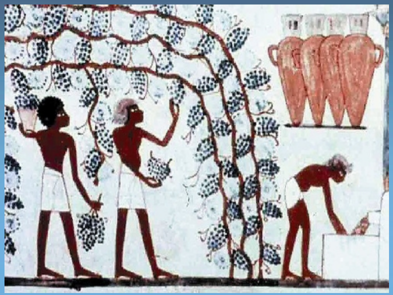
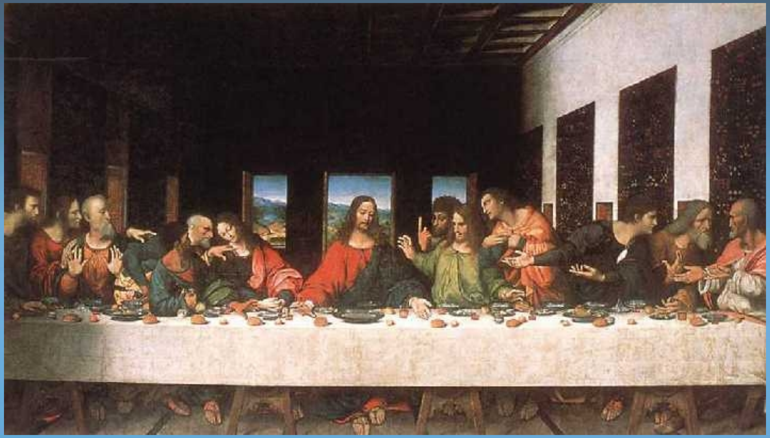
Reconciling competing models: a case study of wine fermentation kinetics

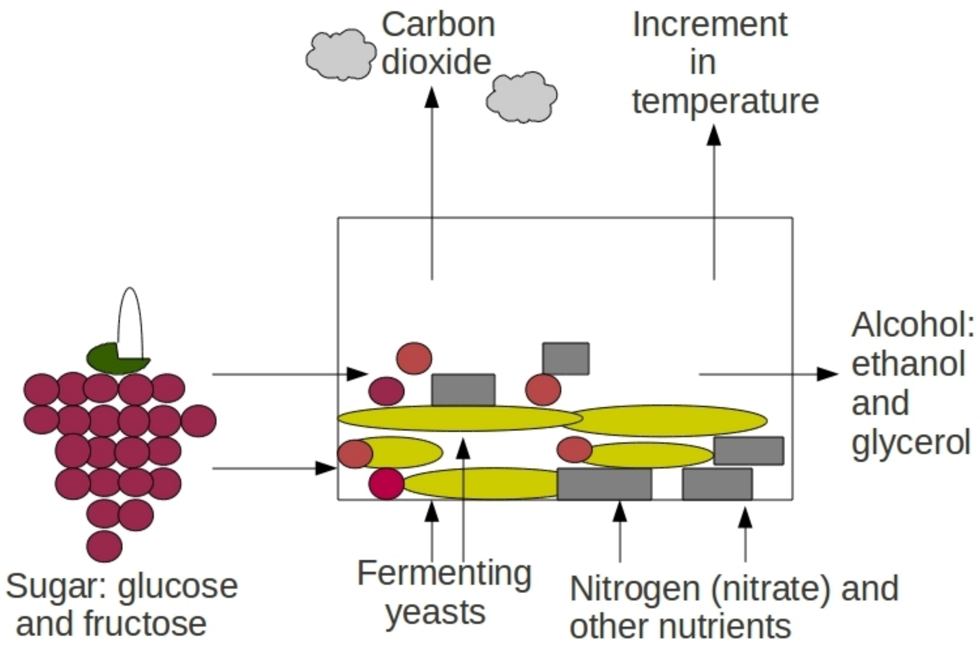
Rodrigo Assar
INRIA Bordeaux project-team MAGNOME
joint with Université Bordeaux and CNRS
Rodrigo.Assar@inria.fr
Felipe A. Vargas, David J. Sherman

ANB 2010: RISC, Castle of Hagenberg, Austria, July 31-August 2





Wine fermentation



$$\frac{dX}{dt} = \mu X_A$$

$$\frac{dX_A}{dt} = \mu X_A - k_d X_A$$

$$\frac{dN}{dt} = -\frac{\mu X_A}{Y_{X/N}}$$

$$\frac{dE}{dt} = \beta X_A$$

Need : more general model that better adapts to different conditions

Need : general method for this kind of reconciliation

Talk Structure

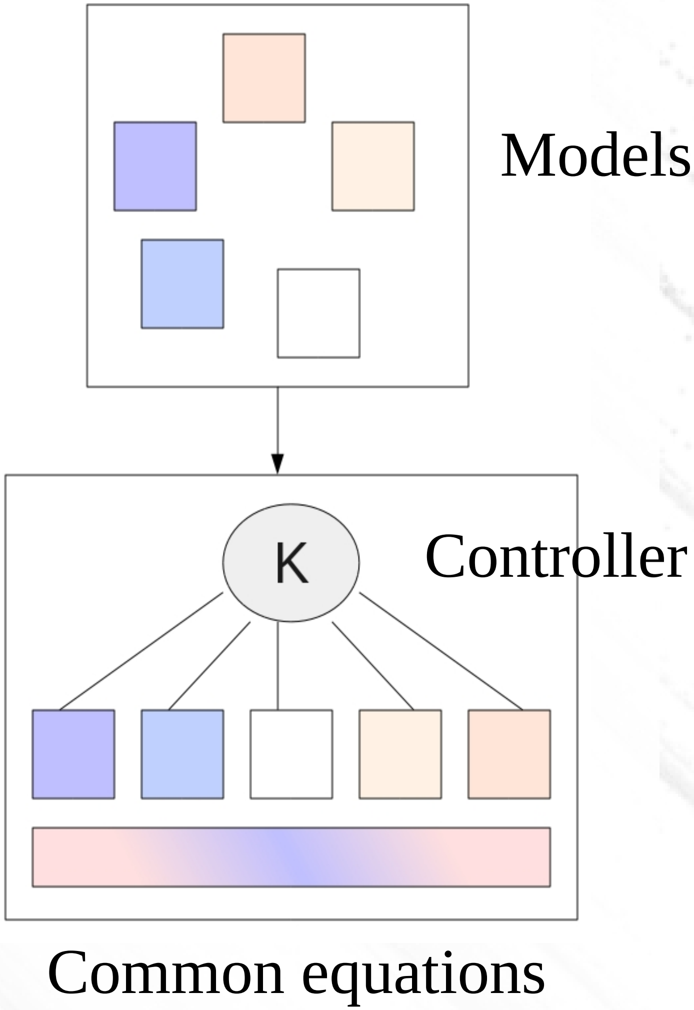
- Introduction
  - Motivation and Goal
- Methods
  - System
  - Analyzed models
  - Steps of the approach
- Results
  - Statistical results
  - Constructive step
  - Analysis
- Summary and Perspectives

Introduction: Motivation

- Wine industry: 5 mill. of tonnes/ year in France
- Losses from fermentation problems (stuck and sluggish): 7 billion of euros/year
- Many different models to explain the wine fermentation process
- Not good results on not training data

Goal

- **Input:** models defined by equations and methods validated on specific training data
- **Goal:** reconciled and generalized model with good predictions of reality, recovering biological behaviors
- **Tool:** combined model selecting best equations for environmental conditions



System

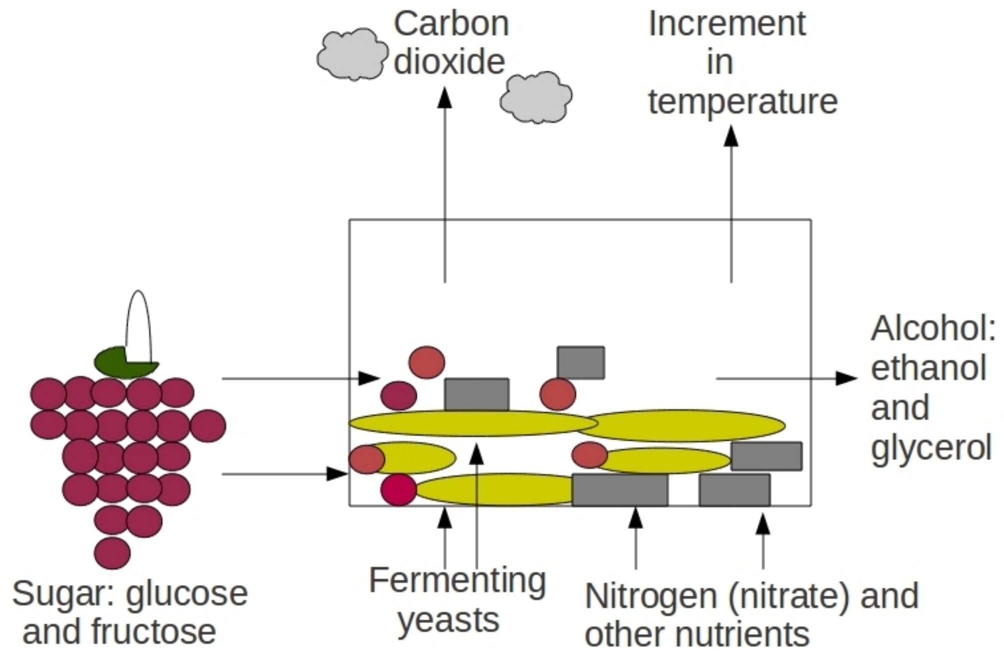
Agent: yeast

Resources: sugar and nitrogen

Other environmental factors: nutrient, pH, yeast populations and flora

Product: alcohol

By-products: carbon dioxide gas, heat



Analyzed models

Coleman et al., 2007

$$\frac{dX}{dt} = \mu X_A$$

$$\frac{dX_A}{dt} = \mu X_A - k_d X_A$$

$$\frac{dN}{dt} = -\frac{\mu X_A}{Y_{X/N}}$$

$$\frac{dE}{dt} = \beta X_A$$

Scaglia et al., 2009

$$\frac{dX}{dt} = \frac{e^{-(CO_2 - CO_{2,95})}}{e^{(CO_2 - CO_{2.95})} + e^{-(CO_2 - CO_{2.95})}} A \mu_m \frac{S}{S + Ks\, B^a} X \left(1 - \frac{X}{A\mu_m \frac{S}{(S+Ks\ B^a)\beta}}\right) + \left[1 - \frac{e^{-(CO_2 - CO_{2.95})}}{e^{(CO_2 - CO_{2.95})} + e^{-(CO_2 - CO_{2,95})}}\right] \left(CX \frac{dS}{dt} - DX\right)$$

$$\frac{dS}{dt} = \frac{1}{Y_{X/S}} \left[-X\left(\mu_m \frac{S}{S + Ks\, B^b} - EX\right)\right] - FX$$

$$\frac{dCO_2}{dt} = G\mu_m \frac{S}{S + Ks\, B^c} X + \frac{d}{dt}\left(H\ \mu_m \frac{S^2}{(S + Ks\, B^d)(S + Ks\, B^e)} X + I\, X\right)$$

$$\frac{dP}{dt} = \frac{1}{Y_{CO_2/P}} \frac{dCO_2}{dt}$$

Pizarro et al., 2007

$$\begin{aligned} &\text{Min}\{-v_j\} \\ &\text{subject to} \\ &T \cdot v = 0 \\ &v_i^{LB} \leq v_i \leq v_i^{UB} \end{aligned}$$

feedback

$$\textit{Glucose}: \frac{d[Glu]}{dt} = -v_{glu} \cdot Xv$$

$$\textit{Ammonia}: \frac{d[NH_4]}{dt} = -v_{NH4} \cdot Xv$$

$$\textit{Viable Cells}: \frac{d[Xv]}{dt} = \mu \cdot Xv$$

$$\textit{Ethanol}: \frac{d[EtOH]}{dt} = v_{EtOH} \cdot Xv$$

$$\textit{Glycerol}: \frac{d[Gly]}{dt} = v_{gly} \cdot Xv$$

Steps of our approach

- **Symbolic step:** obtaining homogenous form, separating linear and secondary effects

- **Statistical step:** classifying the simulation results according to adjustment of experimental results

- **Constructive step:** building the combined model

Identifying effects

Effects represented by coefficient equations:

μ_A, ε_A, σ_A: linear effect of X_A on X, EtOH and S rate.

$\mu^{(1)}$, $\varepsilon^{(1)}$, $\sigma^{(1)}$: linear effect of X on X, EtOH and S.

ε_{CO2}: linear effect of dCO2/dt on EtOH.

$\mu^{(2)}$ and $\sigma^{(2)}$: quadratic effect of X on S.

Rewriting models

Coleman et al., 2007

$$\frac{dX}{dt} = \mu \cdot X_A$$
$$\frac{dX_A}{dt} = (\mu - \tau) \cdot X_A$$
$$\frac{dN}{dt} = -\nu_N \cdot X_A$$
$$\frac{d[EtOH]}{dt} = \nu_{EtOH} \cdot X_A$$
$$\frac{dS}{dt} = -\nu_S \cdot X_A$$

Scaglia et al., 2009

$$\frac{dX}{dt} = (F_\mu \cdot \mu - F_\tau \cdot \tau) \cdot X - F_\mu \cdot \beta \cdot X^2$$

$(F_\mu \cdot \mu - F_\tau \cdot \tau)$ ← $\mu^{(1)}$; $F_\mu \cdot \beta \cdot X$ ← $\mu^{(2)}$

$$\frac{d[EtOH]}{dt} = \frac{1}{Y_{CO_2/EtOH}} \cdot \frac{dCO_2}{dt}$$
$$\frac{dS}{dt} = -((\nu_S + \nu_{S_0}) \cdot X - \frac{0.00002}{Y_{X/S}} \cdot X^2)$$
$$\frac{dCO_2}{dt} = \nu_{CO_2} \cdot X + \frac{d(C02Form)}{dt}$$

Pizarro et al., 2007

$$\frac{dX}{dt} = \mu \cdot X$$

$\mu \cdot X$ ← $\mu^{(1)}$

$$\frac{dN}{dt} = -\nu_N \cdot X$$
$$\frac{d[EtOH]}{dt} = \nu_{EtOH} \cdot X$$
$$\frac{dS}{dt} = -\nu_S \cdot X$$
$$\frac{d[Gly]}{dt} = \nu_{Gly} \cdot X$$

Experimental data

Temperature	Sugar	Nitrogen	Biomass	Ethanol	Sugar
L (0-19 °C)	M (160-240 g/l)	M (50-240 mg/l)			
		H (240-551 mg/l)			1
	H (240-308 g/l)	M			
		H	2		3
M (20-27 °C)	M	M	1	1	1
		H	1	1	6 1
	H	M	1		
		H	1		2
H (28-35 °C)	M	M	1	1	1
		H	2	1	2
	H	M			
		H			1

Pizarro et al., 2007 Malherbe et al., 2004 M-F et al., 2007

- We identified **configuration levels L:** Low, **M:** Moderate, **H:** High
- **Statistical information:** data sets per configuration, standard deviation
- **Biological information:** fermentations of *S. cerevisiae*, different strains

Statistical results

	Local: p-value		Global : r^2	
	Transient	Stable	Transient	Stable
Coleman et al., 2007	0.13	0.15	0.93	0.98
Scaglia et al., 2009	0.14	0.27	0.98	0.98
Pizarro et al., 2007	0.06	0.02	0.99	0.97

Example: MMH configuration, sugar consumption

- **Local criterion:** p-value estimates the probability of error at rejecting that experimental and simulated results coincide, computed at each time point, we show the average
- **Global criterion:** linear correlation index measures the similarity between experimental and simulated log-profiles

Statistical conclusions

- There is no best model for all conditions
- Quality depends on configuration of factors: level of temperature, sugar and nitrogen; temporal phase
- For all the variables there exist models of good adjustment to experimental data

Combined model

$$\frac{dX}{dt} = \mu_A \cdot X_A + \mu^{(1)} \cdot X - \mu^{(2)} \cdot X^2$$

$$\frac{d[EtOH]}{dt} = \varepsilon_A \cdot X_A + \varepsilon^{(1)} \cdot X + \varepsilon_{CO_2} \cdot \frac{dCO_2}{dt}$$

$$\frac{dS}{dt} = -\left(\sigma_A \cdot X_A + \sigma^{(1)} \cdot X - \sigma^{(2)} \cdot X^2\right)$$

Canonical representation

Factors-dependent coefficients

		Configuration of initial conditions								
Coefficient	Meaning	LMH	LHH	MMM	MMH	MHM	MHH	HMM	HMH	HHH
$\mu_A = \frac{max(\mu) \cdot N}{K_N + N}$	linear effect of X_A	-	$\emptyset$	$\forall t$	$\emptyset$	$\emptyset$	$\emptyset$	$\forall t$	$t \leq 30$	-
$\mu^{(1)} = FBA, (F_\mu \cdot \mu + F_\tau \cdot \tau)$	linear effect of X	-	$t > 110$	$t \leq 96$	$\forall t$	$\forall t$	$\forall t$	$\emptyset$	$t > 30$	-
$\mu^{(2)} = F_\mu \cdot \beta$	quadratic effect of X	-	$\emptyset$	$\emptyset$	$t \leq 51$	$t \leq 27$	$\forall t$	$\emptyset$	$\emptyset$	-
$\varepsilon_A = \frac{max(\nu_{EtOH}) \cdot S}{K_S + S}$	linear effect of X_A	-	-	$t \leq 96$	$t \leq 51$	-	-	$t > 300$	$t \leq 30$	-
$\varepsilon^{(1)} = FBA$	linear effect of X	-	-	$\emptyset$	$\forall t$	-	-	$t \leq 300$	$\forall t$	-
$\varepsilon_{CO_2} = \frac{1}{Y_{CO_2/EtOH}}$	linear effect of $\frac{dCO_2}{dt}$	-	-	$\emptyset$	$t \leq 51$	-	-	$\emptyset$	$t \leq 30$	-
$\sigma_A = \frac{\nu_{EtOH}}{Y_{EtOH/S}}$	linear effect of X_A	$\emptyset$	$\emptyset$	$\forall t$	$\emptyset$	-	$t > 107$	$\emptyset$	$t \leq 30$	$\emptyset$
$\sigma^{(1)} = FBA, 0.008 + \frac{max(\mu) \cdot S}{Y_{X/S} \cdot (S + K_S \cdot 93.023^{1.508})}$	linear effect of X	$\forall t$	$\forall t$	$t \leq 96$	$\forall t$	-	$t \leq 107$	$\forall t$	$t > 30$	$\forall t$
$\sigma^{(2)} = \frac{2 \cdot 10^{-5}}{Y_{X/S}}$	quadratic effect of X	$\emptyset$	$\forall t$	$\emptyset$	$t \leq 51$	-	$t \leq 107$	$\emptyset$	$\emptyset$	$t > 103$

Initial condition-dependent equations

$$\frac{dX}{dt} = \mu_A \cdot X_A + \mu^{(1)} \cdot X - \mu^{(2)} \cdot X^2$$

$$\frac{d[EtOH]}{dt} = \varepsilon_A \cdot X_A + \varepsilon^{(1)} \cdot X + \varepsilon_{CO_2} \cdot \frac{dCO_2}{dt}$$

$$\frac{dS}{dt} = -\left(\sigma_A \cdot X_A + \sigma^{(1)} \cdot X - \sigma^{(2)} \cdot X^2\right)$$

Configuration?

MMM

$$\frac{dX}{dt} = \mu^{(1)} \cdot X - 1_{t\leq 51} \cdot \mu^{(2)} \cdot X^2$$

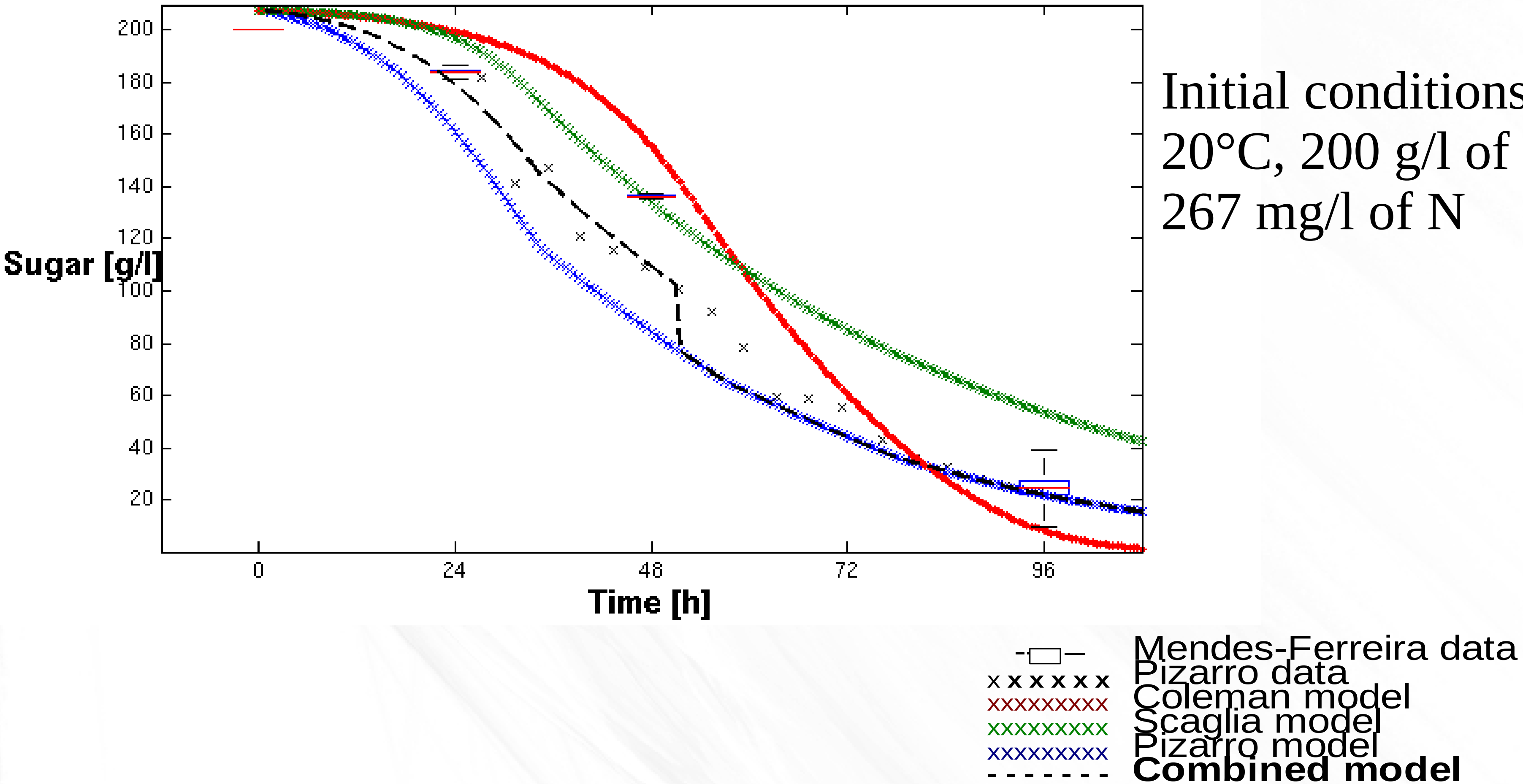
$$\frac{d[EtOH]}{dt} = 1_{t\leq 51} \cdot \varepsilon_A \cdot X_A + \varepsilon^{(1)} \cdot X + 1_{t\leq 51} \cdot \varepsilon_{CO_2} \cdot \frac{dCO_2}{dt}$$

$$\frac{dS}{dt} = -\left(\sigma^{(1)} \cdot X - 1_{t\leq 51} \cdot \sigma^{(2)} \cdot X^2\right)$$

HHH

$$\frac{dS}{dt} = -\left(\sigma^{(1)} \cdot X - 1_{t>103} \cdot \sigma^{(2)} \cdot X^2\right)$$

Example: sugar uptake in MMH



Initial conditions : 20°C, 200 g/l of S, 267 mg/l of N

Summary

- Combined model of wine fermentation kinetics
- Reconciling models steps:
  - Symbolic: obtention of homogenous form. Polynomial for ODE systems
  - Statistical: region-based analysis
  - Constructive:
    - Criterion to select best models
    - Combined model, coefficients depend of factors levels: initial configuration and temporal phase

Perspectives

- Automatically homogenizing and combining of models without revalidating
- Automatically finding regions
- Adding strain-specific effects
- Considering competing populations

Acknowledgements

David J. Sherman

Pascal Durrens

Elisabeth Bon

Nicolás Loira

Natalia Golenetskaya

Anasua Sarkar

Aurelie Goulielmakis

Tiphaine Martin

Alice Garcia

Eduardo Agosin

Biotechnology Laboratory of Pontificia Universidad Católica de Chile

Jean-Marie Sablayrolles

Equipe de Microbiologie et Technologies des Fermentations, INRA

Thanks for your attention!!!

			Coleman						Scaglia						Pizarro					
Temp.	**Sug.**	**Nit.**	Biomass		Ethanol		Sugar		Biomass		Ethanol		Sugar		Biomass		Ethanol		Sugar	
			Trans.	Stable	Trans.	Stable	Trans.	Stable	Trans.	Stable	Trans.	Stable	Trans.	Stable	Trans.	Stable	Trans.	Stable	Trans.	Stable
L	M	M																		
		H					Wrong	Little wrong					Little wrong	Little wrong					Good	Very good
	H	M																		
		H	Wrong	Wrong			Little wrong	Little wrong	Wrong	Wrong			Good	Very good	Wrong	Good			Good	Very good
M	M	M	Good	Little wrong	Good	Wrong	Good	Little wrong	Little wrong	Wrong	Wrong	Wrong	Little wrong	Wrong	Good	Wrong	Wrong	Wrong	Good	Wrong
		H	Little wrong	Wrong	Good	Wrong	Little wrong	Good	Good	Little wrong	Good	Little wrong	Good	Little wrong	Good	Very good	Good	Good	Good	Very good
	H	M	Little wrong	Little wrong					Very good	Wrong					Good	Good				
		H	Little wrong	Wrong			Little wrong	Good	Good	Very good			Good	Little wrong	Little wrong	Good			Little wrong	Little wrong
H	M	M	Very good	Very good	Little wrong	Good	Little wrong	Wrong	Little wrong	Wrong	Little wrong	Wrong	Good	Little wrong	Little wrong	Good	Good	Wrong	Very good	Very good
		H	Very good	Wrong	Very good	Little wrong	Very good	Wrong	Good	Wrong	Very good	Good	Good	Good	Good	Little wrong	Very good	Very good	Good	Very good
	H	M																		
		H					Wrong	Wrong					Little wrong	Good					Very good	Good

- Very good
- Good
- Little wrong
- Wrong

		Coleman model				Pizarro model				Scaglia model				Combined model			
		Transient		Stable		Transient		Stable		Transient		Stable		Transient		Stable	
Config.	Variable	C.1	C.2	C.1	C.2	C.1	C.2	C.1	C.2	C.1	C.2	C.1	C.2	C.1	C.2	C.1	C.2
MMM	*X*	0.006	1	0.006	0.92	0.109	1	0	1	0.012	0.96	0	0.77	0.033	1	0.006	0.92
	EtOH	0.001	1	0	0.97	0	1	0	1	0	1	0	0.93	0.001	1	0	0.97
	S	0.07	0.97	0.001	0.93	0.022	1	0	0.98	0.017	0.95	0	0.98	0.063	1	0.001	0.93
MMH	*X*	0.045	0.98	0.001	0.90	0.006	1	0.483	1	0.113	0.98	0.06	0.99	0.256	0.99	0.483	1
	EtOH	0.047	1	0	0.97	0.001	1	0.053	1	0.048	1	0.149	0.92	0.049	1	0.053	1
	S	0.129	0.93	0.152	0.98	0.140	0.98	0.27	0.98	0.064	0.99	0.015	0.97	0.129	0.99	0.27	0.98
HMM	*X*	0.388	0.99	0.476	0.97	0.102	0.83	0.027	1	0.082	0.99	0	0.87	0.388	0.99	0.476	0.97
	EtOH	0.049	0.95	0.155	0.97	0.129	0.95	0	1	0.08	0.92	0	-0.97	0.129	0.95	0.155	0.97
	S	0.171	0.79	0	-0.7	0.238	1	0.272	0.97	0.107	0.91	0.032	0.73	0.238	1	0.272	0.97
HMH	*X*	0.203	0.97	0	0.28	0.156	0.98	0.048	0.97	0.197	0.96	0	0.91	0.203	0.97	0.048	0.97
	EtOH	0.275	1	0.089	0.80	0.162	0.99	0.214	0.99	0.264	0.98	0.001	1	0.339	0.99	0.214	0.99
	S	0.327	1	0	0.59	0.135	0.98	0.167	0.99	0.197	0.95	0.001	1	0.327	1	0.167	0.99

Very good: local and global criterion very favorable: $p \geq 0.1$ and $corr \geq 0.98$.

Good: one criterion very favorable, other only favorable: $0.05 \leq p < 0.1$ or $0.95 \leq corr < 0.98$

Little wrong: one unfavorable ($p < 0.05$ or $corr < 0.95$) and other favorable or superior.

Wrong: both criteria are unfavorable.

The limit cases: If local criterion is absolutely unfavorable ($p = 0$) we qualified in Wrong, if local criterion is unfavorable (but not absolutely) and global criterion is optimum ($corr = 1$) we considered it Good.

		Configuration of initial conditions								
Coefficient	Meaning	LMH	LHH	MMM	MMH	MHM	MHH	HMM	HMH	HHH
$\mu_A = \frac{max(\mu)\cdot N}{K_N+N}$	linear effect of X_A	-	$\emptyset$	$\forall t$	$\emptyset$	$\emptyset$	$\emptyset$	$\forall t$	$t \leq 30$	-
$\mu^{(1)} = FBA$, $(F_\mu \cdot \mu + F_\tau \cdot \tau)$	linear effect of X	-	$t > 110$	$t \leq 96$	$\forall t$	$\forall t$	$\forall t$	$\emptyset$	$t > 30$	-
$\mu^{(2)} = F_\mu \cdot \beta$	quadratic effect of X	-	$\emptyset$	$\emptyset$	$t \leq 51$	$t \leq 27$	$\forall t$	$\emptyset$	$\emptyset$	-
$\varepsilon_A = \frac{max(v_{EtOH})\cdot S}{K_S+S}$	linear effect of X_A	-	-	$t \leq 96$	$t \leq 51$	-	-	$t > 300$	$t \leq 30$	-
$\varepsilon^{(1)} = FBA$	linear effect of X	-	-	$\emptyset$	$\forall t$	-	-	$t \leq 300$	$\forall t$	-
$\varepsilon_{CO_2} = \frac{1}{Y_{CO_2/EtOH}}$	linear effect of $\frac{dCO_2}{dt}$	-	-	$\emptyset$	$t \leq 51$	-	-	$\emptyset$	$t \leq 30$	-
$\sigma_A = \frac{v_{EtOH}}{Y_{EtOH/S}}$	linear effect of X_A	$\emptyset$	$\emptyset$	$\forall t$	$\emptyset$	-	$t > 107$	$\emptyset$	$t \leq 30$	$\emptyset$
$\sigma^{(1)} = FBA$, $0.008 + \frac{max(\mu)\cdot S}{Y_{X/S}\cdot(S+K_S\cdot 93.023^{1.508})}$	linear effect of X	$\forall t$	$\forall t$	$t \leq 96$	$\forall t$	-	$t \leq 107$	$\forall t$	$t > 30$	$\forall t$
$\sigma^{(2)} = \frac{2\cdot 10^{-5}}{Y_{X/S}}$	quadratic effect of X	$\emptyset$	$\forall t$	$\emptyset$	$t \leq 51$	-	$t \leq 107$	$\emptyset$	$\emptyset$	$t > 103$