System identification of metabolic networks using Nonlinear Bayesian filtering

András Hartmann*,[1], Susana Vinga[1,2] and João M. Lemos[1,3]

[1]INESC-ID - R Alves Redol 9, 1000-029 Lisbon, Portugal [2]FCM-UNL - C Mártires Pátria 130, 1169-056 Lisbon, Portugal
[3]IST-UTL - Av Rovisco Pais, 1049-001 Lisbon, Portugal
*To whom correspondence should be addressed: hdbandi@kdbio.inesc-id.pt

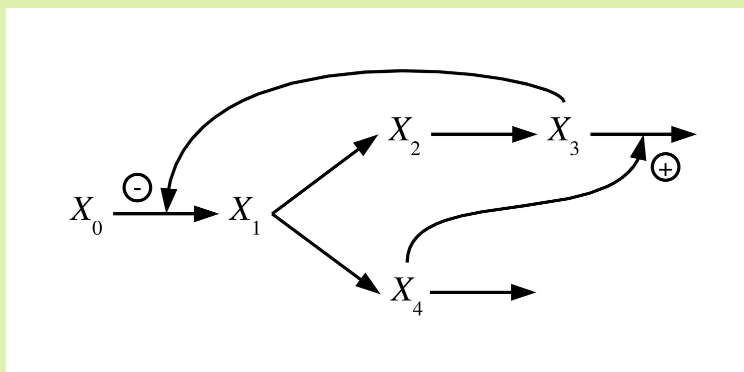
1 Abstract

Understanding dynamic behavior of metabolism is a major challenge in systems biology. To address this challenge, quantitative mathematical models are needed. The metabolite concentrations in living cells can be described by complex systems of nonlinear differential equations with a relative large number of parameters. Our work compares nonlinear Bayesian tools for simultaneous estimation of parameters and states in metabolic networks.

2 Methods

The first step towards applying the above methodology to real data was to develop an unified framework and set up a simulation environment for systems biology data to evaluate the methods in a controlled environment. Here we test the joint filtering approach of the Extended Kalman Filter (**EKF**)[1], the Unscented Kalman Filter (**UKF**)[2] and the Particle Filter (**PF**)[3] with 1000.000 particles.

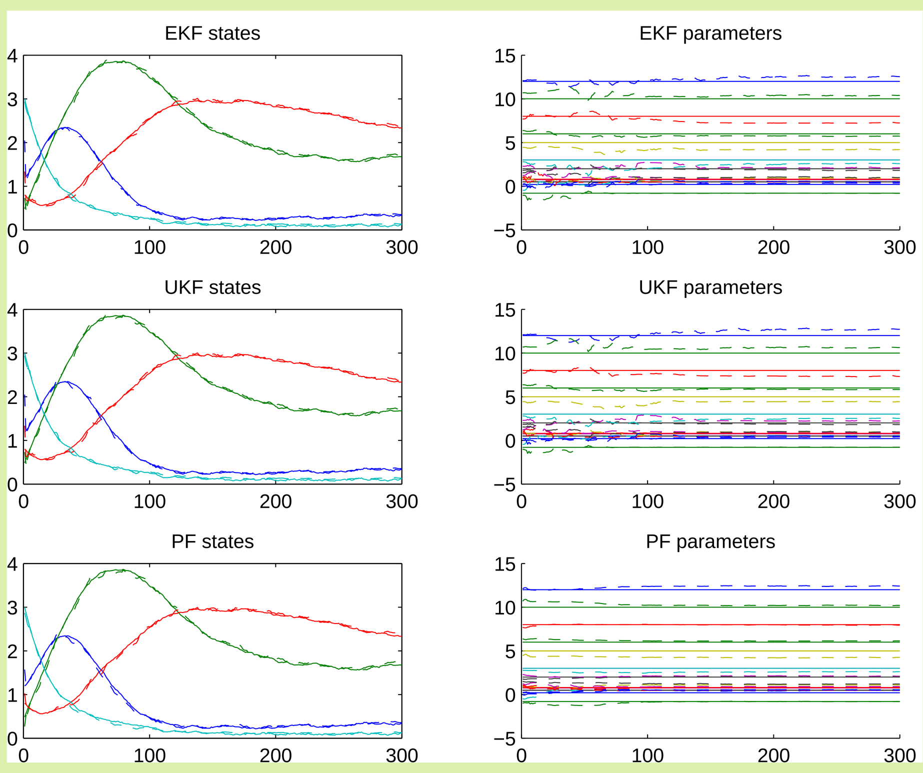
$$\dot{X}_1 = p_1 X_3^{p_2} - d_1 X_1^{d_2}$$
$$\dot{X}_2 = p_3 X_1^{p_4} - d_3 X_2^{d_4}$$
$$\dot{X}_3 = p_5 X_2^{p_6} - d_5 X_3^{d_6} X_4^{d_7}$$
$$\dot{X}_4 = p_7 X_1^{p_8} - d_8 X_4^{d_9}$$



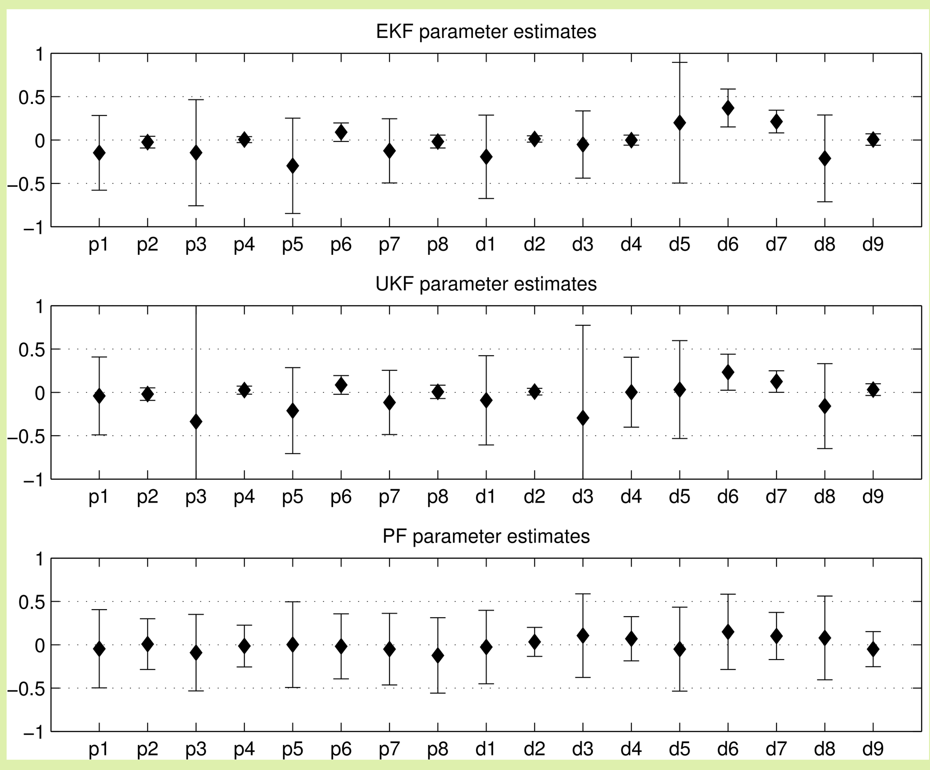
The network used for signal generation is described by a continuous **S-system** with four states and 17 parameters [4].

- Time-series were generated with length of 300 units with identical parameter set
- Additional process and measurement noise were $\mathcal{N}(0, 0.01^2)$ and $\mathcal{N}(0, 0.1^2)$ respectively.
- Sampling frequency for digitalization was 100 Hz
- The tests were run 70 times with different initial values around the true states and parameters.

3 Results and Conclusion



State and parameter identification
The continuous lines show the real states and parameters, the dashed lines are the estimates



Differences between real and estimated values of parameters
Zero here indicates the error free estimate

The methods were compared using Mean Squared Error (MSE) statistics when applied to the synthetic network.

	EKF	UKF	PF
X_1	0.0017	0.0362	0.0307
X_2	0.0025	0.0034	0.1081
X_3	0.0021	0.0024	0.0542
X_4	0.0016	0.0017	0.0113

The table shows average MSE values for the state estimates

Evaluation:

- EKF and UKF performed well on a subset of parameters, and were biased on an another subset
- Using Particle Filter, the mean is closer to the real parameters, but the standard deviation is higher
  - The number of particles is probably still not sufficient according to the number of parameters, but this involves a heavy computational burden.

In conclusion, Bayesian filtering shows promise in parallel identifying states and parameters such systems.

4 References

[1] Welch, G. and Bishop, G. An introduction to the kalman filter Technical report Chapel Hill, NC, USA (1995).
[2] Julier, S. and Uhlmann, J. (1997) *Int. Symp. Aerospace/Defense Sensing, Simul. and Controls* **3**, 26.
[3] Arnaud Doucet, Nando De Freitas, and Neil Gordon, (ed.) June 2001 Sequential Monte Carlo Methods in Practice (Statistics for Engineering and Information Science), Springer, 1 edition.
[4] Voit, E. O. and Almeida, J. July 2004 *Bioinformatics* **20(11)**, 1670–1681.

Acknowledgments

This work was supported by project DynaMo (PTDC/EEA-ACR/69530/2006) FCT, Portugal.