

Fouille de données et apprentissage confrontés à 3 applications

Aide à la décision
en médecine et environnement

Marie-Odile Cordier
Professeur Université Rennes1/ IRISA
Equipe DREAM

Diagnostic, REcommandations d'Actions et Modélisation

Aide à la surveillance et la gestion de systèmes complexes

Marie-Odile Cordier (responsable, Prof. Rennes1)

Véronique Masson (MC Rennes1)

Yves Moinard (CR Inria)

René Quiniou (CR Inria)

Sophie Robin (MC Rennes1)

Laurence Rozé (MC Insa)

Xavier Le Guillou (ATER Rennes1)

Alice Marascu (Postdoc Inria)

Tassadit Bouadi (Doctorante)

Florimond Ployette (Ingénieur)

+ collaborateurs extérieurs :

Philippe Besnard (DR CNRS, IRIT Toulouse)

Christine Largouët (MC Ensar)

Thomas Guyet (MC Ensar)

+ ex-doctorants :

Elisa Fromont, Francois Portet, Ronan Trepos, Alexandre Vautier

+ Tristan Moreau, ex-ingénieur

Acquisition de connaissances et aide à la décision

- **Diagnostic / réparation:** Acquérir des motifs caractéristiques de situations à risque
 - Surveillance de réseaux de télécommunications
 - Surveillance cardiaque
 - Détection d'attaques informatiques
- **Prognostic :** Acquérir des motifs précurseurs de situations à risque
 - Surveillance cardiaque
 - Surveillance d'un troupeau d'élevage
- **Aide à la gestion d'un système complexe :** Découvrir des relations d'influence, de causalité afin d'élaborer des moyens d'action
 - Surveillance d'un bassin versant pour préserver/améliorer la qualité de l'eau
 - pollution par les pesticides
 - pollution par les nitrates

Acquisition de données et aide à la décision

- **1^{er} cas** : extraction de motifs “intéressants” puis utilisation en ligne (le système concerné ne change pas ...)
 - Motifs cardiaques, motifs précurseurs vêlage ...
- **2nd cas** : extraction de motifs “intéressants” + évolution adaptative des motifs en ligne (shift / drift sur les flux)
 - Attaques d'intrusion réseaux, patients cardiaques sous thérapie ...
- **3^{ième} cas** : extraction de motifs en ligne, de manière interactive, pour aider à faire évoluer le système
 - Aider à la gestion d'un bassin versant par modification des pratiques agricoles, aménagement du territoire ...

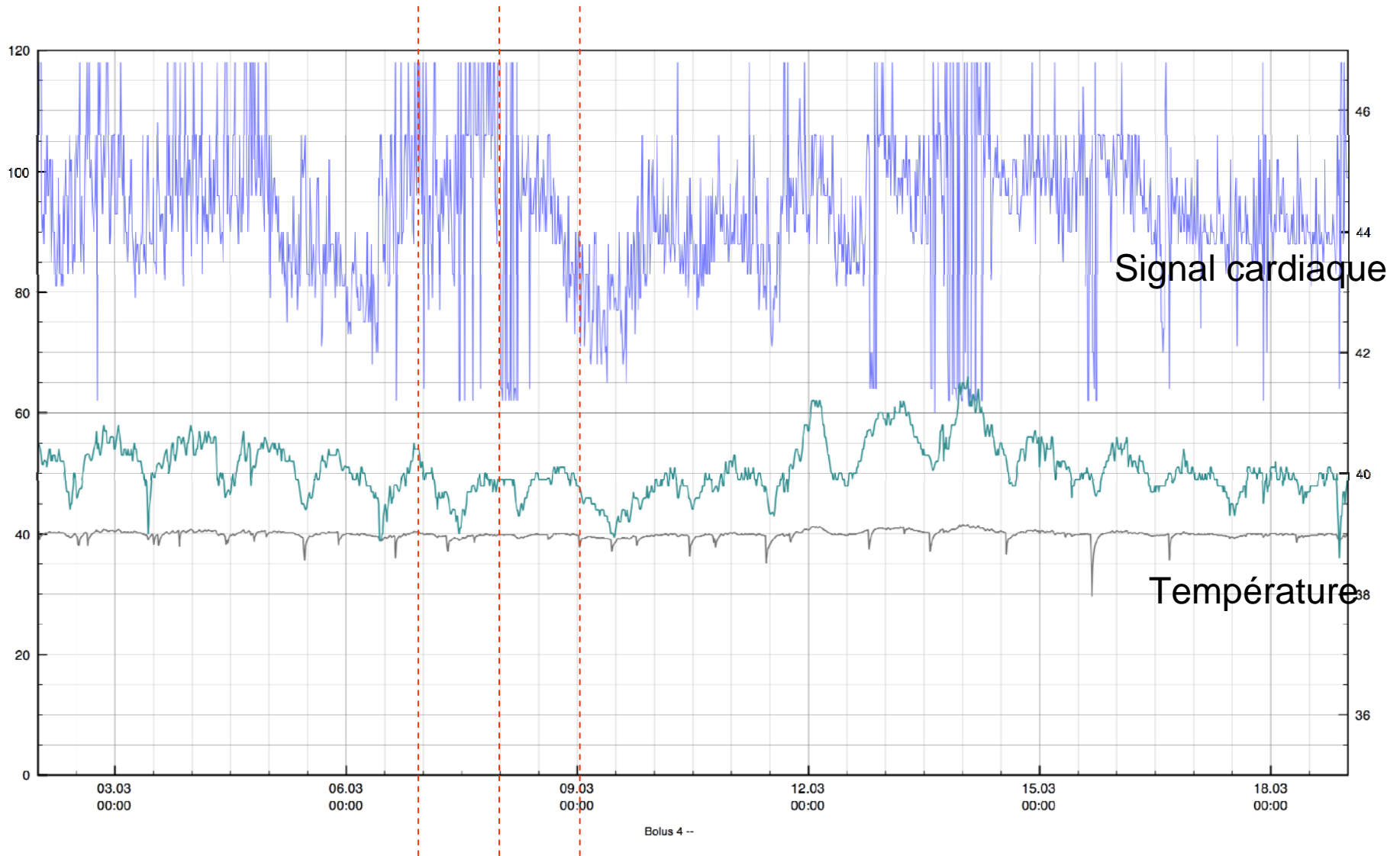
Des données ...

– Comment représenter les données pour permettre l'apparition des éléments intéressants (pépites)?

➤ Rôle de l'expertise

... aux connaissances

Que voit-on?



Fil directeur : 3 applications

- Données :
 - Comment représenter les données pour faire “apparaître” les éléments intéressants?
 - Comment tirer parti de l’expertise existante?

 Prétraitement

- Architecture / Résultats
- Morale

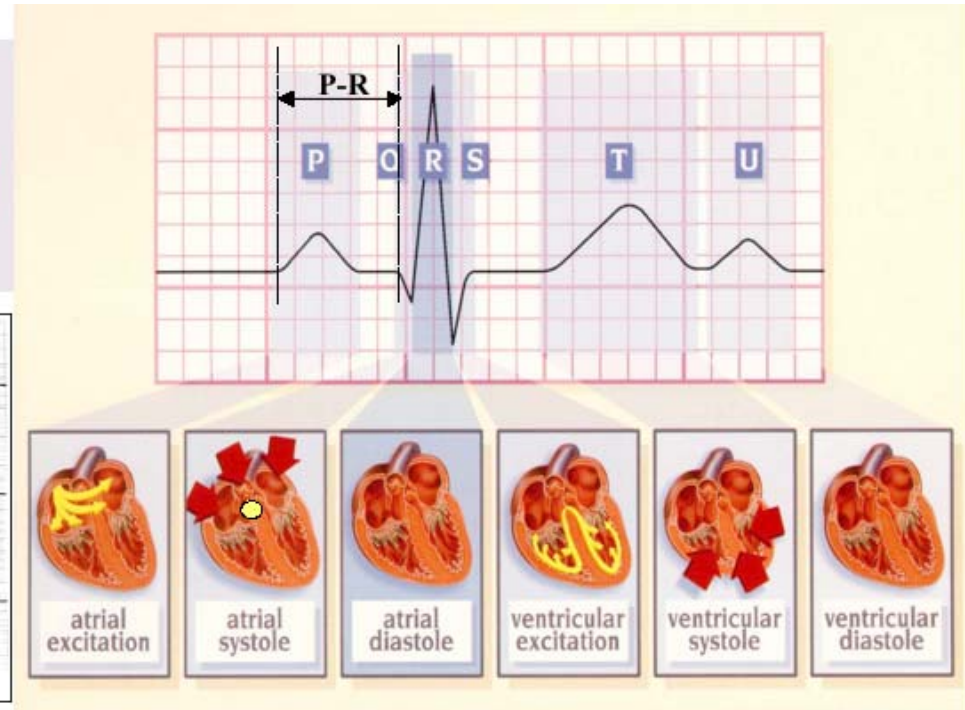
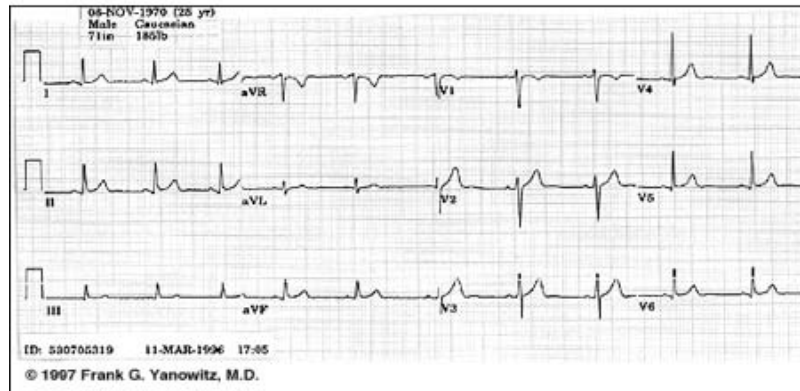
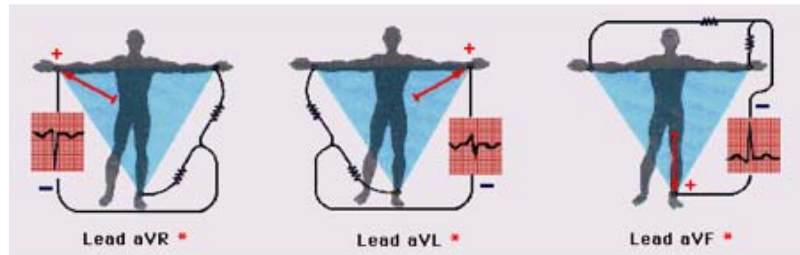
CALICOT

- ACI PISE + RNTS CEPICA en collaboration avec le LTSI Rennes, le CHU Rennes1 et Ela-Medical (1997-2008)
Lucie Callens, Guy Carrault, Marie-Odile Cordier, Elisa Fromont, Philippe Mabo, François Portet, **René Quiniou**, Feng Wang



- Objectif général :
 - Aide à la détection d'arythmies cardiaques dans un contexte hospitalier (USIC). Analyse des signaux (électrocardiogrammes, pression artérielle, respiration) pour émettre des alarmes à destination du personnel médical
- Enjeu :
- éviter les alarmes intempestives (taux important de fausses alarmes)
 - Améliorer la confiance des médecins envers le diagnostic proposé
- Développement de prothèses cardiaques + détection intracardiaque

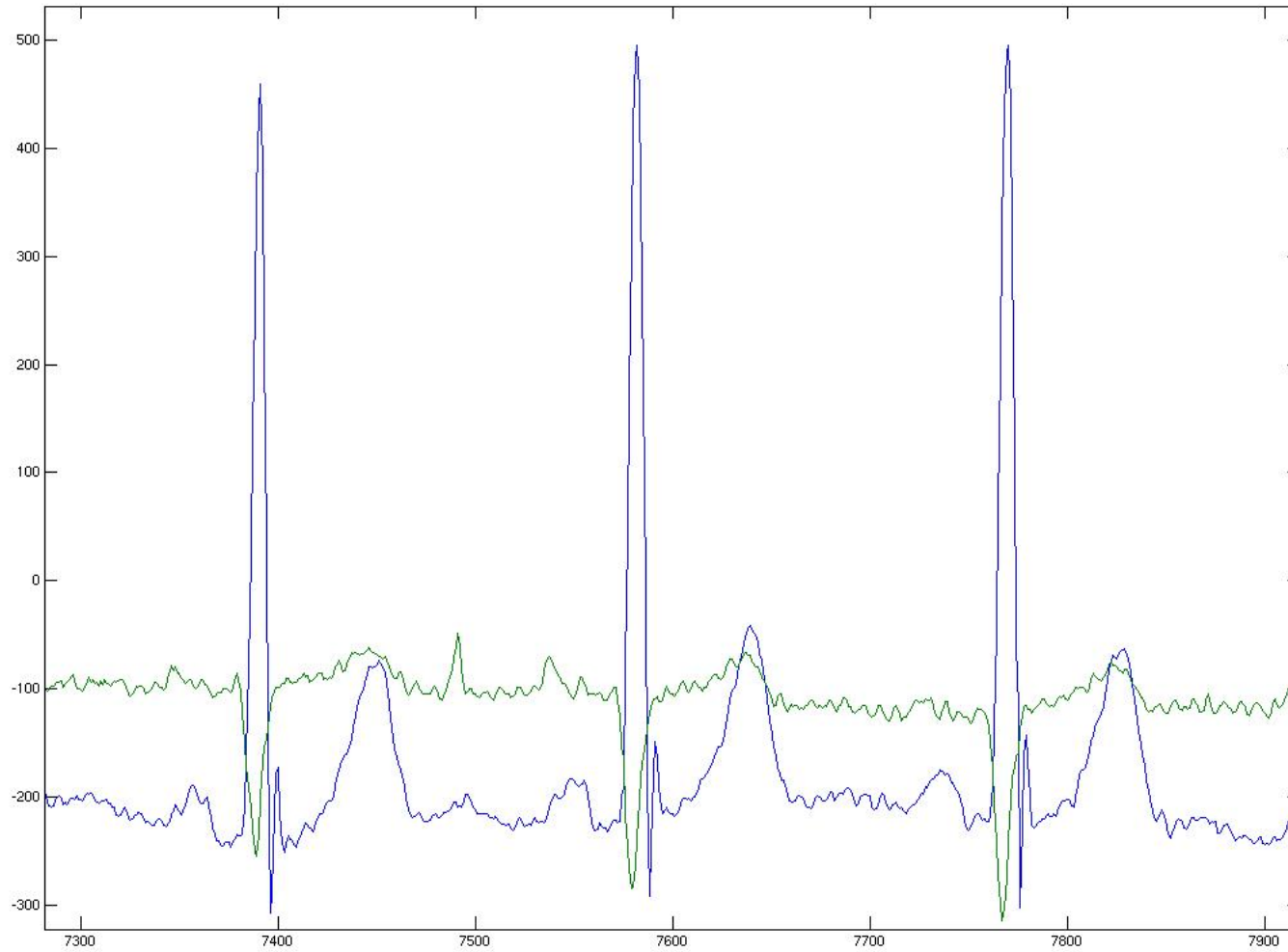
Contexte: monitoring cardiaque



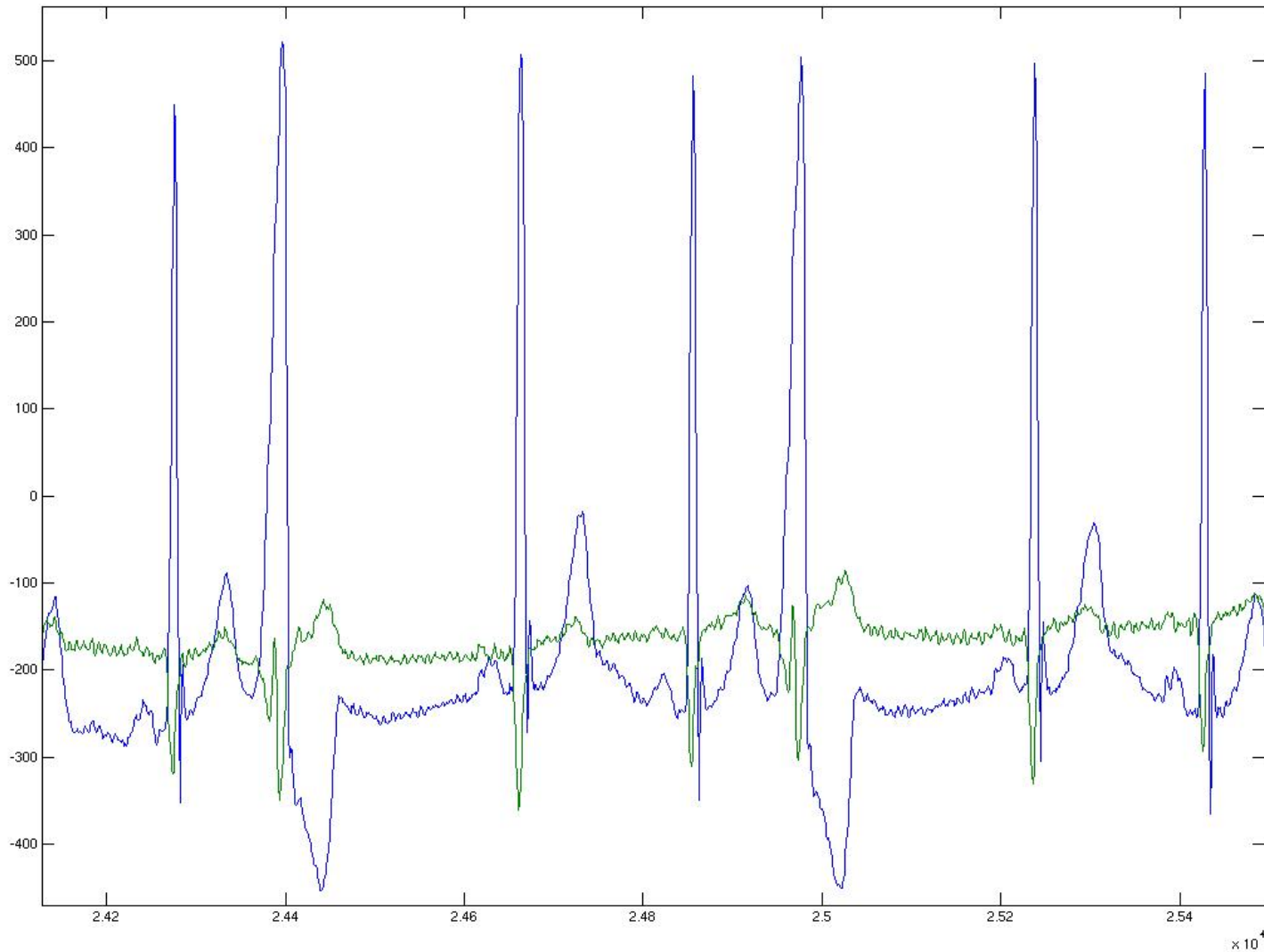
Acquisition de connaissances et aide à la décision

- **Diagnostic / réparation:** Acquérir des motifs caractéristiques de situations à risque
 - Surveillance de réseaux de télécommunications
 - Surveillance cardiaque
 - Détection d'attaques informatiques
- **Prognostic :** Acquérir des motifs précurseurs de situations à risque
 - Surveillance cardiaque
 - Surveillance d'un troupeau d'élevage
- **Aide à la gestion d'un système complexe :** Découvrir des relations d'influence, de causalité afin d'élaborer des moyens d'action
 - Surveillance d'un bassin versant pour préserver/améliorer la qualité de l'eau
 - pollution par les pesticides
 - pollution par les nitrates

Données

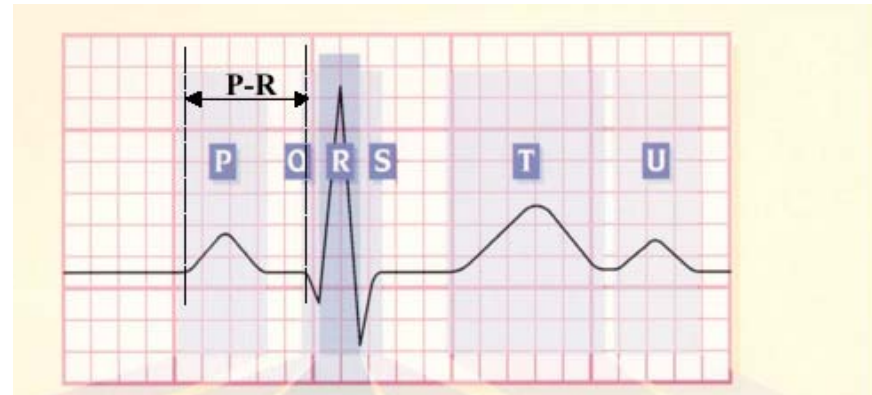


Données



Expertise : importante, formalisée

- Concepts : objets, attributs, contraintes temporelles ...
 - Complexe QRS : (modèle profond)
 - Onde P, Onde T, Onde U
 - Intervalle ...
 - Cycle cardiaque



Expertise : importante, formalisée

- Concepts : objets, attributs, contraintes temporelles ...
 - Complexe QRS : (modèle profond)
 - Onde P, Onde T, Onde U
 - Intervalle ...
 - Cycle cardiaque
- Base de données : signaux annotés (Mimic)
- Situations : décrites (voir doc. pour formation médicale)

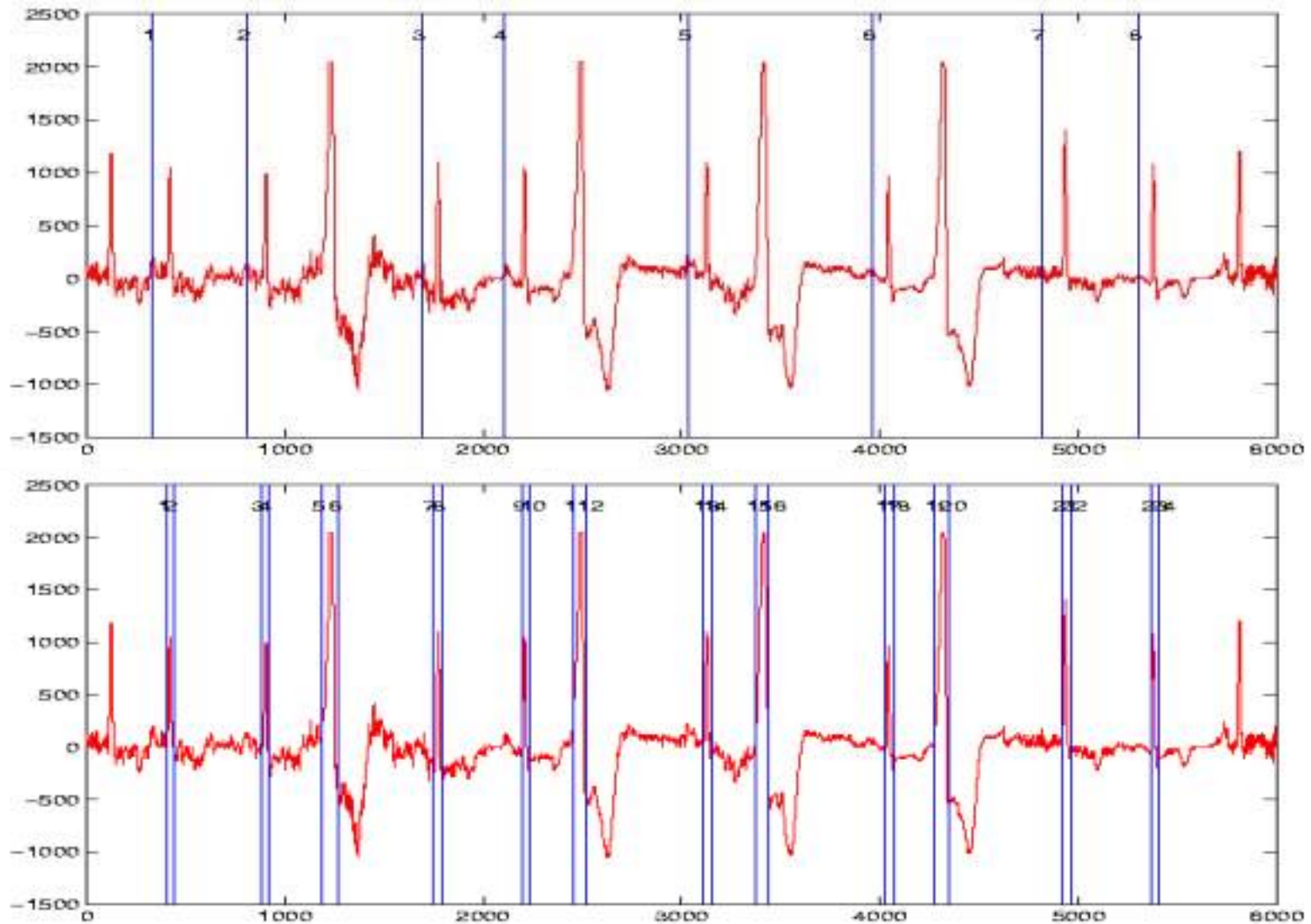
.... autant d'ondes P que de QRS : bradycardie sinusale
plus d'ondes P que de QRS : bloc auriculo-ventriculaire type2
ondes P et QRS dissociés : bloc auriculo-ventriculaire type3

Prétraitement :

des signaux à leur représentation qualitative

- S'appuyant sur l'expertise existante :
 - Identification ordonnée des complexes QRS, des ondes P, des ondes T et U, estimation des distances temporelles entre leurs occurrences, estimation de la durée du cycle ...
 - Utilisation d'algorithmes de traitement du signal existants et dédiés à cette identification

Bigeminy bruité



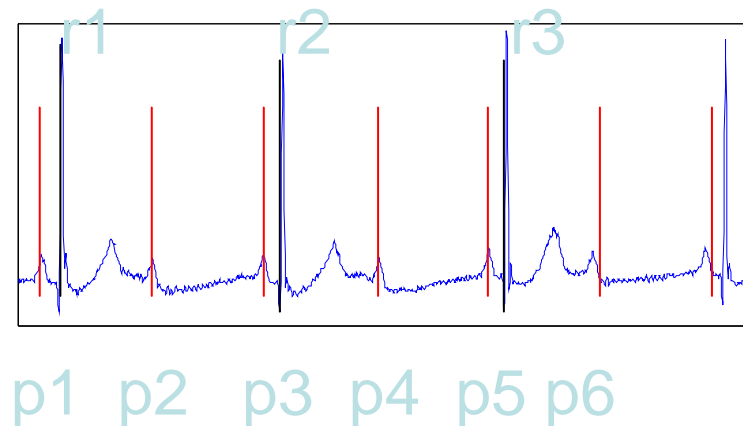
annotations onde P, annotations QRS

Prétraitement :

des signaux à leur représentation qualitative

- S'appuyant sur l'expertise existante :
 - Identification ordonnée des complexes QRS, des ondes P, des ondes T et U, estimation des distances temporelles entre leurs occurrences, estimation de la durée du cycle ...
 - Utilisation d'algorithmes de traitement du signal existants et dédiés à cette identification
- Langage de représentation : Logique du premier ordre
ex : `qrs_complex(Nom, Instant, Morph, Pre_Evt)`

```
begin(exemple-mobitz2_231_1).  
  p(p1, 124, normal, noPreWave).  
  qrs_complex(r1, 295, normal, p1).  
  p(p2, 968, normal, r1).  
  p(p3, 1812, normal, p2).  
  qrs_complex(r2, 1967, normal, p3).  
  p(p4, 2687, normal, r2).  
  p(p5, 3523, normal, p4).  
  qrs_complex(r3, 3669, normal, p5).  
  ...  
end(exemple-mobitz2_231_1).
```



Acquisition des motifs caractéristiques

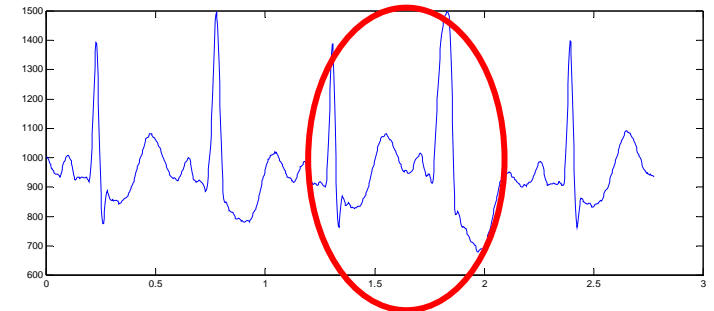
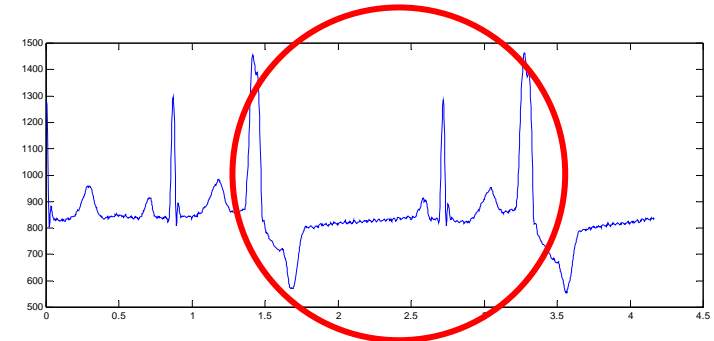
- Utilisation de la PLI :
 - Apprentissage supervisé : existence de données étiquetées, voir BD Mimic
 - Recherche de motifs discriminants : diagnostic
 - Apprentissage relationnel + temporel
 - Biais s'appuyant sur l'expertise existante
- Architecture hors-ligne : base de chroniques

Exemples de règles apprises

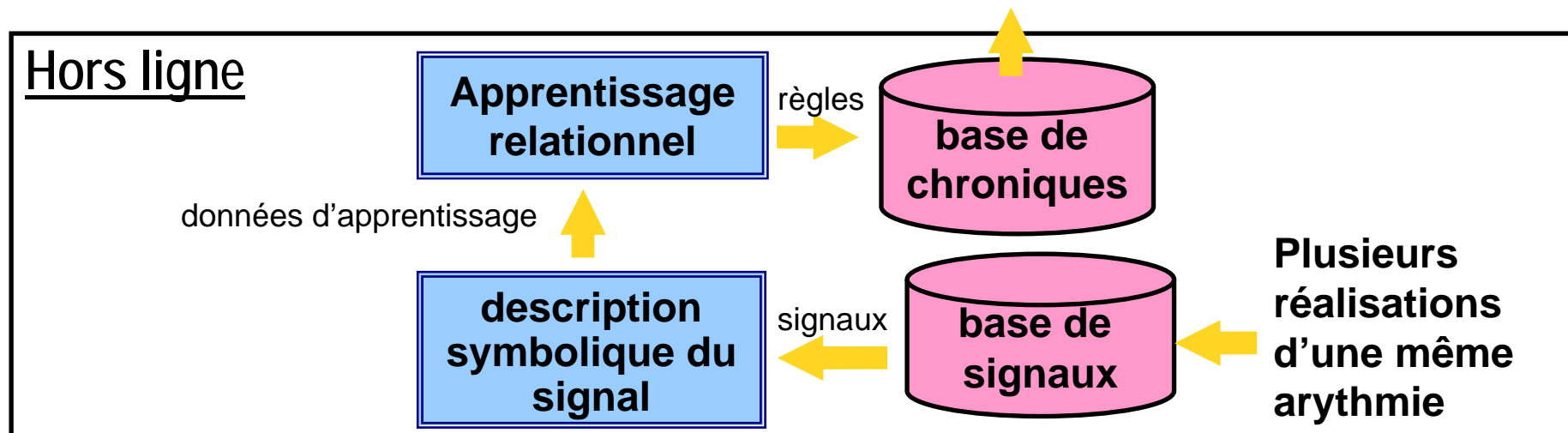
bigeminy :- %[15,0,0,0,0],[5,20,20,20,20]
qrs(R0,abnormal,_),
p(P1,normal,R0), qrs(R1,normal,P1),
qrs(R2,abnormal,R1).

bigeminy :- %[5,0,0,0,0],[12,17,18,18,20]
qrs(R0,normal,_),
p(P1,normal,R0), qrs(R1,abnormal,P1).

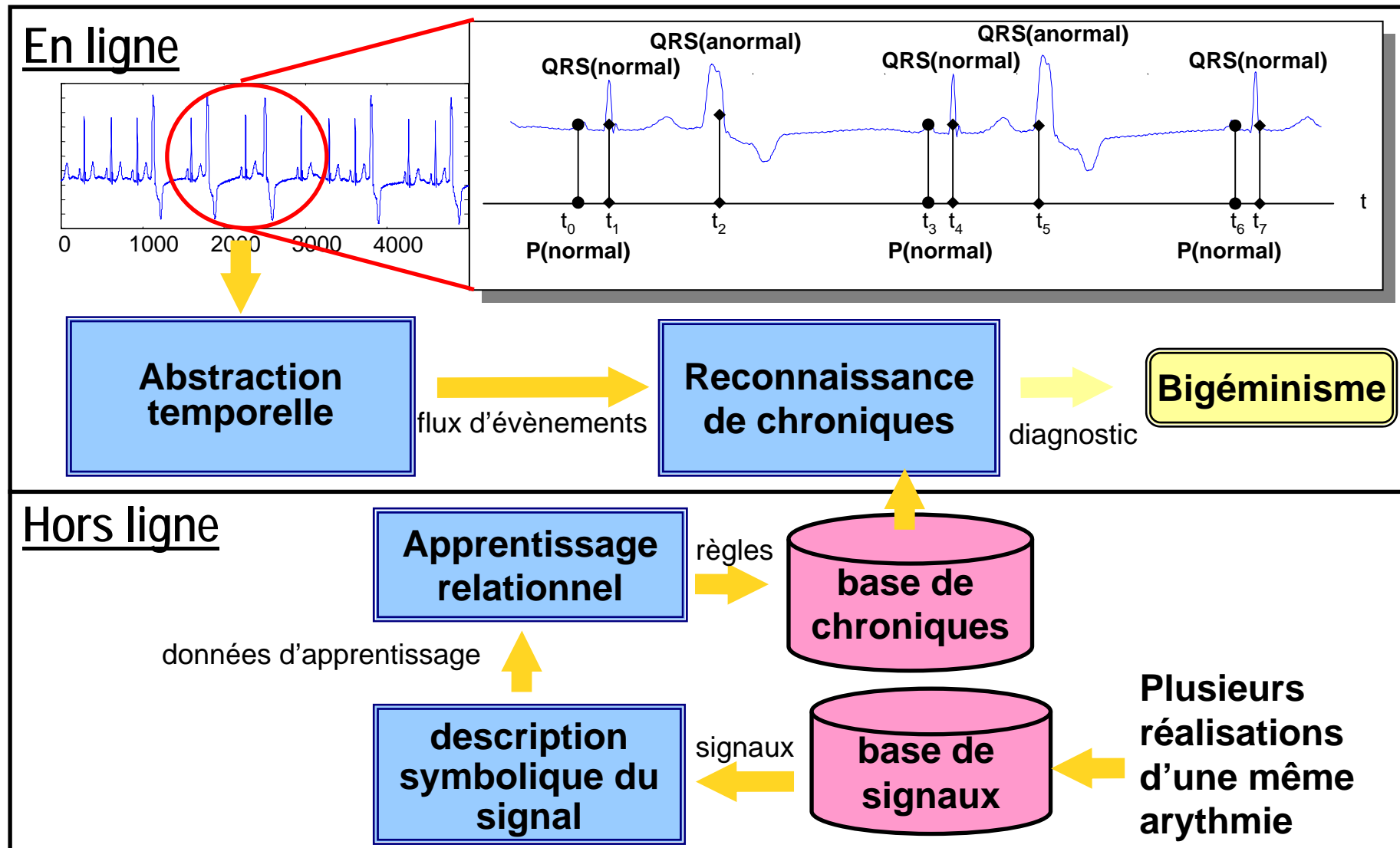
pvc :- [0,0,0,0,20],[20,20,20,20,0]
p(P0,normal,_),qrs(R0,normal,P0),
p(P1,normal,R0),qrs(R1,normal,P1),
qrs(R2,abnormal,R1).



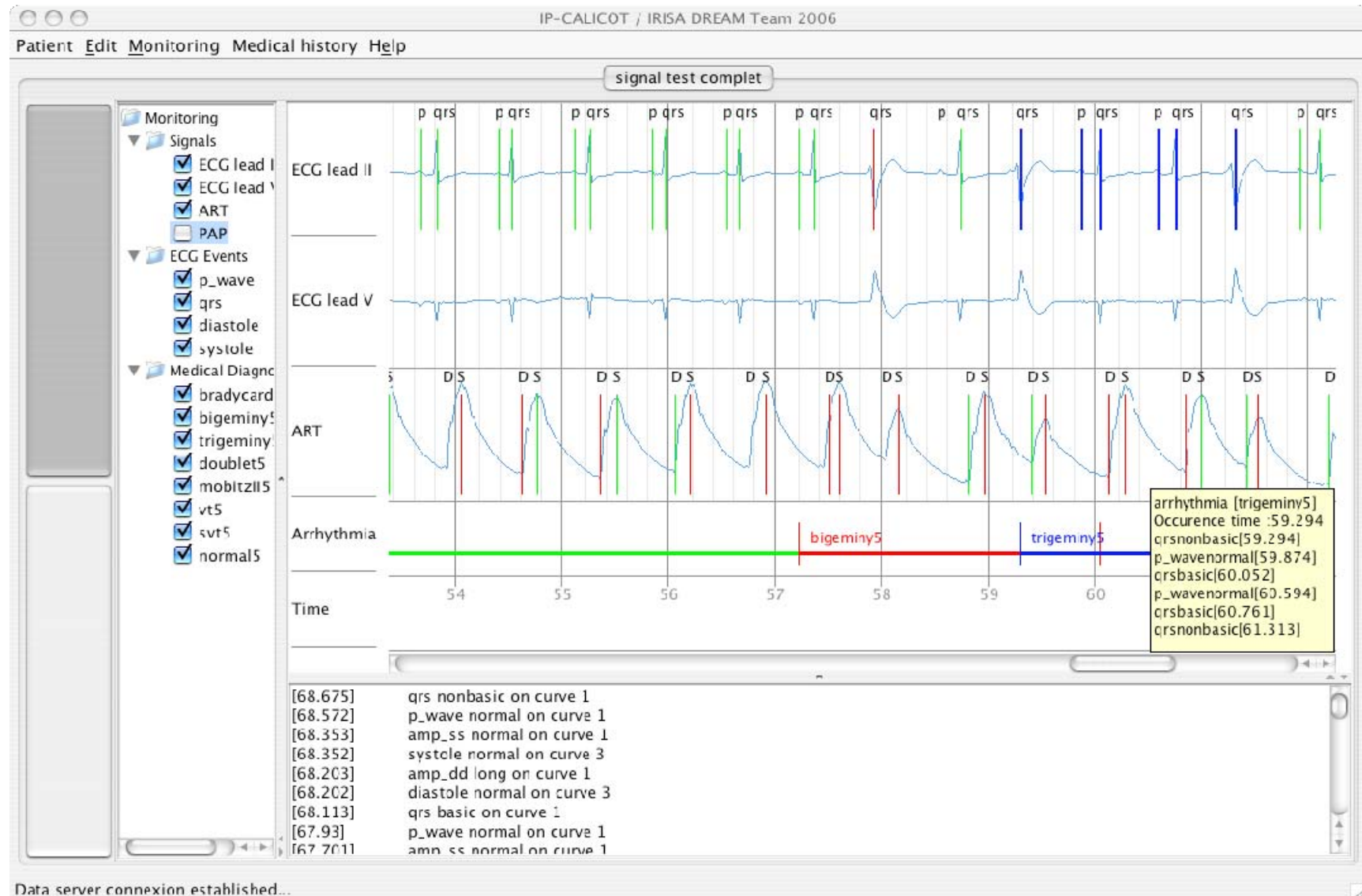
Architecture de Calicot



Architecture de Calicot



Implementation



Morale

- Importance de l'expertise existante :
 - Langage de représentation des données
 - Base d'exemples annotés
 - Biais sémantique lors de l'apprentissage
 - Validation des résultats

A-t-on appris? Découvert quelque chose?

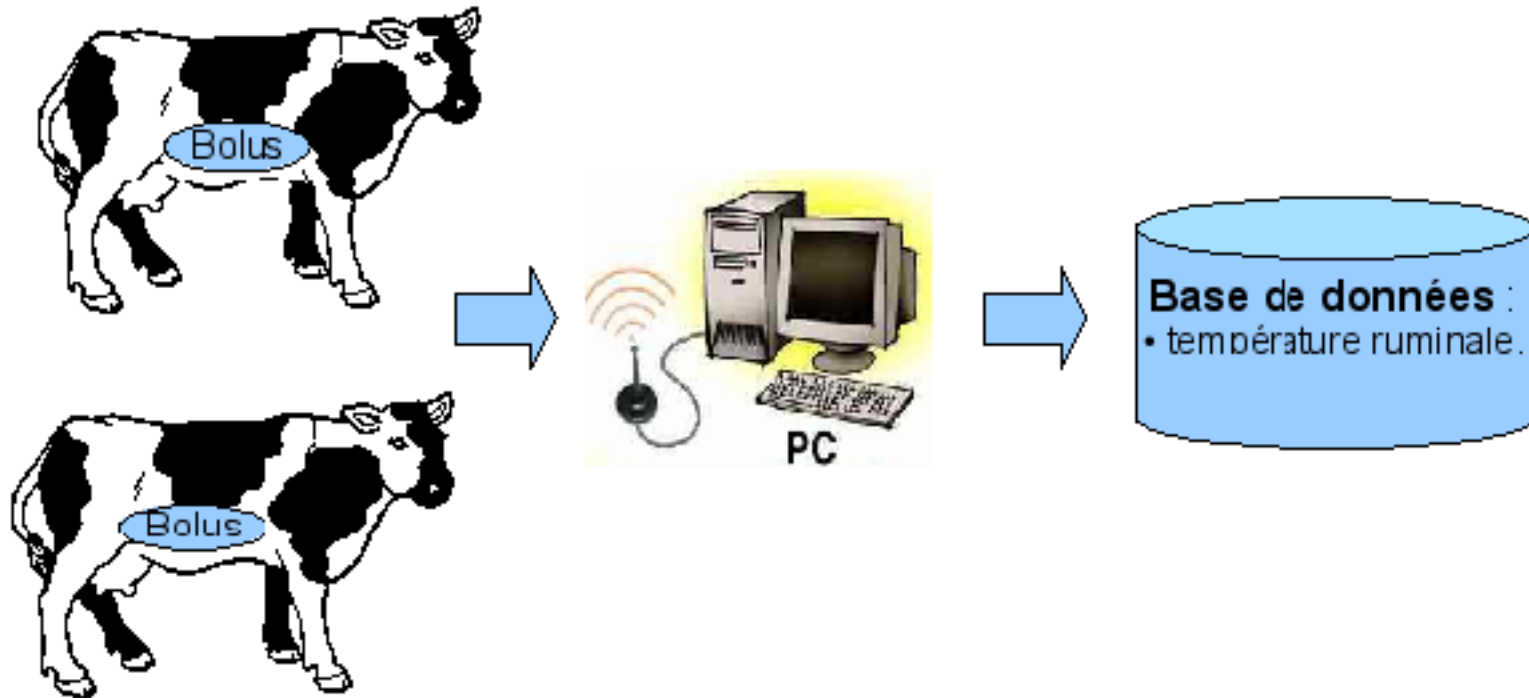
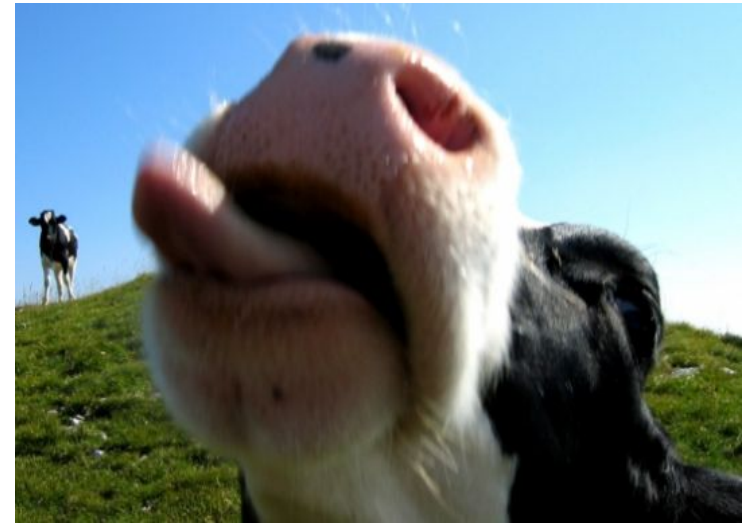
- Connaissances orientées décision
 - Description opérationnelle et efficacité
- Connaissances contextuelles
 - Contexte patient, capteurs, algos ...

Du diagnostic au pronostic ...

- Enjeux:
 - Découvrir des signes avant-coureurs d'arythmies et surtout de fibrillations ventriculaires
 - S'appuyer sur des capteurs nouveaux, en particulier dans le contexte de prothèses cardiaques

➡ Et si pas/peu/moins d'expertise?

Entrons dans le monde des ruminants!



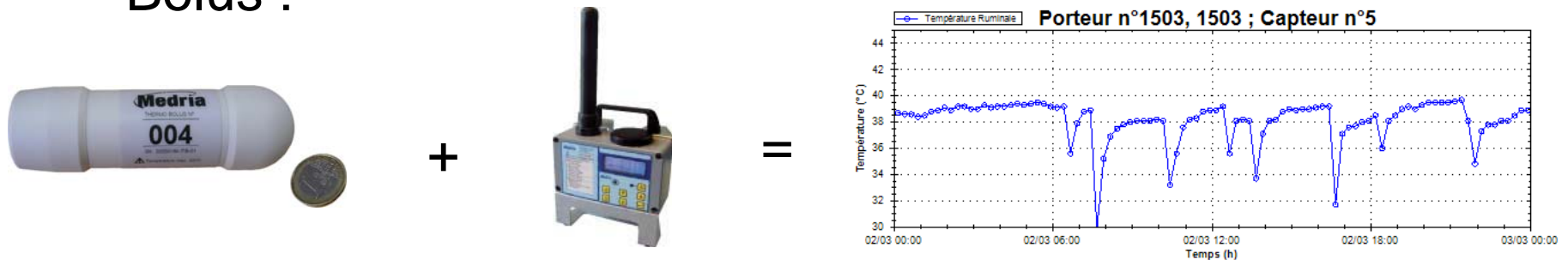
MOZAE : surveillance de troupeau d'élevage

MOnitoring Zootechnique des Animaux d'Elevage (2007-2009)

- Partenaires :

MEDRIA S.A.S, CRAB, INRA (UMR Prod. du Lait), Agrocampus-Ouest, Ecole Nat. Vétérinaire de Nantes (ENVN), Dream/IRISA : René Quiniou, Tristan Moreau, Marie-Odile Cordier)

- Bolus :



- Objectif : Aide à la détection précoce de vêlages, oestrus (chaleurs), maladies hyperthermiques / hypothermiques

- 5 générations de bolus, 533 bolus avalés, 450 vaches concernées + jeunes bovins, 7 élevages commerciaux + 3 fermes expérimentales, ...



**Vous n'aurez jamais été aussi proche
de vos animaux**

Cardio-bolus[®]
Thermo-bolus[®]
Base radio GSM



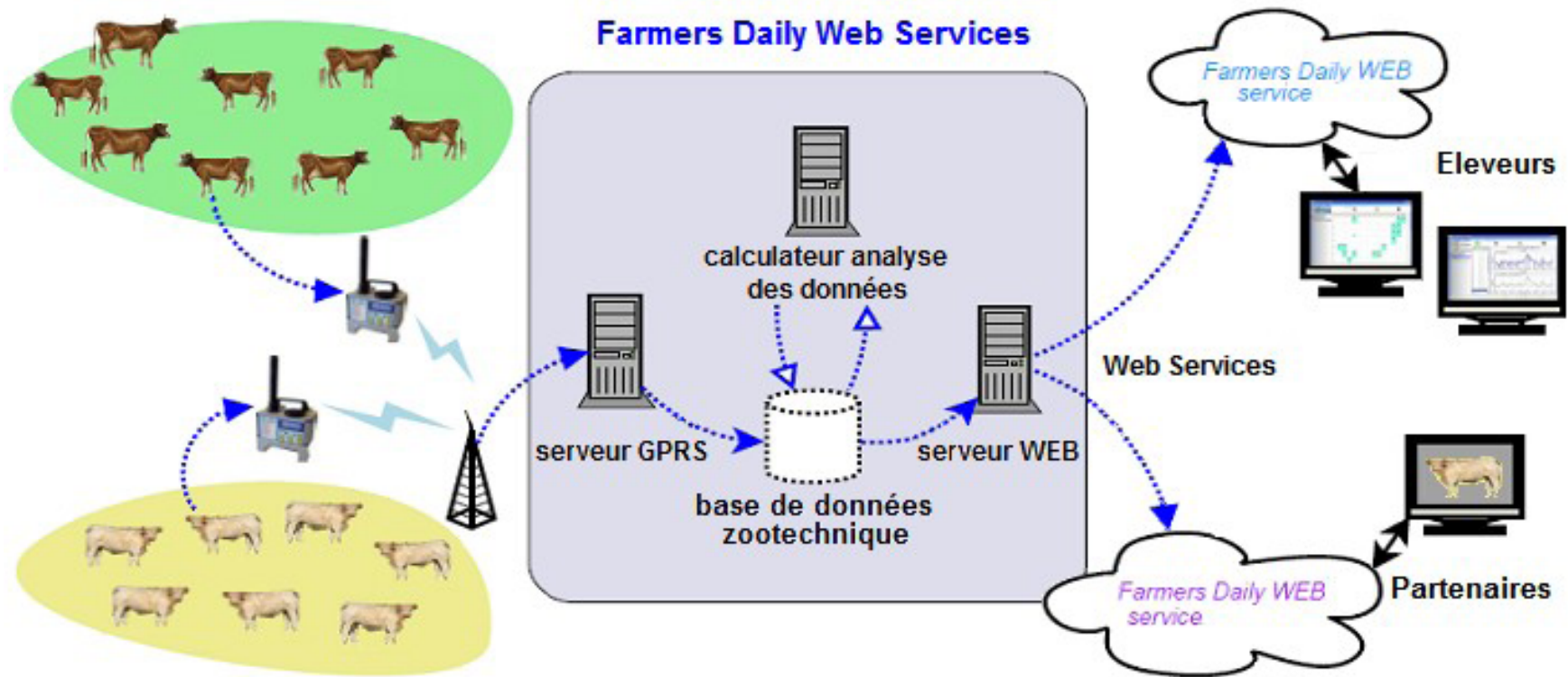
Équipements de suivi
zooteknique

*Un geste une fois
dans la vie
de l'animal !*

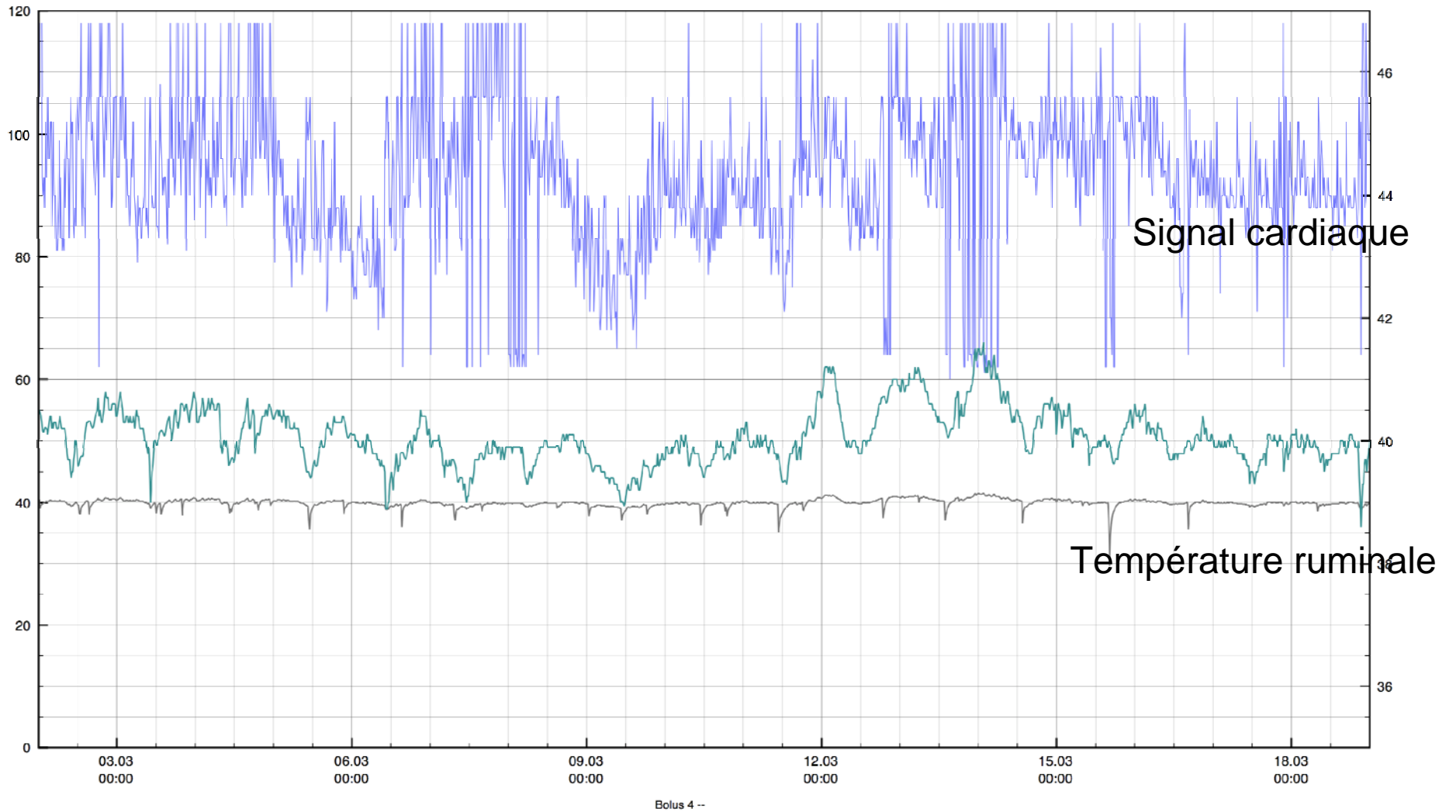
Logiciel d'exploitation
sur PC



Détection des chaleurs, vêlages
et troubles de la santé

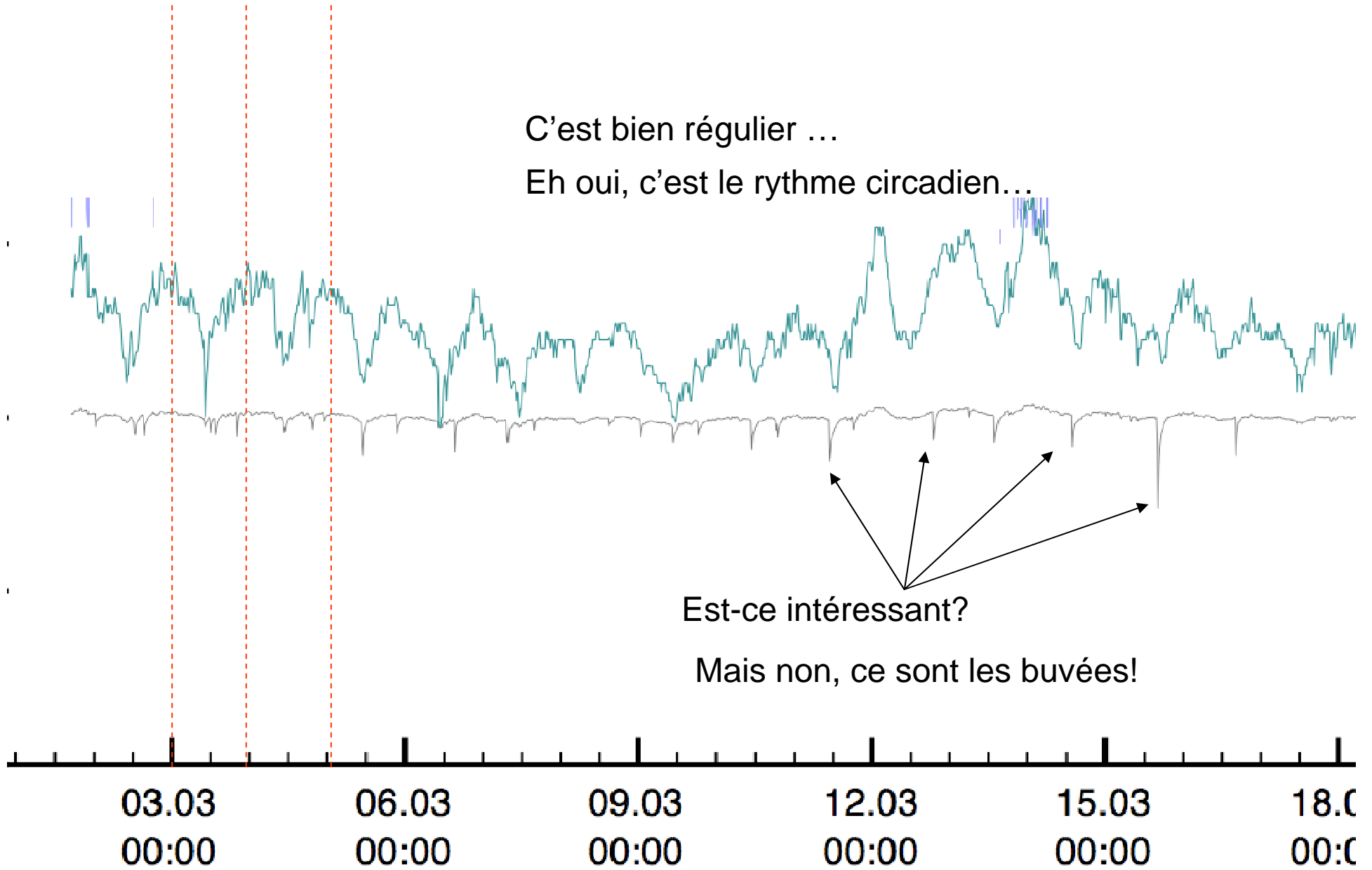


Données



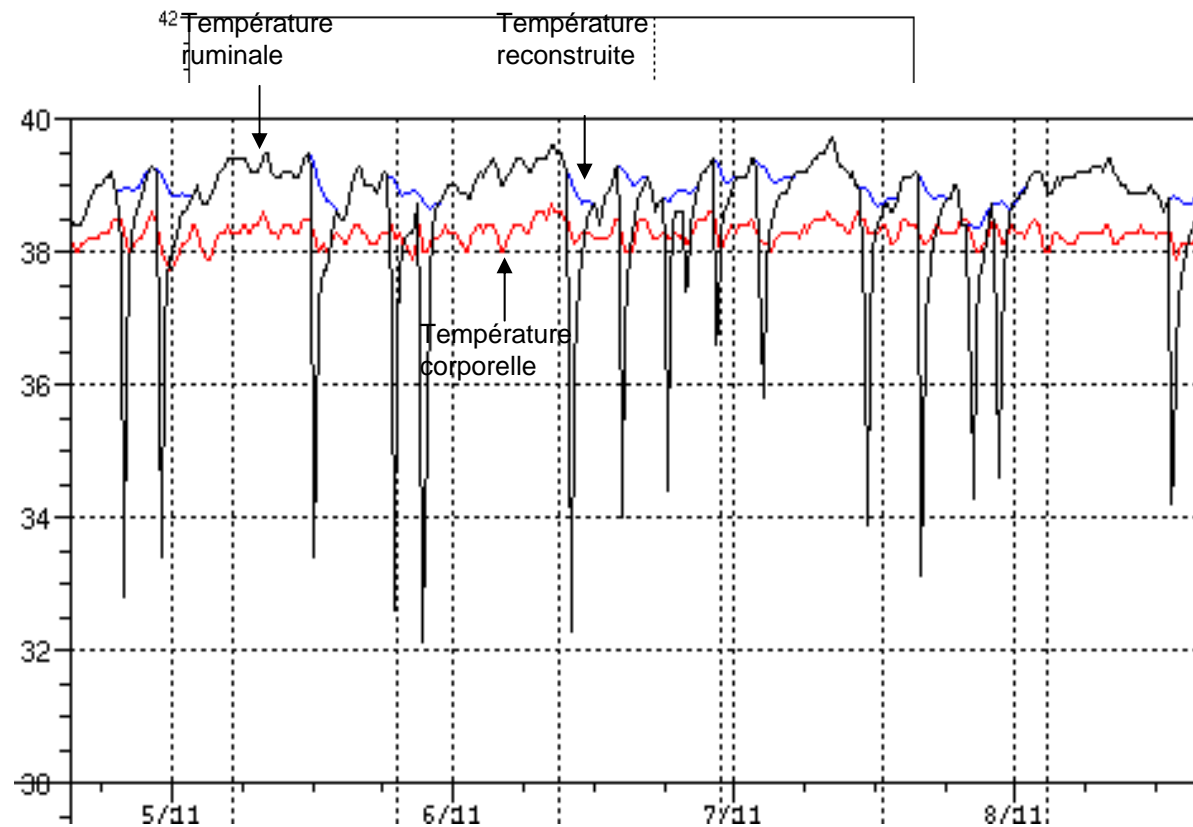
C'est bien régulier ...

Eh oui, c'est le rythme circadien...



Prétraitement : “nettoyer” les données

- Buvées : supprimer le bruit...



Corrélation (après reconstruction) de 0.79 avec température vaginale/rectale

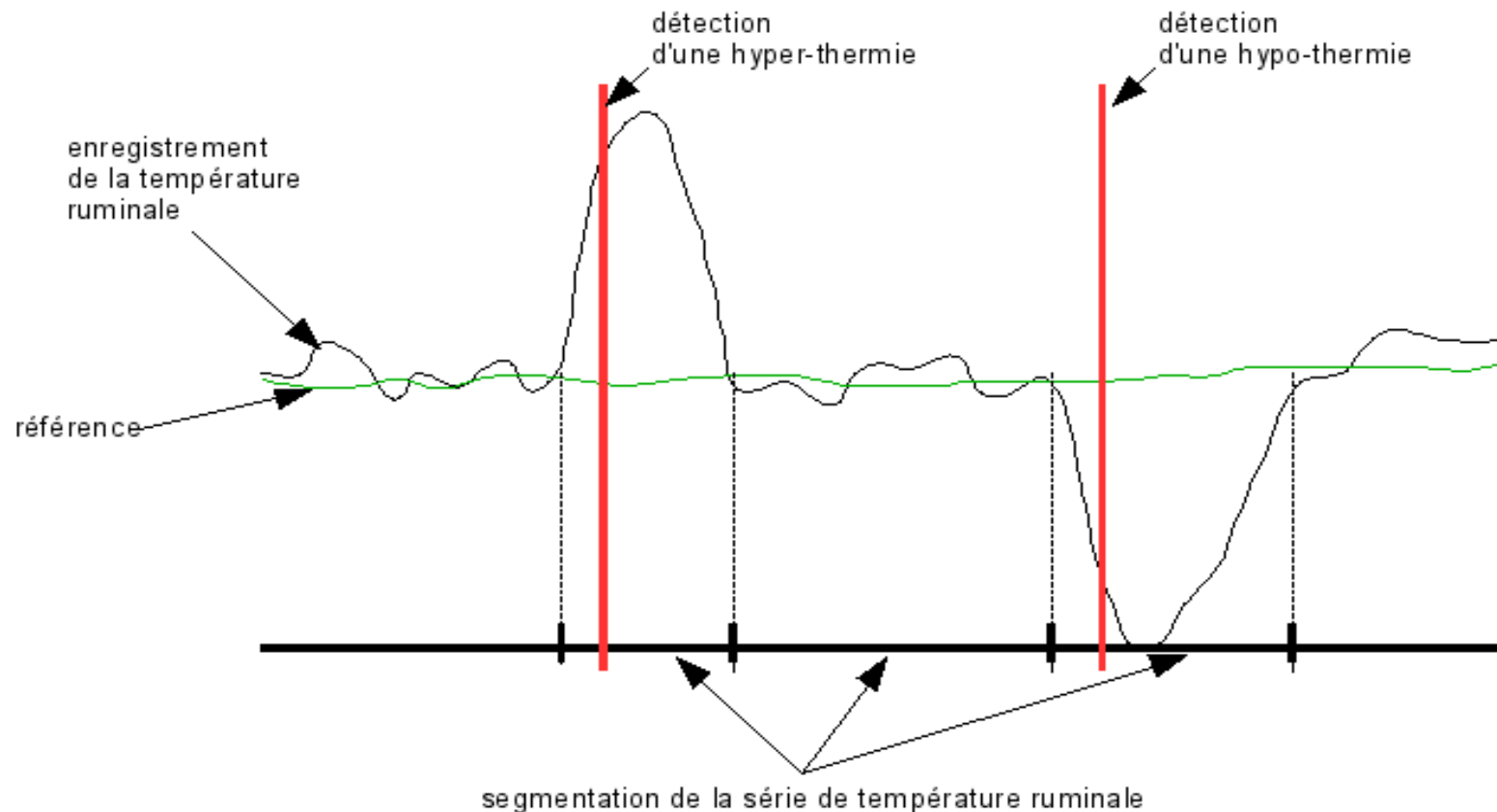
Autres “bruits”?

- “Bruit” : Phénomènes “cachant” ceux que l’on veut découvrir
 - Effet buvée
 - Rythme circadien
 - Effet de la digestion
 - Influence température externe (étable/ prairie)
 - Influence du stress (trains, chiens...)
- Rechercher une référence :
 - Intraindividu
 - intratroupeau

Représentation qualitative des données

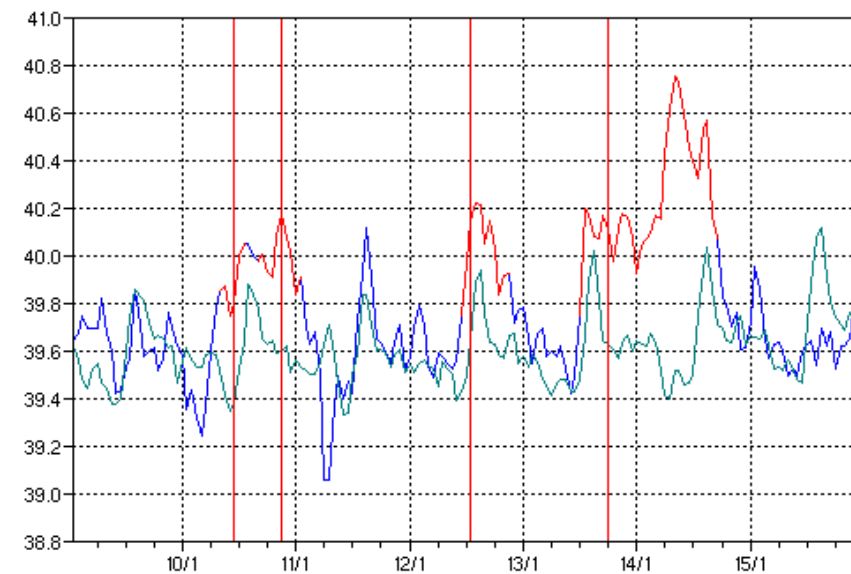
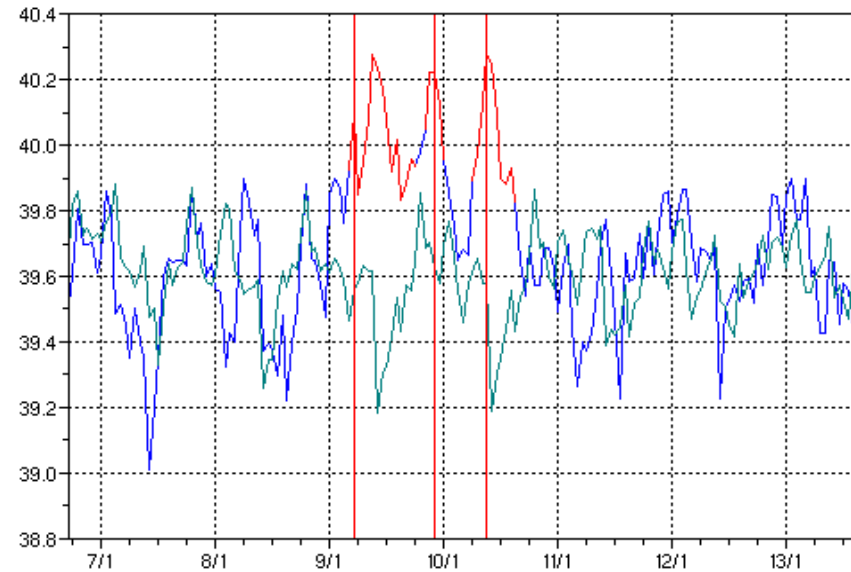
- Concept “intéressant”:
 - Changement de température (hausse/baisse)
 - Tendance
 - Changement brusque, durée du changement
 - Changement significatif : quelle est la référence?
 - Eliminer l’effet du contexte : “désaisonnaliser”
 - Intra-individu : moyenne horaire transversale
 - Inter-individu : moyenne par lot
 - Evaluer la différence donnée / référence
 - Dynamic time-warping (recalage)
 - Somme des erreurs cumulées + seuil: Page-Hinkley

Représentation des données : du signal à un ensemble d'événements

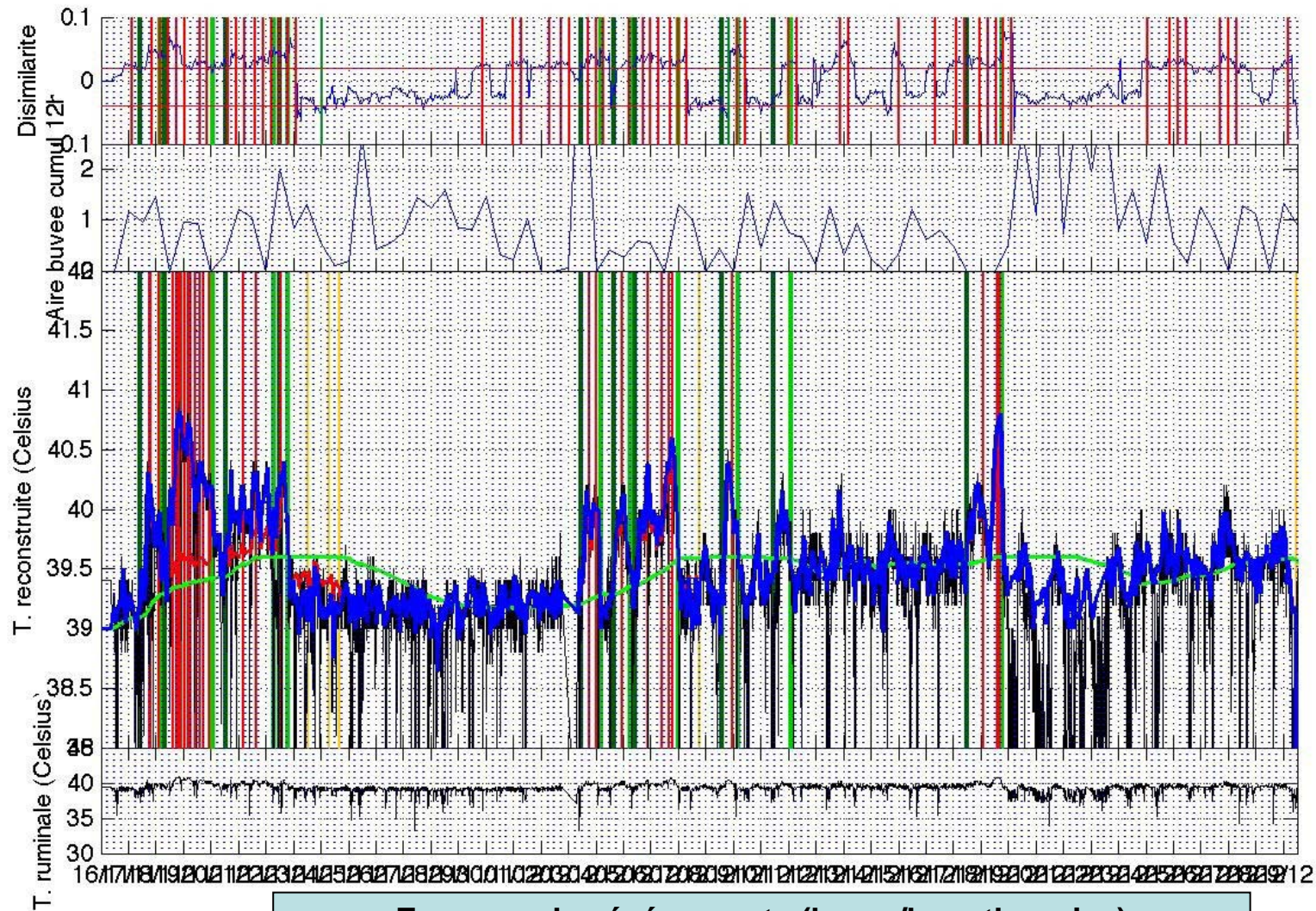


La courbe bleue est la température ruminale, la courbe verte la référence.

Les traits rouges représentent les instants de détections des hyper-hypothermies.



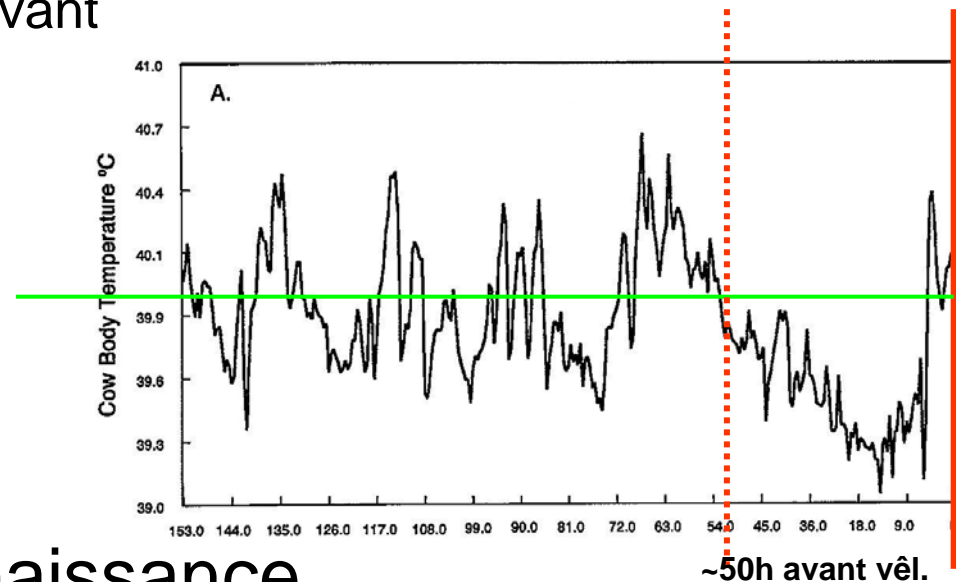
Représentation des données : du signal à un ensemble d'événements



En rouge, les événements (hyper/hypothermies)

Résultats : Vêlage

- Recherche de motifs précurseurs du vêlage
- Expertise littérature (avec imprécisions) :
 - Baisse (0,4 à 0,8°) 2- 3 jours avant; retour à la normale lors du vêlage, ou quelques heures avant



- Confirmation de la connaissance
- Possibilité de mettre en place une alerte fiable pour les vaches à terme

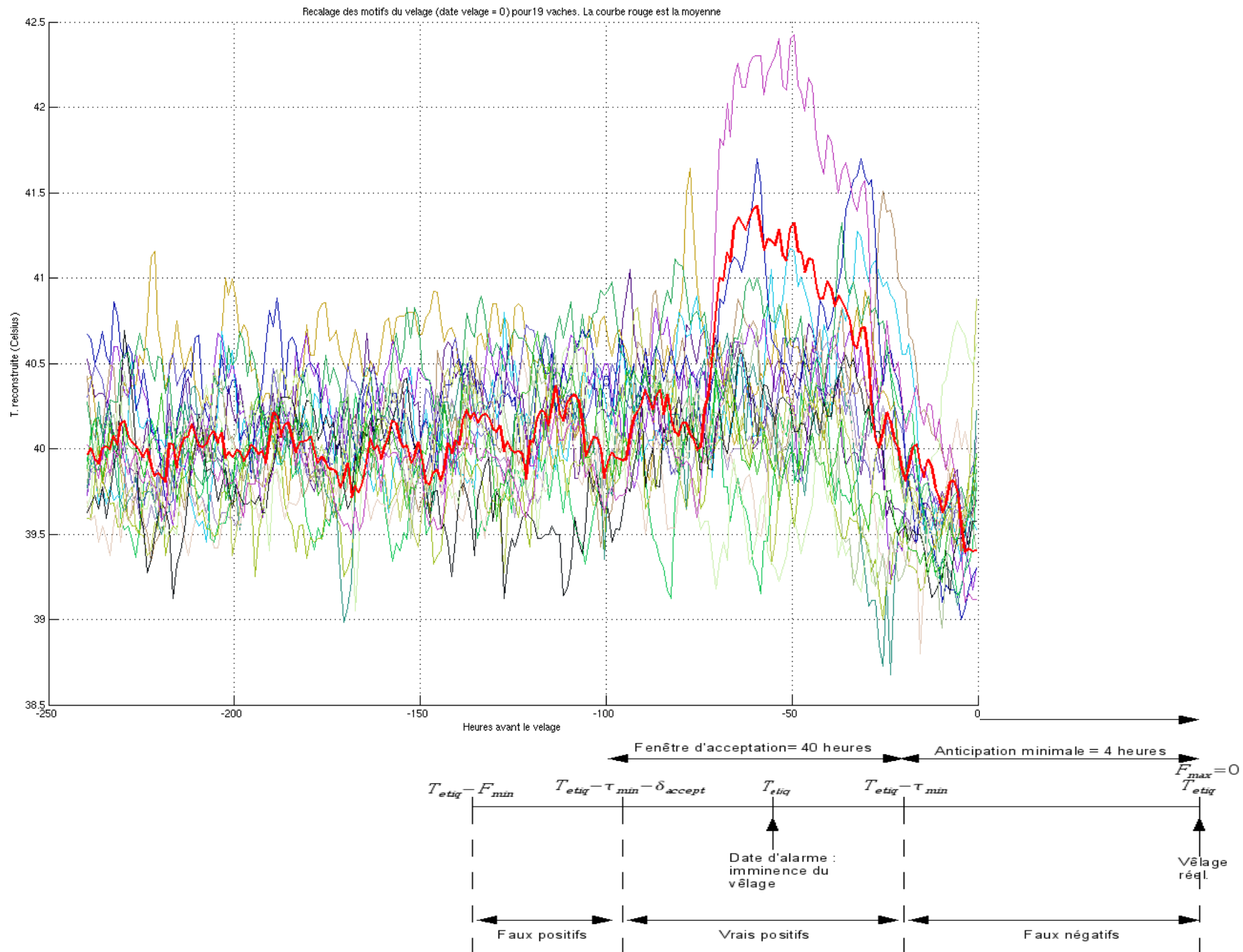


Figure 22 - Illustration de la mesure des performances en détection dans le cas du vêlage

Résultats : oestrus (chaleurs)

- Recherche de motifs précurseurs de l'oestrus pour insémination efficace
- Expertise :
 - Littérature : la température augmente au moment de l'oestrus de 0,3 à 0,9 ° (chaleur!) - La température extérieure aurait plus d'influence sur la température corporelle que l'oestrus (Lewis et Newman 1984)....
 - Signes comportementaux : voir fiche
- Essai de classification hiérarchique et de fouille de séries temporelles
- Infirmité :
 - Pas de “chaleurs” !
 - Combiner avec le signal cardiaque ou podomètre ou mesure de l'activité au collier...

Annexe 2 : Fiche d'enregistrement des chaleurs observées

Observateur	
Vache	
Date	
Heure	

1 - DETECTION	<i>Visuelle</i>	<i>Détecteur</i>	<i>Les 2</i>
2 - AUTRE VACHE en chaleurs		<i>OUI</i> <i>NON</i>	
3 - INTENSITE selon vous	<i>discret</i>	<i>moyen</i>	<i>visible</i>

4 - COMPORTEMENTS observés

- Accepte de se faire chevaucher par la vache N°
- Chevauche ou tente de chevaucher la vache N°
- Chevauche par l'avant la vache N°
- Pose son menton sur la vache N°
- Remue la vulve de la vache N°
- Se frotte à la vache N°
- Vache très agitée
- Présence de glaires translucides à la vulve
- Meuglements
- Baisse significative de la production laitière
- Autres comportements observés

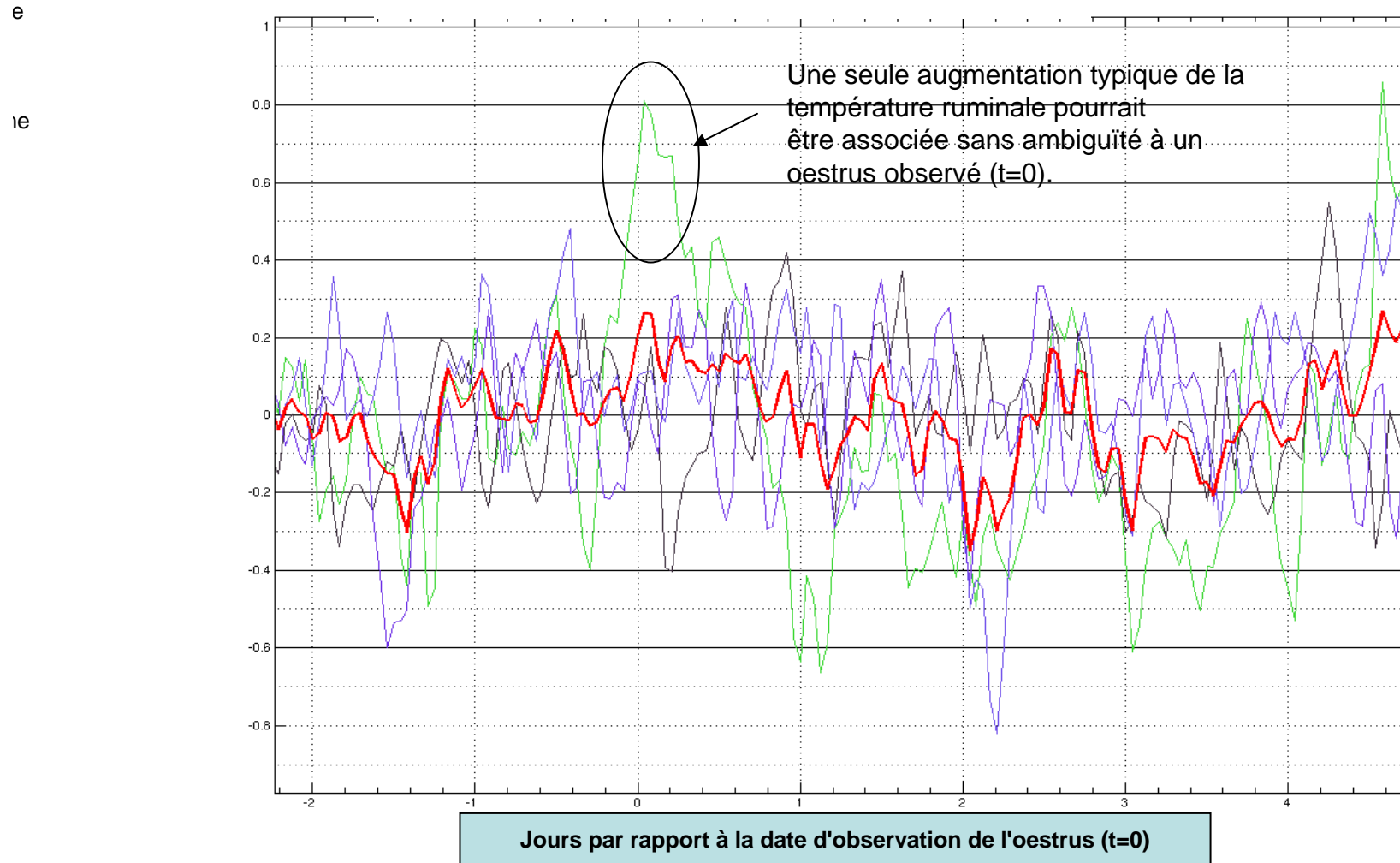
Description :

.....

.....

.....

Oestrus : résultats



Enregistrements de la température ruminale recalées par rapport à la date d'observation de l'oestrus et centrés par rapport à la moyenne. Les vaches ont été sélectionnées sur la base d'un comportement manifeste lors de l'oestrus correspondant à une acceptation du chevauchement. La courbe rouge correspond à la moyenne des autres courbes.

Maladies hyperthermiques

- Recherche de motifs précurseurs de maladies hyperthermiques chez la vache et le jeune bovin
- **Expertise** : « chez les jeunes bovins brusque élévation de la température passant d'une valeur moyenne de 39-39.5 °C à 41 °C et au delà, suivi, éventuellement d'un plateau, puis d'une descente. L'épisode relatif à une hyperthermie peut durer de 5 heures à plusieurs jours . »
- Une “pépité” :
 - signe précoce de la détection de mammites : environ 48h avant les signes classiques connus, grumeaux dans le lait + inflammation
- **Jeunes bovins** : détection plus précoce et meilleure (16%) pour le “bovine respiratory disease” que la détection visuelle par l'éleveur
“Improved detection of bovine respiratory disease in young bulls using rumen temperature boluses” E. Timsit et al. en soumission

Expertise ?

- S'appuyer sur des connaissances “générales”
 - Effet buvée
 - Rythme circadien
 - Effet de la digestion
 - Influence température externe (étable/ prairie)
 - Influence du stress
- Rapatrier l'expertise sur la température vaginale
 - Même modèle profond
 - Corrélation (après reconstruction) de 0.79
- Bibliographie : très hétérogène, sur quelques individus

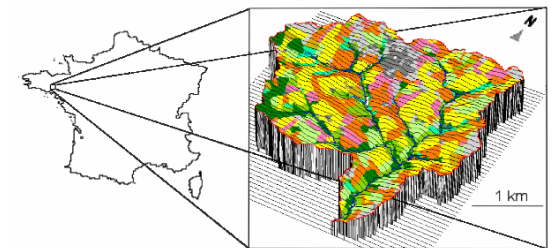
Morale

- Nouveaux capteurs :
 - Rapatrier l'expertise : même modèle profond
- Travail important de prétraitement :
 - Aller-retours entre experts, chercheurs, ingénieurs
 - Modifications des seuils, des références, des descripteurs/indicateurs
 - Et même retour sur le capteur (5 générations de bolus)
 - Nécessitant de l'expertise ...
 - Très dépendant de ce que l'on recherche!
- Existence d'une plate-forme
- Doit-on dire :
 - “Sans expertise : pas de résultat, avec expertise : peu de surprise” ?

Sacadeau (1999-2009)

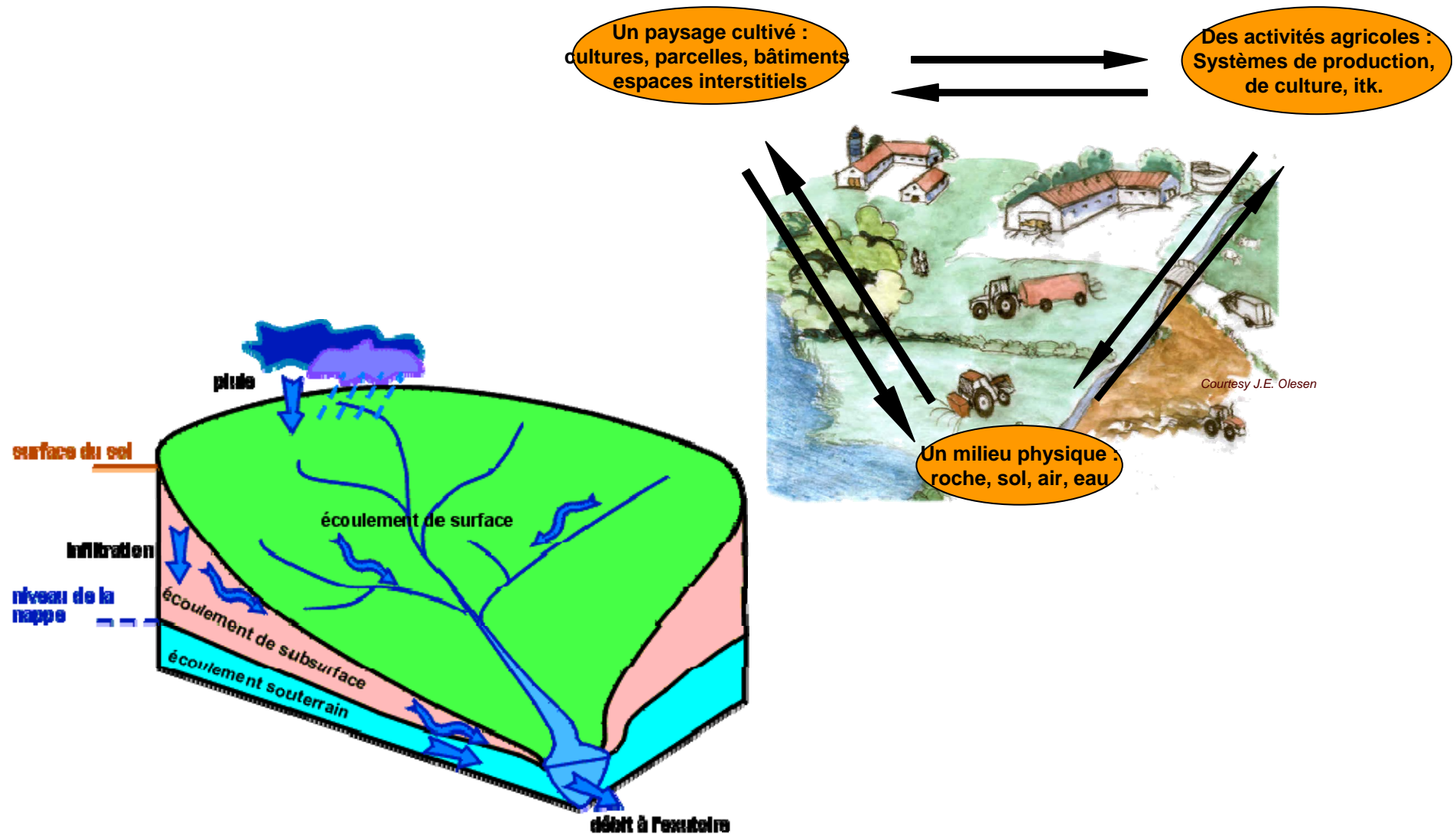
Systeme d'Acquisition de Connaissances pour l'Aide à la Decision sur la qualite de l'EAU

- **Partenaires :**
 - V. Masson, R. Trepos, M.-O.Cordier, Dream/IRISA
 - P.Aurousseau, Ch. Gascuel, Fl. Tortrat INRA-ENSAR / UMR SAS
 - Fr. Garcia, Inra/ BIA, Toulouse
 - M.Falchier, CA 35, D. Heddadj, L. Lebouille CA 56
- **Cadre :** AIP « Aide à la décision – Comment articuler connaissances et actions en agriculture, agroalimentaire et dans l'espace rural »
- **Objectif :** Modélisation du transfert de pesticides dans un bassin versant en vue de la construction d'un outil d'aide à la décision pour la maîtrise de la qualité des eaux
- **Site d'application :** Bassin versant du Frémur



Problématique: émissions de flux de polluants dans les cours d'eau

Bassins versants de taille moyenne (10-50 km²)



Projet SACADEAU

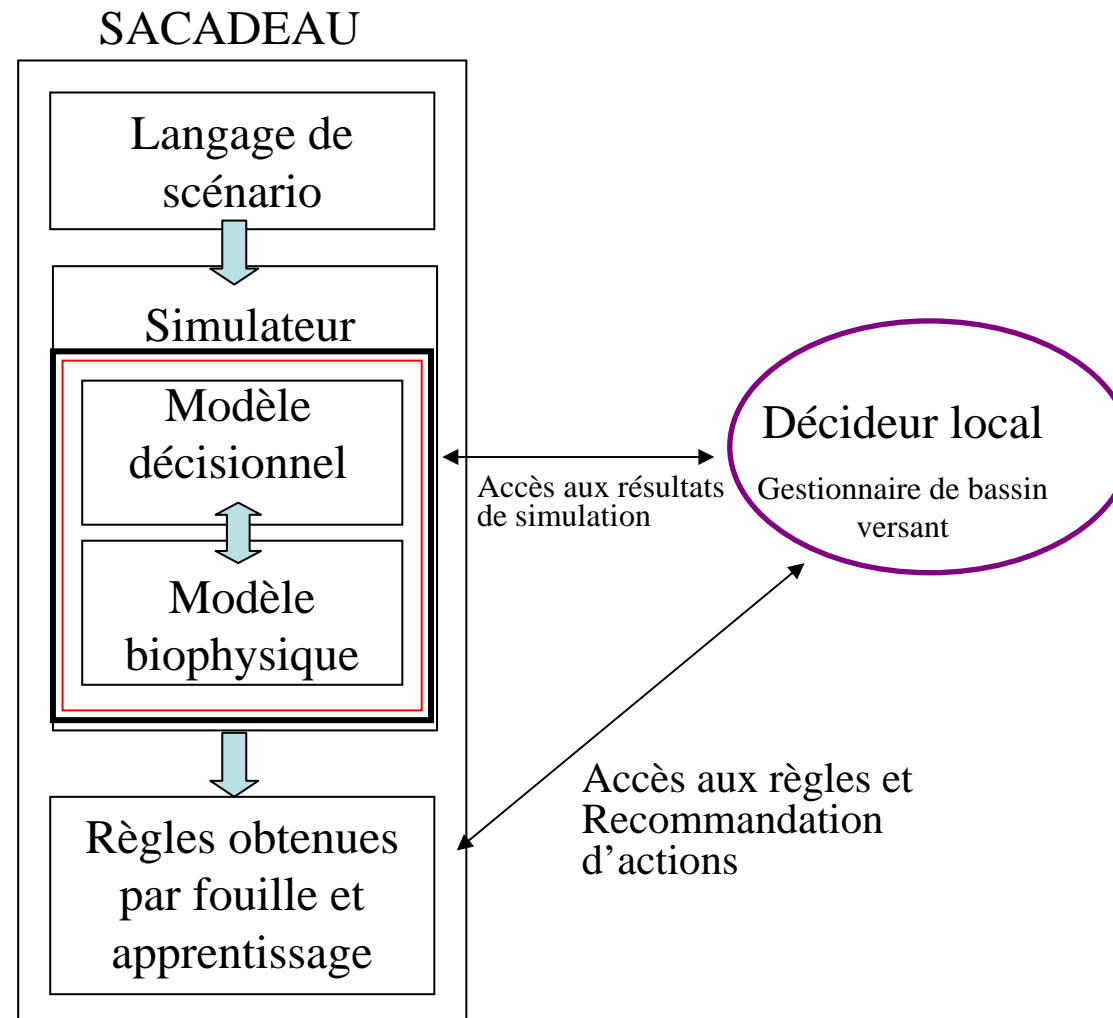
Objectifs : Construire un outil d'aide au conseil sur un bassin versant

- Simuler le transfert des herbicides à travers un bassin versant jusqu'à l'exutoire

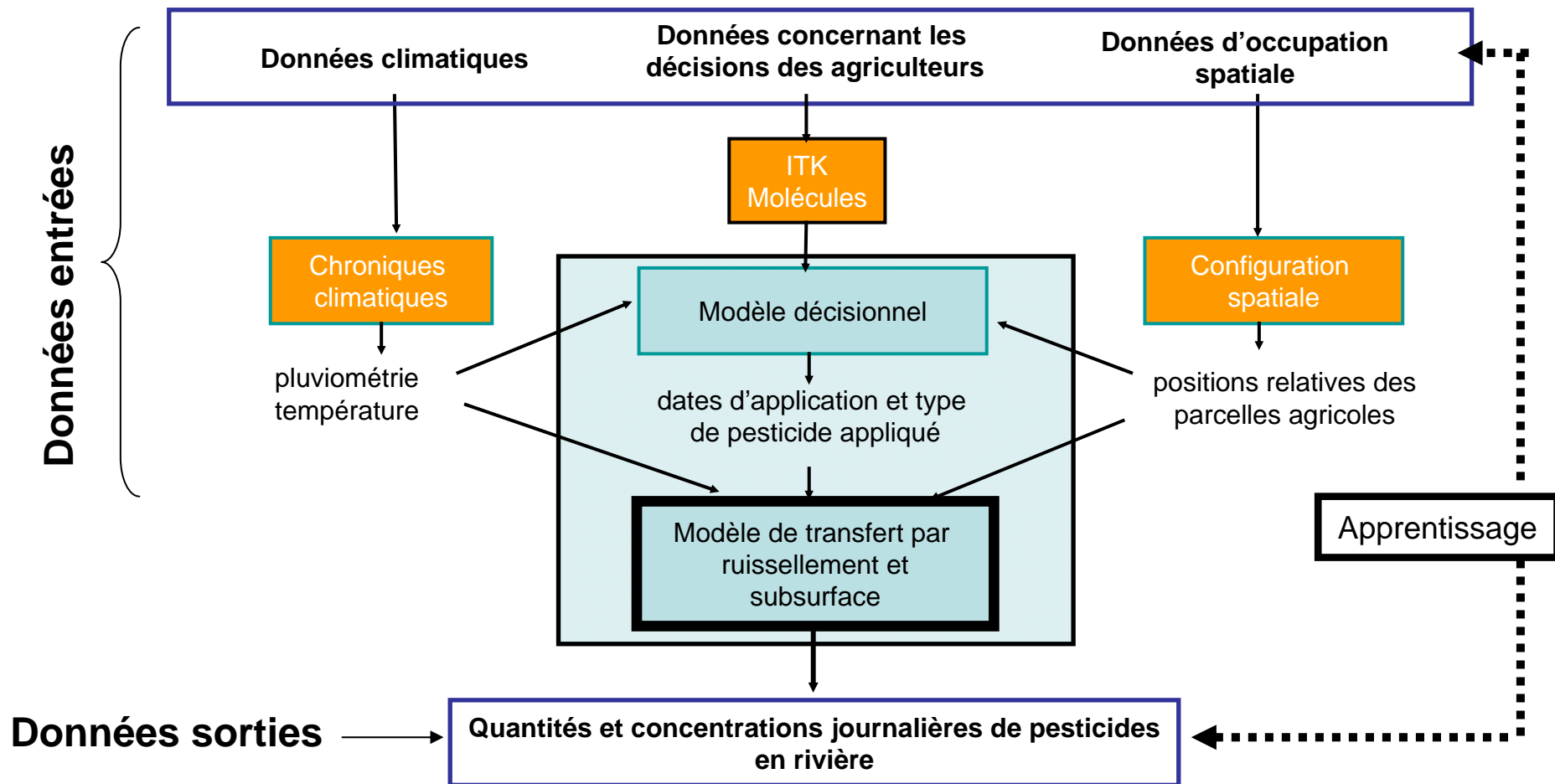
selon le climat de l'année, les stratégies de désherbage, les stratégies d'aménagement (bandes enherbées, haies-talus, fossés), le paysage (topographie, situation des bois, prairies, ...), le paysage cultivé (topographie, arrangement spatial des cultures, ...)

- Analyser les résultats de simulation par des outils de fouille ou d'apprentissage dans le but de dégager des connaissances et des règles de gestion du bassin

Architecture de SACADEAU



Données : entrées-sorties du simulateur/modèle



Données (pas très présentables...)

- #parc data/parcelles_os/os9.txt
- jour id_molecule Qpestriv(en g) Qpestruis(en g) Conc_Exutoire(micro g/l) pluie(en mm)
- 115 12 0.016495 0.019504 0.002565 0.5
- 115 14 0.305884 0.208042 0.036615 0.5
- 116 12 0.011947 0.000000 0.000880 0.0
- 116 14 0.216190 0.000000 0.015928 0.0
- 117 12 0.179061 1.784572 0.128596 6.0
- 117 14 2.567719 14.524073 1.119324 6.0
- 118 12 0.137489 0.000000 0.010246 0.0
- 118 14 1.920554 0.000000 0.143123 0.0
- 119 12 0.215443 1.359740 0.100123 5.5
- 119 14 2.208188 6.010238 0.522387 5.5
- 120 12 0.191183 0.503401 0.038163 2.5
- 120 14 1.704163 1.425853 0.171976 2.5
- 121 12 0.164802 0.120945 0.016690 1.0
- 121 14 1.361941 0.266150 0.095095 1.0
- 122 12 0.137133 0.000000 0.009770 0.0
- 122 14 1.108095 0.000000 0.078948 0.0
- 123 12 0.106024 0.000000 0.008183 0.0
- 123 14 0.837604 0.000000 0.064649 0.0
- 124 12 0.081359 0.000000 0.006677 0.0
- 124 14 0.627507 0.000000 0.051499 0.0
- 125 12 0.058406 0.000000 0.004455 0.0
- 125 14 0.440961 0.000000 0.033634 0.0
- 126 12 0.046387 0.000000 0.003807 0.0
- 126 14 0.341256 0.000000 0.028006 0.0
- 127 12 0.036210 0.000000 0.003089 0.0
- 127 14 0.259672 0.000000 0.022152 0.0
- 128 12 0.028026 0.000000 0.002423 0.0
- 128 14 0.196075 0.000000 0.016950 0.0
- 129 12 0.022227 0.000000 0.002001 0.0
- 129 14 0.151680 0.000000 0.013658 0.0
- 130 12 0.086052 0.320896 0.030327 2.5
- 130 14 1.342308 0.190688 0.114242 2.5
- 130 1 1.009068 0.133026 0.085111 2.5
- 130 4 0.040404 0.035243 0.005637 2.5
- 130 13 0.040097 0.036330 0.005696 2.5
- 131 12 0.077096 0.021539 0.007994 0.5

exu:94,209-numexu:1-parcelle:0-surf:0.000000-mrt:-0.000671-be:0.000000-eau:0.000000,127817400.000000-pest[12]:0.000000,144.393585-pest[14]:0.000000,271.533020-pest[1]:0.000000,1129.599609-pest[4]:0.000000,99.264465-pest[13]:0.000000,56.100643-pest[0]:0.000000,328.667816-pest[3]:0.000000,3.702457-pest[6]:0.000000,106.666252-pest[7]:0.000000,1.583342-pest[11]:0.000000,0.086176-pest[8]:0.000000,2.727302

files :

exu:106,212-numexu:5023-parcelle:1564-surf:0.080000-mrt:9.000000-be:0.000000-eau:172000.000000,6400.000000

files :

exu null

frere:

exu:106,211-numexu:5022-parcelle:3019-surf:0.040000-mrt:4.750000-be:0.000000-eau:89200.000000,0.000000

files :

exu null

frere:

exu:107,211-numexu:5020-parcelle:566-surf:0.040000-mrt:6.250000-be:0.000000-eau:136800.000000,82000.000000

files :

exu:107,212-numexu:5021-parcelle:1564-surf:0.080000-mrt:13.000000-be:0.000000-eau:48800.000000,129600.000000

files :

exu null

frere:

exu null

frere:

exu:107,210-numexu:5018-parcelle:3019-surf:0.040000-mrt:7.250000-be:0.000000-eau:82000.000000,89200.000000

Prétraitement : quelle représentation?

- 1^{ère} expérience : quelle est l'influence du climat et des pratiques agricoles sur la contamination des eaux?
 - comment représenter les résultats de simulations, et donc des entrées et des sorties ?

Prétraitement : quelle représentation?

- 1^{ère} expérience : influence du climat ...

- Entrées :

- Comment décrire la variable explicative : le climat (vis à vis du ruissellement) : typologie à 3 classes

Période mai-juillet	Classe 1	Classe 2	Classe 3
critère 1	< 100 mm	100 – 150 mm	> 150 mm
critère 2	< 3	3 – 6	> 6

Cumul des pluies mai-juillet

Fréquence des pluies > 10mm

- Mais aussi quelles variables pour décrire le contexte : taille des exploitations, type de désherbage, pente des parcelles, molécules herbicides utilisées ...

2 types de sol, 2 stratégies de désherbage, 3 formes de bassin versant.

- Sorties :

- Comment décrire la variable d'intérêt : les chroniques temporelles ?

Prétraitement : quelle représentation? (2)

- 5 classes d'acceptabilité de la teneur en pesticides d'après
 - le nombre et la valeur de concentrations par averse
 - somme des concentrations sur une période de 3 mois (mai-juillet)

Conc. du pic (µg/l)	Nbre de pics	Somme des concentrations (µg/l)					
		0	0 – 0,01	0,01 – 0,1	0,1 – 0,5	0,5 – 2	> 2
0,1 – 0,5	0 - 2		1	2	3	4	5
	3 - 5				4	4	5
	> 5					5	5
> 0,5	> 3					5	5

- Utilisation de la PLI (données étiquetées) + Classification hiérarchique

Connaissances apprises

- 100% des règles proposent, pour expliquer la classe de contamination 0, un climat plutôt sec et la présence d'un sol de forte teneur en matière organique.
- 74% des règles proposent, pour expliquer la classe de contamination 1, la présence d'une bande enherbée sur la quasi totalité du bord du cours d'eau. 74% de règles proposent, pour cette même classe, l'utilisation de molécules nouvelles.
- 59% des règles proposent, pour expliquer la classe de contamination 2, la présence d'une bande enherbée sur quasi totalité du bord du cours d'eau.
- 67% des règles proposent, pour expliquer la classe de contamination 3, la présence d'un sol à faible teneur en matière organique.

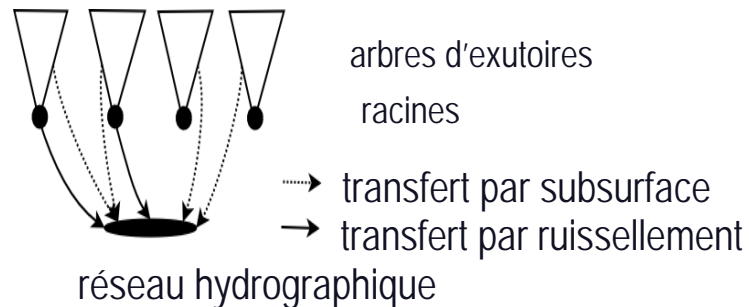
- *Impact de la matière organique dans le sol :*
 - *climat=3 et mat_organ_sol =2% et bande_enherbée =90% et molécules =nouvelles implique classe2*
 - *climat=3 et mat_organ_sol =4-5% et bande_enherbée =90% et molécules=nouvelles implique classe1*
- *Impact de la mise en place de bandes enherbées :*
 - *climat =4 et mat_organ_sol =4-5% et bande_enherbée =0% et molécules =atrazine implique classe 3*
 - *climat =4 et mat_organ_sol =4-5% et bande_enherbée =90% et molécules =atrazine implique classe2*

Prétraitement : quelle représentation?

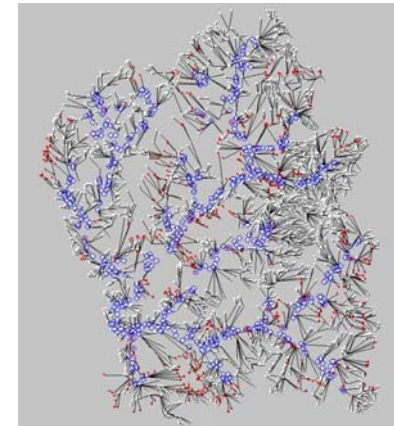
- 2nde expérience : quelle est l'influence de la répartition spatiale des parcelles maïs ?
 - comment représenter les résultats de simulations, et donc des entrées et des sorties ?

Prétraitement : quelle représentation?

- 2^{nde} expérience : influence du spatial...
 - Entrées :
 - Moyenne de neufs climats
 - Comment tenir compte de l'aspect spatial?



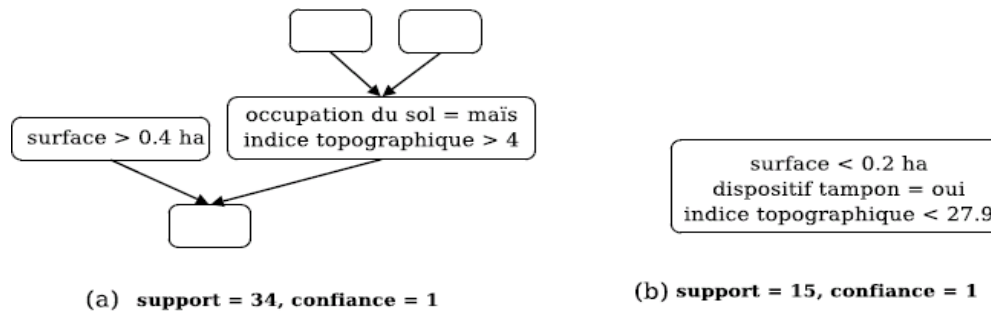
Arbre d'exutoire



- Sorties : taux de transfert d'un arbre d'exutoire, en trois classes fort, moyen, faible

Prétraitement : quelle représentation? (bis)

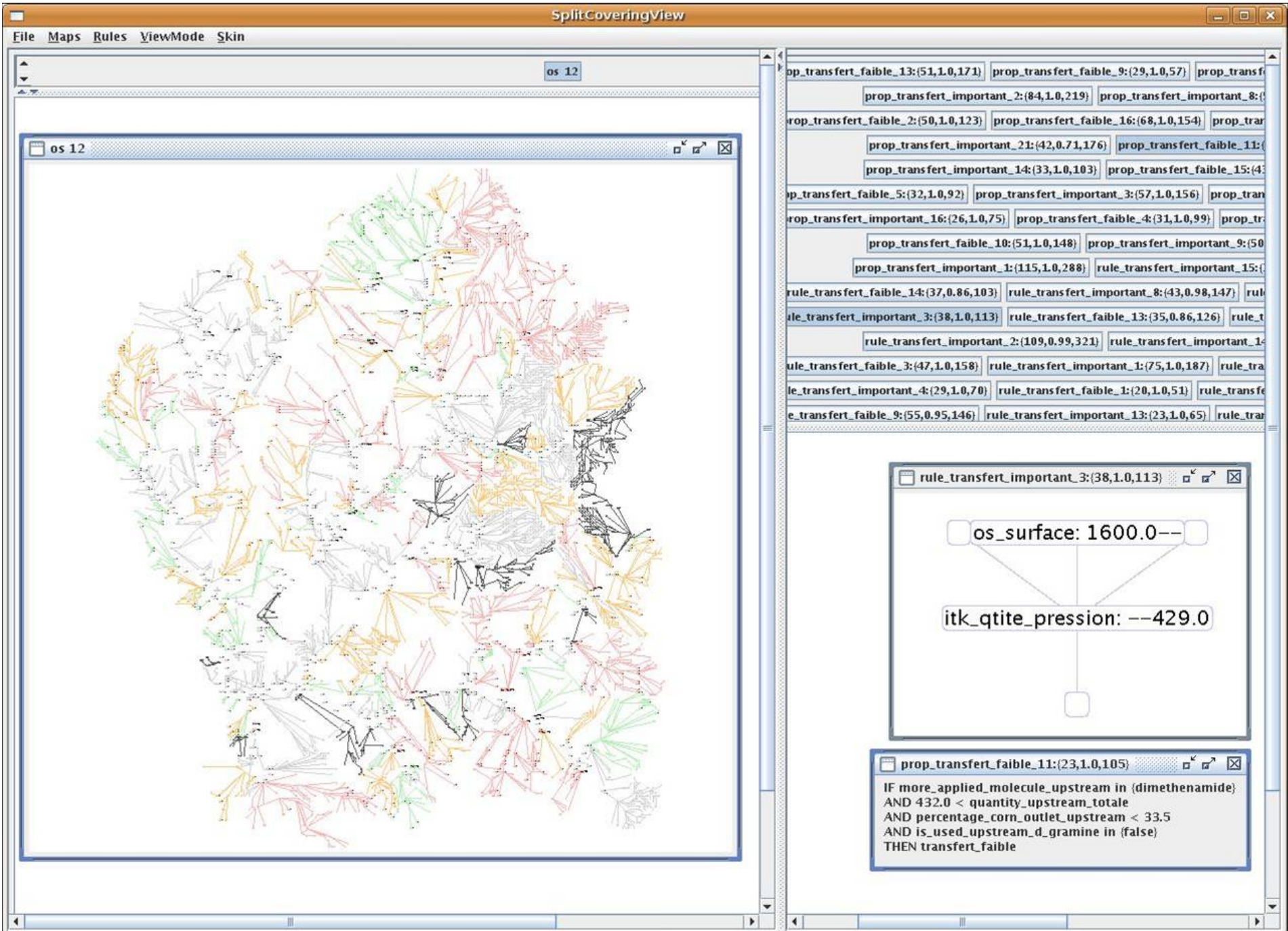
- Apprentissage PLI sur les arbres d'exutoire avec deux classes : fort et faible => motifs d'arbres



- Représentation des arbres par des attributs agrégats puis CN2 : règles attribut-valeur

SI la surface totale de l'arbre < 7,36 ha
ET surface de l'exutoire maïs le plus grand > 1,1 ha
ALORS classe = **transfert_important**

support=32,confiance=1



Morale

- Importance de la représentation des données pour permettre l'apparition des "pépites"
 - Nécessite de l'expertise
 - Importance de la question d'intérêt
- Processus itératif
 - Mise au point : Aller-retours entre experts et chercheurs, et même retour sur les fournisseurs de données (capteurs, modèles)
 - Interaction : l'utilisateur/décideur oriente le processus par la question posée, influe sur le processus par sa réaction aux résultats produits
- Les outils "suivent"
 - requête, fouille, apprentissage ...

Impact de la question posée ...

Sur l'acquisition des connaissances,

mais aussi sur l'acquisition des données

Simulateur : fournisseur de données

- Bien adapté pour l'aide à la décision interactive
- Pose une nouvelle problématique :
 - Comment générer par simulation les données “les meilleures” pour en extraire les connaissances répondant au mieux à une question ?
- notion de base “représentative” :
 - Variables d'impact : contraintes par la question
 - Ex : climat pluvieux, réglementation de distance à la rivière
 - Variables contexte : doivent représenter au mieux toutes les possibilités de l'objet de référence, implicite dans la question
 - Ex: Bassin versant de référence, pratiques habituelles de la région
 - Variable d'intérêt : contraintes par la question (sorties de simulation)

Démarche et résultats sur un exemple

– La question:

- « Quel est l'impact sur le ruissellement des pesticides d'une interdiction de planter du maïs sur certaines parcelles critiques vis-à-vis du chemin de l'eau ? »

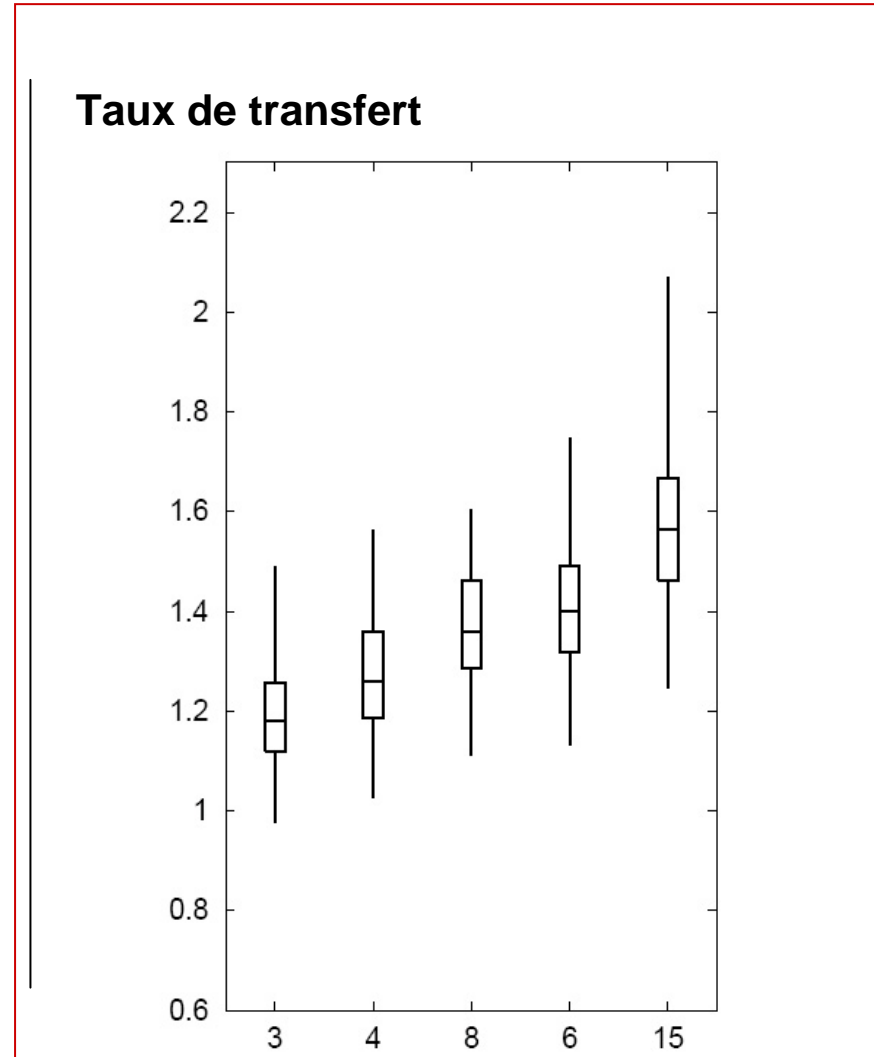
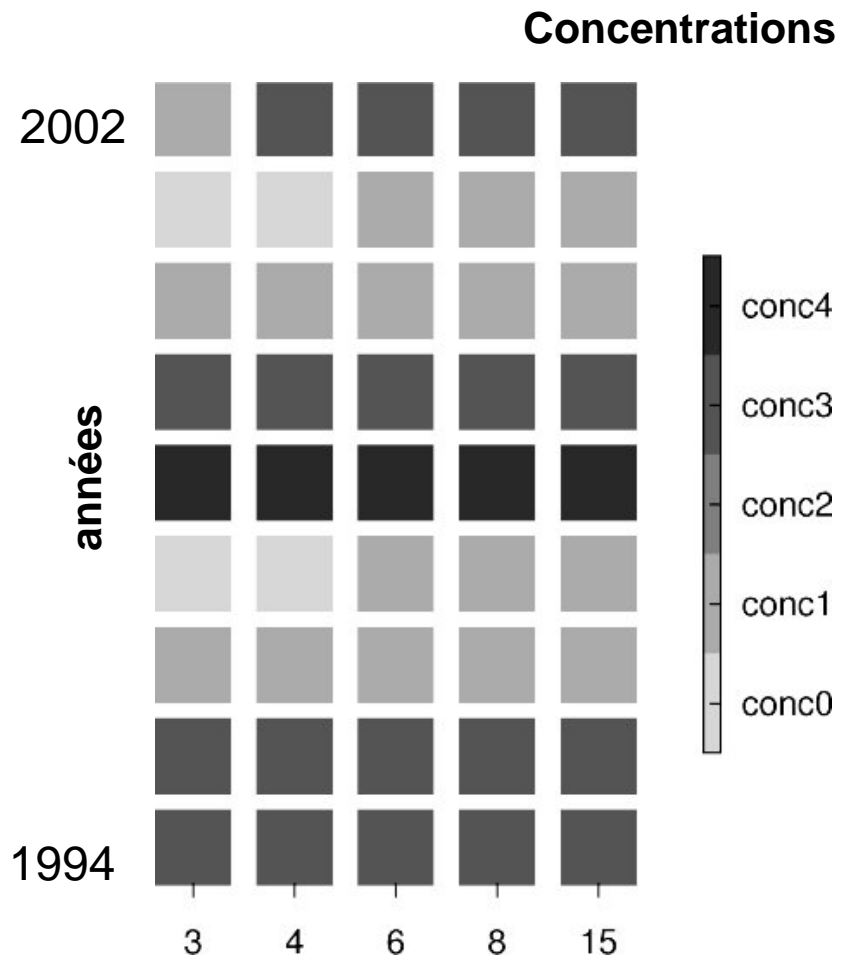
– La traduction:

- Parcelles maïs critiques vis-à-vis du chemin de l'eau
=> parcelle ayant un exutoire ayant plus de k fils dans l'arbre d'exutoires avec $k = 3, 4, 6, 8, 15$ (correspond à 25, 14, 6, 2, 0 % des parcelles maïsables)

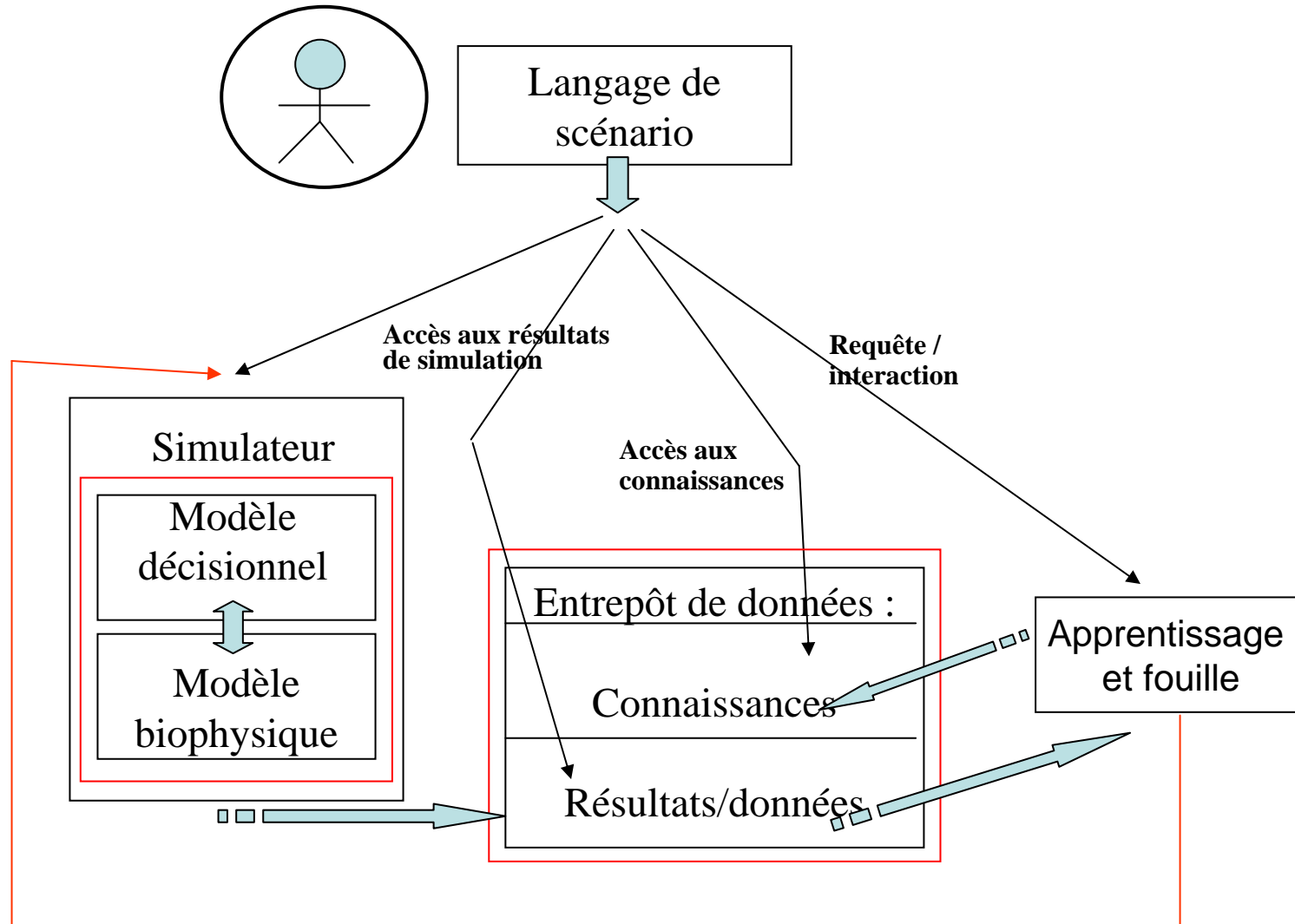
– La simulation:

- Contrainte supplémentaire lors de l'affectation des parcelles maïs (à surface totale constante)

Démarche et résultats sur un exemple



Fouille de données / aide à la décision



Conclusion : acquérir des connaissances dans un contexte d'aide à la décision

- Importance de la phase de prétraitement
 - Peut-on faire sans experts/expertise?
 - Peut-on faire sans savoir ce que l'on cherche ?
- Importante de l'interaction avec le décideur
 - Stockage incrémental des données
 - Apprentissage/fouille incrémental
 - Bases de données inductives
 - Plate-forme de fouille de données (A. Vautier)

Références

- Monitoring cardiaque :
 - G. Carrault, M.-O. Cordier, R. Quiniou, and F. Wang. Temporal abstraction and inductive logic programming for arrhythmia recognition from electrocardiograms. *Artificial Intelligence in Medicine*, 28:231–263, 2003.
 - L. Callens, G. Carrault, M.-O. Cordier, E. Fromont, P. Mabo, F. Portet, and R. Quiniou. Intelligent adaptive monitoring for cardiac surveillance. In Proceedings of *ECAI'08 (18th European Conference on Artificial Intelligence)*, pages 653–657, 2008..
- Projet Sacadeau :
 - R. Trepos, A. Salleb, M.-O. Cordier, V. Masson, and Ch. Gascuel. A distance based approach for action recommendation. In Proceedings of the *European Conference on Machine Learning (ECML'05)*, LNCS 3720, pages 425-436, 2005
 - Ch. Gascuel-Odoux, P. Arousseau, M.-O. Cordier, P. Durand, F. Garcia, V. Masson, J. Salmon-Monviola, F. Tortrat, and R. Trepos. A decision-oriented model to evaluate the effect of land use and agricultural management on herbicide contamination in stream water. *Environmental Modelling and Software*, 24 (2009) 1433–1446, 2009.
- A. Vautier, M.-O. Cordier and R. Quiniou. Towards data mining without information on knowledge structure. In Proceedings of *PKDD'07 (Principles and Practice of Knowledge Discovery in Databases)*, LNAI 4702, pages 300–311, 2007.

Références

- E. Fromont, R. Quiniou, and M.-O. Cordier. Learning rules from multisource data for cardiac monitoring. *International Journal of Biomedical Engineering and Technology (IJBET)*, 2010.
- R. Quiniou, M.-O. Cordier, G. Carrault, and F. Wang. Application of ILP to cardiac arrhythmia characterization for chronicle recognition. In C. Rouveirol and M. Sebag, editors, *ILP'2001*, volume 2157 of *LNAI*, pages 220–227. Springer-Verlag, 2001.
- E. Fromont, M.-O. Cordier and R. Quiniou. Learning from multi source data. Proceedings of *Knowledge Discovery in Databases (PKDD'04)*, *Lecture Notes in Artificial Intelligence 3202*, pages 503–505, Springer, 2004.
- A. Vautier, M.-O. Cordier, and R. Quiniou. An inductive database for mining temporal patterns in event sequences. In Proceedings of the *International Joint Conference on Artificial Intelligence (IJCAI'05)*, pages 1640–1641
- F. Portet, R. Quiniou, M.-O. Cordier and G. Carrault. Learning decision tree for selecting QRS detectors for cardiac monitoring. In Proceedings of *11th Conference on Artificial Intelligence in Medicine (AIME'07)*, *Lecture Notes in Computer Science 4594*, pages 170–174. Springer, 2007.
- M.-O. Cordier, Frederick Garcia, Ch. Gascuel, V. Masson, Jordy Salmon-Monviola, Florent Tortrat, and R. Trepos. A machine learning approach for evaluating the impact of land use and management practices on streamwater pollution by pesticides. In Proceedings of the *International Congress on Modelling and Simulation (MODSIM'05)*, 2005.
- C. Largouet, M.-O. Cordier, and G. Fontenelle. Scenario templates to analyse qualitative ecosystem models. *18th World IMACS Congress and MODSIM09 (International Congress on Modelling and Simulation)*, pages 2129–2135, 2009.