UNIVERSIDAD POLITÉCNICA DE MADRID

E.T.S.I. en Topografía, Geodesia y Cartografía

Departamento de Ingeniería Topográfica y Cartografía

Doctoral Thesis

**Vario-Model for Estimating
and Propagating DEM Vertical
Accuracy: Case of Lebanon**

Wassim Katerji
(Master of Science)

Director: Dr. Mercedes Farjas Abadía
Co-Director: Dr. María Del Carmen Morillo Balsera

2016

Tribunal nombrado por el Sr. Rector Magnífico de la Universidad
Politécnica de Madrid, el día... de..... de 2016.

Presidente:

.....

Secretario:

.....

Vocales:

.....

.....

.....

.....

Suplentes:

.....

.....

Realizado el acto de defensa y lectura de la Tesis el día ... de de 2016
en la E.T.S.I. en Topografía, Geodesia y Cartografía de la Universidad
Politécnica de Madrid

Acknowledgement

In February 2013, I started my PhD studies with a general idea of what are my objectives and why. Since then, I have learned a lot: deepened my knowledge in the geospatial field, developed new skills, and built up my capacity to conduct a scientific research, which summed-up to this thesis. However, this achievement would not have been realized without the help and support of several good people:

First, I would like to express my gratitude and appreciation to my advisor, Prof. Mercedes Farjas Abadia, who have been of a tremendous help and support since day one. I would like to thank you for your encouragement and guiding throughout the first preparatory year, the development of the topic, and in the production of the overall work.

Second, I would like to express my gratitude to my co-advisor, Prof. Maria del Carmen Morillo Balsera, who have been also of tremendous help, especially in the field of geostatistics. Thank you for incorporating this invaluable aspect into my research.

Also, I would like to thank my good friends, Amy Wesley and Wissam Kosseifi, for their review and editing of the final thesis. Your inputs have raised the English writing and logical sequences to a professional level and helped me better share my ideas and findings.

Last but not the least, I would like to thank my family: My wife Hiba, you have always been my inspiration to evolve and progress in all the aspects of life. I would like to thank you for endless support and sacrifices. As for my parents, you have always encouraged me to pursue a higher education, and without your generous support, this PhD would have remained a dream far out of reach.

Abstract

A Digital Elevation Model (DEM) is a 3D continuous representation of the terrain used in Geographic Information Systems (GIS) for visualizing data and performing various analyses. Previous studies have proposed different techniques to assess the accuracy of a DEM before using it. Yet there was a lack when raising the following issues: Over a large coverage area, is the accuracy of a DEM constant, or does it change? And what is the accuracy of the analysis-derived results? This research tackles these two issues and proposes a procedure to generate a varying local accuracy model, called vario-model, and propagates it into the results of the DEM analysis through a proposed slope significance model.

The procedure is tested on three DEMs (ASTER, SRTM, and a local one based on 1/20,000 topographic maps), and validated using a set of 1,105 Ground Control Points (GCPs) over the country of Lebanon. The vario-model is generated by grouping regions of similar Land Use / Land Cover (LULC) and calculating the average local accuracy for each region. To propagate it, the accuracy of the terrain slope is calculated in angular unit, and then the slope significance is calculated as the percentage of the slope uncertainty with the respect to the slope value. The procedure is automated using Python script running on Esri ArcGIS.

Validating the solution showed that there is a pattern in the accuracy change with respect to the LULC, and that the proposed vario-model has a better representation of the DEM accuracy compared to the conventional single-values one by 7, 24 and 13 % for the ASTER, SRTM, and the local DEM respectively. Applying the slope significance into hydrological analysis showed that streams generated from the ASTER and SRTM DEMs in significant areas had their plane accuracy within the threshold of international mapping standards, while in other areas, the accuracy of the results was uncontrollable. Results generated from the local DEM in significant areas had their accuracy within the double of the allowed threshold.

While the topic of geographic data accuracy has already been tackled by previous researches, this study proposed a new innovative technique to represent the accuracy of any given DEM within any geographic extent, in a way that can be propagated to estimate the accuracy of the derived products before performing any analysis and conducting and further assessment. This technique will help researchers and practitioners from the geospatial field in anticipating the quality of their analyses from the beginning, and therefore assessing the usability of the DEM in-hand.

Resumen

Un Modelo Digital de Elevaciones (MDE) es una representación continua del terreno que en los Sistemas de Información Geográficos (SIG) se utiliza para la visualización de datos y para la realización de análisis espaciales. En estudios previos se encuentran propuestas de diferentes técnicas que permiten determinar la precisión de un MDE antes de su utilización. Sin embargo, existe un vacío si se plantean las siguientes cuestiones: sobre una superficie de gran extensión, ¿la precisión del MDE es constante o cambia?; y ¿cuál es la precisión de los resultados de un análisis derivado de un MDE? La presente investigación trata estas dos cuestiones y propone un procedimiento para generar un modelo de precisión local variable, llamado por ello *vario-modelo*, que además permite propagar la cuantificación de la precisión a los resultados de análisis sobre el propio MDE, tomando como parámetro la inclinación de la superficie del modelo original, es decir a través de la pendiente del MDE sobre el que se trabaja.

En la presente investigación, una vez diseñada y justificada la propuesta metodológica, ésta se comprueba sobre tres modelos digitales de elevaciones de Líbano (ASTER, SRTM, y uno local obtenido a partir de mapas topográficos a escala 1/20,000); y posteriormente se valida utilizando un conjunto de 1.105 puntos de control terrestres distribuidos a lo largo de todo el país. El modelo de precisión local variable o *vario-modelo*, se genera agrupando las zonas que tienen un uso o cobertura de suelo similar, y calculando la precisión media local de cada una de ellas. Para propagarlo, se calcula en primer lugar, la precisión de la pendiente del terreno en unidades angulares, y posteriormente la pendiente significativa como un porcentaje de la incertidumbre de la pendiente con respecto al valor de la misma en el modelo. Este procedimiento se automatiza programando el algoritmo en lenguaje Python y creando un script que permita trabajar en la aplicación Esri ArcGIS.

En la validación de la solución propuesta, es decir en la validación del modelo de cambio de precisión en relación a la cobertura del suelo como un nuevo MDE local, se obtiene que este nuevo *vario-modelo* mejora la representación de la precisión del MDE en un 7,24 % para ASTER y en un 13 % para SRTM. Aplicando este nuevo modelo en análisis hidrológicos, se comprueba que los cursos de agua que se generan desde los MDE de ASTER y SRTM en áreas significativas tienen su precisión superficial dentro de los umbrales de los estándares cartográficos internacionales, mientras que, en otras áreas, la precisión de los resultados es incontrolable. Sin embargo, los resultados generados en áreas significativas desde el MDE local o *vario-modelo*, tienen su precisión por encima del doble de los umbrales permitidos.

Mientras que el tema de la precisión de los datos geográficos se ha tratado en investigaciones previas, esta tesis propone una técnica innovadora para representar la precisión de un MDE concreto, creando un modelo de precisión para cualquier extensión geográfica, que además puede propagarse para estimar la precisión de un producto derivado de análisis espacial, antes de que este análisis se realice, permitiendo además evaluaciones adicionales. Los resultados de esta investigación serán de gran ayuda a investigadores y profesionales de la información geoespacial permitiéndoles desde el principio anticipar la calidad de sus análisis espaciales y con ello evaluar la funcionalidad y la utilidad de un MDE específico.

Table of Contents

Chapter 1: Introduction	1
1.1 Challenges and Motivations.....	4
1.2 Research Objectives	6
1.3 Work Dissemination	7
1.4 Thesis Outline	8
Chapter 2: Theoretical Background	13
2.1 Field of Research: DEM Accuracy Assessment	14
2.2 History and Work in Progress.....	15
2.2.1 Accuracy Assessment Methodology	15
2.2.2 Accuracy Standards.....	16
2.2.3 Innovative Methods	17
2.3 Theoretical Principles.....	17
2.3.1 Sources of DEM.....	17
2.3.2 Surface Interpolation Techniques.....	24
2.3.3 Derived Surfaces	27
2.3.4 Applications of Terrain-Based Analysis	30
2.3.5 Land-Use / Land-Cover	34
2.3.6 Geostatistics	37
2.4 Summary	40
Chapter 2 Bibliography.....	43
Chapter 3: Data and Methods.....	51
3.1 Platform and Tools	53
3.2 Area of Study.....	55
3.3 Input Data	60
3.3.1 GCPs.....	60
3.3.2 DEM	63
3.3.3 Land-Use / Land-Cover:.....	67
3.3.4 Pre-Analysis Data Transformation	71
3.4 Exploratory Data Analysis	72
3.4.1 Descriptive Geostatistics of the GCPs	72
3.4.2 Altitude Variation per LULC Class.....	75
3.4.3 Distribution of GCPs per LULC Class.....	77

3.5 Summary	80
Chapter 3 Bibliography	81
Chapter 4: The Accuracy Vario-Model	83
4.1 The Local Accuracy	86
4.2 Calculating the Local Accuracy	89
4.3 Validating the Procedure	91
4.4 Results	94
4.4.1 Individual Accuracy at the GCPs	94
4.4.2 Model 2: Average Accuracy	99
4.4.3 Model 3: Vario-Model	100
4.4.4 Models Assessment	104
4.5 Discussion	108
4.6 Summary	110
Chapter 4 Bibliography	110
Chapter 5: Applications of the Vario-Model	111
5.1 What is a Slope?	113
5.2 Accuracy and Significance of Slope	116
5.3 Application in Hydrology	118
5.3.1 Area of Study: The Litany River	118
5.3.2 Input Data	119
5.3.3 Detecting River Streams	120
5.3.4 The Reference Model	125
5.4 Results	128
5.4.1 Slope Accuracy	128
5.4.2 Slope Significance	129
5.4.3 Correlation Analysis	133
5.4.4 Length of Segments	135
5.4.5 Deviation of Segments	137
5.4.6 Linking Deviation to Significance	139
5.5 Discussion	141
5.5.1 Significant versus Non-Significant	142
5.5.2 Flat Areas	147
5.5.3 Refined Procedure	150
5.6 Summary	153
Chapter 5 Bibliography	155

Chapter 6:	157
Conclusion	157
6.1 Empirical Findings	161
6.2 Theoretical Implication	163
6.3 Challenges and Limitations	165
6.4 Future Work	166
6.5 Final Remarks	167
Appendix	168
Appendix A: Data	170
A1: Conversion Table from SDATL (in French) to CORINE (in English) Classification	170
A2: Descriptive Geostatistics of GCPs Z per LULC Classification	172
A3: Descriptive Geostatistics of GCPs Local Accuracy per LULC Classification	175
Appendix B: Python Scripts	178
B1: Calculate the Local Accuracy for the GCPs	178
B2: Calculate Slope Accuracy	179
Appendix C: Photo Gallery	180
L1: Artificial Surfaces	180
L2: Agricultural Areas	181
L3: Forests and Semi-Natural Areas	182
L5: Water Bodies	184

List of Figures

Figure 1: DTM vs. DSM	14
Figure 2: Difference between ellipsoid and geoid.....	18
Figure 3: Leveling through topographic survey	19
Figure 4: GPS satellites and receivers	20
Figure 5: Stereoscopic satellite acquisition	21
Figure 6: Capture and processing of SAR image	22
Figure 7: Airborne laser scanner	23
Figure 8: Aspect directions (Extracted from ArcGIS 10.2 Help).....	28
Figure 9: Curvature diagram (Extracted from ArcGIS 10.2 Help)	28
Figure 10: Contour lines (Extracted from ArcGIS 10.2 Help).....	29
Figure 11: Viewshed parameters (Extracted from ArcGIS 10. Help)	29
Figure 12: Azimuth and altitude of the sun (Extracted from ArcGIS 10.2 Help)	30
Figure 13: 3D view of slope failure leading to landslide (Xie, Esaki, Zhou, & Mitani, 2003)..	31
Figure 14: Volume calculation between the terrain and a below plane	31
Figure 15: Visibility in military operations (US Naval Research Laboratory, 2015).....	32
Figure 16: Fresnel Zone between two communication antennas	33
Figure 17: Schematic representation of a sample drainage system.....	34
Figure 18: Sample watershed and its components (Extracted from ArcGIS 10.2 Help).....	34
Figure 19: Directional analysis of the semivariogram (Extracted from ArcGIS 10.2 Help)	39
Figure 20: Sample distance analysis using Spatial Analyst (ArcGIS Spatial Analyst Extension, 2015)	53
Figure 21: Sample geostatistical data Exploration (ArcGIS Geostatistical Analyst, 2015)	54
Figure 22: The Python console in ArcMap (Screenshot taken by Author).....	55
Figure 23: Location of Lebanon on the Mediterranean Sea (Map designed by Author).....	56
Figure 24: General topography of Lebanon (Map designed by Author)	57
Figure 25: 3D Perspective views of Lebanon. (A) Beirut and Mount Lebanon, taken from south-west; (B) The Beqaa valley, taken from the south; and (C) The southern hills, taken from south-west (Maps designed by Author).....	57
Figure 26: Triangulation reference points symbols (Left: 1 st order, Right: 2 nd and 3 rd Order)59	
Figure 27: 2 nd order triangulation reference point placed on top of a building in the Dora area at the north-eastern entrance of Beirut city (Photo taken by author)	59
Figure 28: Map showing the distribution of the selected 1,105 GCPs (Map designed by Author)	61
Figure 29: Road cross-section taken at the most northern part of the country. The original GCPs are shown in brown. While the randomly elected ones are shown in bright green (Map designed by Author).....	62
Figure 30: Average Nearest Neighbor result summary for the distribution of the GCPs (Graph generated by Author)	62
Figure 31: Overview of SRTM over Lebanon, with two zoomed views of a sample area, each x10 (Map designed by Author)	63

Figure 32: Overview of ASTER over Lebanon, with two zoomed views of a sample area, each x10 (Map designed by Author)	64
Figure 33: A snapshot of 1/20,000 topographic map Batroun Sheet, showing the 10 meters-interval contour lines.....	65
Figure 34: Overview of TOPO over Lebanon, with two zoomed views of a sample area, each x10 (Map designed by Author)	66
Figure 35: Land-Use / Land-Cover classification as per CORINE level 1 (Map designed by Author)	68
Figure 36: Dominant land covers. From upper left: Open Space (32%), Permanent Crops (16%), Arable Land (15%), Forest (13%), Scrubs and Herbaceous Vegetation (12%), and Urban Fabric (5%) (Maps designed by Author).....	70
Figure 37: 'Create Custom Geographic Transformation' tool in ArcGIS 10.2 (Screenshot taken by Author).....	72
Figure 38: Histogram showing the distribution of the Z values from the GCPs (Screenshot taken by Author).....	73
Figure 39: Trend analysis of the GCPs' Z values (Screenshot taken by Author)	73
Figure 40: Normal QQ Plot showing the distribution of the Z values with respect to the normal one (Screenshot taken by Author).....	74
Figure 41: Semivariogram point cloud and surface of the GCPs' Z values (Screenshot taken by Author)	75
Figure 42: Sample height profile between points i and j (Graph generated by Author)	87
Figure 43: Top view showing the local accuracy between a point i and a set of points js (Graph generated by Author)	88
Figure 44: Sample Voronoy decomposition, showing the neighboring points of a point I (Graph generated by Author)	89
Figure 45: Screenshot of the GCPs table of attributes when assessing the SRTM DEM (Screenshot taken by Author).....	90
Figure 46: Screenshot of the GCPs table of attributes, highlighted the resulting ALA values per GCP (Screenshot taken by Author)	91
Figure 47: Sample profile showing the distribution of the GCPs compared to the average accuracy (Graph generated by Author)	92
Figure 48: Sample profile showing the distribution of the GCPs compared to the vario-model (Graph generated by Author)	93
Figure 49: Descriptive Geostatistics of the GCPs' individual ALA for SRTM (From upper-left corner: a- Histogram, b- Normal QQ-PLot, c- General QQ-Plot with respect to the altitude, and d- General QQ-Plot with respect to the LULC level 2 classification) (Screenshots taken by Author)	95
Figure 50: Descriptive Geostatistics of the GCPs' individual ALA for ASTER (From upper-left corner: a- Histogram, b- Normal QQ-PLot, c- General QQ-Plot with respect to the altitude, and d- General QQ-Plot with respect to the LULC level 2 classification) (Screenshots taken by Author)	96
Figure 51: Descriptive Geostatistics of the GCPs' individual ALA for TOPO (From upper-left corner: a- Histogram, b- Normal QQ-PLot, c- General QQ-Plot with respect to the altitude, and d- General QQ-Plot with respect to the LULC level 2 classification) (Screenshots taken by Author)	97
Figure 52: General QQ-Plots comparing the ALA results of the three models (Left to Right: ASTER vs SRTM, ASTER vs TOPO, and SRTM vs TOPO) (Screenshots taken by Author)	98
Figure 53: Graph showing the variation of the average ALA per LULC for the three models (ASTER in orange, SRM in blue, and TOPO in grey) (Graph generated by Author).....	101
Figure 54: SRTM local accuracy vario-model map (Map generated by Author).....	103

Figure 55: ASTER local accuracy vario-model map (Map generated by Author)	103
Figure 56: TOPO local accuracy vario-model map (Map generated by Author)	104
Figure 57: Histogram and Normal QQ-Plot of the SRTM accuracy deviation from the assumed one (Screenshots taken by Author)	105
Figure 58: Histogram and Normal QQPlot of the SRTM accuracy deviation from the RMSE (Screenshots taken by Author)	105
Figure 4-59: Histogram and Normal QQ-Plot of the SRTM accuracy deviation from the vario- model (Screenshots taken by Author)	105
Figure 60: Histogram and Normal QQPlot of the ASTER accuracy deviation from the assumed one (Screenshots taken by Author)	106
Figure 61: Histogram and Normal QQ-Plot of the ASTER accuracy deviation from the RMSE (Screenshots taken by Author)	106
Figure 62: Histogram and Normal QQ-Plot of the ASTER accuracy deviation from the vario- model (Screenshots taken by Author)	107
Figure 63: Histogram and Normal QQPlot of the TOPO accuracy deviation from the assumed one (Screenshots taken by Author)	107
Figure 64: Histogram and Normal QQ-Plot of the TOPO accuracy deviation from the RMSE (Screenshots taken by Author)	108
Figure 65: Histogram and Normal QQ-Plot of the TOPO accuracy deviation from the vario- model (Screenshots taken by Author)	108
Figure 66: Mathematical 2D graph representing the slope (Graph generated by Author) ...	114
Figure 67: Sample 3x3 matrix to calculate the steepest slope for the central cell with the surrounding cell with the maximum height difference (Extract from ArcGIS 10.2 Help)	115
Figure 68: Mathematical 2D graph of the modelled slope and its error (Graph generated by Author)	117
Figure 69: Map showing the extent of the Litany River Basin (Map designed by Author)....	119
Figure 70: Altitude variation within the river basin according to SRTM (Map designed by Author)	120
Figure 71: Profile view of a sample modeled sink	121
Figure 72: Flow Direction derived from SRTM (Map designed by Author)	122
Figure 73: Flow accumulation derived from SRTM (Map designed by Author)	122
Figure 74: Stream ordering based on Strahler Method (Map designed by Author)	124
Figure 75: Derived streams converted into Vector Polylines (Map designed by Author)	124
Figure 76: Mapping River Stream using IKONOS Satellite Image (Screenshot taken by Author)	125
Figure 77: ASTER-derived Streams' Table of Attributes (Screenshot taken by Author).....	126
Figure 78: Generated Euclidian distance for the Lower Litany stream (Screenshot taken by Author)	126
Figure 79: Streams converted from polylines into points (derived from SRTM) (Screenshot taken by Author).....	127
Figure 80: Results of calculating the deviations of the modeled streams (Map designed by Author)	128
Figure 81: Slope accuracy map of SRTM (Map designed by Author)	130
Figure 82: Slope accuracy map of ASTER (Map designed by Author)	130
Figure 83: Slope accuracy map of TOPO (Map designed by Author)	131
Figure 84: Slope significance map for SRTM (Map designed by Author).....	132
Figure 85: Slope significance for ASTER (Map designed by Author)	132
Figure 86: Slope significance map for TOPO (Map designed by Author)	133

Figure 87: Generated river streams (SRTM in magenta, Aster in orange, and TOPO in light green) versus the reference one (in blue) (Screenshot taken by Author).....	136
Figure 88: Histogram and normal QQ-Plot for the deviation of the nodes generated from SRTM (Screenshots taken by Author).....	138
Figure 89: Histogram and normal QQ-Plot for the deviation of the nodes generated from ASTER (Screenshots taken by Author)	138
Figure 90: Histogram and normal QQ-Plot for the deviation of the nodes generated from TOPO (Screenshots taken by Author)	138
Figure 91: General QQ-Plots comparing the deviation to the significance for the three models	140
Figure 92: Normal QQ-Plots of the significance and deviation for the significant nodes derived from SRTM (Screenshots taken by Author)	143
Figure 93: Normal QQ-Plots of the significance and deviation for the significant nodes derived from ASTER (Screenshots taken by Author).....	143
Figure 94: Normal QQ-Plots of the significance and deviation for the significant nodes derived from TOPO	144
Figure 95: General QQ-Plots comparing the deviation to the significance for the three models significant nodes (SRTM, ASTER, and TOPO) (Screenshots taken by Author).....	144
Figure 96: Distribution of the flat areas as per the TOPO DEM (Map designed by Author)	148
Figure 97: Flat nodes derived from ASTER DEM (Map designed by Author)	149
Figure 98: Normal QQ-Plot for the deviations in flat nodes derived from SRTM (Screenshot take by Author).....	149
Figure 99: Scatter plots comparing the slope accuracy to the deviation at the flat node, SRTM, ASTER, and TOPO respectively (Graphs generated by Author)	150
Figure 100: Binary significance model for SRTM DEM (Map designed by Author)	151
Figure 101: Binary significance model for ASTER DEM (Map designed by Author)	152
Figure 102: Binary significance model for TOPO DEM (Map designed by Author).....	152

List of Tables

Table 1: Correlation matrix of the three DEMs.....	66
Table 2: Total area per LULC class, and its percentage from the area of study	69
Table 3: Variation of altitude per level 1 classes	75
Table 4: Variation of altitude per level 2 classification.....	76
Table 5: Variation of GCPs' Altitude per Level 2 Classification.....	77
Table 6: Optimized Range of Altitudes per Level 2 Classification	79
Table 7: Distribution of GCPs per LULC Level 2 Classes	94
Table 8: Pearson Correlation of the Local Accuracy between the three models using SPSS...	99
Table 9: Descriptive Statistics and RMSE of the Model-Generation GCPs' Subset.....	99
Table 10: Descriptive statistics of the GCPs' ALA per LULC Classification	100
Table 11: Pearson Correlation of the average local accuracy between the three models using SPSS	101
Table 12: RMSE values of three model as per the LULC classification	102
Table 13: Summary of the GCPs accuracy deviations for SRTM	105
Table 14: Summary of the GCPs accuracy deviations for ASTER.....	106
Table 15: Summary of the GCPs accuracy deviations for TOPO.....	107
Table 16: Correlation Matrix of the three DEMs	120
Table 17: Summary of Slope Accuracy.....	129
Table 18: Slope Accuracy Correlation Matrix.....	129
Table 19: Summary of Slope Significance	131
Table 20: Slope Significance Correlation Matrix.....	131
Table 21: Correlation Matrix of SRTM datasets.....	134
Table 22: Correlation Matrix of ASTER datasets.....	134
Table 23: Correlation Matrix of TOPO datasets.....	134
Table 24: Summary of segments' lengths, expressed in meters	135
Table 25: Summary of the nodes' deviation, expressed in meters	137
Table 26: Summary of the nodes' deviation, expressed in cell count	139
Table 27: Summary of the nodes' significance	140
Table 28: Summary of the significant nodes' deviation, expressed in meters.....	142
Table 29: Summary of the significant nodes' significance	143
Table 30: Ranges of deviation in cell count for significant and non-significant nodes.....	145

Chapter 1

Introduction

Geographic information system (GIS) is a computerized system to collect, visualize, manage and analyze geographic data (Bolstad, 2005). GIS is technically considered a specialized information system with a focus on the spatial aspect of data, yet its applications spread over a wide range of fields starting from environmental science, for which GIS was originally created, to engineering, urban planning, public administration, homeland security, law enforcement, demography, business, marketing, etc., and with the booming of the Information and Communication Technology (ICT) sector, GIS became more affordable and easier to use, making it available to a wider market, such as non-governmental agencies, activists, and personal users (McKee, 2002), (Keen, 2014).

From an information perspective, a GIS is considered a decision-support system, which is used by experts and decision-makers from various fields to help them model the problem in-hand they are working to solve, and back up their decisions with a systematic and geo-aware solution (Viavettene, Scholes, Revitt, & Ellis, 2008). This solution is generated by the process of spatially-analyzing data, which are by definition a set of raw collected facts, to extract the exact needed information. Spatial analysis is a complex procedure that can be divided into one or more operations. These operations take many forms depending on the GIS software package used, but they can be categorized into one of three main types: filtering data based on tabular criteria, similar to data analysis in a Database Management System (DBMS); filtering data based on their locations; and generating derived datasets from input ones (Mitchell, 1999). Irrespective of what the objective of a spatial analysis procedure is, the 'correctness' of the result depends on using the correct combination of procedures applied on a reliable data. Assuming that the procedure itself, if properly used, will not induce any additional uncertainty, the reliability of the result is highly dependent on the reliability of the input data, and thus the expression: Garbage In – Garbage Out! (Quinlan, 2015). Data reliability is defined by its accuracy.

In general, the accuracy of a certain measurement refers to how close this measurement is to reality. Theoretically, any measurement of any kind needs to have the same value as its reality; however, in real-life, this is impossible to achieve; instead, any measurement contains in its value some uncertainty that differs from its actual real value (Gaines, 2015). The existence of the uncertainty does not make a measurement useless or false, but at the same time, this uncertainty cannot be totally neglected. For a proper use of any measurement or any piece of datum, its level of uncertainty needs to be properly quantified and taken into consideration during any performed analysis in order to fully make use of its derived results.

Since its beginning, GIS was fully based on 2D planimetric data models. These models, both vector and raster, are based on XY tuples to model the geographical part of the data with respect to a Cartesian system. Another alternative reference is the mathematical modelling of Earth's surface, known as the reference ellipsoid, where the Latitude/Longitude tuple is used to uniquely identify the location of the feature (Longley, Goodchild, Maguire, & Rhind, 2005). In both references, the altitude of the features is not taken into consideration. However, we do not live in a flat world, and all features in it have 3D shapes. To better model these features and their phenomena, several years ago GIS developments started incorporating the third dimension. However, unlike graphics and CAD software were, 3D models are fully represented through polyhedral elements. These models are represented as 2.5D models in GIS, where they are represented through base height and extrusion values, expressed in XYZ tuples (Gold, 2003), (Lenk & Heipke, 2006).

An important aspect of representing the real world in 2.5D is modelling the terrain (Koch & Heipke, 2004). A digital elevation model (DEM) is a 3D continuous representation of a terrain surface. In GIS a DEM can be represented in a grid of equally-sized cells, called 'raster,' or in vector surface based on a network of interconnected triangles of various sizes called Triangular Irregular Network (TIN) (Guth, 2006). A DEM as an end product is only useful for visualizing the terrain in perspective view. However, the uses of a DEM extend far beyond 3D visualization to perform various analyses, such as terrain morphology, volumetric analysis, visibility and line of sight, and hydrologic analysis (Herbei, Ciolac, Popescu, & Ciolac, 2010), (Hutchinson, Xu, & Stein, 2011), (Shingare & Kale, 2013).

1.1 Challenges and Motivations

Concerning geographic data, the accuracy can take several forms, such as: positional or attributes accuracy, logical consistency, completeness, etc. More specifically, positional accuracy can be divided into horizontal or planimetric accuracy (XY) and vertical or altimetric accuracy (Z). All these aspects are important and will affect the data and its derived results in different ways depending on the nature of the data in-hand and the different analytical processes performed (Hunter, Bregt, Heuvelink, De Bruin, & Virrantaus, 2009).

The study of accuracy in all its aspects is a very wide topic that cannot be covered with the proper depth within one research project. Instead, this research will focus on the vertical accuracy of digital terrain model (DTM) and its propagation through terrain-based analysis. But why I have chosen this aspect out of all the others? I did so for the following reasons:

- First of all, with the emergence of 3D GIS, there is a high need for 3D data and 3D models to be the basis of various spatial analyses that go beyond the more traditional planimetric data. DTM forms the base GIS layer for such modeling, where it represents the terrain base height and, in some cases, can also be used to determine the height of features. Furthermore, many terrain-based analyses are based on DTMs and result in useful information, such as volumetric calculations, terrain roughness, 3D routes and surfaces, line of sight, etc.
- Second, this model and its related analyses are emerging greatly in various fields, from basic topographic mapping, hydrology and hydraulics, environment, utility engineering, landscaping, urban and rural planning, risk management, military applications to name a few. These fields fall under the public and private sectors and are emerging greatly in developed and developing countries, even if at different levels.
- Finally, the need for such terrain models to cover vast areas world-wide has caused several agencies from around the world to develop various techniques and are now providing the market various products with different specifications and eventually different accuracies. The main advantages of this wide range of products are: availability of data world-wide; availability of different ranges of competitive prices, including several free products; and the ease and speed of procurement. One main disadvantage arises facing these advantages: given the various available products, with their different specifications, performing the same analysis on either product will give a different result. And this raises the main question: Is there one product optimal for all the situations? As the axiomatic answer is no, it raises the actual question: How to know which product is the most suitable to adapt for a given situation or application? Furthermore, does a specific product have the same accuracy spread equally all over the covered area, or does the accuracy vary?

On a personal level, and throughout my academic and professional experience in the field of GIS, I came across several tasks where a DEM was used to extract altitude data, or to be part of a terrain-based analysis. Specifically, I worked as a GIS consultant between 2010 and 2014 in the Litany River Basin Management Support (LRBMS) Program, which was a USAID-funded project with the aim of assisting the Litany River Authority (LRA), a Lebanese governmental administration, in managing the Litany River Basin' resources and services (environmental protection, irrigation, power generation, etc.). During my tenure in this position, I used different DEM datasets to perform various analysis such as delimiting the exact limit of the river basin, detecting areas of potential inundation risks in the upper river basin based on steady flow flooding simulation, detecting areas and time of arrival in the

lower river basin based on 'Qaraoun' dam failure simulation, generating bathymetric survey of the 'Qaraoun' lake bottom, and calculating lake sedimentation and the new lake volume approximately 50 years after the construction of the dam.

At the end of each of the previously mentioned tasks, a challenging question was always raised by the director of the project: What is the accuracy of the result? This repetitive question to which I did not have the appropriate answer at the time triggered my scientific curiosity to research the appropriate answers to the following questions:

- What is the accuracy of given DEM? And how to systematically calculate it?
- Given the accuracy of the DEM, what is the accuracy of the analysis-derived result?
- Over a large coverage area, is the accuracy of a DEM constant, or does it change?
- If the accuracy of a DEM is variable, is the variation random or dependent on some factors?
- And what factors could affect the variability in the DEM accuracy?

1.2 Research Objectives

When using a DEM on a regional or nation-wide scale, a single product cannot have the same accuracy equally in all the various sub-regions due the variation of the modeled terrain (Aguilar, Aguera, Aguilar, & Carvajal, 2005). Furthermore, when GIS users rely on freely-available DEMs on the web, such as Shuttle Radar Topography Mission (SRTM) and Advanced Spaceborne Thermal Emission and Reflection Radiometer Global DEM (ASTER GDEM) among others, they tend to use them without thorough investigation of their accuracy.

The first objective of this study is to analyze the accuracy of the SRTM, ASTER DEMs, and national DEM that is based on the Lebanese army 1/20,000 topographic maps at a level of one country, Lebanon, with reference to ground control points. The analysis will be based on geostatistical techniques, with the aim of demonstrating that the accuracy of these products is not constant and is dependent on the land / cover land cover (LULC). 'Land use' defines how humans are using the land in order to achieve a certain activity, whereas 'land cover' defines the various physical materials that cover the surface of the Earth (Myint, Gober, Brazel, Grossman-Clarke, & Weng, 2011). While land use and land cover are often coupled and many times used interchangeably, there is a major difference between the two: land use refers to the purpose the land serves, while the land cover refers to the actual cover.

Nerveless, this parameter was chosen because it is assumed that areas with LULC share some common characteristics and will affect the accuracy of the DEM similarly (Thakur, Singh, & Suraiya, 2006).

The second objective is to establish a procedure to estimate the average local accuracy per each land classification and propagate these averages into a continuous coverage that represents the variation of the accuracy among the area of study, called vario-model. While the resulting numbers will be tailored to the country of study, the proposed approach is applicable to any area around the globe, given the required input data is provided.

The third objective is to create a procedure for propagating the vario-model along terrain-based analysis in order to estimate the accuracy of the analysis results. Because there could be many analyses done and different results generated, each having different geometry with its accuracy defined in different dimensions, the proposed procedure of propagation will be based on the concept of significance, expressed as a percentage, and represent how much the result is significant and represents reality. This later model will enable the decision-makers to evaluate their results quantifiably before relying on them as a basis for decision-making.

1.3 Work Dissemination

Short versions and related researches that influenced this work have been disseminated through the following publications and proceedings, sorted in chronological ascending order:

Wassim Katerji, Hoda Maalouf (2011). *GIS Based Deployment Strategies of Wireless Sensors Networks for Forest Fire Surveillance*. In proceeding of the 5th International Conference on Information Technology (ICIT 2011). Al-Zaytounah University, Amman – Jordan.

Fadi Karam, Nabil Amacha, Wassim Katerji, Tom Cheng (2013). *Assessing supplemental irrigation needs using remote sensing data of wheat grown under semi-arid climate of the Bekaa Valley (Lebanon)*. In proceeding of ACSAD/GIZ Regional Conference on Rainwater Harvesting as an Option for Adaptation to Climate Change in the MENA Region”. Beirut – Lebanon.

Wassim Katerji, Mercedes Farjas (2014). *DEM Accuracy Patterns in Land-Use/Land-Cover Classification – Case Study of Lebanon*. In proceeding of 5th GEOBIA Conference. Aristotle University of Thessaloniki, Thessaloniki – Greece.

Wassim Katerji, Mercedes Farjas, and Carmen Morillo (2015). *Propagation of DEM Varying Accuracy into Terrain-Based Analysis*. In proceeding of XIII International Conference on Geographic Information Systems. Madrid – Spain.

Wassim Katerji, Mercedes Farjas, and Carmen Morillo (2016). *DEM Accuracy Patterns in Land-Use/Land-Cover Classification*. Open Geosciences. De Gruyter Open. ISSN: 2391-5447. Poland (Approved for Publication)

As well as in the following workshops and presentations:

Wassim Katerji, Hoda Maalouf (2009). Presentation Entitled: *Combined Strategies in Wireless Sensors Deployment for Optimized Forest Fire Surveillance*. The 16th International Conference of the Lebanese Association for the Advancement of Science (LAAS16). Beirut Arab University, Beirut – Lebanon.

Wassim Katerji (2011). Presentation Entitled: *Geographic Information System and Data Accuracy Assessment*. Lebanese Syndicate of Surveyors, Second Scientific Day “Latest Methods in Geomatics”. Lebanese Canadian University, Aintoura – Lebanon.

Wassim Katerji, Miguel Ángel Manso Callejo, and Ramón Alcarria Garrido (2014). Presentation Entitled: *Positional Accuracy in Smart-Phones and its Effect on LBS Applications*. The 20th International Conference of the Lebanese Association for the Advancement of Science (LAAS20). Lebanese University, Hadath – Lebanon.

Wassim Katerji (2014). Workshop Entitled: *Introduction to Open Source GIS for Scientists and Engineers*. Notre Dame University, North Lebanon Campus. Barsa – Lebanon.

Wassim Katerji (2015). Lecture Entitled: *Digital Terrain Models and Analysis of Local and Global Accuracy*. Presented at the course “3D Modelling and Laser Scanning Systems”, part of the Master Degree program in Geodesy and Cartography, at Universidad Politécnica de Madrid – Spain.

Wassim Katerji (2015). Lecture Entitled: *The Application of Multivariate Statistical Methods to Digital Elevation Models*. Presented at the course “Multivariate Statistical Methods”, part of the Master Degree program in Geodesy and Cartography, at Universidad Politécnica de Madrid – Spain.

1.4 Thesis Outline

Following the Introduction, the remainder of the thesis is divided into the following chapters:

Chapter 2: Theoretical Background, reviews and summarizes all the major published research that is related to the topic of the thesis. The aim of this chapter is to summarize the findings of previous works and concepts on which the thesis is built. This research is grouped under the following topics: Sources of DEM, Terrain-Based Analysis, Data Accuracy, Geostatistics, and Land-Use/Land-Cover.

Chapter 3: Data and Methods, represents the first of three chapters that form the core of this thesis. This chapter describes the tools to be used in this research, the area of study, and the input data. A statistical and geostatistical descriptive analysis is also included to better understand the data before using them in chapters 4 and 5.

Chapter 4: The Accuracy Vario-Model, represents the second of the three chapters that form the core of this thesis. This chapter describes the methodology for defining and implementing the proposed vario-model for representing the varying vertical accuracy of a DEM.

Chapter 5: Propagation of the Vario-Model, represents the third of the three chapters that form the core of this thesis. This chapter describes the methodology for propagating the vario-model to the derived surfaces in order to be able to assess the vario-accuracy of the result of a terrain-based analysis. This chapter also explains how to technically implement the procedure of propagating the vario-model in GIS, similar to the previous chapter.

Chapter 6: Conclusion, summarizes the work done at the different stages of this thesis and reviews its findings and the various faced restrictions and limitations. Finally, it introduces potential future works that could be initiated on top of the findings of this thesis.

At the end of this thesis, and beyond the main chapters, three appendices are included as follows:

Appendix A: Data, describe the physical structure of the input and data including a full listing of all the input elements.

Appendix B: Scripts and Models, lists all the developed scripts and models under GIS to implement tools for generating / automating the described procedures in chapters 4 and 5.

Appendix C: Photo Gallery, lists sample photos taken from different areas around Lebanon in order to visually assist in understanding the reality of each LULC class, especially that these classifications slightly change from one country to another.

Chapter 1 Bibliography

- Aguilar, F., Aguera, F., Aguilar, M., & Carvajal, F. (2005). Effects of Terrain Morphology, Sampling Density, and Interpolation Methods on Grid DEM Accuracy. *Photogrammetric Engineering and Remote Sensing* 71(7), 805-816.
- Bolstad, P. (2005). *GIS Fundamentals: A first text on Geographic Information Systems, Second Edition*. Eider Press.
- Gaines, P. (2015, May 24). *Accuracy, Precision, Mean, and Standard Deviation*. Retrieved from Inorganic Ventures: <http://www.inorganicventures.com/accuracy-precision-mean-and-standard-deviation>
- Gold, C. (2003). But is it GIS? *Journal of Geospatial Engineering* 5(2), 11-26.
- Guth, P. (2006). Geomorphometry from SRTM - Comparison to NED. *Photogrammetric Engineering and Remote Sensing* 72(3), 269-277.
- Herbei, M. V., Ciolac, V. H., Popescu, C., & Ciolac, L. (2010). Modelling of Surfaces in Order to Protect them. *Research Journal of Agricultural Science* 42(3), 607-614.
- Hunter, G., Bregt, A., Heuvelink, G., De Bruin, S., & Virrantaus, K. (2009). Spatial Data Quality: Problems and Prospects. In G. Navratil, *Research Trends in Geographic Information Science, Lecture Notes in Geoinformation and Cartography* (pp. 101-121). Berlin: Springer.
- Hutchinson, M., Xu, T., & Stein, J. (2011). Recent Progress in the ANUDEM Elevation Gridding Procedure. *Geomorphometry*, 19-22.
- Keen, S. (2014, March 26). *GIS for Everyone?* Retrieved from Directions Magazine: <http://www.directionsmag.com/entry/gis-for-everyone/390683>
- Koch, A., & Heipke, C. (2004). Semantically correct 2.5D GIS data — The integration of a DTM and topographic vector data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 509-526.
- Lenk, U., & Heipke, C. (2006). The Radial Topology Algorithm - A New Approach for Deriving 2.5D GIS Data Models. *GeoInformatica* 10.4, 447-468.
- Longley, P., Goodchild, M., Maguire, D., & Rhind, D. (2005). Georeferencing. In *Geographic Information Systems and Science, 2nd Edition* (pp. 115-121). John Wiley and Sons.
- Mckee, F. (2002). The Importance of GIS in Modern Society and the effective partnering of Education and Industry. *5th AGILE Conference on Geographic Information Science*. Palma.
- Mitchel, A. (1999). *The ESRI Guide to GIS Analysis, Volume 1*. Esri Press.
- Myint, S., Gober, P., Brazel, A., Grossman-Clarke, S., & Weng, Q. (2011). Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote Sensing of Environment*.
- Quinion, M. (2015, May 24). *Garbage In, Gargbage Out*. Retrieved from World Wide World: <http://www.worldwidewords.org/qa/qa-gar1.htm>
- Shingare, P., & Kale, S. (2013). Review on Digital Elevation Model. *International Journal of Modern Engineering Research* 3(4), 2412-2418.

Thakur, S., Singh, A., & Suraiya, S. (2006). Comparison of Different Image Classification Technique for Land Use Land Cover Classification: An Application in Jabalpur District in Central India. *International Journal of Remote Sensing and GIS* 1(1).

Viavettene, C., Scholes, L., Revitt, D., & Ellis, J. (2008). A GIS based decision support system for the implementation of Stormwater Best Management Practices. *11th International Conference on Urban Drainage* . Edinburgh.

Chapter 2

Theoretical Background

What is a DEM? What are the sources of DEM? What can be done with it? What is its accuracy? What is land-use/land-cover, and does it relate to DEM? These are some questions that this chapter tries to answer by reviewing and summarizing previous works tackling related topics. By the end of this chapter, I will assess the findings and highlight the gaps in the available publications in fully resolving the problem at hand; these gaps introduce the need for a new, optimized solution that will be detailed in later chapters.

A DEM is the general term that represents a continuous surface representation, mainly referred to a raster. However, in practice, this general representation is divided into two main categories which are different from each other and must be analyzed and interpreted differently (Sreedhar, Muralikrishnan, & Dadhwal, 2015). These two categories are:

- **Digital Terrain Model (DTM):** It represents the bare ground surface without any natural or man-made structures on top of it such as trees or buildings. This category of DEM is combined in conjunction with 3D features in order to represent the actual 3D model.
- **Digital Surface Model (DSM):** It represents a draping surface that combines the ground surface and the tops of all the natural and man-made features. The difference between the DTM and the DSM is visualized in figure 1.

Figure 1: DTM vs. DSM

2.1 Field of Research: DEM Accuracy Assessment

In the fields of science and engineering, accuracy of a measurement represents how close the quantity of measurement is to reality. While in common language, accuracy is often interchanged with precision, they mean two different aspects, as precision refers to

repeatability, that is the degree to which repeated measurements of the same phenomena under the same conditions result in the same value (Taylor, 1997). A DEM is a collection of height measurements of a terrain with respect to a reference system. These measurements, like any other measurement in science or engineering, can never be 100% true. Before using a DEM for any analysis or decision-making, its accuracy must be assessed.

2.2 History and Work in Progress

2.2.1 Accuracy Assessment Methodology

According to (Saati, Arefi, Schmitt, & Stilla, 2011), there are several individual factors that result in the accuracy of a DEM. These factors are: the accuracy of the source data, its density, and distribution; characteristics of the terrain to be modeled; and the interpolation technique used to generate the DEM surface. As a general rule, more accurate and denser input points result in a more accurate DEM (Liu, Zhang, Peterson, & Chandra, 2007).

To assess the vertical accuracy of a DEM, there exist three main techniques that rely on statistics: Root Mean Square Error (RMSE), variance, and 95th percentile. RMSE is a statistical measurement calculated as the square root of the mean of the square's errors of a sample. Variance measures how much the error values are spread; a variance equal to zero means that all the samples have the same value, and small variance indicates that the errors are concentrated around the mean value. As for the 95th percentile, it represents the level of confidence that 95% of the sample's error fall within the proposed statistical model.

$$RMSE = \sqrt{\frac{\sum (Z_m - z_g)^2}{n}} \quad (2-1)$$

$$MAE = \frac{\sum |Z_m - z_g|}{n} \quad (2-2)$$

Where:

- Z_m is the modelled height in the DEM
- z_g is the actual height in reality (Derived from a GCP)
- n is the total number of samples

RMSE is often used to model the accuracy instead of the mean average error (MAE); RMSE gives more weight to the samples with high errors, while MAE gives the same weight for all the samples. Because high errors are undesirable, the model used must be more sensitive to those extreme values and reflect them in modeling the overall accuracy, and, therefore, RMSE is more desirable than MAE (Chai & Draxler, 2014).

2.2.2 Accuracy Standards

There exist several national and international standards that define the norms and accuracy limits of DEM products and link it to its usability. One of the national standards that has international impact is the National Geospatial Program Standards and Specifications, which was developed by the United States Geological Survey (U.S. Geological Survey, 2015). As previously defined, the accuracy of a DEM is defined by its RMSE, which must be calculated using a minimum of 28 control points, 20 interiors and 8 on the edges. The standard divides DEM products into three levels and sets a specific standard for each:

- **Level 1:** Includes standard 7.5-minute DEM product in the US, which have a spatial resolution on 1 arc-second or approximately 30 meters or any equivalent DEM from stereo profiling or image correlation. An RMSE less than or equal to 7 meters is desired, with a maximum of 15 meters allowed.
- **Level 2:** Includes DEMs that were generated from hypsographic and hydrographic data collected using photogrammetry or existing maps. An RMSE of one-half contour interval is the maximum allowed, with no errors greater than one contour interval.
- **Level 3:** Similar to Level 2, but includes additional data to represent the terrain more accurately, such as ridge lines and transportation features. An RMSE of one-third contour interval is the maximum allowed, with no errors greater than two-thirds contour interval.

Another ‘continental’ standard is the European Commission’ Infrastructure for Spatial Information in the European Community (INSPIRE), Data Specifications on Elevation (INSPIRE Thematic Working Group Elevation, 2013). This standard is spread among all the European countries that are part of the European Union. The specification document divides the elevation value depending on the geometry of the data:

- **Vector or TIN:** The maximum allowed vertical RMSE is equal to the contour line interval divided by six. In low reliability areas such as dense forests, the maximum RMSE can be increased by 50%.
- **Raster:** The maximum allowed vertical RMSE is equal to the spatial resolution (referred to as Ground Sample Distance in the documentation) of the raster divided by three.

2.2.3 Innovative Methods

While the method of RMSE is considered by far the standard way of assessing the accuracy of a DEM, there exist other alternative solutions proposed in different studies; another aspect of DEM accuracy, beside its height accuracy, is its spatial resolution. According to (Takagi, 1998), spatial resolution affects the accuracy of the slope inclination, but not the aspect, and accordingly not all the applications on terrain-based analysis will be affected much by the resolution. Also, as proven in (Guo-an, Strobl, Jian-Ya, & Mu-dan, 2001), the RMSE was proven to be dependent on the profile curvature and the spatial resolution on the input DEM. Other methods are proposed in different studies, such as the approximation theory presented in (Liu & Sherba, 2012) which is based on the combination of two components: propagation and interpolation errors. This method showed similar results to the previous conventional methods, but without the need for a large number of GCPs. These methods are researched and proposed in most of the times with the main objective of reducing the number of required GCPs, as collecting GCPs represent a time-consuming and costly operation, if not sometimes impossible due to security reasons or impossibility to reach the areas under study.

Also, several studies have tackled the interpretation of the DEM accuracy variation due to the variation in the land cover. For example, studies on ASTER DEM over the mountainous areas in the Andes, showed that the RMSE varied between 15 and 20 meters in the hilly terrain, and 30 meters in mountainous one (Eckert, Kellenberger, & Itten, 2005). Studies of the Alps region in North Italy show a high correlation between the accuracy of the DEM and the slope of the terrain, where the errors were highest in the steepest slopes (Castrignanò, Buttafuoco, Comolli, & Ballabio, 2006). Also, detailed study on LiDAR-based DEM in rangeland landscapes in Canada showed that the errors were over-estimated in forested areas and under-estimated in meadow habitats, and the increase of slope showed an increase in the height error too, while the LiDAR sampling angle had little impact on the errors (Su & Bork, 2006).

2.3 Theoretical Principles

2.3.1 Sources of DEM

Before tackling the different use of DEMs and its derived product, this section reviews the various methods for measuring and collecting height data, which is the basic unit of the DEM. The four major methods of collection are:

- Ground Surveying
- Stereoscopic Photogrammetry
- Radargrammetry
- Laser Scanning and LIDAR

However, in order to better understand the reference upon which the height value is measured, this sub-chapter starts with a definition of the various reference models and the difference between them.

Reference Model

The measured values in a DEM represent the elevation or height of the surface with respect to some reference. Taking the center of Earth as a reference is neither practical nor meaningful, as the majority of the measurement will show relatively close values. While it is commonly assumed that such measurements are taken with respect to the sea level, they are in fact, and as per (Clynch, 2006), measured with respect to one of two types of references: the ellipsoid or the geoid.

In geodesy, Earth is modeled into a mathematical ellipsoid in order to simplify its complex shape and to provide a systematic reference for positioning. Adopting the ellipsoid as the reference for the height values will result in ellipsoidal or geodetic height. However, geodetic heights are not commonly used or found in maps. Instead, the commonly used heights are referenced to the mean sea level (MSL), technically referred to as the orthometric height. Unlike the ellipsoid which is a constant smooth surface, the modeled MSL, which is referred to a geoid, is based on constant gravity potential and varies throughout the planet (Fotopoulos, Kotsakis, & Sideris, 2003).

Figure 2: Difference between ellipsoid and geoid

As shown in the figure above:

- P represents the point on the terrain whose height needs to be measure
- h represents the geodetic height
- H represents the orthometric height

- N represents the undulation of the geoid, which is the difference between the ellipsoid and geoid surfaces, keeping in mind that this value could be positive or negative

Ground Surveying

The first conventional method for collecting topographic information about a terrain was through ground topographic surveying. As defined in (Farjas, 2012), topographic surveying is the science of measuring distances and angles on the terrain and above features, relative to a reference point and height. In practice, using surveying instruments, a surveyor collects several points covering the area of study relative to well-defined reference points, such as from national geodetic network, then re-computes their absolute XYZ values accordingly. The main output of this method is a set of XYZ points that need to be interpolated in order to generate the modeled continuous surface of the terrain. The various methods of interpolation will be explained in the sub-section of interpolation.

This method provides high accuracy results, where the positional and altitude measurements relative to the reference points have accuracies within the range of millimeters, making it the most precise method available so far. In fact, according to (Leica Geosystems, 2015), a typical modern total station can achieve measurements with an accuracy reaching 1.5 millimeters and 2 parts per million for distances of up to 1,500 meters. However, it is time consuming and requires that the surveyors be available with their equipment on the terrain of study, which causes it to be suitable only for relatively small and accessible areas. Instead, newer technologies, such as surveying GPS and 3D Scanners, are taking over conventional ground surveying (Haddad, 2011).

Figure 3: Leveling through topographic survey

With the evolution of Global Navigation Satellite Systems (GNSS) such as the US's NAVSTAR Global Positioning System (GPS), Russia's GLONASS, EU's Galileo, and China's Beidou, which are based on processing satellite signals to process the location, these technologies are also being used for topographic survey, especially given that many recent receivers have the capability to read and interpret the signals of more than one of the previously mentioned systems to compute the location with the highest possible accuracy (Yu, et al., 2014). With the advancement of differential correction and RTK, the positioning accuracy of such technology has reached 1-2 centimeters in the horizontal measurement and 10-20 centimeters in the vertical (Sridevi, 2015). However, it is important to note that the height value measured in such technology is a geodetic height which is referenced to the ellipsoid of WGS84.

Figure 4: GPS satellites and receivers

Stereoscopic Photogrammetry

Photogrammetry is defined as the process of collection geographic features by measuring them manually on a photograph by visually interpreting it (Gomarasca, 2009). Stereoscopic photogrammetry is based on the concept of stereoscopic viewing, where a user looks at one photo through one eye and another photo through the other eye and the two photos show the same area but taken from different perspectives. This simultaneous vision lets the human combine the two images and generates in the brain a 3D vision out of it. With the advancement of computer technologies, the mean of visualizing stereoscopic images via specialized computer transmitters and glasses, and with the integration of ground control points (GCPs), it is possible to measure 3D features (Linder, 2009).

Figure 5: Stereoscopic satellite acquisition

Stereoscopic photogrammetry provides the means to generate a DEM by collecting a set of mass points, either equally or randomly distributed, in the form of X, Y, Z values, combined with breaklines. These mass points will be interpolated later on to generate the continuous surface (Lemaire, 2008). However, using this technique, it is only possible to measure the uppermost Z value of a given feature. For example, if the stereo images show a dense forest, the generated surface will represent the top of the trees and not the actual terrain. In other words, the generated output will be a DSM and not a DTM (Linder, 2009). In fact, as described in (Höhle, 2009), a DTM with an average vertical accuracy of 13 centimeters was reached by automatic photogrammetric DEM generation from large-frame digital aerial camera. While conventionally, GCPs are always needed and must be collected by the technique of ground surveying, which often result in additional cost and time, there are several studies to minimize the number of needed GCPs, such as in (Khoshelham, 2009) where only one GCP, referred to as tie point, is needed per model without sacrificing the quality of the data.

In addition to the capture of stereoscopic aerial photos, there are several satellite sensors that provide stereoscopic satellite images. The study in (Mukherjee, 2012) showed a DEM generation from CARTOSAT-1 stereoscopic satellite imagery, which has a spatial resolution of 2.5 meters. Using additional GCPs to adjust the output, the resulting product was a 10-meter spatial resolution DEM with an average accuracy of 3.6 meters. Such DEM is suitable for map generation up to a scale of 1:25,000. Another example of stereoscopic satellite imagery based DEM is the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation Model (GDEM), which is a DSM with a spatial resolution of 30 meters and an assumed accuracy varying between 10 and 25 meters. The GDEM product has orthometric heights that are based on the EGM96 geoid. (ASTER Home Page, 2004)

Radargrammetry

Radargrammetry is the concept of collecting data through an active sensor that works on basics of echo, where the sensor periodically transmits a high-powered and short-period signal toward the target surface. The system records the backscattered signal and deduces from it the distance based on the time taken for the signal to hit the target and get back. Transmitting and receiving microwaves have a penetrative advantage over the visible spectrum where it does not get blocked by clouds or high air humidity, especially in tropical areas (Gomarasca, 2009).

Figure 6: Capture and processing of SAR image

This technique, as described in (Neumann, Ferro-Famil, & Reigber, 2010), is used in several applications beyond the capacity of stereoscopic photogrammetry, such as detecting tree canopies and estimating the volume of the forests. In fact, the work done using Polarimetric Interferometric Synthetic Aperture Radar (PolInSAR) resulted in measuring the height of the trees with an average accuracy of 5 meters. As reviewed in (Miller, 2010), a second example of radar satellite sensor is the TerraSAR-X, which is a versatile SAR satellite that provides 2D imaging with 1 meter' spatial resolution and 3D imaging with 2 meters' spatial resolution. The SAR antenna transmits at the frequency of 9.65 GHz and a maximum bandwidth of 300 MHz. A third example of radar-based generated available-free DEM is the Shuttle Radar Topography Mission (SRTM), which is a DSM with a spatial resolution of 90 meters and height accuracy of 20 meters. While height measurements based on orbiting satellite are geodetic, the dataset are processed and available to the public with orthometric heights that are based on the EGM96 geoid. (SRTM Homepage, 2009)

Another advance in Radargrammetry is the airborne Interferometric SAR (InSAR) that provides an intermediate level of details between the airborne LiDAR (explained in the next section) and the space-borne SAR sensors. The study in (Bryan Mercer, 2008) showed the

latest advances in airborne InSAR where such technology provides either a DTM or DSM with 5 meters' spatial resolution. Tests made on heavily forested area gave an average accuracy of 30 centimeters.

A major drawback of this technique is that its resulting images are highly affected multiplicative noise called speckle. Speckle reduces the radiometric quality of the captured images, which directly affects the process of feature extraction and scene recognition and reduces their accuracy. However, as described in (Espinoza Molina, Gleich, & Datcu, 2010), there are several proposed de-speckling methods based on Bayesian approach without affecting the structure of the images.

Laser Scanner and LiDAR

Light Detection and Ranging (LiDAR) is one of the latest technologies for measurement. It is a form of active remote sensing that measure distances by transmitting a laser beam on the target feature and analyzing the reflection. LiDAR sensors can be mounted on airplanes, called airborne LiDAR, or on cars, called terrestrial LiDAR, and can reach vertical accuracy higher than 5 centimeters (LiDAR, 2013). Laser scanners work based on the same concept as LiDAR, with the difference being that the output of the first is an image, while the output of the second is a cloud of point measurements. (Gomarasca, 2009)

LiDAR method overcomes stereoscopic photogrammetry as the laser beam penetrates canopies, giving it the possibility to generate a DTM rather than just a DSM. Also, in dense urban areas, LiDAR provides a full coverage as there are no shadows as in imagery. However, the main challenges in using LiDAR are the processing and filtering of the cloud of points in order to generate the 3D model out of it (Liu X. , 2008), (Hollaus & Höfle, 2010)

Figure 7: Airborne laser scanner

The output product is a DEM with a spatial resolution equal to the spacing between the cloud points. The resolution can be estimated as follows:

$$R = \sqrt{\frac{A}{n}} \quad (2-3)$$

Where:

- R: The output spatial resolution (in linear unit, example: meter)
- A: Area of the scanned region (in area unit, example: meter²)
- n: Total number of collected points

LiDAR-generated DEMs have applications in several fields, where a high-level of detail is needed in order to derive sensitive analysis. An example of LiDAR application is modelling details drainage infrastructure, such as road embankment and drainage ditches, in order to efficiently estimate catchment boundaries and water flow paths, as described in (Lindsay & Dhun, 2015). A second example is detecting land cover classification with a high level of detail in dense urban areas; the analysis of the cloud point goes beyond the 3D geometric component to include point density and patterns, waveform, and urban objects recognition to extract the land cover classification as described in (Yeung Yan, Shaker, & Nagwa, 2015).

2.3.2 Surface Interpolation Techniques

After reviewing the major sources of surface height data, this section will review the major methods for interpolating the height data into a continuous surface in cases where the original data are discrete points and lines.

By definition, interpolation mean the technique of estimating the unknown value at a specific location based on the values of available samples (Yang, Chen, & Li, 2012). Interpolation is based on Tobler's first law of geography: *"Everything is related to everything else, but near things are more related than distant things"*. Accordingly, the interpolation of the value at a location will be affected by the points that are close to it more than those that are far.

There exist several interpolators in major GIS applications, commercial applications such as Esri ArcGIS 3D Analyst extension, and open-source such as GRASS, Quantum GIS, gvSIG, and SAGA. Each interpolator has a specific algorithm for processing the samples and most suitable depending on the number and density of the samples, and the nature of the phenomena under interpolation. However, according to (Li & Heap, 2011), out of 32 types of interpolator, there exit only few that are frequently used. As per Esri 3D Analyst documentation (Booth, 2004), the major available interpolators are:

Inverse Distance Weighted (IDW): This interpolation is based on the local influence of the samples with inverse weight depending on its distance from location of interpolation. In other words, this interpolator applies directly the first law of geography. Also, it has a power parameter that allows the user to control the influence of the samples on the interpolation based on their distance from the point to be estimated. If high power value is set, more emphasis will be placed on the nearest samples, which will result in less smooth surface. On the other hand, if low power value is set, more influence will be given to the surrounding and farther samples, and thus resulting in more smooth surface. It is important to note that this method is based on calculating average values from the samples, which will result in a surface that will not pass exactly by the sample points. Also, this technique does not predict hills and valleys in a terrain (Watson & Philip, 1985). This method is most suitable for modeling distance influenced phenomena, such as marketing and retail sites analysis.

Natural Neighbor: This interpolation is similar in algorithm to IDW; however, the difference is that instead of relying on all the samples within a specific radius, it creates voronoi tessellation polygons and selects the closest samples that form a convex hull around the location of interpolation; the weights are then calculated by proportionate areas. The concept behind this algorithm is to perform an interpolation taking into consideration only a subset of the samples that are 'neighboring' the location of calculation. Similar to IDW, this method does not predict hills and valleys, but unlike it, its resulting surface passes through the sample points (Sibson, 1981). This method is suitable when samples are distributed with uneven density and is considered a good general-purpose non-parametric interpolator.

Spline: This interpolation is based on fitting a minimum-curvature surface through the sample points, similar to bending a rubber sheet to pass by all the points. The technique of minimum-curvature is referred to as thin plate interpolation because it ensures a smooth and continuous surface along with continuous first-derivative surfaces. There exist two variations of Spline: regularized and tension. Regularized Spline generates a smooth surface that ranges higher than the highest sample and lower than the lowest one; therefore, the result's range will go outside the boundaries of the samples' range. On the other hand, Tension Spline generates a less smooth surface that forces this surface to remain within the range of the samples points (Franke, 1982). The decision of what variation on Spline to adopt mainly relates to the range of values in the sample points and whether they represent the full range of the phenomena to be modeled or not. This method is most suitable for gradually varying surfaces such as elevations, water depths, and pollution concentrations. However, it is not suitable at all for phenomena with high variations within short distance because the model will extrapolate out beyond the values range in the samples.

Kriging: This is an advanced geostatistical method that generates a surface from a set of scattered sample points based on several statistical models such as autocorrelation. This interpolation is based on the assumption that there exists a distance or directional spatial correlation between the sample points that describe the variation in the surface. It is a multistep process that includes exploratory statistical analysis, variogram modeling, and surface generation. In addition to the surface generation, Kriging provides the capability to estimate the accuracy of the produced surface. There exist two variations of Kriging: ordinary and universal. Ordinary Kriging is the method used by default, and it is based on the assumption that the constant mean of the sample points is unknown. Logically, this is usually the case unless there is a scientific hypothesis that rejects this assumption based on special cases. On the other hand, the Universal Kriging begins its computation based on the assumption that there exists a global trend that orients the data in the area of study. This trend gets calculated through a polynomial function and subtracts it from the sample points so that the autocorrelation can be built with the random errors derived after the subtraction. Universal Kriging is to be used only in special cases where a trend exists and can be scientifically justified (Oliver, 1990). This method is most suitable for data that have spatial correlation such as geology and soil sciences. Geostatistics will be reviewed in details in section 2.3.6

Topo-to-Raster: This method of interpolation was specifically designed to generate DEMs that are hydrologically correct, in the sense that the interpolator will take into consideration the dynamics of surface water and will generate a modeled surface that respect them. This surface can be achieved by enforcing two main constraints: a properly delineated drainage structure, mainly in terms of simulated boundaries of watersheds, and a properly connected drainage network, mainly in terms of representing accurately ridges and streams (Hutchinson M. , 1988). Unlike the previous four interpolators whose input data are only sample points, this method takes points as mass points or sinks; lines as contours or stream; and polygons as boundaries or lakes.

Triangulated Irregular Network (TIN): Unlike the previous five interpolators whose output is in raster format, TIN method as its name implies generates a triangulated surface composed of adjacent Delaunay triangles, each joining three samples to form its 3D face. Triangles could have variable sizes depending on the distribution and density of the samples, which suitable for samples with varying densities. This is an advantage over raster surfaces that are restricted with fixed spatial resolution. This method is most suitable for high resolution DEM representation where 3D linear and polygon features can be incorporated into the model (Hutchinson M. , 1989), (Li & Yang, 2009). Similar to the previous method (Topo-to-Raster), the generation of a TIN surface takes points as mass points; lines as hard-

lines or soft-lines; and polygons as hard-clip, soft-clip, hard-erase, soft-erase, hard-replace, soft-replace, hard-value-fill, or soft-value-fill.

In (Chen & Yue, 2010) and (Erdogan, 2009), several accuracy tests were made to test how representative interpolated DEMs are of reality. Tests were made on the IDW, Spline and Kriging interpolators. These tests showed that even though the number and distribution of the samples and the interpolator play a role in the accuracy of the final product, this product will always have a loss in accuracy compared to the accuracy of the original raw collected data.

2.3.3 Derived Surfaces

With the advancement of 3D GIS, DEM became a very important dataset, not only for height values that it comprises, but for the rich set of derived surfaces that can be generated from and that are critical in terrain-based analysis. This section will cover the different types of derived surfaces, and the next chapter will cover the different applications that are based on DEM and the analysis of its derived surface

As per Esri 3D Analyst documentation (Booth, 2004), the major available derived surfaces are:

Slope: Slope defines the steepest downhill slope for every location in the DEM. The lower the value is, the flatter the terrain is; the higher it gets, the steeper the terrain becomes. The slope is calculated as percent slope or degree of slope as shown in equation 2-2 and 2-3. The calculation of the slope at each location is made via the average maximum technique applied on a 3x3 cell neighborhood matrix (Burrough & McDonell, Principles of Geographical Information Systems, 1998). The slope can be considered a key derived surface upon which many of other terrain-based analysis are based.

$$\text{Percent Slope} = \frac{\text{rise}}{\text{run}} \times 100 \quad (2-4)$$

$$\text{Degree of Slope} = \tan^{-1} \left(\frac{\text{rise}}{\text{run}} \right) \quad (2-5)$$

Aspect: Aspect defines the direction that the steepest slope faces, calculated for every location in the DEM. The aspect is measured counter-clockwise in degrees and its value ranges between 0° (North) till 360° (North again), passing by 90° (East), 180° (South) and 270° (West). Flat surfaces that have a 0-valued slope will have the aspect value -1 (Burrough & McDonell, Principles of Geographical Information Systems, 1998). Aspect is used in many

applications, such as finding the north-facing slopes for best ski runs, calculating the solar illumination for each region, or identifying flat areas for construction.

Figure 8: Aspect directions (Extracted from ArcGIS 10.2 Help)

Curvature: Curvature defines the second derivative of the input terrain surface, or in other words, it is the slope of the slope. Curvature is used to describe the physical characteristics of the terrain in order to better understand its morphology and how it can effect several studies in hydrology, terrain stability and erosion. There are two measurements of curvature: profile curvature and plan curvature. Profile curvature calculates the concavity of the terrain parallel to the orientation of the slope and is mainly used to calculate the acceleration of flow throughout the terrain. Plan curvature calculates the concavity of the terrain perpendicular to the orientation of the slope and is mainly used to calculate the convergence or divergence of the flow (Kimerling, Buckley, Muehrcke, & Muehrcke, 2011).

Figure 9: Curvature diagram (Extracted from ArcGIS 10.2 Help)

Contour Lines: Contour lines define a set of isolines that join together areas of the same height from the DEM. While contour lines are not used for further data analysis, they are very important datasets for visual interpretation of the terrain in cartographic maps, where the slopes, flat to steep, are estimated visually depending on the spacing between the different lines. Also ridges and valleys can be identified in locations where the contour lines converge or diverge. While contour lines are mainly used to represent terrains, they are also used in meteorological maps to represent variation in air pressure, temperature, and precipitation.

Figure 10: Contour lines (Extracted from ArcGIS 10.2 Help)

Viewshed: Viewshed assesses the visibility from one or more observer locations, not toward a specific target, but to all possible visible areas within the DEM coverage. The result not only identifies the visible and non-visible areas, but also, if there is more than one observer, each area will be identified by which of the observers it is visible to (Turner, Doxa, O'Sullivan, & Penn, 2001). Additionally, there are several parameters that affect the calculation of the viewshed: Spot is the altitude location of the observer; offset is the vertical distance to be added to the spot; azimuth is the horizontal angle range; vertical angle is the vertical angle range; and radius is the range of visibility distance.

Figure 11: Viewshed parameters (Extracted from ArcGIS 10. Help)

Hillshade: Hillshade defines a hypothetical illumination of the terrain by setting the azimuth and altitude of the sun. Depending on the terrain variations, the derived surface will have a varying grayish color, simulating the shadows that were caused by these variations. Hillshade provides a 3D visual effect of the terrain under study in a 2D perspective. The main factors that affect the generation of hillshade is the hypothetical location of the sun in the sky; the location of the sun is identified by the azimuth, which is the horizontal angular direction of the sun and measure from the North clockwise from 0 to 360 degree, and is identified by the altitude, which is the vertical direction of the sun and measured from the horizon, at sunset, up to middle, at noon, from 0 to 90 degrees. Beyond visualization,

hillshade is used in illumination analysis as it models shadows, and each location can be modelled whether it perceives the sun or not any time of the day around the year (Burrough & McDonell, *Principles of Geographical Information Systems*, 1998).

Figure 12: Azimuth and altitude of the sun (Extracted from ArcGIS 10.2 Help)

2.3.4 Applications of Terrain-Based Analysis

After reviewing the major derived surfaces from DEM, this section will review the major application the DEM and its derived surfaces, identify what benefits are generated from such products, and discuss how it will affect the decision in several fields. The use of these products extends to a large number of fields, but this section will focus on four major families of applications. Several examples will be given on each of the different affected fields. These four families of application are: terrain morphology, volumetric analysis, visibility and line-of-sight, and hydrologic analysis.

Terrain Morphology

Terrain morphology, or geomorphology, is the science that studies the origin and historical evolution of the terrain and the physical processes that led to these changes (Chorley, Stanley, & David, 1985). Understanding the terrain characteristic for different fields is very important to assess the most suitable sites and to detect abnormal situation, which can all be derived by analyzing the DEM modeling that terrain (Grohmann, Smith, & Riccomini, 2010), (Prakash, Singh, Singh, & Amit Kumar, 2011).

One major application of terrain morphology is slope stability, which is the study of soil-covered slopes and their potential, withstanding or not, when affected by movement. Slope stability is very important to assess because unstable slopes when faced with any moving phenomena will lead to landslides and mass movements. These hazards result in severe physical damages in land, infrastructure, and construction, and in some extreme cases, lead to human causality. Several studies have proposed different methods on the use of GIS and DEMs for slope stability and landslide analysis, and their risk, such as (Xie, Esaki, Zhou, & Mitani, 2003).

Figure 13: 3D view of slope failure leading to landslide (Xie, Esaki, Zhou, & Mitani, 2003)

Volumetric Analysis

Volumetric analysis comprises the calculation of volumes that is generated, provided a fixed plane on top of below a DEM, or deducted as a difference between two DEMs of the same area but taken at different time snapshots (Booth, 2004).

A major application of volumetric analysis is topographic change detection. This change detection enables scientists and decision makers to understand the changes that took place in the formation of the terrain between two different dates and calculate exactly those changes that would lead to different interpretation at the level of geology, environment, and economy. In fact, several studies have interpreted the use of DEM and GIS in topographic change detection in order to assess the volume of resources that were exploited from a given area, such as coal mining activities as reviewed in (Dong, Fu, & Yoshiki, 2008).

Figure 14: Volume calculation between the terrain and a below plane
(Extracted from ArcGIS 10.2 help)

Another application of volumetric analysis is the calculation of water volumes, mainly accumulated in lakes. In this case, the topography of the lake' bottom is required along with a fixed plan representing the surface of the water volume. Alternatively, as described in (Lua, Ouyangab, Wua, Weic, & Tesemmac, 2013), the volume of the lake can be determined by

analyzing a long-term series of multi-spectral remote sensing data and deriving the bottom surface.

Visibility and Line-of-Sight

Visibility analysis focuses on detecting visible/hidden areas with respect to provided observer locations, while line-of-sight (LOS) focuses on a direct line of visibility between two specific points. Visibility analysis has been already introduced in section 2.3 under Viewshed.

A major application of visibility analysis is in the military and homeland security field. In fact, military operations rely on GIS and geospatial technologies in various aspects of operations ranging from mission command to intelligence, surveillance, and mission support (Baumann, 2014). A special focus is on simulating visibility ranges in order to better plan battles operations or optimally distribute surveillance towers for homeland security and counter-terrorism. In such applications, like the one described in (US Naval Research Laboratory, 2015), the analysis is based on high-detailed DTM coupled with accurate 3D mapping of man-made structures such as buildings, bridges, overhangs, etc. in order to model the reality as realistically as possible.

Figure 15: Visibility in military operations (US Naval Research Laboratory, 2015)

Another application more focused on LOS is telecommunication. In high-frequency radio-wave communications and in order for the receiver to get the signal from the sender, a direct LOS is required between the two devices. Furthermore, the Fresnel zone between the two devices should be clear from any obstruction for the communication to take place appropriately, as shown in the below figure. The Fresnel zone has a specific computation that is dependent on the wavelength of the transmitted signal and the distance between the two devices (Haslett, 2008); however, the exact computation is beyond the scope of this study.

GIS is being used extensively and many specialized tools have been developed and are available in the market and dedicated for planning and simulating fixed wireless networks. Functionalities in such tools include: path profiling, links design and management, power budget analysis, interference analysis, among others, and all these functionalities are primarily based on LOS. One example of the tools in the market is Cellular Expert Radio Links (Cellular Expert, 2015).

Figure 16: Fresnel Zone between two communication antennas

Hydrologic Analysis

By definition, hydrology is the science that studies the dynamics, distribution, and physical quality of water. This science covers water in its different aspects such as groundwater, soil infiltration, surface water flow, precipitation and evaporation, water quality, modeling and simulation, prediction, and water transportation (Anderson & McDonnell, 2005). With water becoming scarcer around the world, compared to the continuous increase in demand, and becoming more prone to pollution, there is a worldwide interest in managing and preserving water resources. Thus, at the international conference on water and the environment that took place in 1992, Integrated Water Resources Management (IWRM) was promoted by the Global Water Partnership as a new process focused on the development and management of water resources in order to achieve a sustainable economic and social welfare (Global Water Partnership, 2015).

GIS is a key tool in performing analysis related to hydrological and IWRM studies, and many special packages are available that contain dedicated tools for such work. However, the most common hydrological analyses in GIS are: delineating drainage systems, delineating watersheds, locating water outlets, and calculating the volume of a reservoir, natural or man-made. These types of analyses are always based on a DEM as an input (Hutchinson, Xu, & Stein, 2011), (Cui, Li, Huang, & Huang, 2014).

A drainage system is the collection of connected streams, rivers, and lakes that transmit water from its sources to the destination. Given that water flows by gravity, the formation

and connectivity of the drainage system is mainly affected by the topography of the area. Other factors include soil type, bedrock type, climate, precipitation, and vegetation cover. These factors influence the drainage system patterns (Pidwirny, 2006).

Figure 17: Schematic representation of a sample drainage system

A watershed or drainage basin is an area where all water sources combine and converge to a single point, usually the lowest in elevation in that area. Defining the boundaries of watershed is very important in several fields: political boundaries, where historically, watershed boundaries had an important impact on defining territories, regions, and sometimes boundaries of a country; hydrology, where the drainage basin is the areal unit for defining independent zones of studies; and environmental science, where the boundary of a watershed defines the contamination area of water pollution (Anderson & McDonnell, 2005).

Figure 18: Sample watershed and its components (Extracted from ArcGIS 10.2 Help)

2.3.5 Land-Use / Land-Cover

Land-use defines how humans are using the land in order to achieve a certain activity, whereas land-cover defines the various physical materials that cover the surface of the Earth. While land-use and land-cover are often coupled and many times used interchangeably,

there is a major difference between the two: land-use refers to the purpose the land serves, while the land-cover refers to the actual cover. For example, a recreational area from the land-use can be classified as forest, grassland, or urban vegetation in land-cover (NOAA, 2014).

The most efficient method for generating LULC is analyzing multi-spectral satellite images through remote sensing in conjunction with reference ground control points used to build and validate the generated model, thus the term supervised classification. Within remote sensing, there are several algorithms to perform the classification, mainly: Mahalanobis, Maximum Likelihood and Minimum Distance. However, according to studies done in (Thakur, Singh, & Suraiya, 2010), the Maximum Likelihood algorithm was found to be the most reliable one, with an overall efficiency of 84%.

These algorithms are considered pixel-based classifications, and they work on a pixel-by-pixel basis. This type of classification is the most widely used, and it is mostly suitable for low-to-medium resolution imagery, with a cell size of 15 meters and more. However, with the advancement of remote sensors, high resolution imagery is becoming more widely available at reasonable costs. According to the research done in (Myint S. , 2011), pixel-based classification gave an overall accuracy of 67.6% on high resolution images. To overcome this low accuracy, a new family of classification was developed, called object-based classification, which detects and interprets objects that are formed of neighboring pixels sharing similar characteristics; this new classification gave an overall accuracy of 90.4%.

There are many ways to create LULC, depending on the area of study, the method of data collection and classification, and the intended use of this dataset. However, in order to work with consistent data that can be easily comparable to other datasets that are generated by other professionals or agencies, a standard classification must be adopted. Unfortunately, no single universal standard exists so far to be used by everybody worldwide. Instead, each country or union of countries has their own standard. This section will review two of the most widely known and used LULC standards worldwide. These are used in the European Union and the United States of America, respectively. Finally, a third standard is introduced, the Lebanese one; although not as well known or wide-spread as the first two, it is the one upon which the data of the area of study is based.

Europe' CORINE

The Coordination of Information on the Environment (CORINE) is a European commission program that was initially developed between 1985 and 1990 to have a land cover database that is unified among Europe in terms of nomenclatures and methodology of generation. After the establishment the European Environment Agency (EEA) by the European Council

and adopting the responsibility of managing and updating CORINE database, CORINE was reviewed and a new update called CORINE 2000 was released in 2000 (European Commission Joint Research Center, 2002).

The Infrastructure for Spatial Information in the European Community (INSPIRE) Directive is a European project to establish a common infrastructure for spatial information throughout Europe. This project tackles 34 data themes, out of them the LULC (INSPIRE Directive, 2014). INSPIRE adopted CORINE classification based on “ISO 19131 Geographic information – data product specification” standard in 2007. It provides multilingual support, open structure to be used by a numerous number of GIS. The data is available in two data structures: vector polygon and raster, both connected to the CORINE code list. The CORINE land cover database consists of three levels of classifications divided as follows: five classes in the first level, which divide into 15 classes in the second level, which finally divide into 44 classes in the third level (European Comission, 2013).

United States’ NLCD

With the ongoing need for data manipulation and analysis on digital geographic data, mainly through GIS, USGS National Mapping Division generated several documents that establish digital cartographic standards, including the land use / land cover database for the entire United States (Robin Feageas, 1983). In 1976, LULC map level 1 was established by Anderson, followed by level 2 in 1983, and was the basis for any potential future level 3 or level 4 classifications. Level 1 is divided into 9 classes, which further divide into 37 classes in level 2.

In 1992, a federal consortium took place to initiate a first-generation National Land Cover Database (NLCD), which was followed by another consortium to establish an updated second-generation NLCD in 2001 with an estimation for completion in 2006 back then. NLCD was developed on top of the previous versions of LULC maps with the intent to have a consistent standard among all the different states in the US. NLCD consists of 29 classes (Homer, Huang, Yang, & Coan, 2004).

Lebanon’ SDATL

The National Physical Master Plan of the Lebanese Territory, or in its original French name “Schéma Directeur d'Aménagement du Territoire Libanais” (SDATL), is a nation-wide master plan with a dataset of GIS layers (vector and raster) that form the base for any small scale mapping. This project was produced by the Lebanese Council of Development and Reconstruction (CDR) and the final version was published on December 2005 (Dar Al-Handasah Consultants, 2005).

One of the produced GIS data is the land use coverage, which consists of four levels of classifications, divided as follows: 7 classes in the first level, which divide into 22 classes in the second level, which further divide into 58 classes in the third level and which finally divide into 114 classes in the fourth level.

After comparing the different standards, SDATL was found very similar to CORINE, the European standard. This similarity is logical as Lebanon is a Mediterranean country like the countries in southern Europe, which all share relatively similar nature and geography, as opposed to the geography of the United States of America. The only major difference between SDATL and CORINE is that SDATL has four levels of detail, while CORINE has only three. However, when reviewing the SDATL list of classes, the majority of the level 3 classes do not have any further classification at level 4. Those which actually have in level 4; their classifications do not provide any added value to the topic under study. Therefore, level 4 classifications have been dropped, and the SDATL-based dataset has been converted to meet the CORINE specifications.

2.3.6 Geostatistics

Statistics is a branch of mathematics that studies data quantitatively, including the collection, analysis, interpretation, and visualization of that data. Statistics can be applied on the whole population of the data. However, in the majority of the cases, it is impossible to collect the whole population of any phenomena to study it; instead, a sample subset is collected and then generalizations that can be applied on the whole population are deduced (Dodge, et al., 2006). Statistics is made up of two methodologies: descriptive and inferential statistics. Descriptive statistics focuses on quantitatively summarizing a collection of data by measuring their central tendency and dispersion. It can be measures via minimum, maximum, mean, median, mode, variance, standard deviation, kurtosis, and skewness. Inferential statistics focuses on deducing properties on the data under study related to the relationships between the different subsets that form the data. This methodology includes testing hypothesis and deriving estimates.

Statistics is applied in any analytical study including analyzing geographical data. However, conventional statistics does not take into consideration the geographical aspect of the data, and therefore omits it, as if the data do not have any geographical or spatial part. However, there exists one branch of statistics, called geostatistics which is more oriented toward geographic data and takes that geographic aspect when performing the statistical analysis (Isaaks & Srivastava, 1989). Geostatistics has the major advantage over statistics because in

the latter, all the samples have the same weight during analysis, while in the former, the samples are stratified according to their geographic distribution.

Currently, there are many fields applying geostatistics such as gas and oil, geology, hydrogeology, hydrology, meteorology, oceanography, geography, forestry, environmental science, landscaping, agriculture, etc. Irrespective of the field, the basics of geostatistics lie in the concept of extending statistical techniques to become based on random function theory that will take into consideration the spatial distribution estimation to model uncertainty and deduce a generalization. Geostatistics extend the interpolation techniques that are taken from spatial analysis (explained in section 2.2), such as IDW, bilinear interpolation, and nearest neighbor.

Geostatistics is highly associated with GIS. In many GIS packages, there exists a geostatistical plug-in or extension, such as the ArcGIS Geostatistical Analyst. This merger between GIS and geostatistics arose toward the late 90s when GIS began to be seen as an analytical tool beyond digital mapping (Burrough, GIS and Geostatistics: Essential Partners for Spatial Analysis, 2001). According to (ESRI, 2001), the Geostatistical Analyst was first introduced in ArcGIS 8.1, with the objective of providing advanced users with a rigorous set of tools for performing spatial data exploration, modeling and predicting errors, and optimal surface generation using Kriging.

Components of Geostatistics

According to (Webster & Oliver, 2007), the major components of performing geostatistical analysis are: generalization, description, interpretation, and control.

Generalization: As previously mentioned in the introduction on statistics, only samples of the population are collected for data analysis and considered as the support of the hypothesis. Sometimes few samples are taken; other times more. However, even if many samples are taken, they will always represent a very small portion of the reality to be studied. Therefore, there will always be an inevitable error present in any study; however, by quantifying the spatial autocorrelation that exists between the samples and generalizing it to the required scale for the study, it is possible to minimize that error and estimate it too.

Description: Similar to descriptive statistics, descriptive geostatistics includes the calculation of means, medians, modes, variances, skewness, frequency distribution, etc. and additionally focuses on calculating the spatial autocorrelation, where the variance is estimated at different distance and direction intervals from the samples. Accordingly, the variation of the covariance can be visualized as a graph with respect to the sample spacing.

Interpretation: Once the description variogram is estimated, the next component in geostatistics is to interpret it. The shape of the points in the variogram explains much about how the phenomena is changing and how adequate the sampling count and distribution was. Computing variograms in different directions help in understanding if the phenomena behave similarly in all directions (isotropy) or is dependent to a specific orientation (anisotropy); this helps in identifying some of the controlling factors that affect the phenomena under study.

Figure 19: Directional analysis of the semivariogram (Extracted from ArcGIS 10.2 Help)

Control: Control is a special component, mostly used when analyzing time-series data. Controlling the processing of analyzing a phenomenon does not mean changing its spatial characteristics because they are based on reality itself; instead, control refers to the modification of the analyst interpretation and response to the results.

Applications of Geostatistics

Geostatistics is used in many fields to predict complex models that fit beyond simplistic linear models. Next are some diverse examples on its use in the various fields:

Geostatistical modeling was used to predict the accuracy of soil mapping in areas away from the samples; two approaches were used and cross-validation of them showed that the multi-Gaussian kriging approach performed better in predicting the uncertainty (Goovaerts, 2001). Another example where geostatistics is used in environmental risk assessment, where two fundamental contributions were made: the spatial distribution of pollutant was quantitatively described along with the estimates of the exposure points. The second contribution was the assessment of the exposure point accuracy especially that the study was built on top of a small set of sample points (Thayer, Griffith, Goodrum, Diamond, & Hassett, 2003).

A specific application of geostatistics that is of most interesting in this thesis is the use of geostatistical methods for assessing the accuracy of a DEM. In fact, several studies, such as (Castrignanò, Buttafuoco, Comolli, & Ballabio, 2006) and (Hengl, Bajat, Blagojevic', & Reuter, 2008), have proposed several stochastic approaches using geostatistics in order to estimate the variation and distribution of the accuracy compared to the conventional method of calculating a single RMSE for the whole area of study. Even some studies, such as (Paredes-Hernandez, Tate, Tansey, Fisher, & Salinas-Castillo, 2010), went beyond assessing the accuracy of a DEM to enhancing it using geostatistical conflation techniques. In this example, two methods were used: ordinary co-kriging and Kriging with and External Drift. Both methods resulted in an increase in the DEM accuracy, but the later method produced the most accuracy.

2.4 Summary

In this chapter, the first reviewed topic was the different technique for collecting height data. These data will be used to generate the DEM. Four major techniques were discussed: ground surveying, stereoscopic photogrammetry, radargrammetry, and laser scanner and LiDAR. With the development and evolution of the last three techniques, the first method is not used anymore for collecting mass 3D data due to its high cost and time consuming operation. However, this technique still provides the most accurate results with a positional accuracy of 1-2 centimeters, so it is still being used to collect GCPs to be used for verification and adjustment. The only other technique that can reach similar accuracy is LiDAR, where recent advancement had led to results with a positional accuracy of 5 approximately centimeters. On the other hand, when the accuracy tolerance is a few meters and the areas to be covered are much bigger, stereoscopic photogrammetry and radargrammetry are still considered the most suitable and affordable, especially in extreme cases where the whole planet is the area of study, such as the scope of work of NASA's SRTM project.

The output data from the different data collection techniques are discrete features, represented as 3D mass points and breaklines; a continuous surface can be generated from these data using interpolation. Six different interpolators were reviewed: IDW, Natural Neighbor, Spline, Kriging, Topo-to-Raster, and TIN. These interpolators are used to generate surfaces for any continuous phenomena. Out of those, the Topo-to-Raster and TIN are the most appropriate ones to generate terrain surfaces, as they support different types of input data beyond sample points, and their algorithms are focused toward terrains.

As previously noted, the use of DEM extends beyond 3D visualization to performing various types of analysis. In performing p analysis, not only is the height of the surface of interest, but also several other terrain characteristics which can be derived from a DEM, such as the slope, aspect, curvature, contour lines, viewshed, and hillshade. Other characteristics can also be derived from these derived surfaces and are specifically defined and generated depending on the field of application.

The fields of DEM application are numerous and expanding more and more with the expansion of the use of GIS and geospatial technologies. In this chapter, four families of applications were reviewed: terrain morphology, volumetric analysis, visibility and line-of-sight, and hydrologic analysis. Examples of real-life applications include slope stability and landslide hazard assessment; topographic change detection; natural resources estimation including water reservoirs; military and homeland security operations including mission command, intelligence, surveillance, and mission control; planning and design of telecommunication infrastructure; and hydrological analysis including IWRM.

On a slightly different topic, LULC was defined and the difference between land-cover and land-use was presented. Methods of generating LULC datasets and their input data were also reviewed, where multi-spectral satellite are the most common data input and studies have shown that Maximum Likelihood algorithm generate the most reliable output. Also, other studies have noted that recent object-based classification give better results than the conventional pixel-based ones. As for the classification itself, there exist several national and international standards to be adopted. In this research, two international standards were reviewed, the European CORINE and the US NLCD. These two were chosen among others because they represent big areas worldwide, and several other countries have had their national standards influenced by one of these. Given that this PhD thesis is focused on a study area in Lebanon, the Lebanese SDTAL standard was also presented; comparing it to the previous two international ones, SDATL is much closer to CORINE rather than NLCD. The result is logical because taking the geographical location of Lebanon into consideration as compared to the geographical areas covered by both standards shows that Lebanon is closer to the European continent and shares similar landscapes, especially to the southern Mediterranean part of it.

Back to the topic of DEM, and in order use it and rely on it in decision-making, its accuracy must be properly calculated. Accordingly, the conventional method for calculating the accuracy was presented, denoted as RMSE, and compared to other innovative ways that take into consideration the different types of land, the spatial resolution of the DEM, and other terrain characteristics. Irrespective of the method to be adopted, and in order for the DEM to be usable, its accuracy must not exceed a certain threshold; this threshold is defined within

mapping standards. Two standards were presented: the USGS' National Geospatial Program Standards and Specifications, and the European INSPIRE. The two standards are defined differently and do not lend themselves to a straightforward comparison, but taking a subset, such as the USGS' Level 3 and INSPIRE' Vector or TIN, shows that the INSPIRE threshold is two times as high as the best one in USGS.

In the last part, geostatistics was introduced and shown it how extends statistics to take into consideration the location and distribution of the samples when performing descriptive or inferential analysis. Also, the different components of geostatistics were discussed: generalization, description, interpretation, and control. Finally, examples were given on the different use of geostatistics in various fields such as soil mapping and environmental risk assessment. Another special case was the use of geostatistics to assess the accuracy of DEM, especially where there are only few sample points available; Studies have shown that this technique resulted in representative results, and even some studies used this technique to enhance the accuracy of the DEM and generate a better, a more accurate version of it.

As a final conclusion, after performing all this research and compiling it into this chapter, it can be concluded that a DEM is not one standard product worldwide: there are many techniques to develop it using different inputs of data. Also, many users from various fields of work analyze it and interpret it differently depending on their application needs. Their applications, as shown in the different previous examples, require highly-sensitive decision making. On the other hand, while there are many studies that proposed different techniques for assessing the accuracy of the DEM before using it, no previous study was found that systematically defines a methodology for assessing the varying accuracy of a DEM and propagates this accuracy to reflect the one of the derived products so that the user of the analysis has the accuracy of the derived product rather than the accuracy of the input DEM only.

Chapter 2 Bibliography

- Anderson, M., & McDonnell, J. (2005). *Encyclopedia of hydrological sciences*. New Jersey: Wiley.
- ASTER Home Page*. (2004, September 7). Retrieved from NASA Jet Propulsion Laboratory: <http://asterweb.jpl.nasa.gov/>
- Baumann, J. (2014, April 7). *Esri Geospatial Technology Speeds Analysis for US Army*. Retrieved from ESRI: <http://www.esri.com/esri-news/releases/14-2qtr/esri-geospatial-technology-speeds-analysis-for-us-army>
- Booth, S. B. (2004). *ArcGIS 9: Using ArcGIS 3D Analyst*. Esri Press.
- Bryan Mercer, Q. Z. (2008). Recent Advances in Airborne InSAR for 3D Applications. *The International Archives of the Photogrammetry, Remote Sensing, and Spatial Information Systems, Vol. XXXVII Part B1*.
- Burrough, P. (2001). GIS and Geostatistics: Essential Partners for Spatial Analysis. *Environmental and Ecological Statistics* 8, 361-377.
- Burrough, P., & McDonell, R. (1998). *Principles of Geographical Information Systems*. New York: Oxford University Press.
- Castrignanò, A., Buttafuoco, G., Comolli, R., & Ballabio, C. (2006). Accuracy assessment of digital elevation model using stochastic simulation. *7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, 490-498.
- Cellular Expert. (2015, May 30). *Products*. Retrieved from Cellular Expert: <http://www.cellular-expert.com/node/206>
- Chai, T., & Draxler, R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? *Geoscientific Model Development* 7, 1247-1250.
- Chen, C., & Yue, T. (2010). A Method of DEM Construction and Related Error Analysis. *Computers & Geosciences* 36, Elsevier, 717-725.
- Chorley, R., Stanley, A. S., & David, S. (1985). *Geomorphology*. London: Methuen.
- Clynch, J. (2006). *Note on Heights Types used in Geodesy*.

- Cui, L., Li, Y. P., Huang, G. H., & Huang, Y. (2014). Effects of digital elevation model resolution on topography-based runoff simulation under uncertainty. *Journal of Hydroinformatics Vol 16 No 6*, 1343-1358.
- Dar Al-Handasah Consultants. (2005). *National Physical Master Plan of the Lebanese Territory – Final Report*. Dar Al-Handasah Consultants.
- Dodge, Y., Cox, D., Commenges, D., Davison, A., Solomon, P., & Wilson, S. (2006). *The Oxford Dictionary of Statistical Terms, 6th Edition*. Oxford University Press.
- Dong, Y., Fu, B., & Yoshiki, N. (2008). DEM generation methods and applications in revealing of topographic changes caused by coal mining activities. *International Conference on Earth Observation Data Processing and Analysis*. Wuhan, China: SPIE.
- Eckert, S., Kellenberger, T., & Itten, K. (2005). Accuracy assessment of automatically derived digital elevation models from aster data in mountainous terrain. *International Journal of Remote Sensing* 26(9), 1943-1957.
- Erdogan, S. (2009). A Comparison of Interpolation Methods for Producing Digital Elevation Models at the Field Scale. . *Earth Surface Processes and Landforms*. John Wiley & Sons, 366-376.
- Espinoza Molina, D., Gleich, D., & Datcu, M. (2010). Gibbs Random Field Models for Model-Based Despeckling of SAR Images. *IEEE Geoscience and Remote Sensing Letters* 7(1), 73-77.
- ESRI. (2001). *ArcGIS Geostatistical Analyst: Statistical Tools for Data Exploration, Modeling, and Advanced Surface Generation* . Redlands, California: ESRI White Paper.
- European Commission. (2013). *D2.8.II.2 INSPIRE Data Specification on Land Cover – Draft Technical Guidelines*. European Commission.
- European Commission Joint Research Center. (2002). *Corine land cover update 2000 – Technical guidelines*. European Environment Agency.
- Falkner, E., & Morgan, D. (2002). Map Accuracy Standards. In *Aerial Mapping: Methods and Application, Second Edition*. Lewis Publishers.
- Farjas, M. (2012). *La Topografía y Sus Métodos: Principios de Investigación*. Madrid: Denboraren Argia.

- Fotopoulos, G., Kotsakis, C., & Sideris, M. (2003). How Accurately Can We Determine Orthometric Height Differences from GPS and Geoid Data? *Journal of Surveying Engineering*, 1-10.
- Franke, R. (1982). Smooth Interpolation of Scattered Data by Local Thin Plate Splines. *Computer and Mathematics with Applications* 8(4), 273-281.
- Global Water Partnership. (2015, May 30). *History of GWP*. Retrieved from Global Water Partnership: <http://www.gwp.org/en/About-GWP/History/>
- Gomarasca, M. (2009). *Basics of Geomatics*. Springer.
- Goovaerts, P. (2001). Geostatistical modelling of uncertainty in soil science. *Geoderma* 103, 3-26.
- Grohmann, C., Smith, M., & Riccomini, C. (2010). Multiscale Analysis of Topographic Surface Roughness in the Midland Valley, Scotland. *IEEE Transactions on Geoscience and Remote Sensing*.
- Guo-an, T., Strobl, J., Jian-Ya, G., & Mu-dan, Z. (2001). Evaluation on the accuracy of digital elevation models. *Journal of Geographical Sciences* 11(2), 209-216.
- Haddad, N. (2011). From ground surveying to 3D laser scanner: A review of techniques used for spatial documentation of historic sites. *Journal of King Saud University - Engineering Sciences*, 109-118.
- Haslett, C. (2008). *Essentials of radio wave propagation*. Cambridge University Press.
- Hengl, T., Bajat, B., Blagojevic', D., & Reuter, H. (2008). Geostatistical modeling of topography using auxiliary maps. *Computers & Geosciences* 34, 1886-1899.
- Höhle, J. (2009). DEM Generation Using a Digital Large Format Frame Camera. *Photogrammetric Engineering & Remote Sensing* 75(1), 87-93.
- Hollaus, M., & Höfle, B. (2010). Terrain Roughness Parameters From Full-Waveform Airborne Lidar Data. *International Archives of Photogrammetry and Remote Sensing* 38(7), 287-292.
- Homer, C., Huang, C., Yang, L. W., & Coan, M. (2004). *Development of a 2001 National Land-Cover Database for the United States*. USGS.
- Hutchinson, M. (1988). Calculation of hydrologically sound digital elevation models. *Third International Symposium on Spatial Data Handling*. Sydney.

- Hutchinson, M. (1989). A new procedure for gridding elevation and stream line data with automatic removal of spurious pits. *Journal of Hydrology* 106, 211-232.
- Hutchinson, M., Xu, T., & Stein, J. (2011). Recent Progress in the ANUDEM Elevation Gridding Procedure. *Geomorphometry*, 19-22.
- INSPIRE Directive. (2014, April 1). Retrieved from European Commission: <http://inspire.ec.europa.eu/>
- INSPIRE Thematic Working Group Elevation. (2013). *D2.8.II.1 INSPIRE Data Specification on Elevation – Technical Guidelines*. European Commision Joint Research Centre.
- Isaaks, E., & Srivastava, R. (1989). *An Introduction to Applied Geostatistics*. New York: Oxford University Press.
- Khoshelham, K. (2009). Role of Tie Points in Integrated Sensor Orientation for Photogrammetric Map Compilation. *Photogrammetric Engineering & Remote Sensing* 75(3), 305-311.
- Kimerling, A., Buckley, A., Muehrcke, P., & Muehrcke, J. (2011). Surface Analysis. In *Map Use: Reading Analysis Interpretation, Seventh Edition* (pp. 346-365). Esri Press.
- Leica Geosystems. (2015, May 27). *Leica FlexLine TSo2/06/09*. Retrieved from Leica Geosystems: http://www.leica-geosystems.us/en/Leica-FlexLine-TSo20609_72053.htm
- Lemaire, C. (2008). Aspects of the DSM Production with High Resolution Images. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 37(4), 1143-1146.
- Li, J., & Heap, A. (2011). A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors. *Ecological Informatics* 6(3), 228-241.
- Li, Y., & Yang, L. (2009). Based on Delaunay Triangulation DEM of Terrain Model. *Computer and Information Science* 2(2), 137-140.
- LiDAR. (2013, January 22). Retrieved from NOAA National Ocean Service: <http://oceanservice.noaa.gov/facts/lidar.html>
- Linder, W. (2009). *Digital Photogrammetry, a Practical Course*. Springer.
- Lindsay, J., & Dhun, K. (2015). Modelling surface drainage patterns in altered landscapes using LiDAR. *International Journal of Geographical Information Science* 29(3), 397-411.

- Liu, X. (2008). Airborne LiDAR for DEM Generation: Some Critical Issues. *Progress in Physical Geography* 32(1), 31-49.
- Liu, X., & Sherba, J. (2012). Accuracy Assessment of Lidar-Derived DEM Based on Approximation Theory. *ASPRS 2012 Annual Conference*. Sacramento, California: ASPRS.
- Liu, X., Zhang, Z., Peterson, J., & Chandra, S. (2007). The effect of LiDAR data density on DEM accuracy. *International Congress on Modelling and Simulation: Land, Water and Environmental Management: Integrated Systems for Sustainability*, (pp. 10-13). Christchurch, New Zealand.
- Lua, S., Ouyangab, N., Wua, B., Weic, Y., & Tesemmac, Z. (2013). Lake water volume calculation with time series remote-sensing images. *International Journal of Remote Sensing* 34(22), 7962-7973.
- Miller, W. P. (2010). The TerraSAR-X Satellite. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 48 No.2.
- Mukherjee, S. (2012). Digital Elevation Model Generation and Retrieval of Terrain Attributes using CARTOSAT-1 Stereo Data. *International Journal of Science and Technology*, Vol. 2 No. 5.
- Myint, S. (2011). Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote Sensing of Environment*.
- Neumann, M., Ferro-Famil, L., & Reigber, A. (2010). Estimation of Forest Structure, Ground, and Canopy Layer Characteristics From Multibaseline Polarimetric Interferometric SAR Data. *IEEE Transactions on Geoscience and Remote Sensing* 48(3), 1086-1104.
- NOAA. (2014, February 24). *What is the difference between land cover and land use?* Retrieved from National Oceanic and Atmospheric Administration: <http://oceanservice.noaa.gov/facts/lclu.html>
- Oliver, M. (1990). Kriging: A Method of Interpolation for Geographical Information Systems. *International Journal of Geographic Information Systems* 4, 313-332.
- Paredes-Hernandez, C., Tate, N., Tansey, K., Fisher, P., & Salinas-Castillo, W. (2010). Increasing the Accuracy of Digital Elevation Models by Means of Gestatistical Conflation. *Accuracy 2010 Symposium*, (pp. 113-116). Leicester, UK.

- Pidwirny, M. (2006). The Drainage Basin Concept. In *Fundamentals of Physical Geography, 2nd Edition*.
- Prakash, D., Singh, C., Singh, P., & Amit Kumar, D. (2011). GIS Based Terrain Modelling of the Area NW of Karimnagar, Southern India. *International Journal of Remote Sensing Applications* 1(2), 6-10.
- Robin Feageas, R. C. (1983). *Land Use and Land Cover Digital Data*. USGS.
- Saati, A., Arefi, H., Schmitt, M., & Stilla, U. (2011). Statistically robust detection and evaluation of errors in DTMs. *Urban Remote Sensing Event* (pp. 305-308). Munich: IEEE.
- Sibson, R. (1981). A Brief Description of Natural Neighbor Interpolation. In *Interpolating Multivariate Data* (pp. 21-36). New York: John Wiley & Sons.
- Sreedhar, M., Muralikrishnan, S., & Dadhwal, V. (2015). Automatic Conversion of DSM to DTM by Classification Techniques Using Multi-Date Stereo Data from Cartosat-1. *Indian Society of Remote Sensing*.
- Sridevi, J. (2015, May 27). *Advances in GPS/GNSS data processing*. Retrieved from Coordinates: <http://mycoordinates.org/advances-in-gpsgnss-data-processing/>
- SRTM Homepage*. (2009, June 17). Retrieved from NASA Jet Propulsion Laboratory: <http://www2.jpl.nasa.gov/srtm>
- Su, J., & Bork, E. (2006). Influence of Vegetation, Slope, and Lidar Sampling Angle on DEM Accuracy. *Photogrammetric Engineering & Remote Sensing* 72(11), 1265-1274.
- Takagi, M. (1998). Accuracy of Digital Elevation Model According to Spatial Resolution. *ISPRS Commission IV Symposium on GIS - Between Visions and Applications*. Stuttgart: ISPRS.
- Taylor, J. (1997). *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*. University Science Books.
- Thakur, S., Singh, A., & Suraiya, S. (2010). Comparison of Different Image Classification Technique for Land Use Land Cover Classification: An Application in Jabalpur District in Central India. *International Journal of Remote Sensing and GIS, Volume 1*.
- Thayer, W., Griffith, D., Goodrum, P., Diamond, G., & Hasset, J. (2003). Application of Geostatistics to Risk Assessment. *Risk Analysis* 23(5), 945-960.

- Turner, A., Doxa, M., O'Sullivan, D., & Penn, A. (2001). From isovists to visibility graphs: a methodology for the analysis of architectural space. *Environment and Planning B: Planning and Design* 28(1), 103-121.
- U.S. Geological Survey. (2015, February 5). *National Geospatial Program Standards and Specifications*. Retrieved from The National Map: <http://nationalmap.gov/standards/>
- US Naval Research Laboratory. (2015, May 30). *Sniper-RT*. Retrieved from US Naval Research Laboratory - Technology Transfer Office: <http://www.nrl.navy.mil/techtransfer/available-technologies/IT/sniper-rt>
- Watson, G., & Philip, D. (1985). A Refinement of Inverse Distance Weighted Interpolation. *Geoprocessing* 2, 315-327.
- Webster, R., & Oliver, M. (2007). Introduction. In *Geostatistics for Environmental Scientists, Second Edition* (pp. 1-8). West Sussex: John Wiley & Sons.
- Xie, M., Esaki, T., Zhou, G., & Mitani, Y. (2003). Geographic Information Systems-Based Three-Dimensional Critical Slope Stability Analysis and Landslide Hazard Assessment. *Journal of Geotechnical and Geoenvironmental Engineering*, 1109-1118.
- Yang, K., Chen, Y.-M., & Li, M.-C. (2012). Accumulated similarity surface for spatial interpolation. *20th International Conference on Geoinformatics* (pp. 1-4). Hong Kong: IEEE.
- Yeung Yan, W., Shaker, A., & Nagwa, E.-A. (2015). Urban land cover classification using airborne LiDAR data: A review. *Remote Sensing of Environment* 158, 295-310.
- Yu, M., Bourne, H., Carroll, M., Yu, J., Taylor, S., Wang, J., . . . Yin, H. (2014). Multi-constellation GNSS observations of equatorial ionospheric scintillation. *General Assembly and Scientific Symposium (URSI GASS)* (pp. 1-4). Beijing: IEEE.

Chapter 3

Data and Methods

This chapter describes the tools used in this research, the area of study, and the input data that was applied. The focus of this chapter will be on statistically and geostatistically describing the area of study and the input data in order to better understand and compare them to the results of chapters 4 and 5.

3.1 Platform and Tools

Esri ArcGIS 10.2 is used as the GIS software package for managing the input data, processing them according to proposed procedure, and analyzing and interpreting the results. ArcGIS was chosen to be used among other available platforms because it is the most widely used GIS package in the world, and it contains the biggest and most comprehensive library of geoprocessing tools (ArcGIS for Desktop, 2015). Along with its core functionalities, two additional extensions are also used: Spatial Analyst and Geostatistical Analyst

- **Spatial Analyst:** ArcGIS Spatial Analyst extension is a set of advanced tools and functions that are used to process and analyze raster-based data and combinations of raster and vector ones. Functions include: converting data between vector and raster; calculating Euclidean and weighted distances; interpolating data into continuous surfaces; generating density maps; analyzing terrains and deriving surfaces such as slopes and aspects; conducting local, focal, and zonal analyses; combining raster datasets through map algebra; among others (ArcGIS Spatial Analyst Extension, 2015).

Figure 20: Sample distance analysis using Spatial Analyst (ArcGIS Spatial Analyst Extension, 2015)

- **Geostatistical Analyst:** ArcGIS Geostatistical Analyst extension is a set of advanced statistical tools and functions that are used to explore spatial data and generate optimal surfaces, using a limited number of features compared to what is needed when using other toolsets and interpolation techniques. Functions include: exploring data variation and spatial relationships; creating optimal statistical methods using multivariate analysis; verifying generated models, and predicting uncertainty; among others (ArcGIS Geostatistical Analyst, 2015).

Figure 21: Sample geostatistical data Exploration (ArcGIS Geostatistical Analyst, 2015)

The used data are organized in the following formats: vector data as points, lines, or polygons are stored in shapefiles; raster data are stored as Esri Grid; and non-geometric tables are stored as dBase files. An alternative solution was to store all the data inside a file geodatabase. A geographic database, or simply geodatabase, is an extended database created by Esri and released with ArcGIS 8.0 in the year 1999 as a managed workspace for storing and managing various geographic datasets in one single location (MacDonald, 2001). A Geodatabase has many advantages such as storing all the layers in a single file; supporting larger data sizes reaching up to 2GB for the personal geodatabase and 1TB for the file geodatabase, that was released with ArcGIS 9.0 in the year 2006; including complex datasets such as topology and network datasets; and supporting other features such as security and authentication, and multi-user editing. However, the major drawback of using a geodatabase is that it is version- and platform-dependent, meaning that a geodatabase can only be shared with users that have the same version of ArcGIS, and in the case of personal geodatabase, that have Microsoft Windows as an operating system (Zeiler, 1999). In order to keep the used data easily sharable among any interested users, the data were not stored in a geodatabase, because it is not possible to predict whether all the potential users have the same version of Esri ArcGIS, if they have ArcGIS and are not using other GIS packages.

To automate and document all the processing that will take place in this study in order to apply the proposed procedures, all the steps are translated into Python scripts. Python is a high-level multi-purpose programming language that is designed with the emphasis of easy and readable code, compared to other languages such as C++ or Java (Python, 2015). This programming language was created by Guido Van Rossum in 1991, and since then it has become well-known and widely used worldwide, especially in the data scientists' area. In fact, according to (CodeEval, 2015), Python was the most popular programming language of the year in 2015, having 31.2% of the submitted code on their website being in Python, out of more than 600,000 submissions done by 2,000 programmers worldwide. ArcGIS, like many other GIS software packages, has adopted Python as a language for automating and extending its functionalities. Python developers can access the libraries of the ArcGIS software through the ArcPy module to develop scripts that can run within ArcGIS Python console or as a stand-alone script using any of the available integrated development environments (IDE).

Figure 22: The Python console in ArcMap (Screenshot taken by Author)

3.2 Area of Study

The objective of this study is to analyze the accuracy of a DEM at a level of a country. The country of choice for this study is the Lebanese Republic: Lebanon is situated on the eastern edge of the Mediterranean Sea. The official area of Lebanon is $10,452\text{Km}^2$.

Figure 23: Location of Lebanon on the Mediterranean Sea (Map designed by Author)

Lebanon as a geographic extent is very interesting to study because, even though it is a very small country ranked 170th in the list of countries by area out of 257 countries, it has a rich variation of landscape. It consists of a narrow coast, followed by a rough chain of mountains (Western Chain), followed by a plain valley (the 'Bekaa' valley), then a less rough chain of mountains (Eastern Chain) toward the eastern side (CIA World Fact Book, 2015). While being a coastal country, with a total of 210 kilometers of shoreline on its west side, the highest peak, called 'The Black Summit', or 'Al Qorneh as Sawda' in Arabic, has an altitude of 3,088 meters above sea level.

Figure 24: General topography of Lebanon (Map designed by Author)

Figure 25: 3D Perspective views of Lebanon. (A) Beirut and Mount Lebanon, taken from south-west; (B) The Beqaa valley, taken from the south; and (C) The southern hills, taken from south-west (Maps designed by Author)

As shown in the 3D perspective views in Figure 25, Lebanon has a different type of landscape in its various regions; in (A), Beirut city and its surrounding suburbs look like they are an extension of the land inside the sea, with almost a flat terrain, except for small bumps. Beyond Beirut is the governorate of Mount Lebanon, where the mountains rise suddenly leaving only a narrow coast. The highest peak in this view is Sannine Mountain, with an altitude of 2,628 meters; in (B), Beqaa valley resides in the middle between the western chain (on the left) and eastern chain (on the right). Being the flattest area in Lebanon, the Beqaa valley has historically been a major agricultural area; and in (C), the mountains from either western or eastern chains, drop in altitude and steepness to become more hills rather than mountains, with a maximum altitude of 940 meters at the Mountain of Aamel.

With reference to WGS84, Lebanon is located in Latitude between 33° and $34^{\circ} 40'$, and in Longitude between 35° and $36^{\circ} 35'$ approximatively. If the Universal Transvers Mercator (UTM) projection were to be used, and according to the UTM grid zones of the world (Morton, 2015), Lebanon would fall in the Zone 36 North, with only a small north-eastern area in the Zone 37 North. However, Lebanon has its own local coordinate system, called the Double Stereographic Projection, with reference to the Clarke 1880 ellipsoid. According to (Georeference.org, 2015), the Double Stereographic projection is an azimuthal, conformal, and polyconic perspective projection. It is called double because two internal projections occur: first the area is conformably projected to a sphere, and then the sphere is stereographically projected into a plane. Clarke 1880 is an ellipsoid created by the British geodesist, Alexander Ross Clarke, and published in his book called *Geodesy* in 1880 (Clarke, 1880).

In 1920, two years after the end of World War I and in accordance to the Sykes-Picot agreement that took place between France and Britain, Lebanon and Syria became a French colony as per the French Mandate for Syria and Lebanon (Fieldhouse, 2007). Somewhere between 1920 and 1944, the French army adopted the Double Stereographic projection for Lebanon and Syria as a base to establish cadastral mapping for the two countries. The city of Tadmur in Syria, which is also known as the historical city of Palmyra, was chosen as the origin of the reference system; therefore, the coordinate system in Lebanon is known to have negative X-values all around the country, and the Y-values are negative in the majority of the country, except in the North where it becomes positive again (Mugnier, 2002).

In 1962, the Directorate of Geographic Affairs (DAG, derived from the French naming: *Directorat des Affaires Géographiques*) was established in the Lebanese Army and tasked with the objective of maintaining the Lebanese geodetic network, collecting geographic data, and compiling several products such as the 1/20,000 topographic maps, providing technical

support for the various ministries and public institution, and providing the army with the required data and support (Lebanese Army, 2015). One important role of DAG is the management and maintenance of the Lebanese geodetic network, which is a set of connected reference points that are used by mappers and land surveyors to reference their geographic data to the local coordinate system that was established by the French army. There exist several orders, each used for specific level of accuracy.

Figure 26: Triangulation reference points symbols (Left: 1st order, Right: 2nd and 3rd Order)
(Lebanese Army, 2015)

Figure 27: 2nd order triangulation reference point placed on top of a building in the Dora area at the north-eastern entrance of Beirut city (Photo taken by author)

3.3 Input Data

For the proposed method to be performed, three inputs are required: a DEM, a set of ground control points (GCPs), and land-cover dataset.

3.3.1 GCPs

The input GCPs must be as a vector point format. At least one GCP must be available within each of the available land-cover classes in the area of study. A set of 1,105 GCPs were collected from various surveying projects, but all using the same land surveying technique based on total stations with the Lebanese Geodetic Network as reference. The height values are based on Ohio State University 1991 Global Geopotential Model (OSU91) geoid, and the dataset is referenced horizontally to the Double Stereographic projection that is based on Clarke 1880 ellipsoid. According to leveling standards in conventional surveying (Moffit, 1998), the maximum allowed error is approximately 1.2 centimeters of height per 1 kilometer of distance between the measured point and the reference one; with a maximum allowed spacing between the points being 3 kilometers, the minimum accuracy of the height of each GCP is about 3-4 centimeters.

Originally, the collected GCPs at the early stages of this study were 9,494 points. However, many of these points were highly concentrated, compared to the area covered, to a point that the author was afraid that the concentration would affect the analysis being done. Therefore, in areas where there was concentration, many points were removed manually without referring to their height or the calculation of their accuracy, so that the reduction process could be considered a random one. This process resulted instead in the 1,105 GCPs. While in controlled simulations in laboratories, the GCPs are collected exactly to the need of the study, many real-life analyses rely on available data instead that are usually not distributed ideally as per the needs; instead the researcher must find a way to stratify these data in order to benefit from them as much as possible. Because this study aims to solve real-life problems, it will follow the case of using the available data instead of collecting the required points optimally.

Figure 28: Map showing the distribution of the selected 1,105 GCPs (Map designed by Author)

As shown in Figure 29, some of the used GCPs are from a road engineering project, where road cross-sections were taken for several roads in the area of Wadi Khaled in the north-eastern edge of the country. For its original use, each cross-section is made up of three to five GCPs, with an average distance between each two consecutive points is 1.5 meters. Back to objective of this research, these GCPs should be ideally distributed equally around the area of study; given that the total area is $10,452 \text{ Km}^2$, each GCP should cover an approximate area of 1.1 Km^2 , and, therefore, the average distance between two GCPs should be approximately $\sqrt{1.1} = 1.05 \text{ Km}$. Comparing this distance to the case of the road cross-section shows that these GCPs are 700 times denser than they should be, and therefore, in such cases, there must be some elimination, especially given that almost all these points are under the same LULC classification. After the elimination process, the average distance between two consecutive GCPs in those dense areas became approximately 400 meters, a dilution of the points' density by about 200 times.

Figure 29: Road cross-section taken at the most northern part of the country. The original GCPs are shown in brown. While the randomly elected ones are shown in bright green (Map designed by Author)

Even after the elimination process, and based on calculating the distribution pattern on the 1,105 GCPs using the Average Nearest Neighbor tool in ArcGIS (Mitchel, 2005), returned a z-score = -38.47, which means that the points are highly clustered with less than 1% likelihood that this pattern is the result of random chance. The observed mean distance between two GCPs was found equal to 621.8 meters, compared to the expected distance of 1,575.1 meters.

Figure 30: Average Nearest Neighbor result summary for the distribution of the GCPs (Graph generated by Author)

The result is logical because, as previously mentioned, these GCPs are a collection of points from various previous field surveying work belonging to different projects, and each project was focused in a specific area. A comprehensive statistical and geostatistical description and analysis of these GCPs, along with their relations with the other input datasets, is described in the next section entitled Exploratory Data Analysis.

3.3.2 DEM

The input DEM must be as a continuous raster format. The used DEMs in this study are the SRTM, the ASTER GDEM, and a local DEM derived from the 1/20,000 topographic maps.

SRTM: As previously described in section 2.1.3, the Shuttle Radar Topography Mission is a DSM generated through the technique of radargrammetry, with a spatial resolution on 90 meters and an assumed vertical accuracy of 20 meters. The height values are based on EGM96 geoid, and the dataset is referenced horizontally to WGS84 ellipsoid (Passini & Jacobson, 2007). Statistically, the SRTM values in the area of study spans between negative 24 and 3,080 meters, with an average height of 1,005.19 meters and a standard deviation of 618.74 meters. A first observation of the data range shows that the DEM assumed that there exist areas on the coastline that are below sea level, defined with negative altitude, and it missed the highest peak in Lebanon, lowering it by 8 meters.

Figure 31: Overview of SRTM over Lebanon, with two zoomed views of a sample area, each x10 (Map designed by Author)

ASTER GDEM: As previously described in section 2.1.2, the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation Model (GDEM) is a DSM generated through the technique of stereoscopic photogrammetry, with a spatial resolution of 30 meters and an assumed vertical accuracy varying between 10 and 25 meters. Similar to SRTM, the height values of the ASTER GDEM are based on WGS84 EGM96 geoid, and the dataset is referenced horizontally to the WGS84 ellipsoid (Reuter, Nelson, Strobl, Mehl, & Jarvis, 2009). Statistically, the ASTER value in the area of study spans between 0 to 3,076 meters, with an average height of 995.74 meters and standard deviation of 620.19 meters. A first observation of the data range shows that while the DEM was able to capture the coastline at altitude zero, it missed the highest peak in Lebanon, lowering it by 12 meters.

Figure 32: Overview of ASTER over Lebanon, with two zoomed views of a sample area, each x10 (Map designed by Author)

TOPO: The TOPO DEM is a DTM generated from the 10-meter interval contour lines collected from the 1/20,000 topographic map that are produced by the DAG in the Lebanese Army (Lebanese Army, 2015). Lebanon is divided into 121 sheets, and each sheet is produced on an A0 paper map, as shown in the topo map below. These contour lines on the maps are generated using the technique of stereoscopic photogrammetry using aerial photos, and the DTM is generated using the topo-to-raster interpolator, as described in section 2.2. The output TOPO DTM has a spatial resolution of 5 meters. The height values are based on the OSU91 geoid, and the dataset is referenced horizontally to the Double Stereographic projection that is based on Clarke 1880 ellipsoid, similar to the input GCPs.

As described in section 2.6, the assumed accuracy of a DEM derived from contour lines as per the US geological survey Level 3 product standard is equal to one-third of the contour interval. In this case, the assumed accuracy is 3.33 meters. Statistically, the TOPO value in the area of study spans between 0 to 3,086 meters, with an average height of 1003.36 meters and standard deviation of 618.90 meters. A first observation of the data range shows that the DEM was able to capture the coastline at altitude zero, similar to ASTER, and the highest peak too by a slight difference of 2 meters.

Figure 33: A snapshot of 1/20,000 topographic map Batroun Sheet, showing the 10 meters-interval contour lines

Figure 34: Overview of TOPO over Lebanon, with two zoomed views of a sample area, each x10 (Map designed by Author)

Comparing the three DEMs shows that they all have almost the same values, with some minor differences, where, for example, the SRTM has negative values, while the other two products do not. Also, as shown in Table 1, the correlation matrix that compares the similarity of each pair of DEMs shows that all of them behave similarly, where each pair have a high positive correlation. This proves that even though the three DEMs were collected differently and at different time stamps, they are representing the same surface, and thus comparing their results especially in term of accuracy is valid.

A more thorough assessment of the correlation matrix shows that the correlation between the ASTER and the SRTM, and between SRTM and TOPO, have slightly lower value than the correlation between ASTER and TOPO, which means that ASTER and TOPO are more coherent among each other and the SRTM is slightly less with them. One potential justification for this differentiation is that ASTER and TOPO were collected through stereoscopic photogrammetry, while SRTM was collected through radargrammetry. In all cases, this difference is very minor and will not be taken into consideration.

Table 1: Correlation matrix of the three DEMs

	ASTER	SRTM	TOPO
ASTER	1.00000	0.99740	0.99856
SRTM	0.99740	1.00000	0.99734
TOPO	0.99856	0.99734	1.00000

3.3.3 Land-Use / Land-Cover

The input Land-Use / Land-cover (LULC) must be as a vector polygon format, with one field containing the name or unique code of the Land-cover classification, as per CORINE standard. CORINE standard was chosen out of several international ones, such as the US NLCD, because the area of study shares common geographic characteristics with the European continent and specifically with the Mediterranean part of it. In fact, Europe and Asia represent a geographic continuity called Eurasia (Rosenberg, 2015). Also, the data are originally classified as per the Lebanese SDATL standard and were published by the Council of Development and Reconstruction (CDR) in 2005, but the LULC layer dating originally 1998. The data are converted to CORINE in order for the results of this study to be comparable to other countries. CORINE, NLCD, and SDATL were described in section 2.5, and the full details on the conversion of the classes between SDATL and CORINE are shown in Appendix A.

Statistically, Table 2 summarizes the total area covered by each LULC class, both level 1 and level 2. Out of the fourteen level 2 classes, the six most dominant ones that make 93% of total area are: open spaces and bare soil which involve mountain tops that get covered seasonally by snow (32%); Permanent crops which mainly involve fruits trees, olive trees, and vineyards (16%); arable lands which involve seasonal varying agriculture (15%); forests (13%); scrubs and herbaceous vegetation (12%); and urban fabrics (5%). Appendix C represents a photo gallery from around Lebanon to visualize the different landscapes that represent the different major LULC classification applied to Lebanon. This photo gallery will help readers of this thesis who are not familiar with the country of study to better understand it, as the definition of the same LULC class could slightly change from one country to another.

Figure 35: Land-Use / Land-Cover classification as per CORINE level 1 (Map designed by Author)

Table 2: Total area per LULC class, and its percentage from the area of study

L 1	L1 Description	L 2	L2 Description	Area (Km²)	
1	Artificial Surfaces	11	Urban Fabric	486.59	5%
		12	Industrial, Commercial and Transport Units	47.56	0%
		13	Mine, Dump and Construction Sites	111.84	1%
		14	Artificial, non-agricultural vegetated areas	2.68	0%
2	Agricultural Areas	21	Arable Land	1,588.36	15%
		22	Permanent Crops	1,605.37	16%
		24	Heterogeneous Agricultural Areas	134.24	1%
3	Forest and Semi-Natural Areas	31	Forests	1,366.00	13%
		32	Scrub and/or Herbaceous Vegetation Association	1,216.18	12%
		33	Open Space with Little or no Vegetation	3,681.50	36%
4	Wetlands	41	Inland Wetlands	2.57	0%
		42	Maritime Wetlands	2.20	0%
5	Water Bodies	51	Inland Waters	14.33	0%
		52	Maritime Waters	0.02	0%

All the types of land covers are dispersed around the area of study. However, as shown in the 6 maps forming Figure 3-17, the dominant ones are more concentrated in specific regions as follows: (A) Open spaces are mostly concentrated on the upper heights on the western and eastern chains of mountains, where the average altitude of these areas is 1,355 meters above the sea. Areas from this altitude and above get covered by snow during the winter, and as it

gets higher, fewer crops and trees are able to grow, which results in bare soil. (B) Permanent crops are mostly concentrated toward the center of the Beqaa valley, where it is famous for vineyards; toward the north in Koura, where it is famous for growing olive trees; and toward the southern coastline going from the Damour plain down south to the city of Sour, where it is famous for growing orange trees. (C) Arable land, which represent in contrast to (B) seasonal crops, is mostly concentrated in the Beqaa valley and in Akkar plain on the upper northern coastline. (D) Forests are mostly concentrated in Mount Lebanon and extending toward the north, and altitude wise are located just beneath the open space in the mountain heights. (E) Scrubs and herbaceous vegetation are somewhat less concentrated than the previous types and are mostly placed between forests and the open spaces. Finally, (F) Urban fabrics are mostly concentrated around the major cities in Lebanon, starting from the capital Beirut in the middle coastline, where the largest urban fabric exists, followed by area surrounding the city of Tripoli on the northern coastline, and the city of Sour on the southern coastline.

Figure 36: Dominant land covers. From upper left: Open Space (32%), Permanent Crops (16%), Arable Land (15%), Forest (13%), Scrubs and Herbaceous Vegetation (12%), and Urban Fabric (5%) (Maps designed by Author)

3.3.4 Pre-Analysis Data Transformation

Before combining and analyzing all the input data, they must all be aligned to the same coordinate system. Given that the most critical input data are the GCPs, the study will adopt their reference system for all the remaining data; all data must be referenced to the double stereographic projection based on Clarke 1880 ellipsoid and referencing the height values with respect to the OSU91 geoid.

The LULC dataset is already under the same projection; ASTER and SRTM must be projected from WGS 1984 into the double stereographic projection. While data projection tools are available in most GIS software packages, it is important to note that while the origin coordinate system is universal and defined in all these packages, the destination one is local and needs to be defined by the user. Even if defined, performing the projection solely will result in a planar deviation of the data between 90 to 100 meters, because the origin and destination coordinate systems are based on different ellipsoids. To overcome this issue, Lebanese land surveyors adopted an additional transformation by creating and using a custom geographic transformation that is added to the projection operation. This custom transformation is defined by the following parameters:

- **Input Coordinate System:** Geographic (WGS1984)
- **Output Coordinate System:** Double Stereographic (Clark1880)
- **Method:** Position Vector
- **X-Axis Translation:** 175.993534m
- **Y-Axis Translation:** 125.164623m
- **Z-Axis Translation:** -244.865805m
- **X-Axis Rotation:** 17.315446"
- **Y-Axis Rotation:** 12.135795"
- **Z-Axis Rotation:** 10.542653"
- **Scale Difference:** -6.123214ppm

Figure 37: 'Create Custom Geographic Transformation' tool in ArcGIS 10.2 (Screenshot taken by Author)

By adopting this additional transformation during data re-projection, the 90-100m shift is reduced to 4- 6 centimeters. Given the spatial resolution of the input datasets (90m for SRTM and 30m for ASTER), this remaining shift is negligible.

As for the height values, the DEMs and the GCPs must have their values referenced to the same geoid in order for their data to be comparable. Also, comparing the assumed accuracy of SRTM and ASTER, the geoid differences are considered minimal and negligible. Furthermore, as it will be defined in a later section, the analysis will be based on the concept of local accuracy, compared to the global one. In this case, the geoid difference will not affect the analysis and can be omitted without the risk of introducing additional uncertainties to the input data.

3.4 Exploratory Data Analysis

3.4.1 Descriptive Geostatistics of the GCPs

Before defining a model to analyze the input DEMs, a first step is to perform a descriptive analysis of the available GCPs with their actual altitudes. These points, or some of them, will be used later on in defining the proposed model. Each GCP has in its attributes five important values: its actual height, denoted Z ; its related LULC level 2 class, denoted as $L2$; the coincident SRTM height, denoted as $SRTM$; the coincident ASTER height, denoted as $aster$; and the coincident TOPO height, denoted as $topo$.

Figure 38: Histogram showing the distribution of the Z values from the GCPs (Screenshot taken by Author)

Figure 39: Trend analysis of the GCPs' Z values (Screenshot taken by Author)

The result shows that these points have a minimum Z of 0.17 meters and a maximum of 2,888.1 meters. The average altitude is 565.14 meters, with respective values for the first and the third quartiles as 63.73 and 871 meters, as shown in the Figure 38.

The trend analysis tool is used to study the global trends in the area of study. The result of the analysis is shown in Figure 39: the green line expresses the tendency of altitude in the West-East orientation. Altitudes begin with the minimum values on the coastline, reach the maximums in the central area, only to re-descend toward the east side of the country, but stay higher on the initial west side. Additionally, the altitudes show a quadratic trend in the West-East and North-South directions. Therefore, the distribution of the GCPs altitudes can be described as non-homogeneous.

Figure 40: Normal QQ Plot showing the distribution of the Z values with respect to the normal one (Screenshot taken by Author)

Next, using the Normal QQ Plot tool, the frequency distribution of the altitudes is compared to the normal distribution. Additionally, the quantile of the distribution of the altitudes is compared to the Gaussian curve. As shown in Figure 40, the straight line represents the normal distribution. The sample points do not follow the normal distribution, and, therefore, adopting all of them would result in biased results in the proposed model. However, as it will be defined in the next sub-section, the generation of the model will be stratified based on the LULC classification.

Finally, structural analysis of the altitudes is performed to establish a relationship between the sampling points and a particular direction of the anisotropy. As shown in Figure 41, the anisotropy of the sample data shows that is the autocorrelation of the altitude variable is different in one or many directions. Therefore, the GCPs are considered as spatially non-stationary, and, therefore, there exists no specific direction for the changes in their altitudes.

Figure 41: Semivariogram point cloud and surface of the GCPs' Z values (Screenshot taken by Author)

3.4.2 Altitude Variation per LULC Class

Beyond the basic descriptive statistics of the LULC dataset that was performed in sub-section 3.3.3, and in order to better understand the geographical distribution of the different LULC classes with respect to the altitude variation, a zonal analysis of the TOPO DEM is performed, taking the LULC dataset as the zones, once using level 1 classes, and second using level 2 classes. The results of the two levels are summarized in the two tables below:

Table 3: Variation of altitude per level 1 classes

	Level 1	MIN	MAX	RANGE	MEAN	STD	MAJORITY	MINORITY
1	Artificial Surfaces	0	2,222	2,222	532.61	435.65	5	2,222
2	Agricultural Areas	0	2,485	2,485	720.90	451.24	864	2,484
3	Forest and Semi-Natural Areas	0	3,086	3,086	1,202.65	626.72	1,410	3,086
4	Wetlands	0	1,718	1,718	550.35	511.78	864	88
5	Water Bodies	0	2,212	2,212	742.26	381.61	830	352

An initial observation of the above table shows that the Forest and Semi-Natural Areas have altitudes ranging on all the altitudes in the whole country, from 0 to 3,086 (refer to sub-

section 3.3.2). This means that a forest or a natural area could be found at any altitude. However, the average altitude of a forest of natural area is 1,202.65 meters, almost 200 meters higher than the country average altitude, which is 1,003.36 meters according to TOPO. A second observation is that all the classes have a minimum altitude of 0 meters, meaning that the country of study has a coastline rich in every type of LULC. It is the maximum altitude that varies between them.

Table 4: Variation of altitude per level 2 classification

	Level 2	MIN	MAX	RANGE	MEAN	STD	MAJORITY	MINORITY
11	Urban Fabric	0	2,193	2,193	548.15	408.89	4	2,110
	Industrial, Commercial and							
12	Transport Units	0	1,536	1,536	211.81	340.13	5	1,308
	Mine, Dump and Construction							
13	Sites	0	2,222	2,222	606.95	517.53	0	2,222
	Artificial, non-agricultural							
14	vegetated areas	0	1,365	1,365	240.23	415.04	5	159
21	Arable Land	0	2,450	2,450	785.10	407.71	864	2,450
22	Permanent Crops	0	2,485	2,485	656.82	476.36	870	2,484
	Heterogeneous Agricultural							
24	Areas	0	2,349	2,349	728.59	518.54	359	2,342
31	Forests	0	2,659	2,659	925.08	504.35	660	2,647
	Scrub and/or Herbaceous							
32	Vegetation Association	0	2,943	2,943	1,054.05	576.11	1,410	2,943
	Open Space with Little or no							
33	Vegetation	0	3,086	3,086	1,354.87	636.57	1,890	3,086
41	Inland Wetlands	1	1,718	1,717	959.89	280.11	864	877
42	Maritime Wetlands	0	397	397	21.75	53.03	3	88
51	Inland Waters	0	1,952	1,952	742.26	380.78	830	352
52	Marine Waters	0	2,212	2,212	728.92	1,038.94	0	2,212

3.4.3 Distribution of GCPs per LULC Class

Using the L2 field in the GCPs layers, the altitudes of the points as per the TOPO DEM are summarized in the table below. The altitudes are taken with respect to the TOPO and not their actual Z in order to compare their results to the work done in the previous sub-section. The distribution of the values with respect to the normal one per each class is computed, and their QQ-Plot for each level 1 or level 2 classification is shown in Appendix A2. A visual inspection of all these QQPlots shows that these points do not have a normal distribution, even per land use class.

Table 5: Variation of GCPs' Altitude per Level 2 Classification

Level 2	GCPs Count	% of Total	MIN	MAX	AVG	STD
11	218	20%	0	1,319	281.34	377.82
12	36	3%	1	883	301.67	407.50
13	20	2%	2	890	242.10	301.90
14	0	0.0%	-	-	-	-
21	353	32%	2	1,820	728.94	313.35
22	212	19%	2	1,680	549.66	417.13
24	25	2%	4	492	39.88	101.12
31	38	3%	79	2,305	783.00	568.21
32	38	3%	40	1,998	474.76	461.68
33	130	12%	1	2,884	698.38	651.00
41	3	0.3%	863	863	863.00	0.00
42	0	0.0%	-	-	-	-
51	32	3%	853	866	862.59	2.93
52	0	0.0%	-	-	-	-

In terms of GCPs distribution per LULC, and comparing the count percentage to the actual coverage percentage of the same class from Table 2, shows that:

- There are three level 2 classes that do not have any GCPs. They are: artificial, non-agricultural vegetated areas (14), Maritime Wetlands (42), and Maritime Waters (52). However, these three types of land cover form almost less than 5 Km², which makes 0.05% of the total area of study. Therefore, in order not to leave any gap, these areas will be assessed in the proposed procedure similar to their average level 1 class.
- There is an excess in GCPs in the following classes: Urban Fabric (11), Industrial, Commercial and Transport Units (12), Arable Land (21), and Inland Waters (41). Conversely, there is a shortage in the GCPs in the following classes: Forests (31), Scrub and/or Herbaceous Vegetation Association (32), and Open Space with Little or no Vegetation (33). The remaining four classes have almost the same percentages.

Despite the discrepancies in the percentages in the majority of the LULC classes, it will not affect the proposed procedure, as it stratifies the analysis per LULC separately. What is more important is that the GCPs in each LULC class represents well the variation of the whole coverage of that class. One way of checking the coverage of the GCPs is to compare the ranges of the GCPs from Table 5 to the ranges of the overall LULC class from Table 4 to assess what is the percentage of the GCPs range compared to the class one. However, before comparing the ranges, it should be taken into consideration that a range from the LULC table could contain outliers which could cause the range percentage of the GCPs to become small and misinterpreted as insignificant.

Therefore, an intermediate range should be calculated for each class, called optimized range, to include only the majority of the data. For a 95% level of confidence, the majority of the data is assumed be grouped two standard deviations around the mean value. Therefore, the new temporary minimum and maximum values are calculated as follows:

$$New\ MIN = MAXIMUM[MEAN - (2 \times STD), Old\ MIN] \quad (3-1)$$

$$New\ MAX = MINIMUM[MEAN + (2 \times STD), Old\ MAX] \quad (3-2)$$

$$Optimized\ RANGE = New\ MAX - New\ MIN \quad (3-3)$$

The new MIN and MAX values are bound by the old ones because it is not logical that the majority of the data have values beyond the limits of the whole population. The GCPs coverage is calculated as follows:

$$GCPs\ Coverage = \frac{GCP\ RANGE}{Optimized\ RANGE} \times 100 \quad (3-4)$$

The below table summarizes the results of the equations 3-3 and 3-4 for all the LULC level 2 classes:

Table 6: Optimized Range of Altitudes per Level 2 Classification

Level 2	GCPs Count	GCPs RANGE	Optimized RANGE	Coverage
11	218	1,319	1,366	97%
12	36	882	892	99%
13	20	888	1,642	54%
14	0	-	1,070	-
21	353	1,818	1,601	114%
22	212	1,678	1,610	104%
24	25	488	1,766	28%
31	38	2,226	1,934	115%
32	38	1,958	2,206	89%
33	130	2,883	2,546	113%
41	3	0	1,120	0%
42	0	-	128	-
51	32	13	1,504	1%
52	0	-	2,212	-

As shown in the above table, seven classes are well represented by the GCPs, with a range 80% or more of the optimized one. These classes make a total of 97% of the area of study, while the remaining six classes that are not well represented only make 3%. Therefore, the shortage in the representation in the GCPs for the latter classes will not have a significant effect on the proposed study.

3.5 Summary

This chapter introduced the tools, area of study, and input data that was used in processing and assessing the proposed model and its usage as it will be described in chapters 4 and 5.

The chosen platform for data processing and analysis was Esri ArcGIS 10.2 with the Spatial Analyst and Geostatistical Analyst extensions. Additionally, python script was used to automate the proposed procedure into a script, for easy and systematic reuse of the procedure on different datasets. While other GIS software packages are available in the market, ArcGIS was chosen because it is the most used worldwide, and its list of functionalities are the most comprehensive. Also, all the vector data are chosen to be stored in shapefiles as it is the most portable GIS format in the market, so that the results are sharable with users of different systems.

The chosen area of study is Lebanon; it is a relatively small country, but, as shown in this chapter, it has a rich variation in its altitude, ranging from zero to 3,088 meters above sea level. Additionally, in different parts of the country, there exist different landscapes, ranging from large plains such as the Beqaa valley, to the steepest mountains such as the western chain of mountains, passing by mild slopes such as the southern hills.

Finally, three inputs were introduced: a set of GCPs, a DEM, and an LULC. For the GCPs, a total of 1,105 points with high positioning accuracy were collected from various sources and combined together. For the DEM, three products were introduced: The SRTM with a spatial resolution of 90 meters, the ASTER with 30 meters, and the TOPO with 5 meters; as for the LULC, a national dataset was chosen from the SDATL publication, but the classifications were converted from the local standard into the European CORINE one, so that the results of any analysis performed on it can be comparable to other datasets from other countries.

The next chapter will cover the proposed procedure for defining the local accuracy vario-model and the implementation of this model into the input data that were previously defined and described in this chapter.

Chapter 3 Bibliography

- CIA World Fact Book. (2015). *Lebanon*. Retrieved from CIA World Fact Book:
<https://www.cia.gov/library/publications/the-world-factbook/geos/le.html>
- Clarke, A. R. (1880). *Geodesy*. Oxford: Clarendon Press.
- CodeEval. (2015, February 9). *Most Popular Coding Languages of 2015*. Retrieved from CodeEval: <http://blog.codeeval.com/codeevalblog/2015>
- Esri. (2015, June 17). *ArcGIS for Desktop*. Retrieved from Esri:
<http://www.esri.com/software/arcgis/arcgis-for-desktop>
- Esri. (2015, June 17). *ArcGIS Geostatistical Analyst*. Retrieved from Esri:
<http://www.esri.com/software/arcgis/extensions/geostatistical>
- Esri. (2015, June 17). *ArcGIS Spatial Analyst Extension*. Retrieved from Esri:
<http://www.esri.com/software/arcgis/extensions/spatialanalyst>
- Fieldhouse, D. (2007). Western imperialism in the Middle East 1914–1958. *Journal of Global History*, 405-406.
- Georeference.org. (2015, June 19). *Double Stereoprojection*. Retrieved from Georeference.org: http://www.georeference.org/doc/double_stereographic.htm
- Lebanese Army. (2015, June 19). *Directorate of Geographic Affairs*. Retrieved from Official Website of the Lebanese Army: http://www.lebarmy.gov.lb/en/d_a_g/
- MacDonald, A. (2001). *Building a Geodatabase*. Esri Press.
- Mitchel, A. (2005). *The ESRI Guide to GIS Analysis, Volume 2*. Redland: Esri Press.
- Moffit, F. (1998). Leveling. In *Surveying, 10th Edition* (pp. 61-127). Addison Wesley Longman.
- Morton, A. (2015, June 19). *UTM Grid Zones of the World*. Retrieved from Distribution Mapping Software: <http://www.dmap.co.uk/utmworld.htm>
- Mugnier, C. (2002). Grids and Datums - The Lebanese Republic . *Photogrammetric Engineering and Remote Sensing*, 977-980.

- Passini, R., & Jacobson, K. (2007). Accuracy Analysis of SRTM Height Models. *ASPRS 2007 Annual Conference*.
- Python. (2015, June 17). *About Python*. Retrieved from Python:
<https://www.python.org/about/>
- Reuter, H., Nelson, A., Strobl, P., Mehl, W., & Jarvis, A. (2009). A first assessment of Aster GDEM tiles for absolute accuracy, relative accuracy and terrain parameters. *Geoscience and Remote Sensing Symposium 5*, 240-243.
- Rosenberg, M. (2015, June 21). *What is Eurasia? Defining the World's Largest Continent*. Retrieved from About Education:
<http://geography.about.com/od/learnabouttheearth/a/What-Is-Eurasia.htm>
- Zeiler, M. (1999). Chapter 4: The Structure of Geographic Data. In *Modeling Our World: The Esri Guide to Geodatabase Design* (pp. 61-74). Redlands, California: Esri Press.

Chapter 4

The Accuracy Vario- Model

This chapter describes the proposed procedure for creating a vario-model estimating the varying local accuracy of a DEM. The analysis of results section will review the proposed model and compare its results to conventional methods of DEM accuracy assessment in order to validate the proposed work. The proposed procedure is described mathematically first, then applied through a Python script running on Esri ArcGIS 10.2. The exact script is available in Appendix B.

4.1 The Local Accuracy

Conceptually, with a specific application to DEM, the local and global accuracy should be the same value. However, this is not true: the two accuracies are different, as several factors that place in the capture and modeling phases of the DEM will increase the global uncertainty solely, while keeping the differences between the various locations intact with respect to each other.

Several factors exist that affect the accuracy of the DEM (Su & Bork, Influence of Vegetation, Slope, and LiDAR Sampling Angle on DEM Accuracy, 2006), but the major ones that would create the difference between the two accuracies are:

- The difference in type of height reference between the model and the control points from the reality
- The difference in reference geoid
- The process of the adjustments in re-projecting the data from one coordinate system to another, especially if the two systems have different height reference
- Uncertainty introduced by the new data (geoid transformations) added to the theoretical model

Furthermore, and back to the main objective of capturing and modeling a DEM, it is to have a continuous representation of a surface, irrespective if it is solely the ground or includes the different natural and man-made structures to be used in analyzing this surface and deriving new datasets that reveal additional information about it and depict further geographic. These analyses will be based not on the individual heights, but on the changes of these heights that form the structure of the surface. Therefore, the local accuracy of the heights is more

representative than the global one that can only be interpreted individually, and thus will be the focus of this study as it could also lead to the assessment of the accuracy of the derived datasets.

To simplify the proposed methodology, two assumptions were made as they have a relative minor effect on the study:

- The used DEMs are considered as low-to-medium accuracy, and the geoid differences are considered minimal and negligible. Therefore, all the input data will be assumed to have the same height reference.
- As previously described, the accuracy of the GCPs is within the range of few centimeters. Also compared to the assumed accuracies of the input DEMs, the error is minimal and these points will be considered references in this method without modeling their accuracies.

Provided that the required data is in the specific format, data can be processed for any area of the world using any GIS software. The local accuracy of the height between two points i and j is calculated as follows:

$$\text{Local Accuracy: } LA_{ij} = |(Z_j - Z_i) - (Zm_j - Zm_i)| \quad (4-1)$$

Where:

- Z_j is the actual height at point j
- Z_i is the actual height at point i
- Zm_j is the modelled height at point j
- Zm_i is the modelled height at point i

Figure 42: Sample height profile between points i and j (Graph generated by Author)

In reality, it is not practical to calculate the relative accuracy between two points only; instead, there should be a possibility to do this calculation for any number of points. Keeping the same equation will raise several issues:

- For each point i , what is the criterion to choose the point j ? If there are different criteria, which one is the best to choose from? And will not this introduce additional uncertainty?
- If the selection criterion is reflexive, then if point j is chosen to calculate the relative accuracy of point i , then point i is to be chosen to calculate the relative accuracy of point j ; this will result in pairs of points having the same relative accuracy, which is misleading.

Instead of relying on a pair of points to calculate the relative accuracy, this study proposes to adopt for each point the average relative accuracy with respect to its surrounding points as follows:

$$\text{Average Local Accuracy: } ALA_i = \sum_{j=1}^n \frac{LA_{ij}}{n} \quad (4-2)$$

Where:

- LA_{ij} is the local accuracy (4-1)
- j is any elements from the subset of points that surround point i

Figure 43: Top view showing the local accuracy between a point i and a set of points j s (Graph generated by Author)

Given that the proposed model must work properly on any set of points, irrespective of their densities and distribution patterns, the surrounding points cannot be selected using a fixed search radius nor a fixed number of closest points; such selection would result in some points having no surroundings, and others including points that are too far from each other,

where these points do not have any tangible spatial autocorrelation. Instead, the proposed work, inspired from the Natural Neighbor spatial interpolation, relies on the Voronoy tessellation to select the neighboring points to calculate ALA.

By definition, Voronoy tessellation is a mathematical method created by Georgy Voronoy to divide a space into a set of sub-regions based on an input set of points. In a 2D system, these sub-regions are defined as a set of adjacent polygons, where each polygon defines the area that is closest to a point with respect to others. Assuming that for each point i , the Voronoy polygon is defined as $V(i)$, the neighboring points will be chosen as those having their Voronoy polygons adjacent to the one of i . Accordingly, for a sample point i , the set of neighboring points are shown in the figure below:

Figure 44: Sample Voronoy decomposition, showing the neighboring points of a point I (Graph generated by Author)

4.2 Calculating the Local Accuracy

The script is developed using Python on top of Esri ArcGIS 10.2. As described next, the procedure is divided into three main steps, and it can be performed using any GIS software package with raster-based spatial analysis tools.

1. **Extract Overlay Data:** For each GCP in the dataset, the overlaying data from the three DEMs and the LULC dataset are extracted and stored as additional fields in the GCPs table of attributes. It can be achieved by creating two new fields in the layer table of attribute, named Z_m for the modeled height from the DEMs and L_2 for the LULC class from the LULC dataset. Then the values are stored using spatial join between each of the related datasets and the GCPs one. The figure below shows the resulting table of attributes for the GCPs layer when assessing the SRTM DEM.

	FID	Shape	X	Y	Z	L2	Zm
	3	Point	-366345.176	-99468.819	17.26	11	21
	4	Point	-366934.381	-99425.721	2.66	11	4
	6	Point	-365299.444	-98460.941	39.261	11	44
	7	Point	-351222.86	-98422.55	449.95	11	437
	15	Point	-366811.227336	-97231.611273	14.749	11	-1
	16	Point	-366533.258336	-97208.116273	16.436	11	6
	17	Point	-366806.656336	-97003.723273	15.319	11	3
	20	Point	-365439.948	-96704.081	17.52	11	19
	21	Point	-365443.796	-96695.617	16.907	11	19
	22	Point	-366848.305336	-96689.152273	17.981	11	5
	25	Point	-367491.823046	-96612.869194	2.79	11	7

Figure 45: Screenshot of the GCPs table of attributes when assessing the SRTM DEM (Screenshot taken by Author)

2. Calculate Local Accuracy: For each GCP in the dataset, the ALA is calculated as described in section 4.1. For consistency, it must be implemented through scripting automation with the following logic. The results are shown in the screenshot in figure below.
 - i. Generate the Voronoy polygon layer that defines the neighborhoods as per the geographical distribution of the points. In ArcGIS, the tool is called '*Create Thiessen Polygons*';
 - ii. For each GCP, search for the neighboring GCPs that have adjacent neighboring Voronoy polygons. This can be achieved by using the '*Select By Location*' tool and using the condition "*Touches Boundary*";
 - iii. For each pair of the central and neighboring GCPs, calculate the LA as defined in equation 4-1, using the mathematics library in the scripting language; and
 - iv. For that central GCP, calculate the ALA as defined in equation 4-2, using the same library as in (iii).

FID	Shape	X	Y	Z	L2	Zr	ALA
3	Point	-366345.176	-99468.819	17.26	11	2	4.309
4	Point	-366934.381	-99425.721	2.66	11		5.518
6	Point	-365299.444	-98460.941	39.261	11	4	4.304
7	Point	-351222.86	-98422.55	449.95	11	43	6.252
15	Point	-366811.227336	-97231.611273	14.749	11		9.481
16	Point	-366533.258336	-97208.116273	16.436	11		10.745
17	Point	-366806.656336	-97003.723273	15.319	11		12.863
20	Point	-365439.948	-96704.081	17.52	11	1	3.19
21	Point	-365443.796	-96695.617	16.907	11	1	8.194
22	Point	-366848.305336	-96689.152273	17.981	11		11.137
25	Point	-367491.823046	-96612.869194	2.79	11		7.334

Figure 46: Screenshot of the GCPs table of attributes, highlighted the resulting ALA values per GCP (Screenshot taken by Author)

3. Compile Data per LULC Class: Statistically calculate the average difference per LULC class and assign this average to all the cells covering each LULC class within a continuous raster covering the whole area of study. This step can be achieved using the following sub-steps:
 - i. Calculate the average ALA for each LULC class, using the Summarize tool; an output summary table is generated;
 - ii. Join the output table to the LULC layer using the *Join Field* tool, and the L2 field as a common field between the table and the layer; and
 - iii. Convert the LULC layer from vector polygon into a raster using the *Polygon to Raster* tool and specify the ALA field for the output raster values. For spatial consistency, define the spatial resolution of the output raster the same as the related input DEM.

4.3 Validating the Procedure

In order to assess the accuracy of the proposed model and how much it enhanced the modeling of accuracy compared to the traditional methods, the height accuracy of the three DEMs were computed at every GCP, then these spot accuracies are compared to three different models:

Model 1: Assumed accuracy: as defined by the DEMs' data providers. For SRTM, the assumed accuracy is 20 meters, for ASTER is between 10 and 25 meters averaged to 17.5 meters, and the TOPO is 3.33 meters. These figures were previously discussed in sub-section 3.3.2.

Model 2: Average accuracy: computed as the average value of the accuracies of the GCPs. The average accuracy is computed as the RMSE of all the GCPs participating in the model generation. The RMSE is calculated as the equation below.

$$RMSE = \sqrt{\frac{\sum (z_m - z_g)^2}{n}} \quad (4-3)$$

Where:

- z_m is the modelled height taken from the DEM
- z_g is the actual height taken from the GCP measurement

Figure 47: Sample profile showing the distribution of the GCPs compared to the average accuracy
(Graph generated by Author)

As shown in Figure 47, the validation GCPs will be plotted on top of the generated RMSE single-model, and the deviation between their spot local accuracy will be compared to the single-model as a whole. Meaning that all the GCPs will be treated as one group.

Model 3: Proposed vario-model: computed as described in the sections 4.1 and 4.2

The deviations from the three defined models is calculated as the difference between the individual local accuracy calculated for the GCP, and the collected values as follows:

$$Accuracy\ Deviation = ALA_{GCP} - ALA_{Model} \quad (4-4)$$

Unlike the case of the single RMSE value and as shown in Figure 47, the GCPs will be grouped by their land cover classification, and the deviations will be calculated with respect to the optimized RMSE for each LULC class.

Figure 48: Sample profile showing the distribution of the GCPs compared to the vario-model (Graph generated by Author)

The results of Model 1 are already existing; for those of Model 2 and 3, they must be calculated using the GCPs. In order to avoid getting ambiguous results, the GCPs need to be divided into two groups: The first group forming 80% of these points will be used to generate Model 2 and 3. The second group forming 20% of these points will be used to validate the three models. In principle, the division of these two groups must be made randomly, but this cannot be applied in this study: Because if the estimation and modelling accuracy is made dependent on the land cover of the area, then the division of the GCPs' group should be made random at the level of each LULC class. After assigning the coincident LULC class for GCP, the number of GCPs that should take part in either of the two groups is summarized in the table below.

Table 7: Distribution of GCPs per LULC Level 2 Classes

Level 2 Class		Total GCPs	Model (80%)	Validation (20%)
11	Urban Fabric	218	174	44
12	Industrial, Commercial and Transport Units	36	28	8
13	Mine, Dump and Construction Sites	20	16	4
21	Arable Land	353	282	71
22	Permanent Crops	212	169	43
24	Heterogeneous Agricultural Areas	25	20	5
31	Forests	38	30	8
32	Scrub and/or Herbaceous Vegetation Association	38	30	8
33	Open Space with Little or no Vegetation	130	104	26
41	Inland Wetlands	3	2	1
51	Inland Waters	32	25	7

4.4 Results

The results of the procedure that was previously defined in section 4.3 applied to the input data that were also described in sections 4.1 and 4.2 are described in this section. The results are divided into three parts: the first part focus on the calculated results of each GCP alone, the second part on the calculated result of the conventional method of RMSE, and the third part on the result of the proposed vario-model.

4.4.1 Individual Accuracy at the GCPs

As previously described in sub-section 4.2, the ALA is calculated at each of the 1,105 GCPs for the three DEMs under study: SRTM, ASTER, and TOPO. The results of the individual ALAs are described next for each of the three models separately first, and in the end, in comparison to each other.

SRTM Local Accuracy:

For the SRTM DEM, the individual ALA value ranges between 0.00 and 42.1 meters, with an average value of 7.00 meters, standard deviation of 6.55 meters, and the index of dispersion equal to 6.14. Looking at the histogram below in Figure 4-8 shows that the data are skewed to the right, which means that even though there are some GCPs with very high ALA values, the majority of them tend to have a low one toward 0.00. Looking at the Normal QQ-Plot shows that the data have a normal distribution toward the mean value and below, but not

toward higher values, which is consistent with the interpretation of the histogram that proves that higher values are not consistent. Using the General QQ-Plot, the distribution of the ALA values is compared to the one of the altitude and the LULC classification: the first plot shows that there is no consistency between the ALA and the altitude, and the second one shows a more disjoint distribution, simply because the LULC classification is as set of discrete numbers rather having continuous values.

Figure 49: Descriptive Geostatistics of the GCPs' individual ALA for SRTM (From upper-left corner: a- Histogram, b- Normal QQ-Plot, c- General QQ-Plot with respect to the altitude, and d- General QQ-Plot with respect to the LULC level 2 classification) (Screenshots taken by Author)

Beyond visual interpretation, and using SPSS statistical package, the curve estimation is calculated for the ALA with the altitude as the independent variable. The estimation is based on the various models: Linear, Quadratic, Cubic, Compound, Growth and Exponential. The Logarithmic, Inverse, Power, and S model were not used because the independent variable contains zero and negative values. Out of the six models used, the Cubic gave the relatively best fit model with a coefficient of determination $R^2 = 0.07$. This proves that there is no correlation between ALA and the altitude.

ASTER Local Accuracy:

For the ASTER DEM, the individual ALA value ranges between 0.00 and 35.82 meters, with an average value of 11.96 meters, standard deviation of 6.46 meters, and the index of

dispersion equal to 3.49. Looking at the histogram below in Figure 46 shows that the data are somehow skewed to the right. Looking at the Normal QQ-Plot shows that the data have a normal distribution. Using the General QQ-Plot, the distribution of the ALA values is compared to the one of the altitude and the LULC classification: the first plot shows that there is no consistency between the ALA and the altitude, and the second one shows a more disjoint distribution, simply because the LULC classification is as set of discrete numbers rather having continuous values, similarly to the results of SRTM.

Figure 50: Descriptive Geostatistics of the GCPs' individual ALA for ASTER (From upper-left corner: a- Histogram, b- Normal QQ-Plot, c- General QQ-Plot with respect to the altitude, and d- General QQ-Plot with respect to the LULC level 2 classification) (Screenshots taken by Author)

Beyond visual interpretation, and using SPSS statistical package, the curve estimation is calculated for the ALA with the altitude as the independent variable. The estimation is based on the various models: Linear, Quadratic, Cubic, Compound, Growth and Exponential. The Logarithmic, Inverse, Power, and S model were not used because the independent variable contains zero and negative values. Out of the six models used, the Cubic gave the relatively best fit model with a coefficient of determination $R^2 = 0.08$. This proves that there is no correlation between ALA and the altitude, similar to SRTM.

TOPO Local Accuracy:

For the TOPO DEM, the individual ALA value ranges between 0.00 and 20.44 meters, with an average value of 4.10 meters, standard deviation of 3.56 meters, and the index of dispersion equal to 3.09. Looking at the histogram below in Figure 51 shows that the data are skewed to the right like the one of the SRTM. This means that even though there are some GCPs with very high ALA values, the majority of them tend to have a low one toward 0.00. Looking at the Normal QQ-Plot shows that the data have a normal distribution toward the mean value and below, but not toward higher values, which is consistent with the interpretation of the histogram that proves that higher values are not consistent. Using the General QQ-Plot, the distribution of the ALA values is compared to the one of the altitude and the LULC classification: the first plot shows that there is no consistency between the ALA and the altitude, and the second one shows a more disjoint distribution, simply because the LULC classification is as set of discrete numbers rather having continuous values, similar to the results of SRTM and ASTER.

Figure 51: Descriptive Geostatistics of the GCPs' individual ALA for TOPO (From upper-left corner: a- Histogram, b- Normal QQ-Plot, c- General QQ-Plot with respect to the altitude, and d- General QQ-Plot with respect to the LULC level 2 classification) (Screenshots taken by Author)

Beyond visual interpretation, and using SPSS statistical package, the curve estimation is calculated for the ALA with the altitude as the independent variable. The estimation is based on the various models: Linear, Quadratic, Cubic, Compound, Growth and Exponential. The

Logarithmic, Inverse, Power, and S model were not used because the independent variable contains zero and negative values. Out of the six models used, the Cubic gave the relatively best fit model with a coefficient of determination $R^2 = 0.08$. This proves that there is no correlation between ALA and the altitude, similar to SRTM and ASTER.

Combining the Three Results:

Comparing the descriptive statistics of the three models:

- All the three models have GCPs with ALA = 0.00, which means that all the models have at least one or more spots where model matches the reality.
- TOPO DEM has the best local accuracy with the lowest average ALA = 4.10 meters, followed by SRTM with ALA = 7.00 meters, and the least is ASTER with ALA = 11.96 meters.
- TOPO DEM has the least dispersed local accuracy with the index of dispersion = 3.09, closely followed by ASTER with the index of dispersion = 3.49, and the most dispersed is SRTM with the index of dispersion = 6.14, which means that the values of SRTM are almost twice as dispersed as those of TOPO and ASTER.

Furthermore, using the General QQ-Plot tool in Geostatistics, the ALA results of the three DEMs are compared each two alone, as shown in the Figure 48. The three plots show that the ALA do change with some consistency among the three models, but not very high one. Using the SPSS statistical package, the Pearson correlation is computed for these values. The results are shown in Table 8. Results show that the three models, as interpreted from the General QQ-Plots, have some positive correlations among each other, with values equals to: $R_{ASTER-SRTM} = 0.653$, $R_{ASTER-TOPO} = 0.741$, and $R_{SRTM-TOPO} = 0.589$. A comparison of the three coefficients shows that the local accuracy of ASTER and TOPO behave more consistently than either of them with SRTM.

Figure 52: General QQ-Plots comparing the ALA results of the three models (Left to Right: ASTER vs SRTM, ASTER vs TOPO, and SRTM vs TOPO) (Screenshots taken by Author)

Table 8: Pearson Correlation of the Local Accuracy between the three models using SPSS

		SRTM	ASTER	TOPO
SRTM	Pearson Correlation	1	.653**	.589**
	Sig. (2-tailed)		.000	.000
	N	1105	1105	1105
ASTER	Pearson Correlation	.653**	1	.741**
	Sig. (2-tailed)	.000		.000
	N	1105	1105	1105
TOPO	Pearson Correlation	.589**	.741**	1
	Sig. (2-tailed)	.000	.000	
	N	1105	1105	1105

** . Correlation is significant at the 0.01 level (2-tailed).

4.4.2 Model 2: Average Accuracy

After describing and interpreting the results of the individual GCPs for the three models in the previous sub-section, 80% of those points are chosen as per the suggested procedure in sub-section 4.3. The chosen points, which make a total of 877 GCPs, are used to compute the overall RMSE as the conventional method for estimating the accuracy of the overall area of study for a certain DEM.

The result of the calculation is summarized in the table below. Accordingly, TOPO DEM has the highest accuracy with an RMSE = 5.40 meters, followed by SRTM with RMSE = 9.65 meters, and the least accurate is ASTER with RMSE = 13.46 meters.

Table 9: Descriptive Statistics and RMSE of the Model-Generation GCPs' Subset

	SRTM	ASTER	TOPO
MIN	0.00	0.00	0.00
MAX	42.10	35.82	20.44
AVG	7.04	11.93	4.09
STD	6.60	6.40	3.52
SMSE	9.65	13.46	5.40

It is important to note, by comparing the results of the selected subset to the results of the full GCPs set that:

- In the three models, the subsets have exactly the same range as the full set.

- The average values in the three models are almost the same in both the subset and the full set, with a mere deviation of 0.6% for the SRTM, -0.3% for the ASTER, and -0.2% for the TOPO.

4.4.3 Model 3: Vario-Model

Before generating the vario-model as a GIS raster layer, the individual full set of GCPs are classified as per their spatially-coincident LULC classification, and the descriptive statistics are calculated for each class separately. The results of this classification are summarized in the table below.

Table 10: Descriptive statistics of the GCPs' ALA per LULC Classification

GCPs	Level 2											
	Class	11	12	13	21	22	24	31	32	33	41	51
	Count	218	36	20	353	212	25	38	38	130	3	32
SRTM Local Accuracy	MIN	0.00	1.80	0.00	0.00	0.00	0.97	0.90	0.00	1.43	0.82	0.95
		27.9	22.6	40.9	30.8	37.7	28.3	38.3	38.8			
	MAX	8	5	4	6	6	2	8	2	42.10	2.14	7.89
				13.4		5.9		14.1	14.5	10.2		2.5
	AVG	7.63	8.23	6	4.46	2	6.51	2	0	9	1.52	8
	STD	5.47	5.16	10.52	4.08	5.67	6.31	11.27	9.06	7.27	0.67	1.60
ASTER Local Accuracy	MIN	0.00	6.81	0.00	0.00	0.00	0.52	3.43	0.00	2.25	9.43	5.00
		31.6		29.0	24.4	31.2	26.6	29.2	35.8		10.0	17.0
	MAX	1	29.67	9	3	5	3	0	2	31.48	0	6
		11.3	14.0	15.7	10.1	11.1		17.1	18.6	15.0		9.6
	AVG	3	9	4	1	4	8.73	2	1	9	9.74	3
				10.4								
TOPO Local Accuracy	STD	7.80	5.76	2	3.27	6.30	8.23	7.51	8.27	6.38	0.29	2.67
	MIN	0.00	0.44	0.00	0.00	0.00	0.24	0.52	0.00	0.88	0.22	0.24
		19.4		20.4	15.8	16.5	11.4					
	MAX	3	18.79	4	3	4	9	14.21	15.66	15.65	0.60	6.22
				2.8	3.6						0.4	
	AVG	5.01	7.83	7.66	0	7	4.37	5.57	6.01	4.56	2	1.67
	STD	4.16	5.17	6.38	2.16	3.22	3.88	3.70	3.90	2.99	0.19	1.45

Comparing the average ALA for the three models per LULC classification is visualized in the graph below in Figure 53. The behavior of the three graphs shows that the average ALA varies similarly between the different land covers for the three models.

Figure 53: Graph showing the variation of the average ALA per LULC for the three models (ASTER in orange, SRM in blue, and TOPO in grey) (Graph generated by Author)

Beyond visual interpretation, and using the SPSS statistical package, the Pearson correlation is computed for the average ALA per LULC class between the three models. As shown in the table below, there is a high positive correlation between ASTER and SRTM, and between SRTM and TOPO, with a Pearson correlation $R = 0.922$ and $R = 0.805$ respectively. A positive correlation also exists between ASTER and TOPO, but is relatively weaker than the previous two, with a Pearson correlation $R = 0.692$

Table 11: Pearson Correlation of the average local accuracy between the three models using SPSS

		SRTM	ASTER	TOPO
SRTM	Pearson Correlation	1	.922**	.805**
	Sig. (2-tailed)		.000	.003
	N	11	11	11
ASTER	Pearson Correlation	.922**	1	.692*
	Sig. (2-tailed)	.000		.018
	N	11	11	11
TOPO	Pearson Correlation	.805**	.692*	1
	Sig. (2-tailed)	.003	.018	
	N	11	11	11

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Accordingly, it can be assumed that the land cover is an important factor in the local accuracy because even though there was not a high positive correlation between the overall GCPs ALA of the three models, as shown in Table 8, the correlation became higher when the data was classified per LULC, as shown in Table 11. Therefore, it can be concluded that even

though the accuracy of the three DEMs are different, it varies similarly when changing from one land cover into another. Given that distribution of the ALA values per LULC have a normal distribution in almost all the classes for the three models, the conclusion can be generalized on any other datasets. For a full list of the Normal QQ-Plot of the ALA values per LULC, please refer to Appendix A3.

Using the same 80% subset as used in sub-section 4.2 to calculate a single-value RMSE for the whole DEM, the RMSE is calculated at each LULC level 2 class for the three DEM. As shown in the table below, instead of a single RMSE, the SRTM now have an RMSE varying between 1.62 and 18.42 meters, the ASTER between 9.61 and 21.42 meters, and TOPO between 0.34 and 7.83 meters.

Table 12: RMSE values of three model as per the LULC classification

Level 2	SRTM	ASTER	TOPO
11	9.75	14.13	6.59
12	8.93	14.81	8.96
13	16.04	19.95	10.77
21	5.96	10.59	3.41
22	8.04	12.29	4.81
24	9.05	12.09	5.56
31	18.42	19.05	7.11
32	18.18	21.42	7.83
33	12.63	16.20	5.31
41	1.62	9.61	0.34
51	3.08	9.80	2.16

Accordingly, as previously explained step 3 in sub-section 4.2, the results in Table 12 are compiled into the input LULC dataset, and the three vario-model are generated, one for each input DEM having the same spatial resolution. The results are shown in the three maps below.

Figure 54: SRTM local accuracy vario-model map (Map generated by Author)

Figure 55: ASTER local accuracy vario-model map (Map generated by Author)

Figure 56: TOPO local accuracy vario-model map (Map generated by Author)

4.4.4 Models Assessment

To test the validity of the vario-models, whether it provides a more representative model of the accuracy compared to the assumed and RMSE values or not, the deviations of the individual GCPs validation subset is calculated as per the proposed procedure in sub-section 4.3 and compared.

SRTM:

For the SRTM, and as shown in the table below, the deviation from the assumed accuracy, which is defined as 20 meters, has a range from -20 meters to -0.57 meters, with an average deviation of -16.07 meters. The variation of this deviation is shown in Figure 57. The deviation from the single-value RMSE has a range from -9.65 meters to 31.29 meters, with an average deviation of -2.82 meters. The variation of this deviation is shown in Figure 58. Finally, the deviation from the proposed vario-model has a range from -18.42 meters to 28.01, with an average deviation of -2.13 meters. The variation of this deviation is shown in Figure 59. Comparing the average deviation from the RMSE and from the vario-model shows that the second one is closer to zero by 24%.

Table 13: Summary of the GCPs accuracy deviations for SRTM

Deviation From:	MIN	MAX	AVG	STD
Assumed Accuracy	-20.00	-0.57	-16.07	3.71
RMSE	-9.65	31.29	-2.82	6.39
Vario-Model	-18.42	28.01	-2.13	5.93

Figure 57: Histogram and Normal QQ-Plot of the SRTM accuracy deviation from the assumed one (Screenshots taken by Author)

Figure 58: Histogram and Normal QQ-Plot of the SRTM accuracy deviation from the RMSE (Screenshots taken by Author)

Figure 4-59: Histogram and Normal QQ-Plot of the SRTM accuracy deviation from the vario-model (Screenshots taken by Author)

ASTER:

For the ASTER, and as shown in the below table, the deviation from the assumed accuracy, which is defined as an average of 17.5 meters, has a range from -17.5 meters to 13.75 meters, with an average deviation of -5.94 meters. The variation of this deviation is shown in Figure

60. The deviation from the single-value RMSE has a range from -13.46 meters to 17.79 meters, with an average deviation of -1.90 meters. The variation of this deviation is shown in Figure 61. Finally, the deviation from the proposed vario-model has a range from -19.95 meters to 18.96 , with an average deviation of -1.73 meters. The variation of this deviation is shown in Figure 62. Comparing the average deviation of from the RMSE and from the vario-model shows that the second one is closer to zero by 7%.

Table 14: Summary of the GCPs accuracy deviations for ASTER

Deviation From:	MIN	MAX	AVG	STD
Assumed Accuracy	-17.50	13.75	-5.94	6.83
RMSE	-13.46	17.79	-1.90	6.83
Vario-Model	-19.95	18.96	-1.73	6.86

Figure 60: Histogram and Normal QQPlot of the ASTER accuracy deviation from the assumed one (Screenshots taken by Author)

Figure 61: Histogram and Normal QQ-Plot of the ASTER accuracy deviation from the RMSE (Screenshots taken by Author)

Figure 62: Histogram and Normal QQ-Plot of the ASTER accuracy deviation from the vario-model (Screenshots taken by Author)

TOPO:

For the TOPO, and as shown in the below table, the deviation from the assumed accuracy, which is defined as 3.33 meters, has a range from -3.33 meters to 16.10 meters, with an average deviation of -0.60 meters. The variation of this deviation is shown in Figure 63. The deviation from the single-value RMSE has a range from -5.4 meters to 14.03 meters, with an average deviation of -1.47 meters. The variation of this deviation is shown in Figure 64. Finally, the deviation from the proposed vario-model has a range from -10.77 meters to 12.84, with an average deviation of -1.27 meters. The variation of this deviation is shown in Figure 65. Comparing the average deviation of from the RMSE and from the vario-model shows that the second one is closer to zero by 13%.

Table 15: Summary of the GCPs accuracy deviations for TOPO

Deviation From:	MIN	MAX	AVG	STD
Assumed Accuracy	-3.33	16.10	0.60	3.71
RMSE	-5.40	14.03	-1.47	3.71
Vario-Model	-10.77	12.84	-1.27	3.73

Figure 63: Histogram and Normal QQ-Plot of the TOPO accuracy deviation from the assumed one (Screenshots taken by Author)

Figure 64: Histogram and Normal QQ-Plot of the TOPO accuracy deviation from the RMSE (Screenshots taken by Author)

Figure 65: Histogram and Normal QQ-Plot of the TOPO accuracy deviation from the vario-model (Screenshots taken by Author)

4.5 Discussion

Based on the found results from the previous section, the following observations can be made:

- The assumed accuracy for the SRTM and the ASTER DEMs had the biggest deviation among the three accuracy models. Moreover, as shown in Table 13, the deviation range of the SRTM is negative, meaning that all the validation GCPs have better accuracy than the assumed one. However, this cannot be fully generalized about the SRTM because, as shown in Figure 53, the deviations do not have a normal distribution especially at the edges of the range. On the other hand, assumed accuracy for the TOPO DEM had the smallest average deviation among the three models. However, similar to the previous deduction, it cannot be fully generalized because, as shown in Figure 59, the deviations do not have a normal distribution either.

- Despite the previous observation and its deductions, it is important to note that the accuracy of the first model, the assumed one, is assumed to be a global one, while the two other models are based on the local one that was previously defined in section 4.1. Therefore, comparing the first model to the other two will not be as useful as comparing the other two to each other.
- Comparing the vario-model to the RMSE shows that the first one is more representative than the others in the three DEMs, where the average deviation was closer to zero, where zero, meaning the accuracy of the model, is equal to the accuracy of the individual validation GCP by 24% for the SRTM, 7% for the ASTER, and 13% for the TOPO. Furthermore, looking at the QQ-Plots in Figures 57, 60, and 63 shows that the deviation of the GCPs from the vario-model of the three DEMs have a normal distribution at the majority of the ranges. This means that a general deduction can be made from their results. That is, areas with similar land cover tend to cause a DEM to have similar or close accuracy compared to land covers, and adopting a vario-model will definitively lead to a better representation of the accuracy than the conventional method of a single RMSE.

With the previous conclusion made, and going back to the results depicted in Table 11 and Figure 14, additional observation can also be made:

- Forest and Scrub and/or Herbaceous Vegetation Association are the two LULC classes that have the lowest local accuracy in the three DEMs, followed by Industrial, Commercial and Transport Units, and Mine, Dump and Construction Sites. This observation is valid because these land covers represent the areas that have the least uniform terrain, with many sudden changes in small areas, so that even DEMs with the highest spatial resolution will always miss some detail.
- On the other hand, wetlands and water bodies are the two LULC classes that have the highest local accuracy in the three DEMs. This observation is also logically valid because these two types of land cover are usually flat due to the nature and mechanics of water. The lack of terrain features and high changes will not cause any DEM to miss important terrain details, and, therefore, will have high accuracy in those areas.

For better interpretation and understanding of the terrain from the different land covers, refer to Appendix C, which is a photo gallery showing sample areas of the main land covers from different parts of the area of study.

While, as far as the research of published articles conducted by the author of this thesis, no similar models have been proposed yet, other previous studies accounted for the different types of land covers when assessing the accuracy of a DEM, even if sometimes not to a formal LULC classification. One of these studies is done by (Sertel, 2010) in Turkey and showed that the ASTER GDEM has high accuracy in mining areas and low agricultural fields and the lowest accuracy in industrial areas. While the exact figures differ between the two studies, both of them showed similar behavior, where the three DEMs under study have lower accuracy in industrial areas compared to the agricultural fields.

4.6 Summary

In this chapter, a new procedure was proposed to define the local accuracy of a DEM. Using the conventional method for assessing the accuracy of a DEM based on a set of GCPs resulted in a local accuracy for each of the three models equal to: 9.65 meters for SRTM, 13.46 meters for the ASTER, and 5.4 for the TOPO. On the other hand, using the proposed vario-model resulted in a range of accuracy as follows: SRTM vario-accuracy ranges between 1.62 meters and 18.42 meters, as shown on the map in Figure 50; ASTER ranges between 9.61 meters and 21.42 meters, as shown in Figure 51; and TOPO ranges between 0.34 meters and 10.77 meters, as shown in Figure 52.

Using another sub-set of GCPs, referred to as the validation GCPs, the assumed accuracy, calculated RMSE, and calculated vario-models were assessed and compared. Results showed that the vario-model provides a better representation of the local accuracy than the conventional single-value RMSE, where the validation GCPs had their average individual accuracy closer to the vario-model than the single value one by 24% for the SRTM, 7% for the ASTER, and 13% for the TOPO.

Chapter 4 Bibliography

- Sertel, E. (2010). Accuracy Assessment of ASTER Global DEM over Turkey. *ISPRS Technical Commission IV*. Orlando, Florida.
- Su, J., & Bork, E. (2006). Influence of Vegetation, Slope, and LiDAR Sampling Angle on DEM Accuracy. *Photogrammetric Engineering and Remote Sensing* 72(11), 1265-1274.

Chapter 5

Applications of the Vario-Model

A DEM, as an end product, is only useful for visualizing the terrain in perspective view. However, the uses of a DEM extend far beyond 3D visualization to perform various analyses such as terrain morphology, volumetric analysis, visibility and line of sight, and hydrologic analysis.

Like the input data, any derived dataset has an accuracy; modeling this accuracy into quantifiable figures is crucial to assess these datasets before using them. Assuming that the processing procedures that generate the derived datasets do not introduce additional uncertainty into the output, and based on the concept of Garbage In – Garbage Out (Quinion, 2015), the accuracy of the input DEM is propagated into the accuracy of the derived dataset. However, the challenge rises as the accuracy of the DEM cannot be assigned as-is to the derived product, mainly because the derived product, as previously mentioned, has a different unit than the input. For example, the accuracy of the slope derived from a DEM with an accuracy of 10 meters cannot be equal to the same value, because the accuracy of the slope should be of the same unit as the slope, which is measured in vertical angle. A more sophisticated derived output is delineated streams, where the accuracy of this output is measured as linear deviation in the plane.

Previous research, such as (Zhou & Liu, 2004), has tackled the accuracy of terrain-based analysis with a focus on various algorithms, but with the assumption of a fixed accuracy for the input DEM, and has been tested using well-defined mathematical models, which do not reflect the true nature of DEMs representing real terrain. The objective of this chapter is to propose a model to propagate the accuracy of a DEM that has been modeled into a vario-model and to assess the accuracy of the products that are the result of performing terrain-based analysis on this DEM in order to quantitatively and systematically assess the accuracies of these products.

5.1 What is a Slope?

Mathematically, a slope of a line in a two-dimensional Euclidean system describes the steepness of this line, denoted as the ratio of the vertical change over the horizontal change (Clapham & Nicholson, 2009). The slope of a line is also denoted as the angle of that slope, which is calculated using trigonometric functions as described in the Figure 66 and equation 5-1 below.

Figure 66: Mathematical 2D graph representing the slope (Graph generated by Author)

$$\alpha = \tan^{-1} \frac{H}{L} \quad (5-1)$$

Where:

- H: Height difference between two points, measured in any linear unit
- L: Plane distance between the two points, measured in the same unit as H
- α : Slope angle, measured in radian

Applied to DEM, the slope function in GIS calculates the slope for each cell in the input DEM as the maximum rate of height change between that cell and its eight surrounding cells based on 3x3 cell neighborhood (Burrough & McDonell, Principles of Geographic Information Systems, 1998). Accordingly, the degree of the slope α will be calculated in a simplified form as follows:

$$\alpha = \tan^{-1} \frac{\Delta_z}{\text{Cell Size}} \quad (5-2)$$

Where

- Δ_z : The difference in height between the center cell and the one with the maximum difference
- Cell Size: The distance between the centers of two adjacent cells, or the spatial resolution.

50	45	50
30	30	30
8	10	10

Figure 67: Sample 3x3 matrix to calculate the steepest slope for the central cell with the surrounding cell with the maximum height difference (Extract from ArcGIS 10.2 Help)

The slope dataset, which is the combination of slope calculation of all the cells that forms the input DEM, is an important indicator to analyze a terrain. The slope dataset is directly used in performing terrain morphology such as landslide and flood hazard detection, construction sites suitability analysis, and landscaping analysis as described in early research such as (Wu, Li, & Huang, 2008) and (Mutluoglu, 2010). Furthermore, several types of analysis are based on various terrain-derived datasets that are based directly or indirectly on the slope dataset rather than on the actual DEM directly. Examples of these derived datasets are:

- **Aspect:** Calculates the direction of the slope at each cell of the terrain. Not only the aspect is derived from the slope, but both datasets are jointly used to perform various analyses.
- **Curvature:** Calculates the plan and profile curvature by calculating the second derivative of terrain, which is the slope of slope. Curvature describes the physical characteristics of the terrain to understand erosion and runoff processes.
- **Hillshade:** Calculates the shaded relief from the surface, taking as parameters the illumination source angle and shadows, which has the objective of providing a 3D perspective vision through a 2D dataset. The calculation of the hillshade is directly based on the slope and aspect datasets of the surface.
- **Flow Direction:** Calculates for each cell the flow from itself to its steepest downslope neighbor. While the algorithm behind the flow direction is based on hydrologic foundations, it is almost the same in nature to the aspect dataset, where both of them are computed based on the slope.
- **Flow Accumulation:** Calculates for each cell the weight accumulated flow into that cell. The calculation of the flow accumulation is directly based on the flow direction

dataset and, therefore, is indirectly based on the slope. Both flow direction and flow accumulation datasets are the basis of hydrological analysis.

Back to the objective of this research and given the direct link between the different presented datasets and the slope, the accuracy and significance of any of these datasets are linked to the accuracy and significance of the slope.

5.2 Accuracy and Significance of Slope

A DEM is a modeled representation of the terrain. Similar to any measurement taken in any field of science and engineering, and irrespective of the technique used for data collection and modeling, the height values stored inside the DEM contain some errors with respect to reality. The error will be referred to the accuracy of the measurement. Therefore, for each cell, the modeled height value can be denoted as follows:

$$Z_{model} = Z_{actual} \pm \varepsilon_z \quad (5-3)$$

Where:

- Z_{model} : The height of the location as modelled in the DEM, expressed in any linear unit,
- Z_{actual} : The actual height of the location in reality, expressed in the same unit as Z_{model} ,
- ε_z : The global or absolute accuracy of each cell of the DEM.

When comparing the height difference between two points, it will be calculated as follows:

$$\Delta Z_{model} = \Delta Z_{actual} \pm \varepsilon_{\Delta z} \quad (5-4)$$

Where:

- ΔZ_{model} : The height difference as modelled in the DEM, expressed in any linear unit
- ΔZ_{actual} : The actual height difference in reality, expressed in the same unit as ΔZ_{model}
- $\varepsilon_{\Delta z}$: The relative or local accuracy for each cell of the DEM with respect to its neighbor

Therefore, the modeled slope will be described as follows:

Figure 68: Mathematical 2D graph of the modelled slope and its error (Graph generated by Author)

Similar to the Equations 5-3 and 5-3, the slope modelled is represented as follows:

$$\alpha_{model} = \alpha_{actual} \pm \varepsilon_{\alpha} \quad (5-5)$$

Combining the Equations 5-1 and 5-5 will result in:

$$\alpha_{model} = \tan^{-1} \frac{H \pm \varepsilon_H}{L} \quad (5-6)$$

By developing the above trigonometric function, the accuracy of the slope is calculated as follows:

$$\varepsilon_{\alpha} = |\tan^{-1}(\tan \alpha_{model} + \tan \varepsilon_{\alpha}) - \alpha_{model}| \quad (5-7)$$

Equation 5-7 applied to the whole terrain on a cell-by-cell basis:

$$\varepsilon_{\alpha} = \left| \tan^{-1} \left(\tan Slope(DEM) + \frac{\varepsilon_{\Delta z}}{Cell\ Size} \right) - Slope(DEM) \right| \quad (5-8)$$

Where:

- $Slope(DEM)$: The slope dataset derived from the input DEM, expressed in radian
- $\varepsilon_{\Delta z}$: The relative accuracy of the DEM, expressed in any linear unit
- $Cell\ Size$: The cell size of the DEM, expressed in same linear unit as $\square_{\Delta}$
- ε_{α} : The accuracy of the slope dataset, expressed in the same angular unit as the slope

The significance of the slope or of any derived value is the level of accuracy to which this modeled value represent the actual one. The significance represents the percentage of how much the modeled value is close to the actual one and is calculated as follows:

If the error of the slope is less than the slope value, then the significance is equal to:

$$Sig(\alpha_{model}) = 1 - \frac{\varepsilon_{\alpha}}{\alpha_{model}} \quad (5-9)$$

Or, if the error is equal to or exceeds the slope value, then the significance is equal to:

$$Sig(\alpha_{model}) = 0 \quad (5-10)$$

5.3 Application in Hydrology

GIS is a key tool in performing digital hydrologic analysis, and many special packages are available that contain dedicated tools for such work. However, the most common hydrological analyses in GIS are: delineating drainage systems, delineating watersheds, locating water outlets, and calculating the volume of a reservoir (natural or man-made). According to (Wechsler, 2007), (Cui, Li, Huang, & Huang, 2014), and (Hutchinson, Xu, & Stein, 2011), the DEM is the most important spatial data source for deriving hydrologic models and performing analysis as it models the topography of the terrain which drives the flow of the surface water.

5.3.1 Area of Study: The Litany River

Lebanon has sixteen rivers both permanent and seasonal. Among these, the Litany River was chosen to be the case study for this project. According to the Litany River Authority (The Characteristics of the Litani River, 2015), the Litany River is the largest river in Lebanon with a total length of 170 km. It is fed by sixteen tributaries, provides irrigation water for the Beqaa and the South, and it has the largest water basin in the country, spanning over approximately 20% of the total country with an area of 2,175 Km².

Terrain-wise, this river is interesting as a case study as it spans over different types of terrain; in fact, the river is divided into two main parts, split by a large artificial lake called the Qaraoun Lake: The Upper Litany starts near Baalbek city in northern part of the Beqaa and flows down south in the middle of the valley until it reaches the Qaraoun dam. This terrain is mainly plain and mostly cultivated and semi-urban areas. The Lower Litany, on the other hand, proceeds from the Qaraoun dam half way south, then veers west toward the Mediterranean Sea, where it ends. While the first part of the Lower Litany has almost the same land cover as the Upper Litany with a slightly rougher terrain, the second part runs with very steep slopes with less cultivated areas until it reaches the coast line, where the terrain flattens again.

Figure 69: Map showing the extent of the Litany River Basin (Map designed by Author)

5.3.2 Input Data

Statistically, SRTM values spans between -2 and 2,611 meters, with an average height of 1,080.06 meters and a standard deviation of 467.87 meters. ASTER values span between 0 to 2,604 meters, with an average height of 1,073.03 meters and a standard deviation of 470.3 meters. Finally, TOPO values span between 0 and 2,622 meters, with an average height of 1,081.4 meters and a standard deviation of 468.53 meters. Comparing the three DEMs shows that TOPO has a slightly bigger range compared to SRTM and ASTER, and the average height too. Additionally, calculating the correlation between them shows that there is a high positive correlation among all of them, as shown in the table below. Comparing the correlation matrix for the Litany River Basin to the one of the overall country, Table 1 in chapter 3 shows that the correlations behaved similarly where in both tables, the pair of ASTER and TOPO has the highest correlation among all the pairs.

Figure 70: Altitude variation within the river basin according to SRTM (Map designed by Author)

Table 16: Correlation Matrix of the three DEMs

	ASTER	SRTM	TOPO
ASTER	1.00000	0.98783	0.99113
SRTM	0.98783	1.00000	0.98050
TOPO	0.99113	0.98050	1.00000

5.3.3 Detecting River Streams

Using the input data described in the above sub-section, this sub-section will cover the required steps to generate the river streams automatically using Esri ArcGIS 10.2 with Spatial Analyst extension.

Step 1: Detect and Remove Sinks

Before using the input DEM for hydrological analysis, it must be checked for unwanted sinks in order to remove them before proceeding with the analysis. By definition, a sink is a sudden depression in the modeled terrain, as shown in the figure below. In some cases, a sink could represent an actual depression in the terrain such as a man-made reservoir, areas protected by supporting walls, or excavations; however, in most of the cases, a sink is due to an error in the DEM. This error could be the result of a faulty sensor reading or an abnormal interpolation of the terrain samples (Mark, 1988).

Figure 71: Profile view of a sample modeled sink

Irrespective of reasons behind erroneous sinks, they must be detected and eliminated, as these sinks will cause abnormal hydrological results and will end up having a discontinuous and deviated drainage network. A sink is detected as a cell whose neighboring cells are higher or when two cells flow into each other causing a hydrological loop.

In ArcToolbox, the *Sink* tool is used to detect and show the user the locations of the sinks. It takes the DEM as an input raster and returns an output raster showing the sink's locations. In the next sub-step, the *Fill* tool is used to fill the sinks that have a depression higher than the specified limit.

Step 2: Calculate Flow Direction

Once the DEM is depression-less, after removing all faulty sinks, the next step is to calculate the flow direction. The flow direction is a derived dataset that calculates for each cell from the input DEM, the direction of its steepest slope toward one of its eight neighboring cells. This eight-direction flow model follows the approach presented in (Jenson & Domingue, 1988). In ArcToolbox, the *Flow Direction* tool takes the DEM as an input and generates the flow direction as a raster output as shown for the SRTM in Figure 71.

Step 3: Calculate Flow Accumulation

After the generation of the flow direction dataset, the flow accumulation is created as the accumulated count of cells into each cell of the original DEM. Cells with high flow accumulation values are considered areas with high flow activities and could be considered as part of the drainage stream network (Jenson & Domingue, 1988).

Figure 72: Flow Direction derived from SRTM (Map designed by Author)

In ArcToolbox, the *Flow Accumulation* tool takes the flow direction as an input, and optionally a weight raster. If a weight raster is used, the output accumulation would be the accumulation of the weights, such as the precipitation, and the unit of the output will be relative to the unit of the weight raster. If no weight raster is used, each cell of the input flow direction will be given the value '1', and the output dataset will not represent the accumulated count of flowing cells into each cell, unit-less, as processed and shown in the figure below.

Figure 73: Flow accumulation derived from SRTM (Map designed by Author)

Step 4: Detect and Order Streams

After the generation of the flow accumulation dataset, the stream is chosen by selecting the cells from the flow accumulation, where the value exceeds a specific threshold. Because the output of the previous step is unit-less and only representing the count of accumulated cells, there is not any standard value to use as a reference. Instead, after several trial-and-error attempts, the threshold used was computed so as to choose the top 2.5% cells with the highest flow accumulation values.

After the streams are detected and stored in a raster as logical value, '1' for stream and '0' for no-stream, these streams need to be ordered to preserve their flow order and to be classified as main or tributary streams. Two methods for stream ordering exist, one created by Strahler in 1957 and another by Shreve in 1966. The Strahler method, which is the most commonly used one, starts with the streams that do not have any tributaries and assigns to them the value '1' and refers to them as first-order segments. The segment order increases whenever two segments of the same order intersect. For example, when two first-order segments intersect, it will create a second-order one, and when two second-order segments intersect, it will create a third-order one, etc. The Shreve method starts the same way as Strahler, but adds up the orders. For example, if a first-order and two-order segment intersect, it will create a third-order segment. Therefore, this method is considered to calculate the magnitude of the streams instead of their order (Tarboton, Bras, & Rodriguez-Iturbe, 1991).

In Arctoolbox, the *Stream Order* tool takes as input the detected streams from the previous sub-step and the flow direction from step 2 and returns the ordered streams with the option to choose the ordering method. The output ordered streams are based on the Strahler method and are shown in the figure below.

Figure 74: Stream ordering based on Strahler Method (Map designed by Author)

Step 5: Convert Streams into Vector

The final step in delineating the river streams is to convert the detected and ordered streams from a raster dataset into vector polyline. In ArcToolbox, the *Stream to Feature* tool converts the raster representation of the streams into vector polyline. It takes as an input the stream raster generated in step 4 and the flow direction generated in step 2 and returns the output stream as polylines. The tool also has an option to simplify the output, where the lines are weeded to reduce the number of vertices based on the Douglas-Puecker Algorithm for line generalization (Tarboton, Bras, & Rodriguez-Iturbe, 1991). The output vector streams are shown in the figure below.

Figure 75: Derived streams converted into Vector Polylines (Map designed by Author)

5.3.4 The Reference Model

In order to test the accuracy of the produced streams from the input DEMs, a reference model is created by 2D digital photogrammetry, where the Litani River and its main tributaries (Berdawni, Ghzayyel, Hafir, Jaiir, Hala – Yahfoufa, and Wadi el Debbe) were digitized as center lines from IKONOS satellite imagery. The background satellite image has a spatial resolution of 0.8 meters. Compared to the ones of TOPO (5 meters), ASTER (30 meters) and SRTM (90 meters), the satellite image is considered high resolution and, therefore, its output will be considered as a reference, and any error is relatively minimal and will be neglected.

Figure 76: Mapping River Stream using IKONOS Satellite Image (Screenshot taken by Author)

To assess the results of the hydrologic analysis, the output stream will be compared to the reference on, in terms of planar displacement. This is achieved by performing the following steps:

1. Assign the name of the segment to each of the generated streams. This step is achieved by adding a new text field in the streams layer and manually naming each of the streams. As shown in the figure below.

Table				
Litani (ASTER)				
FID	Shape	Length	Type	Name
7	Polyline	72449.315706	Main	Lower Litani
0	Polyline	72399.935074	Main	Upper Litani
3	Polyline	11618.716669	Tributary	Berdawni
5	Polyline	2282.793538	Tributary	Ghzayyel
4	Polyline	6499.137797	Tributary	Hafir
1	Polyline	30124.308825	Tributary	Hala - Yahfoufa
2	Polyline	14083.550967	Tributary	Jalir
6	Polyline	12640.072371	Tributary	Wadi el Debbe

Figure 77: ASTER-derived Streams' Table of Attributes (Screenshot taken by Author)

2. Generate Euclidean distance dataset from each reference segment. This is achieved by using the *Euclidian Distance* tool in ArcToolbox, which generates a raster representing the straight-line distance at each location to the river stream as shown in the figure below.

Figure 78: Generated Euclidian distance for the Lower Litany stream (Screenshot taken by Author)

3. Convert the generated river streams from polylines into points, with a distance between each two points equal to the cell size of the related input DEM. This step is achieved using the *Features Vertices to Points* tool in ArcToolbox. The result is a set of points representing each stream as shown in the figure below.

Figure 79: Streams converted from polylines into points (derived from SRTM) (Screenshot taken by Author)

4. Assign for each point the distance to its related reference segment and the significance coincident its location. This last step is achieved by using the *Extract Values to Points* tool in ArcToolbox that takes as an input a point vector layer and a raster dataset and assigns for each point in a new field the coincident value from the raster. This tool is applied twice to the generated points from step 3: once with the distance raster datasets generated in step 2, and the second time with the input significance vario-models. The result is a combination of points derived from the three DEMs and combined with their related significance values and distance to their reference streams as shown in the map below.

Figure 80: Results of calculating the deviations of the modeled streams (Map designed by Author)

5.4 Results

Similar to the structure of chapter 4, this section describes the results of the proposed procedures of defining the slope procedure and significance and the results of the hydrological application of the vario-model that was applied into the area of the Litany River Basin. This section is divided into four main sections: first section for the results of the slope accuracy, the second for the results of the slope significance, the third for the results of the river streams delineation, and the fourth for the correlation of the three previous sub-sections.

5.4.1 Slope Accuracy

After applying the proposed procedure in section 5.2, the slope accuracy of the three DEMs gave the following results: SRTM has the slope accuracy ranging between 0.81 and 11.57 degrees, with an average accuracy of 6.98 degree. ASTER has the slope accuracy ranging between 1.06 and 35.52 degrees, with an average accuracy of 23.6 degrees. And finally, TOPO has the slope accuracy between 0.17 and 65.10 degrees, with an average accuracy of 38.62 degrees, as shown in the table below.

The results show a noticeable difference between the three DEMs, where SRTM has much higher slope accuracy, followed by ASTER, and TOPO, having the least accuracy. This high difference is mainly due to the propagation of the input DEM accuracy and adding to it the cell size which has an inverse weight on the results.

Table 17: Summary of Slope Accuracy

Layer	MIN	MAX	AVG	STD
SRTM	0.81	11.57	6.98	2.32
ASTER	1.06	35.52	23.60	4.47
TOPO	0.17	65.10	38.62	7.35

Also, comparing the behavior of these three slope accuracy models shows that they vary independently of each other, where the coefficient of correlation shows a weak positive correlation among each of the pair of models. As shown in the table below, all the correlations are positive, meaning that there is a high probability that the slope accuracy will either increase or decrease in the three models at the same time. However, the change in the accuracy is not the same in the three models, as the correlations are weak. Furthermore, comparing the values in the matrix shows that TOPO slope accuracy has the least correlation with the other two DEMs.

Table 18: Slope Accuracy Correlation Matrix

	SRTM	ASTER	TOPO
SRTM	1.0000	0.6464	0.2662
ASTER	0.6464	1.0000	0.4894
TOPO	0.2662	0.4894	1.0000

5.4.2 Slope Significance

Finally, the significance of the slope for both DEMs gave the following results: the three DEMs have almost a full range from 0 to 1, with SRTM having the highest average of 0.28, followed by ASTER with 0.04, and TOPO as least significant with 0.02.

Figure 81: Slope accuracy map of SRTM (Map designed by Author)

Figure 82: Slope accuracy map of ASTER (Map designed by Author)

Figure 83: Slope accuracy map of TOPO (Map designed by Author)

Table 19: Summary of Slope Significance

Layer	MIN	MAX	AVG	STD
SRTM	0.00	0.97	0.28	0.29
ASTER	0.00	0.99	0.04	0.13
TOPO	0.00	1.00	0.02	0.09

Similar to the analysis of results in 5.4.1, the slope significance correlation matrix is computed like the one of the slope accuracy in Table 18. The results shown in the table below are also similar to those in Table 18, where all the correlations are positive, however weak.

Table 20: Slope Significance Correlation Matrix

	SRTM	ASTER	TOPO
SRTM	1.0000	0.3858	0.2009
ASTER	0.3858	1.0000	0.3630
TOPO	0.2009	0.3630	1.0000

The map below shows the results of the slope significance for three DEMs:

Figure 84: Slope significance map for SRTM (Map designed by Author)

Figure 85: Slope significance for ASTER (Map designed by Author)

Figure 86: Slope significance map for TOPO (Map designed by Author)

Visual inspection of the above maps clearly shows that using SRTM will lead to significant results for any derived terrain-based analysis in the majority of the country, with the exception of few coastal regions on the western border and in the middle of the Bekaa Valley in the center. On the other hand, using ASTER and TOPO will only lead to significant results on steep slopes. Also, it is noticeable that because the TOPO DEM contains major areas that are flat, it resulted in a lack of significant data due to divide-by-zero error.

5.4.3 Correlation Analysis

In this research, several datasets were involved in the process, as input, intermediate, or output data. These datasets are: DEM, DEM local accuracy, slope, slope accuracy, and slope significance. The Pearson correlation is created for each dataset pair in the three products. The results are summarized in the below three tables 21, 22 and 23.

Table 21: Correlation Matrix of SRTM datasets

	DEM	Local Accuracy	Slope	Slope Accuracy	Slope Significance
DEM	1.0000	0.1973	0.1768	0.1857	0.1097
Local Accuracy	0.1973	1.0000	0.4524	0.9486	0.0890
Slope	0.1768	0.4524	1.0000	0.2138	0.7742
Slope Accuracy	0.1857	0.9486	0.2138	1.0000	-0.1461
Slope Significance	0.1097	0.0890	0.7742	-0.1461	1.0000

Table 22: Correlation Matrix of ASTER datasets

	DEM	Local Accuracy	Slope	Slope Accuracy	Slope Significance
DEM	1.0000	0.0212	0.1488	0.1359	0.0453
Local Accuracy	0.0212	1.0000	0.4081	0.6838	0.1416
Slope	0.1488	0.4081	1.0000	-0.3226	0.7374
Slope Accuracy	0.1359	0.6838	-0.3226	1.0000	-0.4811
Slope Significance	0.0453	0.1416	0.7374	-0.4811	1.0000

Table 23: Correlation Matrix of TOPO datasets

	DEM	Local Accuracy	Slope	Slope Accuracy	Slope Significance
DEM	1.0000	0.0261	0.1621	-0.0766	-0.0305
Local Accuracy	0.0261	1.0000	0.3735	0.4841	0.0709
Slope	0.1621	0.3735	1.0000	-0.5934	0.5368
Slope Accuracy	-0.0766	0.4841	-0.5934	1.0000	-0.4736
Slope Significance	-0.0305	0.0709	0.5368	-0.4736	1.0000

As a first observation, the correlations behaved very similarly for three DEMs, even though the individual values of these datasets were not consistent at all between them. Accordingly, any deduction from these matrices concerning the correlation between any two datasets will be considered product-independent. Some interesting deductions:

- The DEM has a low correlation with all the other datasets. This is due to the fact that the calculation of the slope is based on the relative difference of heights between a cell and its neighbors, irrespective of the cell absolute height, which is generally in reference with the mean sea level.
- In the opposite of the first point, the DEM accuracy is highly correlated to the slope accuracy. In fact, there should be a high correlation between the two datasets, as the second one is dependent on the first, as in the procedure described in section 5.2.
- The most interesting observation is that there is a low correlation between the slope accuracy and the slope significance, while the proposed equation in section 5.2 can be misleading, indicating that there should be a high correlation, as the slope significance is dependent on the slope accuracy. However, the slope significance has a relatively positive high correlation with the actual slope.
- The correlation between the slope accuracy and the local accuracy is affected by the weight of the cell size; however, the cell size has an inverse weight on the correlation between the slope accuracy and the slope itself. Therefore, it can be concluded that depending on the spatial resolution, the bigger the cell size is, the more the slope accuracy is affected by the local accuracy and less by the slope, and the smaller the cell size becomes, the more the slope accuracy become more affected by the slope and less affected by the local accuracy.

5.4.4 Length of Segments

The first step in assessing the results involves looking at the total length of the segments and comparing them to the reference ones. The table below summarizes the lengths that are measured in meters. The overall difference is computed as a percentage with respect to the reference segment.

Table 24: Summary of segments' lengths, expressed in meters

Type	Name	IKONOS	SRTM		ASTER		TOPO	
Main	Lower Litany	86,215.92	76,068.59	12%	72,449.32	16%	83,419.12	3%
Main	Upper Litany	71,452.59	71,090.44	1%	72,399.94	1%	71,116.20	0.5 %

Tributary	Berdawni	10,821.56	12,085.35	12%	11,618.72	7%	9,149.14	15%
Tributary	Ghzayyel	16,531.92	3,230.03	80%	2,282.79	86%	16,801.48	2%
Tributary	Hafir	10,183.13	8,547.21	16%	6,499.14	36%	9,767.64	4%
Tributary	Hala – Yahfoufa	17,594.56	27,122.83	54%	30,124.31	71%	17,233.43	2%
Tributary	Jaiir	11,595.37	14,566.64	26%	14,083.55	21%	9,684.13	16%
Tributary	Wadi el Debbe	12,824.46	12,187.89	5%	12,640.07	1%	13,091.78	2%
		237,219.52	224,898.98	5%	222,097.83	6%	230,262.91	3%

Comparing the results of three DEMs showed that all the results in summary had close results to the reference ones, where the modeled segments were shorter on average by 5%, 6%, and 3% for the SRTM, ASTER, and TOPO, respectively. Also, considering each segment alone, the modeled segments from SRTM and ASTER behaved quasi-similarly in the difference, where between both models out of eight segments, five segments had differences of equal or less than 5%. On the other hand, the ones modelled from TOPO had much lower differences, especially those tributaries (Ghzayyel, Hafir, Hala – Yahfoufa, and Jaiir) in the Beqaa Valley.

Figure 87: Generated river streams (SRTM in magenta, Aster in orange, and TOPO in light green) versus the reference one (in blue) (Screenshot taken by Author)

The Hala – Yahfoufa and Ghzayyel tributaries that are derived from SRTM and ASTER not only have a high discrepancy in the length of the streams, but also the low accuracy in the flat terrain lead to distortion in the connectivity of the streams. In fact, as shown in the Figure 87, the Hala – Yahfoufa tributary is expected to flow south-west until it joins the Upper Litany River at point A, near the town of Delhamieh. However, due to the input DEM inaccuracy, the streams derived from SRTM and ASTER deviated at point B toward the south near the town of Rayaqa and created a fictive stream that will join the Ghzayyel tributary at point C, near the town of Deir Zeinoun. On the other hand, the streams derived from TOPO were more representative of the network.

5.4.5 Deviation of Segments

The second step in assessing the results is the consideration of deviation of the nodes from the reference segments. The table below summarizes the variation of the deviations for the results from the three models.

Table 25: Summary of the nodes' deviation, expressed in meters

	SRTM	ASTER	TOPO
MIN	0.00	0.00	0.00
MAX	9,219.60	2,983.40	2,331.40
AVG	535.69	340.60	140.72
STD	1,247.30	601.88	333.58

The results of the SRTM appeared to have the highest deviation from the reference model, with an average deviation of 535.69 meters, followed by the ASTER with an average one of 340.6 meters. TOPO had the best results with the least deviation with an average one of 140.72 meters. Going into the details of the variations in the three models, as shown the below Figures 88, 89, and 90, the majority of the deviations from the three models are skewed toward zero, and these values do not form a normal distribution. Therefore, at this stage, any conclusion that will be made shall be restricted to these datasets and cannot be generalized.

Vario-Model for Estimating and Propagating DEM Vertical Accuracy: Case of Lebanon

Figure 88: Histogram and normal QQ-Plot for the deviation of the nodes generated from SRTM
(Screenshots taken by Author)

Figure 89: Histogram and normal QQ-Plot for the deviation of the nodes generated from ASTER
(Screenshots taken by Author)

Figure 90: Histogram and normal QQ-Plot for the deviation of the nodes generated from TOPO
(Screenshots taken by Author)

Back to the comparison of the results between the three models, it is obvious that the results derived from SRTM have the highest deviation, and those from TOPO have the least simply because the SRTM has a spatial resolution of 90 meters, compared to the ASTER with 30 meters, which is three times more detailed, and TOPO with 5 meters, which is eighteen times more detailed. To take out this factor, in order to better compare the results, and try to deduce a more general conclusion, the deviation unit is converted from meters into cell count. Accordingly, the values at each node from SRTM will have a new deviation computed as the old one divided by 90 and floored; those from ASTER will be computed as the old

deviation divided by 30 and floored; and finally, those from TOPO will be computed as the old deviation divided by 5 and floored. The computations include the floor operation, which converts the results from a decimal whole number into the largest previous integer. This operation is included because if the deviation between a node and its reference is less than the cell size, then it is assumed that the node is inside the same cell as its reference and, therefore, there is no deviation, because for any raster model, the cell represents the smallest unit and any details within it are not taken into consideration.

For example, if taking the case of SRTM: if node A has a deviation of 55 meters and node B of 150 meters. Converting the deviations of the two nodes as previously described will result in node A having a deviation equal to zero, and node B having a deviation equal to one. This means that node A is inside the same cell as its reference, and, therefore, no deviation should be considered, while node B is one cell away from its reference.

Accordingly, the table below summarizes the variation of the deviations for the results from the three models, expressed in cell count:

Table 26: Summary of the nodes' deviation, expressed in cell count

	SRTM	ASTER	TOPO
MIN	0	0	0
MAX	102	99	466
AVG	5	11	28
STD	14	20	67

Contrary to the results presented in Table 25, SRTM has the least deviation, with an average one of 5 cells, followed by ASTER with an average of 11 cells, and the highest deviation by TOPO with an average one of 28 cells. This means that the deviation in ASTER is almost double that of the one in SRTM, and in TOPO is almost six times greater than the one in SRTM.

5.4.6 Linking Deviation to Significance

After assessing the deviations of the nodes, the third step is to compare the deviations to the slope significance in order to assess whether a correlation exists or not. As shown in the table below, the first sub-step is to assess the significance at the nodes for the three models. Results show that the average significance for SRTM is 0.15, for ASTER is 0.01, and for TOPO is 0.002. This means that at the nodes, ASTER is less significant than SRTM by fifteen times, and TOPO is less significant than SRTM by seventy-five times.

Table 27: Summary of the nodes' significance

	SRTM	ASTER	TOPO
MIN	0.00	0.00	0.00
MAX	0.95	0.76	0.86
AVG	0.15	0.01	0.002
STD	0.27	0.07	0.03

An initial interpretation of the above results compared to those from Table 26, a coarse general observation can be made: that there is a negative correlation between the deviation and the significance, meaning that the higher the significance is, the lower the deviation is expected to be.

Figure 91: General QQ-Plots comparing the deviation to the significance for the three models (SRTM, ASTER, and TOPO) (Screenshots taken by Author)

Using SPSS, the Pearson correlation is computed between the deviation and the significance for the nodes derived from the three DEMs. The results came as follows: for SRTM, the correlation is equal to -0.075; for ASTER is equal -0.215; for TOPO is equal to -0.027. In the three cases, there is almost no correlation, as if these two bits of data variate independently and randomly from each other's.

5.5 Discussion

At this point, the results found and presented in section 5.4 reached a point where there is no correlation between the propagation of the local accuracy and the accuracy of the derived products, where, as shown in sub-section 5.4.6, the end result was that there is almost no correlation between the proposed slope significance and the deviation of the streams in the plane. Before making a final conclusion about the proposed model and its ability or not to predict the accuracy of the derived products, a major observation to be taken into consideration is that the deviation and significance data do not form a normal distribution; therefore, no general deduction can be concluded from the correlation analysis. Therefore, while the dependency between the two data values was not proven, their independency on each other cannot be proven either.

Accordingly, a second review of the above results lead to the discovery of two main deficiencies in the interpreted results. These deficiencies introduced the following two questions that question the validity and generality of the concluded results:

1. Should all the nodes be interpreted as a whole, or should those with significance equal to zero be separated and interpreted alone?
2. The slope significance cannot be computed in the plain where the slope is equal to zero. Should the nodes in the plain have the significance equal to zero, as it happened in the previous results, or should these points be interpreted differently?

The following two sub-sections will address each of the above questions and suggest an enhancement to the original proposed model that should lead to a confident and generalizable conclusion.

5.5.1 Significant versus Non-Significant

Comparing the number of nodes used to analyze and assess the proposed procedure, in terms of the slope significance, shows that in those generated from SRTM, 1,811 nodes out of 2,510 have significance equal to zero, which equates to about 72% of them; in those generated from ASTER, 7,095 nodes out of 7,408 have significance equal to zero, forming 95% of the total; similarly, in those generated from TOPO, 44,325 nodes out of 44,661, forming 99% of the total. It can be concluded that the majority of the nodes from the three models have significance equal to zero. At the same time, this majority has a huge variation in the resulted deviations. Therefore, it can be assumed that this majority plays an important role in preventing the deduction of a strong correlation between slope significance and plane deviation of the streams.

The best solution is to divide the nodes into two sub-sets: the significant sub-set containing the nodes having a significance greater than zero and the non-significant sub-set containing those having significance equal to zero. Because the latter, as previously discussed, has a constant significance and a varying deviation, it is impossible to deduce any correlation. Instead, the focus will be on the significant sub-sets. Accordingly, the assessment of the results that was previously performed in sub-section 5.4.5 will be repeated below for the significant sub-sets only. The results are represented as follows in the two tables below:

Table 28: Summary of the significant nodes' deviation, expressed in meters

	SRTM	ASTER	TOPO
MIN	0.00	0.00	0.00
MAX	230.33	195.26	73.82
AVG	38.13	75.34	10.63
STD	33.31	601.88	333.58

As shown in Table 28, the significant nodes derived from ASTER have the highest average deviation with an average of 75.35 meters, followed by those derived from SRTM with an average deviation of 38.13 meters, and the last are those from TOPO with an average deviation of 10.63 meters. Comparing the results to the overall sets of nodes from the Table 25 shows that the significant sub-sets from the three DEMs have a smaller average deviation from its whole set by approximately four, twenty-four, and one hundred and forty-five times for SRTM, ASTER, and TOPO, respectively.

Table 29: Summary of the significant nodes' significance

	SRTM	ASTER	TOPO
MIN	0.01	0.01	0.01
MAX	0.95	0.76	0.86
AVG	0.53	0.31	0.29
STD	0.26	0.21	0.21

Similarly, as shown in Table 29, the significant nodes derived from SRTM have the highest average significance, with an average of 0.53, followed by ASTER with an average significance of 0.31, and last is TOPO with an average significance of 0.29. Comparing these results to the overall sets of nodes from Table 27 shows that the significant sub-sets from the three DEMs have an average significance higher from its whole set, by approximately fourteen, five, and thirteen times for SRTM, ASTER, and TOPO, respectively.

Furthermore, computing the normal QQ-Plot for the deviation and significance for these three new sub-sets, as shown in the below figures of 92, 93, and 94 for SRTM, ASTER, and TOPO, respectively, shows that all these sub-sets have almost normal distribution, much better than it was for whole sets. Therefore, any conclusion derived from these sub-sets, unlike the whole set, can be generalized.

Figure 92: Normal QQ-Plots of the significance and deviation for the significant nodes derived from SRTM (Screenshots taken by Author)

Figure 93: Normal QQ-Plots of the significance and deviation for the significant nodes derived from ASTER (Screenshots taken by Author)

Figure 94: Normal QQ-Plots of the significance and deviation for the significant nodes derived from TOPO

Comparing the deviation of the nodes with respect to the significance for those nodes, as shown in the General QQ-Plots in Figure 95, shows that linear correlation exists between the two, especially when visually compared to those in Figure 91 when applied to the whole sets.

Figure 95: General QQ-Plots comparing the deviation to the significance for the three models significant nodes (SRTM, ASTER, and TOPO) (Screenshots taken by Author)

Beyond visual interpretation of the graphs using SPSS, the Pearson correlation between the deviation and the significance is computed for the significant sub-sets of the three models.

The correlation came as follows: for SRTM the correlation is equal to -0.036; for ASTER is equal to 0.111; for TOPO is equal 0.01. Similar to the conclusion found in sub-section 5.4.6, the correlation between them is still considered to be weak. However, comparing the ranges of deviation in cell count for the three models, between the significant and non-significant sub-sets shows that, as shown in Table 30, the ranges of all the significant sub-sets makes a small portion of their non-significant complements. In fact, the ranges of the significant sub-sets make only 2%, 6%, and 3% of the non-significant ones for SRTM, ASTER, and TOPO, respectively.

It can be concluded that nodes derived from SRTM that are significant will have a maximum deviation of 2 cells; those derived from ASTER will have a maximum of 6 cells; and those derived from TOPO will have a maximum of 14 cells. Given that the significant points have a normal distribution in the deviations and significance, it can be concluded in general that any derived stream in a significant area will have a maximum deviation of two cells in SRTM, six cells in ASTER, and fourteen cells in TOPO, and, therefore, its plane accuracy will be estimated as such all over Lebanon, and not only in the Litany River Basin. The result follows a pessimistic interpretation of the data by assigning the accuracy as the maximum deviation instead of the conventional approach where the RMSE is calculated. This decision is taken to ensure the highest possibility that all possible deviations will end up to be equal to or less than the set threshold.

On the other hand, non-significant sample points did not have in either DEMs a normal distribution in their deviations, which has much higher ranges, and, therefore, the only conclusion that can be deduced from these sub-set of samples is that the plane accuracy of the derived streams in non-significant areas cannot be anticipated, and these streams cannot be considered reliable without further analyzing them. Instead, it is recommended in these areas to procure other products whose accuracy when propagated, will lead to a significant result.

Table 30: Ranges of deviation in cell count for significant and non-significant nodes

		SRTM	ASTER	TOPO
Significant	MIN	0	0	0
	MAX	2	6	14
Non-Significant	MIN	0	0	0
	MAX	102	99	466

Several standards exist worldwide to control the quality of the mapping products. One of these standards was developed by the American Society for Photogrammetry and Remote Sensing (ASPRS) in 1990. According to ASPRS, the maximum allowable RMSE of horizontal inaccuracies is calculated depending on the class of the data: for class 1, the error expressed in feet is equal to the scale denominator of the product divided by 100. Class 2 error is equal to the twice that of class 1 error, and class 3 error is equal to triple that of class 1 error (Falkner & Morgan, 2002). Considering the highest level of accuracy, the data are considered as class 1. Converting the RMSE into meters from feet would result in:

$$RMSE_{meters} \leq \frac{Scale\ Denominator}{300} \quad (5-11)$$

However, the product in-hand is not expressed in terms of scale, but in spatial resolution. According to (Tobler, 1987), the rule is to divide the denominator of the scale by 1,000 to get the size of the smallest detectable object, expressed in meters; the resolution of the raster used to detect the object is half its size. Accordingly, the resolution of the raster is calculated as:

$$Resolution_{meter} = \frac{Scale\ Denominator}{2000} \quad (5-12)$$

Combining the RMSE of class 1 from ASPRS with the conversion between scale and spatial resolution will result in:

$$RMSE_{meters} \leq Resolution_{meters} \times 6.096 \quad (5-13)$$

Or, in words, the maximum allowable RMSE should be six cells of the input raster product.

Back to the results of the river streams, significant samples have a maximum deviation of two cells for the SRTM, and six cells for the ASTER. This concludes that the streams' generation from any of the two products in a significant area will ensure that the end product will meet the international standards of horizontal mapping, such as the ASPRS standard. As for the TOPO, its derived significant product spans to a maximum error of 14 cells, which is almost double the maximum allowed as per Equation 5-13. In this case, there is not a definitive conclusion, unlike in SRTM and ASTER, in which the derived streams from TOPO in the significant area will ensure that the end product will meet international standards.

Other methods for calculating the horizontal accuracy of derived streams exist; one major method presented in (Sharma, Tiwari, & Bhadoria, 2009) proposes to estimate the shape accuracy of the terrain by calculating the Kappa coefficient of each derived stream with respect to a reference one. While this method, like many others, provides an overall estimation of the accuracy per stream, the proposed method in this thesis focuses on individual spots along the stream and assesses different parts of that same stream separately.

The latter could be more useful when assessing the accuracy of major streams that could extend over different types of landscapes.

5.5.2 Flat Areas

As previously defined in Equation 5-9, the slope significance is inversely dependent on the slope of the terrain. Being in the denominator, the slope significance cannot be computed in flat areas where the slope is equal to zero. Flat areas represent 0.6% of the total area of study in the SRTM, 11% in ASTER, and 15% in TOPO. While these areas are insignificant in SRTM and can be ignored without introducing any major uncertainty to the work, they are significant in ASTER and TOPO, and, therefore, there must be an alternative means to compute their significance.

In fact, as shown in Figure 96, the flat areas as per the TOPO DEM are mostly concentrated in the Beqaa Valley, which spreads up to the Hermel district in the north-eastern part of the country. The second largest flat area is on the northern coast, specifically in the Akkar Plain. Being in the majority plain, these two areas form the major agricultural land in the country. Following those major areas, other plain areas include the capital, Beirut, in the middle coast, the city of Tripoli in the northern coast, and the southern coast in general with the widest in the area of Damour and Sour.

Back to the focused area of study, the Litany River Basin, in the generated nodes of the derived rivers: Those from SRTM, there are 15 nodes that are in flat areas out of 2,510 nodes, or 0.6%; from ASTER, there are 643 nodes out of 7,408, or 9%; and from TOPO, there are 19,507 nodes out of 44,661, or 44%. Comparing the percentages of the flat nodes to the flat areas in all of the country, those from SRTM and ASTER have almost the same percentage; those from TOPO are three times higher.

As shown in Figure 97, out of the nodes generated from ASTER, the majority of those in flat areas are concentrated on the Upper Litany River and its tributaries and toward the end of the Lower Litany, where it flows into the sea.

Figure 96: Distribution of the flat areas as per the TOPO DEM (Map designed by Author)

Using SPSS, the Pearson correlation is computed between the slope accuracy and the deviation. Results came as follows for the three models: SRTM correlation is equal to 0.789; ASTER correlation is equal to -0.22; and TOPO correlation is equal to -0.20. There is a weak correlation for the ASTER and TOPO. As for the SRTM, even though there is a somewhat strong correlation between the two variables, a solid conclusion cannot be deduced because the values of the deviations in SRTM do not have a normal distribution, as shown in the Figure 98.

Figure 97: Flat nodes derived from ASTER DEM (Map designed by Author)

Figure 98: Normal QQ-Plot for the deviations in flat nodes derived from SRTM (Screenshot take by Author)

Furthermore, as shown in the scatter plots in Figure 99 for the three models, it is not possible to deduce a link between the deviation and the slope accuracy. Also, for a given accuracy, there is almost a whole range of deviation; therefore, it can be concluded that irrespective of the slope accuracy, the significance of flat areas can only be equal to zero. This conclusion can be converted into the following logical expression:

$$\text{Slope} = 0 \rightarrow \text{Significance} = 0 \quad (5-14)$$

Figure 99: Scatter plots comparing the slope accuracy to the deviation at the flat node, SRTM, ASTER, and TOPO respectively (Graphs generated by Author)

5.5.3 Refined Procedure

Based on the previous two discussions, two new models superseded the original model for defining the significance as follows:

1. A location on a DEM is considered significant if the slope accuracy is less than its slope.
2. A location is never significant if the slope is equal to zero.

Because in the flat areas from the three models, the nodes all have their slope accuracy greater than zero, then all the areas that should meet the second condition will definitively meet the first one. Therefore, the simplified binary significance model will be computed as per the logical expression:

$$\text{Significance} = \text{Slope Accuracy} < \text{Slope} \quad (5-15)$$

Equation 5-15 applied on the three input datasets will result in dividing the area into two classes of zones: significant zones where the results of analysis are expected to meet mapping standards, and non-significant zones, where the accuracy of the results are highly varying and cannot be predicted. The results of the refined procedure are shown in the three maps in figure 100, 101, and 102 below for the models SRTM, ASTER, and TOPO, respectively.

The area of study was divided as follows: for SRTM, 59% of the area is significant; for ASTER only 14% of the area is significant; and for TOPO, only 9% of the area is significant.

Figure 100: Binary significance model for SRTM DEM (Map designed by Author)

Figure 101: Binary significance model for ASTER DEM (Map designed by Author)

Figure 102: Binary significance model for TOPO DEM (Map designed by Author)

5.6 Summary

In this chapter, a procedure was proposed to propagate the height accuracy of an input DEM in order to estimate the accuracy of its terrain-based analysis derived datasets. To test the proposed procedure, three DEMs were used, the SRTM, ASTER, and TOPO, covering Lebanon as the area of study and vario-models containing the varying relative accuracies of each as computed in chapter 4 to calculate the accuracy and significance of their derived slope datasets.

Tests performed on these two DEMs showed that even though ASTER and TOPO have a higher spatial resolution than SRTM, the derived datasets from the latter has higher significance. Several elements played part in this comparison, mainly the DEMs vario-models, where ASTER was modeled having lower relative accuracy than SRTM. Multivariate spatial analysis highlighted further deduction mainly that the correlation between the different datasets was proven to be product-independent, and, therefore, could be considered applicable to other DEM products. Also, this analysis showed that the slope accuracy is highly related to the DEM accuracy, but the ultimate product that is of interest in this paper – the slope significance – is more related to the actual slope rather than its accuracy.

This will lead to the ultimate conclusion that the analysis performed on relatively low accuracy DEMs applied in steep sloped areas will still lead to significant results, while the flatter areas would require a more accurate DEM to reach the same level of significance. Applied to Lebanon, the ASTER DEM gives significant results only in the steep slopes, whereas the SRTM DEM can be applicable both in the Western and Eastern chains. The coast and the Bekaa Valley, due to their flat nature, would require a more accurate DEM in order to retrieve significant results.

Chapter 5 Bibliography

- Burrough, P., & McDonell, R. (1998). *Principles of Geographic Information Systems*. New York: Oxford University Press.
- Clapham, C., & Nicholson, J. (2009). *Oxford Concise Dictionary of Mathematics*. Addison-Wesley.
- Cui, L., Li, Y. P., Huang, G. H., & Huang, Y. (2014). Effects of digital elevation model resolution on topography-based runoff simulation under uncertainty. *Journal of Hydroinformatics Vol 16 No 6*, 1343-1358.
- Falkner, E., & Morgan, D. (2002). Map Accuracy Standards. In *Aerial Mapping: Methods and Application, Second Edition*. Lewis Publishers.
- Hutchinson, M., Xu, T., & Stein, J. (2011). Recent Progress in the ANUDEM Elevation Gridding Procedure. *Geomorphometry*, 19-22.
- Jenson, S. K., & Domingue, J. O. (1988). Extracting Topographic Structure from Digital Elevation Data for Geographic Information System Analysis. *Photogrammetric Engineering and Remote Sensing 54 (11)*, 1593-1600.
- Mark, D. M. (1988). Network Models in Geomorphology. *Modelling Geomorphological Systems*, 73-79.
- Mutluoglu, Ö. (2010). Investigation of the effect of land slope on the accuracy of digital elevation model (DEM) generated from various sources. *Academic Journal - Scientific Research and Essays Vol. 5(11)*, 1384-1391.
- Quinion, M. (2015, May 24). *Garbage In, Gargbage Out*. Retrieved from World Wide World: <http://www.worldwidewords.org/qa/qa-gar1.htm>
- Sharma, A., Tiwari, K., & Bhadoria, P. (2009). Measuring the Accuracy of Contour Interpolated Digital Elevation Models. *Indian Society of Remote Sensing 37*, 139-146.
- Tarboton, D. G., Bras, R. L., & Rodriguez-Iturbe, I. (1991). On the Extraction of Channel Networks from Digital Elevation Data. *Hydrological Processes 5*, 81-100.
- The Characteristics of the Litani River*. (2015). Retrieved from Litani River Authority: http://www.litani.gov.lb/en/?page_id=63
- Tobler, W. (1987). Measuring Spatial Resolution. *Land Resources Information Systems Conference*, (pp. 12-16). Beijing.

- Wechsler, S. (2007). Uncertainties associated with digital elevation models for hydrologic applications: a review. *Hydrology and Earth System Sciences* 11(4), 1481-1500.
- Wu, S., Li, J., & Huang, G. (2008). Characterization and Evaluation of Elevation Data Uncertainty in Water Resources Modeling with GIS. *Springer Water Resource Management*, 959-972.
- Zhou, Q., & Liu, X. (2004). Analysis of errors in derived slope and aspect related to DEM data properties. *Elsevier Computers & Geosciences* 30, 369-378.

Chapter 6

Conclusion

The main focus of this thesis is about the representation of the accuracy of a DEM and how this representation can be propagated into the datasets that will be derived from that DEM in order to better anticipate its accuracy. This thesis had two main objectives under this general topic: The first objective was to find a suitable representation of the local accuracy of the input DEM, named the accuracy vario-model, which could capture the variation of this accuracy within the area of study. The second objective of this thesis was to develop a generic procedure to apply the accuracy vario-model to the DEM-derived datasets and anticipate the mapping accuracy of their features.

The focus topic of this thesis and its objectives can be considered a very important topic to be discussed in the field of GIS because, while errors are inevitable in any measurement including geographic ones, modelling these errors through the accuracy of these measurements will allow a proper mean to use them in scientific and engineering works and deduce the appropriate decision making accordingly with a suitable level of confidence. DEMs are one specific form of geographic measurements that are being extensively used in many GIS applications such as terrain morphology, volumetric calculations, visibility and line-of-sight, and hydrologic analysis and modeling. These kinds of GIS uses are usually applied to large areas of study. Therefore, adopting the conventional single-value RMSE representation of the accuracy would average a wide range of accuracy variations, and thus the data user would lose a crucial level of detail, which will highly affect his/her decision making.

Also, these applications do not rely solely on the DEM; instead they are based on several derived datasets, whose units defer from the one of the DEM, which is a linear altitude unit. Therefore, knowing the accuracy of the input DEM only is not enough to assess the accuracy of the derived datasets. Instead of assessing these derived datasets separately using supporting field measurements that are usually costly and time consuming, if the terrain itself is accessible, this thesis has proposed a generic model to propagate the accuracy vario-model and anticipate the zones where the results of the analysis would be mapped according to international standards, without the need of any supporting field work.

With reference to *Chapter 1*, here are the initial questions that were raised in this thesis:

- i. What is the accuracy of given DEM? And how to systematically calculate it?
- ii. Given the accuracy of the DEM, what is the accuracy of the analysis-derived result?
- iii. Over a large coverage area, is the accuracy of a DEM constant, or does it change?

- iv. If the accuracy of a DEM is variable, is the variation random or dependent on some factors?
- v. And what factors could affect the variability in the DEM accuracy?

To assess how these questions have been answered throughout this thesis, this conclusion chapter will proceed with the following sections:

- **Empirical Findings:** covers the review of the findings and how do they partially link to the above raised questions
- **Theoretical Implication:** covers the contributions and implications of these findings to the theories and understanding of the topics covered by those questions
- **Limitations:** covers the faced challenges and limitations while doing this research, how they were overcome, and those that restricted the findings
- **Future Work:** covers future topics that could be built on top of the findings

6.1 Empirical Findings

Throughout the chapters that described the work done, mainly chapters 3, 4, and 5, there were several empirical findings that were deduced. These major findings are:

From *Chapter 3*, the chosen AOI at the scale of the countries is relatively small, yet it has a high variation in the altitude and the LULC of the terrain, which makes it a suitable pilot area whose findings could be generalized to other countries around the world.: for example, Spain, which is the second largest country in the European Union, has an area of 504,782 km², a terrain variation from 0 to a peak of 3,718 meters, and an average altitude of 650 meters. Compared to Lebanon, Spain has higher peaks by 20%, higher average by only 5%, but spread over an area that is 48 times larger. Another example is Egypt: the sixth largest country in the Arab world with an area of 1,010,408 km², a terrain variation from -133 to a peak of 2,629 meters, and an average of -15 meters (Below sea level). Compared to Lebanon too, Egypt has an altitude range making only 85% of the Lebanese one, and its average altitude is just below the sea level, meaning that this country, which is 97 times larger than Lebanon, has a majority of flat areas.

Also from *Chapter 3*, the altitude in the DEM was proven to have a low correlation with the LULC classification. This means that a given LULC class could be available at almost any altitude, and therefore, these two datasets can be considered independent from each other's. While this finding by itself does not have any direct implication on the raised questions, it is an important input to support the justification of the latter findings.

From *Chapter 4*, the assessment of the three DEMs showed that there is a high variation in their accuracy at the level of individual GCPs. In fact, the SRTM accuracy had a range from 0 to 111.6 meters for the global accuracy and from 0 to 42.1 meters for the local one, ASTER had a range from 0 to 66.27 meters for global accuracy and from 0 to 35.82 meter for local one, and TOPO had a range from 0 to 44 meters for the global accuracy and from 0 to 20.4 meters for the local one. Generalizing these ranges into a single value for the whole country will cause some regions to be underestimated and other overestimated in terms of their accuracy. Grouping regions by their LULC classification and calculating a single accuracy for each class resulted in a better representation of the accuracy compared to a single value RMSE, even though using the same input GCPs, by 24% for the SRTM, 7% for the ASTER, and 13% for the TOPO.

Also from *Chapter 4*, not only adopting the vario-model for modelling the accuracy showed a better representation for the accuracy for each DEM separately, but also adopting the LULC classification as a parameter to define the regions of similar accuracy was consistent among the three products. Accordingly, it can be concluded that no matter what input DEM is used and irrespective of the actual accuracy values, there is a high probability that regions classified as wetlands will have the highest accuracy, followed by water bodies; arable land; permanent crops; heterogeneous agricultural areas; urban fabric; industrial, commercial, and transport units; open space with little or no vegetation; mines, dumps, and construction sites; forests. Scrub and/or herbaceous vegetation association will have the lowest accuracy. A more generalized conclusion can also be stated as follows: Water and wetland areas will have the highest accuracy, followed by the various agricultural lands, urban areas, and open areas. Forests and herbaceous areas will have the lowest accuracy.

From *Chapter 5*, the proposed procedure to propagate the accuracy vario-model of a DEM was applied into the application of hydrological analysis, where the simplified form of the proposed slope significance dataset divided the area of study into two regions: a significant and a non-significant one. Comparing the deviations in the results of stream delineation with respect to the ASPRS mapping standards showed that those stream generated in significant regions met ASPRS standards in two out of three DEMs, while those generated in non-significant regions had a huge variation in their deviation to a point that they could not meet the standards, and therefore these non-significant streams cannot be used for any further mapping, analysis, nor supporting any decision-making, without being further inspected.

Finally, combined from *Chapters 4 and 5*, adopting the concept of local accuracy instead of the global one showed to be a necessity in order to propagate that accuracy into the derived products, as most of the values in theses derived products are calculated based on the

variation of the altitude of each cell in the input DEM with respect to its neighbors instead of being used solely.

6.2 Theoretical Implication

After reviewing the individual empirical findings from the various chapters, this section will go back to the originally stated questions and summarize an answer for each based on the findings:

Question 1: What is the accuracy of given DEM, and how to systematically calculate it?

From the theoretical background chapter, the accuracy of a given DEM represents how close the altitude in this DEM is to the actual altitude. The accuracy is represented by a number with the same linear unit as the altitude in the DEM, expressed in meters, and this number represents the upper boundary of the absolute difference between the actual and the modelled altitude. The conventional and used method to systematically calculate the accuracy of a DEM is by calculating the RMSE of the individual altitude absolute difference between the model and the reality for a set of GCPs that are collected to assess this DEM. While the global accuracy is calculated as the absolute difference at each point, this research introduced an alternative method where each GCP is compared to its neighboring ones in order to estimate the local accuracy of the GCP with respect to its surroundings. While this alternative method is more complex to perform, it did not provide any added value to the answer for question number one. However, the importance of this new method will be highlighted in the answer to upcoming questions and in the limitations section.

Question 2: Given the accuracy of the DEM, what is the accuracy of the analysis-derived result?

To specifically answer this question, it must include first the intended derived product. Because of the numerous uses of DEM, it is not practical to research a separate solution for each of the potential ones. Instead, a more generic solution is needed that is applicable to any potential derived product. After several assessments, it was noticed that the majority of the terrain-based analyses are based on assessing the altitude variation of each cell with respect to its neighbors, and thus based on the slope. Therefore, the first step in propagating the accuracy of the DEM was to develop a mathematical procedure to calculate the accuracy of the slope based on the mathematical function of the slope. The next step was to calculate how much the error makes or the overall slope, as the higher the error is with respect to the slope, the lower the probability is to ensure that the processing will model the reality

correctly. The result was called the slope significance, and after performing a test of hydrological analysis, the slope significance proved to be a proper indication for the quality of the output. This implication can be generalized for the output of any other terrain-based analysis, as hydrological analysis is considered one of the most complex and multi-steps operations to reach the final output, and if after all these steps, the propagation of the slope significance was preserved, then it will definitively be preserved for simpler and shorter operations.

Question 3: Over a large coverage area, is the accuracy of a DEM constant, or does it change?

Referring back to the empirical findings from the previous section, global accuracy had a range varying from 40 meters to 116 meters for the conventional global accuracy and varying from 20 meters to 66 meters for the proposed local one. Even though the three DEMs have different spatial resolutions and were produced using three different techniques, they all have a large variation in the accuracy at the different GCPs. Therefore, it is clear that the accuracy of any given DEM is not constant at all.

Question 4: If the accuracy of a DEM is variable, is the variation random or dependent on some factors?

In any taken measurement, randomness will always exist. In fact, throughout this thesis, there was always a random component in any researched and established correlation between the accuracy and any other variable, otherwise these correlations would have become perfect ones with a coefficient equal to plus or minus one, which was not the case. However, this does not mean that the variation in the accuracy is totally random; instead there could be many factors, researched or not in this thesis, that play a part in defining the accuracy. Knowing these factors and modelling them would highly assist in predicting the behavior of a DEM accuracy.

Question 5: And what factors could affect the variability in the DEM accuracy?

As mentioned in the previous answer, several factors could exist that will affect the variability of the accuracy. While many would think that the altitude could have an effect on the accuracy of the modelled terrain, this research proved that there is no correlation between the two; instead, one interesting factor that was found affecting is the land cover. In fact, as described in the previous section, using this factor as a parameter to create sub-regions with different DEM accuracy provided a more representative way to model the accuracy compared to the conventional non-parametric one.

6.3 Challenges and Limitations

While processing the data, the author came across several challenges and difficulties that required additional efforts to overcome, and in some cases resulted in restricting the findings of this work.

Some of the major challenges that were faced, and how they were resolved, include:

- **Different spatial references:** The various input datasets were originally based on different coordinate systems, and these systems were based on different ellipsoids. Loading these datasets on the GIS software and letting it re-project them on-the-fly resulted in a considerable shift of 90 meters in the plane between them. To overcome this challenge, an additional custom transformation was used to correct the shift and orientation difference between the two ellipsoids, thus reducing this shift to a negligible few centimeters.
- **Different height references:** The various input datasets are also based on different height references and geoids, which introduced an additional error in the measurement of the altitude in comparison to the GCPs. This issue was resolved by adopting the concept of local accuracy that is based on calculating the error in the altitude differences between two or more GCPs instead of the error in the altitude of the GCP with respect to its height reference.
- **Complexity of the local accuracy calculation:** The calculation of the local accuracy for a given GCP requires selecting all its neighboring GCPs, calculating the local accuracy to each, and finally calculating the average one. Such operation applied to a set exceeding one thousand GCPs can never be done manually without spending a huge amount of time and incorporating the risk of human error in skipping some pairs of GCPs or performing a wrong calculation. To overcome this issue, this procedure was translated into a Python script to easily and systematically repeat the procedure over different datasets.

Other challenges were not resolved and required some additional assumptions to be made. Major unresolved challenges are:

- **Different temporal resolution:** Five main datasets were used in this research. Each of them come from a different source and was captured at a different date: LULC was generated in 1998, SRTM DEM in 2000, ASTER DEM in 2009, and TOPO

in 2007. The GCPs dataset is more complicated as it is a compilation of GCPs from various sources. The exact range of dates is unknown, but they are assumed to be taken between 2005 and 2011. Combining the dates of all these datasets results in a range of thirteen years. This range might introduce some temporal ambiguity, but in assessing a terrain at a level of a country, there is a very low chance that a major change in the terrain might have taken place within this range and could have created a major ambiguity in the performed research. Therefore, to simplify the work, these time differences were ignored, and all the datasets were assumed to be aligned to the same time frame.

- **Irregular distribution of GCPs:** As earlier discussed in sub-section 3.3.1, the collected GCPs do not form a normal distribution in the plane. Again, this is similar to any real situation where there is no possibility to go to the field and collect the optimal number and distribution of GCPs. Instead, the data analyst should make use of what is available and process it in a way so as to remove the effect of the unbalanced distribution.

6.4 Future Work

This work could have never been done if it did not rely on and build on top of previous scientific work. These previous achievements are reviewed and summarized in the theoretical background. Similarly, future scientific works are expected and encouraged to be built on top of the findings of this research. While the future works cannot be all anticipated now, here are some suggestions of potential research that can be made and are directly related to this one:

- **Apply the vario-model to other DEM products:** It would be interesting to apply the concept of the vario-model into other DEM products for the same area and further compare its results to the current findings. One specific product of high interest is the SRTM-2. It is interesting because it is a global product like SRTM and ASTER, yet it has the same spatial resolution like ASTER. SRTM-2 mission was launched in September 2014, and it was expected to be included in this research. However, up until the time of writing this thesis, the Middle East region had still not been released.
- **Apply the vario-model into other countries:** It could also be interesting to apply the vario-model on the same global DEMs, the SRTM and ASTER to other

Mediterranean countries and compare the variation of accuracy for each LULC class among those countries. Countries from outside the Mediterranean region can also be included, but chances are that the land cover would highly differ due to the major changes in the geography around the world.

Apply the vario-model into visibility analysis: This research applied the vario-model and propagated it into hydrological analysis, as it is one type of complex and multi-steps processes. It would be interesting to apply the vario-model to anticipate the accuracy of line of sight and visibility analyses.

- **Translate the procedure to Open-Source GIS:** While this last proposed future work does not form an innovative scientific, an interesting technical work could be to translate the developed code from working under Esri ArcGIS environment into an open-source one, such as QGIS or gvSIG. This task would enable this work to be spread over a wider audience which does not have or work on an Esri platform.

6.5 Final Remarks

Geographic data accuracy is a crucial topic to be taken into consideration when using any data to analyze and deduce and perform a related decision-making. Covering all aspects of accuracy in one thesis is too broad to yield one generic solution. Within this wide, yet sometimes underestimated field of study, this research focused on the vertical accuracy of DEM, and how this accuracy can affect results of DEM-based analyses. While previous research has partially tackled this topic, this one was able to provide a new innovative way to represent the accuracy. The author of this thesis hopes that his contribution to this field opens the door for researchers and practitioners of geospatial sciences and technology to interpret the accuracy of DEM in a more detailed fashion and translate its accuracy to the results of the analysis and eventually to their decision making.

Appendix

Appendix A: Data

A1: Conversion Table from SDATL (in French) to CORINE (in English) Classification

SDATL (Level 3)		CORINE (Level 3)	
111	Tissu urbain continu	111	Continuous Urban Fabric
112	Tissu urbain discontinu	112	Discontinuous Urban Fabric
121	Zone industrielle ou commerciale	121	Industrial or Commercial Units
122	Zone portuaire	123	Port Areas
123	Aéroport	124	Airports
			Road and Rail Networks and Associated
124	Gare ferroviaire	122	Lands
			Road and Rail Networks and Associated
125	Infrastructure routière	122	Lands
126	Bâtiment d'exploitation agricole	121	Industrial or Commercial Units
131	Extraction de matériau (carrière)	131	Mineral Extraction Sites
132	Décharge	132	Dump Sites
133	Remblai en mer	132	Dump Sites
134	Extension urbaine et/ou chantier	133	Construction Sites
135	Terrain urbain vacant	112	Discontinuous Urban Fabric
141	Espace vert urbain	141	Green Urban Areas
142	Grand équipement sportif ou de loisir	142	Sport and Leisure Facilities
211	Culture de plein champ en grande surface	212	Permanently Irrigated Land
	Culture de plein champ en petites surface ou		
212	terasses	212	Permanently Irrigated Land
214	Mitage urbain sur culture de plein champ	212	Permanently Irrigated Land
		22	
221	Oliviers	3	Olive Groves
222	Vignobles	221	Vineyards
		22	
223	Arbres fruitiers à feuilles caduques	2	Fruit Trees and Berry Plantations
		22	
224	Agrumes	2	Fruit Trees and Berry Plantations
		22	
225	Bananiers	2	Fruit Trees and Berry Plantations
		22	
226	Mitage urbain sur verger	2	Fruit Trees and Berry Plantations
		22	
227	Vergers mélangés	2	Fruit Trees and Berry Plantations
231	Cultures intensives en plein champ	212	Permanently Irrigated Land
		24	
232	Cultures intensives sous abri (serres)	2	Complex Cultivation Pattern
		24	
233	Mitage urbain sur culture intensive	2	Complex Cultivation Pattern
234	Cultures intensives en plein air et sous abri	24	Complex Cultivation Pattern

Appendix

	mélangées	2	
		24	
213	Culture de plein champ associée à une autre culture	2	Complex Cultivation Pattern
22		22	
8	Vergers associés à d'autres cultures	2	Fruit Trees and Berry Plantations
		24	
235	Cultures intensives associées à une autre culture	2	Complex Cultivation Pattern
311	Forêt dense de conifères	312	Coniferous Forest
312	Forêt dense de feuillus	311	Broad-Leaved Forest
313	Forêt dense mixte	313	Mixed Forest
314	Mitage urbain sur forêt dense	313	Mixed Forest
321	Forêt clairsemée de conifères	312	Coniferous Forest
322	Forêt clairsemée de feuillus	311	Broad-Leaved Forest
323	Forêt clairsemée mixte	313	Mixed Forest
324	Mitage urbain sur forêt clairsemée	313	Mixed Forest
331	Végétation arbustive	321	Natural Grassland
		32	
332	Végétation arbustive avec arbres dispersés	4	Transitional Woodland-Shrub
333	Mitage urbain sur zonen arbustive	321	Natural Grassland
65		33	
o	Zone incendiée	4	Burnt Area
		33	
410	Végétation herbacée moyennement dense	3	Sparsely Vegetated Areas
42		33	
o	Végétation herbacée clairsemée	3	Sparsely Vegetated Areas
510	Zone humide continentale	411	Inland Marshes
52			
o	Zone humide maritime	421	Salt Marshes
		33	
611	Roche nue	2	Bare Rocks
		33	
612	Mitage urbain sur roche nue	2	Bare Rocks
62		33	
o	Sol nu	2	Bare Rocks
63			
o	Plage	331	Beaches, Dunes, Sands
64			
o	Dune de sable	331	Beaches, Dunes, Sands
711	Plan d'eau	512	Water Bodies
712	Lac collinaire	512	Water Bodies
713	Fleuve ou rivière	511	Water Courses
722	Bassin portuaire	523	Sea and Ocean

A2: Descriptive Geostatistics of GCPs Z per LULC Classification

Using ArcGIS Geostatistical Analyst extension, the Histogram and Normal QQPlot are computed for the GCP' Z values to visualize and interpret their distribution, especially with respect to the normal one. These GCPs are grouped per LULC level 1 and level 2 classification, in order to better understand the behavior of the altitude with respect to the different land covers.

L1: Artificial Surfaces

L11: Urban Fabric

L12: Industrial, Commercial and Transport Units

L13: Mine, Dump and Construction Sites

L2: Agricultural Areas

L21: Arable Land

L22: Permanent Crops

L24: Heterogeneous Agricultural Areas

L3: Forest and Semi-Natural Areas

L31: Forest

L32: Scrub and/or Herbaceous Vegetation Association

L33: Open Space with Little or no Vegetation

L4: Wetlands

L5: Water Bodies

A3: Descriptive Geostatistics of GCPs Local Accuracy per LULC Classification

Similarly, to A2, the Normal QQPlot is computed for the GCP' Local Accuracy of the three DEMs' values to visualize and interpret their distribution, especially with respect to the normal one. These GCPs are grouped per LULC level 2 classification, in order to better understand the behavior of the altitude with respect to the different land covers. For each class the plots from Left to right are for: SRTM, ASTER, and TOPO respectively.

L11: Urban Fabric

L12: Industrial, Commercial and Transport Units

L13: Mine, Dump and Construction Sites

L21: Arable Land

L22: Permanent Crops

L24: Heterogeneous Agricultural Areas

L31: Forest

L32: Scrub and/or Herbaceous Vegetation Association

L33: Open Space with Little or no Vegetation

L4: Inland Wetlands

L5: Inland Waters

Appendix B: Python Scripts

B1: Calculate the Local Accuracy for the GCPs

```
# Import arcpy module
import arcpy

#I/O variables
inputPoints = "C:/Users/Wassim/Documents/PhD/Data/SHP/GCPs_TOPO_Refine.shp"
inputZField = "Z"
inputZmField = "Zm"
tempFolder = "C:/Users/Wassim/Documents/PhD/Data/Temp"
outputFieldName = "LA_TOPO"

#Create Thiessen polygons
arcpy.CreateThiessenPolygons_analysis(inputPoints, tempFolder + "/Thiessen", "ALL")
arcpy.MakeFeatureLayer_management(tempFolder + "/Thiessen.shp", "Thiessen")

print "Thiessen polygons generated"

#Create temp field
arcpy.AddField_management("Thiessen", outputFieldName, "DOUBLE")

print "Temp_Field created"

#For each polygon
curPoly = arcpy.UpdateCursor(tempFolder + "/Thiessen.shp")

for polyRec in curPoly:

    #Search for all adjacent polygons
    arcpy.SelectLayerByLocation_management("Thiessen", "BOUNDARY_TOUCHES", polyRec.shape)
    adjPoly = arcpy.SearchCursor("Thiessen")

    adjCount = 0
    adjSum = 0.0

    print "Calculating for FID: " + str(polyRec.getValue("FID"))

    for adjRec in adjPoly:
        if adjRec.getValue("FID") != polyRec.getValue("FID"):
            #Calculate LA
            adjCount += 1
            adjSum += abs(adjRec.getValue(inputZField) - polyRec.getValue(inputZField) -
adjRec.getValue(inputZmField) + polyRec.getValue(inputZmField))

    #Calculate Average LA
    if adjCount > 0:
        polyRec.setValue(outputFieldName, adjSum / adjCount)
    else:
        polyRec.setValue(outputFieldName, 0.0)

    print "Average LA = " + str(adjSum / adjCount)
    print "-----"

    #save changes
    curPoly.updateRow(polyRec)
    del adjPoly
```

```
del curPoly

print "out of loop"

#Assign Average LA to points
arcpy.JoinField_management(inputPoints, "FID", tempFolder + "/Thiessen.shp", "Input_FID",
"Temp_Field")

print "DONE!!!!!"
```

B2: Calculate Slope Accuracy

```
# Import arcpy module
import arcpy

# Check out any necessary licenses
arcpy.CheckOutExtension("spatial")

# Script arguments
Input_DEM = arcpy.GetParameterAsText(0)
if Input_DEM == '#' or not Input_DEM:
    Input_DEM = "srtm" # provide a default value if unspecified

Output_Slope = arcpy.GetParameterAsText(1)
if Output_Slope == '#' or not Output_Slope:
    Output_Slope = "C:\\Users\\User\\Documents\\ArcGIS\\Default.gdb\\Slope_srtm1" # provide a
    default value if unspecified

Input_DEM_Model = arcpy.GetParameterAsText(2)
if Input_DEM_Model == '#' or not Input_DEM_Model:
    Input_DEM_Model = "srtm_ra" # provide a default value if unspecified

Output_Slope_Accuracy = arcpy.GetParameterAsText(3)
if Output_Slope_Accuracy == '#' or not Output_Slope_Accuracy:
    Output_Slope_Accuracy = "C:\\Users\\User\\Documents\\ArcGIS\\Default.gdb\\rastercalc" #
    provide a default value if unspecified

# Local variables:

# Process: Slope
arcpy.gp.Slope_sa(Input_DEM, Output_Slope, "DEGREE", "1")

# Process: Raster Calculator
arcpy.gp.RasterCalculator_sa("Abs(ATan(Tan(\"%Output Slope%\") + (\"%Input Accuracy Model%\"
/ 1)) - \"%Output Slope%\")", Output_Slope_Accuracy)
```

Appendix C: Photo Gallery

This appendix shows a list of photos taken around the country, that assist in visualizing and understanding the nature of the country in general, and the different LULC classes in specific, as these classes may vary in different parts of the world:

L1: Artificial Surfaces

Figure C-1: Beirut, the capital of Lebanon, a Coastal urban area with high density. Photo taken on April 2015 (L11: Urban Fabric) (Photo taken by Author)

Figure C-2: Coastal urban area with medium density, known as 'Jounieh' bay. Photo taken from 'Harissa' in Mount Lebanon, on July 2011 (L11: Urban Fabric) (Photo taken by Author)

Figure C-3: A major excavation sites for cement production. Photo taken from 'Hamat' in the North, on October 2010 (L13: Mine, Dump and Construction Sites) (Photo taken by Author)

L2: Agricultural Areas

Figure C-4: Large potato field, showing an example of varying cultivation over the year. Photo taken from 'Joub Jannine' in the 'Beqaa' valley, on May 2012 (L21: Arable Land) (Photo taken by Author)

Figure C-5: Vineyard for wine production, showing an example of permanent crops. Photo taken from 'Kifraya' in the 'Beqaa' valley, on August 2011 (L22: Permanent Crops) (Photo taken by Author)

Figure C-6: Combination of green houses and arable land in a rural area. Photo taken from 'Annaya' in Mount Lebanon, on July 2013 (L24: Heterogeneous Agricultural Areas) (Photo taken by Author)

L3: Forests and Semi-Natural Areas

Figure C-7: Cedar trees or ‘Cedrus Libanis’ is a special type of coniferous trees that could age for more than 1000 years and are only available in Lebanon. The cedar tree is the symbol of the country and placed on its flag. Photo taken from the Cedars reserve in ‘Becharre’ in the North, on September 2012 (L31: Forests) (Photo taken by Author)

Figure C-8: Pine forest. One of the most dominant types of trees available in Lebanon. Photo taken from ‘Hammana’ in Mount Lebanon, in July 2012 (L31: Forests) (Photo taken by Author)

Figure C-9: Natural grasslands surrounding cultivated areas. Photo taken in West Beqaa, on August 2011 (L32: Scrub and/or Herbaceous Vegetation Association) (Photo taken by Author)

Figure C-10: Combination of rocks and uncultivated lands. Photo taken from 'Houla' in the South, on November 2014 (L33: Open Space with Little or no Vegetation) (Photo taken by Author)

Figure C-11: 'Qorneh Al Sawda' (The Black Summit) is the highest peak in Lebanon with an altitude of 3,088 meters above the sea level. Photo take from 'Becharre' in the North, on September 2012 (L33: Open Space with Little or no Vegetation) (Photo taken by Author)

Figure C-12: Seasonal snow covering mountain tops. Photo taken from 'El Mzar' ski slopes in 'Kfardebian', in Mount Lebanon, on March 2011 (L33: Open Space with Little or no Vegetation) (Photo taken by Author)

L5: Water Bodies

Figure C-13: 'Hasbani' river. Photo taken at the Lebanese-Israeli border at 'Wazzani' in the South, on July 2011 (L51: Inland Waters) (Photo taken by Author)

Figure C-14: 'Qaraoun' artificial lake. Photo taken from 'West Beqaa' on August 2011 (L51: Inland Waters) (Photo taken by Author)

Figure C-15: The shoreline on the Mediterranean Sea. Photo taken from 'Halat' in Mount Lebanon, on September 2009 (L52: Marine Waters) (Photo taken by Author)

Figure C-16: Selected photos - sites distribution map (Map designed by Author)

This thesis was finalized on September 30, 2016 in
La Escuela Técnica Superior de Ingenieros en
Topografía, Geodesia y Cartografía