



**HAL**  
open science

# Investigating the possibility of forest height/volume estimation using lidar, radar and optical images : case study : Nowshahr Forests in Mazindaran, Iran

Manizheh Rajab Pourrahmati

## ► To cite this version:

Manizheh Rajab Pourrahmati. Investigating the possibility of forest height/volume estimation using lidar, radar and optical images : case study : Nowshahr Forests in Mazindaran, Iran. Electronics. Université Montpellier; University of Teheran, 2016. English. NNT : 2016MONTT288 . tel-01807960

**HAL Id: tel-01807960**

**<https://theses.hal.science/tel-01807960>**

Submitted on 5 Jun 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



University of Montpellier



## Speciality

**Automatic and Microelectronic Systems**  
(Doctoral School: Information, Structures, Systemes)

# **Investigating the possibility of forest height/volume estimation using lidar, radar and optical images**

(Case study: Nowshahr Forests in Mazindaran, Iran)

**By:**

**Manizheh Rajab Pourrahmati**

**Thesis submitted to fulfill the requirements for the degree of PhD**

December 2016

Defence on 19/12/2016, List of jury members:

**Pierre Couteron**, DR, IRD, Cameroun  
Examiner, President of jury

**Clément Mallet**, Engineer IGN, HDR, LaSTIG  
Examiner

**Olivier Strauss**, HDR, MCF, University of  
Montpellier, Examiner

**Nicolas Baghdadi**, DR, Irstea  
Supervisor

**Valéry Gond**, HDR, Cirad  
Advisor

**Mehrez Zribi**, DR, CNRS, Toulouse  
Reviewer

**Pierre-Louis Frison**, HDR, MCF, University of  
Paris-Est Marne-la-Vallée, Reviewer

**Jean Stéphane Bailly**, HDR, AgroParisTech  
Invited member

**Ali A. Darvishsefat**, Prof., University of Tehran  
Co- Supervisor





L'UNIVERSITE DE MONTPELLIER



**Spécialité**

**Systèmes Automatiques et Microélectroniques (SyAM)**

**(l'Ecole Doctorale: Information, Structures, Systemes)**

**Estimation de la hauteur et du volume de la forêt à l'aide du lidar,  
radar et des données optiques  
(Étude de cas: forêts de Nowshahr en Mazindaran, Iran)**

**Par:**

**Manizheh Rajab Pourrahmati**

**Thèse présentée pour satisfaire aux exigences pour le degré de doctorat**

Decembre 2016

Defence on 19/12/2016, Le jury composé de:

**Pierre Couteron**, DR, IRD, Cameroun  
Rapporteur, président du jury

**Clément Mallet**, Ingénieur IGN, HDR, LaSTIG  
Examineur

**Olivier Strauss**, HDR, MCF, Université de  
Montpellier, Examineur

**Nicolas Baghdadi**, DR, Irstea  
Directeur de la thèse

**Valéry Gond**, HDR, Cirad  
co-encadrant de thèse

**Mehrez Zribi**, DR, CNRS, Toulouse  
Rapporteur

**Pierre-Louis Frison**, HDR, MCF, Université  
Paris-Est Marne-la-Vallée, Rapporteur

**Jean Stéphane Bailly**, HDR, AgroParisTech  
Invité

**Ali A. Darvishsefat**, Prof., Université de Tehran  
Co-directeur de la thèse



## Abstract

The importance of measuring forest biophysical parameters for ecosystem health monitoring and forest management encourages researchers to find precise, yet low-cost methods especially in mountainous and large areas. In the present study Geoscience Laser Altimeter System (GLAS) on board ICESat (Ice Cloud and land Elevation Satellite) was used to estimate three biophysical characteristics of forests located in the north of Iran: 1) maximum canopy height ( $H_{\max}$ ), 2) Lorey's height ( $H_{\text{Lorey}}$ ), and 3) Forest volume ( $V$ ). A large number of Multiple Linear Regressions (MLR), Random Forest (RF) and also Artificial Neural Network regressions were developed using two different sets of variables including waveform metrics and Principal Components (PCs) produced from Principal Component Analysis (PCA). To validate and compare models, statistical criteria were calculated based on a five-fold cross validation. Best model concerning the maximum height was an MLR (RMSE=5.0m) which combined two metrics extracted from waveforms (waveform extent " $W_{\text{ext}}$ " and height at 50% of waveform energy " $H_{50}$ "), and one from Digital Elevation Model (Terrain Index: TI). The mean absolute percentage error (MAPE) of maximum height estimates was 16.4%. For Lorey's height, an ANN model using PCs and waveform extent " $W_{\text{ext}}$ " outperformed other models (RMSE=3.4m, MAPE=12.3%). In order to estimate forest volume, two approaches was employed: First, estimating volume using volume-height relationship while height is GLAS estimated height; Second, estimation of forest volume directly from GLAS data by developing regressions between in situ volume and GLAS metrics. The result from first approach (116.3 m<sup>3</sup>/ha) was slightly better than the result obtained by the second approach that is a PCs-based ANN model (119.9 m<sup>3</sup>/ha). But the ANN model performed better in very low (<10 m<sup>3</sup>/ha) and very high (> 800 m<sup>3</sup>/ha) volume stands. In total, the relative error of estimated forest volume was about 26%. Generally, MLR and ANN models had better performance when compared to the RF models. In addition, the accuracy of height estimations using waveform metrics was better than those based on PCs.

Given the suitable results of GLAS height models (maximum and Lorey's heights), production of wall to wall height maps from synergy of remote sensing (GLAS, PALSAR, SPOT5 and Landsat-TM) and environmental data (slope, aspect, classified elevation map and also geological map) was taken under consideration. Thus, MLR and RF regressions

were built between all GLAS derived heights, inside of the study area, and indices extracted from mentioned remotely sensed and environmental data. The best resulted models for  $H_{\max}$  (RMSE=7.4m and  $R_a^2=0.52$ ) and  $H_{\text{Lorey}}$  (RMSE=5.5m and  $R_a^2=0.59$ ) were used to produce a wall to wall maximum canopy height and Lorey' height maps. Comparison of  $H_{\max}$  extracted from the resulted  $H_{\max}$  map with true height values at the location of 32 in situ plots produced an RMSE and  $R^2$  of 5.3m and 0.71, respectively. Such a comparison for  $H_{\text{Lorey}}$  led to an RMSE and  $R^2$  of 4.3m and 0.50, respectively. Regression-kriging method was also used to produce canopy height map with considering spatial correlation between canopy heights. This approach, with the aim of improving the precision of canopy height map provided from non-spatial method, was unsuccessful which could be due to the heterogeneity of the study area in case of forest structure and topography.

**Key words:** Lidar, ICESat GLAS, Alos PALSAR, Optical images, Maximum canopy height, Lorey's height, Forest volume, Iran

## Résumé

L'importance de mesurer les paramètres biophysiques de la forêt pour la surveillance de la santé des écosystèmes et la gestion forestière encourage les chercheurs à trouver des méthodes précises et à faible coût en particulier sur les zones étendues et montagneuses. Dans la présente étude, Le lidar satellitaire GLAS (Geoscience Laser Altimeter System) embarqué à bord du satellite ICESat (Ice Cloud and land Elevation Satellite) a été utilisé pour estimer trois caractéristiques biophysiques des forêts situées dans le nord de l'Iran: 1) hauteur maximale de la canopée ( $H_{\max}$ ), 2) hauteur de Lorey ( $H_{\text{Lorey}}$ ), et 3) le volume du bois ( $V$ ). Des régressions linéaires multiples (RLM), des modèles basés sur les Forêts Aléatoires (FA : Random Forest) et aussi des réseaux de neurones (ANN) ont été développés à l'aide de deux ensembles différents de variables incluant des métriques obtenues à partir des formes d'onde GLAS et des composantes principales (CP) produites à partir de l'analyse en composantes principales (ACP) des données GLAS. Pour valider et comparer les modèles, des critères statistiques ont été calculées sur la base d'une validation croisée. Le meilleur modèle pour l'estimation de la hauteur maximale a été obtenu avec une régression RLM (RMSE = 5.0 m) qui combine deux métriques extraites des formes d'onde GLAS (étendue et hauteur pour une énergie à 50%, respectivement  $W_{\text{ext}}$  et  $H_{50}$ ), et un paramètre issu du modèle numérique d'élévation (Indice de relief TI). L'erreur moyenne absolue en pourcentage (MAPE) sur les estimations de la hauteur maximale est de 16.4%. Pour la hauteur de Lorey, un modèle basé sur les réseaux de neurones et utilisant des CPs et le  $W_{\text{ext}}$  fournit le meilleur résultat avec RMSE = 3.4 m et MAPE = 12.3%. Afin d'estimer le volume du bois, deux approches ont été utilisées: (1) estimation du volume à l'aide d'une relation volume-hauteur avec une hauteur estimée à partir de données GLAS et (2) estimation du volume du bois directement à partir des données GLAS en développant des régressions entre le volume in situ et les métriques GLAS. Le résultat de la première approche (RMSE=116.3 m<sup>3</sup>/ha) était légèrement meilleur que ceux obtenus avec la seconde approche. Par exemple, le réseau de neurones basé sur les PCs donnait un RMSE de 119.9 m<sup>3</sup>/ha mais avec des meilleurs résultats que l'approche basée sur la relation volume-hauteur pour les faibles (<10 m<sup>3</sup>/ha) et les forts (> 800 m<sup>3</sup>/ha) volumes. Au total, l'erreur relative sur le volume de bois est estimée à environ 26%. En général, les modèles RLM et ANN avaient des meilleures performances par rapport aux modèles de FA. En outre, la précision sur l'estimation de la hauteur à l'aide de métriques issues des formes d'onde GLAS est meilleure que celles basées sur les CPs.

Compte tenu des bons résultats obtenus avec les modèles de hauteur GLAS (hauteurs maximale et de Lorey), la production de la carte des hauteurs d'étude par une utilisation combinée de données de télédétection lidar, radar et optique (GLAS, PALSAR, SPOT-5 et Landsat-TM) et de données environnementales (pente, aspect, et altitude du terrain ainsi que la carte géologique) a été effectuée à l'intérieur de notre zone. Ainsi, des régressions RLM et FA ont été construites entre toutes les hauteurs dérivées des données GLAS, à l'intérieur de la zone d'étude, et les indices extraits des données de télédétection et des paramètres environnementaux. Les meilleurs modèles entraînés pour estimer  $H_{\max}$  (RMSE = 7.4 m et  $R_a^2=0.52$ ) et  $H_{\text{Lorey}}$  (RMSE = 5.5 m et  $R_a^2=0.59$ ) ont été utilisées pour produire les cartes de hauteurs. La comparaison des  $H_{\max}$  de la carte obtenue avec les valeurs de  $H_{\max}$  in situ à l'endroit de 32 parcelles produit un RMSE de 5.3 m et un  $R^2$  de 0.71. Une telle comparaison pour  $H_{\text{Lorey}}$  conduit à un RMSE de 4.3m et un  $R^2$  de 0.50. Une méthode de régression-krigeage a également été utilisée pour produire une carte des hauteurs en considérant la corrélation spatiale entre les hauteurs. Cette approche, testée dans le but d'améliorer la précision de la carte de la hauteur du couvert fournie par la méthode non-spatiale, a échoué due à l'hétérogénéité de la zone d'étude en termes de la structure forestière et de la topographie.

**Mots Clés:** Lidar, ICESat GLAS, ALOS PALSAR, images optiques, hauteur maximale de la canopée, hauteur de Lorey, Volume de bois, Iran.

## Table of contents

### Chapter 1:

1. Introduction .....	1
-----------------------	---

### Chapter 2:

2. Materials and data processing .....	11
2.1. Study area .....	11
2.2. Description of data .....	12
2.2.1. Field measurements .....	12
2.2.2. Digital elevation model .....	16
2.2.3. Geological map .....	16
2.2.4. GLAS/ICESat .....	17
2.2.5. ALOS/PALSAR .....	21
2.2.6. Passive optical remote sensing data .....	23
2.3. Data processing and information extraction .....	25
2.3.1. Analysis of in situ data .....	25
2.3.2. Extraction of terrain index, slope, aspect and elevation class maps from DEM ....	32
2.3.3. ICESat GLAS data processing and extraction of metrics .....	34
2.3.3.1. Processing GLAS waveforms .....	34
2.3.3.2. Waveform metrics extraction .....	35
2.3.4. PALSAR data processing and extraction of metrics .....	40
2.3.5. Optical data (Landsat-TM, SPOT5) and extraction of vegetation and texture indices .....	45

### Chapter 3:

3. Research methodology .....	47
3.1. Estimation of maximum ( $H_{max}$ ) and Lorey's heights ( $H_{Lorey}$ ) using ICESat GLAS .....	48
3.1.1. Direct method for estimating $H_{max}$ .....	48
3.1.2. Parametric and non-parametric methods for prediction $H_{max}$ and $H_{Lorey}$ .....	49
3.2. Forest volume estimation using ICESat GLAS .....	53
3.3. Model Validation .....	54

3.4. Production of canopy height map and its validation .....	55
3.4.1. Canopy height map using regression model .....	56
3.4.2. Canopy height map using regression-kriging method .....	56

## **Chapter 4:**

4. Results and discussion .....	59
4.1. Estimation of maximum canopy height ( $H_{max}$ ) using ICESat GLAS .....	59
4.1.1. Direct method for estimation of $H_{max}$ .....	59
4.1.2. Estimation of $H_{max}$ using regression models (MLR, RF and ANN) .....	59
4.1.2.1. Estimation of $H_{max}$ using MLR .....	59
4.1.2.2. Estimation of $H_{max}$ using RF .....	64
4.1.2.3. Estimation of $H_{max}$ using ANN .....	67
4.1.2.4. Discussion on $H_{max}$ estimated using GLAS data .....	69
4.2. Estimation of Mean Lorey's height ( $H_{Lorey}$ ) using ICESat GLAS .....	71
4.2.1. Estimation of $H_{Lorey}$ using MLR .....	71
4.2.2. Estimation of $H_{Lorey}$ using RF .....	73
4.2.3. Estimation of $H_{Lorey}$ using ANN .....	76
4.2.4. Discussion on $H_{Lorey}$ estimated using GLAS data .....	78
4.3. Estimation of forest volume using ICESat GLAS .....	79
4.3.1. Discussion on volume estimated using GLAS data .....	82
4.4. Production of canopy height map .....	85
4.4.1. Canopy height map using regression model .....	85
4.4.2. Canopy height map using regression-kriging .....	91
4.4.3. Discussion on production of canopy height map .....	95

## **Chapter 5:**

5. Conclusion and perspectives .....	97
5.1. Conclusion .....	97
5.1.1. Prediction of maximum height ( $H_{max}$ ) using GLAS .....	97
5.1.2. Prediction of Lorey's height ( $H_{Lorey}$ ) using GLAS .....	98
5.1.3. Effect of terrain index on estimation of height using GLAS .....	99
5.1.4. Prediction of forest volume ( $V$ ) using GLAS .....	99
5.1.5. Uncertainties in prediction of height and volume using GLAS .....	100
5.1.6. Production of Canopy height map .....	102

5.1.7. General conclusion.....	104
5.2. Perspectives .....	106
<b>Chapter 6:</b>	
6. References .....	109

## List of figures

Fig. 1.1. Major carbon pools and fluxes of the global carbon balance (FAO, 2007) .....	1
Fig. 2.1. Location of study area in Iran (top right map) and over Landsat image .....	11
Fig. 2.2. Schematic illustration of location of trees for height measurement in a field plot .....	13
Fig. 2.3. a) Schematic illustration of height measurement based on trigonometry; b) Calculation of horizontal distance from tree over sloped terrain.....	13
Fig. 2.4. illustration of three phases of field work including navigating and locating the plot center, DBH and height measurement.....	15
Fig. 2.5. Schematic of the relationship between geoidal heights and ellipsoidal heights .....	16
Fig. 2.6. Geological map provided by geological survey and mineral exploitation of Iran .....	17
Fig. 2.7. Schematic illustration of GLAS instrument operating from ICESat while orbiting the Earth .....	18
Fig. 2.8. Color composite of SPOT-5 images over study area .....	25
Fig. 2.9. Tree's Height vs. DBH and the best fitted line .....	29
Fig. 2.10. a) Distribution of number of trees per hectare (n/ha), b) forest volume (m <sup>3</sup> /ha) in 60 reference plots .....	31
Fig. 2.11. Histogram of forest volume (m <sup>3</sup> /ha) for 60 reference plots .....	31
Fig. 2.12. Digital elevation model (DEM <sub>10</sub> ) and four extracted maps .....	33
Fig. 2.13. Schematic of laser altimeter pulse spreading over a vegetated area, and a returned waveform .....	37
Fig. 2.14. GLAS waveform and some metrics over a terrain of 25% slope in the study area. 1ns corresponds to 15cm sampling distance in waveform .....	37
Fig. 2.15. a) A schematic illustration of lidar derived vegetation height (H), H <sub>100</sub> , and waveform extent (W <sub>ext</sub> ) for one waveform over flat terrain (solid line) and another over sloped terrain (dash line) (Lee <i>et al.</i> , 2011). b) Impact of slope on lidar height retrieval (Lee <i>et al.</i> , 2011). c,d) A schematic of footprint located over flat (left) and sloped (right) vegetated terrain, respectively .....	39
Fig. 2.16. Information explained by ten first PCs and other PCs .....	40
Fig. 2.17. Gamma naught ( $\gamma^\circ = \gamma_{dB}$ ) before (a) and after (b) LEE and multitemporal filters, and the histogram of backscatters over a forested area (green rectangle) (c) before (blue) and after (red) noise reduction. The standard deviation of $\gamma^\circ$ decreases from 1.42 to 0.44 db .....	42
Fig. 2.18. Eight GLCM texture features extracted from PALSAR-HH .....	44
Fig. 2.19. NDVI and eight GLCM texture features extracted from it .....	46
Fig. 3.1. Overview of forest canopy height and volume estimation using GLAS data .....	47

Fig. 3.2. Overview of forest canopy height map using combination of GLAS, PALSAR, optical images and environmental data .....	48
Fig. 3.3. A schematic of random forest regression .....	50
Fig. 3.4. A schematic of MLP neural network .....	51
Fig. 3.5. Schematic diagram of a 5-fold cross validation .....	55
Fig. 3.6. Illustration of nugget, range and sill components of a semivariogram .....	58
Fig. 4.1. Estimated $H_{\max}$ from GLAS data using direct method versus in situ $H_{\max}$ .....	59
Fig. 4.2. Estimated $H_{\max}$ using MLR based on waveform metrics (model 1, table 4.1) versus in situ $H_{\max}$ .....	61
Fig. 4.3. Estimated $H_{\max}$ using MLR based on waveform metrics (model 2, table 4.1) versus in situ $H_{\max}$ .....	61
Fig. 4.4. Estimated $H_{\max}$ using a) model 4.1, b) model 4.2 versus in situ $H_{\max}$ .....	62
Fig. 4.5. Estimated $H_{\max}$ using stepwise regression of waveform metrics (Eq. 4.3) versus in situ $H_{\max}$ .....	63
Fig. 4.6. Estimated $H_{\max}$ using MLR based on PCs (model 3, table 3.5) versus in situ $H_{\max}$ .....	64
Fig. 4.7. Estimated $H_{\max}$ using RF regression based on waveform metrics (model 1, table 4.3) versus in situ $H_{\max}$ .....	65
Fig. 4.8. Estimated $H_{\max}$ using RF regression based on PCs (model 3, table 4.4) versus in situ $H_{\max}$ .....	66
Fig. 4.9. Estimated $H_{\max}$ using ANN based on waveform metrics (model 1, table 4.5) versus in situ $H_{\max}$ .....	67
Fig. 4.10. Estimated $H_{\max}$ using ANN based on waveform metrics (model 2, table 4.5) versus in situ $H_{\max}$ .....	68
Fig. 4.11. Estimated $H_{\max}$ using ANN based on PCs (model 3, table 4.6) versus in situ $H_{\max}$ .....	69
Fig. 4.12. Comparison of residual errors produced by MLR, RF and ANN for estimation of $H_{\max}$ based on waveform metrics .....	70
Fig. 4.13. Estimated $H_{\text{Lorey}}$ using MLR based on waveform metrics (model 1, Table 4.7) versus in situ $H_{\text{Lorey}}$ .....	72
Fig. 4.14. Estimated $H_{\text{Lorey}}$ using MLR based on PCs; a) model 1, b) model 5 from table 4.9 versus in situ $H_{\text{Lorey}}$ .....	74
Fig. 4.15. Estimated $H_{\text{Lorey}}$ using RF based on PCs versus in situ $H_{\text{Lorey}}$ (a, b, c: model 1 to 3, table 4.10) .....	75
Fig. 4.16. Estimated $H_{\text{Lorey}}$ using ANN based on waveform metrics (model 1, table 4.11) versus in situ $H_{\text{Lorey}}$ .....	76
Fig. 4.17. Estimated $H_{\text{Lorey}}$ using ANN based on waveform metrics (model 1, Table 4.12) versus in situ $H_{\text{Lorey}}$ .....	77
Fig. 4.18. Comparison of residual errors produced by MLR, RF and ANN for estimation of $H_{\text{Lorey}}$ based on PCs .....	78
Fig. 4.19. Correlation between volume and a) $H_{\max}$ , b) $H_{\text{Lorey}}$ .....	79

Fig. 4.20. Estimated volume using $V = 2.6507 H_{Lorey}^{1.5434}$ where $H_{Lorey}$ was estimated from GLAS data versus in situ volume .....	80
Fig. 4.21. Estimated volume using MLR based on waveform metrics versus in situ volume .....	81
Fig. 4.22. Estimated volume using RF based on PCs versus in situ volume .....	81
Fig. 4.23. Estimated volume using ANN based on PCs versus in situ volume .....	82
Fig. 4.24. Trend of forest volume bias (observed value-estimated value) by a) in situ forest volume, and b) terrain slope .....	84
Fig. 4.25. Estimated $H_{max}$ using MLR based on extracted indices from optical images versus reference $H_{max}$ .....	85
Fig. 4.26. Estimated $H_{max}$ using RF regression based on TI and extracted indices from optical and radar data versus reference $H_{max}$ .....	86
Fig. 4.27. Maximum height map produced using RF model .....	87
Fig. 4.28. Comparison of $H_{max}$ map produced from RF model with in situ $H_{max}$ measures at the location of 32 plots .....	88
Fig. 4.29. Estimated $H_{Lorey}$ using RF regression based on extracted indices from optical images and environmental data versus reference $H_{Lorey}$ .....	89
Fig. 4.30. Lorey's height map produced using RF model .....	90
Fig. 4.31. Comparison of $H_{Lorey}$ map produced from RF model with in situ $H_{Lorey}$ measures at the location of 32 plots .....	91
Fig. 4.32. Examples of fitted semivariograms of a) maximum height and b) Lorey's height residuals .....	92
Fig. 4.33. Maximum height map produced using regression-kriging .....	93
Fig. 4.34. Comparison of $H_{max}$ map produced from regression-kriging with in situ $H_{max}$ measures at the location of 32 plots .....	94
Fig. 4.35. Lorey's height map produced using regression-kriging .....	94
Fig. 4.36. Comparison of $H_{Lorey}$ map produced from regression-kriging with in situ $H_{Lorey}$ measures at the location of 32 plots .....	95

## List of tables

Table 2.1. Inventory form for DBH measurement .....	14
Table 2.2. Inventory form for height measurement .....	14
Table 2.3. ICESat operational periods .....	19
Table 2.4. GLAS data products .....	20
Table 2.5. Main characteristics of ALOS-1 PALSAR (Japan Space Center, 2012) .....	23
Table 2.6. Landsat-TM and SPOT 5-HRG spectral bands and their resolutions .....	24
Table 2.7. Distribution of in situ plots in elevation and slope classes .....	26
Table 2.8. Non-linear height-diameter functions .....	27
Table 2.9. Six selected non-linear height-DBH models and their statistical performance .....	28
Table 2.10. Six volume models and their parameters (produced by FRWO) .....	30
Table 2.11. Statistical summary of maximum height ( $H_{\max}$ ), Lorey height ( $H_{\text{Lorey}}$ ) and forest volume ( $V$ ) for 60 in situ plots .....	30
Table 2.12. The main information derived from GLA01 and GLA14 .....	34
Table 2.13. Definition of metrics extracted from GLAS waveforms .....	38
Table 3.1. Some MLR models fitted based on waveform metrics .....	53
Table 4.1. Statistics for five MLR to estimate $H_{\max}$ based on waveform metrics .....	60
Table 4.2. Statistics for three MLR models to estimate $H_{\max}$ based on PCs .....	64
Table 4.3. Statistics of five RF models for estimation of $H_{\max}$ based on waveform metrics .....	65
Table 4.4. Statistics of three RF models for estimation of $H_{\max}$ based on PCs .....	66
Table 4.5. Properties of two ANN models for estimation of $H_{\max}$ based on waveform metrics and the resulted statistics .....	67
Table 4.6. Properties of ANN models for estimation of $H_{\max}$ based on PCs and the resulted statistics .....	68
Table 4.7. Statistics for five MLR for estimation of $H_{\text{Lorey}}$ based on waveform metrics .....	71
Table 4.8. Statistics of three MLR models for estimation of $H_{\text{Lorey}}$ based on PCs .....	72
Table 4.9. Statistics of five RF models for estimation of $H_{\text{Lorey}}$ based on waveform metrics .....	73
Table 4.10. Statistics of three RF models for estimation of $H_{\text{Lorey}}$ based on PCs .....	75
Table 4.11. Statistics of three ANN models for estimation of $H_{\text{Lorey}}$ based on waveform metrics .....	76
Table 4.12. Statistics of three ANN models for estimation of $H_{\text{Lorey}}$ based on PCs .....	77
Table 4.13. Comparison of Lorey's height, number of trees and volume in three couple of plots .....	83



## 1. Introduction

Forests are of extreme importance to humans in many ways. They are watersheds, and have economic, environmental and climate control benefits. Forests and natural areas play a very important role in maintaining natural processes. Forests are one of the biggest reservoirs of carbon. They contain up to 80 percent of the aboveground carbon in the terrestrial communities and around 33 percent of the belowground carbon. So they help to keep the carbon cycle and other natural processes working and help reduce climate change. The relationship between forests and climate change is complex. On one hand forests can mitigate climate change by absorbing carbon, while on the other they can contribute to climate change if they are degraded or destroyed. In turn climatic changes may lead to forest degradation or loss – which intensifies climate change further. Figure 1.1 presents the different carbon pools and fluxes of the global carbon balance.

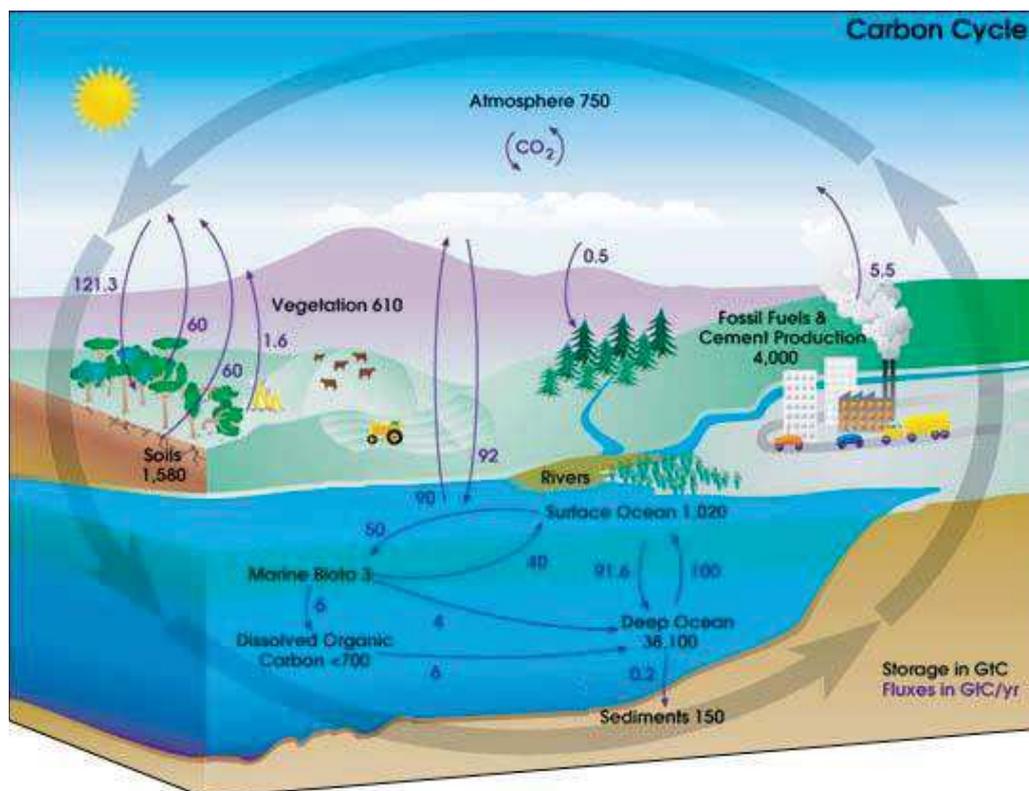


Fig. 1.1. Major carbon pools and fluxes of the global carbon balance (FAO, 2007)

The concentration of carbon dioxide ( $\text{CO}_2$ ) in the atmosphere increased from 285 ppm at the end of the nineteenth century, before the industrial revolution, to about 366 ppm in 1998 as a consequence of anthropogenic emissions of about 405 gigatonnes of carbon (C) ( $\pm 60$  gigatonnes C) into the atmosphere (IPCC, 2001).

Land-use change and soil degradation are major processes for the release of CO<sub>2</sub> to the atmosphere. The increase in greenhouse gases (GHGs) in the atmosphere is now recognized to contribute to climate change (IPCC, 2001). Although uncertainties remain regarding the causes, consequences and extent of climate change, it is believed that human activities are having an impact on the energy balance of the earth. Its influence on the climate is a major concern in the twenty-first century. This concern has led to the 1997 international agreement in Kyoto (the so-called Kyoto Protocol) made by the UNFCCC (United Nations Framework Convention on Climate Change), whereby most countries are committed to reducing their GHG emissions to the atmosphere.

Furthermore, at the Paris climate conference (COP21) in December 2015, 195 countries adopted the first-ever universal, legally binding global climate deal. The agreement sets out a global action plan to put the world on track to avoid dangerous climate change by limiting global warming to well below 2°C. Forests' role in combating climate change was formally recognized in this agreement. This recognition includes formal mention of the Reducing Emissions from Deforestation and Forest Degradation (REDD+) in the agreement text as well as new commitments to increase forest financing (Metzel, 2016).

Issues mentioned above highlight the increasing importance of sustainable forest management. Measuring biophysical parameters of forest is of primary steps for forest ecosystem management. Tree's height has a primary and fundamental importance among all other parameters. In fact the information about vertical structure of forest specifically height is important for ecosystem health assessment, site fertility, volume, biomass and carbon cycle measurement and monitoring (Namiranian, 2007; Cairns *et al.*, 1995).

Different heights are defined and measured for a tree including *total height*, *commercial height*, *trunk height*, *crown height* etc. In forestry, total height is defined as vertical distance between base of tree and top of it (Husch *et al.*, 2003; Namiranian, 2007). Generally, indirect methods are used to measure forest height in the field. The most common way is a tangent method which tree's height (h) is computed through measuring angles of top ( $\alpha$ ) and base ( $\beta$ ) of the tree by a clinometer, and horizontal distance of measurer to the tree (a):  $h = a[\tan \alpha - \tan \beta]$ .

It is not practical to measure the height of all trees in a forest stand. But the height of some trees are measured and then based on relationship between height and diameter at breast height (at 1.3 m above ground) (DBH), the height of all stand trees is estimated (Knowing that, measuring diameter is much easier and cheaper.). Depending on the subject and the objective, the maximum canopy height, mean height, mean Lorey's height or dominant height is measured or computed. The maximum canopy height in a stand is important for

assessing the site quality and trees growth rate. The mean Lorey's height as representative of mean height in uneven-aged stands is also a valuable parameter for forest ecosystem management.

Forest volume, measured in cubic meters per hectare, is considered for forest quantification. Stand volume at a nominated age is related to the site quality. Volume can also be used to estimate biomass quantities (dry weight of forest) and levels of carbon sequestered in the forest. In other words, the data for the forest biomass quantities depend importantly on the ability to measure forest volumes and conversion factors. Scientific researchers study relationship between forest biomass and biodiversity (Vance-Chalcraft *et al.*, 2010; Lasky *et al.*, 2014). Forest carbon estimates are of scientific importance to understand the quantitative role of forest carbon sequestration in earth's climate system (Hamburg *et al.*, 1997; Ding *et al.*, 2011). Changes in forest volume can be a good proxy for changes in forest carbon (Cheng *et al.*, 2013). Hence, volume may ultimately provide the most reliable estimates of deforestation and forest carbon changes (Maccauley *et al.*, 2009).

The most accurate method of measuring standing forest volume is to measure the DBH and the height of each tree (Namiranian, 2007; Maccauley *et al.*, 2009). For a large area of forest, sampling methods are used along with complex equations derived from regression models to estimate the forest volumes (Schreuder *et al.*, 1993; Namiranian, 2007).

For very large heterogeneous forests, measuring forest on the ground, quantitatively and qualitatively, would be prohibitively expensive, time consuming and laborious. This highlights the importance of remotely sensed data in global estimation of forest biophysical parameters. Digital large-scale remote sensing data provide a less expensive option for estimation of forest biophysical parameters over a large area, while potentially also providing accurate and unbiased estimates. Developed remote sensing techniques including photogrammetry (e.g. Miller *et al.*, 2000; Gobakken *et al.*, 2015), Synthetic Aperture Radar (SAR) interferometry (InSAR) and polarimetric interferometry (PolInSAR) (e.g. Balzter *et al.*, 2007; Garestier *et al.*, 2008), and lidar (light detection and ranging) (e.g. Lefsky *et al.*, 2007; Andersen *et al.*, 2006; Chen, 2010; Chen & hay, 2011; Duncanson *et al.*, 2009) make us able to measure three-dimension objects. For instance, Garestier *et al.* (2008) estimated forest canopy height over pine forest stands using P-band PolInSAR data with an RMSE of 2m. Balzter *et al.* (2007) provided a map of canopy height in Monks wood national nature reserve with special heterogeneity of vegetation type and density using dual-wavelength InSAR at X- and L-band. They evaluate the result using airborne imaging lidar data. The RMSE of the lidar canopy height estimates compared to theodolite data was 2.15 m (relative error 17.6%). The RMSE of the dual-wavelength InSAR-derived canopy height model compared to lidar (light detection and ranging) was 3.49 m (relative error 28.5%).

Concerning volume/biomass estimation, methods using radar and optical data are successful in forests with low to medium levels of biomass. Passive optical sensing have shown limited sensitivity to biomass using medium to high resolution imagery when the biomass reaches intermediate levels (150-200 Mg/ha) (Ploton *et al.*, 2011; Lu *et al.*, 2012). This is due to inability of optical data to detect variation in biomass density after complete closure of the canopy top, which can occur from low or intermediate biomass values (depending on forest characteristics). In SAR system, the saturation threshold of radar backscatters increases by the radar wavelength. For instance, L-band SAR systems (wavelength about 25 cm) are limited to low and intermediate biomass levels, with maximum values reaching 150t/ha (e.g. Sandberg *et al.*, 2011; Baghdadi *et al.*, 2014; Attarchi & Gloaguen, 2014). It also depends on the forest characteristics. Improvements in altimetry technology especially lidar led to the most direct measurements of forest structure including height of canopy and forest volume/biomass. To this date, canopy height estimation over large areas is best achieved using lidar data.

lidar is an active remote sensing system not limited to the canopy height, basal area, leaf area index, and canopy cover. It produces and transmits short laser pulses to the surface and objects. The returned pulses are captured by a telescope. Time delay between laser transmission and reception ( $t$ ) is converted to distance ( $D$ ) considering the speed of light ( $c = 3 \times 10^8$  m/s) ( $D = c \times t/2$ ). Since this ability is used for detecting height of objects, the lidar system is also called laser Altimeter. Scanning laser systems may be mounted on different platform; on tripod (terrestrial lidar system), on airplane (airborne lidar system) or on satellite (spaceborne lidar system). lidar was first developed as fixed-position terrestrial instrument for investigating atmospheric composition, clouds and aerosols. These systems produce dense point data with centimeter accuracy and are often used for localized terrain-mapping applications that require frequent surveys. Another type of terrestrial lidar systems uses mobile platforms (water-based and land-based). Data collected from these platforms are highly accurate and are used extensively to map railroads, roadways, airports, buildings, harbors, and shorelines. Airplanes are the most common and cost-effective platforms for acquiring lidar data over vast areas. Most airborne platforms can cover about 50 km<sup>2</sup> per hour and collect a large amount of detailed information for applications that require high-accuracy data (NOAA Coastal Services Center, 2012). Airborne lidar systems are able to produce centimeter accuracy high resolution digital elevation model (DEM) in a relatively small area (Hodgson *et al.*, 2003; Mount *et al.*, 2008). In spaceborne lidar systems, the lidar instrument is mounted on satellite operating in orbits of 700-800 km altitude and collect data over large area. An advantage of spaceborne lidar over airborne is providing global coverage of earth.

Many studies have been performed using airborne lidar to estimate different forest parameters like tree and forest height (Andersen *et al.*, 2006; Chen & hay, 2011; Khorrami *et al.*, 2014), volume (Tonolli *et al.*, 2011; Mohammadi, 2013), basal area (Drake *et al.*, 2002; Mohammadi, 2013), Leaf area index (Zhao & Popescu, 2009; Sabol *et al.*, 2014), and biomass (Gleason & Im, 2012; Takagia *et al.*, 2015).

Generally, there are many researches employing airborne lidar for studying different forest sciences which commonly achieved suitable results but they are site, instrument and species specific (Iqbal, 2010). In Iran, which is subject of the current study, only two studies were performed using airborne lidar in broadleaf mountainous forests: Mohammadi (2013) combined airborne lidar and UltraCam-D digital images to estimate standing volume, basal area and number of trees per hectare. He predicted standing volume and number of trees per hectare using non-parametric statistic method of support vector machine (SVM) with a relative RMSE of 31.4 and 35.5, respectively, and basal area using non-parametric method of random forest (RF) with a relative RMSE of 27.2. Khorrami *et al.* (2014) estimated individual tree's height using airborne lidar with sampling density of 4.5 points/m<sup>2</sup>. They obtained R<sup>2</sup> of 0.96 and 0.95 and RMSE of 1.05 m and 1.48 m for *Acer velutinum* (with height range of 6.8-36 m) and *Carpinus betulus* (with height range of 11-36 m), respectively.

The critical point about airborne lidar is that it is expensive and also the capacity to collect annual data over whole countries does not currently exist. There are some governmental restrictions that prevent access to airspace of sensitive areas or of foreign countries and also physical restrictions in employing aircraft over the remote ice caps and polar regions. While satellite orbits are not subject to these restrictions. Employing spaceborne lidar for estimation of forest biophysical parameters over large extent area was investigated since ICESat (Ice, Clouds, and Land Elevation Satellite) was launched into the space in 2003. The Geoscience Laser Altimeter System (GLAS) onboard ICESat operated for a total of 18 missions during its operational years (2003-2009). GLAS illuminated surface or footprint has a diameter of 70 m in diameter on average, and waveforms were acquired every 170 m along the track. GLAS has been used to retrieve forest canopy height and biomass since 2005 over planted (e.g. Rosette *et al.*, 2008a; Baghdadi *et al.*, 2014) or natural forests including coniferous (e.g. Lefsky *et al.*, 2005; Lefsky *et al.*, 2007; Chen, 2010; Duncanson *et al.*, 2010; Saatchi *et al.*, 2011), deciduous broadleaf (e.g. Lefsky *et al.*, 2007; Mitchard *et al.*, 2012; Los *et al.*, 2012; Khalefa *et al.*, 2013) and mixed coniferous-broadleaf forests (e.g. Sun *et al.*, 2008; Xing *et al.*, 2010; Los *et al.*, 2012). The most concerning point about GLAS data is waveform extent broadening over sloped area (mainly because of the large footprint size, about 70 m), and difficulties of canopy top and ground peak identification

due to mixed vegetation and ground returns (Lefsky *et al.*, 2005; Lefsky *et al.*, 2007; Chen, 2010). Chen (2010) has illustrated terrain slope effects and also plant size and distribution on canopy height estimation. It may occur that the base of tallest tree over a sloped terrain locates above or below the ground elevation peak that is recorded by lidar as representative of the tree base. On the other hand, top of the tallest tree over a sloped terrain may be below the maximum elevation (canopy top peak) recorded by lidar because of the presence of a shorter tree over higher slope. Another possibility is that maximum elevation recorded by lidar which is supposed to be the canopy top, is greater than the tree top. The later condition happens when very short sparse stands are located over sloped terrain. Even if the terrain is simple with approximately constant slope, a non-flat terrain might cause canopy height to be overestimated or underestimated, depending on the spatial distribution of plants within footprints. This is very challenging in our study since Hyrcanian forests in the north of Iran are mostly mountainous with considerable slopes.

The terrain information can be derived from ancillary DEMs (Lefsky *et al.*, 2005; Rosette *et al.*, 2008; Chen, 2010; Xing *et al.*, 2010) or from the waveform itself (Lefsky *et al.*, 2007; Pang *et al.*, 2008). Lefsky *et al.* (2007) proposed multiple transformations of three waveform metrics and then used stepwise regression to develop correction factors for broadened waveform extent (distance between two signals assigned to signal start and signal end) to estimate mean canopy height. Their algorithm estimated forest canopy height with an RMSE of 5 m ( $R^2 = 0.83$ ).

Duncanson *et al.* (2009) improved estimation of canopy height by modeling topography directly from GLAS waveforms and consequently, inclusion of terrain relief in canopy height estimation. They developed a model to estimate maximum relief ( $R^2 = 0.76$ ) and used it for classification of the maximum relief of the area sensed by GLAS. Forest canopy height model was also developed from waveform metrics for three separate relief classes: 0-7 m ( $R^2 = 0.83$ ), 7-15 m ( $R^2 = 0.88$ ) and >15 m ( $R^2 = 0.75$ ). The moderate relief class model resulted better predictions of forest height than the low (increasing in waveform metrics variability by terrain relief) and high (mixing of vegetation and terrain signals in waveforms) relief classes.

Chen (2010) adopted Lefsky *et al.*'s (2005; 2007) methods to retrieve maximum canopy height over mountainous areas (average slope=20°) including two conifers sites of tall and closed canopy and one broadleaf woodland site of shorter and sparse canopy. Three regressions (edge-extent linear and non-linear models and DEM-linear model) were used to remove slope effect on GLAS waveforms. The results showed a better performance for DEM-linear model (the best result was for the broadleaf woodland site with an RMSE of 4.88 m) rather than two edge-extent models. The author stated this result is reasonable

because DEM index (difference between maximum and minimum elevation of airborne lidar DEM within each GLAS footprint) was derived from airborne lidar data, which provides the most direct and precise information about terrain variability.

Lee *et al.* (2011) calculated height difference between real ground elevation and last GLAS waveform return based on size of footprint and terrain slope, and corrected canopy height retrieved from ICESat/GLAS data over slope terrains. They compared GLAS with airborne LVIS (Laser Vegetation Imaging Sensor) lidar heights. LVIS height was calculated as the mean of three tallest LVIS heights (slope-corrected with the same approach of GLAS) within each GLAS footprint. Slope-corrected GLAS vegetation heights matched well with top three LVIS (Laser Vegetation Imaging Sensor) mean (slope-corrected) heights ( $R^2 = 0.64$ , RMSE = 3.7 m). They found vegetation height can be overestimated by 3 m over a  $15^\circ$  slope without slope correction.

Concerning forest volume/biomass estimation, researchers attempted to estimate height from lidar data and then predicting volume and biomass using volume/biomass-height relationships in large extent areas (e.g. Lefsky *et al.*, 2005; Saatchi *et al.*, 2011; Mitchard *et al.*, 2012; Healey *et al.*, 2012; Baghdadi *et al.*, 2014; Asner & Mascaro, 2014). It was also considered to retrieve forest volume/biomass directly from waveform metrics. Boudreau *et al.* (2008), Duncanson (2009) and ZhiFeng *et al.* (2010) estimated above ground biomass (AGB) using Multiple Linear Regression (MLR) between AGB and metrics extracted from GLAS waveforms. Fu *et al.* (2009) and Nelson *et al.* (2009) practiced the same approach in their research where non-parametric technique of neural network was employed.

Although lidar is a promising technique for forest structure measurement, it does not provide wall-to-wall coverage except for small footprint lidar for a small area. Synergistic use of multiple sensors has been used for mapping forest volume/biomass accurately with remote sensing data (e.g. Lefsky *et al.*, 2010; Peterson & Nelson, 2014; Mitchard *et al.*, 2012; ZhiFeng *et al.*, 2010; Quiñones *et al.*, 2011). Peterson & Nelson (2014) produced a forest height map using a combination of spaceborne lidar (ICESat GLAS), airborne lidar, Landsat ETM+ images and field data for Alaska. To do this, forest height was estimated using a multiple linear regression based on waveform metrics extracted from GLAS waveforms and airborne lidar data. The estimated heights were evaluated using field measurement data. The resulting model was applied to all waveforms covering the study area. To spatially extrapolate the GLAS-based canopy height estimates and generate a continuous forest layer, a regression tree (RT) approach was used. GLAS-derived canopy height values were used as dependent variables while extracted values from Landsat composite bands, national elevation dataset DEM and derived slope and aspect, as well as existing vegetation type map were used as independent variables to build the RT models.

These models were then applied to the input geospatial layers to generate spatially continuous maps (30 m resolution) of forest height for all over Alaska.

Mitchard *et al.* (2012) produced AGB map for Gabon's Lope National Park (5000 km<sup>2</sup>) using a combination of terrain-corrected L-band radar (ALOS PALSAR), spaceborne lidar (ICESat GLAS) and ground based data. They estimated Lorey's height based on GLAS waveform features, and predict AGB from AGB-Lorey's height equation developed based on in situ data. An unsupervised classification was performed on original and synthetic bands extracted from PALSAR data to provide a map of vegetation structures (40 classes) for the study area. Then the AGB values derived from GLAS footprints were averaged within each class to produce a 100 m resolution AGB map.

Complex structure of forests in the north of Iran, vertically and horizontally, even and uneven aged stands, presence of diverse broadleaf species, severe topography, etc., brought into question the capability of GLAS data to estimate the forest canopy height and volume in such complexity. However a few literatures evaluated capability of optical images, airborne lidar and radar data for estimating forest volume/biomass in Iran (Khorrami *et al.*, 2008; Kalbi *et al.*, 2013; Mohammadi, 2013) and two studies employed the synergy of optical and radar data to provide biomass map (Attarchi & Gloaguen, 2014; Amini & Tetuko Sri Sumantyo, 2011), there was no investigation using spaceborne lidar or combination of that with other sources of remote sensing data up to now. This study aimed first to investigate capability of GLAS data for estimation of forest canopy height and volume in part of mountainous forests of Iran. To overcome slope effects, numerous parametric and non-parametric regressions were developed based on metrics derived from GLAS waveforms (user defined metrics and metrics derived using statistical method of principal component analysis (PCA)), and also terrain index extracted from 10 and 90 m DEM. Consequently, providing a wall to wall map of forest height was on the agenda. Hence, the best GLAS height models (maximum and Lorey's heights) were applied to all GLAS shots over study area. GLAS estimated heights were then used as reference heights to develop new height models based on indices extracted from radar and optical images and also environmental data. Canopy height maps (maximum and Lorey's heights) were produced using new height models and also regression-kriging method.

In summary, this research pursues following objectives:

- Estimation of maximum canopy height and Lorey's height using ICESat/GLAS data by developing parametric and non-parametric statistical methods between forest height and metrics extracted from GLAS waveforms and DEM.
- Estimation of forest volume using ICESat/GLAS data by developing: 1) volume-height relationship, and 2) parametric and non-parametric statistical methods between forest volume and metrics extracted from GLAS waveforms and DEM
- Producing forest height/volume map using a combination of ICESat/GLAS, ALOS/PALSAR, optical images and environmental data (aspect, slope and geology maps).

To address above objectives following questions were raised:

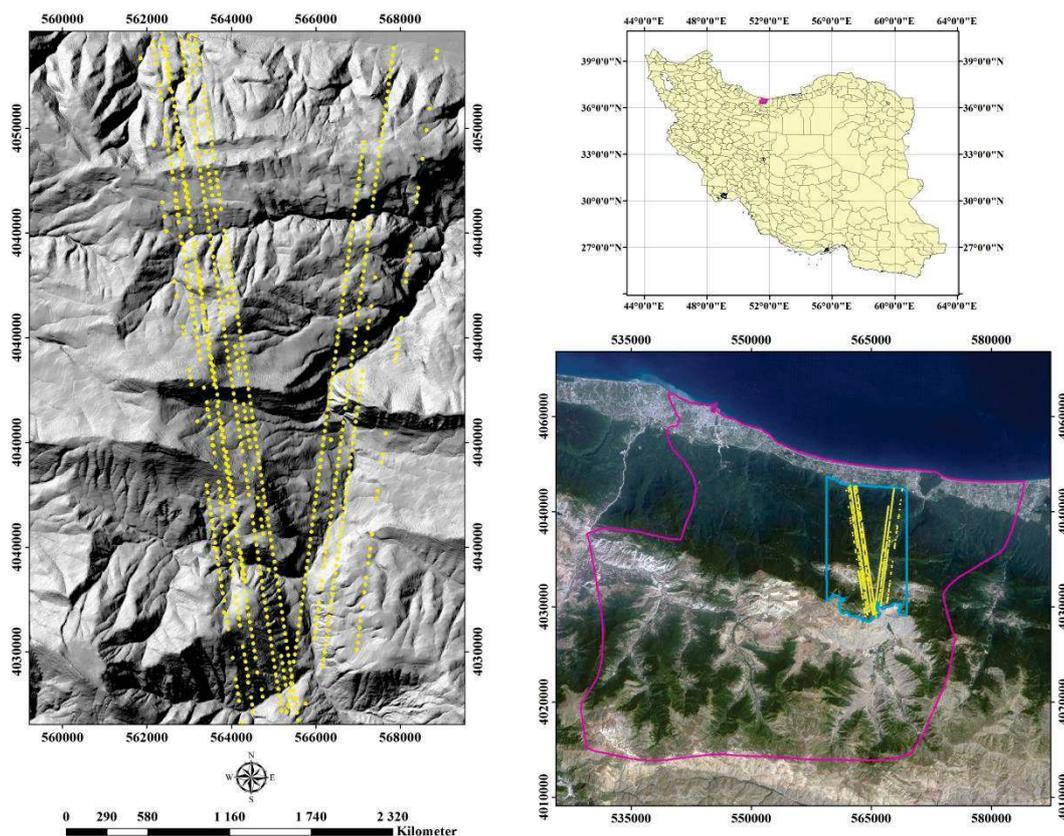
- Does GLAS estimate forest height/volume in mountainous forests of Iran with a suitable accuracy?
- What is preferable statistical method (multiple linear regression, artificial neural network, random; forest) for forest height/volume estimation using GLAS in our study area?
- Whether terrain index will reduce topography effects on GLAS waveform and improve the predictions? Does the resolution of DEM affect performance of the developed model?
- Is there possibility of providing forest height or volume map with a suitable accuracy in such high heterogeneity of horizontal and vertical structure?
- Does regression-kriging improves maps provided from regression models?



## 2. Materials and data processing

### 2.1. Study area

This research was performed in Nowshahr forests, a part of Hyrcanian forests in the north of Iran (Fig. 2.1), located between 36.26 to 36.68 degrees N latitudes and 51.32 to 51.94 degrees E longitudes. It contains temperate deciduous broadleaved forests extended from 100 to 2200 meters altitude above sea level with slopes ranging from flat to greater than 80%. Covering even and uneven aged stands with various species led to a diverse structure across the study site. Depending on the site, the dominant species are oriental beech (*Fagus orientalis*), European hornbeam (*Carpinus betulus*), chestnut-leaved oak (*Quercus castanifolia*), Persian ironwood (*Parotia persica*), oriental hornbeam (*Carpinus orientalis*), and Persian oak (*Quercus macranthera*). Annual mean precipitation is 1200 mm, and average maximum and minimum temperature are 6°C and 25°C, respectively. We focused on part of Nowshahr forest that is covered by lidar footprints.



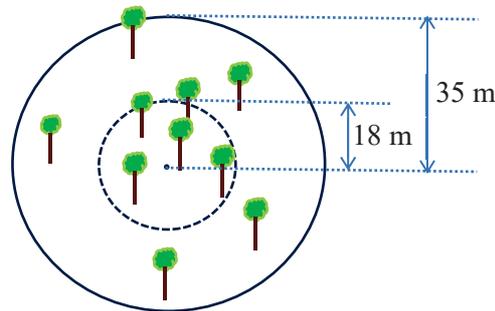
**Fig. 2.1.** Location of study area in Iran (top right map) and over Landsat image (bottom right map): the pink polygon shows the border of Nowshahr and the blue frame is the border of study area. The left map shows lidar footprints over hillshade of the study area.

## 2.2. Description of data

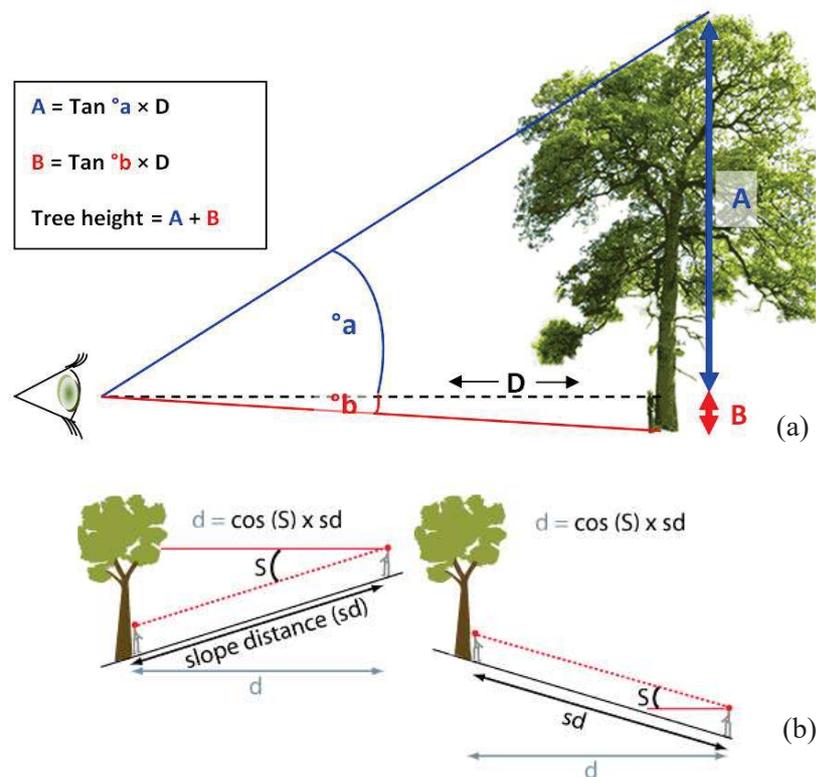
### 2.2.1. Field measurements

In order to estimate forest biophysical parameters (heights and volume) using GLAS, field data were collected during leaf-on seasons as lidar data acquisition dates. In total, 60 GLAS footprints were located on the ground using global positioning system (GPS; Garmin Colorado 300), 33 plots in September 2013 and 27 plots in May 2014. Since the study area includes climax community comprised of low growth deciduous species, trees have almost no or very small growth in height and DBH (diameter at breast height). On the other hand, spaceborne lidar is less sensitive to little changes in forest volume in comparison with very high spatial resolution data like airborne lidar. Therefore, 5-6 year interval between in situ measurement and lidar acquisition was not considered as an important factor influencing the analysis.

As the first step, point layer of GLAS footprints and also roads were added to topographic and slope map of the study area in ArcGIS to consider conditions and accessibility of each plot before going on the field mission. Then geographic coordinates of the plots (location of center of GLAS footprints) were entered in GPS and were next navigated on the ground. DBH of all trees (DBH > 7.5 cm) within a 70 m diameter circle (as large as lidar footprints) were measured using caliper or tape measure in case of large diameter trees. As laser energy decreases towards the margins of the footprint and, consequently, the returned waveform is most representative of the features closest to the footprint center (Harding & Carabajal, 2005; Rosette *et al.*, 2008a; Duong, 2010), this was taken into account through field measurements. So totally 10 dominant heights, 5 within a 36 m diameter circle and 5 in a co-center 70 m diameter circle (outer margin of smaller circle), were measured using a clinometer. The height of the nearest tree to the plot center (or tree with lowest DBH if the nearest is a measured dominant tree), was measured in order to have sufficient trees with different DBH for building height-DBH relationships which will be discussed later in subsection 2.3.1. Figure 2.2 shows a schematic of height measurement plan in a plot. Obtaining tree height requires the use of basic trigonometry:  $h = d[\tan \alpha - \tan \beta]$  where  $h$  is the tree height,  $d$  is the horizontal distance from tree,  $\alpha$  is the angle to the top of tree and  $\beta$  is the angle to the base of tree (Fig. 2.3a). On steep terrain it is almost impossible to accurately determine the horizontal distance from the tree. In situations where the ground is sloped, it is needed to measure slope distance. Once measure slope angle and slope distance was measured, horizontal distance can be calculated (Fig. 2.3b). All required information including environmental parameters such as slope, aspect and elevation level and also biophysical parameters were recorded in inventory forms as presented in table 2.1 and 2.2.



**Fig. 2.2.** Schematic illustration of location of trees for height measurement in a field plot; 5 trees within a 35 m diameter plot and 5 in a co-center 70 m diameter plot (outer margin of smaller circle)



**Fig. 2.3.** a) Schematic illustration of height measurement based on trigonometry; b) Calculation of horizontal distance from tree over sloped terrain

Third mission of field measurement was performed in August 2016 to validate canopy height maps produced using combination of GLAS, PALSAR, optical images (Landsat-TM and SPOT5) and environmental data. Thus, 32 circle plots (each 0.1 hectare) dispersed over part of the study area, were measured in case of DBH of all trees and height of three dominant trees as explained above. Simple random sampling method was used to select and determine plots locations. Figure 2.4 illustrates three phases of field work including navigating and locating plot center, DBH measurement and height measurement.

**Table 2.1.** Inventory form for DBH measurement

Plot Number: 75    Slope: 60%    Aspect: North    Elevation: 1950m    Date: 12/09/2013					
Tree number	Plot		Species	DBH (cm)	Considerations
	Small (r = 18m)	Large (r = 35m)			
1	✓		<i>Carpinus betulus</i>	24.5	
2	✓		<i>Fagus orientalis</i>	50	It is one of ten top trees in the plot.
3	✓		<i>Fraxinus excelsior</i>	34	
4		✓	<i>Fagus orientalis</i>	35	
5		✓	<i>Fagus orientalis</i>	33	
6		✓	<i>Fagus orientalis</i>	32	
⋮					
184		✓	<i>Acer campestre</i>	16	The last measured tree in this plot

**Table 2.2.** Inventory form for height measurement

Plot Number: 75    Slope: 60%    Aspect: North    Elevation: 1950m    Date: 12/09/2013								
Tree number	Species	DBH (cm)	Tan of top of tree angle (%)	Tan of base of tree angle (%)	Slope angle (degree)	Distance to the tree (m)	Plot	
							Small (r = 18m)	Large (r = 35m)
၁	<i>Fagus orientalis</i>	42	+8	-62	29	27.90	✓	
၂	<i>Fraxinus excelsior</i>	32	+36	-22	10	29.10	✓	
၃	<i>Quercus castanifolia</i>	57	+6	-58	27	24.90	✓	
၄	<i>Fagus orientalis</i>	34	+105	+44	26	38.80	✓	
၅	<i>Fagus orientalis</i>	50	+19	-45	23	26.70	✓	
၆	<i>Fagus orientalis</i>	54	+7	-60	29	34.40		✓
၇	<i>Quercus castanifolia</i>	65	+23	-34	16	28		✓
၈	<i>Quercus castanifolia</i>	83	+12	-50	24	28.90		✓
၉	<i>Fagus orientalis</i>	44	+32	-26	11	32.80		✓
၁၀	<i>Fagus orientalis</i>	42	+25	-556	27	31.60		✓
၁၁	<i>Fagus orientalis</i>	31	+105	+53	26	37	✓	

**Considerations:** Tree number 11 is the closest tree to the center of plot



Fig. 2.4. illustration of three phases of field work including navigating and locating the plot center, DBH and height measurement

### 2.2.2. Digital elevation model

Digital elevation model was provided using two sources of data. The first one, Shuttle Radar Topography Mission (SRTM) data sampled at 3 arc-second (about 90 meters). Elevations were measured in meters referenced to the WGS84/EGM96 geoid. As all data used in a research project should have the same coordinate system, including both horizontal and vertical aspects, geoidal heights were transferred to ellipsoidal heights by adding the geoid undulations to geoidal heights ( $DEM_{90}$ ) (Equation 2.1). Figure 2.5 illustrates the relationship between geoidal heights and ellipsoidal heights. The second source of data was digital 1:25000 topographic maps with counter interval of 10 meters and used to produce DEM with 10 meter resolution ( $DEM_{10}$ ).

$$h = H + N \quad (2.1)$$

Where  $h$ ,  $N$  and  $H$  stands for ellipsoidal height, geoid undulation and geoidal height, respectively.

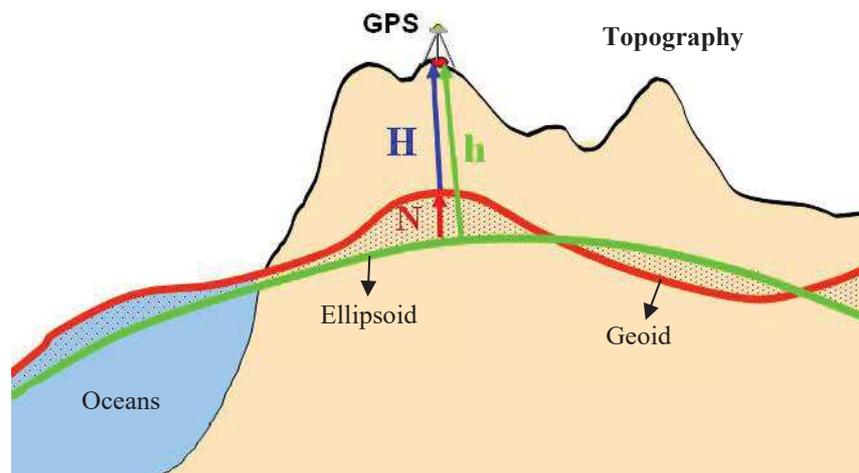
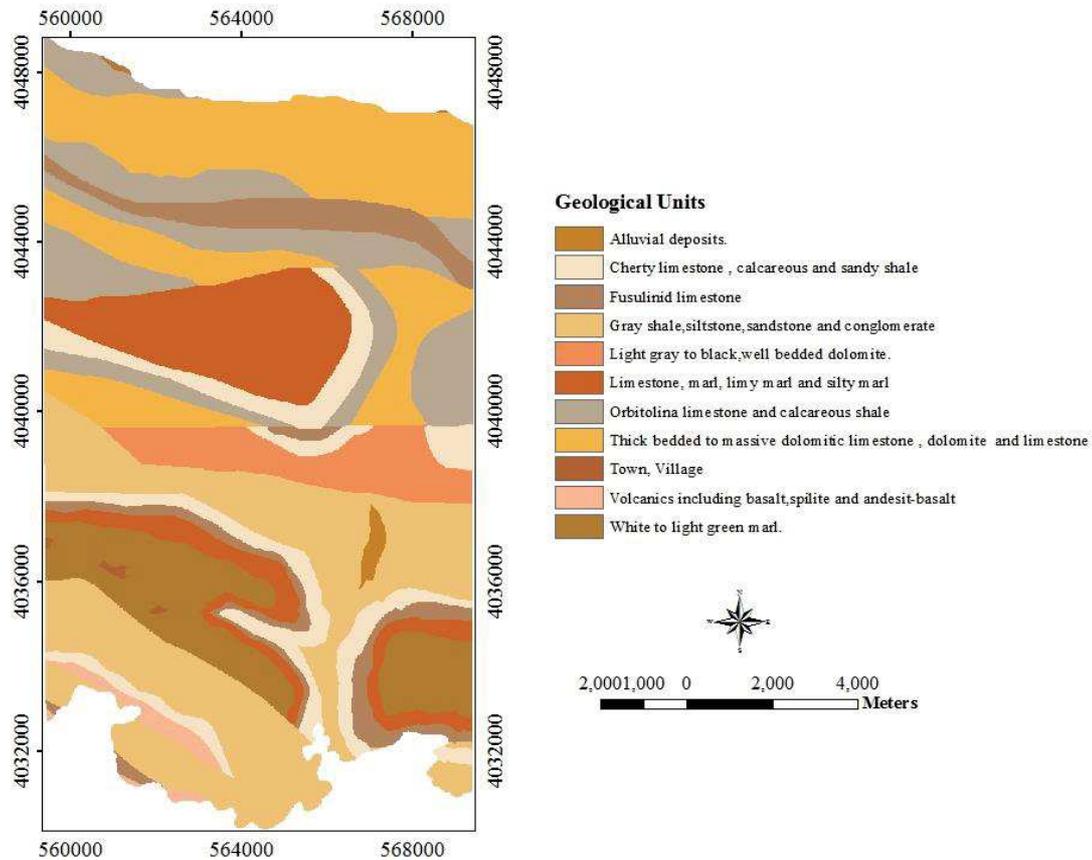


Fig. 2.5. Schematic of the relationship between geoidal heights and ellipsoidal heights

### 2.2.3. Geological map

Rock largely controls the physico-chemical properties of resulting soils which affects tree growth and forest parameters. A geological map produced by geological survey and mineral exploitation of Iran, at the scale of 1:100000, was therefore used in this study. Two sheets of geological map inside the study area were merged and converted to raster layer with different resolutions to match other sources of data. As it is seen in figure 2.6, all classes except one belong to major category of the sedimentary rocks.



**Fig. 2.6.** Geological map provided by geological survey and mineral exploitation of Iran (The white color corresponds to urban and non-forest area)

#### 2.2.4. GLAS/ICESat

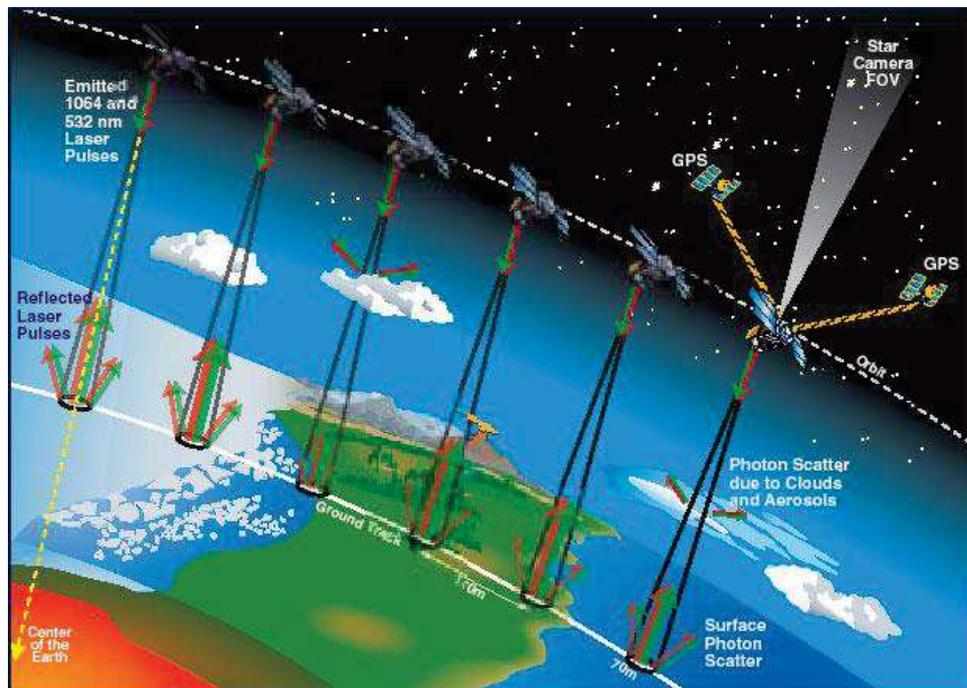
Lidar is an active remote sensing system that produces and transmits short laser pulses to the surface and objects. The returned pulses are captured by a telescope. Time delay between laser transmission and reception ( $t$ ) is converted to distance ( $D$ ) considering the speed of light ( $c = 3 \times 10^8 \text{ m/s}$ ) (Equation 2.2). Since this ability is used for detecting height of objects, the LiDar system is also called laser Altimeter.

$$D = c \times t/2 \quad (2.2)$$

Scanning laser systems may be mounted on different platform; on tripod (terrestrial lidar system), on airplane (airborne lidar system) or on satellite (spaceborne lidar system). In spaceborne lidar systems, the lidar instrument is mounted on satellite operating in orbits of 700-800 km altitude and providing global coverage of earth.

ICESat is an experimental scientific satellite launched by NASA in 2003 to measure mainly ice sheet elevations and its changes over the time and also to provide measurements of cloud and aerosol height profiles, land elevation, and vegetation cover. ICESat moves

26,000 km per hour on orbits at 600 km altitude and 94 degrees inclination to the equator. The Geoscience Laser Altimeter System (GLAS) onboard ICESat consists of three lasers that operate exclusively to measure distance, a Global Positioning System (GPS) receiver, and a star-tracker attitude determination system. The laser transmits short pulses (4 ns) of infrared light (1064 nm) for measuring the elevation of surfaces and dense clouds and visible green light (532 nm) for measuring the vertical distribution of clouds and aerosols (Aronoff, 2005; Pflugmacher, 2008). Laser pulses at 40 times per second illuminate 70 meter diameter footprints on average, spaced at 170-meter intervals along Earth's surface. Separation of the tracks is 15 km at the equator and 2.5 km at 80 degrees latitude (Abshire *et al.*, 2005; NSIDC, 2012). The mean horizontal geolocation accuracy of the ground footprints is less than 5 m for all ICESat missions (NSIDC, 2014). Mean vertical geolocation accuracy was also reported by NSIDC between 0 and 3.2 cm over flat surfaces (NSIDC, 2014). GLAS operated for a total of 18 missions during its operational years (2003-2009). Figure 2.7 illustrates operation of GLAS from ICESat orbiting the Earth. A list of ICESat missions and their operational period is seen in table 2.3 (NSIDC, 2014).



**Fig. 2.7.** Schematic illustration of GLAS instrument operating from ICESat while orbiting the Earth

**Table 2.3.** ICESat operational periods

Laser Identifier	Start Date	End Date	Days in Operation	Duration of repeat orbit cycle (days)
L1A	2003-02-20	2003-03-21	29	8
L1B	2003-03-21	2003-03-29	9	8
L2A	2003-09-25	2003-10-04	10	8
L2A	2003-10-04	2003-11-19	45	91
L2B	2004-02-17	2004-03-21	34	91
L2C	2004-05-18	2004-06-21	35	91
L3A	2004-10-03	2004-11-08	37	91
L3B	2005-02-17	2005-03-24	36	91
L3C	2005-05-20	2005-06-23	35	91
L3D	2005-10-21	2005-11-24	35	91
L3E	2006-02-22	2006-03-28	34	91
L3F	2006-05-24	2006-06-26	33	91
L3G	2006-10-25	2006-11-27	34	91
L3H	2007-03-12	2007-04-14	34	91
L3I	2007-10-02	2007-11-05	37	91
L3J	2008-02-17	2008-03-21	34	91
L3K	2008-10-04	2008-10-19	16	91
L2D	2008-11-25	2008-12-17	23	91
L2E	2009-03-09	2009-04-11	34	91
L2F	2009-09-30	2009-10-11	12	91

It should be noted footprint shape is not fully circular. Laser 3 footprints are moderately elliptical, Laser 2 very elliptical and Laser 1 very elliptical with side-lobe (NSIDC, 2012). The different footprint size and shape make it difficult to describe the surface covered by footprints when all missions' data are employed in a project. This may causes uncertainty in estimations. Within each footprint, laser reflected energy by all intercepting objects and surfaces are collected by a telescope of 1 meter diameter and results a waveform that represents a vertical profile of laser-illuminated surfaces. In the early GLAS missions, energy of each returned pulse was telemetered in 544 bins over ice sheet and land, corresponding to a height of 81.6 m (each bin corresponds to one nanosecond) (Brenner *et al.*, 2003). In highly sloped area or area where feature heights exceed 81.6 m, GLAS waveform would truncate, making it impossible to derive range information. So, in later operations height extent was increased to 150 m (1000 bins) over land, using a “waveform compression scheme” (Harding & Carabajal, 2005).

GLAS collected data were processed by National Snow and Ice Data Center (NSIDC), and 15 products at three levels of corrections; L1A, L1B and L2 were provided (Table 2.4).

**Table 2.4.** GLAS data products

Short name	Long name
GLA01	GLAS/ICESat L1A Global Altimetry Data
GLA02	GLAS/ICESat L1A Global Atmosphere Data
GLA03	GLAS/ICESat L1A Global Engineering Data
GLA04	GLAS/ICESat L1A Global Laser Pointing Data
GLA05	GLAS/ICESat L1B Global Waveform-based Range Corrections Data
GLA06	GLAS/ICESat L1B Global Elevation Data
GLA07	GLAS/ICESat L1B Global Backscatter Data
GLA08	GLAS/ICESat L2 Global Planetary Boundary Layer and Elevated Aerosol
GLA09	GLAS/ICESat L2 Global Cloud Heights for Multi-layer Clouds
GLA10	GLAS/ICESat L2 Global Aerosol Vertical Structure Data
GLA11	GLAS/ICESat L2 Global Thin Cloud/Aerosol Optical Depths Data
GLA12	GLAS/ICESat L2 Antarctic and Greenland Ice Sheet Altimetry Data
GLA13	GLAS/ICESat L2 Sea Ice Sheet Altimetry Data
GLA14	GLAS/ICESat L2 Global Land Surface Altimetry Data
GLA15	GLAS/ICESat L2 Ocean Altimetry Data

Since this research has two main parts (Estimation of forest biophysical parameters using GLAS data and providing height/volume map using synergy of GLAS, PALSAR, optical images and environmental data), GLAS data acquired on October 2007 and October 2008, corresponding to L3I and L3K missions respectively, were used for the first part of this research (estimation of forest biophysical parameters using ICESat GLAS data). All GLAS missions over the study area were also used for the second part of thesis to provide height map using combination of remote sensing data (lidar, radar and optical images) and environmental data. GLA01 and GLA14 among 15 products produced by NSIDC, release 33<sup>1</sup>, were employed to drive forest height and volume.

GLA01 contains intensities of transmitted and received waveforms. GLAS digitizes these intensities as counts (0-255) which were converted to volts using calibration table. It is worth to mention the ordering of the transmitted pulse is in time order. The value of the first sample (bin) is for the sample closest to the spacecraft in time, and the value of the last sample is for the sample farthest from the spacecraft in time. In reverse, the received echo is in time-reversed order. So the value of the first sample is for the sample farthest from the spacecraft in time, and vice versa (NSIDC, 2012).

1- At the moment, there are release 34 data. A new release is created when changes occur in the input data or when improvements are made to the processing algorithms.

GLA14 is a level-2 elevation product derived from level-1 products GLA05 and GLA06. This product provides information about surface elevations. It also includes the laser footprint geolocation and reflectance, as well as geodetic, instrument, and atmospheric correction flags to filter out bad data (more explanation in subsection 2.3.3.1) for range measurements (NSIDC, 2012; Wang *et al.*, 2011). As the transmitted and received pulses are assumed to have a Gaussian shape (Brenner *et al.*, 2003), up to 6 Gaussian peaks were fitted to the waveforms. Hence, the GLA14 contains parameters for these peaks including amplitude, area and standard deviation.

### 2.2.5. ALOS/PALSAR

The word radar stands for radio detection and ranging. In general, radar systems use modulated waveforms and directive antennas to transmit electromagnetic energy into a specific volume in space to search for targets. Objects (targets) within a search volume will reflect portions of this energy (radar returns or echoes) back to the radar. These echoes are then processed by the radar receiver to extract target information such as range, velocity, angular position, and other target identifying characteristics. In this research spaceborne radar data acquired by ALOS-1 (Advanced Land Observing Satellite) was used along with other source of remotely sensed data.

ALOS-1 (Advanced Land Observing Satellite) was launched on January 24, 2006 from the Tanegashima Space Center and completed its operation on 12 May 2011 (Japan Space Systems, 2012). ALOS satellite has three remote-sensing instruments: the Panchromatic Remote-sensing Instrument for Stereo Mapping (PRISM) and for digital elevation models. The Advanced Visible and Near Infrared Radiometer type 2 (AVNIR-2) for precise land coverage observation, and the Phased Array type L-band Synthetic Aperture Radar (PALSAR). PALSAR is an active microwave sensor using L-band frequency (center frequency is 1270 MHz) to achieve cloud-free and day-and-night land observation. It was improved based on Synthetic Aperture Radar (SAR) onboard the first observation satellite, JERS-1. Four operation modes are defined for PALSAR; fine resolution mode, direct downlink mode, scanSAR mode, and polarimetric mode. FB (Fine resolution Beam) mode comprises 18 selections in the off-nadir angle range between  $9.9^\circ$  and  $50.8^\circ$ , each with 4 alternative polarizations: single polarization HH or VV, and dual polarization HH+HV or VV+VH. Out of the 72 possible FB modes, two have been selected for operational use. The direct transmission (or downlink) mode is a contingency backup mode which allows the downlink of the FB mode data to local ground stations in case the high-speed DRTS (Data Relay and Test Satellite) becomes unavailable. ScanSAR is available at a single polarization only (HH or VV) and can be operated with 3, 4, or 5 sub-beams

transmitted in short (14 MHz) or long bursts (28 MHz). Out of the 12 ScanSAR modes available, the sort-burst, HH polarization, 5-beam mode has been selected for operational support. It features a 350 km swath width with an incidence angle range of 18-43°. The 14 MHz polarimetric mode provides the full quad-polarization (HH+HV+VH+VV) scattering matrix with 12 alternative off-nadir angles between 9.7° and 26.2°. Polarization is changed in every pulse of transmission signal, and dual polarization signals are simultaneously received. The operation is limited in lower incident angle in order to achieve higher performances. At the nominal off-nadir angle (21.5°), the swath width is 30 km with 30 m spatial resolution under the maximum data rate condition (240 Mbit/s) (Ito *et al.*, 2001)

Full polarimetry (multi-polarization), off nadir pointing function and other functions of PALSAR improved the accuracy of analyzing geological structure, distribution of rocks and so on, and acquired a lot of effective data for resource exploration and other purposes. At the same time, multi-polarization was effective in acquiring vegetation information, which encouraged the data use in fields such as global and regional observation of vegetation, distinguishing feature on the ground, classification of land use and other purposes (Polychronaki *et al.*, 2013; Attarchi & Gloaguen, 2014; Mermoz *et al.*, 2014). Main characteristics of PALSAR were presented in table 2.5.

PALSAR mosaic data, supplied by JAXA<sup>1</sup>, with 25 m resolution from dates 2007, 2008, 2009 and 2010 were used in this research. It included four layers “HH” and “HV” polarization, “Local incidence angle” and “date” layer from the date of satellite launch (January 24, 2006).

---

1- Japan Aerospace Exploration Agency (JAXA)

**Table 2.5.** Main characteristics of ALOS-1 PALSAR (Japan Space Center, 2012)

Observation Mode		Fine Resolution		Direct Downlink	ScanSAR	Polarimetric	Remarks
Polarization		L band					
		HH or VV	HH + HV or VV + VH	HH or VV	HH or VV	HH + HV+ VV + VH	
Incidence Angle		8° - 60°	8° - 60°	8° - 60°	18° - 36° (3 scan) 18° - 40° (4 scan) 18° - 43° (5 scan)	8° - 30°	Off-nadir Angle: 9.7° - 50.8°
Resolution	Range	10 m*	20 m*	20 m*	100 m*	30 m*	Number of looks of the ScanSAR mode is 8 or more by both range and azimuth.
	Azimuth	10 m (2looks) 20 m (4looks)	10 m (2looks) 20 m (4looks)	10 m (2looks) 20 m (4looks)	100 m	10 m (2looks) 20 m (4looks)	
Swath width		70 km*	70 km*	70 km*	70 km (3 scan) 300 km (4 scan) 350 km (5 scan)	30 km*	
Data Rate		240 Mbps		120 Mbps	120/240 Mbps	240 Mbps	
Radiometric Accuracy		Relative accuracy within 1 scene: < 1dB Relative accuracy within 1 orbiting: < 1.5dB(TBR)					
<p>* Meets under the following off-nadir angle. For under other angle, it is requested to keep similar specification as far as possible.  High Resolution Mode: Off-nadir angle 34.3 deg. (TBR)  Direct Downlink Mode: Off-nadir angle 34.3 deg. (TBR)  ScanSAR Mode: 4th scan (off-nadir 34.1 deg. (TBR))  Polarimetric Mode: Off-nadir angle 21.5 deg. (TBR)  Note: Above descriptions are specifications over the equator.</p>							

### 2.2.6. Passive optical remote sensing data

Optical remote sensing makes use of visible, near infrared and short-wave infrared sensors to form images of the earth's surface by detecting the solar radiation reflected from targets on the ground. Different materials reflect and absorb differently at different wavelengths. Thus, the targets can be differentiated by their spectral reflectance signatures in the remotely sensed images. Vegetation has a unique spectral signature which enables it to be distinguished readily from other types of land cover in an optical/near-infrared image. The reflectance is low in both the blue and red regions of the spectrum, due to absorption by chlorophyll for photosynthesis. It has a peak at the green region which gives rise to the green color of vegetation. In the near infrared (NIR) region, the reflectance is much higher than that in the visible band due to the cellular structure in the leaves. Hence, vegetation can be identified by the high NIR but generally low visible reflectance.

In this research, available cloud free images of two multispectral remote sensing systems of Landsat-TM and spot5, from 2003 onwards (since GLAS data belongs to different mission periods and do not reflect a static moment in time), were selected and analyzed.

The Landsat Thematic Mapper (TM) sensor was carried onboard Landsats 4 and 5 from July 1982 to May 2012 with a 16-day repeat cycle. It possesses seven spectral bands with a

spatial resolution of 30 meters for bands 1 to 5 and 7. Spatial resolution for Band 6 (thermal infrared) is 120 meters, but is resampled to 30-meter pixels. For this study, four dates of cloud free orthorectified images of Landsat5-TM (Table 2.6), held in the USGS archives, were downloaded from EarthExplorer: <http://earthexplorer.usgs.gov>.

SPOT-5 is the fifth satellite in the SPOT series of CNES (Space Agency of France) which was launched on May 2002 and its commercial mission was ended on 27 March 2015. On 2 April 2015, an experimental phase for the mission, SPOT-5 (Take 5), involved the satellite being tasked in a new orbit (lowered by 3 km with 5 day repeat cycle). From that date, and for 5 months until the 15th of September, SPOT5 (Take 5) observed 150 sites every five days with constant observation angles. The data were processed and distributed at CNES. Cloud free orthorectified multispectral bands of SPOT 5-HRG acquired on April and June 2015 were downloaded from the website of Centre d'Etudes Spatiales de la BIOSphère (CESBIO) (<http://www.cesbio.ups-tlse.fr>) and employed in this research (Table 2.6). Figure 2.8 shows a color composite of SPOT data over study area.

**Table 2.6.** Landsat-TM and SPOT 5-HRG spectral bands and their resolutions

Sensor	Spectral bands	Wavelength (micrometers)	Resolution (meters)	Acquisition dates
Landsat 5-TM	Band 1 (Blue)	0.45-0.52	30	
	Band 2 (Green)	0.52-0.60	30	
	Band 3 (Red)	0.63-0.69	30	04/10/2008
	Band 4 (Near-Infrared)	0.76-0.90	30	08/11/2009
	Band 5 (Near-Infrared)	1.55-1.75	30	04/06/2010
	Band 6 (Thermal)	10.40-12.50	120* (30)	29/12/2010
	Band 7 (Mid-Infrared)	2.08-2.35	30	
SPOT-5 (take 5) HRG	Band 1 (Green)	0.50-0.59	10	15/04/2015
	Band 2 (Red)	0.61-0.68	10	20/04/2015
	Band 3 (Near-Infrared)	0.78-0.89	10	19/06/2015
	Band (Shortwave Infrared)	1.58-1.75	10	24/06/2015
* TM Band 6 was acquired at 120-meter resolution, but products are resampled to 30-meter pixels.				

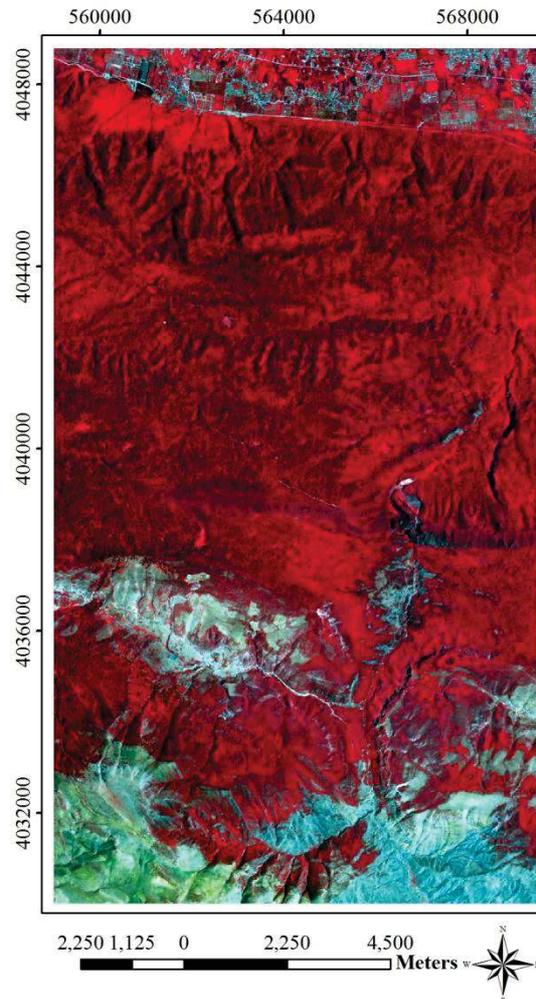


Fig. 2.8. Color composite of SPOT-5 images over study area.

## 2.3. Data processing and information extraction

### 2.3.1. Analysis of in situ data

In situ measurements were carried out in two phases. Phase one includes 60 plots for developing GLAS height and volume models and their validation. It was intended to collect data in all elevation and slope classes, but it was not practical as a reason of lack of GLAS data or lack of forest cover in some elevation range, and also inaccessibility to some area. Phase two includes 32 plots to validate height maps produced from combination of remotely sensed and environmental data. Table 2.7 shows their conditions in terms of elevation and slope.

**Table 2.7.** Distribution of in situ plots in elevation and slope classes

	Elevation (m)			Slope (%)				
	< 200	1000-1500	1500-2000	<10	10-20	20-40	40-50	> 50
Plots (phase 1)	3	19	38	6	11	19	12	12
Plots (phase 2)	0	18	14	3	5	16	3	5

Maximum canopy height was computed using equation:  $h = a[\tan \alpha - \tan \beta]$ . As it was described in section 2.2.1, five top tall trees were measured in a small circle plot of 36 m diameter and five top tall trees in a co-center 70 m diameter circle plot but outer margin of small plot. So the tallest tree inside and outside of the small plot was called  $H_{\max\text{-in}}$  and  $H_{\max\text{-out}}$ , respectively. The tallest one among ten top trees was also selected and called  $H_{\max\text{-total}}$ .

To calculate the height of all trees in each plot (in addition to 11 trees that were measured), a variety of non-linear models relating DBH to height, recommended in different studies were selected and tested (Table 2.8). These relationships were considered for four species as 1) *Fagus orientalis*, 2) *Carpinus betulus*, 3) *Quercus castanifolia*, 4) *Alnus subcordata*, and two groups of species (similar in shape and height) as Group1 included *Tilia begonifolia*, *Acer velutinum*, *Acer cappadocicum*, *Sorbus torminalis* and *Fraxinus excelsior*, and Group2 included *Quercus macranthera*, *Carpinus orientalis*, *Parotia persica* and *Diospyros lotus*. These six categories have been chosen based on six forest volume tables produced by Forests, Range & Watershed Management Organization (FRWO) for northern forests of Iran. To select the best regression model among a number of models, several most commonly used criteria such as adjusted coefficient of determination ( $R^2_a$ ), Root Mean Square Error (RMSE) and Akaike Information Criterion (AIC) were evaluated (Burnham & Anderson, 2002). Besides statistical criteria, biological behavior of models was considered to select the best model.

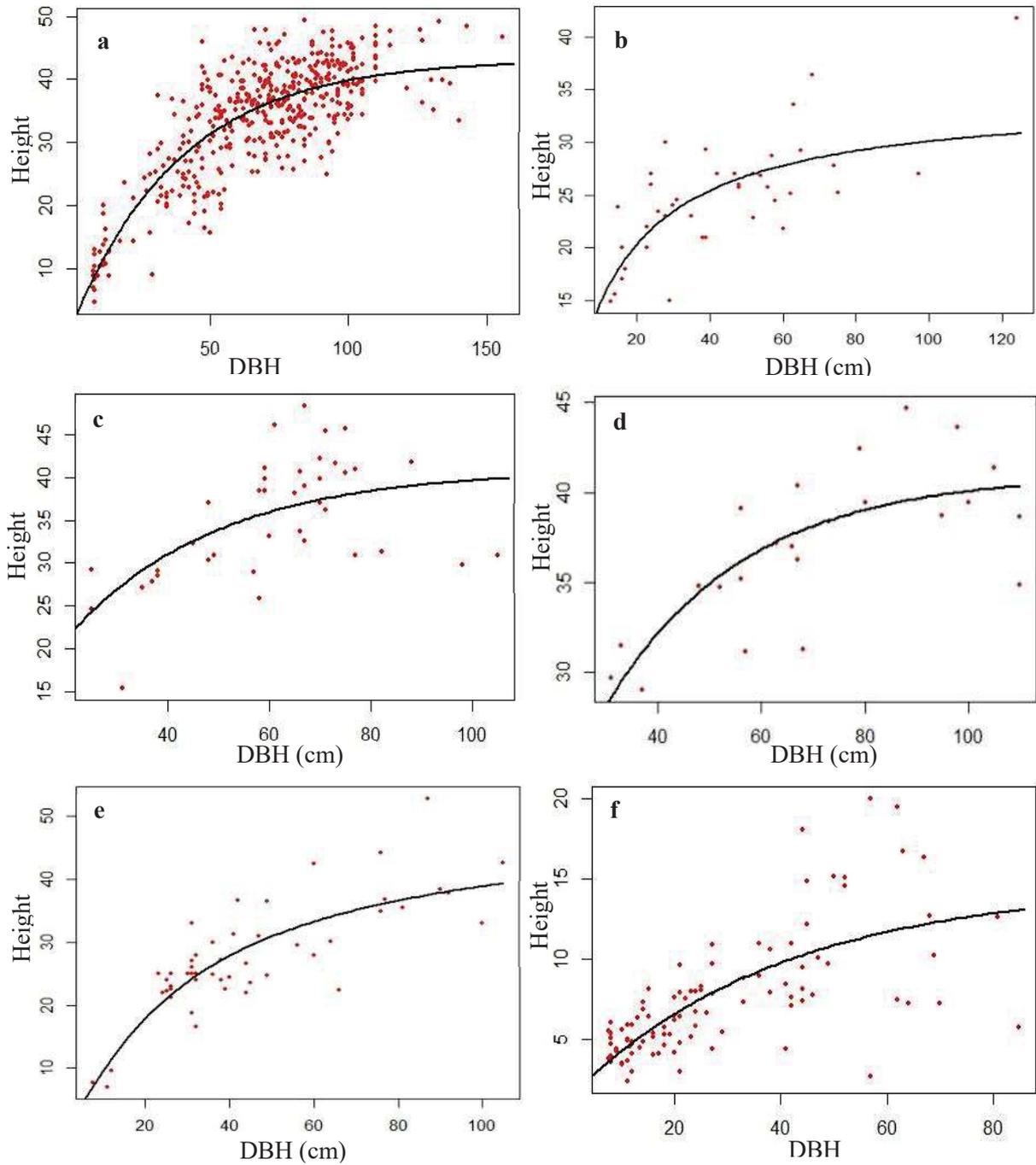
**Table 2.8.** Non-linear height-diameter functions

Model	Reference
$H = 1.3 + a(1 - e^{-b.DBH^c})$	Yang <i>et al.</i> , 1978; Zhang, 1997; Peng <i>et al.</i> , 2001
$H = 1.3 + a(1 - e^{-b.DBH})^c$	Batista <i>et al.</i> , 2001; Zhang, 1997
$H = 1.3 + a \cdot e^{\left(\frac{b}{DBH+c}\right)}$	Huang <i>et al.</i> , 1992; Peng, 1999; Ahmadi <i>et al.</i> , 2013
$H = 1.3 + \frac{a}{(1 + b^{-1}DBH^{(-c)})}$	Ratkowsky & Reedy, 1986; Huang <i>et al.</i> , 1992
$H = 1.3 + a \cdot e^{(-b.DBH^c)}$	Zhang, 1997; Ahmadi <i>et al.</i> , 2013
$H = 1.3 + a \cdot DBH^b$	Huang <i>et al.</i> , 1992 ; Silva Scaranello <i>et al.</i> , 2012; Osman <i>et al.</i> , 2013
$H = 1.3 + e^{\frac{a+b}{DBH+1}}$	Peng, 1999; Silva Scaranello <i>et al.</i> , 2012
$H = 1.3 + a \cdot DBH/b + DBH$	Fang & Bailey, 1998; Ahmadi <i>et al.</i> , 2013
$H = 1.3 + a(1 - e^{-b.DBH})$	Huang <i>et al.</i> , 1992
$H = 1.3 + \frac{DBH^2}{(a + b \cdot DBH)^2}$	Huang <i>et al.</i> , 1992; Silva Scaranello <i>et al.</i> , 2012
$H = 1.3 + a \cdot e^{\frac{b}{DBH}}$	Peng, 1999; Petráš <i>et al.</i> , 2014
$H = 1.3 + 10^a DBH^b$	Huang <i>et al.</i> , 1992; Peng, 1999
$H = 1.3 + \frac{a \cdot DBH}{(DBH + 1)} + b \cdot DBH$	Huang <i>et al.</i> , 1992; Peng, 1999
$H = 1.3 + a \left( \frac{DBH}{(DBH + 1)} \right)^b$	Huang <i>et al.</i> , 1992
$H = 1.3 + e^{-a+b.DBH^c}$	Larsen & Hann, 1987; Wang & Hann, 1988; Colbert <i>et al.</i> , 2002
$H = 1.3 + a(1 - b \cdot e^{-c.DBH})$	Fang & Bailey, 1998; Batista <i>et al.</i> , 2001
$H = 1.3 + \frac{DBH^2}{(a + b \cdot DBH + c \cdot DBH^2)}$	Fang & Bailey, 1998; Peng, 1999
$H = 1.3 + a \cdot DBH^b \cdot DBH^{-c}$	Huang <i>et al.</i> , 1992; Fang & Bailey, 1998
$H = 1.3 + a(1 - b \cdot e^{-c.DBH})^d$	Huang <i>et al.</i> , 1992; Peng, 1999

Table 2.9 presents six best non-linear height-DBH models and their statistical performances developed for six groups of species mentioned above. Figure 2.9 shows the height curve depicted for the six height growth models.

**Table 2.9.** Six selected non-linear height-DBH models and their statistical performance.  $H_{max}$  and DBH stand for maximum height and diameter at breast height, respectively.

Species	Model	Parameters		RMSE (m)	$R^2$
		a	b		
<i>Fagus orientalis</i>	$H_{max} = 1.3 + a(1 - e^{-b.DBH})$	41.794	0.025	5.38	0.65
<i>Carpinus betulus</i>	$H_{max} = 1.3 + a.DBH/b + DBH$	33.039	14.772	3.90	0.48
<i>Quercus castanifolia</i>	$H_{max} = 1.3 + a(1 - e^{-b.DBH})$	39.574	0.035	5.74	0.30
<i>Alnus subcordata</i>	$H_{max} = 1.3 + a(1 - e^{-b.DBH})$	39.698	0.038	3.05	0.47
Group1	$H_{max} = 1.3 + DBH^2/(a + b.DBH)^2$	2.053	0.143	5.24	0.63
Group2	$H_{max} = 1.3 + a(1 - e^{-b.DBH})$	13.263	0.025	2.79	0.44
Group1: <i>Tilia begonifolia</i> , <i>Acer velutinum</i> , <i>Acer cappadocicum</i> , <i>Sorbus torminalis</i> and <i>Fraxinus excelsior</i> Group2: <i>Quercus macranthera</i> , <i>Carpinus orientalis</i> , <i>Parotia persica</i> and <i>Diospyros lotus</i>					



**Fig. 2.9.** Tree's Height vs. DBH and the best fitted line given in table 4-2; a) *Fagus orientalis*, b) *Carpinus betulus*, c) *Quercus castanifolia*, d) *Alnus subcordata*, e) Group1, and f) Group2

Next, the Lorey's height was calculated using equation 2.3. Lorey's height as a mean height of a stand weights the contribution of trees to the stand height by their basal area. Therefore, it is more stable than arithmetic height specifically in uneven-aged stands.

$$H_{\text{Lorey}} = \frac{\sum_{i=1}^n BA_i \times H_i}{\sum_{i=1}^n BA_i} = \frac{\sum_{i=1}^n DBH_i^2 \times H_i}{\sum_{i=1}^n DBH_i^2} \quad (2.3)$$

Where  $H_{\text{Lorey}}$ ,  $BA_i$ ,  $DBH_i$  and  $H_i$  are Lorey's height (m), basal area ( $\text{cm}^2$ ), diameter at breast height (cm), and height (m) of tree  $i$ , respectively, and  $n$  is total number of trees in each plot.

Volume is usually expressed quantitatively as a function of DBH and height (Macauley *et al.*, 2009; Namirianian, 2007). So the selected height-DBH relationships were next used to estimate the height of all trees. Local species level volume equations based on DBH and height developed by FRWO (Table 2.10) were also used to calculate per tree stem volume.

**Table 2.10.** Six volume models and their parameters (produced by FRWO)

Species	Model	Parameters		
		A	b	c
<i>Fagus orientalis</i>	$V = (a.DBH^2 \times H) + (b.DBH^2) + c$	0.000026364	0.000342059	0.0405
<i>Carpinus betulus</i>	$V = a(DBH^2 \times H)^b$	0.000023	1.0432	
<i>Quercus castanifolia</i>	$V = a(DBH^2 \times H)^b$	0.000058	0.9544	
<i>Alnus subcordata</i>	$V = a(DBH^2 \times H)^b$	0.000051	0.9595	
Group1	$V = a(DBH^2 \times H)^b$	0.000056	0.9539	
Group2	$V = a(DBH^2 \times H)^b$	0.000035	1.0058	

Table 2.11 summarizes statistics for 60 measured plots that were used for developing and validating the GLAS height and volume models. Moreover, distribution of number of trees (n/ha) and basal area ( $\text{m}^2/\text{ha}$ ) is seen in figure 2.10. The frequency distribution of the forest volume for 60 GLAS footprints is shown in figure 2.11.

**Table 2.11.** Statistical summary of maximum height ( $H_{\text{max}}$ ), Lorey height ( $H_{\text{Lorey}}$ ) and forest volume ( $V$ ) for 60 in situ plots; Min., Max. and Avg. stands for minimum, maximum and average value, respectively.

Number of plots	$H_{\text{max}}$ (m)			$H_{\text{Lorey}}$ (m)			$V$ ( $\text{m}^3/\text{ha}$ )		
	Min.	Max.	Avg.	Min.	Max.	Avg.	Min.	Max.	Avg.
60	5.3	52.3	36	5.1	36.2	27.4	0.69	996.56	457.44

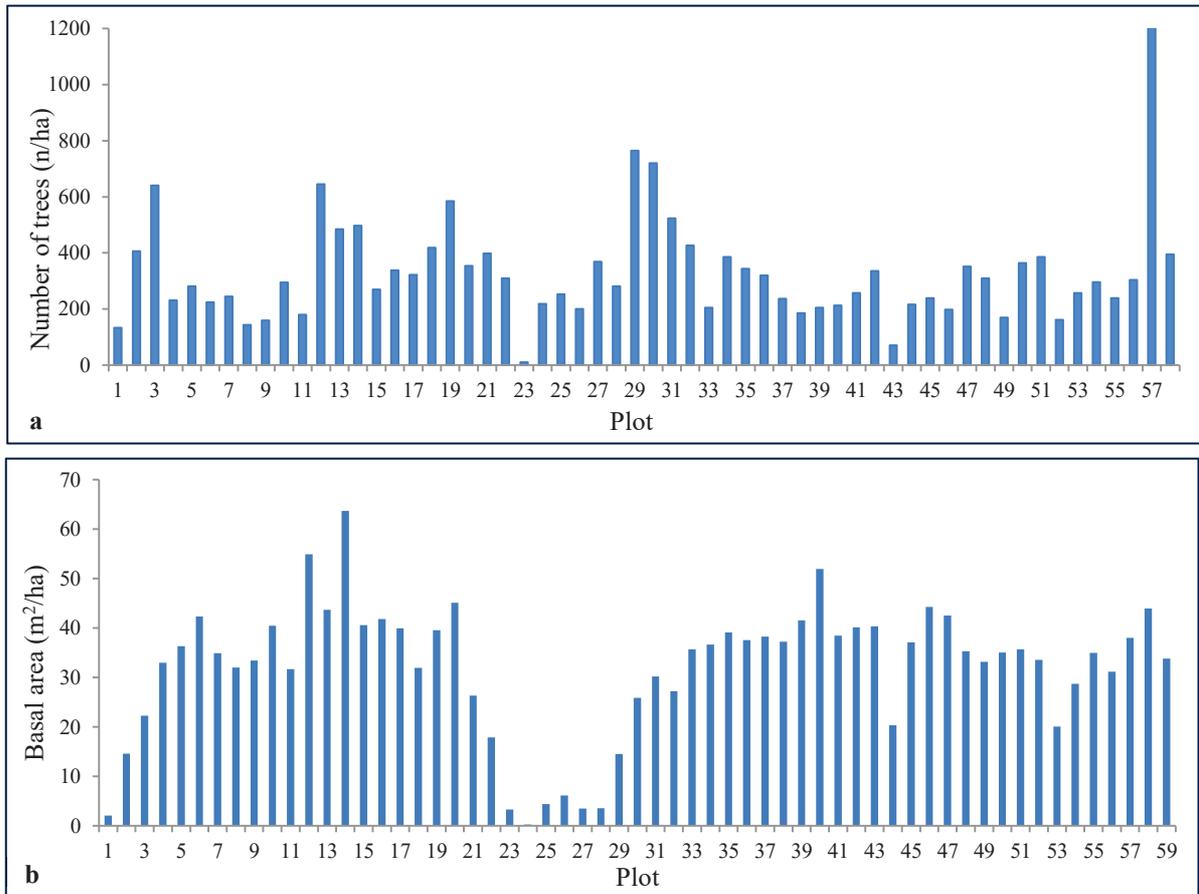


Fig. 2.10. a) Distribution of number of trees per hectare (n/ha), b) forest volume (m<sup>3</sup>/ha) in 60 reference plots

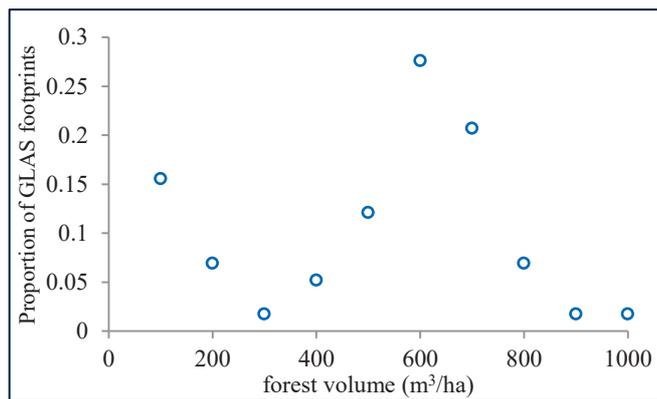


Fig. 2.11. Histogram of forest volume (m<sup>3</sup>/ha) for 60 reference plots

### **2.3.2. Extraction of terrain index, slope, aspect and elevation class maps from DEM**

Terrain Index (TI) was calculated at the location of lidar footprints using: 1) a fine resolution DEM<sub>10</sub> produced based on 1:25000 topographic maps (called TI<sub>10</sub>), and 2) SRTM DEM with 90 meter resolution (called TI<sub>90</sub>). The elevation range within a 7×7 neighborhood of 10m-DEM (Rosette *et al.*, 2008a; Chen, 2010b) and 3×3 neighborhood of 90m SRTM DEM (Baghdadi *et al.*, 2014) at location of each GLAS footprint was considered as TI. The effect of using higher resolution DEM on model performance was also investigated.

Terrain index, slope in degrees, aspect in nine classes (flat, north, north-east, east, south-east, south, south-west, west and north-west) and elevation classification map in six classes ( < 100, 100-300, 300-700, 700-1200, 1200-1600 and 1600 < ) were also produced using DEM<sub>10</sub> for the entire study area. Figure 2.12 shows DEM<sub>10</sub> and four extracted maps from it.

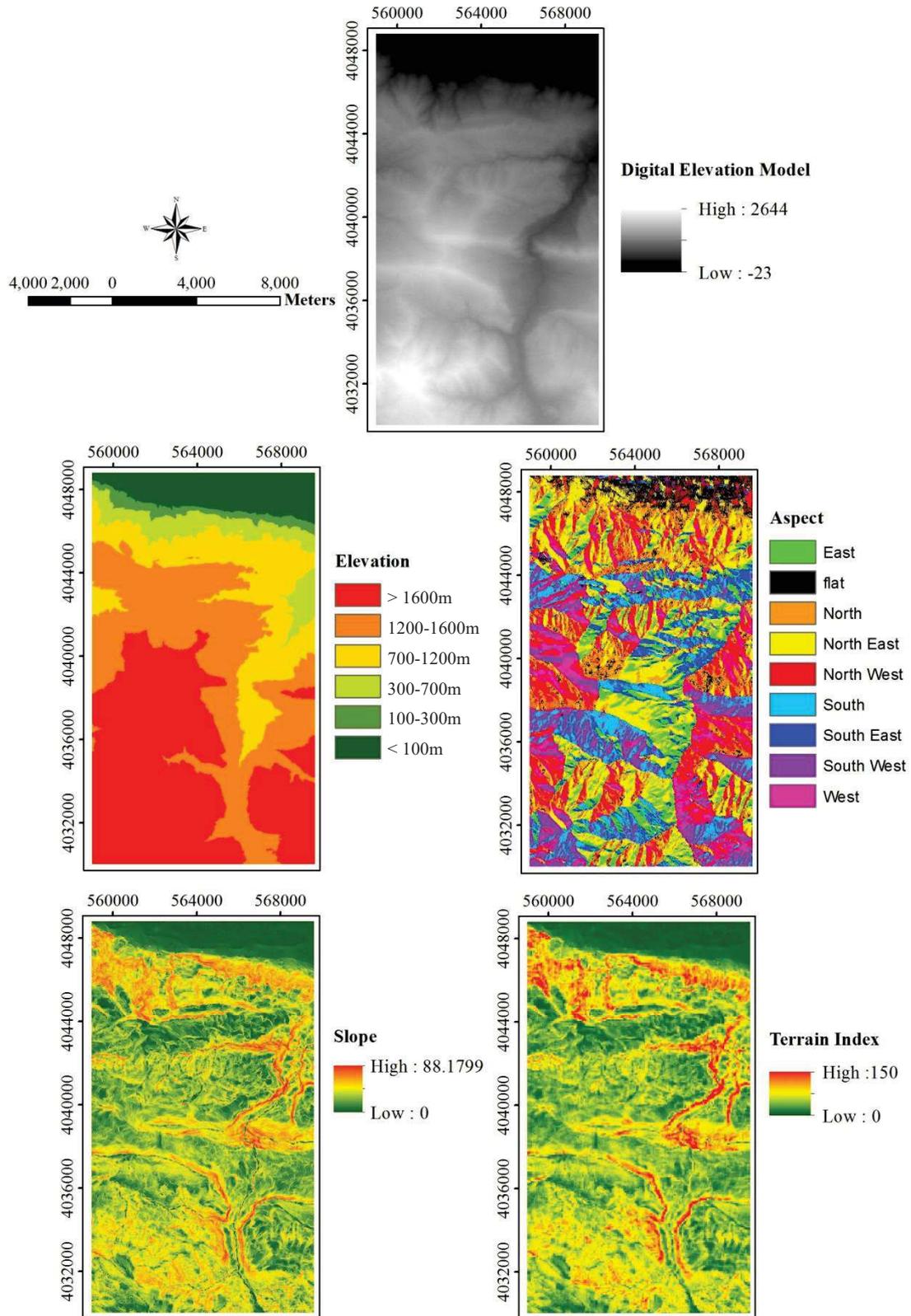


Fig. 2.12. Digital elevation model (DEM<sub>10</sub>) and four extracted maps (elevation and aspect are classified maps)

### 2.3.3. ICESat GLAS data processing and extraction of metrics

#### 2.3.3.1. Processing GLAS waveforms

The GLA01 (Global Altimetry data) and GLA14 (Global Land Surface Altimetry data) data products were converted from binary to the ASCII format using IDLreadGLAS provided by NSIDC. Required information such as latitude, longitude, elevation, centroid elevation, and fitted Gaussian peaks were extracted from GLA14 data, and raw waveforms were extracted from GLA01 data. The most important information derived from these products was summarized in table 2.12.

ICESat/GLAS elevations are referenced to the TOPEX/Poseidon ellipsoid which is 70 cm smaller than WGS84 ellipsoid. For comparison between ICESat, SRTM DEM and in situ data, datasets needed to be available in the same coordinate system. So, GLAS elevations were transformed to the WGS84 ellipsoid by adding 70 cm.

**Table 2.12.** The main information derived from GLA01 and GLA14

#	Flag	product	Description
1	<u>i_rec_ndx</u>	GLA01 GLA14	GLAS Record Index: GLAS is recording 40 waveforms per second; each 40-waveform bunch has a record index
2	<u>i_shot_ctr</u>	GLA01	Shot Counter
3	<u>i_rng_wf</u>	GLA01	The 1064 nm echo waveform digitizer sample output (0-255)
4	<u>i_lat</u>	GLA14	The geodetic latitude of the forty laser spots in the 1 second time frame in micro degree
5	<u>i_lon</u>	GLA14	The longitude of the forty laser spots in the 1 second time frame in micro degree
6	<u>i_elev</u>	GLA14	Surface elevation with respect to the ellipsoid at the spot location determined in mm
7	<u>i_campaign</u>	GLA14	The campaign. i.e.: for campaign L3K, it will be "3K".
8	<u>i_Gamp</u>	GLA14	Amplitude of each Gaussian solved for (up to six) in 0.01 volts
9	<u>i_Garea</u>	GLA14	Area under each of the Gaussians solved for (up to six) in 0.01 volts × ns
10	<u>i_Gsigma</u>	GLA14	Width (sigma) of each Gaussian solved for (up to six) in 0.001ns
11	<u>i_satNdx</u>	GLA14	The count of the number of gates in a waveform which have an amplitude greater than or equal to saturation index threshold
12	<u>i_FRir_qaFlag</u>	GLA14	Indicates the presence of clouds (0-15); 15 = no cloud, 14=likely presence of low clouds, etc.
13	<u>i_4nsBgMean</u>	GLA01	Background Noise Mean Value in 0.01 counts
14	<u>i_4nsBgSDEV</u>	GLA01	The standard deviation of the background noise in 0.01 counts
15	<u>i_maxRecAmp</u>	GLA14	Maximum Amplitude of the Received Echo

Some pre-processes were applied to remove inappropriate and useless waveforms (Chen, 2010b; Hilbert & Schmullius, 2012; Baghdadi *et al.*, 2014):

- 1- Eliminating waveforms affected by cloud; Flag `i_FRir_qaFlag` in GLA14 data indicates the estimated atmospheric conditions over each GLAS footprint using a cloud detection algorithm. To do this, waveforms with `i_FRir_qaFlag=15` were kept and the rest were removed (Chen, 2010b; Duncanson, 2010).
- 2- Eliminating saturated waveforms; `i_satNdx` in GLA14 presents the count of the number of gates in a waveform which have an amplitude greater than or equal to saturation index threshold (`i_satNdxTh`). So only waveforms with `i_satNdx=0` were used for analysis in this study (Chen, 2010b; Hilbert & Schmullius, 2012).
- 3- Removing noisy waveforms with a signal to noise ratio (SNR) lower than 15 (Baghdadi *et al.*, 2014); To calculate SNR, maximum energy of samples from GLA01 (`i_maxRecAmp`) was divided to standard deviation of the background noise (`i_4nsBgSDEV`) recorded in GLA14 data.
- 4- Removing waveforms in which difference between centroid elevation (`i_elev` from GLA14) and corresponding SRTM DEM is greater than 100 meters ( $|\text{ICESat} - \text{SRTM}| > 100$ ) (Baghdadi *et al.*, 2014).

### 2.3.3.2. Waveform metrics extraction

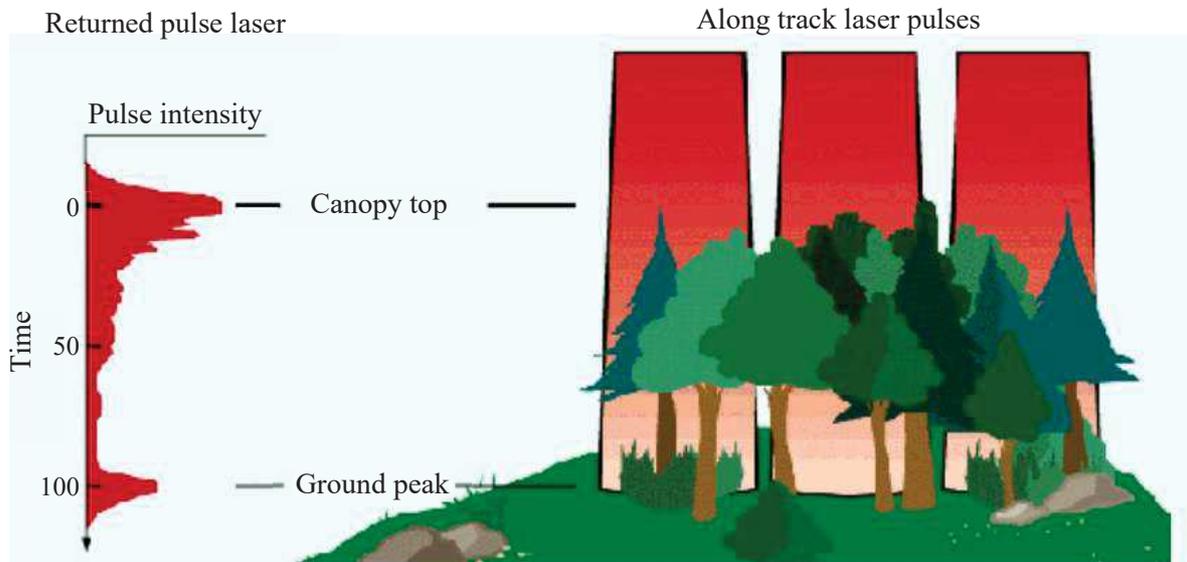
GLAS provides a full waveform of illuminated objects on the surface. Each waveform is a function of vertical structure of illuminated surface and their reflection properties within the footprint. Figure 2.13 illustrates a schematic of transmitted pulses over a vegetated area and a returned waveform. As mentioned before, 544 or 1000 bins (1 bin= 1 ns=15 cm) is recorded for each received waveform. Signal start and end are defined as first and last bins in the waveform where the waveform intensity exceeds background noise threshold,  $n\sigma + \mu$ , where  $\sigma$  and  $\mu$  recorded in GLA01 product are standard deviation and mean background noise respectively, and  $n=0.5, 1, \dots, 5$ . Different thresholds including  $3\sigma + \mu$  (Sun *et al.*, 2008),  $4\sigma + \mu$  (Lefsky *et al.*, 2005),  $4.5\sigma + \mu$  (Baghdadi *et al.*, 2014; Lefsky *et al.*, 2007; Lee *et al.*, 2011) were applied in previous studies. Chen (2010b) tested different thresholds for signal start and end for each three sites from  $2.5\sigma + \mu$  to  $5\sigma + \mu$ . He found that the value of  $n$  for optimal threshold is 3.5 for signal start and 5 for signal end. Therefore, an average value of 4.5 was used in this study. Hilbert & Schmullius (2012) stated that the optimal thresholds might differ according to the waveform types, laser periods or footprint structure. In this research the threshold was set to  $4.5\sigma + \mu$ , the optimum threshold used in the most studies.

The vertical distance between signal start and signal end of a waveform was computed as

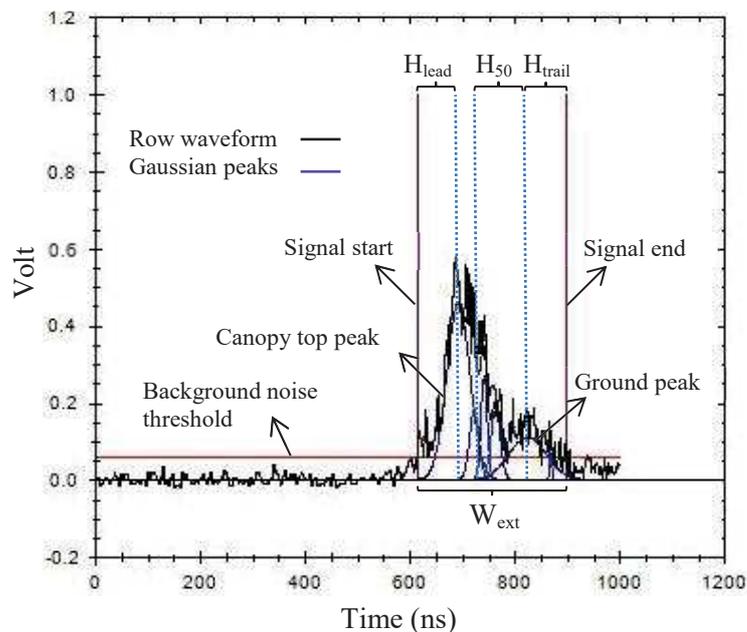
waveform extent ( $W_{\text{ext}}$ ) which could be affected by terrain slope, canopy height and canopy density (Yang *et al.*, 2011).

To identify ground peak as an important feature of waveforms to extract maximum height over flat area (defined as vertical difference between signal start and ground peak) or derive some metrics in waveforms, Gaussian components (up to 6 peaks) are fitted to the raw waveform (Brenner *et al.*, 2003). Duong *et al.* (2009), Sun *et al.* (2008) and Xing *et al.* (2010) have considered the last peak as ground peak. But in dense vegetated area, ground peak may have lower amplitude than adjacent peaks. In this case, it was suggested to consider the peak with highest amplitude as the ground peak even if it is not the last peak (Boudreau *et al.*, 2008). Rosette *et al.* (2008b) considered the centroid of the Gaussian peak with greatest amplitude out of the last two peaks as ground surface. Hilbert & Schullius (2012) applied two approaches to find the ground return; the first was the maximum of the last two Gaussian peaks of a waveform, and the second approach was identifying ground return based on local maxima within the GLA01 waveform. They found both approaches represented the terrain accurately and could be used as basis to calculate the tree height. However the usage of the original waveform (GLA01) is more flexible and meets the ground return more precisely. Chen (2010b) found the stronger peak of the two last one is a better representative of the ground elevation in the studied coniferous site, whereas for the studied woodland site the strongest peak of the five last peaks matched well with the ground elevation. Iqbal (2010) detected the peak with maximum amplitude between signal start and end of a waveform. Then by moving from this point to the signal end, the peak with amplitude one-fifth the maximum amplitude was considered as ground return. If this condition was not met, the peak with maximum amplitude was assigned as ground return. In the current research, we followed last achievements (Baghdadi *et al.*, 2014; Fayad *et al.*, 2014; Rosette *et al.*, 2008a; Chen, 2010b) and we chose the stronger one among two last Gaussian peaks as ground peak. The first Gaussian peak was selected as canopy top. The distance between ground peak and signal start has been defined as maximum canopy height in flat area. The vertical distance from ground peak to signal end and from canopy top to signal start was considered as trail edge and lead edge extents, respectively (Baghdadi *et al.*, 2014; Yang *et al.*, 2011).  $H_{25}$ ,  $H_{50}$ ,  $H_{75}$  and  $H_{100}$  as quartile heights have been extracted from waveforms by calculating the vertical distance between ground peak and position of waveform at which respectively 25%, 50%, 75% and 100% of the returned energy between signal start and end occurs (Nelson *et al.*, 2009; Sun *et al.*, 2008). So the total waveform energy was calculated by summing all the return energies from signal start to end. Starting from the signal end, the position of the 25%, 50%, and 75% of energy were located by comparing the accumulated energy with total energy.  $H_{100}$

is the maximum canopy height as defined above. Figure 2.14 illustrates a GLAS waveform from study area with Gaussian peaks and some extracted metrics. The metrics extracted from GLAS waveforms and their derivatives, used in this research, are listed in table 2.13.



**Fig. 2.13.** Schematic of laser altimeter pulse spreading over a vegetated area, and a returned waveform (NASA, 2015)



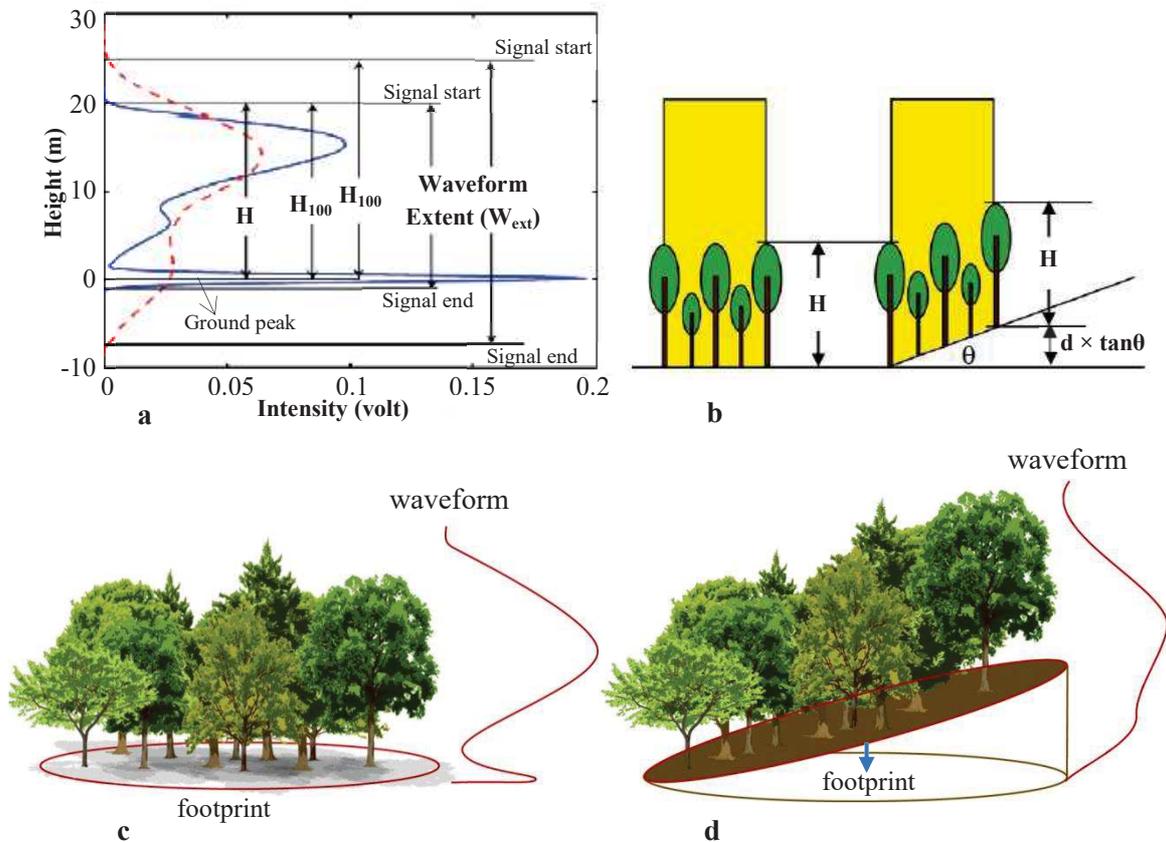
**Fig. 2.14.** GLAS waveform and some metrics over a terrain of 25% slope in the study area. 1ns corresponds to 15cm sampling distance in waveform.

**Table 2.13.** Definition of metrics extracted from GLAS waveforms

Metrics	Definition
$W_{\text{ext}}^n$ , $\ln(W_{\text{ext}})$ , $\exp(W_{\text{ext}})$	Waveform extent
$H_{\text{lead}}^n$ , $\ln(H_{\text{lead}})$ , $\exp(H_{\text{lead}})$	Height of lead edge extent
$H_{\text{trail}}^n$ , $\ln(H_{\text{trail}})$ , $\exp(H_{\text{trail}})$	Height of trail edge extent
$H_{25}^n$ , $\ln(H_{25})$ , $\exp(H_{25})$	Height at which 25% of the returned energy occurs
$H_{50}^n$ , $\ln(H_{50})$ , $\exp(H_{50})$	Height at which 50% of the returned energy occurs
$H_{75}^n$ , $\ln(H_{75})$ , $\exp(H_{75})$	Height at which 75% of the returned energy occurs
$H_{100}^n$ , $\ln(H_{100})$ , $\exp(H_{100})$	Height at which 100% of the returned energy occurs
<b>In:</b> natural logarithm (the logarithm to the base $e=2.718$ ), power $n=0, 0.5, \dots, 3$	

Over mountainous areas with large relief and complex terrain, the peaks from ground and surface objects can be broadened and signals returned from ground and vegetation be mixed. It makes that difficult to identify the ground elevation, and subsequently calculation of metrics which are dependent on the ground location (Chen, 2010b; Lefsky *et al.*, 2005; Zwally *et al.*, 2002; Lee *et al.*, 2011). To clarify, the two most commonly used height metrics to infer vegetation height from lidar are waveform extent (vertical distance between signal start and end) and  $H_{100}$  (vertical distance between signal start and ground peak) (Fig. 2.15a). These metrics would be influenced by different factors such as surface topography, footprint size, forest density and laser pulse energy.

Over flat area  $H_{100}$  equals to the canopy height ( $H$ ), but over sloped terrain it may increase to " $H + d \times \tan\theta$ " (Lee *et al.*, 2011) where  $d$  is lidar footprint diameter and  $\theta$  is the slope angle (Fig. 2.15b). Broadening effect of terrain slope on waveform extent and also Gaussian peaks have been illustrated also in figure 2.15 (b,c,d).



**Fig. 2.15.** a) A schematic illustration of lidar derived vegetation height ( $H$ ),  $H_{100}$ , and waveform extent ( $W_{ext}$ ) for one waveform over flat terrain (solid line) and another over sloped terrain (dash line) (Lee *et al.*, 2011). b) Impact of slope on lidar height retrieval (Lee *et al.*, 2011). c,d) A schematic of footprint located over flat (left) and sloped (right) vegetated terrain, respectively.

In addition to deterministic heuristics (user defined metrics), a non-parametric statistical technique named principal component analysis (PCA) was used to remove noises and reduce dimensionality of waveform signals.

In short, PCA finds a set of synthetic variables (the principal components) that summarizes the original set. It rotates the axis of variation to give a new set of ordered orthogonal axis that summarizes describing proportions of the variations. In fact, the principal components (PCs) are uncorrelated and ordered such that the  $k^{\text{th}}$  PC has the  $k^{\text{th}}$  largest variance among all PCs (Ulfarsson, 2007). The traditional approach is to use the first few PCs in data analysis since they have most of the variation in the original data set.

In this study, lidar signal intensities were used for the PCA analysis. In order to apply PCA, it is necessary to have equal number of samples in all waveforms. So, the length of largest waveform extent was considered as basis (400 samples) and other waveforms were apart from signal start toward signal end till the number of samples reach the base  $W_{ext}$

samples. Since the number of observations (60) is less than the number of samples in the useful part of waveforms (400 samples), it was aimed to reduce the number of samples by selecting one among each ten samples. So PCA was performed using 41 samples as variables using package “FactoMineR” in R (Husson, *et al.*, 2015) to find the main factors (waveform signals) determining most effects on forest canopy height. This procedure of data reduction was also used by Fayad *et al.* (2014) to estimate canopy height using ICESat GLAS data. As it is seen in figure 2.16, three first components had the most information, and explained 77.5% of variance in the data.

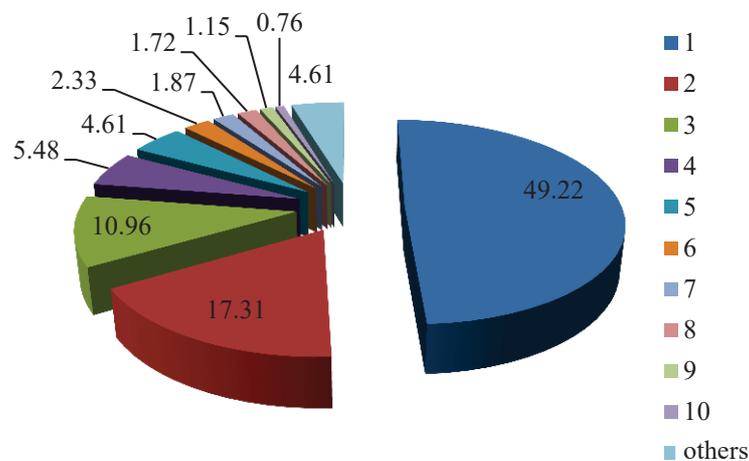


Fig. 2.16. Information explained by ten first PCs and other PCs

In overall, two sets of metrics were extracted from waveform; one, deterministic heuristics representing vertical distance between different positions of waveform and ground peak (represented in Table 2.13), and one non-parametric metrics including principal components produced from PCA. In this dissertation, these metrics are addressed as “waveform metrics” and “PCs”, respectively, however PCs are also metrics derived from waveforms.

#### 2.3.4. PALSAR data processing and extraction of metrics

PALSAR mosaic data which were used in this study have been processed by JAXA using a long mosaicking algorithm described by Shimada & Ohtaki (2010). It includes orthorectification, slope correction and intensity tuning between neighboring strips. Absolute radiometric calibration were performed to drive sigma naught and gamma naught (backscatter coefficient) from each polarization of PALSAR mosaic data (HH and HV), using equations 2.4 and 2.5 (Shimada & Ohtaki, 2010; Mermoz *et al.*, 2014). Equations 2.6 and 2.7 were also used to convert them to the linear scale.

$$\gamma_{dB} = 10 \times \log_{10}(DN^2) + CF \quad (2.4)$$

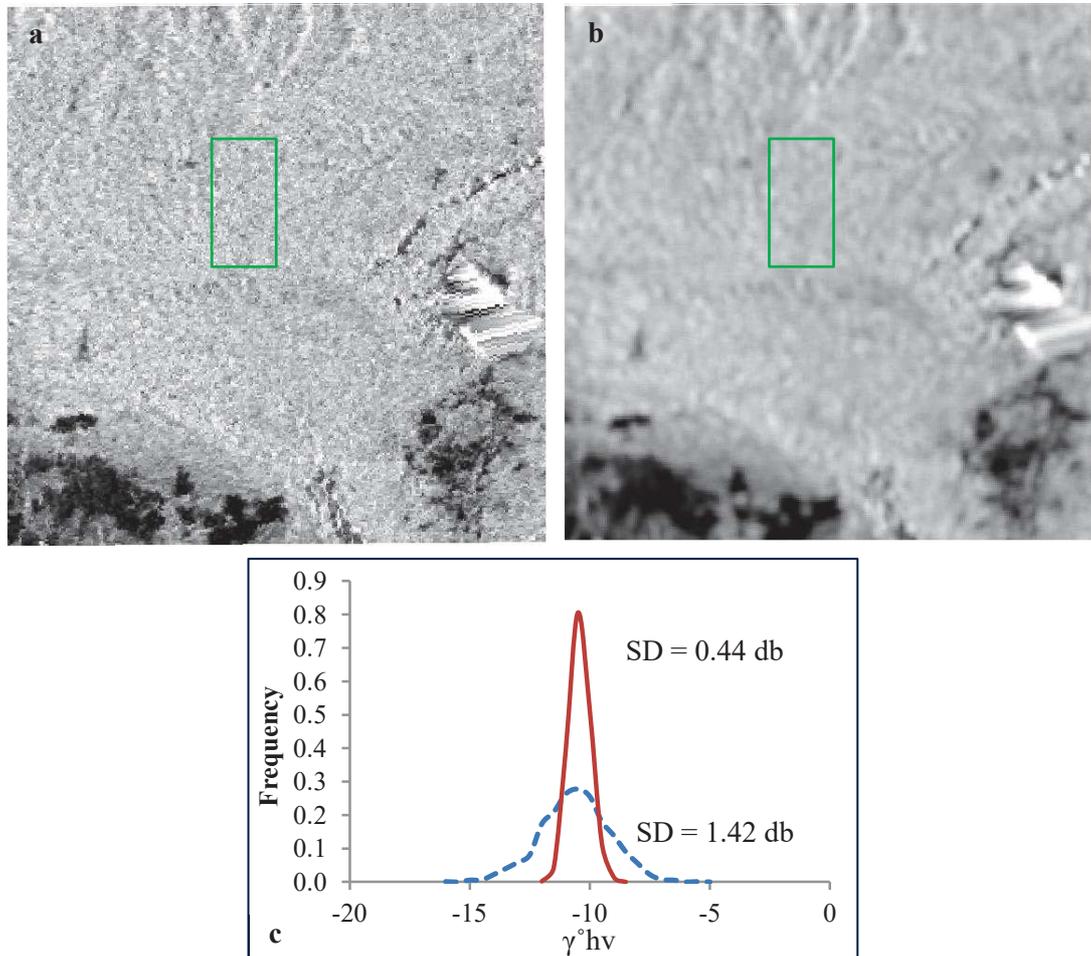
$$\sigma_{dB} = \gamma_{dB} + 10 \times \log_{10}(\cos \theta) \quad (2.5)$$

$$\gamma_l = 10^{(\gamma_{dB}/10)} \quad (2.6)$$

$$\sigma_l = \gamma_l \times \cos \theta \quad (2.7)$$

Where  $\gamma_{dB}$  and  $\sigma_{dB}$  stands for gamma and sigma naught in dB, respectively.  $\gamma_l$  and  $\sigma_l$  stands for gamma and sigma in linear scale, respectively. The  $\theta$  is local incidence angle, DN stands for digital number which is pixel intensity value and CF is constant calibration factor equal to -83 dB.

Speckle noise as defined by Gagnon & Jouan (1997), is a common phenomenon in all coherent imaging systems like laser and SAR imagery. The source of this noise is attributed to random interference between the coherent returns issued from the numerous scatters present on a surface, on the scale of a cell resolution. Speckle noise is often an undesirable effect, and so, speckle filtering turns out to be a critical pre-processing step for detection/classification optimization. There are many speckle reduction techniques. In this study “LEE” as a widely used filter and then “multitemporal” filter (four dates, 2007 to 2010, at HH and HV polarizations) were applied on the data to reduce speckle effects (Attarchi & Gloaguen, 2014; Mermoz *et al.*, 2014). A 5 by 5 window was used for performing LEE filter on  $\sigma_l$  and  $\gamma_l$  in “nest” software. Then multitemporal filter was performed on four images from 2007-2010. Quantitatively the distribution of the backscatter of a homogenous area is reduced significantly after these filters. An illustration of speckle reduction is shown in figure 2.17a,b. The histogram of a forested area before and after filter is seen in figure 2.17c. The standard deviation of the backscatter coefficient (here gamma naught which is shown by “ $\gamma$ ” hereinafter) is decreased from 1.42 to 0.44 decibel.



**Fig. 2.17.** Gamma naught ( $\gamma^{\circ} = \gamma_{\text{dB}}$ ) before (a) and after (b) LEE and multitemporal filters, and the histogram of backscatters over a forested area (green rectangle) (c) before (blue) and after (red) noise reduction. The standard deviation of  $\gamma^{\circ}$  decreases from 1.42 to 0.44 db.

Image texture which is defined as variation of image tones that are related to the spatial distribution of forest vegetation (Roberts *et al.*, 2007) has proved to be capable of identifying different aspects of forest stand structure. Kayitakire *et al.* (2006) showed that textural indices derived from a Grey-level co-occurrence matrices (GLCM) of an IKONOS-2 image are well correlated to the forest structural variables such as age, crown circumference, tree height, stand density and basal area. Nichol & Sarker (2011) and Attarchi & Gloaguen (2014) resulted in heterogeneous forests, texture measures are more sensitive to the canopy structure than spectral reflectance, and are more correlated to forest AGB. Trinder *et al.* (2013) also indicated that the textural-based models are significantly more efficient than spectral-based models for predicting lidar metrics. The texture measures are able to identify significant differences in image texture independently of image contrast (i.e. backscatter). It also increases the saturation threshold and the biomass range that can be measured (Kuplich *et al.*, 2005; Sarker *et al.*, 2012, Cutler *et al.*, 2012).

Hence, in this study GLCM measures were derived from both HH and HV polarization. The GLCM characterizes the texture of an image by calculating how often pairs of pixel with specific values and in a specified spatial relationship occur in an image and then statistical measures are extracted from this matrix. Eight GLCM indices such as mean, variance, homogeneity, contrast, dissimilarity, second moment, entropy and correlation were calculated from this matrix. A window size of  $3 \times 3$ ,  $5 \times 5$ ,  $7 \times 7$ ,  $9 \times 9$  and  $11 \times 11$  pixels with horizontal and vertical offset of one was used for extraction of texture indices. However texture characterization is influenced by window size, it is not easy to select a superior window because an optimal window depends on different characteristics of forest and GLCM attributes are affected by the window size in different ways (Ouma & Tateishi, 2006; Trinder *et al.*, 2013). Consequently, the mean value of each feature from multitemporal data (from 2007 to 2010) was calculated and used. In order to match PALSAR data, optical images and DEM spatially, the produced maps or indices were resampled to 10 m resolution. Figure 2.18 shows mean values of multitemporal data for eight statistical texture features extracted from HH band using GLCM algorithm. As it is seen, variation of “contrast” and “variance” over study area is low. There is no or very low correlation between canopy height/volume and most extracted features at the location of in situ plots. The “mean” and “correlation” showed highest correlation with canopy height/volume. Despite of low correlation of these features with considered parameters, individually, combination of these measures with other variables (extracted from other sources of data) will be considered in the height estimation.

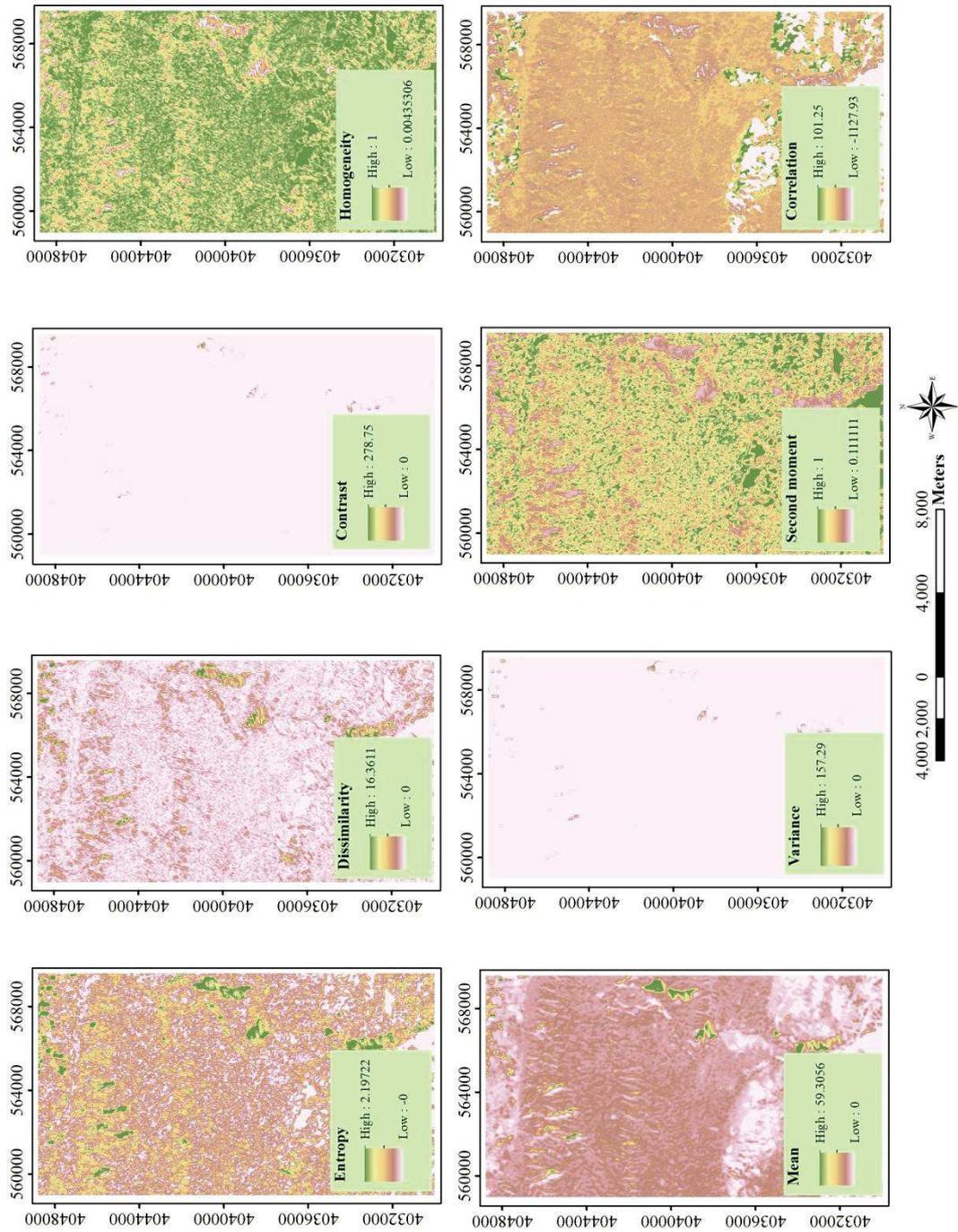


Fig. 2.18. Eight GLCM texture features extracted from PALSAR-HH

### 2.3.5. Optical data (Landsat-TM, SPOT-5) and extraction of vegetation and texture indices

Several studies have shown relationship between forest structure and vegetation indices. Freitas *et al.* (2005) evaluated relationships between forest structure (frequency of multiple-stemmed trees, density of trees, mean and range of tree diameter, mean and range of tree height and average of basal area) and vegetation indices including NDVI (normalized difference vegetation index) and MVI (moisture vegetation index) extracted from Landsat7-ETM+ images in Atlantic rainforest fragments, in southeastern Brazil. They resulted that MVI outperformed in dense humid forests, whereas NDVI is a good indicator of green biomass in deciduous and dry forests. They observed a weaker saturation effect and a higher sensitivity to MVI rather than NDVI over dense canopies in the Atlantic rainforest. Pascual *et al.* (2010) have reported high correlation between NDVI and MVI extracted from Landsat-ETM+ and mean and median lidar derived heights ( $R > 0.6$ ) in pine forests of the Fuenfria Valley in central Spain. Nichol & Sarker (2011) estimated forest biomass based on simple ratio vegetation index (RVI) derived from AVNIR-2 and SPOT-5 with  $R^2$  of 0.59 and 0.39, respectively. They observed a significant improvement in biomass estimation with an  $R^2$  of 0.739 obtained from the combined use of RVI of both sensors.

In this study three vegetation indices including NDVI, MVI and RVI were extracted from Landsat-TM and SPOT5 multispectral bands (Equations 2.8 to 2.10).

$$\text{NDVI} = (\text{NIR} - \text{R}) / (\text{NIR} + \text{R}) \quad (2.8)$$

$$\text{MVI} = (\text{NIR} - \text{MIR}) / (\text{NIR} + \text{MIR}) \quad (2.9)$$

$$\text{RVI} = \text{NIR} / \text{Red} \quad (2.10)$$

Several other maps were also produced using multitemporal data of vegetation indices after resampling TM derived indices to 10 m resolution: 1) minimum, maximum and mean values of NDVIs, MVIs and RVIs (So-called min-ndvi, max-ndvi, mean-ndvi, etc.), and also mean-summer and mean-winter values of each index; 2) eight statistical features including “mean”, “variance”, “homogeneity”, “contrast”, “dissimilarity”, “second moment”, “entropy” and “correlation” derived from GLCM texture analysis on mean NDVI map using  $3 \times 3$ ,  $5 \times 5$ ,  $7 \times 7$ ,  $9 \times 9$  and  $11 \times 11$  window size. Figures 2.16 shows mean NDVI and eight texture features derived from it using GLCM algorithm, respectively. In contrast to the texture features extracted from PALSAR data, almost all texture features from NDVI have moderate to good correlation with forest canopy height/volume. Correlation between “mean” and Lorey’s height, for instance, was 0.68. The “correlation” is the only feature not correlated with  $H_{\text{max}}$  and  $H_{\text{Lorey}}$ . As it is seen in the figure 2.19, the variation of “correlation” values over forested area is very low.

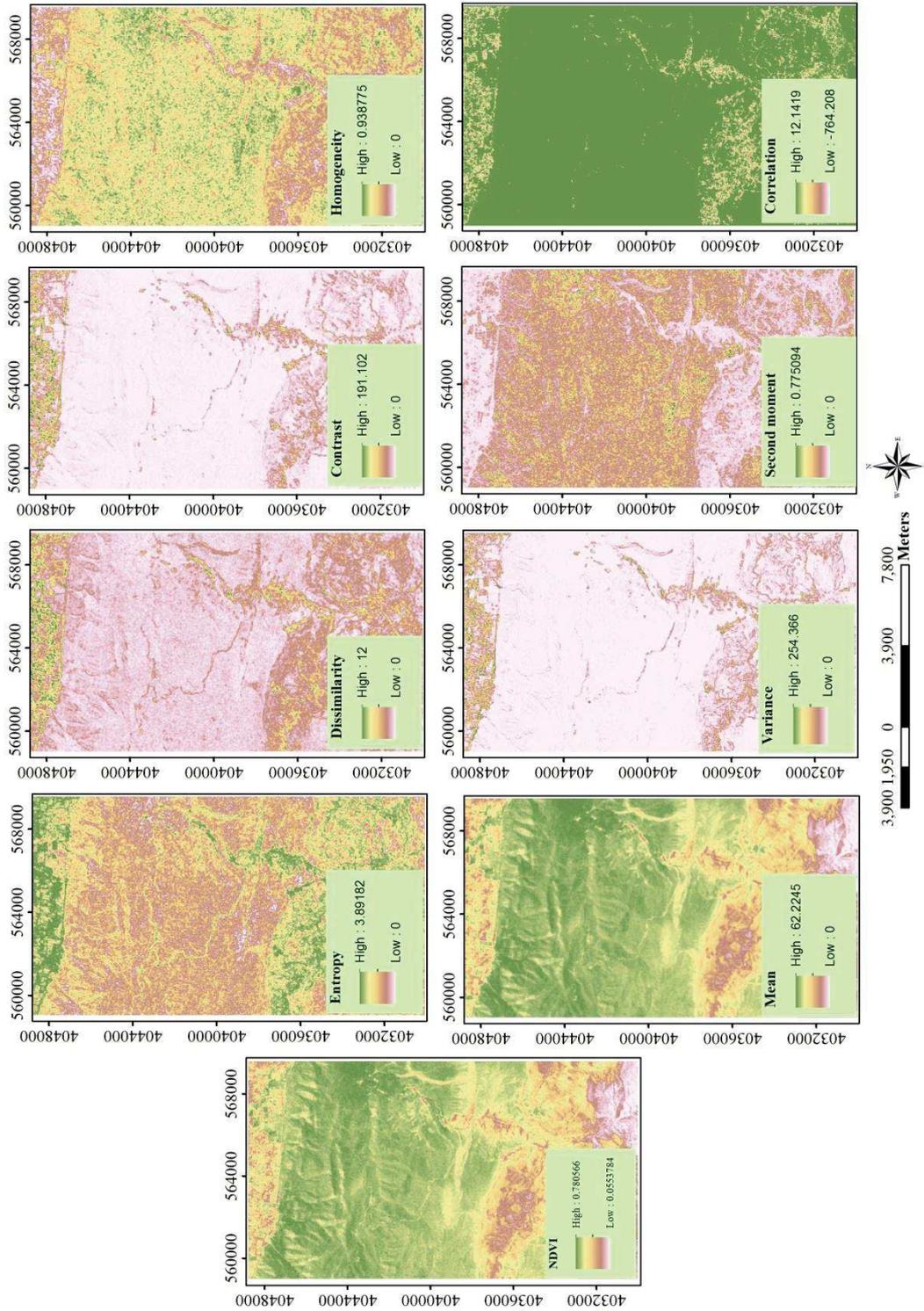
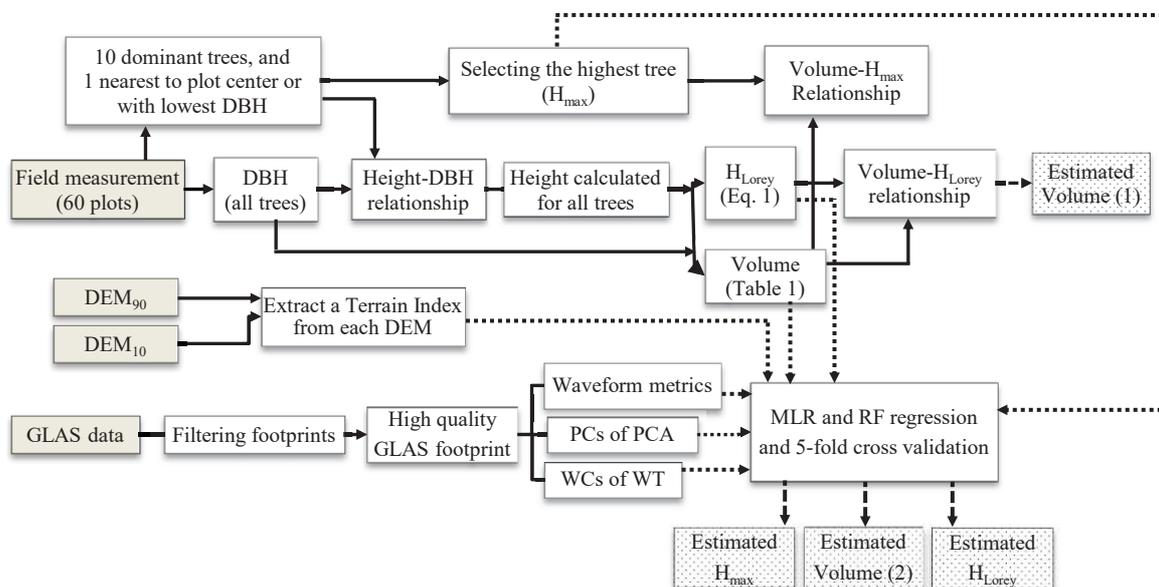


Fig. 2.19. NDVI and eight GLCM texture features extracted from it

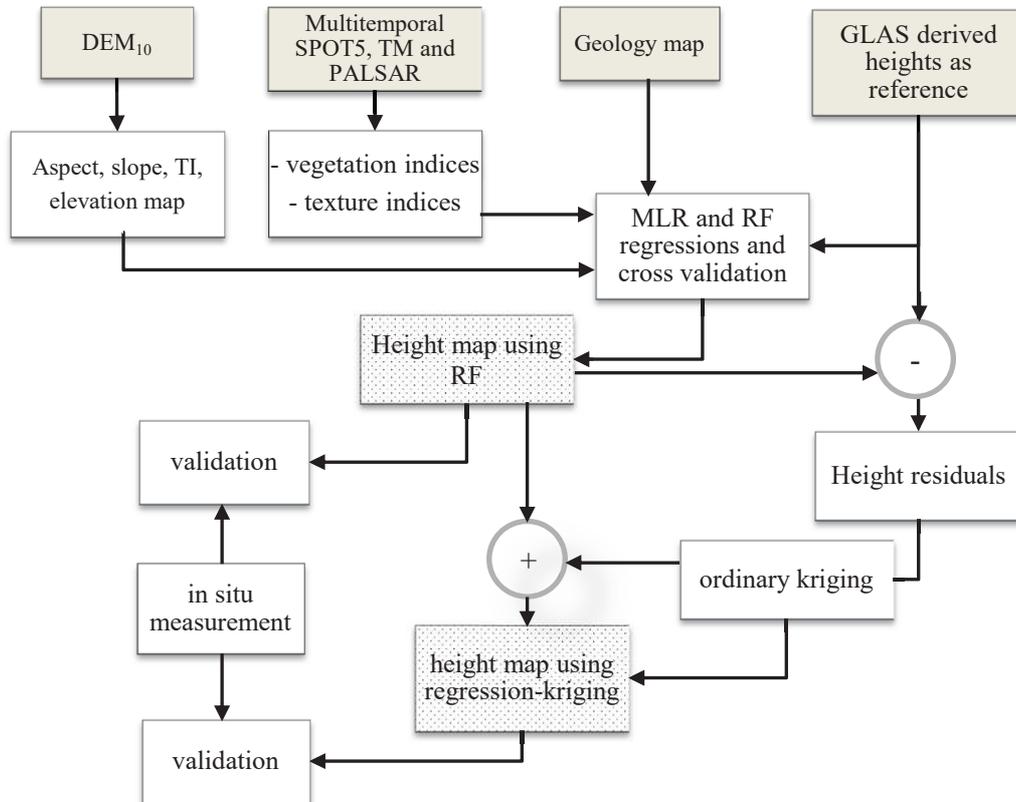
### 3. Research methodology

According to the objectives of this research, the entire process was illustrated in two flowcharts. The flowchart of forest canopy height and volume estimation using GLAS is displayed in figure 3.1 Gray boxes show origin input data, simple white boxes present data preparation processes and dot boxes indicate outputs. Solid lines and arrows indicate intermediate phases of data processing, dot arrows represent forest biophysical parameters ( $H_{\max}$ ,  $H_{\text{Lorey}}$  and Volume) and predictor variables entered in the regressions and finally dashed arrows address final outputs. In the presented flowchart, two general parts are observed; collection of in situ measurement and the related calculations, and GLAS data processing and analysis.

Figure 3.2 shows the process of producing forest canopy height map using combination of GLAS, PALSAR, optical images and environmental data. In this flowchart also, gray boxes show input data, simple white boxes present data preparation processes and dot boxes indicate outputs. More details are described in the following sections of this chapter.



**Fig. 3.1.** Overview of forest canopy height and volume estimation using GLAS data



**Fig. 3.2.** Overview of forest canopy height map using combination of GLAS, PALSAR, optical images and environmental data

### 3.1. Estimation of maximum ( $H_{\max}$ ) and Lorey's heights ( $H_{\text{Lorey}}$ ) using ICESat GLAS

#### 3.1.1. Direct method for estimating $H_{\max}$

Over flat area, estimation of maximum canopy height ( $H_{\max}$ ) is based on vertical difference between the waveform signal start ( $S_s$ ) and the ground peak ( $G_p$ ) (Neuenschwander *et al.*, 2008; Chen, 2010b), and it is calculated using equation 3.1. Vertical resolution of waveforms is 15 cm for GLAS data (Harding & Carabajal, 2005).

$$H_{\max}(\text{in m}) = (G_p - S_s) \times 0.15 \quad (3.1)$$

As it was described in section 2.3.3.2, over non-flat terrain the width of waveform and Gaussian peaks increases by effect of surface roughness. So the distance between signal start and ground peak will not represent the canopy height.

### 3.1.2. Parametric and non-parametric methods for prediction $H_{\max}$ and $H_{\text{Lorey}}$

Over sloped terrain, peaks from ground and surface objects can be broadened and mixed, making identification of ground peak difficult (Lefsky *et al.*, 2005; Pang *et al.*, 2008; Chen, 2010b). Hence it is necessary to find a way to decrease slope impact on waveform. Lefsky *et al.* (2005) and Chen (2010b) used DEM to include topography effects on height estimations. Some researchers derived terrain information from the waveform itself based on metrics such as leading and trailing edge extent ( $H_{\text{lead}}$  and  $H_{\text{trail}}$ , respectively). Lefsky *et al.*, 2007 and Pang *et al.*, 2008 used multiple transforms of the leading and trailing edge extents to model the  $H_{\text{lead}}$  and  $H_{\text{trail}}$  correction factor. In the next step, height correction factor (cf), defined as difference between waveform extent ( $W_{\text{ext}}$ ) and mean tree height ( $H_{\text{mean}}$ ), was regressed against  $H_{\text{lead}}$  and  $H_{\text{trail}}$  correction factor. Then it was used to estimate mean tree height ( $H_{\text{mean}} = W_{\text{ext}} - \text{cf}$ ). In present research, Terrain Index (TI) extracted from two sources of DEM (section 2.3.2) was entered as a predictor to consider the topography condition in the models. The effect of using finer resolution DEM ( $\text{DEM}_{10}$  in replace of  $\text{DEM}_{90}$ ) on performance of models was also investigated.

A large number of multiple linear regression (MLR), Random Forest (RF) and Artificial Neural Network (ANN) models were developed in R, employing different combination of metrics extracted from waveforms (“waveform metrics” and “PCs”) and also TI.

Random forests, as an ensemble learning method developed by Breiman in 2001, operate by constructing a multitude of regression trees (Breiman, 1994). Each tree in the forest is made of a random subset of observations with replacement and also a random of explanatory variables. So, two important parameters in random forest are the number of trees in the forest and the number of variables in the random subset at each node of tree. Prediction of new set of data for regression application would be the average prediction of all trees (Breiman, 2001; Liaw & Wiener, 2002). Therefore, the random forests algorithm (for both classification and regression) contains following steps:

1. Draw  $n_{\text{tree}}$  bootstrap samples from the original data.
2. For each of the bootstrap samples, grow an unpruned classification or regression tree. So that at each node, the best split is chosen among randomly selected variables.
3. Predict new data by aggregating the predictions of the  $n_{\text{tree}}$  trees (i.e., majority votes for classification, average for regression).

Figure 3.3 illustrates a schematic figure explaining how random forest works.

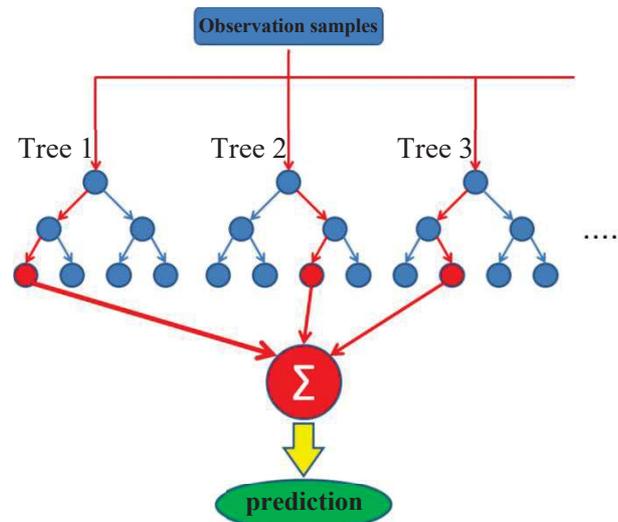


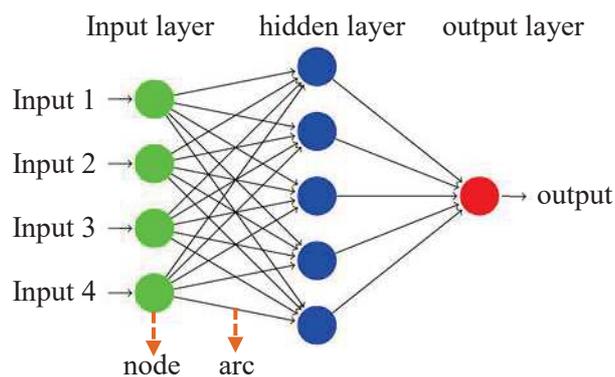
Fig. 3.3. A schematic of random forest regression

Through random sampling of observations, about one-third of them are not used for any individual tree that is called out of the bag, “OOB”, for that tree. The accuracy of a random forest’s prediction can be estimated from these OOB data (Breiman, 2001, Grömping, 2009). It would be possible to calculate variable importance by determining how much worse would be the OOB predictions, if the data for that variable are randomly permuted (Liaw & Wiener, 2002; Wei *et al.*, 2010). In fact, it would be possible to find out what would happen with or without the help of that variable. Variable importance measures produced by RF can also sometimes be useful to build simpler model. One warning about RF is that they are dependent on the training set. If the training set is not actually representing the population, it is supposed to obtain inaccurate predictions for values out of training set (Horning, 2010). RF regressions were developed using different combination of predictors (“waveform metrics” and “PCs”) utilizing “randomForest” package by Liaw & Wiener, 2014 in R.

Artificial neural networks are able to perform nonlinear modeling without a prior knowledge about the relationships between input and output variables. It is also a non-parametric and black-box model. Thus they are a more general and flexible modeling tool for forecasting.

A variety of neural network structures have been developed. In this research a Multilayer perceptron (MLP) as the most popular type of neural networks belong to a general class of structures called “feedforward” have been used. An MLP is composed of several layers of neurons (nodes); the first layer “input layer” for distributing the data into network, the last one “output layer” for extraction the result of the network, and remaining layers between input and output are called hidden layers. There are complete connections between neurons in successive layers. Each neuron, except input layer neurons, is obtained by computing weighted sum of previous layer neurons and applying an activation function. Figure 3.4 shows a MLP neural network consisting one hidden layer.

It utilizes a learning technique called backpropagation for training the network. This kind of ANN is based on supervised learning. The idea of backpropagation algorithm is that output of neural network is evaluated against desired output. If results are not satisfactory, weights between layers are modified and process is repeated until error is small enough. The answer that emerges from a neural network’s weights can be difficult to understand and the network’s training can take longer than certain other methods of machine learning such as random forests.



**Fig. 3.4.** A schematic of MLP neural network

Generally, an MLP is characterized by the number of hidden layers, hidden neurons, output neurons and transfer functions. An MLP combined of one input layer, one output layer and one or more hidden layers. In theory a hidden layer with sufficient number of hidden neurons is capable of approximating any continuous function (Kaastra & Boyd, 1996;

Zhang *et al.*, 1998). The number of neurons in input layer equals to the number of variables, and in output layer depends on the application (usually one output neuron for regressions). There is no formula for setting an optimum number of hidden neurons. Katz (1992) indicated that optimum number of hidden neurons is between one-third and two or three times the number of input neurons. Bailey & Thompson (1990) suggested that the number of hidden neurons for a three layer ANN should be 75% of the number of input neurons.

The relationship between input and output of a neuron or network is determined by an activation function. There are a number of common activation functions in use with ANN including “Linear”, “Gaussian”, “Sigmoid”, “Hyperbolic tangent” etc. Selection of activation function is arbitrary and is usually determined by response variable (Günther & Fritsch, 2010). In this research the hyperbolic tangent and linear function were used as activation function, respectively in hidden and output layers. A numerous ANN models were developed using different combination of input variables (“waveform metrics” and “PCs”), and different number of hidden layers and neurons. The models were developed in R using “monmlp” package (Cannon, 2012).

As mentioned, the idea of using terrain index and edge extents came to remove the broadening effects of sloped terrain. It was questioned if other waveform metrics could improve the result. The selection of predictors (metrics) was initially based on the experience and knowledge about impact of extracted metrics on the forest height (regarding to the literatures). A stepwise regression was also used to select the best combination of predictors. It combines backward elimination and forward selection to reach the best combination of metrics based on AIC criteria. The selected metrics were employed in MLR, RF and ANN models. However numerous models were built using waveform metrics and PCs, a few MLR models developed using waveform metrics were presented in table 3.1 as instance.

**Table 3.1.** Some MLR models fitted based on waveform metrics

#	model	reference
1	$H_E = aW_{\text{ext}} - bTI$	Lefsky <i>et al.</i> , 2005
2	$H_E = aW_{\text{ext}} - bTI - c$	
3	$H_E = a\text{Ln}(W_{\text{ext}}) - bTI - c$	Xing <i>et al.</i> , 2010
4	$H_E = aW_{\text{ext}} - b(H_{\text{lead}} + H_{\text{trail}})$	Chen, 2010b
5	$H_E = aW_{\text{ext}} - bH_{\text{lead}} - cH_{\text{trail}} - d$	Baghdadi <i>et al.</i> , 2014
6	$H_E = aW_{\text{ext}} - bH_{\text{trail}}$	Baghdadi <i>et al.</i> , 2014
7	$H_E = aW_{\text{ext}} - bH_{\text{trail}} - c$	Baghdadi <i>et al.</i> , 2014
8	$H_E = aW_{\text{ext}} - bH_{\text{lead}}$	
9	$H_E = aW_{\text{ext}} - bH_{\text{lead}} - c$	
10	$H_E = aW_{\text{ext}} - bTI + cH_{\text{lead}} - d$	
11	$H_E = aW_{\text{ext}} - bTI - cH_{\text{trail}} - d$	
12	$H_E = aW_{\text{ext}} - bTI - cH_{\text{trail}}$	Baghdadi <i>et al.</i> , 2014
13	$H_E = aW_{\text{ext}} - bTI + cH_{\text{lead}} - dH_{\text{trail}}$	
14	$H_E = aW_{\text{ext}} - bTI + cH_{\text{lead}} - dH_{\text{trail}} - e$	
15	$H_E = aW_{\text{ext}} - bTI + cH_{\text{lead}} - dH_{\text{trail}} + eH_{50} + f$	
16	$H_E = aW_{\text{ext}} + bTI + cH_{50}$	
17	$H_E = aW_{\text{ext}} + bTI + cH_{50} + d$	
18	$H_E = aW_{\text{ext}} - bTI + cH_{50} - dH_{75} + eH_{100} + f$	
19	$H_E = a.W_{\text{ext}}^{(2.5)} + b.W_{\text{ext}}^{(1.5)} + cTI^{(1.5)} + d.\text{Ln}(H_{50}) + e$	
20	$H_E = b.W_{\text{ext}}^{(2)} + b.W_{\text{ext}}^{(1.5)} + cTI + dH_{\text{trail}} + eH_{50} + fH_{75} + gH_{100} + h.H_{75}^{(1.5)} + i.H_{75}^{(2.5)} + j$	

$H_E$ : Estimated height (maximum or Lorey's height); a,b,c,d,e,f,g,h and i are coefficients; TI: terrain index. Other metrics were presented in table 2,8.

### 3.2. Forest volume estimation using ICESat GLAS

Two methods were applied to estimate forest volume. The first method consists of three steps: 1) developing volume- $H_{\text{max}}$  and volume- $H_{\text{Lorey}}$  relationships. The stronger one was chosen to estimate volume next. The common volume-height relationship (Equation 3.2) used in different literatures (Baghdadi *et al.*, 2014; Lefsky *et al.*, 2005; Saatchi *et al.*, 2011; Mitchard *et al.*, 2012; Healey *et al.*, 2012), was calibrated based on collected in situ data; 2) estimating height from GLAS data using best model resulted from subsection 3.1.2. It should be mentioned that if volume- $H_{\text{Lorey}}$  relationship is chosen at the first step, Lorey's height would be estimated from lidar data; and 3) estimating forest volume (V) using chosen volume-height relationship. This method has been used in several studies (Baghdadi *et al.*, 2014; Lefsky *et al.*, 2005; Mitchard *et al.*, 2012; Healey *et al.*, 2012):

$$V = a \cdot H^b \quad (3.2)$$

Where  $V$  is volume in  $\text{m}^3/\text{ha}$  and  $H$  is  $H_{\text{max}}$  or  $H_{\text{Lorey}}$  in m.

The second method estimates forest volume directly from GLAS waveforms (Duncanson, 2009; Zhifeng *et al.*, 2010; Nelson *et al.*, 2009; Hayashi *et al.*, 2015). In fact, a large number of MLR, RF and ANN regressions were developed based on waveform metrics, and PCs to predict forest volume.

### 3.3. Model Validation

A Cross validation allows models to be tested using the full training set by means of repeated resampling; thus, maximizing the total number of points used for testing and potentially, helping to protect against overfitting (Rao *et al.*, 2008). In a k-fold cross validation, the dataset is randomly split into k subsets, and each fold uses one of the k subsets as test and the remaining data for training. This process is then repeated k times (the folds), with each of the k subsamples used exactly once as the validation data. The k results from the folds can then be averaged (or otherwise combined) to produce a single estimation (Fig. 3.5). In this study, based on a 5-fold cross validation, 80 percent of observations were iteratively used for building the models and the remained 20 percent were used for validation. It is worth to note that in RF, as OOB estimate of the error rate is an unbiased estimate of the generalization error, it is not necessary to test the predictive ability of the model using a cross validation procedure (Breiman, 2001). However, in accordance with MLR and ANN and for a more reliable comparison, a 5-fold cross validation was performed. A number of statistics was calculated between predicted parameter from GLAS data (maximum height, Lorey's height or volume) and correspondent in situ measurements. Adjusted coefficient of determination ( $R^2_a$ ) as an indicator of the fit quality (Cameron & Windmeijer, 1995), Root Mean Square Error (RMSE) as a measure of accuracy (Lee *et al.*, 2011), Mean Absolute Error (MAE) as a measure of dispersion (Chai & Draxler, 2014; Willmott & Matsuura, 2005), Mean Absolute Percentage Error (MAPE) as an expression of accuracy in percentage (Makridakis & Hibon, 1995; Hyndman & Koehler, 2005), and Akaike Information Criterion (AIC) as a means for model selection by trading-off between the goodness of fit of the model and the complexity of the model (Burnham & Anderson, 2002) were used to evaluate the result of predictions. The significance of all models was also tested statistically. All above mentioned statistics were presented in equations 3.3 to 3.7.

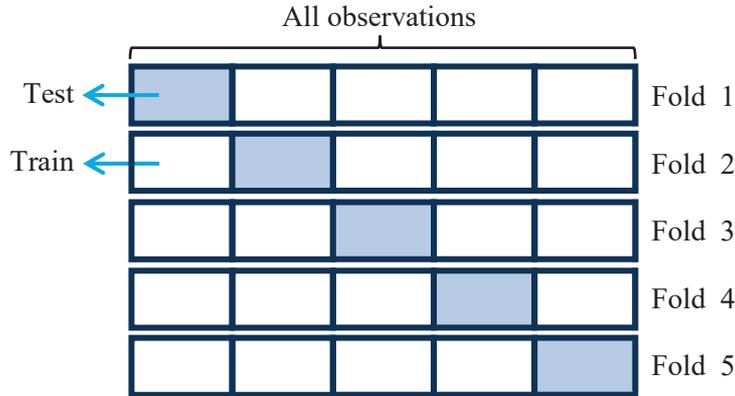


Fig. 3.5. Schematic diagram of a 5-fold cross validation

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (H_i - \hat{H}_i)^2}{n}} \quad (3.3)$$

$$R_a^2 = R^2 - \left( \frac{k \times (1 - R^2)}{n - k - 1} \right) \quad R^2 = 1 - \frac{\sum_{i=1}^n (H_i - \hat{H}_i)^2}{\sum_{i=1}^n (H_i - \bar{H})^2} \quad (3.4)$$

$$MAE = \frac{\sum_{i=1}^n |H_i - \hat{H}_i|}{n} \quad (3.5)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{(H_i - \hat{H}_i)}{H_i} \right| \times 100 \quad (3.6)$$

$$AIC = 2K + n[\ln(RSS/n)] \quad (3.7)$$

Where  $R^2$ : determination coefficient,  $n$ : number of observations,  $H_i$ : in situ height for the plot  $i$ ,  $\hat{H}_i$ : estimated height for the plot  $i$ ,  $\bar{H}$ : mean in situ height,  $K$ : number of predictors in the model,  $RSS$ : residual sum of squares for the fitted model.

### 3.4. Production of canopy height map and its validation

Since GLAS data does not provide continuous coverage of the study area, other sources of remotely sensed data were employed to produce a wall-to-wall height map. Two strategies were used to produce a canopy height map. The first one provides a height map based on a regression model regardless of spatial correlation between the canopy heights. The second one is regression-kriging, a spatial prediction technique that combines a regression of the dependent variable on some predictors with kriging of the regression residuals. The resulting maps were validated using 32 in situ plots dispersed over part of the study area, accidentally.

### 3.4.1. Canopy height map using regression model

Following steps describe the procedure of canopy height map production using regression model:

1) Applying GLAS height model on all GLAS data:

Since canopy height models were built and validated using 60 GLAS waveforms (called GLAS height models hereinafter), the best maximum height and Lorey's height models were employed to estimate these heights at the location of all GLAS footprints over the study area (450 footprints).

2) Developing height model using GLAS heights as reference data and indices extracted from other remote sensing (radar and optical images) and environmental data:

As it was mentioned before, to match all indices extracted from PALSAR and optical images (TM and SPOT) and also DEM-extracted variables spatially, they were resampled to 10 m resolution. Since the average size of GLAS footprints is 70 meter in diameter, mean value of all indices in a 7 by 7 window at the center of GLAS footprints were calculated (mode value for categorical variables). Multiple linear regression (MLR) and Random Forest regression (RF) was used to develop canopy height models which is called second height model hereinafter). For developing MLR models, the most correlated indices were entered in a stepwise regression. Selection of indices for RF models was based on both stepwise regression and importance degree of indices. The main advantage of random forest is its incorporation of continuous or qualitative predictors without making assumptions about their statistical distribution or covariance structure (Breiman, 2001).

3) Selection of best second height model based on statistical criteria through cross validation described in section 3.3.

4) Applying the best second height model (For both  $H_{\max}$  and  $H_{\text{Lorey}}$ ) on the study area and producing maximum and Lorey's height map.

### 3.4.2. Canopy height map using regression-kriging method

Kriging is an advanced geostatistical procedure of interpolation that generates an estimated surface from a scattered set of points with z-values. Kriging assumes that the distance or direction between sample points reflects a spatial correlation that can be used to explain variation in the surface. It is a multistep process which includes exploratory statistical analysis of the data, variogram modeling, creating the surface, and (optionally) exploring a variance surface (Goovaerts, 1997).

Kriging weights the surrounding measured values to derive a prediction for an unmeasured

location. The general formula is formed as a weighted sum of the data (Equation 3.8):

$$z(S_0) = \sum_{i=1}^n \lambda_i z(S_i) \quad (3.8)$$

$z(S_i)$ : the measured value at the  $i^{\text{th}}$  location,  $z(S_0)$ : predicted value at prediction location  $S_0$ ,  $n$ : the number of measured values and  $\lambda_i$ : an unknown weight for the measured value at the  $i^{\text{th}}$  location. In ordinary kriging which is used in this study,  $\lambda_i$  depends on a fitted model to the measured points (the fitted semivariogram function), the distance between the measured points and the prediction location and also the overall spatial arrangement of the measured points.

Regression-kriging involves spatially interpolating the residuals from a non-spatial model using kriging, and adding the results to the prediction obtained from the non-spatial model (Goovaerts, 1997). As described in section 3.4.1, second height model was built to predict canopy height from remotely sensed and environmental data, and a wall to wall canopy height map was produced by applying this model on the entire study area. But it does not take into account the spatial correlation between the canopy heights. The main objective of this section is to consider spatial correlation between the canopy heights in order to improve height map. To do so, semivariogram analysis was applied to the regression residuals (the difference between the predicted and actual values) to quantify the spatial structure of canopy height. This method has been widely used to analyze spatial structures in ecology (Robertson, 1987; Ge *et al*, 2011; Eldeiry & Garcia, 2010). The semivariogram plots the semivariance as a function of the distance between samples using equation 3.9:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(S_i) - z(S_i + h)]^2 \quad (3.9)$$

Where  $\gamma(h)$  is semivariance as a function of distance  $h$ ,  $z(S_i)$  and  $z(S_i+h)$  are the estimated residuals from the regression models at locations  $S_i$  and  $S_i+h$ , a location separated by distance  $h$ ,  $N(h)$  is the total number of pairs of samples separated by distance  $h$ . The empirical variogram, which is a plot of the values of  $\gamma(h)$  as a function of  $h$ , gives information on the spatial dependency of the variable. Once each pair of locations is plotted, a model is fit through them. The coefficients “nugget”, “range” and “sill” are commonly used to describe the model. The nugget is the semivariance at a lag distance of zero. The distance where the model first flattens is known as the range. Sample locations separated by distances closer than the range are spatially autocorrelated, whereas locations farther apart than the range are not. The value at which the semivariogram model attains

the range (the value on the y-axis) is called the sill. A partial sill is the sill minus the nugget. Figure 3.6 shows a schematic of nugget, range and sill components of a semivariogram.

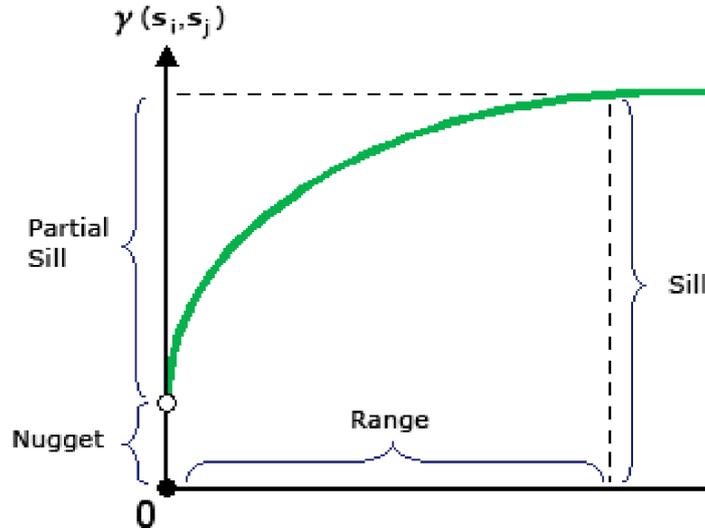


Fig. 3.6. Illustration of nugget, range and sill components of a semivariogram

The Exponential model (Equation 3.10) which fitted the empirical semivariogram was used to define the regression-kriging model.

$$\gamma(h) = S^2 + \sigma^2 \left[ 1 - \exp\left(\frac{-h}{a}\right) \right] \quad (3.10)$$

Where  $S^2$  is the nugget,  $\sigma^2$  the sill, and  $a$  the range of the semivariogram,  $\gamma(h)$ .

Next, the fitted semivariograms were used in the kriging of the canopy height residuals for maximum and Lorey's height using equation 3.8 and then defining the regression-kriging model. As its name indicates, regression-kriging is consisted of a regression part ( $\hat{m}(S_0)$ ) and a kriging part ( $z(S_0)$ ) as shown in the equation 3.11.

$$\hat{z}(s_0) = \hat{m}(S_0) + z(S_0) = \hat{m}(S_0) + \sum_{i=1}^n \lambda_i z(S_i) \quad (3.11)$$

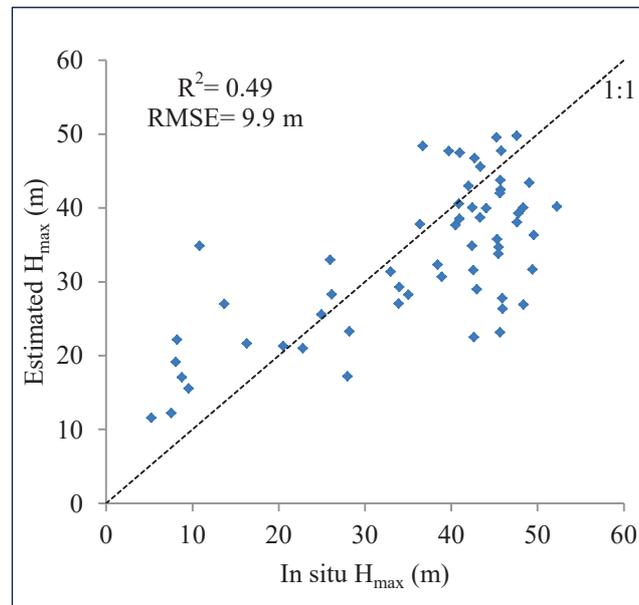
Where  $\hat{z}(s_0)$  is canopy height value using regression-kriging method,  $\hat{m}(S_0)$  is the fitted trend,  $z(S_0)$  is the kriged residual,  $\lambda_i$  are the kriging weights determined by the spatial dependence structure of the residual and  $z(S_i)$  is the residual at location  $S_i$ .

## 4. Results and discussion

### 4.1. Estimation of maximum canopy height ( $H_{\max}$ ) using ICESat GLAS

#### 4.1.1. Direct method for estimation of $H_{\max}$

Maximum canopy height calculated using direct method (vertical difference between signals start and ground peak of GLAS waveform) was compared with in situ  $H_{\max}$ . The correlation between estimated  $H_{\max}$  and  $H_{\max}$ -total was higher rather than the correlation between estimated  $H_{\max}$  and  $H_{\max}$ -in or  $H_{\max}$ -out (refer to section 2.3.1). Figure 4.1 shows estimated  $H_{\max}$  versus  $H_{\max}$ -total. The adjusted coefficient of determination ( $R_a^2$ ) and root mean square error (RMSE) are 0.48 and 9.9, respectively.



**Fig. 4.1.** Estimated  $H_{\max}$  from GLAS data using direct method versus in situ  $H_{\max}$

#### 4.1.2. Estimation of $H_{\max}$ using regression models (MLR, RF and ANN)

In order to reduce the impact of slope on estimation of canopy height, three statistical methods such as multiple linear regressions (MLR), random forest (RF) and artificial neural network (ANN) were used.

##### 4.1.2.1. Estimation of $H_{\max}$ using MLR

Concerning maximum height, the result of regression models using in situ  $H_{\max}$ -total was better than  $H_{\max}$ -in and  $H_{\max}$ -out. Thus hereinafter to the end of GLAS height models result,  $H_{\max}$  refers to  $H_{\max}$ -out. Table 4.1 represents five models developed based on

waveform metrics ( $W_{\text{ext}}$ ) and Terrain index calculated from  $\text{DEM}_{10}$  ( $\text{TI}_{10}$ ). AIC (Akaike Information Criterion) was calculated as a way of selecting the best regression model. The model with lowest AIC score represents the best model. An MLR model combined  $W_{\text{ext}}^{2.5}$ ,  $W_{\text{ext}}^{1.5}$ ,  $\ln(H_{50})$  and  $\text{TI}_{10}^{1.5}$  (model 1) produced the lowest AIC (296.3) and highest accuracy (5.0 m). Based on the MAPE (Mean Absolute Percentage Error), 16.4% of predictions of this model were off (Fig. 4.2). The t-statistics of regression coefficients shows the relative importance of each metric in the model. Based on this statistics,  $\text{TI}_{10}^{1.5}$  and  $W_{\text{ext}}^{1.5}$  contributed most to the model for this set of independent variables.

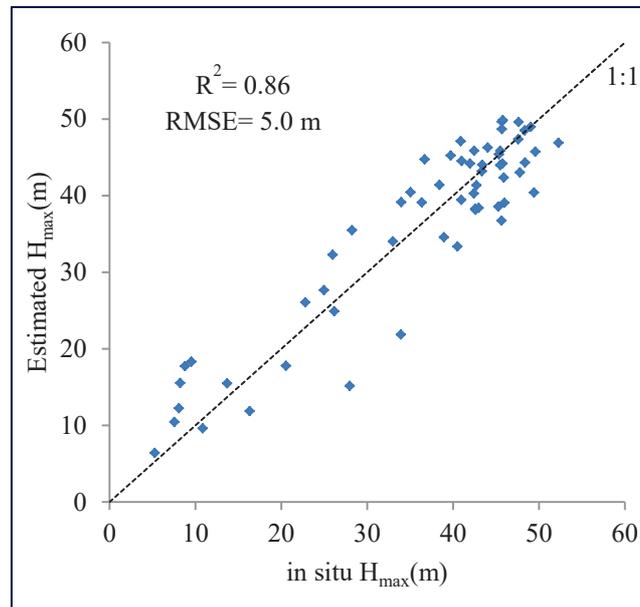
As it is seen in table 4.1, the accuracy of the simplest model (RMSE = 6.3 m) based on two metrics ( $W_{\text{ext}}$  and  $\text{TI}_{10}$ ), was about one meter lower than the accuracy of model 1 (RMSE = 5.0 m). But it should be noticed that the predictor  $H_{50}$  in model 1 is dependent on ground peak identification. So there is uncertainty in extraction of this metric.

However models containing  $H_{\text{trail}}$  and  $H_{\text{lead}}$  produced good result somewhat, but t-statistic showed these predictors are not significant.

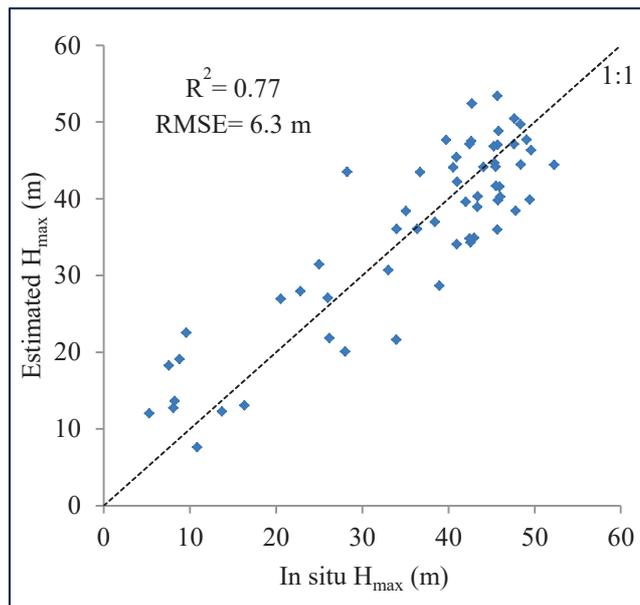
As it is observed in figures 4.2 and 4.3, overestimation and underestimation were decreased largely using regression models 1 and 2 in comparison with direct method (refer to Figure 4.1).

**Table 4.1.** Statistics for five MLR to estimate  $H_{\text{max}}$  based on waveform metrics

#	Model	Coefficients	RMSE (m)	RMSE (%)	$R^2_a$	MAE	MAPE (%)	AIC	P-value
1	$H_{\text{max}} = a.W_{\text{ext}}^{2.5} + b.W_{\text{ext}}^{1.5} + c.\ln(H_{50}) + d.\text{TI}_{10}^{1.5} + e$	a= -0.0042 b= 0.386 c= 3.549 d= -0.052 e= 21.22	5.0	13.8	0.85	4.0	16.4	296.3	3.61e-26
2	$H_{\text{max}} = a.W_{\text{ext}} + b.\text{TI}_{10}$	a= 1.1041 b= -0.4910	6.3	17.5	0.76	5.2	23.0	314.2	2.08e-20
3	$H_{\text{max}} = a.W_{\text{ext}} + b.H_{50} + c.\text{TI}_{10}$	a= 0.93538 b= 0.36011 c= -0.42024	5.8	16.1	0.79	4.7	20.3	307.6	2.10e-22
4	$H_{\text{max}} = a.W_{\text{ext}} + b.\text{TI}_{10} + c.H_{50} + d.H_{75} + e.H_{100} + f$	a= 1.0845 b= -0.3822 c= 0.6642 d= -0.1114 e= -0.2605 f= -0.992	6.1	16.9	0.77	8.3	35.3	312.3	4.00e-21
5	$H_{\text{max}} = a.W_{\text{ext}} + b.\text{TI}_{10} + c.H_{\text{lead}} + d.H_{\text{trail}}$	a= 1.1092 b= -0.4948 c= 0.1067 d= -0.1319	6.4	17.8	0.75	5.3	22.7	316.9	5.67e-20



**Fig. 4.2.** Estimated  $H_{\max}$  using MLR based on waveform metrics (model 1, table 4.1) versus in situ  $H_{\max}$



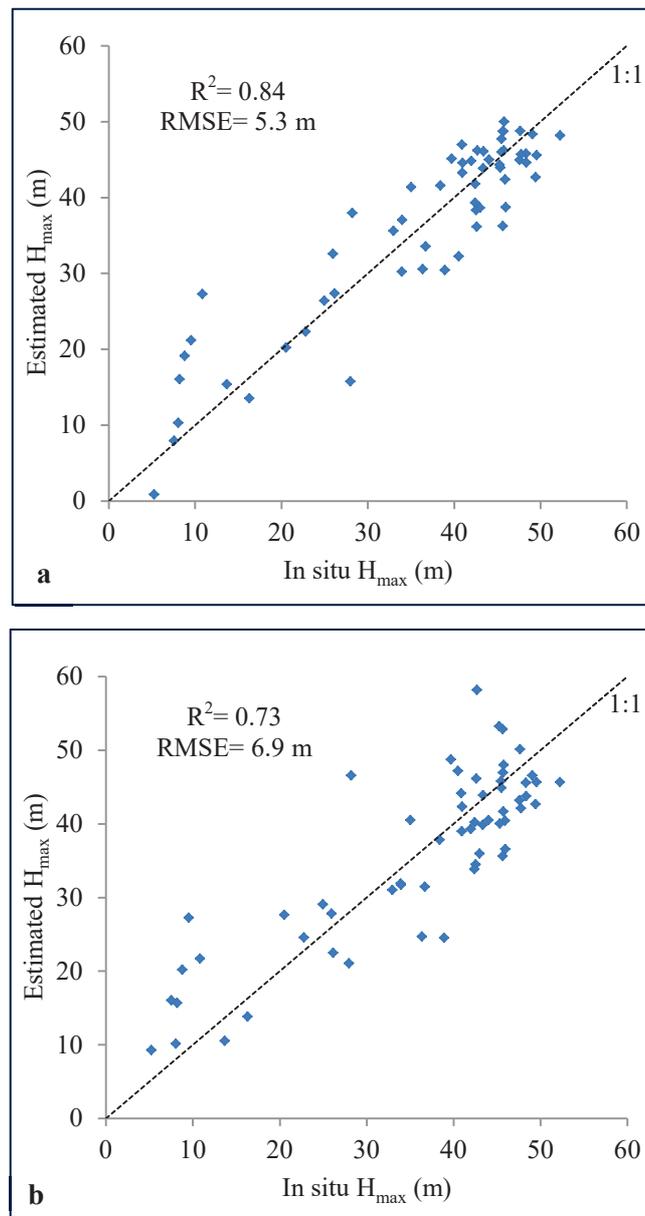
**Fig. 4.3.** Estimated  $H_{\max}$  using MLR based on waveform metrics (model 2, table 4.1) versus in situ  $H_{\max}$

The use of  $TI_{90}$  instead of  $TI_{10}$  in models 1 and 2 led to an  $R^2_a$  of 0.83 and 0.72, and RMSE of 5.3 m and 6.9 m, respectively (Equation 4.1 and 4.2). Figure 4.4 shows the estimated  $H_{\max}$  using equations 4.1 and 4.2 versus in situ  $H_{\max}$ . As it is seen, using  $TI_{90}$  instead of  $TI_{10}$  did not have significant effect on the result of model. Similar outputs were obtained

from  $TI_{10}$  ( $DEM_{10}$ ) and  $TI_{90}$  ( $DEM_{90}$ ) throughout this study. Therefore, only results of models containing  $TI_{10}$  are discussed hereinafter.

$$H_{\max} = -0.00426 W_{\text{ext}}^{2.5} + 0.38712 W_{\text{ext}}^{1.5} - 0.010638 TI_{90}^{1.5} + 4.1842 \ln(H_{50}) - 0.52975 \quad (4.1)$$

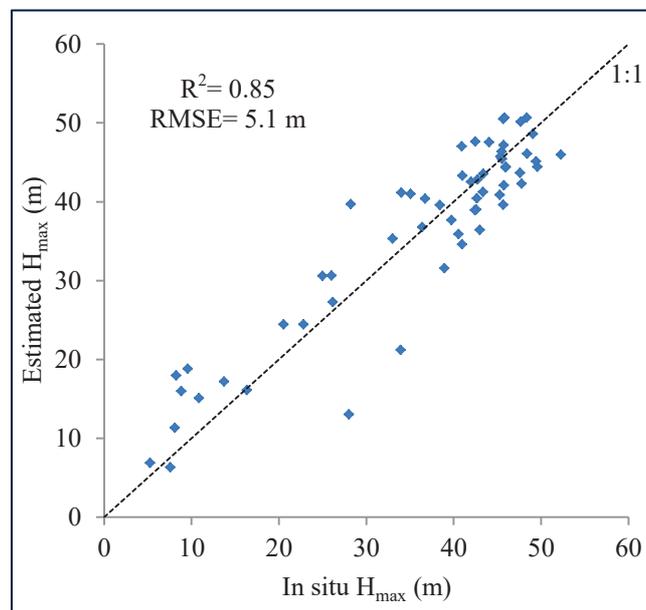
$$H_{\max} = 1.10657 W_{\text{ext}} - 0.16308 TI_{90} \quad (4.2)$$



**Fig. 4.4.** Estimated  $H_{\max}$  using a) model 4.1, b) model 4.2 versus in situ  $H_{\max}$

The model offered by stepwise regression (Equation 4.3) produced an  $R^2_a$  and RMSE of 0.82 and 5.1 m, respectively (Fig. 4.5). As it is seen, the result of stepwise regression is similar to model 1 from table 4.1. But based on principle of parsimony which tends to prevent overfitting and reduction in prediction ability of the model, a model should be simple as much as possible (Vandekerckhove *et al.*, 2014). So however the model resulted from stepwise regression produced good result, it is not preferable because of its complexity.

$$H_{\max} = -0.04336 W_{\text{ext}}^2 + 0.41997 W_{\text{ext}}^{1.5} - 0.48393 \text{TI}_{90} + 1.06764 H_{50} + 10.8052 H_{75} + 0.6756 H_{100} + 0.02297 H_{75}^{1.5} - 2.5984 H_{75}^{1.5} + 0.4155 H_{\text{trail}} - 53.3838 \quad (4.3)$$

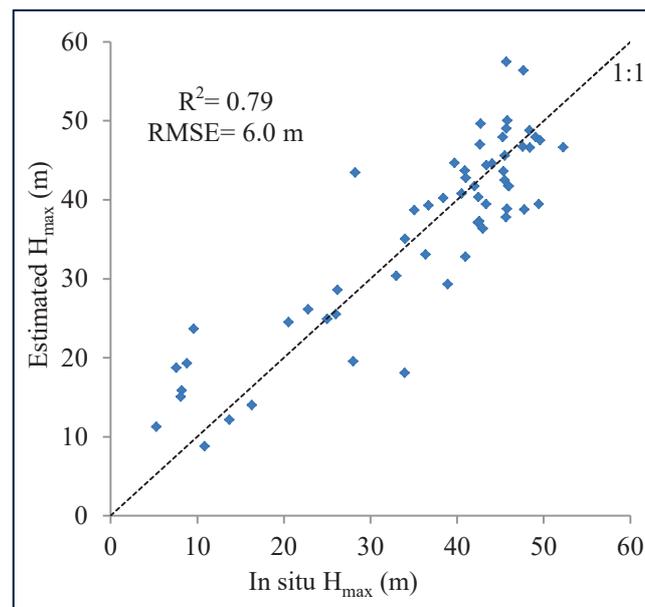


**Fig. 4.5.** Estimated  $H_{\max}$  using stepwise regression of waveform metrics (Eq. 4.3) versus in situ  $H_{\max}$

MLR regressions using all PCs or PCs from stepwise regression (26 PCs) did not produce good result. Three first PCs, explaining 77.5% of data variance, had the highest performance in our models. The statistical result of developing MLR models using three first PCs were presented in table 4.2. The model consisting only three first PCs did not performed well. Adding  $W_{\text{ext}}$  and  $\text{TI}_{10}$  improved the result considerably. The smallest AIC (301.1) belongs to model combining three first PCs,  $W_{\text{ext}}$  and  $\text{TI}_{10}$ . It produced an  $R^2_a$  and RMSE of 0.77 and 6.0 m, respectively, the MAE between predicted and observed height was about 4.7 m, and the prediction error was about 22.1% (Fig. 4.6). Based on the t-statistics of regression coefficients,  $\text{TI}_{10}^{1.5}$  and  $W_{\text{ext}}^{1.5}$  contributed most to the model.

**Table 4.2.** Statistics for three MLR models to estimate  $H_{\max}$  based on PCs

#	Model	coefficient	RMSE (m)	RMSE (%)	$R^2_a$	MAE (m)	MAPE (%)	AIC	P-value
1	$H_{\max} = aPC_1 + bPC_2 + cPC_3 + d$	a= - 5.257 b= 16.065 c= 23.479 d= 35.928	10.8	30	0.29	9.0	46.4	369.0	3.32e-5
2	$H_{\max} = aPC_1 + bPC_2 + cPC_3 + dW_{\text{ext}} + e$	a= 8.1492 b= 7.7948 c= -2.0450 d= 1.3427 e= -21.5505	7.7	21.4	0.63	6.4	28.6	337.3	5.57e-11
3	$H_{\max} = aPC_1 + bPC_2 + cPC_3 + dW_{\text{ext}} + eTI_{10} + f$	a= 4.8863 b= 5.6782 c= -7.1939 d= 1.3460 e= -0.4791 f= -10.5722	6.0	16.6	0.77	4.7	22.1	301.1	2.2e-16

**Fig. 4.6.** Estimated  $H_{\max}$  using MLR based on PCs (model 3, table 3.5) versus in situ  $H_{\max}$ 

#### 4.1.2.2. Estimation of $H_{\max}$ using RF

The five best RF models developed based on waveform metrics were presented in table 4.3. Model 1 containing  $W_{\text{ext}}^{2.5}$ ,  $W_{\text{ext}}^{1.5}$ ,  $\ln(H_{50})$  and  $TI_{10}^{1.5}$  outperformed other models with an  $R^2_a$  and RMSE of 0.72 and 6.8 m, respectively. 28.0% of predictions by this model are off from true values (Fig. 4.7).  $W_{\text{ext}}^{2.5}$  and  $TI_{10}$  have the highest and lowest importance in this model. Models 4 and 5 containing edge extent metrics ( $H_{\text{lead}}$ ,  $H_{\text{trail}}$ ) did not performed as well as other presented models.

**Table 4.3.** Statistics of five RF models for estimation of  $H_{\max}$  based on waveform metrics

#	Importance degree									RMSE (m)	RMSE (%)	$R_a^2$	MAE (m)	MAPE (%)
	$W_{\text{ext}}$	$TI_{10}$	$H_{\text{trail}}$	$H_{\text{lead}}$	$H_{50}$	$\text{Ln}(H_{50})$	$TI_{10}^{1.5}$	$W_{\text{ext}}^{2.5}$	$W_{\text{ext}}^{1.5}$					
1	-	-	-	-	-	1973.6	1348.6	2345.2	2402.9	6.8	18.8	0.72	5.4	28.0
2	5372.6	2596.0	-	-	-	-	-	-	-	7.2	20	0.70	5.7	30.7
3	3366.4	1716.0	-	-	2928.0	-	-	-	-	7.3	20.3	0.69	5.6	30.5
4	2336.6	1270.8	885.7	1308.9	2222.2	-	-	-	-	7.7	21.4	0.70	6.3	35.2
5	3270.0	1830.0	1069.5	1650.6	-	-	-	-	-	8.5	23.6	0.62	6.7	38.8

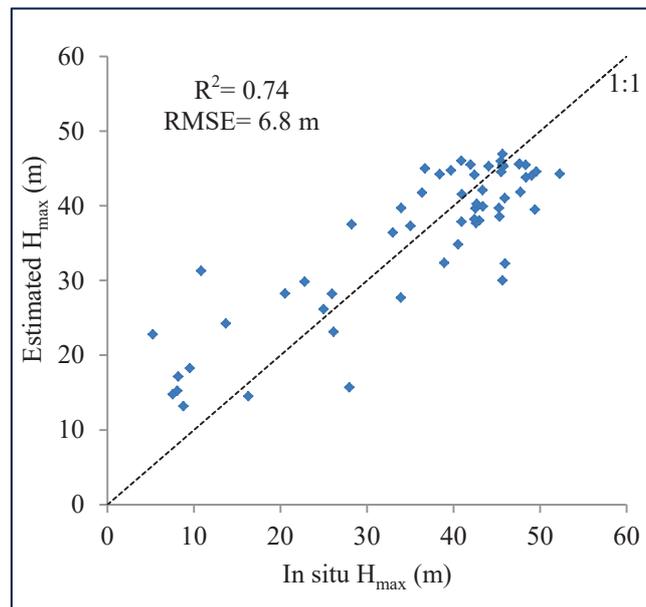
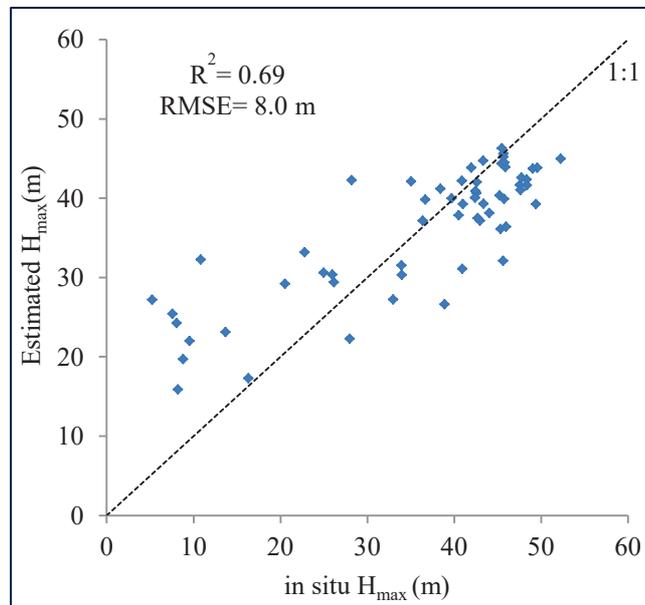
**Fig. 4.7.** Estimated  $H_{\max}$  using RF regression based on waveform metrics (model 1, table 4.3) versus in situ  $H_{\max}$

Table 4.4 shows statistics of three RF models based on PCs. The best result concerning RF regressions based on PCs was generated using the same metrics as MLR (three first PCs,  $W_{\text{ext}}$  and  $TI_{10}$ ) with an  $R^2_a$  and RMSE of 0.66 and 8.0 m, respectively (Fig. 4.8). As it is seen, however  $PC_1$  consists the largest variance of data among all PCs, it has less importance degree rather than  $PC_2$  and  $PC_3$  in the model. This confirms that the informative part of waveform is not always in the first PC.

**Table 4.4.** Statistics of three RF models for estimation of  $H_{\text{max}}$  based on PCs

#	Importance degree					RMSE (m)	RMSE (%)	$R^2_a$	MAE (m)	MAPE (%)
	$W_{\text{ext}}$	$TI_{10}$	$PC_1$	$PC_2$	$PC_3$					
1	-	-	1807.3	1914.4	3839.5	9.3	25.8	0.50	7.4	40.6
2	2920.2	-	1081.0	1338.3	2427.0	8.0	22.2	0.63	6.1	33.1
3	2359.4	1113.7	928.5	1131.8	2153.9	8.0	22.2	0.66	6.1	36.0



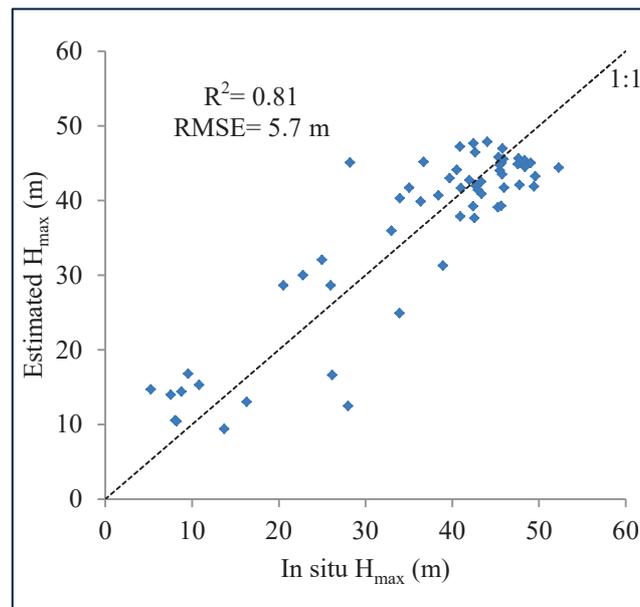
**Fig. 4.8.** Estimated  $H_{\text{max}}$  using RF regression based on PCs (model 3, table 4.4) versus in situ  $H_{\text{max}}$

### 4.1.2.3. Estimation of $H_{\max}$ using ANN

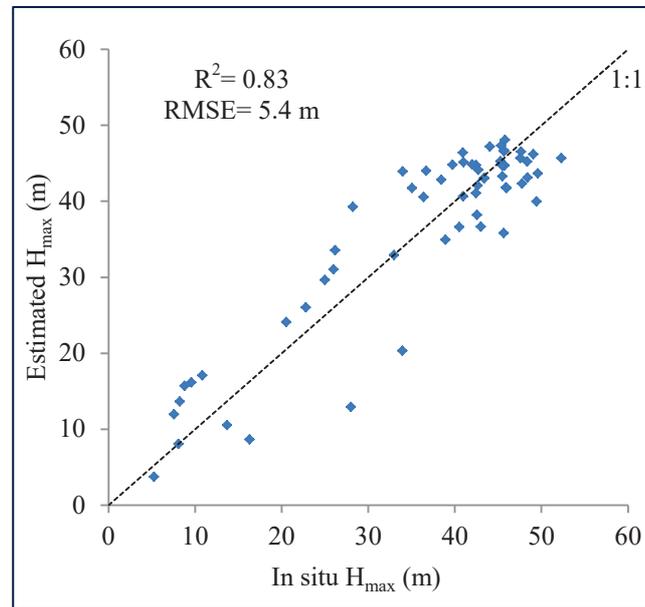
In order to access the optimal structure of ANN, numerous networks with different hidden layers and neurons and also various iteration rates were assessed. As it is seen in table 4.5, three-layer networks (one hidden layer) performed well in predicting forest height. Based on the results, a three layers network with only two metrics “ $W_{\text{ext}}$ ” and “ $TI_{10}$ ” is able to predict maximum height with an RMSE and  $R_a^2$  of 5.7 m and 0.80, respectively (Fig. 4.9). Adding other metrics did not improve the result considerably. As it is seen, an ANN with three metrics “ $W_{\text{ext}}$ ”, “ $TI_{10}$ ” and “ $H_{50}$ ” produced an RMSE and  $R_a^2$  of 5.4 m and 0.82, respectively (Fig. 4.10).

**Table 4.5.** Properties of two ANN models for estimation of  $H_{\max}$  based on waveform metrics and the resulted statistics

#	Input	Properties of network			RMSE (m)	RMSE (%)	$R_a^2$	MAE (m)	MAPE (%)
		Number of hidden layers	Number of hidden neurons	Iteration rate					
1	$W_{\text{ext}}, TI_{10}$	1	2	30	5.7	15.8	0.80	4.6	20.1
2	$W_{\text{ext}}, TI_{10}, H_{50}$	1	2	20	5.4	15	0.82	4.3	17.1



**Fig. 4.9.** Estimated  $H_{\max}$  using ANN based on waveform metrics (model 1, table 4.5) versus in situ  $H_{\max}$

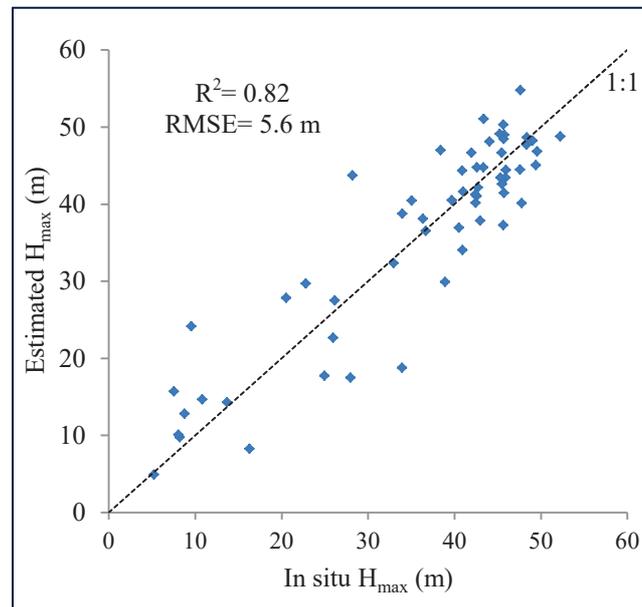


**Fig. 4.10.** Estimated  $H_{\max}$  using ANN based on waveform metrics (model 2, table 4.5) versus in situ  $H_{\max}$

Designing neural networks using all PCs or selected PCs through stepwise regression (26 PCs) led to undesirable results. An ANN model based on three first PCs of PCA produced an RMSE and  $R_a^2$  of 8.8 m and 0.53, respectively. Adding variables  $W_{\text{ext}}$  and  $TI_{10}$  improved the result of predictions (Table 4.6). Figure 4.11 shows estimated  $H_{\max}$  using model 3 versus in situ height.

**Table 4.6.** Properties of ANN models for estimation of  $H_{\max}$  based on PCs and the resulted statistics

#	Input	Properties of network			RMSE (m)	RMSE (%)	$R_a^2$	MAE (m)	MAPE (%)
		Number of hidden layers	Number of hidden neurons	Iteration rate					
1	$PC_1, PC_2, PC_3$	1	3	20	8.8	24.4	0.53	6.6	31.3
2	$PC_1, PC_2, PC_3, W_{\text{ext}}$	1	4	15	6.6	18.3	0.73	5.0	20.3
3	$PC_1, PC_2, PC_3, W_{\text{ext}}, TI_{10}$	1	4	15	5.6	15.5	0.81	4.3	17.4

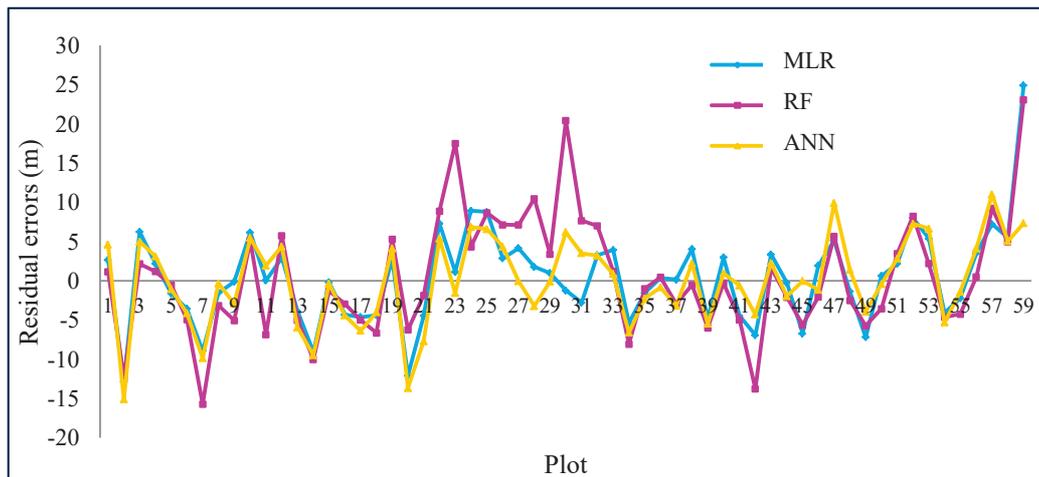


**Fig. 4.11.** Estimated  $H_{\max}$  using ANN based on PCs (model 3, table 4.6) versus in situ  $H_{\max}$

#### 4.1.2.4. Discussion on $H_{\max}$ estimated using GLAS data

Maximum canopy height retrieved from direct method did not show a good result. Canopy height has been overestimated where there are short trees and mostly under-estimated in tall trees locations (Fig. 4.1). Overestimation is expected especially where short trees are located over a sloped terrain. In these conditions, the elevation of the highest object within a footprint is not necessarily at the top of the tallest tree, and could be a shorter tree located in higher elevation or even terrain instead of any vegetation which could occur for sparse canopy over steep terrain (Chen, 2010b). Deep investigation in our field data confirms footprints possessing short trees are located over a sloped terrain (the range of terrain slope for these footprints except one (20%) is between 40-55%) with low forest volume as a proxy of forest density (plots 22-28 and 30 in figure 2.10b). As it was demonstrated in figure 2.10a, number of trees (by ha) is approximately like most plots (except for one plot), but correspondent basal area are very low rather than other plots. The slope problem has been solved greatly using an MLR model combining terrain information with GLAS's waveform metrics ( $W_{\text{ext}}^{2.5}$ ,  $W_{\text{ext}}^{1.5}$ ,  $TI_{10}^{1.5}$ ,  $\ln(H_{50})$ ) (Fig. 4.2) and also an ANN model employing three waveform metrics ( $W_{\text{ext}}$ ,  $TI_{10}$ ,  $H_{50}$ ) (Fig. 4.10). The over and underestimation of height has been also decreased considerable. Generally, based on our outcomes, predicting height of short trees using GLAS data is difficult. This was also reported by Nelson (2010). He showed lack of efficiency of GLAS data to accurately measure forest structure in short-tree sparse forests.

In total, all three regression methods (MLR, RF, ANN) based on waveform metrics produced greater accuracy in comparison with models based on PCs. The performance of the best MLR, RF and ANN models based on waveform metrics was compared in figure 4.12. As it is observed, RF had the weakest performance specially where there are short trees (plots 22-28 and 30).



**Fig. 4.12.** Comparison of residual errors produced by MLR, RF and ANN for estimation of  $H_{\max}$  based on waveform metrics

In terms of using  $TI_{90}$  instead of  $TI_{10}$ , models with  $TI_{10}$  produced just slightly better results. This is contrary to our expectations for producing much more accurate result using local DEM generated from topographic map rather than SRTM DEM. One reason could be that conventional DEMs produced from photogrammetric techniques might not adequately characterize topography over forest areas (NOAA Coastal Services Center, 2012). Conclusively, the SRTM DEM could be an acceptable source of information about terrain variability especially in large extent areas with presence of forest cover. Recent availability to the SRTM DEM<sub>30</sub> for whole world (with more details rather than SRTM DEM<sub>90</sub>) strengthens this deduction.

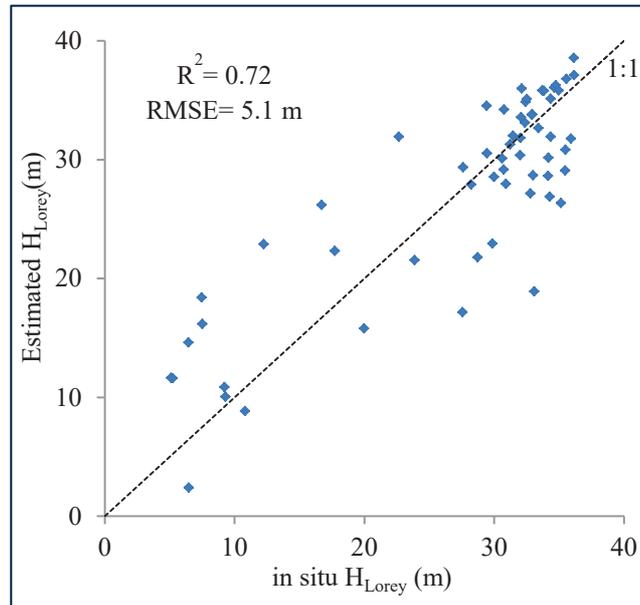
## 4.2. Estimation of Mean Lorey's height ( $H_{Lorey}$ ) using ICESat GLAS

### 4.2.1. Estimation of $H_{Lorey}$ using MLR

Table 4.7 represents some regression models predicting  $H_{Lorey}$ . The first model including  $\ln(W_{ext})$  and  $TI_{10}$  produced the lowest AIC (288.3) with a prediction error of about 24.0% and RMSE of 5.1 m (Fig. 4.13). As mentioned before, significance of coefficients was considered by calculation of t-statistics. However models containing  $H_{50}$ ,  $H_{trail}$  and  $H_{lead}$  (models 4 and 5) produced good result somewhat. But based on t-statistic criterion, coefficients of these metrics are not significant.

**Table 4.7.** Statistics for five MLR for estimation of  $H_{Lorey}$  based on waveform metrics

#	Model	Coefficients	RMSE (m)	RMSE (%)	$R^2_a$	MAE	MAPE (%)	AIC	p-value
1	$H_{Lorey} = a.\ln(W_{ext}) + b.TI_{10} + c$	a= 27.6671 b= -0.3454 c= -67.8802	5.1	18.6	0.71	3.9	24.0	288.3	8.18e-18
2	$H_{Lorey} = a.W_{ext} + b.TI_{10}$	a= 0.8079 b= -0.3252	5.4	19.7	0.70	4.1	23.0	293.1	2.19e-16
3	$H_{Lorey} = a.W_{ext}^{2.5} + b.\ln(W_{ext}) + c.TI_{10}$	a= 0.0007 b= 7.1533 c= -0.3651	5.7	20.8	0.67	4.4	28.7	300.8	2.2e-16
4	$H_{Lorey} = a.W_{ext} + b.H_{50} + c.TI_{10}$	a= 0.7671 b= 0.0876 c= -0.305	5.4	19.7	0.67	4.3	23.5	294.4	1.85e-16
5	$H_{Lorey} = a.W_{ext} + b.TI_{10} + c.H_{lead} + d.H_{trail} + f$	a= 0.7475 b= -0.3408 c= -0.0503 d= 0.0369 f= 3.3432	5.6	20.4	0.63	4.5	25.9	297.7	2.61e-15



**Fig. 4.13.** Estimated  $H_{Lorey}$  using MLR based on waveform metrics (model 1, Table 4.7) versus in situ  $H_{Lorey}$

To estimate Lorey's height based on PCs, like  $H_{max}$ , all PCs or PCs from stepwise regression (20 PCs) were used for prediction. The resulted models produced high error (the RMSE greater than 25 m). MLR models based on only three first PCs, containing 75% of data variance, produced better result (Table 4.8). However based on t-statistics, some coefficients in the model 2 and 3 were not significant.

**Table 4.8.** Statistics of three MLR models for estimation of  $H_{Lorey}$  based on PCs

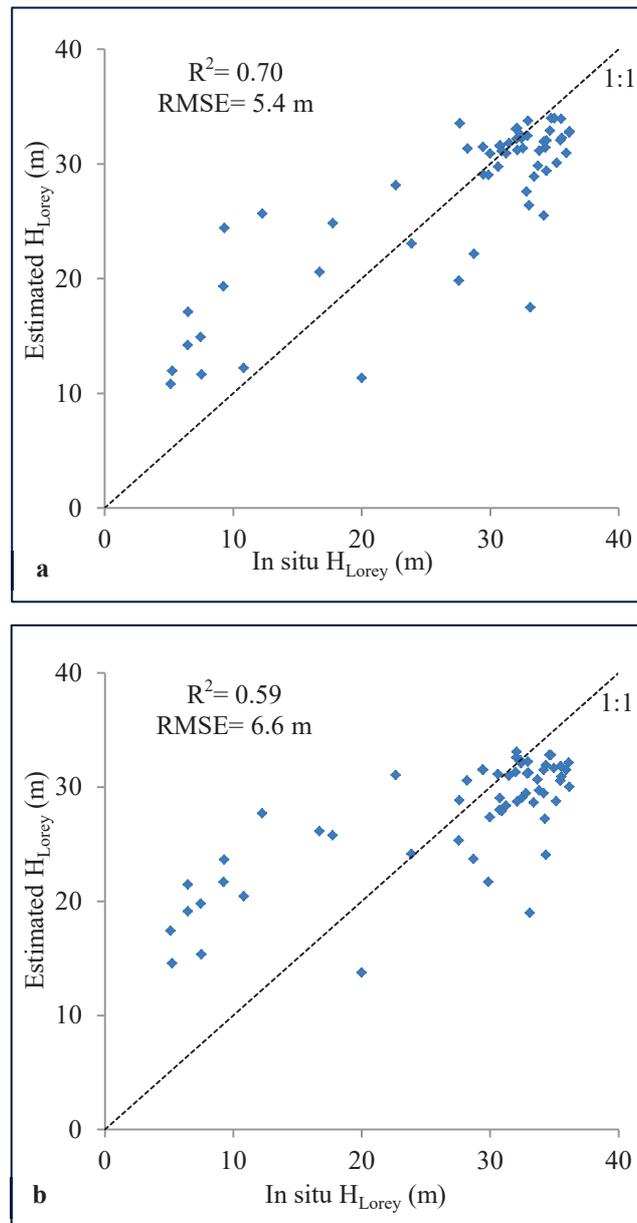
#	Model	Coefficient	RMSE (m)	RMSE (%)	$R^2_a$	MAE (m)	MAPE (%)	AIC	P-value
1	$H_{Lorey} = aPC_1 + bPC_2 + cPC_3 + d$	a= 3.856 b= 12.192 c= 20.590 d= 26.939	7.8	28.4	0.30	6.4	43.9	340.3	1.68e-5
2	$H_{Lorey} = aPC_1 + bPC_2 + cPC_3 + dW_{ext} + e$	a= -7.7682 b= 2.2925 c= 7.6185 d= 0.8996 e= -11.5824	6.6	24.1	0.51	5.2	31.6	320.0	4.19e-9
3	$H_{Lorey} = aPC_1 + bPC_2 + cPC_3 + dW_{ext} + eTI_{10} + f$	a= -4.5647 b= -0.2831 c= 1.5421 d= 0.9006 e= -0.3080 f= -4.4943	5.4	19.7	0.66	4.0	24.1	304.9	1.42e-11

#### 4.2.2. Estimation of $H_{Lorey}$ using RF

The five best RF models developed based on waveform metrics for estimation of  $H_{Lorey}$  were presented in table 4.9. Generally, the four first models produced approximately the same result. Model 1 with four variables (metrics) including  $\text{Ln}(H_{50})$ ,  $\text{TI}_{10}^{1.5}$ ,  $W_{\text{ext}}^{2.5}$  and  $\text{Ln}(W_{\text{ext}})$  produced an RMSE of 5.4 m. Statistic of mean absolute percentage error (MAPE) shows 26.9% of predictions of this model were off. As it is seen, all models presented in this table includes  $\text{TI}_{10}$  or  $\text{TI}_{10}^{1.5}$ , which indicate the importance of this variable in estimation of Lorey's height over sloped area. The prediction error of model 5 built using metrics  $W_{\text{ext}}$ ,  $\text{TI}$ ,  $H_{\text{trail}}$  and  $H_{\text{lead}}$  is greater than the first four models. In fact models containing  $H_{\text{trail}}$  or  $H_{\text{lead}}$  showed less performance in comparison with the others. It could be because of uncertainties in extraction of  $H_{\text{trail}}$  and  $H_{\text{lead}}$  form waveforms broadened by terrain slope (Lefsky *et al.*, 2007). Figure 4.14 shows estimated  $H_{Lorey}$  using model 1 and model 5 versus in situ  $H_{Lorey}$ . As it is seen overestimation and underestimation in model 1 is lower than model 5.

**Table 4.9.** Statistics of five RF models for estimation of  $H_{Lorey}$  based on waveform metrics

#	Importance degree										RMSE (m)	RMSE (%)	$R_a^2$	MAE (m)	MAPE (%)
	$W_{\text{ext}}$	$\text{TI}_{10}$	$H_{\text{trail}}$	$H_{\text{lead}}$	$H_{50}$	$\text{Ln}(H_{50})$	$\text{TI}_{10}^{1.5}$	$W_{\text{ext}}^{2.5}$	$W_{\text{ext}}^{1.5}$	$\text{Ln}(W_{\text{ext}})$					
1	-	-	-	-	-	740.2	490.0	1018.5	1056.8	-	5.4	19.7	0.68	3.9	26.9
2	-	599.8	-	-	-	-	-	1308.3	-	1378.7	5.5	20	0.67	3.9	26.1
3	1527.4	604.2	-	-	1088.4	-	-	-	-	-	5.6	20.4	0.66	4.2	29.8
4	2317.0	861.1	-	-	-	-	-	-	-	-	5.7	20.8	0.66	4.2	28.5
5	1352.7	626.9	437.7	731.8	-	-	-	-	-	-	6.6	24.1	0.56	5.1	37.9

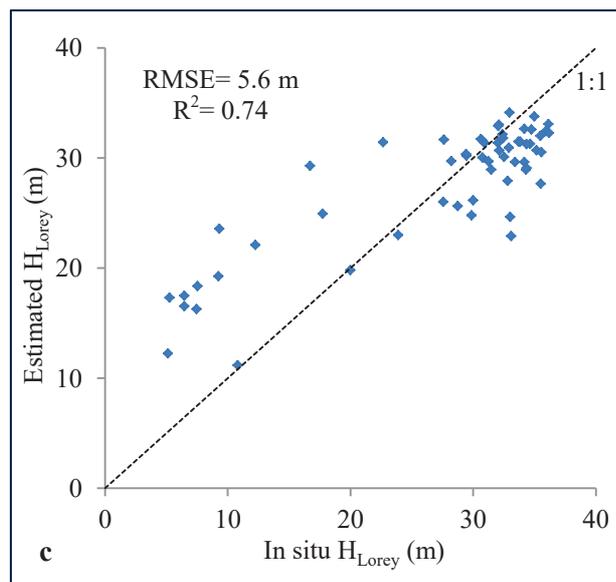
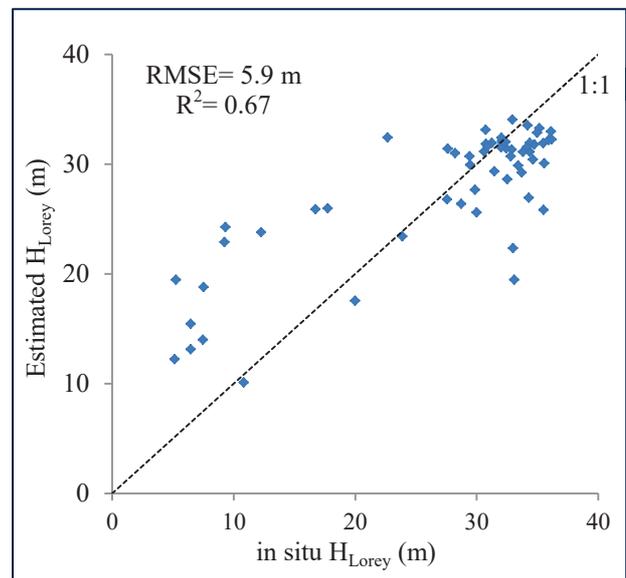
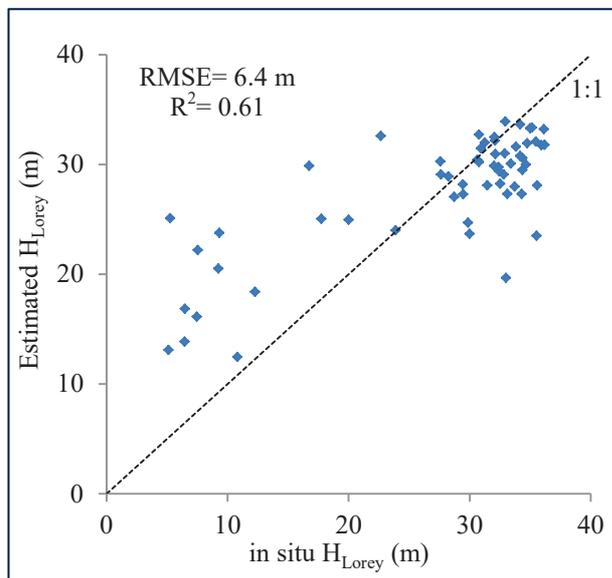


**Fig. 4.14.** Estimated  $H_{Lorey}$  using MLR based on PCs; a) model 1, b) model 5 from table 4.9 versus in situ  $H_{Lorey}$

The RMSE of RF model based on PCs selected by stepwise regression (20 PCs) was high (RMSE= 8.5 m,  $R^2_a = 0.15$ ). Table 4.10 shows statistics of three RF models based on three first PCs and figure 4.15 demonstrates their result versus in situ measurement. The best result was obtained using model 3 which employed metrics  $W_{ext}$ ,  $TI_{10}$  and PCs (Fig. 4.15c) which is similar to the result of RF models based on waveform metrics (Table 4.9). In these models,  $PC_3$  has more contribution to the model than  $PC_2$  and  $PC_1$ , however it contains less variance of the data.

**Table 4.10.** Statistics of three RF models for estimation of  $H_{Lorey}$  based on PCs

#	Importance degree					RMSE (m)	RMSE (%)	$R_a^2$	MAE (m)	MAPE (%)
	$W_{ext}$	TI	PC <sub>1</sub>	PC <sub>2</sub>	PC <sub>3</sub>					
1	-	-	1003.7	1120.0	2178.2	6.4	23.3	0.59	4.8	35.3
2	1448.9	-	681.4	766.0	1366.2	5.9	21.5	0.65	4.3	31.2
3	1276.0	686.7	563.3	615.0	1133.5	5.6	20.4	0.72	4.3	31.3

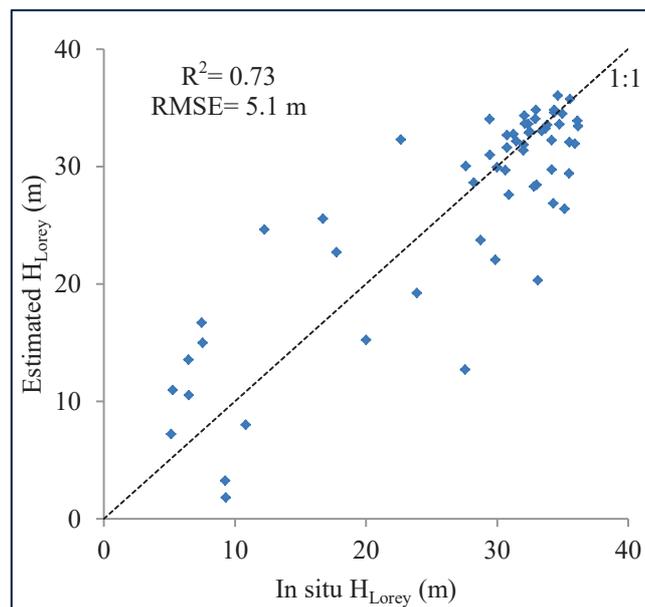
**Fig. 4.15.** Estimated  $H_{Lorey}$  using RF based on PCs versus in situ  $H_{Lorey}$  (a, b, c: model 1 to 3, table 4.10)

### 4.2.3. Estimation of $H_{Lorey}$ using ANN

The optimal structure of ANN was obtained by trial and error. The results showed a three layer network produces suitable outcome in this research. Generally,  $H_{Lorey}$  was predicted with an accuracy of about 5 meters using many types of ANN. Three simple ANN models were presented in table 4.11.  $H_{Lorey}$  resulted from model 1 with two metrics ( $W_{ext}$  and  $TI_{10}$ ) was compared with in situ height in figure 4.16. Based on MAPE statistic, 23.2 percentage of this model predictions are off.

**Table 4.11.** Statistics of three ANN models for estimation of  $H_{Lorey}$  based on waveform metrics

#	Input	Properties of network			RMSE (m)	RMSE (%)	$R_a^2$	MAE (m)	MAPE (%)
		Number of hidden layers	Number of hidden neurons	Iteration rate					
1	$W_{ext}, TI_{10}$	1	2	10	5.1	18.6	0.72	3.7	23.2
2	$W_{ext}^{2.5}, Ln(W_{ext}), TI_{10}$	1	3	10	5.0	18.2	0.72	3.5	20.9
3	$W_{ext}, TI, H_{50}$	1	3	10	5.2	18.9	0.69	4.0	23.5

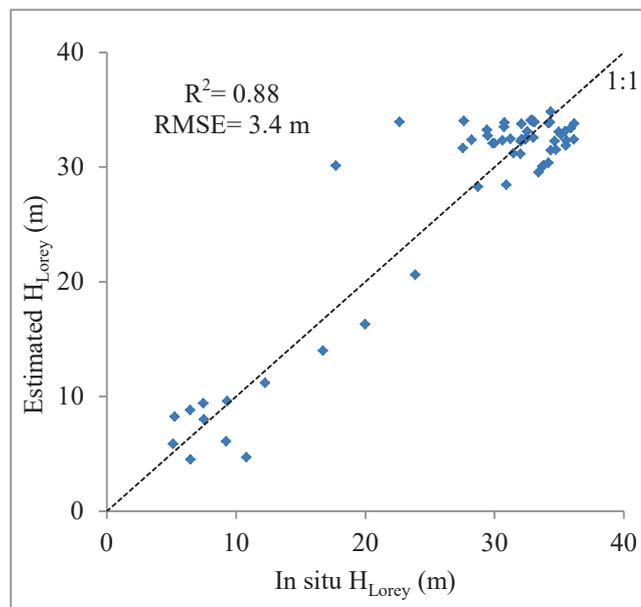


**Fig. 4.16.** Estimated  $H_{Lorey}$  using ANN based on waveform metrics (model 1, table 4.11) versus in situ  $H_{Lorey}$

ANNs developed based on PCs performed slightly better than those based on waveform metrics. A simple neural network employing only three first PCs of PCA produced an RMSE and  $R_a^2$  of 4.7 m and 0.76, respectively (model 1, Table 4.12). Adding  $W_{\text{ext}}$  as input variable improved the result significantly. In other word, an ANN with four inputs including three first PCs and  $W_{\text{ext}}$  (model 2, Table 4.12) estimated  $H_{\text{Lorey}}$  with higher accuracy (RMSE = 3.4 m and  $R_a^2 = 0.87$ ) (Fig. 4.17). 12.3% of predictions by this model are off from true measurements. Adding  $TI_{10}$  did not improve the result comparatively to the model 2.

**Table 4.12.** Statistics of three ANN models for estimation of  $H_{\text{Lorey}}$  based on PCs

#	Input	Properties of network			RMSE (m)	RMSE (%)	$R_a^2$	MAE (m)	MAPE (%)
		Number of hidden layers	Number of hidden neurons	Iteration rate					
1	PC <sub>1</sub> , PC <sub>2</sub> , PC <sub>3</sub>	1	3	35	4.7	17	0.76	3.5	17.4
2	PC <sub>1</sub> , PC <sub>2</sub> , PC <sub>3</sub> , $W_{\text{ext}}$	1	4	50	3.4	12.4	0.87	2.5	12.3
3	PC <sub>1</sub> , PC <sub>2</sub> , PC <sub>3</sub> , $W_{\text{ext}}$ , $TI_{10}$	1	4	30	3.6	13	0.85	2.6	15.5

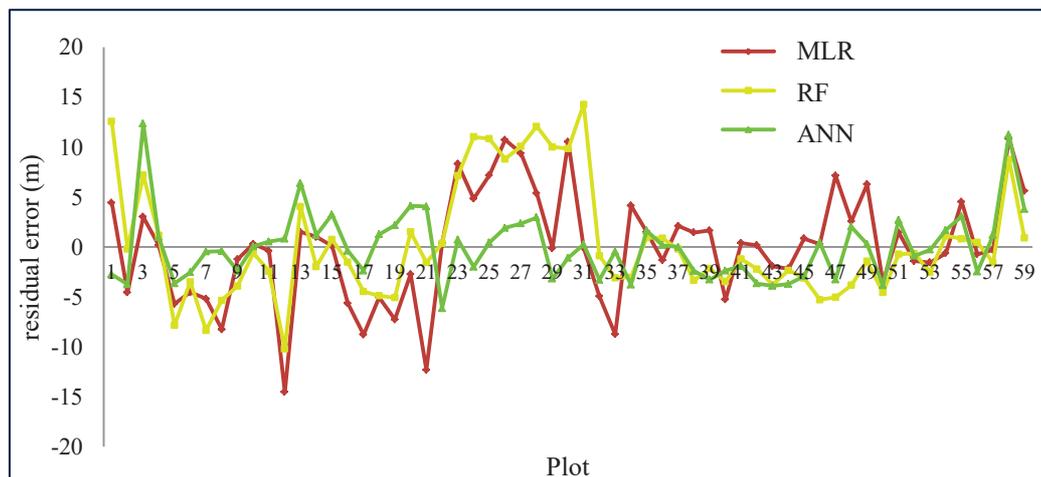


**Fig. 4.17.** Estimated  $H_{\text{Lorey}}$  using ANN based on waveform metrics (model 1, Table 4.12) versus in situ  $H_{\text{Lorey}}$

#### 4.2.4. Discussion on $H_{Lorey}$ estimated using GLAS data

MLR, RF and ANN had approximately similar performance in terms of employing waveform metrics as predictors. In other words, all three methods based on waveform metrics are able to predict Lorey's height with an accuracy of about 5 meters. ANN outperformed two other methods when PCs were used as predictors. The superiority of ANN is considerable in case of using PCs as input variables. An ANN model with four input neurons including three first PCs of PCA and  $W_{est}$  predicted  $H_{Lorey}$  with an accuracy of 3.4 m. The interesting points concerning this model are: firstly, using PCs as input variables which are lack of uncertainties unlike some waveform metrics, especially over sloped terrain; secondly, achieving higher accuracy rather than other models without entering any ancillary data (DEM). In other words, all best models resulted from three statistical methods based on waveform metrics and PCs except ANN based on PCs contained TI extracted from DEM as representative of topography status. Thirdly, in contrast to other models, this model was able to estimate  $H_{Lorey}$  properly even in short sparse and tall dense stands. Figure 4.18 demonstrates performance of the best MLR, RF and ANN models based on PCs.

Fayad *et al.* (2014) estimated canopy height in forest sites of French Guiana that terrain topography is mostly flat. They obtained approximately the same accuracies using MLR and RF models based on waveform metrics and PCs.



**Fig. 4.18.** Comparison of residual errors produced by MLR, RF and ANN for estimation of  $H_{Lorey}$  based on PCs

### 4.3. Estimation of forest volume using ICESat GLAS

As it was explained in section 3.2, two methods were applied to estimate forest volume (V). The first method consists of three steps: 1) developing volume- $H_{\max}$  and volume- $H_{\text{Lorey}}$  relationships ( $V = a \cdot H^b$ ); 2) estimating height from GLAS data using best model resulted from statistical methods explained in section 3.1.2.; and 3) estimating forest volume (V) using chosen volume-height relationship. The second method was estimation of forest volume directly from GLAS waveforms.

Regards to the first method, it was necessary to find out correlation between forest volume and height. Since two heights; maximum height ( $H_{\max}$ ) and mean Lorey's height ( $H_{\text{Lorey}}$ ), were measured in this research, volume- $H_{\max}$  and volume- $H_{\text{Lorey}}$  relationships were developed based on in situ measurements. As it is observed in figure 4.19, there is stronger correlation between volume and  $H_{\text{Lorey}}$  rather than  $H_{\max}$ . And the accuracy resulted from volume- $H_{\text{Lorey}}$  ( $V = 2.6507 H_{\text{Lorey}}^{1.5434}$ ) is higher than volume- $H_{\max}$  ( $V = 2.594 H_{\max}^{1.427}$ ).

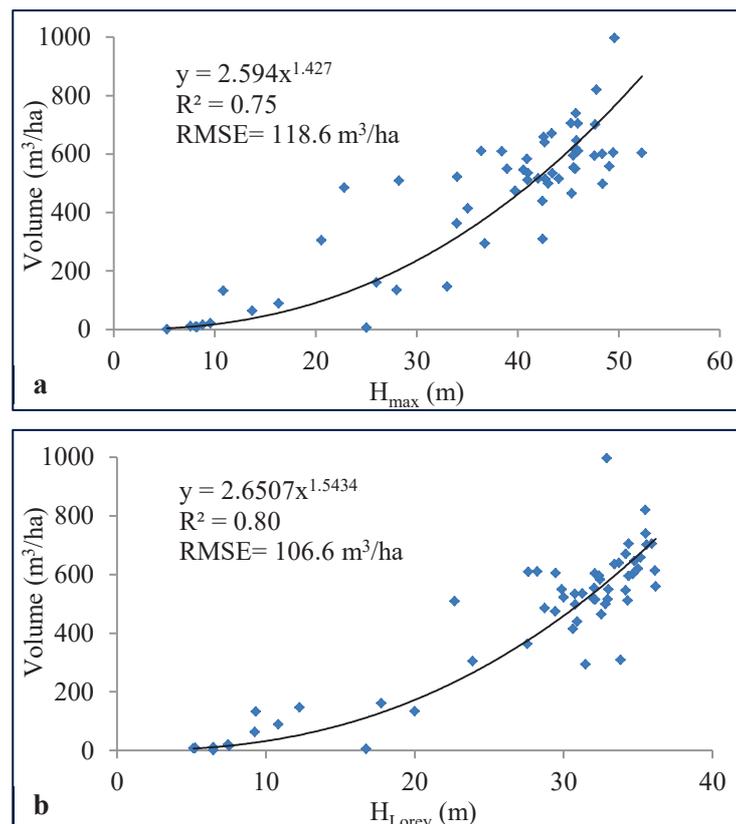
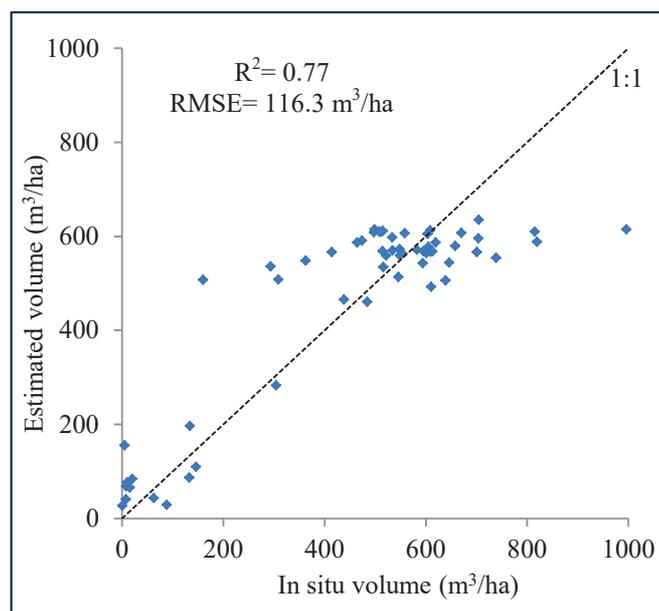


Fig. 4.19. Correlation between volume and a)  $H_{\max}$ , b)  $H_{\text{Lorey}}$

To estimate forest volume, estimated Lorey's heights from GLAS data were replaced in volume- $H_{Lorey}$  equation. Since the best result regards Lorey's height was obtained using an ANN model based on PCs (RMSE = 3.4 m,  $R_a^2 = 0.87$ ), predicted heights using this model were contributed in the  $V = 2.6507 H_{Lorey}^{1.5434}$ . Comparison of estimated volume with in situ volume showed an RMSE and  $R_a^2$  of 116.3 m<sup>3</sup>/ha and 0.77, respectively (Fig. 4.20). The mean absolute error (MAE) of predictions is 83.6 m<sup>3</sup>/ha. This method of volume estimation is called volume- $H_{Lorey}$  hereinafter.

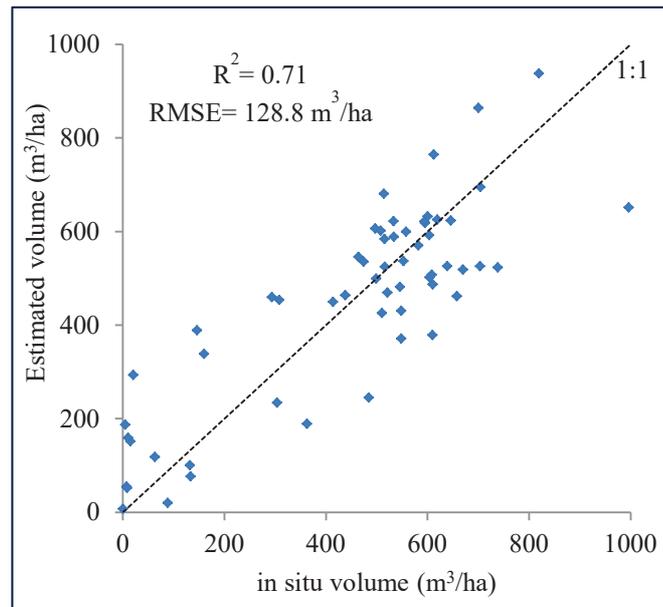


**Fig. 4.20.** Estimated volume using  $V = 2.6507 H_{Lorey}^{1.5434}$  where  $H_{Lorey}$  was estimated from GLAS data versus in situ volume.

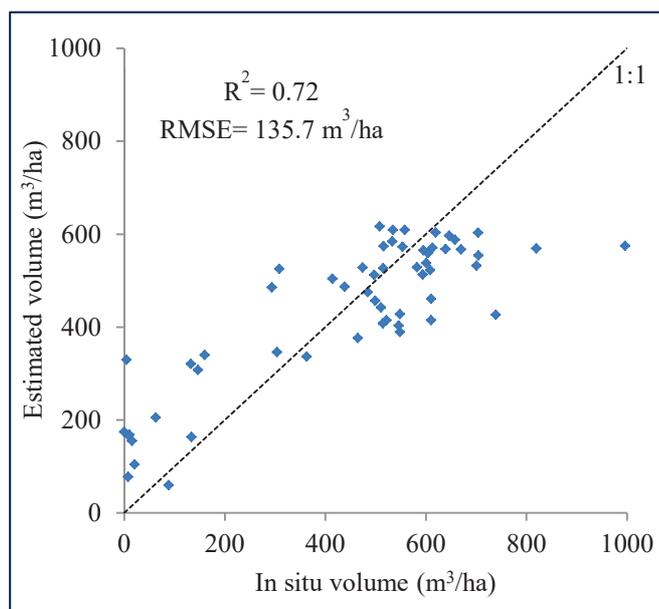
It was under question whether we obtain better result if volume is extracted directly from GLAS data instead of using volume- $H_{Lorey}$  relationship. So, in the second method, a large number of MLR, RF and ANN were developed to estimate forest volume from GLAS waveform.

Concerning MLR models, the best result was obtained based on waveform metrics ( $V = 7.723W_{ext} - 4.406TI_{10} + 1.475H_{lead} + 18.920H_{trail} + 19.482H_{50} - 286.360$ ). It produced an RMSE and  $R_a^2$  of 128.8 m<sup>3</sup>/ha and 0.68, respectively (Fig. 4.21). Moreover an MLR model based on PCs ( $V = -158.252PC_1 + 37.088PC_2 - 252.133PC_3 + 26.421W_{ext} - 5.875TI_{10} - 546.437$ ) produced approximately the same result (RMSE= 131.5 m<sup>3</sup>/ha,  $R_a^2 = 0.67$ ). Regards to RF models, a model employing three first PCs and  $W_{ext}$  outperformed other models with an RMSE and  $R_a^2$  of 135.7 m<sup>3</sup>/ha and 0.70 (Fig. 4.22). In terms of ANN models, best result was also obtained using an ANN based on PCs. A three

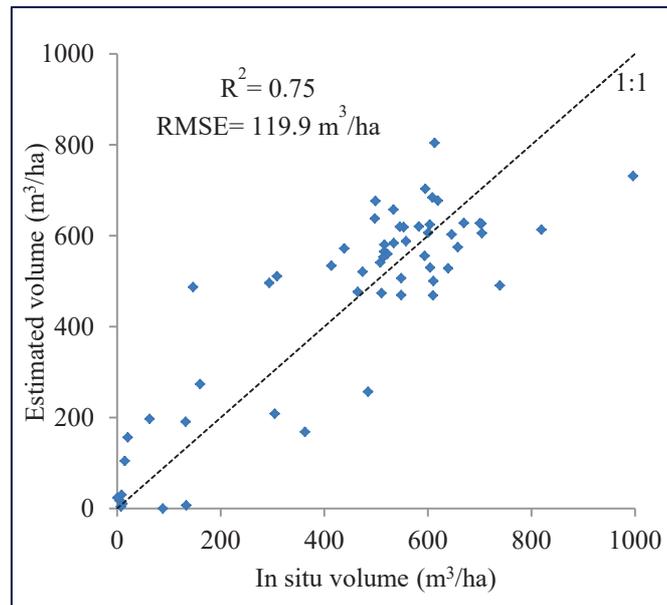
layer network with five hidden neurons, fifteen iteration rate, and four input variables ( $W_{\text{ext}}$  and three first PCs), estimated forest volume with an MAE, RMSE and  $R_a^2$  of 95.5  $\text{m}^3/\text{ha}$ , 119.9  $\text{m}^3/\text{ha}$  and 0.73, respectively (Fig. 4.23). Totally it was possible to estimate volume with an accuracy of 120-135  $\text{m}^3/\text{ha}$  with all three statistical methods (MLR, RF and ANN).



**Fig. 4.21.** Estimated volume using MLR based on waveform metrics versus in situ volume



**Fig. 4.22.** Estimated volume using RF based on PCs versus in situ volume



**Fig. 4.23.** Estimated volume using ANN based on PCs versus in situ volume

#### 4.3.1. Discussion on volume estimated using GLAS data

As observed, forest volume was estimated with an accuracy of  $116.3 \text{ m}^3/\text{ha}$  ( $RMSE\% = 25.4$ ) based on volume- $H_{Lorey}$  relationship, and  $119.9 \text{ m}^3/\text{ha}$  ( $RMSE\% = 26.2$ ) using a neural network model.

Concerning volume- $H_{Lorey}$  method, there were several sources of error that resulted propagation of error and low accuracy of volume estimation. Two main sources were:

- 1) Height-DBH relationships. As known, Lorey's height is not a directly measured parameter on the ground, but is a weighted mean of height. So, to calculate this parameter, height of all trees in each plot were estimated using developed height-DBH relationships.
- 2) It is known that forest volume is a function of both height and diameter as two essential quantitative factors. But in this research, only third dimension of objects is retrievable from GLAS data. So we built  $V = 2.6507 H_{Lorey}^{1.5434}$  with an RMSE and  $R^2_a$  of  $106.6 \text{ m}^3/\text{ha}$  and 0.80, respectively. In other words, even with precise estimation of  $H_{Lorey}$  from GLAS data, we would expect error about  $\pm 100 \text{ m}^3/\text{ha}$  in prediction of volume.

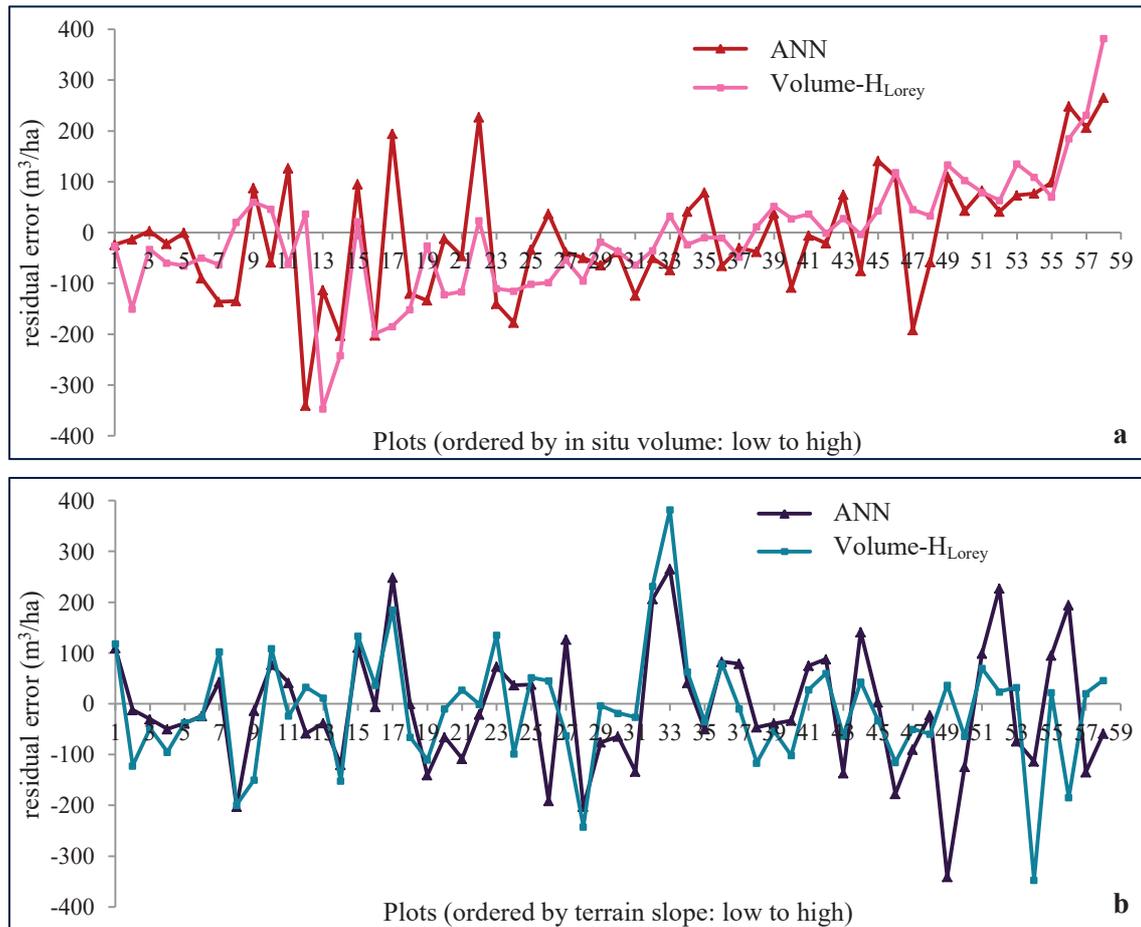
Deep investigation in field inventory data shows the possibility of having same Lorey's height for completely different forest structure which leads to a different forest volume. To better understanding, three couple of plots were compared in terms of Lorey's height (m), number of trees (n/ha) and volume ( $\text{m}^3/\text{ha}$ ) (Table 4.13). As shown, plots with approximately the same Lorey's height have different volumes. It confirms that estimating forest volume only relying on an average height could cause a high discrepancy with reality especially in uneven aged forests.

**Table 4.13.** Comparison of Lorey's height, number of trees and volume in three couple of plots

Plot	Lorey's height (m)	Number of trees (n/ha)	Volume (m <sup>3</sup> /ha)
1	16.7	39	5.2
3	17.7	697	160.5
13	32.9	484	996.6
17	32.8	322	499.2
23	6.5	8	0.7
25	7.5	237	20.7

Concerning statistical methods, the result of an ANN based on PCs is approximately the same as result obtained by MLR model based on waveform metrics. But positive points about ANN model are: 1) It does not need an ancillary data (DEM); 2) It relies on PCs and  $W_{ext}$  which are not prone to uncertainties in contrast with some waveform metrics which have been used in MLR model.

Generally, almost the same accuracy as volume- $H_{Lorey}$  method was obtained using the ANN model. ANN performed slightly better where there exist very low ( $<10$  m<sup>3</sup>/ha) and very high ( $> 800$  m<sup>3</sup>/ha) volume. In most other points ANN produced more error rather than volume- $H_{Lorey}$  method, but points are better dispersed around regression line 1:1 (refer to figure 4.20 and figure 4.23). Residual errors produced by these two methods were sorted based on the lowest to highest values of forest volume (as a proxy of forest density) and terrain slope in figure 4.24. In other words, plot 1 and 58 correspond to the lowest and highest volume in chart a (0.69 and 996.56 m<sup>3</sup>/ha, respectively), flat and sloped terrain (83%) in chart b. As it is observed, there is no obvious trend in volume estimation by increasing in forest density or terrain slope. This confirms that the heterogeneity of forest reduces the ability of lidar data to estimate forest volume. A collection of forest properties including forest type, horizontal and vertical structure of forest, and topographical properties may impress predictions. Consideration of reference plots conditions demonstrated dependency of volume on diverse factors. For instance, it happened to have high number of trees per hectare but low volume and reverse, and also equal number of trees per hectare or equal mean height but different volume (Table 4.13). It is expected to have higher accuracies in homogenous forests.



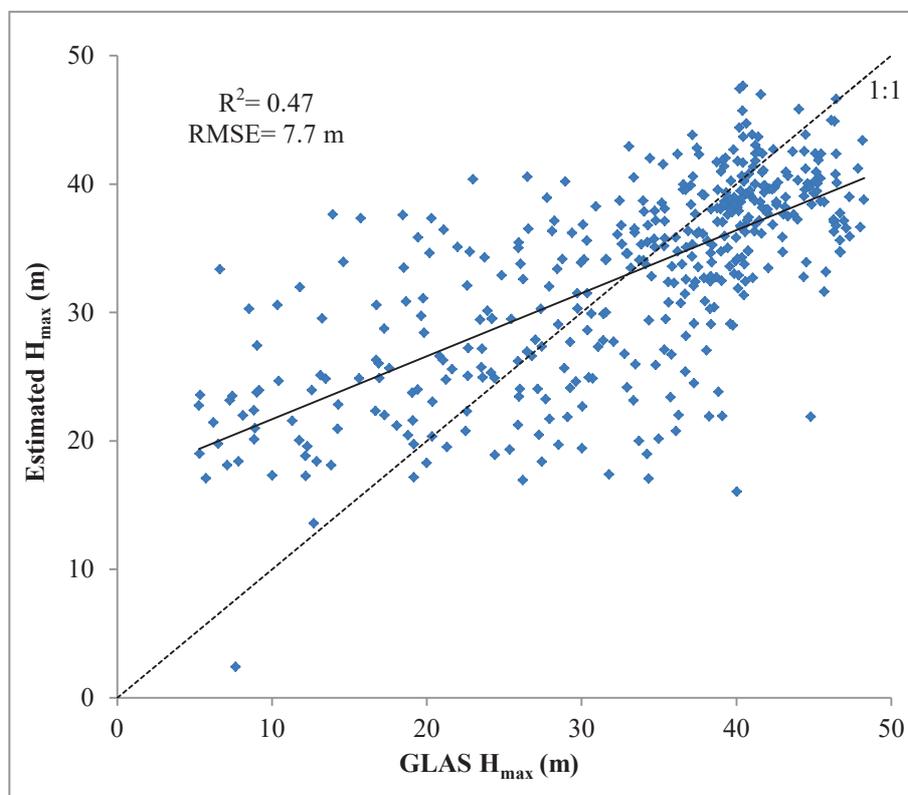
**Fig. 4.24.** Trend of forest volume bias (observed value-estimated value) by a) in situ forest volume, and b) terrain slope (The lowest to highest value on both volume and slope was coded by 1 to 58, respectively. In other words, plot 1 in graph a, as an example, is not necessarily the same plot 1 in graph b.)

#### 4.4. Production of canopy height map

Given the suitable results of GLAS height models regarding to maximum and Lorey's heights (not forest volume), production of wall to wall canopy height maps from synergy of remote sensing (lidar, radar, optical data) and environmental data was taken under consideration. Following sections present relevant results:

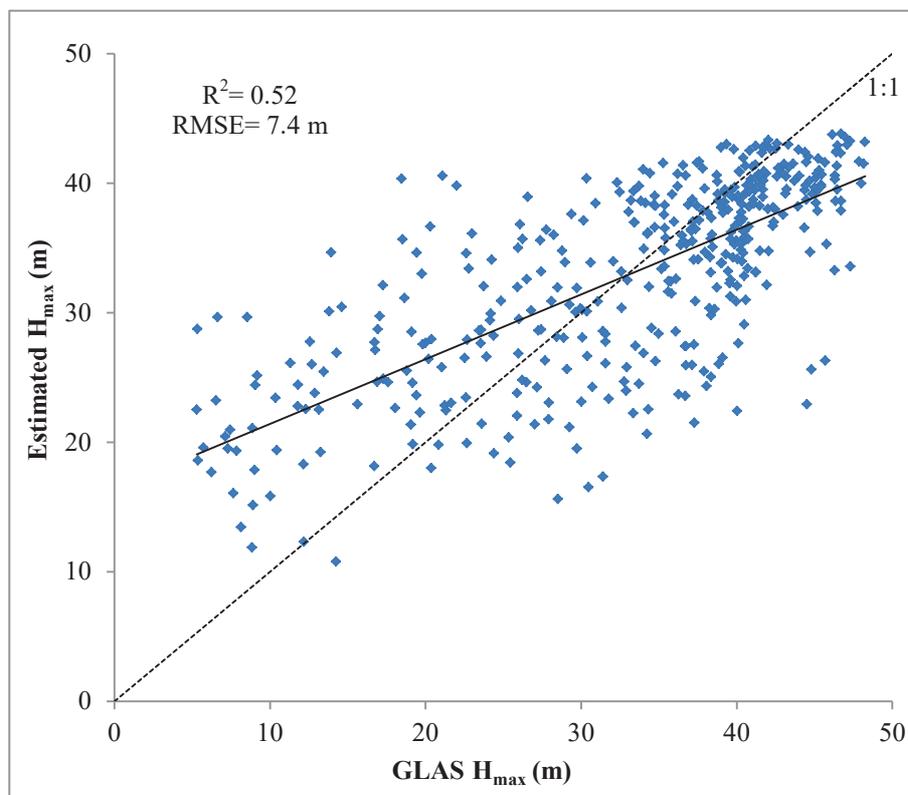
##### 4.4.1. Canopy height map using regression model

In order to produce a wall to wall height map, MLR and RF regressions were built between all GLAS derived heights, inside of the study area, and indices extracted from other remote sensing (radar and optical images) and environmental data. Concerning MLR, the most contributed variables were vegetation indices extracted from Landsat-TM and SPOT5 data including min-rvi, max-ndvi, mean-summer-ndvi and MVI related to October2009, June2010 and April2015 and also mean index from texture analysis (GLCM-mean) on optical images. The resulted MLR model showed an RMSE and  $R_a^2$  of 7.7 m and 0.46, respectively (Fig. 4.25). None of environmental variables had contribution in estimation of maximum height using this model.



**Fig. 4.25.** Estimated  $H_{\max}$  using MLR based on extracted indices from optical images versus reference  $H_{\max}$  (GLAS  $H_{\max}$ )

Concerning RF, the most important variables describing maximum canopy height with lowest error were terrain index (TI), vegetation indices extracted from Landsat-TM and SPOT5 data (min-rvi, mean-summer-ndvi, max-mvi, mvi from dates October2009, June2010 and April2015), mean index from texture analysis of optical images and also texture indices from HH and HV polarization of PALSAR data including mean, correlation and variance (hh-variance, hv-correlation, hv-mean). The RMSE and  $R_a^2$  were 7.4 m and 0.50, respectively. Figure 4.26 shows estimated maximum height using the best random forest regression based on above variables versus reference maximum height derived from GLAS (GLAS  $H_{max}$ ). This model overestimated maximum heights less than 25 m that could be as result of several limitations in height estimation. It should be noticed that the GLAS-based heights were obtained using local GLAS height models developed for a small part of the study site which will lead to height discrepancy especially in heterogeneous forests. The attempt of slope correction in this study was parameterized using only 60 field plots that do not represent all slopes conditions properly, and it could cause the broadening effect of slope on the GLAS waveforms unsolved which will lead to final estimation of height incorrectly.



**Fig. 4.26.** Estimated  $H_{max}$  using RF regression based on TI and extracted indices from optical and radar data versus reference  $H_{max}$  (GLAS  $H_{max}$ )

The fitted RF model was used to produce a wall to wall maximum canopy height map. In order to produce canopy height map with different resolutions, this process were done using variables in different resolutions. The produced maximum canopy height map with 50 meter resolution is seen in figure 4.27. Comparison of  $H_{\max}$  extracted from this map with in situ  $H_{\max}$  at the location of 32 plots, shown in the figure 4.27, produced an RMSE and  $R^2$  of 5.3 m and 0.71, respectively (Fig. 4.28). The overestimation observed in the presented diagram is expected because of firstly, the error involved in the second height model used for preparing the height map (refer to figure 4.26). Secondly, validation of height map is incomplete and is limited by a lack of field observations for many of the forested lands which may have the same height range but different topographical, environmental and structural conditions.

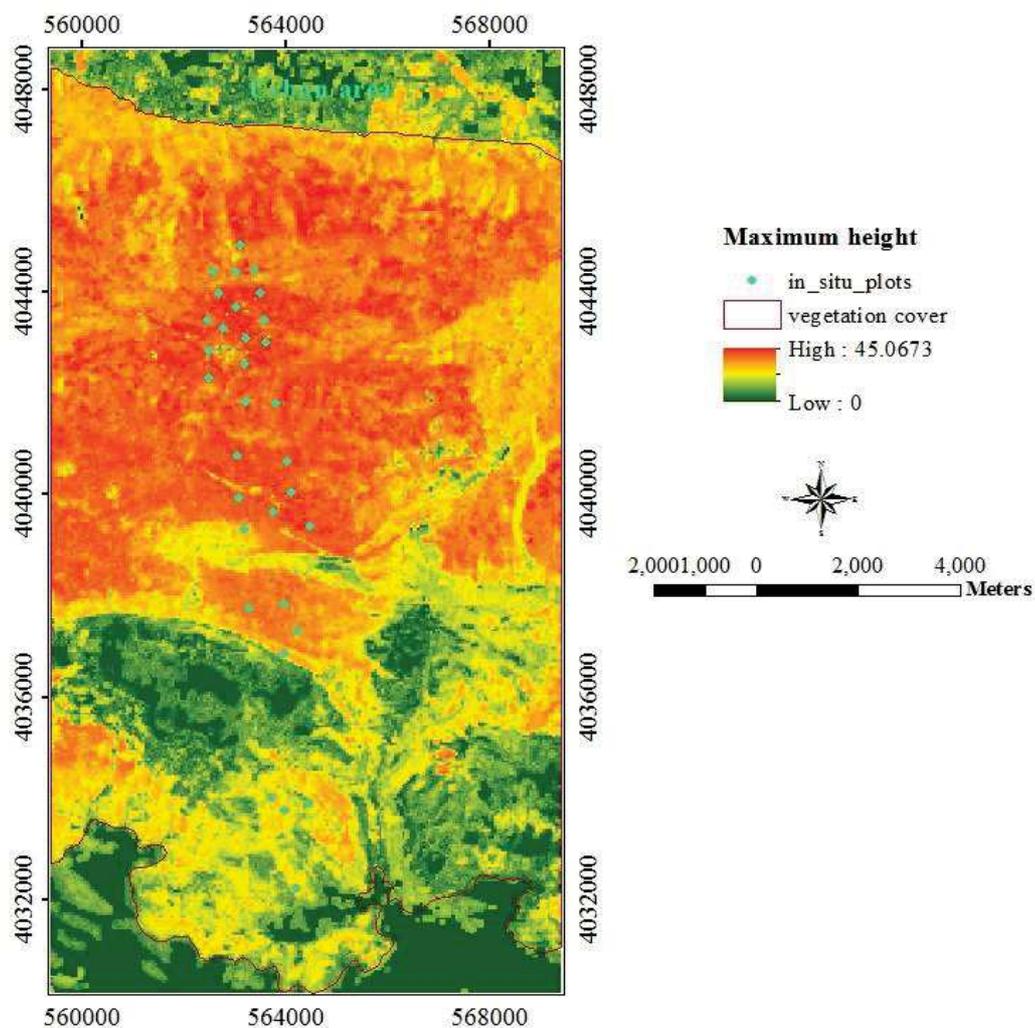
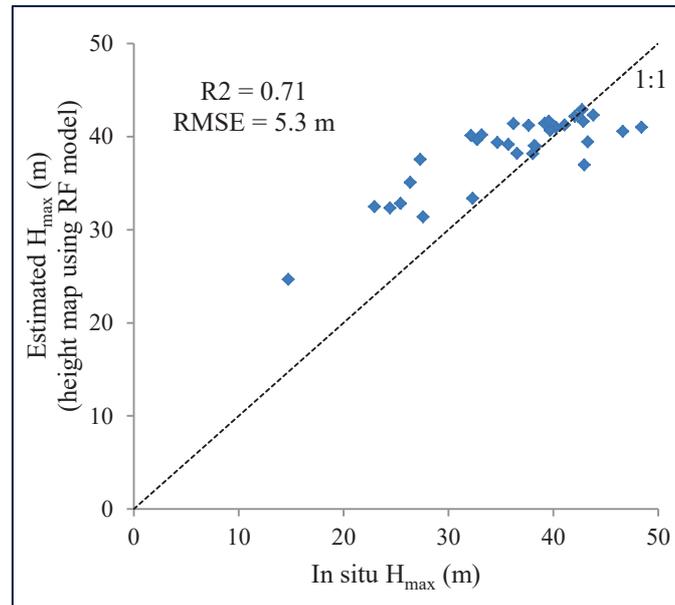
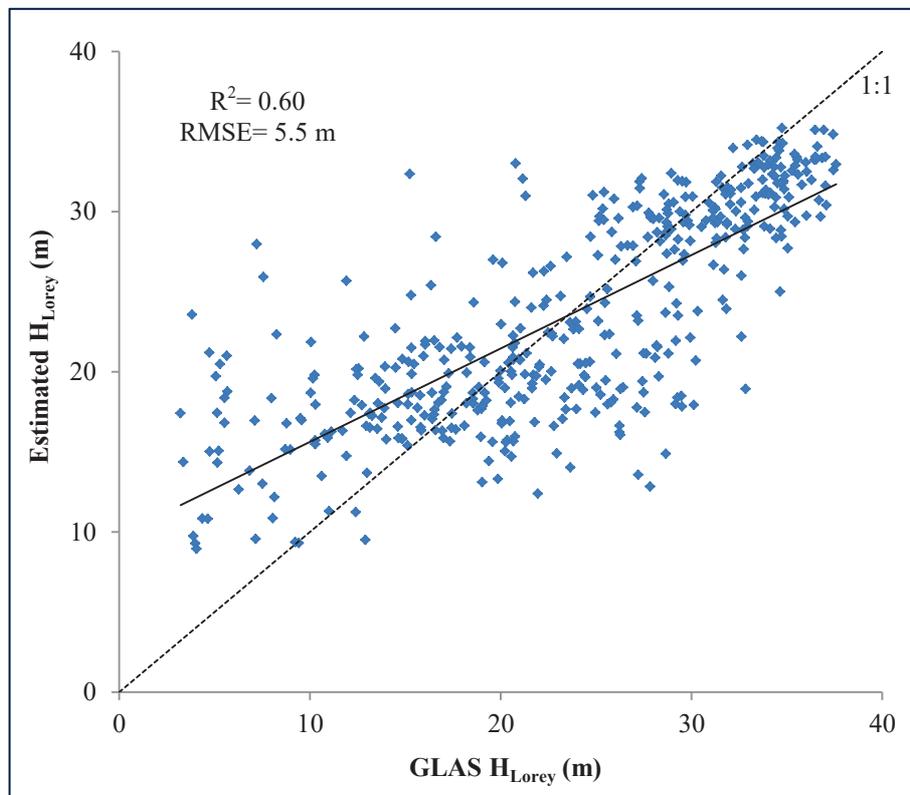


Fig. 4.27. Maximum height map produced using RF model



**Fig. 4.28.** Comparison of  $H_{\max}$  map produced from RF model with in situ  $H_{\max}$  measures at the location of 32 plots

Like maximum height, numerous MLR and RF regressions were built between Lorey's height extracted from GLAS data as reference height and indices extracted from PALSAR, Landsat-TM and SPOT5, and also environmental data. The best result was obtained using an RF model combining terrain index (TI), vegetation indices including min-ndvi, min-summer-ndvi, min-rvi, and RVI related to dates June2010 and April2015, and also texture indices including mean and homogeneity extracted from optical images (TM and SPOT5). Texture indices derived from PALSAR data did not have high importance degree on Lorey's height. This model produced an RMSE and  $R_a^2$  of 5.5 m and 0.59, respectively (Fig. 4.29).



**Fig. 4.29.** Estimated  $H_{Lorey}$  using RF regression based on extracted indices from optical images and environmental data versus reference  $H_{Lorey}$  (GLAS  $H_{Lorey}$ )

The fitted RF model was used to produce a wall to wall Lorey's height map which is observed in figure 4.30. Comparison of  $H_{Lorey}$  extracted from Lorey's height map with true  $H_{Lorey}$  values at the location of 32 in situ plots, shown in the figure 4.30, produced an RMSE and  $R^2$  of 4.3 m and 0.50, respectively. Figure 4.31 shows estimated  $H_{Lorey}$  extracted from Lorey's height map versus in situ  $H_{Lorey}$ .

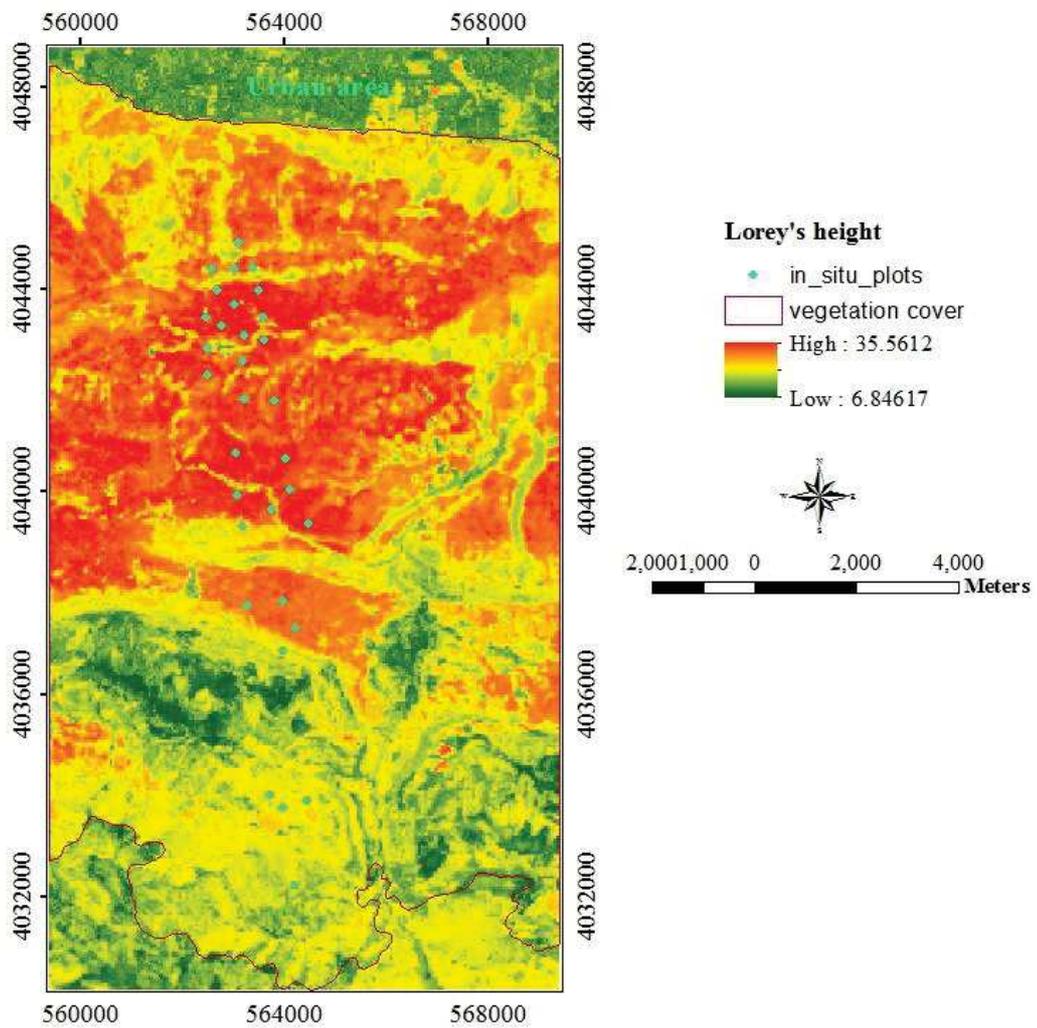
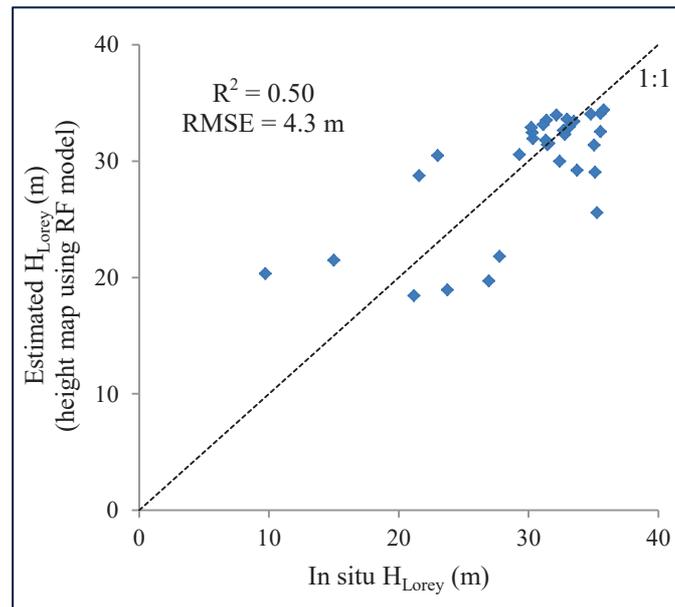


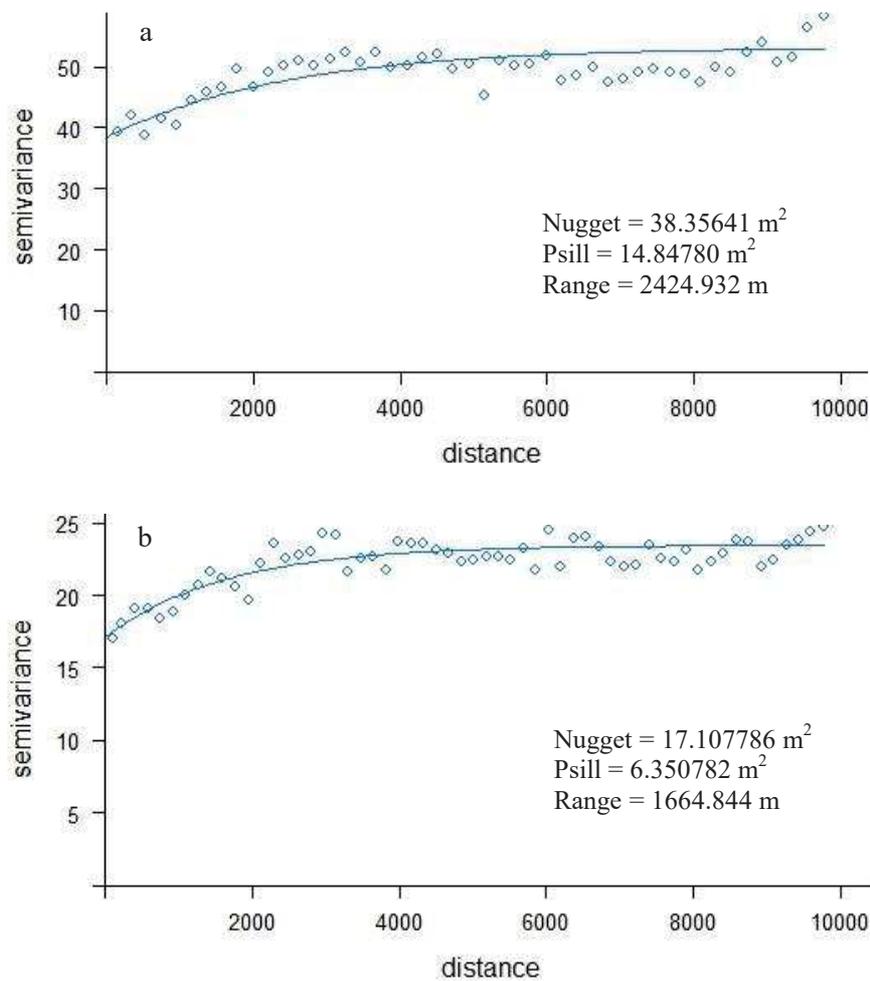
Fig. 4.30. Lorey's height map produced using RF model



**Fig. 4.31.** Comparison of  $H_{Lorey}$  map produced from RF model with in situ  $H_{Lorey}$  measures at the location of 32 plots

#### 4.4.2. Canopy height map using regression-kriging

Regression-kriging method was also used to produce canopy height map with considering spatial correlation between canopy heights. The semivariogram of height residuals (maximum height and Lorey's height) and exponential model fitted on them is observed in figure 4.32. The Nugget, Psill and range obtained from semivariogram of  $H_{max}$  residuals were  $38.35641m^2$ ,  $14.84780m^2$  and  $2424.932m$ , respectively. These coefficients for  $H_{Lorey}$  were  $17.107786m^2$ ,  $6.350782m^2$  and  $1664.844m$ . As known, the nugget effect can be attributed to measurement errors or spatial sources of variation at microscales smaller than the sampling interval (or both). In the presented variograms, the nugget effects look too high that could be as a consequence of sparse data.



**Fig. 4.32.** Examples of fitted semivariograms of a) maximum height and b) Lorey's height residuals

However the residuals of RF models did not exhibit a strong correlation structure (refer to the figure 4.32), regression-kriging method was under consideration to investigate the probability of improving height maps produced using RF models. Thus, kriging of height residuals was performed using kriging weights calculated based on information derived from height residual's semivariogram. This layer was added to the height map produced using RF regression. The resulted height maps were validated using in situ heights on 32 plots in part of study area. The RMSE and  $R^2$  were 5.9 m and 0.72, respectively, for kriged  $H_{\max}$  and 4.3 m and 0.54, respectively, for kriged  $H_{\text{Lorey}}$ , Figures 4.33 and 4.34 show, respectively, maximum canopy height map using regression-kriging method and scatter diagram of  $H_{\max}$  for the validation plots. Figures 4.35 and 4.36 represent the resulted map and validation diagram for the  $H_{\text{Lorey}}$ . As it is seen in the figures, this method did not improve heights. It has been shown earlier that nugget coefficient of the semivariograms

are too high that led to kriging estimates become overly smoothed and consequently incorrect height estimations.

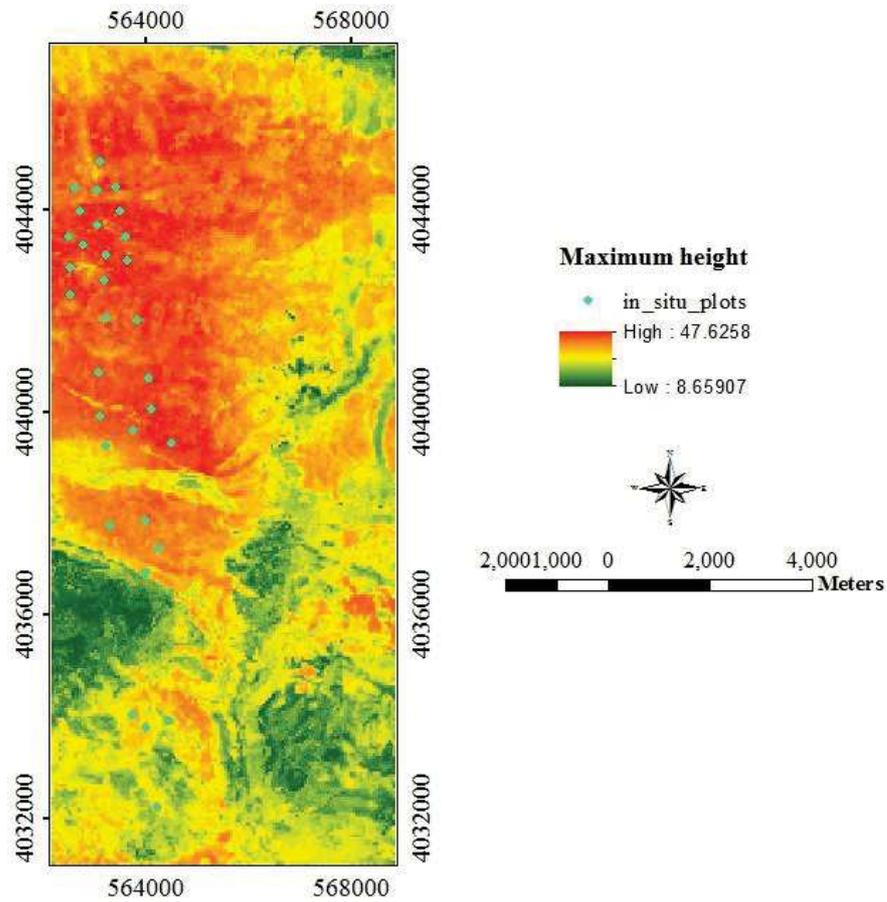


Fig. 4.33. Maximum height map produced using regression-kriging

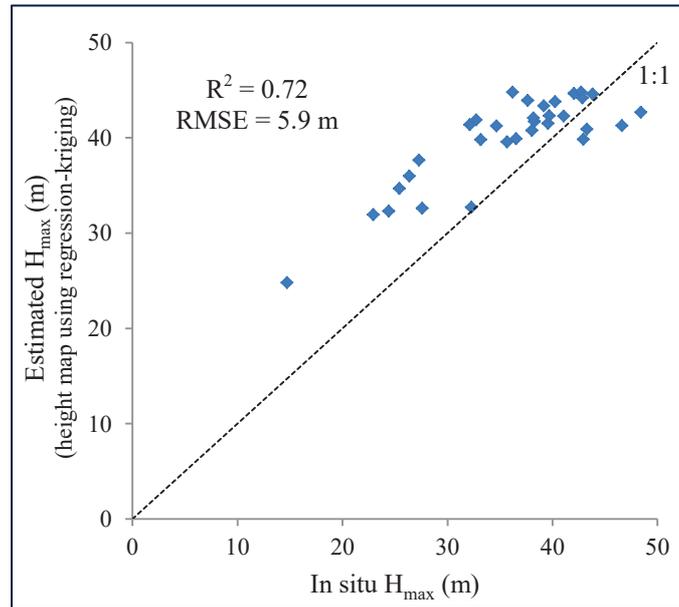


Fig. 4.34. Comparison of H<sub>max</sub> map produced from regression-kriging with in situ H<sub>max</sub> measures at the location of 32 plots

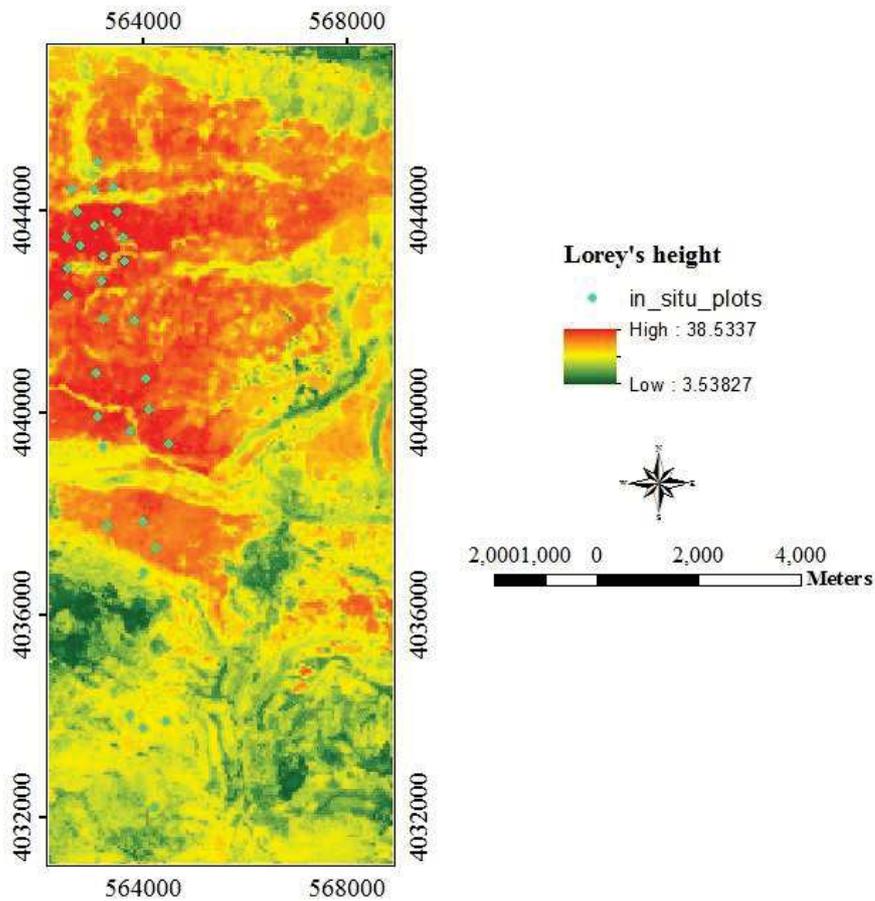
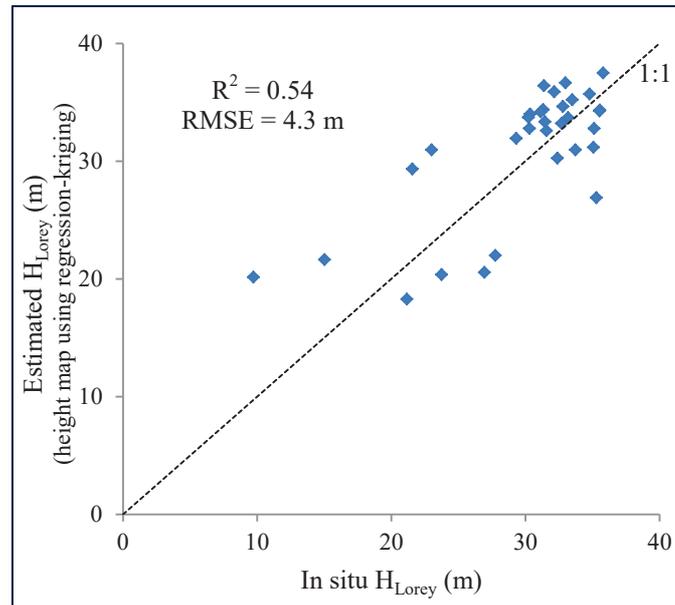


Fig. 4.35. Lorey's height map produced using regression-kriging



**Fig. 4.36.** Comparison of  $H_{Lorey}$  map produced from regression-kriging with in situ  $H_{Lorey}$  measures at the location of 32 plots

#### 4.4.3. Discussion on production of canopy height map

As explained in above sections, developing regression models between GLAS estimated heights (as reference) and other remotely sensed and environmental data was considered as the first step of providing canopy height map. For maximum height, all three vegetation indices (NDVI, RVI and MVI) derived from optical data were important. The “mean” feature from GLCM analysis on mean-summer-NDVI that was highly correlated with forest height individually, showed also high importance degree in the maximum height RF model. Among PALSAR texture features, “mean”, “correlation” and “variance” contributed in the height regression model. However the “variance” was not correlated with forest height individually but it was important for maximum height estimation in combination with other variables. Among environmental data, only TI (terrain index) as a proxy of topography condition was contributed into the model. Other variables including aspect, elevation and geological units did not affect the efficiency of the model noticeably. It is likely to achieve better result if the above environmental data is used in the GLAS-height model (first part of this study). For mean Lorey’s height, PALSAR extracted texture features did not show contribution in the best RF model. Among vegetation indices derived from optical data, only NDVI and RVI, and among texture features derived from mean-summer-NDVI, “mean” and “homogeneity” showed high importance degree. Generally TI and indices extracted from optical images had most contribution in the estimation of forest mean and maximum heights.

However the RMSE in estimation of  $H_{\text{Lorey}}$  (5.5 m) is less than  $H_{\text{max}}$  (7.4 m), both  $H_{\text{Lorey}}$  and  $H_{\text{max}}$  are predictable with similar relative error in the range of 22.8 to 23.5% using random forest regression. These models were used to produce a wall to wall height map and estimate  $H_{\text{max}}$  and  $H_{\text{Lorey}}$  at locations with no GLAS coverage. In total, based on general knowledge about our forest site and visual interpretation, the resulted maps seem logical and reliable at least for large scale studies. Comparison of predicted heights extracted from  $H_{\text{max}}$  and  $H_{\text{Lorey}}$  maps with in situ measurements in small part of study area showed improvement in RMSE for about 2.1 m and 1.2 m for  $H_{\text{max}}$  and  $H_{\text{Lorey}}$ , respectively.

In order to improve  $H_{\text{max}}$  and  $H_{\text{Lorey}}$  maps, spatial correlation of heights were considered using depiction of height residual semivariogram. Refer to figure 4.32, there was no strong spatial correlation between height residuals. The model fitted on the semivariogram was quite flat that could be a consequence of high nugget effect which itself would be resulted from error in observation data (here GLAS-based heights) or low density of data. So, the height maps resulted from regression-kriging did not improve height accuracy.

## 5. Conclusion and perspectives

### 5.1. Conclusion

Measuring biophysical parameters of forest and providing accurate information and knowledge in different scales are vital for forestry plans and ecosystem management. Remote sensing techniques can provide a less expensive and relatively precise outputs in comparison with field measurements especially for large heterogeneous forests and inaccessible area.

This research includes two main parts and aimed to: 1) investigate the capability of GLAS data in estimating forest biophysical parameters including maximum canopy height ( $H_{\max}$ ), Lorey's height ( $H_{\text{Lorey}}$ ) and forest volume ( $V$ ), 2) producing forest height/volume map using integration of ICESat/GLAS, ALOS/PALSAR and optical images and environmental data.

Concerning the first part, numerous MLR, RF and ANN regressions were developed using different sets of metrics including waveform metrics and PCs to estimate each parameter. In situ measurements were carried out to build and validate regression models. In order to overcome slope effect on GLAS waveforms, terrain index (TI) was derived from digital elevation models provided from topographic maps ( $DEM_{10}$ ) and SRTM ( $DEM_{90}$ ) and used as a predictive in regression models.

Regards to the second part, GLAS height models were used to derive  $H_{\max}$  and  $H_{\text{Lorey}}$  at the location of all GLAS data in the study area. Then several MLR and RF regression models were developed between GLAS-based heights as reference and indices extracted from PALSAR, Landsat-TM, SPOT5 and DEM to provide second height models. The resulted models were employed for production of height maps. Consequently regression-kriging procedure was implemented to consider the possibility of height values improvement.

#### 5.1.1. Prediction of maximum height ( $H_{\max}$ ) using GLAS

As expected,  $H_{\max}$  estimated from direct method (vertical distance between the signal start and the ground peak) did not match highly the real heights (RMSE=9.9m, RMSE% =27.5). It could be a cause of sloped terrain and misidentification of ground peak witch has been also stated by several researches (Lefsky *et al.*, 2005; Pang *et al.*, 2008; Chen, 2010b; Xing *et al.*, 2010).

In order to decrease effects of terrain slope on estimation of canopy height, different regression methods were employed. An MLR model combining terrain information with GLAS's waveform metrics ( $W_{\text{ext}}^{2.5}$ ,  $W_{\text{ext}}^{1.5}$ ,  $TI_{10}^{1.5}$ ,  $\text{Ln}(H_{50})$ ) produced  $H_{\max}$  with an

accuracy of 5.0 m (RMSE% = 13.8). An ANN model employing three metrics ( $W_{\text{ext}}$ ,  $TI_{10}$ ,  $H_{50}$ ) estimated  $H_{\text{max}}$  with an accuracy of 5.4 m (RMSE% = 15). The point regards to these two models is employing  $H_{50}$  which is more exposed to uncertainty rather than two other metrics. It is because the extraction of this metric depends on recognition of ground peak which is difficult on steep area and could contain error. An ANN model using only  $W_{\text{ext}}$  and  $TI_{10}$  led to an accuracy of 5.7 m (RMSE% = 16). As observed, these models outperformed direct method (with 13 percent reduction in error of estimation), and over/underestimation of height has been decreased considerably. Although overestimation is still observed in short-trees sparse forest stands ( $< 10$  m). Nelson (2010) has also showed lack of efficiency of GLAS data to accurately measure forest structure in such forests.

Generally models contained TI showed better performance in estimation of forest canopy height. It indicates that TI derived from DEM neutralizes greatly the negative effect of terrain slope on waveform's characteristics. In a study by Chen (2010a) which has been done in three different sites with terrain slope of 20 degrees on average, a simple linear model contains predictors  $W_{\text{ext}}$  and  $TI_{10}$  outperformed two other linear and non-linear models containing  $W_{\text{ext}}$ ,  $H_{\text{lead}}$  and  $H_{\text{trail}}$ .

PCs-based models (MLR, RF and ANN) did not perform as well as models based on waveform metrics. In overall, they produced better result when models include  $W_{\text{ext}}$  and TI in addition to PCs.

Comparison of three statistical methods in estimation of maximum canopy height based on waveform metrics indicated that MLR and RF represented respectively, the best and the worst performance. When the regressions were developed based on PCs, the ANN produced, slightly, better result rather than MLR, and RF did not show good result. These results were in contrast with Fayad *et al.* (2014) that observed approximately the same accuracies in predicting canopy height using MLR or RF models, also waveform metrics or PCs based models. This confirms the local applicability of fitted regressions. It is worth to notice that terrain topography in this research (Planted eucalyptus forests in French Guiana) is mostly flat.

### 5.1.2. Prediction of Lorey's height ( $H_{\text{Lorey}}$ ) using GLAS

A simple three layer ANN model using three first PCs of PCA and  $W_{\text{est}}$  predicted  $H_{\text{Lorey}}$  with an accuracy of 3.4 m (RMSE% = 12.4). While TI was an important variable in most GLAS height models, this model was able to estimate  $H_{\text{Lorey}}$  with high accuracy without necessity of participation the terrain information. Generally, ANN showed better

performance in comparison with MLR and RF when PCs were used as input variables and RF produced less accurate result comparing MLR. All three methods (MLR, RF and ANN) had approximately similar performance in terms of employing waveform metrics as predictors (RMSE about 5 m).

### 5.1.3. Effect of terrain index on estimation of height using GLAS

As discussed before, terrain index (TI) was extracted from two sources of DEM: DEM provided from 10m-topographic maps ( $TI_{10}$ ), and SRTM90 DEM ( $TI_{90}$ ). Generally models containing TI produced higher accuracy. This metric showed highest correlation with canopy height after  $W_{ext}$ , and was important in reduction of broadening effect of terrain slope on waveform metrics.

The result of regression models showed that models containing  $TI_{10}$  performed slightly better than those including  $TI_{90}$ . This is contrary to our expectations for producing much more accurate result using local DEM generated from topographic map rather than SRTM DEM. One reason could be that conventional DEMs produced from photogrammetric techniques might not adequately characterize topography over forest areas (NOAA, 2012). Conclusively, the SRTM DEM may be an acceptable source of information about terrain variability especially in large extent areas with presence of forest cover. Recent availability to the SRTM DEM30 for whole world (with more details rather than SRTM DEM90) strengthens this deduction. However, it is expected to reach higher accuracy using DEM derived from airborne lidar data which has been confirmed by Chen (2010).

### 5.1.4. Prediction of forest volume (V) using GLAS

Concerning volume predictions, two approaches was employed. The first, estimation of volume using volume-height relationship and the second, volume estimation using regressions developed between in situ volume and lidar based metrics. The result of volume- $H_{Lorey}$  ( $116.3 \text{ m}^3/\text{ha}$ ) was slightly better than PCs-based ANN model ( $119.9 \text{ m}^3/\text{ha}$ ) but ANN model performed better in very low ( $<10 \text{ m}^3/\text{ha}$ ) and very high ( $> 800 \text{ m}^3/\text{ha}$ ) volume stands. In total, the relative error of forest volume estimated using GLAS data was about 26%.

The result of this part is better than findings of Nelson *et al.* (2009), the only study on estimating forest volume using GLAS data. They predicted timber volume in central Siberia dominated by coniferous with an  $R_a^2$  of 0.75 and RMSE of  $87 \text{ m}^3/\text{ha}$  using a neural network model employing six metrics extracted from GLAS waveform ( $n=51$ ):  $\bar{h}_{med}$ : a median height which below that cumulative canopy height profile (CHP) is 50% at

maximum;  $h_{2\text{-sun}}$ : a corrected maximum height;  $h_{g1\text{-sun}}$ : height of waveform peak with the maximum amplitude above ground peak,  $f$ : the slope of the line formed by connecting the signal start point with the peak of the uppermost Gaussian return,  $r_{g3}$ : the waveform area under the 3<sup>rd</sup> Gaussian peak, and  $n_g$ : the number of Gaussian peaks in the waveform. Volume average calculated using GLAS/MODIS considering GLAS shots on all slopes ( $n=66119$ ) was about  $172 \text{ m}^3/\text{ha}$  which is about 2.5 times less than mean volume in our study site (mean of  $450 \text{ m}^3/\text{ha}$  in 60 plots). This led to higher relative error, “percentage of RMSE divided by mean volume”, in comparison with our result.

In total, however the developed ANN model improved the accuracy of volume estimations in the extreme low and high volume stands, the residual error is still high in such area. It is worth to notice that our findings are based on only 60 plots mostly (about 75%) over stands with volume ranging from 100 to  $700 \text{ m}^3/\text{ha}$  (figure 2.11, Section 2.3.1). Thus, because of low number of plots in very low and high density stands, the network may not be well trained to learn to differentiate. It is needed to investigate scrupulously while increasing the number of observations, applying other statistical methods or participating ancillary data to enhance the accuracy of forest volume estimation in the future studies.

Following section discusses on possible sources of uncertainties in estimation of forest height and volume.

### **5.1.5. Uncertainties in prediction of height and volume using GLAS**

Two main sources of uncertainties in estimation of forest height and volume using ICESat/GLAS were identified. The first one is related to the field data collection, and the other one concerning errors in extraction of waveform metrics.

The general sources of uncertainty in field data could be:

- Time interval between lidar data acquisition and field measurements which has been ignored in the present study because of having deciduous species with low growth rate in their climax age.
- In situ measurement uncertainty which mainly concerns the uncertainty on trees height due to different factors such as measuring tool, measurement procedure, skill of operator and site geography. Aside from the obvious errors associated with wrong measurements of distances or misreading the angles of top and base of trees with the clinometer, there are several less apparent sources of error that can compromise the accuracy of the tree height calculations. An error occurs where 1) the treetop is offset from the base of the tree, or 2) the top of the tree has been misidentified. Larjavaara & Muller-Landau (2013) compared tree height measured using tangent method (clinometer) with actual heights using towers

adjacent to the trees in a moist tropical forest. They measured trees with five technicians and obtained an RMSE on the tree heights of minimum 2.88 m. They concluded that these methods produced unbiased height estimates but also high level of random error. In our research, we minimized these sources of errors by walking around the tree and viewing it from different angles to distinguish the actual top from other branches. For some trees, measurements were done from different angles by two measurers to compare and justify the measurements. After all, it is expected to have an RMSE of about 2 to 3m on in situ tree height. The DBH was also measured using a caliper with 0.5cm precision.

- Concerning Lorey's height and volume, it was needed to obtain the height of all trees located in each plot. As explained in section 2.2.1, all trees were measured in case of DBH, but only eleven trees in case of height. So, height-DBH relationships were used to address this requirement (Table 2.9, section 2.3.1) which will cause some bias from the true heights. Field volume was also calculated using local volume allometric models provided by FRWO (Table 2.10, section 2.3.1) which may contain uncertainties. Following paragraphs give an overview on the uncertainties associated to the field volume calculation:

As shown in Table 2.9, the accuracy of estimating tree height from DBH is between 3 and 5.7 m. Since the RMSE of volume allometric models has not been reported by FRWO, the precision on the estimation of tree volume (V) using allometric model of *Carpinus betulus*, for instance, is calculated:

$$(V = a(DBH^2 \times H)^b) \quad (\text{refer to equation 2, table 2.10})$$

Where, V is in m<sup>3</sup>, H, and DBH are in m and cm, respectively. The relationship between the precision on the estimation of volume and the precision on height and DBH can be written as equation 5.1:

$$\frac{\Delta V}{V} = 2b \frac{\Delta DBH}{DBH} + \frac{b \Delta H}{H} \quad (5.1)$$

Where  $\Delta V/V$  is the relative precision on the estimation of volume,  $\Delta DBH/DBH$  and  $\Delta H/H$  are the relative error on DBH and height, respectively. The coefficient b is equal to 1.0432 (Table 2.10). For a DBH of 7.5 and 124cm, as minimum and maximum DBH of *Carpinus betulus*, respectively, the estimated height would be 12.4m and 30.8m (refer to height relationship for *Carpinus betulus*, table 2.9). According to a DBH accuracy ( $\Delta DBH$ ) of 0.5cm and height accuracy ( $\Delta H$ ) of 3.9m (Table. 2.9), the maximum and minimum relative error on estimation volume would be 46.7% (for low H and DBH) and 14% (for high H and DBH), respectively.

### 5.1.6. Production of canopy height map

As explained, second height regressions were developed using GLAS-based heights as reference and other remotely sensed and environmental data. In total, RF models outperformed the MLR models in predicting maximum (RMSE = 7.4 m,  $R_a^2 = 0.52$ ) and mean Lorey's height (RMSE = 5.5 m,  $R_a^2 = 0.60$ ) in this stage. Concerning  $H_{\max}$  map produced from RF model, however the accuracy is relatively suitable (RMSE = 5.3 m), comparison of heights extracted from the map with field measurements showed overestimation in samples with maximum height ranging from 23 to 35 m. It is worth to note that the total number of plots included in the assessment is 32 with  $H_{\max}$  ranging from 23 to 48 m (except one plot = 15 m) and lack of samples especially in short sparse stands. Furthermore, the total area represented by the in situ plots is very small as compared to the total area mapped. In total only 3.2 ha of the 15000 ha of forested lands were sampled. Regards to mean Lorey's height, because of lower range of values, the resulted  $H_{\text{Lorey}}$  map showed higher accuracy equal to 4.3 m rather than  $H_{\max}$ . But the relative error for  $H_{\max}$  and  $H_{\text{Lorey}}$  is similar equal to 14.8% and 14.4%, respectively. An advantage of production of  $H_{\text{Lorey}}$  map rather  $H_{\max}$  map is inclusion of less data sources. As observed before, the best regression model describing maximum canopy height includes indices extracted from DEM, PALAR and optical images (Landsat-TM and SPOT5). It is in accordance with research done in east part of Hyrcanian forests of Iran by Attarchi & Gloaguen (2014). They found that the joint use of optical and SAR data increases the reliability of the biomass model, significantly. But for Lorey's height, PALSAR extracted indices did not have contribution in the best regression.

The attempt for improving the precision of canopy height map using regression-kriging was unsuccessful in contrast with the result achieved by Fayad *et al.* (2016) that reported an improvement of about two meters in terms of RMSE for forest canopy height map. In our study, the exponential model fitted on the height residual semivariogram did not show strong spatial correlation which could be as a result of the heterogeneity of the study area in case of forest structure and topography. While Fayad *et al.* (2016) worked on relatively homogeneous flat forests.

In total, there are several limitations to production of height map for the study area. The slope correction attempted here was parameterized using only 60 field plots that do not represent all slopes conditions properly, especially in case of steepest slopes (greater than 60%). The GLAS-based heights were obtained using local GLAS height models developed for a small part of study site which will lead to height discrepancy especially in heterogeneous forests. To ensure that as many GLAS footprints as possible were included in the analysis, the data from all GLAS laser campaigns from September 2003 on were

included and processed using the same algorithms. These data were collected over a more than 5-year period, and therefore do not reflect a static moment in time. Lastly, the validation of the height maps is incomplete and is limited by a lack of field observations for many of the forested lands within study site.

Several other studies have used GLAS data to derive canopy height over large regions, but have typically combined them with 250 m resolution MODIS data rather than 10 to 30 m resolution spot5 and Landsat-TM data. These studies typically developed products that are global assessments of canopy height at a coarse scale (Lefsky, 2010; Simard *et al.*, 2011; Fayad *et al.*, 2016). In this study, we demonstrated GLAS data are also useful in mapping at finer resolution, although subject to the limitations identified herein. However further work needs to be done to thoroughly understand and quantify the various sources of error underlying the lack of correspondence between the field observations and the mapped canopy height values, this map provides a good understanding of the distribution of forest canopy height across the study area in a short time and at the lowest cost.

### 5.1.7. General conclusion

In this research, capability of ICESat GLAS was investigated for estimation of forest maximum canopy height, mean Lorey's height and forest volume in part of Hyrcanian forests of Iran. It was also subjected to provide forest canopy height and volume map if it is possible.

Based on the result, GLAS was able to estimate maximum canopy height and Lorey's height with an accuracy of about 5m and 3.4m, respectively. In contrary to the suitable results related to forest heights (maximum and mean), GLAS data did not meet the desirable achievements in estimation of forest volume. These results address the first main question of this study about ability of GLAS data in estimation of forest biophysical parameters in mountainous heterogeneous forests of north of Iran.

As it was observed in the result, two metrics of waveform extent ( $W_{ext}$ ) and terrain index (TI) had key role in estimation of GLAS-based heights.  $W_{ext}$  as vertical distance between signal start and end of a waveform is directly related to the canopy height over flat area, but it is extended by increasing of the terrain slope. It challenges the height derivation over severe topography. In this research, using regression models and digital elevation model helped to overcome the impact of terrain slope on waveform characteristics. Generally, all regression models containing TI extracted from DEM outperformed models not-including TI. It confirms the importance of this metric in estimation of height using GLAS data which was under question in our study area. Although, further research needs to be done to address the impacts of slope on height recovery in Hyrcanian forests of Iran using GLAS data.

Parametric and non-parametric statistical methods of multiple linear regression, random forest and artificial neural network were employed for GLAS-based height/volume estimation. Concerning the  $H_{max}$ , the best result was obtained using MLR based on waveform metrics. An ANN model based on waveform metrics produced also relatively similar result. Regards to  $H_{Lorey}$ , PCs based neural network models outperformed two other statistical methods. A neural network model based on PCs performed also better than MLR and RF regressions in estimation of forest volume. Thus random forest showed the weakest performance in GLAS based forest parameter derivation in our study area.

As a consequence of desirable results concerning height estimation (both  $H_{max}$  and  $H_{Lorey}$ ), production of height maps was under consideration using regression models and regression-kriging method. Thus a synergy of GLAS, PALSAR, Landsat-TM, SPOT5 and environmental data were used in the analysis. Comparison of resulted maps from regression models with field observations (in a small part of the study area near to GLAS

---

footprints locations) showed promising outcomes. However, further work needs to be done to thoroughly understand and quantify the various sources of error underlying the lack of correspondence between the field observations and the mapped canopy height values. The regression-kriging method did not improve the former height maps that could be as a consequence of heterogeneity of the study area. By this end, the second important question of this research related to the possibility of forest height/volume map production using combination of GLAS and other remotely sensed data was answered.

However terrain slope is very important in estimation of forest parameters using GLAS data, consideration of forest characteristics such as forest type and horizontal and vertical structure of forest may influence the quality of estimations. For instance, it is expected to achieve better results in pure even-aged forest that have simpler vertical structure rather than mixed uneven-aged forest.

With the end of the GLAS data collection in 2009 no new spaceborne lidar data are currently available for updating results. The ICESat-2 mission, which will provide new spaceborne lidar data using the Advanced Topographic Laser Altimeter System, is scheduled for launch in 2017.

## 5.2. Perspectives

- The Hyrcanian forest is a unique natural heritage of global importance. The Caspian region harbors the world's last remaining primary forests of the temperate deciduous forest formation. This research was done on a small area with about 15000 ha forest cover of 1.8 million hectare of Hyrcanian forest. Given the complexity of this forest in terms of vertical and horizontal structures, forest types, forest density, topography etc., it is suggested to do supplementary researches using spaceborn lidar in other parts of mixed broadleaved Hyrcanian forests.
- This research took into account the terrain slope in the estimation of canopy height using GLAS data. Since many other forest characteristics (i.e. forest type, mixture percent and forest age) may affect the quality of estimations, it is interesting to consider these factors in the future studies. However, based on the attained experiences, such a detailed research needs denser coverage of lidar data which is addressed in the upcoming ICESat-2 technology.
- GLAS height models were developed only using leaf-on season lidar data and was applied to all time lidar data. This was a limitation of our research because of low density of GLAS data over study area. As stated by Pang *et al.* (2008), the summer period GLAS waveforms capture the returns from forest canopy. The data from early stage of autumn period still contain enough returns from forest canopy, even with lower intensity. The spring period and late autumn period data contain less signals from forest canopy and difficult to estimate forest height. Therefore, it is expected to improve estimations by performing the analysis separately using leaf-on and leaf-off season data in the deciduous broadleaved forests.
- For the second part of this research (providing forest height map from synergy of GLAS, PALSAR, optical and environmental data) some environmental data such as geological units, aspect and elevation maps were used in addition to TI and slope as predictors in building second height models. Among these variables, only TI showed good contribution in the models. This predictor was also used in the first part of this study which was detecting capability of GLAS data for retrieving mean and maximum height. It is probable that considering other environmental data rather than TI in the first step (GLAS height models), lead to good contribution of them in the second step (second height model and finally height map). However as mentioned before, such investigations require dense coverage of lidar data so that there be enough lidar samples representing different conditions or classes of an environmental data.

- 
- In order to provide forest canopy height map using regression model (non-spatial method), MLR and RF models were developed between GLAS-based heights and pre-mentioned remotely sensed data. The selected models and consequently the resulted maps showed overestimation especially for maximum height. It could be as a result of inability of GLAS in estimating short trees height ( $<10\text{m}$ ) which leads to error in reference heights (less than 10m) estimated from GLAS and consequently the fitted RF model based on optical, radar and environmental data overestimates also tree's height shorter than 20 m. But it is also suggested to test other statistical regression methods.
  - A limitation of this research was lack of field observations describing all conditions of the study area. Indeed, a larger database (ranges of slopes, height, etc.) would be very useful to better understand the limitations of the proposed methods. Rather than laborious, cost and time consuming field work, some points in a mountainous forest are inaccessible. Therefore, it would be interesting to employ airborne lidar to collect information in such condition. By the way, a comprehensive database of field observations needs also a denser coverage of lidar data which is addressed in the upcoming ICESat-2 technology.



## 6. References

- Abshire, J.B., X. Sun, H. Riris, J.M. Sirota, J.F. McGarry, S. Palm, D. Yi & P. Liiva, 2005. Geoscience laser altimeter system (GLAS) on the ICESat mission: On-orbit measurement performance. *Geophysical Research Letters*, 32(21): 1-4.
- Ahmadi, K., S.J. Alavi, M.T. Kouchaksaraei & W. Aertsen, 2013. Non-linear height-diameter models for oriental beech (*Fagus orientalis* Lipsky) in the Hyrcanian forests, Iran. *Biotechnology, Agronomy, Society and Environment*, 17(3): 431-440.
- Amini, J. & J. Tetuko Sri Sumantyo, 2011. SAR and optical images for forest biomass estimation, in *Biomass - Detection, Production and Usage*, Dr. Darko Matovic (Ed.), ISBN: 978-953-307-492-4, InTech, pp 53-74.
- Andersen, H.E., S.E. Reutebuch & R.J. McGaughey, 2006. A rigorous assessment of tree height measurements obtained using airborne lidar and conventional field methods, *Can. J. Remote Sensing*, 32(5): 355-366.
- Aronoff, S., 2005. *Remote Sensing for GIS Managers*, ESRI Press, 487 PP.
- Asner, G.P. & J. Mascaro, 2014. Mapping tropical forest carbon: Calibrating plot estimates to a simple lidar metric. *Remote Sensing of Environment*, 140: 614-624.
- Attarchi, S. & R. Gloaguen, 2014. Improving the estimation of above ground biomass using dual polarimetric PALSAR and ETM+ data in the hyrcanian mountain forest (Iran). *Remote Sensing*, 6: 3693-3715.
- Baghdadi, N., G. Maire, I. Fayad, J.S. Bailly, Y. Nouvellon, C. Lemos & R. Hakamada, 2014. Testing different methods of forest height and aboveground biomass estimations from ICESat/GLAS data in Eucalyptus plantations in Brazil. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(1): 290-299.
- Baghdadi, N., G. Le Maire, K. Osé, Y. Nouvellon, Z. Mehrez, C. Lemos & R. Hakamada, 2014. Evaluation of ALOS/PALSAR L-band data for the estimation of Eucalyptus plantations aboveground biomass in Brazil. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(8): 3802-3811.
- Bailey, D.L. & D. Thompson, 1990. Developing neural network applications, *AI Expert*, 33-41.
- Balzter, H., C.S. Rowland & P. Saich, 2007. Forest canopy height and carbon estimation at Monks wood national nature reserve, UK, using dual-wavelength SAR interferometry. *Remote Sensing Environment*, 108: 224-239.
- Batista, J.L.F., H.T.Z. Couto & M. Marquesini, 2001. Performance of height-diameter relationship models: analysis in three forest types. *Scientia Forestalis*, 60: 149-163.
- Boudreau, J., R.F. Nelson, H.A. Margolis, A. Beaudoin, L. Guindon & D.S. Kimes, 2008. Regional aboveground forest biomass using airborne and spaceborne lidar in Québec. *Remote Sensing of Environment*, 112: 3876-3890.
- Breiman, L., 2001. Random Forests. *Machine Learning*, 45: 5-32.
- Breiman, L. 1994. Bagging predictors, Technical Report No. 421, University of California, 19pp.
- Brenner, A.G., R.J. Zwally, C.R. Bentley, B.M. Csatho, D.J. Harding, M.A. Hofton, J.B. Minster, L. Roberts, J.L. Saba, R.R. Thomas & D. Yi, 2003. Derivation of range and range distributions from laser pulse waveform analysis for surface elevations, roughness, slope, and vegetation heights, Geoscience Laser Altimeter System (GLAS), Algorithm Theoretical Basis Document, Version 4.1: 92pp.

- Burnham, K.P. & D.R. Anderson, 2002. Information and likelihood theory: a basis for model selection and inference, in *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd ed. New York: Springer-Verlag Press, 49-97.
- Cairns, M., J., Barker, R. Shea & P. Haggerty, 1995. Carbon dynamics of Mexican tropical evergreen forests: influence of forestry mitigation options and refinement of carbon-flux estimates. *Interciencia*, 20(6): 401-408.
- Cameron, A.C. & F.A.G. Windmeijer, 1995. R-Squared measures for count data regression models with applications to health care utilization. *Journal of Business and Economic Statistics*, 1-35.
- Cannon, A.J., 2012. Package 'monmlp': Monotone multi-layer perceptron neural network. *Comprehensive R Archive Network (CRAN)*, 1-14.
- Chai, T. & R.R. Draxler, 2014. Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscience Model Development*, 7: 1247-1250.
- Chen, Q., 2010a. Assessment of terrain elevation derived from satellite laser altimetry over mountainous forest areas using airborne lidar data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(1): 111-122.
- Chen, Q., 2010b. Retrieving vegetation height of forests and woodlands over mountainous areas in the Pacific Coast region using satellite laser altimetry. *Remote Sensing of Environment*, 114: 1610-1627.
- Chen, G. & G.J. Hay, 2011. A support vector regression approach to estimate forest biophysical parameters at the object level using airborne lidar transects and QuickBird data. *Photogrammetric Engineering & Remote Sensing*, 77(7): 733-741.
- Cheng, X., M. Yu & T. Wu, 2013. Effect of forest structural change on carbon storage in a coastal metasequoia glyptostroboides stand. *The Scientific World Journal*, 830509: 1-9.
- Colbert, K.C., D.R. Larsen & J.R. Lootens, 2002. Height-diameter equations for thirteen midwestern bottomland hardwood species. *Northern Journal of Applied Forestry*, 19(4): 171-176.
- Cutler, M.E.J., D.S. Boyd, G.M. Foody & A. Vetrivel, 2012. Estimating tropical forest biomass with a combination of SAR image texture and Landsat TM data: An assessment of predictions between regions, *ISPRS Journal of Photogrammetry and Remote Sensing*, 70: 66-77.
- Ding, H., P.A.L.D. Nunes & S. Teelucksingh, 2011. European forests and carbon sequestration services: an economic assessment of climate change impacts, ESE, UNEP, No. 9.
- Drake, J.B., R.O. Dubayah, D.B. Clark, R.G. Knox, J. B. Blair, M.A. Hofton, R.L. Chazdon, J.F. Weishampel, S.D. Prince, 2002. Estimation of tropical forest structural characteristics using large-footprint lidar. *Remote Sensing of Environment*, 79: 305- 319.
- Duncanson, L.I., K.O. Niemann & M.A. Wulder, 2010. Estimating forest canopy height and terrain relief from GLAS waveform metrics. *Remote Sensing of Environment*, 114: 138-154.
- Duncanson, L.I., 2009. Aboveground biomass estimation using spaceborne lidar in managed conifer forests in south central British Columbia, Master of Science dissertation, University of Victoria, 116 pp.
- Duong, H.V., 2010. Processing and Application of Icesat Large Footprint Full Waveform Laser Range Data. A dissertation of doctor of philosophy, Asian Institute of Technology, Thailand, 193pp.

- Duong, H., R. Lindenbergh, N. Pfeifer & G. Vosselman, 2009. ICESat full-waveform altimetry compared to airborne laser scanning altimetry over the Netherlands. *IEEE Transactions on Geoscience and Remote Sensing*, 47(10): 3365-3378.
- Eldeiry, A. & L.A. Garcia, 2010. Comparison of regression kriging and cokriging techniques to estimate soil salinity using Landsat images, *Journal of Irrigation and Drainage Engineering*, 136(6): 355-364.
- Fang, Z. & R.L. Bailey, 1998. Height-diameter models for tropical forests on Hainan Island in southern China. *Forest Ecology and Management*, 110: 315-327.
- Fayad, I., N. Baghdadi, J-S. Bailly, N. Barbier, V. Gond, B. Hérault, M. El Hajj, F. Fabre & J. Perrin, 2016. Regional scale rain-forest height mapping using regression-kriging of spaceborne and airborne lidar data: application on French Guiana, *Remote Sensing*, 8(240): 1-18.
- FAO, 2007. Carbon sequestration in drylands, report on global soil resources 102, Food and Agriculture Organization of the United Nations, 120 pp.
- Fayad, I., N. Baghdadi, J-S. Bailly, N. Barbier, V. Gond, M. El Hajj, F. Fabre & B. Bourguine, 2014. Canopy height estimation in French Guiana with lidar ICESat/GLAS data using principal component analysis and random forest regressions. *Remote Sensing*, 6: 11883-11914.
- Freitas, S.R., M.C.S. Mello & C.B.M. Cruz, 2005. Relationships between forest structure and vegetation indices in Atlantic Rainforest, *Forest Ecology and Management*, 218: 353-362.
- Fu, A., S. Guoqing & G. Zhifeng, 2009. Estimating forest biomass with GLAS samples and MODIS imagery in northeastern China, *Proceedings of SPIE*, 7498: 1-8.
- Gagnon, L. & A. Jouan, 1997. Speckle Filtering of SAR Images - A comparative study between complex-wavelet-based and standard filters, *SPIE Proc. #3169, conference "Wavelet Applications in Signal and Image Processing V"*, San Diego, 1997.
- Garestier, F., P. Dubois-Fernandez & I. Champion, 2008. Forest height inversion using high-resolution P-band Pol-InSAR data. *IEEE, TGRS*, 46: 3544-3559.
- Ge, Y., J.A. Thomasson, R. SUI & J. Wooten, 2011. Regression-kriging for characterizing soils with remote sensing data, *Frontiers of Earth Science*, 5(3): 239-244.
- Ghasemi, N., M.R. Sahebi, A. Mohammadzadeh & R. Bayani, 2011. Mapping biomass of a temperate deciduous forest using SAR data. *Geomatics 90 - National Conference & Exhibition - May 2011, Iran*.
- Gleason, C.J. & J. Im, 2012. Forest biomass estimation from airborne lidar data using machine learning approaches, *Remote Sensing of Environment*, 125: 80-91.
- Gobakken, T., O.M. Bollandsås & E. Næsset, 2015. Comparing biophysical forest characteristics estimated from photogrammetric matching of aerial images and airborne laser scanning data, *Scandinavian Journal of Forest Research*, 30(1): 73-86.
- Goovaerts, P., 1997. *Geostatistics for Natural Resources Evaluation*, Oxford University Press, 483pp.
- Grömping, U., 2009. Variable importance assessment in regression: linear regression versus random forest, *The American Statistician*, 63(4): 308-319.
- Günther, F. & S. Fritsch, 2010. Neuralnet: training of neural networks. *The R Journal*, 2(1): 30-38.
- Hayashi, M., N. Saigusa, Y. Yamagata & T. Hirano, 2015. Regional forest biomass estimation using ICESat/GLAS spaceborne lidar over Borneo, *Carbon Management*, 6(1-2): 19-33.

- Hamburg, S.P., D.G. Zamolodchikov, G.N. Korovin, V.V. Nefedjev, A.I. Utkin, J.I. Gulbe & T.A. Gulbe, 1997. Estimating the carbon content of Russian forests: a comparison of phytomass/volume and allometric projections. *Mitigation and Adaptation Strategies for Global Change*, 2: 247-265.
- Harding, D.J. & C.C. Carabajal, 2005. ICESat waveform measurements of within-footprint topographic relief and vegetation vertical structure. *Geophysical Research Letters*, 32: L21S10.
- Healey, S.P., P.L. Patterson, S. Saatchi, M.A. Lefsky, A.J. Lister & E.A. Freeman, 2012. A sample design for globally consistent biomass estimation using lidar data from the Geoscience Laser Altimeter System (GLAS). *Carbon Balance and Management*, 7:10.
- Hilbert, C. & Ch. Schmullius, 2012. Influence of surface topography on ICESat/GLAS forest height estimation and waveform shape. *Remote Sensing*, 4: 2210-2235.
- Hodgson, M.E., J.R. Jensen, L. Schmidt, S. Schill & B. Davis, 2003. An evaluation of lidar- and IFSAR-derived digital elevation models in leaf-on conditions with USGS Level 1 and Level 2 DEMs, *Remote Sensing of Environment*, 84: 295-308.
- Horning, N., 2010. Random Forests: An algorithm for image classification and generation of continuous fields data sets, International conference on geoinformatics for spatial infrastructure development in earth and allied sciences, 1-6.
- Huang, S., S.J. Titus & D.P. Wiens, 1992. Comparison of nonlinear height-diameter functions for major Alberta tree species. *Canadian Journal of Forest Research*, 22: 1297-1304.
- Husch, B., T.W. Beers, J.A. Kershaw, 2003. Forest Mensuration, 4th Edition, John Wiley & Sons, 456 pp.
- Husson, F., J. Josse, S. Le & J. Mazet, 2015. Package 'FactoMineR': Multivariate exploratory data analysis and data mining. *Comprehensive R Archive Network (CRAN)*, 1-95.
- Hyndman, R.J. & A.B. Koehler, 2005. Another look at measures of forecast accuracy, Department of economics and business, Statistics Monash University, Australia, ISSN 1440-771X.
- IPCC, 2001. Climate change: the scientific basis. Cambridge, UK, Cambridge University Press.
- Iqbal, I.A., 2010. Evaluating the Potential of Icesat/GLAS Data to Estimate Canopy Height in the New Forest National Park, UK., Master of science dissertation, University of Twente, 81 pp.
- Ito, N., T. Hamazaki & K. Tomioka, 2001. ALOS/PALSAR characteristics and status, Proceedings of CEOS Workshop, Tokyo, April 2-5, 2001, CEOS-SAR01-087.
- Japan Space Systems, 2012. PALSAR User's Guide, 2nd edition, PALSAR User Service, 69pp.
- Kaastra, I. & M. Boyd, 1996. Designing a neural network for forecasting financial and economic time series. *Neurocomputing*, 10: 215-236.
- Kalbi, S., A. Fallah, SH. Shataee & DJ. Oladi, 2013. Estimation of forest structural attributes using ASTER data. *Iranian Journal of Natural Resources*, 65(4): 461-474.
- katz, J.O., 1992. Developing neural network forecasters for trading. *Technical analysis for stocks and commodities*, 10(4): 160-168.
- Kayitakire, F., C. Hamel & P. Defourny, 2006. Retrieving forest structure variables based on image texture analysis and IKONOS-2 imagery, *Remote Sensing of Environment*, 102: 390-401.
- Khalefa, E., I.P.J. Smit, A. Nickless, S. Archibald, A. Comber & Heiko Balzter, 2013. Retrieval of Savanna vegetation canopy height from ICESat-GLAS spaceborne lidar with terrain correction. *IEEE Geoscience and Remote Sensing Letters*, 10(6): 1439-1443.

- Khorrani, R., A.A. Darvishsefat & M. Namiranian, 2008. Investigation on the capability of Landsat7 ETM+ data for standing volume estimation of beech stands (Case study: Sangdeh forests). *Iranian Journal of Natural Resources*, 60(4): 1281-1289.
- Khorrani, R.A., Darvishsefat, A.A., Tabari Kochaksaraei, M. and Shataee Jouybari, Sh., 2014. Potential of lidar data for estimation of individual tree height of *Acer velutinum* and *Carpinus betulus*. *Iranian Journal of Forest*, 6(2): 127-140 (In Persian).
- Kuplicha, T.M., P.J. Curran & P.M. Atkinson, 2005. Relating SAR image texture to the biomass of regenerating tropical forests. *International Journal of Remote Sensing*, 26(21): 4829-4854.
- Larjavaara, M. & H.C. Muller-Landau, 2013. Measuring tree height: a quantitative comparison of two common field methods in a moist tropical forest, *Methods in Ecology and Evolution*, 1-9.
- Larsen, D.R. & D.W. Hann, 1987. Height-diameter equations for seventeen tree species in southwest Oregon. Oregon State University, Research Paper 49.
- Lasky, J.R., M. Uriarte, V.K. Boukili, D.L. Erickson, W.J. Kress & R.L. Chazdon, 2014. The relationship between tree biodiversity and biomass dynamics changes with tropical forest succession, *Ecology Letters*, 17: 1158-1167.
- Lee, S., W. Ni-Meister, W. Yang & Q. Chen, 2011. Physically based vertical vegetation structure retrieval from ICESat data: validation using LVIS in White Mountain National Forest, New Hampshire, USA. *Remote Sensing of Environment*, 115: 2776-2785.
- Lefsky, M.A., 2010. Global forest canopy height map from the moderate resolution imaging spectroradiometer and the geoscience laser altimeter system, *Geophysical Research Letters*, 37.
- Lefsky, M.A., M. Keller, Y. Pang, P.B. de Camargo & M.O. Hunter, 2007. Revised method for forest canopy height estimation from Geoscience Laser Altimeter System waveform, *Journal of Applied Remote Sensing*, 1: 1-18.
- Lefsky, M.A., D.J. Harding, M. Keller, W.B. Cohen, C.C. Carabajal, F.D. Espirito-Santo, M.O. Hunter, R. de Oliveira & P.B. de Camargo, 2005. Estimates of forest canopy height and aboveground biomass using ICESat, *Geophysical Research Letters*, 32(22): 1-4.
- Liaw, A. & M. Wiener, 2014. Package 'randomForest': Breiman and Cutler's random forests for classification and regression, CRAN, 1-29.
- Liaw, A. & M. Wiener, 2002. Classification and regression by randomforest. *R News*, 2(3): 18- 22.
- Los, S.O., J.A.B. Rosette, N. Kljun, P.R.J. North, L. Chasmer, J.C. Suarez, C. Hopkinson, R.A. Hill, E. van Gorsel, C. Mahoney & J.A.J. Berni, 2012. Vegetation height and cover fraction between 60°S and 60°N from ICESat GLAS data, *Geoscientific Model Development*, 5, 413-432.
- Lu, D., Q. Chen, G. Wang, E. Moran, M. Batistella, M. Zhang, G. V. Laurin & D. Saah, 2012. Aboveground forest biomass estimation with Landsat and lidar data and uncertainty analysis of the estimates, *International Journal of Forestry Research*, 1-16.
- Macauley, M., D. Morris, R. Sedjo, K. Farley & B. Sohngen, 2009. *Forest Measurement and Monitoring: Technical Capacity and How Good Is Good Enough?*, Washington DC: Resources for the Future.
- Makridakis, S. & M. Hibon, 1995. *Evaluating Accuracy (or Error) Measures*, INSEAD, Fontainebleau, France, 31pp.
- Mermoz, S., Th. Le Toana, L. Villarda, M. Réjou-Méchain & J. Seifert-Granzin, 2014. Biomass assessment in the Cameroon savanna using ALOS PALSAR data. *Remote Sensing of Environment*, 155: 109-119.

- Metzel, R., 2016. Forests in the cop21 climate change agreement: momentum on and mention of forests in the climate change text, Available at: <https://environment.yale.edu/blog/2016/01/forests-in-the-cop21-climate-change-agreement-momentum-on-and-mention-of-forests-in-the-climate-change-text>.
- Miller, D.R., Ch.P. Quineb & W. Hadley, 2000. An investigation of the potential of digital photogrammetry to provide measurements of forest characteristics and abiotic damage, *Forest Ecology and Management*, 135: 279-288.
- Mitchard, E.T.A., S.S. Saatchi, L.J.T. White, K.A. Abernethy, K.J. Jeffery, S.L. Lewis, M. Collins, M.A. Lefsky, M.E. Leal, I.H. Woodhouse & P. Meir, 2012. Mapping tropical forest biomass with radar and spaceborne lidar in Lop'è National Park, Gabon: overcoming problems of high biomass and persistent cloud. *Biogeosciences*, 9: 179-191.
- Mohammadi, J., 2013. Estimation of quantitative characteristics of forest structure using a combination of airborne lidar and digital aerial photos (Case Study: broadleaf forests Shastkalate Gorgan). A dissertation of doctor of philosophy, Gorgan University of Agricultural Sciences & Natural Resources, 241p (In Persian).
- Mount, N., G. Harvey, P. Aplin & G. Priestnall, 2008. Representing, Modeling, and Visualizing the Natural Environment, CRC Press, 416 pp.
- Namiranian, M., 2007. Measurement of Tree and Forest Biometry, 1st Edition, University of Tehran, ISBN: 964-03-5439-2, 574pp (In Persian).
- Nelson, R., 2010. Model effects on GLAS-based regional estimates of forest biomass and carbon. *International Journal of Remote Sensing*, 31(5): 1359-1372.
- Nelson, R., K.J. Ranson, G. Sun, D.S. Kimes, V. Kharuk & P. Montesano, 2009. Estimating Siberian timber volume using MODIS and ICESat/GLAS. *Remote Sensing of Environment*, 113: 691-701.
- Neuenschwander, A.L., T.J. Urban, R. Gutierrez & B.E. Schutz, 2008. Characterization of ICESat/GLAS waveforms over terrestrial ecosystems: Implications for vegetation mapping. *Journal of Geophysical Research*, 113(G02S03): 1-18.
- Nichol, J.E. & Md.L.R. Sarker, 2011. Improved biomass estimation using the texture parameters of two high-resolution optical sensors. *IEEE Transactions on Geoscience and Remote Sensing*, 49(3): 930-948.
- NOAA Coastal Services Center, 2012. lidar 101: An introduction to lidar technology, data, and applications, Revised. Charleston, SC: NOAA Coastal Services Center, 76pp.
- NSIDC, 2014. NSIDC Distributed ICESat GLAS Laser Operations Periods: Latest Release, Available at: [https://nsidc.org/sites/nsidc.org/files/files/glas\\_laser\\_ops\\_attrib.pdf](https://nsidc.org/sites/nsidc.org/files/files/glas_laser_ops_attrib.pdf)
- NSIDC, 2012. GLAS altimetry product usage guidance, Available at: [http://nsidc.org/data/docs/daac/glas\\_altimetry/pdf/NSIDC\\_AltUserGuide\\_Rel33.pdf](http://nsidc.org/data/docs/daac/glas_altimetry/pdf/NSIDC_AltUserGuide_Rel33.pdf).
- Osman, E.M.H., E.Z.A. Idris & E.M.M. Ibrahim, 2013. Modelling height-diameter relationships of selected economically important natural forests species. *Journal of Forest Products & Industries*, 2(1): 34-42.
- Ouma, Y.O. & R. Tateishi, 2006. Optimization of second-order grey-level texture in high-resolution imagery for statistical estimation of above-ground biomass, *Journal of Environmental Informatics*, 8(2): 70-85.
- Pang, Y., M. Lefsky, G. Sun, M. Ellen Miller & Z. Li, 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 Sciences, Vol. XXXVII, Part B7, Beijing, 777-782.

Pascual, C., A. Garcia-Abril, W.B. Cohen & S. Martin-Fernandez, 2010. Relationship between lidar-derived forest canopy height and Landsat images, *International Journal of Remote Sensing*, 31: 1261-1280.

Peng, Ch., 1999. Nonlinear height-diameter models for nine boreal forest tree species in Ontario. Ministry of natural resources, *Forest Research Report*, No. 155, 34pp.

Peterson, B. & K.J. Nelson, 2014. Mapping forest height in Alaska using GLAS, Landsat composites, and airborne lidar. *Remote Sensing*, 6: 12409-12426.

Petráš, R., M. Bošela, J. Mecko, J. Oszlányi & I. Popa, 2014. Height-diameter models for mixed-species forests consisting of spruce, fir, and beech. *Folia Forestalia Polonica*, 56(2): 93-104.

Pflugmacher, D., W. Cohen, R. Kennedy & M. Lefsky, 2008. Regional applicability of forest height and aboveground biomass models for the Geoscience Laser Altimeter System, *Forest Science*, 54(6): 647-657.

Ploton, P., R. Pélissier, N. Barbier, C. Proisy, B.R. Ramesh & P. Coueron, 2013. Canopy texture analysis for large-scale assessments of tropical forest stand structure and biomass, *Treetops at Risk*, 237-245.

Polychronaki, A., I.Z. Gitas, S. Veraverbeke & A. Debien, 2013. Evaluation of ALOS PALSAR imagery for burned area mapping in Greece using object-based classification, *Remote Sensing*, 5: 5680-5701.

Quiñones M., D. H. Hoekman, V. Schut & N. Wielaard, 2011. Above ground biomass map of Kalimantan, SarVision, Germany, Rep. 80087-SV-BMU.

Rao, R.B., G. Fung & R. Rosales, 2008. On the dangers of cross-validation, an experimental evaluation. In: Proceedings of the SIAM International Conference on Data Mining, SDM, Atlanta, Georgia, USA, 588-596.

Ratkowsky, D.A. & T.I. Reedy, 1986. Choosing near-linear parameters in the four-parameter logistic model for radioligand and related assays. *Biometrics*, 42(3): 575-582.

Roberts, J.W., S. Tesfamichael, M. Gebreslasie, J. van Aardt & FB. Ahmed, 2007. Forest structural assessment using remote sensing technologies: an overview of the current state of the art, *Southern Hemisphere Forestry Journal*, 69(3): 183-203.

Robertson, G.P., 1987. Geostatistics in ecology: Interpolating with known variance, *Ecology*, 68(3): 744-748.

Rosette, J.A.B., P.R.J. North & J.C. Suarez, 2008a. Vegetation height estimates for a mixed temperate forest using satellite laser altimetry, *International Journal of Remote Sensing*, 29(5): 1475-1493.

Rosette, J.A.B, P.R.J. North & J.C. Suarez, 2008b. Satellite lidar estimation of stemwood volume; a method using waveform decomposition, *The Photogrammetric Journal of Finland*, 21(1): 76-85.

Saatchi, S.S., H.L., S. Brown, M. Lefsky, E.T. Mitchard, W. Salas, B.R. Zutta, W. Buermann, S.L. Lewis, S. Hagen, S. Petrova, L. White, M. Silman & A. Morel, 2011. Benchmark map of forest carbon stocks in tropical regions across three continents, Proceedings of the National Academy of Sciences USA, 108(2): 9899-9904.

Sabol, J., Z. Patočka & T. Mikita, 2014. Usage of lidar data for leaf area index estimation, *GeoScience Engineering*, 10(3): 10-18.

- Sandberg, G., L.M. Ulander, J.E.S. Fransson, J. Holmgren & T. Le Toan, 2011. L-and P-band backscatter intensity for biomass retrieval in hemiboreal forest, *Remote Sensing of Environment*, 115: 2874-2886.
- Sarker, Md.L.R., J. Nichol, B. Ahmad, I. Busu & A.A. Rahman, 2012. Potential of texture measurements of two-date dual polarization PALSAR data for the improvement of forest biomass estimation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 69: 146-166.
- Schreuder, H.T., T.G. Gregoire & G.B. Wood, 1993, Mensurational aspects of forest inventory, in *Sampling Methods for Multiresource Forest Inventory*, New York, Wiley, pp. 227-315.
- Shimada, M. & T. Ohtaki, 2010. Generating large-scale high-quality SAR mosaic datasets: application to PALSAR data for global monitoring, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 3(4): 637-656.
- Silva Scaranello, M.A., L.F. Alves, S.A. Vieira, P.B. Camargo, C.A. Joly & L.A. Martinelli, 2012. Height-diameter relationships of tropical Atlantic moist forest trees in southeastern Brazil. *Scientia Agricola*, 69(1): 26-37.
- Simard, M., P. Nairara, J.B. Fisher & A. Baccini, 2011. Mapping forest canopy height globally with spaceborne lidar, *Journal of Geophysical Research*, 116: G04021.
- Sun, G.K.J. Ranson, D.S. Kimes, J.B. Blair & K. Kovacs, 2008. Forest vertical structure from GLAS: An evaluation using LVIS and SRTM data, *Remote Sensing of Environment*, 112: 107-117.
- Takagia, K., Y. Yoneb, H. Takahashia, R. Sakaia, H. Hojyoa, T. Kamiuraa, M. Nomuraa, N. Liangc, T. Fukazawad, H. Miyae, T. Yoshidaa, K. Sasaa, Y. Fujinumaf, T. Murayamag & H. Ogumac, 2015. Forest biomass and volume estimation using airborne lidar in a cool-temperate forest of northern Hokkaido, Japan, *Ecological Informatics*, 26(3): 54-60.
- Tonolli, S. & M. Dalponte, L. Vescovo, M. Rodeghiero, L. Bruzzone, D. Gianelle, 2011. Mapping and modeling forest tree volume using forest inventory and airborne laser scanning, *European Journal of Forest Research*, 130 (4): 569-577.
- Trinder, J.C., A. Shamsoddini & R. Turner, 2013. Relating worldview-2 data to pine plantation lidar metrics, *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Antalya, Turkey, 11 – 13 November 2013, Volume II-5/W2.
- Ulfarsson, M.O., 2007. Model Based Principal Component Analysis with Application to Functional Magnetic Resonance Imaging, A dissertation of doctor of philosophy, University of Michigan, 147pp.
- Vance-Chalcraft, H.D., M.R. Willig, S.B. Cox, A.E. Lugo & F.N. Scatena, 2010. Relationship between aboveground biomass and multiple measures of biodiversity in subtropical forest of Puerto rico; *Biotropica*, 42(3): 290-299.
- Vandekerckhove, J., D. Matzke & E.J. Wagenmakers, 2014. Model comparison and the principle of parsimony. In: Busemeyer JR, Townsend Z, Wang J, Eidels A, editors, *Oxford Handbook of Computational and Mathematical Psychology*, Oxford: Oxford University Press. pp: 1-29.
- Wang, X., X. Cheng, P. Gong, H. Huang, Zh. Li & X. Li, 2011. Earth science applications of ICESat/GLAS: a review. *International Journal of Remote Sensing*, 32(23): 8837-8864.
- Wang, Ch.H. & D.W. Hann, 1988. Height-diameter equations for sixteen tree species in the central western Willamette valley of Oregon, Oregon State University, Research Paper 51.
- Wei, Ch.L., G.T. Rowe, E. Escobar-Briones, A. Boetius, T. Soltwedel, M.J. Caley, Y. Soliman, F. Huettmann, F. Qu, Z. Yu, C.R. Pitcher, R.L. Haedrich, M.K. Wicksten, M.A. Rex, J.G. Baguley, J. Sharma, R. Danovaro, I.R. MacDonald, C.C. Nunnally, J.W. Deming, P. Montagna, M. Le´vesque, J.M. Weslawski, M. Wlodarska-Kowalczyk, B.S. Ingole, B.J. Bett, D.S.M. Billett, A. Yool, B.A.

- Bluhm, K. Iken & B.E. Narayanaswamy, 2010. Global patterns and predictions of seafloor biomass using random forests, *PLoS ONE*, 5(12): 1-15.
- Willmott, C.J. & K. Matsuura, 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30: 79-82.
- Xing, Y., A. de Gier, J. Zhang & L. Wang, 2010. An improved method for estimating forest canopy height using ICESat-GLAS full waveform data over sloping terrain: A case study in Changbai mountains, China. *International Journal of Applied Earth Observation and Geoinformation*, 12: 385-392.
- Yang, W., W. Ni-Meister & Sh. Lee, 2011. Assessment of the impacts of surface topography, off-nadir pointing and vegetation structure on vegetation lidar waveforms using an extended geometric optical and radiative transfer model. *Remote Sensing of Environment*, 115: 2810-2822.
- Yang, R.C., A. Kozak & J.H.G. Smith, 1978. The potential of Weibull-type functions as flexible growth curves. *Canadian Journal of Forest Research*, 8(4): 424-431.
- Zhang, G., B.E. Patuw & M.Y. Hu, 1998. Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, 14: 35-62.
- Zhang, L., 1997. Cross-validation of non-linear growth functions for modelling tree height-diameter relationships. *Annals of Botany*, 79: 251-257.
- Zhao, K. & S. Popescu, 2009. lidar-based mapping of leaf area index and its use for validating GLOBCARBON satellite LAI product in a temperate forest of the southern USA, *Remote Sensing of Environment*, 113(8): 1628-1645.
- Zhifeng, G., CH. Hong & S. Guoqing, 2010. Estimating forest aboveground biomass using HJ-1 satellite CCD and ICESat GLAS waveform data. *Science China Earth Sciences*, 53(1): 16-25.
- Zwally, H.J., B. Schutz, W. Abdalati, J. Abshire, C. Bentley, A. Brenner, J. Bufton, J. Dezio, D. Hancock, D. Harding, T. Herring, B. Minster, K. Quinn, S. Palm, J. Spinhirne & R. Thomas, 2002. ICESat's laser measurements of polar ice, atmosphere, ocean, and land. *Journal of Geodynamics*, 34: 405-445.



