Estimation of time-varying growth, uptake and excretion rates from dynamic metabolomics data: Supplementary material

Eugenio Cinquemani [1,*], Valérie Laroute [2], Muriel Cocaign-Bousquet [2], Hidde de Jong [1] and Delphine Ropers [1]

[1]Inria, Centre de Recherche Grenoble – Rhône-Alpes, Montbonnot, France
[2]LISBP, Université de Toulouse, CNRS, INRA, INSA, Toulouse, France

Contents

*to whom correspondence should be addressed

S1 Data interpolation

The EKF is a linearized approximation of the Kalman filter that, in some conditions, may lose the stability properties of the linear Kalman filter. That is, divergence may occur along the dynamic processing of data. One way to favor stability is to operate suitable corrections in-between very sparse measurement times. For especially critical data sets, we implement this option by interpolation of the data in-between measurement points, in a way that reflects common-sense prior expectations about the biological dynamics. Given a species i (biomass or metabolite concentration), consider measurement times t_j and t_{j+1} from $\mathscr{T}_i$, and corresponding measurements $y_i(t_j)$ and $y_i(t_{j+1})$, and measurement error standard deviations $\sigma_i(t_j)$ and $\sigma_i(t_{j+1})$. Consider an interpolation time point $t \in (t_j, t_{j+1})$. We introduce a fictitious, linearly interpolated measurement at t, *i.e.*

$$y_i(t) = y_i(t_j) + \frac{y_i(t_{j+1}) - y_i(t_j)}{t_{j+1} - t_j}(t - t_j).$$

Correspondingly, we introduce a measurement uncertainty $\sigma_i(t)$ that is similar to the standard deviations of the nearest measurements, but reflects the fact that increasing time distance from these points introduces increasing uncertainty about the relevance of $y_i(t)$. In particular, for an inflation factor δ, we define the uncertainty of $y(t)$ as

$$\sigma_i(t) = \sigma_i^*(t) + (\delta - 1) \cdot \bar{\sigma}_i \cdot \frac{2 \cdot \min(t - t_j, t_{j+1} - t)}{t_{j+1} - t_j},$$

where $\sigma_i^*(t)$ is the linear interpolation of $\sigma_i(t_j)$ and $\sigma_i(t_{j+1})$ at t, and $\bar{\sigma}_i$ is their mean value. Thus, for any $\sigma_i(t_j)$ and $\sigma_i(t_{j+1})$, the resulting function $\sigma_i(t)$ is a continuous, piecewise affine interpolation. In particular, if $\sigma_i(t_j) = \sigma_i(t_{j+1}) = \bar{\sigma}_i$, then $\sigma_i(t)$ reaches its largest value, equal to $\delta \cdot \bar{\sigma}_i$, at the mid-time between t_j and t_{j+1}. This definition reflects the typical experimental choice of sparser sampling in periods of slow metabolic changes. Indeed, the rate of increase of $\sigma_i(t)$ away from the closest data point is set to be a function of the time distance from the next measurement, *i.e.* sparser data points correspond to presumably slower fluctuations of the quantities observed. A grid of interpolation times t in-between data points for the new measurements can be chosen by a similar rationale. For an integer oversampling factor of choice S, we consider interpolation at times $t_k = t_j + k \cdot (t_{j+1} - t_j)/(S - 1)$, with $k = 1, \ldots, S - 1$, *i.e.* $S - 1$ uniformly spaced data points are introduced. In this way, for varying j, experimental periods with more frequent measurements are maintained as such in the augmented (interpolated) data set.

For the sake of EKS convergence, this strategy was not needed in the validation example and in the first application of the main text, whereas it was used to process the data from the second application.

S2 Additional estimation results

S2.1 Simulation of diauxic shift experiment

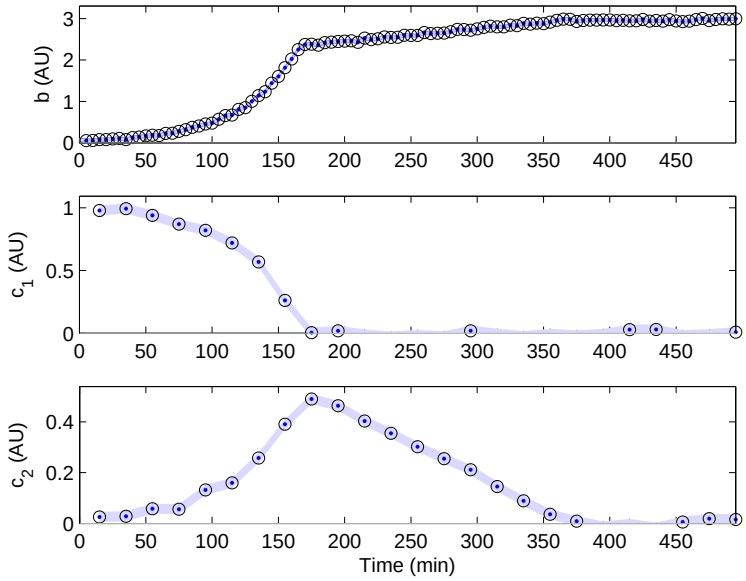


Figure S1: Measurements (circles) from the simulation of a diauxic growth experiment, and their 95% confidence intervals (blue band). The parameters used in the simulation are $(\bar{b}, \bar{c}_1, \bar{c}_2) = (0.05, 1, 0.02)$, $(\bar{\mu}, \underline{\mu}) = (0.0231, 0.0012)$, $(\bar{r}_1, \bar{r}_2, \underline{r}_2) = (0.01, 0.005, 0.001)$, $(\sigma_b, \sigma_{c_1}, \sigma_{c_2}) = (0.02, 0.02, 0.01)$, $T = 5$, $m = 33$ (see main text for the meaning of the symbols). Switching times T_1 and T_2 are implicitly determined by this choice of the parameters.

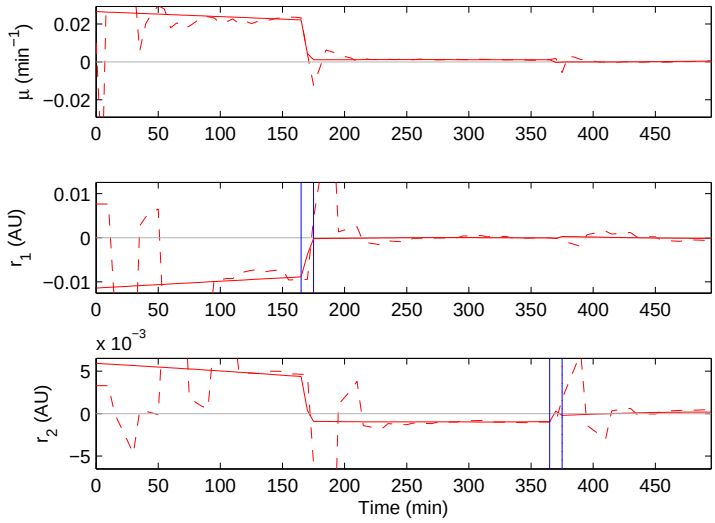


Figure S2: Smoother (solid red) and filter (dashed red) rate estimates in the simulated diauxic growth experiment. Smoother estimates and switching periods (in-between vertical blue lines) are as in Fig. 3 of main text.

S2.2 Simulation of fed-batch experiment

The rate estimation method presented in the paper was also tested on simulated data for a typical fed-batch scenario, where a substrate is added to the culture medium at a time T_{add} in the course of the experiment. As discussed in the main text, this sudden concentration change is not explicitly accounted for in our model. Yet we show here that our method also applies to this scenario.

Fed-batch data was simulated for an experiment comprising a primary and a secondary substrate, as shown in Fig. S3. From time 0 to T_{add}, Eqs (1)–(2) were used for simulation, starting from $b(0) = b^-$ (initial biomass), $c_1(0) = c_1^-$ (initial concentration of primary substrate), and $c_2(0) = 0$ (initial concentration of secondary substrate), with b^- and c_1^- positive constants. Rates were defined as follows. For $g(t) = \max\{c_1(t) - \bar{c}_{1,0}, 0\}$,

$$\mu(t) = \bar{\mu} \cdot g(t),$$

$$r_1(t) = \begin{cases} -\bar{r}_{1,0} - \bar{r}_{1,1}\mu(t), & \text{if } c_1(t) > 0, \\ 0, & \text{otherwise,} \end{cases}$$

$$r_2(t) = \begin{cases} \bar{r}_{2,1} \cdot g(t), & \text{if } c_1(t) > 0, \\ -\bar{r}_{2,2} & \text{if } c_1(t) = 0 \text{ and } c_2(t) > 0, \\ 0 & \text{otherwise,} \end{cases}$$

where $\bar{c}_{1,0}$, $\bar{r}_{1,0}$, $\bar{r}_{1,1}$, $\bar{r}_{2,1}$ and $\bar{r}_{2,2}$ are all positive constants. Thus, growth on the primary substrate and simultaneous excretion of the secondary substrate occur at rates proportional to $g(t) = c_1(t) - \bar{c}_{1,0}$, until $c_1(t)$ falls below threshold $c_{1,0}$. This is followed by the sole consumption of the primary substrate until exhaustion. At this point, the secondary substrate is consumed at a constant rate until (possible) exhaustion. At time $t = T_{add}$, the addition of primary substrate into the medium is simulated by resetting $c_1(t)$ to a positive constant c_1^+ and letting $b(t)$ and $c_2(t)$ unaltered. The system is then simulated from the new state at time T_{add} until a final time T_{max}, in accordance with the same dynamical equations. Measurements of $b(t)$, $c_1(t)$, and $c_2(t)$ are collected at unequally spaced time points in between 0 and T_{max}, corrupted by random measurement errors with standard deviations σ_b, σ_{c_1} and σ_{c_2}. The values of all parameters used in the simulation are given in the caption of Fig. S3. The different phases of utilization or excretion of the two substrates are apparent, as well as the concentration peak corresponding to the addition of primary substrate.

The rate estimation method developed in the main text is then applied to these data, without any knowledge of the specific rate equations. Estimation results are reported in Fig. S4 together with the true (simulated) rates that generated the data. First of all, it is apparent from Fig. 4(a) that the estimation method provides a detailed account of the observed biomass and concentration profiles, including the sudden change in $c_1(t)$ at time $T_{add} = 50$h, even if the latter does not lie within the scope of the model expressed by Eqs. (1)–(2) of the main text. Moreover, from Fig. 4(b), the estimates of $\mu(t)$, $r_1(t)$ and $r_2(t)$ are in excellent agreement with the true rate profiles.

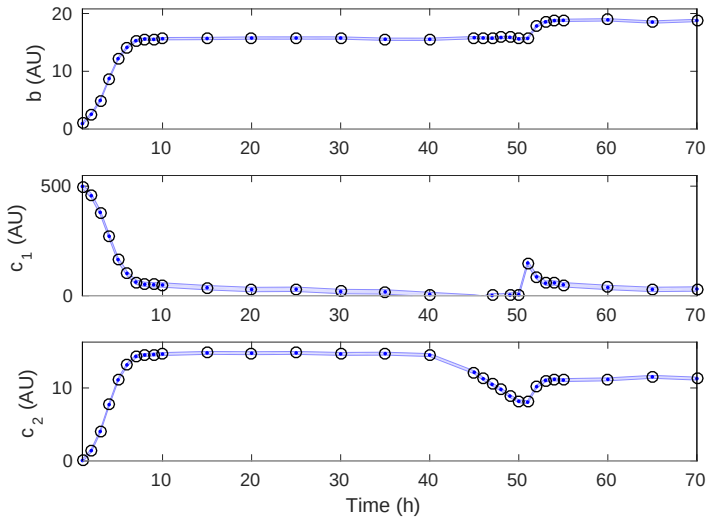
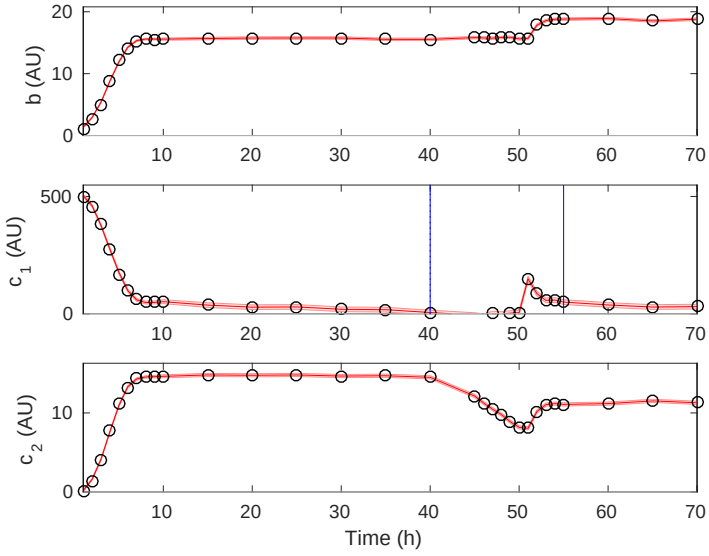
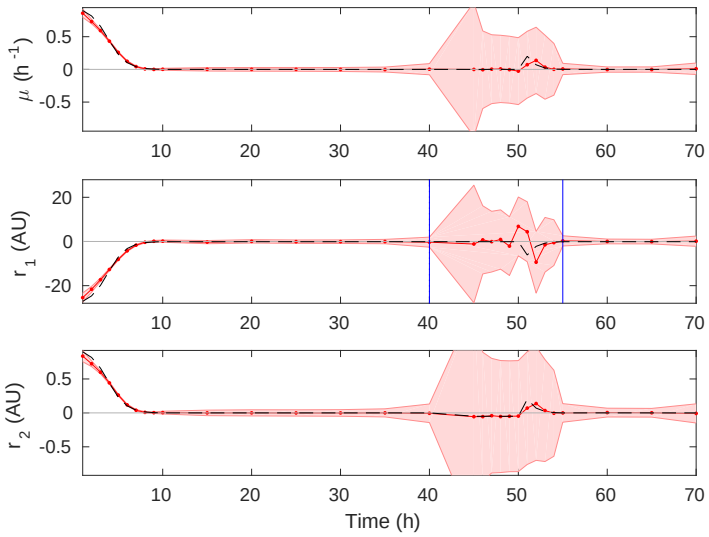


Figure S3: Measurements (circles) from the simulation of a fed-batch experiment, and their 95% confidence intervals (blue band). The parameters used in the simulation are $(T_{add}, T_{max}) = (50, 70)$, $(\bar{\mu}, \bar{c}_{1,0}, \bar{r}_{1,0}, \bar{r}_{1,1}, \bar{r}_{2,1}, \bar{r}_{2,2}) = (0.002, 50, 0.1, 30, 0.002, 0.05)$, $(b^-, c_1^-, c_1^+) = (1, 500, 150)$, $(\sigma_b, \sigma_{c_1}, \sigma_{c_2}) = (0.1, 5, 0.1)$. Time is in hours, biomass and concentrations are in arbitrary units (AU).



(a)



(b)

Figure S4: Estimation of exchange rates by applying the EKS method to a data set obtained by simulating a fed-batch experiment. (a) Simulated data (circles), switching period (in-between vertical blue lines) encompassing exhaustion and subsequent addition of primary substrate, and EKS estimates of biomass and concentration profiles with their 95% credibility intervals (red curves and bands, respectively). Switching period was manually adjusted so as to include T_{add}. Smoothing factors $\gamma^{\circ} = (0.011, 0.284, 0.017)$ were set to 5 times the automatically detected values, whereas the γ^{∞} were left unchanged to the detected values $(2.18, 56.83, 3.50)$. (b) EKS rate estimates for the tuning above, with 95% credibility intervals (red curves and bands, respectively), and simulated rate profiles (dashed black lines).

S2.3 Diauxic growth in *E. coli*

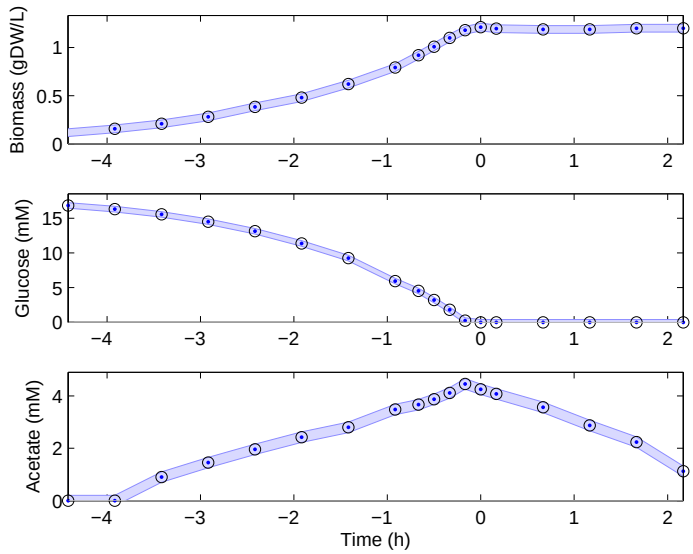


Figure S5: Measurements from diauxic growth on glucose and acetate of *E. coli* (circles and blue dots), and their confidence intervals (size conveyed by the blue shaded bands).

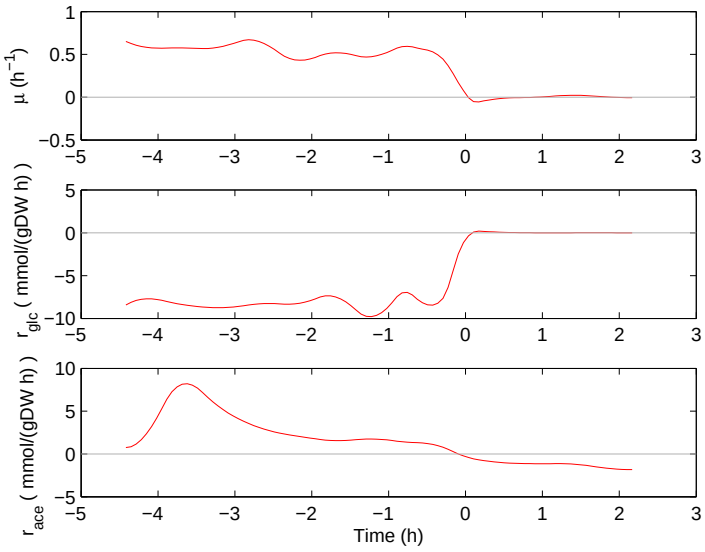
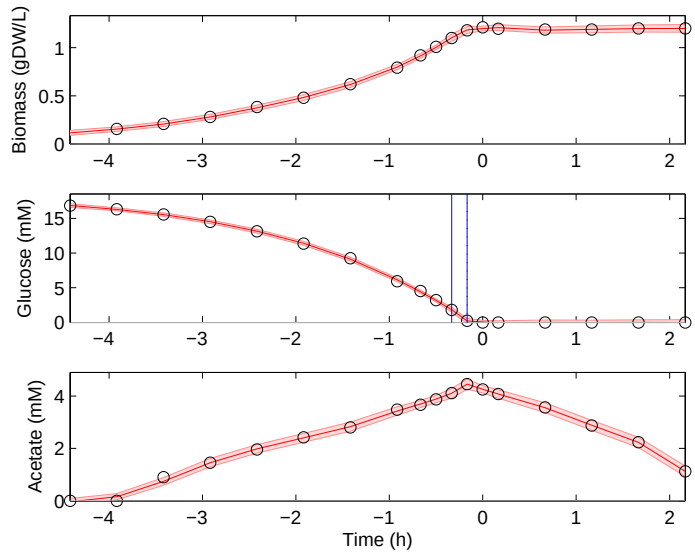
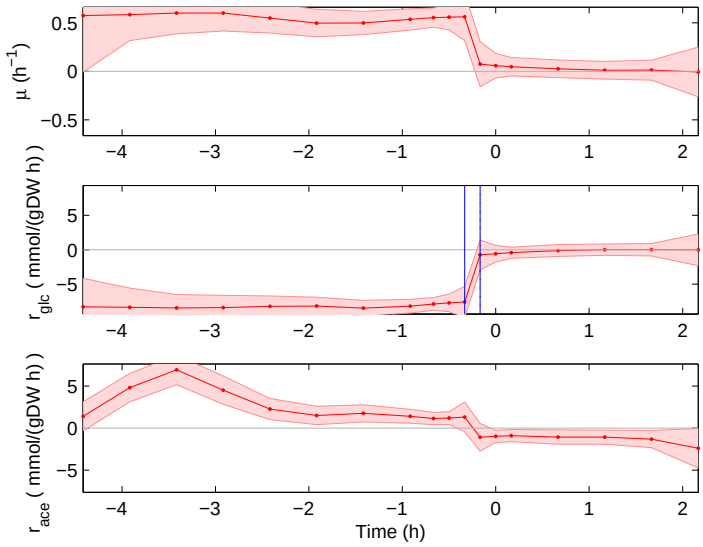


Figure S6: Estimates of rates in *E. coli* diauxic growth experiment, as obtained by cross-validated splining and differentiation (solid red lines).



(a)



(b)

Figure S7: Estimation results for diauxic growth on glucose and acetate of *E. coli* by the fully automated procedure. (a) Data (circles) from Morin *et al.* (2016) and EKS estimates with credibility intervals (solid curve and shaded band) of biomass and concentration profiles. (b) EKS rate estimates and credibility intervals (red curve and shaded band). The smoothing parameters found for biomass, glucose, and acetate are $\gamma_1^\circ = 0.675$, $\gamma_2^\circ = 5.5842$, and $\gamma_3^\circ = 8.4025$, respectively.

S2.4 Fed-batch experiments on *L. lactis*

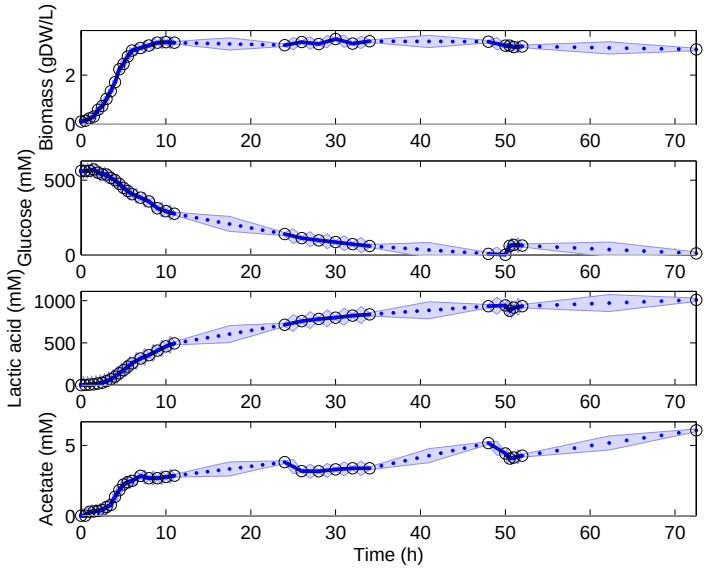


Figure S8: Data from fed-batch experiments with *L. lactis*. Measurements (circles) are interpolated (blue dots) with $S = 10$, i.e. 9 interpolation points in-between measurements, and measurement error inflating factor $\delta = 5$. Shaded blue bands denote confidence intervals on measurements and interpolations.

S3 Experimental procedures

Diauxic growth of the *E. coli* strain K12-MG1655 has been described previously in Morin *et al.* (2016). The other strain used in this study is *Lactococcus lactis ssp. lactis* IL1403. *L. lactis* bacteria were grown in fed-batch culture in semi-synthetic medium containing yeast extract (Sigma-Aldrich) 10 g/L, KH_2PO_4 9 g/L, K_2HPO_4 7.5 g/L, $MgSO_4$,$7H_20$ 0.2 g/L, $MnSO_4$,H_20 0.05 g/L and an initial concentration of glucose of 56 mM. Glucose was added after 50 h, 73 h, and 96 h of culture at the following respective concentrations: 60, 100 and 100 mM.

Cultures were performed at 30°C in a 2L-fermentor (Sartorius, Germany) under air atmosphere with an agitation speed of 250 rpm. The pH was maintained at 6.0 by automatic addition of KOH 10 N.

Bacterial growth was estimated using a Biochrom Libra S4 spectrophotometer at 580 nm (for *L. lactis*, 1 OD unit is equivalent to 0.3 g/L). The bioreactor was inoculated with exponential phase cells from pre-cultures grown in the same medium at an initial OD_{580} of 0.04.

The concentrations of glucose and fermentation products (lactic acid and acetate) were measured in culture supernatants every 30 minutes by high-performance liquid chromatography with a 1200 series system (Agilent Technologies, Waldbronn, Germany) as previously described (Cocaign-Bousquet and Lindley, 1995). Briefly, an HPX87H+ Bio-Rad column was kept at a temperature of $48C$ with H_2SO_4 (5 mM) as the eluent at a flow rate of 0.5 mL min^{-1}. Dual detection was performed by a refractometer and UV analyses.

S4 Metabolic flux analysis

The genome-scale model and constraints used in this study are described in Morin *et al.* (2016). The model is a slightly modified version of the genome-scale reconstruction iAF1260-flux2 of *E. coli* (Feist *et al.*, 2007). We ran metabolic flux analyses with the COBRAv2 toolbox with GLPK as the linear programming solver (Schellenberger *et al.*, 2011), following the procedure described in (Morin *et al.*, 2016). Basically, we determined the intracellular distribution of fluxes consistent with the estimated rates (either the smoothing spline or the EKS estimates), at 7 minutes before the glucose exhaustion, when the smoothing spline and EKS estimates are most different. The optimization problem to solve consists in minimizing the discrepancy between the measured and predicted specific rates, namely the growth rate, the glucose uptake rate and the acetate excretion rate. The minimization is followed by a flux variability analysis (Mahadevan and Schilling, 2003) in order to assess the robustness of the optimal solutions. The lower and upper bounds of the fluxes consistent with the exchange and growth rates are plotted for 15 reactions of carbon central metabolism in Fig. S9. The reactions, as named in the iAF1260-flux2 reconstruction, are from different metabolic pathways: glycolysis (PGI, PFK, ENO, PYK), pentose-phosphate pathway (G6PDH2r, TKT2, RPI), glycogenolysis (PGMT), Krebs cycle and anaplerotic reactions (MDH, CS, PPCK, PPC), and acetate metabolism (PTAr).

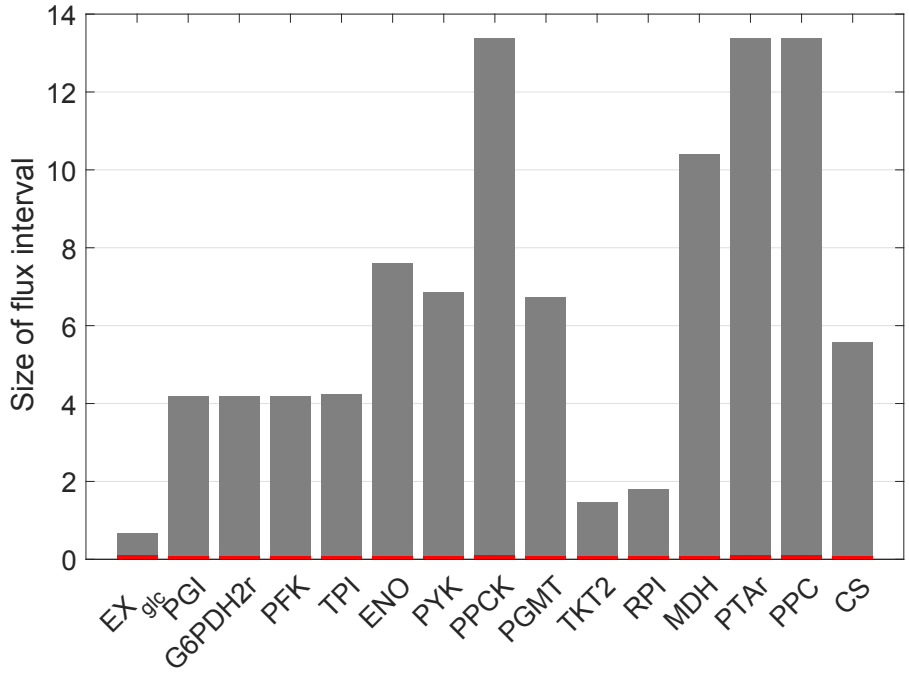


Figure S9: Size of flux intervals obtained by flux variability analysis for the optimal solutions for 15 selected reactions in carbon central metabolism using the smoothing spline (grey) and EKS (red) estimates.

We verified the accuracy of predicted fluxes by comparing their mean value to measured intracellular fluxes in a continuous culture of *E. coli* growing on glucose at a dilution rate of 0.2 h^{-1} (Zhao *et al.*, 2004). The comparison was performed for reactions in the carbon central metabolism for which direct measurements are available: EX_{glc}, PGI, G6PDH2r, PGMT, GAPD, PYK, PDH, PPC, MDH, TKT1, and TKT2. The results are plotted in Fig. S10, and show an excellent correspondence between predicted and measured fluxes.

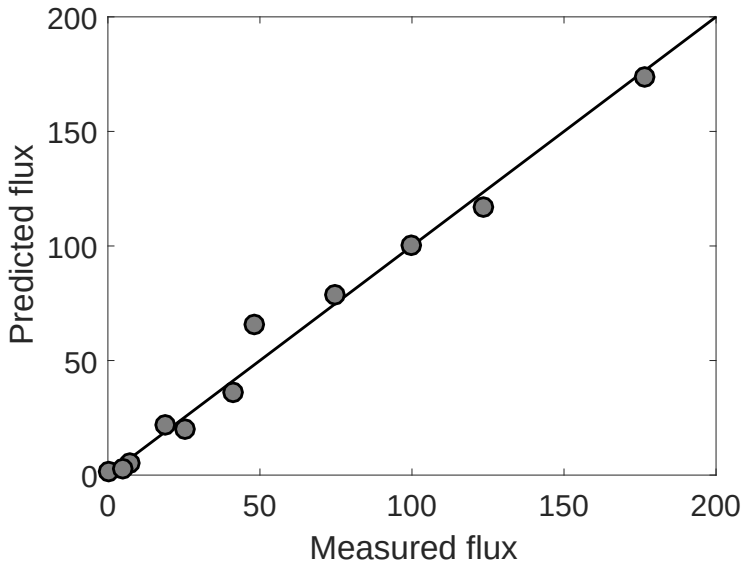


Figure S10: Comparison of predicted and measured intracellular fluxes for eleven reactions in carbon central metabolism (EX_{glc}, PGI, G6PDH2r, PGMT, GAPD, PYK, PDH, PPC, MDH, TKT1, and TKT2) of *E. coli*. Fluxes are given relative to the specific glucose consumption rate.

S5 Model extensions

The model of Eqs (1)-(2) in the main text is based on the assumption that the only changes in the concentration c_i are due to the uptake and excretion rates, and thus applies to batch growth. The model can be easily generalized to the case of the inflow or outflow of mass in the bioreactor. The following modeling scheme, adapted from (Bastin and Dochain, 1990), takes into account these generalizations:

$$\dot{b}(t) = (\mu(t) - d_{in})b(t), \tag{S1}$$

$$\dot{c}_i(t) = r_i(t)b(t) + d_{in}c_{in,i} - d_{in}c_i(t) - g_i(t), \qquad i = 1, \ldots, n, \tag{S2}$$

$$\dot{V} = (d_{in} - d_{out})V, \tag{S3}$$

where b [gDW·L^{-1}] is the biomass concentration in the medium in the bioreactor, c_i [g L^{-1}] the concentration of metabolite i in the medium and in the effluent, $c_{in,i}$ [g L^{-1}] the concentration of metabolite i in the influent, V [L] the volume of the medium, $d_{in} = F_{in}/V$ [h^{-1}] the inflow rate (with F_{in} [L h^{-1}] the influent flow rate), $d_{out} = F_{out}/V$ [h^{-1}] the outflow rate (with F_{out} [L h^{-1}] the effluent flow rate), $g_i(t)$ the (time-varying) rate of outflow of metabolite i in gaseous form or its rate of degradation in the medium.

Notice that Eqs (1)-(2) in the main text are obtained for $d_{out} = d_{in} = g_i(t) = 0$. This corresponds to cultivation in batch mode, without escape of metabolites in gaseous form or extracellular degradation. In case that $d_{out} = d_{in} = d > 0$, with d the dilution rate, the bioreactor operates in continuous mode. Notice that in batch and continuous mode it holds that $\dot{V} = 0$, so that this variable can be omitted from the model. When $d_{out} = 0$ and $d_{in} > 0$, the bioreactor operates in fed-batch mode. In this case, the volume of the medium is not constant.

The extended Kalman smoothing method developed in the main text applies to each of these three cases, with an automated tuning phase that should be adapted to the new form of the model. In particular, the smoothing profiles $\gamma(t)$ (see main text) can equally be defined in terms of pre-estimates of the unknown rate profiles, but the latter should be determined algebraically from spline interpolations of the data (main text Section 3.3.1) in the light of the new form of the dynamic model equations.

S6 Description of estimation code

S6.1 Overview

The estimation method described in the main text, comprising data-driven tuning of the EKS and the EKS itself, has been implemented in Matlab (developed on releases R2014a/R2016a). The software takes as input a structure `data` containing measurements $\mathscr{y}_1, \ldots, \mathscr{y}_{n+1}$ and corresponding measurement times $\mathscr{T}_1, \ldots, \mathscr{T}_{n+1}$, time-invariant or time-dependent measurement error levels $\sigma_1, \ldots, \sigma_{n+1}$, as well as metabolite names and unit labels for visualization purposes. Based on this, it first performs switch detection and determines smoothing profiles and initial state statistics as described in Sec. 3.3 of the main text (internal parameters $L = 10$, $D = 1$, $J = 3$), and then runs the estimation algorithm of main text Sec. 3.2. In accordance with the main text model (1)-(2), the smoother equations are fixed by the number of metabolites n, which is determined from the data themselves. By default, the procedure returns a structure `results` with the estimates of the augmented state z (including $\hat{x}$ and rate estimates $\hat{u}$) at all measurement times $\mathscr{T}$, its estimation error covariance matrix P (including P_x and P_u) at the same times, as well as the pre-fitting results $\tilde{x}$ and $\tilde{u}$. In a structure `settings`, it also returns the EKS settings (*i.e.*, γ_i°, γ_i^∞, switching periods, *etc.*) automatically determined at the tuning stage. These output settings can be inspected and modified by the user, and fed back into a subsequent call of the routine as an additional input accompanying the `data` structure. Alternatively, all or part of these settings can be directly specified by the user and passed to the software (in a `settings` structure) along with the data. The format of `data`, `settings`, `results`, and more details are given in the next section.

S6.2 Implementation

We describe here the overall architecture and usage of the software (refer to the main text for the mathematical symbols). The software comprises the following Matlab function (`.m` files):

`EstimateRates`: The main routine calling the other routines for automatic tuning of the EKS and the computation of the EKS estimates

`PreprocessData`: Called by `EstimateRates`, performs the automatic data-driven tuning of the EKS settings by computing the switching periods and, invoking `EstimateFilteringParameters`, initial state statistics and smoothing factors.

`EstimateFilteringParameters`: Called by `PreprocessData`, performs optimized spline smoothing of the data by a Generalized Cross-Validation approach, computes rate estimates by standard differentiation, and infers initial state statistics and smoothing factors.

`ExtendedSmoother`: Called by `EstimateRates`, it implements the EKS for the computation of state and rate estimates. Uses `ApproximateOneStepDynamics`.

`ApproximateOneStepDynamics`: In-between two measurement times, solves the system of ODEs for the prediction step of the EKF and the sensitivity equations for the precomputation of matrix factors of the Bryson-Frazier smoother. Uses `AugmentedDynamics` and `VariableSmoothing`.

`AugmentedDynamics`: Evaluates the vector field of the augmented systems of ODEs for the biological system dynamics and the Bayesian description of the unknown rate profiles, together with its Jacobian, at a given state value and time instant.

`VariableSmoothing`: Implements the time-varying smoothing profiles of the EKS

For result visualization purposes, `EstimateRates` and `EstimateFilteringParameters` also call the publicly available Matlab function `shadedErrorBar`.

The main function is `EstimateRates`. For the intended usage of the software, this is the only function that the user shall call, defined as follows:

```
function [results,outsettings]=EstimateRates(data,settings)
```

Both `data` and `settings` are required input structures, although `settings` may be an empty matrix or cell array. Both `results` and `outsettings` are optional output structures. Structure `data` must contain the data to be processed and may contain physical variable names for result visualization purposes. Recalling that n denotes the number of experimentally measured metabolites, the fields `fieldname` of `data` (*i.e.* the various entries of the type `data.fieldname`) shall be:

`Measurements`: $(m+1) \times (2+n)$ matrix containing $(m+1)$ measurement times ($\mathscr{T}$, first column) in increasing order and corresponding experimental observations $\mathscr{Y}$ (second column: biomass; column c, from third to last: concentration of metabolite $c-2$. `NaN` entries are admitted for non-available measurements of some entries at given times; no row can have only `NaN` entries.

`ErrorLevels`: $1 \times (1+n)$ matrix containing the standard deviation of measurement errors σ_i for biomass and metabolite concentrations (same ordering as for `Measurements`), if time-independent. Alternatively, $(m+1) \times (1+n)$ matrix containing the standard deviations of measurement errors at different observation times for the different quantities.

Optionally, the following fields may be provided:

`NameLabels`, `UnitLabels`, `RateNameLabels`, `RateUnitLabels`: Cell arrays of $2+n$ strings specifying the names (labels) of time, biomass and metabolite concentrations for `NameLabels`, and corresponding units for `UnitLabels`, and of time, growth rate and metabolite exchange rates for `RateNameLabels`, and corresponding units for `RateUnitLabels`.

Irrespective of `settings`, the output structure `results` is returned with the following fields:

`EstimationTimes`: $M \times 1$ array of times at which estimates are produced ($M \geq m+1$. It equals $m+1$ unless additional times other than measurement times are requested for estimates, see setting `EstimationTimes` further below).

`PredictorEstimates`, `FilterEstimates`, `SmootherEstimates`: $M \times 3 \cdot (n+1)$ matrix of estimates of the augmented state z, each row corresponding to the time in the same row of `EstimationTimes`. `Predictor`, `Filter` and `Smoother` contain estimates $\hat{z}^-$, $\hat{z}$, and $\hat{z}^+$, in the same order. In particular, `SmootherEstimates` contains the desired estimates from the EKS.

`PredictorCovariance`, `FilterCovariance`, `SmootherCovariance`: $3{\cdot}(n{+}1){\times}3{\cdot}(n{+}1){\times}M$ matrices containing the (symmetric) estimation error covariance matrices P^-, P and P^+ corresponding to $\hat{z}^-$, $\hat{z}$, and $\hat{z}^+$ at the M estimation times (in the third dimension).

Depending on `settings`, if spline smoothing is performed for automatic tuning of the smoothing factors and initial rate and state statistics, the following field is also returned:

`Prefits`: A structure with fields as follows:

`Prefits.x0`, `Prefits.u0`: Initial state estimates $\hat{x}_0^-$ and rate estimates $\hat{u}_0^-$, in the same order.

`Prefits.HeavySmoothFact`: $1 \times (n+1)$ array containing the estimated smoothing factors $\gamma_1^\circ, \ldots, \gamma_{n+1}^\circ$.

`Prefits.SpFits`: $1 \times (n+1)$ cell array containing the cross-validated spline fits used for automatic EKS tuning for biomass and the different metabolites

`Prefits.SpRateTimes`: $2 \times (n+1)$ array where every column, say `[Tstart;Tstop]`, with `Tstart<Tstop`, specifies the period over which splines have been used to derive estimates of the rate profile corresponding to that column.

Manual settings for the EKS can be specified with the following optional fields of `settings`:

`AnalysisPeriod`: 2×1 vector with entries `[Tmin ; Tmax]`, providing the initial and final time of the analysis (if absent, automatically determined as whole data span; Only times within data span are considered).

`EstimationTimes`: $m' \times 1$ vector of times at which estimation results have to be returned (in addition to measurement times; only times within `[Tmin ; Tmax]` are considered).

`HeavySmoothFact`: 1×1 smoothing factor for slow rate dynamics. Alternatively, $1 \times n$ vector of smoothing factors, one per rate (if absent, estimated from data).

`LightSmoothFact`: 1×1 smoothing factor for fast rate dynamics. Alternatively, $1 \times n$ vector of smoothing factors, one per rate (if absent, fixed to `1e3*HeavySmoothFact`).

`Tmins, Tmaxs`: $K \times (n+1)$ matrices, where, for any j, `[Tmins(j,i) Tmaxs(j,i)]` is the time period where a switch is authorized. Entries `Tmins(j,i)>=Tmaxs(j,i)` have no effect. See `outsettings` below for a typical structure of this matrix.

`SwitchThresholds`: $1 \times (n+1)$ vector with thresholds for detection of exhaustion of different species. Alternatively, $(m+1) \times (n+1)$ matrix with different thresholds for the different measurement times (if absent, set to `2*ErrorLevels`). Overridden if both `Tmins` and `Tmaxs` are provided.

`SwitchPeriod`: 2×1 vector with entries `[DTminus ; DTplus]`, defining the period of fast rate changes as `[Tswitch+DTminus, Tswitch+DTplus]`, for any automatically detected switching time `Tswitch`. Alternatively, $2 \times (n+1)$ matrix with different values for the different species (if absent, automatically fixed based on measurement times as described in main text). Overridden if both `Tmins` and `Tmaxs` are provided.

`ShowProgress`: Scalar flag, specify 1 for waitbar and other info on the progress of the analysis (default if absent), 0 otherwise.

`HeavySmoothFact`: 1×1 smoothing factor for slow rate dynamics, equal for all species. Alternatively, $1 \times (n+1)$ vector of smoothing factors, one per rate (if absent, estimated from data).

`LightSmoothFact`: 1×1 smoothing factor for fast rate dynamics. Alternatively, $1 \times (n+1)$ vector of smoothing factors, one per rate (if absent, fixed to `1e3*HeavySmoothFact`)

Additional optional fields pertain and trigger linear data interpolation, affecting EKS but not its tuning, as discussed in Sec. S1 above. They are:

`Oversampling`: 1×1 integer number (parameter S of Sec. S1) of uniformly spaced interpolation points from a measurement time (included) to the next (excluded). If equal to 1, in particular, the `NaN` entries of `data.Measurements(:,2:end)` are replaced by linearly-interpolated values, and no interpolated measurement time is added. If absent, no interpolation is operated.

`ErrorInflateFact`: Amplification factor of error standard deviation in-between time points (factor δ of Sec. S1; used in case of data interpolation; if absent, set to `2` by default).

The specified manual settings are transferred to identical fields of `outsettings`. This output structure reports the settings actually used by the software in the processing of `data`, whether manually specified, set by default, or automatically detected. This includes in particular matrices `Tmins` and Tmaxs (see `settings` above), typically not specified by the user, but automatically determined in a format where every row j concerns one switch detected for one species j, with `Tmins(j,i)<Tmaxs(j,i)`, and all other entries of the same row are zero in both matrices. Structure `outsettings` can be easily inspected, modified, and fed back into a new call to `EstimateRates` for manually adjusted data processing.

Scripts showing the usage of this software for the validation study and the two applications in the paper are available together with the software.

References

Bastin, G. and Dochain, D. (1990). *On-line Estimation and Adaptive Control of Bioreactors*. Elsevier, Amsterdam.

Cocaign-Bousquet, M. and Lindley, N. (1995). Pyruvate overflow and carbon flux within the central metabolic pathways of *Corynebacterium glutamicum* during growth on lactate. *Enzyme Microb Technol*, **17**(3), 260–7.

Feist, A., Henry, C., Reed, J., Krummenacker, M., Joyce, A., Karp, P., Broadbelt, L., Hatzimanikatis, V., and Palsson, B. (2007). A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol. Syst. Biol.*, **3**, 121.

Mahadevan, R. and Schilling, C. H. (2003). The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab. Eng.*, **5**(4), 264–276.

Morin, M., Ropers, D., Letisse, F., Laguerre, S., Portais, J., Cocaign-Bousquet, M., and Enjalbert, B. (2016). The post-transcriptional regulatory system csr controls the balance of metabolic pools in upper glycolysis of *Escherichia coli*. *Mol. Microbiol.*, **100**(4), 686–700.

Schellenberger, J., Que, R., Fleming, R. M., Thiele, I., Orth, J. D., Feist, A. M., Zielinski, D. C., Bordbar, A., Lewis, N. E., Rahmanian, S., Kang, J., Hyduke, D. R., and Palsson, B. . (2011). Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat. Protoc.*, **6**(9), 1290–1307.

Zhao, J., Baba, T., Mori, H., and Shimizu, K. (2004). Effect of *zwf* gene knockout on the metabolism of *Escherichia coli* grown on glucose or acetate. *Metab. Eng.*, **6**(2), 164–74.