

# MODELLING OF CHEMICAL REACTIONS IN FOODS: A MULTIRESPONSE APPROACH

Van Boekel, M.A.J.S.

Food Science Group, Wageningen Agricultural University,  
P.O. Box 8129, 6700 EV Wageningen, the Netherlands  
Fax: +31-317-483669, E-mail: Tiny.vanBoekel@IFT.FdSci.wau.nl

**Keywords:** Modelling, multiresponse, kinetics, chlorophyll, quality

## Introduction

The quality of foods depends on several factors. One of these factors is the occurrence of (bio)chemical changes taking place during the post-harvest period and during processing, storage and distribution. In order to optimise quality it is of utmost importance to control (bio)chemical changes as much as possible. That means that the kinetics of relevant chemical changes needs to be studied, i.e. the dependence of reaction rates on concentration of reactants and conditions such as pH and temperature. A complicating factor with foods is that reactions are not simple straightforward chemical reactions. Rather, consecutive and parallel reactions take place, each of which may have an effect on the other. For instance, if in one reaction organic acids are formed as a result of which the pH drops, this decrease in pH may have an effect on another reaction taking place simultaneously. Whenever possible, one should try to model all changes in concentrations of reactants and products in parallel and consecutive reactions that have parameters in common, i.e. multiresponse modelling, as opposed to analysing only one step at a time, i.e. uniresponse modelling (van Boekel, 1996). By applying multiresponse modelling, more realistic models and more accurate parameter estimates will be obtained and this ultimately means better control of (chemical) quality. Some implications of multiresponse modelling will be discussed here and, as an example, the concept will be applied to chlorophyll degradation in vegetables.

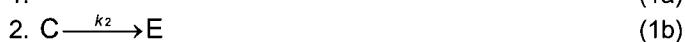
## Multiresponse modelling

Regression models with more than one response variable can be divided in two groups (Kang and Bates, 1990):

1. multivariate linear regression models, which have the same linear functional relationship with independent variables but with different coefficients
2. multiresponse regression models, in which the dependent variables have different (linear or nonlinear) functional relations with the independent variables.

Mechanistic models for chemical reactions in foods are usually specified as nonlinear forms in the parameters, and then multiresponse regression models must be used.

An important aspect for multiresponse modelling is to take variances and covariances of the various responses into account. To clarify this point, a hypothetical reaction scheme is given (later on, a real example will be discussed). Suppose three reactions take place at the same time:



with  $k_i$  as reaction rate constants. Then the following differential equations can be set up:

$$-\frac{d[A]}{dt} = k_1 \cdot [A] \cdot [B] \quad (2a)$$

$$-\frac{d[B]}{dt} = k_1 \cdot [A] \cdot [B] + k_3 \cdot [D] \cdot [B] \quad (2b)$$

$$-\frac{d[C]}{dt} = -k_1 \cdot [A] \cdot [B] + k_2 \cdot [C] \quad (2c)$$

$$-\frac{d[D]}{dt} = -k_1 \cdot [A] \cdot [B] + k_3 \cdot [D] \cdot [B] \quad (2d)$$

$$\frac{d[E]}{dt} = k_2 \cdot [C] \quad (2e)$$

$$\frac{d[F]}{dt} = k_3 \cdot [D] \cdot [B] \quad (2f)$$

These coupled ordinary differential equations (ODE's) can be solved by numerical integration. A well suited algorithm is, for instance, the Gear routine (e.g., Stabler & Chesick, 1978), especially designed for so-called stiff differential equations (in which the parameters may have largely different values, which is frequently the case for kinetic rate constants).

Next, the model (i.e. the integrated rate equations) should be fitted to the experimental data points. The 'natural' procedure for this would seem to be the method of least squares, for instance to minimise for component A:

$$\sum_{u=1}^n (y_A - \eta_A)^2 \quad (3)$$

in which  $u$  (1.. $n$ ) is the number of experimental runs,  $y_A$  the experimental datum points for component A, and  $\eta_A$  the predictions of component A by the model. In the above example, there are several responses at the same time (the concentrations of components A,B,C,D,E,F at each time interval studied). The question is now whether the best fit criterion in the above example is simply to minimise the combined sum of squares for all responses (like for component A in equation 3). It is not always realised that there are several, rather strict, requirements for application of least squares. It turns out that for cases of multiresponse modelling the fit criterion to be used depends on the experimental error structure: the covariance matrix of the experimental errors,  $\Sigma$ , is of importance:

$$\Sigma = \begin{bmatrix} \sigma_{AA} & \sigma_{AB} & \sigma_{AC} & \sigma_{AD} & \sigma_{AE} & \sigma_{AF} \\ \sigma_{BA} & \sigma_{BB} & \sigma_{BC} & \sigma_{BD} & \sigma_{BE} & \sigma_{BF} \\ \sigma_{CA} & \sigma_{CB} & \sigma_{CC} & \sigma_{CD} & \sigma_{CE} & \sigma_{CF} \\ \sigma_{DA} & \sigma_{DB} & \sigma_{DC} & \sigma_{DD} & \sigma_{DE} & \sigma_{DF} \\ \sigma_{EA} & \sigma_{EB} & \sigma_{EC} & \sigma_{ED} & \sigma_{EE} & \sigma_{EF} \\ \sigma_{FA} & \sigma_{FB} & \sigma_{FC} & \sigma_{FD} & \sigma_{FE} & \sigma_{FF} \end{bmatrix} \quad (4)$$

The diagonal elements in the matrix  $\Sigma$  represent the variances of each response (i.e.,  $\sigma_{AA} = \sigma_A^2$ ) and the off-diagonal elements the covariances (i.e.,  $\sigma_{AB} = \rho \sigma_A \sigma_B$  with  $\rho$  the correlation coefficient). In most cases the covariance matrix will be unknown, with the covariances  $\neq 0$  within a run; then the best-fit criterion is not least squares minimisation, but minimisation

of the determinant of the so-called dispersion matrix **C** with elements (Box & Draper, 1965, Hunter, 1967):

$$c_i = \sum_{u=1}^n [y_u^i - \eta_u^i] \cdot [y_u^j - \eta_u^j] \quad (5)$$

in which  $i, j$  is the index of responses ( $i, j=1..r$ ) and  $u$  the index of experimental runs ( $u=1..n$ ). The diagonal elements of matrix **C** correspond to the sum of squares for each of the responses.

If the covariance matrix, equation (5), happens to be known the best-fit criterion is minimisation of (Hunter, 1967):

$$\sum_{i=1}^r \sum_{j=1}^r \sigma^{ij} c_{ij} \quad (6)$$

in which  $\sigma^{ij}$  are the elements of the inverse of the matrix  $\Sigma$ . If, in addition, no correlation exists between responses ( $\sigma^{ij} = 0$  for  $i \neq j$ ) and the variances of the responses are known, minimisation of the following is appropriate (Hunter, 1967):

$$\sum_{i=1}^r \sigma^{ii} \sum_{u=1}^n [y_u^i - \eta_u^i]^2 \quad (7)$$

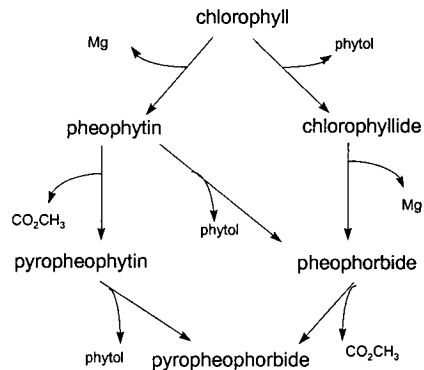
Equation (7) represents the case of weighted least squares. Finally, if  $\sigma^{ii}$  is equal for all responses the minimisation criterion is:

$$\sum_{i=1}^r \sum_{u=1}^n [y_u^i - \eta_u^i]^2 \quad (8)$$

and this is actually the least squares criterion for all responses, analogous to equation (3) for one component. Coming back now to the question whether or not least squares is the best fit criterion, the answer appears to be: only under the (rather strict) conditions that all variances are the same and that covariance's within a run are zero. This is a situation not normally encountered in practice if several responses are measured at the same time. All in all, it turns out that in multiresponse modelling the determinant criterion, i.e., minimisation of the determinant of matrix equation (5), is the best: the multivariate error distribution can be estimated concurrently, whereas it must be prescribed when least squares methods are used (Stewart and Sørensen, 1981). In other words, the method of least squares is not well suited for dealing with multiresponse data.

## Example of multiresponse modelling: Chlorophyll degradation kinetics

Chlorophyll is the principal pigment in green plants and it is easily degraded during processing, resulting in colour changes of the food, hence in quality changes (usually quality loss). The cause of degradation can be enzymic breakdown, pH change, heating or a combination thereof. Examples in which these changes occur are coleslaw, pickles, olives and green peas. A general scheme for chlorophyll degradation is shown in Figure 1 (after Heaton and Marangoni, 1996). In earlier literature, attempts have been made to model chlorophyll degradation as a first order reaction. However, besides chlorophyll concentration itself, the

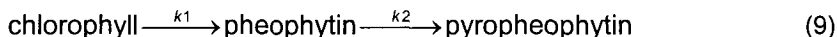


**Figure 1** Degradation scheme of chlorophyll degradation

intermediates shown in Figure 1 can be determined as well and this offers the possibility to apply multiresponse modelling to this problem, as shown below for two cases.

### Chlorophyll degradation in heat processed spinach

Schwartz and von Elbe (1983) studied the degradation of chlorophyll (a as well as b) in heat processed spinach and they suggested the following kinetic scheme (cf. Figure 1):



The authors used the following approach: first they determined  $k_1$  via linear regression by plotting the logarithm of remaining chlorophyll versus time (thus assuming a first-order reaction) and then they used nonlinear regression to fit the model to pheophytin data to obtain  $k_2$  (also assuming a first order reaction) via the equation:

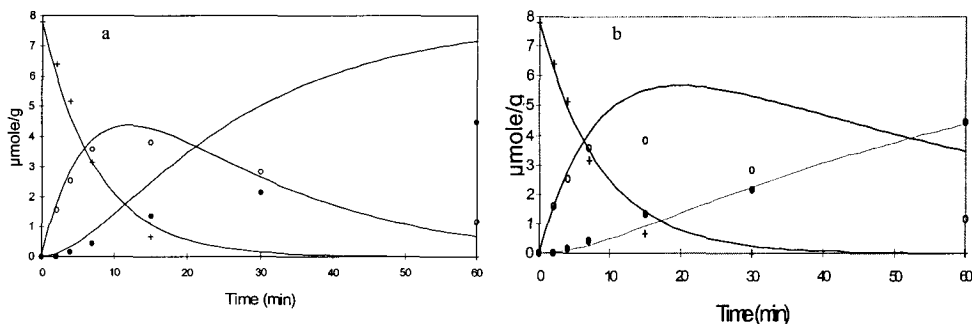
$$[\text{pheophytin}] = [\text{chlorophyll}]_0 \frac{k_1}{k_1 - k_2} (e^{-k_2 t} - e^{-k_1 t}) \quad (10)$$

The data reported by Schwartz and von Elbe (1983) for chlorophyll a were used to apply the multiresponse case for the model depicted in equation (9). For the sake of comparison, the two step approach of Schwartz and von Elbe (1983) was also followed via linear regression as well as nonlinear regression. The software used for multiresponse fitting was the FORTRAN subroutine GREG while the numerical integration of rate equations (ODE's) was done using the FORTRAN subroutine DDASAC (Stewart *et al.*, 1992). The results are in Table 1. The plots are in Figure 2a and 2b for results using the two-step nonlinear regression method and the multiresponse method, respectively.

Several interesting phenomena are apparent from Table 1. First, procedure 1 gave different results than procedure 2 for  $k_1$ . This is most likely due to the linearisation process

**Table 1** Rate constants  $\times 10^{-3} \text{ (s}^{-1}) \pm 95\%$  confidence intervals for chlorophyll degradation in heat processed spinach; data taken from Schwartz & von Elbe (1983) for heating at 121 °C.

Rate constant	procedure 1: linear regression for $k_1$ , nonlinear regression for $k_2$	Procedure2: Nonlinear regression in 2 steps	procedure 3: multiresponse modelling
$k_1$	$2.80 \pm 0.61$	$2.19 \pm 0.72$	$1.91 \pm 0.25$
$k_2$	$0.86 \pm 0.36$	$0.81 \pm 0.18$	$0.27 \pm 0.02$



**Figure 2** Fit of the model in equation (10) using parameters obtained via procedure 2 (a) and procedure 3 (b). + chlorophyll, O pheophytin, • pyropheophytin

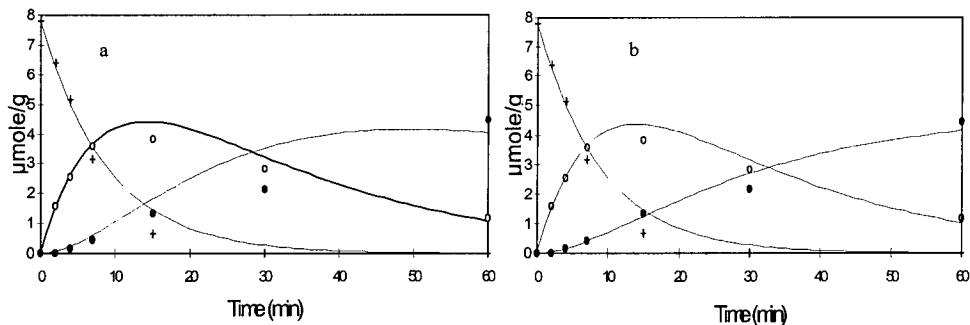
in procedure 1 (logarithmic transformation) which disturbs the error distribution and therefore violates the conditions for regression (van Boekel, 1996). Clearly, non-linear regression is to be preferred. Secondly, multiresponse modelling (procedure 3) gave quite different values than those obtained in procedure 2.

As can be seen in Figure 2a, the 2-step method (procedure 2) results in a reasonable fit for chlorophyll and pheophytin, but strongly overestimates the formation of pyropheophytin, obviously because this procedure does not take the formation of pyropheophytin into account. The multiresponse method on the other hand overestimates the formation of pheophytin at heating times > 10 min (Figure 2b). This lack of fit signals a failure of the model in equation (9). Apparently, further breakdown of the intermediates occurred. This could be breakdown of either pheophytin or pyropheophytin (or perhaps even both). Multiresponse modelling of breakdown of pyropheophytin was less successful (results in Figure 3a, the fit for pyropheophytin was not good). However, multiresponse modelling of additional breakdown of pheophytin gave a very good fit (Figure 3b). This suggests that breakdown of pheophytin (probably into pheophorbide, see Figure 1) occurred rather than that of pyropheophytin. Of course, this hypothesis should be verified by additional experiments, but it shows nicely that kinetic modelling can help in finding realistic reaction mechanisms.

A third point to note from Table 1 is that multiresponse modelling gives much more precise results: the 95% confidence intervals are much narrower than those obtained from the other procedures. This point, important as it is, is much overlooked in kinetic modelling; the precision of the estimated parameters obtained is of utmost importance if one wants to use the estimated rate constants to predict chemical reactions in, and hence quality of, foods.

### Chlorophyll degradation in olives

Another example of chlorophyll degradation was taken from results reported by Minguez-Mosquera *et al.* (1989) for brined olives. A kinetic model for this degradation was recently proposed by Heaton *et al.* (1996), (and further extended by Marangoni (1996) for degradation of pheophorbide into colourless compounds): see Figure 4. Heaton *et al.* (1996) estimated rate constants for the first three steps by solving sequentially the analytically obtained rate equations. The same data were now analysed via the multiresponse approach. An additional remark to be made is that Heaton *et al.* (1996) used mol% in their modelling; however, that creates dependencies in the data because



**Figure 3** Fit of the model in equation (10) with additional breakdown of pyropheophytin (a) or pheophytin (b). + chlorophyll, O pheophytin, • pyropheophytin.

the sum of all responses must then be 100%. This causes numerical problems when using the determinant criterion in multiresponse modelling and ignoring such

dependencies can lead to serious mistakes in interpretation, as shown by Box *et al.* (1973). In general, each response should be determined independently because independent information on each of the responses provides better estimates of the parameters, and it allows a more comprehensive check on the validity of the model (by forcing the data to obey relationships, such as mol%, one forces the data to satisfy theoretical relationships believed to be true). Methods are available to detect dependencies in the data (Box *et al.*, 1973, McLean *et al.*, 1979); the subroutine GREG used here detects such dependencies. For the problem of chlorophyll degradation in olives, raw data were thus used rather than mol%. The results are in Table 2 and Figure 5.

constants differ from the ones obtained by Heaton *et al.* (1996), while the 95% confidence intervals are, again, much better using the multiresponse approach. However, as shown in Figure 5, the fit is reasonable but certainly not perfect, especially not for chlorophyllide. This may be an indication that the kinetic model of Figure 4 needs adjustment. Marangoni (1996) suggested an additional breakdown of pheophorbide into colourless compounds (see Figure 4). This extra reaction step was introduced into the model via rate constant  $k_4$  and fit to the data via multiresponse modelling; results are also displayed in Table 2. The fit for chlorophyllide was marginally improved (results not shown), but the confidence interval for  $k_4$  is quite large compared to the actual parameter value (Table 2), which indicates that this parameter may actually be redundant. The model may therefore need another adjustment. Introduction of a reaction pathway from pheophytin to pheophorbide (cf. Figure 1) did not help: the reaction rate constant obtained was virtually zero and hence redundant.

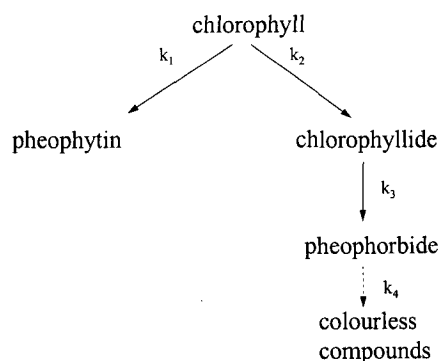


Figure 4 Kinetic model for degradation of chlorophyll in olives

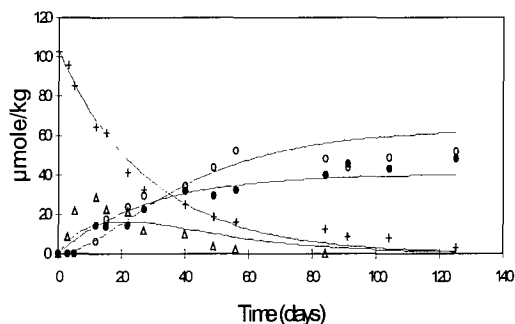


Figure 5 Fit of the model in Figure 4 for breakdown of chlorophyll in olives. +, chlorophyll, ● pheophytin, Δ chlorophyllide, O pheophorbide

**Table 2.** Rate constants ( $\text{day}^{-1}$ )  $\pm$  95% confidence intervals for chlorophyll degradation in brined olives according to the kinetic model of Heaton *et al.* (1996); data were taken from Minguez-Mosquera *et al.* (1989).

Rate constant	Values derived by Heaton <i>et al.</i> (1996)	Values derived by multiresponse modelling for kinetic scheme of Heaton <i>et al.</i> (1996)	values derived by multiresponse modelling for kinetic scheme of Marangoni (1996)
$k_1$	$0.023 \pm 0.011$	$0.014 \pm 0.001$	$0.014 \pm 0.001$
$k_2$	$0.033 \pm 0.028$	$0.021 \pm 0.002$	$0.021 \pm 0.003$
$k_3$	$0.056 \pm 0.043$	$0.065 \pm 0.015$	$0.082 \pm 0.027$
$K_4$	not estimated	not estimated	$0.002 \pm 0.001$

## Conclusion

In conclusion, it may be stated that multiresponse modelling of chemical reactions in foods opens new possibilities:

- it allows rigorous testing of proposed models: lack of fit is much more apparent than when using a uniresponse model. The examples shown indicate that some proposed models for chlorophyll breakdown in vegetables may need adjustment.
- it gives much more precise estimates of kinetic parameters. This is due to use of all the information that is available in the data. This in turn makes it possible to predict quality changes in foods much more reliable.
- it is not necessary to specify the covariance matrix of the responses when using the determinant criterion, whereas this *is* necessary for least squares fitting.

## References

- Box G.E.P. and Draper N.R. (1965): The Bayesian estimation of common parameters of several responses. *Biometrika* **52**, 355-365
- Box G.E.P., Hunter W.G., MacGregor J.F. and Erjavec J. (1973): Some problems associated with the analysis of multiresponse data. *Technometrics* **15**, 33-51
- Heaton J.W. and Marangoni A.G. (1996): Chlorophyll degradation in processed foods and senescent plant tissues. *Trends in Food Science & Technology* **7**, 8-15
- Heaton J.W., Lencki R.W., Marangoni A.G. (1996): Kinetic model for chlorophyll degradation in green tissue. *Journal of Agricultural and Food Chemistry* **44**, 399-402
- Hunter W.G. (1967): Estimation of unknown constants from multiresponse data. *Industrial Engineering Chemistry Fundamentals* **6**, 461-463
- Kang G. and Bates D.M. (1990): Approximate inferences in multiresponse regression analysis. *Biometrika* **77**, 321-331
- Marangoni A.G. (1996): Kinetic model for chlorophyll degradation in green tissue. 2. Pheophorbide degradation to colorless compounds. *Journal of Agricultural and Food Chemistry* **44**, 3735-3740
- McLean D.D., Pritchard D.J., Bacon D.W. and Downie J. (1979): Singularities in multiresponse modelling. *Technometrics* **21**, 291-298.
- Minguez-Mosquera M.I., Garrido-Fernández J. and Gandul-Rojas B. (1989): Pigment changes in olives during fermentation and brine storage. *Journal of Agricultural and Food Chemistry* **37**, 8-11
- Schwartz S.J. and Von Elbe J.H. (1983): Kinetics of chlorophyll degradation to pyropheophytin in vegetables. *Journal of Food Science* **48**, 1303-1306
- Stabler R.N. and Chesick J.P. (1978): A program system for computer integration of multistep reaction rate equations using the Gear integration method. *International Journal of Chemical Kinetics* **10**, 461-469
- Stewart W.E. and Sørensen J.P. (1981): Bayesian estimation of common parameters from multiresponse data with missing observations. *Technometrics* **23**, 131-141
- Stewart W.E., Caracotsios M. and Sørensen J.P. (1992): Parameter estimation from multiresponse data. *American Institute of Chemical Engineers Journal* **38**, 641-650
- Van Boekel M.A.J.S. (1996): Statistical aspects of kinetic modeling for food science problems. *Journal of Food Science* **61**, 477-485 & 489