

Design of Experiments with JMP

Hands-on Case Study:

E. Coli Fermentation

Phil Kay, JMP Learning Manager

The Situation

The case study is from a published example on an E. coli fermentation process for recombinant protein production [Tai 2015]. 10 possible important factors have been identified, including:

- Batch glycerol concentration,
- dissolved oxygen,
- temperature,
- and pH.

The questions we seek to answer are:

- What, if any, effect do these factors have on titer and truncated product?
- How should these factors be controlled to ensure consistent high product titer and minimal production of the truncated product?

Learning Outcomes

First you will design and analyse a huge experiment to look at all possible combinations of each factor at 3 levels. From this you will learn:

1. The overwhelming number of possibilities for bioprocesses with multiple factors.
2. Why you need a smart, strategic approach to exploring the “possibility space”.

And in addition:

3. What you can learn from large datasets with some simple, interactive visual analyses.

Next you will design an efficient and effective experiment to meet the scientific objectives. From this you will learn:

4. How to design and analyse a Definitive Screening Design in JMP Statistical Discovery software.
5. How you can efficiently learn about the key behaviours of a process or system by only testing a small subset of all the possibilities using the “magic” of Design of Experiments.

References

Tai, M., Ly, A., Leung, I. and Nayar, G. (2015), Efficient high-throughput biological process characterization: Definitive screening design with the Ambr250 bioreactor system. *Biotechnol Progress*, 31: 1388-1395. <https://doi.org/10.1002/btpr.2142>

Part 1: A Huge Experiment to Look at All Possibilities

You will design a 3-level, full factorial experiment for 10 factors. This huge experiment is to illustrate the overwhelming number of possibilities in multi-factor processes and systems. **It is not a recommended approach.**

A: Design the Experiment

1. In your JMP journal, click on the links for Responses and Factors to open them as JMP tables.

Part 1: A huge experiment to look at all possibilities!

You are going to design a huge experiment where you will test every possible combination of each factor at 3 levels.

A: Design the Experiment

[Responses](#)

[Factors](#)

[Full Factorial Experiment](#)

The Factors table has ranges for the factor settings that we wish to explore. The first row represents the low settings, the middle row is the mid-point settings, and the third row the high settings. In a real situation you would use the actual ranges of factors but from this published example we only have the coded ranges, -1 to 1.

	Glycerol conc	pH	DO	Bx T	Glycerol feed	Yeast feed	Duration	Prod T	Prod Glycerol Feed	Prod Yeast Feed
1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2	0	0	0	0	0	0	0	0	0	0
3	1	1	1	1	1	1	1	1	1	1

You need to get this information into the Full Factorial Design dialog.

2. Find the menu from any window in JMP. You may need to click on the breadcrumbs, "...", at the top of the window.
3. DOE > Classical > Full Factorial
4. Practise entering responses and factors. These are all continuous factors so set Add N Factors to 10 and click the 'Continuous' button and '3 Level'. Manually enter the names for each factor together with their low, mid and high settings (-1, 0, 1 in this case).

Name	Role	Values
Yeast feed	Continuous	-1 0 1
Duration	Continuous	-1 0 1
Prod T	Continuous	-1 0 1
Prod Glycerol Feed	Continuous	-1 0 1
Prod Yeast Feed	Continuous	-1 0 1

5. Create and save a Factor table. After you have input all factors go into the 'red triangle' menu next to the Definitive Screening Design header and Save Factors. Then go into the File

menu and Save as 'Full Fac Factors'. This Factor setting table can then be loaded from the 'red triangle' menu by selecting 'Load Factors' to avoid having to type everything again if you are creating another similar experiment.

6. Load the responses. Click on the Responses table and then back to the Full Factorial Design window > red triangle menu > Load Responses.

Responses

Add Response Remove Number of Responses...

Response Name	Goal	Lower Limit	Upper Limit	Importance
Titer	Maximize	800	.	.
Truncated Prod	Minimize	.	420	1

7. Click Continue and you will see that there will be 59049 runs!

3x3x3x3x3x3x3x3x3 Factorial

Output Options

Run Order: Randomize

Number of Runs: 59049

Number of Center Points: 0

Number of Replicates: 0

Make Table

Back

8. Click Make Table.

This is your experiment where you will test all possible combinations of each factor at low, mid, and high settings in their ranges. (Experiments are randomised so the order of rows will not be the same.)

	Pattern	Glycerol conc	pH	DO	Bx T	Glycerol feed	Yeast feed	Duration	Prod T	Prod Glycerol Feed	Prod Yeast Feed	Titer	Truncated Prod
1	2113212222	0	-1	-1	1	0	-1	0	0	0	0	.	.
2	1332122131	-1	1	1	0	-1	0	0	-1	1	-1	.	.
3	1231231133	-1	0	1	-1	0	1	-1	-1	1	1	.	.
4	2322122211	0	1	0	0	-1	0	0	0	-1	-1	.	.
5	2332322233	0	1	1	0	1	0	0	0	1	1	.	.
6	1111212323	-1	-1	-1	-1	0	-1	0	1	0	1	.	.
7	3312313231	1	1	-1	0	1	-1	1	0	1	-1	.	.
8	1213312331	-1	0	-1	1	1	-1	0	1	1	-1	.	.
9	2312221233	0	1	-1	0	0	0	-1	0	1	1	.	.
10	2133333121	0	-1	1	1	1	1	1	-1	0	-1	.	.
11	2222222222	0	0	0	1	0	0	0	1	0	1	.	.

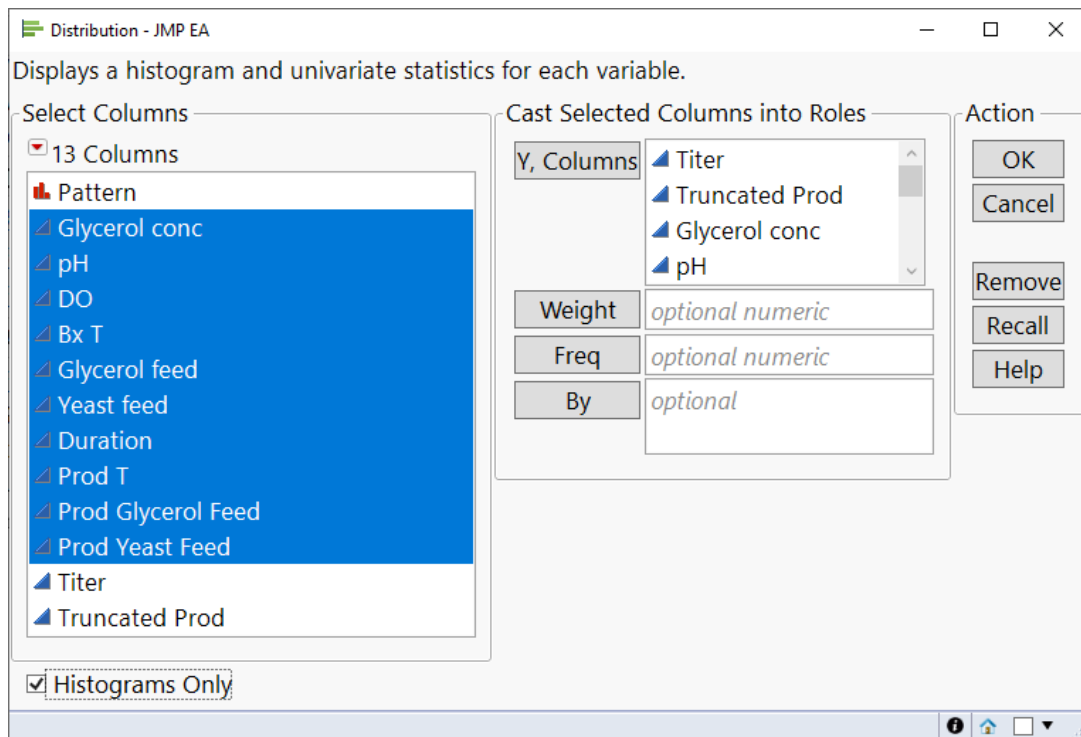
Imagine if you had to go to the lab and carry out all 59049 fermentations according to these factor settings and then record the Titer and amount of Truncated Product. The point is that the possibility space is huge and overwhelming. You can't possibly test all these possibilities, so you will need a smart way to efficiently explore the factor space.

9. Close all JMP windows except the journal and home window. (No need to save any tables)

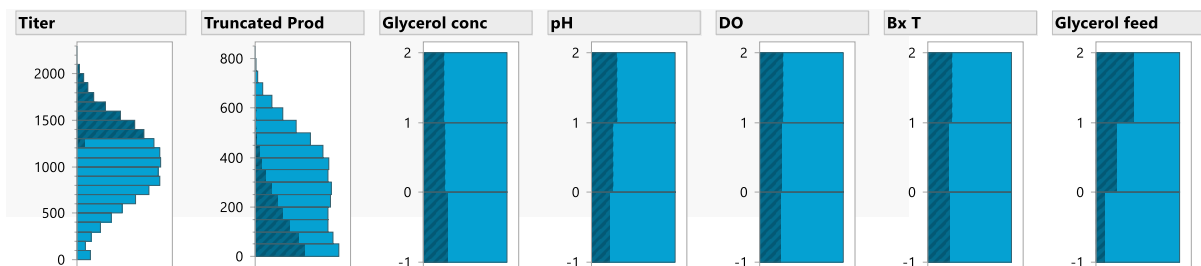
B: Analyse the Experiment

Good news! My minions have carried out the full experiment and recorded the responses for all 59049 runs. (Not really – I simulated the results for each run using the model from the published experiment with some added random noise). You will now analyse this large data set with some simple visual and statistical tools.

1. Click on “Full factorial experiment completed” in your journal
2. Plot interactive distribution of factors and responses. Analyze > Distribution. Select Titer and Truncated Prod and click Y, Columns. Select all the factors and add them to Y, Columns. Check Histograms Only and click OK.

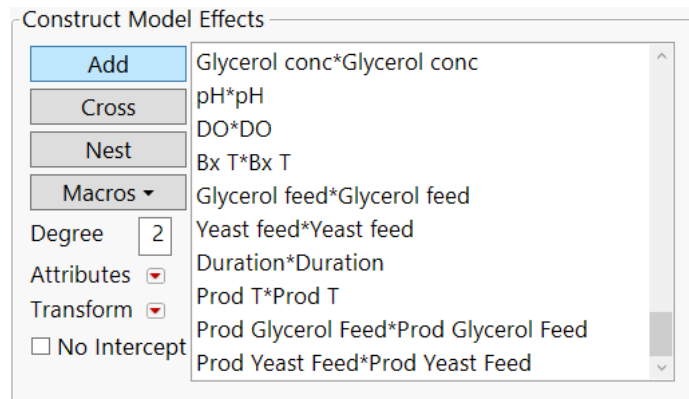


3. Explore how the responses are related to the factors. Click and drag over the bars for high Titer. What patterns do you see? Consider how visual models like this can help to communicate key behaviours in multi-factor processes and systems.

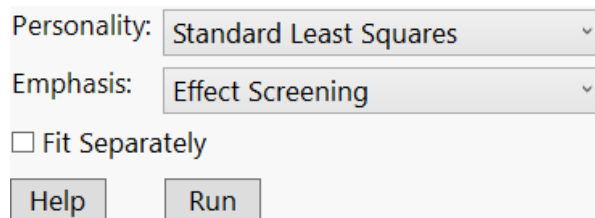


4. Now fit a multiple linear regression statistical model. Analyze > Fit Model. The responses are already in the Y role and the factors are all in the Model Effects. All 2-factor interactions are also specified in the model effects. This Fit Model dialog is pre-completed because this is an experiment designed in JMP.

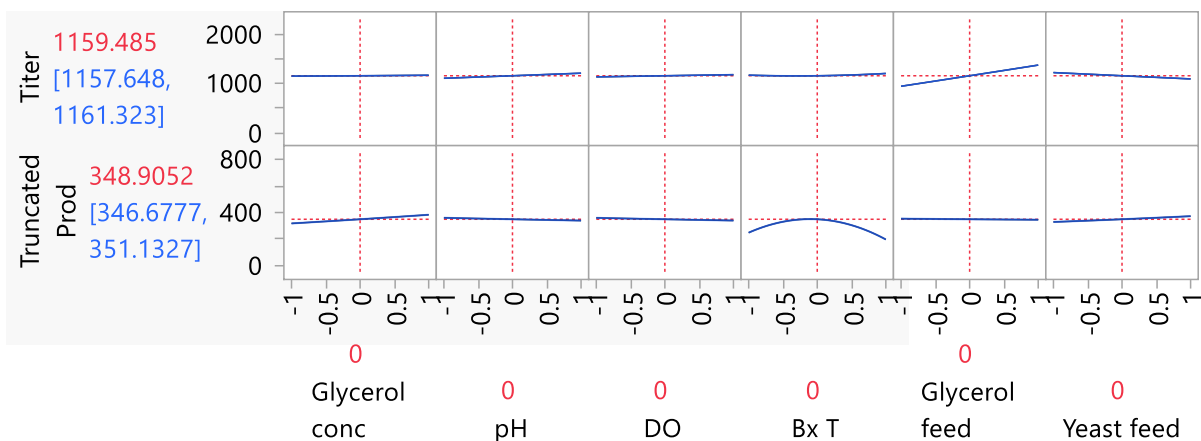
- We want to add the quadratic terms, e.g. pH*pH, that will enable us to model curvilinear behaviours: Select all factor columns and Macros > Response Surface. If you scroll down the effects list box you will find all 10 quadratics.



- Now fit the model. Change Emphasis to Effect Screening (this only changes the reports that you see to a view that I prefer, it does NOT give a different model) and click Run.



- There is a lot to take in on the Least Squares Fit report. We will discuss this. The main parts to focus on are the 'Effect Summary' at the top and 'Prediction Profiler' at the bottom.



Conclusion

We have a very rich understanding of this process now. We know the factors that are most important, how they affect Titer and Truncated Product, and how to control them to ensure the best process. However, it was HUGELY expensive. In the next section we see how to learn much the same information MUCH more efficiently.

Part 2: Efficient and Effective Experimentation with a Definitive Screening Design

Using the same tables of factors and responses you will design a much more efficient experiment to understand the fermentation process. The experiment that you design will have 24 runs instead of 59049 runs! This was the experiment that was used in published example. You will also analyse the real data from the published experiment.

A: Design the Experiment

1. In your JMP journal, click on the links for Responses and Factors to open them as JMP tables. The Responses and Factors are the same as in the previous section.
2. DOE menu > Definitive Screening > Definitive Screening Design. Load the factors and responses as in the previous section (red triangle menu).

Definitive Screening Design

Responses

Add Response Remove Number of Responses...

Response Name	Goal	Lower Limit	Upper Limit	Importance
Titer	Maximize	800	.	.
Truncated Prod	Minimize	.	420	1

Factors

Continuous Categorical Remove Add N Factors 1

Name	Role	Values
Glycerol conc	Continuous	-1 1
pH	Continuous	-1 1
DO	Continuous	-1 1
Bx T	Continuous	-1 1
Glycerol feed	Continuous	-1 1

3. Note that there is a Categorical button which allows 2-level categorical factors to be incorporated into a Definitive Screening Design. We don't have any categorical factors so just click on Continue.
4. We can also add blocks to the Definitive Screening Design if required. The published example had 4 centre-point runs (all factors at mid setting), which we can achieve by choosing Add Blocks with Center Runs.... Set Number of Blocks to 4 and Number of Extra Runs to 0.

Design Options

No Blocks Required

Add Blocks with Center Runs to Estimate Quadratic Effects

Add Blocks without Extra Center Runs

Number of Blocks 4

Number of Extra Runs 0

Make Design

- Click on Make Design. This will generate the Definitive Screening Design. The Design outline shows the 24 runs required for this experiment.

Design												
Run	Block	Glycerol conc	pH	DO	Bx T	Glycerol feed	Yeast feed	Duration	Prod T	Glycerol Feed	Prod Yeast Feed	Prod
1	1	0	1	1	1	1	1	1	1	1	1	1
2	1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
3	2	1	0	-1	-1	-1	-1	1	1	1	1	1
4	2	-1	0	1	1	1	1	-1	-1	-1	-1	-1
5	3	1	-1	0	-1	1	1	1	-1	-1	1	1
6	3	-1	1	0	1	-1	-1	1	1	-1	-1	-1
7	4	1	-1	-1	0	1	1	1	1	-1	-1	-1
8	4	-1	1	1	0	-1	-1	-1	-1	-1	1	1
9	1	1	-1	1	1	0	-1	-1	1	-1	1	1
10	1	-1	1	-1	-1	0	1	1	-1	1	-1	-1
11	2	1	-1	1	1	-1	0	1	-1	1	-1	-1
12	2	-1	1	-1	-1	1	0	-1	1	-1	1	1
13	3	1	1	-1	1	-1	1	0	-1	-1	-1	1
14	3	-1	-1	1	-1	1	-1	0	1	1	-1	-1
15	4	1	1	-1	1	1	-1	-1	0	1	-1	-1
16	4	-1	-1	1	-1	-1	1	1	0	-1	1	1
17	1	1	1	1	-1	-1	1	-1	1	0	-1	-1
18	1	-1	-1	-1	1	1	-1	1	-1	0	1	1
19	2	1	1	1	-1	1	-1	1	-1	-1	0	0
20	2	-1	-1	-1	1	-1	1	-1	1	1	0	0
21	1	0	0	0	0	0	0	0	0	0	0	0
22	2	0	0	0	0	0	0	0	0	0	0	0
23	3	0	0	0	0	0	0	0	0	0	0	0
24	4	0	0	0	0	0	0	0	0	0	0	0

Notice the patterns in the 1s, 0s, and -1s: For each pair of runs, one factor is at the mid-point (0) and the other factors are at the high or low extremes (1, -1) and are a mirror image. We say that they are “foldover pairs”. This symmetry gives the design unique and useful properties, but further discussion of that is beyond the scope of this workshop.

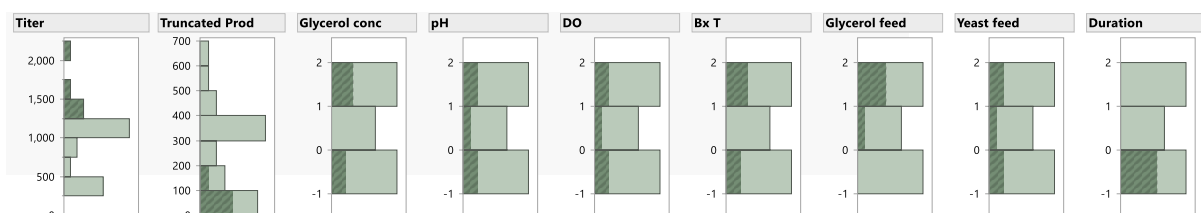
The final four runs are centre-points with all factors at 0. We will ignore the Block column.

- Ensure that Run Order is set to Randomise and click Make Table to create the experimental worksheet as a JMP data table.

In real life you would now have to run this experiment and measure the response for each run. In the published example they used a 24-bioreactor ambr250 to carry out all 24 fermentation runs in parallel. We will analyse this real data next.

B: Analyse the Experiment

- Click on “Completed DSD” in your JMP journal to open the completed experiment table pulled from the published journal article.
- Repeat the visual analysis using Distribution that you did for the huge experiment (uncheck Create Process Capability). Do you see anything different from the huge experiment, other than less data?

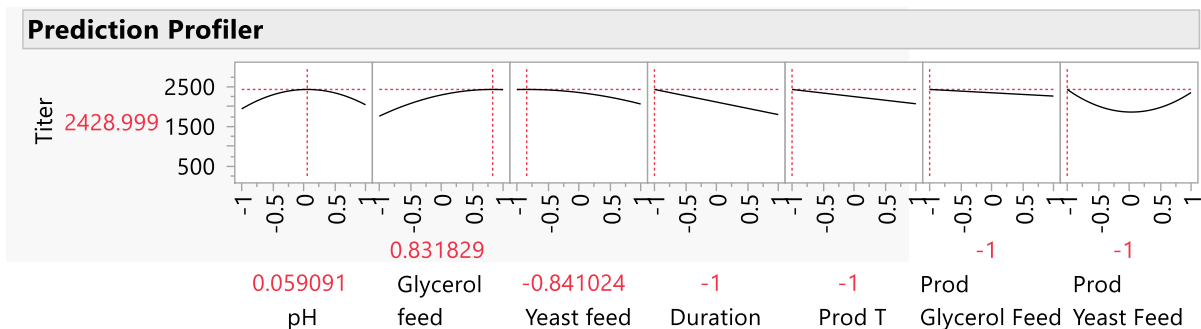


There are several ways you can build a model from the data from this experiment. The easiest is to use the built-in scripts that were created when you designed the experiment.

3. In the top left-hand panel (the script panel) of the table, click on the green triangle next to 'Fit Definitive Screening'.

In the different outlines of the Fit Definitive Screening platform output for each response the active/important factors and terms will be listed. The analysis first screens through to find the important main/linear effects and then looks at the higher order effects in Stage 2. The 'Main Effects Plot' and 'Prediction Profiler' enable visual interpretation of the analysis.

4. Try changing the settings in the prediction profiler to understand how you could improve Titer and reduce Truncated Product. Which factors have the biggest effect? Are there any important interaction effects? What about curvilinear ("quadratic") factor effects?



Having selected the important model effects, we will fit the model in JMP's Fit Model platform.

5. In the Fit Definitive Screening interface click on the Make Model button for Titer.

The Model Specification dialog appears with the effects that Fit DSD determined to be important for Titer.

6. Ensure that Personality is set to Standard Least Squares and Emphasis is set to Effect Screening then click Run.
7. From the JMP output, determine what factors and terms are most important.

Now you will look at a combined model report for both Titer and Truncated Prod.

8. Click on Titer and Truncated Prod Model in your JMP Journal.

Can you find settings that maximise product titer and minimise truncated product?

9. Change factor settings in the prediction profiler to find conditions to maximise Titer and minimise Truncated Prod.

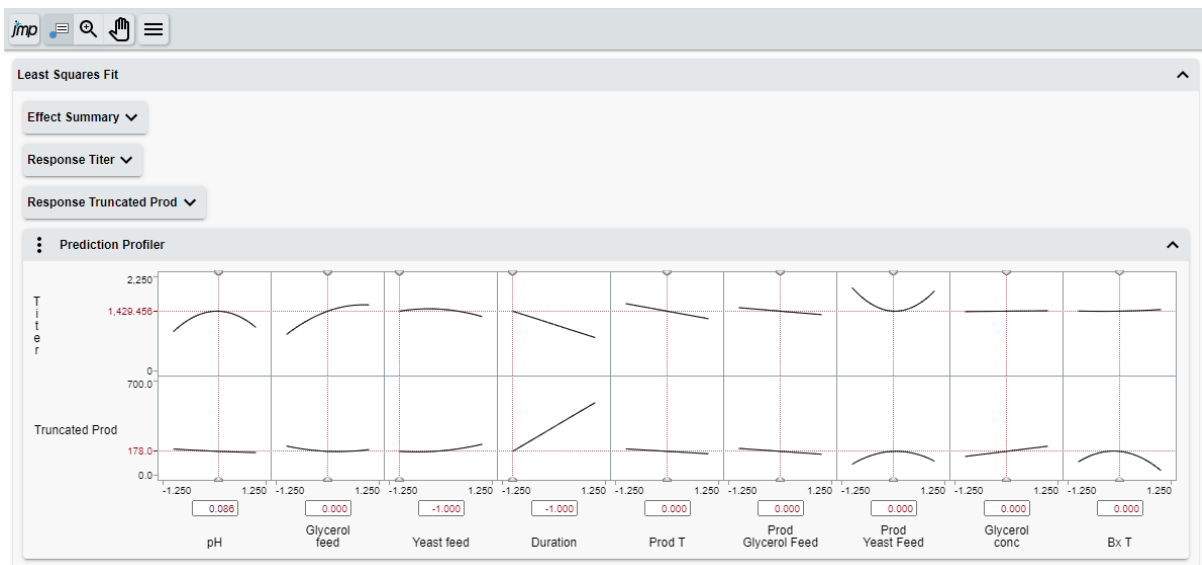
10. From the Prediction Profiler red triangle menu > Optimization and Desirability > Maximize Desirability.

We defined our goals (maximise or minimise) for the two responses when we designed the experiment. JMP can then find the most desirable settings to meet these goals.

Next you will create an interactive report that you can share with colleagues or clients

11. From the model report window: File > Save As > Save as type > Interactive HTML with Data. Save it to any location that you like.

When you open this in a browser you can see how you can interact with the profiler to explore the model.



Conclusion

We have a rich understanding of this process that we can easily share and communicate with colleagues and clients. We know the factors that are most important, how they affect Titer and Truncated Product, and how to control them to ensure the best process. We have learned much the same information as we did with the huge, full factorial experiment, but with MUCH greater efficiency.

The authors of this published example were able to find conditions that resulted in minimal truncated product and a product titer that was well above their original target. And they were able to validate that these conditions worked as expected at larger scales.

Design of Experiments is the most efficient and effective method for empirically learning about multi-factor systems, such as fermentation processes.