IE7215 Simulation Analysis

Biomanufacturing Process Modeling and Modifying

Northeastern University

Group 5

Hongzhao Ma

Quanhan Sun

Shihan Ma

Abstract

In this paper we build a simulation model for a biomanufacturing process that produces two types of bio-drugs. By exploring the biomanufacturing process from previous research work made from Wang, Xie, Martagan, Akcay, and Corlu, we simulate the process using Python, and discovered key factors that have significant influence on the performance of the biomanufacturing process. Assumptions were made for setting up initial attributes values of the simulation based on our investigation. Initial input data were given with limited amounts, so we used bootstrap method to analyze and quantify the input uncertainty. This paper also shows the output value from our simulation model and analyzes the model performance by providing mean values and confidence intervals of interesting attributes. With adjustment on some key factors, for instance, input releasing policy, capacity of main procedures, and number of steps of the Chromatography process. We discovered some significant and insignificant changes brought by adjusting the simulation model and gave suggestions for modifying this biomanufacturing process in order to improve its performance.

Contents

Introduction **1**
1.1 Background 1
1.2 Problem Identification 1
1.3 Purpose 1

Biomanufacturing Process Illustration **2**
2.1 Upstream cell culture process(USP) 3
2.2 Downstream protein purification process(DSP) 4

Model Design and Assumptions **5**
3.1 Arrival Time and Drugs Type 5
3.2 Number of Equipment 5
3.3 Processing time 6
3.4 Thresholds 6
3.5 Releasing Policy 6
3.5 Replication parameters 6

Input Modeling and Input Uncertainty Quantification **7**
4.1 Input Model Fitting 7
4.2 Input Uncertainty Quantification 9

Output Analysis **11**
5.1 Basic Stats of the simulation output 11
5.2 Quantile Analysis 12

Perform design of experiment and Modification **12**

Conclusion **14**

1. Introduction

1.1 Background

With the fast growing speed of the biomanufacturing industry, which is now highly valued and widely developed among the world, there is some necessary work to do for industrial engineers in order to benefit this industry and humankind. The complexity of the biomanufacturing process is one of the problems limiting producing bio-drugs efficiently. Simulation analytics is one of the great methods for better understanding the biomanufacturing process and increasing production rate and good-quality rate. Simulation helps to understand the behavior and outcomes of the biomanufacturing process.

1.2 Problem Identification

The biomanufacturing process, which our group did simulation analysis on, includes two types of products, Type A bio-drug and type B bio-drug. These two types of bio-drugs have mostly the same producing process. But type A drug needs external media for fermentation processes. For type B bio-drug, the external media is not necessary. But we need to prepare the media internally. The production of internal media is included in the biomanufacturing process. Type A bio-drug and type B bio-drug also have different demands, meaning that in the same time period, the amount of type A and type B needed to be produced are different. Then there is some difference of limitations on the capacity of each process. For instance, the number of equipment for each procedure is different and not fixed. It needs investigation to find the impact of equipment amounts on model performance.

1.3 Purpose

There are some clear objectives of this project, which are shown in the bullet points below:

- **Analyze the model performance**
  By analyzing the model outcome, we will know the performance of the model with the current experiment. After changing some initial setups, the model will show better results or some otherwise outcome, which helps us to analyze and improve our model.
- **Identify the key procedures**
  Some procedures are considered to be critical ones that either limit the performance or are important technically. Finding these important procedures will help us improve the model in the right and available direction.

- **Make modification and suggestions**
  The final purpose of any simulation will be improving the system to achieve a better goal, whether it's larger productivity or better product quality. So, modifying the model and investigating each result and comparing the outcomes is very important, because based on that, we can give our suggestions about how to improve the process on materials, equipment and some more details.

2. Biomanufacturing Process Illustration

The whole biomanufacturing process is split into two parts, the upstream cell culture process and downstream protein purification process. And all processes are presented in the *Figure 2-1*:

(1) Preculture bio-drugs
(2) Internal media preparation
(3) Inoculum fermentation
(4) Main fermentation
(5) Centrifuge
(6) Chromatography
(7) Filtration.
(8) Quality Control

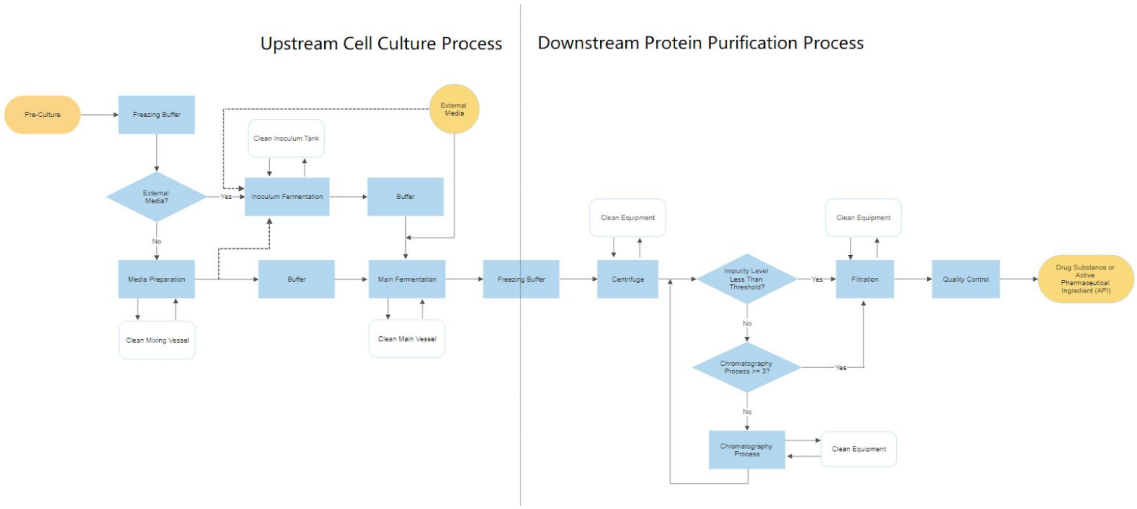

figure 2-1 Biomanufacturing Process

The biomanufacturing process can be split to two part: (1) Upstream cell culture process(USP); (2) Downstream protein purification process(DSP)

2.1 Upstream cell culture process(USP)

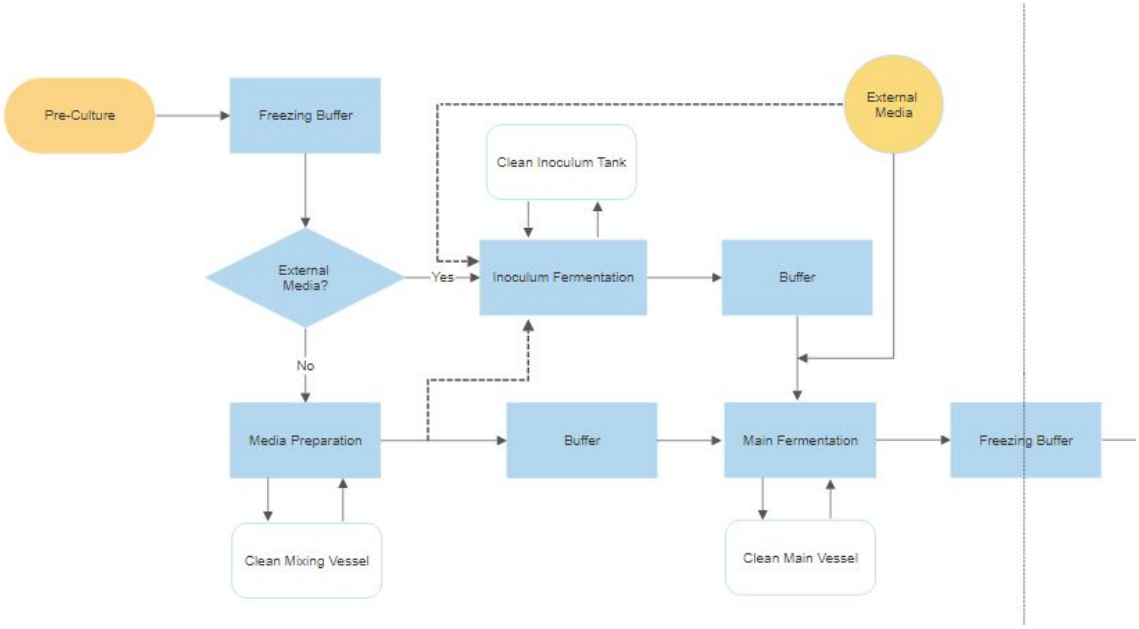

Figure 2-1-1 Upstream cell culture process(USP)

Figure 2-1-1 shows the USP process. The USP starts with the inoculum fermentation. All the antigen will be produced in batches. After the inoculum fermentation is finished, we clean the inoculum tank immediately and start main fermentation at the same time. When the main fermentation is finished, we check the quality of the antigen. If the impurity proportion is higher than a certain threshold, we throw away this batch.

2.1.1 Pre-Culture and Media Preparation

There are two types of antigens, Antigen A and Antigen B, to be produced. Each type of antigen will be produced in batches. For Antigen A, it needs the external media for inoculum fermentation and main fermentation. For Antigen B, the internal media is in need for these two processes. We assume that external media is always enough and doesn't need to be produced in the system. Internal media need to be prepared before the inoculum fermentation and main fermentation. The media preparation equipment can produce both media.

2.1.2 Inoculum Fermentation

The batch will wait for a releasing signal from the USP releasing policy. The inoculum fermentation will start only when there is enough inoculum media. When the process begins, the initial biomass will be inoculated and fermented. The amount of protein remains unchanged in this process. After the fermentation is finished, the material will be moved to the buffer between inoculum fermentation and main fermentation.

2.1.3 Main Fermentation

The main fermentation will start only when there is enough main media. The amount of protein and impurity will increase in this process. After the main fermentation is finished. Quality check will decide whether to discard the batch, if not, the batch will wait in the freeze buffer for the DSP process.

2.2 Downstream protein purification process(DSP)

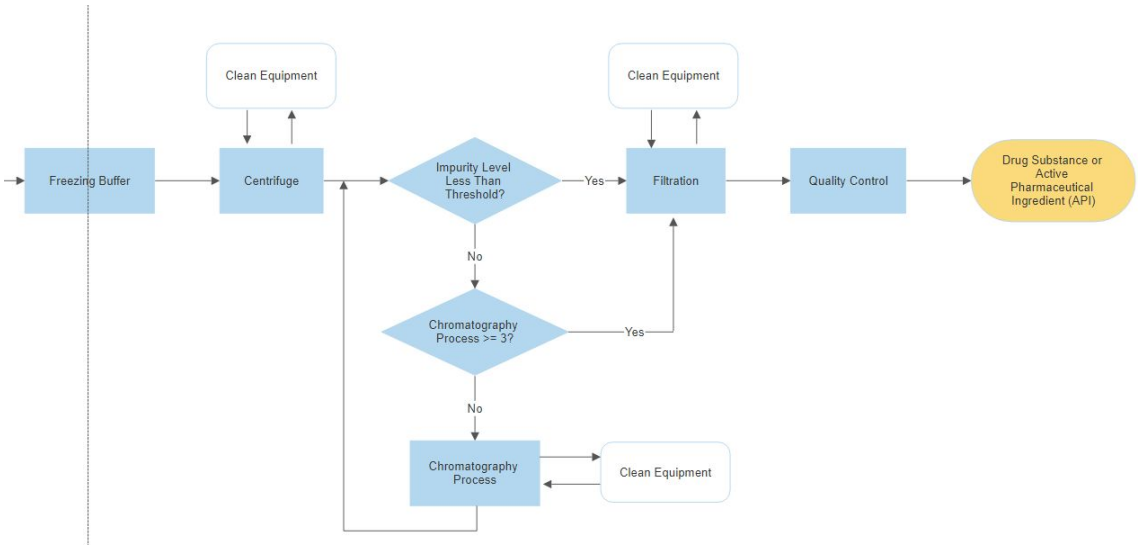

Figure 2-2-1 Downstream protein purification process

Figure 2-2-1 shows the DSP process. In the DSP, when releasing the batches from the freezing buffer, first we will centrifuge the batches and then test for impurity level. If it exceeds the threshold we've set up, then we need to put it into the Chromatography process and after it, we test for impurity again. The batches will be sent to the filtration if they pass the impurity test. The maximum number of chromatography processes is three, which means that when it cycles three times in chromatography processes, the batch will be sent to the next step regardless of whether the impurity level is less than threshold or not. After that are the filtration process and finally the quality control process. The centrifuge, chromatography and filtration process need to clean equipment immediately after finished. In the whole DSP, due to the fragile of the materials, the batches cannot wait between processes, or the batches will be discarded.

2.2.1 Centrifuge

The batch will wait for a signal from the DSP releasing policies to begin centrifuge. In the centrifuge process, the amount of ending protein will remain unchanged and the impurity will decrease by a percentage.

2.2.2 Chromatography

When the batches are operated in the chromatography, the amount of protein and impurity will decrease. To get a higher purity, the chromatography process may be cycled.

2.2.3 Filtration and Quality Control

Filtration process is the final process to increase the purity. In this process, the amount of protein will remain unchanged and the amount of impurity will decrease. After that, the quality control process will decide whether to discard the production or not.

3. Model Design and Assumptions

To build the model, we need some assumptions and set the initial setups for this system.

3.1 Arrival Time and Drugs Type

We assume that the mean arrival time of initial biomass is one day(24 hours), following the exponential distribution. And the probabilities of drugs types are 0.25 and 0.75 for Antigen A and Antigen B respectively.

3.2 Number of Equipment

The number of equipment is shown as *Table 3-1*, which may be changed in our analysis.

Process	# equipment
Inoculum Fermentation	5
Media Preparation	5
Main Fermentation	5
Centrifuge	2
Chromatography	5
Filtration	2
Quality Control	unlimited

Table 3-1 Number of Equipment

3.3 Processing time

We assume the process time of each process is fixed. *Table 3-2* shows the process time of each.

Time(hours)	Antigen A	Antigen B	Clean Equipment
Inoculum Fermentation	24.0	12.0	1.5
Media Preparation	1.0	0.5	0.6
Main Fermentation	72.0	48.0	6.0
Centrifuge	2.5	2.0	0.2
Chromatography	8.0	7.0	1.5
Filtration	2.0	2.0	0.5
Quality Control	2.0	2.o	

Table 3-2 Process Time

3.4 Thresholds

There is a threshold for impurity level both in USP and DSP. They are used for deciding whether a batch of bio-drug is good enough for continuing the producing process. If not, it will be thrown away. The USP impurity threshold is 80%, which means if a batch has more than 80% impurity rather than the desired protein, after the main fermentation process, it will be thrown away. And if a batch has more than 20% impurity after quality control, it will also be discarded. There is also a “No-Wait” constraint in the DSP, that is if a batch starts waiting for any further procedure during DSP, it will be discarded, regarding that bio-drugs must be produced in a strict environment.

3.5 Releasing Policy

We set two releasing policies for the USP and DSP. Only when the number of batches worked in USP is less than 5, the next batch will begin to be processed in USP. And in DSP, only when the number of batches worked in DSP is less than 2, the next batch will be sent to the centrifuge process.

3.5 Replication parameters

The run length and warm up length for this simulation is also modified. *Figure 3-1* shows the relationship between mean cycle time and run length.

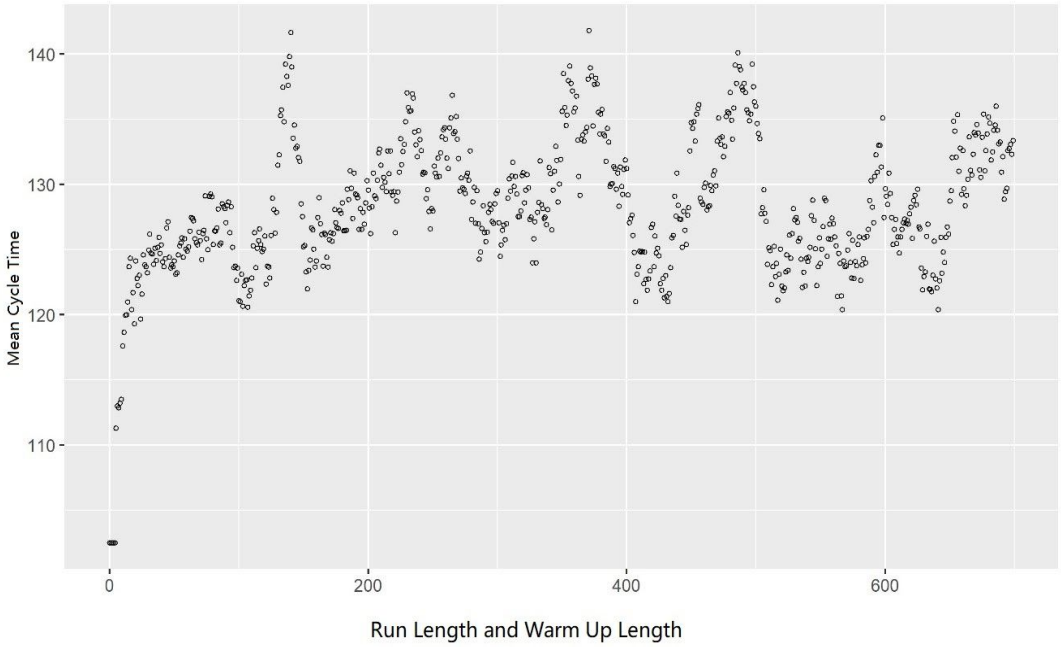

Figure 3-1 Mean Cycle Time of each batch in all replications

After we investigated the impact of run length on the mean cycle time for each batch of bio-drugs, we found that after running 500 batches in each replication, the model will get to a steady state, which is an ideal situation for collecting data. So we set up the total run length to 700 batches, including the first 500 batches for the warm up session. And for the experimentation, we set up the replication number to 100, with relative error equals to 0.013. In each replication, we will collect data from 200 batches.

4. Input Modeling and Input Uncertainty Quantification

4.1 Input Model Fitting

4.1.1 Initial Biomass

For the initial biomass, we only have 20 real-world data. Firstly, we draw the histogram of these two types of biomass as *Figure 4-1-1*.

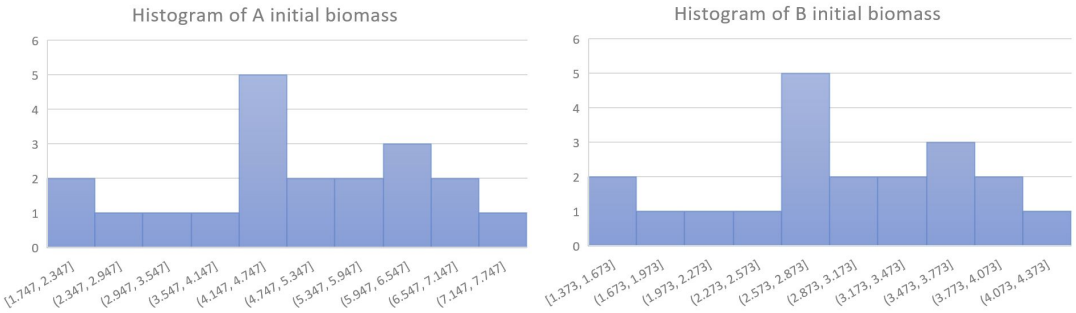

Figure 4-1 Histogram of initial biomass

Since the data doesn't fit any simple theoretical distribution, we decided to apply the empirical distributions in our model. To obtain the continuous distribution, the linearly interpolated ecdf method is used.

4.1.2 Main Fermentation

Main fermentation process will increase the protein and impurity Simultaneously. The relationship between Initial protein X_0, ending protein X_F and ending impurity I_F can be shown as follow

$$X_F = X_0 \cdot e^{\mu T + \varepsilon_P} \text{ and } I_F = X_F \cdot \alpha \cdot e^{\varepsilon_I}$$

where $\varepsilon_P \sim N(0, \sigma_P^2)$ and $\varepsilon_I \sim N(0, \sigma_I^2)$, $T = 72$ and $T = 36$ for type A and type B respectively.

We can take the logarithm of the equations and fit regression

$$logX_F - logX_0 = \mu T + \varepsilon_P \text{ and } log\frac{I_F}{X_F} = log\alpha + \varepsilon_I$$

After fitting regression, we can get $(\mu, \alpha, \sigma_P, \sigma_I)$

For type A drug:

$$X_F = X_0 \cdot e^{0.0249T + \varepsilon_P}$$
$$I_F = X_F \cdot 2.9914 \cdot e^{\varepsilon_I}$$
$$\varepsilon_P \sim N(0, 0.2687^2)$$
$$\varepsilon_P \sim N(0, 0.0966^2)$$

We further performed the Shapiro-Wilk test on the residuals with fitted normal distribution, which gives the test statistics W=0.9453, and p-value 0.3018 for ε_P. And W=0.9724, and p-value 0.8047 for ε_I. Thus, we claim that under current data, we statistically do not reject this model.

For type B drug:

$$X_F = X_0 \cdot e^{0.0221T + \varepsilon_P}$$
$$I_F = X_F \cdot 3.0034 \cdot e^{\varepsilon_I}$$
$$\varepsilon_P \sim N(0, 0.3459^2)$$
$$\varepsilon_P \sim N(0, 0.1192^2)$$

We also performed the Shapiro-Wilk test, which gives the test statistics W=0.9235, and p-value 0.1156 for ε_P and W=0.9726, and p-value 0.8079 for ε_I. Thus, we statistically do not reject this model.

4.1.3 Chromatography

We have m=30 batches real-world data for Q_p and Q_I. The histograms are shown as *Figure 4-1-2* Using @RISK to fit the models, we get that Q_p ~U(0.657, 0.918) , with Chi-squared statistic = 3.600, Q_I ~U(0.242, 0.665), with Chi-squared statistic = 1.600.

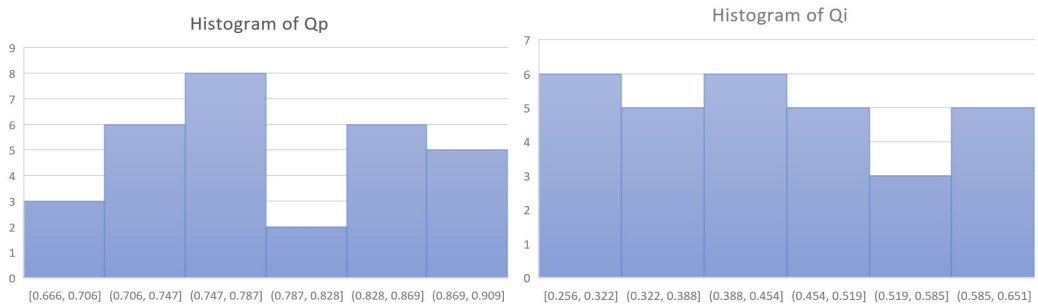

Figure 4-1-2 Histogram of Q_p and Q_I

4.2 Input Uncertainty Quantification

To perform the input uncertainty quantification, we use bootstrap to resample the real-world data and fit the models, then we run the simulation at each bootstrapped input model to study the impact of input uncertainty on system performance estimation. Here we need to bootstrap three data sets, and the outputs are shown as *table 4-2-1, 4-2-2* and *4-2-3*.

Type A							
Output	Mean Protein	Mean Impurity	Mean CycleTime	Total Yield	Var Protein	Var Impurity	Var CycleTime
CI Lower	15.960	3.002	130.740	411.122	52.830	2.445	122.488
CI Upper	16.449	3.094	130.857	424.300	55.452	2.555	127.741
Variance	1.552	0.055	0.090	1130.180	44.739	0.079	179.607
Type B							
Output	Mean Protein	Mean Impurity	Mean CycleTime	Total Yield	Var Protein	Var Impurity	Var CycleTime
CI Lower	4.728	0.889	91.930	363.541	5.313	0.239	121.870
CI Upper	4.866	0.915	92.035	374.361	5.537	0.249	127.026
Variance	0.124	0.004	0.072	761.830	0.328	0.001	173.038

Table 4-2-1 Quantifying impact of Initial Biomass Uncertainty

As the *table 4-2-1* shows, by applying the bootstrapped initial biomass input model, the variance of total yield is pretty large, which means the limited data have a big impact of simulation estimation uncertainty. Thus, more real-world data need to be collected to quantify the batch-to-batch variation.

Type A							
Output	Mean Protein	Mean Impurity	Mean CycleTime	Total Yield	Var Protein	Var Impurity	Var CycleTime
CI Lower	15.793	2.956	130.434	455.687	36.782	1.826	121.738
CI Upper	16.237	3.039	130.558	469.545	41.258	2.019	127.216
Variance	1.280	0.045	0.101	1249.715	130.368	0.243	195.250
Type B							
Output	Mean Protein	Mean Impurity	Mean CycleTime	Total Yield	Var Protein	Var Impurity	Var CycleTime
CI Lower	2.695	0.501	91.305	256.564	1.174	0.058	122.280
CI Upper	3.042	0.568	91.466	275.645	1.840	0.087	127.026
Variance	0.785	0.029	0.168	2369.338	2.879	0.006	146.562

Table 4-2-2 Quantifying impact of Main Fermentation Uncertainty

Type A							
Output	Mean Protein	Mean Impurity	Mean CycleTime	Total Yield	Var Protein	Var Impurity	Var CycleTime
CI Lower	16.151	3.046	130.804	413.212	53.981	2.484	121.764
CI Upper	16.337	3.084	130.954	435.424	55.321	2.551	127.242
Variance	0.226	0.009	0.145	3210.691	11.679	0.029	195.303
Type B							
Output	Mean Protein	Mean Impurity	Mean CycleTime	Total Yield	Var Protein	Var Impurity	Var CycleTime
CI Lower	4.790	0.903	91.972	365.347	5.364	0.242	121.026
CI Upper	4.843	0.914	92.108	385.297	5.497	0.247	125.981
Variance	0.018	0.001	0.121	2590.178	0.114	0.000	159.770

Table 4-2-2 Quantifying impact of Chromatography Uncertainty

The limited data of main fermentation and chromatography process data also have great impact on the variance of total yield. Therefore, all data for input modeling need to be collected.

5. Output Analysis

5.1 Basic Stats of the simulation output

The outputs of two types of drugs are shown in *Table 5-1-1* and *Table 5-1-2*.

Name	Mean	CI lower bound	CI upper bound
ProteinMean	16.320	16.024	16.616
ImpurityMean	3.060	3.003	3.117
CycleTimeMean	130.464	129.847	131.081
ProteinVar	56.674	51.008	62.339
ImpurityVar	2.475	2.216	2.735
CycleTimeVar	105.711	81.582	129.841
Throughput	0.125	0.122	0.127
TotalYield	417.639	401.135	434.143
PassRate	0.511	0.498	0.525
Ino_Utilization	0.136	0.134	0.139
Fer_Utilization	0.500	0.493	0.507
Cen_Utilization	0.048	0.047	0.049
Chr_Utilization	0.233	0.229	0.236
Fil_Utilization	0.049	0.049	0.050

Table 5-1-1 Type A drug output under initial assumption.

Name	Mean	CI lower bound	CI upper bound
ProteinMean	4.828	4.780	4.876
ImpurityMean	0.904	0.893	0.915
CycleTimeMean	91.881	91.355	92.408
ProteinVar	5.390	5.141	5.640
ImpurityVar	0.241	0.228	0.253
CycleTimeVar	122.592	99.843	145.342
Throughput	0.052	0.052	0.053
TotalYield	377.090	369.091	385.090
PassRate	0.520	0.512	0.528
Ino_Utilization	0.136	0.134	0.139
Fer_Utilization	0.500	0.493	0.507
Cen_Utilization	0.048	0.047	0.049
Chr_Utilization	0.232	0.229	0.236

Fil_Utilization	0.049	0.049	0.050

Table 5-1-2 Type B drug output under initial assumption

From the table we find that the utilization of the Chromatography process and Fermentation process is higher than others, and all the utilization rates are low. And the pass rate is only about 0.5, which is very low, so we will perform the sensitive analysis to find the methods to improve pass rate in the next section.

5.2 Quantile Analysis

We also did the quantile Analysis for the output, the results are shown as *Table 5-2-1* and *Table 5-2-2*.

	CycleTimeMean	Impurity rate (%)			ProteinMean	TotalYield	PassRate
95% quantile	135.718	0.162		5% quantile	13.907	267.985	0.404
CI Lower	134.183	0.161		CI Lower	13.443	261.732	0.365
CI Upper	143.370	0.165		CI Upper	14.242	299.074	0.414

Table 5-2-1 Quantile Analysis for Type A

	CycleTimeMean	Impurity rate (%)			ProteinMean	TotalYield	PassRate
95% quantile	97.228	0.1598		5% quantile	4.430	315.719	0.453
CI Lower	96.369	0.1597		CI Lower	4.345	294.906	0.431
CI Upper	100.343	0.162		CI Upper	4.511	325.680	0.456

Table 5-2-1 Quantile Analysis for Type B

For Type A, 95% of the repulications cycle time is less than 135.718 and 95% of impurity rate is less than 16.20%. 95% of the protein mean is greater than 13.907, 95% of the total yield is greater than 267.985.
For Type A, 95% of the repulications cycle time is less than 97.228 and 95% of impurity rate is less than 15.98%. 95% of the protein mean is greater than 4.430, and 95% of the total yield is greater than 315.719.

6. Perform design of experiment and Modification

In this session, we choose three aspects to perform sensitive analysis by adjusting these parameters:

(1) The resources number of Fermentation and Chromatography.
(2) The steps during Chromatography.
(3) The capacity of the Downstream process.

The results of the experiment are as *Table 6-1* and *Table 6-2*:

	Initial	Chr+1	Fer+1	Chr+1,Fer+1	Step+1	Dsp+1
ProteinMean	16.320 ±0.296	16.379 ±0.264	16.310 ±0.287	16.449 ±0.279	14.661 ±0.242	16.374 ±0.295
ImpurityMean	3.060 ±0.057	3.060 ±0.052	3.074 ±0.062	3.084 ±0.061	2.687 ±0.050	3.079 ±0.057
CycleTime Mean	130.465 ±0.617	131.446 ±0.768	130.950 ±0.666	131.029 ±0.555	133.915 ±0.504	117.579 ±19.796
TotalYield	417.640 ±16.504	444.260 ±16.623	417.129 ±16.857	444.558 ±17.436	589.423 ±19.764	418.083 ±18.191
PassRate	0.512 ±0.018	0.541 ±0.013	0.519 ±0.014	0.547 ±0.015	0.811 ±0.012	0.511 ±0.014

Table 6-1 Sensitive Analysis of Type A

	Initial	Chr+1	Fer+1	Chr+1,Fer+1	Step+1	Dsp+1
ProteinMean	4.828 ±0.048	4.824 ±0.051	4.840 ±0.050	4.860 ±0.059	4.329 ±0.039	4.868 ±0.059
ImpurityMean	0.904 ±0.011	0.906 ±0.010	0.910 ±0.011	0.915 ±0.012	0.791 ±0.008	0.911 ±0.012
CycleTime Mean	91.882 ±0.526	92.670 ±0.635	92.036 ±0.538	92.219 ±0.509	94.917 ±0.377	91.961 ±0.433
TotalYield	377.090 ±8.000	384.310 ±8.524	378.188 ±7.976	390.683 ±8.569	522.909 ±9.001	373.642 ±7.511
PassRate	0.520 ±0.008	0.531 ±0.008	0.517 ±0.008	0.534 ±0.009	0.803 ±0.007	0.511 ±0.008

Table 6-2 Sensitive Analysis of Type B

From the result, we find that when we increase the limit of the chromatography process step, the pass rate will increase. Meanwhile, the mean amount of protein decreases, as a result, the total yield increases. More equipment in chromatography will have a little impact on the pass rate. Other changes seem to have no impact on the system. After our analysis, we found that since we set up the number of equipment in main fermentation equal to the USP batches limit and the number of equipment in centrifuge equal to the DSP batches limit, more equipment will not improve the system.

7. Conclusion

This project basically builds the simulation model for a biomanufacturing process for researchers to investigate its performance. Protein level, cycle time, yield rate, pass rate, and impurity level are five most important attributes that could inflex the performance of this biomanufacturing model. Main fermentation and Chromatography are two key procedures in the whole process that will mostly determine the output attributes. The output result shows that adding one more step for the chromatography process can improve total yield and pass rate, which increase productivity of the biomanufacturing process.