GLOBAL FOREST CLASSIFICATION FROM TANDEM-X INTERFEROMETRIC DATA: POTENTIALS AND FIRST RESULTS

Michele Martone, Paola Rizzoli, Benjamin Bräutigam, Gerhard Krieger

Microwaves and Radar Institute – German Aerospace Center, 82234 Weßling (Germany),
Email:Michele.Martone@dlr.de

ABSTRACT

This paper presents a method to generate forest/non-forest maps from TanDEM-X interferometric SAR data. Among the several contributions which may affect the quality of interferometric products, the coherence loss caused by volume scattering represents the contribution which is predominantly affected by the presence of vegetation, and is therefore here exploited as main indicator for forest classification. Due to the strong dependency of the considered InSAR quantity on the geometric acquisition configuration, namely the incidence angle and the interferometric baseline, a multi-fuzzy clustering classification approach is used. Some examples are provided which show the potential of the proposed method. Further, additional features such as urban settlements, water, and critical areas affected by geometrical distortions (e.g. shadow and layover) need to be extracted, and possible approaches are presented as well. Very promising results are shown, which demonstrate the potentials of TanDEM-X bistatic data not only for forest identification, but, more in general, for the generation of a global land classification map as a next step.

1. INTRODUCTION

Land cover classification by means of remote sensing data is of fundamental importance for a broad range of commercial and scientific applications. In particular, the detection of vegetated areas is of great interest for global change research and for relevant uses in agriculture, cartography, geology, forestry, and for regional environmental planning. At a global level, only in the last years high-resolution forest classification maps have been produced [1], [2]. This paper presents a method to generate forest/non-forest maps from TanDEM-X interferometric SAR data. The TanDEM-X mission comprises the two twin satellites TerraSAR-X and TanDEM-X, which fly in close orbit formation at a distance of a few hundred meters, and act as a large single-pass SAR interferometer. TanDEM-X is going to produce a global and consistent Digital Elevation Model (DEM) with an unprecedented accuracy [3] by the end of 2016. The interferometric coherence represents a key quantity for assessing the InSAR performance, describing the amount of noise affecting the interferogram. Among the several error sources which impact the quality of interferometric data, the coherence loss caused by volume scattering represents the contribution which is predominantly affected by the presence of vegetation.

Since the beginning of the TanDEM-X mission (end of 2010), about half a million high-resolution bistatic InSAR scenes have been acquired covering all the Earth's land masses. A single bistatic scene has typically a ground extension of about 30 km in range by 50 km in azimuth. From each of them, quicklook images for several SAR and InSAR quantities (like the coherence matrix or interferometric rawDEM [4]) are generated by applying a spatial averaging process to the corresponding operational TanDEM-X interferometric data in full resolution. They have a ground resolution of about 50 m × 50 m, and are taken into account as input data base for generating quicklook mosaics. Hence, quicklook images of the TanDEM-X coherence-derived volume decorrelation are used as input for applying a classification method based on the fuzzy clustering logic, whose potentials in the context of the TanDEM-X mission have been already shown in [5]. During the whole mission duration, all land masses have been acquired at least twice. In order to further improve the performance, many densely forested areas have been covered up to four times, in a time frame of several years (from 2010 to 2015). Such a unique, timely, and manifold data set can be exploited to get up-to-date information on and to detect possible changes in the forest cover. To improve the classification capabilities, invalid areas, such as water bodies, urban settlements, and regions affected by strong geometrical distortions need to be filtered out. The preliminary classification results shown in this paper are very promising not only for forest identification, but also for the generation of a global land classification map as a next step.

2. VOLUME DECORRELATION FOR FOREST CLASSIFICATION

Several contributions may affect the coherence of TanDEM-X interferometric data [3]. Among them, the volume decorrelation γ_{Vol} describes the impact of multiple scattering within a volume, and is mainly influenced by the presence of vegetation. Hence, given a coherence estimate γ, one can derive γ_{Vol} as [3]

$$\gamma_{Vol} = \frac{\gamma}{\gamma_{SNR} \cdot \gamma_{Quant} \cdot \gamma_{Amb} \cdot \gamma_{Geom} \cdot \gamma_{Temp}},$$
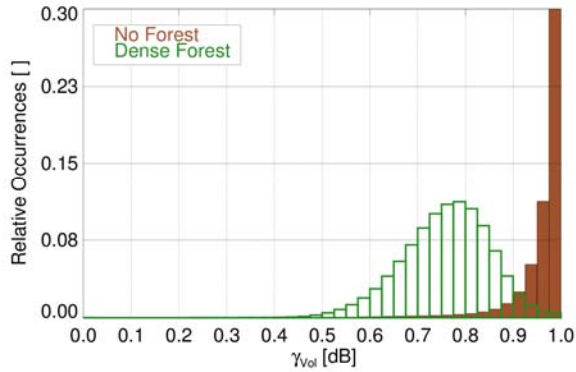
Figure 1: Distribution of the volume decorrelation for densely forested and non-vegetated areas, depicted in green and brown, respectively, estimated from TanDEM-X stripmap single-pol HH bistatic acquisitions over a large region located in the Amazon rainforest. Heights of ambiguity between 40 m and 50 m are considered. For land cover identification, the "vegetation continuous fields tree cover" data set provided by the multispectral sensor Landsat-5 has been used.

being the terms on the right-hand side the decorrelation contributions due to limited signal-to-noise ratio (γ_{SNR}), quantization errors (γ_{Quant}), ambiguities (γ_{Amb}), errors due to relative shift of range and/or Doppler spectra (γ_{Geom}), and temporal decorrelation ($\gamma_{Temp}=1$ for bistatic TanDEM-X acquisitions).

A local incidence angle map is derived from the orbit parameters and from the calibrated RawDEM for each quicklook product, which allows to precisely compensate for possible geometrical decorrelation due to topography. The occurrence distributions of volume decorrelation, estimated from TanDEM-X data for densely vegetated and bare surfaces, are depicted in Fig. 1. For land cover discrimination we have used the freely available "vegetation continuous fields tree cover" data set (30 m resolution) provided by the multispectral sensor Landsat-5 in 2005, providing the percentage of area covered by vegetation (for this example, pixels having a density value smaller than 15% and larger than 65% have been selected as representative for non-vegetated and forested areas, respectively). The two land cover types appear to be well separable and very useful for forest classification purposes. We also investigated the opportunity to exploit the amplitude information for classification purposes. However, the information contained in the HH polarization channel, which has been selected for the global TanDEM-X DEM acquisition, shows to be less reliable, being strongly influenced by several parameters, such as soil roughness or moisture variation. For this reason we consider the coherence information only for classification purposes, as detailed in the next section.
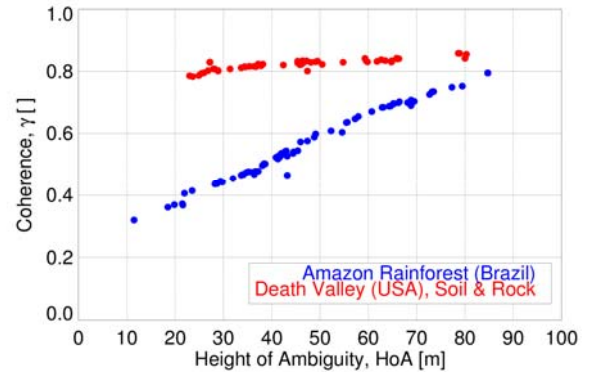

Figure 2: Coherence over height of ambiguity. For the "Death Valley" test site (in red) almost no influence of the baseline on the coherence is observed. On the other hand, for densely forested areas (in blue) the volume decorrelation strongly depends on the acquisition geometry.

3. MULTIPLE FUZZY CLUSTERING FOR FOREST CLASSIFICATION

Fuzzy clustering represents a powerful and effective class of methods, widely used in numerous contexts and applications, such as data mining or pattern recognition. In the most general sense, clustering identifies the task of grouping N input observations together, each one characterized by a set of P features, depending on how similar they are to each other. With fuzzy logic, furthermore, a certain amount of overlap between different clusters is allowed [6]. The observations are divided into c non-empty subsets, called clusters. The degree of affinity of a given observation to each subset can be described by the so-called membership function according to the c-means clustering method [6]. The results are fuzzy c-partitions of the input data set, which contain observations characterized by a high intracluster similarity and a low extracluster one. For our purposes two clusters ($c = 2$) are set, to discriminate between vegetated from non-vegetated areas. As previously explained, for classification purposes the volume decorrelation information only is exploited ($P = 1$). Along with the specific type of forest (characterized by, e.g., their canopy density and tree height), the volume decorrelation is influenced on the specific acquisition geometry. As an example, the interferometric coherence for two test areas characterized by different land covers is shown in Fig. 2, as a function of the height of ambiguity (h_{amb}). The ambiguous height gives information about the phase-to-height sensitivity of the interferogram, $h_{amb}=\lambda r \sin(\theta_{inc})/B_\perp$, being λ the radar wavelength, r the slant range, θ_{inc} the incidence angle, and $B_\perp$ the baseline perpendicular to the line of sight. The two test sites (depicted with different colors) have been repeatedly acquired for long-term monitoring of
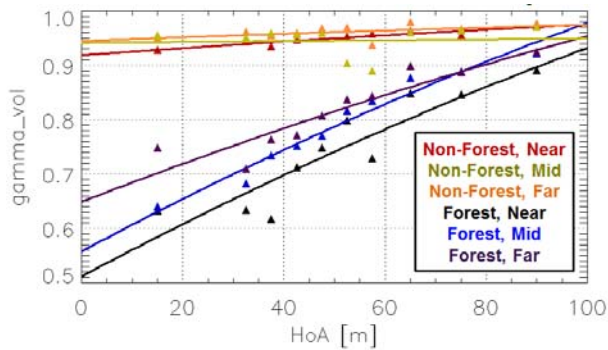
Figure 3: Clusters centers (markers) and MMSE fitting (lines) for multi-clustering classification. The clusters are derived from a large area located in the Amazon rainforest (Brazil), for forest and non-forest areas, and for the corresponding height of ambiguity and the local incidence angle range, depicted with different colors. For the discrimination of near, mid, and far range two thresholds have been set at 35° and 45°.

Figure 4: (Left) GoogleEarth optical image of an area located in the Amazon rainforest, Brazil and (right) forest classification map. The considered area extends by about 55 km in azimuth and 30 km in range.

the interferometric performance since mission start, by exploiting the change over time of the helix formation (i.e. of the interferometric baseline $B_\perp$). The "Death Valley" test site (red dots) is characterized by a non-vegetated, rocky land cover, and shows a substantial stability of the coherence over h_{amb} (and over time). For this test site, the main source of coherence loss (about 0.8) is caused by the limited SNR, which can be assumed constant for the whole time series. On the other hand, the "Amazon rainforest" test site is covered by dense forest and the coherence is consequently affected by the InSAR acquisition geometry. In [7] it is verified that the decorrelation over forest is notably influenced by the local incidence angle. In particular, for shallow angles the surface scattering component becomes dominant, and a high coherence is typically observed. On the other hand, when steeper slopes are imaged (with respect to the sensor's line of sight) the X-band waves are able to penetrate through the canopy, resulting in a stronger volume scattering contribution and, in turn, in a larger coherence loss.

For all the described reasons, it makes sense to perform a partitioning of the original data into S subsets, according to the specific pair of baseline (i.e. height of ambiguity) and incident angle. For the definition of the cluster centers, TanDEM-X bistatic data takes acquired all around the world over different forest types are "trained" by using the forest density information provided by Landsat [1] (see also Fig. 1). For each subset the sample expectation of γ_{Vol} is finally taken as cluster center, as shown in Fig. 3. The markers indicate the cluster centers for forest and non-forest areas, for different heights of ambiguity and local incidence angle ranges, derived from a large area located in the Amazon rainforest (Brazil). For the discrimination of near, mid,

and far range two thresholds have been set at 35° and 45°. For each of them, we applied a minimum mean square error (MMSE) fitting to derive a consistent subcluster distribution, as shown by the continuous lines. Then, for each subset a fuzzy clustering classification algorithm is run independently. The obtained output is the so-called membership function, which describes the probability of a certain observation (pixel) to belong to a cluster and is derived as for the c-means fuzzy clustering algorithm [6]. As a next step one could then try to link this quantity to the local vegetation properties (density, height) which represents an important parameter for several forestry applications such as e.g. biomass estimation.

4. CLASSIFICATION EXAMPLES AND ADDITIONAL INFORMATION LAYERS

The forest classification of a bistatic TanDEM-X acquisition over a vegetated region in Brazil, overlaid on a GoogleEarth optical image, is shown on the right-hand side of Fig. 4. Dense forest and non-forest areas are depicted in dark and light green, respectively, and are clearly distinguishable. On the left-hand side of the figure the optical image of the area is given for comparison. The same classification is depicted for an area over temperate forest in Bavaria (Germany) on the right-hand side of Fig. 5. In general, a good agreement with the optical image (on the left) is observed. The particular values assumed by the membership function strongly depends on how much separable the cluster centers effectively are, which is strictly related to the imaging geometry employed for the specific acquisition (see Fig. 3). Urban areas are highlighted on the right-hand side of Fig. 5 and for their identification a dedicated algorithm which exploits local statistics of the SAR amplitude information (depicted in the centre of the figure as well) has been implemented. Promising results have been obtained so far.
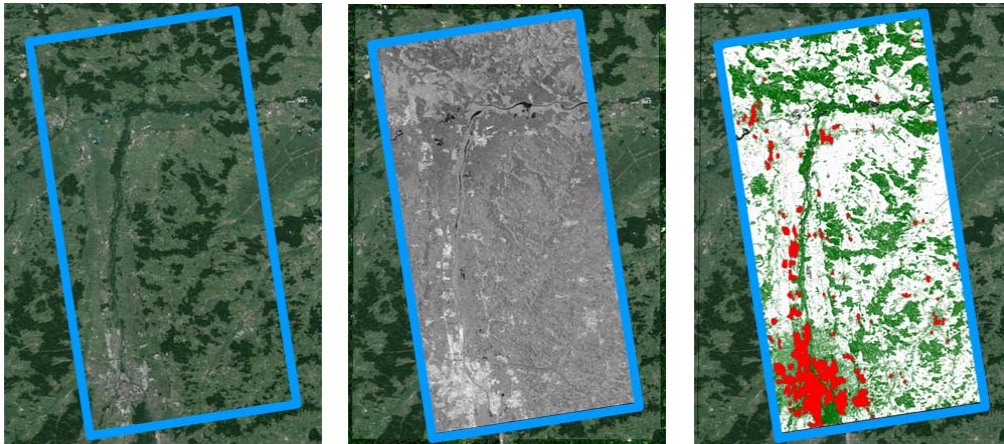
Figure 5: (Left) GoogleEarth optical image of an area located in the Bavaria region, Germany and (right) forest classification map. Urban settlements can be identified by exploiting the amplitude information, which is depicted in the middle for comparison.

Additionally, water areas as well as regions strongly affected by geometrical distortions (such as shadow or layover) are also identified by exploiting the coherence, the amplitude, together with the local incidence angle information directly derived from orbit information and the available TanDEM-X DEM.

The unique and multi-temporal TanDEM-X data set may be also exploited to detect and monitor possible temporal changes in the forest cover. An optical image acquired in the 1970's by the first Landsat satellites over an area located in the Amazon rainforest is shown in Fig. 6 (a). Fig. 6 (b) and (c) show the coherence quicklooks acquired by TanDEM-X over the same area in January 2011 and October 2012, respectively. The effects of about 35 years of deforestation activities are clearly visible. The slightly different baselines explain the higher coherence observed in Fig. 6 (c) with respect to Fig. 6 (b). The forest loss that occurred between the two TanDEM-X acquisitions can be assessed by simply comparing the two coherence maps, as shown by the brown spots in Fig. 6 (d). The red circles in the figures highlight the areas where forest cover changes are most visible.

5. CONCLUSIONS AND OUTLOOK

In this paper we have presented a method to classify forested and non-forested areas from TanDEM-X bistatic data. To do this we exploit fuzzy clustering logic, for which the volume decorrelation derived from the interferometric coherence is used as input feature. Due to the strong influence of the specific acquisition geometry (height of ambiguity and incidence angle) on the considered InSAR quantity, a multi-clustering approach is proposed. To improve the classification accuracy, additional information layers such as urban areas regions, water bodies or difficult terrain need to be identified, and some promising examples have been discussed.

During the whole TanDEM-X mission duration, the global land masses have been acquired at least twice, and many densely forested areas have been covered up to four times in a time span of several years (from 2010 to 2015). Such a unique data set can be exploited for several purposes, such as

1. To get a global up-to-date forest cover information;
2. To improve the classification accuracy by properly mosaicking the multiple coverages;
3. For detecting possible temporal changes in the forest cover (as shown in Fig. 6).

As next steps, the obtained results will be quantitatively validated with existing vegetation maps and by means of external land cover classification data.

6. REFERENCES

1. Global Forest Change, http://earthenginepartners. appspot.com/science-2013-global-forest, [Online 2013].

2. Shimada, M., Itoh T., Motooka T., Watanabe M., Shiraishi, T., Thapa, R. & Lucas, R. *Newglobal forest/nonforest maps from ALOS PALSAR data (2007/2010),* Remote Sens. of Env., vol. 155, pp. 13–31, May 2014.

3. Krieger, G., Moreira, A., Fiedler, H., Hajnsek, I., Werner, M., Younis, M., & Zink M., *TanDEM-X: A satellite formation for high-resolution SAR interferometry*, IEEE Trans. Geosci. Remote Sens., vol. 45, no. 11, pp. 3317–3341, November 2007.

4. Fritz T., Rossi C., Yague-Martinez N. Rodriguez-Gonzalez F., Lachaise M. & Breit, H., *Interferometric processing of TanDEM-X data*, Proc. IGARSS, July 2011, Vancouver, Canada, pp. 2428–2431.

Figure 6: (a): Optical image acquired by the Landsat sensor in the 1970's over a region located in the Amazon rainforest. Coherence quicklook maps of a TanDEM-X acquisition commanded on January 2011 (b) and October 2012 (c). (d): Forest loss map (brown: loss, green, no loss) occurred between the two TanDEM-X acquisitions. The red circles in figures (b)-(d) highlight the areas where forest cover changes are most visible.

5. Rizzoli P., Martone M., & Bräutigam B., *Greenland ice sheet snow facies identification approach using TanDEM-X interferometric data*, Proc. IGARSS, July 2015 Milan, Italy, pp. 2060–2063.

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

7. Martone M., Rizzoli P., Bräutigam B. & Krieger, G., *A method for generating forest/non-forest map from TanDEM-X interferometric data*, Proc. IGARSS, July 2015 Milan, Italy, pp. 2634–2637.