Hybrid parameter learning using Dirichlet distributions for modelling the dynamics of multi-scale phenomena occurring in food processes

[a]Cédric Baudrit*, [b]Pierre-Henri Wuillemin, and [a]Nathalie Perrot

[a]UMR782 Génie et Microbiologie des Procédés Alimentaires. INRA/AgroParisTech, 78850 Thiverval-Grignon, France, email:{cbaudrit,nperrot}@grignon.inra.fr
[b] Laboratoire d'Informatique de Paris VI (CNRS UMR7606) 75016 Paris, France, email:pierre-henri.wuillemin@lip6.fr

Abstract. In life science, transformation processes are some integrations of multiple interacting bio-physicochemical phenomena that occur at different scales. It is very difficult to elaborate mathematical models representing these processes because knowledge stems from various sources of know-how and is tainted with uncertainty. In this context, the concept of dynamic Bayesian networks (DBNs) provides a unifying practical mathematical formalism that makes it possible to describe complex stochastic dynamical systems. However, the definition of DBN parameters or/and network structure requires substantial knowledge which is rarely the case in the context of food processes. In this paper, we consider the problem of parameter learning for a given network structure designed by experts and we present a hybrid approach for learning probabilistic representations from a combination of empirical observations, human expertise and mechanistic models *etc*.The idea is to be able to estimate parameters knowing that we may have one or several sources of knowledge associated with each parameter. Our approach consists in using the framework of Dirichlet distributions initializing parameters by Dirichlet priors and update them by using Bayesian inference and expected *a posteriori* each time new or additional information is available. The modelling of the cheese ripening process, that is still ill-known and complicated to control, illustrates our approach.

Keywords: Dynamic Bayesian networks, food processing modelling, Dirichlet distribution, parameter learning.

1 Introduction

In life science, transformation processes are some integrations of multiple unit phenomena which result in a complex system. They consist of a large number of interacting microbiological and/or physicochemical components, whose aggregate activities are nonlinear and are responsible for the changes of product properties. Decisions related to the management of such processes rely on predictive models that represent the available knowledge about involved phenomena and are able to simulate the different transient and equilibrium states over time. Unfortunately, a lot of transformation

* Corresponding author

processes are too complex to be modelled in one step [19] and such models remain difficult and complicated to implement [6] because the causal relationships between ingredients, physicochemical, microbiological, organoleptic characteristics, on the one hand, and sensory, nutritional and energy properties, on the other, depending on successive process operations remain still widely ill-known in many transformation process technologies. Knowledge stems from various sources of know-how such as expert operators [1, 17], scientific theory [19, 18], experimental trials *etc*. It is expressed in different forms (equations, expert opinions, databases ...), different formats (numeric, symbolic, linguistic ...) and at different scales (microbiological, physicochemical, organoleptic ...). Moreover, faced with the heterogeneous character of knowledge, information is tainted with stochastic and epistemic uncertainty such as randomness, incompleteness, imprecision, vagueness *etc* [5, 7, 12]. The treatment of uncertainties has become crucial in industrial applications and by consequence in decision-making processes. The main questions is how to formulate and couple human knowledge, data measurements, mathematical models into a unifying framework in order to describe the whole system and provide adequate decision-making tools. With this aim in mind, the concept of dynamic Bayesian networks (DBNs) [15] provides a unifying practical mathematical formalism that makes it possible to describe complex stochastic dynamical systems. They are an extension of Bayesian networks (BNs) [13, 16] and generalise the well-known Hidden Markov Models (HMMs) [14].However one of the limitations of dynamic Bayesian networks lies in the definition of parameters (*i.e.* conditional probabilities tables) or/and network structure that requires substantial knowledge which is rarely the case in the context of food processes. In this paper, we consider the problem of parameter learning for a given network structure designed by experts. This paper presents a hybrid approach for learning probabilistic representations from a combination of empirical observations, human expertise and mechanistic models *etc*. Our approach aims to (1) combine multi-sources of knowledge to identify DBN parameters and (2) update our background knowledge with new or additional information to refine DBN parameters with a view to improve the whole model. It consists in using the framework of Dirichlet distributions initializing parameters by using Dirichlet priors that may be assessed from literature, empirical observations, experts opinions, existing models *etc*. Next, parameters are successively updated by using Bayesian inference and expected *a posteriori* each time new or additional information is available and can be formulate into a frequentist form. To illustrate our approach, we consider the framework of the food transformation processes and more specifically the modelling of the Camembert-type soft mould cheese ripening that is still ill-known and complicated to control [8].In Section 2, we introduce the basic notions about dynamic Bayesian networks. In Section 3, we present our approach of hybrid learning. In Section 4, we illustrate the hybrid parameter learning for modelling a Camembert-type cheese ripening process.We end the paper with some concluding remarks and open problems

2 Dynamic Bayesian Networks

DBNs are classical Bayesian networks in which nodes $\{X_i(t), i = 1 \dots n\}$, representing random variables, are indexed by time t. They provide a compact representation of the joint probability distribution P for a finite time interval $[1,\tau]$ defined as follows:

$$P(X(1),\ldots,X(\tau)) = \prod_{i=1}^{n}\prod_{t=1}^{\tau} P(X_i(t)|\text{Pa}(X_i(t))) \tag{1}$$

where $X(t) = \{X_1(t),\ldots,X_n(t)\}$, is called a "slice" and represents the set of all variables indexed by the same time t. This joint probability $P(X(1),\ldots,X(\tau))$ represents the beliefs about possible trajectories of the dynamic process $X(t)$. DBNs assume the *first-order Markov property* which means that the parents of a variable in time slice t must occur in either slice $t-1$ or t :

$$\text{Pa}(X_i(t)) \subset (X(t-1) \cup X(t)) \setminus X_i(t) \tag{2}$$

Moreover, the conditional probabilities are time-invariant (*first-order homogeneous Markov property*):

$$P(X_i(t)|\text{Pa}(X_i(t))) = P(X_i(2)|\text{Pa}(X_i(2))) \text{ ,} \forall t \in [2,\tau] \tag{3}$$

Hence to specify a DBN, we need to define the intra-slice topology (within a time slice), the inter-slice topology (between two time slices), as well as the parameters (*i.e* conditional probabilities, see Equation 3) for the first two time slices. In our context, the structure of a model have been explicitly built on the basis of expert knowledge.

3 Hybrid parameter learning

Assume that $X_i(t)$ are all categorical variables given values ($x_{i1},\ldots,x_{ic_i}$) where c_i is the number of values that node i can take on and let θ^t_{ijk} be the probability that $X_i(t) = x_j$, given that its parents have instantiation x_k (corresponding itself to a vector where k represents the set of parents of i), *i.e.* $\theta^t_{ijk} = P(X_i(t) = x_j|\text{Pa}(X_i(t)) = x_k)$. As we assume the first-order homogeneous Markov property (see Eq.3), we have: $\forall t \in [2,\tau]$, $\theta^t_{ijk} = \theta_{ijk}$. The idea of hybrid parameter learning is to be able to estimate $\theta_{ijk}|(D_1,\ldots,D_m)$ where $D_1,\ldots,D_m$ may correspond to new or additional information as expert opinions, experimental or simulated data *etc*. Our approach consists in using the framework of Dirichlet distributions [9] initializing variables θ^t_{ik} by Dirichlet priors and update θ_{ijk} by using Bayesian inference and expected *a posteriori* each time new or additional information is available and can be formulated into a frequentist form. The Dirichlet prior is the conjugate prior for the multinomial which permits analytical calculations. The Dirichlet prior distribution, $\theta_{ik} \sim \mathcal{D}(\alpha_{i1k},\ldots,\alpha_{ic_ik})$ is defined by

$$P(\theta_{ik}|\alpha_{ik}) = \frac{\Gamma(\sum_{j=1}^{c_i}\alpha_{ijk})}{\prod_{j=1}^{c_i}\Gamma(\alpha_{ijk})}\prod_{j=1}^{c_i}\theta_{ijk}^{\alpha_{ijk}-1} \tag{4}$$

with $\sum_{j=1}^{c_i}\theta_{ijk} = 1$, $\alpha_{ik} > 0$ and $\Gamma(x) = \int_0^{+\infty} e^t t^{x-1} dt$. If we have an available database D in which event $(X_i(t) = x_j, \text{Pa}(X_i(t)) = x_k)$ occurs N^t_{ijk} times, the posterior variable $\theta_{ik}|D$ follows a Dirichlet distribution:

$$\theta_{ik}|D \sim \mathcal{D}(\alpha_{i1k} + \sum_t N^t_{i1k},\ldots,\alpha_{ic_ik} + \sum_t N^t_{ic_ik}) \tag{5}$$

and the expected *a posteriori* (EAP estimate) gives:

$$\hat{\theta}_{ijk} = \frac{\sum_t N^t_{ijk} + \alpha_{ijk}}{\sum_{j=1}^{c_i}(\sum_t N^t_{ijk} + \alpha_{ijk})} \tag{6}$$

In our case, the values

$$\left(\frac{\alpha_{i1k}}{\sum_{j=1}^{c_i}\alpha_{ijk}}, \ldots, \frac{\alpha_{ic_ik}}{\sum_{j=1}^{c_i}\alpha_{ijk}}\right)$$

represent intuitively the *a priori* probabilities for the values of the variable θ_{ik} based in our past experience.

The quantity $\sum_{j=1}^{c_i}\alpha_{ijk}$ may represent the size of (1) a virtual sampling given by expert or (2) a simulated sampling given by mathematical model where model parameters would be tainted with uncertainty. This size can be thus interpreted as a confidence level on experts or mathematical models compared to the database D. That means that more the size will be important more the belief in expert opinion or mathematical model will be important compared to the database D.

3.1 Assessing Dirichlet priors

Next we give some guidelines for choosing the Dirichlet prior distribution according to the kind of available knowledge in order to initiate hybrid learning.

1. Assessing without information.
   In the case where we have not available knowledge over parameters, we decide to use the Dirichlet Prior
   $$\theta_{ik} \sim \mathcal{D}(l, \ldots, l) \tag{7}$$
   inducing an uniform prior over θ_{ik}, *i.e.* $\hat{\theta}_{ijk} = 1/c_i$ and l corresponds to the importance that we want to allocate to our ignorance state.

2. Assessing by means of expert opinion.
   Assume that expert is capable of providing a probability distribution $(p_{ijk})_{j=1,\ldots,c_i}$ over θ_{ik} according to its experience. We may then use the following Dirichlet prior:
   $$\theta_{ik} \sim \mathcal{D}(sp_{i1k}, \ldots, sp_{ic_ik}) \tag{8}$$
   where s corresponds to the confidence level on experts or the imaginary size of database inducing $\hat{\theta}_{ijk} = p_{ijk}$ as expected values. However, it is often beyond the experts ability to specify a full Dirichlet prior over parameters [11].

3. Assessing by means of an experimental database D_e.
   When we have an experimental database D_e, we may use
   $$\theta_{ik} \sim \mathcal{D}(\sum_t N^t_{i1k}, \ldots, \sum_t N^t_{ic_ik}) \tag{9}$$

as Dirichlet prior inducing a prior probability distribution on the values of θ_{ik} given by $\hat{\theta}_{ijk} = \sum_t N^t_{ijk} / \sum_j \sum_t N^t_{ijk}$ corresponding to the maximum likelihood estimation.

4. Assessing by means of a simulated database D_s.
Certain physical or/and biological phenomena, noted X, can be well-known and modelled by mathematical models $\dot{X} = f_a(X, T)$ providing a richer information than an experimental database where X, a and T represent a vector of state variables, model parameters and control variables respectively. However, certain parameters and/or variables are tainted with uncertainty due to the intrinsic variability of phenomena or/and the lack of knowledge about the precise values of parameters. Under such a situation, the traditional attitude is to represent each and every ill-known parameters or inputs by means of probability distributions and to perform a random sampling by using a Monte-Carlo method [10] in order to estimate the uncertainty about model outputs. In this case, we build a simulated database D_s and we use
$$\theta_{ik} \sim \mathcal{D}(\sum_t M^t_{i1k}, \ldots, \sum_t M^t_{ic_ik}) \tag{10}$$
as Dirichlet prior where M^t_{ijk} corresponds to the number of times $(X_i(t) = x_j, Pa(X_i(t)) = x_k)$ occurs in D_s and $\sum_{j=1}^{c_i} \sum_t M^t_{ijk}$ corresponds to the size of database D_s which may be interpreted as the confidence level on mathematical model. We can image to be able to extend this approach by using and combining other mathematical models as stochastic expert systems and/or neural networks *etc*.

3.2 Incremental Bayesian updating

Based on our initial knowledge about θ_{ik} represented by Dirichlet prior distributions, we attempt to update parameters θ_{ijk} with new or additional information $D_1, \ldots, D_m$ by using Bayesian inference and the expected *a posteriori*. Based on the same reasoning leading to Eq. 5, we have

$$\theta_{ik} | (D_1, \ldots, D_m) \sim \mathcal{D}(\alpha_{ik} + \sum_{p=1}^{m} \sum_t N^t_{ik}(p)) \tag{11}$$

where we assume $D_m \perp (D_1, \ldots D_{m-1}) | \theta_{ik}$ and each D_p is formulated into a frequentist form, that is D_p corresponds to a set of virtual or experimental or calculated data with counts $N^t_{ik}(p)$. To summarize, we obtain

$$\hat{\theta}_{ijk} = \frac{\sum_{p=1}^{m} \sum_t N^t_{ijk}(p) + \alpha_{ijk}}{\sum_{j=1}^{c_i} (\sum_{p=1}^{m} \sum_t N^t_{ijk}(p) + \alpha_{ijk})} \tag{12}$$

where $\sum_t N^t_{ijk}(p)$ may be corresponds to:

1. 0 when we have no additional information.
2. the probability sp_{ijk} given by expert with $\sum_j p_{ijk} = 1$ and s corresponds to the virtual size of database.

3. the counts in an experimental database.
4. the counts in a simulated database.

We illustrate our approach by given two examples.

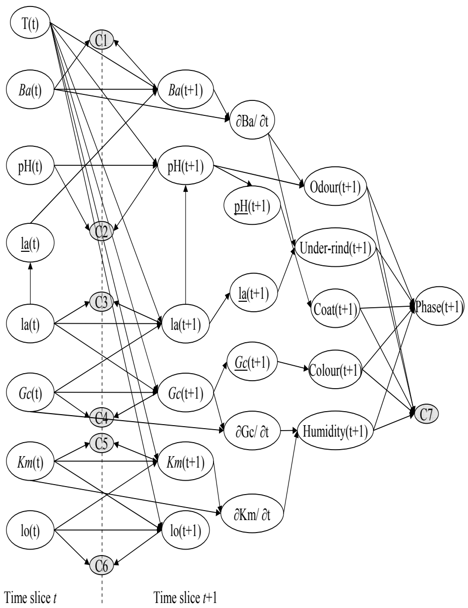

Fig. 1. Dynamic Bayesian network representing the coupled dynamics of micro-organism growth with their substrate consumptions influenced by temperature and involving the sensory changes of cheese during the ripening process. Grey nodes represent expert constraints

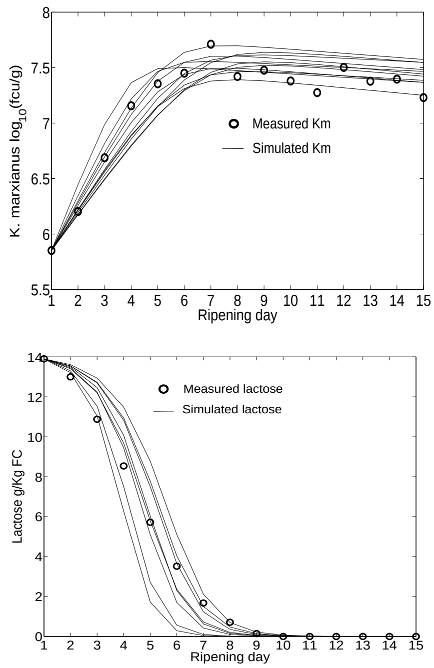

Fig. 2. A random sampling of eleven (resp. eight) possible dynamics of *K.marxianus* (resp. lactose) concentrations resulting form Monte-Carlo simulations at 12^{o}C using the microbial growth model described by Eq. 13.

– Assume that we are faced with none background knowledge about θ_{ik}, we then use a Dirichlet prior distribution $\mathcal{D}(1,\ldots,1)$. If we do not have additional information, Dirichlet prior distribution becomes *a posteriori* which corresponds to a uniform distribution over θ_{ik}.
– Assume that we have a simulated database D_s associated with θ_{ik}, we define a Dirichlet prior distribution $\mathcal{D}(\sum_t M_{i1k}^t,\ldots,\sum_t M_{ic_ik}^t)$ where $\sum_{j=1}^{c_i}\sum_t M_{ijk}^t$ corresponds to the size of D_s. Next, we have two additional information about θ_{ik} namely a probability distribution p_{ik} given by expert and an experimental data D_e. We deduce that $\theta_{ij}|D_e$, "expert" follows the Dirichlet distribution $\mathcal{D}(\sum_t M_{i1k}^t + \sum_t N_{i1k}^t + sp_{i1k},\ldots,\sum_t M_{ic_ik}^t + \sum_t N_{ic_ik}^t + sp_{ic_ik})$ where s corresponds to the

confidence level on expert and $\sum_{j=1}^{c_i} \sum_t N_{ijk}^t$ corresponds to the confidence level on the database D_e.

4 Application to the food process modelling: cheese ripening

To illustrate our approach, we have focused on the modelling of the Camembert-type soft mould cheese ripening that is still ill-known and complicated to control [8]. During the ripening process, cheese represents an ecosystem and a bioreactor where relationships exist between microbiological, physicochemical and organoleptic changes which depend on environmental conditions. From operational and scientific knowledge, baudrit et *al.* [2] defined the structure of a dynamic Bayesian network provid-

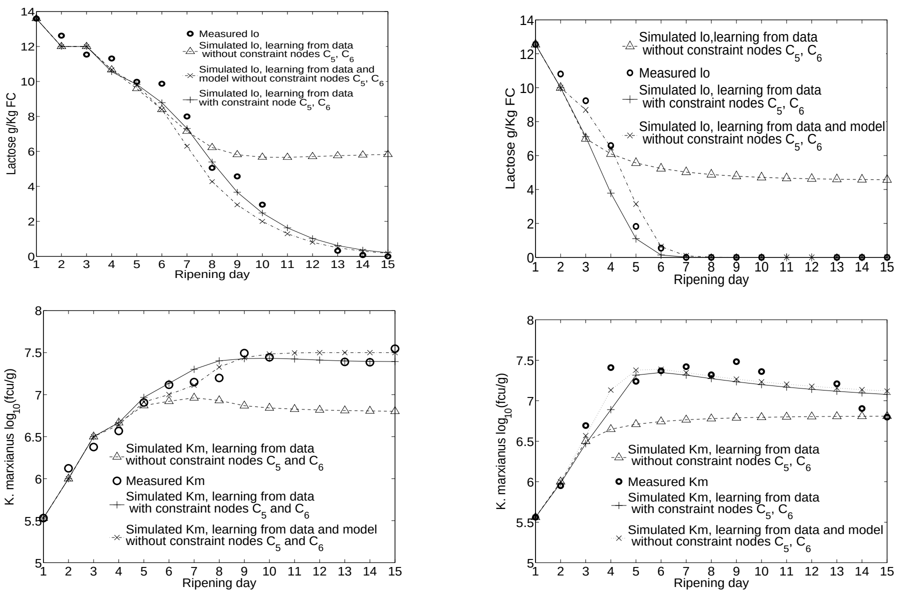

Fig. 3. Measured lactose and *K.marxianus* concentrations versus the mean DBN simulation results according to available knowledge for learning and the taking or not into account constraint nodes (C_5, C_6) for a ripening carried out at 8^{o}C (at left) and 16^{o}C (at right)

ing 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 (see Fig.1). Nodes $C_1, \ldots, C_7$ correspond to constraint nodes which are conditioned to be true (*e.g.* $P(C =' true'|X(t+1) = i, X(t) = j) = 0$ if $j > i$ and 1 otherwise) [4] and allow to represent rules, generally linked to the physical or/and biological conservation laws.In a first time, we use a Dirichlet prior $\mathcal{D}(1)$ for all parameters which are updated from an experimental database D_e filled in from six cheese ripening experimental trials carried out for temperatures varying from T = 8 to 16 ^{o}C. Parameters $P(Km(t+1)|(Km(t), lo(t), T(t)))$ and $P(lo(t+1)|(Km(t), lo(t), T(t)))$ are then updated regardless of the rest of network by using a simulated database D_s (see section 3.2) obtained by Monte-Carlo simulation [10] (see Fig.2) on the following microbial growth model [18]

$$(S)\begin{cases}\dfrac{dKm}{dt} = \mu\dfrac{lo}{K_{lo}+lo}Km - bKm\\ \dfrac{dlo}{dt} = -\dfrac{\mu}{\beta}\dfrac{lo}{K_{lo}+lo}Km\end{cases} \tag{13}$$

where μ (the maximum specific growth rate of Km), K_{lo} (the half saturation constant for growth), b (the decay coefficient) and β (the yield coefficient for Km on lactose) are tainted with uncertainty due to the intrinsic variability of process. Figure 3 display

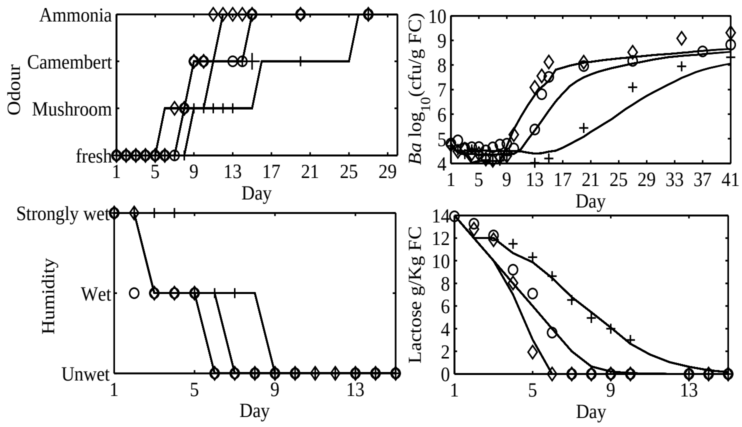

Fig. 4. Predictive mean evolution of *Ba*, la, Odour and Humidity obtained from DBN (in straight line) versus raw data for three different ripening carried out at 8^{o}C (noted +), 12^{o}C (noted $\circ$) and 16^{o}C (noted $\diamond$).

the mean evolutions of simulated lo, Km versus raw data for two ripening carried out at 8^{o}C and 16^{o}C according to (1) learning from database D_e without constraint nodes (C_5, C_6); (2) learning from database D_e with constraint nodes and (3) learning from database D_e and D_s without constraint nodes (C_5, C_6). We see that the two last results are closed to raw data contrary to results obtained without constraint nodes. We can conclude that the coupling of experimental and simulated data allowed to simplify the network structure and to enrich our model. Indeed, although conditioning is a method that comes naturally out of the use of directed graphical models, their introduction to the network may produce side effects which can distort the resulting distribution [4] and makes more complex the model. Figure 4 displays, by way of example, the average evolution of four variables (Ba, la, Odour, Humidity) versus raw data for three ripening carried out at at 8^{o}C, 12^{o}C and 16^{o}C (not available in learning database) and enables to highlight the predictive character of model.

5 conclusion

This paper presents a hybrid approach for learning probabilistic representations from a combination of empirical observations, human expertise and mechanistic models in the context of the food transformation processes. With this hybrid learning, we enriched and refined the previously model established in [2] where we only had single sources of knowledge for different set of nodes. Moreover, the integration of a mechanistic model allowed to replace an expert information (represented by constraint nodes) and resulted in a simplification of the network structure. We hope to extend our approach by integrating incompleteness and imprecision. Indeed, The formalism of DBNs often does not allow us to take epistemic uncertainty inherent to food processes into account in a coherent and relevant way. Faced with partial ignorance, the use of a single probability measure may introduce information that is not in fact available as the uniformity when we use uniform distribution faced a lack of knowledge. For this purpose, imprecise Dirichlet model (initiated by Walley [20]) could be an interesting extension which consists in describing prior uncertainty by a set of Dirichlet priors instead of a single one. To go even further in our approach, studies focusing on credal networks [3] could be an other interesting area of research since they generalize Bayesian networks by allowing each variable to be associated with sets of joint probability measures rather than single probability measures and can be regarded as sets of Bayesian Networks. This tool seems to be a natural extension for dynamic Bayesian networks in order to integrate our partial ignorance in the reconstruction of dynamics and the management of uncertainty.

References

1. Allais, I. and Edoura-Gaena, R. B. and Gros, J. B. and Trystram, G. How human expertise at industrial scale and experiments can be combined to improve food process knowledge and control. *Food Res. Int.* 40, 585-602, 2007.
2. Baudrit, C., Sicard, M., Wuillemin, P.H., Perrot, N. Towards a global modelling of the Camembert-type cheese ripening process by coupling heterogeneous knowledge with dynamic Bayesian networks. *Journal of Food Engineering*, 98(3), 283-293, 2010.
3. Cozman, F. G. Credal networks. *Artificial Intelligence Journal*, vol. 120, 199-233, 2000.
4. Crowley, M., Boerlage, B., and Poole, D. Adding Local Constraints to Bayesian Networks. *In Proceedings of the 20th Conference of the Canadian Society For Computational Studies of intelligence on Advances in Artificial intelligence*. Z. Kobti and D. Wu, Eds. Lecture Notes In Artificial Intelligence, vol. 4509. Springer-Verlag, Berlin, Heidelberg, 344-355, 2007.
5. Dubois, D. Uncertainty theories: a unified view. *IEEE Cybernetic Systems Conference*, Dublin, Ireland, Invited Paper: 4-9, 2007.
6. Georgakis, C. Modern tools of process control: the case of black, gray and white models. *Entropie*, 194, 34-48, 1995.
7. Ferson, S., Ginzburg, L.R. Different methods are needed to propagate ignorance and variability. Reliability Engineering and Systems Safety, 54, 133-144, 1996.

8. Fox, P.F. *Cheese: Chemistry, Physics and Microbiology*. Academic Press Inc, 2004.
9. Geiger, D., Heckerman, D. A characterization of the dirichlet distribution through global and local parameter independence. *The Annals of Statistics*, 25, 13441369, 1997.
10. Gentle, J.E. Monte-Carlo methods. In S. Kotz and N.L. Johnson eds, editors, *Encyclopedia of Statistics*, 5, 612-617. John Wiley and Sons, New York, 1985.
11. O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, R., Garthwaite, P., Jenkinson, D., Oakley, J., and Rakow, T. *Uncertain Judgements: Eliciting Experts Probabilities*. John Wiley & Sons, Inc, 2006.
12. Helton, J.C., J.D. Johnson, and W.L. Oberkampf. 2004. An Exploration of Alternative Approaches to the Representation of Uncertainty in Model Predictions. *Reliability Engineering and System Safety*, 85 (1-3), 39-71, 2004.
13. Finn V. Jensen and Thomas D. Nielsen *Bayesian Networks and Decision Graphs*, Springer-Verlag, 2007.
14. Lawrence R. Rabiner. A tutorial on Hidden Markov Models and selected applications in speech recognition. *Proceedings of the IEEE*, 77 (2), 257-286, 1989.
15. Murphy, K.P. *Dynamic Bayesian Networks: Representation, Inference and learning*. Ph.D. thesis, University of California, Berkeley, 2002.
16. Pearl, J. *Probabilistic Reasoning in Intelligent systems: Networks of Plausible Inference*. Morgan Kaufmann, San Diego, 1988.
17. Perrot, N., Agioux, L., Ioannou, I., Mauris, G., Corrieu, G., Trystram, G. Decision support system design using the operator skill to control cheese ripening-application of the fuzzy symbolic approach. *Journal of Food Engineering*, 64 (3), 321-333, 2004.
18. Sablani SS, Datta AK, Rahman MS,Mujumdar AS. *Handbook of food and bioprocess modeling techniques* . New York: Taylor and Francis, 2006.
19. Tijskens, L.M.M.; Hertog, M.L.A.T.M.; Nicola, B.M. *Food Process Modelling*. Woodhead Publishing, 2001.
20. Walley, P. Inferences from multinomial data: Learning about a bag of marbles. *J.R.Statist. Soc. B*, 58(1), 357, 1996.