Development of a framework for the optimum design of a cooling water system with a time constrained background process

by

Victor Dippenaar

A dissertation submitted in fulfilment of the requirements for the degree
of
Master of Science in Engineering (Chemical Engineering)

Submitted to
School of Chemical and Metallurgical Engineering,
Faculty of Engineering and the Built Environment,
University of the Witwatersrand, Johannesburg, South Africa

Johannesburg,
2022

Declaration of Authorship

I declare that this dissertation is my own unaided work. It is being submitted for the Degree of Master of Science in Chemical Engineering to the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination to any other University.

Victor Dippenaar

2nd day of February in the year 2022

Abstract

Cooling water systems research aims to investigate the interaction between the cooling water network and cooling tower performance. Historically, it has been assumed that the background process operates continuously. This dissertation presents a mathematical framework for the synthesis and optimisation of a cooling water system consisting of multiple cooling towers with a batch background process. The cooling water network is formulated as a series configuration, which is characterised by reuse of cooling water between different cooling water using operations. In achieving this, an overall superstructure is developed which allows for all different combinations between the cooling towers and cooling water network, as well as opportunities to reuse water between cooling operations to be explored. It should be noted that in previous research on cooling water systems, reuse opportunities were constrained only by temperature. Due to the nature of batch processing, reuse opportunities are subject to both temperature and time constraints in this investigation. The mathematical model is developed as a mixed integer nonlinear programming (MINLP) problem in the GAMS platform. The objective function is the maximisation of total annual profit. The MINLP is solved using the BARON solver. The optimal system configuration, operating conditions and design parameters for the cooling towers and cooling water using operations in the network are determined as part of the optimisation of the mathematical formulation. The model was applied to two illustrative examples, a fixed schedule problem and a flexible schedule problem. The use of a sequential optimisation approach was also compared to an integrated optimisation approach. Results from the illustrative examples demonstrated the superiority of the integrated approach, which led to a reduction in the number of cooling towers required from 3 to 2. Total annual profit was increased by 0.45 %, recirculating cooling water was reduced by 42.1 %, average overall cooling tower effectiveness was increased by 25.3 %, and the makeup water required throughout the time horizon was reduced by 6.3 %.

Keywords

Process integration; optimisation; mathematical modelling; operations research; batch processing; cooling towers; utility systems.

Acknowledgements

"Science, as well as, technology will in the near future and in the farther future increasingly turn from problems of intensity, substance, and energy, to problems of structure, organization, information, and control."

- John von Neumann
1949

This work would not have been possible without the guidance throughout the research process provided by my supervisor, Professor Thokozani Majozi. Your insights and constructive criticism is highly valued.

Furthermore, I would like to acknowledge the Sustainable Process Engineering (SPE) research group, including Darren Morar, Melisha Singh, Malebo Mashego, Nyasha Machakaire, Jaimee Jugmohan, Vassili Revelas, Acuman Gcabashe, Idris Sadik, Yannick Moise Ilunga, Simbarashe Magege, Joel Croft, Zhiwei Li, Doris Oke, Jude Bonsu, Nashwa El-Tahhan, Nelisiwe Selane, Muimeleli Mutangwa, Christie Bazolana, Phumlile Kunene, Oludaisi Adekomaya as well as visiting researchers Ayush Nema and Manuel Balanca. The conversations which we shared lead to important contributions reflected in this dissertation.

Finally, I would like to thank the National Research Foundation (NRF) of South Africa for funding the research presented in this dissertation under the Research Chair in Sustainable Process Engineering Grant number 47440.

Publications and Presentations Emanating from this Research

Below is a list of the conference proceedings and journal publications emanating from this research.

Journal Article:
Dippenaar, V. and Majozi, T. (2021) 'Integrated optimisation framework for the design and scheduling of batch cooling water networks'. *Industrial & Engineering Chemistry Research*, 60 (23), pp. 8460–8474.

Conference Proceedings:
Dippenaar, V. and Majozi, T. (2019) 'Optimum design of a cooling water system with a batch background process'. Paper presented at the *22nd Conference on Process Integration, Modelling, and Optimisation for Energy Saving and Pollution Reduction*, Crete, Greece, 20–23 October.

Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Definition
ABC	Artificial Bee Colony
ACO	Ant Colony Optimisation
BARON	Branch And Reduce Optimisation Navigator
BB	Branch and Bound
CIS	Combined Intermediate Storage
CPA	Critical Path Algorithm
DICOPT	Discrete and Continuous OPTimizer
e-NTU	Effectiveness Number of Transfer Units
EA	Evolutionary Algorithms
ERTN	Extended Resource Task Network
FIS	Finite Intermediate Storage
FW	Finite Wait
GA	Genetic Algorithms
GAMS	General Algebraic Modelling System
LP	Linear Programming
m-STN	Modified State Task Network
MINLP	Mixed Integer Nonlinear Programming
MILP	Mixed Integer Linear Programming
MIS	Mixed Intermediate Storage
NIS	No Intermediate Storage
NLP	Nonlinear Programming
OA	Outer Approximation
PIS	Process Intermediate Storage
PSO	Particle Swarm Optimisation
RLT	Reformulation Linearisation Technique
RTN	Resource Task Network
SA	Simulated Annealing
SSN	State Sequence Network
STN	State Task Network
TAM	Time Average Model
TSM	Time Slice Model
UIS	Unlimited Intermediate Storage
UN	United Nations
UW	Unlimited Wait
ZW	Zero Wait

Nomenclature

Symbol	Description	Unit
Sets		
$i \in I$	Tasks	
$j \in J$	Equipment units	
$n \in N$	Cooling towers	
$p \in P$	Time points	
$s \in S$	States	
Subsets		
$i \in I_j(j)$	Tasks that can take place in equipment unit j	
$i \in I_i(s)$	Tasks that use state s as an input	
$i \in I_o(s)$	Tasks that produce state s as an output	
$i \in I_z$	Tasks that require cooling water or produce at least one zero wait state	
$s \in S_i(i)$	States that are consumed by task i	
$s \in S_o(i)$	States that are produced by task i	
Variables		
$Approach(n, p)$	Cooling tower approach temperature	$°C$
$B(n, p)$	Blowdown from cooling tower n at time point p	t/h
$B_f(i, p)$	Batch size of task i that finishes at or before time point p	t
$B_p(i, p)$	Batch size of task i that is being processed at time point p	t
$B_s(i, p)$	Batch size of task i that starts at time point p	t
$B_I(i, s, p)$	Amount of state s used as input for task i at time point p	t
$B_O(i, s, p)$	Amount of state s produced from task i at or before time point p	t
$c_{PO}(p)$	Operational costs during time point p	\$
$c_{VC}(n)$	Variable portion of capital costs associated with cooling tower n	\$
$c_{VC_P}(n, p)$	Variable portion of capital costs associated with cooling tower n at time point p	\$
c_{TC}	Total capital costs	\$
c_{TO}	Total operational costs	\$

c_{TR}	Total raw material costs	\$
$C_R(n,i,p)$	Return cooling water to cooling tower n from task i at time point p	t/h
$C_S(n,i,p)$	Cooling water supplied from cooling tower n to task i at time point p	t/h
CT	Total number of active cooling towers	
CW	Total cooling water flow supplied from all cooling towers	t/h
$D(n,p)$	Drift loss from cooling tower n at time point p	t/h
$E(n,p)$	Evaporation loss from cooling tower n at time point p	t/h
$F_{in}(i,p)$	Cooling water entering task i at time point p	t/h
$F_{out}(i,p)$	Cooling water exiting task i at time point p	t/h
$F_r(i',i,p)$	Reused cooling water from task ii to task i at time point p	t/h
$M(n,p)$	Makeup water supplied to cooling tower n at time point p	t/h
$OS(n,p)$	Total cooling water supplied from cooling tower n at time point p	t/h
$Profit_L$	Linearised total profit	\$
$Profit_P$	Total production profit, without considering utilities	\$
$Profit$	Total profit	\$
$Q_i(i,p)$	Amount of cooling duty provided to task i from time point p	kW
$Q_o(i,p)$	Amount of cooling duty provided to task i until time point p	kW
$Q_u(i,p)$	Amount of cooling duty utilised by task i at time point p	kW
$R(n',n,p)$	Cooling water recycled from cooling tower n' to cooling tower n at time point p	t/h
$Range(n,p)$	Cooling tower range temperature	$^\circ C$
$S_A(s,p)$	Amount of state s available at time point p	t
$S_S(s,p)$	Sales of state s at time point p	t
$t(p)$	Time that corresponds to time point p	h
$t_s(i,p)$	Start time of task i that starts at time point p	h
$t_f(i,p)$	Finish time of task i that starts at time point p	h
$T_{in}(i,p)$	Cooling water temperature into task i at time point p	$^\circ C$
$T_{ret}(n,p)$	Return temperature to cooling tower n at time point p	$^\circ C$
$T_{sup}(n,p)$	Supply temperature from cooling tower n at time point p to cooling water network	$^\circ C$
$T_{out}(i,p)$	Outlet cooling water temperature from task i at time point p	$^\circ C$

$W_f(i,p)$	Binary variable indicating whether task i finishes at or before time point p	
$W_p(i,p)$	Binary variable indicating whether task i is being processed at time point p	
$W_r(i',i,p)$	Binary variable indicating whether tasks i' and i both take place during time point p	
$W_s(i,p)$	Binary variable indicating whether task i starts at time point p	
$y_{VC}(n)$	Binary variable associated with variable cooling tower costing	
$y_{CT}(n)$	Binary variable indicating activity of cooling tower n	
$y_r(i',i,p)$	Binary variable indicating whether cooling water is reused by task i from task i' during time point p	
$Z_f(j,p)$	Binary variable indicating whether a task in $I(j)$ assigned to unit j finishes at or before time point p	
$Z_p(j,p)$	Binary variable indicating whether a task in $I(j)$ assigned to unit j is being processed at time point p	
$Z_s(j,p)$	Binary variable indicating whether a task in $I(j)$ assigned to unit j starts at time point p	
$\gamma_1(n,i,p)$	Linearisation variable for $C_R(n,i,p)T_{out}(i,p)$	
$\gamma_2(i',i,p)$	Linearisation variable for $F_R(i',i,p)T_{out}(i',p)$	
$\gamma_3(i,p)$	Linearisation variable for $F_{in}(i,p)T_{out}(i,p)$	
$\gamma_4(n,i,p)$	Linearisation variable for $C_S(n,i,p)T_{sup}(n,p)$	
$\gamma_5(n,i,p)$	Linearisation variable for $C_R(n,i,p)T_{ret}(n,p)$	
$\gamma_6(n',n,p)$	Linearisation variable for $R(n',n,p)T_{sup}(n,p)$	
$\gamma_7(n',n,p)$	Linearisation variable for $R(n',n,p)T_{ret}(n,p)$	
$\tau(i,p)$	Duration of task i that starts at time point p	h
Parameters		
$B^L(i)$	Lower bound of the batch size of task i	t
$B^U(i)$	Upper bound of the batch size of task i	t
BM	Large number for big-M constraints	
$C_{RS}^U(n)$	Maximum cooling water flow supplied from or returned to cooling tower n	t/h
c_p	Specific heat capacity of water	$J/kg°C$
c_{FC}	Fixed cost associated with existence of a cooling tower	$\$/CT$
$c_{RM}(s)$	Cost of raw material state s	$\$/t$
CC	Cycles of concentration	
$F_{in}^U(i)$	Maximum cooling water flowrate through task i	t/h
H	Time horizon	h
H_Y	Yearly operating time	h

M_T	Total amount of makeup water utilised over the time horizon	t
$OS^U(n)$	Design capacity of cooling tower n	t/h
$Q(i)$	Heat of reaction for task i	kW/t
$S_A^0(s)$	Initial amount of state s available	t
$S_A^U(s)$	Maximum storage capacity of state s	t
T_{amb}	Ambient temperature	$^\circ C$
$T_{ct}(n)$	Cooling water supply temperature from cooling tower n	$^\circ C$
T_{ct}^L	Minimum cooling water supply temperature	$^\circ C$
$T_{in}^U(i)$	Limiting inlet temperature to task i	$^\circ C$
T_{out}^L	Lower limit of outlet cooling water temperature	$^\circ C$
$T_{out}^U(i)$	Upper limit of outlet cooling water temperature from task i	$^\circ C$
T_{ret_A}	Average return temperature to the set of cooling towers over the time horizon	$^\circ C$
$T_{ret}^U(n)$	Maximum return temperature to cooling tower n	$^\circ C$
T_{wb}	Wet bulb temperature	$^\circ C$
$\alpha(i)$	Fixed duration of task i	h
$\beta(i)$	Variable duration of task i	h
$\epsilon(n,p)$	Effectiveness of cooling tower n at time point p	%
$\epsilon_p(p)$	Overall effectiveness of the cooling water system at time point p	%
ϵ_A	Overall effectiveness of the cooling water system averaged over the time horizon	%
$\rho(i,s)$	Mass balance coefficient for the consumption or production of state s in task i	
$\zeta(s)$	Price of state s	$\$/t$

Chapter 1

Introduction

This chapter provides a background to the problem that this investigation aimed to solve. The underlying motivation is explored. The research scope and objectives are then described. Finally, the problem is formalised into a problem statement.

1.1 Background

The water and energy intensive nature of chemical processes has contributed to the depletion of the Earth's natural resources. South Africa is a historically dry country, experiencing an average rainfall of only 500 mm per annum, which is 60% of the global average (Mukheibir and Sparks, 2003). It has been determined that Southern Africa experiences moderate to severe water scarcity for more than half of the year (Mekonnen and Hoekstra, 2016). These findings are demonstrated in Figure 1.1, which depicts the number of months each region experiences blue water scarcity. In this context, blue water refers to fresh surface water and groundwater. Population growth will exacerbate the problem, as there will be an increased demand for limited natural resources. This is of particular concern in Africa, where the population is expected to grow rapidly from 1.256 billion in 2017 to 4.468 billion in 2100 (United

Nations, 2017). It is thus important to use our natural resources sparingly. In the field of chemical engineering, process integration aims to exploit the interactions between units to ensure the efficient use of resources.

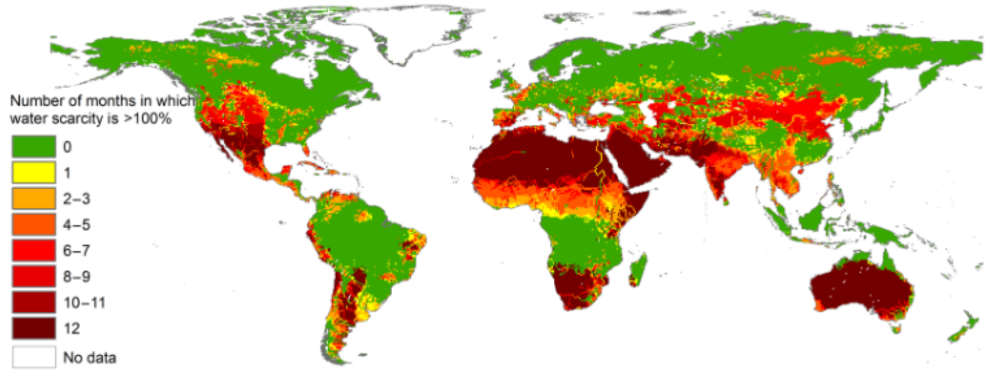


Figure 1.1: The number of blue water scarce months 1996–2015 (Mekonnen and Hoekstra, 2016)

Process integration has become a key research area in the field of chemical engineering since the 1970's energy crisis. Industry was forced to develop innovative methods to reduce energy requirements. Hohmann (1972) investigated the trade-off between heat exchanger area and their associated utility cost. The most successful and widely applied technique developed during this time was termed pinch analysis (Linnhoff and Hindmarsh, 1983).

Insights from pinch analysis were later used to develop methods which could reduce waste and wastewater streams at the outlet of the plant (El-Halwagi and Manousiouthakis, 1989; Wang and Smith, 1994). These methods have been widely applied in industry, both for grassroots and retrofit applications.

The integration of utility systems, which are responsible for meeting heating and cooling demands throughout the plant, has received increasing academic interest. Cooling water systems operate by transferring waste heat from hot processes to cooling water. This water then passes through a cooling tower, where the waste heat is rejected into the environment.

The cooling water system is made up of cooling tower(s), the pumping network, and the cooling water network. The cooling water network consists of hot process streams which require cooling, through cooling water using operations. These operations have historically been arranged in parallel with one another. Thus each cooling water using operation is supplied with cooling water directly from the cooling tower. Figure 2.20 provides a visual illustration of a cooling water system layout.

Cooling towers operate through the mechanism of evaporative cooling. Cooling towers can be divided into two major categories, based on the method used to supply air to the cooling tower. Mechanical draft cooling towers ensure contact between the water and air by means of a fan. The fan can be placed at the top or bottom of the cooling tower. These are known as induced draft and forced draft cooling towers respectively.

Natural circulation cooling towers operate without a mechanical fan. An example of which is the atmospheric cooling tower. These towers allow air to enter through the sides of the cooling tower, thereby cooling the water. Another example is the natural draft cooling tower, which heats the air in the tower by using the heat of the hot water. This lowers the air density, resulting in colder air from outside the tower to flow into the tower. The hot air is then rejected at the top of the tower (Kern, 1950).

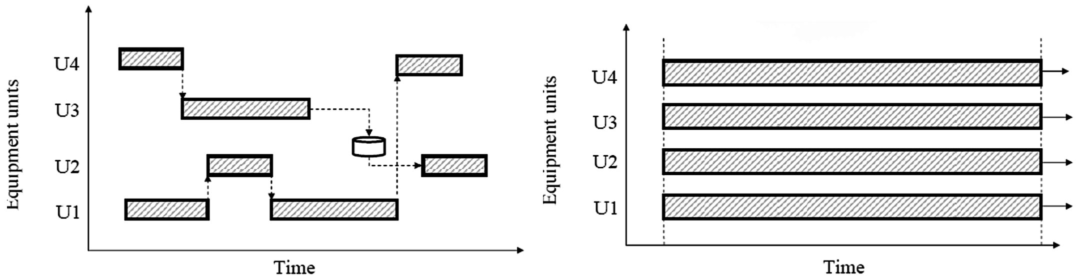
Further, research has investigated the structural reliability of cooling towers (Milford and Schnobrich, 1986; Sudret et al., 2005), optimum sizing of cooling towers (Söylemez, 2001), loss coefficient of various cooling tower fills (Kloppers and Kröger, 2003), heat transfer characteristics (Kloppers and Kröger, 2005), water loss reduction in wet-cooling towers (Lefevre, 1984), the effect of wind on dry-cooling towers (Du Preez and Kröger, 1993) and the effect of ambient air conditions on cooling tower performance (Papaefthimiou et al., 2012).

Initially, research on cooling towers did not consider the cooling water network. The effect of the cooling water network on cooling tower performance was first investigated by Bernier (1994). It was found that cooling tower performance is improved when operating at high cooling water return temperatures and low cooling water flowrates.

Kim and Smith (2001) used this insight in studying the combined mass and heat integration of a cooling water system. They developed a graphical technique to optimise cooling tower performance. Majozi and Moodley (2008) extended their work by developing a mathematical model to minimise the total recirculating cooling water flowrate with multiple cooling towers. The use of multiple cooling towers is common in industrial applications.

Reuse of cooling water between hot processes within the cooling water network naturally leads to an increase in the network pressure drop. This necessitates an increase in the pumping power requirement. Kim and Smith (2003) were the first to attempt the minimisation of pressure drop in the cooling water network. They developed a mathematical model to minimise the network pressure drop based on the critical path analysis (CPA). Their work was later extended to multiple cooling towers (Gololo and Majozi, 2012). The CPA technique was later applied to steam systems to minimise network pressure drop (Price and Majozi, 2010).

The aforementioned cooling water network models operate as continuous processes under steady-state conditions. Very little research has focused on integrating batch chemical processes with utility systems. Batch processes are composed of discrete tasks which follow a set sequence, by which the raw materials are converted to products. One can distinguish batch processes from continuous processes, in that batch operations are active intermittently throughout the time horizon of interest. Conversely, continuous operations are active throughout the time horizon of interest. This distinction is graphically depicted by Figure 1.2.



(a) Batch (b) Continuous

Figure 1.2: Tasks throughout a time horizon for a) batch and b) continuous processes (Majozi, 2010)

Batch processes are becoming increasingly popular due to the flexibility associated with shared equipment, especially in volatile market conditions. Batch processes are well suited to produce low volume complex or specialty chemicals of high value, examples of which can be found in breweries, dairies, paint, food, biochemical, pharmaceutical and agrochemical facilities.

The integration of utility systems with batch processes followed from the design of utility systems under uncertainty (Iyer and Grossmann, 1997). Recent works have integrated batch process with steam systems, which provide for hot utilities (Agha et al., 2010; Behdani et al., 2007). Leenders et al. (2019) integrated a utility system, providing hot and cold utilities, with a batch process. Their utility system did not include a cooling tower. Absorption and compression chillers were used as the cooling media.

Presented in this dissertation is the integrated optimisation of a cooling tower with a batch background process, allowing for cooling water reuse opportunities to be exploited. The benefits of this research for retrofit design include the reduction of makeup water and increasing cooling tower availability. For grassroots design, the capital expenditure for cooling tower capacity is generally reduced.

1.2 Research Motivation

This work is founded on earlier research by Kim and Smith (2001). Their research used the insights from Bernier (1994) to design a cooling water network with a parallel configuration where cooling water reuse is permitted between unit operations. This contrasts the traditional series arrangement where cooling water is supplied directly from cooling towers to each cooling water using operation only. The difference between the parallel and series configurations are graphically depicted by reconfiguring two heat exchangers in Figure 1.3.

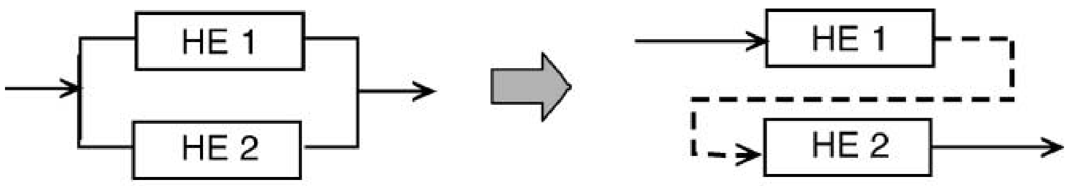


Figure 1.3: Parallel and series cooling water network configuration (Kim and Smith, 2001)

The series cooling water network arrangement leads to a reduction in the required cooling water flowrate and increase in the cooling water return temperature. Both of these conditions lead to improved cooling tower effectiveness as demonstrated by Figure 1.4.

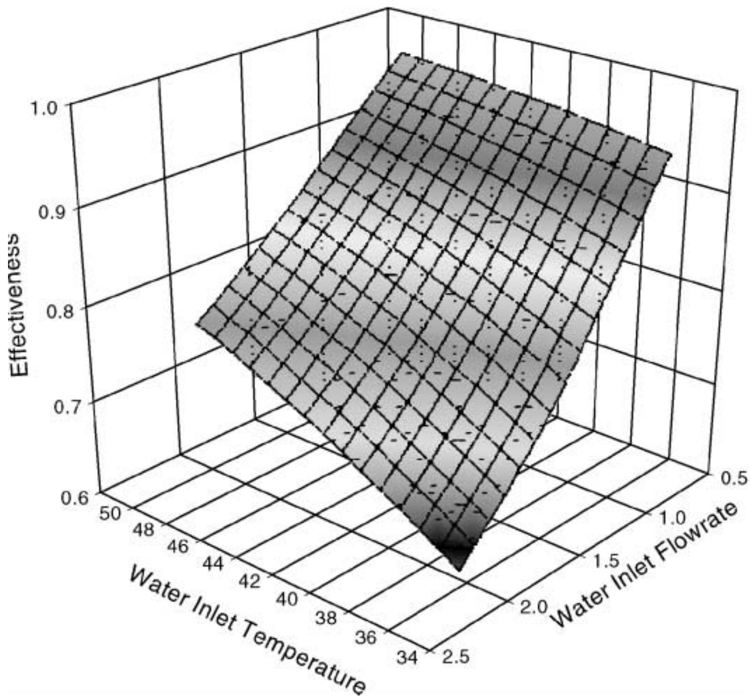
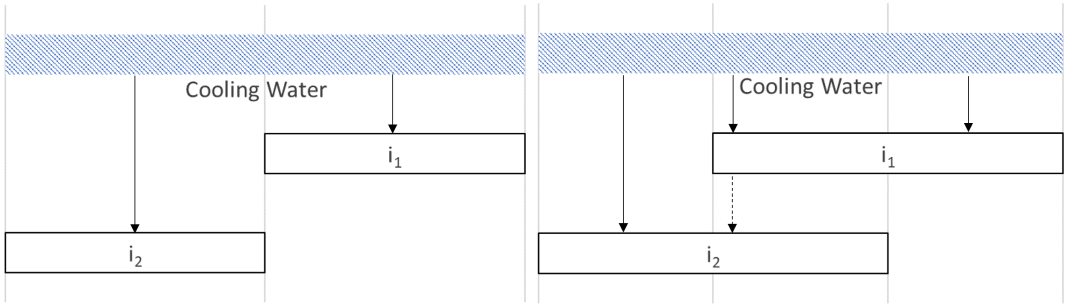


Figure 1.4: Cooling tower effectiveness (Kim and Smith, 2001)

One of the limitations of the current literature available on the synthesis and design of cooling water networks is the assumption that all cooling water using operations operate as continuous processes. Recently, researchers have started exploring utilities integration considering batch background processes. However, the focus of these models appear to be on hot utilities and electricity integration. Additionally, current batch utilities integration models do not consider utility reuse between tasks. The value of utilities integration in batch production facilities is that the production schedule can be adapted to significantly affect utility requirements during the time horizon. The schedule can also be adapted such that tasks which have the potential for utilities integration are active at the same time. This leads to a reduction in utility peak loads, as demonstrated by Figure 1.5.



(a) Original production schedule (b) Adjusted production schedule

Figure 1.5: Impact of production schedule on cooling water reuse

Through the integrated optimisation of the batch schedule and the cooling tower operations, potential savings are unlocked by exploring the interaction of the variables of the two systems. In light of this, the work presented in this dissertation aims to develop a mathematical model for the optimal synthesis and design of a cooling water system operating with a batch background process, where the production schedule is flexible and cooling water reuse is permitted between tasks.

1.3 Scope of Work

The scope of this research is to develop a mathematical model for the optimum synthesis of a cooling tower network operating with batch cooling water using operations. In achieving this the production schedule is first fixed while the cooling water network model is developed and applied to the fixed production schedule. Thereafter a scheduling model is introduced to the cooling water network model. This allows the production schedule to be flexible. The overall model is first solved sequentiallly, i.e. the production schedule is solved independently, maximising the profit associated with production activities only. Thereafter the cooling water network model is applied to the solved production schedule, maximising the overall plant profit. The overall model is also solved in an integrated manner, meaning that the production schedule and the cooling water network is solved in one step maximising the overall plant profit. This provides the opportunity to evaluate the interaction between cooling towers and the cooling water network when the cooling water using operations operate as batch tasks. It also allows for the study of the effect of the cooling water system on the batch production schedule and the contrast between the sequential and integrated optimisation approach.

1.4 Research Objectives

The above discussion highlights the aim of this dissertation, the objectives of which can be summarised as follows.

- Develop a mathematical model which integrates a cooling water network with a multipurpose batch background process.
- Integrate a batch scheduling model with the developed cooling water network.
- Contrast the sequential optimisation approach and the integrated optimisation approach.
- Determine the optimal schedule and operating conditions of the batch network.
- Validate the proposed model with an illustrative example to show the practicality of the developed model.

1.5 Problem Statement

The problem considered in this dissertation can be defined as follows.

- Given:
  (i) Production scheduling data, including the production recipe, unit capacities, material storage capacity, task durations, and the time horizon of interest.

(ii) A set of cooling towers with design specifications, such as maximum design capacity, cooling tower outlet temperature and maximum return temperature.

(iii) A set of cooling water using operations operating in batch mode with limiting temperature requirements and heat duties.

- Determine the maximum annualised profit, considering raw material, capital, and operational costs by exploring a variable production scheduling and reuse opportunities between batch cooling water using operations.

1.6 Dissertation Structure

The remainder of this dissertation is structured as follows.

- Chapter 2 is the literature review, where a detailed survey of the research relevant to this dissertation is discussed. A broad outline of process integration, mathematical optimisation, cooling towers, batch processes, and utilities integration in batch processes is provided.

- The general mathematical model is then formulated in Chapter 3 where the constraints involved in the model are discussed in detail. This chapter includes the scheduling model, cooling water network model, and the linearised model.

- In Chapter 4 the mathematical model is applied to illustrative examples to validate the model and highlight its practical relevance.

- Finally, a conclusion is drawn in Chapter 5 summarising the key findings of the research which has been conducted. A few recommendations are made, highlighting the limitations of this investigation and indicating opportunities for future work.

- References are provided at the end of each chapter.

1.7 References

Agha, M. H., Théry, R., Hetreux, G., Haït, A. and Le Lann, J. M. (2010) 'Integrated production and utility system approach for optimizing industrial unit operations'. *Energy*, 35 (2), pp. 611–627.

Behdani, B., Pishvaie, M. R. and Rashtchian, D. (2007) 'Optimal scheduling of mixed batch and continuous processes incorporating utility aspects'. *Chemical Engineering and Processing: Process Intensification*, 46 (4), pp. 271–281.

Bernier, M. A. (1994) *Cooling tower performance: theory and experiments*. American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., Atlanta, GA (United States).

Du Preez, A. F. and Kröger, D. G. (1993) 'Effect of wind on performance of a dry-cooling tower'. *Heat Recovery Systems and CHP*, 13 (2), pp. 139–146.

Gololo, K. V. and Majozi, T. (2012) 'Complex cooling water systems optimization with pressure drop consideration'. *Industrial & Engineering Chemistry Research*, 52 (22), pp. 7056–7065.

El-Halwagi, M. M. and Manousiouthakis, V. (1989) 'Synthesis of mass exchange networks'. *AIChE Journal*, 35 (8), pp. 1233–1244.

Hohmann, E. C. J. (1972) 'Optimum networks for heat exchange.' PhD thesis. University of Southern California.

Iyer, R. R. and Grossmann, I. E. (1997) 'Optimal multiperiod operational planning for utility systems'. *Computers & Chemical Engineering*, 21 (8), pp. 787–800.

Kern, D. Q. (1950) *Process heat transfer*. Tata McGraw-Hill Education.

Kim, J. K. and Smith, R. (2001) 'Cooling water system design'. *Chemical Engineering Science*, 56 (12), pp. 3641–3658.

Kim, J. K. and Smith, R. (2003) 'Automated retrofit design of cooling-water systems'. *AIChE journal*, 49 (7), pp. 1712–1730.

Kloppers, J. C. and Kröger, D. G. (2003) 'Loss coefficient correlation for wet-cooling tower fills'. *Applied Thermal Engineering*, 23 (17), pp. 2201–2211.

Kloppers, J. C. and Kröger, D. G. (2005) 'A critical investigation into the heat and mass transfer analysis of counterflow wet-cooling towers'. *International Journal of Heat and Mass Transfer*, 48 (3-4), pp. 765–777.

Leenders, L., Bahl, B., Lampe, M., Hennen, M. and Bardow, A. (2019) 'Optimal design of integrated batch production and utility systems'. *Computers & Chemical Engineering*.

Lefevre, M. R. (1984) 'Reducing water consumption in cooling towers'. *Chemical Engineering Progress*, 80 (7).

Linnhoff, B. and Hindmarsh, E. (1983) 'The pinch design method for heat exchanger networks'. *Chemical Engineering Science*, 38 (5), pp. 745–763.

Majozi, T. (2010) *Batch chemical process integration: analysis, synthesis and optimization*. Springer Science & Business Media.

Majozi, T. and Moodley, A. (2008) 'Simultaneous targeting and design for cooling water systems with multiple cooling water supplies'. *Computers & Chemical Engineering*, 32 (3), pp. 540–551.

Mekonnen, M. M. and Hoekstra, A. Y. (2016) 'Four billion people facing severe water scarcity'. *Science Advances*, 2 (2).

Milford, R. V. and Schnobrich, W. C. (1986) 'The effect of cracking on the ultimate strength of reinforced concrete cooling towers'. *Engineering Structures*, 8 (1), pp. 46–54.

Mukheibir, P. and Sparks, D. (2003) *Water resource management and climate change in South Africa: visions, driving factors and sustainable development indicators.* Report for Phase I of the Sustainable Development, Climate Change project. Energy and Development Research Centre (EDRC), University of Cape Town.

Papaefthimiou, V. D., Rogdakis, E. D., Koronaki, I. P. and Zannis, T. C. (2012) 'Thermodynamic study of the effects of ambient air conditions on the thermal performance characteristics of a closed wet cooling tower'. *Applied Thermal Engineering*, 33, pp. 199–207.

Price, T. and Majozi, T. (2010) 'On synthesis and optimization of steam system networks. 3. Pressure drop consideration'. *Industrial & Engineering Chemistry Research*, 49 (19), pp. 9165–9174.

Söylemez, M. S. (2001) 'On the optimum sizing of cooling towers'. *Energy Conversion and Management*, 42 (7), pp. 783–789.

Sudret, B., Defaux, G. and Pendola, M. (2005) 'Time-variant finite element reliability analysis–application to the durability of cooling towers'. *Structural Safety*, 27 (2), pp. 93–112.

United Nations (2017) *World population prospects: The 2017 revision, key findings and advance tables.* United Nations.

Wang, Y. P. and Smith, R. (1994) 'Wastewater minimisation'. *Chemical Engineering Science*, 49 (7), pp. 981–1006.

Chapter 2

Literature Review

This chapter provides a detailed review of the literature relevant to the dissertation. The major sections reviewed include process integration, optimisation, cooling towers, batch processes and the integration of batch and utility systems.

2.1 Process Integration

Process integration aims to unlock improvements in a specified objective function, typically maximising profit or minimising costs, by exploiting the interaction between various units or systems. Chemical engineering applications include heat integration, mass integration and wastewater minimisation.

2.1.1 Heat Exchange Networks

The design of chemical processes focused predominantly on the core reaction and separation processes during the 1960's. The low cost of oil allowed the chemicals industry to make use of external hot utilities for the heating of practically all cold streams. Thereby neglecting any process-process heat transfer opportunities. The energy crisis of the 1970's gave rise to new research interest in the area of process integration techniques (Jiménez-González, 2016). The aim was to reduce the energy consumption of chemical plants by utilising residual heat that has already been generated in the process.

Hohmann (1972) utilised a graphical technique to investigate the trade-off between heat exchanger area and utility cost. The author also introduced the N-1 rule to predict the minimum number of heat exchangers required for a given system. Umeda et al. (1979) targeted minimum utility flowrate by representing all of the hot and cold stream data as a single pair of hot and cold composite curves on a temperature versus heat load graph.

Linnhoff and Hindmarsh (1983) incorporated aspects from various heat integration techniques developed throughout the late 70's and early 80's into a single simple procedure broadly referred to as pinch analysis. Pinch analysis is a conceptual methodology which targets the minimum energy consumption by maximising energy recovery between hot and cold streams within the process (Linnhoff and Flower, 1978a). The overall heat load of all streams in the temperature interval is depicted on a temperature versus heat load graph, referred to as composite curves. Composite curves were introduced by Huang and Elshout (1976) to represent heating and cooling utility requirements.

Figure 2.1 depicts a hot and cold composite curve. The pinch point and need for external heating and cooling utilities have clearly been illustrated. Linnhoff and Flower (1978a) introduced the problem table algorithm as a mathematical approach to locate the pinch temperature and target minimum utility duties. The algorithm relies on a cascaded heat balance performed within multiple temperature intervals based on supply and target temperatures. The temperature interval definition depends on the predefined minimum temperature difference (ΔT_{min}). The optimum ΔT_{min} value is determined by cost analysis, taking into account the interaction of the capital and operational costs. As demonstrated in Figure 2.2, a trade-off exists where capital costs

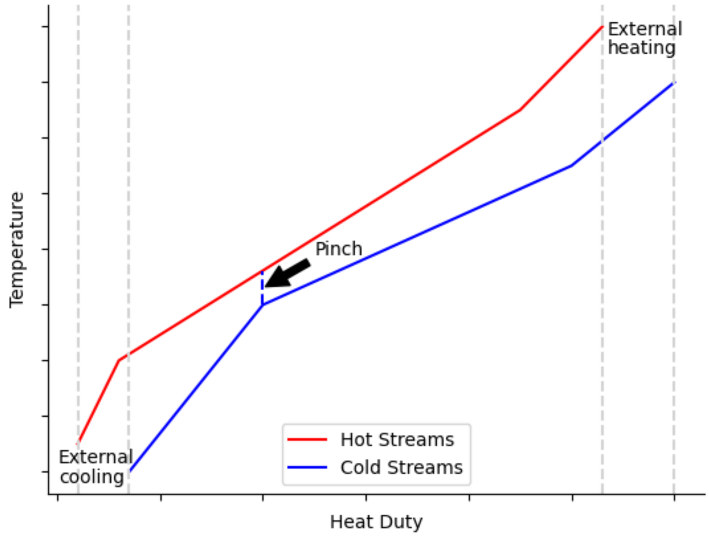


Figure 2.1: Composite curve

decrease with higher ΔT_{min}, since significantly greater amounts of capital is required to bring about improved heat transfer past the optimum point. However, operating costs scale linearly with increases in ΔT_{min}, since more utilities are required when less process-process heat integration takes place.

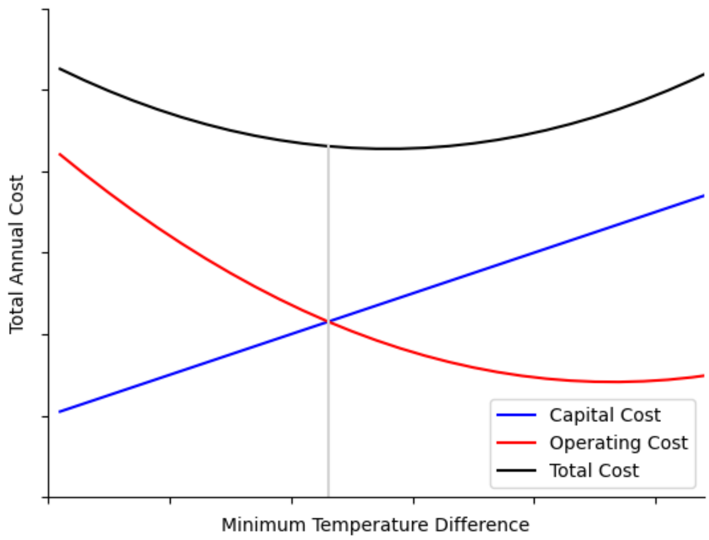


Figure 2.2: Cost curve

Whenever the temperature difference between the hot and cold stream is equal to the minimum temperature difference, a pinch is formed. The pinch represents the most bottlenecked region in the design of the heat exchanger network. The pinch point is clearly demonstrated by the grand composite curve, depicted by Figure 2.3. The main advantage of the grand composite curve is the ability to incorporate multiple utilities at varying temperatures. Pockets where process-process heat integration opportunities exist is also clearly demonstrated by the grand composite curve. In Figure 2.3 these regions are depicted by vertical grey lines. The pinch point divides the graph into two parts, where the heat surplus region is below the pinch and the heat deficit region is above the pinch.

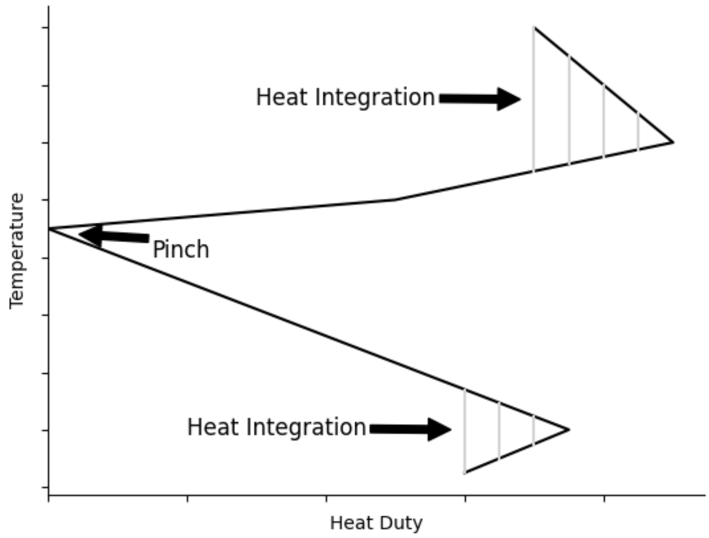


Figure 2.3: Grand composite curve

The heat exchanger network design starts at the pinch, allowing essential stream matches to be made at the most constrained region, and is then developed away from the pinch in two directions, the hot and cold side. Linnhoff and Flower (1978b) developed a grid diagram representing process streams with horizontal lines. This allows for a simplified visual depiction of stream matching and splitting in the HEN. Larger production facilities typically subdivide the plant into various subregions for process integration. This helps eliminate impractical matches due to distance constraints which would necessitate costly additional piping.

2.1.2 Mass Exchange Networks

The rising costs of waste treatment, due to stricter environmental regulation, led to the incorporation of pollution prevention techniques into the process synthesis stage. Mass exchange networks provided a method to transfer valuable pollutants from the waste stream to feed streams, by means of mass transfer operations. The process postulates the exchange of a component from a set of rich streams to lean streams. Various mass separating agents can be selected as a lean stream. Thermodynamic pinch points are identified, which limits the mass transfer between the rich and lean streams. The design of the preliminary network then starts at the pinch and is developed moving away from the pinch (El-Halwagi and Manousiouthakis, 1989). The heat integration techniques were readily extendible to mass integration due to the structural similarity between the governing equation for mass transfer and heat integration. Both Fourier's and Fick's Law are directly proportional to the product of a constant and the driving force causing the phenomena. For this reason, the mass integration pinch is determined by the concentration driving force, as opposed to the temperature driving force as in the case of heat integration. Mass integration is applicable to processes such as absorbers and extractors. This procedure was later automated, by making use of a mathematical model, by El-Halwagi and Manousiouthakis (1990).

2.1.3 Wastewater Minimisation

Further environmental regulations were imposed on wastewater that is discharged into the environment. Wastewater minimisation can be viewed as a specific application of mass integration where water is the only lean stream. Chemical processes produce wastewater when process water comes into contact with contaminants. This typically occurs when water is used for washing operations, liquid-liquid extraction and scrubbing. Wastewater which exceeds the maximum limit of contaminant concentration must undergo costly treatment processes, in operations such as filters and membranes, before being discharged.

Conceptual techniques to minimise wastewater generation, and consequently reduce freshwater demand, were developed by extension of the mass exchange network. The pioneering work in this field is the water pinch, developed by Wang and Smith (1994b). The authors extended the central concepts of pinch analysis to construct a composite curve made up of the limiting concentration profiles of all streams. The composite curve is drawn on a concentration vs mass load plot as demonstrated by Figure 2.4.

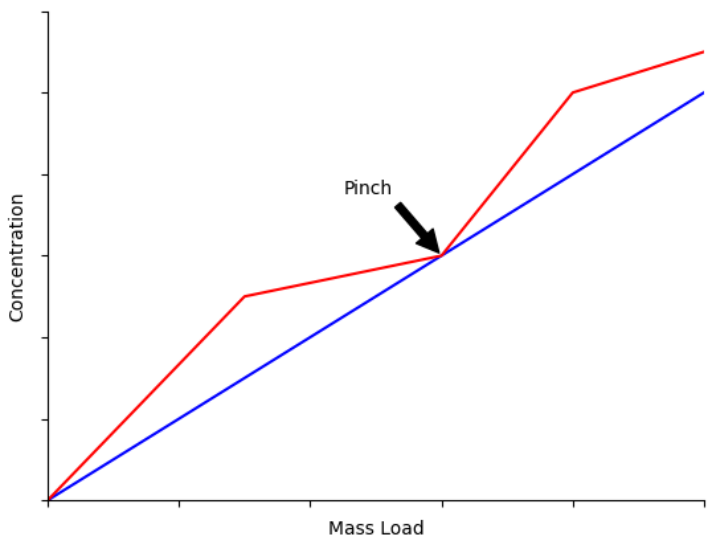


Figure 2.4: Water pinch

The water supply line is drawn as a straight line extending from the origin through the pinch point. The gradient of the water supply line is inversely proportional to the wastewater flowrate. Operations occurring below the pinch point require additional freshwater, whereas operations taking place above the pinch point show potential for water reuse. The authors extended their technique to multiple contaminant streams by consolidating a subnetwork for each contaminant into an overall design (Wang and Smith, 1994a). Wastewater was minimised by utilising water reuse, regeneration-reuse and regeneration-recycling opportunities. Despite the graphical nature of their technique, the authors incorporated constraints due to minimum mass transfer driving forces, fouling and corrosion into the overall problem statement.

Kuo and Smith (1998a) later developed a conceptual and graphical approach, called the water mains method. The mains acted as sources or sinks, dividing the water network into different regions. Operations crossing the boundary of a main is then merged to produce the final water network design. This method integrated the design of the water-using system with the effluent treatment system, thereby minimising water consumption at the lowest effluent treatment cost. The authors later extended the model to include regeneration-reuse and recycling (Kuo and Smith, 1998b).

The pioneering work in the field of water minimisation using mathematical programming is by Takama et al. (1980). They made use of a superstructure, which can be defined as a superset of all feasible solutions. Thus, their model accounted for all reuse and recycle opportunities between the water-using system and the wastewater treatment system. Prior to their research, wastewater treatment systems had been studied independently of the water-using system. They developed an NLP model to minimise wastewater, by maximising reuse and recycle within the network. In solving the problem, they transformed the overall model into a sequence of problems without inequality constraints, and applied a penalty function to the sequence.

Doyle and Smith (1997) developed a model for maximum water reuse in multi-contaminant problems. Their model included practical considerations, such as the option to forbid certain stream matches due to large distances between them or for safety concerns. The model was also formulated as an NLP. A linearisation procedure was implemented by fixing the outlet contaminant concentration. The linearised model was then solved to obtain a starting point for the NLP.

Alva-Argáez et al. (1998) implemented a similar approach to minimise the total annual cost of a water-using industrial system, accounting for pipe length, material of construction, flow velocity, etc. They made use of binary variables to represent every possible connection in the water network. In doing so, they formulated an MINLP model. To solve the MINLP problem, they decomposed the model into a sequence of MILP problems. It was later proven that the mathematical model for single contaminant water allocation problems could be linearised exactly by setting the outlet water concentration at its maximum (Savelski and Bagajewicz, 2000). This is discussed in detail in Section 2.2.2.

Gunaratnam et al. (2005) developed an MINLP model for minimising wastewater consisting of multiple components. The authors considered reuse, regeneration-reuse, and regeneration-recycle of water. A two-stage solution strategy was implemented, whereby the MINLP was decomposed into LP/MILP subproblems which was solved iteratively. The solution to the LP/MILP subproblem was used the starting point for the MINLP model.

2.2 Optimisation

Snyman and Wilke (2018) define mathematical optimisation as the science of calculating the optimal solutions to numerically defined problems, which may be representative of physical reality, manufacturing or management systems. The goal of optimisation is finding the values of variables which lead to the optimal value of the objective function. Variables, within the context of mathematical modelling, can be classified as continuous, integer or binary. Continuous variables can take any real number value within the specified variable bounds. Integer or discrete variables can take any whole number value within the specified bounds. Binary variables are a special case of integer variables, which can take a value of 0 or 1. The interrelationships between the design variables are expressed through the constraints.

Conceptual or mathematical programming techniques can be implemented in solving optimisation problems. Graphical techniques offer the designer of the problem the opportunity to assess the problem more accurately during the synthesis stage. The drawback of graphical techniques are that they are limited to two dimensions and time consuming when evaluating large problems.

Mathematical programming techniques offer more flexibility, as various constraints of practical relevance can be incorporated into the model. Examples of such constraints include pressure drop functions, mass transfer functions and forced/forbidden matches. The drawback of mathematical models is that problems are frequently formulated with nonlinear constraints for which global optimality cannot be ensured. Various linearisation techniques have been developed in literature and are discussed in Section 2.2.2.

Optimisation problems have a minimum of one degrees of freedom. Put differently, a minimum of one variable should be free to vary. The degrees of freedom of a mathematical problem can be determined as the difference between the number of variables and the number of independent constraints in the model. If the degree of freedom is zero, it is considered a simulation problem with one unique solution. If the number of independent constraints exceed the number of variables, the problem is over specified. Thus there are redundant constraints present which should be removed.

Optimisation problems consist of an objective function which is to be maximised or minimised, as well as equality and/or inequality constraints defining the problem. A set of variables which satisfies all of the constraints to an optimisation problem is referred to as a feasible solution. All possible feasible solutions to a particular mathematical problem is represented by the feasible region (Edgar et al., 2001). The optimal solution is then the particular feasible solution within the feasible region where the specific combination of design variables provide the best solution to the objective function. The typical form of a mathematical optimisation problem is illustrated by Equation 2.1 (Williams, 2013).

$$\begin{aligned} &\textbf{Objective:} && minimise f(x,y) \\ &\textbf{Subject to} && h(x,y) = 0 \\ & && g(x,y) \leq 0 \\ & && \forall x \in \mathbb{R}, y \in [0,1] \end{aligned} \tag{2.1}$$

It is clear that the objective of this particular problem is the minimisation of $f(x,y)$ subject to the equality constraints $h(x,y)$ and the inequality constraints $g(x,y)$. A linear programming problem (LP) arises when $f(x,y)$, $g(x,y)$ and $h(x,y)$ are all linear functions. If the objective function or constraints contain any non-linearities, the problem becomes a nonlinear programming problem (NLP). Furthermore, if the problem makes use of integer variables, the problem becomes either a mixed integer linear programming problem (MILP) or a mixed integer nonlinear programming problem (MINLP). An overview of the various optimisation problem types is graphically depicted by Figure 2.5.

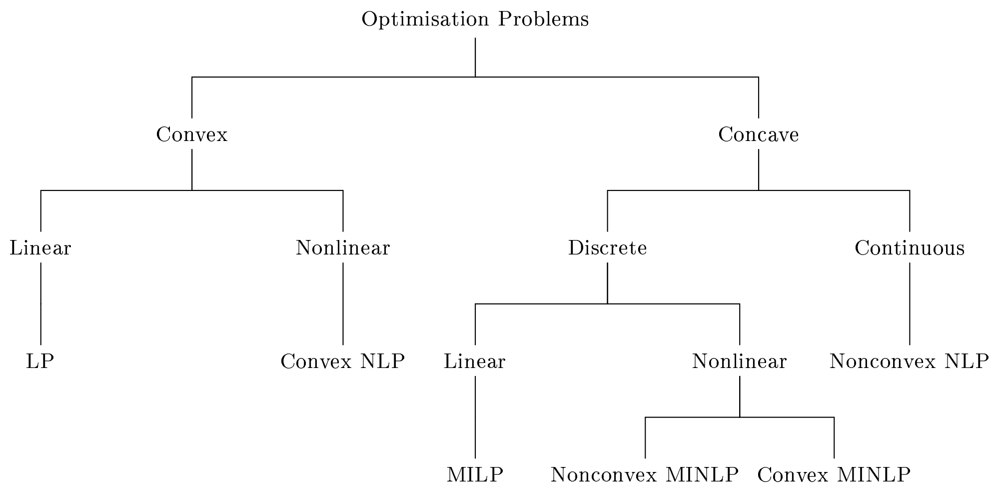
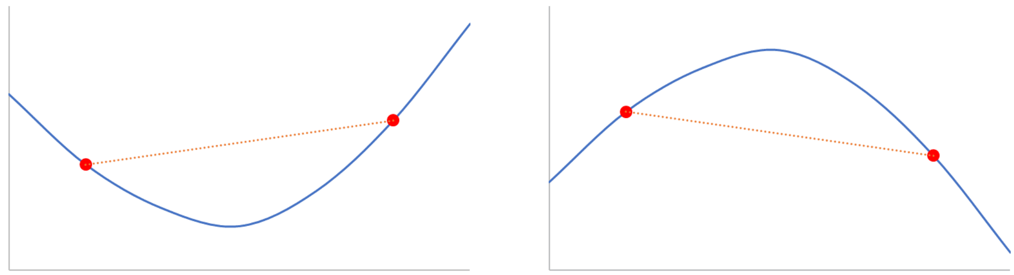


Figure 2.5: Overview of optimisation problem types (Lin et al., 2012)

2.2.1 Convexity

Mathematical problems can exhibit a convex or concave shape. One-dimensional functions are convex if all the points on an arbitry line connecting two points on the function curve is greater than or equal to the function values. Conversely, if all the points on an arbitry line connecting two points on the function curve is less than or equal to the function values, the function is said to be concave. This distinction is illustrated by Figure 2.6. Functions can further be classified as strictly convex or strictly concave. To convert a convex or concave function to a strictly convex or strictly concave function, the inclusive inequality operators should be replaced by exclusive inequality operators. This implies that the strictly concave and strictly convex functions provide a single optimum solution. Nonconvex functions often have multiple locally optimal solutions.



(a) Convex Function (b) Concave Function

Figure 2.6: Comparison of a convex and concave function

The second derivative can be used to determine the convexity of one-dimensional functions as follows

- $f''(x) \geq 0 \quad \forall x \in \mathbb{R} \quad \leftrightarrow f(x)$ is convex.
- $f''(x) > 0 \quad \forall x \in \mathbb{R} \quad \leftrightarrow f(x)$ is strictly convex.
- $f''(x) \leq 0 \quad \forall x \in \mathbb{R} \quad \leftrightarrow f(x)$ is concave.
- $f''(x) < 0 \quad \forall x \in \mathbb{R} \quad \leftrightarrow f(x)$ is strictly concave.

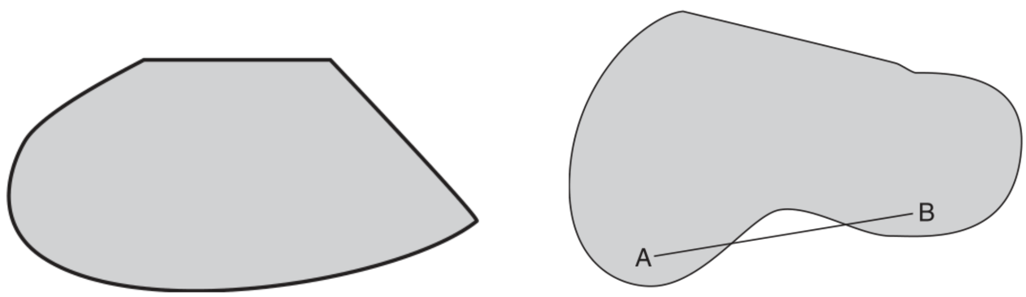
When considering convexity for multi-dimensional functions, a Hessian matrix is required to represent the second derivative. Similar conditions for determining convexity as in one-dimensional functions apply.

- $x^T H x \geq 0 \quad \forall x \in \mathbb{R} \quad \leftrightarrow f(x)$ is convex.
- $x^T H x > 0 \quad \forall x \in \mathbb{R} \quad \leftrightarrow f(x)$ is strictly convex.
- $x^T H x \leq 0 \quad \forall x \in \mathbb{R}, x \neq 0 \quad \leftrightarrow f(x)$ is concave.
- $x^T H x < 0 \quad \forall x \in \mathbb{R}, x \neq 0 \quad \leftrightarrow f(x)$ is strictly concave.

Convenient tests for strict convexity in Hessian matrices can also be implemented. The first of these conditions include that all diagonal elements of the Hessian matrix and the determinants of all leading principal minors must be positive. The second condition is that all the eigenvalues of the

Hessian matrix must be positive. Conversely, when all diagonal elements of the Hessian matrix and the determinants of all leading principal minors are negative or when all the eigenvalues of the Hessian matrix are negative the function is strictly concave.

Regions are said to be convex if an arbitrary continuous line can join any two points within the feasible region. The difference between a convex and nonconvex region is highlighted by Figure 2.7.



(a) Convex Region (b) Nonconvex Region

Figure 2.7: Comparison of a convex and nonconvex region (Williams, 2013)

For the model to be classified as convex, all the functions which make up the model should be convex. The presence of a single nonconvex function would make the entire model nonconvex.

A common challenge faced during the optimisation procedure is the presence of nonlinearities and nonconvexicity within a model. This is due to the presence of bilinear terms, logarithms and exponents. This causes the existence of multiple local minima/maxima which has the potential of trapping the optimisation algorithm and could lead to suboptimal solutions. Problems which are formulated in a convex, linear manner are relatively easy to solve and global optimality of the solution can be guaranteed (Edgar et al., 2001). Similarly, global optimality of the solution can also be guaranteed for convex nonlinear problems, although more iterations are generally required. When considering nonconvex nonlinear problems, formulations of NLP subproblems could result in multiple local optima as shown in Figure 2.8 or the MILP master problem could cut-off from the global optimum. It is therefore important that MINLP be convexified, so that global optimality of the solution can be guaranteed.

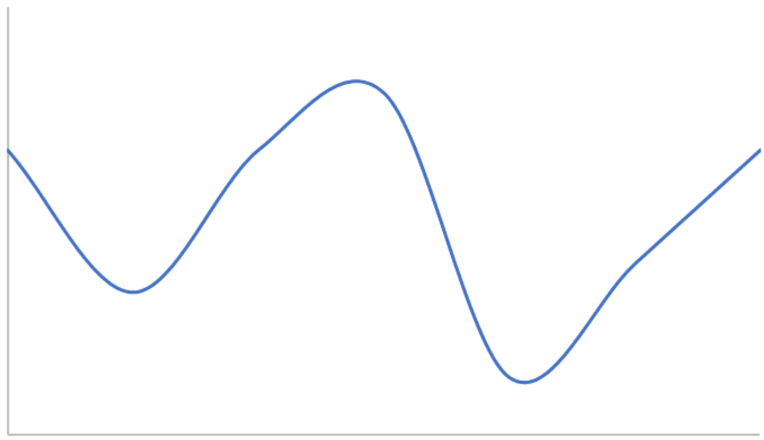


Figure 2.8: Nonconvex NLP subproblem with multiple minima

In solving nonconvex problems, convex estimators are commonly used. This approximates the nonlinear formulations to formulate lower bounding convex NLP/MINLP problems, in the case of minimisation problems, where global optimality of the solution can be guaranteed (Grossmann and Biegler, 2004). The convex estimator linearisation technique involves either directly introducing convex under estimator functions to replace each nonconvex function or by generating new variables and convex constraints by using transformations to approximate the nonconvex function exactly. Figure 2.9 depicts a convex envelope for a nonconvex function.

2.2.2 Relaxation Techniques

Relaxation techniques are implemented to solve difficult modelling problems. Relaxation techniques could be exact or inexact. Exact methods transform the original constraints to a different form, which when solved will produce the same result as the original model. Inexact methods approximate the problem by creating a similar problem that is easier to solve but yields similar

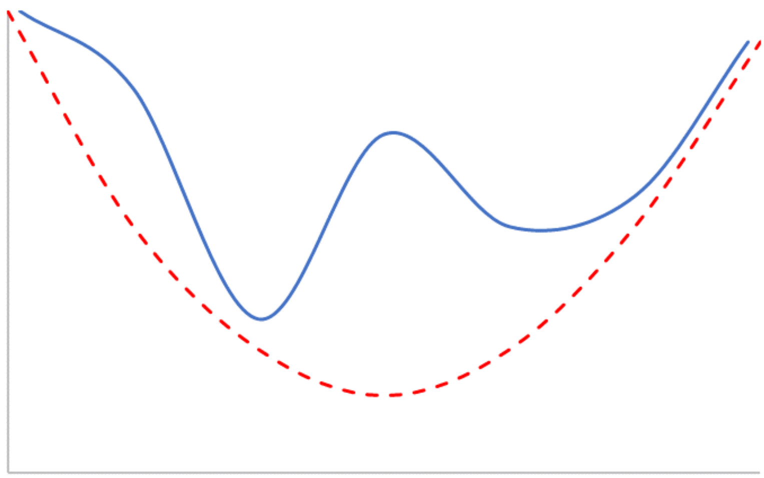


Figure 2.9: Convex envelope for a nonconvex function

results to the original model. The solution to the relaxed problem can then be used as a good starting point for the original model. It is recommended to provide nonlinear models with a good starting point, as this aids in the process of convexification which also reduces the required computational time.

The relaxation methods discussed in this section includes linear programming relaxations, direct linearisations, piecewise linearisation, Glover transforms, McCormick over and under estimators, and reformulation-linearisation techniques.

LP Relaxations

Linear programming relaxations are used to relax MILP problems by removing the integrality constraint on the variables. This results in a LP problem and transforms an NP-hard optimisation problem to a problem solvable in polynomial time. The solution to the LP problem can then be used to gain insight into the original MILP problem.

Direct Linearisation

Direct linearisation is an approach used frequently in water network synthesis problems to convert a nonlinear constraint directly into a linear constraint. This is done by using certain rules of thumb or assumptions on the optimal values of some variables and converting them into parameters. In the field of water network synthesis, Savelski and Bagajewicz (2000) introduced necessary conditions for optimality of water minimisation problems by fixing water outlet concentration of the sinks at the maximum and the concentration of the regeneration units at the minimum for all single contaminant cases. They later extended this work to include multiple contaminants (Savelski and Bagajewicz, 2003).

Direct linearisation of nonlinear models can be used to generate a good starting point for the original problem. An example of such a procedure is the work of Doyle and Smith (1997) where a sequential method that combined LP and NLP optimisation to address the water minimisation problem was developed. The exact NLP model was initialised by the solution from the LP model where outlet concentrations were fixed. This method leads to significant reductions in computational time. However, solutions obtained by this procedure may be suboptimal due to the possible elimination of streams and premature decisions made during the initialisation stage.

Piecewise Linearisation

The piecewise linear approximation technique uses linear functions to approximate nonlinear functions. Figure 2.10 depicts the piecewise linear approximation of a function.
This is achieved by partitioning the domain of a univariate function into several sections. The start or end of each section is termed a break point. The nonlinear function is then approximated by a straight line joining each successive break point. It should be noted that the accuracy of the approximation is greatly dependent on the number of break points. Piecewise linear approx-

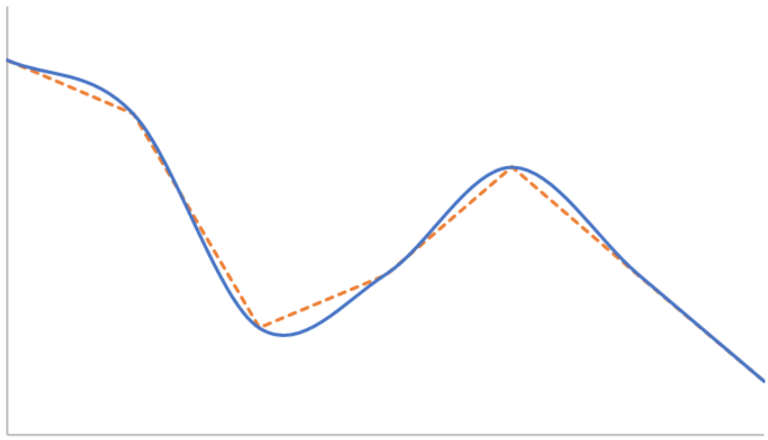


Figure 2.10: Piecewise linearisation

imation can be extended to multivariate functions, by partioning the domain of the function in multiple simplices, followed by an approximation over each simplex with an affine function. This approximation method is inexact and therefore global optimality of the solution cannot be guaranteed.

Glover Transforms

The Glover Transform was developed for the exact linearisation of bilinear terms arising from the product of a continuous and a discrete variable (Glover, 1975). Therefore, this technique does not result in any loss of accuracy and global optimality of the solution can be guaranteed. The method works by replacing the bilinear term with a variable w, as illustrated by Equation 2.2.

$$w = x \cdot y \qquad \forall x \in \mathbb{R}, y \in [0, 1] \tag{2.2}$$

Therefore, we can infer Equations 2.3 and 2.4.

$$w = 0 \qquad if \quad y = 0 \tag{2.3}$$

$$w = x \qquad if \quad y = 1 \tag{2.4}$$

Equations 2.5 and 2.6 are obtained by introducing the upper and lower bounds of continuous variable x.

$$x^L y \leq w \leq x^U y \tag{2.5}$$

$$x - x^U (1 - y) \leq w \leq x - x^L (1 - y) \tag{2.6}$$

It is clear that Equations 2.5 and 2.6 are both linear. This transform relies on the bounds on the continuous variable being known.

McCormick Over- and Underestimators

McCormick (1976) developed a method for deriving over- and underestimator functions, also referred to as convex envelopes, for bilinear terms due to the product of two continuous variables which can be integrated into global optimisation algorithms. The method works as follows, for any bilinearity due to the product of two continuous variables, the condition illustrated by Equation 2.7 holds.

$$w = x \cdot y \qquad \forall x, y \in \mathbb{R} \tag{2.7}$$

From the definition of the upper and lower bounds of each variable, we can derive Equations 2.8–2.11.

$$x - x^L \geq 0 \tag{2.8}$$

$$x^U - x \geq 0 \tag{2.9}$$

$$y - y^L \geq 0 \tag{2.10}$$

$$y^U - y \geq 0 \tag{2.11}$$

By taking the product of each different combination of Equations 2.8–2.11, and substituting in Equation 2.7 the McCormick over- and underestimators can be produced, as per Equations 2.12–2.15.

$$w \geq x^L y + y^L x - x^L y^L \tag{2.12}$$

$$w \geq x^U y + y^U x - x^U y^U \tag{2.13}$$

$$w \geq x^L y + y^U x - x^L y^U \tag{2.14}$$

$$w \geq x^U y + y^L x - x^U y^L \tag{2.15}$$

This relaxation technique is inexact. Thus, global optimality of the solution cannot be guaranteed, however this technique creates a convex solution space as the bilinear terms have all been replaced by linear terms. The bound on the solution space can potentially be weak and is subject to the bounds on the continuous variables x and y. Figure 2.11 illustrates the McCormick Over- and Underestimator technique.

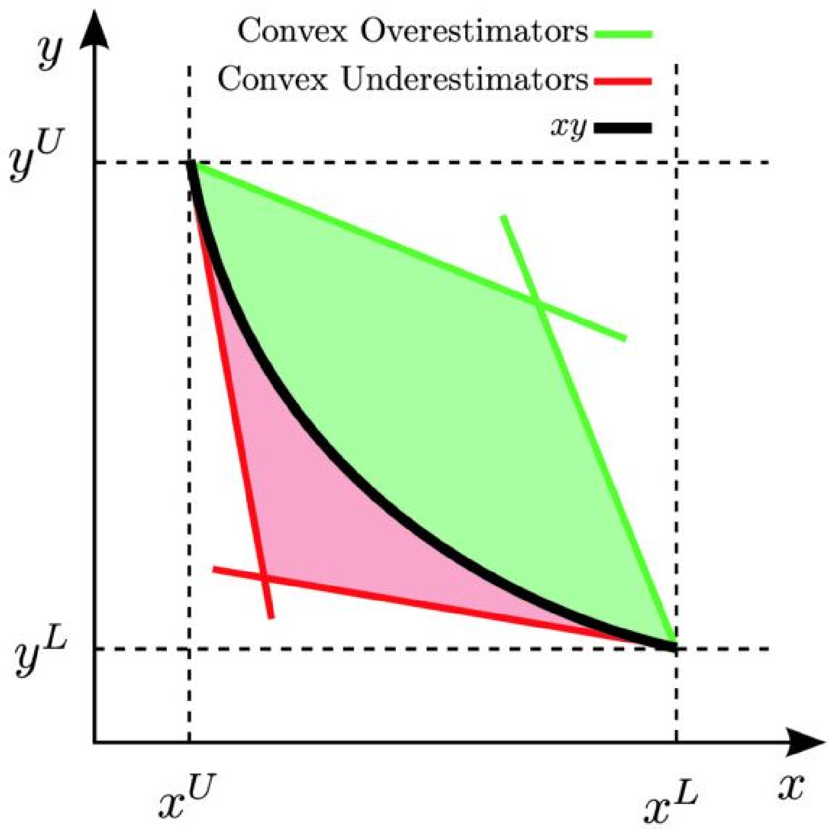


Figure 2.11: McCormick over- and underestimators

Reformulation-Linearisation Techniques

The seminal work for reformulation-linearisation techniques (RLT) for bilinear terms exploiting the special structure of the problem was by Sherali and Alameddine (1992). The procedure can be divided into two steps, namely reformulation and linearisation. The problem is first reformulated by introducing new nonlinear constraints by the multiplication of the constraints by product factors of binary variables, x, and their complements, $1 - x$. These constraints can then effectively be linearised by generating a new variable for each nonlinear constraint. The linearisation step consists of replacing the continuous variables involved in bilinear terms by McCormick (1976) under- and overestimators. Their work was later used by Quesada and Grossmann (1995) for the linear relaxation of bilinear terms by creating a convex solution space. Each bilinear term was replaced by four constraints which contained the upper and lower bounds of the continuous variables. This technique was also included in the branch and bound optimisation algorithm. It was also shown that the solution is globally optimal should the solution to the LP and NLP problem be the same. However, in general this technique is inexact. Thus, global optimality of the solution cannot always be guaranteed.

2.2.3 Stochastic Optimisation

Stochastic or metaheuristic optimisation techniques do not rely solely on the aforementioned mathematical optimisation theory. These techniques typically arrive at solutions relatively quickly and are easy to implement, however global optimality of the solutions cannot be guaranteed given a finite number of iterations. These methods perform calculations on single points, rather than over the entire region of solutions. The term 'stochastic' refers to systems which are based on the theory of probability. Randomness is an inherent aspect of stochastic optimisation procedures, as similar parameters and initial conditions may lead to significantly different outputs. These methods are typically based on physical analogies in order to generate random trial points which then mimics a progression to equilibrium (Grossmann and Biegler, 2004). The following are examples of popular stochastic optimisation techniques.

- Simulated annealing (SA) was introduced by Kirkpatrick et al. (1983) and the origins can be attributed to statistical mechanics. SA makes use of the molecular motion during the cooling of metals allowing for recrystallisation and return of the material to equilibrium in a well ordered solid state of minimal energy.

- Particle swarm optimisation (PSO), first proposed by Eberhart and Kennedy (1995) is inspired by the social behaviour of fish schooling and bird flocking.

- Ant colony optimisation (ACO) is inspired by the nature of ants to find the shortest possible path between their nest and food source. ACO was developed by Dorigo et al. (1996).

- Artificial bee colony optimisation (ABC) was first implemented by Karaboga and Basturk (2007). ABC is inspired by the intelligent foraging behaviour of honey bees. Three groups of bees are explicitly defined, namely employed bees, onlookers, and scouts. This technique has been used for the design of multiple cooling towers (Rao and Patel, 2011).

- Genetic algorithms (GA) was introduced through the work of Holland (1975). These algorithms are a subcategory of evolutionary algorithms (EA), since they draw their analogy from natural selection and reproduction where each generation passes genetic information on to the next. Apart from parental genetic information, each new generation also experiences random mutations, such that the fittest solution to a particular problem is preferentially selected. This technique has been used for batch scheduling (Woolway and Majozi, 2018).

2.2.4 Deterministic Optimisation

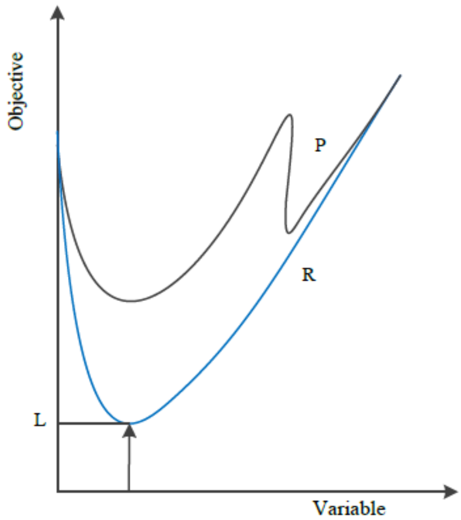
Deterministic optimisation is implemented to find global solutions, within a pre-defined level of tolerance, to optimisation problems by following a rigorous mathematical approach. The output to deterministic models are completely dependent on the dynamics of the system and determined solely by the parameters and initial conditions. Deterministic model also differ from stochastic models in that calculations are performed over regions of the solution space, rather than on single points. These methods typically decompose MINLP problems into subproblems which are then solved to global optimality.

The deterministic optimisation methods discussed in this section include the cutting plane method, branch and bound technique, and outer approximation method.

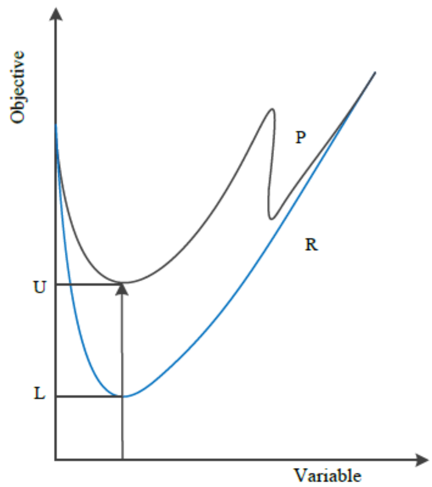
Branch and Bound

Land and Doig (1960) presented the seminal work for a branch and bound (BB) approach to discrete programming. The approach uses tree enumeration and relaxes the integer space into linear subproblems. These subproblems are then solved at each node. Branching entails the iterative subdivision of the feasible domain, whereas bounding is the calculation of the upper and lower bounds on the objective function.

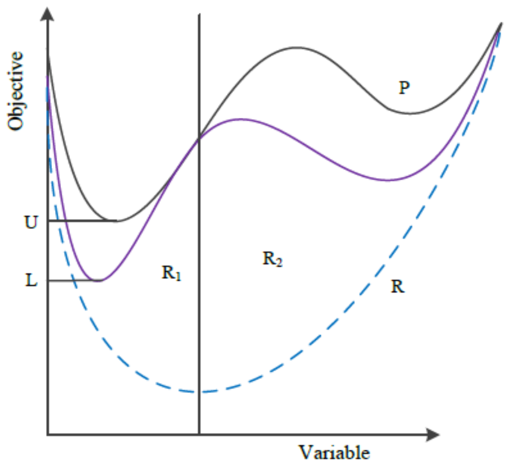
The framework of the branch and bound technique, considering a minimisation problem, is graphically depicted by Figure 2.12 where P is the nonconvex problem and R is its convex relaxation. As seen from Figure 2.12a the convex relaxation is first solved to obtain a lower bound to the problem. Branching rules are applied whilst the lower bound discrete variables remain non-integers.



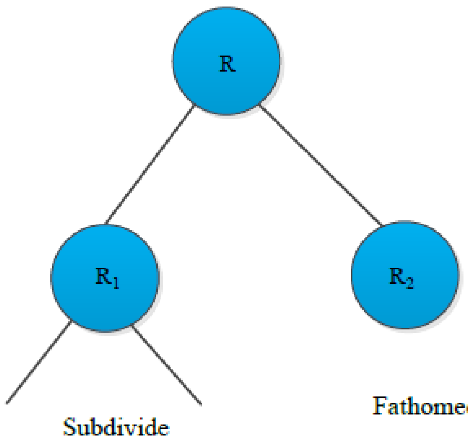
(a) Lower Bounding



(b) Upper Bounding



(c) Region Subdivision



(d) Search Tree

Figure 2.12: Spatial branch and bound procedure (Ryoo and Sahinidis, 1995)

The upper bounds are obtained from the exact evaluation of the objective function within the feasible region, as illustrated by Figure 2.12b. The difference between the problem upper and lower bounds is referred to as the relaxation gap, ϵ. The upper and lower bounds are adjusted iteratively until the relaxation gap meets a predetermined tolerance criterion.

Should the optimality criterion not be met, the value of the lower bound is updated by partitioning the feasible domain into a finite number of subregions to either side of the relaxed solution, as depicted in Figure 2.12c.

If the lower bound of a certain subregion becomes greater than or equal to the current upper bound or if the solution is infeasible, the entire node is fathomed out, as illustrated by Figure 2.12d. Branching on feasible, non-inferior nodes continue until the relaxation gap meets the specified optimality criterion. This ability to eliminate inferior and infeasible branches is the main feature of the branch and bound method.

Branching can take place in either a depth-first or breadth-first manner. The depth-first approach enumerates all nodes on a particular branch from the start to the end node. After the completion of the branch, the approach backtracks to continue with the next branch. This approach is depicted in Figure 2.13.

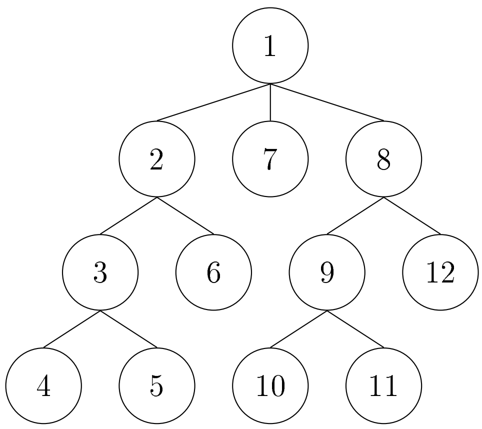


Figure 2.13: Schematic representation of the depth-first approach

The breadth-first approach examines all nodes on the same level and then selects the node with the best solution to continue branching. This approach is illustrated in Figure 2.14.

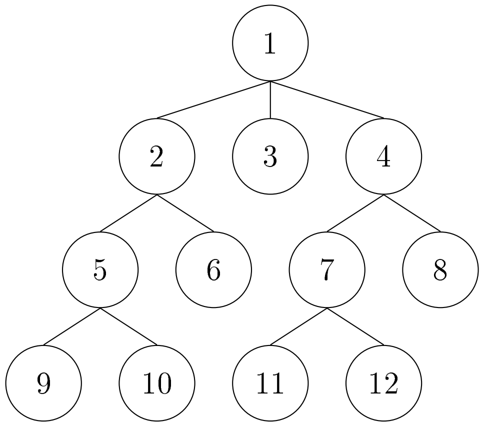


Figure 2.14: Schematic representation of the breadth-first approach

A few disadvantages have been associated with the BB technique for MINLP problems. These include the difficulty in updating NLP subproblems in the tree nodes and the size of the problems.

The concepts of the branch and bound technique, along with optimality and feasibility reduction techniques are implemented in the global optimisation solver, branch and reduce optimisation navigator (BARON). The steps involved in the BARON algorithm is shown in Figure 2.15.

Cutting Plane Method

The cutting plane (CP) method refines a feasible set iteratively, by using linear inequalities, referred to as cuts. The feasible region is thus tightened progressively by means of the introduction of valid cuts (Kelley, 1960). This method was developed to obtain integer solutions to convex MILP problems. Figure 2.16 depicts the process described above. Figure 2.16 depicts a shaded feasible region, where integer solutions are illustrated by the grid and the dot demonstrates the optimal relaxed solution. Should the optimal relaxed solution not be an integer point, a cut is introduced into a section of the

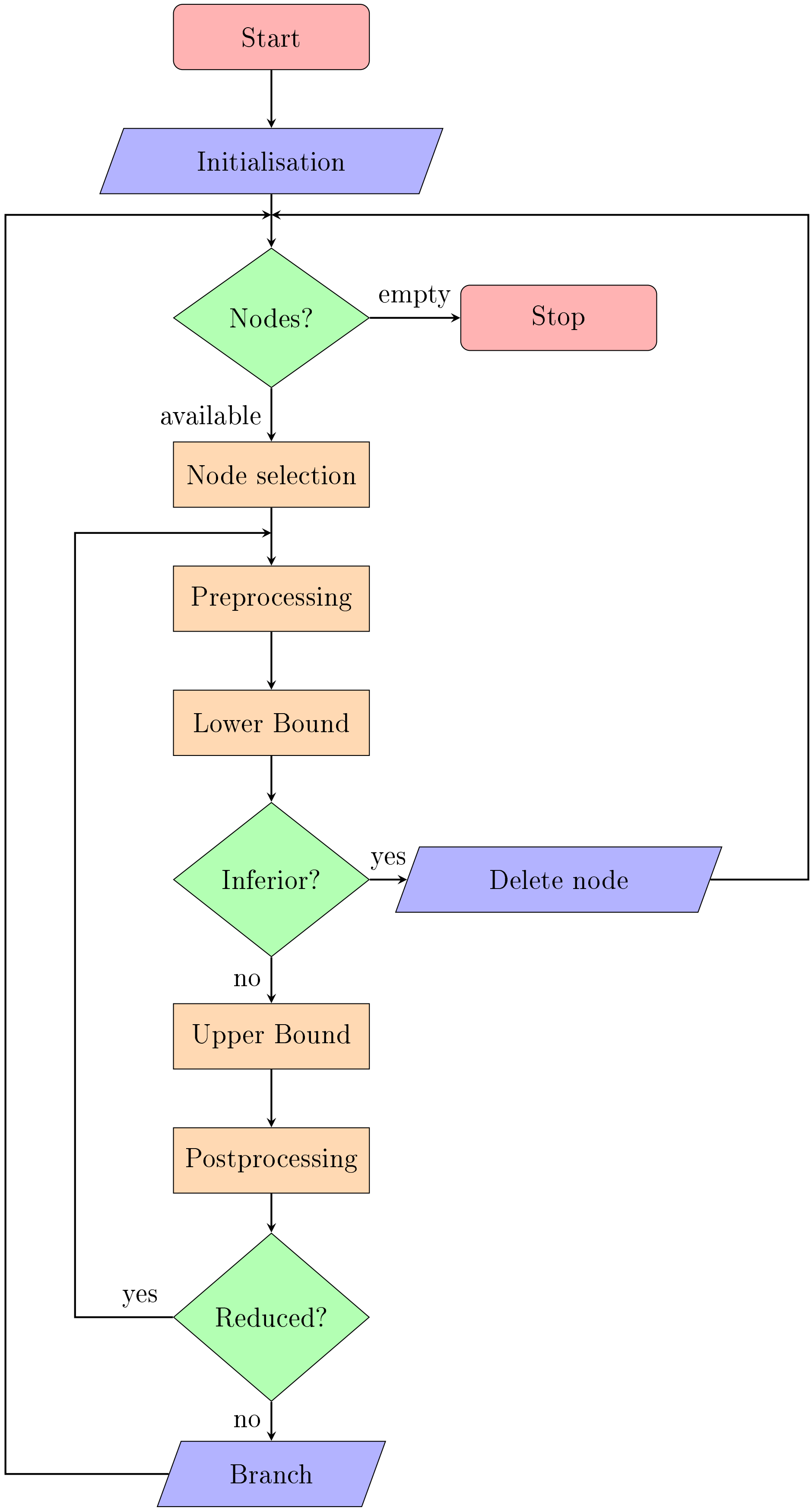


Figure 2.15: BARON algorithm flowchart

polygon near the relaxed optimal solution. This process is repeated until an optimal relaxed solution is found at an integer point, as indicated by Figure 2.16. Global optimalility can be guaranteed for problems which are linear and convex.

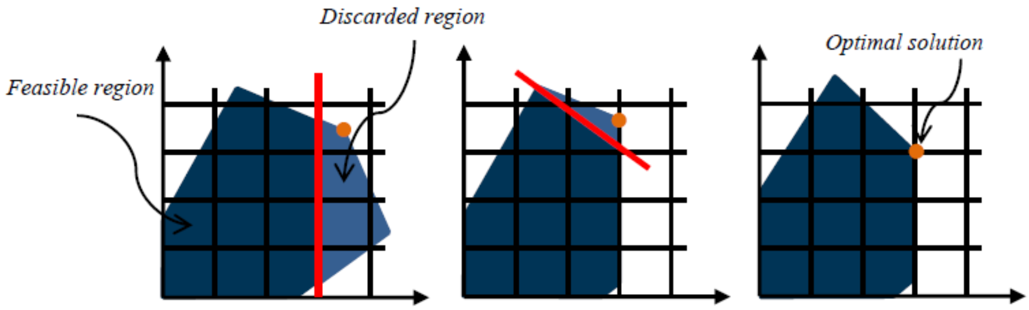


Figure 2.16: Cutting plane theory (Kelley, 1960)

The extended cutting plane (ECP) algorithm was developed to address non-convex MINLP problems using the CP technique (Westerlund and Pettersson, 1995). The relaxed problem is first solved and the optimal solution is tested for integrality. The algorithm terminates, should the optimal solution be an integer point. If not, a cut is introduced near the relaxed optimum point, separating every feasible integer solution from the current relaxed optimal fractional solution. Consequently tightening the formulation and reducing the size of the search space. This procedure is repeated until an integer optimal solution is found.

Outer approximation is an application of the cutting plane method, first introduced by Duran and Grossmann (1986). As with the cutting plane method, this method applies to convex functions and sets. A function is approximated by a polyhedral, which contains the set. This is achieved by iterative linearisations that are successively tightened. The set of linear functions or cuts form an envelope, termed a convex hull. MINLP problems are divided into an MILP master problem, where the OA linearisation approach is used to linearise the nonlinear functions, and NLP subproblems where integer variables are fixed. Solutions to the NLP subproblems provide the MILP master problem with insights to introduce further cuts, as well as an upper bound to the MINLP problem assuming a minimisation problem. The MILP master problem solutions provide the MINLP problem lower bound. A graphical representation of OA applied at four points of a convex function is available from Figure 2.17. OA might require a large number of approximations to adequately represent the feasible region. OA has been implemented in the MINLP solver known as discrete and continuous optimiser (DICOPT). Figure 2.18 provides a graphical representation of the DICOPT algorithm.

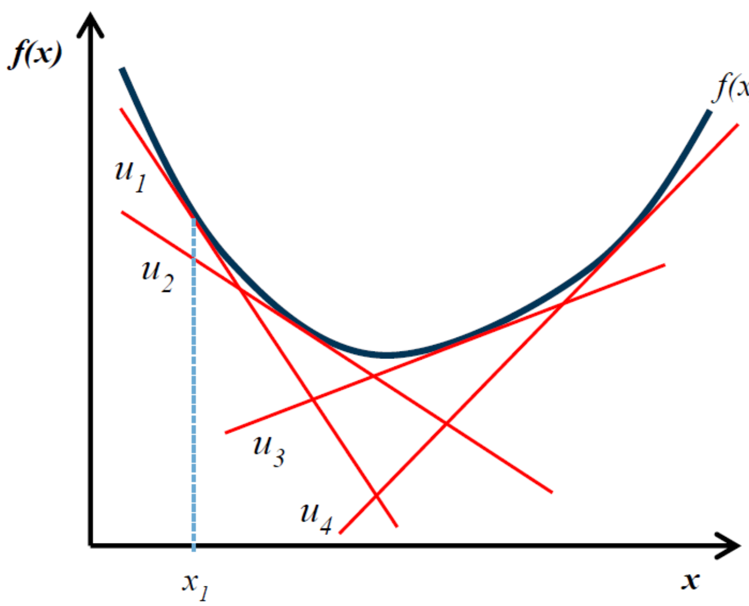


Figure 2.17: Outer approximation of a convex function (Duran and Grossmann, 1986)

Generalised benders decomposition (GBD) is similar to OA, as MINLP problems are separated into an MILP master problem and NLP subproblems. Cuts are obtained from the solution to NLP subproblems and introduced by fixing integer variables of the constraint functions. These two techniques differ in how the MILP master problem is defined. GBD implements the duality theory which requires that only the discrete variables and active inequalities be considered (Duran and Grossmann, 1986). Similar to OA, GBD requires

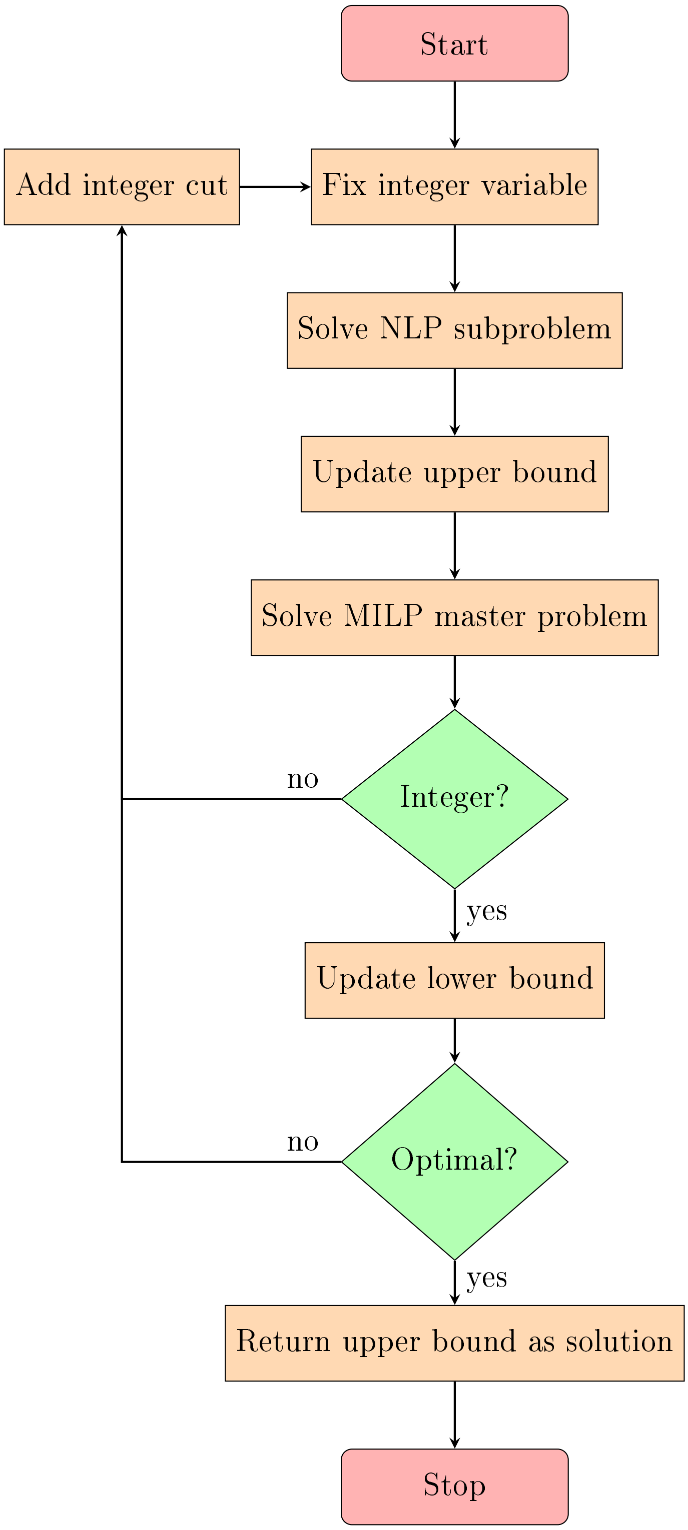


Figure 2.18: DICOPT algorithm flowchart for minimisation problems

a large number of iterations to solve the NLP subproblems and MILP master problem (Viswanathan and Grossmann, 1990). OA results in a lower bound which is tighter, or as tight as the lower bound generated through GBD. For this reason, OA tends to require fewer iterations when compared to GBD.

The techniques mentioned in this section, which are based on the CP method, rely heavily on the solutions to MILP problems, for which powerful algorithms have been developed. However, the disadvantages of these methods include the slow convergence speed, dependence on a good initial point, and global optimality cannot be guaranteed for nonconvex MINLP problems.

2.3 Cooling Towers

Cooling towers are widely used in industry for the rapid cooling of cooling water. The cooling water is typically used to exchange heat with hot processes which could not meet all of its cooling demand through process-process heat integration. A cooling tower is essentially a direct-contact heat exchanger, in which the cooling water return stream is brought into contact with unsaturated air. When the water and unsaturated air are brought into contact, some of the water will evaporate into the air as water vapour. The water vapour joining the air carries the latent heat of vaporisation with it, thereby reducing the temperature of the water. This process is driven by the concentration gradient which exists between the water and air and can be repeated until the air becomes saturated or the vapour pressure of the water and air are equal (Kern, 1950). This mechanism is known as evaporative cooling and accounts for approximately 62.5% to 80% of the total cooling tower heat transfer. The balance is made up of convective heat transfer (Khan et al., 2003; Perry and Green, 1997).

The cooling water can theoretically be cooled up to wet bulb temperature of air, although Li and Pridov (1985) suggests that the cooling water should be at least 2.8℃ above the wet bulb temperature of air. It is estimated that industry can reduce their fresh cooling water consumption by 98% by making use of cooling towers (Kern, 1950).

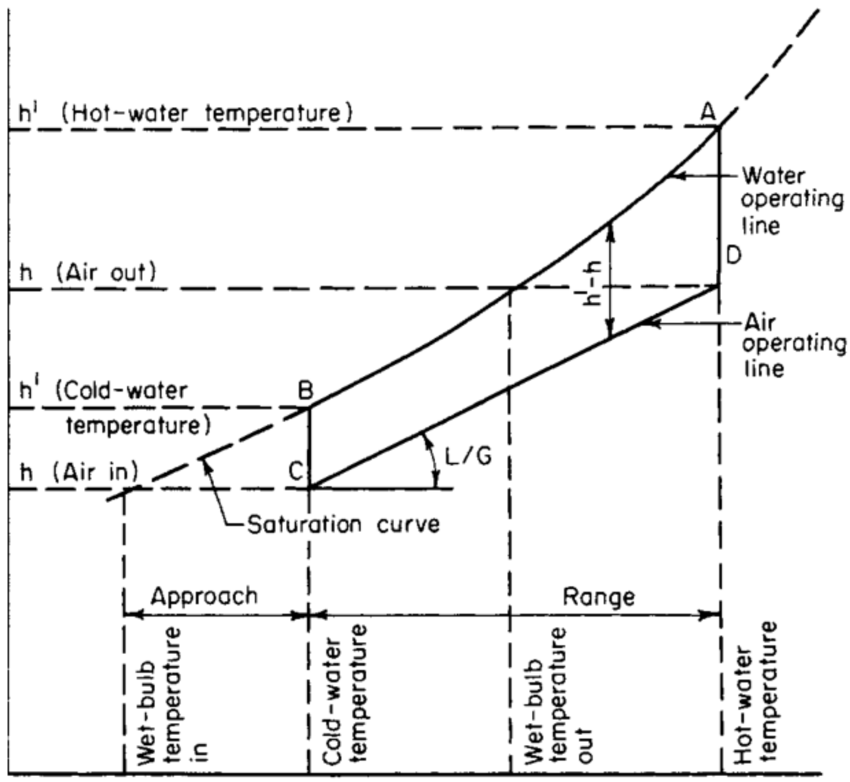


Figure 2.19: Cooling tower process heat balance (Perry and Green, 1997)

Figure 2.19 above provides a visual depiction of the fundamental interactions that occur within the cooling tower. This illustrates the relationship between air and water, as well as the driving potential for counterflow cooling towers. Commonly used terms in the domain of cooling tower optimisation can be inferred from the graphic. For example, it can be seen that the approach temperature is defined as the difference between the cold water temperature exiting the cooling tower and the air wet bulb temperature entering the cooling tower. Similarly, range is defined as the difference between the hot water temperature entering the cooling tower and the cold water temperature exiting the cooling tower.

2.3.1 Types

Cooling towers can be divided into two categories, based on the way that air enters the cooling tower. Mechanical draft cooling towers make use of a fan placed either at the top or bottom of the cooling tower, whereas natural circulation cooling towers do not make use of a mechanical fan.

Natural Circulation Cooling Towers

Natural circulation cooling towers facilitate air-water contact without making use of mechanical devices such as fans. Two types are commonly encountered in industry, these being atmospheric and natural draft cooling towers.

Atmospheric cooling towers allow wind currents to pass through one side of the column. The orientation of the cooling tower varies seasonally based on atmospheric condition and projected wind patterns. These cooling towers are typically much narrower and longer vertically than any of the other types of cooling towers discussed in this section. Due to drift loss occurring over the entire side of these towers, much greater losses are typically experienced.

Natural draft cooling towers operate in a manner analogous to furnace chimneys. Hot water provides heating of air entering the tower. This lowers the air density, leading to a density differential between the air inside and outside of the tower. This phenomenon causes air to flow in at the bottom of the tower and hot, less dense, air to flow out of the top of the tower. It is important to ensure a sufficient temperature difference between ambient air and hot water for these types of cooling towers to operate effectively. Natural draft cooling towers should be sufficiently tall for buyoancy considerations and have a large cross section due to the slow air circulation rate.

Natural circulation towers make use of natural phenomenon to drive process cooling. The benefits of these towers are that they require low operational costs due to the elimination of the cooling tower fan. However, greater requirement of upfront capital investments and reliance on environmental factors have led to these towers being phased out over time (Kern, 1950).

Mechanical Cooling Towers

Mechanical cooling towers ensure contact of hot water droplets and cold air by means of a mechanical device, typically a fan. There are two arrangements which are frequently seen in industry, namely forced draft- and induced draft mechanical cooling towers. The distinction is based on the placement of the fan. Hot water is introduced into the cooling tower from the top end as droplets. Contact is made with the cool air which flows upwards from the bottom of the cooling tower. Forced draft cooling towers place the fan at the bottom entrance of the cooling tower, thereby forcing air through the tower. Fans are placed at the top of the tower inducing air flow in the induced draft arrangement. The induced draft cooling tower has grown increasingly popular in industry due to the reduced air inlet spacial requirements, improvement in air circulation inside the tower, and reduction of air recirculation. Induced draft cooling towers do however experience a greater amount of drift loss and increased pressure drop compared to forced draft cooling towers and subsequently require greater fan power (Kern, 1950).

2.3.2 Fills

Fills are used within the cooling tower to increase the residence time of water within the cooling tower and increase the contact area between air and water. Kloppers and Kröger (2003) evaluated the loss coefficient of splash, trickle, and film type fills. It is recommended that the cooling water return temperature does not exceed 50℃ to ensure the fills are not damaged (Douglas, 1988).

Splash Fills

Splash fills are made up of layers of horizontal bars. Warm recirculating cooling water breaks to form small droplets upon contact with the bars. This provides an increased air-water contact surface. Splash fills are typically made of PVC or wood. Splash fills are inexpensive and can be used in applications with very dirty water.

Trickle Fills

Trickle fills are an adaptation of the splash fills formed into a block shape.

Film Fills

Film fills are made up of textured thin sheets separated by small spacing knobs. This provides a large surface area for the warm water to spread out into a thin film for air-water contact. Film fills are typically made with PVC material. It is important to ensure the recirculating cooling water is clean, as debris can build up on the fill and reduce overall cooling tower performance.

2.3.3 Models

Merkel Method

The prediction of cooling tower performance dates back to Merkel (1926). The author derived the model by considering the mass and energy effects of a falling water droplet surrounded by a thin film of air. The enthalpy difference between the film and ambient air provides the driving force for the heat exchange process. The key assumptions made in the development of the model are as follows:

- The Lewis factor, which relates heat and mass transfer effects, is unity.
- The air exiting the cooling tower is saturated.
- The effect of evaporation on the water flow rate is neglected in the energy balance.

Kloppers and Kröger (2005a) investigated the effect of the Lewis factor on the prediction of cooling tower performance. It was found that an increase in the Lewis factor is associated with increased heat rejection, decreased water outlet temperature and decreased rate of evaporation. Furthermore, the influence of the Lewis factor diminishes when ambient air is hot and humid.

The Merkel method provides a simple method that can accurately predict cooling water outlet temperature, although it does not represent the heat and mass transfer effects within the cooling tower fill accurately (Kloppers and Kröger, 2005a). Mills (1999) found the Merkel method to generally be accurate within 3–5% of exact results up to temperatures of about 60℃. The method makes use of the Merkel number to predict cooling tower performance, as demonstrated by Equation 2.16.

$$Me = \frac{h_d a f L}{G} = \int_{T_{W_{in}}}^{T_{W_{out}}} \frac{c_{pw}}{i_{masw} - i_{ma}} dT_W \tag{2.16}$$

The integral expression represents the area between the saturated air curve and the operating line on a temperature vs enthalpy plot. The Cooling Tower Institute (1997) and British Standard 4485 (1988) recommend that the four-point Chebyshev integration technique be used in determining the Merkel number. Chebyshev integration is very accurate compared to the 100 interval Simpson's rule (Mathews, 1992). It should be noted that Chebyshev integration lacks accuracy when the approach temperature is smaller than approximately 0.56℃ (Kelly, 1976). Oosthuizen (1995) and Mohiuddin and Kant (1996) provide further discussions on Chebyshev integration.

e-NTU Method

Jaber and Webb (1989) later developed the e-NTU (effectiveness-NTU) method, based on the same assumptions as the Merkel method. The authors used the insight that, in essence, a cooling tower operates as a large heat exchanger between the cold air and hot water. This allowed the authors to base their method on the NTU (Number of Transfer Units) technique used in the design of heat exchangers. Due to the commonality of underlying assumptions, the e-NTU method is in the same accuracy range as the Merkel method. The main benefit of the e-NTU method is that the computational difficulty is reduced, as there are no integral functions to evaluate.

Poppe Method

Poppe and Rögener (1991) developed a rigorous method that does not rely on the simplifying assumptions of the Merkel method. The authors utilised a control volume made up of a differential height and cross sectional slice through the cooling tower. The control volume was used as a basis for a mass and enthalpy balance. To more accurately determine the Lewis factor, the authors made use of a correlation by Bosnjaković (1965). The Poppe method does however result in differential equations which are more difficult to solve.

Kloppers and Kröger (2005b) compared the above mentioned models. They concluded that if only water outlet temperature is of importance, the Merkel and e-NTU methods could be used, as all three the methods predict similar water outlet temperature values.

2.3.4 Cooling Water Systems

Recently, cooling towers have been considered as part of the cooling water system. As demonstrated in Figure 2.20, the cooling water system comprises of the cooling tower, cooling water network, and pumping network. Water losses such as evaporation, drift, and blowdown occur during the operation of the cooling water system. Kemmer (1978) determined that drift losses should not exceed 0.2% of the total circulating water flowrate in well-designed cooling towers. An empirical expression to determine the evaporation loss in cooling towers was developed by Perry and Green (1997). Blowdown is used to ensure there is no build-up of contaminants in the cooling water system. Blowdown is expressed in terms of cycles of concentration (CC). This is the ratio of the concentration of soluble component in the blowdown stream to the makeup stream. Li and Pridov (1985) suggests a cycles of concentration value of 2–4 be used. Makeup water is introduced to the system to account for the aforementioned water losses in the cooling water system.

Experimental studies conducted by Bernier (1994) demonstrated that altering the dry bulb temperature while controlling the wet bulb temperature lead to no significant effect on water outlet temperature. Conversely, increasing wet bulb temperature reduced the approach temperature, resulting in increased overall cooling. Arguably, the most significant finding was that a reduction in the recirculating cooling water flowrate, which is concomitant with increased cooling water return temperature, improved the thermal performance of a cooling tower. This finding, demonstrated by Figure 1.4, formed the basis for the optimisation of cooling water systems.

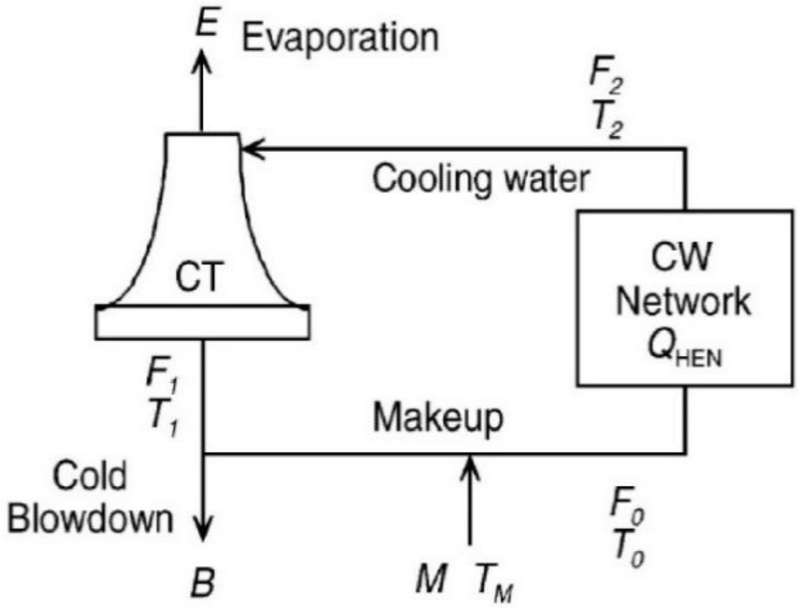


Figure 2.20: Cooling water system (Kim and Smith, 2001)

Castro et al. (2000) developed a nonlinear mathematical model incorporating pressure drop considerations into the cooling water system model. The authors made use of a parallel cooling water network and used correlation functions to predict cooling tower outlet conditions.

Kim and Smith (2001) improved on the work of Castro et al. (2000) by implementing a combined water and energy analysis. The authors developed a graphical method based on pinch analysis to minimise recirculating cooling water flowrate, by maximising cooling water reuse. Their work followed from the aforementioned insight by Bernier (1994), who determined that cooling tower performance improves when operating at reduced cooling water recirculating flowrates. Prior to their work, the cooling water network was traditionally arranged in parallel and supplied by cooling water directly from the cooling tower, as depicted by Figure 1.3.

The authors further mathematically defined cooling tower effectiveness as the ratio of actual heat removed to the theoretical maximum amount of removable heat. In simplifying the model, losses in the form of evaporation and drift were assumed to be negligible. Their procedure was later automated by a mathematical model (Kim and Smith, 2003).

Kim and Smith (2004) later developed a method to reduce the overall makeup water requirement. In doing so, water recovery streams were introduced between wastewater generating processes and the cooling water system. The temperature of the wastewater exiting each unit determined where it was fed into the CWS. Wastewater was added after the cooling tower if the wastewater recycle stream temperature was below the temperature of the cooling water entering the cooling water network. Conversely, if the wastewater recycle stream temperature was above the temperature of the cooling water entering the cooling water network, the water recovery stream was introduced before the cooling tower. This arrangement ensures that additional heat is not added to the cooling water entering the cooling water network.

Khan and Zubair (2001) incorporated drift and evaporative losses in their cooling tower model. The model was complex and included the calculation of the Lewis number. It was concluded that an increase in the Lewis number corresponded with increased cooling tower effectiveness. Khan et al. (2004) extended their work by investigating the effects of fouling on thermal performance. Fouling refers to the deposition of material, such as bio-growth, on the film area.

Ponce-Ortega et al. (2007) developed a mathematical model to synthesise cooling water networks at minimum cost. Their work made use of the stage-wise superstructure approach, introduced by Yee and Grossmann (1990). As depicted by Figure 2.21, the set of hot processes is divided into a series of stages. Each stage behaved similar to a water main, as introduced by Kuo

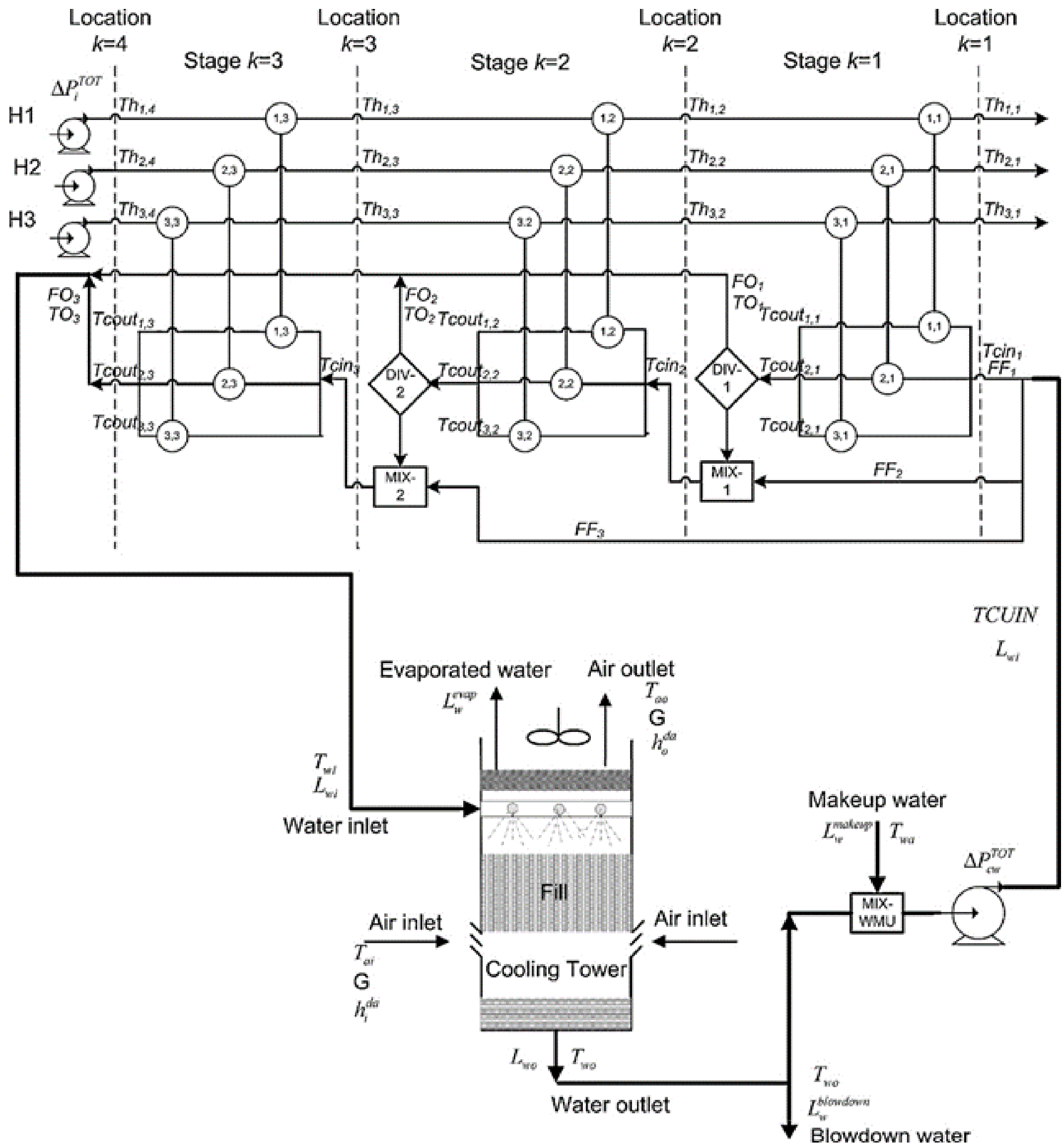


Figure 2.21: Stagewise cooling water system superstructure (Ponce-Ortega et al., 2007)

and Smith (1998b). Thus, a stage could be supplied by cooling water from a stage preceding it or directly from a cooling tower. Ponce-Ortega et al. (2010) later improved on this work, by including a cooling tower model based on the Merkel method. This work was further extended by including various types of pumps and cooling tower fills in the model (Serna-González et al., 2010).

Qureshi and Zubair (2006) developed a cooling water system model which predicts the fouling of packing and investigated the effects on thermal performance. Panjeshahi et al. (2009) expanded on the work of Kim and Smith (2001). They investigated the effects of ozone treatment of the recirculating cooling water. By inhibiting biological growth, the cycles of concentration was increased, resulting in decreased blowdown and makeup. The authors also minimised the total annualised cost of a cooling water system. They developed the advanced pinch design technique by combining pinch integration with mathematical programming. The cooling tower model described by Kröger (2004) was used to predict cooling tower outlet conditions.

The aforementioned cooling water system models all consider a single cooling tower, however in industry it is common practise to make use of multiple cooling towers to meet the cooling demand. Majozi and Nyathi (2007) developed a methodology for the synthesis of cooling water systems made up of multiple cooling towers. They implemented a graphical technique for targeting, and a mathematical programming model for synthesis of the cooling water network. The authors implemented the conditions of optimality introduced by Savelski and Bagajewicz (2000) to linearise the model. Majozi and Moodley (2008) expanded on their work by developing a rigorous mathematical optimisation model.

Gololo and Majozi (2011) extended this model by including a cooling tower model, based on the work of Kröger (2004). This allowed for cooling tower performance to be evaluated. The authors developed a solution algorithm that includes the fourth order Runge-Kutta method for approximating the solution to the differential equations throughout the model. For this reason the authors modelled the cooling towers and cooling tower network using two different platforms, requiring an iterative decomposition procedure to obtain the final solution. Rubio-Castro et al. (2013) extended the work by Ponce-Ortega et al. (2010) to include multiple cooling towers.

More recently, Zhu et al. (2017) optimised the operating parameters of the CWS while maintaining the equipment configuration and network topology. Liu et al. (2018) designed the HEN and CWS simultaneously. The cooling water was treated as a special cold stream. Zheng et al. (2018) developed an MINLP for the simultaneous optimisation of the pumping network and a parallel CWN considering multiple cooling towers. The centrifugal pumps were divided into a main and auxiliary subnetwork. The authors made use of the optimality criterion described by Savelski and Bagajewicz (2000) to linearise their model. It was found that the pumping costs made up more than 60% of the total cost. Ma et al. (2018) later extended the pumping network into the so called multi-loops pump network in order to better satisfy various cooling demands. The authors made use of a stagewise CWN, thereby reducing the number of auxiliary pumps required. Pontes et al. (2019) investigated the effects of ambient temperature on the required cooling tower capacity, paying specific attention to the volume of cooling tower fill and fan power. Li and Li (2020) presented a model to optimise switch-over temperature to facilitate free cooling and thereby enable energy savings through control of cooling tower fans. The model was applied to a data centre case study, resulting in 10% energy savings.

Wu et al. (2021) developed a stochastic NLP model, based on the e-NTU method for cooling tower design, which takes into account the variability of weather by allowing for a design margin in the proposed heat exchanger area.

2.3.5 Pressure Drop Considerations

The use of series connections in cooling water network models, through the implementation of reuse and recycle opportunities, inherently leads to an increase in the network pressure drop. In seeking to minimise recirculating cooling water flowrate, at minimum pressure drop, Kim and Smith (2003) extended their previous work by developing a mathematical model to minimise pressure drop of the cooling water network. The mathematical model was based on the critical path algorithm (CPA) and formulated as an MINLP. The pressure drop was calculated based on the cooling water flowrate for both piping and equipment. Insights from this technique were later used to minimise pressure drop in steam systems (Price and Majozi, 2010) and cooling water systems with multiple cooling towers (Gololo and Majozi, 2012).

2.4 Batch Processes

The aforementioned research on cooling water systems have been applied to continuous background processes. This disregards cooling towers which are coupled with batch background processes. Batch processes can be distinguished from continuous processes by noting that continuous processes operate throughout the time horizon of interest, with a continuous inlet and outlet stream and no set starting or finishing time. Figure 1.2 visually depicts the difference between batch and continuous processes. This mode of operation is suited to products that are produced in bulk, such as petrochemicals.

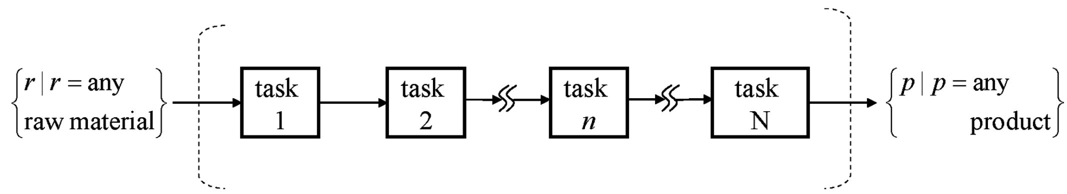
Batch processes operate intermittently throughout the time horizon of interest, without continuous inlet and outlet streams and set starting and finishing times. These processes are ideally suited to produce low-volume, high-value products. Batch processes allow raw materials, intermediates, utilities, and equipment to be shared during production, thereby offering a greater degree of flexibility compared to continuous processes. This flexibility brings about a large degree of complexity to the synthesis and design of batch processes. The large amount of discrete decisions and size of the solution space leads to a combinatorial problem which is computationally demanding. The job shop scheduling problem is analogous to the travelling salesman problem, where cities are equivalent to units and the salesman is equivalent to the jobs. The scheduling problem is regarded as NP-complete, meaning that these problems are not bound by polynomial time. Hence, problem complexity and solution time is exponentially proportional to problem size.

Batch scheduling aims to determine the sequence of tasks in the available processing units and the times at which they occur, which leads to the optimal value of a given objective function. The objective function typically takes the form of either throughput maximisation or makespan minimisation. The majority of techniques developed for batch scheduling are based on mathematical programming. This is due to the relative ease at which the time dimension and various constraints of practical relevance can be incorporated into a mathematical model, as opposed to a graphical technique. There have, however, been research efforts aimed towards developing a graphical technique for batch scheduling, an example of which is the schedule-graph developed by Sanmartí et al. (1998) and later extended by Majozi and Friedler (2006).

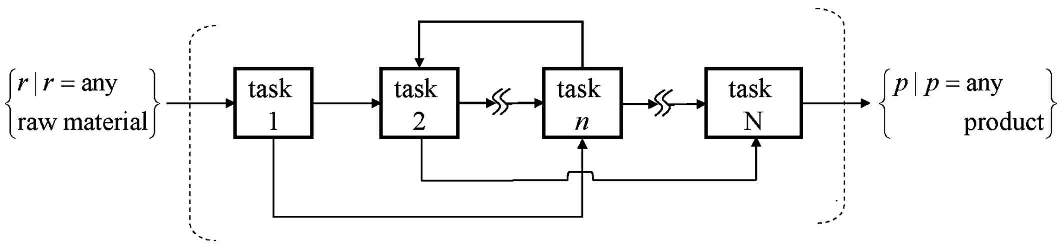
The major contributions to the batch scheduling domain have involved formulations which improved the solution objective functions or decreased solution times. Important considerations to batch scheduling formulations as well as recent developments are discussed in the sections which follow.

2.4.1 Types of Batch Plants

Batch processes can be classified into two distinct categories, based on the manner in which material flows through the processing equipment. In multiproduct batch plants, also referred to as flowshop plants, products are manufactured using the same equipment following the same recipe in successive production campaigns. Thus each unit is dedicated to a specific task. This differs from multipurpose batch plants, also referred to as jobshop plants, where products can be manufactured using different equipment following a different recipe in successive production campaigns. Therefore, equipment units can be shared by various tasks. Multiproduct batch plants can be seen as a subset of multipurpose batch plants. Multipurpose batch plants offer a greater degree of flexibility, however this comes with greater complexity in scheduling. Figure 2.22 depicts the flowsheets of a typical multiproduct and multipurpose batch facility.



(a) Multiproduct Batch Plant



(b) Multipurpose Batch Plant

Figure 2.22: Flowsheet of typical multiproduct and multipurpose batch plants (Majozi, 2010)

2.4.2 Storage and Operational Philosophies

Storage plays an important role in batch processes, as it provides an avenue to overcome the time constraints. This introduce additional flexibility to the given batch plant. The use of storage can further debottleneck the batch schedule and increase equipment utilisation, thereby increasing process productivity. Storage policies for intermediate products are determined by the operational and transfer philosophy which the particular batch plant has adopted. The most important operational and transfer philosophies are briefly discussed in the points that follow.

- **No Intermediate Storage (NIS):** No intermediate storage vessels are available. If intermediates are not used immediately in a successive task, they are temporarily stored in the unit in which they were produced. This reduces the plant area and capital requirements for the batch plant.

- **Common Intermediate Storage (CIS):** A single storage vessel exists to be shared by all intermediates. This requires a smaller amount of area, compared to the situation where a storage vessel is dedicated for each intermediate, however care should be taken to ensure product integrity (Jung et al., 1996).

- **Finite Intermediate Storage (FIS):** Intermediate storage vessels exist between tasks for each intermediate. The capacity of the storage vessels are finite, thus storage cannot be guaranteed.

- **Unlimited Intermediate Storage (UIS):** Intermediate storage vessels exist between tasks for each intermediate. The capacity of the storage vessels are unlimited, thus storage can be guaranteed.

- **Process Intermediate Storage (PIS):** Process units which are not active may be used as intermediate storage vessels, thereby making use of the latent storage potential of the available processing units. This leads to increased capital utilisation of the process units (Pattinson and Majozi, 2010).

- **Mixed Intermediate Storage (MIS):** More than one type of the aforementioned operational philosophies coexist within a particular batch plant. An example of this would be a batch plant which makes use of FIS between certain tasks, while making use of CIS for other tasks (Wiede and Reklaitis, 1987).

- **Zero Wait (ZW):** Intermediates are not allowed to be stored after they have been produced. Intermediates are transferred to the following task immediately after being produced. This transfer philosophy is commonly used for unstable intermediates, where a delay may have a detrimental effect on the quality of the product.

- **Finite Wait (FW):** Intermediates are allowed to be stored for a finite amount of time after they have been produced. This transfer philosophy is commonly used for intermediates which are partially stable and starts decomposition after a fixed amount of time.

- **Unlimited Wait (UW):** Intermediates are allowed to be stored for an unlimited amount of time after they have been produced. This transfer philosophy is commonly used for intermediates which are stable.

2.4.3 Time Representation

Due to the discrete nature of batch processes, accurately capturing the essence of time is an important factor in developing a mathematical model for the scheduling of batch operations. The time horizon can be discretised uniformly or unevenly.

The uniform discretisation of time, also referred to as the discrete time representation, divides the time horizon into a finite number of intervals of uniform duration with a binary variable associated with each interval. The starting and finishing time of each task must coincide with a time interval boundary. To ensure global optimality the time intervals must be sufficiently small. It is suggested that the greatest common factor of processing times be used as the time interval duration. The benefit of the discrete time representation is that it provides all tasks with a common time grid, which simplifies the material balances involving shared resources. Difficulty arises in the industrial implementation of the discrete time representation, as the large amount of unnecessary time intervals leads to an explosive binary dimension which is computationally difficult to solve. The discrete time representation does also not accommodate variable processing times dependent on batch size, as is commonly encountered in industry. The discrete time representation is illustrated in Figure 2.23 and Figure 2.24 provides a depiction of the discrete time grid. Examples of batch scheduling models which make use of the discrete time approach is Kondili et al. (1993), Shah et al. (1993) and Pantelides (1994).

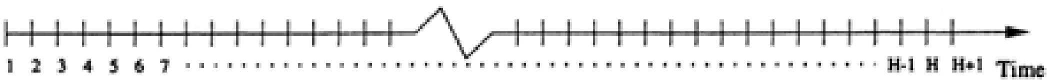


Figure 2.23: Discrete time representation (Floudas and Lin, 2004)

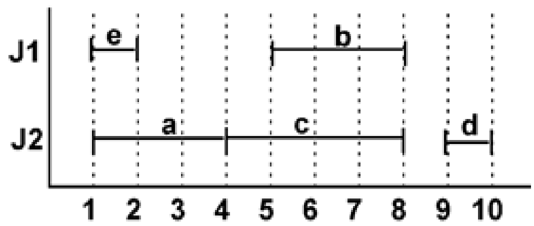


Figure 2.24: Discrete time grid (Méndez et al., 2006)

The uneven discretisation of time, also referred to as the continuous time representation, divides the time horizon into intervals of unknown durations with a specified number of time points. The starting and finishing time of each task must coincide with a time interval boundary. This approach generally leads to smaller problem sizes which require less computational effort, however there is an increased integrality gap associated with the continuous time representation (Floudas and Lin, 2004). The major drawback of this method is that the optimal number of time points have to be determined iteratively.

Seid and Majozi (2012) recently developed a mathematical technique to determine the optimal number of time points. Figure 2.25 demonstrates the continuous time representation.

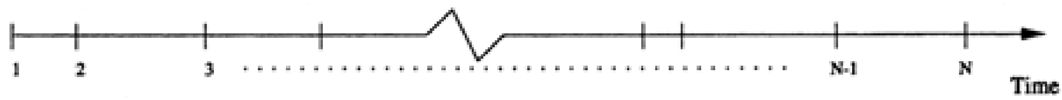


Figure 2.25: Continuous time representation (Floudas and Lin, 2004)

The continuous time formulation can be categorised into slot based and event based models. Slot based models divide the time horizon into a consecutive set of ordered slots of unknown duration. One of the earliest uses of a slot based model is that of Pinto and Grossmann (1995). Event based models represent the time horizon by a series of time instances along the time horizon which signify when a task becomes active. An early example of an event based model is by Zhang and Sargent (1996). By definition n amount of slots will correspond to $n + 1$ events.

A further distinction of scheduling models can be made based on the amount of time grids utilised. A scheduling model can utilise a single, multiple or no time grid. A single time grid entails that there is one common time grid for all the units. Thus if an event point is defined at the start of a particular task taking place in a particular unit, the event point will be common to all operations. This simplifies problems involving shared resources, however the amount of time points required tends to be large for complex problems or problems involving long time horizons. When multiple time grids are used, a separate time grid is defined for each unit. This allows tasks corresponding to the same event point taking place in different units to occur at different absolute times on the time horizon. Scheduling models that do not make use of a time grid are referred to as precedence based models. Precedence based models differ from all the aforementioned scheduling models in that they exploit the sequential processes which some batch plants follow. These models are batch oriented, meaning that states and tasks do not have to be defined explicitly. For this reason, precedence based models are classified as sequentially based models, whereas all previously mentioned models are network based models.

The points which follow state the continuous time formulation models which have been developed. A brief description, depiction of the associated time grid, and an example from literature where this type of time representation has been used are provided.

- **Synchronous Time Grids:** This representation involves a slot based model utilising a single time grid (Sundaramoorthy and Karimi, 2005). Global time points are used to depict the start and end of each task. Figure 2.26 illustrates the use of the synchronous slot based time grid consisting of six time slots.

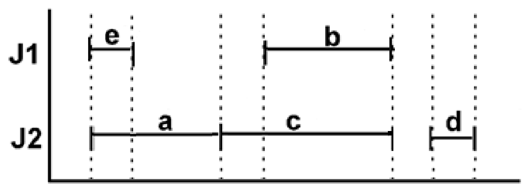


Figure 2.26: Synchronous time grid (Méndez et al., 2006)

- **Asynchronous Time Grids:** This representation involves a slot based model utilising multiple time grids (Karimi and McDonald, 1997). Unit specific time points are used to depict the start and end of each task. Figure 2.27 presents the use of the asynchronous slot based time grid consisting of three time slots.

- **Global Time Points:** This representation involves an event based model utilising a single time grid (Castro et al., 2001; Maravelias and Grossmann, 2003). Global time points are used to depict the start of each task. Figure 2.28 depicts the use of the global time points grid consisting of three five time points.

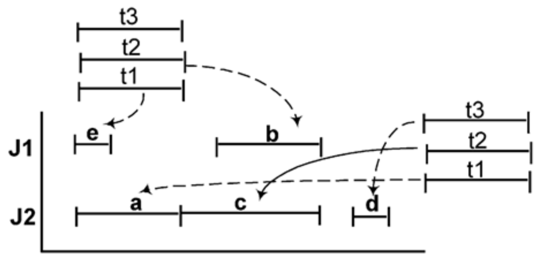


Figure 2.27: Asynchronous time grid (Méndez et al., 2006)

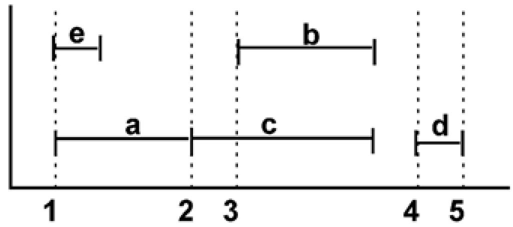


Figure 2.28: Global time points grid (Méndez et al., 2006)

- **Unit Specific Event Based:** This representation involves an event based model utilising multiple time grids (Ierapetritou and Floudas, 1998; Janak et al., 2004). Unit specific time points are used to depict the start of each task. Figure 2.29 presents the use of the unit specific event based time grid consisting of three time points.

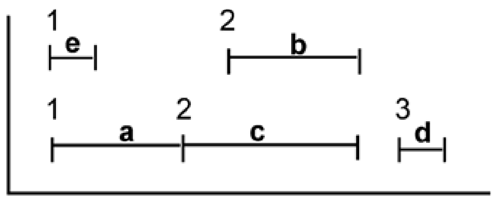


Figure 2.29: Unit specific event based grid (Méndez et al., 2006)

- **Immediate Precedence:** This representation utilises two sets of binary variables to define the allocation and sequencing constraints. The sequencing binary variables becomes active if batch i' immediately follows batch i (Gupta and Karimi, 2003). Figure 2.30 is used to demonstrate the precedence time grid for all of the following precedence based formulations.

- **Global Precedence:** This representation utilises a single sequencing binary variable for each pair of batch tasks that can be allocated to the same unit is required. The sequencing binary variable becomes active if batch i' follows batch i. Subsequent tasks are cumbersome to define and lead to problems with sequence dependent changeover (Méndez et al., 2001).

- **Unit Specific Immediate Precedence:** This representation utilises a single binary variable defines both the allocation and sequencing of batch pairs (Cerdá et al., 1997).

- **Unit Specific Global Precedence:** This representation addresses the issue of sequence dependent changeover by combining the concepts of global precedence with unit specific immediate precedence (Kopanos et al., 2009).

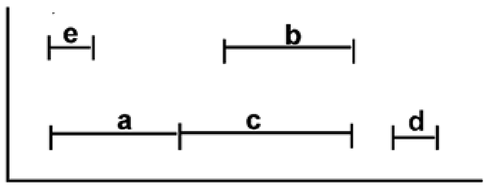


Figure 2.30: Precedence time grid (Méndez et al., 2006)

2.4.4 Flowsheet Representation

The mathematical model for batch scheduling is also influenced by the manner in which the flowsheet is represented. The flowsheet provides the depiction of the production recipe involved in the particular batch process. The four most frequently used flowsheet representations are discussed in the points that follow.

- **State Task Network (STN):** This representation was developed by Kondili et al. (1993). The flowsheet consists of state nodes, task nodes, and arcs. State nodes represent any material with distinct properties and attributes, such as raw material, intermediates, products, etc.

State nodes are depicted by circles. Task nodes represent all unit operations which are responsible for transforming an input state into an output state. Task nodes are depicted by rectangles. The nodes are linked by arcs, which indicate task precedence. Figure 2.31 depicts a typical STN flowsheet.

Figure 2.31: STN flowsheet representation

- **Maximal State Task Network (m-STN):** This representation was generalised by Barbosa-Póvoa and Macchietto (1994) by combining the plant structure characteristics with the process recipe into a single framework. This ensures the unambiguous depiction of the topology and connectivity of the plant, since each equipment unit, as well as resources associated with it, is handled independently. Another feature of the m-STN is that the resource utilisation is explicitly taken account of. The definition and depiction of tasks and states are the same as for the original STN representation.

- **Resource Task Network (RTN):** Pantelides (1994) extended the STN representation to the RTN representation. The flowsheet consists of resource nodes, task nodes, and arcs. Resource nodes represent states, energy, storage, transportation, and manpower. Resource nodes are depicted by circles. Task nodes represent all unit operations which are responsible for transforming a set of input resources to a set of output resources. This definition is similar to that of the STN, but now includes operations such as storage, cleaning and transportation. The nodes are linked by arcs, which indicate task precedence. When considering batch plants involving identical equipment, the RTN is preferred over the STN. A simple example of a RTN flowsheet is presented in Figure 2.32.

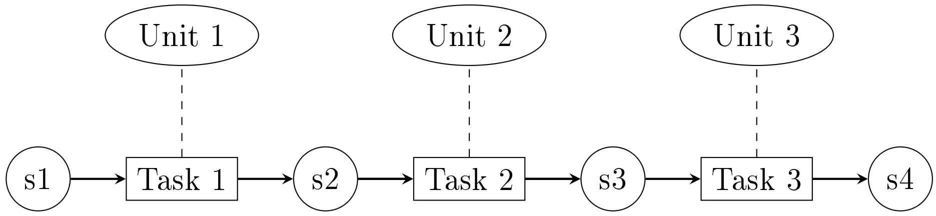


Figure 2.32: RTN flowsheet representation

- **State Sequence Network (SSN):** This representation was introduced by Majozi and Zhu (2001). The flowsheet consists of state nodes and arcs. State nodes are defined as in the STN. Tasks are implied by the arcs, linking an input state to an output state. If multiple states enter a unit, one of the states is declared as the effective state. The implication of the SSN is that binary variables are not assigned to tasks. This results in a mathematical model requiring less binary variables, provided there is not a one-to-one correspondence between states and tasks. A typical representation of the SSN is shown in Figure 2.33.

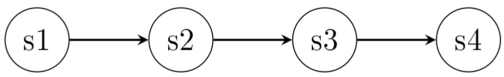


Figure 2.33: SSN flowsheet representation

2.4.5 Batch Scheduling

As mentioned previously, Kondili et al. (1993) developed the STN flowsheet representation. Their MILP formulation was based on the discrete time representation. Tasks were explicitly assigned to units through a dedicated three index binary variable. These two factors resulted in an explosive binary dimension for complex problems.

Schilling and Pantelides (1996) made use of the RTN flowsheet representation in the development of their continuous time formulation. Time points and slots were used to represent the beginning and end of a task. The model took the form of an MINLP, however Glover transforms were used to linearise the model to an equivalent MILP. The MILP resulted in a large integrality gap.

The authors implemented a novel branch and bound algorithm in an attempt to reduce the integrality gap. The technique branched on both discrete and continuous variables.

Ierapetritou and Floudas (1998) attempted to decrease the binary dimension in their batch scheduling model. In doing so the authors decoupled tasks and units. This was achieved by using a separate binary variable for tasks events and unit availability. Thus the total amount of binary variables required for a given problem would be $(i + j) \cdot p$. The number of binary variables could be further reduced in problems which exhibited a one-to-one relationship between tasks and units. Their model was based on the unit specific event based time representation and STN flowsheet representation. The overall model was in the form of an MILP. The FIS storage philosophy was not accommodated by the model.

The SSN flowsheet representation was introduced by Majozi and Zhu (2001). This allowed for a single binary variable to represent the consumption of a state at a particular time point, thereby eliminating the need for task and unit binary variables. Tasks could be inferred where a change in state occurred. In cases where multiple states are consumed in a common unit at the same time, an effective state is defined. This reduced the binary dimension to $s^* \cdot p$ for any given problem. The formulation was based on the continuous time representation and took the form of an MILP. The model was not able to incorporate the FIS storage philosophy.

Maravelias and Grossmann (2003) made use of the global time point time representation and STN in developing their batch scheduling model. The advantage of their model was the relative ease of accommodating shared resources. Sequence dependent changeover was incorporated for different tasks in a common unit. The authors allowed tasks to span over multiple time points to account for the utilisation of various resources and storage policies. Three binary variables were used throughout the model. The large binary dimension has caused this model to struggle with long time horizons.

The work of Ierapetritou and Floudas (1998) was extended by Janak et al. (2004) by considering batch mixing and splitting, sequence dependent changeover, shared storage policies and resource constraints for limited utility availability. The authors made use of an enhanced unit specific event point time representation allowing tasks to span over multiple time points. Storage was provided through dedicated storage vessels as well as idle units performing storage tasks.

Wu and Ierapetritou (2004) developed a cyclical scheduling technique applicable to long term scheduling problems, based on the continuous time representation and STN. Their work can be seen as an adaption of the short term scheduling technique proposed by Ierapetritou and Floudas (1998). The authors paid careful attention to the startup and shut down periods involved in the schedule. The concept driving the idea of cyclical scheduling involves a subschedule, with a time horizon much shorter than the overall time horizon, which can be repeated multiple times. This approach leads to reduced accuracy, compared to the direct scheduling approach, however problem complexity is greatly decreased.

Sundaramoorthy and Karimi (2005) demonstrated that decoupling of task and unit events do not lead to a reduction in the binary dimension compared to the traditional three index STN binary variable. The authors developed an MILP model without big-M constraints, which have been associated with increased computational difficulty. The formulation was based on the synchronous time grid representation, where tasks were allowed to span over multiple time points. The use of a single time grid facilitated shared resources into the scheduling model, however increased the required number of time slots to obtain the optimal solution. The authors also included novel slot based constraints such as time balances. Finally, tasks were allowed to finish before the end of a slot, thereby allowing a processed, non-zero wait, state to be stored in the unit which it was processed in until the end of the current slot. This afforded the model a greater degree of flexibility.

Shaik and Floudas (2009) highlighted the importance of tasks spanning over multiple time points. The authors illustrated that models which do not allow task spanning may yield suboptimal results. Their MILP formulation was based on the unit specific event based time representation. The drawback of this model was that both the number of event points and the amount of points over which any given task was allowed to span had to be determined iteratively.

Susarla et al. (2010) adapted the work of Sundaramoorthy and Karimi (2005) by implementing an asynchronous time grid representation and allowing tasks to span over multiple time points. The model decreased the number of time points required to find the optimal solution. The authors also permitted non-simultaneous material transfer into and out of units. This implies that for tasks which require multiple states, a state can be sent to the unit prior to the task commencing. The state would be stored in the processing unit until all other required states have been sent to the unit for the task to commence. This improved operational flexibility and led to better schedules.

Seid and Majozi (2012) addressed the shortcomings of the model by Majozi and Zhu (2001). The authors utilised the SSN flowsheet representation and the unit specific event based time representation to incorporate constraints which accommodate the FIS storage policy. The formulation also included conditional sequencing. This involves the alignment of a producing and consuming task of an intermediate state only in the event that the consuming task uses the material from the producing task in particular. It was found that the model required fewer time points compared to other contemporary formulations.

Vooradi and Shaik (2012) decreased the model complexity of Shaik and Floudas (2009).The authors reduced the binary dimension of their model by introducing the concept of active tasks. The binary variable responsible for active tasks monitor the time points over which a task is active. The authors further argued that Seid and Majozi (2012) implemented only partial conditional sequencing, since in cases where the conditions are met for the alignment of production and consumption tasks, all of the consumption tasks are aligned with the production task. This would occur even in circumstances where only one of the consumption tasks make use of the produced state. The authors implemented rigorous conditional sequencing, which further reduced the number of time points required for a given scheduling problem.

This highlights the fundamental aspects in the scheduling of batch operations. For a more detailed description of batch scheduling, the reader is referred to the extensive reviews written by Floudas and Lin (2004), Méndez et al. (2006), Shaik et al. (2006) and Harjunkoski et al. (2014).

2.4.6 Heat Integration with Batch Background Processes

Batch processes encountered the same need to reduce energy consumption and waste generation, due to ever more stringent environmental regulation, as their continuous process counterparts. Heat integration in batch processes do however pose an additional difficulty, in that they are constrained by temperature and time. The temperature constraint is also encountered by continuous processes, it ensures that heat cannot be transferred from a heat source to a heat sink if the temperature of the heat sink is greater or equal to the temperature of the heat source. The time constraint is unique to batch processes, it dictates that a heat source cannot transfer heat directly to a heat sink if the heat source and heat sink are not active at a common time interval.

Direct heat integration entails the transfer of heat from a heat source to a heat sink during a time interval where both units are active. Heat storage can be used to bypass the time constraint. To accomplish this, a heat transfer fluid is used to facilitate energy transfer at any time, as long as the heat source took place before the heat sink. This is referred to as indirect heat integration. The use of heat storage thereby improves energy utilisation in batch facilities.

Heat integration models can be applied to batch processes with a pre-determined or variable schedule. When the schedule of the batch plant has been pre-defined, the results from the heat integration model might lead to suboptimal results. This is due to the effect of the batch schedule on heat integration. The optimal batch schedule might entail poor heat integration, or optimal heat integration might necessitate a poor schedule. There is thus a trade-off in terms of batch scheduling, savings from heat integration, heat exchanger sizing, etc. When the batch schedule can vary, the optimisation problem considers these trade-offs and the true optimal solution can be determined. The problem associated with the simultaneous optimisation problem is that the models tend to be very large and complex, leading to computational difficulty.

Initially, the heat integration of batch plants was attempted using graphical methods relying on insights brought about by pinch analysis. Clayton (1986) proposed the time average model (TAM) which assumed continuous behavior of the process, averaging the heat load over the time horizon of the batch schedule.

The time slice model (TSM) developed by Obeng and Ashton (1988) improved this technique by dividing the time horizon into intervals over which the heat load was averaged. Pinch analysis was then carried out on each time slice individually.

Kemp and MacDonald (1987, 1988) provided further improvement with the development of cascade analysis.

The pioneering work with regard to mathematical programming in this research area is Vaselenak et al. (1986). The authors formulated an MILP/heuristic method for maximum heat recovery. The first mathematical model which simultaneously optimised batch scheduling and heat integration is the STN formulated, MINLP model developed by Papageorgiou et al. (1994).

Wang and Smith (1995) adapted the concepts of pinch analysis to develop a graphical approach to heat integration for batch processes. The authors selected time as the primary constraint and temperature as the secondary constraint. Direct and indirect heat integration was considered in the model.

An MILP model was developed, for the minimisation of total costs of a batch process considering direct heat integration, by Vaklieva-Bancheva et al. (1996). The model was restrictive, in that the zero wait condition was applied to all processes.

Majozi (2006a) developed an MILP model based on the continuous time representation and SSN for direct heat integration of a batch process considering both fixed and variable batch sizes. The schedule was flexible and determined as part of the optimisation problem. The formulation required fewer binary variables than models developed prior to their work. The formulation was later extended to consider indirect heat integration where initial storage temperature and storage size are predefined parameters (Majozi, 2009). The work was extended further to optimise initial storage temperature and storage size (Stamp and Majozi, 2011).

Chen and Ciou (2008) formulated an MINLP model for indirect heat integration considering multiple storage vessels. The authors determined that increasing the number of storage vessels does not necessarily improve heat recovery.

Foo et al. (2008) targetted the minimum amount of heat exchanger units required for fixed schedule batch processes. This work was based on similar work for batch mass exchange networks. Targetting was performed both above and below the pinch at each time interval. The authors defined common heat exchangers as the heat exchangers which can be used over multiple time intervals to connect the same hot and cold streams.

Chen and Chang (2009) adapted the work of Majozi (2006a) using the RTN representation to consider one-to-one direct heat integration. Their model incorporated periodic scheduling in addition to the standard short term scheduling approach. The authors further generalised the formulation by allowing heat transfer between tasks to take place at any time where the task is active. This can be distinguished from the case where heat transfer is permitted only from the onset of a task.

Seid and Majozi (2014b) incorporated heat integration into a robust scheduling model (Seid and Majozi, 2012). The model considered direct and indirect heat integration. The formulation made use of the time average model (TAM) such that heat flows were averaged over task processing times. Fewer time points were required, compared to other contemporary models.

Lee et al. (2015) developed a model which considered heat exchange during material transfer between units, as opposed to the traditional in situ heat exchange approach. Intermittently available hot and cold stream pairs were matched during transfer. The authors found that this approach reduced overall processing time requirements, thereby increasing throughput.

Magege and Majozi (2021) introduced the concept of intermittently available material transfer streams. This simplified heat transfer while material is being transferred between units. The authors developed an MINLP model implementing a simultaneous optimisation approach to the scheduling, design and heat integration of a multipurpose batch plant.

2.4.7 Wastewater Minimisation with Batch Background Processes

Shared units in multipurpose batch plants are typically washed between successive tasks to ensure product integrity. This results in a large amount of wastewater generation. Fortunately, techniques have been developed to minimise wastewater generated by the washing operations of batch processes. As with continuous processes, opportunities for reuse, recycling and regeneration exist. Direct reuse entails the immediate use of wastewater generated by one unit to clean another unit. Indirect reuse refers to the storage of water used for cleaning a unit, such that the water can later be used for cleaning of a different unit. This bypasses the time constraint inherent to batch processes. Water can also be recycled back into the same unit. Finally, a regenerator can be used to purify wastewater to a sufficient quality for cleaning purposes.

Wang and Smith (1995) developed a graphical technique which treated concentration as the primary constraint and time as the secondary constraint. Their model was applicable to batch and semi-batch operations.

Almató et al. (1999) developed an NLP model to minimise wastewater by implementing indirect reuse on a fixed schedule batch process. The authors made use of a heuristic method to generate a starting point for the non-linear model.

Majozi et al. (2006) adapted the method developed by Wang and Smith (1995) for processes operating in batch mode only. Both methods are applicable to fixed load problems. Due to the complexity of wastewater minimisation problems involving multiple contaminants, research on wastewater minimisation in batch processes has focused on mathematical techniques.

Majozi (2005a) developed an MILP model which considers direct reuse/recycle for single contaminant fixed load problems. Majozi (2005b) later included the option for indirect reuse. This model was extended to include multiple contaminants (Majozi and Gouws, 2009).

The optimum size of a storage vessel considering a single contaminant was determined by Majozi (2006b) and later extended to multiple storage vessels considering multiple contaminants (Gouws and Majozi, 2008). Adekola and Majozi (2011) included a regenerator in this formulation.

Nonyane and Majozi (2012) presented an approach for wastewater minimisation over long time horizons. The long term cyclical scheduling model introduced by Wu and Ierapetritou (2004) was used to facilitate the direct and indirect reuse of water.

Lee et al. (2014) minimised total interplant wastewater. The plants used in the model were allowed to operate in either batch or continuous mode. Two storage tanks were placed at the entrance and exit nodes of the batch process. This facilitated the modelling of a pseudo-continuous batch process, with regards to the treatment of wastewater. A multi objective function approach was incorporated to minimise the total water storage requirement at minimum wastewater generation.

2.4.8 Combined Heat and Water Minimisation with Batch Background Processes

Recently the simultaneous minimisation of water and energy in production facilities has been explored. This is of particular practical relevance, due to the water required in generating energy, and the energy required to treat and transport water. The interdependence of energy and water is referred to as the water-energy nexus.

Halim and Srinivasan (2011) investigated the water-energy nexus problem within the batch production space. The authors integrated the scheduling of a multipurpose batch plant with direct heat integration and direct water reuse. A sequential optimisation approach was implemented, whereby the process schedule is first optimised. A stochastic search-based integer cut procedure was then implemented to generate alternate schedules. Heat and water integration techniques were then applied to the generated alternate schedules, with the objective of maintaining the optimality of the process schedule.

Adekola et al. (2013) further generalised the solution to the batch water-energy nexus problem, by implemented both direct and indirect heat integration as well as direct and indirect water reuse in their model. The work combined the scheduling model developed by Majozi and Zhu (2001), the wastewater minimisation model of Adekola and Majozi (2011) and the heat integration model of Stamp and Majozi (2011).

Seid and Majozi (2014a) incorporated direct heat and water integration constraints to a robust scheduling model (Seid and Majozi, 2012). The model provided a great degree of flexibility, as units which were heat integrated did not have to share a common starting time. The authors made use of the TAM and TSM techniques in their model, where the time slices were modelled as variables and determined as part of the optimisation.

2.5 Integration of Batch Production and Utility Systems

Utility systems provide external heating and/or cooling to production systems. To ensure the optimal utilisation of resources within the production process, various process integration techniques are incorporated before considering external utilities. Thus, utility systems provide resources to satisfy the heating/cooling demand that could not be met through process integration alone.

The pioneering work of Papoulias and Grossmann (1983) initiated research into the integrated synthesis of production and utility systems. The research in this field has however focused on continuous systems, with integrated synthesis of batch production and utility systems receiving little attention in research until recently.

The earliest work considering utility systems with a batch background process is that of Iyer and Grossmann (1997) who determined the optimal choice of units given a varying hot utility demand due to operation changes in the process schedule. The batch schedule was predefined. A cost penalty was incurred for the startup and shutdown of the boiler.

Behdani et al. (2007) developed an MILP model based on the continuous time representation for batch scheduling with utility considerations. Each unit was assigned a steam, electrical power, and cooling water consumption rate. Utility units were given a maximum capacity, unit cost to produce the utility and a start-up cost. The model placed a much greater emphasis on batch scheduling than on the utility system.

Agha et al. (2010) made use of the RTN formulation and discrete time representation, in integrating a multiproduct batch plant of constant batch size with a steam system. Boiler start-up and shutdown incurred a time penalty. The inlet and outlet temperatures and pressures of the boiler were assumed to be constant. Reuse and recycling of utilities between units were not allowed. The results of the sequential and simultaneous optimisation of their model were compared. Their model struggled to converge, with an average convergence gap of 7.98 %.

Théry et al. (2012) extended the work of Agha et al. (2010) to multipurpose batch plants with variable batch size. The flowsheet was represented by the extended resource task network (ERTN) which the authors developed to provide a clear distinction between batch and continuous tasks.

Recently, Leenders et al. (2019b) developed a mathematical model which explores the synthesis of a batch production and utility system. The scheduling formulation was based on the m-STN framework, as described by Barbosa-Póvoa and Macchietto (1994). The utility system superstructure is based on previous work from their research group and includes a boiler, combined heat and power (CHP) engine, absorption chiller and compression chiller (Voll et al., 2013). Their model resulted in a 5.4 % reduction in total annualised cost (TAC). The authors noted that the energy demand curve experienced less peaks. Significant cost savings were made in the utility system by allowing for a slight increase in the production cost. The authors further developed a heuristic solution procedure, making use of a Stackelberg game, suitable for industrial applications. Stackelberg games are typically formulated with two interacting players, where the leading player makes an initial decision. The subsequent decision made by the follower is then based on the decision made by the leader. For this particular application, pricing information was passed from the utility to the production system. This process was repeated until the problem converged (Leenders et al., 2019a).

It is clear that a greater emphasis has been placed on the integration of hot utilities operating with batch operations. It is worth mentioning that the literature surveyed considered parallel utility systems exclusively. Thus, the reuse of utilities between batch operations has not been considered. The problem is complex due to the discrete nature of batch processes and the necessity of task alignment when considering utility reuse. Furthermore, when integrating cold utilities with batch processes, only the use of absorption and compression chillers have been studied. It is therefore proposed that the integrated optimisation of batch production systems and cooling towers, allowing for utility reuse, be investigated.

2.6 References

Adekola, O. and Majozi, T. (2011) ‘Wastewater minimization in multipurpose batch plants with a regeneration unit: Multiple contaminants’. *Computers & Chemical Engineering*, 35 (12), pp. 2824–2836.

Adekola, O., Stamp, J. D., Majozi, T., Garg, A. and Bandyopadhyay, S. (2013) 'Unified approach for the optimization of energy and water in multipurpose batch plants using a flexible scheduling framework'. *Industrial & Engineering Chemistry Research*, 52 (25), pp. 8488–8506.

Agha, M. H., Théry, R., Hetreux, G., Haït, A. and Le Lann, J. M. (2010) 'Integrated production and utility system approach for optimizing industrial unit operations'. *Energy*, 35 (2), pp. 611–627.

Almató, M., Espuña, A. and Puigjaner, L. (1999) 'Optimisation of water use in batch process industries'. *Computers & Chemical Engineering*, 23 (10), pp. 1427–1437.

Alva-Argáez, A., Kokossis, A. C. and Smith, R. (1998) 'Wastewater minimisation of industrial systems using an integrated approach'. *Computers & Chemical Engineering*, 22 (1), S741–S744.

Barbosa-Póvoa, A. and Macchietto, S. (1994) 'Detailed design and retrofit of multipurpose batch plants'. *Computers & Chemical Engineering*, 18 (11–12), pp. 1013–1042.

Behdani, B., Pishvaie, M. R. and Rashtchian, D. (2007) 'Optimal scheduling of mixed batch and continuous processes incorporating utility aspects'. *Chemical Engineering and Processing: Process Intensification*, 46 (4), pp. 271–281.

Bernier, M. A. (1994) *Cooling tower performance: theory and experiments.* American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., Atlanta, GA (United States).

Bosnjaković, F. (1965) *Technical thermodynamics.* Holt, Rinehart and Winston.

British Standard 4485 (1988) *Water cooling towers, part 2: Methods for performance testing.* British Standards Institution.

Castro, M. M., Song, T. W. and Pinto, J. M. (2000) 'Minimization of operational costs in cooling water systems'. *Chemical Engineering Research and Design*, 78 (2), pp. 192–201.

Castro, P., Barbosa-Póvoa, A. P. F. D. and Matos, H. (2001) 'An improved RTN continuous-time formulation for the short-term scheduling of multipurpose batch plants'. *Industrial & Engineering Chemistry Research*, 40 (9), pp. 2059–2068.

Cerdá, J., Henning, G. P. and Grossmann, I. E. (1997) 'A mixed-integer linear programming model for short-term scheduling of single-stage multiproduct batch plants with parallel lines'. *Industrial & Engineering Chemistry Research*, 36 (5), pp. 1695–1707.

Chen, C.-L. and Chang, C.-Y. (2009) 'A resource-task network approach for optimal short-term/periodic scheduling and heat integration in multipurpose batch plants'. *Applied Thermal Engineering*, 29 (5-6), pp. 1195–1208.

Chen, C.-L. and Ciou, Y.-J. (2008) 'Design and optimization of indirect energy storage systems for batch process plants'. *Industrial & Engineering Chemistry Research*, 47 (14), pp. 4817–4829.

Clayton, R. W. (1986) *Cost reductions on an edible oil refinery identified by a process integration study at Van den Berghs and Jurgens Ltd.* Department of Energy Energy Efficiency Office R&D. Energy Technology Support Unit (ETSU), Harwell University.

Cooling Tower Institute (1997) *CTI code tower standard specifications: Acceptance test code for water-cooling towers.* Cooling Tower Institute.

Dorigo, M., Maniezzo, V., Colorni, A. et al. (1996) 'Ant system: optimization by a colony of cooperating agents'. *IEEE Transactions on Systems, man, and cybernetics, Part B: Cybernetics*, 26 (1), pp. 29–41.

Douglas, J. M. (1988) *Conceptual design of chemical processes.* 1st ed. McGraw-Hill New York.

Doyle, S. J. and Smith, R. (1997) 'Targeting water reuse with multiple contaminants'. *Process Safety and Environmental Protection*, 75 (3), pp. 181–189.

Duran, M. A. and Grossmann, I. E. (1986) 'An outer-approximation algorithm for a class of mixed-integer nonlinear programs'. *Mathematical Programming*, 36 (3), pp. 307–339.

Eberhart, R. and Kennedy, J. (1995) 'Particle swarm optimization'. *Proceedings of the IEEE international conference on neural networks.* Vol. 4. Citeseer, pp. 1942–1948.

Edgar, T. F., Himmelblau, D. M. and Lasdon, L. S. (2001) *Optimization of chemical processes.* McGraw-Hill,

Floudas, C. A. and Lin, X. (2004) 'Continuous-time versus discrete-time approaches for scheduling of chemical processes: a review'. *Computers & Chemical Engineering*, 28 (11), pp. 2109–2129.

Foo, D. C. Y., Chew, Y. H. and Lee, C. T. (2008) 'Minimum units targeting and network evolution for batch heat exchanger network'. *Applied Thermal Engineering*, 28 (16), pp. 2089–2099.

Glover, F. (1975) 'Improved linear integer programming formulations of nonlinear integer problems'. *Management Science*, 22 (4), pp. 455–460.

Gololo, K. V. and Majozi, T. (2012) 'Complex cooling water systems optimization with pressure drop consideration'. *Industrial & Engineering Chemistry Research*, 52 (22), pp. 7056–7065.

Gololo, K. V. and Majozi, T. (2011) 'On synthesis and optimization of cooling water systems with multiple cooling towers'. *Industrial & Engineering Chemistry Research*, 50 (7), pp. 3775–3787.

Gouws, J. F. and Majozi, T. (2008) 'Impact of multiple storage in wastewater minimization for multicontaminant batch plants: Toward zero effluent'. *Industrial & Engineering Chemistry Research*, 47 (2), pp. 369–379.

Grossmann, I. E. and Biegler, L. T. (2004) 'Part II. Future perspective on optimization'. *Computers & Chemical Engineering*, 28 (8), pp. 1193–1218.

Gunaratnam, M., Alva-Argaez, A., Kokossis, A., Kim, J.-K. and Smith, R. (2005) 'Automated design of total water systems'. *Industrial & Engineering Chemistry Research*, 44 (3), pp. 588–599.

Gupta, S. and Karimi, I. A. (2003) 'An improved MILP formulation for scheduling multiproduct, multistage batch plants'. *Industrial & Engineering Chemistry Research*, 42 (11), pp. 2365–2380.

Halim, I. and Srinivasan, R. (2011) 'Sequential methodology for integrated optimization of energy and water use during batch process scheduling'. *Computers & Chemical Engineering*, 35 (8), pp. 1575–1597.

El-Halwagi, M. M. and Manousiouthakis, V. (1989) 'Synthesis of mass exchange networks'. *AIChE Journal*, 35 (8), pp. 1233–1244.

El-Halwagi, M. M. and Manousiouthakis, V. (1990) 'Automatic synthesis of mass-exchange networks with single-component targets'. *Chemical Engineering Science*, 45 (9), pp. 2813–2831.

Harjunkoski, I., Maravelias, C. T., Bongers, P., Castro, P. M., Engell, S., Grossmann, I. E., Hooker, J., Méndez, C., Sand, G. and Wassick, J. (2014) 'Scope for industrial applications of production scheduling models and solution methods'. *Computers & Chemical Engineering*, 62, pp. 161–193.

Hohmann, E. C. J. (1972) 'Optimum networks for heat exchange.' PhD thesis. University of Southern California.

Holland, J. (1975) 'Adaptation in natural and artificial systems: an introductory analysis with application to biology'. *Control and Artificial Intelligence*.

Huang, F. and Elshout, R. (July 1976) 'Optimizing the heat recovery of crude units'. *Chemical Engineering Progress*, 72:7.

Ierapetritou, M. G. and Floudas, C. A. (1998) 'Effective continuous-time formulation for short-term scheduling. 1. Multipurpose batch processes'. *Industrial & Engineering Chemistry Research*, 37 (11), pp. 4341–4359.

Iyer, R. R. and Grossmann, I. E. (1997) 'Optimal multiperiod operational planning for utility systems'. *Computers & Chemical Engineering*, 21 (8), pp. 787–800.

Jaber, H. and Webb, R. L. (1989) 'Design of cooling towers by the effectiveness-NTU method'. *Journal of Heat Transfer*, 111 (4), pp. 837–843.

Janak, S. L., Lin, X. and Floudas, C. A. (2004) 'Enhanced continuous-time unit-specific event-based formulation for short-term scheduling of multipurpose batch processes: Resource constraints and mixed storage policies'. *Industrial & Engineering Chemistry Research*, 43 (10), pp. 2516–2533.

Jiménez-González, C. (2016) 'Embedding sustainability in product and process development-The role of process systems engineers'. In: *Sustainability in the Design, Synthesis and Analysis of Chemical Engineering Processes*. Elsevier, pp. 353–378.

Jung, J., Lee, H.-K. and Lee, I.-B. (1996) 'Completion times algorithm of multiproduct batch processes for common intermediate storage policy (CIS) with nonzero transfer and set-up times'. *Computers & Chemical Engineering*, 20 (6-7), pp. 845–852.

Karaboga, D. and Basturk, B. (2007) 'A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm'. *Journal of Global Optimization*, 39 (3), pp. 459–471.

Karimi, I. A. and McDonald, C. M. (1997) 'Planning and scheduling of parallel semicontinuous processes. 2. Short-term scheduling'. *Industrial & Engineering Chemistry Research*, 36 (7), pp. 2701–2714.

Kelley, J. (1960) 'The cutting plane method for solving convex problems'. *Journal of the Society of Industrial and Applied Mathematics*, 8 (4), pp. 703–712.

Kelly, N. W. (1976) 'A blueprint for the preparation of cross flow cooling tower characteristic curves'. *Paper presented before the Cooling Tower Institute Annual Meeting.*

Kemmer, F. N. (1978) *The Nalco water handbook.* McGraw-Hill.

Kemp, I. C. and MacDonald, E. K. (1987) 'Energy and process integration in continuous and batch processes'. *IChemE Symposium Series.* Vol. 105, pp. 185–200.

Kemp, I. C. and MacDonald, E. K. (1988) 'Application of pinch technology to separation, reaction and batch processes'. *IChemE Symposium Series.* Vol. 109, pp. 239–257.

Kern, D. Q. (1950) *Process heat transfer.* Tata McGraw-Hill Education.

Khan, J., Yaqub, M. and Zubair, M. (2003) 'Performance characteristics of counter flow wet cooling towers'. *Energy Conversion and Management*, 44 (13), pp. 2073–2091.

Khan, J.-U.-R., Qureshi, B. A. and Zubair, S. M. (2004) 'A comprehensive design and performance evaluation study of counter flow wet cooling towers'. *International Journal of Refrigeration*, 27 (8), pp. 914–923.

Khan, J.-R. and Zubair, S. M. (2001) 'An improved design and rating analyses of counter flow wet cooling towers'. *Journal of Heat Transfer*, 123 (4), pp. 770–778.

Kim, J. K. and Smith, R. (2001) 'Cooling water system design'. *Chemical Engineering Science*, 56 (12), pp. 3641–3658.

Kim, J. K. and Smith, R. (2003) 'Automated retrofit design of cooling-water systems'. *AIChE journal*, 49 (7), pp. 1712–1730.

Kim, J.-K. and Smith, R. (2004) 'Cooling system design for water and wastewater minimization'. *Industrial & Engineering Chemistry Research*, 43 (2), pp. 608–613.

Kirkpatrick, S., Gelatt, C. D. and Vecchi, M. P. (1983) 'Optimization by simulated annealing'. *Science*, 220 (4598), pp. 671–680.

Kloppers, J. C. and Kröger, D. G. (2003) 'Loss coefficient correlation for wet-cooling tower fills'. *Applied Thermal Engineering*, 23 (17), pp. 2201–2211.

Kloppers, J. C. and Kröger, D. G. (2005a) 'A critical investigation into the heat and mass transfer analysis of counterflow wet-cooling towers'. *International Journal of Heat and Mass Transfer*, 48 (3-4), pp. 765–777.

Kloppers, J. C. and Kröger, D. G. (2005b) 'Cooling tower performance evaluation: Merkel, Poppe, and e-NTU methods of analysis'. *Journal of Engineering for Gas Turbines and Power*, 127 (1), pp. 1–7.

Kondili, E., Pantelides, C. C. and Sargent, R. W. H. (1993) 'A general algorithm for short-term scheduling of batch operations—I. MILP formulation'. *Computers & Chemical Engineering*, 17 (2), pp. 211–227.

Kopanos, G. M., Laínez, J. M. and Puigjaner, L. (2009) 'An efficient mixed-integer linear programming scheduling framework for addressing sequence-dependent setup issues in batch plants'. *Industrial & Engineering Chemistry Research*, 48 (13), pp. 6346–6357.

Kröger, D. G. (2004) *Air-cooled heat exchangers and cooling towers*. Vol. 1. PennWell Books.

Kuo, W. C. J. and Smith, R. (1998a) 'Designing for the interactions between water-use and effluent treatment'. *Chemical Engineering Research and Design*, 76 (3), pp. 287–301.

Kuo, W.-C. and Smith, R. (1998b) 'Design of water-using systems involving regeneration'. *Process Safety and Environmental Protection*, 76 (2), pp. 94–114.

Land, A. H. and Doig, A. G. (1960) 'An automatic method for solving discrete programming problems'. *Econometrica*, 28, pp. 497–520.

Lee, J.-Y., Chen, C.-L., Lin, C.-Y. and Foo, D. C. Y. (2014) 'A two-stage approach for the synthesis of inter-plant water networks involving continuous and batch units'. *Chemical Engineering Research and Design*, 92 (5), pp. 941–953.

Lee, J.-Y., Seid, E. R. and Majozi, T. (2015) 'Heat integration of intermittently available continuous streams in multipurpose batch plants'. *Computers & Chemical Engineering*, 74, pp. 100–114.

Leenders, L., Bahl, B., Hennen, M. and Bardow, A. (2019a) 'Coordinating scheduling of production and utility system using a Stackelberg game'. *Energy*, 175, pp. 1283–1295.

Leenders, L., Bahl, B., Lampe, M., Hennen, M. and Bardow, A. (2019b) 'Optimal design of integrated batch production and utility systems'. *Computers & Chemical Engineering*.

Li, J. and Li, Z. (2020) 'Model-based optimization of free cooling switchover temperature and cooling tower approach temperature for data center cooling system with water-side economizer'. *Energy and Buildings*, 227, p. 110407.

Li, K. W. and Pridov, A. P. (1985) *Power plant system design*. John Wiley and Sons.

Lin, M.-H., Tsai, J.-F. and Yu, C.-S. (2012) 'A review of deterministic opimization methods in engineering and management'. *Mathematical Problems in Engineering*, 2012 (6-7), pp. 1–15.

Linnhoff, B. and Flower, J. R. (1978a) 'Synthesis of heat exchanger networks: I. Systematic generation of energy optimal networks'. *AIChE Journal*, 24 (4), pp. 633–642.

Linnhoff, B. and Flower, J. R. (1978b) 'Synthesis of heat exchanger networks: II. Evolutionary generation of networks with various criteria of optimality'. *AIChE Journal*, 24 (4), pp. 642–654.

Linnhoff, B. and Hindmarsh, E. (1983) 'The pinch design method for heat exchanger networks'. *Chemical Engineering Science*, 38 (5), pp. 745–763.

Liu, F., Ma, J., Feng, X. and Wang, Y. (2018) 'Simultaneous integrated design for heat exchanger network and cooling water system'. *Applied Thermal Engineering*, 128, pp. 1510–1519.

Ma, J., Wang, Y. and Feng, X. (2018) 'Optimization of multi-plants cooling water system'. *Energy*, 150, pp. 797–815.

Magege, S. R. and Majozi, T. (2021) 'A comprehensive framework for synthesis and design of heat-integrated batch plants: Consideration of intermittently-available streams'. *Renewable and Sustainable Energy Reviews*, 135, pp. 110–125.

Majozi, T. and Zhu, X. X. (2001) 'A novel continuous-time MILP formulation for multipurpose batch plants. 1. Short-term scheduling'. *Industrial & Engineering Chemistry Research*, 40 (25), pp. 5935–5949.

Majozi, T. (2005a) 'An effective technique for wastewater minimisation in batch processes'. *Journal of Cleaner Production*, 13 (15), pp. 1374–1380.

Majozi, T. (2005b) 'Wastewater minimisation using central reusable water storage in batch plants'. *Computers & Chemical Engineering*, 29 (7), pp. 1631–1646.

Majozi, T. (2006a) 'Heat integration of multipurpose batch plants using a continuous-time framework'. *Applied Thermal Engineering*, 26 (13), pp. 1369–1377.

Majozi, T. (2006b) 'Storage design for maximum wastewater reuse in multipurpose batch plants'. *Industrial & Engineering Chemistry Research*, 45 (17), pp. 5936–5943.

Majozi, T. (2009) 'Minimization of energy use in multipurpose batch plants using heat storage: an aspect of cleaner production'. *Journal of Cleaner Production*, 17 (10), pp. 945–950.

Majozi, T. (2010) *Batch chemical process integration: analysis, synthesis and optimization*. Springer Science & Business Media.

Majozi, T., Brouckaert, C. and Buckley, C. (2006) 'A graphical technique for wastewater minimisation in batch processes'. *Journal of Environmental Management*, 78 (4), pp. 317–329.

Majozi, T. and Friedler, F. (2006) 'Maximization of throughput in a multipurpose batch plant under a fixed time horizon: S-graph approach'. *Industrial & Engineering Chemistry Research*, 45 (20), pp. 6713–6720.

Majozi, T. and Gouws, J. F. (2009) 'A mathematical optimisation approach for wastewater minimisation in multipurpose batch plants: multiple contaminants'. *Computers & Chemical Engineering*, 33 (11), pp. 1826–1840.

Majozi, T. and Moodley, A. (2008) 'Simultaneous targeting and design for cooling water systems with multiple cooling water supplies'. *Computers & Chemical Engineering*, 32 (3), pp. 540–551.

Majozi, T. and Nyathi, N. (2007) 'On cooling-water systems design for South African industry: Two recent developments'. *South African Journal of Science*, 103 (5-6), pp. 239–244.

Maravelias, C. T. and Grossmann, I. E. (2003) 'New general continuous-time state-task network formulation for short-term scheduling of multipurpose batch plants'. *Industrial & Engineering Chemistry Research*, 42 (13), pp. 3056–3074.

Mathews, J. H. (1992) *Numerical methods for mathematics, science and engineering*. Prentice-Hall.

McCormick, G. P. (1976) 'Computability of global solutions to factorable nonconvex programs: Part I—Convex underestimating problems'. *Mathematical programming*, 10 (1), pp. 147–175.

Méndez, C. A., Henning, G. P. and Cerdá, J. (2001) 'An MILP continuous-time approach to short-term scheduling of resource-constrained multistage flowshop batch facilities'. *Computers & Chemical Engineering*, 25 (4-6), pp. 701–711.

Méndez, C. A., Cerdá, J., Grossmann, I. E., Harjunkoski, I. and Fahl, M. (2006) 'State-of-the-art review of optimization methods for short-term scheduling of batch processes'. *Computers & Chemical Engineering*, 30 (6-7), pp. 913–946.

Merkel, F. (1926) 'Verdunstungskuhlung'. *VDI Zeitchriff Deustscher Ingenieure*, 70, pp. 123–128.

Mills, A. F. (1999) *Basic heat and mass transfer*. Vol. 2. Prentice Hall Upper Saddle River.

Mohiuddin, A. K. M. and Kant, K. (1996) 'Knowledge base for the systematic design of wet cooling towers. Part I: Selection and tower characteristics'. *International Journal of Refrigeration*, 19 (1), pp. 43–51.

Nonyane, D. R. and Majozi, T. (2012) 'Long term scheduling technique for wastewater minimisation in multipurpose batch processes'. *Applied Mathematical Modelling*, 36 (5), pp. 2142–2168.

Obeng, E. D. A. and Ashton, G. J. (1988) 'On pinch technology based procedures for the design of batch processes'. *Chemical Engineering Research and Design*, 66 (3), pp. 225–259.

Oosthuizen, P. C. (1995) 'Performance characteristics of hybrid cooling towers'. MEng Thesis. Stellenbosch University.

Panjeshahi, M., Ataei, A., Gharaie, M. and Parand, R. (2009) 'Optimum design of cooling water systems for energy and water conservation'. *Chemical Engineering Research and Design*, 87 (2), pp. 200–209.

Pantelides, C. C. (1994) 'Unified frameworks for optimal process planning and scheduling'. *Foundations of Computer Aided Operations*. Cache Publications New York, pp. 253–274.

Papageorgiou, L. G., Shah, N. and Pantelides, C. C. (1994) 'Optimal scheduling of heat-integrated multipurpose plants'. *Industrial & Engineering Chemistry Research*, 33 (12), pp. 3168–3186.

Papoulias, S. A. and Grossmann, I. E. (1983) 'A structural optimization approach in process synthesis—III: Total processing systems'. *Computers & Chemical Engineering*, 7 (6), pp. 723–734.

Pattinson, T. and Majozi, T. (2010) 'Introducing a new operational policy: The PIS operational policy'. *Computers & Chemical Engineering*, 34 (1), pp. 59–72.

Perry, R. H. and Green, D. (1997) *Perry's chemical engineers' handbook*. 7th ed. McGraw-Hill.

Pinto, J. M. and Grossmann, I. E. (1995) 'A continuous time mixed integer linear programming model for short term scheduling of multistage batch plants'. *Industrial & Engineering Chemistry Research*, 34 (9), pp. 3037–3051.

Ponce-Ortega, J. M., Serna-González, M. and Jiménez-Gutiérrez, A. (2007) 'MINLP synthesis of optimal cooling networks'. *Chemical Engineering Science*, 62 (21), pp. 5728–5735.

Ponce-Ortega, J. M., Serna-González, M. and Jiménez-Gutiérrez, A. (2010) 'Optimization model for re-circulating cooling water systems'. *Computers & Chemical Engineering*, 34 (2), pp. 177–195.

Pontes, R. F., Yamauchi, W. M. and Silva, E. K. (2019) 'Analysis of the effect of seasonal climate changes on cooling tower efficiency, and strategies for reducing cooling tower power consumption'. *Applied Thermal Engineering*, 161, p. 114148.

Poppe, M. and Rögener, H. (1991) 'Berechnung von rückkühlwerken'. *VDI-Wärmeatlas*, 37, pp. 1–15.

Price, T. and Majozi, T. (2010) 'On synthesis and optimization of steam system networks. 3. Pressure drop consideration'. *Industrial & Engineering Chemistry Research*, 49 (19), pp. 9165–9174.

Quesada, I. and Grossmann, I. E. (1995) 'Global optimization of bilinear process networks with multicomponent flows'. *Computers & Chemical Engineering*, 19 (12), pp. 1219–1242.

Qureshi, B. A. and Zubair, S. M. (2006) 'A complete model of wet cooling towers with fouling in fills'. *Applied Thermal Engineering*, 26 (16), pp. 1982–1989.

Rao, R. and Patel, V. (2011) 'Optimization of mechanical draft counter flow wet-cooling tower using artificial bee colony algorithm'. *Energy Conversion and Management*, 52 (7), pp. 2611–2622.

Rubio-Castro, E., Serna-González, M., Ponce-Ortega, J. M. and El-Halwagi, M. M. (2013) 'Synthesis of cooling water systems with multiple cooling towers'. *Applied Thermal Engineering*, 50 (1), pp. 957–974.

Ryoo, H. S. and Sahinidis, N. V. (1995) 'Global optimization of nonconvex NLPs and MINLPs with applications in process design'. *Computers & Chemical Engineering*, 19 (5), pp. 551–566.

Sanmartí, E., Friedler, F. and Puigjaner, L. (1998) 'Combinatorial technique for short term scheduling of multipurpose batch plants based on schedule-graph representation'. *Computers & Chemical Engineering*, 22 (Suppl.), S847–S850.

Savelski, M. and Bagajewicz, M. (2003) 'On the necessary conditions of optimality of water utilization systems in process plants with multiple contaminants'. *Chemical engineering science*, 58 (23-24), pp. 5349–5362.

Savelski, M. J. and Bagajewicz, M. J. (2000) 'On the optimality conditions of water utilization systems in process plants with single contaminants'. *Chemical Engineering Science*, 55 (21), pp. 5035–5048.

Schilling, G. and Pantelides, C. (1996) 'A simple continuous-time process scheduling formulation and a novel solution algorithm'. *Computers & Chemical Engineering*, 20, S1221–S1226.

Seid, E. and Majozi, T. (2014a) 'Optimization of energy and water use in multipurpose batch plants using an improved mathematical formulation'. *Chemical Engineering Science*, 111, pp. 335–349.

Seid, E. R. and Majozi, T. (2014b) 'Heat integration in multipurpose batch plants using a robust scheduling framework'. *Energy*, 71, pp. 302–320.

Seid, R. and Majozi, T. (2012) 'A novel technique for prediction of time points for scheduling of multipurpose batch plants'. *Chemical Engineering Science*, 68 (1), pp. 54–71.

Serna-González, M., Ponce-Ortega, J. M. and Jiménez-Gutiérrez, A. (2010) 'MINLP optimization of mechanical draft counter flow wet-cooling towers'. *Chemical Engineering Research and Design*, 88 (5-6), pp. 614–625.

Shah, N., Pantelides, C. C. and Sargent, R. W. H. (1993) 'A general algorithm for short-term scheduling of batch operations—II. Computational issues'. *Computers & Chemical Engineering*, 17 (2), pp. 229–244.

Shaik, M. A. and Floudas, C. A. (2009) 'Novel unified modeling approach for short-term scheduling'. *Industrial & Engineering Chemistry Research*, 48 (6), pp. 2947–2964.

Shaik, M. A., Janak, S. L. and Floudas, C. A. (2006) 'Continuous-time models for short-term scheduling of multipurpose batch plants: A comparative study'. *Industrial & Engineering Chemistry Research*, 45 (18), pp. 6190–6209.

Sherali, H. D. and Alameddine, A. (1992) 'A new reformulation-linearization technique for bilinear programming problems'. *Journal of Global Optimization*, 2 (4), pp. 379–410.

Snyman, J. A. and Wilke, D. N. (2018) *Practical mathematical optimization: Basic optimization theory and gradient-based algorithms.* Vol. 133. Springer.

Stamp, J. and Majozi, T. (2011) 'Optimum heat storage design for heat integrated multipurpose batch plants'. *Energy*, 36 (8), pp. 5119–5131.

Sundaramoorthy, A. and Karimi, I. A. (2005) 'A simpler better slot-based continuous-time formulation for short-term scheduling in multipurpose batch plants'. *Chemical Engineering Science*, 60 (10), pp. 2679–2702.

Susarla, N., Li, J. and Karimi, I. (2010) 'A novel approach to scheduling multipurpose batch plants using unit-slots'. *AIChE Journal*, 56 (7), pp. 1859–1879.

Takama, N., Kuriyama, T., Shiroko, K. and Umeda, T. (1980) 'Optimal water allocation in a petroleum refinery'. *Computers & Chemical Engineering*, 4 (4), pp. 251–258.

Théry, R., Hetreux, G., Agha, M. H., Haït, A. and Le Lann, J. M. (2012) 'The extended resource task network: a framework for the combined scheduling of batch processes and CHP plants'. *International Journal of Production Research*, 50 (3), pp. 623–646.

Umeda, T., Harada, T. and Shiroko, K. (1979) 'A thermodynamic approach to the synthesis of heat integration systems in chemical processes'. *Computers & Chemical Engineering*, 3 (1-4), pp. 273–282.

Vaklieva-Bancheva, N., Ivanov, B. B., Shah, N. and Pantelides, C. C. (1996) 'Heat exchanger network design for multipurpose batch plants'. *Computers & Chemical Engineering*, 20 (8), pp. 989–1001.

Vaselenak, J. A., Grossmann, I. E. and Westerberg, A. W. (1986) 'Heat integration in batch processing'. *Industrial & Engineering Chemistry Process Design and Development*, 25 (2), pp. 357–366.

Viswanathan, J. and Grossmann, I. E. (1990) 'A combined penalty function and outer-approximation method for MINLP optimization'. *Computers & Chemical Engineering*, 14 (7), pp. 769–782.

Voll, P., Klaffke, C., Hennen, M. and Bardow, A. (2013) 'Automated superstructure-based synthesis and optimization of distributed energy supply systems'. *Energy*, 50, pp. 374–388.

Vooradi, R. and Shaik, M. A. (2012) 'Improved three-index unit-specific event-based model for short-term scheduling of batch plants'. *Computers & Chemical Engineering*, 43, pp. 148–172.

Wang, Y.-P. and Smith, R. (1994a) 'Design of distributed effluent treatment systems'. *Chemical Engineering Science*, 49 (18), pp. 3127–3145.

Wang, Y. P. and Smith, R. (1994b) 'Wastewater minimisation'. *Chemical Engineering Science*, 49 (7), pp. 981–1006.

Wang, Y. P. and Smith, R. (1995) 'Time pinch analysis'. *Chemical Engineering Research & Design*, 73 (8), pp. 905–914.

Westerlund, T. and Pettersson, F. (1995) 'An extended cutting plane method for solving convex MINLP problems'. *Computers & Chemical Engineering*, 19, pp. 131–136.

Wiede, W. and Reklaitis, G. (1987) 'Determination of completion times for serial multiproduct processes—3. Mixed intermediate storage systems'. *Computers & Chemical Engineering*, 11 (4), pp. 357–368.

Williams, H. P. (2013) *Model building in mathematical programming.* John Wiley & Sons.

Woolway, M. and Majozi, T. (2018) 'A novel metaheuristic framework for the scheduling of multipurpose batch plants'. *Chemical Engineering Science*, 192, pp. 678–687.

Wu, D. and Ierapetritou, M. (2004) 'Cyclic short-term scheduling of multiproduct batch plants using continuous-time representation'. *Computers & Chemical Engineering*, 28 (11), pp. 2271–2286.

Wu, Z., Lu, Z., Zhang, B., He, C., Chen, Q., Yu, H. and Ren, J. (2021) 'Stochastic Bi-Objective Optimization for Closed Wet Cooling Tower Systems Based on a Simplified Analytical Model'. *SSRN 3901835*.

Yee, T. F. and Grossmann, I. E. (1990) 'Simultaneous optimization models for heat integration—I. Area and energy targeting and modeling of multi-stream exchangers'. *Computers & Chemical Engineering*, 14 (10), pp. 1151–1164.

Zhang, X. and Sargent, R. W. H. (1996) 'The optimal operation of mixed production facilities—a general formulation and some approaches for the solution'. *Computers & Chemical Engineering*, 20 (6-7), pp. 897–904.

Zheng, C., Chen, X., Zhu, L. and Shi, J. (2018) 'Simultaneous design of pump network and cooling tower allocations for cooling water system synthesis'. *Energy*, 150, pp. 653–669.

Zhu, X., Wang, F., Niu, D. and Zhao, L. (2017) 'Integrated modeling and operation optimization of circulating cooling water system based on superstructure'. *Applied Thermal Engineering*, 127, pp. 1382–1390.

Chapter 3

Model Development/Formulation

The mathematical formulation which has been developed can be divided into three distinct sections. These are the batch scheduling model, cooling water system model and the costing model. Figure 3.1 on page 92 depicts the superstructure involved in the development of the mathematical model.

3.1 Batch Scheduling Model

The batch scheduling model presented in this section was developed by Maravelias and Grossmann (2003). The model is in the form of an MILP, making use of a continuous time representation. The time representation is classified as a global time point formulation, where a common time grid exists for all tasks and resources. This allows for ease of handling shared resources.

3.1.1 Assignment constraints

The assignment constraints below dictate the assignment of tasks to unit operations. Constraints 3.1 and 3.2 ensure that a maximum of one task can start or finish in a unit j at any time point p.

$$\sum_{i \in I_j(j)} W_s(i,p) \leq 1 \qquad \forall j \in J, p \in P \tag{3.1}$$

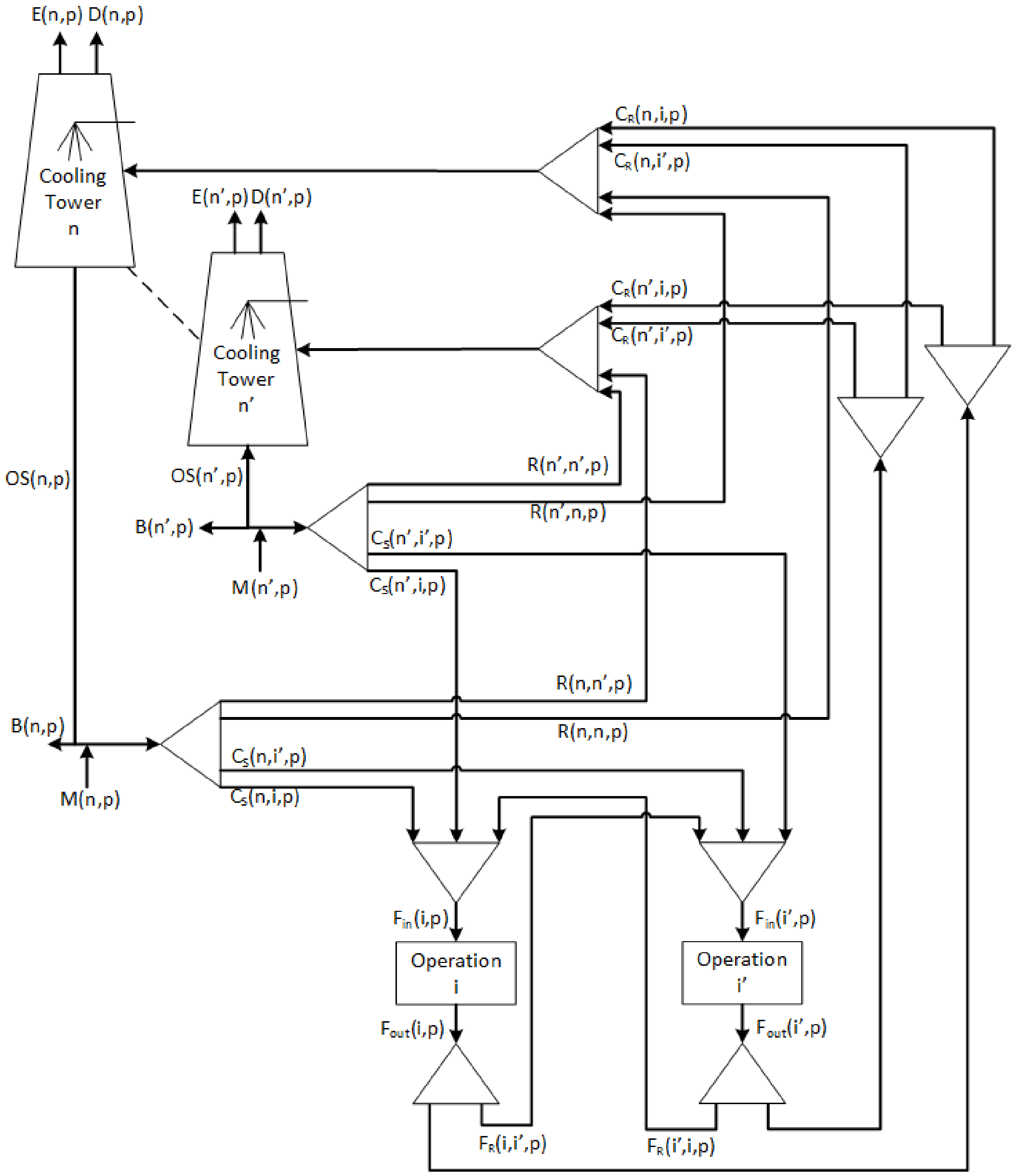


Figure 3.1: Superstructure

$$\sum_{i \in I_j(j)} W_f(i,p) \leq 1 \qquad \forall j \in J, p \in P \tag{3.2}$$

Constraint 3.3 stipulates that all tasks that start must also finish.

$$\sum_{p \in P} W_s(i,p) = \sum_{p \in P} W_f(i,p) \qquad \forall i \in I \tag{3.3}$$

Constraint 3.4 ensures that at most one task can be active in any unit operation j at any time point p.

$$\sum_{i \in I_j(j)} \sum_{p' \in P}^{p' \leq p} \left(W_s(i,p') - W_f(i,p)\right) \leq 1 \qquad \forall j \in J, p \in P \tag{3.4}$$

The conditions that tasks cannot finish at the first time point or start at the final time point are enforced by Constraints 3.5 and 3.6 respectively.

$$W_f(i,p) = 0 \qquad \forall i \in I, p \in P, p = p_1 \tag{3.5}$$

$$W_s(i,p) = 0 \qquad \forall i \in I, p \in P, p = |P| \tag{3.6}$$

3.1.2 Duration and sequencing constraints

The duration and sequencing constraints below are concerned with task duration, the time horizon and the sequencing of the time points. Constraints 3.7 and 3.8 ensure that the first time point occurs at the start of the time horizon and the final time point occurs at the end of the time horizon, respectively.

$$t(p) = 0 \qquad \forall p \in P, p = p_1 \tag{3.7}$$

$$t(p) = H \qquad \forall p \in P, p = |P| \tag{3.8}$$

Constraint 3.9 states that the time at which each time point takes place must be greater than or equal to the time at which the previous time point took place.

$$t(p+1) \geq t(p) \qquad \forall p \in P, p \neq |P| \tag{3.9}$$

The duration constraint is depicted by Constraint 3.10. It is worth mentioning that the duration of each task is made up of a constant processing term, depending on whether the task is active, and a variable processing term, depending on batch size.

$$\tau(i,p) = \alpha(i) W_s(i,p) + \beta(i) B_s(i,p) \qquad \forall i \in I, p \in P \tag{3.10}$$

The finish time of a task at a particular time point is expressed by making use of big-M constraints. These constraints are active only if the particular task i starts at time point p. This can be seen from Constraints 3.11 and 3.12.

$$t_f(i,p) \leq t_s(i,p) + \tau(i,p) + H(1 - W_s(i,p)) \qquad \forall i \in I, p \in P \tag{3.11}$$

$$t_f(i,p) \geq t_s(i,p) + \tau(i,p) - H(1 - W_s(i,p)) \qquad \forall i \in I, p \in P \tag{3.12}$$

Constraint 3.13 uses a big-M constraint to ensure that the finish time of a particular task i at time point p remains unchanged until the next occurrence of task i.

$$t_f(i,p) - t_f(i,p-1) \leq H \cdot W_s(i,p) \qquad \forall i \in I, p \in, p > p_1 \tag{3.13}$$

Constraints 3.14 and 3.15 stipulate that the finish time of a task i at time point p is greater than or equal to the finish time of task i at the previous time point. Should the particular task i take place at time point p, these constraints ensure that the difference in finishing times for the successive time points should at least be equal to the task duration.

$$t_f(i,p) - t_f(i,p-1) \geq \tau(i,p) \qquad \forall i \in I, p \in P, p > p_1 \tag{3.14}$$

$$t_f(i,p) \geq \tau(i,p) \qquad \forall i \in I, p \in P, p = p_1 \tag{3.15}$$

Constraint 3.16 coordinates the time at which time point p occurs to coincide with the start time of any task i at time point p.

$$t_s(i,p) = t(p) \qquad \forall i \in I, p \in P \tag{3.16}$$

Constraints 3.17 and 3.18 are known as the time-matching constraints for the finishing time of a task. The general equation allows for a task i to finish at or before a time point p. However, if the task requires cooling water or produces a zero wait state, the finish time of task i is forced to coincide with time point p.

$$t_f(i,p-1) \le t(p) + H\left(1 - W_f(i,p)\right) \qquad \forall i \in I, p \in P, p > p_1 \tag{3.17}$$

$$t_f(i,p-1) \ge t(p) - H\left(1 - W_f(i,p)\right) \qquad \forall i \in I_z, p \in P, p > p_1 \tag{3.18}$$

3.1.3 Batch size constraints

Batch size constraints impose limits on the amount of material that can be processed in a particular task. Constraints 3.19–3.24 impose minimum and maximum bounds on the batch size of tasks.

$$B_s(i,p) \ge B^L(i,p)\, W_s(i,p) \qquad \forall i \in I, p \in P \tag{3.19}$$

$$B_s(i,p) \le B^U(i,p)\, W_s(i,p) \qquad \forall i \in I, p \in P \tag{3.20}$$

$$B_f(i,p) \ge B^L(i,p)\, W_f(i,p) \qquad \forall i \in I, p \in P \tag{3.21}$$

$$B_f(i,p) \le B^U(i,p)\, W_f(i,p) \qquad \forall i \in I, p \in P \tag{3.22}$$

$$B_p(i,p) \ge B^L(i,p) \left(\sum_{p' \in P}^{p' < p} W_s(i,p') - \sum_{p' \in P}^{p' \le p} W_f(i,p') \right) \qquad \forall i \in I, p \in P \tag{3.23}$$

$$B_p(i,p) \leq B^U(i,p)\left(\sum_{p' \in P}^{p'<p} W_s(i,p') - \sum_{p' \in P}^{p' \leq p} W_f(i,p')\right) \qquad \forall i \in I, p \in P \quad (3.24)$$

Constraint 3.25 is a mass balance over two successive time points. Every task that has started at a previous time point can either continue being processed or finish at the following time point.

$$B_s(i,p-1) + B_p(i,p-1) = B_p(i,p) + B_f(i,p) \qquad \forall i \in I, p \in P, p > p_1 \quad (3.25)$$

The amount of a particular state s consumed or produced by task i at time point p is calculated by Constraints 3.26 and 3.27. The bounds on these variables are set by Constraints 3.28 and 3.29.

$$B_I(i,s,p) = \rho(i,s)\, B_s(i,p) \qquad \forall i \in I, s \in S_i(i), p \in P \quad (3.26)$$

$$B_O(i,s,p) = \rho(i,s)\, B_f(i,p) \qquad \forall i \in I, s \in S_o(i), p \in P \quad (3.27)$$

$$B_I(i,s,p) \leq B^U(i)\, \rho(i,s)\, W_s(i,p) \qquad \forall i \in I, s \in S_i(i), p \in P \quad (3.28)$$

$$B_O(i,s,p) \leq B^U(i)\, \rho(i,s)\, W_f(i,p) \qquad \forall i \in I, s \in S_o(i), p \in P \quad (3.29)$$

3.1.4 Mass balance constraints

The mass balance constraints ensure that the amount of material is conserved over each task. Constraints 3.30 and 3.31 ensure that the amount of state s available at any time point p can either continue being stored or could be used by a task i. Furthermore, should state s be produced from a task i, this state can also be used for further processing.

$$S_A(s,p) + S_S(s,p) = S_A(s,p-1) + \sum_{i \in I_o(s)} B_O(i,s,p) - \sum_{i \in I_i(s)} B_I(i,s,p)$$
$$\forall s \in S, p \in P, p > p_1 \quad (3.30)$$

$$S_A(s,p) + S_S(s,p) = S_A^0(s) + \sum_{i \in I_o](s)} B_O(i,s,p) - \sum_{i \in I_i(s)} B_I(i,s,p)$$
$$\forall s \in S, p \in P, p = p_1 \quad (3.31)$$

Constraint 3.32 imposes the upper bound of available material for each state s.

$$S_A(s,p) \leq S_A^U(s) \qquad \forall s \in S, p \in P \tag{3.32}$$

3.1.5 Tightening constraints

The tightening constraints leads to improved solution times. Constraint 3.33 enforces that the sum of all task durations be less than or equal to the time horizon of interest.

$$\sum_{i \in I_j(j)} \sum_{p \in P} \tau(i,p) \leq H \qquad \forall j \in J \tag{3.33}$$

Constraints 3.34 and 3.35 together enforce that the sum of task durations in a particular slot does not exceed the upcoming time point. In isolation, Constraint 3.34 enforces the summation of the processing times of all tasks i which start on a unit j after $t(p)$ to be less than the total time remaining in the time horizon. Constraint 3.35 forces the durations of tasks which finish processing before time point p to be less than or equal to the time associated with time point p, i.e. $t(p)$.

$$\sum_{i \in I_j(j)} \sum_{p \in P}^{p' \geq p} \tau(i,p') \leq H - t(p) \qquad \forall j \in J, p \in P \tag{3.34}$$

$$\sum_{i \in I_j(j)} \sum_{p' \in P}^{p' \leq p} \left(\alpha\left(i\right) W_f\left(i, p'\right) + \beta\left(i\right) B_f\left(i, p'\right) \right) \leq T\left(p\right)$$

$$\forall j \in J, p \in P \quad (3.35)$$

3.2 Cooling Water Network Model

The basis of the cooling water network model presented in this section was developed by Majozi and Moodley (2008). The model has been adapted to accommodate the batch scheduling model. This involves the calculation of each time-dependent variable at every time point p. Furthermore, constraints for cooling tower losses and makeup water have been introduced into the formulation. The model takes the form of an MINLP problem with bilinear terms which cannot be exactly linearised. To obtain a feasible starting point for the exact MINLP problem, the model is linearised as described by Quesada and Grossmann (1995) and solved prior to the exact model.

3.2.1 Mass balance constraints

The mass balance constraints ensure that the total cooling water at each node is conserved. Constraint 3.36 defines the total cooling water supply as the sum of the cooling water from each cooling tower n.

$$CW = \sum_{n \in N} OS\left(n, p\right) \qquad \forall p \in P, p \neq |P| \tag{3.36}$$

Constraints 3.37 and 3.38 define the outlet and inlet cooling water flow for each cooling tower n at every time point p. Constraint 3.37 stipulates that the sum of the cooling water flowing through a cooling tower n at time point p and the makeup water introduced to cooling tower n at time point p is made up of:

- the cooling water supplied to all tasks from the cooling tower n at time point p

- the total cooling water recycled from cooling tower n to a cooling tower n'
- the water loss due to blowdown from cooling tower n at time point p

$$OS\left(n,p\right)=\sum_{i\in I}C_S\left(n,i,p\right)+\sum_{n'\in N}R\left(n,n',p\right)-M\left(n,p\right)+B\left(n,p\right)$$
$$\forall n\in N,p\in P \quad (3.37)$$

Constraint 3.38 stipulates that the sum of the cooling water flowing through a cooling tower n at time point p and the water losses from cooling tower n at time point p due to drift and evaporation is made up of:

- the cooling water returned from all tasks to the cooling tower n at time point p
- the total cooling water recycled from a cooling tower n' to cooling tower n.

$$OS\left(n,p\right)=\sum_{i\in I}C_R\left(n,i,p\right)+\sum_{n'\in N}R\left(n',n,p\right)-D\left(n,p\right)-E\left(n,p\right)$$
$$\forall n\in N,p\in P \quad (3.38)$$

Constraint 3.39 defines the loss due to evaporation experienced by cooling tower n at time point p. This is an empirical equation developed by Perry and Green (1997). It is worth mentioning that the evaporation rate is dependent on the amount of water entering the cooling tower and the difference between the cooling water inlet and outlet temperatures. This equation is nonlinear due to the multiplication of the continuous variables. The linearised form of the equation used in the initial model can be seen from Constraint 3.73.

$$E(n,p) = 0.00085 \cdot 1.8 \left(\sum_{i \in I} C_R(n,i,p) + \sum_{n' \in N} R(n',n,p) \right) \cdot \left(TW_{in}(n,p) - TW_{out}(n) \right) \qquad \forall n \in N, p \in P \quad (3.39)$$

Kemmer (1978) determined that drift losses should not exceed 0.2% of the total circulating water flowrate in well-designed cooling towers. From Constraint 3.40 it can be seen that a worst case scenario has been assumed for drift losses.

$$D(n,p) = 0.002 \left(\sum_{i \in I} C_R(n,i,p) + \sum_{n' \in N} R(n',n,p) \right) \qquad \forall n \in N, p \in P \quad (3.40)$$

Blowdown is defined in Constraint 3.41. The total blowdown for the cooling tower is dependent on the cycles of concentration.

$$B(n,p) = \frac{E(n,p)}{CC - 1} \qquad \forall n \in N, p \in P \quad (3.41)$$

The makeup water introduced to the cooling water system should equal all of the above defined losses. This is captured by Constraint 3.42.

$$M(n,p) = B(n,p) + D(n,p) + E(n,p) \qquad \forall n \in N, p \in P \quad (3.42)$$

Constraints 3.43 and 3.44 define the cooling water inlet and outlet flow for each task i at time point p. It can be seen that cooling water can be reused from task i to a task i'.

$$F_{in}(i,p) = \sum_{n \in N} C_S(n,i,p) + \sum_{i' \in I}^{i' \neq i} F_r(i',i,p) \qquad \forall i \in I, p \in P \quad (3.43)$$

$$F_{out}(i,p) = \sum_{n \in N} C_R(n,i,p) + \sum_{i' \in I}^{i' \neq i} F_r(i,i',p) \qquad \forall i \in I, p \in P \quad (3.44)$$

Constraint 3.45 states that the cooling water flow entering a task must be the same as the cooling water flow exiting a task.

$$F_{in}\left(i,p\right) = F_{out}\left(i,p\right) \qquad \forall i \in I, p \in P \tag{3.45}$$

Constraints 3.46 and 3.47 are used to determine the amount of active cooling towers. Constraint 3.46 also bounds the cooling water flow passing through each cooling tower n to be equal to or less than the design capacity of each cooling tower.

$$OS\left(n,p\right) \le OS^{U}\left(n\right) y_{CT}\left(n\right) \qquad \forall n \in N, p \in P \tag{3.46}$$

$$CT = \sum_{n \in N} y_{CT}\left(n\right) \tag{3.47}$$

Constraint 3.48 restricts the amount of cooling water recycled directly from cooling tower n to cooling tower n' to be less than or equal to the design capacity of cooling tower n'.

$$\sum_{n' \in N} R\left(n,n',p\right) \le OS^{U}\left(n'\right) \qquad \forall n \in N, p \in P \tag{3.48}$$

Constraint 3.49 bounds the cooling water flow into task i. Additionally, this constraint forces the flow into task i at time point p to be zero, should task i be inactive at time point p.

$$F_{in}\left(i,p\right) \le F_{in}^{U}\left(i\right) \sum_{p' \in P}^{p' \le p} \left(W_s\left(i,p'\right) - W_f\left(i,p'\right)\right) \qquad \forall i \in I, p \in P \tag{3.49}$$

Where $F_{in}^{U}\left(i\right)$ is defined as in Constraint 3.50

$$F_{in}^{U}\left(i\right) = \frac{B^{U}\left(i\right) Q\left(i\right)}{c_p\left(T_{out}^{U}\left(i\right) - T_{in}^{U}\left(i\right)\right)} \qquad \forall i \in I \tag{3.50}$$

Constraints 3.51–3.54 bounds the amount of cooling water reused from task i to a task i' to be less than the upper bound of cooling water flow to both tasks i and i'. Additionally, this constraint forces the reused cooling water at time point p to be zero, should either task i or task i' be inactive at time point p.

$$F_r\left(i',i,p\right) \leq F_{in}^U\left(i\right) \sum_{p' \in P}^{p' \leq p} \left(W_s\left(i,p'\right) - W_f\left(i,p'\right)\right) \qquad \forall i,i' \in I, p \in P, i' \neq i \tag{3.51}$$

$$F_r\left(i',i,p\right) \leq F_{in}^U\left(i\right) \sum_{p' \in P}^{p' \leq p} \left(W_s\left(i',p'\right) - W_f\left(i',p'\right)\right) \qquad \forall i,i' \in I, p \in P, i' \neq i \tag{3.52}$$

$$F_r\left(i',i,p\right) \leq F_{in}^U\left(i'\right) \sum_{p' \in P}^{p' \leq p} \left(W_s\left(i,p'\right) - W_f\left(i,p'\right)\right) \qquad \forall i,i' \in I, p \in P, i' \neq i \tag{3.53}$$

$$F_r\left(i',i,p\right) \leq F_{in}^U\left(i'\right) \sum_{p' \in P}^{p' \leq p} \left(W_s\left(i',p'\right) - W_f\left(i',p'\right)\right) \qquad \forall i,i' \in I, p \in P, i' \neq i \tag{3.54}$$

Reused cooling water is further bounded by Constraints 3.55 and 3.56. These constraints state that the reused cooling water must be less than or equal to the total amount of cooling water entering tasks i or i' at time point p.

$$F_r\left(i',i,p\right) \leq F_{in}\left(i,p\right) \qquad \forall i,i' \in I, p \in P, i' \neq i \tag{3.55}$$

$$F_r\left(i',i,p\right) \leq F_{in}\left(i',p\right) \qquad \forall i,i' \in I, p \in P, i' \neq i \tag{3.56}$$

Constraints 3.57 and 3.58 together prevents multiple instances of cooling water reuse between any two tasks, i.e. cooling water is not allowed to be reused from task i to task i' and then back again to task i.

$$F_r\left(i',i,p\right) \leq F_{in}^U\left(i\right) y_r\left(i',i,p\right) \qquad \forall i,i' \in I, p \in P, i' \neq i \tag{3.57}$$

$$y_r\left(i',i,p\right) + y_r\left(i,i',p\right) \leq 1 \qquad \forall i,i' \in I, p \in P, i' \neq i \tag{3.58}$$

3.2.2 Energy balance constraints

Constraints 3.59–3.62 determine the cooling duty required for each task i at an event point p. The amount of cooling duty introduced by task i during time point p is calculated by constraint 3.59. Similarly, constraint 3.60 describes the amount of cooling duty which is no longer required by task i at the completion of time point p.

$$Q_i\left(i,p\right) = Q\left(i\right) W_s\left(i,p\right) \qquad \forall i \in I, p \in P \tag{3.59}$$

$$Q_o\left(i,p\right) = Q\left(i\right) W_f\left(i,p\right) \qquad \forall i \in I, p \in P \tag{3.60}$$

Constraints 3.61 and 3.62 calculate the total amount of cooling duty required by task i at every time point p.

$$Q_u\left(i,p\right) = Q_u\left(i,p-1\right) - Q_o\left(i,p\right) + Q_i\left(i,p\right) \qquad \forall i \in I, p \in P, p > p_1 \tag{3.61}$$

$$Q_u\left(i,p\right) = Q_i\left(i,p\right) \qquad \forall i \in I, p \in P, p = p_1 \tag{3.62}$$

Constraints 3.63–3.66 are all in nonlinear form due to the multiplication of the continuous variables. The linearised form of these equations used in the initialisation model can be seen from constraints 3.74–3.76.

The energy balance over each task i at time point p is demonstrated in Constraint 3.63. The specific heat capacity is assumed to be constant.

$$Q_u(i,p) + c_p\left(\sum_{n\in N} C_S(n,i,p)\,T_{sup}(n) + \sum_{i'\in I}^{i'\neq i} F_r(i',i,p)\,T_{out}(i',p)\right) = c_p F_{out}(i,p)\,T_{out}(i,p) \qquad \forall i\in I, p\in P \quad (3.63)$$

Constraint 3.64 calculates the inlet temperature to a task i at time point p.

$$T_{in}(i,p) = \frac{\sum\limits_{n\in N} C_S(n,i,p)\,T_{sup}(n,p) + \sum\limits_{i'\in I}^{i'\neq i} F_r(i',i,p)\,T_{out}(i',p)}{F_{in}(i,p)} \qquad \forall i\in I, p\in P \quad (3.64)$$

The temperature of the cooling water returned to cooling tower n at time point p is described by Constraint 3.65.

$$T_{ret}(n,p) = \frac{\sum\limits_{i\in I} C_R(n,i,p)\,T_{out}(i,p) + \sum\limits_{n'\in N} R(n',n,p)\,T_{ct}(n')}{\sum\limits_{i\in I} C_R(n,i,p) + \sum\limits_{n'\in N} R(n',n,p)} \qquad \forall n\in N, p\in P \quad (3.65)$$

Constraint 3.66 describes the supply temperature to task i at time point p after makeup water is introduced into the system.

$$T_{sup}(n,p) = \frac{M(n,p)\,T_{amb} + (OS(n,p) - B(n,p))\cdot T_{ret}(n)}{\sum\limits_{i\in I} C_R(n,i,p) + \sum\limits_{n'\in N} R(n,n',p)} \qquad \forall n\in N, p\in P \quad (3.66)$$

The cooling tower effectiveness is defined as the sum of actual heat transferred by the cooling tower divided by the theoretical maximum heat transfer of the cooling tower. Equation (3.67) defines the effectiveness of cooling tower n at time point p.

$$\epsilon\left(n,p\right)=\frac{H_{a,out}\left(n,p\right)-H_{a,in}\left(n,p\right)}{H_{s,w}\left(n,p\right)-H_{a,in}\left(n,p\right)}\qquad\forall n\in N,p\in P\tag{3.67}$$

Where $H_{a,out}\left(n,p\right)$ is the enthalpy of the air exiting cooling tower n at time point p, $H_{a,in}\left(n,p\right)$ is the enthalpy of the air entering cooling tower n at time point p, and $H_{s,w}\left(n,p\right)$ is the enthalpy of the saturated water in cooling tower n at time point p.

Assuming that the heat capacity rate remains constant over the cooling tower and that the cooling water can be cooled to the wet bulb temperature, the cooling tower effectiveness of cooling tower n at time point p can be written as per Equation (3.68) (Kröger, 2004).

$$\epsilon\left(n,p\right)=\frac{T_{ret}\left(n,p\right)-T_{ct}}{T^{U}_{ret}\left(n\right)-T_{wb}}\qquad\forall n\in N,p\in P\tag{3.68}$$

Constraints 3.69–3.72 provide upper and lower bounds for the temperature related variables.

$$T_{in}\left(i,p\right)\leq T^{U}_{in}\left(i\right)\sum_{p'\in P}^{p'\leq p}\left(W_{s}\left(i,p'\right)-W_{f}\left(i,p'\right)\right)\qquad\forall i\in I,p\in P\tag{3.69}$$

$$T_{out}\left(i,p\right)\leq T^{U}_{out}\left(i\right)\sum_{p'\in P}^{p'\leq p}\left(W_{s}\left(i,p'\right)-W_{f}\left(i,p'\right)\right)\qquad\forall i\in I,p\in P\tag{3.70}$$

$$T_{out}\left(i,p\right)\geq T^{L}_{out}\sum_{p'\in P}^{p'\leq p}\left(W_{s}\left(i,p'\right)-W_{f}\left(i,p'\right)\right)\qquad\forall i\in I,p\in P\tag{3.71}$$

$$T_{ret}\left(n,p\right)\leq T^{U}_{ret}\left(n\right)\qquad\forall n\in N,p\in P\tag{3.72}$$

3.2.3 Linear constraints

The model contains nonlinear constraints due to the multiplication of continuous variables in constraints 3.39 and 3.63–3.66. This renders the model computationally difficult to solve. To provide the exact model with a good starting point, these constraints are linearised below. The linear model is then solved before the exact nonlinear model.

Constraint 3.73 is the linearised form of constraint 3.39. The linearisation variables denoted by γ are shown from constraints 3.81–3.109.

$$E(n,p) = 0.00085 \cdot 1.8 \sum_{i \in I} \left(\gamma_5(n,i,p) - C_R(n,i,p)\, T_{ct}(n) \right) \qquad \forall n \in N, p \in P \quad (3.73)$$

The exact form of constraint 3.63 can be seen in constraint 3.74.

$$Q_u(i,p) + c_p \left(\sum_{n \in N} \gamma_4(n,i,p) + \sum_{i' \in I}^{i' \neq i} \gamma_2(i',i,p) \right) = c_p \gamma_3(i,p) \qquad \forall i \in I, p \in P \quad (3.74)$$

Constraint 3.75 is the linearised form of constraint 3.65.

$$\sum_{i \in I} \gamma_1(n,i,p) + \sum_{n' \in N} R(n',n,p)\, T_{ct}(n') \leq T_{ret}^U(n) \sum_{i \in I} C_R(n,i,p) \qquad \forall n \in N, p \in P \quad (3.75)$$

The exact form of constraint 3.66 can be seen in constraint 3.76.

$$\sum_{i \in I} \gamma_4(n,i,p) + \sum_{n' \in N} \gamma_6(n',n,p) = M(n,p)\, T_{amb} + \left(OS(n,p) - B(n,p) \right) \cdot T_{ct}(n) \qquad \forall n \in N, p \in P \quad (3.76)$$

Constraint 3.77 describes the approximated linearised profit. As can be seen, the operational costs have also been approximated to provide a good starting point for the exact nonlinear objective function.

$$\max Profit_L = \sum_{s\in S}\sum_{p\in P}\zeta\left(s\right)SS\left(s,p\right) - c_{TR} - c_{TC} - \frac{H}{H_Y}\sum_{p\in P}c_{PO}\left(p\right) \tag{3.77}$$

$$W_r\left(i',i,p\right) \leq \sum_{p'\in P}^{p'\leq p}\left(W_s\left(i,p'\right) - W_f\left(i,p'\right)\right) \qquad \forall i,i'\in I, p\in P, i'\neq i \tag{3.78}$$

$$W_r\left(i',i,p\right) \leq \sum_{p'\in P}^{p'\leq p}\left(W_s\left(i',p'\right) - W_f\left(i',p'\right)\right) \qquad \forall i,i'\in I, p\in P, i'\neq i \tag{3.79}$$

$$W_r\left(i',i,p\right) = W_r\left(i,i',p\right) \qquad \forall i,i'\in I, p\in P, i'\neq i \tag{3.80}$$

$\gamma_1\left(n,i,p\right)$ is the linearisation variable for the term $C_R\left(n,i,p\right)T_{out}\left(i\right)$. Constraints 3.81–3.84 are used in defining this term.

$$\gamma_1\left(n,i,p\right) \geq F_{in}^U\left(i\right)T_{out}\left(i,p\right) + C_R\left(n,i,p\right)T_{out}^U\left(i\right) - F_{in}^U\left(i\right)T_{out}^U\left(i\right) \qquad \forall n\in N, i\in I, p\in P \tag{3.81}$$

$$\gamma_1\left(n,i,p\right) \leq F_{in}^U\left(i\right)T_{out}\left(i,p\right) + C_R\left(n,i,p\right)T_{out}^L - F_{in}^U\left(i\right)T_{out}^L \qquad \forall n\in N, i\in I, p\in P \tag{3.82}$$

$$\gamma_1\left(n,i,p\right) \leq C_R\left(n,i,p\right)T_{out}^U\left(i\right) \qquad \forall n\in N, i\in I, p\in P \tag{3.83}$$

$$\gamma_1\left(n,i,p\right) \geq C_R\left(n,i,p\right)T_{out}^L \qquad \forall n\in N, i\in I, p\in P \tag{3.84}$$

$\gamma_2(i',i,p)$ is the linearisation variable for the term $F_r(i',i,p)\,T_{out}(i')$. Constraints 3.85–3.88 are used in defining this term.

$$\gamma_2(i',i,p) \geq F_{in}^U(i)\,T_{out}(i',p) + F_r(i',i,p)\,T_{out}^U(i') - F_{in}^U(i)\,T_{out}^U(i') \qquad \forall i,i' \in I, p \in P \tag{3.85}$$

$$\gamma_2(i',i,p) \leq F_{in}^U(i)\,T_{out}(i',p) + F_r(i',i,p)\,T_{out}^L - F_{in}^U(i)\,T_{out}^L \qquad \forall i,i' \in I, p \in P \tag{3.86}$$

$$\gamma_2(i',i,p) \leq F_r(i',i,p)\,T_{out}^U(i') \qquad \forall i,i' \in I, p \in P \tag{3.87}$$

$$\gamma_2(i',i,p) \geq F_r(i',i,p)\,T_{out}^L \qquad \forall i,i' \in I, p \in P \tag{3.88}$$

$\gamma_3(i,p)$ is the linearisation variable for the term $F_{in}(i,p)\,T_{out}(i)$. Constraints 3.89–3.92 are used in defining this term.

$$\gamma_3(i,p) \geq F_{in}^U(i)\,T_{out}(i,p) + F_{in}(i,p)\,T_{out}^U(i) - F_{in}^U(i)\,T_{out}^U(i) \qquad \forall i \in I, p \in P \tag{3.89}$$

$$\gamma_3(i,p) \geq F_{in}^U(i)\,T_{out}(i,p) + F_{in}(i,p)\,T_{out}^L - F_{in}^U(i)\,T_{out}^L \qquad \forall i \in I, p \in P \tag{3.90}$$

$$\gamma_3(i,p) \leq F_{in}(i,p)\,T_{out}^U(i) \qquad \forall i \in I, p \in P \tag{3.91}$$

$$\gamma_3(i,p) \geq F_{in}(i,p)\,T_{out}^L \qquad \forall i \in I, p \in P \tag{3.92}$$

$\gamma_4(n,i,p)$ is the linearisation variable for the term $C_S(n,i,p)\,T_{sup}(n,p)$. Constraints 3.93–3.96 are used in defining this term.

$$\gamma_4(n,i,p) \geq F_{in}^U(i)\,T_{sup}(n,p) + C_S(n,i,p)\,T_{amb} - F_{in}^U(i)\,T_{amb} \qquad \forall n \in N, i \in I, p \in P \tag{3.93}$$

$$\gamma_4(n,i,p) \leq F_{in}^U(i)\,T_{sup}(n,p) + C_S(n,i,p)\,T_{ct}^L - F_{in}^U(i)\,T_{ct}^L \qquad \forall n \in N, i \in I, p \in P \quad (3.94)$$

$$\gamma_4(n,i,p) \leq C_S(n,i,p)\,T_{amb} \qquad \forall n \in N, i \in I, p \in P \quad (3.95)$$

$$\gamma_4(n,i,p) \geq C_S(n,i,p)\,T_{ct}^L \qquad \forall n \in N, i \in I, p \in P \quad (3.96)$$

where T_{ct}^L is defined by Equation 3.97.

$$T_{ct}^L = \min_{n \in N} T_{ct}(n) \quad (3.97)$$

$\gamma_5(n,i,p)$ is the linearisation variable for the term $C_R(n,i,p)\,T_{ret}(n,p)$. Constraints 3.98–3.101 are used in defining this term.

$$\gamma_5(n,i,p) \geq F_{in}^U(i)\,T_{ret}(n,p) + C_R(n,i,p)\,T_{ret}^U(n) - F_{in}^U(i)\,T_{ret}^U(n) \qquad \forall n \in N, i \in I, p \in P \quad (3.98)$$

$$\gamma_5(n,i,p) \leq F_{in}^U(i)\,T_{ret}(n,p) + C_R(n,i,p)\,T_{ct}^L - F_{in}^U(i)\,T_{ct}^L \qquad \forall n \in N, i \in I, p \in P \quad (3.99)$$

$$\gamma_5(n,i,p) \leq C_R(n,i,p)\,T_{ret}^U(n) \qquad \forall n \in N, i \in I, p \in P \quad (3.100)$$

$$\gamma_5(n,i,p) \geq C_R(n,i,p)\,T_{ct}^L \qquad \forall n \in N, i \in I, p \in P \quad (3.101)$$

$\gamma_6(n',n,p)$ is the linearisation variable for the term $R(n',n,p)\,T_{sup}(n,p)$. Constraints 3.102–3.105 are used in defining this term.

$$\gamma_6(n',n,p) \geq OS^U(n)\,T_{sup}(n,p) + R(n',n,p)\,T_{amb} - OS^U(n)\,T_{amb} \qquad \forall n,n' \in N, p \in P \quad (3.102)$$

$$\gamma_6\left(n', n, p\right) \leq OS^U\left(n\right) T_{sup}\left(n, p\right) + R\left(n', n, p\right) T_{ct}^L - OS^U\left(n\right) T_{ct}^L$$
$$\forall n, n' \in N, p \in P \quad (3.103)$$

$$\gamma_6\left(n', n, p\right) \leq R\left(n', n, p\right) T_{amb} \qquad \forall n, n' \in N, p \in P \qquad (3.104)$$

$$\gamma_6\left(n', n, p\right) \geq R\left(n', n, p\right) T_{ct}^L \qquad \forall n, n' \in N, p \in P \qquad (3.105)$$

$\gamma_7\left(n', n, p\right)$ is the linearisation variable for the term $R\left(n', n, p\right) T_{ret}\left(n, p\right)$. Constraints 3.106–3.109 are used in defining this term.

$$\gamma_7\left(n', n, p\right) \geq OS^U\left(n\right) T_{ret}\left(n, p\right) + R\left(n', n, p\right) T_{ret}^U\left(n\right) - OS^U\left(n\right) T_{ret}^U\left(n\right)$$
$$\forall n, n' \in N, p \in P \quad (3.106)$$

$$\gamma_7\left(n', n, p\right) \leq OS^U\left(n\right) T_{ret}\left(n, p\right) + R\left(n', n, p\right) T_{ct}^L - OS^U\left(n\right) T_{ct}^L$$
$$\forall n, n' \in N, p \in P \quad (3.107)$$

$$\gamma_7\left(n', n, p\right) \leq R\left(n', n, p\right) T_{ret}^U\left(n\right) \qquad \forall n, n' \in N, p \in P \qquad (3.108)$$

$$\gamma_7\left(n', n, p\right) \geq R\left(n', n, p\right) T_{ct}^L\left(n\right) \qquad \forall n, n' \in N, p \in P \qquad (3.109)$$

3.3 Costing Model

The costing constraints are used for the financial analysis of the proposed design. The objective is the maximisation of plant profit, accounting for raw material, operational and capital costs. Constraint 3.110 describes the total raw material cost as the product of the amount of raw material used and the unit cost price of the relevant state.

$$c_{TR} = \sum_{i \in I} \sum_{s \in S} \sum_{p \in P} c_{RM}\left(s\right) B_I\left(i, s, p\right) \qquad (3.110)$$

The operational cost incurred during time period p is demonstrated by Constraint 3.111. This empirical constraint was implemented by Panjeshahi et al. (2009) in calculating the total annualised operational costs. The equation has been adjusted to take makeup water, chemical treatment and blowdown treatment costs into account. Costs associated with the pumps and fan have been ignored.

$$c_{PO}\left(p\right) = \sum_{n \in N} \left(110OS\left(n,p\right) + 2275.132M\left(n,p\right) + 1138B\left(n,p\right)\right) \qquad \forall p \in P \tag{3.111}$$

The variable part of the capital costs is determined for each cooling tower n during time period p in Constraint 3.112. This is an empirical equation obtained from Panjeshahi et al. (2009). The equation calculates an annualised capital cost based on variable cooling tower factors. The constraint has been adjusted to be applicable to batch process operations.

$$\begin{aligned} c_{VC_p}\left(n,p\right) = \frac{H}{H_Y} \cdot & \left(746.749OS\left(n,p\right)^{0.79} \mathit{Range}\left(n,p\right)^{0.57} \mathit{Approach}\left(n,p\right)^{-0.9924} \right. \\ & \left. + \left(0.022T_{wb} + 0.39\right)^{2.447}\right) \qquad \forall n \in N, p \in P \end{aligned} \tag{3.112}$$

Constraints 3.113 and 3.114 are used together in determining the variable portion of the capital costs for each cooling tower n. Since the variable costs fluctuate due to operational changes between time points, the cooling tower should be designed to accommodate the peak demand.

$$c_{VC}\left(n\right) \geq c_{VC_p}\left(n,p\right) \qquad \forall n \in N, p \in P \tag{3.113}$$

$$c_{VC}\left(n\right) \leq c_{VC_p}\left(n,p\right) + BM\left(1 - y_{VC}\left(n,p\right)\right) \qquad \forall n \in N, p \in P, p \neq |P| \tag{3.114}$$

Where,

$$\sum_{p \in P} y_{VC}\left(n,p\right) = 1 \qquad \forall n \in N \tag{3.115}$$

The total capital cost can then be determined as seen in Constraint 3.116. The total capital costs are made up of the sum of the fixed costs associated with the existence of a cooling tower and the variable costs associated with the size of the cooling tower.

$$c_{TC} = \frac{H}{H_Y} \cdot c_{FC} \cdot CT + \sum_{n \in N} c_{VC}(n) \tag{3.116}$$

Constraint 3.117 is used in determining the total operational costs as the sum of the operational costs incurred during each time period p. The equation ensures that the total operational costs are applicable over the time horizon of interest.

$$c_{TO} = \frac{1}{H_Y} \sum_{p \in P} c_{PO} \left(t(p) - t(p-1)\right) \tag{3.117}$$

The profit associated with the production system only is detailed by Constraint 3.118. The profit is defined as the difference between the revenue due to product sales and the raw material cost.

$$\max Profit_P = \sum_{s \in S} \sum_{p \in P} \zeta(s) \, SS(s,p) - c_{TR} \tag{3.118}$$

Constraint 3.119 describes the overall plant profit taking into account raw material, capital and operational costs.

$$\max Profit = \sum_{s \in S} \sum_{p \in P} \zeta(s) \, SS(s,p) - c_{TR} - c_{TC} - c_{TO} \tag{3.119}$$

3.4 References

Kemmer, F. N. (1978) *The Nalco water handbook*. McGraw-Hill.

Kröger, D. G. (2004) *Air-cooled heat exchangers and cooling towers*. Vol. 1. PennWell Books.

Majozi, T. and Moodley, A. (2008) 'Simultaneous targeting and design for cooling water systems with multiple cooling water supplies'. *Computers & Chemical Engineering*, 32 (3), pp. 540–551.

Maravelias, C. T. and Grossmann, I. E. (2003) 'New general continuous-time state-task network formulation for short-term scheduling of multipurpose batch plants'. *Industrial & Engineering Chemistry Research*, 42 (13), pp. 3056–3074.

Panjeshahi, M., Ataei, A., Gharaie, M. and Parand, R. (2009) 'Optimum design of cooling water systems for energy and water conservation'. *Chemical Engineering Research and Design*, 87 (2), pp. 200–209.

Perry, R. H. and Green, D. (1997) *Perry's chemical engineers' handbook*. 7th ed. McGraw-Hill.

Quesada, I. and Grossmann, I. E. (1995) 'Global optimization of bilinear process networks with multicomponent flows'. *Computers & Chemical Engineering*, 19 (12), pp. 1219–1242.

Chapter 4

Illustrative Examples

The applicability of the developed mathematical formulation is demonstrated by two illustrative examples. The first of which is based on a fixed schedule. The second example applies the model to a more general variable schedule operation. The example is solved comparing a sequential optimisation approach with an integrated optimisation approach.

4.1 Example 1: Fixed Schedule

This example has been adapted from Majozi and Moodley (2008) by converting the cooling water using operations from continuous to batch processes. The pre-determined production schedule is depicted by Figure 4.1.

The operation is assumed to occur over a 12 hour time horizon. The objective of this problem is to minimise the recirculating cooling water flowrate by exploring reuse opportunities between the batch cooling water using operations. A maximum of three cooling towers supplying six batch operations with cooling water is assumed. The cooling tower data specifying the cooling tower outlet temperature, maximum design capacity, and maximum cooling water return temperature is provided in Table 4.1.

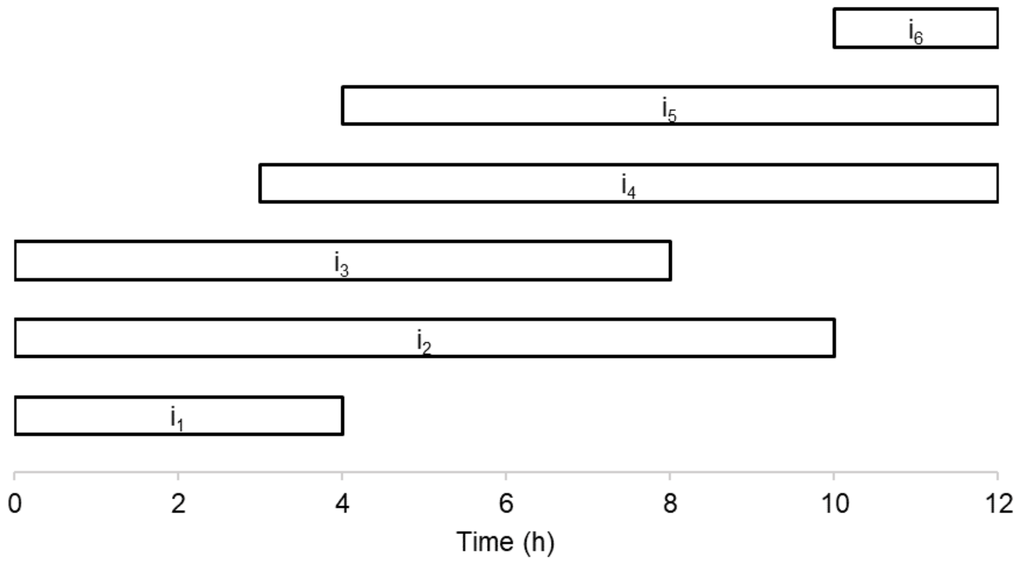


Figure 4.1: Production schedule for example 1.

Table 4.1: Cooling tower data for example 1

Cooling Tower	T_{ret} (℃)	$OS^U(t/h)$	T_{ret}^U (℃)
n_1	20.0	30.0	52.0
n_2	20.0	30.0	52.0
n_3	20.0	30.0	52.0

The cooling water using operation data is specified in Table 4.2. This provides the heat duty, flowrate and limiting inlet and outlet temperatures.

Table 4.2: Cooling water using operation data for example 1

Heat Exchanger	T_{in}^{U} (℃)	T_{out}^{U} (℃)	$CP(kW/$℃$)$	$Q(kW)$
i_1	30.0	45.0	40.7	610
i_2	40.0	60.0	21.0	420
i_3	25.0	50.0	32.0	800
i_4	45.0	53.0	69.4	555
i_5	40.0	55.0	23.0	345
i_6	30.0	45.0	46.7	700

Since the production schedule has been fixed for this particular examples. The batch scheduling constraints can be omitted from the model developed in Chapter 3. As such, the model used to obtain the below results involve only Constraints 3.36 onwards. It is worth noting the effect of the production schedule on the results which follow. In particular, it should be noted that the abrupt changes in the cooling tower return temperature coincides with the time point at which the cooling water system configuration is altered due to a particular operation switching on/off.

The results from this illustrative example have been summarised in Table 4.3. A comparison is made between the base case, where cooling water reuse between batch operations is not permitted, and the optimal case, where cooling water reuse between batch operations is permitted. Two cooling towers are required to meet the total cooling demand. The optimal configuration suggests that cooling towers n_1 and n_2 should be used to supply cooling water for both cases. The recirculating cooling water flowrate reduced by 5.2 % from 67.0 t/h to 63.5 t/h by allowing reuse between cooling water using operations. The total makeup water requirement over the 12 hour time horizon reduced slightly from 40.9 tonnes to 40.8 tonnes. The average return temperature was also increased by 3.2 %.

Table 4.3: Computational results for example 1

Scenario	Cooling Towers	Flowrate (t/h)	Makeup Water (t)
Base Case	2	67.0	40.9
Optimal Case	2	63.5	40.8

Figure 4.2 illustrates the average cooling water return temperature throughout thetime horizon. The return temperature for the base case scenario is compared with the optimal case.

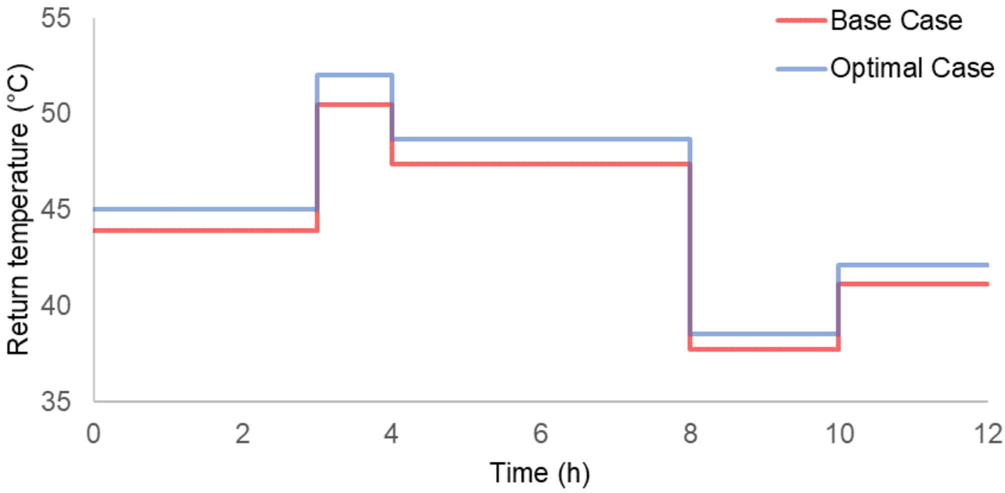


Figure 4.2: Variation of average cooling water return temperature with time for example 1.

It is worth noting that the average cooling water return temperature is greater when cooling water reuse is permitted. These conditions have been shown to result in increased cooling tower effectiveness, as it provides a greater temperature driving force and increases the contact time between the returned cooling water and the air entering the cooling tower (Bernier, 1994).

The model statistics for both the linear and nonlinear subprolems are displayed in Table 4.4. The optimisation problem was formulated in the GAMS 24.8.5 software platform, making use of the DICOPT solver which is discussed in Section 2.2. CPLEX was used to solve the MILP subproblems and CONOPT to solve the NLP subproblems. The model was executed on a computer with a 2.80 GHz Intel Core i7 processor with 8 GB of RAM.

Table 4.4: Model statistics for example 1

Model	Continuous Variables	Binary Variables	Constraints	CPU time (s)
Linear Model	1021	0	2465	0.094
Nonlinear Model	581	3	717	0.094

4.2 Example 2: Variable Schedule

The illustrative example is adapted from a cooling water network problem by Majozi and Moodley (2008) and a scheduling problem by Maravelias and Grossmann (2003). Three scenarios are considered: the base case, sequential approach and integrated approach.

The batch schedule for the base case is determined by optimising the profit due to the production system only, with no consideration given to the utilities. Furthermore, each task is supplied by a specified cooling tower and reuse of cooling water is not permitted between batch operations.

The sequential approach follows a two-step optimisation procedure. In the first step the production schedule is optimised without considering utilities. The schedule is then fixed and the overall system is then optimised, allowing for reuse of cooling water between batch operations and cooling towers supply to any batch operation.

Finally, the integrated approach simultaneously optimises the production and utility systems. Thus, the batch schedule is determined as part of the optimisation while considering utilities. Reuse of cooling water is permitted between batch operations and cooling towers are allowed to supply any batch operation.

Figure 4.3 illustrates the process recipe by means of the STN representation. Three reactors are available for the four reaction processes to take place, by sharing of equipment units. The two raw material and three subsequent intermediate states are converted into two final products.

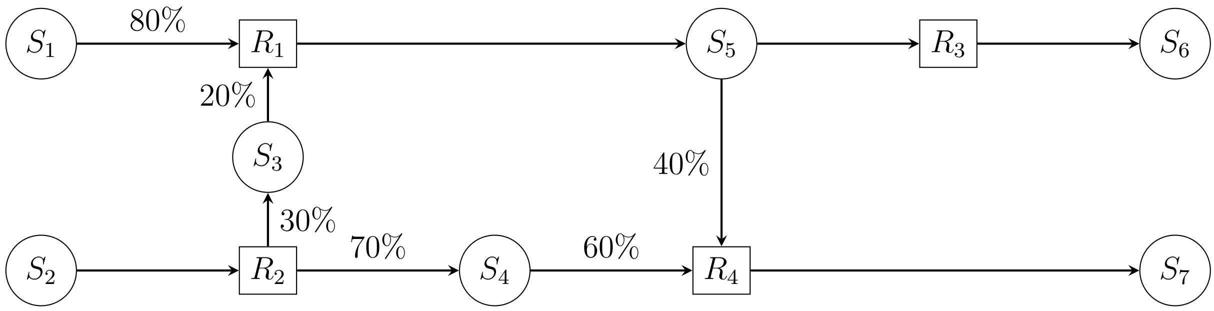


Figure 4.3: Flowsheet for example 2

The state and task related data required for the scheduling problem are available in Tables 4.5 and 4.6.

Table 4.5: State related scheduling data for example 2

State	$S_A^U\,(kg)$	$S_A^0\,(kg)$	$c_{RM}\,(\$/kg)$	$\zeta\,(\$/kg)$
S_1	1000	400	10	0
S_2	1000	400	15	0
S_3	200	0	25	0
S_4	100	0	0	0
S_5	500	0	0	0
S_6	1000	0	0	30
S_7	1000	0	0	40

The data pertaining to the cooling tower temperature bounds and maximum cooling tower design capacities are represented in Table 4.7.

Table 4.6: Task related scheduling data for example 2

Task	$B^L\,(kg)$	$B^U\,(kg)$	$\alpha\,(h)$	$\beta\,(h/kg)$
i_1	40	80	0.5	0.025
i_2	25	50	0.5	0.04
i_3	40	80	0.75	0.0375
i_4	25	50	0.75	0.06
i_5	40	80	0.25	0.0125
i_6	40	80	0.5	0.025

Table 4.7: Cooling tower data for example 2

Cooling Tower	$T_{ct}\,(°C)$	$T_{ret}^U\,(°C)$	$OS^U(t/h)$
n_1	20.0	52.0	30.0
n_2	22.0	52.0	40.0
n_3	25.0	50.0	40.0

Table 4.8 presents data relevant to the cooling water using operations. This includes temperature bounds and heat of reaction data.

Table 4.8: Cooling water using operation data for example 2

Heat Exchanger	$T_{in}^U\,(°C)$	$T_{out}^U\,(°C)$	$Q(kW/t)$
i_1	30.0	45.0	13
i_2	30.0	45.0	13
i_3	40.0	55.0	18
i_4	40.0	55.0	18
i_5	25.0	50.0	6
i_6	45.0	60.0	15

The cycles of concentration, CC, was kept constant at 4. Yearly operating hours, H_Y, is assumed to be $8000h$. Finally, ambient temperature, T_{amb}, and wet bulb temperature, T_{wb}, are specified as $25°C$ and $17°C$, respectively.

The Gantt chart obtained for the base case and sequential optimisation approach is shown in Figure 4.4. This is the optimal production schedule when one considers only the costs associated with the production system *i.e.* product sales and raw material costs, ignoring utility related costs.

The Gantt chart associated with the integrated optimisation approach is depicted in Figure 4.5. It is worth noting that there are no instances in this production schedule where three tasks are active simultaneously, as observed in the base case. This reduces peak demand of cooling water during a single time period.

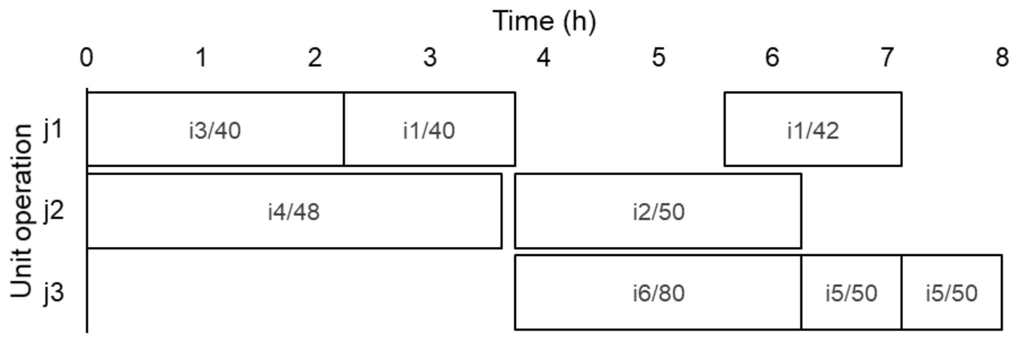


Figure 4.4: Gantt chart for base case and sequential optimisation

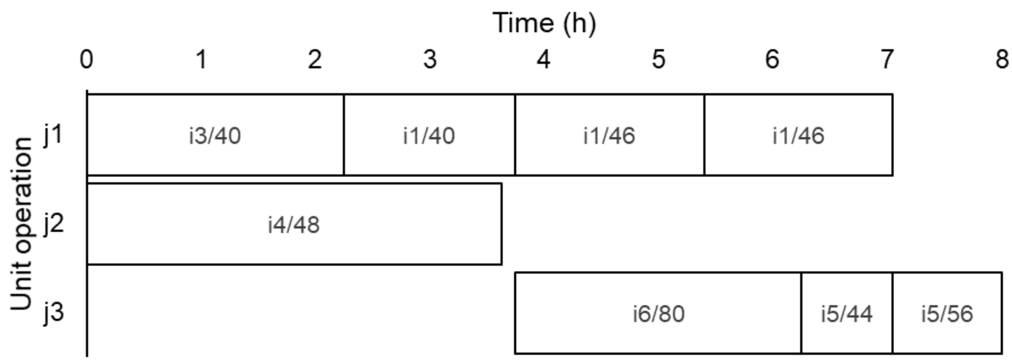


Figure 4.5: Gantt chart for simultaneous optimisation

The process flow diagrams depicting the cooling water system throughout the time horizon are shown in Figures 4.6–4.8. These are associated with the base case, sequential optimisation and integrated optimisation approaches respectively.

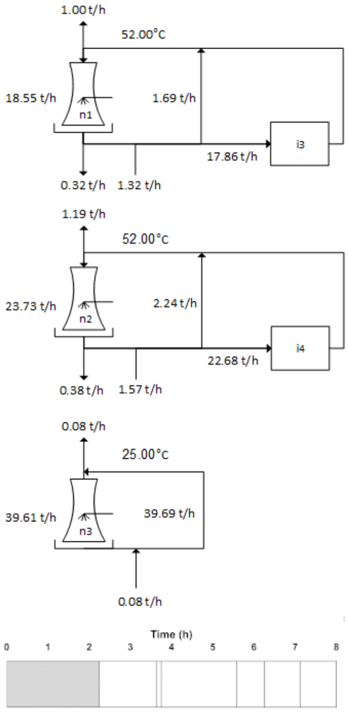
The computational results obtained for the illustrative example are summarised in Table 4.9. The sequential and simultaneous optimisation approach both led to significant improvements over the traditionally implemented base case scenario. However, it is worth mentioning that the simultaneous optimisation approach leads to further improvements over the sequential optimisation approach. This is due to the freedom of a larger amount of variables which produces a larger search space.

Table 4.9: Computational results for example 2

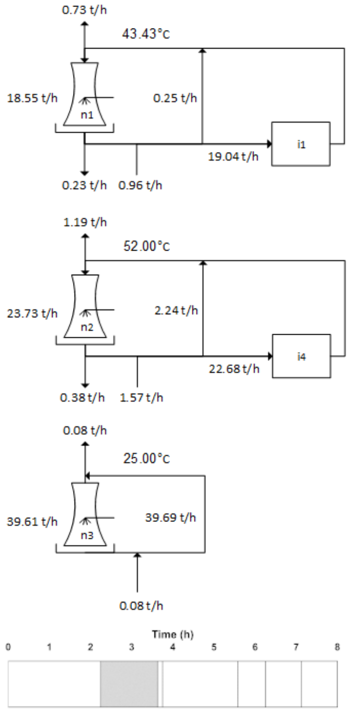
	Base	Sequential	Simultaneous
Profit ($)	3774.93	3785.82	3792.66
Cooling Towers	3	2	2
Cooling Water (t/h)	81.9	63.8	47.5
Makeup Water (t)	21.7	21.2	20.3
Return Temperature ($°C$)	37.6	39.5	44.5
Effectiveness (%)	42.4	52.8	67.7

Information relating to the model statistics for each optimisation approach is summarised in Table 4.10. The optimisation problem was formulated in the GAMS 24.8.5 software platform, making use of the BARON branch and bound solver which is discussed in Section 2.2. The model was executed on a computer with a 2.80 GHz Intel Core i7 processor with 8 GB of RAM.

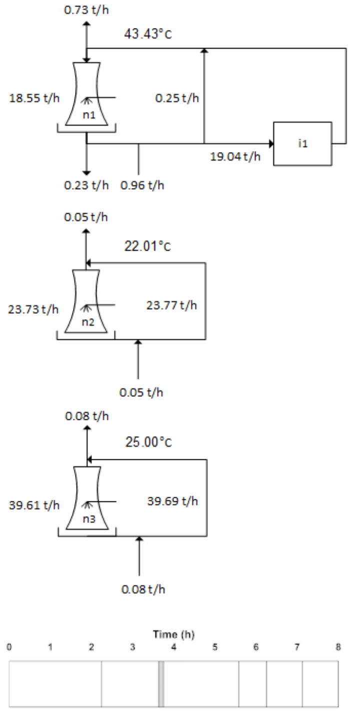
The overall cooling tower effectiveness for each scenario can be visually depicted throughout the time horizon, as observed from Figure 4.9. It is worth noticing that the simultaneous optimisation approach leads to an improved overall cooling tower effectiveness throughout the time horizon. The overall



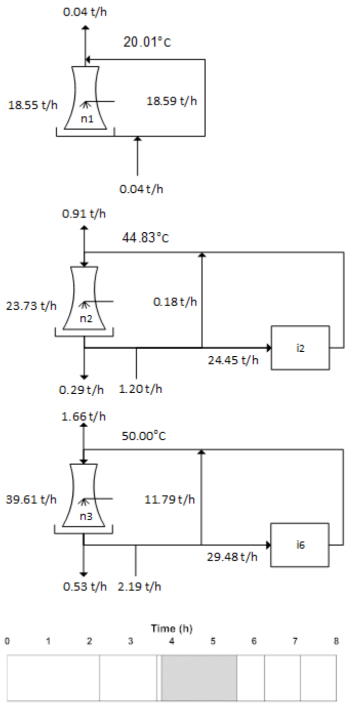
(a) Time point p_1



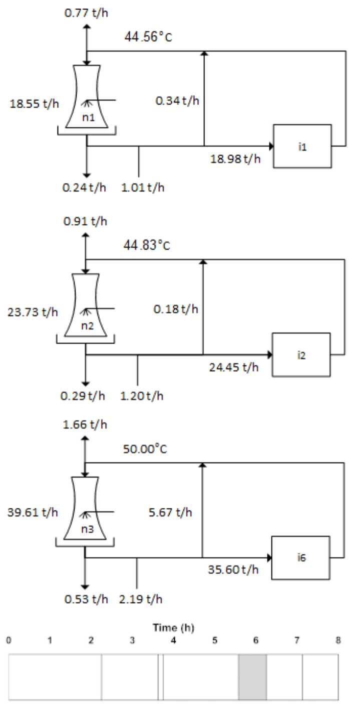
(b) Time point p_2



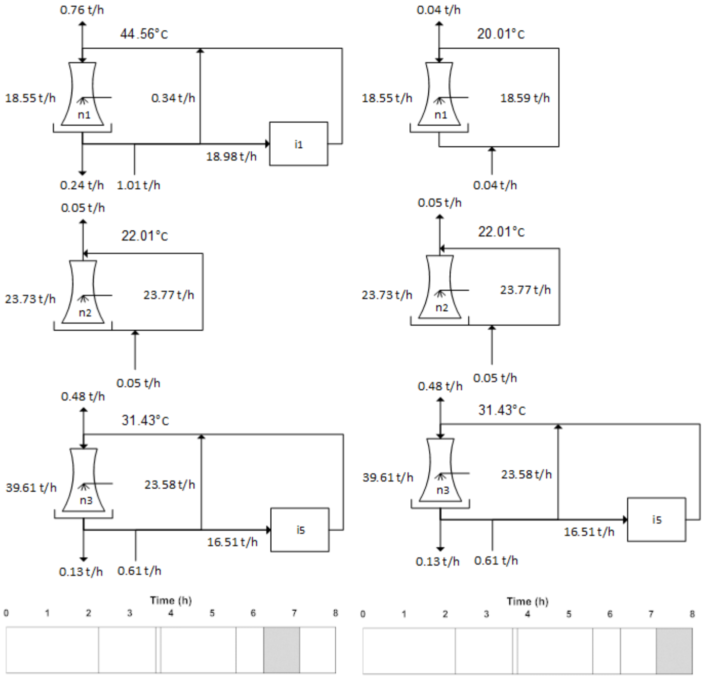
(c) Time point p_3



(d) Time point p_4

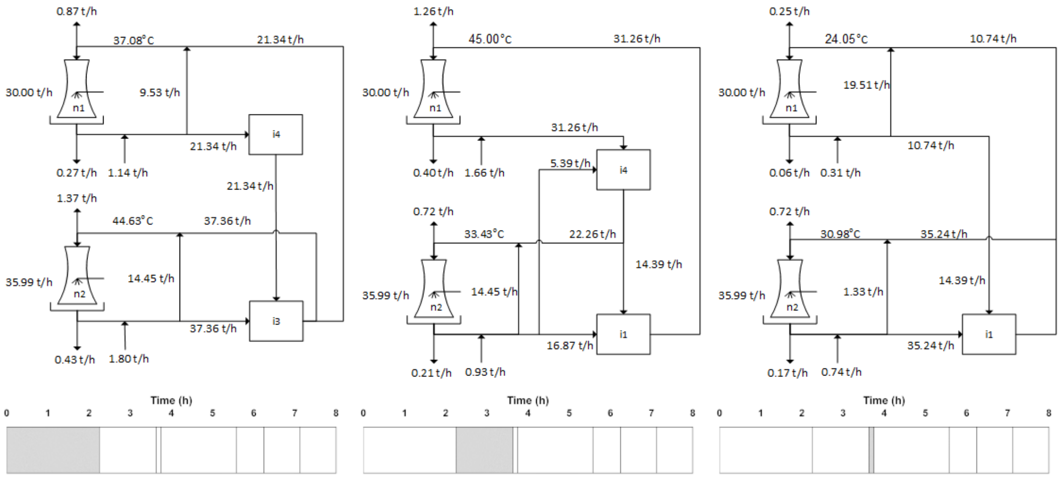


(e) Time point p_5

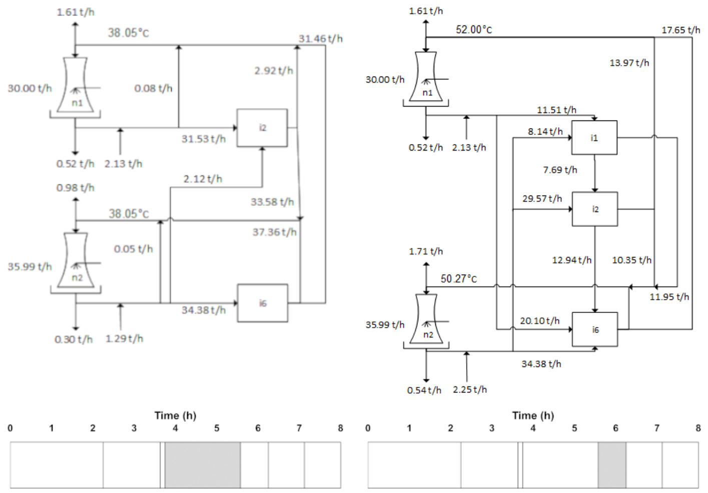


(f) Time point p_6 (g) Time point p_7

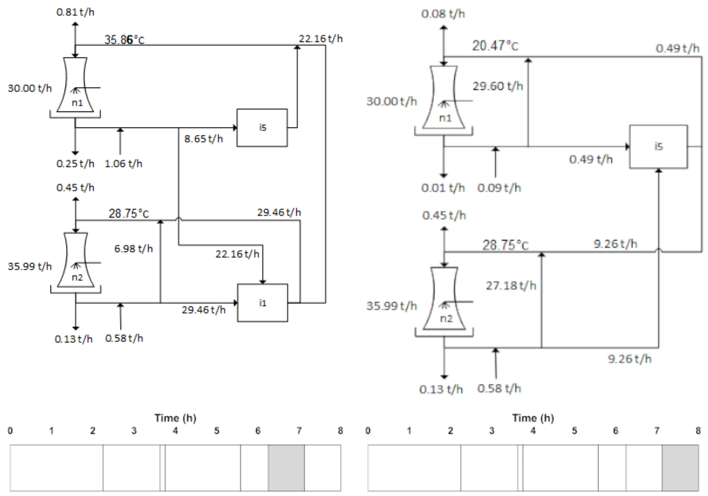
Figure 4.6: Process flow diagram for base case throughout the time horizon



(a) Time point p_1 (b) Time point p_2 (c) Time point p_3

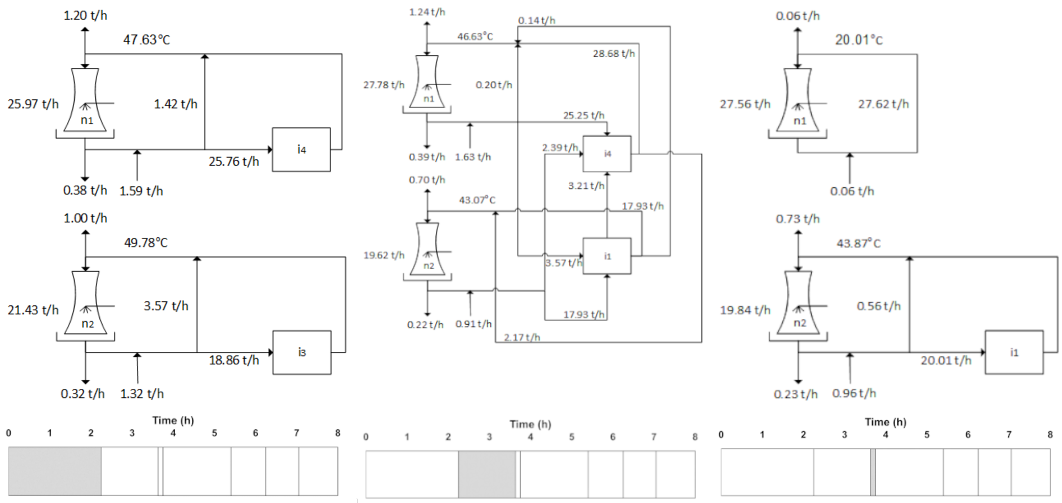


(d) Time point p_4 (e) Time point p_5

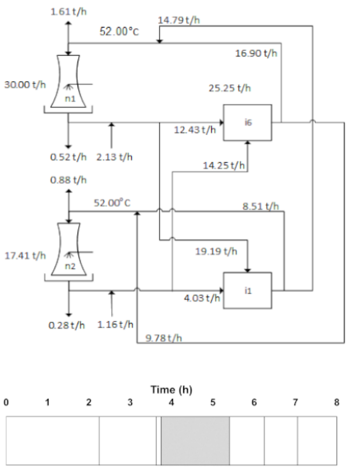


(f) Time point p_6 (g) Time point p_7

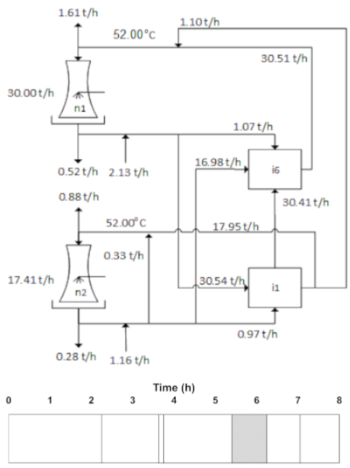
Figure 4.7: Process flow diagram for sequential case throughout the time horizon



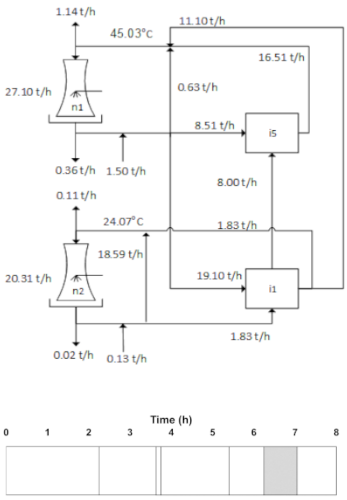
(a) Time point p_1 (b) Time point p_2 (c) Time point p_3



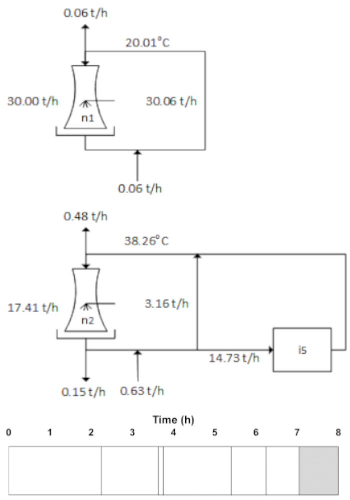
(d) Time point p_4



(e) Time point p_5



(f) Time point p_6



(g) Time point p_7

Figure 4.8: Process flow diagram for integrated case throughout the time horizon

Table 4.10: Model statistics for example 2

	Base	Sequential	Simultaneous
Time points	8	8	8
Continuous variables	2120	2120	2379
Binary variables	360	264	405
Constraints	3825	3825	4274
CPU time (s)	10000	10000	10000
Integrality gap (%)	0.15	0.07	0.50

cooling tower effectiveness for batch processes tends to be comparably lower than continuous operations. This is likely due to the discrete nature of batch processing, whereby there are periods that the cooling towers are operating significantly below their capacity.

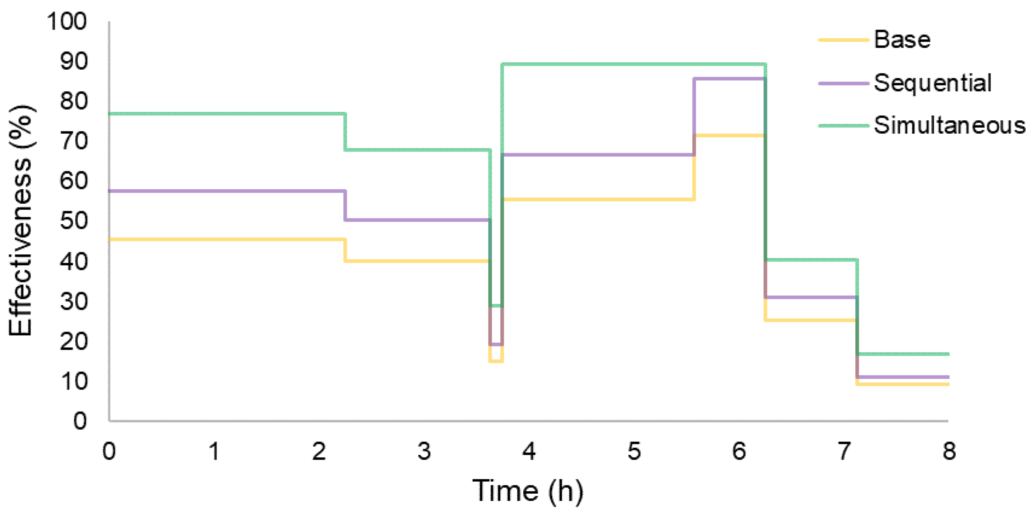


Figure 4.9: Overall cooling tower effectiveness over time horizon of interest

4.3 References

Bernier, M. A. (1994) *Cooling tower performance: theory and experiments.* American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., Atlanta, GA (United States).

Majozi, T. and Moodley, A. (2008) 'Simultaneous targeting and design for cooling water systems with multiple cooling water supplies'. *Computers & Chemical Engineering*, 32 (3), pp. 540–551.

Maravelias, C. T. and Grossmann, I. E. (2003) 'New general continuous-time state-task network formulation for short-term scheduling of multipurpose batch plants'. *Industrial & Engineering Chemistry Research*, 42 (13), pp. 3056–3074.

Chapter 5

Conclusion

The synthesis and optimisation of a batch production system accompanied by a cooling water system has been explored by means of a mathematical formulation in this dissertation. Utility systems integration with production processes have historically considered continuous processes exclusively. This is particularly apparent when looking at cold utilities. The cooling water system is made up of a set of cooling towers operating at different supply temperatures. Two illustrative examples are used to demonstrate the applicability of the mathematical model to problems encountered in industry.

The first example assumes a fixed batch production schedule. Operational improvements are explored through varying cooling tower supply to batch processes and reuse of cooling water amongst tasks. Recirculating cooling water was reduced by 5.2 % and average return temperature to the cooling towers, which is concomitant with cooling tower effectiveness, increased by 3.2 %.

The second illustrative example introduced further complexity by exploring a variable production schedule. Three scenarios were compared. The first scenario involves a traditional base case, where the objective function involves maximisation of revenue due to sale of product only. The second scenario involves a sequential optimisation procedure where the production system is first optimised independent of the utility network, followed by the optimisation of the utility network on the fixed production schedule. Finally, an integrated optimisation approach is considered where the production and utility systems are optimised simultaneously. This provides a comparison between a sequential and integrated optimisation approach.As an output from the optimisation model, the optimal production schedule and operating conditions of the batch network are determined.

The overall model is in the form of an MINLP problem. A linearised model based on the reformulation-linearisation approach is used to initialise the model. In validating and demonstrating the applicability of the developed model, two illustrative models are presented. It was found that the number of cooling towers could be reduced from 3 to 2 leading to a 0.45% increase in the annual profit. Furthermore, the results indicate a 42.1% reduction in recirculating cooling water, a 25.3% increase in the average overall cooling tower effectiveness, and a 6.3% reduction in makeup water required throughout the time horizon when comparing the integrated approach to the base case scenario. It is worth mentioning that the integrated approach outperformed the sequential optimisation approach in all of the above mentioned metrics.

Further opportunities in the integrated design of utility systems and batch production space include the detailed design of the cooling towers, the inclusion of a pumping network to determine the effect of cooling water reuse, inclusion of hot utilities in the design in the form of a steam system, and the integrated design of a batch production facility including process-process heat integration with the utilities system.