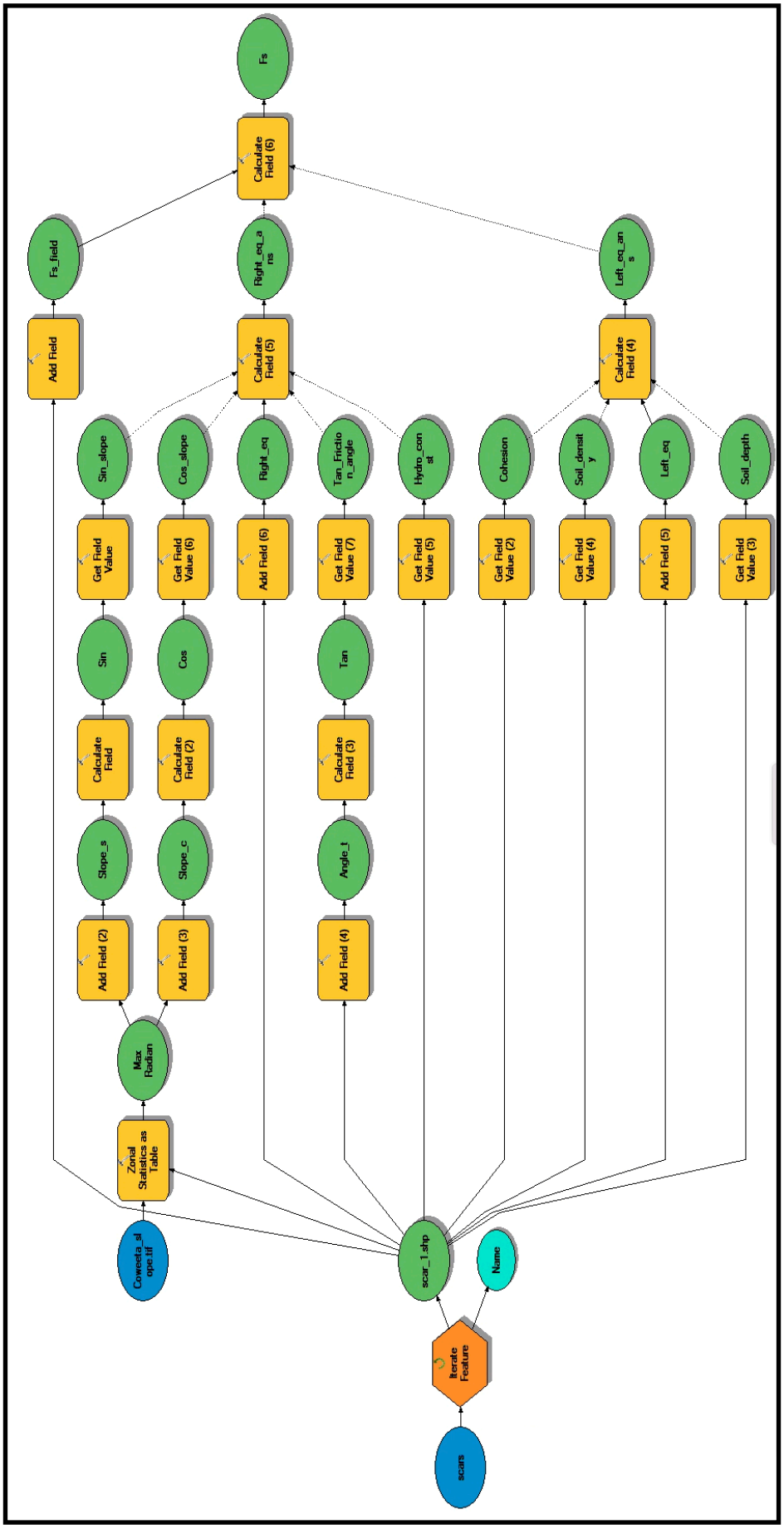
SCHOOL OF GEOGRAPHY COURSEWORK SUBMISSION **COVERSHEET 19/20** *Please complete the following table in full*		
Student number (9 digits):	181084879	
Module code and title:	GEG6320 - Advanced Geospatial Science	
Submission deadline date:	16th December 2020	
Coursework title:	Assessment 2 Portfolio	
Total number of words:	1356	
I declare that this coursework is entirely my own work and contains no instances of plagiarism	X	**Put 'x' in box (left) to confirm**
To help me improve future coursework, I would like feedback on the following aspect of my submission	Code neatness	

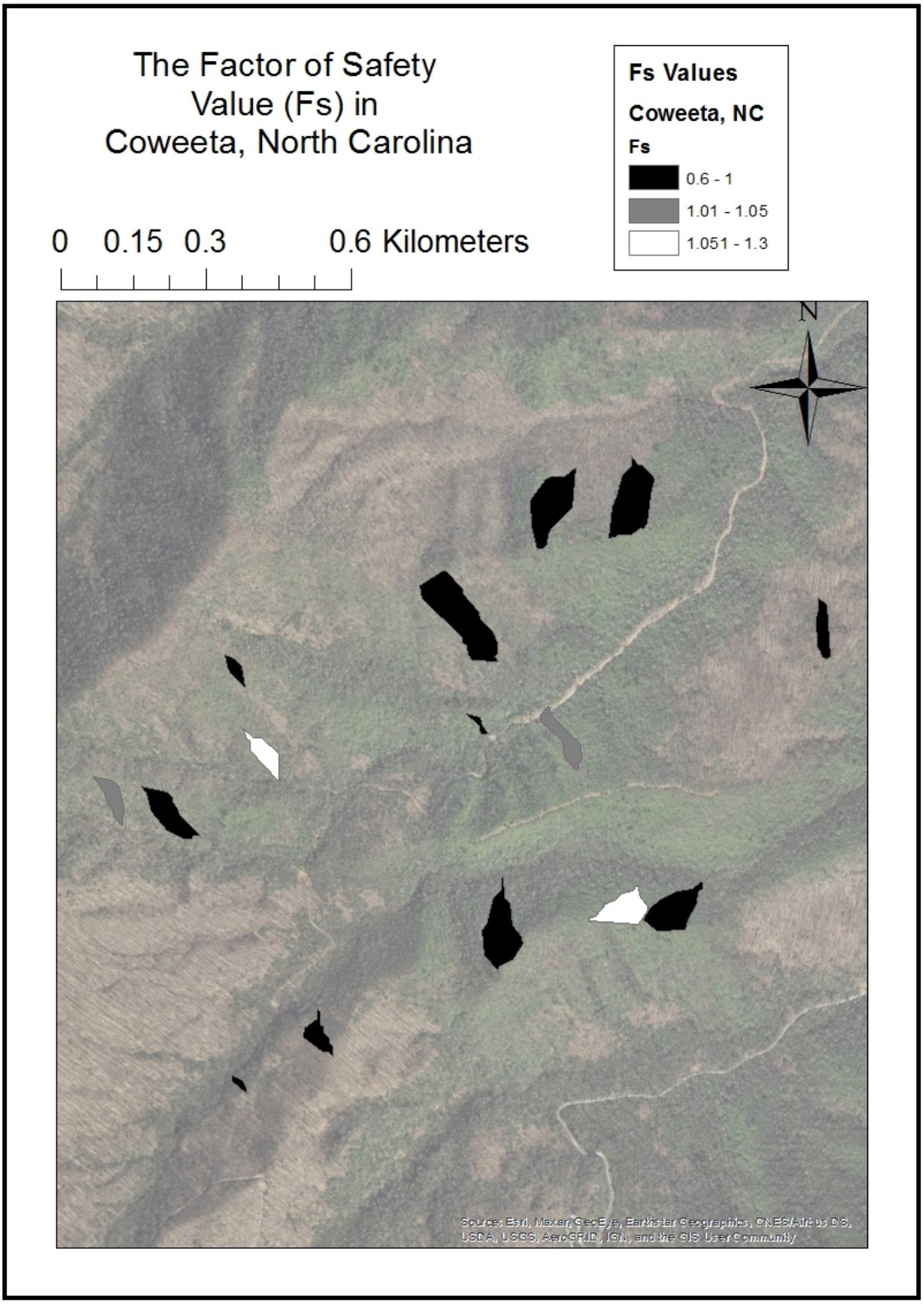
Critically Analysing the Geoprocessing workflows of ArcMap and Python

A Study into the Landslide scars of Coweeta, North Carolina.

ArcMap Model Workflow

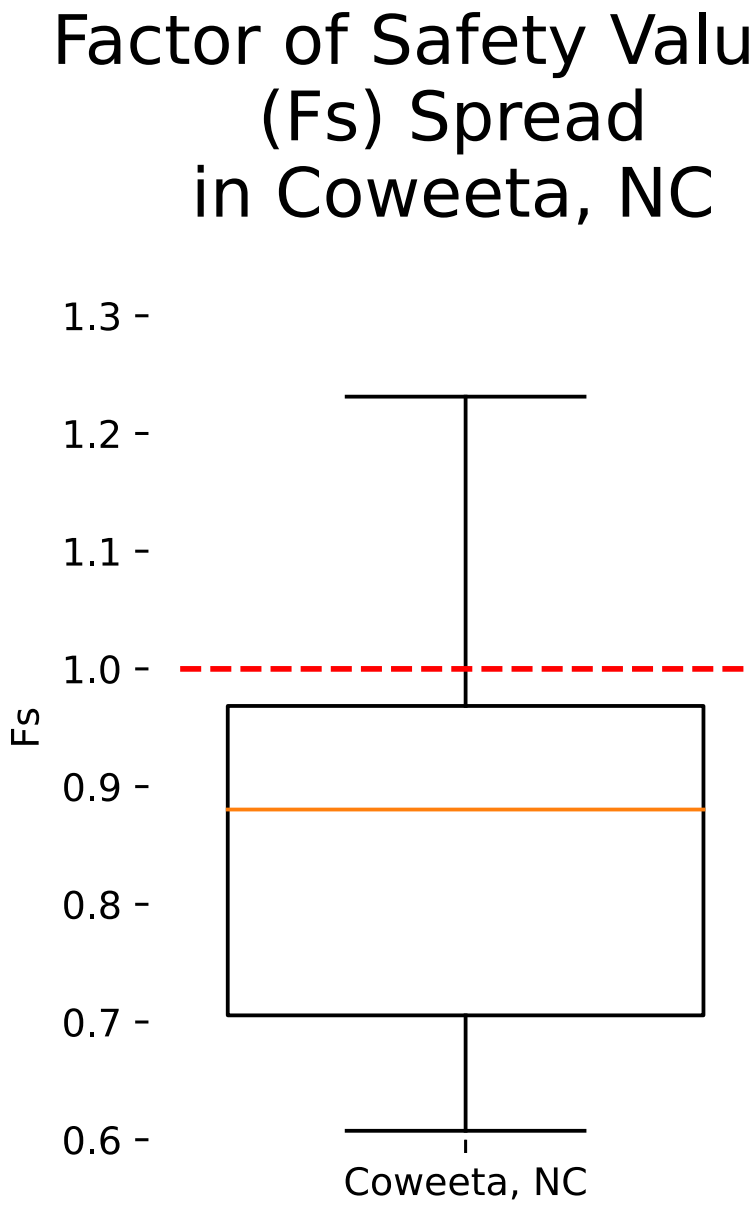


A Map of Fs Values in Coweeta, NC



The majority of Scars have a value less than 1, this constitutes a failure and would require immediate action to alleviate the risk of landslides. Only four scars have a value above 1, two of which still scored within 0.05 of the failure threshold.

A Boxplot of Fs values in Coweeta, NC



The majority of Fs values fall below the failure threshold of 1, as demonstrated by the red line. Fs values range from 0.62 to 1.23. The mean Fs value is 0.88. The lower quartile value is 0.7 and the upper is 0.97, leading to an interquartile range value of 0.27.

The Reproducibility Crisis: a comparison of ArcMap and Python.

Introduction

Geoprocessing workflows are one example of the digitalisation of research and science. The ability to use technology to iterate and automatically work through shapefiles and rasters has significantly reduced the time in which researchers spend on data collection and analysis. However, this has brought about a new set of problems. The ‘inability to reproduce significant results across a wide range of disciplines’ (Grieve at al., p.340, 2020) has led to a reproducibility crisis. There are four main concepts which have arisen with digitalisation that contribute to the current reproducibility crisis are: the use of point and click analysis, largely undocumented workflows and paywalls, which hinder individuals, small business and education from accessing the technology. Grieve et al. stress the benefits of reproducible research as it will gain ‘confidence’, ‘trust’ and ultimately ‘progress’ (Grieve et al., p.343, 2020) in each discipline, if their researchers practice reproducible science. Both Python and ArcMap have benefits and disadvantages when constructing geoprocessing workflows which fall under the aforementioned concepts.

Point and click analysis

When constructing a geoprocessing workflow with ArcMap, a model is created. This model is an example of a low/no code option when completing calculations and analysis. This is a significant advantage over Python as low/no code requires less formal coding education and therefore it is more accessible. However, no/low code methodologies are far less concise than coding methodologies. This leads to a time-consuming methodology construction which, if incorrect or difficult to follow, damages the reproducibility of the study. It should be noted that ArcMap still benefits from coding. It is still beneficial to have an understanding of Python when using ArcMap, as this will decrease the time spent and increase the efficiency of the workflow.

As Python does not use point and click analysis it as its main way of constructing a geoprocessing workflow this appears to be less of an issue. However, to first understand where the data is coming from, organising a file tree and preparing Python for calculations or analysis, some point and click analysis is still required. It is important to remember that neither of these softwares are exclusive in their methodologies and both benefit from coding and point and click analysis where appropriate.

Undocumented workflows

If workflows go undocumented or are poorly documented, the chances of another person or the original researcher being able to understand and reproduce this work is lower. Church et al. reports that ‘>50% [of biomedical researchers] could not reproduce their own results’ (Church et al., p.1, 2020), this is due to poorly documented workflows. Therefore, it is essential document methodologies and workflows the highest extent. Python has clear advantages over ArcMap when documenting workflows. This is due to the ability to markdown and annotate cells of code within the same notebook. This allows the researcher to actively describe why they’re programming that specific code in the same vicinity as the code itself. This makes the

workflow easier to follow for the recipient and increases the chances of reproducibility. ArcMap allows for annotations of cells within its model builder feature. However, this is extremely lacking as the space provided is limited, allowing only a few words per cell. Although this is still useful for visualising the workflow and what is going on, it does not go far enough to demonstrate a methodology on how to construct this workflow again. This means that to fully document the workflow using ArcMap, a larger separate methodology is required.

In a professional setting, undocumented workflows seriously hinder the productivity of companies. The bus factor 'refers to the number of people in your team who can put your project in trouble if they are hit by a bus' (Ahmed, 2018) and is very important to consider when completing work as a team. Particularly in the world of start-ups, where the majority of jobs are contract based and time limited. Should a workflow be created in a professional setting, it is vital that it is reproducible, should the creator not be present. This ultimately is why Python bests ArcMap within this concept, the vital importance of documenting workflows is recognised in Python and it is safer and more productive to utilise that software in that way.

Paywalls

Paywalls are the singular greatest hinderance to reproducible workflows. Paywalls are the pay-to-access softwares where workflows may be constructed, and research is undertaken. Due to its pay-to-access principles, people are restricted from using that specific software. ArcMap, for business purposes, costs $3000 per year to licence. This cost factor when undertaking research reduces the amount of people who are able to reproduce this work, further contributing to the crisis. By contrast, Python is an open-source software, this means that people can openly access it as long as they have a computer. This vastly opens up the possibilities of reproducible workflows if solely done on Python. Python clearly has greater advantages over ArcMap in this specific study as it does not require the specialist abilities of ArcMap. Therefore, it would be more beneficial to use the Python workflow for this specific study.

The issue surrounding pay-for-access feeds into a greater issue surrounding education and access to technology more generally. Reproducible geoscience will be useless if there is not a large proportion of researchers to reproduce the science. Having only a select few who can access the software to reproduce a study also raises intersectional issues. In STEM fields, the overwhelming majority of researchers are men. Therefore, providing access to more people will assist in levelling out this inequality. As python is open access, there are also plenty of free tutorials and educational activities available through GitHub and YouTube. This is making the software even more accessible and leveraging the fact that it is open access.

Conclusion

Overall, Python is a more accessible platform that requires less point and click analysis and is able to document workflows in tandem with programming. For the purposes of this study, it is more highly suited. However, it must be remembered that the visualisation of data is a part of this study, it is impossible to create the same level of cartographic visualisation without the use of ArcMap. Despite the cost and educational barriers that limit ArcMap's reach, the software is high specialised and is able to perform substantial workflows efficiently.

Bibliography

Ahmed, A. (2018) ‘The Bus Factor’, <https://medium.com/tech-tajawal/the-bus-factor-6ea1a3ede6bd> accessed 14th December 2020.

Church, M., Dudill, A., Venditti, J. G., & Frey, P. (2020). Are results in geomorphology reproducible?. Journal of Geophysical Research: Earth Surface, 125(8), e2020JF005553.

Grieve, S. W. D., Clubb, F. J., & Mudd, S. M. (2020). Reproducible topographic analysis. In Developments in Earth Surface Processes (Vol. 23, pp. 339-367). Elsevier.