Tools for Engineering Design 2

Data Processing Individual Project

Introduction

Things like Python scripts and spreadsheets (such as Microsoft Excel) are invaluable in design. They are free-form calculation tools that allow different design options to be rapidly evaluated, tedious hand calculations to be automated, large amounts of data to be manipulated, update calculations and reports as data changes, and verification. The versatility and complexity of these tools, together with how people use them, however, means that they can be very difficult to follow by someone else reviewing them or even the authors themselves (try opening a spreadsheet or a computational notebook you wrote a while back, can you follow easily what happened?).

The aims for this self-led learning assessment are:

a) to build upon the knowledge developed in Programming Skills for Engineers 2 by using Python and its ecosystem of libraries to handle, manipulate and process data for typical engineering calculations;
b) developing a critical understanding of how to process data and check for errors, and communicate about the algorithms about these processes; and
c) to present the calculations and their outputs clearly and concisely, so that they can be easily understood, reviewed, and used by someone who has not been involved in the project before, and in enough detail to give a complete record of the calculations, so that it could be built upon by someone else later.

This project is worth 15% of the total TED2 course marks and it is submitted in two stages:

1. **Formative submission (Thursday 9 February 2023 at 14:00 via Learn)**. This is a voluntary submission, and it does not contribute towards your TED2 course marks. The purpose of this submission is to give you a chance to receive feedback on your approach to project ahead of the summative submission. It should contain **a PDF file with 1 page, with an outline of the plan of attack**. Our advice here is to imagine you are having an informal chat with a colleague and only have an A4 to explain what you are going to do and how. We suggest having hand-written notes describing (1) the overall approach to the project, and (2) calculation steps. Ideally, just enough text alongside hand-drawn diagrams, and a small example with some numbers.
2. **Summative submission (Thursday 2 March 2023 at 14:00 via Learn)**. This is a compulsory submission, and it will be assessed using the criteria shown in Table 1 at the end of the brief. You should **submit a single ZIP file similar to the one you were given as an example**. The ZIP file should contain the key information needed to follow your calculations and produce your reports. The essential information will be: **(1) the Jupyter notebook (.ipynb file), (2) the Jupyter notebook exported to HTML (.html file), (3) the report in the Excel spreadsheet (final .xlsx file), (4) any auxiliary files** that your solution may require. A note about auxiliary files, these include (a) the raw data you were given, but you may have additional files, like (b) empty template for the excel spreadsheet, and/or (c) external diagrams that are referenced within the Jupyter notebook (see the example on Learn).

This is a self-led activity, and our advice is to try to engage early on with this project and to have a go at things by yourself first, using the example provided (files and videos). There are several avenues to receive timely support:

1. Dedicated discussion board on Learn. We expect everyone to give each other support on the task. Staff monitor it and will typically reply to unanswered posts within 2 working days.
2. Surgery hours. Great for in-depth one-to-one or group conversations.
3. Chat during sessions together (mainly for quick questions to not disrupt group work).

The project

A new shipping channel is required due to the construction of a bridge connecting Lachtenstein and Crowrovia. The problem is shown graphically in Figure 1 and Figure 2. A depth survey has been conducted of the region, and the survey data for the existing seabed profile is provide in a raw format in the spreadsheet file available on Learn.

You need to calculate the volume of material to be dredged to form the new shipping channel based upon the survey data. An accurate estimate of volume dredged in m^3 is required to estimate the costs of dredging the shipping channel. Note you do not need to estimate the cost, just the volume.

Note that:

1. The calculation is not trivial and there are many ways of going about it. It is often a good idea to start by solving a simplified version of the problem with pencil and paper before moving on to the real, more complex, data. It is also a good idea to solve the problem in two or more ways to see if results still make sense and are robust to any methodological assumptions you may have introduced in your approach.
2. The survey data comes in a standard format that holds data that is messy and difficult to read by a person. This is very common with large datasets handled by computers. You may assume that:
    a. the survey is always carried out in the same way, meaning that the ship surveys a rectangular domain of the seabed following a rectilinear grid where the spacing between lines are always the same;
    b. the file format of the survey data does not change, meaning that its structure will always be the same for the 'metadata' sheet, and that the table in 'data' will always have the same number of columns and units, although the number of rows may vary.
3. Take care to understand what units and datums are being used in this survey. This is again a common occurrence in practice and can cause problems (see for example the Mars Orbiter Mishap Investigation Board Phase I Report).
4. The data contains errors due to mis-readings and logging issues during the survey. Your approach and/or calculations should be resilient to the worst of these errors.

The intention is that in a few years' time a similar survey will take place (after the initial dredging) to assess if the channel needs to be re-dredged to ensure the required depth is maintained. Your company want to set out your calculations so that this can be done with the minimal of effort in the future for this and other projects.

Presenting the project

It is very important to take care when constructing any set of design calculations so that somebody else can easily understand it. It must be possible for another engineer to understand, check, and if necessary, extend or modify your calculations. Remember that this engineer may not be able to ask you about the content of your files so the methods and where the data is coming from needs to be extremely clear to a future engineer.

Making the calculations is only part of the current exercise; you need **to produce reports** in a neatly formatted Jupyter Notebook and Excel spreadsheet – the rationale for the latter is that colleagues who are not familiar with Python can still review the project for themselves. (Note: that the output data is more than just the final volume to be dredged, but short report on the key input data and process as well).

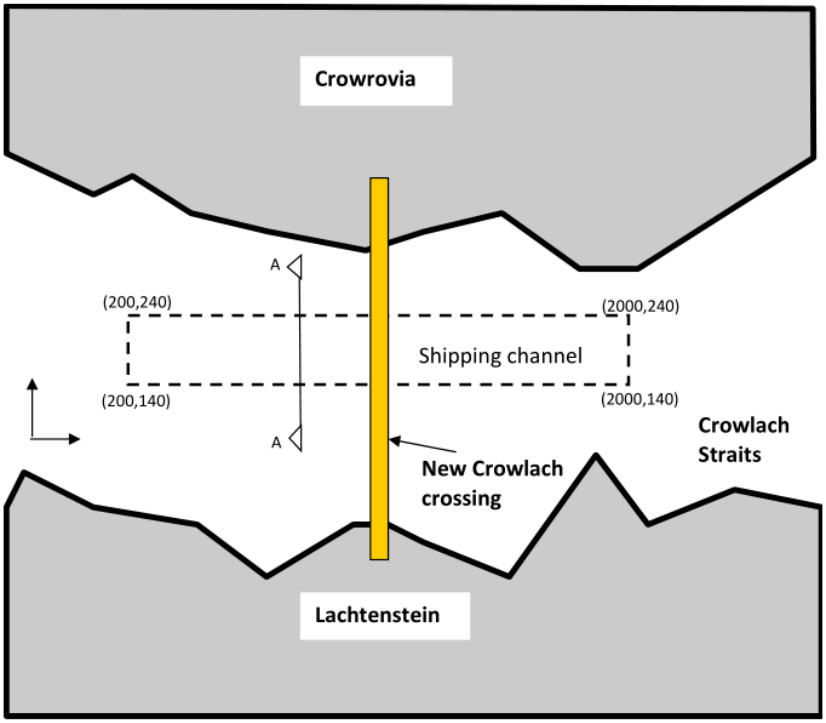

Figure 1: Plan of proposed shipping channel under the new Crowlach crossing. Coordinates in metres from the indicated datum. Not to scale. Section A-A drawn in Figure 2.

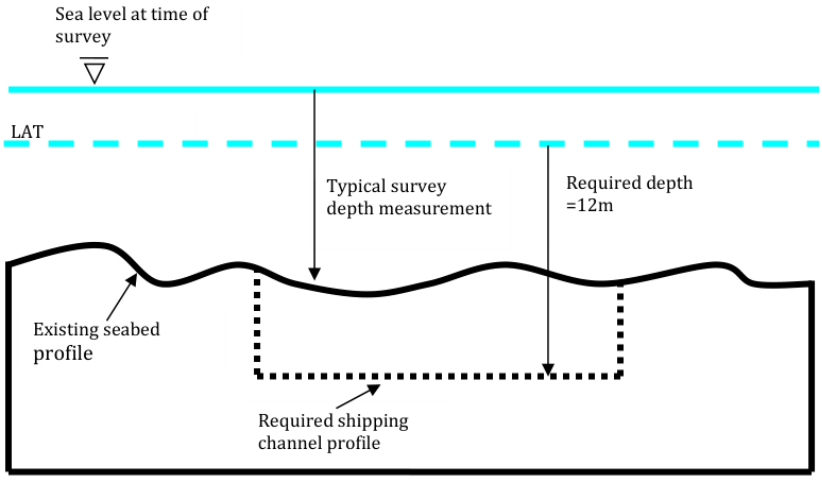

Figure 2: Typical section A-A through the planned Crowlach Straits shipping channel. Lowest astronomical tide (LAT) data taken from Lachtenstien hydrographic office publications.

Hints

1. The formative submission date is not indicative that you should have made progress just up to that point in the project. All things considered for TED2, we recommend that at that stage you are well underway towards the summative submission and have a working solution already implemented. You can, and should, continue working on the project while we review formative submissions.
2. Think carefully about the best way to set out and present the data so that somebody else can easily understand it. For example, the data you have been given is a single list of survey points, but that is not easy to understand because the points have (x,y) coordinates within the shipping channel.
3. Always think about the accuracy of your data and results, and only present to an appropriate number of decimal places and include units (just as with hand calculations).
4. Bear in mind that somebody looking at your notebook or spreadsheet in the future is unlikely to have a copy of this brief. A good set of design calculations will always include general information, concise and well explained calculations and processes (without needing to look in the cell formulae), and will be carefully formatted.
5. The example files (both Python and Excel spreadsheet) showcase a *work in process* to give you an idea of how to think, work and resources that you might find useful. Further work would be needed to clean things up and document things appropriately (particularly in the spreadsheet!).
6. The example notebook contains a basic showcase of the libraries and functions that you might find useful. For example, pd.read_csv is useful to read CSV data and has lots of options in the documentation (click on the link). For this project I would use pd.read_excel since the data comes in an Excel spreadsheet. Similarly, think about the kinds of figures that would be most informative, which are going to be different to the ones used in the example.
7. Asking for help about code is a tricky business.
    1) There is the temptation of thinking that a piece of code that executes without coding errors is working as intended. That is not necessarily the case. Also, sometimes coding errors cover unexpected mistakes one might be making (like `"a" + 1` fails to add `1` to the string `a`), and some other times they are useful safeguards (like `1/0`, division by zero).
    2) Programming languages have mechanisms to report coding errors. See one of the many tutorials available on how to read them for the Python language. Once you know what the coding error is about, it is easier to find help online. For example, `1/0` raises in Python a `ZeroDivisionError` and searching for that error leads to, for example, this page.
    3) When asking someone else help with a problem, it is good practice to create a Minimal, Reproducible Example whenever possible.

Table 1: Assessment Criteria for the summative assessment.

	F	E	D	C	B	A3	A2	A1
Files	No files submitted.	-	-	Partial submission. Only the Jupyter notebook (.ipynb and/or .html) files are submitted or the Excel spreadsheet.	-	-	All necessary files are submitted correctly.	In addition to A2, the folder structure and file naming convention is useful to identify relationship between files and their sequence immediately by someone else.
Calculations	No calculations are presented OR An unrealistic number that should have been judged to be unrealistic as it is overly large/small.	Numerical answer is outside ±10% but still within the boundaries of the problem.	Numerical answer is within ±10%.	Numerical answer is within ±7%. OR Like D, but there is more than 1 calculation method implemented with a reflection on their comparison and with a clear decision on the final value to be used.	Numerical answer is within ±4%. OR Like C, but there is more than 1 calculation method implemented with a reflection on their comparison and with a clear decision on the final value to be used.	An accurate estimate, within ±2%. OR Like B, but there is more than 1 calculation method implemented with a reflection on their comparison and with a clear decision on the final value to be used.	In addition to A3, there is more than 1 calculation method implemented with a reflection on their comparison and with a clear decision on the final value to be used.	In addition to A2, calculations could be directly applicable or easily adaptable to any similar type of dataset (same format, different domain, different sampling resolution, similar errors, data sorted differently).
Notebook	None submitted. OR All calculations are done within Excel (+5 addition at end).	The notebook is unintelligible. It may or may not work as desired, but it could not be reasonably maintained by another person with skills on the area.	The notebook calculates the volume and reports it to the spreadsheet. It also presents information about (1) the task, (2) the approach, (3) the inputs, (4) calculations, (5) the output and (6) overall interpretation. However, the information is presented either explicitly (markdown cells) or implicitly (code) and there are gaps in the information presented. Although the notebook may work as intended, it is confusing and difficult to maintain by another person with skills on the area.	The notebook calculates the volume and produces a report in a spreadsheet. It also presents information about (1) the task, (2) the approach, (3) the inputs, (4) calculations, (5) the output and (6) overall interpretation. Complete but basic information is presented explicitly (markdown cells / code, as appropriate). As a result, another person can follow what is going on and have a superficial understanding of what is going on.	The notebook calculates the volume and produces a report in a spreadsheet. It also presents information about (1) the task, (2) the approach, (3) the inputs, (4) calculations, (5) the output and (6) overall interpretation. Complete and thorough information is presented explicitly (markdown cells / code, as appropriate). As a result, another person can easily follow what is going on and critically appraise results for themselves based on evidence provided.	In addition to B, there has been a clear effort to parametrise the notebook inputs and calculation routines, with appropriate levels of documentation (markdown cells and/or code, as appropriate).	In addition to A1, an effort has been made to clearly visualise key aspects of the task OR sanity checks are performed as necessary to verify the code performs as intended, with the code able to raise informative error messages and halt execution if key assumptions are not met in the dataset or the method implemented.	In addition to A1, an effort has been made to clearly visualise key aspects of the task AND sanity checks are performed as necessary to verify the code performs as intended, with the code able to raise informative error messages and halt execution if key assumptions are not met in the dataset or the method implemented.
Report	No excel output provided. OR If all calculations were done in the spreadsheet, extra points were given at the end according to the accuracy of calculations (Pass +2, Good/Very Good +3, Excellent +4).	The report spreadsheet essentially contains a cell with the estimated volume alongside units and what it represents (this is, the file does not constitute a report).	The report spreadsheet contains a cell with the estimated volume alongside units and what it represents. There has been an attempt to provide contextual information about (1) the project OR (2) the data processing workflow that transformed the input dataset into the final value.	The report spreadsheet contains a cell with the estimated volume alongside units and what it represents. There has been an attempt to provide contextual information about (1) the project AND (2) the data processing workflow that transformed the input dataset into the final value.	The report spreadsheet contains a cell with the estimated volume alongside units and what it represents. The spreadsheet constitutes a complete but basic report about the project and the data processing workflow that transformed the input dataset into the final value. As a result, colleagues can provide general feedback just based on the spreadsheet report, but someone with relevant skills would not be able to fully recreate calculations based on just this report.	The report spreadsheet contains a cell with the estimated volume alongside units and what it represents. The spreadsheet constitutes a complete and thorough report about the project and the data processing workflow that transformed the input dataset into the final value. The spreadsheet includes copies of the input and the revised dataset based on which the calculations were performed. As a result, colleagues can provide specific feedback just based on the spreadsheet report, and someone with relevant skills would be able to almost recreate the notebook based on just this report.	As for A3, but an effort has been made to represent the workflow and the data to the user. This allows those unfamiliar with the project its notebook to interpret the data and represent it as they see fit with only minor changes and edits within Excel. As a result, colleagues can provide specific feedback just based on the spreadsheet report. Someone with relevant skills would be able to recreate and verify calculations based on just this report.	As for A2, but anyone with a general engineering background would be able to follow all aspects of the report and how the final number was calculated.