Design and identification of a lin-log kinetics model for the dynamic metabolic network of Escherichia coli

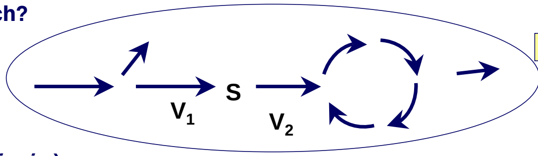
Jo Maertens, Sammy Van Den Broeck, Aditya Bhagwat, Gaspard Lequeux, Peter A. Vanrolleghem
Ghent University - BIOMATH, Department of Applied Mathematics, Biometrics and Process Control, Coupure Links 653, B-9000 Gent, BELGIUM

Introduction

AIM: identify the lin-log kinetic parameters of a dynamic metabolic network model of *E. coli* using data collected during an *in silico* glucose pulse experiment, performed with the model of Chassagnole *et al.* (2002)

WHY the Linlog approach?

Mechanistic kinetics are
- highly nonlinear
- complex
- overparameterized
- based on *in vitro* data (≠ *in vivo*)

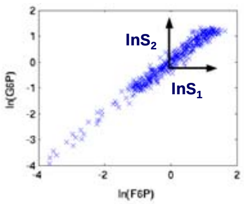


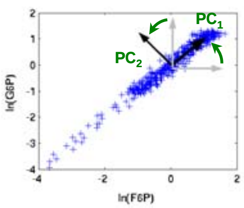
Lin-log kinetics
- linear $v_i = a + b\ln S_1 + c\ln S_2$
- 1 general structure
- multiple linear regression

Three Identification approaches
- Real data → Measured S_i → Fit $V_{mechanistic}$ on S_i → Generate S,v sets, Visser (2002) → (v,S) → Fit Linlog by regression (Y,X)
- Real data → Ideal measurements (S,v), Wu *et al.* (2004) → (v,S) → Fit Linlog by regression
- Real data → Measured S_m (S,σ_S^2) → Calculate dS_m/dt → ($dS_m/dt, S_m$) → Fit Linlog by regression (Y+F,X+E)

Fit Linlog by regression: $Y=X\beta$

Identification procedure





MULTICOLLINEARITY

$Y = b\ln S_1 + c\ln S_2$

= $\ln S_1$ and $\ln S_2$ are correlated
→ Parameters are unidentifiable

How to solve?
- Principal Component Analysis (PCA)

↓ Transformation of Coordinate system

$Y = b`PC_1 + c`PC_2$

- Finds the directions with maximum variability
- PC`s are uncorrelated

Two assumptions of Linear regression are violated

ideal data: $Y=X\beta$

MEASUREMENT ERRORS (F,E) → $Y+F=(X+E)\beta^*$

PRINCIPAL COMPONENTS (PC) → $Y+F=(X+E)\,PC\,\beta^*$ ← Reduce number of PC`s

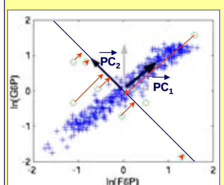
Criteria for reliability Satisfied ? — No → Reduce number of PC`s; Yes ↓

MULTIPLE LINEAR REGRESSION

$\hat{\beta}^{**} \pm \hat{\sigma}_{\beta}$

Bad fit due to data?

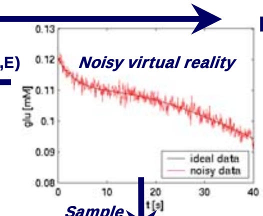
OPTIMAL EXPERIMENTAL DESIGN



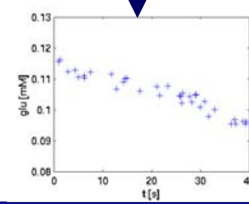
= Increase number of PC`s
e.g. augment the design X with these data X_i o
(see Fig.) whose projection on the PC_2 is as large as possible
→ Increase variability
$Max(|X_i{*}PC_2'|)$

Improved design - Identification procedure

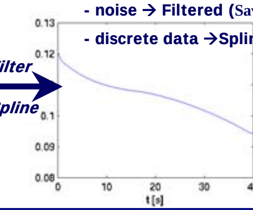
ERRORS-IN-VARIABLES



Sample



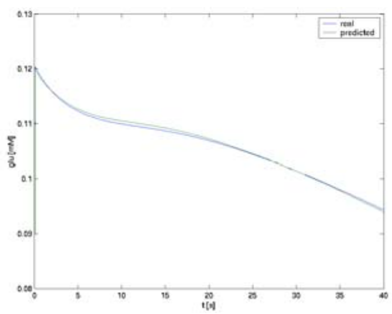
Filter
Spline



$S_m = S + E$ → $X + E$
E = Measurement error, normally distributed $(0,\sigma^2)$

$dS_m/dt = dS/dt + dE/dt$ → $Y+F$
Calculation of the derivative on the basis of the measurements infeasible due to:
- noise → Filtered (Savitzky-Golay filter)
- discrete data → Spline function

Fit Quality



GOOD FIT`s
- MODEL STRUCTURE:
  ‘flexible’ Linlog kinetics
- DATA:
  good, informative data

BAD FIT`s due to
- MODEL STRUCTURE :
  Nonlinearities not catched anymore by linlog approximation due to e.g. additivity
  ln(S → 0)
  not that flexible
  → extrapollation range
- AND/OR
- DATA : too bad, too little data
  → Optimal Experimental Design

Conclusion

- A statistically sound parameter identification has successfully been applied to identify the lin-log kinetic parameters, taking into account multicollinearity and errors-in-variables
- A good fit does not guarantee the uniqueness of the parameter values and does not contribute to the interpretation of the model parameters
- Optimal experimental design can be used to improve the experimental design in view of obtaining more reliable parameter estimates
- Importance of accurate measurements in view of parameter identification asks for a compromise between a large pulse size (signal/noise) and a good approximation by a linear model

Acknowledgements

The first author is Research Assistant of the Fund for Scientific Research – Flanders (Belgium). The authors also wish to thank the Fund for Scientific Research for the financial support in the framework of the F.W.O. - project G.0184.04.

References

Chassagnole, C., Noisommit-Rizzi, N., Schmid, J.W., Mauch, K., Reuss, M., 2002. Dynamic modeling of the central carbon metabolism of E. coli. Biotechnol. Bioeng, 79, 53-72.
Schlosser, P.M., Hoclomb, T., Bailey, J.E., (1993). Determining metabolic sensitivity coefficients directly from experimental data. Biotechnol. Bioeng. 41,1027-1038.
Visser, D., 2002. Measuring & Modeling in vivo kinetics of primary metabolism. PhD dissertation, Technical University Delft, Delft, The Netherlands.
Wu, L., Wang,W., Van Winden, W., Van Gulik, W., Heijnen, J.J., (2004). A new framework for the estimation of controlparameters in metabolic pathways using lin-log kinetics. Eur. J. Biochem. 271,3348-3359.

BIOMATH
Jo.Maertens@biomath.ugent.be - http://biomath.ugent.be/ ˜ MEMO - http://biomath.ugent.be