This is a repository copy of *A sustainability-oriented digital twin of the diamond pilot plant*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/223139/

Version: Published Version

Article:

Ntamo, D., Papadopoulos, I., Omar, C. orcid.org/0000-0002-7839-608X et al. (2 more authors) (2025) A sustainability-oriented digital twin of the diamond pilot plant. Processes, 13 (1). 211. ISSN 2227-9717

https://doi.org/10.3390/pr13010211

Article

A Sustainability-Oriented Digital Twin of the Diamond Pilot Plant

Donald Ntamo, Iason Papadopoulos, Chalak Omar [ID], Payam Soulatiantork [ID] and Mohammad Zandi *[ID]

School of Chemical, Materials and Biological Engineering, The University of Sheffield, Sheffield S1 3JD, UK; dntamo1@sheffield.ac.uk (D.N.); c.omar@sheffield.ac.uk (C.O.); p.soulatiantork@sheffield.ac.uk (P.S.)
* Correspondence: m.zandi@sheffield.ac.uk

Abstract: The pharmaceutical industry is undergoing a significant transition from batch to continuous manufacturing, driven by increasing regulatory requirements and sustainability pressures. Digital twins (DTs) play a pivotal role in facilitating this transition by enabling real-time data visualisation, process optimisation, and predictive analytics. While substantial progress has been made in the development and application of DTs, particularly in industries such as energy and automotive, there remains a critical need for further research and development focused on creating sustainability-oriented digital twins tailored to pharmaceutical processes. In the pharmaceutical sector, DTs are being progressively utilised not only for real-time monitoring and analysis but also as dynamic training platforms for engineers and operators, enhancing both operational efficiency and workforce competency. This paper examines the University of Sheffield's Diamond Pilot Plant (DiPP), a facility showcasing the future of pharmaceutical manufacturing through the integration of Industry 4.0 technologies and advanced sensors. 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 contributes to the field of pharmaceutical manufacturing by providing a foundation for creating energy models and advancing the development of comprehensive DT.

Keywords: twin-screw granulator; energy usage; continuous manufacturing; digital twin; machine learning; Industry 4.0; sustainability; data-driven models; mechanistic models

1. Introduction

In recent years, the pharmaceutical industry has experienced pressure from pharmaceutical regulators, primarily the US FDA and the EU EMA, to implement Quality-by-Design (QbD) principles in its manufacturing processes. The EMA defines Quality-by-Design as follows: "An approach that aims to ensure the quality of medicines by employing statistical, analytical, and risk-management methodology in the design, development, and manufacturing of medicines" [1]. Furthermore, the pharmaceutical industry as well as all other process manufacturing industries have experienced additional pressures to consider the sustainability of their manufacturing processes more carefully [2]. The sustainability of a manufacturing process can be measured in various ways, such as water usage, hazardous waste produced, energy consumption, and others. The energy consumption of processes is of particular importance, as the price of energy has increased rapidly due to geopolitical

instability in Europe, and because governments have started legislating more strictly due to climate change and emissions concerns [3,4]. To address these concerns, the pharmaceutical industry is looking at a transition away from its tried-and-tested batch manufacturing methods and towards continuous manufacturing. Continuous manufacturing in pharma offers real-time data acquisition, enhanced process efficiency, and improved product consistency due to less batch-to-batch variability [5–7]. 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, and 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. The transition to continuous manufacturing would not only help to address the main concerns about sustainability, but it would also alleviate problems that arise during the scale-up stage from the lab towards large-scale manufacturing and issues regarding batch-to-batch variability in the CQAs and KPIs of medicines, as well as the particularly great need for human intervention between batch-to-batch manufacturing, hence also reducing the possibility of human error. Continuous manufacturing would also allow for a quicker time-to-market due to the lack of a need for a scale-up stage. Continuous manufacturing would alleviate all of the above concerns and would also allow for a quicker time-to-market due to the lack of a need for a scale-up stage. This is important for increasing patient safety, as not only would a lesser batch-to-batch variability in CQAs and KPIs mean more consistent medicines, but it would also allow quicker development of new treatments for illnesses [8]. More importantly, continuous manufacturing allows for the use of continuous unit operations, which are generally more efficient than their batch counterparts. However, it was also noted that continuous unit operations have generally higher capital costs, which means payback time would have to be taken into consideration for a business's transition [9].

The development of digital twins has the potential to play a key role in facilitating the adoption of continuous manufacturing within the pharmaceutical industry. Building on successes in energy and automotive sectors, digital twins are now making waves in pharmaceuticals, enabling real-time process monitoring, optimisation through data analysis and visualisation, and training for engineers and operators by modelling process parameter variations and KPIs [10–12]. With regards to the application of digital twins for sustainability analysis, DTs are useful in that they can collect data in real time [13]. The data can then be processed and visualised offline, which means various simulated runs can be executed, allowing engineers to model and optimise for the parameters that yield the most efficient manufacturing process while still producing granules of the right quality for pharmaceuticals. The connection between the digital twin's computer model and the real-world plant also means that model predictive control (MPC) can take place and perform real-time optimisation of the process, allowing for live sustainability optimisation [10–12,14]. This will allow engineers to model a real-world manufacturing process using computer simulations and to monitor how the Critical Quality Attributes (CQAs) react to changes in the input parameters to the process without the need to run costly and laborious experiments. More importantly, it allows us to model the sustainability of the manufacturing processes. The University of Sheffield's Diamond Pilot Plant (DiPP) is designed to be a world-leading Industry 4.0 demonstrator [15,16]. This means that it is equipped with the necessary sensors and live data processing to demonstrate a fully digitalised manufacturing process [17]. The DiPP's Industry 4.0 Technologies (I4.0T) make it the perfect facility to be modelled by a DT, as the live data can be used to perform real-time optimisation of the DT [18,19]. This work aims to take advantage of this technology and develop a sustainable digital twin.

Critical evaluation is of great importance when examining new Industry 4.0 technologies. This is due to the potential for commercial bias inherent in research conducted by companies developing and promoting these technologies. Their focus may naturally be on highlighting the benefits and downplaying the potential limitations. By critically evaluating research, one can ensure a more balanced understanding of Industry 4.0 solutions and their true value proposition for different stakeholders [18]. Consequently, a comprehensive study is crucial to demonstrate the value proposition of continuous production to pharmaceutical companies, particularly in the context of Industry 4.0 technologies [20].

Continuous pharmaceutical manufacturing utilises twin-screw granulators to produce granulated solid particles for subsequent tableting. This process is currently implemented in the DiPP (Diamond Pilot Plant) Consigma-25 pilot plant, a dedicated research facility. By integrating a DT with the physical system, real-time control over process parameters can be achieved, leading to enhanced product sustainability through reduced resource waste. This research aims to address a critical gap in the literature by investigating how big data, specifically concerning twin-screw energy consumption, can be effectively leveraged using the granulator's DT technology. In this research, the DiPP TSG's energy usage is predicted over time, a crucial step towards energy optimisation in next-generation pharmaceutical manufacturing. gPROMS, a computational modelling software, acts as a repository for mechanistic models, serving as a virtual representation of the plant. However, the existing twin-screw granulator model within gPROMS lacks the capability to predict energy consumption. This work enhances the existing gPROMS FormulatedProducts mechanistic model for a TSG by integrating a torque model with it, a feature that is currently absent. The data-driven model presented in this paper serves as a foundation for developing robust mechanistic torque and energy models and providing a valuable platform for validating future mechanistic simulations. This integration will improve the accuracy of the gPROMS TSG model, enabling more realistic simulations. This research will have significant implications for the pharmaceutical industry, facilitating the transition to more efficient and sustainable continuous manufacturing processes.

This paper described the development process of an advanced digital twin of the twin-screw granulator at the Diamond Pilot Plant (DiPP), capable of modelling and optimising the TSG's energy usage. The paper is structured as follows. Firstly, a description of DiPP's key powder process, twin-screw wet granulation is provided. Secondly, a data-driven model of the DiPP's twin-screw granulator model, capable of modelling the energy usage of the TSG unit is presented and evaluated. Following an examination of the current state of digital architecture in pharmaceuticals, this paper will explore the limitations associated with developing and implementing the sustainable digital twin. Finally, the paper concludes by discussing future directions and the overall impact of this technology.

2. Material and Methods

The creation of a digital twin for any manufacturing process is a huge endeavour. Developing a comprehensive digital twin for the ConsiGma-25 (ConsiGma^{TM}, GEA, Kontich, Belgium) system (at the Diamond Pilot Plant would necessitate significant additional resources and extended development time to achieve full implementation [15]. To reduce the vast scope that digital twin creation encompasses, some restrictions must be applied to reduce the problem domain. Sustainability assessment considers a range of environmental factors, including water use, energy consumption, and waste generation [21]. This paper's focus is to analyse the energy usage of the Diamond Pilot Plant. Developing a digital twin solely focused on the DiPP's energy usage would be a significant undertaking. To make the analysis more manageable, this paper will narrow the scope by selecting a single unit operation within the DiPP process for in-depth energy analysis. The DiPP consists of many

unit operations, such as the MODUL P tablet pressing machine, the twin-screw granulator, and the fluidised bed dryer [22–24]. This paper focuses its energy analysis solely on the twin-screw granulator within the Diamond Pilot Plant. 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 our analysis [25–29].

Granulation is a fundamental process in the pharmaceutical industry when it comes to the manufacturing of tablets, which is the reason why a lot of research into granulation has taken place. It is the process of enlargement of dry particles via agglomeration techniques and liquid addition. Granulation is a critical unit operation in the pharmaceutical industry since it produces the necessary wet granules which can then be pressed into medicine tablets for patients [30]. Ever since the introduction of Quality-by-Design by regulators, the pharmaceutical industry has started transitioning away from batch granulation and towards continuous granulation, and the way in which this has been achieved is with twin-screw granulation [28].

Figure 1 shows a diagram of a TSG unit, with powder (i.e., dry granule) feeds, liquid addition feeds, a temperature-control jacket, and conveying and kneading elements. The fundamental design principles of the TSG are the addition of dry granules and liquid lubricant through the feed ports, the conveying of the materials through the conveying stages of the screws, the kneading elements that mix the liquid and the solid, and finally the conveying out of the wet granules through to the outlet port.
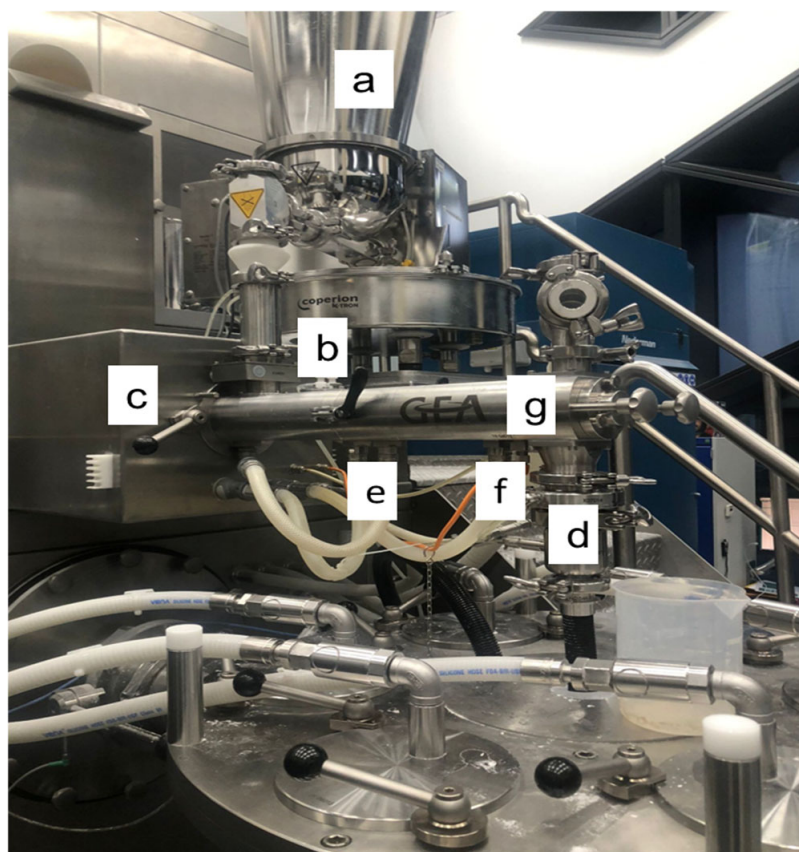

Figure 1. TSG unit at the Diamond Pilot Plant: (a) powder feed; (b) liquid feed; (c) motor unit; (d) granule outlet; (e) temperature-control inlet; (f) temperature-control outlet; (g) TSG barrel.

To create a digital twin (DT) of the twin-screw granulator (TSG) for energy modelling, this paper will first investigate the TSG's operation and identify critical parameters in-

fluencing its energy consumption. This understanding will inform the development of a hybrid DT model. This model will combine data-driven and mechanistic approaches to not only predict TSG energy usage but also perform optimisation analyses, ultimately minimising energy consumption.

A data-driven model for energy usage estimation is proposed in this paper. This approach is chosen due to the mathematical complexity and computational burden associated with traditional Population Balance Models (PBMs) for TSG simulation. The DT's ability to simulate the TSG allows for experimentation through Design of Experiments (DoE). By varying input parameters within the DoE framework, one can effectively model the relationship between these parameters and the TSG's energy usage.

3. Results and Discussions

3.1. Energy Balance on the Twin-Screw Granulator

Before any modelling can take place, an energy balance must be performed around the TSG, which involves listing all input and output energies and enthalpies. Combining all the stream flow rates (F) and energies into an energy balance equation, the following expression can be derived:

$$F_1 h_1 + F_2 h_2 + F_3 h_3 + Q_1 \; = \; F_4 h_{24} + Q_2 \tag{1}$$

An assumption will be made that the enthalpies (h) from the input streams are equal to the enthalpy of the output stream, and no energy is transferred from the granule and liquid streams into the barrel itself. This means that the energy that is provided to the system by the motor is equal to the energy that is dissipated as heat into the barrel of the TSG. Therefore, to model the energy usage of the TSG, it is enough to model the energy usage of the motor only, as it will account for most of the energy being transferred into the system.

3.1.1. Energy Usage of the Motor

Accurate calculation of motor energy consumption requires the integration of instantaneous power measurements over a specific period (t) using Equation (2). This is necessary because motors typically exhibit varying power demands throughout operation:

$$E(t) \; = \; \int_0^{-t} P\,(t)\,dt \tag{2}$$

where P is the power and E is the energy usage. To build the $P\,(t)$ curve, one can either collect enough electrical power measurements through direct measurement or use Equation (3) to calculate the instantaneous power for a given torque, with τ being the torque of the motor, and ω being the speed of the motor.

$$P \; = \; \omega\tau \tag{3}$$

The speed of the motor remains constant throughout a TSG run, while the torque varies with time. To model the power usage of the motor based on the TSG input parameters, the twin-screw granulator inputs need to be linked to a $\tau\,(t)$ curve.

3.1.2. Torque Profile Curve

The torque of the TSG's motor does not stay constant. Rather, it gradually increases with time, and follows a logarithmic-type curve.

To understand the relationship between TSG inputs and torque, using average or rolling average techniques on the torque profile was avoided. Instead, a method that captures the dynamic variations within the curve was employed in this paper. Taking

the average would lead to a loss of information from this curve. The torque curve was modelled as a $y = k \times ln\,(x + 1)$-type function, where k is a constant factor. The model should correlate the input parameters with the k-values.

It is crucial to note that continuous pharmaceutical processes do not run for a few seconds. The curve shown in Figure 2 originates from an undergraduate lab experiment with the TSG, where time limitations were restrictive, which is why the x-axis only stretches up to approximately 350 s. A real-world continuous manufacturing process would run for multiple days. The limitation of the $y = k \times ln\,(x + 1)$ model occurs when plugging in large t values. This is because the model predicts unreasonably large torque values with increasing t, which becomes unrealistic for practical applications. Therefore, caution is advised when using this model for extrapolation, particularly at high t values.
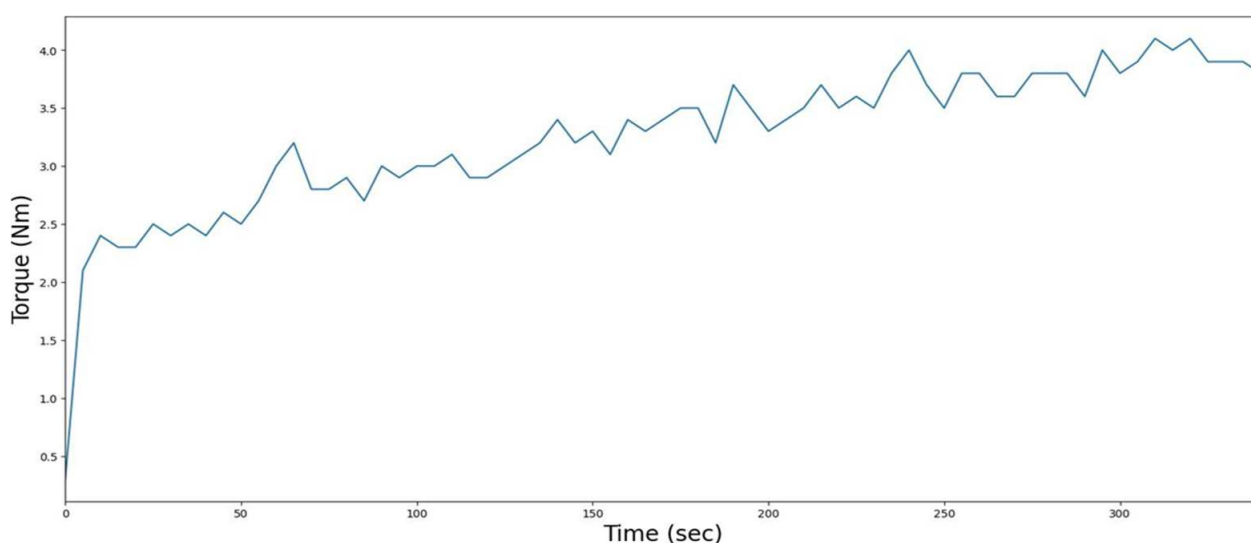

Figure 2. Example of a TSG torque profile curve for the DiPP (at $t = 0$, there is no production and the DiPP is at stoppage).

As shown on the ConsiGma-25 HMI (ConsiGma^{TM}, GEA, Kontich, Belgium), the TSG has four key input parameters that can be set. These are as follows:

- Speed (rpm) of the motor;
- Powder feed rate (kg.h^{-1});
- Liquid feed rate (g.min^{-1}) of the first liquid port;
- Liquid feed rate (g.min^{-1}) of the second liquid port.

By adding together the two liquid flow rates and by dividing the liquid flow rate by the solid flow rate (and converting units appropriately), the information can be condensed down to a singular liquid-to-solid ratio, called the L/S ratio. Hence, Table 1 of the input parameters is as follows:

Table 1. Table of TSG input parameters.

Parameter	Symbol	Unit
Speed	ω	rpm
Powder flow (feed) rate	F	kg.h^{-1}
Liquid-to-solid ratio	λ	kg.kg^{-1}

The energy modelling problem is summarised in the following diagram (Figure 3):

$$E(t) = \int_0^t P(t)\, dt = \int_0^t \omega \times \tau(t)\, dt = \int_0^t \omega \times [\underline{k} \times ln(t+1)]\, dt$$
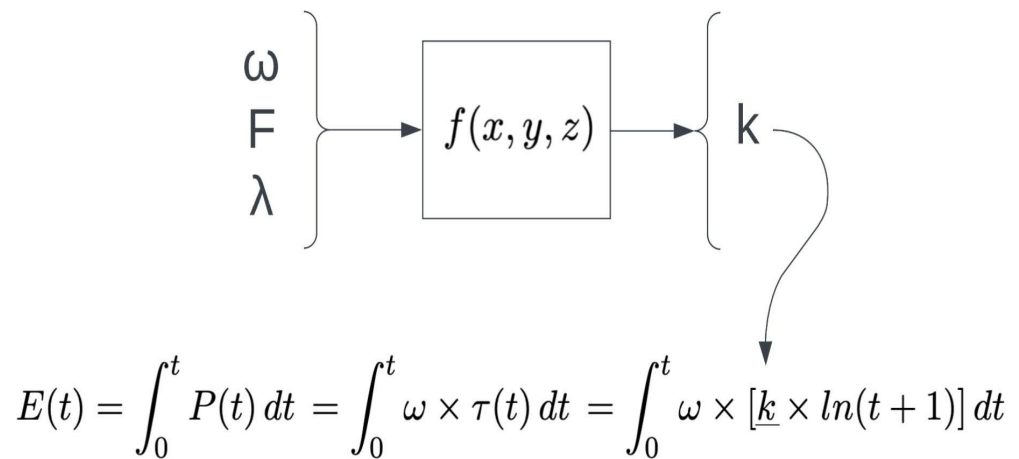
Figure 3. Summary of proposed model.

The three key input parameters (speed, feed rate, and L/S ratio), are plugged into the model to calculate the *k*-value corresponding to the equivalent torque profile curve. The *k*-value is then used in conjunction with the motor's speed to calculate the energy usage for a given time. The speed in *rpm* was converted to an angular velocity value measured in rad.s^{-1}, with the below conversion:

$$\omega(rad.s^{-1}) = \omega(rpm) \times \frac{2\pi}{60} \tag{4}$$

3.2. Data Analysis with Python

As shown in Figure 3, the key objective for the data-driven energy model is formulating a function of *f* (*x*, *y*, *z*) that will relate ω, *F*, and λ with *k*. To relate these variables, past campaign run data of the DiPP were collected and processed. Whenever the DiPP runs, it uses its sensors to collect measurements in real time, and it exports these values in the form of Microsoft Excel spreadsheets (Microsoft 365 Apps, Version 2411). A total of ten campaign spreadsheets from past DiPP runs were used. These runs were from labs conducting undergraduate teaching, as well as from test runs of the DiPP by the technicians responsible for maintaining the DiPP.

For data analysis, Python (version 3.13.1) was employed, utilising libraries such as pandas, SciPy, scikit-learn, and matplotlib to perform all necessary analyses and generate the required plots. The Python energy model code in Python is structured as follows:

- Firstly, importing Excel spreadsheets from past DiPP runs ("campaigns") into Python;
- Then, dropping unnecessary columns (e.g., measurements from unit operations that are not needed during analysis);
- Concatenating all relevant data into a single large dataframe;
- Iterating through the large dataframe and filtering out all periods of TSG inactivity;
- Identifying all periods of TSG activity and slicing up the large dataframe into "slices" of TSG activity containing all relevant data points;
- Taking the τ (*t*) curve from each slice, and performing *y* = *k* × *ln* (*x* + 1) regression on every torque profile curve;
- Dropping all slices where their τ (*t*) curve does not have a satisfactory fit onto a logarithmic regression model;
- Finally, collecting all (ω, *F*, λ) and (*k*) pairs for regression analyses.

This energy model structure is summarised in Figure 4. In the next section, machine learning was used to relate these values to determine *f* (ω, *F*, λ).
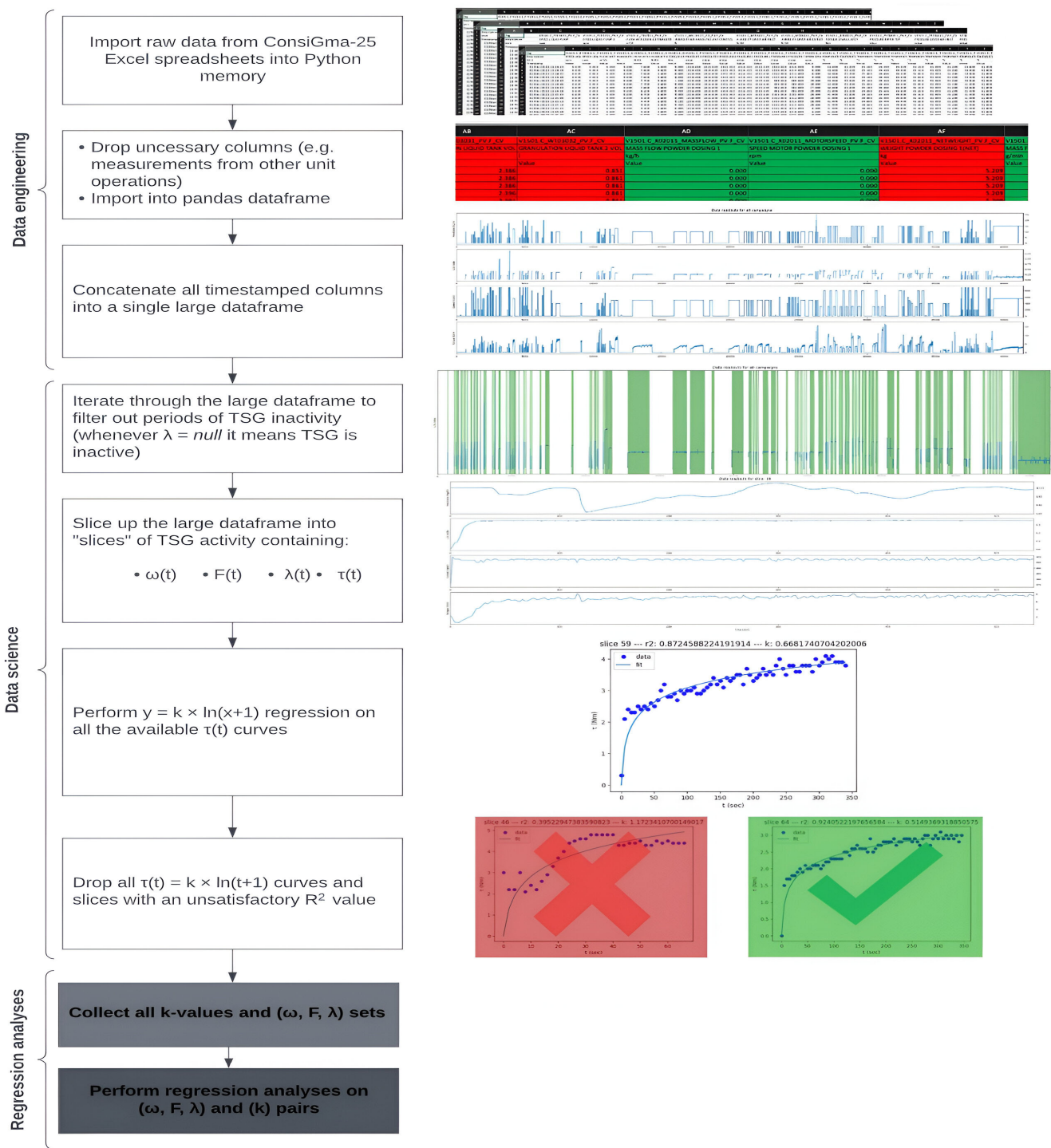
Figure 4. TSG energy modelling workflow in Python: It begins with importing and cleaning campaign data from Excel spreadsheets, removing unnecessary measurements and combining relevant information into a single data frame. Next, the code segments the data by identifying and extracting "slices" representing TSG activity periods. From each slice, the τ(*t*) curve is extracted and analysed using a $y = k \times ln\,(x + 1)$ regression. Slices where the regression fit is unsatisfactory are discarded, highlighted in red. Finally, the remaining data containing (*ω*, *F*, *λ*) and (*k*) pairs are collected for further regression analyses used in TSG energy modelling.

3.2.1. Machine-Learning Regression with Python

The Python library SciKit-Learn was used perform machine-learning regression using a collection of refined slices of data. It is important to note that when performing regression

analysis, a common fault is the overfitting of curves, where all the training data are inputted into the regression solver, thereby resulting in the solver overfitting the inputs and producing a model that may be very good at predicting its own training dataset values, but poor at making correct predictions (in this case, torque predictions). Sklearn provides a very useful function in the code that allows us to split the training dataset into "training" and "testing" sections.

After splitting the dataset appropriately (taking sklearn's recommended values of a 50/50 split for training and testing data), the data were inputted into various regression solvers provided by sklearn as illustrated in Figure 5. The goal was to test a variety of regression models to see which one would provide the best fit for the data. The full list of regression models used is shown in Figure 4, and the results of the regression analysis including RMSE (root-mean-square deviation) and R^2 (R-squared) values are shown in Table 2.

Table 2. Regression analyses results.

Regression 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 [31]	0.93364	0.15756	0.71664	0.35835
Support vector regression (SVR) with RBF kernel [32]	-0.11	0.64455	-0.097	0.70536
K-Nearest-Neighbour [33]	1.0	0.0	-0.29443	0.76591
Neural network [34]	0.95256	0.13321	0.25159	0.58238

As can be seen from Table 2, the best regression model produced was a second-degree polynomial regression. It was expected that a higher-degree polynomial would result in a better fit; however, that turned out to not be the case.

Some more advanced machine-learning techniques were also tested. Random forest, K-Nearest-Neighbour, and SVR with an RBF (radial basis function) kernel were used in the hope that the more sophisticated techniques would result in a better regression model. Random forest produced quite good results; however, the other models proved to be worse than simple linear regression based on their very high RMSE values. While the neural network was able to train well on the data, it was unfortunately very poor at making predictions on the testing data, being beaten by the linear regression model.

Defining a neural network layout requires large expertise in artificial intelligence and neural network design, and it is something that would require input from people with computer science expertise. Since the second-degree polynomial provides sufficient results, it is deemed that the application of neural networks is not necessary, and the design of neural networks is outside the scope of this research.

3.2.2. Final Energy Model

The best model was chosen to proceed with the digital twin creation, and this was the 2nd-degree polynomial regression. A raw mathematical equation (Equation (5)) and its coefficients (Table 3) were exported out of the Python environment, and the result are as follows:

$$k = n_1\,\omega + n_2\,F + n_3\,\lambda + n_4\omega^2 + n_5\,\omega F + n_6\,\omega\lambda + n_7\,F^2 + n_8\,F\lambda + n_9\lambda^2 + C \quad (5)$$
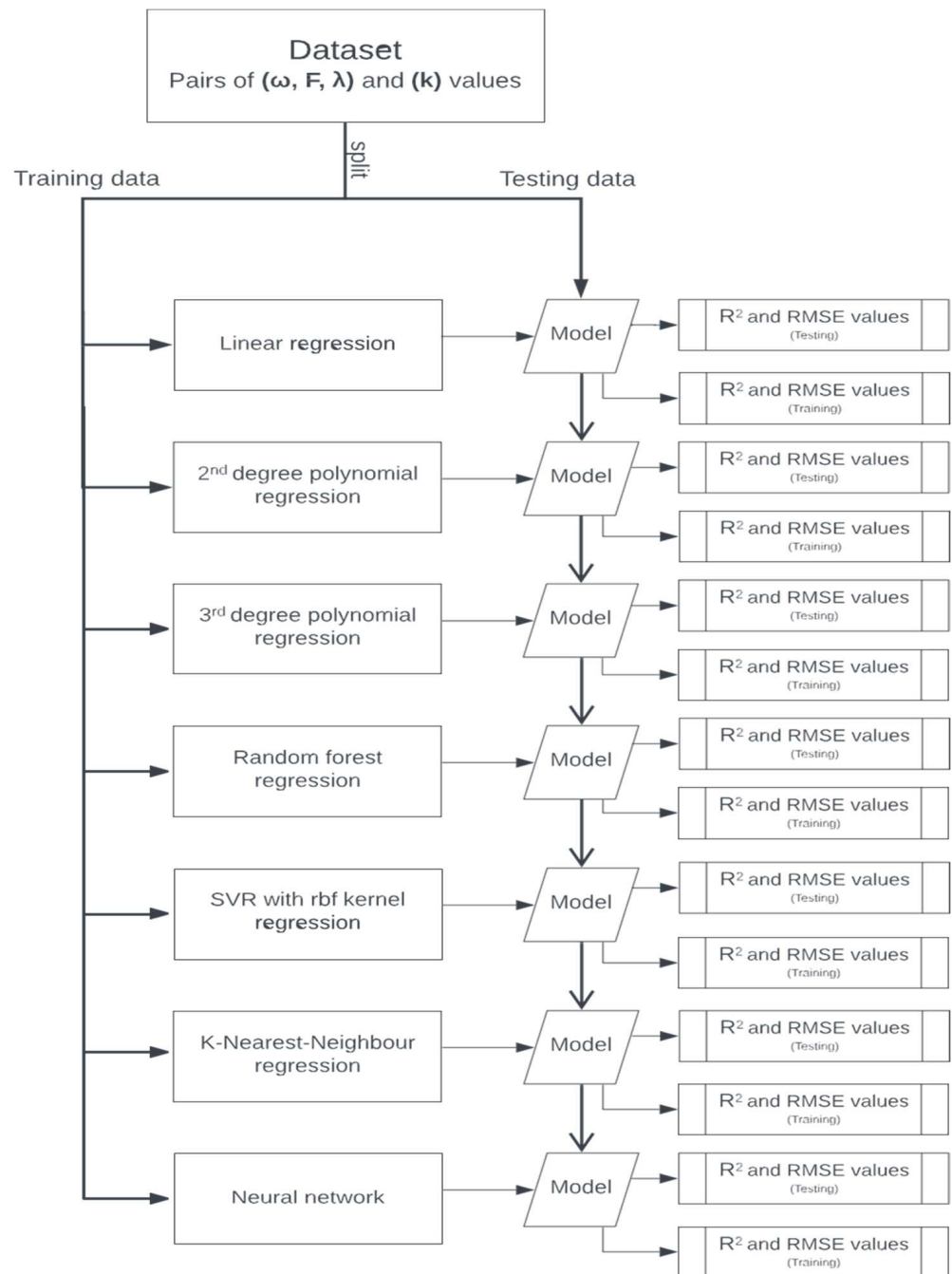
Figure 5. Regression training logic: TSG energy modelling utilises a regression-based approach to analyse torque profiles. To optimise the model, a regression analysis employing the equation $y = k \times ln\,(x + 1)$ is performed on each individual curve. This step is followed by the evaluation of various regression models, including support vector regression (SVR) with a radial basis function (RBF) kernel, a neural network, and a polynomial regression. By comparing their performance, the model that achieves the most accurate fit to the torque profile data is chosen, ultimately leading to more accurate TSG energy modelling.

k-values can hence be calculated, and then used as shown in the equation in Figure 3. It is important to note that this model cannot be extrapolated for all values of ω, F, and λ. This model is only valid for the range it has been trained and tested on, which is as follows:

$$\omega \in [100\text{ rpm}, 900\text{ rpm}] \qquad F \in [5\text{ kgh}^{-1}, 20\text{ kgh}^{-1}] \qquad \lambda \in [0.15, 0.4]$$

Table 3. 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	$7.83459685 \times 10^{-1}$

3.3. Optimisation Testing with SciPy

To demonstrate the capability of the energy model, the optimisation of energy performance indicator (EnPI) values was performed using our model and SciPy. First, an initial guess for the SciPy minimisation solver based off prior standard DiPP values was defined and then input into the gPROMS FormulatedProducts (Version 2023.2.0.55304) flowsheet to obtain an estimate of the PSD. Then, minimisation of the EnPI value was performed using SciPy's minimise function and the solver's results were copied into the gPROMS flowsheet to compare the new PSD.

The EnPI value being optimised was as follows:

$$EnPI \; = \; \frac{E(24\text{h})}{F} \tag{6}$$

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.

An initial guess for the input parameters as follows:

$\omega = 725$ rpm $\qquad\qquad F = 7.5 \text{ kgh}^{-1} \qquad\qquad \lambda = 0.4$

These values were plugged into SciPy, and the results were as follows (Table 4):

Table 4. Minimisation of EnPI results.

	Before Optimisation	After Optimisation
ω	725	724.73
F	7.5	8.607
λ	0.4	0.291
EnPI	8.76×10^6	1.390

3.4. gPROMS FormulatedProducts Mechanistic Model

To complete the hybrid digital twin, the next step is creating a mechanistic model of the TSG with the help of gPROMS FormulatedProducts (Version 2023.2.0.55304). This is a modelling software within the Siemens gPROMS suite specifically designed for modelling the pharmaceutical industry manufacturing process. It contains various models that aid in modelling and optimising pharmaceutical processes, including fluidised bed dryers, dry mills, and various types of granulators, including twin-screw granulation. Furthermore, it provides optimisation functionalities allowing for the minimisation of objective functions, which will be useful in the next section.

With the help of FormulatedProducts and the example project files included in the installation directory, a flowsheet of the process was developed (Figure 6). This flowsheet can calculate the particle size distribution (PSD), and more importantly, thanks to the PSD sensor unit included in the flowsheet, one can determine the median particle diameter d50 value of the wet granules, which is used to describe the lognormal distribution of the granules.
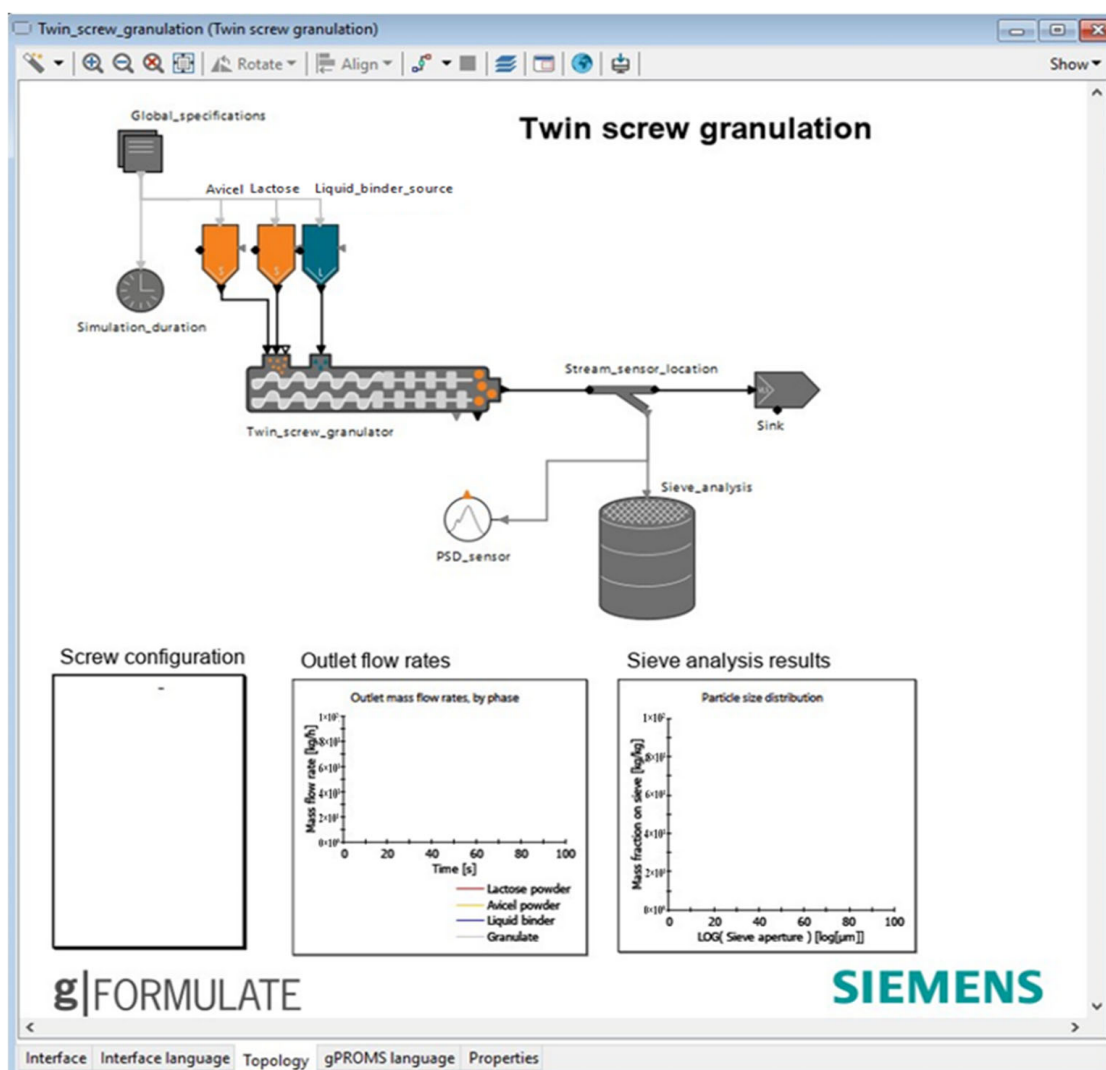

Figure 6. gPROMS FormulatedProducts (Version 2023.2.0.55304) flowsheet.

3.5. Integration of Python and FormulatedProducts Models

The goal of this digital twin 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. The full integration between the Python model and the gPROMS model was not achieved in this paper. The System Programmer Guide documentation indicates the possibility of connecting external models through the Foreign Process Interface (FPI). However, this is restricted to only FORTRAN and C/C++ foreign models, requiring that the user write FORTRAN/C/C++ source code and compile it to an appropriate .dll library (for Microsoft Windows). For the Python program to be correctly interfaced, it would be required that a C "shim" layer be added on top of the Python code, so that the Python program could call the gPROMS model to calculate the d_{50} value and optimise using that value as its boundary condition. Due to time and resource constraints, this was not achieved in this paper.

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 Figures 7 and 8.
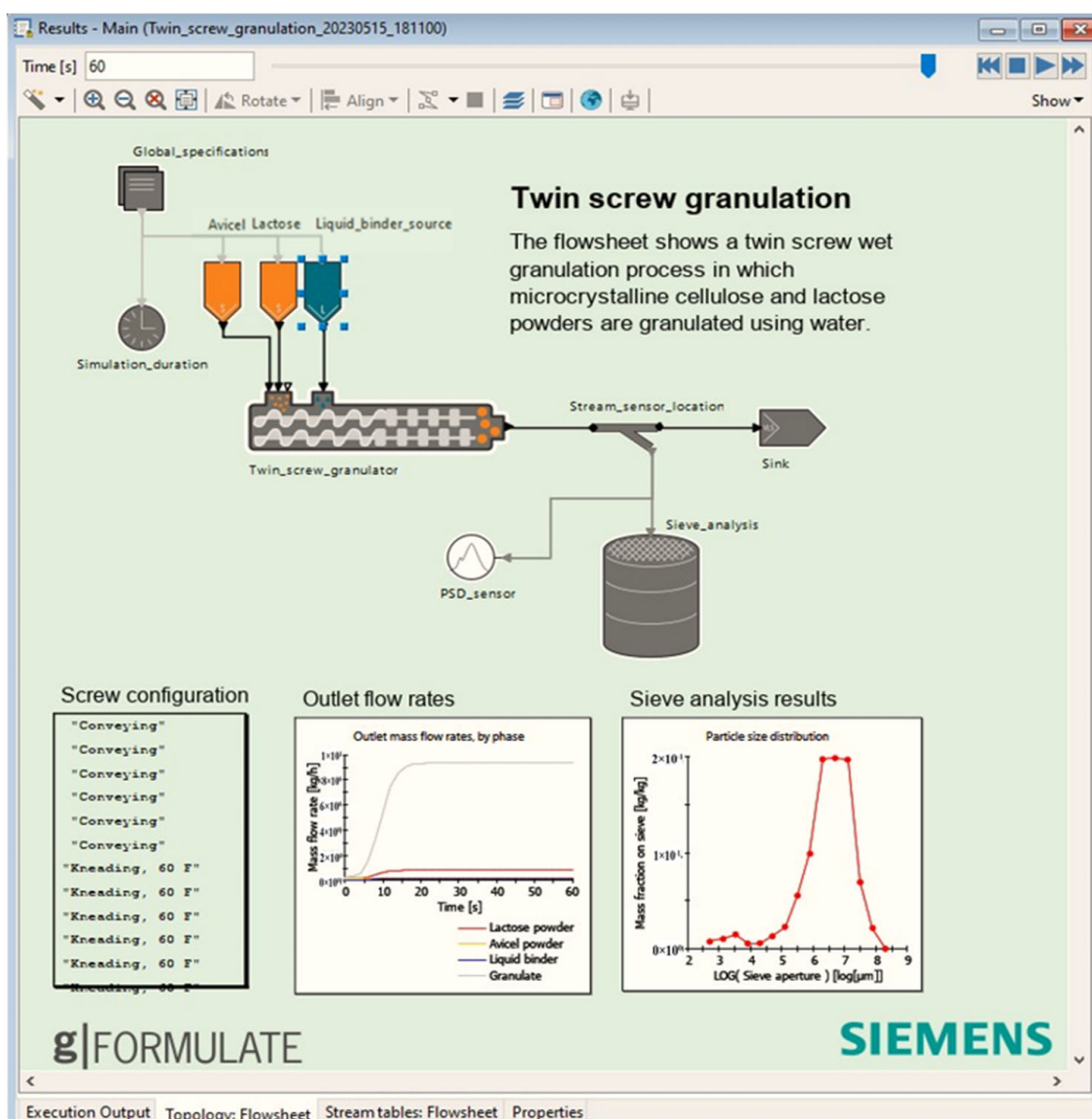

Figure 7. gPROMS PSD calculation of pre-optimisation parameters.

As illustrated in Figures 7 and 8, 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.
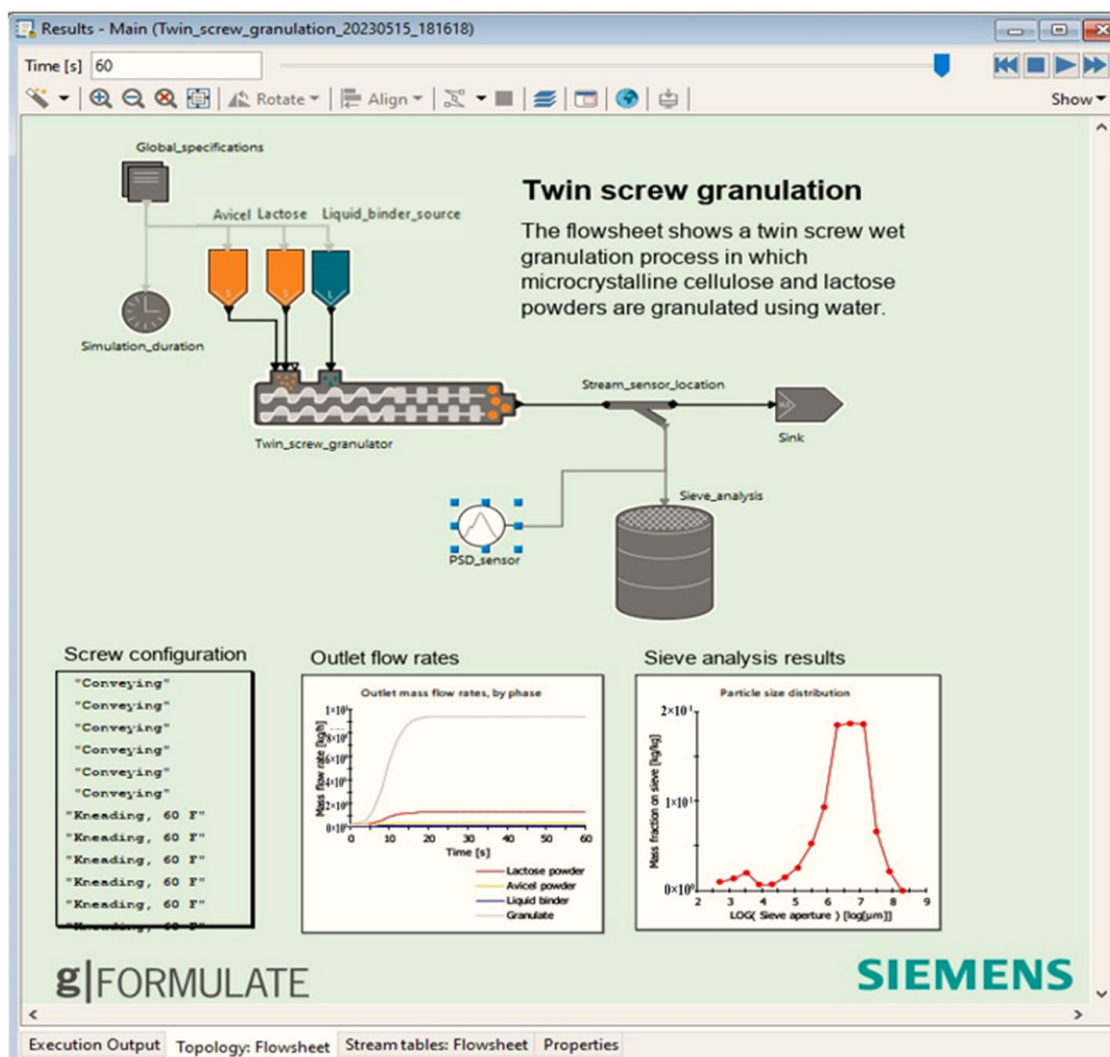
Figure 8. gPROMS PSD calculation of post-optimisation parameters.

4. Future Work

Despite the inability to achieve fully automated integration, this paper was able to establish a workflow for data exchange between the Python and gPROMS models. The documentation available by Siemens for gPROMS FormulatedProducts was very limited, and the software itself was very restrictive, with the TSG model being a heavily protected entity due to commercial interests. This paper successfully optimises the EnPI (24 h) value using d50 as the boundary condition. Future work in this domain should focus on optimisation with respect to additional parameters, such as the barrel temperature and the moisture content of the granules. An optimisation model that can minimise EnPI (24 h) with d50, temperature, and moisture content as its boundary conditions will be a great improvement to the model that has been set out in this paper.

Finally, one of the key aspects of digital twins is their ability to learn and improve in real time from the corresponding real-world process. The energy model currently relies on a manual process of exporting ConsiGma-25 data files in CSV format and storing them in a specific folder ("consigma25_spreadsheets"). This approach is prone to errors and lacks automation. For the digital twin to be compliant with the ISO definition of a digital twin, it would have to continuously learn and improve on its own without requiring human input [35,36]. Another area of future work is fully integrating our digital twin with the Diamond Pilot Plant so that it can automatically retrieve campaign data and learn from

it and improve its models. Such a feature would be extremely useful as it would allow for our model to learn from longer "slices" of TSG activity, as the current model was only trained on shorter "slices" in the range of 50 to 400 s, while real-world processes run for many hours if not days continuously.

This paper successfully performed manual optimisation of the *k*-value energy model using SciPy and compared the pre- and post-optimisation PSD values by calculating them with gPROMS. This resulted in a significant reduction in the EnPI (24 h) value while maintaining the granules' PSD and the d50 value. It is important to note that finding optimal parameters to model the ConsiGma-25's TSG unit will be important research for the future. Detailed research of the components and parameters that make up the mechanistic model are considered out of the scope of this paper, which is why they have not been further researched.

5. Conclusions

The results above show a very accurate data-driven model for predicting the ConsiGma-25's torque profile curve, which can further be used for predicting the TSG's energy usage over time. This model is useful for the pharmaceutical industry as there is little literature on the energy usage of twin-screw granulation. Furthermore, the FormulatedProducts model allows for the prediction of the particle size distribution of the wet granules, which will be useful for future optimisation analysis of the TSG's energy usage. Utilising the SciPy optimisation solver, this paper achieved a significant 100% reduction in the TSG's EnPI (24 h) value. Subsequently, gPROMS FormulatedProducts was employed to verify that this minimisation did not impact the product size distribution (PSD). The analysis revealed a minimal change of only + 3 μm in the d50 value.

The analysis that was performed on prior campaign data of the DiPP resulted in a high-accuracy model, capable of modelling the ConsiGma-25's twin-screw granulator torque profile curve. This paper was also able to use our *k*-value model for predicting the TSG's energy usage over time, which is important for research as it provides a first avenue into energy optimisation for the next generation of pharmaceutical manufacturing. The torque model is also useful in that it can improve on the already existing gPROMS FormulatedProducts model. The current gPROMS TSG model is unable to model either torque or energy, and the work in this paper will help in making the gPROMS twin-screw granulator model better in the future. Future engineers and developers can use this data-driven model as a basis for the development of a robust mechanistic torque and energy model, and it can also be used as a validation platform for future mechanistic models. This will be extremely useful for the pharmaceutical industry, and it will aid greatly in the transition to continuous pharmaceutical manufacturing.

Author Contributions: Methodology, C.O.; Investigation, I.P.; Writing—original draft, D.N.; Supervision, P.S. and M.Z. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

Nomenclature

Acronym	Definition
DT	Digital Twin
QbD	Quality-by-Design
DiPP	Diamond Pilot Plant
IoT	Internet of Things
I4.0	Industry 4.0
I4.0T	Industry 4.0 Technologies
ML	Machine Learning
KPI	Key Point Indicators
CQA	Critical Quality Attributes
FDA	Food and Drug Administration
EMA	European Medicines Agency
PAT	Process Analytical Technology
CM	Continuous Manufacturing
BM	Batch Manufacturing
PSD	Particle Size Distribution
FPM	First Principle Model
HMI	Human–Machine Interface
RMSE	Root-Mean-Square Deviation
API	Active Pharmaceutical Ingredient

Symbols

Symbol	Unit	Parameter
ω	rpm	Speed
F	kg/h	Feed rate
λ	-	Liquid-to-solid ratio
Q	J	Energy
t	s	Time
τ	Nm	Torque
P	W	Power
h	kJ/kg	Enthalpy
d_{50}	μm	Median particle diameter
$EnPI$	kWh^2/kg	Energy performance indicator

References

1. EMA. Quality by Design. May 2020. Available online: https://www.ema.europa.eu/en/human-regulatory/research-development/quality-design (accessed on 7 January 2023).
2. Herrmann, C.; Schmidt, C.; Kurle, D.; Blume, S.; Thiede, S. Sustainability in manufacturing and factories of the future. *Int. J. Precis. Eng. Manuf.-Green Technol.* **2014**, *1*, 283–292. [CrossRef]
3. BCC. Why Is There a Global Energy Crisis and Who Might Suffer Most from It? *BBC*, 30 October 2022.
4. UK Government. Government Report. 2010 to 2015 Government Policy: Energy Demand Reduction in Industry, Business and the Public Sector. 2016. Available online: 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 (accessed on 10 December 2024).
5. Rogers, A.J.; Hashemi, A.; Ierapetritou, M.G. Modeling of Particulate Processes for the Continuous Manufacture of Solid-Based Pharmaceutical Dosage Forms. *Processes* **2013**, *1*, 67–127. [CrossRef]
6. Lee, S.L.; O'Connor, T.F.; Yang, X.; Cruz, C.N.; Chatterjee, S.; Madurawe, R.D.; Moore, C.M.V.; Yu, L.X.; Woodcock, J. Modernizing Pharmaceutical Manufacturing: From Batch to Continuous Production. *J. Pharm. Innov.* **2015**, *10*, 191–199. [CrossRef]
7. Ierapetritou, M.; Muzzio, F.; Reklaitis, G. Perspectives on the continuous manufacturing of powder-based pharmaceutical processes. *AIChE J.* **2016**, *62*, 1846–1862. [CrossRef]
8. Manzano, T.; Langer, G. Getting Ready for pharma 4.0. Data Integrity in Cloud and Big Data Applications. 2018. Available online: https://www.ispe.gr.jp/ISPE/02_katsudou/pdf/201812_en.pdf (accessed on 10 December 2024).
9. Douglas, J.M. *Conceptual Design of Chemical Processes*; McGraw-Hill: New York, NY, USA, 1988; ISBN 978-0070177628.

10. Chen, Y.; Yang, O.; Sampat, C.; Bhalode, P.; Ramachandran, R.; Ierapetritou, M. Digital Twins in Pharmaceutical and Biopharmaceutical Manufacturing: A Literature Review. *Processes* **2020**, *8*, 1088. [CrossRef]

11. Martin, S. What Is a Digital Twin? September 2022. Available online: https://blogs.nvidia.com/blog/2021/12/14/what-is-a-digital-twin/ (accessed on 10 December 2024).

12. IBM. What Is a Digital Twin? December 2020. Available online: https://www.ibm.com/uk-en/topics/what-is-a-digital-twin (accessed on 10 December 2024).

13. Barni, A.; Fontana, A.; Menato, S.; Sorlini, M.; Canetta, L. Exploiting the Digital Twin in the Assessment and Optimization of Sustainability Performances. In Proceedings of the 2018 International Conference on Intelligent Systems (IS), Funchal, Portugal, 25–27 September 2018; pp. 706–713.

14. Veleva, V.; Hart, M.; Greiner, T.; Crumbley, C. Indicators for measuring environmental sustainability: A case study of the pharmaceutical industry. *Benchmarking Int. J.* **2003**, *10*, 107–119. [CrossRef]

15. Ntamo, D.; Lopez-Montero, E.; Mack, J.; Omar, C.; Highett, M.I.; Moss, D.; Mitchell, N.; Soulatintork, P.; Moghadam, P.Z.; Zandi, M. Industry 4.0 in Action: Digitalisation of a Continuous Process Manufacturing for Formulated Products. *Digit. Chem. Eng.* **2022**, *3*, 100025. [CrossRef]

16. Papadopoulos, I. A technical literature review on sustainability-oriented digital twin of the diamond pilot plant. In *Internal Research Report*; University of Sheffield: Sheffield, UK, 2023; Unpublished.

17. Schume, P. Improve Product Quality and Yield with Intelligent, Secure, and Adaptable Manufacturing Operations. May 2020. Available online: https://www.ibm.com/think/topics/ai-for-manufacturing (accessed on 10 December 2024).

18. Bai, C.; Dallasega, P.; Orzes, G.; Sarkis, J. Industry 4.0 technologies assessment: A sustainability perspective. *Int. J. Prod. Econ.* **2020**, *229*, 107776. [CrossRef]

19. He, B.; Bai, K.-J. Digital twin-based sustainable intelligent manufacturing: A review. *Adv. Manuf.* **2020**, *9*, 1–21. [CrossRef]

20. Vanhoorne, V.; Vervaet, C. Recent progress in continuous manufacturing of oral solid dosage forms. *Int. J. Pharm.* **2020**, *579*, 119194. [CrossRef]

21. Grieves, M. SME Management Forum Completing the Cycle: Using PLM Information in the Sales and Service Functions. October 2002. Available online: https://www.researchgate.net/publication/356192963_SME_Management_Forum_Completing_the_Cycle_Using_PLM_Information_in_the_Sales_and_Service_Functions (accessed on 10 December 2024).

22. GEA Group AG. Modul P Tablet Press. 2023. Available online: https://www.gea.com/en/products/tablet-presses/rd-tablet-presses/modul-p-tablet-press/ (accessed on 10 December 2024).

23. GEA Group AG. ConsiGma® Granulation and Drying (GD) Modules. 2023. Available online: https://www.gea.com/en/products/granulators/continuous-granulation-lines/consigma-granulation-drying-gd-modules.jsp (accessed on 9 May 2023).

24. Ryckaert, A.; Ghijs, M.; Portier, C.; Djuric, D.; Funke, A.; Vervaet, C.; Beer, T.D. 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* **2021**, *13*, 293. [CrossRef] [PubMed]

25. Vercruysse, J.; Córdoba, D.D.; Peeters, E.; Fonteyne, M.; Delaet, U.; Van, A.I.; De Beer, T.; Remon, J.; Vervaet, C. Continuous twin screw granulation: Influence of process variables on granule and tablet quality. *Eur. J. Pharm. Biopharm.* **2012**, *82*, 205–211. [CrossRef] [PubMed]

26. Ashish Kumar, I. Experimental and Model-Based Analysis of Twin-Screw Wet Granulation in Pharmaceutical Processes. Ph.D. Thesis, Ghent University, Gent, Belgium, 2015. Available online: https://biblio.ugent.be/publication/6956586 (accessed on 10 December 2024).

27. Morrissey, J. A DEM Study of Wet Granulation in a Twin-Screw Granulator. 2022. Available online: https://www.research.ed.ac.uk/en/activities/a-dem-study-of-wet-granulation-in-a-twin-screw-granulator (accessed on 10 December 2024).

28. Seem, T.C.; Rowson, N.A.; Ingram, A.; Huang, Z.; Yu, S.; Matas, M.D.; Gabbott, I.; Reynolds, G.K. Twin screw granulation—A literature review. *Powder Technol.* **2015**, *276*, 89–102. [CrossRef]

29. McGuire, A.D.; Mosbach, S.; Lee, K.F.; Reynolds, G.; Kraft, M. A high-dimensional, stochastic model for twin-screw granulation—Part 1: Model description. *Chem. Eng. Sci.* **2018**, *188*, 221–237. [CrossRef]

30. Shanmugam, S. Granulation techniques and technologies: Recent progresses. *BioImpacts BI* **2015**, *5*, 55. [CrossRef]

31. Mwiti, D. Random Forest Regression: When Does It Fail and Why? April 2023. Available online: https://neptune.ai/blog/random-forest-regression-when-does-it-fail-and-why (accessed on 10 December 2024).

32. Sreenivasa, S. Radial Basis Function (RBF) Kernel: The Go-to Kernel. October 2020. Available online: https://towardsdatascience.com/radial-basis-function-rbf-kernel-the-go-to-kernel-acf0d22c798a (accessed on 10 December 2024).

33. Teixeira-Pinto, A. Machine Learning for Biostatistics—k-Nearest Neighbours Regression. July 2022. Available online: https://bookdown.org/tpinto_home/Regression-and-Classification/k-nearest-neighbours-regression.html (accessed on 10 December 2024).

34. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

35. VanDerHorn, E.; Mahadevan, S. Digital twin: Generalization, characterization and implementation. *Decis. Support Syst.* **2021**, *145*, 113524. [CrossRef]

36. *BS ISO 23247-1:2021*; Automation Systems and Integration. Digital Twin Framework for Manufacturing—Overview and General Principles. International Organization for Standardization: Geneva, Switzerland, 2021.