Linear Regression Analysis for the EIC Test Stand Thermal Data

Pablo Campero
2023-06

Data was acquired from the EIC test stand measurements and from its equivalent simulated model in Ansys-Fluent. I performed a linear regression model where the silicon pipe temperature from Ansys-Fluent was referred as the independent or explanatory variable and the data from the test stand measurements was referred as the dependent, observed or process variable. The linear regression was used to show the relationship between the two mentioned variable by fitting a line to estimate the value of the response.

I wrote a python program to calculate the linear regression parameters including the including the R-squared value and also the residual plots. The linear regression identifies the equation that produces the smallest differences between all the observed values and the fitted values, see Fig.1., to be precise linear regression finds the smallest sums of squared residuals that is possible for the dataset. The dataset is shown in table 1.

The regression model fits the data well when the differences between the observations and the predicted values are small and unbiased (means in this contend that the fitted values are not systematically too high or too low anywhere in the observation space).

#	SLM	Ansys Temp [°C]	Observed Temp [°C]
1	0	49.4	51.5
2	10	46.99	54
3	20	44.58	49.5
4	50	37.38	41.5
5	100	33.37	34.5
6	150	31.77	30.5
7	200	30.18	29
8	250	29.38	27.5

Table I. Dataset

- **Used linear regression model to show the relationship between the Silicon pipe temperature acquired from the Ansys-Fluent and from the test stand measurements**
- **Written a Python program to calculate the R-squared value**
- **Generated plots to show best-fit line with the scatter points and residuals**
- **Analyzed linear regression model results**

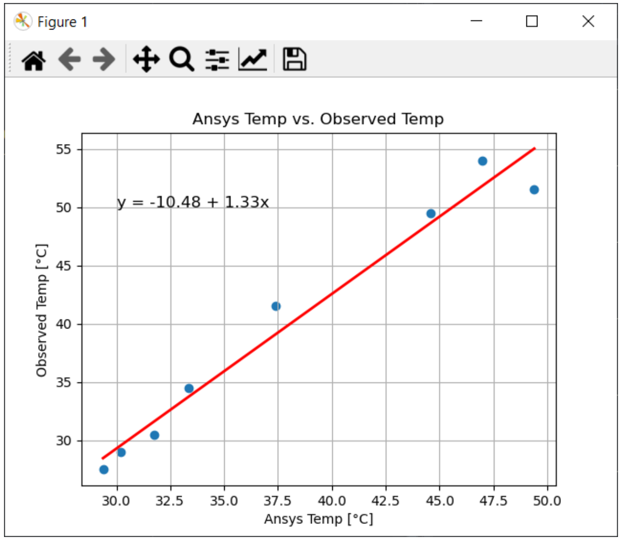

Fig. 1 Regression model plot with the Ansys-Fluent results and observed temperature values, the red straight line in the plot is a best fitted line with a polynomial **y = -10.48+1.33x**

Linear Regression Analysis for the EIC Test Stand Thermal Data

Before assessing numeric measures of goodness of fit, such as the R-squared value, I evaluated the residual plots, residual plots expose the bias model far more effectively than the numeric output by displaying problematic patterns in the residuals. The residual plot is used to assess whether or not a linear regression model is appropriate for a given dataset and to check for heteroscedasticity (same finite variance) of residuals.

I calculated the R-squared value to evaluate the scatter of the data points around the fitted regression line, this value is measured in a range between 0 and 100 %. For the same dataset, higher R-Squared values represents small differences between the observed and the fitted data, and small R-Squared values represents a model that doesn't explains all the variation in the response variable around its mean. The calculated R-squared value for the regression model was 96.72 %. Result shows that the regression model can explain the variation in between the measurements and the simulated Ansys-Fluent results.

Additionally, I generated a QQ plot, which is a graphical tool to assess if the set of data plausibly came from some theoretical distribution such as a normal or exponential. If the data values in the plot fall along a roughly straight line at a 45-degree angle, then the data is normally distributed. From the plot, show in Fig.3., it was determined that the residuals follow a normal distribution.

I completed the data analysis with the implementation of a linear regression model generated in python, which was able to explain the relationship between the test stand measurements and the Ansys-Fluent results for the EIC silicon pipe temperatures. The residual plots and high R-squared values confirmed the goodness-of-fit measured for linear regression model.

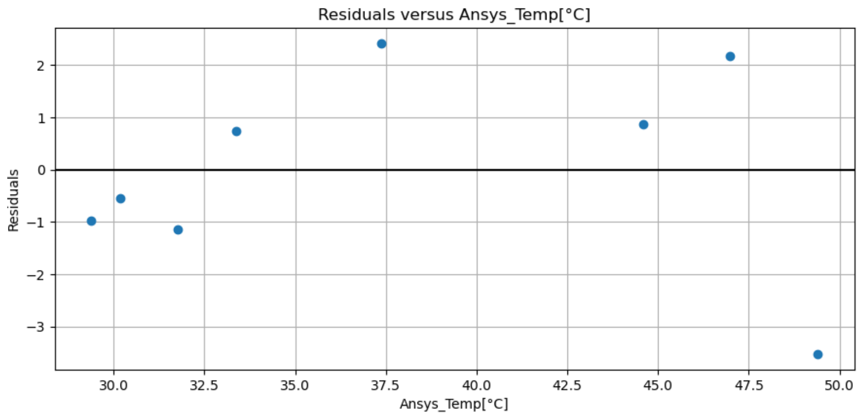

Fig.2. Residual plot

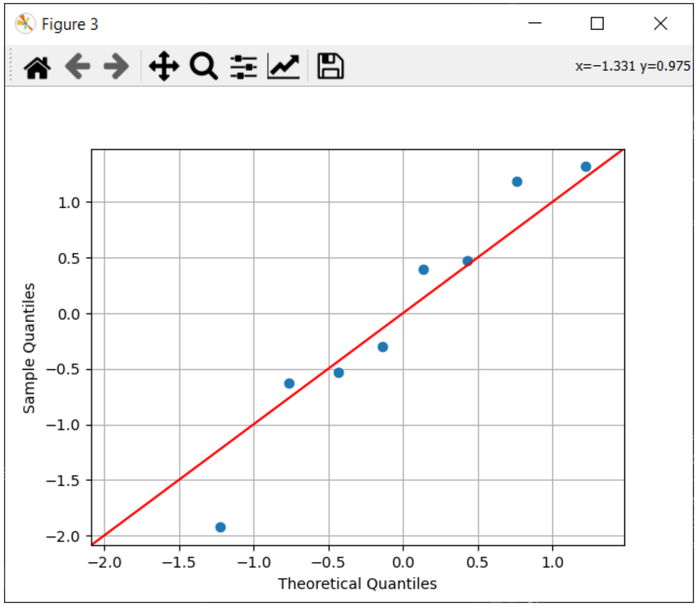

Fig.3. Q-Q Plot