JOURNAL OF CHEMOMETRICS
J. Chemometrics 2002; **16**: 81–88

A phenomenological study of ripening of salted herring. Assessing homogeneity of data from different countries and laboratories

Rasmus Bro[1]*, Henrik Hauch Nielsen[2], Guðmundur Stefánsson[3] and Torstein Skåra[4]

[1]Chemometrics Group, Department of Dairy and Food Science, The Royal Veterinary and Agricultural University, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark
[2]Danish Institute for Fisheries Research, Department of Seafood Research, Ministry of Food, Agriculture and Fisheries, Bldg 121, Technical University of Denmark, DK-2800 Lyngby, Denmark
[3]Icelandic Fisheries Laboratories, Skulagata 4, 121 Reykjavik, Iceland
[4]NORCONSERV, Postboks 327, N-4001 Stavanger, Norway

Received 31 March 2000; Revised 2 January 2001; Accepted 21 March 2001

Data from ripening experiments of herring carried out at three Nordic fishery research institutions in the period 1992–1995 were collected and analyzed by multivariate analysis. The experiments were carried out at different times, with different stocks as raw material, using different types of treatments and analyzed in different laboratories. The question considered here is whether these data can be assumed to be one homogeneous set of data pertaining to ripening of salted herring or whether data from different labs, stocks, etc. must be considered independently. This is of importance for further research into ripening processes with these and similar data. It is shown in this paper that all data can be considered as one homogeneous data set. This is verified using resampling where latent structures are compared between different sample sets. This is done indirectly by testing regression models, that have been developed on one sample set, on other sample sets. It is also done directly by monitoring the deviation in latent structure observed between different sample sets. No formal statistical test is developed for whether samples can be assumed to stem from the same population. Although this can easily be envisioned, it was exactly the need for a more intuitive and visual test that prompted this work, developing different exploration tools that visually make it clear how well the data can be assumed to derive from the same population. Subsequently analyzing the data as one homogeneous group provides new information about factors that govern the ripening of salted herring and can be used in new strategic research as well as in industrial practice.

KEYWORDS: PCA; jackknife; resampling; population homogeneity; cross-validation

1. INTRODUCTION

Salted herring products are of importance for the pelagic fish industry in the Nordic countries. The raw material for the production is herring ripened by salting in barrels for several months. The production is mainly based on tradition and human experience [1]. Although considerable research has been carried out in the Nordic and European community, knowledge of the main factors governing the ripening process is still limited. During ripening, the herring develops a unique taste and texture. A proteolytic degradation of muscle proteins, releasing peptides and free amino acids, and the effect of salt on proteins are believed to contribute considerably to the sensory development. It is well documented that both digestive and muscle proteases participate in proteolysis during ripening [2–5]. However, it is unclear whether the digestive proteases only accelerate the process or lead to specific sensory characteristics. For example, herring salted without intestines also ripens [6]. Small peptides, free amino acids and free fatty acids are released during ripening [4,7] and are believed to contribute to the taste of the salted herring, but no direct link between these compounds and taste has yet been proven. The sensory development is also influenced by other factors such as the salting process [8] and the herring stock used [9].

Several Nordic institutes within fisheries research have been involved in collaborative projects in this field during the last decade. These institutes have carried out a number of salting experiments using herring from different stocks,

* Correspondence to: R. Bro, Chemometrics Group, Department of Dairy and Food Science, The Royal Veterinary and Agricultural University, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark
Contract/grant sponsor: LMC (Center for Advanced Food Studies)
Contract/grant sponsor: AQM (Advanced Quality Monitoring in the Food Production Chain)
Contract/grant sponsor: PMP (Center for Predictive Multivariate Process Analysis)

DOI:10.1002/cem.691

caught in different seasons, etc. A tremendous amount of data are thus available. The data considered here come from storage/ripening experiments with salted herring that have been collected at the Danish Institute for Fisheries Research, Department of Seafood Research, Denmark (abbreviated Denmark/Den in the following), the Icelandic Fisheries Laboratories, Iceland (abbreviated Iceland/Ice) and NOR-CONSERV, Norway (abbreviated Norway/Nor) in the period 1992–1995. The results reported here are part of a larger project described in a separate report [10].

Collective analysis of all existing data from these different institutes has never been carried out before. Combining the data may make it possible to retrieve more detailed information about the ripening process. Some institutes have focused on sensory evaluation of products, others on chemical or physical descriptions of the ripening process. Only by combining the data from different locations can the full picture be obtained. For example, making predictive models between common variables measured in all countries (pH, protein, etc.) and specific variables of e.g. sensory attributes, enzymatic variation, peptides or similar may make it possible to understand the variation in sensory quality on a more fundamental level. This may e.g. lead to a better understanding of the ripening process, development of inexpensive indicator variables of industrial importance, better means for controlling and monitoring the ripening, etc.

These overall aims, however, are not specifically dealt with here. In this paper the first step is taken in this meta-analysis by scrutinizing the measurements which have been carried out in the three countries involved.

These common data are mainly simple physicochemical measures reflecting overall properties of the ripening process. The salting process is mainly characterized by the variation of salt and water in herring and brine over time. This variation is caused by an extraction of water from the herring as a consequence of the dry salting, followed by a diffusion of salt into the herring. This variation is thus assumed to form a strong basic trend in all data sets. However, differences in analysis methods, stocks, treatments, etc. may cause systematic differences between the data from the three countries. Thus the data may not be directly comparable. It is the primary aim here to evaluate this problem first and then subsequently explore the data. Specifically, the following issues are considered.

1. Investigate if variables are coherently measured and sampled across countries/laboratories.
2. Investigate if the data of different fish analyzed in different countries can be described by the same model, such that conclusions pertaining to the data of one country can be transferred to experiments from other countries as well.
3. Investigate if the data reveal patterns of interest to the fish industry for understanding the effect of different treatments and possibly for obtaining inexpensive quality indicators.

If the data sets presented here can be combined into one data set, it suggests that fundamental results from the data can be assumed to be generally applicable for similarly made salted herring.

Table I. Overview of stocks and treatments of herring in the three different countries. For abbreviations see text

Country	Stocks	Processing
Iceland	I	B, C, F, G
Norway	A, N	B, F, G, Fs
Denmark	A, I, K	B, G

2. DATA: RIPENING EXPERIMENTS AND ANALYSES CARRIED OUT

Data from a total of 20 ripening experiments (Denmark, 10; Iceland, Seven; Norway, three) with a total number of 46 trial groups have been analysed. The experiments have been carried out in the period October 1992–November 1995. Four different stocks were used: Kattegat herring (K), Icelandic summer spawning herring (I), Atlanto Scandic spring spawning herring (A) and North Sea herring (N). Five different types of raw material processing were examined in the ripening experiments: beheaded (B), beheaded and gutted (G), fillets (F), skinned fillets (Fs) and clipfish (C). The different stocks and treatments were not distributed evenly between the three countries, as evidenced in Table I. All salting and storage have been done according to the same normal industrial practice in the three countries (see earlier reports [5,6] for details concerning this aspect). All ripening experiments have been carried out at 5 °C.

The present data mainly contain simple chemical and physical analyses that are measured by all three laboratories. These data form the common axis around which to draw inter-country conclusions. The variables are primarily basic measures of overall properties of the brine and the fish, such as analyses of proximate composition. Table II describes the variables. Names of variables are written as abbreviations. The last letter in most of the variable names refers to whether the variable has been measured in brine (B) or in fish muscle (M). For some variables almost a third of the observations are missing, but for most variables only a few observations are missing. The missing variables are generally randomly missing; that is, no overall objective caused specific analyses to be skipped. For some pH measurements it is known, though, that these were skipped at certain points in an experiment time owing to stabilization of the pH. This aspect has not been pursued further here. Mostly, practical time considerations simply prevented all analyses from being performed.

The variables have been measured to give an overall description of the raw material, final product and physical and chemical changes taking place in the herring during the ripening process and, furthermore, to ensure that samples taken for analyses during a ripening experiment are comparable and representative.

Throughout this report, all data are scaled to unit variance and mean-centered across samples. All analyses are carried out in Unscrambler^R 7.5 (CAMO ASA, Norway) and MATLAB^R (The Math Works, Inc.). One extreme sample is removed, leading to a sample set of 265 samples. Occasionally, a few samples are also removed in specific models for various reasons (e.g. owing to many missing elements).

Table II. Explanation and abbreviations of variables used

Abbreviation	Parameter	Description
ProteinM	Protein, muscle	Raw material/product
ProteinB	Protein, brine	Solubilization of protein fragments and salt-soluble protein
AshM	Ash, muscle	Salt uptake (salt content generally 1% lower than AshM)
pHB	pH, brine	Previous studies have concluded that deviating pH values can indicate foul ripening
TCAB	Trichloroacetic acid-soluble nitrogen, brine	Level of small nitrogenous compounds and protein degradation products that are solubilized in brine. Smell of brine is a traditional quality parameter
TCAM	Trichloroacetic acid-soluble nitrogen, muscle	Level of protein degradation (caused by enzymes)
TCAIndexM	Trichloroacetic acid index, muscle	Level of protein degradation relative to total protein content
TCAIndexB	Trichloroacetic acid index, brine	Level of protein degradation relative to total protein content
Water	Water, muscle	
Fat	Fat, muscle	Raw material/product

These samples will not be reported here for brevity. The time-zero samples are left out, as these behave significantly differently from the remaining ones (though similarly across countries).

The data used in this paper are also available at http://www.models.kvl.dk.

3. OUTLINE OF PAPER

The strategy for the multivariate analysis is divided into several steps, each addressed in individual sections in the following.

- A method for assessing population *homogeneity*. An *ad hoc* method is developed with which it is possible to assess whether samples from different subgroups can be considered to stem from the same basic group (population) and thus be assumed to vary according to the same latent variables, i.e. whether the samples are homogeneous with respect to the latent variation in the data.
- Assessing population homogeneity for the data. The method described is then used to assess the given data.
- Further validation of conclusion. Through the use of a split-half approach it is further shown that the conclusions reached above are valid.
- Principal component analysis of data. Having ensured that the data can be described within the same model, the data are superficially analyzed using principal component analysis. This is, on a 'fish technology level', the primary aim of this paper, but owing to the inherent problems, the above chemometric problem has to be solved before this analysis can be pursued.

4. A METHOD FOR ASSESSING POPULATION HOMOGENEITY

To see if and to what extent valid information is present in the data, it is important first to contemplate on the types of deviations expected between samples from different countries. Although some of these deviations are not interesting per se, it is important to be able to evaluate and quantify them.

The basic assumption underlying the following discussion is that within the 10 measured variables there is an underlying latent structure that may be approximated by a bilinear model. Thus the 10 measured variables are manifest variables expressing the variation in a lower number of more fundamental latent variables that may be found as linear combinations of the manifest variables. Therefore a model of all data may be written as

$$X = TP^T + E \quad (1)$$

where X is an $I \times J$ matrix holding the preprocessed data (I samples and J variables). The matrix P is the loading matrix and defines the latent variables. It is of size $J \times F$, where F is the number of latent variables to be chosen. For each latent variable f the fth column of P defines this variable as a weighted sum of all J manifest variables, the jth element being the weight applied to the jth manifest variable.

The principal component analysis (PCA) model states that, indeed, the data are described only by variations of common phenomena, i.e. the loadings P. These loadings should reflect the variation induced by the herring and by the treatment of the herring only. However, it may be anticipated that several sample-specific variations may occur as well. Examples could be the following.

- Samples from different stocks may behave differently and have different biophysiological intrinsic correlations and latent structure.
- Samples taken from the same brine (but at different times) may have variation in common which is not shared by samples from other brines.
- Samples of similar types of treatment may have variation in common that is not shared by samples treated differently.
- Samples from different countries may have their own country-specific variation that arises owing to differences in analytical procedures or in stocks used.

The effects of stocks, brines, treatments and countries are nested and even partly confounded, e.g. because some treatments have only been performed in one country. However, if it can be shown that there is no effect of different countries, then implicitly it is also shown that there is no special variation due to stocks, brine or treatments. Hence, for simplicity, it will first be assessed whether

samples from different countries can be considered to behave according to the same underlying latent variables.

Note that it is explicitly excluded to compare the distributions of the variables here. It can for example be envisioned that a test be devised for assessing if the data from the three different countries follow the same multinormal distribution. However, this does not make sense in this situation. The experiments have been designed such that different seasons, stocks and treatments are used in different countries (see Table I). Thus it is known a priori that the data from the different countries do not follow the same distribution on an empirical/manifest level. That is, empirically estimating the distribution of a certain variable in different countries can give widely different results, because the experimental design is confounded e.g. with the countries, laboratories and technician. For example, fat content in the herring is generally much lower in the Icelandic samples than in the Danish or Norwegian samples. This is so because the Icelandic experiments all are carried out with an Icelandic herring stock that biologically has a lower fat content than the herring stock used for the Danish and Norwegian samples. This, however, does not exclude the possibility that the basic variation in all herring is governed by the same underlying latent phenomena.

There are several ways to verify if the data can be assumed to arise from the same common phenomena. It is natural to seek an ANOVA-like argument for testing this, but this will not be directly pursued for two reasons: there is no explicit design behind the data and there is a clear latent variable structure in the measured data [11]. Although these arguments do not prevent (M)ANOVA-like testing, we will show that there is a powerful and directly appreciable way of verifying the degree of 'commonness'. This approach also provides an indirect illustration of a very common misuse of cross-validation.

A simplified example will be described first to exemplify how commonness or homogeneity will be evaluated. Consider a data set where the height, leg length and shoe size are given for a number of persons. A test is sought to assess whether subsets of persons (e.g. African, European and Asian people) all vary according to the same basic latent variables. Assume that a fourth variable, say weight, is also given. A calibration model may be built for predicting weight from the three main variables using a rank-reduced regression model such as principal component regression or partial least squares regression (PLS) [12]. In this case, principal component regression is the appropriate tool to use, for reasons explained later. Assume, for simplicity and with no loss of generality, that a one-component PCA model can describe the independent data ($X \approx tp^T$) and that the score t is approximately linearly related to the weight. The validity of such a regression model relies on two key aspects. The first is that all samples can be described by the same latent phenomena (in this case only one), and the second is that the relationship between these latent variables and the weight is the same for all persons.

Consider now a situation in which this hypothesized model is determined from persons belonging to different groups. If all persons behave similarly, then it is possible to estimate the score values from persons from a new independent group from the predefined PCA model loadings and subsequently predict the weight of these persons using the regression model. Consider, on the other hand, a situation in which the 'persons' in the new group are all pigs. Disregarding the fact that shoe sizes may be difficult to obtain, it is quite clear that the persons in this group have a fundamentally different relation between the variables height, leg length and shoe size. The small shoe sizes, legs and heights of pigs will lead to scores that are relatively low. The first main conclusion is therefore that if the new subgroup is not of the same fundamental type, then the PCA model will provide misleading scores (and significantly high residuals within the PCA model). More important is the fact that the subsequent regression model will also be misleading. If a human of say 80 kg gets a certain score from his/her shoe size, leg length and height, then this score times the regression coefficient will be approximately 80. However, since a pig of 80 kg gets a much lower score per definition, the weight will naturally be predicted incorrectly low. Thus, with this approach, inconsistencies are revealed in two ways. The PCA model as such may reveal inconsistencies or otherwise the subsequent regression model may reveal inconsistencies.

It follows that one way of assessing homogeneity between countries for the fish data is to compare the predictions of some external variable (such as the weight in the above case) across new countries and across known countries. If the predictions are of similar quality, no fundamental differences exist. Oppositely, if the predictions on data from new countries are significantly poorer than predictions of known countries or the X-residuals are much higher, then differences do exist.

It is important to use principal component regression rather than e.g. PLS. In PLS the stated objective is to include only the most predictive part of the independent data, which is why PLS typically uses fewer components than principal component regression. However, in this case, using only some of the systematic variation in the independent data would impede the conclusion that the same basic variation is present in several types of subsamples. Rather, the conclusion would be valid only for the part of the data used for prediction in the regression model. Hence there would be a risk of incorrectly concluding homogeneity in a situation where the inconsistent part of the data is filtered away by the PLS model.

5. ASSESSING POPULATION HOMOGENEITY FOR THE DATA

5.1. Strategy

The specific procedure for testing the data is the following. The external variable storage time is used as predictand* and

*Distinct curvature in the predictions of storage time indicates a non-linear relationship between time and the data. This seems natural, since the biochemical variation must be assumed to level off in time. The logarithm of time is used instead, excluding the samples with storage time zero. This leads to predictions with no obvious curvature.

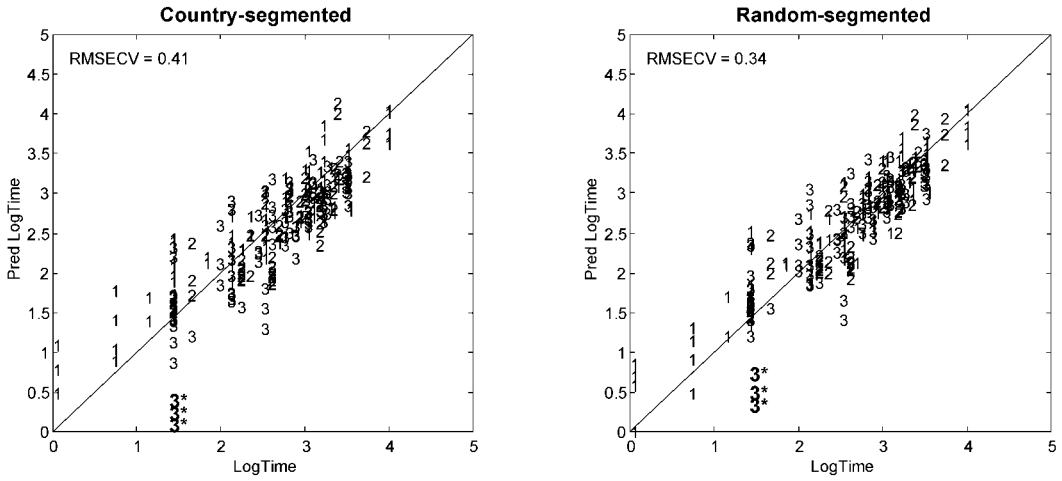

Figure 1. Principal component regression model predicting LogTime from data. Country-segmented (left) and randomly three-segmented (right) cross-validated predictions. Seven principal components used in both cases. The numbers on the samples refer to the countries (1 = Den, 2 = Ice, 3 = Nor).

the data in Table II are used as independent variables. It is noted that time could also be used as one of the independent variables. Using time as a dependent variable is sensible though, because, on one hand, a relatively direct relation between time and the measured variables is assumed to exist and, on the other hand, fundamental differences in this relationship are expected if the data across countries are not homogeneous. After a thorough analysis of the data having time as dependent variable, results will also be shown where time is used as an independent variable and where each of the variables is left out in turn. Principal component regression models are built in such a way that as much variation as possible is explained without sacrificing the predictive quality of the model. To assess how well e.g. time is predicted in new samples, all data pertaining to one country are left out and a principal component regression model is built from the data from the two other countries. This model, given by the loading parameters P and the regression coefficients b (and offsets), is then used for predicting the time of the individual samples from the left-out country. This is repeated for all three countries in turn. The quality of the predictions is an estimate of how well the model generalizes to completely new samples. This cross-validation provides a very severe test of the extrapolation properties of the model.

On the other hand, if some randomly chosen samples are left out, then samples of similar treatment, country, etc. are used both for estimating the parameters as well as for predictions. Thus the quality of random cross-validation provides a measure of the maximal quality predictions for the given data. The difference between the quality of country cross-validation and random cross-validation provides an estimate of the effect (error) introduced by using the model on data from a new country. If the quality of the two models is the same, no country effect is present.

This non-trivial use of cross-validation is of fundamental importance. It provides a very direct way of performing analysis of variance/assessment of effects in a sound and appreciable way. In traditional hypothesis testing, conclusions are based on an 'abstract' mathematical model and on fitted values. Thus bias can be introduced owing to the use of an incorrect model structure, and overly optimistic results can be obtained, because no measure of the inter- and extrapolation and sampling error is available. Both problems are overcome here.

5.2. Results

The most important quality criterion is that the regression model should use most of the information in the independent data. Otherwise, any conclusions drawn will only be valid for a small fraction of the data and hence of no use in this context. For the present regression models the PCA model of the independent data uses seven principal components and describes between 92% and 98% of the variation in individual variables (validated by cross-validation). Therefore conclusions drawn in the following can be assumed to be valid for the whole data set as such.

The random- and country-segmented cross-validated predictions of time are shown in the plots in Figure 1. As can be seen in both plots, the predictions are reasonable, especially given that the chronologically measured storage time cannot be expected to be exactly the same 'biotime' in different experiments. Therefore the results are quite impressive. There seems to be a rather uncertain area at LogTime 1.5. This could be due to the fact that this is approximately the time where the chemical changes, as a result of diffusion of water/salt, are most dramatic. Therefore this is also the time where most samples have been measured. Note also, though, that the visual effect of high variability is mainly caused by the three Norwegian samples labeled with an asterisk in the lower part of the plots.

In the left-hand plot the cross-validated predictions leaving out whole countries are shown. It is remarkable that even though there are differences in treatments, differences in herring stocks as well as possible differences in analytical procedures, the predictions obtained for completely new countries are not much worse than those obtained using samples from the same populations as used for estimating the

Table III. Quantitative results from calibration models for each variable. Random and cross-country cross-validation results are shown (RMSECV/correlation coefficient). For pH in brine, no valid results were obtained

Variable	Random	Cross-country
pH, brine	–	–
Protein, muscle	0.64/0.93	0.82/0.90
Protein, brine	0.58/0.94	0.58/0.94
Water	2.00/0.91	2.54/0.87
Ash, muscle	0.67/0.81	0.84/0.68
Fat	1.73/0.96	2.72/0.89
TCAIndex, muscle	0.96/0.99	1.50/0.98
TCAIndex, brine	5.64/0.94	6.51/0.92
TCA, muscle	0.14/0.99	0.21/0.98
TCA, brine	0.38/0.97	0.37/0.97

model. This is especially remarkable because, in order to predict samples of new countries, there are only samples from two other countries for modeling any country-specific variation.

The positive result is further verified by building calibration models for each of the variables in Table II leaving the particular variable out of the independent set and including LogTime as independent variable. The results are given in Table III, which provides more or less similar results to the ones already given.

The calibration models are based on PCA models that seek to describe all the systematic variation, and the regression coefficients for the prediction problems are all of similar order of magnitude. Thus all variables are important both with respect to the structure in X and with respect to predicting time. If there are country- or experiment-specific variations in the underlying phenomena, it is unlikely that this will not lead to a substantial change in the scores calculated from the PCA model. Additionally, this change would lead to a biased prediction of time. Since the quality of the predictions is not much worse for the country-validated problem, it may be concluded that the samples all behave approximately according to the same latent phenomena. Any conclusion drawn from one part of the data set may therefore be expected to hold for the remaining data.

6. FURTHER VALIDATION OF CONCLUSION

In order to further validate the results, the seven-component PCA model of the data is investigated in more detail. If the above conclusion is correct, then the data taken from one country, with corresponding laboratory, herring stock and treatments, should vary according to the same latent variables as the data from other countries. This means that loadings found from PCA models of different subgroups should span the same subspace. A way to verify this is to look at the jackknife estimate of the uncertainty of the estimates of the loadings [13]. However, even though powerful, jackknifing introduces a risk of misinterpretation, because the resampled estimates are not independent. In order not to introduce any possible error arising on that account, another approach is pursued here. For the data of each country a PCA model is fitted, giving three different PCA models all independent of each other. The 10×7 loading matrices from these three models are then investigated to see if they span the same subspace. This is done by rotating orthogonally the Icelandic and Norwegian loading matrices to maximal agreement with the Danish loading matrix. Thus, if $P^{(ice)}$ is the loading matrix from the PCA model of Icelandic samples, and $P^{(den)}$ and $P^{(nor)}$ are defined likewise, then the optimally rotated $P^{(ice)}$ is given as follows. Let

$$M = (P^{(den)})^T P^{(ice)} \tag{2}$$

and

$$Q = M^T (MM^T)^{-1/2} \tag{3}$$

Then the optimally rotated loading matrix $\tilde{P}^{(ice)}$ is given as [14]

$$\tilde{P}^{(ice)} = P^{(ice)} Q \tag{4}$$

The resulting three sets of first and second loadings are shown in Figure 2. As can be seen, the loadings are very close, confirming that the three subsets of data span the same variation.

This approach of assessing homogeneity is conservative in the sense that the three loading matrices are estimated completely independently, also with respect to treatments, laboratory, etc. However, there is still a risk that too optimistic results are obtained, because the PCA models almost span the complete space (seven components out of 10 variables fitting up to 99% of the variation). If the latter of the seven components are only indicative of noise, however, then these latter components would not be possible to rotate to similar configurations. In Figure 3 the resulting loadings of the sixth and seventh components are shown. As can be seen, these are as close to being identical as those presented in Figure 2.

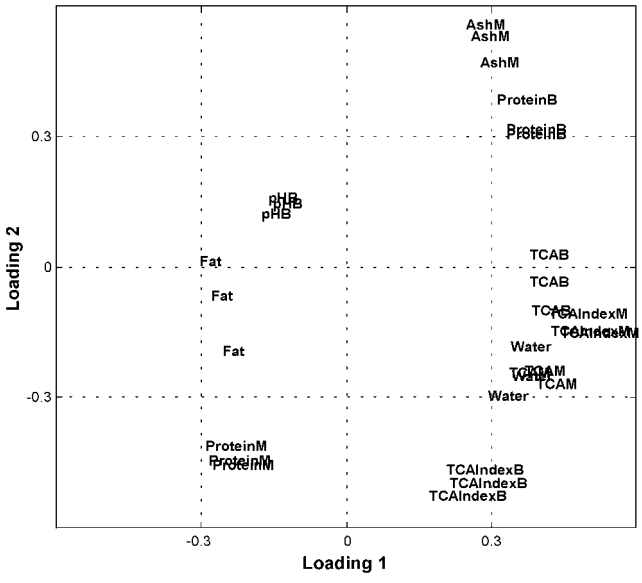

Figure 2. Loadings from three independent PCA models of data from Danish, Icelandic and Norwegian samples upon orthogonal rotation to maximal agreement.

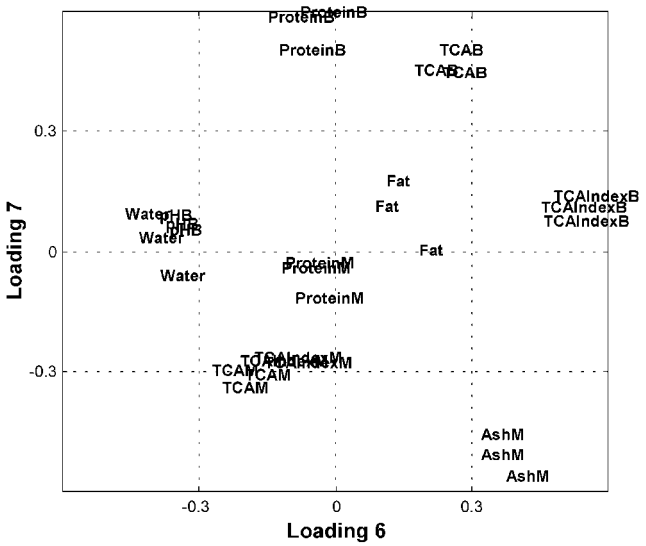

Figure 3. Loadings from three independent PCA models of data from Danish, Icelandic and Norwegian samples upon orthogonal rotation to maximal agreement.

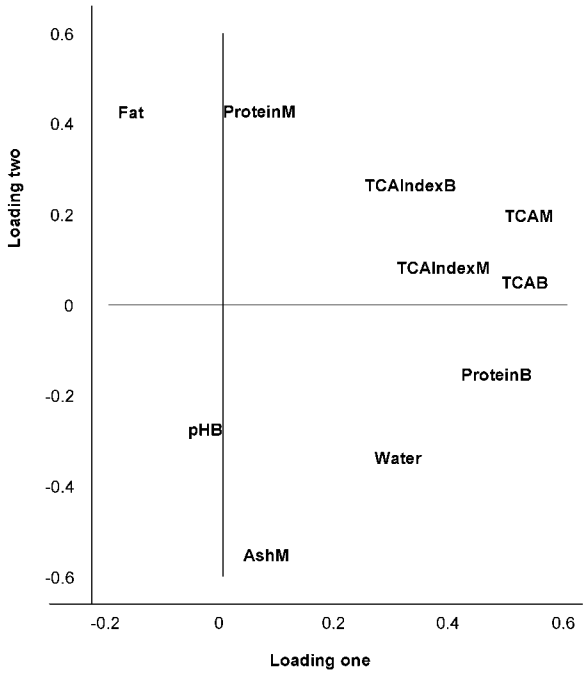

Figure 4. Loading plot from a PCA model of all data except the time-zero samples. 40% of the variation is explained by the first component and 24% is explained by the second. Loadings are normalized, so biplot theory [16] is not applicable.

7. PRINCIPAL COMPONENT ANALYSIS OF DATA

The data have been shown to be consistent across different types and represent the same type of basic variation. Any systematic differences in latent variation arising from the use of different laboratories, treatments, etc. can be assumed to be insignificant and therefore the data may be investigated as one homogeneous group of data.

A PCA model is fitted to the data in order to investigate the nature of the data. A loading scatter plot is shown in Figure 4 of the two first loadings. Many interesting aspects are revealed in this loading plot about the interrelations of the variables. The first two components shown explain 64% of the total variation, hence this plot is quite descriptive of the total variation.

For example, it is seen that the variables Fat and Water appear to be negatively correlated. This is seen by their position near the same line and far position in opposite direction of (0,0) [15,16]. Thus, when Water is high, Fat is low and *vice versa*. This is easily confirmed by plotting the raw data as shown in Figure 5. This is a generally well-known and valid correlation in fresh fish [17] and will therefore not be regarded as an important finding in this study. Nevertheless, it is a simple and powerful demonstration of how easily the multivariate data are explored using multivariate tools.

The variables TCAM and TCAIndexM more or less contain the same information, as evidenced through their high loadings and small interdistances. TCAM is the direct measurement of small nitrogen compounds in muscle and TCAIndexM is TCAM as percentage of the total protein content in muscle, which is constantly high, causing the correlation between the two variables. The correlation between the brine (TCAB) and the muscle (TCAM) is seen because part of the TCA measured is muscle protein that is broken down into smaller fragments. These fragments occur in the brine as well as in the muscle. The correlation between TCAB and ProteinB is probably due to the fact that part of the protein measured in brine comprises low-molecular-weight peptides and free amino acids that are also measured by TCAB. While ProteinM varies almost independently of TCAM, it is negatively correlated with AshM. This negative correlation could be due to a decrease in measured protein caused by the increase in salt content in muscle during storage.

As can be seen, the PCA loading plot immediately provides an overview of all the variables of the data set. The simple conclusions drawn from the loading plot are easily verified, e.g. using a scatter plot as shown in Figure 5.

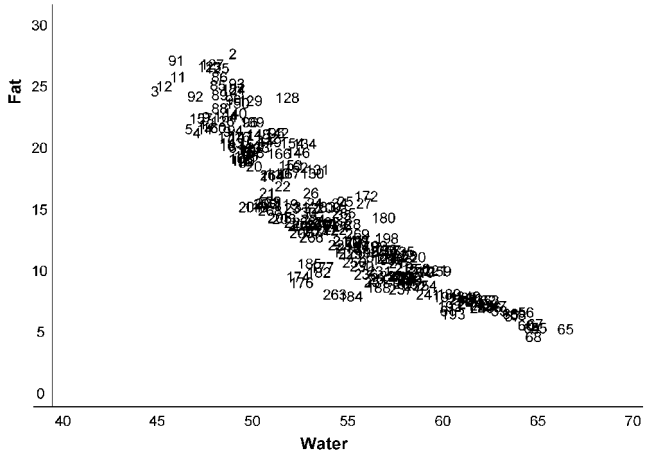

Figure 5. Scatter plot of Fat versus Water (time-zero samples excluded). Correlation is −0.91.

It is quite difficult to comprehend all the detailed facets of the data by looking only at the raw data, but the PCA model provides a condensed simultaneous description of all variables and samples.

8. CONCLUSION

The success of this investigation has important ramifications from both an academic and an industrial point of view. Even though the fish in these experiments come from different stocks, etc. and have different characteristics, it has been verified that, fundamentally, all ripened herring can be described by variations in the same basic latent variables. This has been shown using specially designed cross-validation as well as split-half analysis.

The results obtained are only indicative for salted herring. Even for the raw material for which data are also available, most of the results do not apply. However, for the particular population of salted herring it has amply been demonstrated that the found models and correlations are valid. Combining the present data with additional sensory and chemical profiling data available for a subset of the present data can give new valuable insight into ripening. However, already the correlations found here are of importance. For example, the protein in brine is correlated to TCAM (trichloroacetic acid-soluble nitrogen in muscle), which expresses the degree of protein degradation in salted herring during ripening. Protein degradation in salted herring is known to be an important part of the ripening process, and protein content in brine may therefore be used as a cheap and fast indicator for ripening of salted herring. The correlation between water and fat may also be useful for the industry, because a dry matter/ash analysis is cheaper and less resource-demanding than a fat analysis and will at the same time provide information about the salt content of the product. Both fat and salt are important process and quality parameters of the salted herring.

If e.g. the sensory quality determined only for Norwegian samples can be predicted from the variables measured here, then sensory quality can thus be approximated by a linear combination of these variables. To the extent that the parameters measured here are assumed to provide a full picture of the factors affecting sensory quality, such a sensory model can be applied to data from other countries. Such matters may now be pursued in light of the results obtained here and are highly relevant because additional measurements are available from the different countries.

Acknowledgements

R. Bro gratefully acknowledges financial support provided by LMC (Center for Advanced Food Studies), AQM (Advanced Quality Monitoring in the Food Production Chain) and PMP (Center for Predictive Multivariate Process Analysis) supported by the ministries of research and industry.

REFERENCES

1. Voskresensky NA. Salting of herring. In Fish as Food, vol. III, Borgström G (ed.). Academic Press: New York, 1965; 107–131.
2. Nielsen HH. Proteolytic enzyme activities in salted herring during cold storage. *PhD Thesis*, Danish Institute for Fisheries Research, Technical University of Denmark, Lyngby, 1995.
3. Nielsen HH, Børresen T. The influence of intestinal proteinases on ripening of salted herring. In *Seafood from Producer to Consumer, Integrated Approach to Quality*, Luten JB, Børresen T, Oehlenschläger J (eds). Elsevier Science: Amsterdam, 1997; 293–304.
4. Olsen SO, Skåra T. Chemical changes during ripening of North sea herring. In *Seafood from Producer to Consumer, Integrated Approach to Quality*, Luten JB, Børresen T, Oehlenschläger J (eds). Elsevier Science: Amsterdam, 1997; 305–318.
5. Stefansson G, Nielsen HH, Skåra T, Oehlenschläger J, Schubring R, Luten J, Derrick S, Gudmunsdóttir G. Frozen herring as raw material for spice-salting. *J Sci Food Agric* 2000; **80**: 1319–1324.
6. Stefansson G, Nielsen HH, Gudmunsdóttir G. Ripening of spice-salted herring. TemaNord 1995:613, Nordic Council of Ministers, 1995.
7. Engvang K, Nielsen HH. In situ activity of chymotrypsin in sugar-salted herring during cold storage. *J. Sci. Food Agric.* 2000; **80**: 1277–1283.
8. Gudmunsdóttir G, Stefansson S. Sensory and chemical changes in spice-salted herring as affected by handling. *J. Food Sci.* 1997; **62**: 894–897.
9. Schubring R, Oehlenschläger J. Comparison of the ripening process in salted Baltic and North Sea herring as measured by instrumental and sensory methods. *Eur. Food Res. Technol.* 1997; **205**: 89–92.
10. Nielsen HH, Bro R, Stefansson G, Skåra T. Salting and ripening of herring— collection and analysis of research results and industrial experience within the Nordic countries. TemaNord 1999:578, Nordic Council of Ministers, 1999.
11. Dijksterhuis GB, Gower JC. The interpretation of generalized Procrustes analysis and allied methods. *Food Qual. Prefer.* 1991; **3**: 67–87.
12. Martens H, Næs T. *Multivariate Calibration.* Wiley: Chichester, 1989.
13. Martens H, Martens M. Modified jack-knife estimation of parameter uncertainty in bilinear modelling by partial least squares regression (PLSR). *Food Qual. Prefer.* 2000; **11**: 5–16.
14. Cliff N. Orthogonal rotation to congruence. *Psychometrika* 1966; **31**: 33–42.
15. Gabriel KR. The biplot graphic display of matrices with application to principal component analysis. *Biometrika* 1971; **58**: 453–467.
16. Gower JC. A general theory of biplots. In *Recent Advances in Descriptive Multivariate Analysis*, Krzanowski WJ (ed.). Clarendon: Oxford, 1995; 283–303.
17. Kent M, Lees A, Christie RH. Seasonal variation in the calibration of a microwave fat: water content meter for fish flesh. *Int. J. Food Sci. Technol.* 1992; **27**: 137–143.

