

A user friendly environment for the Spectral Analysis of Geomagnetic Data

Y.N.T.Seshagiri Rao

National Geophysical Research Institute, Hyderabad - 500 007

ABSTRACT

This paper describes a user-friendly Graphical User Interface also called Windows (You click and get something) environment or platform and computational paradigms developed using Java programming language for the spectral analysis of Geomagnetic Data. For the spectral analysis Digital Signal Processing methods such as Fast Fourier Transform (FFT), Cosine and Sine Transforms are used. The FFT transforms a function from the time domain to the frequency domain and is an extremely important and widely used method of extracting useful information from sampled signals. In the design of the system, the methodology used is Unified Modeling Language (UML), an advanced flowchart that uses Use Case Diagrams (UCD). The background and motivation is the automation of the entire process of obtaining spectra analysis.

INTRODUCTION

The system developed aims at solving geomagnetic data spectrum analysis problems in the windows environment. The observatory records the time variations of the earth's magnetic field (Jacobs 1987 and Parkinson 1983), which give us valuable information on the electrical and magnetic state of the upper atmosphere, crust and deep interior of earth. The users require frequency values of the magnetic signal for their work. By using FFT the users get the values of the frequency domain about the availability of the given signal they needed and the details of the wave whether it is Quiet Day, Storm Day, and Solar Flare. Using these frequency values, they calculate the impedance values that are useful in solid earth research. In this system, we use Fast Fourier Transform to convert time domain data into frequency domain under Java Platform in order to increase accuracy, efficiency and to lessen human errors as well as the time of signal processing.

It is interesting and important to develop an automatic system for doing the spectral analysis. Studies of long periodicities in the geomagnetic field can be approached from either a time or frequency domain point of view. This paper presents a power spectrum analysis of data. The advantage of this type of analysis is that the frequency components of the entire spectrum can be investigated. This data is transformed into frequency domain using the FFT program. The naïve approach for using FFT is hard and time consuming. But using the developed system, you can get things by pressing a button.

In previous approaches FFT was written in FORTRAN and FORTRAN was not good enough to develop the GUI system. The Java is a general purpose and universal programming language that came into existence lately. It is used for developing both GUI and computations.

RELATED WORK

After thorough search on the Internet and literature, could not find an integrated Java environment for the spectral analysis of geomagnetic data or any other Geophysical data. One can find many FFT algorithms coded in FORTRAN and C.

SYSTEM DEVELOPMENT

The system mainly consists of two modules: The GUI and computational paradigms. The steps involved in the computational paradigm are the development of algorithm for FFT, Sine and Cosine transforms and coding the algorithms into java language.

FFT ALGORITHM

The Fast Fourier Transform (Cooley & Tukey 1965, Brigham 1988, and Oppenheim & Schaffer 1989) is an optimized computational algorithm to implement the Discrete Fourier Transform to an array of 2^N samples. It allows determining the frequency of a discrete signal, representing the signal in the frequency domain.

The FFT is a mathematical procedure, which can be thought of as transforming a function from time domain to the frequency domain. In digital signal processing, the spectrum of a signal refers to the way energy in the signal is distributed over its various frequency components.

DFT enables to analyze, manipulate and synthesis signals in ways not possible with continuous (analog) Signal Processing. A Discrete signal sequence is a set of values obtained by periodic sampling of a continuous signal in the time domain. In the field of continuous signal processing,

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-j2\pi ft} dt$$

$x(t)$ = some continuous time domain signal is used to transform an expression of continuous time domain function $x(t)$ into a continuous frequency domain function $X(f)$. The evaluation of $X(f)$ expression helps us to determine the frequency content of any practical signal of interest.

Discrete (DFT) is given by,

$$X(m) = \sum_{n=0}^{N-1} x(n) e^{-j2\pi nm/N}$$

$X(n)$ is a discrete sequence of time domain sampled values of continuous variable.

$X(m)$ = the m^{th} DFT output component i.e., $X(0), \dots, X(N-1)$.

m = Index of DFT output in frequency domain, $m = 0, \dots, N-1$.

N = No. of samples of input sequence and the no. of frequency points in DFT output.

The value N is a important parameter because it determines how many input samples are needed, the resolution of frequency domain results and the amount of processing time necessary to calculate the N - point DFT.

The exact frequencies of the different sinusoid depends on both the sampling rate (f_s) at which the original signal was sampled and the number of samples (N).

When the DFT is applied to a discrete signal, the result is a set of sine and cosine coefficients. When sine and cosine waves of appropriate frequencies are multiplied by these coefficients and then added together, the original signal waveform is exactly reconstructed. The sine and cosine waves are the frequency components of the original signal, in the

sense that the signal can be built up from these components. The coefficients determined by the DFT represent the amplitudes of each of these components. The procedure by which the sine and cosine coefficients are calculated is straightforward in principle, although in practice it requires a great deal of computation. To determine each individual coefficient, every one of the sampled values of the signal must be multiplied by the corresponding sampled value of a sine or cosine wave of the appropriate frequency. These products must then be added together, and the result then divided by the number of samples involved giving the value of the coefficient.

If the signal consists of a number of samples N , the DFT requires the calculation of N sine and N cosine coefficients. For each coefficient to be determined, N products of samples of the signal and the appropriate sine or cosine wave must be evaluated and summed. The total number of steps in the computation of the DFT is thus N^2 , each step requiring the evaluation of a sine or cosine function together with a multiplication (and this does not include the calculation of the N products in order to find each coefficient).

Fast Sine and Cosine Transforms

The Fourier transforms of functions can be used to solve differential equations. The most common boundary conditions for the solutions are:

1. They have the values zero at the boundaries, or
2. Their derivatives are zero at the boundaries.

In these instances two more transforms arise naturally, the Sine and Cosine transforms given by,

$$F_k = \sum_{j=1}^{N-1} f_j \sin(\pi jk/N) \quad \text{Sine transform}$$

$$F_k = \sum_{j=0}^{N-1} f_j \cos(\pi jk/N) \quad \text{Cosine transform}$$

where $f_j, j=0, \dots, N-1$ is the data array.

At first blush these appear to be simple the imaginary and real parts respectively of the discrete Fourier Transform. The argument of the sin and cosine differ by a factor of two from the value that would make that so. The sin transform uses sines only as a complete set of functions in the interval from 0 to 2π , and the cosine transform uses cosines only.

In the case of sine transform we extend the data to twice their lengths in such away as to make them

an odd function about $j=N$, with $f_n = 0$,

$$F_{2N-j} = -f_j \quad j=0, \dots, N-1$$

$$F_k = \sum_{j=0}^{2N-1} f_j e^{2\pi i j k / (2N)}$$

the half of this sum from $j=N$ to $j=2N-1$ can be rewritten with the substitution $j=2N-j$.

$$\begin{aligned} \sum_{j=N}^{2N-1} f_j e^{2\pi i j k / (2N)} &= \sum_{j'=1}^N f_{2N-j'} e^{2\pi i (2N-j') k / (2N)} \\ &= -\sum_{j'=0}^{N-1} f_{j'} e^{-2\pi i j' k / (2N)} \end{aligned}$$

This array is of the same dimension as the original. Notice that the first term is symmetric about $j=N/2$ and the second is anti-symmetric. Consequently, when real FFT is applied to y_j , the result has real parts R_k and imaginary parts I_k given by,

$$\begin{aligned} R_k &= \sum_{j=0}^{N-1} y_j \cos(2\pi j k / N) \\ &= \sum_{j=1}^{N-1} (f_j + f_{N-j}) \sin(j\pi/N) \cos(2\pi j k / N) \\ &= \sum_{j=0}^{N-1} 2f_j \sin(j\pi/N) \cos(2\pi j k / N) \\ &= \sum_{j=1}^{N-1} f_j \left\{ \sin \frac{(2k+1)j\pi}{N} - \sin \frac{(2k-1)j\pi}{N} \right\} \\ &= F_{2k+1} - F_{2k-1} \end{aligned}$$

$$\begin{aligned} I_k &= \sum_{j=0}^{N-1} y_j \sin(2\pi j k / N) \\ &= \sum_{j=0}^{N-1} (f_j - f_{N-j}) \frac{1}{2} \sin(2\pi j k / N) \\ &= \sum_{j=0}^{N-1} f_j \sin(2\pi j k / N) \\ &= F_{2k} \end{aligned}$$

Therefore F_k can be determined as follows:

$$F_{2k} = I_k \quad F_{2k+1} = F_{2k-1} + R_k \quad k=0, \dots, (N/2 - 1)$$

The even terms of F_k are thus determined very directly. The odd terms require a recursion, the starting point of which is,

$$F_1 = \sum_{j=0}^{N-1} f_j \sin(j\pi/N)$$

The Sine transform, curiously, is its own inverse. If you apply it twice, you get the original data, but multiplied by a factor of $N/2$.

The Cosine transform is slightly more difficult, but the idea is the same. Now, the auxiliary function is,

$$y_0 = f_0, \quad y_j = \frac{1}{2} (f_j + f_{N-j}) - \sin(j\pi/N)(f_j - f_{N-j})$$

and the same analysis leads to

$$F_{2k} = R_k \quad F_{2k+1} = F_{2k-1} + I_k \quad k=0, \dots, (N/2-1)$$

The starting value for the recursion in this case is

$$F_1 = \sum_{j=0}^{N-1} f_j \cos(j\pi/N)$$

This sum does not appear naturally among the R_k and I_k , and so we accumulate it during the generation of the array y_i .

An additional complication is that the cosine transform is not its own inverse. To derive the inverse of an array F_k , we first compute an array f_l as if the transform were its own inverse,

$$\tilde{f}_l = \sum_{k=0}^{N-1} F_k \cos(\pi k l / N)$$

One easily verifies the relations between f_l and the desired inverse $\tilde{f}_l$,

$$\tilde{f}_0 = N f_0 + \sum_{j \text{ odd}} f_j$$

$$\tilde{f}_0 = \frac{N}{2} f_1 + \sum_{j \text{ even}} f_j \quad \text{for } l \text{ odd}$$

$$\tilde{f}_0 = \frac{N}{2} f_1 + \sum_{j \text{ odd}} f_j \quad \text{for } l \text{ even, } l \neq 0$$

The unknown sums on the right of these equations are determined as follows:

$$C_1 = \sum_{j \text{ even}} \tilde{f}_j = \frac{N}{2} \left[f_0 + \sum_{j=0}^{N-1} f_j \right]$$

$$C_2 = \sum_{l \text{ odd}} \tilde{f}_l = \frac{N}{2} \sum_{j=0}^{N-1} f_j$$

follows that

$$\frac{N}{2} f_0 = C_1 - C_2$$

so,

$$\sum_{j \text{ odd}} f_j = \tilde{f}_0 - 2(C_1 - C_2) \quad \sum_{j \text{ even}} f_j = \frac{2}{N} C_2 - \sum_{j \text{ odd}} f_j$$

Knowing these sums, the desired f_j are now easily recovered from equation.

The FFT is calculated in two parts (Press et al., 1992). The first one transforms the original data array into a bit-reverse order array by applying the bit-reversal method. The second part processes the FFT in $N \cdot \log_2(N)$ operations (application of the Danielson-Lanzcos algorithm). The algorithm starts with an array of complex data. This could be an array of floats in which the data [even index] is the real part and the data [odd index] is the complex part. This algorithm has a complexity of $O(N \cdot \log_2(N))$. Actually, the complexity of the algorithm is a little higher because the data needs to be prepared by bit-reversal. The original array must be transformed in order to perform the Danielson-Lanzcos algorithm. For example, the complex [index] will swap places with the complex [bit-reverse-order-index]. If the index (in binary) is 0b00001, the bit-reverse-order-index will be 0b10000

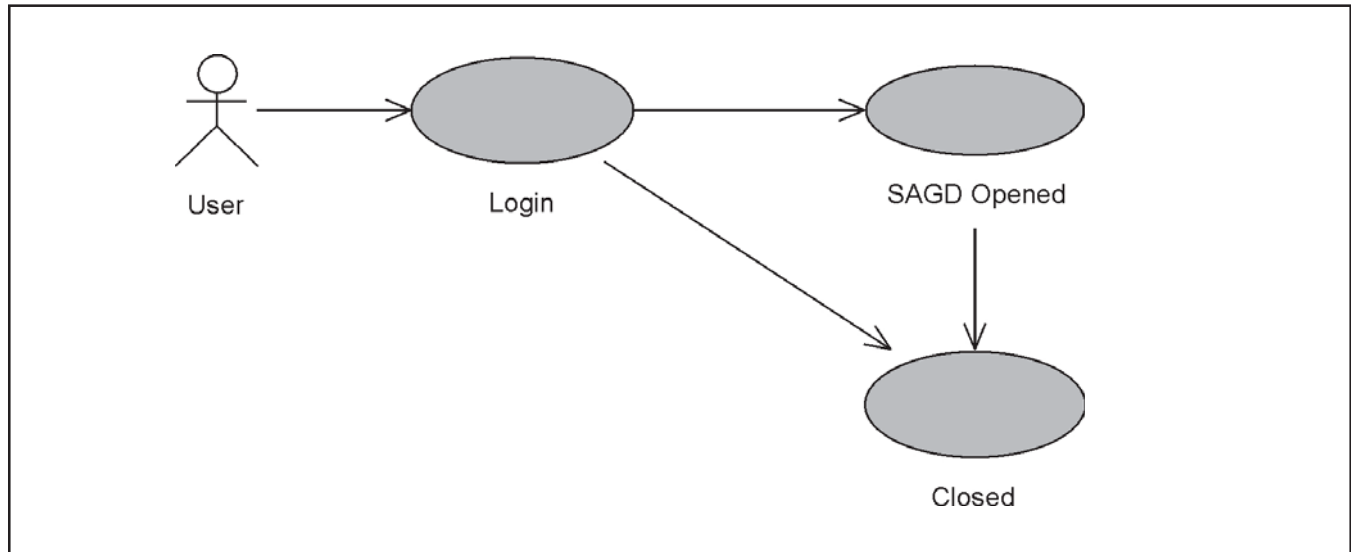
System Design

Computer programming is easier when you use a flowchart. A flowchart is a step-by-step procedure to solve a problem. To write programs in Java, Unified Modeling Language (UML) that uses Use Case Diagrams (UCD) is an efficient way of representing the flow of programming and is an advanced flow chart.

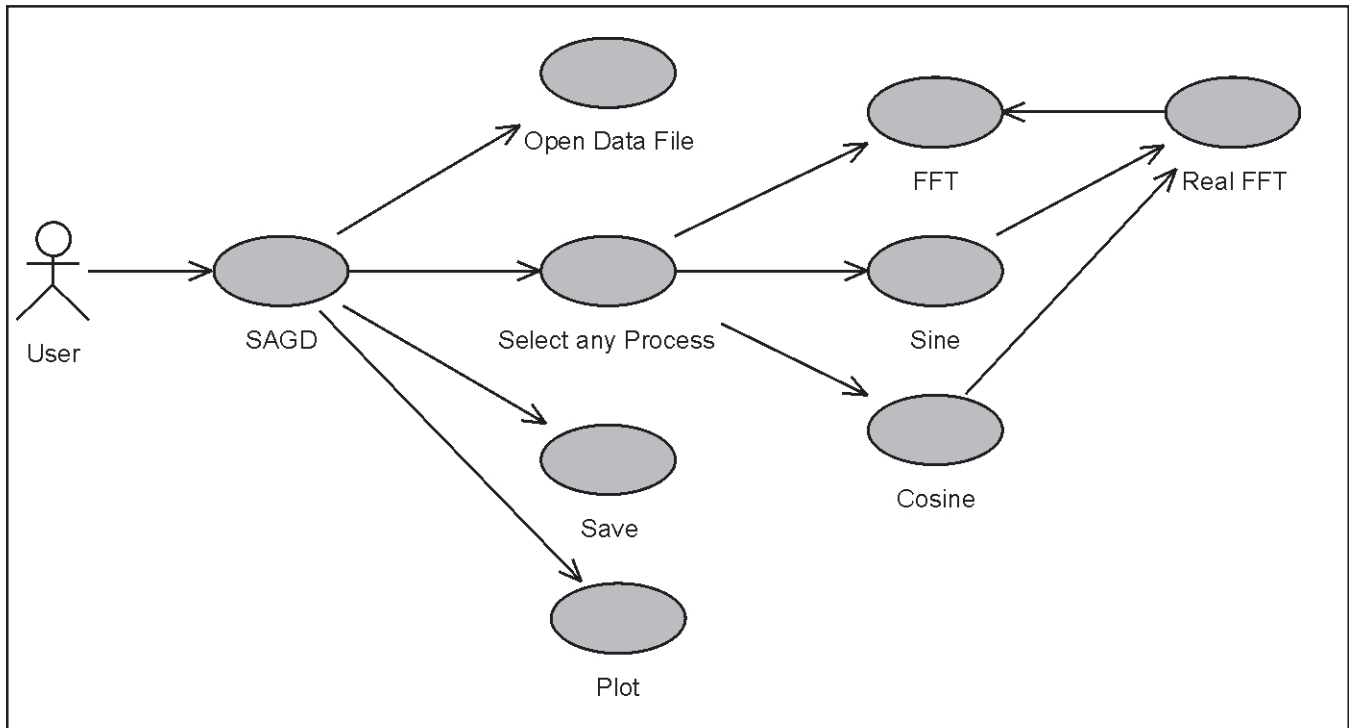
This section presents the design approach for the environment. Unified Modeling Language (UML) (Booch, Rumbaugh & Jacobson 2005) is used in the design of the system. UML is a standard language for specifying, visualizing, constructing and documenting the artifacts of software systems and modeling of large and complex systems. The UML uses mostly graphical notations to express the design of software projects. UML is best preferred for Object oriented programming (OOP) and the system would be implemented in Java, an Object Oriented Programming language.

UML uses Use Case Diagrams (UCD). Use Case diagrams identify the functionality provided by the system (use cases), the users who interact with the system (actors), and the association between the users and the functionality. Use Cases are used in the analysis phase of software development to articulate the high-level requirements of the system.

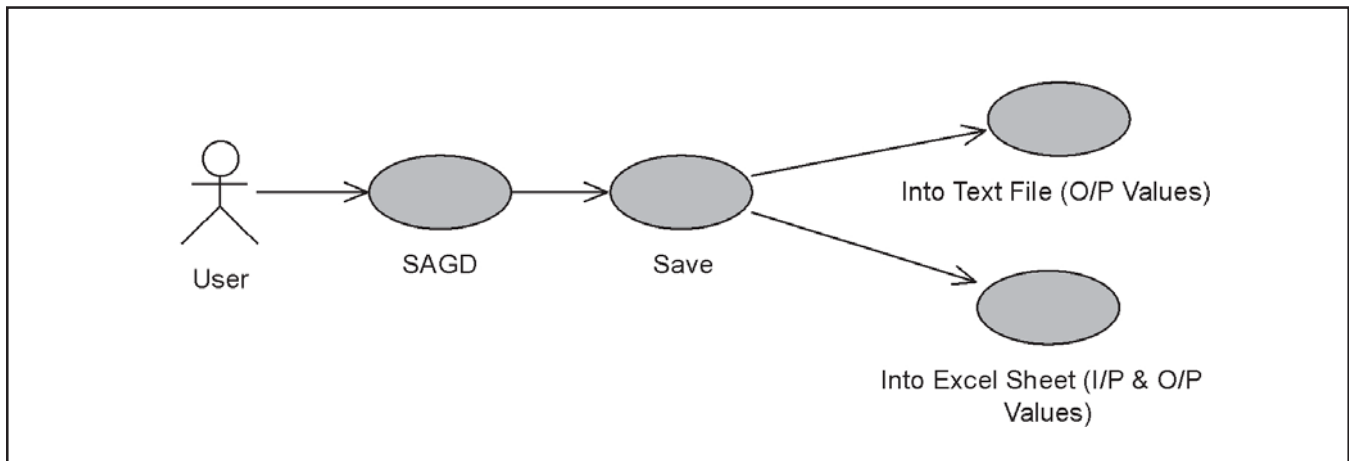
Use Case Diagram for Login



Use Case Diagram for Process

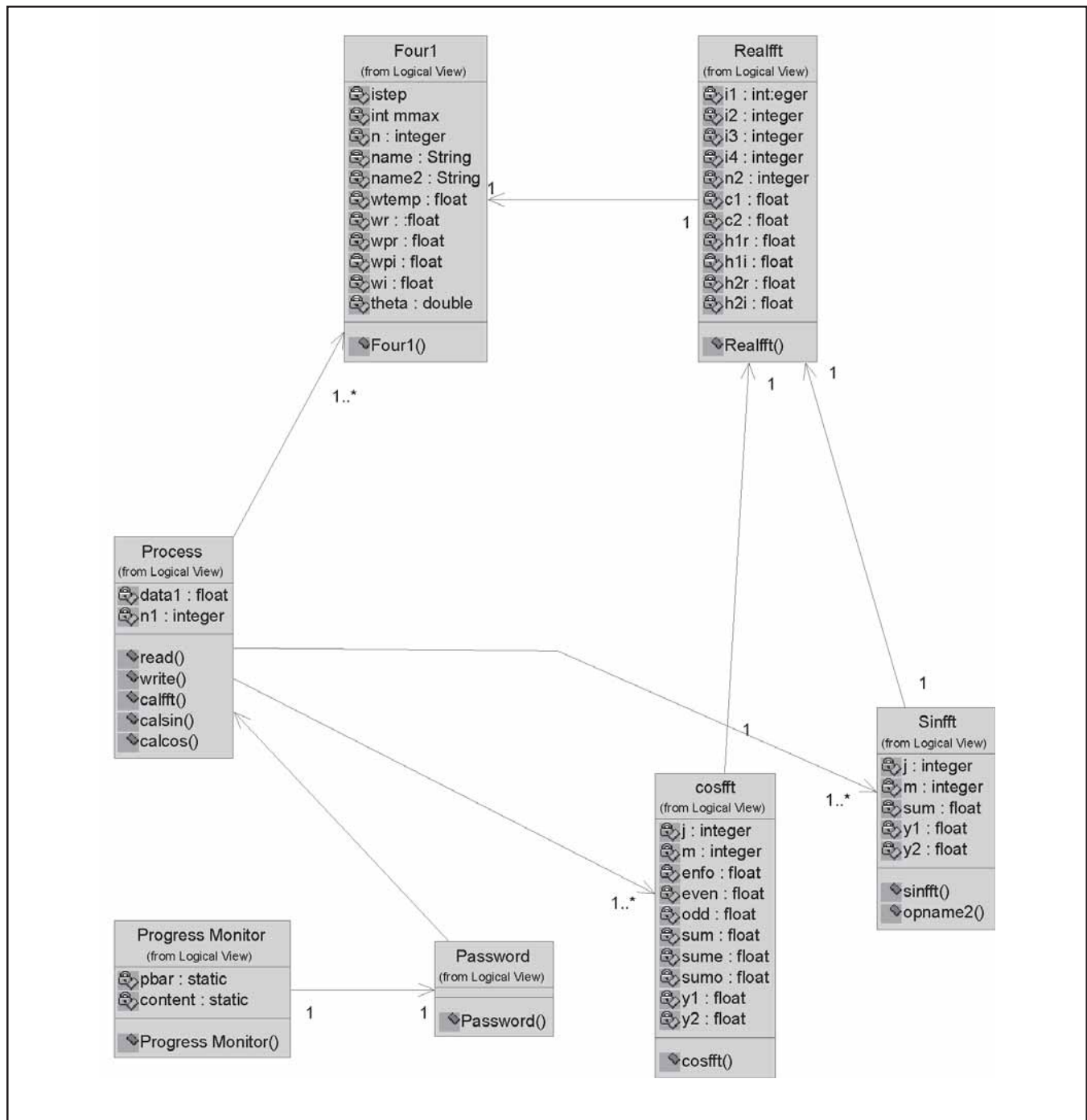


Use Case Diagram for Save



CLASS DIAGRAM:

The Class Diagram is used to refine the Use Case Diagram and define a detailed design of the system. The Class Diagram classifies the actors defined in the Use Case Diagram into a set of interrelated classes. Each Class in the Class Diagram may be capable of providing certain functionalities. These functionalities provided by the Class are termed “Methods” of the Class. Apart from this, each Class may have certain “Attributes” that uniquely identify the Class. A class defines the abstract characteristics of a thing, including the thing’s characteristics (its attributes or properties) and the things it can do (its behaviors or methods) or features).



IMPLEMENTATION AND PERFORMANCE

In the implementation stage of the project Java programming (Roberts, Heller & Ernest 2001 and Deitel & Deitel 2005) is used. Java a universal programming language and has an extensive library of routines. It also supports network computing. Java is intended for writing programs that must be readable in a variety of ways. Java puts a lot of

emphasis on early checking for possible problems, later dynamic checking, and eliminating situations that are error prone.

The Java compiler detects many problems that in other languages would only show up at runtime. The Java interpreters can execute Java byte codes directly on any machine to which the interpreter has been ported. Java contains classes core to the language contributing to the ease of programming.

With Java you can build Graphical User Interface (windows) for user-friendly operations and computational paradigms. The following sample data is used to test the system for obtaining the spectrum and for testing the data.

Sample Data:

The data collected is the H component from the “LA-COUR VARIOMETER “ generated magnetic signals (D, H, Z). The analog charts are also shown in the figures.

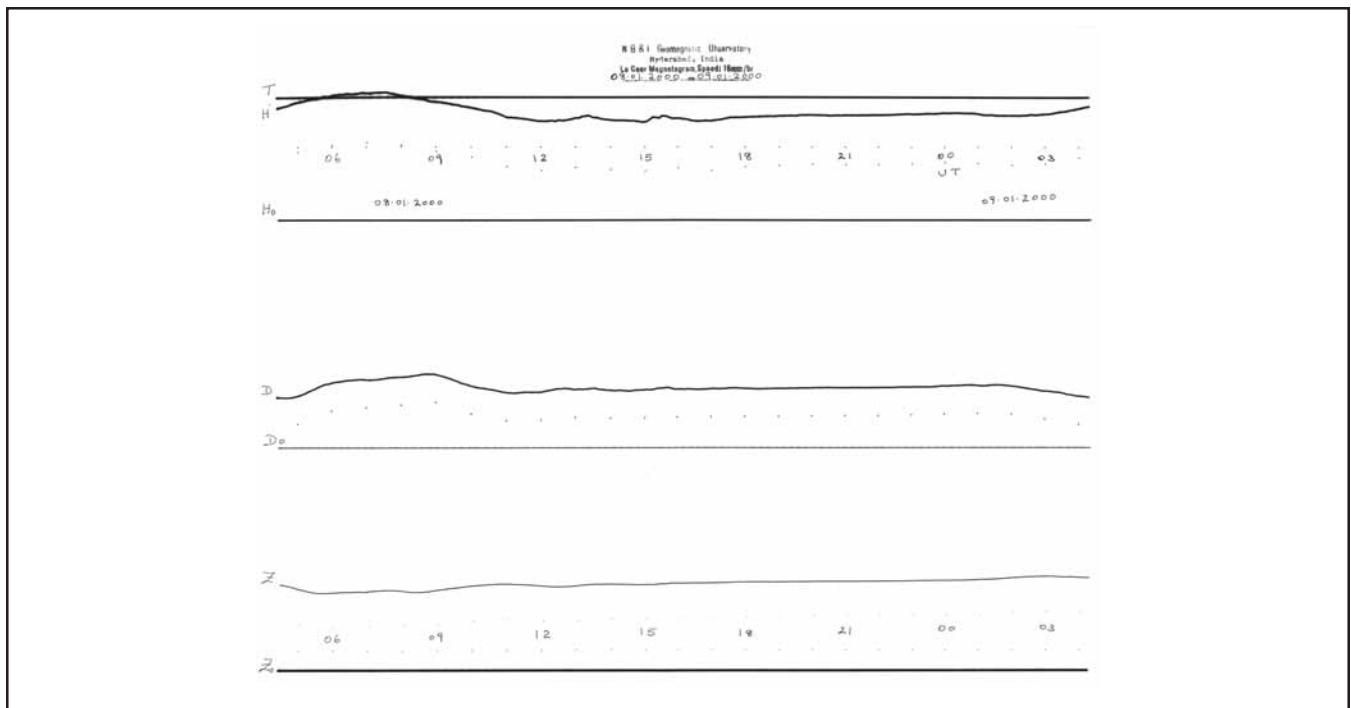
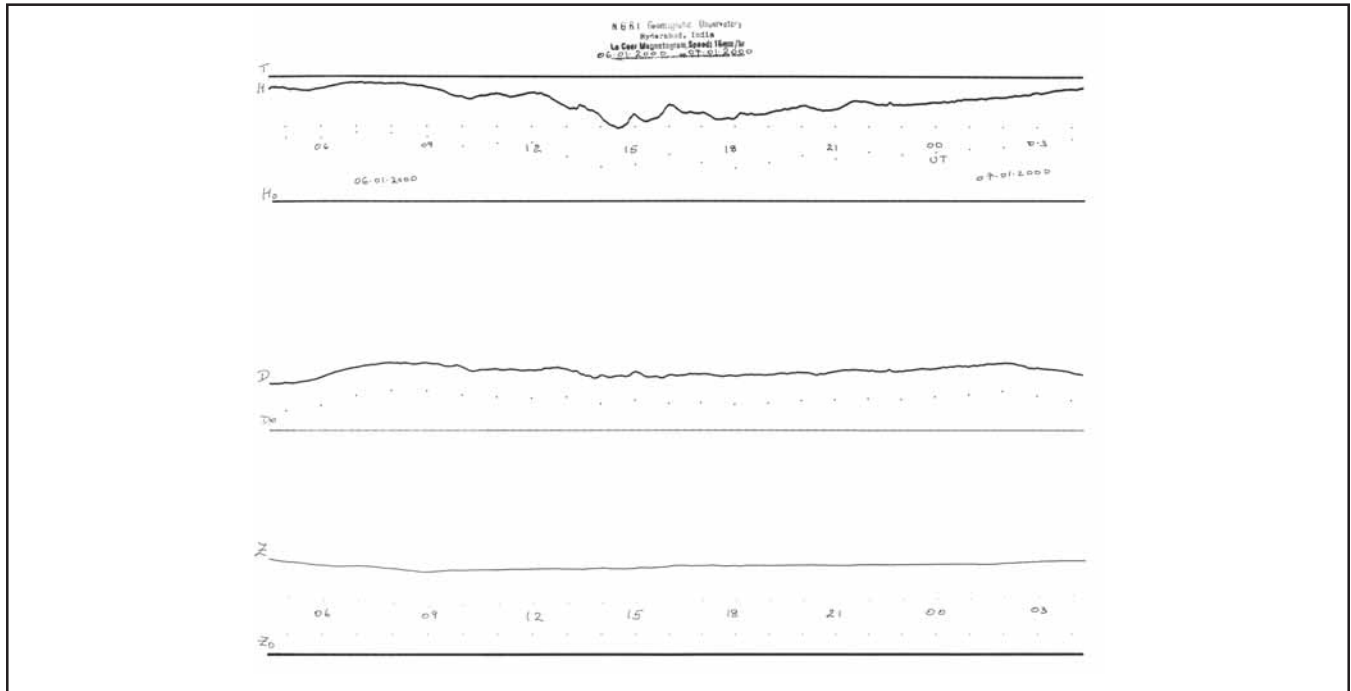
Input Data

48.0	47.8	47.6	47.5	47.6	47.5	47.4	47.4	47.3	47.9	48.3	48.5
48.3	48.0	47.8	47.6	47.6	47.5	47.4	47.2	47.1	47.2	47.4	47.1
47.1	47.1	47.3	47.3	47.4	47.4						

Output Data

Input	FFT output	Sin Transform	Cosine Transform
1.35568E-19	1.64951E-33	1.64951E-33	1.65E-33
2.16927E-33	0.001730573	0	-0.0201
4.13902E-11	0.001858399	5.20663E-12	0
1.706E-07	-0.001724749	-0.0019235	-1.3E-05
5.11125E-07	0.001515063	9.84258E-05	4.88E-07
1.03833E-05	-0.000608912	0.000269517	0.000197
2.16927E-33	-0.001424011	0.00028478	7.04E-06
4.07081E-11	0.001128334	0.00017689	0.000707
1.71531E-07	0.000301156	0.000506755	2.51E-05
4.99145E-10	-0.000887552	0.004431823	-0.00038
1.706E-07	0.000731163	0.000354223	-2.3E-05
2.0445E-06	-0.000485802	0.001864041	-0.00012
1.03833E-05	-0.00128558	0.000192704	-3.1E-05
1.16216E-19	0.001402252	0.001773727	0.000425
1.03829E-05	-0.000769803	6.76446E-05	-4.6E-05
2.16927E-33	0.001280762	0.000544768	0.000163
4.02533E-11	0.00100687	0.000212471	-1.7E-05
1.706E-07	-0.000433692	0.000543091	-0.00018
3.19453E-08	0.001305164	0.000284058	-6E-05
1.03833E-05	-0.000776239	0.001141789	0.000505
4.99145E-10	-0.00031906	0.000154395	-3.6E-05
2.59601E-06	0.000616375	0.001580485	-8.1E-05
2.16922E-33	-0.000280704	-6.2894E-05	-7.8E-05
4.04807E-11	5.0564E-05	0.000632272	-0.00018
1.69669E-07	0.000677925	-9.6299E-05	-0.00011
3.2712E-05	-0.000335178	0.000547115	0.000254
6.48972E-07	-0.000218606	-8.4492E-05	-0.00016
2.16912E-33	0.000541899	0.000975354	0.000303
4.16176E-11	-8.07978E-05	-0.00014241	-5.9E-05
4.25874E-31	0.000134924	0.000297113	1.75E-05

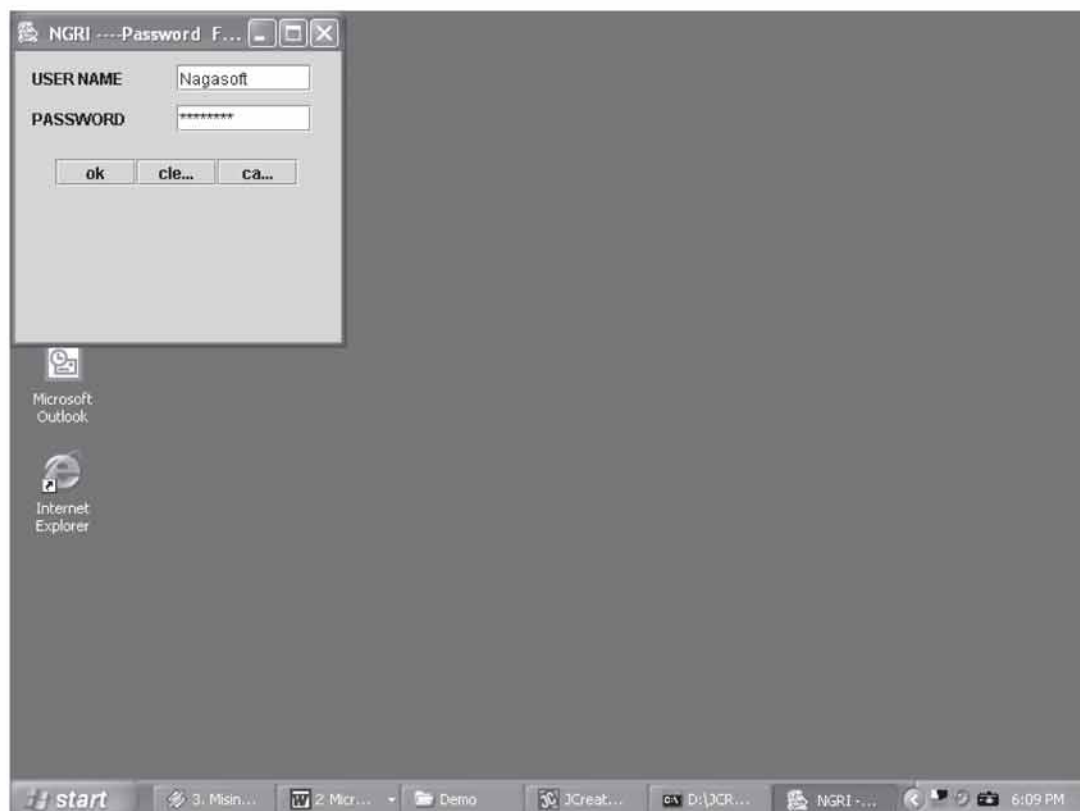
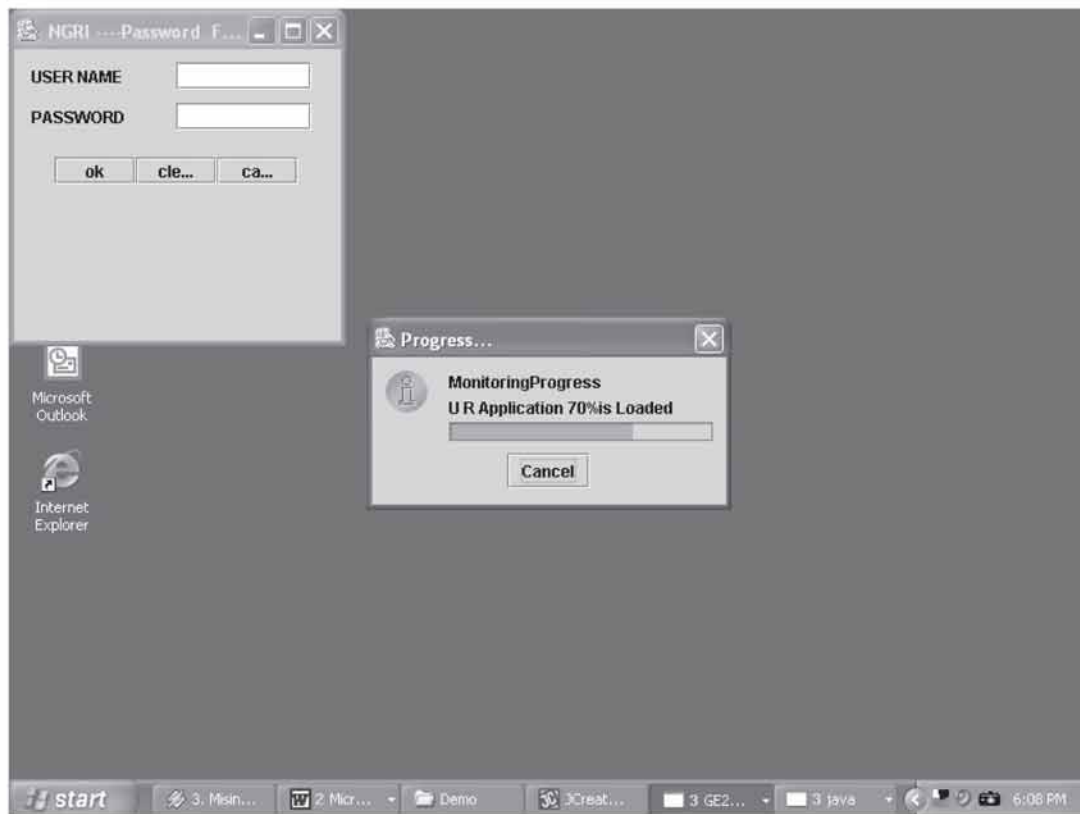
Analog Data

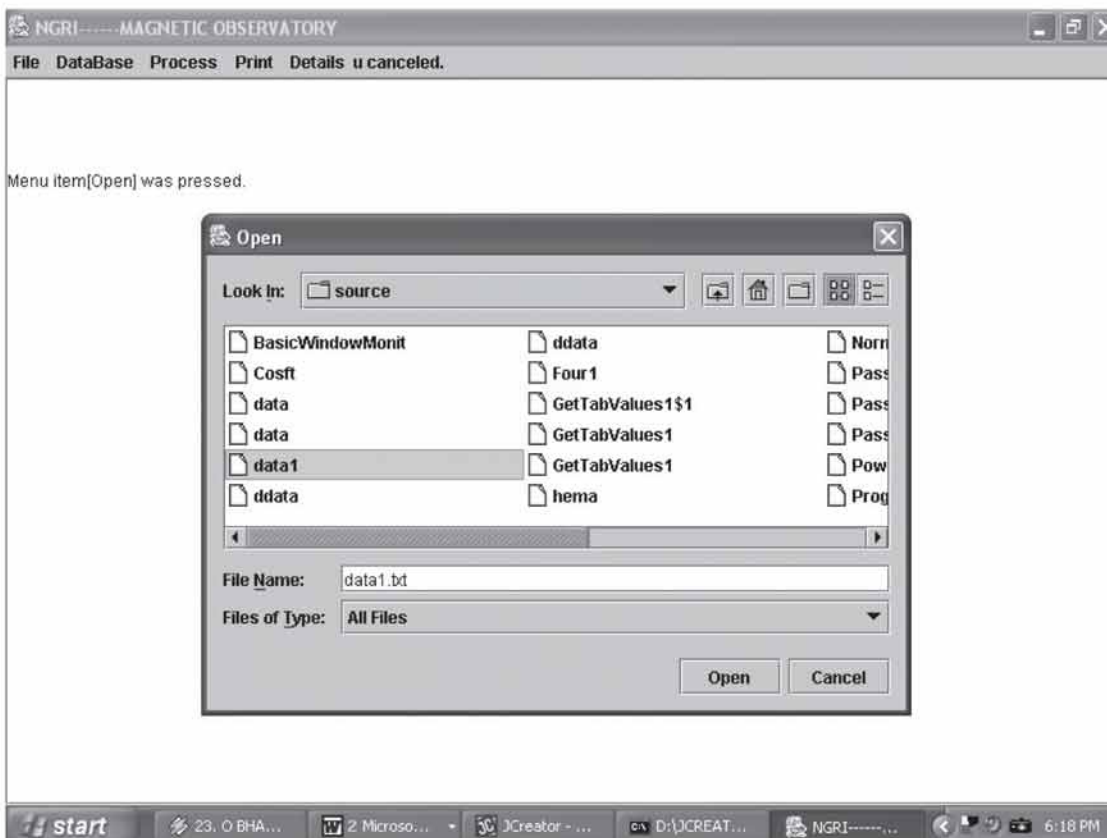
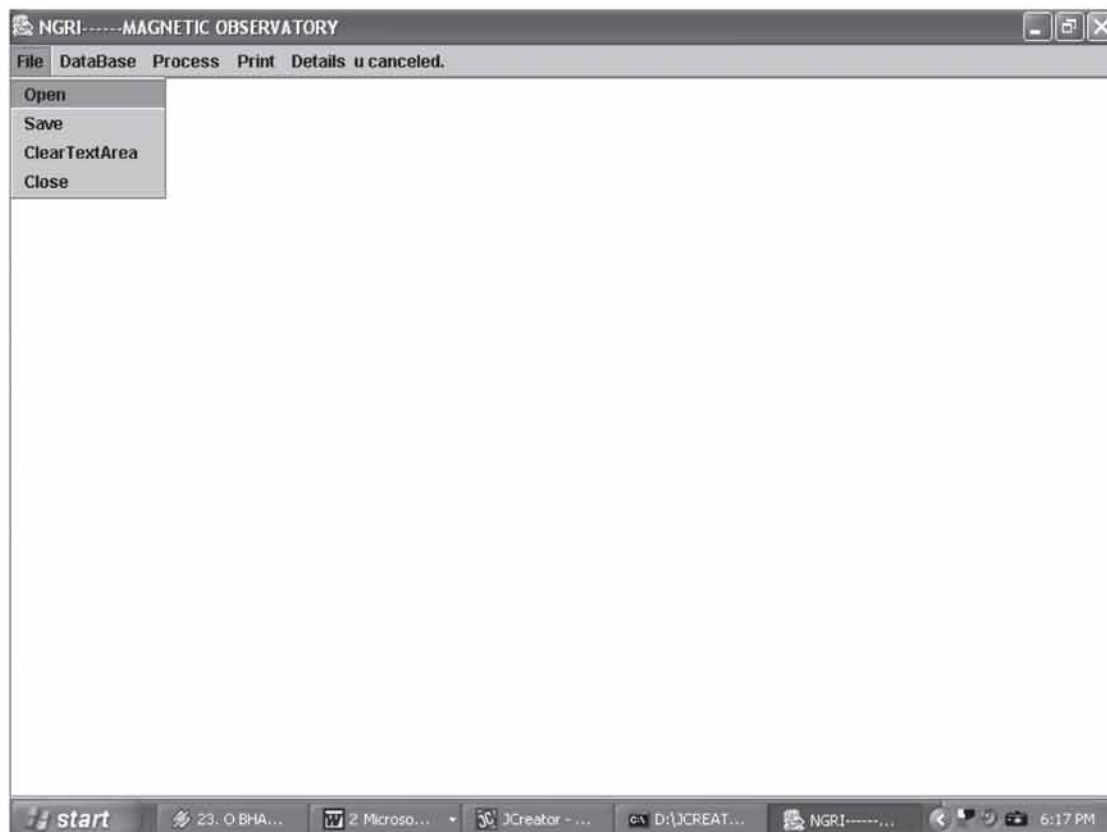


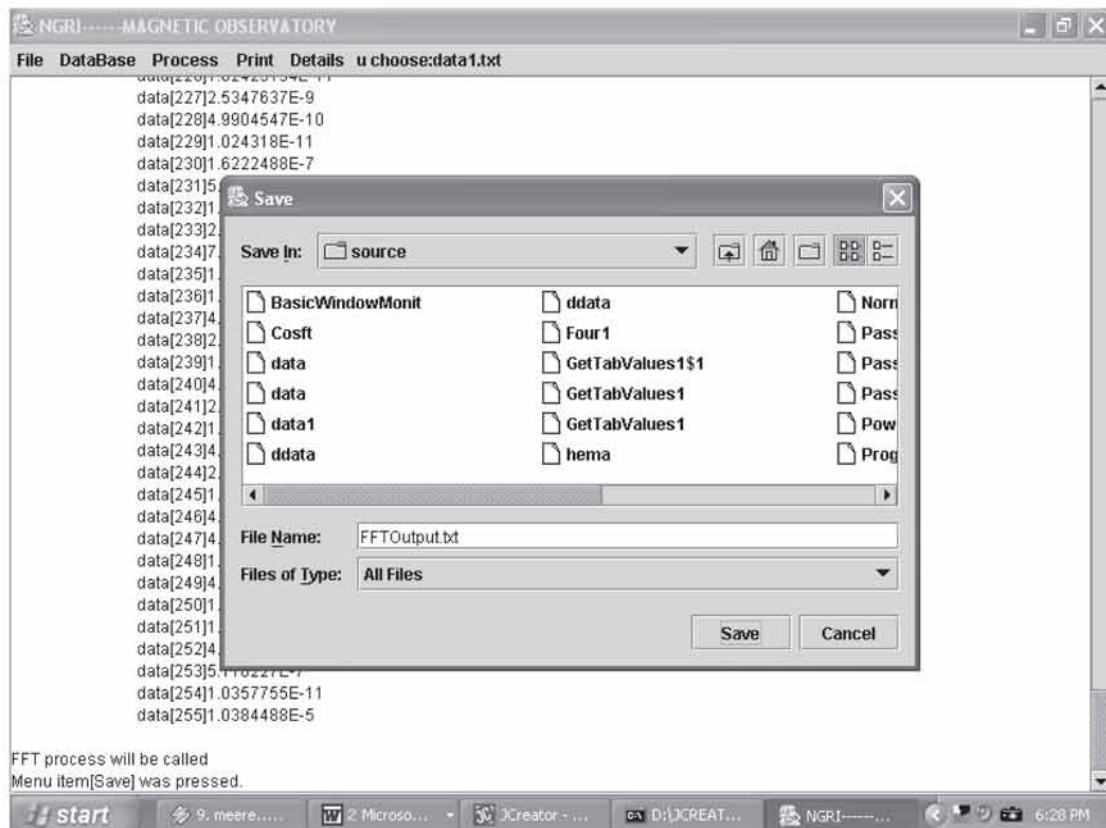
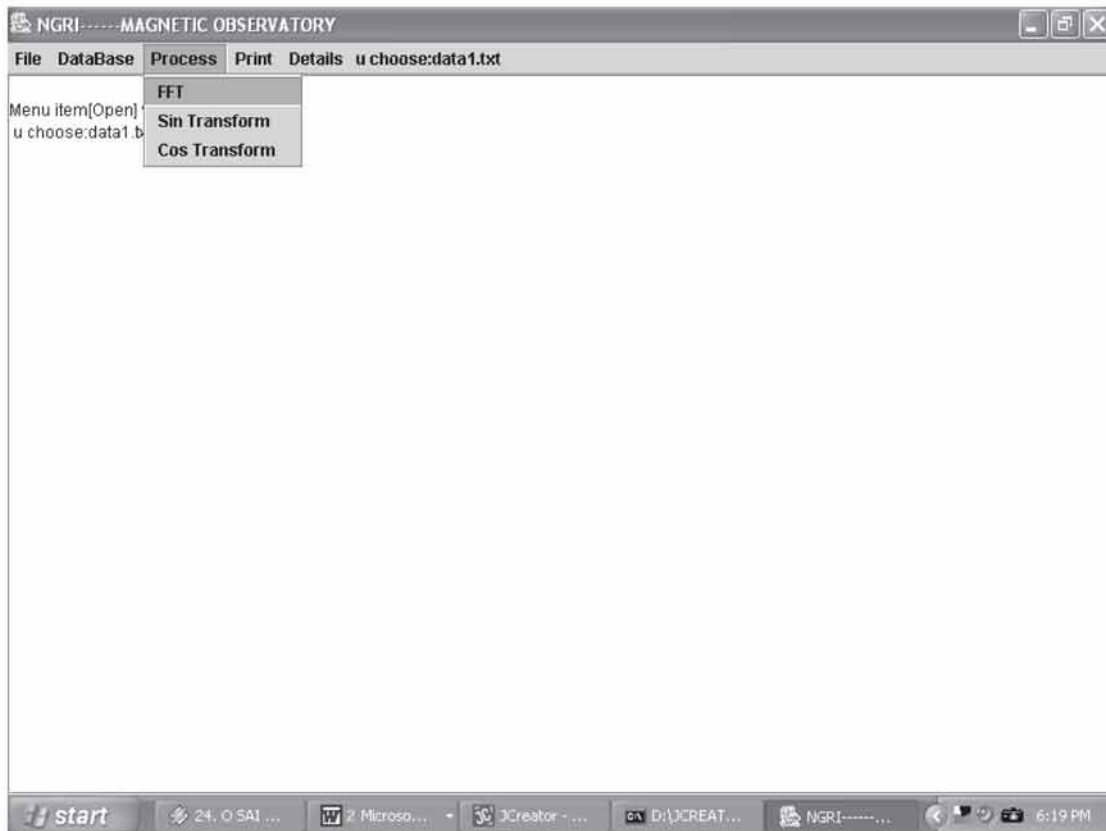
FUTURE WORK

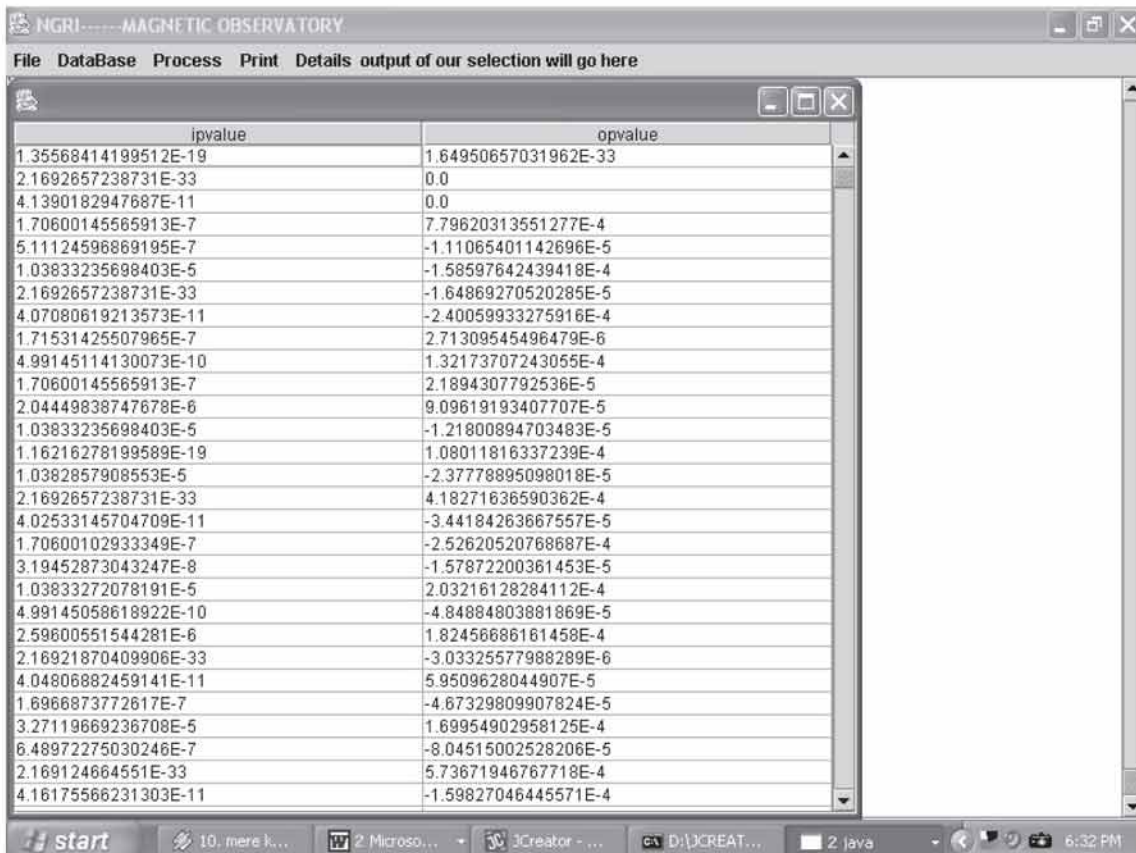
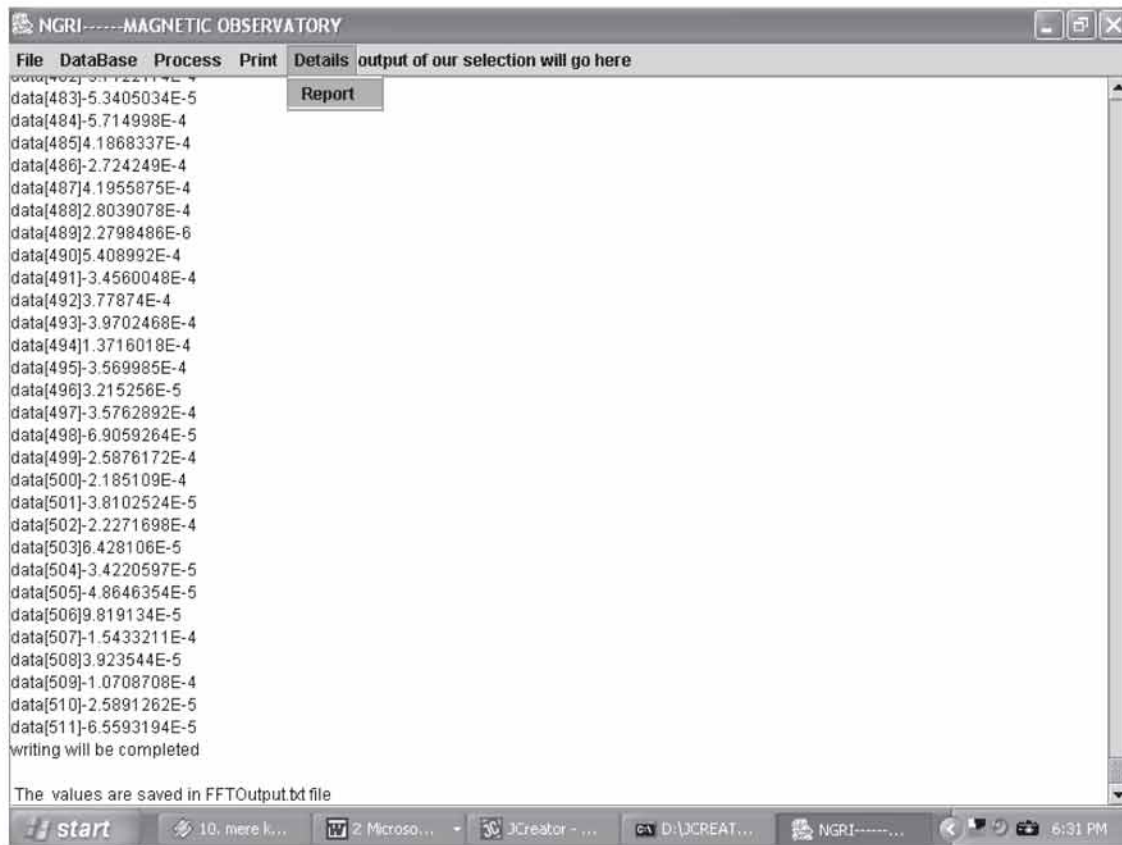
With slight modifications, the integrated environment can be extended to other geophysical data as seismic data, MT data, gravity data and so on depending on the nature of data. A comprehensive geophysical data analysis system can be aimed at to include and switch over from one kind of data to another.

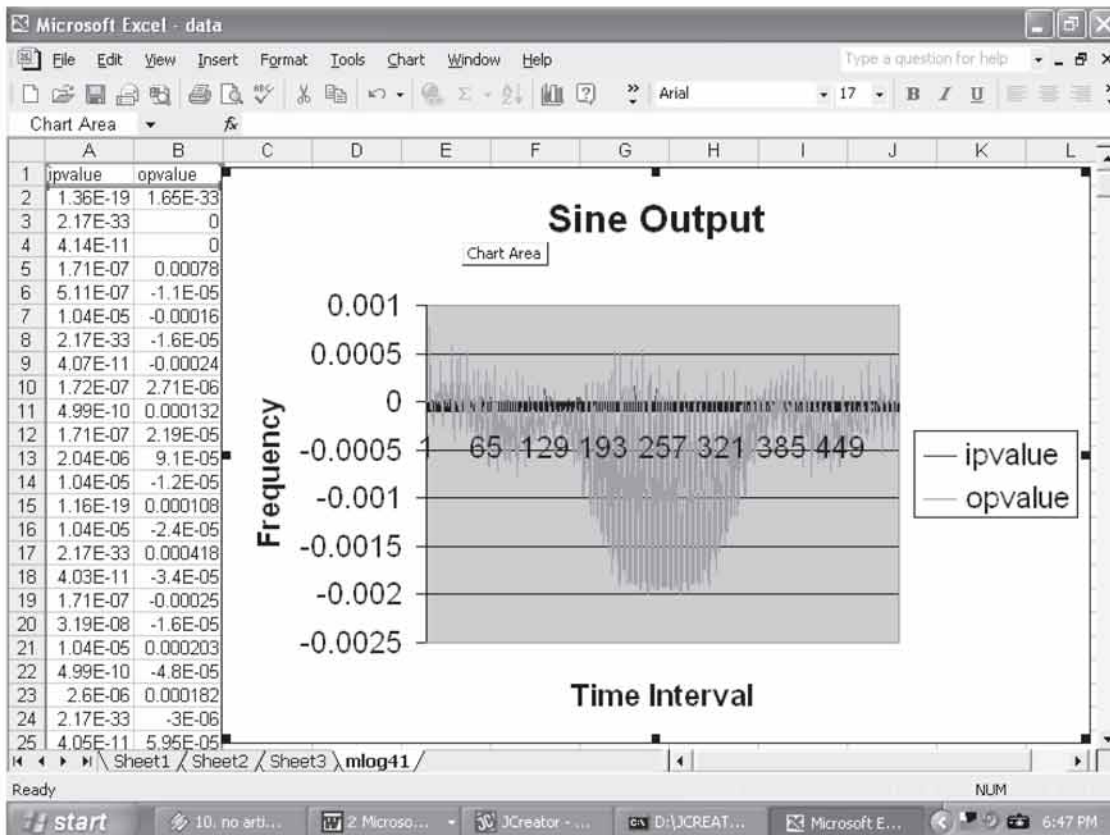
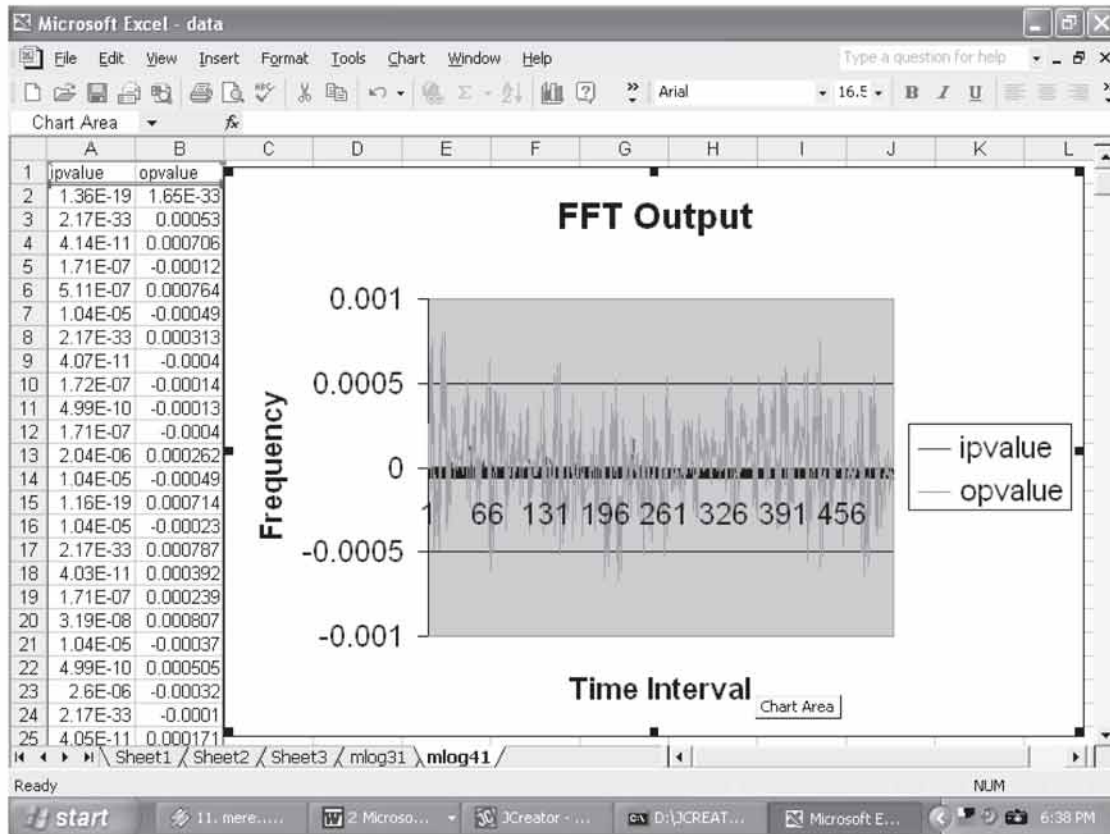
SCREENS for interfacing, inputting the data and obtaining the results.

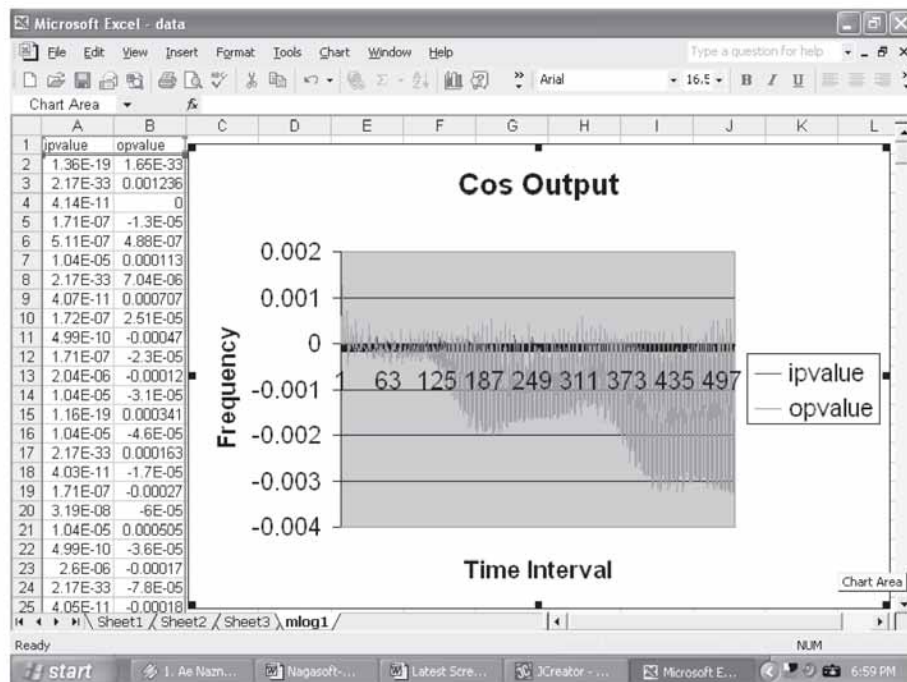












CONCLUSIONS

From the environment created the geomagnetic data signal can be converted in time domain to frequency domain using the FFT, Sine, Cosine, and the power spectrum functions. By conversion of the original signal into frequency values the impedance values can be computed. From those values one can determine the ground characteristics and whether the data is quiet data or disturbed data. One can also find the presence of magnetic storms, their amplitude and phases (Start, Middle, End) etc.

Java is used because of its capability for the development of user-friendly environment such as Graphical User Interface and computations.

FUTURE WORK

This environment can be adaptable with slight modifications to other geophysical data as seismic data, MT data, gravity data and so on.

REFERENCES

Booch, G., Rumbaugh, J. & Jacobson, I., 2005. Unified

Modeling Language User Guide, The (2nd Edition), Addison-Wesley Object Technology Series.

Brigham, E. O., 1988. The Fast Fourier Transform and Applications, Englewood Cliffs, NJ: Prentice Hall.

Cooley, J. W. & Tukey, J. W. "An algorithm for the machine computation of the complex Fourier series," *Mathematics of Computation*, 19, 297-301, Apr. 1965

Deitel, D. H. & Deitel, P. J., 2005. Java How to Program, 5th Edition, Pearson Education.

Jacobs, J. A., (Ed.) Geomagnetism, Vol.1, 1987, Academic Press.

Oppenheim, A. V. & Schaffer, R. W., 1989. Discrete-Time Signal Processing, Prentice Hall, Englewood Cliffs, New Jersey 07632.

Parkinson, W. D., Introduction to Geomagnetism, 1983, Scottish Academic Press.

Press, W. H., Vetterling, W. T., Teukolsky, S. A. & Flannery, B.P., 1992. Numerical Recipes in C: The Art of Scientific Computing, Second Edition, Cambridge University Press.

Roberts, S., Heller, P. & Ernest, M., 2001. Complete Java2 Certification Guide, Second Edition, B P B Publications.

(Accepted 1st September 2006. Received 16th August 2006; in original form 21st July 2006)



Y.N.T. Seshagiri Rao did B.E. (Electronics & Communications) from Govt. (JNTU) College of Engineering, Kakinada, A.P. during 1964-69 and was a National merit scholarship holder. He has been with the National Geophysical Research Institute, Hyderabad since 1973. He worked in the areas of Geophysical Instrumentation, Artificial Neural Networks, Scientific Data Warehousing and wrote software in Java Language and delivered lectures.