

# REPORT

## Literature Review of Methodologies for Development of a High Resolution DEM for the Waste Isolation Pilot Plant (WIPP)

(In Support of Project 2, Task 6: Hydrology Modeling for WIPP,  
Subtask 6.1: Digital Elevation Model and Hydrologic Network)

Prepared by Gisselle Gutierrez-Zuniga, DOE Fellow

**Date submitted:**  
August 30, 2019

**Principal Investigator:**  
Leonel E. Lagos, PhD, PMP®

**Florida International University Collaborators:**  
Yan Zhou, PhD  
Angelique Lawrence, MS, GISP  
Gisselle Gutierrez, DOE Fellow

**Submitted to:**  
U.S. Department of Energy  
Office of Environmental Management  
Under Cooperative Agreement # DE-EM0000598



**TABLE OF CONTENTS**

---

INTRODUCTION.....3  
RELEVANT PUBLICATIONS .....4  
SUMMARY .....66  
REFERENCES .....67

## INTRODUCTION

---

FIU is assisting DOE-EM's Carlsbad Field Office (CBFO) and Pacific Northwest National Laboratory (PNNL) in developing a groundwater-basin model for the Waste Isolation Pilot Plant (WIPP) site using the Advanced Simulation Capability for Environmental Management (ASCEM) modeling toolset. ASCEM is an open source processed based computational framework for prediction of contaminant fate and transport in natural and engineered systems across DOE sites. FIU intends to use the ASCEM groundwater-basin model to assist DOE-EM improve the current understanding of regional and local groundwater flow at the WIPP site, compute the water balance, and derive estimates of groundwater recharge in the post-closure phase. Currently, ASCEM is unable to account for land surface hydrology, which is essential for computing the water balance. The proposed work will therefore require coupling of a state-of-the-art open-source land surface model (LSM) with the groundwater models (GWMs) within the ASCEM toolbox to simulate three-dimensional, unsaturated and saturated water flow.

At the WIPP site, there is a need for an improved understanding of the regional water balance, particularly the relation between Culebra recharge and the intense, episodic precipitation events typical of the monsoon. This relationship is essential for understanding the rate of propagation of the shallow dissolution front, and the impact of land-use changes around the WIPP facility on water levels in compliance-monitoring wells. These types of analyses require a revision of the current site conceptual model to couple surface water and groundwater processes, which requires a high resolution digital elevation model including channels and sink holes to account for surface water routing, and the development of a new mathematical model. FIU therefore plans to develop a high-resolution digital elevation model (DEM) for the WIPP site and surrounding basins using photogrammetric methods to better capture the ground surface topography and local features. A high resolution DEM of the site will enable more accurate delineation and extraction of features such as drainage basins, brine lakes, channels, sink holes and discharge points and will also facilitate future creation of spatially-distributed soil and vegetation classes. The following literature review was therefore conducted as a preliminary step to investigate methodologies for development of a high resolution digital elevation model (DEM) for the WIPP site and surrounding basins to support this effort.

FIU's collaboration with WIPP is projected to be a multi-year effort which will benefit DOE-EM by providing improved estimates of the spatial and temporal patterns of recharge, so that better predictions of halite dissolution and propagation of the shallow dissolution front can be made possible to assess and quantify the potential impact on the WIPP repository performance.

## RELEVANT PUBLICATIONS

---

FIU conducted a literature review of methodologies for development of a high-resolution digital elevation model (DEM) for the WIPP site. In addition, a review of methods for extracting significant topographical features such as sink holes from DEMs was also conducted. The following are some of the journal articles reviewed.

### **METHODOLOGIES FOR DEVELOPMENT OF A HIGH-RESOLUTION DIGITAL ELEVATION MODEL (DEM)**

#### **DEVELOPMENT AND ACCURACY ASSESSMENT OF A HIGH-RESOLUTION DIGITAL ELEVATION MODEL USING GIS APPROACHES FOR THE NILE DELTA REGION, EGYPT**

*El-Quilish, M., El-Ashquer, M., Dawod, G., & El Fiky, G. (2018). Development and Accuracy Assessment of High-Resolution Digital Elevation Model Using GIS Approaches for the Nile Delta Region, Egypt. American Journal of Geographic Information System, 7(4), 107-117.*

This paper describes the development of a high accuracy digital elevation model (DEM) which is needed for the Nile region to investigate the impact of climate change and observing the effects of sea level rise. The first step was the digitizing process which consisted of obtaining available topographic maps of the Nile Delta region as it is covered by 220 topographic maps of scale 1:25,000 with a contour interval of 50 cm. Topographic maps of scales 1:50,000 and 1:100,000 with a contour interval of 1 m were also obtained to complete missing zones of the study areas in the case of damaged topographic maps. Once obtained, georeferencing and projection processes for the horizontal datum were conducted for each topographic map using the ArcGIS projection tool, converting from ETM red belt/GCS Egypt 1907 projection to WGS84. Note, however, that the final DEM was created using a local coordinate system for the study area. These maps were manually digitized by extracting all the available contour lines and elevation points. Once the digitizing process was completed, local DEMs were created using nine different spatial interpolation methods (SIMs).

The authors of this paper emphasized the fact that there is no single interpolation method that is considered to be standard and different methods can be employed depending on the situation and what is most appropriate. In this study, the SIMs used were the Inverse Distance Weighted (IDW), Natural Neighbor Inverse Distance Weighted (NNIDW), Global Polynomial Interpolation (GPI), Spline with Barriers (SWM), Topo to Raster (TTR), Kriging, Radial Basis Function (RBF), Empirical Bayesian Kriging (EBK), and Local Polynomial Interpolation (LPI). Parameters were set to test each SIM to produce the most accurate DEM.

## Methodology

### Vertical Datum Matching

Before conducting any evaluation of the global and local DEMs, the elevation datasets had to have the same horizontal datum into the World Geodetic System 1984 (WGS84) and same vertical datum into the Geodetic Vertical Datum in Egypt (VDE1906) as the Mean Sea Level (MSL) at Alexandria harbor, based on daily reading of high and low water levels during recorded years from 1989 to 1906. It is important for comparison of GCPs with other DEMs to consider conversions of the ellipsoidal heights ( $h$ ) into orthometric heights ( $H$ ) using local geoid undulations ( $N_{VDE1906}$ ). Geoid undulations ( $N$ ) is the difference between the reference ellipsoidal height from GPS ( $h$ ) and the orthometric height from levelling ( $H$ ).

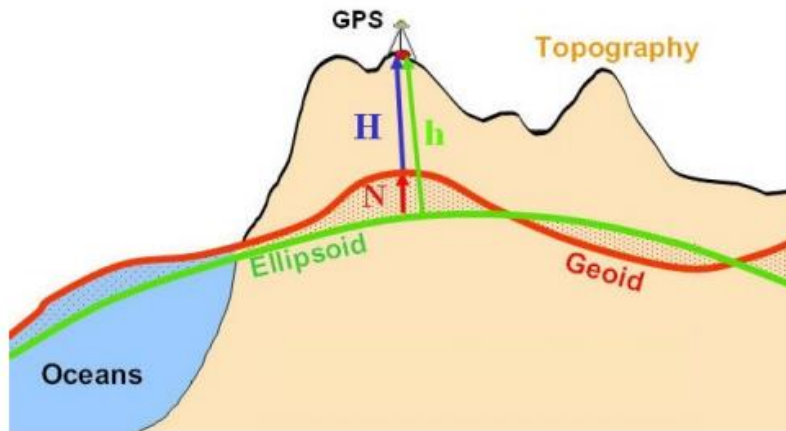


Figure 1. Visual representation of orthometric, geoid, and ellipsoidal heights

The elevation points (DPs) digitized from the topographic maps have vertical datum VDE1906 which is considered the MSL. DP coordinates were also transformed into the horizontal datum WGS94 using ArcGIS. As local DEMs were generated with the DPs, the heights are computed relative to WGS84. The orthometric heights ( $H$ ) were referred to the local geoid model VDE1906. To have the same vertical datum for GDEM, the heights were derived for all 200 GCPs with bilinear interpolation using GIS techniques from the surrounding grid points of the original resolution of each GDEM. Interpolated orthometric heights ( $H_{EGM96}$ ) were vertically referenced to the global geopotential model (Earth Gravitational Model 1996 (EGM96) and transformed to ellipsoidal heights ( $h_{EGM96}$ ) using geoid undulations from EGM96 ( $N_{EGM96}$ ) using the following equation:

$$h_{GDEM} = H_{EGM96} + N_{EGM96}$$

Next, the ellipsoidal heights ( $h_{GDEM}$ ) were transformed to orthometric heights ( $H_{EGY-HGM2016}$ ) using hybrid local geoid model undulations for Egypt using the following equation:

$$h_{EGY-HGM2016} = H_{GSEM} - N_{EGY-HGM2016}$$

Orthometric heights from GPS/levelling measurements were also considered during the modelling process to improve the determination of the hybrid gravimetric geoid over the region. The following table presents the statistical evaluation between the EGM96 global geoid model and the hybrid local geoid model for Egypt.

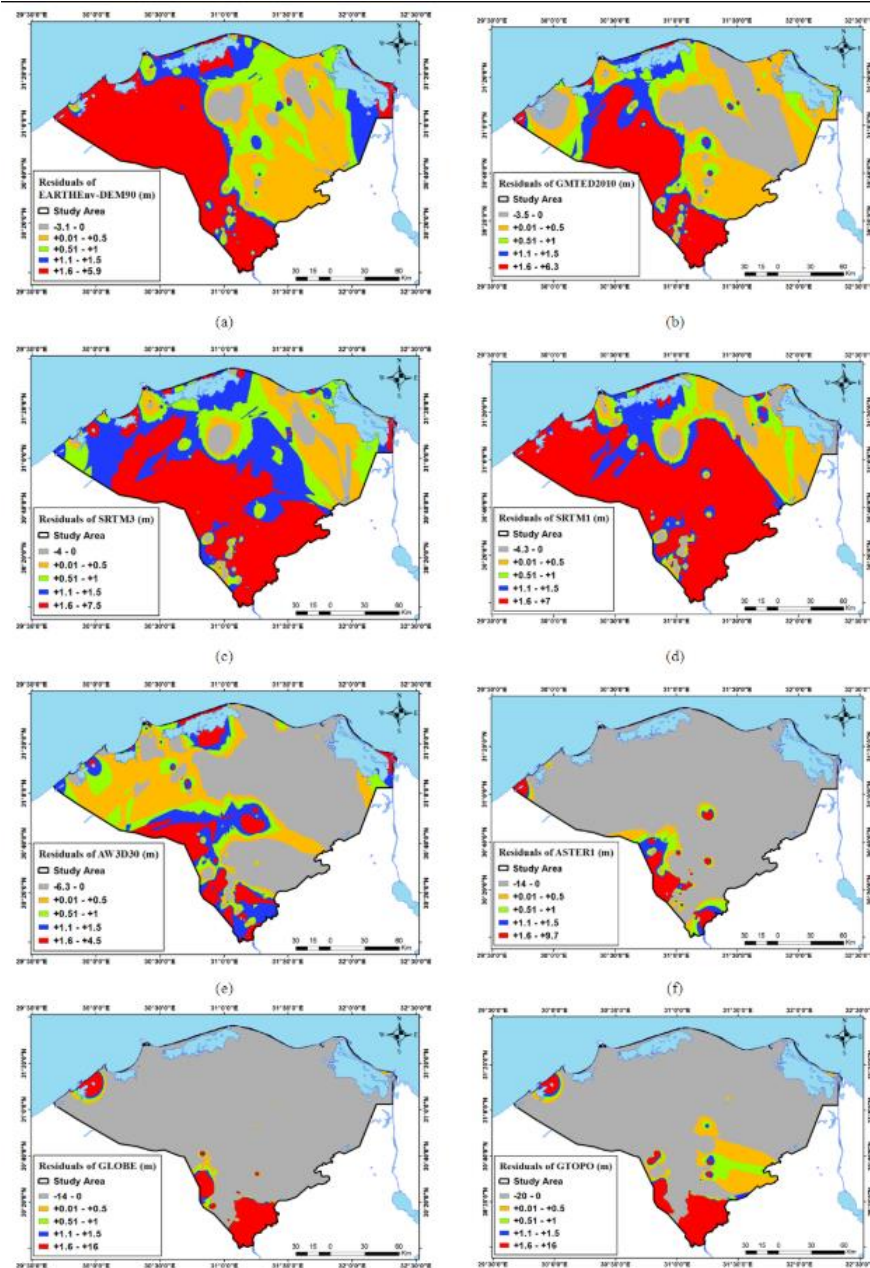
**Table 1. Statistical evaluation for the difference between  $H_{EGM96}$  and  $H_{EGY-HGM2016}$**

Max	Min	Mean	SD
0.15	-0.44	-0.12	0.11

Further evaluations on the GDEMs conduct the statistical calculations on the newly generated orthometric heights ( $H_{EGY-HGM2016}$ ).

*DEMs Validation with GCPs*

Evaluations of the DEMs, using the SIMs, were based on height differences, DEM errors, or the residuals ( $\Delta H$ ) between GCPs ( $H_m$ ) and its corresponding DEM-based elevation ( $H_{DEM}$ ) which were extracted from each of the GDEMs and LDEMs by deriving 200 GCPs through bilinear interpolation at the point location using GIS, shown in figures below.



**Figure 2. Residuals of global DEM in Nile delta study area (a) EARTHenv3 DEM; (b) GMTED2010 DEM; (c) SRTM 3 DEM; (d) SRTM 1 DEM; (e) AW3D30 DEM; (f) ASTER DEM; (g) GLOBE DEM; (h) GTOPO30 DEM**

The residuals ( $\Delta H$ ) were calculated using the following equation:

$$\Delta H = H_m - H_{DEM}$$

Once the residuals were calculated, an outlier detection and removal procedure took place on the raw residual data ( $\Delta H$ ) using a Z score statistical test, as outliers and gross errors in DEMs are

considered as erroneous observations that depart significantly from remaining observations of the sample corrupt statistical accuracy measures. The following equations were used:

$Z_i = (Y_i - \bar{Y})/SD$  and  $|Z_i| > 3SD \rightarrow \text{Outlier}$ , where Y is an observation,  $\bar{Y}$  is the sample mean, and SD is the sample standard deviation. If the absolute value of the calculated  $Z_i$  is greater than three times the sample standard deviation, it is considered an outlier and should be removed. Next, six statistical measures were calculated and evaluated:

1. To compute the correlation coefficient between observed of the control points (X or  $H_m$ ) and its respective elevation from a DEM (Y or  $H_{DEM}$ ), the following equations were used:

$$\rho = \frac{S_{XY}}{\sqrt{S_{XX}S_{YY}}}$$

$$S_{XX} = \sum X^2 - (\sum X)^2 / n$$

$$S_{YY} = \sum Y^2 - (\sum Y)^2 / n$$

$$S_{XY} = \sum XY - (\sum Y)(\sum X) / n$$

Where  $\rho$  represents the correlation coefficient between X and Y,  $\sum X$ ,  $\sum Y$ , and  $\sum XY$  are the summation of the X, Y, and XY variables, and n is the total number of available sample points.

2. The range of  $\Delta H$  is the difference between the maximum and minimum computed from the  $\Delta H$  for each DEM.

$$\text{Range} = \text{Max } \Delta H - \text{Min } \Delta H$$

3. The standard deviation (SD) of  $\Delta H$  is computed with the following equation:

$$SD = \sqrt{\sum (\Delta H)^2 / (n - 1)}$$

where n is the number of utilized points

4. The skewness is calculated by:

$$Sk = \sqrt{n} \sum (Y - \bar{Y})^3 / (\sum (Y - \bar{Y})^2)^{3/2}$$

5. The kurtosis (Kr) is calculated by:



$$Kr = n \sum (Y - \bar{Y})^4 / (\sum (Y - \bar{Y})^2)^2$$

6. Root Mean Square Error (RMSE) is calculated by:

$$RMSE = \sqrt{\sum \Delta H^2 / n}$$

The following table lists the results of each of the following steps when tested with the GDEM and LDEM, respectively:

**Table 2. Statistical measures of GDEM elevations in Delta region (m)**

DEM	Min	Max	Range	Mean	Correlation	SD After Removing Outliers	Kurtosis	Skewness	RMSE	RI
AW3D30	-6.31	4.55	10.86	0.28	0.93	2.18	2.87	-0.52	2.19	4.87
ASTER1 v.2	-14.39	9.69	24.08	-2.57	0.68	4.45	3.05	0.16	5.13	4.13
EARTHEnv-DEM90	-3.13	5.89	9.02	1.51	0.96	1.60	2.89	0.20	2.20	5.47
GLOBE	-14.33	15.70	30.03	-3.14	0.52	6.57	3.13	0.64	7.26	2.93
GMTED2010	-3.52	6.26	9.78	1.22	0.94	1.98	2.93	0.06	2.32	5.47
GTOPO30	-19.95	15.55	35.50	-2.36	0.45	6.94	3.13	0.47	7.32	2.87
SRTM 1	-4.27	7.04	11.32	1.37	0.93	2.24	2.99	-0.06	2.62	5.20
SRTM 3	-4.03	7.43	11.46	1.37	0.95	1.91	3.33	0.22	2.35	5.33

**Table 3. Statistical measures of LDEM elevations in Delta region**

DEM	Min	Max	Range	Mean	Correlation	SD After Removing Outliers	Kurtosis	Skewness	RMSE	RI
IDW	-2.10	3.27	5.36	0.58	0.99	0.85	3.30	-0.23	1.02	9.07
NNIDW	-2.60	3.58	6.18	0.60	0.99	0.90	4.11	-0.14	1.08	6.53
GPI	-15.64	18.85	34.49	0.09	0.70	4.78	6.99	0.16	4.77	4.40
SWB	-2.27	3.06	5.34	0.57	0.99	0.89	3.16	-0.25	1.05	7.47
TTR	-1.17	3.16	4.33	0.65	0.99	0.85	2.82	-0.02	1.07	8.73
Kriging	-1.38	2.81	4.19	0.60	0.99	0.82	2.55	-0.25	1.02	9.27
RBF	-1.94	3.20	5.14	0.58	0.99	0.85	3.10	-0.26	1.03	8.73
EBK	-2.36	3.66	6.02	0.61	0.99	0.96	3.43	-0.04	1.14	6.73
LPI	-12.57	14.22	26.79	0.61	0.94	2.15	30.26	0.27	2.23	5.20

Next, the six statistical measures were categorized in a specific rank which was assigned for each statistical measure based on frequency distribution of the values to assure a uniform distribution

for all the model across the range. For the correlation, range, skewness, RMSE, and standard deviation measures, the ranks range from 1 to 15 based on the ascending order of those measures for each DEM.

**Table 4. Statistical measures ranks**

Rank	Correlation	Range	Skewness	Standard Deviation	RMSE	Kurtosis
1	$0.00 \leq Ri < 0.40$	$50 \leq Ri$	$2.00 \leq  Ri $	$9 \leq Ri$	$24 \leq Ri$	Kr < 3 Ri = 3
2	$0.40 \leq Ri < 0.60$	$40 \leq Ri < 50$	$1.60 \leq  Ri  < 2.00$	$0.8 \leq Ri < 9$	$10 \leq Ri < 24$	
3	$0.60 \leq Ri < 0.70$	$35 \leq Ri < 40$	$1.10 \leq  Ri  < 1.60$	$0.5 \leq Ri < 0.8$	$7 \leq Ri < 10$	
4	$0.70 \leq Ri < 0.75$	$20 \leq Ri < 35$	$0.80 \leq  Ri  < 1.10$	$0.3 \leq Ri < 0.5$	$6 \leq Ri < 7$	Kr = 3 Ri = 8
5	$0.75 \leq Ri < 0.80$	$10 \leq Ri < 20$	$0.70 \leq  Ri  < 0.80$	$0.2 \leq Ri < 0.3$	$5 \leq Ri < 6$	
6	$0.80 \leq Ri < 0.94$	$6.5 \leq Ri < 10$	$0.50 \leq  Ri  < 0.70$	$0.1 \leq Ri < 0.2$	$2 \leq Ri < 5$	
7	$0.94 \leq Ri < 0.95$	$5.5 \leq Ri < 6.5$	$0.40 \leq  Ri  < 0.50$	$0.06 \leq Ri < 0.1$	$1.07 \leq Ri < 2$	
8	$0.95 \leq Ri < 0.98$	$5.0 \leq Ri < 5.5$	$0.30 \leq  Ri  < 0.40$	$0.05 \leq Ri < 0.06$	$1.06 \leq Ri < 1.07$	Kr > 3 Ri = 5
9	$0.98 \leq Ri < 0.99$	$4.5 \leq Ri < 5.0$	$0.20 \leq  Ri  < 0.30$	$0.04 \leq Ri < 0.05$	$1.05 \leq Ri < 1.06$	
10	$0.99 \leq Ri < 1.00$	$Ri < 4.5$	$ Ri  < 0.20$	$Ri < 0.04$	$Ri < 1.05$	

The final step was creating a unique reliability index (RI) and computing it based on the ranks of the statistical measures for each DEM where ( $\Sigma W$ ) denotes the sum of the utilized weight, i.e. 15. It is based on the weighted mean approach, as it ranges from zero as the lowest reliability of a DEM, to a value of 15 which is the highest reliability. RI is computed as:

$$RI = \sum R_i W_j / \sum W$$

**Table 5. Statistical weights**

**Results**

	SD	RMSE	Range	Correlation	Kurtosis	Skewness	Total Weight
Selected Weights ( $W_j$ )	5	4	2	2	1	1	15

It is observed that the GDEMs (EARTHenv-DEM90 and GMTED2010) have the highest RI of 5.47. While the older GDEMs (GLOBE and GTOPO30) had the lowest RI. For LDEMs, the LDEM interpolated from Kriging and IDW SIMs had the highest RI of 9.27, 9.07, respectively.

Finally, the GDEM and LDEM were compared by subtracting the raster maps of the best GDEM and LDEM. The comparison is shown below:

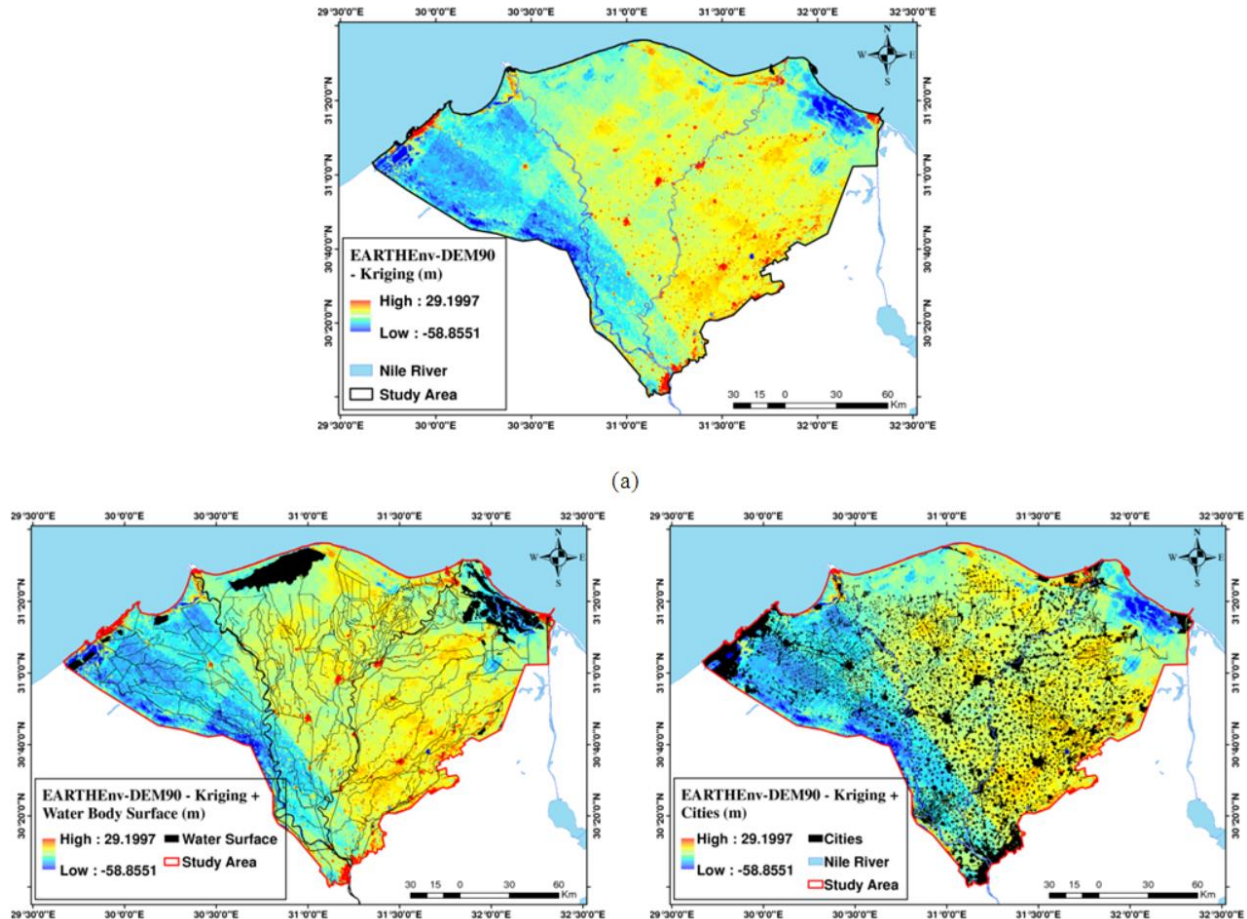


Figure 3. (a) Result of subtracting GDEM EARTHEnv-DEM90 from LDEM Kriging, (b) Using water body surfaces to cover -ve results of subtraction, and (c) Using land use map to cover the +ve results of subtraction

The statistical evaluation of the GDEM and LDEM is shown below:

Table 6. A statistical evaluation for the comparison between the best GDEM and the best LDEM

Max	Min	Mean	SD
29.2	-58.86	-0.02	1.95

## DEVELOPING A FINE-RESOLUTION DIGITAL ELEVATION MODEL TO SUPPORT HYDROLOGICAL MODELING AND ECOLOGICAL STUDIES IN THE NORTHERN EVERGLADES

Xie, Z., Pearlstine, L. & Gawlik, D. E. (2012). *Developing a Fine-Resolution Digital Elevation Model to Support Hydrological Modeling and Ecological Studies in the Northern Everglades*. *GIScience & Remote Sensing*, 49:5, 664-686, DOI: 10.2747/1548-1603.49.5.664.

## Literature Review of Methodologies for Development of a High Resolution DEM for WIPP

This study aimed to integrate elevation survey data and vegetation data at the point and 50 m scales to develop a fine-resolution digital elevation model (DEM) for the northern Everglades. The terrain was divided into lowlands and highlands based on a 50 m scale vegetation map. The DEM in the highlands was interpolated with survey points and later adjusted to connect the vegetation and hydro period (the number of flooded days), while the DEM in the lowlands was interpolated with elevation surveys tagged as lowland types.

The study integrates elevation and vegetation data from USGS High Accuracy Elevation Dataset (HAED) and the South Florida Water Management District (SFWMD) 2004 Vegetation Map to create a 50 m resolution ground elevation model in Water Conservation Area One (WCA1), managed by the U.S. Fish and Wildlife Services at the Arthur R. Marshall Loxahatchee Wildlife Refuge. A 50 m resolution was chosen because it was close to the scale of the slough and ridge structure in the Everglades and it matched the resolution of the SFWMD 2004 Vegetation Map.

### Data

Elevation and vegetation types at elevation points were collected with an Airborne Height Finder (AHF), a field survey built on differential GPS technology, an airborne GPS platform, and a high-tech version of the surveyor's plumb bob. It was specifically designed by USGS to suit the unique surface of the Everglades as it is underwater and obscured by vegetation. Because of the Everglades' terrain, usually common techniques such as LiDAR and Interferometric Synthetic Aperture Radar (IFSAR) were not utilized. The HAED Data is the most accurate elevation data available for the Everglades, collected at 400 m x 400 m grid with general accuracy of AHF of 15 cm.

The SFWMD 2004 Vegetation Map, at the time of this study, was the latest reliable interagency vegetation map. It was produced using a stereoscopic analysis of 1:24,000-scale aerial color – infrared positive transparencies photographed in December 2003.

The daily median water level data from January 1, 2000 to December 31, 2010 was downloaded from the SOFIA website (<http://sofia.usgs.gov/>). The median water level was derived from hourly water level readings at 250 gaging stations. In the case of data gaps, artificial intelligence was used.

From the daily median water-level point data from EDEN project, daily water-level surfaces since 2000 was created using the radial basis function (RBF) interpolation method. The water level surfaces have the same 400 m grid as the EDEN ground DEM that is also available on the SOFIA website.

Water depth data was collected by five research teams during their studies. Depths were collected at 1,515 location from the years 2000-2010.

## 50 m DEM Development

As mentioned above, this study developed a 50 m resolution DEM by HAED data filtering and lowland DEM development with filtered data, highland DEM development with all of HAED data, DEM extraction and merging, DEM assessment, and DEM adjustments. The low lying areas of the Everglades are important because they may hold water in dry conditions and sustain forage fishes for important indicator species (i.e. wading birds, alligators, etc.). The health of indicator species show the success or failure of Everglades's restoration. WCA1 had many small elevated spikes where interpolation may upwardly bias the DEM at low-lying areas. In the release of the 400 m EDEN DEM, a filtering procedure was conducted to remove any HAED point falling on a highland if it was a minority in a 400 m EDEN grid cell. Vegetation types from the SFWMD 2004 Vegetation Map were extracted at HAED points and aggregated into six categories over the 400 m cell. The categories were: (1) Slough or Open Water, (2) Wet Prairie, (3) Sawgrass and Emergent Marsh, (4) Upland, (5) Exotics and Cattails, and (6) Others (wetland shrub and wetland forest). Upland and Others were considered highland because although flooded for most of the year, they are at higher elevation than the surrounding marsh and wet prairie. In this study, a different filtering procedure was implemented using vegetation type data collected at HAED points.

### *Lowland Development*

The HAED vegetation field has 11 types (Alligator Hole, Cattail, Lygodium, Melaleuca, Open Water, Sawgrass, Shrub, Slough, Tree Island, Wet Prairie, and Willow Shrub). This step consisted of selecting HAED points in low-lying areas, or lowland, including: (1) Cattail, (2) Open Water, (3) Sawgrass, (4) Slough, and (5) Wet Prairie. The selected lowland HAED points were then used to develop a DEM with kriging in ESRI ArcGIS 9.3.1. The DEM was then rasterized to 50 m cells, spatially aligned with the 50 m minimum mapping units of the 2004 vegetation map.

### *Highland Development*

The DEM for highlands was developed with all HAED points in WCA1 using kriging in ESRI ArcGIS 9.3.1. All HAED data was used for interpolation because it is not as continuous as the lowlands spatially and most have the form of islands in a lowland matrix. The DEM was rasterized to 50 m cells, spatially aligned with the 50 m minimum mapping units of the 2004 vegetation map.

### *DEM Extraction and Merging*

Once the highland and lowland DEMs were created, they were then merged to form a unified DEM. The 2004 Vegetation Map was utilized to first divide the WCA1 into lowland and

highland regions. The delineated lowland boundaries were used to extract the elevation from the lowland DEM. The two extracted elevation datasets were finally merged into one unified DEM.

*DEM Assessment with Independent Researcher Water Depths*

The unified DEM was then validated and cross validated based on a secondary, independent elevation dataset which was the result of deducing researcher depth measurements from EDEN daily water level surfaces. Because there is a very low slope and slow water flow, the water surface elevation is deemed as flat across a 400 m EDEN cell.

*WCA1 Water-Level Surface Interpolation Model*

A water-level surface model is important to estimate ground elevation from already surveyed depths. The model of the Everglades has been revised by the EDEN team and was concluded to be an improvement. There were higher cross-validation errors in the pseudo-canal stations which were used to represent the abrupt water-level changes across WCA1 and WCA2 which are separated by canals. A water surface model for the sub-area was created for WCA1 by incorporating the conclusions of the previous observation and removing the pseudo-stations.

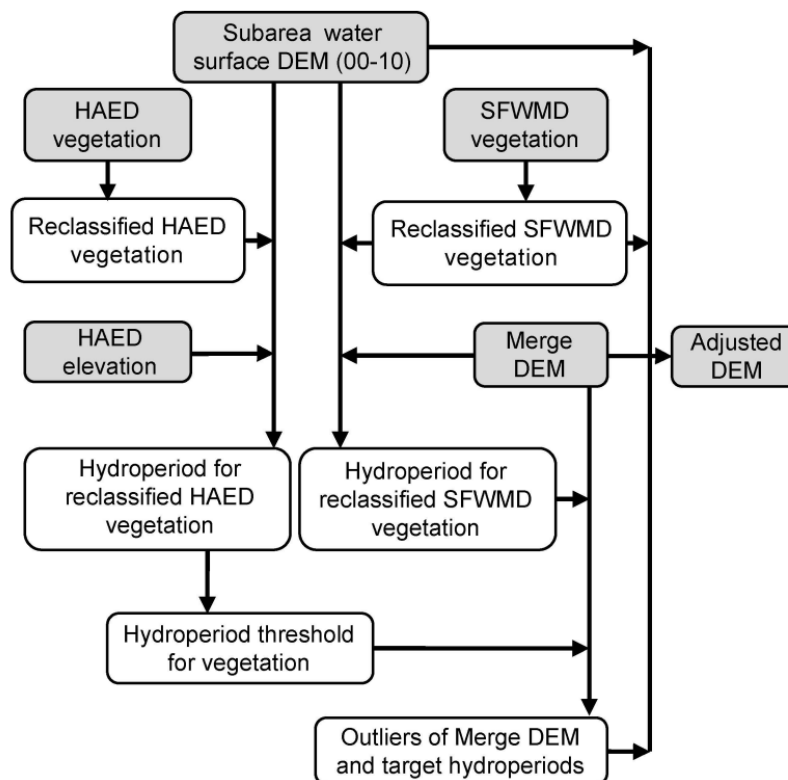


Figure 4. Outline of DEM adjustment workflow

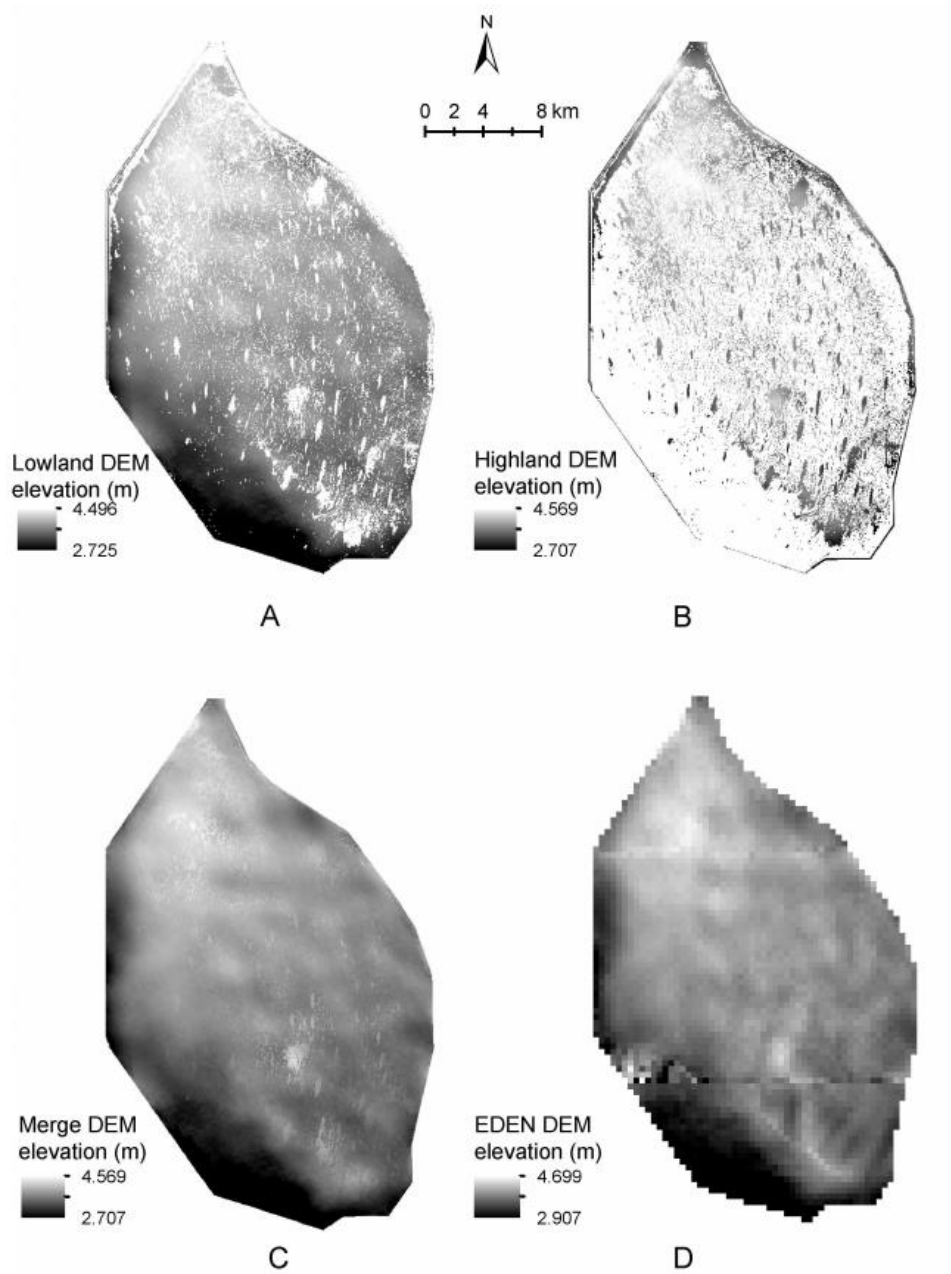
**Results****Table 7. Kriging Models and Cross-Validation Results for the 50 m DEM with Filtered Lowland Data and the 50 m DEM with All Data**

	DEM with lowland data <sup>a</sup>	DEM with all data <sup>b</sup>
<b>Model details</b>		
Kriging method	Universal	Universal
Lag size	400 m	400 m
Number lags	46	46
Trend	1 <sup>st</sup>	1 <sup>st</sup>
Anisotropy	Yes	Yes
Semivariogram model	Spherical (1)	Spherical (2)
Number of points	2285	3536
<b>Cross-validation</b>		
Mean error (m)	-0.003473	-0.003036
Root mean square error (m)	0.1379	0.1589
Average standard error (m)	0.1431	0.1587
Mean standardized (m)	-0.02395	-0.01861
Root mean square standardized (m)	0.9608	0.9978

<sup>a</sup> $0.01535 \times \text{spherical}(18026,13210,306.4) + 0.018527 \times \text{nugget}$ .

<sup>b</sup> $0.019805 \times \text{spherical}(17744,10873,314.1) + 0.022809 \times \text{nugget}$ .





**Figure 5. DEMs extracted with lowland and highland masks for representing the terrain in (a) lowland areas and (b) highland areas. The merged DEM is shown in (c) and the current EDEN DEM in (d).**



## GENERATING HIGH-RESOLUTION DIGITAL ELEVATION MODELS FOR WETLAND RESEARCH USING GOOGLE EARTH™ IMAGERY – AN EXAMPLE FROM SOUTH AFRICA

Hoffmann, E., and Winde, F. (2010) *Generating high-resolution digital elevation models for wetland research using Google Earth™ imagery – an example from South Africa*, *Water SA*, 36:1, 53-68.

The goal of this study was to design a method by which detailed relief/elevation data in Google Earth™ could be converted to a compatible GIS format that could be imported for use in ArcGIS Desktop for development of a digital elevation model for wetland research in South Africa. A procedure was developed to capture Google Earth™ elevation point data and subsequently import them into ArcGIS Desktop. A high-resolution contour map and DEM was then generated in ArcGIS Desktop using the imported elevation data from Google Earth™. An assessment of the quality and reliability of the generated contour map underlying the DEM was then conducted by comparison with the standard CDSM contour data as well as high-resolution relief data generated from airborne survey data. The airborne survey data had an elevation interval of 0.5 m, making it the highest resolution of all the data, so it was used in the assessment after being converted into a GIS-compatible format. Below is a flow path diagram depicting the data sources and steps for generating a high resolution DEM/contour map based on Google Earth™ elevation data and its comparison to other contour datasets.

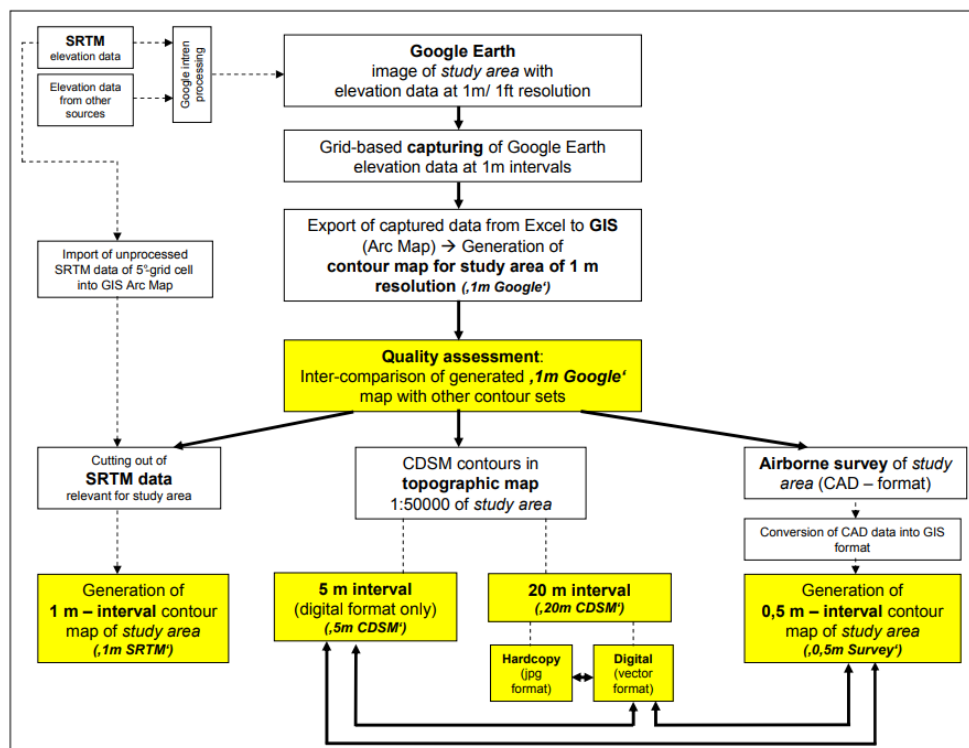
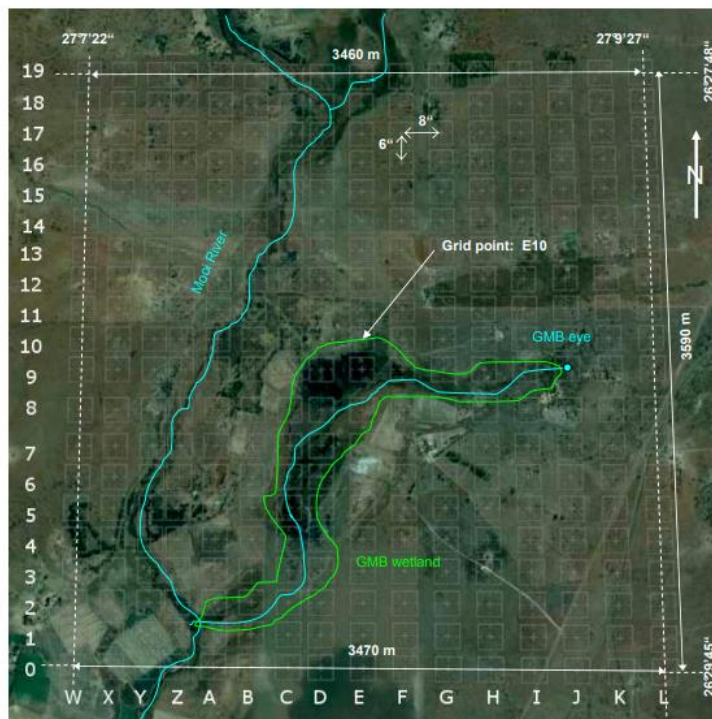


Figure 6. A schematic overview of the process of generating a high-resolution DEM.

### Import of Google elevation data

The first action was to determine the extent of the study area and consider possible hydraulic links between the wetland and the adjacent stream systems, especially overland flow during high flow conditions. The result was the figure below:



**Figure 7. Google Earth™ exported image displaying the target area around the Gerhard Minnebron (GMB) wetland and grid where elevation data at 1 m interval was extracted**

The “display grid option” on Google Earth marked the target area by a grid of markers, with 6” longitudinal and 8” latitudinal difference between points resulting in a total of 16 column and 20 rows. The column were named with capital letters in alphabetical order from west to east. Later, 4 columns were added to the west named W to Z. Together, with the rows 0 to 19 from south to north, each point of the grid received a unique alpha-numerical ID-code. Each grid point was marked by a ‘placemark’, a Google Earth function where additional information (such as elevation) can be recorded and displayed when hovered over with a mouse. Every point where elevation changed by 1m was logged and captured with its coordinates in an Excel sheet. Changes in elevation between grid points were established by slowly moving the cursor along the longitudinal and the latitudinal grid lines until the elevation reading changed. This was done by zooming in to obtain more accurate spatial reference of the point in question. All elevation points were placed between two adjacent grid points (horizontally, east-west, or vertically, north-south) and named by combining the code of the two points separated by an arrow indicating the direction of the cursor movement. After an estimated 30 hours, a total of 2,080 elevation points

were captured covering an area of 12.44 km<sup>2</sup>. Points where the elevation changed by 1m in a W-E direction are classified as ‘horizontals’ while points in a N-S direction are classified as ‘verticals’. In the case where elevation changes more than 1 m between two neighboring grid points, two or more points have the same point name. The coordinates obtained from Google Earth had to be converted from the degree-minute-second format into decimal degrees to be used in ArcView GIS. The conversion was computed with the following equations:

$$\text{DeciDegEast} = \text{deg} + (\text{minute}/60) + (\text{second}/3600)$$

$$\text{DeciDegSouth} = -\text{deg} - (\text{minute}/60) - (\text{second}/3600),$$

Where DeciDegEast and DeciDegSouth are the decimal degree coordinates for the latitudes and longitudes, respectively. Newer versions of Google Earth tools show how latitude and longitude can be changed, avoiding manual calculations.

### **Generating contour maps and DEM in GIS ArcDesktop**

Out of the 2,080 points that were logged, only 1,760 points (where the elevation changed by 1 m) were used to generate a contour map, resulting in a density of one elevation point per 7,069 m<sup>2</sup>. The remaining grid points were omitted because the change in elevation was not by 1m. If the remaining 320 grid points were used, there would have been an error margin of up to 1 m per point. By using the Inverse Distance Weight (IDW) technique of ArcGIS Desktop to interpolate between these points, a 1 m interval contour map was created. Based on the retrieved set of elevation data from Google Earth, a DEM was generated in ArcGIS Desktop, enabling a 3D view of the area.

### **Assessing quality of the generated contour map**

In order to assess the reliability of the newly generated contour map and the DEM of the area, it was compared against two sets of standard topographic data (20 m and 5 m contour levels from CDSM). The two sets of standard topographic data were also compared to determine how much they deviate from each other.

As well as elevation data from an aerial photography survey provided use to this study. The data consisted of 0.5 m contoured elevation data, making it the highest resolution, which in turn was used as the benchmark against which all the datasets were compared. The digital elevation data had to be converted from a computer aided design (CAD) program to a useable format for use in ArcView GIS.

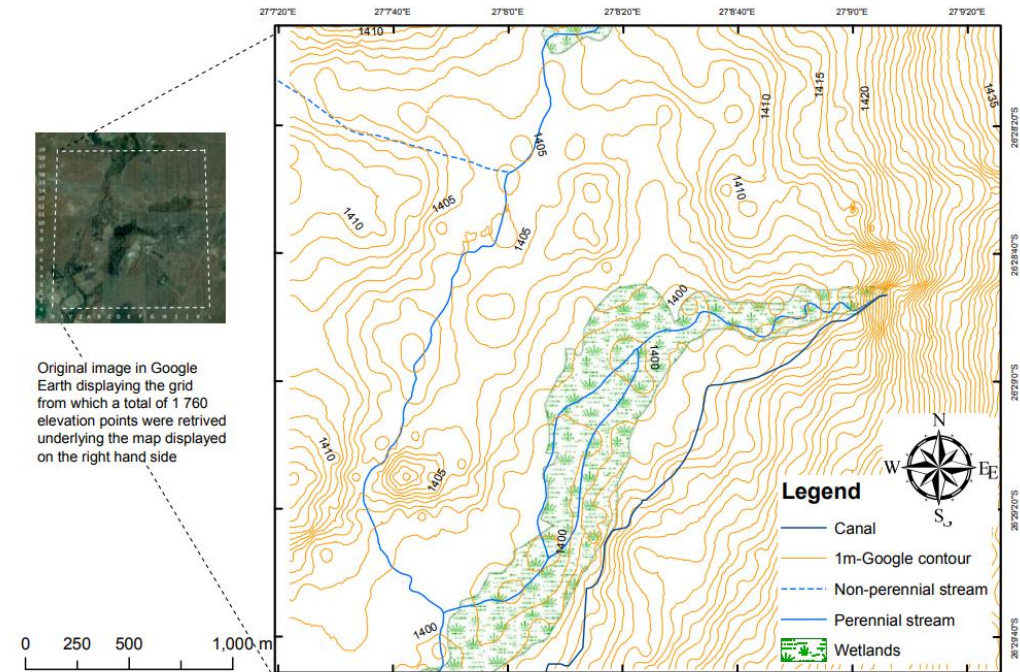


Figure 8. Map of the study area displaying 1m interval contours that were created using IDW interpolation in GIS ArcGIS Desktop based on 1,760 point elevation data retrieved from Google Earth.

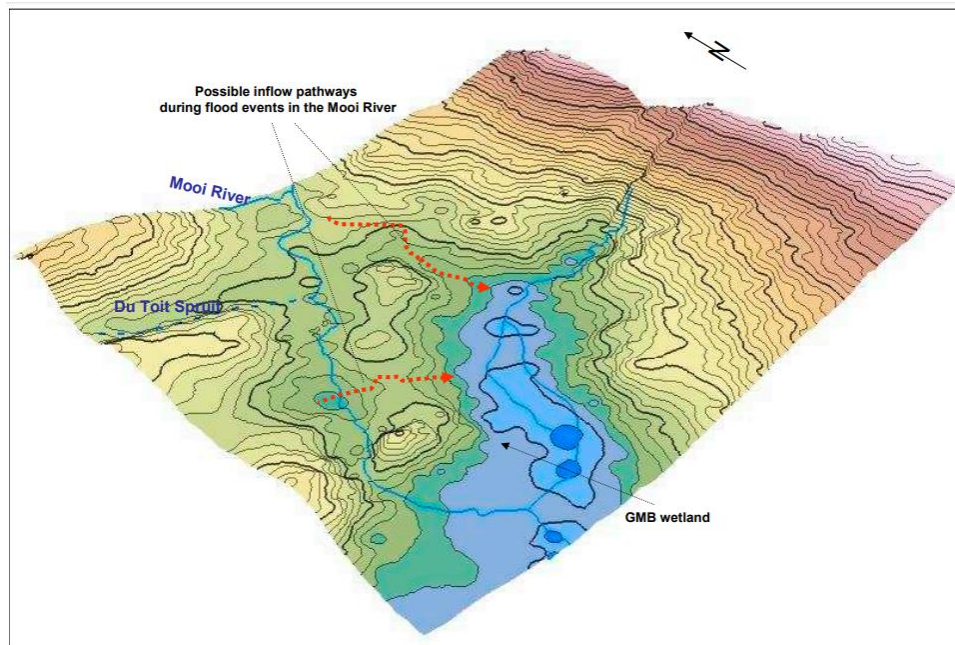
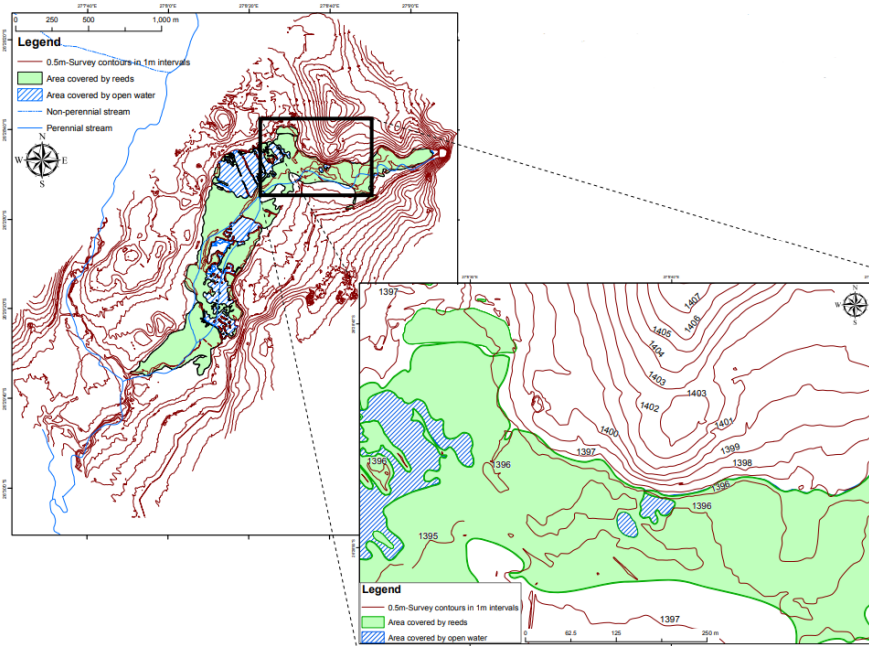


Figure 9. Color-coded DEM (based on IDW interpolation) of the wider study area based on 1 m interval Google Earth elevation data (from the SRTM) showing possible event-driven flow paths.



### Airborne relief mapping (0.5 m contour interval)

Mapping is based on a stereoscopic evaluation of a series of aerial photographs taken from an aircraft at a height of 4000 m above ground level, accurate enough for use in 1:2,000 scale plans. Twelve ground-control points of known elevation and coordinates were used to georeference the aircraft images. Georeferenced stereoscope images were created and a contour map was drawn with these images and a stereo plotter.

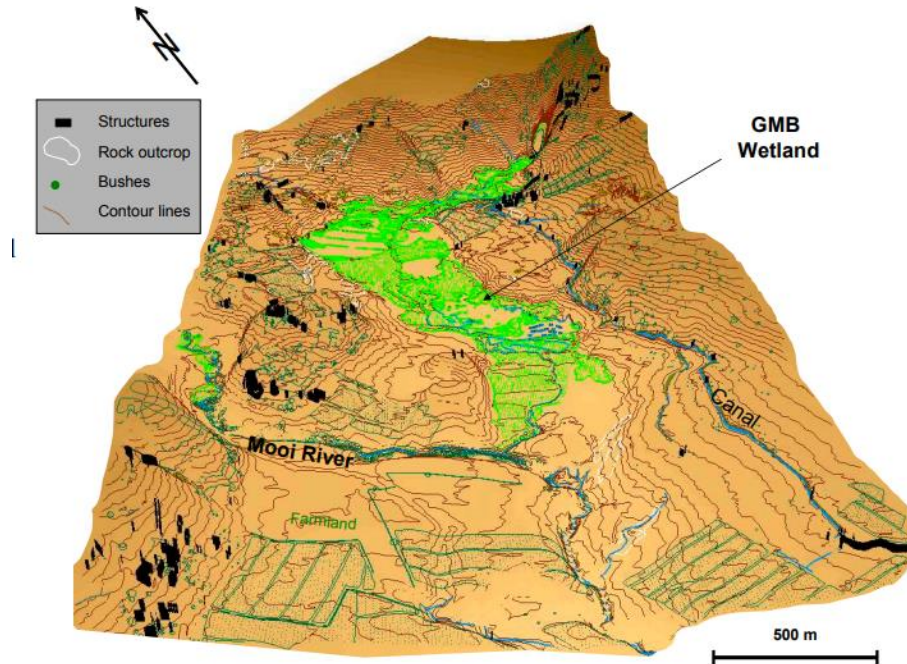


**Figure 10. Map with 0.5 m contour map generated from aerial photography displaying reed cover and open water areas.**

The data from the airborne survey was digitized on a CAD program locking all feature information in a single CAD file (design files, .dgn). Colors and symbols were then utilized to distinguish features.

The conversion of CAD data into GIS-compatible shapefiles was complicated due to the fact that symbols consisting of multiple components, i.e. dashed lines, were identified as separate parts and as a standalone topographic feature. While the single components are automatically grouped into one larger shapefile (e.g. a catchment boundary), they cannot be used for determining enclosed areas. This is only possible once each component of the shapefile is connected to others to form a larger entity (such as the 'merge' function in the 'Editor' tool of ArcGIS Desktop). Before areas enclosed could be determined, the components had to be connected to form a continuous line which could then be transformed into a polygon using the 'buffer' and 'union' functions on ArcGIS Desktop. Once the polygons are created, the size of the enclosed areas can be determined.

Compared to the 1:50,000 topographic map, spatial deviation between the GIS-imported CAD and mapped topographic features was detected. This could be because of inaccuracies in georeferencing and projection. But for the aim of the study, the inaccuracies can be disregarded and deemed tolerable. The DEM produced from CAD data is different from the Google-based DEM in that it was based on CAD generated contours and not on interpolated elevation points.



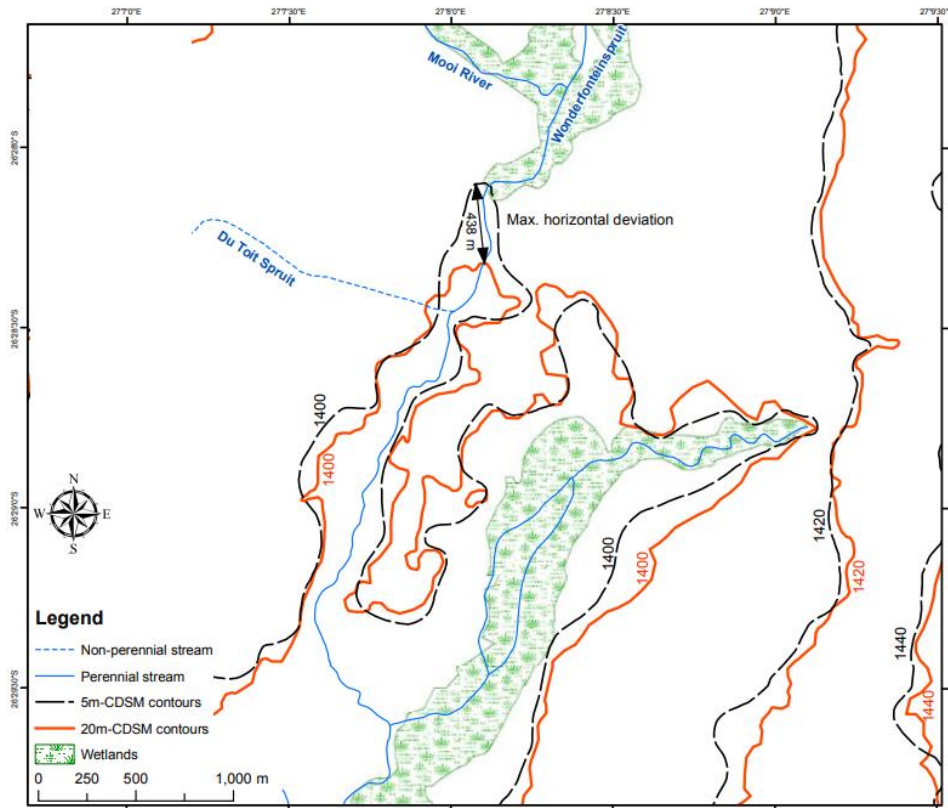
**Figure 11. DEM based on aerial photography survey contour data (0.5 m intervals) including features such as rock outcrops, structures, vegetation, rivers, and canals.**

### **Comparing Google and CDSM data**

Before comparing 20 m interval CDSM contour data with the 5 m CDSM contours, it was observed that the contour data in vector format (i.e. discrete .shp files) showed deviation from identical contours as displayed in scanned images of 1:50,000 topographic maps imported into ArcGIS Desktop. The topographic image's contour lines were used as the base dataset to adjust the digital (vector) contour lines. The adjusted set of the digital 20 m interval contour data was then used for the following comparisons.

### **Deviations between same-source data: 20 m CDSM vs 5 m CDSM contours**

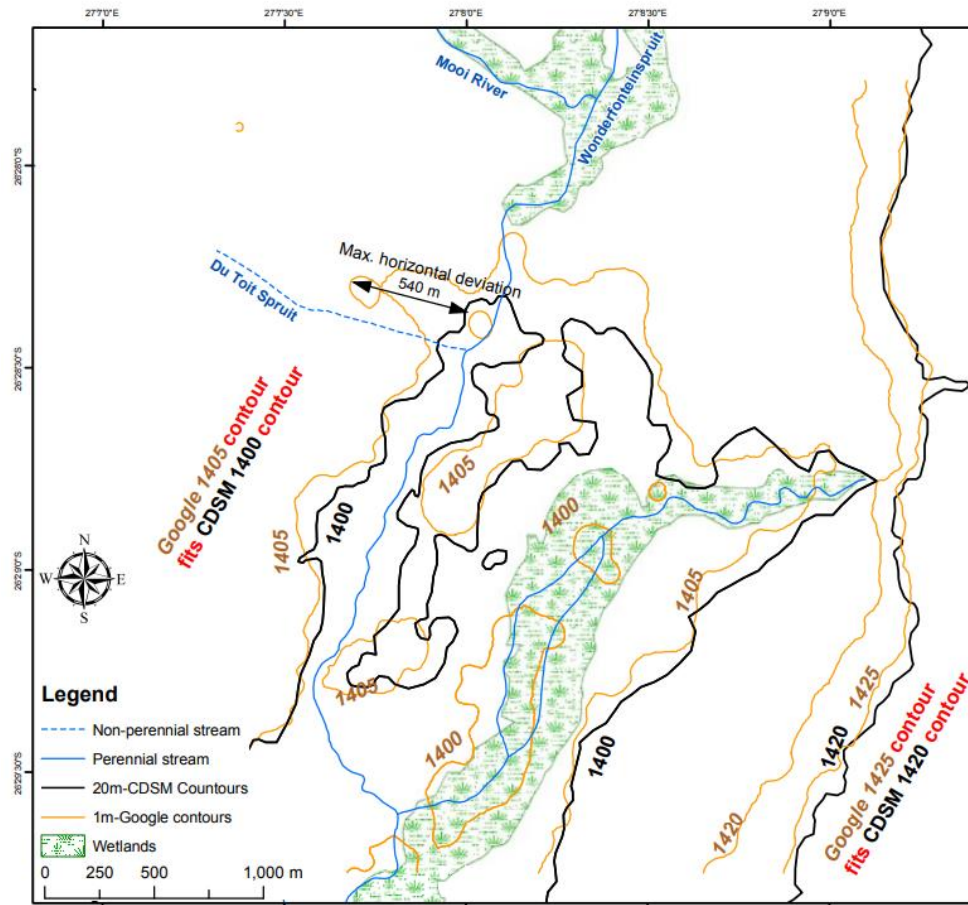
Because of the flat topography, only 3 x 20 m interval contours occur within the larger study areas and can be compared to the 5 m interval contours. When observing the deviation between contour lines, it is noted that identical degrees of deviation in steep terrain results in smaller horizontal distances between contour lines than in flat terrain where the lower topographic gradient results in a larger horizontal distance between contour lines.



**Figure 12. Deviation areas between 20 m and 5 m CDSM contours illustrated using the 1,400, 1,420, and 1,440 m a. m. s. l. contour lines of both sets.**

### **1 m Google contours vs 20 m CDSM contour data**

To compare the 1 m contour map ('Google contours'), only the 2 x 20 m CDS contour lines ('CDSM contours') that are in the study area could be used. Namely, the 1,400 m and the 1,420 m a. m. s. l. lines. Comparing these with the corresponding contour lines, the ones generated on Google, there is a satisfactory fit between the two sets of contours. Google contour lines, however, matched the best with CDSM contours which generally displayed a 5 m lower elevation. It is not known why there is a 5 m offset, but it could be easily adjusted by subtracting the 5 m from all contour heights in the generated Google map. This was the case with the 5 m vs 20 m CDSM contour data, where the largest horizontal deviation occurred at the 1,400 m contour at the confluence area of WFS and Mooi River. Based on the maximum horizontal deviation, it is apparent that the generated Google map of 1 m fits well with the 20 m CDMS contours with a max horizontal deviation of 438 m. It was concluded that the Google generated map displayed an even better fit to the standard 20 m CDSM contour than the 5m CDSM generated map.



**Figure 13.** 1 m Google contours compared with 20 m CDSM contours demonstrating a good fit with a +5m offset in Google data using the 1400 m and 1420 m a.m.s.l. contour lines which show a horizontal misfit of 540 m between the 2 sets of contour lines.

### 1 m Google contours vs. 5 m CDSM contour data

When comparing 1 m Google contours with 5 m CDSM contours, there was an overall good fit between the two data. The largest horizontal deviation occurs on the 1,400 m contour line next to the convergence of the Mooi River and the Du Toit Spruit. As previously mentioned, this is most likely because this area is flat; there are larger horizontal misfits than identical misfits in steeper terrain.



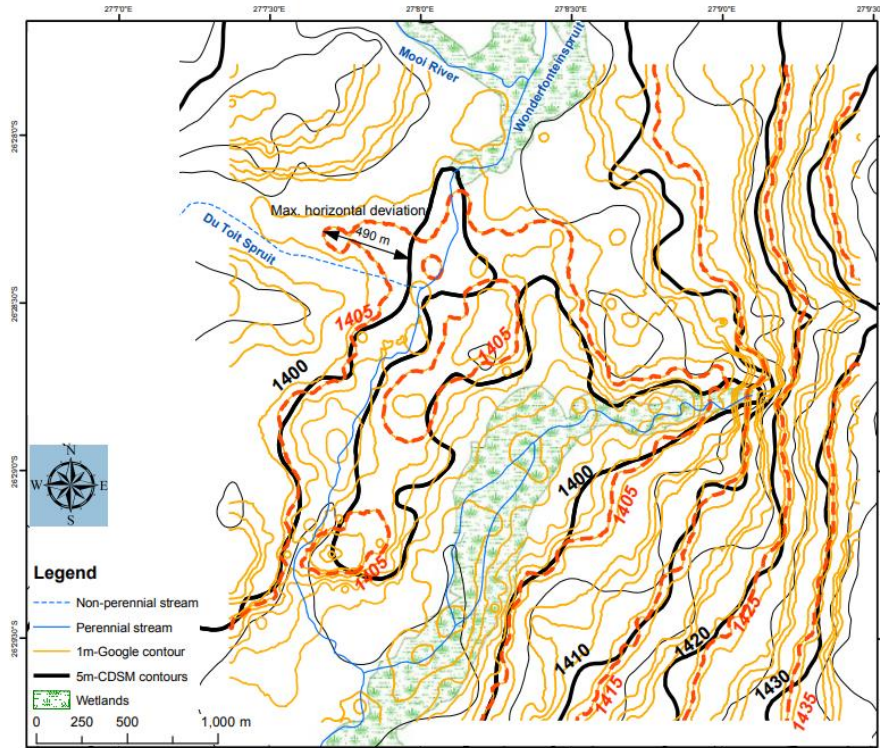


Figure 14. Selected 1 m Google contours (bold dashed line) compared against corresponding 5 m CDSM contours (consider the 5 m offset)

**Comparison of all contour data to a high-resolution airborne survey: 0.5 m contour interval**

The survey was done at two different accuracies with the GMB wetland mapped at a scale of 1:4,000 and the wider study area (the convergence of Mooi River and the WFS) at a scale of 1:8,000. The contour interval for both was 0.5 m. The location of the areas using airborne survey data is shown below.

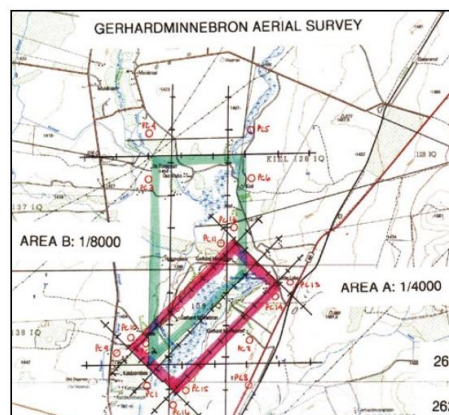
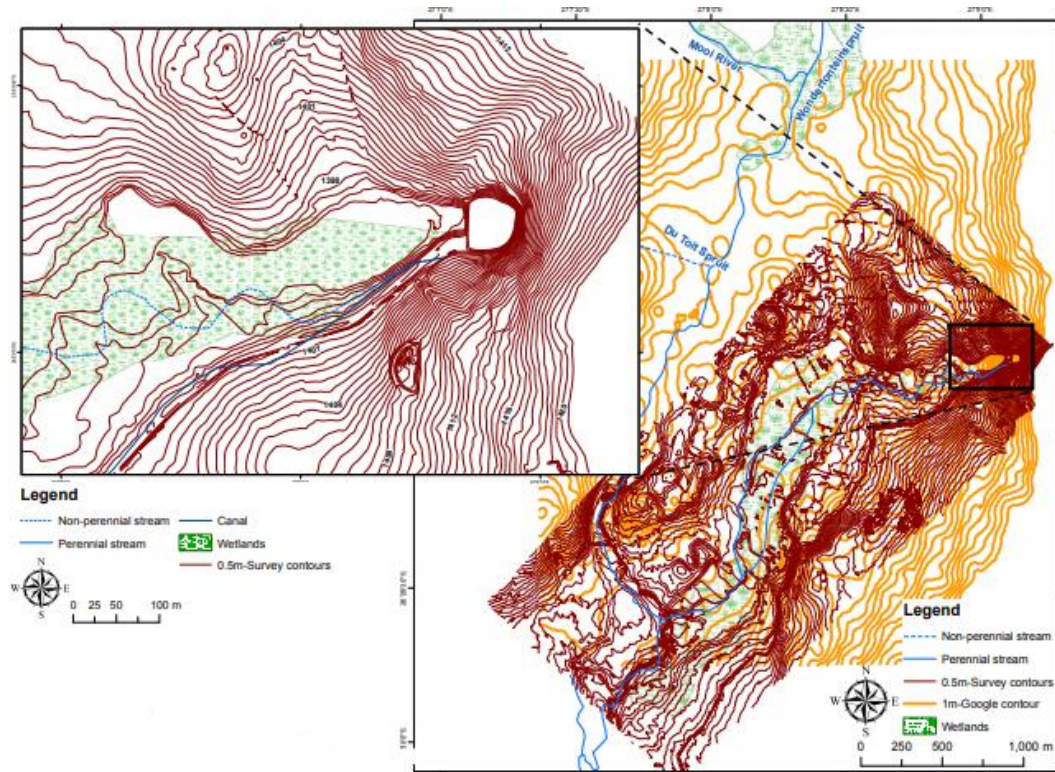


Figure 15. Location of study area using airborne survey data

For the purpose of comparing contour lines, only the area mapped at higher accuracy (1:4,000) covering the wetland was used. The 0.5 m contour interval map produced is shown in the following figure.



**Figure 16. 0.5 m survey contours for the wider study area as well as the area with wetland**

The noticeable lack of contour lines in the wetland demonstrates the flatness of the area, where differences in the micro-relief influencing flow patterns are usually below 0.5 m interval.

### ***20 m CDSM contours***

The aerial photography survey did not require an offset correction when measured against 2 contours from the 20 m CDSM contour dataset, meaning there is good degree of congruence between the 2 contour sets. Because of the limited area at which the resolution was done, only parts of the 20 m contours can be compared against corresponding 0.5 m survey contours. For the area covered, a comparatively small horizontal deviation of max 150 m at the 1,400 contour line indicates a good fit between the two datasets.

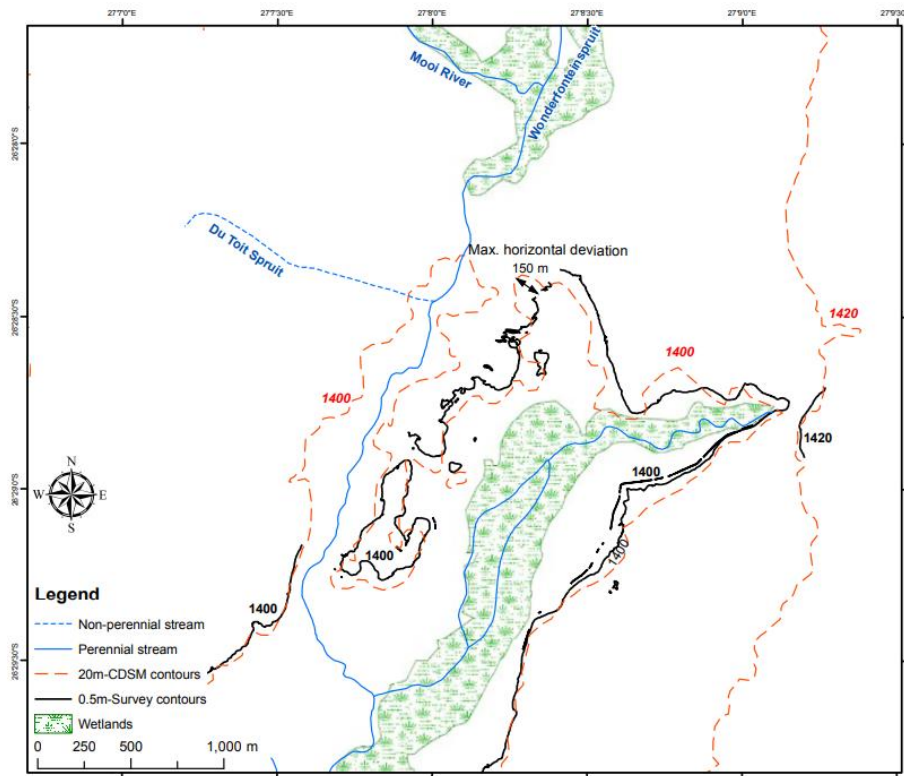


Figure 17. 20 m CDSM contours compared against 0.5 m survey contours

### *5 m CDSM contours*

The 0.5 m survey contours and the three relevant 5 m CDSM contours (i.e. 1400 m, 1405, and 1410 m) match well. The maximum deviation was not associated with the lowest lying contour line in the study area (1,400 m a. m. s. l.) but did occur at the 1,405 m contour where it sums to 140 m.

### *1 m Google contours*

The Google contours had to be corrected by reducing the displayed elevation for each contour by 5 m in order to compare the two datasets. The deviation between the 0.5 m survey and 1 m Google contours was more complicated with a maximum horizontal deviation of 221 m. There is also a difference between the shape of 1 m and 0.5 m contour lines covering the area.



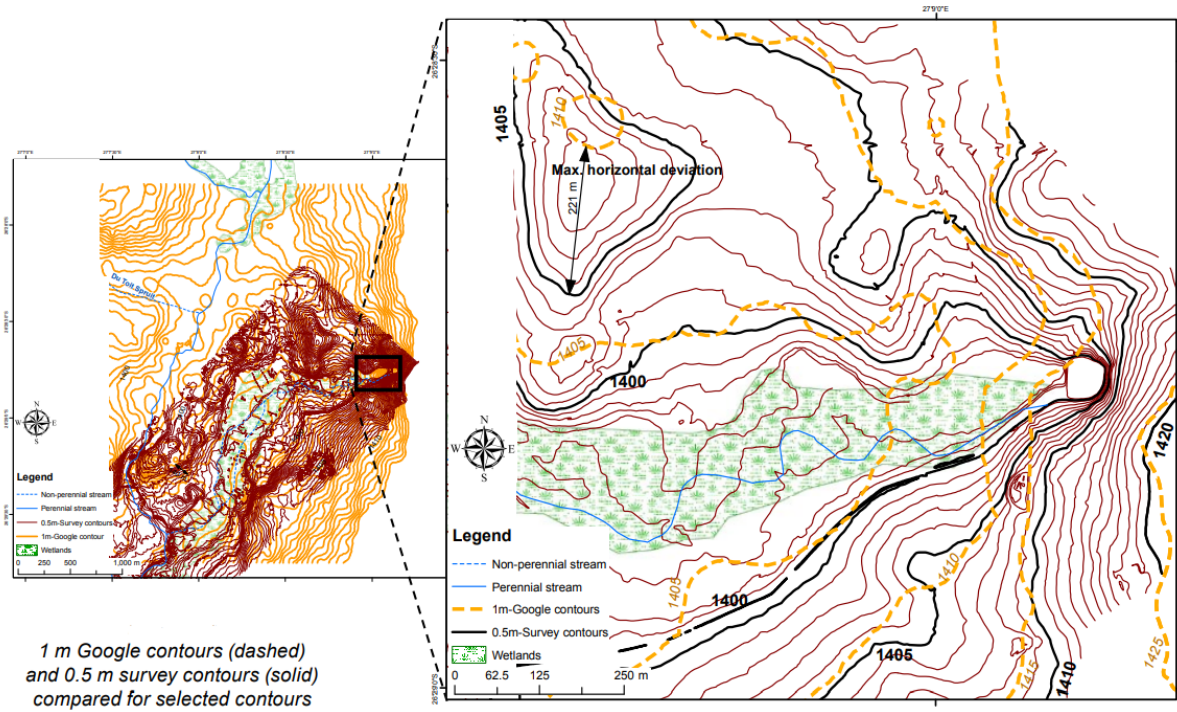


Figure 18. 1 m Google contours (dashed) and 0.5 m survey contours (solid) compared for selected contours

Table 8. Overview of results of the inter-comparison of different contour datasets available for the study area with the generated 1m Google contour map

	1 m SRTM	5 m CDSM (vector shp)	20 m CDSM (image - jpg)	0.5 m survey
1 m Google	<p><b>Figs. 5 and 6:</b></p> <ul style="list-style-type: none"> <li>- SRTM contours are blocky and often open-ended</li> <li>- Horizontal misfit of approx. 50 m between original 5° x 5° grid cell data and those cut out for study area</li> <li>- 1 400 m contour is split in 3 parts and deviates from corresponding Google contour</li> </ul>	<p><b>Fig. 17:</b></p> <ul style="list-style-type: none"> <li>- A vertical off-set of 5 m was detected</li> <li>- After correcting this off-set an overall good fit appeared</li> <li>- Max. horizontal deviation: 490 m (1 400 m contour line)</li> </ul>	<p><b>Fig. 16:</b></p> <ul style="list-style-type: none"> <li>- 2 x contour lines evaluated (1 400/1 420 m)</li> <li>- A vertical off-set of 5 m was detected</li> <li>- After correcting this off-set an overall good fit appeared</li> <li>- Maximal horizontal deviation 540 m (1 400 m contour line)</li> <li>- 1 m Google contours appear somewhat smoother than 20 m CDSM contours</li> </ul>	<p><b>Fig. 22:</b></p> <ul style="list-style-type: none"> <li>- A vertical off-set of 5 m was detected</li> <li>- More complex pattern of deviation: <ul style="list-style-type: none"> <li>→ horizontal misfit max. 221 m (1 405 m contour line)</li> <li>→ shape of contours</li> </ul> </li> <li>- Deviations mainly confined to flat areas; in steep terrain generally good fits</li> </ul>
5 m CDSM (vector - shp)	-	-	<p><b>Fig. 15:</b></p> <ul style="list-style-type: none"> <li>- 3 x contour lines evaluated (1 400/1 420/1 440 m)</li> <li>- Maximal horizontal deviation 438 m (1 400 m contour line)</li> <li>- Positive and negative deviations occur</li> <li>- 5 m contour line appears to be smoother than 20 m line</li> </ul>	<p><b>Fig. 21:</b></p> <ul style="list-style-type: none"> <li>- 3 x contour lines compared (1 400/1 405/1 410 m)</li> <li>- An overall good fit</li> <li>- Max. horizontal deviation: 140 m (1 405 m contour line)</li> </ul>
20 m CDSM (vector - shp)	-	-	<p><b>Fig. 14:</b></p> <ul style="list-style-type: none"> <li>- At 1:50 000 map for study area a horizontal misfit of approx. 30 m was detected (1 400 m contour line)</li> <li>- Positive and negative deviations occur</li> </ul>	<p><b>Fig. 20:</b></p> <ul style="list-style-type: none"> <li>- only 2 x 20 m contour lines could be compared</li> <li>- Overall good fit</li> <li>- Maximal horizontal deviation: 150 m (1 400 m contour line)</li> </ul>

## **DEM DEVELOPMENT FROM GROUND-BASED LIDAR DATA: A METHOD TO REMOVE NON-SURFACE OBJECTS**

*Sharma, M., Paige, G. B., & Miller, S. N. (2010). DEM development from ground-based LiDAR data: A method to remove non-surface objects. Remote Sensing, 2(11), 2629-2642.*

In this study, a method was created to remove vegetation from ground-based LiDAR data to create high-resolution DEMs. Research was conducted on intensively studied rainfall-runoff plots on the USDA-ARS Walnut Gulch Experimental Watershed in Southeast Arizona. LiDAR was used to create 1 cm digital surface models (DSM) for 5 plots. DSMs created from LiDAR data has non-surface objects like vegetation cover. A vegetation removal method was created using a slope threshold and a focal mean filter method to remove vegetation and create a bare earth DEM. It was then validated on a synthetic plot, where rocks and vegetation were added in increments.

According to the study, the removal of non-surface objects to develop DEMs from LiDAR data DSMs can be difficult and time consuming. This study focuses on developing a method to remove non-surface objects from high-resolution ground-based LiDAR DSMs to develop bare earth DEMs for small plots using single returns because of the lack in difference between the first and last returns. This was tested using ground-based LiDAR from rainfall simulator plots and validated on a synthetic laboratory plot where the placement of surface objects was controlled.

### **Methodology**

#### *Study Area*

The study was conducted on the Kendall 112 sub-watershed in the USDA-ARS Walnut Gulch Experimental Watershed (WGEW), Tombstone, Arizona. It is a grass-dominated, small 1.91-ha watershed with an average slope of 9.4%. The rainfall simulator plots are 2 m by 6 m in size and serves as the test bed for development of bare earth DEMs from ground based LiDAR.

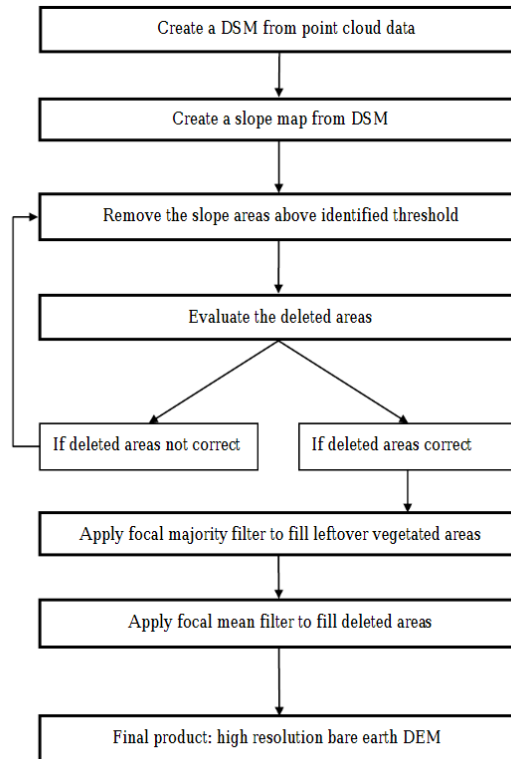
#### *Ground-Based LiDAR Unit and Data Acquisition*

To collect the high-resolution terrain data, an Optech ILRIS 3-D ground-based LiDAR laser scanner was used. The time it takes for the light to travel out to the target and back to the scanner is used to determine the range of the target. The deflection system directs a laser beam in the direction where scanning is taking place. Accuracy of the distance measurements depends on the intensity of the reflected laser light which correlates to the reflectivity of the object. The Optech system is a pulsed system with a vertical accuracy of 0.3 cm, a field view of 40° (horizontal) x 40° (vertical), and a measuring rate of 20 kHz. The laser wavelength is 1500 nm. The laser beam class is Class I with a beam divergence of 1.00974°, and a minimum spot step in X and Y axis of 0.00115°. The raw range accuracy of point cloud data is 7 mm at 100 m and raw positional

accuracy is 8 mm at 100 m. The device can get either the first or last return. The difference can be attributed to a result of the pulse passing through very sparse vegetation or other substance that can cause multiple returns; dense vegetation, rocks, soil, or other hard objects do not generate a second return. The study used the last return to remove any confusion resulting from the occasional pulse traveling through light vegetation and remained consistent throughout the study.

The LiDAR scanner was used to collect surface data in a non-spatially referenced 3-D point cloud for each of the five plots. The short wavelength was useful to ensure accurate distance to any surface object that is in the path of the laser beam. This creates a shadow effect behind the vegetation where no surface data can be recorded. To account for this shadow effect and create a DSM, LiDAR scans were taken from multiple directions and then joined together to create an integrated 3-D dataset of each plot using 3-D data processing software. Common control points were established on each of the plots and were included in all scans for the plot. The common points were then used to align and merge the point cloud data images. These images were then aligned according to their best-fit alignment and comparison using a threshold of 0.01 standard deviation. If there were overlapping areas, they were thinned and composite images were merged to produce a single point cloud DSM for each plot.

The resulting DSM for each plot was subjected to the vegetation removal process to produce a bare earth DEM. After creating a DSM, the first step was to create a slope map from the DSM using GIS algorithms. The second step was to determine the slope threshold for each plot that could be used to identify and remove vegetation. The vegetation areas on each plot have higher local slopes than the adjacent surface topography. The slope threshold range was based on an evaluation of the 5 study plots in Kendall 112 watershed and an additional set of 5 shrub-dominated plots in WGEW and determined by comparing the slope map and photographs for the plots. The vegetated areas were “removed” from the slope map and converted to null values. A limitation of using a simple slope threshold is that the tops of the vegetation demonstrate low slopes but at higher elevation levels than the surrounding surface topography. The “leftover” vegetation areas were identified by checking the elevation in deleted areas as compared to the adjacent non-deleted areas and comparing the DSM with a photograph of the respective plot. The final step consisted of filling the null value areas to create a bare earth DEM. A 5 x 5 focal mean filter (with pixels of size 1 cm<sup>2</sup>) was used. It was specified that only cells with values were to be used to find the focal mean of the target cell. If a NoDATA value existed within a neighborhood of the focal mean filter, then its value would be ignored. This approach minimizes the smoothing effect on the DEM. The NoDATA values were filled with the mean cell value of the neighborhood. The filter was then run several times in an iterative process from the edge of the patch until all of the NoDATA values were filled and a bare earth DEM was generated.



**Figure 19.** Flow diagram depicting the steps used to remove vegetation from DSMs.

### *Validation*

To validate their method, a synthetic experimental plot was created. The synthetic plot of 2 m by 4 m was made in an indoor facility where rocks and other vegetation were manually placed to create pre- and post-vegetation surfaces. The plot was scanned at 5 mm resolution and following the same procedure as the field study. The first scan was with rock cover and no vegetation to create a true bare earth surface model. Vegetation was added to the plot for the next scans to create the DSM. As with the field plots, multiple scans were taken to with and without vegetation and merged together to develop an integrated 3-D bare earth DEM and DSM, respectively.

To evaluate the results, a difference map (DSM minus final DEM) was created to evaluate the elevation values and location of the vegetation areas removed from the DSM to create the final DEM. This allowed the direct comparison of the removed and filled areas to locations where vegetation was placed. The second step was to compare the final DEM to the original bare earth DEM and create an error map using the following equation to show the distribution of error in the final DEM.

$$\% \text{ Error Map} = \frac{(DEM - \text{Final DEM})}{DEM} \times 100$$

## Results and Discussion

the vegetation removal method was carried out on all five of the grass and shrub-dominated plots on the Kendall sub-watershed. A slope threshold range of  $58^{\circ} - 65^{\circ}$  was determined to be appropriate for deletion based on evaluation of the rainfall runoff plots. The slope threshold ranged from  $58^{\circ}$  to  $64^{\circ}$  for the grass-shrub dominated plots in Kendall 112 and  $59^{\circ}$ - $65^{\circ}$  for the additional shrub-dominated plots that were used to test the threshold. It is important to observe that the top of vegetation can show up in the DSM as areas of high elevation-low slope area.

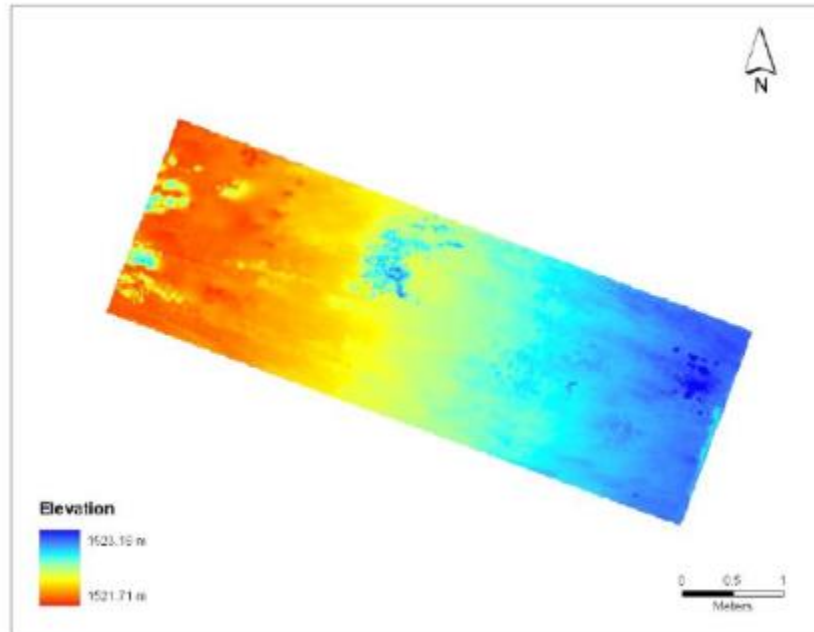


Figure 20. DSM created from XYZ point cloud data collected from ground-based LiDAR unit



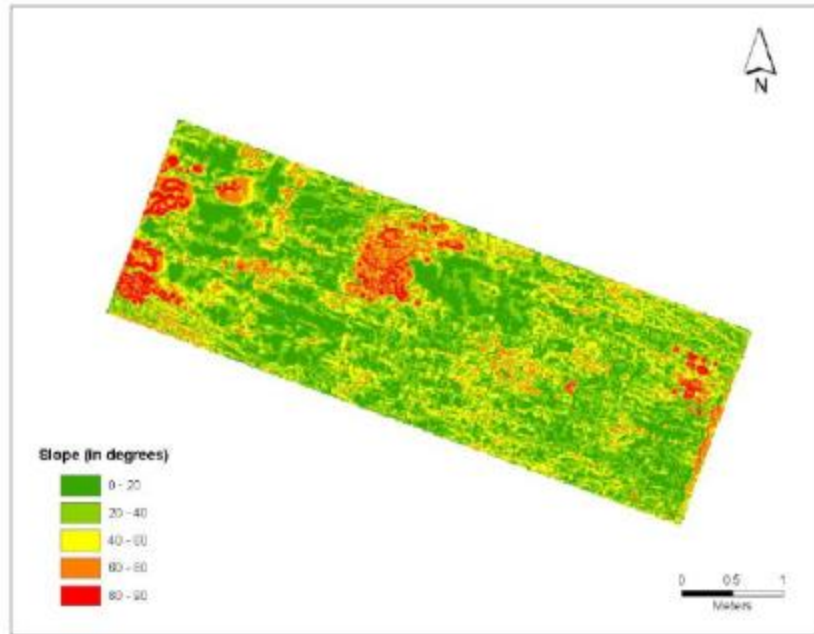


Figure 21. Slope map of the plot. Slope values derived from 1 cm DSM of plot.

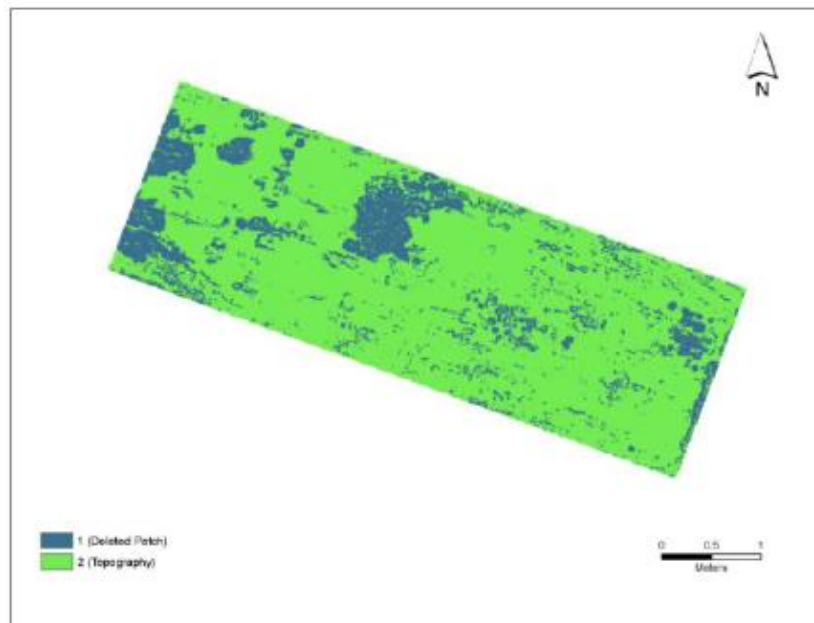
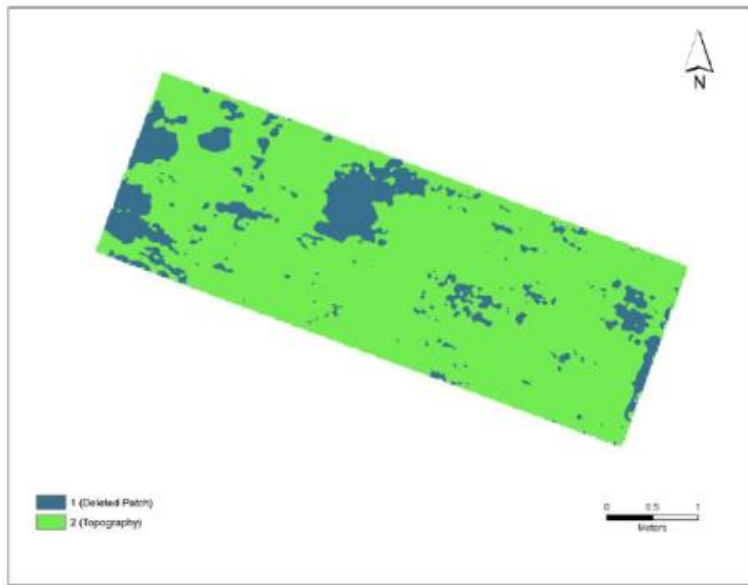


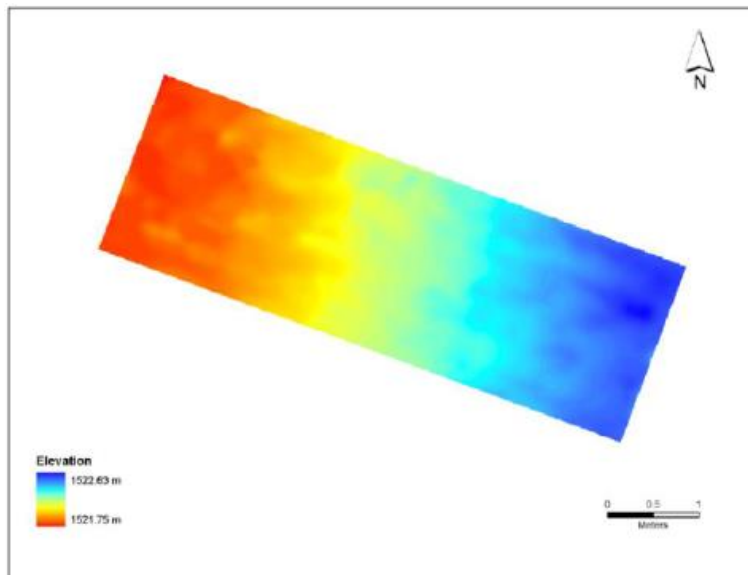
Figure 22. Plot with vegetation areas removed. A slope threshold of 60o was used to delete the vegetation patched. Blue patches were converted to null values.

The final step in vegetation removal consisted of filling the deleted areas in the raster image using a focal mean filter to create the bare earth DEM. Results show that there is a maximum 60 cm elevation, which is related to grass and shrub vegetation that was removed from different areas of the DSM. The highest local elevation value in the DSM was 1,523.17 m compared with

1,522.63 m in the bare earth DEM. The range of elevation removed from shrub-dominated plots was 40-62 cm and from 5 grass-dominated plots was 23-44 cm.



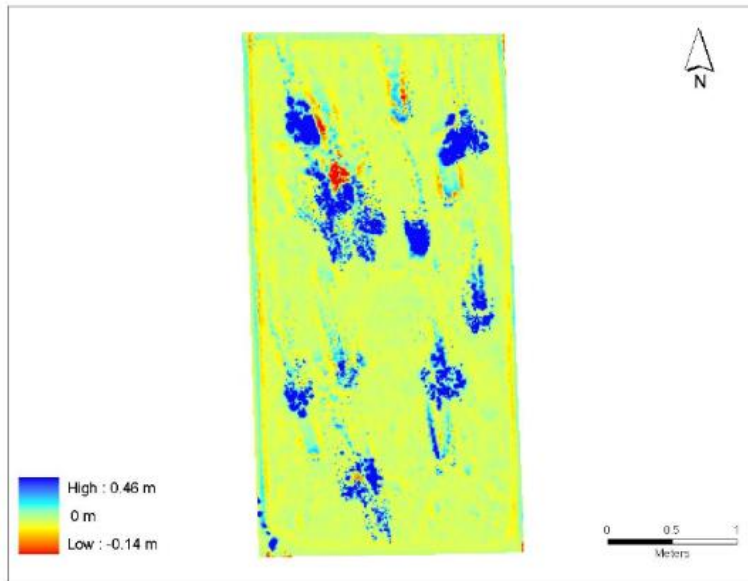
**Figure 23. Raster image after 5 x 5 focal majority was run. Blue patches show removed vegetation areas.**



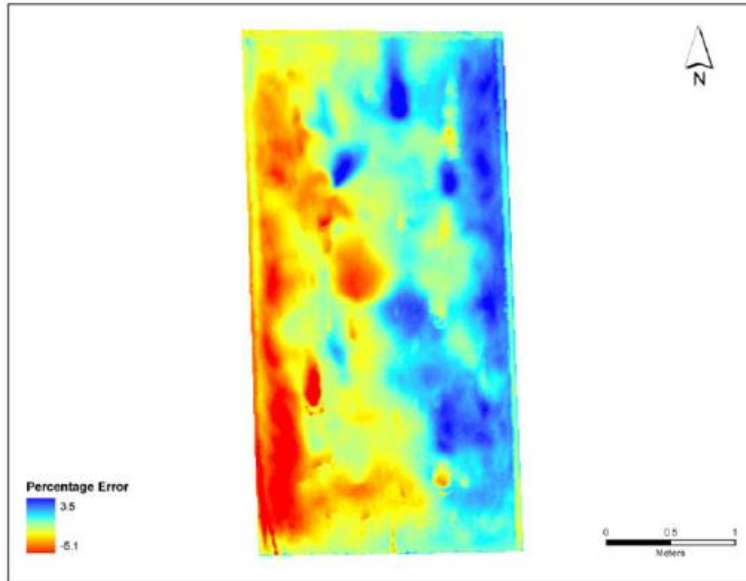
**Figure 24. Bare earth DEM at 1 cm resolution. The vegetation was removed from the DSM and deleted areas filled with focal mean filter.**

**Table 9. Comparison of elevation values between DSM and DEM for the field plot**

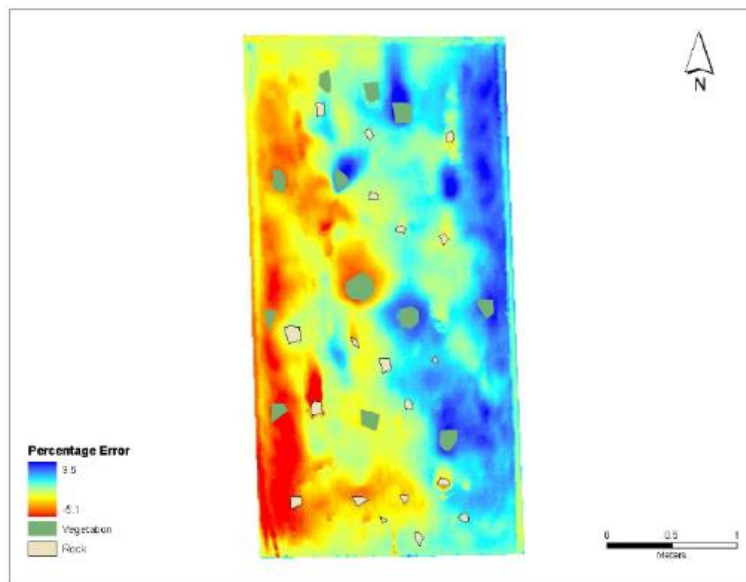
	DSM	DEM
Minimum (m)	1521.71	1521.75
Maximum (m)	1523.17	1522.63
Mean (m)	1522.18	1522.16
Standard Deviation	0.25	0.25



**Figure 25. Difference map showing areas where vegetation was removed from the DSM for the validation plot. The dark blue color indicates areas of higher elevation (height of vegetation). Red color indicates areas that are higher in the final DEM.**



**Figure 26. Percent error map for the final DEM. Dark blue areas indicate an error of 3.5% (4.9 mm) while dark red areas indicate an error of -5% (7.5 mm). Yellow to light blue areas show areas of negligible error.**



**Figure 27. Percent error map for the final DEM showing polygons for known locations of the rocks and vegetation on the validation plot.**

**Table 10. Comparison of elevation values between the DSM and the final DEM for the validation plot.**

	DSM	final DEM
Minimum (m)	3.87	3.92
Maximum (m)	4.78	4.80
Mean (m)	4.34	4.36
Standard Deviation	0.22	0.23

## CREATING HIGH-RESOLUTION DIGITAL ELEVATION MODEL USING THIN PLATE SPLINE INTERPOLATION AND MONTE CARLO SIMULATION, WORKING REPORT

*Pohjola, J., Turunen, J. and Lipping, T. (2009). Creating High-Resolution Digital Elevation Model Using Thin Plate Spline Interpolation and Monte Carlo Simulation, Working Report 2009-56.*

### Thin plate spline interpolation

In this report, thin plate spline interpolation was used to create a high resolution DEM. It is described as if bending a thin metal plate over the desired grid passing through the available source data points by applying the minimum energy principle. The method was implemented using the “tpaps” routine of the Spline-toolbox of the MATLAB software version 7.5 (R2007b). The routine takes the source data points (x,y,z), the x and y coordinates of the grid points of the new DEM and the relaxation parameter p as inputs. The relaxation parameter p determines how strictly the approximated surface follows the source data points and its value is between 0 and 1. If p=1 the surface passes exactly through the z values of the source data points, while in the case p=0 linear interpolation by minimizing the sum of squared errors between the source data point values and the approximated surface is performed. The x and y coordinates of the available source data points have to be in the following equation:

$$X = \begin{bmatrix} x_1 & x_2 & \dots & x_n \\ y_1 & y_2 & \dots & y_n \end{bmatrix}$$

### QR- decomposition

The first step in the algorithm is to find the weights of the source data points. This step involves many matrices which are more feasible if the matrix containing the x and y coordinates of the source data points is first decomposed.

Given an  $m$  by  $n$  matrix ( $m$  rows and  $n$  columns),  $A$ , the QR-decomposition, can be expressed as:

$$A = QR, \quad Q \in \mathbb{R}^{m \times n}, \quad R \in \mathbb{R}^{m \times n}$$

where Q is an orthogonal matrix and R is an upper triangular matrix. For an orthogonal matrix the following property holds:

$$Q^T Q = Q Q^T = I$$

In an upper triangular matrix all the elements below the main diagonal are equal to zero:

$$R = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ 0 & r_{22} & \dots & r_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & r_{nn} \end{bmatrix}$$

*Steps of the thin plate spline interpolation procedure*

The first step is the QR-decomposition of a matrix containing the x and y coordinates of the source data points plus an additional column of ones. The matrix is then modified and denoted by X<sub>1</sub>:

$$X_1 = \begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ \vdots & \vdots & \vdots \\ x_n & y_n & 1 \end{bmatrix}$$

The resulting matrix Q can be written as:

$$Q = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nn} \end{bmatrix}$$

Matrix Q can then divided into two parts, Q<sub>1</sub> and Q<sub>2</sub>, so that Q<sub>1</sub> contains the first three columns of Q and Q<sub>2</sub> contains the remaining columns. Matrices Q<sub>1</sub> and Q<sub>2</sub> can be expressed as:

$$Q_1 = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & a_{n3} \end{bmatrix}, Q_2 = \begin{bmatrix} a_{14} & \dots & a_{1n} \\ a_{24} & \dots & a_{2n} \\ \vdots & \vdots & \vdots \\ a_{n4} & \dots & a_{nn} \end{bmatrix}$$

Next, the collocation matrix is formed containing the x and y coordinates of the source data points. The matrix's purpose is to specify the relative location of the source data points with respect to each other. A matrix A<sub>1</sub>, where the coordinates of each source data point are repeated as many times as there are source data points, is formed first. It has the following form:

$$A_1 = \begin{bmatrix} x_1 & x_1 & \dots & x_1 & x_2 & x_2 & \dots & x_2 & \dots & x_n & x_n & \dots & x_n \\ y_1 & y_1 & \dots & y_1 & y_2 & y_2 & \dots & y_2 & \dots & y_n & y_n & \dots & y_n \end{bmatrix}$$

$A_2$  is formed with the same amount of columns as  $A_1$  but ordered differently,

$$A_2 = \begin{bmatrix} x_1 & x_2 & \dots & x_n & x_1 & x_2 & \dots & x_n & \dots & x_1 & x_2 & \dots & x_n \\ y_1 & y_2 & \dots & y_n & y_1 & y_2 & \dots & y_n & \dots & y_1 & y_2 & \dots & y_n \end{bmatrix}$$

The following operations is performed between  $A_1$  and  $A_2$ : and the following matrix can be created:

$$A = (A_1 - A_2)^2$$

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1s} \\ a_{21} & a_{22} & \dots & a_{2s} \end{bmatrix}$$

The rows of matrix A are now considered into a vector B in the following form:

$$B = [a_{11} + a_{21} \quad a_{12} + a_{22} \quad \dots \quad a_{1s} + a_{2s}]$$

The vector B contains zero corresponding to the cases when the distance of a source data point to itself is considered. The zeros are then changed to ones and the following matrix is calculated:

$$B_1 = [b_1 \ln(b_1) \quad b_2 \ln(b_2) \quad \dots \quad b_s \ln(b_s)]$$

The final collection matrix C is obtained by rearranging the elements of vector  $B_1$  so distances from a certain source data are in the same row. The size of C is  $n$  by  $n$  given the total number of source data points is  $n$ .

The relaxation parameter was considered next. The following was performed and the diagonal of C is filled with the result:

$$p \leftarrow \frac{1-p}{p}$$

Once operations are completed, the weights are evaluated. Among the weights are K1 and K2 and are defined as:

$$K_1 = \left( \frac{ZQ_2}{Q_2^T C Q_2} \right) Q_2^T \quad \text{and} \quad K_2 = \frac{(Z - K_1 C) Q_1}{R_1^T} ,$$

where  $Z$  contains the values of the  $z$ - coordinate of the source data points,  $C$  is the collocation matrix,  $Q_2$  is from the QR-decomposition, and  $R_1$  is a modified version of the triangular matrix  $R$  from QR- decomposition.  $R_1$  can be obtained by removing the rows with zeros from  $R$ .

Given that the number of points is  $k$ , the  $x$  and  $y$  coordinates are placed in matrix  $X_e$ :

$$X_e = \begin{bmatrix} x_1 & x_2 & \dots & x_k \\ y_1 & y_2 & \dots & y_k \end{bmatrix}$$

Another matrix is formed based on the distances between the points to be approximated and the source data points, and repeated as many times as there are points to be approximated. The matrices are subtracted from each other and another matrix with the differences of  $x$  and  $y$  coordinates of the source data points and the points to be approximated in the upper and lower row, respectively. It is then squared and summed to obtain a vector of squared distances between the source data points and the new grid points. The resulting matrix is then multiplied by the natural logarithms and arranged into a matrix with the number of rows equal to the number of source data points. The result is the following:

$$C_e = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1k} \\ w_{21} & w_{22} & \dots & w_{2k} \\ \vdots & \vdots & \dots & \vdots \\ w_{n1} & w_{n2} & \dots & w_{nk} \\ x_1 & x_2 & \dots & x_k \\ y_1 & y_2 & \dots & y_k \\ 1 & 1 & \dots & 1 \end{bmatrix}$$

Lastly,  $K$  and  $C_e$  is multiplied giving the  $z$  coordinates of the points of the new grid.

$$Z_{est} = KC_e$$

$$Z_{est} = \begin{bmatrix} z_{est,1} & z_{est,2} & \dots & z_{est,k} \end{bmatrix}$$



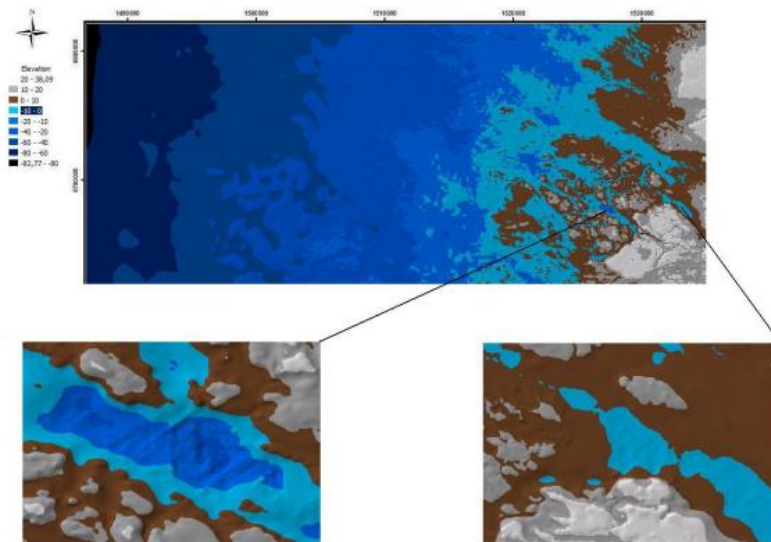
**Development**

**Table 11. Results of comparison between interpolation methods**

Interpolation method	95 % confidence interval of the error (lower limit; upper limit; range) in meters	Number of errors (percentage), exceeding the -2,0 ... 2,0 meter bounds (lower limit; upper limit)
Linear interpolation	-1.41; 1.36; 2.77	16 (0.81%); 15 (0.76%)
Parabola fitting	-3.45; 2.15; 5.60	165 (8.37%); 56 (2.84%)
Cubic interpolation	-0.84; 1.84; 2.68	33 (1.67%); 39 (1.98%)
IDW 1	-2.80; 1.96; 4.76	109 (5.53%); 41 (2.08%)
IDW 2	-2.56; 1.65; 4.21	87 (4.41%); 29 (1.47%)
IDW 3	-2.56; 1.57; 4.13	95 (4.82%); 28 (1.42%)
Averaging	-3.36; 2.36; 5.72	154 (7.81%); 62 (3.14%)
Median	-2.76; 2.56; 5.32	258 (13.08%); 150 (7.61%)
Thin plate spline	-1.53; 1.24; 2.77	21 (1.06%); 18 (0.91%)



**Figure 28. Final DEM (10x10 meters) in raster format**



**Figure 29. Final DEM (10x10 meters) in TIN format**

## CREATING LOW-COST HIGH-RESOLUTION DIGITAL ELEVATION MODELS

*Louhaichi, M., Borman, M. M., Johnson, A. L. and Johnson, D. E. (2003). Creating Low-Cost High-Resolution Digital Elevation Models. Journal of Range Management, 56:1, 92-96.*

This research required a digital elevation model with vertical accuracy (root mean square error) of approximately 10 cm and neither a real time stop and go nor real time kinematic carrier-phase differential GPS was available. Therefore, a method was developed to efficiently generate a high accuracy elevation model for open, relatively level land using a course differential GPS (<\$8000) and a low cost laser level (<\$1000).

### Material and Methods

The procedure developed included both data collection in the field and computer processing.

**Table 12. Field and office computer procedures for developing a DEM**

Location Procedures	
Field	<p><u>Step 1: Identify a reference position</u></p> <ol style="list-style-type: none"> <li>Locate a National Geodetic Survey Ground Control Point, or</li> <li>Create a Temporary Bench Mark                             <ul style="list-style-type: none"> <li>- Phase processing RTSG<sup>1</sup> DGPS<sup>1</sup>, or</li> <li>- Occupy static location for an extended time</li> </ul> </li> </ol> <p><u>Step 2: Establish the Laser Level in the Area to be Surveyed</u></p> <ol style="list-style-type: none"> <li>Determine the height of the laser beam above GCP<sup>1</sup> using leveling rod and receiver</li> <li>Record the X Y<sup>1</sup> locations of the laser lever and laser height identification code (LHID)<sup>1</sup> using the GPS<sup>1</sup> data Logger</li> <li>Enter height of the laser above the GCP<sup>1</sup> or TBM<sup>1</sup> as a comment in the GPS<sup>1</sup> data logger</li> </ol> <p><u>Step 3: Start Grid Sampling</u></p> <ol style="list-style-type: none"> <li>Record location X Y<sup>2</sup> using GPS<sup>1</sup> unit and data logger</li> <li>Record height of the laser beam above the ground as a comment</li> </ol>
Office	<p><u>Step 4: Create Spreadsheet with XYZ<sup>1</sup> Locations</u></p> <ol style="list-style-type: none"> <li>Differentially correct GPS<sup>1</sup> data to give XY<sup>2</sup> locations within 1m</li> <li>Subtract the height above the ground (comment) from the laser beam height to obtain the Z<sup>1</sup>-value</li> </ol> <p><u>Step 5: GIS<sup>1</sup> Applications</u></p> <ol style="list-style-type: none"> <li>Import data into GIS<sup>2</sup> software</li> <li>Interpolate between points using a weighted-averaging, spline, or kriging technique, or a triangulated irregular network (TIN)<sup>1</sup> model</li> </ol> <p><u>Step 6: Produce Digital Elevation Models</u></p> <ol style="list-style-type: none"> <li>Contour maps</li> <li>3-D<sup>1</sup> maps</li> </ol>

<sup>1</sup> Abbreviations used are DGPS (differential global positioning system); GCP (Ground Control Point); GIS (Geographical Information System); GPS (Global Positioning System); LHID (Laser Height Identification); NGS (National Geodetic Survey); RTSG (real-time stop-and-go); TBM (temporary bench mark); TIN (Triangulated Irregular Network); XYZ are longitude, latitude, and elevation.

### Field Setup

Topographic information was collected with two technologies: Coarse-acquisition (C/A) code differential global positioning system for the latitude (Y) and longitude (X) position and laser level with metric leveling rod for elevation (Z), and a Trimble Navigation Pathfinder PRO XR, 12 channel, L1/CA-code differential GPS. With this, X, Y coordinates can be obtained with a root mean square error of 50 to 100 cm in approximately 25 seconds. The laser level system was a Laser Reference Inc. Proshot L4 with a R4 laser receiver and a Crain Enterprise, Inc. CR-5.0M metric leveling rod. The laser level consisted of a working radius of 230 m, with a leveling accuracy of better than 2 cm.

If it is important to determine the true elevation of the laser for reasons other than creating a relative DEM, it is necessary to either find an already existing National Geodetic Survey (NGS) ground control point or establish a benchmark. Locations of the ground control points can be obtained from NGS on CD-ROM or at: <http://www.ngs.noaa.gov/datasheet.html>. In most cases, a temporary benchmark will have to be established and a GPS used to mark its location. In this study, the temporary benchmark was positioned by collecting numerous coarse-acquisition code differential GPS fixes with the GPS antenna in a static position on a tripod at a set height above the point. The coarse acquisition code differential GPS records a position every second and the people of this study normally recorded data for an hour or more to get an accuracy of 60 m. To define absolute elevations of the area, the temporary benchmark is positioned using either traditional survey techniques or carrier-phase differential GPS. The latter requires 2 GPS units, a rover, and a local base station. 10 cm accuracies can be obtained with carrier-phase differential GPS processing by occupying a location for 30 minutes.

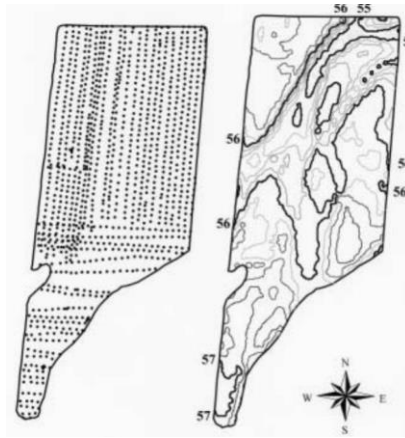
Once positioned, the laser level (source) is set in the area to be surveyed and the height of the laser above the reference point is measured with a leveling rod and laser receiver. The X and Y location of the laser level was recorded and 2 additional pieces of information as components: (1) the laser height identification code and (2) the elevation of the laser beam. The elevation of the rotating laser is determined by measuring its height above the benchmark with a laser receiver attached to a leveling rod and then adding the elevation.

### **Grid Sampling**

Once the laser has been set up on a ground control position, it can rotate 360 degrees and the two person crew moves across the area to be surveyed, stopping to record differential GPS points across the area. At each location, a person must record an X and Y coordinate, while another person adjusts the height of the electronic laser receiver on the calibrated leveling rod to capture the rotating laser beam and measures the height of the laser beam above the ground. This height is then entered into the differential GPS data logger as a comment. It is important to sample other features of interest like hilltops and low spots. To make sure that areas are sampled, each point was marked with foam that lasts until the job is finished. The quality of a DEM will be a function of: (1) how many points are obtained and (2) how the points are positioned. Because it rotates 360 degrees, more than one team can collect data at the same time.

### **Office Processing**

The GPS data was differentially corrected with data from a local base station. Data is then exported to a spreadsheet with each sheet containing all data collected while the laser was at one location and height. Ground elevation is calculated for each sample by subtracting the laser beam to ground distance from the elevation of the laser beam that was determined from the temporary benchmark. Interpolation of a DEM from point data is accomplished with any GIS software packages.



**Figure 30. Sample locations on the left image. A contour map created by interpolation between sample locations is shown on the right.**

### **Limitations**

This technique was developed for measuring land with gently rolling topography, such as found in agronomic fields, wetlands, and research plots. Because the X, Y coordinates have a root mean square error of 50 to 100 cm, it is not appropriate for short, deep cut banks or short, steep escarpments where the cut face must be precisely positioned. Because the leveling rod and receiver unit have a working height from near ground level to 5 m, rugged land can require frequent repositioning of the laser level, which in turn, reduces efficiency and increases cost. Shrubs, trees, rain, fog, and/or excessive dust can limit the technique used in this study.

### **DEVELOPMENT OF A HIGH-RESOLUTION DIGITAL ELEVATION MODEL FOR NEW ZEALAND. LANDCARE RESEARCH CONTRACT REPORT**

*Barringer, J. R. F., Pairman, D., & McNeill, S. J. (2002). Development of a high-resolution digital elevation model for New Zealand. Landcare Research Contract Report: LC0102/170.*

Landcare Research was tasked to develop a high resolution digital elevation model of New Zealand from the national TOPOBASE data that was supplied by the Land Information New Zealand (LINZ). The purpose of this report was to describe the methods being used to develop a national DEM with the use of GIS and LIDAR data. The DEM was created using an interpolation algorithm where data was interpolating from up to four contours or spot heights instead of two, assigning the pixel height based on interpolation within the pixel, tracking distances from the nearest contours using floating point precision for elevations. The developed DEM was then tested against a high resolution reference DEM and showed a consistent bias which overestimated elevation by a mean of six meters.

Landcare Research was able to develop a DEM with the use of interpolation methods and previously developed DEMs but continue to experience spatial inaccuracies when modeling river

valleys or low elevation landforms. The paper notes that interpolation methods vary depending on the focus of the study. Different methods are available to create a DEM such as bilinear interpolation, various spline functions, radial basis functions, etc.). Although it is helpful to use GIS to develop a high resolution, DEMs generated by other organizations which use different software should be referenced as well in order to compare the accuracy of the DEM being created. It is also vital to detail the interpolation methods used in order for other users to be aware of what was used to create the DEM.

### **Interpolation method for Landcare Research 25-m DEM**

The interpolation method was based on the work by David Giltrap for use in a VAX computer system. It was based on the idea that for any regular and complete set of contours, any region will be bounded by contours of at most two different levels. This method makes use of this contour topology by using a neighborhood expansion process to grow into the regions bounded by contours while keeping track of the minimum distance to two different contour levels. Once completed, each pixel within the region will have the minimum distances to the bounding contours at two different levels. The height assigned to the point is a ratio of these two levels in inverse proportion to the minimum distance found. This ensures that both bounding contour levels affect the whole region bounded, even if no line can be drawn from a point to one of the bounding contour levels without crossing the other contour. What was discovered is that mid-slope spot heights and non-regular contours break the topology and because of this they introduced a third height adjacent to the area being interpolated. As a result, artefacts in the surface could be found where the set of two nearest heights changed. To overcome this problem, the procedure was modified to keep track of up to four contours or spot heights adjacent to an area being interpolated.

A contour was considered to have passed through the pixel if it entered the diamond connecting the midpoints of the pixel's four sides, or if there was a node within the pixel. If more than one contour passed through the pixel, then the level assigned was that of the contour passing closest to the pixel's center.

By assigning pixels to contour values where contours met the criteria for passing through the pixel described above, in steeper areas the DEM would have a high proportion of pixels that had elevations that had been assigned but not interpolated; meaning in version 1, there was a tendency to assign pixels of multiples of 20 m or 10 m when contours were dense with respect to the posting distance. In version 2, a new algorithm kept track of the contours running closest to each of the pixel's four edges and assigned the pixel height based on interpolation within the pixel. A suite of space borne radar altimeter data sets over much of the interior ice sheet where surface slopes are low and the absolute accuracy of the radar altimeter is high. Airborne radar data was also used in the development of the DEM to fill in gaps and enhance accuracy. Automation was eased with a GIS based toolkit that was developed by using ARC/INFO analytical and graphical functions and C programming language.

## **DEVELOPMENT OF AN ANTARCTIC DIGITAL ELEVATION MODEL BY INTEGRATING CARTOGRAPHIC AND REMOTELY SENSED DATA: A GEOGRAPHIC INFORMATION SYSTEM BASED APPROACH**

*Liu, H., Jezek, K. C., & Li, B. (1999). Development of an Antarctic digital elevation model by integrating cartographic and remotely sensed data: A geographic information system based approach. Journal of Geophysical Research: Solid Earth, 104(B10), 23199-23213.*

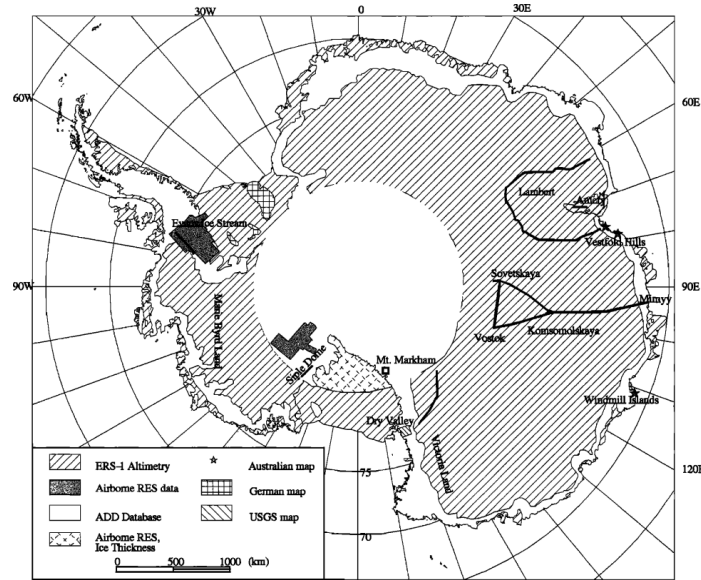
This paper describes the development of a DEM of the Antarctic with a spatial resolution ranging from 200 m to 5 km, generated using a wide selection of topographic data sets. It relied on the availability of the most detailed cartographic data in the Antarctic Digital Database (ADD) and large-scale topographic maps from the USGS and Australian Antarctic Division for mountainous and steeply sloping areas.

### **Data Compilation and Selection**

Scientists at the Byrd Polar Research Center created a high resolution digital elevation model of the Antarctic developed through the use of data that falls into three categories: (1) cartographic data, (2) remotely sensed (RS) data, and (3) survey data. Cartographic maps with the most detailed data including contours, spot heights, and surface structure lines were chosen. Sources included the Antarctic Digital Database (ADD), United States Geological Survey (USGS), Australian Antarctic Division, and a digitized map from Institut fur Angewandte Geodasie of Germany. Remotely sensed data consisted of satellite radar altimetry data and radar echo sounding data. RS data sources included European Remote Sensing (ERS) satellites, SCP airborne radar sounding data, BAS airborne radar echo sounding data, and Ross Ice Shelf Geophysical and Glaciological Survey (RIGGS). Survey data included ground-based survey data and satellite based Global Positioning System (GPS) measurements. Data was selected if it satisfied the following criteria: (1) used GPS data, airborne radar data, and large-scale topographic maps wherever available, (2) used satellite radar altimeter data if slope was less than 0.8 degrees, (3) used the ADD cartographic data for rugged and highly sloped areas, and (4) used satellite radar altimeter data for areas with surface slopes between 0.8 and 1.0 degree (Liu, H., et.al., 1999).

The use of detailed and accurate information from a combination of cartographic data, remotely sensed data, and survey data, enabled the Byrd Polar Research Center to create a DEM using interpolation and merging methods to create an accurate representation of the region. Further improvements could be made using more data sources including digitized topographic maps and radar sensing data.





**Figure 31. Distribution of input data sources used in the final DEM. Dark black lines indicate locations of GPS, RES and geodetic leveling traverses for accuracy assessment.**

## DEM Generation Approaches

### *Spatial Interpolation Algorithms*

Evaluation and choice of interpolation methods are based on three considerations. First, the interpolated surface should agree with the source data well in terms of values and pattern. For contour data, the interpolated grid DEMs were converted back to contour lines using half of the original contour level and then overlaid the derived contour lines on top of the original contour lines to check the consistency. Second, the interpolated surface had to be single-valued, continuous, and smooth at all positions, practically, the visual plausibility of the resulting terrain surface using analytical hill-shading or the simulated stereo images. Third, the computation involved in interpolation must be efficient and fast in view of the size of the Antarctic continent and the correspondingly large amount of data.

### *Interpolation of satellite radar altimeter data*

The satellite radar altimeter data that was obtained already was preprocessed into evenly distributed points with 5 km spacing. The Triangulated Irregular Network (TIN) interpolation method was chosen. It starts with the Delaunay triangulation of satellite altimeter data points to form triangle patches, and then elevation values at nodes of a grid are estimated by evaluating the fitted functions:

$$Z(x, y) = \sum_{j=0}^5 \sum_{k=0}^{5-j} q_{jk} x^j y^k$$

where  $Z(x,y)$  is the interpolated value at the location  $(x,y)$  and  $q_{jk}$  are the fitted coefficients. The result is a smooth and realistic surface due to the fact that the second derivatives of the surface from TIN quantic interpolation are continuous and differentiable.

#### *Interpolation of traverse airborne radar data*

The airborne radar data are anisotropically distributed, namely, densely sampled, along flight lines but widely separated between flight transects. This caused difficulties on the general interpolation methods and so a procedure was developed that combined quadrant neighborhood-based Inverse Distance Weighting (IDW) and TIN methods and achieved a reasonable result. First, airborne radar was filtered and reduced. A super block based 2D searching algorithm was used to partition the data points into an array of square blocks with a width of one fifth of the average distance between flight lines. For each block, only the median point was selected for interpolation. On the basis of the retained data points, a coarse grid with a spacing of half the average distance between flight lines was then interpolated by the IDW algorithm. It uses the 12 closest equally selected from each of four quadrants to determine the elevation value:

$$Z_{t,j} = \frac{\sum_{p=1}^{p=12} Z_p d^{-n}_p}{\sum_{p=1}^{p=12} d^{-n}_p}$$

where  $Z_{i,j}$  is the computed elevation at the node  $(i,j)$  of a grid,  $Z_p$  is the elevation at point  $p$  in the neighborhood,  $d$  is the distance from the node  $(i,j)$  to point  $p$ , and  $n$  is the power factor of distance. The inverse squared distance weighting is used ( $n=2$ ). Finally, a TIN quantic model was constructed using the reduced radar altimeter data together with the IDW derived coarse grid points and interpolated them with a fine DEM grid.

### **METHODS FOR EXTRACTING SIGNIFICANT TOPOGRAPHICAL FEATURES (SINK HOLES) FROM DEMS**

#### **SINKHOLE SUSCEPTIBILITY HAZARD ZONES USING GIS AND ANALYTICAL HIERARCHICAL PROCESS A CASE STUDY OF KUALA LUMPUR AND AMPANG JAYA**

*Rosdi, M. A. H. M., Othman, A. N., Abdul, M. A. M. Z. Z., & Yusoff, Z. M. (2017). Sinkhole Susceptibility Hazard Zones Using GIS and Analytical Hierarchical Process (AHP): A Case Study of Kuala Lumpur and Ampang Jaya. International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences, 42.*

In this study, based in Kuala Lumpur and Ampang Jaya, there has been an increase in sinkhole incidents that have posed a serious threat to human lives and infrastructure. A Sinkhole Hazard Model (SHM) was generated with integration of a GIS framework by applying an Analytical Hierarchical Process (AHP) technique in order to produce a sinkhole susceptibility hazard map

for the area. Five consecutive parameters for main criteria each categorized by five sub classes were selected for this research: (1) Lithology (LT), (2) Groundwater Level Decline (WLD), (3) Soil Type (ST), (4) Land Use (LU), and (5) Proximity to Groundwater Wells (PG). A set of relative weights were assigned to each inducing factor and computed through a pairwise comparison matrix derived from expert judgement. Lithology and Groundwater Level Decline were identified to give the highest impact to sinkhole development.

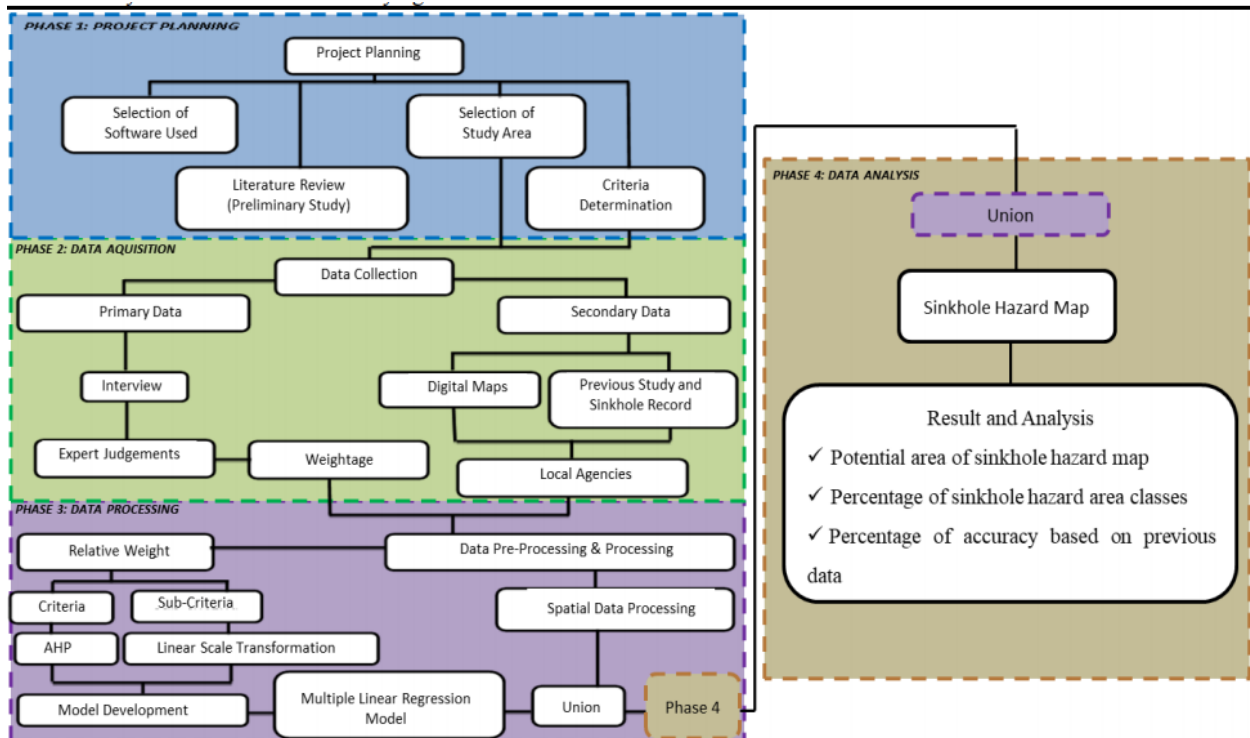


Figure 32. Outline of work flow

The first phase consisted of determining the problem statement and significance of research within the DBKL and MPAJ's area. Next, the data acquisition phase consisted of classifying primary and secondary data. Primary data consisted of interviews with experts and literature reviews. In the third phase, data processing which involves weightage and software analysis determination is carrying out in order to achieve the research objectives/ Finally, the sinkhole hazard model is generated and used to map the susceptible location for sinkhole hazards.

Five parameters were used to identify sinkhole development in Malaysia: (1) bedrock lithology, (2) soil type, (3) water table decline, (4) proximity to groundwater, and (5) land use.

- Digital lithology and soil map was acquired from Mineral and Geoscience Department and was considered as the main source.

## Literature Review of Methodologies for Development of a High Resolution DEM for WIPP

- The topography of the region was obtained using digital topographic map which was obtained from Department of Survey and Mapping Malaysia (JUPEM) and was extracted for land use information.
- Groundwater level was obtained from Mineral and Geoscience Department which includes groundwater level locations.
- For data validation, a sinkhole inventory by the Mineral and Geoscience Department was used.

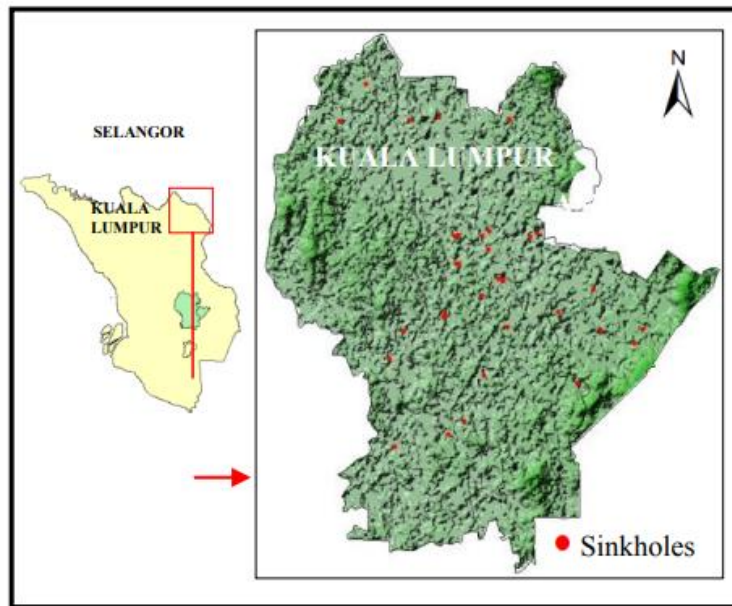


Figure 33. Study area in Kuala Lumpur and Ampang Jaya highlighted

Table 13. Criteria selected to carry out AHP method

Criteria	Unit
Lithology	Type
Soil properties	Type
Groundwater level decline	Meter Cubed (m <sup>3</sup> )
Land Use	Type
Proximity to groundwater well	Meter

According to the study, ranking method, rating method, pairwise comparison method and AHP method are the best suitable method that can be used to study sinkhole formation.

### Data Pre-Processing and Processing

All spatial and attribute data were process throughput map digitizing, editing, and conversion using ArcGIS 10.1. The list of attributes weight of criteria and sub criteria is entered in the

spatial data to classify the values. AHP and linear scale transformation techniques were used in this study to determine the value of relative weight for criteria and sub criteria. The values of criteria were determined from interviews and discussions with geology experts. The result of the weight is used to generate multiple linear regression models in order to produce sinkhole susceptibility hazard maps.

Model development has been preliminary assessed considering relative weights assigned to five selected criteria and to different sub-criteria. A set of criteria was weighted by performing pairwise comparison matrices, referred to in the following table:

**Table 14. Pairwise comparison matrix of criteria chosen**

<b>Criterion</b>	<b>LT</b>	<b>ST</b>	<b>LU</b>	<b>WLD</b>	<b>PG</b>
Lithology (LT)	1	5	7	2	6
Soil Type (ST)	0.200	1	3	0.167	2
Land Use (LU)	0.143	0.333	1	0.200	0.333
Water Level Decline (WLD)	0.500	6	5	1	3
Proximity to Groundwater (PG)	0.167	0.500	3	0.333	1

An important step of AHP is to evaluate the consistency by using the consistency index (CI):

$$CI = \frac{\lambda - n}{n - 1}$$

where  $\lambda$  is the average value of consistency vector and n is the number of parameters. The consistency ratio (CR) is used and uses RI (Random Consistency Index) and is defined below:

$$CR = \frac{CI}{RI}$$

Consistency ratios higher than 0.1 suggest untrustworthy judgements.

Multiple linear regression models were used to generate a series of maps of potential sinkhole hazard areas and the Sinkhole Hazard Model (SHM) and consisted of the following criteria represented below:

SHM = (0.457 x sc\_litho) + (0.109 x sc\_soil) + (0.046 x sc\_lu) + (0.299 x sc\_wld) + (0.090 x sc\_pg), where:

- sc\_litho is standardized score for lithology sub criterion

## Literature Review of Methodologies for Development of a High Resolution DEM for WIPP

- sc\_soil is standardized score for soil type sub criterion
- sc\_lu is standardized score for land use sub criterion
- sc\_wld is standardized score for water level decline sub criterion
- sc\_pg is standardized score for proximity to groundwater sub criterion

The AHP technique was used to analyze complex decision problems taking into account a large number of factors or criteria. The pairwise comparison matrices in the table above were constructed to determine the relative importance of each parameter for sinkhole development with respect to the other. When calculating the weight, every scale factor in each criteria has to be converted to a fraction in order to get the total column value for each cell. The total scale factor is computed with the formula:

$$\Sigma = C^1 + C^2 + C^3 + \dots + C^8$$

where  $\Sigma$  is the total value of every columns variable and C is column variables. The normalized value is obtained with:

$$N = C / \Sigma C$$

where N is Normalize Matrix, C is Criteria Comparison Matrix, and  $\Sigma C$  is total value of every columns variable. The weight (W) was derived by averaging each normalized matrix by the sum of elements in the row, likewise for the other criteria. It was found that based on the weight determination, lithology has the largest weight and the smallest is land use.



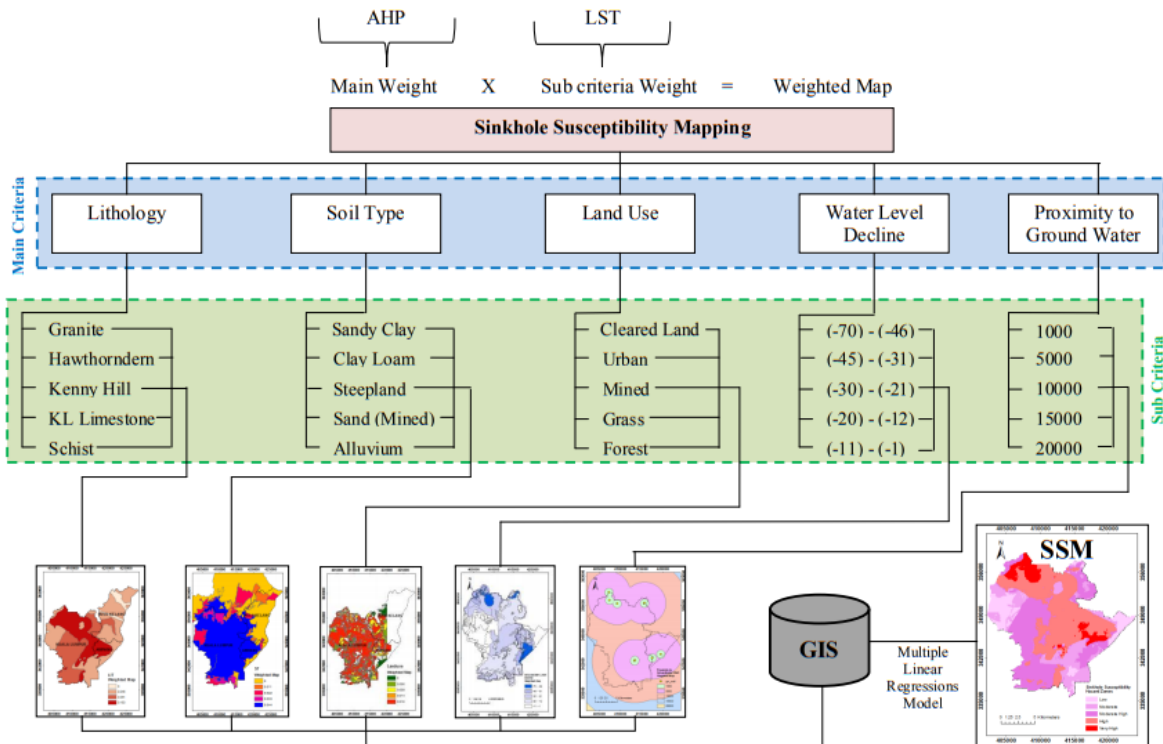


Figure 34. Flow diagram to construct a sinkhole susceptibility map

## AUTOMATIC DELINEATION OF KARST SINKHOLES FROM LIDAR-DERIVED DIGITAL ELEVATION MODELS

Wu, Q., Deng, C., & Chen, Z. (2016). Automated delineation of karst sinkholes from LiDAR-derived digital elevation models. *Geomorphology*, 266, 1-10.

In this study, a localized contour tree method for automated extraction of sinkholes was presented. The proposed method is said to have improved the sinkhole inventory by detecting non-inventoried sinkholes, identifying already inventoried sinkholes, delineating sinkhole boundaries, and characterizing sinkhole morphometric properties. Three times as many sinkholes were identified compared to the already identified sinkholes.

According to this study, previous methods for identification depended on visual interpretation of low resolution topographic maps provided by USGS, aerial images, and on the field research. Similar studies have used LiDAR data to identify sinkholes as well but with different processing methods, such as applying a sink filling method on LiDAR data to create a depressionless DEM and then subtracting the original DEM from the modified DEM to locate depressions. This layer was then processed with different thresholds of elevation differences to locate sinkholes. Other methods included image filtering techniques with kernel windows using focal functions and the “active” contour approach, which delineates sinkhole boundaries based on changes in the flow of

the elevation gradient in the surrounding region around the local minima at potential sinkhole locations.

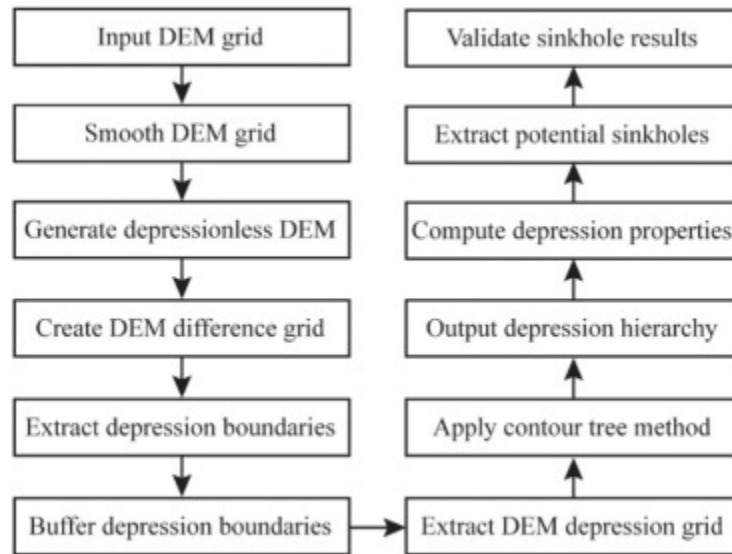


Figure 35. Flowchart of methodology

Once the smoothed DEM was created, they generated vector contours by setting the base contour to be 39 m and the contour interval to be 0.5 m, which is slightly higher than the vertical accuracy of the LiDAR DEM. The contour tree method was then applied to the contours to identify depressions. A minimum depression area of 100 m<sup>2</sup> and minimum depression of 0.5 m was used in order to identify sinkholes larger than 100 m<sup>2</sup> and deeper than 0.5 m.

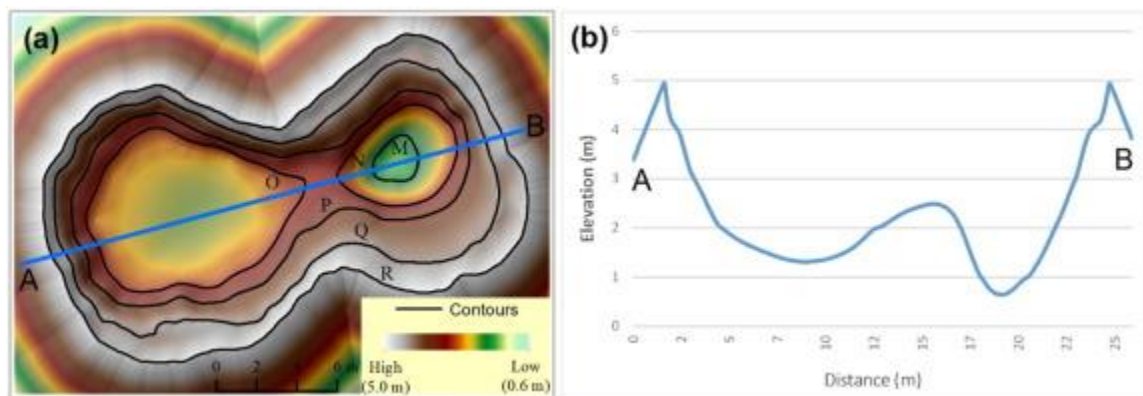
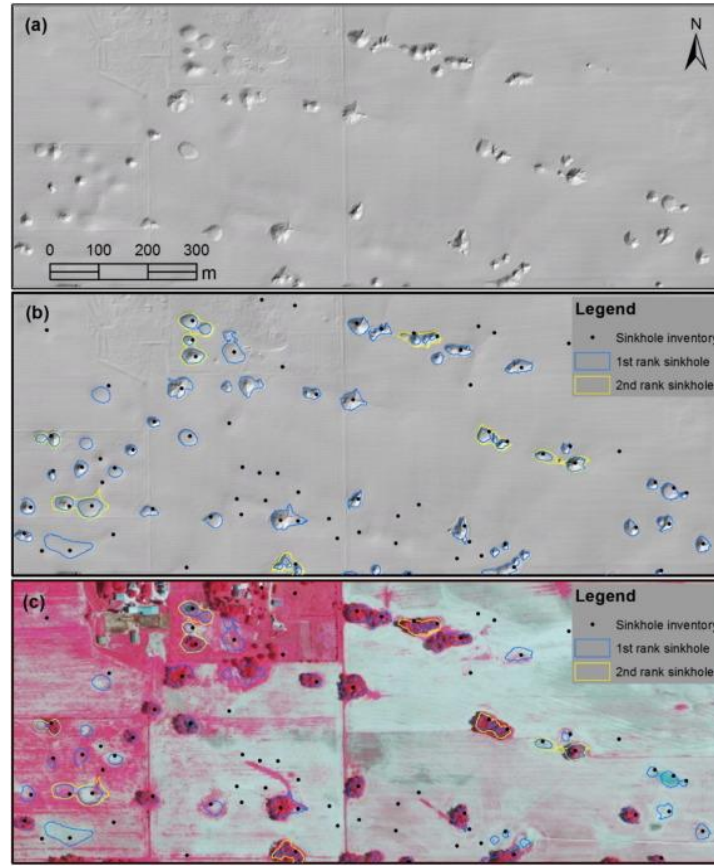


Figure 36. (a) Contours overlain on DEM shaded relief. (b) Elevation profile of the transect A–B shown in (a).



**Figure 37. LiDAR DEM shaded relief (a) and examples of extracted sinkhole boundaries overlain on LiDAR DEM shaded relief (b) and color infrared aerial imagery (c).**

According to the study, an ArcGIS toolbox, Sinkhole Extraction Analyst tool was created and will be freely available for the public to download free of charge in the near future.

## **KARST FEATURES DETECTION AND MAPPING USING AIRPHOTOS, DSMs, AND GIS TECHNIQUES**

*Kakavas, M. P., Nikolakopoulos, K. G., & Zagana, E. (2015, October). Karst features detection and mapping using airphotos, DSMs and GIS techniques. In Earth Resources and Environmental Remote Sensing/GIS Applications VI (Vol. 9644, p. 96440Y). International Society for Optics and Photonics.*

In this study, thirty seven aerial photographs of the study area in Aitolokarnania Prefecture, Western Greece at 1/40,000 scale from the Hellenic Military Geographical Service were obtained and processed to create a mosaic. An orthophoto mosaic with a spatial resolution of 1 m and the respective DSM were derived. Four DEMs were used to identify limestone terrain, describe exokarst features, analyze karst depressions, and detect geological structures that are important to karst development. The first DEM used was the Aster GDEM that was developed by the Ministry of Economy, Trade, and Industry (METI) of Japan and NASA. A Shuttle Radar

Topographic Mission (SRTM) DEM was also used with a resolution of 90 m at the equator and also provided in mosaicked 5 deg x 5 deg tiles for easy download and use. The third DEM used was a DSM created for the Greek Cadastral from aerial imagery with a 5 m resolution and vertical accuracy of 2-3 m. The fourth DEM used was created from digitized contours from 1/50,000 topographic maps with a 20 m pixel size and a vertical accuracy of 10 m.

## **Data Processing**

Five steps were taken for the karst research. Thirty seven airphotos and four software products were used for this study (ERDAS IMAGINE, ArcGIS 10.1, Google Earth).

1. The first step consisted of creating an airphoto mosaic at 1:5,000. Using LEICA photogrammetry suite, thirty seven airphotos were added in a block file and were orthorectified. Ground control points (GCP) and a high accuracy Digital Surface Model (DSM) were used for the orthorectification of the air photos. The air photo mosaic from the Greek Cadastral with a spatial resolution of 0.25 m and its DSM was used as a base map for the orthorectification procedure. The total root-mean-square (rms) error was less than 0.33 pixels. Finally, a DSM was created using the oriented airphotos.
2. In the second step, a bibliographic review of karst features in the wider region of interest was focused on to get more orientated of the research in the in situ observation. Next, Google Earth was used to detect karst characteristics to recognize them in the field work and find the coordinates using GPS.
3. The third step consisted of detecting and mapping possible karst features in 3D view in stereopairs. This was done by using nineteen stereopairs in ERDAS Imaging Stereo Analyst program in order to understand how karst phenomena appear in 3D view. The stereopairs permitted the 3D representation of orthophoto and the fundamental understanding of the topography of the region in order to detect the low points and identify them as karstic or not karstic. To get to this determination, a shapefile was introduced using the points determined from the in situ observation. A new shapefile was then created and contained polygons of possible karst features. By doing this, 277 possible karst characteristics were digitized in stereopairs which lead to detection in the field.
4. The final step consisted of checking the 277 possible karst characteristics which were digitized in stereopairs with in situ observation. To check, the shapefile with the possible karst features was used in ArcGIS and converted to a KML file to add the data to Google Earth. Once introduced into Google Earth, it was easy to identify and relate to the field work. More than 30% of the karst characteristics were checked by the in situ observation. Next, maps were constructed in ArcGIS to represent the appearance of karst features in the study area.

## Literature Review of Methodologies for Development of a High Resolution DEM for WIPP

Initially, the detection and mapping of geologic structures important to karst development was conducted using aerial photo stereo pairs which led to the digitization of 277 karst features that were tested via in situ observation. Next, closed karst features were identified in ArcGIS through the “cut fill” command and the points where there was a reduction of elevation were then converted to polygons and considered representative of “possible” karst features. Finally, through the intersection of data, the “possible” karst features were compared to the digitized karst characteristics from aerial photos.

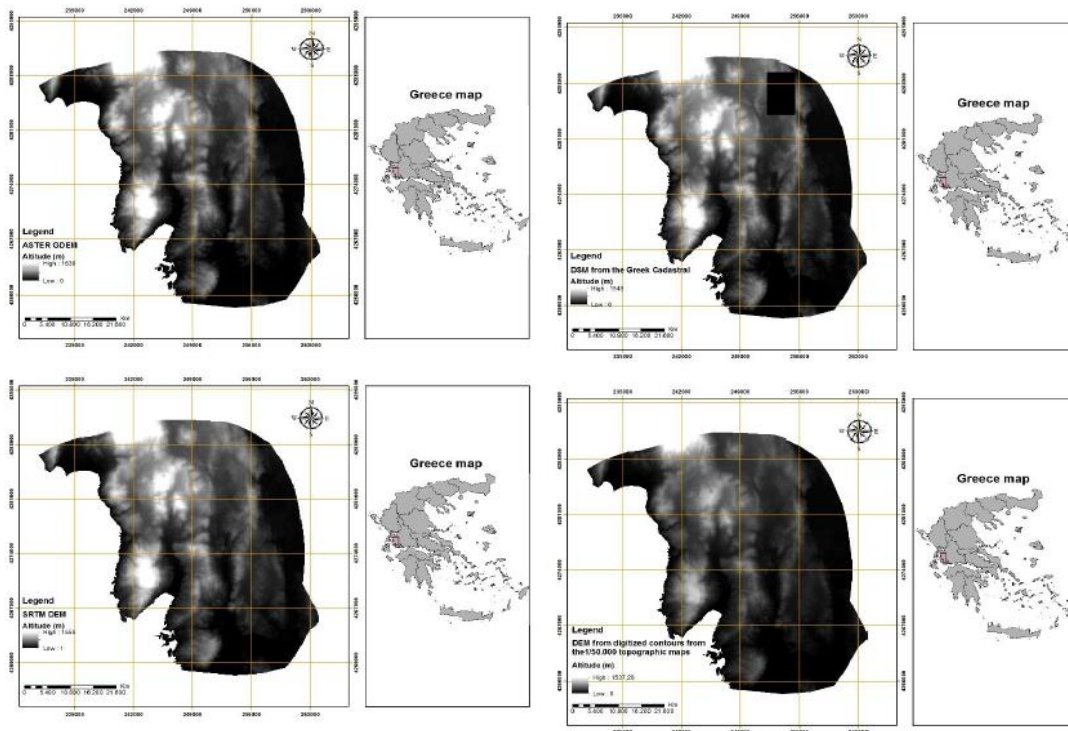


Figure 38. The four different DEMs used in this study



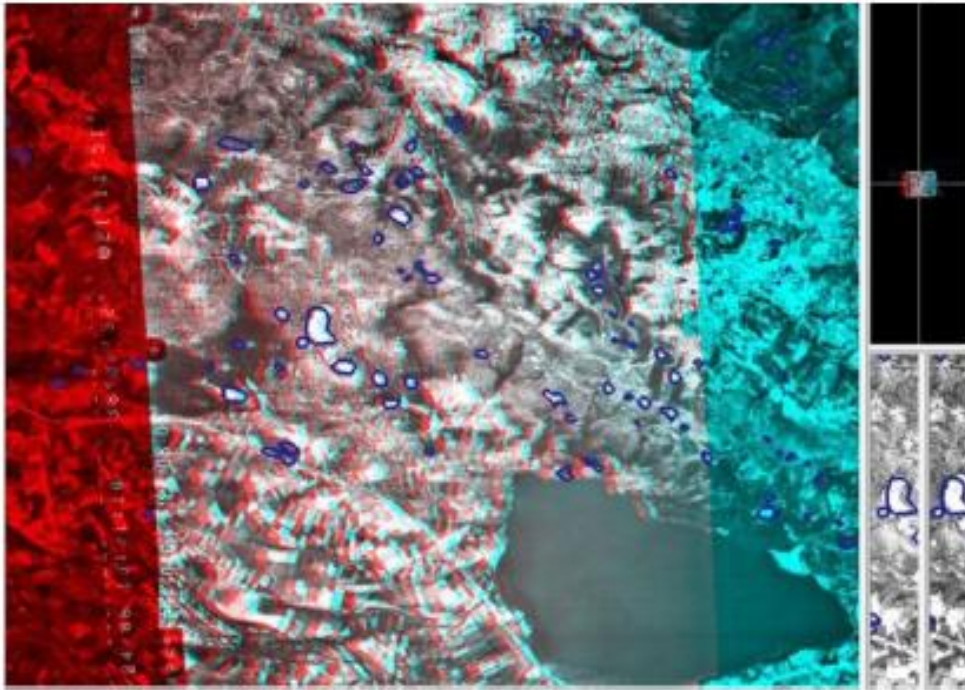


Figure 39. Sinkholes detected.

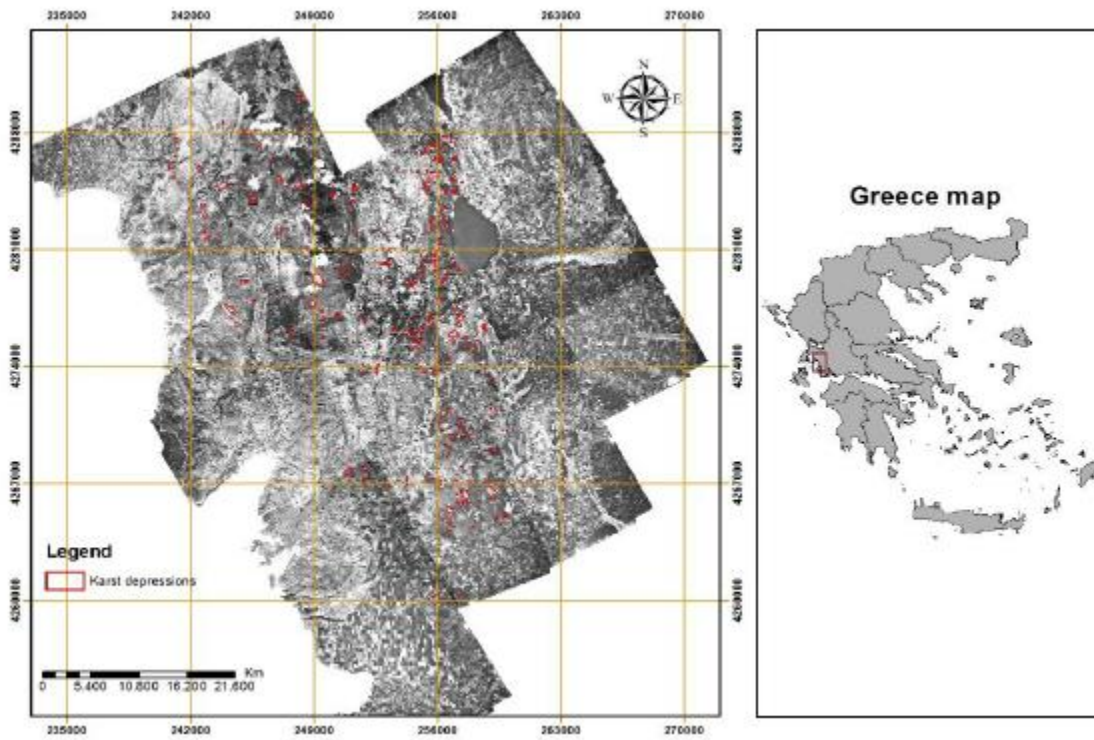


Figure 40. Sinkholes (red points) identified and displayed on the DEM



## **KARST DEPRESSION DETECTION USING ASTER, ALOS/PRISM AND SRTM-DERIVED DIGITAL ELEVATION MODELS IN THE BAMBUÍ GROUP, BRAZIL**

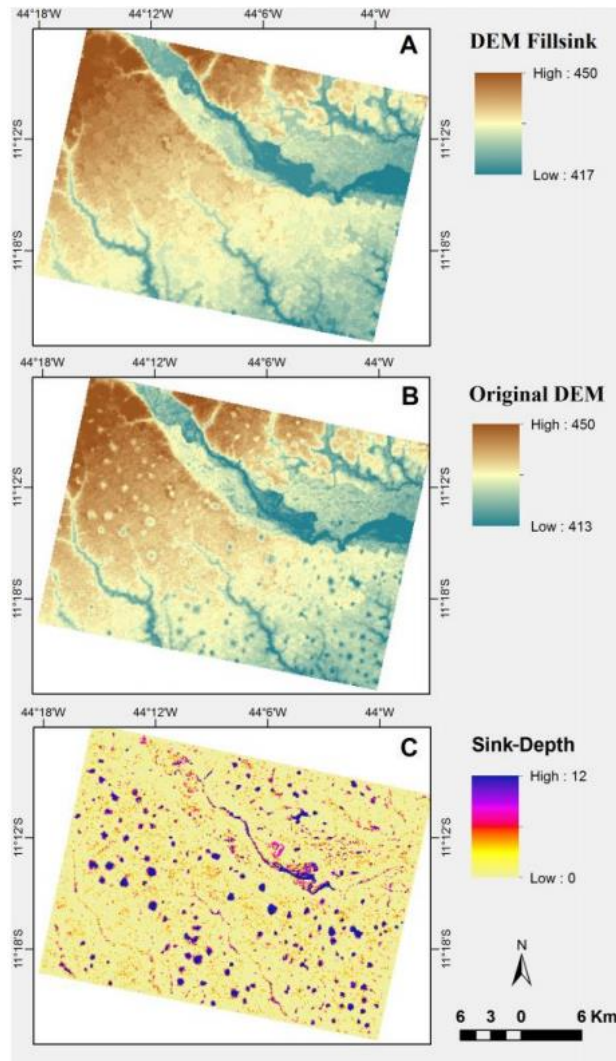
*de Carvalho, O., Guimarães, R., Montgomery, D., Gillespie, A., Trancoso Gomes, R., de Souza Martins, É., & Silva, N. (2014). Karst depression detection using ASTER, ALOS/PRISM and SRTM-derived digital elevation models in the Bambuí Group, Brazil. Remote Sensing, 6(1), 330-351.*

This study investigated the use of ASTER-, SRTM, and ALOS/PRISM derived DEMs to detect natural karst depressions along the Sao Francisco River near Barreiras city, northeast Brazil. Depressions were identified by the following steps: (1) acquiring DEM; (2) sink-depth calculation using the difference between the raw DEM and the corresponding DEM with sinks filled; and (3) eliminating falsely identified karst depressions using morphometric attributes. The DEMs used were the ASTER Global DEM from NASA and Japan's Ministry of Economy, Trade and Industry, the SRTM DEM made by the Consultative Group on International Agricultural Research Consortium for Spatial Information (CGIAR-CSI), and a DEM made for ALOS/PRISM.

1. ASTER-GDEM: Resolution of 1-arc second (30m) with UTM coordinate system and referenced to Earth's geoid using EGM96 geopotential model. This product is generated from a stereo pair of images using SilcAST software and covers 99 percent of Earth's landmass.
2. SRTM: Topographic maps were acquired from a single flight covering 80% of the Earth's surface. The flyover produced 3D models with a resolution of 1 arc-second (30m) and 3 arc-second (90 m) using WGS84 horizontal datum and vertical datum WGS84/EGM96. Vertical accuracy was 5m. It was continuous data acquisition which ensured homogeneous data throughout the globe.
3. The ALOS was created with PRISM, which acquires images with spatial resolution of 2.5 m. "It produces triplet images that achieve along-track stereoscopy by three independent cameras for viewing nadir, forward and backward where the images are acquired in the same orbit at almost the same time. The nadir-looking radiometer can provide coverage 70 km wide, and the forward-looking and backward-looking radiometers each provide coverage 35 km wide."

The methodology included using the Fill Sink tool in ArcMap, extracting the sink depths in the area and subtracting the maps between the sink filled "depressionless" DEM and the original DEM. The depressed areas were given a value of 1 and the remaining areas a value of 0. This binary image was then converted to vector format. The minimum area of depressions corresponded to the spatial resolution of the sensor; however, the polygons depicted natural features as well as pits from surface imperfections. The vectors created therefore had to be further analyzed by defining criteria to separate the dolines from non-depressions. The

delimitation of the non-depressions were derived from threshold values of morphometric attributes, specifically depth, size, and shape. Evaluation of the appropriate threshold values was obtained by comparing maps of already identified dolines with previous mapping of dolines from field validation and interpretation of higher spatial resolution imagery (such as ALOS-PRISM and Google Earth images). Karst features in the area were easily identified through visual interpretation because the area is characterized by natural moist grassy vegetation where the water table approaches the surface for part of the year. A range of different threshold values were used for the minimum sink depth to identify the best threshold value from the maximum accuracy index between manual and automated classification. Identification of true sinkholes was assessed with the intersection of reference and classified polygons.



**Figure 41. Methodological procedures to determine the terrain attribute of sink depth. Digital elevation model (A) (DEM) Fillsink minus (B) original DEM results in the (C) sink-depth distribution.**

## **REMOTE SENSING AND GIS CONTRIBUTION TO THE INVESTIGATION OF KARST LANDSCAPES IN NW-MOROCCO**

*Theilen-Willige, B., Malek, H., Charif, A., El Bchari, F., & Chaïbi, M. (2014). Remote sensing and GIS contribution to the investigation of karst landscapes in NW-Morocco. Geosciences, 4(2), 50-72.*

The goal of this study was to investigate the use of different satellite data, like Landsat, RapidEye, and IKONOS imagery, as well as ASTER- and SRTM-derived digital elevation models (DEMs) for the analysis of karst features. Dolines were identified by visual interpretations based on high resolution images and aerial photographs. Digital image processing of the satellite data was useful to identify areas with higher surface water input, where karstification processes might be more intense than in surrounding areas. ArcGIS tools were used for this purpose by aggregating morphometric, causal factors (lowest and flattest areas) influencing the susceptibility to higher surface water input. Lineament analysis based on different satellite data contributed to the detection of near surface fault and fracture zones with potential for dissolution in subterranean waterways.

When searching for areas that are susceptible to karst processes, causal factors were taken into account, including slope gradient, curvature, lithology, or groundwater table level. Some of the causal factors could be determined from height level maps, slope gradient maps, curvature maps, and the drop raster which is calculated as the difference in the z-value divided by the path between the cell centers.

To visualize these areas, the weighted overlay approach integrated in ArcGIS was used for identification of endangered lowland areas that are susceptible to increased surface water input and possibly flooding due to their disposition by aggregating the causal factors. The susceptibility is calculated by adding every layer with a weighted influence (for example, a slope degree <math><10^{\circ}</math>:30%, lowest local height level: 30%, drop raster: 30%, curvature =0:10%) together and summing all layers. The sum, which can be included into GIS, provides some information on the susceptibility to surface water input and thus, on areas prone to more intense karstification processes. The resulting maps are divided into susceptibility classes which range from 0 to 6, where 6 is the strongest assumed susceptibility to surface water input due to aggregation and summation of causal, morphometric factors in the area. The study basically identified dolines and/or sinkholes through the use of GIS and remote sensing processing tools on satellite imagery and validation through field observation.

## **PREDICTING SINKHOLE SUSCEPTIBILITY IN FREDERICK VALLEY, MARYLAND USING GEOGRAPHICALLY WEIGHTED REGRESSION**

*Doctor, K. Z., Doctor, D. H., Kronenfeld, B., Wong, D. W. S., & Brezinski, D. K. (2008). Predicting Sinkhole Susceptibility in Frederick Valley, Maryland Using Geographically Weighted Regression. 41003(September), 243–256. [https://doi.org/10.1061/41003\(327\)24](https://doi.org/10.1061/41003(327)24)*

In this study, 556 identified sinkholes were analyzed to understand the likelihood of a sinkhole to form in a specific area. Characteristic features considered for predicting the density of sinkholes included clustering of sinkholes, geologic structure, rock type, and the sinkhole proximity to quarries, water bodies, streams, roads, faults, axes of synclines or anticlines, and depth to groundwater. The spatial statistical analysis methods used were K-function, Geographically Weighted Regression (GWR), and Inverse Distance Interpolation. After applying these functions, it was concluded that the proximity to the groundwater table, proximity to fold axes, proximity to faults, and proximity to quarries were the significant factors that greatly influence the formation of sinkholes.

### **K-Function**

The K-function cluster analysis was first used to depict clusters of point distribution over different spatial scales. The K function compares the observed distribution of sinkholes to the same number of random points. It is defined as:

$$K(h) = \frac{A}{N^2} \sum_i \sum_j \frac{I_h(d_{ij})}{w_i}$$

A= Total Area

N= Total Number of Points

$w_i$  = weight at  $i^{\text{th}}$  point

$d_{ij}$  = distance between  $i$  and  $j$  points

$I_h = 0$  when  $d_{ij} > h$  and  $I_h = 1$  when  $d_{ij} < h$

$h$  is the spatial lag which represent the change in scale

$K(h)$  = point density function of a random point

When compared to the point density function of a random point, it will give the difference function  $L(h)$  which is defined as:

$$L(h) = \sqrt{\frac{K(h)}{\pi} - h}$$

### **Density**

It is important to consider sinkhole density to determine areas that are more susceptible to the development of sinkholes. Brezinski (2004) developed the equation to calculate a sinkhole

potential index (SI) where newly collapsed sinkholes and dolines with no evident sinking throat and springs are considered. The equation is:

$$SI = \frac{\frac{\# \text{ of sinkholes}}{\text{mile}^2}}{\frac{\# \text{ of karstic features}}{\text{mile}^2}} = \frac{\# \text{ of sinkholes}}{\# \text{ of karstic features}}$$

It is assumed that the density in a particular unit is homogeneous, but this is unlikely a valid assumption. This equation takes into account a radial buffer area around each sinkhole and is expressed as the number of sinkholes in each buffer zone for which the lag spacing was calculated in the K-function equation. Each sinkhole receives a characteristic equal to density of sinkholes in the surrounding buffer zone. This is then used for future calculations to estimate sinkhole susceptibility with predicted density.

### **Geographically Weighted Regression**

GWR is described to be a robust technique for examining the relation of independent and dependent variables in an area of interest. This technique allows for a regression model to be calibrated separately for each sinkhole and for the parameters of the resulting numerous regression models to be explored in the areas of interest. When key features of the Frederick Valley region were identified, GWR analysis was used to express the significance of the factors and its relation to sinkhole density. GWR estimated local parameters by calculating the weight of each location as represented by the equation:

$$y_i = \beta_0 (u_i, v_i) + \sum_k \beta_k (u_i, v_i) X_{ik} + \epsilon_i$$

where  $(u_i, v_i)$  are where the  $i^{\text{th}}$  points are located. GWR was used using the GWR 3.0 software package and then imported into ArcInfo 9.2.

### **Inverse Distance Weights Interpolation (IDW)**

Although kriging is the most commonly used method for generating smooth surfaces, this study used an IDW spatial interpolation technique, producing a surface map from a GWR analysis to predict where the highest amount of sinkhole density would be located (Figure 42). In this method values are assigned to unknown points by using values of a scattered set of points assuming near locations have more influence on the predicted value than a further location, weighted inversely to the distance. The results of the GWR analysis showed the proximity to quarries, waterbodies, fault lines, fold axes and the groundwater table as being the most significant influential factors for sinkhole development. Future sinkhole developments cannot be detected on the map, but areas prone to sinkhole development are highlighted.

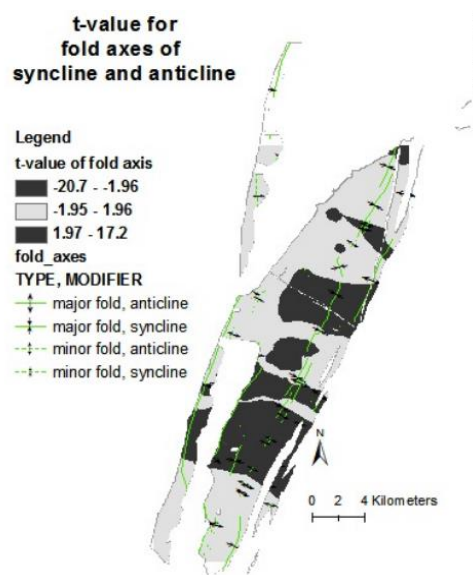


Figure 42. T-value mapping to represent proximity to syncline and anticline fold axes.

## AN EVALUATION OF AUTOMATED GIS TOOLS FOR DELINEATING KARST SINKHOLES AND CLOSED DEPRESSIONS FROM 1-METER LIDAR-DERIVED DIGITAL ELEVATION DATA

Doctor, D. H., & Young, J. A. (2013). *An evaluation of automated GIS tools for delineating karst sinkholes and closed depressions from 1-meter LiDAR-derived digital elevation data. 13th Multidisciplinary Conference on Sinkholes and the Engineering and Environmental Impacts of Karst*, 449–458. Retrieved from

[https://scholarcommons.usf.edu/cgi/viewcontent.cgi?referer=https://www.ecosia.org/&httpsredir=1&article=1156&context=sinkhole\\_2013](https://scholarcommons.usf.edu/cgi/viewcontent.cgi?referer=https://www.ecosia.org/&httpsredir=1&article=1156&context=sinkhole_2013)

In this study, one way of identifying sinkholes was identified using tools integrated into ArcGIS. This involved three steps: 1) filling in sinks to their spill level, 2) determining the flow direction within each pixel once the sinks in the DEM are filled, and 3) determining the flow accumulation of each pixel in the elevation model. The first step in the process (which has already been tested by FIU) was to use the Fill tool in ArcMap, which results in a new elevation surface with all sinks filled to their spill elevation. The process is repeated until all the pixels within the depression are filled and depressions spill over, removing all natural depressions. Poppenga et al. (2011) suggested using a threshold area of depressions specifically for drainage greater than or equal to 1,000 square meters and depth greater than or equal to 1 meter and greater than 0.5 standard deviation of the difference grid. This threshold, however, could result in exclusion of a number of depression artifacts. Once the DEM was filled:

1. The original DEM was subtracted from the filled DEM to generate a fill-difference raster.
2. All values in the fill difference raster greater than 0.10 m were extracted to a new raster.



3. The values in the threshold fill-difference raster were reclassified to an integer type raster where pixels less than 0.1 m were considered as “No Data”.
4. Remaining pixels were converted into polygons.

The resulting polygon layer represents the possible sinks from the fill-difference layer that was created in the steps prior and based on the accuracy of the downloaded DEM. To refine the polygons and classify the ones that are true depressions (and possible sinkholes), a training polygon feature class of known depressions that have been identified in the field and outlined from the LiDAR data is required for comparison. Geometric properties of the true and potential depression polygons were calculated and the results were compared. To calculate the geometric properties:

1. The Zonal Statistics as Table Tool (found in the ArcToolBox → Spatial Analyst Tools → Zonal → Zonal Statistics as Table Tool) was used to extract the MAXIMUM value (depth) from the fill difference raster from the overlying layer of each polygon. The table contains the area of the polygon and a count of the number of pixels per polygon.
2. The results of the zonal statistics table were joined to the attributes of the polygons. A threshold was then applied to remove polygons with a maximum depth value of less than 18 cm.
3. The Zonal Geometry tool (found in the ArcToolbox → Spatial Analyst Tools → Zonal → Zonal Geometry) was used for each of the polygons. Each polygon was attributed with its area, perimeter, and major and minor axis length of an idealized ellipse that contains the polygon.
4. The eccentricity of an ellipse was calculated as a measure of elongation of a potential sinkhole with values between 0 and 1. A threshold value of 0.98 was chosen for eccentricity based on visual examination of the manually delineated polygons and any depression above this was removed.
5. The circularity index was calculated. A threshold value of less than 1.7 was used in order to compare to the elliptical eccentricity results.

The study's results revealed that the semi-automated method of LiDAR and air photo captured 99.5% of the depressions captured in the manually created dataset. For FIU to employ this method, a manually delineated spatial dataset will be required. FIU has reached out to DOE collaborators to determine if this data is available.

## SUMMARY

---

FIU has conducted a literature review to determine an appropriate methodology for development of a high resolution (1-m or sub-meter) DEM of the WIPP and surrounding basins which will then be used for the development of a regional land surface model. Methods to develop a high resolution DEM included the use of Geographic Information Systems (GIS), Google Earth, LiDAR, and different interpolation algorithms. Of the methods investigated, a photogrammetric approach using Unmanned Aerial Vehicles (UAVs) is likely the most appropriate for FIU's research considering its affordability, flexibility, and ease of implementation in areas with limited site accessibility. Due to the cost of acquiring and processing LiDAR data for DEM development, the use of unmanned aerial vehicles (UAVs) to acquire digital photos will be considered. A high resolution DEM will ultimately serve as the stepping stone for development of the regional land surface and hydrological models of the WIPP site and surrounding basins. Once the DEM is created, post processing actions for extracting features of interest, such as sinkholes, will be implemented. The methods found in the literature for delineating and extracting sinkholes include developing an analytical hierarchical process, a localized contour tree method, GIS processes, and implementing statistical approaches. Coupling of the LSM with the ASCEM GWMs will enable more accurate predictions of groundwater flow patterns in the WIPP region. With improved estimates of the spatial and temporal patterns of recharge, predictions of halite dissolution and propagation of the shallow dissolution front will be made possible and the potential impact on repository performance quantified.

## REFERENCES

---

- Barringer, J. R. F., Pairman, D., & McNeill, S. J. (2002). Development of a high-resolution digital elevation model for New Zealand. Landcare Research Contract Report: LC0102/170.
- Brezinski, D.K. (2004). Stratigraphy of the Frederick Valley and its Relationship to Karst. Rep. No. 75, Annapolis: Maryland Department of Natural Resources.
- Chen, H., Oguchi, T., & Wu, P. (2018). Morphometric analysis of sinkholes using a semi-automatic approach in Zhijin County, China. *Arabian Journal of Geosciences*, 11(15). <https://doi.org/10.1007/s12517-018-3764-3>
- de Carvalho Júnior, O. A., Guimarães, R. F., Montgomery, D. R., Gillespie, A. R., Gomes, R. A. T., Martins, É. de S., & Silva, N. C. (2014). Karst depression detection using ASTER, ALOS/PRISM and SRTM-derived digital elevation models in the Bambuí group, Brazil. *Remote Sensing*, 6(1), 330–351. <https://doi.org/10.3390/rs6010330>
- Doctor, D. H., & Young, J. A. (2013). An evaluation of automated GIS tools for delineating karst sinkholes and closed depressions from 1-meter LiDAR-derived digital elevation data. 13th Multidisciplinary Conference on Sinkholes and the Engineering and Environmental Impacts of Karst, 449–458. Retrieved from [https://scholarcommons.usf.edu/cgi/viewcontent.cgi?referer=https://www.ecosia.org/&httpsredir=1&article=1156&context=sinkhole\\_2013](https://scholarcommons.usf.edu/cgi/viewcontent.cgi?referer=https://www.ecosia.org/&httpsredir=1&article=1156&context=sinkhole_2013)
- El-Quilish, M., El-Ashqer, M., Dawod, G., & El Fiky, G. (2018). Development and Accuracy Assessment of High-Resolution Digital Elevation Model Using GIS Approaches for the Nile Delta Region, Egypt. *American Journal of Geographic Information System*, 7(4), 107-117.
- Goodale, C. L., Aber, J. D., & Ollinger, S. V. (1998). Mapping monthly precipitation, temperature, and solar radiation for Ireland with polynomial regression and a digital elevation model. *Climate Research*, 10(1), 35-49.
- Hoffmann, E., and Winde, F. (2010) Generating high-resolution digital elevation models for wetland research using Google Earth™ imagery – an example from South Africa, *Water SA*, 36:1, 53-68.
- Kakavas, M. P., Nikolakopoulos, K. G., & Zagana, E. (2015, October). Karst features detection and mapping using airphotos, DSMs and GIS techniques. In *Earth Resources and Environmental Remote Sensing/GIS Applications VI* (Vol. 9644, p. 96440Y). International Society for Optics and Photonics.
- Kobal, M., Bertoneclj, I., Pirotti, F., Dakskobler, I., & Kutnar, L. (2015). Using lidar data to analyse sinkhole characteristics relevant for understory vegetation under forest cover -

## Literature Review of Methodologies for Development of a High Resolution DEM for WIPP

- Case study of a high karst area in the Dinaric mountains. *PLoS ONE*, 10(3), 1–19. <https://doi.org/10.1371/journal.pone.0122070>
- Launspach, J. (2013). Automated sinkhole extraction and morphological analysis in northeast Iowa using high-resolution LiDAR data. Retrieved from <http://scholarworks.uni.edu/etd%0Ahttp://scholarworks.uni.edu/etd/33>
- Lindsay, J. B., & Creed, I. F. (2006). Distinguishing actual and artefact depressions in digital elevation data. *Computers and Geosciences*, 32(8), 1192–1204. <https://doi.org/10.1016/j.cageo.2005.11.002>
- Liu, H., Jezek, K. C., & Li, B. (1999). Development of an Antarctic digital elevation model by integrating cartographic and remotely sensed data: A geographic information system based approach. *Journal of Geophysical Research: Solid Earth*, 104(B10), 23199-23213.
- Louhaichi, M., Borman, M. M., Johnson, A. L. and Johnson, D. E. (2003). Creating Low-Cost High-Resolution Digital Elevation Models. *Journal of Range Management*, 56:1, 92-96.
- Miao, X., Qiu, X., Wu, S.-S., Luo, J., Gouzie, D. R., & Xie, H. (2013). Developing Efficient Procedures for Automated Sinkhole Extraction from Lidar DEMs. *Photogrammetric Engineering & Remote Sensing*, 79(6), 545–554. <https://doi.org/10.14358/pers.79.6.545>
- Pohjola, J., Turunen, J. and Lipping, T. (2009). Creating High-Resolution Digital Elevation Model Using Thin Plate Spline Interpolation and Monte Carlo Simulation, Working Report 2009-56.
- Poppenga, S.K., Worstell, B.B., Stoker, J.M., and Greenlee, S.K. (2010). Using selective drainage methods to extract continuous surface flow from 1-meter LiDAR-derived digital elevation data. U.S. Geological Survey Scientific Investigations Report 2010–5059, 12 p.
- Rajabi, A., Kim, Y., Kim, S.-H., Kim, Y., Kim, B., & Nam, B. H. (2018). A Preliminary Study on Use of LiDAR Data to Characterize Sinkholes in Central Florida. 23–31. <https://doi.org/10.1061/9780784481585.003>
- Rajabi, A. (2018). Sinkhole Detection and Quantification Using LiDAR Data. 61. Retrieved from <https://pdfs.semanticscholar.org/6e97/ac20592c758b335184edaf30a4ac4a889769.pdf>
- Rosdi, M. A. H. M., Othman, A. N., Abdul, M. A. M. Z. Z., & Yusoff, Z. M. (2017). Sinkhole Susceptibility Hazard Zones Using GIS and Analytical Hierarchical Process (AHP): A Case Study of Kuala Lumpur and Ampang Jaya. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 42.
- Sharma, M., Paige, G. B., & Miller, S. N. (2010). DEM development from ground-based LiDAR data: A method to remove non-surface objects. *Remote Sensing*, 2(11), 2629-2642.

## Literature Review of Methodologies for Development of a High Resolution DEM for WIPP

- Theilen-Willige, B., Malek, H., Charif, A., El Bchari, F., & Chaïbi, M. (2014). Remote sensing and GIS contribution to the investigation of karst landscapes in NW-Morocco. *Geosciences*, 4(2), 50-72.
- Tibouo, F. N. (2016). Use and evaluation of LiDAR for mapping sinkholes in Royal Spring Groundwater Basin.
- Todd, A., & Ivey-Burden, L. (2016). A Method of Mapping Sinkhole Susceptibility Using a Geographic Information System: A Case Study for Interstates in the Karst Counties of Virginia. 299–306. <https://doi.org/10.5038/9780991000951.1048>
- Wall, J., Bohnenstiehl, D., & Levine, N. (2013). Automated sinkhole detection using a DEM subsetting technique and fill tools at Mammoth Cave National Park Methods : LiDAR Processing. 64, 2013.
- Wu, Q., Deng, C., & Chen, Z. (2016). Automated delineation of karst sinkholes from LiDAR-derived digital elevation models. *Geomorphology*, 266 (2016), 1–10. <https://doi.org/10.1016/j.geomorph.2016.05.006>
- Xie, Z., Pearlstine, L. & Gawlik, D. E. (2012). Developing a Fine-Resolution Digital Elevation Model to Support Hydrological Modeling and Ecological Studies in the Northern Everglades. *GIScience & Remote Sensing*, 49:5, 664-686, DOI: 10.2747/1548-1603.49.5.664.
- Zhu, J., Taylor, T. P., Currens, J. C., & Crawford, M. M. (2014). Improved karst sinkhole mapping in Kentucky using LiDAR techniques: A pilot study in Floyds fork watershed. *Journal of Cave and Karst Studies*, 76(3), 207–216. <https://doi.org/10.4311/2013ES0135>