**Comparison Module (*Placeholder*)  
Version 0.0  
16th February 2018  
Oisin Creaner**

This module takes input from a model of telescope performance and compares it with calibration data from real telescopes to generate metrics for the deviation of the model from reality

**Outline**

This module reads in input from a modelling system and input from a real telescope and compares the two against one another. This module assumes that the inputs have been brought to a common format, with common independent variables (e.g. position, time, frequency) and at least one common dependent variable to be compared. Outputs include a plot of the variation of the dependent variable, a calculated value of Root Mean-Square Error (RMSE), and a calculated value for the correlation of the variables. (*These outputs need to be discussed with Onsala*)

**Design Diagram**

**Telescope File**

**Model File**

**Read in Data**

**Dataframe: Model**

**Dataframe: Telescope**

**Merge the data**

**vector: differences**

**Calculate difference**

**Option: could be proportional or normalised differences?**

**Plot differences**

**Image plot File**

**Calculate RMSE**

**Float: RMSE**

**Float: Correlation coefficient**

**Calculate Correlation**

**Option: joining could be carried out by merging the dataframes. Alternatively, we could mandate identically structured inputs**

**Option: these outputs are samples of basic analyses that can be used as a starting point. Would like to add more options**

Figure : Outline of the comparison Module

**Operation**

1. Read in the data from the model file (*for the moment, files are assumed to be CSV format*) and store the contents in a dataframe (*for the moment, I’ve coded this in R*)
2. Read in the data from the telescope file and store the contents in a dataframe
3. Merge the dataframes using an inner join to ensure only data points where a common value(s) for the independent variable(s) exists (*Option: consider including statistics of this operation in the outputs?*)
4. The difference between the two sets of values for the dependent variable is calculated and stored as a vector (*Option: part of the dataframe?*)
5. From this vector, outputs can be calculated. The initial proposals are as follows
   1. A plot of the differences against the independent variable
   2. A calculation of the RMSE
   3. A calculation of the correlation coefficient for the model against the real data

**Assumptions**

1. Model data and telescope data can be made available
2. The datasets have common independent variables with which to match the datasets
3. The datasets have a comparable dependent variable (equivalent property, similar units etc)
4. Option: ensure the datasets are the same size and structure as inputs – if this is the case, this must be considered an assumption