A Machine Learning Approach to PolInSAR and LiDAR Data Fusion for Improved Tropical Forest Canopy Height Estimation Using NASA AfriSAR Campaign Data

Maryam Pourshamsi, Mariano Garcia, Marco Lavalle and Heiko Balzter

Abstract— **This paper investigates the benefits of integrating multi-baseline polarimetric interferometric SAR (PolInSAR) data with LiDAR measurements using a machine learning approach in order to obtain improved forest canopy height estimates. Multiple interferometric baselines are required to ensure consistent height retrieval performance across a broad range of tree heights. Previous studies have proposed multi-baseline merging strategies using metrics extracted from PolInSAR measurements. Here, we introduce the multi-baseline merging using a Support Vector Machine trained by sparse LiDAR samples. The novelty of this method lies in the new way of combining the two datasets. Its advantage is that it does not require a complete LiDAR coverage, but only sparse LiDAR samples distributed over the PolInSAR image. LiDAR samples are not used to obtain the best height among a set of height stacks, but rather to train the retrieval algorithm in selecting the best height using the variables derived through PolInSAR processing. This enables a more accurate height estimation for a wider scene covered by the SAR with only partial LiDAR coverage. We test our approach on NASA AfriSAR data acquired over tropical forests by the L-band UAVSAR and the LVIS LiDAR instruments. The estimated height from this approach has a higher accuracy (r^2=0.81, RMSE = 7.1 m) than previously introduced multi-baselines merging approach (r^2=0.67, RMSE = 9.2 m). This method is beneficial to future spaceborne missions such as GEDI and BIOMASS, which will provide a wealth of near-contemporaneous LiDAR samples and PolInSAR measurements for mapping forest structure at global scale.**

Index Terms—**Forest height, L-band, polarimetric synthetic aperture radar interferometry (PolInSAR), LiDAR (RH100), data fusion, Support Vector Machine (SVM).**

I. INTRODUCTION

Forest height is one of the most important forest biophysical parameters influencing light competition, stand productivity, carbon sequestration and biodiversity [1]. Forest height serves as a proxy to key information on forest ecosystems such as aboveground biomass and biomass change. It can be used to constrain allometric models in order to estimate forest aboveground biomass [2]. Height measurements within a stand provide information on the height distribution which can be used to assess the forest disturbance regime. Monitoring forest height changes over time is critical for understanding forest dynamics as well as detecting logging activities, deforestation and forest degradation.

LiDAR remote sensing, particularly with airborne platforms, has been widely used for estimating forest canopy height from local to regional scales although data availability is constrained by cost of acquisition. In addition, persistent cloud coverage, particularly over tropical areas, can limit LiDAR data availability. Spaceborne LiDAR missions like ICESAT-GLAS (2003-2007) and NASA's forthcoming Global Ecosystem Dynamics Investigation (GEDI) mission aim to overcome these limitations by reducing the spatial density of samples, without producing directly an image, but rather a grid of the world's land surface.

Polarimetric Interferometric Synthetic Aperture Radar (PolInSAR) has emerged as a viable technique for forest canopy height mapping over the past two decades. PolInSAR was first developed and demonstrated with SIR-C L-band data [3, 4]. PolInSAR as a model-based forest height estimation technique has been validated for different radar frequencies (X- C- L- and P- band) from either airborne or spaceborne platforms for boreal, temperate and tropical forests [5-13]. The quality of the PolInSAR-derived height depends on many factors, including a) instrument characteristics, b) temporal baseline, c) spatial baseline and d) forest stand characteristics [8, 14]. Ignoring the effects of instrument characteristics and assuming no temporal decorrelation, PolInSAR inversion performance depends significantly on the selection of the spatial interferometric baseline, or alternatively an appropriate interferometric vertical wavenumber [14]. A single-baseline PolInSAR approach yields an accurate height estimation only for a limited range of forest height. The heights at small baselines are overestimated due to small values of vertical wavenumber, and underestimated at large baselines due to its large values. This causes the interferometric phase to be wrapped which is due to the 2π modulo in the definition of the vertical wavenumber. Additionally, for taller trees, the volumetric decorrelation limits the performance, particularly for the estimation of the ground topography. This typically leads to underestimation of the stand height [14]. For short/sparse vegetation outside the optimal range, the non-volumetric decorrelation limits the performance leading to overestimation of vegetation height. Hence, in order to achieve an accurate result over a wide range of tree heights

Maryam Pourshamsi is with the Centre for Landscape and Climate Research, University of Leicester, Leicester, LE1 7RH, UK (e-mail: mp389@le.ac.uk). Her project sponsored by Engineering and Physical Science Research Council (EPSRC), grant reference: EP/M508081/1.

Dr Mariano Garcia is with the Department of Geology, Geography and Environment, University of Alcalá, Madrid, Spain. (e-mail: mariano.garcia@uah.es)

Dr Marco Lavalle is a Research Scientist with the Radar Science and Engineering group at the Jet Propulsion Laboratory, California Institute of Technology (e-mail: marco.lavalle@jpl.nasa.gov).

Professor Heiko Balzter is with the University of Leicester, Centre for Landscape and Climate Research, and National Centre for Earth Observation (e-mail: hb91@le.ac.uk).

and stand densities, multiple spatial baselines with varying lengths (multiple vertical wavenumbers) are typically employed. The multi-baseline merging approach helps to generate a new image which can cover different types of features on the ground (e.g. trees with different height ranges and density) [15]. There are different approaches introduced in previous studies for merging multiple baselines and obtaining optimum height estimates [10, 16, 17]. In these approaches, only PolInSAR related parameters are used as inputs for merging technique. Although, the estimated height from these multi-baseline merging approaches has higher accuracy than any single baseline approach, the estimated height still can be improved. One of the techniques that can help improving the height estimation is merging the multi-baseline PolInSAR height stack with the LiDAR samples using a machine learning approach. Merging SAR (mostly backscatter) and LiDAR data using machine learning approaches has been already proposed in the literature [18-21]. Hyde, et al. [19] investigated the synergic use of LiDAR, SAR and InSAR datasets using linear regression to estimate forest biomass in a Southwestern Ponderosa Pine forest. Santi, et al. [20] jointly considered Airborne Laser Scanning (ALS) and multi-frequency SAR images in an Artificial Neural Network approach for estimating forest biomass in Mediterranean areas [19]. These studies showed improvements in results when integrating both datasets compared to using either LiDAR or SAR data alone. It is evident that most applications of machine learning for combining SAR backscatter or InSAR parameters with LiDAR for biomass estimation required complete coverage of the study area by both sensors. A different approach was proposed by García, et al. [22] who used SAR backscatter and multispectral data to extrapolate LiDAR measurements of forest canopy height using a machine learning regression approach.

In this paper we explore the use of machine learning for integrating PolInSAR and LiDAR data for improved estimation of tropical forest canopy height under conditions of partial LiDAR coverage. Introducing this new way of combining the two datasets for estimating forest height is the novel contribution of this research.

Some studies explored the joint use of PolInSAR and LiDAR by estimating the underlying topography and the ground phase from the LiDAR-derived Digital Elevation Model (DEM) [23, 24]. Although, the use of LiDAR in these studies returned an accurate estimation of forest height, having LiDAR data in these approaches requires availability of a complete LiDAR coverage which is not always available, especially at regional scales. Hence, it is desirable to develop a synergistic technique that combines the sparse sampling capabilities of LiDAR with the imaging capabilities of PolInSAR that works for incomplete LiDAR coverage.

This paper is motivated by the large amount of PolInSAR and LiDAR data that will be available in the near future from forthcoming missions such as ESA's P-band SAR BIOMASS in 2021 [25], the joint NASA-ISRO L-band SAR (NISAR) in 2021 [26] and DLR's Tandem-L with PolInSAR L-band capabilities in 2023 [27]. GEDI LiDAR is scheduled for launch in late 2018 [28] and is expected to operate for two years. It will sample the Earth's surface with 25 m nominal footprints distributed at 60 m spacing along-track and ~500 m spacing cross-track) [29]. Therefore, in order to better quantify Earth's forest structural parameters, a combination of GEDI footprints with upcoming low-frequency radar missions will be paramount [30]. The new way of combing PolInSAR and LiDAR samples introduced here will be likely beneficial to the science community exploiting future GEDI LiDAR and SAR missions data. Nevertheless, our method applies a stratified random sampling approach which differs from the systematic sampling of GEDI, however, this difference in the sampling scheme should not affect the suitability of our method.

This study aimed to develop a method that applies the synergetic use of LiDAR and multi-baseline PolInSAR that enables a more accurate retrieval of forest canopy height, particularly when the coverage of LiDAR data is incomplete. A Support Vector Machines (SVM) is used to extrapolate LiDAR-based canopy height using PolInSAR inverted parameters. SVM does not make any assumption about the distribution of the data and can capture complex relationships between dependent and the independent variables. A key consideration of this study is that the method needs to work even under conditions with only partial LiDAR coverage. Our method works in situations where a large PolInSAR image coverage and only sparse LiDAR samples are available. The method utilizes the LiDAR samples to inform the retrieval algorithm and improves the height retrieval over the entire PolInSAR scene where the LiDAR data is not available.
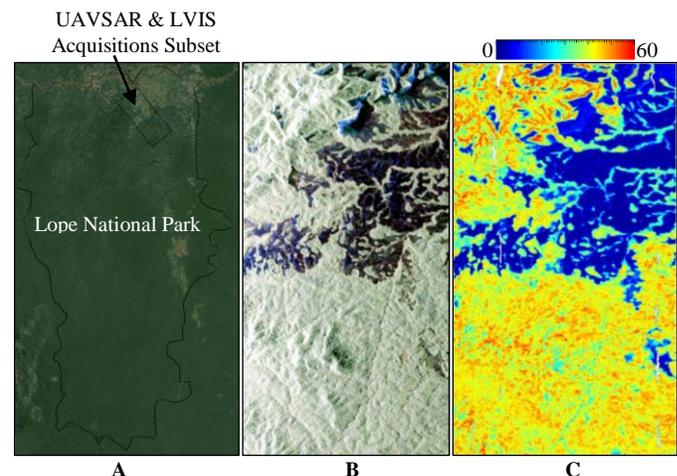
Fig. 1. A: Lope National Park showing the approximate UAVSAR and LVIS acquisition boundary (Google Earth image used for the background image). B: A subset of UAVSAR Pauli decomposition, near range ($\theta=25°$) on the left and far range ($\theta=40°$) on the right side. C: LiDAR RH100 image resampled to the UAVSAR image, indicating that vegetation ranges between 0-15 m over the open areas (blue/magenta) and 15-60 m over dense forest areas (yellow/red).

In Section II, we provide an overview of the AfriSAR campaign, the study site and data acquisition. In Section III, we describe the methods which followed in this study: A) the conventional single-baseline PolInSAR inversion approach, B) PolInSAR multi-baseline merging approach and C) SVM PolInSAR and LiDAR data fusion approach. The results will be given in Section IV and discussion and analysis in Section V. Finally, the conclusions are drawn in Section VI.

II. AFRISAR CAMPAIGN

The AfriSAR campaign was a joint effort between NASA and ESA to collect airborne and field data in multiple sites over Gabonese forests. The first phase of the campaign was carried out in July 2015 with the ONERA's SETHI airborne SAR instrument. The second phase was performed in February-March 2016 with NASA's UAVSAR and LVIS airborne LiDAR, DLR's F-SAR and a ground data campaign led by the University College London (UCL) and the Centre d'Etudes Spatiales de la BIOsphere (CESBIO). Gabonese forests are different from Amazonian or Asian forests as they include diverse forest types, from very tall, dense, rainforests to short, sparse and dry savannas. Heterogeneity in forest types within the same site is a challenge for the estimation of forest height and biomass due to differences in forest structure. Four main study sites within Gabon were identified during the campaign: Mondah, Lope, Mabounie and Rabi. The Lope test site was selected as a supersite due to being so diverse and covering different vegetation types within one area and this makes the observation of such a site challenging. In the scope of this research, we only focus on Lope.
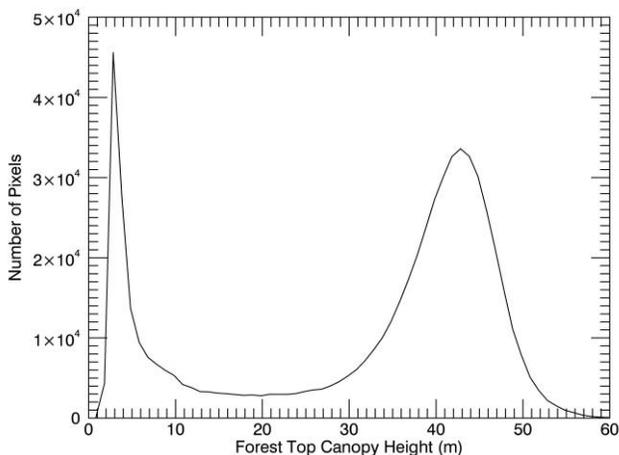
Fig. 2. An insight into forest height stand based on LiDAR RH100. As seen, the area covers two dominant vegetation types, namely short/sparse savannas (0-15 m) and tall/dense forest (15-60 m).

A. Study Site

Lope National Park is located at 0°30′00″S 11°30′00″E in central Gabon and covers an area of 4,910 km². It has been a wildlife reserve since 1946, and a National Park since 2002. The area contains diverse habitats, surrounded by closed-canopy tropical rainforest typical of the Congo Basin, the middle of the site is characterized by savannah and a mosaic of low-biomass forest types [21]. Here, the focus is on a small area of interest in northeastern Lope (Fig. 1).

Fig. 1 and Fig. 2 provide an insight into forest stand heights and forest top canopy height ranges. They clearly indicate that the test site is divided into two different vegetation types. The image B of Fig. 1 is an RGB color composite using the Pauli basis components (HH+VV, 2HV, HH-VV) of the UAVSAR master image. Open areas appear in dark blue and forested areas

in green. The dark blue part covers short/sparse savannas with trees height ranging from 0-15 m. The green part covers tropical forest with different ages: young (<10 years), intermediate (10-24 years), older (25-49 years) and maturing (>50 years) and old-growth with forest height ranges between 15-60 m. The forest is dominated by various tropical species including *Annonaceae*, *Xylopia*, *Aethiopica* , *Staudtii, Burseraceae, Aucoumea, Klaineana, Ochnaceae, Lophira and Alata* [31]. The topography is gently undulating with average slope up to 25%.

B. Radar Acquisitions

During the AfriSAR campaign, L- and P- band SAR data were acquired. Here we focus on the L-band JPL Uninhabited Aerial Vehicle SAR (UAVSAR) [32] acquisitions due to the characteristics of the imaging radar in mapping large areas for ecosystem applications. The UAVSAR instrument operates at 1217.5-1297.5 MHz (L-band) and employs an electronically scanned array antenna to enable robust repeat-pass interferometric measurements [6]. The UAVSAR data were acquired in February 2016 with a nominal flight altitude of 12.5 km, which allows mapping an area of about 22 km wide, with incidence angles ranging from 25° to 65°. The UAVSAR polarimetric single-look-complex (SLC) product has ground range and azimuth resolutions of 2.5 m and 1 m, respectively [7]. The Lope UAVSAR dataset is fully polarimetric and has been acquired by incrementing the aircraft altitude by 20 m at each flight track to achieve a broad range of vertical interferometric wavenumbers and resolve multiple layers of vegetation structure (Table I). UAVSAR data are focused and registered in a stack using precise platform-SRTM DEM [33, 34] geometry.

Table I
SUBSET OF UAVSAR FLIGHTS OVER LOPE USED IN THIS STUDY

Flight	Baseline [m]	Mean k_z [m^{-1}]	Mean HoA [m]	ΔT [min]	Time
1	master	master	master	master	11:11
2	20	0.06	104	22	11:33
3	40	0.12	52	45	11:56
4	60	0.18	35	67	12:18
5	80	0.24	26	90	12:41
6	100	0.30	21	129	13:20
7	120	0.36	17.5	149	13:43

UAVSAR was registered to the DEM using an accurate geometry-based co-location based on airborne look vectors and Doppler information commonly used in interferometry. The DEM then was up sampled to match the full resolution of UAVSAR images.

C. LiDAR Acquisitions

Land Vegetation and Ice Sensor (LVIS) LiDAR data [35] used in this study were acquired on the NASA Langley KingAir

B-200 at an altitude of 7.3 km. Nominal LVIS footprint diameter is 25 m with 9 m separation (overlapping) along track. Among all the metrics that can be extracted from the LVIS waveforms, we used the LVIS RH100 metric, which was generated and distributed by the GSFC LVIS team. The RH100 metrics was provided in slant range geometry with a dimension of 600×1000 (pixels).

LiDAR data was co-located with radar using location information stored in the LVIS metadata. Visual inspection and correlation analysis at full resolution shows highly precise co-location between the different datasets. The original UAVSAR SLC images had a dimension of 3000×29000 (pixels). A multi-looking process was carried out using a factor of 5×29 to transfer the PolInSAR stacks into LiDAR geometry. Both the LVIS and UAVSAR acquisitions used in this study cover the same area (Fig. 1). However, for the data fusion purpose, less than 1% of the available LiDAR samples are used for training, while the remaining pixels of LiDAR are used as an independent dataset for validation purposes.

D. Ground Data Collection

Field data collection was carried out as part of the AfriSAR campaign by the University College of London (UCL) and the Centre d'Etudes Spatiales de la BIOsphere (CESBIO), between January to August 2016 [31]. In total, 12 main plots (50×50 m) were collected, divided into four sub-plots (25×25 m). Some of the main plots are incomplete and miss the sub-plots. Therefore, a total of only 37 subplots are available. The plots include dense tropical forest to sparse savannas. The top canopy height measured for each subplot, is used in this research to validate the SVM estimated top canopy height.

III. METHODOLOGY

A. Single baseline PolInSAR Parameter Inversion

This section describes the PolInSAR inversion scheme based on the Random Volume Over Ground (RVoG) model [4, 36] that was applied in this study to estimate forest canopy height from the UAVSAR data. We started by estimating the complex interferometric coherence γ for all pairs at all polarimetric channels (HH, HV and VV), [6]

$$\gamma = \frac{E(S_1 S_2^*)}{\sqrt{E(S_1 S_1^*)} \cdot \sqrt{E(S_2 S_2^*)}} \quad (1)$$

where S_1 and S_2 are two complex SAR images. The magnitude of γ ranges between 0 and 1 [37, 38] and depends on sensor characteristics, acquisition parameters and vegetation structure. Three main decorrelation sources affect the coherence: temporal decorrelation, thermal noise decorrelation and volume decorrelation [8, 14]. The UAVSAR instrument provides a good signal-to-noise ratio for vegetation applications in all polarimetric channels, which results in a very low thermal noise and phase decorrelation (> 0.95) [39]. Given the structure of the vegetation in the Lope site and the effects of temporal decorrelation at L-band, we ignored the effects of thermal noise decorrelation, which are significantly smaller compared to the combined volume and temporal decorrelation for UAVSAR [43]. Coherence was computed within a rectangular uniform window of 5x29 samples (145 nominal looks). The amplitude of interferometric coherence has values typically greater than 0.85 over open areas in all three polarimetric channels for all pairs, with greatest value in the VV channel (≈ 0.9). Within the forested areas, this value is lower as volumetric interferometric decorrelation tends to be more prominent [6, 23]. On average, coherence in the HV channel is observed to be greater compared to the HH and VV channels within these areas.

The performance of the RVOG model has been widely evaluated using different SAR wavelengths and over different forest conditions [4, 36]. In the RVoG model, forest canopy is idealized as a random volume layer of height h_v over the ground located at $z=z_0$. The ground can be seen through the canopy volume by the interferometer operating at wavelength λ, with a given incident angle θ_0 at range R [6]. The RVOG model predicts the volume coherence (γ_{est}) using:

$$\gamma_{est} = \frac{\int_{z0}^{z0+hv} \exp(\frac{2\sigma hv}{\cos\theta0}) \exp(i \, k_z \, z) dz}{\int_{z0}^{z0+hv} \exp(\frac{2\sigma hv}{\cos\theta0}) dz} \quad (2)$$

Eq.(2) represents a look-up-table (LUT) with priori-known values of all possible extinction coefficient (σ) and heights (hv). The extinction coefficient depends on the density of the canopy elements in the volume and their dielectric constant. We set values between 0-1.5 dB/m for mean extinction coefficient (σ) and 0-60 m for forest heights (hv), which are decided based on the forest types and ground field inspection for Lope. The parameter k_z is the effective vertical wavenumber, which depends on the radar wavelength λ and imaging geometry and it scales the interferometric phase to height:

$$k_z = m \frac{2\pi}{\lambda} \frac{\Delta\theta}{sin(\theta0)} \quad \approx \quad k_z = m \frac{2\pi}{\lambda} \frac{B_\perp}{R \, sin(\theta0)} \quad (3)$$

where $\Delta\theta$ is the angular separation of the two acquisitions in the direction of the resolution cell, $B_\perp$ is the perpendicular baseline, θ_0 is the local incidence angle, R is the slant range distance. The factor m accounts for the acquisitions mode. For monostatic acquisitions $m = 2$, and for bistatic $m = 1$ [9, 14].

Once the volume coherence is predicted through the RVoG model, we need to estimate the volume coherence from the actual PolInSAR stacks. This step includes three main stages: a) mapping the real and imaginary components of the coherence points computed from (1) onto a line in a complex plane, b) estimating the ground phase φ_0 which is the intersection point of the line with the unit circle (Fig. 3) and removing it from the assumed volume coherence and finally c) forest height (h_v) and extinction coefficient (σ) estimation [6]. For the ground phase estimation, we examined both line-circle intersection points:

$$\begin{cases} \text{If } d_1 > d_2, then \, \varphi_0 \, = \, \varphi_1 \\ \text{If } d_1 < d_2, then \, \varphi_0 \, = \, \varphi_2 \end{cases} \quad (4)$$

and considered both solutions as possible valid ground phases in our multi-baseline merging approach, which will be described in the next sections.

At L-band, we assumed HV channel as representative of the volume, and therefore, we subtracted the estimated ground phase from the HV channel to estimate the volume-only

coherence (γ_{Vol}) [5, 6]. This is justified by the fact that the cross-polarimetric channel is sensitive to random orientation of the canopy elements, and based on Lope characteristics, the ground contribution is expected to be low in the HV channel [6, 9, 40].

In the final step, the extinction coefficient and heights can be estimated by finding values at where the differences between predicted RVoG volume-only coherence (γ_{est}) and the estimated volume-only coherence (γ_{Vol}) are minimal.
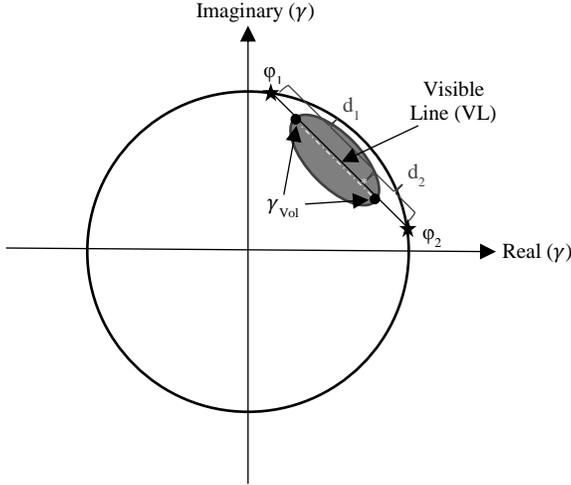
Fig. 3. Unit circle and PolInSAR Inversion scenario: Grey ellipse: coherence region [6, 8, 23], black line: line fitted through the coherence region, white star: an example of a coherence point, black dots: candidates for γ_{Vol}, black stars φ_1 and φ_2: candidates for ground phase, d_1 and d_2: distance candidates between a coherence point and ground phase, white dashed line: visible line of the coherence region (*VL*).

The procedure described above is valid for tree heights not exceeding the interferometric Height of Ambiguity (HoA):

$$\text{HoA} = \frac{2\pi}{k_z} \qquad (5)$$

In the presence of tall trees and large baselines, it is likely that the interferometric phase is wrapped along the vertical direction, which requires an unwrapping process prior to application of the height estimation algorithm. In this paper, however, phase unwrapping is deliberately not applied as baselines potentially affected by wrapped phases are disregarded by the multi- baseline selection and merging algorithm.

B. Multiple Baseline PolInSAR Merging

The procedure described above was applied to each of the single-baseline PolInSAR pairs of UAVSAR data for canopy height estimation. In order to have an accurate estimation of height over a wide range of vegetation types (short/sparse to tall/dense), multiple baselines (different k_z values) are required. One possible approach for merging multiple PolInSAR baselines is based on the cost function introduced in [10, 17]. In this approach, the best height estimation from each individual pair is selected according to the interferometric height precision, which is defined as the standard deviation of the interferometric phase divided by the vertical wavenumber [10]:

$$H_{precision=} -\frac{1}{kz}\varphi_{Int} \quad \text{with} \quad \varphi_{Int} = \sqrt{\frac{1-\gamma^2}{2\,L_{Int}\gamma^2}} \quad (6)$$

where L_{Int} is the number of looks used to estimate the interferometric coherence γ. When the interferometric coherence amplitude is reduced by non-volumetric factors such as temporal decorrelation, the standard deviation of φ_{Int} increases and the precision of estimated height decreases. This implies that the smallest height precision stands for a more reliable parameter inversion results. Between a set of multiple baseline forest height estimations, the approach proposed in [10] will select the height where the height precision is minimal:

$$\begin{cases} H_{precision}^1\left(k_z{}^1, \gamma^1, L_{Int}\right) \\ H_{precision}^2\left(k_z{}^2, \gamma^2, L_{Int}\right) \\ \qquad . \\ \qquad . \qquad\qquad\qquad (7) \\ \qquad . \\ H_{precision}^N\left(k_z{}^N, \gamma^N, L_{Int}\right) \end{cases}$$

where *N* denotes for the number of available baselines.

C. SVM PolInSAR and LiDAR Data fusion

In this section, we introduce an approach based on the synergetic use of LiDAR and multiple PolInSAR baselines using a Least Squares Support Vector Machine (LS-SVM) algorithm [41]. The basis of SVM was established in [42] and has proven to be useful for integrating multisource remote sensing data both for classification [43, 44] and regression applications [22, 45, 46]. In classification problem, SVM attempts to fit an optimal hyperplane to the training samples in a multidimensional feature space based on structural risk minimization. This approach fits the hyperplane by maximizing the margin between the closest training samples, which are known as support vectors. When the classes cannot be linearly separated, the data are mapped into a higher dimensional space, where the classes can be linearly separated, by means of a kernel function [47]. In this study, a radial basis function (RBF) was selected, which is controlled by two parameters, the penalty parameter (C) and the kernel bandwidth (*h*). The parameter *h* determines the size of the kernel whereas C controls the penalty associated to misclassifications in the training data and represents a trade-off between error and model complexity. Large values of C reduce the error but also the generalization ability of the model. A grid search approach with ten-fold cross-validation was used to obtain the optimum values for C and *h*. The first step of this approach consists of searching for the optimum height estimate H_{opt} in the stack of 12 PolInSAR height images previously obtained, by minimizing the absolute difference between PolInSAR height estimates and RH100. The histogram and scatterplot of H_{opt} illustrated in Fig. 4 indicate that in order to obtain an appropriate accuracy for the various height range between 0-60 meters, multiple baselines are required [14]. Fig. 4 also proves that there is at least one height value in the one the 12 PolInSAR heights that is very close to RH100. H_{opt} represents the desired height estimate after merging the available baselines where LVIS RH100 is used as reference height. The SVM classification approach requires a categorical image where each pixel is labeled with a value from 1 to 12 corresponding to the PolInSAR height image with height value closest to RH100. The independent variables

(features) used for the SVM embraced the different parameters used in the estimation of forest height by PolInSAR. These features include the volume coherence (magnitude and phase) and the vertical wavenumber (k_z).
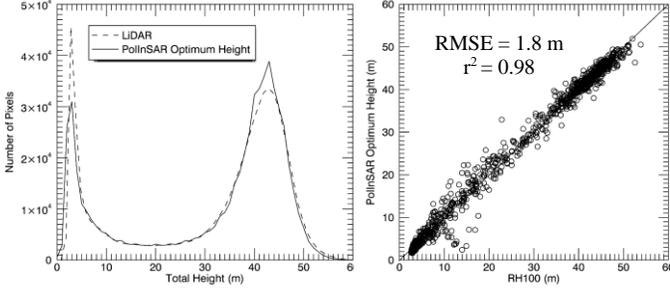
Fig. 4. Validation of the Optimum Height versus LiDAR RH100. The Optimum Height is a merged height based on all the available PolInSAR estimated heights by picking the pixels (one out of 12) which has the minimum absolute difference with the LiDAR RH100.

An analysis of variable importance was carried out to identify the dominant variables in our model. An initial model was trained and validated using only coherence magnitude and phase. Subsequent models were trained by adding sequentially k_z. The improvement in r^2, RMSE and bias obtained after including this variable was evaluated.

Training of the SVM algorithm was based on 5000 samples (less than 1% of the available samples, i.e. image pixels) collected across the images using a stratified random sampling approach. The stratification was based on the RH100 histogram using 5m height bins, thus ensuring that all forest canopy height ranges present in the study area were appropriately sampled based on their frequency. Different training sample sizes were tested but increasing the sample size beyond 5000 samples did not significantly improve the results and increased the time required for training the model. The rest of the pixels of the image (>550000) were used for independent validation. In this way, the generalization capability of the model was tested.
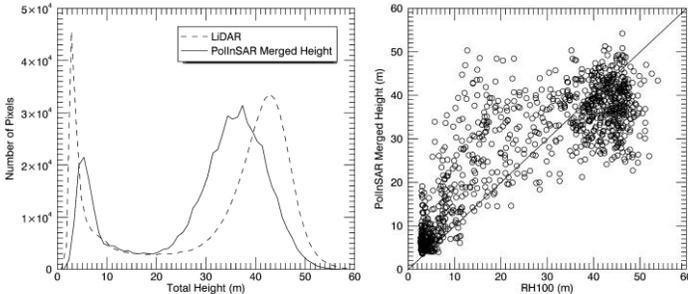
Fig. 5. Validation of the PolInSAR merged height versus LiDAR RH100

IV. RESULTS OF CANOPY HEIGHT ESTIMATION

In this section, we first present the results of single baseline PolInSAR height estimation algorithm described in Section III-A to individual UAVSAR interferometric pairs. This will be followed by the results obtained using a multi-baselines merging approach described in Section III-B and finally the results from PolInSAR and LiDAR data fusion approach which was introduced in Section III-C.

1) Single-baseline PolInSAR Heights: Six nominal interferometric baselines were available: 20, 40, 60, 80, 100 and 120 meters. By considering both possible ground phases for each baseline, we estimated 12 PolInSAR heights in total for each pair using procedures explained in Section III A. It is understood that certain baselines are far from the optimum baseline for a given tree height, potentially providing very low coherence or even wrapped phases that are not appropriate for height estimation with the method described in Section III.A. The estimated height at each baseline was validated versus RH100 individually and Table II reports the errors for all baselines to illustrate how the average errors are dependent on the baseline length, as expected.

Table II
R^2, RMSE AND BIAS FOR THE ESTIMATED HEIGHTS PER INDIVIDUAL BASELINE

Baseline	r^2	RMSE [m]	Bias[m]
B = 20 m, using φ_1	0.28	17.5	-15.7
B = 20 m, using φ_2	0.28	23.1	13.8
B = 40 m, using φ_1	0.45	12.6	-8.8
B = 40 m, using φ_2	0.12	25.8	-22.2
B = 60 m, using φ_1	0.66	11.3	-11.2
B = 60 m, using φ_2	0.53	13.9	-12.8
B = 80 m, using φ_1	0.69	9.5	-8.1
B = 80 m, using φ_2	0.59	12	-10.7
B = 100 m, using φ_1	0.68	9.1	-6.6
B = 100 m, using φ_2	0.63	10.4	-8
B = 120 m, using φ_1	0.71	8.6	-5.5
B = 120 m, using φ_2	0.70	8.7	-7.1

2) PolInSAR Merged Height: We examined the previously introduced multi-baseline PolInSAR combination criterion in [10, 17] which uses a cost-function approach for merging the multiple baselines. In this method, no LiDAR is used, and only PolInSAR-derived parameters contribute to the cost function. These include the volume coherence (magnitude and phase) and the vertical wavenumber [10, 17]. This approach ensures that the phase errors are reduced while sensitivity to height is still high (Fig. 5). The calculated RMSE is 9.2 m and the r^2=0.67.
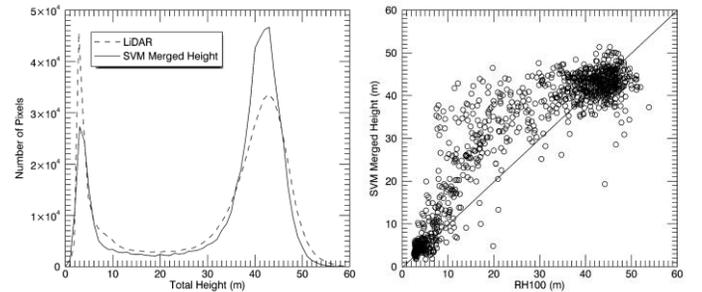
Fig. 6. Validation of the SVM estimated height versus LiDAR RH100

3) SVM Merged Height: In this approach we made a synergetic use of LiDAR and multiple baselines PolInSAR information introduced above to investigate forest height retrieval. The merged height from the SVM approach, performed better than the previous two approaches (Fig. 6), with an increase in r^2 and

decrease in RMSE varying between 1.3 and 18.5 m, and a reduction of the bias by up to 20.8 m.

The overestimation of the height over short/sparse vegetation reduced significantly (~5 m) and underestimation over the tall/dense stand is also reduced (~10 m) in comparison to PolInSAR-only merged height. The estimated height was evaluated against the LiDAR RH100 and model performance assessed in terms of the r^2, RMSE (m) and bias (m) – Table III. We also validated the SVM estimated height versus the ground measured height. To perform this validation, the SVM merged height was georeferenced and resampled to 25×25 m resolution cells, the same size of the field plots (Fig. 7). The SVM estimated height and in situ heights have $r^2 = 0.68$ and RMSE = 7.6 m.

Table III
R^2, RMSE (M) AND BIAS (M) OF THE SVM MERGED HEIGHT

	r^2	RMSE [m]	Bias [m]
Training	0.81	7.17	-1.32
Validation	0.80	7.35	-1.40

V. DISCUSSION

1) Single Baseline PolInSAR Heights: We examined PolInSAR-derived canopy height for different spatial baselines. As anticipated, using a single baseline approach, either short or long, may not always return an accurate height results over a heterogeneous site covering different vegetation types (tall/dense and short/sparse) due to Height of Ambiguity (HoA). The values of k_z play a significant role in PolInSAR performance. For too large k_z values, the sensitivity of the coherence to forest height saturates at a given height, and large heights remain underestimated. For even larger k_z values, phase unwrapping is more likely to happen along the vertical profile. On the other hand, for too small k_z values, decorrelation introduces large height errors, so heights become overestimated [14]. According to (5), large k_z values account for small HoA which indicates better estimation of the height for shorter trees and underestimation of taller heights, while small k_z values account for large HoA. This leads to overestimation of the height. Looking at the results of validation of single baseline PolInSAR heights versus RH100 (Table II) r^2, RMSE and bias are improving from small baselines (20, 40, 60 m) to large ones (80, 100, 120 m). At small baselines, the height becomes highly overestimated due to large HoA, while at larger baselines, the height is underestimated for taller trees and is reasonably well-estimated for shorter trees. The higher accuracy associated with shorter vegetation increased the r^2 values at larger baselines. Therefore, a single baseline may not return an accurate height estimate over a heterogeneous site and one way to overcome this bias is to merge multiple baselines with various k_z values ranging from small to large. Each baseline provides solution space at a certain height value, and with the multiple baseline merging approach, the solution space of one baseline can support the solution space of the other ones [10, 14].

We also carried out further experiments by testing different two candidates of ground phases (φ_1 and φ_2) of the unit circle (Fig.3). Choosing φ_1 and φ_2 as ground phase might return a valid or an invalid outcome, but this only holds true for the high coherences. For these points, the shape of the coherence region is likely to be an ellipse, which normally locates at a side of the unit circle. In this case, the distances candidates between the coherence points and the intersection points are not equal (Fig. 3: $d_1 \neq d_2$), and accordingly, φ_1 and φ_2 as the ground phase are valid or invalid. Our results at single baseline PolInSAR estimated height, totally confirms these arguments. Looking at RMSE and correlation coefficient (r^2) results in Table II, for the smaller baseline (20 and 40 m) which provides higher coherences, the accuracy of estimated height differs from φ_1 to φ_2, meaning that the ground phase with the largest distance from the γ_{HV} is valid for taller vegetation and the ground phase with the shortest distance from the γ_{HV} is valid for shorter vegetation. For larger baselines (60, 80, 100 and 120 m) which give lower coherences, the coherence points locate around the middle of the unit circle, which shape a circular coherence region. In this case the distance between the coherence points and the intersection points are likely to be equal (Fig. 3: $d_1 \approx d_2$), therefore, whatever points are selected as ground phase does not affect significantly on the estimated height. As per results in Table II for larger baselines, there is no significant difference between r^2 and RMSE calculated for the estimated height in using φ_1 or φ_2 as the ground phase, and both return similar results.
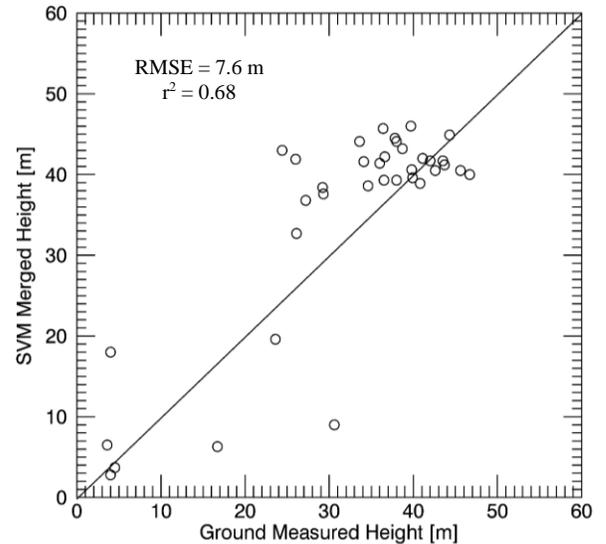
Fig. 7. Validation of the SVM estimated height versus in-situ measured height.

In case of short/sparse vegetation, the density is low, hence, the assumption of largest distance from the γ_{HV} for choosing the intersection point of the unit circle as the ground, might not always pick an accurate ground phase. Radar L-band can penetrate into the open canopy, and the scattering might not necessary be from the top of the canopy. In this case, the ground phase can be the intersection point which has shortest distance from the γ_{HV}. However, this might lead to an underestimation of the height over these areas. If the study site covers different range of vegetation types like the Lope National Park where tall/dense forest and short/sparse savannas exist in the same scene, then both ground phase candidates should be retained.

2) PolInSAR Merged Height: The comparison between PolInSAR multi-baseline merging and LVIS RH100 heights (Fig. 5) confirms that the overall accuracy of the merged product improves significantly compared to the single-baseline right), a small overestimation over open areas and underestimation over forested areas can be observed. The Lope test site is highly affected by topography and terrain slope, and one argument confirming the overestimation/underestimation
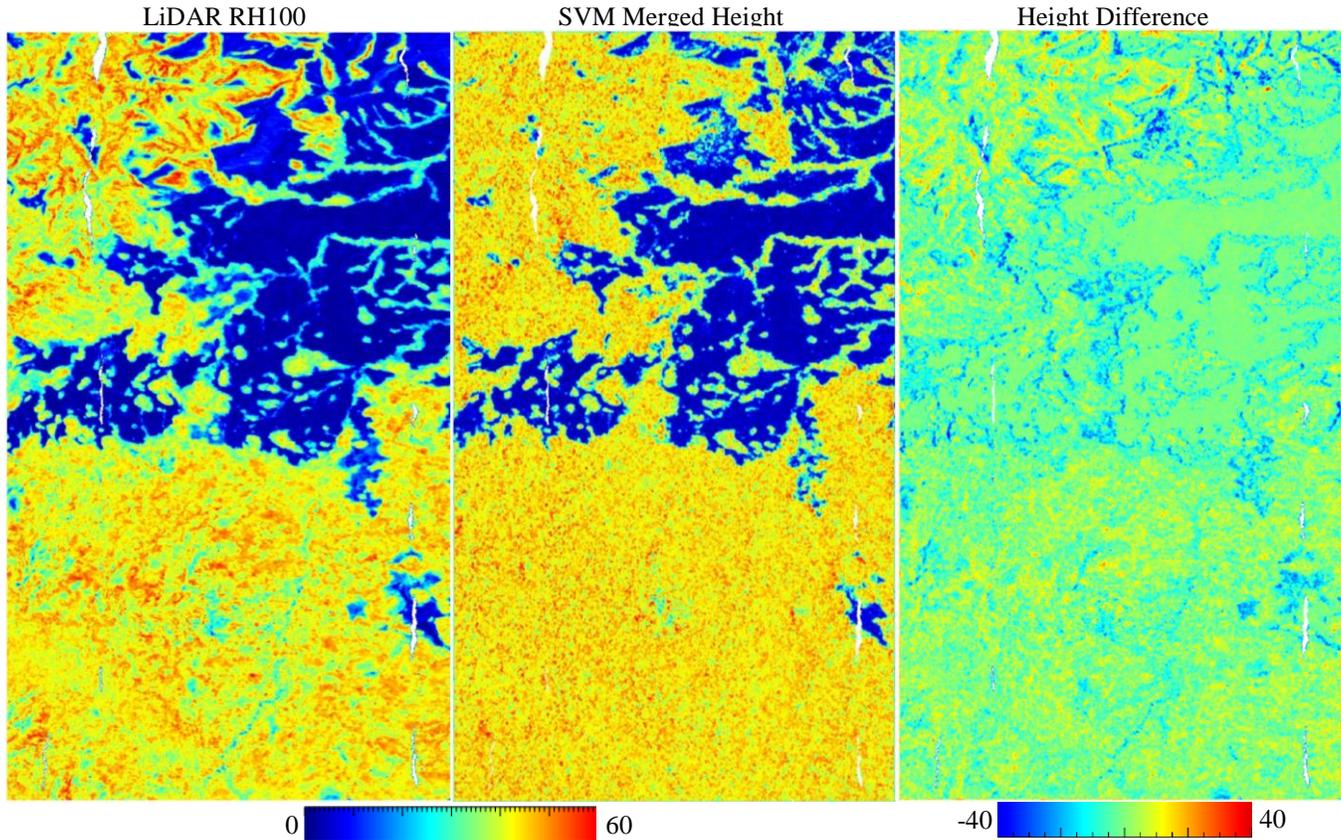
Fig. 8. Left: the LiDAR RH100 image, middle: the SVM merged height and right: the map of height difference between LiDAR RH100 and SVM merged height

height estimate. This is the consequence of the sensitivity of the vertical wavenumber to different tree structural types. As suggested in [10], we used the interferometric height precision as a metric for merging the baselines, which incorporates both the volume coherence and the vertical wavenumber. From Fig. 5, this balance between volume coherence level and baseline length appears to work better for relatively short vegetation (<7 m) and tall trees (>50 m). Within this range, very small / very large baselines are correctly selected by the PolInSAR merging algorithm. However, in the range of 15 to 30 m the PolInSAR-derived height is overestimated because of the sub-optimal baseline selection, i.e. the values of vertical wavenumber are too small or too large for the given tree height and structure.

3) SVM Merged Height: Fusion of LiDAR and PolInSAR parameters improved height estimation compared to merging baselines based on PolInSAR parameters only. The SVM estimated height captured the general trend in vegetation height for the study area, clearly representing the savannah and the closed canopy tropical rainforest vegetation. Nevertheless, it failed to capture some patterns of low vegetation within the closed canopy forest that can be observed in the LiDAR data (Fig. 8) and showed a more random distribution of the forest canopy height. Looking at the map of height difference (Fig. 8

can be related to topographic effects in the UAVSAR images. For terrain and slope correction of the UAVSAR images the SRTM DEM was used. However, the 30 m resolution of SRTM is sometimes insufficient for topographic correction of UAVSAR imagery with 2.5*1 (Rg*Az) m resolution in tropical areas. This then caused an underestimation of height over negative slopes and overestimation over positive slopes. The other reason is that the selected k_z values are too small or too large for the given tree height resulting an overestimation of the mid-range height. However, these values are significantly improved over the PolInSAR-only approach.

Several studies have explored the synergy of LiDAR and InSAR data for forest height estimation, where the LiDAR DEM is used to correct the InSAR ground phase [19, 24]. Nevertheless, these approaches require LiDAR and InSAR measurements continuously acquired over the same area. Qi and Dubayah [24] assessed the impact of extrapolating a simulated GEDI LiDAR DEM with poorer results than those yielded by an airborne imaging LiDAR DEM. Our approach allows an extrapolation of the forest canopy height over areas beyond the LiDAR coverage and can be applied to future satellite LiDAR and radar missions. GEDI data will provide global (between ± 51° latitude) sampling of terrestrial ecosystems. However, in order to provide wall-to-wall

coverage, these data will need to be integrated with other sensors dataset. PolInSAR data provided by future satellite missions such as BIOMASS or NISAR could be used to derive such information using the method presented in this paper. However, further analysis over different forest types is required to prove the robustness of our method for global applications. Differences in the sampling strategy used in this study and the GEDI acquisition plan also require further assessment. GEDI plans to provide systematic sampling with footprints separated by 500 m across-track and 60 m along-track, while in this study a stratified random sampling approach was used, with a minimum, maximum and mean distance between samples of 1, 1146 and 424 pixels in slant range, representing approximately 20 m, 22.9 km and 8.5 km on the ground. Although differences in the sampling scheme could affect the performance of our approach, the GEDI sampling strategy should guarantee the minimum sample size required to achieve accurate results.

Sample size is an important factor in the performance of machine learning algorithms [48], yet our approach proved to be robust with a sample size as small as 1% of the total area, i.e. 5000 pixels. The small sample size highlights the capability of the method for applying it to the in-situ measured height in absence of LiDAR. Where a set of in-situ measured height covering different forest height ranges distributed over a test site is available, it is possible to carry out this approach and train the SVM with the ground-measured height, instead of LiDAR height. However, the LiDAR height is far more accurate than the in-situ measured height, particularly in the tropics where the forest is too dense to collect accurate in-situ height measurements.

The SVM model was based on complex coherence (magnitude and phase) and k_z. An analysis of variables importance showed that using only coherence information the model explained 77% of the variance, the RMSE was 8.46 m, and the bias was -3.29 m. By including k_z in the model, r^2 increased by 3%, whereas RMSE and bias decreased 1.3 m and 2.2 m respectively, both for the calibration and validation datasets. The contribution of k_z to an improvement in the merged height results was negligible for the aggregated statistics, although their inclusion provides additional information over other vegetation types and more heterogeneous canopy structures.

The SVM merged height with r^2 =0.80 and RMSE = 7.35 m has higher accuracy in comparison to the PolInSAR-only merged height with r^2 = 0.67 and RMSE = 9.2 m. In the SVM approach, the overestimation of the height over the short/sparse vegetation is significantly improved whereas this issue remains in place in the PolInSAR-only merged approach. The underestimation of the height over tall/dense forest is reasonably well corrected compared to the previously introduced merging approach, but can still be observed for the tallest forest stands (about 5 m) due to ground level limitation at L-band[40].

In addition to the validation of the SVM-estimated height versus LiDAR, we also carried out a validation analysis by including the ground measured height. Although the number of plots is small (37 plots sized 25×25 m), the results also indicate high accuracy with r^2 = 0.68 and RMSE = 7.6 m. We believe with more measurements this result can be improved further. For this study, we used UAVSAR airborne datasets with six PolInSAR baselines (six different vertical wavenumbers). This helped achieve an accurate height retrieval over a wide range of tree heights and stand densities. Further studies are required to test the robustness of this method using different datasets and for a smaller number of baselines.
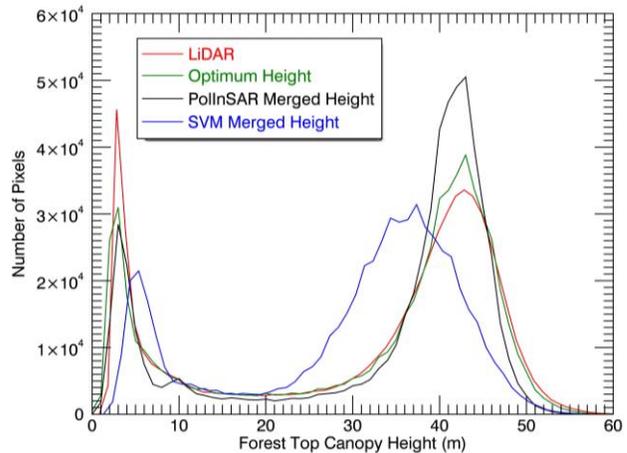
Fig. 9. Histograms of forest height estimated from different approaches discussed in this paper.

VI. CONCLUSION

This paper evaluated the capability of the PolInSAR height estimation using multi-baseline UAVSAR L- Band data over a heterogeneous tropical forest in Gabon. Merging PolInSAR measurements from the multiple baselines improved the result with respect to the single- baseline approach. In this paper, we demonstrated how multiple PolInSAR baselines can be successfully integrated with LiDAR samples using SVM improving previous approaches based directly on PolInSAR descriptors. The proposed approach does not require continuous LiDAR coverage where PolInSAR observations are available. As an outcome, the results presented in this paper support the idea that sparse LiDAR samples can be fused with PolInSAR measurements at the stage of PolInSAR multi-baseline selection, although we recognize that further algorithm assessment is required over forest sites with different vegetation and soil conditions. These results are particularly significant for future space borne missions (GEDI, BIOMASS, TanDEM-L, NISAR), which are expected to provide a wealth amount of LiDAR samples and PolInSAR measurements observations for global forest height and structure mapping.

REFERENCES

[1] A. T. Moles, D. I. Warton, L. Warman, N. G. Swenson, S. W. Laffan, A. E. Zanne, *et al.*, "Global patterns in plant height," *Journal of Ecology,* vol. 97, pp. 923-932, 2009.

[2] E. Pottier, J.-S. Lee, and L. Ferro-Famil, "Advanced Concepts In Polarimetry. Part 2: Polarimetric Target Classification," DTIC Document2005.

[3] S. R. Cloude and K. P. Papathanassiou, "Coherence optimisation in polarimetric SAR interferometry," in *Geoscience and Remote Sensing, 1997. IGARSS'97. Remote Sensing-A Scientific Vision for Sustainable Development., 1997 IEEE International*, 1997, pp. 1932-1934.

[4] S. R. Cloude and K. P. Papathanassiou, "Polarimetric SAR interferometry," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 36, pp. 1551-1565, 1998.

[5] K. P. Papathanassiou and S. R. Cloude, "Single-baseline polarimetric SAR interferometry," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 39, pp. 2352-2363, 2001.

[6] S. Cloude and K. Papathanassiou, "Three-stage inversion process for polarimetric SAR interferometry," *IEE Proceedings-Radar, Sonar and Navigation,* vol. 150, pp. 125-134, 2003.

[7] J. Praks, F. Kugler, K. P. Papathanassiou, I. Hajnsek, and M. Hallikainen, "Height estimation of boreal forest: Interferometric model-based inversion at L-and X-band versus HUTSCAT profiling scatterometer," *IEEE Geoscience and remote sensing letters,* vol. 4, pp. 466-470, 2007.

[8] I. Hajnsek, F. Kugler, S.-K. Lee, and K. P. Papathanassiou, "Tropical-forest-parameter estimation by means of Pol-InSAR: The INDREX-II campaign," *Ieee transactions on geoscience and remote sensing,* vol. 47, pp. 481-493, 2009.

[9] F. Garestier, P. C. Dubois-Fernandez, and I. Champion, "Forest height inversion using high-resolution P-band Pol-InSAR data," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 46, pp. 3544-3559, 2008.

[10] S.-K. Lee, F. Kugler, K. Papathanassiou, and I. Hajnsek, "Multibaseline polarimetric SAR interferometry forest height inversion approaches," in *Proceedings of ESA POLinSAR Workshop*, 2011.

[11] S.-K. Lee, F. Kugler, K. P. Papathanassiou, and I. Hajnsek, "Quantification of temporal decorrelation effects at L-band for polarimetric SAR interferometry applications," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* vol. 6, pp. 1351-1367, 2013.

[12] M. Lavalle and K. Khun, "Three-baseline InSAR estimation of forest height," *IEEE Geoscience and Remote Sensing Letters,* vol. 11, pp. 1737-1741, 2014.

[13] M. Lavalle and S. Hensley, "Extraction of structural and dynamic properties of forests from polarimetric-interferometric SAR data affected by temporal decorrelation," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 53, pp. 4752-4767, 2015.

[14] F. Kugler, S.-K. Lee, I. Hajnsek, and K. P. Papathanassiou, "Forest height estimation by means of Pol-InSAR data inversion: The role of the vertical wavenumber," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 53, pp. 5294-5311, 2015.

[15] A. Moreira, O. Ponce, M. Nannini, M. Pardini, and P. Prats, "Multi-baseline imaging: A vision for spaceborne SAR," in *Radar Symposium (IRS), 2015 16th International*, 2015, pp. 20-29.

[16] S. R. Cloude, "Robust parameter estimation using dual baseline polarimetric SAR interferometry," in *Geoscience and Remote Sensing Symposium, 2002. IGARSS'02. 2002 IEEE International*, 2002, pp. 838-840.

[17] S. Cloude, *Polarisation: applications in remote sensing*: Oxford University Press, 2010.

[18] R. F. Nelson, P. Hyde, P. Johnson, B. Emessiene, M. L. Imhoff, R. Campbell, *et al.*, "Investigating RaDAR–LiDAR synergy in a North Carolina pine forest," *Remote Sensing of Environment,* vol. 110, pp. 98-108, 2007.

[19] P. Hyde, R. Nelson, D. Kimes, and E. Levine, "Exploring LiDAR–RaDAR synergy—predicting aboveground biomass in a southwestern ponderosa pine forest using LiDAR, SAR and InSAR," *Remote Sensing of Environment,* vol. 106, pp. 28-38, 2007.

[20] E. Santi, S. Paloscia, S. Pettinato, G. Fontanelli, M. Mura, C. Zolli, *et al.*, "The potential of multifrequency SAR images for estimating forest biomass in Mediterranean areas," *Remote Sensing of Environment,* vol. 200, pp. 63-73, 2017.

[21] E. T. Mitchard, S. S. Saatchi, L. White, K. Abernethy, K. J. Jeffery, S. L. Lewis, *et al.*, "Mapping tropical forest biomass with radar and spaceborne LiDAR in Lopé National Park, Gabon: overcoming problems of high biomass and persistent cloud," 2012.

[22] M. García, S. Saatchi, S. Ustin, and H. Balzter, "Modelling forest canopy height by integrating airborne LiDAR samples with satellite Radar and multispectral imagery," *International Journal of Applied Earth Observation and Geoinformation,* vol. 66, pp. 159-173, 2018.

[23] F. Kugler, D. Schulze, I. Hajnsek, H. Pretzsch, and K. P. Papathanassiou, "TanDEM-X Pol-InSAR performance for forest height estimation," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 52, pp. 6404-6422, 2014.

[24] W. Qi and R. O. Dubayah, "Combining Tandem-X InSAR and simulated GEDI lidar observations for forest structure mapping," *Remote Sensing of Environment,* vol. 187, pp. 253-266, 2016.

[25] T. Le Toan, S. Quegan, M. Davidson, H. Balzter, P. Paillou, K. Papathanassiou, *et al.*, "The BIOMASS mission: Mapping global forest biomass to better understand the terrestrial carbon cycle," *Remote sensing of environment,* vol. 115, pp. 2850-2860, 2011.

[26] J. Reiche, R. Lucas, A. L. Mitchell, J. Verbesselt, D. H. Hoekman, J. Haarpaintner, *et al.*, "Combining satellite data for better tropical forest monitoring," *Nature Climate Change,* vol. 6, p. 120, 2016.

[27] G. Krieger, I. Hajnsek, K. Papathanassiou, M. Eineder, M. Younis, F. De Zan, *et al.*, "The tandem-L mission proposal: Monitoring earth's dynamics with high resolution SAR interferometry," in *Radar Conference, 2009 IEEE*, 2009, pp. 1-6.

[28] D. B. Coyle, P. R. Stysley, D. Poulios, G. B. Clarke, and R. B. Kay, "Laser transmitter development for NASA's Global Ecosystem Dynamics Investigation (GEDI) lidar," in *SPIE Optical Engineering+ Applications*, 2015, pp. 961208-961208-7.

[29] Qi and Dubayah. (2015, 01/06). *Fusing Next-generation Active Remote Sensing Data for Improved Forest Height and Structure Mapping.* Available: http://cce.nasa.gov/meeting_2015/abs_and_discussions/mtg2015_ab_searchab_id48.html

[30] R. Dubayah, "Crowd-Sourced Calibration: The GEDI Strategy for Empirical Biomass Estimation Using Spaceborne Lidar," in *2015 AGU Fall Meeting*, 2015.

[31] S. Lewis and N. Labrière, "Report for AfriSAR Gabon in situ field campaign," 2016.

[32] S. Hensley, H. Zebker, C. Jones, T. Michel, R. Muellerschoen, and B. Chapman, "First deformation results using the NASA/JPL UAVSAR instrument," in *Synthetic Aperture Radar, 2009. APSAR 2009. 2nd Asian-Pacific Conference on*, 2009, pp. 1051-1055.

[33] C. Brown and K. Sarabandi, "Estimation of red pine tree height using Shuttle Radar Topography Mission and ancillary data," in *Geoscience and Remote Sensing Symposium, 2003. IGARSS'03. Proceedings. 2003 IEEE International*, 2003, pp. 2850-2852.

[34] J. Kellndorfer, W. Walker, L. Pierce, C. Dobson, J. A. Fites, C. Hunsaker, *et al.*, "Vegetation height estimation from shuttle radar topography mission and national elevation datasets," *Remote sensing of Environment,* vol. 93, pp. 339-358, 2004.

[35] J. E. Anderson, L. C. Plourde, M. E. Martin, B. H. Braswell, M.-L. Smith, R. O. Dubayah, *et al.*, "Integrating waveform lidar with hyperspectral imagery for inventory of a northern temperate forest," *Remote Sensing of Environment,* vol. 112, pp. 1856-1870, 2008.

[36] R. N. Treuhaft, S. N. Madsen, M. Moghaddam, and J. J. Zyl, "Vegetation characteristics and underlying topography from interferometric radar," *Radio Science,* vol. 31, pp. 1449-1485, 1996.

[37] R. Touzi, A. Lopes, J. Bruniquel, and P. W. Vachon, "Coherence estimation for SAR imagery," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 37, pp. 135-149, 1999.

[38] S. R. Cloude, "POL-InSAR training course," *Radio Science,* 2005.

[39] P. A. Rosen, S. Hensley, K. Wheeler, G. Sadowy, T. Miller, S. Shaffer, *et al.*, "UAVSAR: A new NASA airborne SAR system for

science and technology research," in *Radar, 2006 IEEE Conference on*, 2006, p. 8 pp.

[40] F. Garestier, T. Le Toan, and P. Dubois-Fernandez, "Forest height estimation using P-band Pol-InSAR data," *Proc. Polinsar 2007,* 2007.

[41] K. Pelckmans, J. De Brabanter, J. A. Suykens, and B. De Moor, "Handling missing values in support vector machine classifiers," *Neural Networks,* vol. 18, pp. 684-692, 2005.

[42] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning,* vol. 20, pp. 273-297, 1995.

[43] B. Koetz, F. Morsdorf, S. Van der Linden, T. Curt, and B. Allgöwer, "Multi-source land cover classification for forest fire management based on imaging spectrometry and LiDAR data," *Forest Ecology and Management,* vol. 256, pp. 263-271, 2008.

[44] J. García-Gutiérrez, D. Mateos-García, M. Garcia, and J. C. Riquelme-Santos, "An evolutionary-weighted majority voting and support vector machines applied to contextual classification of LiDAR and imagery data fusion," *Neurocomputing,* vol. 163, pp. 17-24, 2015.

[45] M. Garcia, S. Saatchi, A. Casas, A. Koltunov, S. Ustin, C. Ramirez, *et al.*, "Quantifying biomass consumption and carbon release from the California Rim fire by integrating airborne LiDAR and Landsat OLI data," *Journal of Geophysical Research: Biogeosciences,* vol. 122, pp. 340-353, 2017.

[46] M. García, S. Saatchi, A. Casas, A. Koltunov, S. L. Ustin, C. Ramirez, *et al.*, "Extrapolating Forest Canopy Fuel Properties in the California Rim Fire by Combining Airborne LiDAR and Landsat OLI Data," *Remote Sensing,* vol. 9, p. 394, 2017.

[47] S. R. Gunn, "Support vector machines for classification and regression," *ISIS technical report,* vol. 14, pp. 85-86, 1998.

[48] F. Fassnacht, F. Hartig, H. Latifi, C. Berger, J. Hernández, P. Corvalán, *et al.*, "Importance of sample size, data type and prediction method for remote sensing-based estimations of aboveground forest biomass," *Remote Sensing of Environment,* vol. 154, pp. 102-114, 2014.

Maryam Pourshamsi received the B.S. degree in Urban Planning from the Yazd University, Yazd, Iran, in 2008 and the M.S.c Degree in Geographical Information System (GIS) from the University of Leicester, Leicester, UK in 2011. She is currently pursuing the Ph.D. degree in Geography at the Centre for Landscape and Climate Research of the University of Leicester, Leicester, UK. Her research interests include processing polarimetric and interferometric radar data for Earth Observation applications.

Mariano Garcia received the Ph.D. degree in Geographical Information Technologies from the University of Alcalá in 2011. He is currently an Assistant Professor at the Department of Geology, Geography and Environment, University of Alcala, Madrid, Spain. His research interests are environmental applications of remote sensing including forest fires and forest structure estimation using multisource data (LiDAR, SAR, passive optical) and machine learning.

Marco Lavalle is a Scientist in the Radar Science and Engineering Section at the Jet Propulsion Laboratory, California Institute of Technology (Pasadena, California). He is currently Principal Investigator for the NASA Terrestrial Ecology Program, and member of the NISAR and UAVSAR Project Science teams. He received the Master degree in Telecommunication Engineering from the University of Rome Tor Vergata (Rome, Italy) in 2006, and the Ph.D. degree in Electrical Engineering from the University of Rennes 1 (Rennes, France) and from the University of Rome Tor Vergata in December 2009. From January 2010 to December 2011, he was recipient of the NASA Postdoctoral Fellowship at the Jet Propulsion Laboratory, California Institute of Technology. His research interests include multi-dimensional image modeling and analysis, characterization of polarimetric-interferometric and tomographic radar signals for Earth science and applications, electromagnetic propagation, scattering theory, inverse theory, statistical estimation and parameter retrieval. Dr. Lavalle received the Student Prize Paper Award at the EUSAR Conference, Friedrichshafen, Germany in 2008.

Heiko Balzter received the degree of Dipl. Ing. agr. in 1994 and Dr. agr. in 1998 by Justus-Liebig-University in Giessen, Germany. He is a research professor and Director of the Centre for Landscape and Climate Research at the University of Leicester in the United Kingdom and leads the Official Development Assistance (ODA) Programme in the National Centre for Earth Observation (NCEO). Prof. Balzter holds the Royal Society Wolfson Research Merit Award (2011) and the Royal Geographical Society's Cuthbert Peek Award 'for advancing geographical knowledge of human impact through earth observation' (2015). He is the UK national representative on the Group on Earth Observations (GEO) Programme Board and has previously worked as Head of Section for Earth Observation for the Centre for Ecology and Hydrology, Monks Wood, UK. His research interests include interactions of the water cycle with ecosystems across multiple spatial and temporal scales, pressures from climate change and land use change on ecosystem services, and the effects of spatial patterns and processes upon biological populations in evolving, three-dimensional landscapes. Prof. Balzter has extensive expertise in Earth observation and remote sensing of the land surface and lakes, particularly of forests.