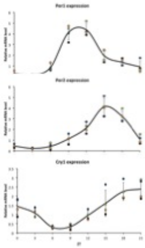
Use of a sparse non-linear model based on local kernels for the inference of regulatory networks

A.Fouchet, J-M. Delosme et F.d'Alché-Buc

Université d'Evry

November 30, 2012

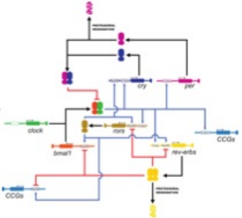
Motivation



Mathematical model of dynamics
(ODE, Markov models)

Estimation algorithm

Objective function
including constraints
(prior knowledge, qualitative constraints)



Choice of model (1/3)

- Described by differential equations
  ($\neq$ Boolean)
- Non-linear
  ($\neq$ LASSO, HMM, etc.)
- Non-parametric
  ($\neq$ S-system)
- Auto-regressive $\dot{\mathbf{x}}(t) = f(\mathbf{x}(t))$

Choice of model (2/3)

- Discrete observations with additive measurement noise $(\mathbf{y}(t) = \mathbf{x}(t) + n(t))_{t=0,1,\ldots,T}$
- Sparsity
  $\dot{x}_i(t)$ depending of a few $x_j(t)$
  - For LASSO
    $$\mathbf{x}(t+1) = A\mathbf{x}(t)$$
  - General additive model [1], used in each dimension
    $$x_i(t+1) = h_i(\mathbf{x}(t)) = \sum_{m=1}^{p} h_{im}(x_m(t)) + b_i$$

Choice of model (3/3)

- Kernel $k(\mathbf{x}, \mathbf{z})$ can build non-linear functions $h(\mathbf{z}) = \sum_{t'=0}^{T-1} w_{t'} k(\mathbf{x}(t'), \mathbf{z})$
- Local kernels, in one dimension, $k_m(\mathbf{x}, \mathbf{z}) = k(x_m, z_m)$
- $k_{tot} = \sum_{m=1}^{p} d_m k_m$, with $\sum_{m=1}^{p} d_m = 1$ and $d_m \geq 0$ for all m

$$h_i(\mathbf{x}(t)) = \sum_{m=1}^{p} h_{im}(x_m(t)) = \sum_{m=1}^{p} d_{im} \left(\sum_{t'=0}^{T-1} w_{t'} k(x_m(t), x_m(t')) \right)$$

- Gaussian kernel $k(x_m, z_m) = \exp\left(-\frac{(x_m - z_m)^2}{2\sigma^2} \right)$

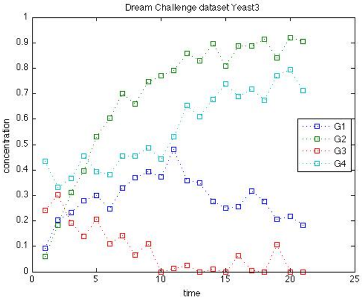
Identification of h

- Cost functional

$$\min C \sum_{t=1}^{T-1} (|y_i(t+1) - h_i(\mathbf{y}(t))| - \epsilon)_+ + \|h_i\|^2_{\mathcal{H}} \qquad (1)$$

  From kernel trick,
  $\|h_i\|^2 = \sum_{t,t'} w_t w_{t'} \left(\sum_m d_{im} k(y_m(t), y_m(t'))\right)$
- Alternate solving in w_t by Lagrangian optimization
- Finding d by reduced gradient descent [2]

Data weighting



- In (1), instead of regularization-error tradeoff term C fixed, we use

$$C_t = C/t$$

Network inferring

Feature selection by multiple kernel learning (FS-MKL)

- Centering and reducing of data
- $d_{i.}$ feature selected for h_i
- $j \rightarrow i$ if d_{ij} in best 10-quantile of $\mathbf{d}$

Hyper-parameter selection

- C, ϵ in grid ; $\sigma = 1$
- Stability [3] : FS-MKL stable w.r.t data perturbation $\mathbf{d}(C, \epsilon, l)$ FS-MKL on subsamples S_l. Similarity

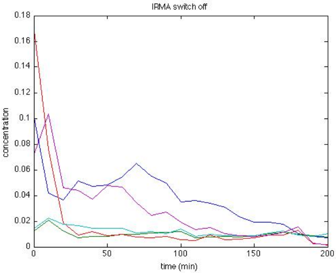
$$s(\mathbf{d}(C,\epsilon,l), \mathbf{d}(C,\epsilon,l')) = \frac{< \mathbf{d}(C,\epsilon,l), \mathbf{d}(C,\epsilon,l') >_{\mathcal{F}}}{\sqrt{\|\mathbf{d}(C,\epsilon,l)\|_{\mathcal{F}}^2 \|\mathbf{d}(C,\epsilon,l')\|_{\mathcal{F}}^2}}$$

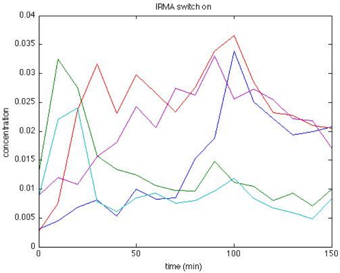
- Subsamples S_l by block-resampling [4]. t_l random integer

$$S_l = (y(t_l), \ldots, y((t_l + 0.80 * T) mod T))$$

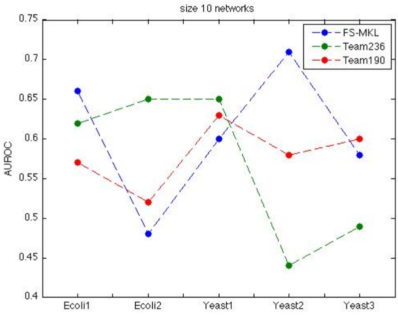
Data sets

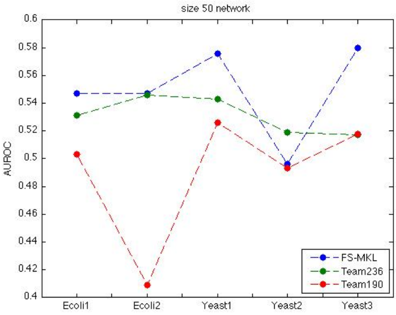
- Challenge DREAM : networks of size 10 (resp. 50) with 4 times series (resp. 23) with 21 time points + steady-state and perturbed data
- IRMA [5] : network of size 5 with 2 time series (with 21 and 16 time points) + steady-state



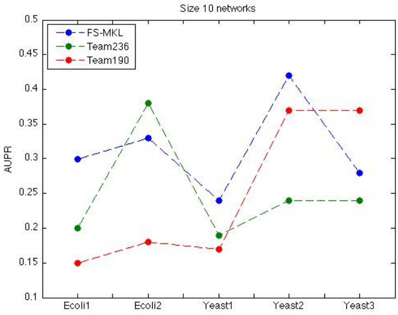


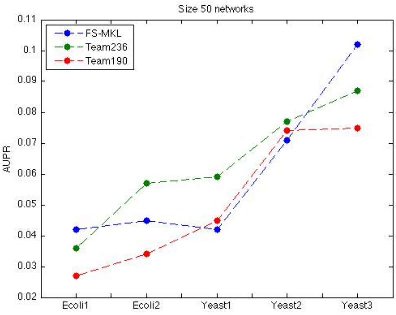
DREAM : AUROC


size 10 networks
AUROC
0.75
0.7
0.65
0.6
0.55
0.5
0.45
0.4
Ecoli1
Ecoli2
Yeast1
Yeast2
Yeast3
FS-MKL
Team236
Team190


size 50 network
AUROC
0.6
0.58
0.56
0.54
0.52
0.5
0.48
0.46
0.44
0.42
0.4
Ecoli1
Ecoli2
Yeast1
Yeast2
Yeast3
FS-MKL
Team236
Team190

DREAM : AUPR


Size 10 networks
FS-MKL
Team236
Team190
AUPR
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
Ecoli1
Ecoli2
Yeast1
Yeast2
Yeast3


Size 50 networks
FS-MKL
Team236
Team190
AUPR
0.11
0.1
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
Ecoli1
Ecoli2
Yeast1
Yeast2
Yeast3

Results : IRMA

IRMA

Method	AUROC_neu	AUPR_neu
FS-MKL	**0.65**	0.63
LASSO	0.57	0.51
LASSO + weight	0.62	0.60
REVEAL	0.64	**0.67**

Conclusion

- Non-linear modeling, with low computational cost, and feature selection for network inference
- Data weighting

Perspectives

- Separate functional costs, allowing introduction of new sources of information
- Consistency of multiple kernel learning with dynamical datas

[1] T Hastie, R Tibshirani - Statistical science, 1986 - JSTOR

[2] A Rakotomamonjy, F Bach, S Canu - Journal of Machine Learning Research, 2008

[3] N Meinshausen, P Bühlmann - Technical report, arXiv: 0809.2932, 2008

[4] Politis, D. N., Romano, J. P. In Exploring the Limits of Bootstrap (R. LePage and L. Billard, eds.) 263–270. 1992

[5] I.Cantone & al, Cell, 2009