Dissertation: Optimising the design of buffer preparation in bioprocessing facilities

MSc in Business Analytics (Part Time) 2015–2017

Sean Tully

University College Dublin

25th August 2017

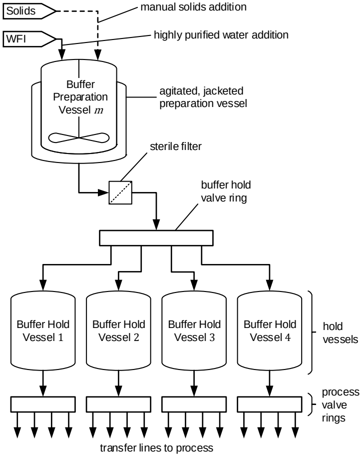
Introduction – Problem Definition

- Bioprocessing involves the synthesis of therapeutic products using biological agents (bacteria or mammalian cells)
- The purification of these products typically requires the use of 10–20 different buffer solutions (*buffers*)
- Calculating the correct number and volume of vessels for buffer preparation in a large-scale facility represents a significant engineering challenge
- The aim of this work was to develop a methodology for calculating buffer vessel requirements, given some process information

Top: Alexion, Dublin. Bottom: Janssen Biologics, Cork (formerly Centocor). Both images ©PM Group.

Buffer Preparation

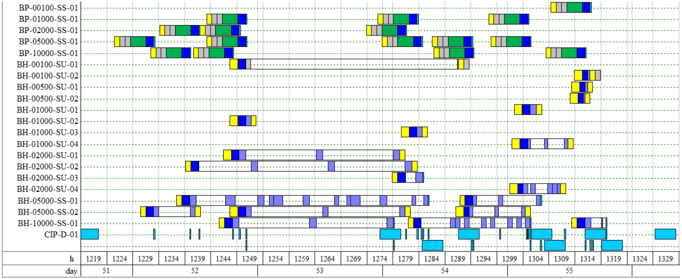
- Buffer preparation requires two vessels
- Each buffer is prepared in a *preparation vessel*, which may be used to prepare many buffers
- It is assumed that each buffer has a dedicated *hold vessel*
- We want to minimise the total cost of vessels in buffer preparation.



Single buffer preparation – equipment

Existing Workflow

- Model production schedule using a software package such as SchedulePro
- Run the model with a dedicated preparation vessel for every hold vessel
- Remove preparation vessels until no more can be removed
  - *involves trial and error*
- Stop when you can't find a way to remove any more vessels
  - *no guarantee that the optimum selection has been found*



Proposition

- Find optimum vessel selection to minimise cost, using Mixed Integer Linear Programming (MILP)

Input Data

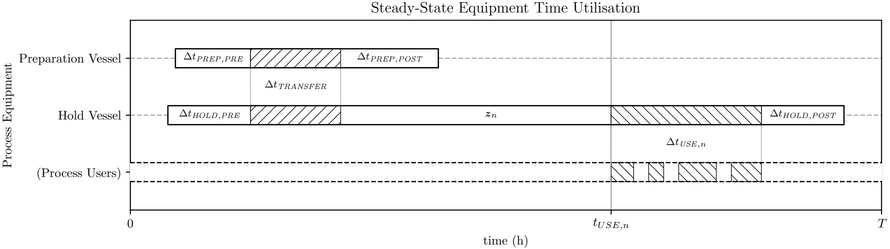
Buffer Data

names	volumes U_n (l)	use start times $t^*_{USE,n}$ (h)	use durations $\Delta t_{USE,n}$ (h)
Buffer #1	24 427.13	76.23	20.56
Buffer #2	5487.29	0.21	49.77
Buffer #3	2588.36	25.78	24.56
Buffer #4	7102.05	46.79	27.77
Buffer #5	1020.87	87.70	36.58
Buffer #6	19 508.79	35.52	58.53
Buffer #7	23 073.55	42.26	39.71
Buffer #8	25 454.10	48.38	43.47
Buffer #9	24 088.67	4.18	55.41
Buffer #10	3172.46	48.31	23.27
Buffer #11	24 752.71	76.38	45.80
Buffer #12	13 445.31	73.93	34.25

Vessel Data

names	volumes V_m (l)	costs c_m (−)
1000 l	1000.0	63.10
2000 l	2000.0	95.64
3000 l	3000.0	121.98
4000 l	4000.0	144.96
5000 l	5000.0	165.72
6000 l	6000.0	184.88
⋮	⋮	⋮
20 000 l	20 000.0	380.73
22 000 l	22 000.0	403.14
25 000 l	25 000.0	435.28
30 000 l	30 000.0	485.59

Buffer Preparation Scheduling



- The volume of each buffer is known, as is it's time of first use and total duration of use
- Some flexibility exists in the hold duration and in vessel selection

Parameters

symbol	short description	value	unit
T	process cycle time	96.0	h
$\Delta t_{PREP,PRE}$	prep pre duration	12.0	h
$\Delta t_{PREP,POST}$	prep post duration	1.5	h
$\Delta t_{TRANSFER}$	transfer duration	2.0	h
$\Delta t_{HOLD,PRE}$	hold pre duration	8.0	h
$\Delta t_{HOLD,POST}$	hold post duration	1.5	h
$\Delta t_{HOLD,MIN}$	minimum hold duration	12.0	h
$\Delta t_{HOLD,MAX}$	maximum hold duration	60.0	h
$f_{MINFILL}$	minimum fill ratio	0.3	–
f_{UTIL}	maximum utilisation ratio	0.8	–

LP Model

Minimise:

$$\boldsymbol{Z} = \sum_{m \in \mathcal{M}} \sum_{p \in \mathcal{P}} c_m \boldsymbol{y}_{mp}$$

Subject to:

$$\sum_{p \in \mathcal{P}} \boldsymbol{x}_{np} = 1 \quad \forall n \in \mathcal{N}$$

$$\sum_{m \in \mathcal{M}} \boldsymbol{y}_{mp} \leq 1 \quad \forall p \in \mathcal{P}$$

$$U_n \boldsymbol{x}_{np} - \sum_{m \in \mathcal{M}} V_m \boldsymbol{y}_{mp} \leq 0 \quad \forall n \in \mathcal{N}, \ \forall p \in \mathcal{P}$$

$$V_{MAX} \boldsymbol{x}_{np} + f_{MINFILL} \sum_{m \in \mathcal{M}} V_m \boldsymbol{y}_{mp} \leq U_n + V_{MAX} \quad \forall n \in \mathcal{N}, \ \forall p \in \mathcal{P}$$

$$\Delta t_{PREP} \sum_{n \in \mathcal{N}} \boldsymbol{x}_{np} \leq f_{UTIL} T \quad \forall p \in \mathcal{P}$$

$$\boldsymbol{z}_n \leq T - \Delta t_{HOLD,PRE} - \Delta t_{TRANSFER} - \Delta t_{USE,n} - \Delta t_{HOLD,POST} \quad \forall n \in \mathcal{N}$$

$$\left.\begin{aligned} 2\boldsymbol{w}_{nkp} - \boldsymbol{x}_{np} - \boldsymbol{x}_{kp} &\leq -2 \\ \boldsymbol{w}_{nkp} - \boldsymbol{x}_{np} - \boldsymbol{x}_{kp} &\geq -1 \end{aligned}\right. \quad \forall n \in \mathcal{N}, \ \forall k \in \mathcal{N}; \ k > n, \ \forall p \in \mathcal{P}$$

$$\left.\begin{aligned} T\boldsymbol{q}_n - \boldsymbol{z}_n &\geq -t_{USE,n} \\ T\boldsymbol{q}_n - \boldsymbol{z}_n &\leq -t_{USE,n} + T \end{aligned}\right. \quad \forall n \in \mathcal{N}$$

LP Model – Continued

Subject to (continued):

$$\left.\begin{aligned} -T\boldsymbol{q}_n + T\boldsymbol{r}_n + \boldsymbol{z}_n &\leq t_{USE,n} + \Delta t_{PREP} \\ -T\boldsymbol{q}_n + T\boldsymbol{r}_n + \boldsymbol{z}_n &\geq t_{USE,n} + \Delta t_{PREP} - T \end{aligned}\right\} \quad \forall n \in \mathcal{N}$$

$$\left.\begin{aligned} T\boldsymbol{q}_n + T\boldsymbol{s}_n - \boldsymbol{z}_n &\leq -t_{USE,n} + \Delta t_{PREP} \\ T\boldsymbol{q}_n + T\boldsymbol{s}_n - \boldsymbol{z}_n &\geq -t_{USE,n} + \Delta t_{PREP} + T \end{aligned}\right\} \quad \forall n \in \mathcal{N}$$

$$\left.\begin{aligned} \boldsymbol{r}_n + \boldsymbol{s}_n - \ \ \boldsymbol{u}_n &\geq 0 \\ \boldsymbol{r}_n + \boldsymbol{s}_n - 2\boldsymbol{u}_n &\leq 0 \end{aligned}\right\} \quad \forall n \in \mathcal{N}$$

$$\left.\begin{aligned} T\boldsymbol{q}_k - T\boldsymbol{q}_n + 2T\boldsymbol{u}_n + 2T\boldsymbol{v}_{nk} - 2T\sum_{p\in\mathcal{P}} \boldsymbol{w}_{nkp} - \boldsymbol{z}_k + \boldsymbol{z}_n &\geq t_{USE,n} - t_{USE,k} + \Delta t_{PREP} - 2T \\ T\boldsymbol{q}_k - T\boldsymbol{q}_n - 2T\boldsymbol{u}_n + 2T\boldsymbol{v}_{nk} + 2T\sum_{p\in\mathcal{P}} \boldsymbol{w}_{nkp} - \boldsymbol{z}_k + \boldsymbol{z}_n &\geq t_{USE,n} - t_{USE,k} - \Delta t_{PREP} + 4T \\ T\boldsymbol{q}_k - T\boldsymbol{q}_n - T\boldsymbol{r}_n + 2T\boldsymbol{u}_n + 2T\sum_{p\in\mathcal{P}} \boldsymbol{w}_{nkp} - \boldsymbol{z}_k + \boldsymbol{z}_n &\geq t_{USE,n} - t_{USE,k} - \Delta t_{PREP} + 4T \\ T\boldsymbol{q}_k - T\boldsymbol{q}_n + T\boldsymbol{s}_n - 2T\boldsymbol{u}_n - 2T\sum_{p\in\mathcal{P}} \boldsymbol{w}_{nkp} - \boldsymbol{z}_k + \boldsymbol{z}_n &\geq t_{USE,n} - t_{USE,k} + \Delta t_{PREP} - 4T \end{aligned}\right\} \quad \forall n \in \mathcal{N},\ \forall k \in \mathcal{N};\ k > n$$

LP Model – Continued

Where:

$$\boldsymbol{q}_n \in \{0,1\} \quad \forall n \in \mathcal{N}$$

$$\boldsymbol{r}_n \in \{0,1\} \quad \forall n \in \mathcal{N}$$

$$\boldsymbol{s}_n \in \{0,1\} \quad \forall n \in \mathcal{N}$$

$$\boldsymbol{u}_n \in \{0,1\} \quad \forall n \in \mathcal{N}$$

$$\boldsymbol{v}_{nk} \in \{0,1\} \quad \forall n \in \mathcal{N},\ \forall k \in \mathcal{N};\ k > n$$

$$\boldsymbol{w}_{nkp} \in \{0,1\} \quad \forall n \in \mathcal{N},\ \forall k \in \mathcal{N};\ k > n,\ \forall p \in \mathcal{P}$$

$$\boldsymbol{x}_{np} \in \{0,1\} \quad \forall n \in \mathcal{N},\ \forall p \in \mathcal{P}$$

$$\boldsymbol{y}_{mp} \in \{0,1\} \quad \forall m \in \mathcal{M},\ \forall p \in \mathcal{P}$$

$$\Delta t_{HOLD,MIN} \leq \boldsymbol{z}_n \leq \Delta t_{HOLD,MAX}; \quad \boldsymbol{z}_n \in \mathbb{R} \quad \forall n \in \mathcal{N}$$

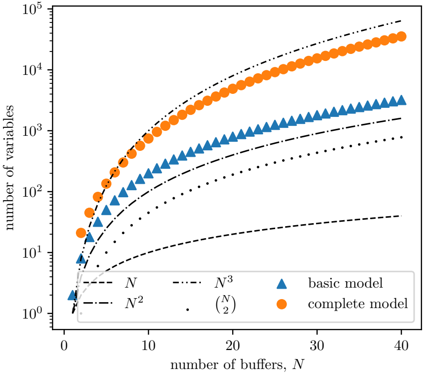
Note: The above holds for a problem with $N = |\mathcal{N}|$ buffers, $M = |\mathcal{M}|$ available vessel sizes and $P = |\mathcal{P}|$ available vessel slots, where $P = N$.

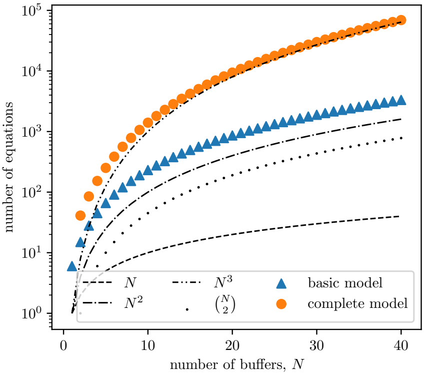
Implementation

- The Complete Model was implemented in python, using the PuLP library which acts as an API to several commercial and open-source mixed-integer linear programming solvers
- The proprietary IBM ILOG CPLEX Optimizer was primarily used to solve the problem
- The open-source CBC solver was also used, though it is considerably slower
- The matplotlib library was used to generate plots
- Both random data and data from real-world examples were examined

Model Complexity

model	no. of variables	no. of equations
basic	$N^2 + NM$	$2N^2 + 3N + 1$
complete	$N\binom{N}{2} + N^2 + \binom{N}{2} + NM + 5N$	$2N\binom{N}{2} + 4\binom{N}{2} + 2N^2 + 12N + 1$





Results – Random Example

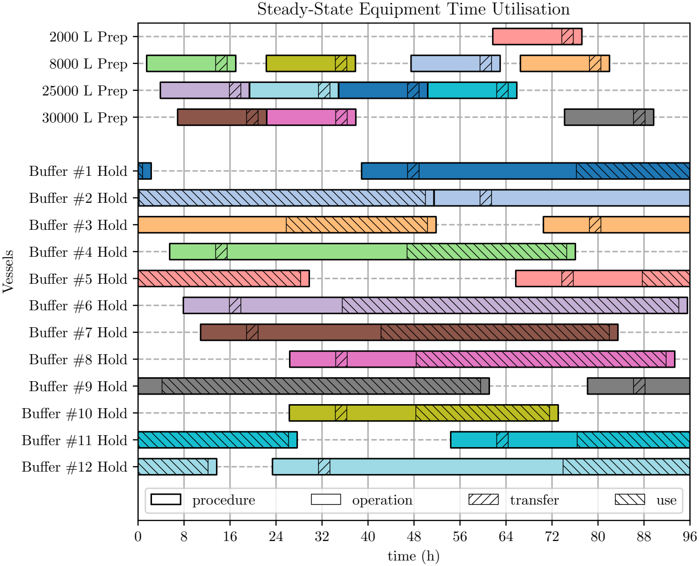
Solving the random example yields the following result:

Required preparation vessels for random example

vessel size
2000 l
8000 l
25 000 l
30 000 l

Note that there may be many ways of assigning the buffers to these vessels and many feasible ways of scheduling these buffers.
A feasible schedule may be represented graphically using an *equipment time utilisation* plot.

Equipment Time Utilisation – Random Example

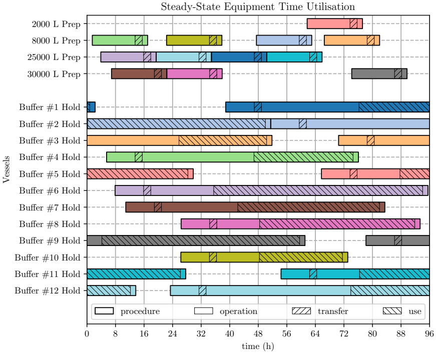


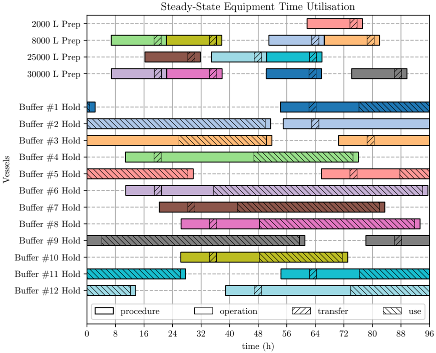
Goal Programming – Reducing Hold Times

A *better* feasible schedule may be obtained using *goal programming*.

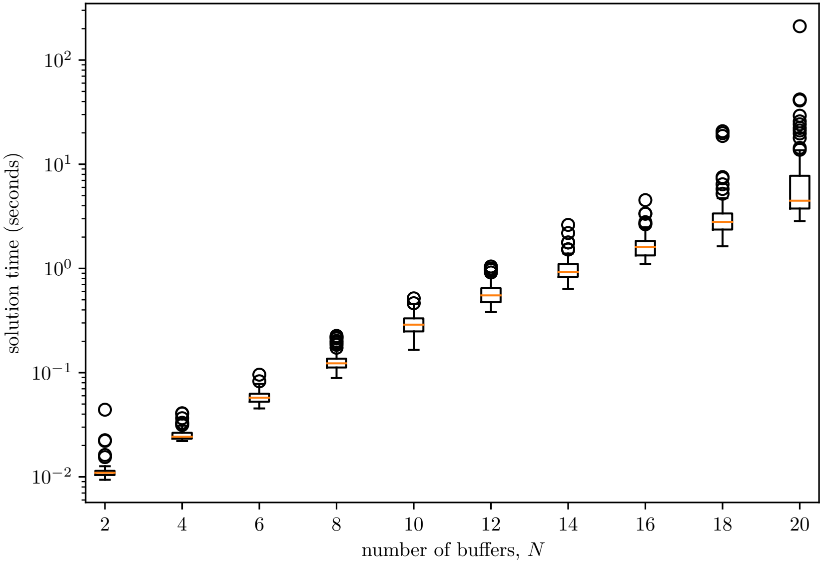
The (original) objective function value is fixed as a parameter and we replace the original objective function with a new one, which seeks to minimise hold times.

In doing so, hold times are minimised, but only to the extent that they don't interfere with the original list of required preparation vessels.





Solution Time



Contribution

- Academic
  - No previous methodologies for solving such a problem could be found in the literature
  - Contributes to the fields of process engineering and linear programming
- Business
  - Reduces time taken to generate early-stage designs
  - Methodology ensures that a solution is provably optimum
  - The ability to carry out the simulation and present visual results showcases expertise that differentiates a design firm's offering from its peers, helping to win further work.

Scope for Further Work

- Add complexity to the model
  - Material compatibility
  - Non-dedicated hold vessels
  - Multi-batch buffers
  - Multi-product facilities
  - Non-standard operations (e.g. buffers for chromatography column repacks)
- Improve solution efficiency
  - Specifying additional constraints may reduce search space
- Further analysis of results
  - Sensitivity analysis
  - Monte-Carlo simulation

Optimising the design of buffer preparation in bioprocessing facilities

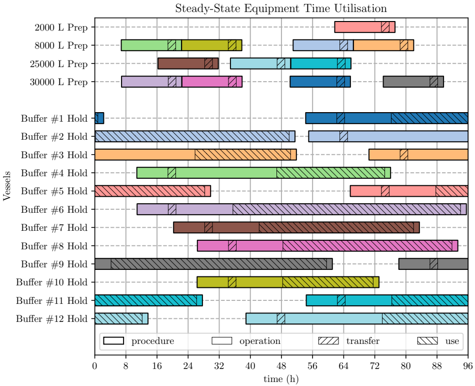
Author: Sean Tully   Supervisor: Prof. Michael O'Neill

Aim:

- Development of a methodology and software tool for finding the optimum number, size and assignment of buffer preparation vessels in the design of a large-scale bioprocess facility

Key Achievements:

- Problem modelled as a MILP problem
- Code developed to solve problem
- Graphical representations generated of feasible production schedules
- Goal programming carried out for further optimisation



Steady-State Equipment Time Utilisation plot showing a feasible schedule for a random problem with 12 buffers, for a single cycle.