USING FREE RESOURCES FOR THE CREATION OF DIGITAL ELEVATION AND GEOGRAPHIC DATA: THE CASE STUDY OF RODRIGUES ISLAND

by

Janique Savy

12004520

Submitted in fulfilment of the requirements for the degree

MSc Geoinformatics

in the Faculty of Natural & Agricultural Sciences

University of Pretoria

Pretoria

December 2022

Department of Geography, Geoinformatics and Meteorology

University of Pretoria

Supervisor

Dr Christel Hansen, Department of Geography, Geoinformatics and Meteorology, University of Pretoria, South Africa

Co-Supervisor

Dr Cilence Munghemezulu, Department of Geography, Geoinformatics and Meteorology, University of Pretoria, South Africa | Agricultural Research Council-Institute for Soil, Climate and Water (ARC-ISCW), Division of Geoinformation Science, Pretoria, South Africa

DECLARATION

I, Janique Dustine Savy declare that the thesis, which I hereby submit for the degree MSc Geoinformatics at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

SIGNATURE:

DATE: Wednesday, 8 February 2023

ABSTRACT

USING FREE RESOURCES FOR THE CREATION OF DIGITAL ELEVATION AND GEOGRAPHIC DATA: THE CASE STUDY OF RODRIGUES ISLAND

by

Janique Dustine Savy

Supervisor:	**Dr Christel Hansen**
Co-supervisor:	**Dr Cilence Munghemezulu**
Department:	**Geography, Geoinformatics and Meteorology**
University:	**University of Pretoria**
Degree:	**MSc Geoinformatics**

The acquisition of spatial data can be a problematic process, especially for geographically isolated areas or those where fieldwork is difficult. It is, therefore, important to explore non-field-based methods of producing geospatial data layers for such areas. Here the use of freely available resources and methods of producing geospatial layers are evaluated with the aim of producing basic basemap features such as contour lines, rivers, towns, and roads. These methods are statistically analysed and validated to ensure the accuracy of the features produced. Rodrigues island (Mauritius), is used as the study area, covering an area of 104 km^2 with the highest peak (Mont Limon) reaching 396 m a.s.l. The island offers a dynamically varied terrain ranging from steep slopes to relatively flat coastal regions, allowing the methodology to be tested over all terrain types. Elevation points were produced using freely available resources, such as Training Center XML (TCX) Converter, and Terrain Zonum Solution. These were interpolated using GIS to create DEMs using two interpolation methods (Inverse Distance Weighted (IDW); Ordinary Kriging). IDW was chosen as a simple interpolation method, Ordinary Kriging as a more statistically robust method. The output DEMs were used as the basis for subsequent data extraction and creation. Hydrological modelling was used to model drainage lines; towns, roads, and dams were manually digitised using the freely available software Google Earth™ as the source. With statistical validation IDW proved to predict elevation values that correspond/correlate more with the elevation values of the control DEM, than those generated from the Ordinary Kriging. However, both methods returned outputs that closely resembled the control DEM and were deemed to be acceptable for data creation. Once all required geospatial layers were produced, they were compiled into a complete basemap and compared to the geospatial data collected by the Surveyor General of Mauritius. Although both maps were similar, multiple areas of differences were identified; these areas were ground truthed to determine and validate the findings. Ultimately it was determined that users can produce basemap features of sufficient accuracy for areas that either do not have geospatial data available or are difficult to access. As such, the framework proposed here may be followed to create basic geospatial layers for other inaccessible areas that exhibit similar geographic characteristics.

Keywords: Basemap, Digital Elevation Model, Inverse Distance Weighting, Kriging, Map, Remote Sensing

ACKNOWLEDGEMENTS

I would like to say a special thank you to my primary supervisor Dr Christel Hansen for all her encouragement throughout this process and for the supervision of my master's studies. Her time and patience with me were greatly appreciated.

I would like to thank my secondary supervisor Dr Cilence Munghemezulu for his added supervision throughout this master's studies.

I would like to say thank you to the University of Mauritius with a special thank you to Dr Boojhawon Ravindra for their assistance with providing data which was crucial for this thesis.

A special thank you goes to my parents Steve and Liesel for making this great opportunity possible and never giving up on me. Their never-ending support is what motivated me to push harder.

Lastly, I would like to thank my family and friends who all helped push me and motivate me.

Table of Contents

Table of Figures

Table of Tables

Table of Equations

CHAPTER 1: INTRODUCTION

1.1 Introduction

Geospatial datasets of known accuracy, precision, currency, and reliability are integral to the management of any geographical area. Geospatial data are needed for achieving many of the Sustainable Development Goals (SDGs) and their indicators (Anderson, et al., 2017; GEO, 2016; Paganini, 2018; UN, 2020; Walter 2020) such as zero hunger (SDG 2), clean water and sanitation (SDG 6), and life on land (SDG 15) (Simelane et al., 2021). This is highlighted by the United Nations initiative on Global Geospatial Information Management (UN-GGIM), which promotes the use of geospatial data in addressing the challenges as defined by the SDGs (Global Partnership for Sustainable Development Data, 2019). Cognisant of the role that geospatial data play in achieving the SDGs, the UN-GGIM has defined 14 geospatial data themes, and the Data4Now program (Global Partnership for Sustainable Development Data, 2019). This program seeks to increase the sustainable use of frameworks, methods and tools that improve the timeliness, coverage, and quality of SDG data through, among others, information sharing.

Information sharing is closely linked to the Fair Data Use principle. This principle refers to the idea that data should be used in an ethical, responsible, and transparent manner, with consideration for privacy and personal data protection, and respect for intellectual property rights (Crawford, 2016, O'Neil, 2016, Domingo-Ferrer, 2017, European Union Commission, 2019, Floridi, 2019). Furthermore, data should be FAIR (Findable, Accessible, Interoperable, and Reusable). It is a set of principles used to ensure that digital data and information are managed in a way that makes it easy to find, use, and reuse. The aim of FAIR is to promote data sharing and collaboration, while ensuring data privacy, security, and ethical considerations (Wilkinson et al., 2016; Fritz et al., 2019). Recent advancements in Web GIS, cloud computing, and mobile GIS, have meant that geospatial data are becoming more collaborative, accessible, open, and easily shared, aligning to FAIR (Fritz, et al., 2019).

Key to the creation, analyses, and interpretation of geospatial data are Geographic Information Systems (GIS) (Sheldon, 2018). GIS have powerful mapping and geovisualisation capabilities with data integration, analysis, and modelling (Nkeki and Asikhia, 2014). Their ability to create, organize, analyse, manage, and display geospatial data, in conjunction with the ability to allow for the interpretation of geographical patterns, makes GIS a useful spatial tool for supporting decision makers relating to many geographical problems, such as development of new infrastructure, the impacts of water courses on a given topography (Tomar & Singh, 2012). However, while the need for reliable geospatial datasets is acknowledged, such datasets are not necessarily available nor of sufficient quality for the intended application of analyses. Furthermore, accurate, reliable, and current data of high resolution are often expensive. This raises the question of how geospatial datasets can be obtained in a less complex, accessible, affordable manner, without compromising on data quality.

The advancement of technology, both hardware and software, increasingly yields more advanced products for the creation of geospatial data. Earth observation derived geospatial datasets can now be produced at a known and greater accuracy, and precision (Simelane et al., 2021). The provision and integration of remote sensing derived data with other GIS tools and methods, thus assists in the production of meaningful geospatial datasets and information (Simelane, et al., 2021). For example, Google Earth™ is an accessible and freely available platform that provides users with access to high spatial resolution satellite imagery. Numerous satellite and aerial imagery are superimposed to produce this virtual

globe program (Daly, 2016). This imagery is combined with digital elevation model (DEM) data captured through NASA's Shuttle Radar Topographic mission (SRTM), augmented by other products including ASTER GDEM2, GMTED2010, and NED (Farr et al., 2007), to provide the user with elevation data of areas of interest (El- Hallaq & Hamad, 2017). However, users are limited in the ways in which they can use these data. Google Earth™ does not allow the user to view or do any analysis on the underlying DEM, as well as in terms of the imagery within Google Earth™ users are unable to perform basic or advanced imagery processing or classifications. Irrespective of this, Google Earth™ is a useful resource for research, school projects or simple visualisation. As such, there has been increasing interest in the accuracy of Google Earth™ and its data available to the public.

Cognisant of the applicability and usefulness of Google Earth™ in geospatial analyses and data collection, this thesis evaluates the suitability of Google Earth™ data once combined with a variety of GIS tools, and remote sensing products to produce a digital geographic dataset of a given area. Furthermore, the focus is on inaccessible areas, for which geographic data are either lacking, or difficult or obtain. Here, the targeted area is Rodrigues Island (Rodrigues), a 104 km^2 island situated approximately 600 km east of Mauritius, and the tenth district of Mauritius (Hantke & Scheidegger, 1998). Rodrigues is geographically isolated, with access provided via commercial flights to the island from Mauritius; all other access is via ocean travel. Mauritius and Rodrigues are some of the most ecologically degraded locations on earth (Sodhi et al., 2013). Degradation of these islands ranges from major habitat loss and severe erosion to detrimental consequences such as the extinction of various species. This has led to a wide range of conservation efforts both on a local and international level. However, an issue within conservation methods already underway is their lack of evidence-based knowledge, which can be obtained using geospatial data (Sodhi et al., 2013). Conservation efforts, management of geographic space, and planning within that space require reliable, accurate, and current geospatial data (Anderson et al., 2017). With the island both difficult to reach, making field-based data collection difficult and/or expensive, and the island having known environmental issues related to erosion that require accurate geospatial data, this thesis aimed at creating a reliable and accurate geospatial dataset for the island. This dataset is subsequently assessed for accuracy precision.

1.2 Research Problem
1.2.1 Problem Statement
Access to accurate, current, and precise geographic data are increasingly becoming important, however, required geographic data are often either not available, or not of sufficient quality. Numerous remote sensing techniques are available to users to derive data for a given area but, these techniques are not always done with a known accuracy rate in mind. Similarly, geospatial methods are not always evaluated for accuracy and precision, often each method relies on numerous other methods to be completed, introducing errors at any stage. As such, error propagation can become problematic if not accounted for. This thesis aimed at testing the suitability of a combination of open and free methods employed throughout to provide users with geospatial dataset at a known accuracy and scale for an area where the creation of such a dataset would be difficult to obtain due to its general inaccessible geographic location.

1.2.2 Research Aim and Objectives
This thesis aimed to create digital geospatial datasets with a known accuracy of a given inaccessible, or hard-to-reach area using remote sensing, geospatial data, and desktop

techniques.

Data derived from these methods are utilised to map areas that are difficult to access or that have not been surveyed. Rodrigues, located in the Indian Ocean, is used as a case study to create such a dataset, due to the island being geographically isolated and therefore relatively inaccessible. Furthermore, only limited geospatial data are currently available for the island. Finally, the island has a highly varied topography and multiple geomorphic zones (slopes, coastlines, steep valleys, and heavily vegetated versus bare areas). This varied topography can affect the accuracy of the digital data created. By assessing different topographic zones, modelled data can be assessed per zone, providing insight into the accuracy of modelled data for different topographies. The accuracy of the output data produced during this thesis underwent validation tests such as correlation analyses, Students t-test, F-test, as well as ground truthing to determine the level of similarity between the control geospatial data and geospatial data created for this thesis.

The study area of Rodrigues was chosen as it posed as a large enough study area that included factors such as varying terrain, costly to travel to and includes inaccessible areas. This study area has, furthermore, been surveyed and geospatial data for the island was thus available to allow for a comparison dataset required for data accuracy assessments.

The aim is achieved through four key objectives. These are:
1. Using satellite and aerial images to create a digital geospatial dataset of Rodrigues.
2. Evaluating the accuracy (high and low) of the created geospatial data, using evaluation methods such as the student t-test, f-test, correlation coefficient and ground truthing.
3. Creating a basemap of Rodrigues.
4. Developing a methodology that can be used to map inaccessible areas at a known and quantifiable accuracy using freely available remote sensing data and geospatial techniques.

1.2.3 Research Questions
The following questions are addressed for the completion of the study:
1. Is it possible to map an area without visiting the area?
2. How does the accuracy of Google Earth's ™ elevation data affect the results of the output DEM's in comparison to the underlying control DEM?
3. What is the accuracy of mapping different topographies (such as slopes, valleys, and the coastline) using free resources?
4. What is the most accurate technique/method to map a specific map feature, both man-made or natural, such as a road or a river?

1.3 Thesis Organisation
This thesis is organised into five chapters. This chapter introduces the topic, whereas CHAPTER 2: LITERATURE REVIEW (pg. 4 onward), provides the background to the topic under investigation. CHAPTER 3: METHODOLOGY (pg. 12 onward) discusses the methods employed for this thesis, including Study Area (pg. 12), Data Requirements (pg. 12 onward), Database Management (pg. 26), Validation (pg. 27), Base Map Compilation (pg. 29), and Ground Truthing of outputs achieved (pg. 29). The same subheadings are utilised in CHAPTER 4: RESULTS AND DISCUSSION (pg. 31 onward). The thesis concludes with CHAPTER 5: CONCLUSION (pg. 58 onward). This chapter also discusses Limitations and Considerations (pg. 58) to the study, as well as Further Areas of Research (pg. 59).

CHAPTER 2: LITERATURE REVIEW

Geographical Information Systems (GIS) are computer-based tools that are used, among others, for the mapping and analyses of geospatial data (Nwauzoma, 2016). These systems allow for data to be analysed and mapped within less time than it would have previously taken to do these processes and analysis manually, enhancing mapping efficiency (Tomar & Singh, 2012). Such datasets range from simple vector files including points, lines, and polygons to the more complex raster files including LiDAR, satellite, and aerial imagery; this also includes Big Data. Software such as the ESRI (Environmental Systems Research Institute) suite of products, including ArcGIS Pro, and QGIS (Quantum Geographic Information System) provide the geoprocessing tools, such as hydrological analysis tools and interpolation tools to yield graphic data representations and geospatial analyses. The widespread distribution of such software along with the availability of high-resolution remote sensing data, has provided the field of mapping with a great capability. These computer-based technologies have led to the alteration of analysis and the improved visualisation techniques of landform data across the globe (Napieralski, et al., 2013), while providing an efficient and cost-effective method of mapping large areas (Simelane et al., 2021).

Geospatial data are data that have been geographically referenced, meaning they have a geometry and location on the Earth's surface; this data forms part of the central component of any GIS (eGyanKosch, 2018). Furthermore, geospatial data form the basis of any basemap. Therefore, to obtain a successful GIS-based investigation it is crucial to have an accurate basemap (Longley, et al., 2015), and as such, accurate geospatial data that forms such basemaps. A combination of geospatial data is effortlessly combined and analysed within a GIS (Folger, 2009). According to Strobl and Nazarkulova (2014), geospatial data is the fuel of GIS and without these data GIS are essentially deemed unsuitable for geospatial analyses. The importance of geospatial data can thus not be underestimated, and this thesis evaluates how geospatial data for areas that have little to no geospatial data available to the public can be created at low cost, using freely available resources and methods.

Geospatial data are not always accessible nor freely available. Furthermore, their quality is not always known nor suitable for the problem at hand. However, with the increase in remote sensing, which allows observing and subsequent data creation of large areas of the earth without requiring field access, geospatial data have been made more accessible, and current. However, while there are multiple sources of free geospatial data these data are usually minimal and not current. As such, such datasets need to be analysed and adapted to improve their quality or to make such datasets suitable for the intended analyses purposes. Public access to these data can also be limited, expensive or come with numerous use constraints, such constraints included not being able to use the data for work purposes and only for a given research project or not being allowed to display the data in one's final findings (Strobl & Nazarkulova, 2014). This raises the concern that, while geospatial datasets might exist for your project or problem, access might be restricted, or the data have poor quality. This is contrary to the FAIR data use principle, and the open-access paradigm.

The open-access paradigm highlights the need to make (any) data, including geospatial data, freely available sans restrictions on use, reuse, and distribution (Borgman, 2015, Rajasekar & Arunachalam, 2015, European Union Commission, 2017, Piwowar et al., 2017). This ensures data and the methods of data creation, analyses, and structure are transparent, allowing for reproducibility (Yiotis, 2005), an integral component of any scientific work (European Union Commission, 2017). Open-access data are also more equitable, in the sense that freely available data reduces the divide between the affluent and poorer users.

This is important when considering that, for example, high-resolution satellite imagery can be expensive. Open-access data also conforms with the FAIR data use principles: Findable, Accessible, Interoperable, and Reusable (Wilkinson et al., 2016; Fritz et al., 2019). Furthermore, open-access data supports the notion of open geographic information science (Peng et al. 2006, Rey 2014; Singleton et al., 2016). This paradigm and the principles discussed above provide a platform for the sharing of data. In places where little data are available, this becomes important. Furthermore, by clearly stating methods of data acquisition and the inclusion of metadata, ensures data are more likely to be used for their intended purposes. Thus, costs are reduced for users requiring such datasets, since data already exist and are available if published under this paradigm.

Geospatial data, once obtained, forms the basis not only of geospatial analyses but also of mapping any given geographic area. Maps produced, depending on their purpose, and intended audience, have multiple uses. For example, topographic mapping has undergone a major evolution from the years 1879 to the present. Between the years 1879 – 1990 topographic mapping was only a map factor operation, which involved obtaining data through field collection and photogrammetry (Paganini, 2018; Walter, 2020). These data were then verified and annotated to create a graphic output (Paganini, 2018; Walter, 2020). This process then evolved into a process that included the use of GIS, where the inclusion of GIS allowed for digital database preparation, and operations for product generation. Overall, the evolution into GIS has led to cheaper, faster, standardised, semiautomated production methods, which are increasingly used to produce mapping outputs (e.g., Paganini, 2018; Walter, 2020).

Mapping can focus on the creation of basemaps. Basemaps (the collection of geospatial data, and the inclusion of orthorectified imagery to provide context by providing reference and orientation data) can be overlaid by other information layers to create any type of map (Jones, 1982). As such, basemaps are important components to any mapping exercise and concomitant geospatial analyses. However, basemapping is often neglected. Yet the increasing technological advances and improved product derived through remote sensing are yielding better basemaps (Aplin, 2003). The most traditional technique to produce basemaps is the use of aerial photography and photogrammetric techniques (Holland & Allan, 2001), based mostly on the fine spatial resolution of aerial photography. However, with the advances seen in the field of remote sensing this method is now not the only option, with satellite imagery increasingly being used to produce basemaps (Jiao et al., 2001). With the emergence of higher spatial resolutions in satellite imagery seeing values of <1 m resolution on panchromatic imagery, and <4 m resolution on multispectral imagery since 1999 (IKONOS satellite), with sub-cm registered today, spatial resolution details are sufficient to produce accurate basemaps compared to the traditional aerial photograph method (Hanley & Fraser, 2001). Although Landsat imagery has been available since before the IKONOS satellite imagery, its resolution of 30 m places it at a limitation to produce basemaps through remote sensing. In contrast, satellites such as IKONOS offer a much higher spatial resolution allowing for more detailed output basemaps (Aplin, 2003). Recently, other products, as those from freely available Sentinel-2 imagery from the European Space Agency (ESA), have increased the ease of obtaining high resolution imagery. Furthermore, satellite products have a known accuracy. For example, the ready-to- download Sentinel-2 products are geometric and radiometrically corrected (Du et al., 2016). With satellite imagery increasingly being made available to the public at both free and proprietary levels at finer spatial resolutions (Aplin, 2003), remote sensing is becoming the preferred method in terms of costs, and currency of data for producing accurate basemaps.

Basemaps provide the basis for other mapping applications (Jones, 1982), and subsequent analyses. They provide a canvas for one's data to be displayed to allow for better interpretation of one's data. It is crucial for the user to choose the correct type of basemap for their given outcome as different basemaps will illustrate different aspects of a particular area. An example of this is a topographic basemap, where the focus is on the terrain and landscape of an area of interest whereas a streets basemap will have more focus on the street network. As such, a basemap can take your data from being random points, lines, or areas on a map to giving your data context, providing for insights into your data (Bounds & Sutherland, 2018). Such context can be, for example, geomorphological mapping, an improvement of simple topographic mapping. Simple topographic maps allow for some geomorphological landforms to be distinguishable but may leave out landforms that are not obvious to the eye. Furthermore, they are unable to illustrate information about the formation of landforms such as their genesis or distribution (Gustavsson, 2006). Yet the mapping of geomorphological landforms is seen as a form of graphical inventory, which can be utilised as a preliminary tool when dealing with land management, and geomorphological risk management, including erosion management (Otto & Smith, 2013; Gustavsson, 2006). Geomorphological maps, either basic or analytical depending on the type and detail of the data to be mapped (Otto & Smith, 2013), can provide baseline data for an unlimited number of sectors such as landscape ecology, forestry, and soil sciences.

As remote sensing data are increasingly used to compile basemaps, such data can also be used to produce, for example, basic geomorphological classification maps. This stands in contrast to geomorphological mapping based on fieldwork methods. Using remote sensing data enhances the objectivity and efficiency of landform measurements and classification (Napieralski et al., 2013). Through the array of available Earth-observing satellites one can clearly visualise different perspectives of the geomorphological landform that occupy a specific region of interest, users can visually identify mountainous regions as opposed to flat open regions, and coastal regions. Integral to these sectors are digital elevation models (DEMs). A DEM forms the basis of a geomorphological map (Otto & Smith, 2013). The DEM depicts geomorphological landforms at a basic level to the accuracy of the spatial resolution of the data. These identified geomorphological landforms can subsequently be used, in conjunction with other data such as rivers and infrastructures (roads and settlements), to manage issues such as erosion. Of note here is that a higher resolution dataset yields better differentiation of landforms.

Using remote sensing data such as satellite images and aerial photography, DEMs can be created (Tomar & Singh, 2012), through either, for example, Triangular Irregular Networks (TINs) or through Grid DEMs. A DEM provides a 3-dimensional illustration of elevation levels, which improves the readers ability to identify basic geomorphological landforms such as ridges (high elevation, narrow contour lines) and valleys (low elevation, u- or v-shaped contour lines). A DEM is an elevation model illustration the "bare" Earth, this model is supposedly free of all nonground objects such as trees and buildings. This type of elevation model is also referred to as a Digital Terrain Model (DTM) (Zhou, 2017). The other type of elevation model available to users is that of a Digital Surface Model (DSM), this is an elevation model that includes the tops of ground features such as treetops and buildings (Zhou, 2017).

A DEM is an inevitable component within the fields of remote sensing and GIS as it reflects the physical surface of the earth thus allowing the user to understand the nature of the terrain. With DEM's being used for numerous reasons such as 3D simulations, estimating river channels, contour maps and determining slope and aspect, there is an emphasis on the

quality of the DEM. There are numerous methods to produce a DEM, the method used by the user is dependent on their available source data and skill level (Table 2.1) (Nelson et al., 2009).

Table 2.1: Digital Elevation Model methodologies and techniques *(Nelson, et al., 2009).*

Method	Data Format	Interpolation	Comments
Airborne laser scanning	Point data	IDW	• Lower cost than photogrammetry
Ground survey	Elevation points	TIN	• Expensive • Time consuming • Not suitable for large areas
High spatial resolution satellite data		Contour interpolation	• Available at low cost • Requires cloud free input data
Light Detection and Ranging (Lidar) satellite imagery	Point values	Kriging	• Produces accurate high-resolution DEM • Unable to penetrate through high-density vegetation • Large dataset harder to process and interpret
Photogrammetric survey with manual interpretation	Contours and Spot Heights	Kriging	• Requires aerial photography • Requires a skilled user • Errors occur over vegetation
Photogrammetric survey with automatic interpretation	Correlated points	Kriging	• Requires aerial photography Issues occur over vegetation and non-ground points
Radar satellite imagery	Raster DEM	None	• Lower cost than photogrammetry • Issues seen with steep slopes and vegetated areas
Stereoscopic satellite imagery	Point data	Correlation for surface points	• Lower cost than photogrammetry • Issues occur over vegetation and non- ground points
Topographic map data	Primarily contours	Kriging	• Cheap and readily available data • Issues seen with steep slopes (16 – 30%) and densely vegetated areas

DEMs were traditionally created using surveyed points or using photogrammetry from aerial photographs (Toz & Erdogan, 2008). The creation of DEMs typically involves the scanning and digitizing of contour lines from topographic maps. This is a time consuming and tedious exercise and allows for human error. With the ever-increasing need for geospatial data within a GIS, DEMs such as SRTM, Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), and Japan Aerospace Exploration Agency Advanced Land Observing Satellite (JAXA ALOS) are freely available to the public. Since 2003, the SRTM DEMs are freely available and easily accessible to the public (Alganci, et al., 2018). These DEMs are available at both 90 m and 30 m intervals with the 30 m DEM a resampled product of the 90

m DEM. Therefore, it is important to know that there may be discrepancies within the accuracy of this dataset (Alganci, et al., 2018). Through a study done by Schumann and Bate (2018) it is acknowledged that although SRTM is widely and freely available to users it is important for users to be made aware of the large vertical errors, which can be experienced over complex topography.

Shuttle Radar Topography Mission (SRTM) is widely compared to the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) DEM (first released in 2009), with both relatively similar (Alganci, et al., 2018). In a study by Kervyn *et al.* (2008), which involved the mapping of topographic features of volcanoes, the two DEMs (ASTER and SRTM) were compared. This study found that although ASTER offered better temporal resolution of 16 days, its ability to characterise morphology and topography was limited by the inclusion of small-scaled artefacts it produces from matching errors, making the SRTM DEM a better choice (Kervyn, et al., 2008). The published absolute vertical accuracy of the SRTM DEM is 16 m, with that of the ASTER DEM 12 – 15 m (Nikolakopoulos, et al., 2006; Alganci, et al., 2018). The horizontal accuracy of the ASTER DEM is approximately 6 m (Alganci, et al., 2018). However, research by Elkhrachy (2018), has shown that much greater vertical accuracies are possible. In his study, the author showed that vertical accuracy between ~5.94 – ~6.87 m is possible, when assessing SRTM data against GPS reference elevations, and topographic elevations respectively. Conversely, vertical accuracies for the ASTER DEM were calculated at ~5.07 –~7.97 m, when compared to GPS reference elevations, and topographic elevations respectively (Elkhrachy, 2018). Since June 2005 Google Earth™ has utilised SRTM elevation data as their elevation baseline. However, several DEMs are used in conjunction with the SRTM DEM at 1-arc spatial resolution (~30 m). This is since the SRTM data has been void filled using ASTER GDEM2, GMTED2010, and NED (Farr et al., 2007). This yields baseline elevation data 5 – 20 times higher in resolution than that of the South African 1:50 000 data obtained from the Chief Directorate Surveys and Mapping (CDSM) (El-Ashmawy, 2016). Relatively recently (first release in 2015) the Japanese Space Agency (JAXA), released a Digital Surface Model (DSM) known as the AW3D–30 m (Alganci, et al., 2018). The original data are captured at 2.5 m intervals. However, the DSM of this resolution is not freely available to the public. Instead, the upsampled 30 m DSM is made freely available (Alganci, et al., 2018). This product, the JAXA Advanced Land Observing Satellite (ALOS) is also known to have slightly higher spatial vertical accuracies (~ 3.28 m) than the ASTER or SRTM DEM (Alganci, et al., 2018; Takaku, et al., 2016).

As illustrated here, many DEMs are freely available to the public with a worldwide coverage via the likes of ASTER, JAXA ALOS, and SRTM at 30 m resolutions. Each of these products has a known spatial accuracy, although such accuracies differ depending on the geographic location, and due to other parameters, such as cloud cover presence in the original images (Alganci, et al., 2018). These accuracies are important to note as when mapping and understanding the landscape of a small area since such DEM data are often not detailed enough, therefore placing a greater need to obtain/create a DEM with a finer scale of less than 30 m resolution. Finally, of the three DEMs discussed above, the SRTM product has been most widely used due to its longer availability. As such, while the JAXA ALOS DEM has a higher vertical spatial resolution, the SRTM DEM is used for validation purposes (refer to Interpolation, pg. 18). The study done by Schumann and Bates (2018) aimed at identifying the limitation of these freely available DEMs as all sources including SRTM, and ASTER were found to have vertical errors that could not resolve microtopographic variations.

In addition to geospatial data being easier to obtain, acknowledgement must be given to the

plethora of tools available to assist this process. For example, Google Earth™ provides one method of obtaining geospatial data at low costs. Google Earth™ is freely available software that allows for the visualisation of all parts of the globe. This software is a combination of various data sources such as satellite and aerial imagery, roads, and boundary data. Although it is freely available, Google Earth™ has limitations with regards to how the user can use and manipulate the data. Studies undertaken by El-Hallaq (2017), and Mohammed et al (2013) have shown that Google Earth™ has ranging vertical and horizontal positional accuracy rates, which should be taken into consideration when using such data. El-Hallaq (2017), questioned the positional accuracy of Google Earth™ over the Gaza Strip and found a positional accuracy of 39.235 m in a north-easterly direction, subsequently recommending that Google Earth™ only be used as a tool for preliminary studies or investigation. In contrast, horizontal accuracy studies have shown that Google Earth's™ elevation data may successfully be used at a small to medium scale with a 1.8 m accuracy (Mohammed, et al., 2013)**.** When making use of such data, although it is freely available to the public, the terms and conditions to the use and publishing of acquired data need to be adhered to. For example, all data created in Google Earth™ belongs to the enterprise and not the individual. Such licensing restrictions has implications on any data created in Google Earth™ and subsequent research. Irrespective of this limitation, many advantages of using the product remain. Some of these key advantages and limitations of Google Earth™ are given in Table 2.2.

Table 2.2: Principal advantages and limitations of Google Earth in geomorphology (Tooth, 2013).

Advantages	Limitations
Basic quantitative measures (for example, distance and slope).	Accuracy of measurements is reliant on the spot height data.
Enables spatial thinking (for example, patterns and comparisons).	Temptation to be uncritically used to overlook other forms of analysis (for example, fieldwork).
Imagery is regularly updated (for most parts of the globe)	Updated imagery does not guarantee improved geomorphological analysis potential.
Import data (users can import their own data for overlay).	Unable to alter map datums and projections.
Local observations and measures within a broad spatial context.	Uncritical use may result in the superficial reliance on this analysis.
Time and cost effective (instant, free access to imagery).	Limited to optical imagery.

A study completed Rusli and Majid (2012) tested the accuracy of Google Earth™ elevation data to create a DEM to conduct watershed delineation within the Sungai Maur Watershed in Malaysia. The watershed area was identified and divided into 36 sections, which were each allocated a marker within Google Earth™ containing both a latitude and longitude attribute. Rusli and Majid (2012) exported their sample points using Terrain Zonum Solution, a free software program used to obtain elevation data (refer to Figure 2.2).

Figure 2.1: Extraction process.
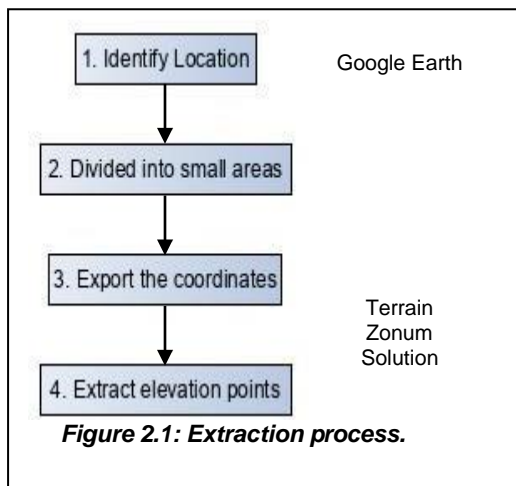
These data were then mapped in ArcGIS Pro as point data containing a latitude, longitude and elevation attribute (x,y,z). The results of this study showed a 0.19% difference in elevation between the Google Earth™ elevation data and the 20 m elevation control data. The comparison results between the Google Earth™ elevation points and a 20 m elevation control data are seen in Table 2.3.

Table 2.3: Height comparisons of two sources of elevation data (Rusli and Majid, 2012).

Area	Source	Mean (m)	Standard Deviation
Flat (0 – 10%)	Google Earth	12.87	6.41
	20 m Interval	14.18	7.36
Hilly (>10%)	Google Earth	62.54	13.15
	20 m Interval	61.94	12.30

The study by Rusli and Majid (2012) illustrates the use of freely available geospatial resources to obtain geospatial data about a particular geographic area. These freely available geospatial resources included Google Earth, Terrain Zonum Solution, with the additional use of ArcGIS 9.3.1. Another such resource is TCX Converter, which is like Terrain Zonum Solution, as applied by Rusli and Maid (2012). Bozek et al. (2016) showed an approximate accuracy of 3 m when comparing data derived from Google Earth™ and TCX Converter, to a LiDAR point cloud. Similarly, the combination of Google Earth™ derived data used in conjunction with TCX Converter has been successfully used for geospatial data creation and analyses (Bozek, et al., 2016; Kumari, 2018; Vangu & Dima, 2018), including the creation of digital contour maps. Marsudi (2017) completed a study comparable to Rusli and Majid (2012). This study differed slightly in the exporting of elevation values for a given latitude and longitude as it made use of a differently freely available software, being TCX Converter.

From the discussion above it can thus be seen that numerous data sources, products, and methods exist that allow for the creation of geospatial data at no or minimal cost. Here, Google Earth™ is used to derive control points and digitising of features; free software such as TCX Converter used to obtain z-values for points created in Google Earth™, GIS such as QGIS used to create and manipulate geospatial data, and results compiled in a basemap, with the aim of creating a useful and suitable base geospatial dataset for regions that are either inaccessible, or where data are too expensive to obtain.

CHAPTER 3: METHODOLOGY

The aim of this thesis is to produce digital geographic data making use of freely available remote sensing and Geographic Information systems (GIS) resources. To achieve this the workflow of Figure 3.3 was followed. The study area is described in Study Area (pg. 12). This chapter describes the workflow of Figure 2.1, focusing on Data Requirements (pg. 15), Database Management (pg. 26), Validation (pg. 27), Base Map Compilation (pg. 29), and Ground Truthing of outputs (pg. 29).
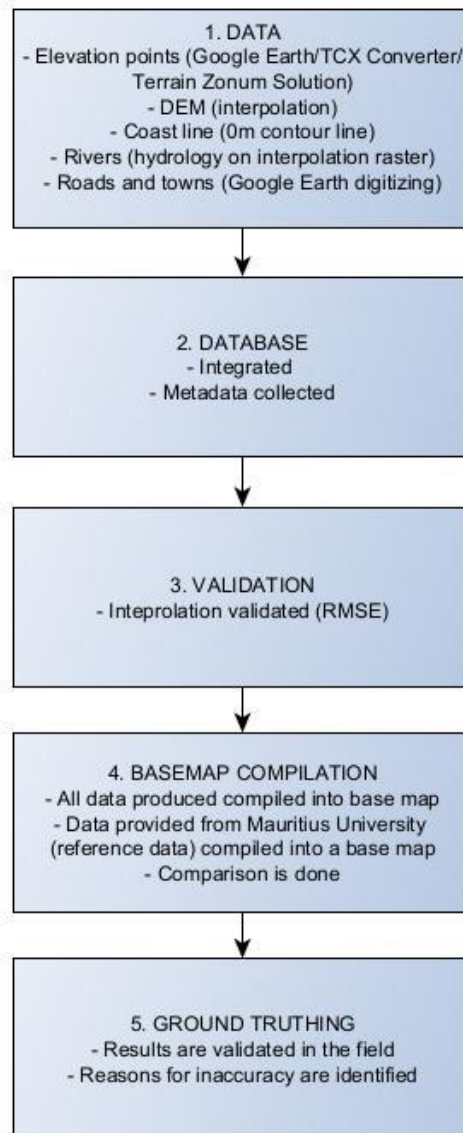
Figure 3.3: Methodology workflow used.

3.1. Type of Research

An empirical quantitative approach was employed, since data were captured using GISc methods, and subsequently validated using ground truthed data. Measurements and observations taken through fieldwork analysis were supplementary to the geospatial data created using the GISc methods. This work further aligns closely to the open-access data paradigm. This paradigm refers to the idea that data should be made freely available to the public without restrictions on use, reuse, or distribution, further aligning to FAIR data use principles (Fritz, et al., 2019). This is based on the belief that open access to data promotes transparency, collaboration, and innovation (Borgman, 2015, Rajasekar & Arunachalam, 2015, European Union Commission, 2017, Piwowar et al., 2017).

3.2. Study Area

Discovered by Arab sailors in the 10th century, and by the Portuguese navigator Don Diego Rodriguez in 1528 (Sodhi, et al., 2013), Rodrigues is a small basaltic island located to the east of Mauritius (Figure 3.4). The island is of volcanic origin with an age of at least 2.5 million years, with an area of approximately 6.5 km by 18 km (104 km^2) and the highest peak (Mont Limon) reaching 396 m a.s.l. Rodrigues is encircled by flat coral reef dotted with small surrounding islands. In 2014 a total of 41 788 people lived on the island at a density of 402 people per km^2 (Unmar & Chinnee, 2015). Its economy is based on tourism, livestock, fisheries, and subsistence agriculture (Middleton & Burney, 2013).
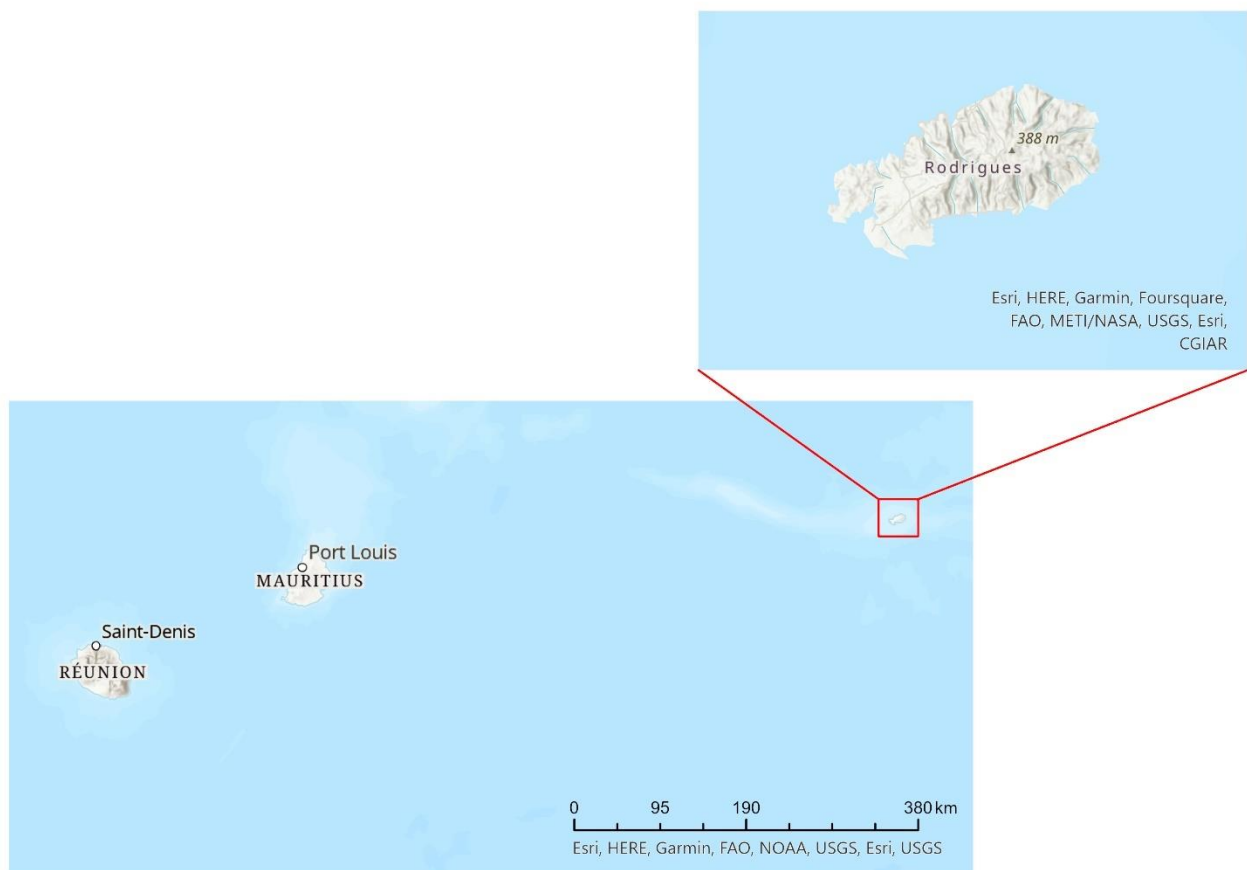
Figure 3.4: Location of Rodrigues in relation to Mauritius.

The island lies along an east-west trending fracture zone and predominately consists of basalt lava with a few cases of calcareous aeolianites (Baxter, et al., 1985). These aeolianites are subject to enhanced erosion, due to the nature of deposition of the calcareous dune sands (Baxter, et al., 1985). Due to this Rodrigues, along with Mauritius, is classed as severely degraded (Sodhi, et al., 2013). This places emphasis on the fact that management strategies need to be set in place to restore the island. Greater knowledge is required on its present state and form. This also requires the mapping of its features, both natural and anthropogenic. To achieve the aim of this thesis, existing geospatial datasets are compared to the created ones. Saddul (2002) classified the geomorphological structure of Rodrigues into the (1) central ridge, (2) inland slopes, (3) coastal environment, (4) western lowland and (5) aeolianite plain, representing one of the few published resources on topographic features of the island. Pasnin et al. (2016), provided a more detailed delineation of the coastal environment using biotope data. The delineations of Saddul (2002) and Pasnin et al. (2016) thus provide reference datasets that can be compared to the data created in this thesis. Furthermore, geomorphological understanding of an area is fundamental to the management of soil erosion and land degradation, prevalent on Rodrigues. As such, the

creation of quality geospatial data, such as Digital Elevation Models (DEMs), hydrological layers, and morphological zones, is integral to decision support and management of this island. The following sections of this chapter describe the workflow of how such a geospatial dataset can be created.

The chosen study site for this thesis, Rodrigues, is severely degraded (Sodhi, et al., 2013), and the creation of basic geomorphological maps may be used as a preliminary tool used for land management strategies, and risk management (Sodhi, et al., 2013; Otto & Smith, 2013). As such, the creation of quality basemaps and geomorphological maps of the island will enhance land management strategies. It must be noted that geomorphology is not the focus of the thesis, but rather that it provides the context for the need of geospatial data. Rodrigues is severely degraded and experiences high erosion rates – a greater understanding of geomorphological zones and processes would assist in mitigating these factors. A geospatial dataset that delineates these zones would then enhance the management of the island.

Thus, this thesis follows the same process of extraction as seen in Figure 2.2 (pg. 10), to obtain elevation data to create a DEM, which was further used to map other features such as hydrology on Rodrigues. As such, Google Earth™, the process as described by Figure 2.2, and the application of GIS methods and tools are used to create a geospatial dataset for Rodrigues. This thesis, furthermore, makes use of both Terrain Zonum Solution, as well as the TCX Converter to obtain georeferenced geospatial data, that can subsequently be used to create a DEM of an area. These two products were chosen to evaluate their user-friendliness when obtaining z-data. These z-values, irrespective of the product used, are based on Google Earth™. Subsequently, geospatial data are created using a variety of methods, including heads-up digitising, and a basemap of the island created.

3.3.   Data Requirements

Figure 3.5 on the next page is an overview of the method used to produce the data needed for the basemap compilation. Relevant steps are numbered on the figure and are referenced as such in the text, for example, data derived from Google Earth™ for eventual input into the TCX Converter is described as 1.1 Google Earth ™ on Figure 3.5.
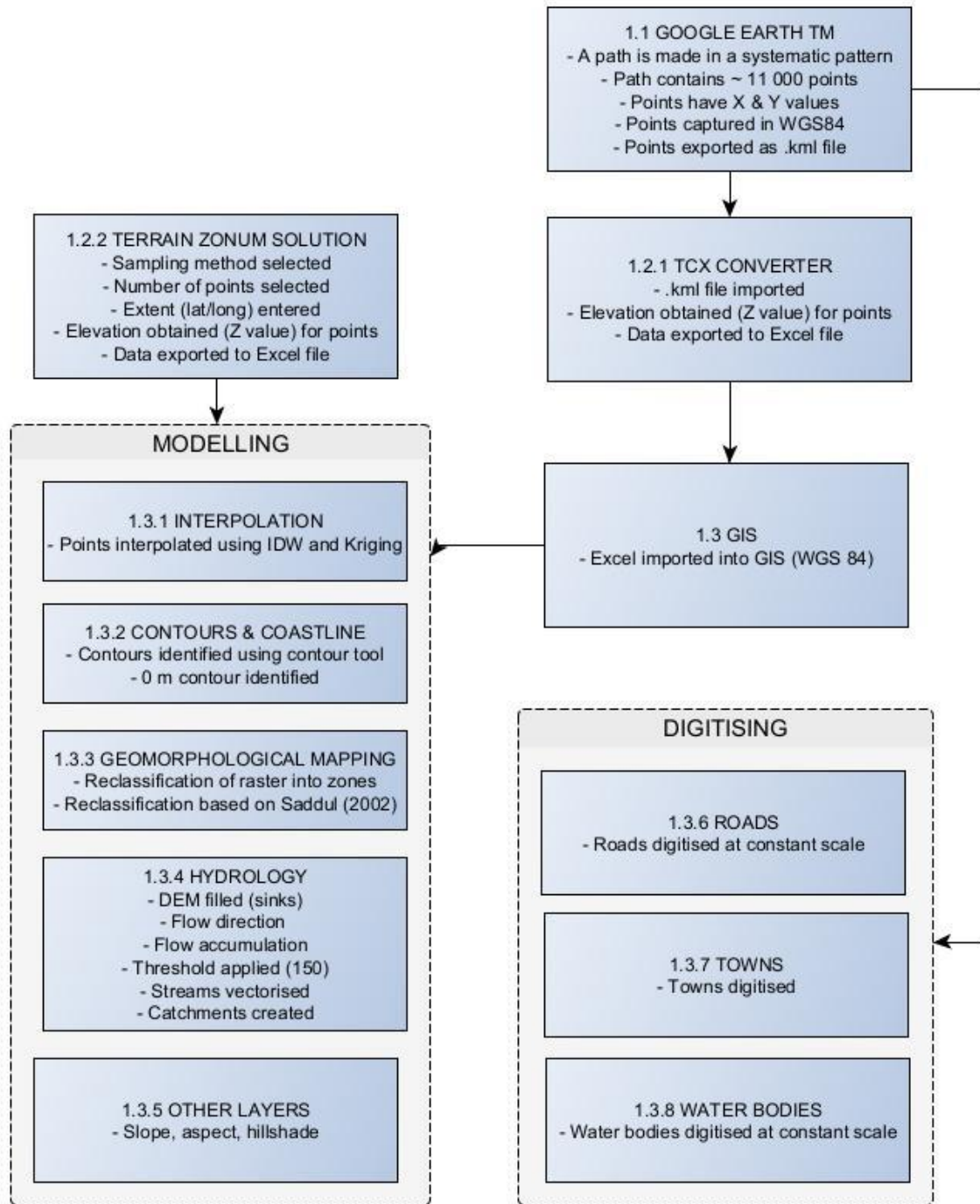
Figure 3.5: Data extraction and creation methodology.

3.3.1 Google Earth™

The aim of this thesis is to use freely available resources to create digital datasets that can be used for geospatial analyses and mapping. Underpinning much of geospatial analyses is a suitable and reliable DEM. To obtain such, Google Earth™ (Figure 3.5 reference 1.1) is used to obtain height values as points that are interpolated to a DEM. Google Earth™ is a product that provides a three-dimensional model of the earth. The platform combines multiple satellite images at various spatial resolutions, allowing the user to view and use the software at any given scale (El-Hallaq & Hamad, 2017). It is important to note that Google Earth™ makes use of interpolation to account for data gaps, yielding data that are not 100% accurate (El-Hallaq & Hamad, 2017). For this study Google Earth™ was used to systematically collect sample points within the boundary of the study site (Rodrigues) to be

used for eventual interpolation to a DEM. The more points obtained the higher the accuracy of the interpolated elevation raster. The *Add Path* tool in Google Earth™ was utilised to obtain these points as it provides a single dataset containing multiple points. The path produced by the *Add Path* tool was exported as a .kml file. A limitation of this process is that although these points have both x and y coordinates, no z-values showing elevation are available. To obtain z-values, an additional freely available product, such as Terrain Zonum Solution or TCX Converter, is used.

3.3.2 TCX Converter

Once points are exported from Google Earth™, TCX Converter[1] (Figure 3.5 reference 1.2.1), a freely available resource, is used to obtain z-values for each point by identifying the elevation for that particular point's location. This is completed by loading the .kml file into TCX Converter and updating altitude values. TCX Converter facilitates obtaining z-values (elevation values) within the WGS 84 coordinate reference frame. The technology utilises x and y coordinates of the .kml file and adds z-values based on the input x and y coordinates (Bozek, et al., 2016; Marsudi, 2017; Vangu & Dima, 2018).

Figure 3.6 (pg. 17) illustrates the TCX Converter interface. The user opens a file (here the .kml file containing all the points from Google Earth™). Once the file containing the points is loaded the user opens the fourth tab labelled *Track Modify* and proceeds to select *Update Altitude*. This button is activated once the converter recognises that data containing a latitude and longitude have been loaded. An Internet connection is required to complete this request as TCX Converter will connect to the elevation data within Google Earth™ at every point containing a latitude and longitude. Once complete TCX Converter produces an Excel file that contains the z-values for all the points that were obtained in Google Earth™. This file forms the input data required to create a DEM of the study area.

---

[1] URL: https://tcx-converter.software.informer.com/2.0/
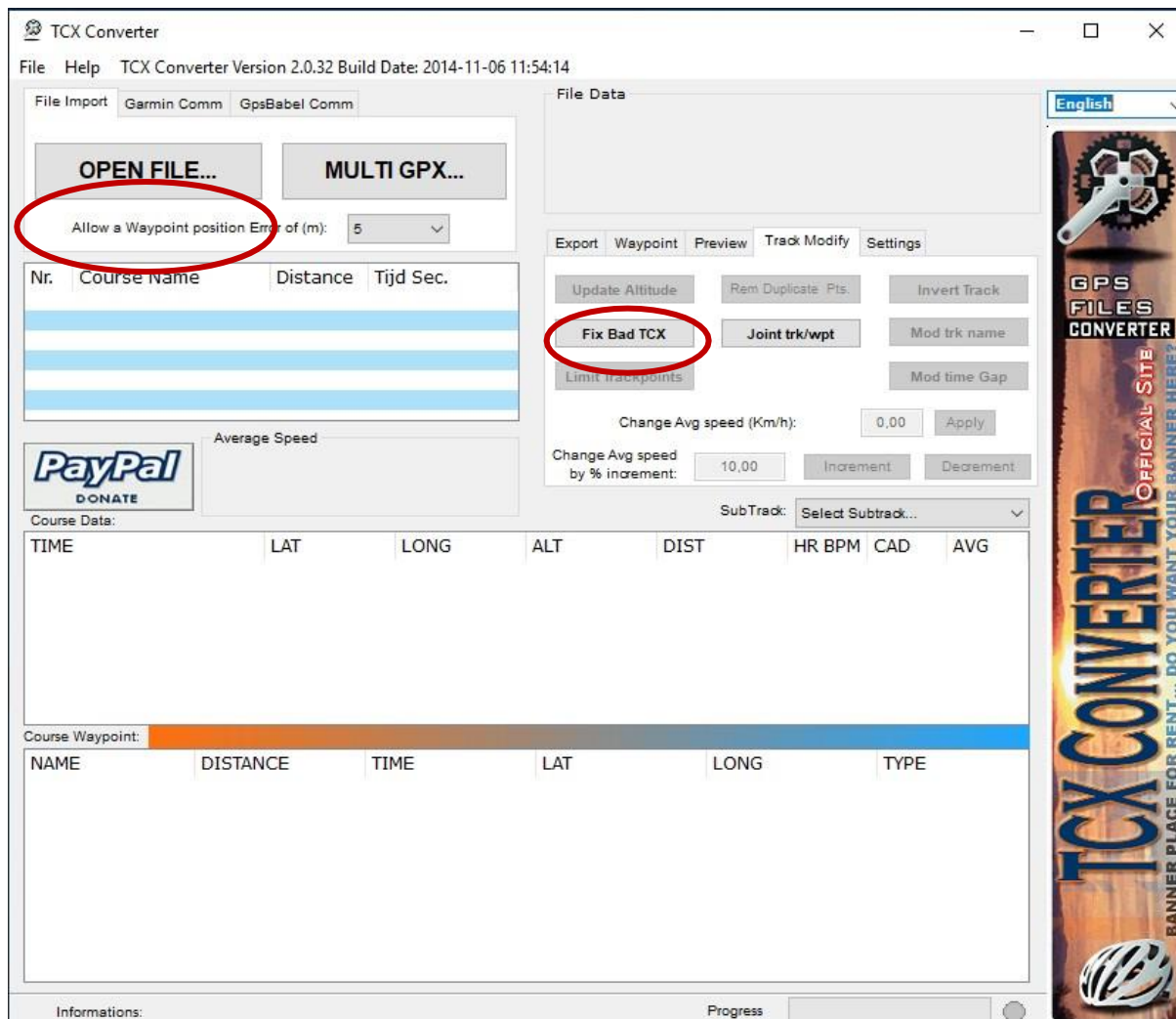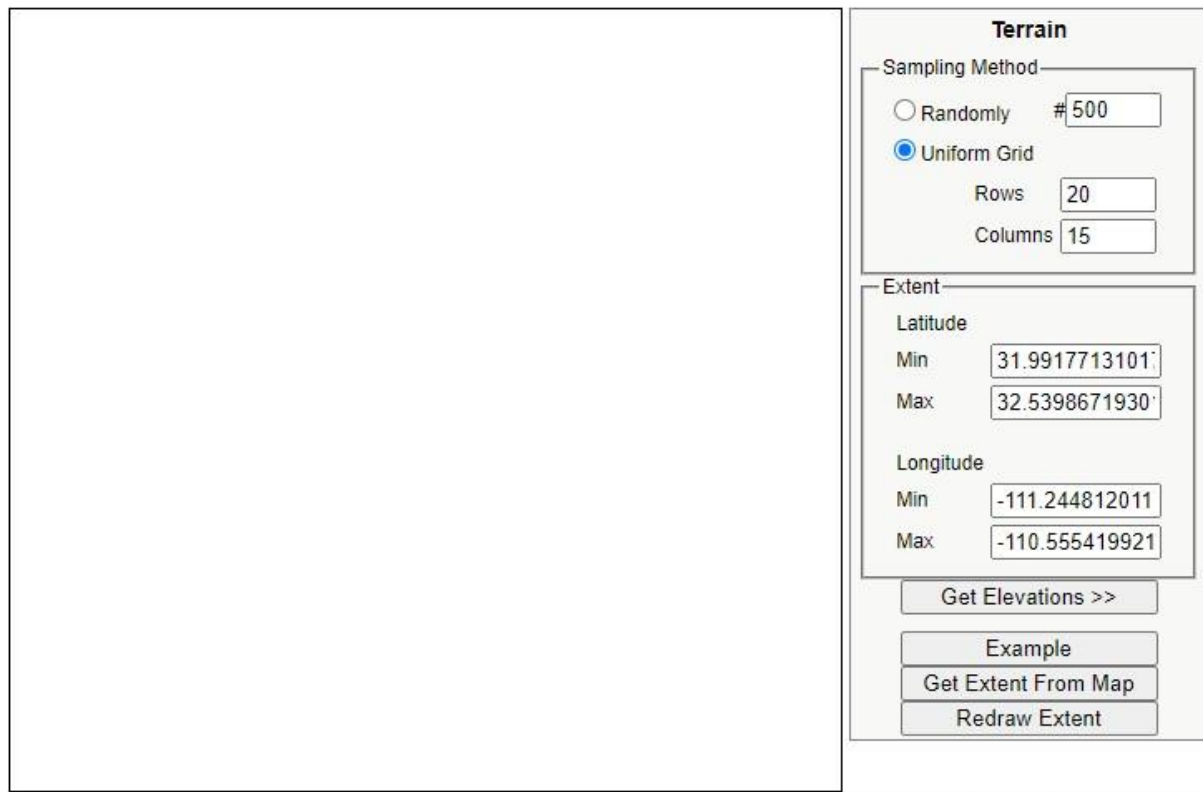
Figure 3.6: Training Center XML (TXC) Converter interface.

3.3.3   Terrain Zonum Solution

Terrain Zonum (Figure 3.5 reference <u>1.2.2</u>) is a browser application[2], to obtain z- values for point locations. This freely accessible web browser does not require the need for XY locations to be added as it produces its own set of point locations based on one of 2 sampling methods, random sampling where the user will specify the amount of sample points or a uniformed grid sampling method where users will specify the grid size. As seen in Figure 3.7 the user has the choice of plotting random sample point where several sample points are to be specified or the sampling can be performed using a uniform grid methodology. The number of rows and columns specified to make the uniform grid will determine how many sampling points are produced. For this study 5 000 randomly sampled points were generated. Once a sampling method is chosen the user specifies the extent of the area that is to be sampled by entering both the latitude and longitude minimum and maximum. When the extent of the study area is determined the user runs the tool by selecting the *Get Elevation* tool. The tool's run time is reliant on the Internet, and the amount of sample points that need to be generated the more points that a user requires the longer the run time of the tool. In comparison to that of the TCX Converter, Terrain Zonum may take longer to produce an output as it is simultaneously completing both Google Earth™ and TCX Converter processing steps.

---

[2]   RL: http://www.zonums.com/gmaps/terrain.php

Figure 3.7: Terrain Zonum Solution Interface.

3.3.4 Geographical Information Systems (GIS)

The output file from Terrain Zonum Solution and TCX Converter is uploaded into a GIS program (Figure 3.5 reference 1.3), such as ArcGIS Desktop, ArcGIS Pro or QGIS and displayed as x and y points that have a known z-value. Points extracted from Google Earth™ use the WGS 1984 (EPSG 4326) coordinate system. As such, points imported into a GIS are assigned to the same system once imported. ArcGIS Pro is proprietary software and as such not freely available, although a trial version can be installed for a limited time of 21 days. QGIS in comparison is freely available. To obtain a geospatial dataset using only freely available resources, QGIS is preferred over ArcGIS Pro.

3.3.5 Interpolation to a DEM

GIS provide various means of interpolating spatial location values and a DEM is created forming the basis for further analyses, such as hydrological modelling. For this study two types of interpolation methods were used (Figure 3.5 reference 1.3.1), these being: 1) Inverse Distance Weighted (IDW), and 2) Ordinary Kriging. These interpolation methods assume that the closer the points are to each other the more similarities and correlations exists between these compared to points that are further apart (Setianto & Triandini, 2013). This can be related back to Tobler's first law of geography that argues that the earth's surfaces change relatively slowly over distances, based on the notion that all things are related but the shorter the distance the more they are related (Waters, 2017). Thus, the shorter the distance between two given locations the more similar they become. IDW assigns cells with a weighting based on the distance of an input cell from the output cell, therefore the greater the distance between two cells the less the influence (Childs, 2004).

This interpolation method should be used only if the sample points are dense enough to accurately capture the extent of the local surface variation (Childs, 2004). The accuracy of this method is affected by the power parameter *p* (Burrough & McDonnell, 1998). This combined with the size of the area and the number of sample points determines the accuracy of the final output. Regardless, this method (Equation 3.1), is a simple-to-use interpolation method, and accommodates for users that have little experience with interpolation.

Equation 3.1: Inverse Distance Weighting (IDW) formula (Burrough & McDonnell, 1998).

$$Z_4 = \frac{\sum_{i=1}^{N} Z_i \cdot d_i^{-n}}{\sum_{i=1}^{N} d_i^{-n}}$$

Where:

Z_0 – *estimation value of variable z in point I*
Z_i – *sample value in point I*
d_i – *distance of sample point to estimated point*
N – *coefficient that determines the weight based on a distance*
n – *total number of predictions per validation case*

Kriging is a more powerful statistical interpolation method that assumes that spatial correlation is defined by the direction or distance between two sample points. This interpolation method is recommended to users that have a better understanding of the method as it requires the user to identify a spatially correlated bias in either direction or distance (Childs, 2004). Once this bias is known the user is required to calculate the sill, nugget, and lag of the data, which form part of the input while performing Kriging. The nugget represents the small-scale variability, and the sill is either the total variance contribution or it is the maximum variability between two points. The lag, which is referred to as the range is the distance at which the variogram starts to level off, represents the point where the data are no longer correlated (Cameron & Hunter, 2002).

Ordinary Kriging is the most widely used kriging technique (Wackernagel, 1995). This technique is described as the best linear unbiased estimator as its estimates are linearly weighted with the main aim of having the mean residual error as zero (Mesić, 2016). Ordinary Kriging determines an estimate of the local constant mean for a given area and then proceeds to use the same method of Simple Kriging on the corresponding residuals (Goovaerts, 1997). It assumes there is no trend, within the dataset, that the random field is locally stationary, and that the mean of the dataset is not known but constant. Weights are derived by solving the system of linear equation, minimizing the expected variance of data values. Ordinary Kriging is performed using Equation 3.2.

Equation 3.2: Formulae used to perform Ordinary Kriging (Goovaerts, 1997).

$$\textbf{A: } \hat{z}(x_0) = \sum_{i=1}^{l} \lambda_i(\ ); \quad \textbf{B: } \hat{\gamma}(h) = \frac{1}{2n} \sum_{i=1}^{n} (z(x_i) - \hat{z}(x_i + h))^2$$

Where:

$\hat{z}$ – estimated value of an attribute at a point of interest x_0
z – the observed value at the sampled point x_i
λ_i – the weight assigned to the sample point
n – the number of sampled points
h – the distance between points
(h) – the semi variogram

The Universal Kriging technique predicts z-values for areas that have not been sampled (Mesić, 2016). This technique acknowledges that mean values are not stationary, therefore, a nonstationary regionalized variable is seen to have both a drift and residual component. Universal Kriging assumes that there exists a functional dependence between the mean (x) and the spatial location, which is determined using the Equation 3.3 (Mesić, 2016).

Equation 3.3: Universal Kriging equation (Mesić, 2016).

$$\mu(x) = \sum_{i=1}^{k} a_l f_l(x)$$

Where:

(x) – Drift
k – Number of functioned that are used to model the drift
a_l – lth coefficient that is estimated from the data
f_l – lth function of spatial coordinates used to describe the drift

Kriging, as a more complex interpolation method, the user needs to have a better understanding of the method. Reasons are, as mentioned above, that Kriging requires the user to analyse the input data prior to performing the interpolation as the user is required to identify values such as the sill, nugget, and lag. These values are data dependent and will change between datasets, requiring greater knowledge and skill on the part of the user. Irrespective of this, Kriging is included as an interpolation method to allow for a comparison to outputs derived from the IDW method. Furthermore, for this thesis Ordinary Kriging was used as the data did not present any inherent trends. This was determined once points had been created in Google Earth™ and z-values extracted using the Terrain Zonum Solution, and TCX Converter (Figure 3.5 reference 1.2.1 and 1.2.2). The resolution of interpolation of both methods was matched to that of the control SRTM DEM.

3.3.6 Contour and Coastline Creation

Contour lines and coastlines (Figure 3.5 reference 1.3.2) form the basis of geospatial analyses and interpretations. Contour lines are formed by connecting all points of equal elevation on a given surface (Marsudi, 2017). These isolines indicate the surface of

the earth giving the user an estimation of the terrain for a given area, allowing for the identification of landforms (Kettunen, et al., 2017). The contour tool is used within GIS to create contour lines using a surface raster such as the DEM produced, the elevation raster is the input, and a contour line interval is determined in meters. For this thesis contour intervals of 20 and 50 meters were chosen because they are the most suitable due to the spatial extent of the study area (see Price, 2006). Contour lines are generated within the extent of the elevation raster since areas without data do not contain elevation values to be connected. As such, the output of this tool is a line feature class that represents the elevations within the input raster. It is important to note that the accuracy of the elevation raster will be indicative of the accuracy of the line feature.

The coastline is determined following contour creation and represents the contour of 0 m a.s.l., *i.e.,* the zero-elevation contour line. Once contour lines are created, GIS are used to extract the zero-elevation contour line using a simple attribute selection, selecting the elevation attribute that is equal to 0 m and creating a new layer from the selection. This 0 m contour line is then used as the admin boundary of Rodrigues within this thesis as it is assuming that all features within this boundary line form part of Rodrigues.

3.3.7   Geomorphological Mapping

Geomorphological units form the basis of much subsequent environmental analyses such as the analysis and understanding of erosion patterns within a given area. As such, geomorphological domains are modelled for Rodrigues (Figure 3.5 reference 1.3.3). Saddul (2002) recognises and classifies the geomorphological structure of Rodrigues into 5 major domains, being the (1) central ridge, (2) inland slopes, (3) coastal environment, (4) western lowland and (5) aeolianite plain. The central ridge is on average 300 m above sea level and is seen as the only watershed and catchment on the island. It spans across from Quatre Vents all the way to Grand Montagne. The majority of Rodrigues makes up the second domain of the inland slopes. These slopes lend themselves to the drainage lines, which move radially outwards from the central ridge. The coastal environment is predominately comprised of extensive coral reefs which expand up to 50 m on the east and up to 9 km in the south west. The Rodrigues coral reef is estimated at 200 km² in area thus making up two thirds of the island's total area.

The remaining two domains occur in the west of Rodrigues, this being the western lowlands and the aeolianite plain. The western lowlands ranging from Plaine Mapou to the west of the Baie du Nord-Anse Grande Var line, including the calcarenite lowlands of Plaine Corail. It is believed that regardless of the rolling topography this area illustrates it does not relate to the original volcanic shield, which emerged from the sea.

Rodrigues has extensive aeolianite and calcarenite plains that cover an area of 3 km² in the south west. This is called Plaine Corail and can be seen between Anse Grande Var and Baie Topaze. The Plaine Corail extends from the sea to the 50 m contour with a gentle slope of 8 to 12 degrees. This plain contains multiple potholes and shallow depressions, which at times may hold water forming water courses however these are more likely to remain dry throughout the year. More aeolianites can be found at Pointe Coton, which have been severely undercut and eroded due to wave action. This plain represents a similar topography to that of Plaine Corail. The Pointe Coton aeolianite plain covers a much smaller area of only 450 – 500 m² and is only 5-6 m a.s.l.
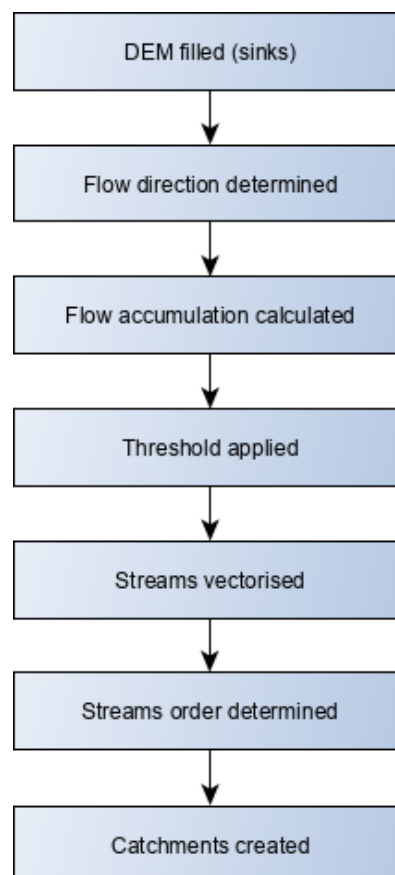
3.3.8 Hydrological Modelling

Interpolation yields a DEM on which multiple hydrological tools within a GIS are performed to produce stream network data for the basemap (Figure 3.5 reference 1.3.4). Hydrological modelling follows a similar process in various GIS products (open source or proprietary), where concepts such as filling discrepancies in the DEM, calculating flow direction and accumulation based on the input DEM, determining streams based on flow accumulation, and the identification of catchments based on the input DEM apply (see Figure 3.8, pg. 23). For example, in ArcGIS Pro hydrological modelling tools are available in the *Hydrology* toolset of the *Spatial Analyst* toolbox. In QGIS, hydrological modelling is available as part of the System for Automated Geoscientific Analyses (SAGA) capabilities (*Terrain Analysis – Hydrology*).

A DEM may contain incorrect sink cells; these are cells that do not have a drainage direction out of them as all surrounding cells are higher than the sink cell. While sinks can occur naturally, such as sinkholes in a karst environment, most sinks are generally incorrect artefacts and need to be filled from a DEM. Sinks are filled using the z-limit, which defines the maximum difference between the sink value and the pouring point. This z- limit determines which sink cells will be filled and which will remain the same (Planchon & Darboux, 2002), to fill erroneous sinks to create a filled DEM, *i.e.,* where the incorrect sinks have been filled to represent the surface of the earth more accurately as modelled by the DEM. A sink is a cell of the DEM that is found to have no surrounding cells with lower elevation values, therefore causing a cavity within the DEM, which in turn will affect the hydrological modelling of that given area.

Once sinks have been filled, flow direction, i.e., the direction of flow from one cell in the raster or input surface based on the steepest descent to surrounding cells within its neighbourhood, is determined. Various algorithms for modelling flow direction exist, such as Multi Flow Direction (MFD), D-Infinity (DINF) or D8. Multi Flow Direction (MFD) partitions the flow from a given cell to all downslope neighbours (Jenson & Domingue, 1988). The partition exponent determines the fraction of flow that is draining to all downslope neighbours and is determined by the local terrain of an area. The MFD modelling algorithm is more complex as there is potential for multiple values to tie into each cell. The DINF flow method makes use of eight triangular facets to determine flow direction to the steepest downslope neighbour. This method produces a floating-point raster, which is represented by a single angle that rotates counterclockwise. The third flow direction method and that employed for this thesis is the eight-direction method (D8). This method attempts to model the flow direction from each given cell to that of its steepest downslope neighbour based on the 8 surrounding cells (Jenson & Domingue, 1988). The output raster produced after filling DEM sinks becomes the input to the flow direction calculation. This provides an output integer raster with each sink assigned a unique value. These unique values have a range that is between one and the total number of sinks. For example, if there are 100 sinks in total these values will range from 1-100.

Flow accumulation is then computed based on the flow direction raster. Based on each cell's flow direction an accumulated flow is calculated according to the accumulated weight of the cells that are flowing into each downslope neighbouring cell (Jenson & Domingue, 1988). Stream channels are identified based on areas of high flow accumulation. Based on the calculated flow accumulation output the user visually sets an accumulation threshold, which is based on the scale / detail of the basemap produced. For this study and following trial and error a threshold of 150 was used as it provided adequate detail for the scale of the

basemap. Stream order, specifically Strahler order, is then determined for the calculated rivers by inputting the rivers created into the Stream Order tool within ArcGIS Desktop. Stream order is an effective way of classifying waterways and integral to understanding and managing the many differences between streams of different sizes. The order of a stream, for example, can provide insight into the amount of sediment potentially transported in an area. In an area with high erosion rates, such as Rodrigues, this is an important consideration since higher water flows, if sediment is available, would transport more sediment than low water flows. Furthermore, knowledge on stream order can assist for designing water quality monitoring and management plans, and for using waterways as natural resources in a sustainable manner. Once Strahler Order is determined, the final output is a river network and catchments dataset that is then integrated into the basic basemap (see Objective 3, pg. 2).

Figure 3.8: Steps used for hydrological modelling in a GIS.

3.3.8.1 Hydrological modelling in ArcGIS Pro

To illustrate the steps used in hydrological modelling, steps executed in ArcGIS Pro 2.8 are detailed in Figure 3.9.
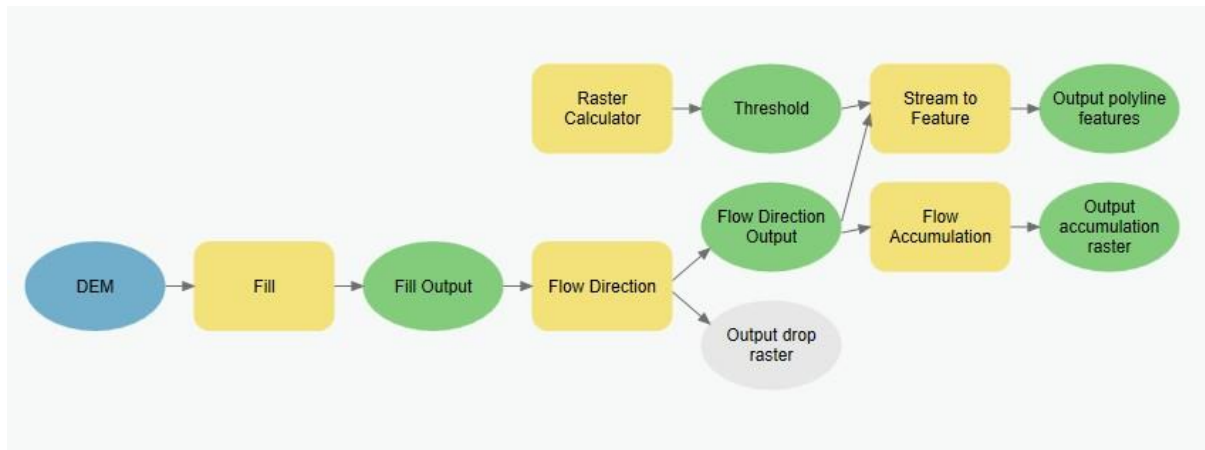
Figure 3.9: Hydrology Workflow Model as built using Model Builder of ArcGIS Pro 2.x.

The *Fill* tool helps eliminate any major discrepancies within the DEM, which may influence the calculated flow direction. Such discrepancies are when there are no cells around a given cell that is lower in elevation. The tool utilizes the z-limit to fill sinks. Once erroneous sinks have been filled *Flow Direction* is run utilising the D8 algorithm. Each grid cell is allocated a flow direction code value when running the Flow Direction tool. These values and directions are East is 1, South-East is 2, South is 4, South-West is 8, West is 16, North-West is 32, North is 64, North-East is 128. Once flow direction is calculated the output is used as the input in calculating flow accumulation, achieved using the *Flow Accumulation* tool. Using this threshold, the *Con* and *Stream to Feature* tools are used to extract areas identified as streams. Strahler Order is then extracted using the *Stream Order* tool, set to calculate Strahler Order using the streams and flow direction raster of the previous steps as inputs. As a last step, catchments are extracted using the *Watershed* tool.

3.3.8.2 Hydrological modelling in QGIS

Like to the section above, steps as executed in QGIS to perform hydrological modelling (Figure 3.10).
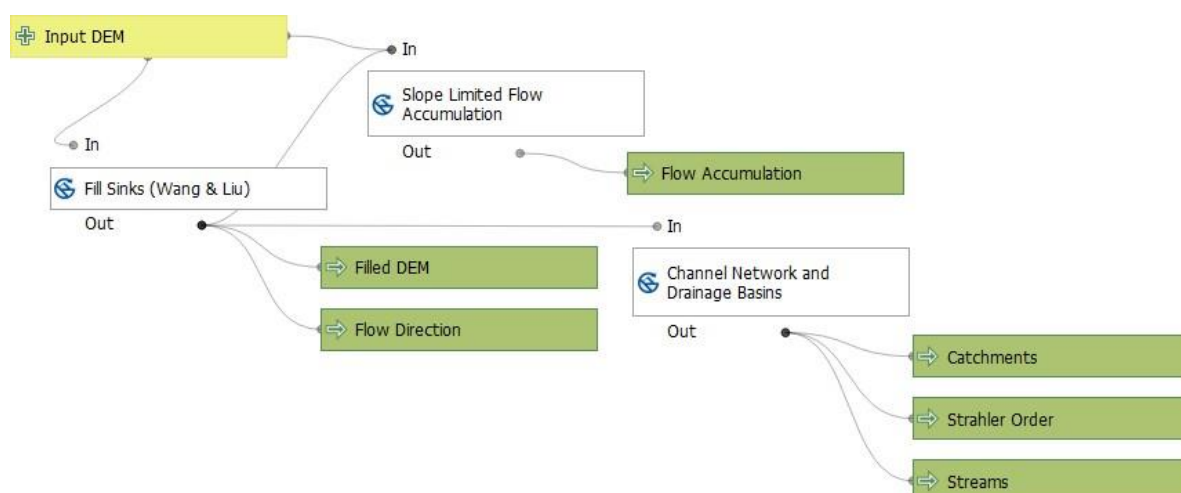
Figure 3.10: Hydrology Workflow Model as built using the Graphic Modeler of QGIS 3.x.

Erroneous sinks are filled, flow direction is calculated, followed by flow accumulation, a drainage line extraction by applying a suitable threshold, drainage line vectorisation, and catchment delineation. Various hydrological modelling tools are available in the SAGA integration to QGIS to fill sinks within a DEM. The Fill Sinks (Wang Liu) module will fill

erroneous sinks and produce a filled DEM, and a flow direction and catchment raster. As such, this module does not require an additional step for flow direction calculations. The Fill Sinks (Wang, Liu 2006) module applies the algorithm of Wang & Liu (2006) to identify and fill sinks. Furthermore, the module creates a hydrologic appropriate DEM, ensuring the preservation of downward slope, which can be user-defined. For this thesis this module is preferred over Fill Sinks XXL (Wang, Liu 2006), since the latter is designed for large datasets, which is not the case for Rodrigues.

Flow accumulation is determined using the *Slope Limited Flow Accumulation* module, based on the work of Freeman (1991). The user can specify a slope threshold at which level all flow is routed through a cell downhill. Streams, Strahler order of streams, and catchments are then extracted using the *Channel Network and Drainage Basins* module, found in the *Terrain Analysis – Channels* menu. Like channel extraction as described in the previous section, a threshold must be specified. This threshold is determined through visual analysis based on the level of detail required for the basemap produced within this thesis Further inputs are the depression free DEM. The threshold is dependent on each individual dataset and needs to be determined by investigating the dataset in question. As before, the threshold applied was 150. No vectorisation of the extracted drainage system is required since a vectorised drainage system is part of the outputs of the *Channel Network* module.

3.3.9   Digitisation

Additional features included in the basemap, such as roads, dams, and towns, are manually digitized (Figure 3.5 reference <u>1.3.6-1.3.8</u>). Using Google Earth™ imagery and visual analysis roads, dams and towns are digitized at a relatively large scale (~1:2000) ensuring all features are clearly visualised. This process means using the *Draw* tool in Google Earth™ to manually draw lines (roads), points (towns) or polygons (dams) to illustrate the features in question. The digitised roads were classed into three categories being 1) main road, 2) secondary road, and 3) roads. Towns were digitised using a vector point feature at the visual centre of each town.

All outputs are saved and exported as .kml files and then imported into GIS by converting. .kml to shapefile, allowing these data to be integrated with the other outputs produced such as the DEMs, rivers and (EPSG 4326) coordinate system. Google Earth™ offers user freely available high-resolution imagery that is cloud free with high temporal resolution with the option to view past imagery via the time scale bar. It offers a user interface (UI) and user experience (UX) interface that is ideal for users that have minimal or do not possess mapping or digitisation skills. Due to these considerations this thesis has, therefore, made use of Google Earth™ for digitisation of its features as opposed to other digitisation platforms.

Digitisation can be subject to errors and these potential errors should be taken into consideration when digitising. A large error to consider is that of missing information resulting in the digitised feature, such as roads, missing road links. To eliminate this potential error, it is important to use the most recent high-resolution imagery with minimal cloud cover made available for a particular area. Outdated imagery may not include newly constructed features; therefore, those features will not be captured. Extensive cloud cover or low-resolution imagery will lead to difficulties visualising these features and, therefore, lead to the features not being captured. Understanding a particular study area will determine whether digitisation is possible within a particular area, a study area that is heavily vegetated and has a high canopy cover will make the visualisation of features for digitisation difficult as opposed

to an area that is characterised by bare ground or lacking vegetation.

3.3.10 Additional Geospatial Layers

Since the project focuses on creating usable geospatial layers of areas where data are not available, inaccessible, or difficult to compile, additional geospatial layers not normally associated with a basemap were created (Figure 3.5 reference 1.3.5). These are a slope surface, an aspect surface, and a hillshade. Slope refers to the change in elevation over the change in horizontal position, whereas aspect is the horizontal direction of the slope of topographic feature, i.e., the direction of maximum gradient of the surface at a particular point. Both slope and aspect are focal functions that are based on the neighbourhood principle and based on first derivatives in terrain analysis (Malaperdas, Panagiotidis 2018). Both slope and aspect are expressed in degrees (for the purpose of this project slope was not calculated as a percentage surface). As such, the slope surface reaches a maximum possible value of 90°, and the aspect surface is measured in degrees in clockwise from north. A hillshade is a shaded relief (3D) based on a DEM based on brightness of terrain reflections given surface and sun location. The illumination source is generally defined at an angle of 45° from the north-west – this position provides the optimal impression of relief in the third dimension. All three outputs are derived from a DEM, as that created in section 3.3.5 Interpolation to a DEM.

3.4.   Database Management

All created data are imported into an integrated database, depending on the GIS product (*for example,* personal file geodatabase for ArcGIS Pro). This ensures a centralised and accessible database for further analyses. Furthermore, all datasets have completed metadata that states the methods used to obtain a given dataset. As such, individual data layers are populated with minimum metadata according to International Organization for Standardization (ISO) 19115: *Geographic Information – Metadata*. This standard aims at providing a clear procedure that is followed to describe a digital geographic dataset. This allows for all users of the dataset to understand what the dataset entails and how it was captured, thus providing the user information as to whether this dataset is relevant and suitable to their study. Finally, data are stored and made accessible via Mendeley Data[3], part of Digital Commons Data, which is a cloud-supported open access data repository that provides a Digital Object Identifier (DOI) for each registered dataset. Data uploaded here becomes available to any researcher or member of the public under the Digital Commons Data license. Data are, furthermore, based on FAIR Data Principles, those of Findable, Accessible, Interoperable, and Reusable (Wilkinson, et al., 2016).

3.5.   Validation

Since many of the basemap features rely on the accuracy of the DEM this study is focused on testing the accuracy of the DEMs produced using the methods described from section 3.3.1 through to 3.3.5 (pgs.22-25). To test this accuracy a hundred random points are generated within the boundaries of Rodrigues. These points are then used to extract z-values from the different DEMs for comparison purposes. Both interpolation methods used (see Interpolation, pg. 18 onward) are statistically analysed to compare the z-values derived from the original Rodrigues control SRTM DEM to those of the interpolated DEMs.

---

4   URL: https://data.mendeley.com/

Several statistical methods are suitable when validating data, in particular methods based on statistical inference. Statistical inference makes use of hypothesis testing, where the null hypothesis (H_0), the alternate hypothesis (H_a), the level of significance (α) and the confidence interval are determined (Briggs, 1977; Till, 1985). For this project a level of significance of 0.05 is used (α = 0.05), since this level generally reflects a realistic level of certainty in physical geography (Briggs, 1977). This level further minimises both Type I (rejecting a true H_0) and II errors (accepting a false H_0) (Till, 1985). Statistical inference used for the project comprises Pearson's product-moment coefficient of linear correlation I (pg. 27), the *F-test* (pg. 27), Student's *t-test* (pg. 28), and determining the Root Mean Square Error (RMSE) (pg. 28).

3.5.1 Pearson's product-moment coefficient of linear correlation I

Correlation analysis is a way of measuring the strength of a linear relationship between two variables, namely x and y. The reliability of y strongly depends on the relationship it has with x, meaning the stronger the relationship the more reliable and accurate the estimate of y is (Wegner, 2016). The strength of using Pearson's product-moment coefficient of correlation lies in that the correlation coefficient I is independent of the units of measurement for the separate populations (Williams, et al., 2006). The null hypothesis assumes that the slope of the line is 0, therefore, there is no relationship, and the populations are not the same. If the slope is not 0 the null hypothesis is rejected, and it is assumed the populations are the same with a slope of -1 or 1. Correlation tests on both Kriging and IDW were completed in relation to the control DEM using Pearson's product-moment correlation coefficient. The correlation coefficient is determined using Equation 3.4.

Equation 3.4: Pearson's Correlation Coefficient formula (Wegner, 2016):

$$r = \frac{n \sum xy - \sum x \sum y}{\sqrt{[n \sum x^2 - (\sum x)^2] \times [n \sum y^2 - (\sum y)^2]}}$$

Where:

r – *Sample correlation coefficient*
x – *Values of the independent variable*
y – *Values of the dependent variable*
n – *Number of paired data points in the sample*

3.5.2 *F*-test

Once the correlation coefficient is determined, the *F*-test is carried out to determine whether the two populations maintain the same variance (Till, 1985; Williams, et al., 2006). This test assumes a normal distribution and is run prior to the parametric *t*-test (Doornkamp & King, 1971). Once similarity of variance has been established, student's *t*-test (see next section) is used to determine whether two populations are statistically similar when population variance is not known (Doornkamp & King, 1971; Till, 1985). The null hypothesis for this test states that population variances are similar. In contrast, the alternative hypothesis states that population variances are dissimilar.

3.5.3 Student's *t*-test

Similarity of populations is tested using a homoscedastic two-sample Student's *t*-test for the equality of means (n is variable; variance is similar) for two independent (unpaired) samples. This test assumes a normal distribution but may be applied for non-normal distributions (Till, 1985). The student's *t*-test is a widely used statistical method to compare the means of two

variables (Al-Achi, 2019). The *t*-test replaces the *Z*-test also known as the Standard Normal Distribution test, when the standard deviation of the population (σ) of the data is unknown (Al-Achi, 2019). The *t*-test determines how significant the difference is between the two data samples, here using a 95% confidence level. The null hypothesis for this test is that the means of the populations are the same between the two sample datasets, thus being proven with a *p*-value that is not less than 0.05. If the *p*-value is less than 0.05 the null hypothesis is rejected, and it is assumed that the population data are not the same. For the relevance of this thesis, it is crucial to determine similarity of interpolated DEMs to the control DEM, to ensure quality and accuracy of data. Null and Alternative hypotheses for the various statistical tests employed are given in Table 3.4 below.

Table 3.4: Hypothesis for statistical tests used.

Statistical Test	Null Hypothesis (H₀)	Alternative Hypothesis (Hₐ)
Correlation coefficient	The slope of the line between the two series is 0, *i.e.,* there is no relationship.	The slope of the line between the two series is closer to either 1 or -1, *i.e.,* a relationship exists.
***F*-test**	The interpolated and control DEM have the same variance, *i.e.,* $\sigma_1 = \sigma_2$.	The interpolated and control DEM variances are not the same, *i.e.,* $\sigma_1 \neq \sigma_2$.
Students *t*-test	The interpolated and control DEM 35eciding35n35 are comparable to each other, *i.e.,* $\mu_1 = \mu_2$.	The interpolated and control DEM measurements are drawn from different 35eciding35n and not comparable to each other, *i.e.,* $\mu_1 \neq \mu_2$.

3.5.4   Root Mean Square Error (RMSE)

To cross validate the accuracy of the DEMs the RMSE is used. The RMSE is an accuracy measure for maps and DEM's (Varga & Bašić, 2015), through the calculation of residuals (the difference between the values from the DEM elevations to that of the control DEM elevation) (Congalton & Green, 2009). These residuals are aggregated into a single measure of power using the Equation 3.5.

Equation 3.5: Root Mean Square Error formula (Congalton & Green, 2009).

$$RMSE = \pm\sqrt{\frac{1}{n}\sum_{i=1}^{n} e^2}, \text{ where } e = v -$$

Where:

$RMSE$ – *Root Mean Square Error*
v_{ri} – *reference elevation at the point i*
v_{mi} – *DEM elevation at the point i*
n – *the number of ground check points.*

The RMSE was calculated in Rstudio (RstudioTeam, 2020). The script to determine the RMSE in Rstudio is given on the next page.

```
Data1 <- read.csv(file.choose(),header = T,sep=",")

IDWstats <- lm(Data1$IDW.Elevation ~ Data1$Rodrigues.Elevation)
summary(IDWstats)
plot(IDWstats)

krigstats <- lm(Data1$Kriging.Elevation ~ Data1$Rodrigues.Elevation)
summary(krigstats)
plot(krigstat)
```

Scatter plots of these points are compiled to produce a trend line that illustrates the R^2 value. This value illustrates the coefficient of determination and is the statistical measure that indicates how close the data are to the fitted regression line: the closer this value is to 100% the more the model explains all the variability of the response data around its mean. To determine the RMSE, data derived are scripted through Rstudio to confirm results, and to determine the residual standard error between datasets. The above process determines whether the DEMs produced are of an acceptable accuracy to be used as a bases for other studies or analyses.

3.6. Base Map Compilation

Saddul (2002) basic classification of landform domains of Rodrigues is used as a basis for object-based analysis of the derived DEMs, as described in Geomorphological Mapping (pg. 21). Datasets created in Contour and Coastline Creation (pg. 20), Hydrological Modelling (pg. 21), and Digitisation (pg. 25) are compiled into a map to produce a suitable basemap (refer to Objective 3, pg. 2). Furthermore, data provided by the University of Mauritius, representing the available geospatial data for the island, is used to create a comparison basemap.

3.7. Ground Truthing

The two completed basemaps (based on created geospatial data and based on data obtained from the University of Mauritius) are used to identify discrepancies between features, and areas of similarity. These identified sites are used as zones where the ground truthing is performed in the field. During a field visit to Rodrigues a handheld GPS (Garmin GPSMAP® 64S) was used to log the coordinates of sampled sites. These sites were systematically selected with three basic concepts: 1) areas where remotely sensed techniques lack data that are found in the Rodrigues geospatial dataset, for example if a dam is not identified yet is featured in the surveyed data; 2) areas where remotely sensed techniques have identified features absent in the Rodrigues dataset, for example stream channels; and 3) areas where both remotely sensed techniques and the Rodrigues dataset have confirmed features. This process allows for validation of the results from the remote sensing data used to produce the basemap. Images were taken at these locations as evidence of research findings. Features such as rivers, dams and roads were visually confirmed during the ground truthing process.

Through the successful completion of the above-mentioned methodology numerous basemap features where produced including features such as DEMs, rivers, contour lines, roads, and dams. The methodology outputs are further reviewed and discussed within the Results and Discussion chapter.

CHAPTER 4: RESULTS AND DISCUSSION

This chapter describes and discusses the results obtained based on the methodology followed in CHAPTER 3: METHODOLOGY. Results are presented under the same subheadings as those of CHAPTER 3: METHODOLOGY.

4.1 Data Requirements

Using the methodology in the Data Requirements section (pg. 15 onward), basemap features were produced. These are presented and discussed in the sections below.
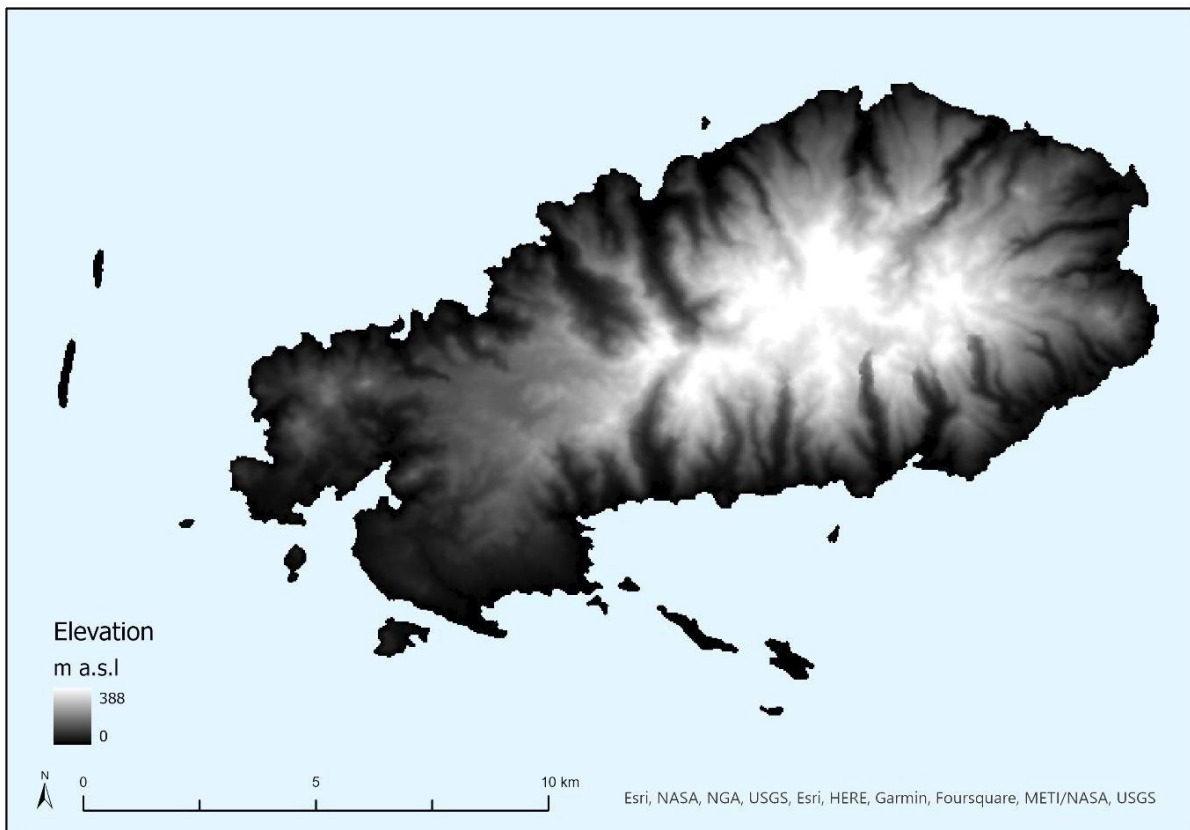
4.1.1 Interpolation

Two types of interpolation were used, namely 1) Inverse Distance Weighted (IDW), and 2) Ordinary Kriging. The results from both methods yield a raster dataset that is like that of a Digital Elevation Model (DEM). Bearing in mind that the aim of the research is to create a digital dataset of an area using easy to follow methods, in addition to free resources, model inputs and thresholds were attempted to be kept to default values. This was done to analyse the accuracy of the output if a user was to only use the default values calculated by the software. To complete Ordinary Kriging, the user is required to specify various values including the nugget, the patrial sill, the lag, and the major range. These values can be either calculated using the Geostatistical Wizard within ArcGIS, which allows for adjustment of values based on minimising the RMSE, and then inputted within the interpolation tool or users can run the kriging interpolation tool without specifying these values to which the software will calculate default values. For this thesis and to ensure the accuracy of the output the Geostatistical Wizard was utilized. Ordinary Kriging was run using the following values: (1) Nugget of 129.6209 (2) Partial Sill of 11463.21 (3) Lag of 0.004888, (4) Major Range of 0.043914.

IDW, not being a geostatistical method, is a simpler interpolation method for users that do not understand statistics and, therefore, requires minimal user inputs or data manipulation. To calculate IDW a power value and radius type is required. For the data within this thesis the default value for the power value is 2 and the radius type is set at variable. Upon further evaluation of the data using the Geostatistical Wizard it was found that these input values were suitable for the data inputs. An evaluation of their suitability, i.e., the method, is given in the 4.2 Validation (pg. 27). Figure 4.11 and Figure 4.12, are the output raster's derived from both methods (Kriging and IDW), respectively.
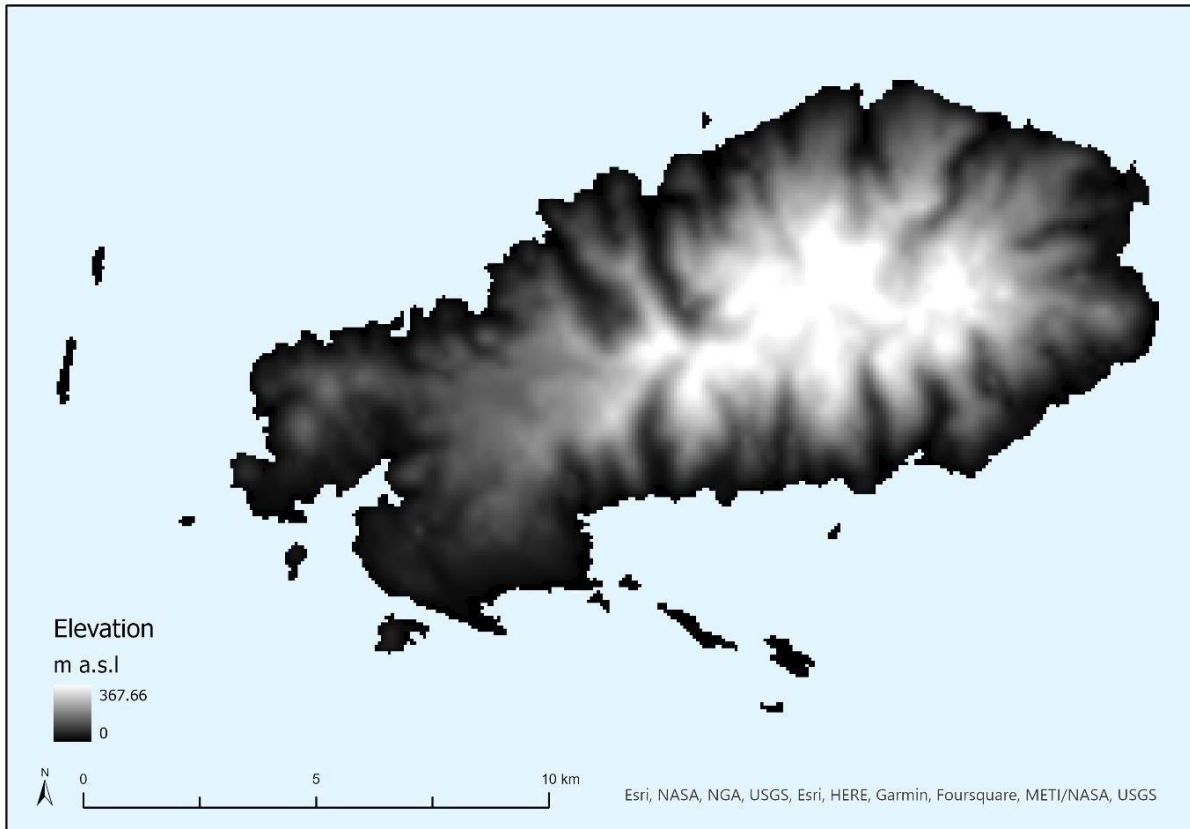
Figure 4.11: Kriging Digital Elevation Model.

*Figure 4.12: Inverse Distance Weighting Digital Elevation Model.*

The output DEMs from the two interpolation methods can be seen in Figure 4.11 and Figure 4.12. Both methods yield similar output when assessed visually. The Kriging DEM is somewhat smoother than that of the IDW DEM, which can be seen to be more defined near the central ridge region of Rodrigues. The IDW interpolation method resulted in a maximum elevation of 388 m above sea level, whereas the Kriging interpolation method resulted in a maximum elevation of 367.66 m a.s.l. This results in a total difference of 20.34 m between the 2 highest points within the interpolation methods. The difference in interpolated values relates to the way the interpolation methods work. IDW interpolates values within the bounds of the input data, i.e., within the minimum and maximum of actual values recorded. Ordinary Kriging interpolation estimates the value of a variable at an unsampled location by computing a weighted average of the known sample points in its neighbourhood and can interpolate values outside the bounds of the input data, i.e., below the minimum value and above the maximum value. A robust method, in the presence of no trend (as applicable here), accounts for autocorrelation. As such, the two methods yield different outputs. However, the interpolated surfaces for both are similar, yielding similar results, illustrating the usefulness of the input data points used for interpolation. When directly compared to the SRTM DEM, seen in Figure 2.1, which illustrates a maximum elevation of 380 m above sea level, the DEM produced using the IDW interpolation method can be classed as more accurate. This is, however, based on a visual assessment and is further validated quantitatively in 4.2 Validation (pg. 27).

4.1.2   Contour Lines and Coastline

The interpolated surface raster derived using IDW and Kriging were used to calculate contours at 20 m and 50 m intervals. The 20 m contours are illustrated in Figure 4.13; the 50 m contour in Figure 4.14 (both on pg. 34). For illustrations purposes the 20 m contour lines was derived using the IDW DEM whereas the 50 m contour lines were derived from the Ordinary Kriging DEM.

The 20 m contour lines derived from the IDW interpolation raster has a maximum contour height of 380 m above sea level, whereas the 50 m contour lines has a maximum elevation of 350 m above sea level. The 30 m discrepancy experienced between the 20 m and the 50 m contour line maximum elevation is due to the contour line interval. If a lower contour interval was used the maximum elevation would closer resemble the maximum heights seen in the Digital Elevation Models discussed in 3.1.1 Interpolation (pg. 17).
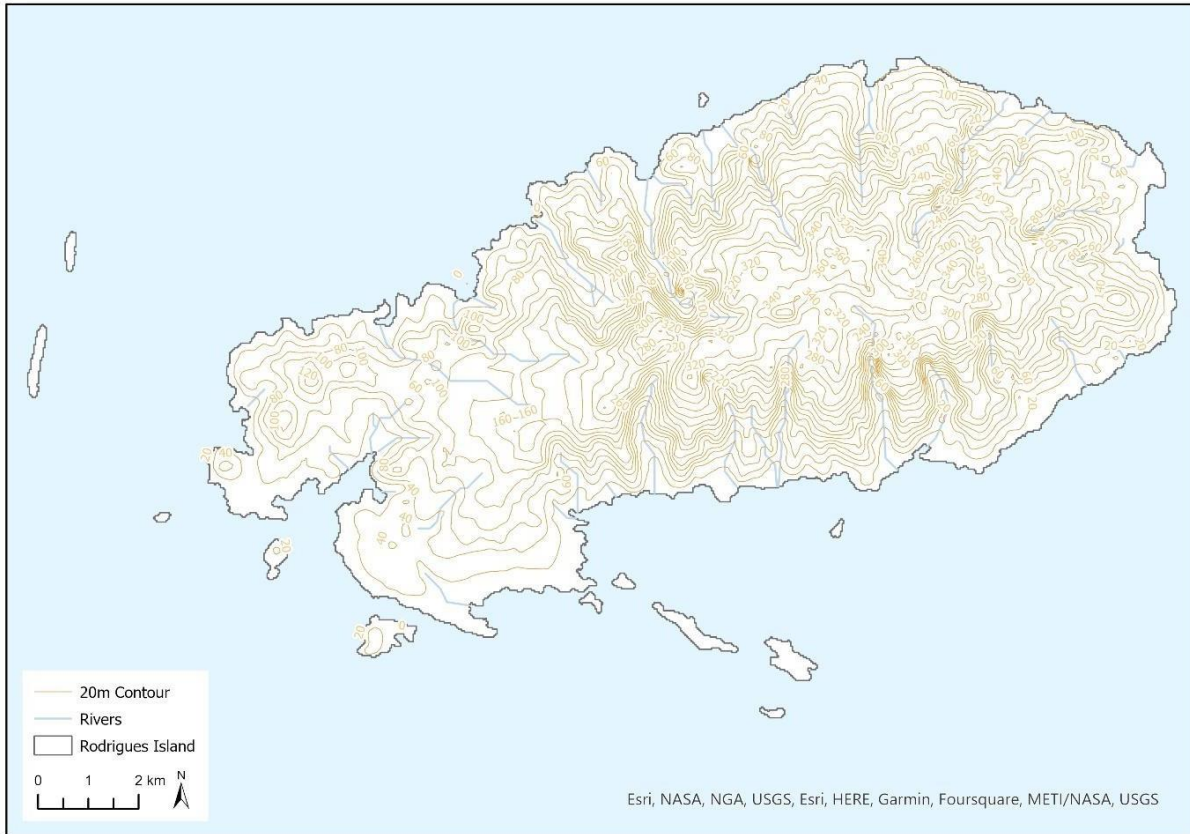
Figure 4.13: Twenty metre contours for Rodrigues.
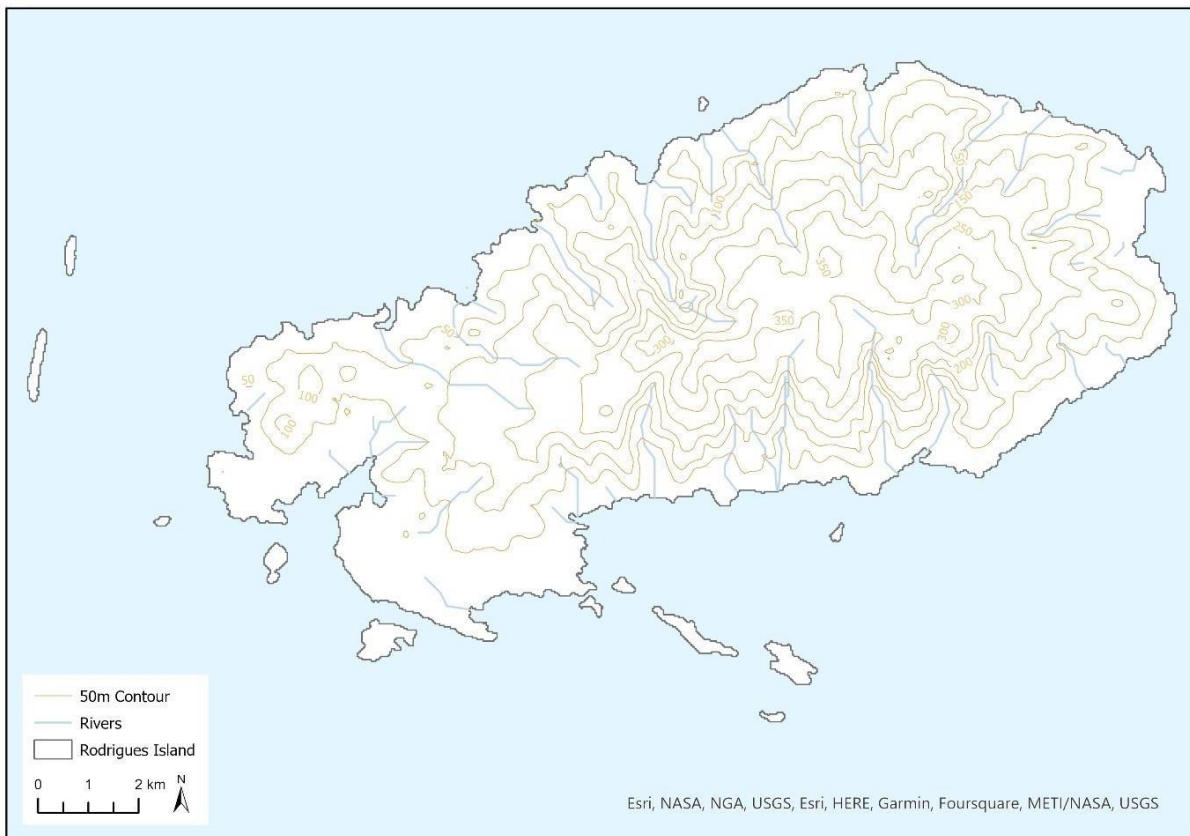
Figure 4.14: Fifty metre contours for Rodrigues.

Numerous factors contribute to the contour interval chosen for a map. These factors include the nature of the topography of the study area 2) the scale of the map, and 3) the extent of the study area. Contour lines at 20 m and 50 m intervals are produced here in accordance these three factors. Rodrigues has a very wide range of topographies as determined by Saddul (2002). These range from relatively flat coastal areas near the western side of Rodrigues to a mountainous region surrounding the central ridge. Rodrigues covers an area of approximately 6.5 km by 18 km (104 km2), which means that the extent of the study area is relatively large compared to the scale at Ih the paper-based basemap is produced. These factors are part of the main influences as to why this thesis used intervals of 20 m and 50 m. Furthermore, if a contour interval lower than 20 m was used the basemap would be cluttered and not visually appealing, thus taking away from the visualisation criteria of a basemap. As to not clutter the final basemap the 50 m interval contour lines were used. This allowed for other features on the basemap to be see while still illustrating the elevation ranges of Rodrigues.

4.1.3 Geomorphological Mapping

A map of the geomorphological zones was created based on the textural description given by Saddul (2002). These geomorphic zones (Figure 4.15), together with hydrology and topography, allow for insight into the landscape dynamics of an area. In the case of Rodrigues, knowing where geomorphological zones lie, in conjunction with rivers and manmade features, provides important information into how erosion can be managed and controlled.
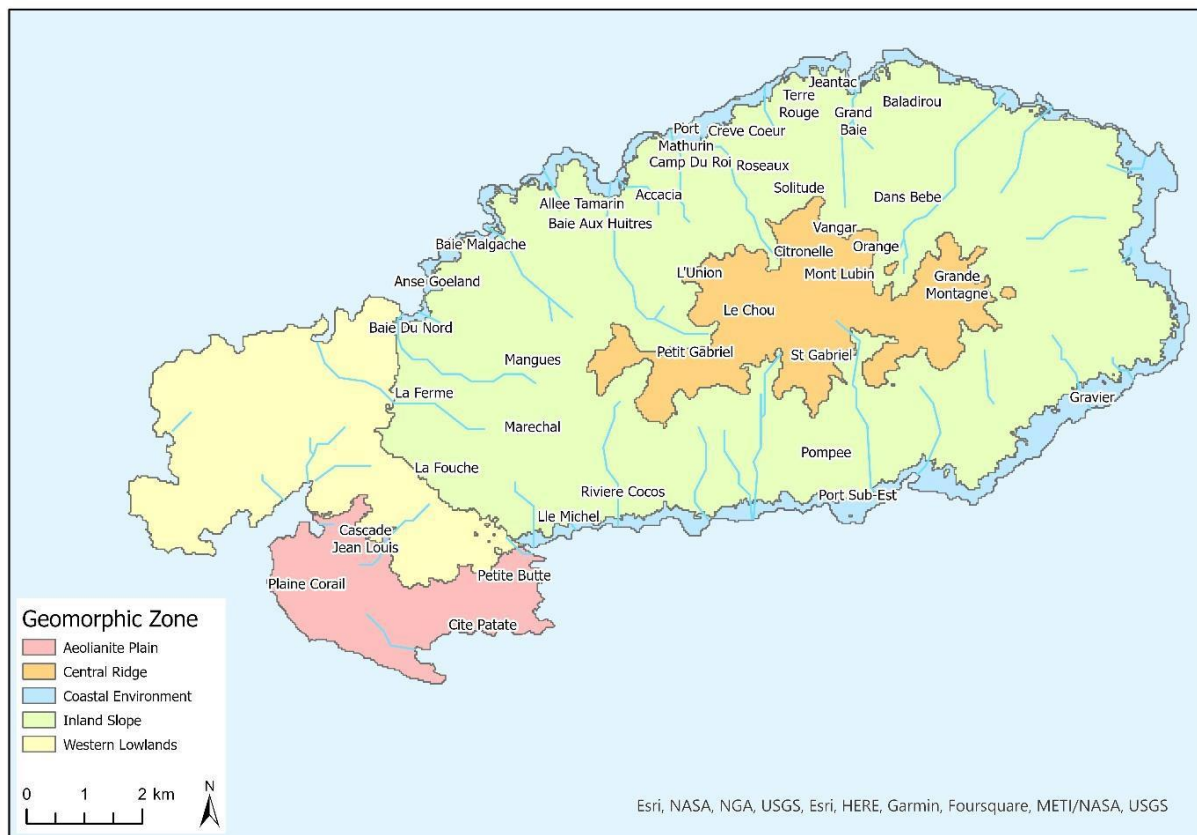
Figure 4.15: Geomorphic Zones of Rodrigues (Saddul, 2002)

4.1.4 Hydrologic Modelling

The outputs of the hydrologic modelling (flow direction, flow accumulation, and major streams once extracted), are provided in Figure 4.16, Figure 4.17 and Figure 4.18.
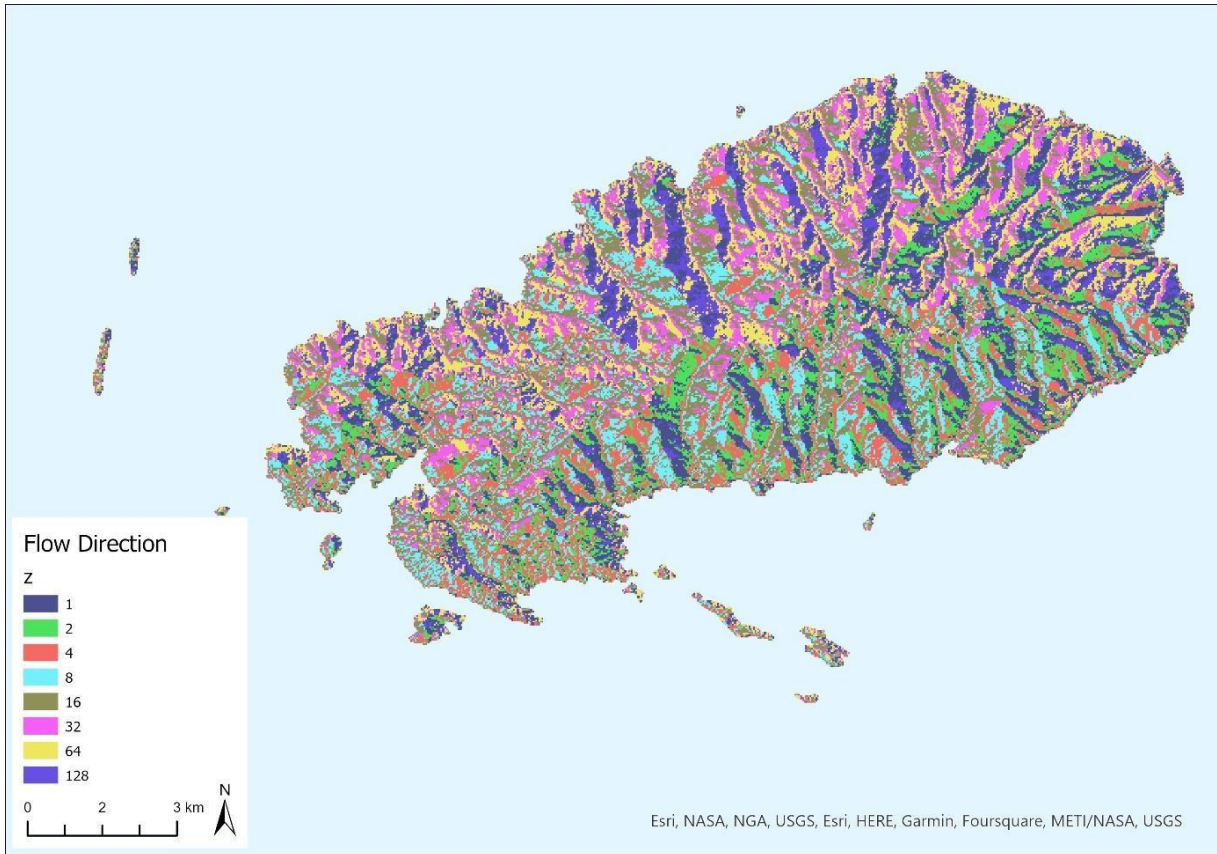
Figure 4.16: Flow direction.
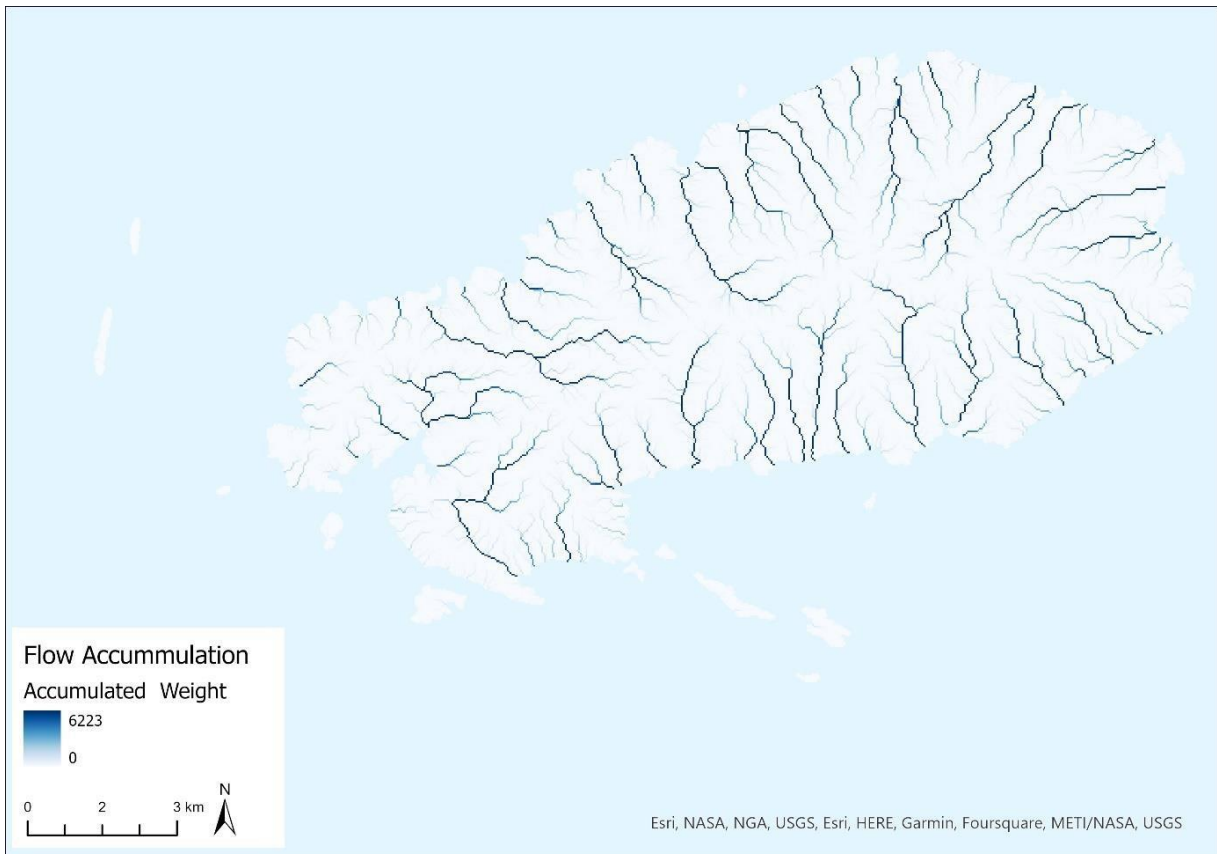
Figure 4.17: Flow accumulation.
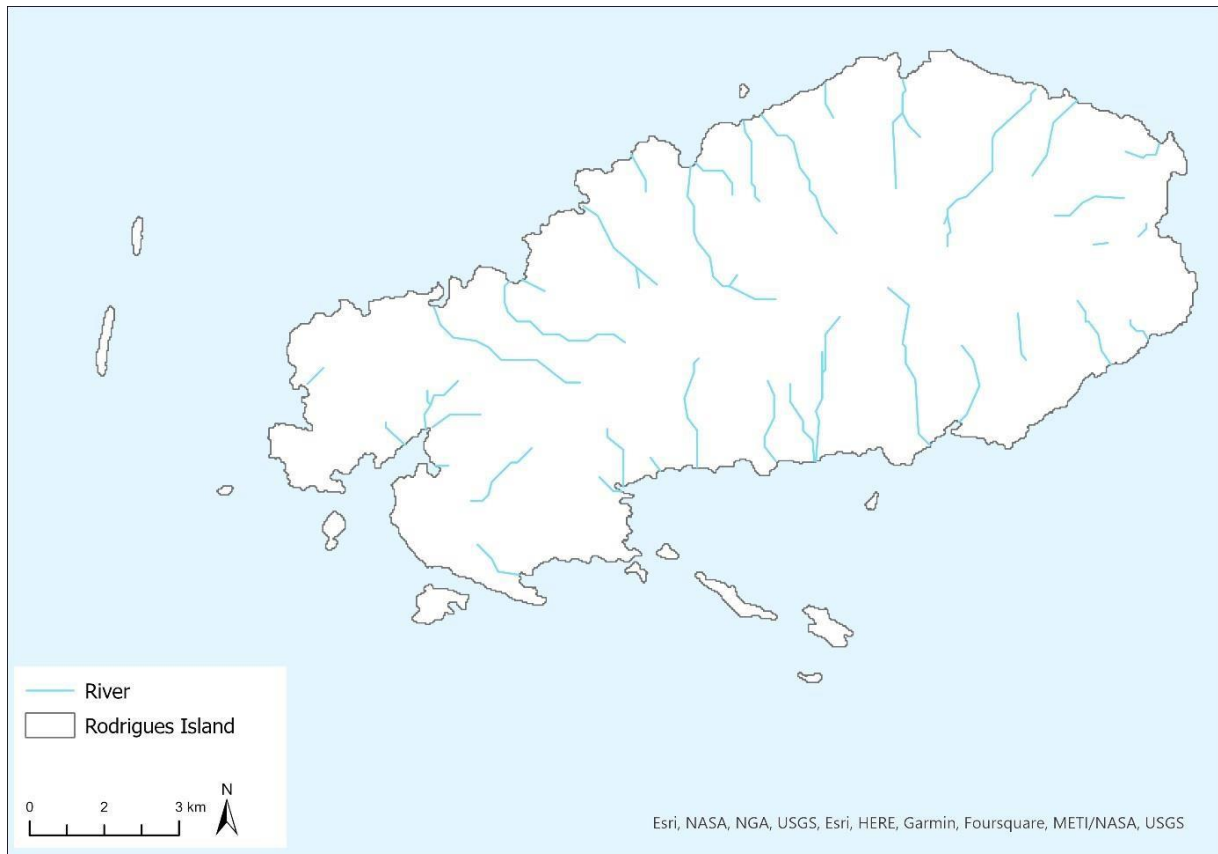
Figure 4.18: Rivers from 150 m threshold.

Rivers follow a radial pattern, flowing from the higher inland areas toward the base level of the sea. This thesis made use of a threshold of 150 when determining the rivers on Rodrigues, this was determined visually. A visual analysis was done when calculating the threshold amount based off the level of detail the produced base map requires. Fewer rivers are evident for the western portion of the island. This area is less varied in topography with wider coastal plains.

4.1.5   Digitisation

The digitisation of these features was manually completed through visualisation of the feature on the Google Earth™ imagery. Figure 4.19 illustrates the digitised roads for Rodrigues symbolised based on the road classifications. Figure 4.20 offers a close-up view of a digitised roads feature in relation to imagery of Rodrigues. As can be seen on the image the digitised roads align with the roads visualised within the imagery back drop, illustrating the benefit of digitising features at a large scale, reducing accuracy error.
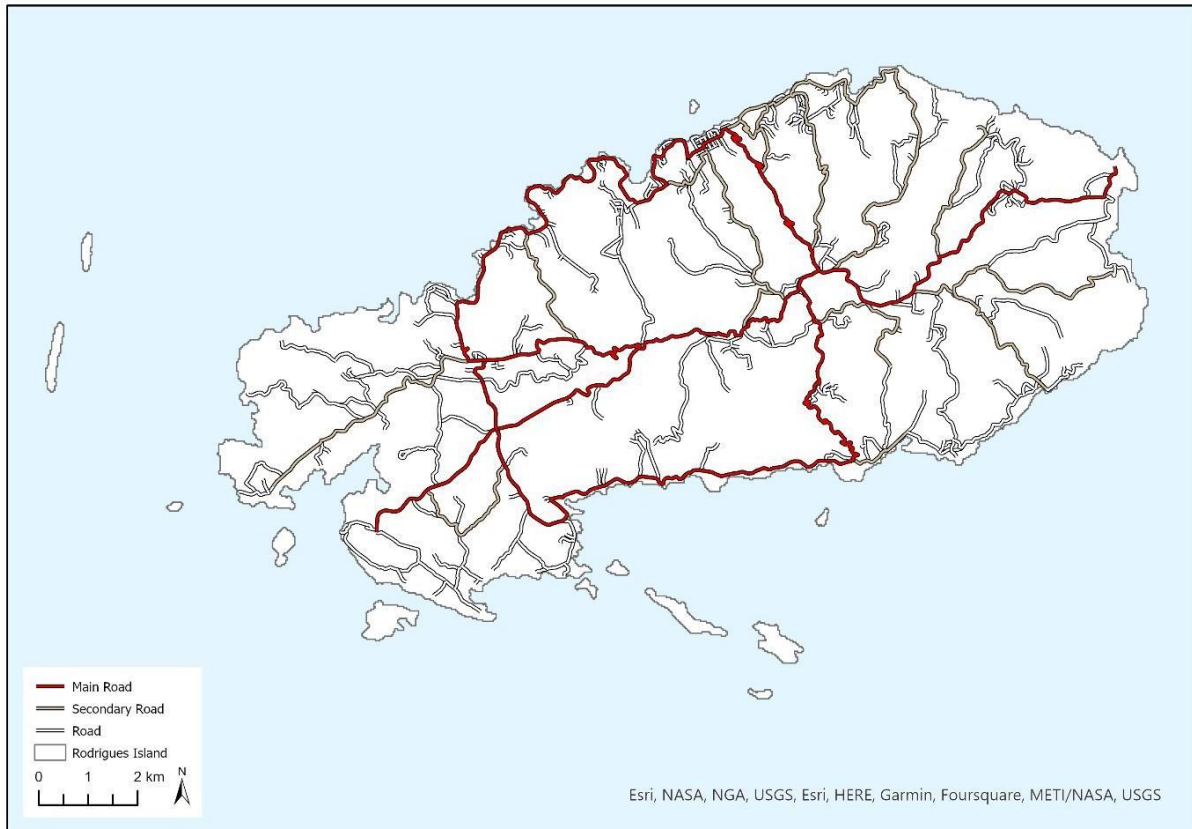
Figure 4.19: Rodrigues digitised roads.

Figure 4.20: Digitised roads with imagery backdrop.

Figure 4.21: Digitised towns with imagery backdrop.

The same process was used when digitising the towns and dams on Rodrigues. Google Earth™ includes place names as a feature one can visualise on the imagery back drop this was used as the bases to identify the towns of which a point feature was then digitised. This thesis further analysed the imagery to confirm all towns were captured ensuring the positional accuracy of the towns feature class. Figure 4.21 illustrates some of the towns captured on Rodrigues.

To digitise the dams, one had to manually scan through every section of the Google Earth™ imagery as these are not features that are labelled and displayed within the software. A polygon feature class of all the dams that were found through the imagery analysis was created. A key criterion for the classification of a visualised water body was that the water body conformed to the above produced river feature classed which were produced following the Hydrological Modelling methodology. If the water body had an associated river water source, it was assumed to be a dam. Figure 4.22 illustrates a dam that was digitised through the visualisation of a water body on the Google Earth™ imagery. A river water source feeding into a dam and later exiting the dam downstream of the dam wall can be seen in Figure 4.22.

Figure 4.22: Rodrigues digitised dams.

4.1.6   Additional Geospatial Layers

Through geospatial tools available in ArcGIS analytics and processing can be done using an input DEM such as slope, aspect and hillshade. These are not requirements for a basemap, however, can be useful to users. These three raster-based images were produced as seen in Figure 4.23, Figure 4.24 and Figure 4.25.
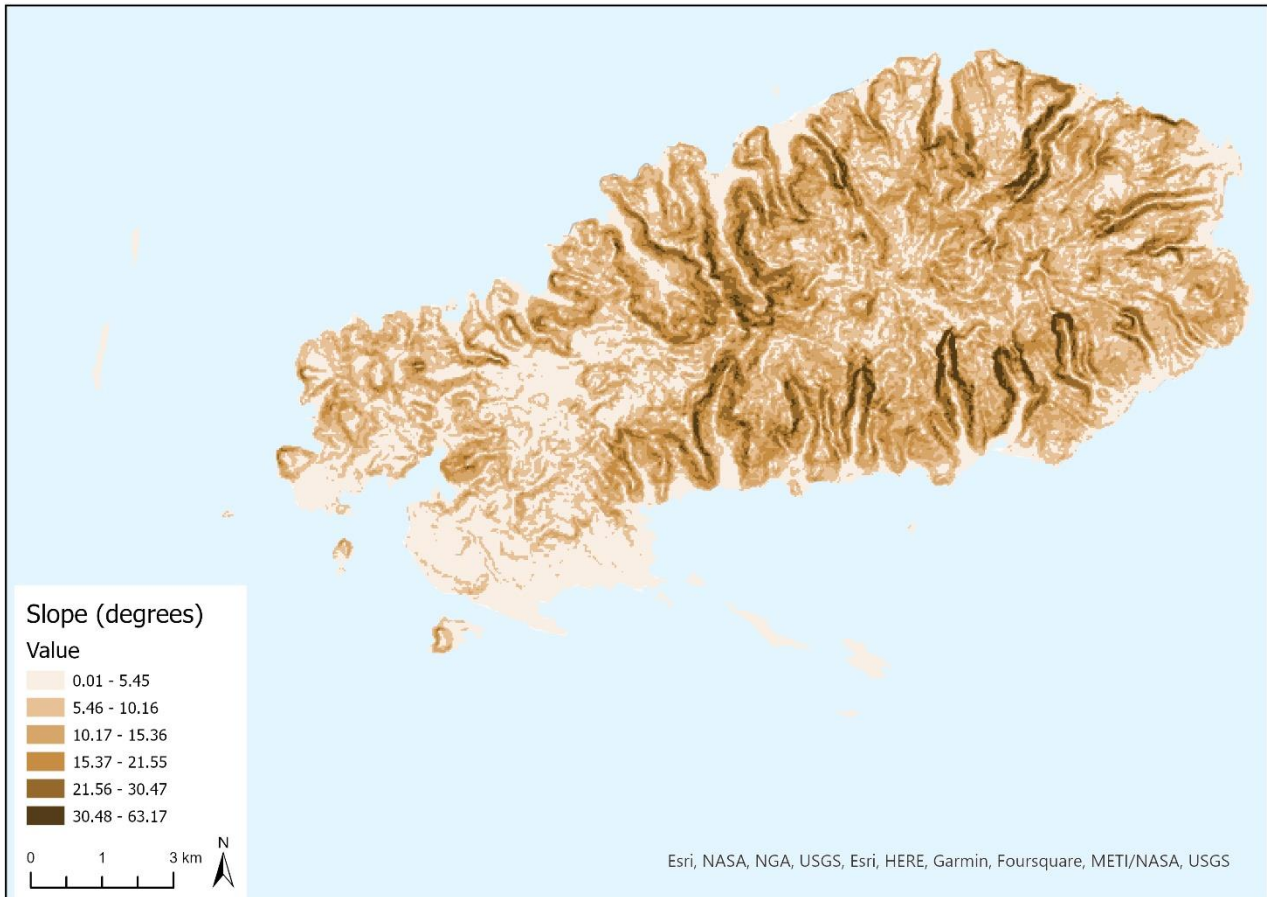
Figure 4.23: Slope in degrees.

Figure 4.24: Aspect

CHAPTER 4: Results and Discussion   47
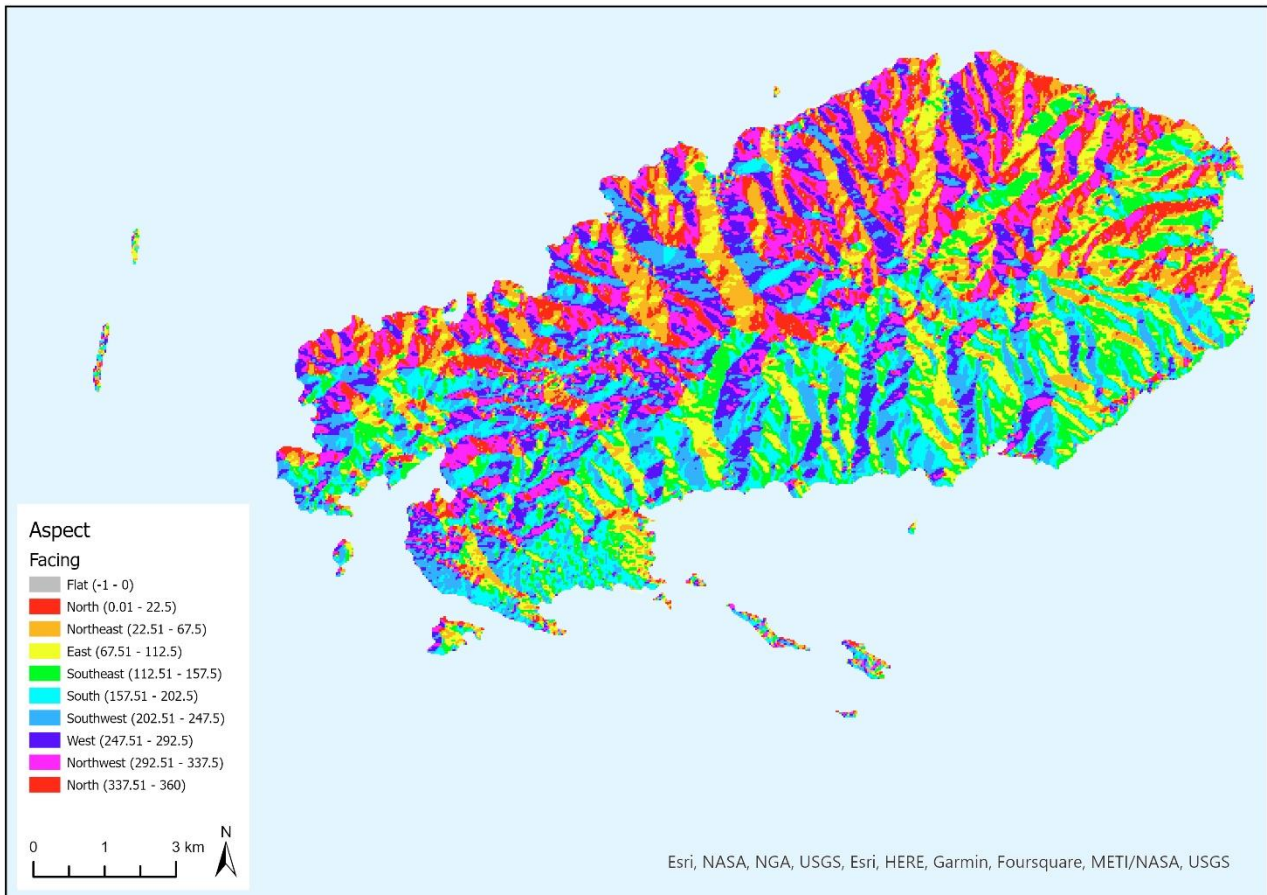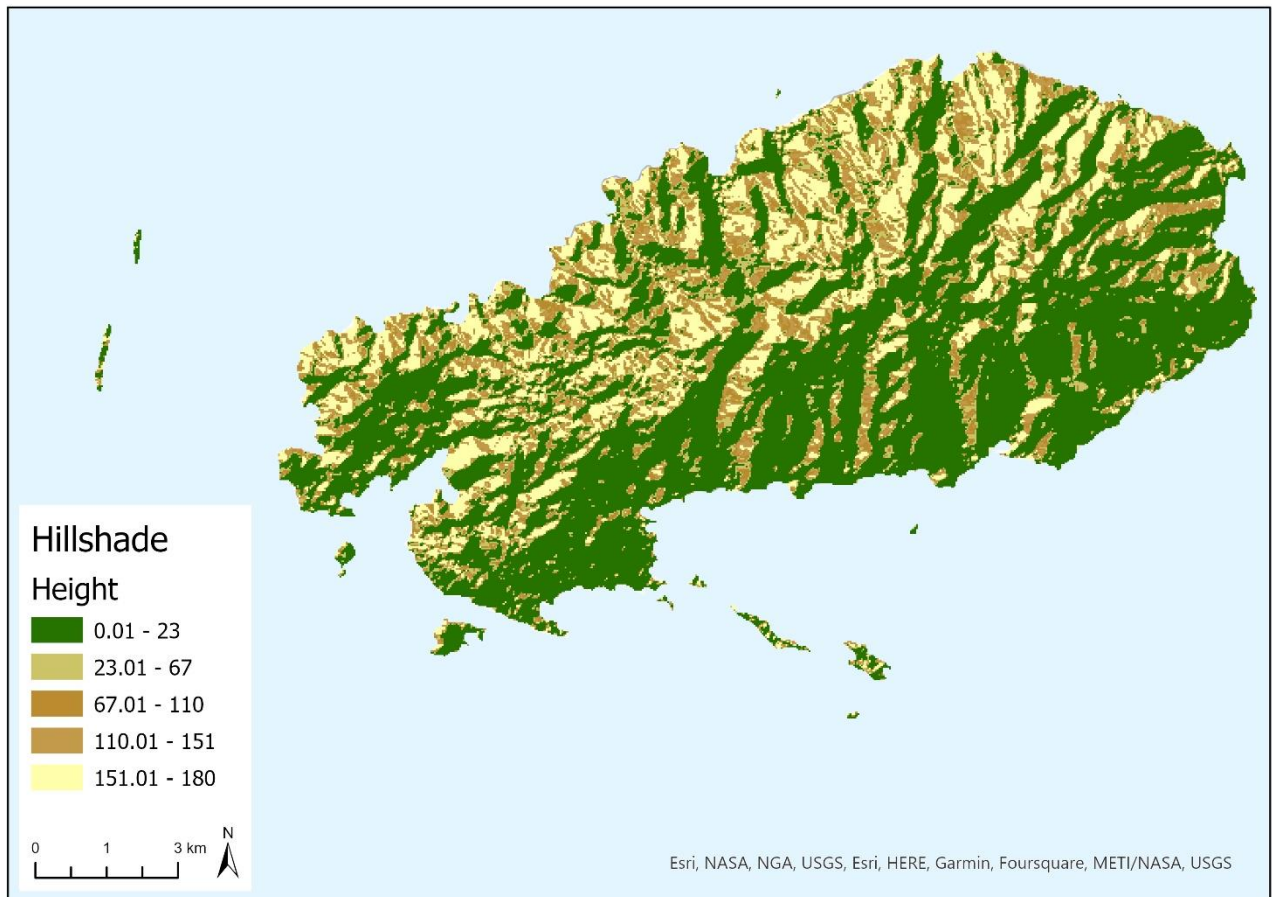
Figure 4.25: Hillshade.

4.2 Database Management

The final database consists of the layers presented in Table 4.5.

Table 4.5: Geospatial data layers of the final database.

Layer	Format	Type	Attributes
Admin Boundary	Polygon Feature	Vector	N.A.
Aspect	TIFF File	Raster	Degree
Contour (20 m)	Line Feature	Vector	Elevation (m)
Contour (50 m)	Line Feature	Vector	Elevation (m)
Dams	Polygon Feature	Vector	N.A.
Hillshade	TIFF File	Raster	N.A.
Geomorphic Zones	Polygon Feature	Vector	Geomorphic Zone
IDW Digital Elevation Model	TIFF File	Raster	Elevation Value (m)
Kriging Digital Elevation Model	TIFF file	Raster	Elevation Value (m)
Main Road	Line Feature	Vector	N.A.
Roads	Line Feature	Vector	N.A.
Secondary Road	Line Feature	Vector	N.A.
Slope	TIFF File	Raster	Degree
Stream	Line Feature	Vector	From Node, to Node, stream order
Town	Point Feature	Vector	Name

All feature layers within the database have completed metadata An example of this metadata is illustrated in APPENDIX A: Metadata Format Example (pg. 64). The metadata provides prospective data users with details pertaining to each layer within the database. These details include 1) the name of the layer, 2) tags which can be used to search for the data, 3) a summary of the data and its purpose, 4) a brief description of the data and what attributes the data has, 5) who the data is accredited to, and 6) the scale range of the data. This metadata structure is in line with the *ISO 19115: Geographic information – Metadata standard,* as previously discussed in Database Management (pg. 26). All created data are made available under the Digital Commons License and the FAIR Data Standards, as previously discussed, on Mendeley Data. The database can be accessed at https://data.mendeley.com/datasets/8sfzmkg7bp/1.

4.3 Validation

Upon following the validation method as described in Validation (pg. 27 onward), the results described below were obtained for the two different DEM's produced through interpolation. These are described separately for IDW and Ordinary Kriging in their respective sections.

4.3.1 Inverse Distance Weighted Validation (IDW)

Inverse Distance Weighted (IDW) interpolation is the less complex interpolation method used in this thesis. As previously discussed, this method made use of all defaults within the GIS tool. ArcGIS software as well as QGIS will provide default values based off of the input data, these values are calculated on the fly by the software based off the users' input data. This accommodates for users that may not have a clear understanding of interpolation. The correlation coefficient for IDW rendered the summary statistics of Table 4.6 and Figure 4.26 (pg. 45). The correlation coefficient was 0.993 at $p<0.0001$; the coefficient of determination evaluated to 0.985.

Table 4.6: Summary statistics for the interpolated IDW DEM and the control DEM.

Variable	Observations	Minimum	Maximum	Mean	Std. deviation
Rodrigues Elevation	100	4,000	359,000	135,580	90,449
IDW Elevation	100	4,311	355,313	135,817	90,571
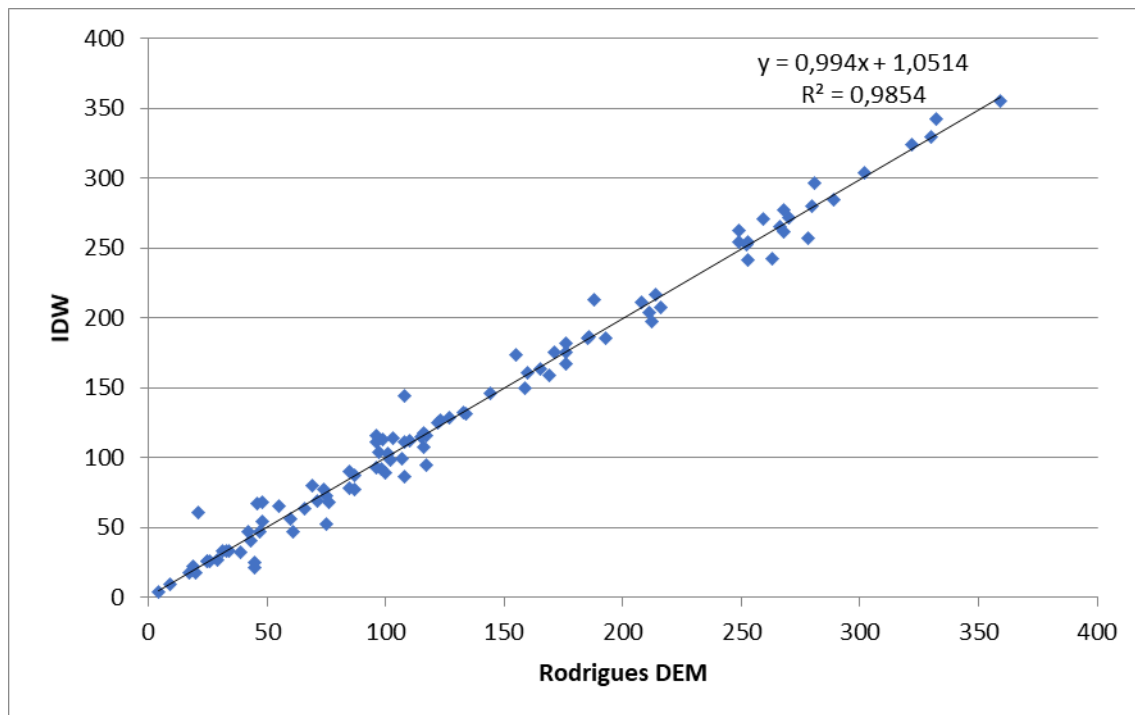
Figure 4.26: Scatter plot between the interpolated IDW DEM and the control DEM.

As seen in the results above the correlation coefficient $r = 0.993$ was obtained. Therefore, the null hypothesis is rejected. The null hypothesis assumes that the slope is 0 and no relationship exists between the two populations (refer to Table 3.4, pg. 28). With the correlation closely resembling 1 or -1 the test shows that the population variables are closely related. In relation to the sample size of 100 values a critical value of 0.165 is required or a *p*-value less than 0.05. This critical value aids in the determination of significance between the two populations. As seen in the results obtained, *r* exceeds this critical value ($p<0.0001$). $R^2 = 0.985$, indicating the high level at which the model explains the variability of the response data around its mean. As such, not only are the two populations highly correlated but they are also significantly correlated.

The *F*-test (Table 4.7) was performed to determine whether the two populations maintained the same variance or not. When the sample data were compared to that of the control Rodrigues data a *p*-value of 0.495 was obtained (null hypothesis cannot be rejected), thus ensuring that the variances between the two populations can be considered the same.

Table 4.7: F-Test Two-Sample for Variances for the interpolated IDW DEM and the control DEM (n = 100; df = 99).

	Interpolated IDW DEM	Control DEM
Mean	135,58	135,82
Variance	8181,074	8203,093
F	0,997	
P(F<=f) one-tail	0,495	
F Critical one-tail	0,717	

Student's two tailed *t*-test was then performed at $p<0.05$, yielding the results of Table 4.8 (pg. 46). The results of the Student's *t*-test yield a *p*-value of 0.985, thus allowing for the acceptance of the null hypothesis that states that the means of the populations are the

same. If a *p*-value greater than 0.05 is obtained the means are deemed to be the same between the two populations. These results show that the means of the IDW interpolation sample is the same as that of the Rodrigues population and that the created DEM (IDW raster) is very similar to that of the control DEM.

Table 4.8: Student's t-test results (independent samples, two-tailed) for the interpolated IDW DEM compared to the control DEM (df = 198).

Statistic	Value
t (Observed value)	-0,019
|t| (Critical value)	1,972
p-value (Two-tailed)	0,985
alpha	0,05

The RStudio output that was used to determine the residual standard error is shown below. This error allows the user to evaluate how much the one population may deviate from the true regression line. The IDW interpolation method returns an error value of 11.01. When related to the type of data being sampled this value illustrates that the values from the IDW created DEM may differ from those values within the control DEM no more than 11.01 m. For the extent of the study areas used in this thesis and the extent of the terrain change on Rodrigues this value of error is accepted as not significant enough to prove the IDW DEM inaccurate. Furthermore, a *p*-value of less than 0.05 (p ~ 0), shows that the residual standard error calculated is highly significant.

```
Call:
lm(formula = Data1$IDW.Elevation ~ Data1$Rodrigues.Elevation)

Residuals:
 Min 1Q Median 3Q Max
-24.774 -4.800 -0.255 3.281 38.944

Coefficients:
 Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.05137 1.99085 0.528 0.599
Data1$Rodrigues.Elevation 0.99399 0.01223 81.248 <2e-16 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.01 on 98 degrees of freedom
Multiple R-squared: 0.9854, Adjusted R-squared: 0.9852
F-statistic: 6601 on 1 and 98 DF, p-value: < 2.2e-16
```

4.3.2   Ordinary Kriging Validation

The same validation method used to validate the IDW interpolation results was used the validate that of the Ordinary Kriging interpolation results. Pearson's correlation coefficient tests for Ordinary Kriging rendered the results of Table 4.9 and Figure 4.27 (pg. 47). The correlation coefficient was 0.988 at *p*<0.0001. The coefficient of determination evaluated to 0.976.

Table 4.9: Correlation coefficient for interpolated Kriging DEM and the control DEM.

Variable	Observations	Minimum	Maximum	Mean	Std. deviation
Rodrigues Elevation	100	4,000	359,000	135,580	90,449
Kriging Elevation	100	4,607	348,881	134,686	89,585
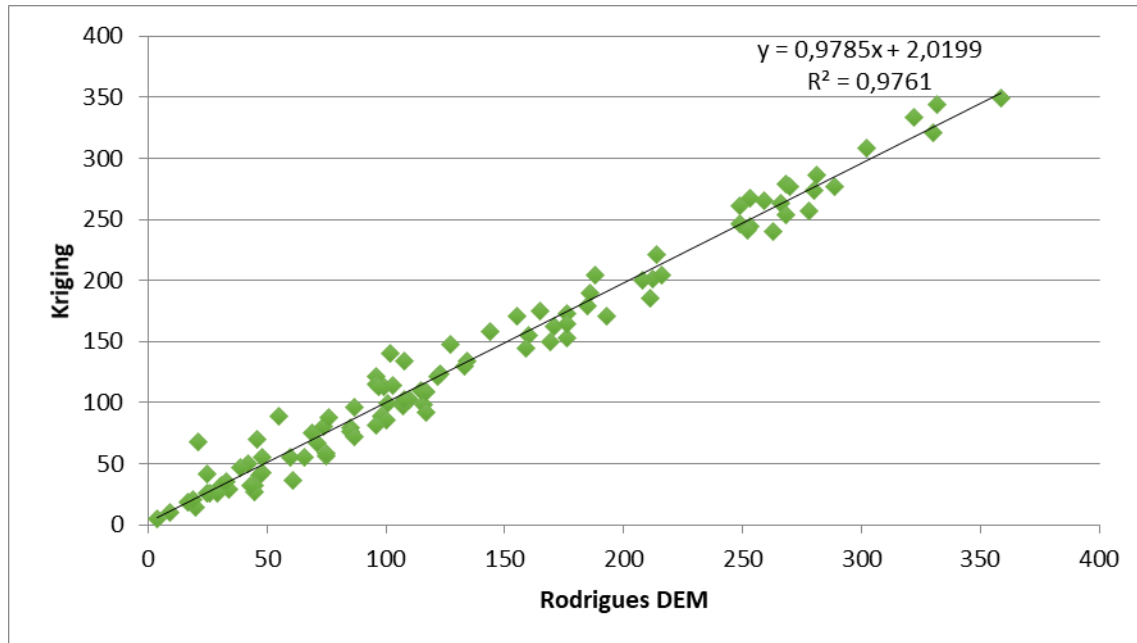
Figure 4.27: Scatter plot between the interpolated Kriging DEM and the control DEM.

As seen in the results above a correlation *r* of 0.988 was obtained. Therefore, the null hypothesis is rejected as the slope is not 0 (refer to Table 3.4, pg. 28), indicating that the populations are correlated. As mentioned in the IDW results a correlation coefficient critical value of 0.165 is required. Like the IDW results the results obtained above illustrate that the population exceeds this critical value, therefore confirming that the relationship between the two populations is significant. This is further confirmed by a *p*-value less than 0.05 ($p<0.0001$). The coefficient of determination (R^2) is close to 100%, illustrating the high level at which the model explains the variability of the response data around its mean.

As completed with the IDW sample the *F*-test was run for the Ordinary Kriging sample to determine whether the two populations maintained the same variance or not. A *p*-value of 0.462 was obtained (Table 4.10), thus ensuring that the variances between the two populations are the same.

Table 4.10: F-Test Two-Sample for Variances for the interpolated Kriging DEM and the control DEM (n = 100; df = 99).

	Interpolated Kriging DEM	Control DEM
Mean	135,58	134,6862
Variance	8181,074	8025,405
F	1,019397	
P(F<=f) one-tail	0,462025	
F Critical one-tail	1,394061	

The results of the Student's two tailed *t*-test (Table 4.11) yield a two-tailed *p*-value of 0.944, thus allowing for the acceptance of the null hypothesis as the means of the populations are the same similar to that of the IDW means. These results show that the mean of the Ordinary Kriging interpolation sample corresponds to that of the Rodrigues DEM and highly significant.

Table 4.11: Student's t-test (independent samples, two tailed) results for the interpolated Kriging DEM compared to the control DEM (df=198).

Statistics	Value
t (Observed value)	0,070
|t| (Critical value)	1,972
p-value (Two-tailed)	0,944
alpha	0,05

Below are the results of the Rstudio script that was used to determine the residual standard error.

```
Call:
lm(formula = Data1$Kriging.Elevation ~ Data1$Rodrigues.Elevation)

Residuals:
 Min 1Q Median 3Q Max
-43.551 -4.818 0.901 6.202 24.755

Coefficients:
 Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.45115 1.78887 0.252 0.801
Data1$Kriging.Elevation 1.00505 0.01108 90.742 <2e-16 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.873 on 98 degrees of freedom
Multiple R-squared: 0.9882,  Adjusted R-squared: 0.9881
F-statistic: 8234 on 1 and 98 DF, p-value: < 2.2e-16
```

The Ordinary Kriging interpolation method returns an error value of 9.873. When related to the type of data being sampled this value illustrates that the values from the Ordinary Kriging created DEM may differ from those values within the Rodrigues DEM no more than 9.873 m. This degree of error is like that of the IDW results discussed previously and can be accepted as not significant enough to prove the Ordinary Kriging inaccurate based on the extent of the study area and terrain.

With a correlation value of 0.993 and a *p*-value of 0.985, IDW interpolation is less varied and more closely related to the variables in the control DEM population than that of Ordinary Kriging, which has a correlation value of 0.988 and a *p*-value of 0.944. In contrast, when looking at the residual standard error obtained from the validation of both interpolation methods Ordinary Kriging is the more accurate interpolation method of the two. However, this does not deem the results obtained from the IDW interpolation any less useful for the method of developing a basemap using freely available resources. Furthermore, both

interpolation methods closely resemble the control DEM, with the residual error within approximately 1 m of each other. This suggests that both DEMs may be used as input to geospatial analyses.

Considering the ease of use of IDW for the uninitiated user, this elevation surface was used as input to contour delineation and river extraction for this project. However, both interpolation methods show strong relationships to the control DEM for Rodrigues. Therefore, it may be assumed that both methods are appropriate to use for this methodology. User experience and understanding of the interpolation methods should be the54ecidingg factor as to which of these two methods are used in a user's study or use case.

4.4 Base Map Compilation

Upon completion of all basemap features, these features were compiled into a single basemap. The completed basemap that was produced using all created geospatial layers based on freely available resources can be seen in Figure 4.28. Figure 4.29, in contrast, displays a basemap created using data provided by the Surveyor General of Mauritius. As with the geospatial database of all layers, produced maps are made available, free of charge, at Mendeley Data (https://data.mendeley.com/datasets/8sfzmkg7bp/1).

The basemap illustrated in Figure 4.28 contains the following layers:

Layers	Methodology
50 m Contour Lines	Contour and Coastline Creation
Dams	Digitisation
Main Roads	Digitisation
Rivers	Hydrological Modelling
Roads	Digitisation
Rodrigues	Contour and Coastline Creation
Secondary Roads	Digitisation
Towns	Digitisation

The basemap illustrated in Figure 4.29 contains the following layers:

Layers	Methodology
Primary Road	Surveyor General
Secondary Road	Surveyor General
Rivers	Surveyor General
50m Contour Lines	Surveyor General
Dams	Surveyor General
Rodrigues	Surveyor General

The basemap compiled contains all the key elements of a basemap at this scale, further details such as land parcels, building footprints, street labels, *etc.* are not required. Through visual comparison both basemaps Figure 4.28 and Figure 4.29**.** are similar with very small differences apparent, this could be due to the way in which the data is captured. These differences between the basemaps were further investigated through Ground Truthing.

Figure 4.28: The topographical basemap created by this research project of Rodrigues

.CHAPTER 4: Results and Discussion

Figure 4.29: Basemap created using the Mauritius Surveyor General data.

4.5 Ground Truthing

Upon completion of the basemap, an additional basemap based on control geospatial data were compiled for comparison purposes. Through this visual comparison many anomalies such as missing data, and additional data were identified within the basemap products using the methodology discussed here. Ground truthing was completed to confirm whether the findings were true or false.

Although Rodrigues is relatively small many of its locations are not easily accessed due to the changing terrain; many areas are heavily vegetated, and many are steeply sloped thus making ground truthing and validation difficult. One of the major differences identified between the two compiled basemaps were that of the rivers and dams, therefore more focus was placed on validation these features.

Through conversations with the residents of Rodrigues and members of the University of Mauritius it was said that the rainy seasons of the island are experienced during the months between February and June with an average rainfall of 729mm, whereas the driest season can be experience during the months of September and January with an average rainfall of 392mm (Mauritius Meteorological Services, 2016). The ground truthing for this thesis took place during the month of July. Although relatively close to the rainy season it was seen that many of the riverbeds around Rodrigues were either dry or held stagnant water. Examples of this can be seen in APPENDIX B: Rodrigues Island Rivers (pg. 65). Within these images one can visually see that none of these rivers have flowing water and, therefore, can be classified as non-perennial rivers. The importance of this is that when modelling hydrology, the methodology considers where water is most likely to flow, therefore, identifying riverbeds whether they currently have flowing water or potentially at some point in the year have flowing water. This finding could be the reasoning why the surveyor data of Rodrigues is missing some river (drainage) lines. Through ground truthing it was confirmed that many of the rivers that were not present within the surveyor data were confirmed rivers and even had the presence of manmade dams, therefore, validating these findings.

In APPENDIX C: Rodrigues Island Dam (pg. 66) one can visualize a fully established and functioning dam, which was one of the largest dams on Rodrigues. This dam is, however, not present within the surveyor data, and neither is the river that is feeding into this dam. As visualized previously in Figure 4.22 (pg. 41), this dam was captured during this thesis methodology through digitisation and its associating river was identified and produced through the hydrological modelling completed within this thesis. All dams and rivers that were accessible during ground truthing process were confirmed, therefore, allowing this thesis to confirm that all rivers and dams identified are correct and the accuracy of this thesis findings are acceptable.

Validating the contour lines and DEMs of this thesis through ground truthing were more challenging due to accessibility of the mountainous regions. Therefore, these geospatial datasets were validated statistically. However, one can also visually analyse the islands terrain and confirm these terrain changes ranging from flat coastal regions to very steep slope angles. This can be seen in APPENDIX D: Rodrigues Island Terrain (pg. 67).

4.6 Assessment of Methodology and Results

A closer look at the methodology and the corresponding results reveals both the positive aspects and the limitations of this thesis.

The methodology employed here was able to produce a basemap of Rodrigues providing users with an accurate representation of the study area. Furthermore, the methodology can be applied to other areas that may be inaccessible. Caution, however, should be used to remain cognisant of the limitations of this study and that this study was executed on a case study. Regardless, through validation and ground truthing of data produced here users have a level of confidence when employing the method followed here to produce their own data. Irrespective of this, the methodology can be adjusted and improved in various aspects. Changes to the methodology can include using different interpolation methods if users have a more advanced understanding of interpolation. Another improvement could be that of automating the process to allow for quicker run time. Unfortunately, this methodology has a heavy reliance on Internet connection for many aspects including the downloading of software, the use of software such as Google Earth™ and the processing of data.

4.6.1   Software Evaluation

The various software's used can pose both positive and negative aspects to this study. This study focused at using mostly freely available software sources. However, processing such as the interpolation methods were completed within ArcGIS Pro, which is proprietary software. The reason for this was due to the author having more experience and familiarity using this software platform. Regardless, all GIS processes can be completed in open-source software such as QGIS, as is detailed in the hydrologic modelling section.

A concern that was identified was the retirement of support to the TCX Converter software platform. The methodology relied on this software to produce elevation points, which were crucial to the production of the DEM's, which formed the basis of this thesis. TCX Convertor can still be downloaded and used for this purposed, however, there are no more updates to the software available. Regardless, Terrain Zonum Solution is another open- source software solution that can complete the same task as that of TCX Convertor.

Of note is that Terrain Zonum Solution is not as easy to use compared to TCX Converter. To illustrate this, issues experienced ranged from limited sample size to final result outputs computing failure. The Terrain Zonum Solutions requires the user to input a minimum and maximum latitude, and longitude for the user to define their extent. In the case of Rodrigues these coordinate values that define the extent are seen in Table 4.12 below.

Table 4.12: Rodrigues map extent

Min Latitude:	-19.6670199
Max Latitude:	-19.7879728
Min Longitude:	63.3215532
Max Longitude:	63.5065044

Terrain Zonum Solution then requires the user to enter the number of sample points. As stated within TCX Converter (pg. 16), a sample size of approximately 11 000 points were used within the TXC Convertor, therefore for a more accurate comparison between both software platforms a sample size of 11 000 within the Terrain Zonum Solution platform would have been ideal. This was, however, not the case as the software produced an error stating that the random sample size was out of range, this error can be seen in Figure 4.30. The sample size then had to be reduced to a maximum of 5 000 points. This already meant that the output of this process would not be as detailed as that of the TCX Converter output.

www.zonums.com says

Error: Check Number of Random Samples

Out of Range

OK

Figure 4.30: Terrain Zonum Solution Sample Size Error

Once all necessary values were inputted Terrain Zonum Solutions started processing the results for the 5 000 sample points. This process relies heavily on Internet access and can be very timely as this process completes two tasks at once: the platform obtains the sample points and their coordinates and then follows through with obtaining the elevation values for each point. Ideally, after the process is complete, the user would be able to download an Excel spreadsheet containing all the sample points with their relative attributes. Although all input values were correct and accepted by the software, Terrain Zonum Solutions failed numerous times to complete the processing of these results producing the error seen in Figure 4.31. Various Internet providers, Internet browsers, and hardware were used to attempt to complete these processes to which all received the same error. This application, therefore, is not recommended over TCX Converter, when evaluated based on user friendliness. Consequently, this thesis determined that even though TXC Converter is not supported anymore as a software with upgrades it is the more reliable freely available software platform when obtaining elevation values for specific coordinate points.

www.zonums.com says

Problem retrieving data:Malformed Server Response Status

OK
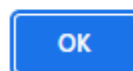
Figure 4.31: Terrain Zonum Solution Processing Error.

4.6.2 Interpolation

This thesis aimed at analysing methods of basemap production that are easy to use and require minimal knowledge of GISc. Therefore, this thesis made use of the IDW interpolation output to produce various other basemap features such as the rivers. To transform a database with elevation values into a DEM one must complete a form of interpolation. There are numerous interpolation methods available to users ranging from simple to complex. Two commonly used interpolation methods (IDW and Ordinary Kriging) were used. IDW is a simple interpolation method requiring very little statistical understanding and is, therefore, can apply to a larger array of users. Ordinary Kriging in contrast requires more statistical understanding by the user. This relates to the user being required to provide the nugget, partial sill, and lag to run this interpolation method.

An extensive validation process (see Validation, pg. 44) was followed to determine the accuracy of the interpolation methods in relation to the control DEM. The same statistical analysis was performed on both the IDW and the Ordinary Kriging DEM results. It was found that IDW is both highly correlated and significantly correlated to the control DEM with a r value of 0.993 and a R^2 value of 0.985. Using a F-test it was determined that the IDW sample alongside the control DEM has a p-value of 0.495, thus the variance between the two samples is to be regarded as equal.

The Ordinary Kriging interpolation method requires more user experience and understanding but when compared to the control DEM it presented with similar results to that of the IDW DEM validation. The Ordinary Kriging DEM sample had a r value of 0.988 and a R^2 value of 0.976. Although these values are slightly below those of the IDW sample they are still within range therefore indicating that the Ordinary Kriging DEM is highly correlated to the SRTM control DEM and is significantly correlated. A p-value of 0.462 was obtained through the F-test therefore claiming that the Ordinary Kriging Sample can be seen as closely correlated to the control sample.

Both interpolation methods are shown to closely resemble that of the control DEM and, therefore, can be confidently used for the purpose of this thesis and its methodology. With the IDW values and correlations being slightly higher than that of the Ordinary Kriging, and this method being easier to use it was the chose interpolation method to perform other geoprocessing analysis such as the hydrological modelling.

This thesis made use of default interpolation values (excepting spatial resolution), when producing the DEM. If a user is to use their own input when completing the interpolation analysis, it is important to consider the effects these values would have on the final output created.

4.6.3   Contour Lines and Coastline
Once a user has produced a DEM a simple geoprocessing tool within any GIS can be used to produce contour lines such as the Contour tool in ArcGIS Desktop. Although the contour interval can be specified to best suit the users need, for the case of Rodrigues and the purpose of this thesis two different intervals were used (20 m and 50 m). These contour lines allow for the quick visual identification and classification of the varying terrain and geomorphological zones of Rodrigues. Typically, within a 1:50 000 map scale the 20 m intervals are used. However, due to the steep terrain of the study area and for the purpose of visualisation without the cluttering of the basemap the 50 m contour interval was used. This contour interval still provided the overall basemap with detail and insight into the islands terrain without cluttering and overwhelming the basemap produced.

4.6.4   Hydrologic Analysis
Following the validation of both interpolation methods the IDW DEM was used to complete the hydrological analysis of Rodrigues. The hydrological modelling makes use of a DEM for its base input. The DEM is filled to ensure there are no sinks within the data that will affect the outcome of the model. Once the DEM is filled, flow direction is calculated to determine the direction in which water is likely to flow (downhill), considering that water will follow a path of least resistance downslope. With the flow direction determined one can calculate the flow accumulation, therefore, identifying stream locations. These stream locations were extracted to form the river channels for the basemap. To extract these river channels a data

threshold is set; this is a value that will change from depending on a given use case depending on the user's detail requirements. For this thesis a threshold of 150 was used. Through visualization of the surrounding elevation and the scale of the basemap to be produced a threshold of 150 was used as this eliminated the unnecessary small rivers that can be seen within the flow accumulation results. This threshold did, however, lead to rivers with sharper rather than gentle curvature that a natural river would present. However, a lower threshold in the case of Rodrigues would have led to many more inaccurate water segments to be left within the data. These are water segments which are identified purely based on the where water is likely to flow according to the flow accumulation output and not necessarily exactly where one can expect the water to start forming rivers on Rodrigues.

4.6.5 Digitisation

Digitisation can be a time-consuming method of data capture. However, it is in many cases necessary to capture certain basemap features such as the roads, towns, and dams. It is important to understand and acknowledge the errors and bias that can be introduced when digitising features. These errors can be avoided when ensuring that the base imagery of a study area is of high spatial resolution and is the most up-to-date. In this thesis Google Earth™ was used as it is a freely available platform that offers users imagery of high spatial resolution and cloud free, ensuring an accurate basis for digitisation. Imagery that has a low spatial resolution or higher levels of cloud cover make it harder for users to see the features that are being digitized therefore introducing errors into the final outputs Throughout the methodology of this thesis these digitized features were validated to provide users with an understanding of the accuracy of their data should this methodology be repeated.

4.6.6 Base Map Compilation

Once all basemap features were produced through the above methodology they were compiled within a basemap and correctly symbolised to provide accurate visualisation and ease of understanding to users. Symbology of these basemp features was based on the basic symbology given to these features for example rivers are symbolized as blue linear features and contour lines are symbolized as brown linear features for easy visual identification. A second basemap was produced using the data received from the surveyor general, this basemap was produced using the same symbology as the basemap produced in this thesis.

Through the visual comparison of the two basemaps differences were identified and then through ground truthing these differences were confirmed to be either true or false. Accessibility of numerous areas was not possible; however, all accessible areas were compared. The ground truthing identified many areas of missing data and discrepancies within the surveyor data basemap and in turn confirmed many of the features produced through this methodology.

After the complete analysis of the methodology and all the features which were produced it in turn confirmed that this methodology yields basic basemap features of high accuracy that can then be repurposed to map similar inaccessible areas. This is of high value to users studying areas that lack spatial data providing these users with a means of producing an accurate basemap without the need for the area in question to be ground surveyed.

4.7 Proposed Framework

The methodology employed here provides a relatively simple approach to create a basic geospatial dataset of most inaccessible and hard to reach areas. Throughout, a focus is made on free resources that are easy to use. While the quality of a dataset improves with higher quality satellite imagery (in terms of spatial resolution), or the skill of the user, the methods discussed allow an inexperienced user to create basic geospatial data. TCX Converter, and Terrain Zonum Solution are freely available and easy to use, although TCX Converter proved to be the more optimal product. Google Earth™ is freely available and integrates a variety of mosaiced satellite imagery, with often high spatial resolution. Data derived from this platform, is thus deemed as high quality for the purpose of this thesis. While digitising requires some skill of the user, executing this step in Google Earth simplifies the process. With IDW yielding improved accuracy to Ordinary Kriging in the case of Rodrigues, the steps are further simplified to the inexperienced GIS user. The proposed framework for the methodology users can follow to create geospatial data fit for their own use cases is illustrated in Figure 4.32.
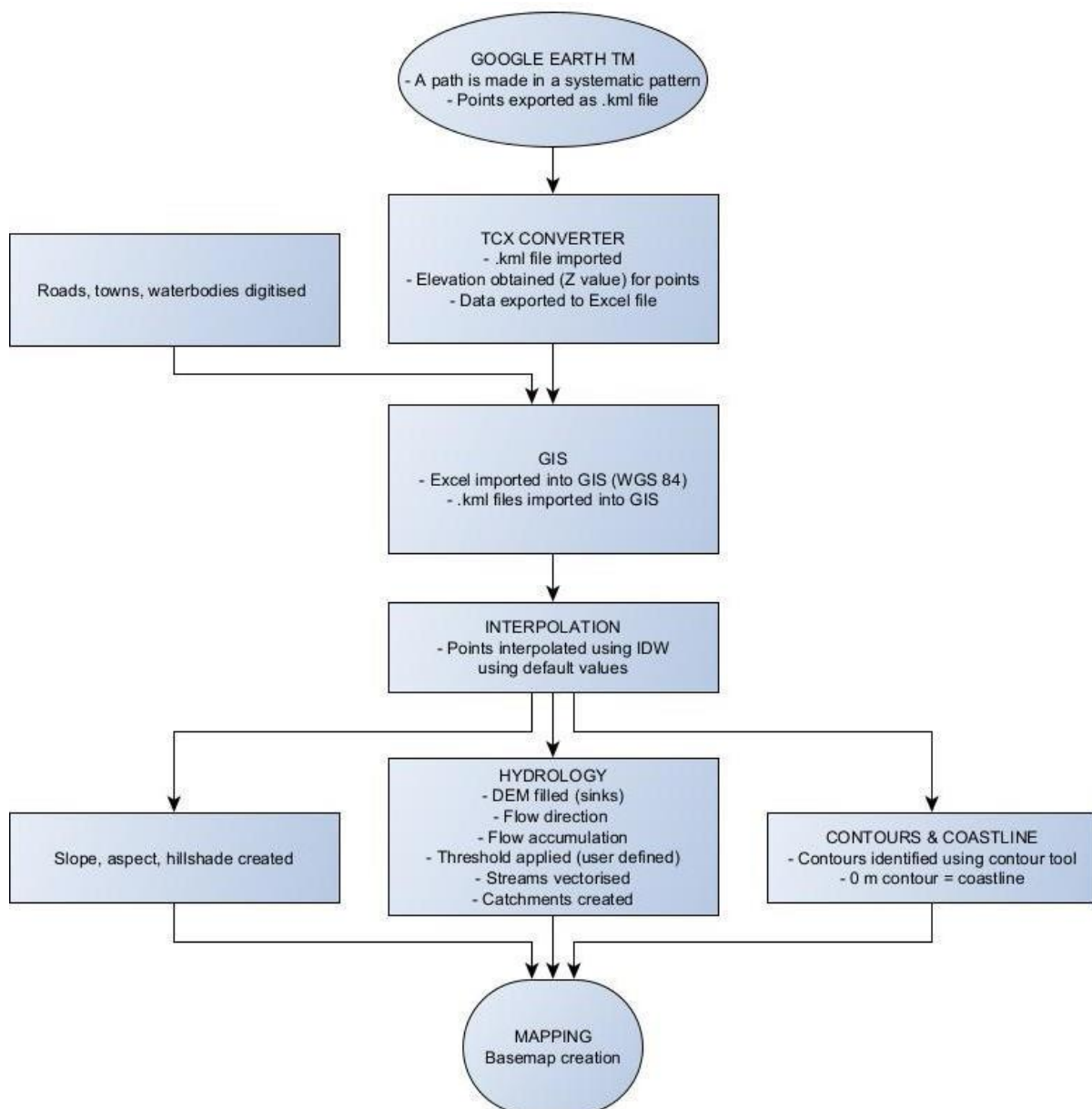

Figure 4.32: Proposed framework.

CHAPTER 5: CONCLUSIONS

Geospatial data, key to understanding of one's surroundings, are increasingly used in research and business. However, data must be accurate, precise, and current, and conform to the appropriate standards to derive the maximum benefits from any particular dataset. Unfortunately, such datasets are on occasion hard to come by. Furthermore, obtaining geospatial data is often difficult due to purchase costs, or because it cannot be collected in person due to areas being inaccessible. As such, a focus should be made on freely available remote sensing techniques and geoprocessing processes that can assist in the creation of geospatial data. This thesis aimed at understanding and formulating these techniques and processes into a comprehensive methodology, which could be replicated by other users and researchers. To ensure a reliable methodology, results underwent validation.

Google Earth™ was the chosen base platform as it is widely known to many users spanning from GIS professionals to non-GIS users. This platform offers users relatively recent satellite imagery at high spatial resolution with 0% cloud cover, therefore, providing a stable foundation of data for users to work with. As such, this platform was used to gather elevation points, which in turn were used to produce a DEM through interpolation. Google Earth™ was used for the digitisation of features such as roads, towns, and dams as the high spatial resolution and the 0% cloud cover ensured that digitisation could be done with minimal probability of error. Although the process of digitisation is slow, it is an easy method to learn and can be highly accurate if a suitable spatial resolution is used to digitize a feature. However. Users are to take into consideration that the accuracy of the digitisation output is dependent of the user's ability to capture said data without missing any features.

A combination of software was used to produce the elevation data needed to produce a DEM, with TCX Converter deemed the more user-friendly application. IDW interpolation, a less complex method to use and understand for users that do not have any statistical or GIS knowledge, proved to produce slightly more accurate results than that of Ordinary Kriging. However, the Ordinary Kriging interpolation method produced accurate results and can be used as an alternative to the IDW interpolation method for users with more statistical and GIS knowledge.

Through the combination of GIS, remote sensing, and Google Earth™ geospatial data can be created to produce a basemap or for the use of spatial analysis. A clearly defined methodology, and data ultimately uploaded to MendeleyData for other users to use, ensures the outputs adhere to FAIR data principles (Wilkinson et al., 2016; Fritz, et al., 2019). Through this methodology it was found that a combination of these techniques proved sufficient for the creation of a reliable and accurate geospatial dataset of an inaccessible area of the earth. This geospatial dataset has been created using predominantly an array of freely available resources (e.g., Google Earth™, TCX Converter, steps explained for QGIS), supporting the notions of open-access data and methods (Borgman, 2015, Rajasekar & Arunachalam, 2015, European Union Commission, 2017, Piwowar et al., 2017). Furthermore, the geospatial dataset maintains a high level of accuracy, and can now be used as a basis for decision support.

5.1 Limitations and Considerations

Usage and data ownership limitations of Google Earth™ need to be acknowledged. While the platform is useful for creating data, data created using this platform falls under specific copyright. Such limitations must be acknowledged when embarking on a project using Google Earth™.

Due to time restrictions, the field visit protocol was impacted. However, the island was traversed over a 3-day period and areas of confirmation and omission evaluated. As such, time restrictions had no real effect on data produced.

TCX Converter is no longer supported with software updates, however users are still able to access the software platform. There are, however, other software platforms that may perform the same process as that of TCX Converter, including Terrain Zonum Solutions.

This thesis made use of the Esri ArcGIS Desktop suit which is a proprietary software platform which is not available to all users. This software does offer a 21-day free trail however this may not be sufficient for users that plan to follow this methodology over many use cases. Therefore, users should consider adapting this methodology to make use of the freely available alternative to ArcGIS Desktop which is that of QGIS.

5.2 Further Areas of Research

A recommendation is to fully automate the processes to produce the output feature classes by allowing for the creation of random points within a GIS product, and subsequently integrating these with TCX Converter or an alternate to this platform. This is followed by the re-importation to a GIS and subsequent analyses. Such automation should be done using Python, which integrates with both ArcGS Pro, and GIS. The methodology should also be compared to other available DTMs, such as the 30 m ASTER DEM, or the 30 m JAXA ALOS DEM. When focusing on hydrologic modelling, it is recommended that the MERIT DEM is used for comparison purposes. Finally, it is recommended that the methodology be assessed against a high accuracy, field-based dataset.

This thesis aimed at creating digital geospatial datasets that had a known accuracy for inaccessible and hard to reach areas. This was set to be done using remote sensing and desktop techniques. This thesis successfully created numerous digital geospatial datasets ranging from rivers, elevation in the form of a DEM and contour lines, roads, towns, geomorphic zones, and numerous other datasets. These digital geospatial datasets were achieved by completing the four key objectives set.

REFERENCES

Al-Achi, A., 2019. The Student's t-Test: A Brief Description. Journal of Hospital and Clinical Pharmacy, 5(1), pp. 1-3.

Albrecht, S., Abbink, D. 2019. Fair data use: A review of ethical and policy considerations. Big Data & Society, 6(1).

Alganci, U., Besol, B., Sertel, E. 2018. Accuracy Assessment of Different Digital Surface Models. *International Journal of Geo-Information,* 7, p. 114.

Anderson, K., Ryan, B., Sonntag, W., Kavvada, A., Friedl, L. 2017. Earth observation in service of the 2030 Agenda for Sustainable Development. *Geo-spatial Information Science,* 20(2), pp. 79-96.

Aplin, P. 2003. Remote sensing: base mapping. *Progress in Physical Geography,* 27, pp. 275-283. Available at: https://www.earthobservations.org/documents/publications/201703_geo_eo_for_2030_agenda.pdf . Accessed 6 February 2023

Baxter, A., Upton, B., White, W. 1985. Petrology and geochemistry of Rodrigues, Indian Ocean. Contributions to Mineralogy and Petrology, 89, pp. 90-101.

Borgman, C. 2015. Big data, little data, no data: scholarship in the networked world. *MIT Press.*

Bounds, P., Sutherland, C. 2018. Perceptual basemaps reloaded: The role basemaps play in eliciting perceptions. *Journal of Linguistic Geography,* 6(2), pp. 145-166.

Bozek, P., Glowacka, A., Litwin, U., Pluta, M. 2016. Using GIS tools to obtain elevation models for the purpose of spatial planning. *Geomatics, Land Management and Landscape,* 2, pp. 19-29.

Briggs, D. 1977. Sources and Methods in Geography –Soil. *London: Butterworths & Co.*

Burrough, P., McDonnell, R. 1998. Creating continuous surfaces from point data. In: Burrough Principles of Geographic Information Systems. *Oxford, UK: Oxford University Press.*

Cameron, K., Hunter, P. 2002. Using Spatial Models and Kriging Techniques to Optimize Long-Term Ground-Water Monitoring Networks: A Case Study. *Environmetrics,* 13, pp. 629-659.

Chang, K. (2019), Introduction to Geographic Information Systems. *Mc Graw Hill*, 9, pp. 280-282

Childs, C. 2004. Interpolating Surfaces in ArcGIS Spatial Analyst. *ArcUser*, July – September, pp. 32-35.

Congalton, R., Green, K. 2009. Assessing the accuracy of remotely sensed data principles and practices. *CRC Press.*

Crawford, K. 2016. The hidden biases in big data. *Communications of the ACM*, 59(1), 104-113.

Daly, E. 2016. Using Google Earth for Field trips and map making. *Irish Journal of Technology Enhanced Learning.* 1(1). Pp. 8-12

Domingo-Ferrer, J. 2017. Fair data: A new paradigm for data protection. *Computer Law & Security Review*, 33(1), pp. 60-66.

Doornkamp, J., King, C. 1971. Numerical analysis in geomorphology –an introduction.

Dramis, F., Guida, D., Cestari, A. 2011. Nature and Aims of Geomorphological Mapping. *Geomorphological Mapping: methods and applications.* Elsevier: London, pp. 39-74.

Du, Y., Zhang, Y., Ling, F., Wang, Q., Li, W., Li, X. 2016. Water bodies' mapping from Sentinel-2 imagery with Modified Normalized Difference Water Index at 10-m spatial resolution produced by sharpening the swir band. *Remote Sensing,* 8(4).

eGyanKosh, 2018. UNIT 4 INTRODUCTION TO GEOSPATIAL DATA. *Concepts of Geospatial Data. https://egyankosh.ac.in/bitstream/123456789/39512/1/Unit-4.pdf.* Accessed 7 February 2023

El-Ashmawy, K., 2016. Investigation of the accuracy of Google Earth elevation data. *Artificial Satellites,* 51(3), pp. 89-97.

El-Hallaq, M. A., Hamad, M. I. 2017. Positional Accuracy of the Google Earth Imagery In The Gaza Strip. *Journal of Multidisciplinary Engineering Science and Technology (JMEST),* 4(5), pp. 7249-7253.

Elkhrachy, I. 2018. Vertical accuracy assessment for SRTM and ASTER Digital Elevation Models: A case study of Najran city, Saudi Arabia. *Ain Shams Engineering Journal*, 9, pp. 1807-1817.

European Union Commission. 2017. Commission Recommendation on access to and preservation of scientific information. *Official Journal of the European Union*, L117, 1-7.

European Union Commission. 2019. Key principles of data governance in the EU. European Commission, *Directorate-General for Justice and Consumers*.

Farr, T., Rosen, P., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., Alsdorf, D. 2007. The shuttle radar topography mission. *Reviews of Geophysics*, 45,(2),

Floridi, L. 2019. The Ethics of Information. *Oxford University Press*

Folger, P. 2009. Geospatial Information and Geographic Information Systems (GIS): Current Issues and Future Challenges*, Congressional Research Service*.

Freeman, G.T. 1991. Calculating catchment area with divergent flow based on a regular grid. *Computers and Geosciences*, 17. pp 413-22

Fritz, S., See, L., Carlson, T. et al. (28 more authors), 2019. Citizen science and the United Nations Sustainable Development Goals. *Nat Sustain,* 2, pp. 922-930.

GEO, 2016. *Earth Observations in Support of the 2030 Agenda for Sustainable Development.* [Online]

Global Partnership for Sustainable Development Data, 2019. *Data for Now.* [Online] Available at: https://www.data4sdgs.org/index.php/initiatives/data-now

Goovaerts, P. 1997. Geostatistics for Natural Resources Evaluation. *New York: Oxford University Press.*

Gustavsson, M., Kolstrup, E., Seijmonsbergen, A. 2006. A new symbol-and-GIS based detailed geomorphological mapping system: Renewal of a scientific discipline for understanding landscape development*, Geomorphology, 77(1-2). Pp 90-111.*

Hanley, H., Fraser, C. 2001. Geopositioning accuracy of Ikonos imagery: indications from

REFERENCES

two dimensional transformations. *Photogrammetric Record,* 17, pp. 317-330.

Hantke, R., Scheidegger, A. 1998. Morphotectinics of the Mascarene Islands. *Annali Di Geofisica,* 41(2), pp. 165-181.

Holland, D., Allan, L. 2001. The Digital National Framework and digital photogrammetry at Ordnance Survey. *Photogrammetric Record,* 17, pp. 291-301.

Jenson, S., Domingue, J. 1988. Extracting Topographic Structure from Digital Elevation Data for Geographic Information System Analysis. *Photogrammetric Engineering and Remote Sensing,* 54(11), p. 1593–1600.

Jiao, Z., Li, X., Wang, J., Yan, G. 2001. Classification-based fusion of IKONOS 1-m high-resolution panchromatic image and 4-m multi-spectral images. *IEEE, Sydney*.

Jones, A. (1982) Manual of Photogrammetry, eds C.C. Slama, C. Theurer and S.W. Hendrikson*, American Society of Photogrammetry*, Falls Church, Va., 1980, Fourth Edition, 12(4)

Kervyn, M., Ernst, G., Goossens, R., Jacobs, P. 2008. Mapping volcano topography with remote sensing: ASTER vs. SRTM,. *International Journal of Remote Sensing*, 29(22), pp. 6515-6538.

Kettunen, P., Koski, C., Oksanen, J. 2017. A design of contour generation for topographic maps with adaptive DEM smoothing. *International Journal of Cartography*, 3(1), pp. 19-30.

Kumari, S. 2018. Planning and Design of Surface Drainage System for Jhilli Chaur (Pusa Farm), Samastipur (Bihar), Pusa (Samastipur): Dr. Rajendra Prasad Central Agricultural University. *London: Butler & Tanner Ltd*.

Longley, P., Goodchild, M., Maguire, D., Rhind, D. 2015 Geographic Information Systems and science. *Chichester: John Wiley & Sons.* Ltd. 4

Malaperdas, G., Panagiotidis, V. (2018). The aspects of Aspect: Understanding land exposure and its part in geographic information systems analysis. *Energy & Environmen*t. 29.

Marsudi, I. 2017. Making a digital contour map. *Advances in Social Science, Education and Humanities Research: Proceedings of the International Conference on Technology and Vocational Teachers (ICTVT 2017),* 102, pp. 41-47.

Mesić, I. 2016. Comparison of Ordinary and Universal Kriging interpolation techniques on a depth variable (a case of linear spatial trend), case study of the Šandrovac Field. *The Mining-Geology-Petroleum Engineering Bulletin,* pp. 41-58.

Middleton, G., Burney, D. 2013. Rodrigues – An Indian Ocean Island Calcarenite: Its History, Study and Management. In: M. Lace & J. Mylroie, eds. *Coastal Karst Landforms.* s.l.:Springer Science + Business Media Dordrecht, pp. 261-276.

Mohammed, N., Ghazi, A., Mustafa, H., 2013. Positional Accuracy Testing of Google Earth. *International Journal of Multidiciplinary Sciences and Engineering,* 4(6), pp. 6- 9.

Napieralski, J., Barr, I., Kamp, U., Kervyn, M., 2013. Remote Sensing and GIScience in Geomorphological Mapping. *San Diego: Academic Press.*

Nelson, A., Reuter, H., Gessler, P. 2009. Chapter 3 DEM Production Methods and Sources. *Developments in Soil Science,* 33, pp. 65-85.

Nikolakopoulos, K., Kamaratakis, E., Chrysoulakis, N. 2006. SRTM vs ASTER elevation products. Comparison for two regions in Crete, Greece. *International Journal of*

REFERENCES

Remote Sensing, 27(21), pp. 4819-4838.

Nkeki, N., Asikhia, M. (2014). Mapping and Geovisualizing Topographical Data Using Geographic Information System (GIS). *Journal of Geography and Geology.* 6. pp. 1-13.

Nwauzoma, A. B. 2016. A review of geographic information systems and digital imaging in plant pathology application. *African Journal of Agricultural Research*, 11, pp. 4172 - 4180.

O'Neil, C. 2016. Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. *Crown/Archetype.*

Otto, J., Smith, M., 2013. Geomorphological mapping. *British Society for Geomorphology .*

Paganini, M., Petiteville, I., Ward, S., et al (2018) Satellite earth observations in support of the sustainable development goals. http://eohandbook.com/sdg/files/CEOS_EOHB_2018_SDG.pdf.

Pampel, H. 2017. Open access to research data in Europe: lessons learned and future directions. *Journal of the Association for Information Science and Technology*, 68(10), 2421-2432.

Pasnin, O., Attwood, C., Klaus, R. 2016. Marine systematic conservation planning for Rodrigues, Island western Indian ocean. *Ocean and Coastal Management*, 130, 213-220.

Peng, R., Dominici, F., Zeger, S., 2006. Reproducible epidemiologic research. American *Journal of Epidemiology*, 163 (9), 783–789

Piwowar, H., Vision, T., Whitlock, M. 2011. Data archiving in ecology and evolution. *Trends in Ecology & Evolution*, 26(8), 430-431.

Planchon, O., Darboux, F. 2002. A fast, simple and versatile algorithm to fill the depressions of digital elevation models. *Catena,* 46(2), pp. 81-100.

Price, M. 2006. Creating Cool Contours Modeling Glacial Terrain with ArcGIS. *ArcUser*, April-June, pp. 48-51.

Rajasekar, A., Arunachalam, S. (2015). The state of open data: historical review, current status, and future prospects. *Annual Review of Information Science and Technology*, 49(1), 606-643.

Rey, S. 2014. Open regional science. *The Annals of Regional Science*, 52 (3), 825–837

RStudio Team, 2020. RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/

Rusli, N., Majid, M., Din, H. (2014) Google Earth's Derived Digital Elevation Model: A comparative assessment with Aster and SRTM Data, *IOP Conference Series: Earth and Environmental Science*, 18

Saddul, P. 2002. Mauritius: A Geomorphological Analysis. *Mahatma Gandhi Institute*. 3. Pp. 1-354

Schumann, G., Bates, P. (2018), The Need for a High-Accuracy, Open-Access Global DEM. *Front. Earth Sci*. 6:225

Setianto, A., Triandini, T. 2013. Comparison of Kriging and Inverse Distance Weighted (IDW) interpolation methods in lineament extraction and analysis. *Journal of Southeast Asian Applied Geology*, 5(1), pp. 21-29.

REFERENCES

Sheldon, A. 2018. Good practices and emerging trends on geospatial technology and information applications for the Sustainable Development Goals in Asia and the Pacific. *Space Applications Section*

Simelane, S., Hansen, C., Munghemezulu, C. 2021. The Use of Remote Sensing and GIS for Land Use and Land Cover Mapping in Eswatini: A Review. *South African Journal of Geomatics*, 10(2), pp. 181-206.

Singleton, D.A., Spielman, S., Brunsdon, C. 2016. Establishing a framework for Open Geographic Information science, *International Journal of Geographical Information Science*, 30(8), 1507-1521.

Sodhi, N., Gibson, L., Raven, P. 2013. Conservation Biology: Voices from the Tropics. 1st ed. *John Wiley & Sons, Ltd*.

Strobl, J., Nazarkulova, A. 2014. Open Geospatial Data: New Opportunities for GIS and GIScience in Central Asia? *Urumqi, XIEG CAS*.

Takaku, J., Tadono, T., Tsutsui, K., Ichikawa, M. 2016. Validation of 'AW3D' global DSM generated from ALOS PRISM. *ISPRS Ann Photogramm Remote Sens. Spat. Inf. Sci.,* 3, pp. 25-31.

Till, R. 1985. Statistical Methods for the Earth Scientist –an introduction. *London: MacMillian Education Ltd.*

Tomar, A., Singh, U., 2012. Geomorphological Mapping Using Remote Sensing and GIS A Tool for Land Use Planning Around Shivpuri City, M.P., India. *IOSR Journal of Computer Engineering,* 5(1), pp. 28-30.

Tooth, S. 2013. Google Earth™ in Geomorphology: Re-Enchanting, Revolutionizing, or Just another Resource? *Treatise on Geomorphology,* 14, pp. 53-64.

Toz, G., Erdoğan, M. (2008). DEM (DIGITAL ELEVATION MODEL) PRODUCTION AND ACCURACY MODELING OF DEMS FROM 1:35.000 SCALE AERIAL PHOTOGRAPHS. *Conference: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*

United Nations, 2020. *Sustainable Development Goals Knowledge Platform.* [Online] Available at: https://sustainabledevelopment.un.org/?menu=1300 Accessed 6 February 2023

Unmar, B., Chinnee, D. 2015. Digest of Statistics of Rodrigues 2014*, Port Louis: Statistics Mauritius.*

Vangu, G., Dima, N. 2018. Generating the Digital Terrain Model based on Google Earth resources. *Annals of the University of Petrosani Mining Engineering,* pp. 98-109.

Varga, M., Bašić, T. 2015. Accuracy validation and comparison of global digital elevation models over Croatia. *International Journal of Remote Sensing,* 36(1), pp. 170-189.

Wackernagel, H. 1995. Ordinary Kriging. In: Multivariate Geostatistics. In: *Multivariate Geostatistics.* Berlin, Heidelberg: Springer, pp. 74-81.

Walter, C. 2020. Sustainable Financial Risk Modelling Fitting the SDGs: Some Reflections, *Sustainability*, 12, pp. 1-28

Wang, L., Liu, H. (2006). An efficient method for identifying and filling surface depressions in digital elevation models for hydrologic analysis and modelling. *International Journal of Geographical Information Science.* 20. 193-213.

Waters, N. 2017. Tobler's First Law of Geography. *In International Encyclopedia of*
REFERENCES

Geography: People, the Earth, Environment and Technology, pp. 1-13.

Wegner, T. 2016. Applied Business Statistics: Methods and Excel-based Applications. 4th ed. *Cape Town: Juta & Company Ltd.*

Wilkinson, M., Dumontier, M., Aalbersberg, I. 2016. The FAIR Guiding Principles for scientific data management and stewardship. 3 ed. *Sci Data.*

Williams, T., Sweeney, D., Anderson, D. 2006. Contemporary Business Statistics with Microsoft® Excel. *Mason: Thomson South-Western.*

Yiotis, K. 2005. The Open Access Initiative: A New Paradigm for Scholarly Communications. *Information Technology and Libraries*, 24(4), 157-162

Zhou, Q. 2017. Digital Elevation Model and Digital Surface Model. *The International Encyclopedia of Geography*

APPENDIX A: Metadata Format Example

APPENDIX B: Rodrigues Island Rivers

APPENDIX C: Rodrigues Island Dam

APPENDIX D: Rodrigues Island Terrain

APPENDIX D