2

Bayesian solutions for food-science problems?

M.A.J.S. van Boekel[#]

Abstract

This paper starts with an overview of some typical food-science problems. In view of the development of safe and healthy food, the use of mathematical models in food science is much needed and the use of statistics is therefore indispensable. Because of the biological variability in the raw materials on the one hand and the complex nature of foods on the other hand food-science problems are characterized by a high degree of uncertainty as well as variability. Consequently, when dealing with these problems Bayesian statistics could be very helpful; however, it is hardly used at all. This paper discusses some possible applications concerning the modelling of food quality and food safety. It is concluded that a Bayesian approach could be quite useful and its potential should be further explored in future research.
Keywords: Food quality; kinetics; modelling; Bayesian statistics

Introduction

Food science is integrating chemical, biochemical, physical, microbiological, physiological and social knowledge in order to control and predict properties of foods as they are influenced by raw materials, ingredients, storage conditions and processing. Foods are in fact very complex inhomogeneous materials in which all kinds of interactions take place. From an experimental point of view they are sometimes considered as ‘dirty materials’ and they are therefore often simplified by using model systems, e.g., a well-defined protein solution instead of milk. It should therefore come as no surprise that many things related to foods are uncertain. On the other hand, the consumer and society ask for safe, healthy and tasty food and they do not appreciate that food scientists are unable to control everything. Therefore, we are in need of models that can help the food scientist to predict and control food quality. Bayesian statistics could be a helpful tool to build such models. However, Bayesian statistics are by and large unknown in the food-science world, and there is much to be learned. It is the purpose of this paper to give an overview of the main problems in food science and to indicate where Bayesian statistics could be helpful. To be sure, this paper does not give a solution in terms of Bayesian approaches, nor does it pretend that Bayesian solutions give the best answer. The main reason for this paper is that the present author believes that statistics should be used to a much larger extent, as indicated before (Van Boekel 1996a; 1996b) and that a Bayesian approach in food science is lacking until now.

[#] Product Design and Quality Management Group, Department of Agrotechnology and Food Sciences Wageningen University, P.O. Box 8129, 6700 EV Wageningen, The Netherlands. E-mail: Tiny.vanBoekel@wur.nl

Food-quality modelling

The most important aspect of food science is that it provides scientific means to discuss quality of foods. Quality is an elusive concept, and a very general definition of quality is 'to satisfy the expectation of a consumer'. This does not only pertain to intrinsic physical product properties but also to extrinsic properties such as emotions and the way a food is produced. Figure 1 gives an impression. A food scientist usually limits him- or herself to measurable product properties, named quality attributes, such as taste, colour, microbial quality, nutritive value (e.g., protein or vitamin content), etc. In view of the discussion later on, it is worthwhile to spend some more words on quality in the eyes of a food scientist. We could specify quality attributes further into performance indicators. For instance, taste is a quality attribute that could be subdivided in sweet, sour, bitter, salt. A performance indicator for salt could be the sodium-chloride content, for sweet it could be the sugar content. It is important to realize that many performance indicators can determine a quality attribute, and also that several performance indicators interact with each other. A quality attribute usually changes as a function of time along the food chain from the primary producer to the consumer; in fact, mostly quality decreases. Only in some cases, such as wine or cheese, does the quality improve in time. A prime task of a food technologist is to reduce this loss in quality as much as possible.

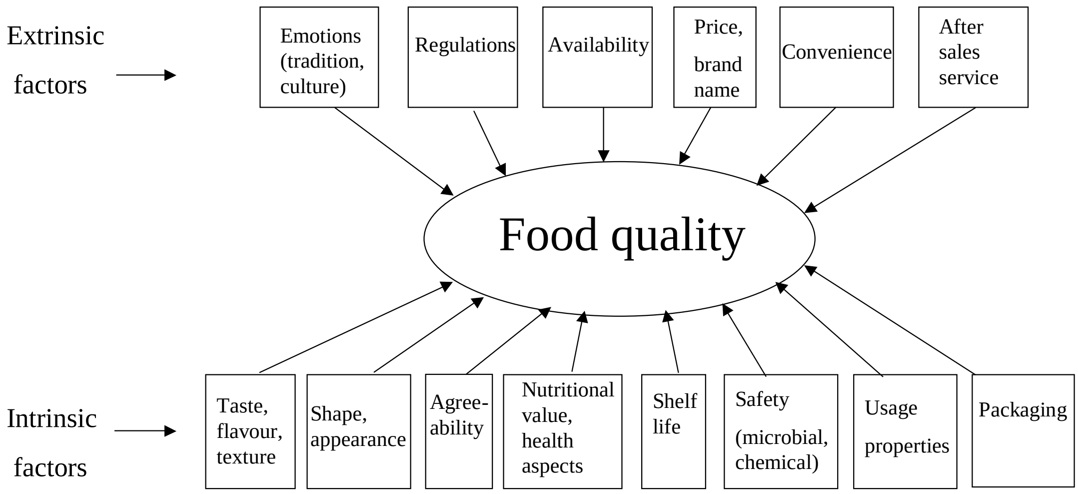

Figure 1. Some intrinsic and extrinsic factors determining food quality

Nowadays it is realized that food quality is not only determined by food-processing operations. Rather, food quality is affected by every operation in the food production chain. Figure 2 gives a very simplified view of this food chain. One of the problems is now that there are several dimensions of quality along the food chain: see Figure 3.

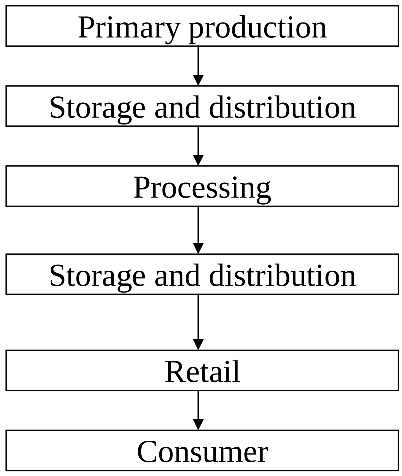

Figure 2. Simplified view of the food production chain

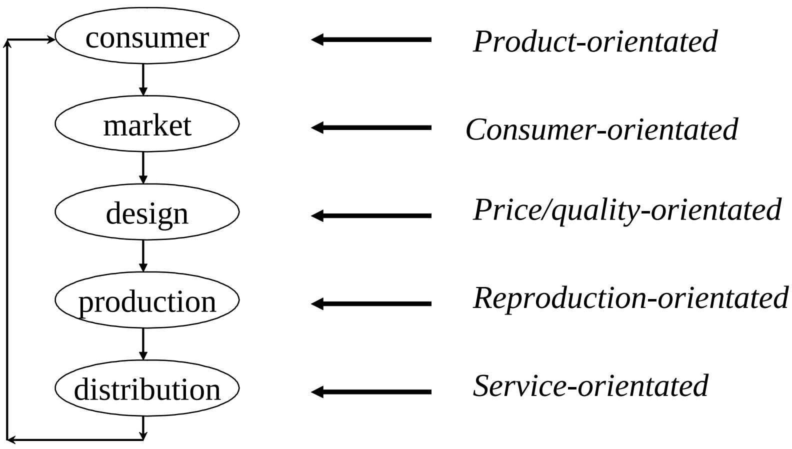

Figure 3. Quality dimensions along the food production chain

Coming back now to the task of a food technologist to reduce loss in quality, he/she needs to know what happens in the various chain elements (Figure 4). In each chain element, factors affecting a quality attribute can have a certain effect. For instance, temperature is a very important factor for almost every quality attribute in every element of the food chain. Other important factors could be gas composition, relative humidity, contamination by micro-organisms or chemicals, etc. When it comes to modelling food-quality attributes, mathematical models are very useful to describe changes as a function of such factors. These models must have a statistical basis when variability and uncertainty characterize the food-quality attributes.

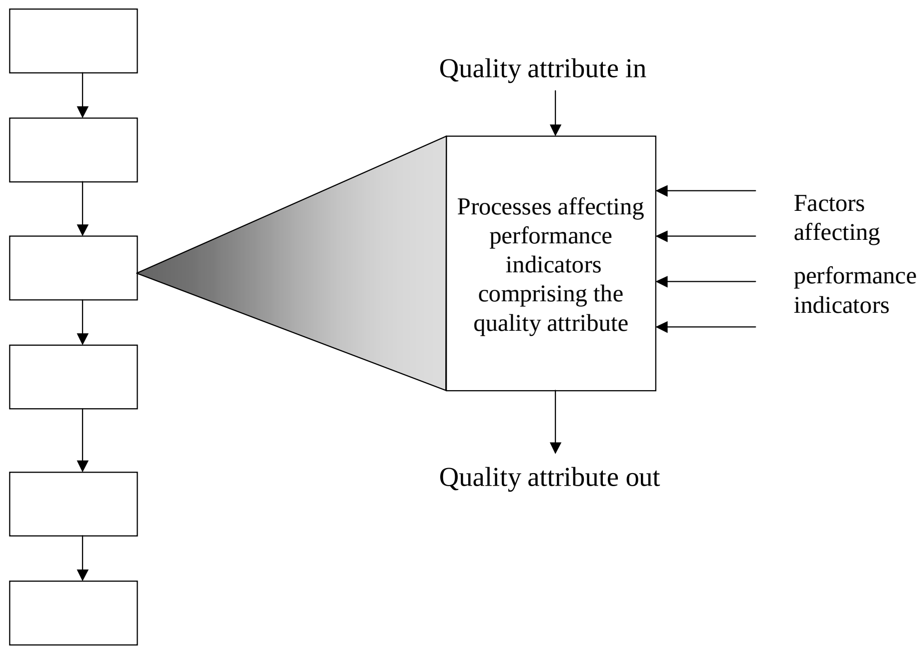

Figure 4. Conceptual model of quality change along the food chain and possible effects in each element of the chain

We will now give some examples where mathematical models describing food-quality changes could be improved by using a Bayesian approach.

Mathematical models to predict food-quality change

As indicated above, models that are able to predict certain quality attributes as a function of relevant product and processing conditions are very useful. Until recently, only relatively simple models were used, mostly based on chemical kinetics. For instance, the degradation of vitamin C can in many cases be described as a so-called first-order reaction at constant temperature:

$$[\text{vitaminC}] = [\text{vitaminC}]_0 \exp(-kt),$$

in which k is the rate constant (s^{-1}) and $[\text{vitamin}]_0$ the initial vitamin-C content, t is the time (s), and the temperature dependence of the rate constant k is modelled via the so-called Arrhenius equation:

$$[\text{vitaminC}] = [\text{vitaminC}]_0 \exp\left(-k_0 t \exp\left(-\frac{E_a}{RT}\right)\right).$$

The parameters k_0 and E_a are the pre-exponential factor and activation energy, respectively. These parameters can be estimated from experiments by regression. This can be done via ‘normal’ statistical approaches to arrive at a point estimate and a confidence interval. There may be two advantages in the use of Bayesian statistics for such models. One advantage could be that a posterior probability distribution is the output rather than a point estimate with confidence interval. This is done in one reference (in water research) for a comparable type of model (Borsuk and Stow 2000), and another one in pharmaceutical research (Lunn et al. 2002). Another possible advantage is model discrimination based on a Bayesian approach. It is the author’s opinion that model discrimination should be more applied than is currently done in the food-science literature. Models that describe changes in foods are usually related to chemical kinetics and microbial-growth models. It is not very common to test several models, and a model is usually selected rather arbitrarily. Model discrimination could be quite useful and helpful to find a better performing model and hence to find a better way to describe food quality. A Bayesian approach seems to be quite powerful (Reilly 1970; Hsiang and Reilly 1971; Reilly and Blau 1974; Stewart, Henson and Box 1996). To the author's knowledge the use of the Bayesian Information Criteria (BIC) for model discrimination has not been applied yet to food-science problems. In general terms BIC has been discussed in Burnham and Anderson (1998), next to Akaike's Information criterion, and it should in that sense be useful in food-science problems as well.

However, a Bayesian approach becomes essential when not just one component is modelled (uniresponse modelling) but rather more reactants and products at the same time, hence when more than one component is modelled in a reaction network. This approach is called multiresponse modelling and the present author has applied it to quite a few cases in food problems (Van Boekel 1996a; 1996b; 1998; 1999; 2000a; 2000b; 2001; Van Boekel and Martins 2002). The advantage of multiresponse over uniresponse modelling is that kinetic models can be tested much more rigorously and the resulting parameter estimates are much more accurate because all the information in responses is taken into account. Suppose we have a reaction of the kind:

$A \rightarrow B \rightarrow C$

and we are able to measure the fate of the components A, B and C in the course of the reaction. The reason that Bayesian statistics is of importance in this case is related to the error structure of experimental data:

$$\Sigma = \begin{bmatrix} \sigma_{AA} & \sigma_{AB} & \sigma_{AC} \\ \sigma_{BA} & \sigma_{BB} & \sigma_{BC} \\ \sigma_{CA} & \sigma_{CB} & \sigma_{CC} \end{bmatrix}$$

Σ represents the experimental variance-covariance matrix, σ_{AA} the variance in concentration of component A, σ_{AB} the covariance between concentrations of component A and B, etc.. Only when this experimental variance-covariance error matrix is known can (weighted) least-squares estimation be applied. In almost all cases such knowledge will not be available. Box and Draper (1965) have provided a Bayesian solution for this problem. The result is that not the sums-of-squares must be minimized but rather the determinant of the so-called dispersion matrix, i.e. the matrix of sums-of-squares for each component (the diagonal elements of the matrix) and the cross-products of the components (the off-diagonal elements). This dispersion matrix C for the number of experimental runs *u* with response *y* and model expectation η is:

$$C = \begin{bmatrix} \sum_{u=1}^{n}(y_A-\eta_A)^2 & \sum_{u=1}^{n}(y_A-\eta_A)(y_B-\eta_B) & \sum_{u=1}^{n}(y_A-\eta_A)(y_C-\eta_C) \\ \sum_{u=1}^{n}(y_B-\eta_B)(y_A-\eta_A) & \sum_{u=1}^{n}(y_B-\eta_B)^2 & \sum_{u=1}^{n}(y_B-\eta_B)(y_C-\eta_C) \\ \sum_{u=1}^{n}(y_C-\eta_C)(y_A-\eta_A) & \sum_{u=1}^{n}(y_C-\eta_C)(y_B-\eta_B) & \sum_{u=1}^{n}(y_C-\eta_C)^2 \end{bmatrix}$$

or:

$$c_{ij} = \sum_{u=1}^{n} (y_u^i - \eta_u^i)(y_u^j - \eta_u^j)$$

So, minimizing the determinant of this matrix gives the best estimates of the parameters involved. The multiresponse approach is used to some extent in chemical engineering, e.g., (Hunter 1967; Box et al. 1973; Ziegel and Gorman 1980; Stewart and Sorensen 1981; Stewart, Caracotsios and Sorensen 1992; Stewart, Shon and Box 1998) but hardly in food science, except by the present author as indicated above. To the author's knowledge, only one dedicated software package deals with such a Bayesian solution (www.athenavisual.com), a preliminary version of which is discussed in Stewart, Caracotsios and Sorensen (1992). It can of course also be programmed using general mathematical software packages such as MathCad, Mathematica and Matlab.

Food safety

The most important food-quality attribute is microbial safety, i.e. the absence of pathogens and toxins. It is taken for granted by the average consumer that food is safe, in other words he/she trusts that he/she runs no risk. Now, it is possible to make food very safe, for instance by heating it intensely. However, the result is an unattractive food that, moreover, has lost much of its nutritional value. The present trend is for minimally processed foods that are fresh-like. This poses a problem for the food technologist: how to preserve fresh-like quality while safeguarding the microbial safety? This brings us into the realm of risk assessment. The field of the so-called predictive microbiology is booming and many semi-mechanistic models are proposed (e.g. McMeekin, Olley and Ross 1993; McMeekin et al. 2002). Such models are used to predict the growth of microbes in foods as a function of temperature, pH, water activity, etc. However useful they are, most of these models are deterministic and cannot handle uncertainty and variability. Obviously, this is needed for a realistic risk assessment. Bayesian statistics has been advocated in this respect (Vose 2000). As indicated by Vose, <u>total uncertainty</u> is due to <u>variability</u> (e.g. caused by biological variation, which is unavoidable) and to <u>uncertainty</u> (i.e. we do not know everything about the system). This is very relevant for foods, which obviously derive from biological materials that are variable in composition and properties, and due to this complexity much is unknown. Pouillot et al. (2003) discussed the use of Bayesian statistics to estimate uncertainty and variability in bacterial growth. A Bayesian approach with respect to food safety has been advocated recently via the use of Bayesian Belief Networks to model the microbial state of a food (Barker, Talbot and Peck 1999; Barker, Talbot and Peck 2002; Carlin et al. 2000). So, these first attempts indicate that a Bayesian approach has potential, and further research is warranted.

There is an increasing awareness with consumers about safety of foods, and quite a few recent food scares sometimes give the impression that there are big problems. Although experts generally agree that the food safety has never been better than in the current situation, consumers just do not believe that. One study has used a Bayesian approach in establishing consumer trust (Böcker and Albrecht 2001). Such an approach is bringing social and natural sciences together.

Product design

Product design should nowadays be consumer-orientated. The reason for this is that the food market is saturated and consumers (at least in the Western world) can make choices from an abundant supply. This implies that only products that attract the consumer's attention will survive in a very competitive market. In other words, the consumer dictates (by his purchasing behaviour) the market. This process is called chain reversal. Rather than a supply-driven market the food market has developed into a demand-driven market. Figure 5 shows this schematically. It implies that each element in the food chain has to be adapted to consumer wishes. This poses enormous problems because these consumer wishes are rather vague, and moreover constantly changing.

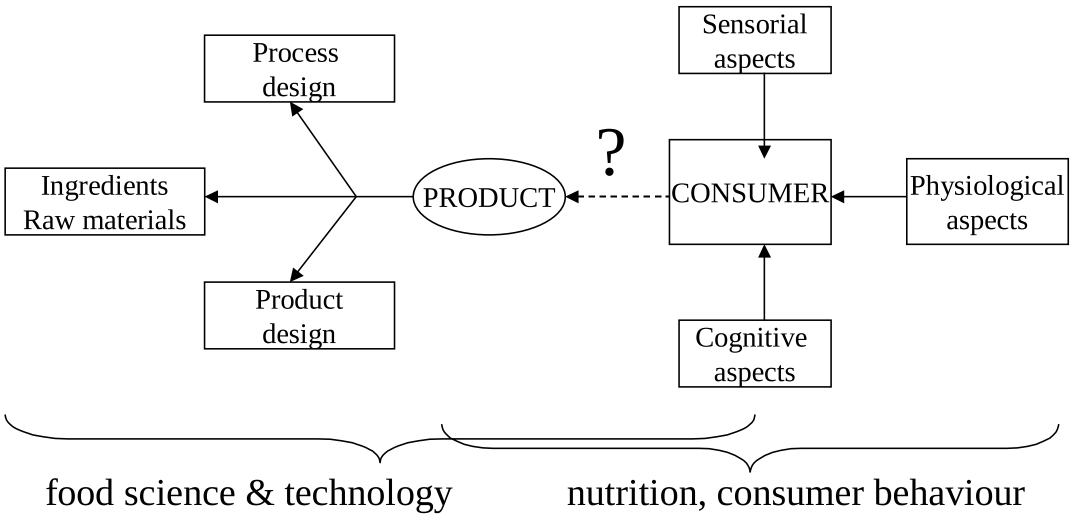

Figure 5. Schematic view of chain reversal

In any case, consumer wishes must in the end be translated into physical product properties. These wishes are driven by complex sensorial attributes in terms of taste, flavour and texture, which are far from clear. The complexity is due to molecular interaction within the food matrix, to molecular interaction with human receptors during eating and to processes taking place in the brain. Sensory evaluation of foods is of prime importance but is notoriously difficult. Bayesian methods might help in this respect, as indicated by some reports (Ming-Hui-Chen, Nandram and Ross 1997; Matsuura et al. 1995; Roussel et al. 2003).

It seems that the effects of uncertainty inherent in sensorial aspects related to food product design could perhaps be modelled via Bayesian Belief Networks. To the author's knowledge only one attempt in this respect has been published (Corney 2000). Figure 6 gives an impression of what such a BBN could look like, though it will be more complicated in practice.

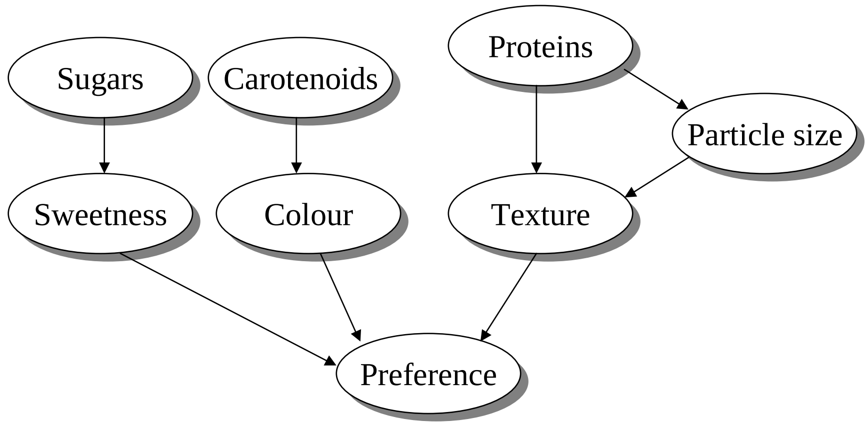

Figure 6. Part of a Bayesian Belief Network for food preference (after Corney 2000).

Expert systems

In food product design many choices and decisions have to be made concerning the use of raw materials, ingredients, processes and the like, taking into consideration food safety and all kinds of legal regulations. Decision-support systems might be quite helpful in food product design. Spiegelhalter et al. (1993) discusses the use of expert systems in general terms. Expert systems should also be quite useful for food-science problems when it comes to food quality management. In the production of foods it is not enough to know about quality of the food per se, but one also must have knowledge about the way the food is actually produced in an industrial environment (Luning, Marcelis and Jongen 2002). Expert systems are then very helpful. A food-related example, namely an expert system for a brewery, has been discussed (Farrow, Goldstein and Spiropoulos 1997). Coupling of decision theory to management systems in the food industry may also have some potential. One such management system is the so-called HACCP (hazard analysis and critical control points). The use of it is now common practice in the food industry. Basically it comes down to identifying critical points with regard to food safety, and then to take appropriate measures to control processes affecting food safety. There seems to be potential to use decision-support systems in the setting up of HACCP and related quality-management systems.

Classification of foods

Classification of foods is something very useful for quality purposes, not least with respect to legal matters, but also to grade foods according to certain quality criteria. Up to now, classification is mostly done by trained experts, obviously with a great deal of subjectivity, and one is in need for more automated and instrumental procedures. Some reports describe Bayesian methods and algorithms that are able to classify foods in this way. Results are already available on olives (Diaz et al. 2000), meat (Lee 2001; Lee and Bertrand 2001), micro-organisms in cheese (Moschetti et al. 2001), honey (Latorre et al. 2000), biscuits (Perrot et al. 1996), beer (Siebert and Stenroos 1989) and cider brandy (Gomis, Tamayo and Alonso 2003). Thus, there seems to be quite some potential for the use of Bayesian approaches in this area.

Conclusion

This paper has attempted to illustrate some typical food-quality problems and how they can be addressed using a modelling approach. From the references cited there seems to be increasing interest in Bayesian solutions, but it is certainly not a generally accepted approach, if only because it is largely unknown. In view of the variability and uncertainty related to food quality and food safety and of the increasing consumer demands it is expected that this will change in the years to come. There is definitely a need for more research on possible applications of Bayesian statistics and modelling in the food-science area.

References

Barker, G.C., Talbot, N.L.C. and Peck, C.C., 1999. Microbial risk assessment for sous-vide foods. *In: Third European symposium on sous vide, Leuven, Belgium*. Katholieke Universiteit Leuven, 37-46.

Barker, G.C., Talbot, N.L.C. and Peck, M.W., 2002. Risk assessment for *Clostridium botulinum*: a network approach. *International Biodeterioration and Biodegradation,* 50 (3/4), 167-175.

Böcker, A. and Albrecht, S., 2001. Risikowahrnehmung und Verbrauchervertrauen nach einem Lebensmittelskandal. *Agrarwissenschaft,* 50, 374-382.

Borsuk, M.E. and Stow, C.A., 2000. Bayesian parameter estimation in a mixed-order model of BOD decay. *Water Research,* 34 (6), 1830-1836.

Box, G.E.P. and Draper, N.R., 1965. The Bayesian estimation of common parameters from several responses. *Biometrika,* 52, 355-365.

Box, G.E.P., Hunter, W.G., MacGregor, J.F., et al., 1973. Some problems associated with the analysis of multiresponse data. *Technometrics,* 15, 33-51.

Burnham, K.P. and Anderson, D.R., 1998. *Model selection and inference: a practical information-theoretic approach*. Springer, New York.

Carlin, F., Girardin, H., Peck, M.W., et al., 2000. Research on factors allowing a risk assessment of spore-forming pathogenic bacteria in cooked chilled foods containing vegetables: a FAIR collaborative project. *International Journal of Food Microbiology,* 60 (2-3 Special Iss. SI), 117-135.

Corney, D.P.A., 2000. Designing food with Bayesian Belief Networks. *In:* Parmee, I. ed. *Adaptive computing in design and manufacture*. Springer-Verlag, 83-94. [http://www.cs.ucl.ac.uk/staff/D.Corney/FoodDesign.html]

Diaz, R., Faus, G., Blasco, M., et al., 2000. The application of a fast algorithm for the classification of olives by machine vision. *Food Research International,* 33 (3/4), 305-309.

Farrow, M., Goldstein, M. and Spiropoulos, T., 1997. Developing a Bayes linear decision support system for a brewery. *In:* French, S. and Smith, J. Q. eds. *The practice of Bayesian analysis*. Edward Arnold, London, 71-106.

Gomis, D.B., Tamayo, D.M. and Alonso, J.J.M., 2003. Evolution of sugars in cider brandy aged in oak barrels: a contribution to its characterization. *Journal of Agricultural and Food Chemistry,* 51 (4), 923-926.

Hsiang, T. and Reilly, P.M., 1971. A practical method for discriminating among mechanistic models. *Canadian Journal of Chemical Engineering,* 49, 865-871.

Hunter, W.G., 1967. Estimation of unknown constants from multiresponse data. *Industrial and Engineering Chemistry Fundamentals,* 6, 461-463.

Latorre, M.J., Pena, R., Garcia, S., et al., 2000. Authentication of Galician (NW Spain) honeys by multivariate techniques based on metal content data. *Analyst,* 125 (2), 307-312.

Lee, D.H., 2001. Comparison of genetic evaluations for weaning weight and carcass traits using threshold model with Bayesian inference and AI-REML in Hanwoo (Korean cattle). *Journal of Animal Science and Technology,* 43, 277-292.

Lee, D.H. and Bertrand, J.K., 2001. Comparison of genetic parameter estimates for carcass traits according to modeling with REML and Gibbs sampling in Korean cattle. *Journal of Animal Science and Technology,* 43, 23-34.

Luning, P.A., Marcelis, W.J. and Jongen, W.M.F., 2002. *Food quality management: a techno-managerial approach*. Wageningen Pers, Wageningen.

Lunn, D.J., Best, N., Thomas, A., et al., 2002. Bayesian analysis of population PK/PD models: general concepts and software. *Journal of Pharmacokinetics and Pharmacodynamics*, 29 (3), 271-307.

Matsuura, K., Hirotsune, M., Hamachi, M., et al., 1995. Modeling of the sensory evaluation of sake by Dempster-Shafer's measure and genetic algorithm. *Journal of Fermentation and Bioengineering,* 79, 45-53.

McMeekin, T.A., Olley, J., Ratkowsky, D.A., et al., 2002. Predictive microbiology: towards the interface and beyond. *International Journal of Food Microbiology,* 73 (2/3), 395-407.

McMeekin, T.A., Olley, J.N. and Ross, T., 1993. *Predictive microbiology: theory and application*. RSP, Taunton. Innovation in microbiology series.

Ming-Hui-Chen, Nandram, B. and Ross, E.W., 1997. Bayesian prediction of the shelf life of a military ration with sensory data. *Journal of Agricultural, Biological and Environmental Statistics,* 1 (3), 377-392.

Moschetti, G., Blaiotta, G., Villani, F., et al., 2001. Comparison of statistical methods for identification of *Streptococcus thermophilus*, *Enterococcus faecalis,* and *Enterococcus faecium* from randomly amplified polymorphic DNA patterns. *Applied and Environmental Microbiology,* 67 (5), 2156-2166.

Perrot, N., Trystram, G., Le Guennec, D., et al., 1996. Sensor fusion for real time quality evaluation of biscuit during baking: comparison between Bayesian and fuzzy approaches. *Journal of Food Engineering,* 29 (3/4), 301-315.

Pouillot, R., Albert, I., Cornu, M., et al., 2003. Estimation of uncertainty and variability in bacterial growth using Bayesian inference: application to *Listeria monocytogenes*. *International Journal of Food Microbiology,* 81 (2), 87-104.

Reilly, P.M., 1970. Statistical methods in model discrimination. *Canadian Journal of Chemical Engineering,* 48, 168-173.

Reilly, P.M. and Blau, G.E., 1974. The use of statistical methods to build mathematical models of chemical reacting systems. *Canadian Journal of Chemical Engineering,* 52, 289-299.

Roussel, S., Bellon-Maurel, V., Roger, J.M., et al., 2003. Fusion of aroma, FT-IR and UV sensor data based on the Bayesian inference: application to the discrimination of white grape varieties. *Chemometrics and Intelligent Laboratory Systems,* 65 (2), 209-219.

Siebert, K.J. and Stenroos, L.E., 1989. The use of multivariate analysis of beer aroma volatile compound patterns to discern brand-to-brand and plant-to-plant differences. *Journal of the American Society of Brewing Chemists,* 47, 93-101.

Spiegelhalter, D.J., Dawid, A.P., Lauritzen, S.L., et al., 1993. Bayesian analysis in expert systems. *Statistical Science,* 8, 219-283.

Stewart, W.E., Caracotsios, M. and Sorensen, J.P., 1992. Parameter estimation from multiresponse data. *American Institute of Chemical Engineers Journal,* 38, 641-650.

Stewart, W.E., Henson, T.L. and Box, G.E.P., 1996. Model discrimination and criticism with single-response data. *Aiche Journal,* 42 (11), 3055-3062.

Stewart, W.E., Shon, Y. and Box, G.E.P., 1998. Discrimination and goodness of fit of multiresponse mechanistic models. *Aiche Journal,* 44 (6), 1404-1412.

Stewart, W.E. and Sorensen, J.P., 1981. Bayesian estimation of common parameters from multiresponse data with missing observations. *Technometrics*, 23, 131-141.

Van Boekel, M.A.J.S., 1996a. Kinetic modelling of sugar reactions in heated milk-like systems. *Netherlands Milk and Dairy Journal,* 50 (2), 245-266.

Van Boekel, M.A.J.S., 1996b. Statistical aspects of kinetic modeling for food science problems. *Journal of Food Science,* 61 (3), 477-485,489.

Van Boekel, M.A.J.S., 1998. Modelling of chemical reactions in foods: a multiresponse approach. *In:* Tijskens, L. M. M. and Hertog, M. L. A. T. M. eds. *Proceedings of the international symposium on applications of modelling as an innovative technology in the agri-food-chain, Model-It, Wageningen, Netherlands, 29 November-2 December, 1998. Acta Horticulturae ISHS,* Vol. 476. 149-155.

Van Boekel, M.A.J.S., 1999. Testing of kinetic models: usefulness of the multiresponse approach as applied to chlorophyll degradation in foods. *Food Research International,* 32 (4), 261-269.

Van Boekel, M.A.J.S., 2000a. Kinetic modelling in food science: a case study on chlorophyll degradation in olives. *Journal of the Science of Food and Agriculture,* 80 (1), 3-9.

Van Boekel, M.A.J.S., 2000b. Modelling of chemical reactions in foods: possibilities and pitfalls. *Czech Journal of Food Science,* 18, 1-4.

Van Boekel, M.A.J.S., 2001. Kinetic aspects of the Maillard reaction: a critical review. *Nahrung/Food,* 45 (3), 150-159.

Van Boekel, M.A.J.S. and Martins, S.I.F.S., 2002. Fate of glycine in the glucose-glycine reaction: a kinetic analysis. *International Congress Series,* 1245, 289-293.

Vose, D., 2000. *Risk analysis: a quantitative guide*. 2nd edn. Wiley, Chichester.

Ziegel, E.R. and Gorman, J.W., 1980. Kinetic modelling with multiresponse data. *Technometrics,* 22, 139-151.