# The Global Forest/Non-Forest Classification Map from TanDEM-X Interferometric Data

Michele Martone, Paola Rizzoli, Carolina González, José-Luis Bueso-Bello, Manfred Zink,
Gerhard Krieger, and Alberto Moreira
Microwaves and Radar Institute, German Aerospace Center, Michele.Martone@dlr.de, Germany

## Abstract

In this paper we present the global Forest/Non-Forest Map derived from TanDEM-X bistatic interferometric synthetic aperture radar (InSAR) data. The global TanDEM-X dataset has been acquired in stripmap single HH polarization mode and covers a time span from 2011 up to 2016. The volume correlation factor (or volume decorrelation), $\gamma_{\mathrm{Vol}}$, derived from the interferometric coherence, quantifies the coherence loss due to multiple scattering within a volume, a mechanism which typically occurs in presence of vegetation. For this reason, the $\gamma_{\mathrm{Vol}}$ has been used as main indicator for the identification of forested areas. Quicklook images, a multi-looked version of the original full-resolution data at a ground resolution of 50 m $\times$ 50 m, have been used as input for the generation of the global product. The mosaicking process of multiple acquisitions is discussed as well, together with the identification of additional information layers, such as urban areas or water bodies. The resulting global forest/non-forest map has been validated using external reference information, as well as with other existing classification maps, and an overall agreement typically exceeding 90% is observed. The global product presented in this paper is intended to be released to the scientific community for free download and usage.

## 1 Introduction

Forests cover about 30% of the Earth's landmasses, represent the dominant terrestrial ecosystem, and are of extreme importance for all living species. Indeed, they play a key-role in climate change dynamics, by continuously absorbing, storing, and converting atmospheric carbon dioxide ($CO_2$) into free oxygen and terrestrially bound carbon, a process which helps for the reduction of atmospheric greenhouse gases concentration. In addition, plants and trees in forested areas mitigate soil erosion, catch rainwater, and are natural watersheds preventing from flood events. Last but not least, forests represent an essential source of energy (such as biomass), food, jobs, and livelihoods in general for many populations on Earth, and serve as natural habitat to a large variety of animal species, preserving biodiversity and healthy ecosystems. However, the existence of forests is nowadays in danger due to alarming degradation and deforestation rates, which have been severely accelerated in the mid-twentieth century, leading to a permanent loss of plants and animal habitats, a reduction in forest carbon stocks, and to an accelerated soil erosion.

For all these reasons, an up-to-date assessment and monitoring of forest resources becomes of crucial importance and, in this scenario, spaceborne remote sensing represents a unique instrument for providing consistent, timely, and high-resolution data at a global scale. Among a number of global forest classification products, which have been released in the last decades (mainly from optical remote sensing systems operating in the visible and near-infrared frequency range), it is worth highlighting the global forest tree cover map produced in 2013 from mosaics of Landsat sensor data at a spatial resolution of 30 m [1], and the global forest/non-forest classification map provided by the L-band SAR sensor ALOS PALSAR at a posting of 25 m [2].

In this paper we present the first global forest/non-forest classification map derived from TanDEM-X InSAR data at X band [3]. The paper is organized as follows: an overview of the proposed method for the detection of forested areas, based on a multi-clustering fuzzy approach applied on the coherence-derived volume correlation factor, is recalled in Section 2. Section 3 focuses on the mosaicking of multiple available observations and on the identification of additional information layers (such as urban settlements or water bodies), derived from TanDEM-X data as well as from external data sources. The resulting global forest/non-forest classification map at 50 m $\times$ 50 m spatial resolution is then presented. The product validation with external reference data and further comparisons with existing land cover maps is discussed in Section 4. Further applications, such as the generation of high-resolution (12 m $\times$ 12 m) forest/non-forest maps, and the opportunity of exploiting such products for deforestation monitoring, are presented in Section 5, demonstrating the unique potentials offered by the TanDEM-X bistatic system for a broad range of commercial and scientific applications. The paper is concluded in Section 6.

## 2 Multi-Clustering Classification

For the generation of the global forest/non-forest map from TanDEM-X interferometric data, a classification

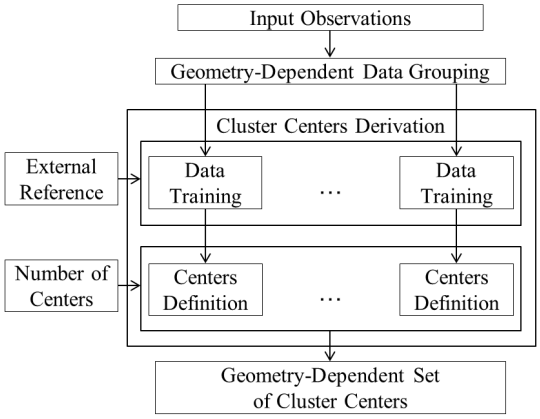


**Figure 1:** Flowchart of the developed method for the determination of a complete set of cluster centers.

method based on a fuzzy clustering algorithm has been developed. For each input scene, it is applied to the volume correlation factor and the algorithm settings are adapted to the specific acquisition geometry. The proposed method consists of different steps, which are summarized in the flowchart in Fig. 1. The output is a complete set of cluster centers, which takes into account the dependency of the volume correlation factor $\gamma_{\mathrm{Vol}}$ on the acquisition geometry. Hence, the classification procedure assigns a different set of cluster centers to each $\gamma_{\mathrm{Vol}}$ value to be classified, depending on the input height of ambiguity $h_{\mathrm{amb}}$ and local incidence angle $\theta_{\mathrm{loc}}$. This is the reason why we refer to a *multi-clustering* classification algorithm for the generation of forest/non-forest maps.

In a general sense, clustering indicates the task of 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 input observations are divided into $c$ nonempty subsets, called clusters. In the present scenario, $c = 2$ (i.e. the *forest* and *non forest* classes), whereas the volume correlation factor is the only feature exploited for classification ($P = 1$). Then, fuzzy-clustering has been introduced in order to allow a certain amount of overlap between different clusters. A well-established algorithm for classification purposes is the $c$-means fuzzy clustering algorithm [4], which introduced the concept of *membership*: according to it, for each input observation a membership function $\hat{\mathbf{U}}$ is defined, which is basically determined by the Euclidean distances between the $k$-th observation and the set of $c$ cluster centers, and describes the probability of an observation to belong to each cluster ($\hat{\mathbf{U}} = [\hat{u}_{ik}] \in [0,1]$, $i = 1, \ldots, c$). Starting from it, we then introduce the concept of *weighted membership* $\hat{\mathbf{U}}_{\mathrm{w}}$, for which the membership is opportunely scaled by the *a priori* probability of a given observation ($y_k$) belonging to the $i$-th cluster. Such a likelihood information is estimated from the distribution of the input observations used for the data training process (see next section). The results are finally $c$ fuzzy partitions of the input observation dataset, each of them containing observations characterized by a high intracluster similarity and a low

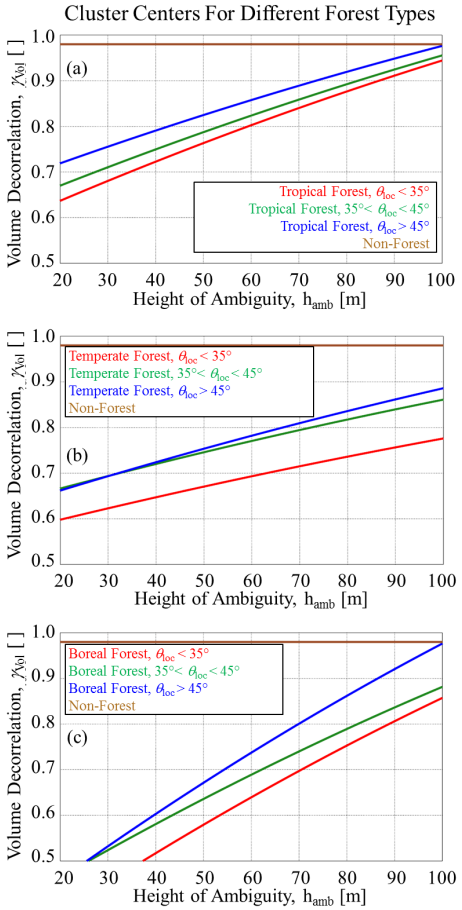


**Figure 2:** Cluster centers ($v_{\mathrm{F}}$, $v_{\mathrm{NF}}$) of the volume correlation factor $\gamma_{\mathrm{Vol}}$ for tropical (a), temperate (b), and boreal (c) forests, respectively, as a function of the height of ambiguity $h_{\mathrm{amb}}$, for near, mid, and far range (depicted with different colors).

extracluster one. As discussed above, the volume correlation factor $\gamma_{\mathrm{Vol}}$ represents the only input feature used for classification purposes. Indeed, the $\gamma_{\mathrm{Vol}}$ term must be estimated from the interferometric coherence by compensating for all decorrelation contributions, as

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

The impact and the evaluation procedure of each decorrelation contribution from TanDEM-X data is discussed in detail in [5]. Since the specific acquisition configuration (namely the incidence angle and the interferometric baseline) directly affects the amount of coherence loss induced by volume scattering, a multi-clustering fuzzy classification approach has been considered. For classification purposes, a training of the input TanDEM-X observables is carried out by means of the Landsat tree cover map [1], and the sample expectations of the corresponding $\gamma_{\mathrm{Vol}}$ distributions for forested and non-forested areas, respectively, are finally taken as corresponding cluster centers $v_{\mathrm{F}}$ and $v_{\mathrm{NF}}$, which are shown in Fig. 2, for tropical, temperate, and boreal forests, respectively. The curves are evaluated as a function of the height of

ambiguity $h_{\mathrm{amb}}$ and for different incidence angle intervals, divided in near ($\theta_{\mathrm{i}} < 35°$), mid ($35° \leq \theta_{\mathrm{i}} < 45°$), and far range ($\theta_{\mathrm{i}} \geq 45°$). Over bare areas, the volume correlation factor does not depend on the particular combination of incidence angles and baselines [6]. For our classification algorithm, a constant value $v_{\mathrm{NF}} = 0.98$ has been selected, represented by the brown horizontal line in the figure (non-forest). An example of the resulting forest/non-forest classification for a TanDEM-X acquisition over temperate forest in Germany is shown in Fig. 3. Areas detected as forest and non-forest are depicted in dark green and in white in the map on the right-hand side, respectively. On the left-hand side, the optical image of the area is given for comparison. The two different land cover types are overall correctly discriminated, and even single lines of trees around agricultural fields are detected (zooms in Fig. 3 (c) to 3 (f)).

# 3 Mosaicking of Overlapping Acquisitions and External Information Layers

Once the weighted membership map has been derived for each bistatic scene, the next step for the generation of a global product consists in properly combining and mosaicking together the large amount of available acquisitions from the TanDEM-X global data set. Typically, at least two global and up to ten regional coverages over mountainous terrain, forests, and sandy desert regions are available; overall, more than 500,000 bistatic scenes have been considered for the current work. For each scene, extending 30 km by 50 km, the weighted membership matrix is used as input data for the mosaicking process.
Moreover, other available information at scene level, such as layover and shadow layers or the local incidence angle and height of ambiguity are used in the mosaicking process, too. In order to further improve the final classification accuracy, additional information layers are applied by exploiting external maps in a final post-processing step, which are shortly summarized in the following:

- *Urban Areas*: identified by means of the *Global Urban Footprint* (GUF), derived from full-resolution TanDEM-X data backscatter information [7]. The GUF is a binary classification map (city/non-city) and is freely available at a resolution of 2.8 arcsec (75 m - 85 m).
- *Water Bodies*: filtered by using the freely available global map of open permanent water bodies, obtained from the Land Cover (LC) project of the Climate Change Initiative (CCI), provided by ESA, at a resolution of 150 m [8].
- *Tree Line*: the tree line corresponds to the virtual altitude limit at which trees can grow. We derived a global tree line map exploiting the digital elevation information provided by TanDEM-X in combination with the CCI-LC map [8]. Furthermore, over mountainous regions, areas affected by geometrical distortions are identified by means of a scene-based *geometry mask*.
- *Deserts*: in order to avoid misclassification over sandy areas (characterized by poor backscatter levels), desert regions are separately filtered out by using the ESA CCI-LC map [8].

The final TanDEM-X global binary forest/non-forest map is shown in Fig. 4. It has a spatial resolution of 50 m $\times$ 50 m and is divided into geocells of 1° by 1° in latitude and longitude, in accordance with the TanDEM-X global DEM format. Forested and non-forested areas are depicted in green and white, respectively. After applying the external layers, urban areas are depicted in black and water bodies in light blue.

# 4 Validation and Comparisons

To validate the global classification product we have used as reference a highly accurate lidar-optic forest map available for the state of Pennsylvania (USA) [10]. Eight geocells have been considered in an area extending from 40°N to 42°N in latitude and from 77°W to 81°W in longitude, and for all validated geocells the obtained accura-


**Figure 3:** (a), (b) GoogleEarth optical image and TanDEM-X forest/non-forest classification map of an area located in Germany. (c), (d) Zoom in of the area within the light blue rectangle in (a) and (b); the area extends by about 3 km $\times$ 2 km. (e), (f) Zoom in of the area within the light blue rectangle in (c) and (d); the area extends by about 1 km $\times$ 1 km.

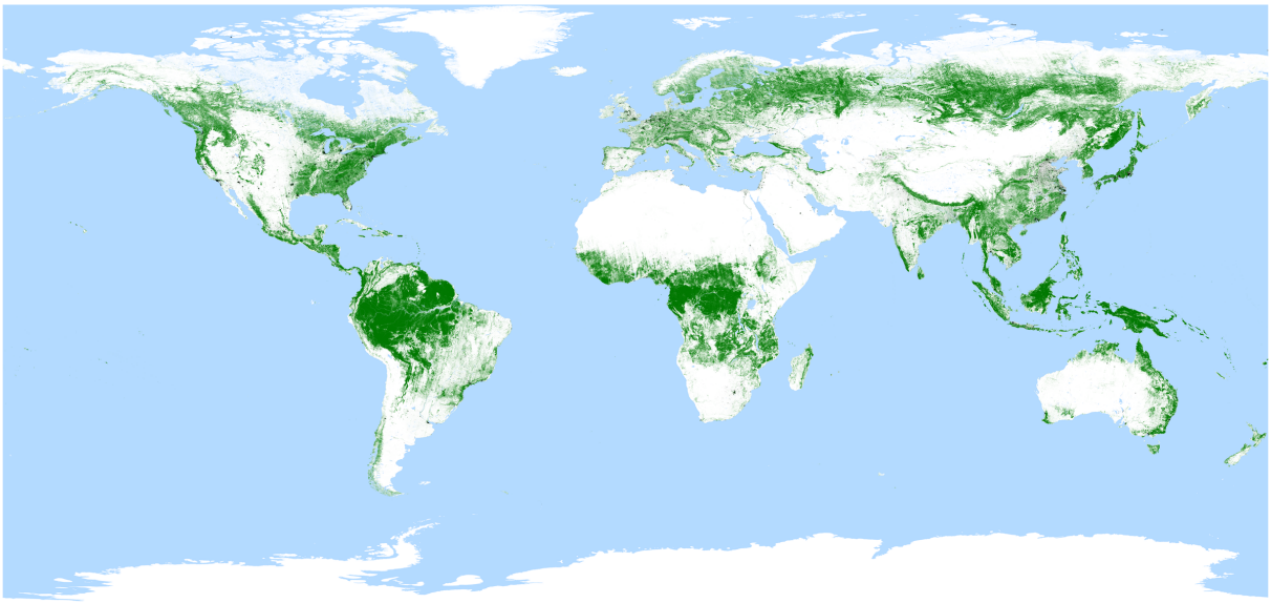

**Figure 4:** Global TanDEM-X forest/non-forest map at 50 m $\times$ 50 m sampling. Forested regions are depicted in green, non-forested areas in white, urban settlements in black, and water bodies in light blue.

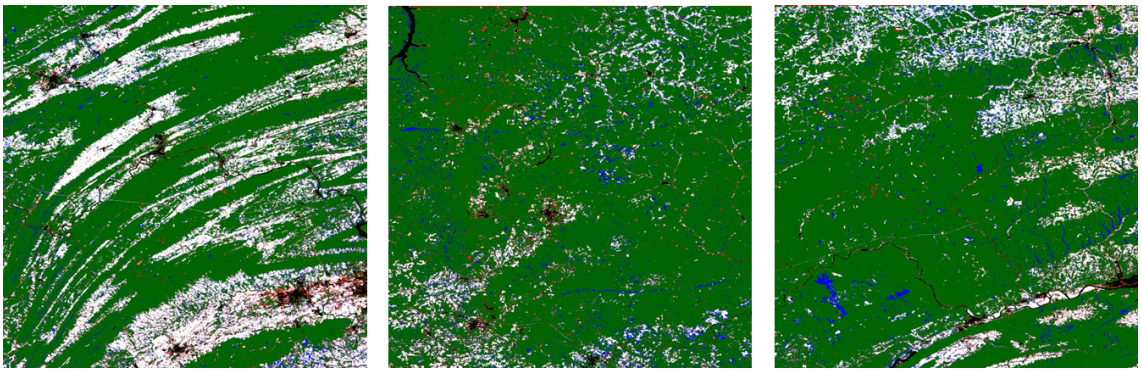

**Figure 5:** Confusion map for three geocells located in Pennsylvania (USA), between the TanDEM-X forest/non-forest map and the lidar-optic binary map used for validation. TP (i.e. both forest) are depicted in green, TN (both non-forest) in white, FP (TanDEM-X only forest) in red, and FN (TanDEM-X only non-forest) in blue. Classification errors are highly clustered: false positives (in red) are mainly located in correspondence of densely to sparse built-up areas, whereas false negatives (in blue) are clustered in correspondence of small water bodies and of areas affected by rugged terrain, such as slopes and valleys. The resulting accuracy is 91%, 92%, and 93%, respectively.

cy is in the range between 85% and 93%. Fig. 5 shows for three geocells the binary comparison between the TanDEM-X forest/non-forest map and the Pennsylvania tree density map, where the true positive (TP), true negative (TN), false positive (FP), and false negative (FN) pixels are depicted with different colors. It can be noticed that the classification errors are highly clustered. In particular, false positives (in red) mainly occur in correspondence of dense to sparse built-up areas, whereas false negatives (in blue) are concentrated over narrow rivers, small lakes and ponds, or rugged terrain, such as slopes and valleys, which cause the occurrence of geometric distortions. Such remaining error sources would be further mitigated by employing an urban and a water mask at higher resolution, as well as by means of a larger stack of acquisitions with different viewing geometry.

We have additionally compared the TanDEM-X forest/non-forest map to other available forest maps, such as the Copernicus High Resolution Layers (HRL) forest density map [11]. Fig. 6 shows the agreement for 145 geocells over Europe. The horizontal light blue line indicates the mean value of agreement, which is of about 86%. By considering mountainous regions only (i.e. the Alps and Southern Germany), a mean agreement of 84% is obtained (red line). Over flat regions (yellow line), the agreement is of about 90%. Moreover, a comparison to the Landsat tree cover map [1] has been carried out for about 70 geocells in the Amazon rainforest, and the resulting agreement between both maps is shown in Fig. 7. The mean agreement of 90% (solid line) is partially affected by forest changes observed between the two maps (mainly due to deforestation activities), which explains the large deviation (dashed lines), and in particular the lower values (around 80%) obtained in few cases.

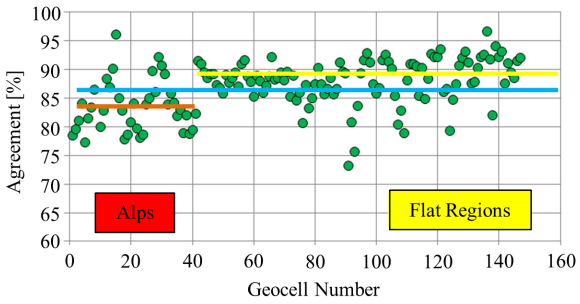


**Figure 6:** Agreement between TanDEM-X and Copernicus HRL forest density for 145 analyzed geocells located in Germany and Eastern Europe. The light blue line shows the overall average agreement, the red and yellow lines show the average agreement for the Alpine and flat regions, respectively.

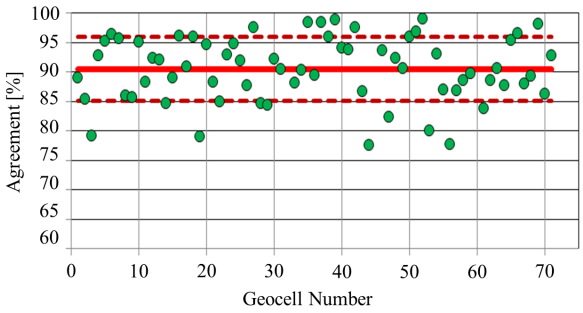


**Figure 7:** Agreement between TanDEM-X and Landsat for about 70 geocells located in the Amazon rainforest, Brazil. The solid and the dashed horizontal lines indicate the mean and the standard deviation of the agreement distribution, respectively.

# 5 Further Applications

High-resolution forest/non-forest maps from TanDEM-X data can be exploited to further improve the classification detail and accuracy, opening new opportunities for the monitoring of forested areas. In particular, the high accuracy and reliability offered by the TanDEM-X forest classification map can be exploited for the monitoring of forest changes. This goal is accomplished by using stacks of repeated acquisitions (time series), especially over areas characterized by illegal deforestation activities. As an example, in Fig. 8 (a) the vegetation map provided by Landsat is shown for a region located in the state of Rondonia, in Brazil, extending by about 30 km $\times$ 20 km. This map was generated using data acquired in 2009, i.e. before the launch of TanDEM-X. Fig. 8 (b) depicts the forest map over the same area generated using TanDEM-X data acquired in 2011 and processed at an interferometric resolution of 12 m. First logging activities (the narrow "tracks" in the middle of the scene) are already noticeable. In Fig. 8 (c) the same map is obtained from TanDEM-X data acquired in 2013 (resolution is again 12 m), and the increase of deforested areas is clearly visible. Fig. 8 (d) shows in red the forest losses that occurred between the two TanDEM-X acquisitions (corresponding to an area of about 20 km$^2$), verifying the great potentials of exploiting X-band single pass SAR interferometry for the monitoring of forested areas. Finally, Fig. 8 (e) shows the confusion matrix between the Landsat map of 2009 (Fig. 8 (a)) and the TanDEM-X one of 2013 (Fig. 8 (c)). New deforested areas are indicated in blue, while green and white areas correspond to stable forested and non-forested regions. This example demonstrates the potentials for exploiting multi-sensor data fusion in order to monitor changes in the forest cover occurred over larger time spans.

The use of alternative processing strategies, such as, e.g., the nonlocal filtering method, to further improve the resolution as well as the classification accuracy are currently being investigated [12] and will be objective of future publications.

# 6 Conclusions and Outlook

In this paper we present the first global Forest/Non-Forest Map derived from the TanDEM-X InSAR data set. Given its sensitivity to the presence of vegetation, the volume correlation factor has been used as input for a multi-clustering classification algorithm based on fuzzy logic. For global classification purposes, so-called quicklooks images at a ground pixel spacing of 50 m $\times$ 50 m have been used. The final product has been compared and validated with existing vegetation maps and by means of external land cover classification data. An accuracy/agreement typically around 90% has been obtained for a variety of forest types and terrains. Potentials for high-resolution forest mapping and monitoring have been presented as well. The global TanDEM-X classification mosaics presented in this paper will be released to the scientific community for free download and usage.

# References

[1] M. C. Hansen, P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. O. Justice, and J. R. G. Townshend *High-Resolution Global Maps of 21st-Century Forest Cover Change*, Science, 342 (6160), pp. 850-853, 2013.

[2] 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, 2014.

[3] P. Rizzoli, M. Martone, C. Gonzalez., C. Wecklich, D. Borla Tridon, B. Bräutigam, M. Bachmann, D. Schulze, T. Fritz, M. Huber, B. Wessel, G. Krieger, M. Zink, and A. Moreira, *Gen-*

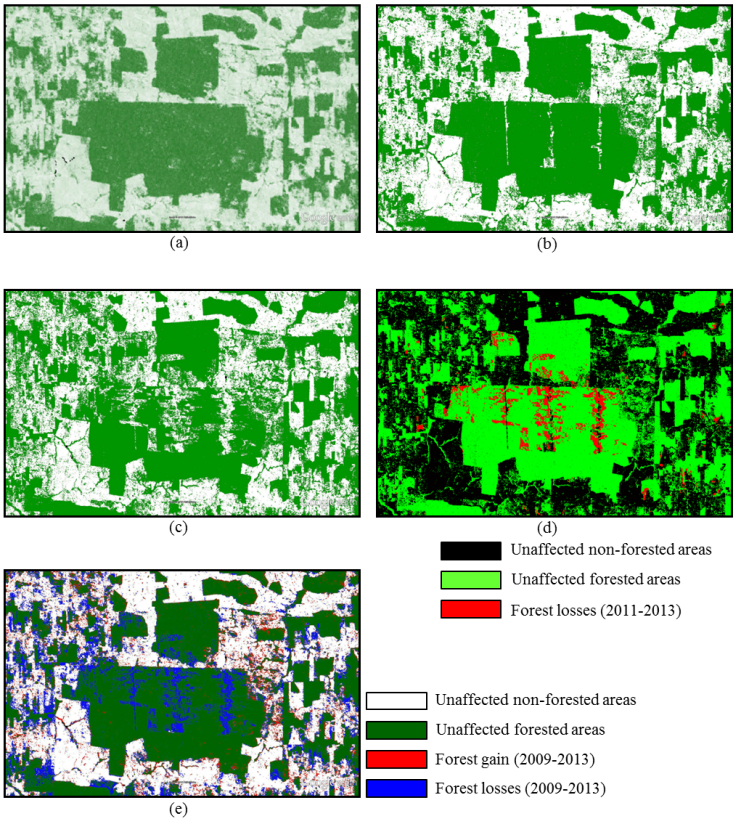


**Figure 8:** (a) Forest/non-forest map over an area in the Rondonia state (Brazil) produced with Landsat data from 2009. The area extends by about 30 km $\times$ 20 km. (b) and (c) Forest/non-forest map over the same area generated from TanDEM-X data acquired in 2011 and 2013, respectively, where the effects of large deforestation activities are clearly visible. (d) Forest losses (in red) occurred between the two TanDEM-X maps (2011-2013). (e) Confusion matrix between the Landsat map (2009) in (a) and the TanDEM-X one (2013) in (c). Areas affected by deforestation are depicted in blue, while green and white regions are classified as forested and non-forested in both maps, respectively. The few red pixels indicate a gain in the forest cover.

*eration and performance assessment of the global TanDEM-X digital elevation model*, ISPRS J. of Photog. and Remote Sens., Vol. 132, pp. 119-139, 2017.

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

[5] M. Martone, P. Rizzoli, B. Bräutigam, and G. Krieger: *The global forest/non-forest map from TanDEM-X interferometric SAR data*, Remote Sens. of Env., Vol. 205, pp. 352-373, 2018.

[6] M. Martone, P. Rizzoli, B. Bräutigam, and G. Krieger: *Forest Classification from TanDEM-X Interferometric Data by means of Multiple Fuzzy Clustering*, EUSAR, Hamburg (Germany), 2016.

[7] T. Esch, H. Taubenböck, A. Roth, W. Heldens, A. Felbier, M. Thiel, M. Schmidt, A. Müller, and S. Dech: *TanDEM-X mission - New perspectives for the inventory and monitoring of global settlement patterns*, Journal of Applied Remote Sensing, Vol. 6 (1), pp. 1-21, 2012.

[8] G. Kirches, M. Santoro, J. Wevers, M. Boettcher, C. Brockmann, C. Lamarche, S. Bontemps, and P. Deofurny: *Land Cover CCI - Product User Guide - Version 2*, ESA Public Document CCI-LC-PUG, Issue 2.4., 2015.

[9] C. Körner: *Climatic tree lines: conventions, global patterns, causes*, Erdkunde, Vol. 61, pp. 316-324, 2007.

[10] J. O'Neil-Dunne, S. MacFaden, A. Royar, M. Reis, R. Dubayah, and A. Swatantran: *An Object-Based Approach to Statewide Land Cover Mapping*, Proceedings of ASPRS 2014 annual conference, Louisville, KY, USA, 2014.

[11] T. Langanke, G. Büttner, H. Dufourmont, D. Iasillo, M. Probeck, M. Rosengren, A. Sousa, P. Strobl, and J. Weichselbaum: *GIO land (GMES/Copernicus initial operations land) High Resolution Layers (HRLs) - summary of product specifications*, 2016.

[12] F. Sica, M. Martone, and P. Rizzoli: *InSAR forest/non-forest classification exploiting non-local pixel similarities*, Proceedings of FRINGE, Helsinki, Finland, 2017.