CLAIMS

1. A method for processing lipid contents of at least one oil sample and simulating at least one training sample, the method comprising:

receiving a first dataset of values associated with lipid contents of the at least one oil sample, wherein the lipid contents of the at least one oil sample comprises a first lipid content and a second lipid content;

applying a multivariate mixture model to the first dataset of values to simulate an intermediate dataset, wherein the multivariate mixture model is capable of accounting for at least a multivariate variance between the value associated with the first lipid content and the value associated with the second lipid content; and

simulating the at least one training sample based on at least part of the simulated intermediate dataset,

wherein the at least one simulated training sample is represented by a second dataset of values associated with lipid contents of the at least one simulated training sample, and

wherein the step of simulating the at least one training sample comprises:

selecting from the at least part of the simulated intermediate dataset and drawing a mixing proportion from a probability distribution to obtain the at least one simulated training sample, the probability distribution comprising a Dirichlet distribution.

2. The method of claim 1, wherein the multivariate mixture model comprises a Gaussian mixture model or a multivariate Gaussian mixture model.

3. The method of claim 1 or 2, wherein the multivariate mixture model is constructed and constrained by a geometric characteristic.

4. The method of claim 3, wherein the geometric characteristic comprises a diagonal distribution with equal volume and equal shape.

5. The method of any one of claims 1 to 4, wherein the step of applying the multivariate mixture model to the first dataset of values comprises:

learning groups of lipid parameters from the first dataset of values; and

simulating, based on the groups of lipid parameters, the intermediate dataset.

6. The method of claim 5, wherein each group of lipid parameters comprises a vector of average values representing likely lipid percentages and a covariance matrix representing a spread within and across the likely lipid percentages.

7. The method of any one of claims 1 to 6, wherein the at least one oil sample comprises a pure oil sample.

8. The method of any one of claims 5 to 7, wherein the step of simulating the at least one training sample comprises performing a Monte Carlo simulation.

9. The method of any one of claims 1 to 8, wherein the number of dimensions of the Dirichlet distribution is more than or equal to 2.

10. The method of any one of claims 1 to 9, wherein the number of dimensions of the Dirichlet distribution is 9 or more.

11. The method of any one of claims 1 to 10, wherein the lipid contents of the oil sample comprise fatty acid contents of the oil sample obtainable using a gas chromatography – flame ionization detector or photonic detection.

12. A method for predicting a blending formula of a blended oil sample, the method comprising:

receiving a dataset of values associated with lipid contents of the blended oil sample;

providing a single prediction model capable of generating a prediction of at least two constituent oils in the blended oil sample; and

processing the at least part of the dataset using the single prediction model,

wherein the single prediction model is trained by a training dataset, wherein the training dataset comprises a second dataset of values associated with lipid contents of the at least one training sample simulated by a method of any one of claims 1 to 11.

13. The method of claim 12, wherein the single prediction model comprises a deep neural network.

14. The method of claim 12 or 13, further comprising generating the prediction of at least two constituent oils in the blended oil sample.

15. The method of claim 14, wherein the prediction comprises at least a type or a percentage amount of each of the at least two constituent oils.

16. A method for identifying a blind oil sample, the method comprising:

receiving a dataset of values associated with lipid contents of the blind oil sample; and

applying a multivariate mixture model comprising groups of lipid parameters to the dataset of values to identify the blind oil sample,

wherein the groups of lipid parameters are learned by the multivariate mixture model based on lipid contents of at least one preceding oil sample, wherein the lipid contents of the at least one preceding oil sample comprise a first lipid content and a second lipid content, wherein the multivariate mixture model is capable of accounting for at least a multivariate variance between a value associated with the first lipid content and a value associated with the second lipid content, and wherein the lipid contents of the at least one preceding oil sample comprise lipid contents of at least one oil sample processed by a method of any one of claims 1 to 11.

17. A method for processing lipid contents of at least one oil sample to simulate at least one blend sample and determining compliance of the at least one blend sample with a prescribed requirement, the method comprising:

receiving a first dataset of values associated with lipid contents of the at least one oil sample;

simulating the at least one blend sample based on at least part of the first dataset, wherein the at least one simulated blend sample is represented by a second dataset of values associated with lipid contents of the at least one simulated blend sample; and

determining whether each value of the second dataset for each simulated blend sample falls within a pre-determined range of a corresponding lipid content in accordance with the prescribed requirement such that if it is determined affirmative, the simulated blend sample complies with the prescribed requirement, and if it is determined negative, the simulated blend sample fails compliance with the prescribed requirement,

wherein the step of simulating the at least one blend sample comprises selecting from the at least part of the first dataset and drawing a mixing proportion from a probability distribution to obtain the at least one simulated blend sample, the probability distribution comprising a Dirichlet distribution.

18. The method of claim 17, wherein the step of simulating the at least one blend sample comprises selecting from the at least part of the first dataset and drawing a plurality of mixing proportions from the probability distribution to obtain a plurality of simulated blend samples; and wherein each mixing proportion differs from another mixing proportion within a range between a maxima and a minima of the probability distribution.

19. The method of claim 17 or 18, wherein the prescribed requirement comprises a national standard for oil blends implemented in China.

20. A computer readable storage medium comprising computer readable instructions operable when executed by a computer to carry out at least one of the followings:

process lipid contents of at least one oil sample and simulate at least one training sample, the computer readable instructions configured to perform a method of any one of claims 1 to 11;

predict a blending formula of a blended oil sample, the computer readable instructions configured to perform a method of any one of claims 12 to 15;

identify a blind oil sample, the computer readable instructions configured to perform a method of claim 16; and

process lipid contents of at least one oil sample to simulate at least one blend sample and determine compliance of the at least one blend sample with a prescribed requirement, the computer readable instructions configured to perform a method of any one of claims 17 to 19.

21. An apparatus or system comprising:

a receiving unit configured to receive a dataset of values associated with lipid contents of at least one oil sample, or a blended oil sample, or a blind oil sample;

a memory for at least storing a multivariate mixture model capable of accounting for at least a multivariate variance between the lipid contents of the at least one oil sample, a single prediction model capable of generating a prediction of a blending formula of the blended oil sample; a probability distribution model capable of simulating a blend sample based on at least part of the dataset; and

a processor configured to access

the multivariate mixture model stored in the memory to carry out at least one of the followings:

perform steps of a method of any one of claims 1 to 11 for processing the lipid contents of the at least one oil sample and simulating at least one training sample,

perform steps of a method of claim 16 for identifying the blind oil sample;

the single prediction model stored in the memory to perform steps of a method of any one of claims 12 to 15 for generating a prediction of the blending formula of the blended oil sample; and

the probability distribution model stored in the memory to perform steps of a method of any one of claims 17 to 19 for processing lipid contents of the at least

one oil sample to simulate at least one blend sample and determining compliance of the at least one blend sample with a prescribed requirement.