

ME40358 – Assessment 1c

Sensitivity Analysis 2020/21

Dr Joseph M Flynn

Background Information

A company in charge of managing a wind farm is embracing the art of data science. As such, they wish to extract data from the sensors on one of their wind turbines and conduct a sensitivity analysis to identify parameters that have a significant effect on the generated power. Although they think that they have a good idea about what might have an effect, they are taking an open-minded approach. And are willing to consider any of the parameters as being significant.

Part 1 of the Assignment

You are one of the data analysts working within the company and your assignment is as follows:

- i. Download the wind turbine dataset from Moodle. Please note that this is real wind turbine data, captured by the French firm, Engie (link for data download: <https://moodle.bath.ac.uk/mod/resource/view.php?id=799451>)
- ii. Download the list of variables names so that you can identify what each column in the dataset represents (link for variable name download: <https://moodle.bath.ac.uk/mod/resource/view.php?id=885727>)
- iii. Select five variables that you think might have an effect on the generated power. This might be based on intuition or some of your own research.
- iv. Cleanse the data to remove any data points that you think might have a negative impact on the analysis. Some suggestions include missing data points, outliers and data points that are a result of the control system and not the physics that determines the wind turbine performance. Clearly describe all of the reasons for removing data points in your report.
- v. Using the cleansed dataset, conduct a sensitivity analysis using any of the techniques from the notes. It is highly recommended that you use more than one method, and marks will be awarded for trying some of the more advanced techniques. For example, a submission that only creates scatter plots is not going to score highly.
- vi. Describe your decision making process for using the sensitivity analysis techniques that you have employed. Why was it a valid or sensible choice, and why did you choose these techniques instead of some of the others.
- vii. For each technique that you employ, clearly show the outputs from the analysis. This might include correlation coefficients, regression coefficients or sensitivity indices.

Once again, you do not need to write a full, formal report. Simply describe the steps you have taken, the motivations for your choices, and clearly tabulate or plot the results in each case.

Part 2 of the Assignment

A data analyst at *Company A* has fitted two mathematical models to wind turbine data captured from one of their products. This model has been fitted using MATLAB. As a result of industrial espionage, a competing company, *Company B*, has gained access to this model. However, the analyst at *Company A* has taken reasonable security measures and obfuscated their code (it cannot be viewed). However, *Company B* are still hoping to gain some insights into the performance of *Company A*'s product. *Company B* has already worked out that the model operates by using five model parameters (for this exercise, it does not matter what these variables are). They have also deduced that the output of the model is the rotor power of *Company A*'s product.

To use the stolen MATLAB function, you must type the following into MATLAB:

1. five model parameters, which are presumably related to some operational parameters (e.d. wind speed), formatted so that each input forms one column of an $N \times 5$ matrix;
2. a string that identifies which of the two mathematical models is going to be used (a string of either '1' or '2');
3. an employee ID number between 1 and 22 (your individual employee ID numbers are at the end of this document, and you must use your allocated number).

As such, the code that calls the model will be formatted as follows in MATLAB if the five model parameters are A, B, C, D and E, model '1' is being used and the employee ID is 22. N is the number of values supplied for each model parameter, which also corresponds to the number of rows in the matrix.

$$power = TurbineModel_2020 \left(\begin{bmatrix} A_1 & B_1 & C_1 & D_1 & E_1 \\ A_2 & B_2 & C_2 & D_2 & E_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ A_N & B_N & C_N & D_N & E_N \end{bmatrix}, '1', 22 \right)$$

For models 1 and 2, use variance-based sensitivity analysis to identify the First order Effects (S_i) and the Total Effect indices (S_{Ti}).

For this part of the assignment, you must upload your MATLAB code that conducts the variance-based sensitivity analysis, and also your tabulated results in the following strict format:

	Employee ID = ##			
	Model 1		Model 2	
Parameter	S_i	S_{Ti}	S_i	S_{Ti}
A				
B				
C				
D				
E				

Finally, you must include a short written section on what you have deduced about the nature of the model and the interactions between model parameters inside the model

Employee ID	Student Name
1	Ege Arabul
2	Gabriele Butkeviciute
3	Sam Butter
4	Jerry Chung
5	Tom Clarke
6	Joel Cullis
7	Jerome De Saulles
8	Loic Estier
9	Alberto Guerra Martinuzzi
10	Rosie Knight
11	Jack Martin
12	Conor Moffatt
13	Callum Morrison
14	Andrei Petrus
15	Finin Quincey
16	Benedict Ratcliffe
17	Sam Riddell-Webster
18	James Russ
19	Hannah Thomas
20	Helene Verhaeghe
21	Donny Wong
22	Darshi Shah