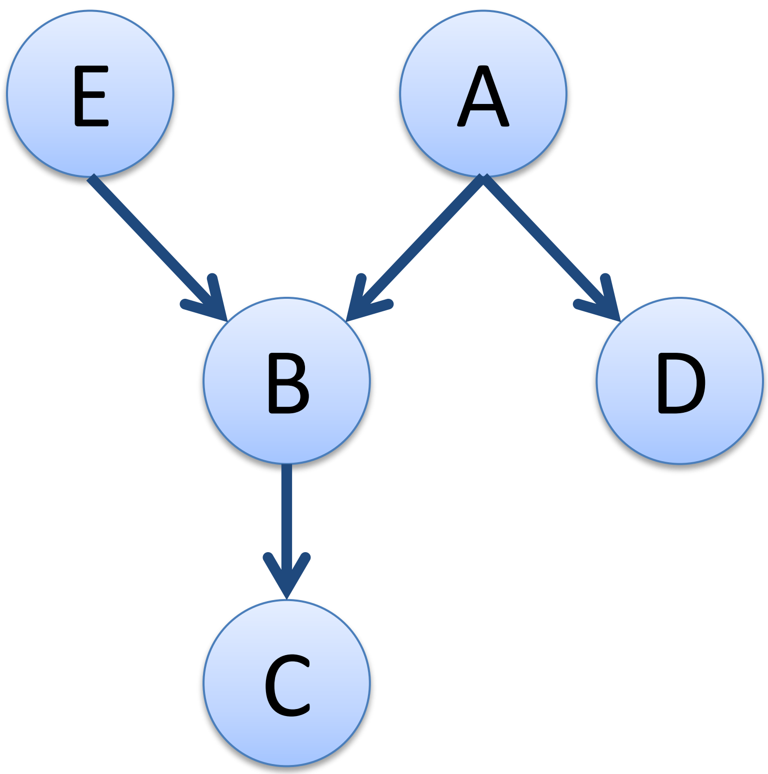
An Introduction to Bayesian Networks

Alberto Tonda, PhD
Researcher at Team MALICES

INRA
SCIENCE & IMPACT

Objective

Basic understanding of what **Bayesian Networks** are, and where they can be applied.
Example from food science.



Outline

- Introduction
- Basic concepts of probability
- Bayesian Networks
- A case study: Camembert cheese ripening

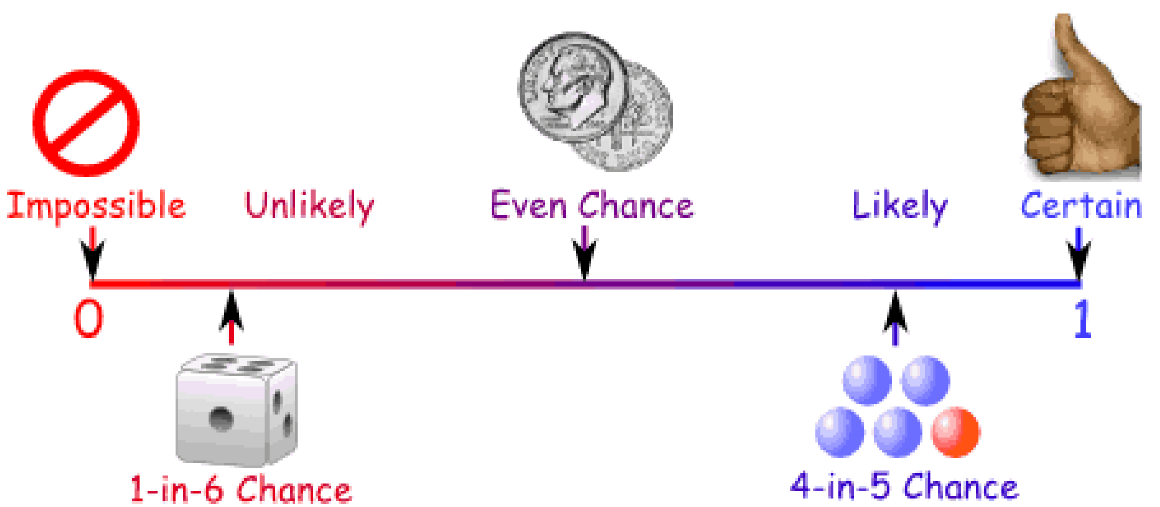
Link to slides: **http://goo.gl/bvwM6O**

Introduction

- Why should you care about Bayesian Networks (BNs)?
  - Probabilistic models
  - Understandable by humans
  - Built from data and human expertise
  - Include both quantitative and qualitative variables

Introduction

- BNs are **probabilistic models**
  - Instead of a unique response…
  - …you get the *probability* of an outcome
  - They can work with *incomplete* information!



Introduction

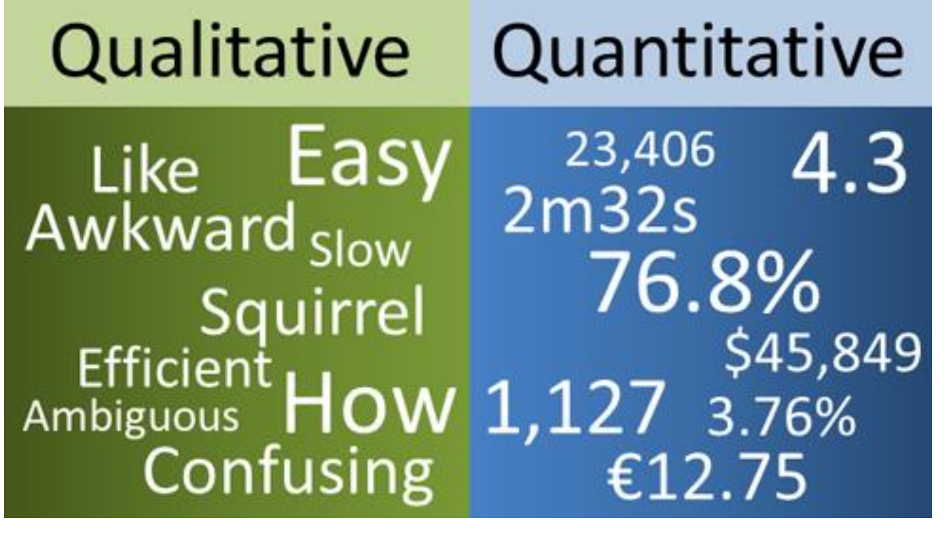
- BNs can be **understood by humans**
  - Graphical models
  - Arcs representing relationships between variables
  - Other models are “black boxes” (e.g., NN)

Introduction

- BNs can be **built automatically** or **manually**
  - By algorithms, starting from experiments
  - By experts, using their knowledge
  - Both: built by algorithm, validated by expert


Introduction

- **Qualitative** and **quantitative** variables
  - In the same network!
  - Link the *flavor* to the *concentration* in microbes
  - Extremely useful for *complex systems*

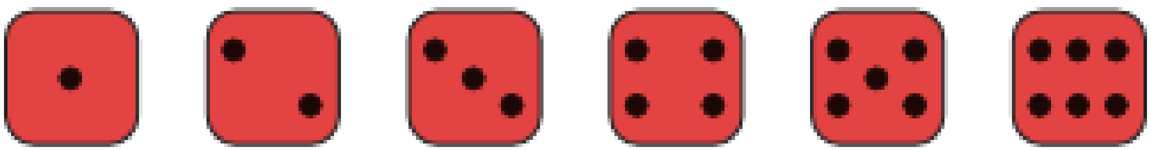


Introduction

- Applications, applications everywhere
  - Classification (anti-spam filters, diagnostics, …)
  - Modeling (simulations, predictions, modeling of players, …)
  - Engineering, gaming, law, medicine, risk analysis, finance, computational biology, bio-informatics…

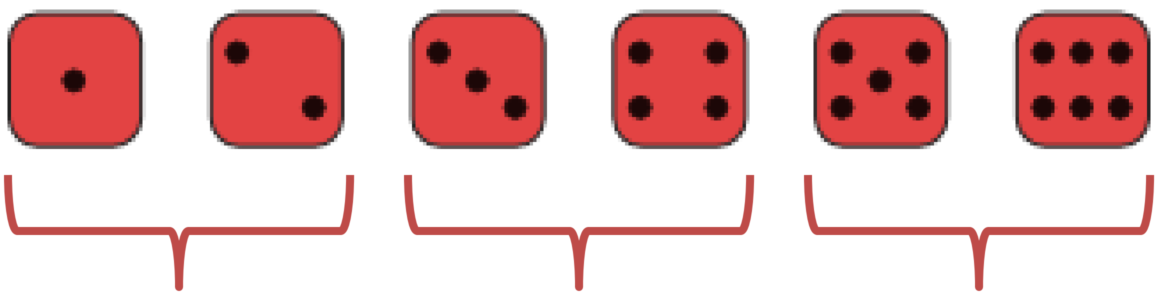
Basic concepts of probability

- Probabilities for **discrete** events
  - Rolling a die! (result *d*)
  - Probabilities for (1 or 2), (3 or 4), (5 or 6)?



Basic concepts of probability

- Probabilities for **discrete** events
  - Rolling a die! (result *d*)
  - Probabilities for (1 or 2), (3 or 4), (5 or 6)?



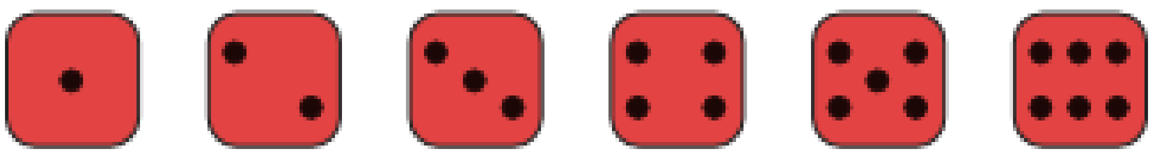
P(d=**1or2**) 2/6 -> 1/3

P(d=**3or4**) 2/6 -> 1/3

P(d=**5or6**) 2/6 -> 1/3

Basic concepts of probability

- Probabilities for **discrete** events
  - Rolling a die! (result *d*)
  - Probabilities for (1 or 2), (3 or 4), (5 or 6)?



P(d=**1or2**) 2/6 -> 1/3

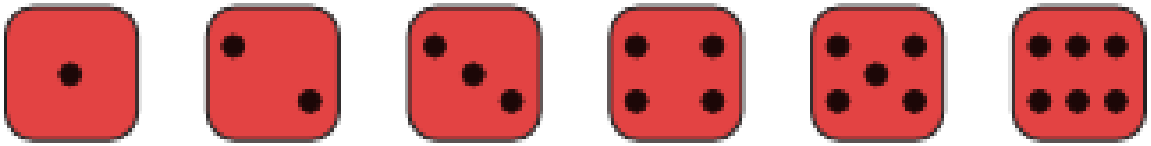
P(d=**3or4**) 2/6 -> 1/3

P(d=**5or6**) 2/6 -> 1/3

P(d=**1or2**) + P(d=**3or4**) + P(d=**5or6**) = 1

Basic concepts of probability

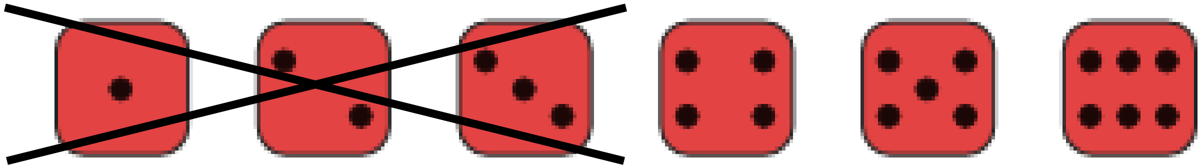
- Conditional probability
  - Probability for any of the 3 events is 33%
  - Would that change with **more information**?



Event	Probability
d=1or2	0.33
d=3or4	0.33
d=5or6	0.33

Basic concepts of probability

- Conditional probability
  - For example, what if we knew that the result *d* was bigger than 3?



Event	Probability
d=1or2	**??**
d=3or4	**??**
d=5or6	**??**

Basic concepts of probability

- Conditional probability
  - For example, what if we knew that the result *d* was bigger than 3?

$$P(d{=}1or2\,|\,\mathbf{d>3}) = 0$$
$$P(d{=}3or4\,|\,\mathbf{d>3}) = 0.33$$
$$P(d{=}5or6\,|\,\mathbf{d>3}) = 0.66$$

Event	Probability
d=1or2	**0**
d=3or4	0.33
d=5or6	**0.66**

Basic concepts of probability

- Combining P
  - Yeast concentration (Y)
  - Bact. Concentration (B)
  - Aroma (A)

- Parameters
  - 2 x 2 x 3

$$\sum_{i,j\,k} P(Y = i, B = j, A = k) = 1$$

Yeast (Y)	Bacteria (B)	Aroma (A)	P
Weak	Weak	Strawberry	0.2
Weak	Weak	Camembert	0.05
Weak	Weak	Ammonia	0.005
Weak	High	Strawberry	0.005
Weak	High	Camembert	0.05
Weak	High	Ammonia	0.2
High	Weak	Strawberry	0.05
High	Weak	Camembert	0.1
High	Weak	Ammonia	0.005
High	High	Strawberry	0.005
High	High	Camembert	0.1
High	High	Ammonia	0.23

Basic concepts of probability

- P(Y,B|**A=Strawberry**)

Yeast (Y)	Bacteria (B)	Aroma (A)	P
Weak	Weak	Strawberry	0.2
~~Weak~~	~~Weak~~	~~Camembert~~	~~0.05~~
~~Weak~~	~~Weak~~	~~Ammonia~~	~~0.005~~
Weak	High	Strawberry	0.005
~~Weak~~	~~High~~	~~Camembert~~	~~0.05~~
~~Weak~~	~~High~~	~~Ammonia~~	~~0.2~~
High	Weak	Strawberry	0.05
~~High~~	~~Weak~~	~~Camembert~~	~~0.1~~
~~High~~	~~Weak~~	~~Ammonia~~	~~0.005~~
High	High	Strawberry	0.005
~~High~~	~~High~~	~~Camembert~~	~~0.1~~
~~High~~	~~High~~	~~Ammonia~~	~~0.23~~

Basic concepts of probability

- P(Y,B|**A=Strawberry**)

Y	B	P
Weak	Weak	0.2 / 0.26 = **0.769**
Weak	High	0.005 / 0.26 = **0.019**
High	Weak	0.05 / 0.26 = **0.193**
High	High	0.005 / 0.26 = **0.019**

Yeast (Y)	Bacteria (B)	Aroma (A)	P
Weak	Weak	Strawberry	0.2
~~Weak~~	~~Weak~~	~~Camembert~~	~~0.05~~
~~Weak~~	~~Weak~~	~~Ammonia~~	~~0.005~~
Weak	High	Strawberry	0.005
~~Weak~~	~~High~~	~~Camembert~~	~~0.05~~
~~Weak~~	~~High~~	~~Ammonia~~	~~0.2~~
High	Weak	Strawberry	0.05
~~High~~	~~Weak~~	~~Camembert~~	~~0.1~~
~~High~~	~~Weak~~	~~Ammonia~~	~~0.005~~
High	High	Strawberry	0.005
~~High~~	~~High~~	~~Camembert~~	~~0.1~~
~~High~~	~~High~~	~~Ammonia~~	~~0.23~~

Basic concepts of probability

- Bayes' Theorem

$$P(H \mid E) = \frac{P(E \mid H) \cdot P(H)}{P(E)}$$

- Syntax
  - H = Hypothesis
  - E = Evidence
- Meaning: belief in H before and after taking into account E
- In many practical cases $P(H|E) \propto P(E|H) \cdot P(H)$

Basic concepts of probability

- Bayes' Theorem: Example*
  – Three production machines A1, A2, A3
  – Probability of having a piece produced by An
    - $P(A1) = 0.2$; $P(A2) = 0.3$; $P(A3) = 0.5$
  – Probability of a defective piece
    - $P(D|A1) = 0.05$; $P(D|A2) = 0.03$; $P(D|A3) = 0.01$
  – What is the probability of $P(A3|D)$?

*from Wikipedia

Basic concepts of probability

- Bayes' Theorem: Example*
  - Three production machines A1, A2, A3
  - Probability of having a piece produced by An
    - P(A1) = 0.2 ; P(A2) = 0.3; P(A3) = 0.5
  - Probability of a defective piece
    - P(D|A1) = 0.05; P(D|A2) = 0.03; P(D|A3) = 0.01
  - What is the probability of P(A3|D)?

$$P(H \mid E) = \frac{P(E \mid H) \cdot P(H)}{P(E)}$$

*from Wikipedia

Basic concepts of probability

- Bayes' Theorem: Example*
  - Three production machines A1, A2, A3
  - Probability of having a piece produced by An
    - P(A1) = 0.2 ; P(A2) = 0.3; P(A3) = 0.5
  - Probability of a defective piece
    - P(D|A1) = 0.05; P(D|A2) = 0.03; P(D|A3) = 0.01
  - What is the probability of P(A3|D)?

$$P(H \mid E) = \frac{P(E \mid H) \cdot P(H)}{P(E)}$$

*from Wikipedia

Basic concepts of probability

- Bayes' Theorem: Example*
  - Three production machines A1, A2, A3
  - Probability of having a piece produced by An
    - P(A1) = 0.2 ; P(A2) = 0.3; P(A3) = 0.5
  - Probability of a defective piece
    - P(D|A1) = 0.05; P(D|A2) = 0.03; P(D|A3) = 0.01
  - What is the probability of P(A3|D)?

P(D) = P(D|A1) * P(A1) +
P(D|A2) * P(A2) +
P(D|A3) * P(A3) =
0.024

$$\cdots) = \frac{P(E \mid H) \cdot P(H)}{P(E)}$$

*from Wikipedia

Basic concepts of probability

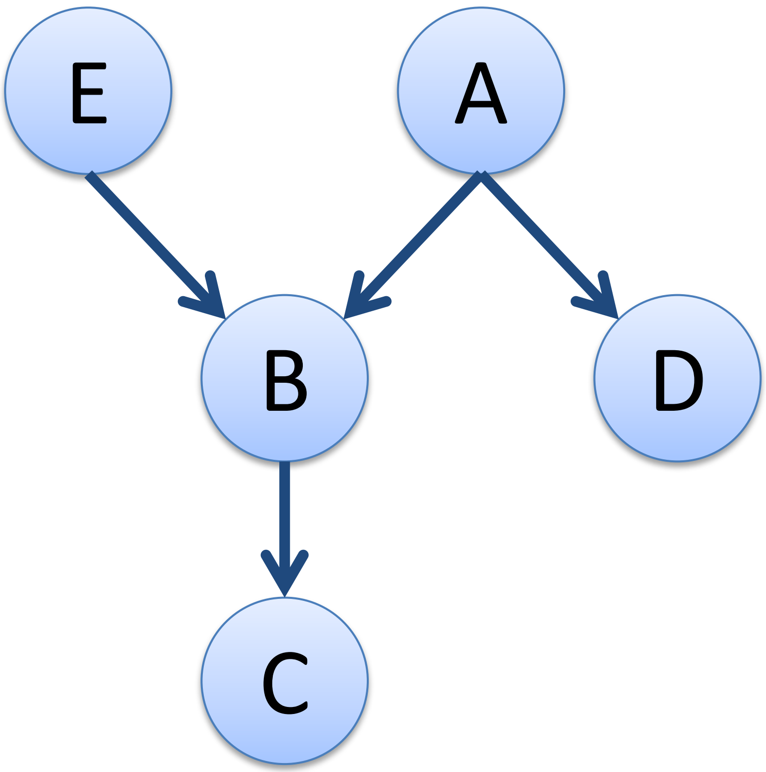
- Bayes' Theorem: Example*
  - Three production machines A1, A2, A3
  - Probability of having a piece produced by An
    - P(A1) = 0.2 ; P(A2) = 0.3; P(A3) = 0.5
  - Probability of a defective piece
    - P(D|A1) = 0.05; P(D|A2) = 0.03; P(D|A3) = 0.01
  - What is the probability of P(A3|D)?

P(D) = P(D|A1) * P(A1) + P(D|A2) * P(A2) + P(D|A3) * P(A3) = **0.024**

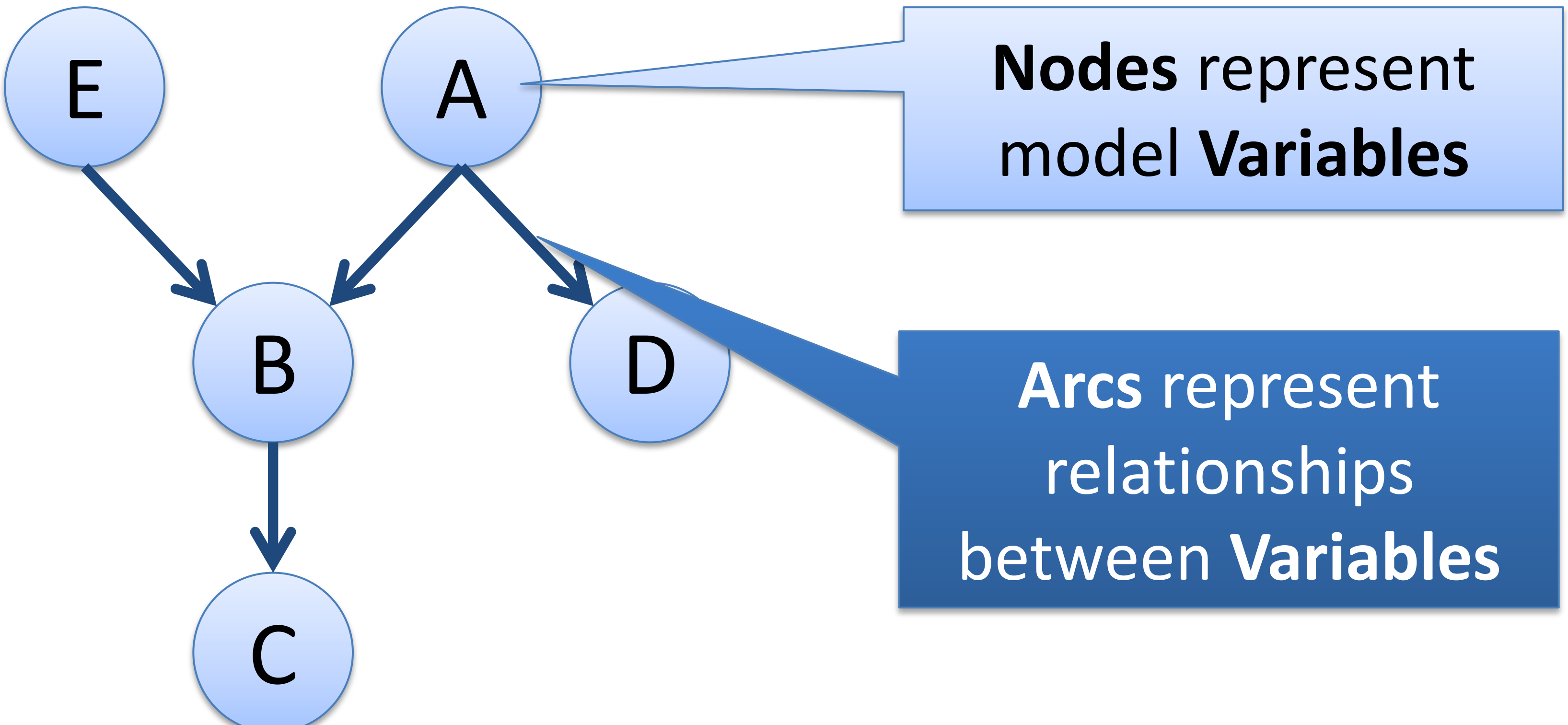
$$\ldots E) = \frac{P(E \mid H)\, P(H)}{P(E)} = \frac{0.01 * 0.5}{0.024} = 0.21$$

*from Wikipedia

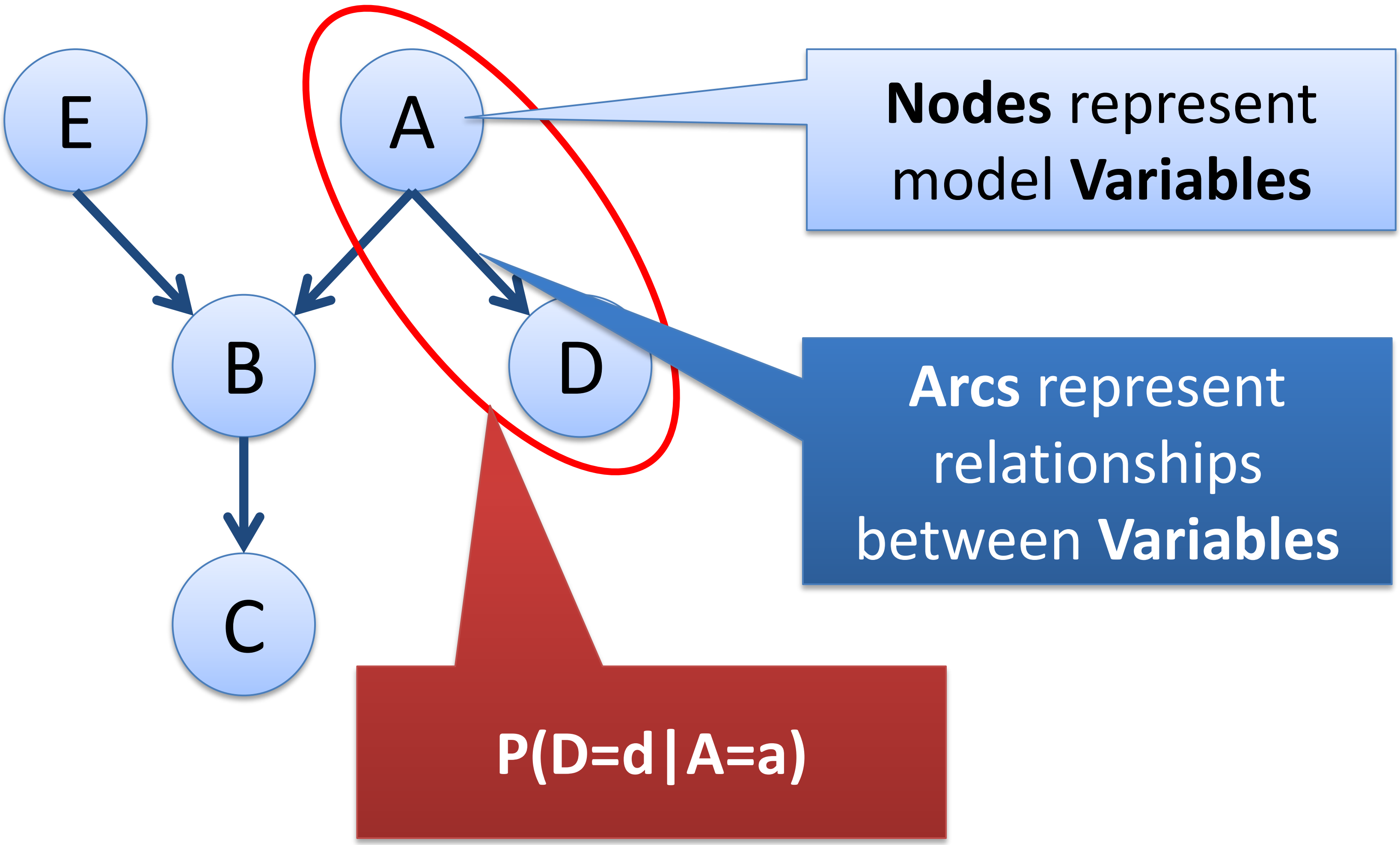
Bayesian Networks



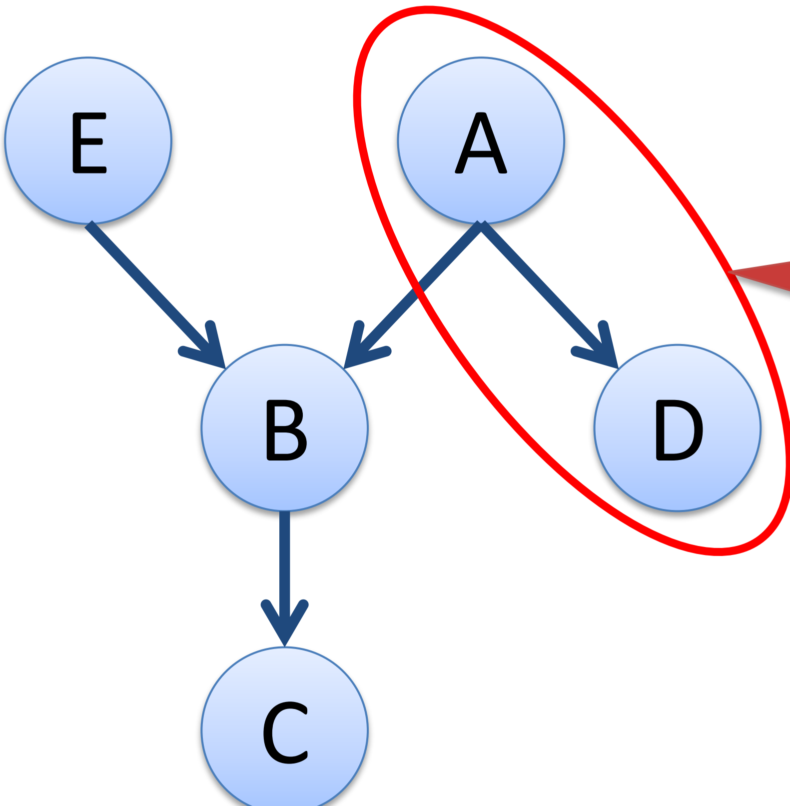
Bayesian Networks



Bayesian Networks

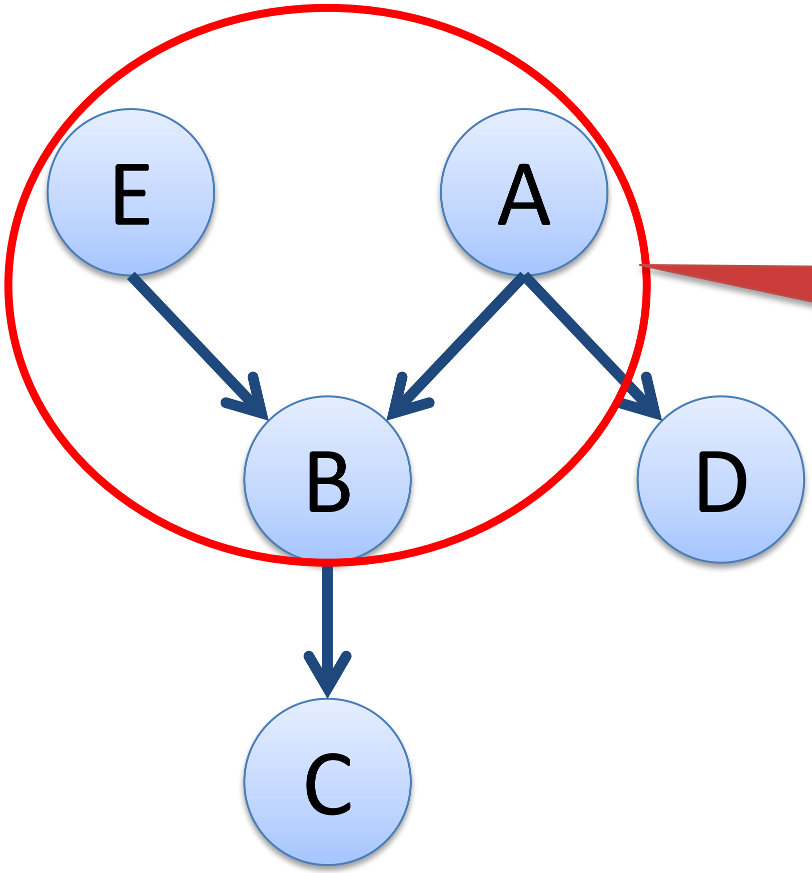


Bayesian Networks



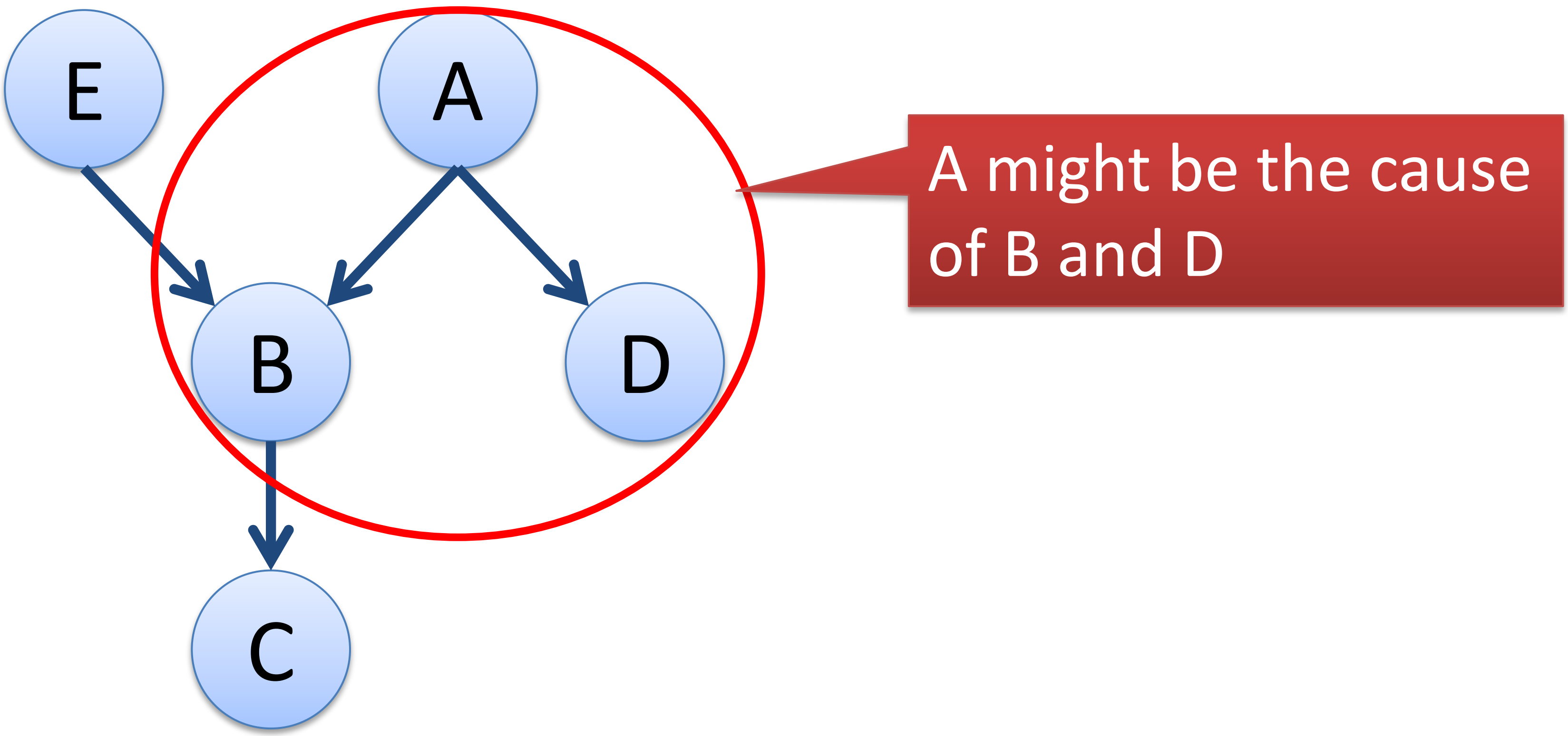
This does not imply that D depends on A; just that we know or *suspect* a connection

Bayesian Networks

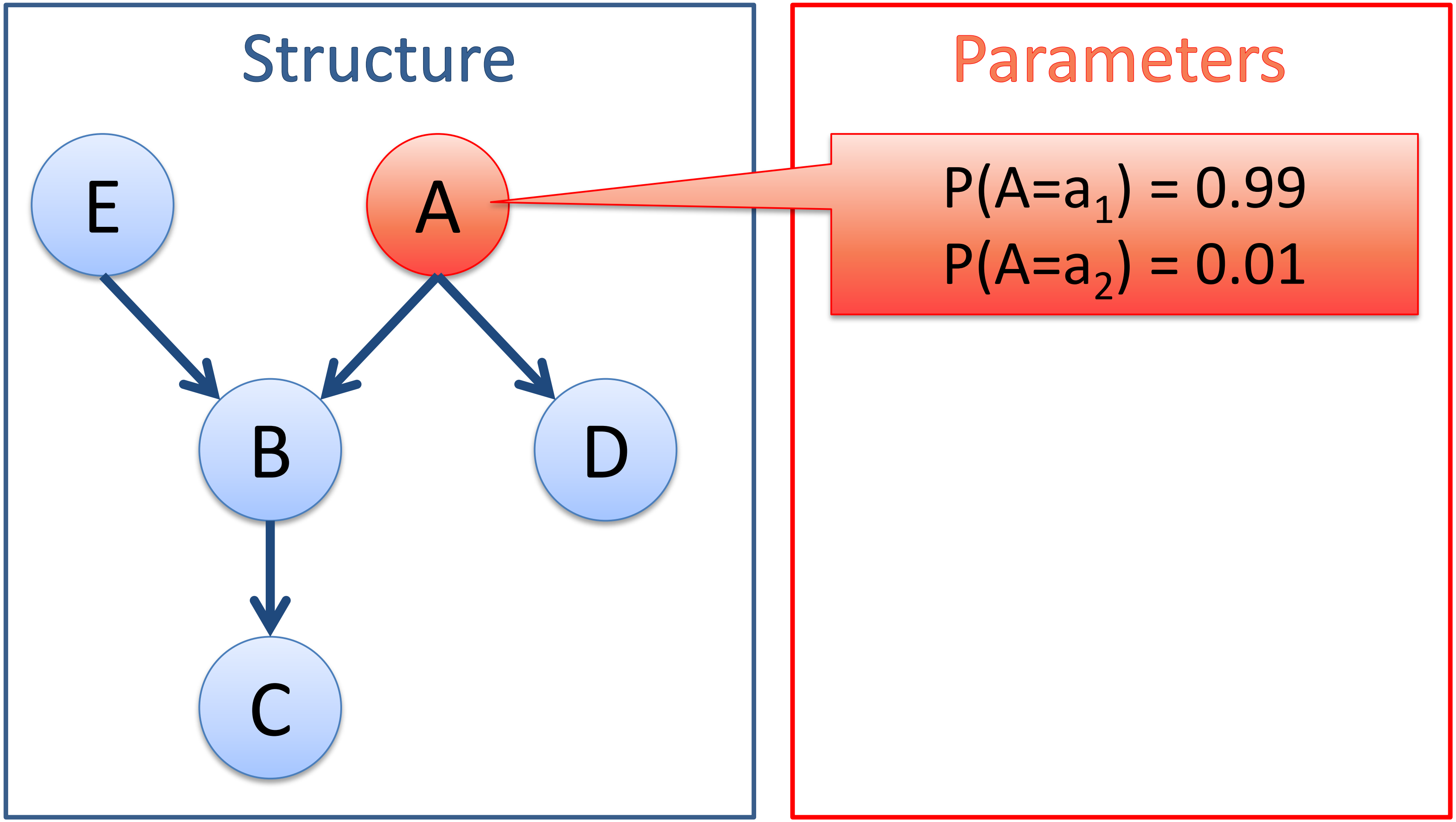


B has multiple possible causes, in this case E and A.

Bayesian Networks

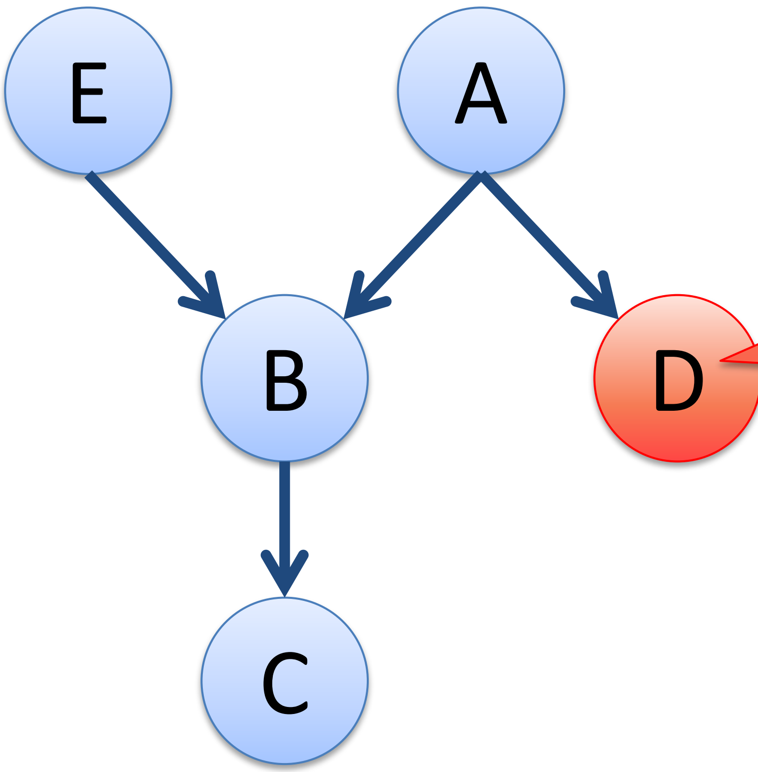


Bayesian Networks



Bayesian Networks

Structure



Parameters

$P(D=d_1|A=a_1) = 0.8$

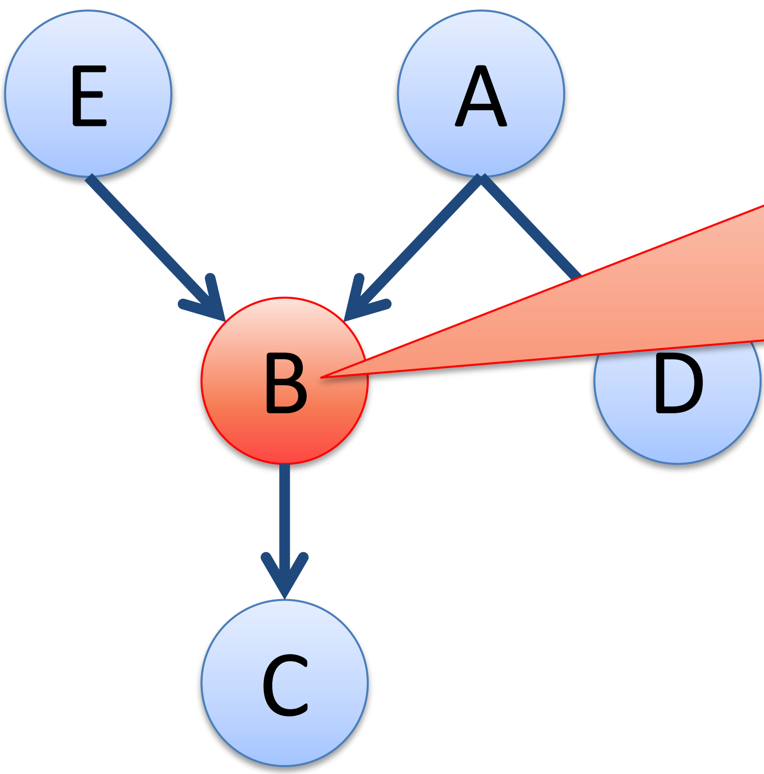
$P(D=d_2|A=a_1) = 0.2$

$P(D=d_1|A=a_2) = 0.7$

$P(D=d_2|A=a_1) = 0.3$

Bayesian Networks

Structure



Parameters

$P(B=b_1|A=a_1,E=e_1) = 0.5$

$P(B=b_2|A=a_1,E=e_1) = 0.5$

$P(B=b_1|A=a_1,E=e_2) = 0.9$

$P(B=b_2|A=a_1,E=e_2) = 0.1$

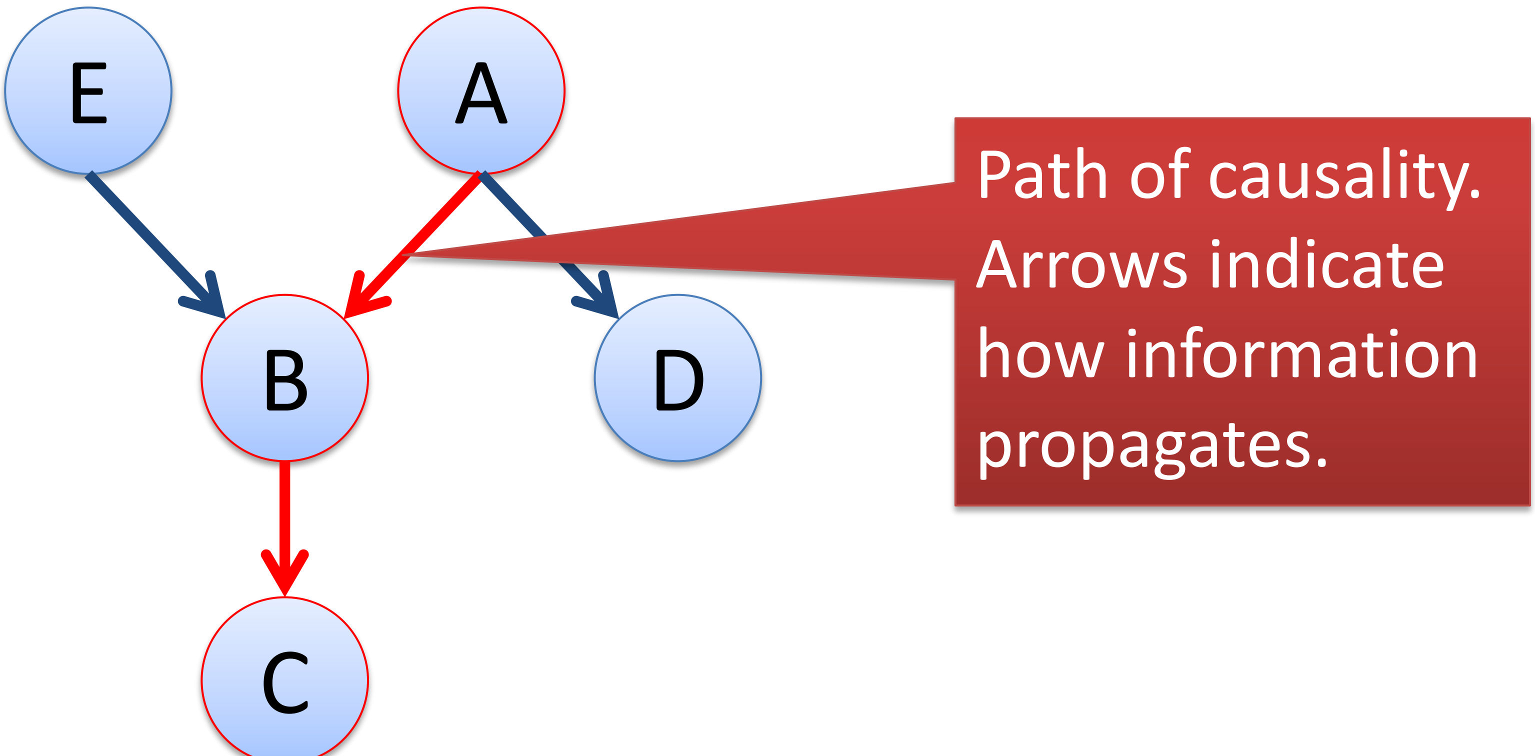
$P(B=b_1|A=a_2,E=e_1) = 0.4$

$P(B=b_2|A=a_2,E=e_1) = 0.6$

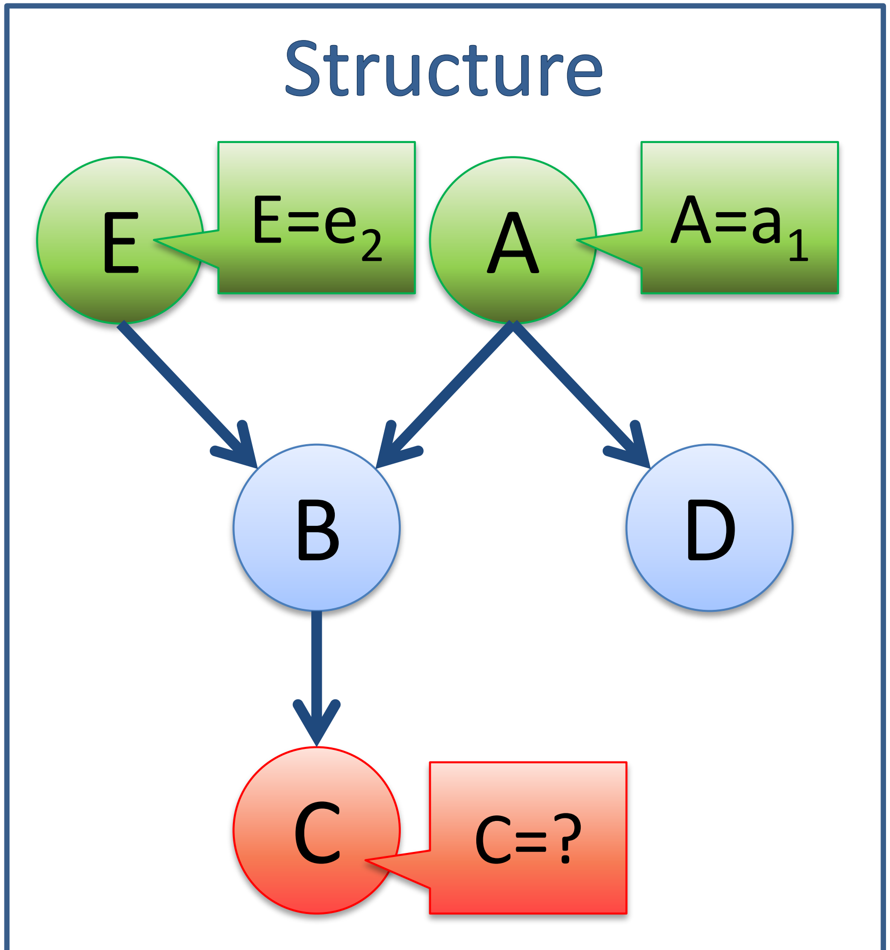
$P(B=b_1|A=a_2,E=e_2) = 0.2$

$P(B=b_2|A=a_2,E=e_2) = 0.8$

Bayesian Networks



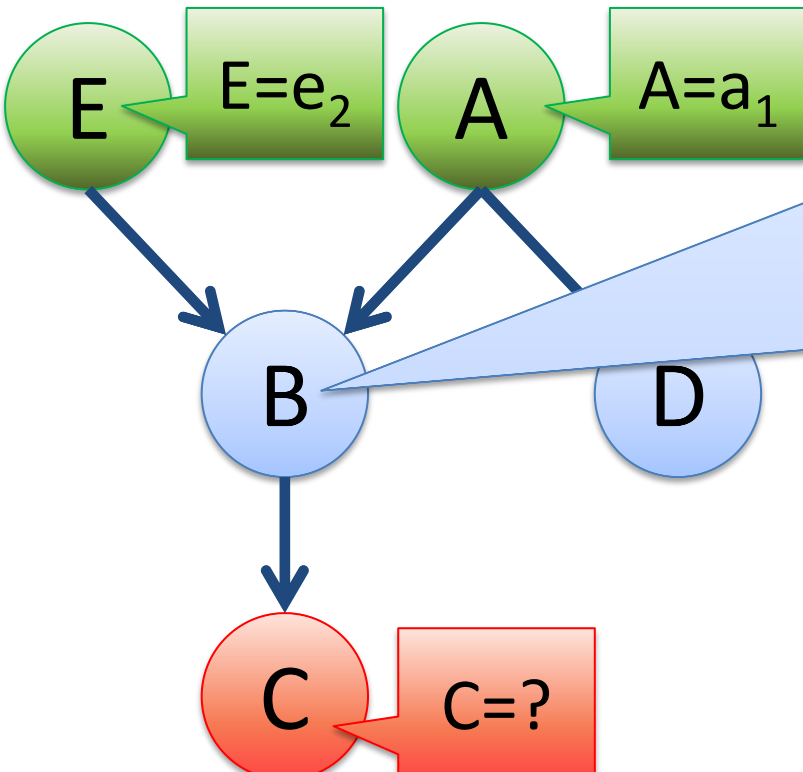
Bayesian Networks: Inference



Parameters

Bayesian Networks: Inference

Structure



Parameters

$P(B=b_1|A=a_1,E=e_1) = 0.5$

$P(B=b_2|A=a_1,E=e_1) = 0.5$

$P(B=b_1|A=a_1,E=e_2) = 0.9$

$P(B=b_2|A=a_1,E=e_2) = 0.1$

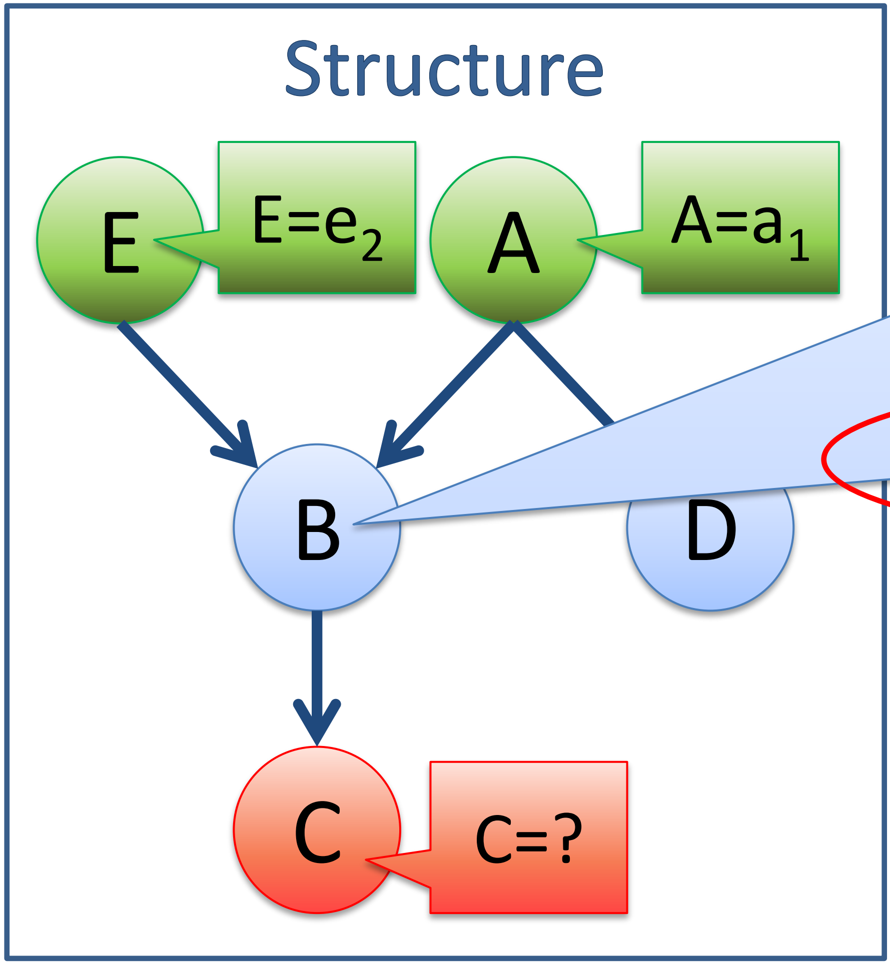
$P(B=b_1|A=a_2,E=e_1) = 0.4$

$P(B=b_2|A=a_2,E=e_1) = 0.6$

$P(B=b_1|A=a_2,E=e_2) = 0.2$

$P(B=b_2|A=a_2,E=e_2) = 0.8$

Bayesian Networks: Inference



Parameters

$P(B=b_1 | A=a_1, E=e_1) = 0.5$

$P(B=b_2 | A=a_1, E=e_1) = 0.5$

$P(B=b_1 | A=a_1, E=e_2) = 0.9$

$P(B=b_2 | A=a_1, E=e_2) = 0.1$

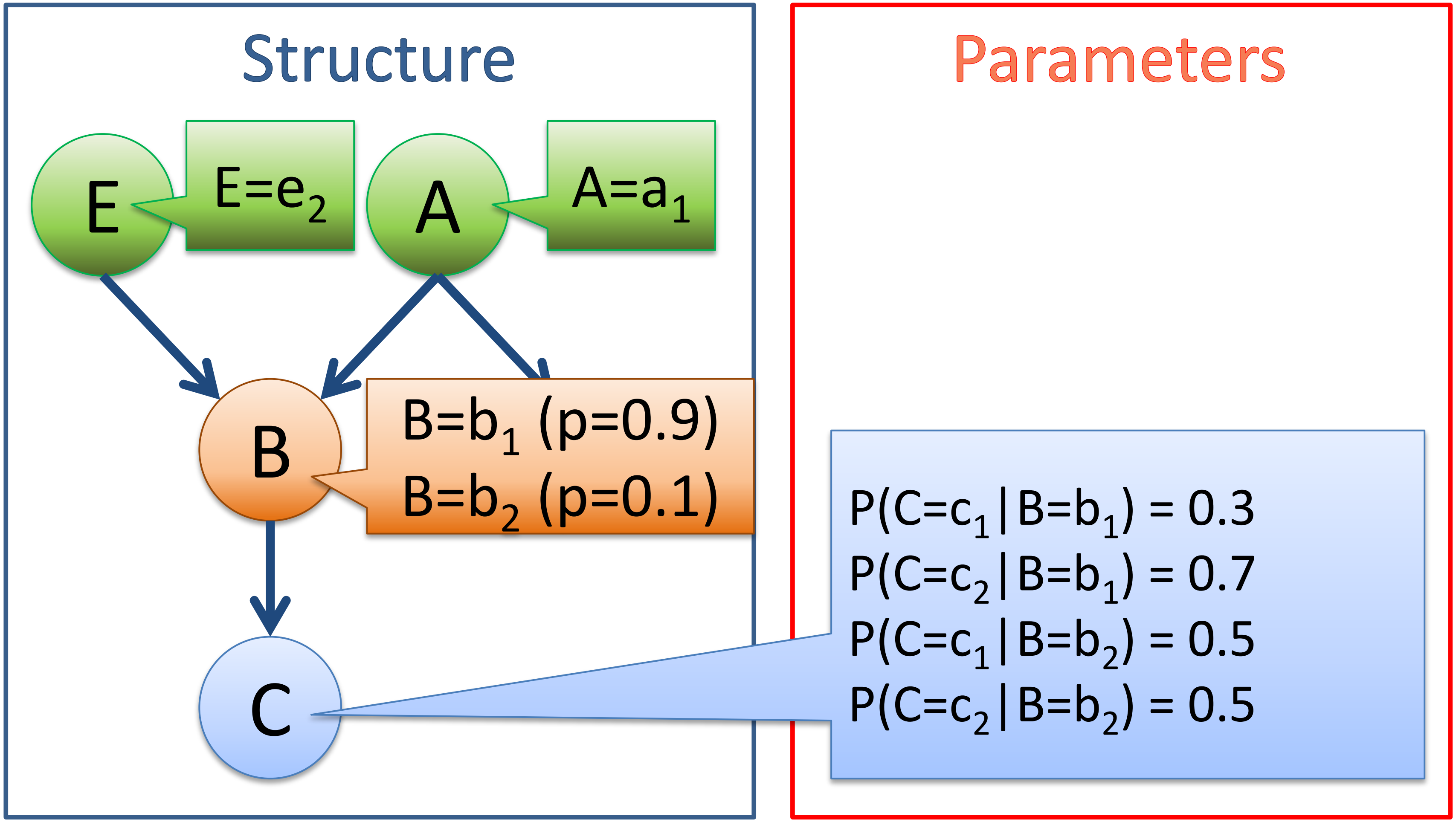
$P(B=b_1 | A=a_2, E=e_1) = 0.4$

$P(B=b_2 | A=a_2, E=e_1) = 0.6$

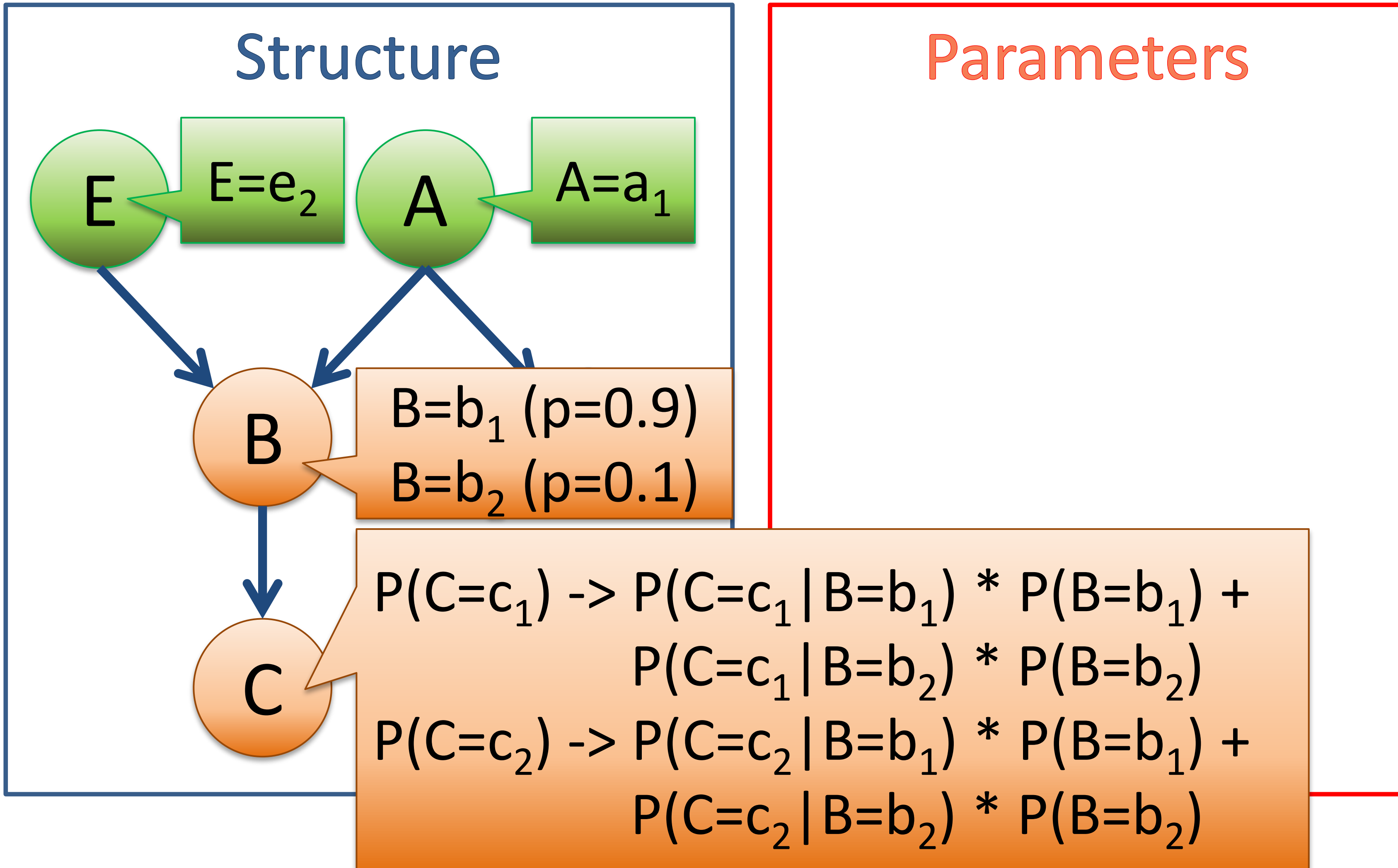
$P(B=b_1 | A=a_2, E=e_2) = 0.2$

$P(B=b_2 | A=a_2, E=e_2) = 0.8$

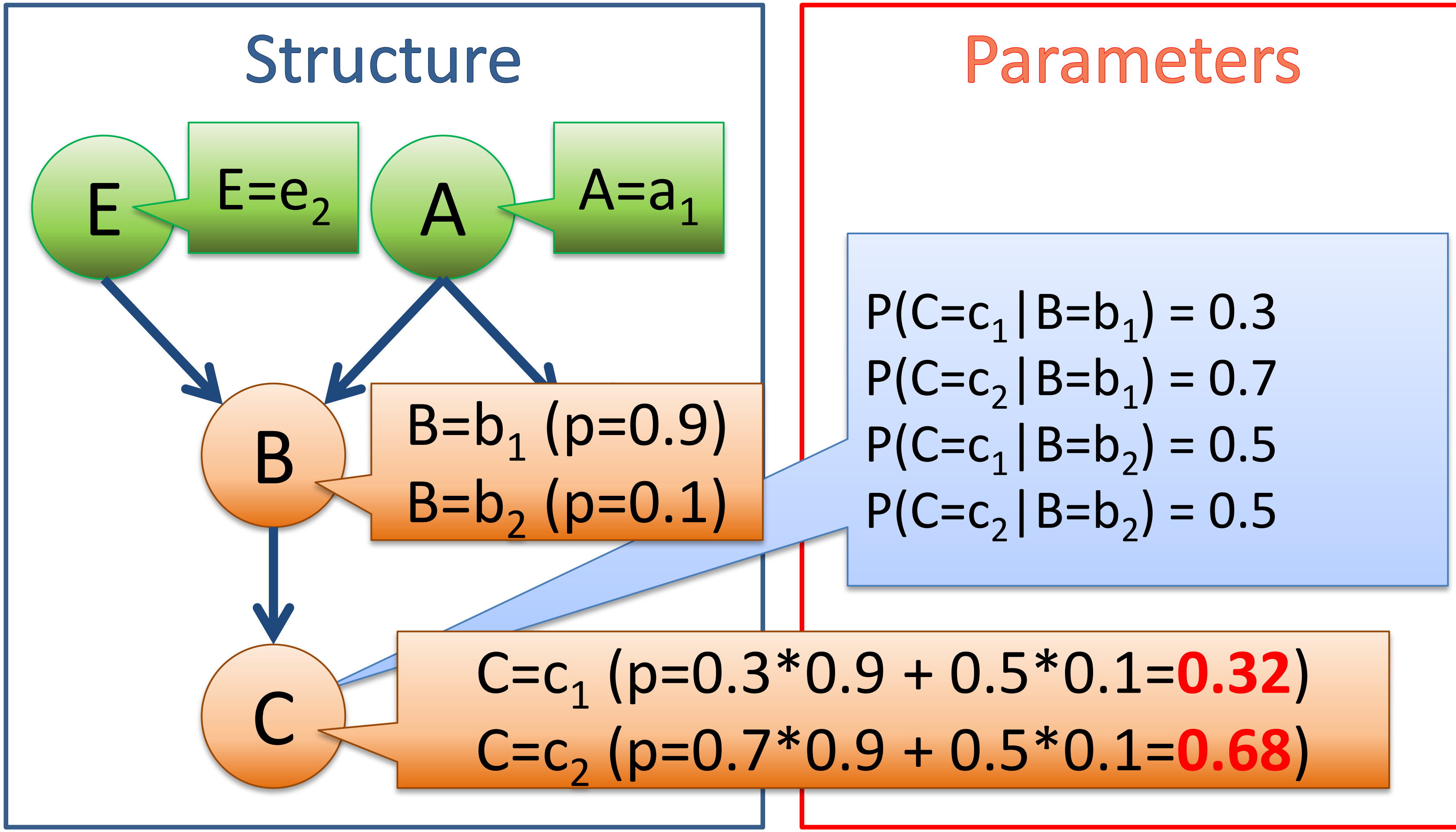
Bayesian Networks: Inference



Bayesian Networks: Inference

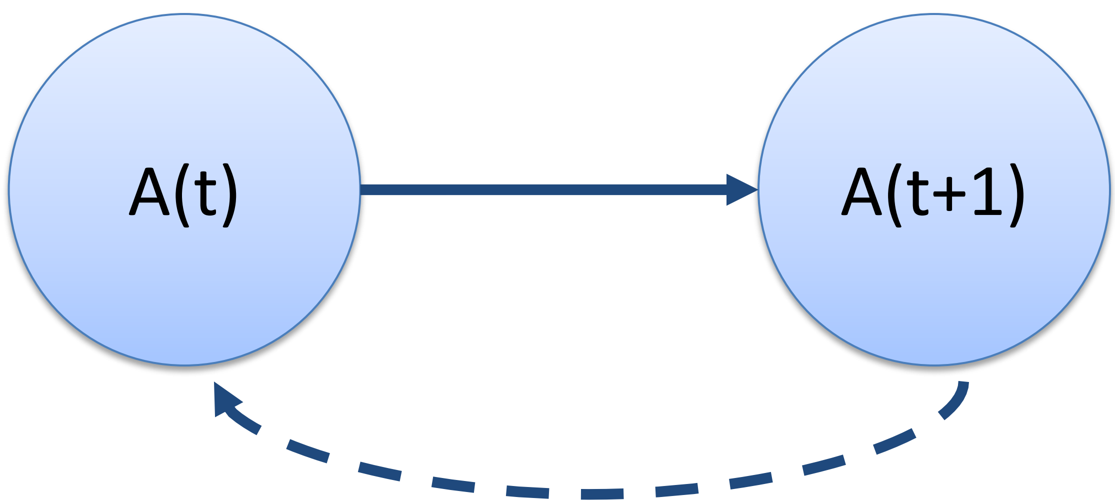


Bayesian Networks: Inference



Bayesian Networks: Dynamic BNs

- Evolution in time
  - Some variables at time **t**, others a time **t+1**
  - Values most probable for **t+1** can be “re-used”
  - With “re-used” values, obtain new predictions
  - In this way, a dynamic is produced

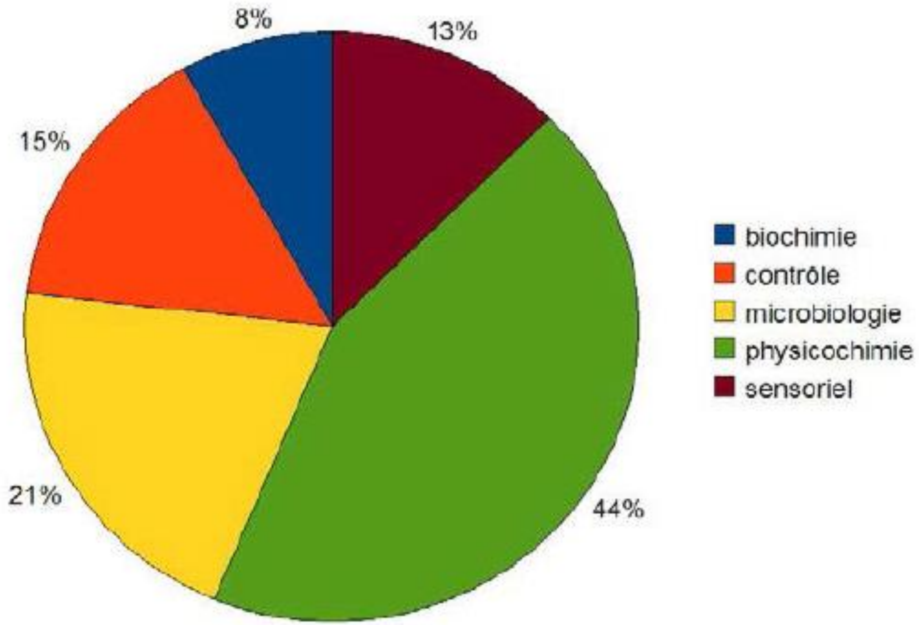


Bayesian Networks: and more!

- Several other interesting properties
  - Can be retrained with new evidence (anti-spam)
  - …both automatically and manually
  - New nodes can be added to existing structures
  - …and much more!

Case study: Camembert

- 41 days of ripening
  - 15 days in ripening room
  - 26 days packed, at 4°C
- 112 studies as of October 2009



Case study: Camembert

- Complex system (ecosystem, bioreactor)
- Research lines and models
  - Development of microbes
  - Link microbial activity - sensorial properties
  - Physical-chemical phenomena
  - Ripening control through expert systems

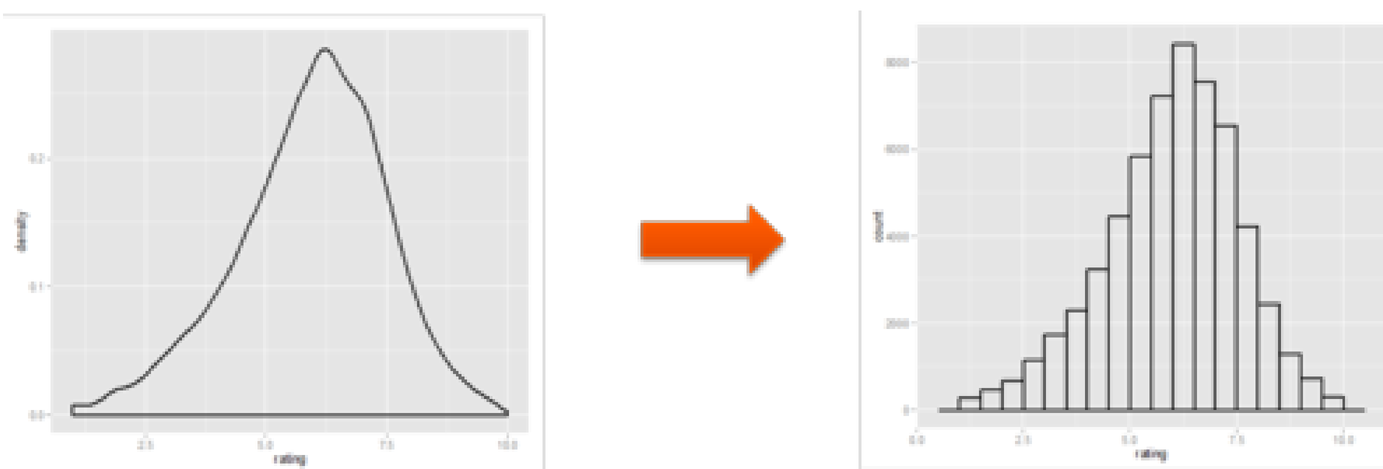
- **No global view of the process!**

Case study: Camembert

- Camembert cheese ripening process
  - Quantitative variables: pH, temperature, …
  - Qualitative variables: odor, under-rind, coat, …
  - Data from heterogeneous sources
  - Dynamic BN (DBN): t -> t+1

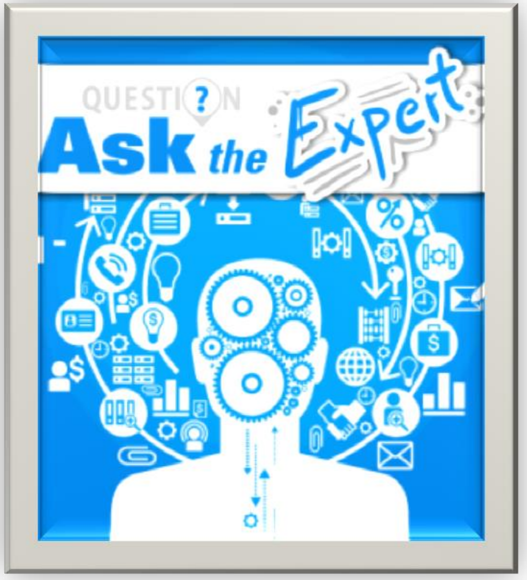
Case study: Camembert

- Quantitative variables
  - Discretize into intervals
  - Meaningful values for the intervals!

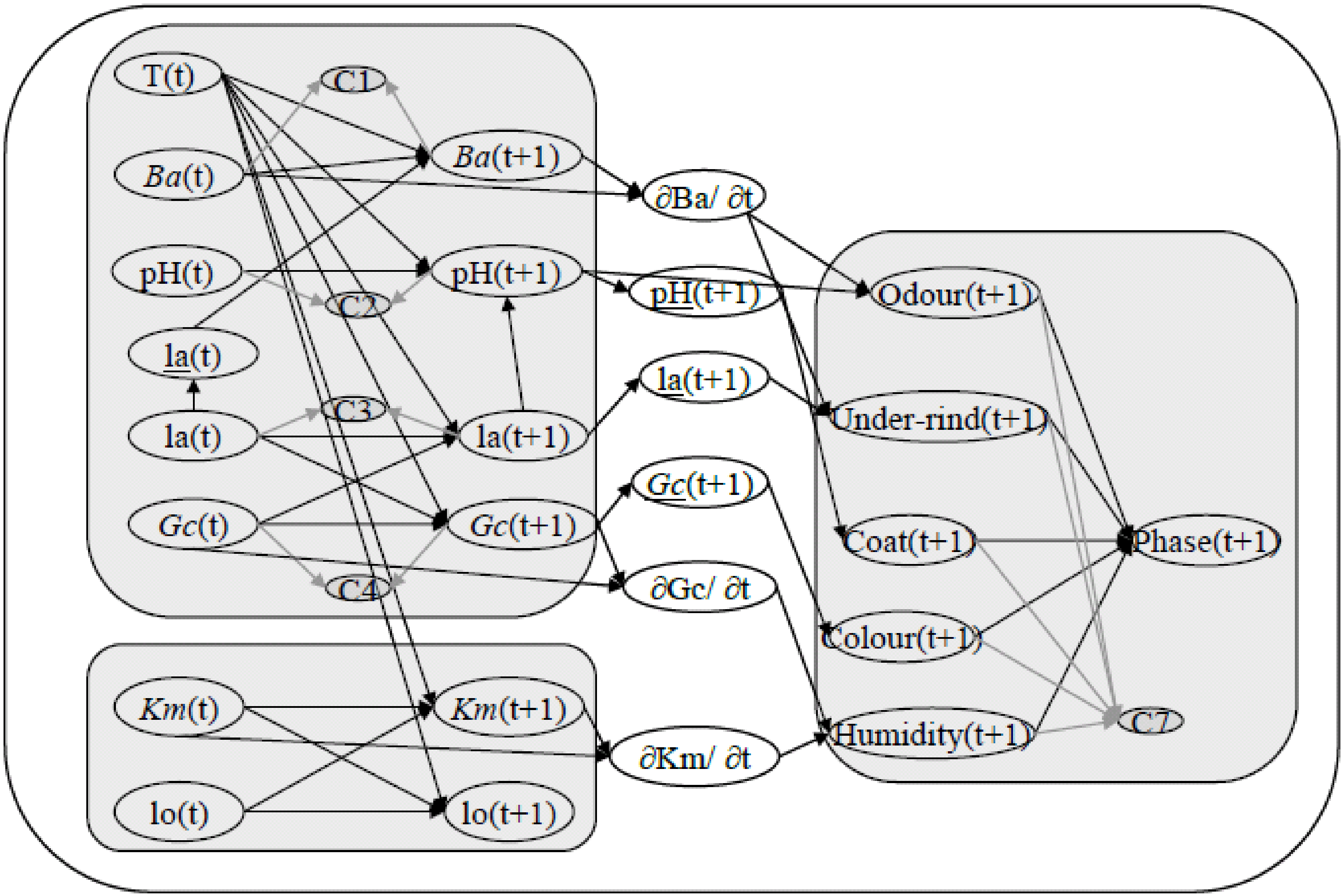


Case study: Camembert

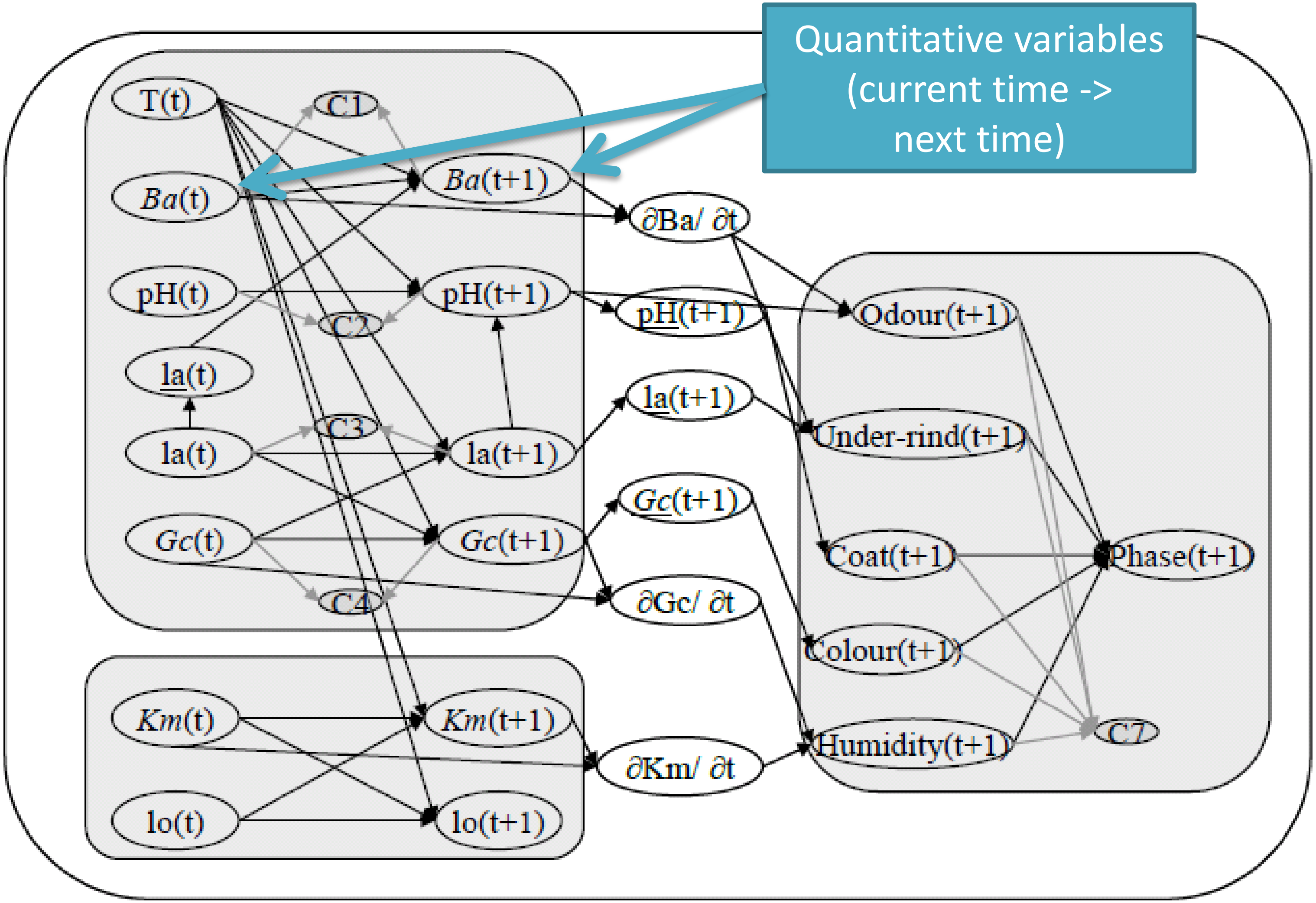
- Qualitative variables
  - Ask experts
  - Link their judgment to interval of values
  - Different experts might have different judgment!



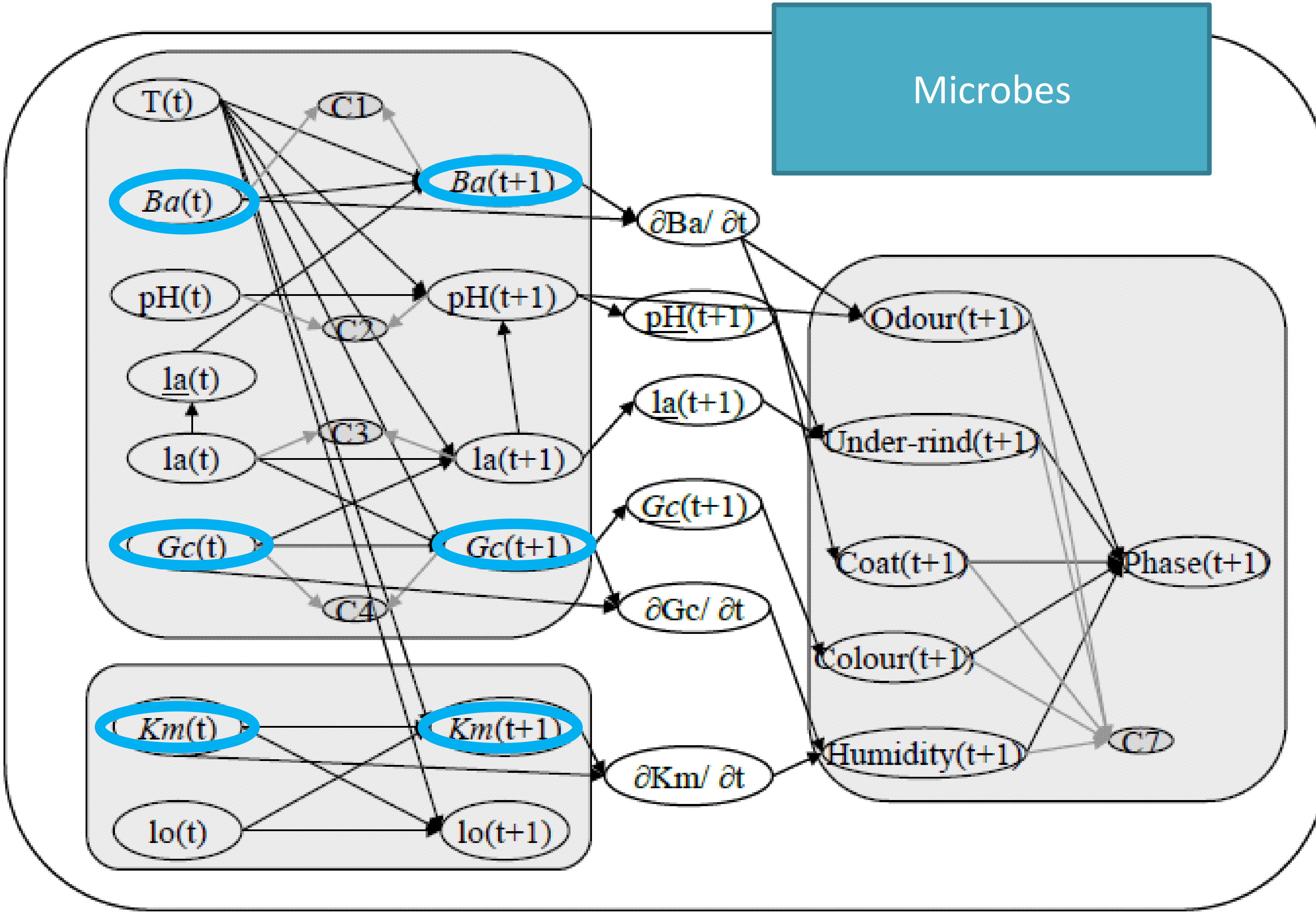
Case study: Camembert



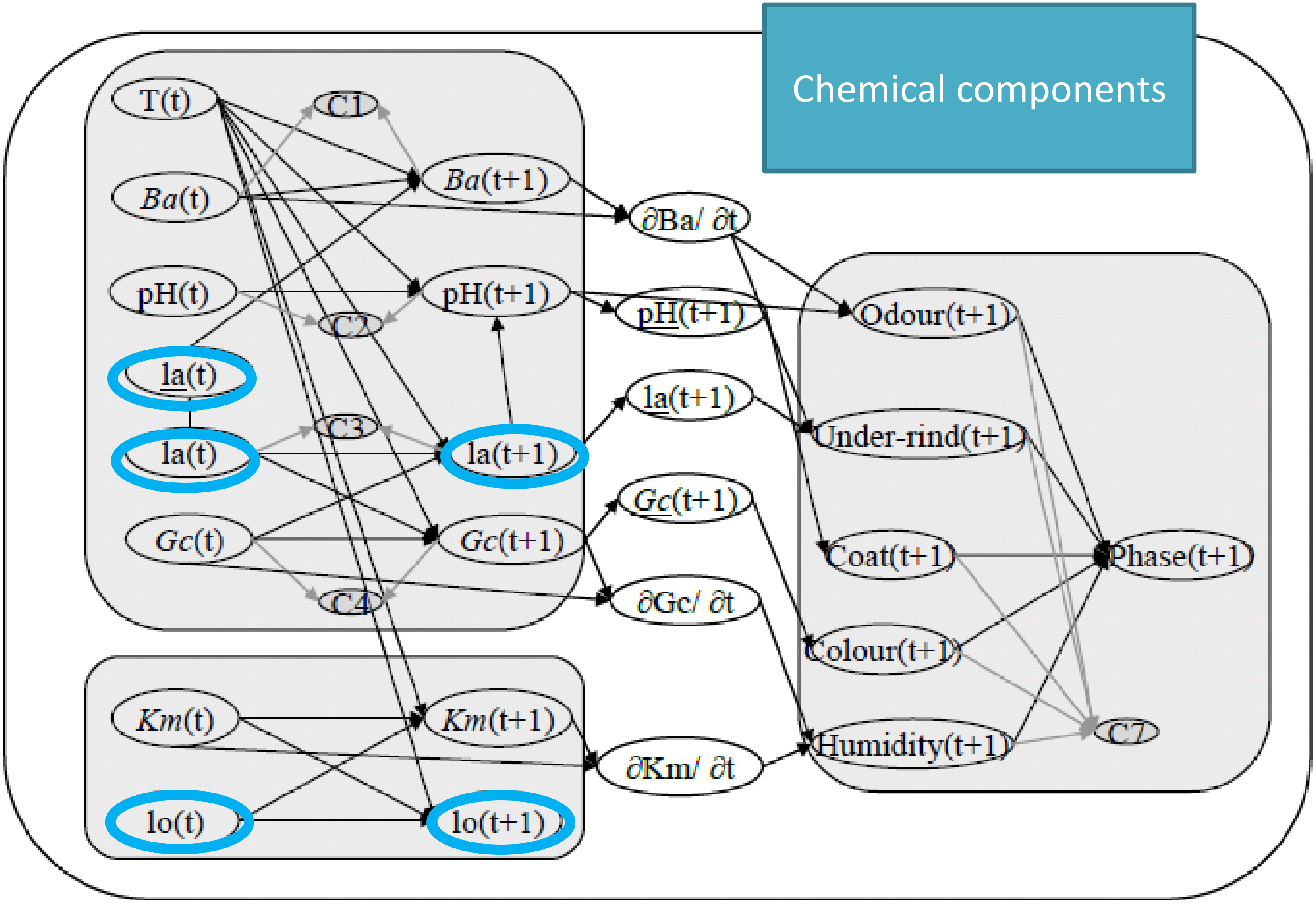
Case study: Camembert



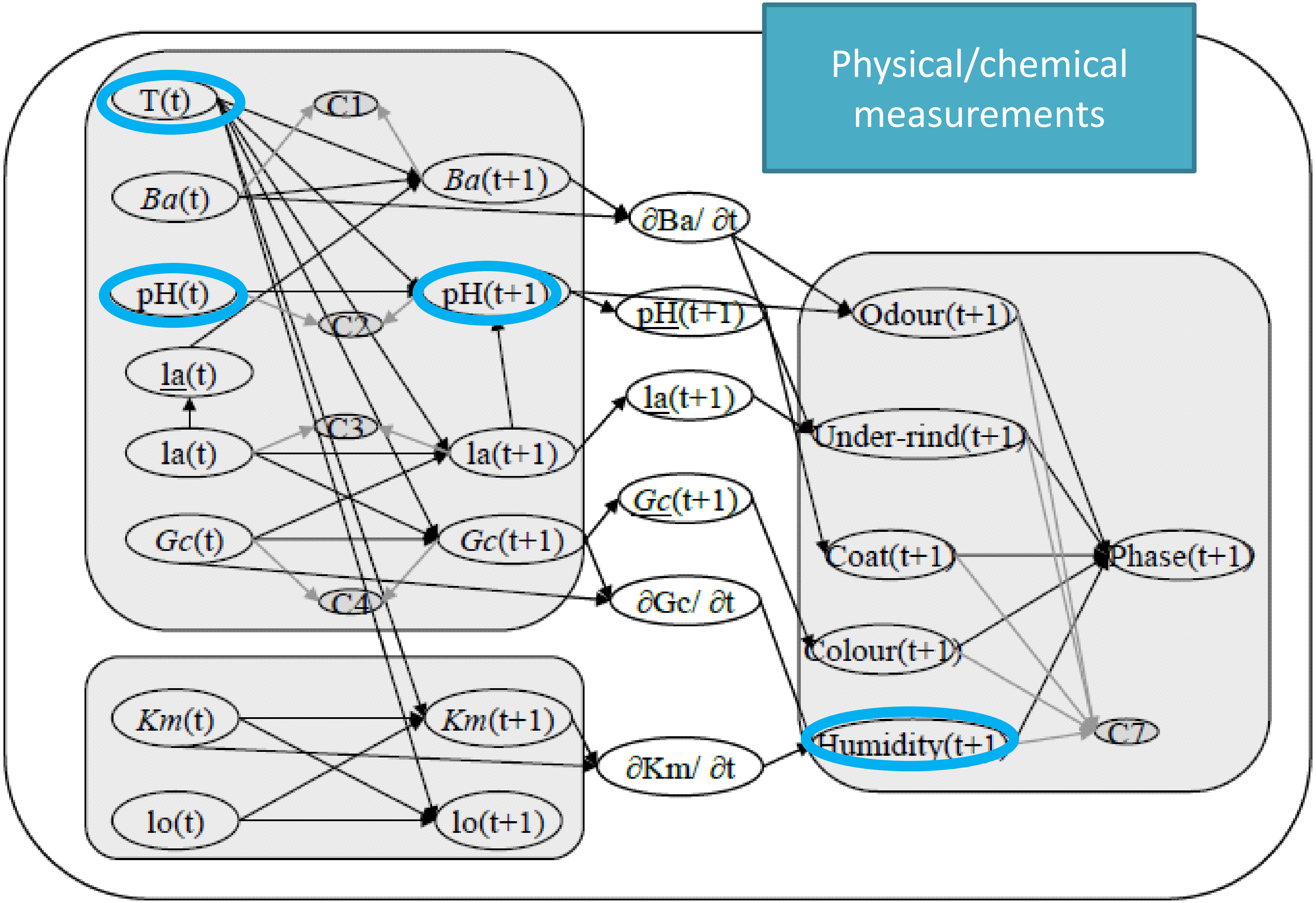
Case study: Camembert



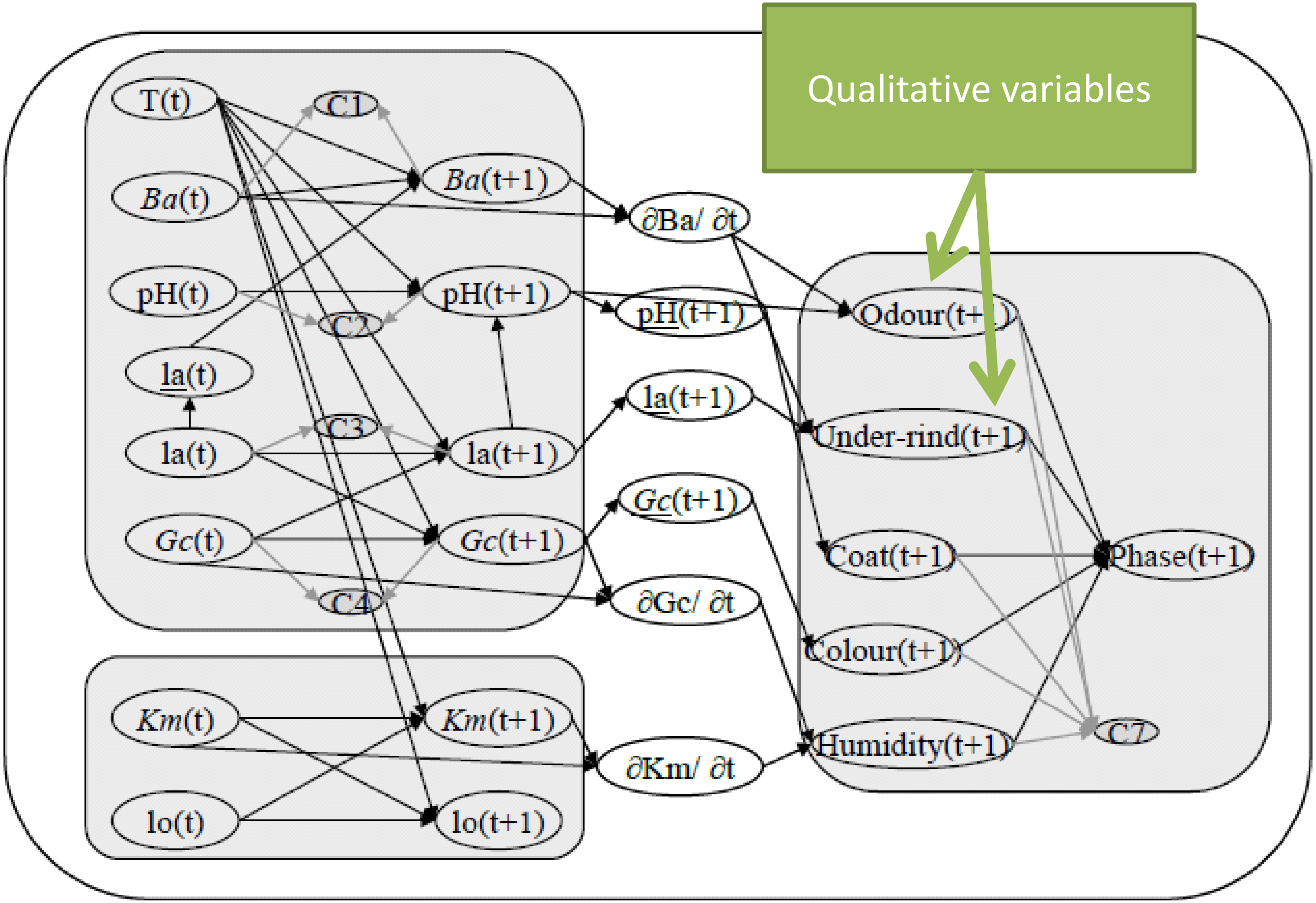
Case study: Camembert



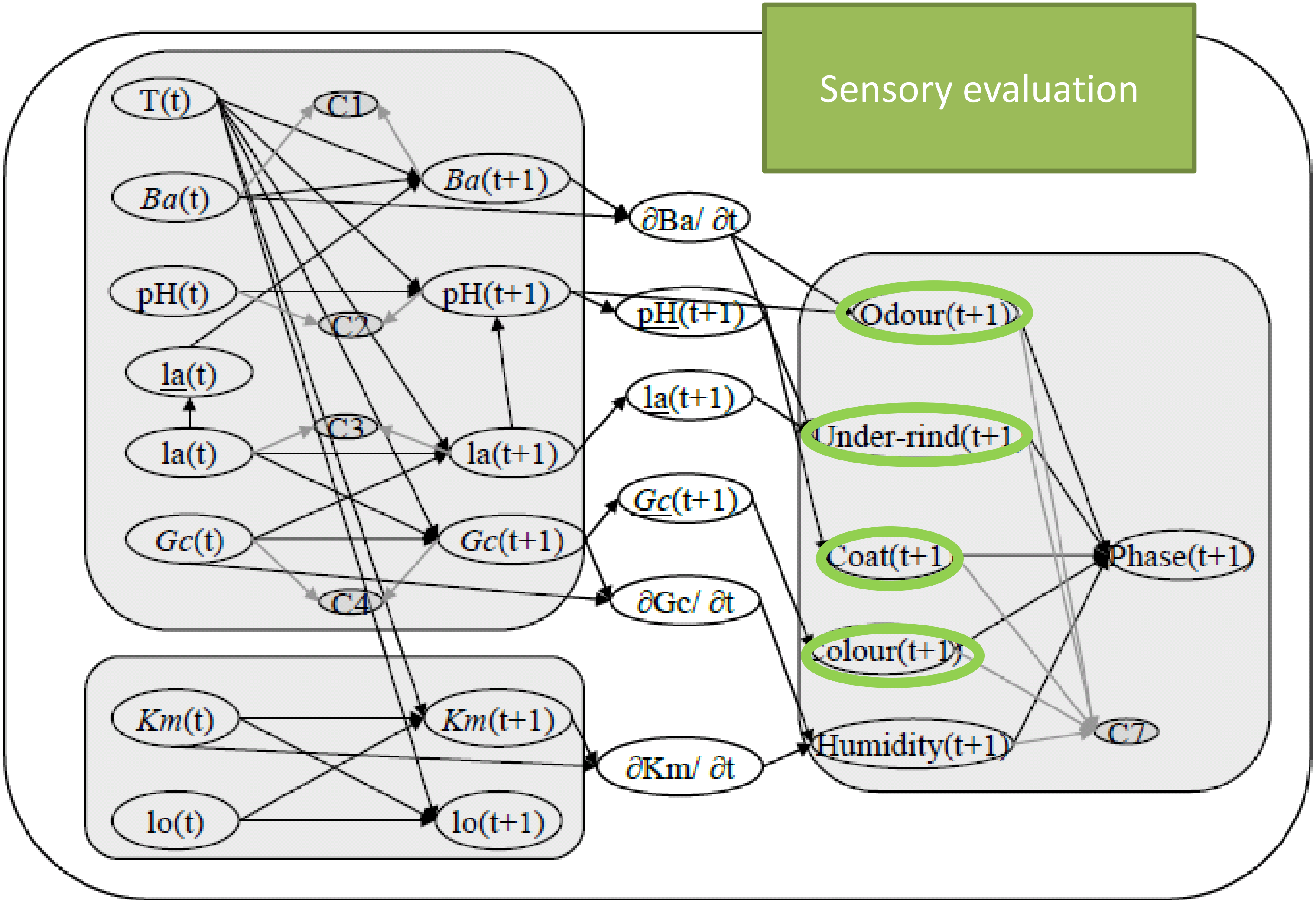
Case study: Camembert



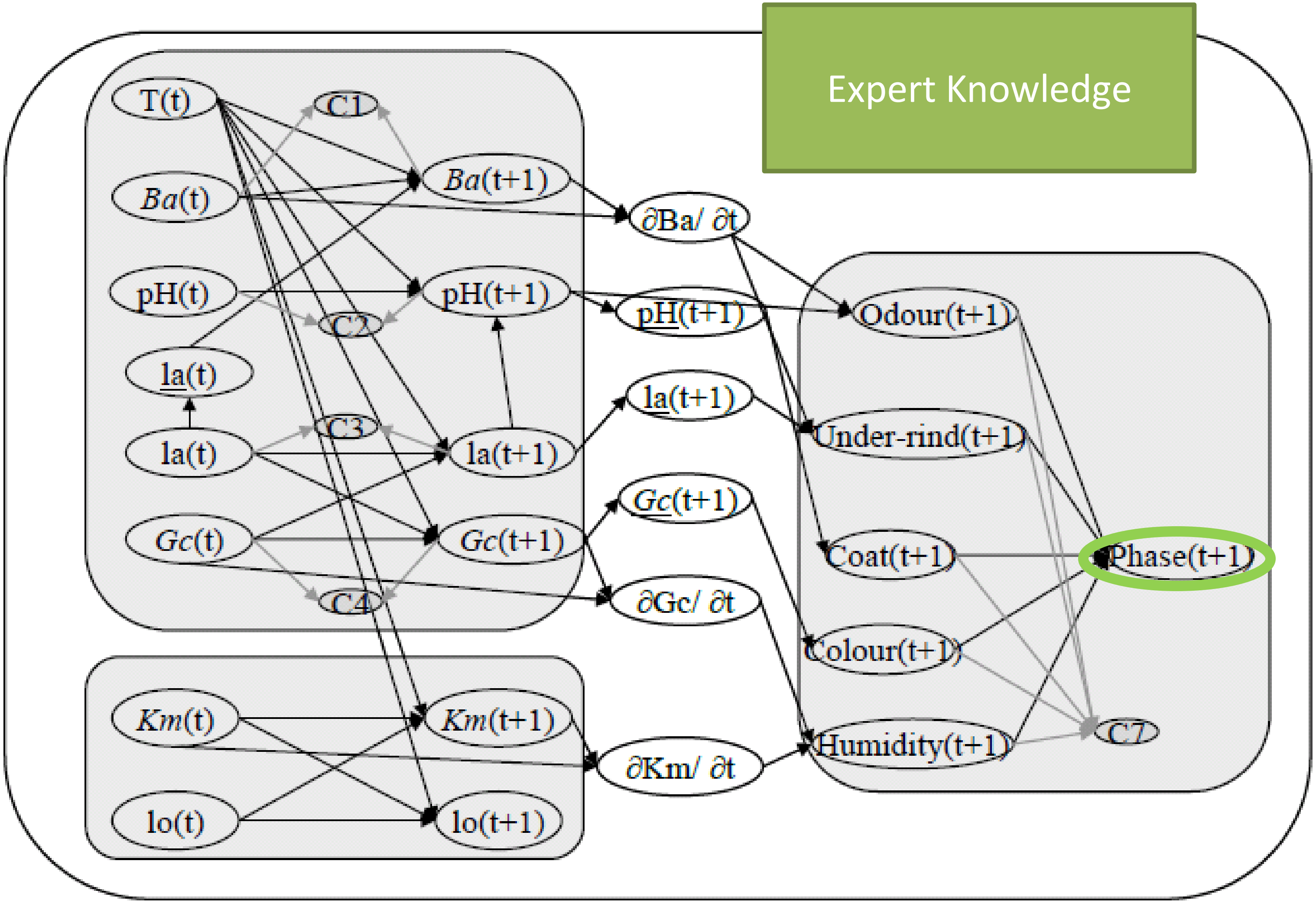
Case study: Camembert



Case study: Camembert



Case study: Camembert



Case study: Camembert

- Ripening
- 4 distinct phases
- Expert knowledge

Day ~15

Day 1

Day >30

Case study: Camembert

1. Evolution of humidity
2. Development of under-rind + "champignon" aroma
3. Development of crust + creamy consistency
4. "Ammonia" aroma + brown color on crust

Day 1

Day ~15

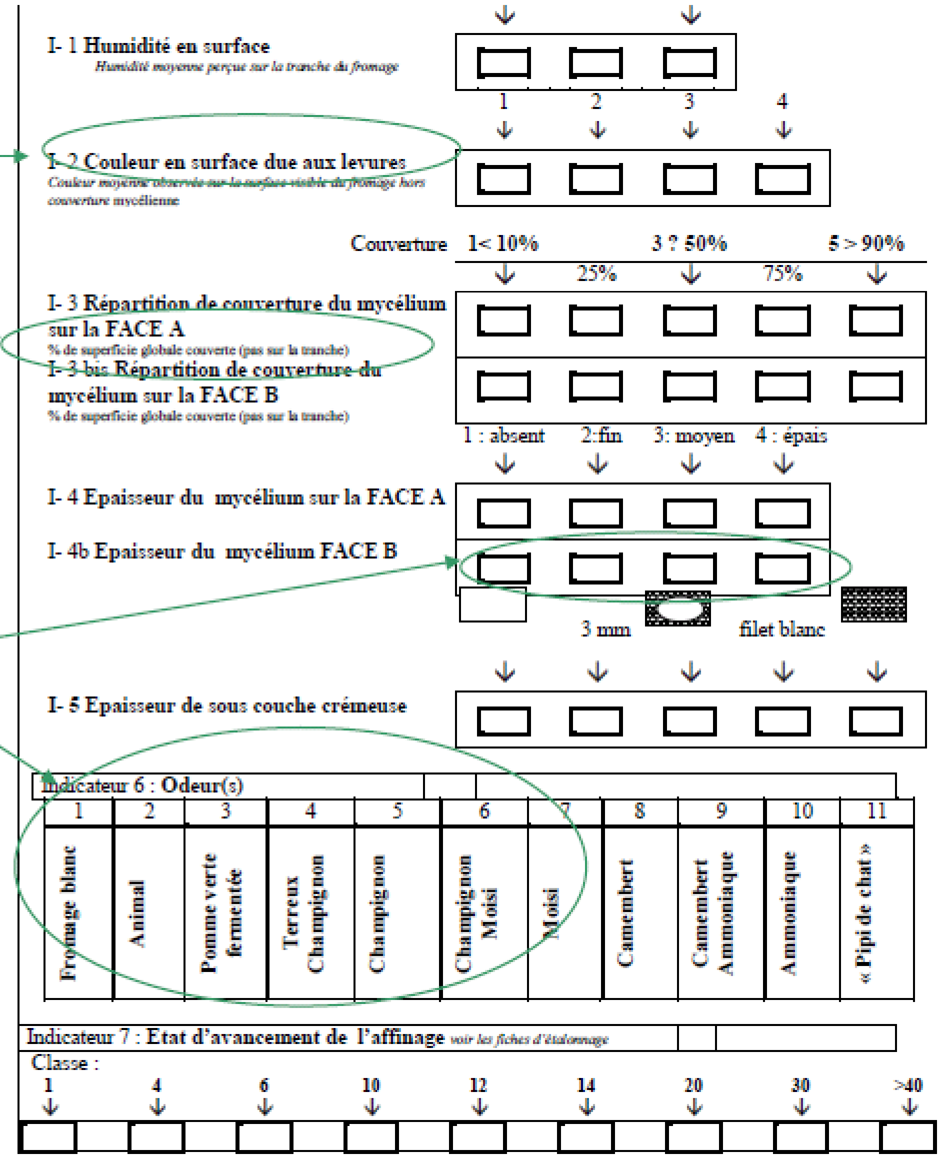
Day >30

Case study: Camembert

Sensory criteria

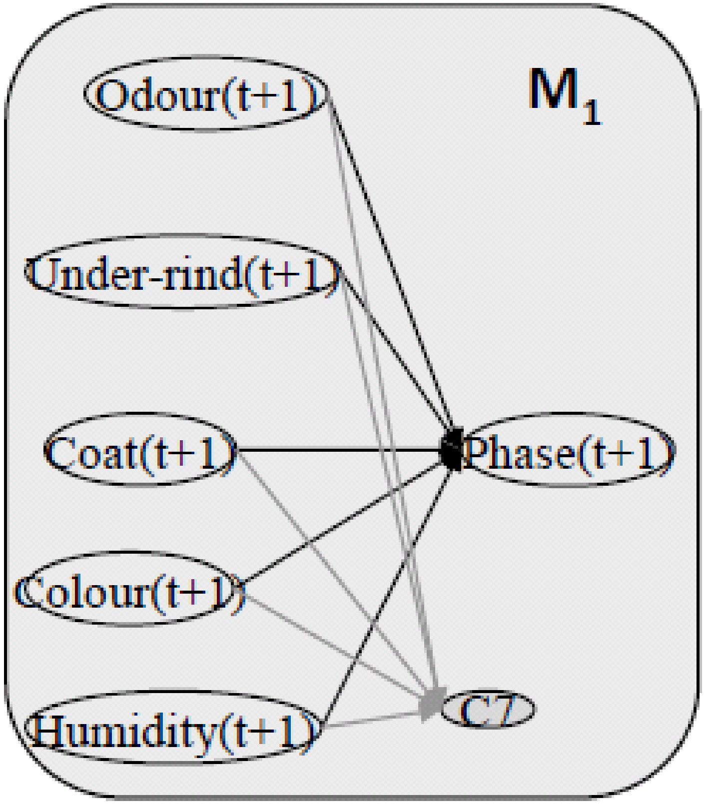
Evaluation protocol

Symbolic scale



Case study: Camembert

- Final result

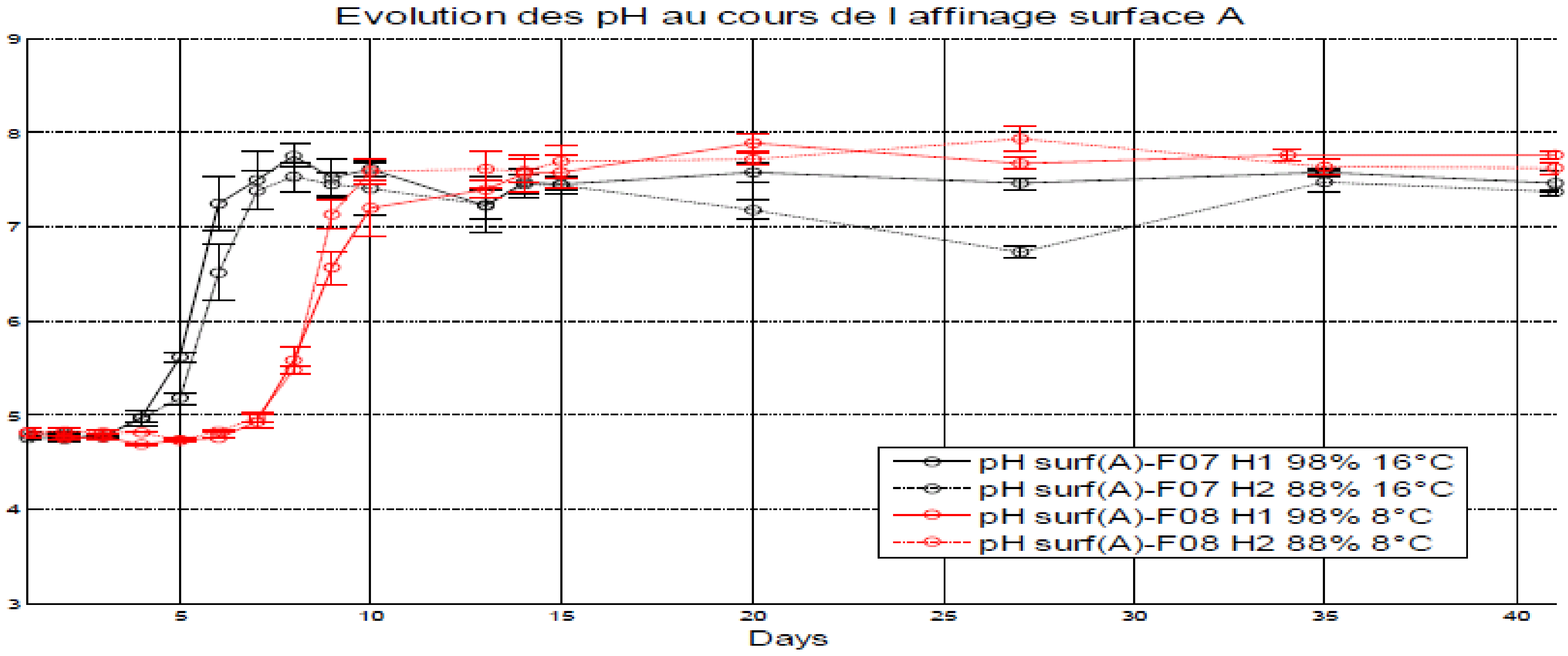


Case study: Camembert

- Experimental data
  - Measurable quantities (pH, T, la, …)
  - From continuous to discrete values
  - Choose appropriate discretization

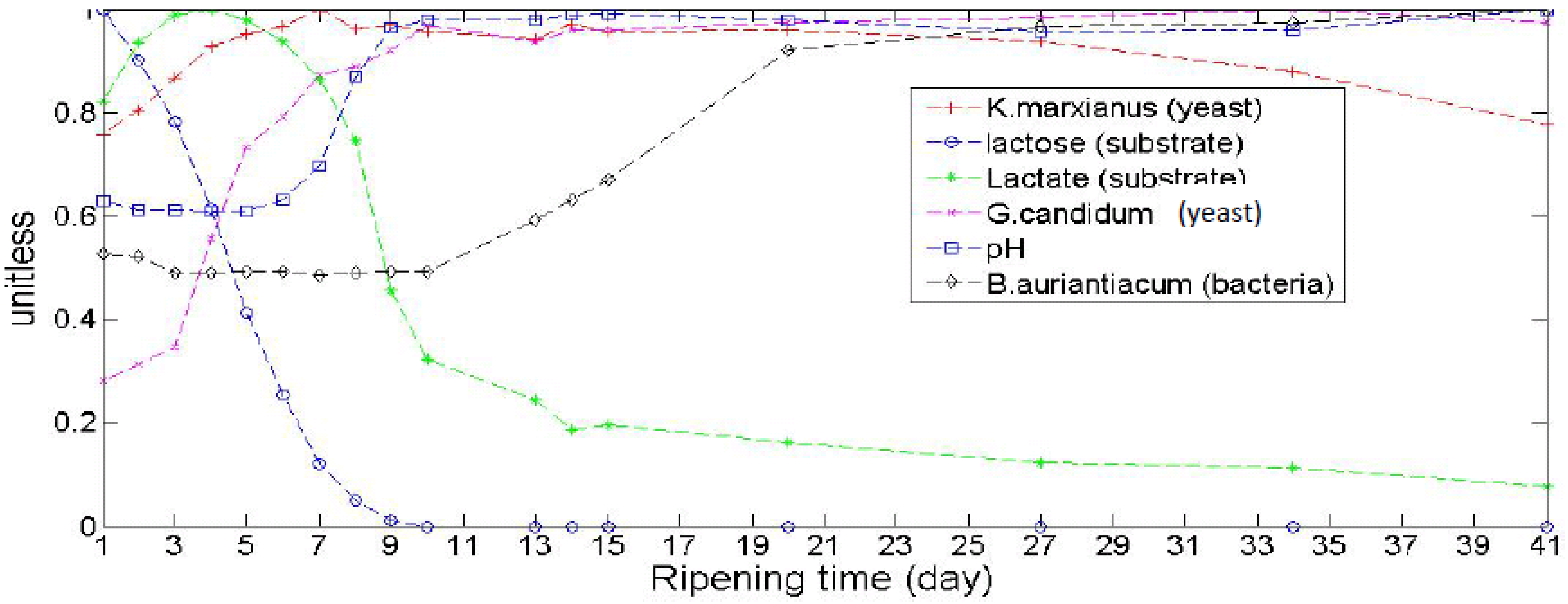
Case study: Camembert

- pH



Case study: Camembert

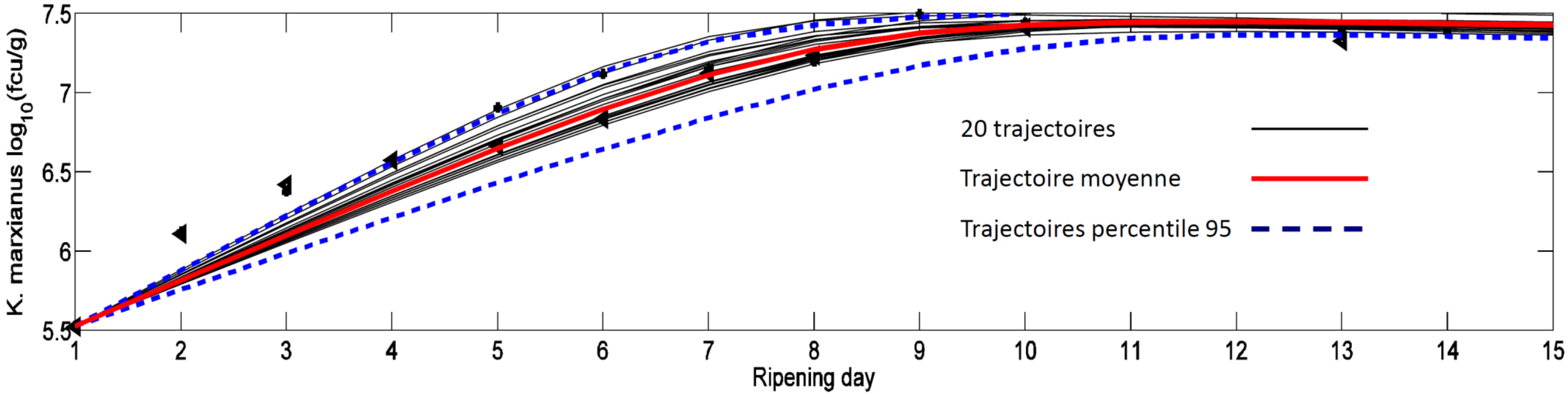
- Microbes and chemical components



Case study: Camembert

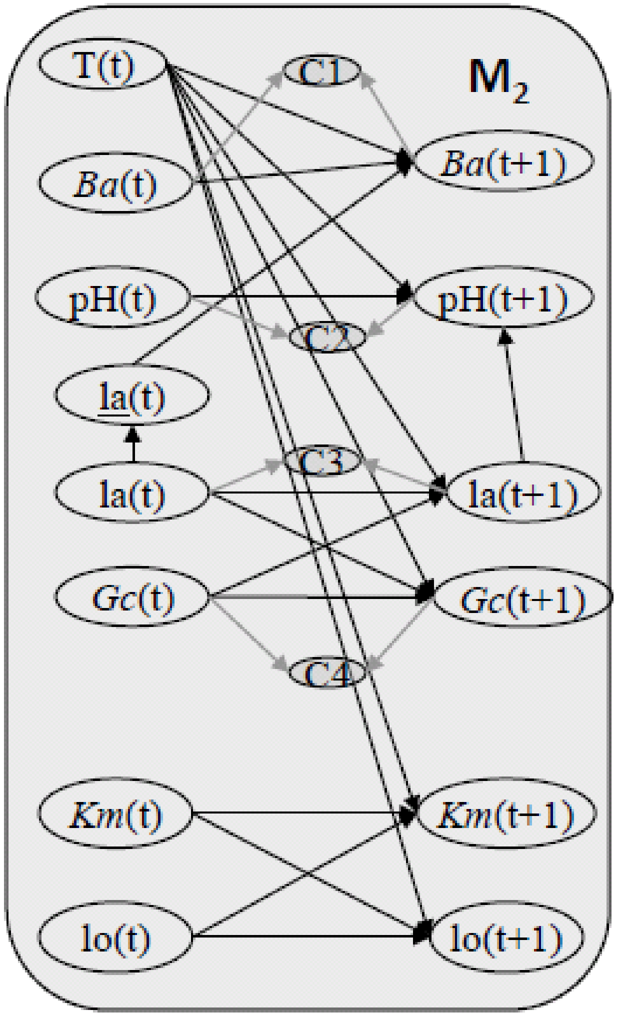
- Existing models

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



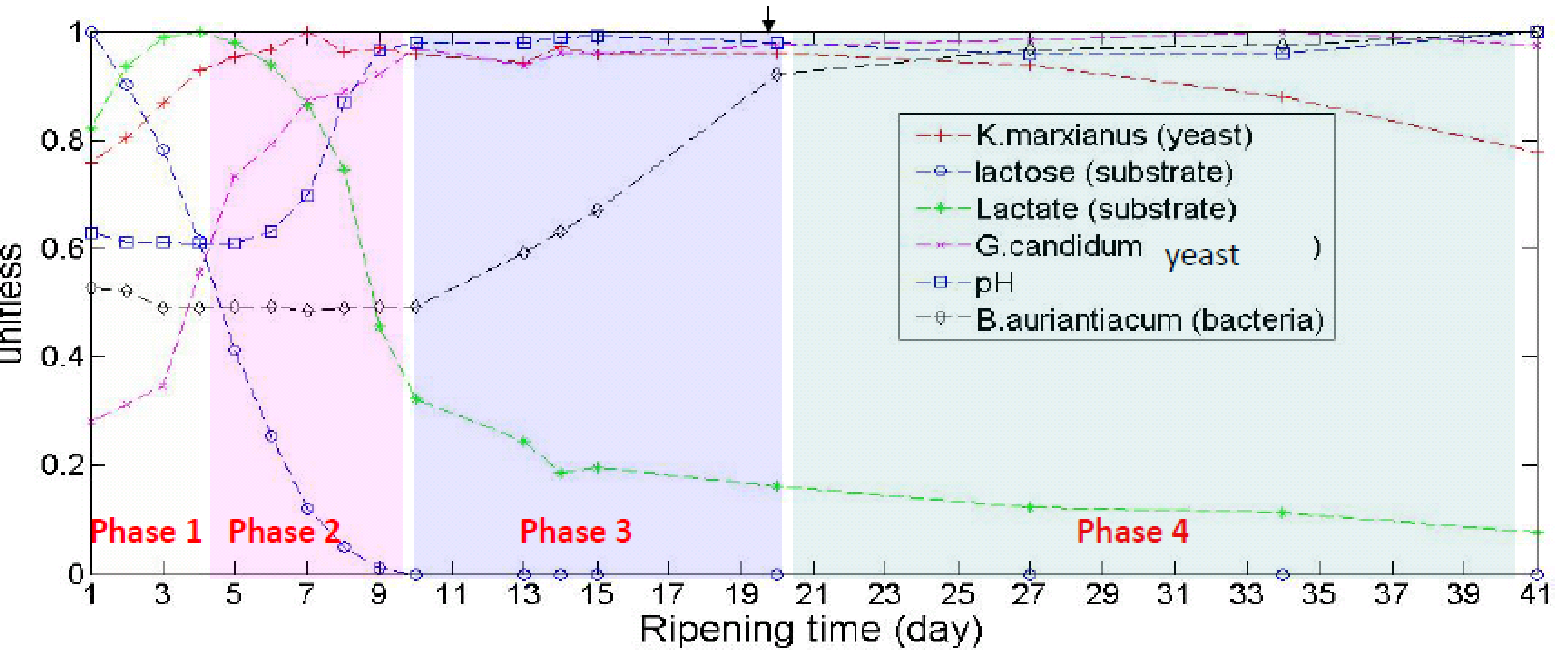
Case study: Camembert

- Final result



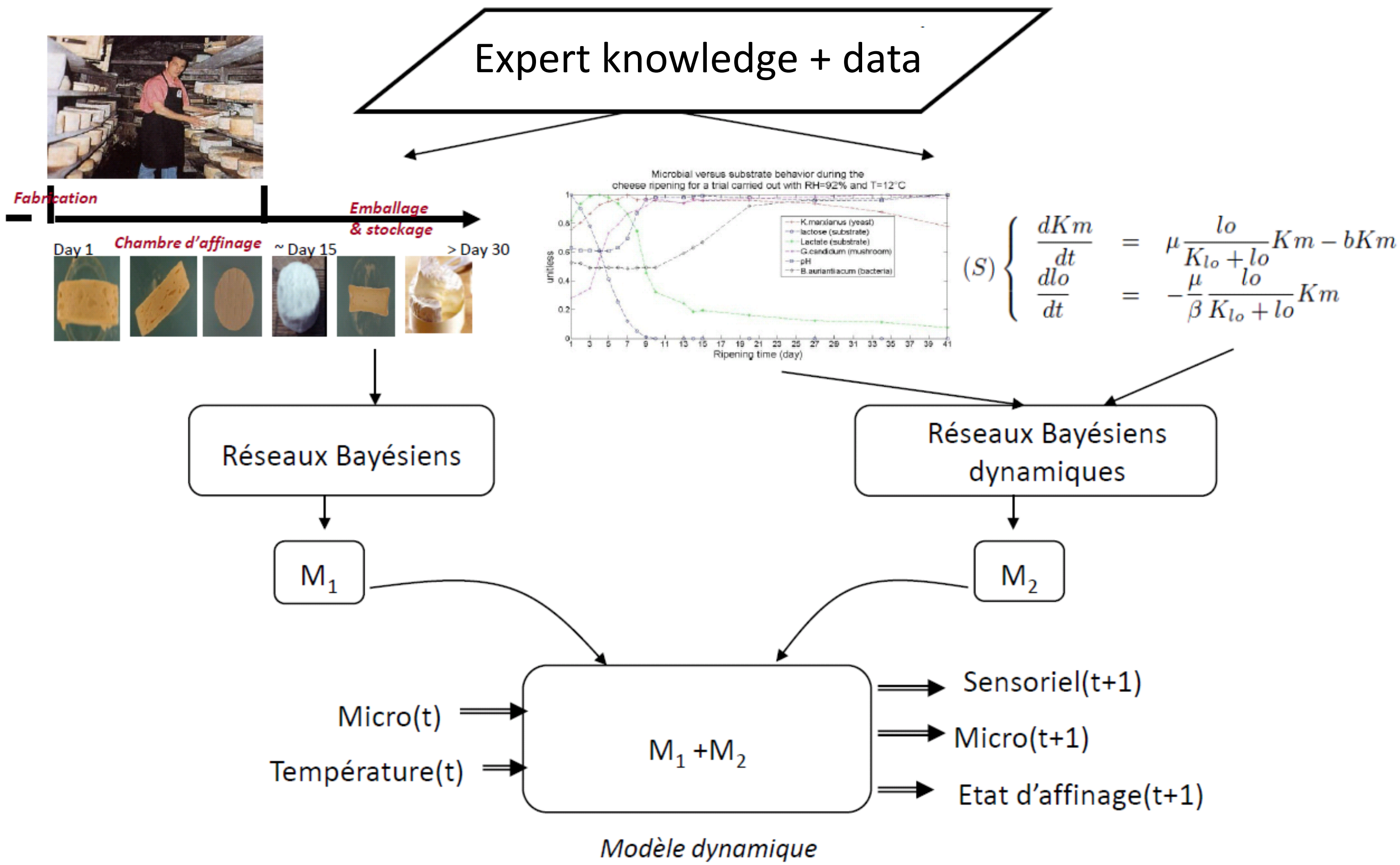
Case study: Camembert

- Finally, link the two parts!

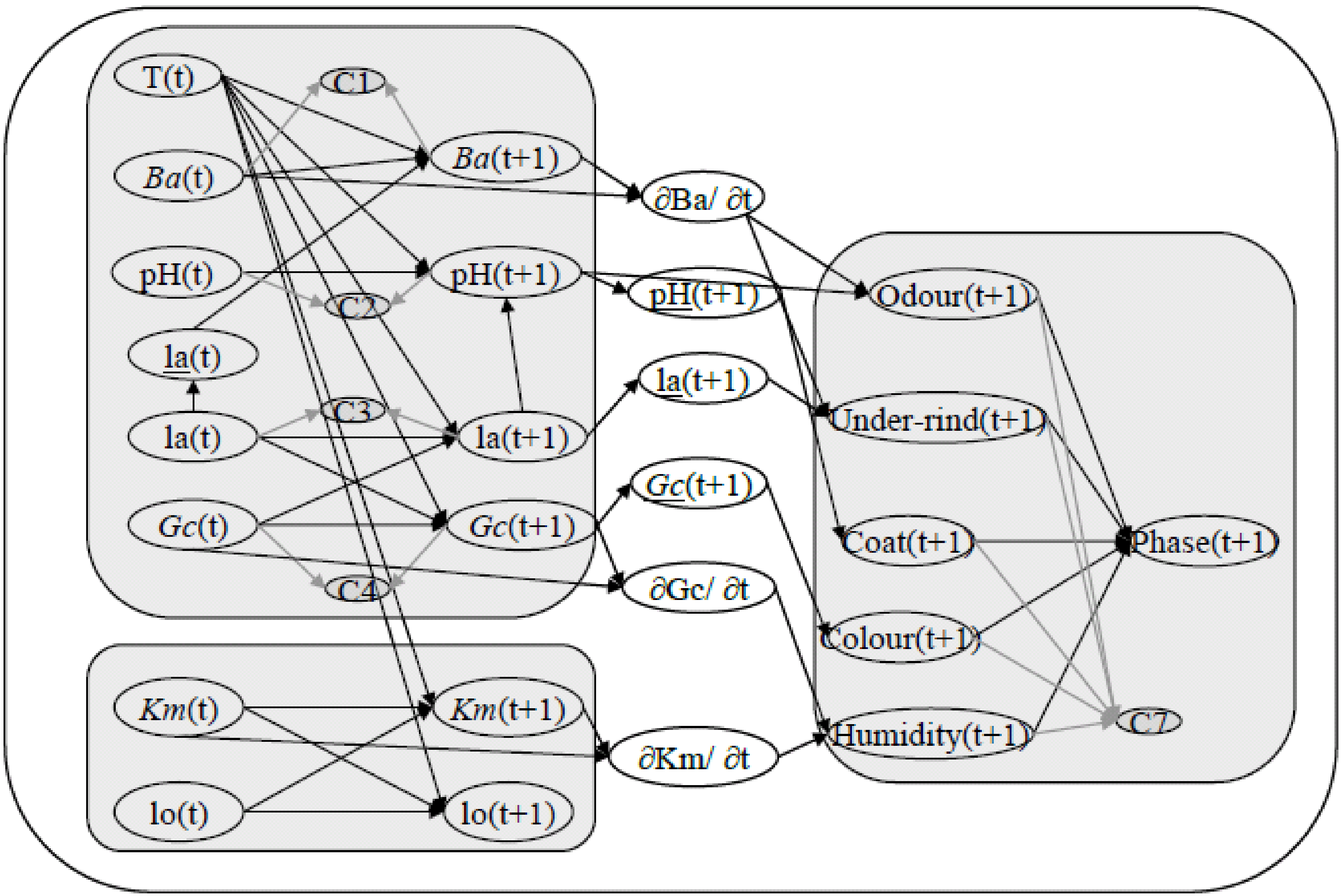


Case study: Camembert

- Expert knowledge was prominently used



Case study: Camembert

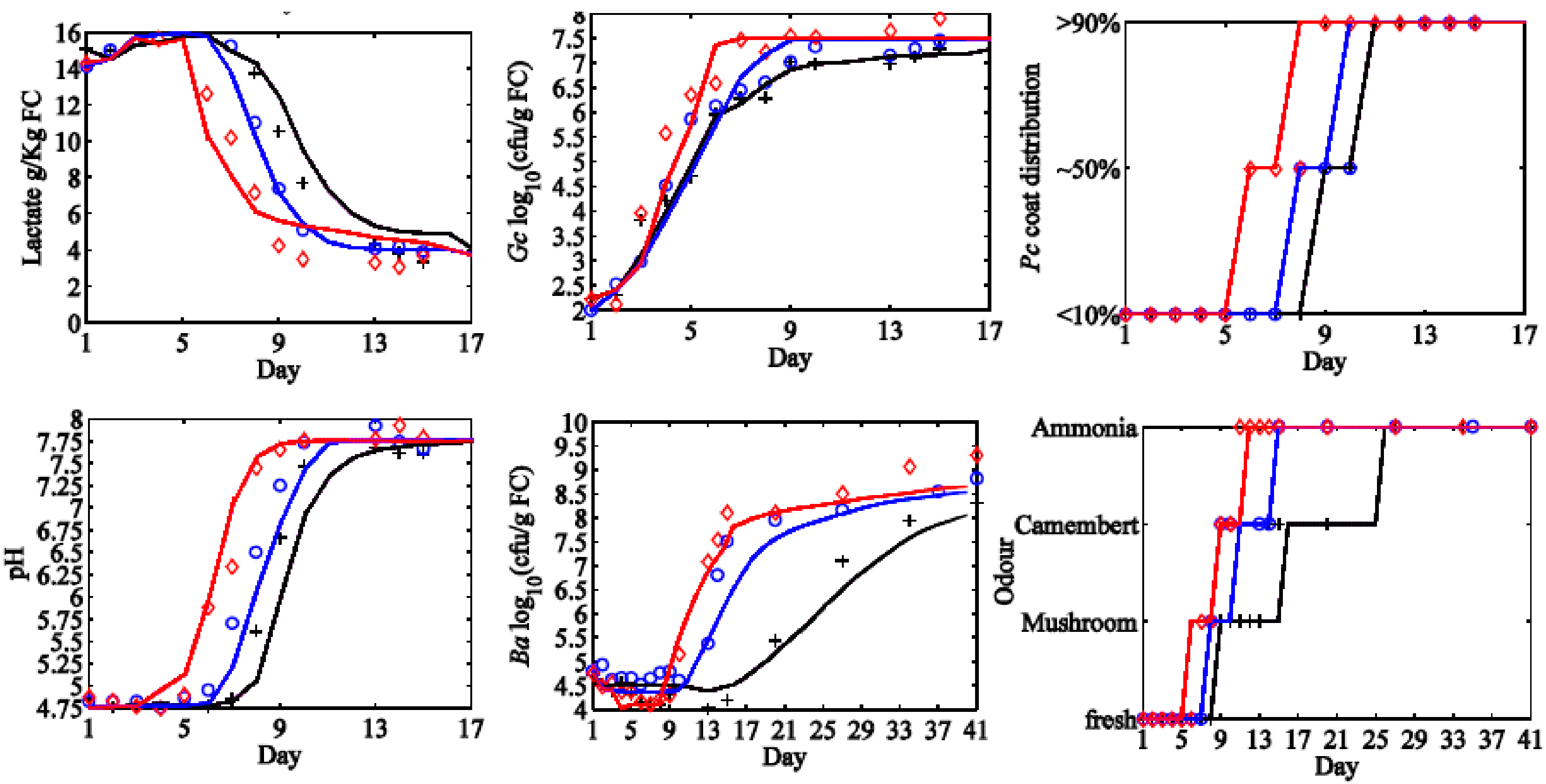


Case study: Camembert

- Now, it's time to test the model!
  - Set initial values T(0), Gc(0), …, Km(0)
  - Temperature is set from outside
  - All other values are re-injected (DBN)
  - We observe the final phase prediction

Case study: Camembert

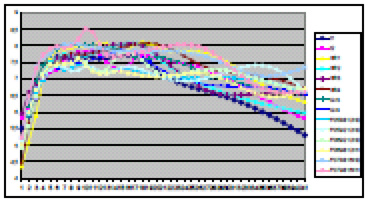
- Compare model with experimental data, for three settings (**8°C**, **12°C**, **16°C**)



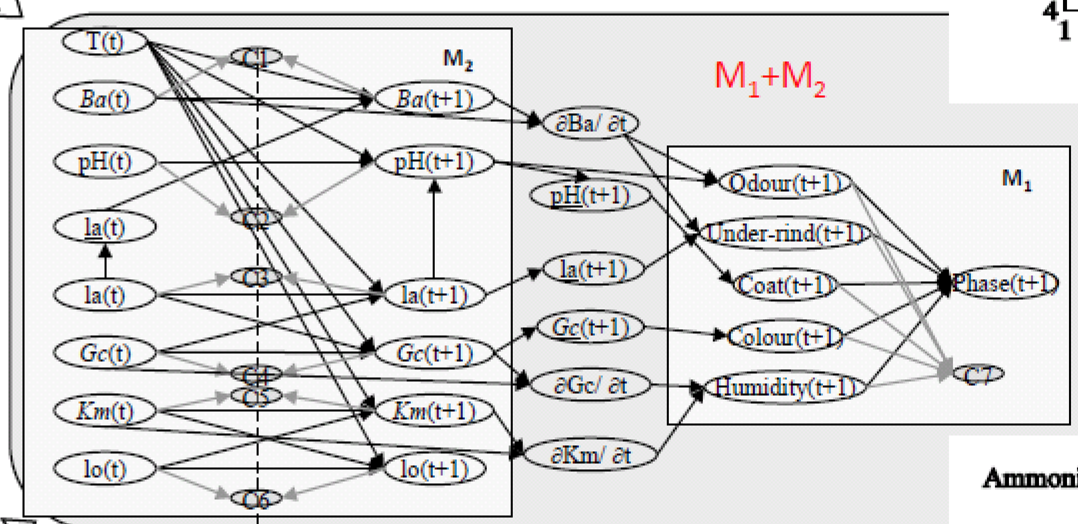
Case study: Camembert

Expert

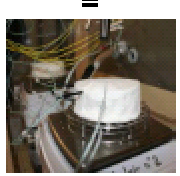
Données



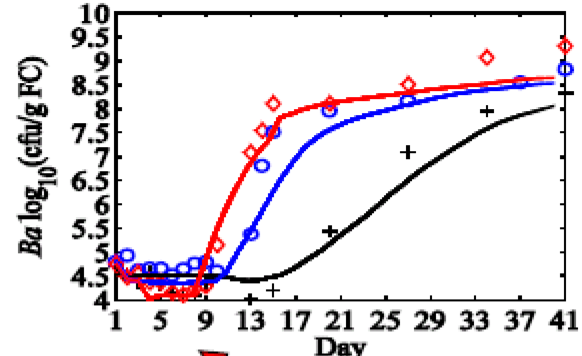
Sensoriel



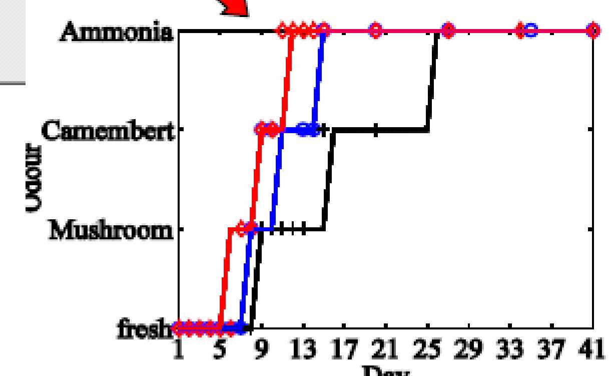
Littérature



Expériences



Taux de prédiction du modèle : 85%



Conclusions

- BNs are useful when
  - Quantitative and qualitative data in one model
  - Some relationships are not completely known
  - Data from heterogeneous sources
  - Need to add non-coded expert knowledge inside the model

Conclusions

- Cases where BNs might not be that useful
  - Only quantitative variables
  - Need for deterministic results
  - Well known phenomena

QUESTIONS?

Références

- **Les Réseaux Bayésiens**
  P. Naïm, P.H. Wuillemin, Ph. Leray, O. Pourret, A. Becker (Eyrolles).

- **Probabilistic reasoning in Intelligent Systems : Networks of plausible inference**
  J. Pearl (Morgan Kaufman).

- **An introduction to Bayesian Networks**
  F. Jensen (Springer Verlag).

- **Probabilistic Networks and Expert Systems**
  R.G. Cowell & al. (Springer Verlag)

- **Learning Bayesian Networks**
  R. Neapolitan (Prenctice Hall)

- **Learning in Graphical Models** - Jordan M.I. ed. (Kluwer)

Expert knowledge integration to model complex food processes.
Application on the camembert cheese ripening process (Elsevier, 2011)
http://www.sciencedirect.com/science/article/pii/S0957417411004763