View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Institute of Transport Research:Publications

PRODUCTION OF A GLOBAL FOREST/NON-FOREST MAP UTILIZING TANDEM-X INTERFEROMETRIC SAR DATA

Christopher Wecklich, Michele Martone, Paola Rizzoli, José-Luis Bueso-Bello, Carolina Gonzalez, Gerhard Krieger

Microwaves and Radar Institute, German Aerospace Center, Christopher.Wecklich@dlr.de

ABSTRACT

In this paper we describe the method that has been implemented to derive the forest/non-forest maps from TanDEM-X interferometric synthetic aperture radar (InSAR) data, globally acquired in stripmap single polarization (HH) mode. Among the several observables systematically provided by the TanDEM-X system, the volume decorrelation contribution, derived from the interferometric coherence, shows to be consistently sensitive to the particular land cover type, and is therefore used as an input data set for applying a classification method based on a fuzzy clustering algorithm. Since the considered InSAR quantity strongly depends on the geometric acquisition configuration, namely the incidence angle and the interferometric baseline, a multi-clustering classification approach is used. Once the Forest/NonForest classification for individual acquisitions is generated, overlapping acquisitions are mosaicked together to improve the resulting accuracy. The final step in the Forest/NonForest map production is to apply a binary Forest/Non-Forest decision and the decision threshold is found through comparison with similar data and statistical analysis. Verification and validation of the final product will be accomplished through comparison to other forest maps. In summary, this paper covers the processing and production status of the global TanDEM-X Forest/Non-Forest map which will be made available to the scientific community in 2017.

Index Terms—Synthetic aperture radar (SAR), SAR interferometry (InSAR), TanDEM-X mission, volume decorrelation, forest classification.

1. INTRODUCTION

A precise and up-to-date knowledge of land cover information is of great importance for a wide range of scientific and commercial purposes. Specific to this paper is the identification and the monitoring of vegetated areas which is critical for a variety of applications, such as agriculture, cartography, geology, forestry, global change research, as well as for regional planning. In this paper we present a method to generate forest classification maps from TanDEM-X interferometric SAR data. The TanDEM-X mission comprises the two twin satellites TerraSAR-X and TanDEM-X with the main goal of producing a global and consistent Digital Elevation Model (DEM) with an unprecedented accuracy, by exploiting single-pass SAR interferometry [1]. Several observables are systematically provided by the TanDEM-X system, and of great importance to this paper is the volume correlation factor, which quantifies the coherence loss due to volume scattering and represents the contribution which is predominantly influenced by the presence of vegetation. Since the beginning of the TanDEM-X mission (end of 2010), about half a million of high-resolution bistatic single polarization (HH) scenes covering all the Earth's land masses have been acquired and processed. From these, quicklook images representing several SAR and InSAR quantities (like backscatter and coherence maps, or the calibrated RawDEM) are generated at a ground pixel spacing of about 50 m x 50 m by applying a spatial averaging process. Working with such reduced resolution data makes feasible the exploitation of the TanDEM-X dataset on a global scale, as it keeps the computational load low. In this paper, we extend and apply the method presented in [2] and [3], which are revisited in sections 2 (volume correlation factor extraction) and 3 (multi-clustering classification). In sections 4 and 5, the method for mosaicking the overlapping observations into a single map and the calculation of the threshold for the binary forest/non-forest decision is described. In the following section, the validation approach and global forest map final product are presented. The paper is concluded in Section 7 with a final summary.

2. EXTRACTING VOLUME CORRELATION FACTOR FROM TANDEM-X SAR DATA

Several contributions cause a coherence loss in TanDEMX interferometric data [4] and, assuming statistical independence, can be factorized as follows:

$$\gamma = \gamma_{SNR} \cdot \gamma_{Quant} \cdot \gamma_{Amb} \cdot \gamma_{Range} \cdot \gamma_{Azimuth} \cdot \gamma_{Temp} \cdot \gamma_{Vol} \quad (1)$$

The terms on the right-hand side of the equation above describe the decorrelation due to: limited signal-to-noise ratio (γ_{SNR}), quantization errors (γ_{Quant}), ambiguities (γ_{Amb}), baseline decorrelation (γ_{Range}), errors due to relative shift of Doppler spectra ($\gamma_{Azimuth}$), temporal decorrelation (γ_{Temp}),

and volume scattering (γ_{Vol}). This last term, here called the volume correlation factor, represents the contribution which is predominantly affected by the presence of vegetation. Given a coherence estimate, the volume correlation factor contribution can be found solving Equation 1 for γ_{Vol} as shown below:

$$\gamma_{Vol} = \frac{\gamma}{\gamma_{SNR}\cdot\gamma_{Quant}\cdot\gamma_{Amb}\cdot\gamma_{Range}\cdot\gamma_{Azimuth}\cdot\gamma_{Temp}} \quad (2)$$

The impact and the estimation procedure for each correlation contribution in the TanDEM-X data is discussed in detail in [2]. For each quicklook product a local incidence angle map is derived from the orbit parameters and the calibrated RawDEM. This allows us to improve the estimation accuracy and, in particular, to precisely compensate for possible geometrical decorrelation due to the presence of topography. To demonstrate the effectiveness of this procedure, we have used the 30m resolution global vegetation continuous fields tree cover data from the multispectral sensor Landsat-5 developed in 2005 and freely available online [5], which provides the percentage of area covered by vegetation. Pixels having a value smaller than 15% and larger than 65% have been selected as representative for non-vegetated and forested areas, respectively. The corresponding occurrence distribution of volume correlation factor estimated from TanDEM-X data is depicted in Figure 1 for densely vegetated areas and for bare surfaces. This figure clearly shows that this information is a powerful indicator for forest classification purposes.

3. CLASSIFICATION VIA FUZZY LOGIC

Clustering entails grouping a set of objects coming from N input observations $\mathbf{Y} = [\mathbf{y}_k]$ ($k = 1,\ldots,N$), each one characterized by a set of P features, depending on how similar they are to each other. The observations are then divided into c nonempty subsets, called clusters and in fuzzy-clustering a certain amount of overlap between different clusters is allowed [6]. A membership function $\mathbf{U}$ is defined, which describes the probability of an observation belonging to each cluster ($\widehat{\mathbf{U}} = [\hat{u}_{ik}] \in [0,1], i = 1,\ldots,c$). The results are c fuzzy partitions of the input observation data set, which contain observations characterized by a high intracluster similarity and a low extracluster one. In the case of TanDEM-X data, we exploit the volume correlation factor information only (i.e. $P = 1$). The number of clusters is set to two, to discriminate forest (F) from non-forest (NF) areas. The cluster centers $\{\upsilon_F, \upsilon_{NF}\}$ are identified by their feature $\{\gamma_{Vol,F}, \gamma_{Vol,NF}\}$. In [2] it is verified that the coherence loss over forest is notably influenced by the specific incidence angle. For this reason a partitioning of the original data into S subsets is performed, depending on the specific pair of baseline $B_\perp$ (i.e. height of ambiguity) and incident

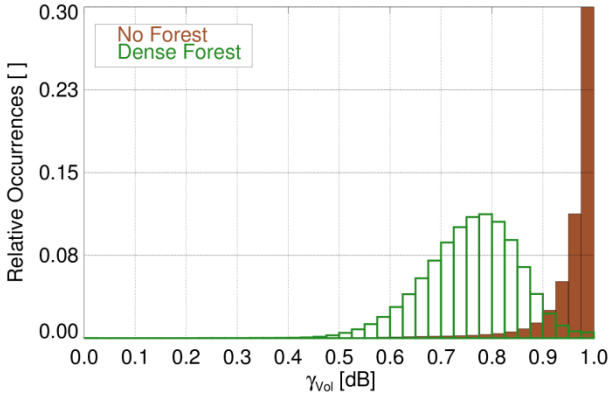

Figure 1: Occurrence distribution of volume correlation factor contribution estimated for densely forested and non-vegetated areas, depicted in green and brown, respectively.

angle θ_{inc}. Therefore, an observation (pixel) k is associated to the i-th subset if

$$B_{\perp\,k} \in [B_{\perp\,i,\min}, B_{\perp\,i,\max}], \quad (3)$$

$$B_{inc\,k} \in [B_{inc\,i,\min}, B_{inc\,i,\max}]. \quad (4)$$

An important step in the cluster algorithm is the definition of the cluster centers, each one identified by a P-dimensional tie-point vector υ_i (P being the number of features). For their determination TanDEM-X bistatic data takes acquired over a large region located in the Amazon rainforest are "trained" by using the forest density information provided by Landsat [5] (see also Figure 1). For each subset i, the sample expectation of γ_{Vol} is taken as cluster center. Figure 2 shows the cluster centers as points for forest and non-forest areas, for different heights of ambiguity as well as the local incidence angle ranges. Near, mid, and far range are identified with angles steeper than 35º, between 35º and 45º, larger than 45º, respectively. A minimum mean square error fitting is applied for more continuous distribution, which is indicated by the continuous lines. In addition to defining the cluster centers for rainforest, the same process is used for defining cluster centers over temperate and boreal forests and all three

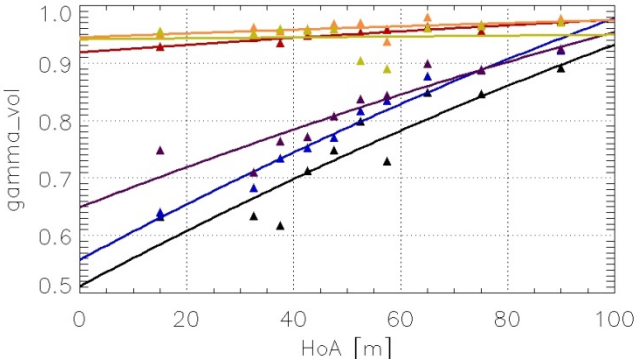

Figure 2: The points indicate the volume correlation factor cluster centers for a large region located in the Amazon rain forest (Brazil), and calculated according to (2).

cluster centers are used in the final global forest map production.

The output of the clustering is a weighted membership which can be interpreted as a sort of probability of a pixel to be "covered" by vegetation. Quality weights are introduced depending on the training set statistics.

4. MOSAICKING INTO A SINGLE FOREST MAP

For an area of interest, multiple overlapping forest observations can be used to create a single forest map. To mosaic the overlapping forest observations together, a measurement of the quality for each overlapping pixel is needed. Returning to Figure 2, we see that with a decrease in HoA comes an increase in the separation between forest and non-forest classifications. Thus a lower HoA has a better reliability in the classification. To this end, the following metric is used in the mosaicking of overlapping pixels.

$$\Delta M_i \triangleq \frac{1}{\Delta\gamma_{Vol}\cdot\gamma_{SNR}} - 1 \qquad (6)$$

The correlation factor due to limited signal-to-noise ratio (γ_{SNR}) is included in the above metric to reduce the effects of poor SNR that can be found on the edges of an acquisition. Using the metric in (6), a weighted mean method can be utilized to combine all overlapping membership pixels into a combined membership pixel of the forest map with the weighting and combined membership defined as.

$$\alpha_i \triangleq \Delta M_i^{-2} \cdot \left(\sum_{k=1}^{N} \Delta M_k^{-2}\right)^{-1} \qquad (7)$$

$$\widetilde{M}_{Comb} = \sum_{i=1}^{N} \alpha_i \widetilde{M}_i \qquad (8)$$

Using equations (6) through (8) above, a forest map with pixel values ranging between 0 (No Forest) and 1 (Forest) is created and the left side of Figure 3 below shows the results of the mosaicking process for a section located at N40W078 in Pennsylvania.

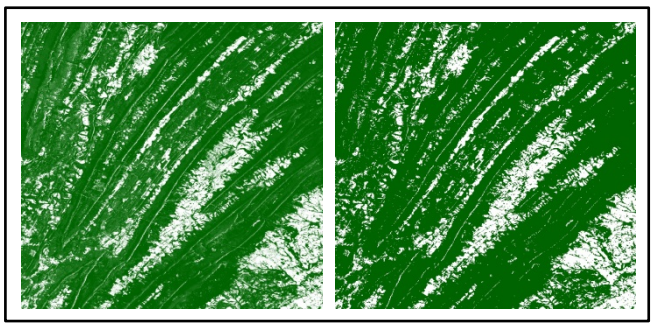

Figure 3: Pennsylvania Test site for mosaicking located at N40W078. Left shows the mosaicked forest map. Right shows the Forest map after the binary decision is applied

5. BINARY THRESHHOLD CALIBRATION

The mosaicking process described in the previous section produces a Forest/Non-Forest map with pixel values ranging between 0 and 1, and the last processing step for the final product will be to convert this into a binary Forest/Non-Forest map. To accomplish this, a threshold must be established and as a test case for temperate forests, the TanDEM-X Forest/Non-Forest dataset is compared to the Lidar/Optical dataset covering Pennsylvania [7] with the focus of this calibration tile N40W078. As the Lidar/Optical (L/O) data has a resolution of 1m x 1m, the data is first scaled to the TanDEM-X forest resolution of 50m x 50m where each pixel represents the percent of 1m x 1m L/O forest pixels that overlap it. Furthermore, only a subset of the tile N40W078 is used in order to exclude urban areas and farmland. The left side of Figure 4 shows the 2-demensional histogram of the pixel-by-pixel comparison between the TanDEM-X and Lidar/Optical datasets. In this figure, the bottom left corner is the case where both datasets find Non-Forest and the top right is when both find Forest. The other two corners are the cases when the two datasets disagree. From this histogram, the φ coefficient is constructed for each forest percent value to evaluate the changes in both datasets as shown below:

$$\varphi = \frac{TP\cdot TN - FP\cdot FN}{\sqrt{P\cdot RP\cdot RN\cdot N}} \qquad (9)$$

Where:

TP = True Positive	TN = True Negative
FP = False Positive	FN = False Negative
P = Total Positive	N = Total Negative
RP = Total Ref. Positive	RN = Total Ref. Negative

The results of the φ coefficient are shown on the right side of Figure 4. From the φ coefficient plot, the optimal threshold is found to be 60% for the TanDEM-X dataset and this corresponds to a forest coverage of 25% in the L/O dataset. The result of a binary decision application using this threshold to the forest map is shown in on the right side of Figure 3 for the same section as the left side.

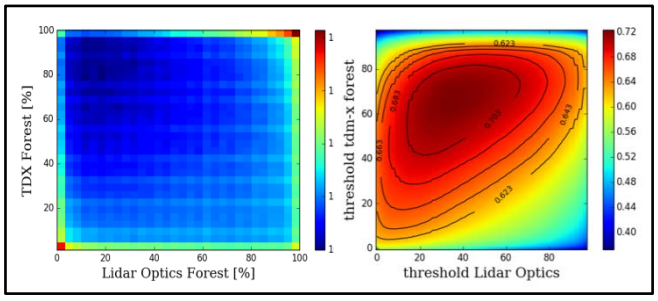

Figure 4: Left side showing the 2D histogram of TanDEM-X and Lidar/Optics Forest pixel data. Right side is the φ coefficient graph showing the optimal threshold to be 60% for TanDEM-X and 25% for Lidar/Optical

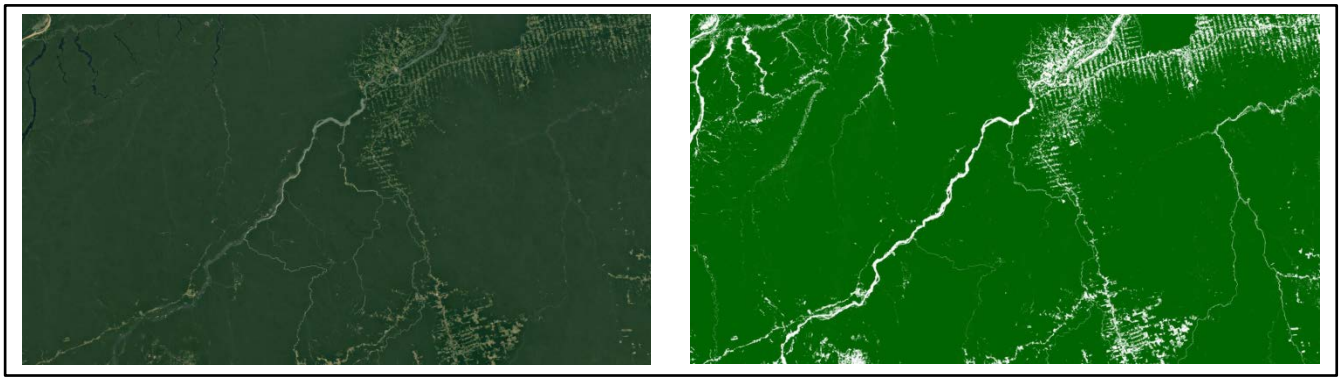

Figure 5: The Amazon rainforest with Google Earth Optical Image (Left) and TanDEM-X Forest/Non-Forest map (Right)

6. VERIFICATION AND FINAL PRODUCT

Verification of the final dataset quality will be performed firstly, by a performance comparison against existing Forest/Non-Forest maps such as Landsat [5], Palsar [8], and CORINE Standard [9]. Secondly, the dataset will be validated against the Pennsylvania Lidar/Optic dataset and the CORINE high resolution Forest map in Europe.

Figure 5 shows the application of the Forest classification method described in this paper for a region over the Amazon rainforest. On the left is the optical image provided by Google Earth and on the right side is the corresponding TanDEM-X forest/non-forest map.

The final Forest/Non-Forest dataset product will provide a global coverage based on TanDEM-X interferometric SAR data. The resolution of this TanDEM-X forest map will be 50m x 50m with each pixel being A binary Forest/Non-Forest value. It is expected to finish the TanDEM-X Forest map dataset by mid-2017. It will be made freely accessible to the scientific community.

7. CONCLUSION

In this paper, a method for the generation of a Forest/non-forest map from TanDEM-X interferometric SAR data has been proposed. It has been shown that from the TanDEM-X quick looks with a resolution of 50m x 50m meters the volume correlation factor can be extracted and used for vegetation identification. Fuzzy-clustering has been introduced as a method to translate from the volume correlation factor to a weighted membership in forest / non-forest clusters. A method for reliably mosaicking the overlapping observations into a final forest map has been presented as well as the determination of the binary threshold between forest and non-forest. The validation of the final product has been described in the last section.

8. REFERENCES

[1] G. Krieger, A. Moreira, H. Fiedler, I. Hajnsek, M. Werner, M. Younis, and M. Zink, "TanDEMX: A Satellite Formation for High-Resolution SAR Interferometry", IEEE Trans. Geosci. Remote Sens., Vol. 1, N. 11, pp. 3317–3341, Nov. 2007.

[2] M. Martone, P. Rizzoli, B. Bräutigam, and G. Krieger, "A method for generating forest/nonforest map from TanDEM-X interferometric data", IEEE Int. Geosci. Remote Sens. Symp., Milan (Italy), July 2015.

[3] M. Martone, P. Rizzoli, B. Bräutigam, and G. Krieger, "Forest/non-forest classification from TanDEM-X Interferometric Data by means of Multiple Fuzzy Clustering", European Conference on Synthetic Aper-ture Radar, Hamburg, Germany, pp 681-686, 2016

[4] M. Martone, B. Bräutigam, P. Rizzoli, C. Gonzalez, M. Bachmann, and G. Krieger, "Coherence Evaluation of TanDEM-X Interferometric Data", ISPRS J. of Photogr. Remote Sens., Vol. 73, pp. 21-29, Sep. 2012.

[5] J. O. Sexton, X.-P. Song, M. Feng, P. Noojipady, A. Anand, C. Huang and D.-H. Kim, K. M. Collins, S. Channan, C. DiMiceli, and J. R. Townshend, "Global, 30-m resolution continuous fields of tree cover: Landsat-based rescaling of MODIS Vegetation Continuous Fields with lidar-based estimates of error", International Journal of Digital Earth, Vol. 6, N. 5, pp. 427-448, 2013.

[6] J. Bezdek, R. Ehrlich, and W. Full, "FCM: the fuzzy c-means clustering approach", Computers and Geosciences, Vol. 10, pp. 191-203, December 1984.

[7] Pennsylvania Statewide High-Resolution Tree Canopy, University of Vermont Spatial Analysis Laboratory, http://www.pasda.psu.edu/uci/DataSummary.aspx?dataset=3170, (Visited on 23.12.16)

[8] M. Shimada, T. Itoh, T. Motooka, M. Watanabe, T. Shiraishi, R. Thapa, and R. Lucas, "New global forest/non-forest maps from ALOS PALSAR data (2007-2010)", Remote Sens. of Env., Vol. 155, pp. 13-31, May 2014.

[9] CORINE Land Cover Database 2012, - http://www.eea.europa.eu/data-and-maps/data/clc-2012-vector (Visited on 23.12.16)