

A NEURO-GENETIC APPROACH FOR RAPID ASSESSMENT OF ABOVE GROUND BIOMASS: AN IMPROVED TOOL FOR MONITORING THE IMPACT OF FOREST DEGRADATION

ABHIJEET KUMAR PARMAR
MARCH, 2012

SUPERVISORS:

Dr. Sarnam Singh
Dr. Y.A. (Yousif) Hussin



**A NEURO-GENETIC APPROACH FOR RAPID ASSESSMENT
OF ABOVE GROUND BIOMASS:
AN IMPROVED TOOL FOR MONITORING THE IMPACT OF
FOREST DEGRADATION**

ABHIJEET KUMAR PARMAR

Enschede, The Netherlands, March, 2012

Thesis submitted to the Faculty of Geo-Information Science and Earth
Observation of the University of Twente in partial fulfilment of the
requirements for the degree of Master of Science in Geo-information
Science and Earth Observation.

Specialization: Natural Hazards and Disaster Risk Management

SUPERVISORS:

Dr. Sarnam Singh
Dr. Y.A. (Yousif) Hussin

THESIS ASSESSMENT BOARD:

Prof. Dr. Ir. A. Alfred Stein (Chair)
Shri M.L. Srivastava, Joint Director, FSI (External Examiner)
Prof. Dr. V.G. (Victor) Jetten (ITC)
Dr. C.J. (Cees) van Westen (ITC)
Dr. Y.A. (Yousif) Hussin (Supervisor)
Dr. Sarnam Singh (Supervisor)
Dr. P.K. Champati Ray (Observer)
Dr. (Nicholas) Hamm (Observer)

DISCLAIMER

This document describes work undertaken as part of a programme of study at the Faculty of Geo-Information Science and Earth Observation of the University of Twente. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the Faculty.

*...To my family
I am lucky to be one of you!*



ABSTRACT

The potential to extract vertical forest structural attributes such as maximum tree height and AGB from full waveform large footprint spaceborne ICESat GLAS LiDAR data has been analysed using neuro-genetic approach of modelling the waveform signal over dry deciduous forests in central part of India. The results showed that the maximum tree height and AGB can be directly modelled from the GLAS waveform derived indices. The study used four parameters viz. *wDistance*, *R75*, *H75* and *R25* derived from LiDAR waveform for predicting AGB and maximum tree height.

The modelling results predicted maximum tree height with a RMSE of 1.8 m compared to the RMSE of 3.4 m and AGB with a RMSE of 29 t/ha compared to 30.45 t/ha as obtained by Sun *et al.*, (2007). Besides waveform modeling constraints, the critical factors that may have affected the model's performance were study area conditions like terrain slope, less number of field samples and location of tree with maximum height within the sampled footprint and short stand height.

The main issues confronting the future studies lies in further improving the precision of the prediction model in areas of steep slope, varied forest types and extending the footprint measurements to global estimates.

Keywords: [ICESat GLAS, LiDAR, footprint, AGB, maximum tree height, neuro-genetic approach, *wDistance*, *R75*, *H75*, *R25*]

ACKNOWLEDGEMENTS

Thank you,

My sincerest gratitude for your support, patience and guidance:

- ② Dr. Sarnam Singh, Academic Supervisor, IIRS, for his continuous support and guidance of my MSc research, for his extraordinary scientific intuition which made him a constant source of ideas and inspirations.
- ② Dr. Yousif A. Hussin, Academic Supervisor, ITC, University of Twente, for his advice, supervision and immense knowledge, for his critical comments and suggestions for improving my thesis.
- ② Dr. P.S. Roy, Director, IIRS, for providing excellent research environment and infrastructure to carry forward my research.
- ② Dr. P.K. Champati Roy, Head, GSGH Division & IIRS Course Director (NHDRM), for his constant support and providing critical inputs for making this course an invaluable experience.
- ② Drs. Michiel C.J. Damen, ITC Course Director (NHDRM), for imparting his knowledge and experience as well as offering world class education right on our doorsteps.
- ② All IIRS and ITC faculty and staff who transferred their immense knowledge and skill for exceptionally enriching my intellectual maturity.
- ② Dr. Michael Lefsky, Colorado State University, for his willingness to critically review my research method, results and answering my unintelligent questions about computer programming.
- ② Dr. Guoqing Sun, GSFC, NASA and Dr. Hieu Duong, Colorado State University, for their qualified remarks which helped me in refining the research model and results.
- ② David Korn, NSIDC User Service, University of Colorado, for providing data and crucial inputs for data processing and analysis.
- ② ITC Alumnus, Mr. David Gwenzi, Mr. Junjie Zhang and Mr. Irfan Akhtar Iqbal for their invaluable advice and discussion throughout my research phase.

For your time and cooperation:

- ② Mr. Lalji Singh Rawat, Director, Madhav National Park and Mr. D.S. Chauhan, Kuno Palpur Wildlife Sanctuary, for facilitating the field work arrangements.
- ② Computer Maintenance Cell, IIRS, for their excellent software and hardware support.
- ② Ram Singh, Mohan, Najim and all other forest staff for assisting me in field data collection.

For your faithful proof reading skills:

- ② Mr. Apurv Chauhan, my oldest friend, for his precious time to proof read this thesis and giving his invaluable insights about the composition and structure of the document.

To my friends, old, new and future, throughout the globe

And above all, for your love and support

Mom, Sister, and Brother Amit

Sincerely,

(Abhijeet Kumar Parmar)

TABLE OF CONTENTS

1. INTRODUCTION	1
1.1. Background.....	1
1.2. Research context.....	1
1.2.1. A review of global forest biomass assessment methods.....	1
1.2.2. Advent of LiDAR remote sensing for forest biomass and carbon stock studies.....	2
1.2.3. ICESat GLAS full waveform large footprint LiDAR in forestry.....	4
1.2.4. Neuro – Genetic Approach for extracting vegetation variables from LiDAR measurements	5
1.2.4.1. Artificial neural networks	6
1.2.4.2. Evolutionary algorithm and Evolutionary artificial neural network.....	8
1.3. Problem statement	9
1.4. Research objectives	10
1.5. Outline of the thesis	11
2. MATERIALS AND METHODS.....	13
2.1. Study area.....	13
2.1.1. Selection of the study area.....	13
2.1.2. Description of the study area	13
2.2. Data available	15
2.2.1. ICESat GLAS data.....	15
2.2.2. Other reference dataset and materials	15
2.3. Field data collection	16
2.4. Research method	17
2.4.1. GLAS waveform processing	18
2.4.1.1. GLAS waveform pre-processing.....	18
2.4.1.2. Waveform conversion.....	18
2.4.1.3. Waveform normalization.....	18
2.4.1.4. Detection of effective waveform signal	20
2.4.1.5. Waveform smoothing	21
2.4.1.6. GLAS waveform initial parameter estimation	22
2.4.1.7. GLAS waveform Gaussian fitting.....	23
2.4.1.8. GLAS waveform parameters derivation.....	25
2.5. Data analysis	28
2.5.1. Neuro-Genetic Model Prediction.....	28
2.5.2. Validation	28
2.5.3. AGB change trend analysis for monitoring forest degradation	28
3. RESULTS AND DISCUSSION.....	29
3.1. Preliminary analysis of GLAS derived waveform parameters	29
3.2. Neural Network input parameter optimization using EA	30
3.2.1. Results.....	30
3.2.2. Validation	31
3.3. Neuro-Genetic Approach	33
3.3.1. Multi-Layer Perceptron (MLP) Neural Network.....	33
3.3.1.1. Results and validation	33

3.3.2. EANN based AGB & tree height prediction model	36
3.3.2.1. Results and validation.....	36
3.4. Comparative analysis of GLAS predicted AGB and AGB estimates from past inventory	37
3.5. Discussion.....	39
4. CONCLUSIONS AND RECOMMENDATIONS	43
4.1. Conclusions	43
4.2. Recommendations	44
REFERENCE.....	45
APPENDICES.....	50

LIST OF FIGURES

Figure 1-1: Overview of Discrete Return LiDAR and Waveform LiDAR Systems	3
Figure 1-2: ICESat GLAS Waveform	4
Figure 1-3: An elementary ANN structure.....	6
Figure 1-4: Feedforward MLP Network Topology.....	7
Figure 1-5: Process flow of EA.....	8
Figure 1-6: Total Biomass Carbon and Uncertainty in their Estimation.....	10
Figure 2-1: Location of the Study Area in India showing GLAS Footprints Track.....	14
Figure 2-2: Sampling Plots within GLAS Footprint	16
Figure 2-3: Research Method Flow Diagram.....	17
Figure 2-4: GLAS Waveform before Normalization.....	19
Figure 2-5: GLAS Waveform after Normalization	19
Figure 2-6: Actual GLAS Waveform signal defined based on Threshold Value	20
Figure 2-7: Noisy GLAS Waveform.....	21
Figure 2-8: Smoothed GLAS Waveform.....	22
Figure 2-9: Identified Peaks (red circle) in the GLAS Waveform	23
Figure 2-10: Results of the Gaussian Decomposition of the GLAS Waveform.....	24
Figure 2-11: GLAS Waveform Parameters as Noise Threshold Level, Signal Start, Signal End, Ground Return and Canopy Height	27
Figure 2-12: GLAS Waveform Quartile Based Parameters [H25, H50 and H75]	27
Figure 3-1: Scatter Plot of the GLAS Waveform Derived Parameters.....	29
Figure 3-2: Input and Target Parameter for EA.....	30
Figure 3-3: Best Team Selection Result	30
Figure 3-4: Input and Target Parameters for Validation.....	32
Figure 3-5: Best Team Selection Result	32
Figure 3-6: MLP Network Topology for Tree Height and AGB Prediction.....	33
Figure 3-7: GLAS Height <i>vs</i> Field Height Predicted by MPL.....	33
Figure 3-8: Relative Importance of Waveform Derived Input Parameters.....	34
Figure 3-9: GLAS Derived AGB <i>vs</i> Field AGB.....	35
Figure 3-10: Relative Importance of GLAS Derived Indices for AGB Estimation.....	35
Figure 3-11: GLAS Height <i>vs</i> Field Height Predicted by EANN.....	36
Figure 3-12: GLAS Derived AGB <i>vs</i> Field Measured AGB by EANN Model.....	37
Figure 3-13: AGB Estimate Comparison (Past Estimate <i>vs</i> GLAS Predicted).....	38
Figure 3-14: Field photographs showing illegal cutting of tress	38
Figure 3-15: R ² <i>vs</i> Number of Generation Plot for AGB and Tree Height Prediction Model	41

LIST OF TABLES

Table 2-1: ICESat GLAS Specifications.....	15
Table 2-2: Description of GLAS waveform derived parameters	25
Table 3-1: Generation Impact Factor of GLAS Derived Waveform Parameter.....	31
Table 3-2: GLAS Derived Waveform Parameters Input Combinations.....	31
Table 3-3: GLAS Derived Waveform Parameters Input Combinations (Validation).....	32
Table 3-4: MLP Tree Height Predictor Network Statistics	34
Table 3-5: MLP AGB Predictor Network Statistics	34
Table 3-6: EANN Tree Height Prediction Model Statistics.....	36
Table 3-7: EANN based AGB Predictor Model Statistics	37

LIST OF ABBREVIATIONS

Abbreviations	Expansions
AGB	Above G round B iomass
ANN	A rtificial N eural N etwork
ASCII	A merican S tandard C ode for I nformation I nterchange
BRDF	B idirectional R eflectance D istribution F unction
DBH	D iameter at B reast H eight
DGA	D ynamic G enetic A lgorithm
DRD	D iscrete R eturn D evelops
EA	E volutionary A lgorithm
EANN	E volutionary A rtificial N eural N etwork
EO	E arth O bservation
FWHM	F ull W idth at H alf M aximum
GLAS	G eoscience L aser A ltimeter S ystem
GPS	G lobal P ositioning S ystem
GA	G enetic A lgorithm
HOME	H eight O f M edian E nergy
ICESat	I ce, C loud, and L and E levation S atellite
IDL	I nteractive D ata L anguage
IUFRO	I nternational U nion of F orest R esearch O rganizations
LAI	L eaf A rea I ndex
LASER	L ight A mplification by S timulated E mission of R adiation
LiDAR	L ight D etection and R anging
MLP	M ulti L ayer P erceptron
MSL	M ean S ea L evel
NASA	N ational A eronautics and S pace A dministration
NDVI	N ormalized D ifference V egetation I ndex
Nd:YAG	N eodymium: Y ttrium A luminium G arnet
NGAT	N SIDC G LAS A ltimetry E levation E xtraction T ool
NIR	N ear I nfra- R ed
NSIDC	N ational S now a nd I ce D ata C enter
PolInSAR	P olarimetric I nterferometry S ynthetic A perture R adar
RBF	R adial B asis F unction
RMSE	R oot M ean S quare E rror
SAR	S ynthetic A perture R adar
VCL	V egetation C anopy L iDAR
VHR	V ery H igh R esolution
VI	V irtual I nstrument
WRD	W aveform R ecording D evelops

1. INTRODUCTION

1.1. Background

Role of forest in climate change has incited the need of accurate carbon pool estimation for their scientific management (Kwak *et al.*, 2005). Accurate estimates of forest carbon stock are necessary to understand the global carbon dynamics. There is a growing need for development of systematic and cost effective methodology for global forest inventory to estimate carbon stock and analyse land use changes (Rosenqvist *et al.*, 2003). The precise quantification of forest carbon stock will result into efficient carbon accounting and trading system for effective implementation of international climatic conventions (Gibbs *et al.*, 2007).

The structural attributes of forest ecosystems like diameter of tree at breast height (DBH) has been widely used in estimation of above ground biomass (AGB) (Keller *et al.*, 2001). Estimation of AGB from DBH in field is quite inefficient in terms of time and money (Houghton, 2005). On the other hand remote sensing with spatial and temporal coverage capability proves to be a promising technique in AGB estimation. Several studies have attempted to utilize passive optical and active microwave remote sensing tool to retrieve forest biophysical parameters for estimation of AGB and carbon stock (Dobson *et al.*, 1995; Viergever *et al.*, 2008; Wijaya & Gloaguen, 2009).

Recent advances in remote sensing techniques like LASER altimetry, PolInSAR and very high resolution (VHR) imagery have been used for estimation of vegetation structural attributes with reasonable accuracy (Wang and Ouchi, 2010; Benson *et al.*, 2011; Neumann *et al.*, 2011). Moreover, direct retrieval of forest attributes like canopy height provides new avenues for enhanced forest degradation monitoring and management strategies.

1.2. Research context

1.2.1. A review of global forest biomass assessment methods

The forest ecosystem acts as global carbon sink (Anaya *et al.*, 2009). International Union of Forest Research Organizations (IUFRO) recognized forest biomass as one of the most significant areas of priority (Zhao & Zhou, 2005) which makes AGB as an important forest structural attribute to explore. AGB represents the major portion of total forest biomass. Hence, forest biomass change is treated as an important factor of climate change by Kyoto Protocol.

Traditional AGB inventory links ground measurements and tree parameters to their biomass using species-specific allometric relationship (Jenkins *et al.*, 2003) and further extrapolates these measurements to entire forest stand (Hall *et al.*, 2006). The field based method is useful for small areas but inefficient for regional or national scale biomass estimate because of high cost involved and uncertainty associated with geographically distributed biomass data and equations.

Remote sensing technology has been extensively used for vegetation and ecosystem studies. Typical remote sensing systems such as passive optical systems and active systems like radar provided a battery of ecological applications like estimation of leaf area index (LAI) and AGB (Lefsky *et al.*, 2002). Passive optical remote sensing utilizes spectral reflectance properties of forested area to relate with structural attributes of forest ecosystem. The capability of optical sensors for estimation of AGB is limited by their inability to retrieve vertical structural information of forest. Synthetic Aperture Radar (SAR) and LiDAR have extended a vertical dimension to forest modeling and AGB estimation. Radar sensor operates in the microwave portion of the electromagnetic spectrum (~ 1 cm to 100 cm). Longer wavelength and weather-independent operability of radar sensors have been utilized for biomass estimation in various forest types like coniferous forests (Dobson *et al.*, 1992) and mixed deciduous forests (Ranson & Sun, 1994). However, the problem of signal saturation limits the applicability of radar sensors in areas with high biomass content (Waring *et al.*, 1995) and heterogeneous forest conditions (Imhoff, 1995). Moreover, the ability of SAR to estimate AGB is further limited by high canopy density as dense canopy doesn't allow enough SAR radiations to penetrate through them (Dobson *et al.*, 1992).

1.2.2. Advent of LiDAR remote sensing for forest biomass and carbon stock studies

Use of Light Detection and Ranging (LiDAR) remote sensing or laser altimetry has facilitated incredible advancement in understanding of vegetation structural parameters. LiDAR has unique capability to provide information on vertical structure of forest ecosystem (Drake *et al.*, 2002). LiDAR uses light pulses generated in visible or near infrared range to measure the travel time from laser sensor to target and back (Patenaude *et al.*, 2005). The analysis of echo of the scattered light pulse by the target is used to derive various forest vertical structural parameters (Wu & Xing, 2010). LiDAR is independent of the signal saturation and biomass density limitations (Lefsky *et al.*, 1999a) as it measures the physical attributes of the forest rather than the spectral properties.

There are basically two categories of LiDAR systems: Discrete Return Devices (DRD) which measures time elapsed between emission and return of laser pulse (Hudak *et al.*, 2002; Evans *et al.*, 2006) resulting in 3D point cloud and Waveform Recording Devices (WRD) which captures continuous energy return from every emitted laser pulse (Patenaude *et al.*, 2005) as shown in Figure 1-1. LiDAR systems are also classified on the basis of the width of the laser pulse: large-footprint system having diameter of the laser beam greater than 5 m on the ground and small-footprint system having diameter less than 50 cm (Bortolot & Wynne, 2005). LiDAR systems can also be classified based on the height of the platform as airborne having sensor height less than 2 km and spaceborne LiDAR with sensor height of 600 km (Figure 1-1).

In context with forest AGB estimation studies, large footprint, waveform LiDAR is more efficient than small footprint, discrete return LiDAR (Drake *et al.*, 2002) as they fully digitize the return laser pulse, thus providing enhanced vertical structural information of forests. However, these systems have high associated cost and relatively low spatial coverage, and consequently are inappropriate for AGB estimation studies at regional or global scale. Hence attention has been shifted to the satellite based platforms. National Aeronautics and Space Administration (NASA) 's ambitious Vegetation Canopy LiDAR (VCL) Mission which was scheduled for launch in January 2000 aimed at three-dimensional mapping of the land surface structure of the entire Earth including land cover canopy. VCL promised to provide unprecedented global data for modeling, monitoring, and predicting the state of Earth's ecosystem and provide key inputs for climate modeling and predictions (Dubayah *et al.*, 1997). The once proposed Earth Observation (EO) mission "Carbon 3D" for global biomass mapping using VCL with a Bidirectional Reflectance Distribution Function (BRDF) imager (Hese *et al.*, 2005) which was scheduled for launch in

2009 also failed to materialize. The ICESat Geospatial Laser Altimeter System (GLAS), launched in January 2003, is the only operational spaceborne LiDAR available till date. Hence, for the success of future spaceborne LiDAR missions, a thorough analysis of GLAS waveform is required.

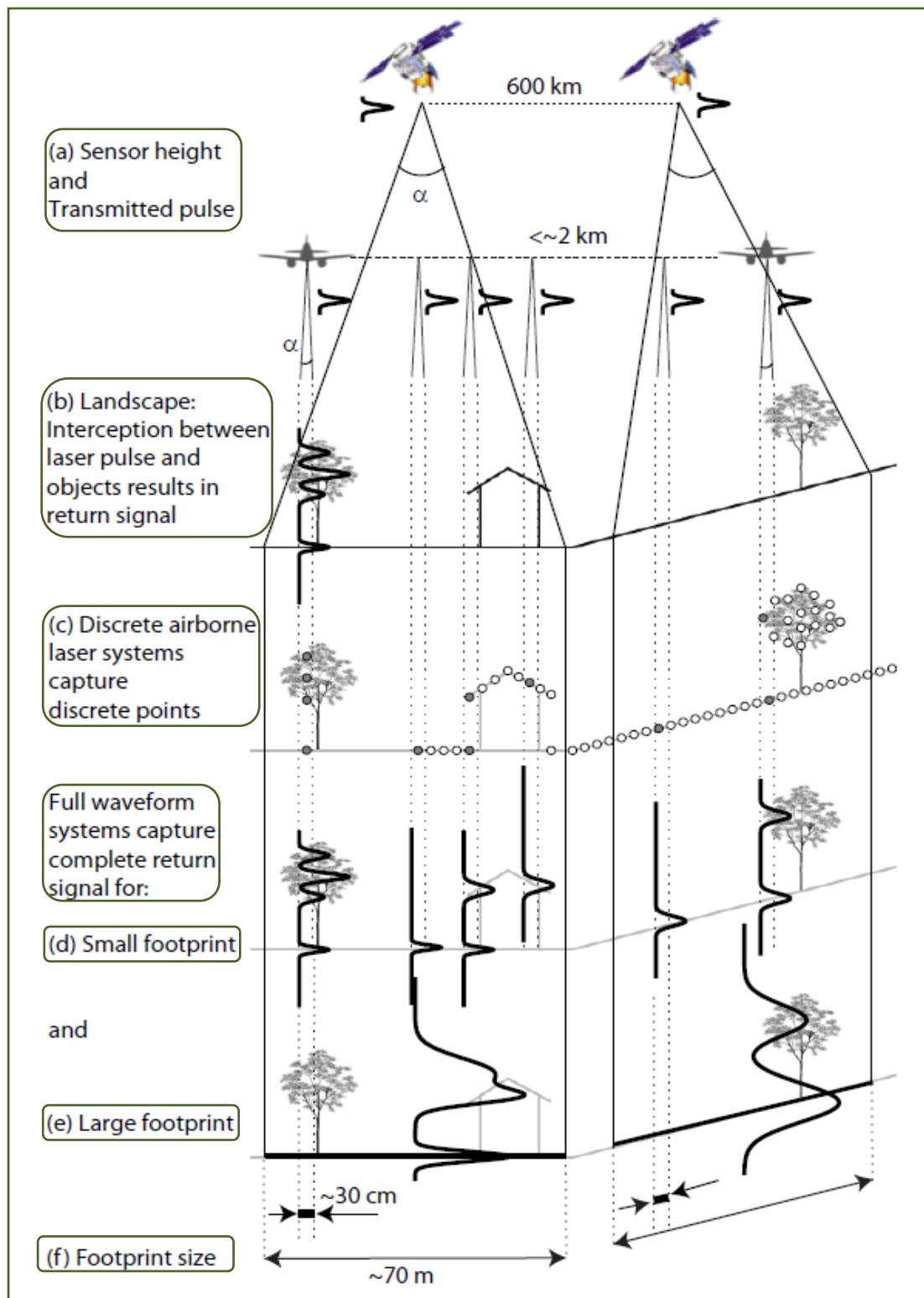


Figure 1-1: Overview of Discrete Return LiDAR and Waveform LiDAR Systems
[Source: Figure adapted from Duong (2010)]

1.2.3. ICESat GLAS full waveform large footprint LiDAR in forestry

NASA's GLAS on board Ice, Cloud and Land Elevation Satellite (ICESat) is designed to measure ice-sheet elevation changes, land surface topography profile and canopy heights (Zwally *et al.*, 2002). GLAS consists of three Nd:YAG lasers each of which has a 1064 nm laser channel for surface altimetry (NSIDC, 2012). The first laser failed shortly after the start of its operational period in 2003 and the ICESat's science mission ended in February 2010 with the failure of its third laser instrument thereby completing 18 laser-operation campaigns successfully during the whole lifespan.

GLAS transmits laser pulses of 4 ns width which is equivalent to 60 cm (1 ns = 15 cm) in surface elevation at a rate of 40 shots per second (40 Hz). All the 40 shots are telemetered 544 bins over the land surface corresponding to a height of 81.6 m (544 bins x 15 cm/bin) over ground (Brenner *et al.*, 2003) for laser L1 and 1000 bins corresponding to 150 m (1000 bins x 15 cm/bin) for laser L3 (Harding and Carabajal, 2005). Laser L2 operates in both the height ranges (Duong *et al.*, 2009). The approximate diameter of GLAS footprint is 70 m spaced at 175 m along track on the ground surface. The footprints are normally elliptical in shape. However, the ellipticity of the footprints have varied throughout the laser operation campaigns (Pang *et al.*, 2008) starting from moderately elliptical, very elliptical and very elliptical with side lobe for laser 3, laser 2 and laser 1, respectively (NSIDC, 2011a). The return waveform represents the vertical distribution of surface features within the footprint area.

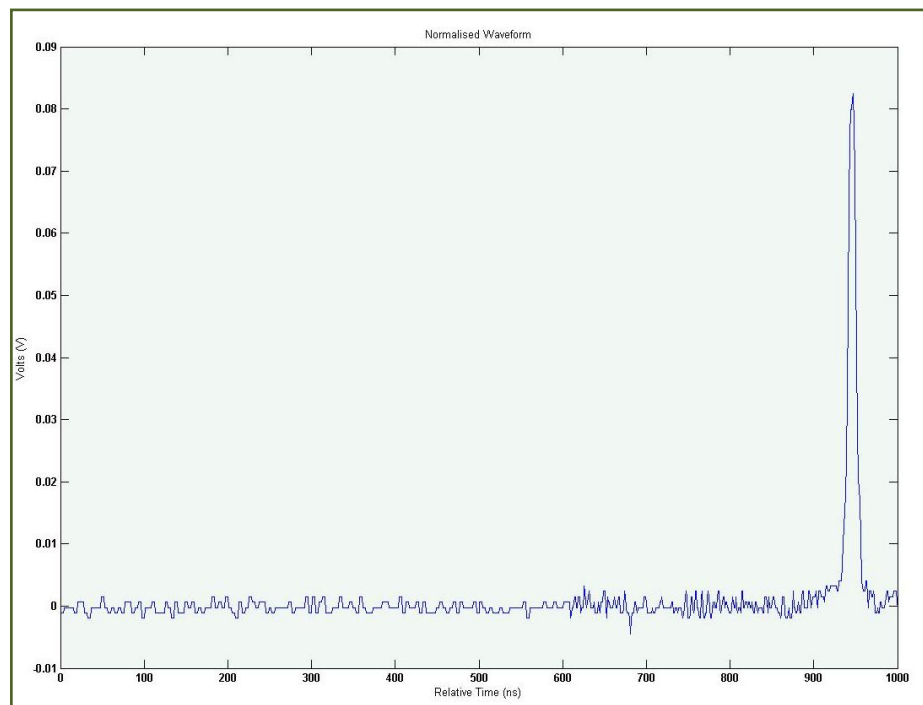


Figure 1-2: ICESat GLAS Waveform

Many studies have demonstrated that the breadth of the GLAS waveform can be used for AGB estimation in relatively flat and homogeneous forest area (Harding and Carabajal, 2005; Lefsky *et al.*, 2005; Rosette *et al.*, 2008). However, in areas with moderate to high slope, the vertical extent of waveform (distance from the beginning to end of the signal) increases as the function of the slope and footprint size which results in pulse broadening (Lefsky *et al.*, 2005). As a consequence, the ground return and canopy return can occur at the same elevation making interpretation of the waveform complex.

Several studies have attempted to model the forest vertical structural attributes from GLAS waveform metrics (Harding and Carabajal, 2005; Lefsky *et al.*, 2007; Boudreau *et al.*, 2008; Rosette *et al.*, 2008; Sun *et al.*, 2008). Lefsky *et al.* (2007), used the waveform extent, leading edge extent (distance from signal start to half of the waveform maximum power) and trailing edge extent (distance from signal end to half of the waveform maximum power) to retrieve canopy height with R^2 value of (83 %). Sun *et al.* (2008), demonstrated that energy quartile of GLAS waveform can be used to estimate canopy height. The position of 25 %, 50 %, 75 % and 100 % energy quartiles are calculated on the basis of the percentage of total reflected energy starting from the signal end. These energy quartiles represent the ground surface, not the signal end, hence the ground peak needs to be located efficiently. Boudreau *et al.* (2008), used terrain index from SRTM data and slope between signal start and first Gaussian peak to predict AGB.

Gaussian decomposition is widely used to retrieve information from GLAS waveform metrics. GLAS waveform is thought to be composed of sum of individual Gaussian return from each surface feature within the footprint (Blair and Hofton, 1999). ICESat science processing team have developed waveform fitting algorithm which fits up to six Gaussian curves to each waveform for land regions. The location of the last Gaussian peak is likely to represent the ground return. However, it has been found that the last Gaussian peak may not always represent the ground return, and in some cases the second lowest peak is a better representation of the ground elevation (Rosette *et al.*, 2008). The number of the Gaussian curves in a waveform is a function of canopy and surface feature properties within the footprint. Hence waveforms having higher noise level could not be decomposed effectively using limited number of Gaussian modes and may result into missing the last mode (Duong *et al.*, 2009). This limitation shadows the utility of Gaussian decomposition method for canopy and surface feature characterization. This problem calls for a detailed investigation of curve fitting methods which can account for such deficiencies.

1.2.4. Neuro – Genetic Approach for extracting vegetation variables from LiDAR measurements

Rapid advances in earth observation and remote sensing technologies have enabled data capture with enhanced spatial, temporal and spectral coverage (Asrar and Dozier, 1994). These advancements have to be supplemented with development of efficient and robust data analysis techniques and algorithms which can account for a wide variety of data in order to retrieve vegetation structural parameters. The vegetation structural attributes such as AGB, leaf area index, canopy height, forest age, etc. are continuous in nature which implies that a functional relationship can be developed between the remotely sensed data and the ancillary ground data (Kimes *et al.*, 1998). Hence, a significant amount of research is being conducted to develop efficient method to extract continuous vegetation attributes accurately.

Several approaches have been investigated for retrieving the continuous vegetation parameters which mainly falls under three categories viz. linear models, nonlinear models and physically based models (Kimes *et al.*, 1998). Linear models represent simple linear relationship between independent variables (remotely sensed data) and dependent variables (AGB, canopy height, etc.) shown in equation (1.1):

$$Y = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \dots + \beta_i Z_i + \varepsilon \quad (1.1)$$

where Y is the predicted variable, Z_i is a function of basic independent variable X_i and ε is error factor (Kimes *et al.*, 1998). The performance of the aforementioned model for predicting vegetation parameters is limited as the relationship between the remotely sensed signal and vegetation attributes are nonlinear (Jakubauskas, 1996). To overcome this limitation, transformation techniques such as vegetation indices are being used to correlate remote sensing data with vegetation variables. NDVI, ratio of simple combination

of NIR and red bands (Rouse *et al.*, 1974), is widely used for vegetation monitoring studies (Goward *et al.*, 1991; Huete *et al.*, 1994). Nonlinear and physically based modeling approaches have also been used to extract vegetation variables. Physically based models attempt to extract desired vegetation variables from measured vegetation parameters (Kimes *et al.*, 1998). These models use inversion and numerical optimization technique for extraction of vegetation variables making these models numerically and computationally intensive (Pinty *et al.*, 1990). For the effective implementation of all these models and to achieve accurate results, the researcher must have a clear understanding of the functional relationship between the input variables and output parameters. Hence there is a growing need for an innovative tool which is able to understand and learn the complex functional relationships underlying in the collected data, develop the functional basis of input and desired output, and should be able to adapt and evolve itself with variation occurring in the input data.

1.2.4.1. Artificial neural networks

Artificial Neural Network (ANN), a system inspired by functioning of biological neuron seems to be the one-stop solution having ability to learn relationships using the given data and to predict the desired output from the given input (Wasserman, 1989). This remarkable capability of ANN is attributed to its ability to mimic the unique information processing characteristics of biological neuron such as parallelism, fault tolerance and nonlinearity. Even if the ANNs are biologically inspired, they are essentially nonlinear statistical models (Tettamanzi and Tomassini, 2001).

ANNs consist of neurons or nodes grouped into layers which receive some input, add them up and produce output according to some simple threshold function (Grönroos, 1998). Layers are interconnected through connection channels or links. Links have weights whose values are determined during the training phase. Weights are multiplied to the node output to produce desired results (Figure 1-3). The weights associated with connection channels are either inhibitory or excitatory in nature. The way how a network is joined using connection channels or links is referred to as network topology or network architecture. There are various kinds of neural network architecture such as Multi Layer Perceptron (MLP), Radial Basis Function (RBF), Hopfield network, Self-organizing map and recurrent network.

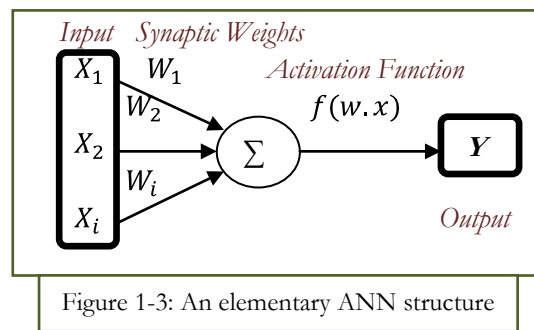


Figure 1-3: An elementary ANN structure

The feedforward MLP network topology (Figure 1-4) is used in this study because of its desirable computational and approximation capabilities (Cybenko, 1989; Hecht-Nielsen, 1989). The MLP architecture maps input to output through processing performed by series of interconnected nodes. The layers between input and output are termed as hidden layers and neurons contained within them as hidden nodes because of their indirect connection with external input and output layers (Azzini, 2005). Each node implements a transfer function expressed in equation (1.2):

$$O = f_i(\sum_{i=1}^n w_i x_i) \quad (1.2)$$

where f is the activation function. In many studies sigmoid function is used as activation function represented by equation (1.3):

$$f(x) = \frac{1}{1 + e^{-(bx-c)}} \quad (1.3)$$

The applicability of the sigmoid activation function is limited due to its differentiability (Azzini, 2005). Moreover, it compresses all the outputs between the range (0 and 1) which further reduces its efficiency (Kimes *et al.*, 1998).

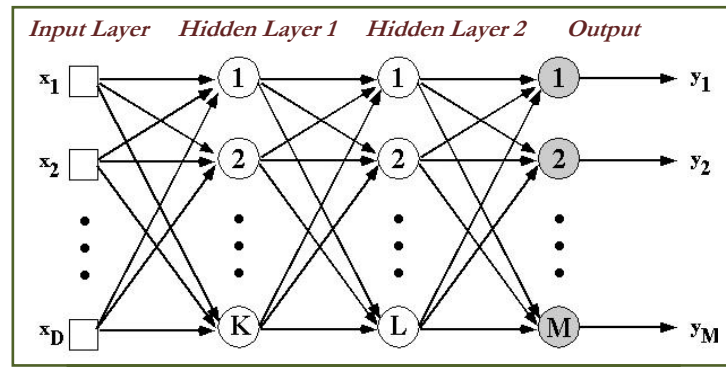


Figure 1-4: Feedforward MLP Network Topology

[Source: Figure adapted from (Roebber *et al.*, 2003)]

Once an appropriate network topology has been identified, the ANN has to be trained in order to produce desired output. Essentially, training is an iterative process which modifies the weights of ANN thus producing best possible nonlinear approximation relationship based on network architecture. This modification minimizes the error between predicted and the actual output. Hence a thorough investigation is required to design robust training rule for finding optimum weight. There are mainly three categories of learning rules that have been adopted to train the neural network viz. supervised learning, unsupervised learning and reinforcement learning. In supervised learning method, the training data consist of a number of input/output (I/O) training patterns. This rule aims to determine the set of weights which minimizes the error by modifying and updating the weights at each training step. The first and most widely used supervised learning rule is backpropagation (Werbos, 1974, 1994; Rumelhart *et al.*, 1986) which evolved as a standardized technique for training MLP network architecture. Backpropagation uses gradient descent method to extract network weights (Moller, 1993; Kimes *et al.*, 1998) which is essentially a least square optimization technique. Unsupervised learning rule self-organizes the input data presented and extract their emergent relationships. No external feedback is provided pertaining to the accuracy of I/O mapping.

The quality and integrity of data are some of the critical aspects of neural network research. The data should be a good representation of all the variation of the problem under investigation. The data should have enough diversity so as to produce training and validation data set. If a network has too many degrees of freedom, it can easily adapt itself to the training data and the process is termed as *overfitting*. This means that even if a network is able to predict the output with a greater accuracy, it fails to perform efficiently with previously unseen inputs. In other words, the generalization capability of the ANN model is reduced.

In order to overcome the problem of overfitting, a method called *early stopping* (Moody, 1994; Bishop, 1995; Sarle, 1995) is adopted. In this approach, the training pattern data is divided into three categories viz. training, test and validation data. The training data set is used for the learning process and then at certain intervals, the ANN model so developed is evaluated with the validation data and a model producing the best result is selected. Test data is used to verify the generalization capability of the ANN (Gupta *et al.*, 2003).

Therefore, a representative data population, robust training method and optimized network topology are essential for the efficient performance of ANN model. This research endeavours to address these issues by adopting a Neuro – Genetic approach or Evolutionary ANN for input data selection and optimization of network topology for extracting forest structural variables.

1.2.4.2. Evolutionary algorithm and Evolutionary artificial neural network

Evolutionary algorithms (EA) or genetic algorithms (GA) are system of evolutionary techniques inspired by the idea of the natural evolution and genetics (Tettamanzi and Tomassini, 2001; Eiben and Smith, 2003). This field of investigation is quite nascent and attempts to bring together concepts that form the basis of Darwinian evolution. The idea of GA is first proposed by John Holland in 1960s. He envisioned genetic code as sequence of binary values or *chromosomes*. EA utilizes chromosomes for problem solving. First, desired outputs are encoded into the chromosomes. Then new populations are evolved using genetic operators. The common genetic operators used are reproduction, crossover and mutation (Afandizadeh and Kianfar, 2009). Reproduction involves direct transfer of chromosomes to next generation. Crossover selects two parent chromosomes and produces offspring chromosomes using combination of parents. Mutation is random operator which acts by introducing new genetic material in the chromosomes thereby maintain the variability of population. The so evolved chromosome population after transferring the positive characteristics of each generation is evaluated on the basis of their fitness which is computed using a fitness function. The whole process is summarized as shown in Figure 1-5.

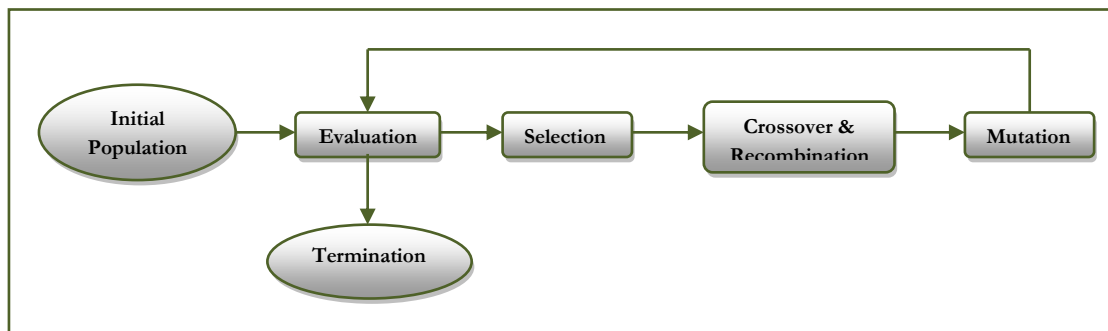


Figure 1-5: Process flow of EA

EA approach has been used to solve complex optimization problem and a number of applications have been found in enhancing the efficiency of ANN. EA can be used to find the relevant set of input data from a large number of input variables which significantly increases the performance of ANN. This combined framework of ANN and EA has been referred to as Evolutionary ANN (EANN). EANN is characterized by natural evolution as another fundamental form of adaptation in addition to learning (Yao, 1991, 1999). EANN is distinguished by their ability to adapt itself with changing environment that is they can modify their network topology and learning rule without human intervention (Azzini, 2005). EANN can automatically evolve itself to an optimized ANN with appropriate topology and weights without

manual trial and error process. A large number of input data increases the complexity and computation time of ANN. The dimensionality of the input data can be reduced using EANN without a loss in the performance (Brill *et al.*, 1992). EA searches for a subset of input data that performs synergistically to produce the best result (Kimes *et al.*, 1998). Network architecture optimization is yet another important issue where EANN have been applied with significant accuracy. There is a lack of systematic method to design a network with optimal architecture for performing a given task automatically. EANN uses constructive and destructive approach for automation of network topology optimization (Yao, 1999). Constructive approach starts with a small network and subsequently builds layers, links and nodes until a desired topology is achieved (Yao and Liu, 1997). Destructive method starts with a large network and subsequently discards layers, connections and nodes until the best possible architecture is achieved (Mozer and Smolensky, 1989).

EANNs have several unique capabilities which facilitate the extraction of forest structural variables from remotely sensed signals. Some studies have utilized ANN for predicting vegetation variables. Pierce *et al.*, (1994) used MLP network to predict trunk density, average trunk diameter, and average trunk height of Loblolly pine using AirSAR data. Kimes *et al.*, (1996) developed a MLP network to extract forest age using TM and topographic data.

The merits of ANN and EA are now discussed in details and advancements made in this area are briefly reviewed. However, there have not been any studies attempting to use EANN approach for AGB estimation using full waveform large footprint LIDAR. The only research that attempted to extract AGB and tree height from full waveform LiDAR using ANN was by Sun *et al.*, (2007). The present research aimed to assess the capability of EANN for predicting tree height and AGB using large footprint full waveform spaceborne LiDAR. This study also endeavours to serve as benchmark for current and future AGB modeling studies using Neuro – Genetic approach. Moreover, the rapid estimation of AGB with enhanced accuracy using a combination of EANN and LiDAR will help to monitor the forest degradation by comparing temporal change in AGB content.

1.3. Problem statement

Forest degradation has resulted in qualitative and quantitative loss of vegetation cover and reduced productivity (FAO, 2006). Continuous anthropogenic pressure for exploitation of forest products has resulted in degradation of forests. This problem is acute as it is critically reducing the ability of the forest land to sequester carbon due to reduced forest cover. Furthermore, significant degree of uncertainty exists in carbon estimates in areas with low biomass carbon (Figure 1-6) which called for a more robust and accurate method to estimate AGB for effective monitoring of carbon pool in such regions. Hence, reasonably accurate, cost effective and rapid AGB estimation tools are needed to effectively monitor the changes in carbon stock due to degradation of forests.

ICESat / GLAS data products with its unprecedented accuracy and global coverage have been used for land cover classification and biomass estimation (Lefsky *et al.*, 2005; Boudreau *et al.*, 2008). Most of the GLAS waveform derived AGB estimation studies were conducted in locations having flat terrain. In areas of greater slope or complex topography, deriving reasonable estimate of canopy height is affected by pulse broadening (Harding and Carabajal, 2005). In such conditions, there are robust possibilities to explore waveform processing methods in order to accurately retrieve canopy height information contained in the GLAS waveform which in turn will enhance the accuracy of AGB estimate.

Furthermore, as explained earlier, EANN have unique capabilities which can be used for extracting continuous forest variable such as tree height and AGB. Thus far, suitability of EANN for AGB estimate has not been documented, which emphasized their in-depth investigation.

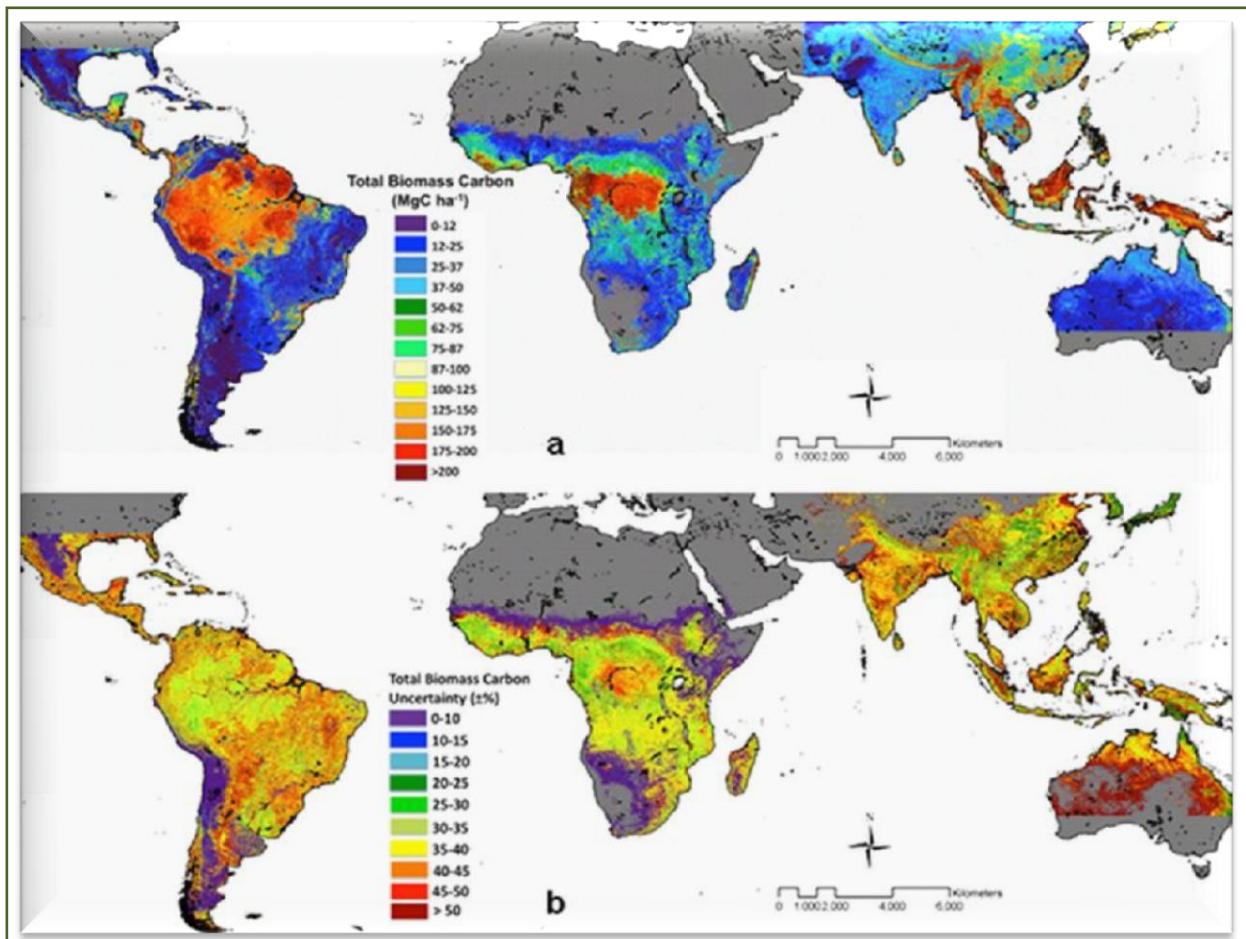


Figure 1-6: Total Biomass Carbon and Uncertainty in their Estimation

[Source: Figure adopted from (Saatchi et al., 2011)]

1.4. Research objectives

The main objective of the research is to investigate the possibility of using EANN in combination of large footprint full waveform LiDAR to extract tree height and AGB with enhanced precision, thus extending the applicability of full waveform large footprint LiDAR and to facilitate the effective monitoring of forest degradation through analysis of temporal changes in AGB content.

The specific objectives are as follows:

- To develop a methodology to consistently model tree height and AGB from ICESat GLAS waveform derived metrics.

Research questions:

- To what extent is the EANN approach is capable of predicting tree height and AGB using GLAS waveform derived metrics?
- To identify and establish a reliable combination of ICESat GLAS waveform derived parameters for reasonable estimate of tree height.

Research question:

- Which GLAS waveform derived parameters are relevant for extracting tree height?
- To analyse the change trend of AGB estimate for monitoring the impact of forest degradation.

Research question:

- What is the change in AGB as compared with past AGB estimate in the study area?

1.5. Outline of the thesis

The whole document has been organised into four chapters. *Chapter 1* presents background, research context, problem statements and objectives of the research. The description of the study area, data and method used are described in *chapter 2* and the results so obtained after the implementation of the method is discussed, analysed and presented in *chapter 3*. *Chapter 4* constitutes a thesis conclusion, remaining research issues and recommendations for future efforts to further investigate the utility of EANN in the field of LiDAR remote sensing for extraction of forest vegetation structural parameters.

2. MATERIALS AND METHODS

2.1. Study area

2.1.1. Selection of the study area

To fulfil the objectives of present study, a forest location is required that offers scope for sufficient biomass estimation and also with varying topography and forest composition. Another important criterion for the selection of the study site is ICESat GLAS waveform data, whose footprints can be accessed on the ground for the purpose of field data collection. The research is therefore planned to be conducted in forest regions around Shivpuri and Sheopur districts of Madhya Pradesh province, Central India (Figure 2-1). Also, some studies on biomass estimation (Roy and Ravan, 1996; Kale *et al.*, 2004) was previously conducted by Indian Institute of Remote Sensing (IIRS), Dehradun, India, hence some necessary ancillary data is already available.

2.1.2. Description of the study area

The study area is located in Madhya Pradesh state of India between (24°50` to 25°55`) N and (77°15` to 78°30`) E at an altitude of 462 m MSL. The total geographic area of Sheopur and Shivpuri is 16,883 km² with 5,971 km² of forest cover. The study area has semiarid to arid climate with annual mean daily temperature of 23 °C and average annual rainfall of 895 mm (Kale *et al.*, 2004). The general physiography of the terrain in the study area is typical of Central Indian highlands, interspersed with woodlands and scrubby vegetation.

The dominant forest type in the region is tropical dry deciduous with some patches of moist deciduous type. The forest composition of the study area is mostly of mixed type, including tropical dry deciduous tree species like *Anogeissus pendula*, *Boswellia serrata*, *Acacia catechu*, *Ziziphus xylopyra*, *Elaeodendron glaucum*, *Bauhinia racemosa*, *B. monosperma*, *Lannea coromandelica*, *Anogeissus latifolia*, *Diospyros melanoxylon*, *Buchanania lanzan*, *Aegle marmelos*, *Madhuca indica* and *Terminalia bellerica* (Kale *et al.*, 2004).

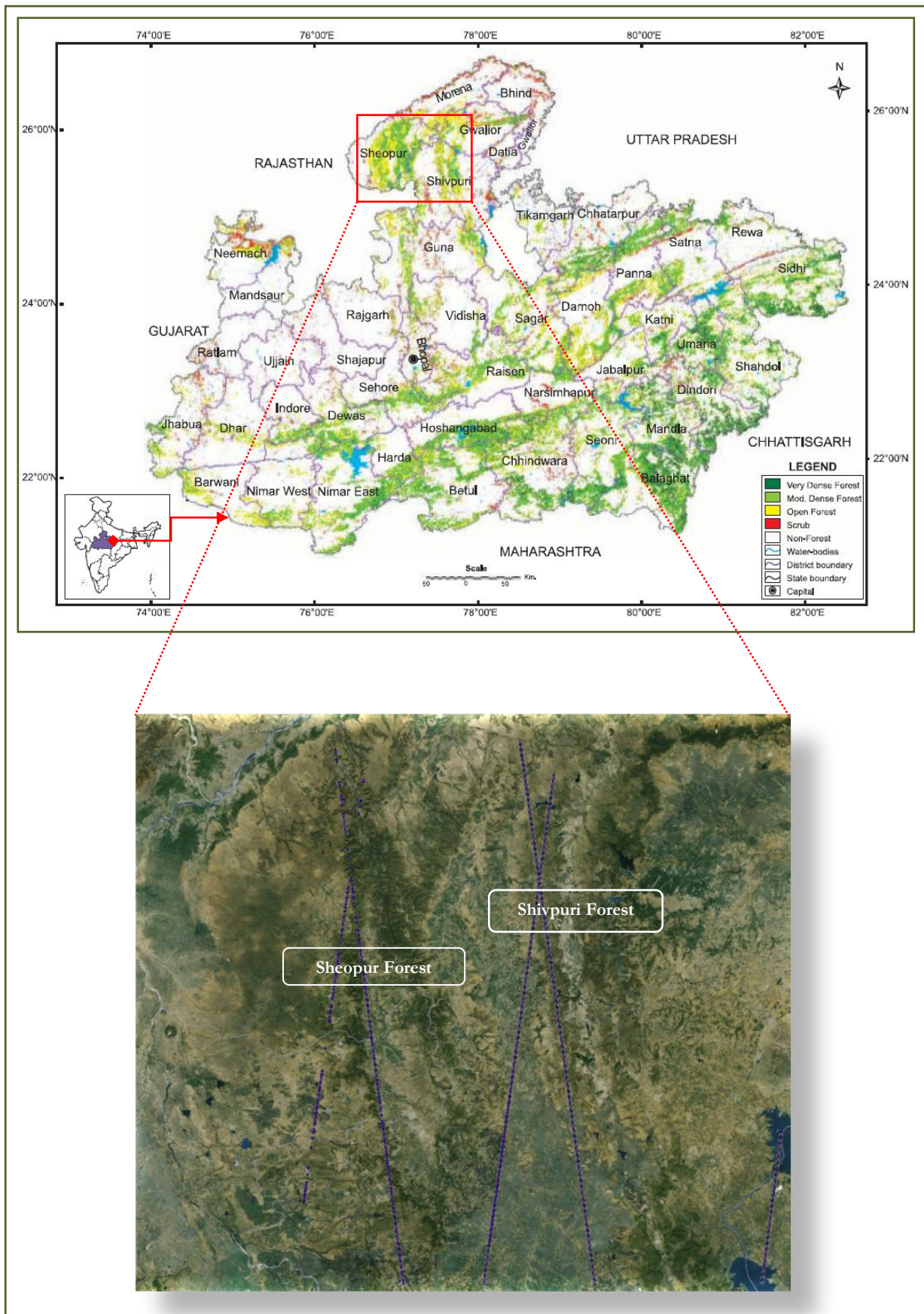


Figure 2-1: Location of the Study Area in India showing GLAS Footprints Track

2.2. Data available

Necessary description of the data used in this study is in the following sections.

2.2.1. ICESat GLAS data

GLAS data used in this study is acquired for the period (11/04/2009 – 09/03/2009) from L2E laser campaign. GLAS has fifteen data products (GLA01 to GLA15) out of which two data products (GLA01 and GLA14) of release 33 are used in this research. The data is procured from National Snow and Ice Data Center (NSIDC) through ICESat GLAS data subsetter (NSIDC, 2011b).

GLA01 is Level 1A Global Altimetry data product which contains raw waveform. It also contains information about index number, echo intensity and shot number. GLA14 (Level 2 Global Land Surface Altimetry data) product is derived from GLA05 and GLA06 which contains information regarding region specific surface altimetry. It also gives precise geolocation of the footprint center along with the height elevation and centroid of the waveform. The record number, shot time and shot number is a common field across all GLAS product which is used to relate the different GLAS products. The relevant specifications of ICESat GLAS are listed in Table 2-1.

Table 2-1: ICESat GLAS Specifications (Zwally *et al.*, 2002)

ICESat GLAS Parameter	Specification on Land Surface
Wavelength	1064 nm
Laser Pulse Rate	40 Hz
Average Footprint Diameter	~ 70 m
Laser Pulse Width	4 ns
Vertical Sampling Resolution	0.15 m
Surface Ranging Accuracy	5 cm
Footprint Geolocation Accuracy	6 m
Footprint Spacing Along Track	~ 170 m
Laser Beam Divergence	110 μ rad

2.2.2. Other reference dataset and materials

Other reference data, software and field instruments used in this research are:

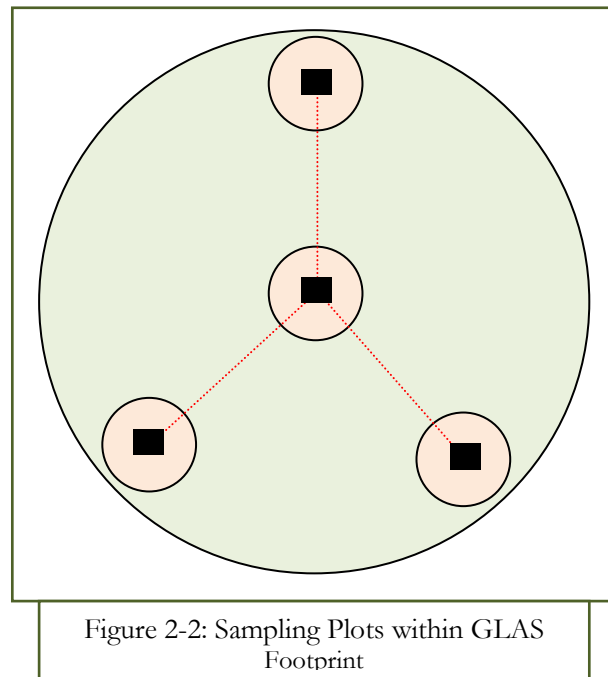
- **Survey of India Toposheet:** Survey of India topographic sheets [Toposheet Number: 54G(5,6,9,10,11,13,14,15,16)] scaled 1:50,000 are used for the purpose of field work.
- **Software required:** MATLAB, IDL, LabView, Python, Statistica, LabView SignalExpress, Emergent, JNEAT, DTREG
- **Field instruments:** GPS, Haga Altimeter, Laser Hypsometer, Silva Ranger Compass, measuring tape, field data collection form

2.3. Field data collection

Field work was carried out in Shivpuri and Sheopur forest region from 07/01/2012 to 19/01/2012. Before commencing with the field work, the GLAS footprints to be sampled for tree height and AGB estimation were identified. Out of 130 footprints, 40 were randomly selected to obtain a representative sample for field data collection.

Selected GLAS footprints were located on the ground using a handheld Garmin GPS with an accuracy level of ± 5 m. The footprints are elliptical in shape. For the convenience of laying sample plots in forest, the footprint shape was assumed to be circular with a diameter (73.5 m) equal to the average of semi-major (95 m) and semi-minor (52 m) axis of the elliptical footprint. After the center of the GLAS footprint was located, four sampling plots with a radius of 7.5 m, one in center and three on the footprint margin (22.5 m to north, south-east and south-west to the center) (Figure 2-2) ((Sun *et al.*, 2007) were laid.

The present study is focused on prediction of tree height and AGB from GLAS waveform, so diameter at breast height (DBH) and tree height of all trees with DBH > 10 cm within the sample plot were measured and recorded. The tree species information was also collected for estimating the field AGB using species-specific allometric equation developed by Singh and Misra, (1975) and Kale *et al.*, (2004) for the study area. The detailed field data collection form can be found in Appendix.



2.4. Research method

The overall research method followed is shown in Figure 2-3:

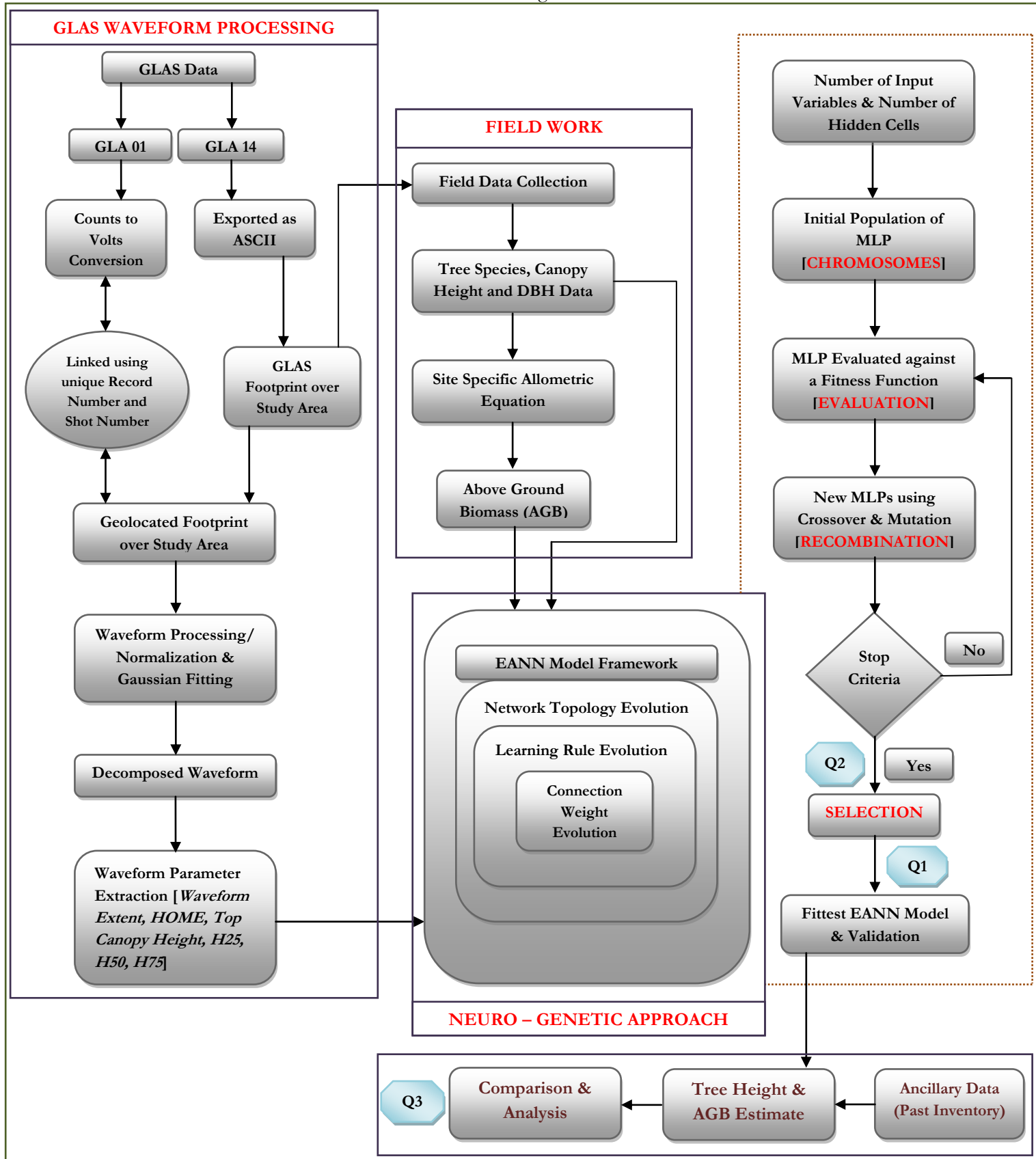


Figure 2-3: Research Method Flow Diagram

2.4.1. GLAS waveform processing

2.4.1.1. GLAS waveform pre-processing

Pre-processing of raw GLAS data obtained from NSIDC GLAS Subsetter was essential for carrying out further analysis and parameter extraction. The method implemented to process GLAS waveform data is adopted from Zhang *et al.*, (2009) and Duong, (2010).

2.4.1.2. Waveform conversion

ICESat GLAS data (GLA01 and GLA14) is distributed in binary format which has to be converted into ASCII format for further processing. The conversion was carried out using IDL routines *read_glas_file.pro* developed by National Snow and Ice Data Center (NSIDC, 2012). The elevation and geolocation information was extracted from GLA14 data using NSIDC GLAS Altimetry Elevation Extraction Tool (NGAT). The resultant waveform data which was originally in counts (from 1 - 256) was converted into voltage units for subsequent analysis. The count to volt conversion was implemented using IDL routine *read_gla01_wf.pro* developed by NSIDC.

2.4.1.3. Waveform normalization

The resultant voltage waveform (Figure 2-4) is normalized to facilitate the comparison of waveforms captured in different time periods. The normalization is done by dividing each energy value V_i by the total received energy V_T , at instant i represented as equation (2.1):

$$V_{Norm}(i) = \frac{V_i}{V_T} \quad \text{with} \quad V_T = \sum_{i=1}^N V_i \quad (2.1)$$

where N represents the number of waveform bins. In the present study $N = 1000$ bins. The area under any waveform becomes equal to 1 after normalization (Figure 2-5).

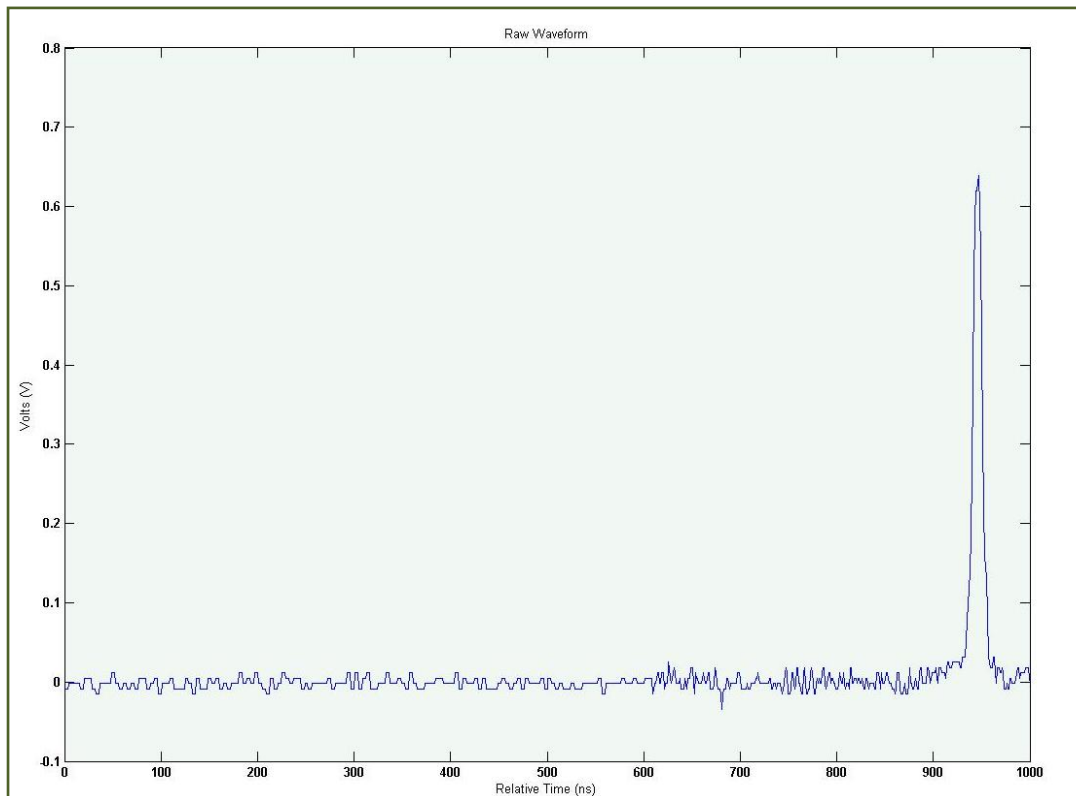


Figure 2-4: GLAS Waveform before Normalization

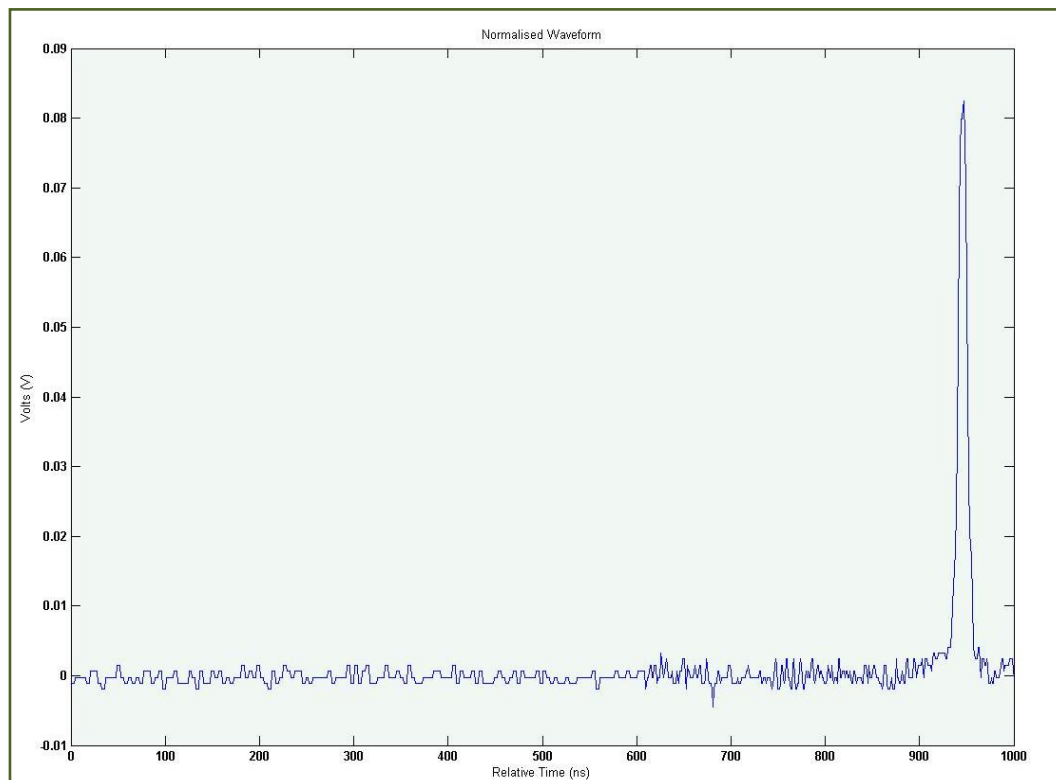


Figure 2-5: GLAS Waveform after Normalization

2.4.1.4. Detection of effective waveform signal

The waveform detection system continuously measures the return signal. Hence it is necessary to extract actual waveform signal from the continuous time series (Duong, 2010). This process is implemented by considering the position of the amplitude of the GLAS waveform signal firstly exceeds a certain noise threshold level.

In the present study, it has been observed during the visualization of the data that the actual waveform signal (Figure 2-6) often starts after 850th bin (850 ns) within 1000 bins. Hence the first 850 bins have been used for the calculation of mean (M_N) and standard deviation (δ_N) of the noise, as expressed in the equations (2.2) and (2.3):

$$M_N = \sum_{i=1}^{850} \frac{V_i}{850} \quad (2.2)$$

$$\delta_N = \sqrt{\sum_{i=1}^{850} \frac{(V_i - M_N)^2}{850 - 1}} \quad (2.3)$$

where (V_i) is the amplitude of the (i^{th}) bin in the waveform. The threshold value ($N_{Threshold}$) for differentiating between the noise and the actual waveform signal is determined as the mean (M_N) plus four times the standard deviation (δ_N) (Lefsky *et al.*, 2005) as represented by the equation (2.4):

$$N_{Threshold} = M_N + 4 * \delta_N \quad (2.4)$$

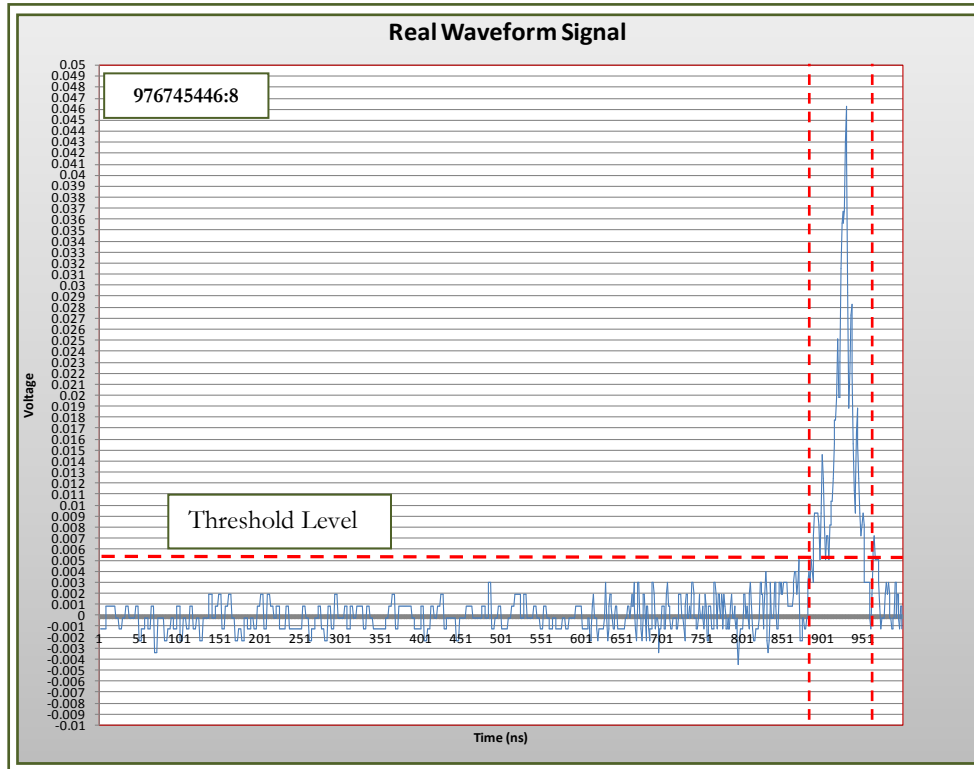


Figure 2-6: Actual GLAS Waveform signal defined based on Threshold

2.4.1.5. Waveform smoothing

The smoothing of the waveform is necessary for the elimination of noise (Figure 2-7) and estimation of initial parameters of the waveform such as location of waveform peaks. The smoothing is carried out using Gaussian filter (Duong, 2010) and implemented using LabView SignalExpress environment. The width of the Gaussian kernel is represented in terms of the standard deviation (δ) of the Gaussian function (Duong, 2010) and described as Full Width at Half Maximum (FWHM) which is linked to the standard deviation as given by equation (2.5):

$$FWHM = \delta * \sqrt{8 * \log 2} \quad (2.5)$$

which approximates to $[2.35 * \delta]$. Smoothing of waveform separates small amplitude signal from larger amplitude signal with similar frequencies (Figure 2-8).

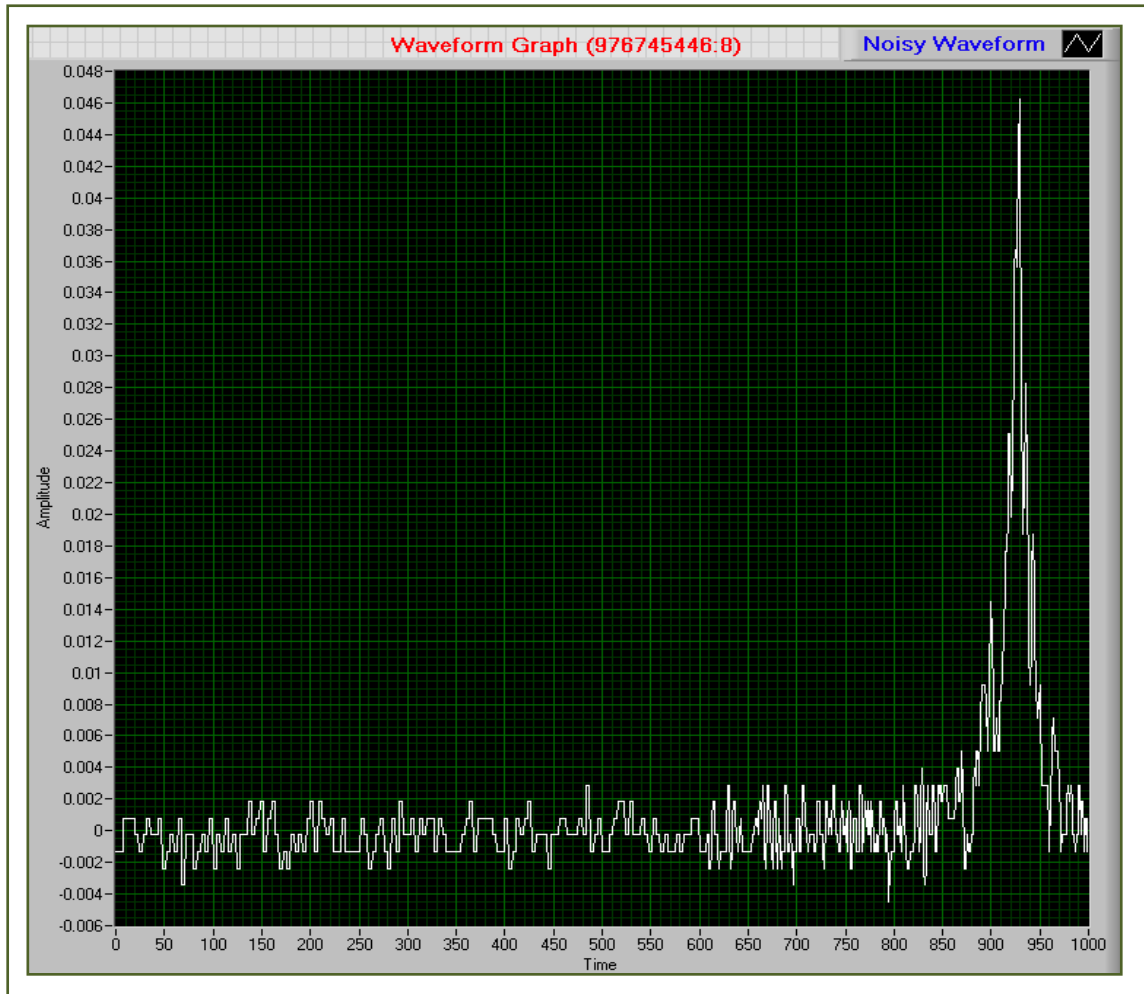


Figure 2-7: Noisy GLAS Waveform

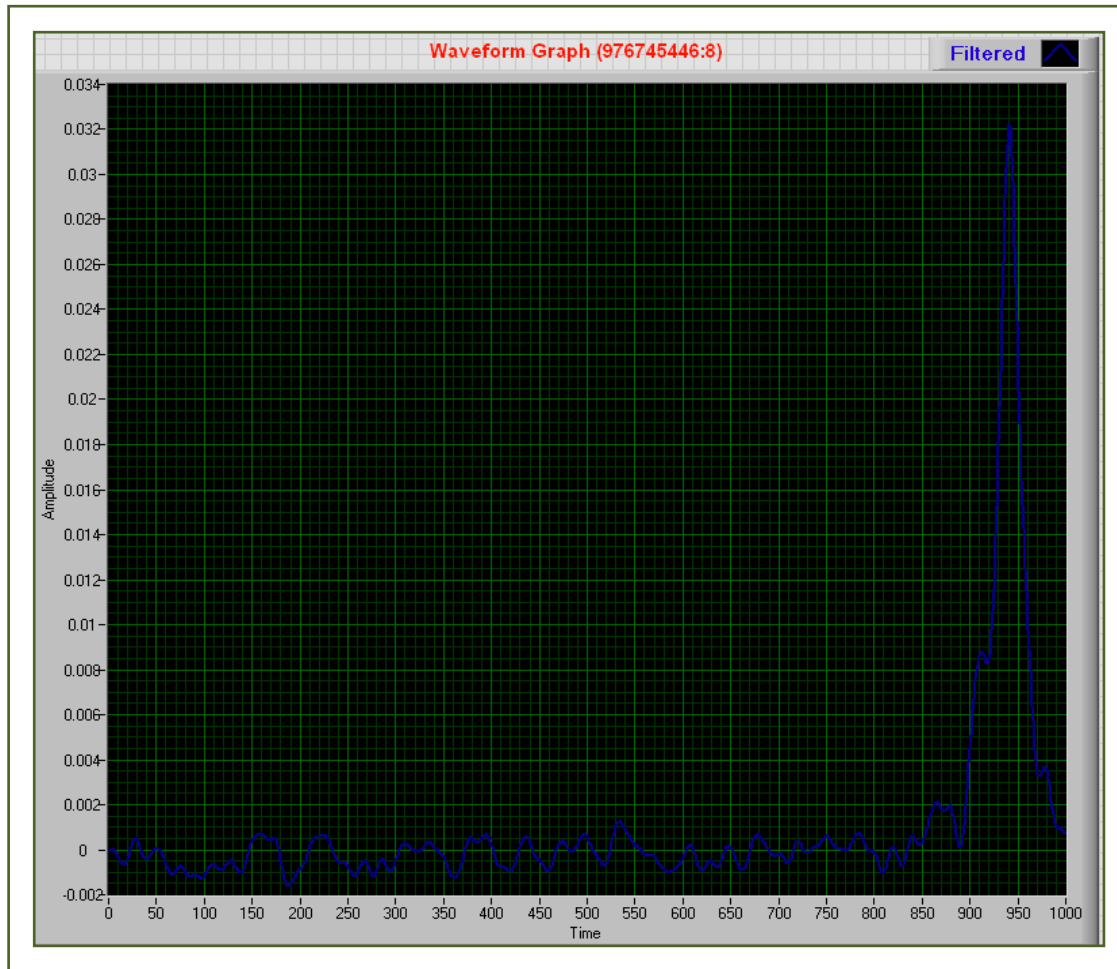


Figure 2-8: Smoothed GLAS Waveform

2.4.1.6. GLAS waveform initial parameter estimation

For Gaussian curve fitting to GLAS waveform, it necessary to extract some initial parameters of the waveform like peak locations and their corresponding amplitudes. The peak detection strategy adopted is based on the fact that the peaks are the highest locations between the valleys. Hence there are lower value points around the peaks. A search window of 5 ns (5 bins) is applied to detect the peaks. The window moves from the beginning to the end of the signal and detected the waveform peak by calculating the difference between two succeeding amplitude values and marked a peak when the resultant difference value changes from positive to negative. All the peaks were identified in this manner (Figure 2-9). The method was implemented in LabView. A generic Virtual Instrument (VI) developed for waveform peak detection will be provided in the appendix.

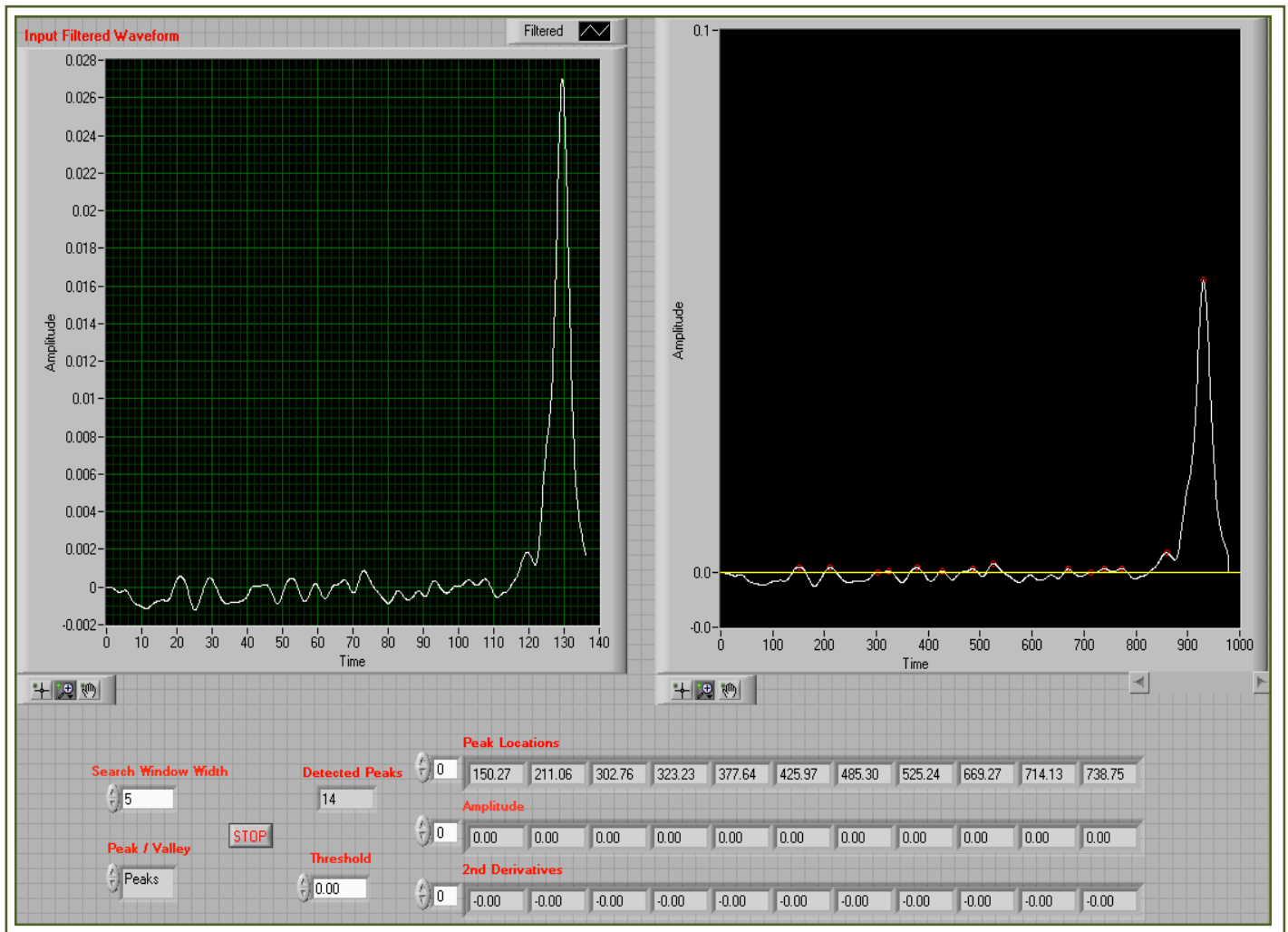


Figure 2-9: Identified Peaks (red circle) in the GLAS Waveform

2.4.1.7. GLAS waveform Gaussian fitting

The curve fitting step attempts to capture the trend of the signal by assigning a single homogeneous function across the entire range of the signal. The most general method for fitting any model to the signal is a kind of “trial and error” technique in which the parameters of the function are adjusted until it fits the signal with minimum possible error. Hence, the technique involved the selection of the best function model, initial guessing of function parameters and trial and error approach for fitting the model to the signal with desired accuracy.

Therefore, for selecting the best function to fit the waveform, following assumptions and deductions were made. For a reasonable approximation of the complex characteristic of the ground footprint, it was assumed that all the reflecting surface features behave as Lambertian scatters. Since the transmitted GLAS pulse is assumed to be Gaussian and the surface feature is also Gaussian, hence the telemetered waveform is expected to be Gaussian (Brenner *et al.*, 2003). Following this assumption, the normalized GLAS waveform was fitted with Gaussian curve. The fitting step involved modeling of smoothed waveform with Gaussian components using the algorithm developed by Duong, (2010).

The smoothed waveform is sum of series of Gaussian components expressed as equation (2.6):

$$W(t) = \varepsilon + \sum_{m=1}^N W_m(t), \text{ with } W_m(t) = A_m * e^{\frac{-(t-\mu_m)}{2\sigma_m^2}} \quad (2.6)$$

where $W(t)$ is the waveform at time t , ε is the noise level, $W_m(t)$ is the contribution of m^{th} Gaussian component, N is the total number of Gaussian components in the waveform, A_m is the amplitude of the m^{th} Gaussian component, μ_m represents the location (mean) and σ_m its width (standard deviation).

For Gaussian modeling of GLAS waveform, a nonlinear least square estimation technique is applied for computing the function parameters in equation (2.6). This is done by fitting the theoretical model to the observed waveform in such a way that the difference between the theoretical model and observation is minimised in the least square sense. Figure (2-10) shows the results of the Gaussian modeling of GLAS waveform.

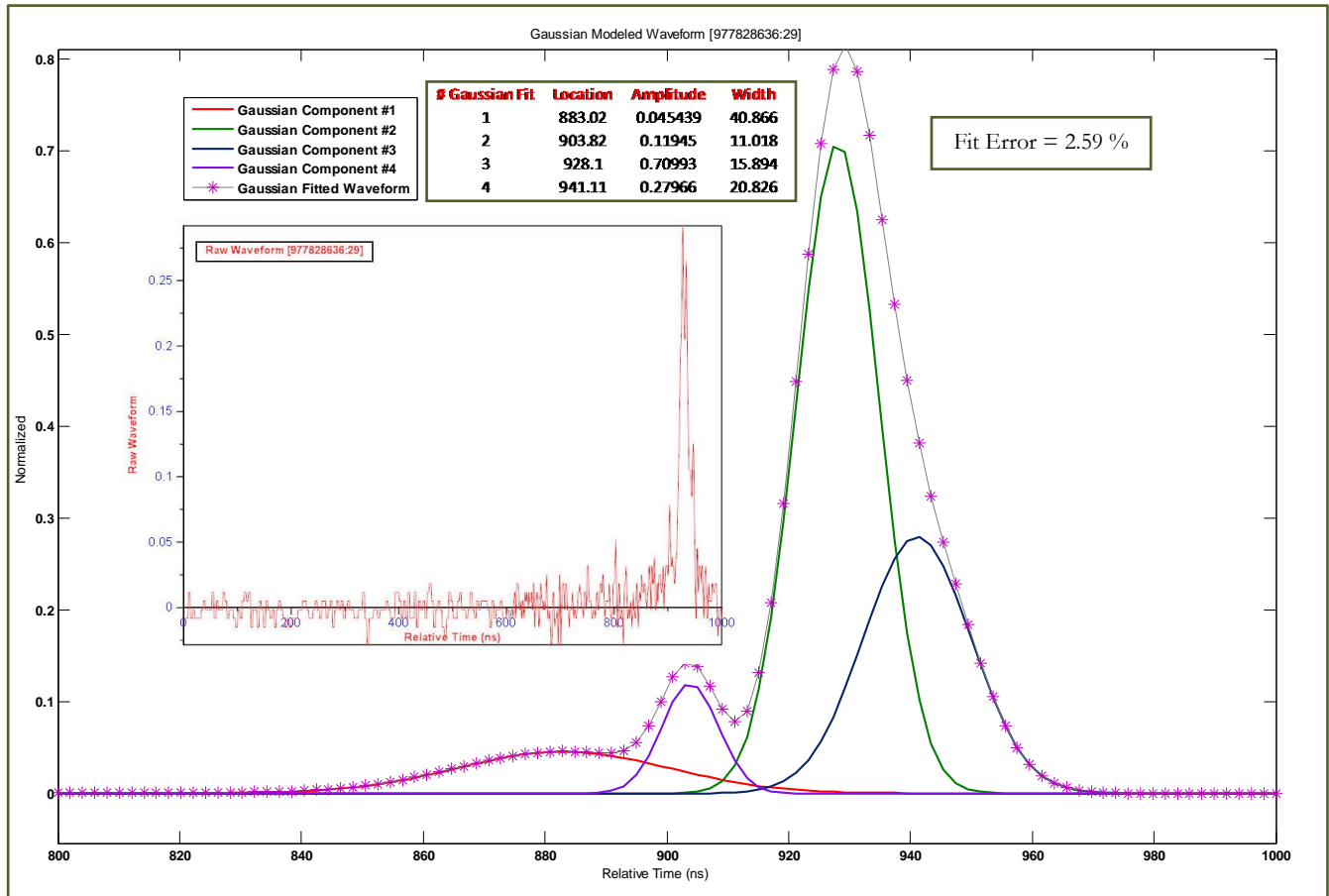


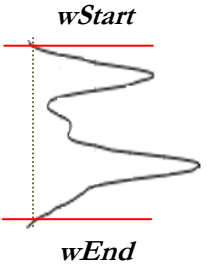
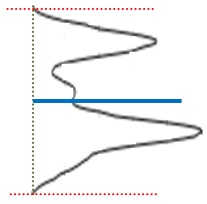
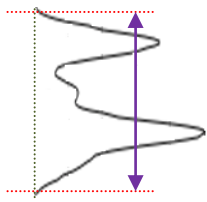
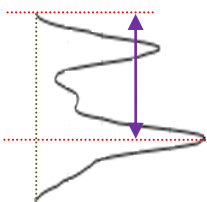
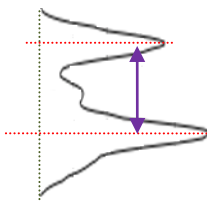
Figure 2-10: Results of the Gaussian Decomposition of the GLAS Waveform

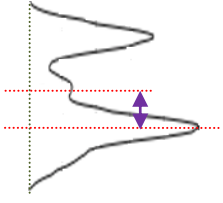
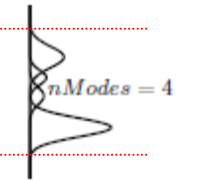
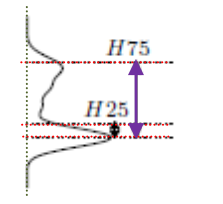
Four significant modes were detected in the waveform. The telemetered waveform may have a longer trailing edge caused due to atmosphere and cloud conditions referred to as ringing effect (Fricker *et al.*, 2005; Duong, 2010). This may result in detection of false modes that do not represent actual surface feature.

2.4.1.8. GLAS waveform parameters derivation

After the smoothing and Gaussian fitting of GLAS waveform, some relevant waveform derived parameters were calculated for the purpose of extracting tree height and AGB. The parameters were identified and selected based on studies carried out by Duong, (2010) and other related works. The waveform derived metrics are listed in Table 2-2.

Table 2-2: Description of GLAS waveform derived parameters [(Source: Duong, 2010)]

Waveform Parameters	Definition	Physical Significance	Visualization
Waveform signal start ($wStart$) and waveform signal end ($wEnd$)	Position where waveform first/last crosses above/below a certain threshold level	$wStart$: Highest interception point between surface and transmitted pulse. $wEnd$: Lowest elevation reflected from earth surface. Relevant for surface feature height extraction like tree height	
Waveform centroid ($wCentroid$)	Position where return pulse energy is divided in two equal parts.	Represents mean elevation within the footprint.	
Waveform extent ($wExtent$)	Distance between signal start and signal end.	Represents maximum tree height and maximum canopy height.	
Waveform distance ($wDistance$)	Distance from signal start to peak of the last Gaussian mode	Represents top tree height and top canopy height.	
Peak Distance ($wpDistance$)	Distance between first and last peak	Represents average tree height	

Waveform Parameters	Definition	Physical Significance	Visualization
Height of Median Energy (H_{OME})	Distance from peak of the ground return to the waveform centroid	Sensitive to change in vertical arrangements of tree canopy and degree of canopy openness	
Number of Gaussian Fits ($nModes$)	Number of Gaussian components derived from nonlinear least squares estimation	Signifies number of height levels corresponding to object and earth surface	
x % Quartile height (H_x) [H25, H50, H75, H100]	Height at which x % of the return energy occurs.	$H50$ equals the $wCentroid$.	
x % Quartile height to Waveform Distance ratio (R_x) [R25, R50, R75]	x % Quartile height divided by waveform distance	Normalizes the effect of different canopy heights	

The parameters listed in Table 2-2 were extracted from the modelled GLAS waveform. The location of waveform signal beginning and signal end (Figure 2-11) were determined by locating the leftmost and rightmost position where the amplitude of the waveform firstly exceed the noise threshold level as calculated using equation (2.4). After delineating the signal start and end boundary, the waveform metrics related to tree height such as waveform extent, waveform distance and peak distance (Figure 2-11) were determined for extracting ground return, tree top and tree height from GLAS waveform.

To detect the proposed forest structural attributes the location of the highest peak of the Gaussian component in the right half of the waveform was considered as a reference and identified as ground return. The canopy return is defined here as the first peak above the noise threshold level after the signal start. Further, the tree height was calculated as the distance between the canopy return and canopy return (Figure 2-11). Similarly, tree heights were extracted from all the modelled GLAS waveform. Quartile based GLAS waveform parameters, H25, H50 and H75, which represents the height of 25%, 50% and 75% return energy from canopy (Figure 2-12) were extracted from the waveform. The height quartiles were used to calculate R25, R50 and R75 which is given by dividing H25, H50 and H75 from tree height obtained from GLAS waveform respectively.

All the Gaussian modelled GLAS waveform derived parameters listed above are expected to be related to the forest structural attributes like tree height and AGB. As proposed earlier, one of the objectives of this study is to identify the best combination of waveform parameters for reasonable estimation of tree height and AGB. To achieve this aim, the powerful methodology of EANN (refer to section 1.2.4.2) is implemented to achieve the optimised combination of inputs for extracting reasonable outputs.

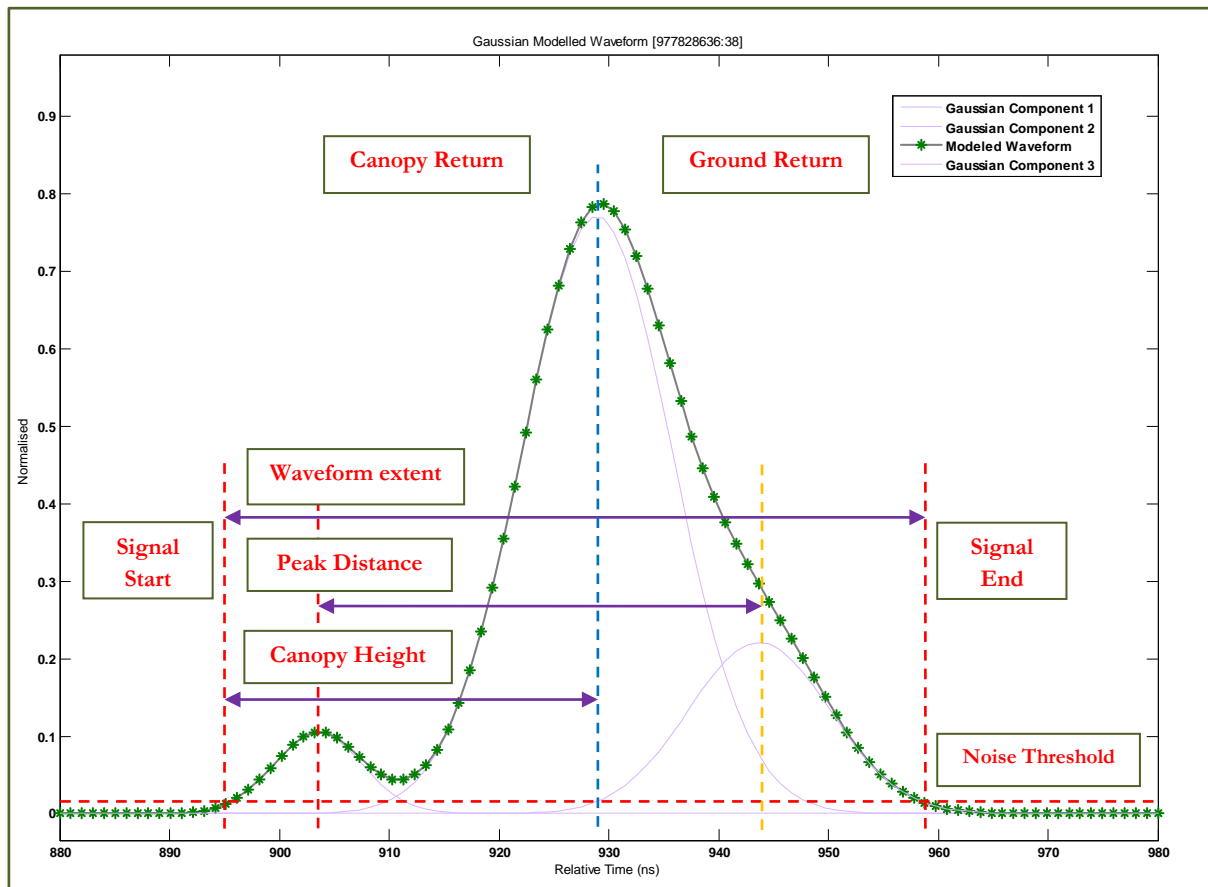


Figure 2-11: GLAS Waveform Parameters as Noise Threshold Level, Signal Start, Signal End, Ground Return and Canopy Height

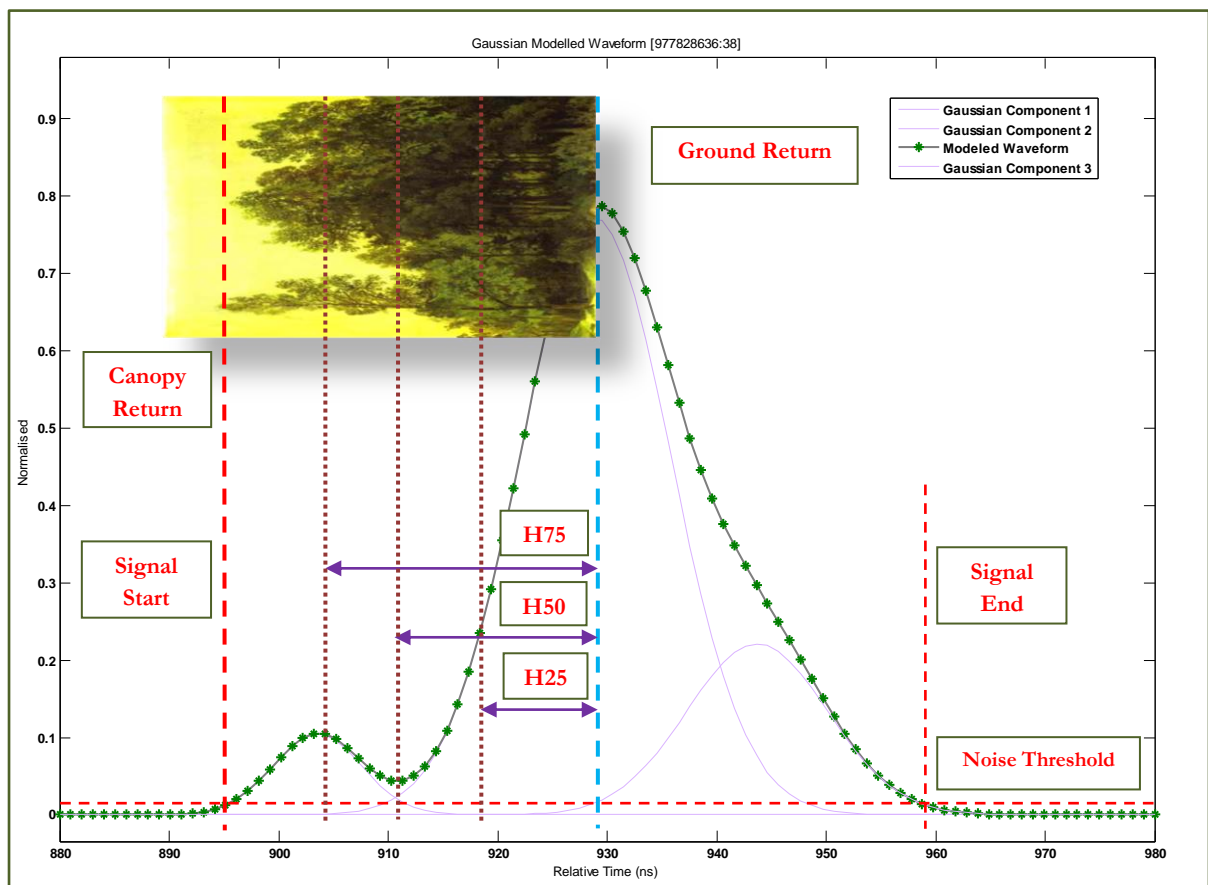


Figure 2-12: GLAS Waveform Quartile Based Parameters [H25, H50 and H75]

2.5. Data analysis

In order to find out the relationship between the GLAS waveform derived parameters and forest structural attributes, namely tree height and AGB, a neuro-genetic approach was adopted to examine the best combination of waveform derived input parameters which have the best potentiality to estimate tree height and AGB.

2.5.1. Neuro-Genetic Model Prediction

In the present study, the Discipulus and DTREG are used for implementation of EANN to extract tree height and AGB from waveform derived parameters. The simulation environment has modules for data transformation, selection, training, testing, optimization and validation.

GLAS waveform derived parameters selected for this study are [*wDistance*, *R25*, *R50*, *R75*, *H25*, *H50*, *H75*] which constituted the initial population of chromosomes. Each of the parameter is used as input to GA in Discipulus to find the best combination of input parameters for predicting tree height and AGB.

After evolution of the best input combination, the parameters so selected were used to train EA neural network to predict tree height and AGB. For testing the efficiency of EANN over ANN, a MLP model was also developed for comparing the outputs of two models. As discussed earlier (section 1.2.4.1), the performance of the MLP is based on efficient training method. The aim of the training process is to find the optimum set of weight values that will result into the output to approximate with the desired values. Training involves selection of hidden layers, number of neurons in each hidden layer, finding a globally optimal solution and validation of the network to test for overfitting. The training method used in the study is the BP (Back-propagation) with *logistic* activation function in hidden layer. Several variations of the BP network architecture were tested in order to search for the appropriate architecture. After repeated tests, a network with one hidden layer with 10 neurons was selected for predicting tree height and one hidden layer with 7 neurons for predicting AGB.

2.5.2. Validation

Cross-validation method (Kohavi, 1995; Krogh and Vedelsby, 1995) is selected for validating the output of GA based input parameter optimization. Cross – validation was performed to minimize the problem of overfitting and selecting the most relevant input parameters for NN model in order to achieve a reasonable prediction of tree height and AGB. Cross-validation was carried out by dividing the validation data into a number of subsets and sequentially testing each subset with the GA output from remaining subsets. For this study the validation data was grouped into 10 subsets for the purpose of cross-validation. Further, the predictability of EANN is compared with the output derived from MLP model.

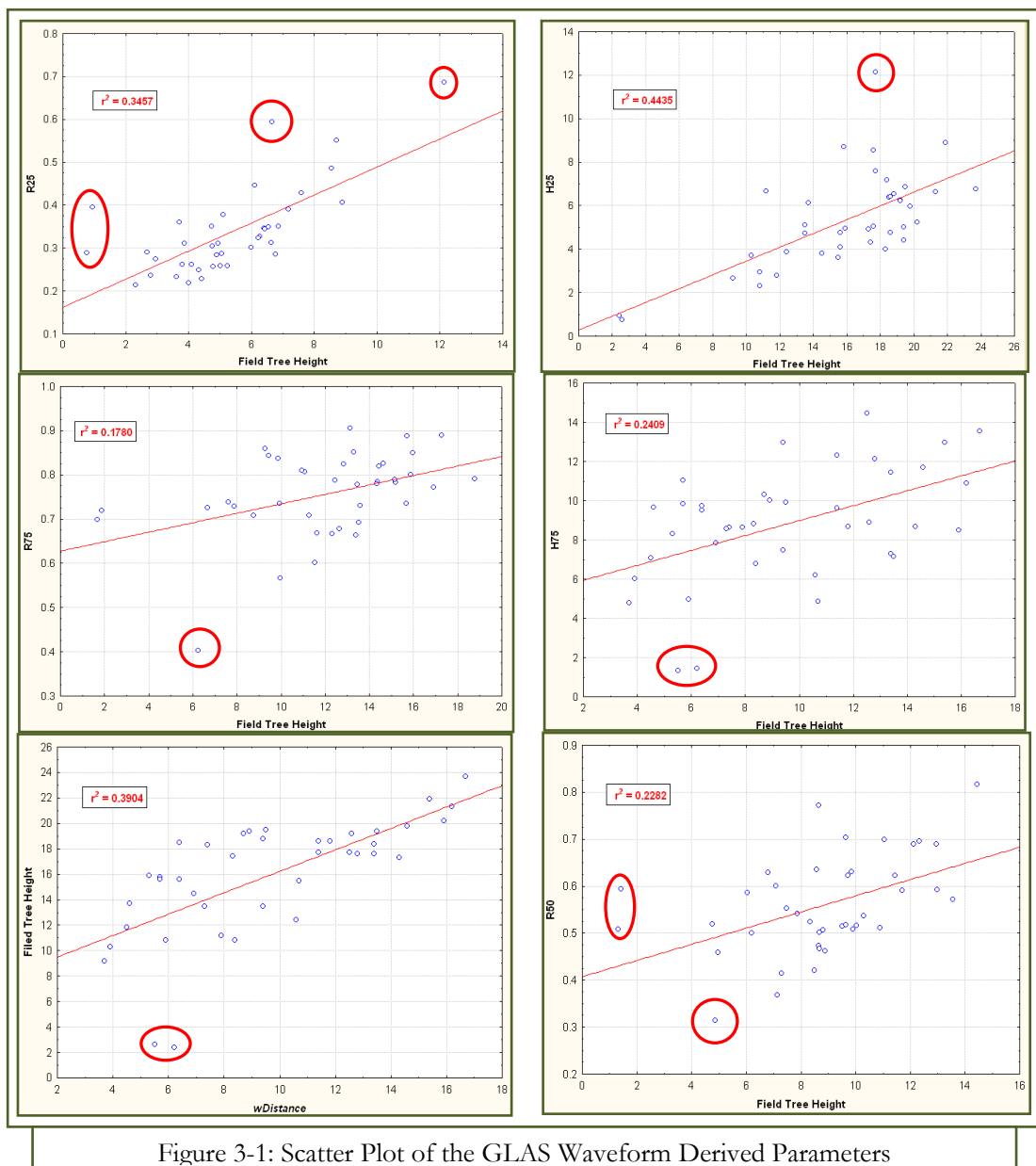
2.5.3. AGB change trend analysis for monitoring forest degradation

The GLAS predicted AGB is compared with the AGB estimates of the study area carried out by Roy and Ravan, (1996) for analysing the change in AGB. Based on the negative or positive change trend, AGB estimate of forest provided a preliminary indication of forest degradation or increase in forest cover.

3. RESULTS AND DISCUSSION

3.1. Preliminary analysis of GLAS derived waveform parameters

Some statistical analysis was carried out on GLAS waveform derived indices in order to get a preliminary overview of the relationship between indices and tree height. Figure 2-13 shows some results of the statistical analysis between waveform parameters and field tree height. It is clear that the variability of data is high or there may be some outliers. Further, none of the waveform parameter alone is capable of estimating tree height with reasonable accuracy. Therefore, various combinations of input metrics were tested to optimize the predictability of the model.



3.2. Neural Network input parameter optimization using EA

3.2.1. Results

As the project focused on estimation of forest structural attribute for rapid assessment of AGB, it becomes imperative that the variables required for model input should be easily extracted and produce results with reasonable accuracy. Hence, an evolutionary approach is implemented for selection and optimization of inputs.

The GLAS derived waveform metrics as proposed earlier (Section 2.5.1) [H25, H50, H75, *wDistance*, R25, R50, R75] corresponding to [Input (0-6) respectively; Figure (3-2)] were subjected to EA as input population and field tree height as target parameter. The inputs were optimized after evolving through a number of generations and best performing individuals were selected based on fitness function (Figure 3-3). The optimization, selection and validation results of EA are presented in this section.

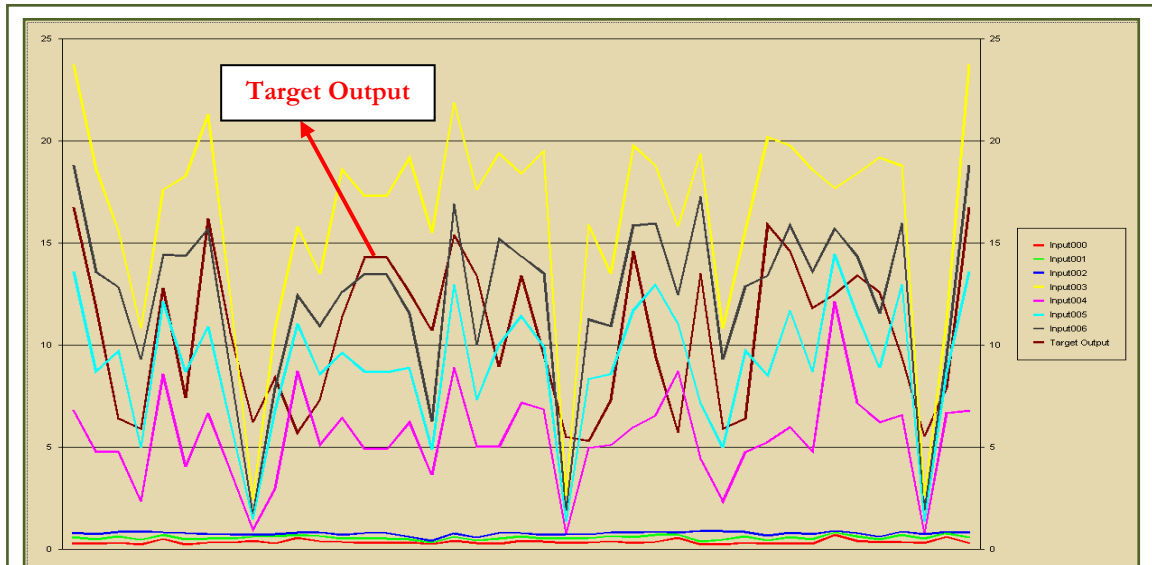


Figure 3-2: Input and Target Parameter for EA

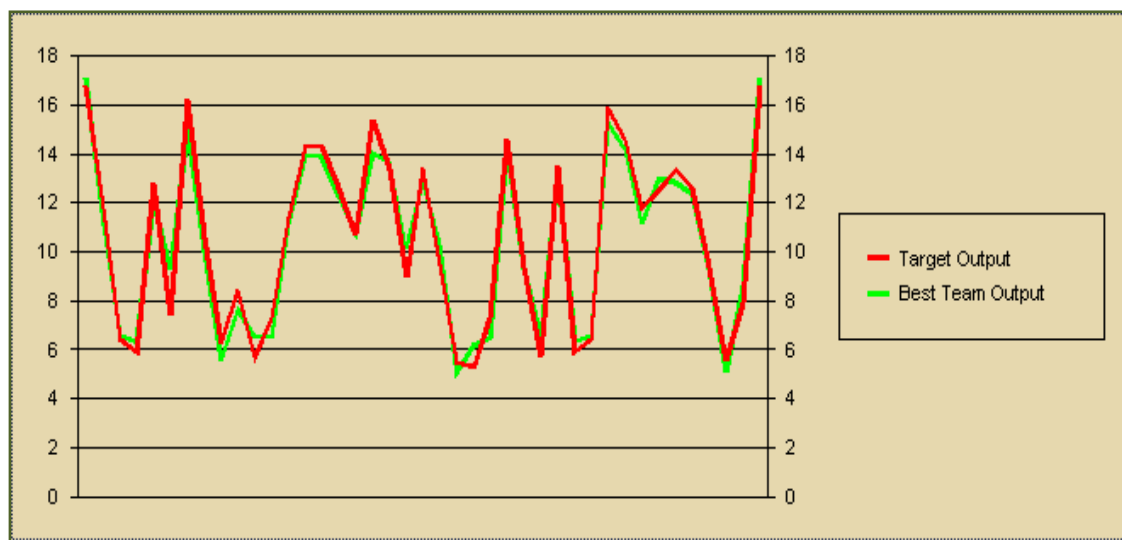


Figure 3-3: Best Team Selection Result

As inferred from the Figure (3-2), the *wDistance* curve (yellow line) resembles closely with the target output curve, hence have a greater chance of inheritance to the next generation followed by *R75* (black line). This inference is confirmed by the impact factor (Table 3-1) of each input parameter.

A total of 20000 generations were tested with a crossover rate of 67.76 and mutation rate of 76.62. The input impact or contribution of each waveform derived parameter to the next generation is presented in Table 3-1.

Table 3-1: Generation Impact Factor of GLAS Derived Waveform Parameter

Waveform Derived Parameters	Impact
<i>R25</i>	0.42454
<i>R50</i>	0.28953
<i>R75</i>	0.67839
<i>H25</i>	0.13353
<i>H50</i>	0.04475
<i>H75</i>	0.38397
<i>wDistance</i>	0.82681

As evident from Table (3-1), waveform distance (*wDistance*) has maximum impact followed by *R75* and *R25*. Several random combination of were generated and evaluated by a fitness function based on impact values. Some proposed combinations of input parameter obtained using EA are given in Table 3-2.

Table 3-2: GLAS Derived Waveform Parameters Input Combinations

Team Number	Fitness	Waveform Parameter Combination	Team Rank
1	0.731903863042	<i>wDistance</i> , <i>R25</i> , <i>R50</i> , <i>R75</i>	2
2	0.459517925978	<i>wDistance</i> , <i>H25</i> , <i>H50</i> , <i>H75</i>	5
3	0.695538009167	<i>wDistance</i> , <i>R25</i> , <i>H50</i> , <i>R75</i>	4
4	0.652529847622	<i>R25</i> , <i>R75</i> , <i>H25</i> , <i>H75</i>	3
5	0.845910310742	<i>wDistance</i> , <i>R25</i> , <i>R75</i> , <i>H75</i>	1

The EA results show that the team number 5 having the highest fitness value evolved as the best combination and is selected as the input parameter for ANN model. It was followed by team 1, 4, 3 and 2.

3.2.2. Validation

The findings of the EA were validated using cross validation technique (Section 2.5.2). The results of the validation constitute this section. A 10 – fold cross validation is adopted with training and validation split of 80 % and 20 %. The data set is divided into 10 folds with each fold containing training and validation set. The results of the cross validation reconfirms the training output with team 1 having the highest fitness value. The *wDistance* curve (Figure 3-4) again shows similar variation pattern with target output (brown line) followed by *R75*. The fitness curve of best team and target output is shown in Figure 3-5. The variation (blue circle) in fitness may be due to presence of some outliers in validation data.

Table 3-3: GLAS Derived Waveform Parameters Input Combinations (Validation)

Team Number	Fitness	Waveform Parameter Combination	Team Rank
1	2.875527410507	<i>wDistance, R25,R50,R75</i>	1
2	2.234132289886	<i>wDistance, H25,H50,H75</i>	5
3	2.264853715897	<i>wDistance, R25,H50,R75</i>	3
4	2.236797809601	<i>R25,R75,H25, H75</i>	4
5	2.268614768982	<i>wDistance, R25, R50,R75</i>	2

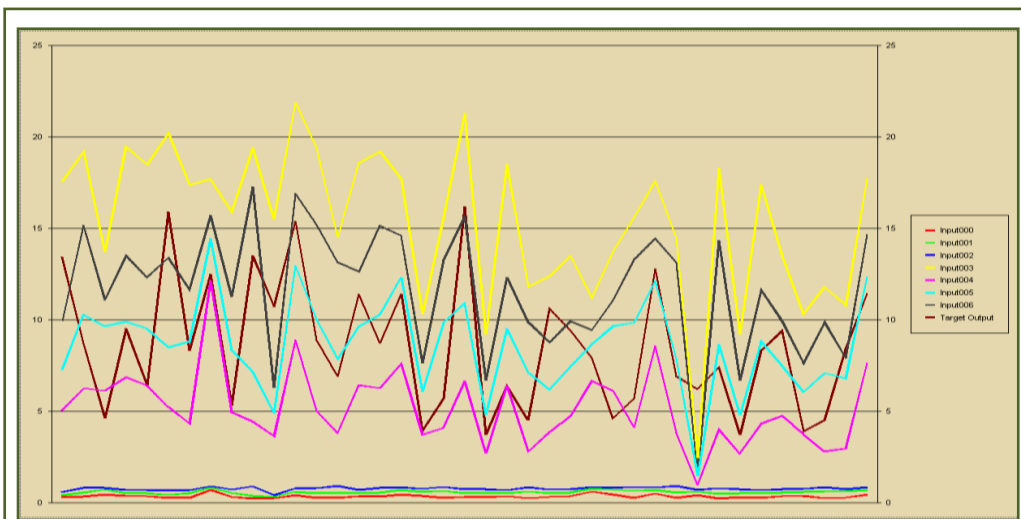


Figure 3-4: Input and Target Parameters for Validation

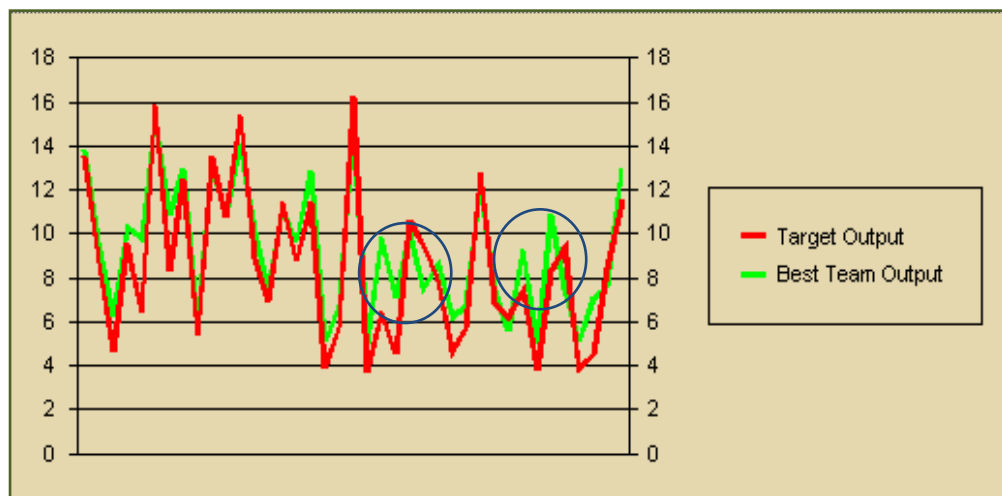


Figure 3-5: Best Team Selection Result

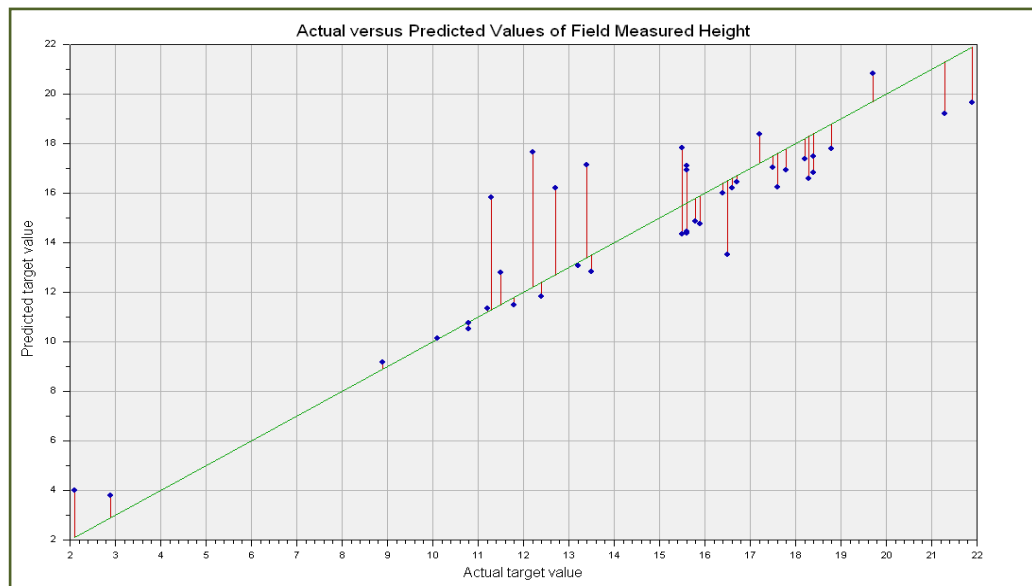
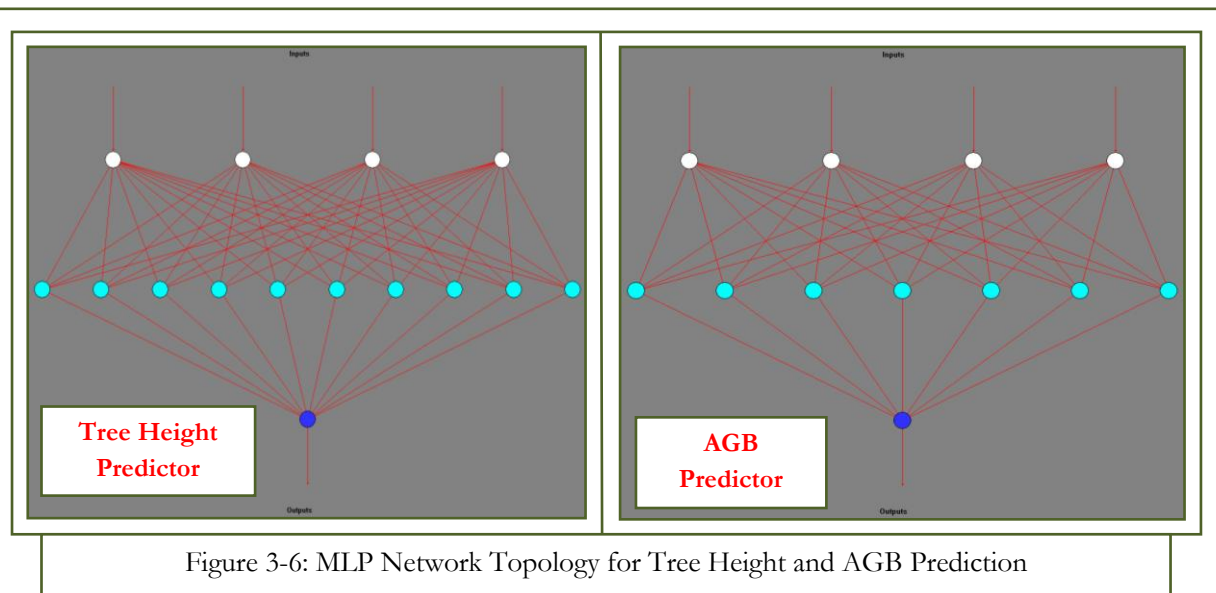
Table (3-3) shows the validation output of EA. The best combination of selected input matches with the evaluation and selection results. Hence the obtained combination can be used as input for ANN modeling for predicting tree height and AGB.

3.3. Neuro-Genetic Approach

3.3.1. Multi-Layer Perceptron (MLP) Neural Network

3.3.1.1. Results and validation

The MLP neural network is used for predicting tree height and AGB from GLAS waveform metrics. The results produced by MLP are described in this section. Figure (3-6) shows the MLP network topology with one hidden layer containing 10 nodes for tree height prediction and one hidden layer with seven nodes for AGB prediction. Figure (3-7) shows the correlation between field maximum tree height measurements and height predicted by trained MLP model using GLAS waveform derived parameters.



Vital statistics for assessing the performance of the model is given in Table 3-4. The coefficient of determination signifies that how much better we can do in predicting the tree height using MLP model. A R^2 value of 0.79892 means that 79 % of variability observed in GLAS derived tree height can be explained by field measured tree height.

Table 3-4: MLP Tree Height Predictor Network Statistics

MLP Model Statistics	Training Data	Validation Data
(R^2)	0.79892 (79.892%)	0.75139 (75.139%)
RMSE	1.8145077	2.0696915

The relative importance of GLAS waveform derived metrics in predicting tree height is shown in Figure 3-8. It is clearly indicated that *wDistance* is closely related with tree height followed by H75 height quartile.

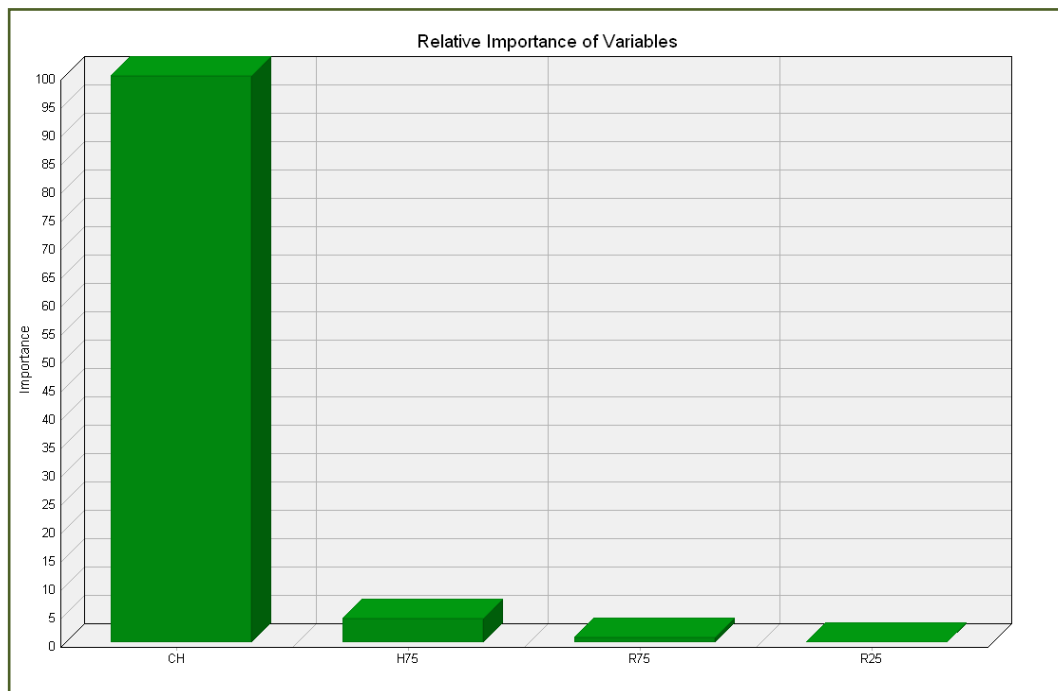


Figure 3-8: Relative Importance of Waveform Derived Input Parameters

A MLP model with one hidden layer containing seven nodes was used for predicting AGB from GLAS waveform derived parameters. Figure (3-9) represents the correlation between field measured AGB and AGB predicted by trained MLP model from GLAS waveform derived indices. Table (3-5) gives the statistics for analysing the model performance.

Table 3-5: MLP AGB Predictor Network Statistics

MLP Model Statistics	Training Data	Validation Data
(R^2)	0.66073 (66.073%)	0.63139 (63.139%)
RMSE	40.445836	43.811122

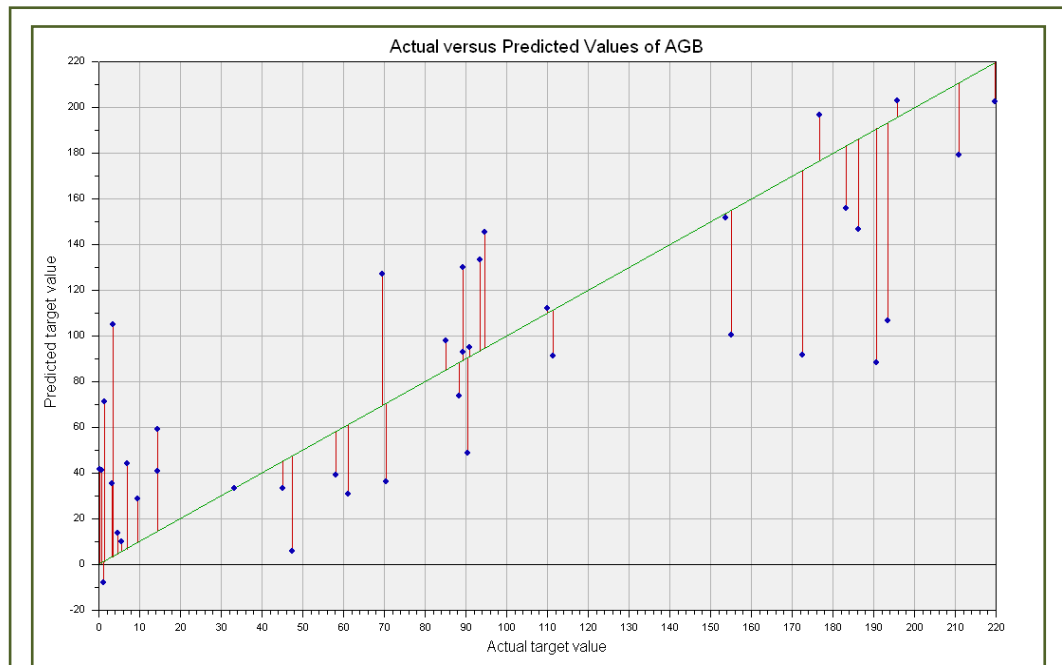
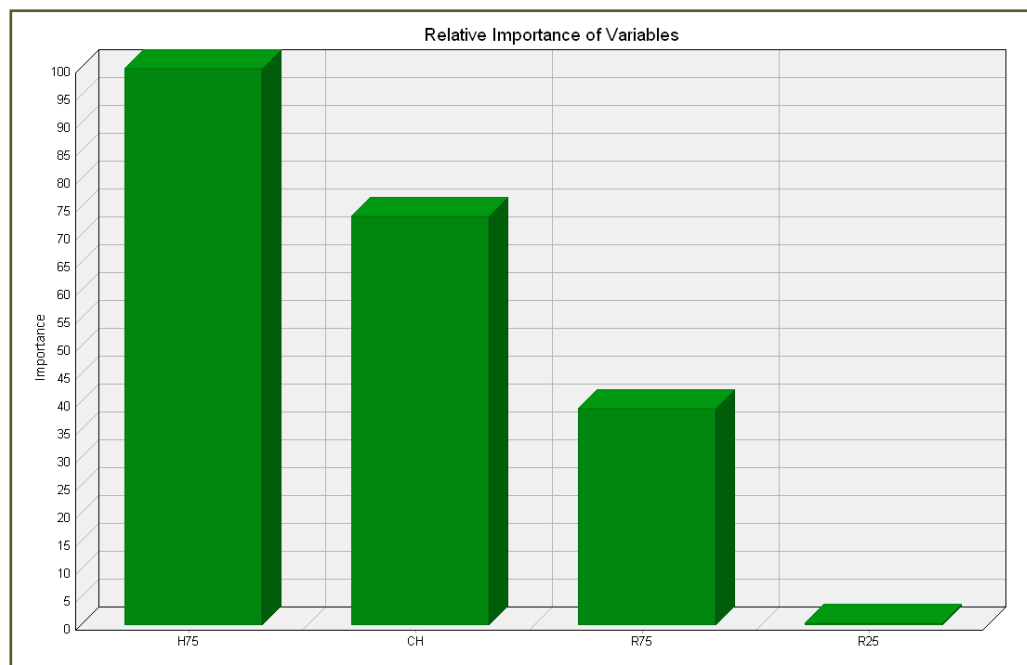
Figure 3-9: GLAS Derived AGB *vs* Field AGB

Figure 3-10: Relative Importance of GLAS Derived Indices for AGB Estimation

The relative order of importance of GLAS derived waveform indices are represented in Figure 3-10. The most important parameter is H75 height quartile followed by *wDistance*.

3.3.2. EANN based AGB & tree height prediction model

3.3.2.1. Results and validation

The results of the EANN model for predicting tree height and AGB constitutes this section. Figure (3-11) shows the correlation between tree height measured in field and height predicted by the waveform derived parameters using EANN model.

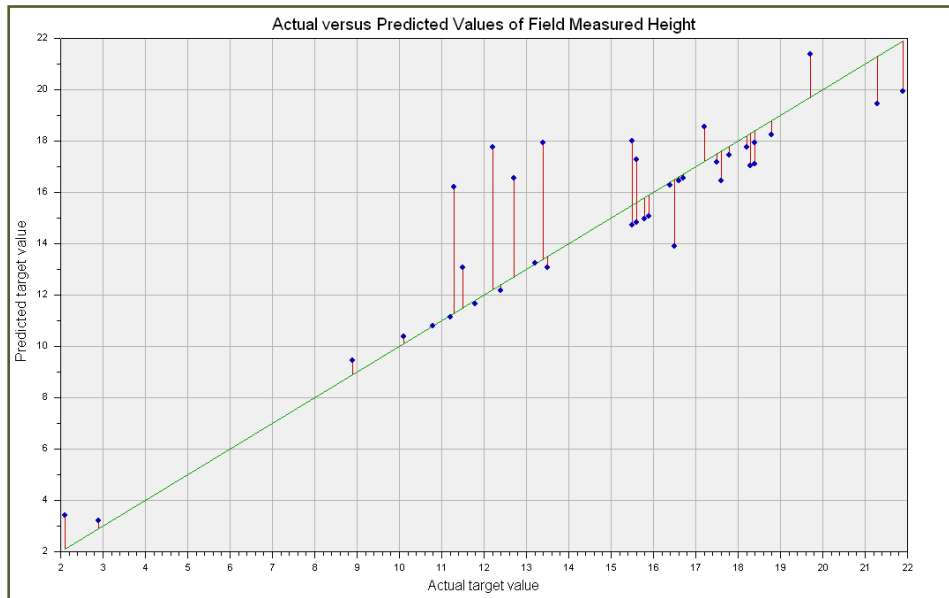


Figure 3-11: GLAS Height vs Field Height Predicted by EANN

Table (3-6) shows the parameters for evaluating the performance of the model.

Table 3-6: EANN Tree Height Prediction Model Statistics

EANN Network Parameters	Training Data	Validation Data
(R²)	0.80785 (80.785%)	0.78412 (78.412%)
RMSE	1.8195845	1.9286752

The model thus obtained after evolution and selection for predicting tree height from GLAS waveform variable is represented in equation (3.1).

$$Height_{Tree} = \frac{wDistance}{\sqrt{[(2 * R^2) + 1.265]}} \quad (3.1)$$

The resulting equation explains 80.7 % variability in predicted tree height from waveform derived metrics.

Figure (3-12) shows the correlation between the predicted AGB using EANN from GLAS derived indices. The model performance evaluation parameters such as R^2 as shown in Table (3-7) indicates that the predictability of EANN has improved as compared with the MLP predicted AGB.

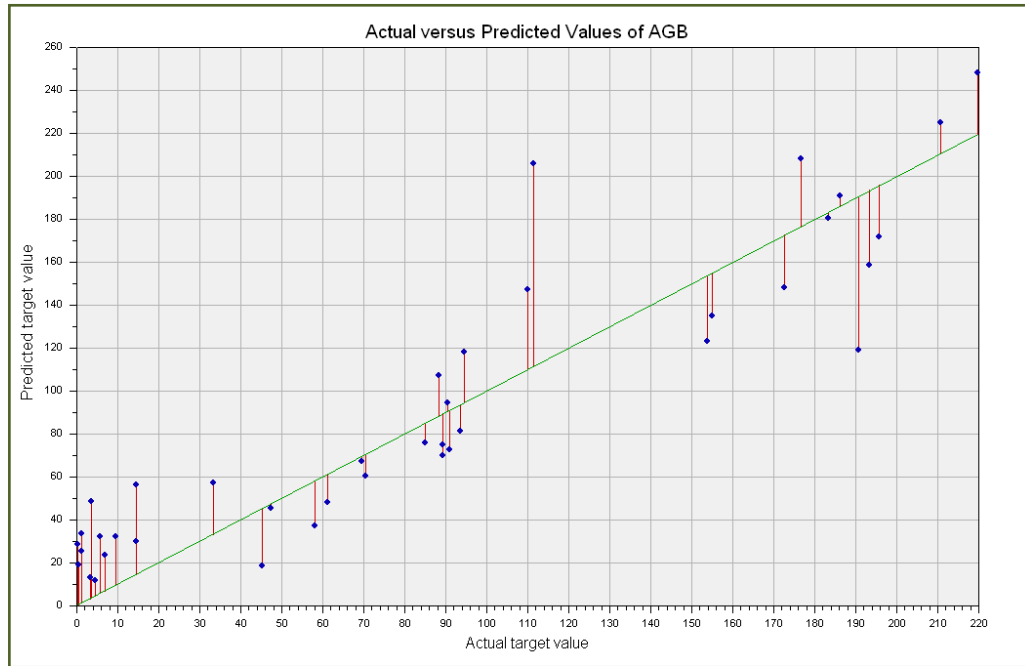


Figure 3-12: GLAS Derived AGB vs Field Measured AGB by EANN Model

Table 3-7: EANN based AGB Predictor Model Statistics

EANN Network Parameters	Training Data	Validation Data
(R^2)	0.69132 (69.132%)	0.64152 (64.152%)
RMSE	29.032746	33.041511

AGB estimation model thus obtained using EA for assessment of AGB is expressed in equation. This equation explains 69 % variability in AGB prediction based on GLAS derived parameters.

$$AGB = (0.5 * H75 * wDistance)^2 - 2.87 \quad (3.2)$$

3.4. Comparative analysis of GLAS predicted AGB and AGB estimates from past inventory

The GLAS predicted AGB estimate was compared with the AGB estimate of the study area based on the past study carried out by Roy and Ravan, (1996) in Madhav National Park (Figure 3-13) situated in Shivpuri. As the past data was available only for the Shivpuri region so GLAS footprints falling inside the Madhav National Park only was considered and analysed for AGB change.

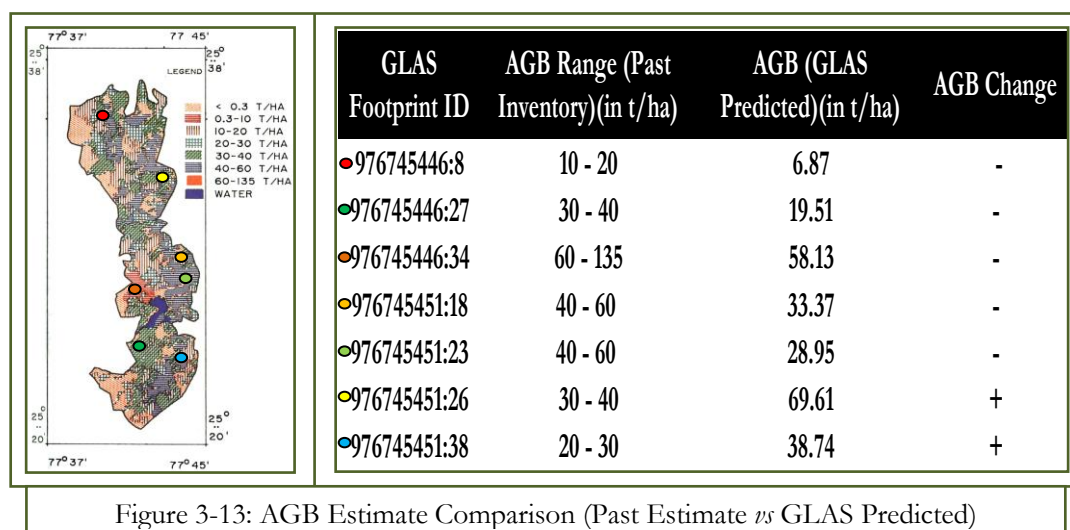


Figure 3-13: AGB Estimate Comparison (Past Estimate vs GLAS Predicted)

As evident from the Figure (3-13), in most of the parts of the Madhav National Park there is a loss of AGB. This fact was also confirmed during field observation as there is a lot of illegal cutting of the trees (Figure 3-14). Hence, a degrading state of forest can be inferred from the negative change in AGB estimate. However a more appropriate comparison would consist of AGB estimates as predicted by GLAS data from two different time periods. This would provide an improved and quick method for monitoring the impact of forest degradation.



Figure 3-14: Field photographs showing illegal cutting of trees

3.5. Discussion

The present study focussed on further investigating the applicability of the full waveform large footprint spaceborne LiDAR for extraction of forest structural attributes like tree height and AGB. Owing to its unique capability to extract forest vertical structural information, this is an ideal tool for monitoring regional carbon stock with enhanced precision. There have been several studies on AGB estimation using spaceborne LiDAR (Lefsky *et al.*, 2005, Lefsky *et al.*, 2007) but the application of evolutionary techniques for AGB assessment are yet to be explored.

The present research attempted to develop a new method to retrieve tree height and AGB from large footprint full waveform LiDAR and to serve as a benchmark study for development of EA models for AGB studies. To achieve this purpose, the raw waveform signal was decomposed into Gaussian components, and ground and canopy return from the waveform were delineated. Based on previous studies (Lefsky *et al.*, 2007; Sun *et al.*, 2007, 2008; Zhang *et al.*, 2009; Duong, 2010), several waveform parameters assumed to be related with vertical structural attributes of the forest were extracted from processed waveforms. The proposed set of waveform derived indices is [*H25*, *H50*, *H75*, *R25*, *R50*, *R75*, *wDistance*]. In order to identify the most relevant set of indices, GA was applied to select the best set combination of parameters for extracting tree height and AGB. The identified set of metrics was [*wDistance*, *R75*, *R25* and *H75*]. After the selection of the best input parameters, EANN approach was used to predict tree height and AGB. To test the superiority of the neuro-genetic approach, a conventional ANN model, MLP was also used for prediction of tree height and AGB. Based on the findings of EA model, it was observed that the model is able to explain 80.7 % of the variability in prediction of tree height and 69.13% variability in prediction of AGB which is a bit higher than the results of the MLP model having R^2 values of 79.8% and 66 % for tree height and AGB estimation respectively.

The result of tree height prediction model using EA found that the most useful waveform derived indices to predict average tree height were *wDistance* and *R75*. Of these, *wDistance* had the highest correlation. The *wDistance* increases with the canopy height. *R75* decreases with the canopy height as more energy is reflected from upper canopy leaving less energy for reflection from ground.

Sun *et al.*, (2007) proposed an ANN model to predict maximum tree height and AGB using GLAS waveform derived metrics on flat terrains. The model used 5 input parameters derived from total length of waveform, quadratic mean canopy height, 25% and 50% waveform quartiles, with 3 nodes in hidden layer for predicting tree height (Sun *et al.*, 2007). The model was able to explain 75% variability in prediction of tree height with a RMSE of 3.4 m. The AGB prediction model which was proposed by Sun *et al.*, (2007) used six input parameters derived from total length of waveform, top canopy height, 25% and 75% waveform quartiles, with 10 nodes in hidden layer. The model was able to explain 73% of variability in AGB prediction with a RMSE of 30.43 Mt/ha. They recommended that similar results could be obtained in areas with steep slope. The results obtained from the present study showed a slight improvement over the results of Sun *et al.*, (2007) as 6 GLAS footprint measurements were obtained from areas with moderate slope. As the aim of this study was to serve as a benchmark model for further studies using EA for tree height and AGB prediction, the results are promising in this context and further studies are required to validate the approach in areas with steep slope.

The analysis of error statistics showed that there is no significant improvement observed in tree height prediction using EA model as compared to MLP model. However the prediction has improved in case of AGB estimation using EA model having a reported R^2 of 69 %. There is no observation that the input variable data in the training period have been over-fitted that is, the performance of the model does not

deteriorates much significantly during validation period as evident from the R^2 values of training and validation phase. Further, there are no absolute deciding criteria for a good value of RMSE. It depends on the method of our measurements. The tree height values in our study area ranges from 3.7 m to 19 m. The RMSE as obtained from EA model is 1.8 m which is less than 50% of the minimum value of tree height and 25% of the mean and range of the field measured tree heights. Hence the RMSE value is reasonable as the possible error in the minimum tree height value is less than 50%. The R^2 values are also reasonably high. This agreement can be explained with the fact that the majority of the trees within the sampled footprints are of uniform heights, hence there are no possible residual error compensation among short and the tall stands.

However, the predictability of the AGB model is relatively poor. This underperformance of model in AGB estimate may be attributed to the density of the forest stand. The majority of the study area consists of open forest due to lot of trespassing and illegal felling of the forest stand by local inhabitants. In open forest conditions, the competition for sunlight and nutrient resources are less as compared to the closed forest. As a consequence, the trees show more gain in DBH as compared to height (Taiz and Zeiger, 2011). AGB is a function of DBH and less related to height. It can be inferred that on relatively flat terrain, the $wDistance$ is a function of tree height and not DBH. Hence a high value of RMSE is observed. The waveform derived metrics used in this work relies mainly on height attributes and, mainly on maximum tree height within the sampled footprint. This consideration may result in larger sampling error in the field condition as inferred by Lefsky *et al.*, (2007). The future studies should investigate the average tree height given by peak distance of modelled waveform, which is considered to have stronger correlation with AGB (Means *et al.*, 1999; Lefsky *et al.*, 1999b; Anderson *et al.*, 2006, Anderson *et al.*, 2008).

The performance of the GA depends on the stochastic operators such as selection, recombination, crossover and mutation. Our GA used only one crossover and mutation rate of 67% and 76% respectively to generate next generation. The selection of appropriate crossover and mutation rate is critical (Grefenstette, 1986; Schaffer and Morishima, 1987) and often it is difficult to determine the appropriate values and the choice is mainly carried out using a trial-and-error procedure. The same constraint inherits the evolutionary model for tree height and AGB prediction with a crossover rate of 47% and mutation rate of 58%. These factors may account for the possible error in AGB and tree height estimates. Furthermore, the choice of suitable crossover and mutation operator is also important. This study uses *single point* crossover operator and *change* mutation function to generate the optimized input subset. Future studies should test other robust mutation and crossover function such as *multipoint* crossover or *swapping* mutation function. In addition to this the robust fitness function is necessary for evaluating and selecting the best individual from the population. This study has used default fitness function of mean square error provided in the software both for input optimization and prediction model. The choice and formulation of fitness function is still an area under investigation providing a wide scope for further research.

Besides all these constraints, the results of GA for input optimization seems to be reasonable as in the prediction models, approximately 50% of variance is explained within first few generations (Figure 3-14) showing that the input population have related trend. Also, it can be deduced that a higher crossover and mutation rate may result into more accurate outputs. However this inference needs to be tested with further studies.

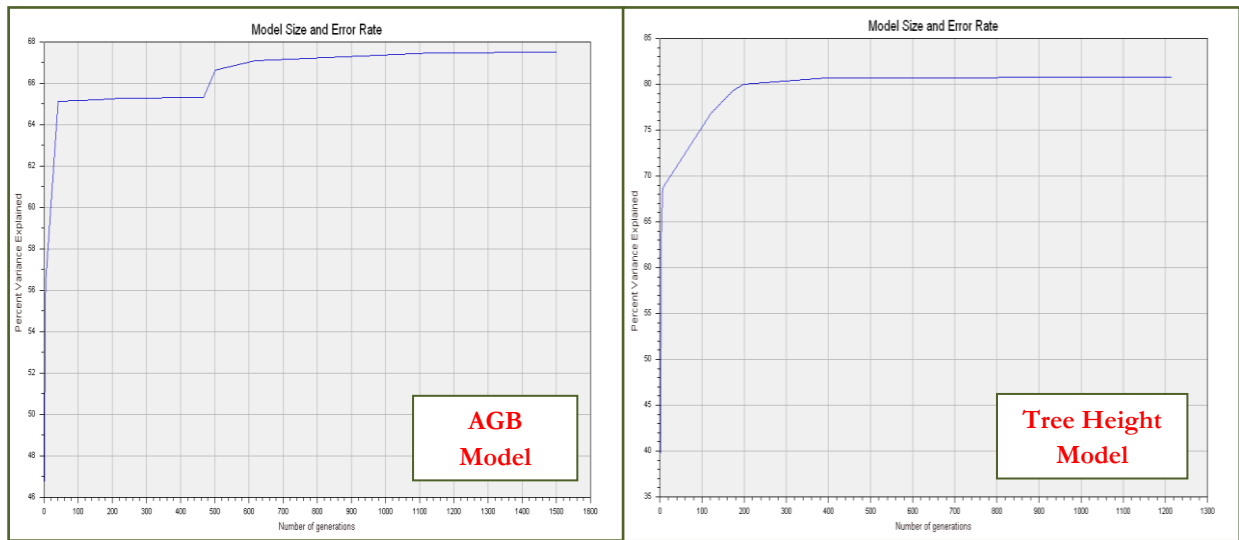


Figure 3-15: R^2 vs Number of Generation Plot for AGB and Tree Height Prediction Model

All the aforementioned considerations and modeling efforts would be effective if our understanding of GLAS waveform modeling and extraction of proposed indices are accurate. One of the major obstructions in application of GLAS is multi-surface scattering within the footprint. A clear and thorough understanding of underlying properties of waveform signal propagation and their interaction with surface feature is necessary for accurate interpretation of GLAS waveform and derived indices (Jutzi and Stilla, 2006; Wagner *et al.*, 2006; Reitberger *et al.*, 2008). The nature of the return laser pulse corresponds to the material composition and the geometric properties of the reflecting surface feature and the energy distribution of the emitted laser signal. All these factors are of prime importance and should be given a due consideration while modeling the return laser pulse. In context with the forest tree height and AGB modeling, these modifications and improvement in the modeling of the waveform may enhance the distinguishability between the canopy return and the return from other surface features.

The return waveform pulse is a convolution function of transmitted waveform or system response and surface feature response (Jutzi and Stilla, 2006; Neuenschwander, 2008; Duong, 2010). Hence waveform deconvolution is required to remove the system response for extracting the surface feature response. Duong, (2010) proposed a novel technique for deconvolution of waveform using a Wiener filter and inverse fast-Fourier transformation. The author concluded that deconvolution method can accurately measure the difference between top and bottom of the object thereby making height estimations more accurate.

The promises of reasonable prediction of AGB using large footprint full waveform spaceborne LiDAR, an important deduction of this work, opens up a new paradigm for monitoring forest degradation. An example is included in this study, by comparing AGB estimate from past studies and GLAS predicted AGB. The AGB change trend in the study area showed that there is a loss of forest cover. The observations made during field work corroborate the inference. As Madhav National Park is located close to the human settlement, the local inhabitants are dependent on the forest for fuel, fodder and other needs. Moreover, the presence of commercially important tree species such as *Acacia catechu* attracts traders resulting in rapid loss of forest cover. A rapid estimation technique for AGB assessment will provide more frequent AGB stock estimates, thus enabling the authorities effectively monitor the forest degradation. Furthermore, the multi-temporal nature of spaceborne LiDAR can provide data for periodic assessments

of AGB and subsequently carbon stock thus supporting the ongoing effort for implementation of uniform and transparent global carbon credit system.

Pertaining to a short MSc research period, it was not feasible for the author of this work to incorporate all the improvements suggested in the discussion. The idea was to test the evolutionary approach for extracting vertical forest attributes using large footprint full waveform spaceborne LiDAR. Hence, a very basic model was implemented in all the modeling procedure and much consideration was given to simplicity of the approach rather than complex waveform parameter derivation and intricate neuro-genetic modeling. The outcomes of the study fulfilled the overall objective and gave a reasonable estimate of tree height and AGB and meanwhile opening up a gateway for future investigations with more accurate model, more resources and more time.

4. CONCLUSIONS AND RECOMMENDATIONS

4.1. Conclusions

In this work, a method for extracting tree height and AGB from ICESat GLAS waveform is developed. It has been demonstrated from this study that tree height and AGB can be directly retrieved from GLAS derived waveform metrics with reasonable accuracy. Hence, the aim of investigating the applicability of neuro-genetic approach for predicting tree height and AGB from ICESat GLAS thus extending the applicability of full waveform large footprint spaceborne LiDAR to facilitate the effective monitoring of forest degradation through analysis of temporal changes in AGB content. Hence the study was successful to achieve all the predefined objectives and answer all the associated research questions.

Research Question 1: *To what extent is the EANN approach is capable of predicting tree height and AGB using GLAS waveform derived metrics?*

Answer: From the outcomes of the study, it has been concluded that EANN model so developed is able to predict tree height with a RMSE of 1.8 m and AGB with RMSE of 29 t/ha. The study also revealed that the GLAS footprints can provide a reasonable estimate of AGB level. The model has been developed from the observations and measurements obtained from the GLAS footprint in areas with relatively gentle to moderate slope. Hence the model is expected to perform better in flat terrain and the performance in areas with steep slope is still a subject of further investigation. One primary observation made from the visual analysis of GLAS waveform obtained from footprints from moderate terrain areas is that the number of Gaussian peaks obtained after the waveform decomposition and modeling may prove useful in tree height estimation in areas of steep slope as the number of peaks increases with the complexity of the surface terrain.

Research Question 2: *Which GLAS waveform derived parameters are most relevant for extracting tree height?*

Answer: The indices identified after application of GA to GLAS waveform derived parameters are *wDistance*, *R75*, *H75* and *R25*. These parameters were identified on the basis of the generation impact factor that is by assessing the relative contribution of each parameter to the final target value which is tree height. *wDistance* is found to have maximum impact hence more related to the tree height.

Research Question 3: *What is the change in AGB as compared with past AGB estimate in the study area?*

Answer: The comparison shows that the AGB estimate as predicted by GLAS data has decreased as compared to the AGB estimate from the past inventory. Though this comparison is only basic and indicative, but do signifies a degrading forest and more accurate change analysis can be done by utilizing the periodic data availability from spaceborne LiDAR and enhancing the accuracy of AGB prediction models based on these LiDAR systems.

The outcomes of this study demonstrates that with subsequent improvements and more representative studies, full waveform large footprint spaceborne LiDAR can be used for AGB and carbon stock

inventories on global scale with enhanced precision. The analysis of the study framed up issues that needed critical considerations to improve the accuracy of model developed in this study. The product of this research was influenced by several factors such as period of GLAS data acquisition and the period of the present research work differed by approximately 2 years. This may be a possible source of error. Forest management practices and conditions of study area have a direct impact on the model performance. Hence a thorough consideration on all the aspects raised in the present work is required before developing a robust model for global biomass monitoring.

4.2. Recommendations

All the future investigations carried out should consider the following issues in order to improve the ongoing efforts of this research theme:

- Use of Dynamic Genetic Algorithm (DGA) is recommended for AGB modeling as it has the capability to use more than one mutation and crossover operators for evolving next generation thus enhancing the efficiency of the GA approach.
- More robust signal processing techniques can be implemented other than Gaussian modeling for extracting waveform derived metrics more accurately. Waveform deconvolution is suggested in subsequent modeling effort to achieve a clear differentiation among reflecting surface features.
- So far estimation of AGB has been carried out on footprint level only. Future works should consider integration of different sensor products such as VHR imagery, PolInSAR, airborne LiDAR in combination with geostatistical techniques for extending the AGB estimate over the entire region.
- More number of representative field samples should be considered, especially samples with different slope classes for effective AGB modeling and validation.
- The future work based on the themes of the present study should be tested under different scenarios and varied field conditions for ensuring their repeatability and transferability.

LIST OF REFERENCES

- Afandizadeh, S.H., Kianfar, J., 2009. A Hybrid Neuro-Genetic Approach to Short-Term Traffic Volume Prediction. *International Journal of Civil Engineering*, 7, 41–48.
- Anaya, Chuvieco, E., Palaciosorueta, A., 2009. Aboveground biomass assessment in Colombia: A remote sensing approach. *Forest Ecology and Management*, 257, 1237–1246.
- Anderson, J., Martin, M., Smith, M., Dubayah, R., Hofton, M., Hyde, P., Peterson, B., Blair, J., Knox, R., 2006. The use of waveform LiDAR to measure northern temperate mixed conifer and deciduous forest structure in New Hampshire. *Remote Sensing of Environment*, 105, 248–261.
- Anderson, J.E., Plourde, L.C., Martin, M.E., Braswell, B.H., Smith, M.L., Dubayah, R.O., Hofton, M.A., Blair, J.B., 2008. Integrating waveform LiDAR with hyperspectral imagery for inventory of a northern temperate forest. *Remote Sensing of Environment*, 112, 1856–1870.
- Asrar, G., Dozier, J., 1994. EOS: science strategy for the earth observing system. *American Institute of Physics*.
- Azzini, A., 2005. *A New Genetic Approach for Neural Network Design and Optimization*. Universita degli Studi di Milano, Milan, 151.
- Benson, M., Pierce, L., Bergen, K., Sarabandi, K., Kailai Zhang, Ryan, C., 2011. *Forest structure estimation using SAR, LiDAR, and optical data in the Canadian Boreal forest*, In: Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International. Presented at the Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International, IEEE, pp. 2609–2612.
- Bishop, C.M., 1995. *Regularization and complexity control in feed-forward networks*, In: Proceedings International Conference on Artificial Neural Networks ICANN'95. Birmingham, pp. 141–148.
- Blair, J.B., Hofton, M.A., 1999. Modeling laser altimeter return waveforms over complex vegetation using high-resolution elevation data. *Geophysical Research Letters*, 26, 2509–2512.
- Bortolot, Z.J., Wynne, R.H., 2005. Estimating forest biomass using small footprint LiDAR data: An individual tree-based approach that incorporates training data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 59, 342–360.
- Boudreau, J., Nelson, R.F., Margolis, H.A., Beaudoin, A., Guindon, L., Kimes, D.S., 2008. Regional aboveground forest biomass using airborne and spaceborne LiDAR in Quebec. *Remote Sensing of Environment*, 112, 3876–3890.
- Brenner, A., Zwally, H., Bentley, C., Csatho, B., Harding, D., Hofton, M., Minster, J., Roberts, L., Saba, J., Thomas, R., others, 2003. Geoscience Laser Altimeter System (GLAS): Derivation of Range and Range Distributions from Laser Pulse Wave form Analysis for Surface Elevations, Roughness, Slope, and Vegetation Heights. *Algorithm Theoretical Basis Document-Version 4.1*, 4, 92.
- Brill, F.Z., Brown, D.E., Martin, W.N., 1992. Fast generic selection of features for neural network classifiers. *IEEE Transactions on Neural Networks*, 3, 324–328.
- Cybenko, G., 1989. Approximation by superposition of a sigmoidal function. *Mathematics of Control, Signals, and Systems (MCSS)*, 2, 303–314.
- Dobson, M.C., Pierce, L.E., Bergen, K., Kelndorfer, J., Ulaby, F.T., 1995. *Retrieval of above-ground biomass and detection of forest disturbance using SIR-C/X-SAR*, In: Geoscience and Remote Sensing Symposium, 1995. IGARSS '95. "Quantitative Remote Sensing for Science and Applications", International. Presented at the Geoscience and Remote Sensing Symposium, 1995. IGARSS '95. "Quantitative Remote Sensing for Science and Applications", International, pp. 987–989 vol.2.
- Dobson, M.C., Ulaby, F.T., LeToan, T., Beaudoin, A., Kasichke, E.S., Christensen, N., 1992. Dependence of radar backscatter on coniferous forest biomass. *IEEE Transactions on Geoscience and Remote Sensing*, 30, 412–415.
- Drake, J.B., Dubayah, R.O., Clark, D.B., Knox, R.G., Blair, J.B., Hofton, M.A., Chazdon, R.L., Weishampel, J.F., Prince, S., 2002. Estimation of tropical forest structural characteristics using large-footprint LiDAR. *Remote Sensing of Environment*, 79, 305–319.
- Drake, J.B., Dubayah, R.O., Knox, R.G., Clark, D.B., Blair, J.B., 2002. Sensitivity of large-footprint lidar to canopy structure and biomass in a neotropical rainforest. *Remote Sensing of Environment*, 81, 378–392.
- Dubayah, R., Blair, J.B., Bufton, J.L., Clark, D., Jaja, J., Knox, R., Luthcke, S.B., Prince, S., Weishampel, J., 1997. *The vegetation canopy LiDAR mission*, In: Proceedings of Land Satellite Information in the

- Next Decade, II: Sources and Applications. Bethesda (MD): American Society of Photogrammetry and Remote Sensing, pp. 100–112.
- Duong, H., Lindenbergh, R., Pfeifer, N., Vosselman, G., 2009. ICESat full-waveform altimetry compared to airborne LASER scanning altimetry over the Netherlands. *IEEE Transactions on Geoscience and Remote Sensing*, 47, 3365–3378.
- Duong, V.H., 2010. *Processing and Application of ICESat Large Footprint Full Waveform Laser Range Data*. Technische Universiteit, Delft, 213
- Eiben, A.E., Smith, J.E., 2003. *Introduction to evolutionary computing*, 1st Ed, XVI. Springer Verlag, New York, USA.
- Evans, D.L., Roberts, S.D., Parker, R.C., 2006. LiDAR A new tool for forest measurements? *The Forestry Chronicle*, 82, 211–218.
- FAO, H., 2006. Global forest resources assessment 2005. *FAO Forestry Paper*, 147.
- Fricker, H., Bassis, J., Minster, B., MacAyeal, D., 2005. ICESat's new perspective on ice shelf rifts: The vertical dimension. *Geophysical Research Letters*, 32, L23S08.
- Gibbs, H.K., Brown, S., Niles, J.O., Foley, J.A., 2007. Monitoring and estimating tropical forest carbon stocks: making REDD a reality. *Environmental Research Letters*, 2, 13.
- Goward, S.N., Markham, B., Dye, D.G., Dulaney, W., Yang, J., 1991. Normalized difference vegetation index measurements from the Advanced Very High Resolution Radiometer. *Remote sensing of environment* 35, 257–277.
- Grefenstette, J.J., 1986. Optimization of control parameters for genetic algorithms. *Systems, IEEE Transactions on Man and Cybernetics*, 16, 122–128.
- Grönroos, M.A., 1998. *Evolutionary design of neural networks*. University of Turku, Finland, 90
- Gupta, M.M., Jin, L., Homma, N., 2003. Static and dynamic neural networks: from fundamentals to advanced theory. *Wiley-IEEE Press*, USA.
- Harding, D.J., Carabajal, C.C., 2005. ICESat waveform measurements of within-footprint topographic relief and vegetation vertical structure. *Geophysical Research Letters* 32, L21S10.
- Hecht-Nielsen, R., 1989. *Theory of the backpropagation neural network*, In: Proceedings of International Joint Conference on Neural Networks (IJCNN)'89. IEEE, Washington D.C., p. I-593– I-605.
- Hese, S., Lucht, W., Schmullius, C., Barnsley, M., Dubayah, R., Knorr, D., Neumann, K., Riedel, T., Schroter, K., 2005. Global biomass mapping for an improved understanding of the CO₂ balance—the Earth observation mission Carbon-3D. *Remote Sensing of Environment* 94, 94–104.
- Houghton, R.A., 2005. Aboveground Forest Biomass and the Global Carbon Balance. *Global Change Biology* 11, 945–958.
- Hudak, A.T., Lefsky, M.A., Cohen, W.B., Berterretche, M., 2002. Integration of LiDAR and Landsat ETM+ data for estimating and mapping forest canopy height. *Remote Sensing of Environment* 82, 397–416.
- Huete, A., Justice, C., Liu, H., 1994. Development of vegetation and soil indices for MODIS-EOS. *Remote Sensing of Environment* 49, 224–234.
- Imhoff, M.L., 1995. Radar backscatter and biomass saturation: ramifications for global biomass inventory. *IEEE Transactions on Geoscience and Remote Sensing*, 33, 511–518.
- Jakubauskas, M.E., 1996. Thematic Mapper characterization of lodgepole pine seral stages in Yellowstone National Park, USA. *Remote Sensing of Environment* 56, 118–132.
- Jenkins, J., Chojnacky, D., Heath, L., Birdsey, R., 2003. National-scale Biomass Estimators for United States Tree Species. *Forest Science* 49, 12–35.
- Jutzi, B., Stilla, U., 2006. Range determination with waveform recording laser systems using a Wiener Filter. *ISPRS Journal of Photogrammetry and Remote Sensing* 61, 95–107.
- Kale, M., Singh, S., Roy, P.S., Ghole, V.S., 2004. Biomass equations of dominant species of dry deciduous forest in Shivpuri district, Madhya Pradesh. *Current Sciences* 87, 683–687.
- Keller, M., Palace, M., Hurtt, G., 2001. Biomass estimation in the Tapajos National Forest, Brazil: Examination of sampling and allometric uncertainties. *Forest Ecology and Management* 154, 371–382.
- Kimes, D., Nelson, R., Manry, M., Fung, A., 1998. Review article: Attributes of neural networks for extracting continuous vegetation variables from optical and radar measurements. *International Journal of Remote Sensing* 19, 2639–2663.

- Kimes, D.S., Holben, B.N., Nickeson, J.E., McKee, W.A., 1996. Extracting forest age in a Pacific Northwest forest from Thematic Mapper and topographic data. *Remote Sensing of Environment* 56, 133–140.
- Kohavi, R., 1995. *A study of cross-validation and bootstrap for accuracy estimation and model selection*, In: International Joint Conference on Artificial Intelligence. pp. 1137–1145.
- Krogh, A., Vedelsby, J., 1995. Neural network ensembles, cross validation, and active learning. *Advances in neural information processing systems* 231–238.
- Kwak, D.A., Lee, W.K., Son, M.H., 2005. *Application of LiDAR and Digital Aerial Photograph for precise forest inventory*. Presented at the Proceedings of 2005 ESRI International User Conference, ESRI, San Diego, CA, USA.
- Lefsky, M.A., Cohen, W.B., Acker, S.A., Parker, G.G., Spies, T.A., Harding, D., 1999a. LiDAR remote sensing of the canopy structure and biophysical properties of Douglas-Fir Western Hemlock Forests. *Remote sensing of environment* 70, 339–361.
- Lefsky, M.A., Cohen, W.B., Parker, G.G., Harding, D.J., 2002. LiDAR Remote Sensing for Ecosystem Studies. *BioScience* 52, 19–30.
- Lefsky, M.A., Harding, D.J., Keller, M., Cohen, W.B., Carabajal, C.C., Del Bom Espirito-Santo, F., Hunter, M.O., de Oliveira, R., 2005. Estimates of forest canopy height and aboveground biomass using ICESat. *Geophysical Research Letters*. 32, L22S02.
- Lefsky, M.A., Keller, M., Pang, Y., De Camargo, P.B., Hunter, M.O., 2007. Revised method for forest canopy height estimation from Geoscience Laser Altimeter System waveforms. *Journal of Applied Remote Sensing* 1, 013537.
- Means, J.E., Acker, S.A., Harding, D.J., Blair, J.B., Lefsky, M.A., Cohen, W.B., Harmon, M.E., McKee, W.A., 1999. Use of large-footprint scanning airborne LiDAR to estimate forest stand characteristics in the Western Cascades of Oregon. *Remote Sensing of Environment* 67, 298–308.
- Moller, M.F., 1993. A scaled conjugate gradient algorithm for fast supervised learning. *Neural networks* 6, 525–533.
- Moody, J., 1994. Prediction risk and architecture selection for neural networks. In J.F.V. Cherkassky and H. Wechsler (Editors), *From Statistics to Neural Networks: Theory and Pattern Recognition Applications*. NATO ASI Series F Computer and System Sciences. Springer-Verlag 136, 147–147.
- Mozer, M.C., Smolensky, P., 1989. Using relevance to reduce network size automatically. *Connection Science* 1, 3–16.
- Neuenschwander, A.L., 2008. Evaluation of waveform deconvolution and decomposition retrieval algorithms for ICESat/GLAS data. *Canadian Journal of Remote Sensing* 34, 240–246.
- Neumann, M., Saatchi, S.S., Ulander, L.M.H., Fransson, J.E.S., 2011. Parametric and non-parametric forest biomass estimation from PolInSAR data. *IEEE*, pp. 420–423.
- NSIDC [WWW Document], 2011a. . GLAS Altimetry Product Usages Guidance. URL http://nsidc.org/data/docs/daac/glas_altimetry/pdf/NSIDC_AltUserGuide_Rel29.pdf
- NSIDC [WWW Document], 2011b. . ICESat / GLAS Data Subsetter. URL http://nsidc.org/forms/glas_subset_form.html
- NSIDC [WWW Document], 2012. . NSIDC. URL http://nsidc.org/data/icesat/laser_op_periods.html
- Pang, Y., Lefsky, M., Sun, G., Miller, M., Li, Z., 2008. Temperate forest height estimation performance using ICESat GLAS data from different observation periods. The International Archives of the Photogrammetry, *Remote Sensing and Spatial Information Science* 37, 777–782.
- Patenaude, G., Milne, R., Dawson, T.P., 2005. Synthesis of remote sensing approaches for forest carbon estimation: reporting to the Kyoto Protocol. *Environmental Science & Policy* 8, 161–178.
- Pierce, L., Sarabandi, K., Ulaby, F., 1994. Application of an artificial neural network in canopy scattering inversion. *International Journal of Remote Sensing* 15, 3263–3270.
- Pinty, B., Verstraete, M.M., Dickinson, R.E., 1990. A Physical Model of the Bidirectional Reflectance of Vegetation Canopies. *Journal of Geophysical Research* 95, 11–767.
- Ranson, K.J., Guoqing Sun, 1994. Mapping biomass of a northern forest using multifrequency SAR data. *IEEE Transactions on Geoscience and Remote Sensing*, DOI - 10.1109/36.295053 32, 388–396.
- Reitberger, J., Krzystek, P., Stilla, U., 2008. Analysis of full waveform LIDAR data for the classification of deciduous and coniferous trees. *International journal of remote sensing* 29, 1407–1431.
- Roebber, P.J., Bruening, S.L., Schultz, D.M., Cortinas Jr, J.V., 2003. Improving snowfall forecasting by diagnosing snow density. *Weather and forecasting* 18, 264–287.

- Rosenqvist, A., Milne, A., Lucas, R., Imhoff, M., Dobson, C., 2003. A review of remote sensing technology in support of the Kyoto Protocol. *Environmental Science & Policy* 6, 441–455.
- Rosette, J., North, P., Suárez, J., 2008. Vegetation height estimates for a mixed temperate forest using satellite laser altimetry. *International Journal of Remote Sensing* 29, 1475–1493.
- Rouse, J., Haas, R.H., Schell, J.A., Dering, D.W., Harlan, J.C., 1974. Monitoring the vernal advancement and retrogradation of natural vegetation (No. Type III), Final Report. *NASA/GSFC, Greenbelt, Maryland, USA*.
- Roy, P., Ravan, S., 1996. Biomass estimation using satellite remote sensing data—An investigation on possible approaches for natural forest. *Journal of Biosciences* 21, 535–561–561.
- Rumelhart, D.E., Hintont, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. *Nature* 323, 533–536.
- Saatchi, S.S., Harris, N.L., Brown, S., Lefsky, M., Mitchard, E.T.A., Salas, W., Zutta, B.R., Buermann, W., Lewis, S.L., Hagen, S., Petrova, S., White, L., Silman, M., Morel, A., 2011. *Benchmark map of forest carbon stocks in tropical regions across three continents*. In: Proceedings of the National Academy of Sciences 108, 9899.
- Sarle, W.S., 1995. *Stopped training and other remedies for overfitting*, In: Proceedings of the 27th Symposium on the Interface.
- Schaffer, J.D., Morishima, A., 1987. *An adaptive crossover distribution mechanism for genetic algorithms*, In: Proceedings of the Second International Conference on Genetic Algorithms on Genetic Algorithms and Their Application. pp. 36–40.
- Singh, K.P., Misra, R., 1975. Structure and functioning of natural, modified and silvicultural ecosystems of eastern Uttar Pradesh (Final Technical Report), MAB Research Project. *Banaras Hindi University, Varanasi, UP, India*.
- Sun, G., Ranson, K., Masek, J., Fu, A., Wang, D., 2007. *Predicting tree height and biomass from GLAS data*, In: 10th International Symposium on Physical Measurements and Signatures in Remote Sensing. Davos, Switzerland.
- Sun, G., Ranson, K.J., Kimes, D.S., Blair, J.B., Kovacs, K., 2008. Forest vertical structure from GLAS: An evaluation using LVIS and SRTM data. *Remote Sensing of Environment* 112, 107–117.
- Taiz, L., Zeiger, E., 2011. Plant Physiology Online Fifth Edition [WWW Document]. URL <http://5e.plantphys.net/>
- Tettamanzi, A., Tomassini, M., 2001. Soft computing: integrating evolutionary, neural, and fuzzy systems. *Springer Verlag*, Berlin, Germany.
- Viergever, K.M., Woodhouse, I.H., Marino, A., Brolley, M., Stuart, N., 2008. *SAR Interferometry For Estimating Above-Ground Biomass Of Savanna Woodlands In Belize*, In: Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International. Presented at the Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International, p. V – 290–V – 293.
- Wagner, W., Ullrich, A., Ducic, V., Melzer, T., Studnicka, N., 2006. Gaussian decomposition and calibration of a novel small-footprint full-waveform digitising airborne laser scanner. *ISPRS Journal of Photogrammetry and Remote Sensing* 60, 100–112.
- Wang, H., Ouchi, K., 2010. A Simple Moment Method of Forest Biomass Estimation From Non-Gaussian Texture Information by High-Resolution Polarimetric SAR. *IEEE Geoscience and Remote Sensing Letters* 7, 811–815.
- Waring, R.H., Way, J.B., Hunt, E.R., Morrissey, L., Ranson, K.J., Weishampel, J.F., Oren, R., Franklin, S.E., 1995. Imaging radar for ecosystem studies. *BioScience* 45, 715–723.
- Wasserman, P.D., 1989. Neural computing: theory and practice. *Van Nostrand Reinhold Co.*, New York.
- Werbos, P., 1974. *Beyond Regression: New Tools for Prediction and Analysis in the Behavioural Sciences*. Harvard University, MA, USA
- Werbos, P.J., 1994. The roots of backpropagation: from ordered derivatives to neural networks and political forecasting. *Wiley-Interscience*, New York.
- Wijaya, A., Gloaguen, R., 2009. *Fusion of ALOS Palsar and Landsat ETM data for land cover classification and biomass modeling using non-linear methods*, In: Geoscience and Remote Sensing Symposium, 2009 IEEE International, IGARSS 2009. Presented at the Geoscience and Remote Sensing Symposium, 2009 IEEE International, IGARSS 2009, p. III–581–III–584.

- Wu, H., Xing, Y., 2010. *Wavelet Transform and its Application to ICESat-GLAS Full Waveform Data*, In: Intelligence Information Processing and Trusted Computing (IPTC), 2010 International Symposium On. pp. 373–377.
- Yao, X., 1991. *Evolution of connectionist networks*, In: Preprints of the Int'l Symp. on AI, Reasoning & Creativity. Queensland, Australia, pp. 49–52.
- Yao, X., 1999. *Evolving artificial neural networks*. Proceedings of the IEEE 87, 1423–1447.
- Yao, X., Liu, Y., 1997. A new evolutionary system for evolving artificial neural networks. *IEEE Transactions on Neural Networks*, 8, 694–713.
- Zhang, J., de Gier, A., Xing, Y., Sohn, G., 2009. Derivation of forest type information from large-footprint spaceborne LiDAR waveforms, In: *Sihvalaser*. Texas, USA, p. 12.
- Zhao, M., Zhou, G., 2005. Estimation of biomass and net primary productivity of major planted forests in China based on forest inventory data. *Forest Ecology and Management* 207, 295–313.
- Zwally, H., Schutz, B., Abdalati, W., Abshire, J., Bentley, C., Brenner, A., Bufton, J., Dezio, J., Hancock, D., Harding, D., others, 2002. ICESat's laser measurements of polar ice, atmosphere, ocean, and land. *Journal of Geodynamics* 34, 405–445.

APPENDICES

Appendix I: Field Data Collection Form

Form I-A

AGB Estimation Using Spaceborne Waveform LiDAR

FIELD DATA COLLECTION FORM, MNP, Shivpuri, MP

CATEGORY 1: SITE AND OBSERVER
State: _____ District: _____ Plot GPS Point Name = _____
Area / Location (any landmark): _____
Site Centre GPS co-ordinates: Lat: _____° _____' _____" Long: _____° _____' _____"
Sample Site- Plot No.: _____ / C/T/P1/P2 Date: _____/_____/_____ 2012 Time: _____
G.P.S. Reading (HHDD: MM:SS and WGS 84): Lat: _____° _____' _____" Long: _____° _____' _____"
Observer: _____ Altitude: Site Plot: _____ m / _____ m
Slope (°): _____
Aspect: N/E/S/W/NE/SE/SW/NW Topography (General observations): _____

CATEGORY 2: FOREST AND PLOT DESCRIPTION IN GENERAL:
Forest Type: _____
Visual Evidence of Disturbance: Lopping/ Cutting/ Fire/ Grazing _____

CATEGORY 3: QUANTITATIVE MEASUREMENTS
Crown Density (%) (Make small hole in leaf paper and count sky or crown cover hits and steps, count which is ever is less in count across both diagonals):
NE-SW: Steps: _____ Hits: _____ (Canopy/Sky)
NW-SE: Steps: _____ Hits: _____ (Canopy/Sky)

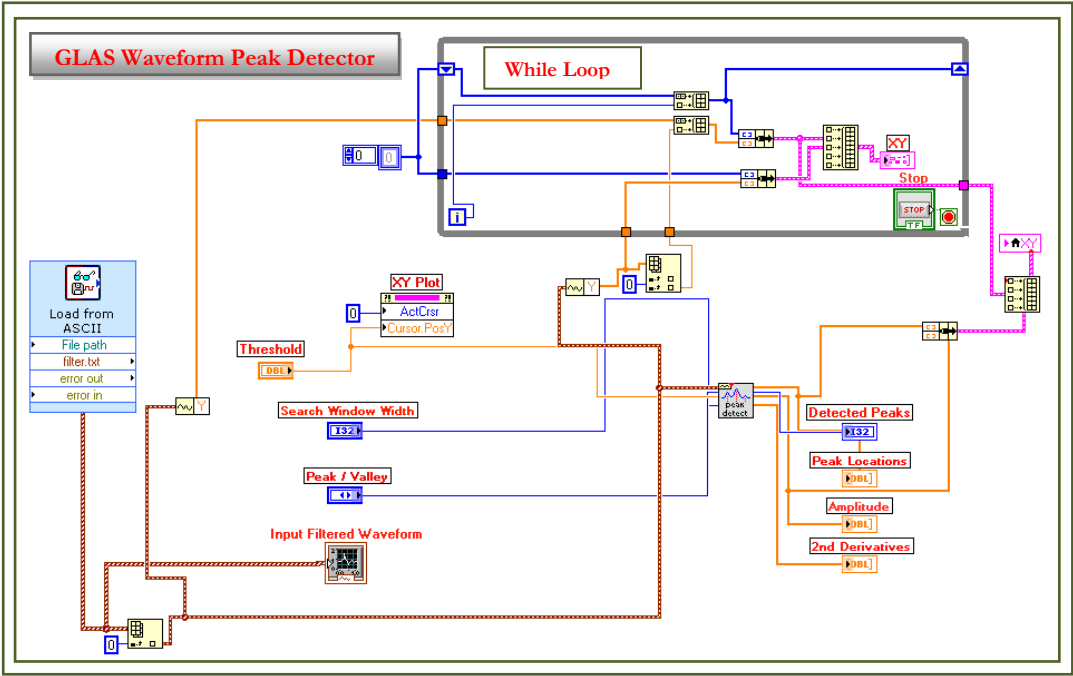
Size of Sample Plot for Trees (L2: 73.5 m & L3: 54 m) / For Footprints having greater variability: [4 sub plots 9(1 at centre / radius = 10 m and 3 at periphery / radius = 5 m)
(a) Pl. use steel measuring tape; (b) for DBH at 1.37m from ground make sticks of 1.37 m to fix the height, in plans from any direction but on slopes from upper side; (c) fix the metallic tag at the base of the tree preferably concealed and hide with stones) – Pl record DBH up to mm level.

Tree id.	Species	DBH (≥ 10 cm) all plants [cm]	Height (m) 1 st forking	Height (m) at ultimate forking
1				
2				
3				
4				
5				
6				
7				

8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				
25				
26				
27				
28				
29				
30				
31				
32				
33				
34				
35				
36				
37				
38				
39				
40				
41				
42				
43				
44				
45				
46				
47				
48				

Appendix II: Software

A. Waveform Peak Detection [LabView Virtual Instrument Outline]



B. DTREG [*EANN Application*]

Model

PNN/GRNN | RBF Network | GMDH | Cascade Correlation | Discriminant Analysis | K-Means Clustering | Linear Regression | Logistic Regression | Factor Analysis | Class labels | Initial split | Category weights | Misclassification | Missing data | Variable weights | DTL | Scoring | Translate | Misc. | Design | Encryption | Data | Variables | Validation | Time series | Decision Tree | TreeBoost | Decision Tree Forest | SVM | GEP | Multilayer Perceptron

General | Functions | Evolution | Linking | Constants

--- Gene Expression Programming (GEP) and Symbolic Regression ---

Type of model to build: **Gene Expression Programming**

Model building parameters:

- Population size: 50
- Max. tries for initial pop.: 10000
- Genes per chromosome: 4
- Gene head length: 8
- Maximum generations: 2000
- Gen. without improvement: 1000
- Stop if fitness reaches: 1.000
- Max. minutes for training:

Fitness function: **Root relative squared error (RRSE)**

Precision (hit tolerance): 0.01

Selection range: 100

Expression simplification:

- ☒ Do algebraic simplification
- ☐ Parsimony pressure: 1.0002
- ☒ Try to simplify after training
- Simplification generations: 500
- Gen. without improvement: 200
- Max. minutes simplifying:

Model testing and validation:

- ☐ No validation, use all data rows
- ☐ Use variable to select validation rows
- ☐ Random percent: 70
- ☒ V-fold cross-validation: 10
- ☐ Leave-one-out validation

How to handle missing predictor variable values:

- ☐ Don't use rows with missing predictors
- ☒ Replace missing predictors with medians
- ☐ Use surrogate variables

Options:

- ☐ Compute importance of variables

Expression simplifier

Model

--- Multilayer Perceptron Neural Networks (MLP) ---

Type of model to build: **Multilayer Perceptron**

Number of network layers: **3 layers (1 hidden)**

Automatic hidden layer neuron selection:

- ☒ Automatically optimize hidden layer 1
- Min: 2 | Max: 20 | Step: 1
- Max. steps without change: 9
- % rows to use for search: 100
- Cross validate, folds: 4
- Hold-out sample %: 70
- ☐ Use training data

Number of neurons for hidden layers:

Layer 1: | Layer 2: 1

Overfitting detection & prevention:

- ☒ Use test data to detect overfitting
- % training rows to hold out: 20
- Max. steps without change: 10

Model testing and validation:

- ☐ No validation, use all data rows
- ☐ Use variable to select validation rows
- ☐ Random percent: 70
- ☒ V-fold cross-validation: 10
- ☐ Leave-one-out validation

How to handle missing values:

- ☐ Don't use rows with missing values
- ☒ Replace missing values with medians
- ☐ Use surrogate variables

Options:

- ☐ Compute importance of variables

Hidden layer activation function: **Logistic**

Output layer activation function: **Linear**

Conjugate gradient parameters:

- Num. convergence tries: 4
- Maximum iterations: 10000
- Iterations without improvement: 100
- Convergence tolerance: 1.000e-005
- Min. improvement delta: 1.000e-006
- Min. gradient: 1.000e-006
- Max. minutes execution time: 0

Training method:

- ☒ Scaled conjugate gradient (recommended)
- ☐ Traditional conjugate gradient

☐ Write progress report to project log

OK Cancel Apply

Appendix III: Field Photographs



