Investigation of Error Patterns in Geographical Databases

Final Technical Report

ODURF Project # 105271 - end date 7/31/02

NASA NAG-1-2330

October 21, 2002

Submitted to NASA Langley Research Center

Principal Investigators:
David Dryer, Ph.D.
Derya A. Jacobs, Ph.D.

Graduate Research Assistants:
Gamze Karayaz
Chris Gronbech

TABLE OF CONTENTS

Appendix 1-Data Acquisition of USGS 7.5' DEM
Appendix 2-Select Visualizations for Sample Airports
Appendix 3-Select Statistical Output for Sample Airports
Appendix 4- Contour and Error Representation Visualizations for
All Sampled Airports

1. Introduction and Background

The objective of the research conducted in this project is to develop a methodology to investigate the accuracy of Airport Safety Modeling Data (ASMD) using statistical, visualization, and Artificial Neural Network (ANN) techniques. Such a methodology can contribute to answering the following research questions:

Over a representative sampling of ASMD databases, can statistical error analysis techniques be accurately learned and replicated by ANN modeling techniques? This representative ASMD sample should include numerous airports and a variety of terrain characterizations.

- Is it possible to identify and automate the recognition of patterns of error related to geographical features?
- Do such patterns of error relate to specific geographical features, such as elevation or terrain slope?
- Is it possible to combine the errors in small regions into an error prediction for a larger region?
- What are the data density reduction implications of this work?

ASMD may be used as the source of terrain data for a synthetic visual system to be used in the cockpit of aircraft when visual reference to ground features is not possible during conditions of marginal weather or reduced visibility. In this research, United States Geologic Survey (USGS) digital elevation model (DEM) data has been selected as the benchmark. Artificial Neural Networks (ANNs) have been used and tested as alternate methods in place of the statistical methods in similar problems. They often perform better in pattern recognition, prediction and classification and categorization problems. Many studies show that when the data is complex and noisy, the accuracy of ANN models is generally higher than those of comparable traditional methods.

In preliminary research, data was gathered for five airports that were used for training of ANN models. The data encompassed a 60 by 60 statute mile square region of interest (ROI) around an airport composed of geographic point pairs of elevation postings from both the ASMD and USGS databases. Before the development of ANN models, an

error value was computed for each point for the entire airport ROI using the following formula:

$$ASMD\ error = USGS\ elevation - ASMD\ elevation$$

With the computed errors, an average error over the entire ROI was calculated as used as an evaluating metric. The errors ranged from six meters to forty-two meters as shown below:

Asheville Regional (NC)	= 42.91 meter
McClellan–Palomar (CA)	=30.07 meter
Delta Municipal (UT)	=16.31 meter
Scappoose Industrial Airpark (OR)	=11.73 meter
Denver (CO)	= 6.34 meter

It is claimed that ninety percent of the elevations in the ASMD are within thirty meters of the actual elevation. According to the average ASMD errors, four of the five airports fall within thirty meters. These predetermined errors were used for comparison purposes in the development of ANN models. The project was conducted over two years and includes three phases. The accomplishments and the summary of results and findings are described in the following sections.

2. Project Phase I – August 2000 – October 2000

The following objectives have been accomplished during Phase 1 of the project in the first year:

- Literature search on the use of statistical models and ANN models
- Investigate error in geographical databases related to geographical characteristics

The literature review included ANNs associated with error and error modeling in geographical databases, as well as the classical ANN application areas. The goal was to

investigate different types of ANN algorithms and their appropriate application areas. A number of studies indicate that Backpropagation, Learning Vector Quantization, Self Organizing Map, Reinforcement, Hopfield, and Modular Network models may be best suited for this research in determining the error in geographical terrain databases. It was also concluded that Gaussian initialization may be the best way to handle the complex/noisy data contained in these databases. As a result of the literature review, these six ANN algorithms have been investigated for use in testing the airport data from the preliminary research.

3. Project Phase II – November 2000 – July 2001

The following objectives were achieved during Phase II in the first year:

- Investigate ANN application development software packages and identify one that is suitable for this project
- Generate data files for the five airports for training and testing of the ANN models
- Explore various ANN algorithms to best determine the error in the ASMD database

After the initial survey of ANN development packages, two software packages were considered: NeuroSolutions® and NeuralWorks Professional II/PLUS. It was decided that NeuralWorks Professional II/Plus included features that were appropriate for the needs of this project.

3.1 ANN Data Generation

For each of the five airports, one training file and one testing file were generated. Each of the training and testing files included two input patterns and one output pattern. Input pattern-1 included the elevation values from the ASMD database, and input pattern-2 included the elevation values from the USGS database. The output pattern was the predetermined error between the two elevations from the ASMD and USGS databases.

4

The ASMD and USGS elevation values (input values) were normalized between "0" and "1" when *sigmoid transfer function* was used during training. The formula below was used for normalization:

$$Y = \frac{1}{(max - min)} \times (elevation - min)$$

where   y = scaled value

max = maximum value of the related data

min = minimum value of the related data

The software automatically normalized the data between "-1" and "+1" when the *TanH transfer function* was used during training. Furthermore, the output values were coded as binary based on the following assumption criteria:

a) If USGS elevation > ASMD elevation, then the output value (error) is coded as "1", indicating an alarm or *risk error* in the ASMD elevation.

b) If USGS elevation < ASMD elevation, then the output value (error) is coded as "0", indicating no alarm or no risk in the ASMD elevation.

c) If USGS elevation = ASMD elevation, then the output (error) is coded as "0", indicating no alarm or no risk in the ASMD elevation.

3.2. Development of ANN Models

Based on the literature review results, the following ANN algorithms have been selected and investigated to design and develop ANN models to determine the ASMD error:

- Backpropagation

- Learning Vector Quantization
- Self Organizing Map
- Reinforcement
- Hopfield
- Modular

There were approximately 100 ANN models developed, trained and tested using the above six ANN algorithms. The Modular ANN models yielded the best results. With these results, it was decided to further investigate and enhance the Modular Networks model in order to improve the results achieved in error recognition. Modular Neural Networks consist of a group of networks (referred to as "local experts") competing to learn different aspects of a problem. A gating network controls the competition and learns to assign different regions of the data space to different local expert networks. The learning rule tends to encourage competition among local experts for different regions of the input space.

The best (optimal) Modular ANN model, among the 100 models, had one input layer (with two processing elements), one-hidden layer (with one processing element), and one output layer (with one processing element). For gating networks, one hidden layer with four processing elements and one output layer with three processing elements were used in the model. Various ANNs were trained and tested with more than one hidden layer. However, the results showed that one hidden layer ANN models performed better than the others with two or more. Momentum values were varied between 0.4 and 0.9, and ANN models with 0.7 momentum yielded best results. Similarly, those models with the *Extended-Delta-Bar-Delta Learning Rule* resulted in better performance.

Although the literature suggests that the *TanH transfer function* may be the best in real world applications, the number of ANN models developed in this research showed that the *sigmoid transfer function* yielded better results for this problem domain. For initialization and noise generation, a *Gaussian Distribution* was used. Although the software can automatically set the epoch equal to the number of vectors in the training file, the epoch was manually set up to 500. In order to set the error convergence criteria,

the software's RMS instrument was used and a threshold value of 0.001 was selected for convergence criteria. Finally, all the layers within the model were interconnected.

3.3 Results and Conclusions

The testing results for each of the five airports are shown in Tables 1, 2, and 3 along with the predetermined average error. These testing results demonstrate that Modular Network models are viable in determining the error in the ASMD database. The predetermined average error for each airport was used for comparison with the average error predicted by the ANN model.

Table 1. Testing Results-Asheville and Delta Airports

Asheville Regional Airport	Delta Municipal Airport
Number of Iterations: 7500	Number of Iterations: 7500
Transfer Function: Sigmoid	Transfer Function: Sigmoid
Learn Rule: Extended-Delta-Bar-Delta	Learn Rule: Extended-Delta-Bar-Delta
Momentum: 0.7	Momentum: 0.7
RMS: 0.0857	RMS: 0.0583
Actual Output Average: 40.77491	Actual output average: 15.16636
Desired Output Average: 37.89625	Desired output average: 10.5373
Predetermined Error: 42.91 meter	Predetermined Error: 16.31 meter

Table 2. Testing Results-Denver and Scappoose Airports

Denver International Airport	Scappoose Industrial Airpark
Number of Iterations: 7500	Number of Iterations: 7500
Transfer Function: Sigmoid	Transfer Function: Sigmoid
Learn Rule: Extended-Delta-Bar-Delta	Learn Rule: Extended-Delta-Bar-Delta
Momentum: 0.7	Momentum: 0.7
RMS: 0.0812	RMS: 0.0239
Actual Output Average: 8.44611	Actual output average: 9.05
Desired Output Average: 4.86944	Desired output average: 10.03
Predetermined Error: 6.34 meter	Predetermined Error: 11.73 meter

Table 3. Testing Results-McClellan-Palomar Airport

McClellan-Palomar Airport
Number of Iterations: 7500
Transfer Function: Sigmoid
Learn Rule: Extended-Delta-Bar-Delta
Momentum: 0.7
RMS: 0.0867
Actual Output Average: 39.389
Desired Output Average: 32.078
Predetermined Error: 30.07 meter

4. Phase III August 2001-July 2002

During the first year, an Artificial Neural Network (ANN) model was developed and tested with data for five airports. This Modular type ANN had one input layer, one hidden layer and one output layer. It was able to predict an average error for geographic point elevation pairs for a region of interest around an airport based on the data modeling assumptions of:

- If USGS elevation > ASMD elevation, then the output value (error) is coded as "1", indicating a risky error in the ASMD elevation.

- If USGS elevation < ASMD elevation, then the output value (error) is coded as "0", indicating no alarm or no risk in the ASMD elevation.

- If USGS elevation = ASMD elevation, then the output (error) is coded as "0", indicating no alarm or no risk in the ASMD elevation.

Based on the results achieved in the first year of the project, the investigators continued to enhance the models and test more airports in order to ensure validity of the models. In the second year, the project objectives were:

- Generate more data files for model development and testing
- Design, development and analysis of statistical model
- Design, development and analysis of ANN Model(s)
- Refinements to the selected best models
- Statistical analysis of the results to determine best ANN model(s) and/or statistical models

4.1 Airport Data File Acquisition

In the accomplishment of the objectives for Phase III, it was necessary to gather more geographic data for a newly selected set of airports. The five original airports chosen were picked to gather a variety of East Coast, Midwest and West Coast airports. The next set of airports was chosen to try and capture a varied range of terrain elevation characteristics. A good spectrum of geographic features was sought. The aim is to investigate geography from coastlines to mountain ridges, valleys to high elevations and peaks, and varying degrees of sloping terrain to large differences in elevation within a

selected region of interest covering a range of high positive elevation values to negative values. In the selection of these next airports, Sectional Aeronautical Charts published by the U.S Department of Transportation, Federal Aviation Administration and the National Aeronautical Charting Office were used to get a preliminary look at the terrain and help map out the region of interest before moving on to replicate it with computer applications. The procedures for the acquisition of data were modified slightly from the prior phase of research due to enhancements in utilized computer applications. The methodology was basically the same but there are some key details that should be noted to ensure accurate data acquisition and reduce rework of a very labor intensive and time consuming process. A checklist of procedures, supporting developed documents and necessary references are attached in Appendix 1.

The next set of airports contained three samples. They were Lake Havasu City Airport, John Wayne-Orange County Airport and Palm Springs Regional Airport. These airport data files covered a ROI 60 by 60 square statute miles around an airport reference point (ARP) and contained geographic latitude, longitude pairs of ASMD elevations to their associated USGS elevations. The preferred spacing of the elevation postings for the construction of the USGS ROI was 30 meters but some of the airports had only 10 meter spacing available for certain 7.5' x 7.5' Digital Elevation Models (DEMs) that composed the selected area. The drawback to this spacing variation in USGS source data was that a large data density of points had to be searched in matching corresponding geographic points. This added to processing time in the construction of sample files and had a definite impact on storage of the preprocessed files and on the amount of processing runs necessary for building the final sample. The USGS file of the entire ROI often had to be partitioned into smaller elements to stay within the bounds of the statistical error analysis program limitations.

4.2 Modular ANN Model Testing for New Airports

The three new airport data files were formatted and then testing files were developed for processing with the Modular ANN model. The testing results that are presented in Table 4 and Table 5 demonstrate this models' accuracy.

JOHN WAYNE-ORANGE COUNTY	LAKE HAVASU CITY
Number of iterations: 7500	Number of iterations: 7500
Transfer Function: Sigmoid	Transfer Function: Sigmoid
Learn Rule: Extended-Delta-Bar-Delta	Learn Rule: Extended-Delta-Bar-Delta
Momentum:0. 7	Momentum:0. 7
RMS: 0.0442	RMS: 0.0496
Actual Output Average: 21.80	Actual Output Average: 32.87
Desired Output Average: 19.1835	Desired Output Average: 23.7240
Predetermined Error: 23.96	Predetermined Error: 23.51

Table 4.Testing results-John Wayne and Lake Havasu Airports

PALM SPRINGS REGIONAL
Number of iterations: 7500
Transfer Function: Sigmoid
Learn Rule: Extended-Delta-Bar-Delta
Momentum:0. 7
RMS: 0.0570
Actual Output Average: 63.97
Desired Output Average: 67.0852
Predetermined Error: 57.60

Table 5.Testing results-Palm Springs Airport

4.3.New Data Modeling Assumption

National Imagery and Mapping Agency's (NIMA) Digital Terrain Elevation Data 1 (DTED1) files serves as the source data for the ASMD models. The data vertical accuracy objective of this DTED1 is stated by NIMA as +- 30 Meters at 90% linear error (LE). The full resolution 3 arc second DTED have a vertical accuracy of ± 30 meters LE at the 90% confidence level. If the error distribution is assumed to be Gaussian with a mean of zero, the statistical standard deviation of the errors is equivalent to the root mean square error (RMSE). Under those assumptions, vertical accuracy expressed as ± 30 meters linear error at 90% can also be described as an RMSE of 18 meters. Concerning the Airport Safety Model (ASM), which is based on this data, no greater accuracy is implied or should be assumed.

Absolute vertical accuracy (in meters):

Data source	RMSE	LE at 90%
DTED	18	± 30

It is claimed that ninety percent of the elevations in ASMD are within thirty meters of the actual real world elevation. In this project, USGS data is used as a more accurate baseline standard for comparison for actual real world elevations. USGS source data was selected due to its unclassified availability as well as a higher resolution and accuracy claims when compared to DTED1. The following USGS vertical elevation accuracy statement supports this, "For a 7.5 minute DEM derived from a photogrammetric source, ninety percent must have a RMSE of seven meters or better and ten percent are in the eight to fifteen meter range" (Fact Sheet102-96). Based on these accuracy claims, USGS

vertical accuracy (7 m RMSE) does appear to have a higher vertical accuracy standard than the ASMD/DTED1 vertical accuracy (18 m RMSE).

Using USGS as an elevation data standard or baseline, the ASMD error (the USGS elevation minus ASMD elevation) should fall within the ± 30-meter range to reach a conclusion that this is an accurate ASMD elevation point. Due to aircraft safety considerations, our analysis needed to identify ASMD elevations that were significantly greater than the USGS standard points. Therefore, the following data modeling assumptions are added to our research during the second year.

- If the ASMD error is greater than +30 meters, it is dangerous and coded as "1"
- If the error is equal to or smaller than +30 meters, it is safe and coded as "0"

4.4 Visualization and Statistical Analysis

For each airport data file sample, analysis was done on the elevation error between the given ASMD value and the corresponding discovered USGS value. Histograms of the errors were created and statistical measures of the error are presented in tabular format. Denver International Airport is used to illustrate the error analysis, which is presented in Figure 1 and Table 6 with the rest of the airports presented in the discussion of analysis and results.

Bin	Frequency
-43	1
-35.1	2
-27.2	15
-19.3	131
-11.4	661
-3.5	3134
4.4	15138
12.3	12860
20.2	4358
28.1	1809
36	720
43.9	265
51.8	159
59.7	59
67.6	22
75.5	22
83.4	14
91.3	7
99.2	6
107.1	11
115	4
More	0

Denver International Airport
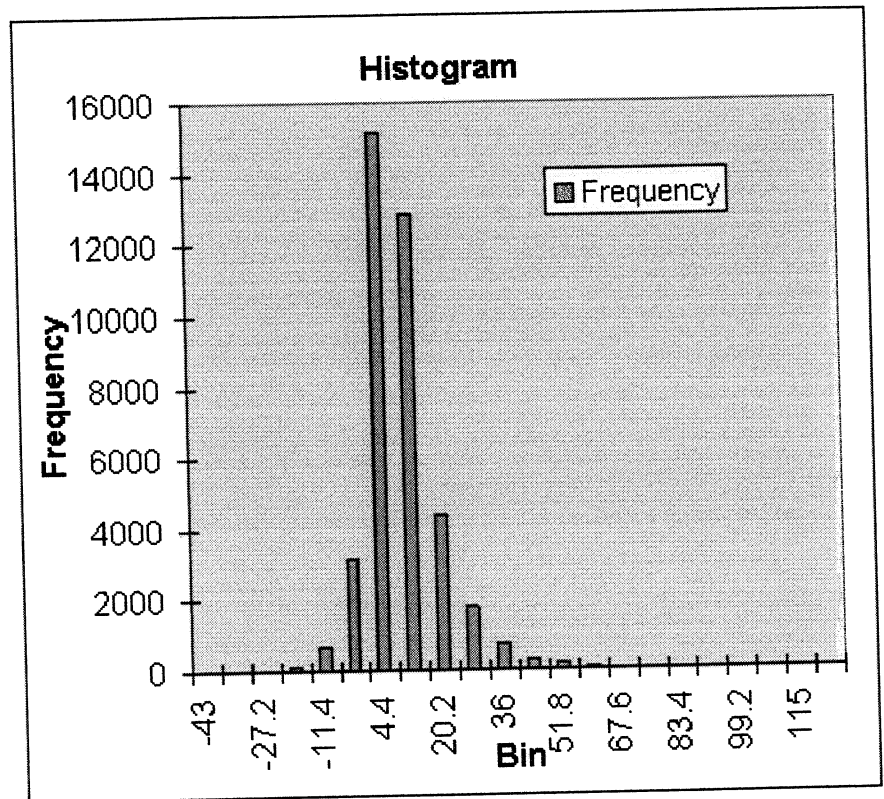
Figure 1. ASMD Error Histogram

Mean Error	6.34311386
Median	5
Mode	4
Standard Deviation	10.2597157
Sample Variance	105.261766
Range	158
Minimum	-43
Maximum	115
Count	39398

Table 6. ASMD Error Statistical Measures

Additionally with our assumption of a dangerous value having an ASMD error greater than +30m, the total number of dangerous and safe elevations and the percentages

of each were calculated and the data for Denver International is presented in Table 7 with the rest of the airports presented in the discussion of analysis and results.
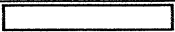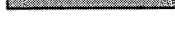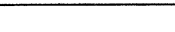
# Safe values	38324
# Danger values	1074
Total	39398
% Safe	97.27%
% Dangerous	2.73%

Table 7. Safe/Dangerous ASMD elevation points

To support the statistical aspect and follow-on ANN models, a visualization component was added to this phase of the project. ARCVIEW® GIS 3.2 from the Environmental Systems Research Institute Inc. was utilized for this portion of the project for its ability to display both two dimensional and three dimensional representations of geographic data.

Each airport data file acquired during the project was decomposed and formatted into files of the ASMD elevations, USGS elevations and Error values. Two-dimensional views and three-dimensional scenes of this data and varying combinations of this data were built. The 2-dimensional views and 3-dimensional scenes were composed of both point features and Triangulated Irregular Networks (TINs). The point features were necessary when there were voids in the data. Some files were not continuous areas due to a lack of available data within the USGS database or due to bodies of water for which no DEM data exists. Again, Denver International Airport (DEN) is used to illustrate the utilization of visualizations and the capabilities of ARCVIEW®. More visualization for other airports will be included in Appendix 2 and are addressed in a discussion of analysis and results.

Denver International Airport

Color	Elevation(m)
	2172-2274
	2070-2172
	1968-2070
	1866-1968
	1765-1866
	1663-1765
	1561-1663
	1459-1561
	1357-1459
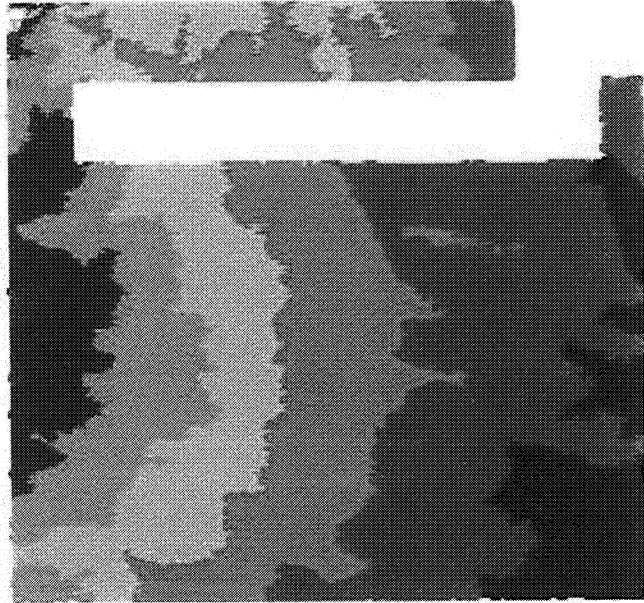
2-Dimensional Representation of DEN from ASMD

3-Dimensional TIN of DEN from ASMD
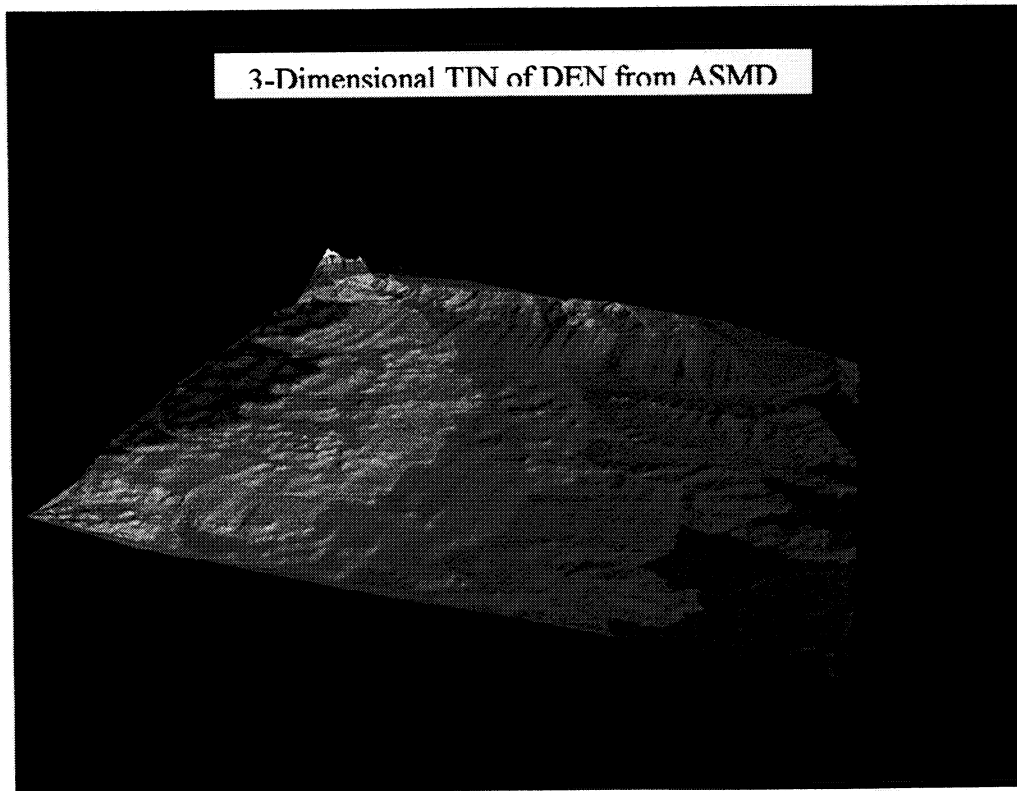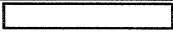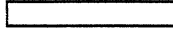
Figure 2. Visualizations of DEN from ASMD
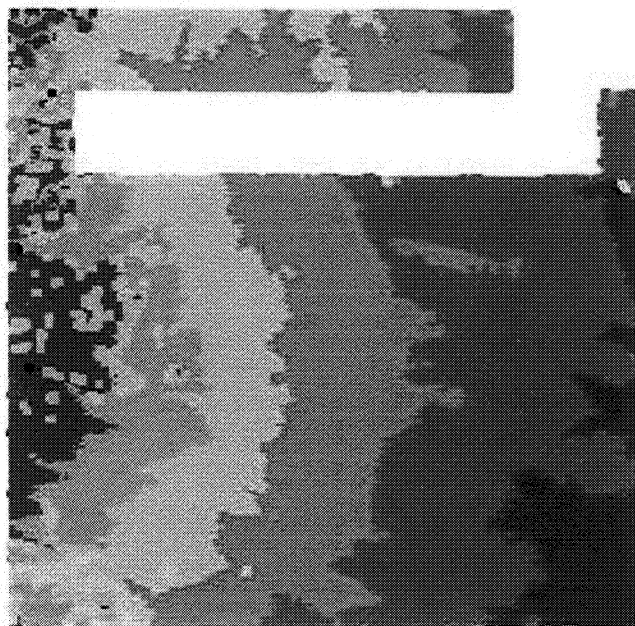
2-D view of ASMD Error v. USGS Elevation

Color	Elevation (m)
	2180-2283
	2076-2180
	1973-2076
	1870-1973
	1766-1870
	1663-1766
	1560-1663
	1456-1560
	1353-1456

Color	Error (m)
	115 - 50
	50 - 30
	30 - -30*
	-30 - -43

* This color band is transparent to so that values within limits show through.

Figure 3. Visualizations of DEN from USGS data and ASMD Error data
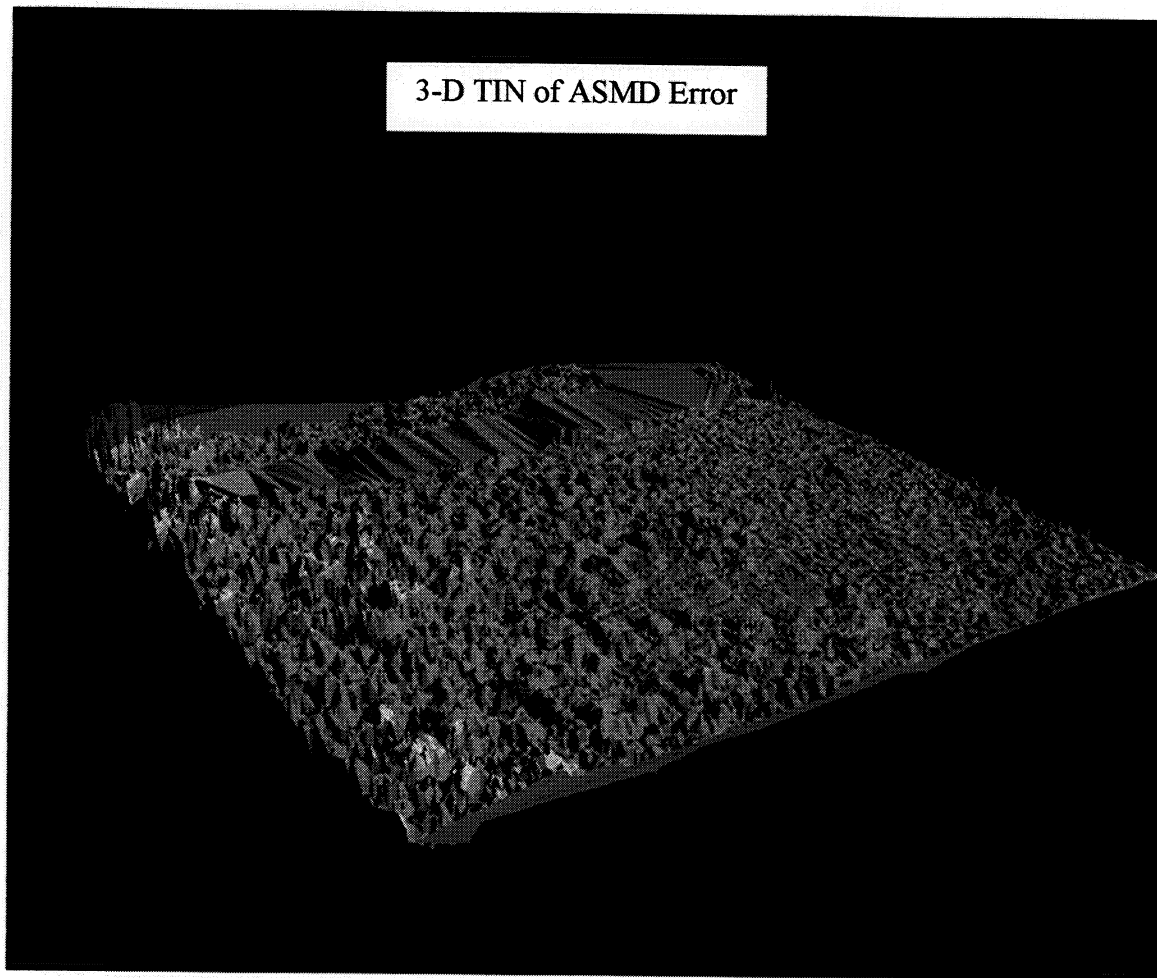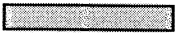
Color	Elevation (m)
	50-115
	30-50
	-30 - 30
	-43 - -30

Figure 4. Visualization of ASMD Error data for DEN

Figure 5. Visualizations of DEM with USGS Elevation and ASMD Elevation data

ASMD Elevation Legend	
Color	Elevation (m)
	1357-1459
	1459-1561
	1561-1663
	1663-1765
	1765-1866
	1866-1968
	1968-2070
	2070-2172
	2172-2274

USGS Elevation Legend	
Color	Elevation (m)
	1353-1456
	1456-1560
	1560-1663
	1663-1766
	1766-1870
	1870-1973
	1973-2076
	2076-2180
	2180-2283

3-D TIN of USGS Elevation overlaid ASMD Elevation
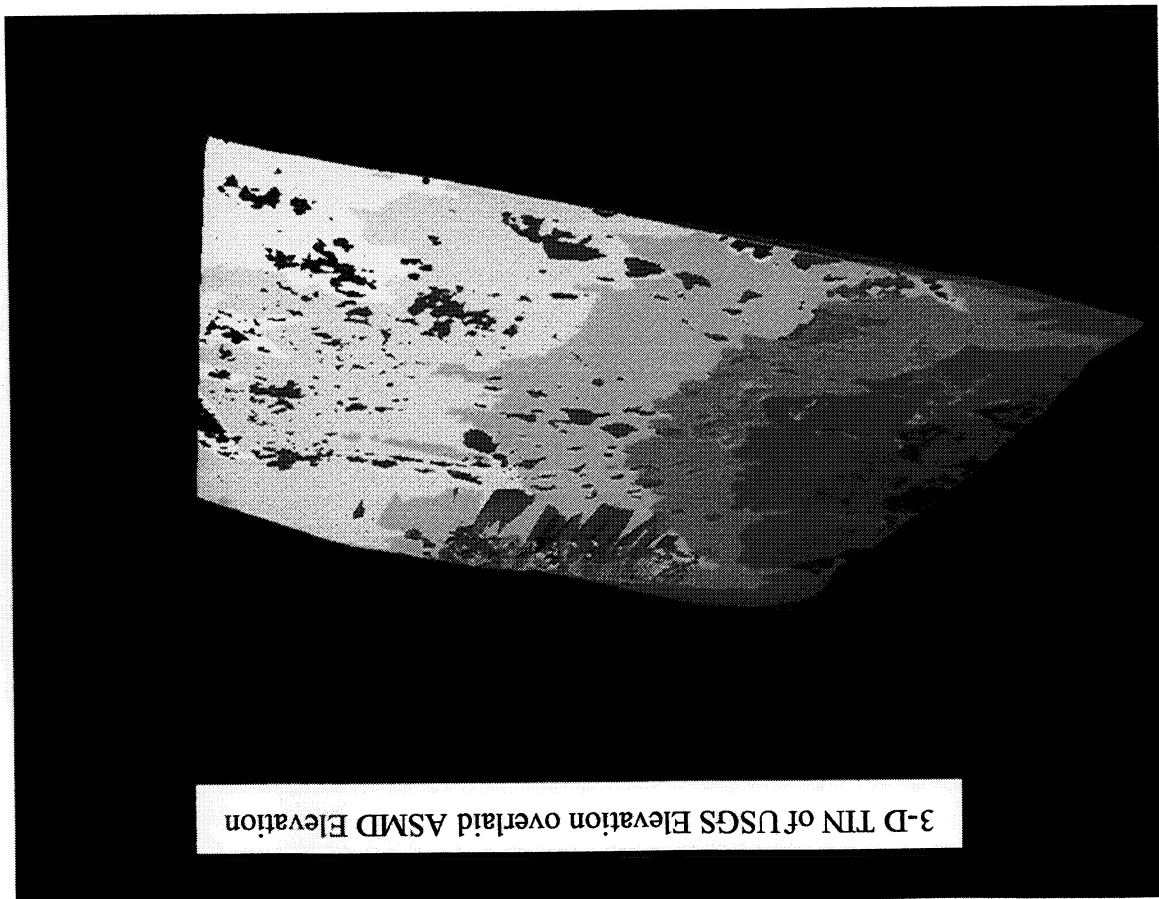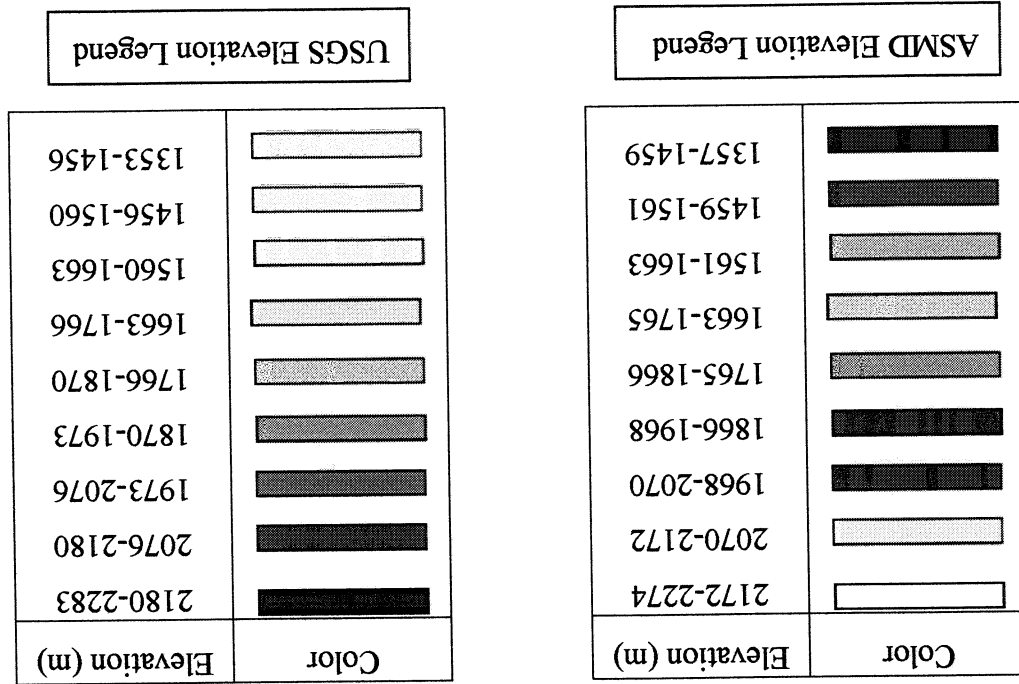
The description, analysis and conclusions are discussed for Denver International and the other airports in a follow on section.

4.5 New Modeling Assumptions with the Modular ANN Model

The Modular ANN was trained and tested with the new modeling assumptions. The RMS results are presented in Figure 6. The results varied from 0.3043 to 0.5079 and the average RMS for the eight airports was 0.4099. This is considered a high RMS result that lead to the conclusion that another ANN model should be considered.
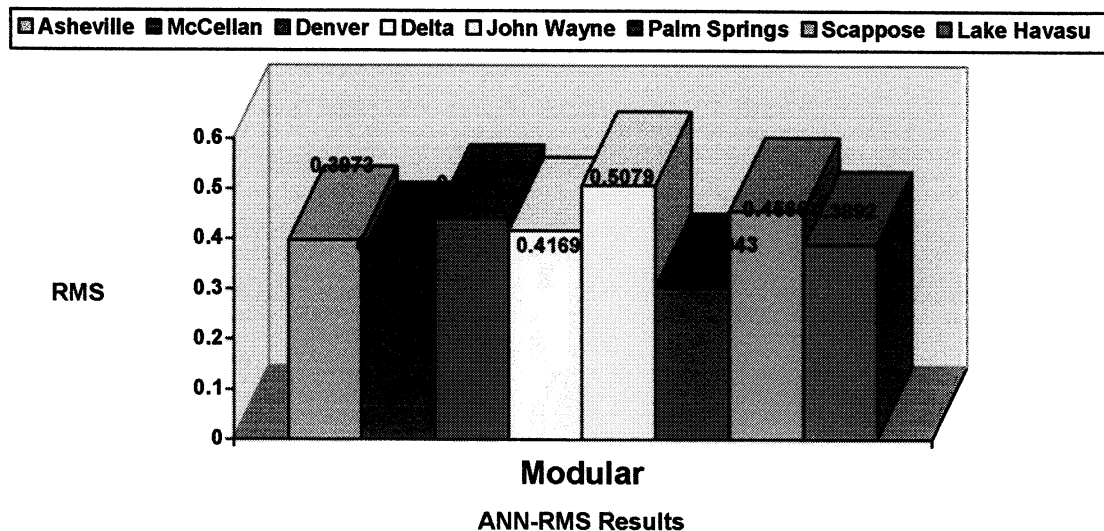
Figure 6. ANN RMS Results

4.6 Backpropagation Research Summary

The research for another candidate ANN model, which could produce better RMS results, focused on backpropagation networks for the reasons described below. The backpropagation-learning paradigm is very popular among neural network researchers

due to its capability of recognizing large amount of data. These networks are consistent estimators of binary classifications under similar assumptions. There are potential advantages of backpropagation identified in the literature for the type of complex data analysis, which is required in the investigation and error comparisons of large elevation databases. It is a general-purpose non-linear regression technique, which attempts to minimize global error. Any multi-dimensional function can in theory be synthesized by a back-propagation network. It can provide very compact distributed representations of complex data sets.

Preliminary work has shown that one of the central issues of a back-propagation type network is to set up an appropriate learning rate. It is important to keep learning rate low, although it can lead to very slow learning. Therefore, the momentum is used for faster learning with a low learning rate. Another problematic area is to appropriately set up the number of hidden layers and processing elements.

Therefore, after reviewing the literature, the following research questions have been followed to design the optimum architecture.

- How does learning rate and momentum affect ANN learning? What is the best combination for training for this geographical database?

- How does the hidden layer size affect the learning? What are the optimal processing elements in the hidden layers for this application?

The investigation of these research questions is now discussed.

4.7. Backpropagation ANN Data Generation

A new set of data was generated for the training and testing processes for each airport. The ASMD and USGS elevation values were scaled between "0" and "1" for

inputs and the output (error) values were coded as binary based on the modeling assumptions adopted in Phase III:

- If the error is greater than 30 meters, it is coded as "1"

- If the error is less than or equal to 30 meters, it is coded as "0"

In this study, one training file and one testing file were used for each of 17 airports. Although testing and training files are randomly selected samples from the data, both training and testing files are reasonably representative of an entire ASMD database's area sample points. Therefore, the training/testing data file sizes are considered independent for each airport and the hidden elements, layer size, learning rate and momentum are dependent.

The initial network weights were set up randomly between −10 and 10 and then on-line training was used for each case. The network was able to update the network weight after training. In all of the trials, the sigmoid function was used since data was scaled and coded between "0" and "1".

4.8 The effect of learning rate and momentum on ANN learning

Figure 7 shows a simple standard Backpropagation network that is developed with one input layer, two hidden layers and one output layer.

For the experiments the network was built with the input layer having three processing elements, the hidden layer one processing element and the output layer one processing element. The Delta learning rule was used for training with the sigmoid transfer function.
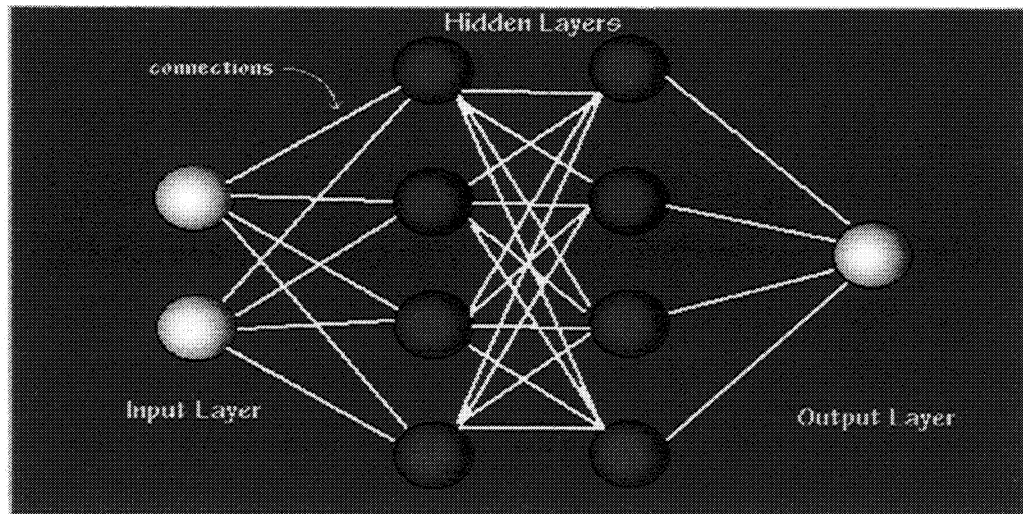
Figure 7. Sample Backpropagation network

Experiments were conducted to iteratively determine momentum and learning rate

parameters that would perform well using number of iterations and an acceptable RMS as

measures of performance. When data is coded based on the 0 and 1 criteria, the research

shows, a lower RMS corresponds to better performance. Therefore RMS was chosen as a

critical measure of network performance. The Asheville Regional airport is used in these

experiments. In previous model runs it was the most challenging airport to model and

database elevation errors had the largest elevation error values. The momentum

parameter, which makes the weight distribution properly, was determined first as is

shown in previous backpropagation research studies. The experimental results can be

found in table 8,9, 10 and in figures 8 and 9. For these experiments, RMS convergence

criterion as 0.010 was used for error determination; therefore, the number of iterations

was used to determine best level of parameters. The learning rate was set as 0.9 and the

momentum was changed between 0.3-0.9. The results of this experiment showed that

when the learning rate was fixed at 0.9; the momentum value of 0.6 had a significantly

influence reducing the number of iterations. Then, the momentum value was set at 0.6 and the learning rate was changed between 0.3-0.9.

Learning rate	0.9	0.9	0.9	0.9	0.9	0.9	0.9
RMS	0.010	0.010	0.010	0.010	0.010	0.010	0.010
Momentum	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Iteration	4436	7233	7445	4244	4491	4456	9548

Table 8. ANN Parameter Sensitivity Analysis of Asheville Airport – Learning Rate of .9, Variable Momentum
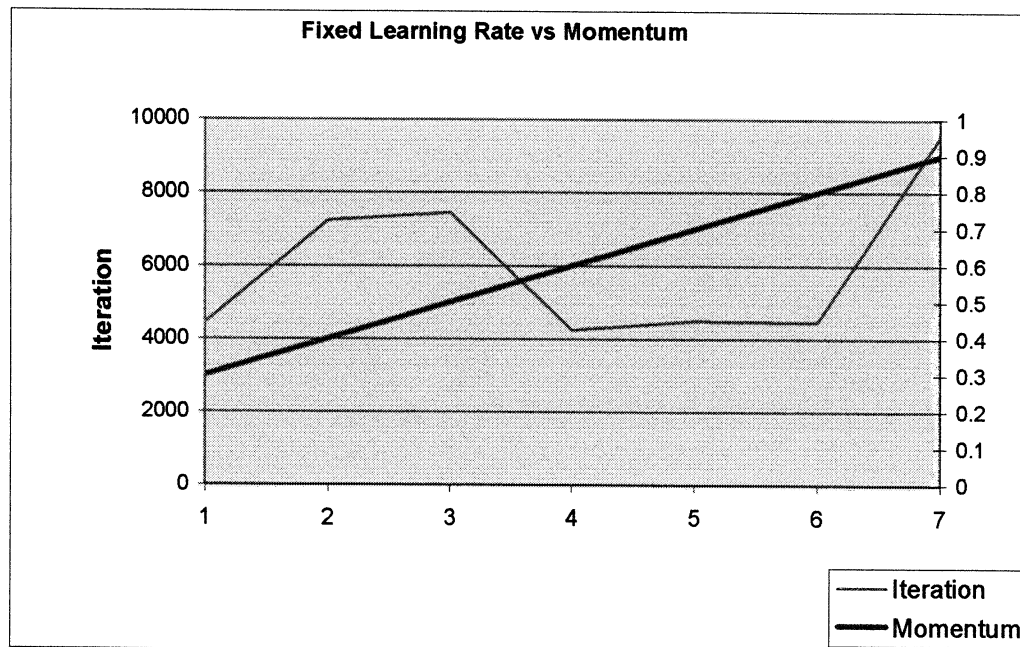
Figure 8. Asheville Airport: Varying Iterations and Momentum Values for a Fixed Learning Rate (.9)

Learning rate	0.3	0.4	0.5	0.6	0.7	0.8	0.9
RMS	0.010	0.010	0.010	0.010	0.010	0.010	0.010
Momentum	0.6	0.6	0.6	0.6	0.6	0.6	0.6
Iteration	4460	3467	4218	4128	4165	4329	4207

Table 9. ANN Parameter Sensitivity Analysis of Asheville Airport Momentum of .6, Variable Learning Rate
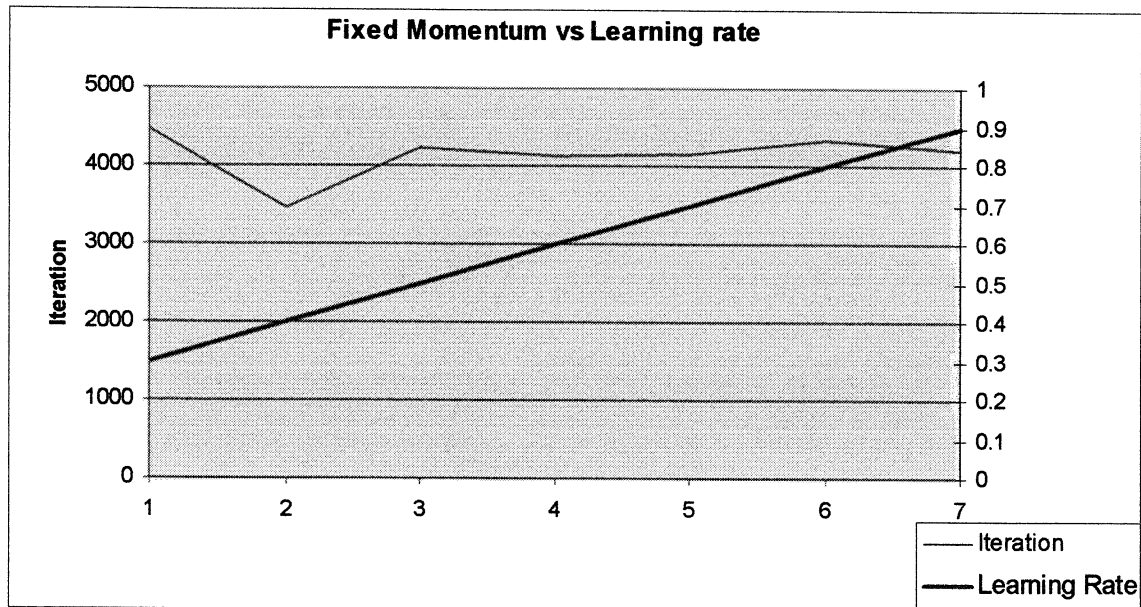
Figure 9. Asheville Airport: Varying Iterations and Learning Rate Values for a Fixed Momentum (.6)

As shown in the Table 8 and Table 9, the learning rate with 0.9 yielded the best result with 0.6 momentum value. The learning rate with 0.4 yielded the best result with 0.6 momentum value in two experiments. After determining that, it was tested with McClellan airport. The results can be found in Table 10.

Learning rate	0.9	0.4
RMS	0.4795	0.4793
Momentum	0.6	0.6
Iteration	8938	5917

Table 10. McClellan-Palomar Airport ANN Backpropogation Results

The results showed that 0.6 is an appropriate number for momentum. Although a small (0.4) learning rate yielded best result, there is barely difference between two learning rates based on RMS. Based on the above experimental results, it was determined the best parameter values for momentum was 0.6, and for learning rate was 0.4 using the iteration size as a determining criteria. It is concluded that larger momentum works best with smaller learning rate for the threshold ASMD vs. baseline USGS error analysis being conducted.

4.9 The effect of hidden layer size and processing elements in the hidden layers

The literature shows that when the network has hidden layers, the results of training might depend on the random weights. Therefore, as mentioned earlier in this report, the initial network weights were set between −10 and 10 and then on-line training was used for each case.

Having determined the momentum and learning rate, layer parameters in the network were explored, including the number of hidden layers and number of processing elements per layer. There is no theoretical limit on the number of hidden layers, but some research has been done which indicates that a maximum of four hidden layers are required to solve complex pattern recognition problems. Each layer should be fully

connected to the succeeding layer. During learning, information is propagated back through the network and used to update the connection weights.

In order to determine layer parameter combinations that performed well, the iterations were fixed at a reasonable level of 35,000 and the lowest RMS value was used as the measure of performance. To start this process, a second layer was added to the network and the number of processing elements was increased.

Table 11 summarizes the testing results using Asheville airport data. Each hidden layer was tried with various processing elements starting from 1 to 10. As shown in the Table 11, the network that has more hidden layer with three processing elements yields better result than a simple network with one hidden layer and one processing element in the layer. The best (lowest RMS) network is highlighted in Table 11, which has two hidden layers, with three processing elements in the first hidden layer and two processing elements in the second hidden layer.

1st hidden layer	2nd hidden layer	RMS	1st hidden layer	2nd hidden layer	RMS
1	1	0.2780	4	3	0.2754
2	1	0.2776	4	4	0.2731
2	6	0.2722	5	5	0.2725
3	1	0.2803	6	1	0.2813
3	2	0.2707	10	1	0.2757
3	3	0.2754	6	2	0.2766
3	4	0.2751	8	6	0.2750
3	5	0.2747	10	10	0.2762

Table 11. Backpropogation Processing elements-Asheville Airport
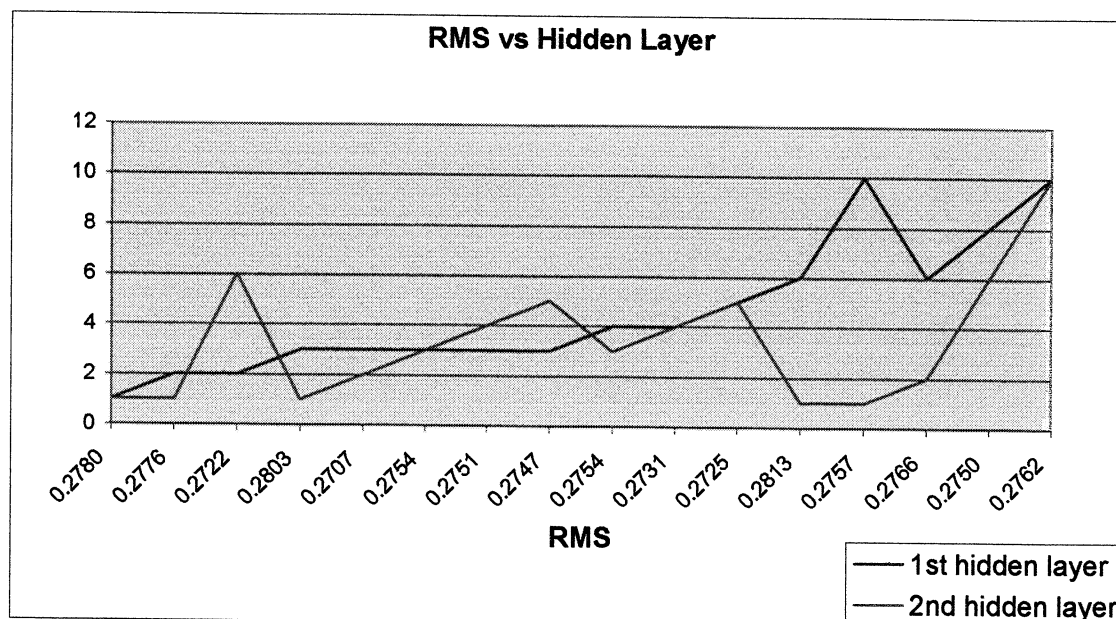
Figure 10. RMS vs. Number of Processing Elements in Hidden 1ˢᵗ and 2ⁿᵈ Layers

From Table 11 it can be concluded that first hidden layer with large number of processing elements does not yield good results. Therefore, from the above trials, it was determined that three processing elements in the first hidden layer and two processing elements in the second hidden layer produce the best observed results. This network model with an optimized number of layers and processing elements was then validated using Denver and McClellan Airports. The results showed a significant reduction in RMS values. For example, McClellan Airport changed from 0.4793 shown in Table 10, to the 0.1793 RMS value shown in Table 12.

Based on the testing results, the following RMS results were achieved:

Airport	Asheville Airport	Denver Airport	McClellan Airport
RMS	0.2707	0.2191	0.1793

Table 12. Improved model RMS results

4.10 Analysis, Results and Airport Data File Discussion

With the design of the Artificial Neural Network model accomplished, an analysis

and discussion of findings can begin. The first issue is an analysis of the viability of the

Backpropagation ANN for the given airport data. To address this, a comparison of this

network will be done with the Modular network investigated in Phase II. In Phase III, the

modeling analysis criteria was changed to address ASMD elevation not only above or

below the USGS elevation, but the ASMD error outside a tolerance of +30m of the

baseline USGS database. Figure 11 represents the RMS values using the data modeling

criteria of Phase II. Both the Modular Network and Backpropagation Network were

tested using this modeling criteria.
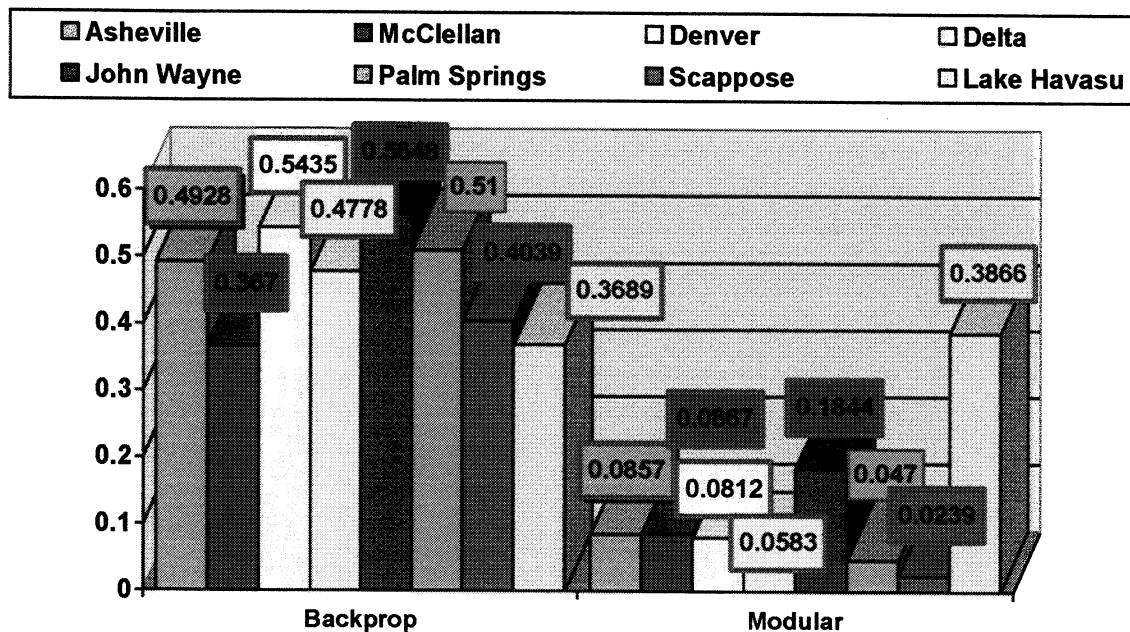
Figure .11. ANN RMS Results with Phase II Error Pattern Criteria

Based on the Modular network results from Figure 11, it is concluded that the

Modular type of network model is more successful in recognizing the error pattern in

geographical databases where the error is classified as risky or not with the criteria of the

29

ASMD elevation lower than the baseline USGS elevation. The network properly

classified 100 percent of the patterns, but some of the airports results had high RMS

values (e.g., Lake Havasu City), when compared to the other airports when the Modular

network is used. Lake Havasu has the highest average error for the prescribed ROI

(57.60 meters) that could be a factor in the high RMS.

The Backpropagation results from Figure 11, yielded high RMS values. It is

concluded that Backpropagation may not be capable of recognizing the pattern in the

error database when using the data modeling assumption of Phase II. In other words, it

might not be possible from the results to interpret efficiently whether the

Backpropagation network learned risky areas when these areas are defined as any ASMD

elevation merely exceeding the USGS elevation.

Figure 12 represents the RMS values using the data modeling assumptions

adopted in Phase III.
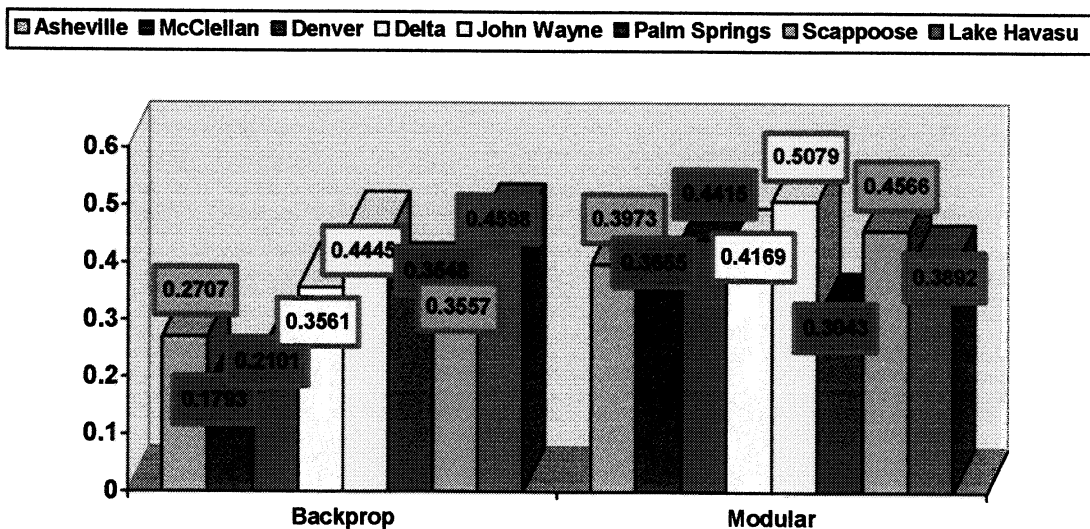
Figure .12. ANN RMS Results with Phase III Error Pattern Criteria

From Figure 12, it can be concluded that Backpropagation is better able to recognize the error pattern in the database using the data modeling assumptions of Phase III where an error threshold of thirty meters is considered. The results show that Lake Havasu and John Wayne Airport yield a little higher RMS values than the others. For the outputs illustrated in Figure 12, a quick discussion of the input data is necessary. The input data was just a random sample taken from the ROI of each airport shown. This detail could have an influence on the results and this will be addressed later.

Before any conclusions are made with respect to the RMS of specific airports, it is useful to analyze the data visually. This can help identify potential underlying factors influencing ANN results. Since Denver International Airport (DEN) was introduced earlier in section 4.4, illustrating the elements of the visualization component, this airport will be discussed first. The 2-D representation of Figure 13 shows a 3600 square mile area around DEN with areas of missing data. This missing data is of no consequence, with the qualification that the summary of results is for the area covered by the data represented by each airport file and not of a ROI with defined dimensions of 60 by 60 statute miles. The view of interest is that of the USGS standard with contour lines as shown in Figure 13. A depiction of contour lines on the elevation view of USGS data shows that terrain depicted around this airport is not greatly sloped. Utilizing a contour interval of 100 meters, it can be seen that the airport is on a large plain at high altitude with a gently rolling slope. This is an important characteristic influencing the comparison of geographic data and the ANN viability.
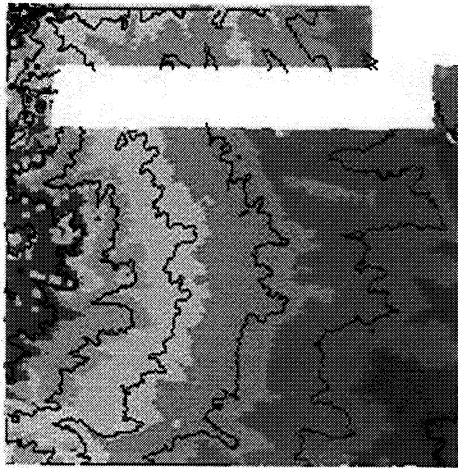
Figure 13. 2-D USGS Visualization with Contours

The minimally sloped terrain lends to the capability of the neural network model. For

Denver International (DEN), the RMS was .2101. Using Denver International Airport

databases, a sample file of data was taken randomly from the ROI and processed to

discover how well the ANN mimicked the true output in categorizing geographic

positions above or below a designated threshold. For this process the data was coded

binary with ones and zeros representing dangerous (>30m error) or safe (<=30m error)

positions. Figure 14 is a visual representation of the ANN sample file.
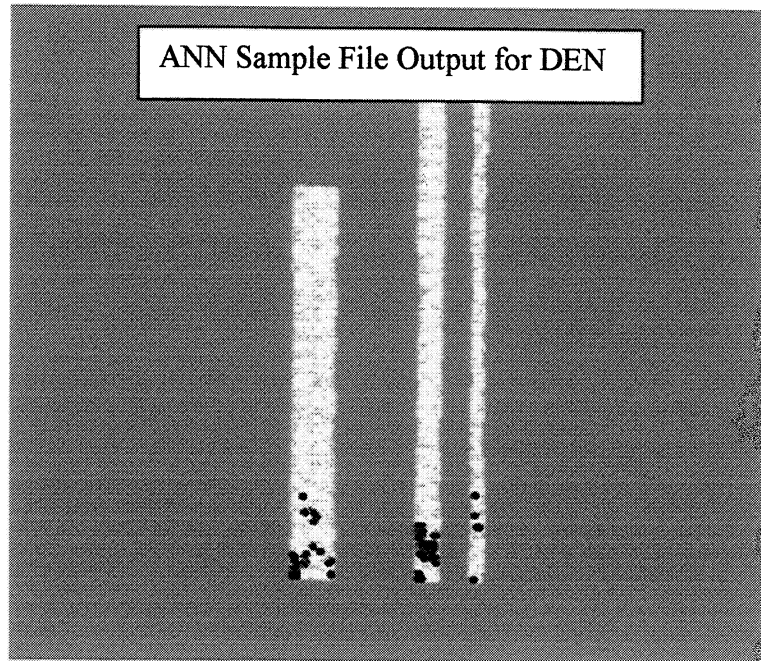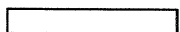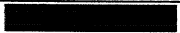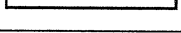
Figure 14. Visual Representation of DEN Sample file ANN Output

Figure 15. ANN Output v. Expected Output
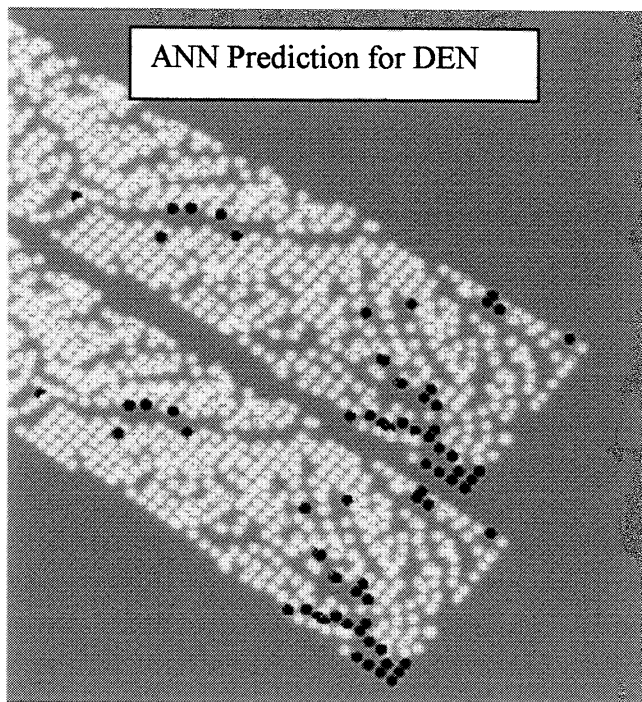
Color	Input
⬛	1
⬜	0

Expected Output Legend

Color	Output
⬛	0.9 - 1.0003
⬛	.75 - 0.9
⬛	.25 - .75
⬜	0.1 - .25
⬜	-.0004 – 0.1

ANN Output Legend

Figure 15 is a visual representation of the sample file ANN output when viewed with the

desired output. Figure 15 contains two bands in a zoomed view of the sample file. The

upper band is the expected output and the lower band is the ANN output. The legend

displays the coded output of values <=30m, coded 0, as white and the values >30m,

coded 1, in black. The ANN output is presented in a grayscale.

Figure 16 is a graphical representation of the ANN output from the sample file for

DEN.

Bin	Frequency
-0.000443	1
0.0495921	5258
0.0996272	0
0.1496623	0
0.1996974	0
0.2497325	0
0.2997676	0
0.3498027	0
0.3998378	0
0.4498729	0
0.499908	0
0.5499431	0
0.5999782	0
0.6500133	0
0.7000484	0
0.7500835	0
0.8001186	0
0.8501537	0
0.9001888	0
0.9502239	0
1.000259	113
More	0
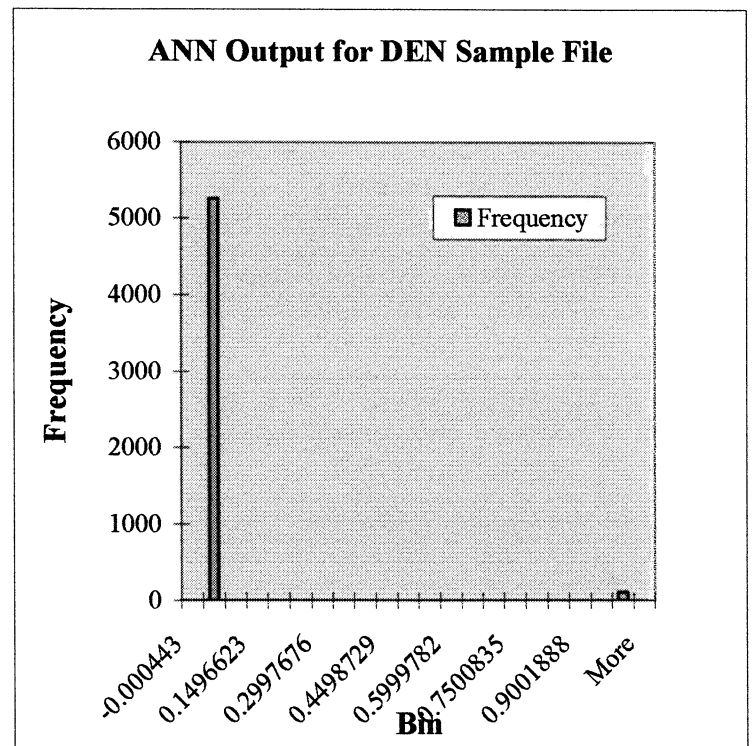
Figure 16. Graphical Representation of ANN Output

The visual display and graph shows the ANN output was essentially zero on the low side

and one on the high side. These are the values for which the ANN was designed. The

ANN correctly mimicked the desired output and performed as expected.

For the eight airports illustrated in Figure 12, the ANN correctly distinguished between safe elevation points and dangerous elevation points every time. This was not always correctly identified with a classification of one or zero. In some instances the classification was on the order of .8 on the high side and .2 on the low side. Although the classification was different, the distinction was absolute and the match of safe points to the low value and dangerous points to the high value was exact. One correlation can be seen in the RMS. With an RMS above .2101, the ANN output was .8 and 2. This is illustrated by the Asheville Regional Airport (AVL) sample file. The complete sample file for AVL with both the expected (coded) outputs and ANN outputs are shown in Figures 17 and 18. The color scales for these figures are the same as previous ANN displays.
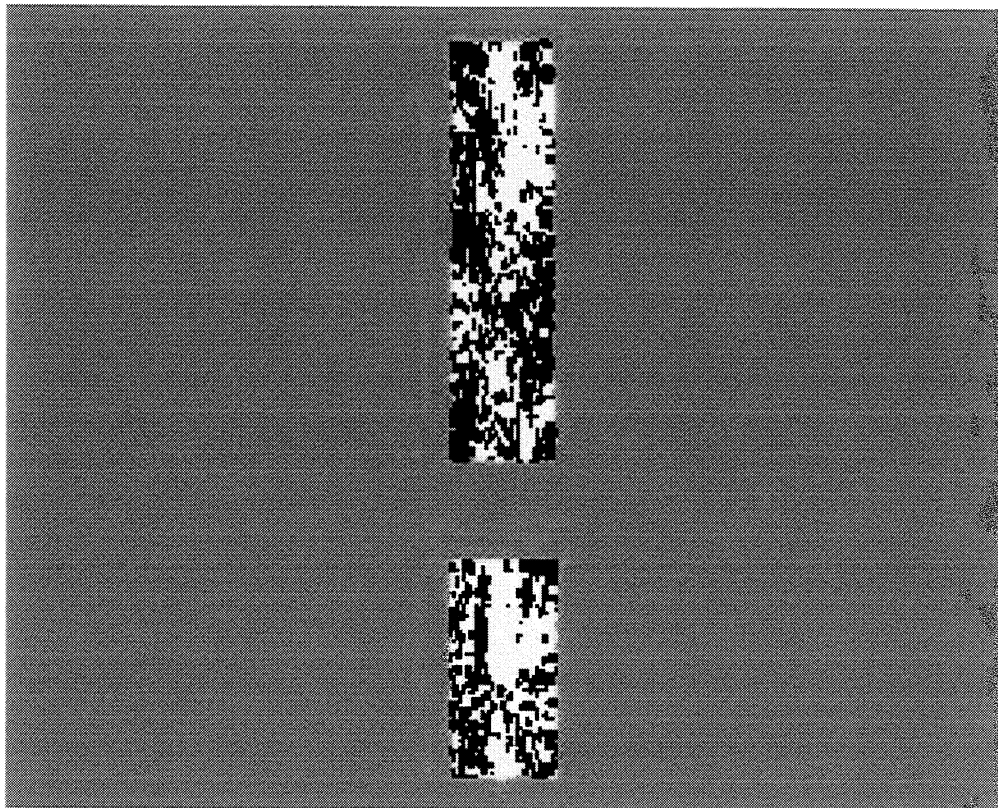
Figure 17. AVL Sample File Expected Output
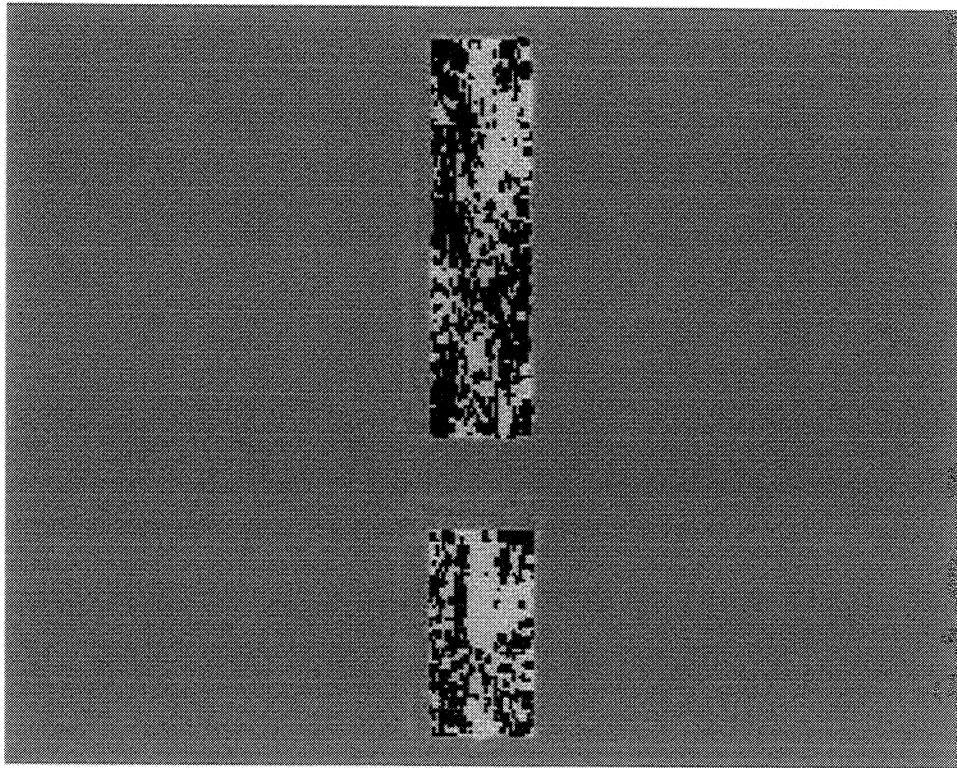
Figure 18. AVL Sample File ANN Output

The grayscale output of Figure 18 shows the different classification of safe and

dangerous elevation points, in this case at about .2 and .8. It also illustrates the distinct

separation of values and the correlation of ANN output to expected output. This can also

be supported by a graphical representation similar to Figure 16 but with the data for

AVL.

Now there is an ANN output of safe and dangerous points, an RMS, and a link of

the terrain to the RMS. Also we have shown statistical metrics of mean error, standard

deviation of the error and percentage of dangerous points within a ROI. There also are

many different aspects of the visualizations of the ROI. All of these elements will be

used to determine whether two separate samples of geographic data are essentially similar

within predetermined tolerances. Additionally these metrics will be used to draw some

relationships between the geographic characteristics and associated errors between databases.

A complete listing of pertinent results for all the airports studied is illustrated next. The data will be broken down into two distinct groups. The first group represents airports with data covering an area up to 60 by 60 statute miles. Again some of the ROI had voids due to a lack of data but this is not a factor as explained previously. The airport data files of this group were very large and time consuming to obtain and manipulate for processing by the ANN. The ANN outputs for the airports of group 1 were for random samples taken from the complete data file. The second group of airports is composed of data that covers only half the area as listed in the first group. The second group was collected, as such, to enable the acquisition of a larger set of different airports and also to allow the complete processing of the airport data file by an ANN.

Airport	Mean (m)	Standard Deviation	Percentage of Danger Elevations	RMS	ANN Output Classes
1S4	11.732	34.268	23.38	.3557	0.8,0.2
AVL	39.771	39.405	48.67	.2707	0.8,0.2
CRQ	30.069	34.889	39.75	.1793	1,0
DEN	6.34	10.36	2.73	.2101	1,0
DTA	16.313	28.342	16.07	.3531	0.8,0.2
HII	23.514	32.489	25.23	.4598	0.8,0.2
PSP	57.608	50.920	64.30	.3548	0.8,0.2
SNA	23.963	35.272	30.43	.4445	0.8,0.2

Table 13 Group One Summary of Metrics

Airport	Count of Elevation Postings	Qualitative Visual Description	Visual Representation
1S4	46051	Elevation from 0-1363m. One mountain with possible river btwn rough terrain.	Appendix 1& Appendix 4
AVL	44499	Elevation from 91-2039m. 2 mtns and dominated by rough terrain.	Appendix 1& Appendix 4
CRQ	34577	Elevation from sea level to 1873m. Coastal. One isolated mountain	Appendix 1& Appendix 4
DEN	39398	Elevation from 1357-2283m. Light slope to beginnings of mountain area.	Figs. 2,3,4,5, 13 & Appendix 4
DTA	42819	Elevation from 1380-3083m. Flat up to a mountain ridge.	Appendix 1& Appendix 4
HII	45174	Elevation from 108-2317m. About 5 scattered mtns. Valleys in between.	Appendix 1& Appendix 4
PSP	46062	Elevation from −70 to 3506m. Two mountain ridges with a valley between.	Appendix 1& Appendix 4
SNA	30273	Elevation from sea level to 1735m. Coastal. Steep climb to mountain ridge.	Appendix 1& Appendix 4

Table 14. Group One Summary of Metrics

Airport	Mean (m)	Standard Deviation	Percentage of Danger Elevations	RMS	ANN Output Classes
48V	13.28	26.86	13.91	.3521	0.8,0.2
ASE	91.90	54.89	89.63	.7627	0.8,0.2
COS	19.44	29.29	17.41	.3810	0.8,0.2
EGE	65.25	45.29	76.84	.7094	0.8,0.2
JAC	70.73	56.35	75.14	.7012	0.8,0.2
LGU	40.20	45.59	44.11	.5534	0.8,0.2
MSO	65.58	48.15	73.11	.6931	0.8,0.2
SLC	57.59	54.41	51.16	.5899	0.8,0.2
FLG	25.20	77.04	25.53	.4430	0.8,0.2

Table 15. Group Two Summary of Metrics

Airport	Count of Elevation Postings	Qualitative Visual Description	Visual Representation
48V	22443	Elevation from1426-2980m. Flat to gradual slope up to steep mountain ridge	Appendix 4
ASE	22246	Elevation from 1950-4340m. Most elevation >3000m. Dominated by mtns.	Appendix 4
COS	22317	Elevation from 1531-3771m. Flat terrain to high mountain range.	Appendix 4
EGE	22201	Elevation from1855-3730m. Very rough mountains of varying degrees of slope.	Appendix 4
JAC	22956	Elevation from 1778-4199m. Rough and mountainous.	Appendix 4
LGU	22240	Elevation from 1283-3038m. Rough and mountainous around a plateau.	Appendix 4
MSO	22373	Elevation from 851-2792m. Rough and mountainous.	Appendix 4
SLC	16827	Elevation from 1280-3501m. Half flat due to lake progressing to steep mountain range.	Appendix 4
FLG	41131	Elevation from996-4399m. One high mountain. Few other rough areas.	Appendix 4

Table 16. Group Two Summary of Metrics

As Denver International has been introduced in previous discussions first, we'll start with it here. Statistical analysis, in Table 7, shows that only 1074 of 39398 points in the ASMD are dangerous when compared to the USGS data. Ninety-seven percent of the ASMD longitude, latitude, elevation triples are within a +30 meter tolerance. Also exhibited in Table 6, the average ASMD error of the compared elevations is only 6 meters and the standard deviation is only 10 meters. Examining visualizations, Figure 3 is a 2-D view of USGS longitude, latitude, elevation triples overlaid with ASMD error values. Only a small number of errors above our threshold are shown. The error values within ±30 meters are displayed as transparent to demonstrate good areas of elevation values. Additionally, the points outside of the −30m error value are plotted and these are easily evident in this perspective. Figure 4 shows a 3-D TIN of the error between the ASMD and the USGS database. When the data is presented as a TIN the areas of missing data from previous views are masked. These just need to be remembered but they don't affect the analysis and as shown can be worked around with differing views. The dominance of the light blue shading throughout supports the conclusion of low error between the ASMD and USGS. The ANN output is absolute between the dangerous and safe elevation values, classifies them as (1,0) and has a low RMS of .2101. The criteria for an acceptable match between the two databases is outlined as mean error <30m, error standard deviation <18m and dangerous elevation coverage <10%. ANN support is offered by a low RMS, an absolute distinction between dangerous and safe values and classification of these values at 1 and 0. The data for DEN positively supports that the ASMD for this airport is a safe representation of actual conditions within the prescribed

tolerance. Denver International airport is the only accepted match between the ASMD and USGS when verified by all methods employed.

Also with respect to the RMS from Figure 13 and Figure 51, a correlation can be shown between the slope of the terrain and the elevation errors. A low RMS is linked to the large space between the contour lines and the sparseness of red shading in the visualization.

The rest of the discussion will illustrate characteristics or factors that would support a rejection of the match between the ASMD and USGS data for each airport. The next airport of interest is McClellan-Palomar (CRQ). It stands out as a candidate for a match because the ANN model correctly classified the safe and dangerous values, predicted them at 1,0 and the RMS for the model was low. This alone is not enough to accept the ASMD data. All the other metrics for this airport are above acceptance criteria.. An explanation can be found in the visualization of Figure 50 and the qualitative description in Table 14 of McClellan-Palomar. Depicted is an area moderately sloped with an isolated mountain. Remember that CRQ is from the Group 1 airports where the ANN model was tested on a random sample of the data in the ROI. This just happens to be a case where the random sample was not representative. The sample data covers a good-sized area but just a small portion of it contains that representative of the isolated mountain where the terrain is steeply sloped. It has been concluded and supported that the RMS is greatly affected by sloping terrain. The lack of inclusion of it in the ANN model for this airport influenced the unsupportive ANN output. This issue will be dealt with later with the Group 2 airports when the ANN model will process the entire ROI. This factor just surfaced on this individual airport and doesn't seem to be a factor on the

other 7 airports in Group 1. The visualization aspect was a strong tool when the disagreement between the ANN output and the statistical output was evident. Utilizing the acceptance criteria listed, none of the rest of the airports in Group 1 can be considered a match between the ASMD and USGS data. ??

The rest of the discussion will focus on the Group 2 airports. These airports are much more representative of terrain impacted areas and contain more slope/steepness and general surface roughness. An obvious characteristic of all these airports' results is that the ANN RMS' are all above .3500 and none of the outputs could be classified to the design criteria with dangerous and safe values of 1 and 0. However, the ANN model did correctly and absolutely distinguish safe and dangerous elevations. The model has been successful and unfailing here. A look at the graphics in Appendix 4 supports the fact that the RMS of the ANN is related to the steepness of the terrain. The magnitude of the RMS coincides with the slope of the terrain and with its coverage. The significance of the ANN's ability to illustrate the measure of slope and the elevation errors related to the sloping topography is important in concluding or rejecting a match between terrain databases. Without this insight a possible errant determination could be made. For instance two possible examples where an inaccurate conclusion could be formed are exhibited in Tri-County (48V) airport and City of Colorado Springs (COS) airport. If the only metric utilized was that chosen in the preliminary research, you would be led to believe that for these two airports the ASMD and USGS are similar. The output of the ANN contradicts this assumption and further evidence is gained from the visual representations in Appendix 4. The areas are dominated by a flat or gradually sloping landscape and then move to a prominent steep mountain range. These mountain ranges

have a noteworthy amount of error associated with them and their disregard is perilous. So far we have discussed a good match of databases and some questionable and inconsistent evidence. A solid illustration of a bad match is Aspen-Pitkin County/Sardy field (ASE). The ANN distinguished between safe and dangerous values but had a very high RMS and the predicted output was in the neighborhood of .2 and .8. This would lead to a rejection of matching databases. This is further enhanced by the statistics and the visualization of a predominantly red ROI indicating error greater than the threshold over a large portion of the data that is representative of very rough and sloping mountainous terrain.

To cover a more complete spectrum in this research, included in the Group 2 airports is a ROI of 60 by 60 statute miles that was processed with the valid ANN model. Flagstaff-Pulliam (FLG) airport was the experiment. The RMS of the ANN suggests that the databases don't match. From this number it may be hypothesized that there are some features with elevation error related to sharp sloping topography. The area covered by the ROI is not completely rough or mountainous but does have some elements of moderate slope and not vast areas of flat or gradual landscape. It may not be hazardous terrain throughout but there are prominent features that elevate the danger factor for this area.

This airport illustrates an added benefit and a byproduct of the research. A look at the statistics shows a very high standard deviation for the errors. Also a histogram of the data shows a decent number of very large outliers. A two-dimensional and three-dimensional view of the error shows a huge error in a distinct area of the ROI. This was

explored and a discrepancy was found in the USGS elevation data. This instance was not just specific to FLG and some irregularities were discovered in CRQ and AVL.

Additionally the visualization aspect can be used to support the research of Phase II. Visualizations can be constructed illustrating actual USGS elevations overlaid on the ASMD elevations (e.g., Fig 5). The intent of such a graphic is to show where the ASMD elevations are greater than the USGS elevations. Red tinted areas can show USGS elevations greater than ASMD elevations and are classified as risky. This does not by itself identify the airport ROI out of a +30m tolerance as investigated in Phase III, but it can give insight into the ASMD being within a threshold. If the area is predominantly colored with the terrain elevation color scale then the ASMD error will be negative and obviously under a positive threshold. The fact that a predominantly red shaded graphic is not indicative of a bad ASMD to USGS comparison is illustrated well by the analysis of DEN (Fig. 5) which is mostly red, but the data shows 97% of the points are under the +30 threshold and the mean ASMD error is only 6 meters.

5.Conclusions

Artificial neural networks have arisen from analogies with models of the way that human might approach pattern recognition tasks, although they have developed a long way from the biological roots. Great claims have been made for these procedures, and although few of these claims have withstood careful scrutiny, neural network methods have had great impact on pattern recognition practice. A theoretical understanding of how they work is still under construction. (Ripley, 1995).

Artificial neural networks proved viable in pattern recognition in terrain databases. The ANN models investigated were able to discern areas of acceptable data and areas of

unacceptable hazardous data. Additionally, they were able to classify error between two similar databases as safe or dangerous with respect to a chosen criteria or error threshold. The strength of the classification seemed to be linked to certain aspects of the terrain. The ability of the ANN to offer this underlying indication was noteworthy. This indicator brought light to weaknesses of other methods and metrics used in comparison of terrain data. A substantial enhancement to the use of ANNs was a visualization aspect. It complimented the results of the ANN output and allowed the discovery of a link between the strength of the ANN classification and the degree of slope within the chosen area of interest. Furthermore, the visualizations also added support to the criticism of statistical metrics. An elevated predominantly flat terrain with either an isolated mountain or a ridgeline covering a minor percentage of the area of interest could give a low average error for a relatively large area. The steep mountainous area, although not dominant, is significant. The large number of low errors offset the small concentration of large errors outputting a low average that could mask the hazard area and lead to the belief that the two databases are similar within tolerances. A picture of the errors reveals the large concentration of high errors that are very dangerous in their part of the area of interest. The ANN also alluded to this with its output of an RMS value greater than .2 and an inability to classify points at absolute values of zero or one.

There is a considerable overlap between the fields of neural networks and statistics. While neural networks are often defined in terms their algorithms or implementations, statistical models are usually defined in terms of their results. There are many ideas in research literature in order to explain why to use ANNs rather than any of the statistical methods. Depending on the answer sought, some popular ANN models is useful in the

same situation as a regression, where: the number of inputs is fairly large, many of inputs are relevant, but most of the predictive information lies in a low-dimensional subspace.

Other findings exhibited in the research investigation pertain to data density reduction and sampling of the region of interest. In one instance a random sample from the ROI was not a true indicator of the entire ROI. The random sample was not representative and gave false results. It is imperative that the sample be representative. The random choice of points may exclude major terrain features that are a threat to aircraft safety. The data density was dictated by the ASMD. This was not nearly as good as the 10 m spacing or even the 30 m spacing of the USGS. This could have been a factor in many of the mismatches. An added benefit of the visualization aspect of the research is the discovery of outliers and error in the underlying data used as the standard. This was demonstrated in isolated areas in 4 of the examined airports.

6.Further Research and Discussions

Using 17 sets of airport data that aids visualization, our results may help to explain if the data is suitable to use for cockpit simulation on particular airports. As mentioned earlier, some of the airports yielded high RMS results.

The sample size and sampling areas are also other factors that might influence on the results. A research question that can be further explored is whether training and testing sets should be chosen by randomly or by some systematic algorithm. Those two factors can be explored for the further research. If it is then we go back the question ". Does training and testing sets should be chosen by randomly or by some systematic algorithm? and suggest some further research on it. Incorporation of ANN in a larger database accuracy methodology.

A larger research question deals with the use of ANN for a more comprehensive role in terrain database correlation analysis methodologies. In this project, ANN models did show promise for simulating aggregate statistical analyses, which are part of an overall assessment of database correlation. Of even greater importance is the role the human analyst provides in terrain database correlation judgments, which deal with the cognitive synthesis of statistical measures, as well as analyzing comparative visualizations of aggregate and local database areas. By capturing such analyst processes, and feeding such human database pattern recognition changes and conclusions into ANN models, assessments can be made of ANN for a role in helping to higher-level cognitive recommendations to analysts, based on previous similar analysis procedures.

7. References.

Antognetti, Paolo; Milutinovic Veljko, 1991. *Neural Networks, Concepts, Applications, and Implementations :Volume IV.* New jersey : Prentice Hall

German, Gordon,1999. *Neural Network Classifiers for GIS data: Improved search strategies.* GeoComputation 99.

Hunter, Gary J. and Goodchild, Michael F. *Dealing with Error in Spatial Databases: A Simple Case Study:* Photogrammetric Engineering &Remote Sensing V.61, No.5, May 1995, p.p. 529-537.

Kadaba ,Nagesh ; Nygard, Kendall ; Juell , Paul L.; Kangas, Lars J., 1989. *Modular Back-Propagation Neural Networks for Large Domain Pattern Classification.* International Joint Conference on Neural Networks Conference proceedings Volume II pg.607.

Lawrence, Steve; Giles, Lee; Tsoi, Chung Ah ,1996. *What size neural Network Gives Optimal Generalization? Convergence Properties of Backpropagation*(Report No. Lisboa P.J.G.; Taylor M.J.,1993. *Techniques and Applications of neural Networks.* England : Ellis Horwood

Looney, Carl G., 1997. *Pattern recognition Using Neural Networks:Theory and Algorithms for Engineers and Scientists.* New York: Oxford University Press

McCord N.M.;Illingworth W.T.1991. *A Practical Guide to Neural Nets. Massachusetts :* Addison-Wesley Publishing Company, Inc.

Pariente, Dillon ,1994. *Geographic Interpolation and Extrapolation by Means of Neural Networks.* INSA de Lyon, France. EGIS Foundation

Principe C. Jose; Euliano R. Neil, Lefebvre W. Curt , 1999. *Neural and Adaptive Systems :Fundamentals through Simulation.* New York: John Wiley & Sons, Inc.

Ripley, B.D.1996. *Pattern Recognition and Neural Networks.* New York: Cambridge University Press

Schalkoff, Robert J.,1992. *Pattern Recognition: Statistical, Structural and Neural Approaches.* New York:John Wiley & Sons, Inc.

UMIACS-TR-96-22 and CS-TR-3617). College Park, Maryland, Institute for Advanced Computer Studies, University of Maryland.

Appendix 1.

Data acquisition of United States Geologic Survey (USGS) 7.5' Digital Elevation Model (DEM) data.

A checklist has been developed for the construction of a region of interest around a desired airport of geographic point pairs of elevation values from United States Geologic Survey (USGS) data and Airport Safety Modeling Data (ASMD).

Data Acquisition

1. Airport Selection
2. Airport Reference Point
3. Determination of Region of Interest (ROI) and calculation of edge boundaries for ROI
4. Determination of central DEM and county in which airport is located
5. Production of a template of inclusive USGS DEMS
6. Download of data from the web
7. Determination of other counties represented in DEM template, if necessary
8. Review associated .TXT files of downloaded DEMs
9. Unzip files of each DEM
10. Merge compatible 1:24000 7.5' DEMs utilizing MICRODEM ©
11. View merged DEMs for quality assurance
12. Export merged DEMs into files of xyz triples
13. Merge and format conversion of xyz triples into a size and format useable within Visual Basic © data point matching and error computation program.
14. Import output of Visual Basic © program into Excel © for statistical analysis

Amplifying information for each checklist item above follows:

1. Chosen from the list of ASMD airports contained in the US_INDEX file of the original data.

2. Located in the header of the ASMD airport data file.

3. Can be taken from the expanded index document included in this Appendix or calculated if other than 30 nm. The determination of the edge boundaries does not have to be exact but must be large enough to contain the entire ROI. This can also be estimated in the completion of step 5.

4. This information can be found through the web site-
http://edcwww.cr.usgs.gov/Webglis/glisbin/finder_map.pl?dataset_name=MAPS_LARGE.

5. This is accomplished through the following web site-
http://geonames.usgs.gov/pis/gnis/web_query.gnis_web_query_form.
Construct a template similar to that of Figure 1 until all DEMs have been filled in to cover the area depicted by the ROI.

6. This is completed through the web site-
www.gisdatadepot.com/dem/demdownload.html. The free data are the 1:24,000 7.5' DEMs in SDTS format. They are in either 10m or 30m spacing. Download time for 10m spacing is much greater than that for 30m, i.e. 15 minutes vice 2 minutes with a broadband connection. Also, download the accompanying .txt file.

7. In the construction of the tiled ROI, oftentimes the area covers multiple counties and possibly other states. If you exhaust the DEMs in one county you must return to the website listed in step 4 to discover the other counties required. If a state border is crossed that information can be seen on the map view of the website listed in step 5.

8. A review of the .txt file for each downloaded DEMs is necessary to find out certain information necessary for successful merging later. Items of importance are spacing of 10m or 30m; z value in ft or meters; UTM zone; Data level 1 or 2; Vertical datum and Horizontal datum.

9. Procedures for Unzip of the .tar.gz files are given in a tutorial on the web link ftp://sdts.er.usgs.gov/pub/sdts/articles/ascii/sdts_tutorial.txt Although the website for DEM download has been changed since the production of this tutorial in 1998, it offers pertinent information related to versions of WINZIP When unzipped, 19 files will be extracted. The actual data is contained in the file with the ####celo.ddf extension.

10. MICRODEM © ver 5.1 rev12/11/2001 and ver 5.12 rev 1/7/2002 was used to convert the SDTS format to a MICRODEM format and then to XYZ triples of longitude, latitude, elevation. The data manipulation menu was used with the features of merge and export being the primary functions. The data obtained from the USGS DEM .txt file is necessary here to get a successful merge. The merged DEMs must be in the same UTM zone. Datum transformation and geographic coordinates are accomplished automatically by MICRODEM to that set in the options tab. All the baseline parameters should be set prior to entering the data manipulation menu. Some versions of MICRODEM automatically convert the different z values of feet and meters to all meters. In this project only DEMs of similar spacing, UTM, z, horizontal and vertical datum were merged together. Some DEMs were just merged individually for format conversion. Also, with 10m spacing a limited number of DEMs were merged to keep the files to a workable size after export. Too many 10m DEMs significantly impacted processing time both in MICRODEM and in follow on applications.

11. In MICRODEM, each merged DEM was viewed to ensure a good product before continuing to the follow on export and other processing. The next steps are time consuming and a good product here eliminates lengthy rework.

12. Export of the DEMs into XYZ triples of longitude, latitude, elevation in meters. Often the entire ROI of DEMs could not be exported in one file. It was either too large or had differing parameters as listed in the descriptive text files. When necessary the DEMs were merged and exported by rows. The convention with the lower left corner being the starting point was followed in naming files and building the entire ROI.

13. Two computer programs developed in prior research were used in this step. A Demmerge program is used to consolidate the output XYZ triples files of MICRODEM. This was necessary for those DEMs with differing descriptive information before the merge and export but now all data is of the same format after the export. Also the Demmerge program reformats the data to that required for the next process. Again, it should be noted that the 10 m spacing has an impact on processing time. Also, the size of a merged file could outgrow that capable by the next computer application. Parameters will go out of bounds. The Demmerge program gives an indication of the number of points in the file. Typical size ranges from 4 to 7 million points. 10 million is the limit. When an ROI exceeds 10 million points the next process is just repeated until all the individual pieces of an ROI for a single airport are finished. For a 30 mile ROI with many 10 m DEMs the Demmerge and follow on processing is very lengthy. With a computer with an average speed processor it could take a few days running continuously to complete a single ROI. After the Demmerge program, the next process is pair- wise geographic location matching of elevations from the USGS database and the ASMD database. This program was also developed in prior research. It was modified slightly because of a newer version of MICRODEM. The need for a datum transformation no longer exists with that capability added to MICRODEM. Also some features still in development in the previous research were remarked out to increase processing time.

14. The final step is just import of the raw data output from the Visual Basic programs into Excel spreadsheets for further analysis and use by follow on applications. A sample of the final product is shown below:

LAKE HAVASU CITY
ASMD ROI average elevation is 507.9962013

Counter	asmdlat	asmdlong	asmdelev	asmderror	usgselev	USGS posts
1	34.1792	114.8017	630	94	724	302
2	34.1833	114.8017	758	46	804	300
3	34.2	114.8017	822	28	850	282

Homer Mtn	East of Homer Mtn	Mt Manchester	Davis Dam SE	Oatman	Mount Nutt	Kingman SW	Kingman SE	Hualapai Peak	Dean Peak	Bottleneck Wash
Homer	Bannock	Needles NW	Needles NE	Boundray Cove	Warm Springs	Yucca NW	Yucca NE	Wabayuma Peak	Hibernia Peak	Pilgram Wash
West of Flat Top Mtn	Flat Top Mtn	Needles SW	Needles	Warm Springs SW	Warm Springs SE	Yucca NW	Yucca SE	Creamery Canyon	Diamond Joe Peak	Gunsight Canyon
Step Ladder Mts NW	Step Ladder Mts NE	Monumental Pass	Whale Mtn	Topock	Franconia	Buck Mtns	Buck Mtns NE	Dutch Flat	Beecher Canyon	Wikieup NW
Step Ladder Mts SW	Step Ladder Mts NE	Snaggletooth	Chemehuevi Peak	Castle Rock	Lake Havasu City North	Crossman Peak	Buck Mtns SE	Ducth Flat SW	Dutch Flat SE	Groom Spring
West of Mohawk Spring	Mohawk Spring	Savahia Peak NW	Savahia Peak NE	Havasu Lake	Lake Havasu City South	Standard Wash	Mohave Springs	Castaneda Hills	McCraken Peak	Signal
Martins Well	Mopah Peaks	Savahia Peak SW	Savahia Peak	Whipple Mtns SW	Whipple Wash	Gene Wash	Monkeys Head	Castaneda Hill SW	Centennial Wash	Rawhide Wash
Sablon	Horn Spring	Vidal NW	Vidal Junction	Parker NW	Parker	Cross Roads	Osborne Well	Planet	Swansea	Reid Valley
Arica Mtns	Rice	Grommet	Vidal	Parker SW	Parker SE	Black Peak	Bobs Well	Powerline Well	Butler Pass	Butler Well

Figure 19. Dem Template for Lake Havasu City ROI

Appendix 2

Select Visualizations for Sample Airports

1S4 SCAPPOOSE INDUSTRIAL AIRPARK

AVL ASHEVILLE REGIONAL

CRQ MC CLELLAN-PALOMAR

DTA DELTA MUNI

HII LAKE HAVASU CITY

PSP PALM SPRINGS REGIONAL

SNA JOHN WAYNE AIRPORT-ORANGE COUNTY

1S4   SCAPPOOSE INDUSTRIAL AIRPARK

Color	Elevation (m)
	1205-1356
	1055-1205
	904-1055
	753-904
	603-753
	452-603
	301-452
	151-301
	0-151

2-Dimensional Representation of 1S4 from ASMD

3-Dimensional TIN of 1S4 from ASMD
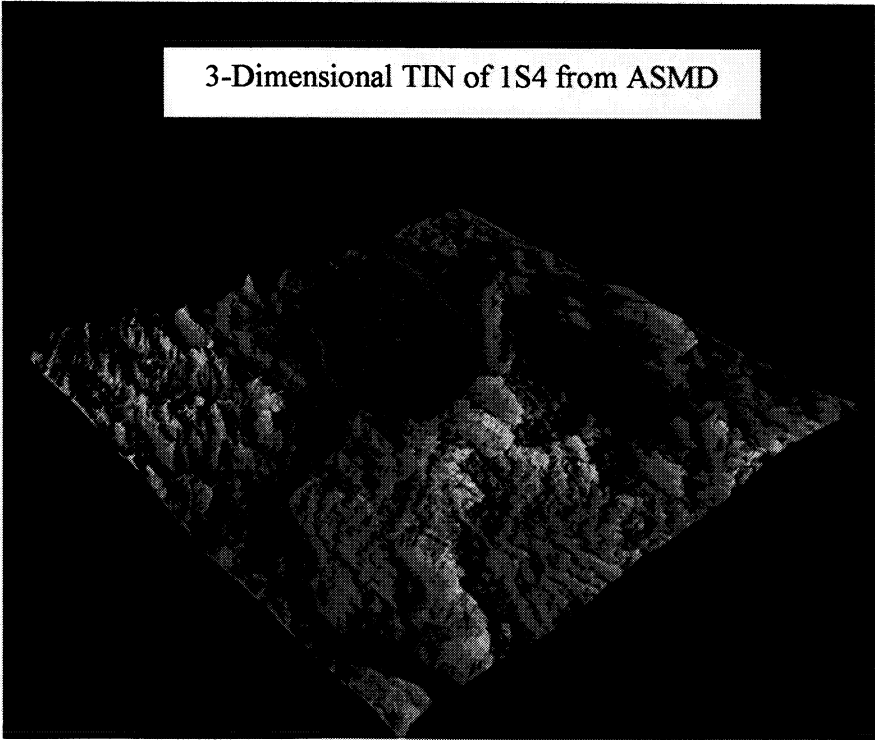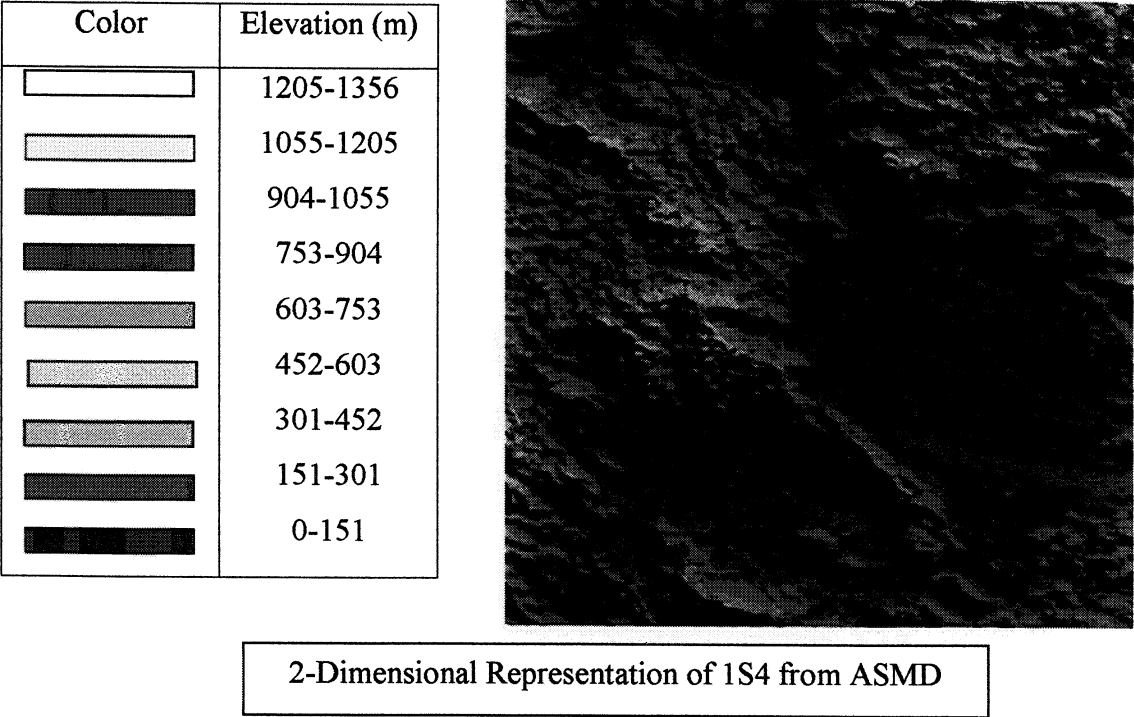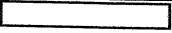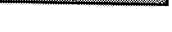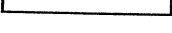
Figure 20. Visualizations of 1S4 from ASMD
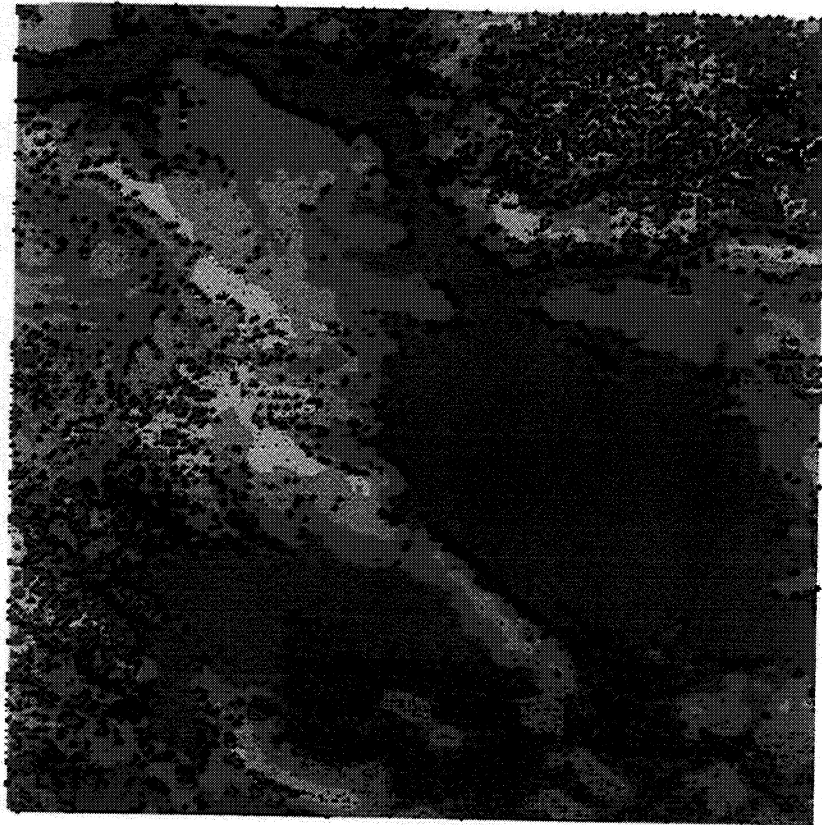
2-D view of ASMD Error v. USGS Elevation

Color	Elevation (m)
	1212-1363
	1060-1212
	909-1060
	757-909
	606-757
	454-606
	303-454
	151-303
	0-151

Color	Elevation (m)
	150-744
	50-150
	30-50
	-30 - 30*
	-43 - -30

*This color band is transparent to so that values within limits show through.

Figure 21. Visualizations of 1S4 from USGS Data and ASMD Error

58

3-D TIN of ASMD Error
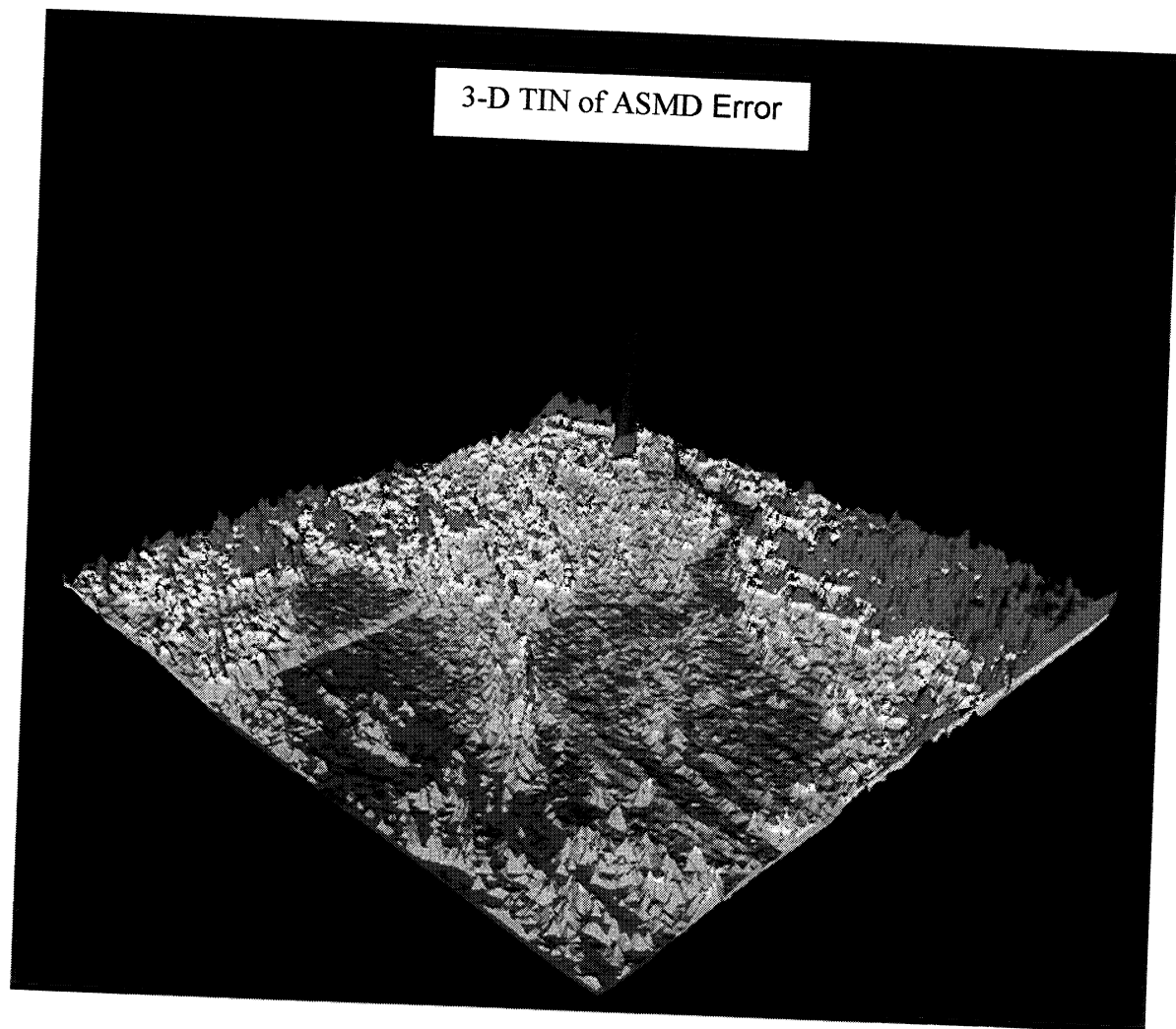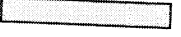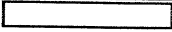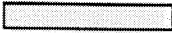
Color	Elevation (m)
	150-744
	50-150
	30-50
	-30 - 30
	-43 - -30

Figure 22. Visualizations of 1S4 from ASMD Error Data

AVL ASHEVILLE REGIONAL

Color	Elevation (m)
	1821-2018
	1624-1821
	1427-1624
	1230-1427
	1033-1230
	836-1033
	639-836
	442-639
	245-442
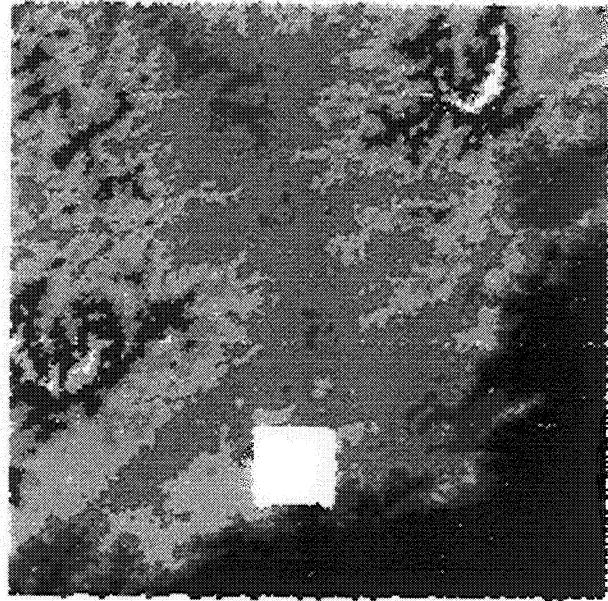
2-Dimensional Representation of AVL from ASMD

3-Dimensional TIN of AVL from ASMD
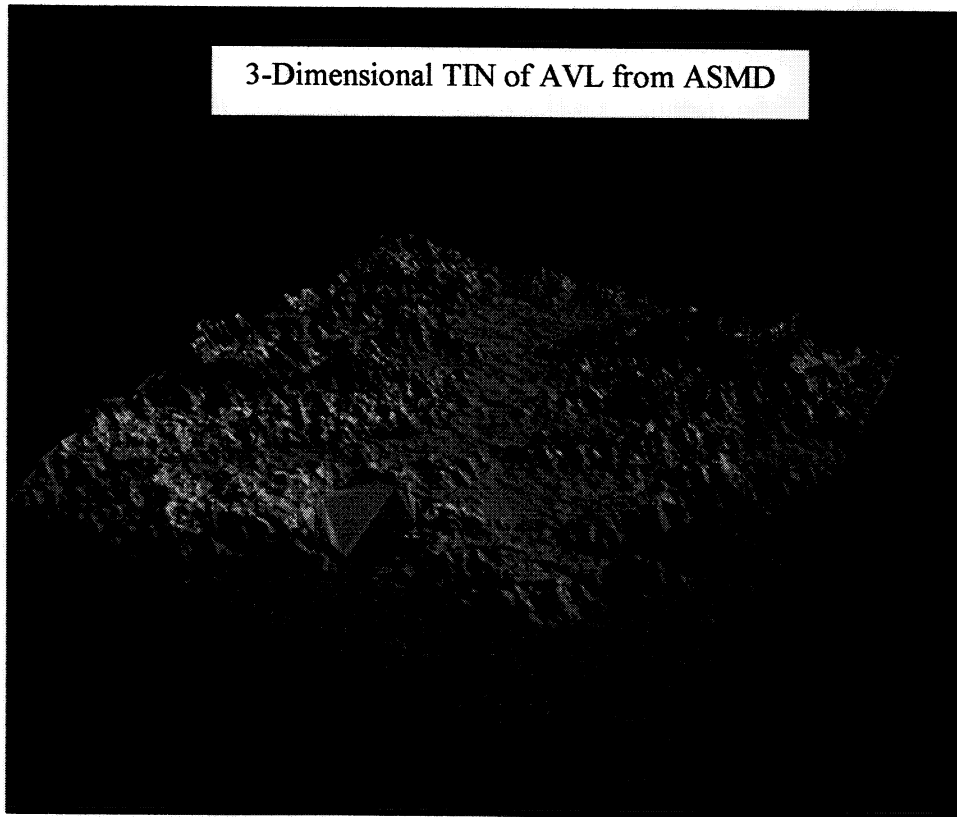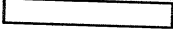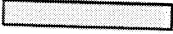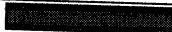
Figure 23. Visualizations of AVL from ASMD
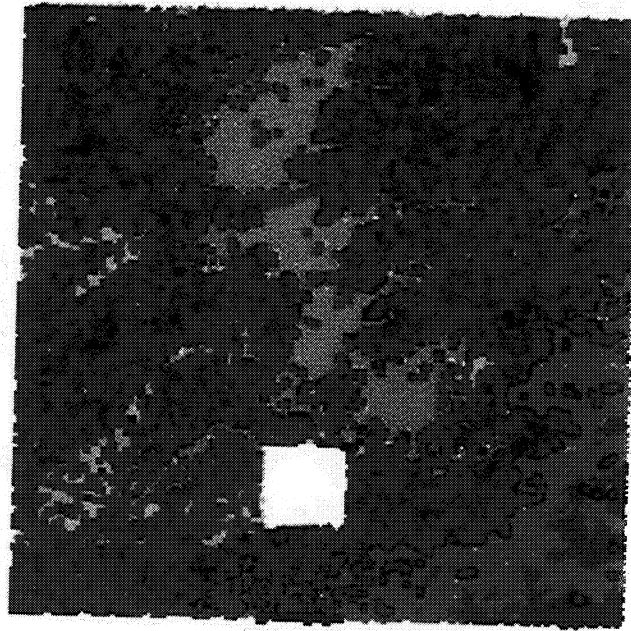
Color	Elevation (m)
	1823-2039
	1606-1823
	1390-1606
	1173-1390
	957-1173
	740-957
	524-740
	307-524
	91-307

Color	Elevation (m)
	150-274
	50-150
	30-50
	-30-30*
	-266 - -30

*This color band is transparent to so that values within limits show through.

Figure 24. Visualizations of AVL from USGS Data and ASMD Error
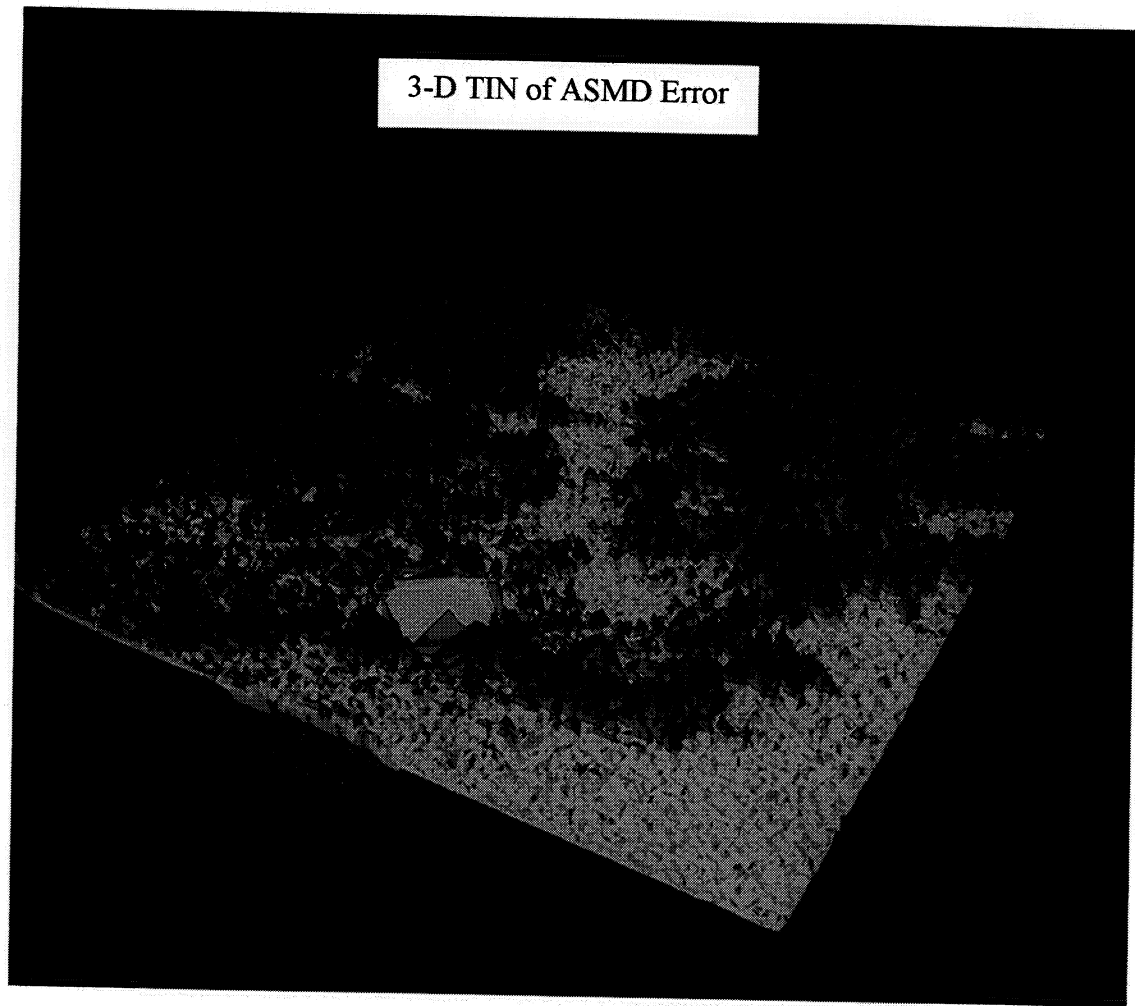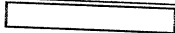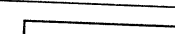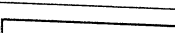
Color	Elevation (m)
	150-274
	50-150
	30-50
	-30-30
	-266 - -30

Figure 25. Visualizations of AVL from ASMD Error Data

Figure 26. Visualizations of AVL from ASMD and USGS Data

CRQ MC CLELLAN-PALOMAR

Color	Elevation (m)
	1641-1846
	1436-1641
	1231-1436
	1026-1231
	820-1026
	615-820
	410-615
	205-410
	0-205
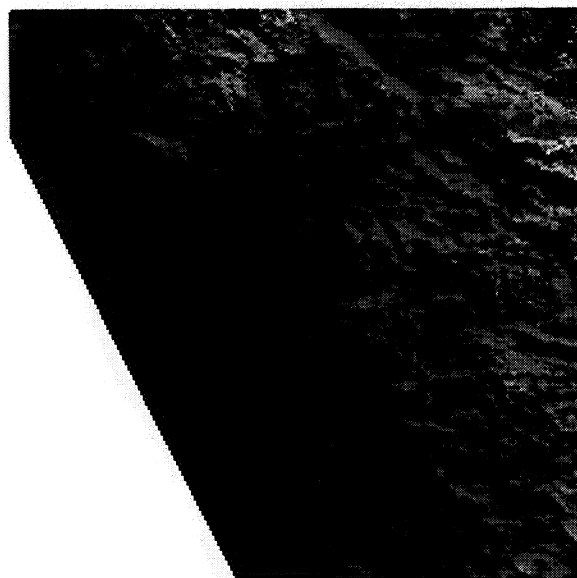
2-Dimensional Representation of CRQ from ASMD

3-Dimensional TIN of CRQ from ASMD
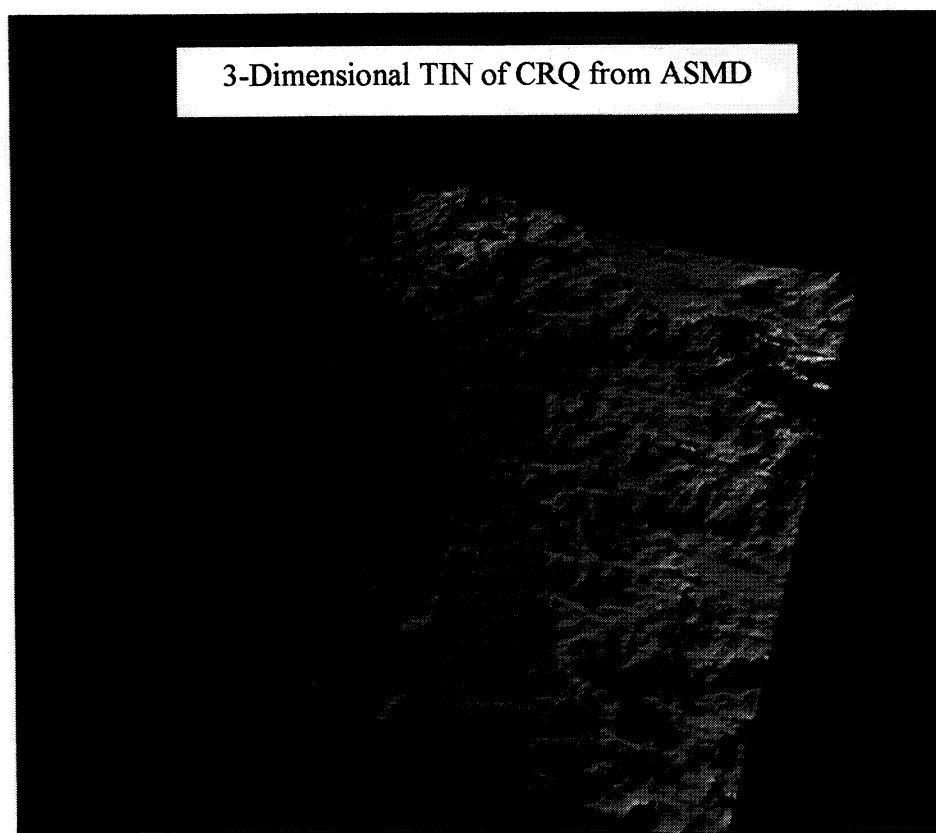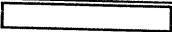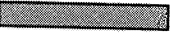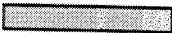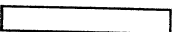
Figure 27. Visualizations of CRQ from ASMD
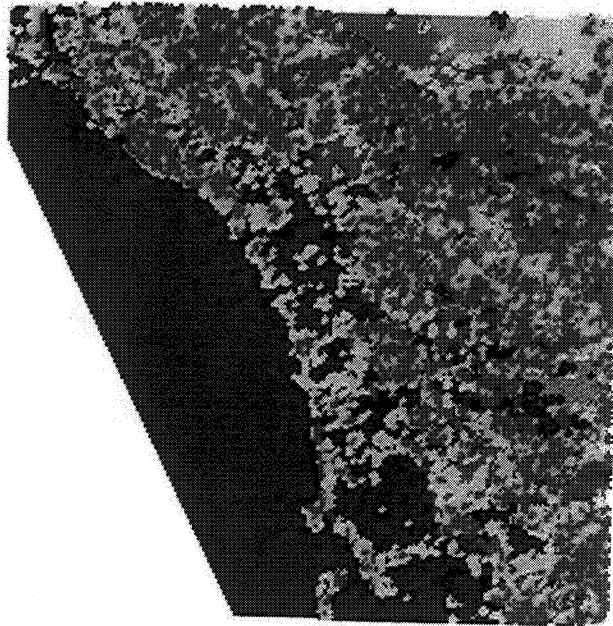
2-D view of ASMD Error v. USGS Elevation

Color	Elevation (m)
	1665-1873
	1457-1665
	1249-1457
	1041-1249
	832-1041
	624-832
	416-624
	208-416
	0-208

Color	Elevation (m)
	150-608
	50-150
	30-50
	-30 - 30*
	-154 - -30

* This color band is transparent to so that values within limits show through.

Figure 28. Visualizations of CRQ from USGS Data and ASMD Error
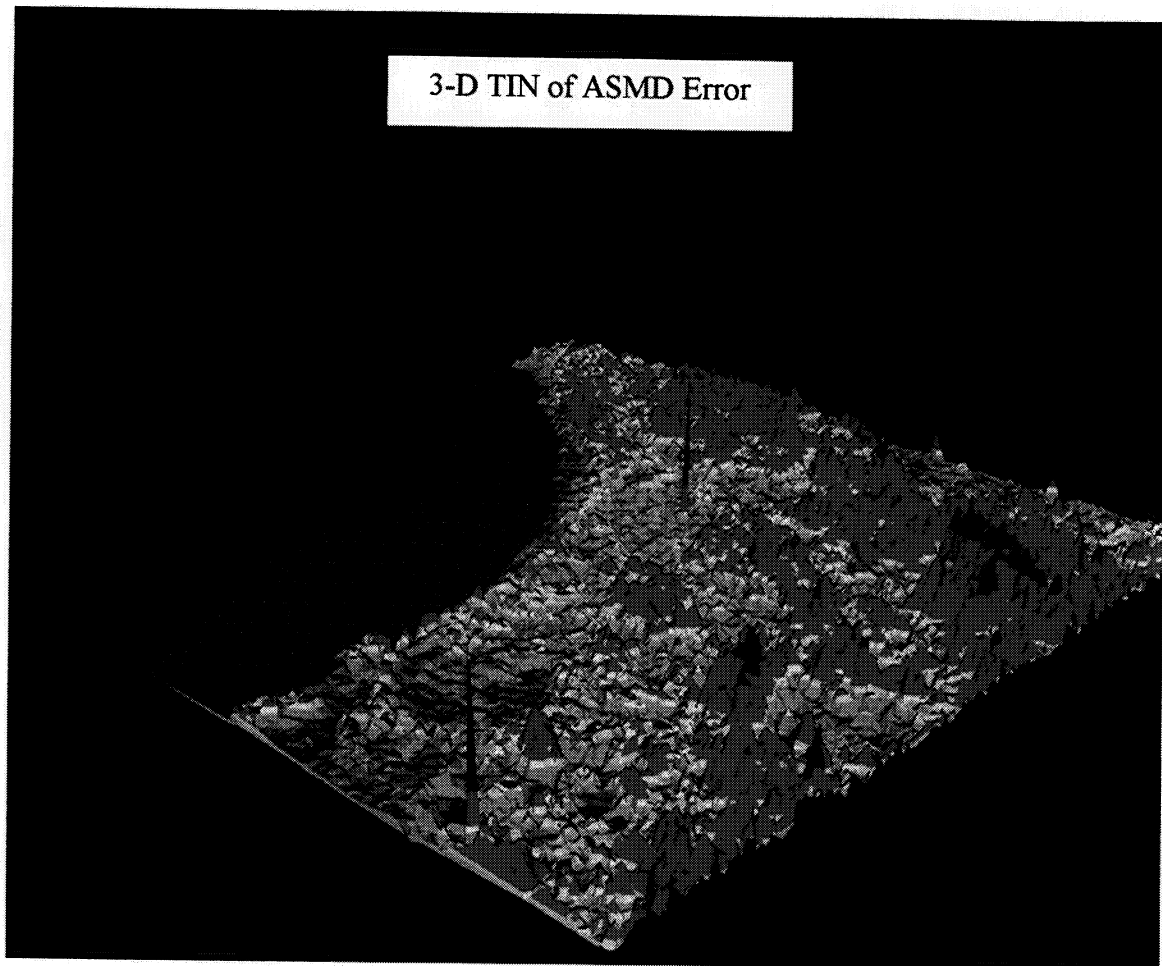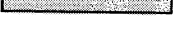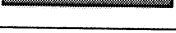
Color	Elevation (m)
	150-608
	50-150
	30-50
	-30-30
	-154 - -30

Figure 29.Visualizations of CRQ from ASMD Error Data
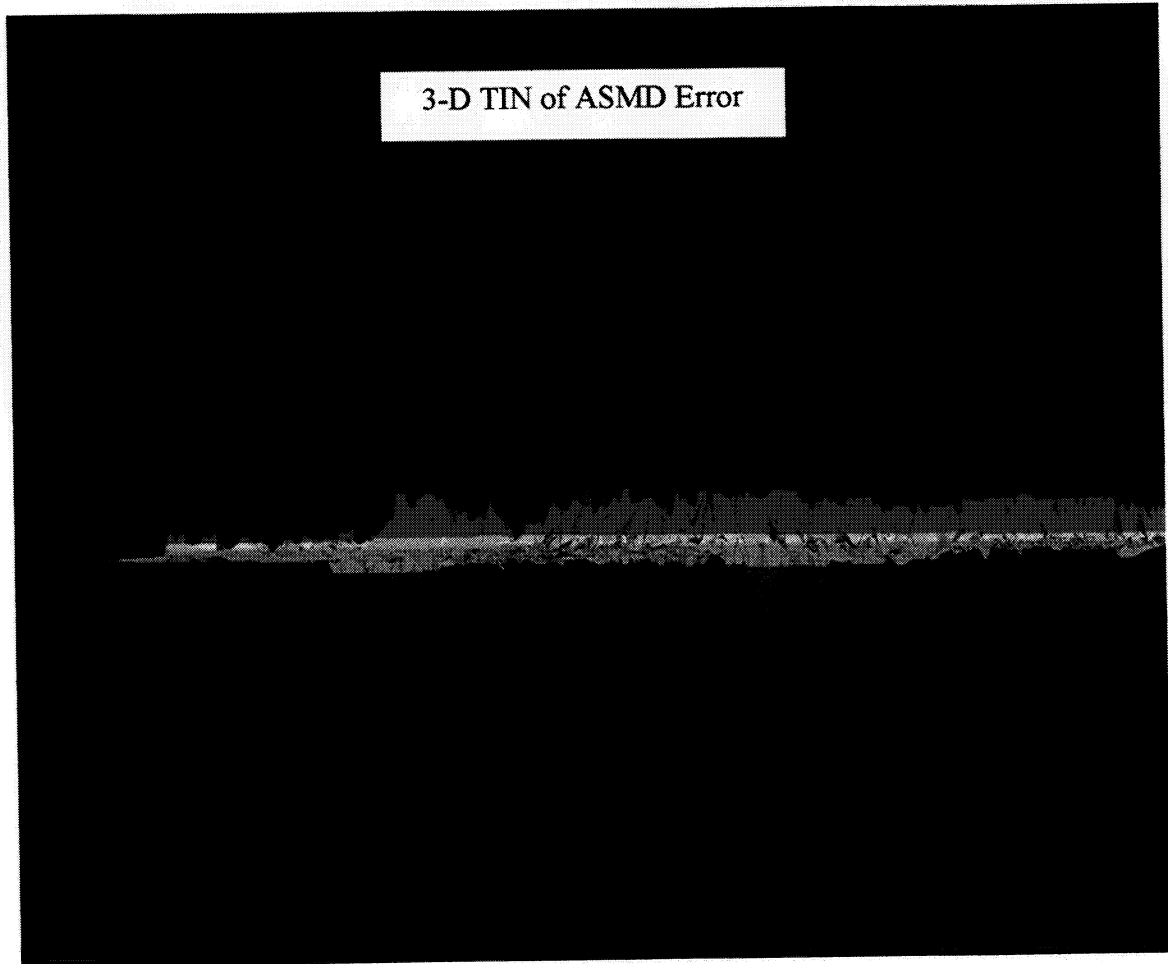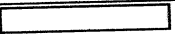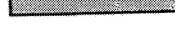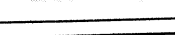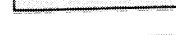
Color	Elevation (m)
	150-608
	50-150
	30-50
	-30-30
	-154 - -30

Figure 30.Visualizations of CRQ from ASMD Error Data

Figure 31.Visualizations of CRQ from ASMD and USGS Data

DTA DELTA MUNI

Color	Elevation (m)
	2864-3050
	2678-2864
	2492-2678
	2306-2492
	2119-2306
	1933-2119
	1747-1933
	1561-1747
	1375-1561
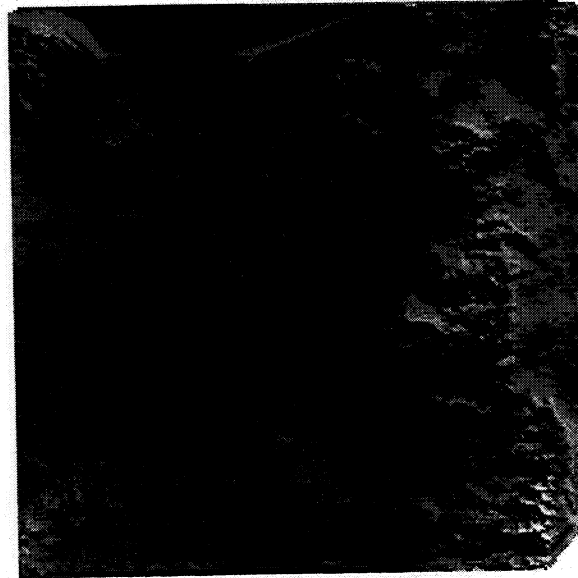
2-Dimensional Representation of DTA from ASMD

3-Dimensional TIN of DTA from ASMD
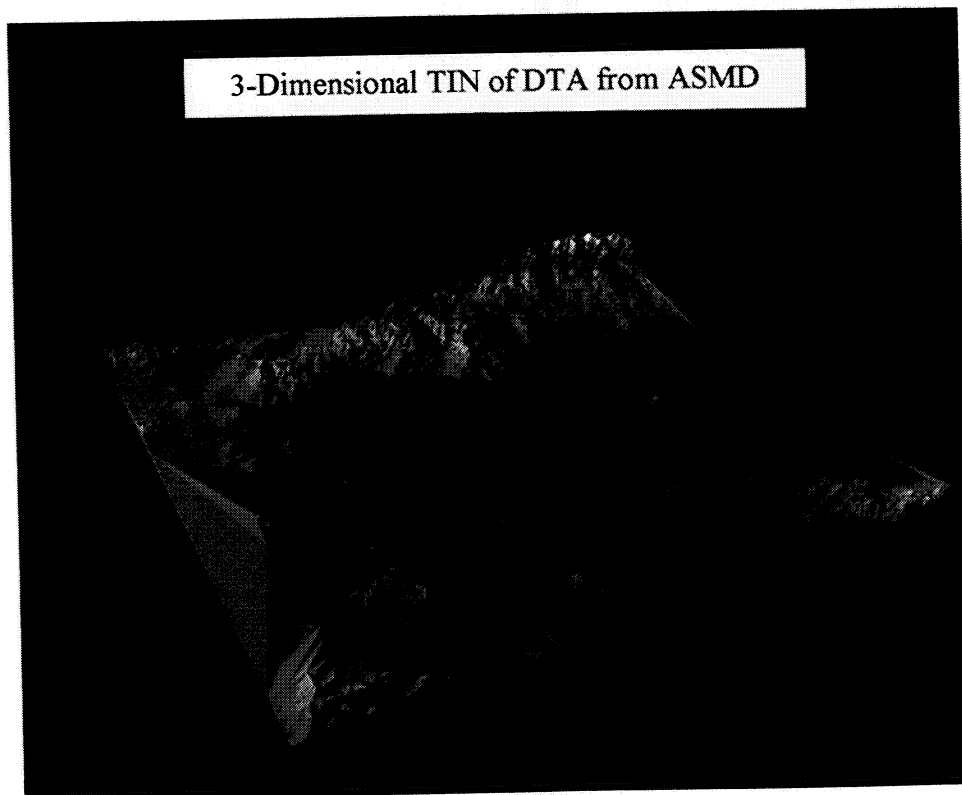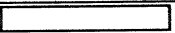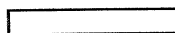
Figure 32.Visualizations of DTA from ASMD
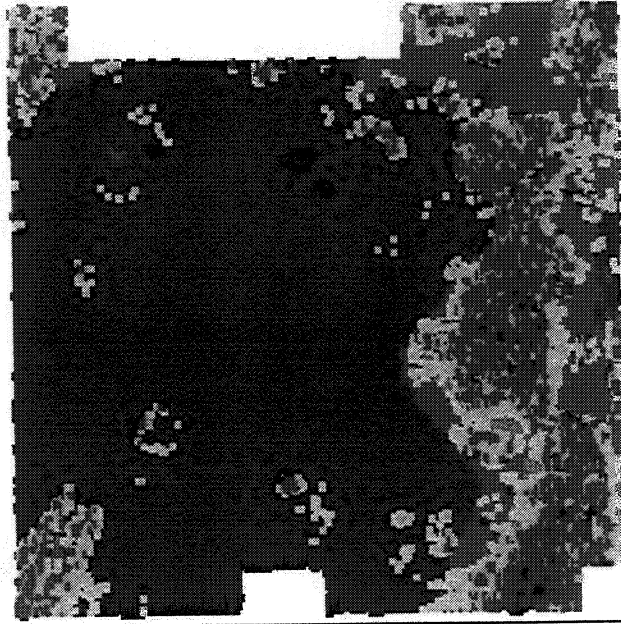
2-D View of ASMD Error v. USGS Elevation

Color	Elevation (m)
	2894-3083
	2705-2894
	2515-2705
	2326-2515
	2137-2326
	1948-2137
	1758-1948
	1569-1758
	1380-1569

Color	Elevation (m)
	150-342
	50-150
	30-50
	-30 - 30*
	-49 - -30

* This color band is transparent to so that values within limits show through.

Figure 33. Visualizations of DTA from USGS Data and ASMD Error
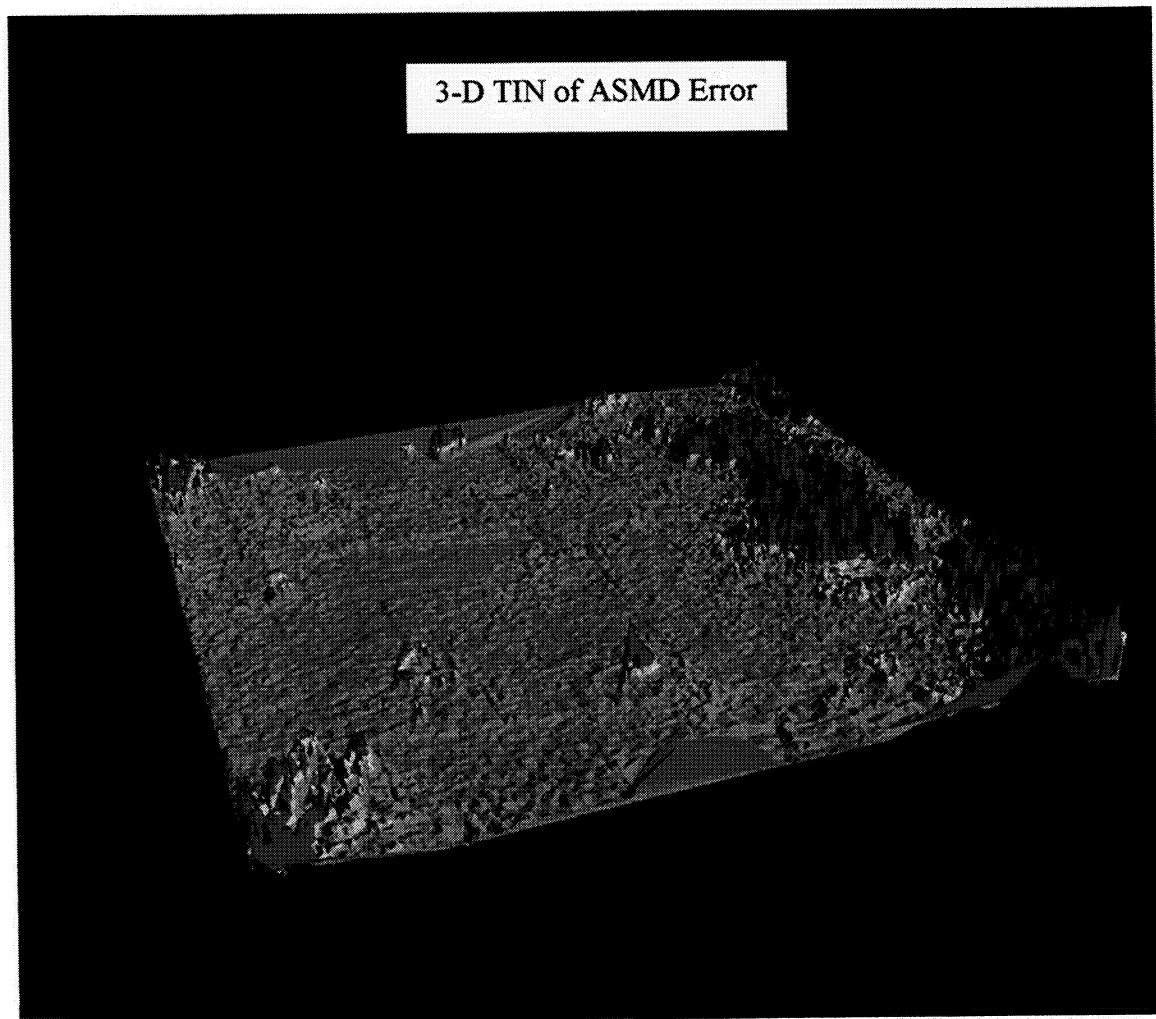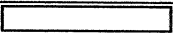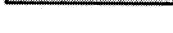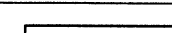
Color	Elevation (m)
	150-342
	50-150
	30-50
	-30-30*
	-49 - -30

Figure 34. Visualizations of DTA from ASMD Error Data

Color	Elevation (m)		Color	Elevation (m)
	2864-3050			2894-3083
	2678-2864			2705-2894
	2492-2678			2515-2705
	2306-2492			2326-2515
	2119-2306			2137-2326
	1933-2119			1948-2137
	1747-1933			1758-1948
	1561-1747			1569-1758
	1375-1561			1380-1569

ASMD Elevation Legend	USGS Elevation Legend

Figure 35.Visualizations of DTA from ASMD and USGS Data

HII   LAKE HAVASU CITY

Color	Elevation (m)
	2045-2286
	1804-2045
	1562-1804
	1321-1562
	1080-1321
	839-1080
	597-839
	356-597
	115-356
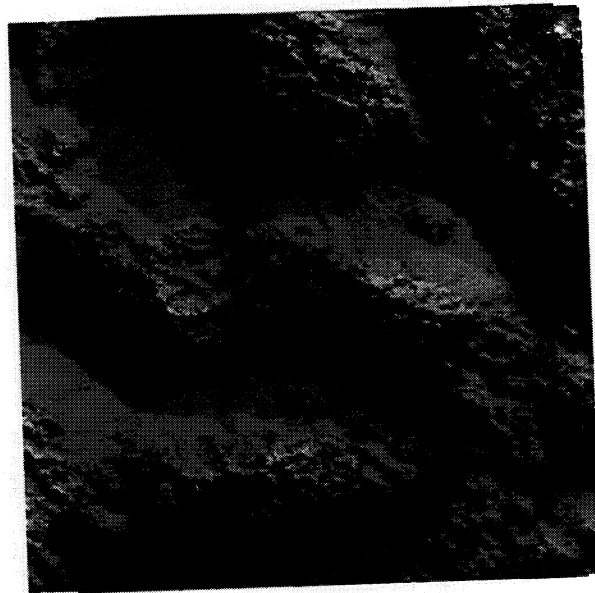
2-Dimensional Representation of HII from ASMD

3-Dimensional TIN of HII from ASMD
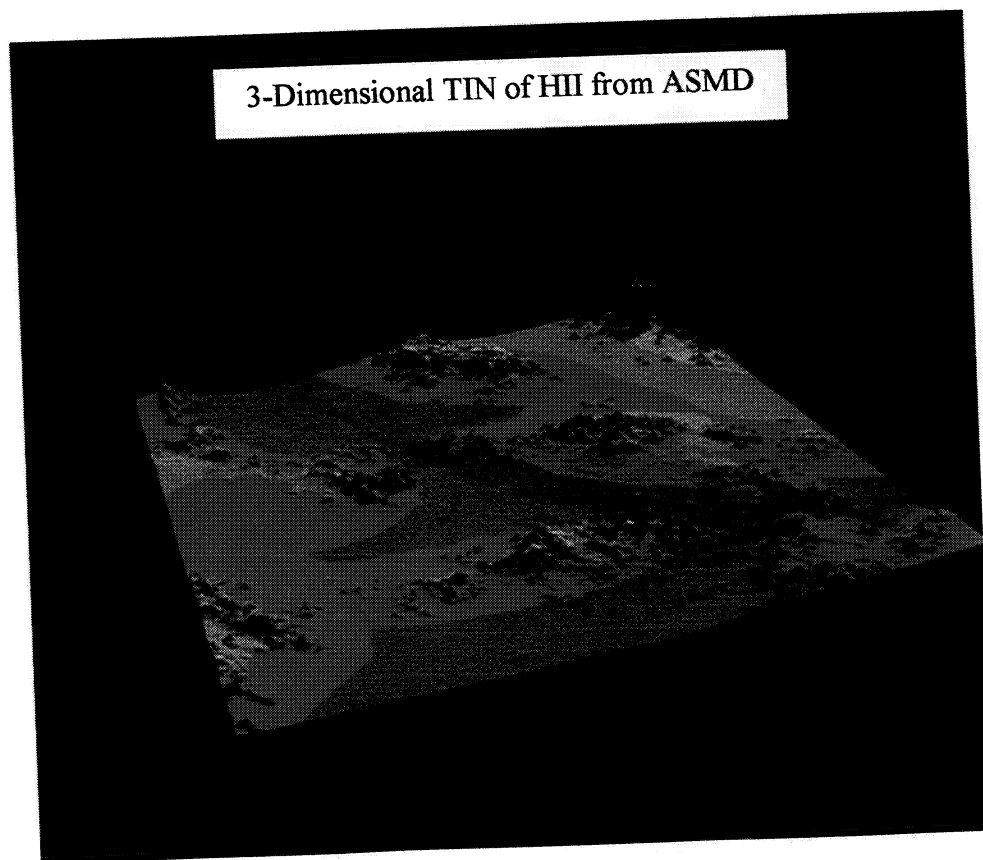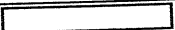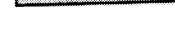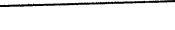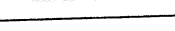
Figure 36.Visualizations of HII from ASMD
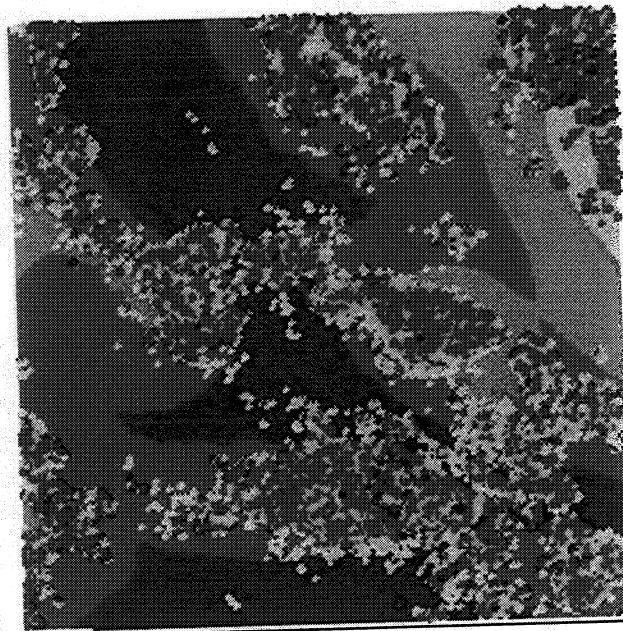
2-D view of ASMD Error v. USGS Elevation
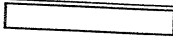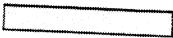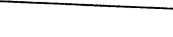
Color	Elevation (m)
	2072-2317
	1826-2072
	1581-1826
	1335-1581
	1090-1335
	844-1090
	599-844
	353-599
	108-353

Color	Elevation (m)
	150-365
	50-150
	30-50
	-30 - 30*
	-46- -30

*This color band is transparent to so that values within limits show through.

Figure 37. Visualizations of HII from USGS Data and ASMD Error
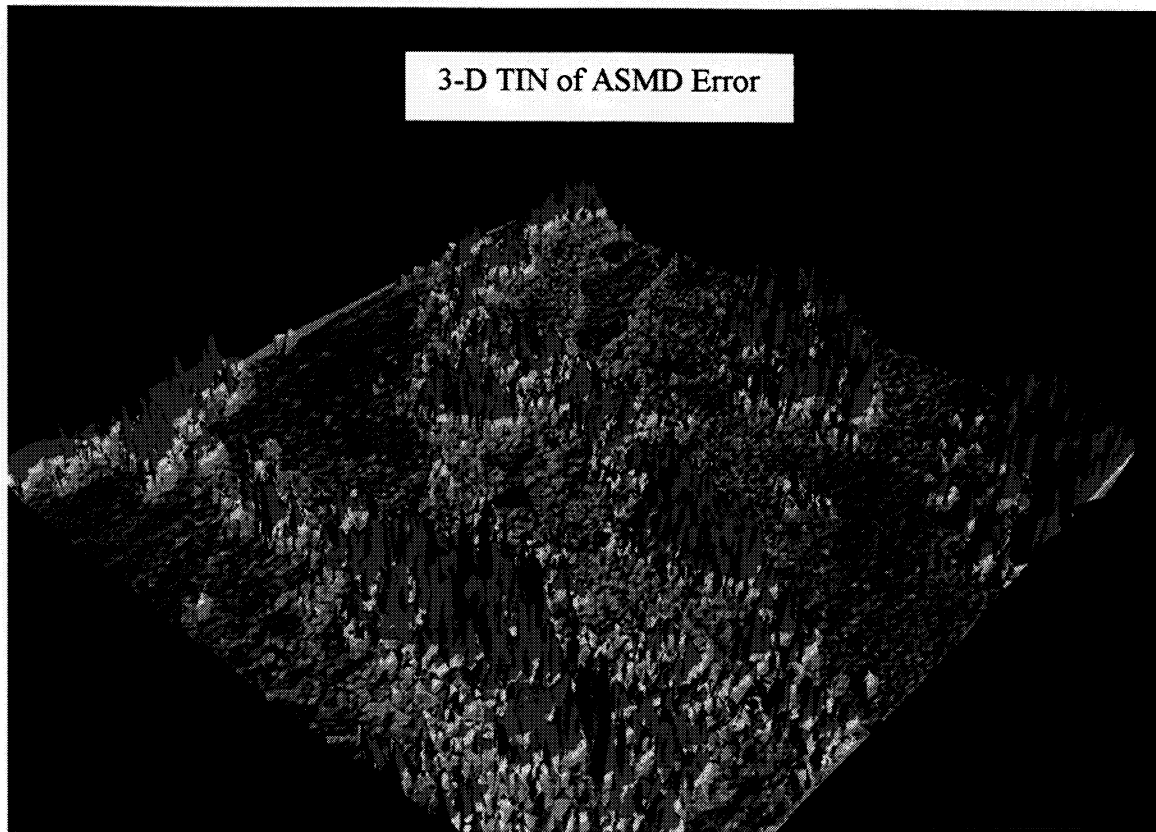
Color	Elevation (m)
	150-365
	50-150
	30-50
	-30 - 30
	-46 - -30

Figure 38. Visualizations of HII from ASMD Error Data

Figure 39. Visualizations of HII from ASMD and USGS Data

PSP   PALM SPRINGS REGIONAL

Color	Elevation (m)
	3057-3448
	2666-3057
	2276-2666
	1885-2276
	1494-1885
	1103-1494
	713-1103
	322-713
	-69-322
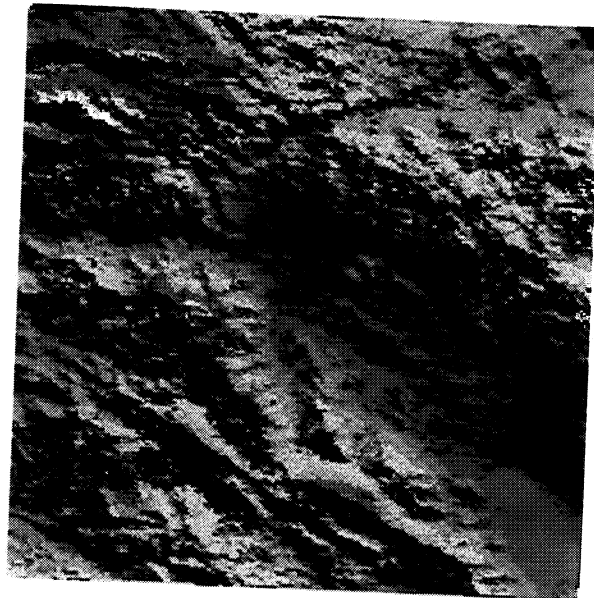
2-Dimensional Representation of PSP from ASMD

3-Dimensional TIN of PSP from ASMD
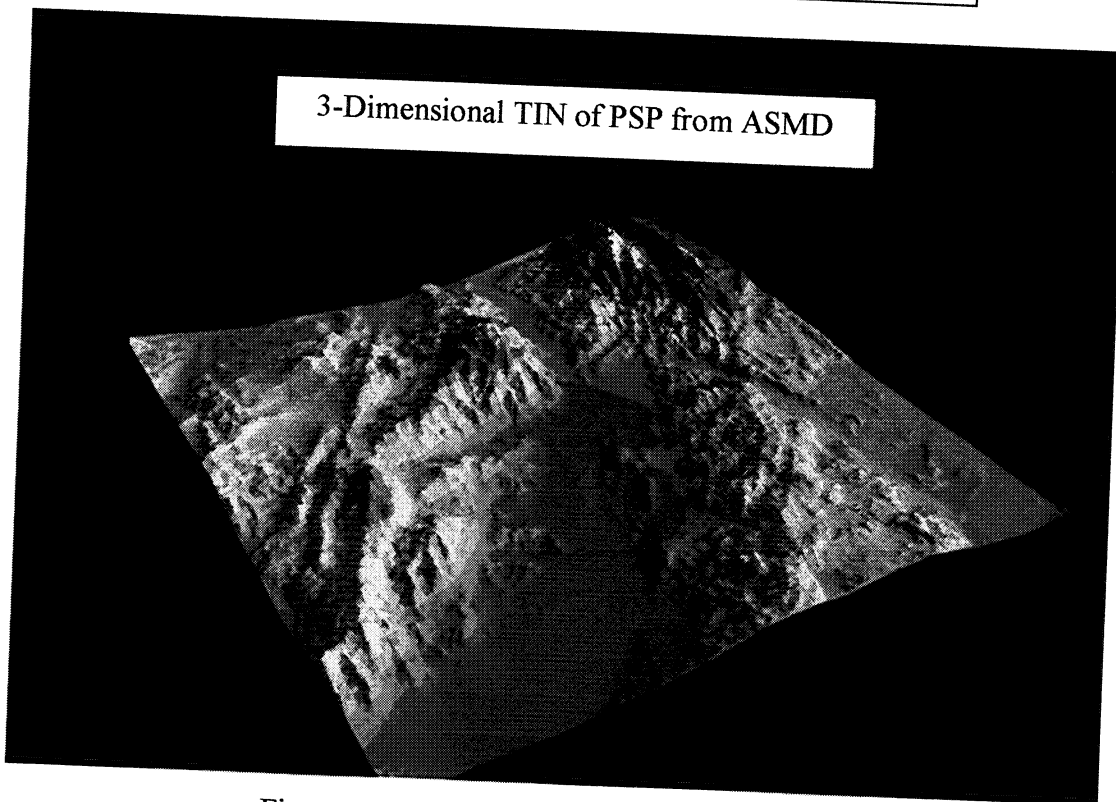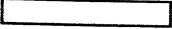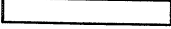
Figure 40. Visualizations of PSP from ASMD
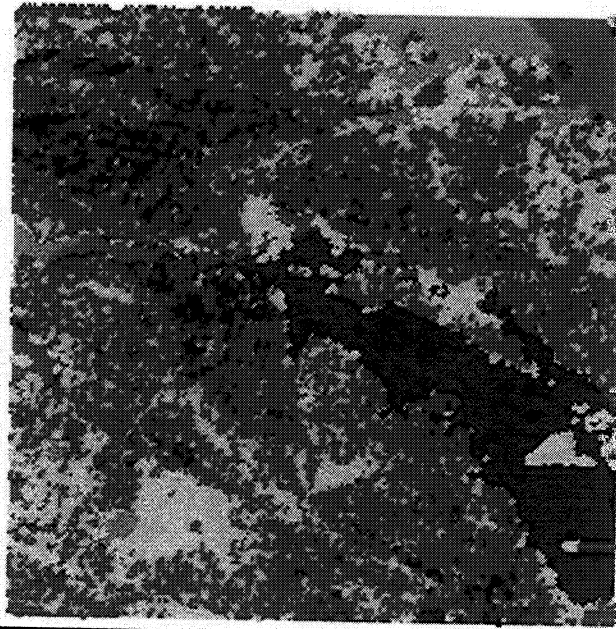
2-D view of ASMD Error v. USGS Elevation

Color	Elevation (m)
	3109-3506
	2711-3109
	2314-2711
	1917-2314
	1519-1917
	1122-1519
	725-1122
	327-725
	-70-327

Color	Elevation (m)
	150-357
	50-150
	30-50
	-30 - 30*
	-127 - -30

*This color band is transparent to so that values within limits show through.

Figure 41. Visualizations of PSP from USGS Data and ASMD Error
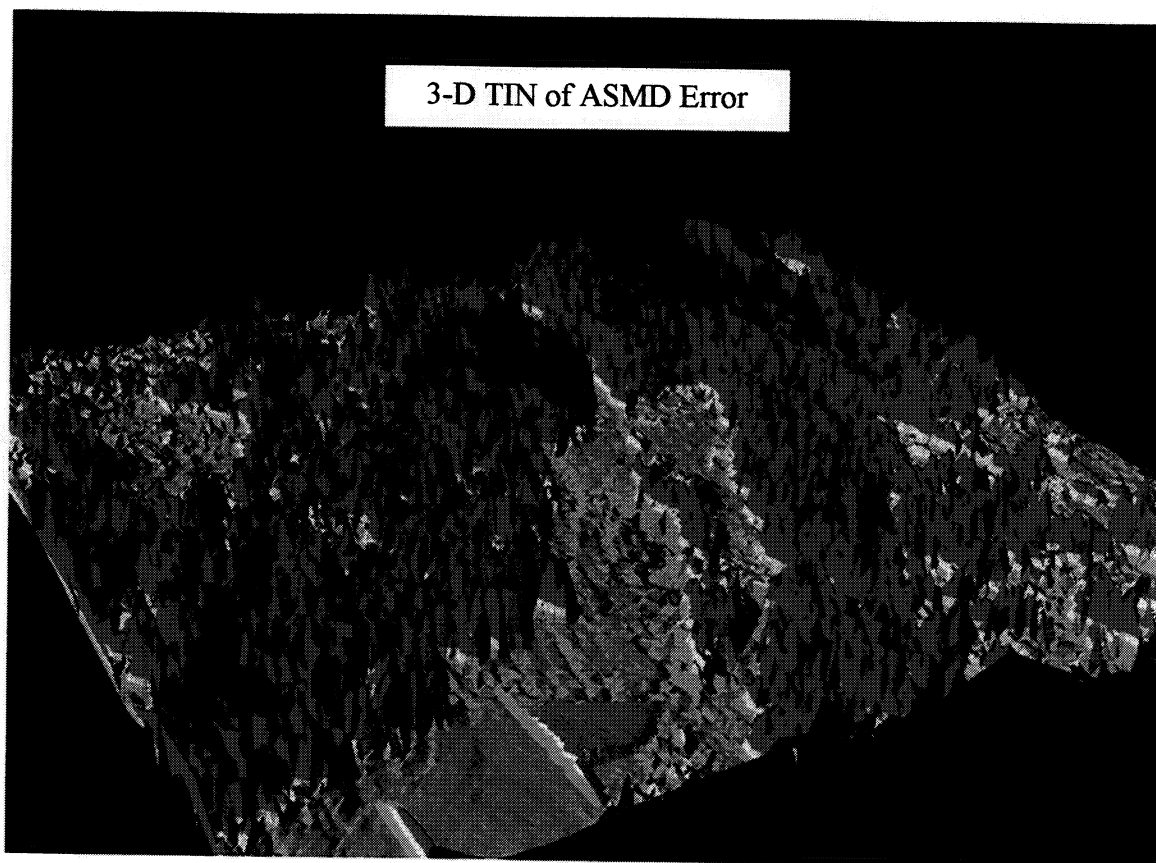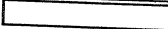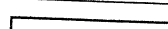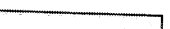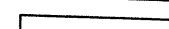
Color	Elevation (m)
	150-357
	50-150
	30-50
	-30-30
	-127 - -30

Figure 42. Visualizations of PSP from ASMD Error Data

Figure 43. Visualizations of PSP from ASMD and USGS Data

SNA   JOHN WAYNE AIRPORT-ORANGE COUNTY

Color	Elevation (m)
	1541-1734
	1349-1541
	1156-1349
	993-1156
	771-993
	578-771
	385-578
	193-385
	0-193

2-Dimensional Representation of SNA from ASMD

3-Dimensional TIN of SNA from ASMD
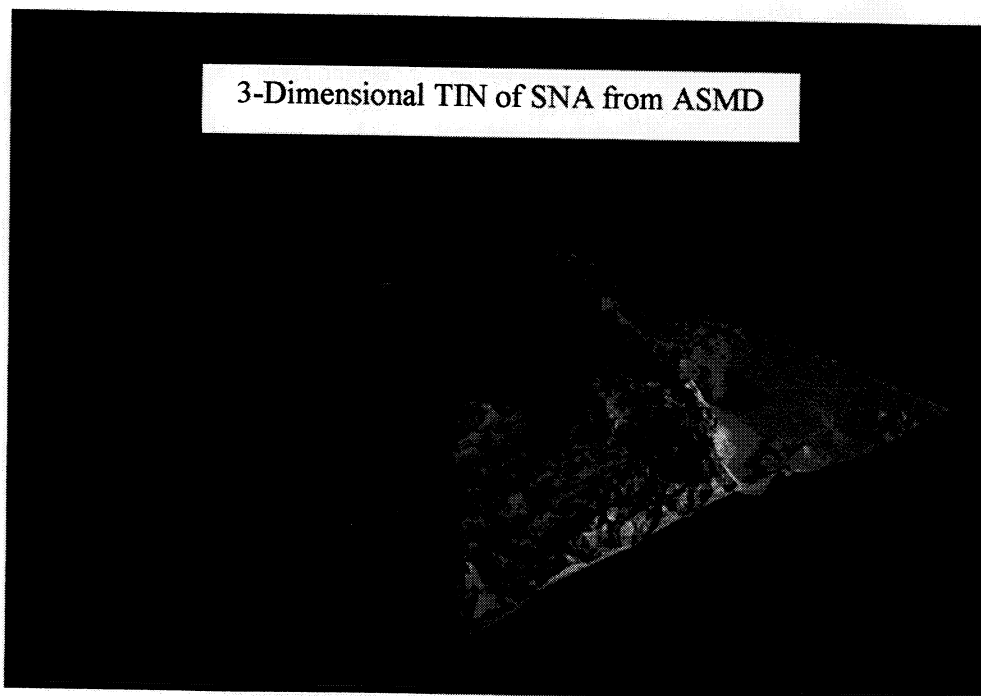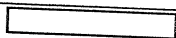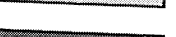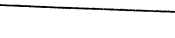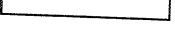
Figure 44. Visualizations of SNA from ASMD
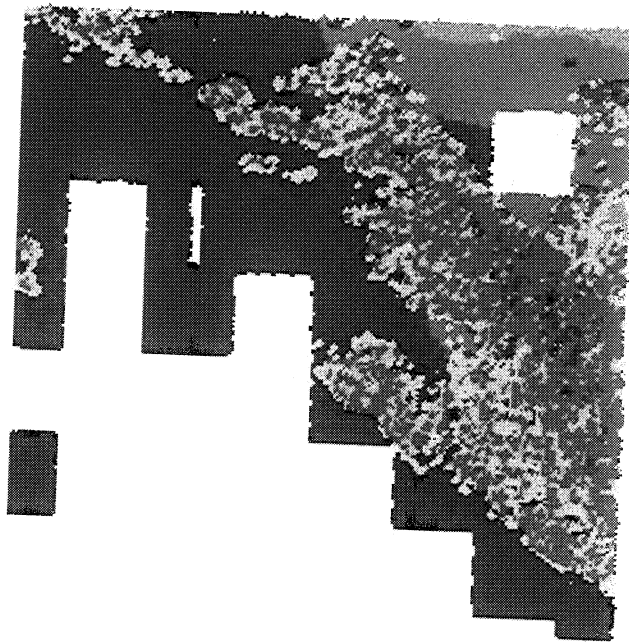
2-D View of ASMD Error v. USGS Elevation

Color	Elevation (m)
	1542-1735
	1349-1542
	1157-1349
	964-1157
	771-964
	578-771
	386-578
	193-386
	0-193

Color	Elevation (m)
	150-248
	50-150
	30-50
	-30 - 30*
	-119 - -30

*This color band is transparent to so that values within limits show through.

Figure 45. Visualizations of SNA from USGS Data and ASMD Error
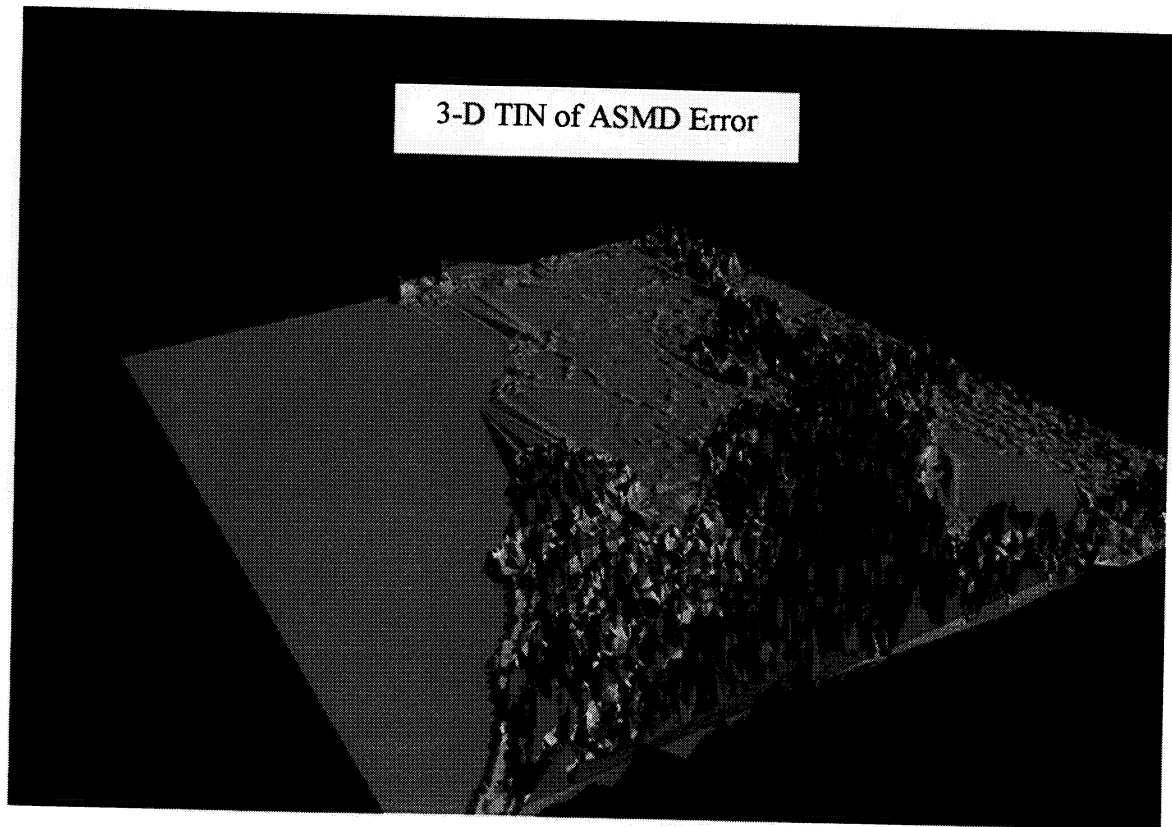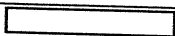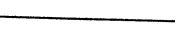
Color	Elevation (m)
	150-248
	50-150
	30-50
	-30-30
	-119 - -30
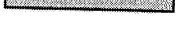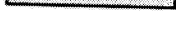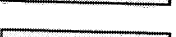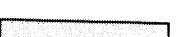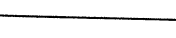
Figure 46. Visualizations of SNA from ASMD Error Data

Color	Elevation (m)	Color	Elevation (m)
	1541-1734		1542-1735
	1349-1541		1349-1542
	1156-1349		1157-1349
	993-1156		964-1157
	771-993		771-964
	578-771		578-771
	385-578		386-578
	193-385		193-386
	0-193		0-193

ASMD Elevation Legend	USGS Elevation Legend

Figure 47. Visualizations of SNA from ASMD and USGS Data

Appendix 3

Select Statistical Output for Sample Airports

1S4   SCAPPOOSE INDUSTRIAL AIRPARK

AVL   ASHEVILLE REGIONAL

CRQ   MC CLELLAN-PALOMAR

DTA   DELTA MUNI

HII   LAKE HAVASU CITY

PSP   PALM SPRINGS REGIONAL

SNA   JOHN WAYNE AIRPORT-ORANGE COUNTY

1S4   SCAPPOOSE INDUSTRIAL AIRPARK

Bin	Frequency
-72	1
-31.2	3490
9.6	21360
50.4	15997
91.2	4058
132	913
172.8	174
213.6	42
254.4	6
295.2	2
336	0
376.8	0
417.6	0
458.4	0
499.2	0
540	0
580.8	0
621.6	0
662.4	4
703.2	2
744	2
More	0
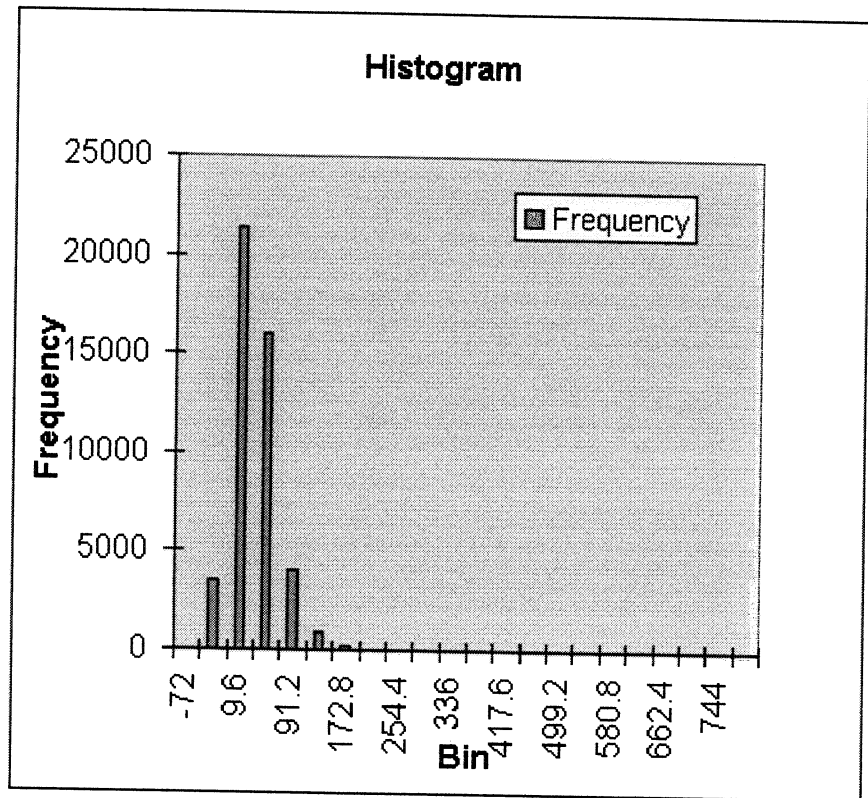
Mean Error	11.732167
Median	6
Mode	-7
Standard Deviation	34.267535
Sample Variance	1174.2639
Range	816
Minimum	-72
Maximum	744
Count	46051

# Safe values	35285
# Danger values	10766
Total	46051
% Safe	76.62%
% Dangerous	23.38%

AVL   ASHEVILLE REGIONAL

Bin	Frequency
-266	2
-239	20
-212	26
-185	9
-158	0
-131	0
-104	0
-77	0
-50	3
-23	49
4	6647
31	16467
58	8934
85	6241
112	3721
139	1592
166	558
193	176
220	44
247	5
274	5
More	0
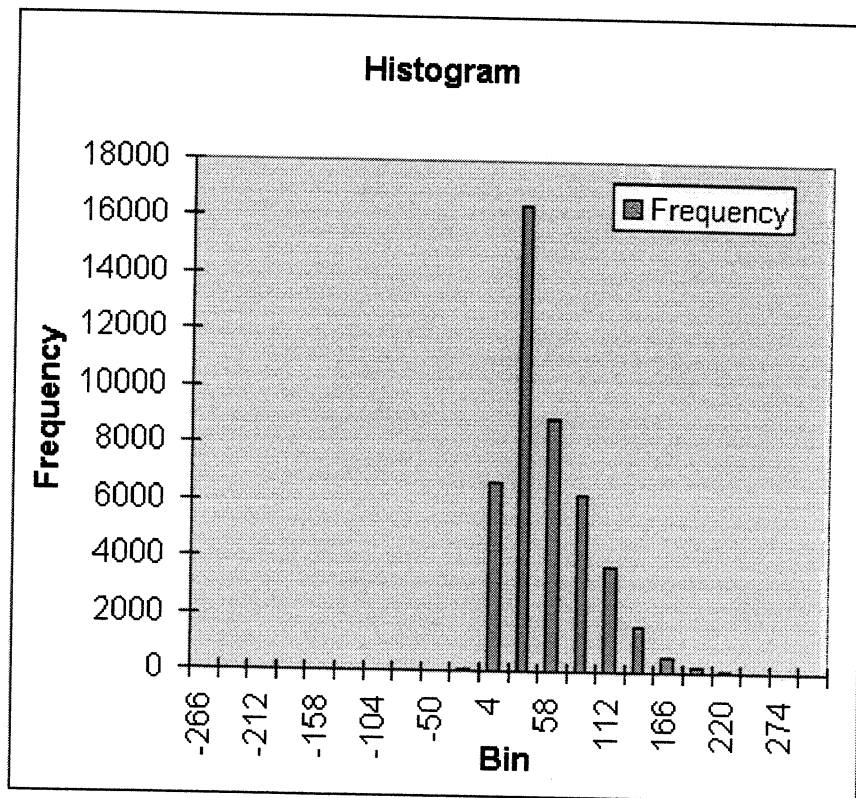
Mean Error	39.7714106
Median	29
Mode	6
Standard Deviation	39.4052989
Sample Variance	1552.77758
Range	540
Minimum	-266
Maximum	274
Count	44499

# Safe values	22841
# Danger values	21658
Total	44499
% Safe	51.33%
% Dangerous	48.67%

CRQ MC CLELLAN-PALOMAR

Bin	Frequency
-154	1
-115.9	11
-77.8	21
-39.7	178
-1.6	2071
36.5	20771
74.6	7958
112.7	2511
150.8	773
188.9	210
227	60
265.1	8
303.2	1
341.3	1
379.4	0
417.5	0
455.6	0
493.7	0
531.8	0
569.9	0
608	2
More	0

Histogram
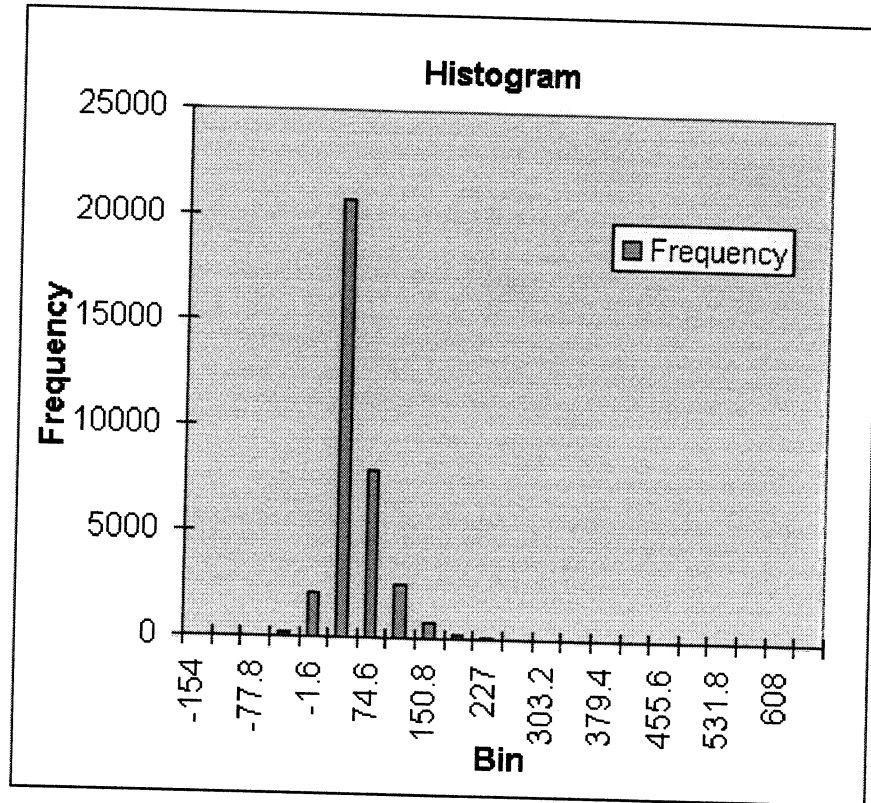
Mean Error	30.0694392
Median	22
Mode	0
Standard Deviation	34.8891955
Sample Variance	1217.25597
Range	762
Minimum	-154
Maximum	608
Count	34577

# Safe values	20834
# Danger values	13743
Total	34577
% Safe	60.25%
% Dangerous	39.75%

DTA DELTA MUNI

Bin	Frequency
-49	1
-29.45	26
-9.9	614
9.65	23809
29.2	11265
48.75	2933
68.3	1560
87.85	919
107.4	673
126.95	427
146.5	289
166.05	162
185.6	81
205.15	34
224.7	10
244.25	10
263.8	3
283.35	2
302.9	0
322.45	0
342	1
More	0

Histogram
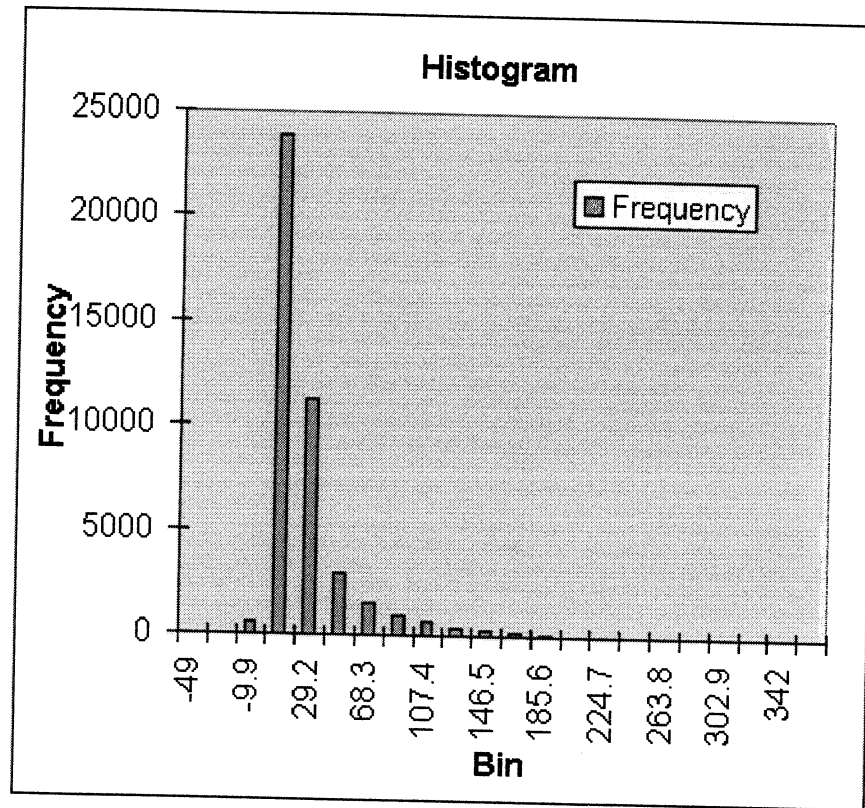
Mean Error	16.312945
Median	7
Mode	1
Standard Deviation	28.341519
Sample Variance	803.24169
Range	391
Minimum	-49
Maximum	342
Count	42819

# Safe values	35939
# Danger values	6880
Total	42819
% Safe	83.93%
% Dangerous	16.07%

HII LAKE HAVASU CITY

Bin	Frequency
-46	1
-25.45	15
-4.9	2080
15.65	23451
36.2	10049
56.75	3963
77.3	2418
97.85	1344
118.4	816
138.95	452
159.5	233
180.05	151
200.6	98
221.15	42
241.7	25
262.25	14
282.8	10
303.35	4
323.9	5
344.45	0
365	3
More	0
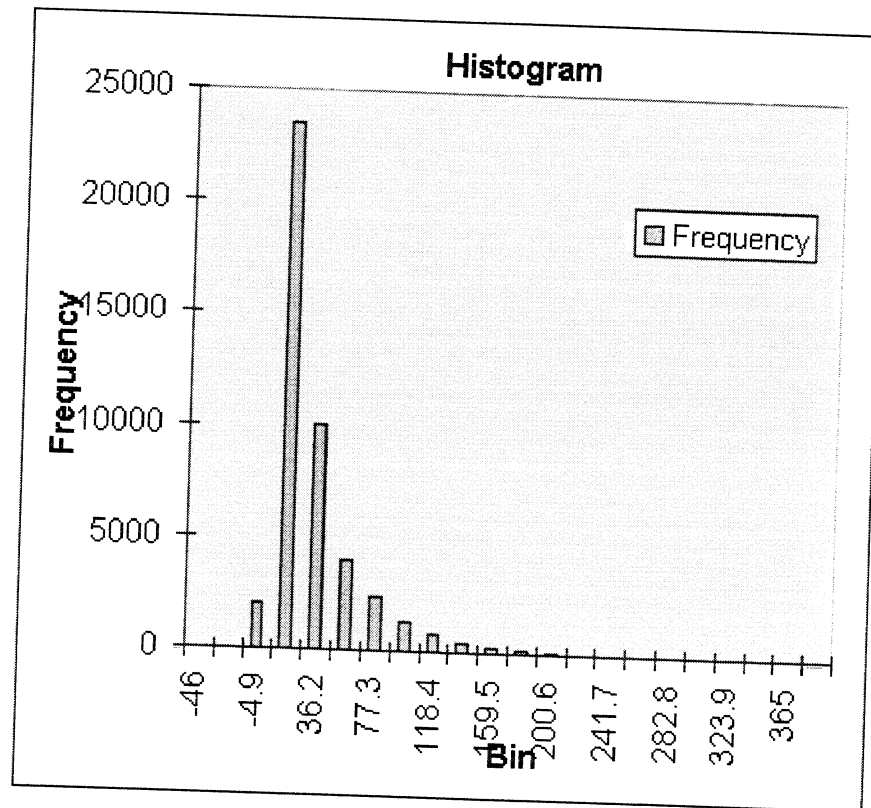
Mean Error	23.5138575
Median	13
Mode	0
Standard Deviation	32.4888211
Sample Variance	1055.52349
Range	411
Minimum	-46
Maximum	365
Count	45174

# Safe values	33775
# Danger values	11399
Total	45174
% Safe	74.77%
% Dangerous	25.23%

PSP   PALM SPRINGS REGIONAL

Bin	Frequency
-127	1
-102.8	5
-78.6	25
-54.4	78
-30.2	201
-6	1438
18.2	8428
42.4	11373
66.6	8169
90.8	6089
115	4270
139.2	2541
163.4	1590
187.6	862
211.8	484
236	246
260.2	140
284.4	69
308.6	33
332.8	13
357	7
More	0
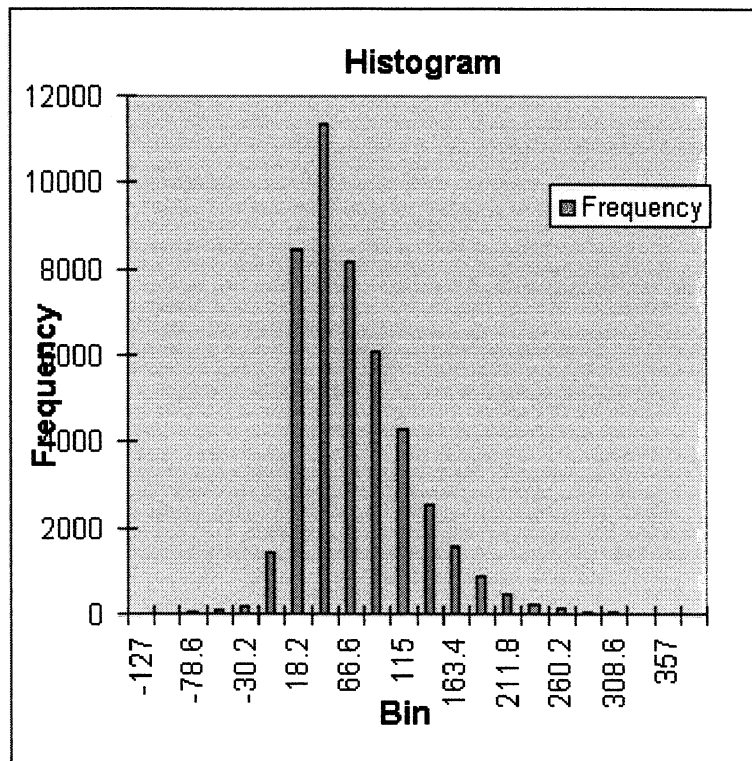
Mean Error	57.6077244
Median	47
Mode	25
Standard Deviation	50.9197299
Sample Variance	2592.81889
Range	484
Minimum	-127
Maximum	357
Count	46062

# Safe values	16445
# Danger values	29617
Total	46062
% Safe	35.70%
% Dangerous	64.30%

SNA   JOHN WAYNE AIRPORT-ORANGE COUNTY

Bin	Frequency
-119	3
-100.65	1
-82.3	3
-63.95	12
-45.6	11
-27.25	59
-8.9	443
9.45	15631
27.8	4245
46.15	3755
64.5	2376
82.85	1443
101.2	873
119.55	542
137.9	398
156.25	250
174.6	116
192.95	77
211.3	19
229.65	11
248	5
More	0

Histogram — Frequency vs Bin
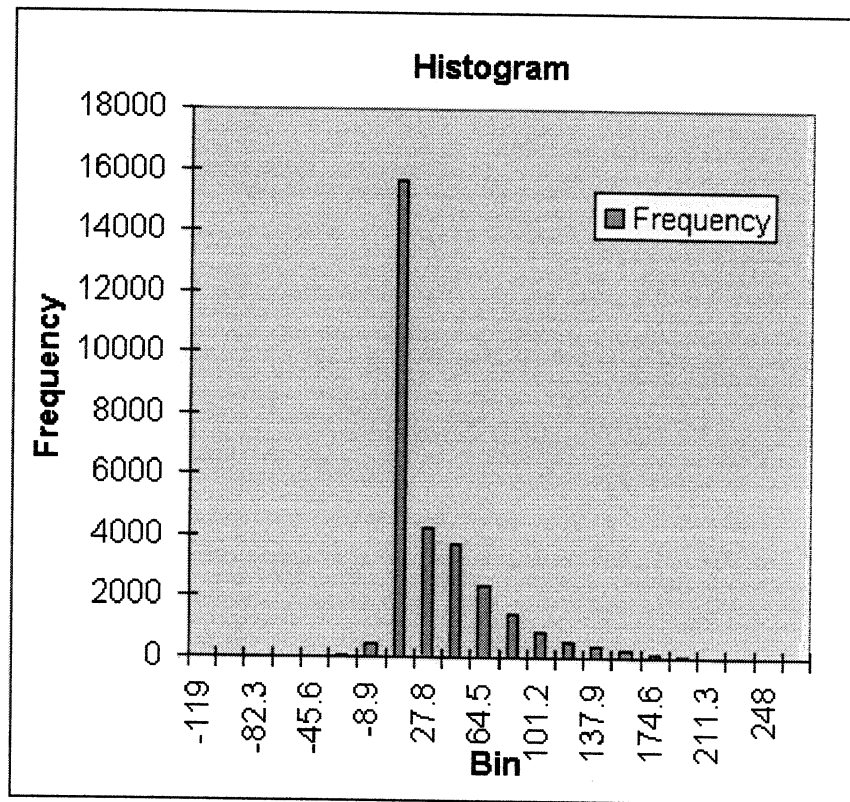
Mean Error	23.9629042
Median	7
Mode	0
Standard Deviation	35.2723403
Sample Variance	1244.13799
Range	367
Minimum	-119
Maximum	248
Count	30273

# Safe values	21062
# Danger values	9211
Total	30273
% Safe	69.57%
% Dangerous	30.43%

Appendix 4

Contour and Error Representation Visualizations for All Sampled Airports

1S4   SCAPPOOSE INDUSTRIAL AIRPARK

AVL   ASHEVILLE REGIONAL

CRQ   MC CLELLAN-PALOMAR

DEN   DENVER INTERNATIONAL AIRPORT

DTA   DELTA MUNI

HII   LAKE HAVASU CITY

PSP   PALM SPRINGS REGIONAL

SNA   JOHN WAYNE AIRPORT-ORANGE COUNTY

48V   TRI-COUNTY

ASE   ASPEN PITKIN

COS   CITY OF COLORADO SPRINGS

EGE   EAGLE COUNTY

JAC   JACKSON HOLE

LGU   LOGAN CACHE

MSO   MISSOULA

SLC   SALT LAKE CITY

FLG   FLAGSTAFF PULLIAM

Figure 48. SCAPPOOSE INDUSTRIAL AIRPARK Contour and Error Representation
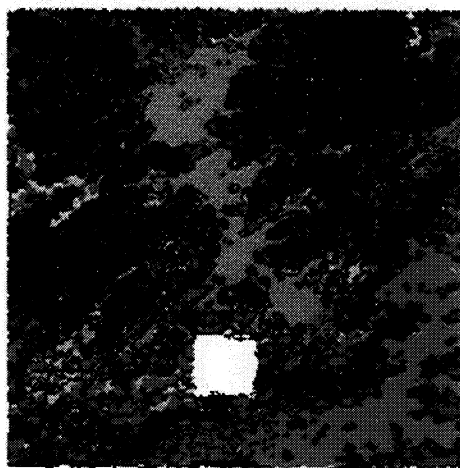
Figure 49. ASHEVILLE REGIONAL Contour and Error Representation
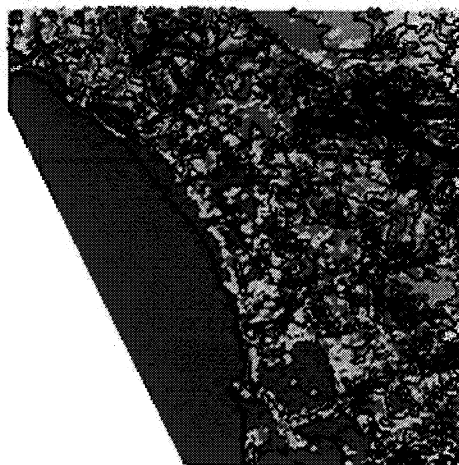
Figure 50.MC CLELLAN-PALOMAR Contour and Error Representation
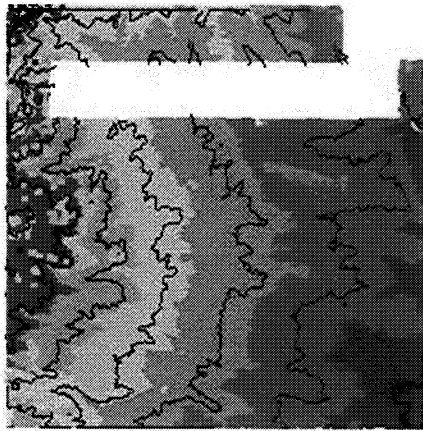
Figure51. DENVER INTERNATIONAL AIRPORT Contour and Error Representation
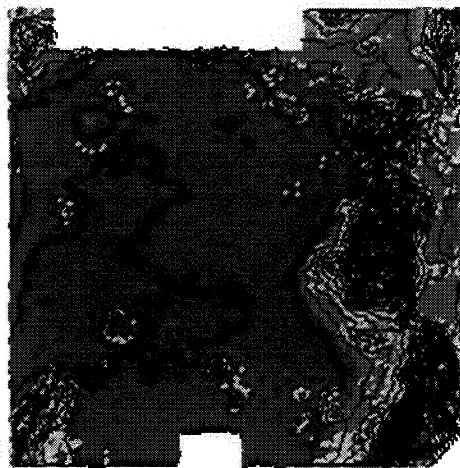

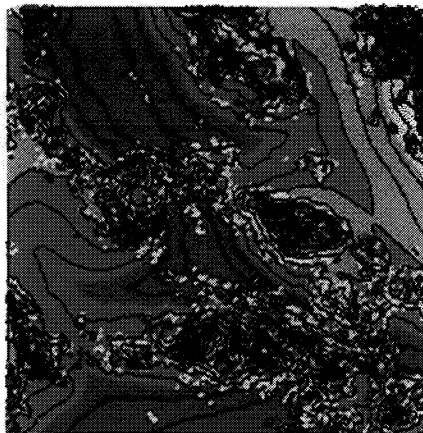Figure 52.DELTA MUNI Contour and Error Representation

Figure 53.LAKE HAVASU CITY Contour and Error Representation

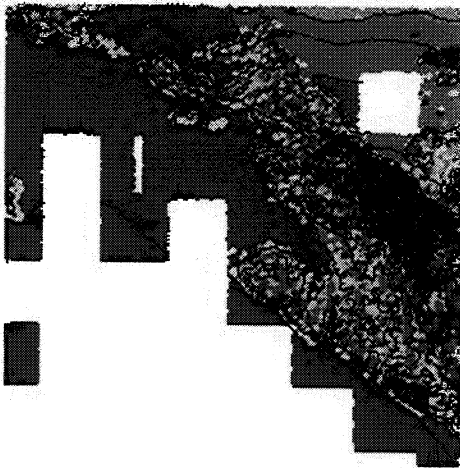Figure 54. PALM SPRINGS REGIONAL Contour and Error Representation

Figure 55.JOHN WAYNE AIRPORT-ORANGE COUNTY Contour and Error Representation
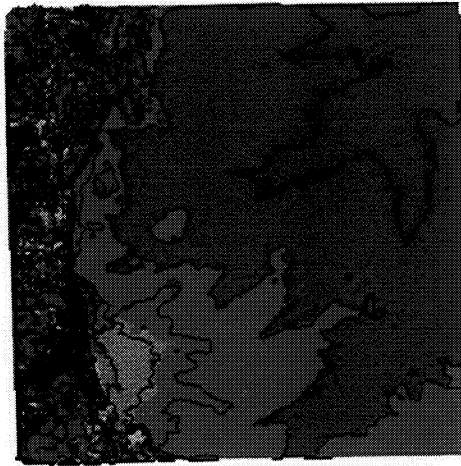
Figure 56. TRI-COUNTY Contour and Error Representation
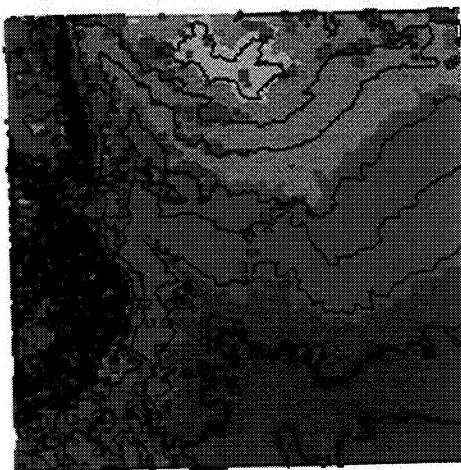

Figure 57.ASPEN PITKIN Contour and Error Representation

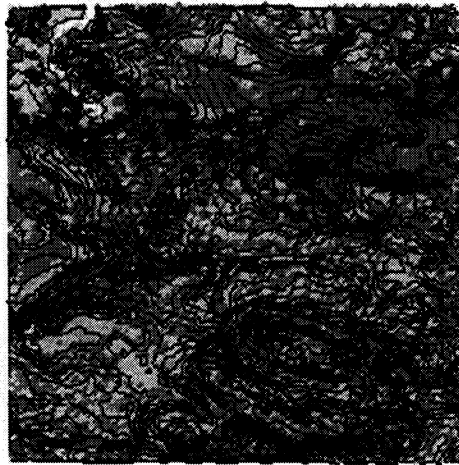Figure 58.CITY OF COLORADO SPRINGS Contour and Error Representation
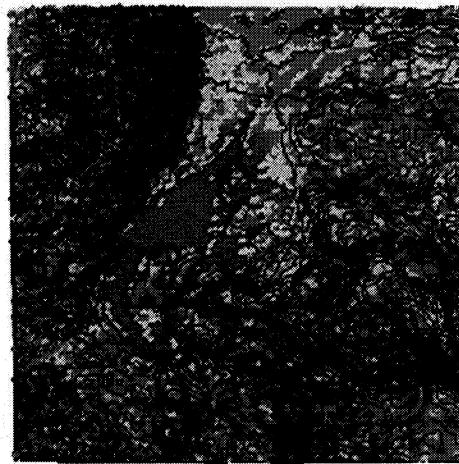
Figure 59. EAGLE COUNTY Contour and Error Representation
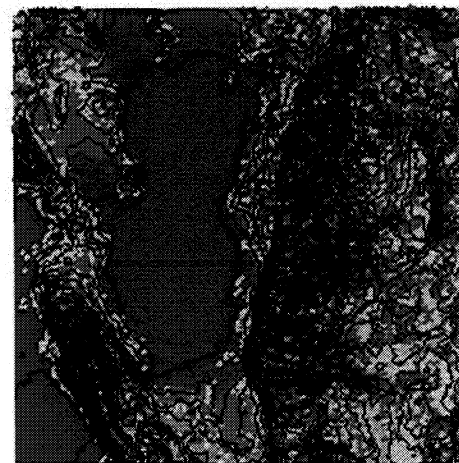

Figure 60.JACKSON HOLE Contour and Error Representation

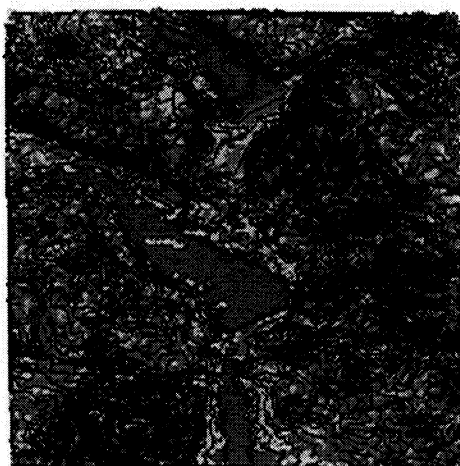Figure 61.LOGAN CACHE Contour and Error Representation

Figure 62.MISSOULA Contour and Error Representation
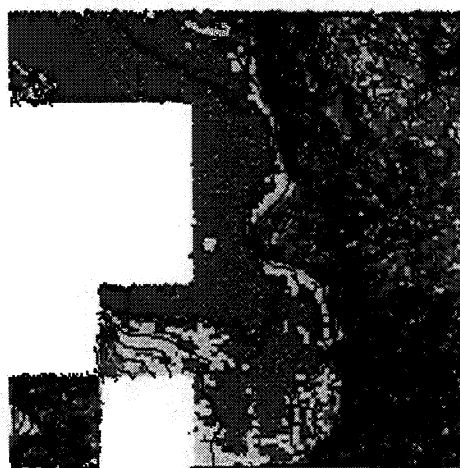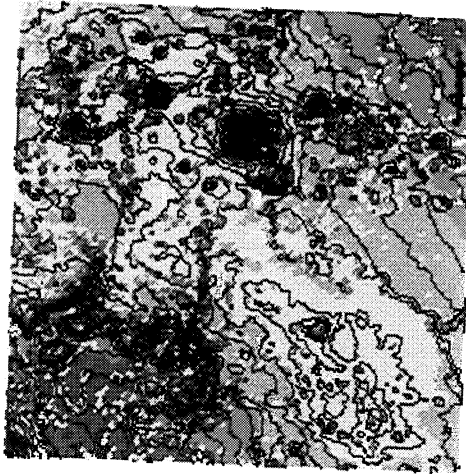

Figure 63. SALT LAKE CITY Contour and Error Representation

Figure 64. FLAGSTAFF PULLIAM Contour and Error Representation