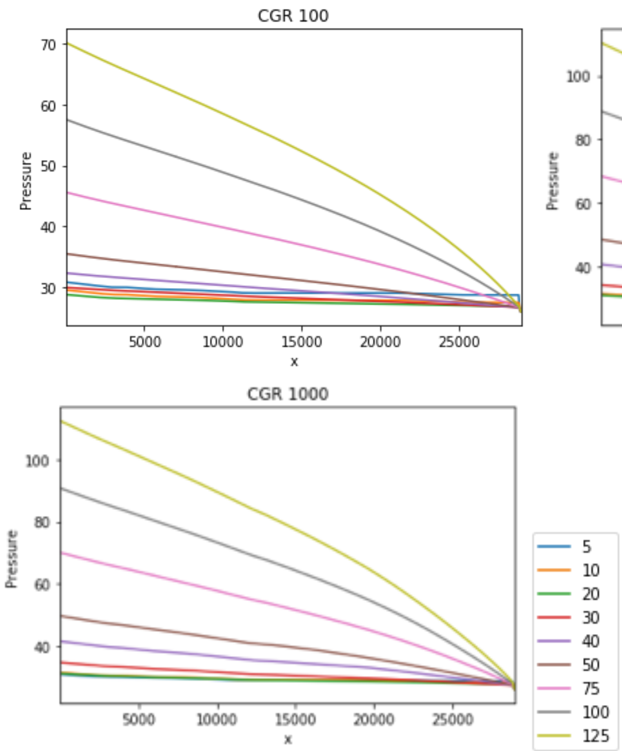
Table of Contents

Exploratory data analysis

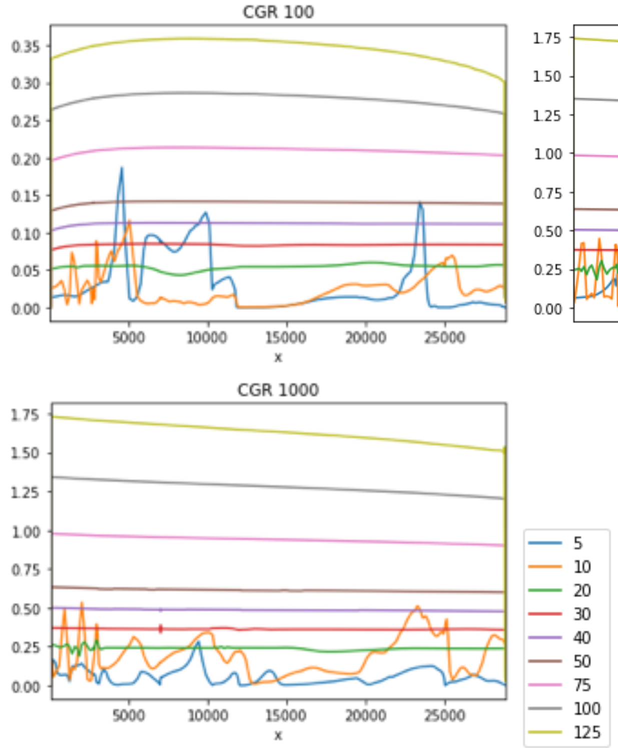
Pressure



Across the length of the pipeline, the pressure drops and eventually converges to a specific value for all 3 CGR. Interestingly, the trends and values corresponding to CGR 500 and CGR 1000 are very similar.

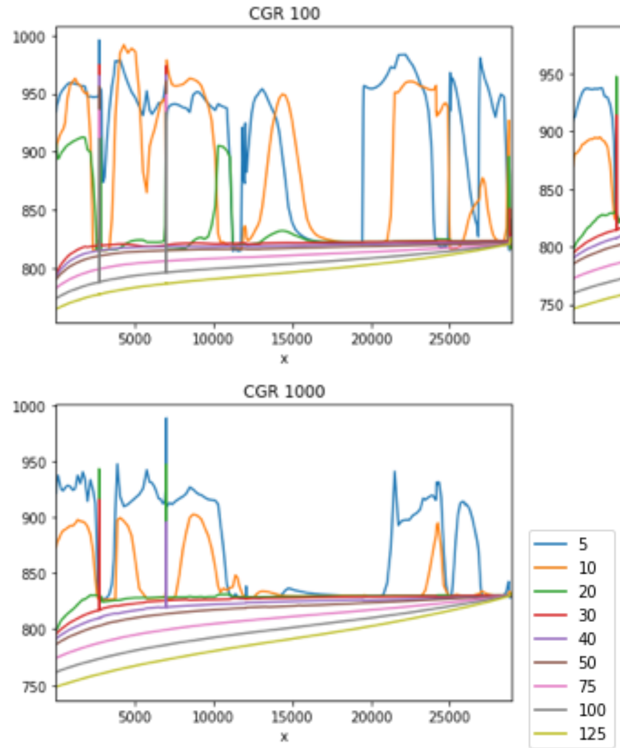
Liquid phase

Superficial liquid velocity

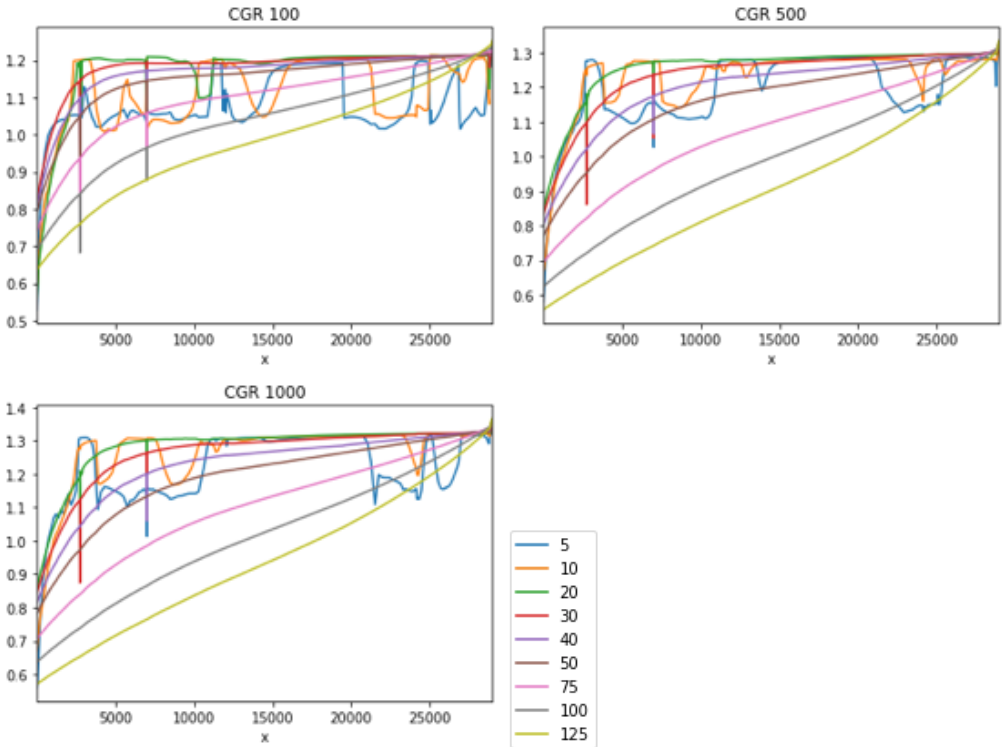


At low gas rates (5 and 10 MMscf/d), the liquid velocity fluctuates intensely and at some locations it even becomes negative. If negative velocity is left untreated, then the mixture density, viscosity, and Reynolds number will turn out negative - which does not make any physical sense since those values are scalar. To handle this, I converted any negative velocity values to their absolute values.

Liquid density



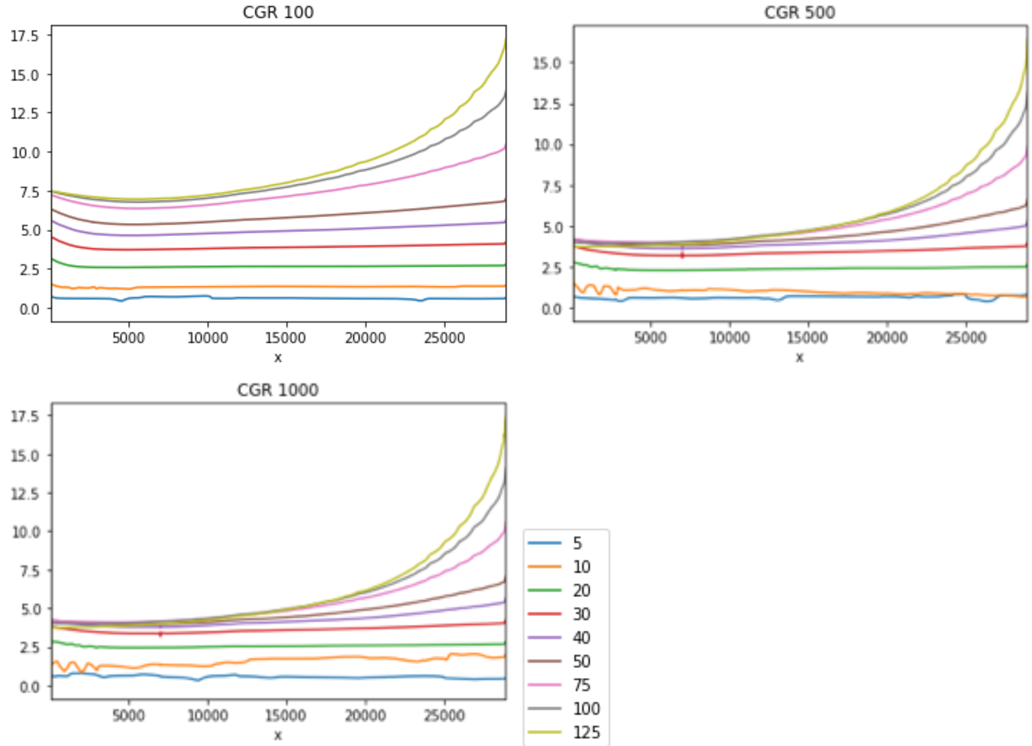
Liquid viscosity



Similar fluctuations in the variables are observed at lower gas rates. Both for liquid density and velocity we see a convergence to a specific value at the end of the pipeline.

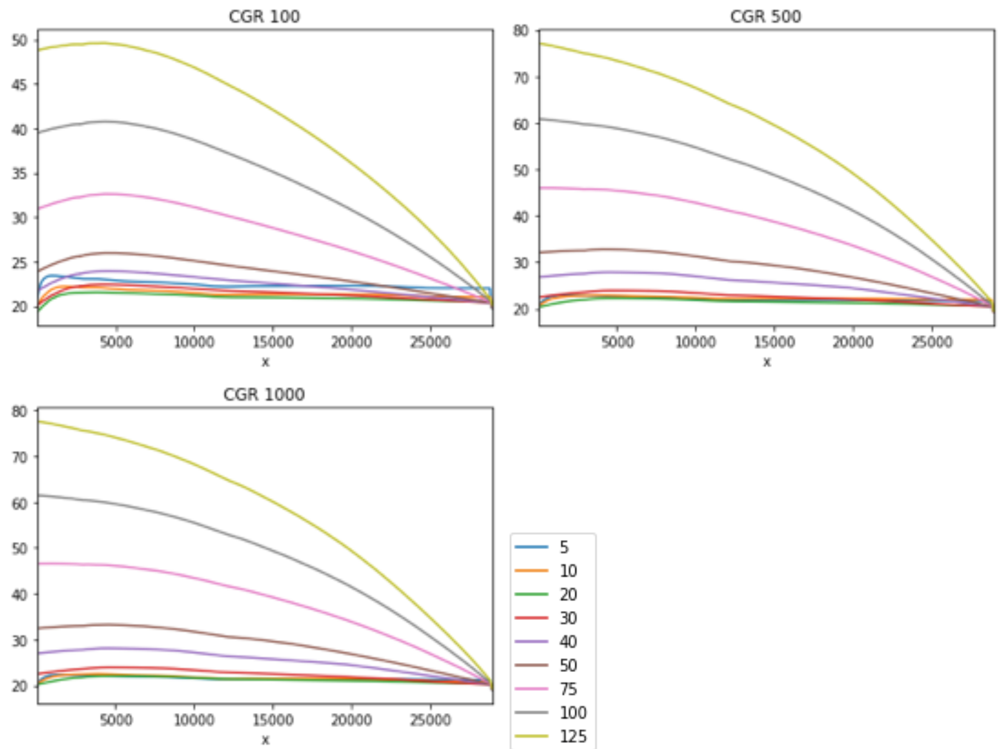
Gas phase

Superficial gas velocity

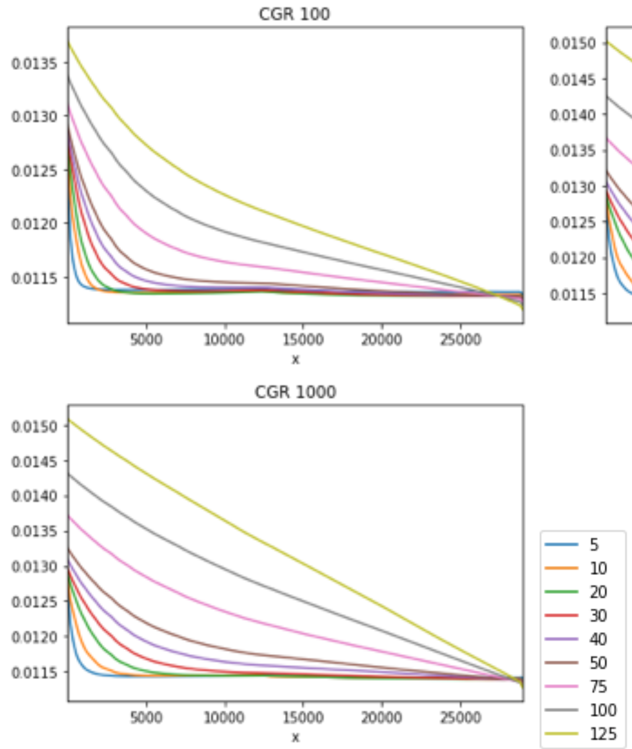


Towards the end of the pipeline (x>22000), the gas velocity appears to diverge quickly for higher gas rates (>75 MMscf/d)

Gas density



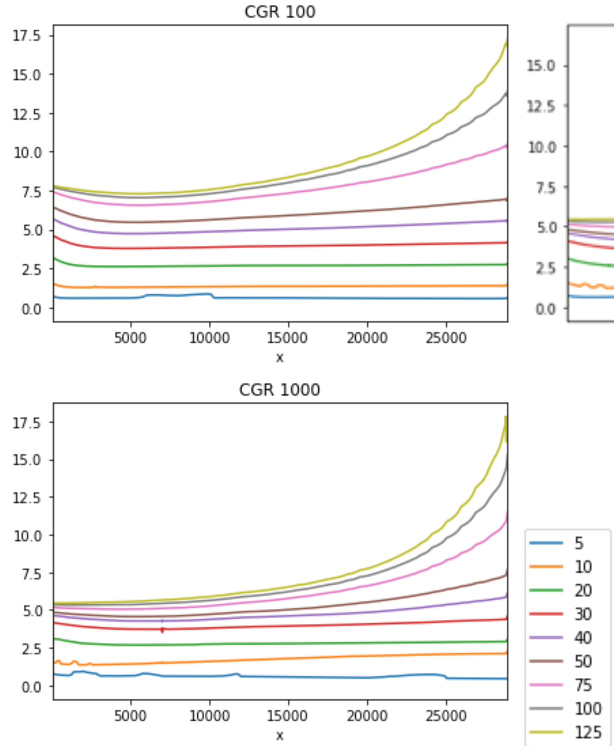
Gas viscosity



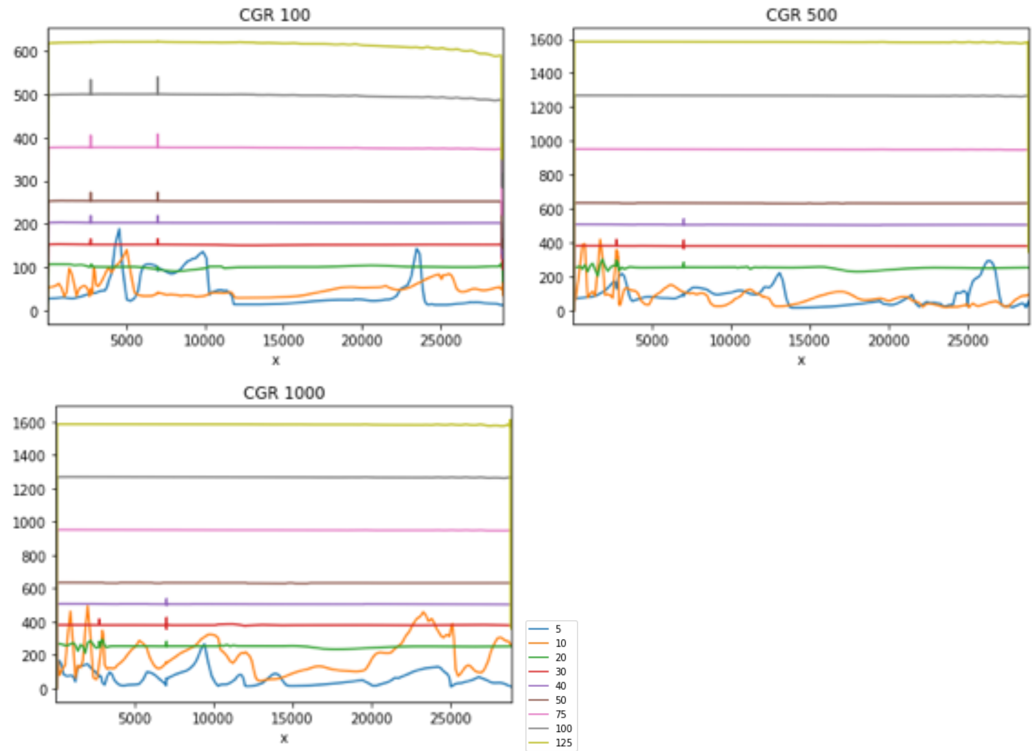
The gas density and viscosity appear to converge towards a specific value towards the end of the pipeline. No fluctuations in density and viscosity were observed at low gas rates, unlike the liquid phase. This is reasonable given the larger quantities in the volume of the gas present.

Mixture

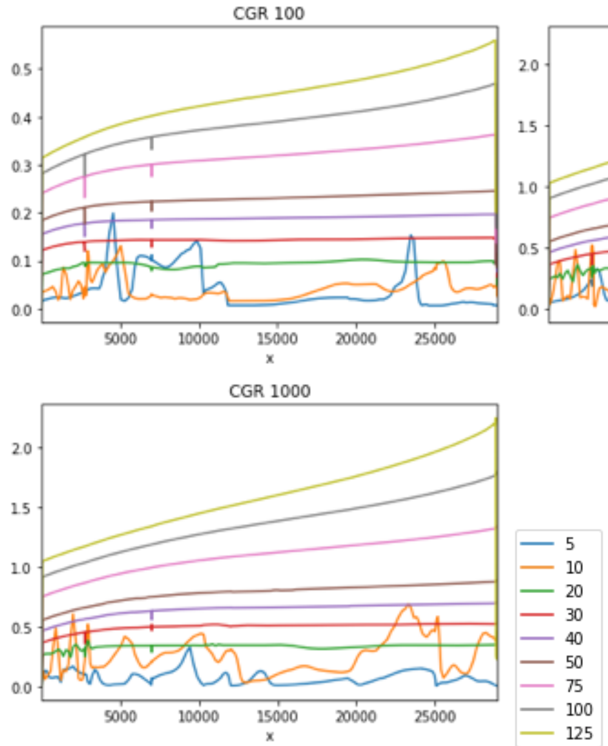
Superficial velocity



Density

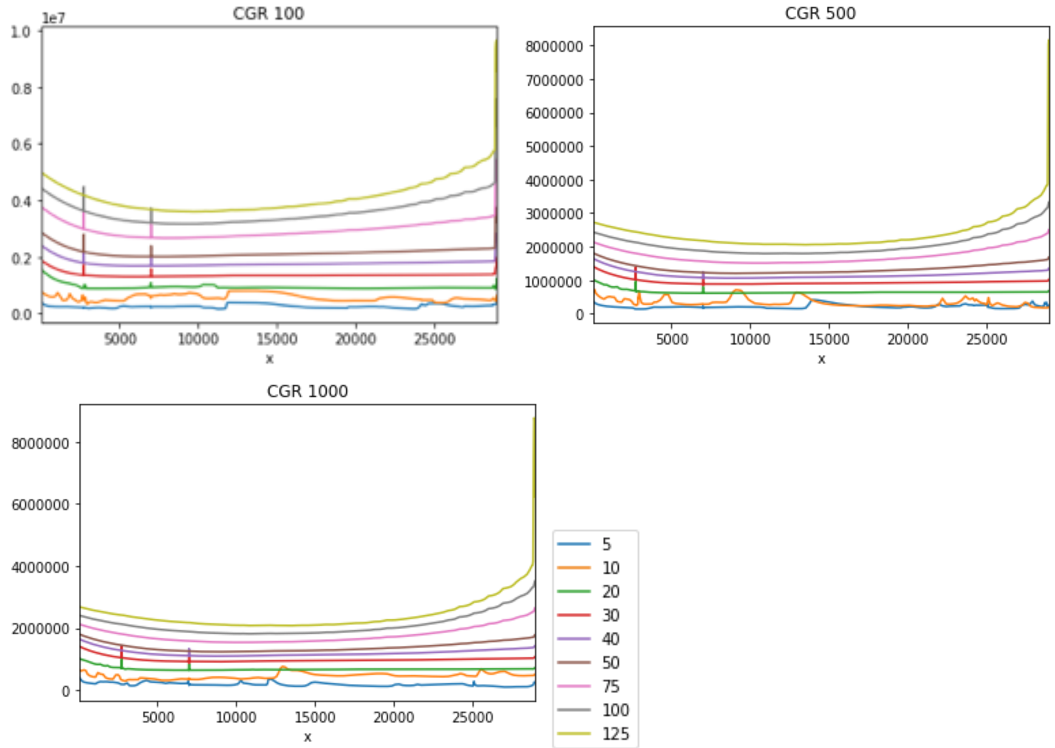


Viscosity



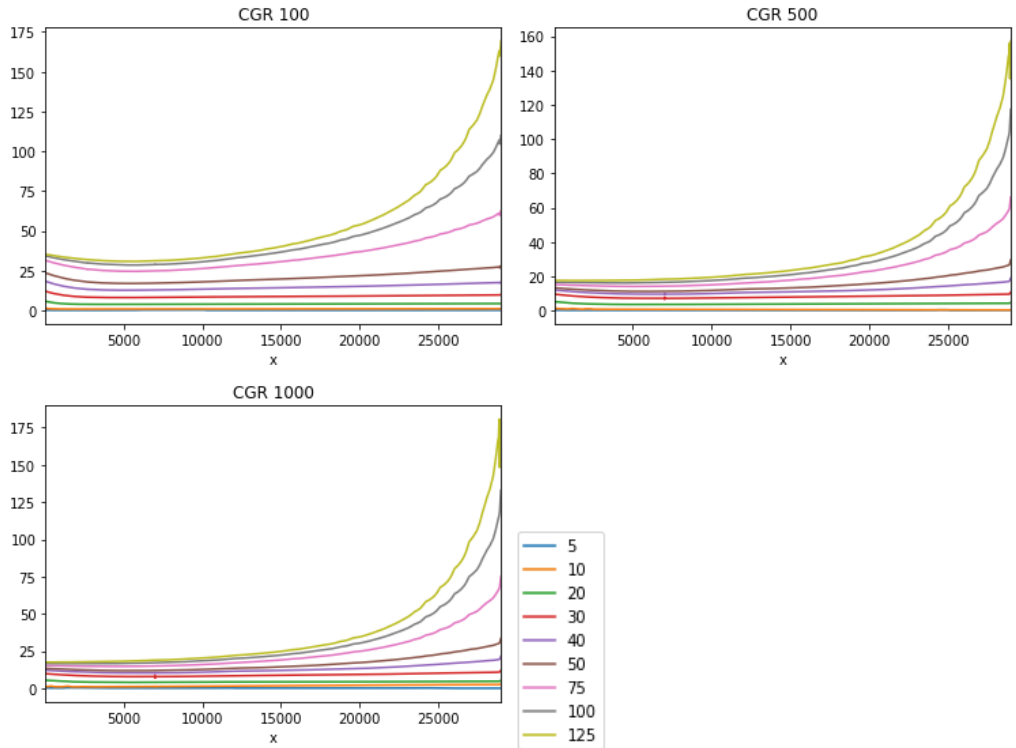
The mixture velocity diverges for larger gas rates (>75 MMscf/d). The mixture density remains relatively the same across the length of the pipeline - with the exception of trends corresponding to lower gas rates which fluctuate. The mixture viscosity diverges at larger gas rates (>75 MMscf/d), while at small gas rates (5,10) the trend fluctuates. Note: at the end of the pipeline, values of viscosity and density show abrupt drops for higher gas rates, which affects some of the derived quantities like Reynolds number.

Reynolds number



Sudden jumps in Reynolds number are observed at the end of the pipeline for larger gas rates, which were caused by sudden drops in various quantities at the denominator of the Reynolds number formula. The rows with outliers have been removed to prevent them from affecting the regression later.

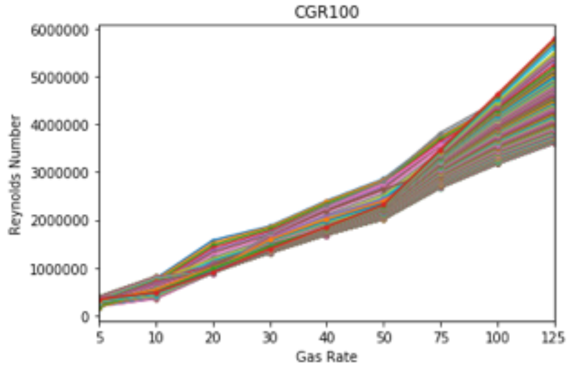
Froude number

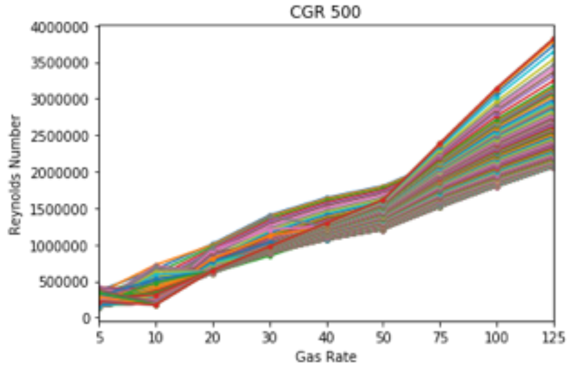


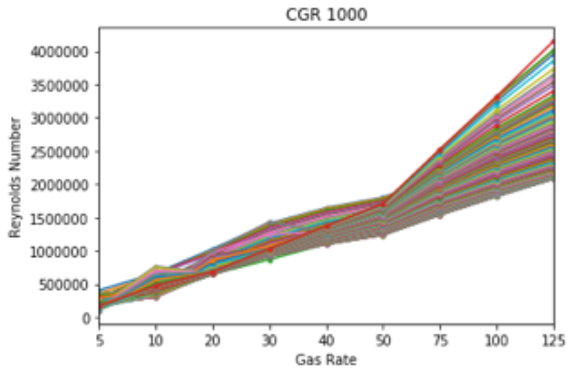
Towards the end of the pipeline the Froude number appears to diverge rapidly for higher gas rates (>75 MMscf/d).

Reynolds, Froude and Gas Rate at various locations across the pipeline

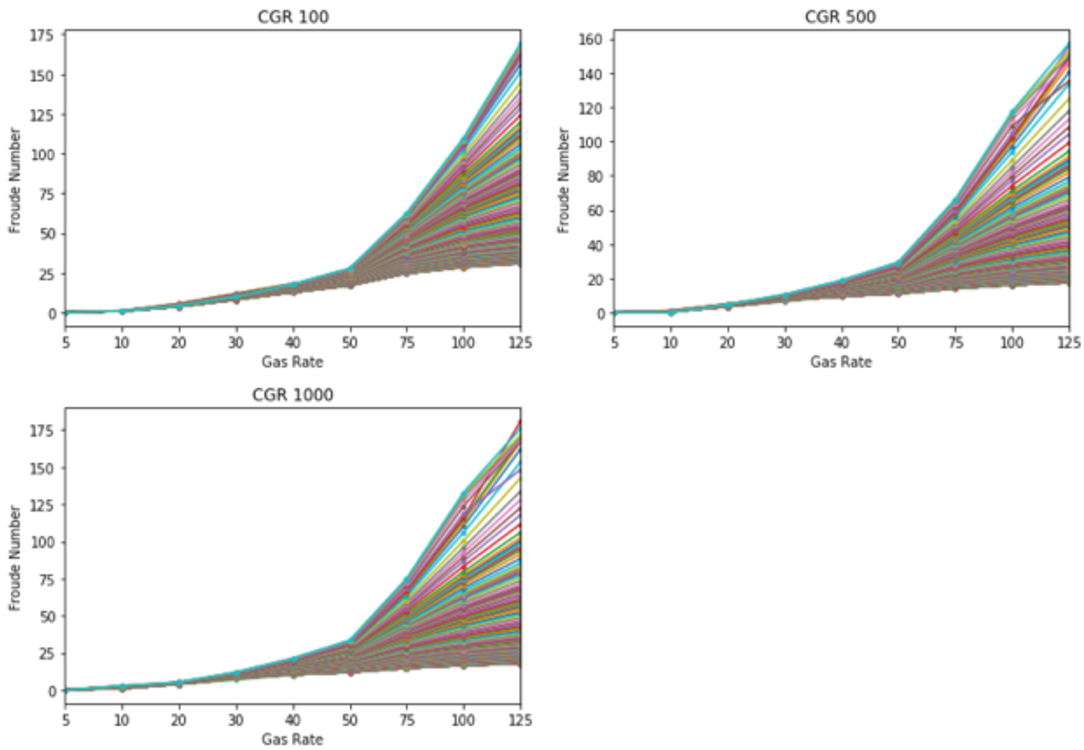
Reynolds-Gas Rate







Froude-Gas Rate



Each line in the above diagrams represents a location along the pipeline. The legend was omitted due to its size.

In the graphs we can see an overall positive increase of the Reynolds and Froude numbers as the gas rate increases. For gas rates >50 MMscf/d the rate of increase appears to be more intense for the majority of locations along the pipeline. The trend appears to vary for different locations across the pipeline. That is reasonable to some extent, given the inhomogeneous nature of the gas/liquid phase.

In the next section, we take a suubset of these curves, plot them individually, and fit functions to assess which type of function fits best overall.

Regression analysis

The simulation outputs provide information at various locations along the pipeline (270 distinct locations in total). To visualise regression, I will sample the dataset for conciseness (i.e., select a subset of those 270 locations). Then, plot the trend for that output and fit a function.

To perform the regressions I used the '[lmfit'](lmfit) package and the '[scipy.optimize'](scipy.optimize) library in python. Both methods, minimize error using the *Levenberg-Marquardt* least squares technique. The following regressions have been performed:

- exponential model (through the lmfit package)
- polynomial model (through the lmfit package)
- power-law model (through the lmfit package)
- logistic model (through the lmfit package)
- logarithmic model (through the scipy.optimize library)

To assess the goodness of fit, I plotted the residuals from each regression. A good fit would imply that the residuals approximate the random errors that make the relationship between the explanatory variables and the response variable a statistical relationship. Therefore, if the residuals appear to behave randomly, that suggests that the model fits the data well. However, a non-random structure in the residuals, implies that the model fits the data poorly.

Additionally, I quantified the sum of squared error (SSE) of each fit, to assess the overall goodness of fit.
The code and detailed plots for each function, alongside their respective plot of residuals and SSE can be found in the notebooks: **cgr100.ipynb, cgr500.ipynb, cgr1000.ipynb**.

Some functions are better suited for locations closer to the start of the pipeline, whilst some others are better suited halfway through or towards the end of the pipeline. This is something unsurprising given the nature of the liquid/gas phase transformations taking place in the pipeline.
Deciding on a universal model that captures trends throughout the length of the pipeline, appears to be a complex task.

In summary, the best models for fitting the Reynolds number are polynomials (e.g. quadratic or higher) and logarithmic regression. At some locations along the pipeline, the logarithmic fit is better than the higher order polynomial and vice-versa. This can be justified given the inhomogeneity in the pipeline as a result of the dual phase flow.
As a note, the higher the order of the polynomial, the better the apparent fit on existing data. However, that needs to be treated with caution as it could lead to overfitting – i.e., failure of the model to generalise well to new data. To assess model overfitting, more data are needed to perform cross-validation.

The best fit for Froude number appeared to be the higher order polynomial. While analysing the trends, I also experimented with the fitting of a logistic function which is defined by a sigmoid shape. The fit appeared good at some locations; however, the higher order polynomial appeared to have more randomly dispersed residuals and a lower sum of squared error, overall.

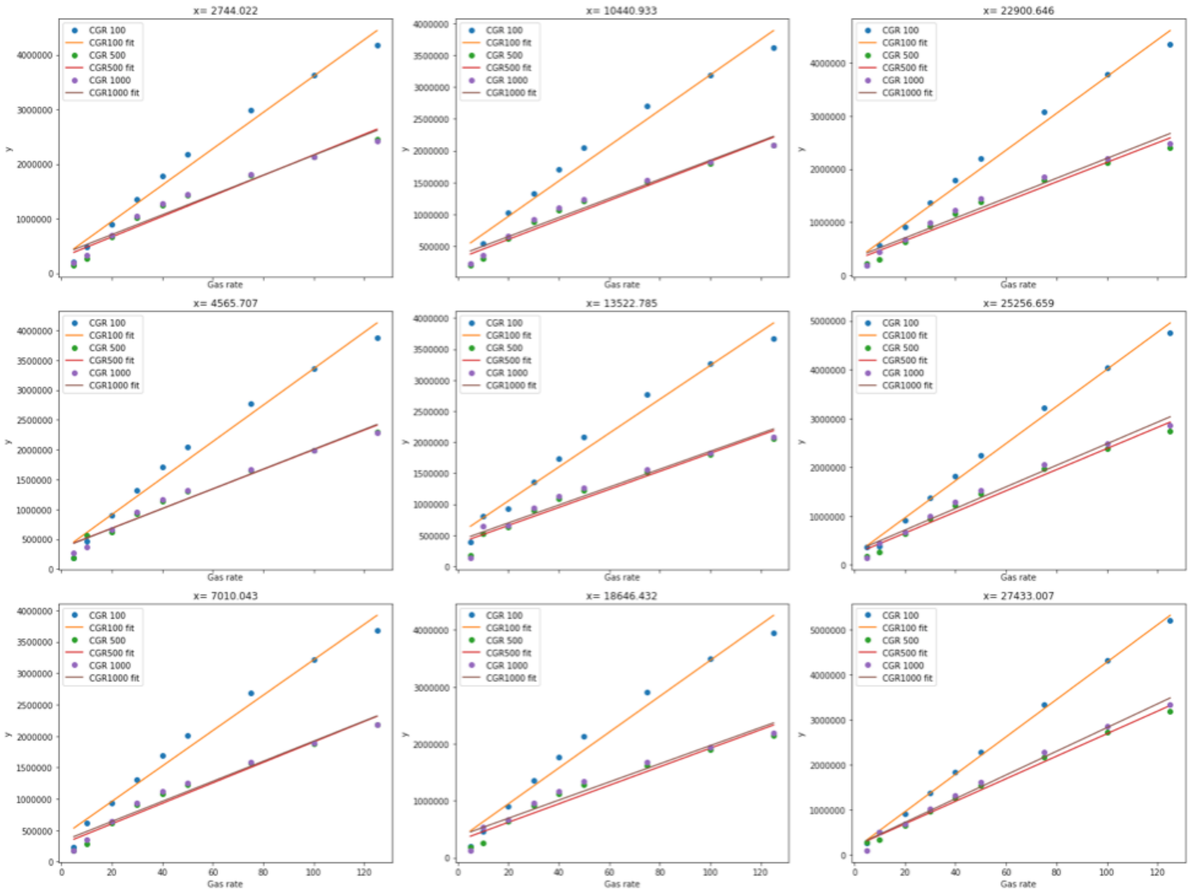
Below I am summarising the fitting of different models on Reynolds and Froude number, for the 3 CGR data as a function of Gas rate. The code for the plots that follow can be found in the **cgr100-500-1000.ipynb** notebook.

From the plots below, it is interesting to notice the proximity and similarity in the trends between CGR 500 and CGR 1000.

Reynolds-Gas Rate

Polynomial Regression

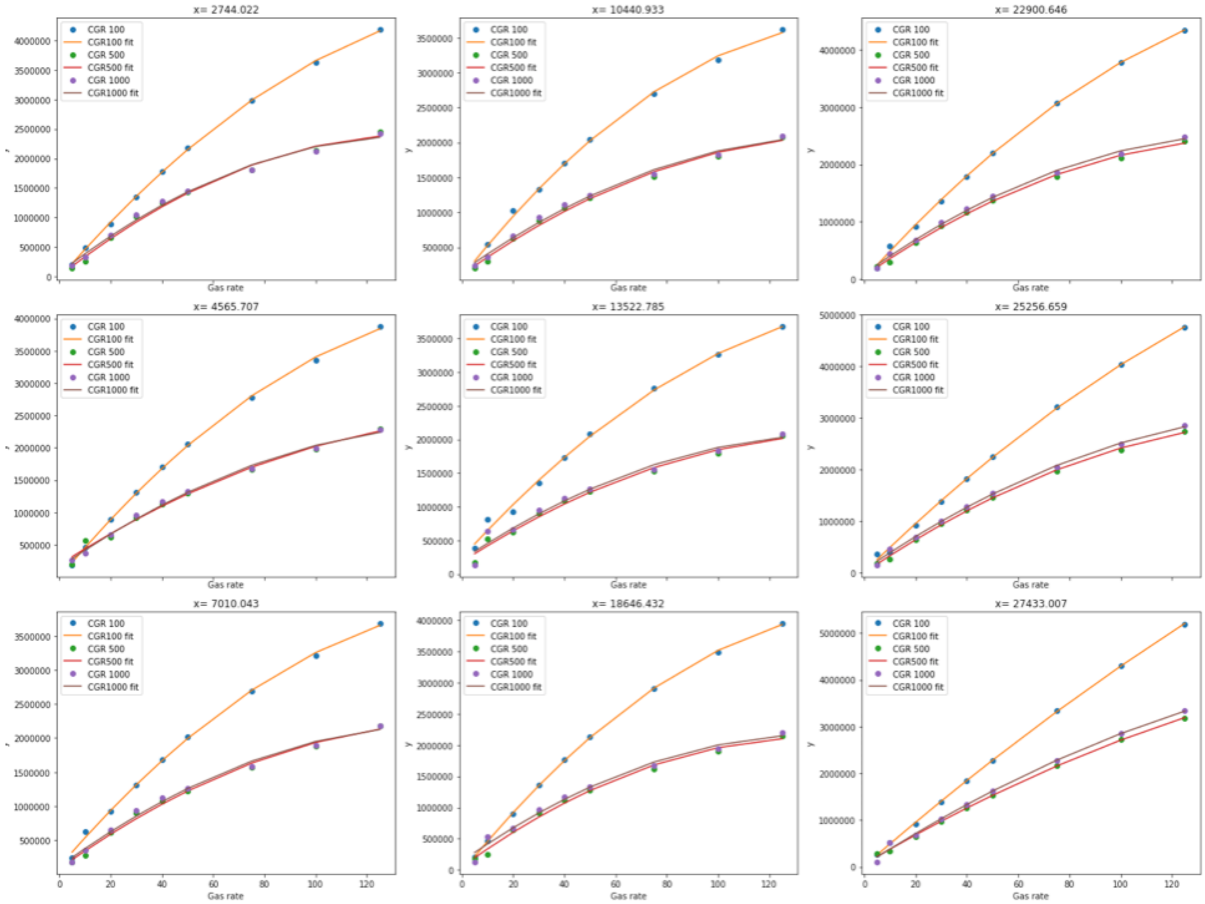
Degree 1



Reynolds-Gas Rate

Polynomial Regression

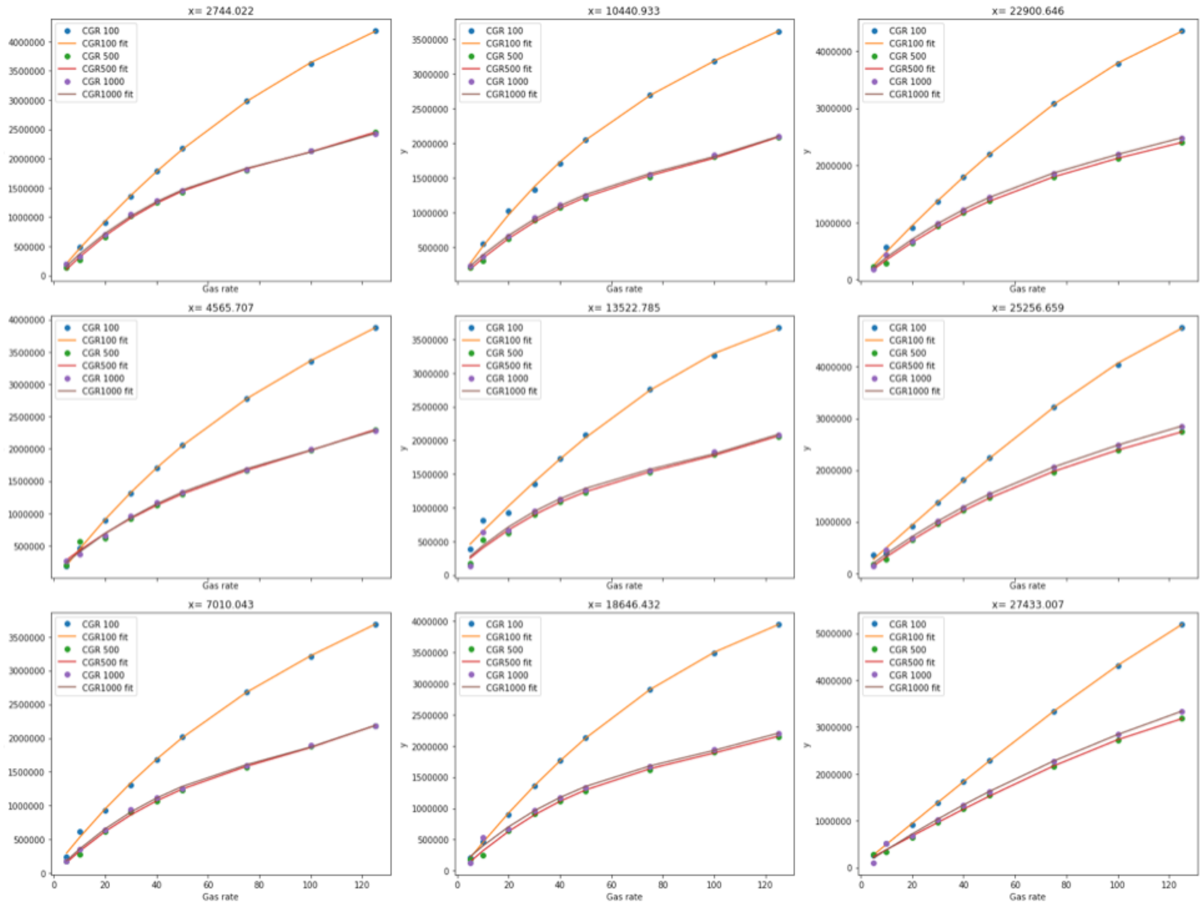
Degree 2



Reynolds-Gas Rate

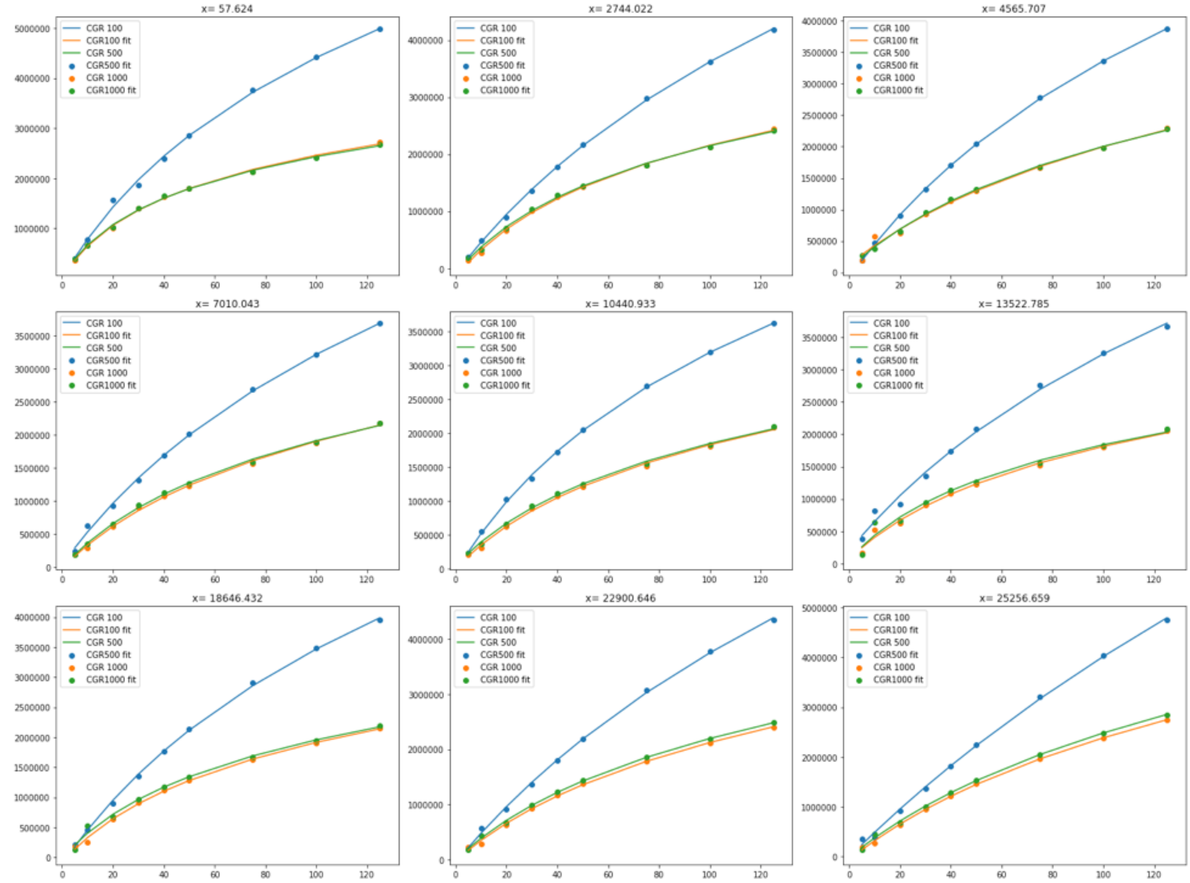
Polynomial Regression

Degree 3



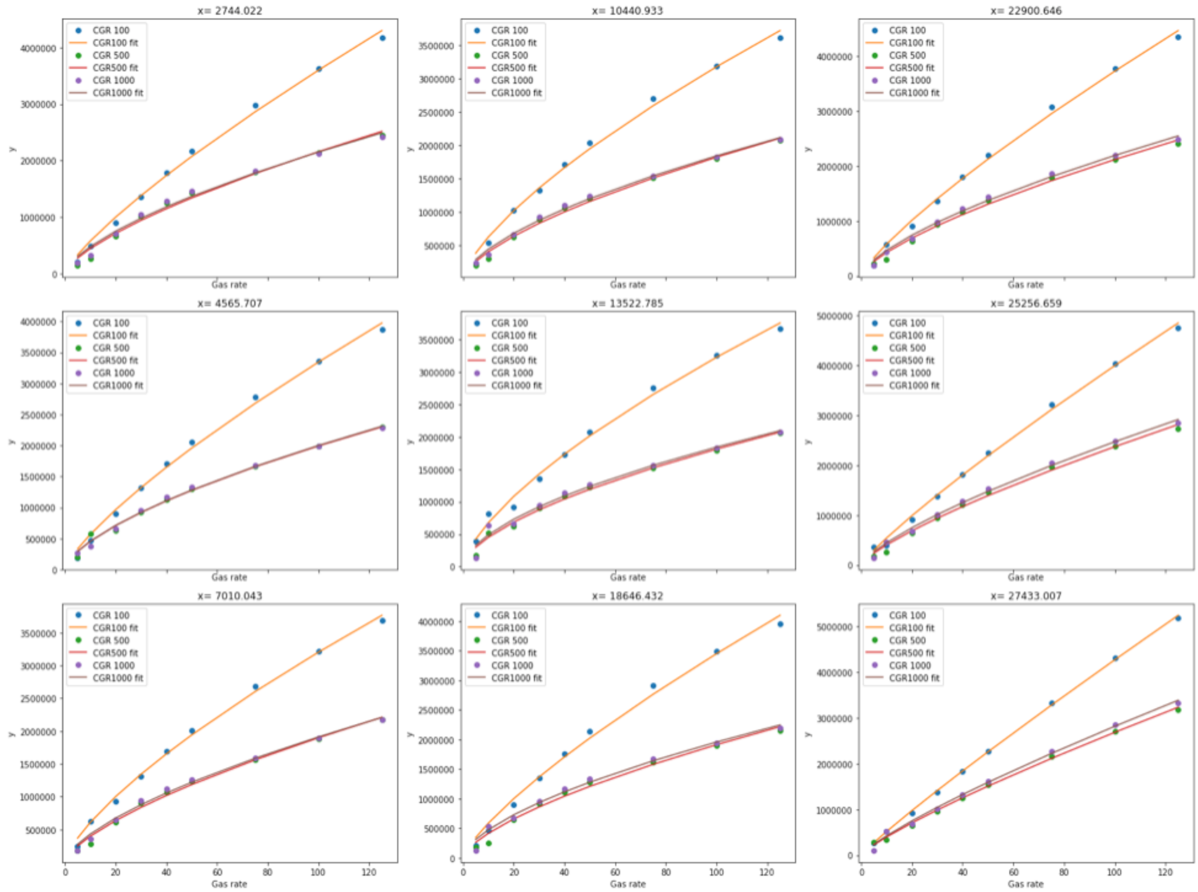
Reynolds-Gas Rate

Logarithmic Regression



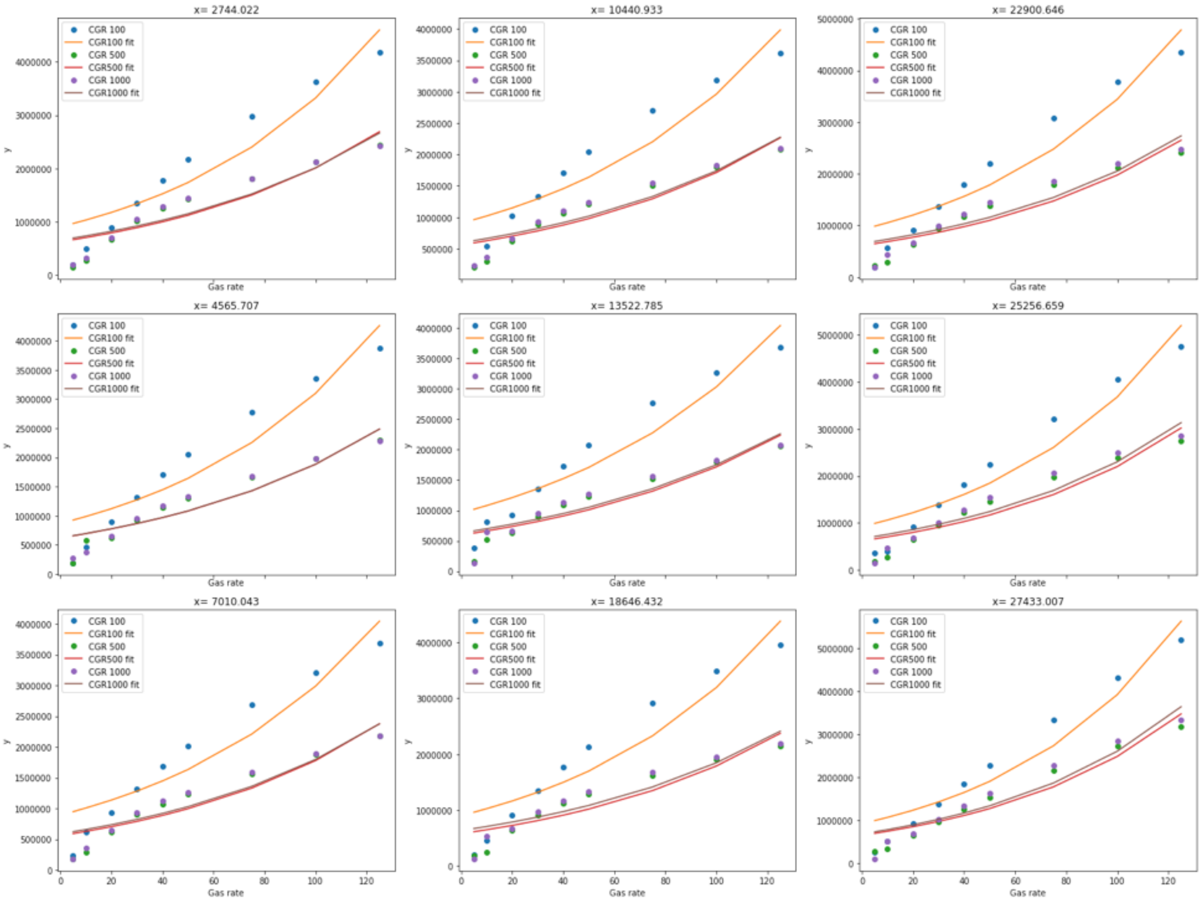
Reynolds-Gas Rate

Power-law Regression



Reynolds-Gas Rate

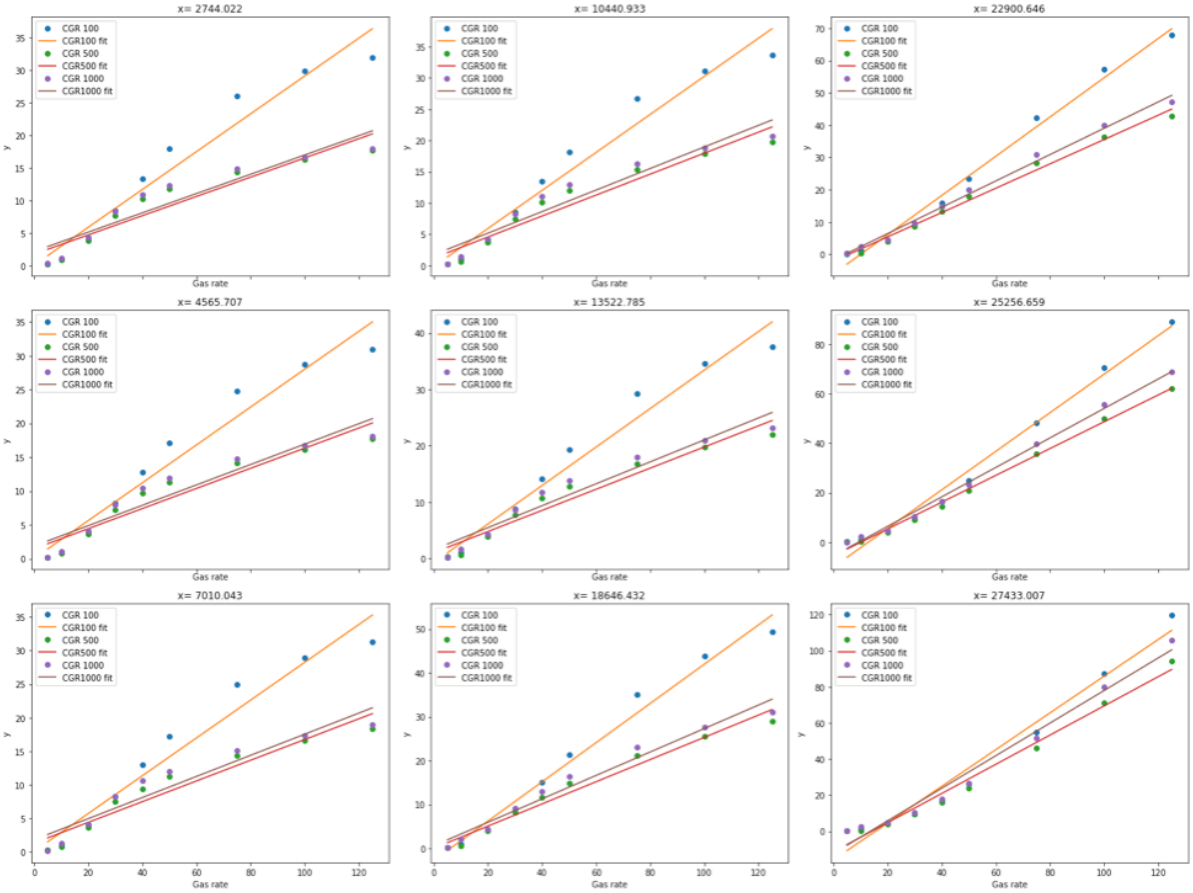
Exponential fit Regression



Froude-Gas Rate

Polynomial Regression

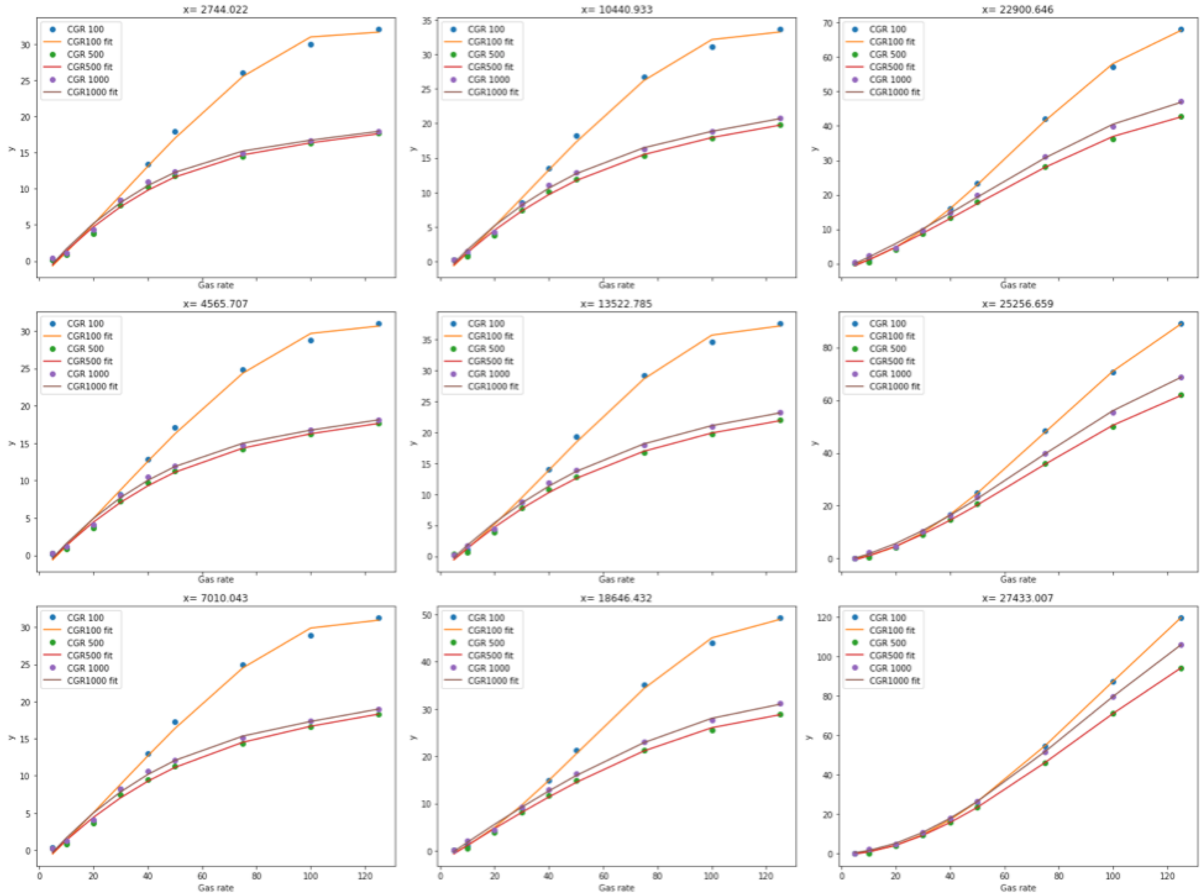
Degree 1



Froude-Gas Rate

Polynomial Regression

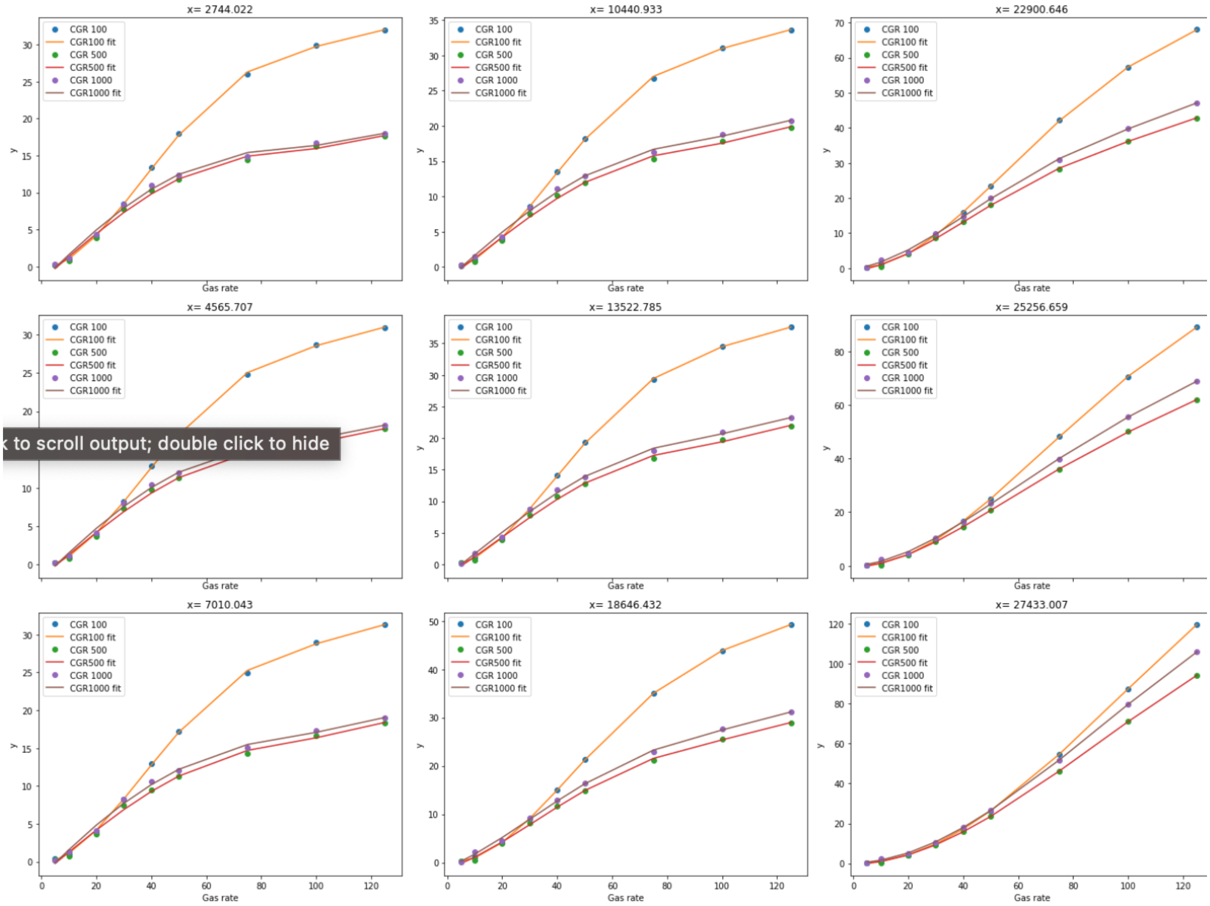
Degree 3



Froude-Gas Rate

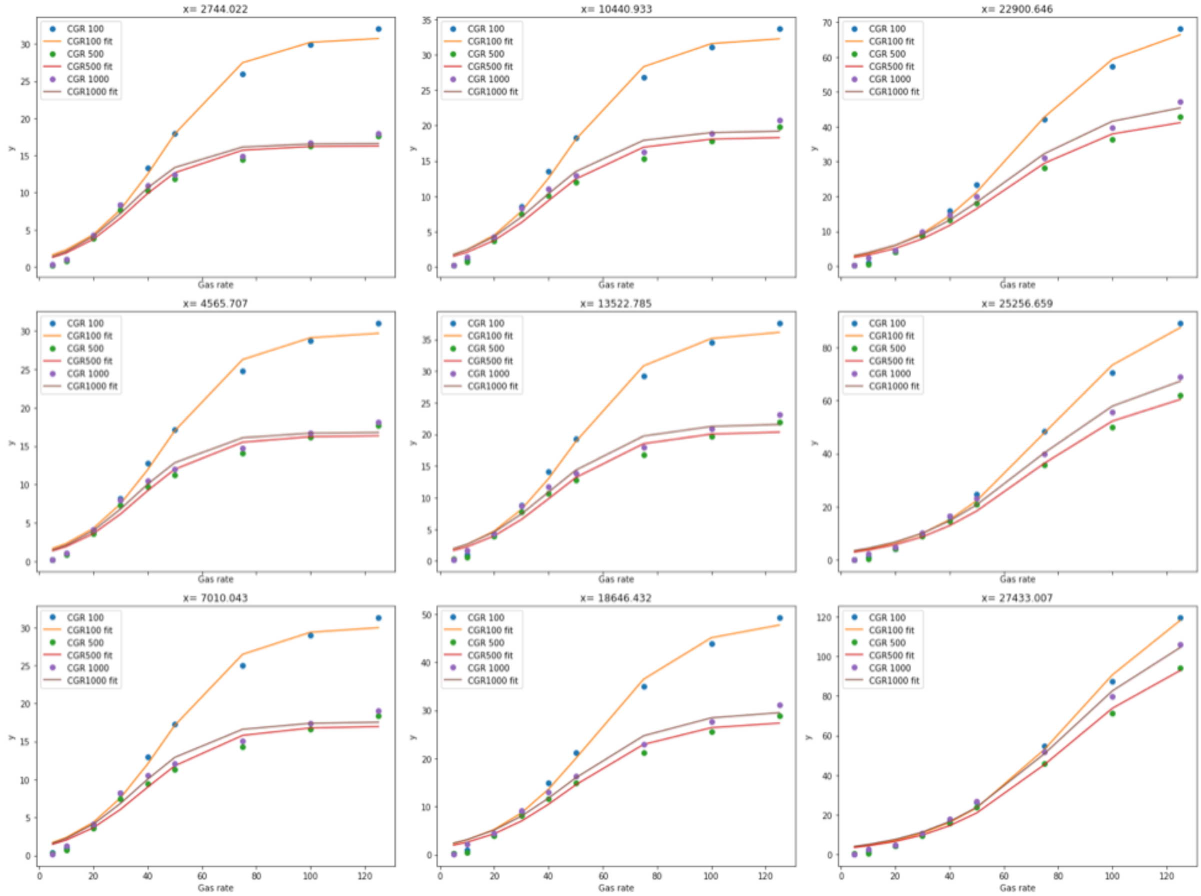
Polynomial Regression

Degree 4



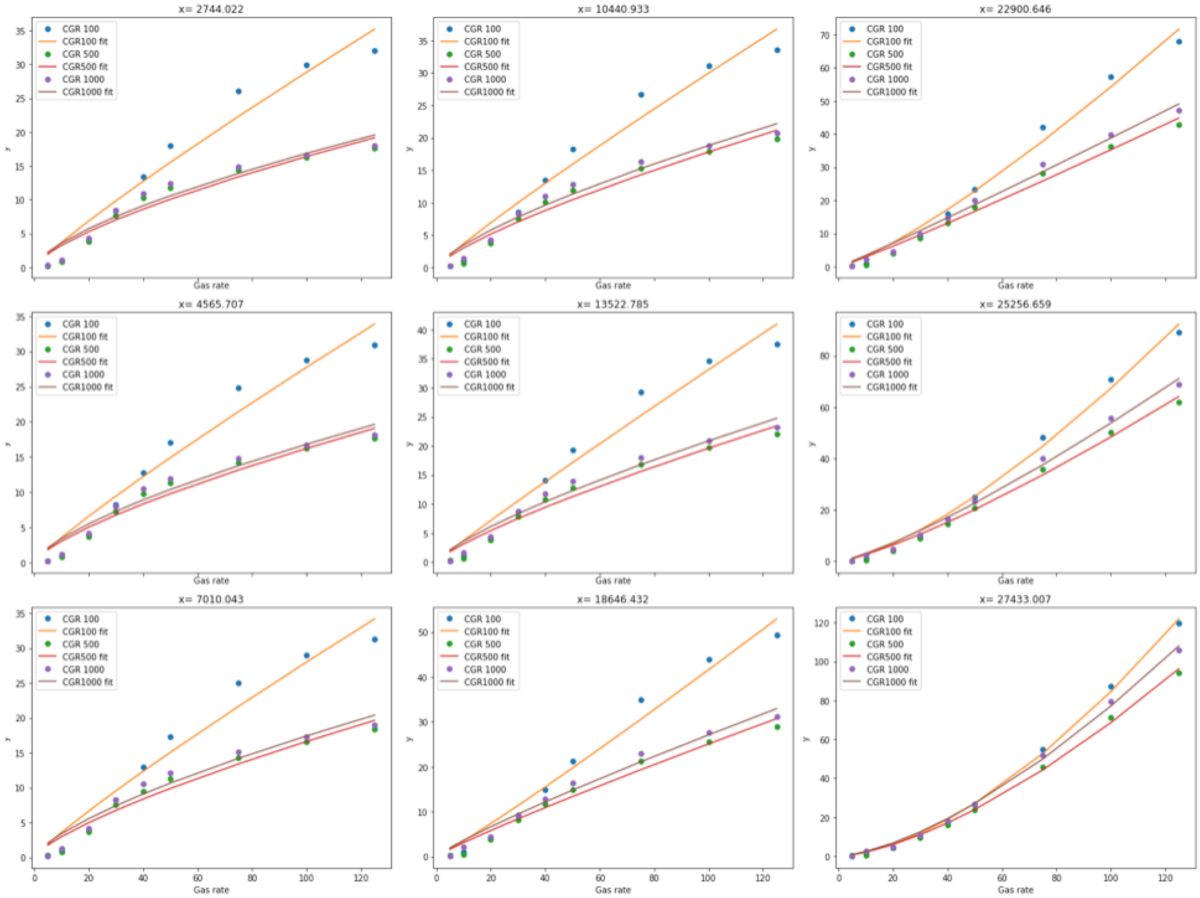
Froude-Gas Rate

Logistic function regression



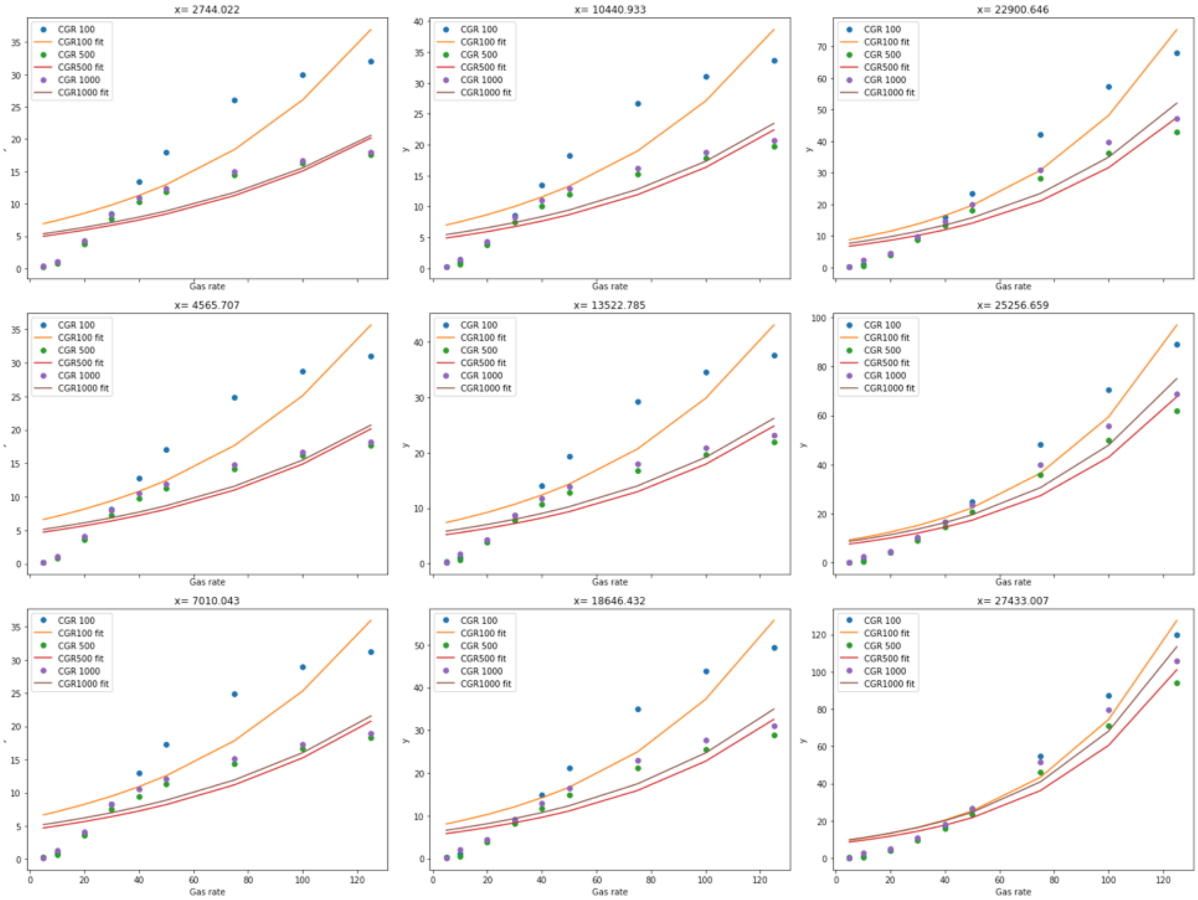
Froude-Gas Rate

Power Law function regression



Froude-Gas Rate

Exponential function regression



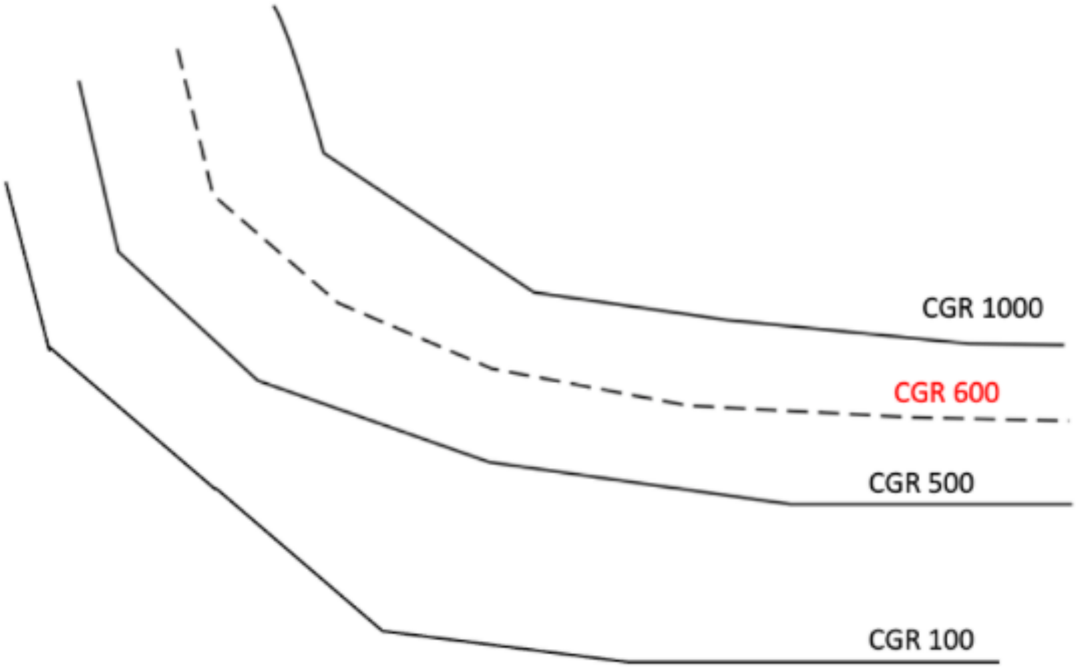
Optional Exercise

Due to time limitations, I will outline my approach to the optional exercise.

Given that the data are unavailable for CGR 600, we need to understand how the data from CGR100, CGR500 and CGR 1000 can help us infer or calculate values of various quantities for CGR 600.

After plotting the data points for pressure, liquid and gas velocities as a function of gas rate, corresponding to each CGR, it is clear that the trends between CGR500 and CGR1000 are very close to each other. A rough estimation of the values of pressure and velocities would be taking the average of values at a gas rate of 75 MMscf/d for CGR 500 and CGR1000, in order to calculate the pressure at a gas rate 75 MMscf/d for CGR 600.

A more sophisticated approach would be to interpolate between lines, like so:



We begin by identifying an appropriate regression function for the pressure, liquid and gas velocities (please see bottom of the notebook cgr100-500-1000.ipynb).

After finding appropriate functions we then create a custom function interpolating between the data points of the available datasets.

The above methods assume that the CGR trends follow an order (e.g. the response value at CGR 100>CGR 500>CGR 1000 or CGR 100<CGR 500<CGR 1000