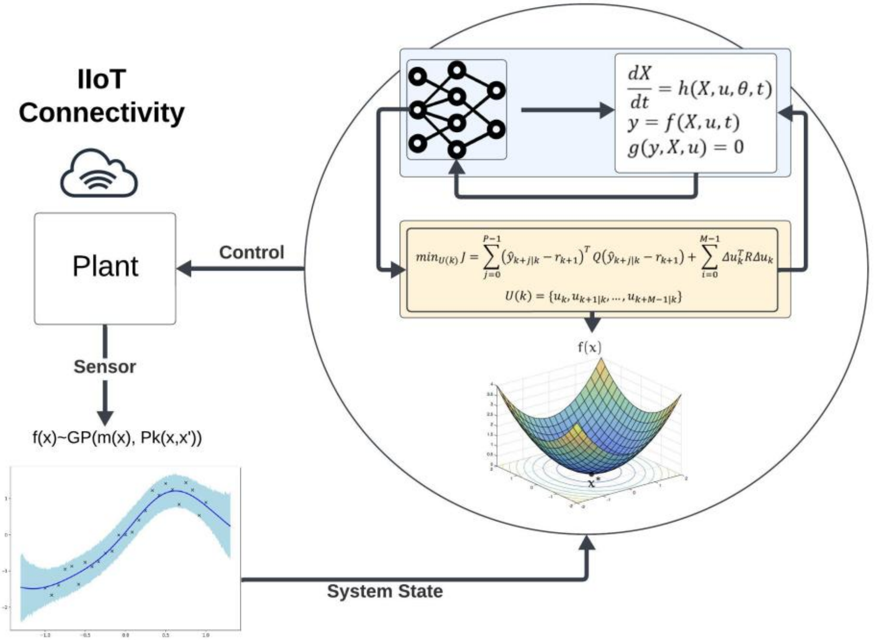
Bioprocessing 4.0: Minimising Cost of mRNA Vaccine Manufacturing



Kesler Isoko [a,b], Mahdi Ahmed [b], Joseph Middleton [b], Joan L. Cordiner [b], Zoltan Kis [b,c*], Peyman Z. Moghadam [a*]

a Department of Chemical Engineering, University College London, Torrington Place, London WC1E 7JE, United Kingdom

b School of Chemical, Materials and Biological Engineering, University of Sheffield, Mappin Street, Sheffield S1 3JD, UK

c Department of Chemical Engineering, Imperial College London, South Kensington Campus, London SW7 2AZ, UK

uceckis@ucl.ac.uk

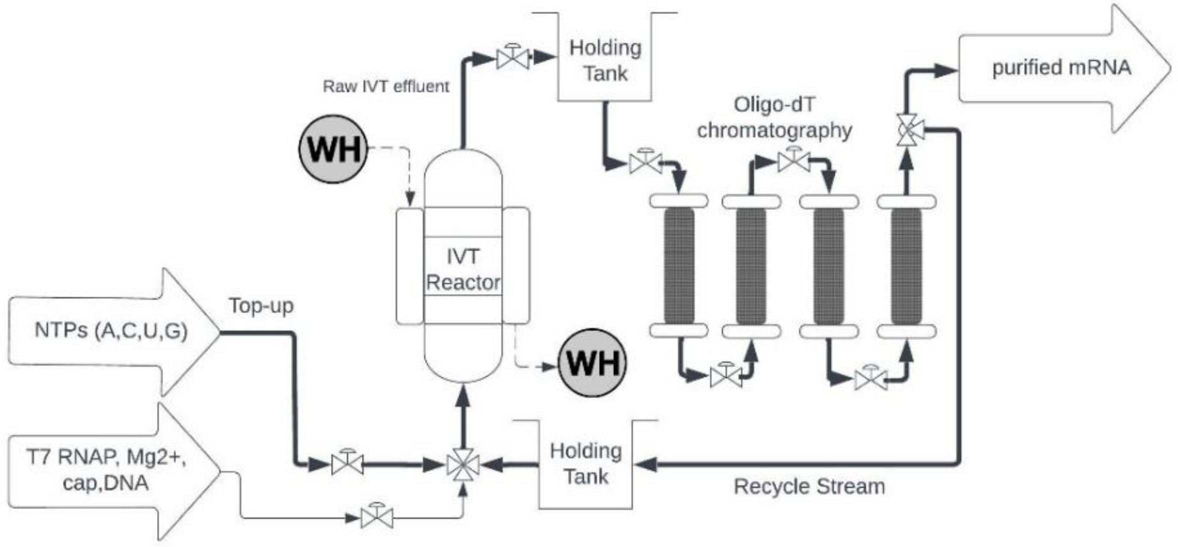
The Goal: Recycling Expensive Raw Materials

Opportunity

- Recycling expensive reagents could significantly reduce the cost of mRNA (2-5x)

Problem

- However, this process is still very manual, and the high dead-times, low observability make it challenging to control



(Nair et.al., 2025 ~ draft; patent - PCTGB2024/052819) ~ filed

Control Setup in mRNA Recycling

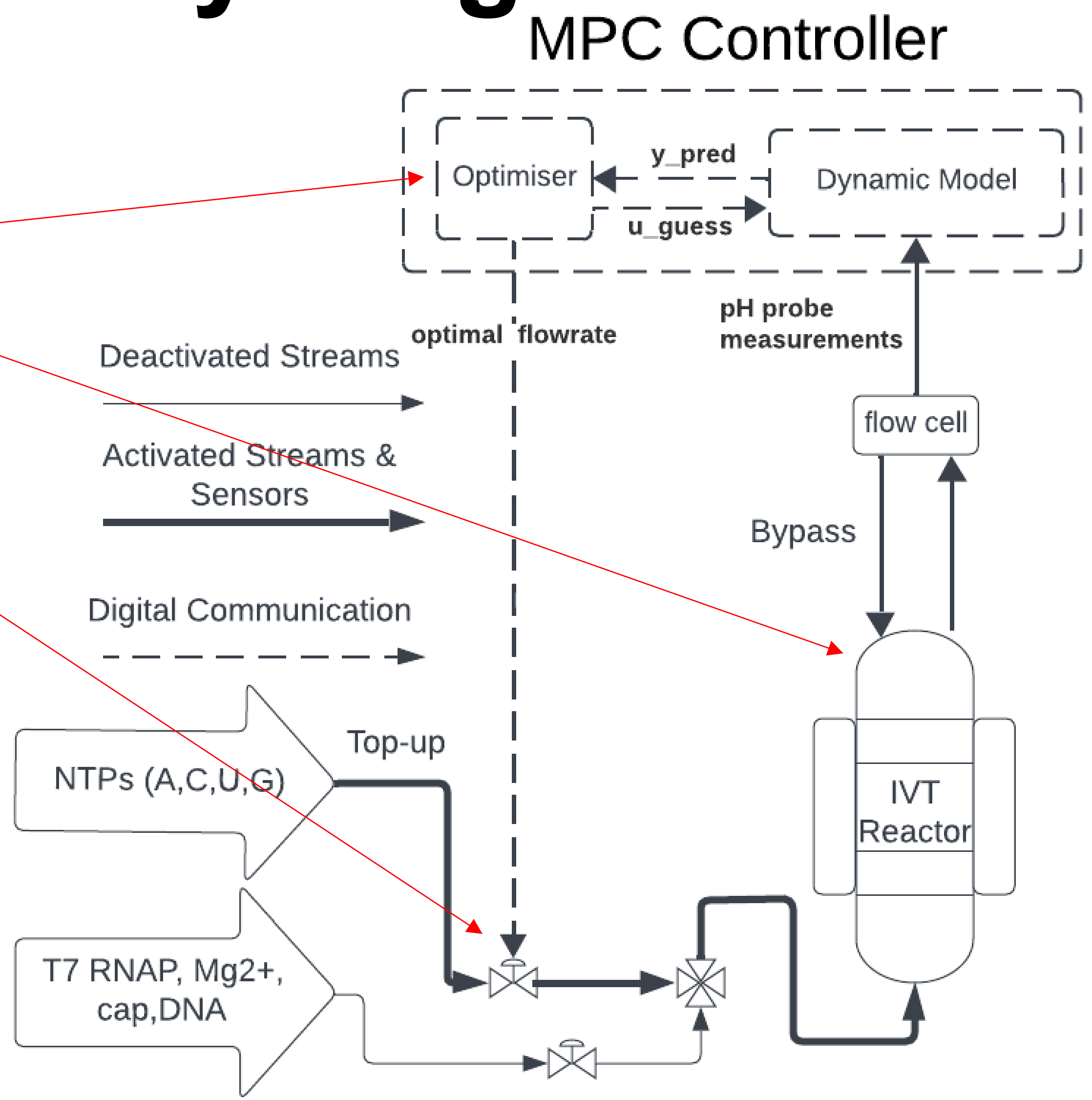
Variable Names	Notation	Example
Prediction Horizon	P	Number of steps the model predicts
Control Horizon	M	Number of control actions evaluated
Control Outputs	Y	RNA yield
Reference Trajectory	r	Desired Yield Response
State Variables	x	Other Components i.e. DNA Template
Control Inputs	U	NTPs Valve Position

Measurements

- pH can be used to monitor the reaction

Objective

- Our goal is to feed NTP such that we maximise yield with minimum side-effects such as pyrophosphate ions production



Control Setup: Dynamic Model

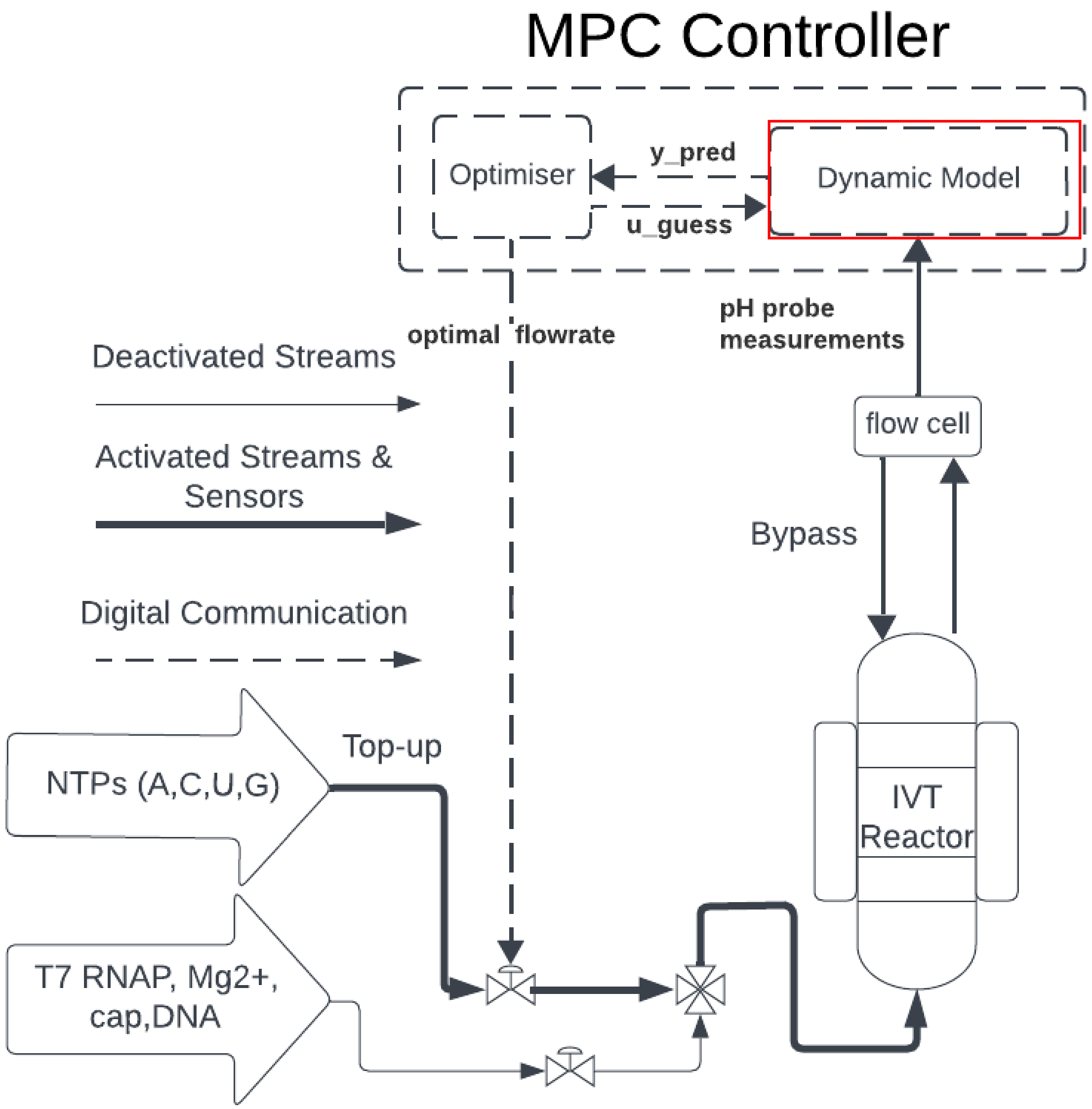
Variable Names	Notation	Example
Prediction Horizon	P	Number of steps the model predicts
Control Horizon	M	Number of control actions evaluated
Control Outputs	Y	RNA yield
Reference Trajectory	r	Desired Yield Response
State Variables	x	Other Components i.e. DNA Template
Control Inputs	U	NTPs Valve Position

Measurements

- pH can be used to monitor the reaction

Objective

- Our goal is to feed NTP such that we maximise yield with minimum side-effects such as pyrophosphate ions production



Dynamic model development: Data Driven

Purely data-driven (System identification)

- AutoRegressive with eXogenous inputs (ARX)
- Nonlinear AutoRegressive with eXogenous inputs (NARX)
- Neural Networks (RNN, LSTMs, etc…)

Purely Mechanistic (with state estimation)

- Extended Kalman Filters (EKF), Unscented Kalman Filters (UKF), Kalman Filters (KF)
- Process Analytical Technologies (PAT)
- Soft sensors

Hybrid semi-parametric modelling

- Serial
- Parallel

Pros of data-driven

- No first principal knowledge required
- Linear models generates a quadratic optimisation programme (convex) that can be solved reliably online

Cons of data-driven

- ARX Might not capture nonlinear dynamics leading to instability
- Requires enough step response data or access to plant to calibrate dead-time, rise-time and steady states parameters

Control Setup: Dynamic Model

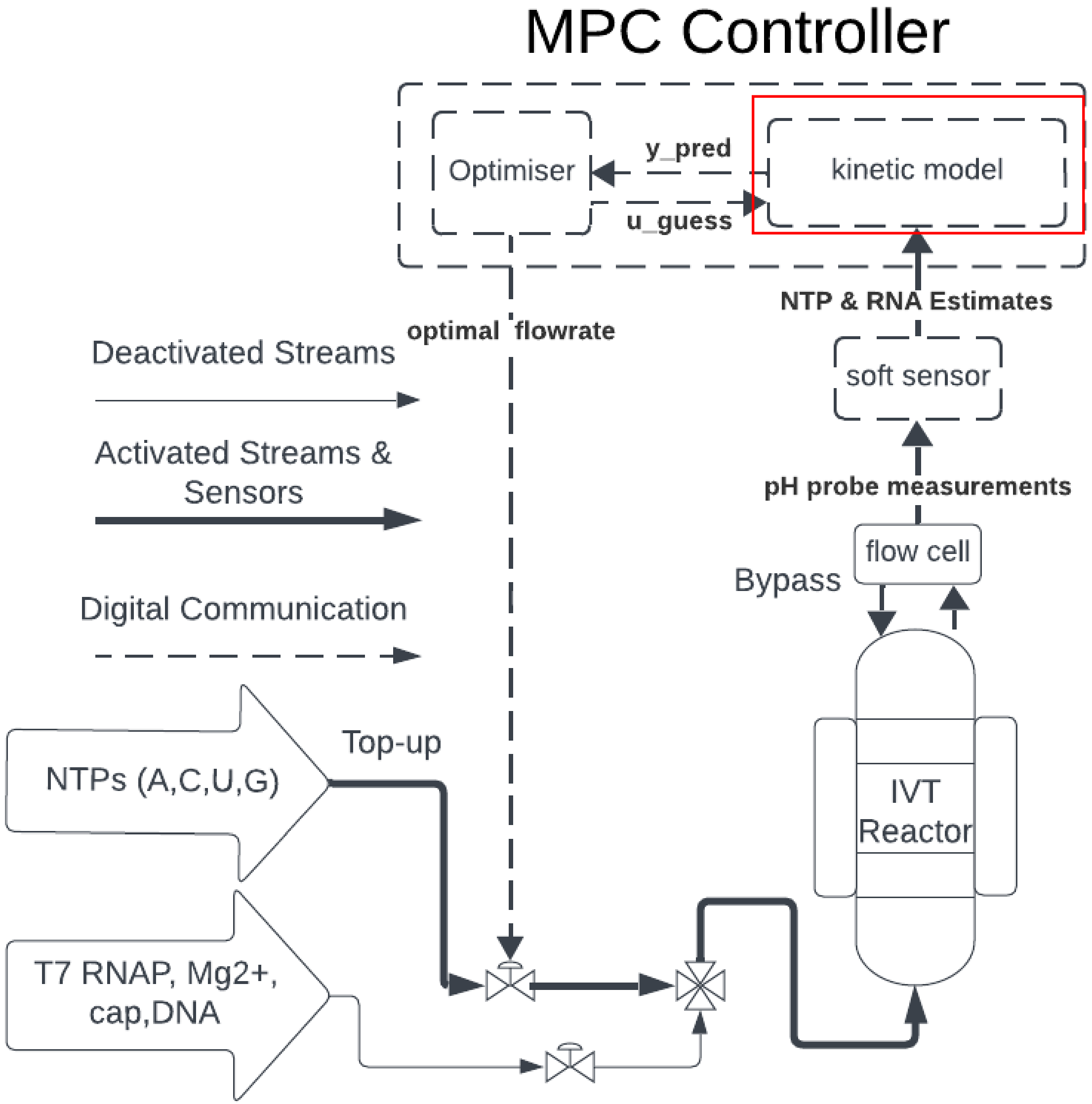
Variable Names	Notation	Example
Prediction Horizon	P	Number of steps the model predicts
Control Horizon	M	Number of control actions evaluated
Control Outputs	Y	RNA yield
Reference Trajectory	r	Desired Yield Response
State Variables	x	Other Components i.e. DNA Template
Control Inputs	U	NTPs Valve Position

Measurements

- pH can be used to monitor the reaction

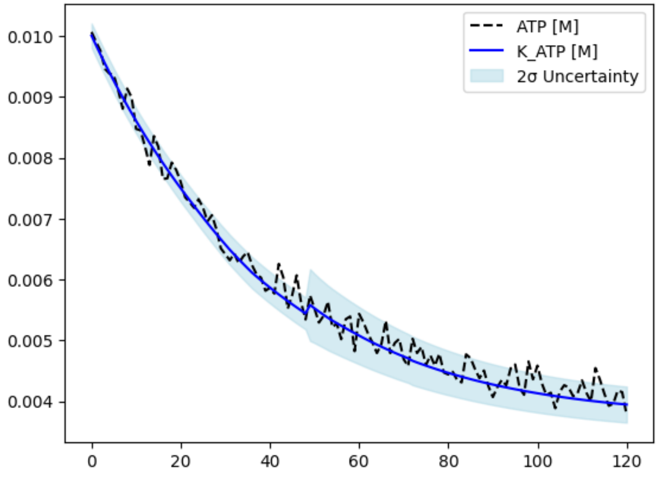
Objective

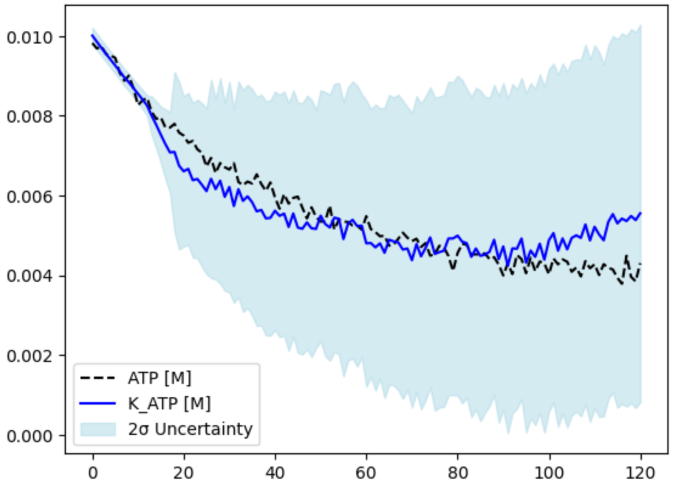
- Our goal is to feed NTP such that we maximise yield with minimum side-effects such as pyrophosphate ions production



Dynamic model development: Mechanistic

- **Purely mechanistic**
- Moving Horizon Estimation
- PAT
- Soft sensor for key variables





- $$\begin{aligned}
\frac{d[\mathrm{RNA}]_{\mathrm{tot}}}{dt} &= V_{\mathrm{tr}} - V_{\mathrm{deg}} - (2 \cdot V_{ds} + V'_{ds}) \\
\frac{d[\mathrm{PPi}]_{\mathrm{tot}}}{dt} &= N_{\mathrm{all}} \cdot V_{\mathrm{tr}} - V_{\mathrm{precip}} - V_{\mathrm{hydrolysis}} \\
\frac{d[\mathrm{ATP}]_{\mathrm{tot}}}{dt} &= -N_{\mathrm{a}} \cdot V_{\mathrm{tr}} \\
\frac{d[\mathrm{GTP}]_{\mathrm{tot}}}{dt} &= -N_{\mathrm{g}} \cdot V_{\mathrm{tr}} \\
\frac{d[\mathrm{CTP}]_{\mathrm{tot}}}{dt} &= -N_{\mathrm{c}} \cdot V_{\mathrm{tr}} \\
\frac{d[\mathrm{UTP}]_{\mathrm{tot}}}{dt} &= -(N_{\mathrm{u}} \cdot V_{\mathrm{tr}} + N_{ds,\mathrm{u}} \cdot V'_{ds}) \\
\frac{d[H]_{\mathrm{tot}}}{dt} &= N_{\mathrm{all}} \cdot V_{\mathrm{tr}} \\
\frac{d[\mathrm{T7RNAP}]_{\mathrm{tot}}}{dt} &= -k_d[\mathrm{T7RNAP}]_{\mathrm{tot}} \\
\frac{d[\mathrm{Mg}]_{\mathrm{tot}}}{dt} &= -2 \cdot V_{\mathrm{precip}} \\
\frac{d[\mathrm{dsRNA}]}{dt} &= 2 \cdot V_{ds} + V'_{ds}
\end{aligned}$$

$$V_{\mathrm{tr}} = k_{\mathrm{app}}[\mathrm{DNAT7RNAPSpd}][\mathrm{Mg}]\frac{[\mathrm{MgNTP}]}{1 + k_1[\mathrm{Mg}] + k_2[\mathrm{MgNTP}]}$$

$$V_{\mathrm{deg}} = (k_{\mathrm{ac}}[H]^{n_{\mathrm{ac}}} + k_{\mathrm{ba}}[OH]^{n_{\mathrm{ba}}} + k_{\mathrm{Mg}}[\mathrm{Mg}]^{n_{\mathrm{Mg}}})[\mathrm{RNA}]^{n_{\mathrm{RNA}}}$$

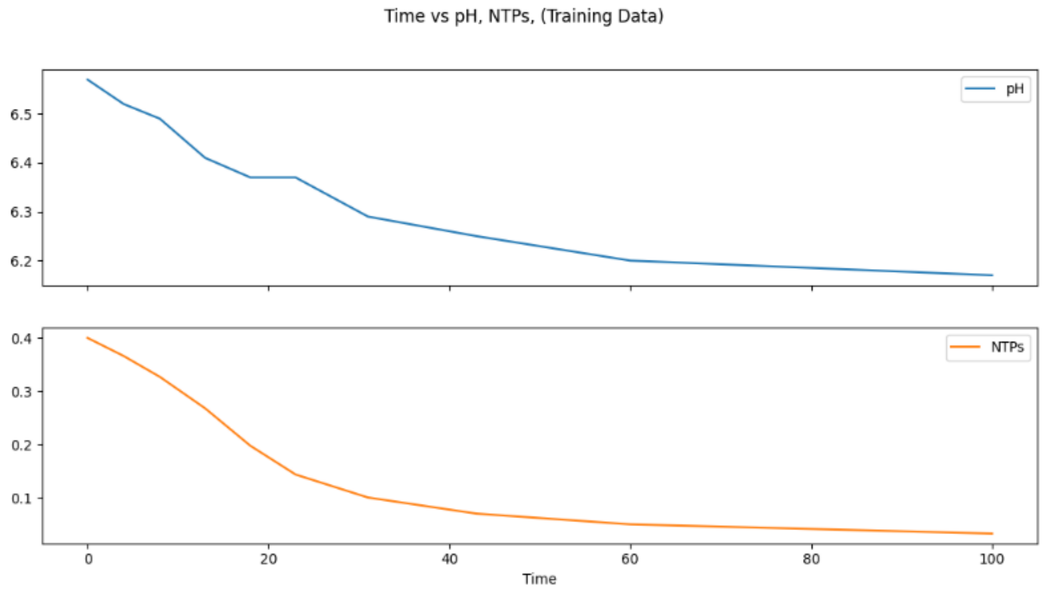
$$N_{\mathrm{all}} = N_{\mathrm{a}} + N_{\mathrm{g}} + N_{\mathrm{c}} + N_{\mathrm{u}}$$

$$V_{\mathrm{precip}} = k_{\mathrm{precip}}([\mathrm{MgPPi}] - [\mathrm{MgPPi}]_{\mathrm{sp}})$$

$$V_{\mathrm{hydrolysis}} = \frac{[PPi] \cdot [PPase]}{k_{\mathrm{PPase}} + [PPi]}$$

- Linearisation techniques lead to poor performance (EKF)
- UKF is accurate however requires proper calibration and is too slow to be integrated in optimisation loop

Developing a PAT sensor



$$pH = pK_a + \log\left(\frac{[A^-]}{[HA]}\right)$$

Pros

- Provides accurate estimates of NTP & RNA

Cons

- Numerical instability of the log in python due to floating point errors
- Sensitive to noisy measurements

Soft sensor development: Physics-informed Gaussian Process Regression

$$f(\mathbf{x}) \sim \mathcal{GP}\left(m_f(\mathbf{x}), k_f(\mathbf{x}, \mathbf{x}')\right)$$

$$\kappa(x_i, x_j) = \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\sigma^2}\right)$$

$$pH = pK_a + \log\left(\frac{[A^-]}{[HA]}\right)$$

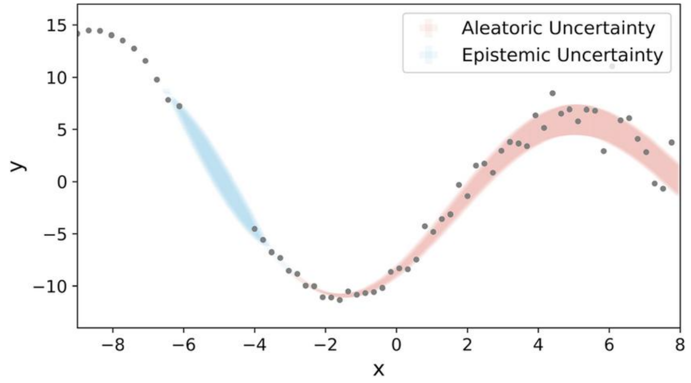
$$k(x_i, x_j) = exp\left(\frac{\left\|\log(1 + |x_i - pK_a|) - \log(1 + |x_j - pK_a|)\right\|^2}{2\sigma^2}\right) + \alpha * \delta_{ij}$$

Requirements

- Keep the physics
- Sampling efficient
- Robust to noisy measurements

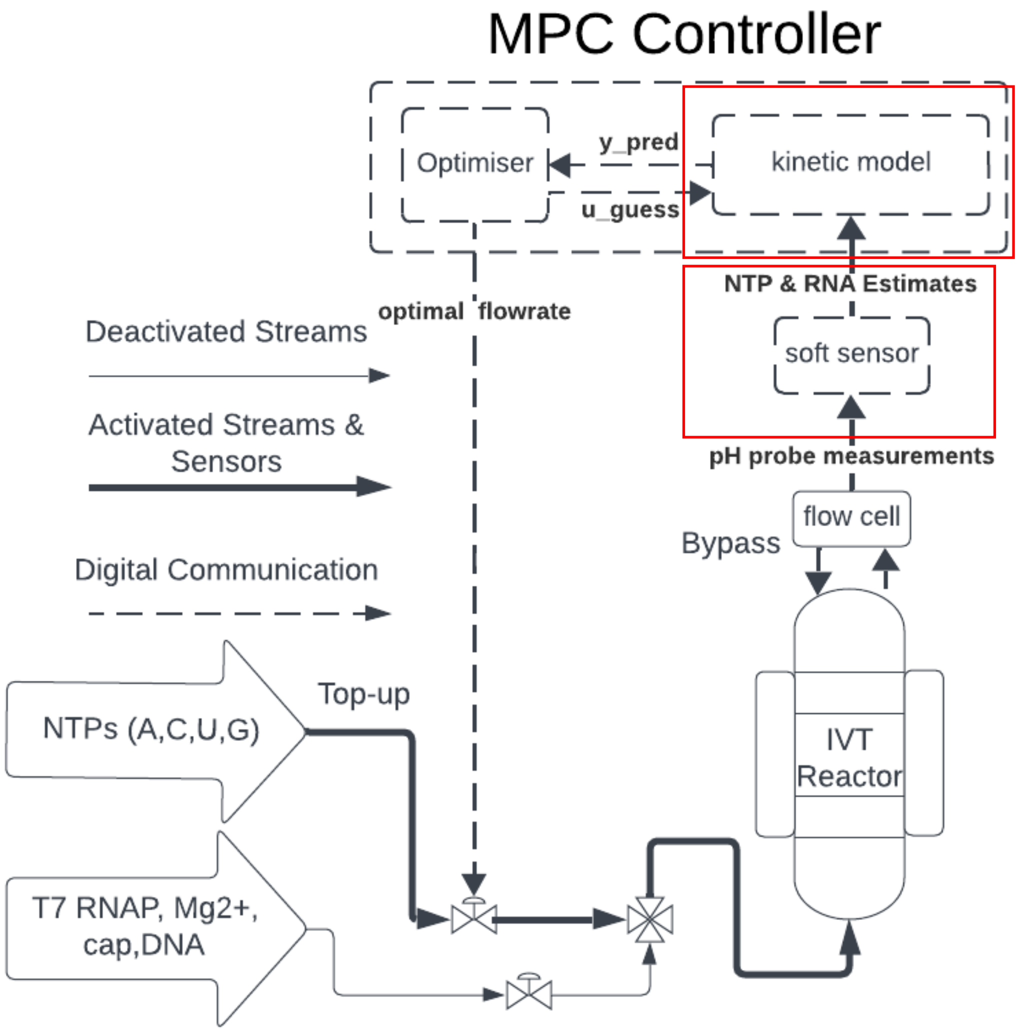
Gaussian Processes

- Follow a Bayesian formulation
- Require little data
- Uncertainty Quantification



Soft sensor development: Physics-informed Gaussian Process Regression

- Two models deployment (complicated setup)
- Error can carry over leading to unnecessary instability
- Kinetic model performance was unsatisfactory



Control Setup: Hybrid Model

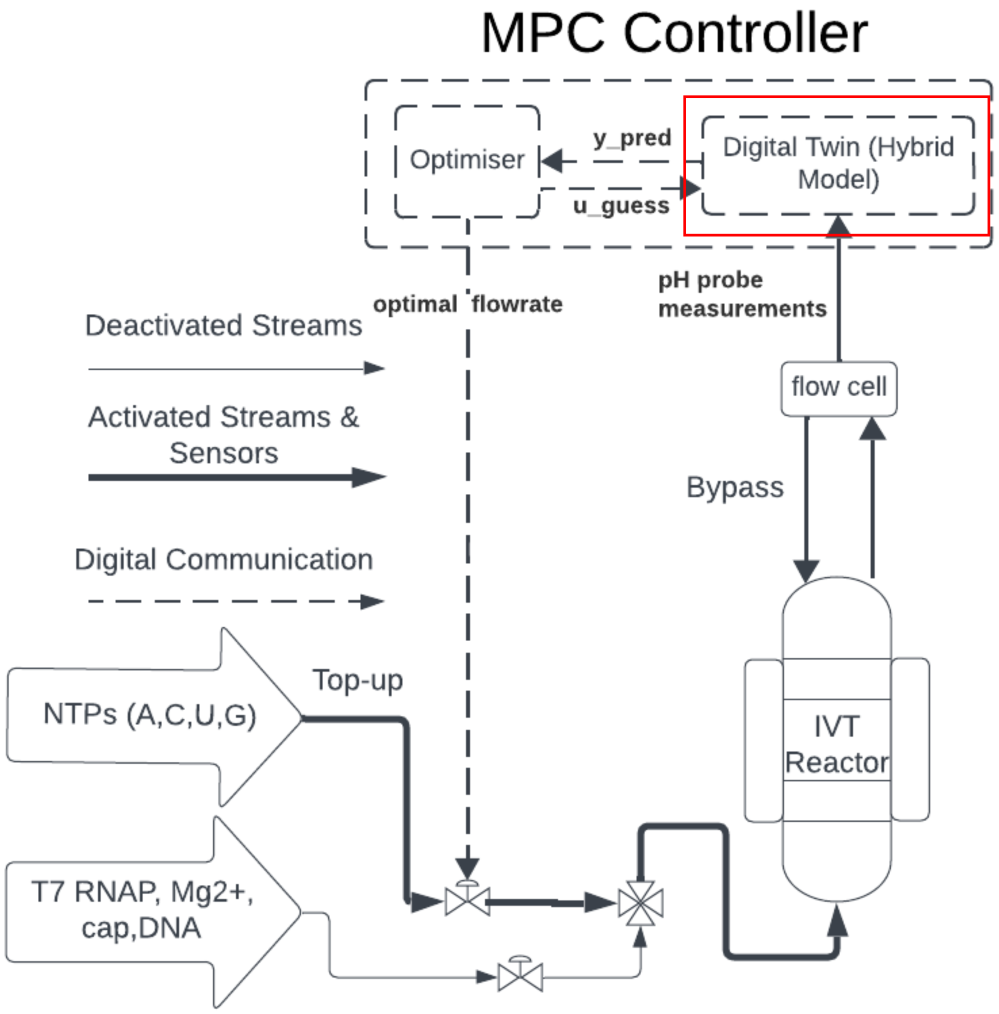
Variable Names	Notation	Example
Prediction Horizon	P	Number of steps the model predicts
Control Horizon	M	Number of control actions evaluated
Control Outputs	Y	RNA yield
Reference Trajectory	r	Desired Yield Response
State Variables	x	Other Components i.e. DNA Template
Control Inputs	U	NTPs Valve Position

Measurements

- pH can be used to monitor the reaction

Objective

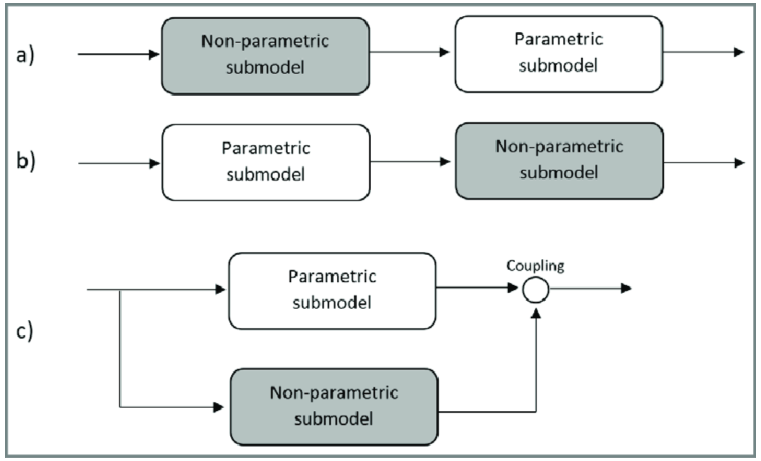
- Our goal is to feed NTP such that we maximise yield with minimum side-effects such as pyrophosphate ions production



Dynamic model development: Hybrid Model

Hybrid semi-parametric model

- Serial structure (ensemble)
- Parallel structure (discrepancy, hierarchical)

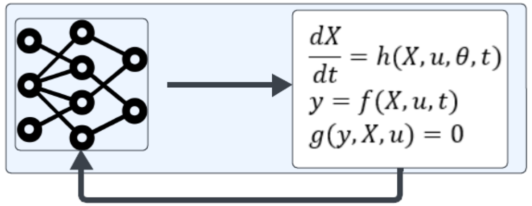


Methodology

$$
\begin{aligned}
\frac{d[\mathrm{RNA}]_{\mathrm{tot}}}{dt} &= V_{\mathrm{tr}} - V_{\mathrm{deg}} - (2 \cdot V_{ds} + V'_{ds}) \\
\frac{d[\mathrm{PPi}]_{\mathrm{tot}}}{dt} &= N_{\mathrm{all}} \cdot V_{\mathrm{tr}} - V_{\mathrm{precip}} - V_{\mathrm{hydrolysis}} \\
\frac{d[\mathrm{ATP}]_{\mathrm{tot}}}{dt} &= -N_{\mathrm{a}} \cdot V_{\mathrm{tr}} \\
\frac{d[\mathrm{GTP}]_{\mathrm{tot}}}{dt} &= -N_{\mathrm{g}} \cdot V_{\mathrm{tr}} \\
\frac{d[\mathrm{CTP}]_{\mathrm{tot}}}{dt} &= -N_{\mathrm{c}} \cdot V_{\mathrm{tr}} \\
\frac{d[\mathrm{UTP}]_{\mathrm{tot}}}{dt} &= -(N_{\mathrm{u}} \cdot V_{\mathrm{tr}} + N_{ds,u} \cdot V'_{ds}) \\
\frac{d[H]_{\mathrm{tot}}}{dt} &= N_{\mathrm{all}} \cdot V_{\mathrm{tr}} \\
\frac{d[\mathrm{T7RNAP}]_{\mathrm{tot}}}{dt} &= -k_d[\mathrm{T7RNAP}]_{\mathrm{tot}} \\
\frac{d[\mathrm{Mg}]_{\mathrm{tot}}}{dt} &= -2 \cdot V_{\mathrm{precip}} \\
\frac{d[\mathrm{dsRNA}]}{dt} &= 2 \cdot V_{ds} + V'_{ds}
\end{aligned}
$$

$$f(\mathbf{x}) \sim \mathcal{GP}\left(m_f(\mathbf{x}), k_f(\mathbf{x}, \mathbf{x}')\right)$$

$$\kappa(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$$



$$WMSE = \frac{1}{T}\sum_{t=1}^{T} \frac{(c_t^* - c_t)^2}{\sigma_t^2}$$

Levenberg-Marquardt (LMM) algorithm

Evaluated ALL the relevant ML techniques

- Support Vector Regression
- Bayesian Regression
- Gaussian Processes
- Xgboost, Random Forest, LightGBM
- Linear Regression, ElasticNet
- Partial Least Squares

Control Setup: Optimiser

Variable Names	Notation	Example
Prediction Horizon	P	Number of steps the model predicts
Control Horizon	M	Number of control actions evaluated
Control Outputs	Y	RNA yield
Reference Trajectory	r	Desired Yield Response
State Variables	x	Other Components i.e. DNA Template
Control Inputs	U	NTPs Valve Position

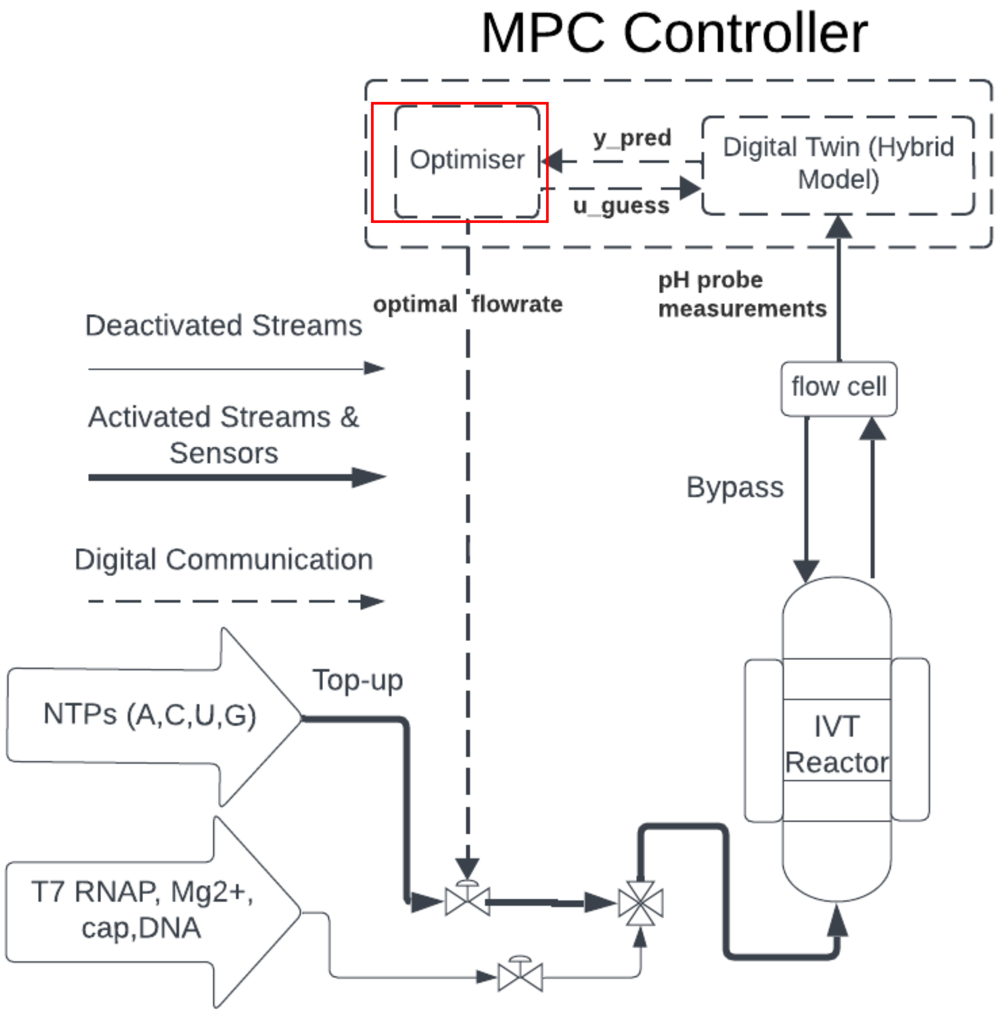
Measurements

- pH can be used to monitor the reaction

Objective

- Our goal is to feed NTP such that we maximise yield with minimum side-effects such as pyrophosphate ions production

$$min_{U(k)} J = \sum_{i=0}^{M-1} \Delta u_k^T W_1 \Delta u_k + \sum_{j=0}^{P-1} \left(\hat{y}_{k+j|k} - r_{k+1}\right)^T W_2 \left(\hat{y}_{k+j|k} - r_{k+1}\right) + \sum_{i=0}^{M-1} u_k^T W_3 u_k$$



MPC Controller Design: Tuning Parameters

Metrics

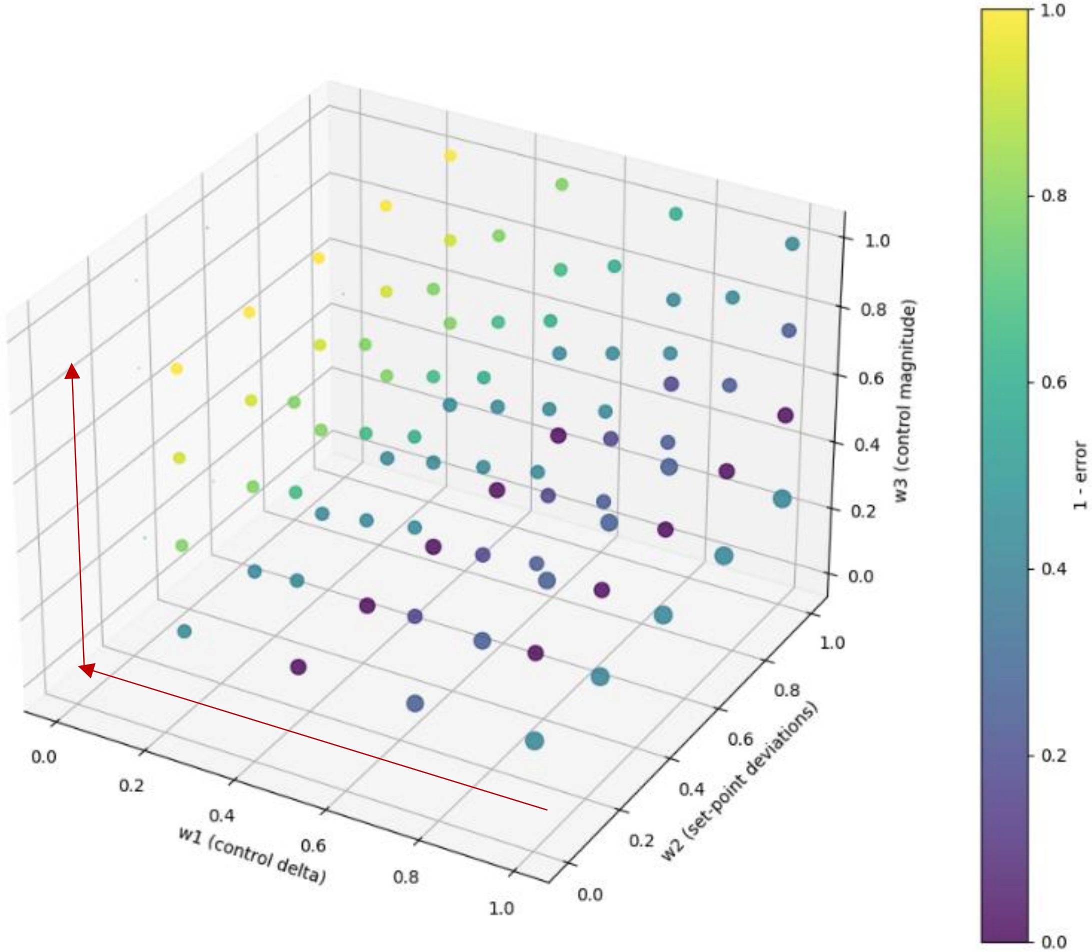
- Colour – Steady state performance
- Size – Control Effort

Factors

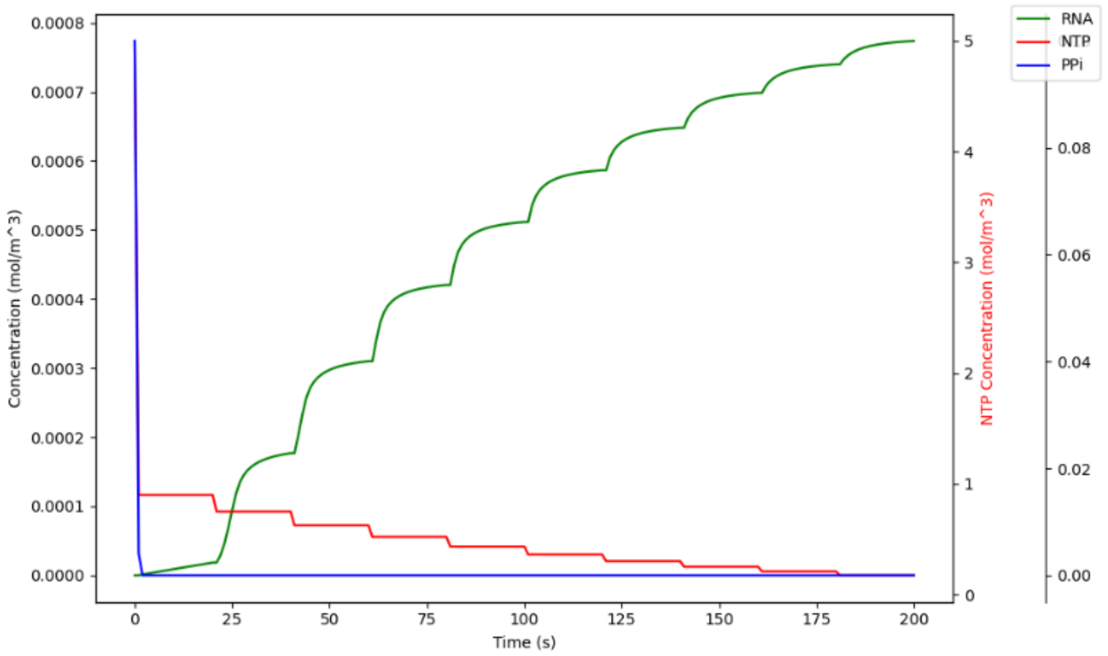
- W_1– control delta – Significant impact
- W_2– reference trajectory – No impact on performance
- W_3– control magnitude – Significant impact

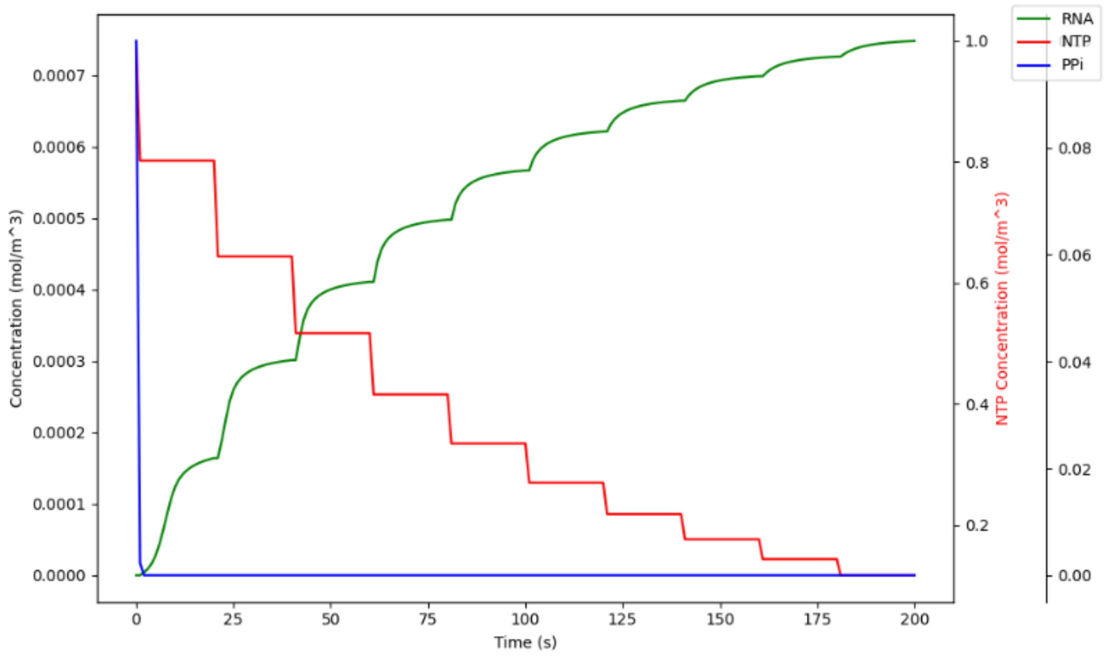
Note

- Reference trajectory range was not physically meaningful therefore, we will need to adopt economic MPC



MPC Controller Design: Control Effort





Methodology

- W_1– control delta
- W_2– reference trajectory
- W_3– control magnitude

$$min_{U(k)} J = \sum_{i=0}^{M-1} \Delta u_k^T W_1 \Delta u_k + \sum_{j=0}^{P-1} \left(\hat{y}_{k+j|k} - r_{k+1}\right)^T W_2 \left(\hat{y}_{k+j|k} - r_{k+1}\right) + \sum_{i=0}^{M-1} u_k^T W_3 u_k$$

Control Setup: Connectivity

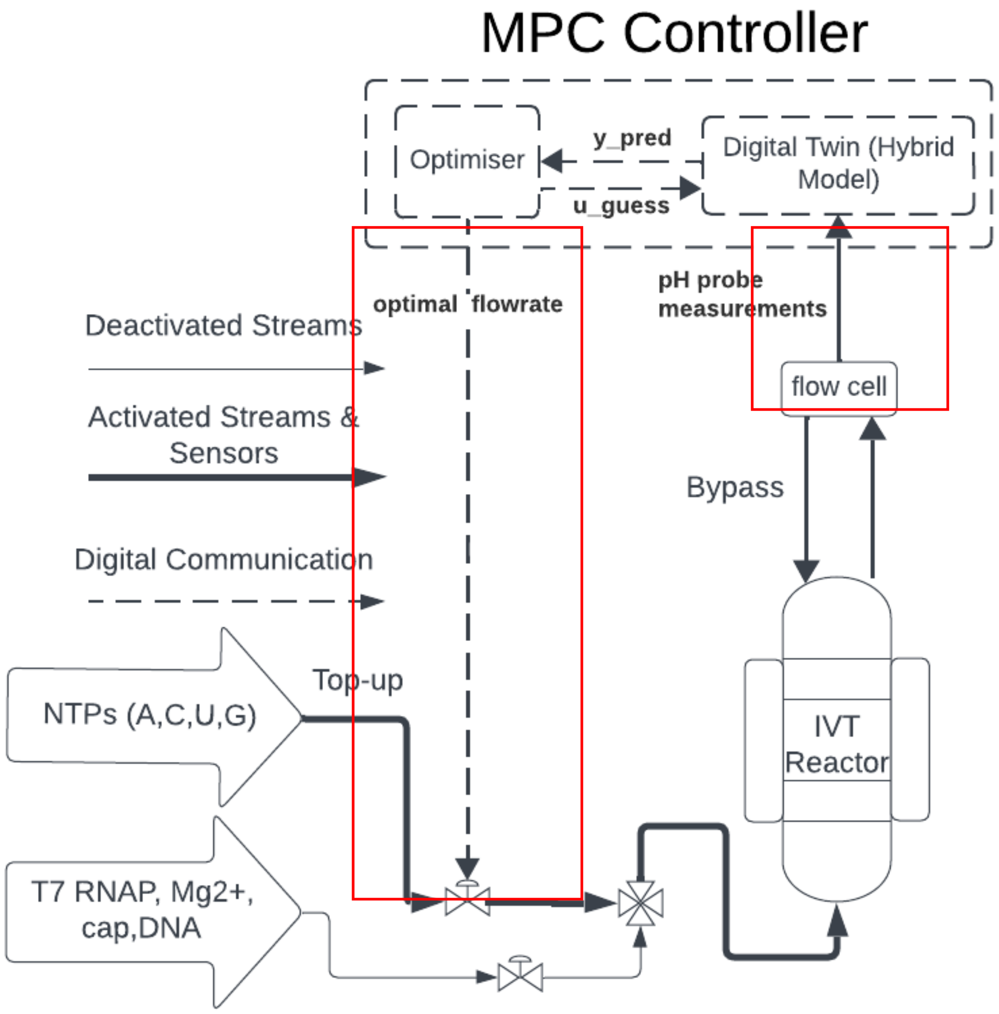
Variable Names	Notation	Example
Prediction Horizon	P	Number of steps the model predicts
Control Horizon	M	Number of control actions evaluated
Control Outputs	Y	RNA yield
Reference Trajectory	r	Desired Yield Response
State Variables	x	Other Components i.e. DNA Template
Control Inputs	U	NTPs Valve Position

Measurements

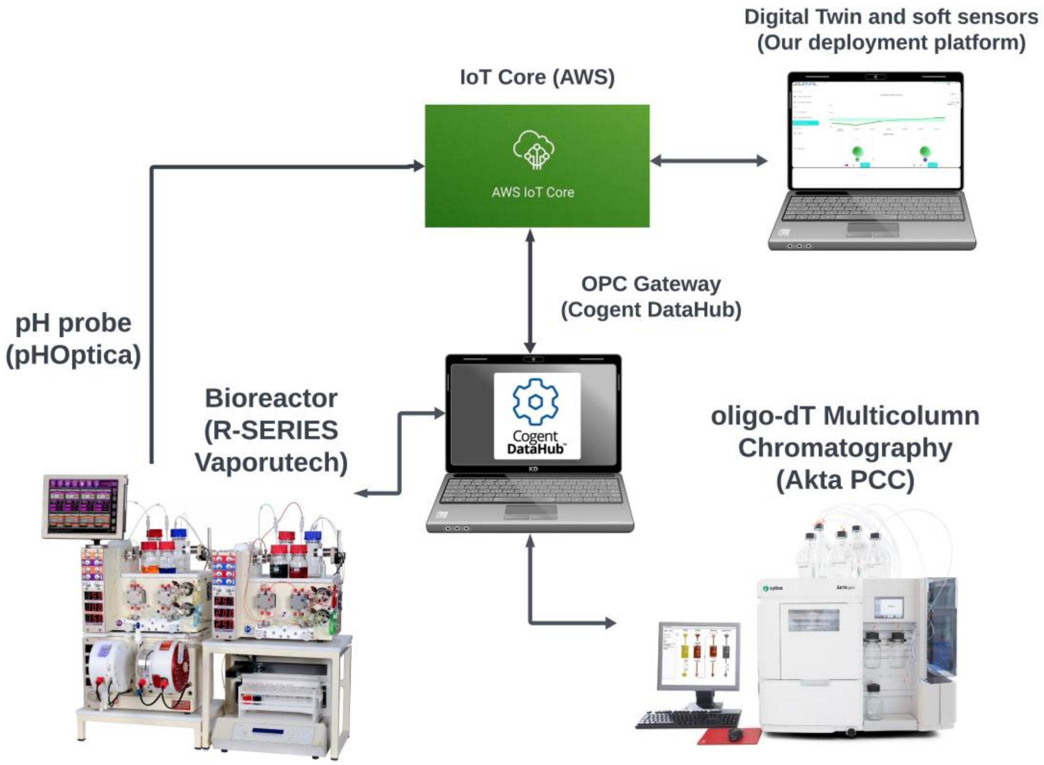
- pH can be used to monitor the reaction

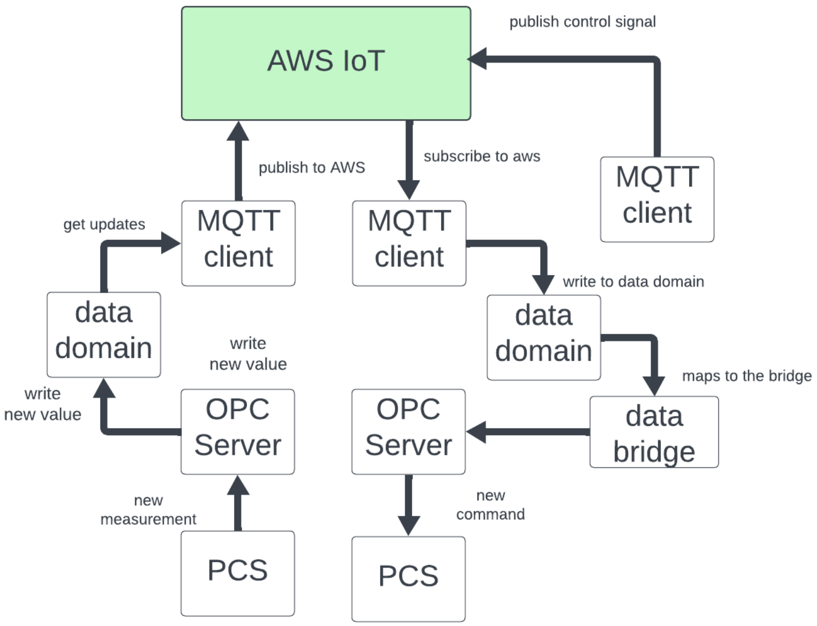
Objective

- Our goal is to feed NTP such that we maximise yield with minimum side-effects such as pyrophosphate ions production



Industrial IoT Platform: Model Deployment and Connectivity





Methodology

- OPC – for interoperability
- MQTT – for lightweight robust communication

IIoT Platform Test Feedback

Test Type I errors (False Positives)

Step 1: Increase Mg2+ flowrate to 1.25 L/min. **(SAFE)**

Step 2: Check new RNA yield prediction (≈7 g/L) within acceptance ranges so reaction continues.

Test Type II errors (False Negative)

Step 3: Increase Mg2+ flowrate to 2.00 L/min. **(PRECIPITATION)**

Step 4: Check new RNA yield prediction (≈5 g/L) outside acceptance ranges.

Step 5: The model shuts down the reaction (Set minimum flowrate to 50 μL/min to prevent bad batch)

IIoT Platform Soft Sensor Deployment

Results

Hybrid model

- accuracy: 85% in forecasting RNA yield

Soft sensor

- accuracy: 90% in estimating current NTPs & RNA concentrations

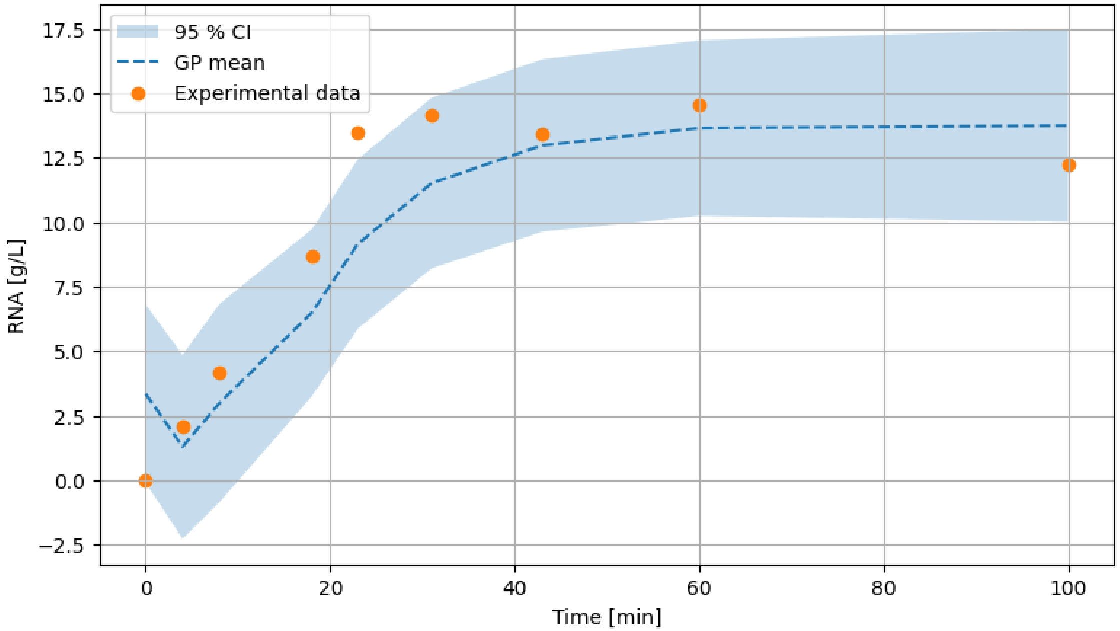
IIoT Platform Features:

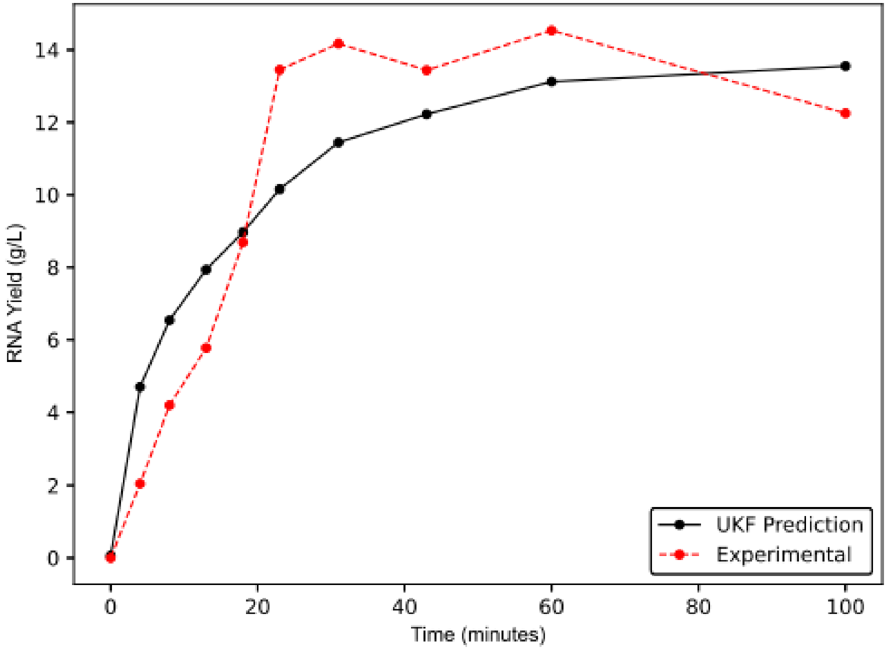
- model deployment
- two-way data connectivity
- real-time visualisation

MPC Controller

- Reduce Costs by 31% vs recycling w PID (85% vs Traditional)

Results Hybrid Model vs Mechanistic Model

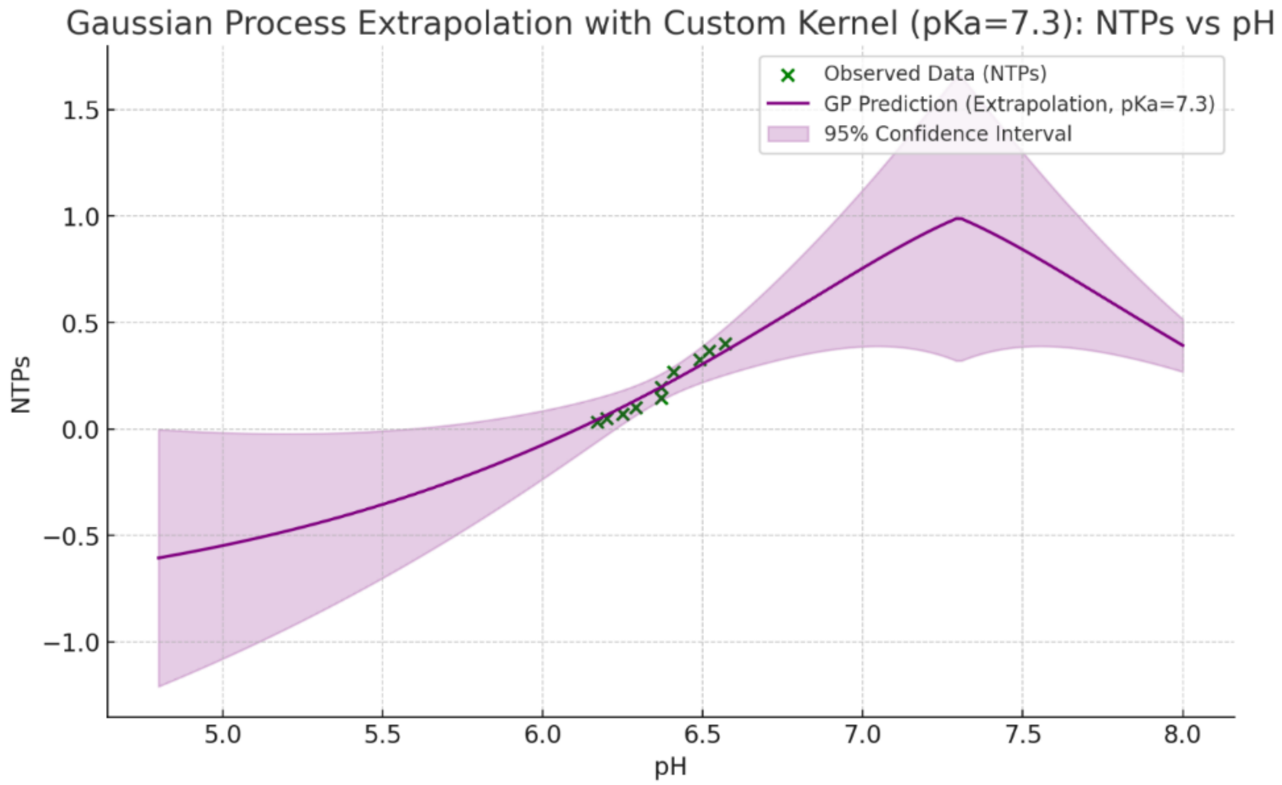




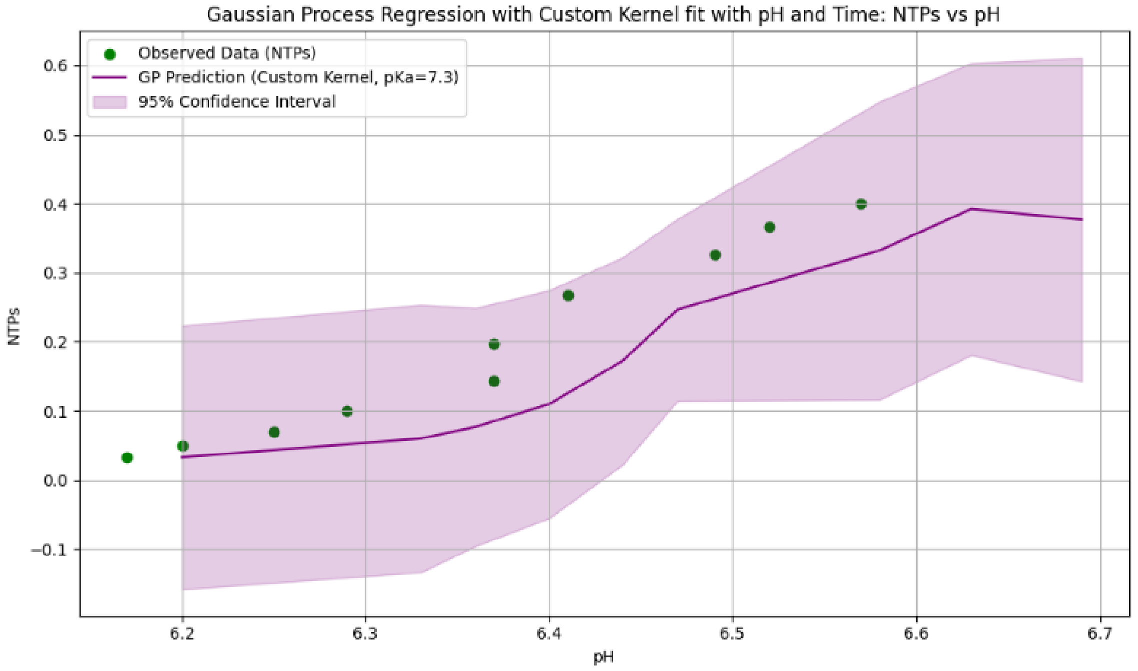
- Hybrid model predictions on unseen data

- Mechanistic model predictions on unseen data

Results: soft sensor



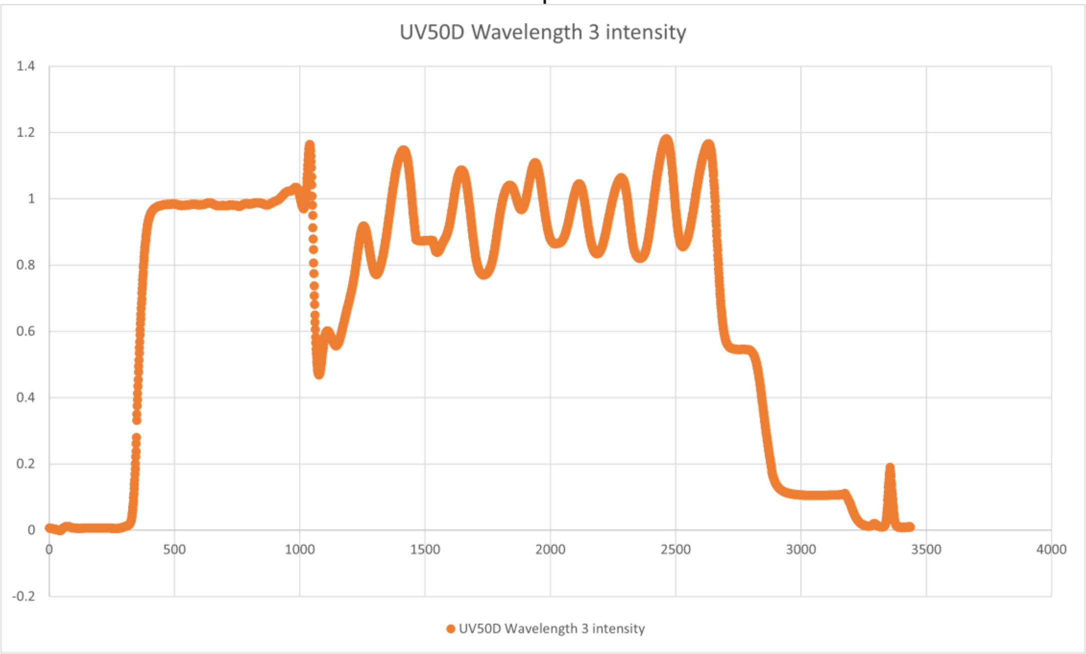
Ceiling constrained successfully!



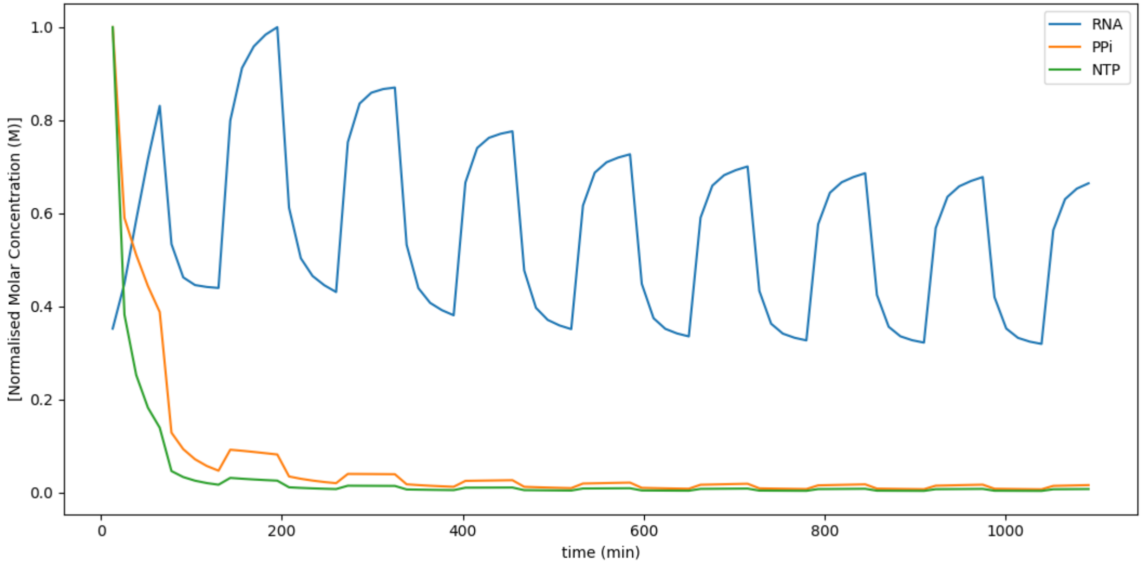
Non-linear relationship between pH and NTPs

Results: Plant Simulator

MPC Response Profile

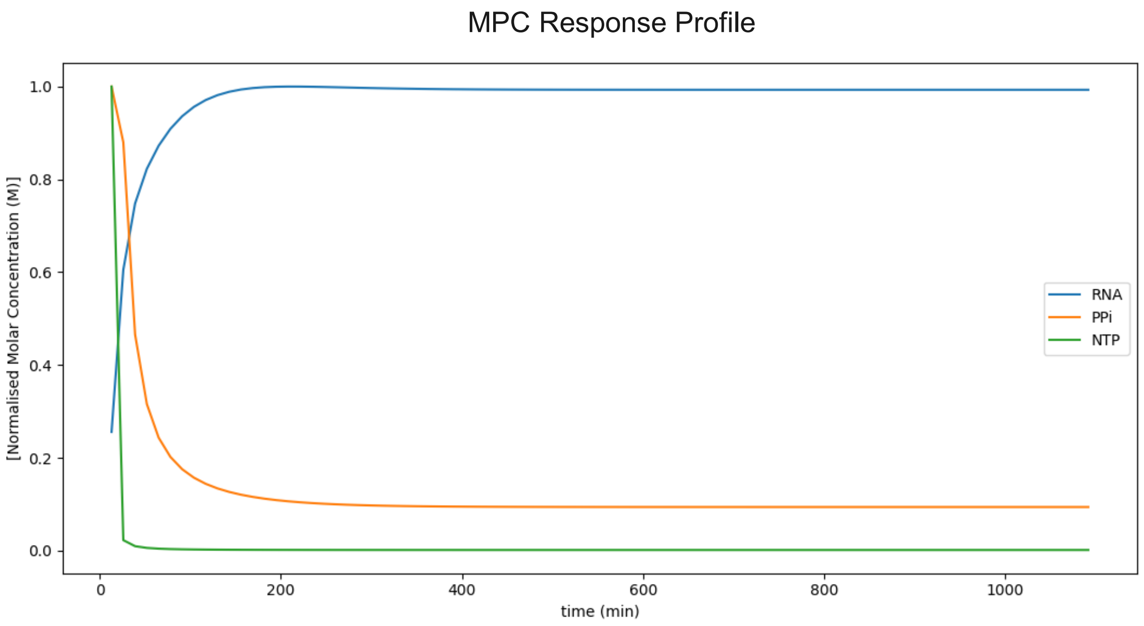


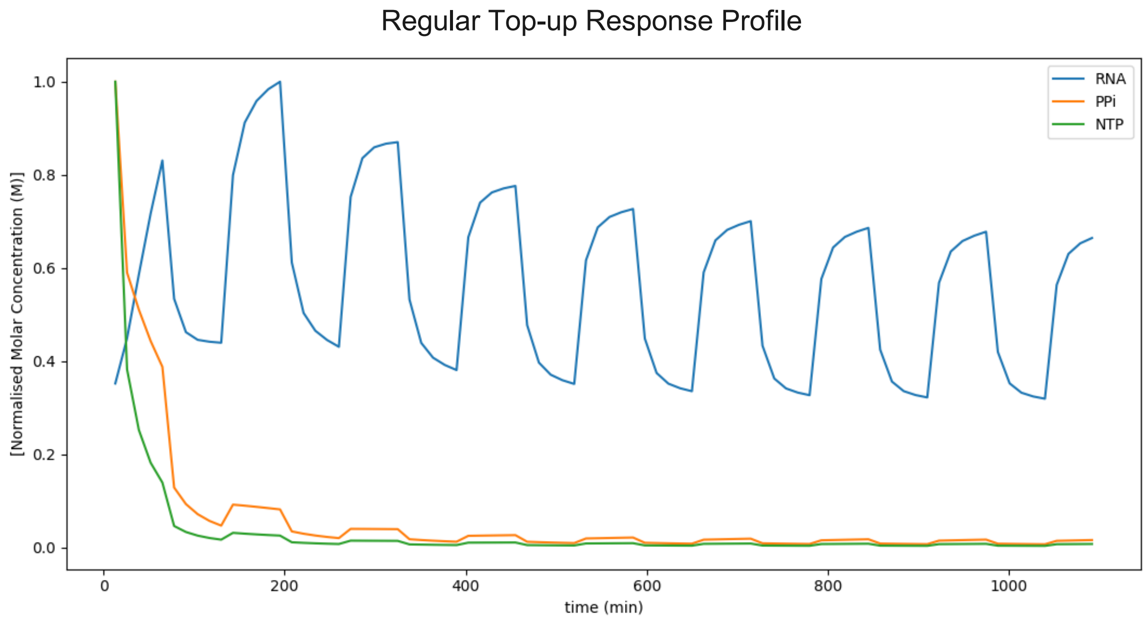
Regular Top-up Response Profile



- Plant simulation captures oscillatory dynamics in experimental data
- Quantitative evaluation of the plant required

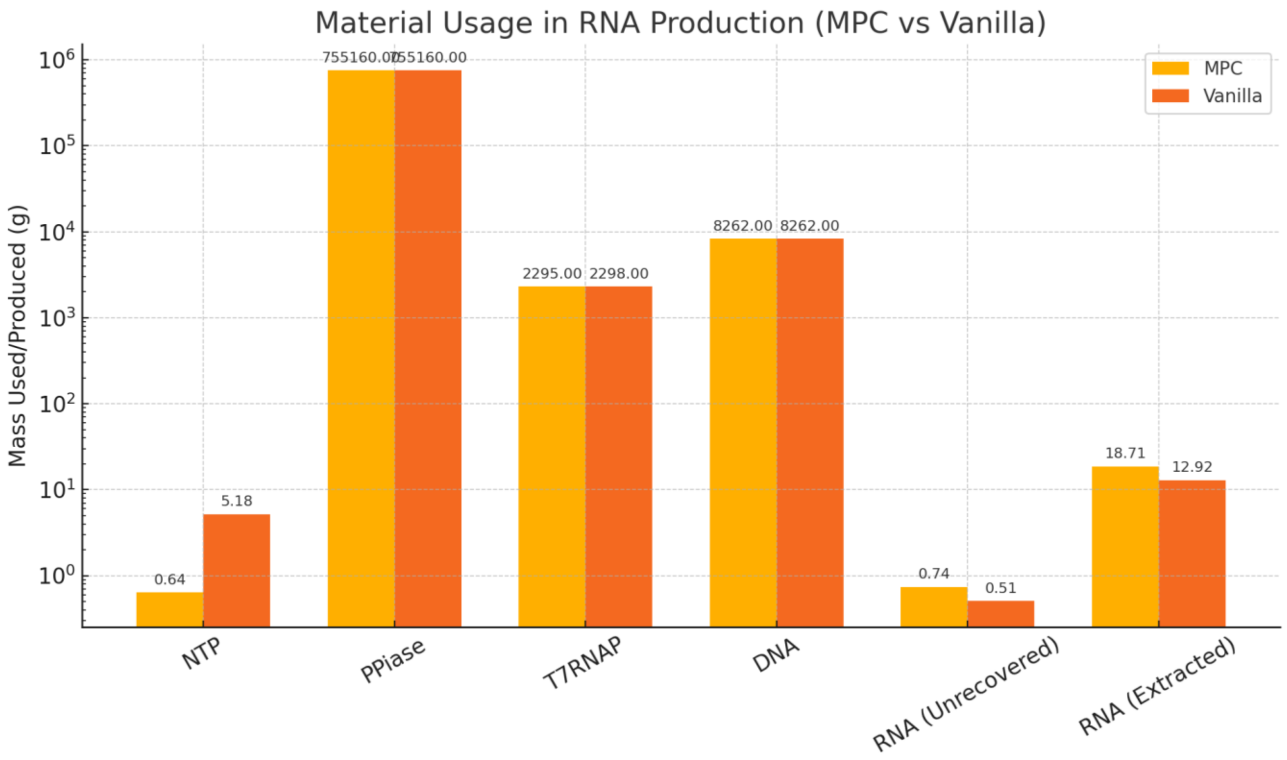
Results: Multi-objective optimisation

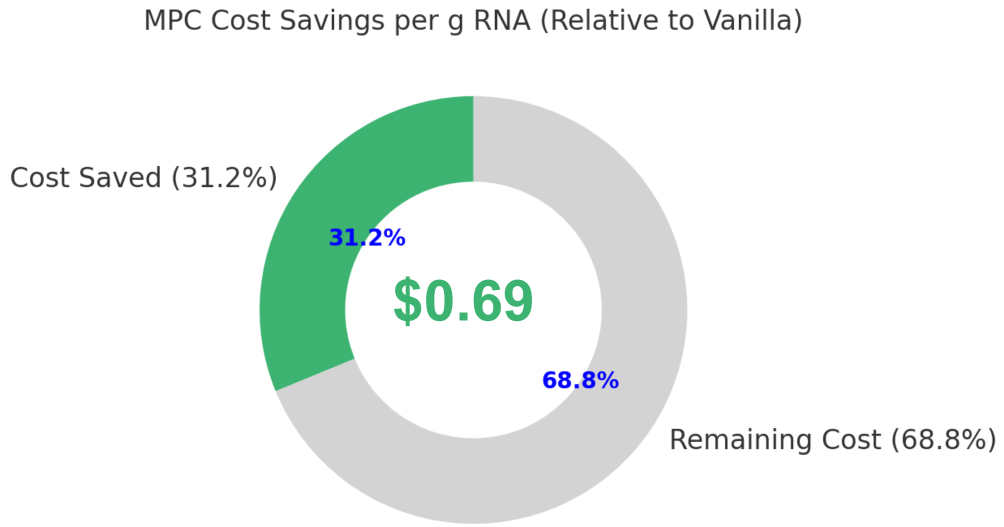




- Model Predictive Controller minimises impurity (PPi)
- MPC controller also maximises yield and maintains optimal operations
- MPC controller achieves better stability and faster steady-state

Results: Multi-objective optimisation





- **MPC costs $0.69 on the dollar compared to standard recycling**
- **Compared to traditional (no recycling) is < 15% of the cost**

- MPC makes use of manipulated variables (NTPs) much more effectively
- This leads to higher yields of RNA using less reagent

Summary & Future Work

Summary

- Developed dynamic hybrid model and soft sensors for real-time predictions
- Achieved multi-objective optimisation
- Tested the control strategy in simulation demonstrating cost savings
- Deployed dynamic model in predictive mode via a custom IIoT Platform

Future Work

- Implement Economic MPC with GP-ARX model in the close-loop
- Use Bayesian Optimisation to tune controller parameters
- Deploy full-controller

Acknowledgments
Thanks for listening
Any Questions?

Funding: Wellcome Leap R3 Programme, CEPI, H. Walter Stern Scholarship

Collaborators: Kesler Isoko, Mahdi Ahmed, Joseph Middleton, Joan L. Cordiner, Zoltan Kis, Peyman Z. Moghadam