Bridging the Digital Gap in Continuous Pharmaceutical Manufacturing:

From Complex Models to Digital Twins

Donald Ntamo

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

School of Chemical, Materials and Biological Engineering

The University of Sheffield

March 2025

Declaration

I, Donald Ntamo, confirm that the thesis is my own work. I am aware of the University's guidance on the Use of Unfair Means (www.sheffield.ac.uk/ssid/unfair-means). This work has not been previously presented for an award at this, or any other, university.

Donald Ntamo

Date: 26/03/2025

Publications And Conferences

Publications

Throughout my PhD research, I have actively engaged in disseminating my findings through journals like digital chemical engineering and MIDP on topics leveraging Industry 4.0 techniques like machine learning, cloud techniques, big data and data analysis, and digital sensing for digital twin applications such as visual design of experiments, soft sensing and sustainability optimisation.

1. Ntamo, D., Montero, E., Omar, C., Highett, M., Moss, D., Soulatintork, P., Moghadam, P., Zandi, M. and Mitchell, N., "Industry 4.0 in action: Digitalisation of a continuous process manufacturing for formulated products," Digital Chemical Engineering, vol. 3, p.100 025, Jun. 2022. DOI: 10.1016/j.dche.2022.100025.

This paper aims to share the latest development of an advanced digital twin of a continuous wet granulation and tableting process at The University of Sheffield. These include the delivery of a digital platform consisting of an Advanced Process Control system (APC), mechanistic model platform and industrial IoT platform for data analytics and visualisation. The combined solution aligns with the concepts of Industry 4.0 by providing a digital twin, cloud integration, sophisticated statistical, as well as hybrid and mechanistic models. The models are in turn, used for soft-sensors, Model Predictive Control and Optimisation algorithms to predict and control product Quality Attributes. Through an examination of existing digital twin architectures, this paper identifies a high-fidelity digital twin gap and subsequently develops a new architecture, presented in Chapter 7, to address this deficiency.

2. Ntamo, D.; Papadopoulos, I.; Omar, C.; Soulatiantork, P.; Zandi, M. A Sustainability-Oriented Digital Twin of the Diamond Pilot Plant. Processes 2025, 13, 211. https://doi.org/10.3390/pr13010211.

This paper focuses on developing a data-driven model to predict energy consumption in a twin-screw granulator (TSG) within the DiPP. The model, based on second-degree polynomial regression, demonstrates strong predictive accuracy with R-squared values exceeding 0.8. By optimising energy performance indicators, this work aims to improve the sustainability of pharmaceutical manufacturing processes. This research, detailed in Chapter 6,

advances the development of comprehensive pharmaceutical manufacturing Digital Twins by laying the groundwork for integrating energy models.

Conference presentations

I have also actively shared these insights at prominent conferences listed below. These presentations have been instrumental in solidifying my findings through interactions with pharmaceutical manufacturing industry companies such Siemens and Process Engineering Systems.

1.Ntamo D., Moghadam, P., Soulatiantork P., Zandi M., (2021) Developing an advanced digital twin for continuous pharmaceutical manufacturing, IChemE, Advances in the Digitalisation of the Process Industries 2021, (Online).

2.Ntamo D. and Mo Zandi, M., (2022) Developing a maintenance free advanced digital twin for continuous pharmaceutical manufacturing, Advanced process modelling forum 2022; Digital design and operations: towards sustainability in the process industry, England, London.

3.Ntamo D. and Zandi M., (2023) Using global system analysis to create fast solving models from rigorous models, Chem Eng Day 2023, The University of Belfast, Northern Ireland, Belfast.

4.Ntamo D., Papadopoulos I., Soulatiantork P., Zandi M., (2023) A sustainability-oriented Digital Twin of the Continuous Manufacturing Pharmaceutical Pilot Plant, Chem Eng Day 2024, Imperal College London, England, London.

Acknowledgements

The creation of this thesis would not have been possible without the support of numerous individuals. would like to express my sincere gratitude to my supervisors, Professor Mohammad Zandi and Dr Payam Soulatiantork, for their unwavering support, insightful guidance, and valuable contributions throughout my research journey. They helped me maintain a broad perspective on my research, ensuring its relevance to the modern-day manufacturing industry. The Government of Botswana's funding made this work possible. My colleagues at the University of Sheffield's Particulate Research Group provided invaluable support and camaraderie throughout my PhD journey. Additionally, I would like to thank Dr Chalak, Mr Martin Highett, and Dr Daniele Monaco for their technical assistance with the diamond plant's operations and control systems. Their expertise was essential to the success of my research. Thank you to the students who helped refine, discuss, and experimentally evaluate the work done during this project, including Iason Papadopoulos, Thomas Warnke, Vladislavs Kuliss and Max Wedekind.

I am grateful to Siemens, PSE, and Perspective Engineering for developing the software essential to creating an advanced digital twin. Their ongoing support, particularly from Dr Stefan Bellinghausen of PSE, who provided technical assistance with PSE's gPROMS formulated products software, was instrumental in the completion of this research. I also would like to acknowledge the contributions of my co-authors: N. Mitchell from Siemens Process Systems Engineering, J. C. Mack and E. Lopez-Montero from Perceptive Engineering Limited, and Omar, M. I. Highett, D. Moss, P. Soulatiantork, P. Z. Moghadam, I. Papadopoulos, and M. Zandi from the University of Sheffield, and. Their insightful reviews, constructive feedback, and expert analysis were invaluable in shaping the final publication.

Lastly, I would like to express my appreciation to my friends and family for their unwavering emotional support throughout this endeavour.

Nomenclature

Symbols	Description	Units
$a, a'\ b, b', c, c'$	Conveying and kneading element breakage kernel empirical parameters	-
A_p	Surface area of particles expose for drying.	m^2
B_{M1} and B_{M2}	Breakage functions	-
c_w	Specific heat capacity	$kJ.K^{-1}$
C_{LV}	Set of species in the vapour or liquid phase for vaporisation	-
DYS	Dynamic yield stress	kPa
d_{50}	Median particle diameter	μm
$EnPI$	Energy Performance Indicator	kWh^2
$F_p^{in}\ and\ F_p^{out}$	Inlet and outlet flow rate of phase p	$kg.s^{-1}$
$F_v^{\text{in}}\ and\ F_v^{\text{ out}}$	Inlet and outlet flow rate of vapour phase	kgs^{-1}
$G_i(v)$	Layering growth rate	$\mu m.s^{-1}$
$G_i(x)$	Layering linear growth rate	$\mu m.s^{-1}$
$\Delta H_{drying,i,p}$	Enthalpy change rate from drying	kJs^{-1}
$h_{i,p}^{in}$ and $h_{i,p}^{out}$	Specific enthalpy of phase p in the inlet and outlet stream.	$kJ.kg^{-1}$
$h_{i,v}^{\text{ in}}\ and\ h_{i,v}^{\text{ out}}$	Specific enthalpy of the vapour phase in the inlet and outlet stream	$kJ.kg^{-1}$
H_p	Total enthalpy holdup in phase p	kJ
j_i	Mass flux of species i	$kg.m^{-2}s^{-1}$

$k_{i,\,\text{comp}}$	Consolidation rate constant in compartment i	s^{-1}
k_b	The bonding capacity of the material	-
k_{cc}	Compressibility constant	-
$k_{c,i}$	Mass transfer coefficient for species i	ms^{-1}
l_c	Milling selection rate kernel critical size	mm
$\dot{l}_i$	Liquid mass flow rate into compartment i	$kg.h^{-1}$
l_{sc}	Milling screen size	mm
$\dot{m}_g$	Granule mass flow rate	$kg.h^{-1}$
$\dot{m}_p$	Powder mass flow rate	$kg.h^{-1}$
M_i	Total mass of fine powder, granules, and liquid in the compartment i	kg
$M_{i,p}$	Total mass holdup of species i in phase p	kg
$M_{i,v}$	Total mass of species in the vapour phase.	kg
$P_{b,i}(v)$	Breakage fragment distribution function volume v in compartment i	-
p	Applied pressure	kN
p_0	Theoretical pressure required to produce a zero-porosity compact.	kN
PFN	Powder feed number	-
$\dot{Q}_{pv}\ and\ \dot{Q}_{vw}$	Heat flux interaction with	$kJ.m^{-2}$

	particles, vapour and wall	
$R_{drying,i,p}$	Drying rate of phase p lost to the vapour phase.	$kJ.s^{-1}$
$R_i(\varepsilon)$	Consolidation rate in compartment i	s^{-1}
s^*	Droplet pore penetration	-
T_w	Wall temperature	K
V_{drop}	Single drop volume	m^3
$x_{\text{vap},i}$ and $x_{\text{vap},j}$	Mass fraction of species i and j in the vapour phase.	$kg.kg^{-1}$
$x_{i,p}^{in}$ and $x_{i,p}^{out}$	Mass fraction of phase p in the inlet and outlet stream.	$kg.kg^{-1}$
$x_{i,v}^{\text{in}} and\ x_{i,v}^{\ \text{out}}$	Mass fraction of vapour phase in the inlet and outlet stream.	$kg.kg^{-1}$
x_{lc}	Lower bound granule diameter	mm
x_{uc}	Upper bound granule diameter	mm
$y_{w,i}$	Moisture content in compartment i	$\%$
$Y_{eq,i}$ and $Y_{bulk,i}$	Equilibrium and bulk dry basis moisture content of species i in the gas phase	$kg.kg^{-1}$
α	Reduction in apparent screen size due to the tangential movement of the particles in the mill	-
β	Weibull function scale parameter	-
γ	Weibull function shape parameter	-
ε_t	Tablet porosity	-
$\eta_{flow,p}{}^{in(r)}$ and $\eta_{flow,p}{}^{out(r)}$	Rate number of particles of radius r entering and exiting the unit	s^{-1}
$\eta p(r)$	Number of particles of radius r	-

$\lambda \, and \, k$	Layering kernel empirical parameter	-
μ_{drop}	Drop size	mm
$\mu_{\text{drop, mean}}$	Mean droplet size	mm
ρ_g	Density of the gas phase.	$kg.m^{-3}$
σ_{drop}	Drop size standard deviation	mm
$\sigma^2_{i,T,y}$	Total variance attributed to factor I on response y through all factors	-
$\sigma^2_{i,y}$	Variance attributed to factor i on response y	-
σ_t	Tablet Tensile strength	MPa
σ_0	Tensile strength at zero porosity	MPa
τ_i	Dispersion time constant in compartment i	s
ϕ	Granule mass fraction	-

Abbreviations

ANN	Artificial Neural Network
APC	Advanced Process Control
API	Active Pharmaceutical Ingredient
APIs	Active Pharmaceutical Ingredients
AR	Augmented Reality
BM	Batch Manufacturing
CE	Conveying Element
CFD	Computational Fluid Dynamics
CM	Continuous manufacturing
CPPs	Critical Process Parameters
CQAs	Critical Quality Attributes
DAEBDF	Differential Algebraic Equation Backward Differentiation Formulae
DEM	Discrete Element Method
DiPP	Diamond Pilot Plant
DoE	Design of Experiments
DT	Digital twin
EMA	European Medicines Agency
EnPI	Energy Performance Indicator
EPA	Environmental Protection Agency
FBD	Fluidised Bed Drier
FDA	Food and Drug Administration
FEM	Finite Element Method
gFP	gPROMS FormulatedProducts
GSA	Global System Analysis
HMI	Human Machine Interface
I4.0	Industry 4.0
IoT	Internet of Things
IT	Information Technology
KE	Kneading Element
KPI	Key Point Indicators

MCR-ALS	Multivariate Curve Resolution Alternating Least Squares
MDD	Model Driven Design
MDM	Model-driven models
ML	Machine Learning
MPC	Model Predictive Control
MXLKHD	Maximum Likelihood
NIR	Near infrared
OT	Operational technology
PAT	Particle Analytical Tool
PBM	Population balance model
PCA	Principal Component Analysis
PDAEs	Partial Differential and Algebraic Equations
PLC	Programmable Logic Controller
PLS	Partial Least Squares
PSD	Particle Size Distribution
QbD	Quality by Design
RBFN	Radial Basis Function Network
RMSE	Root-Mean-Squared Error
SQP	Sequential Quadratic Programming
SVR	Support Vector Regression
TSWG	Twin Screw Wet Granulation
VR/AR	Virtual reality

Abstract

This research explores the gap between the conceptual concept of High-Fidelity Digital Twins (DTs) in pharmaceutical manufacturing and their practical implementation. Process modelling and simulation play a pivotal role within the high-fidelity DT framework, providing tools to analyse, and optimise complex processes. However, studies indicate that formulated products mechanistic models are often stand alone, impractical for digital application due to the complexity of calibration and maintenance. Previous utilised limited capability data driven models, creating a need for more efficient and adaptable modelling techniques. This thesis aims to bridge this gap by demonstrating the feasibility of implementing high-fidelity DT in wet granulation manufacturing line.

Novel aspects and research questions are addressed in this thesis by following objectives:

- Designing and implementing a comprehensive high-fidelity DT framework for Pharmaceutical Manufacturing Processes.
- Enhancing model development speed for high-fidelity Models.
- Leveraging ML to create fast solving models.
- Propose a faster path to model Maintenance.
- Designing a workflow for integrating high fidelity models into DT framework.
- Developing sustainability-oriented DT by integrating digital twin and energy management strategies.

By addressing these key points, this thesis contributes to the advancement of DT technologies in the pharmaceutical industry. Future directions of this research include novel digital solutions such as Digital Model Driven Control (DMDC) and virtual Design of Experiments (DOE).

Key contributions of this thesis include:

- The development of a robust and scalable high fidelity DT framework for continuous manufacturing processes, integrating process modelling and industry 4.0 technologies.

- It proposes a new approach that leverages surrogate modelling, hybrid modelling, and global system analysis (GSA). This approach enabled faster deployment of adaptable models that require minimal computational resources (40% calibration parameter reduction, simulation time less than 5 seconds).
- It showcases a successful application and integration of hybrid energy models (with testing R^2 values greater than 0.8) by optimising the Energy Performance Indicator (EnPI) using plant data.

Contents

List of figures

List of tables

Chapter 1: Introduction

1.1 Background

Rising demand for product quality, faster time-to-market, medication shortages, cheaper operational costs, personalised treatment, and easier regulatory compliance have all pushed the pharmaceutical sector to adopt Industry 4.0 technology (Stegemann, 2016; Sarkis et al., 2021). Rapid technological advancement is required to suit the changing expectations of consumers. While other businesses have adopted Industry 4.0 technology, the pharmaceutical industry has been hesitant to adopt (Litster and Bogle, 2019). Pharmaceutical processes, which are frequently batch-based, are less efficient than those in other chemical industries (Litster and Bogle, 2019; Lee et al., 2015). Many pharmaceutical operations are based on antiquated automation systems that lack IoT capabilities. Functions are often managed independently, with little interaction between unit functions for sophisticated process control. Furthermore, the complexities of solid product production make it difficult to establish well-defined unit operations for a wide range of goods (Litster and Bogle, 2019; Su et al., 2019).

To maintain competitiveness, industries must adopt innovative digital technologies to enhance productivity, foster innovation, and improve profitability. Major economies such as the US, China, and the EU are actively implementing smart manufacturing strategies to optimise pharmaceutical production processes (Litster and Bogle, 2019). As market demands evolve, the healthcare industry is compelled to consider the entire drug lifecycle, encompassing environmental impact, energy consumption, and supply chain sustainability (Pharmafile, 2020). While the WHO's GMP guidelines prioritise product quality, they overlook environmental considerations (Pharmafile, 2020). The pharmaceutical industry, particularly in small molecule drug production, is notoriously inefficient, with waste-to-product ratios reaching as high as 800:1 (Zhang and Wormr, 2019). Energy consumption, a significant contributor to both environmental impact and operational costs, accounts for a substantial portion of pharmaceutical plant expenses (Zhang and Wormr, 2019).

The escalating costs associated with developing and launching new drugs, particularly in research and development, have compelled the pharmaceutical industry to explore cost-effective strategies (Suresh and Basu, 2008). Continuous manufacturing emerges as a promising solution to offset these expenses (Lee et al., 2015). However, no industry is willing to uproot their existing manufacturing workflows to move to a new regime without being adequately convinced that such a transition is worthwhile. The inertia that exists in the current batch manufacturing regime is very large. It is mostly because pharmaceutical companies have gained regulatory approval for their products based on batch manufacturing techniques, making it a well-established and trusted approach. It is up to universities and research scientists to produce the necessary research to convince pharmaceutical companies that a transition to continuous manufacturing is worthwhile. Research indicates that the reduced costs associated with continuous manufacturing can incentivise both brand-name and generic drug manufacturers to invest in this technology (Schaber et al., 2011). Recent advancements in process analytical technologies and digital manufacturing have significantly enhanced our understanding of pharmaceutical processes and product quality (Teżyk et al., 2016). Given the intricate nature of formulations and processes, frequent in-process control measurements are essential to ensure product quality (Chen et al.,2020). Oral solid dosage forms, the most common drug delivery method, involve combining active pharmaceutical ingredients (APIs) with excipients through various processes guided by patient safety, manufacturability, end-user requirements and the therapeutic effect (Lopez et al., 2025).

1.2 Motivations

The COVID-19 pandemic has significantly disrupted global supply chains (GSCs) across all sectors (Xu et al., 2020). While functional products typically experience stable supply and demand, the pandemic has transformed some products, such medicine into innovative products with highly volatile demand and supply dynamics (Xu et al., 2020). Global governments-imposed lockdown measures to curb the spread of the virus had a severe impact on the pharmaceutical business. Supply chain disruptions, labour shortages, and economic losses compounded these challenges. As a result, boosting efficiency while preserving quality,

changing rapidly, and being adaptable in manufacturing is a difficult challenge for the pharmaceutical business.

The COVID-19 pandemic highlighted the critical need for flexible, demand-responsive manufacturing methods that minimise human intervention throughout the production process. According to Figure 1.1 ,It found that 47% of businesses plan to increase their investments in IoT, with a main reason behind the increase being that while companies have a limited history with IoT, IoT implementers produce a predictable return on investment (ROI) within a specified timeframe. I4.0, IoT, cloud technology, and big data, offers a solution by enabling intelligent production, leading to smart supply chains, smart working practices, smart manufacturing, and smart products. A virtual product developed during the industry 4.0 era, the smart supply chain, can facilitate intelligent production.

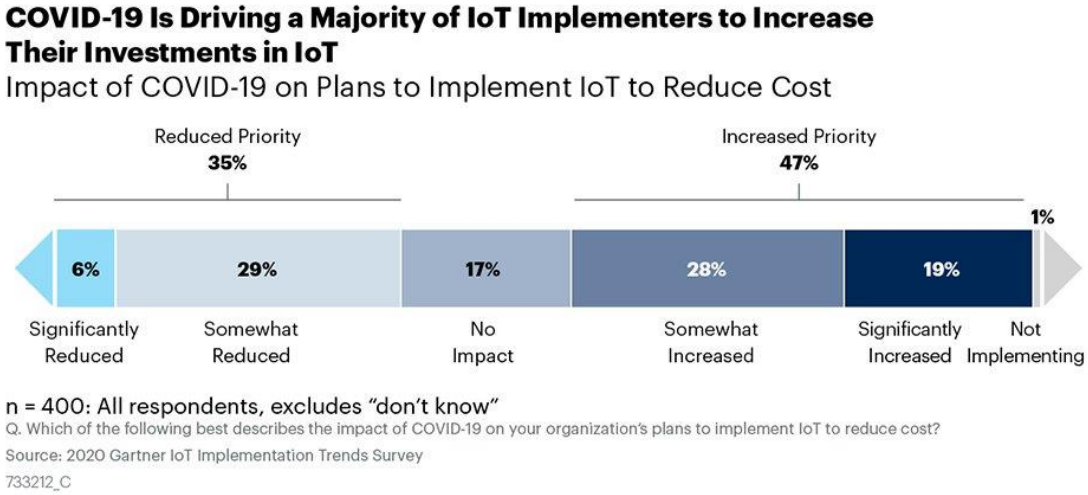


Figure 1.1 Survey analysis: Covid-19 use cases driving IoT adoption, powered by innovations like AI and digital twins, conducted online from June through July 2020, with 402 respondents across the US, the UK, Germany, Australia, Singapore and India.

Environmental sustainability is emerging as a critical factor in the pharmaceutical industry's success. As market demands evolve, healthcare systems are under increasing pressure to adopt a holistic approach to the entire drug lifecycle, including its environmental impact. Digitalisation and IoT are the key components to improve sustainability. Sustainability can be achieved by prioritising digitisation as a necessary component of business-driven efficiency efforts. A significant challenge in pharmaceutical manufacturing is the high volume of waste generated. For small molecule drugs, the average waste-to-product ratio is a

staggering 200:1, with some products reaching ratios as high as 800:1 (Martin, 2021). This disparity is even more concerning in the biopharmaceutical sector, where waste-to-product ratios can reach a shocking 10,000:1 (Martin, 2021). These statistics highlight the immense inefficiency within the current manufacturing processes. To illustrate the real-world impact, consider a 2009 report by the Environmental Protection Agency (EPA) which revealed that US pharmaceutical companies alone generated a staggering 530 million tons of waste (Martin, 2021). This waste not only represents a significant financial burden but also raises environmental concerns associated with disposal and potential contamination.

While there is growing interest in integrating Industry 4.0 technologies into pharmaceutical manufacturing, their practical application remains a novel endeavour. Despite claims of successful integration, empirical evidence supporting complete cross-unit technological harmony is scarce. Incompatible platforms, models, and systems persist as significant barriers to seamless integration. Given the proliferation of emerging Industry 4.0 technologies with diverse compatibility standards, a systematic approach is essential to design digital solutions that are rapid, cost-effective, and adaptable to the dynamic pharmaceutical landscape. This thesis will also serve as a steppingstone in that approach, addressing the identified issues.

Drug manufacturing issues can be effectively addressed with digital twin technology. High Fidelity Digital twins may dramatically enhance efficiency by enable rapid modelling and simulation, expediting the development, and testing of products and processes as well as increasing productivity through automation and predictive control. Sensors and other Industry 4.0 technology allow for real-time tracking of production performance and product data throughout the product lifecycle. This allows for smart packaging, optimised supply chains, and sustainable process development by minimising experimentation time and waste, maintaining consistent quality, and enabling a "quality by design" approach. Furthermore, digital twins can improve process quality and dependability while monitoring sustainability parameters throughout the product lifetime.

Three key strategies that can propel businesses towards sustainable development in the digital age:

- **Data-driven optimisation:** Leveraging the power of data analytics to identify areas for reducing energy consumption and lowering carbon emissions.
- **Cross-industry collaboration:** Fostering collaboration across sectors and disciplines to accelerate progress. This collective effort can revitalise connected infrastructure, unlock energy efficiency opportunities, solidify sustainability commitments, and promote renewable energy usage.
- **Unlocking IoT potential:** Embracing IoT to unleash opportunities for enhanced process and energy efficiency. As digital transformation continues to escalate, integrating IoT solutions will be crucial for adapting to the demands of a sustainable future.

1.3 Pharmaceutical Manufacturing in the Drug Timeline

The pharmaceutical drug life cycle encompasses several stages, including drug research and development, drug production and scale-up, and drug distribution. Before regulatory approval and large-scale manufacturing, drugs undergo rigorous clinical trials and studies (Hoang, 2015). While the unit operations for producing Active Pharmaceutical Ingredients (APIs) and liquid formulations share similarities with biomanufacturing, the production of oral solid dosage forms necessitates specialised equipment like blenders, twin-screw wet granulators, and tablet presses (secondary processing).

Pharmaceutical supply chains play a crucial role in ensuring the timely availability of drug inventory for distribution to healthcare providers and patients. The typical duration of a pharmaceutical drug product life cycle is 10-15 years (Hoang, 2015). During the manufacturing stage, extensive and internal checks are done before it is supplied to the consumer. Before reaching the market, drug products undergo extensive quality control and regulatory scrutiny. This includes formulation development, process design, and optimisation, followed by rigorous testing and validation at various stages of the manufacturing process. The manufacturing process itself must adhere to stringent Good Manufacturing Practices (GMP)

guidelines, a comprehensive system designed to minimise risks and ensure product quality. GMP encompasses a wide range of quality control measures, including documentation, training, equipment calibration, and environmental monitoring.

To ensure product consistency and safety, pharmaceutical manufacturers must carefully manage the entire supply chain, from the procurement of raw materials to the distribution of finished products. This involves optimising inventory levels, minimising lead times, and mitigating potential disruptions. By effectively managing supply chain operations, pharmaceutical companies can help to ensure that patients have access to the medications they need, when they need them. Section 501(a)(2)(B) of the Federal Food, Drug, and Cosmetic Act under cGMP products states that “A drug shall be deemed to be adulterated if the methods used in, or the facilities or controls used for, its manufacture, processing, packing, or holding do not conform to or are not operated or administered in conformity with current good manufacturing practice to assure that such drug’’(Production and process controls overview of CGMP,2015). GMP regulations outline specific actions to control how drug ingredients and final products are made. This includes producing the intended dosage form (like tablets or injections), checking materials throughout the process, and verifying that the manufacturing steps work as intended.

1.4 Continuous Manufacturing (CM)

The pharmaceutical industry has witnessed a shift from traditional batch operations to CM, particularly in specialised areas (Lerapetritou et al., 2016). This transition offers significant advantages, including improved efficiency, flexibility, and reduced costs (Sarkis et al., 2021). Additionally, continuous manufacturing enables the exploration of new chemistries that may not be feasible with batch processes. Recent advancements in technology have further fuelled interest in advanced manufacturing techniques to improve drug quality, address shortages, and accelerate time-to-market.

CM is a key approach to process intensification, facilitating the implementation of digital twins in the pharmaceutical industry. One major benefit of continuous manufacturing is the potential for enhanced in-process control, leading to

reduced out-of-specification batches and minimised waste. CM enables a seamless flow of product between unit operations, allowing for real-time monitoring and control of product quality throughout the process. By leveraging in-line quality control systems and closed-loop feedback mechanisms, which rely on advanced Process Analytical Technology (PAT) and digital data processing, continuous manufacturing offers enhanced process control and reduced variability.

In addition to its flexibility and cost-saving advantages, continuous manufacturing provides a significant benefit by enabling real-time monitoring and control of Critical Quality Attributes (CQAs) and Critical Process Parameters (CPPs). This real-time monitoring capability is crucial for implementing advanced control systems and ensuring consistent product quality.

1.5 Smart Manufacturing in the Pharmaceutical Industry

Smart manufacturing, a cornerstone of I4.0, leverages advanced technologies such as big data, advanced analytics, and AI to optimise manufacturing processes. By integrating these technologies, smart manufacturing enables intelligent, efficient, and responsive operations, ultimately leading to sustainable process design and personalised healthcare solutions.

Key features that underpin smart manufacturing include:

- **Near Real-Time DTs**: Virtual representations of physical assets and processes that enable real-time monitoring, analysis, and optimisation.
- **PAT-Based Control**: Advanced control strategies that utilise real-time process data to optimise performance and ensure product quality.
- **Continuous Data Acquisition**: The collection and integration of data from various sources, including sensors, equipment, and systems, to provide a comprehensive view of the manufacturing process.
- **Continuous Global Modelling and Data Analysis**: The development and application of advanced models to analyse large volumes of data, identify trends, and make informed decisions.

The pharmaceutical industry has experienced a transformative shift in recent years, driven by the adoption of advanced manufacturing technologies such as Quality by Design (QbD) as depicted in Figure 1.2.

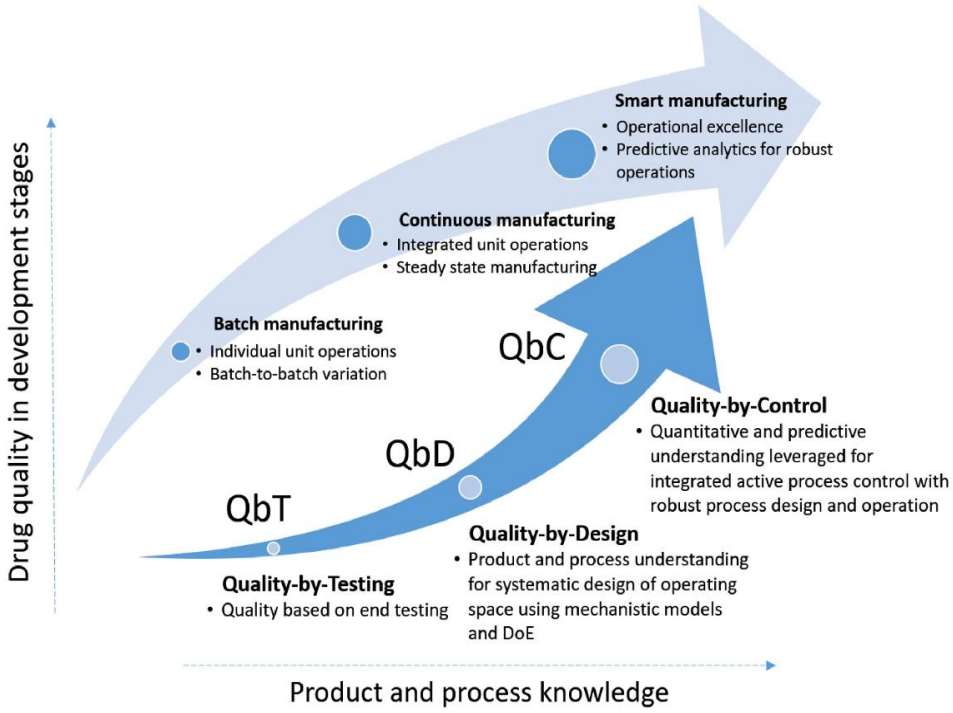


Figure 1.2 Advancing Quality Assurance: A Journey from Testing to Design and Control (Su et al., 2022).

These innovations have enabled manufacturers to significantly enhance product quality, reduce production costs, and accelerate time-to-market. To further optimise manufacturing processes and improve decision-making, the industry has embraced the concept of digital twins. By integrating data from various sources, including sensors, manufacturing equipment, and enterprise systems, digital twins create a virtual representation of physical assets and processes. This virtual representation enables real-time monitoring, analysis, and simulation, providing valuable insights for predictive maintenance, process optimisation, and quality control.

The development and implementation of digital twins have been facilitated by advancements in data analytics, AI, and cloud computing technologies. These technologies enable the collection, storage, and analysis of vast amounts of data, allowing for the identification of patterns, trends, and anomalies. By leveraging these insights, manufacturers can make data-driven insights and decisions to improve efficiency, reduce waste, and enhance overall operational performance.

1.6 Process Digital Twin and Enabling Technologies

While the concept of digital twins has evolved over time, various terms like "Digital model" and "Digital shadow" have been used to describe them (Kritzinger et al., 2018). However, these terms often differ in the extent of data integration between the physical and virtual product. A Process Digital Twin (DT) should seamlessly integrate with real-time data bilaterial exchange and provide functional outputs, as illustrated in Figure 1.3 (Chen et al., 2020; Kritzinger et al., 2018).

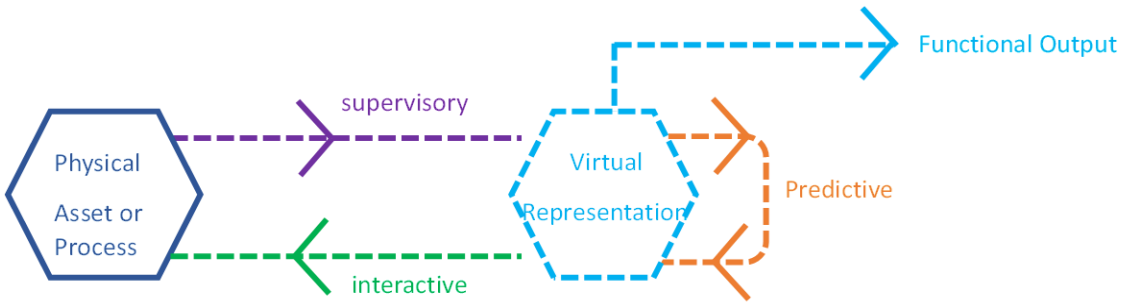


Figure 1.3 Categorisation of a Digital Twin. Taken from: Advanced Manufacturing Research Centre, 2020.

The conceptual framework for a digital twin involves a synergistic interaction between real-world components and their virtual counterparts, seamlessly connected through a robust cloud platform. This cloud-based infrastructure harnesses the power of innovative technologies like IoT and cloud computing to enable complex processing, sophisticated data analysis, secure data management, and seamless device connectivity. The architectural design of the digital twin elucidates the roles of each layer, the intricate relationships between them, and the mutual benefits they derive. Furthermore, digital twins can leverage the insights gleaned from virtual simulations to make informed decisions and address real-world challenges, thereby optimising operations and enhancing overall performance.

Table 1.1 provides an overview of essential digital twin enabling technologies required for the successful implementation of digital twins in the pharmaceutical industry. To construct a comprehensive digital twin encompassing physical

entities, virtual models, data management, services, and connectivity, the integration of various technologies is crucial.

Table 1.1 Digital Twin Enabling Technologies in the Pharmaceutical Industry.

Service platforms	*Data management platforms*	*Modelling and simulation platforms*
Optimisation • Simulink • FLEXSIM • Beacon • Predix • PharmaMV	Data visualisation • Tableau • Data V • FLOT	Geometric modelling • Auto CAD • UG • SolidWorks
Simulation • 3dMAX • Abaqus • Simulink	Data fusion • MATLAB • Predix • Beacon	Physical modelling tools • Simulink • ANSYS • Solid works • gPROMS
Diagnosis • MATLAB • ANSYS • 3D experience	Data transmission • MindSphere • Predix • Thingworx • Aspera	Behavioural modelling tools • Twin Builder • ADAMS • Composer

The accuracy of virtual models significantly impacts the effectiveness of digital twins. Therefore, rigorous validation, verification, and accreditation processes, along with optimisation algorithms, are essential to ensure model accuracy and reliability. Model evolution technologies are indispensable for keeping virtual models up to date with changes in the physical world. Furthermore, application software and service-oriented architecture technologies are necessary to deliver essential services. Sophisticated methodologies, including machine learning, deep learning, and advanced statistical modelling, coupled with robust data fusion algorithms, are essential for deriving actionable insights from heterogeneous datasets. Real-time monitoring of physical assets and processes necessitates the utilisation of advanced visualisation technologies. Finally, internet technologies and interaction technologies facilitate seamless connectivity and user interaction with the digital twin.

1.7 Challenges

Industry 4.0 delivers many benefits to process industries; however, some barriers impede digital transformation. These include:

- **Legacy Processes and Long Development Cycles**: Traditional, resource-intensive development cycles and entrenched legacy processes create significant roadblocks to fully leveraging the potential of digital twins in the pharmaceutical industry (Subramaniam, 2020; Huang et al., 2021).
- **Lack of Comprehensive Models:** Currently, there is no single, robust model that effectively captures the complex relationships between Critical Process Parameters (CPPs) and Critical Quality Attributes (CQAs) across all unit operations within an integrated pharmaceutical manufacturing process (Chen et al., 2020). Additionally, implementing these comprehensive models within digital simulations can be computationally expensive and difficult to deploy online (Chen et al., 2020).
- **Inaccurate Models Hinder Progress**: Digital twins rely heavily on accurate models for simulation and real-time analysis. Unfortunately, the continued presence of inaccurate models remains a significant barrier to realising the full potential of digital twins in pharma (Zhang and Yan, 2019).
- **Underutilised Data**: Research suggests that a staggering 70% of pharmaceutical manufacturing data remains uncollected due to insufficient sensor deployment (Manzano & Langer, 2018). This represents a significant missed opportunity and hinders the development of robust digital twins.
- **Limited Machine-to-Machine Communication**: The lack of sensors and intelligent devices for machine-to-machine interaction and collaboration significantly restricts the potential for expanding production flexibility within a digital twin framework (Gunes et al., 2014).
- **Heterogeneous Software**: Standardisation of equipment software is crucial for seamless integration within a digital twin environment. Developing a standardised interface or file format would address the current challenge posed by the heterogeneity of equipment software across various machines (Chen et al., 2020).
- **Strict Regulatory Landscape**: The pharmaceutical industry operates under rigorous regulatory frameworks. This often translates into a longer adoption timeline for innovative tools like Advanced Process Control (APC) and continuous manufacturing, which can diagnose process variabilities, compared to other sectors (Gunes et al., 2014; Huang et al., 2021).

1.8 Aim and Objectives

Thesis aim

Using complex fundamental models for online monitoring and exploration in pharmaceutical manufacturing can significantly enhance process control, improve product quality, and reduce downtime. This thesis aims to provide a framework and case study evidence to harness the power of complex fundamental models for DT creation, inline monitoring, and exploration, leading to more efficient, dependable, and cost-effective pharmaceutical manufacturing processes.

The following novel aspects and research questions are addressed in this work:

1. How are digital twins developed and used to support real-time, autonomous decision-making in complex, multi-operation environments where data availability, processing power, and network latency are significant constraints?
2. Can the implementation timeline and efficiency of integrating complex fundamental models for online monitoring and exploration in manufacturing be optimised through streamlined model development/model selection, calibration, validation, and integration processes?
3. What are the slowest steps in the process of using complex fundamental models for online monitoring/real-time optimisation and exploration in manufacturing, and how can these bottlenecks be resolved to improve efficiency?
4. Can Machine learning and big data play important roles in DT?

Novelty and Contributions of This Work

This work focuses on several key advancements in the field of digital twins for pharmaceutical processes:

1. Accelerated Deployment of Integrated Mechanistic Models:

- Traditional development and validation of mechanistic models can be a slow and iterative process. This work proposes novel framework called AFRAME to streamline this process, potentially leading to:
    - **Reduced development time** through techniques like sensitivity analysis and surrogate modelling. These approaches identify the most critical parameters for model calibration, reducing the number of experiments required and accelerating model development.
    - **Improved model efficiency** by utilising computationally efficient surrogate models in place of complex mechanistic models for specific applications.

To develop a versatile modelling framework capable of creating hybrid models suitable for usage in a digital twin, certain specific requirements must be met. The main criteria for the model and its framework are as follows:

- The model must be able to effectively use a dataset that is continuously expanding.
- The model must enable rapid development.
- The process of creating and updating the model should involve minimal manual effort.
- The model should be computationally efficient for training and evaluation.
- The framework used for modelling should apply to a wide range of processes.

2. Leveraging AI, IoT, and Hybrid Modelling:

- This work develops new techniques and framework to integrate various technologies:
    - **Artificial Intelligence (AI):** Techniques like machine learning can be used for tasks such as outlier detection, data-driven parameter estimation, and model improvement. This can result in faster model building and adaptation to changing process conditions.
    - **Internet of Things (IoT):** Dense sensor networks can provide real-time data for model calibration, validation, and updating, leading to more accurate and adaptive models.

- **Hybrid Modelling:** Combining mechanistic models with data-driven models (e.g., AI) can create more robust and adaptable models that capture both the fundamental process behaviour and the complexities of real-world data.

These technologies offer significant advancements compared to traditional methods, potentially leading to faster and more accurate model development, improved process control, and enhanced product quality and manufacturing efficiency.

3. Maintenance-free universal DT can be adopted to any process quickly:

- The concept of a "maintenance-free" digital twin is ambitious but can be achieved through strategies like:
  - **Self-learning Models:** Implementing AI techniques that enable models to learn and adapt from real-time data can potentially reduce the need for manual intervention and updates.
  - **Automated Data Validation and Model Calibration:** Automating data cleaning and model calibration procedures can minimise the need for manual adjustments, leading to a more self-sustaining digital twin framework.

While the concept of a completely maintenance-free digital twin may be challenging to achieve in practice, this work aims to develop a framework that minimises manual intervention and maximizes model autonomy.

4. Digital Operations Deployment with Cloud and Systems Engineering:

- This work explores the integration of these models with digital operations such as:
  - **Virtual Design of Experiments (DoE):** Utilising models within a virtual environment allows for cost-effective experimentation and optimisation of process parameters without the need for physical experiments.

Evidence for the effectiveness of these approaches comes from the growing body of research in process modelling and control. Cloud technology plays a crucial role by providing scalable computing resources to oversee the demands of complex models and real-time data processing. Systems engineering principles emphasise a comprehensive approach, ensuring seamless integration of models and data across different platforms within the digital environment.

5. Sustainability oriented digital twins

This research integrates a comprehensive sustainability model into the pharmaceutical process, addressing economic considerations often overlooked in previous studies. By adopting a system engineering approach, the project takes a holistic view of the entire pharmaceutical process rather than focusing on individual operations in isolation. This holistic approach enables the identification of optimisation opportunities across the entire process. To fully understand and optimise continuous manufacturing processes, it is essential to consider not only the individual unit operations but also the intricate interactions between these units and their impact on the final product. By analysing these interdependencies, we can develop more efficient and sustainable manufacturing strategies.

The following provides a comprehensive summary of the primary objectives of this research that guided the thesis's focus and methodology:

- Designing and implementing a comprehensive high-fidelity DT Framework for pharmaceutical manufacturing processes.
- Enhancing model development speed for high-fidelity Models.
- Leveraging ML to create fast solving models.
- Propose a faster path to model Maintenance
- Designing a workflow for integrating high fidelity models into DT framework.
- Developing sustainability-oriented DT.

1.9 Thesis scope

Creating a comprehensive digital twin for a complex manufacturing process like the ConsiGma-25 line, installed at the Diamond Pilot Plant (DiPP) (Figure 1.4), is a significant undertaking. Such a digital twin would require the integration of

various data sources, advanced modelling techniques, and sophisticated simulation tools. To narrow down the extensive range of tasks involved in creating a digital twin, it is necessary to impose some limitations to shrink the issue area.

The initial step in developing a sustainability-oriented digital twin involves defining the specific sustainability metrics to be focused on. Sustainability can be assessed through various parameters, including:

- Water usage
- Energy consumption
- Waste generation.

In this thesis, I have chosen to prioritise energy usage as the primary focus for the digital twin. Even with this specific focus, building a comprehensive digital twin for the DiPP remains a substantial undertaking, requiring significant effort and resources.

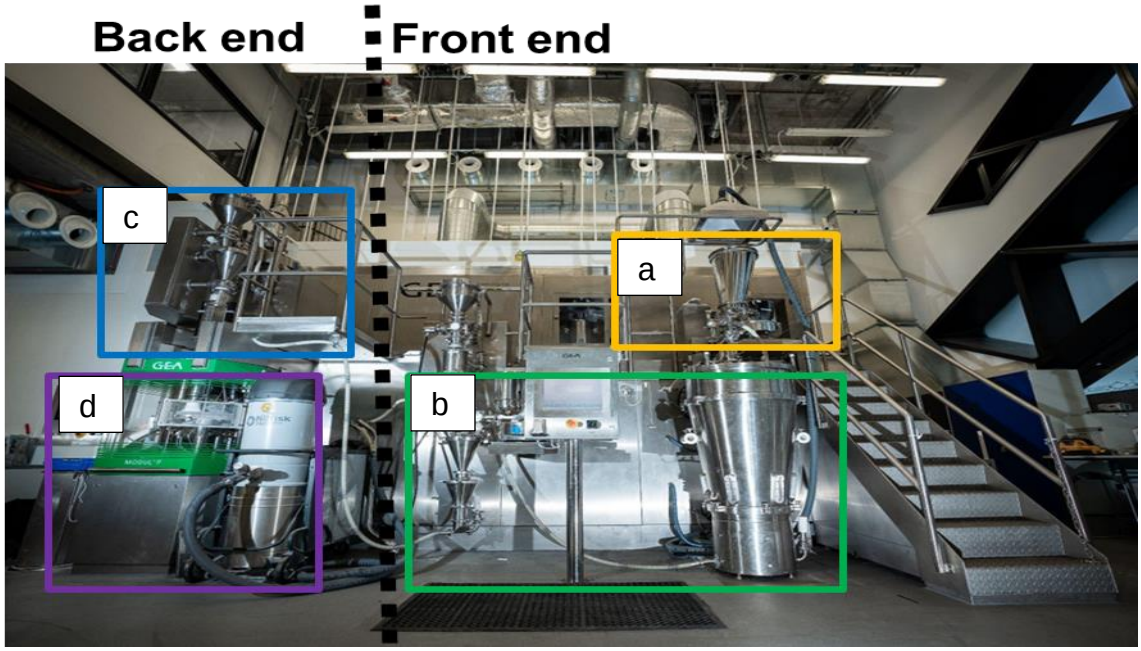


Figure 1.4 Setup of a fully integrated ConsiGma-25 by GEA Pharma System's powder to tablet line at the University of Sheffield: (a)Twin screw granulator, (b)Conditioning unit, (d)Blender and (d)Tablet press.

The DiPP comprises numerous unit operations, including the blender, twin-screw wet granulation (TSWG) unit, Fluidised Bed Dryer (FBD), and tablet press (GEA Group AG, 2023; Ryckaert, 2021). To make the project feasible, further narrowing of the scope is necessary. One approach is to divide the Consigma line into two distinct parts: back end, and front end. By focusing on

specific sections of the line, I can create a more manageable and focused of the research. The TSWG unit was chosen for model development and energy analysis, extending the application of the framework to additional unit operations, specifically those in the back-end processes of the case studies. This selection is driven by two key factors. Firstly, the twin-screw granulator is known for its significant energy consumption. Secondly, a substantial body of scientific research exists on the fundamentals of twin-screw granulation, which provides a strong foundation for my analysis (Vercruysse et al.,2012; Kumar, 2015; Altair Engineering Inc., Whitepaper, 2022; Seem et al.,2015; McGuire et al.,2017).

1.10 Thesis Map

In summary, offering an accelerated method for deployment of fast solving, cost-effective and adaptive models is the focus of this research. Figure 1.5 provide a map to navigate the lay out of this thesis.

Chapter 1 sets the stage by providing an overview of the current state of DT technology in the pharmaceutical industry, identifying key drivers for its adoption, and outlining the specific objectives of this research.

Chapter 2 delves deeper, conducting a comprehensive literature review to identify relevant insights and knowledge gaps. This chapter explores various applications of DT across different industries, examines advancements in DT components within pharmaceutical manufacturing, and provides a comprehensive summary of I4.0 technologies and how they can be exploited.

Chapter 3 outlines the experimental methodology, including the design rationale, types of experiments conducted, analytical techniques employed, and the utilisation of cloud technologies for data collection and storage. Chapters 4 through 8 present and discuss the findings of the research.

Finally, **Chapter 9** outlines future research directions, including online model deployment, Digital Model driven Control, and the implementation of virtual design of experiments.

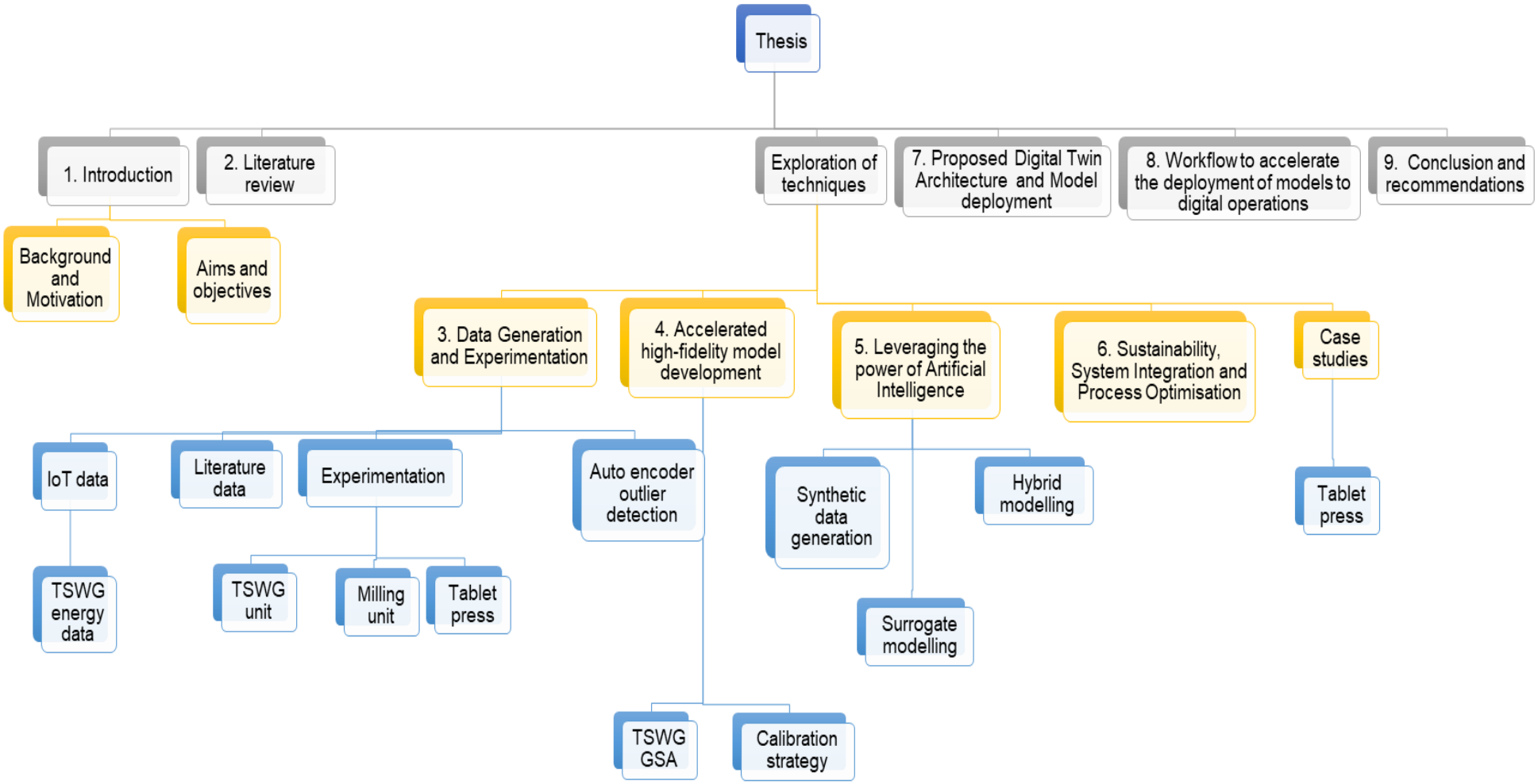


Figure 1.5 Thesis Report Map

Chapter 2: Literature review

2.1 Introduction

Industry 4.0 research has concentrated on the implementation of its revolutionary technologies, which allow for the creation of DTs in a variety of industries, including process manufacturing. Several current reviews have examined Industry 4.0 technology and the use of DTs in various industries, including discrete and process manufacturing. This section seeks to explore the unique obstacles and limitations to the introduction of DTs in the pharmaceutical business. To do this, the study will look at the present status of the pharmaceutical industry and undertake a comprehensive evaluation of existing DT uses in the sector.

This research will look at numerous case studies to highlight current gaps in the research landscape. A comprehensive study of relevant works will be carried out, with a special emphasis on current I4.0 technologies and breakthroughs that might help to build unique digital twin framework in the pharmaceutical business. This section will also look at modelling techniques for pharmaceutical processes ranging from twin-screw granulation to tableting. It will also look at the present status of research and industry applications for digital twins in pharmaceutical manufacturing, such as sensor technologies, model development approaches, and data strategies for integration. Finally, the section will discuss the problems and opportunities for further research in this topic.

2.2 Pharmaceutical Manufacturing Processes

Various granulation techniques, such as high-shear granulation, roller compaction, fluidised bed granulation, and twin-screw granulation, have been developed to process solid particles. Figure 2.1 illustrates a comprehensive pharmaceutical tablet manufacturing process, beginning with the feeding of Active Pharmaceutical Ingredient (API) and excipients, both subject to identity verification. The process then progresses through blending to achieve homogeneity, with in-line monitoring ensuring proper mixing.

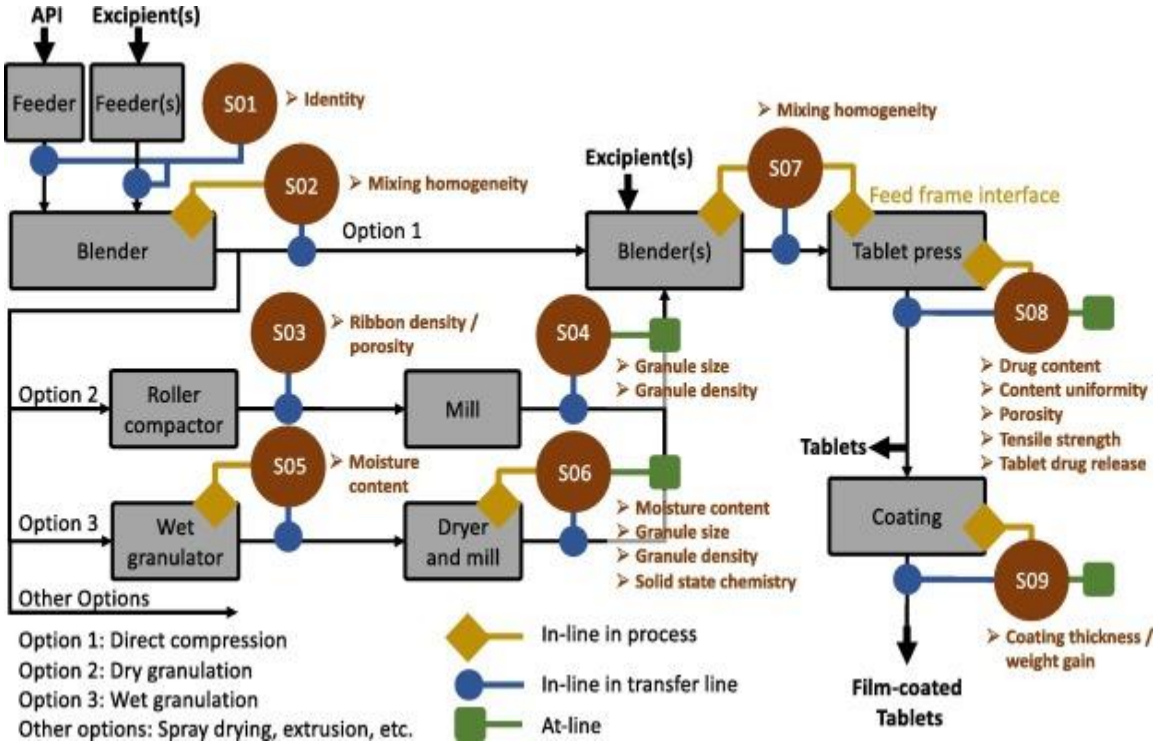


Figure 2.1 Process Flow Diagrams: Key Testing Points and Properties of Interest in Direct Compression, Dry Granulation, and Wet Granulation (Markl et al., 2020)

Granulation, a critical step to enhance flowability and compressibility, presents several options: direct compression for suitable blends, dry granulation using a roller compactor with ribbon property checks, or wet granulation involving wet mass processing, drying, and milling, all monitored for moisture content and granule characteristics. Milling refines the granules to the desired size and density, followed by secondary blending, again with homogeneity checks, before the final compression into tablets. This tableting stage involves rigorous monitoring of drug content, uniformity, porosity, tensile strength, and drug release. An optional coating process may then be applied, with coating parameters carefully controlled. Throughout the process, in-line, at-line, and in-transfer line monitoring points (S01-S09) are strategically placed to ensure quality control, covering various parameters from material identity to finished tablet attributes, with additional processing options like spray drying and extrusion also noted. Compared to direct compression, dry and wet granulation are more prevalent in industrial settings as they improve powder flowability and compressibility, enabling higher drug loads and better content uniformity, especially for low-dose formulations (Ennis, 2016). Key granule characteristics,

including PSD and porosity, are influenced by various macroscopic rate mechanisms such as nucleation, layering, breakage, and consolidation.

The pharmaceutical industry is increasingly transitioning from traditional batch processes to continuous manufacturing (CM), which offers significant advantages in terms of efficiency and flexibility (Lee et al.,2015). Twin-screw wet granulation (TSWG) is a promising continuous process that can be readily integrated into continuous manufacturing lines. TSWG offers several benefits (Teżyk et al., 2016) including:

- Negligible residence time
- Excellent granule properties,
- Superior mixing of APIs

According to European Public Assessment Reports, wet or dry granulation is a common step in over 70% of tablet formulations (Leane et al., 2018). TSWG requires significantly less water to produce granules of the desired size (Megarry et al., 2020). Additionally, TSWG granules tend to be less spherical and more porous than those produced by high-shear granulation (Teżyk et al., 2016; Matsui and Watano, 2018; Fülöp et al., 2021). These unique granular characteristics can enhance tablet tensile strength due to increased fragmentation during compaction (Matsui and Watano, 2018).

An integrated tablet line, as depicted in Figure 2.2, offers a seamless and efficient approach to solid dosage form manufacturing. Excipients and active pharmaceutical ingredients (APIs) are individually fed into a continuous inline blender using a series of precise gravimetric feeders, ensuring accurate dosing and consistent blend composition. The resulting blend is then transferred to a twin-screw granulator, where it undergoes a controlled wet granulation process to form granules of the desired size and shape. The wet granules are subsequently transferred, either gravimetrically or pneumatically, to a semi-continuous drying unit, where excess moisture is removed. The dried granules are then milled to achieve the desired PSD, optimising their flowability and compressibility. In the next stage, the dried and milled granules are blended with additional granular excipients, such as magnesium stearate, which acts as a

lubricant. These excipients are added using gravimetric feeders to maintain precise dosing and ensure consistent blend composition.

The final blend is then fed into a rotary tablet press, where it is compressed into tablets of the desired shape and weight. The produced tablets may undergo further processing, such as coating, to enhance their stability, appearance and dissolution properties.

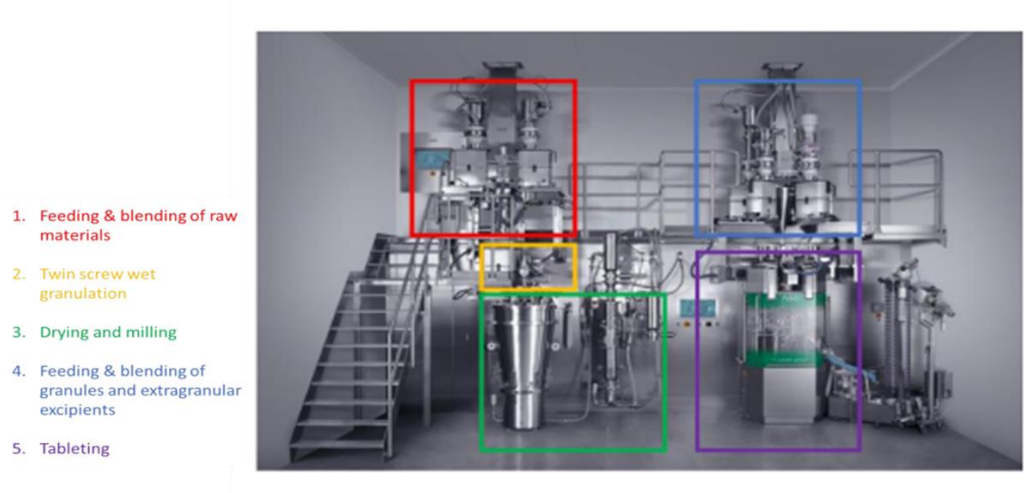


Figure 2.2 GEA Pharma Systems' Consigma Powder-to-Tablet Lines (Portier et al., 2021).

While CM and digital technologies have been successfully implemented in industries such as petrochemical and food manufacturing for several decades, their adoption in the pharmaceutical industry was initially hindered by regulatory hurdles and a lack of clear guidelines (Thompson, 2015; Vervaet and Remon, 2005 Ding, 2018). However, recent years have witnessed significant progress in the pharmaceutical industry's adoption of these technologies, driven by several key factors:

- **Advancements in Equipment and Technology**: The development of fully integrated manufacturing lines by equipment manufacturers like GEA Pharma Systems, L.B. Bohle, and Glatt has significantly accelerated the adoption of continuous manufacturing technologies (Teżyk et al., 2016).
- **Regulatory Support**: Regulatory agencies like the FDA, EMA, and ICH have played a crucial role in facilitating the adoption of novel manufacturing techniques by providing clear guidelines and regulatory

frameworks (Lerapetritou et al., 2016; U.S. Food and Drug Administration, 2004).

- **Industry-Academia Collaboration**: Numerous collaborations between industry and academia, through consortia such as CESPE, C-SOPS, CMAC, and the European Consortium for Continuous Pharmaceutical Manufacturing, have fostered innovation and accelerated research in the field of continuous manufacturing.

As a result of these factors, major pharmaceutical companies such as Janssen, Vertex, Lilly, and Pfizer have successfully obtained market approval for various medications manufactured using continuous processes like direct compression and TSWG (U.S. Food and Drug Administration, 2018a, 2018b). The increased interest of the industry in continuous manufacturing has also led to several partnerships with academic institutes and consortia. These collaborations have been one of the key drivers of increased research output in this subject over the last 10 years as shown in Figure 2.3.

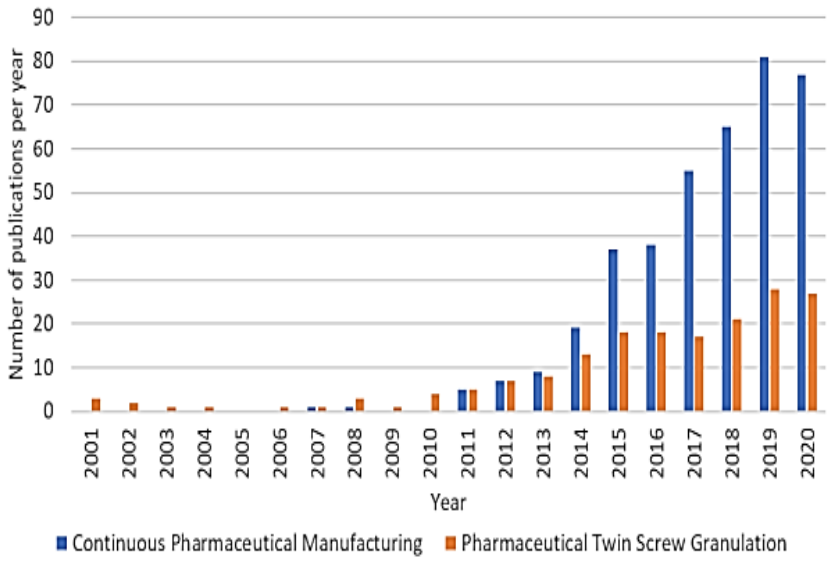


Figure 2.3 Tracking the Evolution of Research in Continuous Pharmaceutical Manufacturing and TSWG (Portier et al., 2021).

2.3 Advanced Process Modelling of the Wet Granulation Manufacture of Pharmaceutical Tablets

This section delves into the various processing units employed for wet granulation process of pharmaceutical formulations for solid dosage forms. It discusses their

advantages, disadvantages, and the application of mechanistic modelling approaches to simulate their behaviour.

While previous research primarily focused on the impact of process settings and equipment configurations on critical quality attributes (CQAs) (Vercruysse et al., 2012, 2015; Fonteyne et al., 2015; Meier et al., 2017), recent studies have emphasized the importance of formulation and raw material properties, as well as regulatory considerations like PAT implementation and control strategies (Teżyk et al., 2016; Madarász et al., 2018; Portier et al., 2020).

A significant limitation of existing research on TSWG is its focus on specific formulations, hindering the generalisation of findings and their application to diverse formulations. To address this limitation and facilitate the effective implementation of TSWG, the development of robust mechanistic models for continuous granulation processes is crucial. These models can provide valuable insights into the underlying mechanisms, enabling process optimisation, control, and scale-up. Regardless of the specific technology used, wet granulation processes share fundamental mechanisms that govern the final product's characteristics, such as particle size distribution, bulk, and envelope densities, and flowability. Figure 2.4 provides a visual representation of how these mechanisms can be modelled.

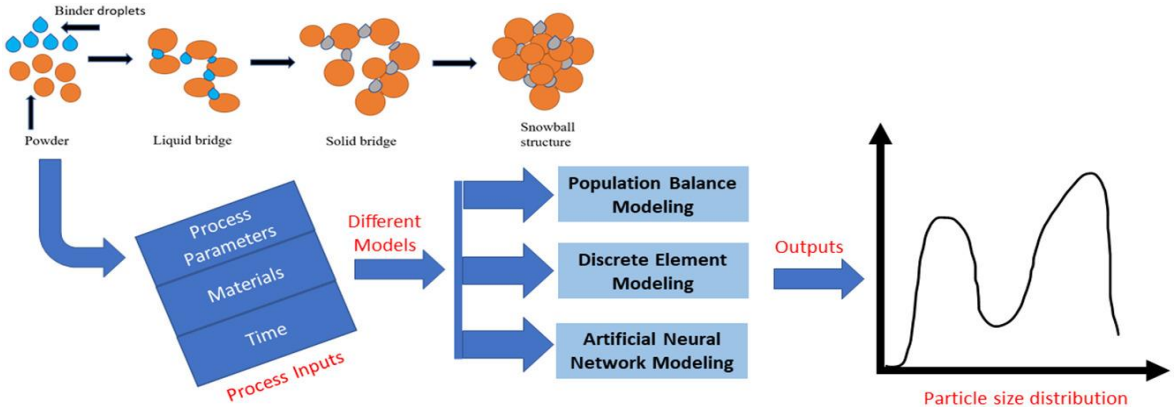


Figure 2.4 Three different modelling techniques including PBM, DEM, and data driven for wet granulation processes (Singh et al., 2022).

Modelling approaches can be broadly categorised into mechanistic, data-driven, and hybrid modelling. Mechanistic modelling techniques, such as the Discrete Element Method (DEM), Finite Element Method (FEM), and Computational Fluid Dynamics (CFD), are commonly employed in pharmaceutical manufacturing to

simulate particle-level and bulk behaviour. While DEM is effective in capturing detailed particle-level interactions, its high computational cost limits its practical application. However, advancements in High-Performance Computing (HPC) and cloud computing have enabled the integration of DEM simulations into overall process models, potentially enabling near-real-time simulations (Chen et al., 2020). Table 2.1 summarises the discussion of CFD, PBM, DEM and FEM, highlighting their key features and applications in pharmaceutical manufacturing (Arthur et al., 2024).

CFD and FEM are widely used to model fluid flow in pharmaceutical processes, including API drying and fluidised bed operations. These techniques are also extensively used in biopharmaceutical manufacturing (Chen et al., 2020). To address the limitations of purely mechanistic and data-driven approaches, hybrid modelling techniques have emerged. These techniques combine the strengths of both approaches, reducing computational cost and data requirements while improving model accuracy and predictive capabilities. Population balance modelling (PBM) is a semi mechanistic computationally efficient method for modelling blending and granulation processes. The material balance in the wet granulation process has been modelled in terms of population balance equations (PBEs), which track the change in particle size over time for different classes. PBEs are the established framework for particulate systems with distinct and evolving particle size distributions (PSDs). Moreover, PBEs are nearly indispensable where the rate processes depend on the particle sizes and compositions. Therefore, a semi-mechanistic PBE model has been developed in which the equations were developed from the first principles, but some experimental data and/or material properties were used as input parameters to the rate equations.

Table 2.1 Discussion of CFD, PBM, DEM and FEM, highlighting their key features and applications in pharmaceutical manufacturing

Modelling Technique	*Focus*	*Key Applications in Pharmaceuticals*	*Modelling outputs*
Computational Fluid Dynamics (CFD)	Fluid flow, heat transfer and mass transfer	Reactor design, mixing, bioreactor optimisation, HVAC systems	Provides insights into fluid dynamics within equipment; crucial for processes involving liquids or gases
Particle Population Balance Modelling (PBM)	Evolution of particle size distributions and population properties	Crystallization, granulation, milling, powder processing, continuous manufacturing optimisation of particle attributes	Critical for predicting and controlling particle size, a key quality attribute in many pharmaceutical products; valuable for real-time control.
Discrete Element Modelling (DEM)	Individual particle interactions (collisions, friction, etc.) in granular materials	Powder flow in blenders, tablet compression, granular material handling	Provides detailed insights into particle-level behaviour; enhances simulation of powder processing and granular material handling.
Finite Element Modelling (FEM)	Mechanical behaviour of solid objects (stress, strain, deformation)	Tablet compression, packaging integrity, medical device design, structural analysis	Simulates mechanical behaviour of products and equipment; ensures product quality and safety; essential for devices and packaging

PBM Suitability for Digital Twins:

- Critical Quality Attributes:
    - Particle size distribution is often a critical quality attribute (CQA) in pharmaceutical products. PBM allows for real-time prediction and control of this CQA within a digital twin.
- Process Understanding:
    - PBM provides a deep understanding of the underlying mechanisms that govern particle behaviour, enabling better process optimisation and control.
- Continuous Processes:

    - In continuous manufacturing, where steady-state conditions are desired, PBM can predict how changes in process parameters affect particle size distribution over time, allowing for proactive adjustments.
- Integration:
    - PBM models can be integrated with other models (like CFD) within a digital twin to provide a comprehensive representation of the process. For example, CFD can provide the hydrodynamic environment, while PBM models the particle behaviour within that environment.
- Real time optimisation:
    - Because of the ability of PBM to predict particle attributes, it allows for real time adjustments to the process, thus enabling real time optimisation.

Equation 2.1 presents a one-dimensional population balance.

$$\frac{\partial}{\partial t} n_i(v,t) + \frac{\partial}{\partial v}[n_i(v,t)G_i(v)] = B_{nuc,i}(v,t) + B_{break,i}(v,t) - D_{break,i}(v,t) + \frac{n_i(v) - n_{i-1}(v)}{\tau_i} \quad 2.1$$

Where:

- v is granule volume and t is time.
- $n_i(v,t)$ is granule number density of species with volume v at time t.
- $G_i(v)$= Growth rate by layering and consolidation
- $B_{nuc,i}(v,t)$ is birth rate of granules due to nucleation
- $B_{break,i}(v,t)$ and $D_{break,i}(v,t)$ is birth and death of granules due to breakage.
- τ_i is mean residence time.

Each term in the equation can be described using rate processes mechanism kernels which are based on established mechanistic understanding of twin-screw wet granulation (TSWG). The right side of the equation describes the birth and death terms for breakage. This term is normally used the describes milling

process where breakage is the significate mechanism. The added of the $B_{nuc,i}(v,t)$ term in the right of the equation considers the birth of particle due to nucleation and described the process in the twin granulator where a liquid binder is added. In systems where only coating/growth occurs, the derived equation turns in the following Equation 2.2.

$$\frac{\partial}{\partial t}n_i(v,t) + \frac{\partial}{\partial v}[n_i(v,t)G_i(v)] = 0 \qquad 2.2$$

Two-dimensional PBM considers as internal coordinates in the gist of the framework, granule size (v) in volume, and porosity ε.2D-PBM is an extension of the 1D-PBM where Consolidation phenomenon related to porosity is added to the one-dimensional model. Accurate population balance modelling (PBM) hinges on selecting the right rate process mechanism kernels. This requires a deep, experimentally validated understanding of the system, enabling one to refine existing kernels or create new ones for precise representation.

2.3.1 Twin Screw Wet Granulation Modelling

This involves developing mathematical models that capture the complex physical and chemical phenomena occurring within the granulator shown in Figure 2.5, including mass and heat transfer, fluid flow, particle interactions, and chemical reactions.

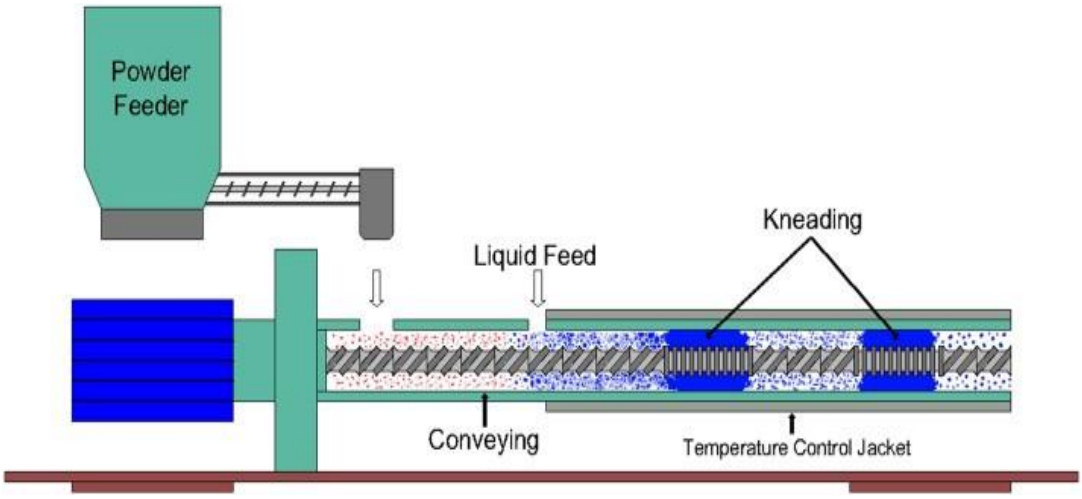


Figure 2.5 Diagram of typical TSWG unit (Seem et al., 2015).

Figure 2.5 illustrates a schematic diagram of a twin-screw wet granulator (TSWG) unit. The TSWG consists of various components, including powder feed ports for the introduction of dry granules, liquid addition ports for the

introduction of liquid binders, a temperature control jacket to maintain optimal processing conditions, and conveying and kneading elements to transport and mix the materials.

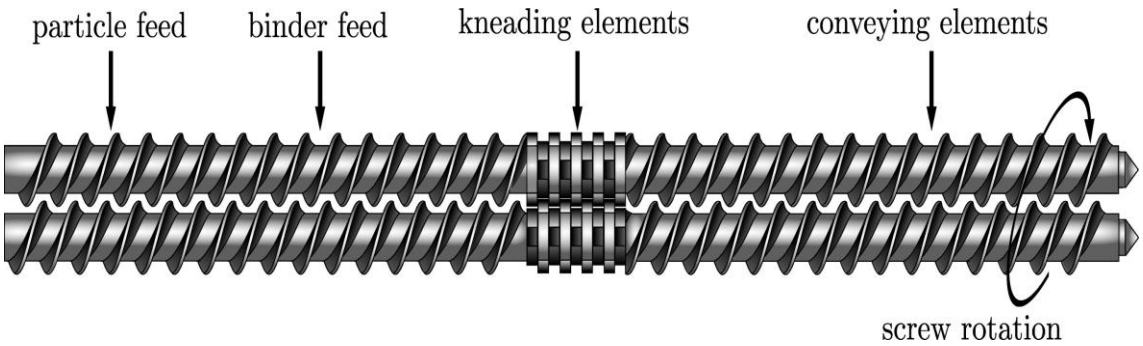


Figure 2.6 Diagram of TSG screw elements (McGuire et al., 2018).

Figure 2.6 illustrates a diagram of the screw elements and a theoretical screw configuration suitable for industrial TSG applications. The versatility of screw configurations is limitless, allowing for customisation to suit various pharmaceutical applications. By carefully designing the screw profile, including the pitch, helix angle, and kneading block design, it is possible to optimise the granulation process for specific product requirements.

Several population balance models (PBMs) have been developed to simulate twin-screw wet granulation processes, including those by Iveson, Litster, and Ermis, Cameron et al., Barrasso et al., and Wang et al. Wang et al. proposed a compartmental model to investigate the impact of screw configuration on granule properties, highlighting its major influence on particle size distribution (Wang et al., 2021). This emphasises the importance of incorporating screw configuration into models to enable optimisation.

To accurately model pharmaceutical granulation processes, it is essential to consider multidimensional factors. Shirazian et al. developed a 2D population balance model that considered particle size and liquid composition. Their model focused on the key mechanisms of coalescence and breakage, demonstrating that increasing screw speed leads to more uniform liquid distribution (Shirazian et al.,2019). Additionally, they observed that aggregation is more prevalent in conveying zones, while breakage dominates in kneading zones at higher liquid-to-solid ratios. However, as Pinto et al. noted, reducing the dimensionality of the

PBM can simplify the model and improve computational efficiency (Pinto et al.,2007). To achieve near-real-time simulation, the use of reduced-order modelling techniques, such as aggregating granule dimensions, is recommended. Hybrid PBM-DEM models have been employed to enhance model accuracy while maintaining computational efficiency (Barrasso, Tamrakar, and Ramachandran, 2015). Additionally, semi-empirical hybrid models that incorporate material properties into process models have been developed to investigate their impact on residence time distribution and process parameters for various powder processing unit operations (Toson et al., 2018; Bostijn et al., 2019).

2.3.2 Fluidised Ded Dryer Modelling

In a fluidised bed dryer, wet granules are fluidised by channelling high-pressure hot gas through a perforated bed. This process suspends the granules in the hot gas stream, as depicted in Figure 2.7. The granules' moisture content vaporises, and the resulting moisture is carried away by the hot gas. The inlet temperature and flow of the dry gas are crucial factors in optimising energy usage in an FBD. By carefully adjusting these parameters, drying time and hot gas temperatures can be minimised, leading to energy savings while maintaining the desired product moisture content.

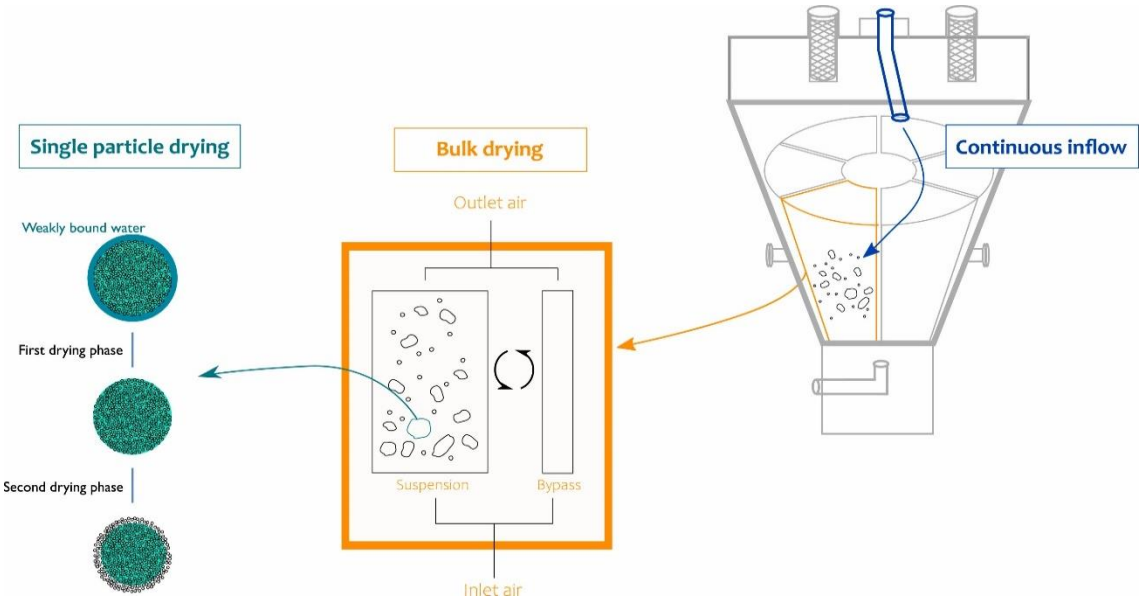


Figure 2.7 Schematic representation of Mechanistic Model of the ConsiGma Fluidised Bed Dryer (Vandeputte et al., 2023).

Numerous studies have delved into the mechanistic aspects of drying wet, porous granules in fluidised bed dryers. A primary focus of these investigations has been to accurately determine the moisture content of granules in real-time using inline near-infrared (NIR) spectroscopy, as demonstrated by Chablani et al. in their work on the C25 dryer (Chablani et al.,2011).

More recently, Stauffer et al. shed light on the critical role of wet granule properties in ensuring stable drying processes. They emphasised that excessive accumulation of fine particles on the drier filter surface, resulting from suboptimal initial granule properties, can destabilise the drying process (Stauffer et al., 2019). De Leersnyder et al. conducted a comprehensive study to understand how various drying process parameters impact the quality attributes and breakage behaviour of granules. Their research provides valuable insights into optimising drying conditions to maintain granule integrity (De Leersnyder et al., 2018). Building upon this knowledge, Ryckaert et al. investigated the breakage and attrition mechanisms that granules undergo during different stages of the drying process in a dryer, including transport, filling, and drying (Ryckaert et al., 2021). Their work offers a deeper understanding of the physical forces that can lead to granule degradation (Ryckaert et al., 2021).

When examining average drying times for different cells targeting the same final moisture content, Grelier et al observed that cells 5 and 6 consistently achieved the target moisture content faster, cells 2 and 3 were the slowest, and cells 1 and 4 exhibited intermediate drying times, as depicted in Figure 2.8 (Grelier et al., 2022). These findings suggest that cells operating under identical process parameters experience variations in heat distribution.

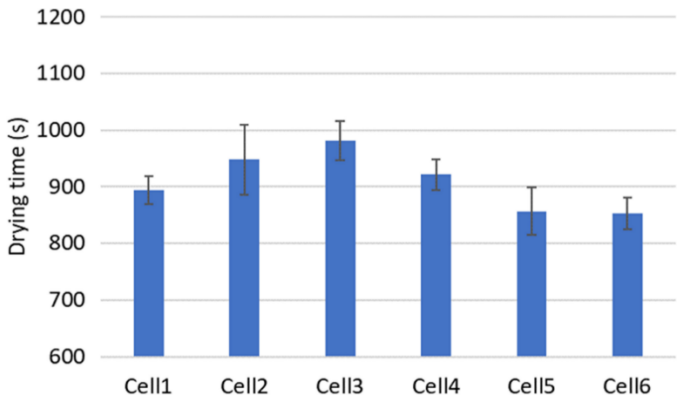


Figure 2.8 Drying Time Comparison for Different Dryer Cells (Grelier et al., 2022).

2.3.3 Milling Processes

Milling is a ubiquitous operation in pharmaceutical manufacturing, employed at various stages of the production process. For instance, it is used to break down wet granules into smaller particles (Akkisetty et al., 2010; Vanarase and Muzzio, 2011). While several milling devices exist, conical screen mills are widely adopted in the pharmaceutical industry. These mills feature an impeller that rotates above a perforated steel screen, effectively breaking down particles and allowing smaller particles to pass through the screen (Teżyk et al., 2016).

The inherent continuous nature of the milling process makes it highly compatible with continuous manufacturing lines. This seamless integration is possible because milling can serve as a bridge between discrete unit operations. Consequently, milling can be readily incorporated into continuous production lines, enhancing overall process efficiency. The size reduction of pharmaceutical compacts that occurs during continuous milling can be quantitatively modelled using the continuous PBM. This mathematical framework allows for the analysis of mass balance for particles of specific sizes, providing valuable insights into the milling process and its impact on particle size distribution. Assuming the milling process occurs in a well-mixed environment, the continuous PBM can be expressed as an integrodifferential equation. This equation quantifies the change in mass distribution over time with respect to particle size (Amini et al., 2020).

Barrasso et al. investigated the performance of a conical screen mill operating under continuous choke conditions, a common configuration in continuous manufacturing lines. They explored the relationship between various granulate properties, such as ribbon density, impeller speed, feed rate, and screen size, and the residence time of the material within the milling cone (Barrasso et al., 2013). By conducting a series of experiments, they developed a predictive PBM capable of accurately predicting the PSD.

2.3.4 Tablet Press Modelling

Tablet pressing is a multi-step process as shown in Figure 2.9 that involves filling the die with powder, pre-compressing the powder, applying the main compression force to form the tablet, and ejecting the tablet from the die .The

weight of the tablet is primarily controlled by adjusting the dosing position of the powder feeder. However, factors such as variations in powder bulk density and filling time can influence the final tablet weight (Su et al., 2019). The minimum tablet thickness is determined by the punch displacement, which is manually set.

Crucially, the maximum main compression force is directly correlated to the amount of powder present in the die, which is equivalent to the tablet weight. This interdependence between the pre-compression and main compression stages creates a closed-loop system (Su et al., 2019).

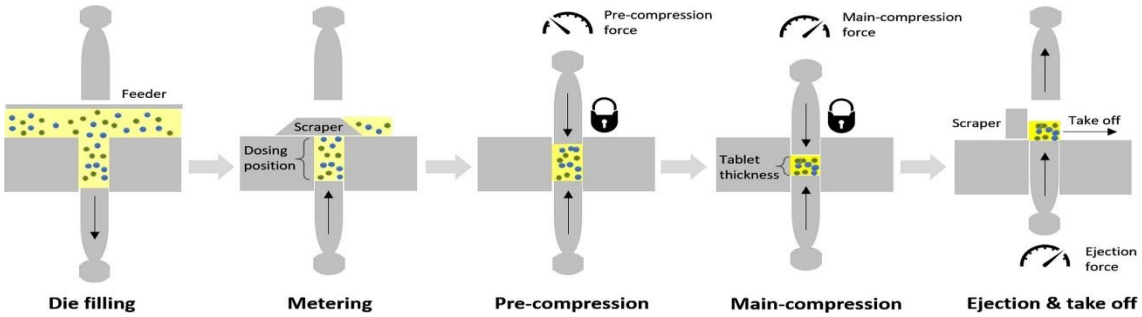


Figure 2.9 A Breakdown of the Major Steps a Rotary Tablet Press (Su et al., 2019).

Gavi and Reynolds have developed an extensive model of the roll compaction tablet manufacturing process, which integrates various unit operations, including the tablet press, to explore the intricate relationships between CPPs and product performance (Gavi and Reynolds, 2014). Their work demonstrates that a mechanistic model for tablet manufacturing can be developed with a substantial reduction in experimental calibration compared to traditional statistical approaches., enabling quantitative design space exploration even with limited or unavailable experimental materials, particularly in early development. To assess tablet performance, the model incorporates models for tensile strength model and disintegration rate model.

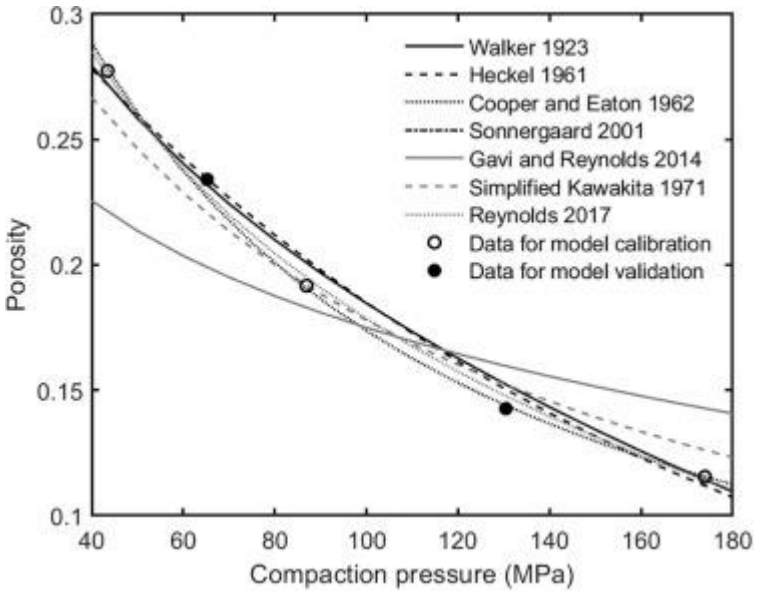


Figure 2.10 Model Calibration and Validation: Tablet Porosity (Wang et al., 2021).

Tablet tensile strength and porosity are two crucial parameters that are significantly influenced by the tablet pressing process. Figure 2.10 illustrates the performance of various models used to calibrate and validate tablet porosity predictions using on experimental data. Among the models evaluated, the Sonnergaard (2001) model consistently outperforms others, such as the Cooper and Eaton and Kawakita models, in terms of fitting quality for both calibration and validation datasets. While the Gavi and Reynolds (2014) model tends to underestimate porosity, the simplified one-parameter Kawakita model is less accurate than two-parameter models. The three-parameter Sonnergaard model exhibits slightly better performance than the four-parameter Cooper and Eaton model, highlighting the importance of selecting appropriate fitting parameters. The Reynolds et al. (2017) model, with its simple form and good fit, offers a practical choice for two-parameter porosity models. The Reynolds et al. (2017) porosity model is defined by the two curial parameters namely, compressibility constant and pressure at zero porosity. The compressibility constant quantifies how much a material's volume changes in response to a change in pressure. Pressure at zero-porosity" refers to the pressure that would be required to achieve this perfectly dense state. This is often extrapolated from experimental data, as truly achieving zero porosity can be very difficult. Overall, two-parameter models demonstrate sufficient accuracy when compared to experimental data, as depicted in Figure 2.10.

While Gavi and Reynolds's model provides a valuable framework for understanding the relationship between process parameters and tablet tensile strength, additional studies have further enriched our knowledge in this area. AI-driven models, such as neural networks and machine learning algorithms, have been employed to predict tablet tensile strength based on a vast array of process parameters and material properties (Wang et al., 2020). These models often outperform traditional statistical methods in terms of accuracy and predictive power. The type and concentration of binders used in tablet formulation significantly affect tensile strength. Studies have investigated the interactions between different binders and the matrix formed during compaction, shedding light on the mechanisms by which binders contribute to tablet integrity (Liu et al., 2022). Disintegrants, while essential for tablet dissolution, can also influence tensile strength. The distribution and interaction of disintegrants within the tablet matrix can affect its mechanical properties. DoE techniques have been widely used to optimise tablet manufacturing processes with respect to tensile strength (Wu et al., 2021). By systematically varying process parameters and analysing the resulting tablet properties, researchers can identify optimal conditions for achieving desired tensile strength while minimising variability. Advancements in process monitoring technologies, such as near-infrared spectroscopy and image analysis, have enabled real-time assessment of tablet properties during manufacturing. This information can be used to implement closed-loop control systems and ensure consistent tablet quality. The literature on tablet tensile strength continues to evolve, with new studies exploring advanced modelling techniques, the role of excipients, and process optimisation strategies. By incorporating these findings into tablet manufacturing practices, pharmaceutical companies can improve product quality, reduce variability, and enhance patient safety.

2.3.5 Literature Gaps and Opportunities

Process rate mechanism kernels have become a cornerstone in modelling various unit operations within the pharmaceutical manufacturing process, including:

- **Twin-Screw Granulators:** Kernels can simulate granule growth, breakage, and densification during the wet granulation process.
- **Driers:** Models can represent moisture evaporation and particle interactions during drying.
- **Mills:** These kernels capture particle size reduction mechanisms like grinding and breakage.
- **Blenders:** Mixing behaviour and potential segregation issues can be addressed through kernel-based models.
- **Tablet Presses:** Compaction and densification processes during tablet formation can be modelled using appropriate kernels.

Strengths of Kernel-Based Models:

Limitations of Mechanistic Models:

- **Fundamental Principles**: Process rate mechanism kernels rely on fundamental physical and chemical principles governing particle behaviour. This provides a deep understanding of the underlying mechanisms within each unit operation.
- **Particle Attribute Dependency**: Advanced kernels can account for the influence of particle size, shape, and other attributes on the process rate. This allows for more accurate simulations compared to simpler models.
- **Commercial Software Availability**: Several commercial software packages, such as gPROMS Formulated Products and ASPEN, offer pre-built kernels for various processes.
- **Data-Driven Adaptation**: Specific model formulations (kernels) exist for different rate processes, but a systematic approach for selecting the most appropriate one for a given application and design space is currently lacking. Here, data-driven approaches that can incorporate real-time or historical data offer an advantage. As new data becomes available, the model can be updated and continuously improved through techniques like machine learning.

- **Computational Efficiency:** The detailed nature of mechanistic models can lead to computational hurdles, making them potentially sluggish for

real-time applications within digital twins or continuous manufacturing processes.

- **Calibration Complexity:** The intricate nature of these models can make calibration a challenging task, requiring significant expertise and potentially extensive experimental data.
- **Maintenance Challenges:** Maintaining and updating complex mechanistic models can be resource-intensive, especially in rapidly evolving environments.

For each rate process, different kernels/models have been formulated over the years, however, there is no systemic way of selecting which to use for given applications and design space. For adaptive modelling capability, the modelling approaches that can incorporate data are advantageous as new data can be used to update the models.

While mechanistic models may not be ideal for direct implementation in certain digital applications due to their computational and maintenance demands, their value lies in process design and optimisation. They offer a deep understanding of fundamental mechanisms and can be used to develop simpler surrogate models or inform the selection of appropriate kernels for specific applications. This highlights the importance of a tiered modelling approach, where complex mechanistic models are used for initial design and optimisation, and then simpler, data-driven models are derived for real-time control within digital twins.

2.4 Systems Engineering Approach

The application of a systems engineering approach has become crucial in addressing the complex challenges within this highly regulated sector. The key concepts of systems engineering include interdisciplinary Integration which promotes the integration of various disciplines, such as chemical engineering, mechanical engineering, and information technology, to create a holistic understanding of pharmaceutical manufacturing processes. A fundamental aspect of systems engineering is requirements engineering. This involves capturing and managing requirements at different stages of the pharmaceutical

product lifecycle, ensuring alignment with regulatory standards and market demands.

2.4.1 The V-model

To effectively navigate the intricate landscape of product development, manufacturing processes, and supply chain management, this research adopted a system engineering approach, as illustrated in Figure 2.11. This structured framework offers a holistic perspective by considering all interconnected aspects of the system, from initial design to final production and delivery. The V-model in systems engineering provides a structured framework, visually represented as a "V," that emphasises both development and rigorous testing throughout the project lifecycle. Beginning with the left side of the "V," the process initiates with the translation of user needs into detailed system requirements, followed by architectural and module design, culminating in coding and unit implementation. This decomposition phase is mirrored on the right side of the "V" by a corresponding verification and validation process, starting with unit testing and progressing through integration, system, and acceptance testing. This systematic approach ensures traceability between each development stage and its associated testing phase, facilitating early defect detection and correction. While the V-model's rigid structure may not suit projects with rapidly changing requirements, it excels in scenarios requiring well-defined specifications, safety-critical systems, and strict regulatory compliance, ultimately contributing to improved system quality through its focus on comprehensive verification and validation.

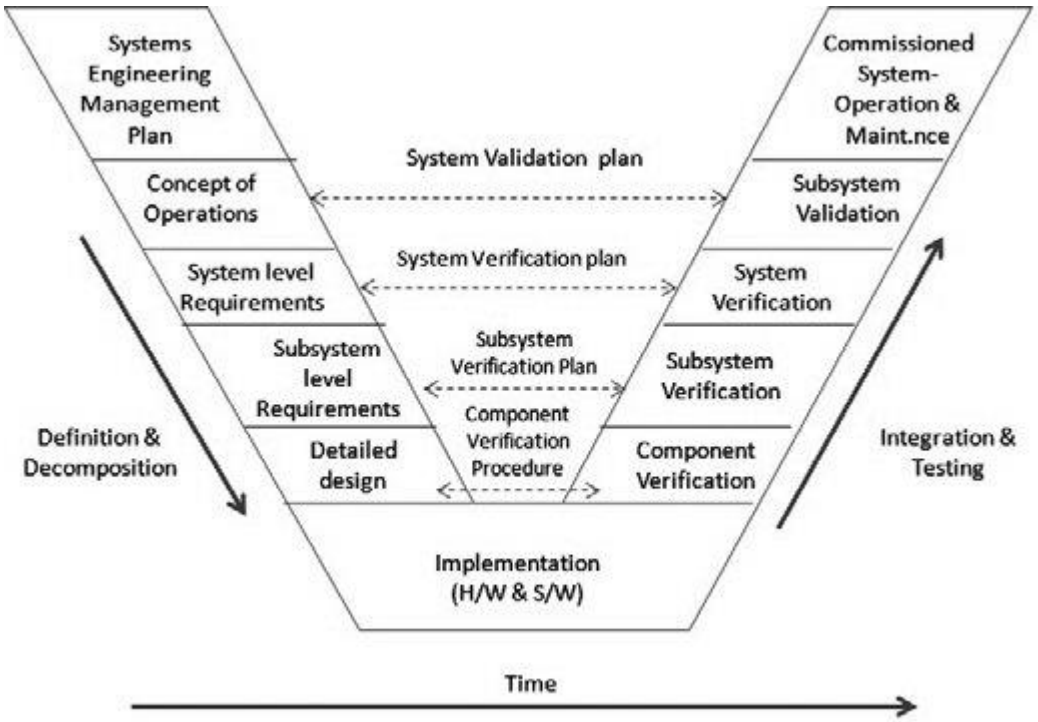


Figure 2.11 The "V" Model of systems engineering (Kumar et al., 2016).

Systems engineering emphasises risk management throughout the product development lifecycle as illustrated in Figure 2.12 (Markl et al., 2020). This includes identifying and mitigating risks related to process variability, raw material quality, and regulatory compliance. Systems engineering provides a framework for effective life cycle management of pharmaceutical processes. This involves continuous improvement, adaptation to changing regulatory requirements, and integration of modern technologies.

In the context of pharmaceutical manufacturing, the V-model proves invaluable for the structured development of digital twins. The initial stages, mirroring the left side of the "V," involve meticulously defining user requirements, such as predicting product quality or optimising process parameters, and translating these into detailed system specifications for the digital twin. Architectural design then focuses on defining the twin's components, including data acquisition, model development, and simulation capabilities. As development progresses, the right side of the "V" ensures rigorous validation: unit testing verifies individual model components, integration testing checks data flow and model interactions, system testing evaluates the digital twin against real-world scenarios, and acceptance testing confirms its ability to meet specific pharmaceutical manufacturing objectives, like predicting critical quality attributes or enabling process

optimization. This structured approach ensures traceability, facilitates early identification of discrepancies between the digital twin and the physical process, and supports the development of robust and reliable digital twins for enhanced pharmaceutical manufacturing efficiency and quality control.

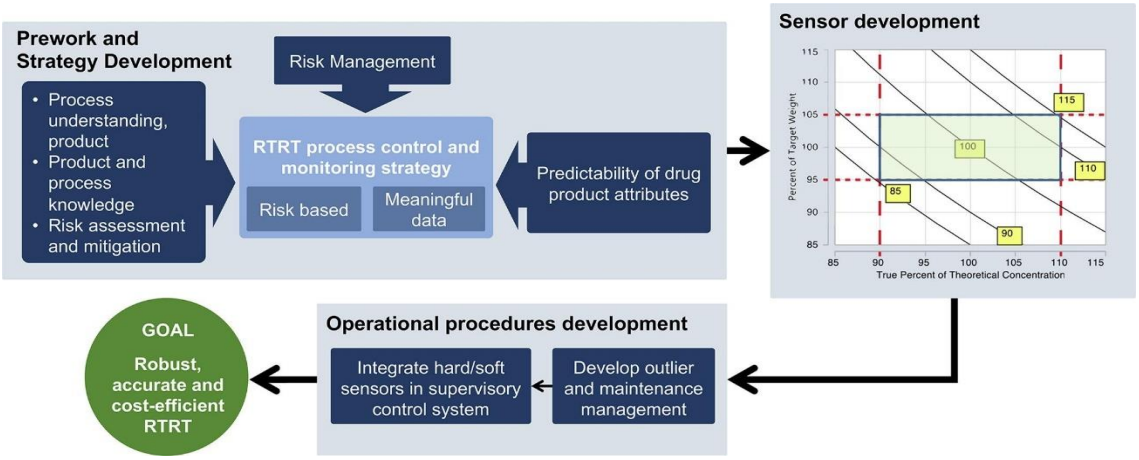


Figure 2.12 Monitoring Real-Time Drug Release from Pharmaceutical Tablets (Markl et al., 2020).

2.4.2 Process Systems Engineering techniques.

QbD principles, derived from systems engineering, have gained prominence in the pharmaceutical industry (Rathore, A.S., 2014). QbD is defined in the ICH Q8 guideline as ‘a systematic approach to development that begins with predefined objectives and emphasises product and process understanding and process control, based on sound science and quality risk management (Guidance for Industry, 2009). The approach focuses on designing quality into the product and process, results to greater consistency and reliable manufacturing outcomes. PAT, aligned with systems engineering, involves the real-time monitoring and control of pharmaceutical processes. This enhances process understanding, facilitates timely decision-making, and contributes to the production of high-quality pharmaceuticals. Systems engineering principles are applied to optimise pharmaceutical supply chains. This involves the integration of suppliers, manufacturers, and distributors to enhance overall efficiency and reduce time-to-market. The use of modelling and simulation tools allows pharmaceutical engineers to analyse and optimise manufacturing processes. Dynamic models help in predicting system behaviour, identifying potential bottlenecks, and optimising production throughput. Flowsheet modelling is a powerful tool that can be used to integrate and simulate the entire manufacturing process. By

representing the interconnected units and operations, this approach enables the prediction of process dynamics, operational costs, and product performance based on the properties of the raw materials and the specific conditions under which each unit operates.

System energy engineering, flow sheeting, and reverse engineering are interconnected disciplines that play crucial roles in optimising energy systems. Reverse engineering illustrated in Figure 2.13 is the process of analysing a system to understand its components and how they interact, often with the goal of improving or replicating it. By combining these approaches, engineers can identify energy inefficiencies, develop innovative solutions, and design sustainable energy systems that meet the needs of a growing and evolving world.

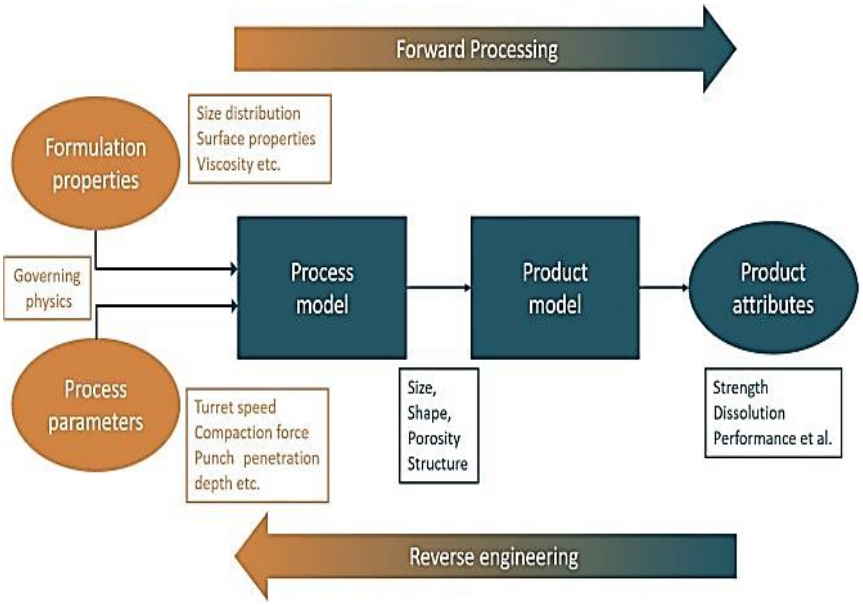


Figure 2.13 Formulated Product Reverse Engineering: A Process of Inferring Formulation Properties from Product Attributes (Litster J.,2019).

System engineering approaches includes:

- Flowsheet modelling
- Sensitivity analysis
- Design space identification.
- System optimisation.

By analysing a process, one can uncover its strengths, weaknesses, and bottlenecks, which is crucial for designing effective control strategies across the entire manufacturing cycle using these approaches.

Process Systems Engineering (PSE) techniques, such as equation-based simulation, flexibility analysis, and constrained global optimisation, have been instrumental in designing robust and cost-effective manufacturing processes across various industries. By combining feasibility and flexibility analyses, algorithms can identify optimal process design alternatives that are both resilient and economical.

2.4.2.1 Sensitivity analysis

Sensitivity analysis, a powerful tool in process engineering, helps pinpoint the process parameters that have the most significant impact on product quality. By understanding these critical parameters, engineers can develop effective control strategies to ensure consistent product quality and minimise deviations. Variance-based and Morris' method are two common techniques for global sensitivity analysis. The Variance-Based Method utilises Monte Carlo simulation techniques to quantify the sensitivity of a model's output to variations in its input parameters. These sensitivity indices, often represented as Sobol indices, can be visualised in heat maps as shown in Figure 2.14 (Metta et al, 2019), providing a clear and intuitive representation of the relative importance of each input factor. By analysing these heat maps, researchers and engineers can identify the most influential parameters and prioritise efforts for model calibration, uncertainty quantification, and optimisation.

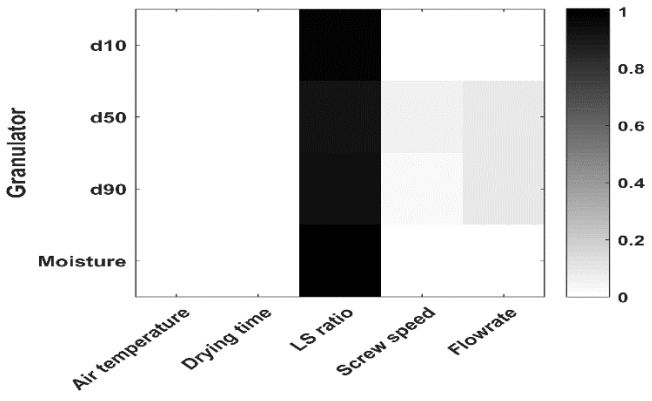


Figure 2.14 Assessing the Sensitivity of Granulation Process Parameters (Metta et al, 2019).

Table 2.2 Summary of past works on systems engineering applications in pharmaceutical industry.

Author(s)	*Title*	*Journal/Book*	*Year*
Calhan, S.D., Eker, E.D. and Sahin, N.O.	Quality by design (QBD) and Process Analytical Technology (PAT) applications in pharmaceutical industry'	European Journal of Chemistry	2017
International Council for Harmonisation	Ich harmonised tripartite guideline	International conference on harmonisation of technical Requirements for registration of pharmaceuticals for human use	2009
Swami, A. et al..	A critical review on recently used PAT in pharmaceutical industry	Research Journal of Pharmacy and Technology	2023
Food and Drug Administration (FDA)	Risk Management Guidance for Pharmaceutical Development	FDA report	2020

2.4.3 Literature Gaps and Opportunities

The traditional approach to pharmaceutical development and manufacturing often involves compartmentalised processes, leading to inefficiencies and potential challenges. To address this, the adoption of a systems engineering approach offers a significant advantage. The adoption of a process systems engineering approach in the pharmaceutical manufacturing industry offers a structured framework to address the complexities associated with product development, manufacturing processes, and supply chain management. The integration of interdisciplinary knowledge, coupled with advanced modelling and risk management techniques, contributes to enhanced product quality, regulatory compliance, and overall efficiency in the pharmaceutical sector.

Systems engineering provides a structured framework that considers the entire pharmaceutical value chain, from product development and process design to manufacturing execution and supply chain management. This holistic view allows for the identification and optimisation of interdependencies between these stages, leading to a more streamlined and efficient overall process.

A core tenet of systems engineering is the integration of interdisciplinary knowledge. This fosters collaboration between various teams, including:

- Research and Development
- Process Engineering
- Quality Assurance
- Regulatory Affairs

Systems engineering emphasises the use of advanced modelling techniques to simulate and predict process behaviour. This allows for:

- **Virtual process optimisation**: Identifying and refining processes virtually before implementation, minimising costs and risks associated with physical experimentation.
- **Proactive Risk Management**: Systems engineering frameworks encourage the identification and mitigation of potential risks early in the development phase, reducing the likelihood of costly delays or product recalls.

By implementing a system engineering approach, pharmaceutical companies can expect to achieve improved process control and risk management contribute to a higher level of product quality and consistency. Early integration of regulatory considerations reduces the risk of non-compliance issues later in the development process. A holistic approach fosters process optimisation, leading to increased production efficiency and reduced costs. Faster identification and mitigation of risks, coupled with efficient process design, can accelerate the time it takes to bring new drugs to market. Overall, a systems engineering approach offers a powerful framework for optimising the entire pharmaceutical manufacturing ecosystem, leading to significant benefits for patients, manufacturers, and regulators alike.

2.5 Digital Twin Background

Industry 4.0, “the current wave of technological advancement in manufacturing, is characterised by the integration of cyber-physical systems, the Internet of Things, cloud computing, and cognitive computing to create smart factories” (Industry 4.0 and the fourth industrial revolution explained, 2006). This revolution

goes beyond mere automation; it involves the seamless integration of digital technologies into production processes, enabling machines to generate data autonomously and make informed decisions. A prime example of this is predictive maintenance, where machines can anticipate potential failures and schedule maintenance proactively (as illustrated in Figure 2.15). This data-driven approach not only streamlines production but also enhances efficiency, reduces downtime, and improves overall competitiveness.

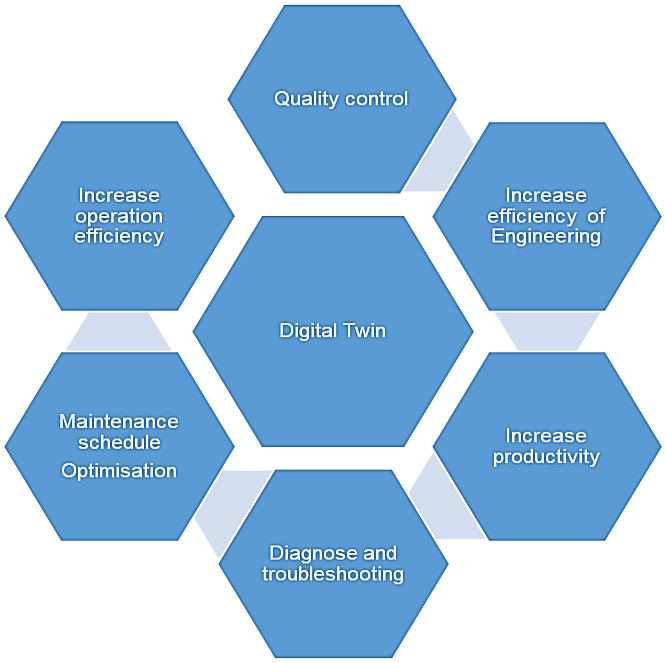


Figure 2.15 Digital twin typical applications.

According to Ślusarczyk, the primary objectives of adopting I4.0 are to enhance operational efficiency, effectiveness, and automation. Nakagawa et al., 2021, highlight that the rigid boundaries between the different layers of the industrial automation pyramid have become less distinct in response to the increased flexibility demanded by I4.0. This has led to the emergence of a more interconnected and decentralised network, characterised by interoperability among various components.

Big data, with its capacity to capture, store, and analyse vast datasets, forms the bedrock of a novel simulation approach in the era of I4.0 (Roblek et al., 2016; Xu and Li, 2018; Cimino et al, 2019). Researchers are actively looks for ways to enhance data visualisation techniques. Han et al. have developed a framework for visualising real-time 3D models using VR/AR. At the University of Sheffield, a web-based data visualisation tool has been created to offer an integrated solution

for data analytics and visualisation. This tool empowers scientists, operators, and managers to gain deeper insights into their data by providing a unified platform for analysis and visualisation (Allen et al., 2021).

Recent advancements in Industry 4.0 across various process industries hold the potential to significantly enhance the capabilities of existing digital twins in the pharmaceutical industry. As outlined in Table 2.3, these technologies have been successfully applied in diverse sectors. Leading companies like GE, PTC, Siemens, ANSYS, and Dassault have pioneered the use of digital twins, leveraging historical data on process parameters, power consumption, and equipment wear and tear to make real-time recommendations for optimising operational costs. This thesis aims to build upon these advancements by incorporating historical data, sustainability considerations, and product performance metrics into the digital twin framework. By doing so, this would enable real-time optimisation of pharmaceutical manufacturing processes, leading to improved efficiency, reduced costs, and enhanced product quality.

Table 2.3 The Application of Digital Twins in Diverse Industries (Chen et al., 2020 ; Cimino et al., 2019).

Industry	*Service*	*Technology used*
Civil engineering	Dassault created a Digital Twin of Singapore with the 3D Experience Platform for urban planning and construction.	3D Experience
Health care	Dassault developed a human heart digital twin	Virtual model using finite element-based modelling
Marine	DNV GL created a virtual ship to reduce operational viability, and therefore improving safety across the lifecycle of the physical ship.	
Energy	GE created a virtual wind farm to enhance maintenance planning, dependability, and energy output.	Virtual model component based on Predix

2.6 Digital Twin architecture; Studies on digital twin tools

Significant advancements in digital twin technology have emerged, including the development of hybrid models, the integration of soft sensors, and the utilisation of commercial IoT platforms like Insight hub and Amazon Web Services. These platforms empower users to visualise, analyse, and manage data on cloud servers.

Qi et al. proposes a comprehensive five-dimensional digital twin model to facilitate the widespread adoption of digital twins (Qi et al.,2021):

$$M_{DT} = (pe, vm, ss, dd, cn) \quad 2.3$$

Where pe are physical entities, vm are virtual models, ss are services, dd is Digital twin data, and cn are connections.

- **Physical Entities:** The real-world objects or systems that the digital twin represents.
- **Virtual Models:** The digital representations of the physical entities, capturing their properties, behaviour, and interactions.
- **Services:** The functionalities and capabilities provided by the digital twin, such as simulation, optimisation, and predictive maintenance.
- **Digital Twin Data:** The data generated, collected, and processed by the digital twin, including sensor data, simulation results, and historical records.
- **Connections:** The communication channels and interfaces that enable the integration of the physical entities, virtual models, services, and data.

This model expands upon Grieves' original digital twin model by incorporating the dimensions of services and connections, further enhancing its versatility and applicability. In the following sections, this review will delve into each aspect of this new digital twin model.

2.6.1 Studies on Digital Twin Tools Table of References

The provided references shed light on the current state and future trends of data-driven modelling, PAT, and IoT in pharmaceutical manufacturing. These technologies are shaping the industry's future by enabling data-driven decision-making, process optimisation, and real-time monitoring. The references also highlight the importance of data reliability, regulatory compliance, and overall process integrity, emphasising the need for robust data governance practice.

Table 2.4 Studies on digital twin tools – References list

Author(s)	*Title*	*Journal/Book*	*Year*
	Data driven modelling and Process Analytical technology		
Dong Y. et al.	Data-driven modelling methods and techniques for pharmaceutical processes	The AAPS Journal	2023
Dorota-Owczarek	"Applications of Machine Learning in Drug Manufacturing"	Nexocode	2019
Calhan S.D., Eker E.D. and Sahin N.O.	"Process Analytical Technology: A Comprehensive Review"	European Journal of Chemistry	2017
Sacher S. et al.	PAT implementation for advanced process control in solid dosage manufacturing – A practical guide	International Journal of Pharmaceutics	2022
Nagy B. et al.	Application of artificial neural networks in the process analytical technology of Pharmaceutical Manufacturing—A Review'	The AAPS Journal	2022
	Modelling and Simulation		
Ghahramani, P.	Modelling and simulation applications in drug development process	Translational Medicine	2016
Zilong W.	Simulation-based process analysis and optimization	Rucore	2018
Jagtap, K., Chaudhari, B. and Redasani, V.	Quality by Design (QBD) concept review in pharmaceuticals	Asian Journal of Research in Chemistry	2022
Schiemer R. et al.	An adaptive soft-sensor for advanced real-time monitoring of an antibody-drug conjugation reaction	Journal of Biotechnology and Bioengineering	2023
	IoT Applications in Pharmaceutical Manufacturing		
Bassolillo S.R. et al.	" Decentralized Mesh-based model predictive control for swarms of uavs', Sensors	Sensors	2020

Chen et al.	Internet of things (iot)—blockchain-enabled pharmaceutical supply chain resilience in the post-pandemic era	Frontiers of Engineering Management	2022
M. Pachayappan, N. Rajesh, G. Saravanan,	Smart logistics for pharmaceutical industry based on Internet of Things (IoT)	International Journal of Computer Science and Information Security	2016
A. K. Pundir, J. D. Jagannath, and L. Ganapathy	" IMPROVING SUPPLY CHAIN VISIBILITY USING IoT-INTERNET OF THINGS	Annual Computing and Communication Workshop and Conference	2019
Data Management and Data Contamination in Pharmaceutical Manufacturing Industry			
Charoo N.A., Khan M.A. and Rahman Z.	Data Integrity Issues in pharmaceutical industry: Common observations, challenges, and mitigations strategies	International Journal of Pharmaceutics	2023
Kakad S.B., Kolhe M.H., Dukre TP	A Review on Pharmaceutical Validation	International Journal of Pharmaceutical Quality Assurance	2020
Par, C. and Leeds M.	A highly efficient robust design under data contamination	, Computers & Industrial Engineering	2016
Chen Y. et al.	An integrated data management and Informatics Framework for Continuous Drug Product Manufacturing Processes: A case study on two pilot plants	International Journal of Pharmaceutics	2023

2.6.2 Process Modelling and Simulators

Process modelling and simulation play a pivotal role and essential element of DT in the pharmaceutical manufacturing industry, providing tools to design, analyse, and optimise complex processes. This literature review examines the key principles, applications, and advancements in process modelling and simulators within the pharmaceutical sector.

Process modelling involves the creation of mathematical representations of pharmaceutical manufacturing processes. This includes understanding and describing the behaviour of various unit operations, reaction kinetics, and mass transfer phenomena. In pharmaceutical manufacturing, process modelling aids in the design of efficient and scalable processes. It allows for the prediction of critical quality attributes, the identification of optimal operating conditions, and the exploration of alternative manufacturing scenarios. Beyond the traditional particle-based approach, gPROMS FormulatedProducts (gFP) has been integrated with multi-physics simulations to consider factors like heat transfer, moisture diffusion, and chemical reactions within the granulator. This allows for a more comprehensive understanding of the granulation process and its sensitivity to process parameters. gFP has been combined with DoE to efficiently explore the design space of twin-screw granulation and identify optimal process conditions. By simulating various combinations of process parameters, researchers can optimise for factors like granule size distribution, yield, and energy consumption (Arthur et al., 2023).

Process modelling aligns with the principles of QbD, emphasising the systematic development of processes to ensure product quality. QbD-driven models contribute to a deeper understanding of the relationships between process parameters and product attributes. Simulation tools enable the dynamic representation of pharmaceutical processes, providing a virtual environment for testing and optimisation. Simulators facilitate real-time process monitoring and control, allowing pharmaceutical manufacturers to respond promptly to deviations and optimise manufacturing operations. This capability is crucial for achieving consistent product quality. Simulation-based training programs enhance the skills of pharmaceutical operators and engineers. Virtual environments enable hands-

on experience in dealing with complex manufacturing scenarios, reducing the learning curve for new processes. The literature reviewed underscores the diverse applications of process modelling, its integration with QbD principles, and the role of simulators in real-time monitoring, control, and operator training.

Advances in process understanding are further facilitated by first principle and hybrid modelling of the system. Hybrid modelling approaches, which combine the strengths of DNN and mechanistic models, offer a promising solution for simulating complex pharmaceutical production processes while mitigating computational costs and data scarcity. Gargalo et al. presented a hybrid model structure (Figure 2.16) that integrates these two modelling paradigms. Nielsen et al. (2020) demonstrated that hybrid models can capture process dynamics more accurately than traditional mechanistic models, making them a valuable tool for pharmaceutical process understanding and optimisation (Gargalo et al., 2020).

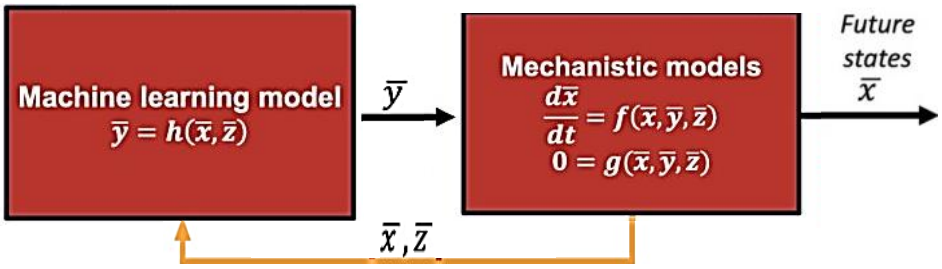


Figure 2.16. Combining the Best of Both Worlds: A Hybrid Machine Learning and Mechanistic Model (Gargalo et al., 2020).

2.6.3 Data Driven Modelling and PAT

In the dynamic landscape of pharmaceutical manufacturing, the integration of advanced technologies such as data-driven modelling and PAT has become imperative. This season explores the key concepts, applications, and impact of these approaches on enhancing efficiency, quality, and regulatory compliance in the pharmaceutical manufacturing sector. Data-driven modelling involves the use of data analytics, machine learning, and statistical techniques to extract valuable insights from process data. In pharmaceutical manufacturing, this approach enables the identification of patterns, correlations, and predictive models for improved decision-making.

Data-driven modelling is applied in various aspects of pharmaceutical manufacturing, including process optimisation, quality prediction, and predictive maintenance. It facilitates real-time monitoring and control, leading to increased process robustness and product consistency. Challenges associated with data quality, integration, and interpretation are addressed through advanced data-driven techniques. Opportunities for innovation arise as pharmaceutical companies leverage big data to gain deeper insights into their manufacturing processes.

PAT involves the systematic application of process analytics to monitor, control, and optimise pharmaceutical manufacturing processes. It aims to ensure the quality of the final product by providing real-time information and facilitating continuous improvement. PAT tools, including spectroscopy, chromatography, and sensors, are integrated into manufacturing processes to enable real-time analysis as illustrated in Figure 2.17.

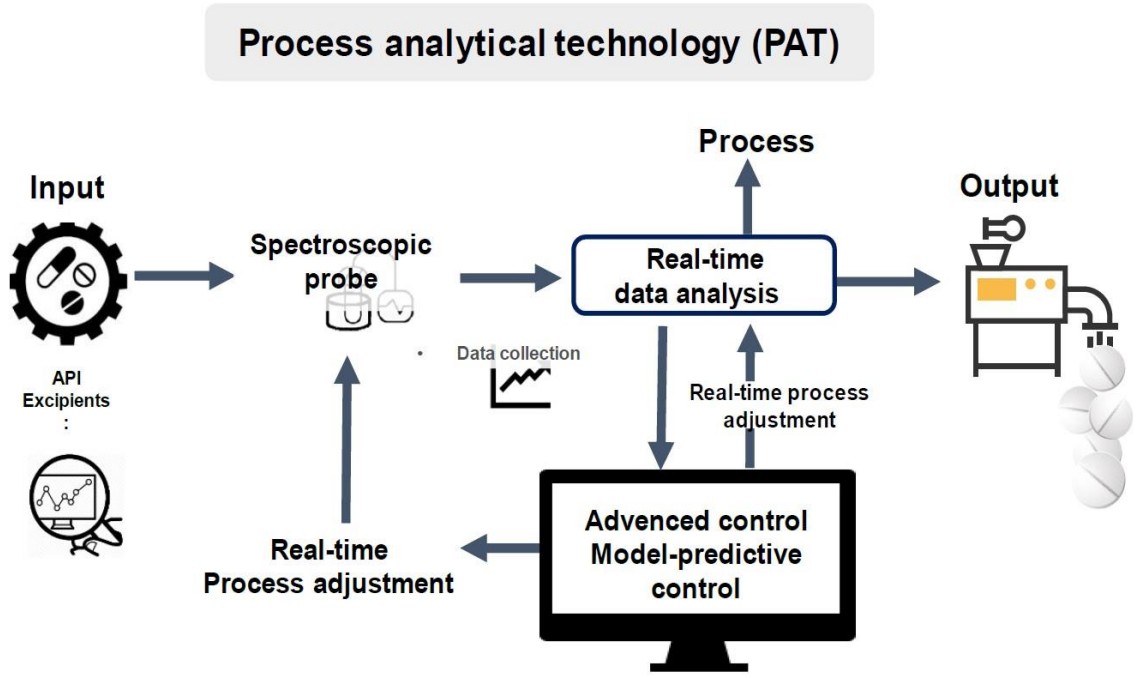


Figure 2.17 PAT for Real-Time Process Monitoring and Control (Kim et al., 2021).

This leads to enhanced process understanding, reduced variability, and improved product quality. PAT aligns with regulatory initiatives such as QbD, emphasising the proactive design of quality into pharmaceutical manufacturing processes. It supports compliance with regulatory requirements by providing a scientific basis for process understanding and control. The integration of data-driven modelling

and PAT in pharmaceutical manufacturing represents a paradigm shift towards a more proactive, efficient, and quality-centric approach. These technologies contribute to enhanced process understanding, control, and compliance with regulatory standards.

Figure 2.18 showcases different types of soft sensor commonly employed in various industrial sectors. While traditional soft sensing techniques like Principal Component Analysis (PCA) and Artificial Neural Networks (ANN) remain widely used, hybrid approaches are gaining popularity due to their ability to combine the strengths of multiple methods. Another noteworthy technique that has recently garnered significant interest from soft sensor developers is Support Vector Machine (SVM)-based regression (Kadlec et al., 2009).

Vargas et al. present a comprehensive study on near-infrared calibration models and chemometric techniques. Their research focuses on analysing different strategies for estimating blend homogeneity in real-time control utilising PAT. The study highlights the potential of Partial Least Squares (PLS), Principal Component Analysis (PCA), and Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS) as promising methods for estimating multiple parameters.

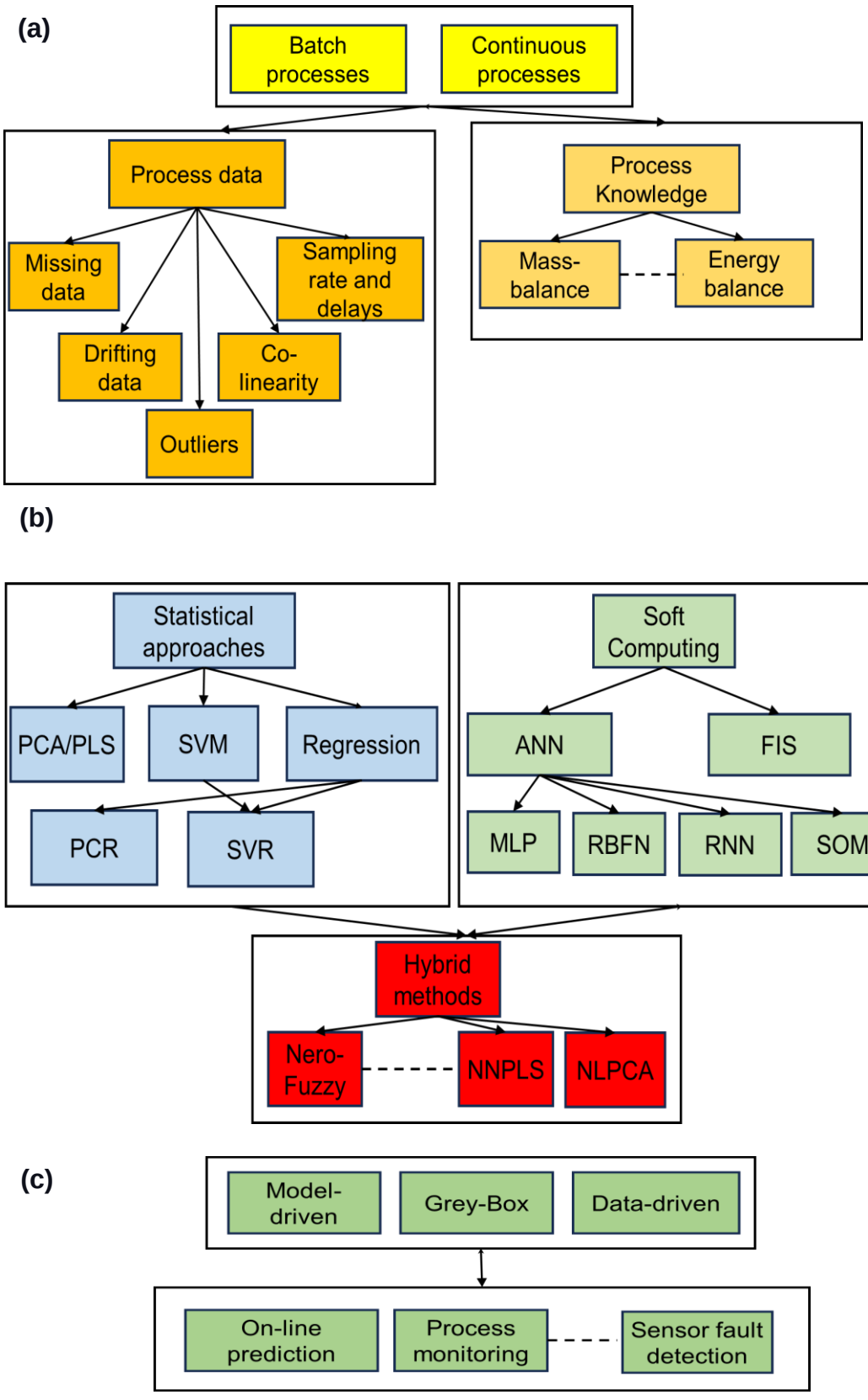


Figure 2.18 Soft Sensor modelling techniques:(a) Mechanistic, (b) Computational and (c) Hybrid (Kadlec et al., 2009).

Ding et al. proposed a soft sensor technology using Partial Least Squares (PLS) to estimate temperature confidence intervals from pyrometer data (Ding et al., 2011). PLS is a robust statistical technique for multivariate analysis, capable of handling multicollinearity and missing data, common challenges in industrial settings. Given the limited availability of machine data samples, PLS is a superior choice compared to methods like Principal Component Analysis (PCA) and Ordinary Least Squares (OLS). Its ability to construct stable regression models with small datasets makes it particularly suitable for this application.

Rogina et al. conducted a study to explore the use of neural networks and linear soft sensors in estimating product quality during the crude distillation process (Rogina et al., 2011). Neural networks, often described as "black box" models due to their complex internal structure, are composed of interconnected layers of simple linear functions. These layers work together to process input data and generate output predictions. By training on large datasets, neural networks can learn to recognize complex patterns and relationships within the data, enabling them to make highly accurate predictions, even in the presence of noise or uncertainty.

$$z_j^{(l)} = \sum_{i=1}^{N_l} w_{ji}^{(l)} x_i^{(l-1)} + b_j^{(l)} \text{ for } j = 1, \dots, J_l \qquad 2.4$$

where $z_j^{(l)}$ is activation j at layer l, $x_i^{(l-1)}$ is the input from layer $l-1$ to layer l, $w_{ji}^{(l)}$ is a weight parameter at layer l, $b_j^{(l)}$ is a bias parameter.

Desai et al. compared Multilayer Perceptron (MLP), Radial Basis Function Networks (RBFN), and Support Vector Regression (SVR) models for soft sensor applications (Desai et al., 2006). SVR outperformed the others due to its ability to find global minima more accurately. While Principal Component Analysis (PCA) is a strong statistical method, it can struggle with datasets that change significantly over time, a common issue in pharmaceutical production. Partial Least Squares (PLS), on the other hand, can produce reliable and consistent predictions, even with diverse training data, making it well-suited for integrated pharmaceutical production lines.

2.6.4 Enterprise Systems & Cloud Manufacturing

The adoption of IoT in pharmaceutical manufacturing presents a transformative paradigm, enhancing real-time monitoring, supply chain visibility, and operational efficiency. This literature review provides insights into the diverse applications of IoT in the pharmaceutical sector, the challenges faced, and the opportunities for leveraging IoT technologies for improved compliance and competitiveness. The integration of IoT technologies has revolutionised various industries, and the pharmaceutical manufacturing sector is no exception. IoT devices enable real-time monitoring of critical parameters such as temperature, pressure, and humidity during pharmaceutical manufacturing processes. This ensures adherence to optimal conditions and facilitates timely interventions to prevent deviations. IoT sensors and tracking devices provide greater visibility throughout the pharmaceutical supply chain. Monitoring the transportation and storage conditions of raw materials, intermediates, and finished products helps to reduce the risk of product degradation.

The data management platform includes production and simulation data, data transmission devices, protocols, and databases. Aside from process prediction, dynamic data analysis, and optimisation, the platform has powerful data visualisation features. These systems employ technologies that allow users to view, evaluate, and monitor data on the cloud. Real-time monitoring requires a model to provide simulation or forecast results in near-real-time, synchronised with sensor data, as seen in Figure 2.19. While hybrid and semi-empirical modelling techniques have this potential, they are usually trained and implemented offline.

Modelling techniques that include data are advantageous for adaptive modelling capabilities since fresh data may be used to improve the models. Several commercial IoT solutions, such as Siemens' MindSphere enable data flow from the plant to the cloud, allowing for remote access and integration with prediction tools (Industrial Edge from Siemens adds benefits from the cloud at the field level Press Company Siemens, 2018).

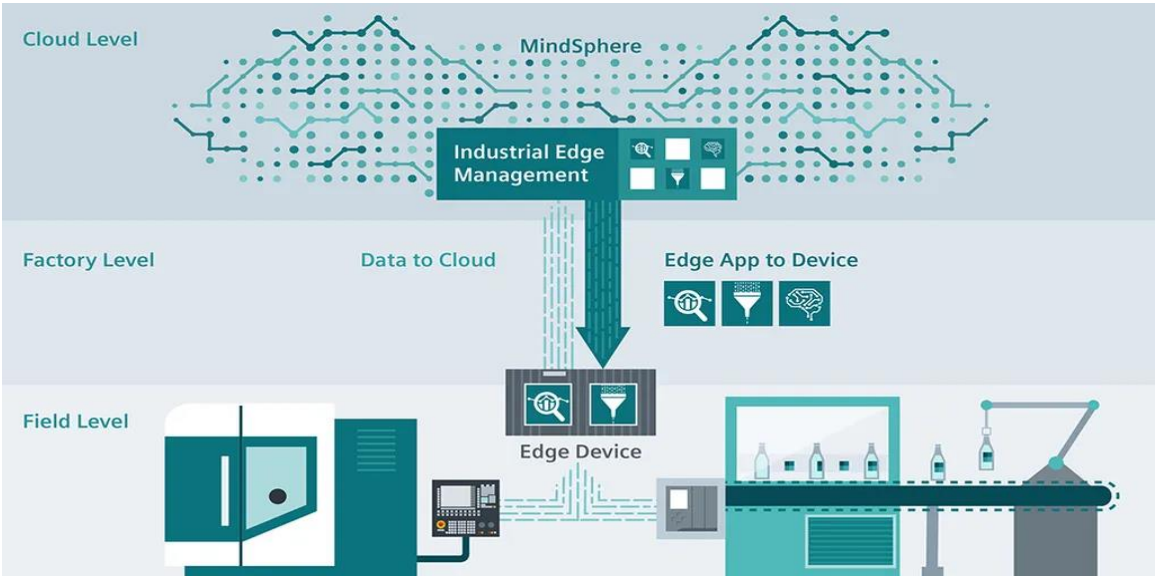


Figure 2.19 Data Flow: Field to Cloud with MindSphere/Insight hub (Industrial Edge from Siemens adds benefits from the cloud at the field level Press Company Siemens, 2018).

Challenges and Opportunities

The growth of IoT devices raises worries over data security and privacy. Pharmaceutical companies must implement robust cybersecurity measures to safeguard sensitive information and comply with regulatory requirements. Integrating IoT technologies with existing manufacturing systems can be challenging. Ensuring seamless connectivity and interoperability between IoT devices and legacy equipment requires careful planning and implementation. Meeting regulatory requirements, particularly in terms of data integrity and documentation, is crucial. Pharmaceutical manufacturers adopting IoT must ensure compliance with industry-specific regulations such as Good Manufacturing Practice (GMP) and Good Distribution Practice (GDP).

2.6.5 Advanced Process Control (APC) system

While traditional control loops, often based on PID controllers, are effective for maintaining process variables at setpoints, they are limited in their ability to optimise overall process performance. Advanced Process Control (APC) and Model Predictive Control (MPC) offer significant advancements. APC techniques, such as multivariable control and model predictive control, can optimise multiple variables simultaneously, considering constraints and disturbances.

Model Predictive Controllers (MPCs) are a cornerstone of Advanced Process Control (APC) systems. MPC uses a predictive model to anticipate future process behaviour and make optimal control decisions. By proactively adjusting control actions, MPC can improve process efficiency, product quality, and reduce energy consumption compared to traditional control strategies.

They often rely on data-driven models to regulate processes. However, building and validating these models can be time-consuming and resource-intensive, requiring dedicated process experiments. To mitigate this, a promising approach involves integrating MPC with digital models enhanced by machine learning techniques. This integration allows for more efficient model development and validation, reducing the need for extensive process trials. By minimising this cost function (Equation 2.5), the MPC controller determines the optimal control inputs to steer the process towards the desired operating conditions while considering constraints on control inputs and process variables (Wong et al, 2018).

$$J = \sum_{k=1}^{p} (\hat{y}_k - y_k^*)' Q_y (\hat{y}_k - y_k^*) + \sum_{k=0}^{m-1} \Delta u_k' Q_u \Delta u_k \qquad 2.5$$

The cost function is composed of two primary components: the actuation target term and the actuation move term.

Where:

- p is the prediction horizon.
- m is the control horizon.
- $\hat{y}_k$ the prediction of the state vector for time step k .
- $y_k^* \, is$ the set-point at time step k.
- $u_k$is the manipulated variable for timestep k.
- Δu_k is the discrete-time rate of change of the manipulated variable which corresponds to the control action size at time-step k.
- Q_u and Q_y are weight matrices.

While the pharmaceutical industry has historically been cautious about adopting advanced process control technologies like Model Predictive Control (MPC), the

success of companies like Fonterra Co-operative Group Ltd. highlights the significant benefits of such implementations. By embracing MPC, Fonterra achieved a substantial ROI of over 60% and experienced notable improvements in productivity, product quality, and energy efficiency (Radspinner and Tormollen 2022). These results demonstrate the potential for MPC to revolutionise pharmaceutical manufacturing processes, leading to increased profitability, reduced environmental impact, and enhanced product quality (Radspinner and Tormollen 2022).

Singh et al. proposes a strategy to integrate control platforms, such as DeltaV, with simulation tools like gPROMS using intermediary software like MS Excel and gORUN (Singh et al., 2014). This integration enables the implementation of Digital Model Predictive Control (DMPC), where the flowsheet model interacts directly with the control system. DMPC offers numerous benefits, including reduced material waste, minimised operational downtime, and fewer experimental trials. By leveraging digital models, DMPC optimises control strategies without solely relying on real-time plant data. This approach, as highlighted by Reynolds and Singh et al., has the potential to significantly enhance the efficiency and productivity of pharmaceutical manufacturing processes (Reynolds, 2019) and Singh et al., 2014).

Reynolds presented a study demonstrating the integration of PharmaMV, a control platform, with gPROMS, a mechanistic modelling platform from Process Systems Enterprise Ltd., to enable model-based digital control of twin-screw granulation (Reynolds, 2019). By creating a task within PharmaMV to facilitate data exchange, the study successfully integrated the two systems. The resulting implementation showcased precise control of critical quality attributes (CQAs) using advanced process control (APC) techniques, ultimately leading to reduced process variability. This integration highlights the potential of combining advanced modelling and control technologies to optimise pharmaceutical manufacturing processes in real time.

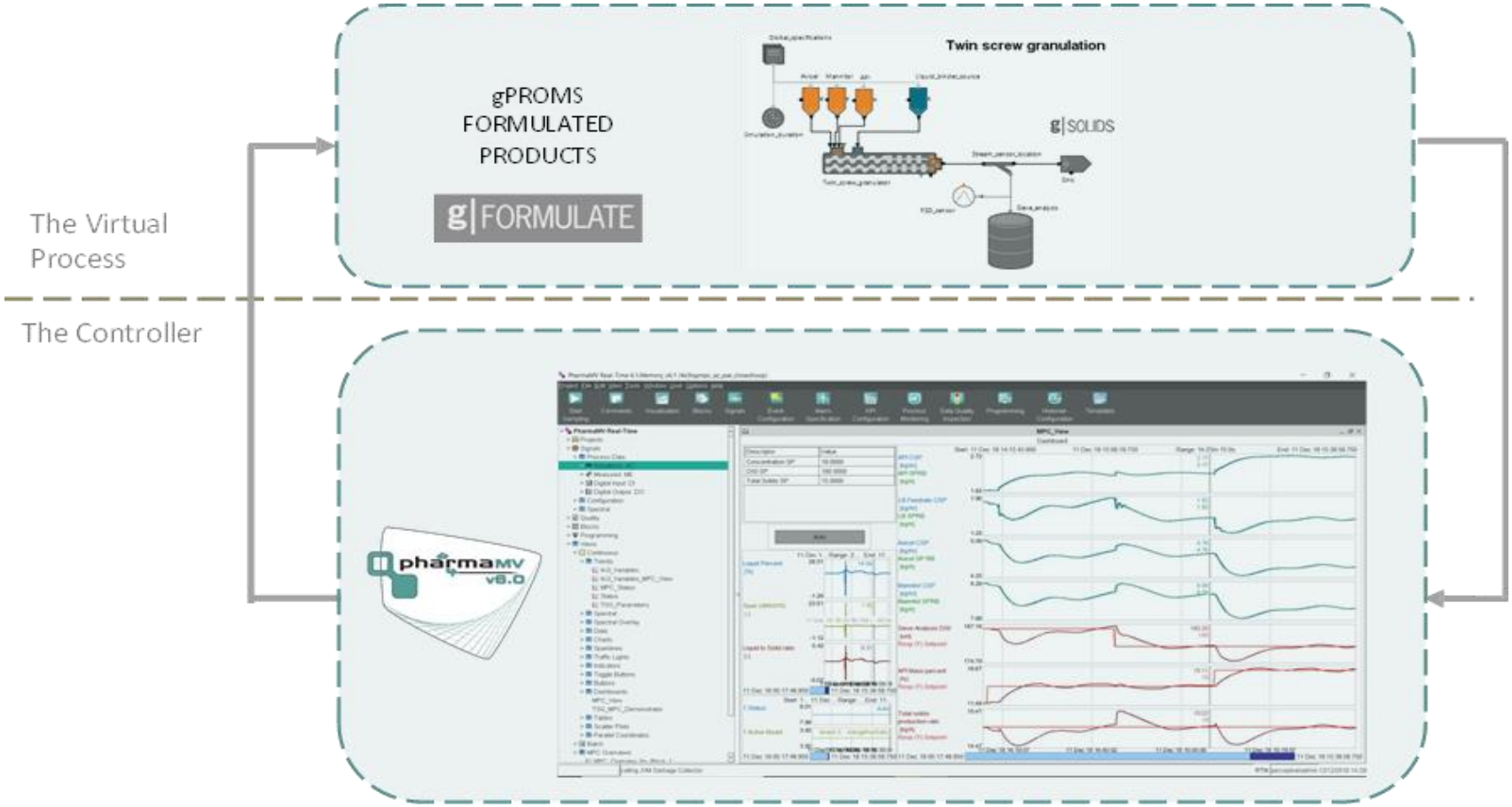


Figure 2.20 gPROMS /Pharma MV Integration (Reynolds, 2019).

2.6.6 Data Management and Data Contamination in Pharmaceutical Manufacturing Industry

This literature review explores the current state of data management, challenges in data contamination, and strategies to mitigate risks in the pharmaceutical sector. In the pharmaceutical manufacturing industry, robust data management practices are vital for ensuring product quality, regulatory compliance, and operational efficiency. Simultaneously, preventing data contamination is crucial to maintain the integrity of information (CloverDX, 2023). Integrated data platforms streamline information flow across the manufacturing lifecycle as shown in Figure 2.21. These platforms enhance data accessibility, accuracy, and traceability, contributing to better decision-making and compliance with regulatory standards.

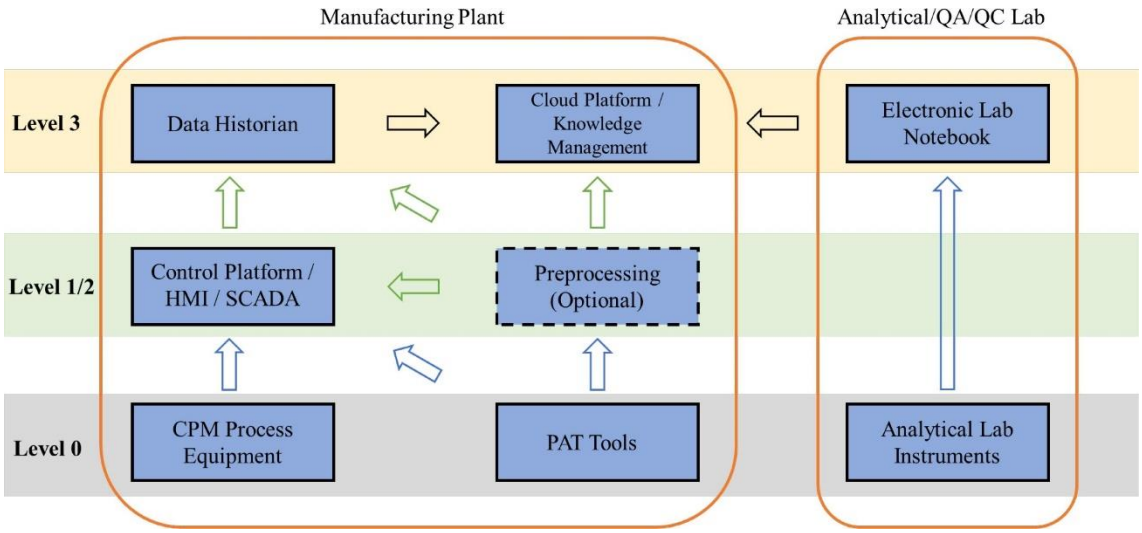


Figure 2.21 General data infrastructure for pilot plants (Chen et al., 2023)

Robust data security measures, including encryption, access controls, and regular audits, ensure the confidentiality and integrity of pharmaceutical data. Compliance with regulations such as Good Manufacturing Practice (GMP) is a critical aspect of data management (Charoo et al., 2023). Implementation of advanced analytics and machine learning in data management allows for predictive maintenance, process optimisation, and quality prediction. These technologies contribute to data-driven decision-making and process improvements (Sudarshan et al,2020).

Data Contamination Challenges and Mitigation

- Regular data validation and quality checks help identify and rectify erroneous entries or inconsistencies. Automated validation processes can be implemented to ensure data accuracy and reliability.
- Establishing strong data governance frameworks and standardising data formats and entry protocols contribute to preventing contamination. This ensures consistency and reliability in the collected data.

Effective data management and contamination control are pivotal for pharmaceutical manufacturing success.

2.6.7 Literature gaps and opportunities

The pharmaceutical industry has made significant strides in adopting digital technologies, but challenges remain in fully realising the potential of DTs. While advancements in PAT technologies, data infrastructure, and modelling techniques have improved, gaps in PAT precision, model accuracy, and real-time data processing hinder the integration of DT modules. Additionally, the complexity of pharmaceutical products and processes makes it difficult to create accurate virtual representations. Furthermore, the industry's stringent regulatory environment and the need for comprehensive, interconnected control systems pose additional challenges. To overcome these obstacles, the industry should invest in advanced instrumentation, iterative optimisation techniques, and machine learning algorithms. By addressing these challenges, the pharmaceutical industry can unlock the full potential of digital twins, leading to improved efficiency, quality, and sustainability.

2.7 Chapter Conclusion

This chapter delved into the conceptualisation of DTs in manufacturing and the existing gap between the ideal DT concept and its practical implementation. The findings revealed a diverse understanding of DTs, ranging from simple digital replicas to complex, real-time, and self-learning systems. While the potential benefits of DTs, such as enhanced product design, optimised manufacturing processes, and improved asset management, are significant, several challenges hinder their full realisation. Key challenges identified include data quality and

integration, modelling complexity, real-time synchronisation, security and privacy concerns, and the lack of standardised global frameworks and methodologies. To bridge the gap between the ideal and the practical, researchers and practitioners must collaborate to develop robust DT frameworks using advanced data analytics techniques and invest in cybersecurity measures. Additionally, fostering a culture of innovation and experimentation within organisations can accelerate DT adoption and drive tangible benefits.

One prominent gap identified in the literature is the limited integration of smart manufacturing principles and technologies tailored specifically for formulated products. While smart manufacturing approaches have gained traction in various industries, their application to formulated products, which often involve complex formulations, stringent quality control requirements, and batch-based production processes, remains underexplored.

Additionally, existing literature focuses on individual components of the manufacturing process, such as automation systems or data analytics tools, rather than providing comprehensive solutions that address the end-to-end manufacturing workflow for formulated products. This fragmented approach hinders the development of holistic smart manufacturing frameworks tailored to the intricacies of formulated product manufacturing. Furthermore, there is a lack of research that explores the potential synergies between smart manufacturing technologies and emerging trends such as Industry 4.0, IoT, and AI in the context of formulated products. Integrating these advanced technologies into the manufacturing process has the potential to revolutionise productivity, quality assurance, and resource efficiency, yet their application to formulated products remains unexplored in the existing literature. Digital operations in pharmaceutical manufacturing leverage models for enhanced efficiency, quality, and decision-making. This literature review also explores methods and strategies to expedite the deployment of models in the pharmaceutical sector, focusing on accelerated implementation and improved operational outcomes.

Implementing a unified data architecture ensures seamless integration of diverse datasets from different manufacturing stages. This enables a holistic view of the manufacturing process, facilitating the deployment of models for predictive

maintenance, process optimisation, and quality control. Leveraging edge computing brings computational power closer to the data source, reducing latency. Edge analytics enable real-time processing, allowing models to run directly on manufacturing equipment. This accelerates decision-making and responsiveness in pharmaceutical operations. Implementing automated processes for model training and validation expedites the development cycle. Continuous learning techniques, coupled with automated validation protocols, ensure that models remain accurate and effective in dynamic manufacturing environments. Cloud platforms provide scalability and flexibility for deploying models across pharmaceutical manufacturing facilities. Cloud-based solutions facilitate rapid deployment, accessibility, and collaborative model development, enabling seamless integration with existing systems.

Meeting regulatory requirements is crucial in pharmaceutical manufacturing. Implementing pre-validated models, adhering to GAMP, and leveraging regulatory-compliant cloud solutions are strategies to address compliance challenges. Ensuring data quality and effective integration are critical for successful model deployment. Employing data cleansing techniques, utilising standardised data formats, and employing interoperable platforms help overcome challenges associated with diverse data sources. Managing organisational change and providing adequate training are essential components of successful model deployment. Establishing robust change management protocols and offering training programs ensure that personnel can effectively use and interpret model outputs.

Accelerating the deployment of models for digital operations in pharmaceutical manufacturing is crucial for realising the full potential of Industry 4.0. The methods discussed in this literature review, from unified data architectures to cloud-based solutions, offer practical strategies to overcome challenges and expedite the implementation of models. The references provided contribute to a comprehensive understanding of the current landscape and best practices for deploying models in the pharmaceutical manufacturing industry.

Overall, the literature review highlights the need for further research and development efforts aimed at bridging the gap between conventional process manufacturing approaches and the unique requirements of formulated products. Future research directions should focus on exploring advanced DT architectures, investigating the impact of AI and machine learning on DT capabilities, and developing user-centric interfaces for effective DT interaction. By addressing these challenges and capitalising on emerging technologies, the manufacturing industry can unlock the full potential of DTs and achieve a new era of digital transformation.

Chapter 3: Experimentation and Data Exploration

3.1 Introduction

The DiPP at the Faculty of Engineering houses large-scale equipment for various industries, with a particular focus on pharmaceuticals. This facility includes a continuous crystallisation unit, a filter dryer, and an industrial-scale GEA Consigma 25 powder-to-tablet line. The continuous oscillatory baffled crystalliser (COBC) and continuous filter dryer produce high-purity crystals and drug substances for use in drug production. Figure 3.1 illustrates the unit operations of the Consigma 25 line, which is divided into front and back-end sections. This continuous wet granulation production line can operate at a nominal throughput of 25 kg/h. The front end consisting of the twin screw wet granulator, segmented dryer and the mill was used to develop the accelerated modelling approach for digital operation and a case study was performed on the Back end which consists of the tablet press.

The DiPP's continuous powder processing plant highlights the importance of complex particulate and formulated products in modern chemical engineering. Equipped with an advanced control and monitoring system, Pharma MV, the plant's dedicated industrial control room enables cutting-edge advanced process control. A collaborative effort between the University of Sheffield, Perceptive Engineering, and Siemens has led to the development of an innovative solution to digitise the entire continuous drug manufacturing process, from crystallisation to tablet pressing. This initiative positions the DiPP as a world-leading Industry 4.0 demonstrator (Ntamo et al., 2022). The collaboration’s primary goal is to adopt a data-centric approach and automation, while also providing a platform for testing and validating potential IoT applications in real-world scenarios.

The line commences with a hopper and a loss-in-weight feeder that transport pre-blended powder to the TSWG unit. In the TSWG unit, granules are formed by combining the powder with the suitable binder. The granulator utilises two screws equipped with kneading and conveying parts to facilitate the mixing process, applying varying levels of force depending on the arrangement. The moist granules are subsequently conveyed by gravity to the segmented FBD, a dryer consisting of six segments where the granules are dried under predetermined

drying conditions. The dehydrated granules are subsequently conveyed by a pneumatic system to the granules conditioning unit, where the moisture content is measured using NIR technology (Near Infra-Red, Fibre Optic FP710e, NDC Technology, Essex, UK).

After milling the granules in a cone mill, an NIR probe, a PAT tool, monitors the moisture content to ensure product quality. The system either retains the product or diverts it to waste based on the NIR probe's feedback. Subsequently, the granules are transferred to two additional blending units for incorporation of further substances and lubricants. The final stage involves the rotary tablet press.

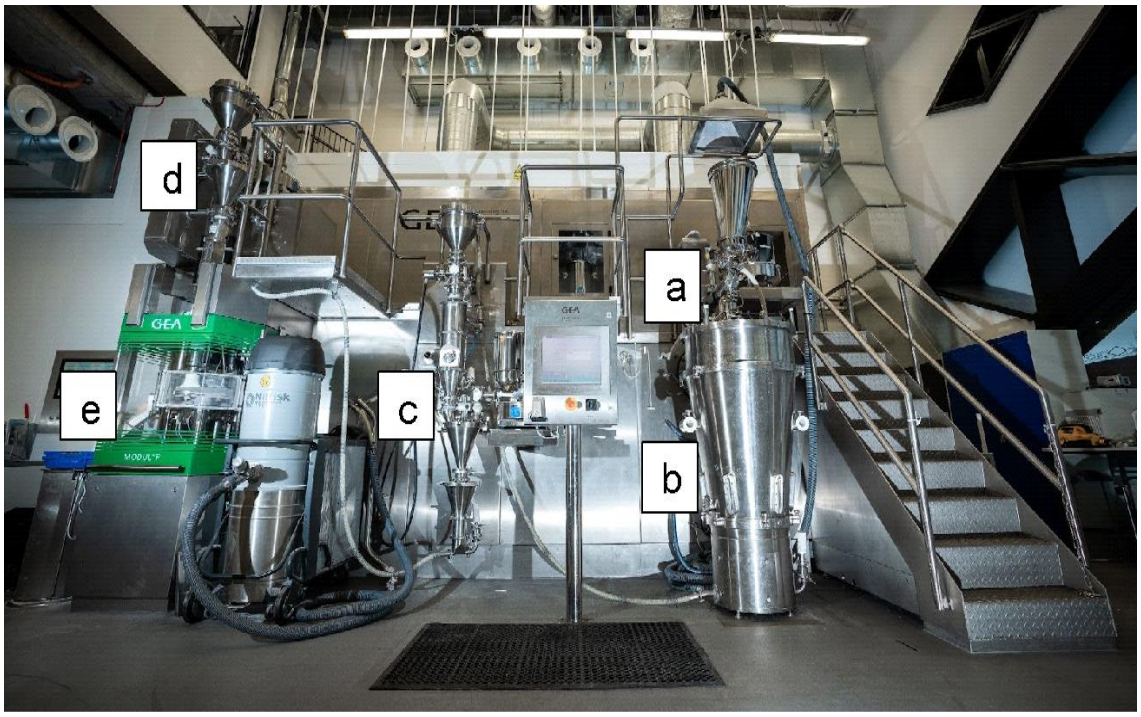


Figure 3.1 The Consigma 25 line at Diamond Pilot Plant (DiPP): (a) Twin screw wet granulator, (b) Segmented fluid bed dryer, (c) Cone mill, (d) Blender and (e) Tablet press.

Data scientists employ exploratory data analysis (EDA) to delve into datasets, often employing data visualisation techniques to summarise key characteristics. EDA's primary goal is to facilitate data analysis without making prior assumptions. It helps identify errors, locate outliers, and discover hidden relationships between variables. Data visualisation enables users to grasp the essential features of complex datasets shown in Figure 3.2 (process data, engineering data, quality logs, and historical data) more intuitively than by simply examining data tables.

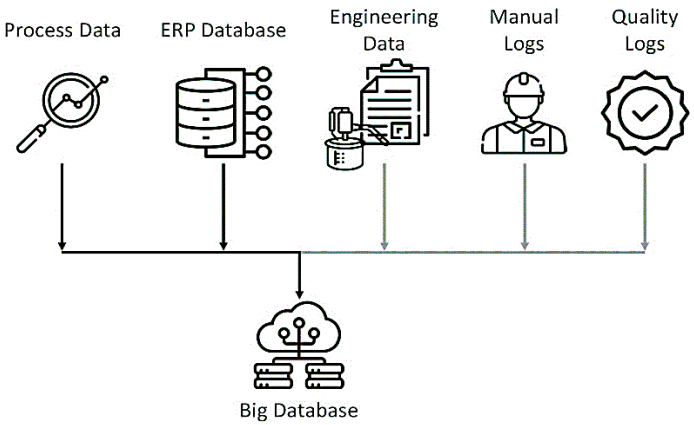


Figure 3.2 Manufacturing data sources (Chukwu Christian Onyemaechi, 2024).

Visualisation can show aspects of the data missed by the descriptive statistics. In this case, scatter plots can identify the presence of non-linearities, outliers and leverage points. Visualisation can reveal data features that descriptive statistics cannot capture. In this scenario, scatter plots might reveal the existence of nonlinearities, outliers, and leverage points (D. Ntamo et al.,2021).

Digital twins rely on information from the product they are simulating. They rely on it both in the original design of the digital twin and its ongoing modifications. Given the frequent updates to digital twin models and the vast amount of data they generate, working with them requires significant data management capabilities. All this data must be kept and tagged effectively and traceable. This requires assets to be versioned and monitored during development.

To ensure data quality and prevent the propagation of errors, it is crucial to implement checks at every stage of the data infrastructure, as illustrated in Figure 3.3. By treating each pipeline step as an independent component, data quality issues can be identified and addressed early on. Validation, which involves applying rules to each record to verify accuracy, is essential. In this context, RSME is used to assess model performance. Entries that satisfy verification will advance to the next phase, while those that fail are flagged for review, correction, or removal. Profiling, which involves creating statistical summaries of datasets, helps identify potential data quality issues and allows for the detection of early signs of data degradation. Outlier detection can be employed to remove problematic data from the entire dataset.

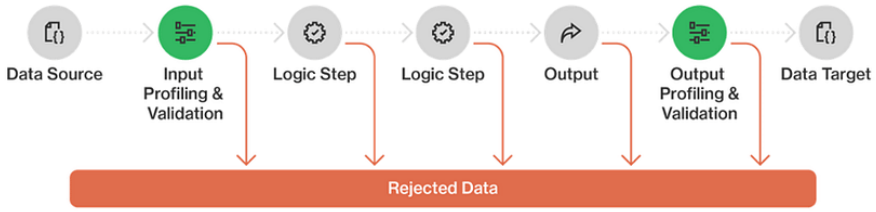


Figure 3.3 Data processing workflow (CloverDX, 2023).

Prior to analysis, data underwent preliminary cleaning to handle missing or null values. This process included the application of several strategies: complete case deletion, imputation with a fixed value, global or local mean imputation, linear interpolation, and forward fill (hold last good value).

3.2 Experimental Planning and preparation

There are several critical quality attributes (CQAs) for pharmaceutical granules which include granule size distribution, flowability and moisture content. These CQAs of granules will eventually affect the CQAs of the final tablets including mass, content uniformity, strength, and dissolution profile. For each drug product that is registered with the regulator agency (European Medicines Agency (EMA), U.S. Food and Drug Administration (FDA)), a specified value and range for each CQA is specified, ultimately for patient safety. If a CQA falls outside this range, the batch of tablets must be discarded. The experiments were designed around controlling these attributes for a seamless operation of the physical product in a cost-effective manner. A factorial design of experiments was used to identify which parameters, and parameter interactions have a significant effect on CQAs and sustainability KPI. In addition, this section provides specific information on our experimental design, such as the duration of equipment usage per experiment and the number of duplicate trials. To optimise the experimental design, factors such as specific objectives, the number of input variables and their interactions, and the statistical power and efficiency were considered. Screening designs were selected due to their cost-effectiveness. Common screening designs include two-level full factorial designs, fractional factorial designs, and Placket-Burman designs. These designs enable the evaluation of numerous input parameters with fewer experiments. However, they have significant limitations that may hinder a comprehensive understanding of input-output relationships.

3.3 Materials

A 20 Litre tumble blender (Inversina-Bioengineering, Wald, Switzerland) was used to prepare the formulation pre-blend. The blender was operated for 15 minutes at 25 rpm. Then, the pre-blend was transferred to the gravimetric loss-in-weight feeder (KT20, K-Tron Soder, Niederlenz, Switzerland). The blended powder composed of 72% lactose, 24% MCC and 4% PVP. Before being fed into the granulation unit particle size of the formulation was analysed using sieves to provide material properties and results are shown in Figure 3.4.

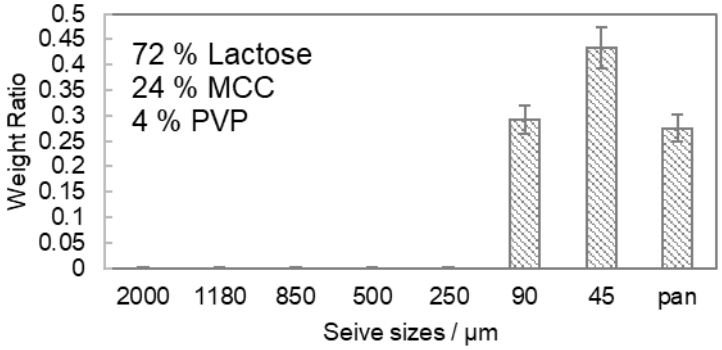


Figure 3.4 Pre-Blend formulation size analyses.

3.4 Experimentation

3.4.1 TSWG Experiments

3.4.1.1 Equipment setup

This experiment uses the Front end of the ConsiGma25 plant consisting of the main hopper where the pre-blended powder is stored, loss in weight feeder, TSWG unit, segmented FBD, NIR probe, conical milling unit, and blender. Since only the granulator will be needed in this experiment, the granulator illustrated in Figure 3.5 was disconnected from the whole line and ran without the dryer and mill.

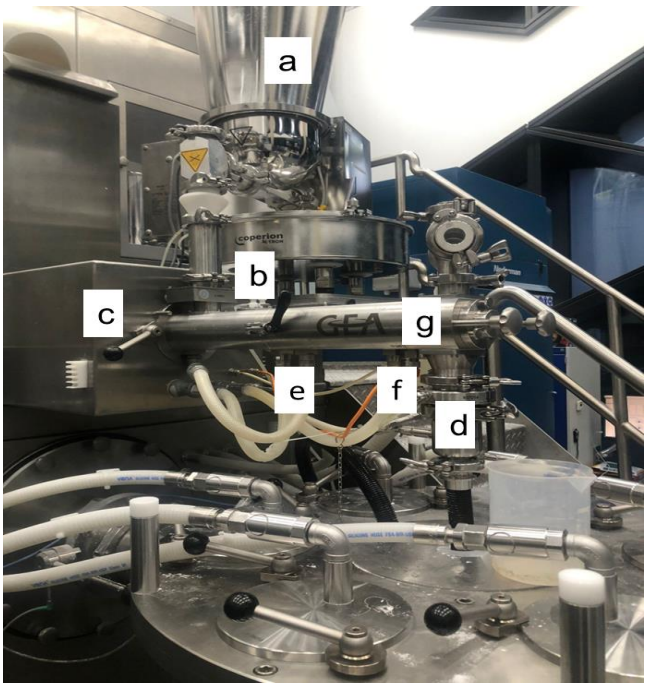


Figure 3.5 TSWG unit at the Diamond Pilot Plant: (a) Powder Feed, (b) Liquid Feed, (c) Motor unit, (d) Granule outlet, (e)Temperature control inlet, (f)Temperature control outlet and (g) TSWG barrel.

Fresh water was used as the granulation fluid, and it was delivered into the granulator using two peristaltic pumps (Watson Marlow, Cornwall, UK) to guarantee exact dosing rates. Figure 3.6 depicts the screw configuration, which comprised of five compartments: three conveying elements and two kneading elements. The kneading components were placed at a 60° forward stagger angle, separated by conveyer elements.

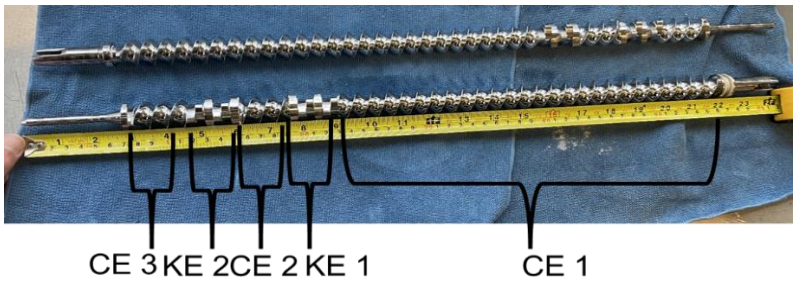


Figure 3.6 Twin screw granulator screw configuration at the DiPP.

The jacket of the granulator barrel was kept at 25°C. The liquid-to-solid ratio, powder feed rate, and screw speed have been varied from 0.15 to 0.35, 5 to 20 kg/h, and 200 to 500 rpm, respectively. The wetted granules are collected from the TSWG outlet, dried overnight at DiPP's room temperature (22°C), and then

characterised. The granule size distribution was evaluated using a sieve shaker (Figure 3.7) with screen sizes ranging from 60 µm to 2800 µm.

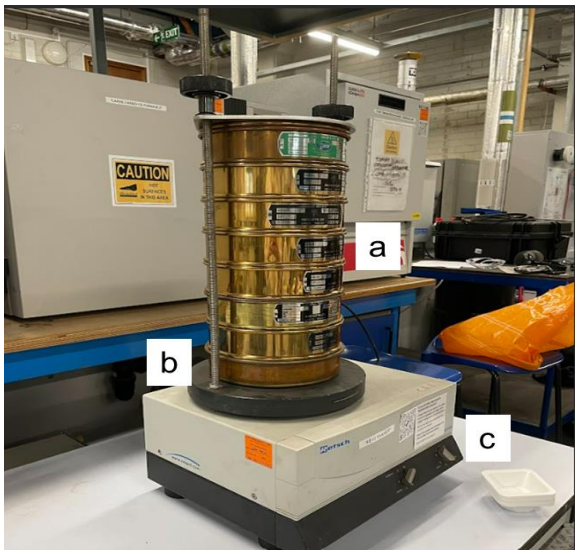


Figure 3.7 Retsch Sieve Shakers AS 200 control: (a) Sieves, (b) Sieve shaker and (c) Amplitude and timer knobs.

The ideal parameters for sieving time and amplitude/speed depend on the material to be sieved. They have a crucial influence on the sieving result. Usually, national and international standards such as ISO 3310 and ASTM E11 and internal regulations provide plenty of product-specific information about sieve analyses and the corresponding parameters. The sieves were mechanically shaken with an amplitude of 1 mm every ten seconds for a total of three minutes to provide sufficient movement to separate particles without causing excessive wear on the sieves or damage to the sample and to prevent particles from becoming lodged in the sieve mesh. Three minutes have been determined through testing as an optimal time to achieve a complete separation of particle sizes for the type of material being tested.

3.4.1.2 Design of Experiments and Sampling strategy

A custom randomised design matrix was employed, utilising centre points to investigate process variables. To ensure robust statistical analysis, each experimental run was performed with three repeats. The sampling rate was meticulously controlled, with five 3g samples collected every 2 minutes. Furthermore, to capture the process dynamics after initial stabilisation, sample collection commenced five minutes after the start of each run. This approach

allowed for the generation of comprehensive data, facilitating a thorough understanding of the process behaviour across the defined experimental space.

3.4.1.3 Results and Discussion

Figure 3.8 presents the granule size distribution as a mass fraction plot, normalised by the number of regarded granules on each sieve.

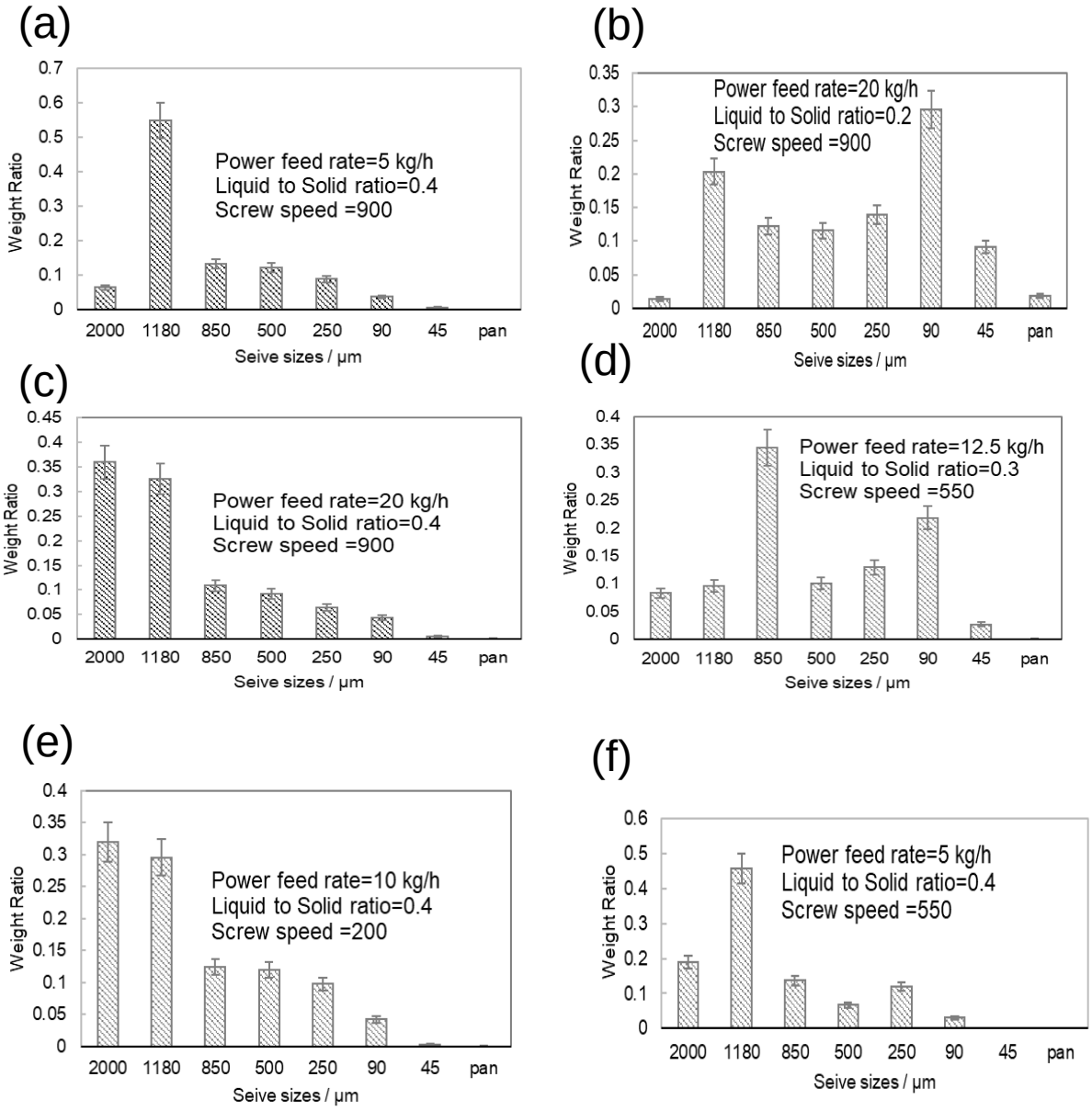


Figure 3.8 Sieve analysis particle size distribution results for different runs: (a) Run1, (b) Run 2, (c) Run 3, (d) Run 4, (e) Run 5, (f) Run 6 and (g) Run 7.

All experimental runs exhibited an increase in final granule size compared to the starting material. Run 2, with a powder feed rate of 20 kg/h, screw speed of 900 rpm, and a liquid-to-solid ratio of 0.2, resulted in a bimodal size distribution,

indicating two distinct particle populations. The two bimodal distributions show that there were mixed proportions of small and large granules in these samples due to liquid to solid ratio compared to an even distribution of sizes, demonstrated by the monomodal distribution. Increasing the liquid-to-solid ratio to 0.4 in run 3 led to a higher d50 (median particle size) and lower d10 (10th percentile particle size), accompanied by a shift towards a narrower, monomodal distribution (single peak). Such observation is in accordance with the literature report where a higher L/S ratio results in a narrower, monomodal size distribution (Dhenge et al., 2010). Interestingly, runs 1 and 6 displayed similar distributions despite a significant increase in screw speed, suggesting minimal influence of screw speed on particle size distribution in this range. This aligns with the findings from Metta et al.'s process parameter sensitivity study. In contrast, runs 2 and 4, with lower liquid-to-solid ratios, exhibited bimodal distributions likely due to a higher proportion of fine particles. These observations suggest that powder feed rate and liquid-to-solid ratio significantly impact granule size dispersion, whereas screw speed, within the tested range, appears to have a lesser effect.

3.4.2 Drier and Milling Experiments

3.4.2.1 Equipment setup

This section delves into the effect of intermediate CQAs on tablet properties, particularly focusing on the fluidised bed dryer and milling unit, to gain a comprehensive understanding and optimise the entire tableting process. The Drier experiment uses the Front end of the ConsiGma25 plant consisting of the main hopper where the pre-blended powder is stored, loss in weight feeder, twin screw wet granulator, fluidised bed dryer (Figure 3.9) and NIR probe.

The drying process is a critical step in pharmaceutical manufacturing, as it directly impacts the quality and stability of the final product. In particular, the moisture content of granules is a key critical quality attribute (CQA) that significantly influences various product properties. Excess moisture can lead to chemical degradation, poor powder flowability, suboptimal compaction behaviour, delayed disintegration, and dissolution rates, and increased microbial growth. All these factors have direct implications for patient safety. Therefore, precise control of the drying process is essential to ensure that the final product meets stringent

quality standards. The amount of moisture remaining in the granules prior to tableting is determined by the drying conditions within the fluidised bed dryer and the efficiency of mass and heat transfer within the unit.

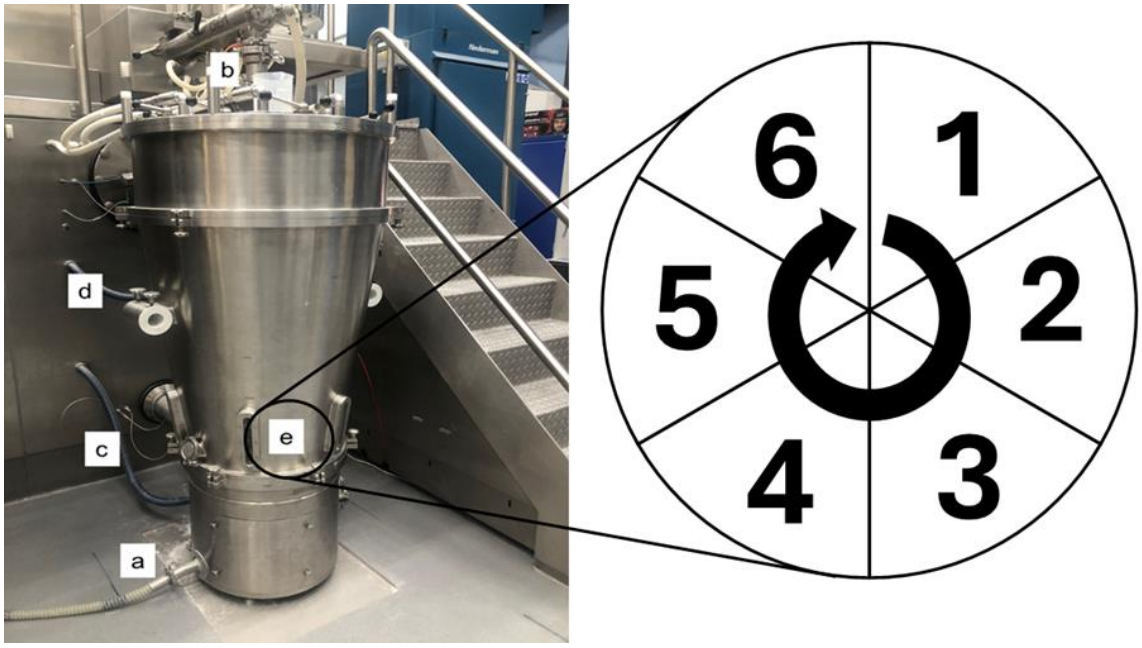


Figure 3.9 DiPP Fluidised bed drier: (a) Dried granule outlet, (b) Wet granule inlet, (c) Drying air inlet tube, (d) Drying air outlet tube and (e) 6 segmented drying cells.

3.4.2.2 Design of Experiments

This experiment investigates the impact of drying air temperature on the final product moisture content. The granulator is operated at standard screw speed, powder feed rate and Liquid to Solid ratio of 500 rpm, 10 kg/h and 0.25 respectively. The drying air flow rate was kept as 360 m^3/hr at atmospheric pressure. The dry air temperature was varied from 40°C to 70°C, as detailed in Figure 3.10.

3.4.2.3 Results and discussion

As expected, the results demonstrate that higher drying air temperatures lead to a faster decrease in granule moisture content during the drying phase (240 s to 840 s). From 0 to 240 s, the wet granules were filled into the wet cell from the granulator.

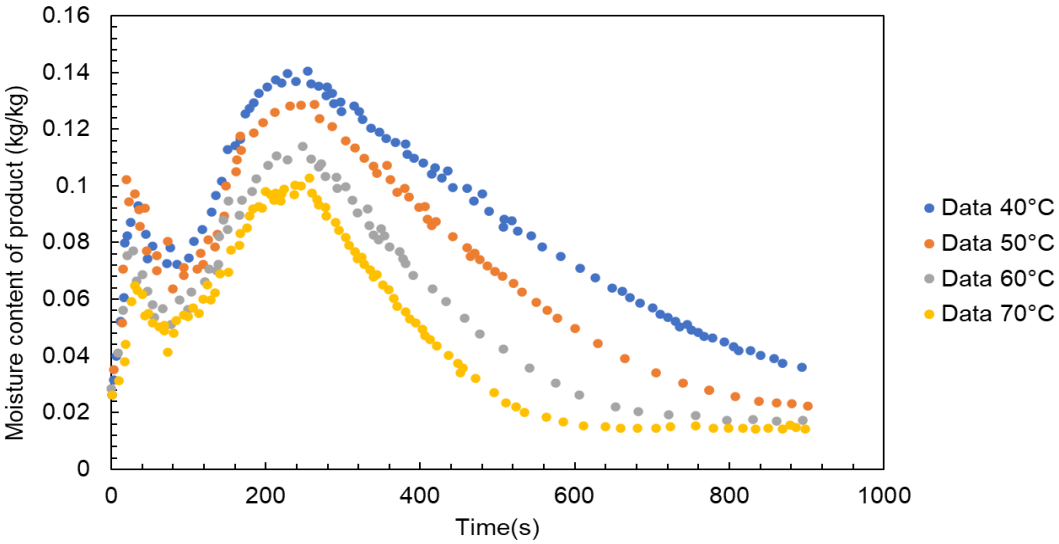


Figure 3.10 Drying experiment results at different air inlet temperature (Wang et al,2022).

3.4.2.4 Milling output

While an ideal liquid-to-solid ratio in the twin-screw granulator can lead to a narrower particle size distribution (PSD), achieving both optimal mean size and narrow PSD can be challenging. Therefore, minimising PSD becomes a crucial objective in the subsequent milling process. Dried granules were crushed in a conical screen mill for 100 seconds to ensure steady-state operation.

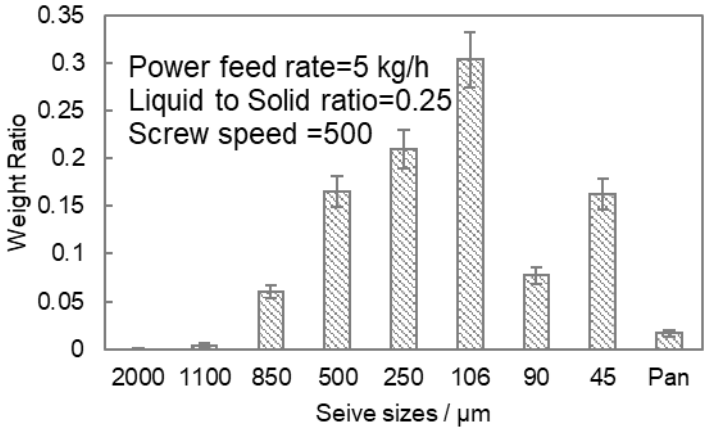


Figure 3.11 Sieve analysis results of products.

The milled product was then collected and characterised using sieve analysis to determine its PSD. The results of these experiments are presented in Figure 3.11.These findings will be used to develop a model for predicting milling outputs.

Through a comprehensive preliminary analysis of the Front-end Dryer and Mill unit, optimal operating conditions, such as drying air temperature, were identified. These parameters were meticulously maintained throughout the subsequent analysis of tablet press CPPs and CQAs, as detailed in the following section. By establishing and adhering to these optimal conditions, I aim to ensure consistent product quality and efficient process performance.

3.4.3 Tablet Press Experiments

3.4.3.1 Equipment setup

The ConsiGma25 Back End consists of the rotary press (tabletting machine). Figure 3.12 shows the schematic diagram of ConsiGma25 and description of all parts.

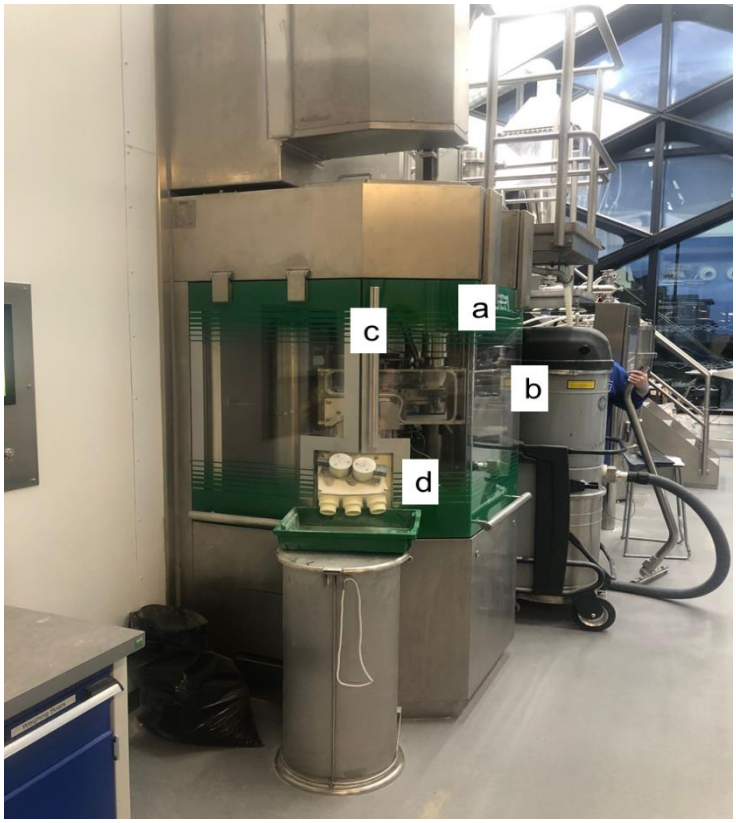


Figure 3.12 Rotary tablet press: (a) Feeding point, (b) Feeding frame, (c) Feeding paddles and (d) Ejection point.

The rotary press allows powders to be compacted straight into tablets. To ensure that the tablet press produces excellent tablets with the requisite key quality attributes (CQAs), the powder mix before the compression stage must be dry, flowable, and compressible. In addition, all ingredients in the blend must be

evenly distributed within each tablet that is produced. As a lubricant, magnesium stearate is added to the mixture to enhance die filling before compression, lessen powder adhesion to the die and punches, and facilitate tablet ejection. The filling of powders to the die occurs in the feeding frame using two feeding paddles. The powder then undergoes a pre-compaction step where the air is removed, and particles rearrange themselves. The powder undergoes deformation, fragmentation, and bonding during the final compression stage following pre-compaction. This forms the tablet, which is subsequently expelled by the lower punch during the ejection step.

3.4.3.2 Design of experiments

The experiment was focused on investigating the effect of the compaction force, mass, and turret speed in the final compaction stage on tablet properties. When investigating the effect of compaction force, the Magnesium stearate (MgSt) concentration was fixed at 1%. This mixture was then transferred to the tablet press to be compressed into tablets at variable tablet masses, turret speeds, and final compaction forces, as outlined in Table 3.1.

Table 3.1 Tablet press process parameters.

Tablet Compaction Conditions	*Value*	*Unit*
Over fill cam	12	mm
Fill cam dosing mass	500-700	mg
Displacement on the final compaction	0.3	mm
Turret speed	20-80	rpm
Pre-compaction force	2	kN
Compaction force	2-20	kN

Each run was allowed to operate for 10 minutes, and 10 tablets were collected every 2 minutes. After each run was completed,10 random tablet samples were collected from the batch for further analysis. These tablets were characterised by measuring their breaking force, thickness, mass, and diameter using a SOTAX tablet hardness tester (MT-50, SOTAX, Switzerland) illustrated in Figure 3.13.

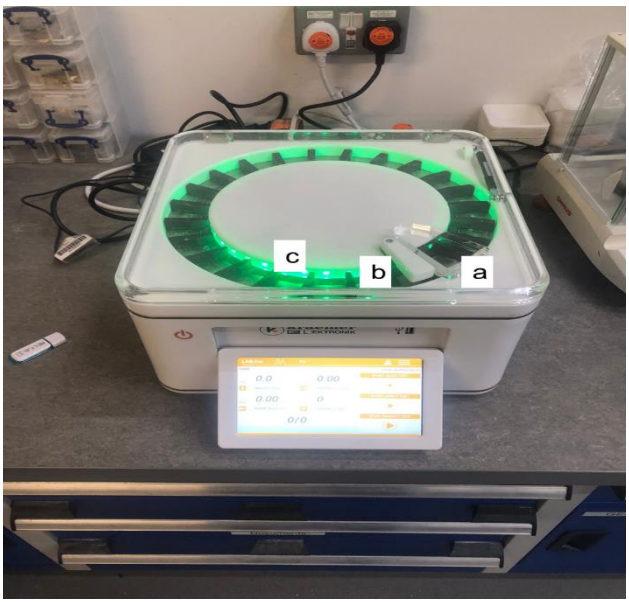


Figure 3.13 SOTAX tablet hardness tester: (a) Diameter and breaking force analyser, (b)Thickness analyser and (c) Slot to insert tablet for characterisation.

The tablet weight was measured in grams using a four decimal high-precision balance. The fill cam height, a crucial process parameter that influences the amount of powder filled into the die cavity, was determined through initial experimentation to achieve the target tablet mass by determining the mass of 10 tablets for each fall cam height and dividing them by 10 maintaining the error at ±0.0005 g. The results of this experimentation are presented in Figure 3.14.

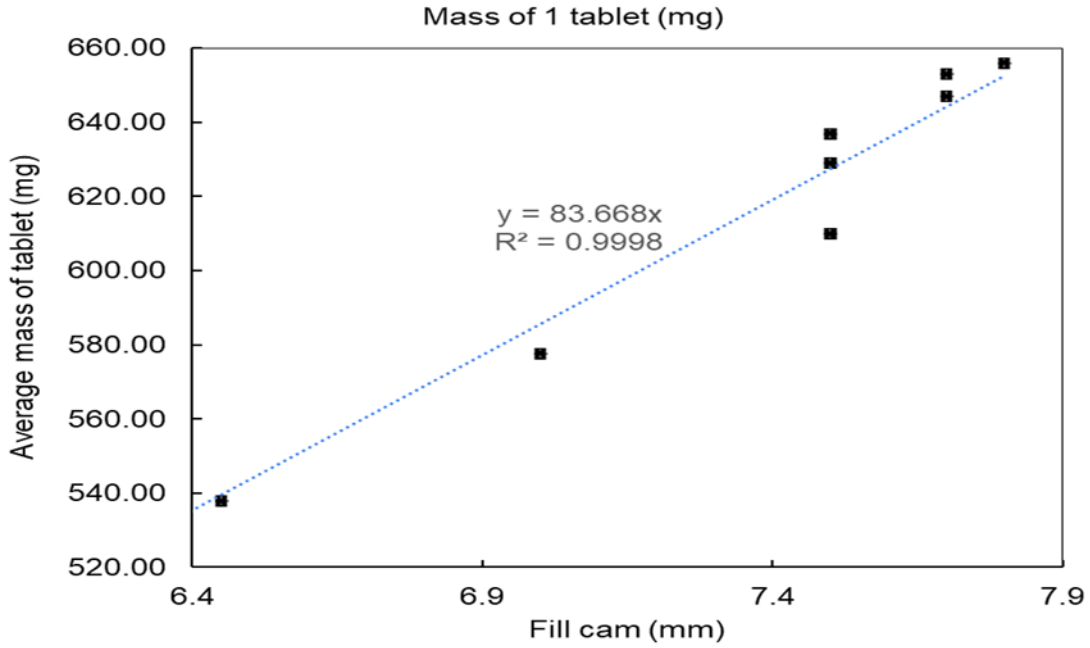


Figure 3.14 Mass of single tablet vs Fill cam.

The tensile strength of a tablet is an important attribute as the tablet needs to be mechanically strong enough to withstand further handling such as film–coating, packaging, transport and end-use by the patient, but to be weak enough to break apart in the human body and so release its contents. Pitt et al. developed Equation 3.1 to estimate the tensile strength σ_t of convex-faced elongated tablets, specifically accounting for their unique geometry (Pitt et al ,1989).

$$\sigma_t = \frac{10F}{\pi d^2 \left(\frac{2.84h}{D} - \frac{0.126h}{(h-2h_{cap})} + \frac{3.15(h-2h_{cap})}{d} + 0.01\right)} \quad 3.1$$

This equation relates tablet breaking force F,its the diameter d and tablet thickness h measured using the hardness tester to its tensile strength. Punch cap depth h_{cap} for the tablet is measured to be 1.21 mm. This equation was consequently used to calculate the tablet tensile strength in this thesis.

The ConsiGma-25 HMI provides a user-friendly interface for precise control over the tablet compression process. Four key input parameters can be adjusted to influence the final tablet characteristics. These parameters are crucial for achieving desired Critical Quality Attributes (CQAs) such as tablet tensile strength, disintegration time, and dissolution profile.

***Motor Speed*:** This parameter dictates the rotational speed of the tablet press's compression tooling. Higher motor speeds generally lead to increased production throughput but can also impact tablet uniformity and potentially generate excessive heat. Precise control over motor speed is essential for maintaining consistent tablet quality at the desired production rate.

***Compression Force*:** This parameter directly controls the force applied to the powder bed during the compression stage. The compression force significantly influences tablet hardness and tensile strength. Insufficient force may result in friable tablets, while excessive force can lead to capping or lamination.

***Tablet Mass*:** This parameter defines the target weight of individual tablets. Precise control over tablet mass is critical for ensuring consistent dosage and product uniformity. The HMI facilitates accurate adjustment of tablet mass,

allowing for fine-tuning of the filling process to achieve the desired target weight and minimise weight variation.

To comprehensively assess the impact of key process parameters on the Critical Quality Attribute (CQA) of tablet tensile strength, a full factorial Design of Experiments (DOE) was implemented. This approach enabled the systematic evaluation of all possible combinations of pre-defined low and high levels for each parameter. The resulting tablet tensile strength for each experimental run, representing a unique combination of parameter levels in Table 3.2, was subsequently measured and is presented in Table 3.3. This methodology allowed for the determination of both main effects and interaction effects of the parameters on tablet tensile strength.

Table 3.2 Tablet press DoE process parameters combinations.

Exp	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
Speed (rpm)	20	80	20	80	20	80	20	80
Compression Force (kN)	2	2	20	20	20	20	2	2
Tablet Mass (mg)	545	538	546	539	651	647	650	640

3.4.3.3 Results and discussion

To achieve the target tablet mass, the corresponding fill cam height was adjusted. The relationship between tablet target mass and fill cam height is depicted in Figure 3.14. For each experimental run, 10 tablet samples were collected at two-minute intervals to ensure a representative sample of the population. Each experiment was replicated three times. Table 3.3 presents the statistical analysis for one of these replicates. A more in-depth analysis of all three replicates for each experiment, along with an important discovery related to data contamination, is discussed in the Data Contamination section.

Compaction force, turret speed and target mass of a single tablet were varied from 2 to 20 kN, 20 to 80 rpm, and 500 to 700 rpm respectively as shown in Table 3.3. have a significant effect on the tablet tensile strength with p value less than 0.05. The linear model provides the aces the impact of each parameter has high adjusted r squared value of 0.991 suggest that a simple correlation compare the twin screw ANOVA.

Table 3.3 Tablet press tensile strength results Exp 1, Exp 2, Exp 3 and Exp 4.

	Tablet strength (MPa)					
Exp	*Mean*	*Std. Deviation*	*Range*	*25th percentile*	*50th percentile*	*75th percentile*
1	1.63E-01	8.89E-03	2.32E-02	1.52E-01	1.62E-01	1.72E-01
2	1.56E-01	1.09E-02	3.42E-02	1.45E-01	1.55E-01	1.65E-01
3	2.06E+00	1.25E-01	4.18E-01	2.05E+00	2.10E+00	2.13E+00
4	1.86E+00	6.95E-02	2.75E-01	1.82E+00	1.86E+00	1.89E+00
5	1.93E+00	7.59E-03	3.28E-01	1.90E+00	1.94E+00	1.97E+00
6	1.80E+00	3.95E-03	2.10E-01	1.74E+00	1.80E+00	1.84E+00
7	1.71E-01	9.47E-05	2.71E-02	1.61E-01	1.70E-01	1.78E-01
8	1.78E-01	6.53E-05	2.15E-02	1.69E-01	1.80E-01	1.84E-01

The experimental design varied three key process parameters: compaction force (2 to 20 kN), turret speed (20 to 80 rpm), and target tablet mass (500 to 700 mg), as detailed in Table 3.2. A linear regression model was employed to assess the individual contributions of each parameter. The model exhibited a high adjusted R-squared value of 0.991, indicating a strong correlation between the parameters and tensile strength. This suggests that, in this specific context, a simple linear model effectively captures the relationships, potentially offering a more straightforward analysis compared to a two-way ANOVA.

Table 3.4 Tablet press ANOVA Model summary

					Change Statistics			
Model	*R*	*R Square*	*Adjusted R Square*	*Std. Error of the Estimate*	*R Square Change*	*F Change*	*df1*	*df2*	*Sig. F Change*
1	.996[a]	.992	.991	.0821808	.992	3013.900	3	76	<.001

a. Predictors: (Constant), Turret Speed, Compaction Force, Tablet Mass

Statistical analysis, illustrated in Table 3.5, revealed a significant impact of increasing compaction force, turret speed, and tablet mass on tablet tensile strength ($p < 0.05$). Within the linear regression model, compaction force exhibited a substantially greater coefficient compared to turret speed and tablet mass, indicating its dominant influence on tablet tensile strength.

Table 3.5 Tablet press ANOVA Tests of Between-Subjects Effects

Dependent Variable: Tensile Strength					
Source	*Type III Sum of Squares*	*df*	*Mean Square*	*F*	*Sig.*
Corrected Model	61.286a	7	8.755	2160.376	<.001
Intercept	86.379	1	86.379	21314.372	<.001
Tablet Mass	0.032	1	0.032	7.860	0.006
Compaction Force	60.884	1	60.884	15023.483	<.001
Turret Speed	0.148	1	0.148	36.638	<.001
Tablet Mass ×Compaction Force	0.060	1	0.060	14.795	<.001
Tablet Mass × Turret Speed	0.009	1	0.009	2.122	.150
Compaction Force × Turret Speed	0.149	1	0.149	36.814	<.001
Tablet Mass × Compaction Force × Turret Speed	0.004	1	0.004	0.923	0.340
Error	0.292	72	0.004		
Total	147.957	80			
Corrected Total	61.578	79			

a. R Squared = .995 (Adjusted R Squared = .995)

This observation aligns with the direct relationship between compaction force and tablet mechanical strength. While turret speed and tablet mass are critical for ensuring consistent dosage and product uniformity, their impact on tensile strength is less pronounced. This finding is consistent with the trends depicted in Figure 3.15, which illustrates the effects of increasing compaction force, turret speed, and tablet mass on tablet tensile strength.

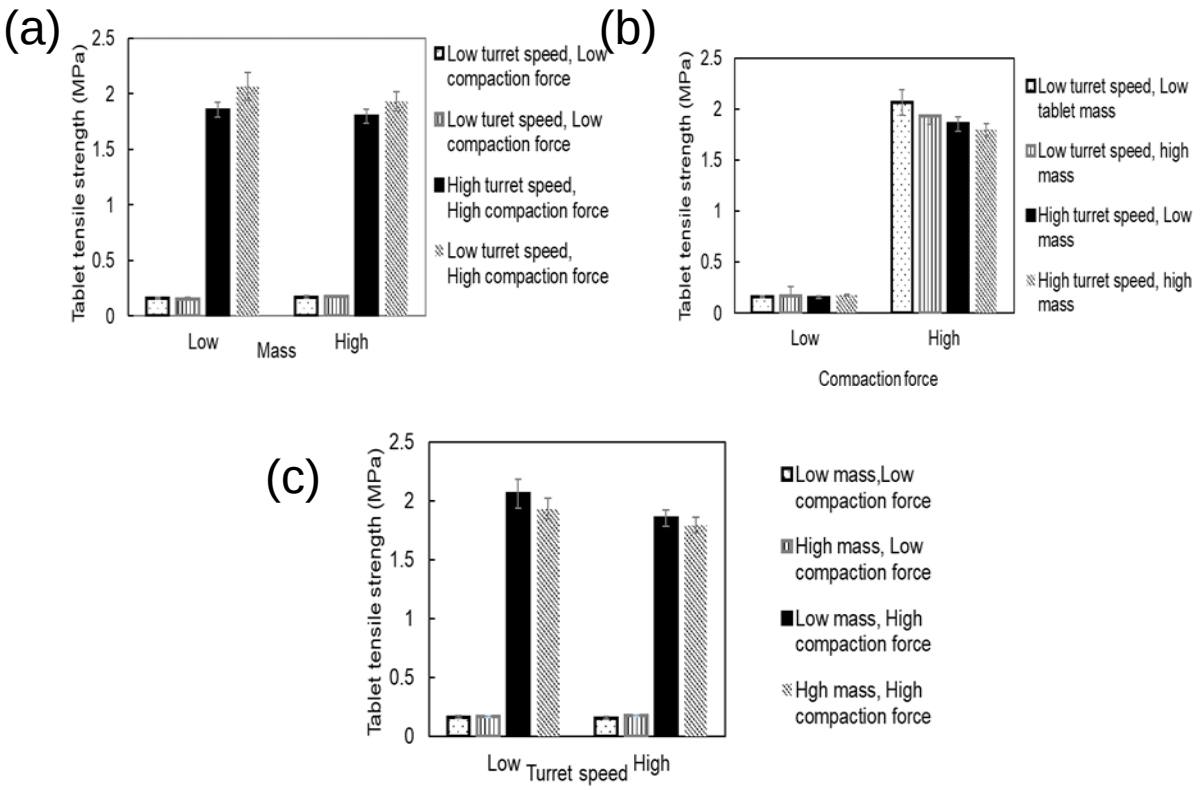


Figure 3.15 Effect of different process parameters on tablet tensile strength: (a) Single tablet mass(mg), (b)Compaction Force (kN) and (c) Turret speed (rpm).

3.5 IoT Historical Energy data

A primary objective of this research was to establish a seamless connection between physical assets and the cloud, enabling real-time data management and analysis. To achieve this, two Siemens nano boxes were integrated with the Perceptive Engineering's PharmaMV control system, facilitating the transmission of data from all process units to the Insights Hub account. The Siemens Insights Hub, an industrial IoT as-a-service platform, was employed for real-time data acquisition, monitoring, and analysis, and to facilitate the creation of bespoke applications. Data generated from the physical process is captured by Perceptive Engineering's PharmaMV and uploaded to the cloud in real-time, providing accessible and actionable insights to users both online and offline.

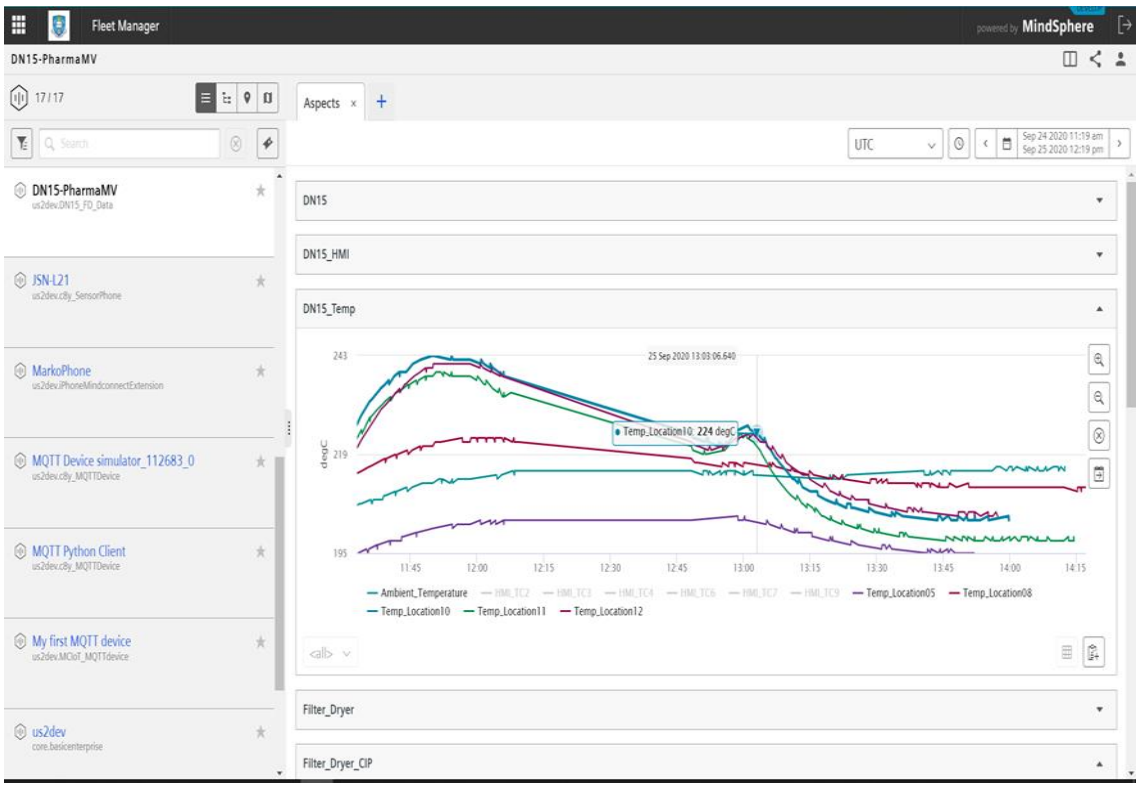


Figure 3.16 Example of real-time readings of the DiPP Data from MindSphere (Ntamo et al,2022).

Process data were acquired from previous DiPP campaign runs csv folders showed in Figure 3.16. When the DiPP is turned on, it employs its sensors to collect real-time readings, which it then outputs in Microsoft Excel spreadsheet format. Our investigation utilised ten campaign spreadsheets from prior DiPP runs, including undergraduate lab sessions and DiPP test runs conducted by maintenance personnel. For this data study, Python version 3.10 was used, together with the library's pandas, scipy, scikit, and matplotlib, to execute all the necessary analyses and plot the findings. The Python energy model code is structured into a series of steps to process and analyse the data from past DiPP runs:

- Data Import: Excel spreadsheets containing historical data from DiPP campaigns are imported into the Python environment.
- Data Cleaning: Unnecessary columns, such as those related to irrelevant unit operations, are removed from the dataset to streamline the analysis.
- Data Concatenation: Relevant data from multiple spreadsheets is combined into a single, unified data frame for efficient processing.
- Data Filtering: The large data frame is iterated through to identify and filter

out periods of inactivity in the TSWG unit.

- Data Slicing: The filtered data frame is divided into smaller, focused datasets, or "slices," each representing a period of unit activity. These slices contain all relevant data points for further analysis.

This structured approach to data processing lays the foundation for subsequent Twin screw granulator energy modelling and analysis in Figure 3.17.

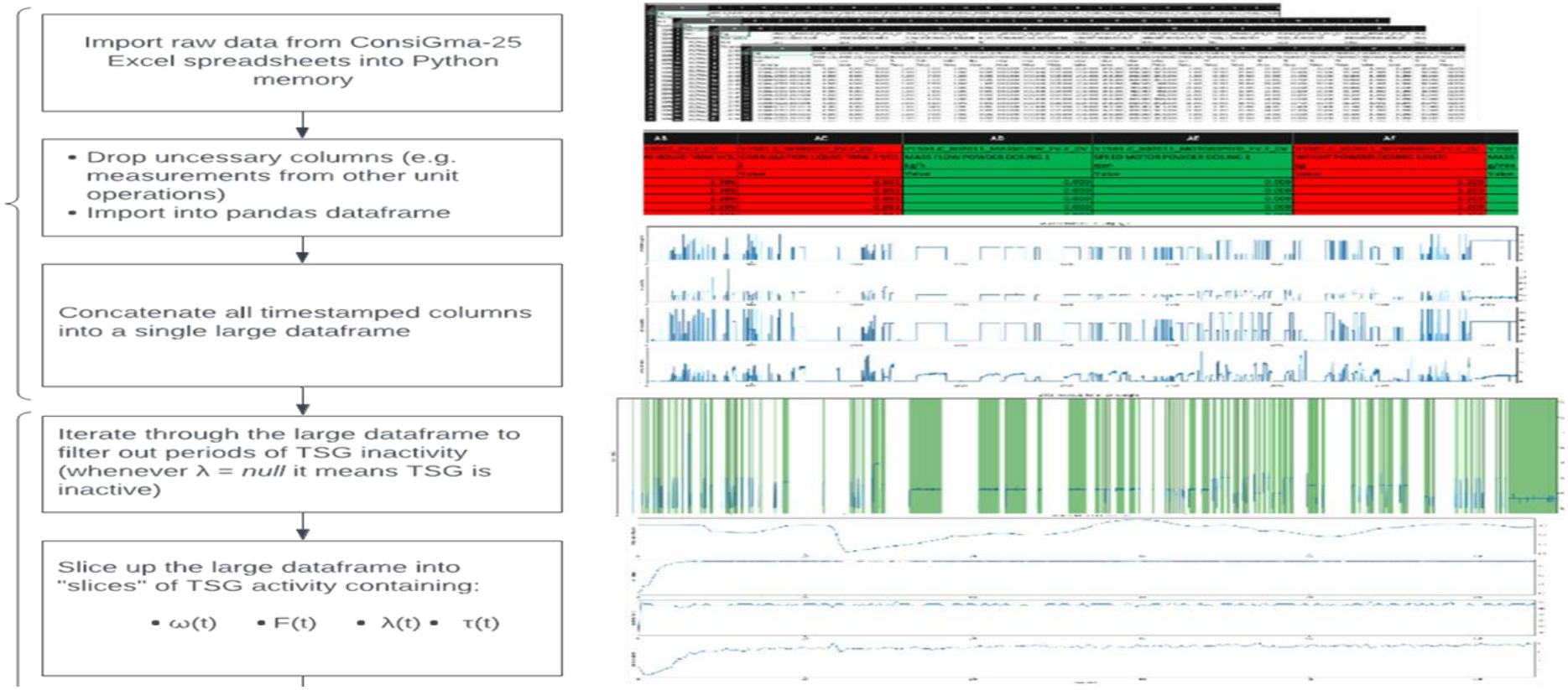


Figure 3.17 Workflow for TSWG energy modelling. Using the τ (t) curve from each slice, perform y = k × ln (x + 1) regression on each torque profile curve. The purpose is to evaluate a range of regression models, including SVR with RBF kernel, neural network, and polynomial to determine which one would offer the best fit to the data (Ntamo et al.,2025).

3.5.1 Twin Screw Granulator Energy IOT data

Table 3.6 summarises the steady-state torque values for each run. These values were obtained after running the twin screw wet granulator for 10 minutes to ensure steady state operation. The results indicate a strong influence of powder feed rate on torque. Further detailed analysis planned for later chapters. The results of torques were taken every 2 minutes for 10 minutes and averaged to ensure consistent results. Low screw speed of 200 rpm combined with a high powder flow rate above 10 kg/h results in higher torque above 7 Nm as shown in Table 3.6. The Maximum Liquid to Solid ratio (L/S) is 0.45 to prevent paste formations.

Table 3.6 Twin screw wet steady state torque results.

Run	*Powder Feed Rate (kg/h)*	*Liquid to Solid Ratio*	*Liquid flow rate (g/min)*	*Screw Speed rpm*	*Torque Nm*
1	12.5	0.3	62.50125	550	3.2
2	5	0.4	33.334	900	3.3
3	5	0.2	16.667	200	2.8
4	5	0.4	33.334	900	3.1
5	10	0.2	33.334	200	7.6
6	20	0.2	66.668	900	2.3
7	20	0.4	133.336	900	4.5
8	10	0.2	33.334	200	7.4
9	12.5	0.3	62.50125	550	3.3
10	20	0.2	66.668	900	2.5
11	5	0.2	16.667	900	2
12	5	0.4	33.334	200	3.4
13	10	0.4	66.668	200	7.7
14	5	0.4	33.334	200	4.5
15	5	0.2	16.667	900	2.1
16	20	0.4	133.336	900	5
17	5	0.4	33.334	550	2.7
18	5	0.2	16.667	550	2.1
19	20	0.2	66.668	900	3
20	10	0.4	66.668	200	10

Preliminary data analysis

Preliminary data analysis was conducted to identify the factors exerting the strongest influence on torque, a crucial parameter for calculating the energy consumption of the twin-screw granulator. Linear regression was applied to the

torque data presented in Table 3.6. To quantify the contribution of each factor and their interactions, a two-way ANOVA was performed.

As with simple linear regression and one-way ANOVA, the two-way ANOVA requires certain assumptions:

- The data points are independent.
- The residuals are normally distributed.
- The variance of the residuals is similar at different levels of the factors.

Table 3.7. Torgue regression model R square values.

SUMMARY OUTPUT	
Regression Statistics	
Multiple R	0.7002589
R Square	0.4903625
Adjusted R Square	0.1930739
Standard Error	2.0471475
Observations	20

Interactions up to the third order were considered. While the adjusted R-squared value of 0.5 in Table 3.7 is not ideal, it is sufficient for identifying the primary activity metric or driving factor. The ANOVA results (Table 3.8) indicate that only powder flow rate and screw speed have a statistically significant impact on torque (p-value < 0.05).

However, the significance of the regression output over the residual line plot suggests the need for either more data or a more sophisticated regression method. To address this, further investigation will involve the integration of IoT data and advanced regression algorithms, which will be explored in subsequent sections.

Table 3.8 Test of Between-Subjects Effects

	ANOVA				
	Degree of Freedom	*Sum of Square*	*Means of Square*	*F*	*Sig*
Regression	7	48.39	6.91	1.65	0.21
Residual	12	50.29	4.19		
Total	19	98.68			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	
Intercept	5.03	0.61	8.30	2.57E-06	
Powder Feed Rate	2.00	0.91	2.21	0.048	
Liquid Flow Rate	0.58	0.70	0.83	0.42	
Screw Speed	-1.89	0.67	-2.80	0.016	
Powder Feed Rate × Liquid Flow Rate	0.34	0.94	0.35	0.73	
Powder Feed Rate ×Screw Speed	-1.61	0.92	-1.74	0.11	
Liquid Flow Rate ×Screw Speed	-0.55	0.70	-0.79	0.45	
Powder Feed Rate × Liquid Flow Rate ×Screw Speed	-0.29	0.97	-0.30	0.77	

3.5.2 Fluidised bed dryer IOT data

Figure 3.18 and Figure 3.19 below exhibit a visual depiction of the data acquired during the plant runs the DiPP Consigma-25 fluidised bed dryer. The temperature of the product in a fluidised bed dryer closely approaches the temperature of the dryer's cells due to heat transfer. As the drying process progresses, thermal equilibrium is reached, minimising the temperature difference between the granules and the surrounding environment. Figure 3.18 exhibits that initially the temperature of the dryer for cells 1 to 3 are all at around 40°C. This is because the temperature of the vapour fed into the fluidised bed dryer is run until it reaches a constant temperature. The constant temperature for the case below is 40°C. From the start of operation, the temperature drops to its lowest after 260 seconds. This temperature drop is a result of the gradual addition of wet granules into the fluidised bed dryer. The wet granules fed into the cell are at a temperature of 21.8°C which was recorded before they were fed into the fluidised bed dryer. This

lower temperature of the wet granules being fed into the cells drops the temperature to 22.4°C.

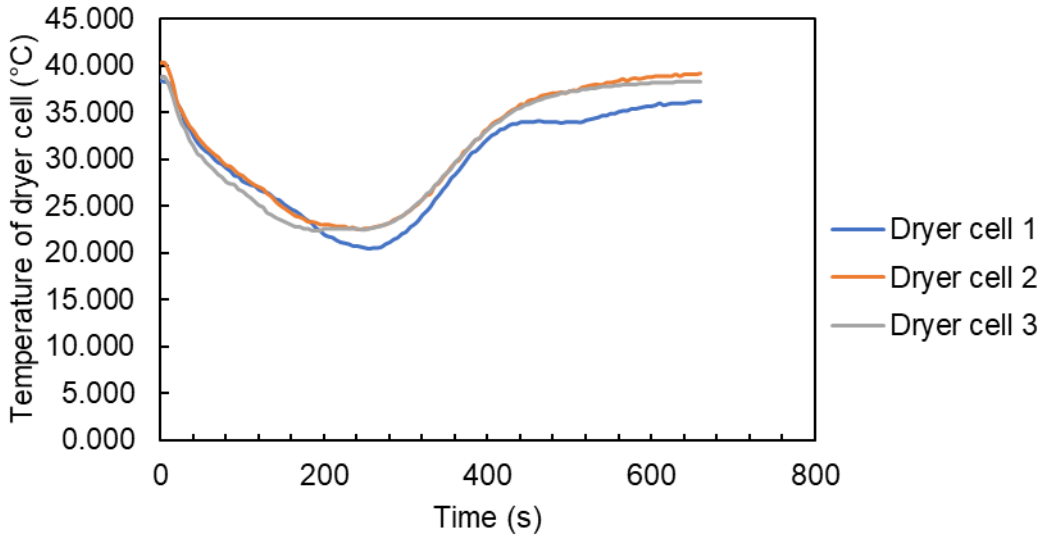


Figure 3.18 Temperature of dryer cell over time with drying air temperature at 40oC and L/S ratio of 0.18.

The temperature of the cells starts to rise around 250 seconds and continues until 620 seconds. At this point, the granules' moisture content starts to drop, which causes their temperature to rise. Following 620 seconds, the temperature begins to stabilise and approaches the hot vapour's temperature consistently. At this point, the temperature of the hot vapour and the moist granules reach a thermal equilibrium. As a result, causing the temperature of the cell to level off. At the liquid/solid ratio of 0.18 each repeat of the experiment at the same temperature shows a similar trend with little variation. These variances result from the fluidised bed dryer's inconsistent insulation. Therefore, the irregularities result from heat escaping from the unit, which implies that the heating process is not totally consistent. The dryer cell's temperature does not quite reach the vapour temperature because heat dissipation to be the surrounding.

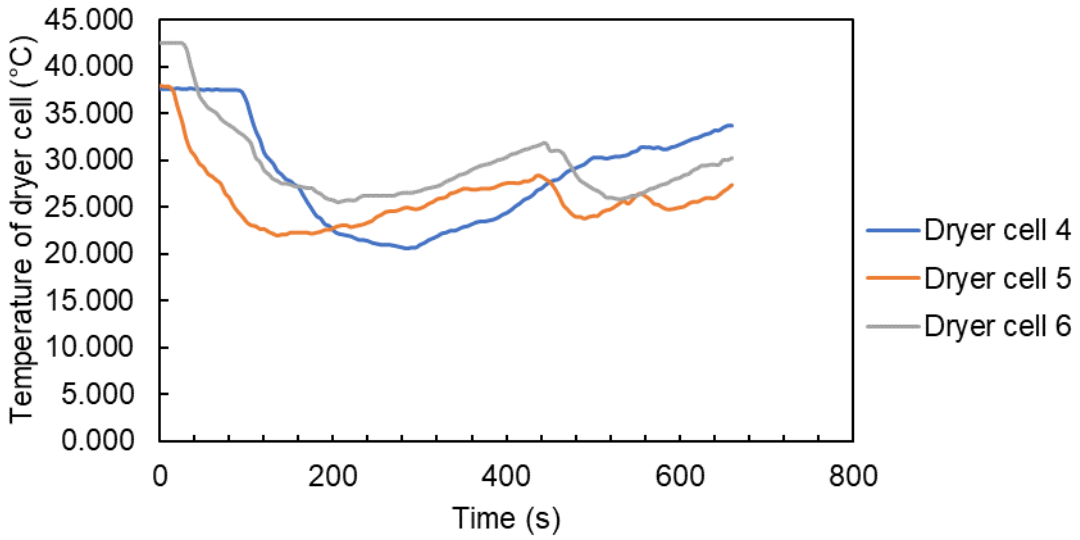


Figure 3.19 Temperature of dryer cell over time with vapour temperature at 40oC and L/S ratio of 0.3.

Figure 3.19 depicts the temperature profiles of the fluidised bed dryer cells when operating with a liquid- solid ratio of 0.3 and a vapour temperature of 40°C. More liquid is introduced while maintaining a consistent granule quantity. The primary patterns in Figure 3.18 and Figure 3.19 are similar, showing a temperature decline at first followed by an increase to the vapour temperature. Nevertheless, the cell's final temperature decreases with a higher L/S ratio. Additionally, the experiment's variability increases significantly with each repetition. This is since a growing L/S ratio makes it more difficult to understand processes like wall deposition and particle aggregation

3.6 Data Contamination

Bad data can arise from a variety of sources such as typos or incorrect input, and software-related issues, such as inconsistent data structures, format mismatches, or missing data. These errors can significantly impact the quality and reliability of data. Additionally, software issues like inconsistent data structures or format mismatches, or even missing data, can further compromise data quality.

To mitigate the impact of bad data and improve data analysis, techniques like box plots can be employed. A box plot can be used to visually represents the distribution of a dataset, providing insights into its central tendency and dispersion. By calculating key statistical measures such as the minimum, maximum, and interquartile range (IQR), box plots effectively highlight outliers

and anomalies within the data. This visual representation aids in identifying potential data quality issues and making informed decisions regarding data cleaning and preprocessing.

The box plot approach begins by sorting and scoring the residuals. The sorted residuals are then divided into four groups of equal size. Each category contains 25% of the total residuals. The lines separating the groups are known as quartiles. The box plot graph algorithm analyses each value in a dataset and flags "High Outlier" and "Low Outlier".

Table 3.9 Sieve analysis Granule PSD for twin screw wet granulator with operating at screw speed of 900 rpm, liquid to solid ratio of 0.4 and powder feed rate of 20 kg/h.

		Statistics	
		Run 3	Run 3 repeat
Mean		0.14	0.14
Std. Deviation		0.14	0.16
Variance		0.020	0.026
Range		0.35	0.43
Percentiles	25	0.043	0.023
	50	0.092	0.082
	75	0.33	0.30

Table 3.10 Hardness tester tablet strength results for tablets produced at operating turret speed of 20 rpm, single tablet target mass liquid of 500 mg and compaction force of 2kN.

Exp 1 repeat 2			*Exp 1 repeat 3*		
Tablet strength (MPa)					
N	Valid	1.00E+01	N	Valid	1.00E+01
	Missing	0.00E+00		Missing	0.00E+00
Mean		1.56E-01	Mean		1.63E-01
Std. Deviation		9.92E-03	Std. Deviation		2.01E-02
Variance		9.85E-05	Variance		4.04E-04
Range		2.95E-02	Range		6.51E-02
Percentiles	25	1.48E-01	Percentiles	25	1.48E-01
	50	1.54E-01		50	1.59E-01
	75	1.63E-01		75	1.74E-01

Using the box plot approach, twin-screw granule experimental outliers for run 3 would be found at mass fraction larger than 0.86 and less than -0.10 (or zero), and tablet press experimental outliers for run 1 at tensile strength greater than 0.21 and less than 0.10. But this is a rudimentary method of finding anomalies. A digital twin deal with real-time data, large data, and soft sensors; thus, a more adaptable and reliable approach is needed.

3.6.1 Tablet press Autoencoders

Neural networks, including autoencoders, and different deep learning architectures are often used to detect anomalies in complex datasets. LSTMs, a form of RNN, are frequently employed in sequence-based anomaly detection. They are well-suited to jobs requiring the order and relationships of data points, making them useful for time-series data. GANs are made up of a generator and a discriminator that is adversarially trained. In the context of anomaly detection, the generator seeks to provide data that is comparable to regular training data, whereas the discriminator distinguishes between actual and created data. Anomalies can be found by examining situations in which the discriminator fails to distinguish between genuine and produced data. These deep learning approaches can handle high-dimensional and complicated datasets, making them appropriate for anomaly identification in numerous fields including finance, cybersecurity, healthcare, and industrial processes. Implementation details and model hyperparameters may vary based on the specific characteristics of the data and the nature of the anomalies you are trying to detect. R. D. Cook created Cook's squared distance, another popular approach for diagnosing outliers, in 1977(Cook, 1977).

$$\mathrm{CD}_{(i)}^{2} = \frac{\sum_{j=1}^{n} \left(\hat{y}_j - \hat{y}_{j(i)}\right)^2}{p \times MSE} \qquad 3.2$$

Where:

- $\hat{y}_j$ is the predicted value for the $j(th)$ observation,
- $\hat{y}_{j(i)}$ is the predicted value for the $j(th)$ observation without the $i(th)$ observation in the model,
- p is the number of predictors, and
- MSE is the Mean Squared Error of the model.

This demonstrates how each observation might alter the regression model. A high Cook's squared distance $\mathrm{CD}_{(i)}^{2}$ suggests that the $i(th)$ observation has a significant impact on regression coefficients and may be regarded an outlier. A CD value greater than 1 is generally considered an anomaly. The fundamental disadvantage of Cook's squared distance is that it is sensitive to outliers at the extremes of the data range but fails to detect outliers in the middle of the range.

Like R. D. Cook method, Autoencoders are neural networks that learn to encode and decode input data (Torabi et al., 2023). During training, the network learns how to recreate regular patterns. Anomalies can then be identified by looking for high reconstruction errors in situations that differ greatly from the taught patterns (Torabi et al., 2023). Synthetic data was generated from both anomalous and normal datasets using a Standard Scaler for standardisation. A basic autoencoder neural network was designed and implemented in Python version 3.10.11 using Keras. The model is trained on normal data to provide a representation. Reconstruction errors i.e. RMSE are computed throughout the full dataset. Anomalies are recognised using a predetermined threshold.

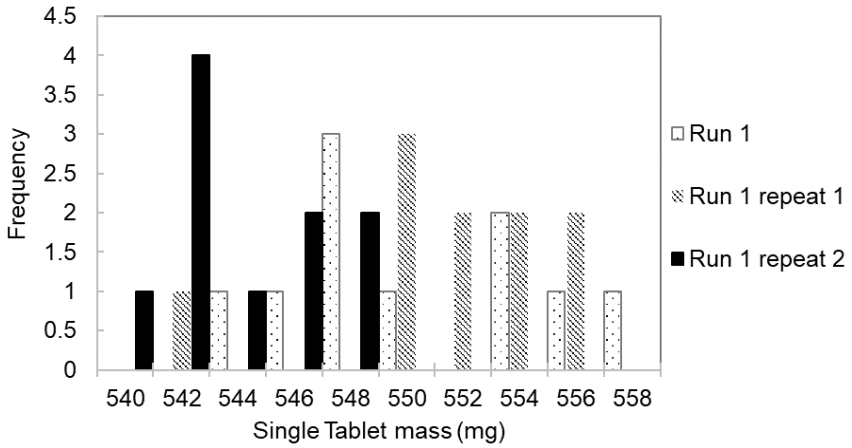


Figure 3.20 Hardness tester mass results for tablets produced at operating turret speed of 20 rpm, single tablet target mass liquid of 500 mg and compaction force of 2kN.

For each run shown in Figure 3.20 , ten samples were collected and examined. To confirm that the results were repeatable, the experiment was performed three times and re-analysed. The graph depicts the outcomes of a single-run tablet mass distribution. The three-distribution graphs are different, indicating that there is another variable influencing tablet strength that must be controlled.

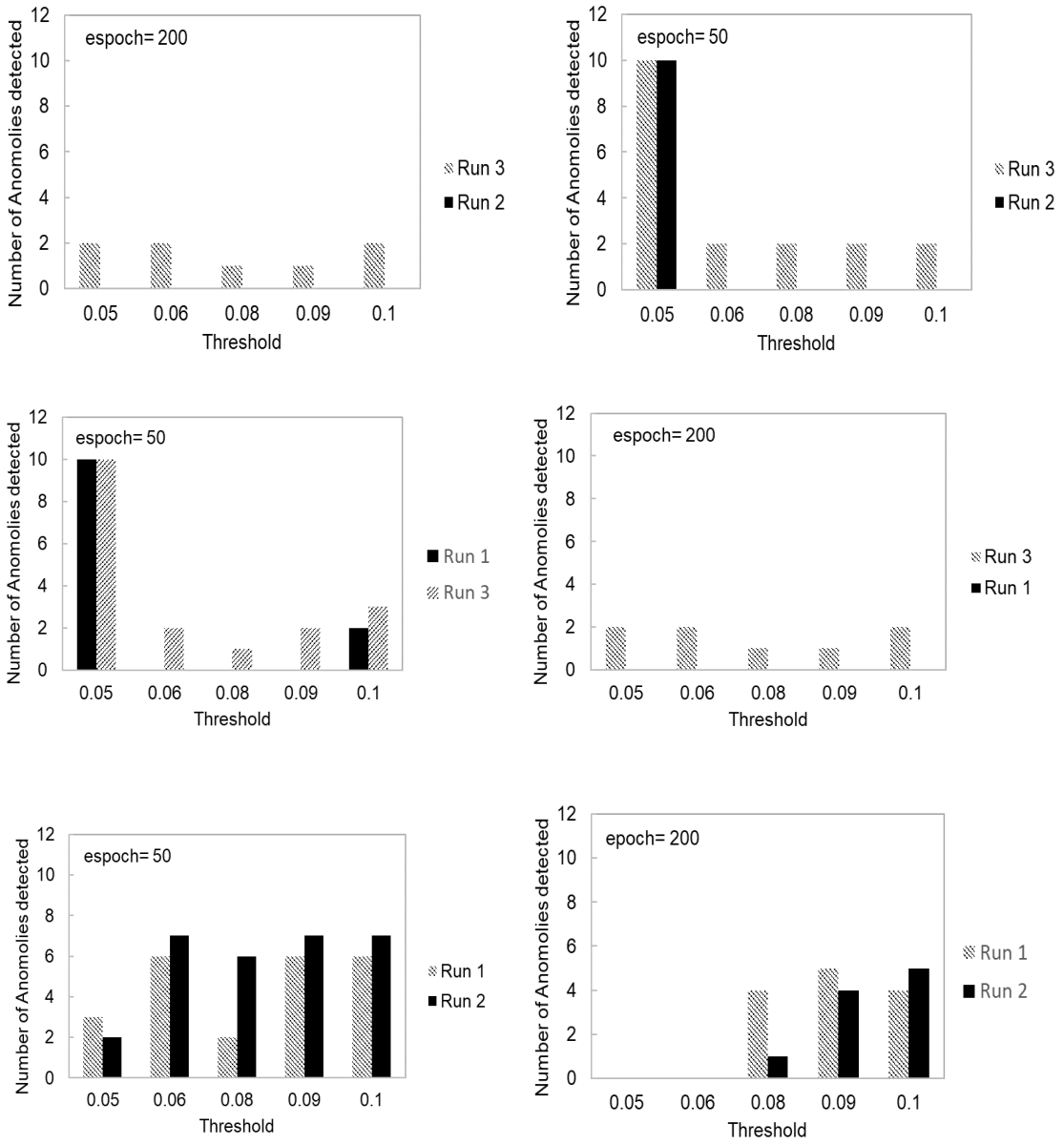


Figure 3.21 Number of anomalies detected by autoencoder at different threshold and epoch per repeat.

An autoencoder found in appendix A was used to detect and delete outliers before calibration. Two dummy data points (7,2) were added in run 3 tensile strength results. The results of autoencoder outlier identification are presented in Figure 3.21. By varying the threshold and epoch, the autoencoder may detect anomalous data as well as outliers. Increasing the threshold has no significant impact on the amount of anomalous data detected. Conversely, a lower threshold, such as 0.05, identifies more data points as anomalies because it is stricter than the Root Mean Square Error (RMSE). The two bad data were identified in every instance showing how full prove this approach is. When autoencoder identification was performed between runs 2 and 1 in the absence of dummy outliers, run 2 data points were classified as anomalies more than run

1 data points, implying that the two runs are statistically distinct. Changing the auto encoder's threshold allows you to find distributions that are statistically distinct in addition to finding outliers.

3.7 Chapter Conclusion

Digital twins rely on information from the product they are simulating, both in the original design and ongoing modifications. However, working with many models and regularly updated data requires managing a massive amount of data, which must be kept, tagged, and traceable. This requires assets to be versioned and monitored during development. Consumer-focused manufacturers of formulated goods recognise the importance of transitioning to smart and continuous production but are sceptical about digital technologies due to the absence of industrial-scale demonstrations complying with standards like 21 CFR Part 11. A new framework was built and tested at the DiPP; an industrial-scale demonstrator compliant with regulations.

This chapter has explored the multifaceted approach to data sourcing that empowers the deployment of models for digital operation. Data was meticulously gathered from three key sources:

- **IoT-connected devices:** By leveraging sensor technology embedded within machines and throughout production lines, real-time operational data was captured. This data provides valuable insights into machine health, process performance, and environmental conditions.
- **Literature review:** Existing research and industry best practices documented in scientific publications and reports were carefully examined. This knowledge base serves as a foundation for understanding established techniques and identifying potential gaps that machine learning models can address.
- **Experimentation:** Controlled experiments were conducted to generate specific data sets tailored to the needs of the digital operation models. This allows for focused training and validation of the models, ensuring their effectiveness in real-world scenarios.

Following this comprehensive data collection effort, the raw data underwent rigorous analysis and pre-processing steps. This crucial stage involved

techniques like data cleaning to remove inconsistencies and errors, data transformation to ensure consistency in format and units, and feature engineering to extract relevant features that will be most useful for model training. By meticulously preparing the data, I ensured its seamless integration into the digital architecture, enabling the models to learn from the rich information and deliver accurate and insightful results.

In robust design, the general assumption is that experimental data are normally distributed without contamination. However, this assumption may not hold in practice, and an outlier-resistant approach is needed to address this problem. Autoencoder, compared to Cook's squared distance, is fast and adaptive, making it a useful technique to accelerate the deployment of models to digital operation.

Chapter 4: Enhancing Model Development Speed

4.1 Introduction

Despite the theoretical feasibility of rigorous, first-principles models for most processes, practical challenges frequently impede their real-time implementation in digital operations. Specifically, these models can be computationally intensive and lack robustness, hindering real-time implementation. This chapter addresses these issues by accelerating the digital deployment of high-fidelity models for twin-screw wet granulation.

Global sensitivity analysis was used in this study as an essential component for successful process model development and calibration efforts, particularly when involving many inputs and kinetic variables that can affect model outputs. The remainder of the chapter is focused on validating the results of global sensitivity analysis and coming up with better calibration strategies. Finally, the limitations associated with the development and application of the rigorous model will be discussed, followed by a presentation of future research directions, and concluding remarks.

The proposed model development is divided into two parts. The first step focuses on defining the problem, setting up the model, and formalising the model training problem. This step requires some prior process expertise. This information must be used to identify critical process variables and select an appropriate model structure, considering both known and unknown dynamics within the system. If the trained model proves to be insufficiently accurate, it is most likely owing to assumptions established during the initial step. The essential process variables are identified, assessed in terms of sensing technology and sampling frequency, and then classified. The main model structures are established, including preprocessing functions, governing equations, and constitutive equations.

The second stage of the modelling framework focuses on data gathering, model training, and prediction. This phase is fully dependent on the requirements and model structures established in the first phase of the framework and will thus require no further inputs. This step may be automated if data gathering is possible utilising online and/or inline sensors. The model is trained, verified, and tested using pre-processed data, yielding an estimate of the model calibration

parameters. The model calibration parameters are utilised to make forecasts using raw sensor data and programmed control actions.

GSA is crucial in process modelling by systematically identifying the most influential design parameters (Jung and Taflanidis, 2023). By identifying these critical parameters, the engineering team can prioritize their efforts to optimize the factors that have the most significant impact on the final product and the production process. This approach offers valuable insights into system behaviour and can be utilised to guide various tasks, including dimensionality reduction, uncertainty quantification, and optimal decision-making (Jung and Taflanidis, 2023). Furthermore, the systems engineering approach incorporates feasibility analysis, enabling a comprehensive evaluation of various design options. This analysis considers not only technical feasibility but also economic and logistical factors. By rigorously assessing the practicality and cost-effectiveness of potential solutions, the team can make informed decisions that ensure a robust and cost-efficient process design.

4.2 Material and Methods

Siemens's gPROMS FormulatedProducts (gFP) 2023.2.1 has been widely designed and employed for simulations and digital applications in the pharmaceutical industry and has been popularly referred to in several publications. While academic research like "Modelling of Particulate Processes for the Continuous Manufacture of Solid-Based Pharmaceutical Dosage Forms" explores new methods, Siemens' gPROMs offers a mature solution backed by extensive industry experience (Rogers AJ. Et al, 2013 ; Matsunami, K. et al., 2023 ; Davidopoulou, C. and Ouranidis, A., 2022). Their global reach and Fortune 500 customer base solidify their position as a reliable partner (PSE: Sanofi Joins Systems-based Pharmaceutics Alliance,2019). Additionally, compliance with regulations like EU GMP and FDA 21 CFR Part 11 ensures data integrity and system security, making gPROMs a trustworthy choice for digital twin implementation in the pharmaceutical industry (Good Manufacturing Practice, Siemens.com).

gFP 2023.2.1 offers models for simulating and optimising pharmaceutical processes, including various types of granulators, such as twin-screw

granulators, fluidised bed dryers, dry mills, and tablet press. By integrating the process model developed in gFP with the online PharmaMV suite, the study can leverage Digital Model Predictive Control (DMPC) to enhance process control and optimisation. Furthermore, it provides optimisation functionalities allowing for the minimisation of objective functions. The software also employs high computing and parallel worker that would allow the production for than 2000 simulations per sitting quickly for GSA.

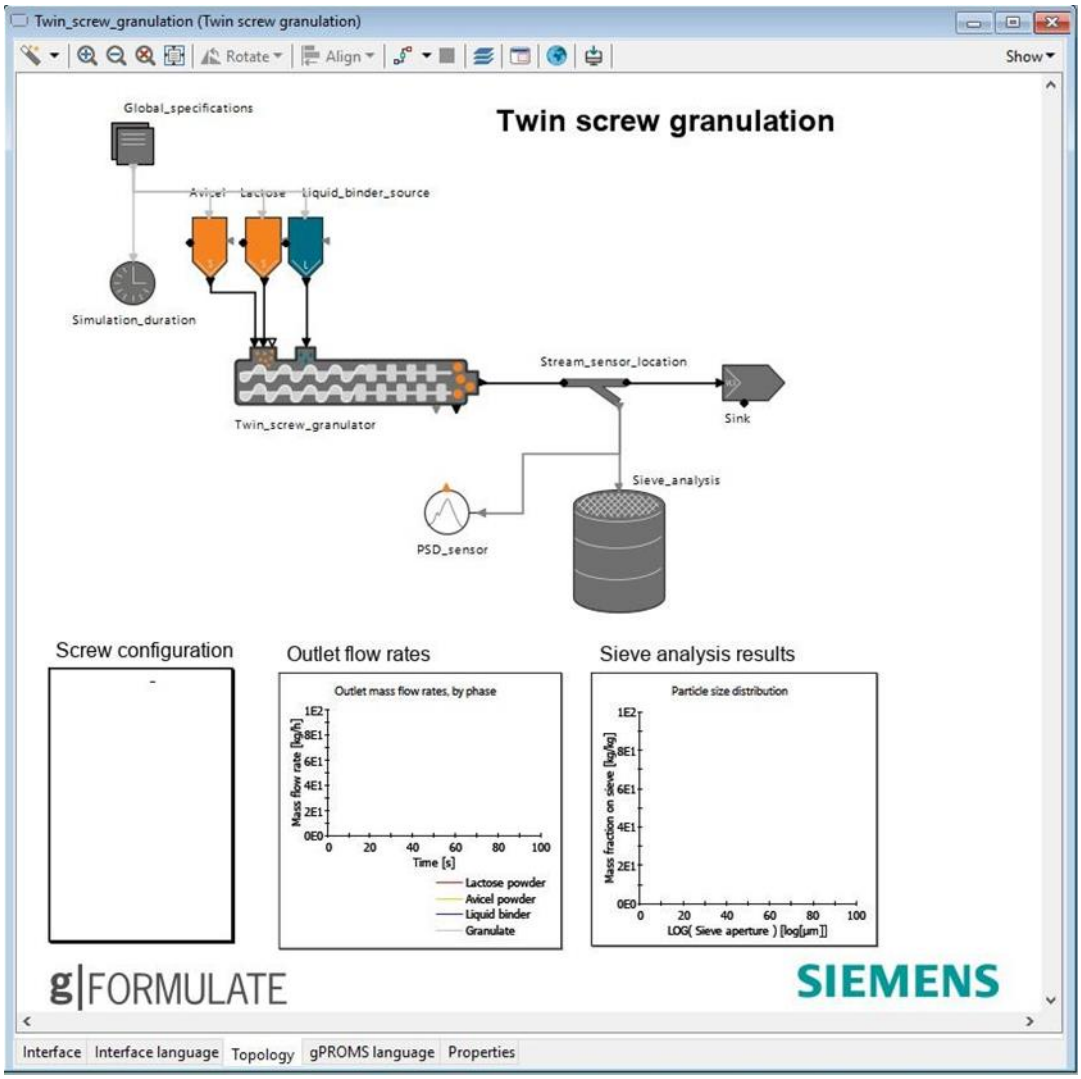


Figure 4.1 gPROMS FormulatedProducts flowsheet.

The flowsheet illustrated in Figure 4.1 features a PSD sensor unit, which allows for the determination of the outlet granule PSD.

4.3 Mechanistic Model Development

Given the screw configuration's significant impact on granule size distribution, a compartmental approach is typically used to evaluate material transport along the barrel. Each axial compartment is assigned an average residence time based on the screw elements. The inlet flow rate in one compartment is equal to the outlet flow rate in the last neighbouring compartment as described by Equation 4.1.

$$\dot{M}_{out} = \frac{M}{\tau} \tag{4.1}$$

Where $\dot{M}_{out}$is the outlet flow rate for each compartment; M is the mass holdup and τ is the residence time in each compartment. The inlet flow rate of a compartment is equal to the outlet flowrate of the previous one. However, there is an exception for the first compartment where the inlet flow rate is specified based on the inlet feed rate of materials. The following models, selected to describe the PBM rate mechanism, are grounded in the fundamental processes of nucleation, breakage, layering, and consolidation. These models represent advancements over previous iterations, incorporating updated mechanisms to provide a more comprehensive understanding of the PBM process.

Drop nucleation, adapted from Barrasso and Ramachandran (2015), involves the formation of droplets with a lognormal size distribution, which penetrate a bed of fine powder and replace interparticle voids. The probability density function $f(\mu_{\text{drop}})$ is given as:

$$f(\mu_{\text{drop}}) = \frac{1}{\mu_{\text{drop}}} \frac{1}{\sigma_{\text{drop}}\sqrt{2\pi}} \exp\left(-\frac{\left(\ln \mu_{\text{drop}} - \mu_{\text{drop, mean}}\right)^2}{2\sigma_{\text{drop}}^2}\right) \tag{4.2}$$

Where $\mu_{\text{drop, mean}}$ is the mean droplet size and σ_{drop} is the standard deviation. The rate of nuclei formation in compartment i, $B_{nuc,i}$ is determined by droplet size μ_{drop}, bed porosityϵ_{bed}, and maximum pore saturation using Equation 4.3 by assuming that the rate of nuclei formation is equal to the rate of droplet addition to the powder fines (Barrasso and Ramachandran, 2015).

$$B_{nuc,i} = \frac{\dot{m} l_i}{V_{\text{drop}}} \delta(v - v_{nuc}) \tag{4.3}$$

Where $\dot{l}_i$ is flow rate of liquid added to the compartment I, $\dot{m}$ is mass in the compartment, V_{drop} is volume of a single liquid drop, v is granule volume and v_{nuc}

is saturated granule of volume. The nuclei volume is determined from Equation 4.4 which assumes that each nucleus is fully saturated with liquid and has a porosity equal to that of the bed.

$$v_{nuc} = \frac{V_{\text{drop}}}{s^* \epsilon_{\text{bed}}} \qquad 4.4$$

Where $s^*$is the droplet pore penetration and ϵ_{bed} is bed porosity

Granule breakage, modelled after Wang et al. (2020), considers chipping and fragmentation mechanisms in conveying and distributive mixing elements. A Weibull distribution is used to describe the cumulative breakage function $P_{b,KE}$for the distributive kneading element due to its versatility in modelling various failure or breakage behaviours, while a bimodal distribution represents the course and fine particles resulting from chipping in the conveying element (Wang et al., 2020).

Conveying element

The breakage probability $P_{b,CE}$ was found to increase with an increase in Powder feed number, PFN and granule size and a decrease in dynamic yield stress of the granule, DYS. Taking these parameters into account along with the upper and lower granule size, x_{lc} and x_{uc}limits for breakage, an empirically-derived selection function for conveying elements CE may be written as Equation 4.5.

$$P_{b,CE} = \begin{cases} 1 & x \geq x_{uc} \\ a \times exp\left(-\frac{DYS}{b} \times \frac{1}{PFN}\right) \times (x/x_{lc})^c & x_{lc} < x < x_{uc} \\ 0 & x \leq x_{lc} \end{cases} \qquad 4.5$$

Where a , b and c are fitting parameters in breakage kernel and are lower and upper granule size bound. To represent the distribution of course and fine particles that result from chipping in CEs, a bimodal distribution is proposed:

$$B_{M,CE} = \nu B_{M1} + (1 - \nu) B_{M2} \qquad 4.6$$

where ν is the proportion of fine and coarse particles and B_{M1} and B_{M2} are the modified Weibull distributions.

Kneading element

The breakage probability in Kneading element, $P_{b,KE}$ is found to be a strong function of the screw geometry, DYS, and the powder feed number (PFN). We

propose the following empirically derived three parameter model for the breakage selection function.

$$P_{b,CE} = \begin{cases} 1 & x \geq x_{uc} \\ 1 - a' \times exp\left(-PFN \times \frac{b'}{DYS} \times \left(\frac{x}{x_{lc}}\right)^{c'}\right) & x_{lc} < x < x_{uc} \\ 0 & x \leq x_{lc} \end{cases} \quad 4.7$$

A modified Weibull distribution is proposed to describe the breakage fragment size distribution via fragmentation and is described by Equation 4.8

$$B_{M,KE} = exp\left(-\beta \times \left(\frac{x}{x_{uc}}\right)^{\gamma}\right) \quad 4.8$$

where $B_{M,KE}$ is the cumulative size distribution and β is the scale parameter and γ is the shape parameter in the Weibull function.

Layering, based on Cameron et al. (2005), involves the deposition of fine powder particles onto wet granule surfaces, leading to an increase in granule size and described by the Equation 4.9 below:

$$G_i(x) = G_m \frac{M_{i,P}}{kM_{i,g} + M_{i,p}} exp\left[-\lambda\left(y_{w,i} - y_{wc}\right)^2\right] \quad 4.9$$

Where $G_i(x)$ is the linear size-based growth rate, G_m is layering maximum linear growth rate, λ and k are fitting parameter in layering kernel, $M_{i,g}$ is Mass of granules in compartment i ,$M_{i,p}$ is mass of fine powder in compartment i, $y_{w,i}$ is moisture content and y_{wc} is critical moisture content. Granule growth rate by layering, $G_i(v)$ related to linear size-based growth rate using Equation 4.10.

$$G_i(v) = 6.9v^{\frac{2}{3}}G_i(x) \quad 4.10$$

The consolidation model, adapted from Iveson et al. (1996), accounts for the reduction in particle size due to decreasing internal porosity. This model considers the combined effect of intra-granule void and internal liquid, leading to a decrease in liquid and internal pore space. The mass of other granulate components remains conserved, and the loss of liquid from the granulate phase is balanced by an increase in external liquid (Iveson et al., 1996). The consolidation rate is characterised by an exponential decrease in porosity ε_i, approaching a minimum value ε_{min} (Cameron et al., 2005) and described by 4.11.

$$R_i(\varepsilon) = -k_{i,comp}(\varepsilon_i - \varepsilon_{min}) \quad 4.11$$

Where $R_i(\varepsilon)$ is rate of change of granule porosity in compartment i and $k_{i,\,\text{comp}}$ is compaction rate constant, s^{-1}

A total of 25 parameters listed in Table 4.1 are to be calibrated. These parameters fall into two main categories: material properties and equipment geometry. Material properties, such as the Dynamic Yield Stress, influence the behaviour of the material during the granulation process. Equipment geometry, such as the screw design, affects the breakage and consolidation of granules.

Table 4.1 Model calibration parameters

Rate Mechanism	*Parameters to be estimated*	*Symbol*	*Units*
Consolidation	Consolidation minimum porosity	ε_{min}	-
	Consolidation rate constant	k_i	s^{-1}
Layering	Layering critical moisture content	y_{wc}	$kgkg^{-1}$
	Layering kinetic parameter λ	λ	-
	Layering kinetic parameter k	k	-
	Layering maximum linear growth rate	G_m	μms^{-1}
Nucleation	Drop nucleation mean initial droplet size	$\mu_{\text{drop, mean}}$	mm
	Drop nucleation pore penetration	v_{nuc}	$m^{-3}m^{-3}$
	Drop nucleation standard of initial droplet size	σ_{drop}	mm
Breakage	Breakage rate constant in CE	$B_{M,cE}$	s^{-1}
	Breakage rate constant in KE	$B_{M,KE}$	s^{-1}
	Breakage Dynamic Yield Strength (DYS) in CE	DYS	kPa
	DYS in KE	DYS	kPa
	Breakage fitting parameter a in KE	a'	-
	Breakage fitting parameter c in KE	c'	-
	Breakage fitting parameter c in CE	c	-
	Maximum breakage granule size in KE	x_{lc}	μm
	Maximum breakage granule size in CE	x_{lc}	μm
	Minimum breakage granule size in KE	x_{uc}	μm
	Minimum breakage granule size in CE	x_{uc}	μm
	Breakage shape parameter in KE	γ	-
	Breakage scale parameter in KE	β	-
	Powder Feed Number (PFN) in KE	PFN	-
	PFN in CE	PFN	-
	Fraction of fines in CE	ν	-

The relationship between droplet size and granule size is linear, with larger droplets leading to larger granules. However, as granule size increases, there is a higher likelihood of breakage, resulting in an increase in the number of fine particles. The particle size distribution (PSD) of the granules is influenced by the droplet size distribution. A broader droplet size distribution leads to a wider granule size distribution, with a slightly lower frequency mode.

Increasing the pore saturation reduces the number of granules produced at a given liquid-to-solid ratio, as more liquid is required to form each granule nucleus. The geometric constraints of the twin-screw granulator limit the maximum achievable granule size, known as the maximum critical size. Additionally, there is a minimum critical size below which granules are not susceptible to breakage.

The screw configuration used in the experiments consisted of five compartments: three CE and two KE. Metta et al. conducted a comprehensive study on wet granulation processes, employing Global Sensitivity Analysis (GSA) to identify critical process parameters (CPPs) (Metta et al., 2019). By utilising Sobol indices, they were able to quantify the impact of various input parameters on the output variables of the process (Metta et al, 2019). The results of this analysis, presented in Figure 4.2, provide valuable insights into the factors that significantly influence the process outcomes.

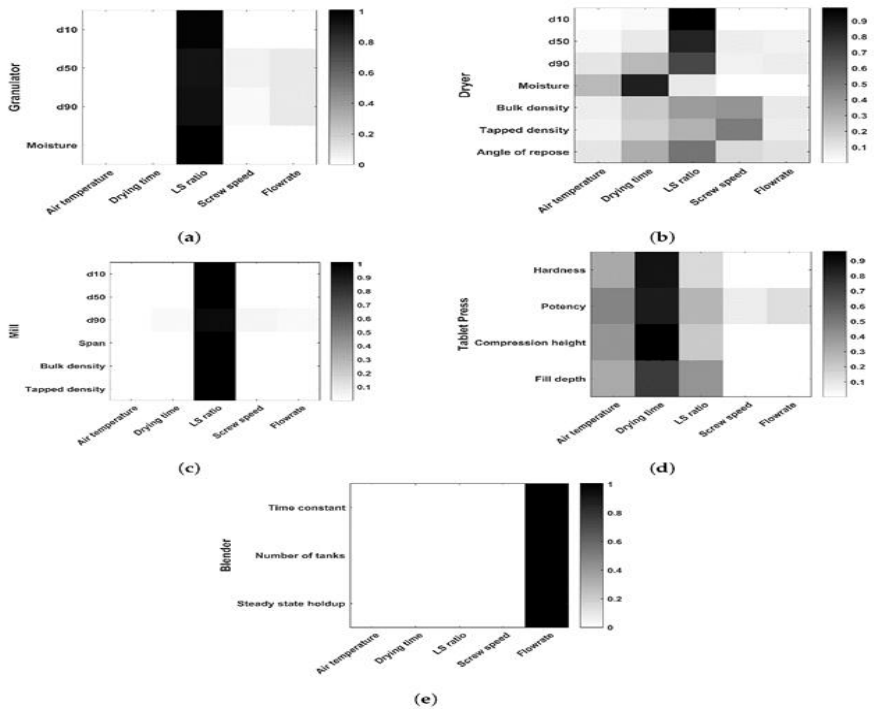


Figure 4.2 Sensitivity analysis plot for wet granulation process parameters per unit operation (Metta et al, 2019).

The three parameters (powder feed rate, L/S ratio, and screw speed) were chosen because research supports their strong influence on torque and PSD (Metta et al,2019).

Table 4.2 Twin screw wet operational boundaries.

Parameter	*Range*
Screw speed (rpm)	200 to 900
Powder feed rate (kg/hr)	5 to 25
Liquid feed rate (g/min)	15 to 208

The maximum Liquid to Solid ratio (L/S) to make granules and prevent paste is 0.45. Low screw speed combined with a high powder flow rate results in higher torque and it may exceed the safety limit of the equipment, therefore the machine may stop, and you should avoid these conditions. The blended powder, consisting of 72% lactose, 24% microcrystalline cellulose (MCC), and 4% polyvinylpyrrolidone (PVP), was used in this research for safety reasons due to its non-hazardous composition, which is commonly found in paracetamol tablets. Before being fed into the granulation particle size of the formulation was analysed using sieves to provide material properties and results and were input to gFP. While CAMSIZER offers superior detail and automation, sieve shakers provide a reliable, cost-effective, and standardised method for basic particle size analysis. For many applications where high precision shape analysis is not needed, the sieve shaker is perfectly adequate.

4.4 Rate Mechanism Global Sensitivity Analysis (GSA)

Variance-based sensitivity indices are utilised to assess the impact of individual rate mechanism parameters on critical quality attributes (CQAs) within the specified operational range. This method, grounded in Saltelli's approach, employs the formulae proposed by Saltelli et al. (2010) to estimate the sensitivity indices. By quantifying the sensitivity of CQAs to variations in model parameters, this analysis enables a deeper understanding of the underlying process mechanisms and aids in identifying the most influential factors. The indices were computed using the Monte Carlo method. A significant limitation of variance-based sensitivity analysis is its high computational cost. For this case study, a quasi-random sampling method to cover the whole operational space with over 50 to 2000 samples was used. Estimating the sensitivity indices requires many model evaluations. The main assumption when carrying out analysis is that rate

mechanisms are additive, and the factors are each interactive but are not correlated.

The study conducts sensitivity analysis on each rate mechanism based on the defined process kernels in PBM given below while holding all other rate mechanisms constant. Lower and upper bounds are defined by shifting each parameter by 20% of their initial reference value taken from the literature (Wang et al, 2021). The process parameters were specified within the equipment's operational limits, i.e. screw speed operating conditions of 200- 900 rpm, powder feed rate operating conditions of 5-25 kg/hr, and Liquid to solid ratio operating conditions of 0-0.45. The DEM-informed parameters were carefully adjusted to cover a wide range of values obtained from numerous DEM simulations documented in the literature. This broad parameter range ensures that the model can capture a wide range of potential behaviours and outcomes, enhancing its predictive capabilities (Wang et al, 2021).

Table 4.3 Breakage Rate mechanism parameter reference value, lower and upper boundaries (Wang et al,2021).

Parameters		*Reference value*	*Low value*	*High value*
Selection function	Breakage rate	CE: 4.5 s^{-1} KE :6 s^{-1}	CE: 0.9s^{-1} KE: 1.2s^{-1}	CE:22.5s^{-1} KE:30s^{-1}
	Minimum critical size	CE: 1000 µm KE: 1000um	CE: 800 µm KE: 800 µm	CE: 1200 µm KE: 1200 µm
	Maximum critical size	CE: 3500 µm KE: 3200 µm	CE: 2800 µm KE: 2560 µm	CE: 4200 µm KE: 3840 µm
	DYS	10kPa	5kPa	15kPa
	PFN	0.022	0.011	0.045
	Size exponent	CE: 2 KE: 10	CE: 1 KE: 5	CE: 4 KE:20
	Parameter a	CE: 23 KE: 0.9	CE: 23 KE: 0.8	CE: 23 KE: 1.0
	Parameter b	CE: 74.5 KE: 0.06	CE: 74.5 KE: 0.03	CE: 74.5 KE: 0.12
	Proportion	CE: 0.05 KE: 1	CE: 0.01 KE: 1	CE: 0.25 KE: 1
Breakage function	Scale parameter	CE:2.54e10 KE: 1.0	CE:2.54e10 KE: 0.2	CE:2.54e10 KE: 5.0
	Shape parameter	CE: 7.23 7.23 KE: 1.5	CE:7.23 KE:0.3	CE:7.23 KE:7.5

The first order effect, that is, the direct effect of factor i on response y is given by Equation 4.12.

$$S_i = \frac{\sigma_{i,y}^2}{\sigma_y^2} \quad 4.12$$

The total effect, that is, the total effect of factor i on response, y is given by Equation 4.13

$$S_{i,T} = \frac{\sigma_{i,T,y}^2}{\sigma_y^2} \quad 4.13$$

The first-order effect index quantifies the direct impact of each input factor on the output variance. If S_i= 0, then the ith factor has no direct influence on the output; However, a small value of S_i does not necessarily imply a non-influential factor. A higher value of S_i, the first-order effect index for the ith factor, indicates a greater influence of that factor on the output. Conversely, if S_i= 0 is zero, the ith factor has no direct impact on the output. However, a small S_i value does not necessarily imply that the factor is unimportant. It may still have a significant influence through its interactions with other factors. To account for both direct and indirect effects, the total effect index, $S_{i,T}$, is used. $S_{i,T}$ represents the total contribution of the ith factor to the output variance, including its first-order effect and all higher-order effects resulting from interactions with other factors. CQAs used to describe the output granules are particle size distribution (PSD), D10, D50, and D90 particle sizes.

By examining the total sensitivity index value for each rate mechanism parameter, one can identify the most influential parameters that significantly impact the model output. This information is crucial for determining which parameters should be included in the parameter estimation process to ensure accurate model predictions. By focusing on these critical parameters, one can optimise the model and improve its predictive capabilities. For an impactful parameter, the criteria state the average total sensitivity index value > 0.1. The average total sensitivity index value is an average of D10 total sensitivity index, D50 total sensitivity index, and D90 total sensitivity index. Typically, there is a minimum number of samples that produce accurate results. The optimal number of samples is case-specific and depends on system non-linearity, system complexity, factors, and responses selected as well as their number and method. To check for convergence, the analysis is executed for different numbers of samples. By plotting the key metric

of interest, the average total effect index, against the number of samples for each rate mechanism. In the first instance, the sample size will be increased from 50 to 2000 samples until the sensitivity index converges.

Spaghetti plots were used for the initial screening of the input factors to reduce calibration effort. Rankings of the input factors will be generated according to their relative contribution to the output variability to come up with a calibration strategy.

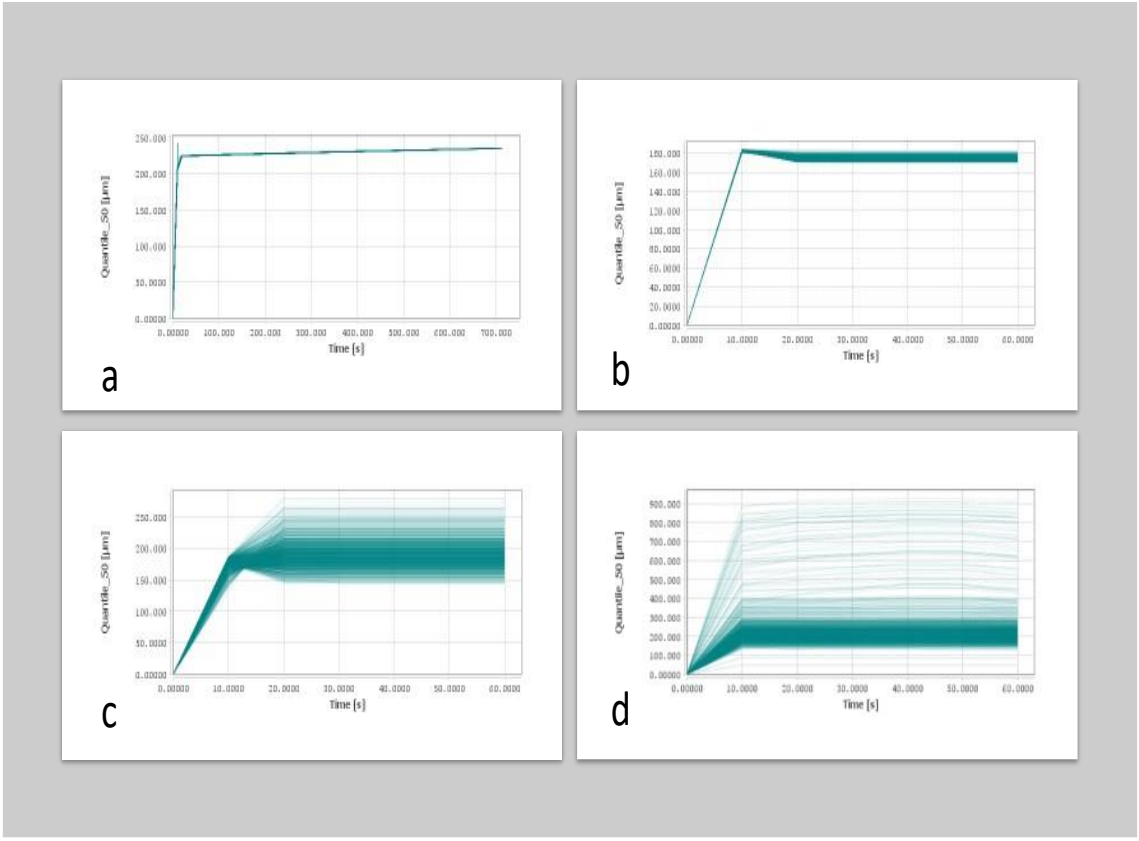


Figure 4.3 Quantile 50 Output granule particle size distribution for different rate mechanisms (a) consolidation; (b) nucleation; (c) layering; and (d) breakage.

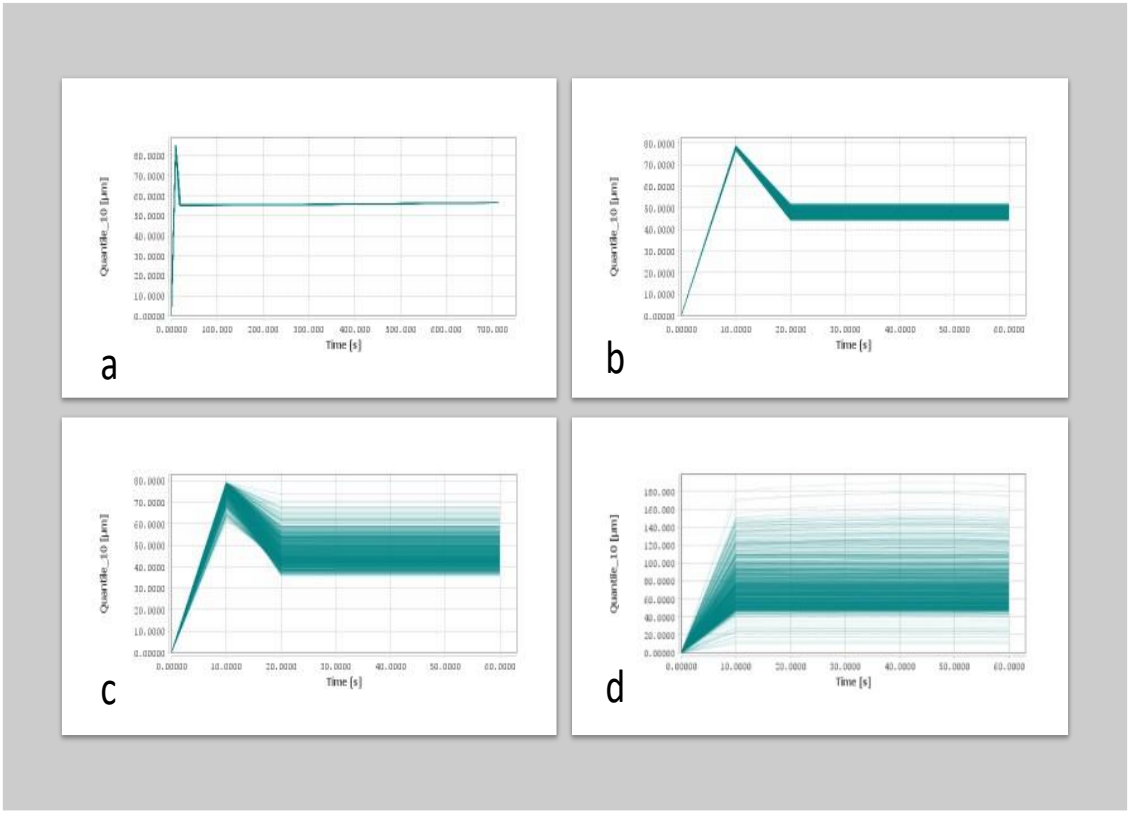


Figure 4.4 Quantile 10 Output granule particle size distribution for different rate mechanisms (a) consolidation; (b) nucleation; (c) layering; and (d) breakage.

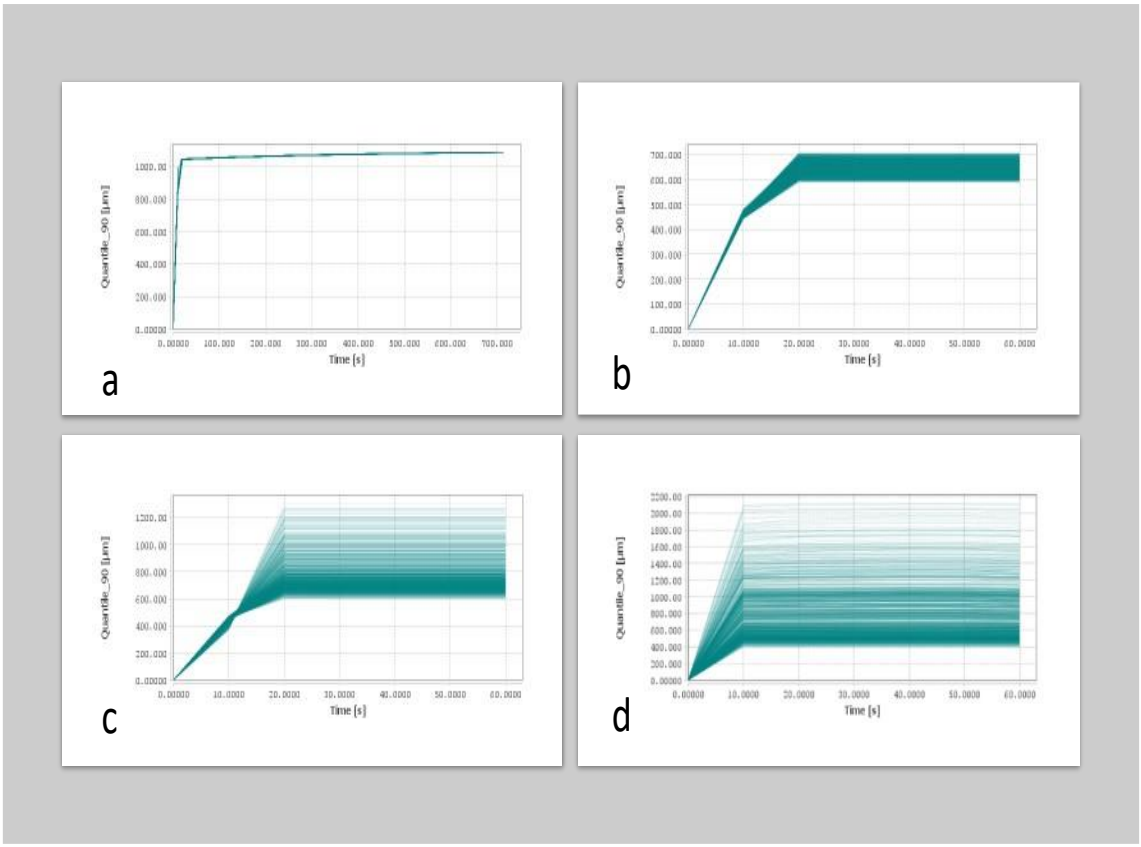


Figure 4.5 Quantile 90 Output granule particle size distribution for different rate mechanisms (a) Consolidation; (b) Nucleation; (c) Layering; (d) Breakage.

Figure 4.3, Figure 4.4 and Figure 4.5 shows spaghetti plots to display out values of a response (D10, D50, D50) as a function of time, for all samples of the rate mechanism input space. The graphs show that layering has the biggest impact on D50 following nucleation, then breakage. Consolidation has an insignificant effect on D50.

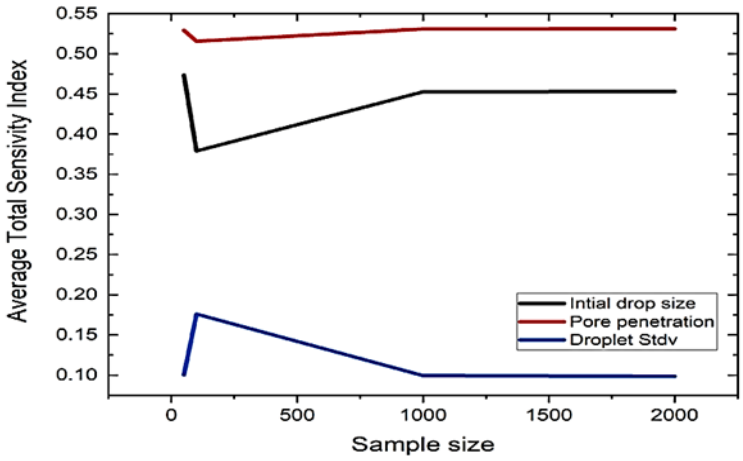


Figure 4.6 Nucleation parameters (Initial drop size, Pore penetration and Droplet standard deviation (Std Dev) average total sensitivity index at different sample sizes.

Figure 4.6 shows how the average total sensitivity index changes with sample size for nucleation kinetic parameters. Around 1000 samples, there is convergence. However, even at a low sample size, the rank is still consistent. An assumption can be made that the same trend with be observed for layering and breakage parameters. The total sensitivity index for the model's parameters was calculated using a sample size of 2000.

Figure 4.7 through Figure 4.10 present the results of a Global Sensitivity Analysis (GSA) for the 25 kinetic parameters to be estimated in the model. These tables report the total sensitivity index (TSI) and total average sensitivity index (TAAS) for each parameter at three different particle size fractions: D10 (representing the smallest 10% of particles), D50 (median particle size), and D90 (largest 10% of particles). Interestingly, the analysis does not reveal a clear trend in terms of which particle size fraction is most affected by a specific parameter. This suggests that the impact of each parameter might be more nuanced and may vary across the entire particle size distribution.

Furthermore, a significant observation emerges from the data: all consolidation parameters exhibit sensitivity values close to zero. This finding aligns with the

general understanding that consolidation processes primarily influence granule strength and density rather than directly affecting their size distribution. This insight allows us to confidently exclude consolidation parameters from the calibration process, streamlining the effort.

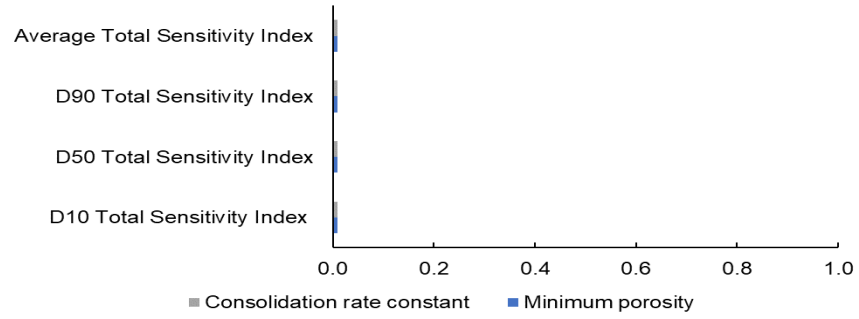


Figure 4.7 Consolidation parameters sensitivity index values.

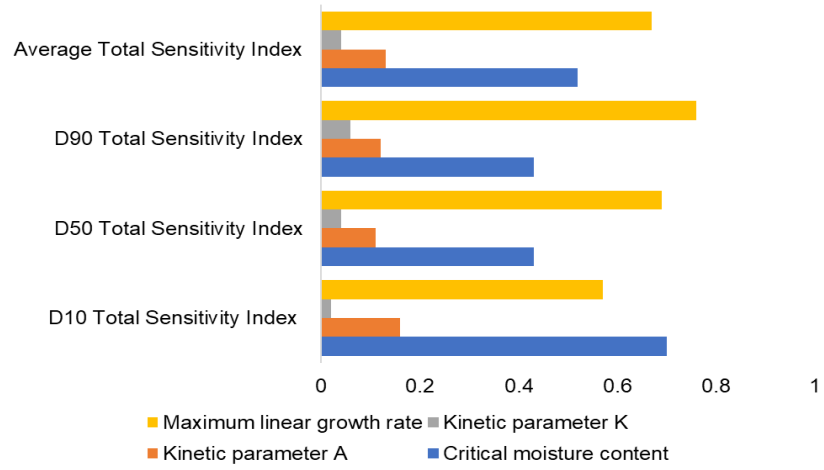


Figure 4.8 Layering parameters sensitivity index values.

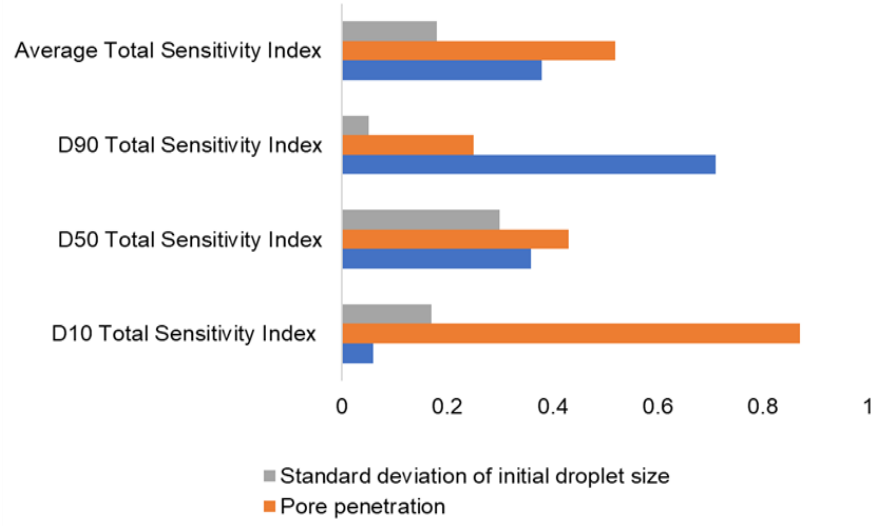


Figure 4.9 Nucleation parameters sensitivity index values.

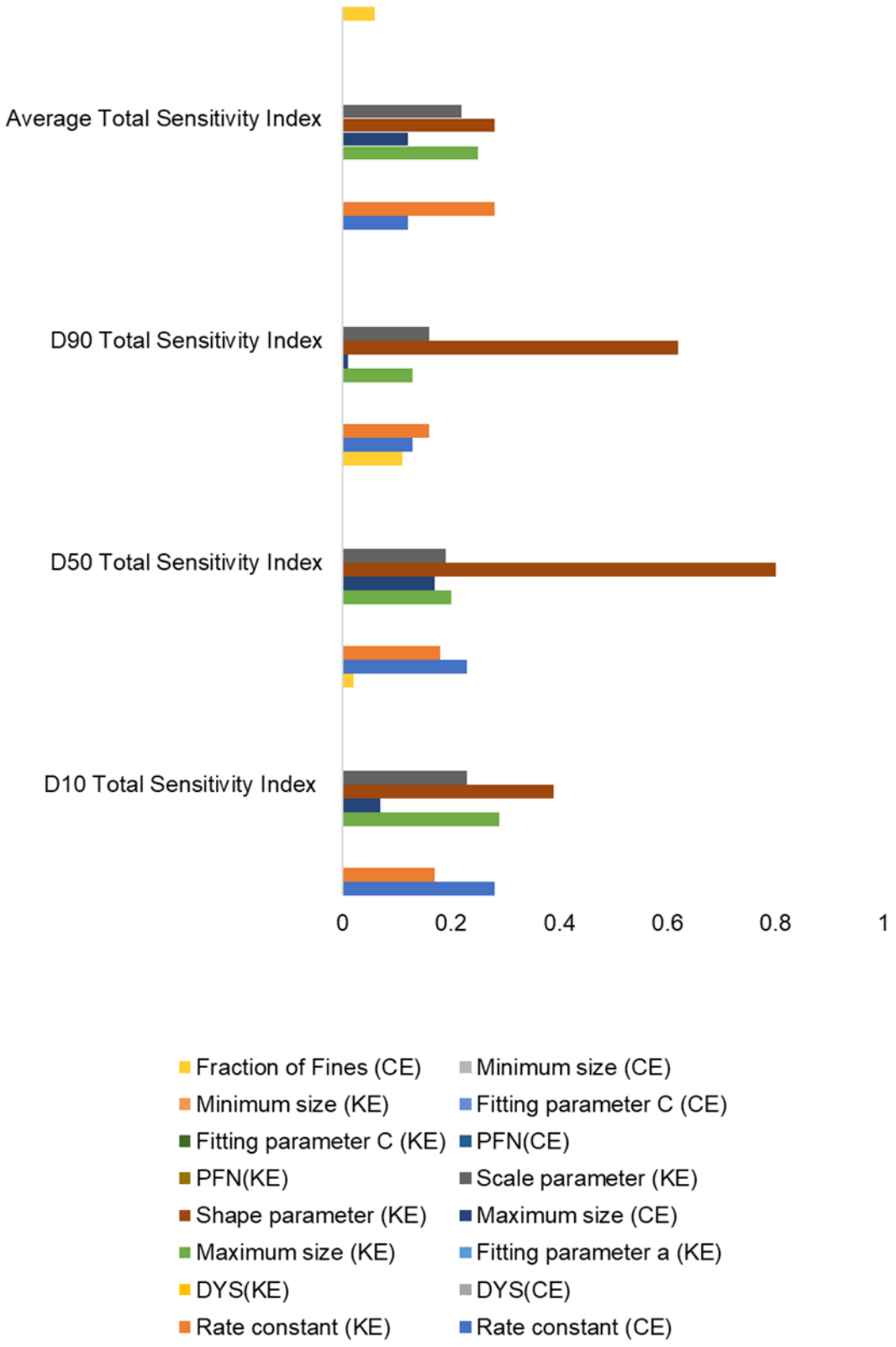


Figure 4.10 Breakage parameters sensitivity index values.

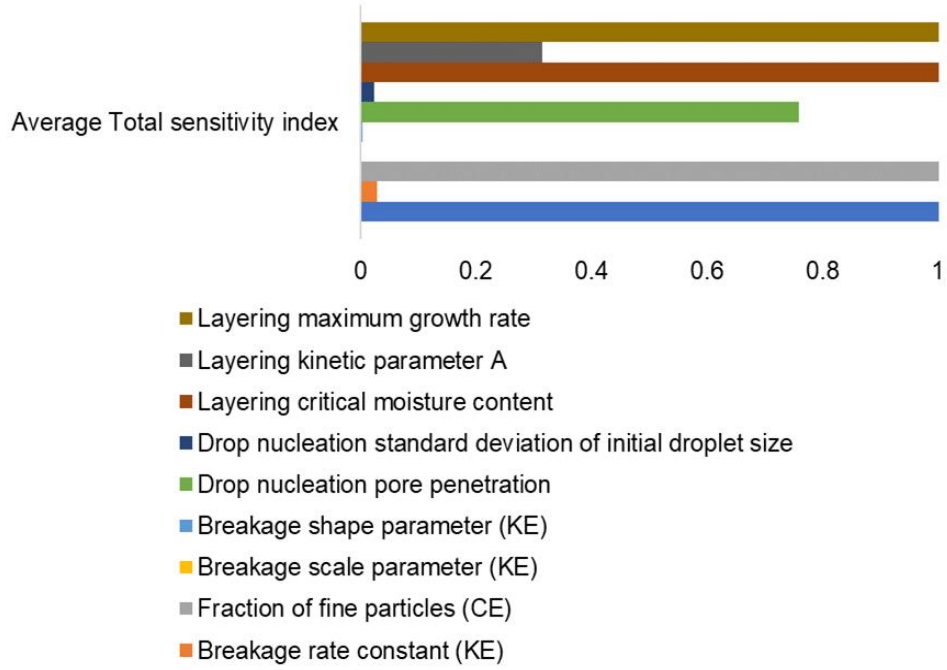


Figure 4.11 Summary of the influence of TSWG kernel on the output PSD.

Figure 4.11 summarises the key takeaways from the GSA. It consolidates the information from previous figures and identifies the ten most important TSWG parameters based on their sensitivity index values for various outputs. These ten parameters are the primary candidates for calibration during characterisation tests or parameter estimation procedures. The ranking of parameters provided in

*Figure 4.11*offers valuable guidance in selecting appropriate calibration methods. Notably, the layering parameter's minimum moisture content parameter stands out with the highest average total sensitivity index (2.8978). This suggests that it will likely have the most significant influence during calibration, making it a prime candidate for focused calibration efforts. DYS, Upper breakage critical size in KE, PFN and Minimum breakage critical size in CE are dependent on material properties and equipment geometry properties and therefore will not be used in analysis.

Using machine learning (ML) to calculate sensitivity indices for models with many parameters can indeed speed up the process significantly. Sensitivity analysis aims to identify how changes in the inputs of a model affect its outputs. Traditional methods like Monte Carlo simulations can become computationally expensive, especially with many parameters like breakage parameters. ML models can be

trained to approximate the interconnected relationships between inputs and outputs of the original model. These surrogate models can then be used for sensitivity analysis, which is usually much faster than directly evaluating the original model. Techniques like PCA or autoencoders can reduce the dimensionality of the parameter space while preserving important features. However, it's crucial to acknowledge the inherent trade-off: the computational gains achieved through surrogates and dimensionality reduction come at the cost of potential inaccuracies in the sensitivity analysis results.

Sampling Techniques: Machine learning algorithms can employ advanced sampling techniques such as Quasi-Monte Carlo Sampling to efficiently explore the parameter space, reducing the number of model evaluations needed. Machine learning algorithms can be parallelised to take advantage of distributed computing resources, further speeding up the calculation process. By leveraging these techniques, sensitivity analysis for models with many parameters can be carried out much quicker, enabling faster decision-making and model refinement.

4.5 Developing an accelerated calibration strategy

While significant advancements have been made in characterising material properties, challenges persist in accurately estimating certain parameters within the defined population balance model, particularly those with empirical foundations or those that are difficult to measure precisely. Nevertheless, precise parameter estimation is crucial for ensuring the model's predictive accuracy. The Parameter Estimation Module in gPROMS Formulated Products is utilised to estimate these parameters by comparing model predictions to experimental data.

Model calibration and validation are essential steps in establishing the predictive capabilities of the model. Calibration involves optimising parameter values to minimise discrepancies between model predictions and experimental data. Validation, on the other hand, ensures that the calibrated model can accurately predict outcomes for conditions beyond the initial experimental design space.

The TSWG model is defined by a system of partial differential and algebraic equations (PDAEs). The DAEBDF numerical solver, specifically designed to handle large, sparse systems of equations, is employed to solve these PDAEs. The Maximum Likelihood (MXLKHD) solver, based on sequential quadratic programming (SQP), was utilised to find the global optimum. The Maximum

Likelihood algorithm is given as below, consists of three contributory terms, i.e. constant term, variance term and residual term.

$$\Phi = \frac{N}{2} ln\,(2\pi + \frac{1}{2} \underset{\theta}{min} \left\{ \sum_{i=1}^{NE} \sum_{j=1}^{NV_i} \sum_{k=1}^{NM_{ij}} \left[ln\left(\sigma_{ijk}^2\right) + \frac{\left(\tilde{z}_{ijk} - z_{ijk}\right)^2}{\sigma_{ijk}^2} \right] \right\} \quad 4.14$$

The optimal parameter values are determined by minimising the contribution of residual terms, which represent the difference between predicted and measured values for all relevant variables. Once the model has been calibrated, it is used to predict PSD values in twin-screw granulation.

4.5.1 Model Calibration Table Matrix

Calibrating TSWG models often presents a significant challenge due to the high number of rate mechanism parameters involved, typically exceeding 25. This complexity makes it difficult to isolate and understand the individual impact of each parameter on model behaviour under specific operating conditions. Attempts at global calibration, employing all parameters and/ or all experimental runs simultaneously, resulted in immediate errors. Consequently, current practices rely on setting a threshold for a sensitivity index to selectively choose which process parameters to use (usually Liquid to solid ratio due to its high correlation with PSD) for calibration. However, this approach inherently compromises accuracy. As powder flow rate and screw speed are critical to modelling twin-screw energy consumption, their inclusion is necessary. This novel approach calibrates twin-screw granulator parameter data points within the design space, akin to the ‘Steam Tables methodology’ which provides specific volume, enthalpy, internal energy, and entropy for water across a range of temperatures and pressures. Calibration values for intermediate process conditions can be extrapolated from these established points.

Table 4.4 includes the parameters used in parameter estimation. In all, 10 parameters for estimate were chosen through global system analysis. Four trials (runs 2, 3, 4, and 5) are chosen for model calibration, while the remaining runs (runs 1 and 7) will be utilised as external data for validation to assess the predictive capacity of the calibrated model. Given that our dataset comprises

independent observations, I opted for random splitting of the datasets. This strategy is particularly suitable because the datasets are independent, as it prevents any potential bias introduced by chronological ordering. Unlike time-series splitting, which is crucial for data with inherent time dependencies, random splitting allows for a more unbiased evaluation of the model's generalisation capabilities on unseen data. While k-fold cross-validation offers a robust evaluation, especially for smaller datasets, the dataset's size and independent nature made simple random splitting an efficient and effective method for assessing model performance.

By using Run 2 as a baseline and adjusting process parameters in Run 4 (decreasing powder flow and screw speed, increasing L/S ratio), I observed a slight increase in breakage parameters (fraction of fine and scale) and a decrease in the layering parameter. These changes aligned with the observed shift in product size distribution towards a narrower, unimodal peak. Conversely, increasing L/S ratio, powder feed rate, and screw speed in Run 3 led to a reduction in the breakage fraction of fine and an increase in the layering minimum moisture content. This corresponded to a wider size distribution with an increase in d50 and a decrease in d10. When Run 3 was used as the baseline, decreasing powder flow and screw speed in Run 5 resulted in reduced breakage and nucleation parameters, but increased layering parameters. These changes were associated with the development of a bimodal size distribution. Understanding how changes in operating conditions influence calibration parameters can significantly reduce calibration time. By establishing a correlation between process parameters and physical calibration parameters, I eliminated the need for lengthy and unpredictable optimisation solvers. This approach, which aims to minimise the amount of calibration required, will be explored further in the next section.

Table 4.4 presents the results of calibrating the model using the optimal values obtained for the estimated parameters. As illustrated in Figure 4.12, this approach yielded a fitted granule size distribution that demonstrates a high degree of agreement with the experimental data.

Table 4.4 Optimised parameters for run 2,3,4 and 5.

Mechanism	*Parameters*	*Initial value*	*Optimised value*			
			Run 2	run 3	Run 4	run 5
Nucleation	Droplet mean size Standard deviation	0.45	0.3562	0.1640	0.001	0.01
	Droplet pore penetration	0.4	0.5397	0.3131	0.2750	0.2743
Breakage	Breakage rate (CE)	4.5	1.5	1.5	1.5	1.5
	Breakage rate (KE)	6	2.5	2.5	2.5	2.5
	Scale parameter (KE)	1	0.8821	1.5464	2	2.0049
	Shape parameter (KE)	1.5	1.3956	1.9418	0	3.0074
	Fraction of fines	0.05	0.0518	-0.0083	0.1	-0.0002
Layering	Growth rate	25	15.7916	13.4049	0	176.2680
	Minimum moisture content	0.2	0.0623	0.2420	-0.001	-0.001
	Empirical parameter A	10	6.9523	12.5622	20	20.049
Linear variance model	Alpha	0.05	0.0024	0.0024	0.0251	0.0252
	Beta	0.01	0.0006	0.0006	0.0051	0.0051

The fitted model accurately captures the overall shape of the distribution, although a slight underestimation of the peak proportion is evident. RMSE is used for measuring the difference between an actual value and the value predicted by a model. The value is used to show how accurate a model is in predicting the target value. Although, the values of RMSE may range between zero and infinity, however, an RMSE value between 0.2 and 0.5 shows how accurate the model is in predicting the actual value.

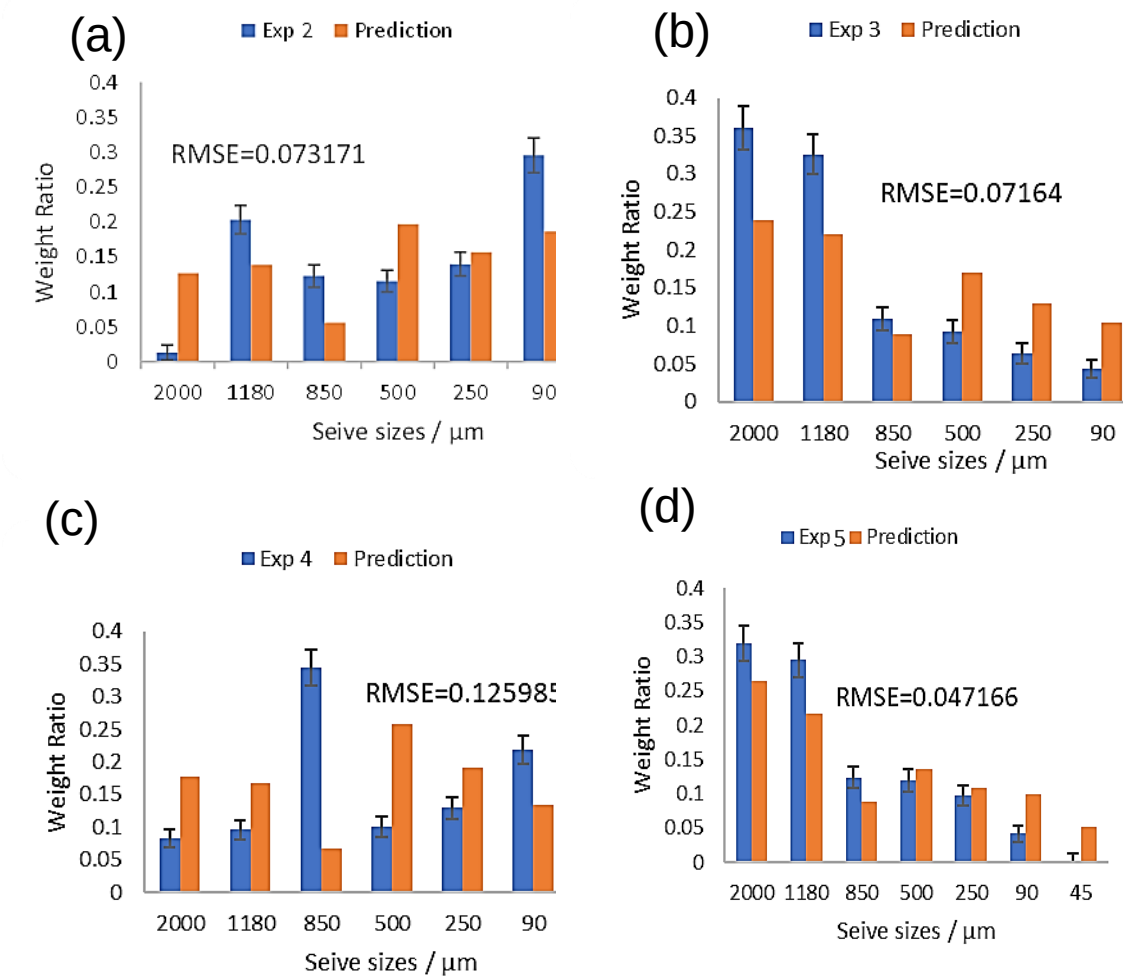


Figure 4.12 Calibration Predictions: (a) Run 2, (b) Run 3, (c) Run 4 and (d) Run 5

When an RMSE value is closer to 0, it is assumed that the model has a better performance. However, when the RMSE value is greater than 1, it means that the model couldn't find a solution, or it is not accurate, or it is not a better fit with the target value. It's important to note that the Root Mean Square Error (RMSE) values vary between runs. Run 4 exhibits the highest RMSE, which can be attributed to the abrupt change in the weight ratio between the 500 µm and 850 µm sieve sizes. Conversely, run 5 has the lowest RMSE, indicating a closer fit between the model prediction and the experimental data for this run. Interestingly, both Runs 2 and 4 possess bimodal particle size distributions. However, run 2 achieves a better fit compared to Run 4. This can be explained by the more gradual variations in the weight ratio observed in Run 2 as the sieve size increases, which aligns better with the underlying assumptions of the model.

4.5.2 Linking Rate Mechanism Parameters with CPPs

The new calibration table format allowed developing models that explicitly link formulation and process properties, reducing the number of calibration parameters further. Trying to consider of all process parameter for calibration is cause massive in error in the predictions. This error was later reduced by reducing the number of calibration parameter using global system analysis and performing calibration by range. Considering all process consider is essential to optimisation of energy usage, however if particle size distribution optimisation is the main objective, calibration using L/S only is sufficient. Results of calibration using L/S experiments are shown in Figure 4.13. These results are taken from Wang et al paper (Wang et al, 2021). Another way of reducing the calibration efforts is through mechanistic understanding.

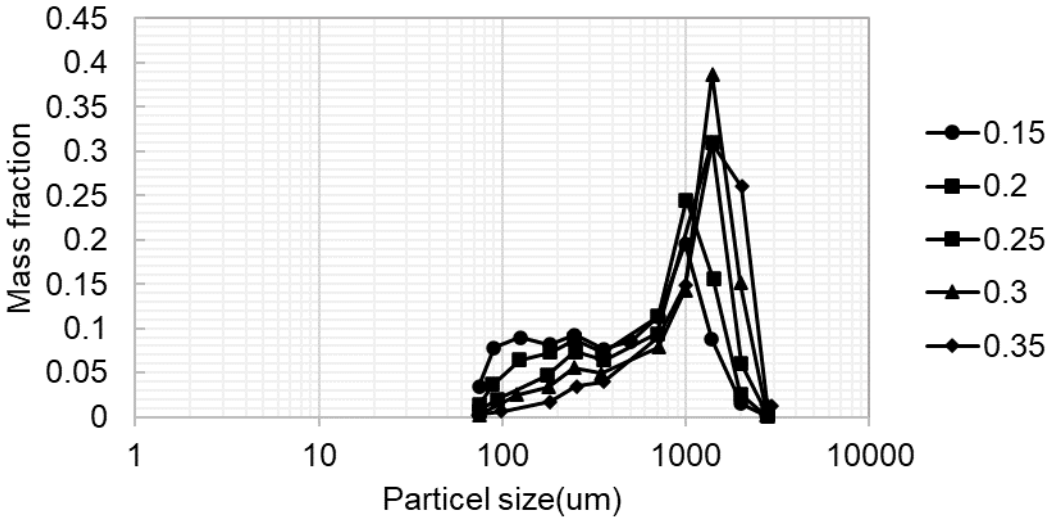


Figure 4.13 Twin screw wet granulator sieve analysis results. Taken from Wang et al,2021.

This section explored the formulation of models that can directly link formulation properties, such as the L/S ratio, with key process parameters like droplet size and standard deviation. This method aims to reduce the number of calibration parameters required for accurate model predictions. The investigation revealed an interesting relationship: granule size appears to increase linearly with droplet size, suggesting a potential dependence of nucleus size on initial droplet size. Consequently, it can be hypothesised that increasing the L/S ratio would lead to larger droplets due to a higher volume of liquid available for atomisation. This, in turn, could result in larger granules and potentially higher standard deviations in granule size distribution.

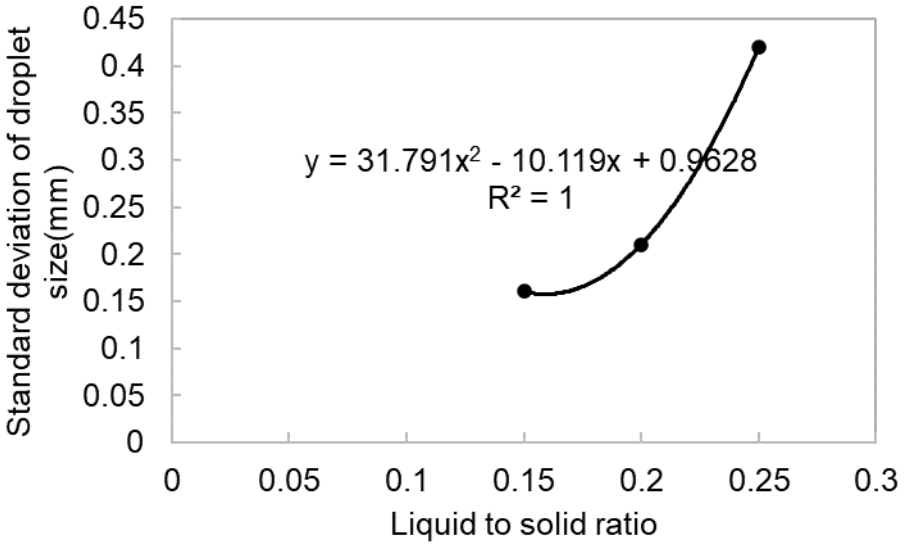


Figure 4.14 Droplet size and Liquid to solid ratio correlation.

To validate these hypotheses, the standard deviation of droplet size was calibrated for each L/S ratio. For L/S ratios ranging from 0.15 to 0.25, the calibration results, presented in Figure 4.14, reveal an exponential correlation between the L/S ratio and the standard deviation of droplet size. This finding provides valuable insights into the mechanistic understanding between L/S ratio and nucleation processes. While further research is needed to fully elucidate the behaviour at higher L/S ratios, this initial exploration demonstrates the potential of linking formulation properties with process parameters to improve model accuracy and reduce calibration complexity.

4.5.3 Model Validation

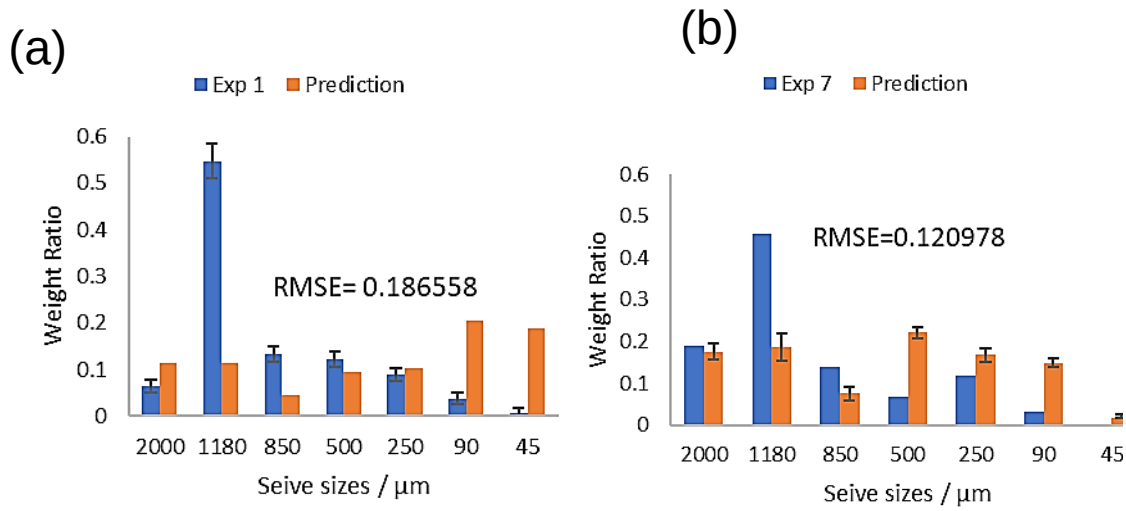


Figure 4.15 Run 2 Model validation using: (a) Run 1 and (b) Run 6

Figure 4.15 illustrate the comparison between the predicted PSD obtained from the Run 2 calibrated TSWG model and the actual measured values from external experiments i.e. exp 1,6 and 7. Overall, the model does not align well with the experimental data, unsuccessfully capturing the general shape of the PSD. However, a consistent underestimation of the peak proportion is observed across all figures at particle size of 1180 um. This indicates that the model may be slightly conservative in predicting the amount of material at the peak size range. Exp 1 with the high peak size range has this the high RMSE. This is because run 1 and 2 does not a similar structure.

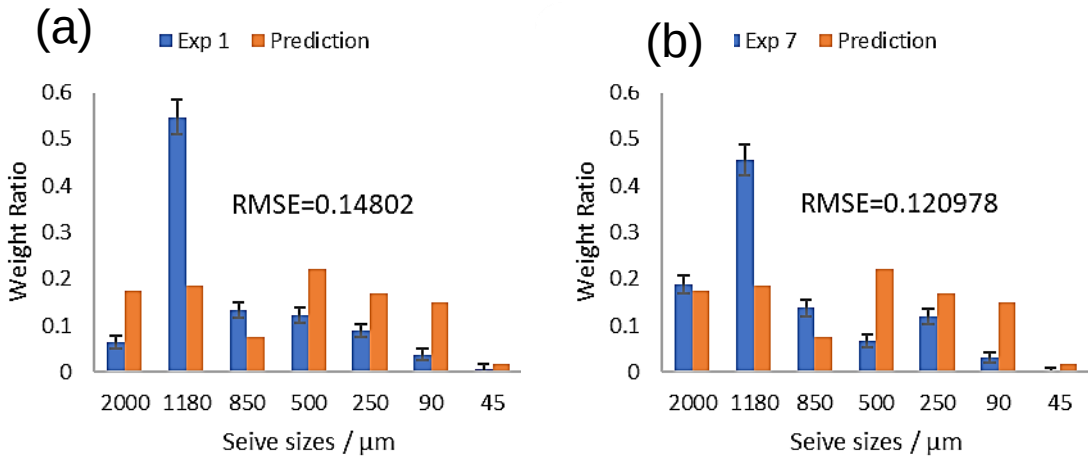


Figure 4.16 Run 3 Model validation using: (a) Run 1, (b) Run 6 and (c) Run 7.

Figure 4.16 illustrate the comparison between the predicted granule size distribution (PSD) obtained from the Run 3 calibrated TSWG model and the actual measured values from external experiments (exp 1,6 and 7). Overall, the model aligns well with the experimental data, successfully capturing the general shape of the PSD. However, a consistent underestimation of the peak proportion is observed across all figures. This indicates that the model may be slightly conservative in predicting the amount of material at the peak size range. Exp 1 with the high peak size range has a better fit compared in run 3 model because they have the same unimodal structure. Consider having different calibration model for bimodal and unimodal particle distribution shape.

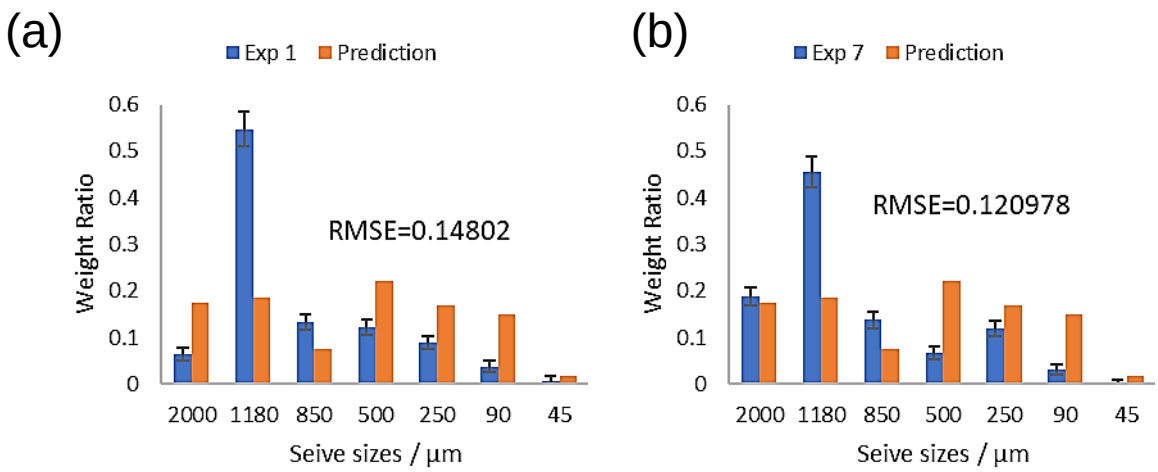


Figure 4.17 Run 4 Model validation using: (a) Run 1, (b) Run 6 and (c) Run 7.

Figure 4.17 illustrate the comparison between the predicted granule size distribution (PSD) obtained from the Run 4 calibrated TSWG model and the actual measured values from external experiments. Overall, the model aligns well with the experimental data, successfully capturing the general shape of the PSD.

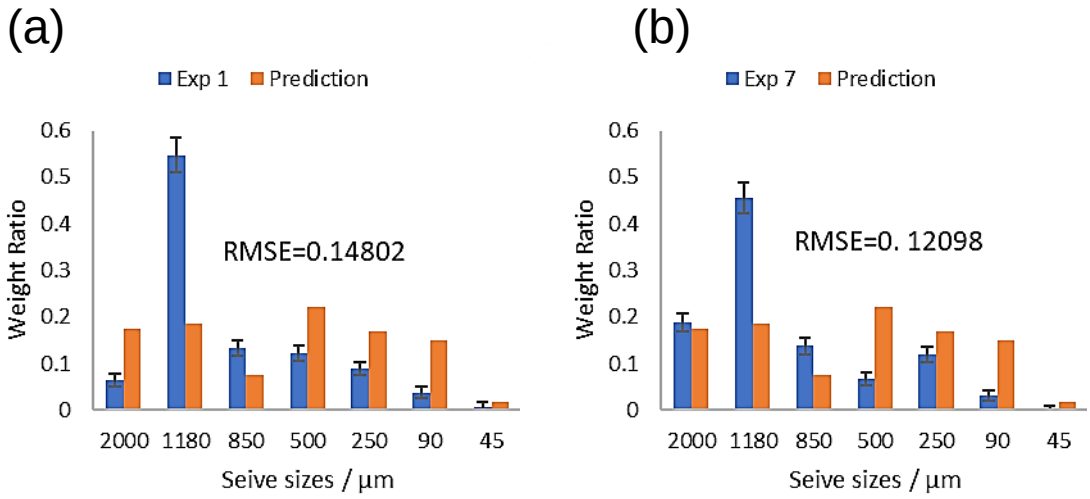


Figure 4.18 Run 5 Model validation using: (a) Run 1, (b) Run 6 and (c) Run 7.

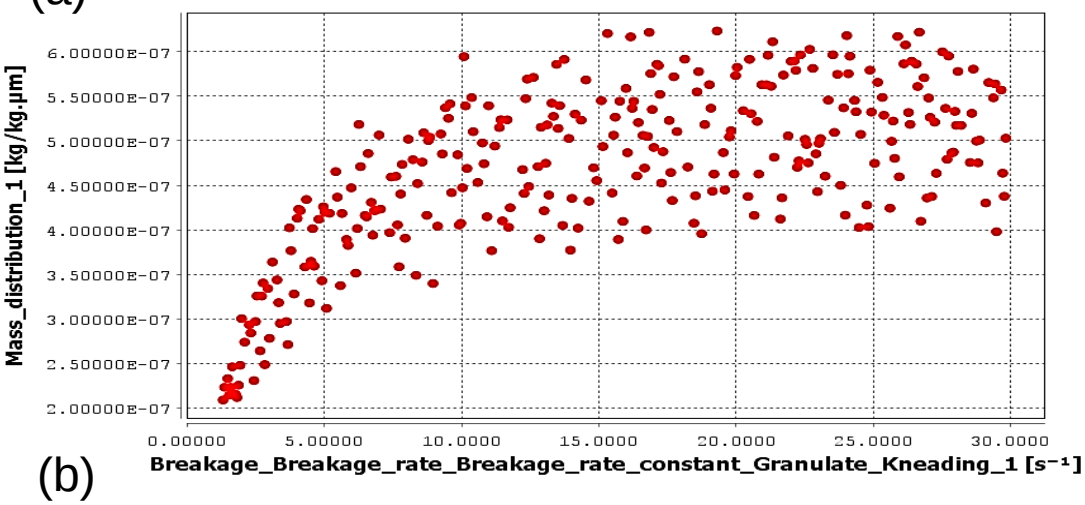
Figure 4.18 illustrate the comparison between the predicted PSD obtained from the RUN 5 calibrated TSWG model and the actual measured values from external experiments. Overall, the model aligns well with the experimental data, successfully capturing the general shape of the PSD. However, a consistent underestimation of the peak proportion is observed across all figures. This indicates that the model may be slightly conservative in predicting the amount of material at the peak size range. However, run 6 calibration stands out as demonstrating the strongest agreement with the validation data compared to the

other calibrations. This suggests that the parameter set used in run 5 calibration might be particularly well-suited for this specific experimental dataset. Interestingly, the Root Mean Square Error (RMSE) values for the predicted PSDs appear to be very similar, regardless of the specific calibration parameters used. This suggests that the model's performance might not be highly sensitive to variations in these parameters within the tested range. Run 7 RMSE values remain contained across the 4 different TSWG model. The suggest that this is the best it can get given the 10 chosen calibrations parameters and operation range.

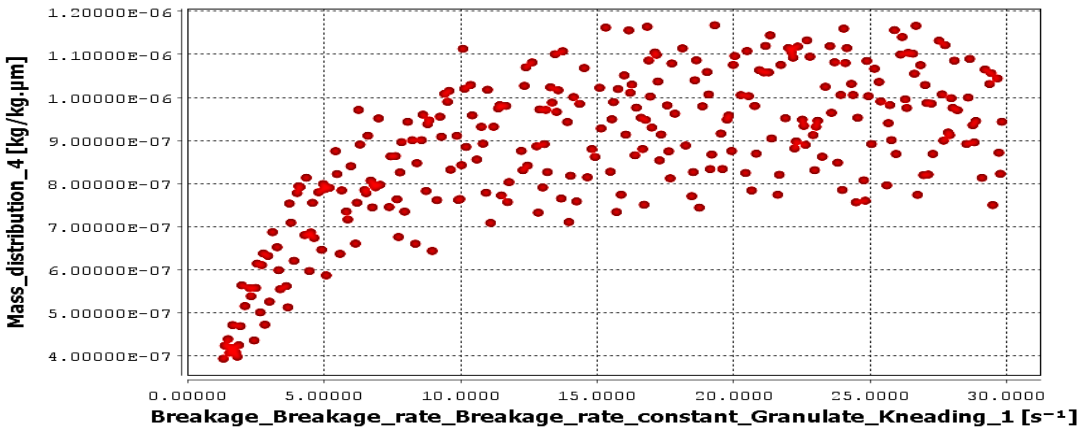
4.5.4 Calibration by range

Normal calibration has a significantly large error due to the rapid change in granule weight ratio between successive sieve sizes.

(a)



(b)



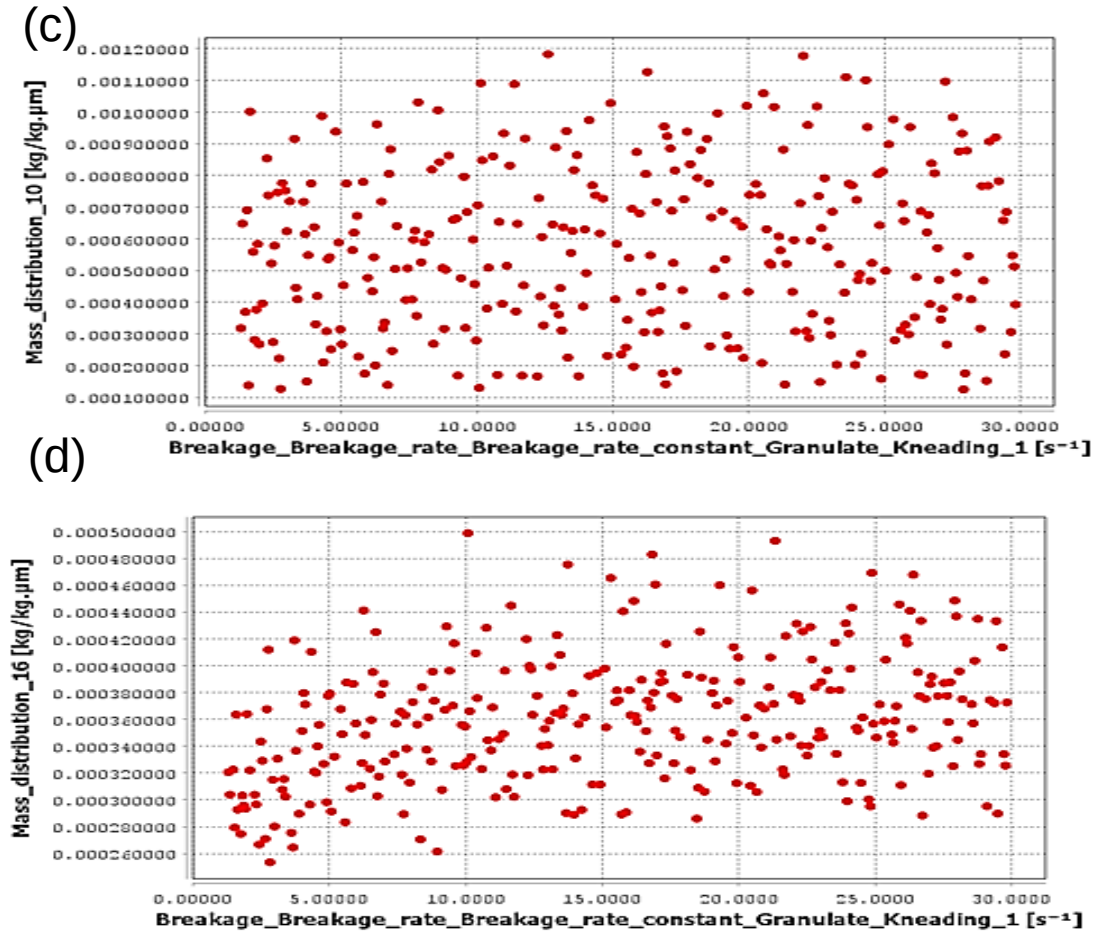


Figure 4.19 Effects of breakage rate constant particle size mass distribution: (a) Particle size 0.01µm, (b) Particle size 0.09 µm, (c) Particle size 6.85 µm and Particle size 1000 µm).

Figure 4.19 graphs below show the effects of different breakage rate constants on the mass distribution of particle sizes. For small particle size distributions, there is a clear trend that increasing breakage rate increases the mass fraction however, for large particle size distributions, it is random. This explains why sudden changes in mass density between particle ranges make calibration difficult.

Figure 4.20 Typical forms of the granule particle size distribution.

By splitting the distribution into parts, it is possible to increase the accuracy. The distribution observed in the experimental data has a similar shape shown below

in Figure 4.20. The particle size distribution can be split into three parts to avoid rapid change in weight ratios.

Table 4.5 Optimised parameters for run 2 range 1,2 and 3.

Mechanism	*Parameters*	*Run 2 Optimised Value*		
		Range 1	Range 2	Range3
Nucleation	Droplet mean size Standard deviation	0.3562	0.1639	0.0010
	Droplet pore penetration	0.5397	0.3131	0.2750
Breakage	Breakage rate (CE)	1.5000	1.5000	1.5000
	Breakage rate (KE)	2.5000	2.5000	2.5000
	Scale parameter (KE)	0.8821	1.5464	2.0000
	Shape parameter (KE)	1.3956	1.9418	0.0000
	Fraction of fines	0.0518	-0.0083	0.1000
Layering	Growth rate	15.7916	13.4049	0.0000
	Minimum moisture content	0.0623	0.2420	-0.0010
	Empirical parameter A	6.95227	12.5622	20.0000
Linear variance model	Alpha	0.0024	0.0024	0.0251
	Beta	0.0006	0.0006	0.0051

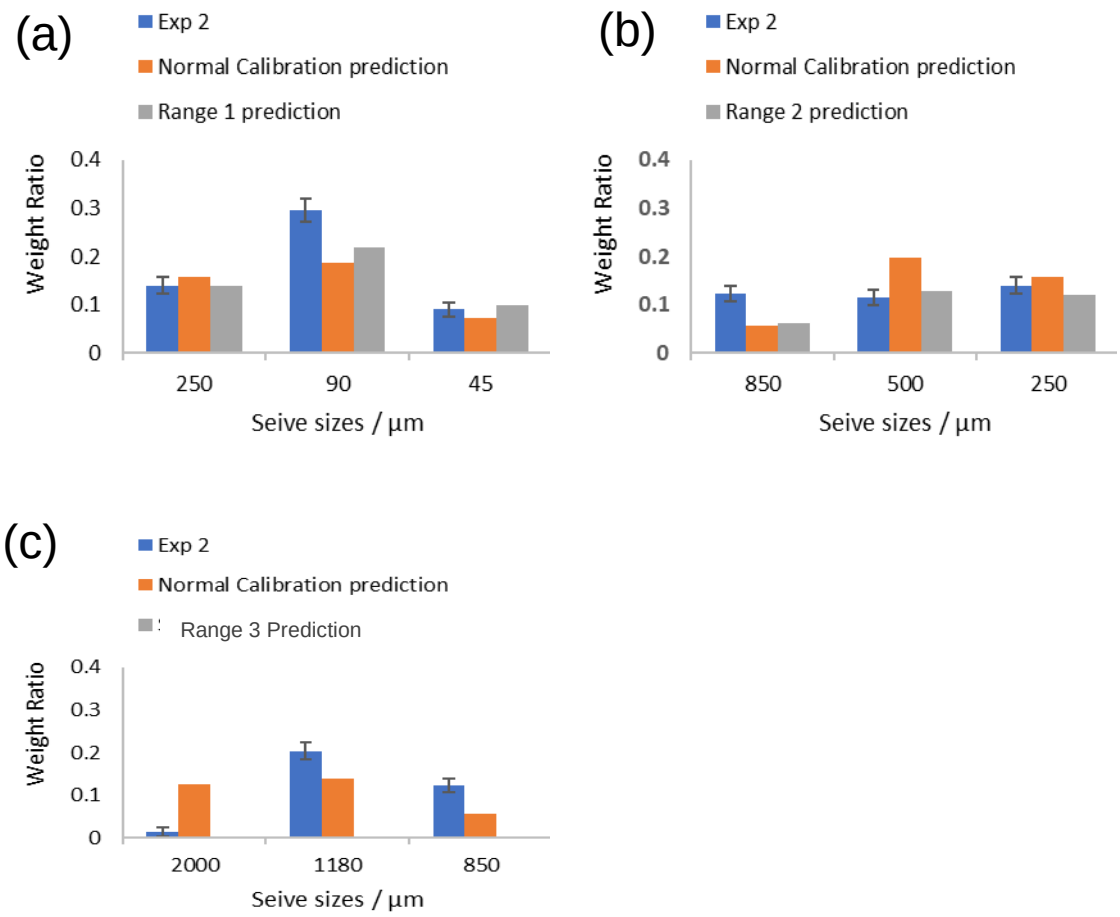


Figure 4.21 Run 2 Prediction by range: (a) Range 1 Prediction, (b) Range 2 Prediction and (c) Range 3 Prediction.

Table 4.6 Optimised parameters for run 3 range 1,2 and 3.

Mechanism	*Parameters*	*Run 3 Optimised Value*		
		Range 1	Range 2	Range 3
Nucleation	Droplet mean size Standard deviation	0.3562	0.1639	0.0010
	Droplet pore penetration	0.5397	0.3131	0.2750
Breakage	Breakage rate (CE)	1.5000	1.5000	1.5000
	Breakage rate (KE)	2.5000	2.5000	2.5000
	Scale parameter (KE)	0.8821	1.5464	2.0000
	Shape parameter (KE)	1.39564	1.9418	0.0000
	Fraction of fines	0.0518	-0.0083	0.1000
Layering	Growth rate	15.7916	13.4049	0.0000
	Minimum moisture content	0.0623	0.2420	-0.0010
	Empirical parameter A	6.9523	12.5622	20.000
Linear variance model	Alpha	0.0024	0.00238705	0.0251
	Beta	0.00063	0.00062	0.0051

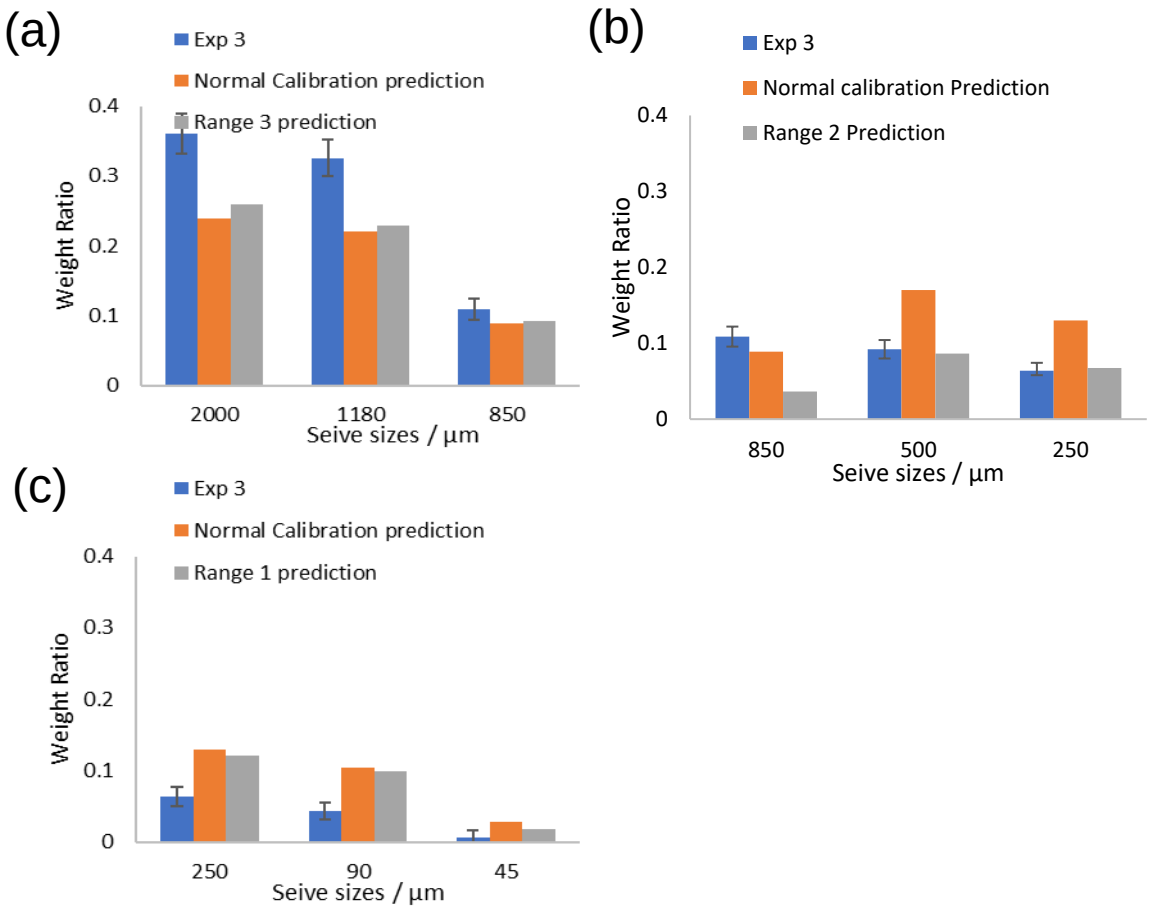


Figure 4.22 Run 3 Prediction by Range: (a) Range 1 Prediction, (b) Range 2 Prediction and (c) Range 3 Prediction.

Table 4.7 Optimised parameters for run 4 range 1,2 and 3.

Mechanism	*Parameters*	*Run 4 Optimised value*		
		Range 1	Range 2	Range 3
Nucleation	Droplet mean size Standard deviation	0.3562	0.1639	0.0010
	Droplet pore penetration	0.5396	0.3131	0.2750
Breakage	Breakage rate (CE)	1.5000	1.5000	1.5000
	Breakage rate (KE)	2.5000	2.5000	2.5000
	Scale parameter (KE)	0.8821	1.5464	2.0000
	Shape parameter (KE)	1.3956	1.9418	0.0000
	Fraction of fines	0.0518	-0.0083	0.1000
Layering	Growth rate	15.7916	13.4049	0.0000
	Minimum moisture content	0.0623	0.2420	-0.0010
	Empirical parameter A	6.9523	12.5622	20.000
Linear variance model	Alpha	0.0024	0.0024	0.02505
	Beta	0.0006	0.0006	0.00505

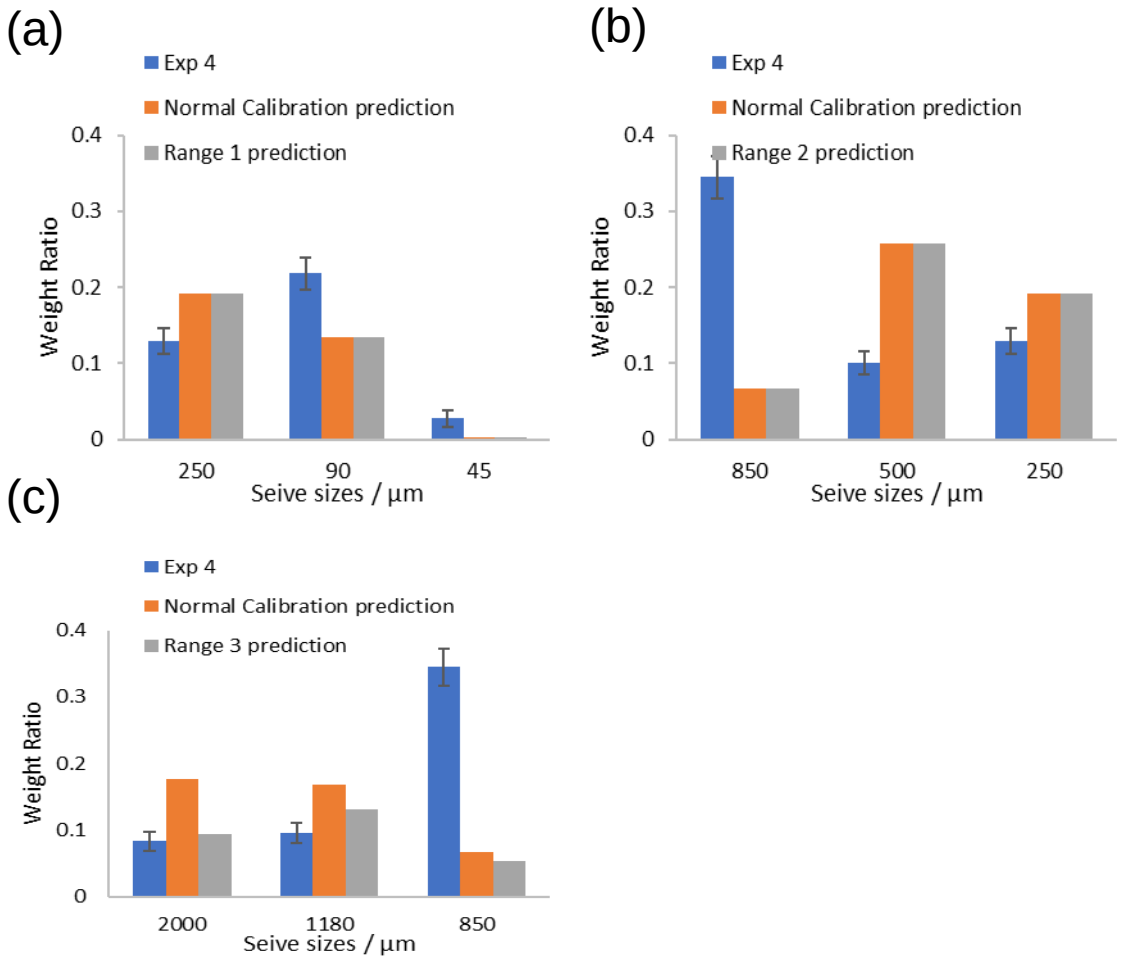


Figure 4.23 Run 4 Prediction by Range: (a) Range 1 Prediction, (b) Range 2 Prediction and (c) Range 3 Prediction.

Table 4.8 Optimised parameters for run 5 range 1,2 and 3.

Mechanism	*Parameters*	*Run 4 Optimised value*		
		Range 1	Range 2	Range 3
Nucleation	Droplet mean size Standard deviation (Std Dev)	0.3562	0.1639	0.0010
	Droplet pore penetration	0.5396	0.3131	0.2750
Breakage	Breakage rate (CE)	1.5000	1.5000	1.5000
	Breakage rate (KE)	2.5000	2.5000	2.5000
	Scale parameter (KE)	0.8821	1.5464	2.0000
	Shape parameter (KE)	1.3956	1.9418	0.0000
	Fraction of fines	0.0518	-0.0083	0.1000
Layering	Growth rate	15.7916	13.4049	0.0000
	Minimum moisture content	0.0623	0.2420	-0.0010
	Empirical parameter A	6.9523	12.5622	20.0000
Linear variance model	Alpha	0.0024	0.0024	0.0250
	Beta	0.0006	0.0006	0.0050

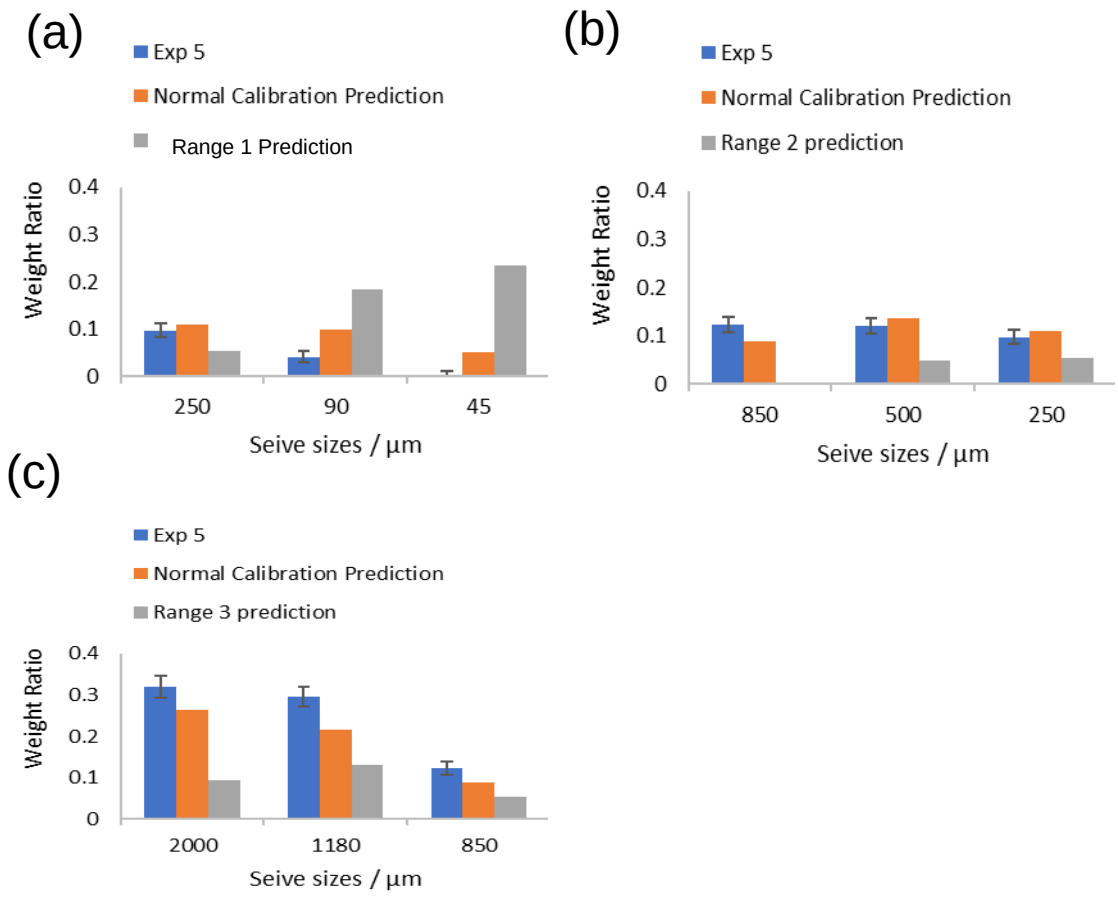


Figure 4.24 Run 5 Prediction by range: (a) Range 1 Prediction, (b) Range 2 Prediction and (c) Range 3 Prediction.

Figure 4.21 through Figure 4.24 showcase a promising outcome – a noticeable improvement in prediction accuracy for particle size distributions (PSDs) achieved through calibration. This approach demonstrates the model's ability to effectively handle PSDs that deviate from the typical symmetrical, bell-shaped curve (normal distribution) commonly encountered in many processes. These non-symmetrical distributions, often characterized by bimodality (two peaks) or other deviations, can be challenging for traditional models to predict accurately. However, by incorporating calibration techniques, the TSWG model demonstrates its potential to overcome this limitation and provide reliable predictions for a wider range of real-world PSD scenarios.

4.6 Back end-Case study-DiPP's ConsiGma 25-line Back end

4.6.1 Mechanistic model

The Reynolds et al. model assigns each tablet phase an independent tensile strength, determined solely by compaction pressure and phase properties, and unaffected by other compacted phases. The Reynolds et al. model uses the Ryshkewitch-Duckworth equation to model the compatibility of a directly compressed powder (Equation 4.15). The tablet tensile strength model connects the tablet's attributes to its tensile strength (Wu et al., 2006).

$$\sigma_t = \sigma_{\mathrm{t0}} e^{-k_b \varepsilon_t} \qquad 4.15$$

Where:

- σ_{t0} represent the tensile strength at zero porosity.
- $k_b$is the bonding capacity of the material.
- ε_t is tablet porosity.

The tensile strength at zero porosity and bonding capacity are primary parameters that depend on the properties of the individual phases within the tablet. Tensile strength at zero porosity is a theoretical value representing the tensile strength a material would possess if it were completely dense, with no pores. This value is often determined by extrapolating experimental data that relates tensile strength to porosity. bonding capacity describes a material's ability to form bonds, while tensile strength at zero porosity represents the material's ideal strength without any weakening effects from pores. The overall tablet tensile strength is calculated as the geometric mean of the tensile strengths of these

individual phases. This model effectively predicts tablet tensile strength at moderate compression forces. However, it falls short in capturing the decrease in tensile strength observed at higher compression forces.

A modified version of the Gurnham Equation 4.16 is used to represent the compressibility of a powder in a tablet press:

$$\varepsilon_t = -\frac{1}{k_{cc}} \ln\left(\frac{p}{p_0}\right) \qquad 4.16$$

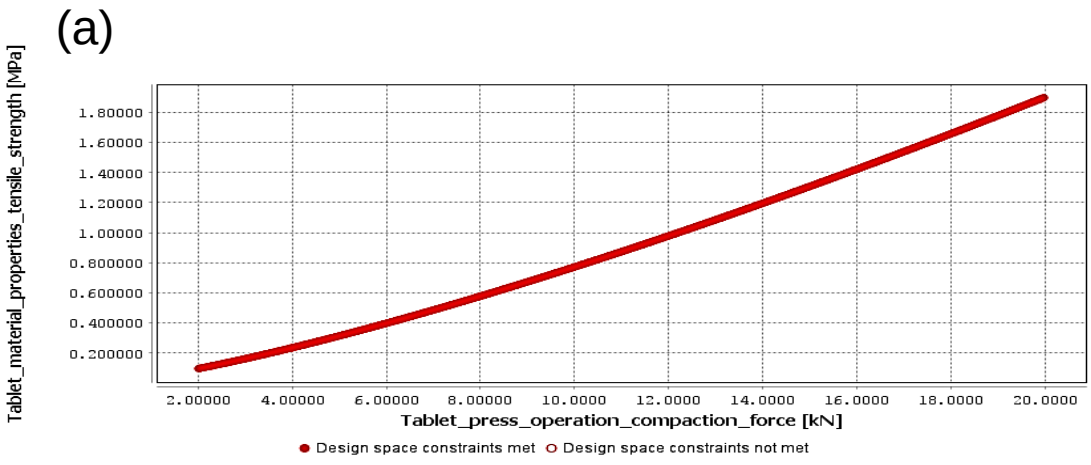
Where k_{cc} is the compressibility constant for confined compression, p is the applied pressure and p_0 is the theoretical pressure required to produce a zero-porosity compact. The compressibility parameters, k_{cc} and p_0, are determined for each component in the formulation using experimental tablet porosity as a function of tablet compaction pressure data. In order to account for changes in formulation, specifically API mass fraction, the Reynolds et al. compressibility model was applied to Equation 4.12 to take account of the relative contribution of the constitutive components.

Effect of tablet press CPPs on the tablet tensile strength

Parameters that can be varied that might impact tablet strength, combined with safe limits of operation include:

Table 4.9 Tablet press Standard Operating Conditions

Parameter	*Range*
Compression Force (kN)	2 to 20
Tablet mass (mg)	500 to 700
Turret speed (rpm)	20 to 80



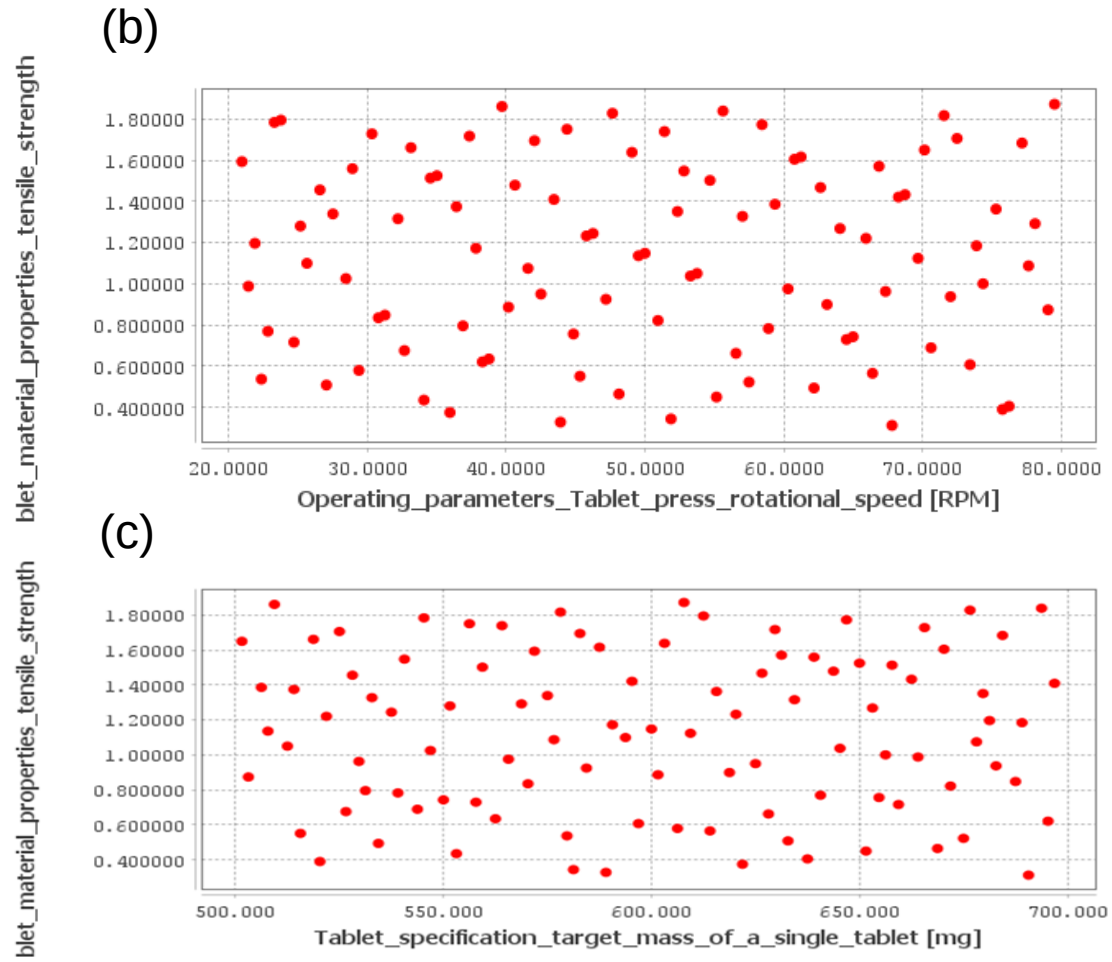


Figure 4.25 Effect of tablet press CPPs on the tablet tensile strength: (a) Compaction force, (b)Turret speed and (c) Die mass.

The analysis of Figure 4.25 revealed a critical trend: increasing compaction force leads to a corresponding increase in tablet tensile strength. This finding aligns with established principles of tablet compression, where higher force translates to denser tablets with greater resistance to breaking. Interestingly, the data suggests that die mass and turret speed do not exhibit a clear directional effect (trend) on tensile strength within the investigated range. This absence of a trend for these parameters simplifies the calibration process.

By focusing our calibration efforts primarily on the well-defined relationship between compaction force and tensile strength, we can develop a more targeted model. This approach allows us to calibrate the model's response to compaction force while maintaining fixed values for die mass and turret speed within their relevant operating ranges. As a result, the calibration becomes more efficient and computationally less demanding.

4.6.2 Rate Mechanism GSA of The Back End

Table 4.10 show tensile strength sensitivity index for the 5 kinetic parameters to be estimated. Dispersion coefficient is an equipment related parameter were used to predict the concentration of the blend in the feeding frame. Other parameters; Bonding capacity. Compressibility constant, compaction pressure at zero porosity and tensile strength at zero porosity are found in The Ryshkewitch–Duckworth equation is used for relating the tensile strength to the tablet porosity.

Equipment Dispersion coefficient parameter has sensitivity value close to zero, which confirms that it does not affect tensile strength at the operation range average. The ranking will help to understand the calibration method to be used. Compressibility constant of blend mixture has the highest impact with sensitivity index of 0.377 which suggests that it will be the most viable during calibration.

Table 4.10 Tablet press rate mechanism parameters sensitivity index values

Tablet press rate mechanism parameters	*Sensitivity Index*
Equipment Dispersion coefficient	0.0000
Bonding capacity of blend mixture	0.3521
Compaction pressure at zero porosity of blend mixture	0.1876
Compressibility constant of blend mixture	0.3768
Tensile strength at zero porosity of blend mixture	0.1198

4.6.1 Calibration

Figure 4.25 reveals a critical insight into the factors influencing tablet tensile strength. The data suggests that compaction force exhibits a trending effect on tensile strength, while the impact of turret speed and tablet mass appears to be more random or scattered. This finding is significant because it allows us to leverage calibration to improve model predictions. Since compaction force has a clear directional influence, I can adjust this parameter within the model for a given turret speed and tablet mass to achieve a more accurate prediction of tensile strength.

The calibration strategy employed this concept. Runs 1 and 3 were used for calibration set 1, focusing on the relationship between compaction force and tensile strength at specific turret speed and tablet mass settings. Similarly, runs 5 and 7 were used for calibration set 2. Runs 2, 4, 6, and 8 were then designated for the validation data to evaluate the model's accuracy with external data.

Table 4.11 Optimised parameters (Using run 1 and 3 for calibration 5 and 7 for calibration 2

Mechanism	*Parameters*	*Initial value*	*Optimised value*	
			1	2
Feed frame	Dispersion coefficient	10	49.9547	49.6753
Table press material properties	Bonding capacity	8	8.1702	8.1590
	Compressibility constant	10	9.7933	9.8626
	Compaction pressure at zero porosity	500	1035.46	1024.93
	Tensile strength at zero porosity	20	8.65912	8.05484

Figure 4.26 presents the results of calibrating the model using the optimal values obtained for the estimated parameters (Table 4.11). The difference between the model is the Compaction pressure at zero porosity. The suggest changing target tablet mass and turret speed has a direct correlation to the rate parameter. As illustrated in Figure 4.26 and Figure 4.27, this approach yielded a fitted granule size distribution that demonstrates a high degree of agreement with the experimental data .

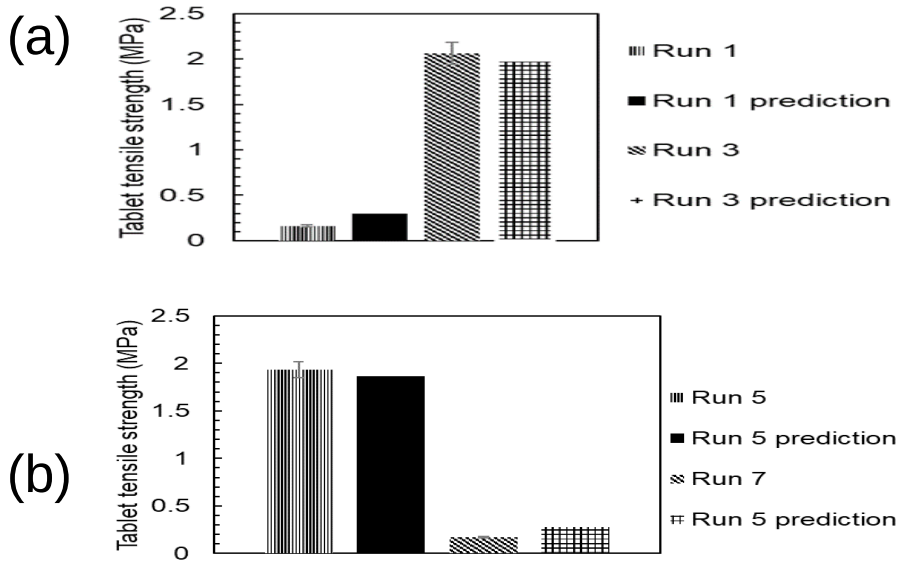


Figure 4.26 Model prediction vs Experimental data: (a) Run 1 and Run 3 and b) Run 5 and Run 7.

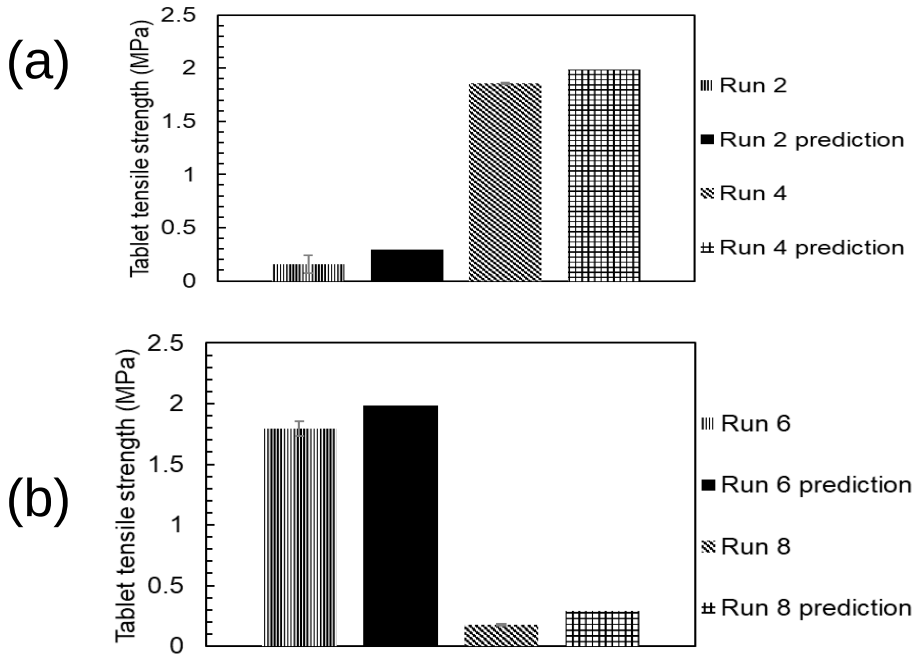


Figure 4.27 Model validation: (a) Run 2 and Run 4 model validation using Run 1 and 3 calibration and (b) Run 6 and Run 8 model validation using Run 5 and 7 calibration.

4.6.2 Linking Rate Mechanism Parameters with CPPs

GSA was conducted to determine which process parameter has the most important influence on the kinetic parameters of the model. The results in Figure 4.28 suggest that the sensitivity of the kinetic parameters is most likely driven by the value of this specific process parameter. The next step involved investigating whether the change in kinetic parameter sensitivity is substantial within a relevant range for this process parameter.

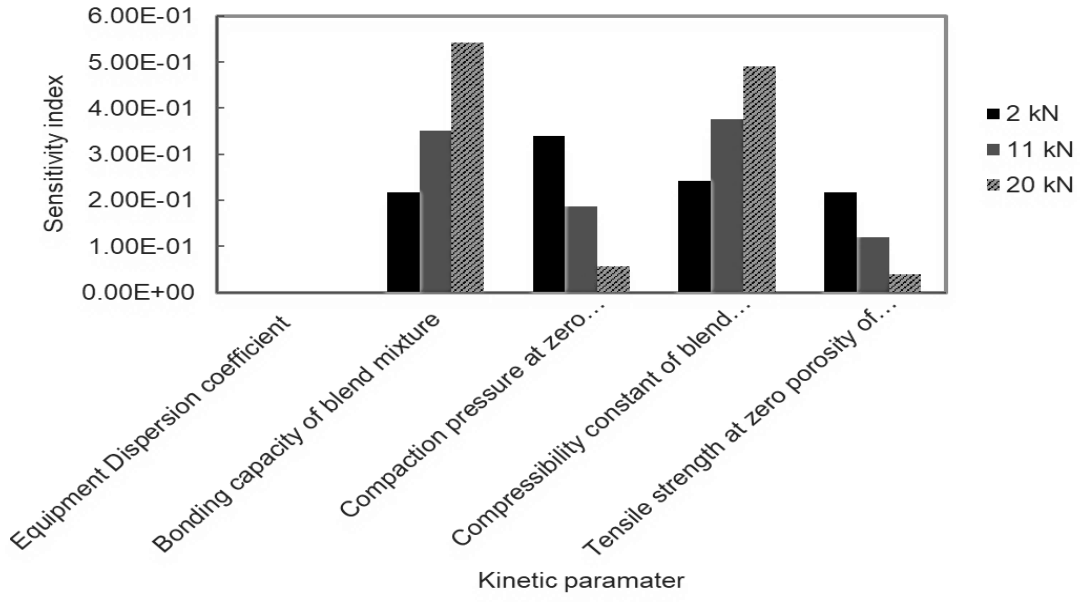


Figure 4.28 Effect of changing tablet press compaction force on sensitivity.

Focusing on the compressibility constant, a key parameter in the model, the analysis revealed that it indeed increases with increasing compaction force. Calibration of the compressibility constant was then performed for each experimental run. Interestingly, the calibration parameter values showed minimal change when the compaction force was increased from 2 kN to 4 kN in Figure 4.29. However, a clear linear relationship emerged between the compressibility constant and compaction force when the range was expanded from 2 kN to 4 kN.

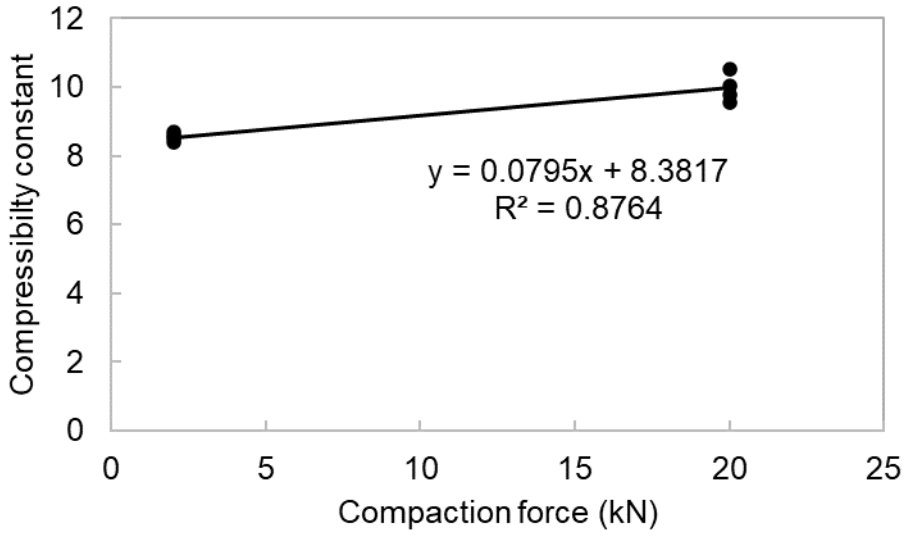


Figure 4.29 Compressibility constant and Compaction force mechanistic correlation.

This observation provides a valuable starting point for developing a mechanistic relationship between compressibility constant and compaction force. The observed trend suggests that higher compaction forces might lead to a denser packing of particles, thus influencing the material's resistance to further compression, which is reflected in the compressibility constant.

4.7 Chapter conclusion

Twin-screw wet granulation (TSWG) models often require calibration of over 25 rate mechanism parameters, making it challenging to pinpoint their individual effects on specific operating conditions. Sensitivity analysis becomes crucial for efficient process model development and calibration, especially when numerous input variables can significantly alter the model's outputs. This research introduces a novel approach to streamline the calibration stage by systematically identifying critical parameters. The analysis revealed that processes like nucleation, breakage, and layering exert the most significant influence on granule

development, while consolidation processes have a comparatively weaker impact. This valuable insight facilitated a 60% reduction in calibration parameters, focusing efforts on the ten most critical ones.

Furthermore, this study highlights the importance of considering PSD throughout the calibration process. By segmenting the particle size range and utilising all available data points within each run, the study revealed distinct sensitivity trends for breakage calibration parameters at different size fractions. This finding emphasizes the need for a nuanced approach to calibration that accounts for the complexities of particle size distribution within the model.

This section explored the development of models that can directly link formulation and process properties, such as the liquid-to-solid (L/S) ratio, with key process parameters like droplet size and standard deviation. This approach aims to reduce the number of calibration parameters needed for accurate model predictions. The investigation revealed an interesting relationship: granule size appears to increase linearly with droplet size, suggesting a potential dependence of nucleus size on initial droplet size. For L/S ratios ranging from 0.15 to 0.25, the calibration results, reveal an exponential correlation between the L/S ratio and the standard deviation of droplet size. This finding provides valuable insights into the mechanistic correlation between L/S ratio and nucleation processes. While further research is needed to fully elucidate the behaviour at higher L/S ratios, this initial exploration demonstrates the potential of linking formulation properties with process parameters to improve model accuracy and reduce calibration complexity. Similar efforts were undertaken to optimise the calibration of the compressibility constant, a key parameter in tablet pressing. This study demonstrated a linear correlation between compaction pressure and the compressibility constant, with the latter being the most sensitive parameter (total sensitivity index of 0.3). By performing calibration at different compaction forces, a direct link was established between these two variables with an absolute error of 0.4. This approach effectively reduces the number of parameters requiring calibration through computationally intensive techniques like Maximum Likelihood Estimation (MLE).

Porosity plays another crucial role in tablet quality will be investigated in future. It significantly impacts tensile strength and is considered a CQA for tablets. The wet granulation process, particularly the consolidation step, significantly

influences tablet porosity. During consolidation, the granules are pressed together, reducing void spaces within the tablet matrix, and ultimately affecting its overall porosity.

Chapter 5: Leveraging Artificial Intelligence to Create Fast Solving Models

5.1 Introduction

The chemical and biochemical sectors are experiencing a paradigm shift, driven by the urgent need for sustainable production methods. Continuous processing, with its inherent efficiency and reduced waste generation, has emerged as a frontrunner in this transformation. However, this shift comes with its own set of challenges. Continuous processes are inherently more complex and prone to disruptions compare BMs. Fortunately, the pharmaceutical industry is also experiencing a parallel digital revolution that offers solutions to these challenges. A multitude of new business opportunities are emerging, fuelled by advancements in digital technologies. DTs have attracted substantial industrial and research attention as a means of transitioning into continuous processing and bio-manufacturing while also allowing for more product customisation. It promises reduced time to market, enables manufacturing process optimisation, and is regarded as a critical enabler of continuous manufacturing. However, developing DTs now very expensive and lengthy. When it comes to shortening the time to market, the time required is particularly challenging.

Digital manufacturing is the deployment of Information Technology (IT) and Operational technology (OT), cyber-physical systems, widespread and centralised connectivity, sensing and actuation, and the use of data and analytics across the value chain to provide actionable insights into manufacturing and supply chain (Javaid et al., 2023). Models are crucial to Digital manufacturing as actively support control strategies and process optimisation. One limitation of data-driven models, such as DNN is their lack of interpretability, making it difficult to understand the underlying mechanisms and reasoning behind their predictions. Such models consist of explicit relations between inputs and outputs meaning that they execute very rapidly and robustly over the domain for which they are fitted. For all input combinations within a given domain, it can be guaranteed that these models can predict the detailed models with a specified accuracy. This constriction for AI-based models can be offset using first-principles knowledge whenever possible.

Artificial neural networks (ANNs) are a type of machine learning model that utilize an input layer, hidden layer, and output layer. Connections between these layers are assigned weights, which are adjusted during the training process to effectively map input data to output data. While widely used, ANNs can suffer from overfitting, a phenomenon where the model becomes overly complex and performs well on the training data but poorly on testing data. However, more recent techniques, such as Bayesian regularised artificial neural networks, have been developed to mitigate this issue (Ekins, 2016). Deep learning tools, including powerful libraries like TensorFlow and Keras, are readily available in open-source software environments with large computing power like R and Python, making it easier to implement and train complex neural networks.

A major limitation of mechanistic models is their computational intensity, making them inappropriate for real-time applications like MPC. MPC requires models that can predict future process behaviour within seconds to make timely decisions. Therefore, faster model alternatives are necessary for industrial implementation. Surrogate modelling offers a potential solution to reduce time of process simulators by replacing complex mechanistic models with machine learning-based models. The effectiveness of this surrogate modelling framework has been demonstrated through successful application to steady-state simulations. This work explores the potential of surrogate modelling, a specific type of hybrid approach, to accelerate the deployment of digital twins in chemical and biomanufacturing. Surrogate models leverage machine learning to create computationally efficient alternatives to rigorous first-principles models. This allows for:

- Faster simulations: Surrogate models can significantly reduce the computational time required for process simulations, enabling more rapid optimisation and decision-making.
- Real-time applications: The speed of surrogate models makes them suitable for online applications such as model predictive control (MPC), which can further improve process performance.
- Early-stage integration: The flexibility of surrogate models allows them to be incorporated into digital twins even during the early stages of process development.

This chapter focuses on developing a surrogate modelling methodology for a wet granulation process - a crucial step in many pharmaceutical manufacturing workflows. By applying various machine learning algorithms, I aim to create useful and computationally efficient surrogate models that can be seamlessly integrated into the digital twin framework. This chapter delves deeper into the specific machine learning algorithms used, the development of the hybrid model, and the limitations associated with this approach. It also explores potential future directions for research in this rapidly evolving field.

5.2 Materials and Methods

First-principle-model flowsheet developed produces high fidelity results. However, such models can be very computationally demanding which inhibits their use in digital applications. Surrogate Modelling, a method to simplify the model yet retain sufficient accuracy can used for use cases where the model must be solved very quickly. Such surrogate models consist of explicit relations between inputs and outputs, meaning that they execute very rapidly in a robust over the domain for which they are fitted.

The key objective of the developed model is to achieve predictive accuracy and through proper sampling methods the model can generate a reasonable amount of pseudo data. These data sets are used to train a more computationally efficient and or reliable surrogate. The resulting surrogate model can typically be solved much faster than the original model, but its predictive accuracy is limited by the extent of the sampling carried out to obtain the training data and the data-driven method used to generate the model.

The workflow for developing a surrogate model flow sheet can be grouped into four stages.

- First, a working flow sheet or custom model mainly based on 1st principle was created.
- Then a Global System Analysis tool was to generate extensive synthetic input output data from the first principles model.
- The data was then used to generate the surrogate models.

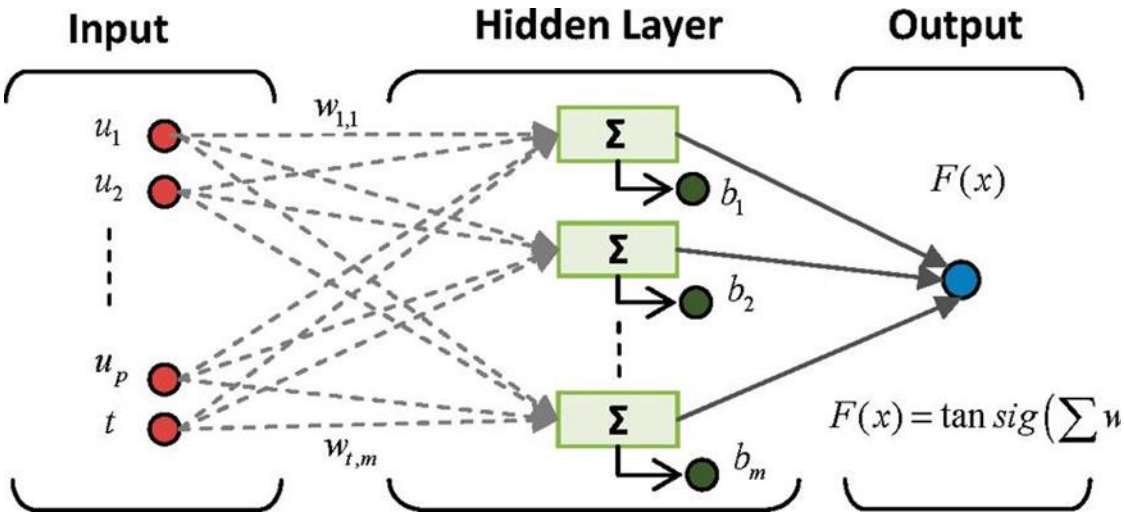


Figure 5.1 Machine learning algorithm structure; time t included for dynamic modelling and excluded for steady state modelling.

$$z_j^{(l)} = \sum_{i=1}^{N_l} w_{ji}^{(l)} x_i^{(l-1)} + b_j^{(l)} for\ j = 1, \dots, J_l \quad 5.1$$

where $z_j^{(l)}$is activation j at layer l, $x_i^{(l-1)}$is the input from layer $l-1$ to layer l, $w_{ji}^{(l)}$ is a weight parameter at layer l, $b_j^{(l)}$ is a bias parameter.

Neural networks can be thought of as layers of computation units, neurons with connections between consecutive layers. The network is comprised of one input layer with a pre-determined number of neurons which are equal to the number of model input parameters , one or more hidden layers with a custom number of neurons, and one output layer as illustrated in Figure 5.1. In the network configuration, the user can see the input and output layers, which are automatically generated by gPROMS and can add one or more hidden layers. gPROMS currently supports numerous types of different types of ANN activation functions, either linear or non-linear. The non-linear activation functions include:

- Rectified Linear unit
- Sigmoid
- Hyperbolic Tangent.

5.3 Data Generation

Synthetic data generation offers a powerful approach to address challenges in machine learning, particularly when real-world data is scarce (Klein & Bergmann, 2018). This technique involves leveraging validated simulation models to create

realistic, yet controlled, data sets. This data can then be combined with operational data from real systems to provide a richer training ground for MLs.

Synthetic data, generated artificially, offers significant value in scenarios where collecting real-world data is difficult or cost prohibitive, even in data-scarce environments (Saxena et al., 2008). Synthetic data generated from simulations can serve as a safe and cost-effective substitute, allowing researchers to explore these rare events and train machine learning models for failure prediction. Beyond expanding the data range and enhancing model performance in scarce-data regimes, synthetic data offers additional benefits. It can be particularly useful in overcoming class imbalance issues commonly encountered in classification problems. By generating synthetic data points for underrepresented classes, I can create a more balanced training set, leading to more robust and generalisable models.

In essence, synthetic data generation emerges as a versatile tool for empowering machine learning in diverse scenarios. It offers a safe, controlled, and cost-effective approach to data augmentation, ultimately leading to more robust and generalisable machine learning models. After completing the preprocessing and split, I have a dataset ready to be fed into the surrogate model. This training data will allow the model to learn the complex relationships between the input features and the desired outputs.

5.4 Evaluation Criteria

A crucial step in developing any machine learning model is evaluating its accuracy. The model's predictive capability is assessed by comparing its predictions to observed or validated synthetic values. Common metrics used for this evaluation include Mean Absolute Error (MAE), Relative Absolute Error (RAE), Root Mean Square Error (RMSE), and R-squared (R^2).

$$MSE = \frac{\sum_i |f_i - y_i|}{n} \quad 5.2$$

$$RAE = \frac{\sqrt{\sum_i (f_i - y_i)^2}}{\sqrt{\sum_i (y_i)^2}} \quad 5.3$$

$$RMSE = \sqrt{\frac{\sum_i (f_i - y_i)^2}{n}} \quad 5.4$$

$$R^2 = 1 - \frac{\sum_i (f_i - y_i)^2}{\sum_i (f_i - \bar{y})^2} \quad 5.5$$

f and y refer to predicted and observed values, respectively. i also refers to the set of experimental runs and n refers to the number of measurements.

5.5 Surrogate modelling

This step focuses on preparing the data for training the deep learning surrogate model. The data originates from a global system analysis conducted using a previously calibrated and validated mechanistic model using optimised value in Table 5.1 using the operational boundaries in Table 5.2.

Table 5.1 Twin screw wet Optimised parameters (Wang et al,2022)

Mechanism	*Parameters*	*Optimised value*
Nucleation	Droplet mean size Standard deviation	0.52
	Droplet pore penetration	0.40
Breakage	Breakage rate in CE	1.5
	Breakage rate in KE	2.5
	Scale parameter in CE	2.54e10
	Shape parameter in KE	8.23
	Fraction of fines in CE	0.4
Layering	Growth rate	2666.5

Table 5.2 Twin screw wet granulator operational boundaries

Parameter	*Lower bound*	*Upper bound*
Screw speed(rpm)	200	900
Powder feed rate(kg/hr)	5	25
Liquid rate(kg/hr)	0.9	11.25

The Maximum Liquid to Solid ratio (L/S) to make granules and prevent paste is 0.45. Low screw speed combined with a high powder flow rate results in higher torque and it may exceed the safety limit of the equipment, therefore the machine may stop, and you should avoid these conditions.

5.5.1.1 Surrogate model Configuration

To develop the machine learning model, a series of steps were undertaken. Key parameters influencing the performance of the twin-screw machine were identified and incorporated into the model. Table 5.3 provides a detailed overview of these parameters, including their significance and how they were integrated into the model.

By carefully selecting and tuning these parameters, the machine learning model was optimised to accurately predict and control the twin-screw machine's behaviour, leading to improved process efficiency and product quality. The preprocessing stage involves transforming and manipulating the raw data to ensure it's suitable for the deep learning surrogate model. This includes:

- **Feature engineering:** Creating new features from existing data or combining existing ones to improve model performance.
- **Normalisation or standardisation**: Scaling the data to a common range to prevent certain features from dominating the training process (Table 5.3).
- **Handling missing values:** Employing techniques like imputation or deletion to address any missing data points.
- **Train-Test-Validation Split**

Once the data is pre-processed, I split it into three distinct sets for training, testing, and validation. The specific ratio of this split depends on the total data size. Here's a breakdown of each set's role:

1.Training set (largest portion): Used to train the deep learning surrogate model by exposing it to correlations within the data.

2.Test set (smaller portion): Evaluate the trained model's performance on external data.

3. Validation set (optional, smaller portion): Used to fine-tune hyperparameters (model settings) during the training process. This helps prevent overfitting.

Table 5.3 Surrogate model configuration.

Surrogate model inputs			
Symbols	*Variables*		
U1	Liquid mass flow rate	kg/h	
Surrogate model output			
Y1	D10	µm	
Y2	D50	µm	
Y3	D90	µm	
Y4	Mass fraction screen 1(2000 µm)	kg/kg	
Y5	Mass fraction screen 1(1180 µm)	kg/kg	
Y6	Mass fraction screen 1(850 µm)	kg/kg	
Y7	Mass fraction screen 1(500 µm)	kg/kg	
Y8	Mass fraction screen 1(250 µm)	kg/kg	
Y9	Mass fraction screen 1(90 µm)	kg/kg	
Y10	Mass fraction screen 1(45 µm)	kg/kg	
Scaling			
Input scaling			Normalised
Output scaling			Normalised
Number of points			
...used for training			100
...used for cross validation			200
Neural Network			
Number of hidden layers			2
Layer	Activation function		Number of neurons
Hidden layer 1	Tanh		30
Hidden layer 2	Tanh		30
Output	Linear		1

5.5.1.2 Model evaluation

The following Table 5.4 provides a detailed evaluation of each set, training, and validation. This evaluation includes metrics such as RMSE and R^2, allowing for a clear comparison of their performance. By analysing these results, one can gain valuable insights into strengths and weaknesses of each set influencing performance areas for improvement.

Table 5.4 PSD Distribution Machine learning evaluation table.

Accuracy metrics	*Dataset*	*MAE*	*RAE*	*RMSE*	*R^2*
D10	Training	0.0181078	0.000530503	0.0226737	0.998985
	Cross validation	0.0184147	0.000538770	0.0230374	0.998973
D50	Training	0.57	0.00442550	0.679053	0.951188
	Cross validation	0.57	0.00449420	0.689237	0.948575
D90	Training	0.242677	0.000283546	0.280787	0.999966
	Cross validation	0.244845	0.000286712	0.283553	0.999966

Once the machine learning model has demonstrated robust performance in controlled environments, the next crucial step is to deploy it into real-world applications. This process, known as operationalisation, involves integrating the model into existing systems and workflows to deliver practical value.

In this specific case, the developed machine learning models found in appendix C were seamlessly integrated into the gPROMS process flowsheet as custom models. This integration enabled the models to provide real-time predictions and insights, directly influencing decision-making and optimising process operations.

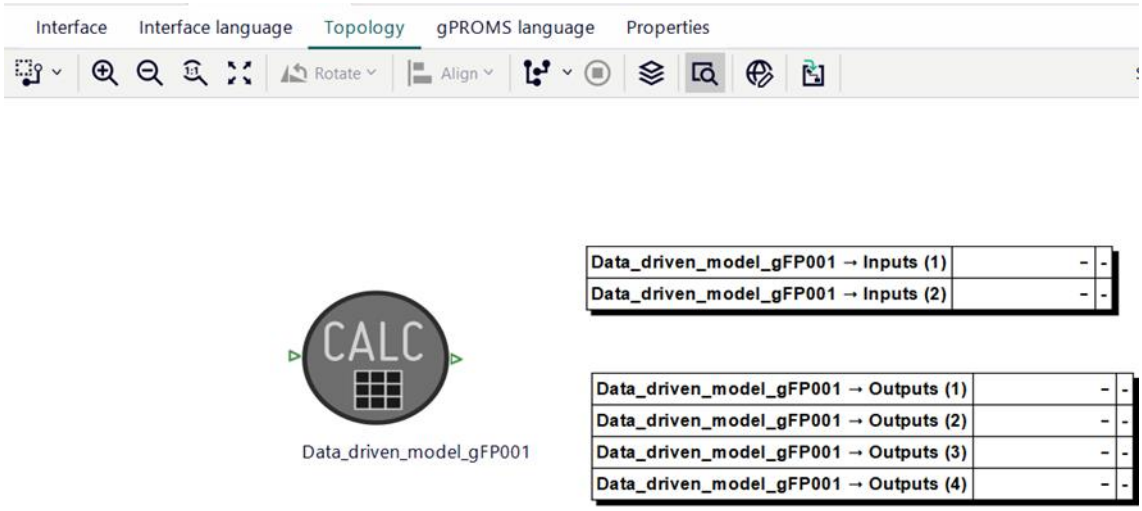


Figure 5.2 Process Flowsheet with Integrated ML Models.

The predictive capabilities of the deployed models were evaluated, and the results are illustrated in the following figures.

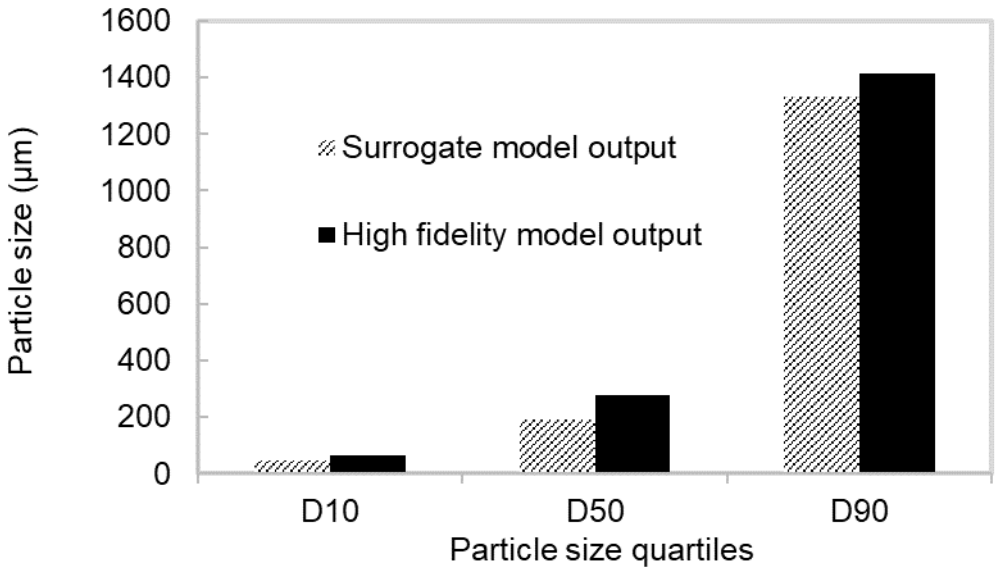


Figure 5.3 PSD Quartile prediction; surrogate model vs high fidelity model.

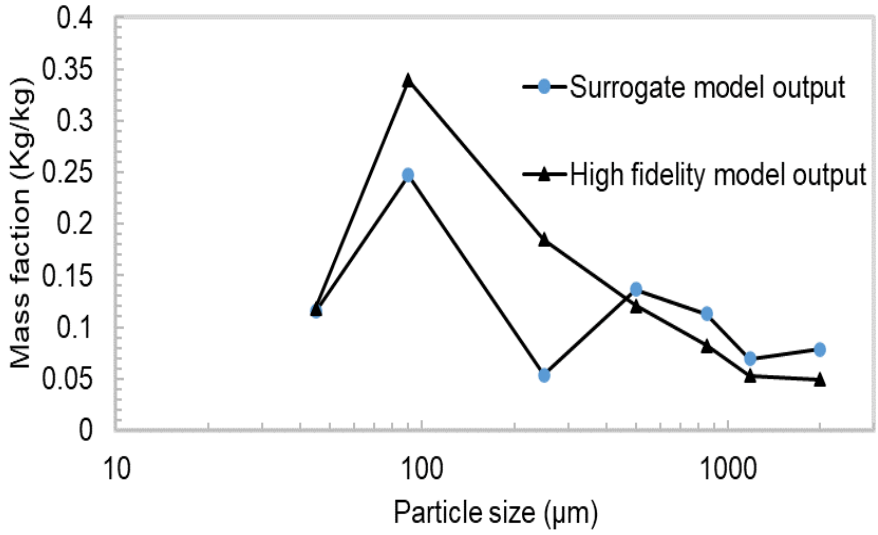


Figure 5.4 PSD prediction; surrogate model vs High fidelity model.

Table 5.5 Simulation stats; ML surrogate model prediction and High-fidelity model prediction

Model output	*Surrogate*	*High fidelity model*
Number of equations	10	54271
Simulation time (s)	0.40672	83.7564

By operationalising the machine learning models, significant benefits were realised, including:

- **Enhanced decision-making:** Real-time predictions empowered operators and engineers to make informed decisions, leading to improved process efficiency and product quality.

- **Optimised process control:** The models enabled precise control of process variables, minimising deviations, and maximising yields.
- **Predictive maintenance:** By anticipating potential equipment failures, maintenance costs were reduced, and downtime was minimised.
- **Accelerated product development:** The models facilitated rapid experimentation and optimisation of product formulations.

Overall, the successful operationalisation of these machine learning models demonstrates the transformative potential of AI in the manufacturing industry. By bridging the gap between research and practical application, these models have the power to revolutionise processes, drive innovation, and create sustainable competitive advantages.

5.6 Hybrid Modelling: A Faster Path to Model Maintenance

The first step is to identify the most impactful kinetic parameters influencing the process. This typically involves analysing the underlying reaction mechanisms and pinpointing the parameters that have the most significant effect on the Critical Quality Attribute (CQA) of interest. In this case, the CQA is the average particle size produced by the twin-screw granulator.

Once the key kinetic parameters are identified, the next step involves generating input and output data for these impactful factors. Performing experiments were not used to generate the data in this case during the high complexity. This data was obtained from leveraging existing data from published research or industry databases and utilising simulation software to generate virtual data under controlled conditions.

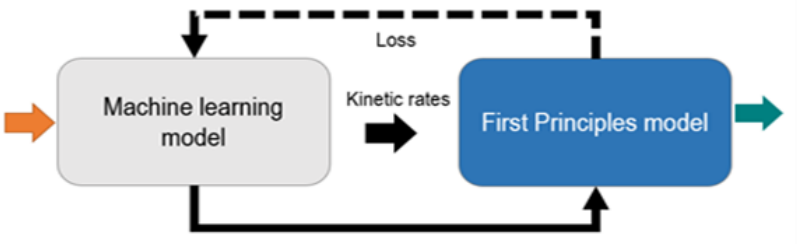


Figure 5.5 Hybrid model (ML and PBM) algorithm structure.

5.6.1.1 Surrogate model configuration

To assess the impact of changing the kinetic parameters on the average particle size, a GSA was performed. As shown in Table 5.7, GSA allows us to quantify the influence of each parameter on the CQA and identify the parameters with the greatest effect. This information is crucial for focusing modelling efforts on the most critical factors and ultimately developing a more accurate and efficient model.

Table 5.6 Twin screw wet granulator estimated parameter GSA ranking.

High impact parameters	*Average Total sensitivity index*
Breakage rate constant (CE)	1.4423(1)
Breakage rate constant (KE)	0.0298(7)
Fraction of fine particles (CE)	1.3442(2)
Breakage scale parameter (KE)	0.0009(10)
Breakage shape parameter (KE)	0.0031(9)
Drop nucleation pore penetration	0.7595(5)
Drop nucleation Std Dev of initial droplet size	0.0236(8)
Layering critical moisture content	2.8978(4)
Layering kinetic parameter λ	0.3144(6)
Layering maximum growth rate	1.1758(3)

A kinetic parameter machine learning model found in Appendix B was developed and trained using default hyperparameters, including learning rate, optimiser, and loss function.

Table 5.7 Twin screw wet granulation rate mechanism parameter ML Surrogate model Configuration.

Surrogate model inputs			
Variables			
U1	Liquid mass flow rate	$kg.h^{-1}$	
U2	Mass distribution on size 1	$kg.kg^{-1}\mu m^{-1}$	
U3	Mass distribution on size 2		
U4	Mass distribution on size 3		
U5	Mass distribution on size 4		
U6	Mass distribution on size 5		
U7	Mass distribution on size 6		
U8	Mass distribution on size 7		
U9	Mass distribution on size 8		
U10	Mass distribution on size 9		
U11	Mass distribution on size 10		
U12	Mass distribution on size 11		
U13	Mass distribution on size 12		
U14	Mass distribution on size 13		
U15	Mass distribution on size 14		
U16	Mass distribution on size 15		
U17	Mass distribution on size 16		
U18	Mass distribution on size 17		
U19	Mass distribution on size 18		
U20	Mass distribution on size 19		
U21	Mass distribution on size 20		
Surrogate model output			
Y1	Breakage rate constant on CE		
Y2	Breakage rate constant on KE		
Y3	Fraction of fine on CE		
Y4	Layering maximum linear growth rate		
Y5	Pore penetration		
Y6	Std deviation of initial droplet size		
Scaling			
Input scaling			Normalised
Output scaling			Normalised
Number of points			
...used for training			100
...used for cross validation			200
Neural Network			
Number of hidden layers			2
Layer	Activation function		Number of neurons
Hidden layer 1	Tanh		30
Hidden layer 2	Tanh		30
Output	Linear		1

5.6.1.2 Model evaluation

In the final step, the model's performance was evaluated by simulating its predictions on the test data and comparing the results to expected outcomes. Additionally, the simulation time was measured.

Table 5.8 Twin screw wet granulation breakage parameter surrogate model prediction vs High fidelity model calibration.

Surrogate model output			
Y1	Breakage rate constant on CE	39.4669	1.5
Y2	Breakage rate constant on KE	19.1587	2.5
Y3	Fraction of fine on CE	-0.6226	0.4

Table 5.9 Simulation stats; Twin screw wet granulation layering surrogate model prediction vs High fidelity model calibration.

Surrogate model output			
Y4	Layering growth rate	2557.346	2666.5

Table 5.10 Breakage kinetic parameter neural network evaluation.

Accuracy metrics	*Dataset*	*MAE*	*RAE*	*RMSE*	R^2
Breakage rate constant on CE	Training	2.39267	0.221972	3.01623	0.759478
	Cross validation	2.50472	0.221653	3.06428	0.743218
Breakage rate constant on KE	Training	1.16414	0.0778259	1.40392	0.973640
	Cross validation	1.16527	0.0821685	1.42362	0.970333
Fraction of fine on CE	Training	0.0271737	0.0615636	0.0391970	0.975856
	Cross validation	0.0297212	0.0759655	0.0479952	0.964419

Table 5.11 Layering kinetic parameter neural network evaluation.

Accuracy metrics	*Dataset*	*MAE*	*RAE*	*RMSE*	R^2
Layering maximum linear growth rate	Training	1874.63	0.330813	2184.92	0.0111535
	Cross validation	1874.69	0.330293	2193.18	0.00926839

The machine learning model developed to predict breakage rate constants in the CE and KE , as well as the layering growth rate, exhibited a high degree of error as shown in Table 5.10 and Table 5.11. The calculated breakage fraction of fine kinetic parameter, produced a negative result, indicating a physically unrealistic outcome. This anomaly arises from the model's formulation, which does not incorporate constraints to enforce non-negativity. The high error can also be attributed to the inherent challenges of solving inverse problems. Unlike forward

problems with well-defined unique solutions, inverse problems may have high sensitivity to input variations. Small changes in the input data can lead to significant changes in the predicted parameters, making the model unstable. Figure 4.19 in chapter 4 illustrates the influence of varying breakage rate constants on the particle size distribution. For smaller particle sizes, there's a clear trend: increasing the breakage rate constant leads to a higher mass fraction of smaller particles. However, for larger particle sizes, the effect appears more random, suggesting a complex interplay with other factors.

Our analysis also identified layering maximum growth rate, breakage rate constant in CE, and breakage rate constant in KE as the most sensitive kinetic parameters. This sensitivity contributes to the high prediction errors observed in the model. Despite these limitations, the model can still provide an initial estimate of the kinetic parameters, serving as a starting point for further refinement.

Furthermore, the error can be significantly reduced by increasing the training data size. As demonstrated in the experiment, increasing the training samples for the nucleation rate parameter model from 100 to 250 resulted in an 0.99 R-squared value (a measure of model fit). This highlights the importance of using a substantial and representative dataset for training machine learning models.

Table 5.12 Twin screw wet granulation nucleation surrogate model prediction vs High fidelity model calibration.

	Surrogate model output		
Y5	Pore penetration	0.68	0.52
Y6	Std deviation of initial droplet size	0.712	0.40

Table 5.13 Drop nucleation kinetic parameter neural network evaluation.

Accuracy metrics	*Dataset*	*MAE*	*RAE*	*RMSE*	R^2
Pore penetration	Training	0.00250387	0.00846911	0.00333945	0.998990
	Cross validation	0.00263801	0.00920760	0.00358751	0.998838
Stvd Dev of initial droplet size	Training	0.00342300	0.00960545	0.00438456	0.991983
	Cross validation	0.00350624	0.00961306	0.00443592	0.992709

While PBMs are the traditional approach for modelling twin-screw granulation, this study explores the potential of artificial neural networks (ANNs) as a complementary tool. PBMs excel at tracking particle property changes throughout

the process, predicting key outputs like PSD, liquid content, and porosity. The developed ANN, however, demonstrates promise for rapid and reliable prediction (Table 5.14), even if its accuracy is currently lower than established PBM methods.

Table 5.14 Simulation stats; ML surrogate model prediction and High-fidelity model calibration.

Model output	*Surrogate*	*High fidelity model*
Number of equations	10	54271
Simulation time (s)	0.40672	83.7564

This advantage in speed and reliability makes ANNs suitable for implementing model predictive control (MPC) strategies. Furthermore, the ANN can be coupled with PBM models to accelerate the calibration process during continuous pharmaceutical manufacturing development.

5.7 Back-end case study

5.7.1 Surrogate modelling

Figure 5.6 reveals that compaction force is the primary factor influencing tablet tensile strength. This aligns with the findings of Mate et al., who investigated the effects of turret speed, compaction force, and target mass (fill cam height) on tensile strength. Since turret speed and target mass have negligible impacts, I simplified the surrogate model by excluding them. While a clear linear trend exists between compaction force and tensile strength, other models are considered for further analysis.

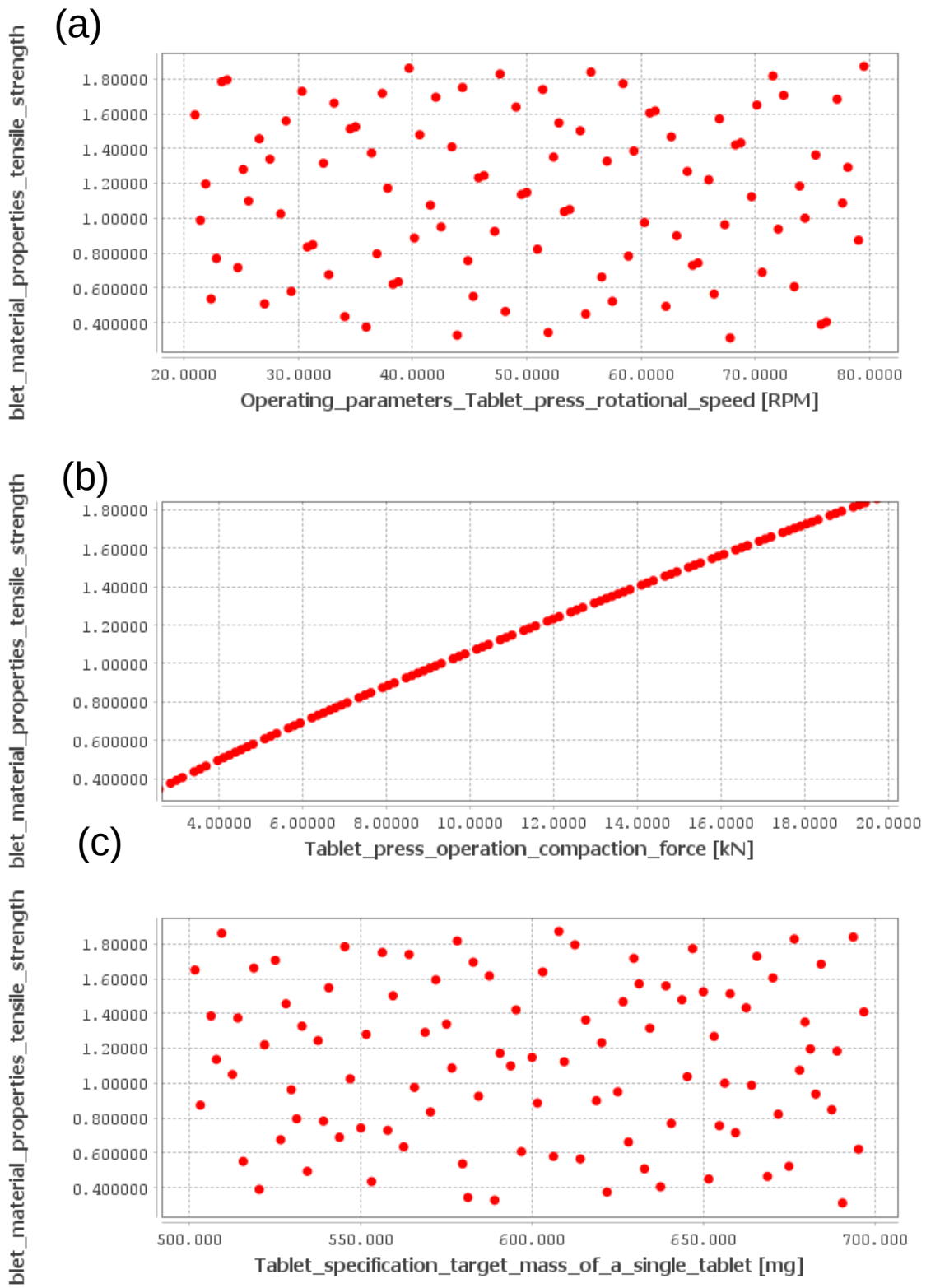


Figure 5.6 Effects of table press process parameters on tablet tensile strength: (a) Tablet press turret speed, (b) Compaction force and (c) target mass of a single tablet mass.

The choice of a surrogate model depends on the complexity of the relationship between input parameters and the output. In this case, the dominance of compaction force suggests that a simpler model might be sufficient. Linear regression, for example, could effectively capture this linear relationship.

5.7.1.1 Surrogate model configuration: Automated Learning of Algebraic Models (ALAMO) Algorithm

This approach excels at identifying low-complexity models from limited data. It selects the best combination of basic functions to build a model, accounting for potential interactions between these functions. While ALAMO struggles with large datasets, it could be a valuable option if computational limitations exist. The final surrogate model will be recommended by ALAMO based on a trade-off between model complexity and residual error.

The disadvantage of ALAMO algorithm lies its weakness in performing large training data sets. The advantage of the neural network generator over ALAMO is that it can perform efficient nonlinear regression and is therefore capable of fitting highly complex data if correctly set up. The large number of parameters that represent the network allows for fitting a much more complex model than might be expected, and these parameters scale well with large data sets. However, the neural networks generator may struggle with small training sets where the underlying structure is poorly communicated.

By evaluating these options, I can choose the most suitable model for accurately predicting tensile strength based on compaction force, while considering factors like data size and potential non-linearities.

5.7.1.2 Model evaluation

Table 5.15 Tensile strength and compaction force correlation using experimental data.

Regression model	*R*	*R^2*	*Adjusted R^2*	*Std. Error*
Linear	0.997	0.993	0.992	0.082
Logarithmic	0.997	0.993	0.992	0.082
Quadratic	0.997	0.993	0.992	0.082
Inverse	0.997	0.993	0.992	0.082
Cubic	0.997	0.993	0.992	0.082

Table 5.16 Tensile strength and compaction force linear correlation stats (experimental data).

ANOVA					
	Sum of Squares	*dof*	*Mean Square*	*F*	*Sig.*
Regression	6.088	1	6.088	909.119	<0.001
Residual	0.040	6	0.007		
Total	6.129	7			

Table 5.17 Tensile strength and compaction force linear correlation coefficients (experimental data).

Coefficients					
	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
Compaction force	0.097	0.003	0.997	30.152	<0.001
Constant	-0.027	0.046		-0.594	0.574

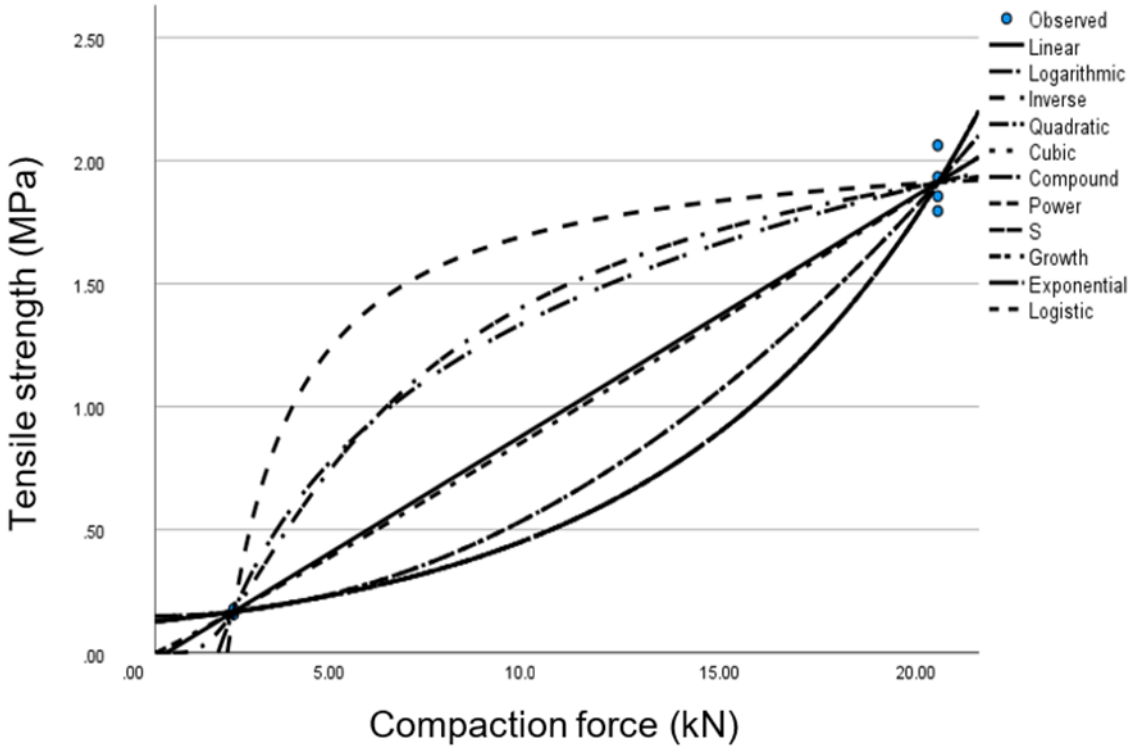


Figure 5.7 Tensile strength and compaction force correlation (experimental data).

Table 5.18 Tensile strength and compaction force correlation using synthetic data.

Regression model	*R*	*R^2*	*Adjusted R^2*	*Std. Error*
Linear	0.999	0.998	0.998	0.019
Logarithmic	0.996	0.952	0.951	0.099
Quadratic	1.000	1.000	1.000	0.005
Inverse	0.877	0.770	0.764	0.219
Cubic	1.000	1.000	1.000	0.003

As shown in Table 5.18, the quadratic regression polynomial yielded the best model performance. While a higher-degree polynomial might potentially improve the fit, it could also lead to overfitting, compromising the model's ability to generalise to new data.

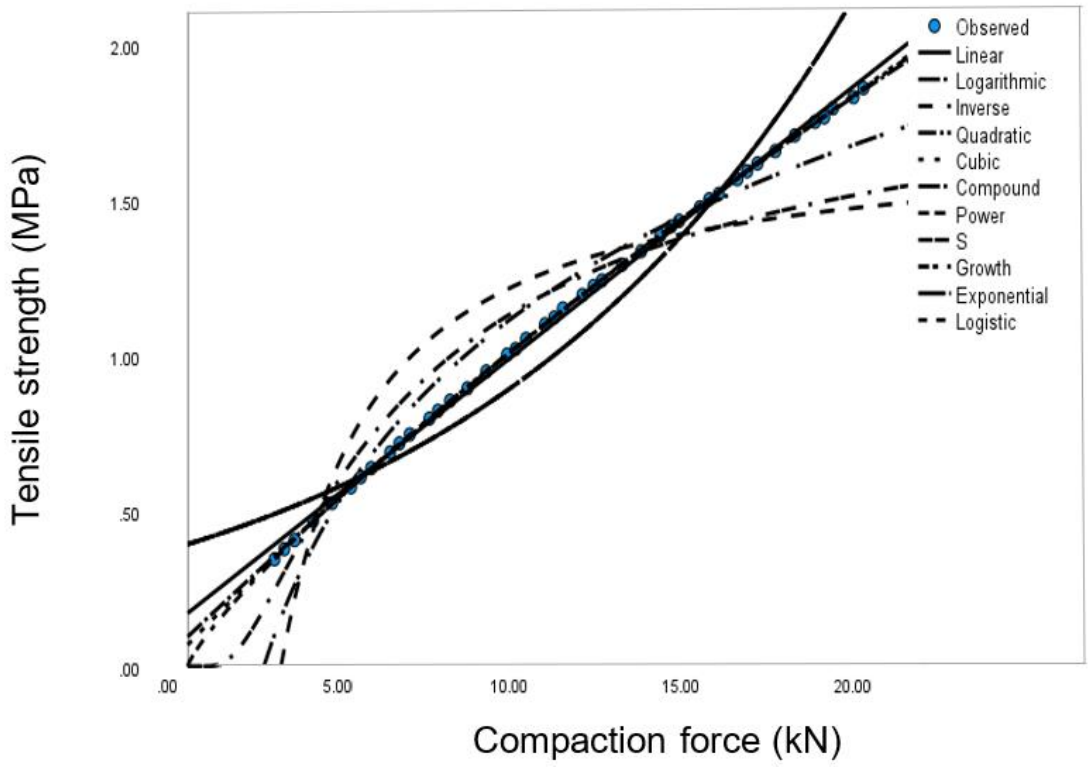


Figure 5.8 Tensile strength and compaction force curve correlation (synthetic data).

Linear regression analysis exhibited a high error in the constant, as illustrated in Figure 5.7. This can be attributed to insufficient data encompassing the entire operating space of the process. By combining experimental data with generic or uncalibrated mechanistic models, I was able to significantly reduce the error as shown by Figure 5.8. Mechanistic models offer valuable insights into the

underlying physical and chemical processes. They can be used to predict trajectories and fill in gaps within the limited data collected from experiments. In essence, these models act as flexible, data-driven tools that can be extrapolated beyond the available data points, leading to a significant reduction in the amount of experimental data required. However, determining the optimal amount of experimental data remains crucial. Finding the right balance ensures model accuracy while minimising the resource burden of extensive experimentation.

5.7.2 Hybrid modelling

The section showcase developed and trained kinetic parameter machine learning models for the tablet press case study. Default Hyper-parameters like learning rate, optimiser and loss function were used.

5.7.2.1 Surrogate model configuration

Table 5.19 Tablet press rate mechanism parameter ML Surrogate model Configuration.

Surrogate model inputs			
Variables			
U1	Tensile strength	MPa	0.17
Surrogate model output			
Y1	Dispersion coefficient (separated)		
Y2	Bonding capacity		
Y3	Compressibility constant		
Y4	Compaction pressure at zero porosity		
Y5	Tensile strength at zero porosity		
Scaling			
Input scaling			Normalised
Output scaling			Normalised
Number of points			
...used for training			250
...used for cross validation			50
Neural Network			
Number of hidden layers			2
Layer	Activation function		Number of neurons
Hidden layer 1	Tanh		30
Hidden layer 2	Tanh		30
Output	Linear		1

5.7.2.2 Model Evaluation

Table 5.20 Tablet press rate mechanism parameter neural network models evaluation.

Accuracy metrics	*Dataset*	*MAE*	*RAE*	*RMSE*	R^2
Bonding capacity of the blend mixture					
	Training	0.505096	0.0854604	0.601818	0.442595
Tensile strength at zero porosity					
	Training	0.563881	0.109668	0.652825	0.54183
Compaction pressure at zero porosity					
	Training	39.9148	0.11477	45.4056	0.0310863
Dispersion coefficient (Separate Surrogate model)					
	Training	0.00670800	0.00110336	0.00778191	0.999907
Compressibility constant					
	Training	0.505096	0.0854604	0.601818	0.442595

Table 5.21 Tablet press rate parameter surrogate model prediction vs High fidelity model calibration.

Surrogate model input		*Integrated Surrogate model output*	*Separate Surrogate model output*	*Optimal calibrated model output*	*Reference values*
Y1	Dispersion coefficient	N/A	47.1552	49.6753	10
Y2	Bonding capacity	8.41221	N/A	8.15895	8
Y3	Compressibility constant	7.26545	N/A	9.86257	10
Y4	Compaction pressure at zero porosity	414.218	N/A	1024.93	500
Y5	Tensile strength at zero porosity	5.99069	N/A	8.05484	20

Analysis revealed that only tensile strength at zero porosity and bonding capacity from the blending model yielded high R-squared values. This finding aligns with the parameters exhibiting high sensitivity indices and strong correlations with tensile strength, confirming their importance in predicting final product characteristics.

Separating the kinetic model and limiting the number of interactions between input factors led to an increase in model accuracy, as illustrated in the Table 5.21. However, the resulting predicted values for compaction pressure at zero porosity and tensile strength at zero porosity fell outside the established reference range. This discrepancy suggests an inaccurate surrogate model prediction. Therefore, it's necessary to develop a more robust and reliable boundary decision strategy

that goes beyond a simple 20% deviation from the reference values. Further investigation is required to identify a more appropriate approach for defining appropriate boundaries.

5.8 Chapter Conclusion

Developing a sufficiently accurate model is one of the most time-consuming and resource-intensive aspects of building a DT. Nowadays it is possible to build rigorous, first principles for most processes. However, with the advent of digital operations applications, such rigorous models are required for the deployment of online applications. Two common problems can arise the models developed in Chapter 4 are too computationally demanding for the intended application and lack robustness. To address this challenge, hybrid modelling approaches. These approaches combine accurate nonlinear surrogate models derived from large-scale physics-based models with machine learning techniques. This integration offers an opportunity to explore the design space and incorporate fundamental process knowledge into operational decision-making systems. In this study, fundamental modelling technique, PBM, was combined with data driven methods such as ML.

Calibrated mechanistic models can be used to create surrogate models that execute rapidly and robustly over a domain. First-principle-model flowsheets produce high-fidelity results, but they can be computationally demanding, making them unsuitable for digital applications. The objective is to achieve predictive accuracy, and through proper sampling methods, pseudo data can be generated to train a more computationally efficient and reliable surrogate. This chapter showcased as Twin-screw wet machine learning surrogate models developed from the calibrated mechanistic model in Chapter 4 which required only 10 equations and a simulation time of 0.41 s to produce particle size distribution prediction results with accuracy of 0.998 R^2 value.

This section further showcased a developed a hybrid twin screw wet granulation model by coupling PBM and ANN, which is less accurate than other mechanistic models for TSWG prediction. ML can be used to provide an estimate kinetic rate parameter for a calibration range in the first instant before coupling with PBM. The developed modelling methodology offers a fast and efficient approach to calibration in continuous pharmaceutical manufacturing. Here ML surrogate

model for the prediction of kinetics rate is developed to rapidly estimate kinetics rate when operating conditions are changed. However, machine learning models for breakage rate in CE and KE and layering growth rate exhibit high error. Solving inverse problems is challenging due to solutions not existing or unique or being sensitive to slight changes in inputs. To reduce error, increasing the training data can reduce the error by increasing the training sample for layering from 100 to 250 and thereby increasing the R^2 value by 900%.

As a final step, I simulated the model results using the test data to inspect if the model is close to predicting CQAs and compare the simulation time. The surrogate model simulation time was less than 5 seconds. Data driven dynamic model importances is of no significance using the approach in the section. It can be replaced by combining steady state experimental data and generic/uncalibrated mechanistic models. Uncalibrated mechanistic models are used to provide trajectory and fill gaps in steady state data and therefore act as robust data driven models that can be extrapolated.

Ensuring a digital twin remains relevant (fit for purpose) requires continuous updates to the process model for two reasons: improved model performance through incorporating new data for better predictions, and adapting to process drift where real-world behaviour changes over time. Ideally, a balance of automation and human oversight is achieved. A major challenge in setting up a digital twin is the time investment required for an accurate process model. This includes both the initial development and ongoing maintenance throughout the process lifecycle. Hybrid modelling offers a potential solution to expedite initial model creation, especially for processes lacking complete first-principles knowledge. As seen in the case studies, hybrid models can accelerate development. However, this comes at a cost: a potentially weaker understanding of the underlying process and potentially reduced model robustness. Some hybrid models can be readily updated by simply retraining with new data through automation, minimising manual intervention. However, significant process changes or revisions to assumptions might necessitate manual adjustments. Finding the optimal update frequency is crucial, balancing keeping the model current with managing computational costs (training these models can be expensive, especially with large datasets). Strategies like error thresholds or time-based intervals could be used to determine update frequency, and

automation of data collection and retraining (integrating sensors with the physical process and scheduling retraining based on criteria) can further enhance efficiency.

This study highlights the significant time savings achieved through reusable model structures in hybrid modelling. A generic first principles-based model, like the population balance model used here, facilitates easy adaptation to new processes governed by similar phenomena. Notably reusing the model structure drastically reduced development time. Only minor adjustments to the ANN's hyperparameters were necessary, while the core structure remained unchanged.

Adding more process variables through additional sensors can improve model accuracy but requires a cost-benefit analysis. Collecting new data, generating a new model structure, and retraining the model are necessary, making this feasible only if the additional variables significantly enhance predictions. Furthermore, the largest training cost, evaluating parameter gradients, is unlikely to be reduced significantly as the kinetic parameter provides a good starting point. Therefore, a careful evaluation of potential accuracy gains should be balanced against the additional data collection, model development, and computational resources required.

Chapter 6: Sustainability and System Integration

6.1 Introduction

The pharmaceutical industry is undergoing a significant transformation towards smart manufacturing to enhance supply chain integration and improve overall sustainability. The highly competitive market demands a holistic view of the drug life cycle, encompassing environmental impact, energy consumption, resource utilisation, and the overall societal implications of drug production. To support sustainability optimisation and analysis within the pharmaceutical industry, the goal is to create a sustainability-oriented DT of the DiPP.

Sustainable manufacturing seeks to integrate processes to create high-quality products while minimising energy and material consumption. The goal is to reduce the negative impacts of manufacturing on society and the environment throughout the entire production lifecycle. A DT enables offline analysis and simulation of the plant, eliminating the need for costly and time-consuming physical experiments. By providing a virtual representation, the digital twin facilitates the exploration of different operating conditions, identification of potential bottlenecks, and development of strategies to improve energy efficiency, reduce waste, and minimise environmental impact.

This research aims to develop energy usage models for the DiPP to aid the pharmaceutical industry in optimising energy consumption. By integrating these models with mechanistic models, I improved process sustainability while maintaining product quality. This study presents hybrid and data-driven models of the DiPP's Consigma 25 line, capable of predicting energy usage. Through sustainability analysis and optimisation, I identified optimal EnPI values, leading to significant reductions in material and energy consumption. It also adopts a holistic approach to the drug life cycle by integrating the entire process and employing universal GSA to identify areas for improvement.

6.2 Front end

6.2.1 TSWG Energy usage dynamic modelling

A pharmaceutical company making one of the products in a continuous powder to tablet line has recently noticed that the wear rate of the motor of the granulator is beyond the accepted range. Wear is difficult to measure directly

online, but I expect the wear rate to correlate with the measured torque on the motor. The energy that is provided to the system by the motor. The torque of the TSWG's motor gradually increases with time, and follows a logarithmic curve as shown in Figure 6.1.

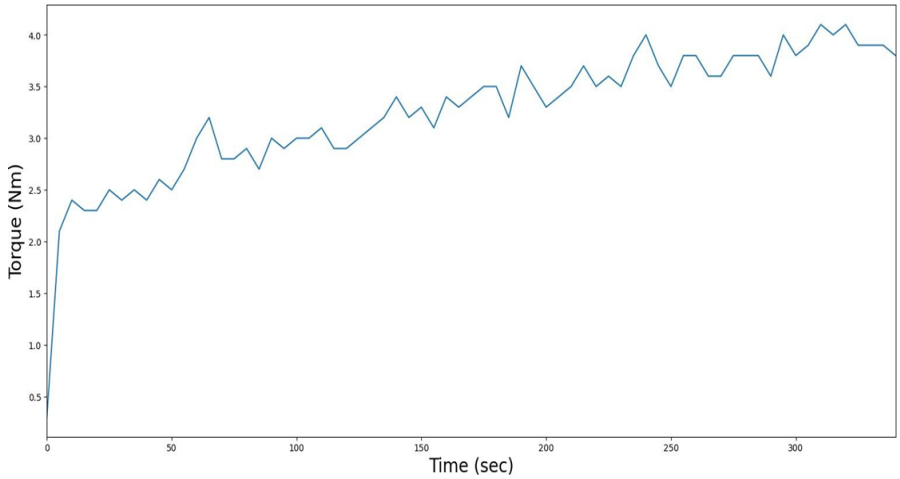


Figure 6.1 Typical of a TSWG torque profile curve for the DiPP.

The profile captures the dynamic variations in twin screw granulator operation. Torque curve was modelled as ln type function, with k being a factor. The curve depicted in Figure 6.1 is derived from experiments conducted on the TSWG in DiPP laboratory from 2018 when Insight hub was first connected. The torque model's domain is the time interval from 0 to 350 seconds. Therefore, caution is advised when using this model for extrapolation, particularly at high t values.

TSWG has three CPPs that can be set. These are:

- Screw speed (rpm)
- Powder feed rate (kg/h)
- Liquid to solid ratio

The energy modelling problem can be summarised by the following integral equation:

$$k = f(\omega, F, \lambda) \tag{6.1}$$

$$E(t) = \int_0^t p(t)dt = \int_0^t \omega \times \tau(t)dt = \int_0^t \omega \times [k \times \ln(1+t)]dt \tag{6.2}$$

The three key parameters were used to determine the k-value from the torque

curve depicted in Equation 6.1, and which was then used to calculate the energy usage(Equation 6.2) for a given time *t*. The data was inputted into python to perform various regression solvers to see which model performed better.

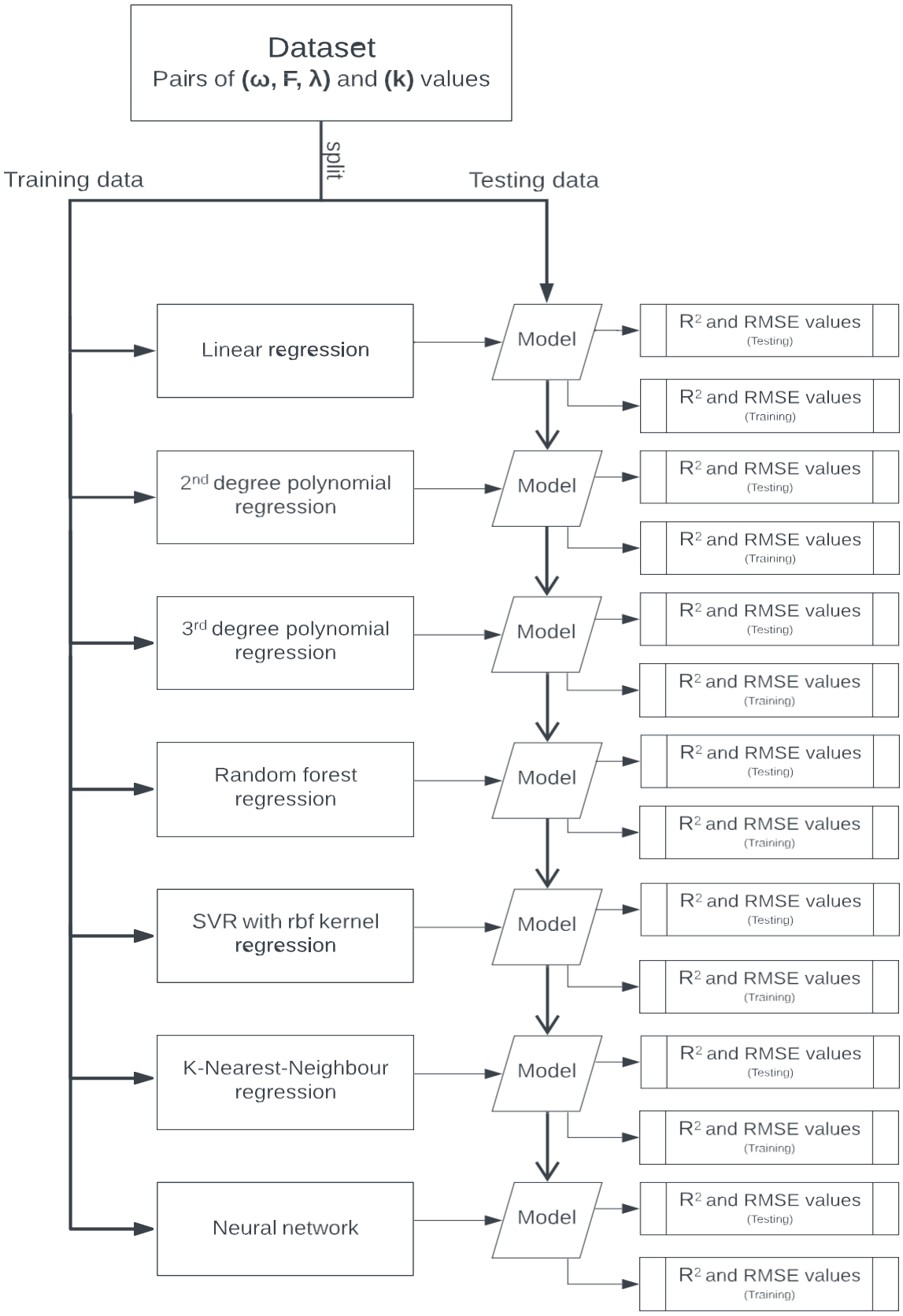


Figure 6.2 Regression training logic (Ntamo et al., 2025).

After evaluating various machine learning models, the 2nd degree polynomial regression model was determined to be the most suitable because it has a third highest training R^2 of 0.91 and highest testing R^2 values of 0.82 (Table 6.1).

Table 6.1 Regression analyses results.

Model	*Training R^2*	*Training RMSE*	*Testing R^2*	*Testing RMSE*
Linear	0.65681	0.35833	0.77118	0.32201
2nd degree polynomial	0.91168	0.18178	0.81638	0.28847
3rd degree polynomial	0.98270	0.08044	-133	7.81
Random forest	0.93364	0.15756	0.71664	0.35835
SVR with rbf kernel	-0.11	0.64455	-0.097	0.70536
K-Nearest-Neighbour	1.0	0.0	-0.29443	0.76591
Neural network	0.95256	0.13321	0.25159	0.58238

The mathematical Equation and its corresponding coefficients were exported from the Python environment, providing a clear and concise representation of the model (Table 6.2).

$$k = n_1\,\omega + n_2\,F + n_3\,\lambda + n_4\omega^2 + n_5\,\omega F + n_6\,\omega\lambda + n_7\,F^2 + n_8\,F\lambda + n_9\lambda^2 + C \qquad 6.3$$

Table 6.2 Polynomial coefficients

Coefficient	*Value*
n_1	$-1.13097020 \times 10^{-2}$
n_2	$3.27713175 \times 10^{-1}$
n_3	$3.42615181 \times 10^{+0}$
n_4	$6.17107414 \times 10^{-6}$
n_5	$-3.71204881 \times 10^{-4}$
n_6	$2.32487761 \times 10^{-2}$
n_7	$1.84521172 \times 10^{-3}$
n_8	$-1.53832945 \times 10^{-1}$
n_9	$-1.42015029 \times 10^{+1}$
C	$7.83459685 \times 10^{-1}$

6.2.2 Optimisation Testing with SciPy

To demonstrate the capability of the energy model, the optimisation of energy performance indicator (EnPI) value was performed using the model and SciPy. First, an initial guess for the SciPy minimisation solver based off prior standard DiPP values was defined.

The EnPI value being optimised was as follows:

$$EnPI = \frac{E(24h)}{F} \quad 6.4$$

Twenty-four hours was chosen as a baseline for the optimisation time as it represents a standard unit day for the industry. Here, EnPI is a crucial metric in this analysis. Minimising energy usage alone would be misleading, as the solver would simply reduce the input feed rate to the twin-screw granulator (TSG) to zero, resulting in no wet granules (and therefore tablets) being produced. This EnPI value of the ratio of energy over feed rate means that the solver will maximise the feed rate while minimising the energy. Reference input parameters are screw speed of 725 rpm powder feed rate of 7.5 kgh^{-1} and Liquid to solid rate of 0.4. These values were plugged into SciPy, and the results were as follows.

Table 6.3 Minimisation of EnPI results

Parameter	*Units*	*Before optimisation*	*After optimisation*
ω	rpm	725	724.73
F	Kg/h	7.5	8.607
λ	-	0.4	0.291
EnPI	kWh^2/kg	8.76×10^6	1.390

The results in Table 6.3 show a significant reduced in energy usage of the twin screw granulator.

6.2.3 Integration of Python and FormulatedProducts Models

The goal of this section is to be able to minimise the energy usage of the TSG while maintaining the CQAs (i.e., particle size distribution, PSD) of the wet granules that are ejected from the TSG's outlet, hence resulting in a lower overall energy usage by integrating Python with gPROMS FormulatedProducts. While full integration between the Python and gPROMS models was not achieved in this research, the gO: Python package offers a powerful pathway for future integration. This library enables seamless control of gPROMS simulations directly from Python, treating them as functions with defined inputs and outputs. Regardless of complexity or type, gO: Python supports both steady-state and dynamic simulations of any gPROMS flowsheet or custom model.

Python code can effectively interact with and leverage the capabilities of the gPROMS model for various calculations. gPROMS model was linked to IOT

driven model generated in python using surrogate modelling. An alternative route was that links Python with gPROMS was explored in next chapter. This route links python to gPROMS indirectly through Perceptive Engineering ‘s Pharma MV, an inline control platform.

The two sets of parameters (initial parameters and optimisation parameters) were manually inputted into the gPROMS flowsheet to compare the difference in PSD between the two, and the following results were produced in Figure 6.3 and Figure 6.4.

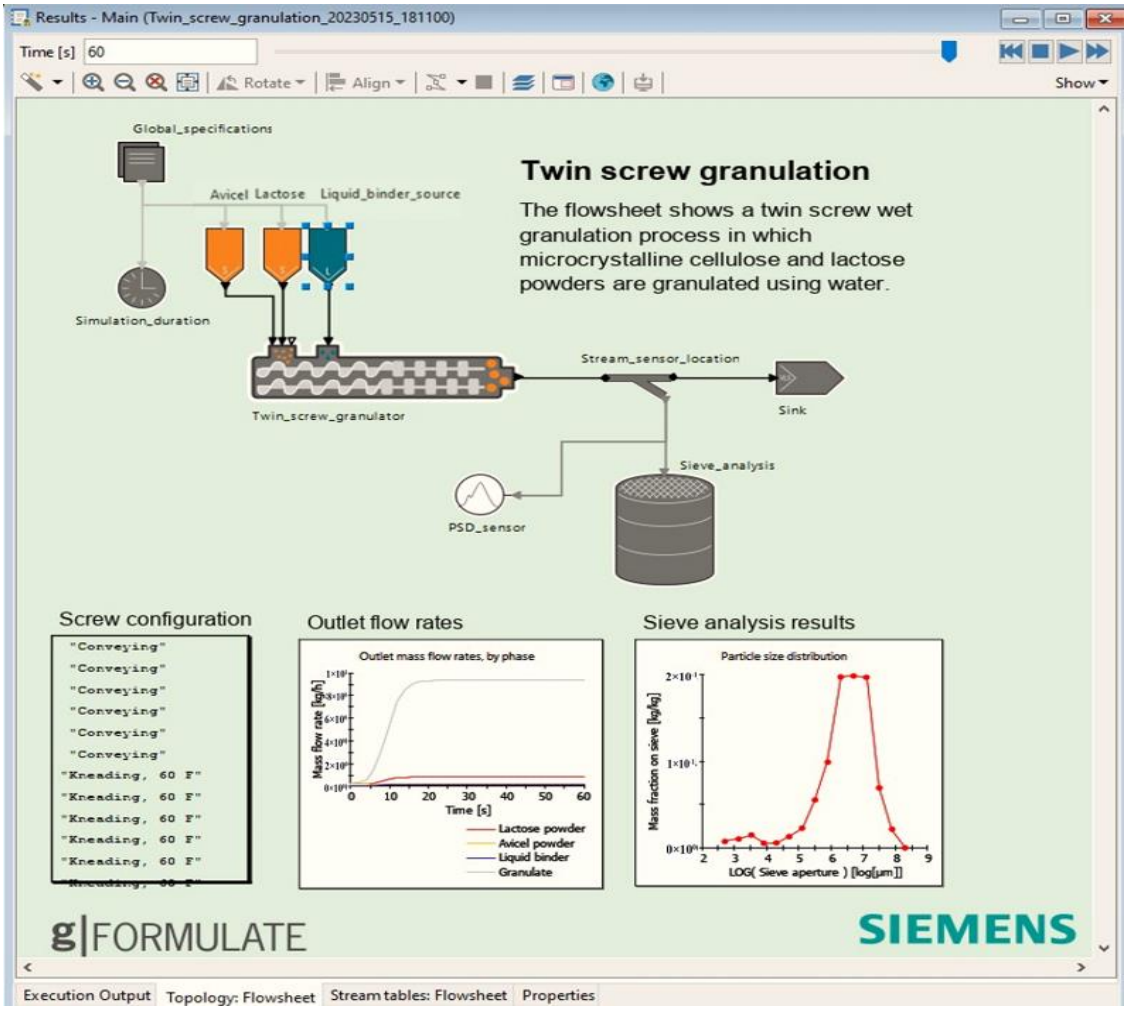


Figure 6.3 gPROMS PSD calculation of pre-optimisation parameters (Ntamo et al.,2025).

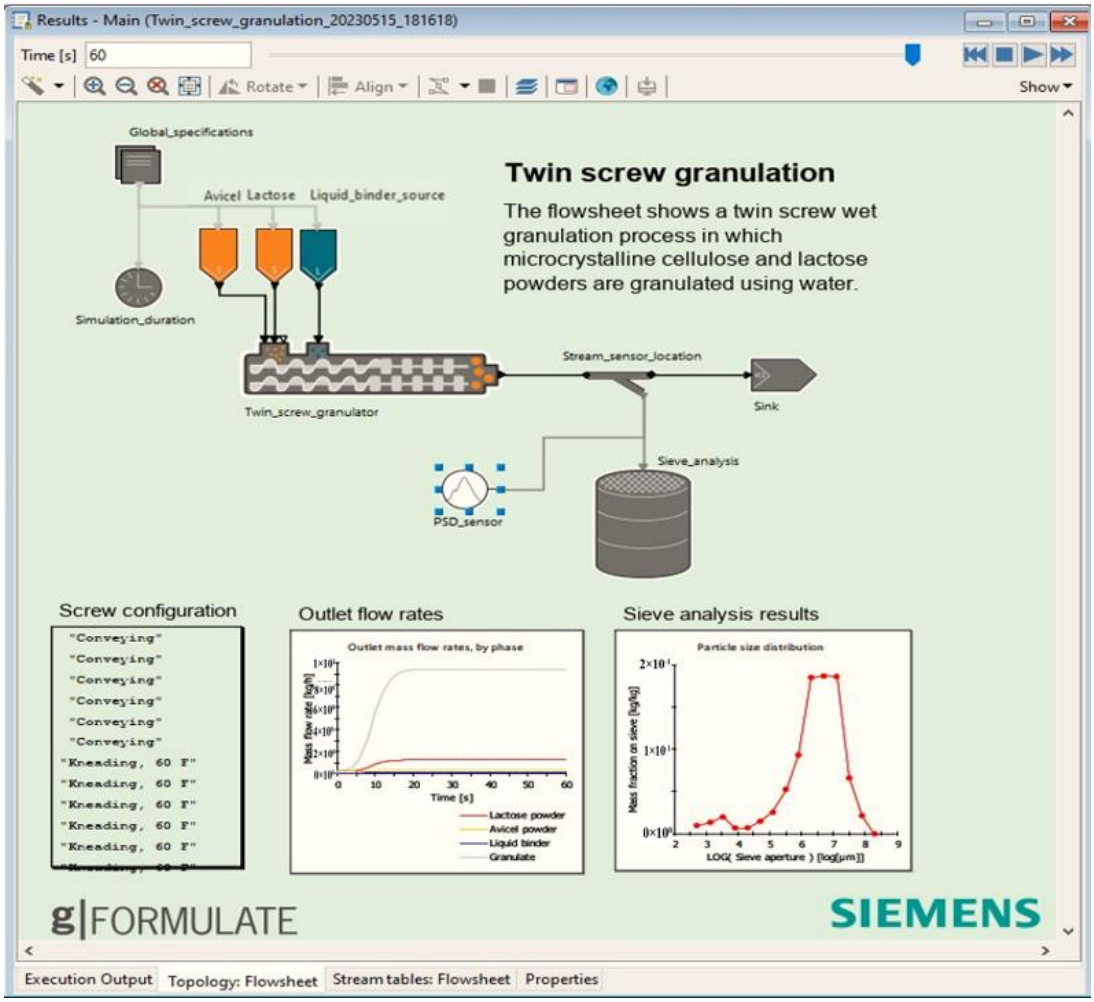


Figure 6.4 gPROMS PSD calculation of post-optimisation parameters (Ntamo et al.,2025).

As illustrated in Figure 6.3 and Figure 6.4, gPROMS FormulatedProducts enabled us to verify that minimising EnPI had a negligible impact on the particle size distribution (PSD). The d50 value shifted by only +3 µm, demonstrating the robustness of the PSD to EnPI reduction.

6.2.4 Front end's Mean-Variance Optimisation

Minimising particle size distribution is a key objective in the milling process. One approach is to optimise the liquid-to-solid ratio such that a narrower distribution is produced at the twin screw wet granulator.

The span of the resulting granules, which reflects the distribution of particle sizes, can be calculated using the following equation.

$$span = -29.097L^2 + 200.05L + 264.17 \qquad 6.5$$

$R^2 = 0.9992$

D43, can be calculated using the following equation.

$$D43 = -9.3181L^2 + 96.86L + 309.92 \quad 6.6$$

$R^2 = 1$

Where L is the liquid flow rate and powder flow rate is kept constant at 10 kg/h.

The above equations represent surrogate models derived from a GSA conducted in gFP. The liquid-to-solid ratio is calculated by dividing the liquid flow rate (L) by 10. Increasing the D43 parameter broadens the particle size distribution (PSD) of the board, while decreasing the span results in a finer granule size. To simultaneously achieve large granules with a narrow PSD, a multi-objective optimisation problem was implemented in Python. However, gFP can only minimize one variable at a time. To address this limitation, Mean-Variance Optimisation was solved using the SciPy optimisation library in Python, resulting in an optimal value of 1.5 kg/h. Simulations of the optimal process parameters in gFP demonstrated that a liquid-to-solid ratio of 0.15 minimises the span and produces a narrower particle size distribution, as illustrated in Figure 6.5. This optimisation strategy not only reduces energy consumption but also has the potential to enhance product quality.

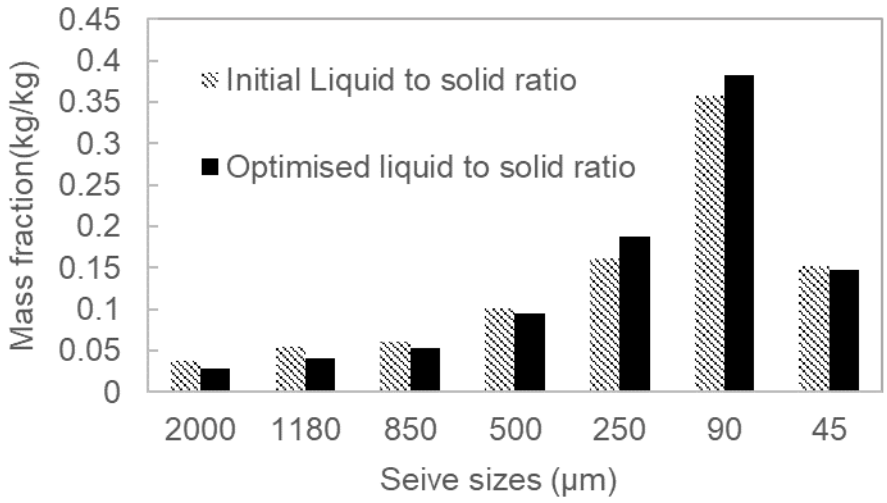


Figure 6.5 Optimised PSD distribution.

6.3 Back-end optimisation case study

6.3.1 Energy modelling

The energy consumption of a tablet press is a complex function of several variables, including the force applied to the die, the speed of the press, and the efficiency of the motor. This energy consumption can be quantified through the

power consumption equation, which relates the force F and speed of the press v to the electrical power p required to operate it.

$$E = Fvt \qquad 6.7$$

Here:

- E is the energy usage per rotation in (j/rev).
- F is the force applied by the tablet press (kN) .
- v is the speed of the tablet press (ms^{-1}).

Assuming a throughput of 1 die per rotation, the EnPI (overall energy usage per unit mass and revolution, ($kj/(mg.rev)$ for a tablet press can be expressed as:

$$E = \frac{1}{m}Fv \qquad 6.48$$

Here:m is the mass of the die in mg

It's important to note that this is a simplified formula, and the actual energy usage may involve more complex considerations such as motor efficiency, auxiliary equipment, and other factors. For a more accurate assessment, you may need detailed specifications from the tablet press manufacturer or conduct a more comprehensive energy audit.

6.3.2 Optimisation of Energy Performance Indicator for tabletting process

Optimising tablet press energy usage (appendix E) resulted in a trade-off with production efficiency. While a lower compaction force of 16 kN reduces energy consumption, it necessitates a lower bound turret speed of 20 rpm (which compromises production rate) and an upper bound target tablet mass of 700 mg (which results in high material usage).

Table 6.4 Minimisation of EnPI results.

Variable	*Optimal value*
EnPI	0.47
Compaction force (kN)	16.37
Mass of single tablet (mg)	700.00
Tensile strength (MPa)	1.60
Height of single tablet (mm)	8.37
Turret speed (rpm)	20.00

6.3.3 Dwell Time Consideration

Dwell time, another critical compaction parameter, significantly influences tablet mechanical strength. During this phase, compacts experience a constant strain state, allowing time-dependent phenomena like viscoelastic flow to occur within the tablet microstructure. The importance of dwell time for deformation has been extensively researched. Researchers have quantified the extent of plastic flow during this phase by analysing stress relaxation curves (compaction force decay) and deformation parameters from high-speed tableting force-time profiles. Prolonging dwell time often leads to substantial improvements in mechanical strength due to a favourable shift in the compact's elastic-plastic equilibrium towards a more plastic state.

The dwell phase's influence on tablets isn't isolated. It interacts with the material properties being compressed (constituents) and other compaction parameters like strain rate during consolidation. Understanding these interactions, especially in high-speed tableting, is crucial. Research on dwell time alongside formulation and process variables provides valuable insights into the compaction process under real-world manufacturing conditions.

While increasing dwell time can be beneficial, it shouldn't compromise production rate. On rotary presses, the tangential velocity (deformation rate) is determined by the punch's speed and the compression roller diameter. This velocity is related to press speed, die table circumference, and the die table's rotational speed. As this velocity increases, compression and decompression rates rise, while overall compression time decreases. For future work, consider carrying out experiments will be carried out to optimise the dell time by changing compression displacement.

6.4 System integration Fluidised bed dryers and Milling process integration

Several critical quality attributes (CQAs) significantly influence the final characteristics of pharmaceutical tablets. These CQAs of the granules, including granule size distribution, flowability, and moisture content, ultimately affect the CQAs of the final tablets, such as mass, content uniformity, strength, and dissolution profile.

This section delves into the mechanistic model of the fluidised bed dryer and milling unit into order in to integrate the whole tableting process. This model will be used to investigate how changes in various control parameters and intermediate CQAs of the granules (e.g., outlet temperature, drying time, milling speed, final granule size) impact a key final product property - tensile strength. By analysing these relationships, I aim to optimise the drying and milling processes to achieve consistent production of high-quality tablets.

6.4.1 Modelling the Fluidised Bed Drying Process

The governing mechanism for the continuous fluidised bed drying process is based on the model developed by Burgschweiger and Tsotsas (2002). This model assumes:

- Negligible particle elutriation: Particles are not entrained and carried away by the drying air.
- Accounted for mass accumulation: The model considers the accumulation of both solid material and vapor within the drying unit.
- Perfectly mixed phases: Both the particle and gas phases are assumed to be well-mixed, ensuring uniform contact and heat transfer.
- Negligible particle shrinkage: The model does not account for size changes in the particles due to drying.

The key equations that make up the segmented fluidised bed dryer are found in the literature and are written below:

For the solid phase:

Change rate of species within the unit given the assumption of prefect mixing and no reactions:

$$\frac{dMi,p}{dt} = F_p^{\,in}\, x_{i,p}^{in} - F_p^{\,out}\, x_{i,p}^{out} - R_{drying,i,p} \qquad 6.9$$

Where Mi,p is the total mass holdup of species i in phase p; $F_p^{\,in}$ and $F_p^{\,out}$ are the inlet and outlet flow rate of phase p; $x_{i,p}^{in}$ and $x_{i,p}^{out}$ are the mass fraction of phase p in the inlet and outlet stream; $R_{drying,i,p}$is the drying rate of phase p which is lost to the vapour phase.

As for the energy balance for the particle phase, it is assumed that the pressure is constant in the unit and the contents in the unit are incompressible. Likewise, the rate of energy accumulation is given by:

$$\frac{dH_p}{dt} = F_p^{in}\, h_{i,p}^{in} - F_p^{out}\, h_{i,p}^{out} - \Delta H_{drying,i,p} \qquad 6.10$$

Where H_p is the total enthalpy holdup in phase p in the unit; $h_{i,p}^{in}$ and $h_{i,p}^{out}$ are the specific enthalpy of phase p in the inlet and outlet stream; $\Delta H_{drying,i,p}$ is the enthalpy change rate resulting from drying.

Unlike the population balance equation in TSG, there are no particle evolution kernels, i.e. agglomeration and breakage defined in the fluid bed dryer. The evolution of particle size distribution is only governed by the inlet and outlet flow rate, and it can be given as:

$$\frac{\partial \eta p(r)}{\partial t} = \eta_{flow,p}{}^{in(r)} - \eta_{flow,p}{}^{out(r)} \qquad 6.11$$

Where $\eta p(r)$ is the number of particles of radius r in the unit; $\eta_{flow,p}{}^{in(r)}$ and $\eta_{flow,p}{}^{out(r)}$ are the rate number of particles of radius r entering and exiting the unit respectively.

For the vapour phase:

Mass balance for the vapour phase in the control volume:

$$\frac{dM_{i,v}}{dt} = F_v^{\text{in}}\, x_{i,v}^{\text{in}} - F_v{}^{\text{out}}\, x_{i,v}{}^{\text{out}} - R_{drying,i,v} \qquad 6.12$$

Where $M_{i,v}$ refers to the total mass of species i in the vapour phase v; F_v^{in} and $F_v{}^{\text{out}}$ are the inlet and outlet flow rate of vapour phase v; $x_{i,v}^{\text{in}}$ and $x_{i,v}{}^{\text{out}}$ are the mass fraction of vapour phase v in the inlet and outlet stream; $R_{drying,i,v}$ is the drying rate of phase p which is lost to the vapour phase.

Energy balance of the vapour phase:

$$\frac{dH_v}{dt} = F_v{}^{\text{in}}\, h_{i,v}{}^{\text{in}} - F_v{}^{\text{out}}\, h_{i,v}{}^{\text{out}} + \Delta H_{drying,i,p} - \dot{Q}_{pv} - \dot{Q}_{vw} \qquad 6.13$$

Where H_v is the total enthalpy in the vapour phase; $h_{i,v}{}^{\text{in}}$ and $h_{i,v}{}^{\text{out}}$ are the

specific enthalpy of the vapour inlet and outlet stream; $\dot{Q}$ is the heat flux and the subscript of $\dot{Q}$ indicates the interaction between the phases of particle, vapour and wall.

Energy balance at the wall:

$$c_w \frac{dT_w}{dt} = \dot{Q}_{pW} + \dot{Q}_{vW} - \dot{Q}_{We} \qquad 6.14$$

given that c_w is the specific heat capacity and T_w is the wall temperature. The subscript p, W and e of $\dot{Q}$ indicates the interaction between the phases of particle, wall and environment.

Drying rate:

$$R_{\text{drying}} = A_p \times j_i i \in C_{LV} \qquad 6.15$$

Where R_{drying} is the drying rate; A_p is the surface area of particles exposed for drying; $j_i i$ is the mass flux of species i and C_{LV} is the set of species in the vapour or liquid phases for vaporisation.

Max flux:

$$j_i = v_i \rho_g k_{c,i} \left(Y_{eq,i} - Y_{bulk,i}\right) i \in C_{LV} \qquad 6.16$$

Where v_i and ρ_g are the normalised single particle drying rate and the density of gas phase; $k_{c,i}$ is the mass transfer coefficient for species i; $Y_{eq,i}$ and $Y_{bulk,i}$ are the equilibrium and bulk dry basis moisture content of species i in the gas phase respectively.

Bulk dry basis moisture content:

$$Y_{\text{bulk},i} = \frac{x_{\text{vap},i}}{\sum_{j \in C_V - C_{LV}} x_{\text{vap},j}} \qquad 6.17$$

Where $x_{\text{vap},i}$ and $x_{\text{vap},j}$ are the mass fraction of species i and j in vapour phase; C_V is the set of species present in the vapour phase whilst C_{LV} is the set of species in the vapour and liquid phases which can vapourise.

Heat and Mass Transfer Coefficient

The mass transfer coefficient $k_{c,i}$ is calculated from the bulk Sherwood number as below:

$$Sh_{bulk,i} = \frac{k_{c,i} d_p}{D_i} i \in C_{LV} \qquad 6.18$$

Where $Sh_{bulk,i}$ is the bulk Sherwood number; d_p and D_i are the diffusion coefficient of species and the particle diameter.

To calculate the resulting heat flow rate to the population $Q_{pp}$with residence time τ, the equation is used.

$$\partial Q_{pp} \partial \tau = n(\tau) \alpha_{pw} A_{ps} \left[T_p - T_p(\tau) \right] \qquad 6.19$$

Where α_{pw} is the heat transfer coefficient between particle surface and wall; A_{ps} is surface area of the particles in the suspension phase and T_p is particle temperature. The details of heat transfer can be found in Burgschweiger and Tsotsas's paper (Burgschweiger and Tsotsas, 2002).

CQA of primary concern for the drying process is moisture content. High moisture levels in the final granules can be detrimental to patient safety. Therefore, a soft sensor for moisture content is strategically placed just before the feed frame to ensure real-time monitoring and control. While several drying parameters can influence tablet strength, such as drying air humidity, drying time, airflow rate, and temperature, this study focuses on manipulating drying air temperature due to its relative ease of control within the system.

*T*able 6.5 Fluidised bed dryer standard operating boundary.

Parameter	*Range*
Drying air temperature (°C)	70-80

Drying air humidity is very difficult to control, measure and normally variable so it will be considered a parameter to be estimated. The mass transfer coefficient in the Burgschweiger and Tsotsas (2002) model is derived from the bulk Sherwood number for externally limited regimes. The equilibrium dry basis moisture content is determined from the sorption isotherm, which relates moisture content to drying air humidity. The critical moisture content, a key parameter, needs to be determined to accurately model the drying process. Table 6.6 shows the optimum calibration value for these parameters taken from Wang et al (2022)

Table 6.6 Fluidised bed dryer Optimised parameters (Taken from Wang et al,2022).

Parameters	*Optimised value*
Critical moisture content (kg/kg)	1.013
Bulk Sherwood correction	0.3668
Drying air humidity (%)	0.1

Model validation is already caried out in Wang et al,2022 paper and therefore will not be carriable in this thesis.

6.4.2 CUSUM Chart Based Optimisation

In this section, I propose integrating optimisation methodology based on data exploration, known as Cumulative Sum control charts to the Digital twin framework of fluid bed dryer. Inline energy usage IOT data (product granule temperature and moisture content) of the dryer in Chapter 3 has been analysed using the Cumulative Sum control charts to find the areas where energy consumption can be reduced without compromising the product quality.

A CUSUM analysis conducted that maintaining a vapor temperature of 60°C during the drying process, when the granule liquid-to-solid ratio was 0.18 or 0.3 (Runs 5 and 6 in Figure 6.6), resulted in improved energy efficiency compared to other conditions. However, due to the absence of data for temperatures between 50°C and 60°C, this optimal temperature is an estimated value.

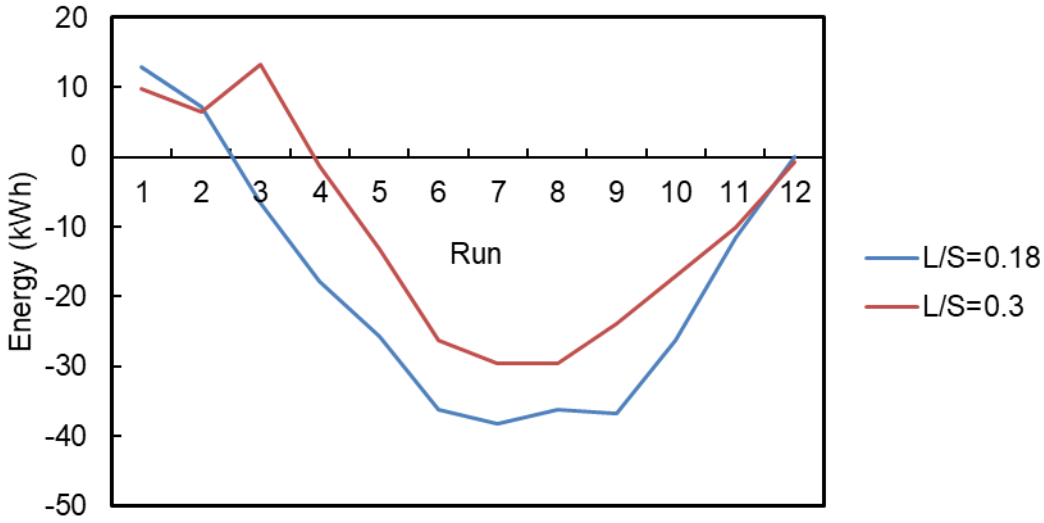


Figure 6.6 CUSUM analysis of energy consumption of the DiPP fluidised bed dryer for two L/S ratios of 0.18 and 0.3.

6.4.3 Modelling the Milling process

CQAs considered for the milling process is content uniformity/span. Non inform tablets are no safe for patient. As a precaution are a %content sensor is place just before the feed frame. Parameters that can be varied that might impact tablet strength, combined with safe limits of operation include:

Table 6.7 Milling unit standard operating condition boundaries.

Parameter	*Value*
Powder flow rate (kg/h)	5-25

6.4.3.1 Mill Aperture Size Estimation

The breakage dynamics in the milling process are represented by the continuous population balance equation.

$$\frac{\partial n(v,t)}{\partial t} = \int_{v}^{\infty} b(v,u)S(u)n(u,t)du - S(v)n(v,t) \qquad 6.20$$

where n is the number density function, S is the selection or breakage rate kernel and b is the fragment distribution.

Considering the selection rate kernel, the critical size, l_c, can be approximated as:

$$l_c = \alpha l_{sc}$$

Where l_{sc} is the size of the apertures in the mill screen and α is an adjustable parameter, descriptive of the reduction in apparent screen size due to the tangential movement of the particles in the mill

To simplify the analysis, ideal separation and negligible breakage are assumed in this thesis's milling process model therefore the calibration parameter became the aperture size. The aperture size, a crucial factor in determining the output particle size distribution (PSD), is estimated non-invasively, offering a practical alternative to dismantling the mill. Fouling and granule accumulation on the aperture will likely affect the actual flow rate compared to a clean aperture. This information, along with the initial mill design size of two meters, was used as a starting point for the calibration process. Breakage is another key parameter influencing particle size distribution. While not considered in this initial model development, it will be explored in future work for a more complete modelling of the milling process. Since I am focusing on estimating a single parameter

(aperture size), a global sensitivity analysis (GSA) was not deemed necessary in this context. However, GSA might be employed in future studies involving multiple control variables and calibration parameters.

Table 6.8 Milling unit optimised parameters.

Parameters	*Initial value*	*Optimised value*
Aperture size (mm)	2	1.496

6.4.3.2 Model Evaluation

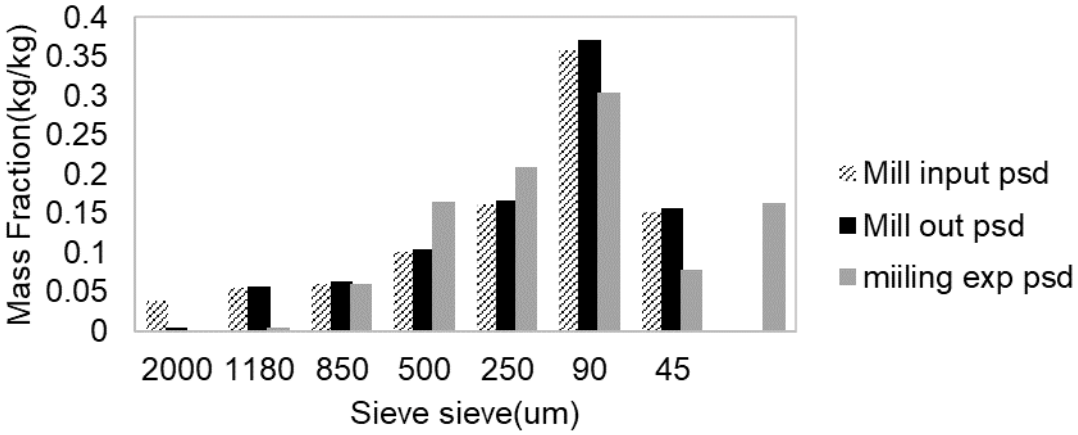


Figure 6.7 Milling Model prediction vs Experimental data.

The initial mill sieve aperture size of 2 mm was based on the manufacturer's specification. However, through calibration, an optimal value of 1.496 mm was determined (Table 6.8), which, despite being smaller, more accurately predicts experimental data output (Figure 6.7). This suggests that the 1.496 mm aperture is an apparent value, likely influenced by fouling over time.

The decision of whether to model the output using the manufacturer's specification (assuming a rigorous cleaning regime) or the apparent size (which perfectly predicts current output) involves a trade-off. While the apparent size offers immediate accuracy, it may not be sustainable long-term without consistent cleaning. Conversely, the manufacturer's specification may require more frequent cleaning to maintain accuracy but could potentially lead to more stable long-term performance.

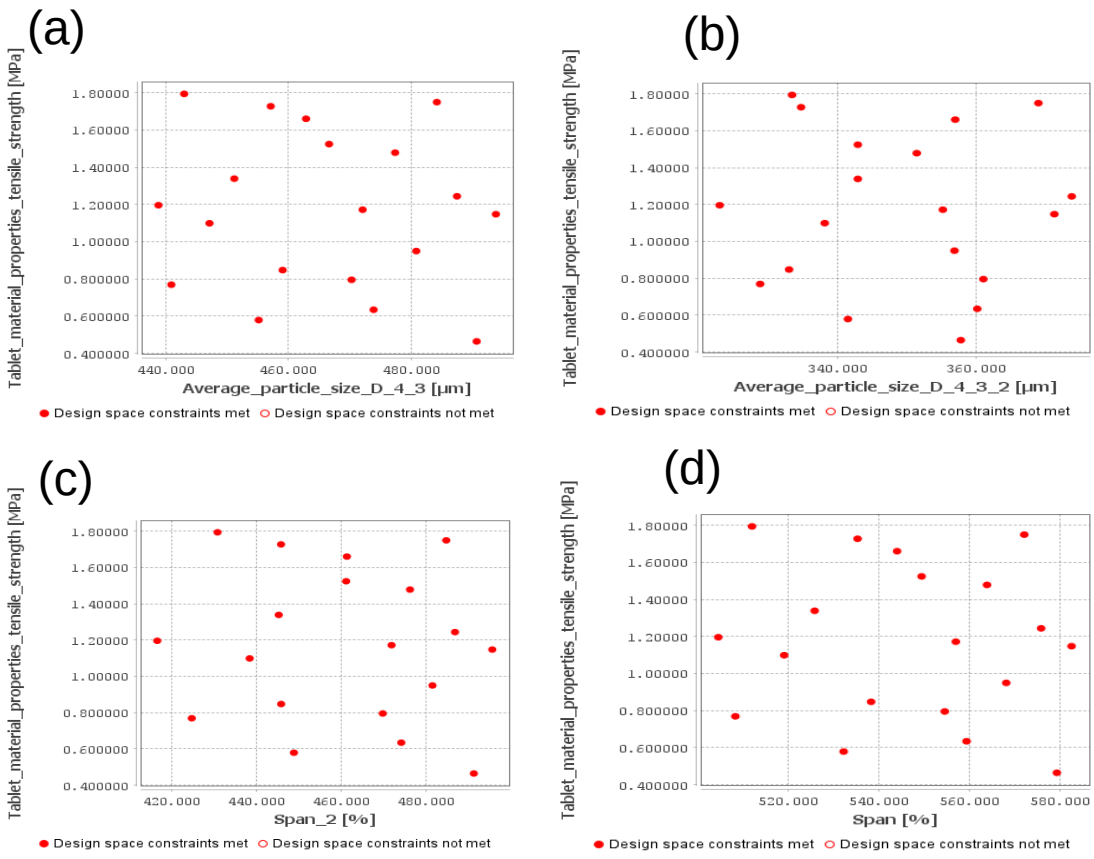
6.5 Effect of intermediate CQAs on final CQAs

Analysis from Table 6.9 and Figure 6.8 revealed that the liquid mass flow set point has minimal influence on both the final tensile strength and mass fraction of tablets. Similarly, the drying air temperature set point exhibited little to no impact on the mass fraction before compaction and the final tablet tensile strength.

Table 6.9 Effect of process intermediate and finalise tablet CQA.

	Parameter		
Total Sensitivity index	*Liquid mass flow set point (kg/s)*	*Drying air Temperature set point (°C)*	*Tablet compaction force (kN)*
D43 after granulation	1.217714		
Span after granulation	1.220287		
D43 after milling	1.214998	0.616937	
Span after milling	1.220457	0.75007	
Drying moisture content	1.214961	0.630789	
Mass fraction before compaction	0	7.37E-07	
Tablet Tensile strength	0	0	1.173837

However, it's important to acknowledge that these findings may not represent the entire picture. Further investigation is necessary to assess the impact of these parameters under different operating conditions or in conjunction with other process variables.



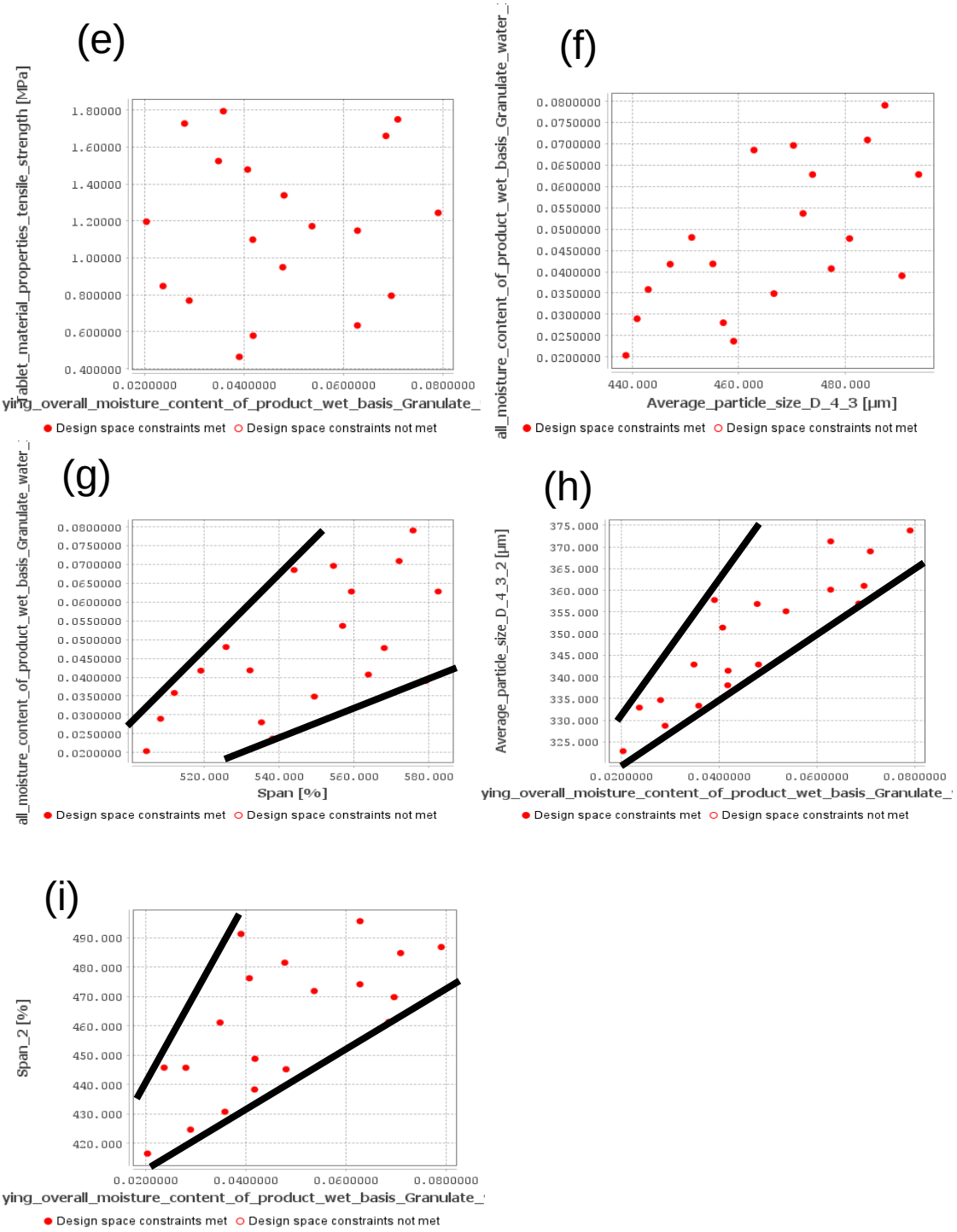


Figure 6.8 Effect of intermediate CQA on the finalise tablet CQA (tensile strength) based on calibrated DiPP model: Twin screw wet granulation output Average PSD and Milling output Average PSD(D43_2), Moisture content, Twin screw wet granulation unit span and Mill output PSD span (Span_2).

6.5.1.1 Flowability consideration

While the impact of individual backend CQAs on final tensile strength appears minimal as shown in the figures, their interactions with each other and their influence on the overall process are crucial for energy optimisation. Here's a

breakdown of how key backend CQAs can affect both flowability and energy consumption:

Larger granules often exhibit better flowability due to reduced interparticle forces and increased void space. This translates to less energy required for material handling and processing. Conversely, a wider size distribution (higher span) can lead to flow issues and inefficiencies, requiring additional energy input. High moisture content can negatively impact flowability by promoting interparticle adhesion and friction. This necessitates increased energy expenditure during material handling and blending. Maintaining optimal moisture content is essential for ensuring good flow while minimising energy consumption. Higher compaction pressures during tablet manufacturing can improve tensile strength but may also lead to reduced flowability due to increased interparticle bonding. Striking a balance between desired tablet properties and efficient powder flow is crucial for optimising energy use within the overall process. Therefore, while the direct effect of individual backend CQAs on final tensile strength might be subtle, understanding their combined influence on flowability and overall process efficiency is critical for achieving optimal energy consumption during tablet manufacturing.

To develop a specific equation, experimental data and statistical analysis may be required to identify correlations between these factors and powder flowability. The equation could take the form of a multivariable regression model, where coefficients are determined through experimental validation. Additionally, factors such as particle shape, surface roughness, and electrostatic charge may also influence powder flowability and could be included in the equation if relevant. Here's a simple example of how you can relate flowability to particle size distribution using an empirical equation:

$$HR = k_{psd} \times (D10/D90)^{\alpha} \qquad 6.21$$

Where:

HR is the flowability index of the powder.

k_{psd} is a constant that depends on the specific powder and experimental conditions.

α is an exponent that determines the sensitivity of flowability to particle size distribution.

In this equation, the ratio $D90/\ D10$ represents the breadth of the particle size distribution. A higher ratio indicates a wider distribution of particle sizes, which may affect flowability. The exponent *α* controls the degree to which the particle size distribution influences flowability. The constant *k* can be determined through experimental calibration using known powder samples with varying particle size distributions. The exponent *α* can also be determined empirically based on the specific characteristics of the powder and the manufacturing process. It's important to note that this equation is a simplified model and may not capture all the complexities of powder behaviour.

In this equation below, the term $1 - Moisture\ Content$ represents the dryness of the powder, with higher values indicating lower moisture content. The exponent *β* controls the degree to which moisture content influences flowability.

$$Flowability\ = k_{Moisture\ Content} \times (1 - Moisture\ Content)^{\beta} \qquad 6.22$$

Where:

- $k_{\text{Moisture Content}}$ is a constant that depends on the specific powder and experimental conditions.
- β is an exponent that determines the sensitivity of flowability to moisture content.

The constant *k* can be determined through experimental calibration using known powder samples with varying moisture contents. The exponent *β* can also be determined empirically based on the specific characteristics of the powder and the manufacturing process. It's important to note that this equation is a simplified model and may not capture all the complexities of powder behaviour. Additionally, other factors such as particle shape, and surface properties can also influence powder flowability and should be considered in more comprehensive models.

Developing an online sensing tool for flowability requires integrating sensors and data acquisition systems into the manufacturing process. This tool could utilise various techniques such as laser diffraction, acoustic spectroscopy, or image analysis to continuously monitor powder flow properties in real-time. By analysing the data collected from these sensors, the online sensing tool can provide valuable feedback to operators, enabling them to optimize process parameters and maintain consistent powder flowability throughout production. Overall, flowability is a critical parameter in pharmaceutical manufacturing, and

developing an online sensing tool can enhance process control and ensure product quality and consistency.

6.6 Chapter conclusion

This section demonstrates data-driven models for predicting the energy consumption of the Consigma-25 line. This model is particularly valuable to the pharmaceutical industry, as there is limited literature on energy usage in this specific context. Furthermore, the gFP models allows for the prediction of the Particle Size Distribution of the wet granules, which will be useful for future optimisation analysis of the TSWG's energy usage.

By leveraging Python's SciPy optimisation library, the research can significantly reduce the EnPI value of the DiPP units. The current gPROMS TSWG model is unable to model energy consumption. The work produced in this research will help in making the gPROMS model better for the future of net zero. Researchers can utilise these data-driven models as a starting point for comprehensive mechanistic energy models and PBM integration. This will be highly beneficial to the pharmaceutical industry, facilitating the transition to continuous sustainable manufacturing.

This section also explores the relationship between backend Critical Quality Attributes (CQAs) and their impact on energy consumption in tablet manufacturing. While individual backend CQAs may not directly influence final tensile strength their combined effects on flowability, and the overall process are largely crucial for energy optimisation. This section also explored optimising the tablet press by integrating it with the milling and drying process. While reducing the compaction force to 16 kN successfully lowered energy consumption, it did come with production efficiency trade-offs. A lower compaction force necessitates a slower turret speed (limited to 20 rpm) to maintain tablet quality. This, in turn, compromises the production rate. Additionally, a lower compaction force often requires a higher target tablet mass (around 700 mg) to ensure proper bonding, which leads to increased material usage.

Chapter 7: Designing and Implementing a Comprehensive Digital Twin Framework for Pharmaceutical Processes

7.1 Introduction

DT is a virtual asset of a physical system that serves as its counterpart in the digital world (Aheleroff et al., 2021). This architecture facilitates a comprehensive understanding of the physical system by capturing its core elements, relationships, and functionalities. The platforms must be integrated and exchanging data to be classified as digital twin as shown in Figure 7.1.

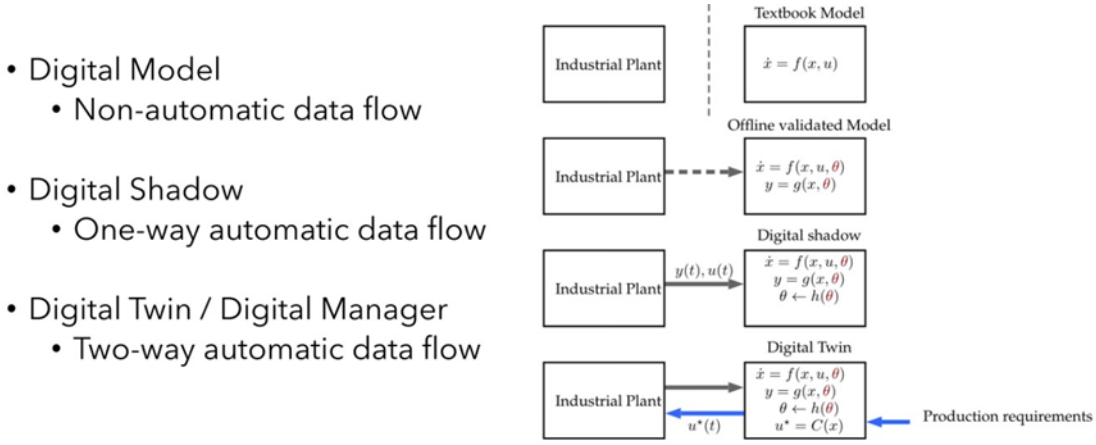


Figure 7.1 Digital Twins Classification (Yildiz et al,2020).

Mostafa et al. (2020) propose a six-layer DT architecture to describe how various components collaborate:

1. **Physical Layer**: This layer encompasses the actual physical equipment, sensors, and processes being monitored. Sensors collect real-time data and convert it into a digital format for computer processing.
2. **Intake Layer**: Raw data received from the physical layer is received and formatted in this layer to ensure compatibility with subsequent processing steps.
3. Persistence Layer: This layer stores and manages all collected data, including historical information and real-time measurements. Security

considerations are paramount in this layer because of to the sensitive nature of data.

4. **Inference Layer**: Data analysis and processing occur in this layer. Machine learning algorithms can be used to extract insights, identify trends, and generate predictions based on existing data. Fault tolerance mechanisms ensure the robustness of the system by handling potential errors during analysis. Pre-aggregation of frequently accessed data can also be implemented to optimise query performance.

5. **Service Layer**: This layer acts as the interface between the DT and users, providing structured access to data through well-defined APIs. Data can be categorised as raw measurements, metadata (information about the data), or predictions derived from machine learning models.

6. **Consumption Layer**: The final layer focuses on user interaction. Dashboards and visualisation tools present data insights in an easily understandable format. Automation can be integrated to detect anomalies and send corrective information back to the persistence layer, ultimately influencing the physical aaset through adjustments in real-time.

This six-layer architecture provides a comprehensive framework for building digital twins, enabling continuous monitoring, analysis, and optimisation of physical processes.

7.2 Assessment of Existing System Architecture

1. *Implementation of an advanced hybrid MPC–PID control system using PAT tools into a direct compaction continuous pharmaceutical tablet manufacturing pilot plant*.

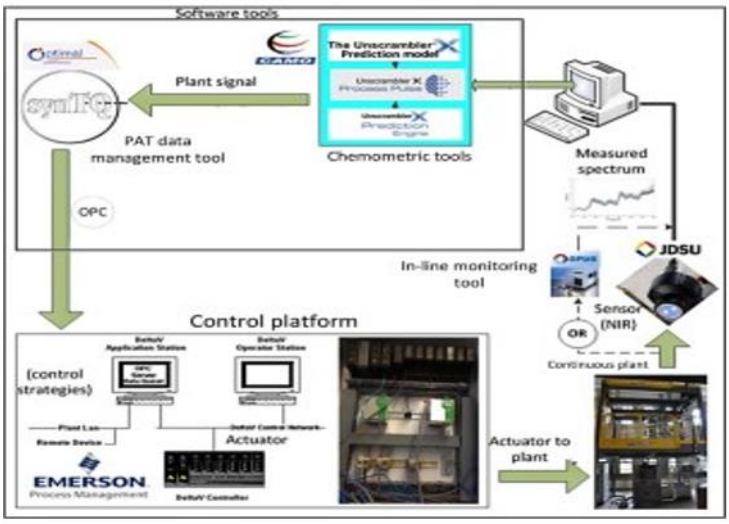


Figure 7.2 Control hardware and software integration. Taken from (Ravendra Sigh et al, 2014).

This work proposes an advanced hybrid MPC and Proportional-Integral-Derivative (PID) control architecture for a continuous direct compaction tablet manufacturing process. This approach integrates real-time inline/online monitoring tools, and a supervisory control layer based on PCA. The proposed hybrid MPC-PID scheme is expected to lead to enhanced control loop performance of critical quality attributes (CQAs) compared to traditional PID control methods.

Limitations:

Real-time implementation of MPC can require significant computational resources, potentially limiting its application in resource-constrained environments. The effectiveness of MPC heavily relies on the quality and quantity of historical data used for model training. Insufficient or inaccurate data may lead to suboptimal control performance. Developing and maintaining accurate process models for complex manufacturing processes can be challenging and computationally expensive.

Successful implementation of the proposed architecture necessitates a deep understanding of the underlying process dynamics and the selection of appropriate control variables and measurements. Rigorous validation and potential refinement of the control strategy may be necessary to ensure optimal performance across various operating conditions and potential disturbances. While the proposed hybrid control architecture holds promise for improved CQA

control in continuous tablet manufacturing, these limitations need to be carefully considered during implementation and further research.

2. *A Systematic Framework for Data Management and Integration in a Continuous Pharmaceutical Manufacturing Processing Line.*

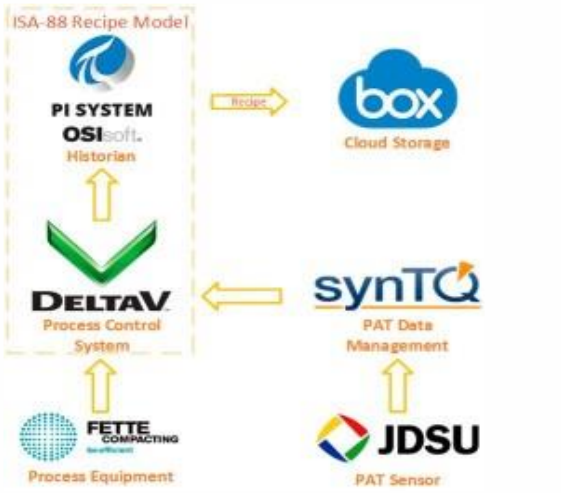


Figure 7.3 Data flow in continuous tablet manufacturing. Taken from Coo et al.,2018.

This work proposes a novel approach that integrates the recipe model with a process control system (PCS), data historian, and Electronic Laboratory Notebook (ELN) system. By centralising recipe data within these systems, it becomes readily available for export to cloud storage. This readily accessible data can serve as a reliable and consistent source for various downstream applications, including:

- **Data Visualisation**: Centralised data facilitates the creation of informative dashboards and reports for enhanced process monitoring and analysis.
- **Data Analytics**: Large datasets can be leveraged for advanced analytics techniques, leading to deeper insights into process behaviour and potential areas for improvement.
- **Process Modelling**: Recipe data can be used to develop or refine process models, enabling simulation and optimisation strategies.

Limitations:

- **Data Standardisation**: Successful integration across different systems often requires data standardisation to ensure compatibility and eliminate potential errors during data transfer.

- **Data Security**: Safeguarding sensitive recipe information within a cloud-based environment necessitates robust data security measures.
- **Data Quality**: The quality and consistency of the data exported to the cloud ultimately depend on the quality of the initial recipe data within the PCS, data historian, and ELN systems. Inaccurate or incomplete data can lead to misleading results in downstream applications.
- **User Expertise**: Utilising advanced data analytics techniques and process modelling tools may require specialised user expertise within the organisation.

3. *Advanced Process Control: Bridge the Gap.*

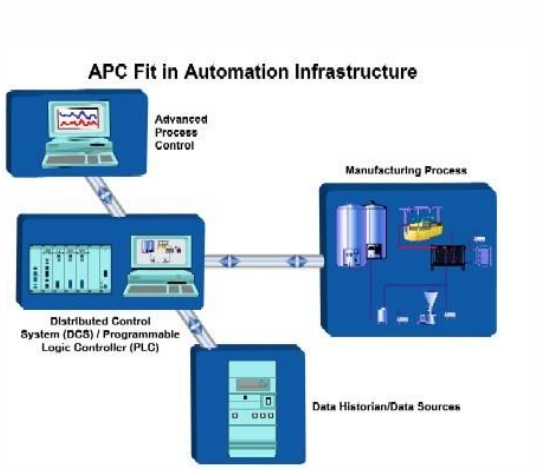


Figure 7.4 APC Fit in Automation Infrastructure. Taken from Advanced Process Control: Bridge the Gap | Pharmaceutical Processing.

This study highlights the potential advantages of implementing advanced process control (APC) in the pharmaceutical industry. The case of Fonterra Co-operative Group Ltd. demonstrates significant benefits, including a return on investment (ROI) exceeding 60%, alongside improvements in productivity (5%), product quality (50%), energy efficiency (5%), and reduced product variability. However, it's important to acknowledge limitations associated with widespread adoption of APC in pharmaceuticals, as mentioned by Radspinner and Tormollen (2022).

These limitations include:

The pharmaceutical industry operates under strict regulatory guidelines. Implementing new control strategies often requires rigorous validation and approval processes, potentially slowing down adoption. Many pharmaceutical companies rely on existing, legacy manufacturing systems that may not be readily

compatible with advanced control technologies. Upgrading infrastructure can be a significant investment. Pharmaceutical manufacturing processes can be intricate and involve complex interactions between variables. Developing accurate and robust models for APC can be challenging. Some pharmaceutical companies may have a culture that prioritizes experience and intuition over data-driven approaches. Overcoming this resistance might require training and change management initiatives. While the Fonterra case study offers a compelling example of the potential benefits of APC, these limitations suggest that broader adoption within pharmaceuticals will necessitate addressing these challenges. Further research exploring best practices for overcoming these hurdles and implementing APC successfully in this specific industry would be valuable.

7.3 DiPP Proposed Digital Architecture

The new architecture allows the synergy between the chosen platforms with no incompatible format. The new structure will be built on the already existing structure. This initiative involves the delivery of a comprehensive digital platform that integrates an Advanced Process Control (APC) system, a mechanistic model platform, and an industrial IoT platform for data analytics and visualisation. This integrated solution embodies the principles of Industry 4.0 by providing a digital twin, enabling cloud integration, and leveraging sophisticated statistical, hybrid, and mechanistic models. This digital platform will empower industries to optimise processes, improve efficiency, and make data-driven decisions.

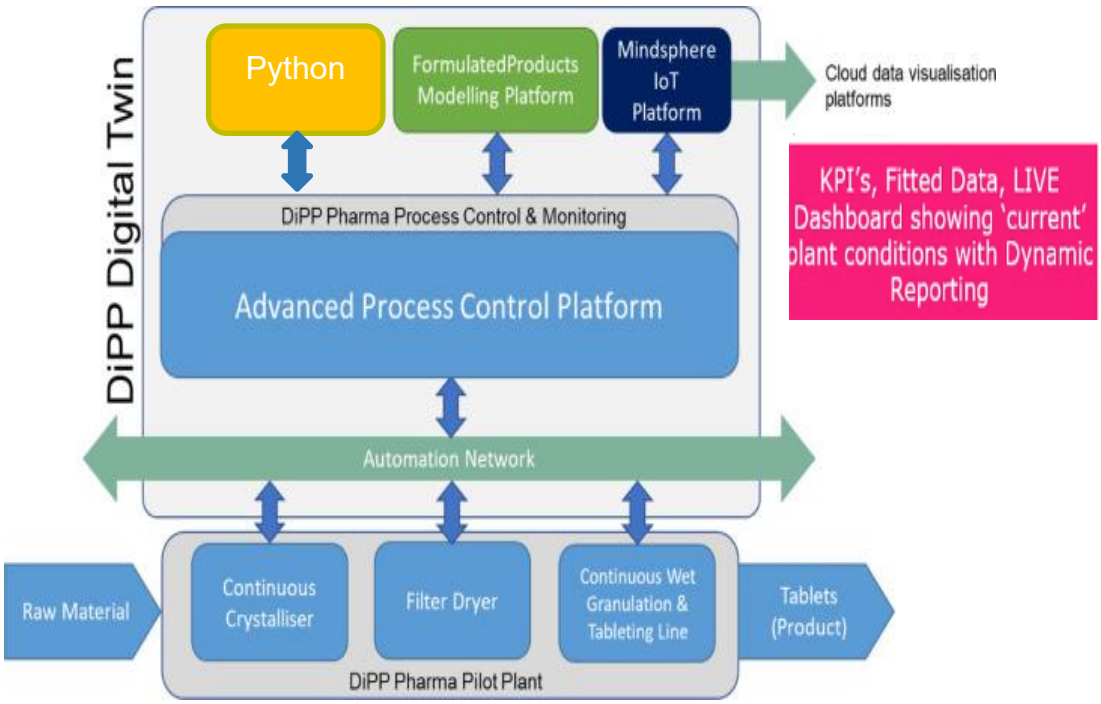


Figure 7.5 DiPP proposed digital architecture.

The proposed architecture, illustrated in Figure 7.5, establishes a comprehensive DT by integrating various platforms and facilitating seamless data exchange between them. This integrated approach enables:

Predictive Modelling and Monitoring: A connection is established between gPROMS FormulatedProduct and MindSphere, allowing for the utilisation of process models for predictive purposes and real-time monitoring of the manufacturing process.

- Digital Model Predictive Control (DMPC): The gPROMS FormulatedProduct platform connects to PharmaMV, enabling the implementation of DMPC strategies for optimising process control based on real-time and predicted data.
- Data-Driven Modelling with Deep Learning: gPROMS FormulatedProduct integrates with Python TensorFlow, facilitating the development and deployment of data-driven models using deep learning techniques. This allows for leveraging large datasets to extract insights and improve process understanding.

7.4 Enterprise Systems Integration and Cloud technology

By fostering communication between these platforms, the architecture creates a unified digital representation of the physical manufacturing process, fulfilling the core concept of a digital twin. This integrated DT offers significant advantages for optimising process performance, product quality, and overall production efficiency. To link Python and gPROMS, one can use the gPROMS Python API (Application Programming Interface). The gPROMS Python API allows one to interact with gPROMS Process in Python scripts, enabling one to perform various tasks such as running simulations, accessing model parameters, and extracting simulation results. For this case, python is already interface with PharmaMV and can be link to gPROMS via PharmaMV, linking the Multivariate Model (PharmaMV) with the process simulation capabilities provided by gPROMS Formulated Products.

One must first gain a complete understanding of the architecture and capabilities of the platform involved. This includes understanding the data formats, communication protocols, and interfaces supported by each system. Define the data exchange requirements between each platform especially, PharmaMV and gPROMS Formulated Products. This involves identifying the variables, setpoints, and process data that need to be exchanged. Check if each platform provides specific interfaces or connectors for integration. This might involve exploring options such as APIs (Application Programming Interfaces), file-based exchange, OPC (OLE for Process Control), or other communication standards. Based on the identified data exchange requirements and integration interfaces, develop scripts or functions that handle the communication the platforms. This could involve using a scripting language like Python, or any other language supported by the systems. Establish two-way communication between PharmaMV and gPROMS Formulated Products. This includes sending relevant data between platform for simulation and for analysis. Ensure that the integration meets the real-time constraints of your application. Data is transmitted to MindSphere via Edge devices for storage. MindSphere offers opportunities for real-time and historical data analysis and visualisation (Donald Ntamo, 2022). Mindsphere connect (Figure 7.6), is designed to encrypt the data before it is sent over the Internet, and it is behind a firewall, the potential surface area for a cyber-attack is significantly reduced.

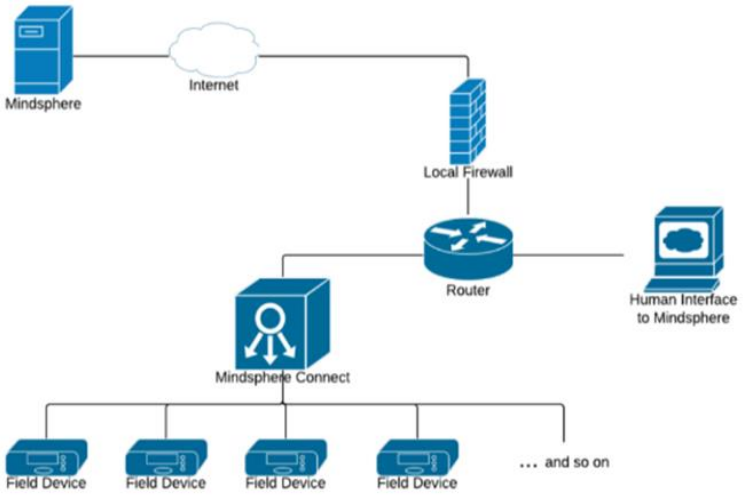


Figure 7.6 illustration of basic network topology needed to utilise real time data with MindSphere.

After developing and calibrating the mechanistic and fast solving surrogate model of the DiPP in chapter 4 and 5, the models were integrated into the PharmaMV APC system for different range of application including virtual design of experiments. gPROMS is integrated to PharmaMV by creating a configuration file that can be imported into Perceptive APC to match up signals and variables which were used to create dashboards as shown in Figure 7.7 and Figure 7.8.

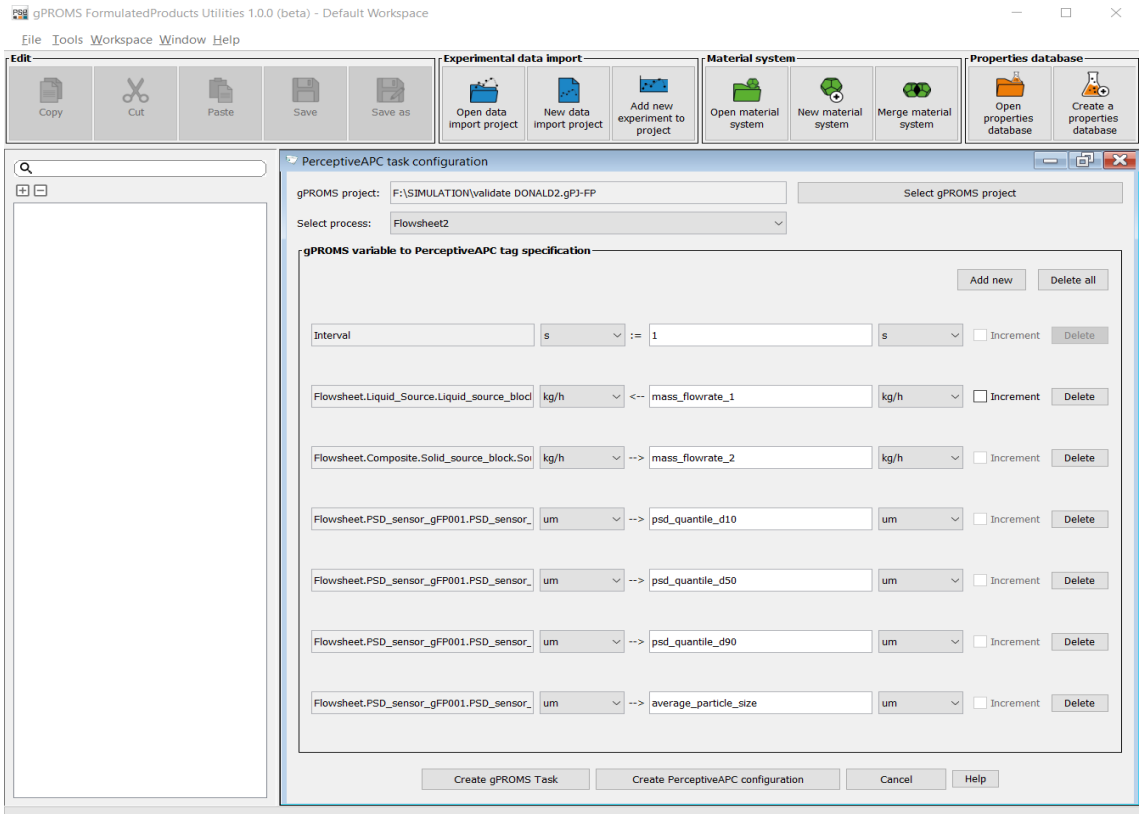


Figure 7.7 Creating a configuration file in gPROMS Utilities that can be imported into Perceptive APC to match up signals and variables.

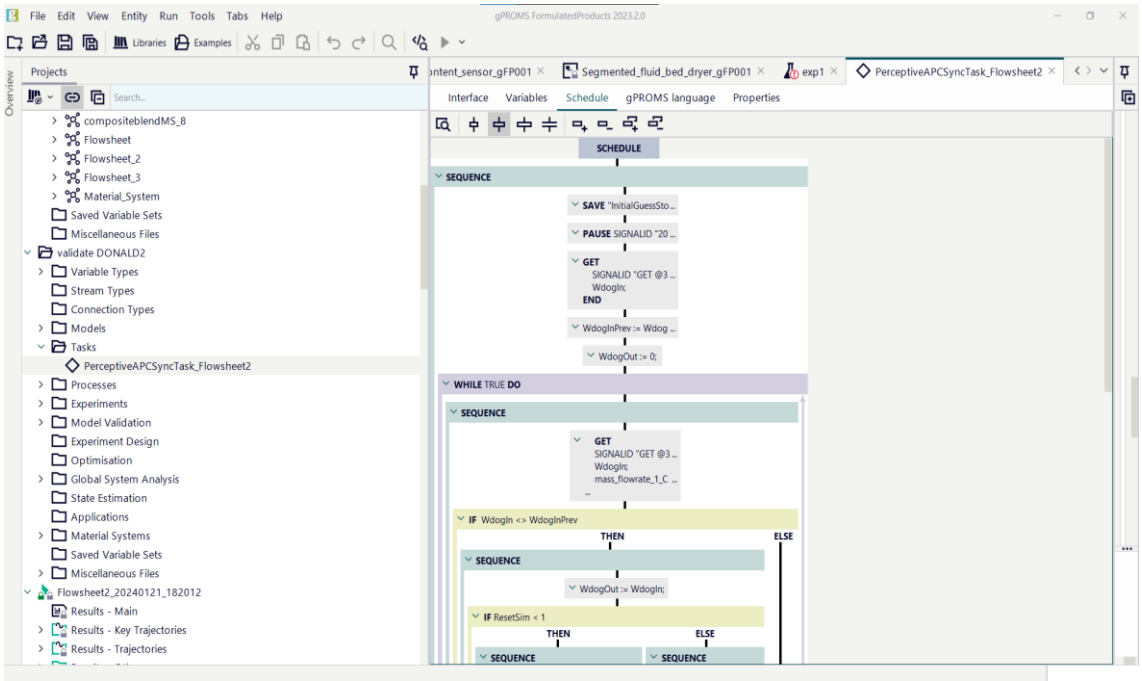


Figure 7.8 Code in gPROMS process to send and receive variables from PharmaMV to allow PharmaMV to read and overwrite values in gPROMS Formulated Products

Once the project is opened in PharmaMV and gPROMS has been integrated into the project, select Start Sampling to initiate the simulation. Select Tools → Interfaces. The gPROMS flowsheet can now be simulated. Synchronisation is complete. This can be confirmed by observing that the readback signals every 6 seconds in PharmaMV as shown in Figure 7.9. Setpoints and the corresponding response variables also change as a result.

(a)

5. The 5, 10, 16, 25, 50, 75, 84, 90 and 95% quantiles of the PSD

Property	Value at time 6.00000 s	Units
Quantile("05%")	356.027	µm
Quantile("10%")	462.569	µm
Quantile("16%")	543.992	µm
Quantile("25%")	648.067	µm
Quantile("50%")	940.232	µm
Quantile("75%")	1518.17	µm
Quantile("84%")	1948.60	µm
Quantile("90%")	2426.05	µm

(b)

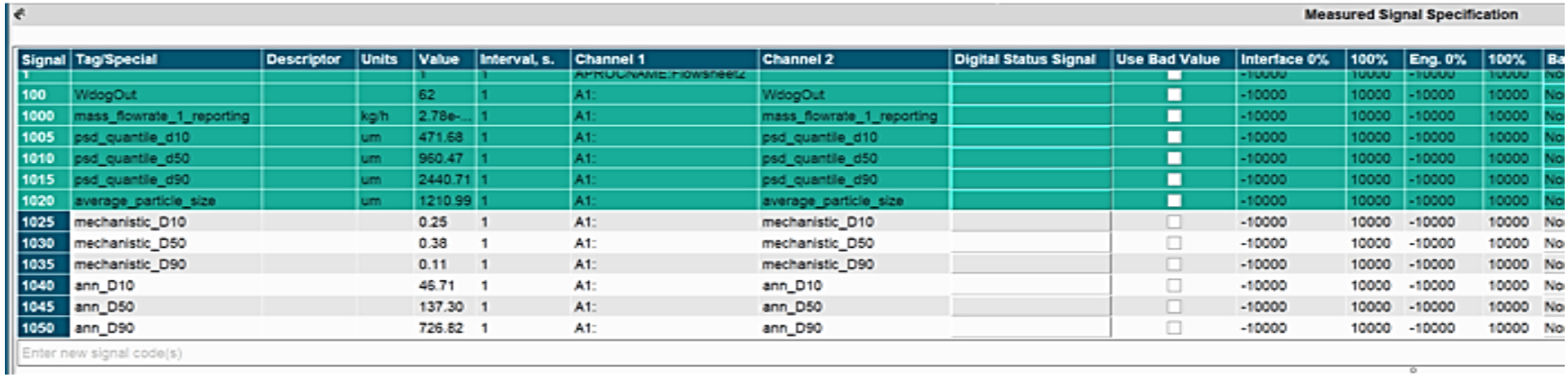


Figure 7.9 Twin screw granulator measured values in: (a) gPROMSFormulatedProducts after 6 second of running the simulation and (b) PharmaMV after 6 seconds of sampling.

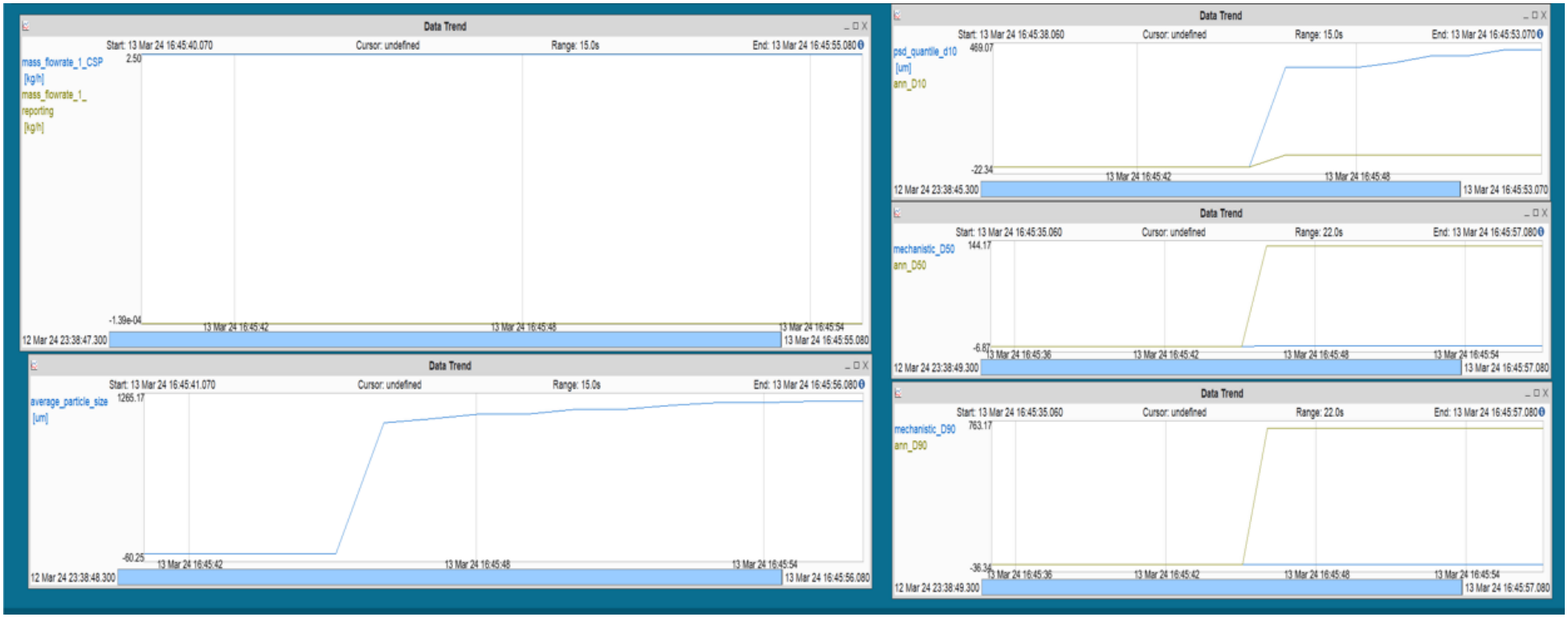


Figure 7.10 Pharma MV dashboard developed from monitoring twin screw gPROMSFormulatedProducts models' signals

This thesis demonstrated that collected data, also referred to as historical data, can be exploited by statistical and machine learning techniques to obtain additional information that can be used to make decisions towards more efficient and sustainable process operation in the pharmaceutical industry where quality is the main concerns in chapter 6. These models can now be used to provide an inline prediction of energy usage using the workflow shown in Figure 7.11. This was done by connecting the online process parameters from Perceptive Engineering's Pharma MV with the twin screw granulator energy model in python.

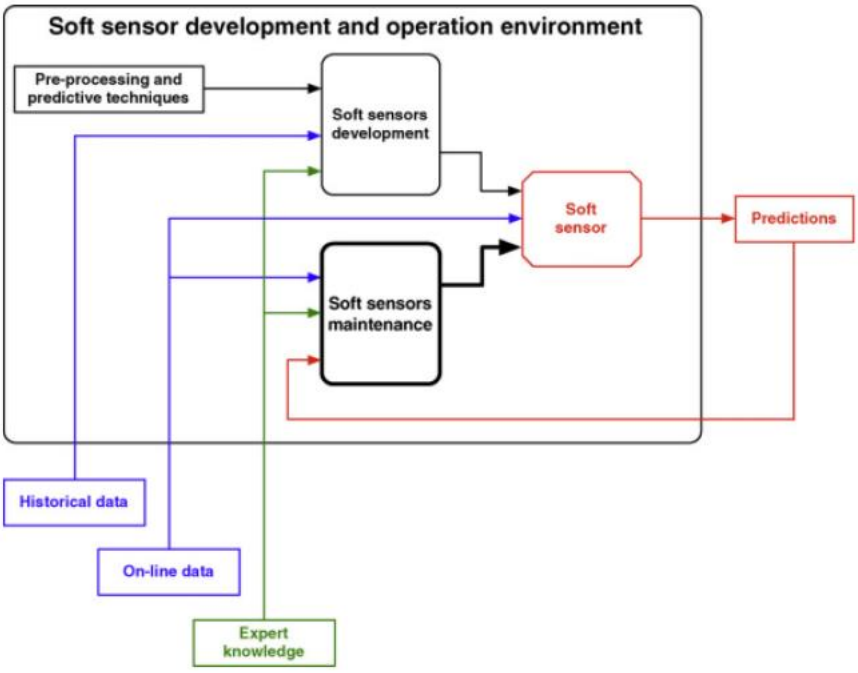


Figure 7.11 Soft sensor deployment in Perceptive Engineering's Pharma MV.

This was done by setting input to the model to actuator signals and energy usage out the measured signals in the inline monitoring tool, ie Perceptive Engineering's Pharma MV as shown by Figure 7.12.

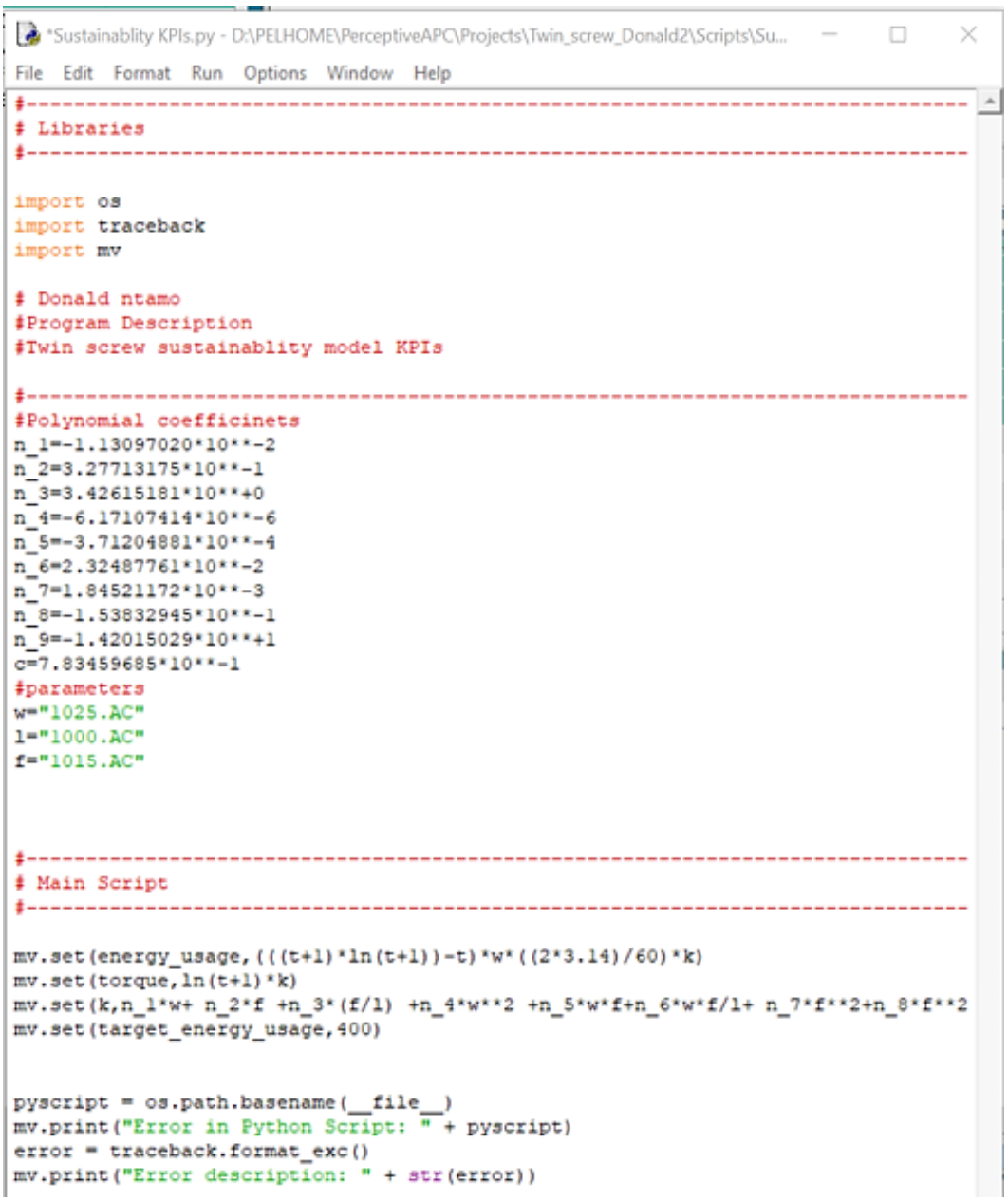


Figure 7.12 New Python energy model integrated into DiPP's in line monitoring and control tool, Perceptive Engineering's PharmaMV signals via mv.set function.

In conclusion, to facilitate advanced real time process control, a sophisticated hybrid modelling architecture was implemented. This involved the deployment of a gPROMS first-principles mechanistic model, computationally efficient deep learning surrogate models, and an energy consumption model onto the plant system. The integration was orchestrated using Perceptive Engineering's Pharma MV software suite, specifically leveraging its mv.set function within a custom Python script, as visually depicted in Figure 7.12. This deployment enabled the seamless transfer of model outputs as manipulated variables, directly influencing the physical process and achieving closed-loop control.

7.5 Chapter conclusion

This chapter has explored the development and implementation of a novel digital architecture platform (DiPP) for pharmaceutical manufacturing. DiPP offers a flexible, customisable, and scalable solution, allowing seamless data transmission across individual process units and entire production lines. This paves the way for a truly integrated and data-driven manufacturing environment.

The chapter showcased the power of cloud technology in data analysis. By leveraging the cloud, DiPP facilitates real-time data collection, processing, and analysis. This empowers manufacturers to gain deeper insights into their processes, identify process trends, and optimize production efficiency. Furthermore, this chapter demonstrated the successful deployment of an interactive mechanistic model dashboard within Perceptive Engineering's Pharma MV, a leading line monitoring tool. This innovative integration provides operators with a real-time view of process simulations alongside actual production data. This enables proactive decision-making and facilitates process adjustments based on predicted outcomes. DiPP's capabilities extend further by incorporating data-driven models identified through machine learning algorithms. These models offer highly accurate predictions that can be used to automate complex process variables. This level of automation not only reduces human error but also optimizes process performance by maintaining critical parameters within desired ranges. The chapter also highlighted DiPP's compatibility with existing industry tools. The seamless integration with Perceptive Engineering's Perceptive Engineering's PharmaMV, gPROMS mechanistic models, and Python energy models exemplifies DiPP's versatility and ability to enhance existing workflows. This interoperability fosters a truly connected manufacturing ecosystem.

In conclusion, this chapter has presented a compelling vision for the future of pharmaceutical manufacturing. By embracing a digital architecture platform like DiPP, manufacturers can unlock a new era of process optimisation, enhanced efficiency, and data-driven decision-making. This innovative approach paves the way for a more agile, responsive, and cost-effective production environment, ultimately leading to the consistent production of high-quality pharmaceuticals.

Chapter 8: Fast-Track Deployment: A Workflow for Integrating Models into Digital Operations

8.1 Introduction

When dealing with digital twins, several design and development tools are employed, as demonstrated. It is critical to establish standards across projects so that the process may be copied and applied to a different physical asset.

8.2 Standard workflow

The standard validation cycle involves an iterative procedure as shown in Figure 8.1 below. Once the designed experiments have been performed and the chosen measurements have been collected and imported, they can be used to generate better estimates of the parameters. The accuracy of the model and the quality of the estimated parameters can be quantified through cross-validation and uncertainty propagation analysis. If necessary, the model with the updated parameters can then be used to design further experiments and the cycle is repeated until the desired level of accuracy is attained.

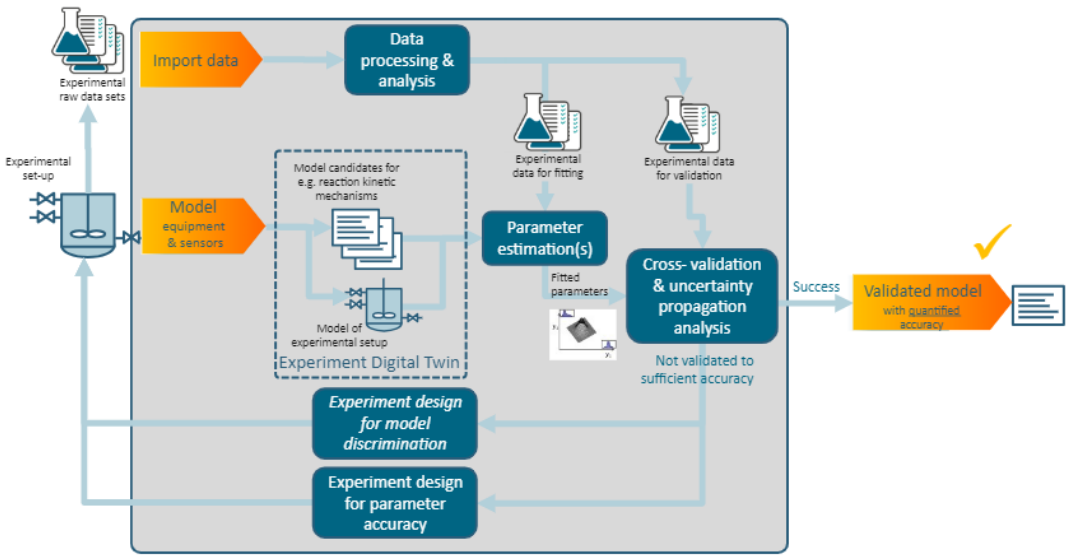


Figure 8.1 Iterative model development and validation procedure (Siemens Industry Limited,2024).

8.3 Limitations of the Standard Workflow

The current workflow for model development and parameter estimation has some limitations:

- **Time-consuming:** Iterative experimentation and analysis can be a slow process, especially when dealing with complex models and large datasets.
- **Human Bias:** Selection of experimental designs and interpretation of results can be influenced by human bias.
- **Limited Exploration:** Traditional methods might not explore the full design space efficiently, potentially missing optimal solutions.

Here's how Machine Learning (ML) can be used to address these limitations:

- **Faster Experiment Design:** ML algorithms can analyse historical data to suggest informative experiments for efficient parameter estimation.
- **Reduced Bias:** Automated analysis through ML helps to minimise human bias in interpreting results.
- **Exploration Optimisation:** ML techniques like active learning can prioritise the most informative experiments, accelerating the process of finding optimal parameters.

8.4 Online Model Deployment Vs Research and Development (R&D)

The traditional workflow for model development and parameter estimation, while valuable, has limitations. Iterative experimentation can be time-consuming and susceptible to human bias. Machine Learning offers significant advantages by suggesting informative experiments, reducing bias through automated analysis, and optimising exploration strategies. The standard workflow primarily focuses on the research and development phase. For the model to be truly impactful, it needs to be deployed online for real-world use. Online deployment considerations include ensuring robust performance, handling data drift, and maintaining model security.

The standard workflow described above focuses primarily on the R&D phase of model building. However, for the model to be truly valuable, it needs to be deployed online for real-world use:

- **Online Deployment:** This involves integrating the developed model into a production environment, where it can be utilised for real-time predictions, control strategies, or other desired applications.

- **Considerations:** Challenges of online deployment include ensuring robust performance, handling data drift (changes in data patterns over time) and maintaining model security.

8.5 Integrated Workflow: AFRAME

A system engineering approach can significantly accelerate the deployment of models for digital applications. The process begins with a comprehensive engineering management plan. This involves identifying stakeholders, understanding their high-level requirements and expectations, and clearly defining the project's background, motivation, and scope. Next, system-level requirements are established. This defines how well the system should perform in terms of functionality, interfaces, environment, reliability, and safety. Historical data and documentation are analysed to inform these requirements. A well-organised data storage and retrieval system can significantly expedite this step. The concept of operations (CONOPS) stage identifies industry best practices and selects the most suitable ideas to build the digital application architecture. This selection considers the project's functional outputs and target platforms. Data-driven subsystem requirements are then defined, involving data acquisition, cleansing, and management protocol design. Component detail design focuses on building fast-to-deploy, adaptive, and cost-effective models that solve problems efficiently. Verification and validation ensure the models meet requirements and deliver the intended functionality. Finally, the system is commissioned for operation, with clear maintenance guidelines provided.

Here are the standard steps for utilising engineering models for online deployment explored in this thesis:

Understanding the Process: Before implementing any monitoring system, it's crucial to have a deep understanding of the manufacturing process. This includes understanding the various variables involved, their interactions, and how they affect the final product.

Developing a Complex Fundamental Model: Developing a mathematical model that describes the fundamental physics, chemistry, or mechanics of the manufacturing process. This model should capture the interaction between different variables and how they influence the process outcomes. Depending on

the sophistication of the process, this model could range from simple empirical equations to complex CFD simulations.

Data Collection: Collecting data from sensors installed throughout the manufacturing process. These sensors should capture relevant variables such as temperature, pressure, flow rates, chemical concentrations, etc. Ensure that the data collected is of high quality and sufficient for model calibration and validation.

Calibration and Validation: Calibrating the complex fundamental model using historical data to ensure that it accurately represents the behaviour of the manufacturing process. Validate the model by comparing its predictions with real-world observations under various operating conditions.

Online Monitoring System: Implementing an online monitoring system that continuously collects real-time data from sensors and feeds it into the complex fundamental model. The model should then analyse this data to monitor the current state of the process and predict future behaviour.

Anomaly Detection: Using the complex fundamental model to detect anomalies or deviations from expected behaviour in real-time. This could include detecting sudden changes in process variables, drifts from normal operating conditions, or the onset of potential faults or failures.

Process Optimisation and Control: Using the insights gained from the complex fundamental model to optimise the manufacturing process in real-time. This could involve adjusting process parameters to improve product quality, increase yield, reduce energy consumption, or minimise waste generation.

Exploration and Scenario Analysis: Using the complex fundamental model for scenario analysis and exploration of alternative operating conditions. This could involve simulating the effects of changes in process parameters, equipment configurations, or raw material properties to identify opportunities for improvement or risk mitigation.

Integration with Control Systems: Integrating the online monitoring and exploration system with the manufacturing plant's control systems to enable closed-loop control. This allows for automatic adjustment of process parameters based on real-time feedback from the complex fundamental model, leading to adaptive and self-optimising manufacturing processes.

Continuous Improvement: Continuously updating and refining the complex fundamental model based on new data and insights gained from ongoing monitoring and exploration. This iterative process ensures that the model remains accurate and effective in guiding decision-making and process optimisation efforts.

Identifying the slowest steps in the process outlined for using complex fundamental models for online monitoring and exploration in manufacturing can help in prioritising efforts for optimisation. While each step plays a crucial role in the successful implementation of complex fundamental models for online monitoring and exploration in manufacturing, focusing on optimising the slowest steps can help streamline the overall process and accelerate the realisation of benefits. Strategies such as parallelising tasks, leveraging automation and advanced modelling techniques, and allocating sufficient resources to critical activities can help mitigate bottlenecks and improve efficiency.

The following are the slowest step identified in the standard workflow:

- Model Development
- Calibration and Validation
- Integration with Control Systems
- Continuous Improvement

The summary provides solution/ techniques explored in this thesis to improve and accelerate the steps above:

- GSA may be used to assist and supplement model calibration by giving information on how changes in input parameters affect the model's fit to experimental data. The sensitivity analysis findings were used to rank and filter input parameters, as well as to identify relevant sections of the input variability domain.
- To further minimise the number of experiments to be carried out, repetition was carried out only at the centre points. The linear variance model was estimated and applied at other runs.
- Mechanistic models can be used to create surrogate models that execute rapidly and robustly over a domain. First-principle-model flowsheets produce high-fidelity results, but they can be computationally demanding,

making them unsuitable for digital applications. The objective is to achieve predictive accuracy, and through proper sampling methods, pseudo data can be generated to train a more computationally efficient and reliable surrogate.

- This thesis explored the setting up of models that can directly link formulation properties, such as the L/S ratio, with key process parameters like droplet size and standard deviation. This approach aims to reduce the number of calibration parameters needed for accurate model predictions.
- This study explored the potential of artificial neural networks (ANNs) as a complementary tool. PBMs excel at tracking particle property changes throughout the process, predicting key outputs like PSD. The developed ANN, however, demonstrates promise for rapid and reliable prediction, even if its accuracy is currently lower than established PBM methods. This advantage in speed and reliability makes ANNs suitable for implementing model predictive control (MPC) strategies. Furthermore, the ANN can be coupled PBM models to speed up the calibration procedure during continuous pharmaceutical manufacturing development.
- This thesis also demonstrated DiPP's in-line monitoring and control tool, Perceptive Engineering's PharmaMV and gPROMS mechanistic model synchronisation and signal readback and Python energy model integration.

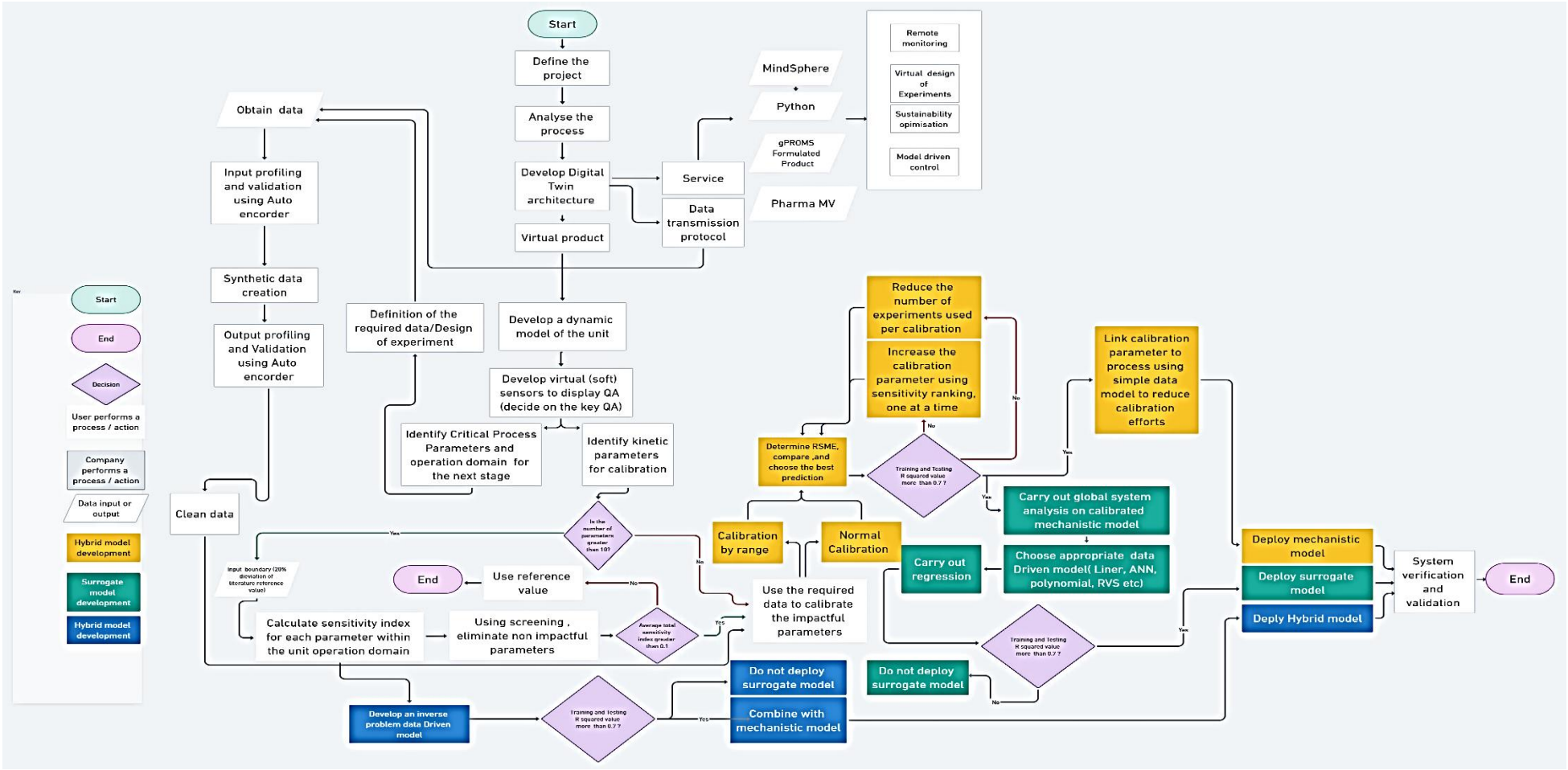


*Figure 8.2 Proposed Accelerated Framework for the deployment of advanced models for digital operation (**AFRAME**).*

8.6 Chapter Conclusion

Traditional process modelling methods often involve significant time and resource investment. Additionally, most existing research has focused on offline model development, utilising static models trained on predetermined datasets. These methods struggle in online environments due to limitations in computational power and the ability to handle continuously expanding datasets. To address this challenge, a novel workflow was developed, incorporating three key elements:

- Surrogate modelling: This technique utilises computationally efficient models to simulate the behaviour of complex physical systems.
- Hybrid modelling: This approach combines physics-based models with data-driven models to leverage the attributes of both.
- Global sensitivity analysis (GSA): This technique identifies the most influential factors affecting the process, allowing for a focus on the most critical parameters.

By integrating these methods, the new workflow offers significant advantages:

Reduced development time: Surrogate modelling and data-driven approaches often require less time and resources compared to developing full-fledged physical models.

Improved model adaptability: The inclusion of data-driven elements allows the model to adapt to changes in the process or operating conditions.

Faster model deployment: Utilising surrogate models and GSA enables faster deployment of operational models for real-time optimisation.

Scalability and continuous learning: The new workflow leverages models capable of utilising continuously expanding datasets, facilitating online learning and model updates.

To effectively implement such a framework within a digital twin environment, the model and its development process must meet specific criteria:

Online learning capabilities: The model should effectively learn and adapt from a continuously growing dataset collected in real-time.

- Rapid development and minimal manual effort: The framework should enable rapid model development and updates with minimal manual intervention.
- Computational efficiency: Both training and evaluation processes should be computationally efficient for real-time application.
- Process versatility: The framework should be agile and adaptive to a wide range of manufacturing processes.

By incorporating these elements, the proposed workflow, AFRAME, paves the way for developing versatile and adaptive modelling frameworks suitable for creating robust digital twins in various manufacturing applications. This approach can significantly improve process optimisation, product quality, and overall production efficiency. This innovative workflow was then applied to the tablet press process at the Diamond plant. Traditionally, developing such a model could take a significant amount of time. However, by leveraging the new approach, the tablet press model deployment was significantly accelerated, with completion achieved in just one month. This demonstrates the efficiency and effectiveness of the proposed workflow in bringing process models to fruition.

Chapter 9: Future Work

This research has the potential to significantly accelerate the adoption of digital twins in the chemical and biomanufacturing sectors. By enabling faster and more cost-effective development, advanced models can pave the way for a future of:

- Sustainable manufacturing: Digital twins can optimise processes for reduced waste and energy consumption.
- Enhanced product customisation: Manufacturers can leverage digital twins to cater to specialised markets without compromising sustainability.
- Improved process control: Real-time insights from digital twins can lead to more consistent and efficient biomanufacturing operations.

9.1.1 Model Driven Digital Control

This work will allow the digital twin to take advantage of the Digital Model driven digital Control using MPC shown in Figure 9.1.

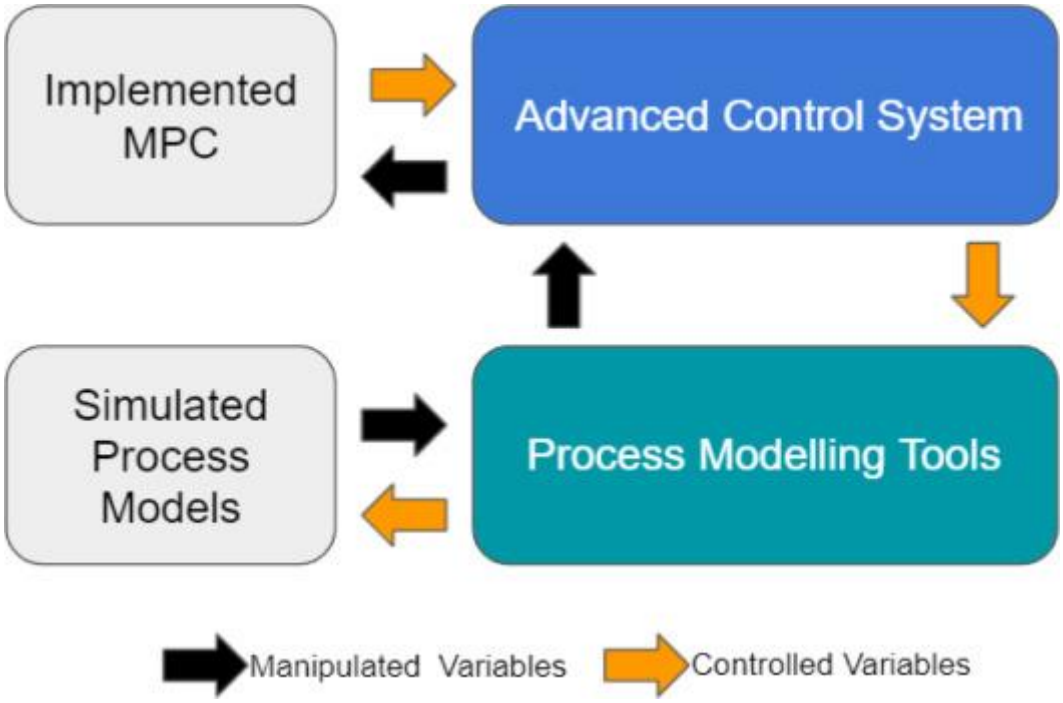


Figure 9.1 Integration of the flowsheet model and control platform. The synchronisation of the two platforms will allow data flow (manipulated and control variable data) between flowsheet model and control platform for digital model prediction control (DMPC)(Ntamo et al.,2022.)

A hybrid strategy, leveraging the integration of PharmaMV, gPROMS FormulatedProducts, and MindSphere, offers a rapid and cost-effective in silico approach to closed-loop controller development. This digital design-based methodology contrasts with traditional data-driven approaches, which necessitate step response testing and a fully operational plant. By utilising high-fidelity mechanistic models within gPROMS FormulatedProducts, the Diamond Pilot Plant's unit operations can be simulated, significantly reducing experimental effort during Advanced Process Control (APC) strategy development.

PharmaMV enables the transmission of controller output (actuator signals) to the process model, providing input for calculating new control variable signals. This Digital Model Predictive Control (MPC) framework facilitates controller parameter tuning, model identification for MPC implementation, and evaluation of data-driven control systems. Notably, this approach minimizes material waste by allowing control strategy design independent of Process Analytical Technology (PAT) data, reduces operational downtime, and lowers experimental costs (Reynolds, 2019; Singh et al., 2014). This work will also involve the addition of control elements like push buttons linked to control instructions/algorithms of the virtual model, Input, output/display elements for data. These graphic user Interfaces (GUI) elements control the virtual simulation and online connection to the physical system.

9.1.2 Virtual Design of Experiments

Virtual design of experiments (VDOE) represents a cutting-edge industrial application where mechanistic models are integrated into online deployment systems, revolutionising the way processes are developed and optimised. By leveraging detailed mechanistic models that simulate the fundamental physical and chemical phenomena of a process, VDOE allows for comprehensive virtual experimentation without the need for extensive physical trials. This integration facilitates real-time process optimisation and control, enabling industries to predict outcomes, identify optimal operating conditions, and swiftly adapt to changes. The online deployment of these models ensures continuous monitoring and adjustment, enhancing efficiency, reducing costs, and minimising downtime. This approach not only accelerates the development cycle but also enhances the robustness and reliability of industrial processes, driving innovation and competitiveness in various sectors.

DoEs are set up as a Block in the Development system under Modelling facilities in Perceptive Engineering's PharmaMV. The DoE Manager is used to configure the DoE after the Block has been defined.

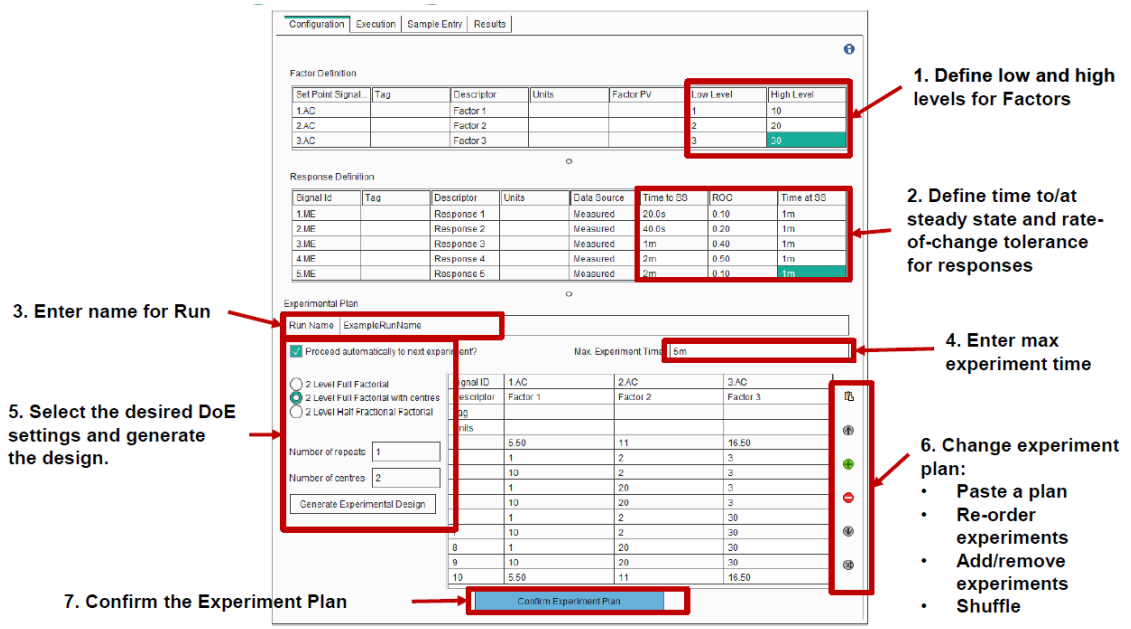


Figure 9.2 DoE Manager Configuration Tab in Perceptive Engineering's PharmaMV

9.1.3 Supply Chain Integration

This research further proposes a platform aims to provide a centralised repository for real-time data sharing across the supply chain, fostering collaboration and ensuring transparency. New users can register their accounts with the platform, providing necessary information such as company name, contact details, and role within the supply chain. Registered users can log in securely using their credentials to access the platform's features and data. The dashboard provides an overview of key performance indicators (KPIs), including production volumes, inventory levels, order status, and delivery schedules. Users can customise widgets to display specific metrics and data relevant to their roles and responsibilities. All relevant data, including product formulations, specifications, production schedules, and quality control reports, are stored in a centralised repository accessible to authorised users. Real-time updates, ensures that stakeholders have access to the most current information promptly. Users can share documents, reports, and other files directly within the platform, streamlining communication and information exchange. The platform integrates seamlessly with existing supply chain systems to enable data synchronisation and

automation. Advanced analytics tools enable users to analyse trends, identify insights, and make decisions to optimise processes and performance. Users can generate custom reports based on their specific requirements, with options for scheduling and automated delivery.

9.1.4 Lab Experimental Data Integration

Data was extracted from process data, historical data from the pilot plant DiPP and literature in the study and hence avoided the need to from scale up. Translating lab experimental data to pilot plant models involves a systematic approach to ensure accurate and efficient scale-up. Initially, comprehensive data from lab experiments, such as temperature, pressure, flow rates, and reaction times, are meticulously gathered. Dimensionless numbers like Reynolds (Re) and Damköhler (Da) are employed to maintain geometric, kinematic, and dynamic similarities between scales. Adjustments in operational conditions, including flow rates and mixing efficiency, are made to replicate lab conditions at the pilot scale. Pilot trials validate these models, providing critical insights for further adjustments. Throughout this process, safety assessments and feasibility studies ensure that the scale-up is not only technically sound but also economically viable. This iterative approach, backed by continuous data collection and analysis, bridges the gap between lab-scale experiments and full-scale industrial applications.

Chapter 10: Conclusion and Recommendations

The convergence of simulation, ML, statistical algorithms, and advanced hardware like high-performance computing, coupled with affordable sensors and cloud storage, is poised to revolutionise traditional engineering and production in manufacturing. Moving from a digital model to a digital twin requires exploring computationally efficient training methods for models using large datasets. This involves balancing learning outdated process dynamics with building a robust model. A promising approach lies in utilising the proposed adaptive and flexible modelling strategy for continuous learning. Several problems must be solved to establish a seamless integration of simulation models, statistics, and machine learning with control and monitoring platform. On the one hand, algorithmic advances must be made to ensure that hybrid algorithms can take advantage of the best of both worlds, retaining the high predictive accuracy and computationally cheap nature of data-driven models while also incorporating elements of interpretability, physical law confirmation, and simulation model trustworthiness. User-friendly software implementations should be offered for wider adoption in related domains and industrial applications. For example, in the field of machine learning, deep-learning tools like TensorFlow and Kera have democratised the technology, allowing beginner users to readily alter underlying programmes and apply them to specific use cases in a relatively short period.

PBM has been widely used to model blending and granulation processes due to its lower computational power requirements. TSWG latest PBM models have more than 25 rate mechanism parameters to calibrate, and their effects on specific operational conditions are difficult to determine. This work has demonstrated that sensitivity analysis can be a critical component for effective process model creation and calibration efforts, particularly when numerous input variables can alter model outputs. A new approach has been introduced to save time and speed up the calibration stage by systematically confirming which parameters are critical. The sensitivity analysis results were utilised to rank, filter, and map the region of the input variability space that is significant. A fresh unique calibration technique has been created and tested in which parameter estimation is conducted according to the order of parameter rankings until the model accuracy is met. The result of this study shows that nucleation, breakage, and layering have the greatest impact on granule development, whereas

consolidation processes have a much less impact. The new approach reduced the number of calibration parameters by 60%, leaving just ten crucial parameters to calibrate.

Furthermore, this work has demonstrated how to improve calibration accuracy successfully using flexible and adaptive calibration techniques such as calibration by range, by ranking, hybrid adaptive model reference. This work also successfully linked formulation and process properties using data driven methods and calibration correlations instead of relying on purely experimental methods. This has provided a good starting point for mechanistic understanding between mechanistic rate constants and process parameters. Furthermore, this work has demonstrated how to improve calibration accuracy successfully by breaking the particle distribution size range when utilising all the available data points in each run. Trying to consider all process parameter for calibration can cause massive error in the predictions.

This thesis has demonstrated the use of IOT historical data, experimental data, and literature data along to side to solving modelling deployment problems. To model TSWG energy usage, past campaign from 2021 were used to model torque against screw speed, powder flow and Liquid solid ratio. Research indicates that a significant portion, approximately 70%, of pharmaceutical manufacturing data is not captured and subsequently lost. (Manzano & Langer, 2018). This study demonstrated the use of an autoencoder machine learning algorithm to detect outliers in data when upgrading models. When it comes to using big data for manufacturing, data contamination is a major risk. The presented outlier detection system is quite resilient and adapts to changes in operational requirements. Using threshold settings, autoencoders have been shown to detect severe outliers, and system flaws, and invalidate experimentation assumptions in new data before applying a digital model.

Surrogate models are becoming increasingly used in the field of chemical process modelling. Their strength comes in their capacity to replace rigorous basic models, which speeds up computationally costly tasks such as superstructure optimisation and dynamic non-linear process prediction (Shokry et al., 2020; Granacher et al., 2021: Keßler et al., 2019). This thesis has demonstrated use of synthetic data can be generated from simulation models which can then be

combined with operational data available from pilot plants for training models. Mechanistic models can be used to accelerate the creation of surrogate models that execute very rapidly and robustly over the domain for which they are fitted. First-principle-model flowsheet developed produces high-fidelity results. However, such models can be very computationally demanding which inhibits their use in digital applications. The key objective of the developed model is to achieve predictive accuracy and through proper sampling methods, the model can generate a reasonable amount of pseudo data. These data sets are used to train a more computationally efficient and or reliable surrogate. For example, Twin screw wet machine learning surrogate model produced required 10 equations and a simulation time of 0.40672 s to produce particle size distribution results of 0.998 R^2 value while the PB model required 542127 equations and a simulation time of 83.7564 s. While the developed ANN demonstrated lower accuracy compared to other mechanistic models for predicting TSWG unit, it offers a faster and more reliable approach. Machine learning can be used to initially estimate the kinetic rate parameters for the range in which calibration is taking place.

Normally, the modelling cycle involves an iterative procedure. Once the designed experiments have been performed and the chosen measurements have been collected and imported, they can be used to generate better estimates of the parameters. Cross-validation and uncertainty propagation analysis can help to quantify the model's correctness and the quality of the predicted parameters. The new approach, which uses surrogate modelling, hybrid modelling, and global system analysis to speed the deployment of cost-effective, quick, and adaptable models, was subsequently applied to the tablet press process at the Diamond facility. This thesis demonstrated DiPP's in-line monitoring and control tool, Perceptive Engineering's PharmaMV and gPROMS mechanistic model synchronisation and signal readback and Python energy model integration.

In conclusion, manufacturing models have evolved over time, but they are still exclusively employed in the engineering stage due to their complexity for real-time applications. Companies have access to all models and data but do not understand how to use them. Several models have been published, but firms do not know how to implement or incorporate them into everyday operations in a timely and effective manner.

This thesis laid the groundwork for the application of engineering to digital applications. From the standpoint of industrial adoption, there must be an understanding of the potential effect of such an integrated vision, as well as the human resources required to carry out the project, anticipated project completion deadlines, expected outcomes, and return on investment.

Bibliography

Ahmadiyeh, F., Sajedi-Amin, S., Kafili-Hajlari, T. and Naseri, A. (2022). Roadmap for outlier detection in univariate linear calibration in analytical chemistry: Tutorial review. Journal of Chemometrics, 37(1). doi: https://doi.org/10.1002/cem.3460.

Akkisetty, P.., Lee, U., Reklaitis, G.V. and Venkatasubramanian, V., (2010). Population Balance Model-Based Hybrid Neural Network for a Pharmaceutical Milling Process. Journal of Pharmaceutical Innovation, 5(4), pp.161–168. doi: https://doi.org/10.1007/s12247-010-9090-2.

Altair Engineering Inc., Whitepaper, 2022 “A DEM Study of Wet Granulation in a Twin-Screw Granulator. [Online]. Available: https://www.research.ed.ac.uk/en/activities/a-dem-study-of- wet-granulation-in-a-twin-screw-granulator.

Amini, H., Palahnuk, H. and Akseli, I. (2020) “Population balance modeling (PBM) of ribbon milling in pharmaceutical roller compaction process,” Powder Technology, 376, pp. 438–457. Available at: https://doi.org/10.1016/J.POWTEC.2020.08.036.

Arthur, T.B. and Rahmanian, N. (2024) ‘Process simulation of twin-screw granulation: A Review’, Pharmaceutics, 16(6), p. 706. doi:10.3390/pharmaceutics16060706.

Arthur, T.B., Sekyi, N.K.G., Rahmanian, N. and Pu, J. (2023). Process Simulation of Twin-Screw Granulator: Effect of Screw Configuration on Size Distribution. Chemical Engineering & Technology. doi: https://doi.org/10.1002/ceat.202200539.

Bai, C., Dallasega, P., Orzes, G. and Sarkis, J., “Industry 4.0 technologies assessment: A sustainability perspective,” International Journal of Production Economics, vol. 229, p. 107 776, November 2020. DOI : 10.1016/j.ijpe.2020.107776.

Barni, A., Fontana, A., Menato, S., Sorlini, M. and Canetta, L., Exploiting the Digital Twin in the Assessment and Optimization of Sustainability Performances. IEEE, Sep. 2018. DOI: 10.1109/ is.2018.8710554. [Online]. Available: https://doi.org/10.1109/is.2018.8710554.

Barrasso, D., El Hagrasy, A., Litster, J.D. and Ramachandran, R. (2015). Multi-dimensional population balance model development and validation for a twin-screw granulation process. Powder Technology, 270, pp.612–621. doi: https://doi.org/10.1016/j.powtec.2014.06.035.

Barrasso, D., Oka, S., Muliadi, A., Litster, J.D., Wassgren, C. and Ramachandran, R. (2013). Population Balance Model Validation and Prediction of CQAs for Continuous Milling Processes: toward QbDin Pharmaceutical Drug Product Manufacturing. Journal of Pharmaceutical Innovation, 8(3), pp.147–162. doi: https://doi.org/10.1007/s12247-013-9155-0.

Barrasso, D., Tamrakar, A. and Ramachandran, R. (2015) “Model order reduction of a multi-scale PBM-DEM description of a wet granulation process via ANN,” Procedia Engineering, 102, pp. 1295–1304. Available at: https://doi.org/10.1016/J.PROENG.2015.01.260.

Bassolillo, S.R., D’Amato, E., Notaro, I., Blasi, L. and Mattei, M. (2020). Decentralized Mesh-Based Model Predictive Control for Swarms of UAVs. Sensors, 20(15), p.4324. doi: https://doi.org/10.3390/s20154324.

BBC, “Why is there a global energy crisis and who might suffer most from it?” BBC, 30 October 2022. [Online]. Available: https://www.bbc.co.uk/news/world-63430824.

Benjamin, B.J. and Madras, G. (2003) “Analytical solution for a population balance equation with aggregation and fragmentation,” Chemical Engineering Science, 58(13), pp. 3049–3051. Available at: https://doi.org/10.1016/S0009-2509(03)00159- 3.

Bostijn, N., Dhondt, J., Ryckaert, A., Szabó, E., Dhondt, W., Van Snick, B., Vanhoorne, V., Vervaet, C. and De Beer, T. (2019). A multivariate approach to predict the volumetric and gravimetric feeding behavior of a low feed rate feeder based on raw material properties. International Journal of Pharmaceutics, [online] 557, pp.342–353. doi: https://doi.org/10.1016/j.ijpharm.2018.12.066.

Boukouvala, F., Muzzio, F.J. and Ierapetritou, M.G. (2011) ‘Dynamic data-driven modeling of Pharmaceutical Processes’, Industrial & Engineering Chemistry Research, 50(11), pp. 6743–6754. doi:10.1021/ie102305a.

Calhan, S.D., Eker, E.D. and Sahin, N.O. (2017) ‘Quality by design (QBD) and Process Analytical Technology (PAT) applications in pharmaceutical industry’, European Journal of Chemistry, 8(4), pp. 430–433. doi:10.5155/eurjchem.8.4.430-433.1667.

Cameron, I.T., Wang, F.Y., Immanuel, C.D. and Stepanek, F. (2005). Process systems modelling and applications in granulation: A review. Chemical Engineering Science, 60(14), pp.3723–3750. doi: https://doi.org/10.1016/j.ces.2005.02.004.

Chablani, L., Taylor, M.K., Mehrotra, A., Rameas, P. and Stagner, W.C. (2011). Inline Real-Time Near-Infrared Granule Moisture Measurements of a Continuous Granulation–Drying–Milling Process. AAPS PharmSciTech, 12(4), pp.1050–1055. doi: https://doi.org/10.1208/s12249-011-9669-z.

Charoo, N.A., Khan, M.A. and Rahman, Z. (2023) ‘Data Integrity Issues in pharmaceutical industry: Common observations, challenges and mitigations strategies’, International Journal of Pharmaceutics, 631, p. 122503. doi: 10.1016/j.ijpharm.2022.122503.

Chaudhury, A., Barrasso, D., Pandey, P., Wu, H. and Ramachandran, R. (2014). Population Balance Model Development, Validation, and Prediction of CQAs of a High-Shear Wet Granulation Process: Towards QbD in Drug Product Pharmaceutical Manufacturing. Journal of Pharmaceutical Innovation, 9(1), pp.53–64. doi: https://doi.org/10.1007/s12247-014-9172-7.

Chen, H., Engkvist, O., Wang, Y., Olivecrona, M. and Blaschke, T. (2018). The rise of deep learning in drug discovery. Drug Discovery Today, [online] 23(6), pp.1241–1250. doi: https://doi.org/10.1016/j.drudis.2018.01.039.

Chen, X., He, C., Chen, Y. and Xie, Z. (2022). Internet of Things (IoT)—blockchain-enabled pharmaceutical supply chain resilience in the post-pandemic era. Frontiers of Engineering Management, 10(1). doi: https://doi.org/10.1007/s42524-022-0233-1.

Chen, Y., Sampat, C., Huang, Y.-S., Ganesh, S., Singh, R., Ramachandran, R., Reklaitis, G.V. and Marianthi Ierapetritou (2023). An integrated data management and informatics framework for continuous drug product

manufacturing processes: A case study on two pilot plants. 642, pp.123086–123086. doi: https://doi.org/10.1016/j.ijpharm.2023.123086.

Chen, Y., Yang, O., Sampat, C., Bhalode, P., Ramachandran, R. and Ierapetritou, M. (2020). Digital Twins in Pharmaceutical and Biopharmaceutical Manufacturing: A Literature Review. Processes, [online] 8(9), p.1088. doi: https://doi.org/10.3390/pr8091088.

Chukwu Christian Onyemaechi. Application Of Data Analytics in Process Predictions, Analysis, Management & Visualization Using Microsoft Power BI . TechRxiv. November 13, 2024.

Cimino, C., Negri, E. and Fumagalli, L. (2019) “Review of digital twin applications in manufacturing,” Computers in Industry, 113, p. 103130. Available at: https://doi.org/10.1016/J.COMPIND.2019.103130.

CloverDX (2023) ‘Building data pipelines to handle bad data: How to ensure data quality’, Cloverdx.com [Preprint]. Available at: https://doi.org/106293229876/1731679122631.

Cook, R.D. (1977). Detection of Influential Observation in Linear Regression. Technometrics, 19(1), p.15. doi: https://doi.org/10.2307/1268249.

Davidopoulou, C. and Ouranidis, A. (2022) ‘Pharma 4.0-artificially Intelligent Digital Twins for solidified Nanosuspensions’, Pharmaceutics, 14(10), p. 2113. doi:10.3390/pharmaceutics14102113.

De Leersnyder, F., Vanhoorne, V., Bekaert, H., Vercruysse, J., Ghijs, M., Bostijn, N., Verstraeten, M., Cappuyns, P., Van Assche, I., Vander Heyden, Y., Ziemons, E., Remon, J.P., Nopens, I., Vervaet, C. and De Beer, T. (2018). Breakage and drying behaviour of granules in a continuous fluid bed dryer: Influence of process parameters and wet granule transfer. European Journal of Pharmaceutical Sciences, [online] 115, pp.223–232. doi: https://doi.org/10.1016/j.ejps.2018.01.037.

Desai, K., Badhe, Y., Tambe, S.S. and Kulkarni, B.D. (2006). Soft-sensor development for fed-batch bioreactors using support vector regression. Biochemical Engineering Journal, 27(3), pp.225–239. doi: https://doi.org/10.1016/j.bej.2005.08.002.

Dhenge R.M. , Fyles R.S., Cartwright J.J. , Doughty D.G., Hounslow M.J., Salman A.D., Twin screw wet granulation: Granule properties, Chem. Eng. J., 164 (2010), pp. 322-329, 10.1016/j.cej.2010.05.023

Ding, J., Qu, L., Hu, X. and Liu, X. (2011). Application of Temperature Inference Method Based on Soft Sensor Technique to Plate Production Process. Journal of Iron and Steel Research International, 18(3), pp.24–27. doi: https://doi.org/10.1016/s1006-706x(11)60032-0.

Dong, Y., Yang, T., Xing, Y., Du, J. and Meng, Q. (2023). Data-Driven Modeling Methods and Techniques for Pharmaceutical Processes. Processes, [online] 11(7), p.2096. doi: https://doi.org/10.3390/pr11072096.

Dorota-Owczarek (2021) Process Analytical Technology (PAT) and Artificial Intelligence: The expanding impact of ML in drug manufacturing, nexocode. Available at: https://nexocode.com/blog/posts/process-analytical-technology-pat-with-ai-in-biopharmaceutical-manufacturing/ (Accessed: 04 June 2024).

Douglas J. M., Conceptual design of Chemical Processes. 1988, ISBN: 978-0070177628.

Drug Product Process Development & Optimization | Lonza (2022). Available at: https://pharma.lonza.com/offerings/oral-drug-products-and-intermediates/processdevelopment-and-optimization (Accessed: February 3, 2022).

Ekins, S. (2016) 'The next era: Deep Learning in Pharmaceutical Research', Pharmaceutical Research, 33(11), pp. 2594–2603. doi:10.1007/s11095-016-2029-7.

EMA. "Quality by design." (May 2020), [Online]. Available: https:/ / www.ema.europa.eu / en / human-regulatory/research-development/quality-design (visited on 07/01/2023).

Fonteyne, M., Correia, A., De Plecker, S., Vercruysse, J., Ilić, I., Zhou, Q., Vervaet, C., Remon, J.P., Onofre, F., Bulone, V. and De Beer, T. (2015). Impact of microcrystalline cellulose material attributes: A case study on continuous twin screw granulation. International Journal of Pharmaceutics, 478(2), pp.705–717. doi: https://doi.org/10.1016/j.ijpharm.2014.11.070.

Food and Drug Administration (FDA). (2020). Risk Management Guidance for Pharmaceutical Development. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/q9r1-quality-risk-management.

Fülöp, G., Domokos, A., Galata, D., Szabó, E., Gyürkés, M., Szabó, B., Farkas, A., Madarász, L., Démuth, B., Lendér, T., Nagy, T., Kovács-Kiss, D., Van der Gucht, F., Marosi, G. and Nagy, Z.K. (2021). Integrated twin-screw wet granulation, continuous vibrational fluid drying and milling: A fully continuous powder to granule line. International Journal of Pharmaceutics, 594, p.120126. doi: https://doi.org/10.1016/j.ijpharm.2020.120126.

Gargalo, C.L., Udugama, I., Pontius, K., Lopez, P.C., Nielsen, R.F., Hasanzadeh, A., Mansouri, S.S., Bayer, C., Junicke, H. and Gernaey, K.V. (2020). Towards smart biomanufacturing: a perspective on recent developments in industrial measurement and monitoring technologies for bio-based production processes. Journal of Industrial Microbiology & Biotechnology, 47(11), pp.947–964. doi: https://doi.org/10.1007/s10295-020-02308-1.

Gavi, E. and Reynolds, G.K. (2014) 'System model of a tablet manufacturing process', Computers & Chemical Engineering, 71, pp. 130–140. doi: 10.1016/j.compchemeng.2014.07.026.

GEA Group AG. "CONSIGMA® GRANULATION AND DRYING (GD) MODULES." (2023), [Online]. Available: https: / / www.gea.com / en / products / granulators / continuous - granulation - lines / consigma-granulation-drying-gd-modules.jsp (visited on 09/05/2023).

GEA Group AG. "MODUL P TABLET PRESS." (2023), [Online]. Available: https://www.gea.com/ en/products/tablet-presses/modul-p-tablet-press.jsp (visited on 09/05/2023).

Ghahramani, P. (2016a) 'Modeling and simulation applications in drug development process', Translational Medicine, pp. 55–65. doi:10.1016/b978-0-12-803460-6.00003-9.

Good manufacturing practice (GMP) (no date) siemens.com Global Website. Available at: https://www.siemens.com/global/en/industries/pharmaceutical-life-

science-industries/pharma-industry/good-manufacturing-practice.html?gclid=CjwKCAiAtt2tBhBDEiwALZuhAE-Y_PaSV_cHNrM7i5ZzGVEDu04puT8cKb40Qklo8iTjYjJ1Q7qQhRoCF68QAvD_BwE&acz=1&gad_source=1#ComplianceResponses (Accessed: 30 January 2024).

Good manufacturing practice (GMP) resources (2019) ISPE. Available at: https://ispe.org/initiatives/regulatory-resources/gmp (Accessed: 15 May 2024).

Grelier, A., Matej Zadravec, Johan Remmelgas, Forgber, T., Colacino, F., Pilcer, G., Stauffer, F. and Hörmann-Kincses, T. (2022). Model-Guided Development of a Semi-Continuous Drying Process. Pharmaceutical Research, 39(9), pp.2005–2016. doi: https://doi.org/10.1007/s11095-022-03361-4.

Grieves, M., 2014. Digital twin: manufacturing excellence through virtual factory replication. White paper, 1(2014), pp.1-7.

Grieves, M., Sme management forum completing the cycle: Using plm information in the sales and service functions, October 2002. [Online]. Available: https://www.researchgate .net /publication/356192963_SME_Management_Forum_Completing_the_Cycle_Using_PLM_Information_in_ the_Sales_and_Service_Functions.

Guidance for Industry: Q8(R2) Pharmaceutical Development, ICH Harmonised Tripartite Guideline, Step 4, August 2009

Gunes, V., Peter S., Givargis T., Vahid F., et al. (2014) “A Survey on Concepts, Applications, and Challenges in Cyber-Physical Systems. (2014). KSII Transactions on Internet and Information Systems, 8(12). doi: https://doi.org/10.3837/tiis.2014.12.001.

He B. and Bai K.-J., “Digital twin-based sustainable intelligent manufacturing: A review,” Advances in Manufacturing, vol. 9, no. 1, pp. 1–21, May 2020. DOI: 10.1007/s40436-020-00302-5. [Online]. Available: https://doi.org/10.1007/s40436-020-00302-5.

Herrmann, C., Schmidt C., Kurle, D., Blume, S., and Thiede, S., “Sustainability in manufacturing and factories of the future,” International Journal of Precision

Engineering and Manufacturing- Green Technology, vol. 1, no. 4, pp. 283–292, October 2014. DOI: 10.1007/s40684-014-0034-z.

Hoang, H. (2015) An NDA at the FDA Understanding the Drug Approval Process. Huang, J. et al. (2021) “ AIChE PD2M Advanced Process Control workshop-moving APC forward in the pharmaceutical industry ,” Journal of Advanced Manufacturing and Processing, 3(1). Available at: https://doi.org/10.1002/amp2.10071.

Ierapetritou, M., Muzzio, F., and Reklaitis, G., (2016) “Perspectives on the continuous manufacturing of powder-based pharmaceutical processes,” AIChE Journal, 62(6), pp. 1846–1862. Available at: https://doi.org/10.1002/AIC.15210.

Industrial Edge from Siemens adds benefits from the cloud at the field level | Press | Company | Siemens (2018). Available at: https://press.siemens.com/global/en/pressrelease/industrial-edge-siemens-addsbenefits-cloud-field-level (Accessed: February 25, 2022).

Industry 4.0 and the fourth industrial revolution explained (2006). Available at: https://www.i-scoop.eu/industry-4-0/ (Accessed: December 9, 2021).

International Council for Harmonisation (ICH). (2022). ICH Q8(R2) Pharmaceutical Development. https://database.ich.org/sites/default/files/Q8_R2_Guideline.pdf

International Organization for Standardization, Geneva, CH, Standard, October 2021, “Automation systems and integration. Digital twin framework for manufacturing - Overview and general principles. BS ISO 23247-1:2021 [Online]. Available: https :// www. iso. org / standard / 75066.html.

ir. Ashish Kumar, “Experimental and model-based analysis of twin-screw wet granulation in pharmaceutical processes,” Ph.D. dissertation, Ghent University, 2015. [Online]. Available: https://biomath.ugent.be/sites/default/files/2018-05/PhDThesis_AshishKumar_0.pdf.

Is Pharma doing enough on sustainability? | Pharmafile (2020). Available at: http://www.pharmafile.com/news/553923/pharma-doing-enough-sustainability (Accessed: October 13, 2021).

Iveson, S.M., Litster, J.D., and Ennis, B.J. (1996) ‘Fundamental studies of granule consolidation part 1: Effects of binder content and binder viscosity’, Powder Technology, 88(1), pp. 15–20. doi:10.1016/0032-5910(96)03096-3.

Jagtap, K., Chaudhari, B. and Redasani, V. (2022) ‘Quality by Design (QBD) concept review in pharmaceuticals’, Asian Journal of Research in Chemistry, pp. 303–307. doi:10.52711/0974-4150.2022.00054.

Javaid, M., Haleem, A., Singh, R.P. and Suman, R. (2022). An integrated outlook of Cyber-Physical systems for Industry 4.0: Topical practices, architecture, and applications. Green Technologies and Sustainability, 1(1), p.100001. doi: https://doi.org/10.1016/j.grets.2022.100001.

Jerry Martin, P., 2021. Improving Sustainability in Pharmaceutical Manufacturing - Pharmaceutical Processing World. [online] Pharmaceutical Processing World. Available at: <https://www.pharmaceuticalprocessingworld.com/improving-sustainability-in-pharmaceutical-manufacturing/> [Accessed 26 April 2021].

Jung, W. and Taflanidis, A.A. (2023) ‘Efficient global sensitivity analysis for high-dimensional outputs combining data-driven probability models and dimensionality reduction’, Reliability Engineering & System Safety, 231, p. 108805. doi: 10.1016/j.ress.2022.108805.

Kadlec, P., Gabrys, B. and Strandt, S. (2009) “Data-driven Soft Sensors in the process industry,” Computers & Chemical Engineering, 33(4), pp. 795–814. Available at: https://doi.org/10.1016/J.COMPCHEMENG.2008.12.012.

Kakad SB, Kolhe MH, Dukre TP. A Review on Pharmaceutical Validation. International Journal of Pharmaceutical Quality Assurance. 2020;11(3):338-342.

kanncaa1 (2019) Time series prediction tutorial with EDA, Kaggle. Available at: https://www.kaggle.com/code/kanncaa1/time-series-prediction-tutorial-with-eda (Accessed: 28 February 2024).

Kapur, P.C. (1972) “Kinetics of granulation by non-random coalescence mechanism,” Chemical Engineering Science, 27(10), pp. 1863–1869. Available at: https://doi.org/10.1016/0009-2509(72)85048-6.

Kim E.J., Kim J.H., Kim M. S., Jeong S.H., and Choi D.H., (2021). Process Analytical Technology Tools for Monitoring Pharmaceutical Unit Operations: a Control Strategy for Continuous Process Verification. Pharmaceutics, [online] 13(6), p.919. doi: https://doi.org/10.3390/pharmaceutics13060919.

Kritzinger, W., Karner, M., Traar, G., Henjes, J. and Sihn, W. (2018). Digital Twin in manufacturing: A categorical literature review and classification. IFAC-PapersOnLine, [online] 51(11), pp.1016–1022. doi: https://doi.org/10.1016/j.ifacol.2018.08.474.

Kumar, K.A., Saravanakumar, M., Joseph, J. and Ramanathan, H. (2016). Generative Model for Conceptual Design of Defence Equipment. Defence Science Journal, 66(1), p.81. doi: https://doi.org/10.14429/dsj.66.9105.

Leane M., Pitt K., Reynolds G.K., Dawson N., Ziegler,I., Szepes A., Crean A.M., Dall Agnol R. (2018). Manufacturing classification system in the real world: factors influencing manufacturing process choices for filed commercial oral solid dosage formulations, case studies from industry and considerations for continuous processing. Pharmaceutical Development and Technology, 23(10), pp.964–977. doi: https://doi.org/10.1080/10837450.2018.1534863.

Lee, S.L., O'Connor, T.F., Yang, X., Cruz, C.N., Chatterjee, S., Madurawe, R.D., Moore, C.M.V., Yu, L.X. and Woodcock, J. (2015). Modernizing Pharmaceutical Manufacturing: from Batch to Continuous Production. Journal of Pharmaceutical Innovation, 10(3), pp.191–199. doi: https://doi.org/10.1007/s12247-015-9215-8.

Lerapetritou, M., Muzzio, F. and Reklaitis, G., "Perspectives on the continuous manufacturing of powder-based pharmaceutical processes," AIChE Journal, vol. 62, no. 6, pp. 1846–1862, March 2016. DOI: 10.1002/aic.15210.

Litster, J., and Bogle, I. D. L. (2019). Smart Process Manufacturing for Formulated Products. In Engineering (Vol. 5, Issue 6, pp. 1003–1009). Elsevier Ltd. https://doi.org/10.1016/j.eng.2019.02.014

Liu, X., et al. (2022). Influence of binder type and concentration on tablet tensile strength and disintegration time. AAPS PharmSciTech, 23(1), 1-10.

Lopez, F.L., Ernest, T.B., Tuleu, C. and Gul, M.O. (2015). Formulation approaches to pediatric oral drug delivery: benefits and limitations of current platforms. Expert Opinion on Drug Delivery, 12(11), pp.1727–1740. doi: https://doi.org/10.1517/17425247.2015.1060218.

Madarász L., Nagy Z. K., Hoffer, I., Szabó B., Csontos, I., Pataki, H., Barna D.B., Szabó B., Csorba, K. and Marosi, G. (2018). Real-time feedback control of twin-screw wet granulation based on image analysis. International Journal of Pharmaceutics, 547(1-2), pp.360–367. doi: https://doi.org/10.1016/j.ijpharm.2018.06.003.

Manoharan, S. (2018). Smart Supply Chain Management using Internet of Things (IoT). International Journal of Systems, Control and Communications, 9(2), p.1. doi: https://doi.org/10.1504/ijscc.2018.10008373.

Manzano, T. and Langer, G., Getting ready for pharma 4.0. data integrity in cloud and big data applications, 2018. [Online]. Available : https://www.ispe.gr.jp/ISPE/02_katsudou/pdf/201812_ en.pdf.

Markl, D., Warman, M., Dumarey, M., Bergman, E.-L., Folestad, S., Shi, Z., Manley, L.F., Goodwin, D.J. and Zeitler, J.A. (2020). Review of real-time release testing of pharmaceutical tablets: State-of-the art, challenges and future perspective. International Journal of Pharmaceutics, 582, p.119353. doi: https://doi.org/10.1016/j.ijpharm.2020.119353.

Martin.S. “What is a digital twin?” (Sep. 2022), [Online]. Available: https://blogs.nvidia.com/ blog/2021/12/14/what-is-a-digital-twin/. ShanmugamS., “Granulation techniques and technologies: Recent progresses,” BioImpacts, vol. 5, no. 1, pp. 55–63, August 2017. DOI: 10.15171/bi.2015.04.

Matsui, Y. and Watano, S. (2018) “Evaluation of properties of granules and tablets prepared by twin-screw continuous granulation and comparison of their properties with those by batch fluidized-bed and high shear granulations,” Journal of the Society of Powder Technology, Japan, 55(2), pp. 86–94. Available at: https://doi.org/10.4164/SPTJ.55.86.

Matsunami, K., Miura, T., Yaginuma, K., Tanabe, S., Badr, S. and Sugiyama, H. (2023). Surrogate modeling of dissolution behavior toward efficient design of

tablet manufacturing processes. Computers & Chemical Engineering, 171, p.108141. doi: https://doi.org/10.1016/j.compchemeng.2023.108141.

McGuire, A.D., Mosbach, S., Lee, K.-O., Reynolds, G.P. and Reynolds, G.P. (2018). A high-dimensional, stochastic model for twin-screw granulation – Part 1: Model description. 188, pp.221–237. doi: https://doi.org/10.1016/j.ces.2018.04.076.

Megarry, A., Taylor, A., Gholami, A., Wikström, H. and Tajarobi, P. (2020). Twin-screw granulation and high-shear granulation: The influence of mannitol grade on granule and tablet properties. International Journal of Pharmaceutics, 590, p.119890. doi: https://doi.org/10.1016/j.ijpharm.2020.119890.

Meier, R., Moll, K.-P., Krumme M. and Kleinebudde P. (2017). Impact of fill-level in twin-screw granulation on critical quality attributes of granules and tablets. European Journal of Pharmaceutics and Biopharmaceutics, 115, pp.102–112. doi: https://doi.org/10.1016/j.ejpb.2017.02.010.

Meier, R., Thommes, M., Rasenack, N., Krumme, M., Moll, K.-P. and Kleinebudde, P. (2015b). Simplified formulations with high drug loads for continuous twin-screw granulation. International Journal of Pharmaceutics, 496(1), pp.12–23. doi: https://doi.org/10.1016/j.ijpharm.2015.05.060.

Metta, N., Ghijs, M., Schäfer, E., Kumar, A., Cappuyns, P., Van Assche , I., Singh , R., Ramachandran , R., De Beerr, T., Ierapetritou, M., & Nopens, I. (2019). Dynamic Flowsheet Model Development and Sensitivity Analysis of a Continuous Pharmaceutical Tablet Manufacturing Process Using the Wet Granulation Route. Processes 2019, Vol. 7, Page 234, 7(4), 234. https://doi.org/10.3390/PR7040234

Mortier, S.T.F.C. et al. (2011) “Mechanistic modelling of fluidized bed drying processes of wet porous granules: A review,” European Journal of Pharmaceutics and Biopharmaceutics, 79(2), pp. 205–225. Available at: https://doi.org/10.1016/J.EJPB.2011.05.013.

Nagy, B., Galata, D.L., Farkas, A. and Nagy, Z.K. (2022). Application of Artificial Neural Networks in the Process Analytical Technology of Pharmaceutical Manufacturing—a Review. The AAPS Journal, 24(4). doi: https://doi.org/10.1208/s12248-022-00706-0.

Nakagawa, E.Y. et al. (2021a) “Industry 4.0 reference architectures: State of the art and future trends,” Computers & Industrial Engineering, 156, p. 107241. Available at : https://doi.org/10.1016/J.CIE.2021.107241.

Nakagawa, E.Y. et al. (2021b) “Industry 4.0 reference architectures: State of the art and future trends,” Computers & Industrial Engineering, 156, p. 107241. Available at: https://doi.org/10.1016/J.CIE.2021.107241.

Nielsen, R.F., Nazemzadeh, N., Sillesen, L.W., Andersson, M.P., Gernaey, K.V. and Mansouri, S.S. (2020). Hybrid machine learning assisted modelling framework for particle processes. Computers & Chemical Engineering, 140, p.106916. doi: https://doi.org/10.1016/j.compchemeng.2020.106916.

Ntamo, D., Montero, E., Omar, C., Highett, M., Moss, D., Soulatintork, P., Moghadam, P., Zandi, M. and Mitchell, N., 2022. Digitalisation of a Continuous Process Manufacturing for Formulated Products: Adding IoT capabilities to a legacy process. Digital Chemical Engineering. vol. 3, p.100 025, Jun. 2022. DOI: 10.1016/j.dche.2022.100025.

Ntamo, D.; Papadopoulos, I.; Omar, C.; Soulatiantork, P.; Zandi, M. A Sustainability-Oriented Digital Twin of the Diamond Pilot Plant. Processes 2025, 13, 211. https://doi.org/10.3390/pr13010211

O’Halloran, J. Nearly Half of Firms to Increase Investments in IoT Despite the Impact of COVID-19. Available online: https://www.computerweekly.com/news/252491333/Nearly-half-of-firms-to-increase-investments-in-IoT-despite-the-impact-of-Covid-19 (accessed on 30 May 2024)

Schume, P. “Improve product quality and yield with intelligent, secure, and adaptable manu- facturing operations.” (May 2020), [Online]. Available: https://www.ibm.com/blogs/internet- of-things/iot-manufacturing-ready/ (visited on 08/01/2023).

Pachayappan M., Nelavala Rajesh, G. Saravanan, (2016) Smart logistics for pharmaceutical industry based on Internet of Things (IoT) Vol. 14 CIC 2016 Special Issue International Journal of Computer Science and Information Security, ISSN 1947-5500.

Park, C. and Leeds, M. (2016) ‘A highly efficient robust design under data contamination’, Computers & Industrial Engineering, 93, pp. 131–142. doi: 10.1016/j.cie.2015.11.016.

Pascoal, F. and Silveira, J. (2013). Sustainable machining - correlation of the optimization by minimum energy, minimum manufacturing time and cost of production.

Pitt, K.G., Newton, J.M., Richardson, R. and Stanley, P. (1989). The Material Tensile Strength of Convex-faced Aspirin Tablets. Journal of Pharmacy and Pharmacology, 41(5), pp.289–292. doi: https://doi.org/10.1111/j.2042-7158.1989.tb06458.x.

Pitt, K.G., Newton, J.M., Richardson, R. and Stanley, P. (1989). The Material Tensile Strength of Convex-faced Aspirin Tablets. Journal of Pharmacy and Pharmacology, 41(5), pp.289–292. doi: https://doi.org/10.1111/j.2042-7158.1989.tb06458.x.

Portier C., Pandelaere K., Delaet U., Vigh T., Kumar A., Pretoro G.D, Thomas De Beer T., Chris Vervaet C., Vanhoorne V.Portier, C. et al. (2020) “Continuous twin screw wet granulation: Influence of process and formulation variables on granule quality attributes of model formulations.,” International Journal of Pharmaceutics, 576, pp. 118981–118981. Available at: https://doi.org/10.1016/J.IJPHARM.2019.118981.

Portier, C., ervaet, C. and anhoorne,. (2021) “Continuous Twin screw wet Granulation: A Review of Recent Progress and Opportunities in Formulation and Equipment Design,” Pharmaceutics 2021, Vol. 13, Page 668, 13(5), p. 668. Available at: https://doi.org/10.3390/PHARMACEUTICS13050668.

Production-and-process-controls--overview-of-CGMP-(2015) ... Available at: https://www.fda.gov/files/drugs/published/Production-and-Process-Controls--Overview-of-CGMP-Regulations-and-Regulatory-Expectations.pdf (Accessed: 15 May 2024).

Pugliese, R., Regondi, S. and Marini, R. (2021) ‘Machine learning-based approach: Global Trends, Research Directions, and regulatory standpoints’, Data Science and Management, 4, pp. 19–29. doi: 10.1016/j.dsm.2021.12.002.

Pundir, A. K., Jagannath, J. D. and L. Ganapathy, "IMPROVING SUPPLY CHAIN VISIBILITY USING IoT-INTERNET OF THINGS," 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 2019, pp. 0156-0162, doi: 10.1109/CCWC.2019.8666480.

Qi, Q., Tao, F., Hu, T., Anwer, N., Liu, A., Wei, Y., Wang, L. and Nee, A.Y.C. (2019). Enabling technologies and tools for digital twin. Journal of Manufacturing Systems, 58. doi: https://doi.org/10.1016/j.jmsy.2019.10.001. Qi,

Rathore, A.S. (2014) 'QBD/Pat for bioprocessing: Moving from theory to implementation', Current Opinion in Chemical Engineering, 6, pp. 1–8. doi: 10.1016/j.coche.2014.05.006.

Reynolds, G. (2019) Application of hybrid models for Advanced Process Control of a Twin screw wet Wet Granulation Process. London. Roblek, Meško, M. and Krapež, A. (2016) "A Complex iew of Industry 4.0:" https://doi.org/10.1177/2158244016653987, 6(2). Available at: https://doi.org/10.1177/2158244016653987.

Roblek, Meško, M. and Krapež, A. (2016) "A Complex view of Industry 4.0:" https://doi.org/10.1177/2158244016653987, 6(2). Available at: https://doi.org/10.1177/2158244016653987.

Rogers, A. J. (2015). ROCESS SYSTEMS ENGINEERING METHODS FOR THE DEVELOPMENT OF CONTINUOUS PHARMACEUTICAL MANUFACTURING PROCESSES.

Rogers, A., Hashemi, A. and Ierapetritou, M. (2013). Modeling of Particulate Processes for the Continuous Manufacture of Solid-Based Pharmaceutical Dosage Forms. Processes, 1(2), pp.67–127. doi: https://doi.org/10.3390/pr1020067.

Rogina, A., Šiško, I., Mohler, I., Ujević, Ž. and Bolf, N. (2011). Soft sensor for continuous product quality estimation (in crude distillation unit). Chemical Engineering Research and Design, 89(10), pp.2070–2077. doi: https://doi.org/10.1016/j.cherd.2011.01.003.

Ryckaert, A., Ghijs, M., Portier, C., Djuric, D., Funke, A., Vervaet, C. and De Beer, T. (2021). The Influence of Equipment Design and Process Parameters

on Granule Breakage in a Semi-Continuous Fluid Bed Dryer after Continuous Twin-Screw Wet Granulation. Pharmaceutics, 13(2), p.293. doi: https://doi.org/10.3390/pharmaceutics13020293.

Sacher S., Poms J., Rehrl J., Khinast J.G., Sacher, S. et al. (2022) 'PAT implementation for advanced process control in solid dosage manufacturing – A practical guide', International Journal of Pharmaceutics, 613, p. 121408. doi: 10.1016/j.ijpharm.2021.121408.

Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M. and Tarantola, S. (2010). Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. Computer Physics Communications, [online] 181(2), pp.259–270. doi: https://doi.org/10.1016/j.cpc.2009.09.018.

Sarkis, M., Bernardi, A., Shah, N. and Papathanasiou, M.M. (2021). Emerging Challenges and Opportunities in Pharmaceutical Manufacturing and Distribution. Processes, [online] 9(3), p.457. doi: https://doi.org/10.3390/pr9030457.

Schaber, S.D., Gerogiorgis, D.I., Ramachandran, R., Evans, J.M.B., Barton, P.I. and Trout, B.L. (2011). Economic Analysis of Integrated Continuous and Batch Pharmaceutical Manufacturing: A Case Study. Industrial & Engineering Chemistry Research, 50(17), pp.10083–10092. doi: https://doi.org/10.1021/ie2006752.

Schiemer, R., Weggen, J.T., Schmitt, K.M. and Hubbuch, J. (2023b). An adaptive soft-sensor for advanced real-time monitoring of an antibody-drug conjugation reaction. Biotechnology and Bioengineering, 120(7), pp.1914–1928. doi: https://doi.org/10.1002/bit.28428.

Sebastian Escotet-Espinoza, M., Foster, C.J. and Ierapetritou, M. (2018) "Discrete Element Modelling (DEM) for mixing of cohesive solids in rotating cylinders," Powder Technology, 335, pp. 124–136. Available at: https://doi.org/10.1016/J.POWTEC.2018.05.024.

Seem, T. C., Rowson, N. A., Ingram A. et al., "Twin screw wet granulation — a literature review," Powder Technology, vol. 276, pp. 89–102, May 2015. DOI: 10.1016/j.powtec.2015.01.075.

Shanmugam, S., “Granulation techniques and technologies: Recent progresses,” BioImpacts, vol. 5, no. 1, pp. 55–63, August 2017. DOI : 10.15171/bi.2015.04.

Shirazian, S., Kuhs, M., Darwish, S., Croker, D. and Walker, G.M. (2017). Artificial neural network modelling of continuous wet granulation using a twin-screw extruder. International Journal of Pharmaceutics, 521(1-2), pp.102–109. doi: https://doi.org/10.1016/j.ijpharm.2017.02.009.

Siemens Industry Limited (2024), gPROMS FormulatedProducts Documentation.

Singh, M., Shirazian, S., Ranade, V., Walker, G.M. and Kumar, A. (2022). Challenges and opportunities in modelling wet granulation in pharmaceutical industry – A critical review. Powder Technology, 403, p.117380. doi: https://doi.org/10.1016/j.powtec.2022.117380.

Singh, R., Sahay, A., Muzzio, F., Ierapetritou, M. and Ramachandran, R. (2014). A systematic framework for onsite design and implementation of a control system in a continuous tablet manufacturing process. Computers & Chemical Engineering, 66, pp.186–200. doi: https://doi.org/10.1016/j.compchemeng.2014.02.029.

Ślusarczyk, B. (2018) “Industry 4.0 – Are we ready?” Polish Journal of Management Studies, 17(1), pp. 232–248. Available at: https://doi.org/10.17512/PJMS.2018.17.1.19.

Soos, M., Sefcik, J. and Morbidelli, M. (2006) “Investigation of aggregation, breakage and restructuring kinetics of colloidal dispersions in turbulent flows by population balance modelling and static light scattering,” Chemical Engineering Science, 61(8), pp. 2349–2363. Available at: https://doi.org/10.1016/J.CES.2005.11.001.

Stauffer, F., Vanhoorne, V., Pilcer, G., Chavez, P.-F., Vervaet, C. and De Beer, T. (2019). Managing API raw material variability in a continuous manufacturing line – Prediction of process robustness. International Journal of Pharmaceutics, 569, p.118525. doi: https://doi.org/10.1016/j.ijpharm.2019.118525.

Stegemann, S. (2016) “The future of pharmaceutical manufacturing in the context of the scientific, social, technological and economic evolution,” European Journal of Pharmaceutical Sciences, 90, pp. 8–13. Available at : https://doi.org/10.1016/J.EJPS.2015.11.003.

Su, Q., Bommireddy, Y., Shah, Y., Ganesh, S., Moreno, M., Liu, J., Gonzalez, M., Yazdanpanah, N., O’Connor, T., Reklaitis, G.V. and Nagy, Z.K. (2019a). Data reconciliation in the Quality-by-Design (QbD) implementation of pharmaceutical continuous tablet manufacturing. International Journal of Pharmaceutics, [online] 563, pp.259–272. doi: https://doi.org/10.1016/j.ijpharm.2019.04.003.

Su, Q., Ganesh, S., Moreno, M., Bommireddy, Y., Gonzalez, M., Reklaitis, G.V. and Nagy, Z.K. (2019). A perspective on Quality-by-Control (QbC) in pharmaceutical continuous manufacturing. Computers & Chemical Engineering, 125, pp.216–231. doi: https://doi.org/10.1016/j.compchemeng.2019.03.001.

Suresh, P. and Basu, P.K. (2008) “Improving pharmaceutical product development and manufacturing: Impact on cost of drug development and cost of goods sold of pharmaceuticals,” Journal of Pharmaceutical Innovation, 3(3), pp. 175–187. Available at: https://doi.org/10.1007/S12247-008-9043-1.

Swami, A., Shivani Chakankar, Chavan, P., Limaye, D. and Amol Tagalpallewar (2023). A Critical Review on Recently Used PAT in Pharmaceutical Industry. Research Journal of Pharmacy and Technology, pp.4436–4442. doi: https://doi.org/10.52711/0974-360x.2023.00724.

Tężyk, M., Milanowski, B., Ernst, A. and Lulek, J. (2015). Recent progress in continuous and semi-continuous processing of solid oral dosage forms: a review. Drug Development and Industrial Pharmacy, 42(8), pp.1195–1214. doi: https://doi.org/10.3109/03639045.2015.1122607.

Thompson, M.R. (2015) “Twin screw wet granulation – review of current progress,” http://dx.doi.org/10.3109/03639045.2014.983931, 41(8), pp. 1223–1231. Available at: https://doi.org/10.3109/03639045.2014.983931.

Torabi, H., Mirtaheri, S.L. and Greco, S. (2023) ‘Practical autoencoder based anomaly detection by using vector reconstruction error’, Cybersecurity, 6(1). doi:10.1186/s42400-022-00134-9.

Toson, P., Siegmann, E., Trogrlić, M., Hermann Kureck, Khinast, J.G., Dalibor Jajčević, Doshi, P., Blackwood, D.O., Alexandre Bonnassieux, Daugherity, P.D. and Mary (2018). Detailed modeling and process design of an advanced continuous powder mixer. International Journal of Pharmaceutics, 552(1-2), pp.288–300. doi: https://doi.org/10.1016/j.ijpharm.2018.09.032.

U.S. Food and Drug Administration (2018a) FDA approves abemaciclib as initial therapy for HR-positive, HER2-negative metastatic breast cancer | FDA. Available at: https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approvesabemaciclib-initial-therapy-hr-positive-her2-negative-metastatic-breast-cancer (Accessed: October 25, 2022).

U.S. Food and Drug Administration (2018b) FDA approves new treatment for patients with acute myeloid leukemia | FDA. Available at: https://www.fda.gov/newsevents/press-announcements/fda-approves-new-treatment-patients-acute-myeloidleukemia (Accessed: October 25, 2022).

UK Government, Government report, 2016. [Online]. “2010 to 2015 government policy: energy demand reduction in industry, business and the public sector,” Available: https://www. gov.uk/government/publications/2010- to- 2015-government- policy- energy- demand- reduction- in-industry-business-and-the-public-sector/2010-to-2015-government-policy-energy-demand- reduction-in-industry-business-and-the-public-sector.

Vanarase, A.U. and Muzzio, F.J. (2011) “Effect of operating conditions and design parameters in a continuous powder mixer,” Powder Technology, 208(1), pp. 26–36. Available at: https://doi.org/10.1016/J.POWTEC.2010.11.038.

Vandeputte T., Ghijs M., Hauwermeiren D.V., Schultz E.D.S, Schäfer E., Stauffer, F., De Beer T., Nopens I. (2023) ‘Mechanistic modeling of semicontinuous fluidized bed drying of pharmaceutical granules by incorporating single particle and bulk drying kinetics’, International Journal of Pharmaceutics, 646, p. 123447. doi: 10.1016/j.ijpharm.2023.123447.

VanDerHorn, E. and Mahadevan, S., “Digital twin: Generalization, characterization and implementation,” Decision Support Systems, vol. 145, p. 113 524, Jun. 2021. DOI: 10.1016/j.dss. 2021.113524. [Online]. Available: https://doi.org/10.1016/j.dss.2021.113524.

VanDerHorn, E. and Mahadevan, S., “Digital twin: Generalization, characterization and implementation,” Decision Support Systems, vol. 145, p. 113 524, Jun. 2021. DOI: 10.1016/j.dss. 2021.113524. [Online]. Available: https://doi.org/10.1016/j.dss.2021.113524.(E. VanDerHorn and S. Mahadevan, 2021)

Vanhoorne, V. and Vervaet, C. (2020). Recent progress in continuous manufacturing of oral solid dosage forms. International Journal of Pharmaceutics, 579, p.119194. doi: https://doi.org/10.1016/j.ijpharm.2020.119194.

Vargas, J.M., Roman-Ospino, A.D., Sanchez, E. and Romañach, R.J. (2017) “Evaluation of Analytical and Sampling Errors in the Prediction of the Active Pharmaceutical Ingredient Concentration in Blends from a Continuous Manufacturing Process.” Available at: https://doi.org/10.1007/s12247- 017-9273-1.

Veleva, V., M. Hart, M., T. Greiner, T. and C. Crumbley, C., “Indicators for measuring environmental sustainability,” Benchmarking: An International Journal, vol. 10, no. 2, pp. 107–119, April 2003. DOI: 10.1108/14635770310469644. [Online]. Available: https://doi.org/10.1108/14635770310469644.

Vercruysse, J., Burggraeve, A., Fonteyne, M., Cappuyns, P., Delaet, U., Van Assche, I., De Beer, T., Remon, J.P. and Vervaet, C. (2015). Impact of screw configuration on the particle size distribution of granules produced by twin screw granulation. International Journal of Pharmaceutics, 479(1), pp.171–180. doi: https://doi.org/10.1016/j.ijpharm.2014.12.071.

Vercruysse, J., Córdoba Díaz, D., Peeters, E., Fonteyne, M., Delaet, U., Van Assche, I., De Beer, T., Remon, J.P. and Vervaet, C. (2012). Continuous twin screw granulation: Influence of process variables on granule and tablet quality. European Journal of Pharmaceutics and Biopharmaceutics, [online] 82(1), pp.205–211. doi: https://doi.org/10.1016/j.ejpb.2012.05.010.

Vervaet, C. and Remon, J.P. (2005) “Continuous granulation in the pharmaceutical industry,” Chemical Engineering Science, 60(14), pp. 3949–3957. Available at: https://doi.org/10.1016/J.CES.2005.02.028.

Wang, L. G., Omar, C., Litster, J., Slade, D., Li, J., Salman, A., Bellinghausen, S., Barrasso, D., & Mitchell, N. (2022). Model driven design for integrated twin screw wet granulator and fluid bed dryer via flowsheet modelling. International Journal of Pharmaceutics, 628, 122186. https://doi.org/10.1016/J.IJPHARM.2022.122186.

Wang, L., Pradhan, S.U., Wassgren, C., Barrasso, D., Slade, D. and Litster, J.D. (2020). A breakage kernel for use in population balance modelling of twin-screw granulation. Powder Technology, 363, pp.525–540. doi: https://doi.org/10.1016/j.powtec.2020.01.024.

Wang, L.G., John P. Morrissey b, Barrasso D.,Slade D., Clifford S., Reynolds G., Ooi J.Y ,Litster J.D.Wang, L.G. et al. (2021) “Model driven design for twin screw wet granulation using mechanistic-based population balance model,” International Journal of Pharmaceutics, 607. Available at: https://doi.org/10.1016/j.ijpharm.2021.120939.

Wang, L.G., Omar, C., Litster, J.D., Li, J., Mitchell, N., Bellinghausen, S., Barrasso, D., Salman, A. and Slade, D. (2021). Tableting model assessment of porosity and tensile strength using a continuous wet granulation route. International Journal of Pharmaceutics, 607, p.120934. doi: https://doi.org/10.1016/j.ijpharm.2021.120934.

Wang, Y., et al. (2020). Prediction of tablet tensile strength using artificial neural networks. Powder Technology, 369, 113-123.

Wang, Zilong, (2018) Simulation-based process analysis and optimization - rucore. Available at: https://rucore.libraries.rutgers.edu/rutgers-lib/57755/PDF/1/play/ (Accessed: 04 June 2024).

What is a digital twin? (December. 2020), [Online]. Available: https:/ / www. ibm .com / uk - en / topics/ what - is- a- digital-twin.

Wilson, D., Wren, S. and Reynolds, G. (2011) “Linking Dissolution to Disintegration in Immediate Release Tablets Using Image Analysis and a Population Balance Modelling Approach.” Available at: https://doi.org/10.1007/s11095-011-0535-1.

Xu, L. da, Xu, E.L. and Li, L. (2018) “Industry 4.0: state of the art and future trends,” https://doi.org/10.1080/00207543.2018.1444806, 56(8), pp. 2941–2962. Available at: https://doi.org/10.1080/00207543.2018.1444806.

Xu, Z., Elomri, A., Kerbache, L. and El Omri, A. (2020). Impacts of COVID-19 on Global Supply Chains: Facts and Perspectives. IEEE Engineering Management Review, 48(3), pp.153–166. doi: https://doi.org/10.1109/EMR.2020.3018420.

Yeardley, A.S., Bellinghausen, S., Milton, R.A., Litster, J.D. and Brown, S.F. (2021). Efficient global sensitivity-based model calibration of a high-shear wet granulation process. Chemical Engineering Science, 238, p.116569. doi: https://doi.org/10.1016/j.ces.2021.116569.

Yildiz, E., Møller, C. and Bilberg, A. (2020) ‘Virtual factory: Digital Twin Based Integrated Factory Simulations’, Procedia CIRP, 93, pp. 216–221. doi: 10.1016/j.procir.2020.04.043.

Yoon S.,Galbraith S., Cha B., Liu H. Yoon, S. et al. (2018) “Flowsheet modeling of a continuous direct compression process,” Computer Aided Chemical Engineering, 41, pp. 121–139. Available at: https://doi.org/10.1016/B978-0-444-63963-9.00005-1.

Zhang, E. and Wormr, M. (2019) Improving Sustainability in Pharmaceutical Manufacturing Efficiency in a Pharmaceutical Manufacturing Environment Production Facility.

Zhang, H., Zhang, G. and Yan, Q. (2019) “Digital twin-driven cyber-physical production system towards smart shopfloor,” Journal of Ambient Intelligence and Humanized Computing, 10(11), pp. 4439–4453. Available at: https://doi.org/10.1007/S12652-018-1125-4.

Appendix A: Anomaly detection (Autoencoder Python Code)

```
import numpy as np
import pandas as pd
import tensorflow as tf
from tensorflow import keras
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler

# Generate synthetic data (replace this with your own dataset)
np.random.seed(42)
normal_data = np.random.normal(loc=0, scale=1, size=(1000, 10))
anomaly_data = np.random.normal(loc=5, scale=1, size=(50, 10))

# Create a DataFrame with normal and anomaly data
data = np.vstack([normal_data, anomaly_data])
df = pd.DataFrame(data, columns=[f'feature_{i}' for i in range(data.shape[1])])

# Standardize the data
scaler = StandardScaler()
df_scaled = pd.DataFrame(scaler.fit_transform(df), columns=df.columns)

# Split the data into training and testing sets
X_train, X_test = train_test_split(df_scaled, test_size=0.2, random_state=42)
```

```
# Build the autoencoder model
input_dim = X_train.shape[1]

model = keras.Sequential([
    keras.layers.InputLayer(input_shape=(input_dim,)),
    keras.layers.Dense(units=8, activation='relu'),
    keras.layers.Dense(units=input_dim, activation='linear')
])

model.compile(optimizer='adam', loss='mean_squared_error')

# Train the autoencoder on normal data
model.fit(X_train, X_train, epochs=50, batch_size=32, shuffle=True,
validation_data=(X_test, X_test))

# Use the trained autoencoder to predict on the entire dataset
predictions = model.predict(df_scaled)

# Calculate reconstruction errors (mean squared error between input and output)
mse = np.mean(np.square(df_scaled - predictions), axis=1)

# Set a threshold for anomaly detection (you may adjust this threshold based on your
data)
threshold = 2.0

# Identify anomalies based on the threshold
anomalies = df[mse > threshold]

# Print or visualize the anomalies
print("Anomalies:")
```

Appendix B: Kinetic Parameter TSWG ML Code

```
import seaborn as sns

import tensorflow as tf
from tensorflow import keras
import numpy as np
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D
import pandas as pd
from matplotlib import cm
from sklearn.model_selection import train_test_split

# Loading the data as a pandas dataframe
df2 = pd.read_csv("kinetic predictions.csv")
df2.dropna(inplace=True)
df2.plot(x="mass_fraction_on_screen_1",y=["nucleation"],figsize=(15,6))
plt.xlabel('mass_fraction_on_screen_1 in Micrometres')
plt.ylabel('Pore saturation ')
plt.show()
```

```
X_train = np.concatenate([np.expand_dims(df2.mass_fraction_on_screen_1,axis=-1)
                    np.expand_dims(df2.mass_fraction_on_screen_2,axis=-1),
                    np.expand_dims(df2.mass_fraction_on_screen_3,axis=-1),
                    np.expand_dims(df2.mass_fraction_on_screen_4,axis=-1),
                    np.expand_dims(df2.mass_fraction_on_screen_5,axis=-1),
                    np.expand_dims(df2.mass_fraction_on_screen_6,axis=-1),
                    np.expand_dims(df2.mass_fraction_on_screen_7,axis=-1),
                    np.expand_dims(df2.mass_fraction_on_screen_8,axis=-1),
                    np.expand_dims(df2.mass_fraction_on_screen_9,axis=-1),
                    np.expand_dims(df2.mass_fraction_on_screen_10,axis=-1),
                    np.expand_dims(df2.mass_fraction_on_screen_11,axis=-1),
                    np.expand_dims(df2.mass_fraction_on_screen_12,axis=-1),
                    np.expand_dims(df2.mass_fraction_on_screen_13,axis=-1)
                                ],axis=1)

y_train =np.concatenate([np.expand_dims(df2.nucleation,axis=-1),
                np.expand_dims(df2.breakage,axis=-1),
                 np.expand_dims(df2.Layering,axis=-1)
                 ],axis=1)



model = tf.keras.Sequential(
   [
       tf.keras.layers.Dense(20, activation="elu", name="Hidden_layer_one"),
        tf.keras.layers.Dense(20, activation="elu", name="Hidden_layer_two"),
        tf.keras.layers.Dense(20, activation="elu", name="Hidden_layer_three"),
        tf.keras.layers.Dense(20, activation="elu", name="Hidden_layer_four"),
        tf.keras.layers.Dense(20, activation="elu", name="Hidden_layer_five"),
       tf.keras.layers.Dense(10, activation="elu", name="Hidden_layer_six"),
```

```
model.compile(optimizer='Adam', loss='mse',metrics=['accuracy'])


#model.compile(loss="mse", optimizer=tf.keras.optimizers.SGD(lr=1e-9,
momentum=0.8))
history= model.fit(X_train, y_train, epochs=20000, batch_size=27)
prediction=model.predict(x=X_train)
print(X_train.shape)
print(y_train.shape)
print(prediction.shape)
print(prediction[:,0])
print(prediction[:,1])
print(prediction[:,2])
loss_train = history.history['loss']
epochs = range(1,20001)
plt.plot(epochs, loss_train, 'g', label='Training loss')
plt.title('Training loss')
plt.xlabel('Epochs')
plt.ylabel('Loss')
plt.legend()



plt.show()
plt.plot(df2.mass_fraction_on_screen_1, y_train[:,2], 'g', label='Training')
plt.plot(df2.mass_fraction_on_screen_1, prediction[:,2], 'r', label='Prediction')
plt.show()
from sklearn.metrics import mean_squared_error
import math



MSE = mean_squared_error(y_train[:,2],  prediction[:,2])
```

Appendix C: TSWG ML Code

```
import tensorflow as tf
import numpy as np
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D
import pandas as pd
from matplotlib import cm

#Make Plotly figure
import plotly
import plotly.graph_objs as go

df2 = pd.read_csv("Twinscrewdatamachinelearning.csv")

fig1 = go.Scatter3d(x=df2.Powder_Feed_Rate_kg_per_h,
                    y=df2.Liquid_to_Solid_Ratio,
                    z=df2.Torque_Nm,
                    marker=dict(opacity=0.9,
                                reversescale=True,
                                colorscale='Blues',
                                size=5),
                    line=dict (width=0.02),
                    mode='markers')

#Make Plot.ly Layout
mylayout = go.Layout(scene=dict(xaxis=dict( title="powder feed"),
                                yaxis=dict( title="liquid to solid ratio"),
                                zaxis=dict(title="Torque")),)



#Plot and save html
```

```
model = tf.keras.Sequential(
    [
        tf.keras.layers.Dense(200, activation="elu", name="Hidden_layer_1"),
        tf.keras.layers.Dense(100, activation="elu", name="Hidden_layer_2"),
        tf.keras.layers.Dense(1, name="Hidden_layer_3"),
    ]
)

model.compile(optimizer='Adam', loss='mean_squared_error')
x_train = np.concatenate([np.expand_dims(df2.Powder_Feed_Rate_kg_per_h,axis=-
1),
                          np.expand_dims(df2.Liquid_to_Solid_Ratio,axis=-1)
                  ],axis=1)
y_train = np.expand_dims(df2.Torque_Nm,axis=-1)

model.fit(x=x_train, y=y_train, epochs=10000, batch_size=30, verbose=1)



df2.Powder_Feed_Rate_kg_per_h = np.linspace(0,20,7)
df2.Liquid_to_Solid_Ratio= np.linspace(0.0,0.4,7)

x_test1, x_test2 =
np.meshgrid(df2.Powder_Feed_Rate_kg_per_h,df2.Liquid_to_Solid_Ratio)

x_test = np.concatenate([np.expand_dims(x_test1.flatten(),axis=-1),
                          np.expand_dims(x_test2.flatten(),axis=-1)],axis=1)
y_test = model.predict(x=x_test).reshape([7,7])

fig = plt.figure()
ax = fig.gca(projection='3d')
```

```
ax.scatter(df2.Powder_Feed_Rate_kg_per_h,df2.Liquid_to_Solid_Ratio,df2.D50,line
width=0, antialiased=False,)

ax.set_xlabel('Powder feed rate(kg/h)')

ax.set_ylabel('liquid to solid ratio')

ax.set_zlabel('Torque Nm')

plt.show()


fig2 = plt.figure(figsize = (16, 10))

plt.xlabel('Epoch Number')

plt.ylabel("Loss Magnitude")

plt.plot(history.history['loss'])


plt.show()
```

Appendix D : TSWG EnPI Optimisation Code

```
import numpy as np
import scipy.constants
import scipy.integrate as integrate
from scipy.optimize import minimize, LinearConstraint
ln = np.log
def kval(rpm, F, ls_ratio):
    n1 = -1.13097020 * 10**(-2)
    n2 = +3.27713175 * 10** (-1)
    n3 = +3.42615181 * 10**(+0)
    n4 = +6.17107414 * 10**(-6)
    n5 = -3.71204881 * 10**(-4)
    n6 = +2.32487761 * 10** (-2)
    n7 = +1.84521172 * 10** (-3)
    n8 = -1.53832945 * 10**(-1)
    n9 = -1.42015029 * 10**(+1)
    k = + n1 * rpm
        + n2 * F
        + n3 * ls_ratio
        + n4 * rpm**2
        + n5 * rpm * F
        + n6 * rpm * ls_ratio
        + n7 * F**2
        + n8 * F * ls_ratio
        + n9 * ls_ratio**2
    return k
```

```
def e_of_t(rpm, k, t):
    angular_velocity = (rpm * 2 * scipy.constants.pi) / 60
    energy = integrate.quad(lambda x: angular_velocity * k * ln(x+1), 0, t)
    return energy[0]
def EnPI(rpm, F, ls_ratio):
    k = kval(rpm, F, ls_ratio)
    energy_24hr = e_of_t(rpm, k, 24*60*60) # EnPI for 24hr run
    enpi = energy_24hr / F
    return enpi
def f(x):
    return EnPI(x[0], x[1], x[2])
# ----------- main()
# Initial guess - taken from past DiPP run
x0 = [725, 7.5, 0.4]
k = kval(x0[0], x0[1], x0[2])
print("kval is: " + str(k))
print("EnPI (24hr) is: " + str(f(xO)))
print("\n")
cons1=(
    {'type': 'ineq',
    'fun': lambda x: x[0] - 100}, # rpm - 100 > 0
    {'type': 'ineq',
    'fun': lambda x: 800 - x[0]}, # 800 - rpm > 0
    {'type': 'ineq',
    'fun': lambda x: x[1] - 3}, #F - 3 > 0

    {'type': 'ineq',
    'fun': lambda x: 15 - x[1]}, # 15 - F > 0
    {'type': 'ineq',
    'fun': lambda x: x[2] - 0.1}, # ls - 0.1 > 0
```

```
res = minimize(f, xO, constraints=cons1, method="cobyla")
print(res)
print("\n")
print("the solver results are:")
print(" - rpm: " + str(res.x[0]))
print(" - F: " + str(res.x[1]))
print(" - ls: " + str(res.x[2]))
print(" - k: " + str(res.x[2]))
print(" - EnPI: " + str(f(res.x)))
```

Appendix E : Tablet Press EnPI Optimisation Code

```
from scipy.optimize import minimize
import numpy as np
import matplotlib.pyplot as plt

# Define the objective function to minimize: x
def objective_function(vars):
    y,w,t,f,h,s,e =vars

    return e # Minimize x

def equality_constraint_1(vars):
    y,w,t,f,h,s,e =vars
    return  83.668*h-w



def equality_constraint_2(vars):
    y,w,t,f,h,s,e =vars
    return  ((y*s)/w)-e

def equality_constraint_3(vars):
    y,w,t,f,h,s,e =vars
    return (0.0873*y+0.1708)-t

def equality_constraint_4(vars):
    y,w,t,f,h,s,e =vars

    return  10*f-(t*(3.14*(12*0.001)**2*((2.84*h/12)-((0.126*h)/(h-2*1.21))+(3.15*(h-
1.21)/12)+0.01))*0.000001)
```

```
# Define initial guess and bounds for each variable
initial_guess = [11,600,1.65,13,5,50,76]  # Initial guess for x
bounds = [(2, 20)]  # Bounds for x

# Add bounds for y,w,t,f,h,s,e
bounds.extend([(500, 700), (1.6, 1.7), (0, 20), (0, 20), (20,80 ),(0,None)])

# Define the constraints as a list of dictionaries
constraints = [{'type': 'eq', 'fun': equality_constraint_1},  # Equality constraint 2
               {'type': 'eq', 'fun': equality_constraint_2},  # Equality constraint 2
               {'type': 'eq', 'fun': equality_constraint_3},  # Equality constraint 3
               {'type': 'eq', 'fun': equality_constraint_4}]  # Equality constraint 4

options = {'maxiter': 10000}
objective_values = []
variable_values = []
def callback(xk):
    objective_values.append(objective_function(xk))
    variable_values.append(xk)

# Minimize x subject to equality constraints
result = minimize(objective_function, initial_guess,method='SLSQP',
bounds=bounds,options=options, constraints=constraints,callback=callback)

# Print the optimal solution
print(result)
print("Optimal value of x:", result.x[0])
print("Optimal value of x:", result.x[1])
print("Optimal value of x:", result.x[2])
print("Optimal value of x:", result.x[3])
print("Optimal value of x:", result.x[4])
```

```
# Plot variable trajectories
variable_values = np.array(variable_values)



num_variables = len(initial_guess)
plt.figure(figsize=(15, 70*num_variables))
for i in range(num_variables):
    plt.subplot(num_variables, 1, i+1)
    plt.plot(variable_values[:, i], label=f'Variable {i+1}')
    plt.xlabel('Iteration')
    plt.ylabel(f'Variable {i+1} Value')
    plt.title(f'Variable {i+1} Trajectory')
    plt.legend()
    plt.grid(False)
plt.subplots_adjust(hspace=10)
plt.tight_layout()
plt.show()
```