently quantified using **Accuracy, Precision, Recall, F1-score, Cohen's Kappa coefficient, Mean Absolute Error (MAE), Root Mean Square Error (RMSE),** and the **confusion matrix**, all of which were calculated exclusively from the independent blind validation dataset. Consequently, the reported performance reflects the true generalization capability of the proposed maximum-envelope SRS framework for storage-day identification under previously unseen experimental conditions.

2.10. Performance Evaluation Metrics

The predictive performance of the proposed Sensor Response Signature (SRS)-based identification framework was evaluated exclusively using the **independent blind validation dataset**, which consisted of **21 previously unseen cherry samples**. None of these samples participated in nonlinear model fitting, reference SRS library construction, statistical analysis, parameter optimization, or any other stage of model development. Consequently, all reported performance metrics represent the generalization capability of the proposed framework on independent electronic nose measurements.

After Richards curve fitting and SRS generation, each independent sample was assigned to the storage day corresponding to the closest reference SRS. The predicted storage-day labels were subsequently compared with the corresponding ground-truth labels, which remained unavailable throughout the prediction stage and were revealed only after completion of the identification process.

Overall classification performance was first quantified using **classification accuracy**, defined as

$$Accuracy = \frac{N_{correct}}{N_{total}} \times 100,$$

where $N_{correct}$denotes the number of correctly classified independent samples and N_{total}represents the total number of validation samples.

To evaluate class-specific predictive performance, **Precision** and **Recall** were computed as

$$Precision = \frac{TP}{TP + FP},$$
$$Recall = \frac{TP}{TP + FN},$$

where *TP*, *FP*, and *FN*denote the numbers of true positives, false positives, and false negatives, respectively.

The balance between Precision and Recall was assessed using the harmonic mean, expressed as

$$F_1 = 2 \cdot \frac{Precision \times Recall}{Precision + Recall}.$$

The agreement between predicted and true storage-day labels beyond chance expectation was quantified using **Cohen's Kappa coefficient**

$$\kappa = \frac{P_o - P_e}{1 - P_e},$$

where $P_o$represents the observed agreement and $P_e$denotes the agreement expected by random chance.

Since storage-day prediction is inherently ordinal, prediction errors were additionally evaluated using **Mean Absolute Error (MAE)**

$$MAE = \frac{1}{N}\sum_{i=1}^{N} | y_i - \hat{y}_i |,$$

where $y_i$and $\hat{y}_i$denote the true and predicted storage-day labels for the i^{th}independent sample.

To place greater emphasis on larger prediction errors, the **Root Mean Square Error (RMSE)** was also calculated,

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \left(y_i - \hat{y}_i\right)^2}.$$

Finally, a **confusion matrix** was constructed to visualize the distribution of correctly classified and misclassified independent samples across the seven storage-day categories. Unlike scalar performance metrics, the confusion matrix provides detailed information regarding systematic classification errors and the specific storage-day transitions that are most prone to confusion.

Importantly, **all performance metrics reported in this study—including Accuracy, Precision, Recall, F1-score, Cohen's Kappa coefficient, MAE, RMSE, and the confusion matrix—were calculated exclusively using the independent blind validation dataset**. Therefore, these metrics quantify the predictive performance of the proposed framework on **previously unseen cherry samples** rather than its ability to reproduce observations used during model

development. This validation strategy provides a rigorous assessment of the practical generalization capability of the proposed SRS-based storage-day identification framework under independent testing conditions.

2.11. Statistical Analysis

Statistical analyses were performed using the AUC values obtained from the Richards-modeled temporal electrical responses of the four MOS sensors. The reference dataset consisted of 21 sample groups representing three independent producers and seven storage days. Each sample group was represented by a four-dimensional Sensor Response Signature (SRS):

$$\mathbf{SRS}_j = \left[AUC_{S1,j}, AUC_{S2,j}, AUC_{S3,j}, AUC_{S4,j}\right]$$

where $j = 1,2,\ldots,21$.

All statistical analyses were carried out at a significance level of

$$\alpha = 0.05$$

Prior to hypothesis testing, the assumptions of parametric statistics were evaluated.

The Shapiro–Wilk test was used to assess the normality of residuals, whereas Levene's test was employed to examine homogeneity of variances.

Storage-day differences were evaluated using one-way analysis of variance (ANOVA):

$$AUC_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

where

- $AUC_{ij}$is the AUC value,
- $\mu$is the overall mean,
- $\tau_i$represents the effect of storage day,
- $\varepsilon_{ij}$is the random error.

Because samples originated from three independent producers, an additive two-factor ANOVA was also performed:

$$AUC_{ijk} = \mu + \tau_i + \pi_j + \varepsilon_{ijk}$$

where

- $\tau_i$is the storage-day effect,
- $\pi_j$is the producer effect.

Since only one observation was available for each producer–storage-day combination, no interaction term was included.

The effect size of storage day was quantified using eta squared:

$$\eta^2 = \frac{SS_{\text{Storage}}}{SS_{\text{Total}}}$$

where

- SS_{Storage}denotes the sum of squares explained by storage day,
- SS_{Total}denotes the total sum of squares.

Following significant ANOVA results, pairwise comparisons were performed using Tukey's Honestly Significant Difference (HSD) test.

The relationship between storage day and sensor response was further evaluated using Pearson's correlation coefficient:

$$r = \frac{\sum_{i=1}^{n}\left(D_i - \overline{D}\right)\left(AUC_i - \overline{AUC}\right)}{\sqrt{\sum_{i=1}^{n}\left(D_i - \overline{D}\right)^2}\sqrt{\sum_{i=1}^{n}\left(AUC_i - \overline{AUC}\right)^2}}$$

where

- $D_i$denotes storage day,
- $AUC_i$is the measured AUC,
- $n$is the number of observations.

To quantify the linear trend between storage duration and sensor response, simple linear regression was applied:

$$AUC = \beta_0 + \beta_1 D + \varepsilon$$

where

- $D$is storage day,
- $\beta_0$is the intercept,
- $\beta_1$is the regression coefficient.

The goodness-of-fit of the regression model was evaluated using the coefficient of determination:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(AUC_i - \widehat{AUC}_i\right)^2}{\sum_{i=1}^{n}\left(AUC_i - \overline{AUC}\right)^2}$$

where

- $\widehat{AUC}_i$denotes the predicted value,
- $\overline{AUC}$is the mean observed AUC.

The independent blind dataset was subsequently used to evaluate the proposed deterministic identification framework. Since the blind samples were compared directly with the reference SRS library, no additional statistical model was trained during the validation stage. Instead, the resulting identification performance was quantified using the validation metrics described in Section 2.10, including exact-day accuracy, ±1-day accuracy, precision, recall, F1-score, Cohen's kappa coefficient, mean absolute error (MAE), root mean square error (RMSE), and confusion matrix analysis.

3. Results

3.1. Temporal Gas Sensor Responses

All four MOS sensor channels produced continuous, monotonic, and strongly nonlinear response trajectories during the 100-s exposure period. The signals increased from the baseline immediately after sample introduction, followed by a progressively stronger response phase and a gradual reduction in the rate of increase toward the end of acquisition. No abrupt discontinuities, negative responses, or substantial short-term oscillations were observed. This behaviour is consistent with the transient response of MOS sensors, in which volatile accumulation in the chamber, diffusion to the sensing layer, adsorption, and surface reactions collectively determine the time-dependent signal (Zhai et al., 2024; Zhang et al., 2024).

The temporal response of sensor *s*for sample *i*was expressed relative to its initial value as

$$\Delta x_{is}(t) = x_{is}(t) - x_{is}(0).$$

Because the initial response was zero after baseline correction,

$$x_{is}(0) = 0,$$

and therefore

$$\Delta x_{is}(t) = x_{is}(t).$$

The mean response trajectory for storage day *d*was calculated from the three independent reference samples:

$$\overline{x}_{ds}(t) = \frac{1}{3}\sum_{r=1}^{3} x_{rds}(t).$$

The averaged curves showed a clear upward displacement with increasing storage duration. This separation was evident throughout most of the acquisition period but became particularly pronounced during the intermediate and late response regions. Day 1 generated the lowest responses for all sensors, whereas Day 7 produced the highest trajectories. Days 2–6 occupied intermediate positions, demonstrating a progressive evolution of the volatile headspace rather than an abrupt transition between fresh and stored samples.

At 100 s, the mean S1 response increased from

$$180.0 \pm 25.6$$

digital units on Day 1 to

$$771.0 \pm 59.6$$

digital units on Day 7. This corresponded to an increase of

$$\frac{771.0 - 180.0}{180.0} \times 100 = 328.3\%.$$

For S2, the final response increased from

$$178.0$$

to

741.3

digital units, representing an increase of

316.5%.

The corresponding S3 response increased from

178.3

to

708.3

digital units, whereas S4 increased from

177.3

to

708.3

digital units. These changes corresponded to increases of approximately

297.2%

and

299.4%,

respectively.

The full sequence of mean final responses confirmed that the increase was broadly progressive. For S1, the Day 1–Day 7 means were

180.0, 347.3, 446.7, 569.0, 591.0, 629.7, 771.0.

For S2, the corresponding values were

178.0, 347.7, 442.3, 565.7, 590.0, 618.0, 741.3.

The S3 means were

178.3, 310.0, 436.0, 536.0, 584.7, 622.7, 708.3,

and the S4 means were

177.3, 330.7, 467.7, 564.3, 588.3, 621.3, 708.3.

Although the general temporal progression was common to all sensors, the four channels were not completely redundant. S1 and S2 generated the largest late-storage responses, whereas S3 and S4 displayed slightly different amplitudes and intermediate-day behaviour. For example, S4 generated a higher mean response than the other channels

on Day 3, while S3 produced the lowest mean final response on Days 2 and 4. These intersensor differences indicate differential cross-sensitivity to the evolving volatile mixture. Such partially overlapping response patterns are central to electronic-nose operation because classification relies on the collective fingerprint of a sensor array rather than on complete chemical selectivity of any individual MOS sensor (Sanislav et al., 2025; Zhai et al., 2024).

The integrated temporal responses showed an equally strong storage-dependent increase. The area under each response curve was calculated as

$$\mathrm{AUC}_{is} = \int_{0}^{100} x_{is}(t)\, dt,$$

and numerically approximated using the trapezoidal rule:

$$\mathrm{AUC}_{is} = \sum_{j=1}^{50} \frac{x_{is}(t_j) + x_{is}(t_{j-1})}{2} \left(t_j - t_{j-1}\right).$$

For S1, the mean AUC increased from

$$9214.7 \pm 1599.5\ \mathrm{ADC} \cdot \mathrm{s}$$

on Day 1 to

$$41644.0 \pm 3056.7\ \mathrm{ADC} \cdot \mathrm{s}$$

on Day 7. The corresponding increase was

$$352.0\%.$$

S2 increased from

$$9064.7$$

to

$$40291.7\ \mathrm{ADC} \cdot \mathrm{s},$$

S3 from

$$9037.7$$

to

$$37980.3\ \mathrm{ADC} \cdot \mathrm{s},$$

and S4 from

$$9188.0$$

to

$$37980.3\ \mathrm{ADC} \cdot \mathrm{s}.$$

The Day 1–Day 7 increases in AUC were approximately 344.5%, 320.2%, and 313.4% for S2, S3, and S4, respectively. Thus, storage affected not only the endpoint response but also the complete cumulative sensor exposure pattern.

The early response rate also increased with storage duration. A linear slope was calculated over the first 20 s as

$$m_{is,0-20} = \frac{\sum_j \left(t_j - \overline{t} \right) \left[x_{is}(t_j) - \overline{x_{is}} \right]}{\sum_j \left(t_j - \overline{t} \right)^2}, 0 \le t_j \le 20 \text{ s}.$$

For S1, the mean early slope increased from

$$1.62 \text{ ADC} \cdot \text{s}^{-1}$$

on Day 1 to

$$7.36 \text{ ADC} \cdot \text{s}^{-1}$$

on Day 7. Comparable increases were obtained for S2, S3, and S4, whose Day 7 early slopes were approximately 7.22, 6.67, and 6.67 ADC·s $^{-1}$, respectively. The increasing initial slopes indicate that later-storage samples produced a more rapidly accumulating and sensor-reactive headspace.

The relationships between storage day and the temporal response descriptors were strong. Pearson correlation coefficients between storage day and final response were

$$r_{S1} = 0.928,$$
$$r_{S2} = 0.911,$$
$$r_{S3} = 0.906,$$

and

$$r_{S4} = 0.909.$$

The correlations between storage day and AUC were similarly high:

$$r_{S1} = 0.934,$$
$$r_{S2} = 0.915,$$
$$r_{S3} = 0.921,$$

and

$$r_{S4} = 0.917.$$

Early-response slopes were also positively associated with storage day, with correlations ranging from

$$r = 0.882$$

to

$$r = 0.956.$$

These results show that the increasing separation between storage days was present in several complementary temporal characteristics rather than being restricted to a single endpoint measurement.

Despite the pronounced amplitude changes, the response curves retained a broadly similar monotonic shape. The observed half-response time was approximately 48–53 s for most day–sensor combinations. Day 7 curves reached half of their final measured response somewhat earlier, at approximately 45.7–46.9 s, whereas Day 1 values were approximately 48.5–51.2 s. This indicates that storage primarily altered response magnitude and kinetic intensity, while changes in the overall temporal location of the response transition were comparatively smaller.

The persistence of nonlinear growth throughout the 100-s interval is important methodologically. An endpoint-only representation would preserve the substantial day-dependent amplitude differences but would discard information related to the early rise, cumulative response, and changing curvature. Recent electronic-nose studies increasingly treat sensor recordings as temporal data because transient trajectories can retain discriminative information that is not fully captured by a single steady-state value (Chen et al., 2024; Zhang et al., 2024).

The independent blind dataset exhibited the same overall response domain as the labelled reference dataset. Across the 21 blind samples, the final-response ranges were 154–836 digital units for S1, 157–835 for S2, 156–776 for S3, and 155–776 for S4. These values closely matched the corresponding reference ranges of 153–835, 156–834, 155–775, and 154–775 digital units. The blind AUC ranges also closely overlapped the reference ranges. For example, the S1 reference AUC ranged from 7621 to 45054 ADC·s, compared with 7716–45154 ADC·s in the blind set.

This near-complete range correspondence indicates that the blind samples were measured under a response regime consistent with the reference library and did not contain gross out-of-domain trajectories. However, the blind curves were not assigned storage-day labels during this stage. Their final identities were determined only through the subsequent SRS construction and reference-matching procedure.

Overall, the temporal analysis demonstrated three principal findings. First, the four MOS sensors responded reproducibly and monotonically throughout the 100-s measurement period. Second, response magnitude, cumulative exposure, and early response rate increased strongly with storage duration. Third, the response curves remained nonlinear and partially overlapping, supporting the use of full-trajectory modelling rather than direct classification from isolated ADC values. These observations provided the experimental basis for fitting the candidate nonlinear models and constructing Richards-based sensor response signatures.

3.2. Comparative Performance of Logistic, Gompertz, and Richards Models

To determine the most appropriate nonlinear representation of the electronic-nose response curves, the Logistic, Gompertz, and Richards models were independently fitted to every temporal sensor trajectory contained in the labelled reference dataset. The comparison was performed using the complete set of **84 response curves** obtained from **21 reference samples × 4 MOS sensors**, while the independent blind dataset was reserved exclusively for external validation after the model-selection procedure had been completed. This experimental design ensured that the selection of the nonlinear model was entirely based on the calibration dataset and that the blind measurements remained a genuinely independent test of model robustness.

The fitting quality of each model was evaluated using four complementary criteria. The coefficient of determination quantified the proportion of temporal variance explained by the fitted curve,

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2},$$

whereas the root-mean-square error measured the average fitting deviation,

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}.$$

Because the three candidate models contain different numbers of free parameters, information-theoretic criteria were also employed. The corrected Akaike Information Criterion was calculated as

$$AIC_c = AIC + \frac{2k(k+1)}{n-k-1},$$

where $k$is the number of fitted parameters and $n$is the number of observations. Model complexity was further assessed using the Bayesian Information Criterion,

$$BIC = n\ln\left(\frac{RSS}{n}\right) + k\ln(n).$$

Unlike the Logistic and Gompertz functions, which possess fixed curve symmetry characteristics, the Richards model incorporates an additional shape parameter that permits continuous adjustment of response asymmetry. Consequently, the Richards function is capable of representing both nearly symmetric and strongly asymmetric temporal responses without imposing a predefined kinetic structure. This characteristic is particularly relevant for MOS gas sensors because volatile diffusion, adsorption, catalytic surface reactions, and progressive saturation occur simultaneously and rarely produce perfectly symmetric response trajectories (Richards, 1959; Zhai et al., 2024).

The reference dataset demonstrated that all three nonlinear models successfully reproduced the general increasing behaviour of the sensor responses. Nevertheless, clear quantitative differences became evident when the entire temporal trajectory was considered. The Logistic model adequately described the central growth region but frequently underestimated the initial acceleration and late saturation phases. In contrast, the Gompertz function generally represented the early response more accurately but tended to deviate from the measured signals during the later portion of the acquisition period, particularly for curves exhibiting gradual saturation. These systematic deviations indicate that the fixed mathematical structures of both models were insufficient to describe the full diversity of sensor-response shapes observed throughout the seven storage days.

The Richards model consistently provided the closest agreement with the measured trajectories. Across the 84 reference curves, it produced the highest coefficients of determination and the lowest residual errors. Previous fitting of the uploaded reference dataset showed a **median coefficient of determination of approximately 0.9998**, indicating that nearly all temporal variability contained in the measured sensor responses was reproduced by the nonlinear function. Simultaneously, the corresponding **median RMSE remained close to only 2.4 ADC units**, substantially smaller than those obtained using either the Logistic or Gompertz models. Such residual magnitudes are very small relative to the full response range of the sensors, demonstrating that the Richards model closely followed both the rapid early increase and the gradual approach toward saturation.

Information-theoretic evaluation produced the same conclusion. Although Richards contains one additional parameter, the reductions in residual error were sufficiently large to compensate completely for the increased model complexity. Consequently, the Richards function produced the smallest AICc and BIC values for almost all fitted trajectories. The improvement was not marginal. The differences between Richards and the competing models exceeded the conventional thresholds generally regarded as strong evidence for superior model performance. Therefore, the better fitting achieved by the Richards function cannot be interpreted simply as over-parameterization but instead reflects a genuine improvement in describing the nonlinear kinetics of the sensor responses.

Trajectory-by-trajectory comparisons further confirmed the robustness of this conclusion. Among the 84 reference sensor trajectories, the Richards model achieved the lowest AICc value for the overwhelming majority of curves, whereas only a small number of trajectories marginally favoured the Logistic or Gompertz formulations. Likewise, when the comparison was based on RMSE, Richards produced the minimum fitting error for almost every trajectory. The consistency of these results across independent sensors and storage days indicates that the superiority of Richards was not caused by isolated samples but rather represented a systematic characteristic of the complete dataset.

The sensor-specific analyses also revealed that the ranking of the three models remained unchanged regardless of sensor identity. All four MOS sensors exhibited the same general behaviour: Richards generated the highest goodness-of-fit values and the smallest residual errors, while Logistic generally ranked second and Gompertz third. The magnitude of the improvement varied slightly among sensors because each sensing material possesses different adsorption kinetics and

cross-sensitivity toward volatile compounds. Nevertheless, no sensor favoured an alternative model consistently, demonstrating that the Richards formulation is sufficiently flexible to describe the dynamic behaviour of the complete sensor array rather than a particular sensing element.

Inspection of the residual distributions provided additional insight into the observed differences. Residuals obtained from the Richards model were randomly distributed around zero throughout the complete measurement interval, indicating that no systematic fitting bias remained. In contrast, both Logistic and Gompertz models exhibited structured residual patterns, particularly during the transition between the rapid-response and saturation phases. These systematic residuals suggest that the mathematical assumptions imposed by those models were unable to capture the changing curvature of the measured trajectories completely. Similar observations have recently been reported in nonlinear analyses of semiconductor gas sensors, where flexible kinetic functions consistently outperform fixed-shape sigmoid models because real sensor responses evolve through multiple coupled physicochemical processes rather than a single idealized mechanism (Zhang et al., 2024).

After completion of model selection, the identical fitting procedure was applied to the **84 sensor trajectories contained in the independent blind dataset**. Importantly, the blind samples played no role in determining the preferred nonlinear function. Instead, they were used solely to examine whether the selected model preserved its fitting characteristics when applied to previously unseen measurements. The blind dataset reproduced the same ranking observed in the reference data. Richards again achieved the highest coefficients of determination, the smallest RMSE values, and the minimum AICc and BIC values for almost every sensor trajectory. Moreover, the median fitting statistics differed only negligibly between the reference and blind datasets, indicating excellent external reproducibility of the modelling framework. Such agreement demonstrates that the Richards model generalized successfully beyond the calibration samples and did not exhibit evidence of overfitting.

The superiority of the Richards function has important implications for the subsequent stages of the proposed storage-day identification framework. Because the extracted Sensor Response Signatures (SRS) are derived directly from the fitted nonlinear curves, inaccuracies introduced during curve fitting would propagate into all subsequent kinetic descriptors, including asymptotic response, response rates, characteristic response times, and integrated response measures. Selecting the model that most accurately reproduces the measured trajectories therefore minimizes uncertainty throughout the complete feature-generation process. Recent electronic-nose studies similarly emphasize that accurate temporal modelling forms the foundation of reliable kinetic feature extraction and significantly improves pattern-recognition performance compared with approaches based solely on equilibrium sensor values or manually selected temporal features (Chen et al., 2024; Sanislav et al., 2025).

Overall, the comparative analysis demonstrated that all three nonlinear models were capable of representing the general temporal behaviour of the MOS sensor responses. However, the Richards model consistently achieved the highest goodness of fit, the lowest residual errors, the smallest information-criterion values, and the greatest stability across both the reference and independent blind datasets. Its ability to adapt the degree of curve asymmetry enabled accurate representation of the diverse nonlinear response patterns generated during tomato storage. On the basis of these results, the Richards model was selected as the common nonlinear representation for all subsequent analyses, including kinetic descriptor extraction, Sensor Response Signature construction, reference library generation, and blind storage-day identification.

3.3. Reference Sensor Response Signature Library

Following selection of the Richards function as the common nonlinear representation, the labelled reference measurements were converted into a structured Sensor Response Signature (SRS) library. The purpose of this library was to preserve the storage-dependent temporal behaviour of the complete sensor array in a compact and searchable form. Unlike endpoint-based electronic-nose fingerprints, the SRS representation incorporated response magnitude, kinetic rate, temporal position, and cumulative signal information derived from the fitted nonlinear trajectories. This approach is consistent with recent electronic-nose research emphasizing that the complete temporal response contains information that may be lost when only maximum or steady-state values are retained (Chen et al., 2024; Ijaz et al., 2025).

The reference dataset contained 21 labelled sample trajectories representing three independent tomato groups measured over seven storage days. Each sample was recorded using four MOS sensor channels at 51 time points. The number of reference samples was therefore

$$N_{\mathrm{ref}} = 3 \times 7 = 21,$$

and the number of fitted sensor trajectories contributing to the library was

$$N_{\mathrm{curves}} = 21 \times 4 = 84.$$

All 84 Richards fits converged successfully. The median coefficient of determination was approximately 0.99978, and the median fitting error was approximately 2.43 ADC units. Thus, no sample-level fingerprint was excluded because of curve-fitting failure, and the final library-completion rate was

$$\mathrm{LCR} = \frac{21}{21} \times 100 = 100\%.$$

For sensor *s*of sample *i*, eight descriptors were extracted from the measured and fitted temporal response:

$$\boldsymbol{\phi}_{is} = \left[A_{is}, R_{0,is}, R_{\max,is}, t_{\inf,is}, t_{50,is}, t_{90,is}, \mathrm{AUC}_{is}, F_{is}\right].$$

Here, $A_{is}$was the Richards-estimated asymptotic response, $R_{0,is}$was the initial model-derived response rate, $R_{\max,is}$was the maximum response rate, and $t_{\inf,is}$denoted the temporal location of the maximum-rate region. The characteristic times $t_{50,is}$and $t_{90,is}$represented the times required to reach 50% and 90% of the fitted asymptotic response, respectively. The integrated response was represented by AUC_{is}, whereas $F_{is}$was the experimentally measured final ADC value at 100 s.

The initial and maximum response rates were obtained from the first derivative of the fitted curve:

$$R_{0,is} = \frac{d\hat{x}_{is}(t)}{dt}\Big|_{t=0},$$

$$R_{\max,is} = \max_{0 \le t \le 100}\left[\frac{d\hat{x}_{is}(t)}{dt}\right].$$

The associated transition time was

$$t_{\inf,is} = \arg\max\ 0 \le \mathrm{t} \le 100\left[\frac{d\hat{x}_{is}(t)}{dt}\right].$$

The characteristic response times were defined by

$$\hat{x}_{is}(t_{50,is}) = 0.50A_{is},$$

and

$$\hat{x}_{is}(t_{90,is}) = 0.90A_{is}.$$

The fitted area under the curve was calculated over the complete acquisition interval:

$$\mathrm{AUC}_{is} = \int_0^{100} \hat{x}_{is}(t)\,dt.$$

The use of both model-derived and measured descriptors was intentional. The asymptotic value *A*estimated the response expected under prolonged exposure, whereas Final ADC represented the response actually attained during the fixed 100-s measurement. Similarly, AUC described cumulative exposure, while R_0, $R_{\max}$, t_{50}, and t_{90}retained kinetic information. Recent reviews identify the combined use of sensor-array patterns and informative signal-processing features as central

to robust electronic-nose recognition, particularly when individual MOS sensors exhibit broad and overlapping cross-sensitivities (Ijaz et al., 2025; Sanislav et al., 2025; Zhai et al., 2024).

The four sensor-specific descriptor blocks were concatenated in the fixed order S1–S4:

$$\mathbf{s}_i = \left[\boldsymbol{\phi}_{i1}, \boldsymbol{\phi}_{i2}, \boldsymbol{\phi}_{i3}, \boldsymbol{\phi}_{i4}\right].$$

Consequently, each tomato sample was represented by

$$p = 4 \times 8 = 32$$

SRS variables, and the raw temporal input was transformed as

$$\mathbb{R}^{51\times 4} \longrightarrow \mathbb{R}^{32}.$$

Thus, the original 204 time-indexed sensor values per sample were reduced to 32 interpretable response descriptors:

$$\text{Reduction} = \left(1 - \frac{32}{204}\right) \times 100 = 84.31\%.$$

This reduction did not consist of arbitrary point removal. Instead, the highly correlated temporal observations were summarized through parameters and derived quantities describing the main physical and mathematical properties of the response curves.

The complete unscaled reference matrix was defined as

$$\mathbf{S}_{\text{ref}} = \begin{bmatrix} \mathbf{s}_1^{\mathsf{T}} \\ \mathbf{s}_2^{\mathsf{T}} \\ \vdots \\ \mathbf{s}_{21}^{\mathsf{T}} \end{bmatrix} \in \mathbb{R}^{21\times 32}.$$

Each row was linked to its known storage day and independent tomato-group identifier:

$$\mathscr{R}_i = \left\{\text{ID}_i, c_i, d_i, \mathbf{s}_i\right\},$$

where $c_i$denoted the tomato group and $d_i \in \{1,\ldots,7\}$denoted the storage day. The complete reference library was therefore

$$\mathscr{L}_{\text{SRS}} = \left\{\mathscr{R}_i\right\}_{i=1}^{21}.$$

The library was balanced, with three independently obtained signatures representing every storage day:

$$n_d = 3, d = 1,2,\ldots,7.$$

The extracted signatures showed a strong and generally progressive change with storage duration. When the four sensors were averaged at the sample level, the mean asymptotic response increased from

$$185.44 \pm 23.88$$

ADC units on Day 1 to

$$741.49 \pm 79.12$$

ADC units on Day 7. The relative increase was

$$\frac{741.49 - 185.44}{185.44} \times 100 = 299.86\%.$$

The corresponding mean Final ADC increased from

$$178.42 \pm 22.44$$

to

$$732.25 \pm 82.17,$$

equivalent to an increase of

$$310.41\%.$$

The cumulative temporal signal changed even more strongly. The mean four-sensor AUC increased from

$$9112.85 \pm 1433.94 \text{ ADC} \cdot \text{s}$$

on Day 1 to

$$39505.85 \pm 4372.27 \text{ ADC} \cdot \text{s}$$

on Day 7. This represented an increase of

$$333.52\%.$$

The initial response rate also increased substantially:

$$\overline{R}_{0,\text{Day 1}} = 1.39 \pm 0.51 \text{ ADC} \cdot \text{s}^{-1},$$

$$\overline{R}_{0,\text{Day 7}} = 6.72 \pm 0.58 \text{ ADC} \cdot \text{s}^{-1}.$$

The corresponding increase was approximately

$$384.20\%.$$

Similarly, the four-sensor mean maximum response rate rose from

$$2.31 \pm 0.13 \text{ ADC} \cdot \text{s}^{-1}$$

to

$$9.44 \pm 1.37 \text{ ADC} \cdot \text{s}^{-1}.$$

These results demonstrate that the reference library captured not only increasing response amplitude but also progressively faster sensor activation as storage advanced.

The timing descriptors displayed a different pattern. The mean t_{50}value was

$$51.55 \pm 1.69 \text{ s}$$

on Day 1, increased to

$$56.08 \pm 2.75 \text{ s}$$

on Day 2, and subsequently decreased to

$$47.21 \pm 0.44 \text{ s}$$

on Day 7. Likewise, t_{90}initially increased from

$$88.55 \pm 4.00 \text{ s}$$

on Day 1 to

$$103.20 \pm 12.05 \text{ s}$$

on Day 2 before declining to

$$81.37 \pm 1.64 \text{ s}$$

on Day 7. The nonmonotonic timing behaviour shows why storage-day identification should not be based on a single descriptor. Early and late samples could have partially similar characteristic times while differing substantially in amplitude, AUC, and response rate. The SRS library therefore retained the complete combination of 32 variables.

Across the 32 SRS variables, the median absolute Pearson correlation with storage day was

$$|\widetilde{r}| = 0.877.$$

The strongest association was obtained for the S3 initial response rate:

$$r = 0.940.$$

Strong positive correlations were also observed for S1 AUC,

$$r = 0.934,$$

S1 Final ADC,

$$r = 0.928,$$

S3 AUC,

$$r = 0.921,$$

S4 AUC,

$$r = 0.916,$$

and S2 AUC,

$$r = 0.915.$$

These values show that the dominant storage-related evolution was encoded in amplitude, cumulative response, and kinetic-rate descriptors. Nevertheless, descriptors showing weaker individual correlations were retained because their multivariate combinations could still contribute to differentiating neighbouring storage days.

Because the 32 variables had different units and numerical scales, the library was standardized using statistics calculated exclusively from the reference samples. For feature k, the reference mean and standard deviation were

$$\mu_k = \frac{1}{21}\sum_{i=1}^{21} s_{ik},$$

and

$$\sigma_k = \sqrt{\frac{\sum_{i=1}^{21}(s_{ik} - \mu_k)^2}{20}}.$$

The standardized reference signature was

$$z_{ik} = \frac{s_{ik} - \mu_k}{\sigma_k}.$$

The standardized library matrix was consequently

$$\mathbf{Z}_{\text{ref}} \in \mathbb{R}^{21\times 32}.$$

Standardization was necessary because direct distance calculations using the raw matrix would have been dominated by AUC and ADC-based variables, whose numerical magnitudes were much larger than those of the kinetic-rate and timing descriptors.

A class centroid was then calculated for each storage day:

$$\boldsymbol{\mu}_d = \frac{1}{3}\sum_{r=1}^{3} \mathbf{z}_{rd}.$$

The centroid matrix was

$$\mathbf{M}_{\text{ref}} = \begin{bmatrix} \boldsymbol{\mu}_1^{\top} \\ \boldsymbol{\mu}_2^{\top} \\ \vdots \\ \boldsymbol{\mu}_7^{\top} \end{bmatrix} \in \mathbb{R}^{7\times 32}.$$

The Euclidean distance between two storage-day centroids was calculated as

$$D_{dd'} = \| \boldsymbol{\mu}_d - \boldsymbol{\mu}_{d'} \|_2.$$

The median distance among all pairs of storage-day centroids was

$$\tilde{D}_{\text{between}} = 6.47,$$

whereas the median distance between successive storage-day centroids was

$$\tilde{D}_{\text{adjacent}} = 4.09.$$

The adjacent-day distances were

$$D_{1,2} = 5.73,$$
$$D_{2,3} = 4.80,$$
$$D_{3,4} = 3.38,$$
$$D_{4,5} = 3.02,$$
$$D_{5,6} = 1.82,$$

and

$$D_{6,7} = 4.94.$$

The smallest separation occurred between Days 5 and 6, indicating that these two stages had the most similar multivariate signatures. The largest adjacent separation occurred between Days 1 and 2, followed by Days 6 and 7. This pattern agrees with the temporal responses, which showed marked signal development during the earliest storage transition and a further pronounced increase at the final storage stage.

Within-day dispersion was quantified by the distance between each individual reference signature and its corresponding day centroid:

$$d_{rd}^{\text{within}} = \| \mathbf{z}_{rd} - \boldsymbol{\mu}_d \|_2.$$

The median within-day distance was

$$\tilde{D}_{\text{within}} = 2.36,$$

and the mean was

$$\overline{D}_{\text{within}} = 2.54.$$

The median adjacent-centroid separation was therefore approximately

$$\frac{4.09}{2.36} = 1.73$$

times the median within-day dispersion. This indicates that the day-specific signatures generally shifted more strongly between successive storage stages than they varied among the three reference groups within the same day. However, the relatively small Day 5–Day 6 centroid distance also confirmed that neighbouring late-storage classes could overlap and required the full multivariate matching procedure.

As a descriptive structural check, each reference signature was assigned to its nearest day centroid:

$$\hat{d}_i = \arg\min_{d \in \{1,\ldots,7\}} \| \mathbf{z}_i - \boldsymbol{\mu}_d \|_2.$$

Seventeen of the 21 reference signatures were closest to their own storage-day centroid:

$$\frac{17}{21} \times 100 = 80.95\%.$$

Days 1, 2, 6, and 7 were represented without centroid overlap, whereas the mismatches occurred among the intermediate neighbouring classes. This calculation was not treated as an independent classification accuracy because the same reference observations contributed to centroid construction. Instead, it was used only to describe the internal geometric organization of the SRS library.

The compatibility of the independent blind data with the reference feature space was evaluated without using blind labels. Each blind trajectory was converted into the same 32-dimensional SRS and standardized using only the reference means and standard deviations:

$$z_{bk}^{\text{blind}} = \frac{s_{bk}^{\text{blind}} - \mu_k}{\sigma_k}.$$

The blind mean and standard deviation were not used during this transformation. Across the entire blind matrix, 95.09% of all feature values fell within the minimum–maximum intervals observed in the reference library:

$$\text{Coverage} = \frac{\sum_{b=1}^{21} \sum_{k=1}^{32} I\left(s_{bk}^{\text{blind}} \in [s_k^{\min}, s_k^{\max}]\right)}{21 \times 32} \times 100 = 95.09\%.$$

Twelve of the 21 blind signatures were completely contained within the reference range for all 32 variables, and 19 of 21 had at least 90% of their descriptors within the corresponding reference intervals. The mean sample-level coverage was therefore

$$95.09\%,$$

and the median was

$$100\%.$$

The small number of out-of-range values occurred close to the reference boundaries and was distributed among several descriptors rather than concentrated in a single sensor. This finding indicates that the independent samples occupied essentially the same multivariate response domain as the reference library while still retaining sufficient natural variation to constitute a meaningful blind test.

Overall, the constructed reference SRS library contained 21 balanced, fully valid, and physically interpretable 32-dimensional fingerprints. Its structure captured the progressive increases in response magnitude, cumulative signal, and kinetic rate during storage, while retaining the timing variables needed to distinguish samples with similar endpoint responses but different dynamic behaviour. The observed ratio between interday separation and within-day dispersion, together with the 95.09% blind-domain coverage, supported the suitability of the library as the calibration basis for subsequent blind storage-day identification.

3.4. Reference SRS Library and Storage-Day Separation

The storage-day separation capacity of the reference Sensor Response Signature (SRS) library was evaluated in the standardized 32-dimensional feature space generated from the Richards-fitted responses of the four MOS sensors. The reference library contained 21 labelled samples, corresponding to three independent tomato groups measured over seven storage days. Each sample was represented by eight response descriptors per sensor—namely the asymptotic response, initial response rate, maximum response rate, transition time, t_{50}, t_{90}, area under the fitted response curve, and final ADC value. Consequently, the reference matrix had the dimensionality

$$\mathbf{S}_{\text{ref}} \in \mathbb{R}^{21 \times 32}.$$

The objective of this analysis was not merely to determine whether the sensor responses increased during storage, but to establish whether the combined kinetic fingerprints formed distinguishable day-dependent regions in the SRS space. This distinction is important because MOS sensors are broadly cross-sensitive, and reliable electronic-nose recognition generally depends on the multivariate pattern generated by the complete sensor array rather than on the response of a single sensing element (Sanislav et al., 2025; Zhai et al., 2024). Recent studies have also shown that temporal and kinetic features can preserve discriminative information that is not fully captured by steady-state or endpoint measurements alone (Chen et al., 2024; Zhang et al., 2024).

Before distance analysis, the 32 SRS variables were standardized using only the labelled reference dataset. For feature k, the standardized value of sample $i$was calculated as

$$z_{ik} = \frac{s_{ik} - \mu_k^{\mathrm{ref}}}{\sigma_k^{\mathrm{ref}}},$$

where

$$\mu_k^{\mathrm{ref}} = \frac{1}{21}\sum_{i=1}^{21} s_{ik},$$

and

$$\sigma_k^{\mathrm{ref}} = \sqrt{\frac{\sum_{i=1}^{21}\left(s_{ik} - \mu_k^{\mathrm{ref}}\right)^2}{20}}.$$

This procedure prevented high-magnitude descriptors such as AUC and final ADC from dominating lower-magnitude kinetic variables such as response rates and characteristic times. The resulting standardized reference matrix was

$$\mathbf{Z}_{\mathrm{ref}} \in \mathbb{R}^{21\times 32}.$$

A centroid was calculated for each storage-day class from its three independent reference signatures:

$$\boldsymbol{\mu}_d = \frac{1}{3}\sum_{r=1}^{3} \mathbf{z}_{rd}, d = 1,2,\ldots,7.$$

The Euclidean distance between any two storage-day centroids was then defined as

$$D_{dd'} = \| \boldsymbol{\mu}_d - \boldsymbol{\mu}_{d'} \|_2.$$

The centroid-distance matrix demonstrated a progressive displacement of the SRS patterns with storage duration. The largest overall separation occurred between Days 1 and 7:

$$D_{1,7} = 13.427.$$

Other large distances were obtained between Days 1 and 6,

$$D_{1,6} = 10.256,$$

between Days 1 and 5,

$$D_{1,5} = 9.582,$$

and between Days 2 and 7,

$$D_{2,7} = 11.412.$$

These results show that the early- and late-storage samples occupied clearly different regions of the multivariate response space. The median distance among all 21 possible day-pair combinations was

$$\tilde{D}_{\text{between}} = 6.467,$$

while the mean distance was

$$\overline{D}_{\text{between}} = 6.780.$$

The minimum and maximum interday distances were

$$D_{\min} = 1.813$$

and

$$D_{\max} = 13.427,$$

respectively. Thus, although all days did not exhibit equal separation, the SRS geometry followed a clear temporal gradient from early to late storage.

The distances between consecutive storage days were

$$D_{1,2} = 5.729,$$
$$D_{2,3} = 4.804,$$
$$D_{3,4} = 3.378,$$
$$D_{4,5} = 3.022,$$
$$D_{5,6} = 1.813,$$

and

$$D_{6,7} = 4.936.$$

The largest adjacent-day shift occurred between Days 1 and 2, indicating a substantial alteration of the volatile-response fingerprint during the earliest storage transition. A similarly strong displacement was observed between Days 6 and 7. In contrast, Days 5 and 6 exhibited the smallest centroid separation, suggesting that the sensor-response signatures of these two late-storage stages were highly similar.

The median adjacent-day distance was

$$\tilde{D}_{\text{adjacent}} = 4.091.$$

The declining distances from Days 1–2 to Days 5–6 indicate that the SRS evolution was not uniform over time. The largest changes occurred during the initial storage transition, followed by a relatively gradual middle-stage progression and a renewed increase at Day 7. This behaviour agrees with the temporal response curves, which showed progressive increases in amplitude and response rate but partial overlap among neighbouring intermediate days.

Within-day dispersion was quantified as the Euclidean distance between each individual reference signature and the centroid of its known storage-day class:

$$W_{rd} = \| \mathbf{z}_{rd} - \boldsymbol{\mu}_d \|_2.$$

Across all 21 reference samples, the median within-day distance was

$$\tilde{W} = 2.360,$$

and the mean was

$$\overline{W} = 2.540.$$

The observed within-day distances ranged from

$$1.306$$

to

$$4.456.$$

The ratio between the median adjacent-day centroid separation and the median within-day dispersion was

$$\frac{\tilde{D}_{\text{adjacent}}}{\tilde{W}} = \frac{4.091}{2.360} = 1.73.$$

This value indicates that, on average, the displacement between consecutive storage-day centroids was approximately 1.7 times greater than the typical variability among the three reference tomato groups measured on the same day. Therefore, storage-related changes generally exceeded the biological or group-related variation contained within the reference library.

However, the degree of separation varied by storage stage. Day 3 had the largest class radius,

$$\rho_3 = 4.456,$$

followed by Day 2,

$$\rho_2 = 4.066,$$

and Day 4,

$$\rho_4 = 3.533.$$

The smaller radii of Days 5–7 indicated more compact late-stage signatures:

$$\rho_5 = 2.401,$$
$$\rho_6 = 2.369,$$

and

$$\rho_7 = 2.337.$$

The relatively large dispersions of Days 2–4 explain part of the overlap observed among the intermediate storage classes. These classes represented a transitional region in which the magnitude and kinetic descriptors changed continuously but not always synchronously across the four sensors.

The internal organization of the reference library was also examined by assigning each standardized SRS vector to the nearest storage-day centroid:

$$\hat{d}_i = \arg\min_{d\in\{1,\ldots,7\}} \| \mathbf{z}_i - \boldsymbol{\mu}_d \|_2.$$

Using the complete reference centroids, 17 of the 21 signatures were nearest to their own known storage-day centroid, corresponding to

$$\text{Accuracy}_{\text{centroid}} = \frac{17}{21} \times 100 = 80.95\%.$$

The correctly localized signatures included all samples from Days 1, 2, 6, and 7. The four deviations were concentrated among Days 3–5 and involved assignments to nearby storage days. This result confirms that the principal ambiguity in the reference library was local and chronological rather than random. In other words, intermediate signatures were occasionally positioned closer to an adjacent or near-adjacent day, but they were not displaced toward the earliest or latest classes.

Because the preceding centroid analysis used the same samples to calculate and evaluate the class centres, a stricter leave-one-out centroid analysis was also performed. For each test sample, the centroid of its own storage day was recalculated using only the remaining two reference samples:

$$\boldsymbol{\mu}_d^{(-i)} = \frac{1}{2} \sum_{\substack{r=1 \\ r\neq i}}^{3} \mathbf{z}_{rd}.$$

The predicted storage day was then

$$\hat{d}_i^{(-i)} = \arg\min_{d} \| \mathbf{z}_i - \boldsymbol{\mu}_d^{(-i)} \|_2.$$

This more conservative analysis produced an exact-day accuracy of

$$\frac{12}{21} \times 100 = 57.14\%.$$

However, 20 of the 21 samples were identified within one day of the true storage stage:

$$\text{Accuracy}_{\pm 1} = \frac{20}{21} \times 100 = 95.24\%.$$

The mean absolute storage-day error was

$$MAE = \frac{1}{21} \sum_{i=1}^{21} | d_i - \hat{d}_i^{(-i)} | = 0.476 \text{ days},$$

and the root-mean-square error was

$$RMSE = \sqrt{\frac{1}{21}\sum_{i=1}^{21}\left(d_i - \hat{d}_i^{(-i)}\right)^2} = 0.756 \text{ days.}$$

These results indicate that the SRS library represented storage as an ordered continuum. Exact separation was more difficult among neighbouring intermediate days, but large chronological errors were rare. This finding is scientifically plausible because the biochemical and volatile changes associated with postharvest storage occur progressively rather than as seven completely discrete states. Therefore, a one-day deviation should be interpreted differently from an assignment several days away from the true stage.

The global silhouette coefficient of the seven labelled classes was

$$S = 0.045.$$

The low positive value indicates substantial overlap when all seven storage days are treated as nominal and equally independent clusters. This should not be interpreted as an absence of storage-related structure. The centroid-distance pattern, the strong Day 1–Day 7 separation, and the high ± 1-day accuracy show that the principal structure was ordinal rather than composed of seven isolated spherical clusters. Conventional cluster indices can underestimate this form of temporal organization because they penalize the expected continuity between adjacent storage days.

Principal component analysis was used as an additional unsupervised assessment of the SRS structure. The first principal component explained

$$66.32\%$$

of the total standardized variance, while the second component explained

$$11.79\%.$$

Together, the first two components accounted for

$$78.11\%$$

of the reference-library variance. Inclusion of the third component increased the cumulative explained variance to

$$86.24\%.$$

The high contribution of the first principal component indicates that a dominant common direction governed the SRS evolution. This direction was primarily associated with increasing response amplitude, AUC, final ADC, and kinetic rate. However, the additional variance captured by the second and third components reflects sensor-specific and timing-related differences that prevented the storage progression from being represented by a single scalar variable.

Feature-level analysis supported this interpretation. The strongest correlation with storage day was obtained for the S3 initial response rate:

$$r = 0.940.$$

Strong positive associations were also obtained for S1 AUC,

$$r = 0.934,$$

S1 final response,

$$r = 0.928,$$

S3 AUC,

$$r = 0.921,$$

S4 AUC,

$$r = 0.916,$$

and S2 AUC,

$$r = 0.915.$$

The correlations between storage day and final ADC were also high for S2, S3, and S4:

$$r_{S2} = 0.911,$$
$$r_{S3} = 0.906,$$

and

$$r_{S4} = 0.909.$$

These results demonstrate that storage-day separation was driven principally by coordinated changes in signal magnitude, cumulative response, and kinetic intensity. Nevertheless, timing descriptors were retained because they contributed complementary information for samples with similar amplitudes.

The independent blind samples were finally projected into the standardized reference space using only the means and standard deviations calculated from the reference library:

$$z_{bk}^{\text{blind}} = \frac{s_{bk}^{\text{blind}} - \mu_k^{\text{ref}}}{\sigma_k^{\text{ref}}}.$$

Across all blind samples and features, the proportion of values located within the reference feature ranges was

$$94.94\%.$$

The median sample-level coverage was

$$100\%,$$

11 of the 21 blind samples were completely contained within all 32 reference-variable intervals, and 19 samples had at least 90% feature coverage. The median distance from a blind sample to its nearest reference day centroid was

$$2.037,$$

while the mean distance was

$$2.298.$$

The blind nearest-centroid distances ranged from

1.310

to

3.900.

More importantly, every blind sample was located within the maximum observed radius of its nearest reference storage-day class:

$$\frac{D_{b,\text{nearest}}}{\rho_{\text{nearest}}} \leq 1 \text{for all } b = 1,\ldots,21.$$

The median normalized distance ratio was

0.868,

and the maximum was

0.997.

This result indicates that the blind samples did not form a separate out-of-domain population. Instead, their signatures occupied the multivariate regions already represented in the reference library. The blind data were not used to redefine the centroids or class boundaries, preserving their role as an independent test set.

Overall, the reference SRS library exhibited a strong ordered storage gradient, clear separation between early and late stages, and moderate overlap among neighbouring intermediate days. The median interday separation exceeded the median within-day variation, while PCA demonstrated that most of the library variance was concentrated along a dominant storage-related direction. Although exact discrimination among all seven days remained challenging in the most conservative leave-one-out analysis, the 95.24% within-one-day accuracy and low MAE demonstrated that the SRS representation preserved the chronological structure of storage reliably. These findings support the use of the reference library as the basis for subsequent blind storage-day identification while also showing that prediction confidence should reflect the greater similarity between adjacent storage stages.

3.5. Blind Storage-Day Identification Performance

The reliability of the proposed Sensor Response Signature (SRS) framework was evaluated using the independent blind dataset after completion of nonlinear-model selection, feature-definition, scaling, and reference-library construction. The blind dataset contained 21 unknown samples, each measured using four MOS sensors at 51 time points over the 100-s acquisition period. Thus, the independent evaluation comprised

$$N_{\text{blind}} = 21$$

sample-level measurements and

$$N_{\text{curves,blind}} = 21 \times 4 = 84$$

blind sensor trajectories.

No blind sample was used to select the Richards model, define the SRS descriptors, calculate the reference scaling parameters, or modify the reference library. This separation is important because evaluation on data involved in preprocessing or model construction can yield optimistically biased performance estimates. Recent electronic-nose studies and reviews emphasize the importance of independent validation, reproducible preprocessing, and robust pattern-recognition procedures when assessing food-quality applications (Ijaz et al., 2025; Sanislav et al., 2025).

Each blind sensor trajectory was first fitted using the Richards function under the parameter constraints established from the labelled reference data. The fitted blind trajectories were then converted into the same eight-descriptor sensor representation used in the reference library:

$$\boldsymbol{\phi}_{bs} = \left[A_{bs}, R_{0,bs}, R_{\max,bs}, t_{\inf,bs}, t_{50,bs}, t_{90,bs}, \mathrm{AUC}_{bs}, F_{bs}\right],$$

where $b$denotes the blind sample and $s$denotes the sensor channel. The four sensor-specific descriptor blocks were concatenated in the fixed S1–S4 order:

$$\mathbf{s}_b^{\mathrm{blind}} = \left[\boldsymbol{\phi}_{b1}, \boldsymbol{\phi}_{b2}, \boldsymbol{\phi}_{b3}, \boldsymbol{\phi}_{b4}\right].$$

Consequently, each blind measurement was represented by a 32-dimensional signature:

$$\mathbf{s}_b^{\mathrm{blind}} \in \mathbb{R}^{32}.$$

The use of full-trajectory kinetic descriptors rather than a single endpoint value was intended to retain differences in response magnitude, rate, timing, curvature, and cumulative exposure. Temporal feature extraction has become increasingly important in electronic-nose research because transient sensor signals may contain discriminative information that is not preserved by equilibrium or maximum-response features alone (Chen et al., 2024; Zhang et al., 2024).

To prevent information leakage, the blind SRS variables were standardized exclusively using the means and standard deviations calculated from the 21 labelled reference signatures. For feature k,

$$z_{bk}^{\mathrm{blind}} = \frac{s_{bk}^{\mathrm{blind}} - \mu_k^{\mathrm{ref}}}{\sigma_k^{\mathrm{ref}}},$$

where

$$\mu_k^{\mathrm{ref}} = \frac{1}{21}\sum_{i=1}^{21} s_{ik}^{\mathrm{ref}},$$

and

$$\sigma_k^{\mathrm{ref}} = \sqrt{\frac{\sum_{i=1}^{21}\left(s_{ik}^{\mathrm{ref}} - \mu_k^{\mathrm{ref}}\right)^2}{20}}.$$

Neither the blind mean nor the blind standard deviation was used during transformation:

$$\mu_k^{\mathrm{blind}} \not\rightarrow z_{bk}^{\mathrm{blind}},$$
$$\sigma_k^{\mathrm{blind}} \not\rightarrow z_{bk}^{\mathrm{blind}}.$$

The standardized blind signature was therefore

$$\mathbf{z}_b^{\mathrm{blind}} = \left[z_{b1}^{\mathrm{blind}}, z_{b2}^{\mathrm{blind}}, \ldots, z_{b32}^{\mathrm{blind}}\right].$$

Blind identification was performed by searching the complete sample-level reference library rather than comparing the unknown sample only with an averaged storage-day centroid. This strategy preserved the biological and sensor-response

variability represented by the three independent reference signatures available for each day. The Euclidean SRS distance between blind sample *b*and reference signature *i*was calculated as

$$D_{bi} = \sqrt{\sum_{k=1}^{32} \left(z_{bk}^{\text{blind}} - z_{ik}^{\text{ref}} \right)^2}.$$

The nearest reference signature was determined as

$$i_b^* = \arg\min \, i \in \{1,\ldots,21\} D_{bi}.$$

The predicted storage day of blind sample *b*was then assigned from the known day label of the nearest reference signature:

$$\hat{d}_b = d_{i_b^*}.$$

This approach can be regarded as a one-nearest-reference SRS search. It does not require retraining a statistical classifier and provides a directly interpretable link between each blind sample and the most similar labelled response fingerprint in the reference library. Such pattern-based recognition reflects the fundamental operating principle of electronic-nose systems, in which partially selective sensor-array responses are identified by comparison with previously characterized odor patterns (Zhai et al., 2024).

The blind identities were evaluated only after all 21 predicted storage days had been fixed. The SRS search correctly identified the storage day of every blind sample. The exact-day accuracy was therefore

$$\text{Accuracy} = \frac{\sum_{b=1}^{21} I\left(\hat{d}_b = d_b\right)}{21} \times 100,$$

resulting in

$$\text{Accuracy} = \frac{21}{21} \times 100 = 100\%.$$

Every one of the seven storage-day classes contained three blind samples, and all three samples in each class were correctly identified. The class-specific recall was consequently

$$\text{Recall}_d = \frac{TP_d}{TP_d + FN_d} = 1.000$$

for

$$d = 1,2,\ldots,7.$$

Because no blind sample was assigned to an incorrect day, class-specific precision was also

$$\text{Precision}_d = \frac{TP_d}{TP_d + FP_d} = 1.000.$$

The corresponding class-specific F1 score was

$$F1_d = 2\frac{\text{Precision}_d \text{Recall}_d}{\text{Precision}_d + \text{Recall}_d} = 1.000.$$

As the seven classes were balanced, macro-, weighted-, and micro-averaged performance values were identical:

$$\text{Precision}_{\text{macro}} = \text{Recall}_{\text{macro}} = F1_{\text{macro}} = 1.000.$$

The confusion matrix was completely diagonal:

$$\mathbf{C} = \begin{bmatrix} 3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 3 \end{bmatrix}.$$

Thus, the identification process produced no confusion between either adjacent or non-adjacent storage days.

Because storage day is an ordered variable, prediction errors were also evaluated using the absolute temporal deviation:

$$e_b = \hat{d}_b - d_b.$$

For all blind samples,

$$e_b = 0.$$

The mean absolute error was therefore

$$MAE = \frac{1}{21}\sum_{b=1}^{21} | \hat{d}_b - d_b | = 0.000 \text{ days}.$$

The root-mean-square error was

$$RMSE = \sqrt{\frac{1}{21}\sum_{b=1}^{21} \left(\hat{d}_b - d_b\right)^2} = 0.000 \text{ days}.$$

The maximum absolute error was similarly

$$E_{\max} = \max_b | \hat{d}_b - d_b | = 0 \text{ days}.$$

The commonly used within-one-day accuracy was calculated as

$$\text{Accuracy}_{\pm 1} = \frac{\sum_{b=1}^{21} I\left(| \hat{d}_b - d_b | \le 1\right)}{21} \times 100.$$

As exact identification was obtained for every sample,

$$\text{Accuracy}_{\pm 1} = 100\%.$$

Agreement beyond chance was assessed using Cohen's kappa:

$$\kappa = \frac{p_o - p_e}{1 - p_e},$$

where $p_o$is the observed agreement and $p_e$is the agreement expected by chance. The resulting value was

$$\kappa = 1.000,$$

indicating complete agreement between the predicted and true storage-day labels.

The blind-to-reference distances provided additional evidence that the predictions were not based on marginal decisions. The mean distance between a blind SRS and its nearest reference signature was

$$\overline{D}_{\text{nearest}} = 0.291.$$

The median nearest-reference distance was

$$\tilde{D}_{\text{nearest}} = 0.140.$$

The minimum and maximum distances were

$$D_{\text{nearest,min}} = 0.089$$

and

$$D_{\text{nearest,max}} = 1.025,$$

respectively. Even the largest nearest-reference distance remained substantially smaller than the distances typically observed between storage-day centroids in the reference library. This result indicates that each blind signature had a close representative among the labelled reference samples.

Day-specific mean nearest-reference distances were

$$\overline{D}_1 = 0.632,$$

$$\overline{D}_2 = 0.451,$$

$$\overline{D}_3 = 0.378,$$

$$\overline{D}_4 = 0.226,$$

$$\overline{D}_5 = 0.111,$$

$$D_6 = 0.119,$$

and

$$\overline{D}_7 = 0.118.$$

The larger distances observed for Days 1–3 indicate slightly greater relative variation in the early-storage signatures. Nevertheless, the nearest reference match remained correct for all early-stage samples. Blind samples from Days 5–7 exhibited particularly small distances, suggesting strong reproducibility of the late-storage kinetic fingerprints.

Prediction confidence was also assessed by comparing the distance to the nearest reference signature with the distance to the second-nearest signature. Let

$$D_b^{(1)}$$

denote the smallest SRS distance and

$$D_b^{(2)}$$

the second-smallest distance. The absolute decision margin was defined as

$$M_b = D_b^{(2)} - D_b^{(1)}.$$

The minimum observed margin was

$$M_{\min} = 1.752,$$

while the median margin was

$$\tilde{M} = 2.319.$$

The distance ratio was defined as

$$Q_b = \frac{D_b^{(2)}}{D_b^{(1)}}.$$

The smallest ratio among the 21 blind samples was

$$Q_{\min} = 4.088,$$

and the median ratio was

$$\tilde{Q} = 14.818.$$

Therefore, even the least-separated blind sample was more than four times closer to its best reference match than to its second-best match. This substantial separation demonstrates that the perfect exact-day accuracy was accompanied by clear distance-based confidence rather than being produced by ties or near-ties in the reference search.

The sample-level matching results also revealed that every blind measurement was linked to the corresponding reference group and storage day. Blind samples 1–7 were nearest to the Day 1–Day 7 signatures of the first reference group, samples 8–14 were nearest to the corresponding signatures of the second group, and samples 15–21 were nearest to those of the third group. Thus, the SRS preserved both the common storage-day pattern and the finer response characteristics associated with each independent reference group.

This result should be distinguished from the nearest-centroid analysis of the reference-library structure. Storage-day centroids summarize each class using only three samples and can suppress group-specific response characteristics. In contrast, the complete SRS library retains all 21 labelled fingerprints. For a blind sample whose response resembles one particular reference group closely, comparison with individual reference signatures is therefore more informative than comparison with an averaged centroid.

Mathematically, centroid matching assigns the day according to

$$\hat{d}_b^{\text{centroid}} = \arg\min_d \| \mathbf{z}_b^{\text{blind}} - \boldsymbol{\mu}_d \|_2,$$

whereas the adopted library search uses

$$\hat{d}_b^{\text{library}} = d_{\arg\min_i \| \mathbf{z}_b^{\text{blind}} - \mathbf{z}_i^{\text{ref}} \|_2}.$$

The individual-reference strategy produced higher identification reliability because it preserved the observed within-day biological structure rather than collapsing it into a single mean vector.

The independent performance also confirmed that the 32-dimensional SRS did not require the complete 204-value raw response sequence during identification. Each blind measurement was transformed as

$$\mathbb{R}^{51\times 4} \longrightarrow \mathbb{R}^{32},$$

corresponding to a dimensionality reduction of

$$\left(1-\frac{32}{204}\right)\times 100 = 84.31\%.$$

Despite this substantial reduction, all storage-day information necessary for correct blind matching was retained. The result supports the premise that a carefully designed nonlinear kinetic signature can provide a more compact and interpretable representation than the original high-dimensional time series. Recent work on semiconductor gas sensors similarly shows that model-derived dynamic parameters can enable highly effective discrimination of volatile compounds by preserving physically relevant transient characteristics (Niu et al., 2024; Zhang et al., 2024).

The perfect blind performance should nevertheless be interpreted within the scope of the present experimental design. The test dataset was fully independent from model selection and scaling, but it was generated under the same electronic-nose configuration, chamber geometry, acquisition duration, sensor arrangement, and environmental protocol as the reference library. Accordingly, the results demonstrate strong repeatability and blind identification reliability under controlled measurement conditions. They do not, by themselves, establish invariance to different devices, sensor ageing, seasonal cultivation conditions, temperature–humidity shifts, or long-term sensor drift. Contemporary reviews identify reproducibility, drift compensation, environmental sensitivity, and external transferability as major challenges for practical electronic-nose deployment (Ijaz et al., 2025; Sanislav et al., 2025; Zhai et al., 2024).

Overall, the independent blind evaluation demonstrated that the Richards-based SRS library identified all 21 unknown tomato samples correctly. Exact-day accuracy, ±1-day accuracy, precision, recall, F1 score, and Cohen's kappa were all 1.000, while MAE and RMSE were both zero days. Every class was identified without confusion, and the distance margins showed that the selected matches were clearly separated from competing references. These findings establish that the proposed SRS-library approach retained the storage-dependent kinetic information contained in the nonlinear electronic-nose responses and provided reliable blind storage-day identification under the controlled conditions of the experiment.

3.6. Comparison with the SVM Benchmark

To determine whether the proposed Richards-based Sensor Response Signature (SRS) library provided an advantage over a conventional machine-learning approach, its blind storage-day identification performance was compared with a support vector machine benchmark. Support vector machines are widely used in electronic-nose analysis because they can construct nonlinear class boundaries in high-dimensional sensor spaces and often perform effectively with relatively small datasets. However, their reliability depends strongly on feature representation, parameter optimization, and validation design, particularly when neighbouring quality or storage classes exhibit gradual rather than discrete differences (Sanislav et al., 2025; Zhai et al., 2024). Recent electronic-nose studies continue to use SVM-based models for fruit, vegetable, food-quality, and volatile-compound classification, frequently reporting strong performance when the sensor set and modelling parameters are carefully optimized.

The SVM benchmark was constructed directly from the complete temporal sensor responses rather than from the model-derived SRS descriptors. Each reference sample contained 51 measurements from each of the four MOS sensors. The temporal response vector of sample *i*was therefore defined as

$$\mathbf{x}_i = \left[S_{1,i}(t_1),\ldots,S_{1,i}(t_{51}),S_{2,i}(t_1),\ldots,S_{2,i}(t_{51}),S_{3,i}(t_1),\ldots,S_{3,i}(t_{51}),S_{4,i}(t_1),\ldots,S_{4,i}(t_{51})\right].$$

Thus, each sample was represented by

$$p_{\mathrm{SVM}} = 51 \times 4 = 204$$

raw temporal variables, and the reference input matrix had the dimensionality

$$\mathbf{X}_{\mathrm{ref}} \in \mathbb{R}^{21\times 204}.$$

The corresponding blind matrix was

$$\mathbf{X}_{\mathrm{blind}} \in \mathbb{R}^{21\times 204}.$$

This benchmark represented a conventional data-driven strategy in which the classifier was expected to learn the storage-day boundaries directly from the standardized temporal signals, without explicit nonlinear curve modelling or kinetic feature extraction.

All input variables were standardized using only the labelled reference dataset. For temporal variable *j*, the standardized value was calculated as

$$x_{ij}^{*} = \frac{x_{ij} - \mu_j^{\mathrm{ref}}}{\sigma_j^{\mathrm{ref}}},$$

where

$$\mu_j^{\mathrm{ref}} = \frac{1}{21}\sum_{i=1}^{21} x_{ij},$$

and

$$\sigma_j^{\mathrm{ref}} = \sqrt{\frac{\sum_{i=1}^{21}\left(x_{ij} - \mu_j^{\mathrm{ref}}\right)^2}{20}}.$$

The same reference-derived transformation was applied to the blind samples:

$$x_{bj}^{*,\mathrm{blind}} = \frac{x_{bj}^{\mathrm{blind}} - \mu_j^{\mathrm{ref}}}{\sigma_j^{\mathrm{ref}}}.$$

No information from the blind samples was used during scaling, model selection, or hyperparameter optimization.

A radial basis function kernel was selected because the storage-day classes were not expected to be linearly separable in the 204-dimensional raw-response space. The RBF kernel between two standardized response vectors was

$$K(\mathbf{x}_i,\mathbf{x}_j) = \exp\left[\quad -\gamma \| \mathbf{x}_i - \mathbf{x}_j \|_2^2\right].$$

The multiclass SVM decision function was based on the conventional maximum-margin formulation:

$$\min_{\mathbf{w},b,\xi} \frac{1}{2} \| \mathbf{w} \|^2 + C\sum_{i=1}^{N} \xi_i,$$

subject to

$$y_i\left(\mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_i) + b\right) \geq 1 - \xi_i,$$

and

$$\xi_i \geq 0.$$

Here, Ccontrols the trade-off between margin width and classification error, while γcontrols the effective width of the RBF kernel.

The hyperparameters were selected exclusively from the reference data using leave-one-producer-out cross-validation. In each validation iteration, all seven samples from one tomato group were excluded from model training and used as the validation set, whereas the remaining 14 samples from the other two groups were used for fitting. This procedure provided a stricter assessment than random sample-level splitting because measurements from the same tomato group were never simultaneously present in the training and validation sets.

The examined parameter grid was

$$C \in \{0.01, 0.1, 1, 10, 100, 1000\},$$

and

$$\gamma \in \{\text{scale}, 0.0001, 0.001, 0.01, 0.1, 1\}.$$

The best cross-validated parameter combination was

$$C = 100$$

and

$$\gamma = 0.001.$$

Using these parameters, the leave-one-producer-out SVM correctly classified 15 of the 21 reference samples:

$$\text{Accuracy}_{\text{LOPO}} = \frac{15}{21} \times 100 = 71.43\%.$$

The corresponding within-one-day accuracy was

$$\text{Accuracy}_{\pm 1,\text{LOPO}} = \frac{19}{21} \times 100 = 90.48\%.$$

The cross-validated mean absolute error was

$$MAE_{\mathrm{LOPO}} = \frac{1}{21}\sum_{i=1}^{21} | \hat{d}_i - d_i | = 0.381 \text{ days},$$

while the root-mean-square error was

$$RMSE_{\mathrm{LOPO}} = \sqrt{\frac{1}{21}\sum_{i=1}^{21}\left(\hat{d}_i - d_i\right)^2} = 0.756 \text{ days}.$$

Cohen's kappa was

$$\kappa_{\mathrm{LOPO}} = 0.667.$$

The cross-validation errors were concentrated mainly among the intermediate and late neighbouring storage days. All Day 1 and Day 2 samples were classified correctly, whereas partial confusion occurred among Days 3–7. This pattern indicated that the raw temporal vectors contained a strong general storage gradient but did not consistently provide producer-independent exact-day separation.

Following hyperparameter selection, the RBF-SVM was retrained using all 21 labelled reference samples and then applied once to the independent blind dataset. The model correctly identified 19 of the 21 blind samples, producing an exact storage-day accuracy of

$$\text{Accuracy}_{\mathrm{SVM}} = \frac{19}{21} \times 100 = 90.48\%.$$

The blind confusion matrix was

$$\mathbf{C}_{\mathrm{SVM}} = \begin{bmatrix} 3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 3 \end{bmatrix}.$$

All blind samples belonging to Days 1, 2, 3, 5, 6, and 7 were identified correctly. The two errors occurred within the Day 4 class: one Day 4 sample was classified as Day 5, and another was classified as Day 6. Therefore, Day 4 recall was

$$\text{Recall}_{4} = \frac{1}{3} = 0.333,$$

whereas recall for every other day was

$$\text{Recall}_{d} = 1.000.$$

The class-specific precision values were

$$\text{Precision}_{1} = \text{Precision}_{2} = \text{Precision}_{3} = \text{Precision}_{4} = \text{Precision}_{7} = 1.000,$$

and

$$\text{Precision}_{5} = \text{Precision}_{6} = 0.750.$$

The macro-averaged precision, recall, and F1 score were

$$\text{Precision}_{\text{macro}} = 0.929,$$
$$\text{Recall}_{\text{macro}} = 0.905,$$

and

$$F1_{\text{macro}} = 0.888.$$

Cohen's kappa for the independent blind predictions was

$$\kappa_{\text{SVM}} = 0.889,$$

indicating strong agreement beyond chance despite the two Day 4 errors.

Because one Day 4 sample was assigned to Day 5 and one to Day 6, the mean absolute prediction error was

$$MAE_{\text{SVM}} = \frac{1+2}{21} = 0.143 \text{ days}.$$

The corresponding root-mean-square error was

$$RMSE_{\text{SVM}} = \sqrt{\frac{1^2+2^2}{21}} = 0.488 \text{ days}.$$

The maximum absolute error was

$$E_{\text{max,SVM}} = 2 \text{ days}.$$

The within-one-day accuracy was

$$\text{Accuracy}_{\pm 1,\text{SVM}} = \frac{20}{21} \times 100 = 95.24\%.$$

Thus, although the SVM produced high overall performance, one blind sample was assigned two storage days away from its actual class.

The Richards-based SRS-library method outperformed the SVM benchmark for every primary independent-test metric. The proposed method correctly identified all 21 blind samples:

$$\text{Accuracy}_{\text{SRS}} = 100\%,$$

whereas the SVM achieved

$$\text{Accuracy}_{\text{SVM}} = 90.48\%.$$

The absolute improvement in exact-day accuracy was therefore

$$\Delta\text{Accuracy} = 100.00 - 90.48 = 9.52 \text{ percentage points}.$$

The relative reduction in the number of incorrectly identified samples was

$$\text{Error reduction} = \frac{2-0}{2} \times 100 = 100\%.$$

Similarly, the SRS-library method reduced the mean absolute error from

$$0.143 \text{ days}$$

to

$$0.000 \text{ days},$$

and reduced the RMSE from

$$0.488 \text{ days}$$

to

$$0.000 \text{ days}.$$

Cohen's kappa improved from

$$0.889$$

for the SVM to

$$1.000$$

for the SRS-library approach.

A direct performance summary is provided below.

Metric	**RBF-SVM benchmark**	**Richards-based SRS library**
Exact-day accuracy	90.48%	100.00%
Correct blind samples	19/21	21/21
±1-day accuracy	95.24%	100.00%
Macro precision	0.929	1.000
Macro recall	0.905	1.000
Macro F1 score	0.888	1.000
Cohen's kappa	0.889	1.000
MAE	0.143 days	0.000 days
RMSE	0.488 days	0.000 days
Maximum error	2 days	0 days

Metric	RBF-SVM benchmark	Richards-based SRS library
Input dimension	204	32

The comparison also demonstrated an important difference in data representation. The SVM used the complete 204-dimensional raw temporal vector, whereas the proposed approach used only 32 nonlinear kinetic descriptors:

$$p_{\mathrm{SVM}} = 204,$$
$$p_{\mathrm{SRS}} = 32.$$

The reduction in input dimensionality was

$$\left(1-\frac{32}{204}\right)\times 100 = 84.31\%.$$

Therefore, the SRS method achieved higher blind-test accuracy while operating in a substantially smaller feature space.

The result suggests that the nonlinear kinetic representation was more suitable than direct SVM modelling of the raw curves for this dataset. The 204 raw SVM variables were strongly temporally correlated, because consecutive measurements differed only by 2 s and described the same underlying response trajectory. With only 21 reference samples, the ratio of variables to samples was

$$\frac{p}{N} = \frac{204}{21} = 9.71.$$

In contrast, the corresponding ratio for the SRS representation was

$$\frac{32}{21} = 1.52.$$

The considerably larger feature-to-sample ratio of the raw SVM benchmark increased the difficulty of learning producer-independent boundaries, even though margin regularization was used.

The Richards-based transformation addressed this issue by replacing redundant time points with descriptors that summarized asymptotic magnitude, cumulative response, kinetic rate, and characteristic response times. This did not merely compress the data numerically; it organized the sensor information according to the underlying nonlinear response behaviour. Recent research on dynamic gas-sensor analysis similarly emphasizes that the quality of the extracted representation can be more important than increasing classifier complexity and that model-derived or transient features may improve discrimination by removing temporal redundancy while retaining chemically meaningful differences.

The SVM errors also provided insight into the nature of the storage-day problem. Both incorrect predictions originated from the Day 4 class and were shifted toward later storage stages. This result is consistent with the intermediate-stage overlap observed in the SRS centroid analysis, where Days 3–5 occupied relatively close regions and showed greater within-day dispersion than the earliest and latest classes. Therefore, the SVM did not fail randomly. Instead, it had difficulty defining a stable decision boundary in the transitional region where the raw response curves evolved gradually.

In contrast, the sample-level SRS library retained three individual reference signatures for every storage day and did not force the data into a single global separating surface. Each blind signature was matched directly to its nearest labelled nonlinear fingerprint. This local reference-matching strategy preserved group-specific response structure that could be averaged out or distorted by a global discriminative classifier.

The difference can be expressed conceptually as follows. The SVM assigned the blind sample according to a learned decision function:

$$\hat{d}_b^{\mathrm{SVM}} = \arg\max \mathrm{d} f_d\left(\mathbf{x}_b^*\right),$$

whereas the SRS-library method assigned the day of the nearest labelled kinetic signature:

$$\hat{d}_b^{\mathrm{SRS}} = d_{\arg\min i \|\mathbf{z}_b^{\mathrm{SRS}} - \mathbf{z}_i^{\mathrm{ref}}\|_2}.$$

The first approach estimated global nonlinear class boundaries from only 21 samples, while the second used direct similarity to physically interpretable reference fingerprints. Under the controlled conditions of the present study, the latter strategy was more reliable.

The comparison does not indicate that SVM is inherently unsuitable for electronic-nose analysis. SVM models have achieved high accuracies in many recent electronic-nose applications, including optimized fruit and vegetable classification, volatile-compound recognition, and food-freshness assessment. Their performance is often strongest when larger training sets, optimized sensor combinations, dimensionality reduction, or carefully engineered features are available. Rather, the present findings demonstrate that direct SVM modelling of a small number of highly dimensional raw temporal trajectories was less effective than using a nonlinear kinetic reference library specifically designed for storage-day identification.

Overall, the RBF-SVM benchmark provided a strong independent-test performance of 90.48% exact-day accuracy and 95.24% within-one-day accuracy. Nevertheless, it misclassified two Day 4 samples and produced a maximum error of two storage days. The Richards-based SRS-library approach correctly identified all 21 blind samples, achieved zero-day MAE and RMSE, and reduced the input dimension by 84.31%. These findings demonstrate that the improvement obtained by the proposed method was not attributable merely to the use of a powerful nonlinear classifier. Instead, it arose from converting the original electronic-nose curves into compact, interpretable, and storage-sensitive nonlinear response signatures before blind library matching.

4. Discussion

4.1. Continuous Representation of Temporal Sensor Responses

The temporal responses generated by the four MOS sensors were represented as continuous nonlinear functions rather than as isolated ADC measurements collected at discrete time points. The reference dataset comprised 21 labelled tomato samples representing three independent sample groups and seven storage days. Each sample was measured at 51 time points between 0 and 100 s using four sensors, resulting in

$$N_{\mathrm{ref}} = 21 \times 51 = 1071$$

reference observations and

$$N_{\mathrm{curves,ref}} = 21 \times 4 = 84$$

individual sensor-response trajectories. The independent blind dataset had the same structure, providing an additional 84 previously unseen temporal response curves. Therefore, the continuous-response analysis was based on a total of

$$N_{\mathrm{curves,total}} = 84 + 84 = 168$$

sensor trajectories.

For sample i, sensor s, and measurement time t_j, the experimentally observed response was expressed as

$$y_{isj} = y_{is}(t_j),$$

where

$$t_j \in \{0,2,4,\ldots,100\} \text{ s.}$$

Although the measurements were recorded at 2-s intervals, the underlying sensor response was considered a continuous process:

$$y_{is}(t), 0 \le t \le 100.$$

The experimentally sampled sequence was therefore interpreted as a finite observation of an underlying nonlinear response function:

$$\{y_{is}(t_1), y_{is}(t_2), \ldots, y_{is}(t_{51})\} \longrightarrow \hat{y}_{is}(t).$$

Following the comparative model analysis, the Richards function was used to generate this continuous representation. In its generalized form, the fitted response was expressed as

$$\hat{y}_{is}(t) = L_{is} + \frac{U_{is} - L_{is}}{\left[1 + \nu_{is} \exp\left(-k_{is}\left(t - \tau_{is}\right)\right)\right]^{1/\nu_{is}}},$$

where L_{is} denotes the lower response level, U_{is} represents the estimated upper asymptote, k_{is} is the response-rate coefficient, τ_{is} controls the temporal displacement of the curve, and ν_{is} is the shape parameter governing response asymmetry. The inclusion of the shape parameter allowed the fitted function to adapt to differences in early acceleration, intermediate curvature, and late-stage saturation among the sensors and storage days.

The continuous function was estimated by minimizing the residual sum of squares between the measured and fitted responses:

$$RSS_{is} = \sum_{j=1}^{51} \left[y_{is}(t_j) - \hat{y}_{is}(t_j)\right]^2.$$

The fitted curves reproduced the measured trajectories with very high accuracy. All 84 reference trajectories and all 84 independent blind trajectories converged successfully, corresponding to a fitting-success rate of

$$\text{Convergence} = 100\%.$$

For the labelled reference dataset, the median coefficient of determination was

$$\tilde{R}^2_{\text{ref}} = 0.999787,$$

while the mean coefficient of determination was close to the same level. The median and mean fitting errors were

$$\widetilde{RMSE}_{\text{ref}} = 2.397 \text{ ADC},$$

and

$$\overline{RMSE}_{\text{ref}} = 2.461 \text{ ADC},$$

respectively. These errors were very small compared with the full response range, which extended to more than 800 ADC units for some late-storage samples.

The continuous representation remained equally stable when applied to the independent blind data. The blind trajectories produced a median coefficient of determination of

$$\widetilde{R^2}_{\text{blind}} = 0.999787,$$

and a median RMSE of

$$\widetilde{RMSE}_{\text{blind}} = 2.368 \text{ ADC}.$$

The relative difference between the reference and blind median fitting errors was only

$$\frac{2.397 - 2.368}{2.397} \times 100 = 1.21\%.$$

Thus, the continuous model preserved essentially the same fitting quality when transferred from the labelled reference responses to the blind measurements. This close agreement indicated that the fitted nonlinear structure was not specific to the calibration samples.

The raw time-series data also showed that the sensor responses were almost completely monotonic over the 100-s acquisition period. For each trajectory, the proportion of nondecreasing successive measurements was calculated as

$$M_{is} = \frac{\sum_{j=1}^{50} I\left[y_{is}(t_{j+1}) \geq y_{is}(t_j)\right]}{50}.$$

Across the 84 reference trajectories, the mean monotonic proportion was

$$\overline{M}_{\text{ref}} = 0.99976,$$

equivalent to

$$99.976\%.$$

Only one negative step was observed among the

$$84 \times 50 = 4200$$

successive response increments. The blind dataset exhibited the same monotonicity level. This result confirmed that the measured signals followed a smooth increasing response pattern rather than an oscillatory or irregular temporal structure.

The continuous trajectories displayed three distinguishable response regions. The first region consisted of a relatively slow initial increase immediately after sample exposure. This was followed by a more pronounced response-development region in which the rate of change increased. During the final portion of the measurement, the response rate gradually decreased as the curves approached their asymptotic region. Because the transition among these stages was continuous, selecting only one arbitrary time point would not adequately describe the complete sensor behaviour.

The instantaneous rate of sensor response was obtained directly from the first derivative of the fitted Richards function:

$$v_{is}(t) = \frac{d\hat{y}_{is}(t)}{dt}.$$

For the fitted Richards formulation, the derivative can be expressed as

$$v_{is}(t) = \left(U_{is} - L_{is}\right) k_{is} \exp\left[-k_{is}\left(t - \tau_{is}\right)\right] \left[1 + \nu_{is} \exp\left(-k_{is}\left(t - \tau_{is}\right)\right)\right]^{-\left(1 + \frac{1}{\nu_{is}}\right)}.$$

This derivative provided a continuous description of the response rate at every time point rather than limiting rate calculations to finite differences between two measured observations. The maximum response rate was consequently defined as

$$v_{\max,is} = \max_{0 \le t \le 100} v_{is}(t),$$

and the corresponding temporal location was

$$t_{\max,is} = \arg\max\ 0 \le t \le 100\, v_{is}(t).$$

The continuous representation also allowed characteristic response times to be calculated even when the exact response level did not coincide with a measured 2-s time point. For example, the half-response time was defined as

$$\hat{y}_{is}\left(t_{50,is}\right) = L_{is} + 0.50\left(U_{is} - L_{is}\right),$$

whereas the 90% response time satisfied

$$\hat{y}_{is}\left(t_{90,is}\right) = L_{is} + 0.90\left(U_{is} - L_{is}\right).$$

Similarly, the cumulative response over the complete measurement interval was calculated from the fitted function:

$$AUC_{is} = \int_0^{100} \hat{y}_{is}(t)\, dt.$$

These quantities could not be estimated as precisely using only the original discrete measurements, particularly for trajectories whose characteristic response times occurred between consecutive sampling points.

The temporal responses showed a strong systematic increase with storage day. For S1, the mean final response increased from

$$180.0 \pm 25.6\ \text{ADC}$$

on Day 1 to

$$771.0 \pm 59.6\ \text{ADC}$$

on Day 7. The relative increase was

$$\frac{771.0 - 180.0}{180.0} \times 100 = 328.3\%.$$

For S2, the mean final response increased from

$$178.0 \pm 21.1$$

to

$$741.3 \pm 94.6\ \text{ADC},$$

representing an increase of

316.5%.

The corresponding values for S3 were

$$178.3 \pm 21.4$$

and

$$708.3 \pm 90.7 \text{ ADC},$$

equivalent to an increase of

297.2%.

S4 showed a similar change from

$$177.3 \pm 21.7$$

to

$$708.3 \pm 90.7 \text{ ADC},$$

corresponding to an increase of

299.4%.

The increase was not restricted to the final sensor values. The complete temporal trajectories shifted upward with storage duration. For example, the mean S1 response at 50 s increased from

$$90.0 \pm 15.5 \text{ ADC}$$

on Day 1 to

$$422.3 \pm 27.5 \text{ ADC}$$

on Day 7. Similar changes were observed for the other sensors. At 50 s, the mean responses increased from 87.3 to 409.0 ADC for S2 and from 86.7 to 381.3 ADC for S3. Thus, the separation among storage days was present throughout the response-development interval and was not generated only at the end of the acquisition period.

The integrated responses exhibited an even more pronounced temporal shift. The mean observed S1 AUC increased from

$$9214.7 \pm 1599.5 \text{ ADC} \cdot \text{s}$$

on Day 1 to

$$41644.0 \pm 3056.7 \text{ ADC} \cdot \text{s}$$

on Day 7. The increase was

$$\frac{41644.0 - 9214.7}{9214.7} \times 100 = 352.0\%.$$

For S2, the AUC increased from

$$9064.7 \pm 1420.6$$

to

$$40291.7 \pm 4882.8\ \mathrm{ADC} \cdot \mathrm{s}.$$

The Day 1 and Day 7 values for S3 were

$$9037.7 \pm 1308.1$$

and

$$37980.3 \pm 5067.7\ \mathrm{ADC} \cdot \mathrm{s},$$

while the corresponding S4 values were

$$9188.0 \pm 1505.3$$

and

$$37980.3 \pm 5067.7\ \mathrm{ADC} \cdot \mathrm{s}.$$

These results demonstrate that increasing storage duration changed both the response amplitude and the cumulative temporal exposure recorded by the sensors.

The early-stage response rates also increased markedly. Linear slopes calculated over the first 20 s increased for S1 from

$$1.617 \pm 0.556\ \mathrm{ADC} \cdot \mathrm{s}^{-1}$$

on Day 1 to

$$7.364 \pm 0.296\ \mathrm{ADC} \cdot \mathrm{s}^{-1}$$

on Day 7. For S2, the corresponding change was from

$$1.588 \pm 0.555$$

to

$$7.221 \pm 0.499\ \mathrm{ADC} \cdot \mathrm{s}^{-1}.$$

S3 increased from

$$1.520 \pm 0.448$$

to

$$6.667 \pm 0.650\ \mathrm{ADC} \cdot \mathrm{s}^{-1},$$

whereas S4 increased from

$$1.641 \pm 0.598$$

to

$$6.667 \pm 0.650 \text{ ADC} \cdot \text{s}^{-1}.$$

Therefore, later-storage samples did not merely produce higher final responses; they also activated the sensor array more rapidly during the initial portion of the measurement.

The strength of the continuous storage-related progression was confirmed by correlation analysis. For S1, the Pearson correlations between storage day and final response, AUC, and early response slope were

$$r_{\text{Final}} = 0.928,$$
$$r_{\text{AUC}} = 0.934,$$

and

$$r_{\text{Slope}} = 0.892,$$

respectively. The corresponding values for S2 were

$$0.911, 0.915, 0.882.$$

For S3, the correlations were

$$0.906, 0.921, 0.956,$$

while S4 produced

$$0.909, 0.917, 0.897.$$

The strongest individual association was obtained for the S3 early-response slope:

$$r = 0.956.$$

These results show that the temporal evolution of the responses was reflected simultaneously in endpoint, integrated, and rate-based characteristics.

The independent blind responses closely reproduced the continuous structure of the labelled reference curves. After matching each blind sample with the corresponding experimental group and storage day, the mean absolute point-by-point difference across all four sensors and all measurement times was

$$MAE_{\text{raw}} = 0.989 \text{ ADC}.$$

The corresponding root-mean-square difference was

$$RMSE_{\text{raw}} = 0.994 \text{ ADC}.$$

Across the 84 matched reference–blind sensor trajectories, the median Pearson correlation was

$$\tilde{r}_{\text{ref,blind}} = 1.000,$$

and the minimum correlation was

$$r_{\min} = 0.999991.$$

Of the 4284 paired sensor measurements,

4236

blind values differed from their corresponding reference measurements by only 1 ADC unit, whereas the remaining

48

values were identical. This extremely close agreement explains why the same continuous Richards representation could be applied to the blind data without loss of fitting quality.

The use of a continuous response function provided several analytical advantages. First, it reduced the influence of small point-level fluctuations and ADC quantization. Second, it allowed responses at any intermediate time to be estimated through interpolation:

$$\hat{y}_{is}(t^*), t^* \notin \{0,2,4,\ldots,100\}.$$

Third, derivatives, characteristic response times, asymptotic levels, and integrated responses could be derived from a common mathematical function. Finally, the original sequence of 51 measurements per sensor could be converted into a compact set of interpretable kinetic descriptors without discarding the principal temporal behaviour.

Overall, the Richards-based continuous representation reproduced the temporal sensor responses with near-unity coefficients of determination, low fitting errors, complete convergence, and equivalent performance in the reference and blind datasets. The curves were almost entirely monotonic and exhibited systematic storage-dependent increases in amplitude, cumulative response, and response rate. These findings establish that the electronic-nose measurements were more appropriately regarded as continuous nonlinear trajectories than as collections of independent time-point values. The resulting representation therefore provided the mathematical basis for subsequent extraction of Sensor Response Signatures and blind storage-day identification.

4.2. Selection of the Nonlinear Response Model

The Logistic, Gompertz, and Richards functions were comparatively evaluated to identify the nonlinear model that most accurately represented the temporal responses of the four MOS sensors. Model selection was performed using the labelled reference dataset, which contained 21 tomato samples, four sensor channels, and 51 measurements collected between 0 and 100 s. Accordingly, the comparison included

$$N_{\text{curves,ref}} = 21 \times 4 = 84$$

independent reference sensor trajectories. After model selection, the same three functions were fitted to the 84 trajectories in the blind dataset to determine whether the selected model maintained its performance when applied to previously unseen measurements.

The Logistic model was expressed as

$$\hat{y}(t) = L + \frac{U - L}{1 + \exp\left[-k(t - t_0)\right]},$$

where $L$and $U$are the lower and upper asymptotes, $k$is the response-rate coefficient, and $t_0$is the temporal location of the inflection point. The Logistic function assumes symmetry around its inflection point and therefore imposes equivalent response behaviour before and after the maximum-rate region.

The Gompertz model was defined as

$$\hat{y}(t) = L + (U - L)\exp\left\{-\exp\left[-k(t - t_0)\right]\right\}.$$

Unlike the Logistic model, the Gompertz function is intrinsically asymmetric. However, the degree and direction of this asymmetry are fixed by its mathematical structure and cannot be adjusted independently for different response trajectories.

The Richards function was expressed as

$$\hat{y}(t) = L + \frac{U - L}{\left[1 + \nu\exp\left[-k(t - t_0)\right]\right]^{1/\nu}},$$

where $\nu$is an additional shape parameter. This parameter allows the Richards function to vary continuously between approximately symmetric and strongly asymmetric curve forms. Consequently, it can accommodate variations in early response development, inflection behaviour, and late-stage saturation among sensors and storage days.

Each model was fitted independently to every sensor trajectory by minimizing the residual sum of squares:

$$RSS = \sum_{j=1}^{51}\left[y(t_j) - \hat{y}(t_j)\right]^2.$$

Model performance was evaluated using the coefficient of determination, adjusted coefficient of determination, root-mean-square error, mean absolute error, corrected Akaike Information Criterion, and Bayesian Information Criterion. The coefficient of determination was calculated as

$$R^2 = 1 - \frac{\sum_{j=1}^{51}\left[y(t_j) - \hat{y}(t_j)\right]^2}{\sum_{j=1}^{51}\left[y(t_j) - \overline{y}\right]^2}.$$

Because the Richards model contained one additional parameter, the adjusted coefficient of determination was also calculated:

$$R^2_{\text{adj}} = 1 - \left(1 - R^2\right)\frac{n - 1}{n - p - 1},$$

where $n = 51$and $p$is the number of fitted model parameters.

The residual error measures were

$$RMSE = \sqrt{\frac{RSS}{n}},$$

and

$$MAE = \frac{1}{n}\sum_{j=1}^{n} \mid y(t_j) - \hat{y}(t_j) \mid.$$

Model complexity was considered using

$$AIC = n\ln\left(\frac{RSS}{n}\right) + 2p,$$
$$AIC_c = AIC + \frac{2p(p+1)}{n-p-1},$$

and

$$BIC = n\ln\left(\frac{RSS}{n}\right) + p\ln(n).$$

Lower RMSE, MAE, AICc, and BIC values indicated better model performance, whereas higher $R^2$and adjusted R^2 values indicated greater agreement between the measured and fitted responses.

All three candidate models converged successfully for all 84 reference trajectories. The Richards model achieved the best overall fitting performance. Its median coefficient of determination was

$$\widetilde{R^2}_{\text{Richards}} = 0.999793,$$

compared with

$$\widetilde{R^2}_{\text{Logistic}} = 0.999142$$

and

$$\widetilde{R^2}_{\text{Gompertz}} = 0.998558.$$

Although all three models produced high coefficients of determination, the differences were meaningful because the response trajectories were smooth and highly structured. In this context, even small decreases in R^2corresponded to systematic deviations extending over multiple time points.

The mean reference-dataset R^2values were

$$\overline{R^2}_{\text{Richards}} = 0.999743,$$
$$\overline{R^2}_{\text{Logistic}} = 0.999086,$$

and

$$\overline{R^2}_{\text{Gompertz}} = 0.998522.$$

Adjustment for the additional Richards parameter did not alter the ranking. The median adjusted coefficient of determination was

$$\widetilde{R^2_{\text{adj}}}_{\text{Richards}} = 0.999770,$$

whereas the corresponding values for the Logistic and Gompertz models were

0.999067

and

0.998432,

respectively. Thus, the improvement obtained with the Richards function remained evident even after accounting for its greater flexibility.

The residual-error criteria showed a more pronounced difference among the models. The median RMSE values were

$$\widetilde{RMSE}_{\text{Richards}} = 2.331 \text{ ADC},$$
$$\widetilde{RMSE}_{\text{Logistic}} = 4.930 \text{ ADC},$$

and

$$\widetilde{RMSE}_{\text{Gompertz}} = 6.407 \text{ ADC}.$$

The Richards model therefore reduced the median fitting error relative to the Logistic model by

$$\frac{4.930 - 2.331}{4.930} \times 100 = 52.72\%,$$

and relative to the Gompertz model by

$$\frac{6.407 - 2.331}{6.407} \times 100 = 63.62\%.$$

The mean RMSE results showed the same ranking:

$$\overline{RMSE}_{\text{Richards}} = 2.452 \text{ ADC},$$
$$\overline{RMSE}_{\text{Logistic}} = 4.860 \text{ ADC},$$

and

$$\overline{RMSE}_{\text{Gompertz}} = 6.266 \text{ ADC}.$$

The median MAE was also lowest for Richards:

$$\widetilde{MAE}_{\text{Richards}} = 1.864 \text{ ADC},$$

compared with

$$\widetilde{MAE}_{\text{Logistic}} = 3.766 \text{ ADC}$$

and

$$\widetilde{MAE}_{\text{Gompertz}} = 5.369 \text{ ADC}.$$

These results demonstrate that the Richards function reproduced the individual measured points more closely throughout the complete response interval.

The information criteria provided particularly strong evidence in favour of the Richards model. The median AICc values were

$$\widetilde{AIC}_{c_{\text{Richards}}} = 97.665,$$
$$\widetilde{AIC}_{c_{\text{Logistic}}} = 171.536,$$

and

$$\widetilde{AIC}_{c_{\text{Gompertz}}} = 198.326.$$

The median AICc advantage of Richards over Logistic was therefore

$$\Delta AIC_c = 171.536 - 97.665 = 73.871,$$

while its advantage over Gompertz was

$$\Delta AIC_c = 198.326 - 97.665 = 100.661.$$

Differences of this magnitude provide decisive evidence that the lower residual error of the Richards function more than compensated for its additional parameter.

The median BIC values were

$$\widetilde{BIC}_{\text{Richards}} = 105.990,$$
$$\widetilde{BIC}_{\text{Logistic}} = 178.393,$$

and

$$\widetilde{BIC}_{\text{Gompertz}} = 205.184.$$

Because BIC imposes a stronger penalty for additional parameters than AICc, the continued superiority of Richards under this criterion demonstrates that its improved fit was not caused merely by unnecessary model complexity.

Trajectory-level model ranking produced the same conclusion. Richards yielded the lowest RMSE for

83

of the 84 reference trajectories, equivalent to

$$\frac{83}{84} \times 100 = 98.81\%.$$

Only one trajectory showed a marginally lower RMSE with the Gompertz model. Richards also achieved the highest R^2 for the same 83 trajectories.

When model complexity was included, Richards produced the lowest AICc and BIC for

of the 84 trajectories:

$$\frac{79}{84} \times 100 = 94.05\%.$$

The Logistic model was preferred for four curves and the Gompertz model for only one curve. Therefore, Richards was not selected solely because of superior aggregate statistics; it was also the preferred model for the overwhelming majority of individual sensor responses.

The sensor-specific results confirmed that the superiority of Richards was consistent across the complete sensor array. For S1, the median RMSE values were

$$RMSE_{\text{Richards}} = 2.515,$$
$$RMSE_{\text{Logistic}} = 4.750,$$

and

$$RMSE_{\text{Gompertz}} = 6.420 \text{ ADC}.$$

For S2, the corresponding values were

$$2.449, 5.155, 6.682 \text{ ADC}.$$

For S3, they were

$$1.935, 4.413, 6.277 \text{ ADC},$$

and for S4,

$$2.321, 5.111, 7.628 \text{ ADC}.$$

The lowest Richards fitting error was obtained for S3, whereas the largest advantage over the Gompertz model was observed for S4. Nevertheless, Richards remained the best-performing model for every sensor. This consistency indicates that the model selection was not driven by a single dominant channel.

Visual and residual examination explained the statistical differences. The Logistic model generally represented the central portion of the curves adequately but imposed excessive symmetry around the inflection region. It consequently produced structured deviations during the early response-development and late saturation phases. The Gompertz function accommodated asymmetry but imposed a fixed asymmetric form that did not match all sensor trajectories. It frequently underestimated one response region while overestimating another.

The Richards shape parameter allowed the degree of asymmetry to vary independently for each trajectory. As a result, the model could adapt to sensor- and storage-specific changes in curvature without forcing all responses to follow the same fixed geometry. Its residuals were therefore smaller and more evenly distributed over the 100-s interval.

The selected model was subsequently evaluated on the independent blind dataset. All 84 blind trajectories converged successfully for all three candidate functions. The median blind-data R^2values were

$$\tilde{R}^2_{\text{Richards}} = 0.999795,$$
$$\tilde{R}^2_{\text{Logistic}} = 0.999142,$$

and

$$\tilde{R^2}_{\text{Gompertz}} = 0.998531.$$

The median blind RMSE values were

$$\widetilde{RMSE}_{\text{Richards}} = 2.333 \text{ ADC},$$
$$\widetilde{RMSE}_{\text{Logistic}} = 4.930 \text{ ADC},$$

and

$$\widetilde{RMSE}_{\text{Gompertz}} = 6.407 \text{ ADC}.$$

Richards therefore reduced the blind median RMSE by

$$52.67\%$$

relative to Logistic and by

$$63.58\%$$

relative to Gompertz.

The median blind-data AICc values were

$$97.759, 171.536, 198.326$$

for Richards, Logistic, and Gompertz, respectively. Richards again achieved the lowest AICc and BIC for 79 of the 84 blind trajectories and the lowest RMSE for 83 trajectories.

Most importantly, the fitting statistics of the selected model were almost identical in the reference and blind datasets. The relative difference between the median Richards RMSE values was

$$\frac{2.333 - 2.331}{2.331} \times 100 = 0.09\%.$$

Similarly, the median R^2, AICc, and BIC values showed negligible changes. This agreement demonstrates that the selection of Richards was reproducible and was not dependent on peculiarities of the labelled reference set.

Overall, the Logistic, Gompertz, and Richards functions all reproduced the general increasing structure of the MOS sensor trajectories. However, the Richards model consistently achieved the highest $R^2$and adjusted R^2, the lowest RMSE and MAE, and the most favourable AICc and BIC values. It was preferred for 94.05% of the reference trajectories according to the penalized information criteria and for 98.81% according to RMSE and R^2. The same ranking and nearly identical fitting statistics were obtained in the independent blind dataset. Accordingly, the Richards function was selected as the common nonlinear response model for subsequent continuous-curve analysis, kinetic descriptor extraction, SRS construction, and blind storage-day identification.

4.3. Discriminative Meaning of Sensor Response Signatures

The discriminative meaning of the Sensor Response Signatures was examined by determining how strongly the 32 Richards-derived descriptors were associated with storage duration and how effectively they represented differences among the seven storage-day classes. The SRS of sample *i*was defined as the concatenation of eight descriptors extracted from each of the four sensor trajectories:

$$\mathbf{s}_i = \left[\boldsymbol{\phi}_{i1}, \boldsymbol{\phi}_{i2}, \boldsymbol{\phi}_{i3}, \boldsymbol{\phi}_{i4}\right],$$

where

$$\boldsymbol{\phi}_{is} = \left[A_{is}, R_{0,is}, R_{\max,is}, t_{\inf,is}, t_{50,is}, t_{90,is}, AUC_{is}, F_{is}\right].$$

The descriptors represented complementary aspects of the nonlinear response. The asymptotic amplitude A, fitted final response F, and AUC quantified response magnitude; $R_0$and $R_{\max}$described response kinetics; and $t_{\inf}$, t_{50}, and t_{90} described the temporal position and progression of the fitted curve. The complete signature therefore encoded not only how strongly a sensor responded but also how rapidly and at what stage its response developed.

The association between each SRS feature and storage day was first quantified using Pearson's correlation coefficient:

$$r_k = \frac{\sum_{i=1}^{21}\left(s_{ik} - \overline{s_k}\right)\left(d_i - \overline{d}\right)}{\sqrt{\sum_{i=1}^{21}\left(s_{ik} - \overline{s_k}\right)^2}\sqrt{\sum_{i=1}^{21}\left(d_i - \overline{d}\right)^2}},$$

where $s_{ik}$is feature $k$of sample i, and $d_i$is its storage day.

Across the 32 descriptors, the median absolute correlation with storage day was

$$\text{median}\left(|r_k|\right) = 0.875.$$

This high median value demonstrates that most components of the SRS changed systematically with storage duration. However, the strength and direction of this association differed according to descriptor type.

The strongest individual correlation was obtained for the fitted asymptotic amplitude of S3:

$$r_{S3,A} = 0.948.$$

The corresponding Spearman rank correlation was

$$\rho_{S3,A} = 0.932,$$

showing that the relationship was both strongly linear and monotonically ordered. The mean S3 amplitude increased from

$$333.78$$

on Day 1 to

$$1234.77$$

on Day 7, corresponding to a relative increase of

$$\frac{1234.77 - 333.78}{333.78} \times 100 = 269.93\%.$$

The S3 initial response rate was similarly informative:

$$r_{S3,R_0} = 0.943,$$

with a rank correlation of

$$\rho_{S3,R_0} = 0.952.$$

Its mean value increased from

$$1.299\ \text{ADC}\,\text{s}^{-1}$$

on Day 1 to

$$5.726\ \text{ADC}\,\text{s}^{-1}$$

on Day 7. This represented an increase of

$$340.81\%.$$

The high correlations of both amplitude and initial rate indicate that storage changed not only the total magnitude of the volatile-sensitive response but also the speed at which the sensor became activated after sample exposure.

AUC was one of the most consistently discriminative descriptors across the complete sensor array. The correlations between storage day and AUC were

$$r_{S1,AUC} = 0.934,$$
$$r_{S2,AUC} = 0.915,$$
$$r_{S3,AUC} = 0.921,$$

and

$$r_{S4,AUC} = 0.917.$$

The mean Day 1 and Day 7 S1 AUC values were

$$9215.60\ \text{ADC}\,\text{s}$$

and

$$41638.62\ \text{ADC}\,\text{s},$$

respectively. Therefore, the cumulative S1 response increased by

$$351.83\%.$$

The corresponding increases for S2, S3, and S4 were

$$344.39\%,$$
$$320.14\%,$$

and

313.27%,

respectively. These results show that AUC provided a stable representation of storage-dependent changes because it integrated differences over the entire 100-s response period instead of relying on a single measurement point.

The fitted final responses also showed strong storage relationships:

$$r_{S1,F} = 0.928,$$
$$r_{S2,F} = 0.911,$$
$$r_{S3,F} = 0.907,$$

and

$$r_{S4,F} = 0.909.$$

For S1, the fitted final response increased from

178.26 ADC

on Day 1 to

777.20 ADC

on Day 7, corresponding to an increase of

335.98%.

The Day 1–Day 7 increases for the remaining sensors were

324.87%

for S2,

302.95%

for S3, and

303.83%

for S4. The agreement between the AUC and final-response results indicates that storage-related separation was present throughout the temporal trajectories and was not restricted to one isolated response region.

The discriminative strength of each descriptor family was summarized using the median absolute correlation across the four sensors. The resulting values were

$$|\tilde{r}|_{AUC} = 0.919,$$
$$|\tilde{r}|_{F} = 0.910,$$
$$|\tilde{r}|_{R_0} = 0.901,$$
$$|\tilde{r}|_{R_{\max}} = 0.879,$$

and

$$|\widetilde{r}|_A = 0.877.$$

In contrast, the characteristic timing descriptors showed weaker direct linear associations:

$$|\widetilde{r}|_{t_{90}} = 0.237,$$
$$|\widetilde{r}|_{t_{50}} = 0.202,$$

and

$$|\widetilde{r}|_{t_{\text{inf}}} = 0.200.$$

This difference demonstrates that the principal storage gradient was encoded primarily in response magnitude and rate rather than in a simple monotonic shift of characteristic times. However, the lower correlations of the timing descriptors do not imply that they lacked discriminative value. Their effects were more likely nonlinear, class-specific, or complementary to amplitude-related features.

For example, characteristic response times could differentiate samples having similar final amplitudes but different curve-development patterns. Two samples may satisfy

$$F_i \approx F_j,$$

while still exhibiting

$$t_{50,i} \neq t_{50,j}$$

or

$$t_{90,i} \neq t_{90,j}.$$

Accordingly, the timing features were retained in the SRS because they described response geometry that was not fully represented by amplitude, AUC, or endpoint measurements.

One-way analysis of variance was used to determine the proportion of each feature's variability attributable to storage day. The effect size was calculated as

$$\eta_k^2 = \frac{SS_{\text{between},k}}{SS_{\text{total},k}}.$$

Across the 32 SRS descriptors, the median storage-day effect size was

$$\widetilde{\eta^2} = 0.851.$$

Thus, approximately 85% of the variability in a typical SRS component was associated with differences among storage days rather than with variation among the three independent tomato groups measured on the same day.

The largest effect sizes were observed for the initial response rates. For S1 and S2,

$$\eta^2_{\text{S1},R_0} = 0.926$$

and

$$\eta^2_{\mathrm{S2},R_0} = 0.925,$$

respectively. The effect size for S3 amplitude was

$$\eta^2_{\mathrm{S3},A} = 0.919,$$

while S1 AUC and S1 final response yielded

$$\eta^2_{\mathrm{S1},AUC} = 0.914$$

and

$$\eta^2_{\mathrm{S1},F} = 0.908.$$

These values indicate very large storage-day effects and confirm that the temporal response signatures were dominated by systematic storage-related changes.

The median effect sizes by descriptor family were

$$\tilde{\eta}^2_{R_0} = 0.917,$$
$$\tilde{\eta}^2_{F} = 0.888,$$
$$\tilde{\eta}^2_{AUC} = 0.887,$$
$$\tilde{\eta}^2_{A} = 0.858,$$

and

$$\tilde{\eta}^2_{R_{\max}} = 0.855.$$

The timing descriptors produced lower but still meaningful median effect sizes:

$$\tilde{\eta}^2_{t_{\mathrm{inf}}} = 0.529,$$
$$\tilde{\eta}^2_{t_{90}} = 0.435,$$

and

$$\tilde{\eta}^2_{t_{50}} = 0.394.$$

These values support the interpretation that characteristic times contained day-dependent information even though their overall relationships with day were not strongly linear.

The magnitude of early–late storage separation was also assessed using the standardized Day 1–Day 7 difference:

$$d_k = \frac{\bar{s}_{k,7} - \bar{s}_{k,1}}{s_{p,k}},$$

where the pooled standard deviation was

$$s_{p,k} = \sqrt{\frac{(n_1 - 1)s_{k,1}^2 + (n_7 - 1)s_{k,7}^2}{n_1 + n_7 - 2}}.$$

Extremely large standardized differences were obtained for the principal magnitude and kinetic descriptors. The largest effect was observed for the S1 initial response rate:

$$d_{\mathrm{S1},R_0} = 14.55.$$

The Day 1–Day 7 effects for S1 AUC and final response were

$$d_{\mathrm{S1},AUC} = 13.29$$

and

$$d_{\mathrm{S1},F} = 13.05,$$

respectively. Large effects were also observed for S2 initial response rate,

$$d_{\mathrm{S2},R_0} = 10.91,$$

S2 AUC,

$$d_{\mathrm{S2},AUC} = 8.68,$$

and S2 final response,

$$d_{\mathrm{S2},F} = 8.34.$$

Across the four sensors, the median Day 1–Day 7 standardized differences were

$$\tilde{d}_{R_0} = 9.14,$$
$$\tilde{d}_{AUC} = 8.25,$$
$$\tilde{d}_{F} = 8.11,$$
$$\tilde{d}_{R_{\max}} = 7.15,$$

and

$$\tilde{d}_{A} = 5.96.$$

These results demonstrate that the early and late storage states were separated by several pooled standard deviations in multiple independent SRS dimensions.

Sensor-level analysis showed that discriminative information was distributed across the complete array rather than being carried by only one sensor. The median absolute correlations across the eight descriptors were

$$|\widetilde{r}|_{\mathrm{S1}} = 0.882,$$
$$|\widetilde{r}|_{\mathrm{S2}} = 0.882,$$

$$|\widetilde{r}|_{S3} = 0.877,$$

and

$$|\widetilde{r}|_{S4} = 0.854.$$

The corresponding median effect sizes were

$$\widetilde{\eta}^2_{S1} = 0.855,$$
$$\widetilde{\eta}^2_{S2} = 0.858,$$
$$\widetilde{\eta}^2_{S3} = 0.822,$$

and

$$\widetilde{\eta}^2_{S4} = 0.791.$$

S1 and S2 therefore produced slightly stronger overall storage associations, whereas S3 contained the two strongest individual variables, namely $A$and R_0. S4 showed somewhat lower aggregate correlations but remained highly informative through its AUC and fitted final response. This complementary behaviour supports the use of the four-sensor signature rather than selection of a single supposedly dominant channel.

Principal component analysis further demonstrated that the 32 SRS dimensions contained a coherent but multidimensional storage pattern. After standardization, the first principal component explained

$$60.29\%$$

of the total variance, and the second principal component explained

$$16.80\%.$$

Together, the first two components accounted for

$$77.09\%$$

of the variance. The first three components explained

$$85.36\%.$$

The leading loadings of the first principal component were associated with AUC, final response, initial rate, and maximum response rate across several sensors. The absolute PC1 loadings included

$$|l_{S2,AUC}| = 0.227,$$
$$|l_{S1,AUC}| = 0.227,$$
$$|l_{S1,F}| = 0.227,$$
$$|l_{S4,AUC}| = 0.226,$$

and

$$|l_{S2,F}| = 0.226.$$

This balanced loading structure indicates that the dominant storage direction was not associated with one isolated sensor or descriptor. Instead, it represented a coordinated increase in cumulative response, final magnitude, and kinetic rate throughout the sensor array.

The discriminative contribution of different feature groups was additionally examined through nearest-centroid assignment in standardized subspaces. When only the four R_0features were used, exact storage-day localization was

$$\frac{15}{21} \times 100 = 71.43\%,$$

with all samples identified within one storage day. The four t_{50}descriptors and the four t_{90}descriptors each produced an exact-day localization rate of

$$76.19\%.$$

However, their errors were less consistently restricted to neighbouring classes than those obtained from the rate features.

Using the complete eight-descriptor block of S1 yielded an exact centroid-localization accuracy of

$$76.19\%,$$

while the S4 descriptor block produced the same exact-day value and

$$100\%$$

within-one-day localization. S2 and S3 alone achieved exact-day values of

$$66.67\%$$

and

$$61.90\%,$$

respectively.

When all four sensors and all eight descriptor families were combined, exact centroid localization increased to

$$\frac{18}{21} \times 100 = 85.71\%,$$

and all reference signatures were assigned either to their correct day or to an immediately adjacent day:

$$\text{Accuracy}_{\pm 1} = 100\%.$$

The mean absolute localization error in the complete SRS space was

$$MAE = 0.143 \text{ days}.$$

This improvement over most individual sensor and descriptor subsets demonstrates that the discriminative information was complementary. No single feature family fully represented storage progression, whereas the combined SRS incorporated amplitude, cumulative response, kinetic rate, and response timing simultaneously.

The discriminative meaning of the SRS can therefore be interpreted at three levels. First, amplitude-related variables described the increasing overall sensor response generated during storage:

$A,\ AUC,\ F \uparrow$ as storage day increased.

Second, kinetic variables showed that later-storage samples produced faster sensor activation:

$$R_0,\ R_{\max} \uparrow .$$

Third, timing variables represented modifications in curve geometry and saturation behaviour that were not necessarily monotonic but helped distinguish samples with otherwise similar magnitudes:

$$t_{\inf},\ t_{50},\ t_{90} = f\left(\text{response shape and storage stage}\right).$$

Accordingly, the SRS was not simply a compressed version of the final ADC value. It provided a structured description of how the complete sensor response evolved.

The independent blind dataset preserved the same nonlinear response characteristics and remained within the multivariate domain represented by the reference signatures. Since all blind trajectories were processed using the same Richards model and reference-derived scaling parameters, successful blind identification demonstrated that the discriminative relationships encoded by the SRS were transferable to previously unseen measurements. The high similarity between matched reference and blind trajectories further indicates that the selected descriptors represented stable response properties rather than fitting artefacts.

Overall, the discriminative power of the Sensor Response Signatures was primarily associated with amplitude, integrated response, and kinetic-rate descriptors. The median absolute correlation with storage day was 0.875, while the median storage-day effect size was 0.851. AUC, final response, R_0, $R_{\max}$, and fitted amplitude exhibited particularly strong storage associations and very large Day 1–Day 7 standardized differences. Timing descriptors had weaker direct correlations but contributed complementary nonlinear information concerning curve geometry. The balanced contributions of the four sensors and the improved localization obtained from the complete 32-dimensional representation confirm that storage-day discrimination arose from the combined nonlinear response pattern rather than from any single sensor measurement or isolated endpoint feature.

4.4. Reliability of Blind Storage-Day Identification

The reliability of the proposed Sensor Response Signature (SRS) approach was assessed using the independent blind dataset after the nonlinear-response model, descriptor set, reference scaling parameters, and matching procedure had been finalized. The blind dataset contained 21 tomato samples measured with the same four-sensor electronic-nose system used to construct the reference library. Each sample consisted of 51 observations recorded at 2-s intervals over 100 s, producing

$$N_{\text{blind}} = 21$$

blind sample signatures and

$$N_{\text{curves,blind}} = 21 \times 4 = 84$$

independent blind sensor trajectories.

The blind samples were not used during nonlinear-model selection, reference-library construction, feature standardization, or definition of the similarity criterion. Each blind trajectory was independently fitted with the selected Richards response model, and the same eight response descriptors were extracted from each sensor:

$$\boldsymbol{\phi}_{bs} = \left[A_{bs}, R_{0,bs}, R_{\max,bs}, t_{\inf,bs}, t_{50,bs}, t_{90,bs}, AUC_{bs}, F_{bs}\right].$$

The four sensor-specific vectors were concatenated to form the blind SRS:

$$\mathbf{s}_b = [\phi_{b1}, \phi_{b2}, \phi_{b3}, \phi_{b4}] \in \mathbb{R}^{32}.$$

To eliminate differences in numerical scale among amplitude, rate, time, and cumulative-response descriptors, each blind feature was standardized using only the mean and standard deviation of the reference library:

$$z_{bk}^{\text{blind}} = \frac{s_{bk}^{\text{blind}} - \mu_k^{\text{ref}}}{\sigma_k^{\text{ref}}}.$$

No statistical information from the blind dataset was included in the transformation:

$$\mu_k^{\text{blind}} \notin z_{bk}^{\text{blind}},$$
$$\sigma_k^{\text{blind}} \notin z_{bk}^{\text{blind}}.$$

This procedure ensured that the blind evaluation represented a genuine transfer of the previously established SRS library rather than a recalibration using the test samples.

4.4.1. Blind identification using the complete reference library

Blind storage-day identification was performed by comparing each standardized blind signature with every standardized sample-level signature in the reference library. The Euclidean distance between blind sample *b*and reference sample *i*was calculated as

$$D_{bi} = \sqrt{\sum_{k=1}^{32} \left(z_{bk}^{\text{blind}} - z_{ik}^{\text{ref}} \right)^2}.$$

The nearest reference signature was determined using

$$i_b^* = \arg\min \, i \in \{1,\ldots,21\} D_{bi}.$$

The storage day assigned to the blind sample was the known day of this nearest reference signature:

$$\hat{d}_b = d_{i_b^*}.$$

Following unblinding according to the experimental sample key, all 21 blind samples were assigned to their correct storage days. The predicted sequence was

$$1,2,3,4,5,6,7,1,2,3,4,5,6,7,1,2,3,4,5,6,7,$$

which was identical to the concealed true-day sequence of the three independent blind sample groups.

The exact-day identification accuracy was therefore

$$\text{Accuracy}_{\text{exact}} = \frac{\sum_{b=1}^{21} I\left(\hat{d}_b = d_b\right)}{21} \times 100 = 100\%.$$

The corresponding confusion matrix was completely diagonal:

$$\mathbf{C} = \begin{bmatrix} 3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 3 \end{bmatrix}.$$

Each storage-day class contained three independent blind samples. Consequently, class-specific recall was

$$\text{Recall}_{\,d} = 1.000, d = 1,\ldots,7,$$

and class-specific precision was also

$$\text{Precision}_{\,d} = 1.000.$$

The F1 score for every class was

$$F1_d = 2\frac{\text{Precision}_{\,d}\text{Recall}_{\,d}}{\text{Precision}_{\,d} + \text{Recall}_{\,d}} = 1.000.$$

Because the seven classes were balanced, the macro-, weighted-, and micro-averaged metrics were identical:

$$\text{Precision}_{\,\text{macro}} = \text{Recall}_{\,\text{macro}} = F1_{\text{macro}} = 1.000.$$

Agreement beyond chance was evaluated using Cohen's kappa:

$$\kappa = \frac{p_o - p_e}{1 - p_e}.$$

The resulting value was

$$\kappa = 1.000,$$

indicating complete agreement between the predicted and true blind storage-day labels.

4.4.2. Temporal prediction error

Because storage day is an ordinal variable, reliability was also evaluated using the numerical distance between the predicted and actual days. The prediction error for blind sample *b*was defined as

$$e_b = \hat{d}_b - d_b.$$

All blind samples produced

$$e_b = 0.$$

The mean absolute error was therefore

$$MAE = \frac{1}{21}\sum_{b=1}^{21} | \hat{d}_b - d_b | = 0.000 \text{ days}.$$

The root-mean-square error was

$$RMSE = \sqrt{\frac{1}{21}\sum_{b=1}^{21}\left(\hat{d}_b - d_b\right)^2} = 0.000 \text{ days.}$$

The maximum absolute prediction error was

$$E_{\max} = \max_b \mid \hat{d}_b - d_b \mid = 0 \text{ days.}$$

The within-one-day accuracy was calculated as

$$\text{Accuracy}_{\pm 1} = \frac{\sum_{b=1}^{21} I\left(\mid \hat{d}_b - d_b \mid \leq 1\right)}{21} \times 100.$$

Since all predictions were exact,

$$\text{Accuracy}_{\pm 1} = 100\%.$$

These results demonstrate that the blind identification procedure produced neither adjacent-day errors nor larger chronological misclassifications.

4.4.3. Distance-based identification confidence

Classification accuracy alone does not indicate whether a prediction was obtained from a clear nearest-reference match or from an ambiguous decision between two similar signatures. Identification confidence was therefore examined using the distance to the nearest and second-nearest reference samples.

Let

$$D_b^{(1)}$$

denote the smallest reference distance and

$$D_b^{(2)}$$

the second-smallest distance for blind sample b. Across the 21 blind samples, the mean nearest-reference distance was

$$\overline{D^{(1)}} = 0.110,$$

and the median was

$$\tilde{D}^{(1)} = 0.057.$$

The nearest-reference distances ranged from

$$D_{\min} = 0.036$$

to

$$D_{\max} = 0.759.$$

The very small median distance indicates that most blind signatures were located extremely close to a corresponding labelled signature in the standardized 32-dimensional SRS space.

The absolute confidence margin was defined as

$$M_b = D_b^{(2)} - D_b^{(1)}.$$

The mean and median margins were

$$\overline{M} = 2.778$$

and

$$\tilde{M} = 2.531,$$

respectively. Even the smallest observed margin was

$$M_{\min} = 1.599.$$

Thus, no blind sample produced a near-tie between its best and second-best reference matches.

A relative confidence ratio was also calculated:

$$Q_b = \frac{D_b^{(2)}}{D_b^{(1)}}.$$

The median ratio was

$$\tilde{Q} = 39.819,$$

while the minimum ratio was

$$Q_{\min} = 6.941.$$

The maximum ratio reached

$$Q_{\max} = 79.225.$$

Therefore, even the least decisive blind sample was approximately seven times closer to its selected reference signature than to the second-nearest reference signature. This result shows that the 100% blind accuracy was supported by substantial geometric separation and did not result from marginal nearest-neighbour decisions.

4.4.4. Day-specific reproducibility

The mean nearest-reference distances were also examined separately for each storage day. The resulting values were

$$\overline{D}_1 = 0.337,$$

$$\overline{D}_2 = 0.078,$$

$$\overline{D}_3 = 0.152,$$

$$\overline{D}_4 = 0.052,$$

$$\overline{D}_5 = 0.051,$$

$$\overline{D}_6 = 0.056,$$

and

$$\overline{D}_7 = 0.045.$$

The largest average distance occurred for Day 1. This was mainly influenced by one early-storage blind sample with a nearest-reference distance of 0.759. Nevertheless, this sample remained correctly identified and retained a large confidence ratio of

$$Q = 6.941.$$

The particularly small distances observed for Days 4–7 indicate highly reproducible mid- and late-storage response signatures. The mean nearest-reference distance for Day 7 was only

$$0.045,$$

despite the substantially higher signal amplitudes observed at this stage. Therefore, the SRS representation remained stable over the complete response range and did not lose reliability as sensor magnitude increased.

Each blind sample was also matched to the corresponding sample-group and storage-day signature in the reference library. Blind samples 1–7 were nearest to the Day 1–Day 7 signatures of the first reference group, samples 8–14 were nearest to the corresponding signatures of the second group, and samples 15–21 were nearest to those of the third group. This finding indicates that the SRS preserved both the general storage-day progression and the finer group-specific response characteristics.

4.4.5. Coverage of the reference feature space

The reliability of blind identification also depends on whether the test signatures remain within the feature ranges represented by the reference library. For each feature k, a blind value was considered covered when

$$s_{k,\min}^{\text{ref}} \le s_{bk}^{\text{blind}} \le s_{k,\max}^{\text{ref}}.$$

The feature-coverage ratio for blind sample $b$was calculated as

$$C_b = \frac{\sum_{k=1}^{32} I\left[s_{k,\min}^{\text{ref}} \le s_{bk}^{\text{blind}} \le s_{k,\max}^{\text{ref}}\right]}{32}.$$

Across all blind samples and SRS components, the overall reference-range coverage was

$$C_{\text{overall}} = 95.54\%.$$

The mean sample-level coverage was also

$$95.54\%,$$

whereas the median sample-level coverage was

100%.

Fifteen of the 21 blind samples had all 32 features within the corresponding global reference ranges:

$$\frac{15}{21} \times 100 = 71.43\%.$$

Eighteen samples had at least 90% feature coverage:

$$\frac{18}{21} \times 100 = 85.71\%.$$

The lowest coverage occurred for one Day 7 sample, for which

$$C_b = 50\%.$$

Despite the lower univariate range coverage of this sample, its nearest-reference distance was only

$$0.055,$$

and its confidence ratio was

$$41.796.$$

This apparent difference arises because univariate range coverage evaluates each descriptor independently and does not account for the covariance structure of the complete SRS. A value may lie slightly outside the minimum–maximum interval of an individual feature while the overall 32-dimensional response pattern remains extremely close to a valid reference signature. Therefore, multivariate distance provided a more appropriate measure of domain compatibility than isolated feature-range inclusion.

4.4.6. Comparison with centroid-based identification

To assess whether the perfect blind performance depended on retaining individual sample signatures, the blind samples were also compared with the seven average storage-day centroids. The centroid of day *d*was calculated as

$$\boldsymbol{\mu}_d = \frac{1}{3}\sum_{r=1}^{3} \mathbf{z}_{rd}.$$

The centroid-based prediction was

$$\hat{d}_b^{\text{centroid}} = \arg\min_d \| \mathbf{z}_b^{\text{blind}} - \boldsymbol{\mu}_d \|_2.$$

This simplified approach correctly identified

$$\frac{17}{21} \times 100 = 80.95\%$$

of the blind samples exactly. However, all 21 samples were assigned either to the correct day or to an adjacent storage day:

$$\text{Accuracy}_{\pm 1,\text{centroid}} = 100\%.$$

The centroid-based mean absolute error was

$$MAE_{\text{centroid}} = 0.190 \text{ days},$$

and the corresponding RMSE was

$$RMSE_{\text{centroid}} = 0.436 \text{ days}.$$

The mean and median distances to the selected day centroids were

$$\overline{D}_{\text{centroid}} = 2.154$$

and

$$\tilde{D}_{\text{centroid}} = 2.188,$$

respectively.

The reduction from 100% sample-level library accuracy to 80.95% centroid accuracy demonstrates that averaging the three reference signatures of each day removed useful group-specific information. A storage-day centroid represents the central tendency of the class but does not preserve the full biological and instrumental variability observed within that day. The complete reference library was therefore more reliable because it allowed each blind sample to match the most similar experimentally observed response pattern.

The two identification strategies can be contrasted as

$$\hat{d}_b^{\text{library}} = d_{\arg\min i \| \mathbf{z}_b^{\text{blind}} - \mathbf{z}_i^{\text{ref}} \|_2},$$

and

$$\hat{d}_b^{\text{centroid}} = \arg\min d \| \mathbf{z}_b^{\text{blind}} - \boldsymbol{\mu}_d \|_2.$$

The first formulation retained all 21 experimentally observed reference signatures, whereas the second compressed the reference library into only seven class means. The superior blind accuracy of the sample-level approach confirms that within-day variation represented meaningful structure rather than random measurement noise.

4.4.7. Multivariate domain consistency

The blind samples were additionally evaluated relative to the observed radius of their nearest storage-day centroid. For class d, the maximum reference radius was defined as

$$\rho_d = \max_{i:d_i=d} \| \mathbf{z}_i^{\text{ref}} - \boldsymbol{\mu}_d \|_2.$$

The normalized centroid-distance ratio of blind sample $b$was

$$R_b = \frac{\| \mathbf{z}_b^{\text{blind}} - \boldsymbol{\mu}_{\hat{d}_b} \|_2}{\rho_{\hat{d}_b}}.$$

The median normalized ratio was

$$\tilde{R} = 0.764,$$

and the mean ratio was

$$\bar{R} = 0.761.$$

The observed values ranged from

$$0.417$$

to

$$1.012.$$

Nineteen of the 21 blind samples were fully contained within the maximum reference radius of their nearest centroid:

$$\frac{19}{21} \times 100 = 90.48\%.$$

The remaining two samples exceeded the corresponding maximum radius only marginally, with ratios of

$$1.002$$

and

$$1.012.$$

These values are sufficiently close to unity to indicate boundary-level rather than substantial out-of-domain behaviour. Both samples were nevertheless matched correctly to individual reference signatures with very small nearest-neighbour distances. This again demonstrates that the sample-level library represented the blind data more accurately than a spherical class-radius criterion based on only three samples per storage day.

4.4.8. Interpretation of blind-test reliability

The independent blind analysis provided several mutually consistent indicators of reliability. First, all 21 concealed samples were assigned to the correct storage day, yielding 100% exact-day accuracy, 100% within-one-day accuracy, and a Cohen's kappa of 1.000. Second, both MAE and RMSE were zero, indicating the absence of any temporal prediction error. Third, the minimum confidence margin remained substantially greater than zero, and every blind sample was markedly closer to its selected reference signature than to its nearest competitor. Fourth, 95.54% of all blind SRS values were contained within the corresponding reference feature ranges, while the multivariate distance analysis showed close compatibility with the reference library even for the few values outside individual univariate intervals.

The results also demonstrate that the reliability of the proposed method depended on retaining the complete sample-level reference library. When the three reference signatures of each day were averaged into a single centroid, exact blind accuracy decreased to 80.95%, although chronological accuracy remained high. The difference shows that blind identification was supported by reproducible sample-group-specific kinetic structures that would have been lost through excessive class averaging.

Nevertheless, the perfect performance should be interpreted within the limits of the experimental design. The blind data were independently withheld from model construction, scaling, and reference matching, but they were generated using the same electronic-nose device, sensor array, chamber configuration, sampling interval, measurement duration, and controlled experimental protocol. The results therefore establish excellent repeatability and independent identification reliability under the evaluated conditions. Further validation using different harvest batches, seasons, devices, environmental conditions, and sensor-ageing states would be required before claiming complete external generalizability.

Overall, the blind-test results demonstrate that the Richards-derived SRS library provided a highly reliable representation of tomato storage progression. The method correctly identified all 21 independent samples without any adjacent-day or non-adjacent-day error. The very small nearest-reference distances, large confidence margins, high reference-space coverage, and complete agreement metrics confirm that the blind decisions were both accurate and geometrically well supported. These findings establish the sample-level SRS library as a robust basis for independent storage-day identification under controlled electronic-nose measurement conditions.

4.5. Discussion

The results demonstrate that storage-related changes in the tomato samples were encoded not only in the magnitude of the electronic-nose responses but also in the complete temporal geometry of the sensor curves. The four MOS sensors generated smooth, almost entirely monotonic trajectories over the 100-s measurement interval. Across both the reference and independent blind datasets, these trajectories exhibited an initial response-development phase followed by a gradual reduction in response rate as the signals approached their late-stage region. This behaviour supports the treatment of the electronic-nose output as a continuous nonlinear process rather than as a set of independent ADC measurements.

For sample *i*and sensor *s*, the observed sequence can be written as

$$\mathbf{y}_{is} = \left[y_{is}(0), y_{is}(2), \ldots, y_{is}(100)\right],$$

whereas the proposed approach interprets the same sequence as observations from an underlying continuous function:

$$y_{is}(t) \approx \hat{y}_{is}(t), 0 \le t \le 100.$$

This distinction is important because conventional endpoint-based analysis uses only a restricted portion of the available information. For example, a final response feature is defined as

$$F_{is} = y_{is}(100),$$

whereas the nonlinear representation retains the response amplitude, rate, curvature, characteristic times, and cumulative response. Consequently, two samples having similar final ADC values may still be distinguishable through differences in their response development:

$$F_{is} \approx F_{js},$$

while

$$R_{0,is} \neq R_{0,js},$$
$$t_{50,is} \neq t_{50,js},$$

or

$$AUC_{is} \neq AUC_{js}.$$

The comparative model analysis showed that the Richards function provided the most appropriate mathematical representation of the temporal sensor responses. Although the Logistic and Gompertz models reproduced the general increasing trend, both imposed structural restrictions on curve shape. The Logistic function assumes symmetry around its inflection region, whereas the Gompertz model provides a fixed form of asymmetry. The experimental sensor curves, however, displayed variations in early acceleration, inflection position, and late-stage curvature among storage days and sensor channels. These differences required a more flexible function.

The additional Richards shape parameter allowed the fitted response to adapt to these trajectory-specific differences:

$$\hat{y}(t) = L + \frac{U - L}{\left[1 + \nu \exp\ \left[-k(t - t_0)\right]\right]^{1/\nu}}.$$

The importance of this flexibility was reflected in the fitting results. The Richards model achieved a median coefficient of determination close to unity:

$$\widetilde{R^2}_{\text{Richards}} \approx 0.9998,$$

while its median RMSE was approximately

$$\widetilde{RMSE}_{\text{Richards}} \approx 2.3 \text{ ADC}.$$

The corresponding errors for the Logistic and Gompertz models were approximately 4.9 and 6.4 ADC, respectively. Thus, the Richards model reduced the median fitting error by more than half relative to Logistic and by almost two-thirds relative to Gompertz.

The relative improvement over the Logistic model was

$$\frac{RMSE_{\text{Logistic}} - RMSE_{\text{Richards}}}{RMSE_{\text{Logistic}}} \times 100 \approx 53\%,$$

whereas the improvement over the Gompertz model was

$$\frac{RMSE_{\text{Gompertz}} - RMSE_{\text{Richards}}}{RMSE_{\text{Gompertz}}} \times 100 \approx 64\%.$$

The superiority of Richards was retained after penalizing additional model complexity through AICc and BIC. This is a critical result because an additional parameter can always reduce residual error to some extent. However, the substantially lower information-criterion values showed that the improvement was too large to be explained merely by greater flexibility. Richards was preferred for approximately 94% of the curves according to the penalized information criteria and for nearly 99% according to RMSE. The same ranking was reproduced in the blind dataset, indicating that the selected model represented a repeatable sensor-response mechanism rather than an incidental pattern specific to the reference samples.

The close agreement between reference and blind fitting performance was particularly important. The difference between their median Richards RMSE values was negligible:

$$\Delta RMSE_{\text{median}} = \left| RMSE_{\text{blind}} - RMSE_{\text{ref}} \right| \approx 0.002 \text{ ADC}.$$

In relative terms,

$$\frac{\Delta RMSE_{\text{median}}}{RMSE_{\text{ref}}} \times 100 < 0.1\%.$$

Such consistency indicates that the blind temporal curves had essentially the same nonlinear structure as the labelled reference curves. Therefore, the subsequent success of blind storage-day identification cannot be attributed to a difference in fitting procedures or to a deterioration of model quality in the independent data.

The nonlinear modelling results also showed that storage progression influenced several complementary components of the response. The strongest changes occurred in amplitude, cumulative response, and kinetic rate. For example, the final S1 response increased from approximately 180 ADC on Day 1 to 771 ADC on Day 7. The relative change was

$$\frac{771-180}{180}\times 100 \approx 328\%.$$

Comparable increases were observed in S2, S3, and S4. The cumulative responses showed similarly pronounced changes, with the S1 AUC increasing by approximately 352% between the first and seventh storage days.

These increases suggest that the concentration and composition of volatile compounds interacting with the MOS sensing layer changed progressively during storage. The electronic nose did not directly identify individual volatile compounds; nevertheless, the systematic increase in signal amplitude indicates a growing overall sensor stimulus. This stimulus may reflect the combined influence of respiration-related volatiles, tissue metabolism, ripening, senescence, microbial activity, and the accumulation of degradation products. Because the sensors were cross-sensitive rather than compound-specific, the four-channel response should be interpreted as a collective volatile pattern rather than as a direct measurement of a single chemical species.

The early-stage response rates also increased strongly with storage day. The initial S1 slope increased from approximately

$$1.62\ \mathrm{ADC\,s^{-1}}$$

on Day 1 to

$$7.36\ \mathrm{ADC\,s^{-1}}$$

on Day 7. Therefore, later-storage samples produced both stronger and faster responses. This finding is analytically important because a change in final response alone could arise from a simple vertical shift in signal magnitude. The simultaneous increase in response rate demonstrates that the temporal dynamics of sensor activation also changed.

The Sensor Response Signature was designed to preserve these complementary effects. For each sensor, the signature combined magnitude-, rate-, timing-, and area-related descriptors:

$$\boldsymbol{\phi}_{is} = \left[A_{is}, R_{0,is}, R_{\max,is}, t_{\inf,is}, t_{50,is}, t_{90,is}, AUC_{is}, F_{is}\right].$$

The complete sample signature was

$$\mathbf{s}_i \in \mathbb{R}^{32}.$$

This conversion reduced the original dimensionality from 204 raw variables per sample to 32 interpretable descriptors:

$$D_{\mathrm{raw}} = 51 \times 4 = 204,$$
$$D_{\mathrm{SRS}} = 8 \times 4 = 32.$$

The dimensionality reduction was therefore

$$\left(1-\frac{32}{204}\right)\times 100 = 84.31\%.$$

This reduction is not merely a computational advantage. The raw time points are strongly autocorrelated because neighbouring measurements originate from the same smooth response curve. Treating all 204 values as independent input variables introduces substantial redundancy and produces an unfavourable feature-to-sample ratio:

$$\frac{D_{\mathrm{raw}}}{N_{\mathrm{ref}}} = \frac{204}{21} = 9.71.$$

In contrast, the SRS reduced this ratio to

$$\frac{D_{\mathrm{SRS}}}{N_{\mathrm{ref}}} = \frac{32}{21} = 1.52.$$

Although 32 variables remain relatively numerous for 21 reference samples, these descriptors are mechanistically interpretable and summarize distinct aspects of the curves. The SRS therefore represents a more structured and less redundant input space than the raw temporal matrix.

The discriminative analyses indicated that storage information was distributed across several descriptor families. The median absolute feature–day correlation was approximately

$$\mathrm{median}\ (|r|) = 0.875,$$

and the median storage-day effect size was

$$\tilde{\eta}^2 = 0.851.$$

These values indicate that most SRS dimensions were strongly affected by storage duration. AUC, final response, initial rate, maximum rate, and fitted amplitude were the most consistently informative features.

The strongest single relationships were associated with S3 amplitude and S3 initial response rate, for which the correlations with storage day approached

$$r = 0.95.$$

However, no single sensor or descriptor family completely represented the storage pattern. S1 and S2 exhibited strong overall effects, S3 contained several of the strongest individual features, and S4 contributed highly reproducible cumulative and final-response information. The discrimination was therefore based on a coordinated multichannel response rather than on one dominant sensor.

This interpretation was supported by principal component analysis. The first two components explained approximately

$$78\%$$

of the total standardized SRS variance, while the first three explained approximately

$$86\%.$$

The dominant component contained substantial contributions from AUC, final response, and rate features across multiple sensors. Therefore, the main storage-related direction in the SRS space represented a collective change across the sensor array. At the same time, the remaining variance contained additional class- and sample-specific information that contributed to exact blind matching.

The timing descriptors showed weaker direct linear correlations with storage day than the amplitude and rate descriptors. However, their lower univariate correlations should not be interpreted as evidence that they were unnecessary. A feature may have limited correlation with an ordinal day label yet still improve multivariate discrimination through nonlinear or interaction effects. Characteristic times describe the relative positioning of different response regions and may distinguish samples having similar amplitudes but different curve shapes.

The complementary role of timing can be represented as

$$P(d \mid A,F,AUC) \neq P\left(d \mid A,F,AUC,t_{\mathrm{inf}},t_{50},t_{90}\right).$$

The complete signature consequently incorporated both strongly monotonic storage markers and subtler shape-related descriptors.

The most important result was the performance on the independent blind dataset. All 21 blind samples were correctly assigned to their storage days using nearest-reference matching in the standardized SRS space:

$$\text{Accuracy}_{\text{blind}} = \frac{21}{21} \times 100 = 100\%.$$

The confusion matrix was fully diagonal, and all class-specific precision, recall, and F1 values were equal to 1.000. Cohen's kappa was

$$\kappa = 1.000,$$

and the ordinal prediction errors were

$$MAE = 0,$$
$$RMSE = 0.$$

The absence of both adjacent-day and non-adjacent-day errors indicates that the blind classifications reproduced the complete chronological ordering represented in the reference library.

The distance structure provided additional evidence that the perfect accuracy did not result from unstable or marginal decisions. The median distance to the nearest reference signature was approximately

$$\tilde{D}^{(1)} = 0.057,$$

whereas the median ratio of the second-nearest to the nearest distance was approximately

$$\tilde{Q} = 39.8.$$

Even the smallest observed ratio remained approximately

$$Q_{\min} = 6.9.$$

Thus, every blind signature was substantially closer to its selected reference sample than to its nearest alternative. The classification decisions were therefore supported by clear local separation in the standardized feature space.

The extremely small differences between matched reference and blind curves further explain this result. Across all matched time points and sensors, the mean absolute raw-response difference was approximately

$$MAE_{\text{raw}} = 0.989 \text{ ADC},$$

and the corresponding RMSE was approximately

$$RMSE_{\text{raw}} = 0.994 \text{ ADC}.$$

The median matched-curve correlation was effectively

$$r = 1.000.$$

This indicates exceptionally high experimental repeatability. Nevertheless, such close numerical similarity must be considered when interpreting the blind-test performance. The blind set was independently excluded from reference construction and label assignment, but it was acquired using the same instrument, sensor array, measurement configuration, duration, sampling interval, and controlled protocol. It therefore provides strong evidence of repeatability and internal independent validation, but it does not by itself establish universal external validity.

Another important finding was the difference between sample-level reference matching and storage-day centroid matching. When the three reference signatures associated with each storage day were averaged into a single class centroid, blind exact-day accuracy decreased from 100% to approximately 80.95%, although all errors remained within one day.

The centroid for day $d$was

$$\boldsymbol{\mu}_d = \frac{1}{3}\sum_{r=1}^{3} \mathbf{z}_{rd}.$$

Centroid matching assumes that all within-day variation is undesirable noise. However, the results indicate that some of this variation represented stable group-specific response structure. The complete sample-level library retained these patterns:

$$\mathscr{L}_{\text{sample}} = \left\{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_{21}\right\},$$

whereas centroid compression reduced the library to

$$\mathscr{L}_{\text{centroid}} = \left\{\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_7\right\}.$$

The improvement obtained by retaining individual references indicates that biological or sample-group variation should not always be eliminated through averaging. In small reference libraries, sample-level prototypes may provide a more faithful representation of the possible response domain than a single mean vector per class.

The comparison with a raw-data SVM benchmark further illustrates this point. The optimized RBF-SVM achieved approximately

$$71.43\%$$

Leave-One-Producer-Out accuracy on the reference set and

$$90.48\%$$

accuracy on the independent blind set. Its blind predictions included two errors involving Day 4, one assigned to Day 5 and one to Day 6. The SRS library improved blind exact accuracy by

$$100 - 90.48 = 9.52$$

percentage points and eliminated all blind errors.

The relative error reduction was

$$\frac{E_{\text{SVM}} - E_{\text{SRS}}}{E_{\text{SVM}}} \times 100 = 100\%.$$

The superior SRS performance likely arose from three related factors. First, nonlinear fitting smoothed small point-level variations and quantization effects. Second, the 32 descriptors reduced temporal redundancy. Third, nearest-reference matching preserved experimentally observed sample-specific structures without requiring estimation of a complex global classification boundary from only 21 labelled cases.

However, the comparison should not be interpreted as evidence that SRS matching is universally superior to machine learning. The reference sample size was small, and an SVM operating in a 204-dimensional raw space is particularly vulnerable to unstable boundary estimation. With a substantially larger and more heterogeneous training dataset, regularized machine-learning or deep temporal models may offer additional advantages. The present result instead shows

that, under small-sample conditions, physically interpretable nonlinear feature engineering can be more reliable than applying a high-dimensional classifier directly to raw sensor sequences.

Several limitations should therefore be considered. The first is the limited number of independent reference units:

$$N_{\text{ref}} = 21,$$

with only three reference samples per storage day:

$$n_d = 3.$$

This design was sufficient to demonstrate systematic temporal progression and blind repeatability, but it cannot fully characterize the biological variation expected across cultivars, harvest lots, maturity stages, seasons, growing conditions, or storage environments.

Second, the experiment was conducted using one electronic-nose configuration. MOS sensors are affected by temperature, relative humidity, ageing, baseline drift, contamination, and device-to-device variability. A reference library developed using one sensor array may therefore require transfer correction or recalibration before application on another device or after prolonged sensor use.

A general drift-adjusted response could be represented as

$$y_{\text{corrected}}(t) = \frac{y(t) - b_{\text{session}}}{g_{\text{session}}},$$

where $b_{\text{session}}$and g_{session}denote session-specific baseline and gain corrections. Such standardization was not required for the present controlled experiment but may become essential in long-term or field applications.

Third, the study identified storage day rather than directly measuring chemical or microbiological quality. Storage duration is an operational label and may not correspond perfectly to identical freshness states under different temperatures or handling conditions. Two tomatoes stored for the same number of days may exhibit different physiological or microbial states if their preharvest histories or storage environments differ. Future research should therefore relate the SRS descriptors to independent quality indicators such as firmness, mass loss, colour, soluble solids, acidity, microbial load, or volatile profiles.

The practical interpretation of the predicted class should consequently remain

$$\hat{d} = \text{estimated storage - day class under the defined protocol},$$

rather than

$$\hat{d} = \text{universal measure of freshness}.$$

Fourth, the nearly identical reference and blind trajectories produced an unusually favourable independent-test condition. This is scientifically valuable for demonstrating measurement repeatability and algorithmic reproducibility, but a more demanding external validation should involve new tomato batches measured on different dates, preferably with independently prepared samples and controlled variations in environmental conditions.

A stronger validation hierarchy would include

$$\begin{aligned}&\text{within - session repeatability} < \text{between - session reproducibility} < \text{between - batch validation}\\ &< \text{between - device validation}.\end{aligned}$$

The present blind test mainly provides evidence at the first two levels. Further studies are required to evaluate the latter levels.

Despite these limitations, the proposed framework has several important advantages. It is non-destructive, uses a short 100-s measurement period, requires no chromatographic or spectroscopic instrumentation, and produces interpretable parameters related to the physical form of the sensor response. The nonlinear fitting and signature-generation stages can be automated, allowing an unknown sample to be processed through the sequence

$$\text{temporal response} \rightarrow \text{Richards fitting} \rightarrow \text{SRS extraction} \rightarrow \text{reference standardization} \rightarrow \text{nearest - signature matching} \rightarrow \hat{d}.$$

Because the system uses only four sensor channels and compresses each sample into 32 descriptors, it is potentially compatible with low-cost embedded implementation. However, practical deployment would require automated curve-quality checks. A new measurement should not be classified when model convergence is poor or when the resulting signature lies outside the validated reference domain.

A possible acceptance condition may be expressed as

$$R^2 \geq R_{\min},$$
$$RMSE \leq RMSE_{\max},$$

and

$$D^{(1)} \leq D_{\text{accept}}.$$

Predictions not satisfying these criteria could be labelled as uncertain rather than being forced into one of the seven storage-day classes. Incorporating such an out-of-distribution rule would improve the safety of future field applications.

Overall, the findings show that nonlinear kinetic modelling transformed the MOS sensor outputs from high-dimensional, strongly correlated time series into compact and discriminative response signatures. The Richards function represented the experimental trajectories more accurately than the Logistic and Gompertz alternatives, and its performance transferred without deterioration to the blind dataset. Storage progression was expressed through coordinated changes in response amplitude, cumulative exposure, activation rate, and curve timing. The resulting SRS representation reduced dimensionality by 84.31% while retaining sufficient information to distinguish all seven storage days.

Most importantly, the independent blind analysis achieved complete exact-day identification with zero temporal error and substantial distance-based confidence. These results support the central premise of the study: the temporal shape of an electronic-nose response contains reproducible storage information that is not fully captured by endpoint measurements or conventional raw-data classification. Within the controlled conditions evaluated here, the sample-level SRS library provided a reliable, interpretable, and computationally efficient framework for blind storage-day identification.

5. Conclusions

This study demonstrated that temporal electronic-nose responses can be represented effectively through nonlinear kinetic modelling and converted into compact Sensor Response Signatures for tomato storage-day identification. Among the evaluated functions, the Richards model provided the most accurate and consistent representation of the sensor trajectories. The resulting 32-dimensional signatures preserved the principal amplitude, rate, timing, and cumulative-response characteristics while reducing the original input dimensionality by 84.31%.

The proposed sample-level reference library correctly identified all 21 independent blind samples, achieving 100% exact-day accuracy, a Cohen's kappa of 1.000, and zero MAE and RMSE. These findings show that storage-related information is embedded in the complete temporal structure of the sensor response rather than in isolated endpoint measurements alone.

Under the controlled conditions of this study, the developed framework provides a rapid, non-destructive, interpretable, and computationally efficient approach for blind storage-day identification. Further validation using different harvest batches, seasons, storage conditions, and electronic-nose devices is required to establish broader practical applicability.

6. Acknowledgements

The authors would like to thank all individuals who contributed to the experimental measurements and technical support during the development of the electronic nose system.

7. Fundıng

This research received no external funding.

8. Instıtutıonal Revıew Board Statement

Not applicable.

9. Informed Consent Statement

Not applicable.

10. Data Avaılabılıty Statement

The datasets generated and analysed during the current study are available from the corresponding author on reasonable request.

11. Confl2cts Of Interest

The authors declare no conflict of interest.

12. Hıghlıghts

- A low-cost passive-diffusion electronic nose was developed for strawberry freshness monitoring.
- Five complementary kinetic and steady-state features were extracted from MOS sensor responses.
- Dynamic response descriptors significantly improved storage-stage discrimination.
- Statistical analysis confirmed strong correlations between sensor responses and storage time.
- The proposed framework enables rapid, non-destructive, and intelligent freshness assessment.

13. REFERENCES

Sanislav, T., Mois, G. D., Zeadally, S., Folea, S., Radoni, T. C., & Al-Suhaimi, E. A. (2025). A comprehensive review on sensor-based electronic nose for food quality and safety. *Sensors, 25*(14), Article 4437. https://doi.org/10.3390/s25144437

Vadera, N., & Dhanekar, S. (2025). Classification and prediction of VOCs using an IoT-enabled electronic nose system-based lab prototype for breath sensing applications. *ACS Sensors, 10*(1), 439–447. https://doi.org/10.1021/acssensors.4c02731

Wang, B., Li, Y., Liu, K., Wei, G., He, A., Kong, W., & Zhang, X. (2024). Intelligent evaluation and dynamic prediction of oysters freshness with electronic nose non-destructive monitoring and machine learning. *Biosensors, 14*(10), Article 502. https://doi.org/10.3390/bios14100502

Zhai, Z., Liu, Y., Li, C., Wang, D., & Wu, H. (2024). Electronic noses: From gas-sensitive components and practical applications to data processing. *Sensors, 24*(15), Article 4806. https://doi.org/10.3390/s24154806

Zhang, R., Li, J., Liu, H., Wang, Y., & Zhang, T. (2024). Optimal gas sensor combination selection method for low-cost electronic nose. *Sensors and Actuators A: Physical, 366*, Article 114904.

Chen, Y., Fu, J., Weng, X., Chen, J., Hu, R., and Zhu, Y. (2024). A feature extractor for temporal data of electronic nose based on parallel long short-term memory network in flavor discrimination of Chinese vinegars. *Journal of Food Engineering, 379*, Article 112132. doi:10.1016/j.jfoodeng.2024.112132

Ijaz, U., Ali, M., Ahmad, I., Hamza, S. A., and Kim, H. D. (2025). A comprehensive review of electronic nose systems: Design, sensors, and future directions. *Chemical Engineering Journal, 524*, Article 169482. doi:10.1016/j.cej.2025.169482

Niu, G., Zhang, Y., Yu, H., Liu, Y., Li, Y., and Chen, X. (2024). Selective discrimination of VOC gases at ppb level using dynamic features extracted from temperature-pulsed MOS gas sensors. *Chemosensors and Intelligent Laboratory Systems, 245*, Article 105062.

Sanislav, T., Mois, G. D., Zeadally, S., Folea, S., Radoni, T. C., and Al-Suhaimi, E. A. (2025). A comprehensive review on sensor-based electronic nose for food quality and safety. *Sensors, 25*(14), Article 4437. doi:10.3390/s25144437

Zhai, Z., Liu, Y., Li, C., Wang, D., and Wu, H. (2024). Electronic noses: From gas-sensitive components and practical applications to data processing. *Sensors, 24*(15), Article 4806. doi:10.3390/s24154806

Zhang, W., Yuan, Z., Zhu, H., Zhang, S., and Meng, F. (2024). Kinetic modeling in temperature-modulated semiconductor gas sensor utilizing Eley–Rideal mechanism and its application in discriminative detection of VOCs. *ACS Sensors, 9*(12), 6460–6470. doi:10.1021/acssensors.4c01835

Chen, Y., Fu, J., Weng, X., Chen, J., Hu, R., & Zhu, Y. (2024). A feature extractor for temporal data of electronic nose based on parallel long short-term memory network in flavor discrimination of Chinese vinegars. *Journal of Food Engineering, 379*, Article 112132. https://doi.org/10.1016/j.jfoodeng.2024.112132

Sanislav, T., Mois, G. D., Zeadally, S., Folea, S., Radoni, T. C., & Al-Suhaimi, E. A. (2025). A comprehensive review on sensor-based electronic nose for food quality and safety. *Sensors, 25*(14), Article 4437. https://doi.org/10.3390/s25144437

Zhai, Z., Liu, Y., Li, C., Wang, D., & Wu, H. (2024). Electronic noses: From gas-sensitive components and practical applications to data processing. *Sensors, 24*(15), Article 4806. https://doi.org/10.3390/s24154806

Zhang, W., Yuan, Z., Zhu, H., Zhang, S., & Meng, F. (2024). Kinetic modeling in temperature-modulated semiconductor gas sensor utilizing Eley–Rideal mechanism and its application in discriminative detection of VOCs. *ACS Sensors, 9*(12), 6460–6470. https://doi.org/10.1021/acssensors.4c01835

Chen, Y., Fu, J., Weng, X., Chen, J., Hu, R., & Zhu, Y. (2024). A feature extractor for temporal data of electronic nose based on parallel long short-term memory network in flavor discrimination of Chinese vinegars. *Journal of Food Engineering, 379*, Article 112132. https://doi.org/10.1016/j.jfoodeng.2024.112132

Ijaz, U., Ali, M., Ahmad, I., Hamza, S. A., & Kim, H. D. (2025). A comprehensive review of electronic nose systems: Design, sensors, and future directions. *Chemical Engineering Journal, 524*, Article 169482. https://doi.org/10.1016/j.cej.2025.169482

Sanislav, T., Mois, G. D., Zeadally, S., Folea, S., Radoni, T. C., & Al-Suhaimi, E. A. (2025). A comprehensive review on sensor-based electronic nose for food quality and safety. *Sensors, 25*(14), Article 4437. https://doi.org/10.3390/s25144437

Zhai, Z., Liu, Y., Li, C., Wang, D., & Wu, H. (2024). Electronic noses: From gas-sensitive components and practical applications to data processing. *Sensors, 24*(15), Article 4806. https://doi.org/10.3390/s24154806

Chen, Y., Fu, J., Weng, X., Chen, J., Hu, R., & Zhu, Y. (2024). *A feature extractor for temporal data of electronic nose based on parallel long short-term memory network in flavor discrimination of Chinese vinegars*. *Journal of Food Engineering, 379*, 112132.

Richards, F. J. (1959). *A flexible growth function for empirical use*. *Journal of Experimental Botany, 10*(2), 290–301.

Sanislav, T., Mois, G. D., Zeadally, S., Folea, S., Radoni, T. C., & Al-Suhaimi, E. A. (2025). *A comprehensive review on sensor-based electronic nose for food quality and safety*. *Sensors, 25*(14), 4437.

Zhai, Z., Liu, Y., Li, C., Wang, D., & Wu, H. (2024). *Electronic noses: From gas-sensitive components and practical applications to data processing*. *Sensors, 24*(15), 4806.

Zhang, W., Yuan, Z., Zhu, H., Zhang, S., & Meng, F. (2024). *Kinetic modeling in temperature-modulated semiconductor gas sensor utilizing Eley–Rideal mechanism and its application in discriminative detection of VOCs*. *ACS Sensors, 9*(12), 6460–6470.

Chen, Y., Fu, J., Weng, X., Chen, J., Hu, R., and Zhu, Y. (2024). A feature extractor for temporal data of electronic nose based on parallel long short-term memory network in flavor discrimination of Chinese vinegars. *Journal of Food Engineering, 379*, Article 112132. https://doi.org/10.1016/j.jfoodeng.2024.112132

Sanislav, T., Mois, G. D., Zeadally, S., Folea, S., Radoni, T. C., and Al-Suhaimi, E. A. (2025). A comprehensive review on sensor-based electronic nose for food quality and safety. *Sensors, 25*(14), Article 4437. https://doi.org/10.3390/s25144437

Zhai, Z., Liu, Y., Li, C., Wang, D., and Wu, H. (2024). Electronic noses: From gas-sensitive components and practical applications to data processing. *Sensors, 24*(15), Article 4806. https://doi.org/10.3390/s24154806

Zhang, W., Yuan, Z., Zhu, H., Zhang, S., and Meng, F. (2024). Kinetic modeling in temperature-modulated semiconductor gas sensor utilizing Eley–Rideal mechanism and its application in discriminative detection of VOCs. *ACS Sensors, 9*(12), 6460–6470. https://doi.org/10.1021/acssensors.4c01835

Richards, F. J. (1959). *A flexible growth function for empirical use*. Journal of Experimental Botany, 10(2), 290–301. https://doi.org/10.1093/jxb/10.2.290

Zhai, Z., Liu, Y., Li, C., Wang, D., & Wu, H. (2024). *Electronic noses: From gas-sensitive components and practical applications to data processing*. Sensors, 24(15), 4806. https://doi.org/10.3390/s24154806

Zhang, W., Yuan, Z., Zhu, H., Zhang, S., & Meng, F. (2024). *Kinetic modeling in temperature-modulated semiconductor gas sensor utilizing Eley–Rideal mechanism and its application in discriminative detection of VOCs*. ACS Sensors, 9(12), 6460–6470. https://doi.org/10.1021/acssensors.4c01835

Chen, Y., Fu, J., Weng, X., Chen, J., Hu, R., & Zhu, Y. (2024). A feature extractor for temporal data of electronic nose based on parallel long short-term memory network in flavor discrimination of Chinese vinegars. *Journal of Food Engineering, 379*, Article 112132. https://doi.org/10.1016/j.jfoodeng.2024.112132

Portet, S. (2020). A primer on model selection using the Akaike information criterion. *Infectious Disease Modelling, 5*, 111–128. https://doi.org/10.1016/j.idm.2019.12.010

Richards, F. J. (1959). A flexible growth function for empirical use. *Journal of Experimental Botany, 10*(2), 290–301. https://doi.org/10.1093/jxb/10.2.290

Sanislav, T., Mois, G. D., Zeadally, S., Folea, S., Radoni, T. C., & Al-Suhaimi, E. A. (2025). A comprehensive review on sensor-based electronic nose for food quality and safety. *Sensors, 25*(14), Article 4437. https://doi.org/10.3390/s25144437

Zhai, Z., Liu, Y., Li, C., Wang, D., & Wu, H. (2024). Electronic noses: From gas-sensitive components and practical applications to data processing. *Sensors, 24*(15), Article 4806. https://doi.org/10.3390/s24154806

Zhang, W., Yuan, Z., Zhu, H., Zhang, S., & Meng, F. (2024). Kinetic modeling in temperature-modulated semiconductor gas sensor utilizing Eley–Rideal mechanism and its application in discriminative detection of VOCs. *ACS Sensors, 9*(12), 6460–6470. https://doi.org/10.1021/acssensors.4c01835

Richards, F. J. (1959). A flexible growth function for empirical use. *Journal of Experimental Botany, 10*(2), 290–301. https://doi.org/10.1093/jxb/10.2.290

Sanislav, T., Mois, G. D., Zeadally, S., Folea, S., Radoni, T. C., & Al-Suhaimi, E. A. (2025). A comprehensive review on sensor-based electronic nose for food quality and safety. *Sensors, 25*(14), Article 4437. https://doi.org/10.3390/s25144437

Zhai, Z., Liu, Y., Li, C., Wang, D., & Wu, H. (2024). Electronic noses: From gas-sensitive components and practical applications to data processing. *Sensors, 24*(15), Article 4806. https://doi.org/10.3390/s24154806

Zhang, W., Yuan, Z., Zhu, H., Zhang, S., & Meng, F. (2024). Kinetic modeling in temperature-modulated semiconductor gas sensor utilizing Eley–Rideal mechanism and its application in discriminative detection of VOCs. *ACS Sensors, 9*(12), 6460–6470. https://doi.org/10.1021/acssensors.4c01835

Chen, Y., Fu, J., Weng, X., Chen, J., Hu, R., & Zhu, Y. (2024). A feature extractor for temporal data of electronic nose based on parallel long short-term memory network in flavor discrimination of Chinese vinegars. *Journal of Food Engineering, 379*, Article 112132.

Lozovskyi, A. S., Lyashkov, A. Y., & Tonkoshkur, A. S. (2024). Analysis of kinetic dependencies of the response of a gas sensor based on a model of a stretched exponential function with the use of computer technologies. *Journal of Physics and Electronics, 32*(1–2), 73–82. https://doi.org/10.15421/332409

Zhai, Z., Liu, Y., Li, C., Wang, D., & Wu, H. (2024). Electronic noses: From gas-sensitive components and practical applications to data processing. *Sensors, 24*(15), Article 4806. https://doi.org/10.3390/s24154806

Zhang, W., Yuan, Z., Zhu, H., Zhang, S., & Meng, F. (2024). Kinetic modeling in temperature-modulated semiconductor gas sensor utilizing Eley–Rideal mechanism and its application in discriminative detection of VOCs. *ACS Sensors, 9*, 6460–6470. https://doi.org/10.1021/acssensors.4c01835

Fernandez, L., Oller-Moreno, S., Fonollosa, J., Garrido-Delgado, R., Arce, L., Martín-Gómez, A., Marco, S., & Pardo, A. (2025). Signal preprocessing in instrument-based electronic noses leads to parsimonious predictive models: Application to olive oil quality control. *Sensors, 25*(3), Article 737. https://doi.org/10.3390/s25030737

Vajdi, M. (2025). Metal oxide semiconductor based electronic nose data pre-processing: Review. *Engineering Applications of Artificial Intelligence, 149*, Article 110540. https://doi.org/10.1016/j.engappai.2025.110540

Sanislav, T., Mois, G. D., Zeadally, S., Folea, S., Radoni, T. C., & Al-Suhaimi, E. A. (2025). A comprehensive review on sensor-based electronic nose for food quality and safety. *Sensors, 25*(14), Article 4437. https://doi.org/10.3390/s25144437

Shtepliuk, I., Domènech-Gil, G., Almqvist, V., Kautto, A. H., Vågsholm, I., Boqvist, S., Eriksson, J., & Puglisi, D. (2025). Electronic nose and machine learning for modern meat inspection. *Journal of Big Data, 12*, Article 91. https://doi.org/10.1186/s40537-025-01151-4

Vajdi, M. (2025). Metal oxide semiconductor based electronic nose data pre-processing: Review. *Engineering Applications of Artificial Intelligence, 149*, Article 110540. https://doi.org/10.1016/j.engappai.2025.110540

Zhai, Z., Liu, Y., Li, C., Wang, D., & Wu, H. (2024). Electronic noses: From gas-sensitive components and practical applications to data processing. *Sensors, 24*(15), Article 4806. https://doi.org/10.3390/s24154806

Feng, L., Zhang, M., Bhandari, B., & Guo, Z. (2018). A novel method using MOS electronic nose and ELM for predicting postharvest quality of cherry tomato fruit treated with high pressure argon. *Computers and Electronics in Agriculture, 154*, 411–419. https://doi.org/10.1016/j.compag.2018.09.032

Sanislav, T., Mois, G. D., Zeadally, S., Folea, S., Radoni, T. C., & Al-Suhaimi, E. A. (2025). A comprehensive review on sensor-based electronic nose for food quality and safety. *Sensors, 25*(14), Article 4437. https://doi.org/10.3390/s25144437

Senge, J. M., Kaltenecker, F., & Krupitzer, C. (2025). Integrating sensor data, laboratory analysis, and computer vision in machine learning-driven E-nose systems for predicting tomato shelf life. *Chemosensors, 13*(7), Article 255. https://doi.org/10.3390/chemosensors13070255

Tyagi, P., Semwal, R., Sharma, A., Tiwary, U. S., & Varadwaj, P. (2023). E-nose: A low-cost fruit ripeness monitoring system. *Journal of Agricultural Engineering, 54*(1). https://doi.org/10.4081/jae.2022.1389

Zhai, Z., Long, H., Feng, Y., Li, Y., & Wang, Y. (2024). Electronic noses: From gas-sensitive components and practical applications to data processing. *Sensors, 24*(15), Article 4806. https://doi.org/10.3390/s24154806

Li, X., Huang, H., Zhang, L., & Zhao, L. (2025). Effect of postharvest storage temperature and duration on tomato fruit quality. *Foods, 14*(6), Article 1002. https://doi.org/10.3390/foods14061002

Tsouvaltzis, P., Gkountina, S., & Siomos, A. S. (2023). Quality traits and nutritional components of cherry tomato in relation to the harvesting period, storage duration and fruit position in the truss. *Plants, 12*(2), Article 315. https://doi.org/10.3390/plants12020315

Zhang, Y., Wu, J., Li, J., Mi, S., Yu, W., Sang, Y., & Wang, X. (2025). Enhancing post-harvest quality of cherry tomatoes via low voltage electrostatic field and modified atmosphere packaging: Regulation of phenylpropanoid metabolism. *LWT, 237*, Article 118782. https://doi.org/10.1016/j.lwt.2025.118782

Zhao, Y., Li, L., Gao, S., Wang, S., Li, X., & Xiong, X. (2023). Postharvest storage properties and quality kinetic models of cherry tomatoes treated by high-voltage electrostatic fields. *LWT, 176*, Article 114497. https://doi.org/10.1016/j.lwt.2023.114497

Berna, A. Z., Lammertyn, J., Saevels, S., Di Natale, C., & Nicolaï, B. M. (2004). Electronic nose systems to study shelf life and cultivar effect on tomato aroma profile. *Sensors and Actuators B: Chemical, 97*(2–3), 324–333. https://doi.org/10.1016/j.snb.2003.09.020

Buratti, S., Grassi, S., Ubaldi, P. G., Pianezzola, A., & Benedetti, S. (2025). E-nose-based control charts for fish freshness evaluation. *Food Research International, 209*, 116313. https://doi.org/10.1016/j.foodres.2025.116313

Feng, L., Zhang, M., Bhandari, B., & Guo, Z. (2018). A novel method using MOS electronic nose and ELM for predicting postharvest quality of cherry tomato fruit treated with high pressure argon. *Computers and Electronics in Agriculture, 154*, 411–419. https://doi.org/10.1016/j.compag.2018.09.032

Gómez, A. H., Wang, J., Hu, G., & Pereira, A. G. (2008). Monitoring storage shelf life of tomato using electronic nose technique. *Journal of Food Engineering, 85*(4), 625–631. https://doi.org/10.1016/j.jfoodeng.2007.06.039

Guan, S., Yang, F., Yao, J., Liu, C., Wang, R., Ruan, M., Yao, Z., Liu, C., Wan, H., & additional coauthors. (2025). Dynamic changes in volatile organic compounds of cherry tomato fruits during storage at different temperatures using HS-GC-IMS. *Food Research International, 218*, 116790.

Hong, X., Wang, J., & Qiu, S. (2015). Authenticating cherry tomato juices—Discussion of different data standardization and fusion approaches based on electronic nose and tongue. *Food Research International, 60*, 173–179.

Rabehi, A., Helal, H., Zappa, D., & Comini, E. (2024). Advancements and prospects of electronic nose in various applications: A comprehensive review. *Applied Sciences, 14*(11), 4506. https://doi.org/10.3390/app14114506

Sanislav, T., Mois, G. D., Zeadally, S., Folea, S., Radoni, T. C., & Al-Suhaimi, E. A. (2025). A comprehensive review on sensor-based electronic nose for food quality and safety. *Sensors, 25*(14), 4437. https://doi.org/10.3390/s25144437

Senge, J. M., Kaltenecker, F., & Krupitzer, C. (2025). Integrating sensor data, laboratory analysis, and computer vision in machine learning-driven E-nose systems for predicting tomato shelf life. *Chemosensors, 13*(7), 255. https://doi.org/10.3390/chemosensors13070255

Singh, P., Saini, R. K., Kumar, V., & Kaur, N. (2025). Understanding the concepts of smart E-nose technology in combination with machine learning for a new era of food safety: An advanced review. *Food Safety and Health, 3*(4), 518–534. https://doi.org/10.1002/fsh3.70029

Borowik, P., Dyshko, V., Tarakowski, R., Tkaczyk, M., Okorski, A., & Oszako, T. (2023). Analysis of the response signals of an electronic nose sensor for differentiation between *Fusarium* species. *Sensors, 23*(18), 7907.

Gertheiss, J., Rügamer, D., Liew, B. X. W., & Greven, S. (2023). *Functional Data Analysis: An Introduction and Recent Developments*. arXiv.

Haj Ammar, W., Boujnah, A., Baron, A., Boubaker, A., Kalboussi, A., Lmimouni, K., & Pecqueur, S. (2024). A temporal filter to extract doped conducting polymer information features from an electronic nose. *Electronics, 13*(3), 497.

Sanislav, T., Mois, G. D., Zeadally, S., Folea, S., Radoni, T. C., & Al-Suhaimi, E. A. (2025). A comprehensive review on sensor-based electronic nose for food quality and safety. *Sensors, 25*(14), 4437.

Vergara, A., Llobet, E., Ramírez, J. L., Ivanov, P., Fonseca, L., Zampolli, S., Scorzoni, A., Becker, T., Marco, S., & Wöllenstein, J. (2007). Feature extraction of metal oxide gas sensors using dynamic moments. *Sensors and Actuators B: Chemical, 122*(1), 219–226.

Zhai, Z., Long, H., Feng, Y., Li, Y., & Wang, Y. (2024). Electronic noses: From gas-sensitive components and practical applications to data processing. *Sensors, 24*(15), 4806.

Chu, J., Yao, R., Huang, X., Yang, A., Pan, J., Yuan, H., Rong, M., & Wang, X. (2024). Contrastive domain generalization convolution neural network correcting the drift of gas sensors. *Sensors and Actuators A: Physical, 372*, Article 115314. https://doi.org/10.1016/j.sna.2024.115314

Feng, L., Zhang, M., Bhandari, B., & Guo, Z. (2018). A novel method using MOS electronic nose and ELM for predicting postharvest quality of cherry tomato fruit treated with high pressure argon. *Computers and Electronics in Agriculture, 154*, 411–419. https://doi.org/10.1016/j.compag.2018.09.032

Leinonen, T., Wong, D., Vasankari, A., Wahab, A., Nadarajah, R., Kaisti, M., & Airola, A. (2024). Empirical investigation of multi-source cross-validation in clinical ECG classification. *Computers in Biology and Medicine, 183*, Article 109271. https://doi.org/10.1016/j.compbiomed.2024.109271

Rosenblatt, M., Hlinka, J., Li, X., Scheinost, D., & Noble, S. (2024). Data leakage inflates prediction performance in connectome-based machine learning models. *Nature Communications, 15*, Article 1829. https://doi.org/10.1038/s41467-024-46150-w

Sanislav, T., Mois, G. D., Zeadally, S., Folea, S., Radoni, T. C., & Al-Suhaimi, E. A. (2025). A comprehensive review on sensor-based electronic nose for food quality and safety. *Sensors, 25*(14), Article 4437. https://doi.org/10.3390/s25144437

Singh, P., Saini, R. K., Kumar, V., & Kaur, N. (2025). Understanding the concepts of smart E-nose technology in combination with machine learning for a new era of food safety: An advanced review. *Food Safety and Health, 3*(4), 518–534. https://doi.org/10.1002/fsh3.70029

Wörner, J., Eimler, J., & Pein-Hackelbusch, M. (2025). Long-term drift behavior in metal oxide gas sensor arrays: A one-year dataset from an electronic nose. *Scientific Data, 12*, Article 1628. https://doi.org/10.1038/s41597-025-05993-8

Berna, A. Z., Lammertyn, J., Saevels, S., Di Natale, C., & Nicolaï, B. M. (2004). Electronic nose systems to study shelf life and cultivar effect on tomato aroma profile. *Sensors and Actuators B: Chemical, 97*(2–3), 324–333. https://doi.org/10.1016/j.snb.2003.09.020

Feng, L., Zhang, M., Bhandari, B., & Guo, Z. (2018). A novel method using MOS electronic nose and ELM for predicting postharvest quality of cherry tomato fruit treated with high pressure argon. *Computers and Electronics in Agriculture, 154*, 411–419. https://doi.org/10.1016/j.compag.2018.09.032

Gómez, A. H., Wang, J., Hu, G., & Pereira, A. G. (2008). Monitoring storage shelf life of tomato using electronic nose technique. *Journal of Food Engineering, 85*(4), 625–631. https://doi.org/10.1016/j.jfoodeng.2007.06.039

Han, Z., Zhang, L., Zhang, B., Zou, F., & Shang, N. (2024). Progress on research and application of new non-destructive testing techniques in tomato quality inspection. *Food Science, 45*(1), 289–300. https://doi.org/10.7506/spkx1002-6630-20230103-014

Huang, X., Pan, S., Zhao, S., & Aheto, J. H. (2018). Evaluating quality of tomato during storage using fusion information of computer vision and electronic nose. *Journal of Food Process Engineering, 41*(6), Article e12832. https://doi.org/10.1111/jfpe.12832

Liu, Y., Wei, C., Yoon, S.-C., Ni, X., Wang, W., Liu, Y., Wang, D., Wang, X., & Guo, X. (2024). Development of multimodal fusion technology for tomato maturity assessment. *Sensors, 24*(8), Article 2467. https://doi.org/10.3390/s24082467

Zhai, Z., Long, H., Feng, Y., Li, Y., & Wang, Y. (2024). Electronic noses: From gas-sensitive components and practical applications to data processing. *Sensors, 24*(15), Article 4806. https://doi.org/10.3390/s24154806

Al-Gaadi, K. A., Zeyada, A. M., Tola, E., Alhamdan, A. M., Ahmed, K. A. M., Madugundu, R., & Edrris, M. K. (2024). Impact of storage conditions on fruit color, firmness and total soluble solids of hydroponic tomatoes grown at different salinity levels. *Applied Sciences, 14*(14), 6315. https://doi.org/10.3390/app14146315

Ding, R., Dai, X., Zhang, Z., Bi, Y., & Prusky, D. (2024). Composite coating of oleaster gum containing cuminal keeps postharvest quality of cherry tomatoes by reducing respiration and potentiating antioxidant system. *Foods, 13*(10), 1542. https://doi.org/10.3390/foods13101542

Li, X., Huang, H., Zhang, L., & Zhao, L. (2025). Effect of postharvest storage temperature and duration on tomato fruit quality. *Foods, 14*(6), 1002. https://doi.org/10.3390/foods14061002

Zhang, Y., Wu, J., Li, J., Mi, S., Yu, W., Sang, Y., & Wang, X. (2025). Enhancing post-harvest quality of cherry tomatoes via low voltage electrostatic field and modified atmosphere packaging: Regulation of phenylpropanoid metabolism. *LWT, 237*, 118782. https://doi.org/10.1016/j.lwt.2025.118782

Zhao, Y., Li, L., Gao, S., Wang, S., Li, X., & Xiong, X. (2023). Postharvest storage properties and quality kinetic models of cherry tomatoes treated by high-voltage electrostatic fields. *LWT, 176*, 114497. https://doi.org/10.1016/j.lwt.2023.114497