MATERIALS AND METHODS

This study was conducted in five parts:

1. Development of quantitative models for predicting the chemical and physical properties of Thai commercial fish sauces.
2. Development of qualitative models for classifying Thai fish sauce based on their total nitrogen content
3. Categorization of Thai fish sauces based on sensory characteristics and principal component analysis
4. Correlation between sensory, chemical and physical properties of Thai fish sauces studied by NIR spectroscopy combined with chemometrics
5. Development of quantitative models for predicting the sensory properties of Thai fish sauces.

Materials

Two types of Thai commercial fish sauces, pure and mixed fish sauces, were purchased from the markets around Kasetsart University, Bangkok, Thailand. In this study, both types of fish sauces were purchased two times for two sample sets that belong to the different study. The descriptions of each set are as following:

(1) Set 1 consisted of 100 fish sauce samples (65 pure fish sauce samples and 35 mixed fish sauce samples). All samples were produced during the same period of January 2004. This set was used for the study of part 1 and 2.

(2) Set 2 consisted of 20 fish sauce samples (12 pure fish sauce samples and 8 mixed fish sauce samples). All samples were produced during the same period of January 2006. This set was used for the study of part 3 to 5.

All samples were kept in tightly sealed plastic bottles and stored at 4 °C for 3 weeks before measuring their properties.

Methods

1. Development of Quantitative Models for Predicting the Chemical and Physical Properties of Thai Commercial Fish Sauces

A total of 100 Thai fish sauce samples were used. The chemical and physical analyses as well as NIR analysis of samples were investigated. To develop the quantitative models, the chemical and physical data were used as dependent variables (Y-data). While the NIR spectral data were used as independent variables (X-data). Partial least squares (PLS) regression with the aid of wavelength interval selection methods were used to develop the models.

1.1 Chemical and Physical Analyses

1.1.1 Chemical Analysis

The total nitrogen, sodium chloride, and pH were determined by standard methods (AOAC, 2000). The pH of the fish sauce was determined by direct measurement of the test sample with the pH meter (Model HM-165, TOA Electronic, Tokyo, Japan). Reducing sugar was determined by Nelson-Somogyi reducing sugar method (Somogyi, 1952). The reference methods for the chemical analyses are described in Appendix A.

1.1.2 Physical Analysis

Density and baume values of the fish sauces were determined directly at 25 °C using a density meter (DMA 35n, Anton Paar, Austria). Total soluble solids and refractive index of the fish sauces were measured directly at 25 °C with an Abbe refractometer (Model 2T, ATAGO, Japan). The color of fish sauce was determined by measuring the L* (black-white component, luminosity), a* (+red to-green component), and b* (+yellow to –blue component) values using a spectrophotometer (CM-3500d, Minolta, Japan), according to the CIE Lab scale. Samples were pipetted into a path

length of 5 mm quartz cuvette cell. The sample was illuminated with D65-artificial daylight (10° standard angle).

All chemical and physical analyses were performed in triplicate, and then mean values of the triplicate determination were calculated. These data were imported into the NIR spectral data files and used in calibration and validation studies.

1.2 NIR Spectra Acquisition

NIR transflectance spectra were obtained from 1100 to 2500 nm at a 2 nm interval using an InfraAlyzer 500 spectrometer (Bran+Luebbe, Norderstedt, Germany) and a 0.3 mm British cup. The fish sauce samples were incubated at 29°C in a water bath prior to the NIR measurements. The spectra data were transferred into JCAMP.DX format and imported into the Unscrambler® software (version 7.8: CAMO AS, Trondheim, Norway) and into files for the MATLAB® software (version 5.3: The MathWorks, USA) for the data analysis. The spectra were not subjected to any preprocessing.

1.3 Model Analysis

Both data from section 1.1 (Y-data) and section 1.2 (X-data) were used to develop the quantitative models. The models were constructed by using PLS calibration models with the aid of wavelength interval selection methods. The purpose of wavelength interval selection is to find out the suitable input wavelength variables for developing the calibration models. Two wavelength interval selection methods named i) moving window partial least squares regression (MWPLSR) and ii) searching combination moving window partial least squares (SCMWPLS) were applied. The block diagram of quantitative modeling process for chemical and physical properties is shown in Figure 17.

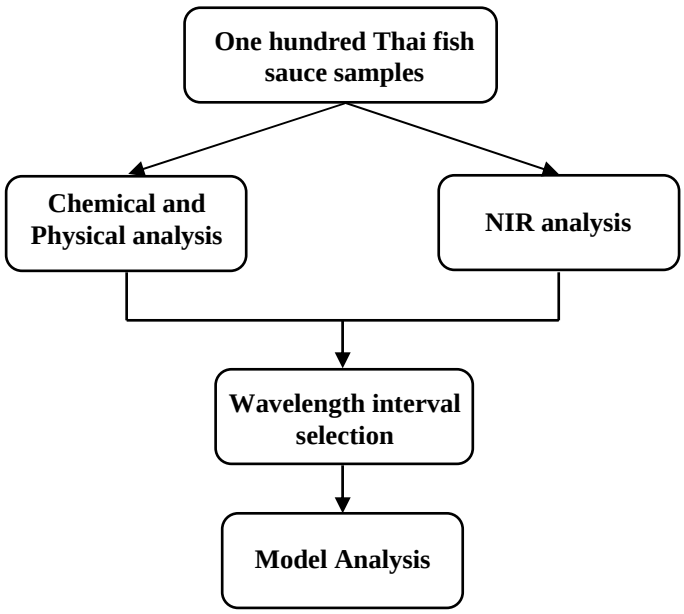

Figure 17 Block diagram of quantitative modeling process for chemical and physical properties.

The spectra data of 100 samples were split randomly into two sets for modeling the concentrations of chemical and physical parameters. The calibration set consisted of 70 samples while the prediction set contained 30 samples.

PLS regression was utilized to develop calibration models by using i) the whole spectra region, ii) the informative regions, iii) the direct combination of informative regions and iv) the optimized combination of informative regions. The informative regions and the optimized combination of informative regions were obtained by MWPLSR and SCMWPLS, respectively. MWPLSR with the window size of 20 spectra points and SCMWPLS were carried out by use of in-house-written program in MATLAB® software ver. 5.3 (The MathWorks,USA).

PLS models were performed by the Unscrambler® software (version 7.8: CAMO AS, Trondheim, Norway). The random cross-validation was used to develop and evaluate these regression models. The lowest root mean square error prediction in cross validation (RMSECV) was used to choose the optimum number of PLS factors

in the model. The maximum number of PLS factors were set at 10 which usually sufficient to adequately describe a model (Clark *et al*., 2004). The RMSECV was calculated as the equation:

$$\text{RMSECV} = \sqrt{\frac{\sum_{i=1}^{n} (C_{\text{NIRi}} - C_{\text{REFi}})^2}{n}}$$

where n is the number of samples included in the calibration matrix or the calibration set, C_{REF} the concentration in the sample as measures by the reference methods and C_{NIR} the concentration as predicted by PLS from the NIR spectrum.

Once optimized on the calibration set, the prediction model was applied to spectra in the prediction set. The predictive accuracy of PLS models was compared in terms of the root mean square error of prediction (RMSEP), which was calculated as the equation:

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n} (C_{\text{NIRi}} - C_{\text{REFi}})^2}{n}}$$

where n is the number of samples included in the prediction matrix or the prediction set, C_{REF} the concentration in the sample as measures by the reference methods and C_{NIR} the concentration as predicted by PLS from the NIR spectrum.

Finally the model with the lowest RMSEP will be selected as final model. Correlation coefficients (R) between the predicted and the measured value are calculated for the prediction set.

2. Development of Qualitative Models for Classifying Thai Fish Sauces Based on their Total Nitrogen Content

A total of 100 Thai fish sauce samples were used to develop the qualitative models for classifying Thai fish sauces into three groups based on their total nitrogen content. Five different supervised pattern recognition techniques were applied to develop the classification models. They were i) Linear discriminant analysis (LDA), ii) Factor analysis-Linear discriminant analysis (FALDA), iii) Soft Independent Modeling of Class Analog (SIMCA), iv) K Nearest Neighbors (KNN) and v) Artificial Neural Networks (ANNs). The selected wavelengths for total nitrogen content obtained by SCMWPLS method were used as input variables (X-data). The output variables (Y-data) were the category memberships of samples for standard pure fish sauce (SPF), standard mixed fish sauce (SMF), and out of standard fish sauce (OF). The block diagram of classification modeling process as showed in Figure 18.

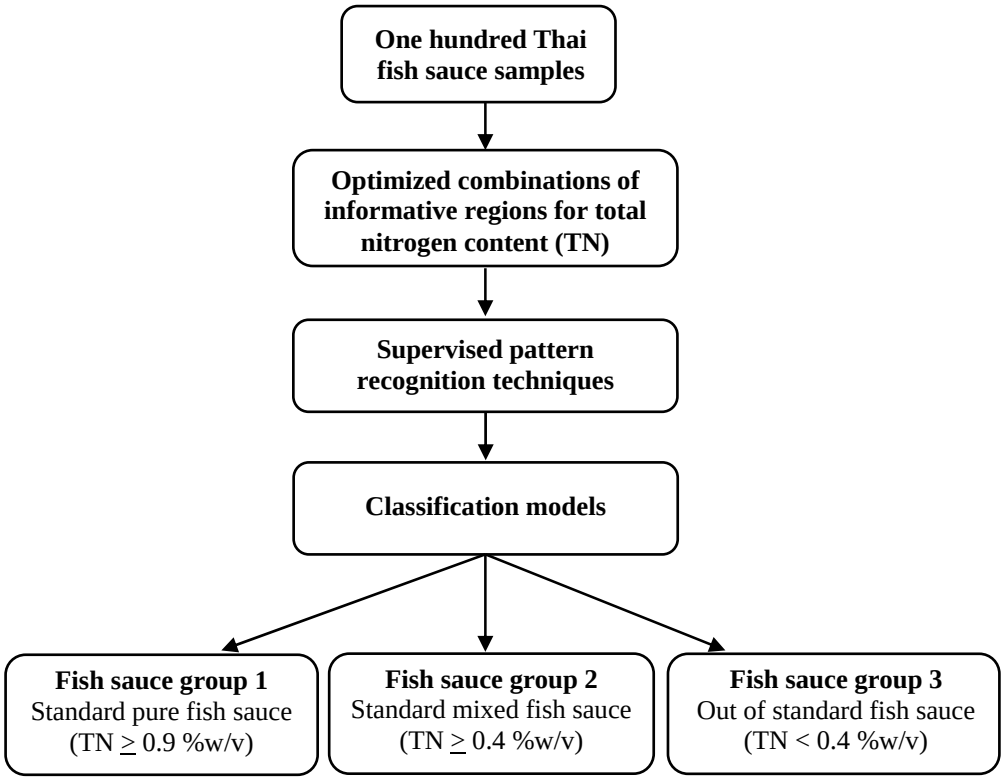

Figure 18 Block diagram of classification modeling process.

The memberships of samples for standard pure fish sauce (SPF), standard mixed fish sauce (SMF), and out of standard fish sauce (OF) were categorized based on the total nitrogen content as shown in Figure 18. For classification analysis, the samples were ascribed dummy variable values of 1, 2, and 3 depending on SPF-, SMF-, OF-groups, respectively. To develop the qualitative models, the samples were split into training set and test set. The memberships of three groups with different content of total nitrogen were randomly separated into both sets. The supervised pattern recognition techniques used in this study are the followings:

2.1 Linear Discriminant Analysis (LDA)

LDA is a parametric method based on maximizing the variance between categories and minimizing the variance within categories. This method renders a number of orthogonal linear discriminant functions equal to the number of categories minus 1 (Groot *et al*., 1999; Moreda-Pineiro *et al*, 2003). In this study, LDA was carried out on the NIR spectral data using SPSS® software. Full cross-validation was used to validate the developed models. The F-values for entering and removing stepwise method were determined to be 3.84 and 2.71, respectively.

2.2 Factor Analysis-Linear Discriminant Analysis (FALDA)

The purpose of FALDA technique is to combine the data reduction method with LDA in order to reduce the number of input variables. The data reduction method used in this study is Factor analysis (FA). In FALDA, the factors were extracted from the correlation matrix composed of informative spectra by using FA. The extraction and rotation methods used in this study were principal component analysis (PCA) and Varimax with Kaiser Normalization methods, respectively. Subsequently, the extracted factor scores were used as input variables (X-data) for developing the LDA model. The F-values for entering and removing stepwise method were determined to be 3.84 and 2.71, respectively. Both FA and LDA were performed by SPSS® software.

2.3 Soft Independent Modeling of Class Analog (SIMCA)

SIMCA is a parametric method based on making a principal component analysis (PCA) model for each class in the training set (Iizuka and Aishima, 1999; Blanco and Pages, 2002). In this study, PCA was applied to NIR spectra in order to identify the principal components (PCs). The SIMCA models were constructed for each class in the training set with an optimal number of PCs. New data can be classified into the appropriate group on the basis of the distance of the new data to each of the class models. The classification results can be examined by Cooman's plot, which shows the distance between samples and the center of each group. SIMCA was performed by the Unscrambler® software (ver. 7.8: CAMO AS, Trondheim, Norway) and used to discriminate the similarity or dissimilarity among samples at the 95% confidence level.

2.4 K Nearest Neighbors (KNN)

KNN is a non-parametric method. In this method, the distance between an unknown sample of the prediction set and each of the objects of the training set was computed. The matrix of distances of the prediction set samples to all other points of the training set is computed (Roggo *et al*., 2003a). In this study, KNN was performed by use of in-house-written program in MATLAB® software ver. 5.3 (The MathWorks, USA). The correlation coefficient was used to measure the similarity of an object to its neighbors (Wu and Massert, 1997). The K neighbors of an unknown sample are the training samples with the highest correlation coefficient. The optimum number of K was determined by leave-one-out cross-validation method. This method was performed for K = 1 to 15. The number K which gives the highest corrective classification rate was chosen.

2.5 Artificial Neural Networks (ANNs)

ANNs is a non-linear and non-parametric method. It is composed of several layers of neurons: input, hidden and output layers. A neuron is a processing unit of which inputs are transformed by an activation function into the outputs (Blanco *et al*., 2000; Roggo *et al*., 2003a). In this study, ANNs was trained by using a back-error propagation algorithm. In order to acquire the optimal architecture of neural network, several ANNs systems with different number of hidden nodes varied from 1 to 4 were tested. The target vector describing the belongingness of the object to a class was set to the values of 1, 2 and 3 for corresponding to SPF-, SMF- and OF- groups, respectively. The actual output was assigned to a class by rounded each value to nearest integer (1, 2 or 3). The performance of ANNs was evaluated on the prediction set. The model yielding highest corrective classification rate with a small number of hidden node was selected as the optimal one. The networks were trained for at least 5000 epochs or until the error measurement approached 0.001. ANNs was performed by MATLAB® software ver. 5.3 (The MathWorks, USA).

The performance of the LDA, FALDA, SIMCA, KNN and ANNs classification models was assessed by calculating the proportion of correctly classified samples in the training and test sets.

3. Categorization of Thai Fish Sauces Based on Sensory Characteristics and Principal Component Analysis

The sensory characteristics of twenty fish sauces consisted of pure fish sauces and mixed fish sauces were investigated using the generic descriptive analysis method. Subsequently, all samples were categorized based on their sensory properties using the combination of cluster analysis and principal component analysis (PCA).

3.1 Samples

A total of twenty fish sauces were available used in this study. According to their bottles, twelve samples were pure fish sauces (P) and the remaining of eight samples were mixed fish sauces (M). All samples were transferred to tightly sealed plastic bottles and stored at 4 °C before measurement of their sensory properties.

3.2 Panelists

Twelve panelists (aged 22-28 years) from the Department of Product Development, Faculty of Agro-Industry, Kasetsart University, participated in this study. These panelists had been trained on the generic quantitative descriptive analysis method described in Lawless and Heymann (1998). They had some experience on descriptive tests and ability to perceive difference between test samples. The sensory panelists were selected according to guidelines in ISO (1993).

3.3 Sample Evaluation

All twenty Thai commercial fish sauces were examined during the orientation session to develop terms and definitions for describing sensory characteristics of Thai fish sauces. Panelists were trained in the definition of the attributes for 16 hours before testing samples. Five sessions were held for sample evaluation. Four samples were presented monadically in random during each session. Each panelist was served approximately 20 mL of a fish sauce sample in a 60-mL

plastic cup with cover slip at room temperature (25°C). Panelists evaluated intensities and scores of each attributes on a 15-point numerical scale divided into 0.5 increments, with 0 meaning none and 15 meaning extremely strong. Panelist cleaned their palates with mineral water and unsalted crackers between sample evaluations. A ten minute break was taken between sample analyses.

3.4 Sensory Data Analysis

Sensory data obtained by trained panel were analyzed in 3 ways:

(1) Two-way analysis of variance (ANOVA) was performed to determine significant differences in sensory quality between fish sauces samples. The mean scores of the fish sauce samples were considered. Two factors, the panelist and types of Thai fish sauce, were studied. ANOVA was performed using SPSS® software. The sensory attribute scores which significantly discriminated between samples were subjected to cluster analysis and principal component analysis.

(2) Cluster analysis was applied to the sensory attribute scores which significantly discriminated between samples. The attributes were also analyzed, standardize, by Hierarchical cluster analysis (HCA) with Euclidean distance and average linkage using SPSS® software. This analysis was carried out in order to identify clusters of samples having a similar sensory profile.

(3) Principal component analysis (PCA) of mean rating for each sensory attribute was used to illustrate the relationship among variables and samples. The PCA can be translated as finding the linear combination of the initial variables that contribute most to making the samples different from each other (Camo, 1996). PCA was conducted on the covariance matrix using the sensory attributes found to be significantly different with ANOVA. PCA was performed by the Unscrambler® software (version 7.8: CAMO AS, Trondheim, Norway). Full cross-validation was used to validate the model (Martens and Naes, 1989).The sensory attributes were standardized prior to the analysis by weighing variables by their standard deviations.

For interpretation of the results, the optimal number of the principal component (PCs) that explained most of the information in the data was determined (that is, model with a total residual variance close to 0 or a large total explained variance) (Suwansri *et al.*, 2002).

4. Correlation between Sensory, Chemical and Physical Properties of Thai Fish Sauces Studied by NIR Spectroscopy Combined with Chemometrics

The same samples used for categorization of fish sauces based on the sensory characteristics study were also used in this study.

4.1 Sensory Evaluation

According to the categorization of fish sauces based on the sensory characteristics study from the study in Part 3, the sensory scores were imported into the NIR spectral data files in order to investigate the relationships.

4.2 NIR Spectra Acquisition

NIR transflectance spectra were obtained from 1100 to 2500 nm at a 2 nm interval using an InfraAlyzer 500 spectrometer (Bran+Luebbe, Norderstedt, Germany) and a 0.3 mm British cup. The fish sauce samples were incubated at 29°C in a water bath prior to the NIR measurements. The spectra data were transferred into JCAMP.DX format and imported into the Unscrambler® software (version 7.8: CAMO AS, Trondheim, Norway). The spectra were not subjected to any preprocessing.

4.3 Correlation between Sensory Attributes and NIR Spectra

PCA was applied to reduce the number of NIR spectral data, to establish the relationships between the NIR data and sensory attributes, and to detect the most important factors of variability. The large numbers of NIR spectral data were reduced as loadings scores before interpret the relationships between the data. PCA was

performed by the Unscrambler® software (version 7.8: CAMO AS, Trondheim, Norway). Full cross-validation was used to validate the model (Martens and Naes, 1989).The sensory attributes and NIR loadings scores were standardized prior to the analysis by weighing variables by their standard deviations. For interpretation of the results, the optimal number of the principal components (PCs) that explained most of the information in the data was determined (that is, model with a total residual variance close to 0 or a large total explained variance) (Suwansri *et al.*, 2002).

4.4 Prediction of Chemical and Physical Properties using NIR Spectra

According to the development of quantitative models for chemical and physical properties of fish sauces in the study Part 1, the developed SCMWPLS models which yielded the correlation coefficients (R) larger than 0.70 were applied to predict the chemical and physical properties in twenty Thai fish sauce samples using their NIR spectra. The predicted chemical and physical properties were used to investigate the relationships with the sensory descriptive attributes. The block diagram for prediction of chemical and physical properties is shown in Figure 19.

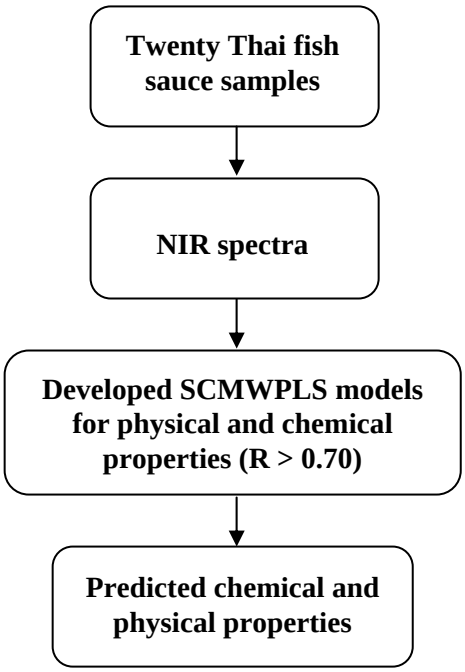

Figure 19 Block diagram for prediction of chemical and physical properties with developed SCMWPLS models using NIR spectra.

4.5 Correlation between Sensory Attributes and Predicted Chemical and Physical Properties

PCA was applied separately to establish the relationships between sensory and predicted chemical data, and sensory and predicted physical data. PCA was performed by the Unscrambler software® (version 7.8: CAMO AS, Trondheim, Norway). Full cross-validation was used to validate the model (Martens and Naes, 1989).The sensory attributes and predicted data were standardized prior to the analysis by weighing variables by their standard deviations. For interpretation of the results, the optimal number of the principal components (PCs) that explained most of the information in the data was determined (that is, model with a total residual variance close to 0 or a large total explained variance) (Suwansri *et al.*, 2002).

5. Development of Quantitative Models for Predicting the Sensory Properties of Thai Fish Sauces.

The same samples used for categorization of fish sauces based on the sensory characteristics study were also used in this study.

5.1 Sensory Evaluation

According to the categorization of fish sauces based on the sensory characteristics study from the study in Part 3, the sensory scores were imported into the NIR spectral data files and used as dependent variables (Y-data) in quantitative modeling.

5.2 NIR Spectra Acquisition

NIR transflectance spectra were obtained from 1100 to 2500 nm at a 2 nm interval using an InfraAlyzer 500 spectrometer (Bran+Luebbe, Norderstedt, Germany) and a 0.3 mm British cup. The fish sauce samples were incubated at 29°C in a water bath prior to the NIR measurements. The spectra data were transferred into JCAMP.DX format and imported into the Unscrambler® software (version 7.8: CAMO AS, Trondheim, Norway). The spectra were not subjected to any preprocessing. In order to investigate the relationships between the various samples, PCA was applied to the NIR spectra data.

5.3 Model Analysis

Both data from section 5.1 (Y-data) and section 5.2 (X-data) were used to develop the quantitative models. The models were constructed by using PLS calibration models with the aid of wavelength interval selection methods. The purpose of wavelength interval selection is to find out the suitable input wavelength variables for developing the calibration models. Two wavelength interval selection methods named i) moving window partial least squares regression (MWPLSR) and ii)

searching combination moving window partial least squares (SCMWPLS) were applied. The block diagram of quantitative modeling process for sensory properties is shown in Figure 20.

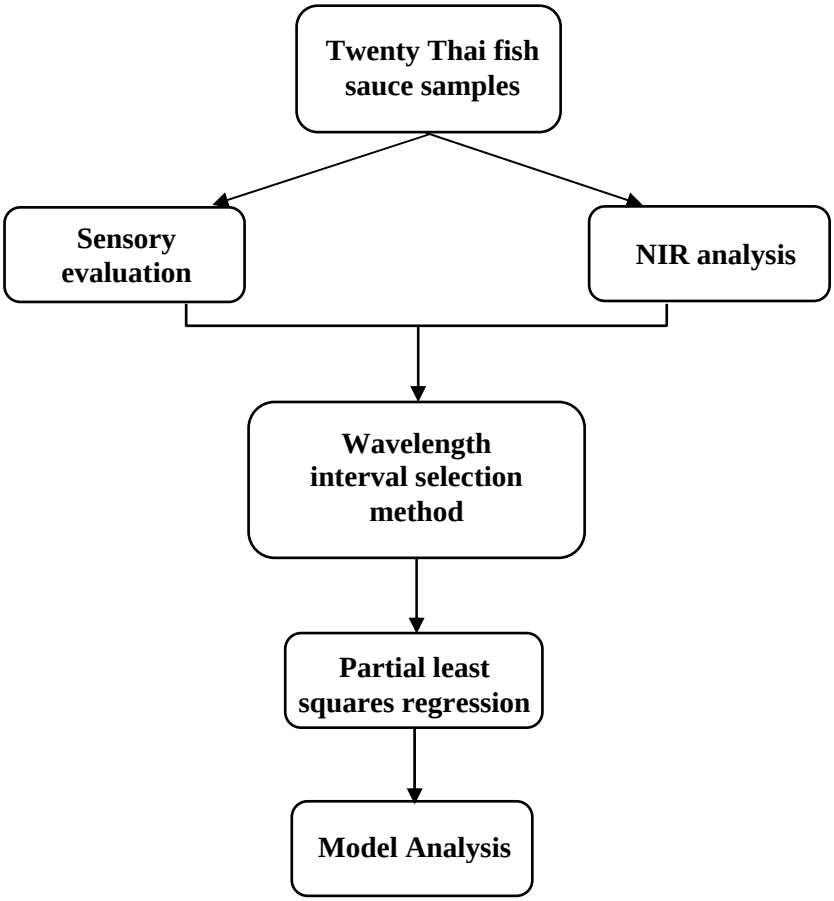

Figure 20 Block diagram of quantitative modeling process for sensory properties.

PLS regression was utilized to develop calibration models by using i) the whole spectra region, ii) the informative regions, iii) the direct combination of informative regions and iv) the optimized combination of informative regions. The informative regions and the optimized combination of informative regions were obtained by MWPLSR and SCMWPLS, respectively. MWPLSR with the window size of 20 spectra points and SCMWPLS were carried out by use of in-house-written program in MATLAB® software ver. 5.3 (The MathWorks,USA).

PLS models were performed by the Unscrambler® software (version 7.8: CAMO AS, Trondheim, Norway). The random cross-validation was used to develop and evaluate these regression models. The lowest root mean square error prediction in cross validation (RMSECV) was used to choose the optimum number of PLS factors in the model. The maximum number of PLS factors were set at 10 which usually sufficient to adequately describe a model (Clark *et al*., 2004). The RMSECV was calculated as equation:

$$\text{RMSECV} = \sqrt{\frac{\sum_{i=1}^{n} (C_{\text{NIRi}} - C_{\text{REFi}})^2}{n}}$$

where n is the number of samples included in the calibration matrix or the calibration set, C_{REF} the concentration in the sample as measures by the reference methods and C_{NIR} the concentration as predicted by PLS from the NIR spectrum.

In order to compare between the different established models for sensory attributes, the performance of models was quantified by determining the squared correlation coefficient (R^2), the ratio of prediction to deviation (RPD), and the range error ratio (RER). The RPD and RER are defined for a specific sensory characteristic (Mouazen *et.al*, 2005; Downey *et al.,* 2005). These ratios were calculated as equations:

RPD = Standard deviation / RMSECV

RER = (maximum-minimum)/RMSECV

The RPD ratio is desired to be larger than 2 for good calibration. In addition, a RPD ratio less than 1.5 indicates poor predictions and can not be used for further prediction. Practical utility of the calibration can also be assessed by using the RER which higher than 5.0 (Karoui *et al.*, 2006).