# Project Kojak Summary
# Predicting Industrial Chiller Performance

Dana Lindquist
December 11, 2018

**Project Summary**

Industrial utility plants generate chilled water to buildings for manufacturing or air-conditioning. They generally consist of several chillers which facilitate the heat transfer between a loop which sends cold water to the buildings which comes back warmer and a second water loop to dissipate the heat in cooling towers. This project is predicting the performance curves for these industrial chillers.

The data comes from Optimum Energy, a local Seattle company that optimizes chiller plants. They collect data on chiller plants and suggest operating points for the plant for the lowest energy consumption. At this time they do not use machine learning for the optimization.

The dataset consists of 30 York YK dual-pass chillers in 8 distinct configurations ranging from 853 to 1875 Tons (a Ton is a unit of refrigeration power - the energy required to take 1 Ton of water at 0°C to freezing in 24 hours). The goal of this project is to predict performance curves based on characteristics of the chillers.

**Data**

The data from Optimum Energy was provided in csv files. Ultimately, the data was separated to create one file for each individual chiller. Points when the chiller was in alarm, not running or just starting or stopping were removed from the dataset.

For each chiller there were timestamped points for the following:

kW/Ton
Load
Chilled Water Return Temperature
Condenser Water Supply Temperature
Temperature Lift
Ton
Evaporator Approach Temperature
Condenser Approach Temperature
Flow Rate
Compresser Superheat
Inlet Guide Vane position
* Refrigerant Level

* do not have data for all Chillers in dataset

For each chiller I was also provided provided the following chiller characteristics:

Model Number
RatedTons
VarSpeed
Rated kW
* Rated FLA
Condenser & Evaporator
    * Minimum Flow
    Design Flow
    * Maximum Flow
    Pressure Drop
    Entering Water Temperature
    Leaving Water Temperature

* do not have data for all Chillers in dataset

**Methodology**

One chiller

I started by looking at just one distinct chiller configuration and performed a linear regression to determine the performance curves. I had some background information from Optimum Energy and expected the kW/Ton (the target) to be a function of Load, Lift (ΔT) and their higher powers. They assume that all of these chillers operate with the same performance curves but are aware that this is not always the case.

I started with the linear regression

$$kW/Ton = C_1 + C_2{*}Lift + C_3{*}Load + C_4{*}Lift^2 + C_5{*}Load^2 + C_6{*}Load{*}Lift + C_7{*}Load^2{*}Lift + C_8{*}Load{*}Lift^2$$

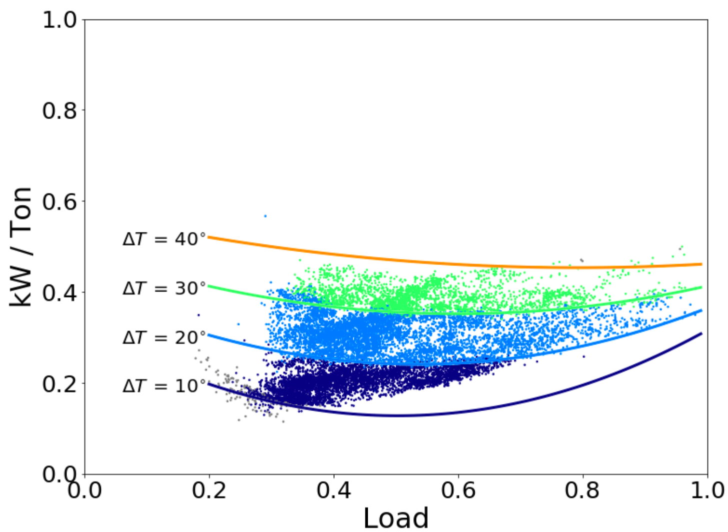
Using LASSO L1 Regularization I determined that the $Lift^2$ terms did not contribute to the linear regression and could be dropped so the linear regression was now

$$\begin{aligned} kW/Ton &= C_1 + C_2{*}Lift + C_3{*}Load + C_5{*}Load^2 + C_6{*}Load{*}Lift + C_7{*}Load^2{*}Lift \\ &= (C_1 + C_2{*}Lift) + (C_3 + C_6{*}Lift){*}Load + (C_5 + C_7{*}Lift){*}Load^2 \end{aligned}$$

Which is a quadratic for constant Lift.

The S plant had four identical Chillers. I had the most data points for the first chiller so I trained the model on Chiller S1 and tested on Chillers S2, S3 and S4. The $R^2$ values for the training and test sets were just barely under 1.0. This model predicted the chiller characteristics very accurately.

| | Accuracy |
|---|---|
| Chiller1 (training set) | 0.987 |
| Chiller 2 | 0.984 |
| Chiller 3 | 0.968 |
| Chiller 4 | 0.988 |




<u>Multiple Chillers</u>

The goal of the project was to predict the linear regression coefficients for the chiller performance curves based on the chiller characteristics.

I then added to the linear regression terms

- Rated Tons
- Variable Speed (0 for no and 1 for yes)
- Condenser Design Flow
- Condenser Design Pressure Drop
- Condenser Entering Water Temperature
- Condenser Leaving Water Temperature
- Evaporator Design Flow
- Evaporator Design Pressure Drop
- Evaporator Entering Water Temperature
- Evaporator Leaving Water Temperature

And the above times Load, $Load^2$, Lift, Load*Lift and $Load^2$*Lift

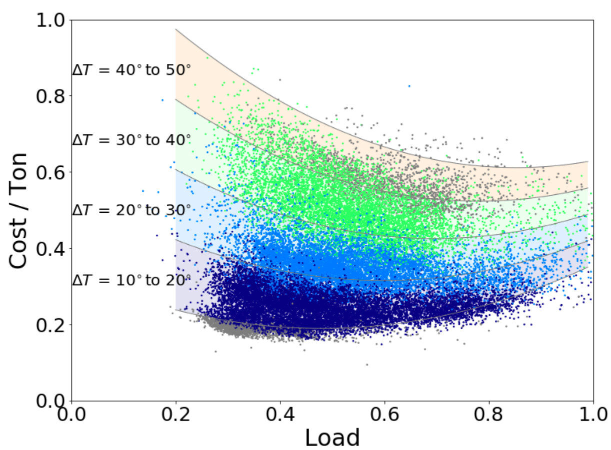
Using LASSO L1 Regularization I determined that the terms worth keeping were

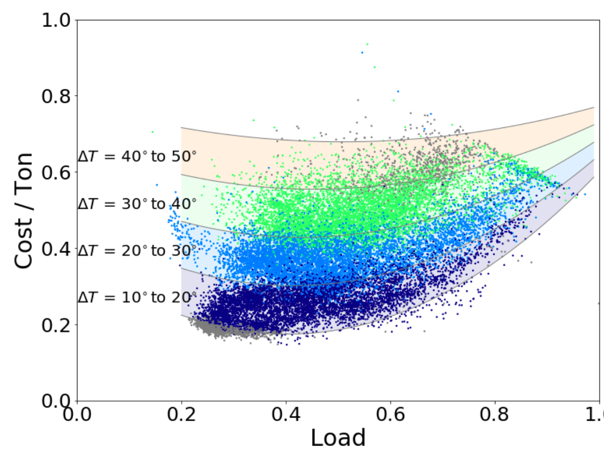
Rated Tons
Variable Speed (0 for no and 1 for yes)
Condenser Design Flow
And the above times Load and $Load^2$

As well as
Evaporator Design Flow and Evaporator Pressure Drop times Load and $Load^2$.

With these terms I was able to predict the performance curves for different chillers.

Below are two different 900 ton chillers, B1 and B4. The bands are 10° ranges in ΔT Lift. The dots are the actual data for these chillers. The model was able to predict the performance curves at just over $R^2 = 0.9$




<u>PYMC3</u>

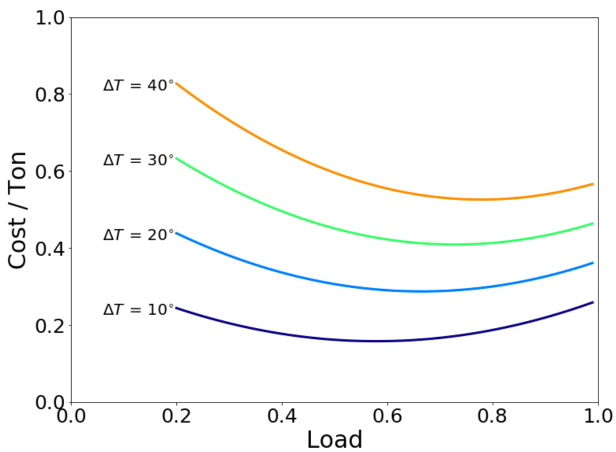
Since the chiller characteristics are constants for each chiller, I would like to have a model with nested linear regressions.

$$\begin{bmatrix} | \\ kW/Ton \\ | \end{bmatrix} = \begin{bmatrix} | & | & | & | & | & | \\ 1 & Load & Load^2 & Lift & Load \cdot Lift & Load^2 \cdot Lift \\ | & | & | & | & | & | \end{bmatrix} \begin{bmatrix} A \\ B \\ C \\ D \\ E \\ F \end{bmatrix}$$

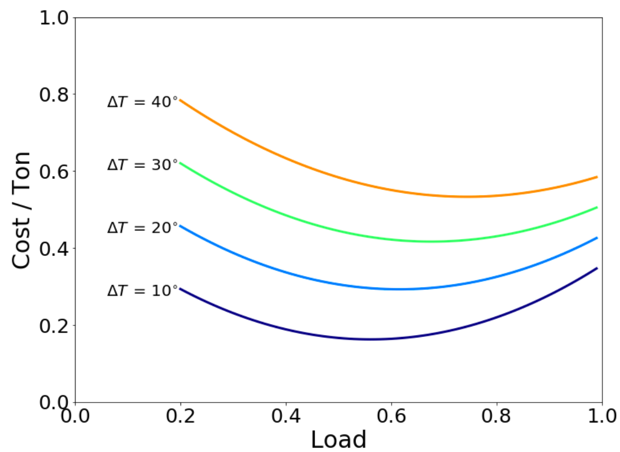
Where

$$
\begin{bmatrix} A \\ B \\ C \\ D \\ E \\ F \end{bmatrix} = \begin{bmatrix} a_1 & a_1 & a_3 & \dots \\ b_1 & b_1 & b_3 & \dots \\ c_1 & c_1 & c_3 & \dots \\ d_1 & d_1 & d_3 & \dots \\ e_1 & e_1 & e_3 & \dots \\ f_1 & f_1 & f_3 & \dots \end{bmatrix} \begin{bmatrix} | \\ Chiller \\ Characteristics \\ | \end{bmatrix}
$$

This model can be created using a Bayesian hierarchical model in PYMC3. I was able to re-create the linear regression model for one type of chiller using PYMC3 but did not have the time to implement the heuristic model which would have included the nested linear regressions. Below are the performance curves for the B1 plant using both methods.




Linear Regression Model




PYMC3 Model

It was necessary to normalize and center the data for the pymc3 model using StandardScaler

<u>Plant Optimization</u>

Given performance curves for a chiller, it is now possible to determine the best way to operate a plant (a collection of chillers) to run at the lowest cost.

Using optimize.minimize from the scipy library, it was possible to determine the configuration which would minimize the Cost (kW) for a collection of chillers.

As a constraint, assume a given total Tons and ΔT Lift is required. You can feed optimize.minimize a set of chiller curves determined by the linear regression model and the constraint that the tons from each chiller must add to the given total tons.

For example, assume 2400 tons are required at 30° ΔT Lift.  If you have two B1 900 ton chillers and two C1 1200 ton chillers then to get the minimum kW you should run

| | | |
|---|---|---|
| 1 | B1 900 Ton Chiller | Don't use |
| 2 | B1 900 Ton Chiller | 522 tons |
| 3 | C1 1200 Ton Chiller | 939 tons |
| 4 | C1 1200 Ton Chiller | 939 tons |

**Conclusions and Future Extensions**

It was possible to predict the performance curves for a chiller given the chiller characteristics.  For future work I would like to

1. Incorporate more chillers in the dataset.  This would require getting more data from Optimum Energy
2. Further the work using PYMC3 to add the model characteristics.  This should be fairly straight forward but will require some compute time to tune the model.

I will be presenting this work to Optimum Energy in January.

**Packages Used**

pandas

numpy

scikit-learn
    LinearRegression
    train_test_split
    Lasso
    mean_squared_error
    StandardScaler

spicy
    Optimize
matplotlib

pickle