IMPERIAL

Advancing Bioprocessing with EnKF: From State Estimation to Knowledge Transfer

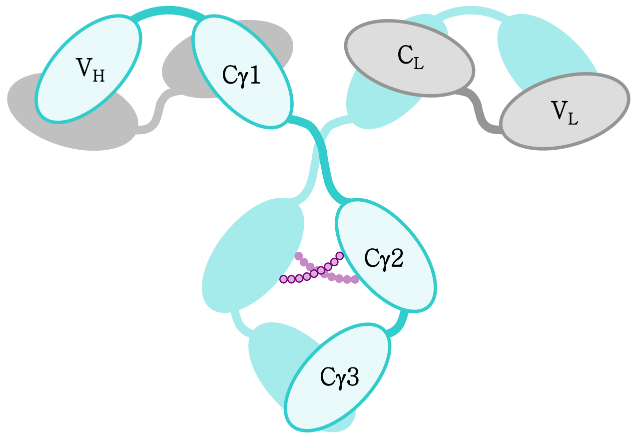
Luxi Yu [a] ; Antonio del Rio Chanona [a] ; Cleo Kontoravdi [a]

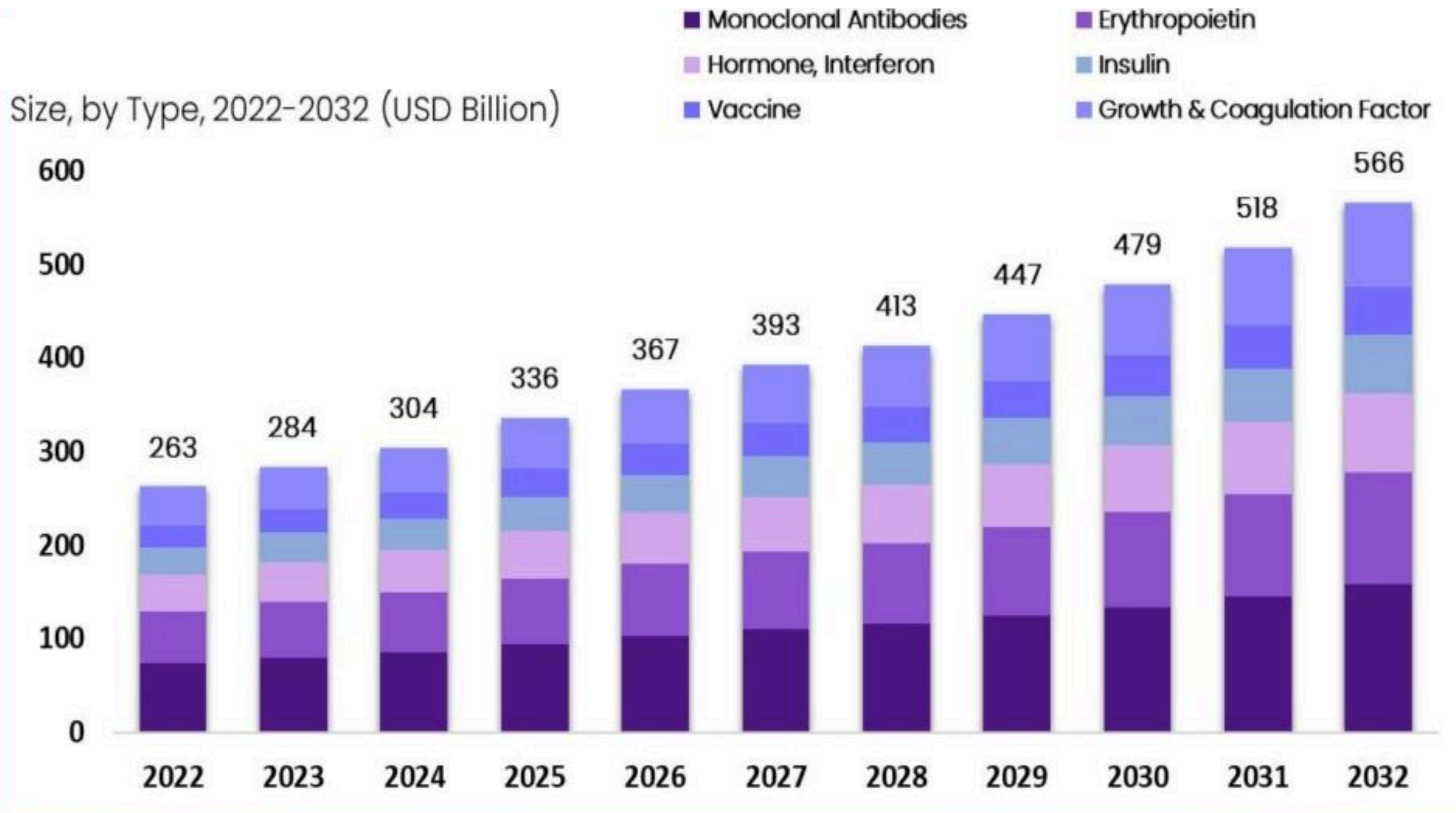
a. Sargent Centre for Process Systems Engineering, Department of Chemical Engineering, Imperial College London, United Kingdom.

Global biopharmaceuticals market

- Dominated by Monoclonal Antibodies (mAb)

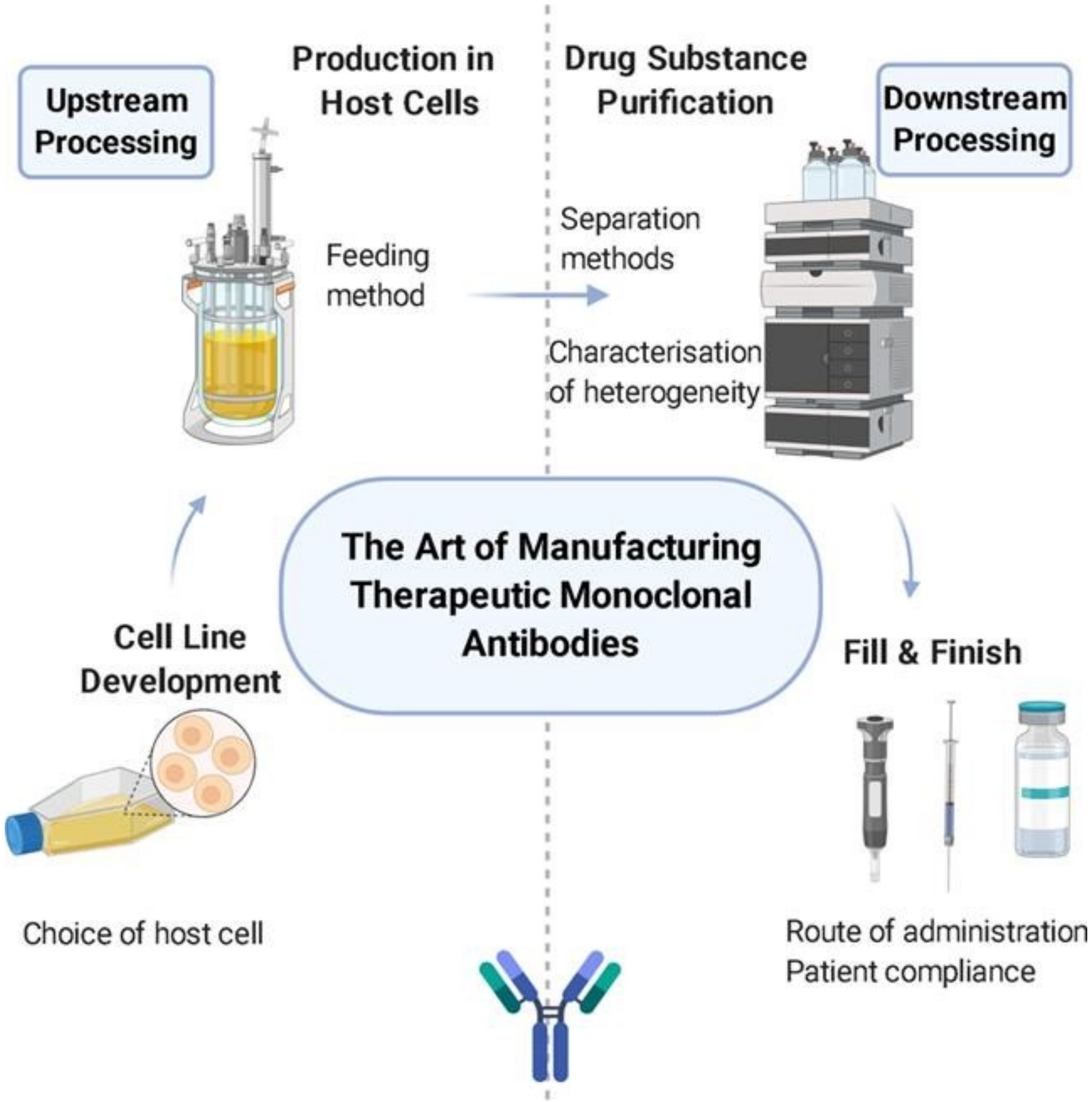
- Highly specific targeting





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.



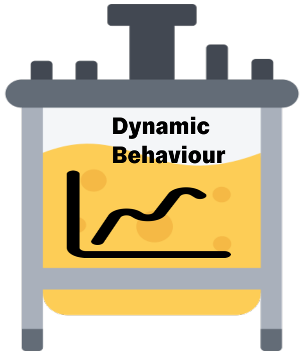
 Carrara, Stefania C. et al, *International Journal of Pharmaceutics* (2021), Volume 594

Mechanistic kinetic modelling for CHO cell process optimization

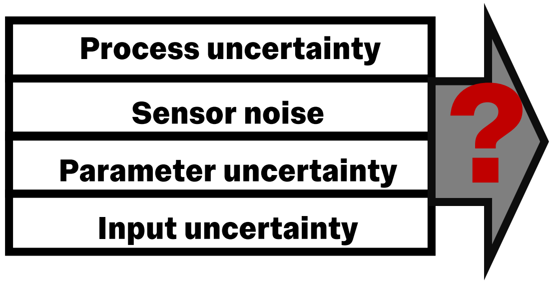
Mechanistic kinetic modeling is a mathematical tool that is derived from first-principles in biological systems.

- Provides insights into cell growth, death, and metabolism.
- Enables accurate predictions for bioprocess optimization.
- Improves productivity and product quality.

Challenges in applying mechanistic models in cell cultures

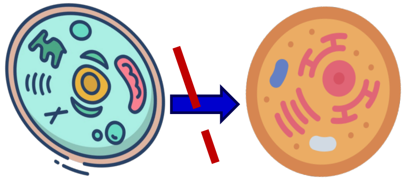


Fixed model parameters can't adapt to dynamic changes



Does not account for various sources of uncertainties explicitly

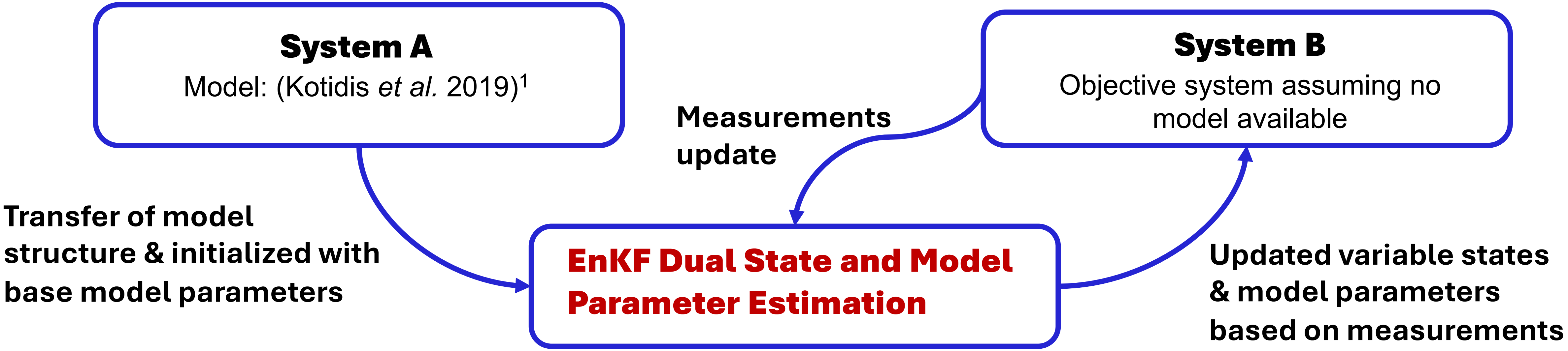
Difficult to handle process unreliability and batch-to-batch inconsistency



Hard to achieve knowledge transfer across cell lines, scales etc.

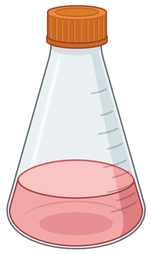
An adaptative approach : Transferring knowledge across systems with EnKF

- The EnKF can estimate system states and model parameters **for a new System B using a single dataset**, based on an existing model initially designed for **System A**.
- EnKF allows dynamic updates of states and model parameters, also explicitly representing uncertainty.



Kotidis, P. *et al.* *Biotechnol Bioeng* (2019) 116:1612–1626

What are our System A and System B?



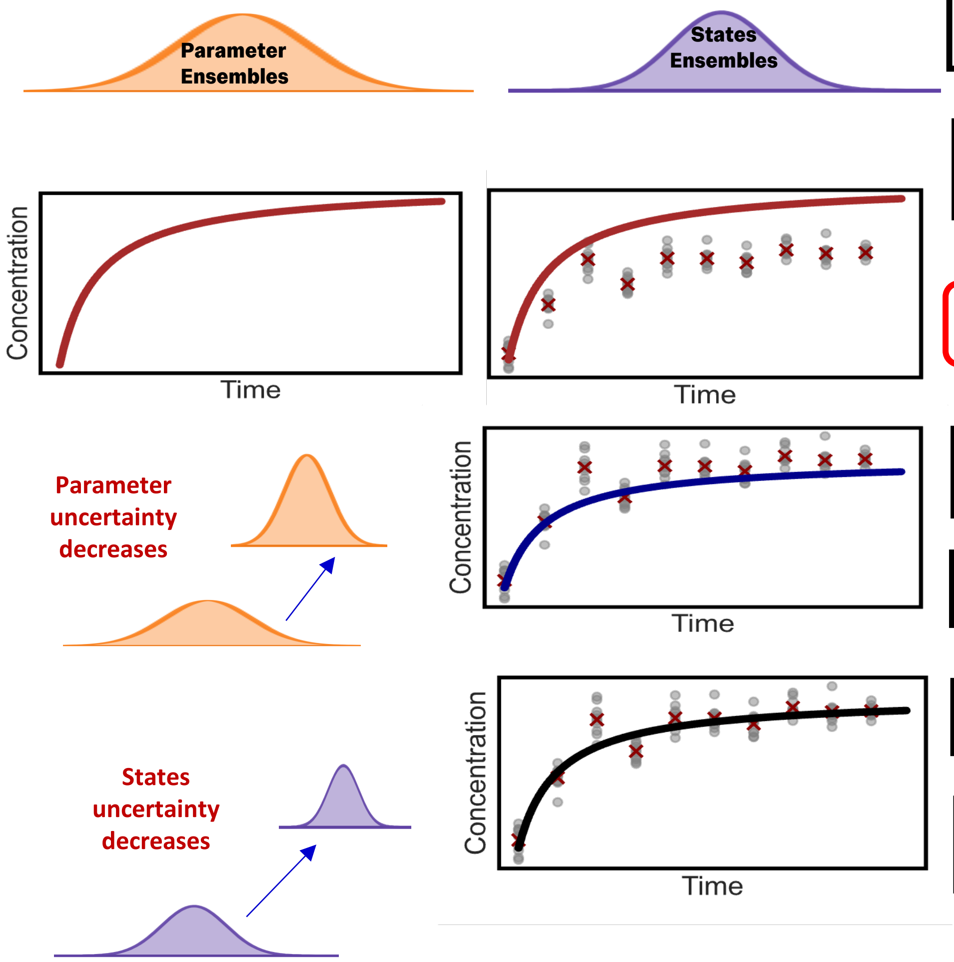
Shake flask
Cell line A
Product: IgG 1
Temperature: 36.5 °C

	What is kept the same ?	**What is different ?**
Dataset 1:	Cell line A, product, temperature, feed	Scale
Dataset 2:	Cell line A, product, feed	Scale, temperature
Dataset 3:	Scale, feed, temperature	Cell line B, product
Dataset 4:	Scale, temperature	Cell line B, product, feed
Dataset 5:	Scale, temperature	Cell line B, product, feed

System A:
Model available

System B:
No model available

EnKF Workflow



Generate state and parameter distributions to account for uncertainty and system differences

Predict model states with preliminary mechanistic model $x_{t+1}^{i-} = f\left(x_t^{i+}, \theta_{t+1}^{i-}\right) + \varepsilon_t^i\,, \quad \varepsilon_t^i \sim N(0, \Sigma_t^x)$

Measurements available?

No → **Use the same model parameters at the next time point** → back to predicting model states

Yes ↓

Generate measurements replicates within experimental error bars $z_{t+1}^i \sim N(z_{t+1}, \Sigma_{t+1}^z)$

Update model parameters based on measurements $\theta_{t+1}^{i+} = \theta_{t+1}^{i-} + K_{t+1}^{\theta}\left(z_{t+1}^i - x_{t+1}^{i-}\right)$

Predict model states using the updated parameters $x_{t+1}^{i-} = f\left(x_t^{i+}, \theta_{t+1}^{i+}\right) + \varepsilon_t^i\,, \quad \varepsilon_t^i \sim N(0, \Sigma_t^x)$

Update states estimate based on measurements $x_{t+1}^{i+} = x_{t+1}^{i-} + K_{t+1}^{x}\left(z_{t+1}^i - x_{t+1}^{i-}\right)$

End of Cell Culture?

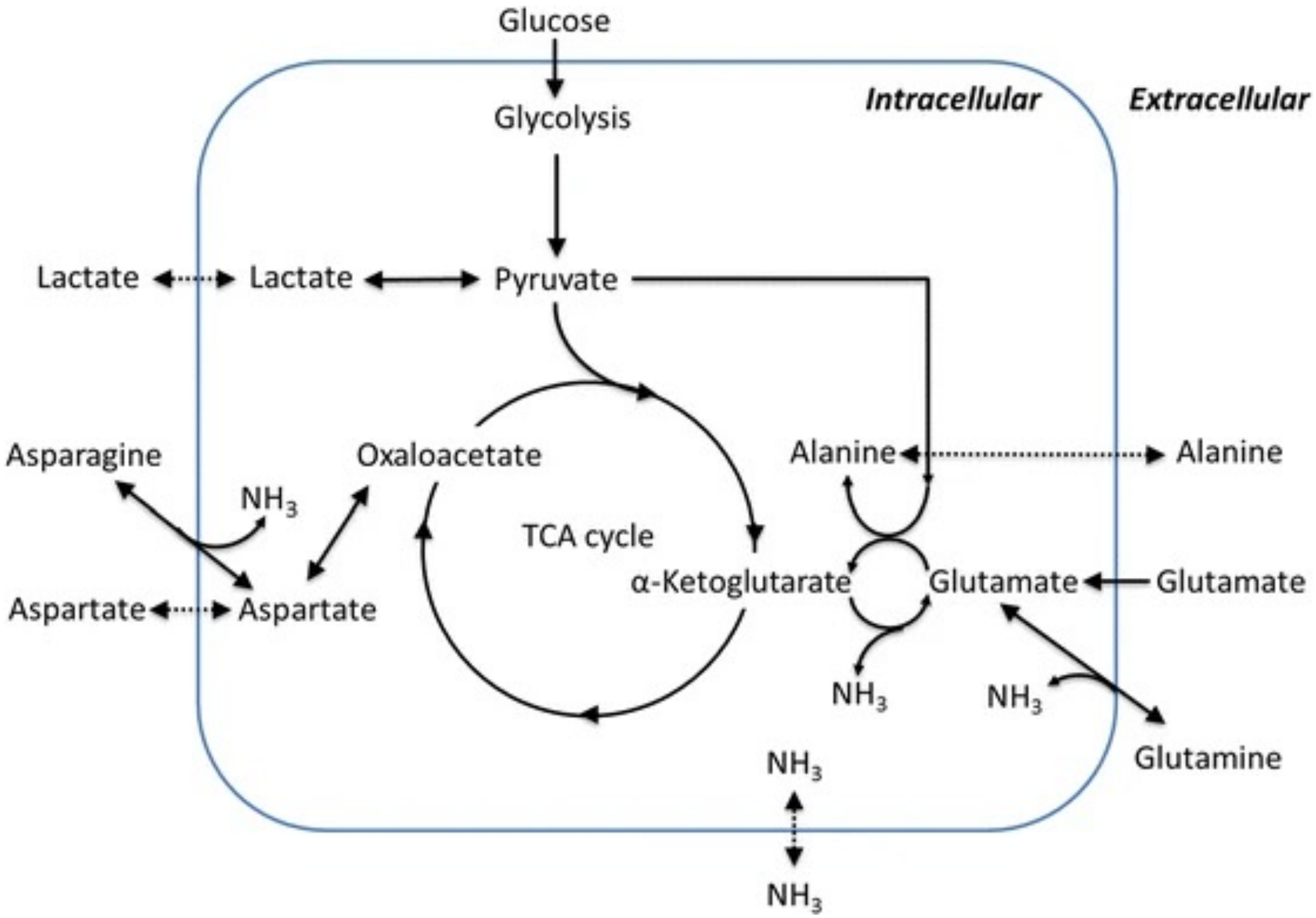
No → $t = t + 1$ → back to predicting model states

Yes ↓

Finish

Case study: Understanding lactate metabolism

- **Byproduct:** CHO cells convert excess glucose into lactate during rapid growth or limited oxygen.
- **Reutilisation:** Under favourable conditions, CHO cells can later consume lactate as an alternative energy source.
- **Process impact:** Controlling lactate dynamics is essential for optimal cell growth, product yield, and quality.



Pan., X *et al. Cytotechnology (2017) 69:39–56*

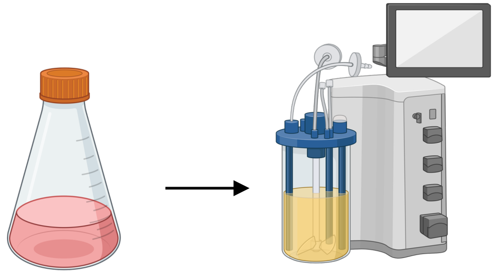
Lactate metabolism in mathematical forms

$$\frac{d(V[Lac])}{dt} = q_{Lac}VX_v - F_{out}[Lac]$$

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

- [Lac] is system state lactate concentration.
- The ODE is the lactate material balance, correlated with rest of the cell culture system through cell density, X_v.
- q_{Lac} is the internal lactate metabolism term, also coupling to the system through cell growth, μ.
- q_{Lac} is also strongly interacting with glucose through $Y_{Lac/Glc}$ and q_{Glc}.
- **Lac_{max1} and Lac_{max2} are lactate consumption activation constants.**

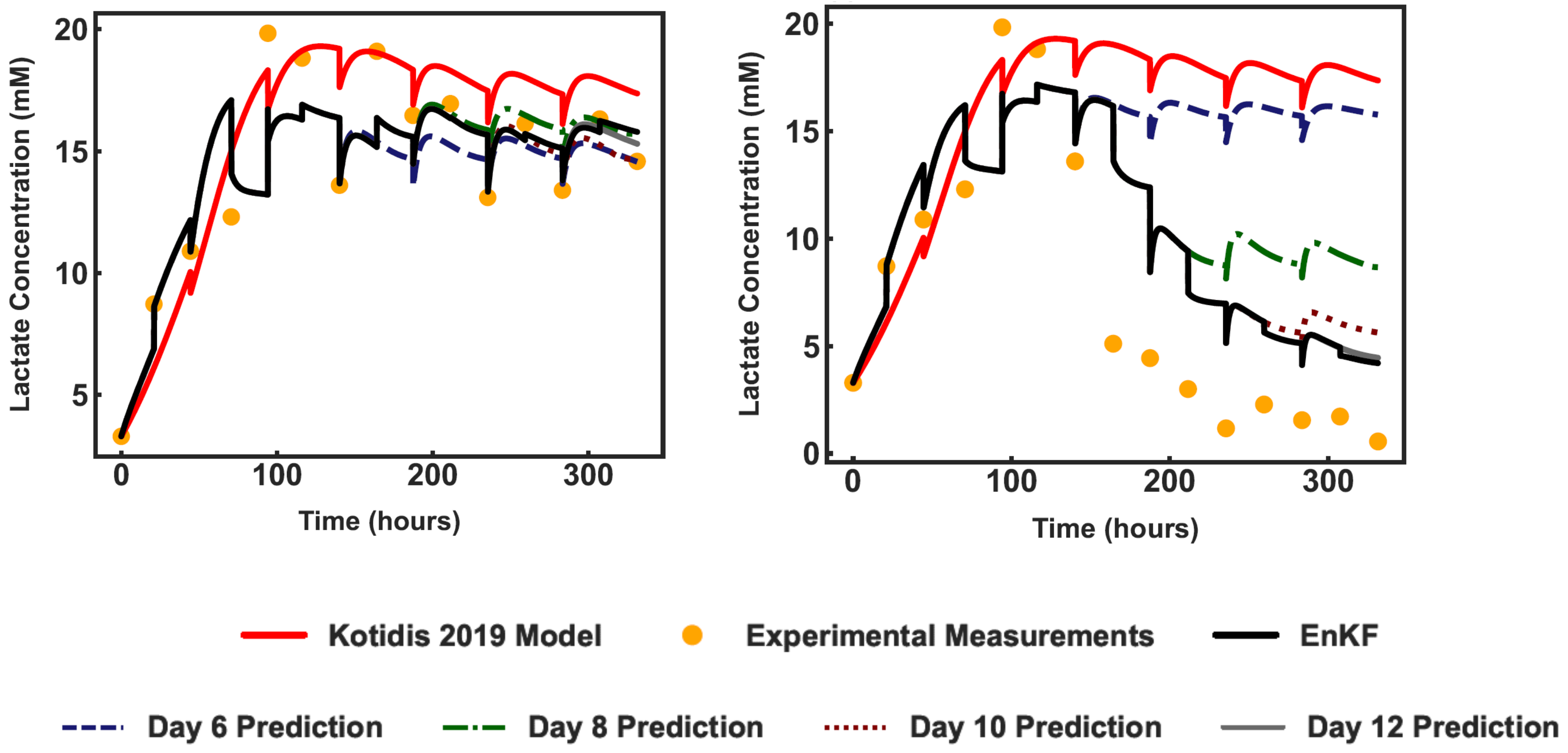
Dataset 1 & 2 - Scaling up from shake flask to bioreactor



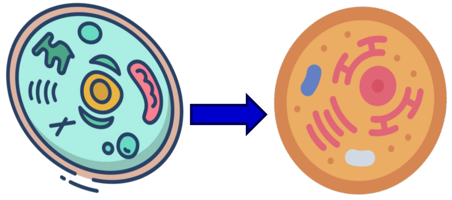
Shake flask Bioreactor

Bioreactor 36.5 °C, feed type 1

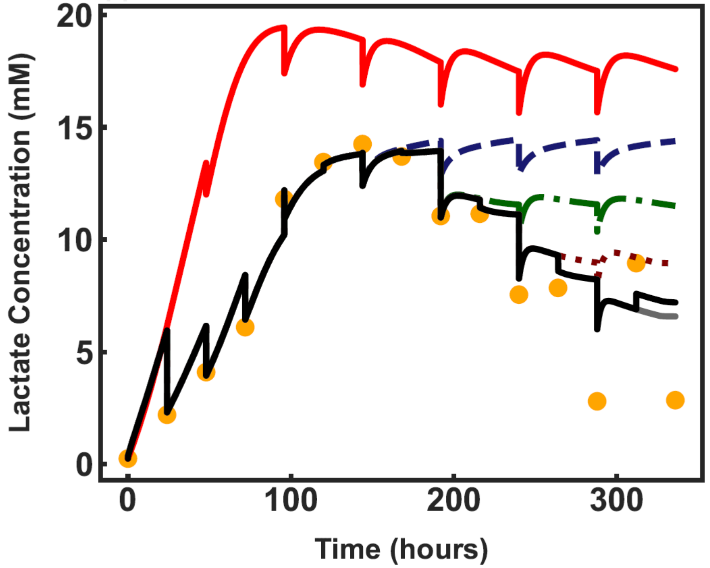
Bioreactor 32 °C, feed type 1



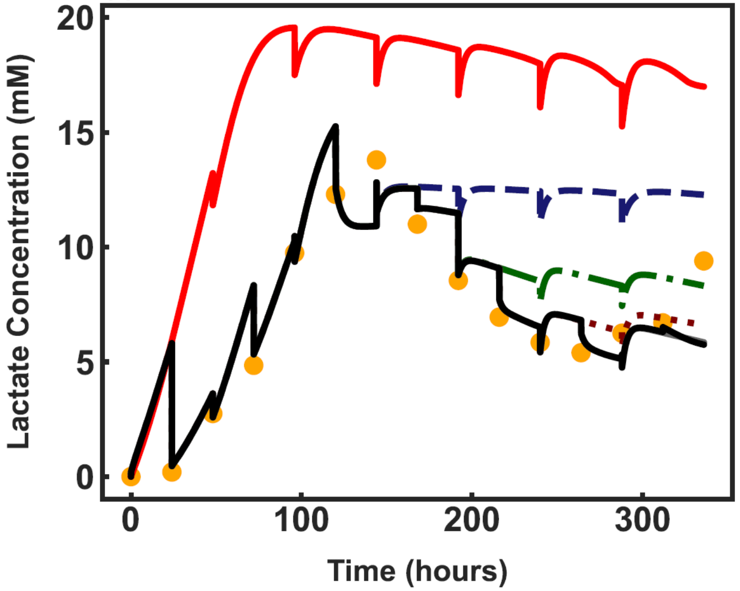
Dataset 3, 4 & 5 – Different cell line and feeds, same scale



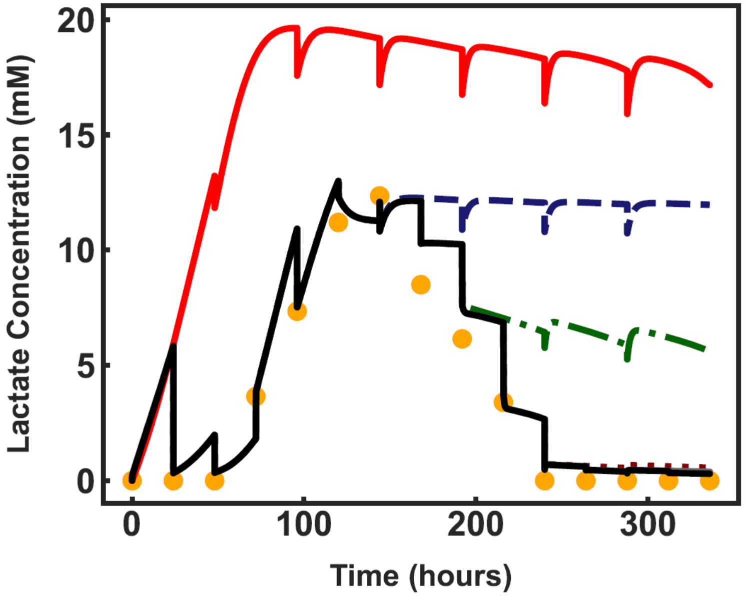
Different cell line, same feed type 1



Different cell line & feed type 2



Different cell line & feed type 3



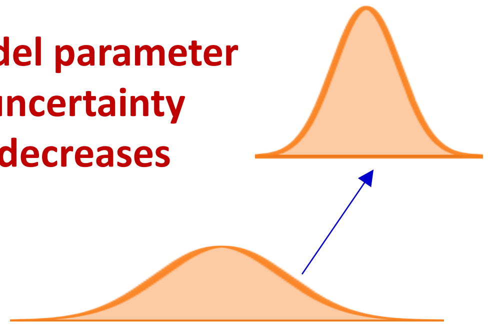
Kotidis 2019 Model — Experimental Measurements — EnKF

Day 6 Prediction — Day 8 Prediction — Day 10 Prediction — Day 12 Prediction

Dynamic Evolution of Model Parameters for Biological Insights: Knowledge Transfer from System A to System B

- Initial parameter are based on model parameters for **System A**, with uncertainty.
- Model estimates become more confident as more measurements are incorporated from the new **System B**, model parameters uncertainty reduce.

Model parameter uncertainty decreases

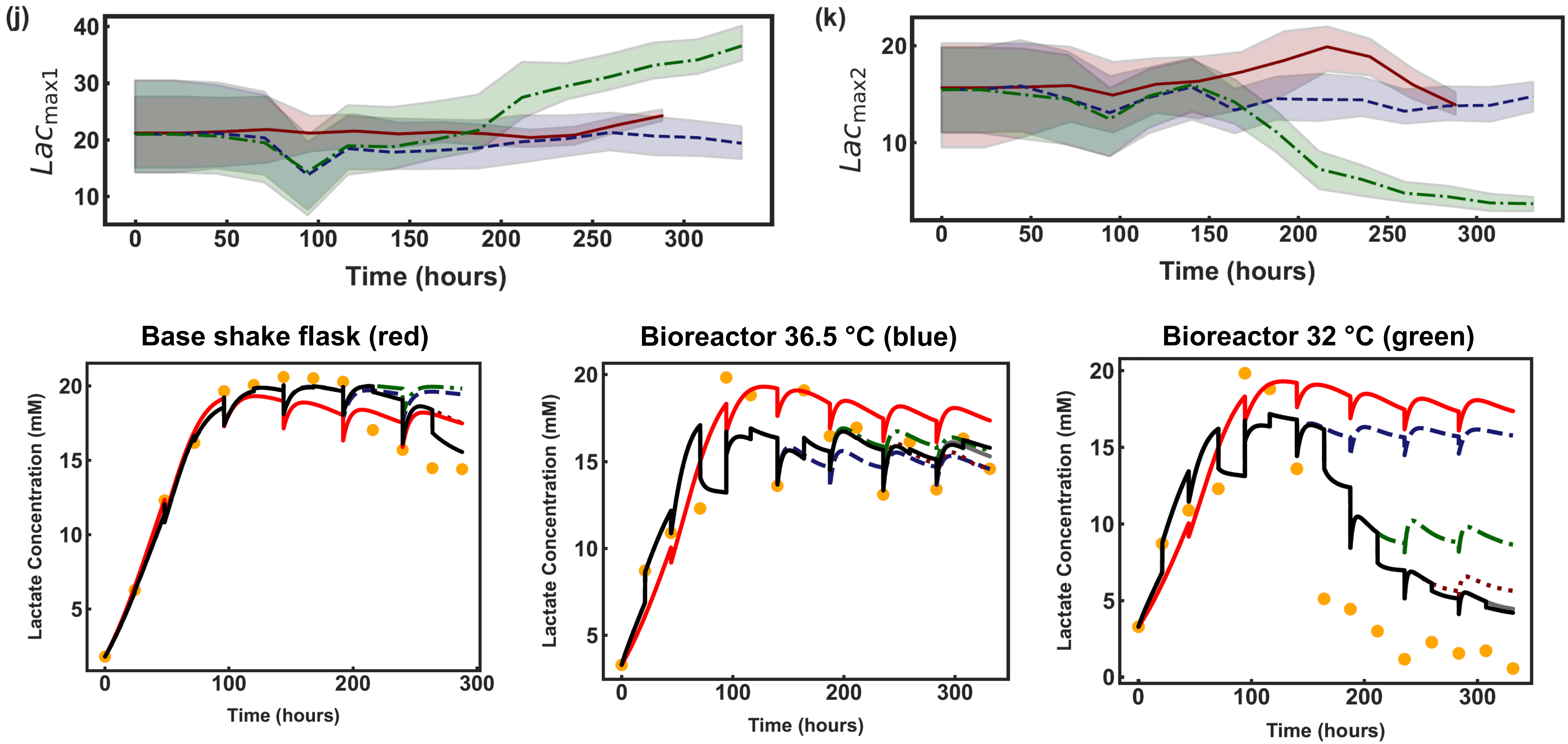


Something very controversial… No parameter covariance inflation is applied.

Why does it work in bioprocessing context?

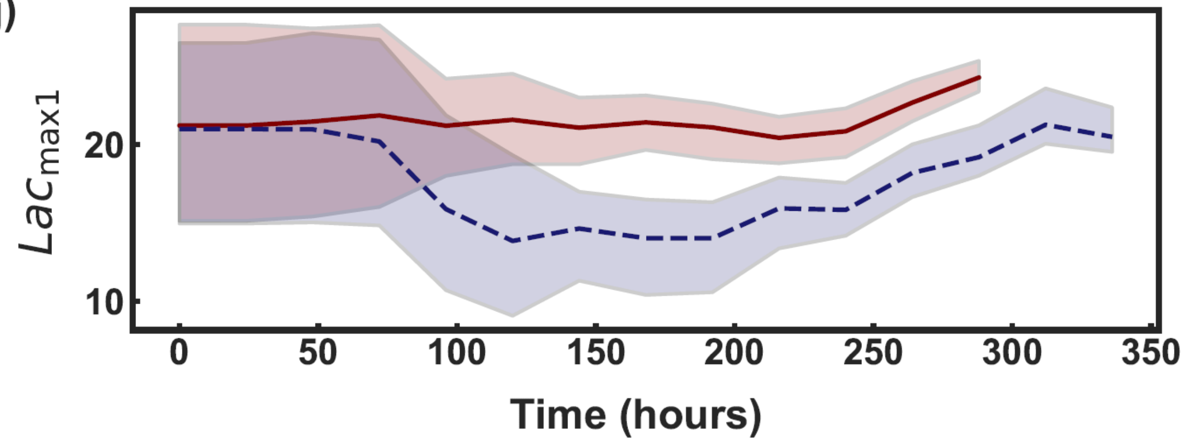
- Very sparse dataset, 12-14 observation updates for entire run.
- Process development stage, biological understanding more important than accuracy.
- Mitigate ensemble collapse by setting large uncertainty spread at the beginning, computational time not a bottleneck due to slow bioprocesses.
- Recursive parameter updates without inflation becomes a dynamic parameter sensitivity analysis for biological understanding.

Effect of scale and temperature on lactate metabolism through parameter ensemble propagation

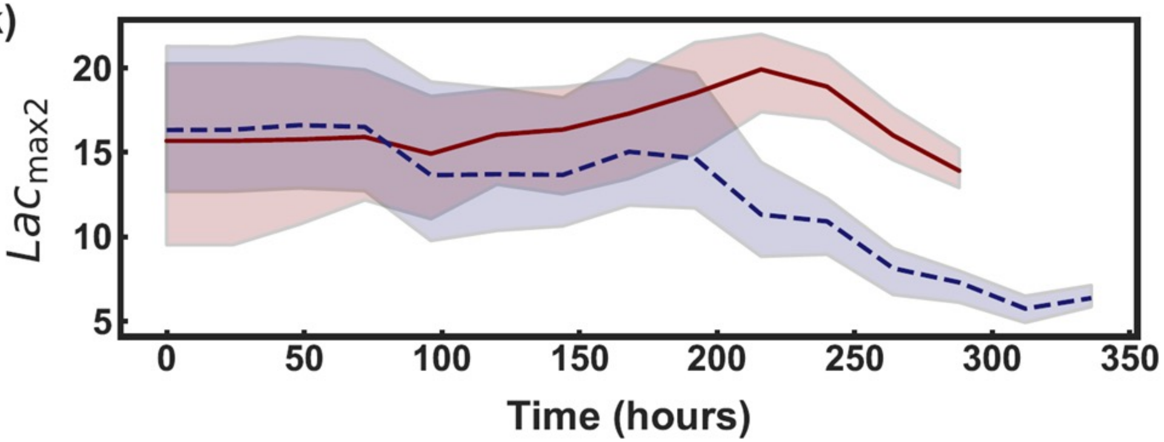


Effect of cell line on lactate metabolism through parameter ensemble propagation

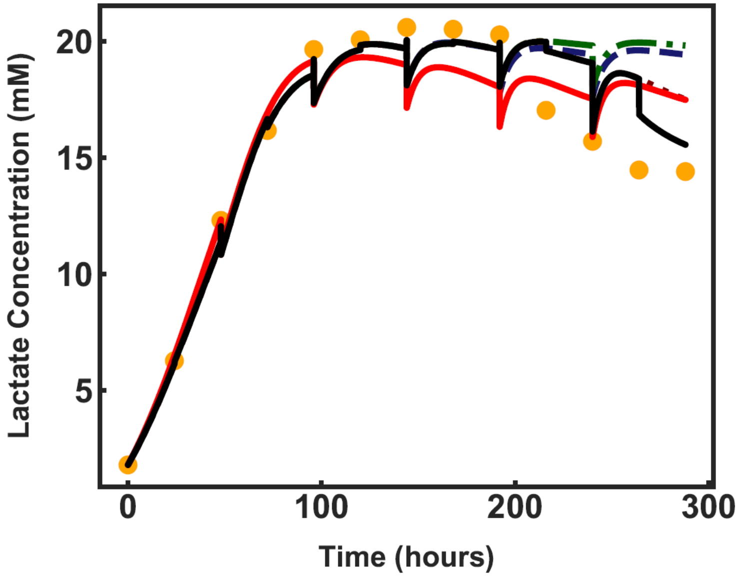
(j)



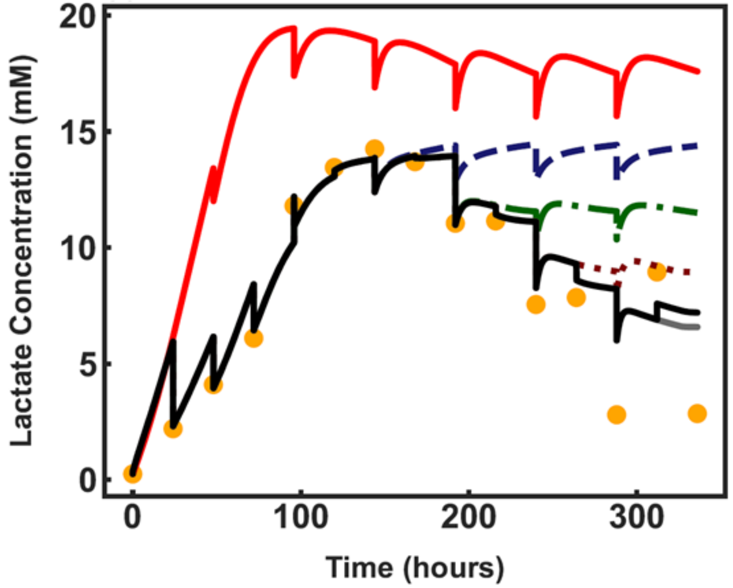
(k)



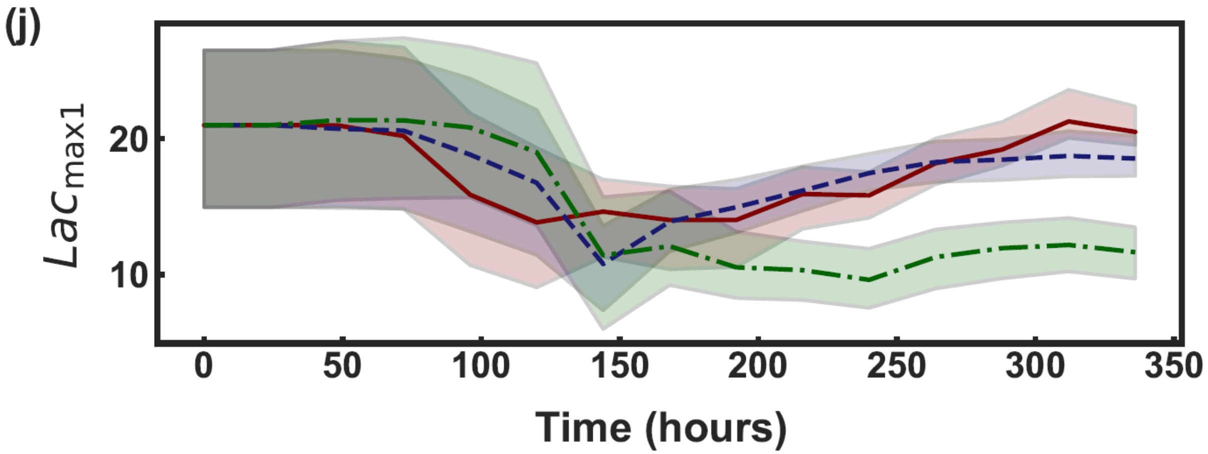
Cell line A (red)

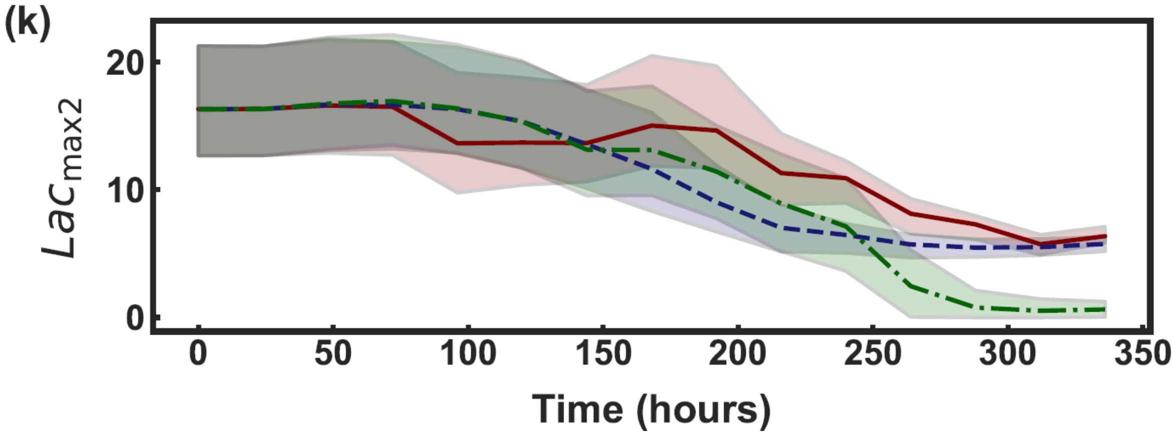


Cell line B (blue)

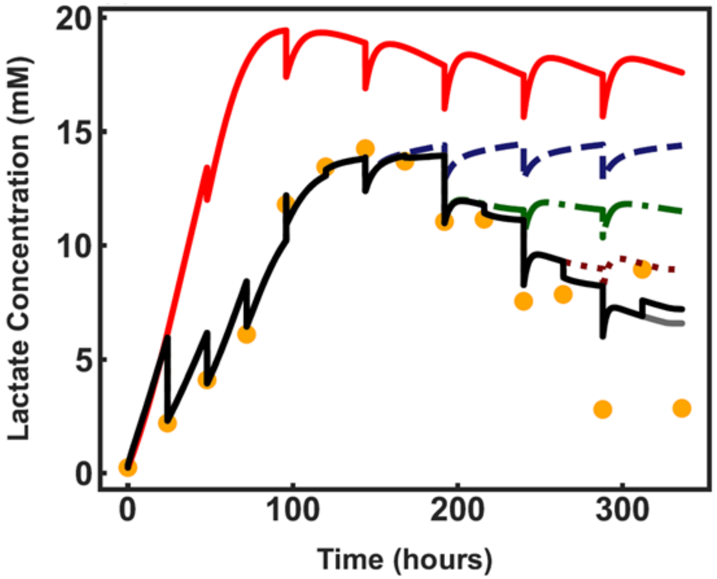


Effect of cell line and feed on lactate metabolism through parameter ensemble propagation

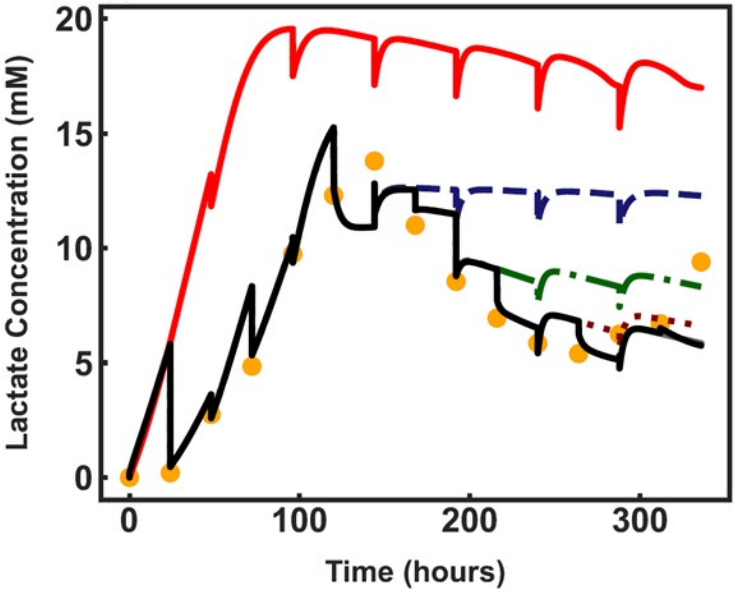




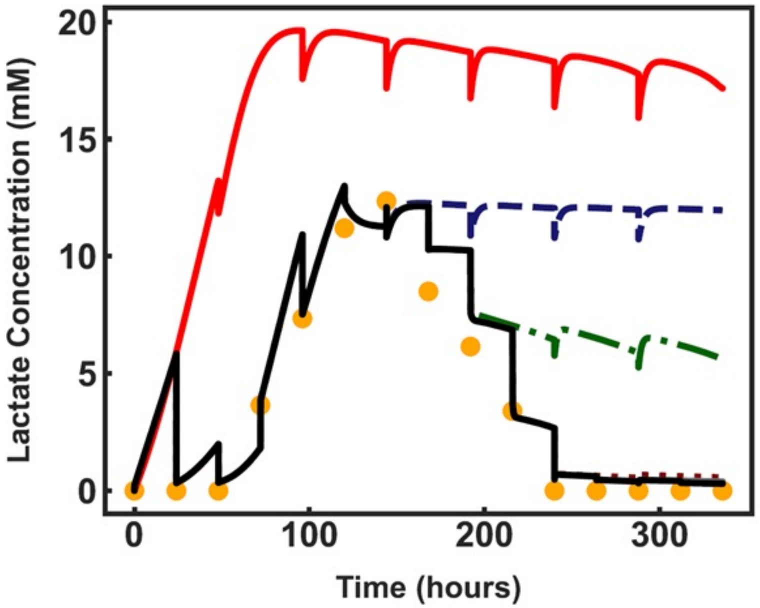
Cell line B, feed type 1 (red)



Cell line B, feed type 2 (blue)



Cell line B, feed type 3 (green)



Key Takeaways

- **Real-Time Adaptation:** EnKF enables model adaptation across different systems using minimal experimental data.
- **Uncertainty Quantification:** EnKF explicitly accounts for uncertainties (e.g., input/output, process variability, sensor noise) to improve prediction reliability.
- **No parameter ensemble inflation:** Changes in the parameter ensemble spread only comes from measurements of the new system, serve as dynamic sensitivity analysis for model parameters.
- **Biological Insights:** System understanding such as metabolic shifts through natural parameter ensemble propagation.
- **Future work : parameter covariance inflation if more frequent observations, more accurate prediction for manufacturing settings required.**

IMPERIAL

Thank you

Please ask me (hard) questions to help me prepare for my viva

Scan this for my email and LinkedIn Page.

Also find me at poster session for more details