# Combining Autosmooth and Alternating Regression for the Estimation of S-System Parameters

**Wolfgang Beyer[a], Ana Teresa Freitas[a,b], João Miranda Lemos[a,b], Marco Vilela[c], Jonas Almeida[c], Susana Vinga*[a,d]**

[a] *INESC-ID Instituto de Engenharia de Sistemas e Computadores: Investigação e Desenvolvimento,* Portugal
[b] *IST Instituto Superior Técnico - Universidade Técnica de Lisboa,* Portugal
[c] *Univ. Texas MDAnderson Cancer Center*, USA
[d] *FCM/UNL Faculdade de Ciências Médicas – Universidade Nova de Lisboa,* Portugal

wolfi.b@gmx.at
svinga@algos.inesc-id.pt

## 1 Abstract

Two recently published methods are combined for the task of generating an S-system model of a biochemical pathway based on time series data of metabolite concentrations: *Autosmooth* [1] for smoothing the raw time course data and estimation of slopes and *alternating regression* [2] for the actual estimation of the model parameters. The results show that this method improves the identification of the dynamic model of the metabolic pathway when compared with previous techniques based on direct estimation from the systems of nonlinear differential equations.

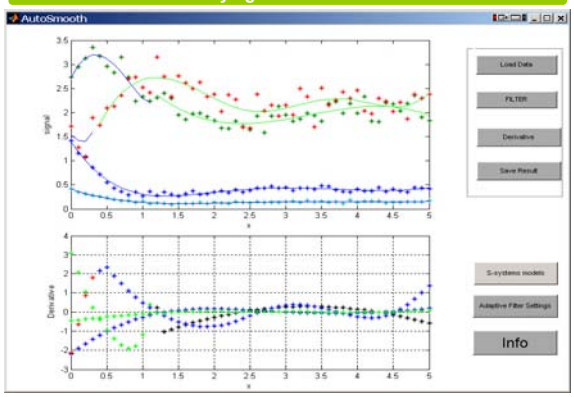
## 2 Introduction

An important challenge in systems biology is the dynamic modeling and optimization of metabolic networks. This task relies on the availability of accurate time course data of metabolite concentrations in living cells generated by methods of modern molecular biology such as in vivo nuclear magnetic resonance (NMR). The use of Biochemical Systems Theory and, in particular, the S-systems modeling framework, reduces the model building to a parameter estimation task. This requires the numerical solution of a system of nonlinear differential equations, which is both difficult and computationally expensive.

## 3 Methods

### Autosmooth [1]

- Separates noise from signal
- Estimates slopes of the signal
- Based on Eilers' extension of the Whittaker filter
- Uses information-theoretic error function
- Able to handle time-varying error structures




Autosmooth can be downloaded from http://sourceforge.net/projects/autosmooth

### S-systems

One equation for each metabolite $X_i$ ($i$=1…$n$)

$$\dot{X}_i = \alpha_i \prod_{j=1}^{n} X_j^{g_{ij}} - \beta_i \prod_{j=1}^{n} X_j^{h_{ij}}$$

Rate constants; Kinetic orders; Production term; Degradation term; Metabolite concentrations

### Decoupling

$n$ coupled differential equations ⇨ $n$ x $N$ algebraic equations

$$S_i(t_k) \approx \alpha_i \prod_{j=1}^{n} X_j^{g_{ij}}(t_k) - \beta_i \prod_{j=1}^{n} X_j^{h_{ij}}(t_k)$$

Slope; Production term; Degradation term

### Alternating Regression [2]

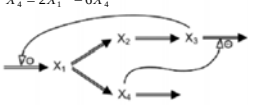
- Cycles between two phases of multiple linear regression (alternating between estimating production term and degradation term parameters)
- Computationally fast due to multilinearity
- Dependence on initial parameter estimations

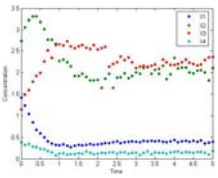
## 4 Results

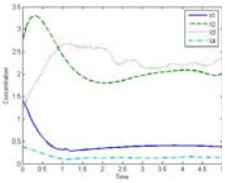
### Procedure

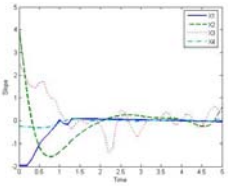
Generate synthetic data → Autosmooth → Alternating Regression → Simulate resulting network

$$\dot{X}_1 = 12X_3^{-0.8} - 10X_1^{0.5} \qquad X_1(t_0) = 1.4$$
$$\dot{X}_2 = 8X_1^{0.5} - 3X_2^{0.75} \qquad X_2(t_0) = 2.7$$
$$\dot{X}_3 = 3X_2^{0.75} - 5X_3^{0.5}X_4^{0.2} \qquad X_3(t_0) = 1.2$$
$$\dot{X}_4 = 2X_1^{0.5} - 6X_4^{0.8} \qquad X_4(t_0) = 0.4$$



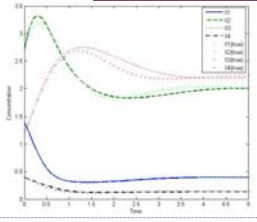







$$\dot{X}_1 = (2.07 + 61.57i)X_3^{0.06i} - (2.17 + 61.61i)X_1^{-0.05i}$$
$$\dot{X}_2 = (5.33 + 0.35i)X_1^{1.11-0.53i} - (0.21 + 0.54i)X_2^{1.70-0.94i}$$
$$\dot{X}_3 = (-1.70 + 165.64i)X_2^{-0.02i} - (-1.80 + 165.63i)X_3^{0}X_4^{0}$$
$$\dot{X}_4 = (-0.19 + 31.86i)X_1^{-0.01i} - (1.15 + 31.85i)X_4^{0.02i}$$

Complex parameter values make no biochemical sense



### Summary

| Data Points | Slopes | Alternating Regression Results | |
|---|---|---|---|
| | | Parameter Values | Curve Fit |
| Exact | Exact | Converge to true values | Perfect |
| Noisy | Exact | Converge to true values (as long as σ is small enough) | Perfect |
| Exact | Autosmooth | Converge to different values | Perfect |
| Noisy | Autosmooth | Converge to complex values | Reasonable |
| Autosmooth | Autosmooth | Converge to complex values | Reasonable |

## 5 Conclusions

- Autosmooth and alternating regression perform very well in a noise-free environment, but adding noise leads to difficulties
- Qualitative predictions about the behavior of the system under different conditions are possible
- Manual fitting may improve the results
- Parameter estimation is still the bottleneck in metabolic modeling using a top-down approach

### References

- Vilela M, Borges C, Vinga S, Vasconcelos AT, Santos H, Voit E and Almeida JS. *BMC Bioinformatics* **2007**, 8:305.
- Chou I, Martens H, Voit E. *Theor Biol Med Model* **2006**, 3:25.

### Acknowledgments

The authors acknowledge financial support by projects DynaMo (PTDC/EEA-ACR/ 69530/2006, FCT) and MaGiC (IE02ID01004, INESC-ID).