Knowledge Engineering, a Useful Tool for Integrating Food Chain

Cédric Baudrit, Patrice Buche, Guy Della Valle, Sébastien Destercke, Kamal Kansou, Amadou Ndiaye, Nathalie Perrot, Rallou Thomopoulos, Gilles Trystram

Knowledge Engineering, a useful tool for integrating food chain

C. Baudrit[1], P. Buche[2], G. Della Valle[3 *], S. Destercke[2], K. Kansou[3,4], A. Ndiaye[4], N. Perrot[1], R. Thomopoulos[1], G. Trystram[5].

* 33(0)240675000, 33(0)240675043, dellaval@nantes.inra.fr

[1] JRU 782, AgroParisTech, INRA-CEPIA, Thiverval Grignon, 78850 France

[2] INRA - CEPIA / LIRMM / EPI GraphIK, 2 place Viala, Montpellier, 34000 France

[3] INRA- CEPIA, UR 1268 BIA, rue de la géraudière, Nantes, 44316 France

[4] JRU 927, INRA-CEPIA, CNRS, Univ. Bordeaux 1, Talence, 33405 France

[5] JRU 1145, INRA-CEPIA, CNAM, AgroParisTech, Massy, 91744 France

Abstract. *Issued from cognitive sciences, computer sciences and applied mathematics, Knowledge (K) Engineering encompasses modelling and can be roughly defined by three steps: acquisition of available K, representation of acquired K and computational use of represented K for simulation, validation and optimisation purposes. These different steps are illustrated for various applications to food industry like the durum wheat chain, dough mixing for breadmaking, cheese ripening, based on specific approaches like decision trees and conceptual graphs, qualitative reasoning, Bayesian networks and optimisation under uncertainty and multi-criteria flexible querying, respectively. The application of such approaches opens prospects for the virtual design of food products which will be of help for the sustainable production of high quality foods.*

Keywords: bread, complex system, cheese, durum wheat, modeling, multi-disciplinary, multi-scale.

Introduction

Foods are now developped in response to new demands of consumers, concerned by environmental and nutritional issues. An evolution of the know how in food industry is needed, although it is difficult to simultaneously improve the qualification of the staff and upgrade the technological level of production lines. It is therefore important to support the capacity of developing practices of quality management and technology choices within the food processing chain. Conversely, on the scientific side, during food processing, the close interaction between continuous structural changes and transfer mechanisms impairs the complete modeling of coupled physical, chemical and microbiological phenomena. So, despite the increase of scientific papers in this area and the progresses of our understanding on multi-scale food structural changes, the knowledge is fragmented and incomplete (Perrot *et al.*, 2006). The whole food processing chain can be viewed like a complex system, like for recent scientific issues in biology. We propose to adapt and implement concepts able to take into account this complexity. Their application relies on tools able to take explicitly into account the fragmented and heterogeneous knowledge available on the dynamics of the process, with uncertainty on the global behaviour of the system. Recently, some of these tools (Monte Carlo, Neural and Bayesian Networks, fuzzy logic, expert systems...) have been implemented in various applications, such as immunology (Cohen and Harel, 2007), systems engineering (Beckerman, 2000), bioinformatics (Desiere *et al.*, 2001). In our context, these tools have to be tested for the multiscale dynamic reconstruction of the processes of food models by an Integrated Knowledge Model (IKM) as it has been envisaged in the frame of the Dream EU project (Fig.1).

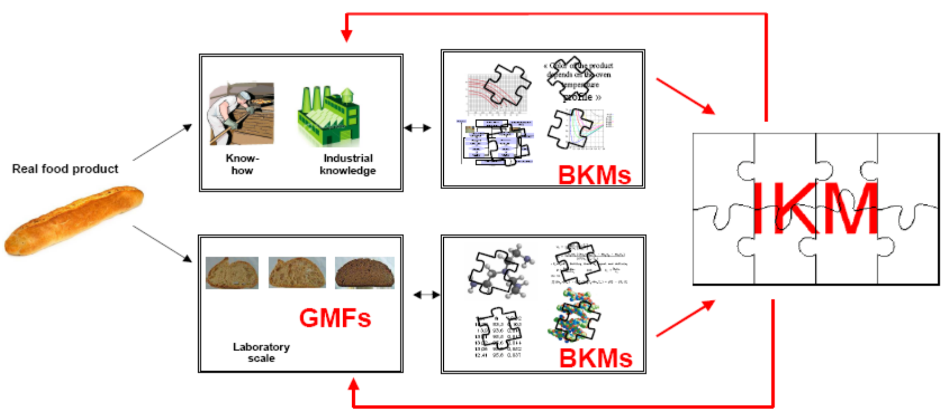


Figure 1: Example of model integration (IKM) in the EU-Dream project (GMF= generalized model foods; BKM = basic knowledge model).

Generic model foods (GMF) are realistic foods considered to be representative of a whole class of foods through variations of process and composition. Then, cognitive maps of technical knowledge (know-how) on processing of the selected GMFs are to be drawn and integrated with the numerical models, namely BKMs. These last models may be those currently set in the food engineering area, and based on physical laws, chemical kinetics… usually formalized by ordinary and partial differential equations (see for instance Bimbenet et al., 2007); but they are also those available from technological know-how (expert's rules), which have thus to be formalized. The purpose of Knowledge Engineering is to build this IKM and this paper illustrates how it can be applied to the management of food chain processing by considering the following examples: durum wheat chain, breadmaking, cheese ripening. They are presented in this order of decreasing sizes of knowledge grain and scale of the considered domain, so that the relevant tools may also be described briefly.

Decision support system to manage the durum wheat chain

The need for food security has triggered the development of tools combining models with databases in the area of predictive microbiology like Sym'Previus (Haemmerlé et al., 2007) but, up to our knoweledge, there are no tools available to manage a whole food chain until food products, integrating heterogeneous information sources on nutritional, sensory and technological aspects. Noting the basic role of cereals in the food of mankind, the management of durum wheat products has recently been addressed, taking into account experimental data, from scientific literature, and expert statements describing commonly admitted mechanisms in a qualitative way (Thomopoulos et al., 2009). Since this approach is not based on predetermined models, a specific learning technique has been used, namely decision trees.

Expert knowledge representation and exploitation

Since we address a complete food chain, we have to build a system to handle disparate information presented under various forms (quantitative, descriptive…) and referring to very different domains (processing operations, product quality…). These issues may be addressed by using tools for the representation of expert knowledge like the conceptual graph model and rules. The latter well illustrate the link "If…then…" , like for instance (Fig.2): "Drying a pasta product in which peroxydase is active will yield a pasta with brown color". Here "drying" is a unit operation and "color" a qualitative variable.

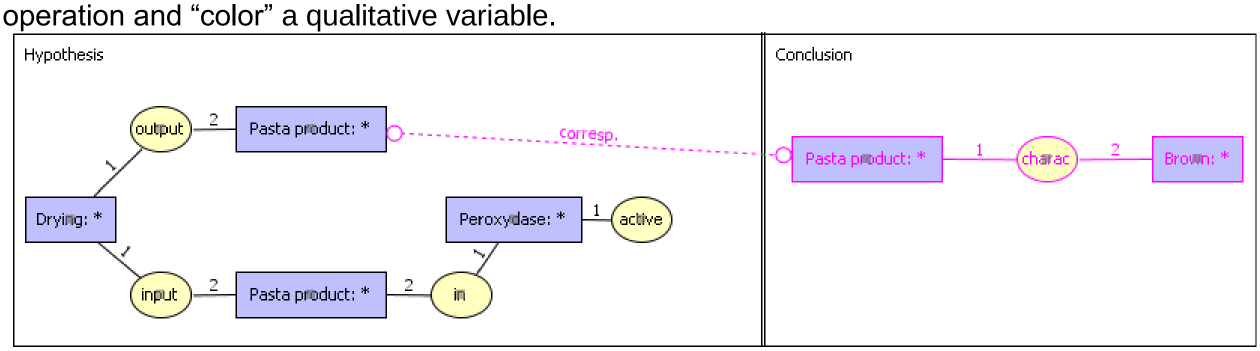


Figure 2: Example of use of expert knowledge: an expert rule in the conceptual graph model.

These rules can be applied either in forward chaining in order to predict an output result in terms of quality, or in backward chaining to determine the possible conditions that can lead to the expected properties (reverse engineering).

Experimental data exploitation by decision trees

The management of the food chain may be performed using decision trees, which can be viewed as a collection of rules implementing its different variables. In the decision tree, leaves represent the average value (for a continuous variable) or class (for a symbolic variable), whereas the branch represent the conjunction of inputs that lead to this classification (or value). After the trees have been learnt from the data set, we will use them in their predictive form in order to predict the classification, or average value, from input parameters. The knowledge management system involved 29 unit operations of the durum wheat chain and 56 quality variables characterizing the various families of products (precooked grain, couscous, pasta…).

Among those, let us examine the impact of the “cooking in water” operation on the “vitamin content” quality, for which 145 experimental results have been reported from 11 references. On the basis of vitamin initial composition and previous processing steps, once the most discriminating parameters are determined (here type of vitamin), a family of decision trees is generated. One of them is depicted in Fig.3. It provides the average value of “vitamin decrease” (%) as terminal leaves, the box plot below indicates the distribution of this variable around the average value.

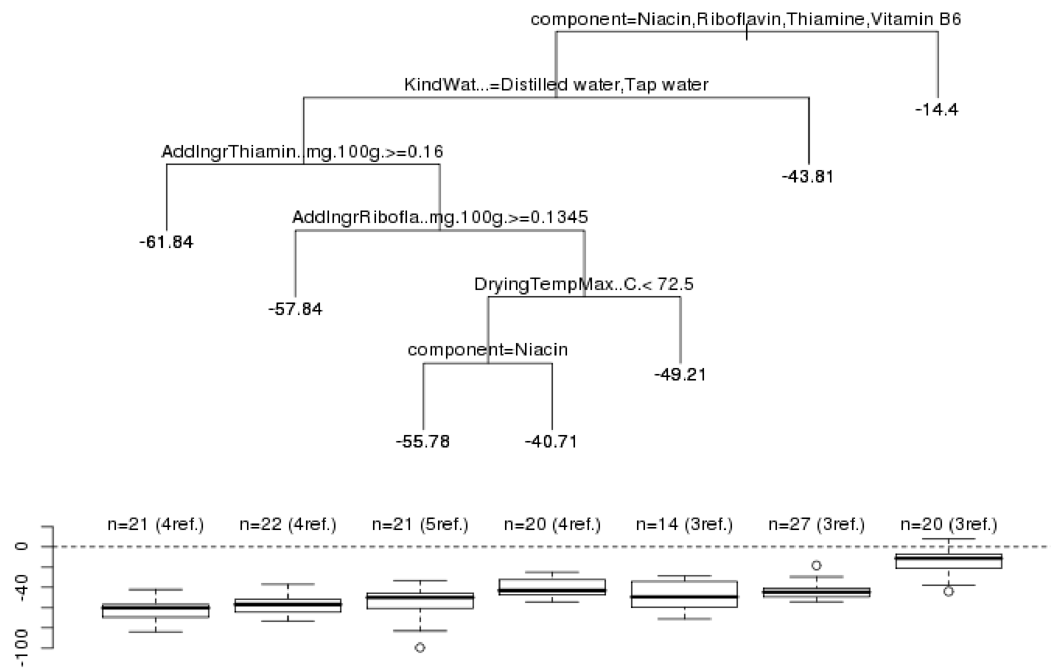


Figure 3: Example of descriptive decision tree learnt from the “cooking in water” data set.

Synergy between expert knowledge and experimental data

Expert knowledge and experimental data complete each other in two ways: (1) a confrontation of both is performed by testing the expert rules using a confidence rate; exceptions to the rule may be identified in order to build new rules (Thomopoulos, 2008); (2) an interactive procedure is proposed (Johnson et al., 2010) to improve the decision tree results through adequate knowledge elicitation.

Formalization of expert knowledge in French breadmaking

Among cereal processing chain, breadmaking has a specific part and the tool described by Young (2007), namely ‘Bread Advisor’ is one of the first software based on expert knowledge that proposes information about processing, and diagnoses about possible defaults. Starting from expert knowledge in French breadmaking, but also considering scientific results, we have developed a knowledge based system (KBS) based on formal algebraic representation of breadmaking operations allowing to predict the state of dough or bread from the inputs and processing conditions. In addition we have also proposed a glossary of terms of quality pasta and French bread loaves.

Knowledge and technological backgrounds: craft processing of wheat flour

French breadmaking has long been a traditional activity relying on craftsman's manual skills. Today, part of know-how is automated and it is also an industrial activity, which demonstrates a good knowledge of ingredients and baking process (Roussel and Chiron, 2002). This industrialization has been favoured by the work of breeders and geneticists for stable wheats for breadmaking, but it involves the risk of supply uniformity and economic dependance for the craft, in the long term. In addition, causal relationships between the physico-chemical ingredients and sensory and nutritional characteristics of bread according to the sequence of unit processes (mixing, ..., cooking) remain ill known. As part of any industrial process, knowledge related to the explanation or the implementation of these processes involve a great deal of tacit knowledge. The elicitation of this knowledge requires a scientific analysis of the processes involved, through a formalization of the results of this analysis, but also a knowledge survey from the different actors involved in these processes. The use of a common formalization of these two components allows to build a comprehensive knowledge base. This base will be both the evidence of current practices and the starting point for investigating future practices. The scientific objectives were to make knowledge emerging on accessible scientific issues and to develop a formal tool for representing algebraically expert knowledge expressible as rules to facilitate the integration of knowledge.

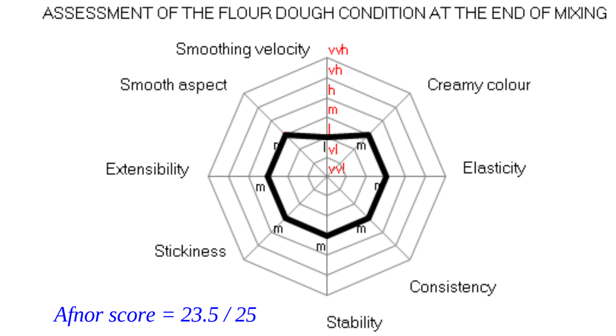
Approach and results: qualitative algebra and knowledge based system

We developed an algebra (Q-algebra) for writing as a qualitative function each set of rules for evaluating a characteristic of the wheat flour dough and bread. It has been used to model the states of the dough at the end of first mixing and end of texturing operations, the two successive steps of mixing (Ndiaye et al., 2009). The state of the dough at the end of first mixing is influenced by the characteristics of the ingredients (%flour, water, protein and pentosan content…); the state of the dough at the end of texturing is defined by the following descriptors: smoothing velocity SV, smooth aspect SA, Extensibility Ext, stickiness Stic, Stability Stab, Consistency Cons, Elasticity Elas and Creamy Colour CC. This state is influenced by its consistency (w) at texturing start, by the target temperature at the end of mixing (x), and by the mixer settings: the difference in linear velocity between the arm and bowl (y) and the expected heat dissipated during texturing (z). These words for describing behavior of the dough and bread were selected thanks to a glossary of terms defining dough quality and bread baking, developed in French and available on the Web (Roussel et al., 2010). The terms of language at different levels – empirical, technological and scientific - were identified by: (1) expressing explicitly semantic relationships between terms from different levels of knowledge, (2) accounting for rheological knowledge that can both describe the behavior of dough and suggest instrumental methods where only sensory assessments exist.

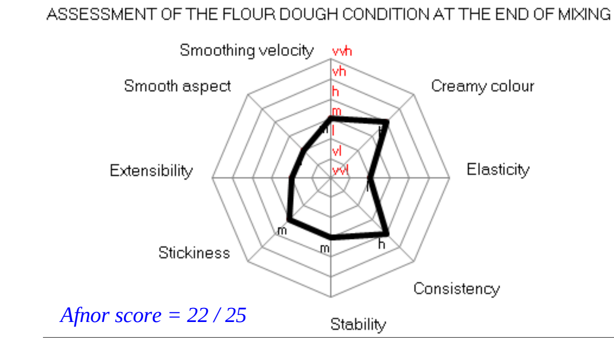
$$
\begin{aligned}
&\mathrm{SV} &&\approx (1\otimes w)\oplus \alpha(y,z)\\
&\mathrm{SA} &&\approx \top((1\otimes w)\oplus(1\otimes\perp(x))\oplus\top(z)\oplus(1\otimes\perp(z)))\\
&\mathrm{Ext} &&\approx (1\otimes\perp(w))\oplus(1\otimes\perp(x))\oplus\top(z)\oplus(1\otimes\perp(z))\\
&\mathrm{Stic} &&\approx \perp((1\otimes w)\oplus\perp(\perp(x)\oplus z))\\
&\mathrm{Stab} &&\approx \perp((1\otimes w)\oplus\perp(x)\oplus z)\\
&\mathrm{Cons} &&\approx w\oplus(1\otimes\perp(x))\oplus(1\otimes z)\\
&\mathrm{Elas} &&\approx (1\otimes\perp(w))\oplus x\oplus\top(z)\\
&\mathrm{CC} &&\approx (1\otimes z)
\end{aligned}
$$

Figure 4: Qualitative functions of dough state descriptors after mixing (Ndiaye et al., 2009)

The functional writing of the set of expert rules allows to calculate the state of a dough according to the state of the inputs of the operation and its settings (Fig. 4). The qualitative determination of the function corresponding to a set of rules is one of the main difficulties of this formalization, a purpose for which an algorithm for automatic determination of qualitative functions is being developped. Once implemeted in the KBS, it allows to determine the states of the inputs of the processing operation (Fig.5).



(a)



(b)

Figure 5: Examples of the KBS outputs: predictions of dough state at the end of texturing, starting from a standard consistency at the end of first mixing (350 ≤ w≤ 450 UB), a standard target temperature (22 ≤ x ≤ 25 °C), an average difference of velocity y and a heat dissipation z (a) medium and (b) low. Scale ranges from vvh (very excessive) to vvl (very insufficient).

Qualitative algebra is also useful to prospect expert's reasoning in order to build KBS (Kansou et al., 2008), which, in the near future, can be also used to determine the settings of the operation, and the dough composition, that achieve a desired state of dough (reverse engineering) like for the design of composite materials (Michaud et al., 2009).

Modelisation and optimisation of the cheese ripening process

After breadmaking, cheese manufacturing is certainly the most representative area of food industry in France, and it processes half of the milk produced in this country. In spite of this industrial importance, soft cheese like Camembert is an ecosystem and a bioreactor difficult to assess in its entirety. Despite extensive research conducted on this product, knowledge remains fragmentary and incomplete and no model provides a comprehensive representation of the process. In this context, we have used dynamic Bayesian networks to model the network of interactions occuring at different scales and reconstruct its dynamics (Baudrit et al., 2010). A model of cheese mass loss (Helias et al., 2007) was then considered for optimizing the ripening process. In this purpose, a viability kernel representing a compromise between production costs and ripened cheese quality was computed (Sicard et al., 2009).

Coupling heterogeneous knowledge with dynamic Bayesian networks (DBNs)

The concept of DBNs provides a practical mathematical formalism that enables to describe dynamical complex systems tainted with uncertainty. DBNs are an extension of classical Bayesian networks that rely on probabilistic graphical models in which nodes representing random variables are indexed by time. They are very useful tools for combining expert knowledge with data at different levels of knowledge, where the structure can be explicitly built on the basis of expert knowledge and conditional probability, quantifying dependence between variables, can be automatically learned without *a priori* knowledge on the basis of a dataset.

From operational and scientific knowledge, Baudrit *et al*. (2010) defined the structure of a DBN providing a qualitative representation of the coupled dynamics of microorganism behaviour (*Kluyveromyces marxianus* (Km), *Geotrichum candidum* (Gc), *Brevibacterium aurantiacum* (Ba) with their substrate consumptions (lactose (lo), lactate (la)) influenced by temperature (T) and involving the sensory changes (Odour, Under-rind, Coat, Colour and Humidity) of cheese during ripening (Fig.6).

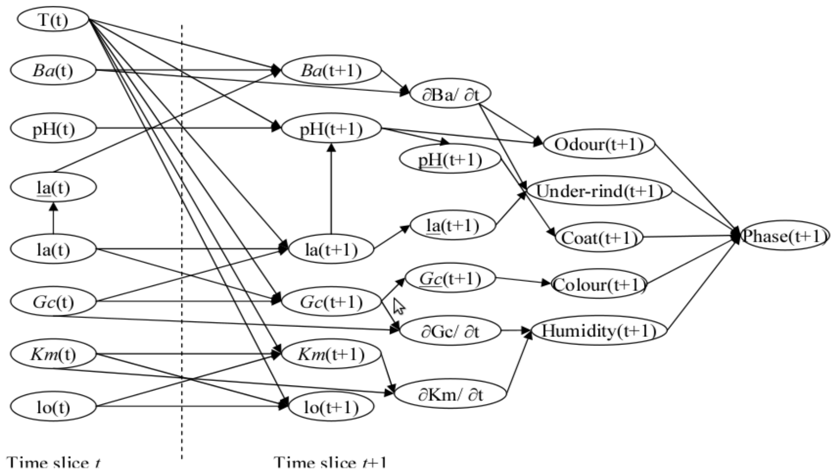
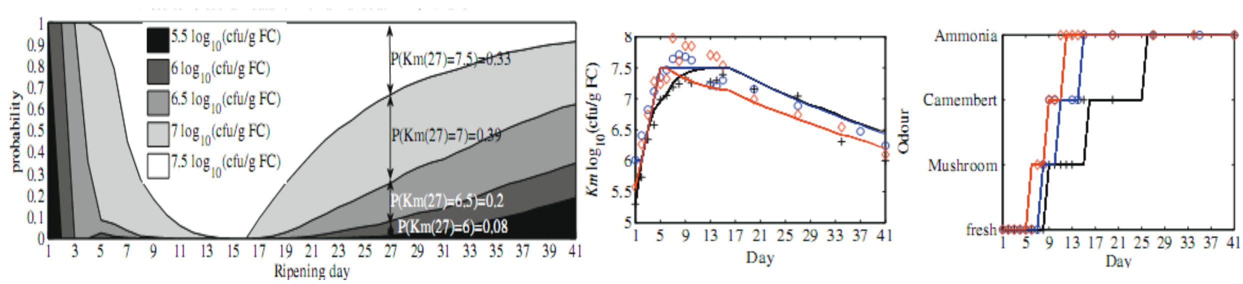


Figure 6: Dynamic Bayesian network representing the coupled dynamics of micro-organism growth with their substrate consumptions influenced by temperature and involving the sensory changes of cheese during the ripening process.

After the learning step to define conditional probability distributions from experimental trials with various temperature and humidity, DBNs inferences can be carried out in order to simulate the behaviour of microbial activities associated with sensory development, for instance the beliefs of the possible trajectories of the yeast *Km* during ripening at 8°C (Fig.7a). This figure means, for instance, that at the 27th day of ripening, the concentration of Km has a probability of 39% to be $\approx 10^7$cfu/g FC and that it cannot be lower than 3.10^5.



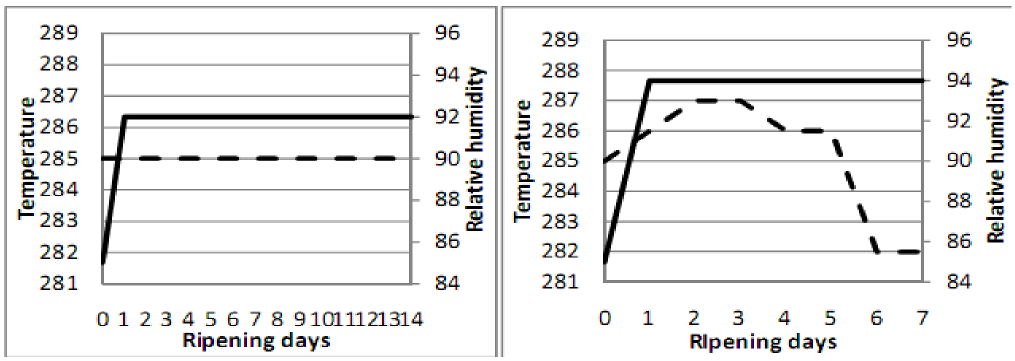
(a) (b) (c)

Figure 7: DBN results of (a) Km(t) probability distribution at 8°C, predictive evolutions of (b) Km microbial growth and (c) odour, versus experimental data for ripening performed at T= 8 (+), 12 (o) and 16°C (◊), RH=98%.

From these results of DBNs simulations, the mean evolution of *Km* (Fig.7b) as well as the modal evolution of odour properties(Fig.7c) may be estimated and compared to experimental data. The model was thus shown to be able of (1) coupling and integrating heterogeneous knowledge at different scales; (2) predicting the evolution of microbial activities and sensory properties with an overall average adequacy rate of about 85% to experimental data.

Process optimization by the viability theory and geometric calculus

Viability theory aims at controlling a dynamical system, here cheese during ripening, in order to maintain it in a given set of evolutions, namely the viability kernel. The viability kernel was defined by associating a target on cheese mass at the end of ripening ($\approx$280g) and constraints on microorganisms respiration. It was then computed thanks to a classical heat&mass transfer model, proposed by Helias et al. (2007) that predicts cheese mass, surface temperature and the respiration of the microorganisms. Meanwhile, the cost trajectories, involving the number of control variations and the ripening time are computed to define the compromise between cheese quality and energy consumption saving (Sicard et al., 2009). Then, optimal trajectories, lowering cost, were found; among those, one reached the mass target after 8 days ripening, and the results of its control variations are presented in Fig. 8b, to be compared to the conventional control performed for 12 days ripening (Fig. 8a), without quality loss.



(a) (b)

Figure 8: Control (T(K) ---, RH(%) —) of ripening chambers, (a) conventional in industry and (b) optimal computed applying viability approach.

Whilst setting a higher humidity (94%), it imposes a daily change of temperature between 14 and 9°C. These controls have been applied to real ripening chambers; the analysis of processed cheese gave sensory results very close to those obtained under classical conditions (12°C, 92%). This is an example of the application of the reverse engineering approach to a single food processing operation.

Conclusions

Food processing represents a complex system with incomplete knowledge and numerous interactions, that are not fully available from the scientific literature, but also rely largely on technological expertise. Managing such a system is a real challenge, that can be addressed by Knoweldge Engineering. After having described briefly the principles of this approach, we have presented some of its recent applications to food processing and described briefly the relevant tools implemented in each application. By integrating models together, they may build an integrated mathematical model (IKM) for multistage dynamic reconstruction of foods which can in turn be implemented for reverse engineering. The application of such approach opens prospects for the virtual design of food products, which will be of help for the sustainable production of high quality foods.

Acknowledgements

The authors are grateful to all participants of the Cepia Division's Knowledge Engineering Network and of the PNRA-Incalin program for which the support of ANR is also acknowledged.

References

Baudrit, C., M. Sicard, P.H. Wuillemin, N. Perrot. 2010. Towards a global modelling of the Camembert-type cheese ripening process by coupling heterogeneous knowledge with dynamic Bayesian networks. *Journal of Food Engineering.* **98**: 283-293.

Beckerman, L.P. 2000. Application of complex systems science to systems engineering. *Systems engineering*. **3**: 96-103.

Cohen, I.R., Harel, D. 2007. Explaining a complex living system: dynamics, multi-scaling and emergence. *Journal of the royal society interface*. **4**: 175-182.

Desiere, F., German, B., Watzke, H., Pfeifer, A. & Saguy, S. 2001. Bioinformatics and data knowledge: the new frontiers for nutrition and foods. *Trends in Food Science & Technology.* ***12***: 215-229.

DREAM EU Project -7th framework programme. 2009. Design and development of REAlistic food Models with well-characterised micro- and macro-structure and composition. Available at: http://dream.aaeuropae.org/

Haemmerlé O., Buche P. & Thomopoulos R. 2007. The MIEL system: uniform interrogation of structured and weakly-structured imprecise data. *Journal of Intelligent Information systems*, **29**: 279–304.

Helias, A. , Mirade, P. S., & Corrieu, G. 2007. Modeling of camembert-type cheese mass loss in a ripening chamber: Main biological and physical phenomena. *Journal of Dairy Science*, **90**: 5324–5333.

Johnson, I., Abécassis, J., Charnomordic, B., Destercke, S. & Thomopoulos, R. 2010. Making Ontology-Based Knowledge and Decision Trees interact: an approach to enrich knowledge and increase expert confidence in data-driven models. In *Proceedings of KSEM'10, Lecture Notes in Artificial Intelligence*, 6291: 304-316, Springer.

Kansou, K., Della Valle, G. & Ndiaye, A. 2008. Qualitative modelling to prospect expert's reasoning. Frontiers in Artificial Intelligence and Applications **179:** 94-105.

Ndiaye, A., Della Valle, G. & Roussel, P. 2009. Qualitative modelling of a multi-step process: the case of French breadmaking. *Expert Systems with Applications* **36**: 1020-1038.

Perrot, N., Ioannou, I., Allais, I., Curt, C., Hossenlopp, J., & Trystram, G. 2006. Fuzzy concepts applied to food product quality control: a review. *Fuzzy Sets and Systems* **157**: 1145-1154.

Michaud, F., P. Castera, C. Fernandez and A. Ndiaye (2009). Meta-heuristic Methods Applied to the Design of Wood-Plastic Composites, with Some Attention to Environmental Aspects. *Journal of Composite Materials* **43**: 533-548.

Roussel, P. & Chiron H. 2002. Les pains français. Evolution, qualité, production. Vesoul, France: Mae Erti Ed, 433p.

Roussel, P., H. Chiron, G. Della Valle & A. Ndiaye. 2010. Recueil de connaissances sur les descripteurs de qualité des pâtes et des pains ou variables d'état pour la panification française. http://www4.inra.fr/cepia/Editions/glossaire-pains-francais.

Sicard M., Perrot, N., Baudrit, C., Reuillon, R., Bourgine, P., Alvarez, I. & Martin, S. 2009. The viability theory to control food processes. *Eur. Conf Complex Systems*, Univ. Warwick (UK).

Thomopoulos, R. 2008. Learning Exceptions to Refine a Domain Expertise. In *Encyclopedia of Data Warehousing and Mining – 2nd Edition*, pp. 1129-1136. Hershey, PA, USA: Information Science Reference.

Thomopoulos, R., Charnomordic, B., Cuq, B. & Abecassis, J. 2009. Artificial intelligence-based decision support system to manage quality of durum wheat products. *Quality Assurance and Safety of Crops & Foods* **1**:179-190.

Young, L.S. 2007. Application of Baking Knowledge in Software Systems. In *Technology of Breadmaking 2nd Edition*, pp. 207-22, Cauvain, S.P. & Young, L.S., ed. New York, USA: Springer Science+Business Media, LLC.