Specifications

Item		Description
Analysis functions	Univariate analysis	t-test Mann-Whitney U-test ANOVA (analysis of variance)
	Multivariate analysis	PCA (Principal Component Analysis) PLS-DA
	Discriminant analysis	Support Vector Machine (SVM) Random Forest
	Other	Dynamic grouping
Display functions	Multivariate analysis	Peak Matrix Box Plot ROC AUC Score/Loading Plot Dendrogram
	Discriminant analysis	Discriminant analysis results (group and score) Unknown samples superimposed on Score Plot
Input-output data	Input	Chromatogram summary file ASCII format peak lists JCAMP format peak lists mzML format peak lists (only supports Centroid mode, 1 scan/per file, data uncompressed) Note: File name must only contain half-width alphanumeric characters.
	Output	Peak lists (txt format) Analysis results (xlsm format) Graph screenshot
Operating system	OS	Windows® 11 Professional 64bit (English/Chinese) Windows® 10 Professional 64bit (English/Chinese)

LC-MS

Product

MALDI-TOF MS

Product

GC-MS

Product

DPiMS™

Product

eMSTAT Solution, DPiMS, Smart Metabolites Database, Smart Aroma Database, and GCMS-TQ are trademarks of Shimadzu Corporation or its affiliated companies in Japan and/or other countries.
Windows is either a registered trademark or a trademark of Microsoft Corporation in the United States and/or other countries.

Shimadzu Corporation
www.shimadzu.com/an/

For Research Use Only. Not for use in diagnostic procedures.
This publication may contain references to products that are not available in your country. Please contact us to check the availability of these products in your country.
Company names, products/service names and logos used in this publication are trademarks and trade names of Shimadzu Corporation, its subsidiaries or its affiliates, whether or not they are used with trademark symbol "TM" or "®".
Third-party trademarks and trade names may be used in this publication to refer to either the entities or their products/services, whether or not they are used with trademark symbol "TM" or "®".
Shimadzu disclaims any proprietary interest in trademarks and trade names other than its own.

The contents of this publication are provided to you "as is" without warranty of any kind, and are subject to change without notice. Shimadzu does not assume any responsibility or liability for any damage, whether direct or indirect, relating to the use of this publication.

© Shimadzu Corporation, 2024 / First Edition: August 2018, Printed in Japan 3655-02404-PDFIK, C146-E377A

C146-E377A

SHIMADZU
Excellence in Science

Statistical Analysis Software for Analytical Instruments

eMSTAT Solution

Enabling statistical analysis of chromatography and direct ionization mass spectrometry data regardless of user experience

eMSTAT Solution™

Sample group differentiation—such as when comparing against benchmark products in the product development phase or in sample pass/fail assessment during quality control—can be difficult when relying on conventional qualitative and quantitative analytical techniques. The process typically involves multivariate analysis using large amounts of chromatogram or MS spectrum data to search for markers that can help differentiate between sample groups, and to visualize sample characteristics using these markers.
It also requires modeling using training data and discriminant analysis to scientifically determine the characteristics of unknown samples, including their condition (e.g., deterioration), type, and place of origin.
The eMSTAT Solution statistical analysis software is equipped with a range of features that make multivariate analysis more accessible and user-friendly for everyone—even when analyzing large volumes of various analytical data (chromatograms and MS spectra) .

Features

Statistical Analysis Mode for Intuitive Data Analysis Operations

The software's Statistical Analysis Mode contains extensive, easy-to-use functions for univariate and multivariate analysis that enable even users without statistical analysis expertise to identify marker peaks for sample differentiation and discrimination.

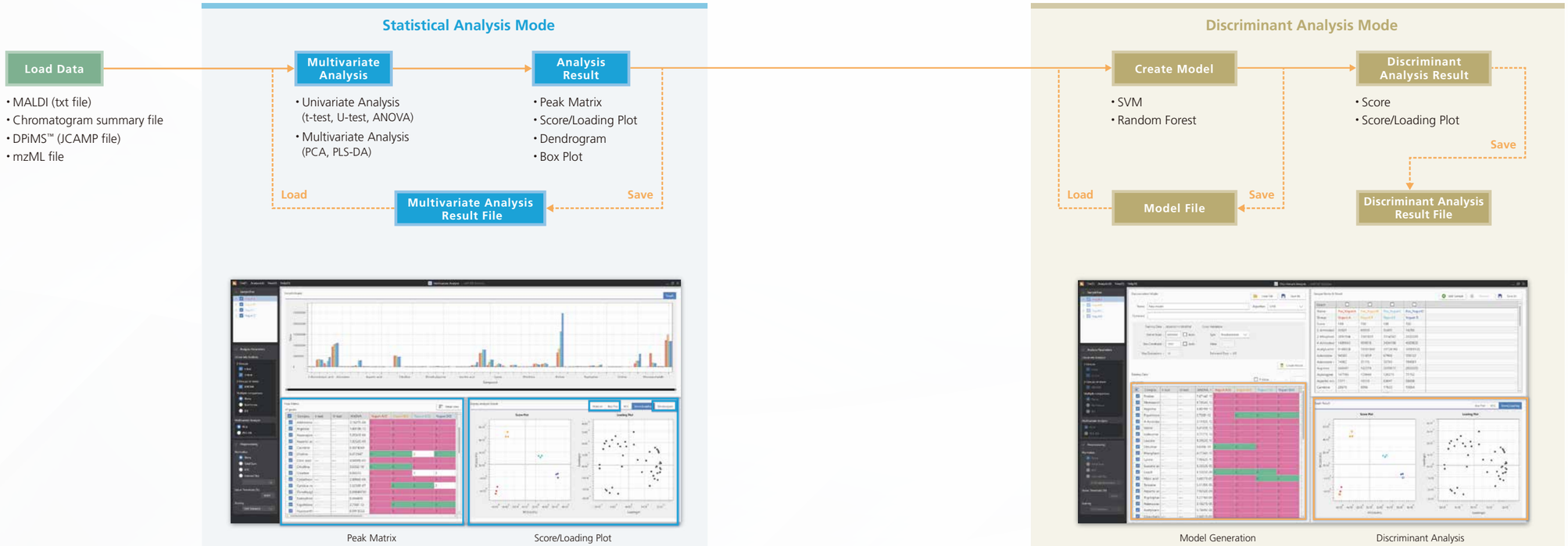
Dynamic Grouping Function for Flexible Sample Grouping

Leveraging the software's quality information library, the dynamic grouping function facilitates marker discovery by enabling flexible sample grouping in various combinations.

Discriminant Analysis Mode Supports Grouping of Unknown Samples

Discriminant models for discrimination of unknown samples can be created using marker peaks identified in Statistical Analysis Mode. With single-step mode switching, statistical data obtained in Statistical Analysis Mode can be used in Discriminant Analysis Mode.

Workflow Using eMSTAT Solution



Peak Matrix　　Score/Loading Plot

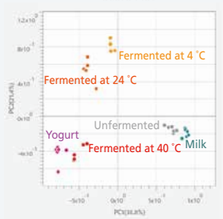
Model Generation　　Discriminant Analysis

Statistical Analysis Mode for Intuitive Data Analysis Operations

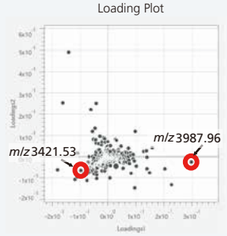
Proper statistical analysis typically requires a certain degree of user expertise and poses a steep learning curve. The eMSTAT Solution statistical analysis software eliminates these challenges with a suite of intuitive operating features. The software enables a wide range of analyses from sample differentiation and identification of useful marker peaks in Statistical Analysis Mode to discriminant modeling of unknown samples in Discriminant Analysis Mode.

Statistical Analysis Mode

Score Plot

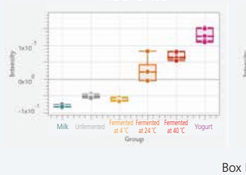


Loading Plot

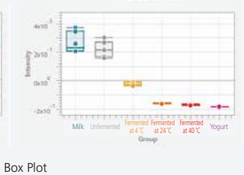


Comparison of yogurts with different fermentation conditions (PCA)

m/z 3421.53



m/z 3987.96



Box Plot

When searching for characteristic sample components, sample differences can be visualized by using principal component analysis (PCA) to comprehensively handle all measurement data and identify inherent differences.

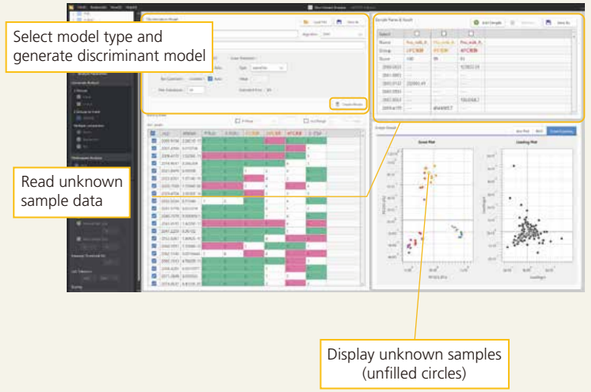
After selecting each sample group's characteristic peaks from the score plot and loading plot, differences between groups can be confirmed via the box plot.

In the score plot (top left), different yogurt samples are clearly classified according to the degree of fermentation. The loading plot (top right) pairs the positional relationship of each sample against those on the score plot and shows the peaks (i.e., compound names for chromatogram data and *m/z* for MS spectrum data) that aid in classifying each sample group.

The components characterizing each group can then be quantitatively compared using the box plot.

The peak *m/z* 3421.53 located in the lower left of the loading plot is detected with particular intensity for the yogurt and 40 °C-fermented samples. Meanwhile, the peak *m/z* 3987.96 located in the lower right of the loading plot has a strong intensity in the unfermented and milk samples that diminishes as the level of fermentation increases. These results suggest that *m/z* 3421.53 can be used as a fermentation marker while *m/z* 3987.96 can be used as a non-fermentation marker. Analytical efficiency can be improved by registering such markers in the Annotation List.

Discriminant Analysis Mode



Discriminant analysis of unknown samples can be performed by generating models using known sample data (training data).

After generating a discriminant model and reading the unknown sample data, the software then displays the sample groups by data attribute.

Users can select Support Vector Machine or Random Forest for the discriminant model algorithm.

Dynamic Grouping Function for Flexible Sample Grouping

In sample group differentiation, analysis samples can be grouped according to attribute (raw materials, reaction conditions, normal/abnormal, manufacturer, type, etc.) and then searched for characteristic group markers. The eMSTAT Solution software's Dynamic Grouping Function effectively enables the creation of new sample groups while retaining existing group information.

Analytical Conditions

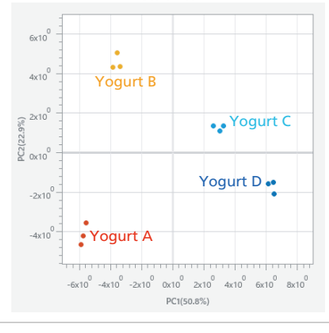
Instrument: LCMS-8045/8050/8060NX
Sample: Commercially available yogurt

Using the ion-pair-free LC/MS/MS method included in Shimadzu's LC/MS/MS Method Package for Primary Metabolites Ver. 3, the system can comprehensively analyze a total of 153 compounds, including sugars, in addition to hydrophilic metabolites essential for metabolomic analysis, such as amino acids, organic acids, nucleosides, and nucleotides.

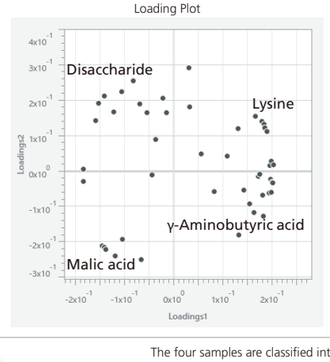
Group samples by product

Grouping of yogurt samples by product and identification of characteristic marker compounds

Score Plot



Loading Plot



Dynamic grouping

The four samples are classified into two groups.
Yogurt A and Yogurt B in Company X,
Yogurt C and Yogurt D in Company Y

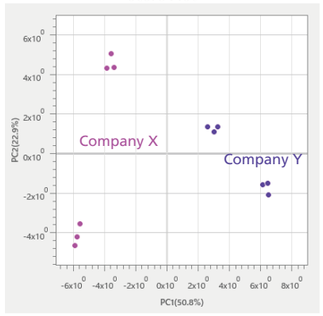
Rearranged Groups

Group Name	A+B	C+D
Company X	+	-
Company Y	-	+

Search for characteristic group markers

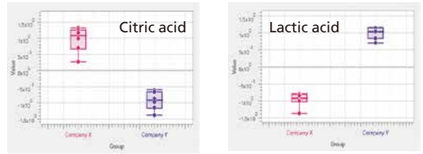
Once the yogurt samples have been grouped, the compounds differentiating each group can be identified using the Peak Matrix *p*-value (t-test and U-test for 2 groups and ANOVA for 3 or more groups). The box plot shows that the levels of Citric acid and Lactic acid compounds with a *p*-value of < 0.05 are higher in the Company X and Company Y samples, respectively.

Score Plot



Peak Matrix

Compound	Annotat	t-test	U-test	ANOVA	Company X	Company Y
Lactic acid		1.5638E-08	0.0021645	---	6	6
Arginine		1.716E-08	0.0021645	---	6	6
Citrulline		6.4082E-08	0.0021645	---	0	6
Tyrosine		1.6226E-07	0.0021645	---	6	6
Aspartic acid		5.8596E-07	0.0021645	---	6	6
Citric acid		5.4604E-06	0.0021645	---	6	6
Dimethylglycine		2.3815E-05	0.0021645	---	6	6

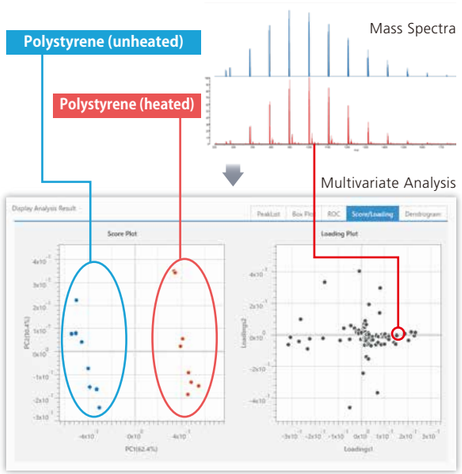


Box Plot

eMSTAT Solution
Statistical Analysis Software for Analytical Instruments

Discriminant Analysis Mode Supports Grouping of Unknown Samples

Identifying Degraded Synthetic Polymers



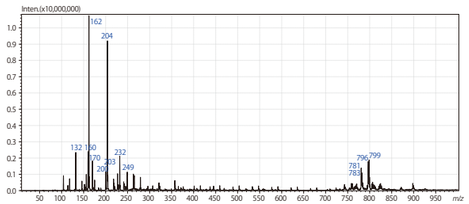
Polystyrene (heated and unheated) polymers were separated into two groups (heated or unheated) by multivariate analysis (PLS-DA) of the MALDI mass spectrum (Score Plot). A Loading Plot can be used to confirm which peaks (marker peaks) affect the differences between the two groups.

Experimental Conditions
- Instrument: MALDI-8020
- Sample: Polystyrene 1,000 (heated for 2 hours at 200 ℃ or unheated)
- Matrix: Dithranol
- Cationizing Agent: AgTFA
- Sample Preparation: The polystyrene (heated and unheated) dissolved in THF, matrix solution, and cationizing agent were spotted onto a sample plate.

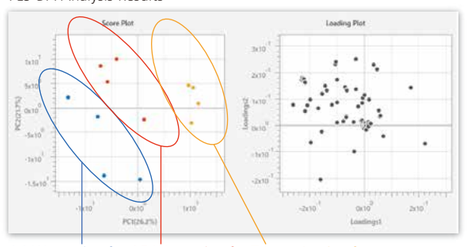
Using marker peaks identified by multivariate analysis to create a discriminant model and discriminate between heated and unheated polymers in a polystyrene mass spectrum, obtained separately, resulted in correct discrimination of all polymers.
By using eMSTAT Solution in combination with MALDI mass spectrometry, which can easily measure samples with large molecular weight synthetic polymers and a wide variety of other samples, such as protein, fat, or sugar samples, can be easily differentiated.

Easy Differentiation for Beef Classification

Direct Probe Ionization-MS Mass Spectrum



PLS-DA Analysis Results



Extracts from commercial beef (Tasmanian and A5/A4 grade Wagyu beef) were analyzed in a DPiMS system mass spectrometer. The resulting spectra were then analyzed by PLS discriminant analysis. A Score Plot confirms grouping into three groups and a Loading Plot confirms which metabolite peaks affect grouping.

Experimental Conditions
- Instrument: DPiMS system
- Sample: Edible beef
- Ionization Solution: 50% aqueous ethanol solution
- Sample Preparation: 10 µL of the suspension from homogenizing the beef for human consumption in the 50% aqueous ethanol solution was dripped onto a sample plate.

With eMSTAT Solution, spectra obtained by convenient metabolite analysis in a DPiMS system spectrometer can be used to easily differentiate between differences in food, plant, and other samples, and screen for information about metabolites that contribute to differentiation.

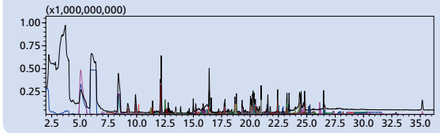
Discriminant Analysis of Beef Meats

In addition to chromatogram and MS spectrum data, the eMSTAT Solution software can read various data formats (JCAMP, ASCII, mzML, etc.), supporting the integrated analysis and multi-faceted assessment of various types of data.

Analytical Conditions
Instrument: GCMS-TQ8040 NX
Sample: Beef meat
Using the Smart Metabolites Database™ Ver. 2, this system is capable of analyzing about 500 primary metabolites, including organic acids, sugars, nucleic acids, fatty acids, and amino acids. Using the Smart Aroma Database™, it can also analyze about 500 aroma compounds with SPME Arrow.

A discriminant analysis was successfully performed on an unknown meat to determine its farm of origin based on approximately 600 variables (i.e., GC/MS, UV, meat lipid measuring device, microwave-drying moisture meter, and visual assessment).
In addition to determining the origin, the discriminant model revealed the characteristic variables of Farmer 2 (e.g., the roast-smelling aroma compound benzothiazole, the primary metabolite linoleic acid).

GC-MS
488 primary metabolite compounds and 476 aroma compounds were measured



Meat lipid measuring device
Oleic acid and monounsaturated fatty acids were measured.

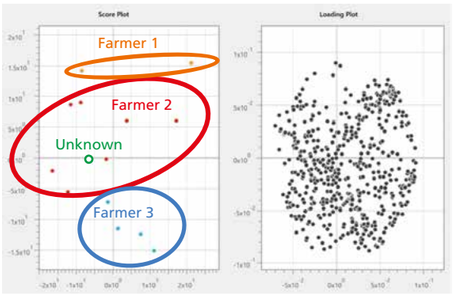
Visual assessment
Assessment of grading score, marbling, BMS, BFS, BCS, etc.

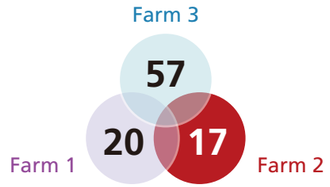
UV-Visible spectrophotometer
Determination of glycogen concentration

Microwave-drying moisture meter
Analysis of crude fat content, crude protein content, and water content

Approximately 600 variables

Random forest discriminant model

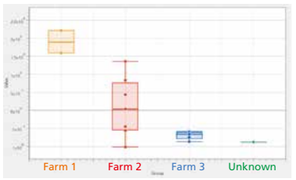




The number of characteristic variables in each farm (i.e., ANOVA p-value of <5 %)

Select	☐
Name	Unknown (Farm 2)
Group	Farm 2
Score	70

Aroma compound
Benzothiazole (*p*-value < 0.0018)



Primary metabolite
Linoleic acid (*p*-value < 0.0000043)

eMSTAT Solution
Statistical Analysis Software for Analytical Instruments