INTEGRATION AND COMPRESSED REPRESENTATION OF GEOSPATIAL DATA USING NEURAL NETWORKS FOR 3-D PROFILING OF SUBSURFACE CONTAMINATION LEVELS

Berrin Tansel, Ibrahim N. Tansel, Wei Y. Bao

Florida International University
Miami, Florida

Outline

- Introduction
- Theoretical Background
- Modeling underground contamination with neural networks
- Results and Discussion
- Conclusions

Introduction

- Current situation: World is complex
- Complexities increase because:
  - More data are collected
  - Regulations and responsibilities increase
  - People are aware about the effect of various contaminants
  - We feel more responsible about the world we want to leave to the next generations

Introduction

- Objective:
  - Accurately assess the situation (Estimate the contamination levels)
  - Arrive the best decision (for possible purchases, cleaning operations, warning people, evaluation of feasibility)
  - Adequately estimate the value, cost and timing

Introduction

- Extensive data sets are available from multiple resources:
  - Contamination levels from randomly distributed wells
  - Satellite pictures at different wavelengths to obtain extensive information
- Powerful multipurpose tools such as neural networks could be used:
  - For direct correlating of inputs and the output(s) without writing analytical model
  - For cross correlating the information
  - Use of historic information easily for predictions

Introduction to Neural Networks

- Objective
  - Development of generic software which will learn
  - Perform very difficult tasks for computers
  - Use parallel processing hardware for speed and resistance against partial damage
- Accomplishments
  - Found many applications from character recognition to driving vehicle (others such as control of an airplane, credit rating, stock price estimation)
- Limitations
  - Hardware did not make the expected progress
  - Engineers want to understand the phenomena and establish the governingl rules

Introduction

- Backpropagation (BP) type neural network (NN) was found as a powerful mapping tool.
- A BP-NN with three layers is capable to represent complex information
- These layers include input, hidden and output layers
- If the geographic coordinates and depth are given as inputs, and concentration level of a contaminant is represented as output

Introduction

- Training process: uses the discrete data collected from the groundwater monitoring wells,
- To predict the contaminant levels: NN can be used at other coordinates around a site to produce continuous
- Output can be represented as 3-D contour maps of contamination profiles

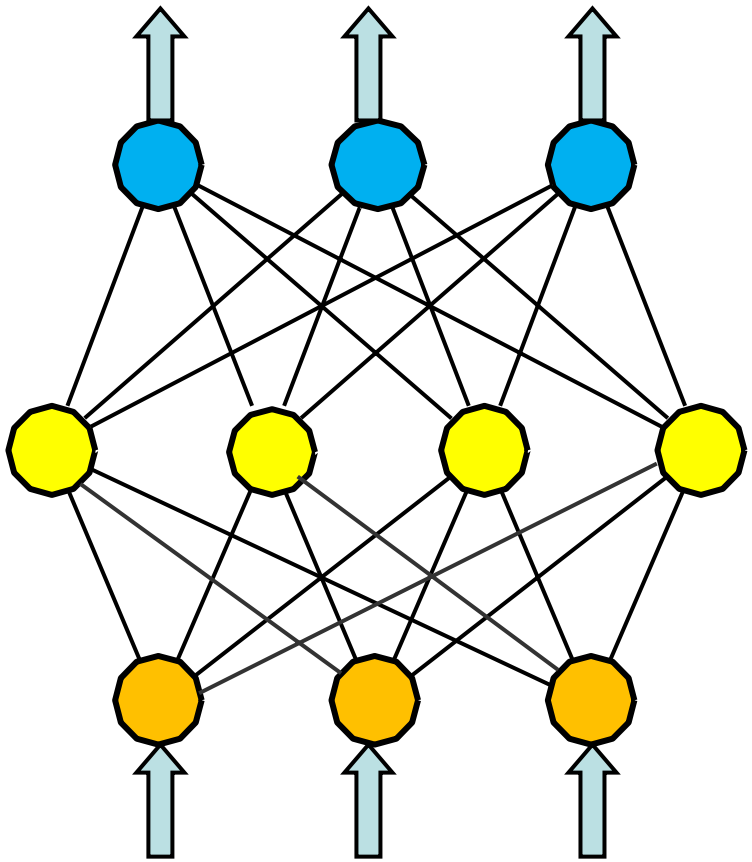
Theoretical Background BP-NN

- Neural Network is a computational model that mimics the constructions and operations of a biological brain to solve problems.
- Construct a network by using some nodes (neurons) connected together by some connection channels (nerves).

Theoretical Background BP-NN

- The most popular neural network
- Can be used for mapping and classification
- Can be forced to generalize and work well with small number of available cases
- Limitations: Requires many iterations (very slow), user determines number of hidden layers and nodes, cannot judge its own experience

Theoretical Background BP-NN



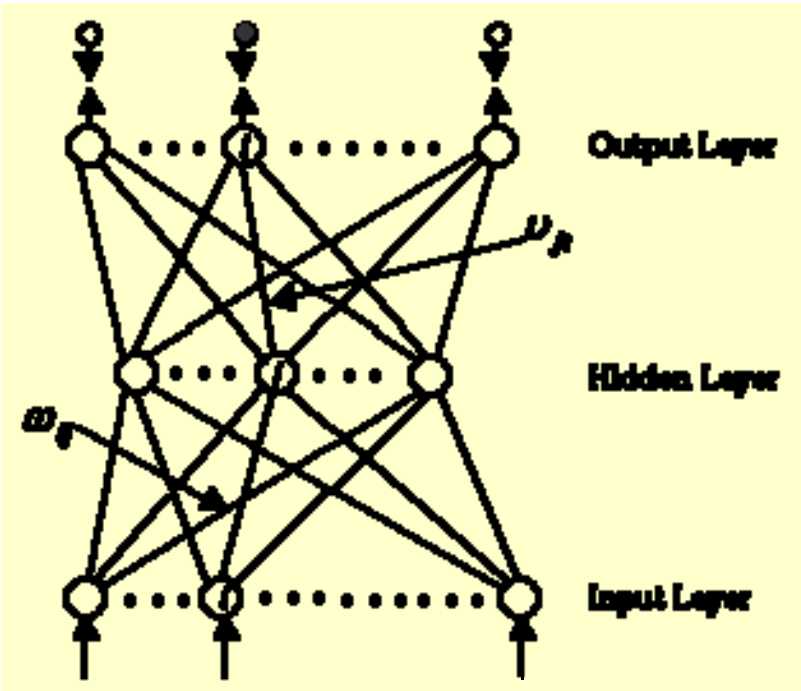
Output layer

Hidden layer

Input layer

Theoretical Background BP-NN

- Architecture



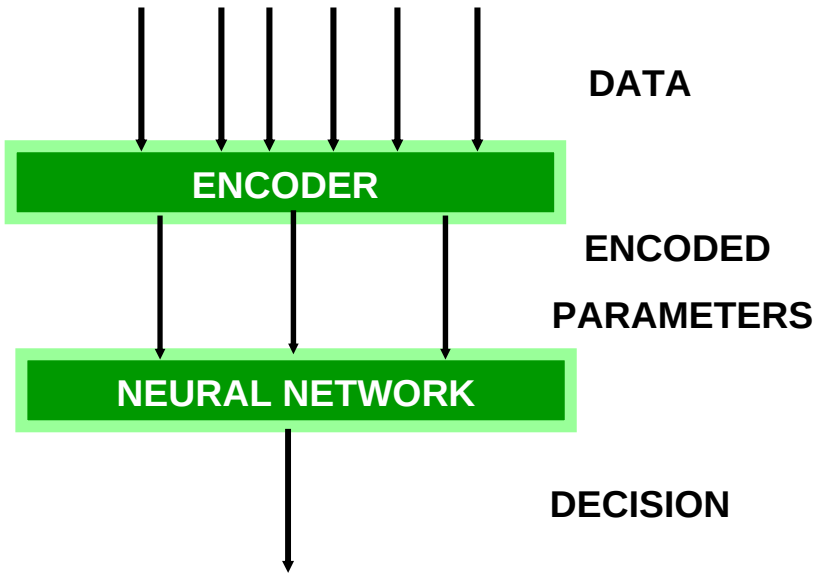
$$O_{pj} = \frac{1}{1 + e^{-(Sum(w_{ji}\ o_{pi} + \theta_i))}}$$

Most commonly used sigmoid activation function

Figure is taken from "http://www.inrialpes.fr/sharp/people/guo/Programming/programmingcontent/nnProgramming/node4.html" Equation is taken from Rumelhart "Learning Internal Representations by Error Propagation"

Theoretical Background BP-NN

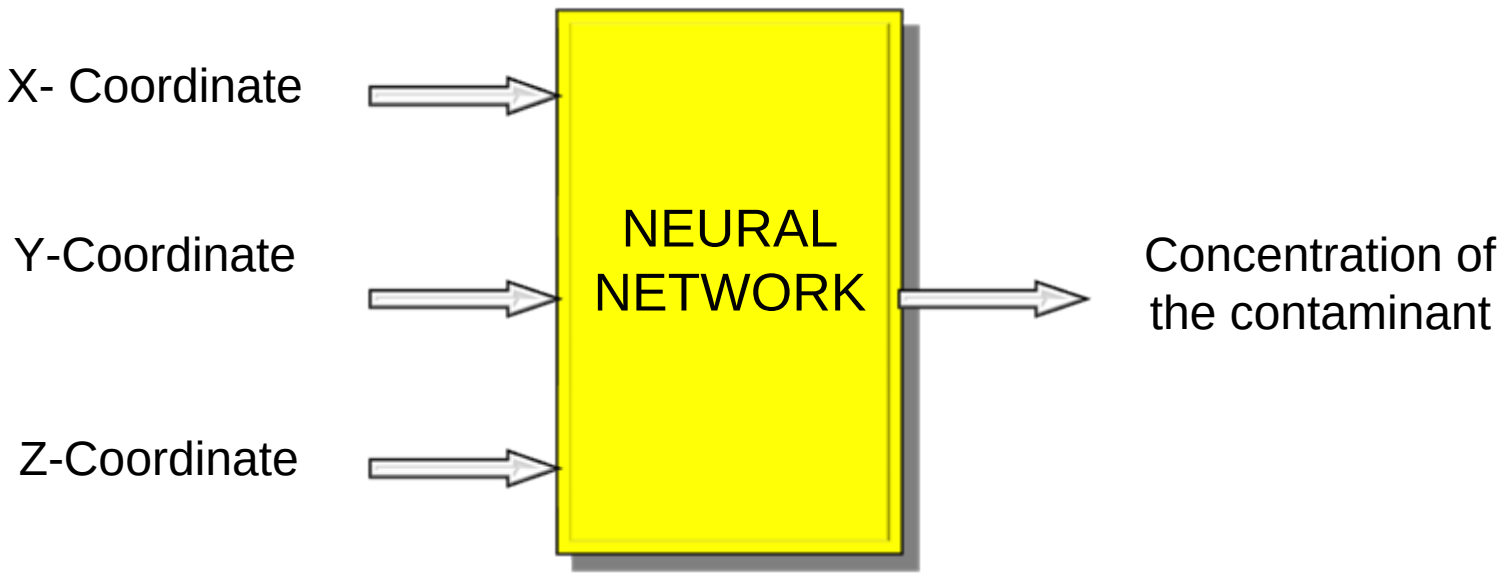
- Simplify their task by using encoder



- Train and test
  - Use 80% of the data for training
  - Test the performance of the neural network on the remaining 20%
  - Use simulated data

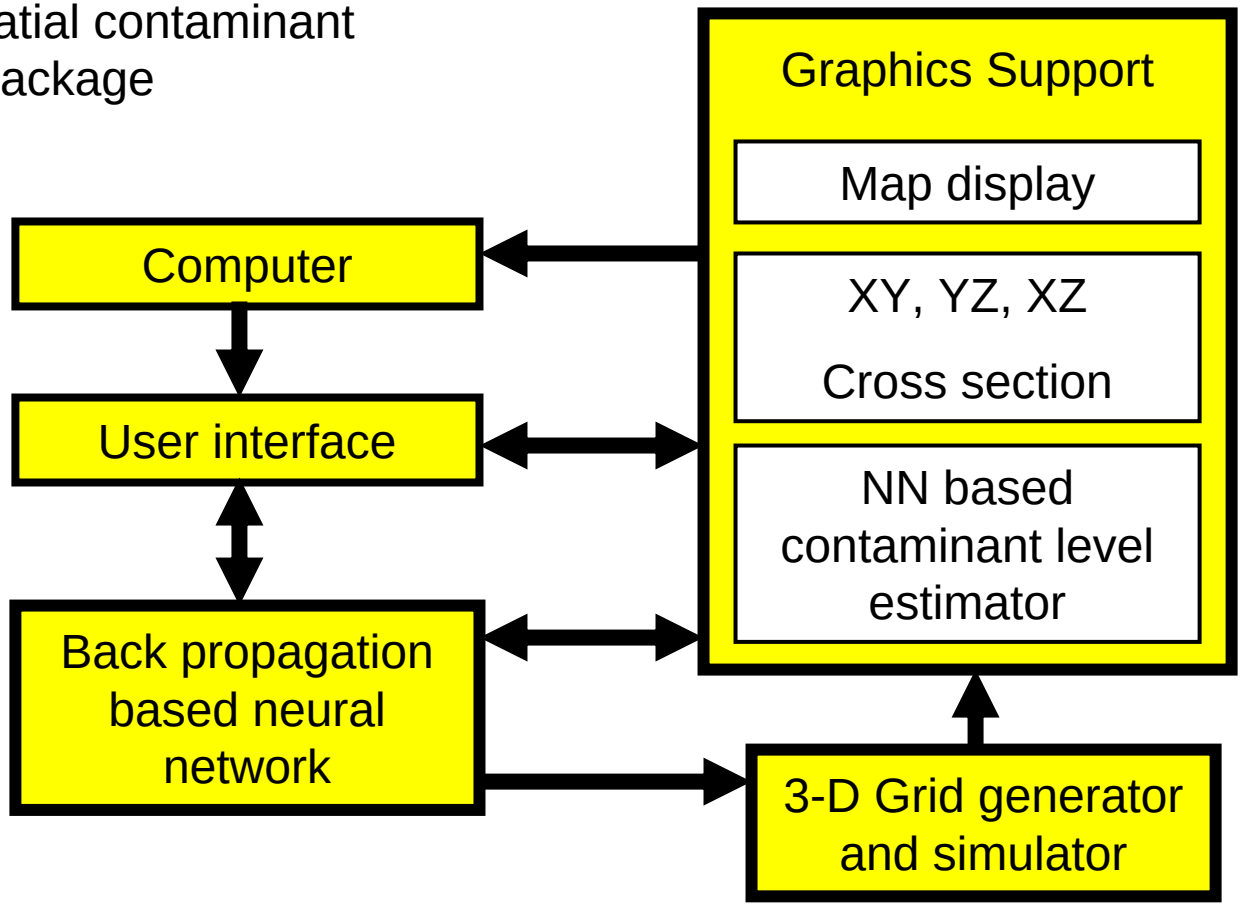
Modeling underground contamination with BP-NN

- Neural network



Modeling underground contamination with BP-NN

BP-NN geospatial contaminant visualization package (GCVP)

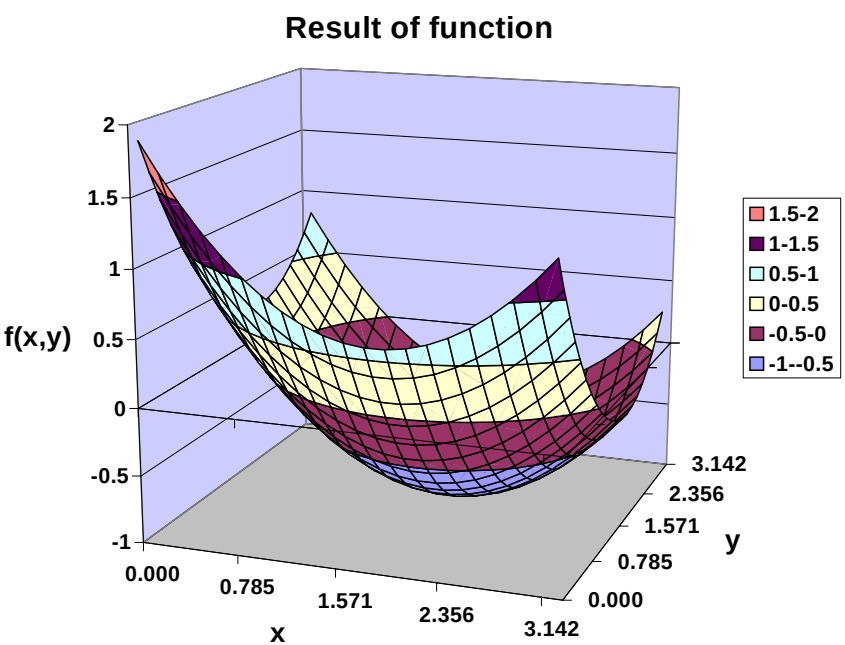


Results and Discussion

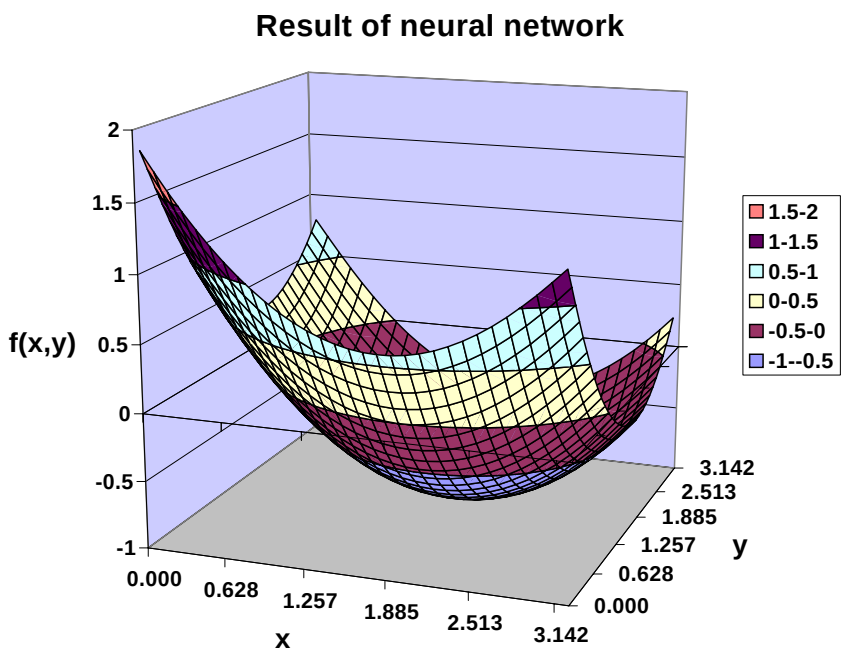
- Evaluate the performance of the BP-NN by using simulated data
- Model the underground contmination
  - Estimate the contamination at the desired 3-D coordinates
  - Demonstrate the distribution of the contamination at the considered cross-sections
  - Use the data from the satellites

Results and Discussion

$$f_2 = 0.4(x-1.8)^2 + 0.4(y-2)^2 - 1$$



441 Cases

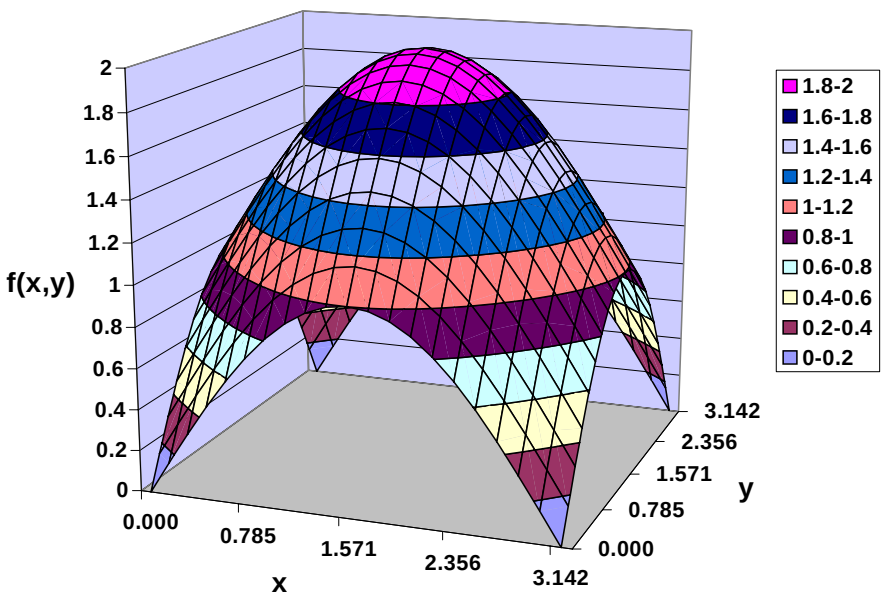


961 Cases, Error 0.15%

Results and Discussion

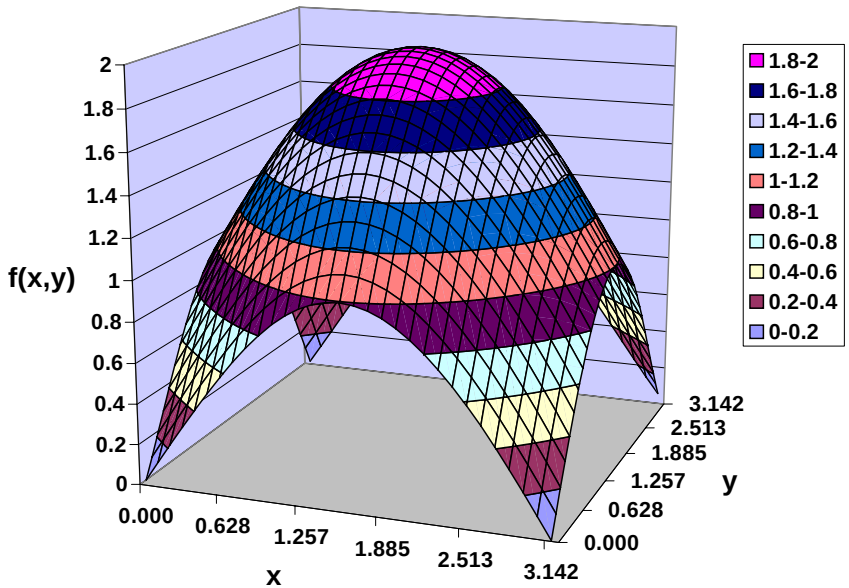
$$f_1 = sin(x)+sin(y)$$

Result of function



441 Cases

Result of neural network



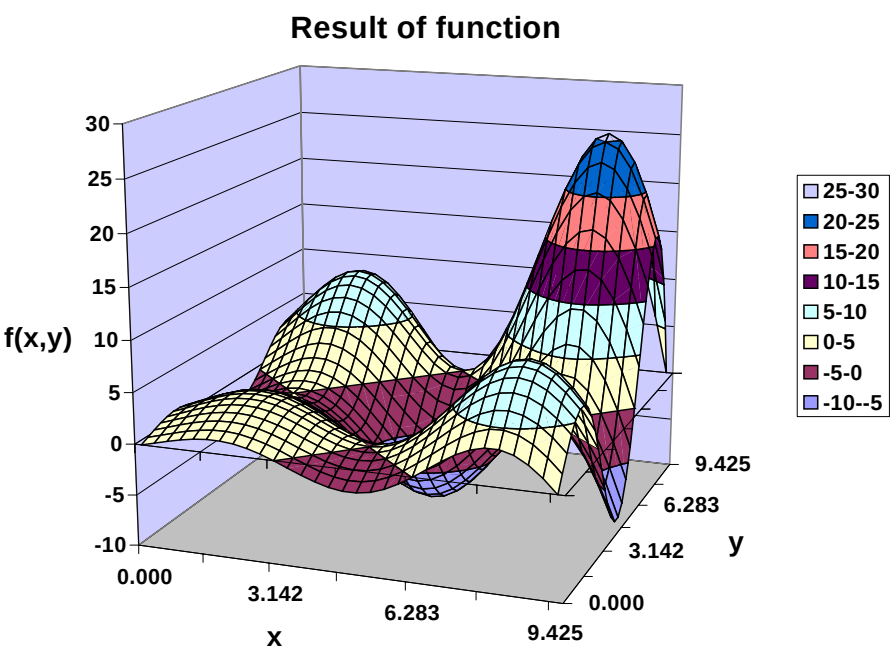
961 Cases, Error 0.26%

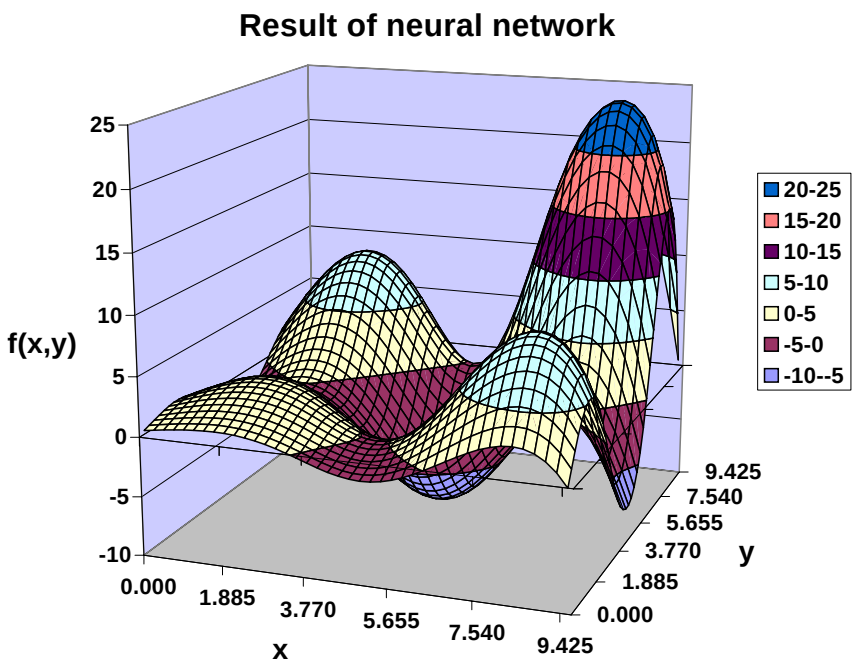
Results and Discussion

Representing the following function by using Neural Networks

$$f(x, y)=e^{(x+y)/(2\pi)}(\sin(x)+\sin(y))$$

Results and Discussion



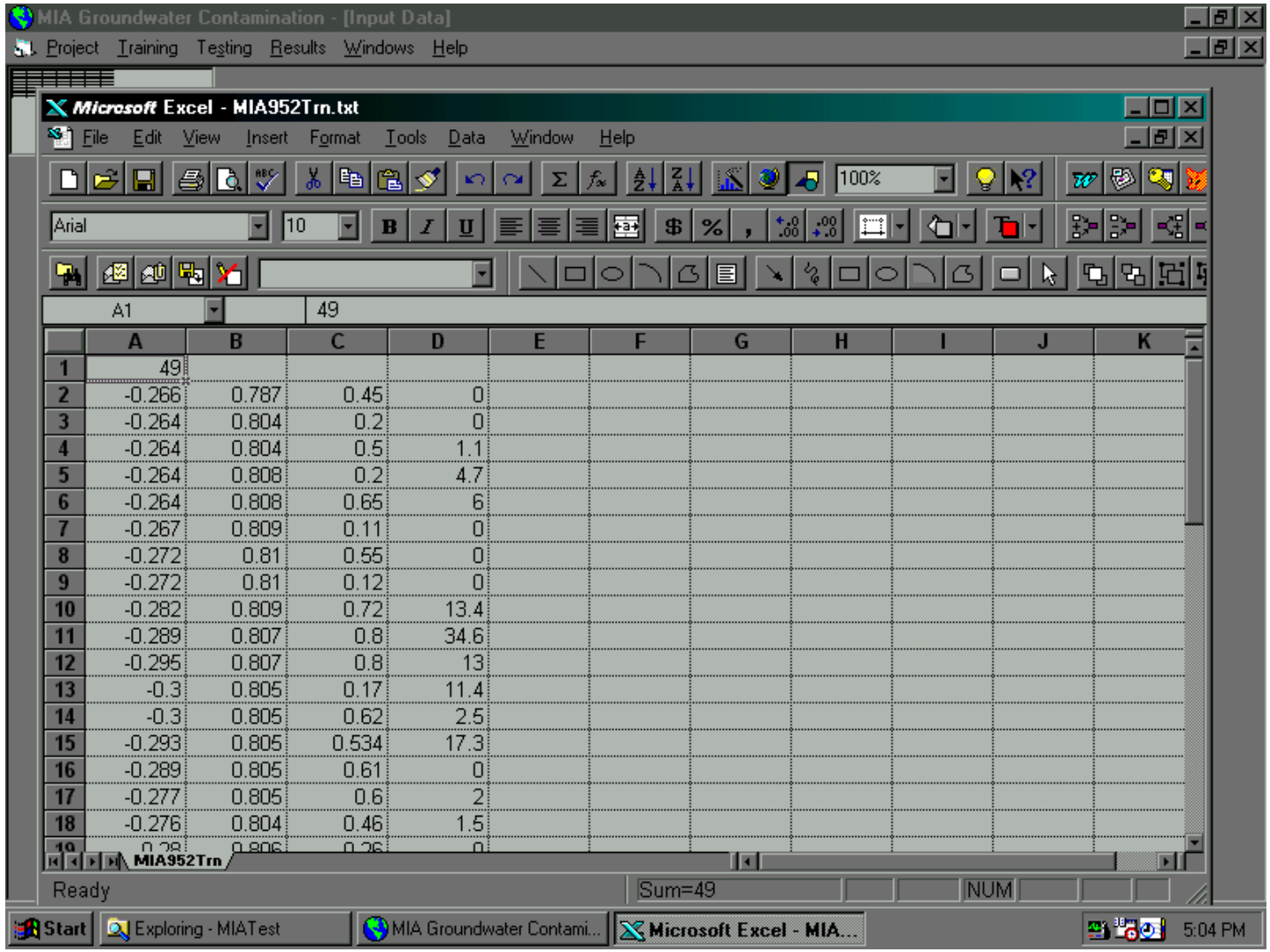


	Number of cases	Error
Training	961	0.833%
Testing	1681	0.823%

Results and Discussion

- The graphical displays demonstrate the following:
  - The Excel-based data input
  - The digitized map of a simulated site
  - Three gray-scale plots of the UCVP

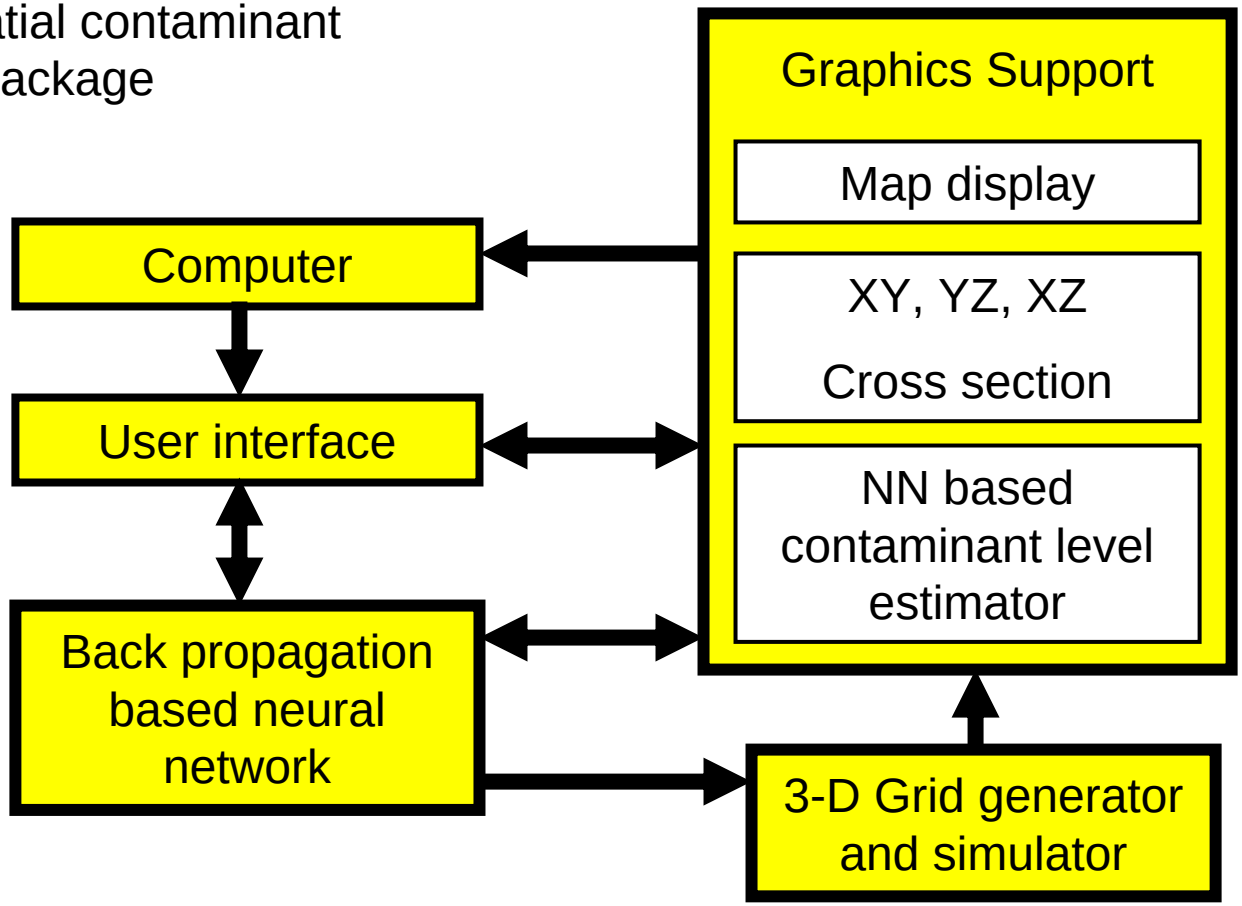
Results and Discussion



	A	B	C	D
1	49			
2	-0.266	0.787	0.45	0
3	-0.264	0.804	0.2	0
4	-0.264	0.804	0.5	1.1
5	-0.264	0.808	0.2	4.7
6	-0.264	0.808	0.65	6
7	-0.267	0.809	0.11	0
8	-0.272	0.81	0.55	0
9	-0.272	0.81	0.12	0
10	-0.282	0.809	0.72	13.4
11	-0.289	0.807	0.8	34.6
12	-0.295	0.807	0.8	13
13	-0.3	0.805	0.17	11.4
14	-0.3	0.805	0.62	2.5
15	-0.293	0.805	0.534	17.3
16	-0.289	0.805	0.61	0
17	-0.277	0.805	0.6	2
18	-0.276	0.804	0.46	1.5
19	-0.28	0.806	0.26	0

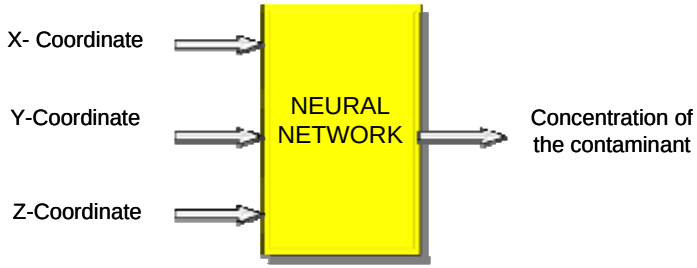
Results and Discussion

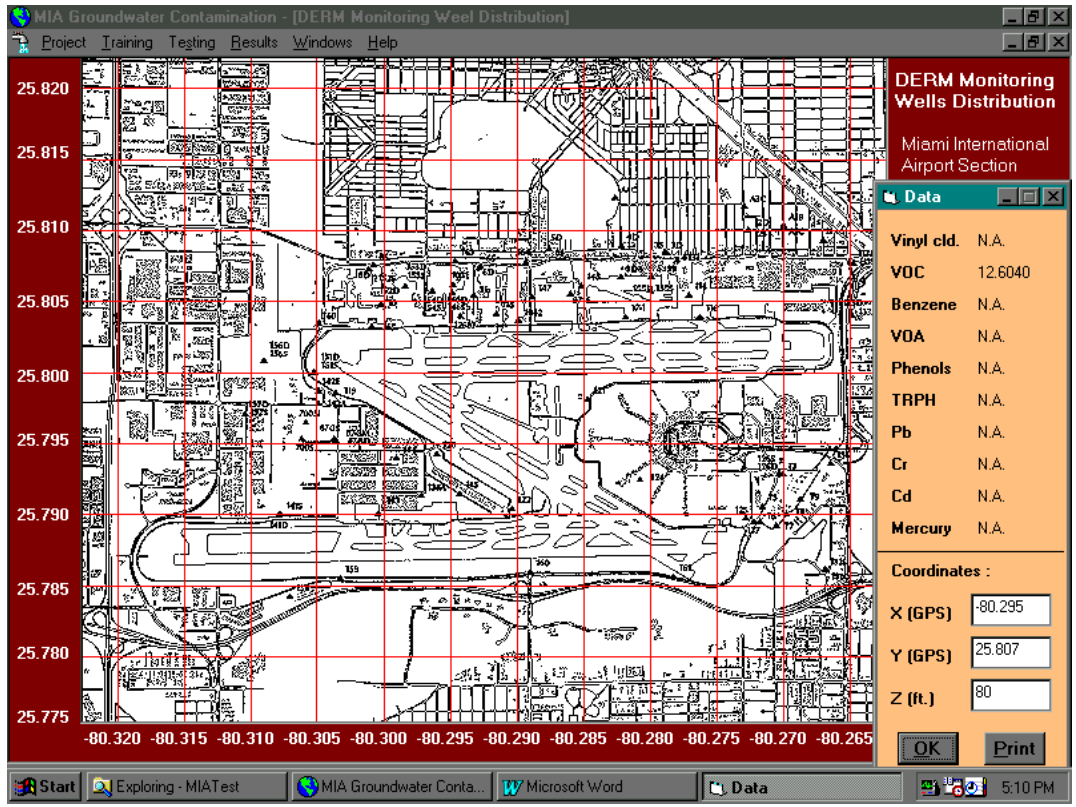
BP-NNgeospatial contaminant visualization package (GCVP)



Results and Discussion

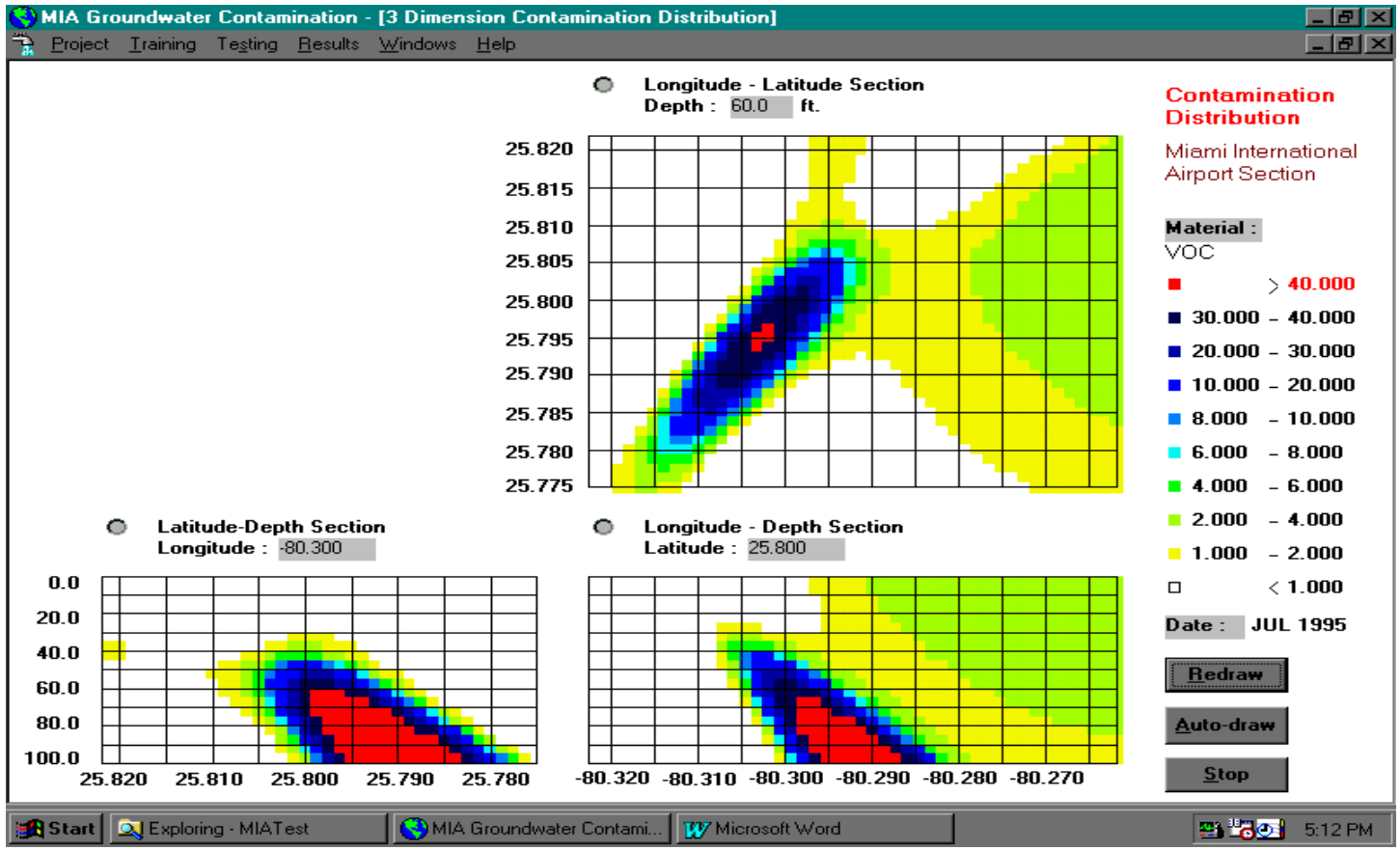
- Display



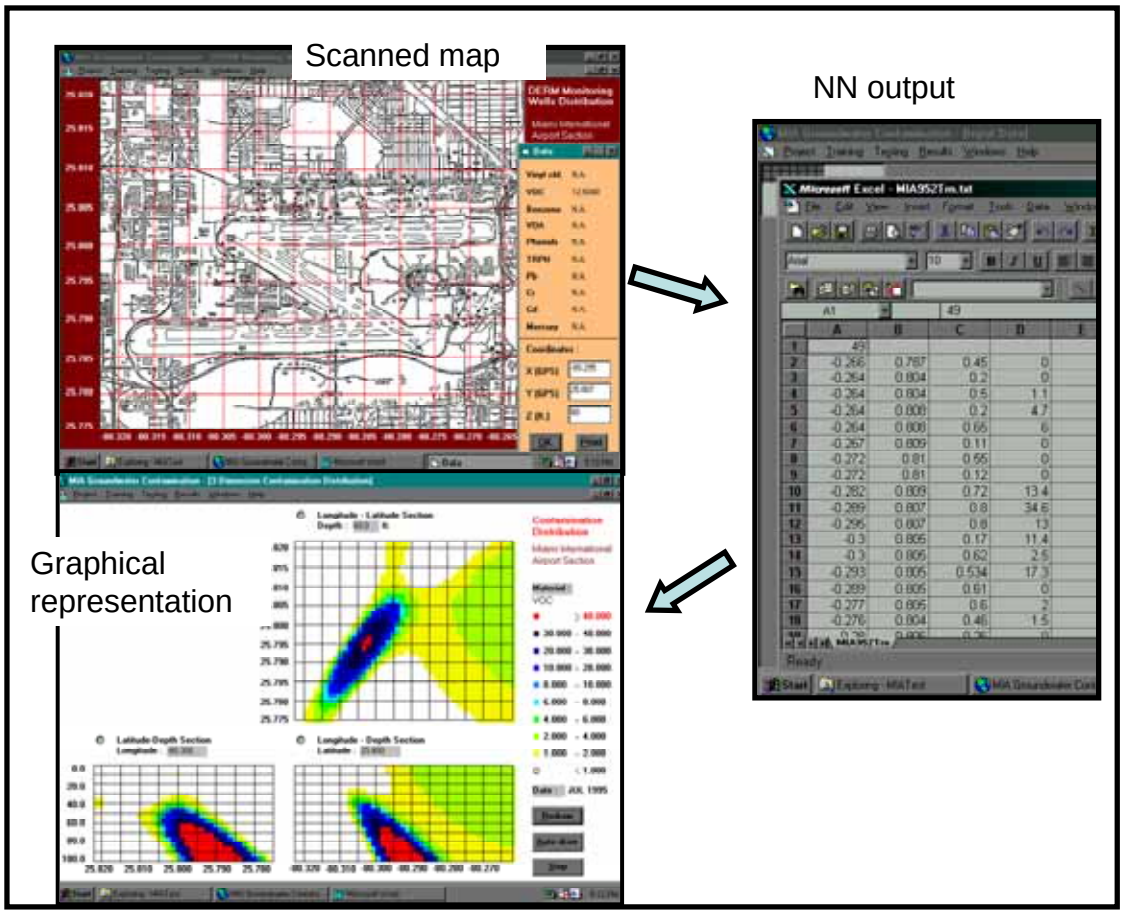


Results and Discussion

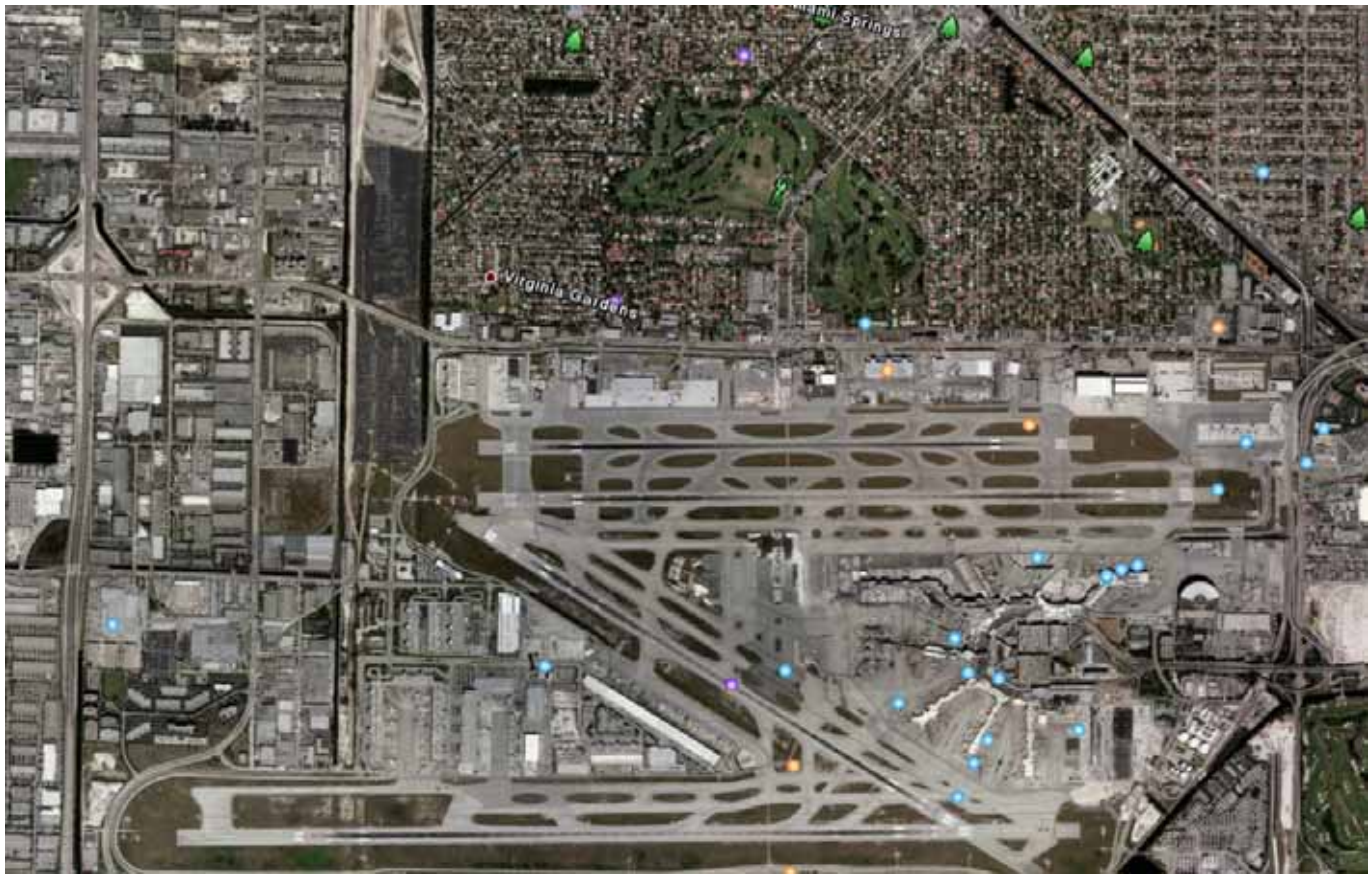
- Display



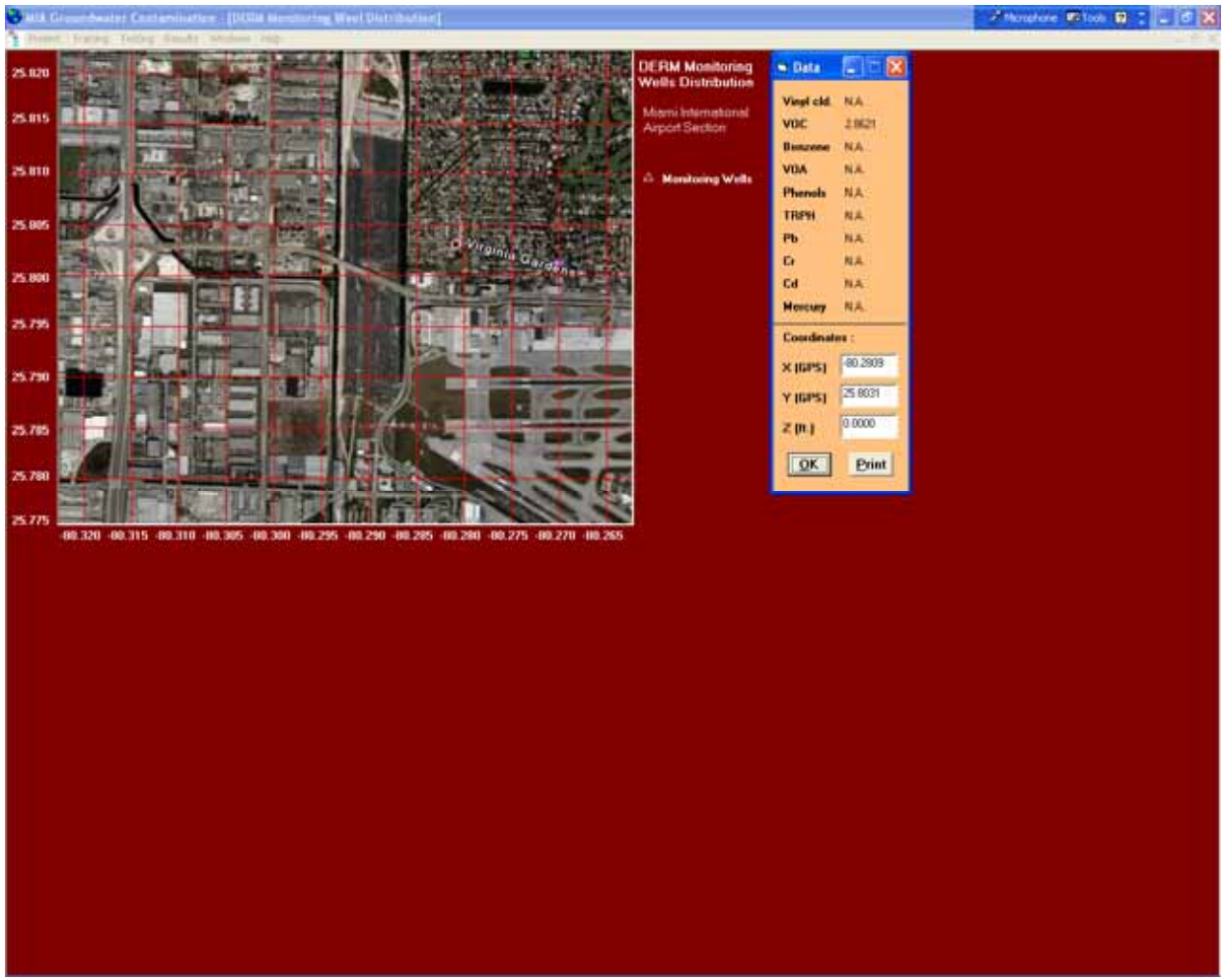
Results and Discussion



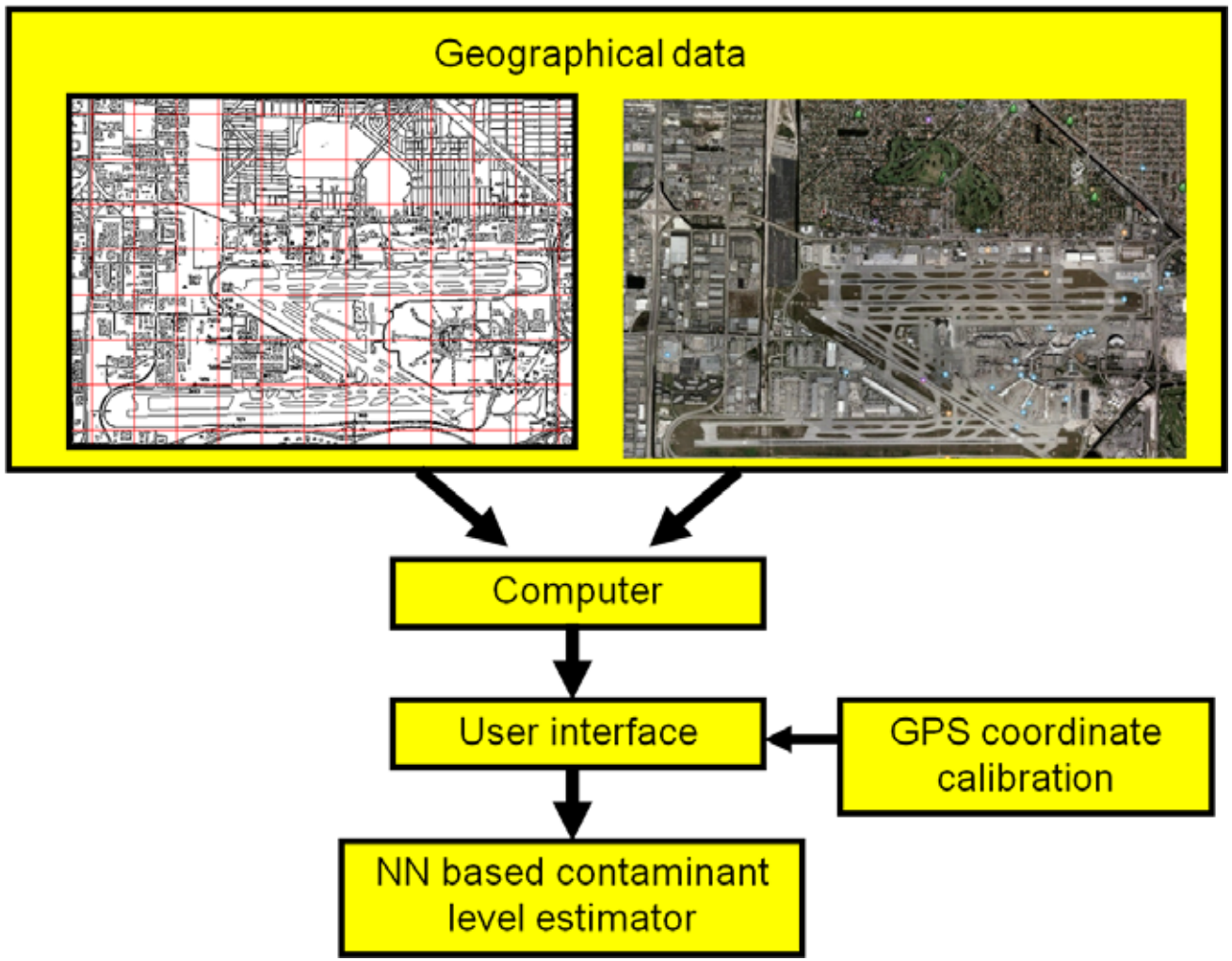
Results and Discussion



Results and Discussion



Results and Discussion



Conclusions

- Neural networks learned the 2 input – 1 output simulated data accurately:
  - Less than 1% error on the test data it never saw before

- NN are an excellent mapping tool:
  - For estimation at any desired location
  - Compression (Thousands of data of a database may be represented with the parameters of a NN)
  - Generation of contamination level distribution with any desired resolution
  - Excellent computational efficiency once the model is estimated

Conclusions

- NN effectively represented the underground contamination – no need for considering any analytical model
- Based on the estimated contamination intensity profiles NN performed well

Conclusions

- The GCVP allows the user to visualize the contamination easily
  - Contamination at any concerned location
  - Variation of the intensity at any cross-section
- The information may be enhanced with the data obtained from satellites
  - Allowing the user to correlate contamination with satellite visible light pictures
  - Additional information may be obtained from satellite images obtained at different wavelengths particularly if the surface and vegetation is effected

Conclusions

- NN are very promising tools for effective analysis of Geospatial data
  - For representation of information in a compact way
  - Correlate the data from different databases to estimate other parameters
  - Estimation of performance, process time and feasibility by training the NN with historic data