19 Determination of the Composition of Foodstuffs Using Microwave Dielectric Spectra

Frank Daschner, Reinhard Knöchel

Christian Albrechts University of Kiel
Kaiserstrasse 2; 24143 Kiel; Germany

19.1 Introduction

Water is often added to foodstuffs during processing. This is inevitable when the raw materials have to be cleaned. Furthermore the nutritiousness is retained and oxidation of the surface is reduced when the raw materials are wet. Because water is inexpensive it is sometimes added deliberately in order to increase the financial profit. For quality control purposes there is a need to measure the water content of the foodstuffs. Furthermore the determination of other constituents like protein, fat, and salt is needed. There are many measurement systems available to measure the composition of foodstuffs, but they have the disadvantage that they can only measure one constituent and they are relatively expensive. In addition they are very time consuming. A measurement in real time in order to control the production process is not possible. Generally these methods alter the materials or they are destructive. In short there is a need for inexpensive, fast, non-destructive, and compact systems for measuring the composition of foodstuffs.

The method followed here to realize such a system uses the fact that a change in the composition of the foodstuffs also alters dielectric spectra in the microwave region. The problem is to relate changes in the dielectric spectrum to the composition of the material. Kent et al have already carried out investigations in this area. They used principal components analysis and regression [1, 2]. Other methods to described here are used for comparison purposes. At first a dielectric modeling is applied. Due to the high complexity of foodstuffs it is difficult to get precise results with this approach. Therefore multivariate calibration methods are applied. The measured dielectric data are used in a multiple linear regression. An orthogonalization of the measured data is carried out with partial least squares regression to improve the performance. Non-linear processing of the data is described using artificial neural networks. Finally all these methods are compared.

19.2 Experiments

The water content of flesh can be increased by dipping it in polyphosphate solution and saline solution. The two treatment parameters are the dipping time and the concentration of the solution. Because they have a different influence on the dielectric spectra of the samples, both parameters have to be taken into account. For example, the ionic conductivity of the samples will increase with the concentration of the solution.

19.2.1 Preparation of the Samples

In the framework of a project financed by the European Commission (FAIR CT97-3020) many experiments with poultry, pork, and fish samples were carried out. Two test series were selected to demonstrate the methods which were utilized. For the first experiment prawns (*Pandalus borealis*) were peeled and treated in a polyphosphate solution with different concentrations (0%, 0.5%, 1%, and 2%) and for different times (0 h, 2.5 h, 24 h, and 48 h). The temperature of the solution was 4–5°C. After the treatment the prawns were sieved and gently dried by dabbing with a paper towel. The amount of added liquid was determined by weighing the sample before and after the treatment. The samples were minced and halved. One part was used for the dielectric measurements while the other part was sent to an external laboratory which determined the composition of the samples using the conventional methods mentioned above. These measurements served as reference for the methods discussed here.

In the second experiment herring (*Clupea harengus*) was treated with a solution with constant polyphosphate content (2%) while the NACl concentration was varied (0%, 0.5%, 2%, and 3%). The treatment time was 0 h, 2.5 h, 24 h, and 48 h again. In comparison to the prawn samples the herring samples had a naturally high variation of fat content.

19.2.2 Dielectric Measurements

The dielectric measurements were carried out using an open-ended coaxial line with a diameter of 3 mm. This sensor has been well investigated and has become the standard for measuring dielectric spectra with liquid and soft materials [3–5]. The dielectric probe kit HP85070 in combination with an HP8510 automatic network analyzer was used for the measurements. The dielectric spectra of the samples were measured in the frequency range from 200 MHz to 12 GHz at 31 frequency points with logarithmic separation. Each measurement was repeated five times. Outliers caused by bad contacts between the probe and the sample were removed. The means of the permittivity values were calculated and used subsequently in order to remove the influence of any remaining inhomogeneities in the samples.

Generally the permittivity of a sample depends on its temperature. Therefore the dielectric measurements were taken at a sample temperature of 3°C, 8°C, and 20°C and 30°C. A total number of 91 data sets were obtained in the prawn test series and 52 in the herring test series.

19.3 Qualitative View on the Influence of the Sample Treatment

Due to the fact that the investigated foodstuffs have relatively high losses in the measured frequency range, the dielectric spectra are complex. Before the processing of the measured data is discussed in detail a qualitative view on the influence of the treatment of the samples is presented. The dielectric spectra of some prawn samples are shown in Fig. 19.1. On the left hand side is the Cole–Cole diagram while on the right hand side the curves of both the real and imaginary part (losses) of the permittivity are plotted vs. frequency. The selected samples are:

1. The untreated sample (solid line), liquid uptake: 0%.
2. Short treatment with weak solution (dashed line), liquid uptake: 5.5%.
3. Long treatment with pure water (dot-dashed line), liquid uptake: 10.9%.
4. Long treatment with strong solution (dotted line), liquid uptake: 11%.

The sample which was treated in pure water has lower losses at low frequencies in comparison to the untreated sample. This is explained by the fact that ions were diffused into the water. Hence the ionic conductivity decreases. With the highly treated sample (4) the opposite effect is observable: the losses increase at lower

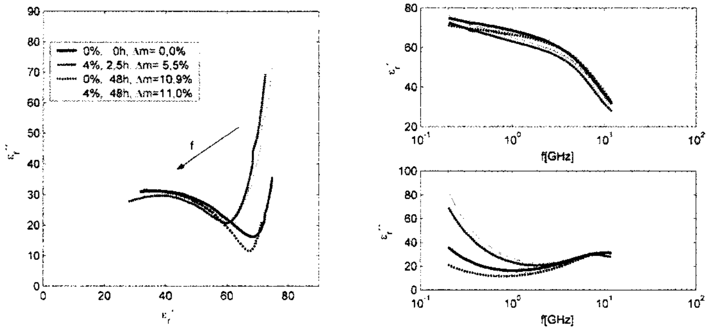

Fig. 19.1. Effects of the treatment of the samples on the dielectric spectra of prawns. The measurement temperature was 3°C

frequencies because ions diffuse from the solution into the tissue. With sample 2 it is observable that the polyphosphate accelerates the liquid uptake. With the real part of the permittivity it is difficult to detect any evident effects of the treatment.

However, the treatment has an influence on the shape of the dielectric spectrum. Signal processing for the determination of the composition of the samples has to extract the hidden information from the changes in the shape. Remarkably the samples which were treated for 48 h, (3) and (4), have nearly the same amount of liquid uptake although their dielectric spectra have a different shape. The signal processing has to tolerate such relatively great variations in the same prediction value.

19.4 Dielectric Modeling

Foods are generally complex stuffs. First of all flesh is inhomogeneous. It consists of different kinds of tissues like muscles, blood vessels, connective and fatty tissue. The main constituent of flesh is water. Part of it lies within the cells and part occurs in the extras cellular space. But flesh also consists of many other organic and inorganic substances, some of which have a complex interaction. A Parts of the water is bound by hydrophilic groups of organic molecules. The free water is rotationally hindered by other molecules. The dispersion of water lies within the range of microwave frequencies depending on its temperature. Smaller peptides and amino acids also have its relaxation frequencies in the microwave region [6].

The microwave dielectric spectrum is mainly formed by water. However, due to the complex interaction it is difficult to create a sufficiently accurate model for the dielectric behavior of flesh. Nevertheless, it should be possible to determine the existence of constituents which do not have its relaxation frequency in the measured frequency range – provided that they have an interaction with water. As mentioned above, salts increase the ionic conductivity. Hence at lower frequencies the losses also increase. In spite of this direct effect one cannot distinguish between the kinds of diluted salts.

The first method discussed here tries to apply a dielectric model of the flesh. The model parameters are taken from the measurements. A correlation between these parameters and the ingredients is used to determine the composition of the samples. The model applied in the following text consists only of three components: the free water (γ-dispersion), the bound water (δ-dispersion), and the salt content:

$$\varepsilon_r(\omega) = \varepsilon_\infty + \frac{\varepsilon_{s\gamma} - \varepsilon_\infty}{1 + (j\omega\tau_\gamma)^{(1-\alpha_\gamma)}} + \frac{\varepsilon_{s\delta} - \varepsilon_\infty}{1 + (j\omega\tau_\delta)^{(1-\alpha_\delta)}} - j\frac{\sigma}{\varepsilon_0\omega} \tag{19.1}$$

Where ε_∞ is the permittivity at infinite frequency, ε_s is the static permittivity, ω is the angular frequency, τ is the relaxation time, α describes the distribution of the relaxation time, and σ is the conductivity. Both the free and bound water are described by a Cole–Cole-relaxation and the salt content is considered in the third term. ε_∞ is fixed at the value of 4.6 for both dispersions..

The seven model parameters are calculated from the measured dielectric spectra using a non-linear curve fitting procedure. This procedure is described in [6] and based upon the Gauss–Newton method. The computational effort of the curve fitting is relatively high ($\approx 2\times10^6$ floating point operations for one curve fitting).

After the calculation of the model parameters, the samples were divided randomly, two-thirds into the calibration group, and one third into the validation group. The model parameters of the calibration group are used as regressors in a multiple linear regression (MLR). A composition value (z_c, e.g. water content) is described as the linear combination of the model parameters:

$$\vec{z}_c = A_c\vec{\beta} + \vec{e}_c\,. \tag{19.2}$$

The error vector is $\vec{e}_k$ while the regressor matrix is built in the following way:

$$A_c = \begin{bmatrix} 1 & \varepsilon_{s\gamma 1} & \tau_{\gamma 1} & \alpha_{\gamma 1} & \varepsilon_{s\delta 1} & \tau_{\delta 1} & \alpha_{\delta 1} & \sigma_1 & T_1 \\ & & & & \vdots & & & & \\ 1 & \varepsilon_{s\gamma t} & \tau_{\gamma t} & \alpha_{\gamma t} & \varepsilon_{s\delta t} & \tau_{\delta t} & \alpha_{\delta t} & \sigma_t & T_t \end{bmatrix}. \tag{19.3}$$

Each row consists of the model parameters of one sample of the calibration group. The number of samples in the calibration group is t. A least squares estimation for the vector of coefficients can be estimated by [7]

$$\hat{\vec{\beta}} = (A_c^T A_c)^{-1} A_c \vec{z}_c\,. \tag{19.4}$$

The composition value of the samples in the calibration and validation groups is estimated by

$$\hat{\vec{z}}_{c,v} = A_{c,v}\hat{\vec{\beta}}\,. \tag{19.5}$$

The calibration procedure described above has to be performed for each composition value[1]. The quality of the prediction is evaluated by the root mean square error[2]

$$RMSE_{c,v} = \sqrt{\frac{\vec{e}_{c,v}^{\,T} \cdot \vec{e}_{c,v}}{t}} \tag{19.6}$$

and the coefficient of determination R^2

$$R^2 = 1 - \frac{\vec{e}_c^{\,T} \cdot \vec{e}_c}{\sum_{i=1}^{t} (z_{c,i} - \bar{z}_c)^2}. \tag{19.7}$$

The results obtained from the dielectric modeling are shown in Table 19.1. By examining R^2 one can see that the performance of this method is disappointing. Only for the prediction of the salt content of the herring samples is $R^2 > 90\%$ obtained. With the prawn samples only the water and protein content is predicted with moderate accuracy.

Table 19.1. Results of the dielectric modelling procedure.

Prawns	R^2	$RMSE_c$	$RMSE_v$
Liquid uptake	0.745	2.75% of m_0	3.37 % of m_0
Fat	0.616	0.064 % abs.	0.077 % abs.
Protein	0.889	0.59 % abs.	0.73 % abs.
Water	0.852	0.61 % abs.	0.70 % abs.
Salt	0.771	0.047 % abs.	0.061 % abs.
Herring	R^2	$RMSE_c$	$RMSE_v$
Liquid uptake	0.634	6.50 % of m_0	5.89 of m_0
Fat	0.696	1.35 % abs.	1.67 % abs.
Protein	0.650	1.28 % abs.	1.06 % abs.
Water	0.623	2.19 % abs.	2.29 % abs.
Salt	0.955	0.125 % abs.	0.264 % abs.

---

[1] It is also possible to expand Eqs. 19.2, 19.3, and 19.5 to matrix equations. But for reasons of clarity this was not carried out here.

[2] The RMSE and R^2 for the method described later are determined in the same way.

Obviously the poor performance can be explained by the simplicity of the model used, which considers only three components. Now one could expand the model to improve the accuracy. But there are many reasons why this strategy does not succeed:

- Water has the most intense influence on the dielectric spectrum of flesh. All other polar molecules have only a minor influence on it.
- The dispersions of the diverse small protein molecules are widely spread[3]. Hence there is no clear relaxation frequency despite its distribution is considered with the parameter α.
- The interactions between the components are complex, possibly unknown as yet, and it is not possible to consider all of them.
- The measurement accuracy is not high enough to acquire all the details of the model. The dominant source of error with the open-ended coaxial line is the contact between the probe and the sample. Furthermore the material is not homogeneous enough.

For these reasons the determination of the model parameters of the bounded water is critical. In Fig. 19.2 the confidence intervals of all the model parameters of all the prawn samples are shown. They were calculated with a 95% level as described in [6]. This means the probability that the true model parameter is within the calculated confidence interval is 95%. In the figure one can see for some samples that the distributions of the relaxation times (α_γ and α_δ) are not determined reliably. Due to the numerical instability of the curve fitting procedure some of the determined relaxations times of the bounded water (τ_δ) samples are useless.

Another method to check the model parameters is to calculate the correlation between the them[4]. If the correlation is high, although they are physically independent this is a warning signal of the low reliability of the determination of the model parameters. For example, the correlation factor between the dielectric increment of the free water ($\Delta_\gamma = \varepsilon_{s\gamma} - \varepsilon_\infty$) and the distribution factor of relaxation time of the bounded water (α_δ) is unrealistic for one sample: $\rho = -0.99$. In comparison to other polar molecules the bounded water i.e. has nevertheless a determinable influence on the dielectric spectrum. But it is already difficult to measure its model parameters.

In short, it is not reasonable to expand the model with other components. Although the three-component model is primitive not all its parameters can be determined reliably. On the other hand, the use of a very primitive model can lead to a loss of information during the curve fitting procedure.

---

[3] Leucine: 56 MHz; glycine: 3.23 GHz [6].

[4] The calculation procedure is also described in [6].

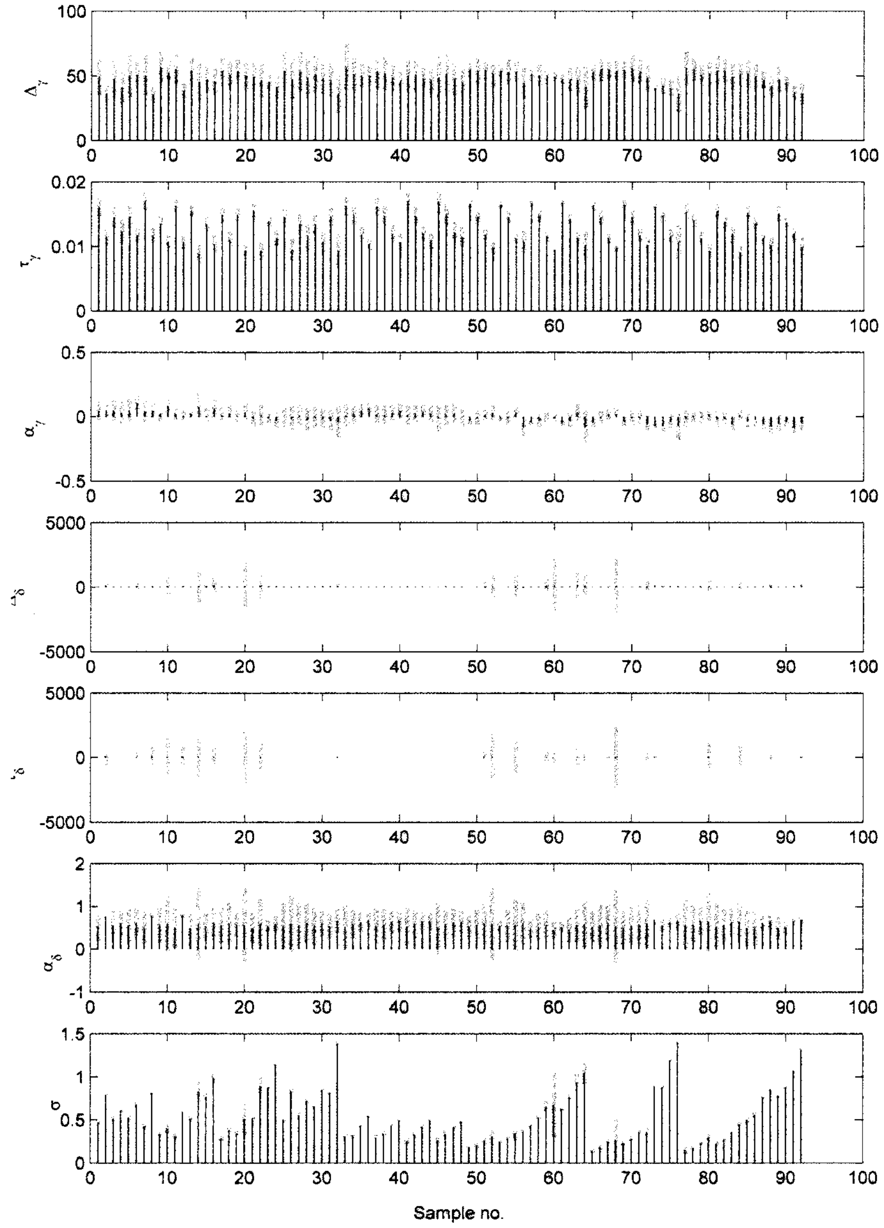

Fig. 19.2. Confidence intervals (95%) of all model parameters of all prawn samples. The determined parameters are shown as thin black lines while the confidence intervals are plotted as thick gray lines

19.5 Direct Processing of the Dielectric Data in a Multiple Linear Regression

The measurement problem is generalized in the following text. It is necessary to find a function that describes the relationship between the dielectric spectrum and the composition values:

$$z = f(\varepsilon(\omega,T)). \tag{19.8}$$

It is preferable to use the dielectric data directly as regressors in a MLR instead of the tediously determined, but inaccurate, model parameters. The procedure is nearly the same as that described above (Eqs. (19.2–19.5)). Only the matrix of regressors has to be changed. It is now built from the real and imaginary parts of the dielectric data and the temperature:

$$A_{k,v} = \begin{bmatrix} 1 & \varepsilon'_{r1(\omega_1)} & \cdots & \varepsilon'_{r1(\omega_n)} & \varepsilon''_{r1(\omega_1)} & \cdots & \varepsilon''_{r1(\omega_n)} & T_1 \\ & & & & \vdots & & & \\ 1 & \varepsilon'_{rt(\omega_1)} & \cdots & \varepsilon'_{rt(\omega_n)} & \varepsilon''_{rt(\omega_1)} & \cdots & \varepsilon''_{rt(\omega_n)} & T_t \end{bmatrix} \tag{19.9}$$

The calibration equation Eq. (19.5) estimates a composition value by a linear combination of the dielectric data and the temperature. The problem here is that the dielectric data are too collinear, i.e. the adjacent frequency points are relatively

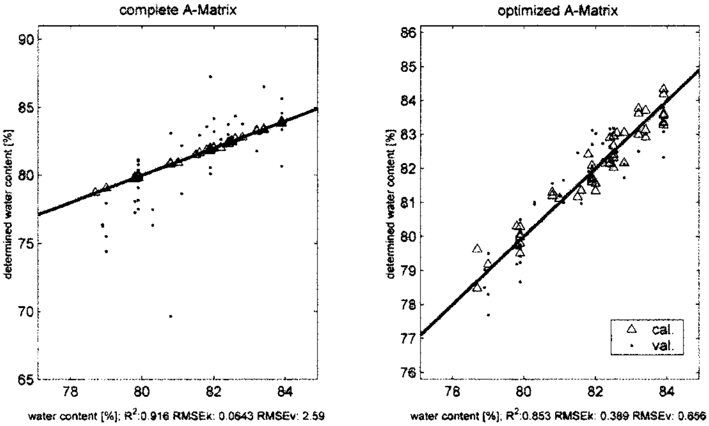

Fig. 19.3. Determination of the water content of prawns using the dielectric data as regressors in MLR. Left hand side: overfitting caused by collinear data. Right hand side: reduced matrix

highly correlated. For this reason the matrix inversion in Eq. (19.4) is numerically unstable [8]. In this case overfitting occurs. (Overfitting means that the calibration equation only treats the calibration group and the performance is bad for the validation group.) The effect of the collinearity is shown in Fig. 19.3.

A primitive method to reduce the collinearity is to thin out the matrix of regressors (19.9). That means only every second, third, etc. column is used. The optimum configuration is found when $RMSE_v$ has a minimum. This procedure is carried out for every composition value and the results are shown in Table 19.2. In comparison to the performance of the dielectric modeling a definite improvement results. However, the reduction of the collinearity has been taken arbitrarily. A more fashionable and precise method is used in the next section.

Table 19.2. Results obtained from the optimized regressor matrix composed of dielectric data

Prawns	R^2	$RMSE_c$	$RMSE_v$
Liquid uptake	0.827	2.26 % of m_0	3.18 % of m_0
Fat	0.394	0.080 % abs.	0.086 % abs.
Protein	0.910	0.53 % abs.	0.65 % abs.
Water	0.905	0.48 % abs.	0.58 % abs.
Salt	0.772	0.047 % abs.	0.052 % abs.
Herring	R^2	$RMSE_c$	$RMSE_v$
Liquid uptake	0.738	5.49 % of m_0	4.31 % of m_0
Fat	0.817	1.05% abs.	1.13 % abs.
Protein	0.586	1.39% abs.	0.73% abs.
Water	0.792	1.63% abs.	1.44% abs.
Salt	0.938	0.147% abs.	0.224% abs.

19.6 Elimination of the Collinearity using Partial Least Squares Regression

Kent et al [1, 9] proposed the use of principal components analysis (PCA) and principal components regression (PCR) for the processing of dielectric data. The principal components are an orthogonal transformation of the measured data. Hence the collinearity is removed completely. The criterion of the orthogonal transformation is to maximize the variance of the principal components, where the first one has the greatest variance, sorted in descending order. Normally the first few principal components have nearly the total variance of the data and the last one contains only the influence of noise. Hence PCA can be used for data reduction purposes. After the transformation some of the first components are used as regressors in MLR. One problem is to make a good selection for this purpose. If too many components are used, overfitting can also occur. Another problem is that the transformation is done completely independently of the composition values.

Partial least squares regression (PLSR) performs the calibration more directly. The composition values are considered during the orthogonalization. Originally PLSR was developed to process economic data, and it was more or less developed intuitionally. But Martens and Naes have used this technique also for near infrared spectroscopy [8]. In [10] PLSR is also used to determine the water content of wheat using microwave transmission measurements.

PLSR reduces the data to a set of data what is called "hidden path variables." The PLSR algorithm used here is called "PLS1" and is described in [8]. The structure of this method is as follows:

1. After subtracting the means the measured data are weighted in such a way that the covariance between them and the composition values is maximal.
2. A factor is defined as the projection of the data onto the vector of weights.
3. There then follows a regression analysis between the input variables and the composition values and the factor. The parts described by this linear regression are subtracted from the input data and composition value. These new data are used in the next iteration.
4. The algorithm is repeated until a specific number of factors are calculated.

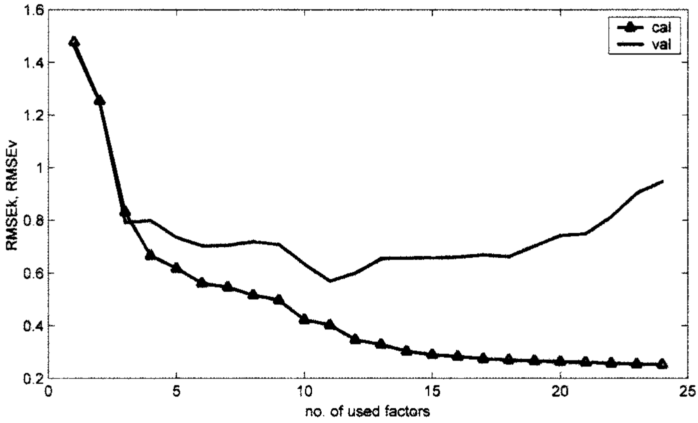

Fig. 19.4. Relationship between the RMSE of the calibration and validation groups for the determination of the water content of the prawn samples and the number of factors used in the PLSR.

Hence the only degree of freedom is the number of factors to be calculated. If too many are used overfitting occurs again. This effect is shown in Fig. 19.4. The RMSE of the calibration group decreases with the number of factors used. First, the RMSE of the validation group also decreases, and a minimum is reached at 12. But if more than 12 factors are used the RMSE_v increases. That means overfitting appears. The optimal number of factors has to be determined empirically for each composition value. However, in comparison to the more or less arbitrary deletion of columns of the regressor matrix described in the previous section, the determination of a minimum is much easier. This optimum can be determined more reliably the greater is the number of available data sets in the calibration and the validation groups.

The results obtained with PLSR are shown in Table 19.3. They are better than those obtained with the dielectric modeling but comparable to those of the dielectric data used directly as regressors described in the previous section. For the herring test series high values of R^2 are observable with PLSR, but this is only because of a higher overfitting.

Table 19.3. Comparison of the results obtained with PLSR

Prawns	R^2	RMSE_c	RMSE_v
Liquid uptake	0.744	2.76 % of m_0	3.27 % of m_0
Fat	0.436	0.078 % abs.	0.086 % abs.
Protein	0.936	0.45 % abs.	0.66 % abs.
Water	0.942	0.38 % abs.	0.57 % abs.
Salt	0.845	0.039 % abs.	0.052 % abs.
Herring	R^2	RMSE_c	RMSE_v
Liquid uptake	0.968	1.92 % of m_0	3.79 % of m_0
Fat	0.914	0.72 % abs.	1.19 % abs.
Protein	0.955	0.46 % abs.	0.80 % abs.
Water	0.981	0.49 % abs.	1.48 % abs.
Salt	0.945	0.139 % abs.	0.228 % abs.

19.7 Non-linear Data Processing Using Artificial Neural Networks

Apart from the curve fitting procedure of the dielectric modeling all calibration equations are a linear combination of the input variables. But the unknown function $z = f(\varepsilon(\omega, T))$ may be non-linear. One method to approximate an unknown function is the use of artificial neural networks (ANNs). Bartley et al used ANNs for the determination of the water content of wheat using free-space transmission measurements from 10 to 18 GHz [11]. Before such networks can be used successfully, a suitable architecture has to be chosen. Even if such a fundamental architecture is found many degrees of freedom still remain. For a functional approximation the use of multi-layer feed forward (MLFF) networks is recommended in the literature [12].

The architecture of the MLFF ANN used is shown in figure 19.5. The configuration of the displayed network proves very well suited for the application discussed here. It has one hidden layer which contains 5–10 neurons. The activation functions of the neurons in the hidden layer are non-linear (*tansig* function) while those of the output layer are linear.

It has been shown by Kreinovich and other authors that such types of ANNs are able to estimate unknown functions with a limited number of discontinuities at any desired accuracy [14].

If this architecture is chosen various degrees of freedom are left:

- The number of hidden layers.
- The number of neurons of the hidden layer.
- The kind of activation function of the neurons of each layer.

A more complex function to be approximated requires a greater number of neurons, hidden layers, and training data sets. Unfortunately the only guidelines

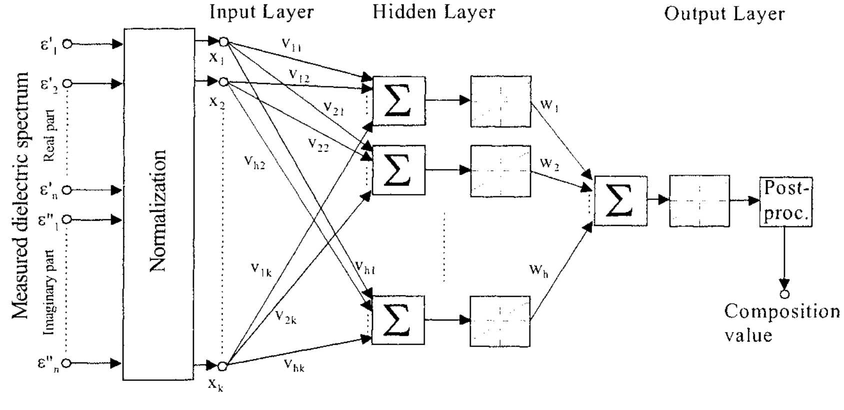

Fig. 19.5. Architecture of the MLFF network

known are to choose all these parameters. Therefore one has to evolve intuition in designing such a network by using trial and error.

19.7.1 Training of the Artificial Neural Network

After a fundamental architecture has been chosen the ANN has to be trained. That requires the determination of the weights (*w* and *v*) of the connections between the inputs and neurons in order to minimize the error of the function approximation. The ANNs discussed here are trained with back-propagation, which is a descent gradient algorithm. It uses a performance function which delivers the sum of the squared residuals as output values. The absolute minimum of this performance function has to be found by moving the network weights along the negative gradient of the performance function. The new calculated weights are used in the next iteration.

If, for example, five complex values and the temperature are used as input variables and a network consists of one hidden layer with eight neurons and one output variable is used, the performance function is a surface in a space with 88+1 dimensions. The surface may have a complex curvature which generally increases with the size of the network. Therefore the gradient algorithm may converge only to a local minimum or it may not converge at all.

Another problem is the requirement of random numbers to be used as starting values of the weights. Hence the success of a training exercise depends on those starting values. From that point of view this kind of ANN is not strictly deterministic. A further disadvantage is that ANNs have the tendency of overfitting. That means the network imprints only the training data and loses its required properties of generalization. The more complex the network, the higher is the risk of overfitting. To avoid overfitting the method of "Early Stopping" [13] was used. With this method the network output is observed if the validation set is used as the input. If its performance decreases while that of the calibration data output still increases the training is stopped.

The training was performed using the efficient Levenberg–Marquardt algorithm which is implemented within the "Neural Network Toolbox" of MATLAB. The different network parameters were varied in order to find an optimal configuration. Because of the sensitivity of the training with respect to the randomly selected starting values, the training was repeated 60 times in order to find a good pass.

19.7.2 Optimal Architecture Found

Using an oversized network is punished by a bad convergence of the training. For the application discussed here one hidden layer is sufficient. The number of neurons in the hidden layer should be between 5 and 10. If the number is below 5 the performance is not optimal. If it is increased above 10 the training becomes unstable again, the calculation effort rises, and the risk of overfitting increases. Quoting a range rather than a precise number is deliberate. The reason is that there is not a clear optimum. As mentioned above because of the random starting values of the weights the results also have variations. Apart from this, small changes in the number of neurons in the hidden layer have no dramatic effects on the results. Such a general robustness is also an appreciated quality of the ANN.

The use of a non-linear activation function is necessary if an MLFF network is required to approximate non-linear functions. Predominantly the *tansig* functions are used as non-linear activation functions [13].

The experiment results in the experience that a linear activation function is the best choice for the neurons of the output layer. If the *tansig* function is used for the output layer the convergence of training and performance is worse in comparison to the application of the linear activation function.

19.7.3 Results Obtained with ANN

The results obtained using the best found configuration are given in the diagrams of Fig. 19.6 (prawn), Fig. 19.7 (herring), and Table 19.4. On the abscissa of the diagrams the composition values determined by the reference methods are plotted while the predicted values by the ANN are shown on the ordinate. Hence a perfect prediction would hit the quality line.

With the prawn test series an improvement of the $\mathrm{RMSE_v}$ is observable. Especially with the estimated liquid uptake the accuracy is conspicuously higher. But with the herring test series the training of the ANN was not stable. Therefore the coefficient of determination R^2 is relatively poor. This is explainable by the low number of calibration data sets of this test series (herring: only 52; prawn: 92). In short the training of an MLFF ANN needs more effort and more calibration data sets. But the higher effort leads to a better prediction of the composition values.

Table 19.4. Results obtained with ANNs

Prawns	R^2	$\mathrm{RMSE_c}$	$\mathrm{RMSE_v}$
Liquid uptake	0.919	1.55 % of m_0	2.59 % of m_0
Fat	0.800	0.046 % abs.	0.076 % abs.
Protein	0.939	0.43 % abs.	0.53 % abs.
Water	0.933	0.41 % abs.	0.46 % abs.
Salt	0.792	0.045 % abs.	0.047 % abs.
Herring	R^2	$\mathrm{RMSE_c}$	$\mathrm{RMSE_v}$
Liquid uptake	0.503	6.45 % of m_0	5.04 % of m_0
Fat	0.304	1.61 % abs.	1.15 % abs.
Protein	0.366	1.42 % abs.	0.92 % abs.
Water	0.605	1.64 % abs.	1.62 % abs.

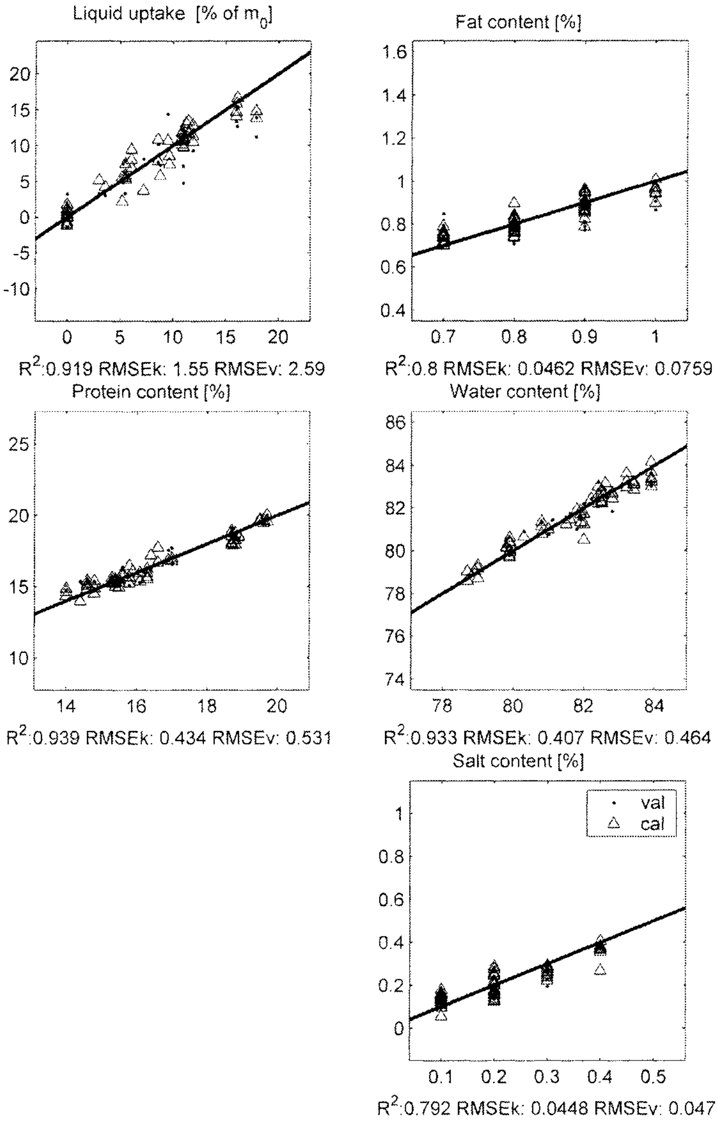

Fig. 19.6. Results obtained with ANNs for the prediction of the composition of prawn

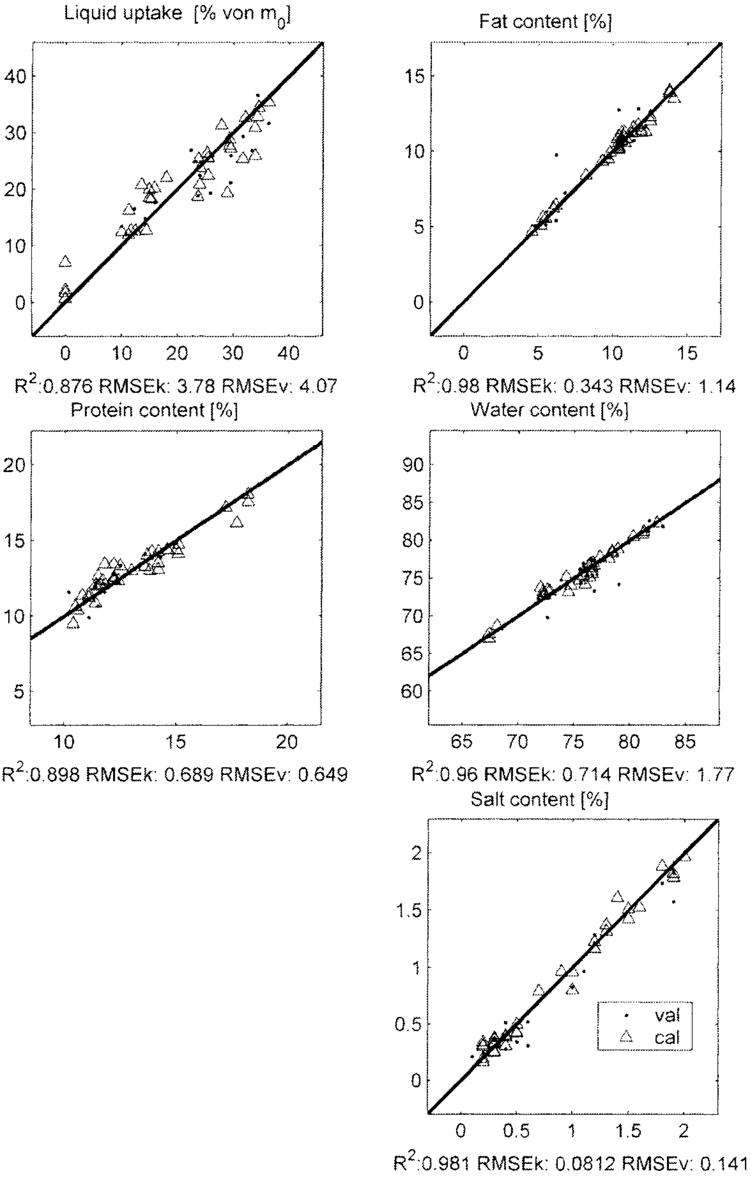

Fig. 19.7. Results obtained with ANNs for the prediction of the composition of herring

19.8 Comparison of Methods

For the comparison of methods for the prediction of the composition of foodstuffs using dielectric spectra three criteria are used:

1. Accuracy.
2. Calculation effort for calibration and validation.
3. Number of data sets needed for calibration and training.

In order to measure the calculation effort the *flops* function of MATLAB was used. In figure 19.8 the calculation effort for the complete calibration procedure of all methods is displayed versus the $\mathrm{RMSE_c}$. One can see that the dielectric modeling has the worst performance although the calculation effort is relatively high. This high effort is caused by the curve fitting procedure. The highest calculation effort has to be done during the training of the ANN. But here the effort pays because the performance has the best results. In comparison, the direct processing of the dielectric data in a MLR is relatively effortless.

However, for the practice it is more important to know how costly the prediction of the composition values of a sample is. While the calibration does not have to be done in real time, a realized system may be controlled by a relatively simple microcontroller which has less computational power. In Fig. 19.9 the calculation effort for the estimation of the water content of one prawn sample is plotted versus the $\mathrm{RMSE_V}$. Here the dielectric modeling also has the highest computational effort while the performance is relatively low. The performance of the direct processing of the dielectric data in a MLR and PLSR is similar. The effort of the MLR is lower. But with PLSR the orthogonalization proceeds more directly. The best results are taken with the ANN. The computational effort is higher but lower than with the dielectric modeling. With nearly 2,000 floating point operations it should be possible to use ANN in real time.

With all methods expect the ANN it is possible to make a successful calibration with fewer calibration data sets. If only a relatively small number of calibration data sets are available the internal cross-validation may be used [8]. But ANNs need a relatively high number of calibration data sets. However, the higher effort leads to better results.

19.8.1 General Evaluation of the Methods Discussed

The enforced experiments and discussed data processing methods showed that it is in principle possible to determine important composition values of foodstuffs using microwave dielectric spectra in combination with a multivariate processing of the measurement data. But because this method works in an indirect way two requirements have to be fulfilled:

1. The range of a composition value has to be wide enough.
2. The reference methods have to be accurate and should have a high resolution.

For example, these requirements are not given for the prediction of the fat content of the prawn samples. The range was only from 0.7% to 1.0% and the resolution of the reference method[5] was 0.1%. The results are accordingly moderate. This is the same case for the prediction of the salt content of prawns. The range was only 0.1% to 0.4% and the resolution of the reference method[6] was 0.1% too. But with the herring test series these requirements are fulfilled for the prediction of the fat and salt content. The fat content of the herring samples varied between 6% and 14.1%. Due to the treatment with different saline solutions the salt content varied between 0.1% and 2%. Hence the performance is comparatively better.

The quality of the reference methods has a key role for the methods discussed here. An indirect measurement method cannot be better than its reference method. But for complex items like food it is also difficult to make accurate measurements with the reference method. For example, the reference method for the prediction of the protein content may be disputed. Proteins are a whole class of constituents with multifaceted properties. The determination of the protein content by the reference method works with the measurement of the nitrogen content[7]. The protein content is then estimated using a more or less arbitrary factor (the nitrogen factor). The prediction of the water content is also critical[8]. Not only does water vaporize during the procedure, but so too do other volatile substances. The results obtained with ANN reach the accuracy of the reference method but they cannot better it.

The performance could be improved with another sensor. Because it is not easy to make a good contact between the open-ended coaxial line and the sample, and

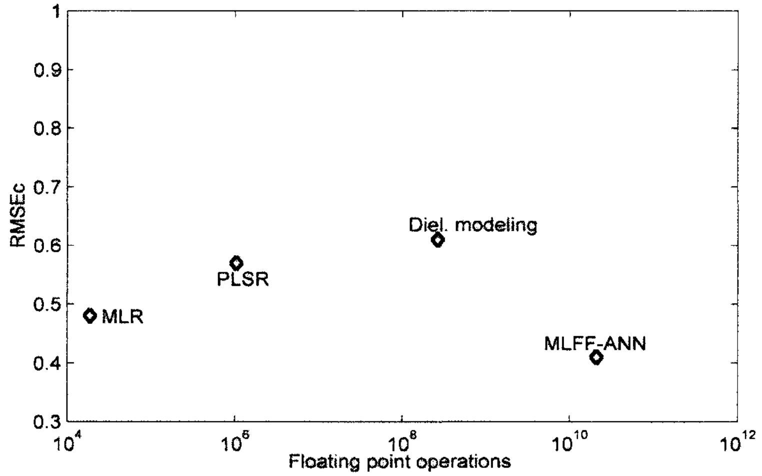

Fig. 19.8. Comparison of the complete calculation effort of the calibration (water content of prawn)

---

[5] Fat: NMR (SLV 195:9).

[6] Salt: conductometry.

[7] Protein: AOAC No. 999.15 and 990.03.

[8] Water: oven drying (104°C, 104 h).

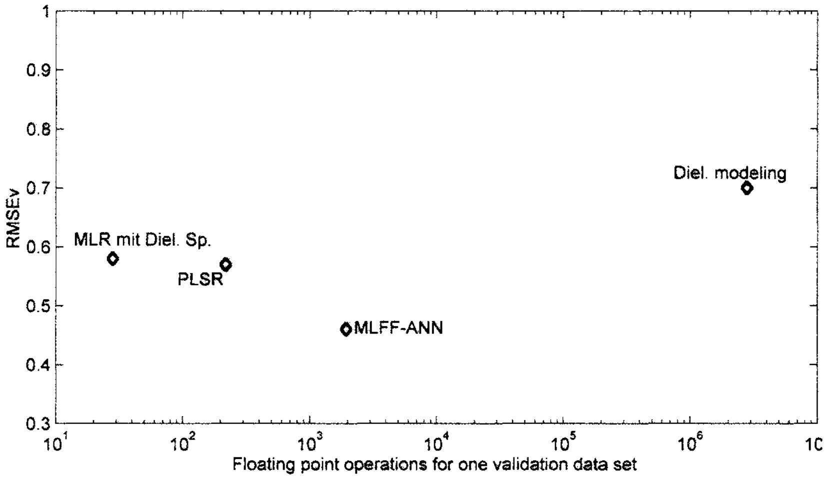

Fig. 19.9. Comparison of the calculation effort of the validation respectively the application of the methods for one dataset (water content of prawn).

because flesh is an inhomogeneous material, it is desirable to have a sensor with a higher sensitive area. The multivariate methods do not necessarily need the permittivity as an input variable because various other arrangements are possible as sensors. Because the measured *S*-parameters could be used as input variables directly, there is no need to create a model of the arrangement. It is only important that the sensor is unambiguously sensitive enough to changes in the composition.

Another aspect is the question of how far the prediction of the protein and salt content is independent of the prediction of the water content. The treatment of the samples leads to the addition of water. This means that the other components are diluted. Water has a dominant influence on the microwave dielectric spectrum of flesh. Hence there is a risk that the protein and fat content is only apparently predicted by the use of the negative correlation between the water content and the fat and protein content. Furthermore, there is a complex interaction between the water and the other constituents as mentioned above. This complex interaction could make it possible to detect other components if they modify the dielectric behavior of the water.

In Fig. 19.10 the correlation between water and protein and between water and fat is shown. The correlation factors are $\rho = -0.75$ and $\rho = -0.9$ approximately. In order to test how far the prediction of the protein and fat content depends on the negative correlation they were determined using the predicted water content (PLSR) and a linear regression only. In this case the coefficient of correlation decreases to $R^2 = 0.64$ (protein) and $R^2 = 0.53$ (fat). That means not only does the negative correlation deliver information about the fraction of protein and fat, but the hidden information in the dielectric spectrum is also effective.

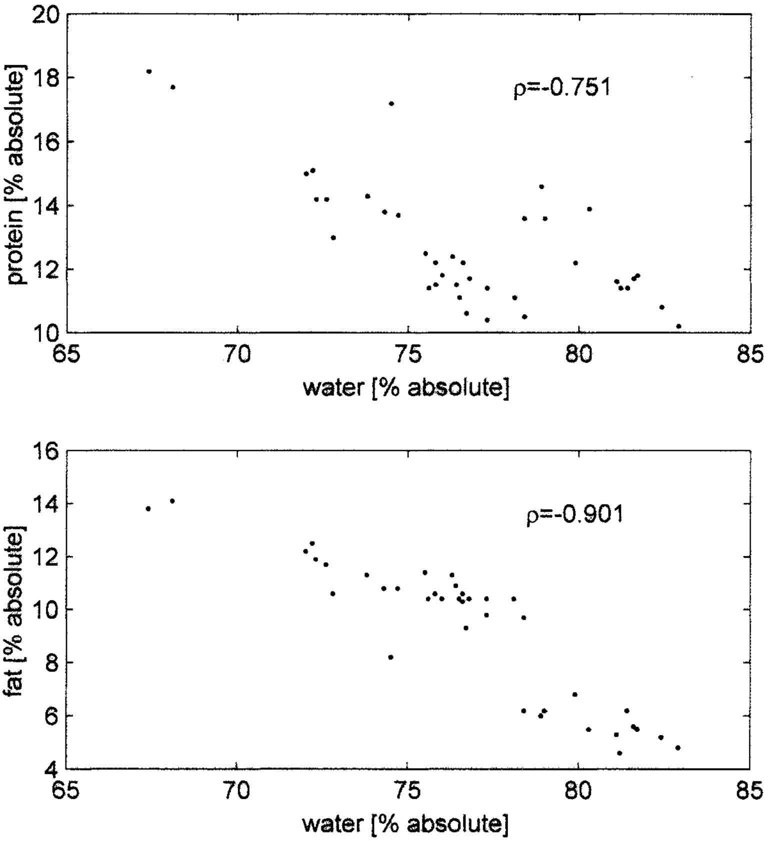

Fig. 19.10. Correlation between the water and the protein and fat content (herring).

Because the salt content causes higher losses at lower frequencies it has a direct and strong influence on the dielectric spectra. Hence the problem discussed above does not exist with the prediction of the salt content.

19.9 Conclusions

The determination of the composition of foodstuffs using conventional methods is time consuming, alters or destroys the samples, and needs many expensive devices. Therefore these methods are not applicable on the whole for the control of production processes. The approach suggested here is based on the influence of the ingredients on the microwave dielectric spectrum. The samples of two test series were diluted with water and their dielectric spectra were measured in the frequency range between 200 MHz and 12 GHz using an open-ended coaxial line.

A qualitative view on the measurements data showed that the dielectric spectra change with the amount of added water. Several methods were investigated to extract the hidden information about the composition of the samples from the dielectric spectrum.

The first method used a simple dielectric model. However, the performance of this method was not satisfactory and the computational effort was relatively high. Statistical analyses result in the perception that it is not reasonable to expand the dielectric model because the measurement accuracy is not good enough to detect small details of a more complex model. It was easier and more accurate to process the measured dielectric data directly with multiple linear regression (MLR). Due to the fact that the data have a high collinearity there is the problem that numerical instability occurs.

This problem can be avoided by the use of principal components analysis (PCA) or partial least squares regression (PLSR). If it is assumed that the unknown function which describes the relation between the permittivity and the composition of the samples might be non-linear the use of multi-layer feed forward artificial neural networks (MLFF ANNs) can be used to approximate it. The computational effort during calibration and training is relatively high and more calibration data are needed. But the best results were obtained with this method. If the number of calibration data sets is below approximately 50 the use of PLSR should be preferred.

It was also shown that besides water other important ingredients like protein, fat, and salt can be determined. In fact the reference methods cannot be exceeded. But all constituents can be measured in one step and in a short time by the multivariate processing of the microwave dielectric spectra.

Acknowledgments

The contributions of and many discussions with Dr. Mike Kent are greatly appreciated. This study has been carried out with financial support from the Commission of the European Communities, Agriculture and Fisheries (FAIR) specific RTD program FAIR CT97-3020 "Added Water in Foods". This study does not necessarily reflect the Commission's views and in no way anticipates its future policy in this area.

References

1. Kent M, Anderson D (1996) Dielectric studies of added water in poultry meat and scallops, J Food Eng 28: 239–259
2. Kent M (1999) Simultaneous determination of composition and other material properties by using microwave sensors. In: Sensors update, Wiley-VCH, Weinheim, vol 7, pp 4–25

3. Gajda G, Stuchly S (1983) An equivalent circuit of an open-ended coaxial line. IEEE Trans Instrum Meas IM-32(4): 506–508
4. Stuchly M, Brady M, Stuchly S, Gajda G (1982) Equivalent circuit of an open-ended coaxial Line in a lossy dielectric. IEEE Trans Instrum Meas IM-31(2): 116–119
5. Marsland TP, Evans S (1987) Dielectric measurement with an open-ended coaxial probe, IEE Proc 134(4): 341–349
6 Grant EH, Sheppard RJ, South GP (1978) Dielectric behaviour of biological molecules in solution. Oxford University Press, Oxford
7. Hartung J, Epelt B (1995) Multivariate Statistik, Lehr- und Handbuch der angewandten Statistik, 5. Auflage, R. Oldenbourg Verlag, Munich
8. Martens H, Naes T (1989): Multivariate Calibration. Wiley, Chichester
9. Kent M, Knöchel R, Daschner F, Berger UK (2000): Composition of foods using microwave dielectric spectra. Eur Food Res Technol no 210: 359–366
10. Archibald DD, Trabelsi S, Kraszewski AW, Nelson SO (1998): Regression analysis of microwave spectra for temperature-compensated and density-independent determination of wheat moisture content. Appl Spectrosc 52(11): 1435–1446
11. Bartley PG, Nelson SO, McClendon RW, Trabelsi S (1999): Determination of moisture content in wheat using an artificial neural network. In: 3rd workshop on electromagnetic interaction with water and moist substances, Athens, GA, pp 74–78
12. Patterson D (997) Künstliche neuronale Netze, Prentince Hall Verlag, Haar
13. Neural Network Toolbox, User's Guide, Version 4, The Mathworks Inc., 2000
14 Kreinovich VY(1991): Arbitrary nonlinearity is sufficient to represent all functions by neural networks: a theorem. Neural Networks 4: 381–383