Imperial College London

Sargent Centre for Process Systems Engineering

Online State and Dynamic Parameter Estimation in Biotherapeutic Production through Ensemble Kalman Filtering

19th international EnKF workshop (2024)

Luxi Yu

Supervisors: Professor Cleo Kontoravdi; Dr. Ehecatl Antonio del Rio Chanona

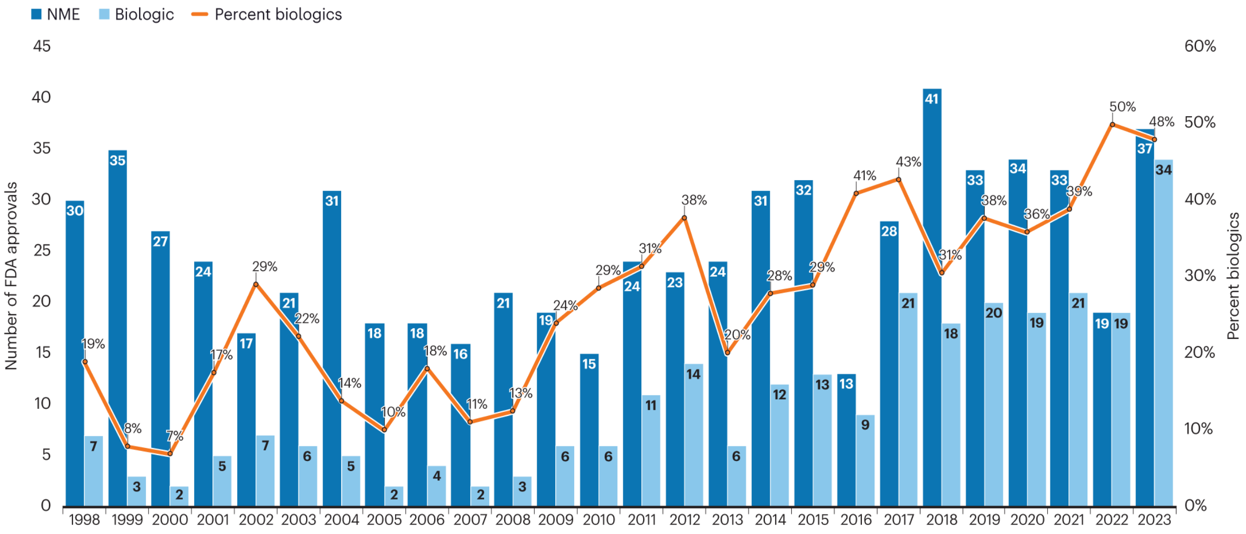
Overview

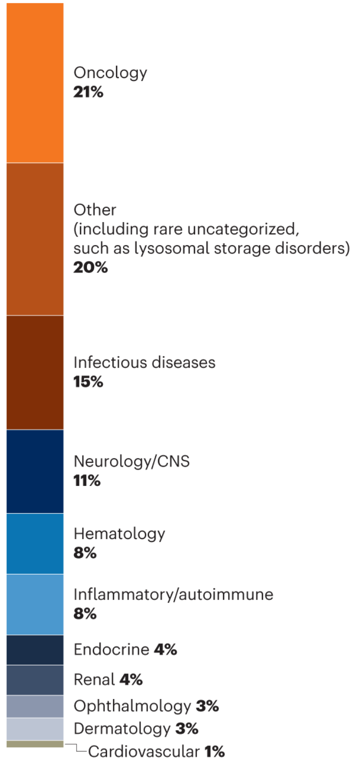
- Introduction of biotherapeutics and bioprocessing
- Applying mathematical modelling and challenges
- Motivation for employing EnKF for
  Combined State and Parameter Estimation
- Case Study: Applying EnKF for
  Online State and Dynamic Parameter Estimation of a Mild Hypothermia Process

- **Please help-challenges ahead**

Biologics: Record-breaking FDA approvals

Biologically derived drugs

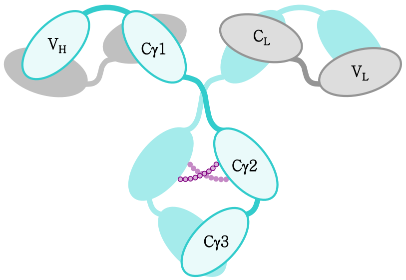


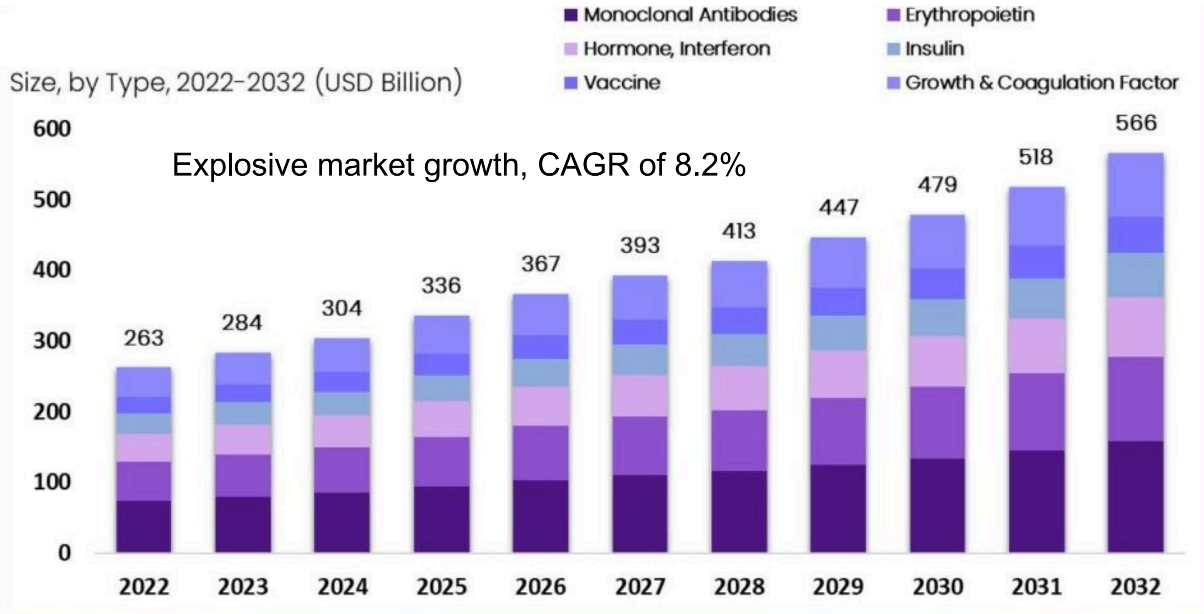


Senior, M. Fresh from the biotech pipeline: record-breaking FDA approvals. *Nat Biotechnol* **42**, 355–361 (2024).

Global Biopharmaceuticals Market

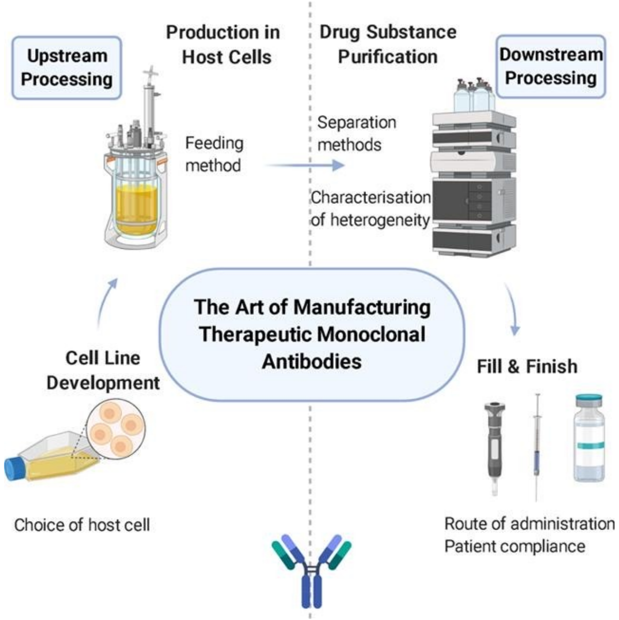
- Dominated by **Monoclonal Antibodies (mAb)**
- Highly specific targeting





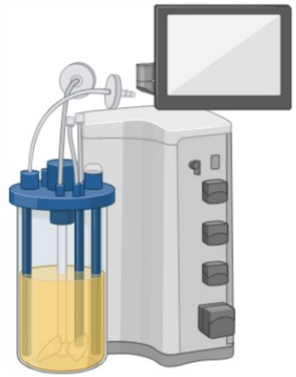
Bioprocessing of therapeutic proteins in mammalian cells

- Industrial production of therapeutic proteins, rely on living cells
- Mammalian cells are favoured due to compatibility to human bodies
- Nearly 70% of therapeutic proteins are produced in Chinese Hamster Ovary (CHO) cells.

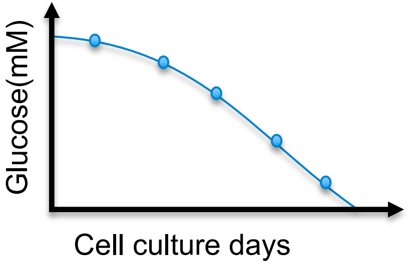


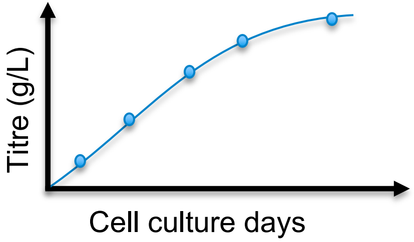
Stefania C. Carrara, Michael Ulitzka, Julius Grzeschik, Henri Kornmann, Björn Hock, Harald Kolmar, International Journal of Pharmaceutics 2021, Volume 594,

Mechanistic model development



Data Collection





Model Development

$$\mu = \mu_{max} \prod_{i=1}^{N.C.} \left(\frac{C_i}{K_{M,i} + C_i} \right)$$

$$q_i = \alpha_i \left(\frac{C_i}{K_{M,i} + C_i} \right)$$

$$q_p = \alpha_p \sum_{i=1}^{N.C.} \alpha_i \, q_i$$

Monod Kinetics

$$\frac{d(VX_V)}{dt} = \mu V X_V - Q_{out} X_V$$

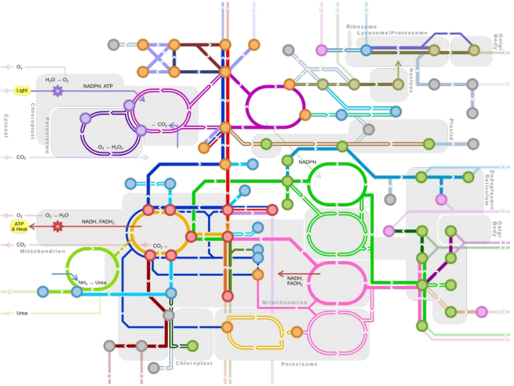
$$\frac{d(VC_{i,in})}{dt} = Q_{in} C_{i,in} - Q_{out} C_i + q_i V X_V$$

$$\frac{d(VP)}{dt} = q_p V X_V - Q_{out} P$$

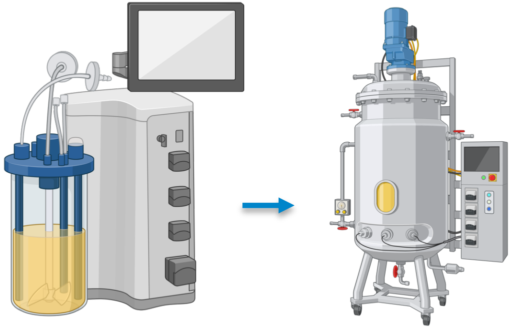
Material Balances

- Apostolos Tsopanoglou, Ioscani Jiménez del Val,, Current Opinion in Chemical Engineering,Volume 32,2021,100691,.
- Created with BioRender.com

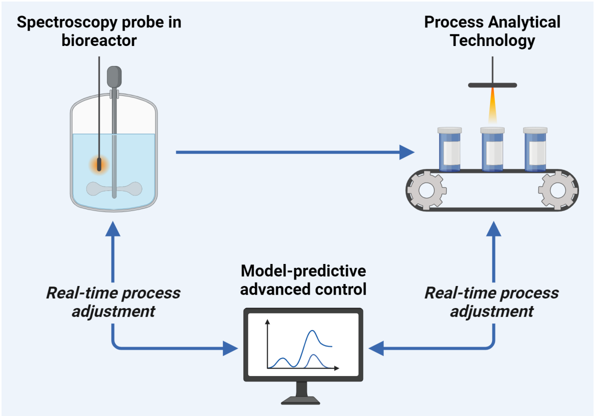
Challenges in applying mechanistic models



Complexity of biological systems:
- non-linear system dynamics
- parameter sensitivity and identifiability
- Inherent variability but fixed parameter



Challenges in knowledge transfer:
- specific to cell line, experimental condition and scale
- high quality and comprehensive data required for validation

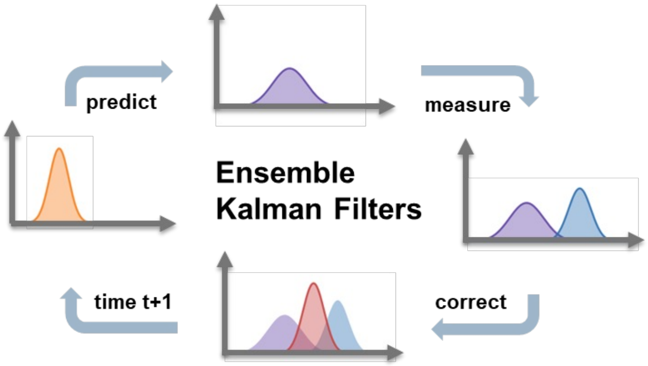


Integration with process control:
- require models to be adaptable and responsive to real-time data
- models need to be robust to deal with uncertainty and variability

Motivation for applying Ensemble Kalman Filter (EnKF)

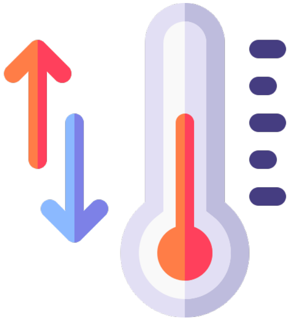
Ideally…

- Avoid reparametrize the model from scratch for a different system
- Identify uncertainty explicitly:
  - Fundamental differences – e.g. different cell lines/metabolic pathways
  - Does it come from biological variability?- e.g. same cell line
  - Does it come from sensor noise? Parameters? Inputs? Outputs?
- Knowledge transfer – parameters that needs to be adjusted/kept same
- Robust model for online real time control, allow dynamic adaptation of states and parameters



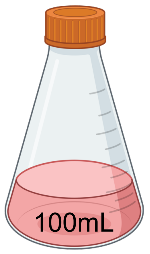
Imperial College London

Case Study: Temperature Downshifts in Bioprocessing



- Mild hypothermia: temperature downshift half-way through the production run
- Common practice in industry to enhance cell culture productivity
- Slow down cell growth/death, higher viability of cells, higher titre and less impurities
- Challenge:
  - highly specific mechanistic model, no temperature dependence understanding
  - time invariant parameters unsuitable for flexible temperature downshift regime

Case Study: Temperature Downshifts in Bioprocessing

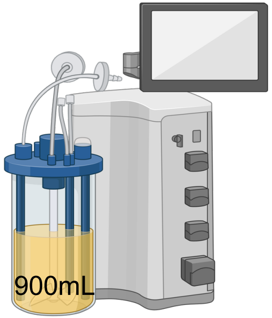


Shake flask

- Cell-line: CHO-T127
- Product: IgG
- Fed-batch (feed every 2 days)

- Volume: 100mL
- No pH, antifoam control
- Temperature: 36.5 °C

- Model: (Kotidis *et al.* 2019)



Bioreactor

- Cell-line: CHO-T127
- Product: IgG
- Fed-batch (feed every 2 days)

- Volume: 900mL
- pH, antifoam control
- Day 6 temperature downshift: 32 °C

EnKF Combined States and Parameter Estimation
(True: Sou *et al.* 2017 32 °C)

No. of states: 8; No. of parameters: 24

EnKF for dual state and parameter estimation-algorithm

- Parameter evolution, sample parameters $\theta_{t+1}^{i-} = \theta_t^{i+} + \tau_t^i, \quad \tau_t^i \sim N(0, \Sigma_t^\theta)$
- State ensemble prediction with sampled parameters $x_{t+1}^{i-} = f(x_t^{i+}, u_t^i, \theta_{t+1}^{i-})$
- Update parameter ensemble members according to standard EnKF $\theta_{t+1}^{i+} = \theta_{t+1}^{i-} + K_{t+1}^\theta(y_{t+1}^i - \hat{y}_{t+1}^i)$

where Kalman gain for parameters is $K_{t+1}^\theta = \Sigma_{t+1}^{\theta y}[\Sigma_{t+1}^{yy} + \Sigma_{t+1}^y]^{-1}$

i : Ensemble member

i^- : Forecasted

i^+ : Updated

θ : Model parameters

y : Process measurements

x : Process states

K : Kalman gain

EnKF for parameter estimation

- State ensemble prediction with updated parameters $x_{t+1}^{i-} = f(x_t^{i+}, u_t^i, \theta_{t+1}^{i+})$
- Update state ensemble members $x_{t+1}^{i+} = x_{t+1}^{i-} + K_{t+1}^x(y_{t+1}^i - \hat{y}_{t+1}^i)$

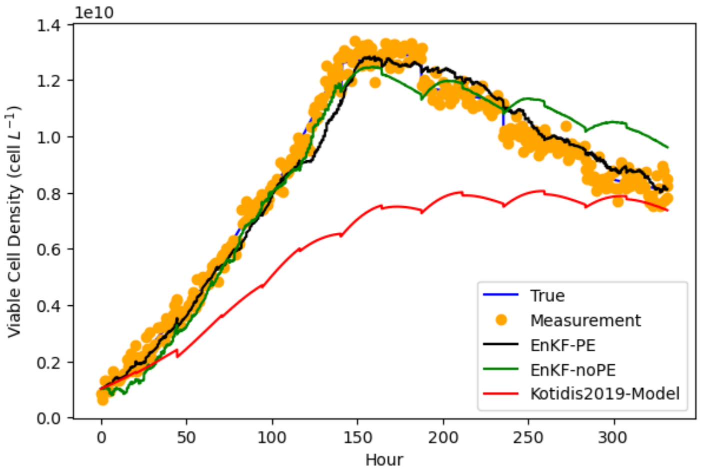
where Kalman gain matrix for state $K_{t+1}^x = \Sigma_{t+1}^{xy}[\Sigma_{t+1}^{yy} + \Sigma_{t+1}^y]^{-1}$

Standard EnKF

Moradkhani *et al*. Dual state–parameter estimation of hydrological models using ensemble Kalman filter; Advances in Water Resources 2005

Cell growth and death

$$\frac{d(VX_v)}{dt} = (\mu - \mu_{deat}\ )VX_v - F_{out}X_v$$

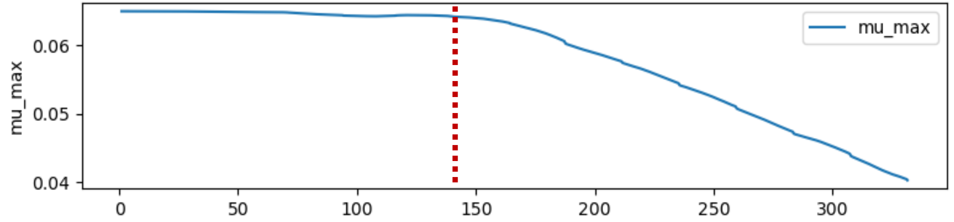


Temperature Downshift

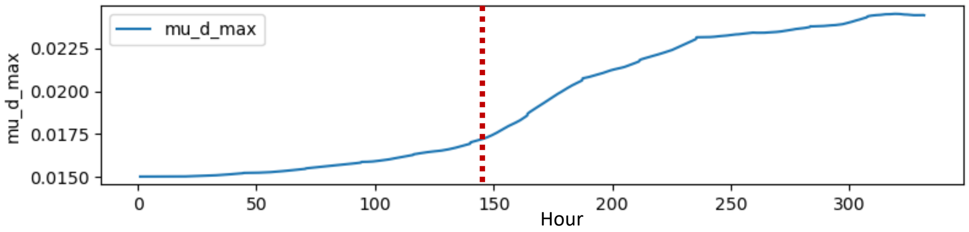
$$\mu = \boxed{\mu_{max}} f_{\lim} f_{inh}$$

$$\mu_{death} = \boxed{\mu_{death,max}} \left(\frac{[Amm]}{[Amm]+K_{d,Amm}} + \frac{[Urd]}{[Urd]+K_{d,Urd}} \right)$$

Growth



Death

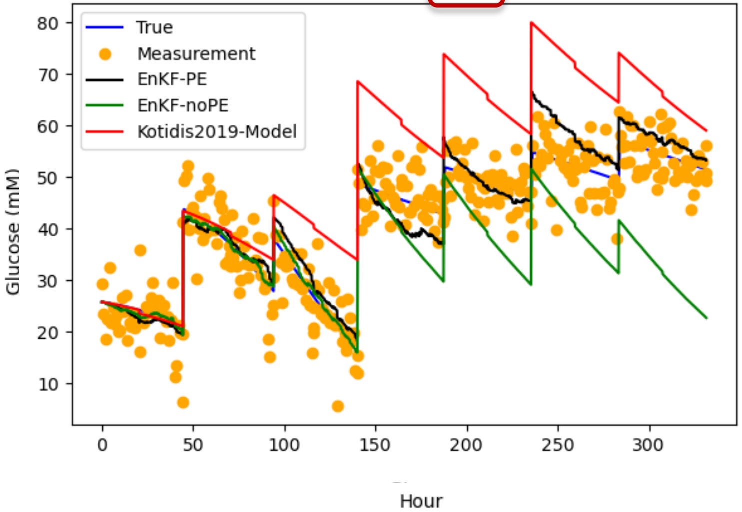


Glucose: main nutrient for cell growth

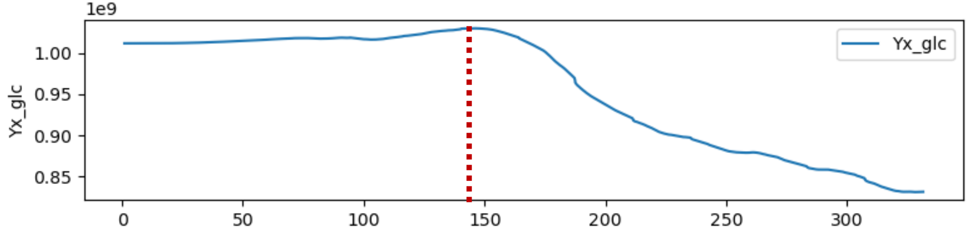
Temperature Downshift

$$\frac{d(V[Metabolite])}{dt} = F_{in}[Metabolite_{feed}] - F_{out}[Metabolite] + q_{Glc} V X_v$$

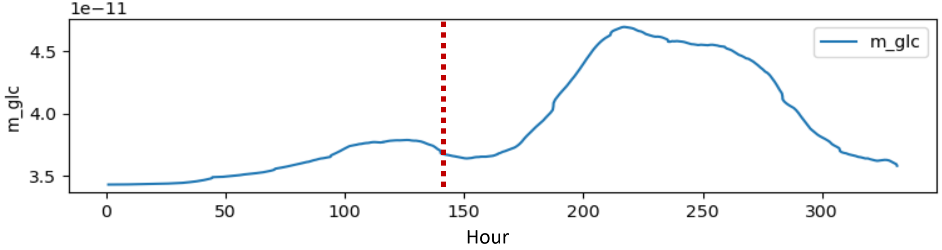
$$q_{Glc} = \left(-\frac{\mu}{\boxed{Y_{X_{Glc}}}} - \boxed{m_{Glc}}\right)$$



Yield of biomass from glucose



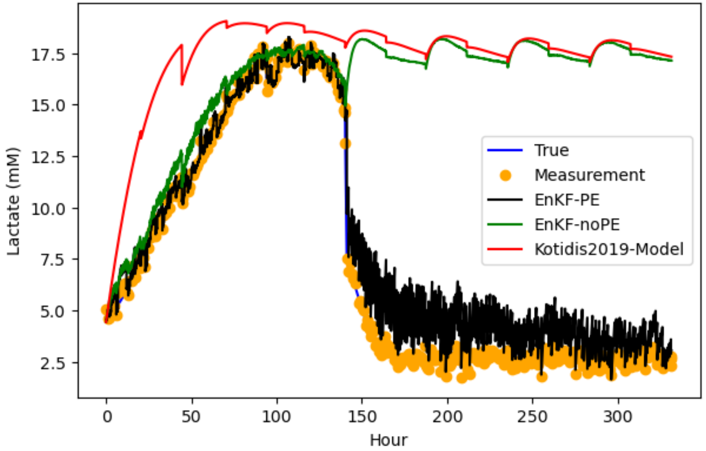
Consumption towards other pathways



Lactate: by product of glucose metabolism, contribute to cell death

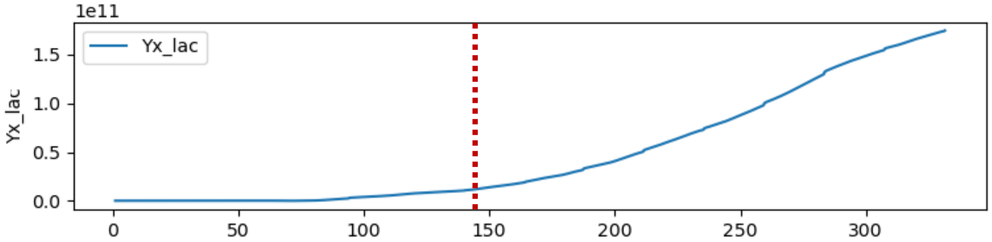
$$\frac{d(V[Metabolite])}{dt} = F_{in}[Metabolite_{feed}] - F_{out}[Metabolite] + q_{Lac} V X_v$$

$$q_{Lac} = \left(\frac{\mu}{\boxed{Y_{X_{Lac}}}} - \boxed{Y_{Lac/Glc}} q_{Glc}\right)\frac{(Lac_{max1} - [Lac])}{Lac_{max1}} + m_{lac}\frac{Lac_{max2} - [Lac]}{Lac_{max2}}$$

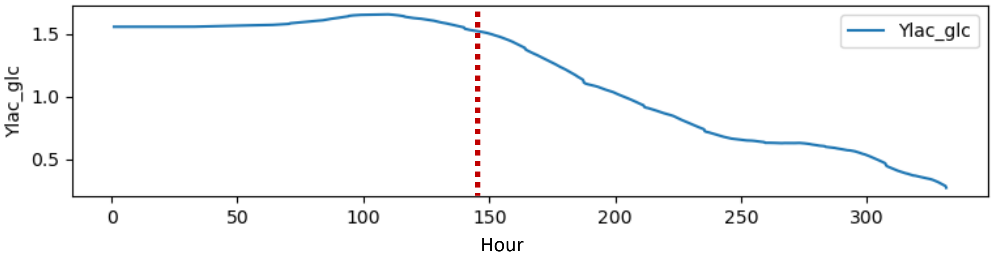


Temperature Downshift

Yield of biomass from lactate

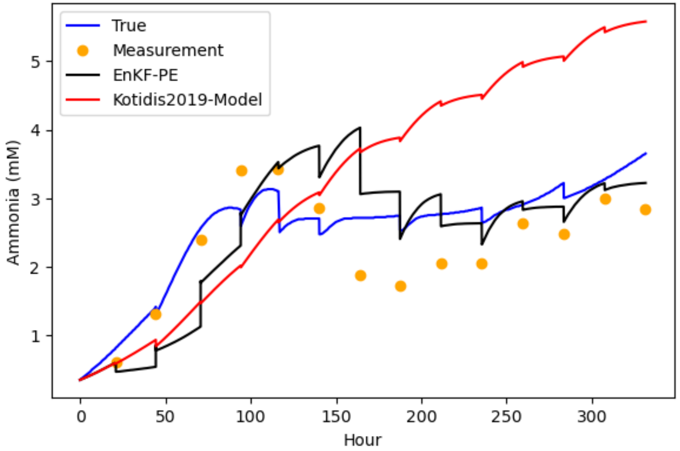


Yield of lactate from glucose

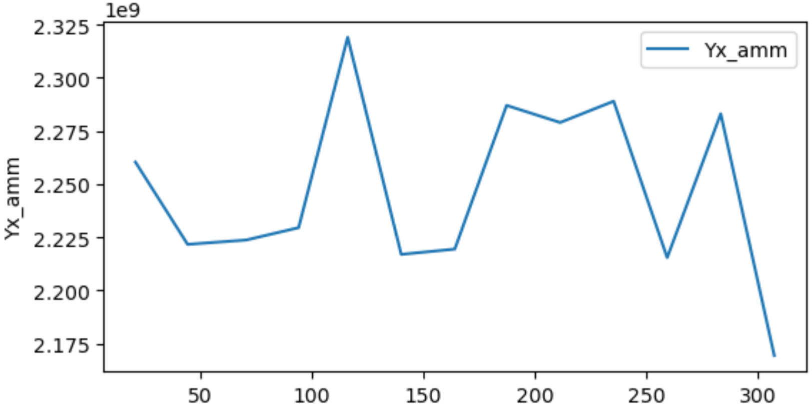


Imperial College London

Question 1: Real process data: sparse measurement

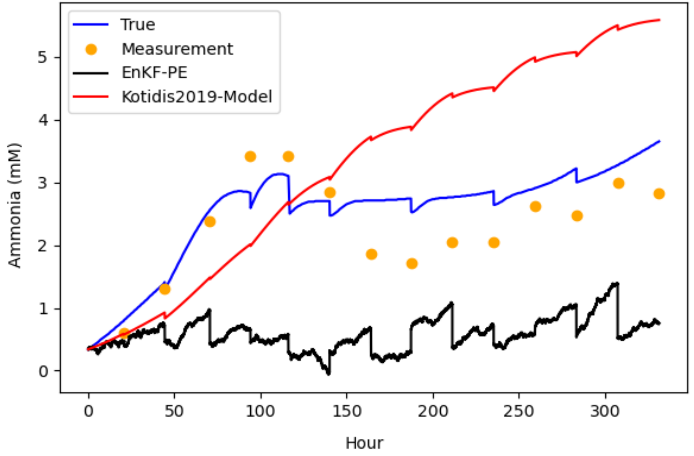


Made a mistake in code but 'worked' : at each step the current ensemble is discarded, new ensemble created based on current mean value



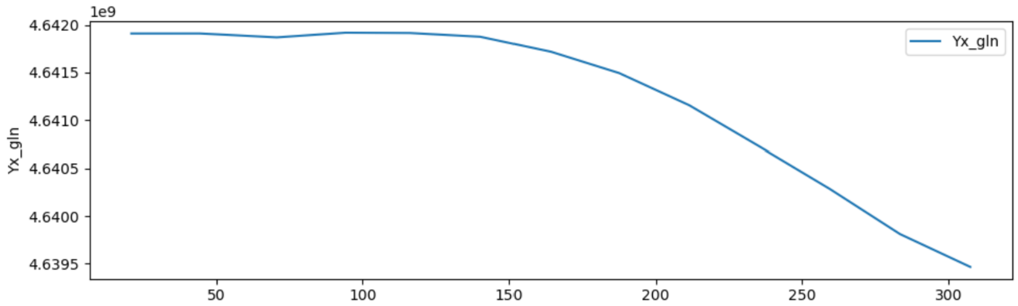
Is this parameter evolution purely based on noise?

Question 1(continued): Real process data, sparse measurement



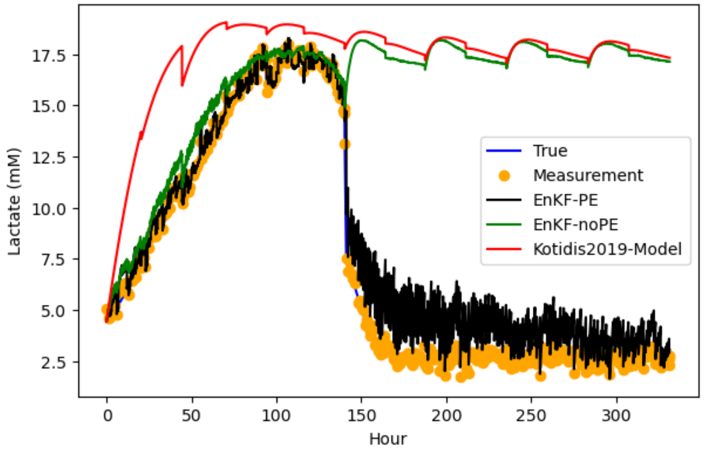
Propagate the ensemble of parameters: every ensemble member follow random walk, does not work.

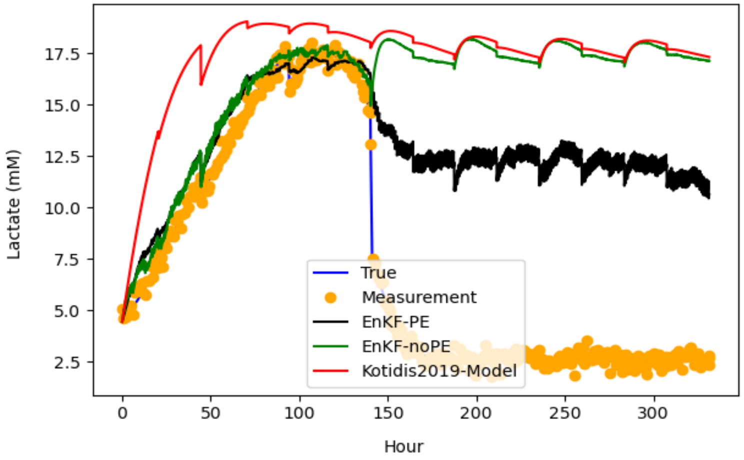
- Is this a problem of tuning only?



- How to assess if PE is needed?
- If PE is not needed, can we use ensemble parameter estimation method for understanding the dynamic variation of system? –parameter no longer fixed
- How to overcome limitation of sparse measurement?

Question 2: For very different profile, how to smooth the curve when EnKF heavily depends on observations?





Question 3: How to assess which parameters require reparameterization and a systematic way for tuning? At least a good starting point?

Thank you very much for the help!