



**HAL**  
open science

# Unifying parameter learning and modelling complex systems with epistemic uncertainty using probability interval

C. Baudrit, S. Destercke, Pierre-Henri Wuillemin

► **To cite this version:**

C. Baudrit, S. Destercke, Pierre-Henri Wuillemin. Unifying parameter learning and modelling complex systems with epistemic uncertainty using probability interval. *Information Sciences*, 2016, 367-368, pp.630-647. 10.1016/j.ins.2016.07.003 . hal-01346202

**HAL Id: hal-01346202**

**<https://hal.sorbonne-universite.fr/hal-01346202>**

Submitted on 18 Jul 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

14 Unifying parameter learning and modelling complex  
 15 systems with epistemic uncertainty using probability  
 16 interval

17 C. Baudrit<sup>a,\*</sup>, S. Destercke<sup>b</sup>, P.H. Wuillemin<sup>c</sup>

18 <sup>a</sup>INRA, I2M, USC 1368, F-33400 Talence, France.

19 <sup>b</sup>UTC, HEUDIASYC, UMR6599, F-60205 Compiègne cedex, France

20 <sup>c</sup>Sorbonne Universites, UPMC, Univ Paris 06, UMR 7606, LIP6, Paris, France

21 CNRS, UMR 7606, CNRS, Paris, France

---

22 **Abstract**

Modeling complex dynamical systems from heterogeneous pieces of knowledge varying in precision and reliability is a challenging task. We propose the combination of dynamical Bayesian networks and of imprecise probabilities to solve it. In order to limit the computational burden and to make interpretation easier, we also propose to encode pieces of (numerical) knowledge as probability intervals, which are then used in an imprecise Dirichlet model to update our knowledge. The idea is to obtain a model flexible enough so that it can easily cope with different uncertainties (i.e., stochastic and epistemic), integrate new pieces of knowledge as they arrive and be of limited computational complexity.

23 *Keywords:* Dynamic credal networks, imprecise probability, Dirichlet  
 24 model, knowledge integration, uncertainty, modelling.

---

25 **1. Introduction**

26 Firms and industrials of all sectors have to face up new challenging situ-  
 27 ations. On the one hand, citizens as well as public authorities have stronger  
 28 demands in terms of quality, safety, ... and on the other hand, they must  
 29 adapt to the increase of population, global warming and the depletion of  
 30 fossil resources. This means, among other things, that industrial projects  
 31 have to integrate sustainability from local to world scale in their conception.  
 32 Possessing adequate tools to model their systems is likely to make the task  
 33 easier.

34 In order to provide relevant conclusions and recommendations, such tools  
 35 should be able to integrate as much available knowledge as possible, how-  
 36 ever heterogeneous it is, both in terms of nature (e.g., qualitative expert  
 37 knowledge vs statistical data) and quality (different precision or degrees of  
 38 reliability). Such systems are also complex, meaning that the modeling tool

---

\*Corresponding author

Preprint submitted to *Electric Power Systems Research* on July 4, 2016  
 Email addresses: cbaudrit@u-bordeaux.fr (C. Baudrit),  
 sebastien.destercke@hds.utc.fr (S. Destercke), pierre-henri.wuillemin@lip6.fr  
 (P.H. Wuillemin)

39 must be able to cope with different scales (e.g., molecular to macroscopic)  
 40 and with dynamic, time-varying processes. Current researches rely on the  
 41 development of mathematical tools [53, 6] capable of helping decision-makers  
 42 to deal with uncertainties, linked for instance to meteorological variations,  
 43 to expert reliability, *etc.* To summarize, ideal modeling tools should be able  
 44 to deal with:

- 45 • heterogeneous sources of knowledge (Web, data warehouse, experts,  
 46 ...)
- 47 • mathematical formalisms used by different disciplines (differential equa-  
 48 tions, graphs, cognitive maps, ...)
- 49 • various manipulated scales (molecular, cellular, population, ...)
- 50 • different forms of uncertainty [32, 36, 40] (natural randomness, impre-  
 51 cision in expert opinions, data scarcity, vagueness, ...)

52 In this paper, we propose dynamic credal networks as a possible answer  
 53 to these challenging tasks to describe complex dynamical systems tainted  
 54 with stochastic and epistemic uncertainty. As an extension of dynamic  
 55 Bayesian networks (DBNs) [51], their network structure provides an intu-  
 56 itively appealing interface for human experts to model highly-interacting sets  
 57 of variables, resulting in a qualitative representation of knowledge. Stochas-  
 58 tic and epistemic uncertainties pertaining to the system are then taken into  
 59 account by quantifying dependence between variables by means of convex  
 60 sets of conditional probability distributions. The concept of DCNs makes it  
 61 possible to combine different sources of information, from qualitative expert  
 62 knowledge to experimental data.

63 In this paper, we are specifically interested in the problem of param-  
 64 eter learning for a given network structure (assumed to be known), when  
 65 faced with heterogeneous knowledge. Indeed, while DCN are very attractive  
 66 modeling tools, they also come with a number of challenges, such as how to  
 67 control their computational tractability, or how to combine efficiently and  
 68 easily various pieces of information. For example, how to combine simu-  
 69 lations coming from stochastic differential equations with an experimental  
 70 database, both offering information for the same parameters? We propose  
 71 to use an imprecise Dirichlet model [7] as a model of the conditional prob-  
 72 abilities, and probability intervals as a common uncertainty model to treat  
 73 different pieces of knowledge. Once transformed, these information pieces  
 74 gradually increment the set of prior distributions according to the received  
 75 knowledge, using the Generalized Bayes rule each time additional informa-  
 76 tion arrives. Lower and upper expected *a posteriori* (EAP) are then used as  
 77 probability bounds to draw inferences from the network. The combination  
 78 of information is done through a weighted average, allowing us to weigh the  
 79 importance of the different sources of knowledge.

80 Section 2 details the material regarding imprecise probabilities as well  
 81 as the proposed updating scheme of a given parameter set. We then de-  
 82 scribe in Section 3 how various common sources of information can be  
 83 transformed into probability intervals. Section 4 presents how we extend  
 84 Dynamic Bayesian Networks to sets of conditional probabilities, while Sec-  
 85 tion 5 illustrates the whole approach on a real-case scenario involving cheese  
 86 ripening.

## 87 2. Imprecise probabilities and Dirichlet model

88 Let  $X$  be a variable<sup>1</sup> taking its values on the finite set  $\mathcal{X} = \{x_1, \dots, x_n\}$ ,  
 89 and  $p : \mathcal{X} \mapsto [0, 1]$ ,  $\sum_{x \in \mathcal{X}} p(x) = 1$  be a probability mass function over  $\mathcal{X}$ .  
 90  $p(X)$  will denote the vector mass function, while  $p(x)$  will denote the value  
 91 taken by  $p$  for  $X = x$ . Such a mass function defines a measure  $P_X(A) =$   
 92  $\sum_{x \in A} p(x)$  for all  $A \subseteq \mathcal{X}$ .

### 93 2.1. Imprecise probability and credal sets

94 In general, identifying a single probability modelling our uncertainty  
 95 about some variable  $X$  requires a lot of data and/or knowledge. When such  
 96 knowledge is not available, a safer option is to model our uncertainty by  
 97 convex sets of probabilities, often called *credal sets* [47, 61, 2]. A credal  
 98 set associated with  $X$ , denoted  $K(X)$ , is a convex set of probability masses  
 99 over  $\mathcal{X}$ .  $K(X)$  represents the uncertainty about the unknown value of the  
 100 variable  $X$ . From  $K(X)$  are defined upper and lower probability measures  
 101 of an event  $A \subseteq \mathcal{X}$  as

$$\overline{P}_X(A) = \sup_{p \in K(X)} \sum_{x \in A} p(x), \quad \underline{P}_X(A) = \inf_{p \in K(X)} \sum_{x \in A} p(x). \quad (1)$$

102 and, in particular, for any element  $x \in \mathcal{X}$  we will have that the upper and  
 103 lower probabilities are given by

$$\overline{p}(x) = \sup_{p \in K(X)} p(x), \quad (2)$$

$$\underline{p}(x) = \inf_{p \in K(X)} p(x) \quad (3)$$

104 In a subjectivist tradition, the lower probability  $\underline{P}_X(A)$  can be interpreted  
 105 as the maximal price one would be willing to pay for the gamble which pays  
 106 1 unit if event  $A$  occurs (and nothing otherwise) [61].  $\underline{P}_X(A)$  is therefore a  
 107 measure of evidence in favour of event  $A$ , or in other words how much  $K(X)$   
 108 supports event  $A$ , while  $\overline{P}_X(A)$  measures the lack of evidence against  $A$ .  
 109  $K(X)$  can also be given a robust interpretation, in which it models imperfect

<sup>1</sup>We adopt notations similar to those of [2, Ch.9] and [24].

110 knowledge of a precise, possibly frequentist, probability  $p$ . A credal set  
 111  $K(X)$  contains a set  $\mathcal{Ext}(K(X))$  of extreme probability masses, always finite  
 112 in this paper, corresponding to the vertices of  $K(X)$ . Geometrically,  $K(X)$   
 113 may be equivalently specified by the convex hull (denoted  $CH$ ) of the set  
 114  $\mathcal{Ext}(K(X))$ , i.e.

$$K(X) = CH\{\mathcal{Ext}(K(X))\}. \quad (4)$$

115 The *vacuous* credal set

$$K_v(X) = \{p(X) : p(x) \geq 0, \forall x \in \mathcal{X}, \sum_{x \in \mathcal{X}} p(x) = 1\} \quad (5)$$

116 that includes all probability masses over  $\mathcal{X}$  plays an important role, as it  
 117 models total ignorance, and should be the starting point of any model. We  
 118 refer to Walley [61, Sec. 5.5.] for a discussion about uniform probability  
 119 distribution not being a good model of ignorance.

120 In this paper, we will also be especially interested in particular credal  
 121 sets  $K(X)$  specified by means of *interval probability*

$$K(X) = \{p(X) : p(x) \in [l_x, u_x], 0 \leq l_x \leq u_x \leq 1, \sum_{x \in \mathcal{X}} p(x) = 1\}. \quad (6)$$

122 Indeed, such credal sets that focus over bounds of singletons have the advan-  
 123 tage to be easier to manipulate, simulate and represent than general ones,  
 124 while remaining expressive enough (they include both the vacuous and the  
 125 precise models). We refer to De Campos *et al.* [11] for a detailed exposition,  
 126 and will only limit ourselves to necessary elements in this paper.

*Example 1.* Consider an example with three possibilities  $\mathcal{X} = \{x_1, x_2, x_3\}$   
 (e.g., the working states of a system such as "failing", "degraded function-  
 ing", "fully functioning"), and assume that previous experiments result in  
 the following intervals

$$p(x_1) = [0; 0.2], \quad p(x_2) = [0.3; 0.4], \quad p(x_3) = [0.4; 0.6].$$

The credal set  $K(X)$  is the set of all precise probabilities  $P(X) = (p(x_1), p(x_2), p(x_3))$   
 within these interval bounds. Here  $K(X)$  is a polytope defined by the convex  
 hull of its four vertices in a three dimensional space:

$$K(X) = CH\{(0, 0.4, 0.6); (0.2, 0.3, 0.5); (0.2, 0.4, 0.4); (0.1, 0.3, 0.6)\}.$$

127 Finding these vertices can be done by using classical tools of convex ge-  
 128 ometry [39], or by using algorithms proper to a given representation (an  
 129 Algorithm is provided by De Campos *et al.* [11]). The set  $K(X)$  is repre-  
 130 sented in Figure 1 in barycentric coordinates.

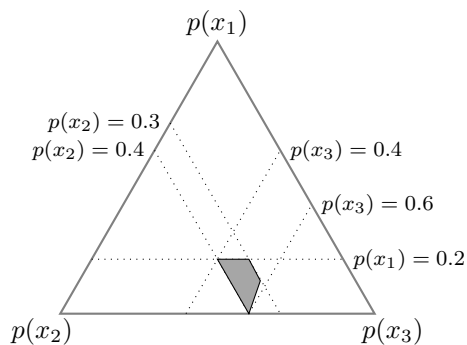


Figure 1: Example 1 credal set in Barycentric coordinates.

## 131 2.2. Robust Dirichlet model to learn $K(X)$

132 An important question is how the credal set  $K(X)$  can be instantiated  
 133 from actual evidence, or in other words how can we go from an initially  
 134 vacuous knowledge towards a more precise state of knowledge. An in-  
 135 strumental tool to do that is to use a robustified version of the Dirichlet  
 136 model, also commonly referred to as the Imprecise Dirichlet Model (IDM)  
 137 [62, 7, 8, 60]. The basic model is based on two hyper-parameters: a positive  
 138 real value  $s_0$  associated to the strength of prior knowledge, and a vector  
 139  $\epsilon_0 = (\epsilon_0(x_1), \dots, \epsilon_0(x_n))$  associated to our initial beliefs about the probabili-  
 140 ties of occurrence of elements  $x_i$ .

141 Let  $\theta = (\theta_1, \dots, \theta_n)$  be a vector of chances such that  $\theta_i$  corresponds to  
 142 the chance that  $X = x_i$ . The prior distribution of vectors  $\theta$  given by a  
 143 Dirichlet model is then

$$\text{Dir}(s_0; \xi_0)(\theta) = \frac{\Gamma(s_0)}{\prod_{i=1}^n \Gamma(s_0 \xi_0(x_i))} \prod_{i=1}^n \theta_i^{s_0 \xi_0(x_i) - 1} \quad (7)$$

144 where  $\Gamma$  is the gamma function. A very easy way to make this model imprecise  
 145 is to let the vector  $\epsilon_0$  become imprecise, and more precisely to consider  
 146 the set of Dirichlet models

$$\mathcal{M}_{(s_0; \xi_0)} = \{\text{Dir}(s_0; \xi_0)(\theta) : \xi_0 \in \mathcal{T}\} \quad (8)$$

147 with

$$\mathcal{T} = \{\xi_0 : 0 < \xi_0(x_i) < 1, \sum_{i=1}^n \xi_0(x_i) = 1\} \quad (9)$$

148 the open  $(n - 1)$ -dimensional unit simplex. When  $\xi_0$  is precise, the first  
 149 moments of  $\text{Dir}(s_0; \xi_0)$  are given by  $E(\theta_i | (s_0; \xi_0)) = \xi_0(x_i)$ , and they can be  
 150 used as estimates of  $p(x_i)$ , *i.e.*

$$E(\theta_i | (s_0; \xi_0)) = \xi_0(x_i) = p(x_i). \quad (10)$$

151 When starting from a vacuous prior knowledge  $\xi_0 \in \mathcal{T}$ , the bounds over the  
152 first moments become

$$\underline{E}(\theta_i | (s_0; \xi_0)) = \min_{\xi_0 \in \mathcal{T}} \xi_0(x_i) = 0 \quad (11)$$

153 and

$$\bar{E}(\theta_i | (s_0; \xi_0)) = \max_{\xi_0 \in \mathcal{T}} \xi_0(x_i) = 1. \quad (12)$$

154 The credal set corresponding to these bounds is then the vacuous one (5).

155 We may then receive additional information from various  $m$  sources. A  
156 convenient way to encode this information is as a couple  $s_k, \mathcal{P}_k, k = 1, \dots, m$ ,  
157 with  $\mathcal{P}_k \subseteq \mathcal{T}$  a convex polytope providing information about the possible  
158 chances  $\theta_i$ , and  $s_k \in \mathbb{R}^+$  modelling the strength of the information. We can  
159 then update the Dirichlet modelling our uncertainty about  $\theta | (s_k; \xi_k)_{k=0}^m$  into

$$\mathcal{M}_{(s_k; \xi_k)_{k=0}^m} = \{\text{Dir}((s_k; \xi_k)_{k=0}^m)(\theta) : \xi_k \in \mathcal{P}_k \forall k\}. \quad (13)$$

160 We can then use the posterior first moments to make inferences on chances  
161  $\theta_i$

$$E(\theta_i | (s_k; \xi_k)_{k=0}^m) = p(x_i) = \frac{\sum_{k=0}^m s_k \xi_k(x_i)}{\sum_{k=0}^m s_k} \quad (14)$$

162 As information  $\mathcal{P}_k$  are imprecise, we again obtain bounds in the form

$$\underline{E}(\theta_i | (s_k; \xi_k)_{k=0}^m) = \underline{p}(x_i) = \frac{\sum_{k=0}^m s_k \underline{\xi}_k(x_i)}{\sum_{k=0}^m s_k}, \quad (15)$$

$$\bar{E}(\theta_i | (s_k; \xi_k)_{k=0}^m) = \bar{p}(x_i) = \frac{\sum_{k=0}^m s_k \bar{\xi}_k(x_i)}{\sum_{k=0}^m s_k}. \quad (16)$$

163 where

$$\underline{\xi}_k(x_i) = \inf_{\xi_k \in \mathcal{P}_k} \xi_k(x_i) \quad (17)$$

$$\bar{\xi}_k(x_i) = \sup_{\xi_k \in \mathcal{P}_k} \xi_k(x_i). \quad (18)$$

164 These bounds then induce an updated credal set

$$K_{(s_k; \xi_k)_{k=0}^m}(X) = \left\{ p : p(x_i) \in \left[ \frac{\sum_{k=0}^m s_k \underline{\xi}_k(x_i)}{\sum_{k=0}^m s_k}, \frac{\sum_{k=0}^m s_k \bar{\xi}_k(x_i)}{\sum_{k=0}^m s_k} \right] \right\} \quad (19)$$

165 that we can use as new knowledge. In practice,  $s_0$  can be interpreted as  
166 the number of "unseen" data, and  $s_k = s_0$  means that the  $k$ th information  
167 source has as much importance as our initial uncertainty.

168 *Remark 1.* The exact updated credal set

$$\tilde{K}_{(s_k; \xi_k)_{k=0}^m}(X) = \left\{ \frac{\sum_{k=0}^m s_k \xi_k}{\sum_{k=0}^m s_k} : \xi_k \in \mathcal{P}_k, \forall k = 1, \dots, m \right\} \quad (20)$$

169 is a subset of  $K_{(s_k; \xi_k)_{k=0}^m}(X)$ , i.e.,  $\tilde{K}_{(s_k; \xi_k)_{k=0}^m}(X) \subseteq K_{(s_k; \xi_k)_{k=0}^m}(X)$ . The set  
 170 (19) is thus an outer-approximation. Yet, the main advantages of using  
 171 probability bounds as a basic representation are that

- 172 • their number of extreme points is bounded and relatively low, even  
 173 when combining them through a weighted average. This is in general  
 174 not the case if we consider averaging of heterogeneous simple repre-  
 175 sentations: if we denote  $|\mathcal{Ext}(\mathcal{P}_k)|$  the number of extreme points of  
 176 the  $k$ th item of information, then their (Minkowsky) sum  $\sum_{k=0}^m s_k \mathcal{P}_k$   
 177 may have as much as  $\prod_{k=0}^m |\mathcal{Ext}(\mathcal{P}_k)|$  extreme points, an exponentially  
 178 growing number;
- 179 • they are easy to explain and to represent graphically (e.g., as imprecise  
 180 histograms), therefore offering a convenient way to communicate with  
 181 domain experts or users not specialized in mathematics or computer  
 182 science. This is not the case of more complex representations such as  
 183 belief functions (Section 3.4);
- 184 • except for requiring a finite space, they do not require specific assump-  
 185 tions, such as the existence of an ordering between elements;
- 186 • they are expressive enough so that they can go from a fully precise  
 187 probability to the complete ignorance model.

188 None of the other common practical models of information reviewed in Sec-  
 189 tion 3 have all these advantages at once, making probability bounds a quite  
 190 convenient model. Given this, using probability bounds seem a good general  
 191 starting point in applications, not preventing one from investigating refined  
 192 solutions if the results are unsatisfactory.

193 Of course, in some cases  $K_{(s_k; \xi_k)_{k=0}^m}(X)$  may be a poor outer-approximation,  
 194 however we shall see in Section 5 that it does not necessarily lead to com-  
 195 pletely void conclusions. Previous studies [1] also suggest that this kind of  
 196 approximation may be in average reasonable.

197 *Example 2.* Let  $\mathcal{X} = \{x_1, x_2, x_3\}$  and  $\mathcal{P} = \{\xi : \xi(x_2) \geq \xi(x_1), 1/2 \geq$   
 198  $\xi(x_3), \sum_{i=1}^3 \xi(x_i) = 1\}$  be an item of information. The credal set  $\tilde{K}_\xi(X)$   
 199 over  $\{x_1, x_2, x_3\}$  obtained by (20) has four extreme points  $\{(0, 1, 0), (0, 0.5, 0.5),$   
 200  $(0.5, 0.5, 0), (0.25, 0.25, 0.5)\}$  that are also extreme points of  $K_\xi(X) = \{p :$   
 201  $p(x_i) \leq \max_{\epsilon \in \mathcal{P}} \epsilon(x_i)\}$ . However, the probability  $(0.5, 0.25, 0.25)$  is an ex-  
 202 treme point of  $K_\xi(X)$  but not of  $\tilde{K}_\xi(X)$ .



203 **3. Review of practical sources of information**

204 Building  $K_{(s_k; \xi_k)_{k=0}^m}(X)$  requires to obtain elements of information  $\mathcal{P}_k$ .  
 205 In this section, we review different practical models and the bounds they  
 206 induce over  $\xi_k(x_i)$ . We will also provide small examples illustrating what  
 207 kind of information they can model. For the sake of brevity, we will denote  
 208  $\xi(x_i)$  by  $\xi^i$  in this section. Note that our information is initially queried on  
 209 observed values  $x_i$ , to be then transferred as knowledge on the parameters  $\xi^i$ .  
 210 Hence we will consistently refer to knowledge about  $\xi^i$ , and to observation  
 211 or information about  $x_i$ .

212 *3.1. Precise evaluations*

213 The most simple models is when the knowledge  $\mathcal{P}$  is given by a precise  
 214 vector, in which case  $\xi^i = f_i$  is a precise number, and we have

$$\bar{\xi}^i = \underline{\xi}^i = f_i \quad (21)$$

215 A classical way to obtain such precise evaluations is when observing  $\mathbf{m} =$   
 216  $(m_1, \dots, m_n)$  experiments, where  $m_i$  is the number of times  $x_i$  was observed.  
 217 In such a case, a classical choice is to take as strength  $s = m = \sum_{i=1}^n m_i$   
 218 and  $\mathcal{P}$  is the vector  $\mathbf{f} = (f_1, \dots, f_n)$  where  $f_i = m_i/m$ .

219 *Example 3.* Assume we can observe three possibilities  $x_1, x_2, x_3$  (e.g., sever-  
 220 ity of a disease, importance of a bacterial population), and we observed 3  
 221 times  $x_1$ , 6 times  $x_2$  and one time  $x_3$ . We then have

$$f_1 = 0.3, f_2 = 0.6, f_3 = 0.1 \text{ and } s = m = 10 \quad (22)$$

222 Note that the more observations we accumulate, the stronger becomes  
 223 this piece of knowledge.  $s$  can also be modulated to reflect the reliability  
 224 of data. Note that this model is a degenerated case of probability intervals,  
 225 and can therefore be exactly represented in our framework.

226 *3.2. Numerical possibility distributions and fuzzy subsets*

227 A possibility distribution  $\pi$  is simply a mapping from  $\{\xi^1, \dots, \xi^n\}$  to  
 228  $[0, 1]$ , with at least one element  $\xi^i$  such that  $\pi(\xi^i) = 1$  [33]. In practice,  
 229 we can see distribution  $\pi$  as an ordering  $1 = \pi(\xi^{(1)}) \geq \dots \geq \pi(\xi^{(n)})$  of the  
 230 elements  $x_1, \dots, x_n$ , from the most plausible to the least plausible one.

231 Another instrumental way is to encode the possibility distribution through  
 232 the necessity measure  $N$ . This necessity measure  $N$  is such that

$$N(A_{(i)} = \{\xi^{(1)}, \dots, \xi^{(i)}\}) = 1 - \pi(\xi^{(i+1)}) \quad (23)$$

233 with  $\pi(\xi^{(n+1)}) = 0$ .  $N(A_{(i)})$  can be associated to a lower probability bound  
 234 of event  $A_{(i)}$ . In particular, the sets  $A_{(i)}$  can be interpreted as nested sets  
 235 with an associated lower confidence, these nested sets being built by starting

236 from the most plausible element  $\xi^{(1)}$  and incrementally including the less  
 237 plausible ones. Note that we may have  $\pi(\xi^{(i)}) = \pi(\xi^{(i+1)})$ , in which case  
 238 elements  $x_{(i-1)}$  and  $x_{(i)}$  would be in the same confidence set.

239 In practice, an expert can provide a possibilistic information by giving  
 240 confidence bounds over a collection of nested sets. Let  $A_1 \subseteq \dots \subseteq A_m$  be  
 241 such sets with associated confidence levels  $\alpha_1 \leq \dots \leq \alpha_m$ , then it encodes  
 242 the knowledge  $\mathcal{P}_\pi = \{\xi : \sum_{k=1}^i \xi^{(k)} \geq \alpha_i, \forall i\}$ . From the knowledge on sets  
 243  $A_1, \dots, A_m$ , one can always come back to an associated distribution  $\pi$  using

$$\pi(\xi^k) = \min_{i: \xi^k \in A_i} 1 - \alpha_{i-1} \quad (24)$$

244 with  $\alpha_0 = 0$ .

245 Another possibility is to use the formal equivalence between a possibility  
 246 distribution  $\pi$  and a fuzzy set having  $\pi$  for membership function. This means  
 247 that an expert conveying information in the form of linguistic assessment [64]  
 248 can also be modelled by possibility distributions. Deriving bounds on  $\xi^i$  from  
 249  $\mathcal{P}_\pi$  using the possibility distribution  $\pi$  is very easy, as

$$\underline{\xi}^i = 1 - \max_{\xi \neq \xi^i} \pi(\xi) \quad (25)$$

250

$$\bar{\xi}^i = \pi(\xi^i) \quad (26)$$

251 *Example 4.* Assume that an expert is interrogated about the temperature  
 252 in a room that can be in three states  $x_1, x_2, x_3$ . Expert judges that  $x_2$  is  
 253 the most plausible state, then  $x_3$  and  $x_1$ , meaning that  $\xi^{(1)} = \xi^2, \xi^{(2)} =$   
 $\xi^3, \xi^{(3)} = \xi^1$ . The expert provides the following confidence values:

$$N(\{\xi^2\}) = 0.5$$

$$N(\{\xi^2, \xi^3\}) = 0.8$$

$$N(\{\xi^2, \xi^3, \xi^1\}) = 1$$

254 which means that the expert has a confidence 0.5 that  $x_2$  will be the observed  
 255 state, a confidence 0.8 that the observed state will be either  $x_2$  or  $x_3$ , and  
 256 finally is certain that the only observable states are  $x_1, x_2, x_3$ . From these  
 257 values can be deduced the values of the corresponding possibility distribution  
 258  $\pi(\xi^1) = 0.2, \pi(\xi^2) = 1, \pi(\xi^3) = 0.5$ .

259 Alone, possibility distributions will often be simpler than probability  
 260 intervals: they require less information (one value per element) and will  
 261 have a maximal number of  $2^{|\mathcal{X}|-1}$  extreme points [56]. Yet the average of  
 262 multiple sets  $\mathcal{P}_{\pi_1}, \dots, \mathcal{P}_{\pi_m}$  would no longer be a possibility distribution, and  
 263 the corresponding number of extreme points could explode. Also, possibility  
 264 distributions cannot model precise probabilities, unless they are degenerate  
 265 ones.

266 *3.3. Probability boxes and clouds*

267 A probability box [35]  $\underline{F}, \overline{F}$  is an imprecise cumulative distribution. It  
 268 can be modelled by two discrete non-decreasing functions  $\underline{F}$  and  $\overline{F}$  from  
 269  $(\xi^1, \dots, \xi^n)$  to  $[0, 1]$  such that  $\underline{F}(\xi^i) \leq \overline{F}(\xi^i)$  for all  $i$  in  $\{1, \dots, n\}$  and  
 270  $\underline{F}(\xi^n) = \overline{F}(\xi^n) = 1$ . The values  $\underline{F}(\xi^i), \overline{F}(\xi^i)$  are interpreted as the following  
 271 bounds

$$\underline{F}(\xi^i) \leq \sum_{i=1}^n \xi^i \leq \overline{F}(\xi^i)$$

272 and we can denote by  $\mathcal{P}_{\underline{F} \leq \overline{F}}$  the knowledge modelled by a p-box. A p-box  
 273 information provides us with estimates about the cumulated probabilities of  
 274 events of the kind  $\{x_1, \dots, x_i\}$ , hence assuming that the ordering induced  
 275 by the indices do make sense.

276 In the case of p-boxes, the bounds over  $\xi^i$  are very easy to determine  
 277 [59], and are equal to

$$\underline{\xi}^i = \max(0, \underline{F}(\xi^i) - \overline{F}(\xi^{i-1})) \quad (27)$$

278

$$\overline{\xi}^i = \overline{F}(\xi^i) - \underline{F}(\xi^{i-1}) \quad (28)$$

279 with the convention  $\underline{F}(\xi^0) = \overline{F}(\xi^0) = 0$ .

280 *Example 5.* Assume we have to assess the how likely it is that a bacterial  
 281 population is below some threshold, or how likely it is that a component may  
 282 function for a given period of time. The population sizes or time intervals  
 283 may be discretized into  $x_1, x_2, x_3$ . Assume the following p-box has been  
 284 given as information

$$\underline{F}(\xi^1) = 0.2, \underline{F}(\xi^2) = 0.7 \text{ and } \overline{F}(\xi^1) = 0.5, \overline{F}(\xi^2) = 0.9.$$

285 From it we can deduce the bounds

$$\underline{\xi}^1 = 0.2, \underline{\xi}^2 = 0.2, \underline{\xi}^3 = 0.1 \text{ and } \overline{\xi}^1 = 0.5, \overline{\xi}^2 = 0.7, \overline{\xi}^3 = 0.3.$$

286 P-boxes usually rely on the fact that the set  $(\xi^1, \dots, \xi^n)$  is naturally  
 287 ordered, and provide confidence bounds over sets of the kind  $\{\xi^1, \dots, \xi^i\}$ .  
 288 However, one possibility is to extend this notion by considering that values  
 289  $\xi^i$  follows an arbitrary ordering  $\xi^{(1)} \leq \dots \leq \xi^{(n)}$  (for example, from the least  
 290 to the most plausible element) and to ask to the expert to provide upper and  
 291 lower confidence bounds about the fact that the truth lies in  $\{x_{(1)}, \dots, x_{(i)}\}$ ,  
 292 thus obtaining  $\underline{F}(\xi^{(i)})$  and  $\overline{F}(\xi^{(i)})$ . As in principle any ordering can be used,  
 293 this is indeed a generalization of p-boxes, known as comonotonic clouds [31].  
 294 In particular, in the case where  $\underline{F}(\xi^{(i)}) = 0$  for any  $i$ , we retrieve the notion  
 295 of possibility distribution as a special case.

296 Up to now, what is the maximal number of extreme points of a p-box  
 297 structure and how to efficiently enumerate them remains an open prob-  
 298 lem. However, as p-boxes are a special case of belief functions, one can

299 use (potentially sub-optimal) algorithms and methods applicable to belief  
 300 functions [17]. It is also clear that the maximal number of such points is  
 301 bounded above by the maximal number of extreme point of a belief func-  
 302 tion ( $n!$ ). Classical p-boxes suffer from the fact that a natural order must  
 303 exist on  $\mathcal{X}$ , and when no such order exists, then the average of generalized  
 304 p-boxes relying on different orders will not be a p-box.

### 305 3.4. Belief functions and random sets

306 Formally, a random set or belief function, initially introduced by Demp-  
 307 ster [30] and Shafer [57], is defined as a positive mapping  $\nu : 2^{\{\xi^1, \dots, \xi^n\}} \rightarrow$   
 308  $[0, 1]$  from the power set of  $\{\xi^1, \dots, \xi^n\}$  to the unit interval, such that  
 309  $\nu(\emptyset) = 0$  and  $\sum_E \nu(E) = 1$ . From this mapping can then be defined proba-  
 310 bility bounds  $Bel(A), Pl(A)$  for any event that are equal to

$$Bel(A) = \sum_{E, E \subseteq A} \nu(E) \text{ and } Pl(A) = \sum_{E, E \cap A \neq \emptyset} \nu(E) = 1 - Bel(A^c) \quad (29)$$

311 that induce an information  $\mathcal{P}_\nu$  such that

$$\mathcal{P}_\nu = \left\{ \xi : \sum_{E \subseteq A} \nu(E) \leq \sum_{\xi^i \in A} \xi^i \leq \sum_{E \cap A \neq \emptyset} \nu(E), \forall A \right\} \quad (30)$$

312 In particular, this means that given a function  $\nu$ , the bounds over elementary  
 313 events are given by

$$\xi^i = Bel(\{\xi^i\}) = \nu(\{\xi^i\}) \quad (31)$$

314

$$\bar{\xi}^i = Pl(\{\xi^i\}) = \sum_{\xi^i \in E} \nu(E) \quad (32)$$

315 Belief functions are instrumental to model frequencies of imprecise observa-  
 316 tions, for example when multiple exclusive options can be chosen in surveys,  
 317 or when some sensors sometimes send back imprecise observations. They  
 318 also include p-boxes, comonotonic clouds and possibilities as special cases.

319 *Example 6.* Assume again that we can meet four different situations  $x_1, x_2, x_3$   
 320  $, x_4$ . Out of 20 observations,  $x_1, x_2, x_3, x_4$  were each perfectly observed re-  
 321 spectively 3, 2, 5, 6 times, we observed 3 times the set  $\{x_2, x_3, x_4\}$  (excluding  
 322  $x_1$ ) and 2 times the set  $\{x_1, x_2, x_3\}$ . Such observations can be modelled on  
 323  $\xi^1, \xi^2, \xi^3, \xi^4$  by the mass

$$\nu(\{\xi^1\}) = 3/20, \nu(\{\xi^2\}) = 2/20, \nu(\{\xi^3\}) = 5/20, \nu(\{\xi^4\}) = 6/20,$$

324

$$\nu(\{\xi^1, \xi^2, \xi^3\}) = 2/20, \nu(\{\xi^2, \xi^3, \xi^4\}) = 3/20.$$

325 From this, we can for example deduce  $\xi^3 = 0.25$  and  $\bar{\xi}^3 = 0.5$ .

326 Belief functions are general enough to deal with a lot of practical as-  
 327 sessments, and share the properties of probability intervals that an average  
 328 of belief functions is still a belief function. However, providing an intuitive  
 329 graphical representation of a belief function is challenging, and their use may  
 330 quickly lead to computational issues (e.g., their number of extreme points  
 331 can be as high as  $\mathcal{X}!$  [50])

### 332 3.5. Fuzzy random variables

333 Fuzzy random variables have been given different interpretations in the  
 334 literature, depending on the nature of the fuzzy elements. For example, a  
 335 fuzzy random variable can be seen as a random phenomenon with precise  
 336 observations that are fuzzy in nature, or as a random phenomenon with  
 337 imprecise observations. We refer to [19, 21, 22] for a detailed discussion. In  
 338 this paper, Fuzzy random variables are interpreted as conditional possibility  
 339 measures [4, 21], which consist in putting positive masses, not on subsets,  
 340 but on possibility distributions. They can be modelled by a set  $\pi_1, \dots, \pi_k$   
 341 where each distribution receives probability mass  $p(\pi_i)$ . As each  $\pi_i$  can in  
 342 turn be turned into a mass function  $\nu_{\pi_i}$  defined this time over subsets, it  
 343 is always possible to come back from a fuzzy random variable to a classical  
 344 mass function, simply by computing for any subset  $E$  the value

$$\nu(E) = \sum_{i=1}^k p(\pi_i) \nu_{\pi_i}(E).$$

345 We obtain a weighted random sampling of subset  $E$  defining a belief function  
 346  $\nu$ . Fuzzy random variables in this context may be cast into the framework  
 347 of belief functions leading to the same formal advantages and disadvantages  
 348 of them (see Section 3.4). Fuzzy random variables can result, for instance,  
 349 from Monte-Carlo simulations of physical models mixing possibilistic and  
 350 probabilistic uncertainty [4], or from the random observation of fuzzy sets  
 351 (modelling an ill-calibrated scale, for instance [18]).

### 352 3.6. Summary of types of knowledge

353 A final type of knowledge simply consists in directly providing bounds  
 354 over the values of possible observations  $x_i$ . This means specifying, for each  
 355  $\xi^i$ , the bounds  $\underline{f}_i = \underline{\xi}^i$  and  $\bar{f}_i = \bar{\xi}^i$ . Such bounds are formally equivalent to  
 356 probability intervals [11].

357 There are multiple ways to derive such bounds: for instance by instan-  
 358 tiating multinomial confidence intervals over observations, by requiring lin-  
 359 guistic opinions of the type "probable", "very probable" from the experts  
 360 and then translating them into numerical evaluations [54], by simply re-  
 361 quiring numerical evaluations from the experts, when having imprecise his-  
 362 tograms, ...

363 Table 1 summarises the most common type of practical information one  
364 can meet, to what type of information they correspond and how can be  
365 computed the lower/upper values  $\underline{\xi}^i$  and  $\bar{\xi}^i$ . These are the values (used as  
366 a common mathematical tool) that are then combined and integrated into  
367 the learning process developed in Section 2.2.

ACCEPTED MANUSCRIPT

Model	Usual type of information	$\underline{\xi}^i$	$\bar{\xi}^i$
Precise values	Sample/simulation	$f_i$	$f_i$
Possibility	Lower confidence on nested sets Linguistic assessments	$1 - \max_{\xi \neq \xi^i} \pi(\xi^i)$	$\bar{\xi}^i = \pi(\xi^i)$
P-boxes and clouds	Lower/upper confidence on nested sets	$\max(0, \underline{F}(\xi^{(i)}) - \bar{F}(\xi^{(i-1)}))$	$\bar{F}(\xi^{(i)}) - \underline{F}(\xi^{(i-1)})$
Belief functions	Imprecise sample	$\nu(\xi^i)$	$\sum_{\xi^i \in E} \nu(E)$
Fuzzy random variable	Fuzzy sample	$\sum_{i=1}^k p(\pi_i) \nu_{\pi_i}(\xi^i)$	$\sum_{\xi^i \in E} \sum_{i=1}^k p(\pi_i) \nu_{\pi_i}(E)$
Probability bounds	Linguistic assessments Multinomial confidence regions	$\underline{f}_i$	$\bar{f}_i$

Table 1: Summary of the different types of collectible information

#### 368 4. Robust dynamic probabilistic graphical models

369 When modeling complex systems, we are not interested in a single vari-  
 370 able, but in multiple variables interacting with each others and evolving  
 371 over time. In theory, our knowledge about these variables, their interaction  
 372 and evolution can be represented by a credal set defined over the Cartesian  
 373 product of the corresponding spaces.

374 In practice, we need tool to represent these interactions, and to sim-  
 375 plify the daunting task of specifying a full joint model. Credal networks  
 376 are graphical (directed) models that aims at encoding our knowledge about  
 377 variable interactions and at splitting the full joint into multiple, simple con-  
 378 ditional models. This section introduces them, as well as their dynamical  
 379 extension.

##### 380 4.1. Credal networks

381 Let  $\mathbf{X} = (X_1, \dots, X_n)$  be a discrete random vector associated with the  
 382 joint probability mass function  $p(\mathbf{X})$  defined over  $\prod_{i=1}^n \mathcal{X}_i$ . Let  $K(\mathbf{X})$  be the  
 383 closed convex set of multivariate probability mass functions describing our  
 384 knowledge of  $\mathbf{X}$ .

385 A credal network (CN) [24, 23] is an extension of Bayesian networks  
 386 (BNs) where imprecision is introduced in probabilities by means of credal  
 387 sets [47]. When working with probability sets rather than precise proba-  
 388 bilities, the notion of stochastic independence can be extended in several  
 389 ways [20]. Within graphical models, the most commonly used extension is  
 390 *strong independence* (also called a *type 1 product* of the marginals in [61]),  
 391 that induces the strong extension. It can be interpreted as a robust model  
 392 of a precise yet ill-known BN. Under the *strong extension* [23] hypothesis,  
 393 the joint credal set  $K(\mathbf{X})$  over  $\Omega_{\mathbf{X}}$  may be formulated as:

$$394 \quad K(\mathbf{X}) = CH \left\{ p(\mathbf{X}) : p(\mathbf{X}) = \prod_{i=1}^n p_i, p_i \in K_i \right\} \quad (33)$$

395 where  $p_i = p(X_i | \mathbf{U}_i)$ ,  $\mathbf{U}_i$  denotes the set of parent nodes of the node  $X_i$  and  
 396  $K_i = K(X_i | \mathbf{U}_i)$  is the closed convex set of probability mass function for  
 397 the random variable  $X_i$  given  $\mathbf{U}_i$ . As mentioned in Section 2, it is sufficient  
 398 to focus on  $\mathcal{E}xt(K(X_i | \mathbf{U}_i))$  in Eq. (33).

399 In this work, we focus on the notion of strong independence and its ex-  
 400 tension to dynamical models, as this is the most widely used independence  
 401 notion within graphical models and the one that fits the best with a robust  
 402 interpretation of probability sets. Other independence notions that may  
 403 even have asymmetrical versions such as epistemic irrelevance remain com-  
 404 putationally intractable [25, 49], except for specific network structures [9]  
 that are usually less complex than the one generally considered here.



405 *4.2. Dynamic credal networks*

406 Let  $\mathbf{X}(\mathbf{t}) = (X_1(1), \dots, X_n(1), \dots, X_1(\tau), \dots, X_n(\tau))$  be a discrete ran-  
 407 dom vector process associated with the joint probability function  $p(\mathbf{X}(\mathbf{t}))$   
 408 defined over  $\prod_{t=1}^{\tau} \prod_{i=1}^n \mathcal{X}_i(t)$ . Let  $K(\mathbf{X}(\mathbf{t}))$  be the closed convex set of mul-  
 409 tivariate probability mass functions for  $\mathbf{X}(\mathbf{t})$ .

410 A dynamic credal network (CDN) [41] is a dynamic Bayesian network  
 411 (DBNs) [51] where conditional probabilities  $p(X_i(t) | \mathbf{U}_i(t))$  (noted  $p_i^t$ ) are  
 412 replaced by credal sets  $K(X_i(t) | \mathbf{U}_i(t))$  (noted  $K_i^t$ ). It is therefore a time-  
 413 sliced model that can be used to describe a dynamic process or system<sup>2</sup>.

414 We assume the same *first-order Markov property* as for DBN, meaning  
 415 that parents only originate from the same or previous time slice, and also  
 416 that conditional models remain the same at each times slice, that is

$$K(X_i(t) | \mathbf{U}_i(t)) = K(X_i(2) | \mathbf{U}_i(2)), \forall t \in \llbracket 2, \tau \rrbracket. \quad (34)$$

417 Therefore, specifying the graphical structure of a DCN requires the same  
 418 effort as the one of a DBN (that is, specifying only two consecutive time  
 419 slices) but allows the user to provide conditional credal sets rather than  
 420 probabilities if these latter cannot be reliably estimated (from data and/or  
 421 experts).

422 *4.2.1. Independence in DCN*

423 Extending DBN to DCN requires to specify which kind of independence  
 424 we consider within and also between each time-slice. We remind that we  
 425 will only consider extensions relying on the strong independence (33). The  
 426 most straightforward extension is to simply apply strong independence to  
 427 the whole network, i.e.,

$$K(\mathbf{X}(\mathbf{t}))_{st} = CH \left\{ p(\mathbf{X}(\mathbf{t})) : p(\mathbf{X}(\mathbf{t})) = \prod_{i=1}^n \prod_{t=1}^{\tau} p_i^t, p_i^t \in K_i^t \right\} \quad (35)$$

We call this extension the *dynamic strong extension* and it is worth notic-  
 ing that we can have  $p_i^t \neq p_i^{t'}$  for  $t, t' \in \llbracket 2, \tau \rrbracket$ . That is, we do not assume  
 probabilities within each time-slice to be identical. However, when stepping  
 to dynamic models, Condition (34) allows us to use the notion of *repeti-*  
*tive independence* (also called a *type 2 product* of the marginals in [61]).  
 This condition states that if two variables  $X, Y$  have the same set of pos-  
 sible outcomes, that is  $\mathcal{X} = \mathcal{Y}$ , and can be assumed to be governed by the  
 same probability distribution belonging to  $K(X)$ , then the joint credal set  
 $K(X, Y)$  is :

$$K(X, Y) = CH\{p(X)p(X) : p(X) \in K(X)\}. \quad (36)$$

<sup>2</sup>It should be noted that the network itself is static, but is used to represent a dynamic process.

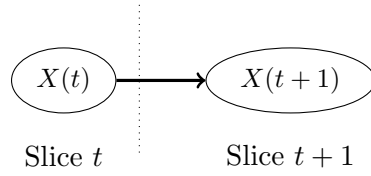


Figure 2: A simple dynamical graphical model

428 Adapting this notion of independence to DCN, so that probabilities of each  
 429 time slice are assumed to be identical, leads to a second extension, i.e.,

$$K(\mathbf{X}(\mathbf{t}))_{rp} = CH \left\{ \begin{array}{l} p(\mathbf{X}(\mathbf{t})) : p(\mathbf{X}(\mathbf{t})) = \prod_{i=1}^n \prod_{t=1}^{\tau} p_i^t, \\ p_i^2 \in K_i^2 \text{ and } p_i^t = p_i^2 \forall t \in \llbracket 2, \tau \rrbracket \end{array} \right\} \quad (37)$$

430 that we call the *dynamic repetitive extension*. We have  $K(\mathbf{X}(\mathbf{t}))_{rp} \subseteq K(\mathbf{X}(\mathbf{t}))_{st}$ ,  
 431 as  $K(\mathbf{X}(\mathbf{t}))_{rp}$  is more constrained. In practice, the strong extension assumes  
 432 that the dynamic network is ill-defined and that its behaviour can change  
 433 between time slices, while the repetitive extension assumes that we seek a  
 434 precise classical DBN who is partially known.

*Example 7.* Consider the very simple example where  $\mathcal{X} = \{0, 1\}$  and the  
 2-slice network given in Figure 2, which is nothing else than a two-state  
 imprecise Markov chain, and an observed value  $X(1) = 1$ . Assume further-  
 more that we have three time slices ( $\tau = 3$ ), that  $X(1) = 1$  is observed, and  
 that

$$\begin{aligned} p(X(t) = 1 | X(t-1) = 0) &= 0.8, \\ p(X(t) = 1 | X(t-1) = 1) &\in [0.2, 0.5]. \end{aligned}$$

435 That is, the transition rates from state 0 are precisely known, but not the  
 436 one from state 1 (although staying in state 1 is clearly less likely). The  
 437 different extreme points over  $(X(1), X(2), X(3))$  resulting from the strong  
 438 and repetitive extension are summarized in Table 2, in which we adopt the  
 439 notation  $x(t)$  for  $X(t) = 1$  for simplification purposes. Each cell of the table  
 440 corresponds to a precise network obtained by a specific selection of extreme  
 441 points. The non-specified transition probabilities can be retrieved by the  
 442 formula  $p(X(t) = 1 | X(t-1) = 1) = 1 - p(X(t) = 0 | X(t-1) = 1)$ .

#### 443 4.2.2. Inference algorithms in DCN

444 (D)CNs can be queried as in (D)BNs to get information about the state of  
 445 a variable given evidence about other variables, with respect to the chosen  
 446 network *extension*. However, the use of credal sets makes the updating  
 447 problem much harder, as it becomes an optimization problem. As such,  
 448 the computation of the lower bound on  $p(\mathbf{X}_Q | \mathbf{X}_E)$  requires to minimize a  
 449 fraction containing polynomials :

Strong extension				Repetitive extension			
100	101	110	111	100	101	110	111
$P(x(t) x(t-1)) = 0.5$				$P(x(2) x(1)) = 0.5, P(x(3) x(2)) = 0.2$			
0.1	0.4	0.25	0.25	0.1	0.4	0.4	0.1
$P(x(t) x(t-1)) = 0.2$				$P(x(2) x(1)) = 0.2, P(x(3) x(2)) = 0.5$			
0.16	0.64	0.16	0.04	0.4	0.4	0.1	0.1

Table 2: Simple DCN extreme probabilities

$$p(\mathbf{X}_Q(t) | \mathbf{X}_E(t)) = \min_{p(\mathbf{X}(t)) \in K(\mathbf{X}(t))_\omega} \frac{\sum_{X_i(t) \in \mathbf{X}(t) \setminus \mathbf{X}_Q(t) \cup \mathbf{X}_E(t)} \prod_{i=1}^n \prod_{t=1}^{\tau} p_i^t}{\sum_{X_i(t) \in \mathbf{X}(t) \setminus \mathbf{X}_E(t)} \prod_{i=1}^n \prod_{t=1}^{\tau} p_i^t} \quad (38)$$

450 with  $p(\mathbf{X}(t)) \in K(\mathbf{X})_\omega$  belonging to the *dynamic strong extension* ( $\omega = st$ )  
451 or *dynamic repetitive extension* ( $\omega = rp$ ) of the network. An upper bound  
452 can be obtained by maximizing (38). It is known that such a minimum (or  
453 maximum) is obtained at a vertex of the *dynamic strong/repetitive extension*.  
454 Depending on (1) the structure of network, (2) the number of modality  
455 of variables and (3) the chosen extension (strong/repetitive), the updating  
456 problem will be more or less complex to solve. Because inferences are already  
457 hard in static credal networks, little work has been done on DCNs [41].  
458 By unrolling a two-time slice network over  $T$  time steps, the number of  
459 possible vertex combinations goes from  $\prod_{i,t=0} \#\mathcal{E}xt(K_i^t) \prod_{i,t=1} \#\mathcal{E}xt(K_i^t)$  in  
460 the case of repetitive independence, to  $\prod_{i,t=0} \#\mathcal{E}xt(K_i^t) \prod_{i,t=1} \#\mathcal{E}xt(K_i^t)^{\tau-1}$   
461 in the case of strong independence. Given the potential number of vertices,  
462 approximate algorithms seem more appropriate regarding DCNs.

463 Many algorithms, exact and approximate, have been proposed to deal  
464 with CN. Some are generalizations of well known (D)BNs algorithms. Among  
465 the approximate algorithms, there are those that compute inner bounds, i.e.  
466 bounds that are enclosed by the exact ones, outer bounds, which enclose  
467 the exact ones, and those that perform randomly. The 2U algorithm [34]  
468 performs an exact rapid inference in the case of binary tree-shaped (D)CNs  
469 with the assumption of *strong independence*. The CCM transformation [15]  
470 turns a (D)CN into a (D)BN by adding transparent nodes before performing  
471 an Maximum A Posteriori (MAP) estimation over the latter to find the best  
472 combination of vertices. It has the same complexity as credal network in-

473 ference, that is  $NP^{PP}Complete$ , and performs poorly with separately spec-  
 474 ified credal networks such as the one we used during our trials (because  
 475 of the sheer number of vertices). Optimization techniques such as branch  
 476 and bound over local vertices of credal sets [27, 13] are also well suited to  
 477 medium-sized networks and can be stopped at any time to give an approxi-  
 478 mate answer. Other algorithms are based on a variable elimination scheme  
 479 from (D)BNs, such as Separable Variable Evaluation [26, 55] which keeps  
 480 the separately specified credal sets as separated as possible during propaga-  
 481 tion, and can be mapped to an integer or a multi-linear program [29, 28].  
 482 Regarding binary and DAG-shaped (DAG : Directed Acyclic Graph) credal  
 483 networks, algorithm L2U (Loopy 2U) [44] (similar to LBP (Loopy Belief  
 484 Propagation) [63]) produces either inner or outer approximations. Its effi-  
 485 ciency is due both to the bounded cardinality of variables and to ignoring  
 486 loops. Another way to handle credal sets complexity is to represent them  
 487 by simpler means. Variational methods [43, 42] choose a family of functions  
 488 to approximate the exact combination of credal sets to decrease compu-  
 489 tational costs. Those functions are optimized according to some criteria  
 490 until convergence and the inference is then realized in the network with the  
 491 original credal sets replaced by the new found functions. The A\R(+)  
 492 algorithm [27] uses interval probability arithmetic to approximate credal sets  
 493 in a propagation scheme in tree-shaped networks (with the use of some ad-  
 494 ditional constraints limiting the information loss in its enhanced version).  
 495 The intervals produced are outer bounds of the real ones. Although those  
 496 algorithms are fast in medium-sized network, they either produce too many  
 497 approximations or are too complex to work with DCNs. Another popular  
 498 family of approximate algorithms producing inner bounds is based on Monte-  
 499 Carlo sampling [38]. Several methods have been proposed to better guide  
 500 the search (simulated annealing [12], genetic algorithms [14]) among the ver-  
 501 tices of the (conditional) local credal sets, but they require some tuning for  
 502 more accurate results, otherwise they can lead to poor approximations.

503 Although there exist several inference algorithms, none allows to do infer-  
 504 ence, in a realistic and practical way, on networks capable of representing  
 505 global complex system of Life Sciences. In further inferences, we used a  
 506 simple Monte-Carlo sampling algorithm [38] which has the advantage to be  
 507 a good starting point, as it applies with the same easiness to *dynamic repet-*  
 508 *itive* and *strong extensions* (with a faster convergence for *dynamic repetitive*  
 509 *extension*).

#### 510 4.2.3. Robust parameter learning

511 Let  $p_{ijk}^t$  be the probability that  $X_i(t) = x_k$ , given that its parents have  
 512 instantiation<sup>3</sup>  $x_j$  (corresponding itself to a vector where  $j$  represents the

---

<sup>3</sup>Possible values of variables according to its discretization.

513 vector of parents of  $i$ ), *i.e.*

$$p_{ijk}^t = p(X_i(t) = x_k | \mathbf{U}_i(t) = x_j) \quad \begin{array}{l} i = 1, \dots, n \\ j = 1, \dots, c_i \\ k = 1, \dots, r_i \end{array} \quad (39)$$

514 where  $r_i$  is the number of values that node  $i$  can take and  $c_i$  is the number  
515 of distinct configurations of  $\mathbf{U}_i(t)$ . Parameter learning consists in estimat-  
516 ing  $p_{ij}^t$  faced with available information [45, 10]. For the sake of clarity,  
517 parameters  $p_{ij}^t$  will be denoted  $p_{ij}$  since parameters  $p_{ij}^t$  are time-invariant in  
518 the case of *repetitive extension* assumption and it is sufficient to only con-  
519 sider information limited to each time slice in the case of *strong extension*  
520 assumption. According to section 2.2, for all  $i \in \{1, \dots, n\}; j \in \{1, \dots, c_i\}$   
521 the credal set  $\tilde{K}_{(s_i; \xi_i)_{l=0}^m}(X_i | \mathbf{U}_i = x_j)$  may be approximated by using the  
522 outer credal set  $K_{(s_i; \xi_i)_{l=0}^m}(X_i | \mathbf{U}_i = x_j)$  defined by

$$K_{(s_i; \xi_i)_{l=0}^m}(X_i | \mathbf{U}_i = x_j) = \left\{ p_{ij} : p_{ijk} \in [\underline{p}_{ijk}, \bar{p}_{ijk}], \sum_k p_{ijk} = 1 \right\} \quad (40)$$

523 where  $[\underline{p}_{ijk}, \bar{p}_{ijk}]$  is estimated and updated from Eq. (19) according to the  
524 available sources of knowledge  $(S_0, \dots, S_m)$ .

#### 525 4.3. Practical robust parameter learning example

526 Wood is essentially composed of cellulose (denoted  $C$ ) that is a polymer  
527 whose quantity characterizes the nature of wood (denoted  $T$ ) namely hard-  
528 wood or softwood. Imagine that we want to determine the kind of wood  
529 according to its chemical composition tainted with uncertainties, that is we  
530 are interested in  $\bar{P}(T|C)$ . For the sake of clarity, we choose  $C = \{x_1 =$   
531  $20\%, x_2 = 40\%, x_3 = 60\%\}$  meaning that there is 20%, 40% or 60% of cel-  
532 lulose inside wood,  $T = \{x_1 = \text{Soft}, x_2 = \text{Hard}\}$  and all sources  $s_i$  have the  
533 same confidence level, *i.e.*  $s_i = 1$  for all  $i$ . We thus need to estimate the  
534 following parameters:

$$p_{jk} = p(T = x_k | C = x_j) \quad (41)$$

535 according to the available knowledge described in the following. The credal  
536 sets  $K(T|C = x_j)$  are initialized by

$$K_{s_0}(T|C = x_j) = \{p_j : p_{jk} \geq 0, \sum_k p_{jk} = 1\}, \quad \forall j = 1, \dots, 3 \quad (42)$$

537 1. Precise measures are provided  $\{(20, \text{Soft}), (20, \text{Hard}), (40, \text{Hard}), (60, \text{Soft})\}$   
538 leading to update by Eq. (19)

539 •  $K_{(s_0, s_1)}(T|C = 20) = \{p_1 : \frac{1}{4} \leq p_{11} \leq \frac{3}{4}, p_{12} = 1 - p_{11}\},$

- 540       •  $K_{(s_0, s_1)}(T|C = 40) = \{p_2. : 0 \leq p_{21} \leq \frac{1}{2}, p_{22} = 1 - p_{21}\},$   
 541       •  $K_{(s_0, s_1)}(T|C = 60) = \{p_3. : \frac{1}{2} \leq p_{31} \leq 1, p_{32} = 1 - p_{31}\}.$

542 2. A first expert says that the more cellulose there is, the harder the  
 543 wood. This information may be formalized by means of the following  
 544 fuzzy numbers or possibility distribution (see Section 3.2):

- 545       •  $\pi(T = \text{Hard}|C = 20) = 0.5, \quad \pi(T = \text{Soft}|C = 20) = 1$   
 546       •  $\pi(T = \text{Hard}|C = 40) = \pi(T = \text{Soft}|C = 40) = 1$   
 547       •  $\pi(T = \text{Hard}|C = 60) = 1, \quad \pi(T = \text{Soft}|C = 60) = 0.5$

548 meaning for instance that  $P(T = \text{Hard}|C = 20) \leq 0.5$  leading to  
 549 update

- 550       •  $K_{(s_0, s_1, s_2)}(T|C = 20) = \{p_1. : \frac{1}{3} \leq p_{11} \leq \frac{5}{6}, p_{12} = 1 - p_{11}\},$   
 551       •  $K_{(s_0, s_1, s_2)}(T|C = 40) = \{p_2. : 0 \leq p_{21} \leq \frac{2}{3}, p_{22} = 1 - p_{21}\},$   
 552       •  $K_{(s_0, s_1, s_2)}(T|C = 60) = \{p_3. : \frac{1}{3} \leq p_{31} \leq \frac{5}{6}, p_{32} = 1 - p_{31}\}.$

553 3. A second expert provides more accurate estimation in terms of confi-  
 554 dence

- 555       •  $P(T = \text{Soft}|C = 20) \geq 95\%,$   
 556       •  $P(T = \text{Hard}|C = 40) \geq 60\%,$   
 557       •  $P(T = \text{Hard}|C = 60) \geq 95\%.$

558 which can be modeled again by a possibility distribution. This leads  
 559 to update

- 560       •  $K_{(s_0, \dots, s_3)}(T|C = 20) = \{p_1. : 0.49 \leq p_{11} \leq 0.875, p_{12} = 1 - p_{11}\},$   
 561       •  $K_{(s_0, \dots, s_3)}(T|C = 40) = \{p_2. : 0 \leq p_{21} \leq 0.6, p_{22} = 1 - p_{21}\},$   
 562       •  $K_{(s_0, \dots, s_3)}(T|C = 60) = \{p_3. : 0.25 \leq p_{31} \leq 0.64, p_{32} = 1 - p_{31}\}.$

563 4. Defective sensors and measurements provide joint imprecise observa-  
 564 tions, summarized in Table 3 and producing a joint belief function  
 565 (Section 3.4).

From this information lower and upper probability bounds over pa-  
 rameters are given by

$$Bel(T = t|C = c) = \frac{Bel(T = t, C = c)}{Bel(T = t, C = c) + \sum_{t' \neq t} Pl(T = t', C = c)}$$

$$Pl(T = t|C = c) = \frac{Pl(T = t, C = c)}{Pl(T = t, C = c) + \sum_{t' \neq t} Bel(T = t', C = c)}$$

	Focal sets	Type		
		Soft	Hard	{Soft,Hard}
Cellulose	20	1	2	0
	40	4	6	1
	60	0	10	1
	{20, 40}	10	5	1
	{20, 60}	0	0	5
	{40, 60}	1	3	10
	{20, 40, 60}	0	0	8

Table 3: Focal sets occurrences

For example

$$Bel(T = \text{Soft}|C = 20) = \frac{1/68}{1/68 + 21/68} = 0.045$$

$$Pl(T = \text{Soft}|C = 20) = \frac{24/48}{24/48 + 2/48} = 0.926$$

566 Credal set  $K_{(s_0, \dots, s_4)}$  is then updated by

- 567 •  $K_{(s_0, \dots, s_4)}(T|C = 20) = \{p_1 : 0.4 \leq p_{11} \leq 0.89, p_{12} = 1 - p_{11}\},$
- 568 •  $K_{(s_0, \dots, s_4)}(T|C = 40) = \{p_2 : 0.021 \leq p_{21} \leq 0.65, p_{22} = 1 - p_{21}\},$
- 569 •  $K_{(s_0, \dots, s_4)}(T|C = 60) = \{p_3 : 0.2 \leq p_{31} \leq 0.67, p_{32} = 1 - p_{31}\}.$

## 570 5. Real-life case study

571 To illustrate the feasibility and practical use of our approach in a real  
572 case, we have focused on the ripening process of the Camembert type soft  
573 mould cheese that represents an ecosystem and a bioreactor difficult to ap-  
574 prehend from a global point of view [37, 52]. Based on recent works carried  
575 out by Baudrit *et al.* [5]; Sicard *et al.* [48], a simplified sub-structure  
576 of dynamic Bayesian networks has been extracted (see Figure 3) provid-  
577 ing a qualitative representation of the coupled dynamics of yeast behaviour  
578 *Kluyveromyces marxianus* ( $Km$ , colony forming unit/g of Fresh Cheese in  
579 decimal logarithmic scale) with its lactose substrate ( $lo$ , g/Kg of Fresh  
580 Cheese) influenced by temperature ( $T$ , °C) inside the ripening chamber and  
581 involving odour changes ( $Od = \{\text{Fresh, Mushroom, Camembert}\}$ ).

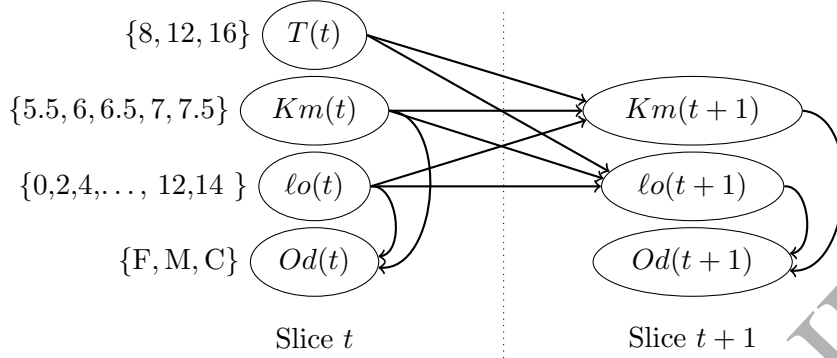


Figure 3: Structure of the dynamic credal network and the values of each variables representing the coupled dynamics Km growth versus lo consumptions influenced by temperature involving odour changes during the cheese ripening process (F=Fresh,M=Mushroom,C=Camembert).

### 582 5.1. Parameter learning

Assuming *repetitive extension* for computational reason mentioned in Section 4.2.2, we present, in the following, how parameters

$$\begin{aligned}
 \mathbf{p}_1 &= p(Km(1)), \\
 \mathbf{p}_2 &= p(lo(1)), \\
 \mathbf{p}_3 &= p(T(1)), \\
 \mathbf{p}_4 &= p(Km(2)|(Km(1), lo(1), T(1))), \\
 \mathbf{p}_5 &= p(lo(2)|(Km(1), lo(1), T(1))), \\
 \mathbf{p}_6 &= p(Od(1)|(Km(1), lo(1)))
 \end{aligned}$$

583 may be estimated by using the robust hybrid parameter learning when we  
 584 have several sources of knowledge (denoted  $S_i$ ) tainted with stochastic and  
 585 epistemic uncertainty.

#### 586 1. Initialization ( $S_1$ ).

587 All DCN parameters are initialized by:

- 588 • An experimental database  $S_{experiments}$  of six cheese ripening trials  
 589 carried out for temperatures varying from  $T = 8$  to  $16$  °C is  
 590 available.
- 591 • the vacuous credal sets leading to bracket parameters by  $[0,1]$   
 592 when no information is available.

593 With  $s_1$  corresponding to the confidence level about experimental trials  
 594  $S_{experiments}$ , according to (19) we have:

$$p_{ijk} \in \left[ \frac{s_1 f_{ijk}}{s_0 + s_1}, \frac{s_0 + s_1 f_{ijk}}{s_0 + s_1} \right] \quad (43)$$



595 where  $f_{ijk}$  represents the observed frequency corresponding to sample  
596 information in Table 1 and linked to Section 3.1.

597 2. Integration of partial mechanistic model tainted with uncertainties ( $S_2$ ).  
598 The yeast  $Km$  is one of the dominant species in the yeast flora of  
599 Camembert cheeses and its principal activity is the consumption of  
600 lactose ( $lo$ ) [46]. Models to determine the growth of microorganisms  
601 have been studied in the fermentation industry [58], and the descrip-  
602 tion of the growth of  $Km$  is obtained by performing material balances  
603 on biomass  $Km$  and lactose  $lo$  [3]:

$$(S) \begin{cases} \frac{dKm}{dt} &= \mu \frac{lo}{K_{lo} + lo} Km - b \cdot Km \\ \frac{dlo}{dt} &= -\frac{\mu}{\beta} \frac{lo}{K_{lo} + lo} Km \end{cases} \quad (44)$$

604 where  $\mu$  (the maximum specific growth rate of  $Km$ ),  $K_{lo}(T)$  (the  
605 half saturation constant for growth),  $b$  (the decay coefficient) and  
606  $\beta$  (the yield coefficient for  $Km$  on lactose), depending on tempera-  
607 ture, are tainted with stochastic and epistemic uncertainties, due to  
608 the natural variability of yeast population and the imperfection of  
609 the model. The background knowledge about parameters  $\mathbf{p}_1$ ,  $\mathbf{p}_2$ ,  $\mathbf{p}_4$   
610 and  $\mathbf{p}_5$  are then updated regardless of the rest of network by using  
611 a simulated database  $S_{simulated}$  resulting from Monte Carlo simula-  
612 tion coupled to interval analysis [4] leading to manage a joint ran-  
613 dom set  $([Km(t), \overline{Km}(t)], [lo(t), \overline{lo}(t)], T(t))_l$  associated with mass  $\nu_l =$   
614  $1/\#S_{simulated}$  such that for instance

$$\mathbf{p}_{4jk} \in \left[ \frac{s_1 f_{4jk} + s_2 \xi_{4jk}}{s_0 + s_1 + s_2}, \frac{s_0 + s_1 f_{4jk} + s_2 \bar{\xi}_{4jk}}{s_0 + s_1 + s_2} \right] \quad (45)$$

615 where

$$\xi_{4jk} = \frac{bel(j, k)}{bel(j, k) + \sum_{l \neq k} pl(l, j)} \quad \text{and} \quad \bar{\xi}_{4jk} = \frac{pl(j, k)}{pl(j, k) + \sum_{l \neq k} bel(l, j)} \quad (46)$$

616 and

$$pl(j, k) = \sum_{\substack{l, [km(t+1), \overline{km}(t+1)]_l \cap \{km_k\} \neq \emptyset \\ [km(t), \overline{km}(t)]_l \cap \{km_j\} \neq \emptyset \\ [lo(t), \overline{lo}(t)]_l \cap \{lo_j\} \neq \emptyset \\ T_l(t) = T_j}} \nu_l \quad (47)$$

617

and

$$\begin{aligned}
bel(j, k) = & \sum_{\substack{l, \{km_k\} \subseteq [km(t+1), \overline{km}(t+1)]_l \\ \{km_j\} \subseteq [km(t), \overline{km}(t)]_l \\ \{lo_j\} \subseteq [lo(t), \overline{lo}(t)]_l \\ T_l(t) = T_j}} \nu_l \quad (48)
\end{aligned}$$

618

This kind information is linked to Sections 3.4, 4.3 and corresponds to Belief functions in Table 1.

619

620

### 3. Integration of expert knowledge, ( $S_3$ ).

621

622

In cheese ripening, as in every complex food process, most of the control measures are performed on the basis of the expert's sensory perceptions. Indeed, experts have in mind the ripening process that they oversee and they are able to explain part of the complex reactions through their perception of quality changes [16]. Expert elicitation [48] informs us that during the exponential growing of the yeast  $Km$ , a characteristic fresh or lactic odour is released. Mushroom odour appears when the concentration of the yeast  $Km$  begins to stabilize and typical Camembert odour appears when the population of  $Km$  begins to decay. From this qualitative information, general rules may be deduced such as "it is impossible to have a Camembert odour with a weak (*resp.* high) concentrations of  $Km$  (*resp.*  $lo$ )". That means for several combinations of  $Km$  and  $lo$  concentrations, likely values about variable *Odour* may be formalized by means of possibility distributions  $\pi_{Odour}(\cdot|Km, lo)$ . That is:

623

624

625

626

627

628

629

630

631

632

633

634

635

636

- When there is a high (*resp.* weak) concentration of lactose (*resp.* the yeast  $Km$ ), having a fresh odour is the most plausible state, followed by Mushroom and Camembert odours, which can be formalized as the following possibility distribution:

$$\pi_{Odour}(\text{Fresh}|j) = 1$$

$$\pi_{Odour}(\text{Mushroom}|j) = 0.8$$

$$\pi_{Odour}(\text{Camembert}|j) = 0.2$$

637

where  $j = (Km \leq 6.5, lo \geq 8)$ .

- When there is a medium concentration of lactose and  $Km$ , the Mushroom odour is the most plausible state but we cannot exclude having Fresh or Camembert odours, formalized by:

$$\pi_{Odour}(\text{Fresh}|j) = 0.2$$

$$\pi_{Odour}(\text{Mushroom}|j) = 1$$

$$\pi_{Odour}(\text{Camembert}|j) = 0.2$$

638 where  $j = (Km = 7, 2 < lo < 8)$ .

- When having very weak (*resp.* high) concentration of lactose (*resp.* the yeast  $Km$ ), the Camembert odour is the most plausible state, followed by Mushroom and Fresh odours, formalized by:

$$\begin{aligned}\pi_{Odour}(\text{Fresh}|j) &= 0.2 \\ \pi_{Odour}(\text{Mushroom}|j) &= 0.8 \\ \pi_{Odour}(\text{Camembert}|j) &= 1\end{aligned}$$

639 where  $j = (Km > 7, lo < 2)$ .

640 Parameter  $\mathbf{p}_6$  is then updated by using

$$p_{6jk} \in \left[ \frac{s_1 f_{6jk} + s_3 \xi_{6jk}}{s_0 + s_1 + s_3}, \frac{s_0 + s_1 f_{6jk} + s_3 \bar{\xi}_{6jk}}{s_0 + s_1 + s_3} \right] \quad (49)$$

641 where

$$\bar{\xi}_{6jk} = \pi_{Odour}(k|j) \text{ and } \xi_{6jk} = 1 - \max_{l \neq k} \pi_{Odour}(l|j) \quad (50)$$

642 This kind of information is linked to the Section 3.2 and corresponds  
643 to possibilistic model in Table 1.

## 644 5.2. Inference results and discussion

645 We attempt to estimate the lower and upper mean time evolution of  $Km$ ,  
646  $lo$  and  $Odour$  for a temperature control according to the previous parameter  
647 learning. That is

$$\underline{E}(X(t)|\mathbf{U}(t)) = \sum_k x_k \underline{p}(X(t) = x_k | \mathbf{U}(t)) \quad (51)$$

648 for the lower bounds where  $X$  may be  $Km$ ,  $lo$ ,  $Odour$ ;  $\mathbf{U}(t) = (lo(0), Km(0)$   
649  $, T(0), \dots, T(t))$  and

$$\underline{p}(X(t) = x_k | \mathbf{U}(t)) = \inf_{p \in K(X(t)|\mathbf{U}(t))} p(X(t) = x_k | \mathbf{U}(t)) \quad (52)$$

650 by assuming  $s_0 = s_1 = s_2 = s_3 = 1$  and the *repetitive independence*, since  
651 we assume that transition probabilities remain the same along the process  
652 (there is no reason to assume a change in the bacteria population behaviour),  
653 but are ill-known due to insufficient experiments and information. Figure  
654 4 displays the lower and upper simulated mean evolution of  $Km$ ,  $lo$ ,  $Odour$   
655 versus experimental data over the cheese ripening carried out at 12°C each  
656 time a source of information is added. Supported by Table 4, we may observe  
657 that the imprecision of simulated results well decreases (characterized by the  
658 surface in gray).

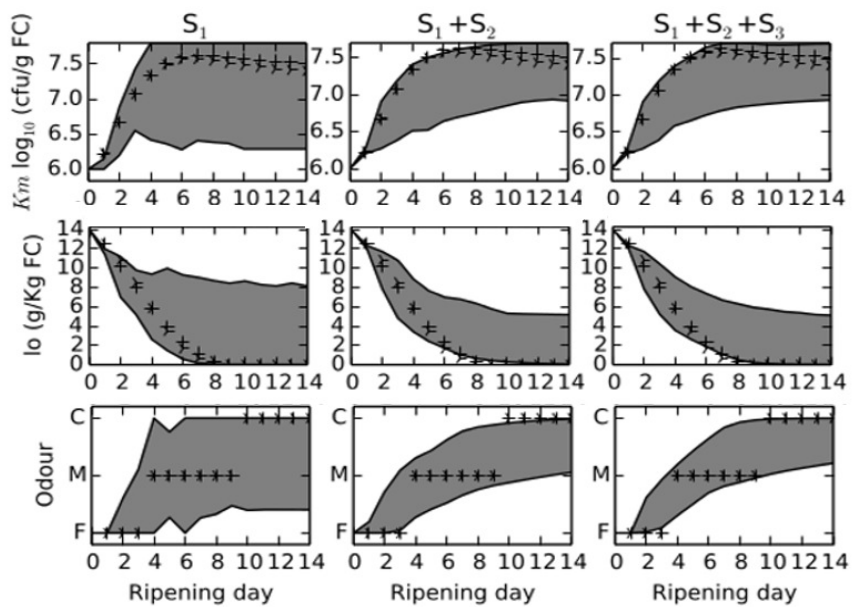


Figure 4: Incremental DCN average simulation versus raw data (dotted) of  $Km$ , lactose ( $lo$ ) and Odour for a ripening carried out at and  $T=12^{\circ}$  C each time a new source of information is integrated.

	$Km$	lactose	Odour
Source 1	18.62	105.38	20.91
Source 1 & 2	11.37	71.06	13.77
Source 1& 2& 3	11.03	71.32	10.99

Table 4: Area between the lower and upper bounds of the simulated mean time evolution

## 659 6. Conclusion

660 There are complex dynamical processes for which no deterministic model  
661 describing the complete process exists. In such cases, dynamic credal net-  
662 works are convenient models that allow to include expert knowledge, data  
663 and variable interaction in a single framework. They allow a faithful repre-  
664 sentation of incomplete knowledge or scarce data, that are inherent to the  
665 complexity of bio-physicochemical phenomena occurring in Life Sciences. In  
666 this paper, we attempted to implement a practical methodology coupling  
667 interval analysis and Dirichlet model in the framework of dynamical credal  
668 networks for building mathematical model capable of representing complex  
669 systems. Moreover, the concepts of dynamic repetitive and strong exten-  
670 sions have been proposed. While the latter can be seen as a straightforward  
671 extension of classical credal networks, the former considers repetitive inde-  
672 pendence to allow the model to preserve a temporal regularity. Methodology  
673 has been applied to a simplified real-case study concerning microbial popu-  
674 lation growth involving sensory evolution during cheese ripening. These ex-  
675 periments have shown that including information reduces imprecision about  
676 result simulations. Next tools should consider to manipulate, to combine  
677 convex sets in order to not lose information during incremental parameter  
678 learning. In further works, DCNs should enable us to determine the contri-  
679 bution of imprecision and/or incompleteness on the outcomes of a model in  
680 order to know if an ambiguous answer is due to a lack of information or due  
681 to a random phenomenon. That is, we plan to develop refined sensitivity  
682 analysis techniques based on their use. They should thus determine key  
683 variables and/or key phenomena for which it will be necessary to acquire  
684 more information. Finally, we also plan to investigate their usefulness in  
685 determining optimal commands.

## 686 References

- 687 [1] A. Antonucci and F. Cuzzolin. Credal sets approximation by lower  
688 probabilities: application to credal networks. In *Computational Intel-*  
689 *ligence for Knowledge-Based Systems Design*, pages 716–725. Springer,  
690 2010.
- 691 [2] T. Augustin, F.P.A Coolen, G. de Cooman, and M.C.M. Troffaes. *In-*  
692 *troduction to imprecise probabilities*. John Wiley & Sons, 2014.
- 693 [3] D. Barba, F. Beolchini, G. Del Re, G. Di Giacomo, and F. Veglió.  
694 Kinetic analysis of *kluveromyces lactis* fermentation on whey: batch  
695 and fed-batch operations. *Process Biochemistry*, 36(6):531 – 536, 2001.
- 696 [4] C. Baudrit, I. Couso, and D. Dubois. Joint propagation of probability  
697 and possibility in risk analysis: Towards a formal framework. *Interna-*  
698 *tional Journal of Approximate Reasoning*, 45(1):82–105, 2007.

- 699 [5] C. Baudrit, M. Sicard, P.H. Wuillemin, and N. Perrot. Towards a global  
700 modelling of the camembert-type cheese ripening process by coupling  
701 heterogeneous knowledge with dynamic bayesian networks. *Journal of*  
702 *Food Engineering*, 98(3):283–29, 2010.
- 703 [6] C. Baudrit, P.H. Wuillemin, and N. Perrot. Parameter elicitation in  
704 probabilistic graphical models for modelling multi-scale food complex  
705 systems. *Journal of Food Engineering*, 115(1):1 – 10, 2013.
- 706 [7] J.M. Bernard. An introduction to the imprecise Dirichlet model for  
707 multinomial data. *International Journal of Approximate Reasoning*,  
708 39(2-3):123–150, June 2005.
- 709 [8] J.M. Bernard. The imprecise Dirichlet model. *Int. J. Approx. Reason-*  
710 *ing*, 50(2):201–203, 2009.
- 711 [9] J. De Bock and G. De Cooman. State sequence prediction in imprecise  
712 hidden markov models. In *Proceedings of the seventh International*  
713 *Symposium on Imprecise Probabilities: Theory and Applications*, pages  
714 159–168, 2011.
- 715 [10] W. Buntine. A guide to the literature on learning probabilistic networks  
716 from data. *Knowledge and Data Engineering, IEEE Transactions on*,  
717 8(2):195–210, 1996.
- 718 [11] L.M. De Campos, J.F. Huete, and S. Moral. Probability intervals :  
719 a tool for uncertain reasoning. *International Journal of Uncertainty*,  
720 2(2):167 – 196, 1994. Fuzziness and Knowledge-Based Systems.
- 721 [12] A. Cano, J. Cano, and S. Moral. Convex sets of probabilities propaga-  
722 tion by simulated annealing on a tree of cliques. In *In: Proceedings of*  
723 *Fifth International Conference on Processing and Management of Un-*  
724 *certainty in Knowledge-Based Systems (IPMU 1994*, pages 4–8, 1994.
- 725 [13] A. Cano, M. Gómez, S. Moral, and J. Abellán. Hill-climbing and  
726 branch-and-bound algorithms for exact and approximate inference in  
727 credal networks. *Int. J. Approx. Reasoning*, 44(3):261–280, 2007.
- 728 [14] A. Cano and S. Moral. A genetic algorithm to approximate convex sets  
729 of probabilities. In *Proc. of the Int. Conference on Information Pro-*  
730 *cessing and Management of Uncertainty in Knowledge-Based Systems*,  
731 pages 859–864, 1996.
- 732 [15] A. Cano and S. Moral. Using probability trees to compute marginals  
733 with imprecise probabilities. *International Journal of Approximate Rea-*  
734 *soning*, 29(1):1 – 46, 2002.

- 735 [16] W. G. Chase and H. A. Simon. Perception in chess. *Cognitive Psychol-*  
736 *ogy*, 4(1):55 – 81, 1973.
- 737 [17] A. Chateauneuf and J.-Y. Jaffray. Some characterizations of lower prob-  
738 abilities and other monotone capacities through the use of möbius in-  
739 version. *Mathematical social sciences*, 17(3):263–283, 1989.
- 740 [18] I. Couso and D. Dubois. On the Variability of the Concept of Variance  
741 for Fuzzy Random Variables. *IEEE Transactions on Fuzzy Systems*,  
742 17(5):1070–1080, 2009.
- 743 [19] I. Couso, D. Dubois, and L. Sánchez. *Random Sets and Random Fuzzy*  
744 *Sets as Ill-Perceived Random Variables: An Introduction for Ph. D.*  
745 *Students and Practitioners*. Springer, 2014.
- 746 [20] I. Couso, S. Moral, and P. Walley. A survey of concepts of independence  
747 for imprecise probabilities. *Risk Decision and Policy*, 5:165–181, 2000.
- 748 [21] I. Couso and L. Sanchez. Higher order models for fuzzy random vari-  
749 ables. *Fuzzy Sets and Systems*, 159:237–258, 2008.
- 750 [22] I. Couso and L. Sánchez. Upper and lower probabilities induced by a  
751 fuzzy random variable. *Fuzzy Sets and Systems*, 165(1):1 – 23, 2011.  
752 Theme: Fuzzy intervals and applications.
- 753 [23] F. Cozman. Credal networks. *Artificial Intelligence*, 120(2):199–233,  
754 2000.
- 755 [24] F. Cozman. Graphical models for imprecise probabilities. *International*  
756 *Journal of Approximate Reasoning*, 39(2-3):167–184, June 2005.
- 757 [25] F. Cozman. Separation properties of sets of probability measures.  
758 *CoRR*, abs/1301.3845, 2013.
- 759 [26] J.C.F. da Rocha and F.G. Cozman. Inference with separately specified  
760 sets of probabilities in credal networks. In *Proceedings of the Eighteenth*  
761 *conference on Uncertainty in artificial intelligence*, UAI’02, pages 430–  
762 437, San Francisco, CA, USA, 2002. Morgan Kaufmann Publishers Inc.
- 763 [27] J.C.F. da Rocha, F.G. Cozman, and C.P. de Campos. Inference in  
764 polytrees with sets of probabilities. In *Proceedings of the Nineteenth*  
765 *conference on Uncertainty in Artificial Intelligence*, UAI’03, pages 217–  
766 224, San Francisco, CA, USA, 2003. Morgan Kaufmann Publishers Inc.
- 767 [28] C.P. de Campos and F.G. Cozman. Inference in credal networks us-  
768 ing multilinear programming. In *Proceedings of the 2nd Starting AI*  
769 *Researchers Symposium*, pages 50–61, 2004.

- 770 [29] C.P. de Campos and F.G. Cozman. Inference in credal networks through  
771 integer programming. In *Proceedings of the Fifth International Symposi-*  
772 *um on Imprecise Probability: Theories and Applications*, pages 145–  
773 154, 2007.
- 774 [30] A.P. Dempster. Upper and lower probabilities induced by a multivalued  
775 mapping. *Annals of Mathematical Statistics*, 38:325–339, 1967.
- 776 [31] S. Destercke, D. Dubois, and E. Chojnacki. Unifying practical uncer-  
777 tainty representations. ii: Clouds. *International Journal of Approx-*  
778 *imate Reasoning*, 49(3):664–677, 2008.
- 779 [32] D. Dubois. Uncertainty theories: a unified view. In *Cybernetic Systems,*  
780 *Dublin (Ireland)*, pages 4–94, <http://www.ieee.org/>, 2007. IEEE.
- 781 [33] D. Dubois, H.T. Nguyen, and H. Prade. Possibility theory, probability  
782 and fuzzy sets: misunderstandings, bridges and gaps. . In D. Dubois  
783 and H. Prade, editors, *Fundamentals of Fuzzy Sets*, The Handbooks of  
784 Fuzzy Sets Series, pages 343–438. Kluwer, Boston, Mass., 2000.
- 785 [34] E. Fagioli and M. Zaffalon. 2U: an exact interval propagation al-  
786 gorithm for polytrees with binary variables. *Artificial Intelligence*,  
787 106(1):77 – 107, 1998.
- 788 [35] S. Ferson, L. Ginzburg, V. Kreinovich, D.M. Myers, and K. Sentz. Con-  
789 structing probability boxes and dempster-shafer structures. Technical  
790 report, Sandia National Laboratories, 2003.
- 791 [36] S. Ferson and L. R. Ginzburg. Different methods are needed to propa-  
792 gate ignorance and variability. *Reliability Engineering & System Safety*,  
793 54(2-3):133 – 144, 1996. Treatment of Aleatory and Epistemic Uncer-  
794 tainty.
- 795 [37] P.F. Fox. *Cheese - Chemistry, Physics and Microbiology; 3rd ed.* Else-  
796 vier, San Diego, CA, 2004.
- 797 [38] J.E. Gentle. Monte carlo methods. *Encyclopedia of Statistical Sciences*,  
798 2006.
- 799 [39] B. Grünbaum. Convex polytopes. 1967. *Interscience, New York*, 1967.
- 800 [40] J.C. Helton and W.L. Oberkampf. Alternative representations of epis-  
801 temic uncertainty. *Reliability Engineering & System Safety*, 85(1-3):1 –  
802 10, 2004. Alternative Representations of Epistemic Uncertainty.
- 803 [41] M. Hourbracq, P.H. Wuillemin C. Baudrit, and S. Destercke. Dynamic  
804 credal networks: introduction and use in robustness analysis. In F. Coz-  
805 man, T. Dencoux, S. Destercke, and T. Seidenfeld, editors, *ISIPTA '13:*



- 806 *Proceedings of the Eighth International Symposium on Imprecise Prob-*  
807 *ability: Theories and Applications*, pages 159–168, Compiègne, 2013.  
808 SIPTA.
- 809 [42] J.S. Ide and F.G. Cozman. Approximate Inference in Credal Networks  
810 by Variational Mean Field Methods. In *International Symposium on*  
811 *Imprecise Probabilities and Their Applications*, pages 203–212, 2005.
- 812 [43] J.S. Ide and F.G. Cozman. Approximate algorithms for credal networks  
813 with binary variables. *International Journal of Approximate Reasoning*,  
814 48(1):275 – 296, 2008. Special Section: Perception Based Data Mining  
815 and Decision Support Systems.
- 816 [44] J.S. Ide and Cozman F.G. Ipe and l2u: Approximate algorithms for  
817 credal networks. In *Proceedings of the second starting AI Researcher*  
818 *Symposium*, pages 118–127. IOS Press, 2004.
- 819 [45] S.L. Lauritzen. The em algorithm for graphical association models with  
820 missing data. *Computational Statistics & Data Analysis*, 19(2):191–201,  
821 1995.
- 822 [46] M.N. Leclercq-Perlat, F. Buono, D. Lambert, E. Latrille, E. Spinnler,  
823 and G. Corrieu. Controlled production of camembert-type cheeses. part  
824 i: Microbiological and physicochemical evolutions. *Journal of Dairy*  
825 *Research*, 71:346–354, 8 2004.
- 826 [47] I. Levi. *The Enterprise of Knowledge: An Essay on Knowledge, Credal*  
827 *Probability, and Chance*. MIT press, 1983.
- 828 [48] S. Mariette, C. Baudrit, M.N. Leclerc-Perlat, P.H. Willemin, and  
829 N. Perrot. Expert knowledge integration to model complex food pro-  
830 cesses. application on the camembert cheese ripening process. *Expert*  
831 *Syst. Appl.*, 38(9):11804–11812, 2011.
- 832 [49] D.D. Mauá, C.P. de Campos, A. Benavoli, and A. Antonucci. Proba-  
833 bilistic inference in credal networks: new complexity results. *Journal*  
834 *of Artificial Intelligence Research*, 50(1):603–637, 2014.
- 835 [50] E. Miranda, I. Couso, and P. Gil. Extreme points of credal sets gener-  
836 ated by 2-alternating capacities. *International Journal of Approximate*  
837 *Reasoning*, 33(1):95–115, 2003.
- 838 [51] K. Murphy. *Dynamic Bayesian Networks: Representation, Inference*  
839 *and Learning*. PhD thesis, UC Berkeley, Computer Science Division,  
840 July 2002.
- 841 [52] N. Perrot, L. Agioux, I. Ioannou, G. Mauris, G. Corrieu, and G. Trys-  
842 tram. Decision support system design using the operator skill to control

- 843 cheese ripening application of the fuzzy symbolic approach. *Journal of*  
844 *Food Engineering*, 64(3):321 – 333, 2004.
- 845 [53] N. Perrot, I.C. Trelea, C. Baudrit, G. Trystram, and P. Bourguine. Mod-  
846 elling and analysis of complex food systems: State of the art and new  
847 trends. *Trends in Food Science & Technology*, 22(6):304 – 314, 2011.
- 848 [54] A. Piatti, A. Antonucci, and M. Zaffalon. Building knowledge-based  
849 systems by credal networks: a tutorial. *Advances in Mathematics Re-*  
850 *search*, 11, 2010.
- 851 [55] J.C.F. Rocha and F.G. Cozman. Evidence propagation in credal net-  
852 works: An exact algorithm based on separately specified sets of prob-  
853 ability. In G. Bittencourt and G.L. Ramalho, editors, *Advances in*  
854 *Artificial Intelligence*, volume 2507, pages 376–385. 2002.
- 855 [56] G. Schollmeyer. On the number and characterization of the extreme  
856 points of the core of necessity measures on finite spaces. In *ISIPTA '15:*  
857 *Proceedings of the Ninth International Symposium on Imprecise Prob-*  
858 *ability: Theories and Applications*, pages 277–286, Pescara, 2015.  
859 SIPTA.
- 860 [57] G. Shafer. *A mathematical Theory of Evidence*. Princeton University  
861 Press, New Jersey, 1976.
- 862 [58] L.M.M. Tijskens, L.A.T.M. Hertog, and B.M. Nicolai. *Food Process*  
863 *Modelling*. Woodhead Publishing Series in Food Science, Technology  
864 and Nutrition. Woodhead, 2001.
- 865 [59] M. C. M. Troffaes and S. Destercke. Probability boxes on totally pre-  
866 ordered spaces for multivariate modelling. *International Journal of*  
867 *Approximate Reasoning*, 52(6):767–791, 2011.
- 868 [60] L.V. Utkin. Probabilities of judgments provided by unknown experts  
869 by using the imprecise dirichlet model. In *Risk, Decision and Policy*,  
870 9(4):391 - 400, page 400, 2004.
- 871 [61] P. Walley. *Statistical reasoning with imprecise probabilities*. Chapman  
872 and Hall London, 1991.
- 873 [62] P. Walley. Inferences from multinomial data: learning about a bag of  
874 marbles. *Journal of the Royal Statistical Society. Series B (Method-*  
875 *ological)*, 58(1):3–57, 1996.
- 876 [63] J. S. Yedidia, W. T. Freeman, and Y. Weiss. Generalized belief propa-  
877 gation. In *NIPS*, volume 13, pages 689–695, 2000.
- 878 [64] L.A. Zadeh. Fuzzy sets. *Information Control*, 8:338–353, 1965.

879 **Vitae**

Cédric Baudrit is a researcher for the French Nationale Institute for Agricultural Research (INRA), in knowledge representation and reasoning. He is interested in the development of mathematical tools capable of (1) integrating fragmented heterogeneous knowledge stemming from different sources; (2) taking into account stochastic and epistemic uncertainty in order to model global complex system. In 2005, he received the Ph.D. degree in computer science from Université Paul Sabatier, in Toulouse (France). In 2002, he holds a master in Applied Mathematics and Computer Science from the University of Orléans, France.



Destercke Sebastien graduated in 2004 as an engineer from the Faculté Polytechnique de Mons in Belgium. In 2008, he earned a Ph.D. degree in computer science from Université Paul Sabatier, in Toulouse (France). He now works as a researcher for the French National Research Center (CNRS), in the Heuristic and Diagnostic of Complex Systems (Heudiasyc) Laboratory. His main research interests are in the field of uncertainty reasoning with imprecision-tolerant models (DS theory, imprecise probabilities, possibility theory, ...), with a focus on issues related to reliability and risk analysis, decision making and machine learning..



Pierre-Henri Wullemmin is Associate Professor in the DESIR (Decision, Intelligent System and operational Research) department of LIP6 (UPMC, CNRS). His research field concerns Graphical Models for uncertainty reasoning and decision making. He is interested both in theoretical aspects - e.g. improving inference algorithms and PGM formulation, structure and parameters learning methods, credal networks - and in applications - e.g. object oriented framework for modeling PGM, sensitivity analysis, troubleshooting, (PO)MDPs.