

Master's Degree Project Proposal:  
Optimizing Recycled Butanol Feed with Machine Learning

Steve Austin  
Louisiana State University  
Master of Science in Industrial Engineering.

Project Advisor:  
Dr. Gerald Knapp

## Introduction

This paper proposes the use of machine learning technology to optimize the feed rate of recycled butanol fed to the alumina purification section of a Ziegler Alcohol production facility. The goal of the optimization is to minimize the amount of recycled butanol injected into the system, while maintaining effective process conditions and meeting product quality specifications. The process conditions create a complex multivariate input criteria for the computation, which makes it a prime candidate for regression machine learning models to generate a function for a predictive process control module.

Machine learning (ML) is a “function approximation [technique that] can be viewed as a search through a space of hypotheses (possible representations of functions) for one that best fits a set of data [5].” ML develops a model to best fit a large dataset to predict future outcomes of the input variables. There are various applications of ML and scenarios where it is effective, but this paper predicates on the premise that very large amounts of data are too complex for human computation. Further discussion of machine learning is found in the Methodology section of this paper.

The Ziegler Alcohol process synthesizes fatty alcohols from ethylene using an organoaluminium compound [4]. The front-end process yields aluminum alkoxide, which is introduced to a hydrolysis reactor with water to separate the organic and inorganic compounds [4]. The compounds are phase separated in the hydrolysis reactor. The inorganic alumina/water (slurry) phase contains dissolved and entrained alcohol. The alcohol must be removed from the slurry phase to prevent environmental air emission excursions, prevent alcohol production losses, and reduce carbon in the final alumina product.

Recycled butanol is used to remove entrained heavy alcohols from the alumina slurry, and it too must be removed from the slurry before it is sent to the alumina process section. The slurry is sampled routinely at the purification section and at the downstream alumina production section. The results from the slurry samples indicate the number of heavy alcohols present. Final stage alumina product is also sampled routinely and analyzed to identify carbon content in the alumina.

The recycled butanol is injected into the alumina slurry purification section to extract entrained heavy alcohols from the slurry. Currently, recycled butanol feed rate is adjusted as a ratio to the alkoxide feed rate. According to operators familiar with the ratio controller, the system is operated at approximately 0.80 recycled butanol feed rate to alkoxide feed rate. Adjustments to the ratio controller are made when quality deviations are noticed by the control board operator in the alcohol samples analyzed but the QC every six hours.

## Literature Review

Previous research related to the same process area and sections was provided by this project's advising process engineer. The literature does not address the specific issues defined in this proposal, but it does describe data collection techniques used to build the operational regression models for those projects. Additional research and literature may be beneficial in later studies and will be included in the Literature Review section of the final project as required. Current comparative literature includes:

1. TSR R&D 22-037 01-22: Z1 Crystallite Size Control Model
2. TSR R&D 22-005 01-22, Z2 Crystallite Size Control Model
3. TSR R&D 22-022 01-22: Z2 Crystallite Size Control Model Based Upon TI030-53150 Reactor Slurry Temperature

The research within these reports is thorough and predicts the intended process variable with a high degree of accuracy. The techniques used in the listed research apply regression analysis through MS Excel Solver. The research indicates effective predictions were calculated using two input variables. Given the success of the data collection methods used in those studies, similar collections will be explored when developing a dataset for the ML algorithm of this project.

While the previous research found success using Excel Solver, the extent of data that will be used for this project will be better suited using a regression-based machine learning calculation. Also, the preliminary empirical evidence suggests this recycle butanol problem may require significantly more input variables. A larger multivariate input condition for regression correlation determination is better served by machine learning than Excel's Solver routine.

## Objectives

The primary objective of this project is to use machine learning for developing a model that accurately predicts the optimal recycled butanol ratio, based on various input parameters.

The optimal ratio is considered the least amount of recycled butanol required to maintain:

- effective process conditions,
- product quality specifications,
- and zero environmental impact.

## Methodology

Machine learning is useful in solving a variety of problems. In this project our problem involves developing a complex prediction function to optimize the rate of an auxiliary feed in a chemical process. These parameters require the use of regression type machine learning models. ML will search for the function that best fits the data it is given, but the ML task must be stated to ensure extract, transform and load actions generate the best available learning/testing dataset. A ML task is given by  $\langle P, T, E \rangle$ , where P = performance, T = task, and E = experience. For this project the task is defined as follows:

T: Predict optimal recycled butanol feed

P: Reduction in recycled butanol feed

E: Database of historical process conditions and database of sample results

### A. Data collection

Historical data of process conditions is available via Aspen SQLplus. In process and finished product sample data is available on LabWare. Data will be extracted from both databases and merged according to corresponding time data. The Aspen SQLplus database is designed to accept queries through Python API's. Given the size of the dataset that it available and the improved performance of ML over Excel solver for large data, the data will be queried and loaded into the ML environment through Python scripting language. LabWare may require extraction into a .CSV file, then merged into the Python environment.

Some data collection points to consider are based off previous research assumptions and assumptions determined through the work on this project. It is the intent of this project to collect data from stable operating conditions. These conditions will be identified by consistent levels and flows through the system. Significant deviations from stable conditions will be analyzed for cause and removed as necessary. Controlling for variability will be used to generate a baseline function, but it's effectiveness will also be evaluated against other than stable conditions.

### B. Data preprocessing

Missing values will be handled by removing its feature vector from the dataset. Outliers will be investigated to determine if they exist because of measurement or collection instrument anomalies. If they are found to originate from anomalies, they will also be discarded.

If features exhibit large skews in the preprocessed state, then transformation may be considered. The transformation will be necessary to prevent violation of the normality assumption of standard linear regression.

Normalization will be used to prevent one feature from overshadowing another. A feature with a very large range will have more effect on the ML training than one on the same dataset with a much smaller range. Normalization is a technique that scales the features to the same range so that the ML algorithm can approach them all with equal bias. The preferred method that will be used in this study is to normalize the features to Z-score equivalents.

### C. Feature engineering

The current control scheme assumes a relation between the primary feed and the recycled butanol feed. This assumption will be maintained in preliminary data collection and maintained throughout the project unless another variable enables better prediction performance. These variables and other variables analyzed in the Z1 crystallite size research will be considered to develop the training and test data. The variables considered in the previous research are:

- T, Hydrolysis Temperature, °F
- Alkoxide Feed
- Recycled Butanol Feed
- Slurry Flow, pph
- Slurry Density, from the butanol stripper bottoms stream.

There are several known process conditions that will affect the quality of the data and must be addressed to achieve a higher degree of success. Any change to the recycled butanol ratio controllers set point will take more than one sample cycle to observe all downstream process changes due to the cumulative residence time incurred by the slurry in the hydrolysis reactor, the alumina purification section, and the surge tanks of the alumina production units. Therefore, times of changes to the ratio controller's set point may need to be adjusted to times of sample collections to better observe changes in the process.

The aluminum alkoxide feed is defined by an *m-value*, which gives the average number of carbon chains present, across the entire distribution of alcohols present in the feed. Since this metric is the average across a Gaussian distribution, it's possible that feed compositions can skew to heavier alcohols, which implicates the need for significantly different recycled butanol rates for the same rates of aluminum alkoxide feed. The time variable, feed composition, and other variables discussed later, will be mined from the historical database, and controlled for stable unit conditions to provide an effective dataset for the chosen machine learning algorithm.

The assumption garnered through the process of generating this proposal is the effect of the composition of the alkoxide feed. Research indicates front end processing is optimized for alcohol chain distribution in the alkoxide feed. These parameters will be analyzed against sample results showing alcohol distributions. Care must be taken as the alkoxide is surged to a storage tank where the in-process distribution will be altered by the storage tank's distribution. A better sample point may be an alcohol content sample from the hydrolysis reactor, if one exists. The variables in the front-end process that may affect the composition

will be discussed with operators and engineers familiar with those sections. Further discussions with process subject matter experts may lead to the inclusion or exclusion of other variables.

#### **D. Model selection**

A regression ML model will be used to analyze the data collected for learning. This project will start with the assumption that the model is linear and use a linear regression model for initial training cycles. However, the prediction may perform better with non-linear regression models. Therefore, both linear and non-linear models will be evaluated on the dataset. The following is a list of models to train:

- Linear regression (lm)
- Generalized linear models (glm)
- SVM regression (ksvm)
- Neural Networks (neuralnets)
- Nonlinear models (nlm)

#### **E. Model training and evaluation**

**Training** will take place in the Python environment because of its efficiency with a large dataset. Large datasets will produce the best performance from the ML models. To prevent the model from becoming biased on the training data is to implement cross-validation. Cross validation divided the data into multiple train/test sets instead of just one set. This gives a better indication of the model's performance on new, unseen data.

**Evaluation** of the model will begin with calculating R-Squared values, which provides the fraction of total variance explained by the model. From the R-squared metric an F-distribution is derived, and from the F-distribution a P-value is derived. The P-value will allow for conclusion of the existence of a linear relationship. If this model's input will require a multivariate combination, then an adjusted R-squared calculation will better describe the models fit to the dataset.

Further determination of performance with a regression model can be observed with confidence intervals. The confidence intervals indicate the difference between the predicted values may be from the true underlying model. This metric will demonstrate the accuracy and reliability of the model.

### **Expected Results**

The expected accuracy of the machine learning model is difficult to speculate given the complexity of and expected number of input variable. General knowledge about the process from subject matter experts is that the recycled butanol is being used in excess because operators rather than the upset conditions that occur when there is not enough recycled butanol in the system. Therefore, it is the stance of this paper that significant reductions in recycled butanol will be attained and observed through the development of DCS controller using the machine learning generated function.

## **Conclusion**

This project intends to use machine learning to optimize the feed rate of recycled butanol in an alumina purification section of a Ziegler alcohol processing facility. The facility maintains adequate historical data to achieve a high level of performance from the model. Data will be collected, transformed, and loaded into the model using the Python scripting language and embedded SQL queries. The significance of this project is its potential for increased energy efficiency through reduced demands for recycled butanol. The lessons learned in this project and potential energy saving functions can be applied to other process in this facility and others thereby significantly reducing greenhouse gas emissions across the industry.

## References

- [1] D. Barclay, TSR R&D 22-005 01-22, Z2 Crystallite Size Control Model, January 18, 2022
- [2] D. Barclay, TSR R&D 22-022 01-22, Z2 Crystallite Size Control Model Based Upon TI030-53150 Reactor Slurry Temperature, Mar. 2, 2022
- [3] D. Barclay, TSR R&D 22-037 01-22: Z1 Crystallite Size Control Model
- [4] [https://en.wikipedia.org/wiki/Ziegler\\_process](https://en.wikipedia.org/wiki/Ziegler_process)
- [5] G. Knapp, Industry 4.0 slides, Module 7 Section 1