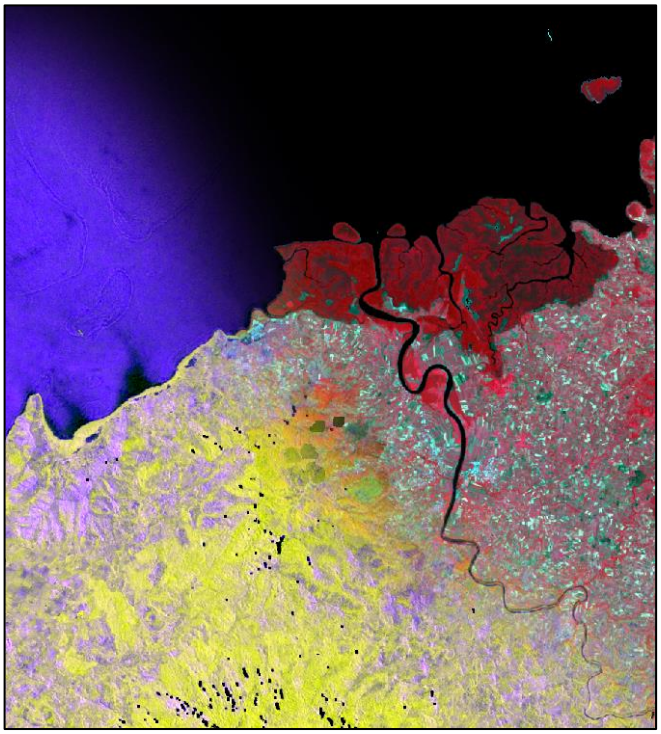
Centre for Geo-Information

Thesis Report GIRS-2014-38

# Integration of Landsat and SAR time series for near real-time deforestation monitoring

Johannes Eberenz



28-10-2014

WAGENINGEN UNIVERSITY
WAGENINGEN UR

II

# Integration of Landsat and SAR time series for near real-time deforestation monitoring

Johannes Eberenz

Registration number 86 11 11 214 080

Supervisors:

Dr. Jan Verbesselt
Johannes Reiche, Msc

A thesis submitted in partial fulfilment of the degree of Master of Science
at Wageningen University and Research Centre,
The Netherlands.

28-10-2014
Wageningen, The Netherlands

Thesis code number: GRS-80436
Thesis Report: GIRS-2014-38
Wageningen University and Research Centre
Laboratory of Geo-Information Science and Remote Sensing

# Abstract

Monitoring forest cover change in near real-time is crucial for timely detection of deforestation. Integration of remote-sensed medium-resolution optical and SAR data can lead to denser time-series in tropical regions with high cloud cover, and thus potentially improve detection speed and accuracy. I developed methods for near real-time deforestation monitoring with integrated multi-temporal Landsat NDVI and ALOS PALSAR L-band HVHH backscatter ratio. Change detection was based on BFAST monitor and data-driven thresholds; fusion was performed at data- and decision level. The methods were validated with 3-monthly reference from an evergreen plantation site on Fiji (2859 ha). Impact of increased cloud cover was studied by testing the methods also with an artificially increased level of missing NDVI data. NDVI – HVHH data fusion at decision-level with threshold-based change detection was found to improve the detection accuracy and speed slightly, when using the original NDVI data (ca. 53% missing data): The overall accuracy with fused data reached 94.4%, compared to 93.8% when using the only SAR data. The detection time-lag could be reduced by 0.2 month at equal accuracy. For 90% missing optical data, the advantages of data-fusion were even smaller. Motivated by short-comings of the common accuracy measures for multi-temporal near real-time change detection, I also propose a statio-temporal change detection accuracy assessment method.


**Keywords:** Multi-sensor data fusion, ALOS PALSAR, Landsat, Time series, Change detection, Near real-time, Deforestation, Sensor interoperability, BFAST monitor, MulTiFuse, Multi-temporal change validation

# Foreword and Acknowledgements

*“You can't connect the dots looking forward; you can only connect them looking backwards. So you have to trust that the dots will somehow connect in your future.” (Jobs 2005)*

Near real-time change detection takes a rather literal approach to these words, in my case with mixed results. Luckily, many people helped me not to lose that trust.

First of all I want to thank my supervisors Johannes Reiche and Jan Verbesselt for guiding my research. They were always open for stimulating discussions and gave me lot of important input without restricting my freedom. Johannes Reiche also contributed the pre-processed RS data. Benjamin Brede and Ruby Rose Bye were kind enough to proofread the manuscript, all spelling mistake in this latest version were added by me after this.

This research would not have been possible without the detailed reference data kindly provided by Fiji Pine Limited. On a more personal level, I am grateful to the staff of the Management and survey division of the Fiji Forestry department, the SOPAC GIS group and GIZ Fiji and the Fiji rucksack Club. They taught me a lot about Fiji and its forest and made me feel welcome in the Pacific during my internship which led to this thesis. Vinaka vaka levu!

I also want to mention the open source community for providing most all of my research tools, only to mention R and its spatial packages, RStudio and gdal.

Finally I want to thank my dear friends and family, who supported me during this period in many ways.

# Index

# List of Tables



# List of Figures



The title-picture shows a detail of a false colour Landsat 7 ETM+ scene (WRS Path 75, Row 72) from 2004 blended into a ALOS PALSAR FBD image (Track 310, Frame 6830) from 2010. The study area is located in the South of the centre of the image.

# Abbreviations

| | |
|---|---|
| ALOS | Advanced Land Observing Satellite |
| BFAST | Breaks For Additive Season and Trend |
| BFM | BFAST monitor |
| CE | Commission Error |
| DETER | Brazilian Real Time System for Detection of Deforestation |
| ETM+ | Landsat Enhanced Thematic Mapper |
| F | forest |
| FORMA | Forest Monitoring for Action |
| HVHH | Ratio of co-polarised and cross-polarised SAR backscatter, in this study specifically time series stacks derived from ALOS PALSAR |
| k | threshold factor for data-driven thresholds |
| LEDAPS | Landsat Ecosystem Disturbance Adaptive Processing System |
| MAD | median absolute derivation |
| MD | missing data |
| MTL | mean time-lag |
| NDVI | normalized difference vegetation index, in this work specifically NDVI time series stack derived from Landsat ETM+ 7 |
| $NDVI_{org}$ | NDVI data with original missing data level (53% on average) |
| $NDVI_{90MD}$ | NDVI data with 90% missing data level |
| NF | non-forest |
| NRT | near real-time |
| n | threshold for number of confirming observations |
| OA | Overall accuracy |
| OE | Omission error |
| PALSAR | Phased Array type L-band Synthetic Aperture Radar |
| RS | remote sensing |
| SAR | Synthetic Aperture Radar |
| SLC | Landsat scan line corrector |
| TM | Landsat Thematic Mapper |

# 1 Introduction

Monitoring forest cover change in near real-time (NRT) is crucial for timely detection of deforestation (Lynch et al. 2013; Xin et al. 2013). These alerts can support law enforcement and thereby contribute to the reduction of illegal deforestation (Assunção, Gandour, and Rocha 2013; Wheeler et al. 2014).

Remote sensing (RS) based time series are the only data streams that can provide repetitive observations of vast forest areas (Lynch et al. 2013). The 'near' real-time aspect of RS-based monitoring depends on the revisiting time of the sensor and the data availability. In order to minimize the change detection lag, every incoming satellite observation should be analysed promptly when it is acquired.

## 1.1 Background

Current operational RS-based NRT monitoring systems for tropical forest are the Brazilian Real Time System for Detection of Deforestation (DETER, http://www.obt.inpe.br/deter/) and the Forest Monitoring for Action (FORMA) system employed by the World Resources Institute Global Forest Watch platform (http://www.globalforestwatch.org). Both systems use coarse optical data (MODIS-like) as RS-input. This renders fortnightly updates. However, small scale changes are missed due to the coarse spatial resolution (Anderson et al. 2005; Hammer, Kraft, and Wheeler 2009; Wheeler et al. 2014).

At medium (Landsat-like) resolution, RS-imagery has successfully been used for detecting smaller scale tropical forest cover changes (De Sy et al. 2012). Landsat imagery is commonly used for regional mapping of forest cover changes (De Sy et al. 2012; Hansen and Loveland 2012; Lu, Li, and Moran 2014) and in operational annual mapping of deforestation (the Brazilian PRODES system, Wheeler et al. 2014). However, missing data due to persistent cloud cover is a limiting factor for monitoring tropical areas (Lehmann et al. 2012; Verbesselt, Zeileis, and Herold 2012; Walker et al. 2010), where cloud cover rates of more than 80% are common (Herold 2009). This reduces the number of valid observations per year and thus often inhibits rapid change detection. Additionally, clouds that are not masked out completely often lead to errors through change commission (Reiche et al. accepted; Zhu and Woodcock 2014a). Using multi-temporal cloud filtering can reduce this problem (Zhu and Woodcock 2014a; Zhu and Woodcock 2014b), but typically increases change detection lag (Reiche et al. accepted).

Data from spaceborne SAR sensors is not impaired by cloud cover and is therefore seen as the ideal complement for optical data in tropical forest monitoring (Herold 2009; Lu, Li, and Moran 2014; De Sy et al. 2012; Walker et al. 2010). Compared to C- and X-band, L-band penetrates deepest into the canopy and is more sensitive to forest cover changes (Luckman et al. 1997). In particular, L-Band backscatter from the ALOS PALSAR sensor has successfully been used for forest cover mapping (Attarchi and Gloaguen 2014; Erasmi and Twele 2009; Kuplich 2006; Vaglio Laurin et al. 2013; Walker et al. 2010; Wijaya and Gloaguen 2009). Together with the follow-up mission ALOS-2 PALSAR-2 it's the only available spaceborne L-Band system. Results indicate the suitability of PALSAR backscatter for operational tropical deforestation monitoring (Almeida-Filho et al. 2009)

with accuracy comparable to Landsat based monitoring (Walker et al. 2010). Before the failure of ALOS in 2011, Brazil planned to launch an ALOS PALSAR based NRT deforestation monitoring system for the Amazon to complement their optical monitoring systems (Butler 2010). ALOS PALSAR was the first SAR mission with a repetitive global observation strategy. However, this comprises only a few observations per location annually (Rosenqvist et al. 2014), falling behind the revisit rate of optical sensors at comparable spatial resolution.

## 1.2 Problem Definition

Recently, a number of time series based approaches for NRT change detection has been developed (see e.g. Hammer, Kraft, and Wheeler 2009; Verbesselt, Zeileis, and Herold 2012; Zhu and Woodcock 2014a; Zhu, Woodcock, and Olofsson 2012). This follows a general trend towards multi-temporal change detection, exploiting the full temporal detail of the RS-based data streams available today (Hansen and Loveland 2012). Some of these approaches are fully data-driven and therefore require no training data or detailed knowledge of the expected change. This makes them more widely applicable (Verbesselt, Zeileis, and Herold 2012; Zhu and Woodcock 2014a). However, all introduced approaches were implemented solely with data from a single optical sensor (Reiche et al. accepted).

Monitoring deforestation in NRT using a single optical or SAR sensor at Landsat-like spatial resolution is limited by the relatively low number of valid observations per year in the available time series. This leads to late detection of changes. Integrating SAR and optical data can generate temporally finer time series and therefore decrease the delay of change detection (Hussain et al. 2013; Lehmann et al. 2012; Reiche et al. accepted; Zhang 2010).

Integration of SAR with Landsat data has been shown to improve tropical land and forest cover mapping (Vaglio Laurin et al. 2013; Walker et al. 2010; Wijaya and Gloaguen 2009; Erasmi and Twele 2009; Kuplich 2006; Attarchi and Gloaguen 2014). However, fusion of SAR and optical RS time series is not trivial, challenges include the need for accurate co-registration, different temporal and spectral variations in the time series and images acquired on different dates (Reiche et al. accepted). The use of SAR – optical integration for multi-temporal forest cover change detection is thus seen as an advanced technique with a research need (Hussain et al. 2013) for which only few approaches exist (see e.g.: Lehmann et al. 2012; Reiche et al. accepted).

SAR – optical data fusion can be performed at data-, feature- or decision-level (Pohl and Van Genderen 1998):

- Data-level refers to integration of measured physical parameters at pixel-level
- Feature-level fusion means integration of regions or objects derived from the individual sources
- In decision-level fusion the different data-sources are processed individually and only the derived information (e.g. Land-Cover changes) is integrated.

While data-level fusion methods dominate in forest classification (see e.g. Attarchi and Gloaguen 2014; Kuplich 2006; Vaglio Laurin et al. 2013), other approaches such as feature- and decision-level fusion have also been applied successfully in change detection (Lehmann et al. 2012; Reiche et al. 2013).

Reiche et al. (accepted) present a pixel-based data-level fusion approach (MulTiFuse) to integrate optical and SAR data for forest cover change detection. The times-series are related and fused using an optimized weighted regression model, followed by unsupervised change detection using the BFAST monitor (BFM) algorithm developed by Verbesselt et al. (2012). Lehmann et al. (2012) derive multi-temporal forest probability maps from ALOS PALSAR and Landsat time series and integrate them at decision-level using a Bayesian framework. However, neither approach is geared towards NRT monitoring: the former is only applicable if observations before and after the change event are available, while the latter is updated only annually and requires the full temporal data for each pixel as training data.

## 1.3 Research Objective and Research Questions

The objective of my thesis is to develop an approach for integrating Landsat NDVI (NDVI) and PALSAR HVHH backscatter ratio (HVHH) time series for NRT deforestation detection.

1. Which pixel-based multi-temporal change detection method is better suited for RS based NRT deforestation monitoring with NDVI and HVHH time series: i) BFM or ii) data-derived thresholds?
2. How can HVHH - NDVI data fusion at A) data-level and B) decision-level be applied for deforestation monitoring in NRT?
3. Can the developed fusion approach improve the spatial and temporal accuracy of NRT deforestation monitoring compared to using single-sensor data?
4. How are the results of monitoring with fused and single-sensor data affected by increased levels of missing data in the optical time series?

# 2 Data

## 2.1 Study Area and Reference Data

The study area comprises the 14059 ha Lololo plantation lease, a managed softwood plantation located in the North West of Viti Levu Island, Fiji (Lat. 17.32° S, Lon. 177.37° E). Evergreen *Pinus caribea* is harvested in cycles of 15 – 20 years; complete stands are logged and then replanted. This clear-cut logging was used here as a proxy for deforestation.

As reference data I used detailed inventory data provided by Fiji Pine Ltd. It covers quarterly harvesting from 2000 to 2013 and planting activity from 1975 onwards. A forest cover map for 01/2008 was derived, covering 2859 ha of stands planted before 01/2000 and not logged until 01/2008. As shown in Figure 1, about 50 percent of this area is logged during the study period from 01/2008 to 09/2010, while the other 50 percent remain standing.

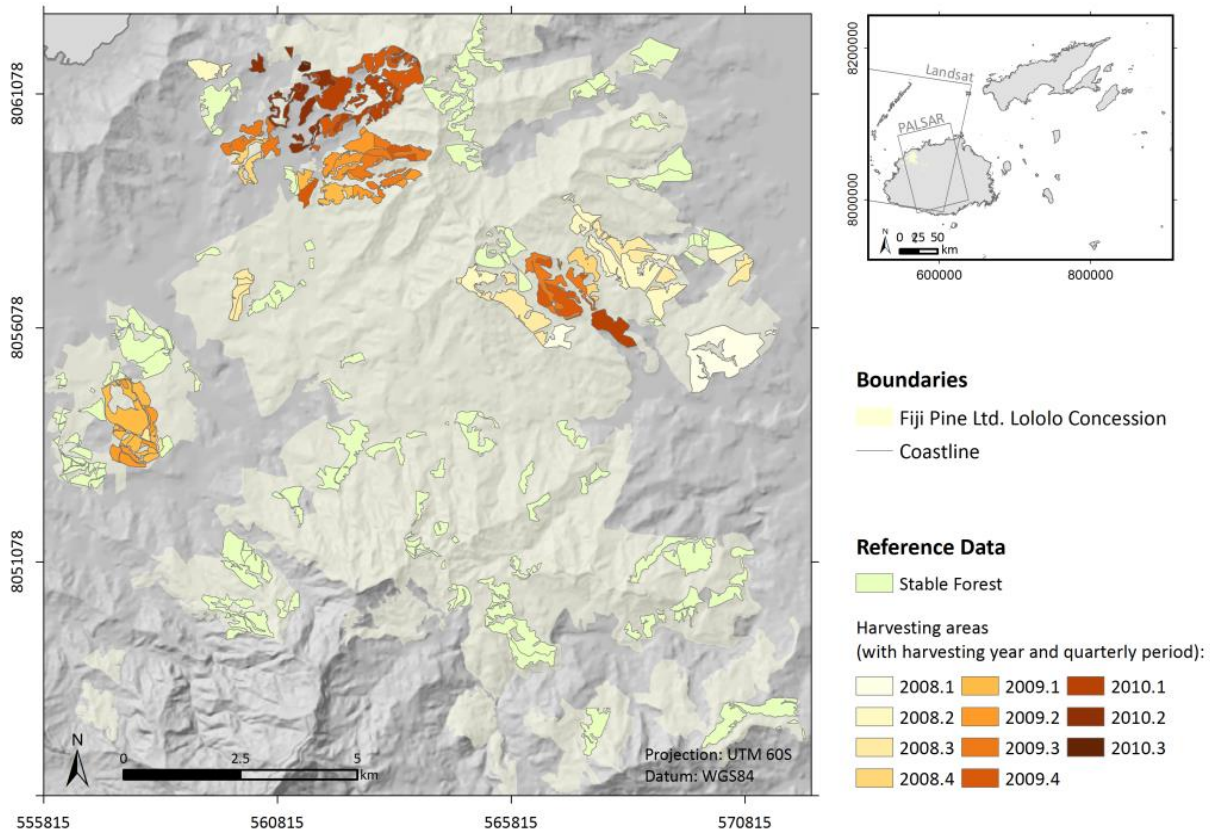



*Figure 1: Overview of study area : Fiji Pine Lololo plantation lease. Reference data for harvested stands (with harvesting year and quarterly period) and stable forest are depicted in different colours.*

## 2.2 Remote Sensed Data

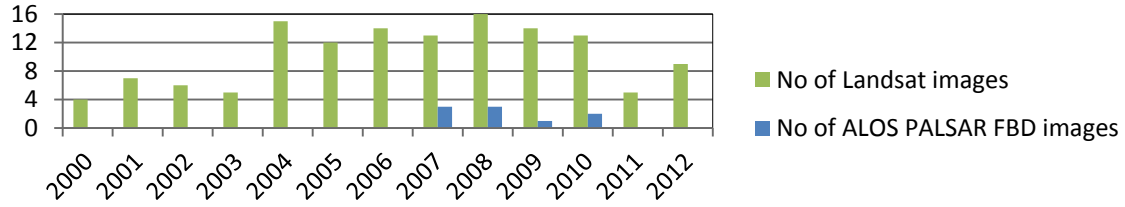



*Figure 2: Number of observations per year for the RS-data: Landsat 7 ETM+ (WRS Path 75, Row 72) and ALOS PALSAR FBD (Track 310, Frame 6830) for 2000-2012*

The RS-data comprises two time series stacks: Landsat NDVI and ALOS PALSAR Fine Beam Double (FBD) HVHH ratio backscatter for the years 2000 to 2012, Figure 2 depicts the images per year for each sensor.

Both time series stacks were provided pre-processed as described in Reiche et al. (under review). For the Landsat data this includes cloud masking with Fmask (Zhu and Woodcock 2012) and atmospheric correction using Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS, Masek, J.G. et al. 2013). The original data $NDVI_{org}$ has on average 53% missing data (MD) per pixel over the study area, caused by cloud masking and the Landsat 7 Scan Line Corrector failure. In order to study the effects of increased cloud cover, additional MD was introduced by random exclusion of observations, leading to data with 90 percent MD per pixel ($NDVI_{90MD}$).

For the ALOS PALSAR data the standard pre-processing (multi-looking, radiometric calibration, topographic normalization and geocoding) was followed by adaptive multi-temporal SAR filtering in order to reduce speckle (Quegan and Yu 2001). The 25m PALSAR pixels were resampled to match the 30 m Landsat pixels, to enable data fusion. Areas with missing data due to layover and shadow effects were excluded from the analysis.

# 3 Methods

## 3.1 Overview

The research comprised the simulation of NRT scenarios (see Section 3.2), pixel based development and testing of NRT change detection (research question 1, see Sections 3.3) and data fusion methods (research question 2, see Sections 3.4). The methods were validated area-wide with the reference data (see section 3.5). I compared validation results of integrated and single sensor monitoring (research question 3) at different levels of missing optical data (research question 4). In an iterative process the validation results were used to further develop the most promising methods (see Figure 2 for a schematic illustration).

All developing steps were conducted in R (R Development Core Team 2013), supported by ArcGIS and Google Earth for visualization.

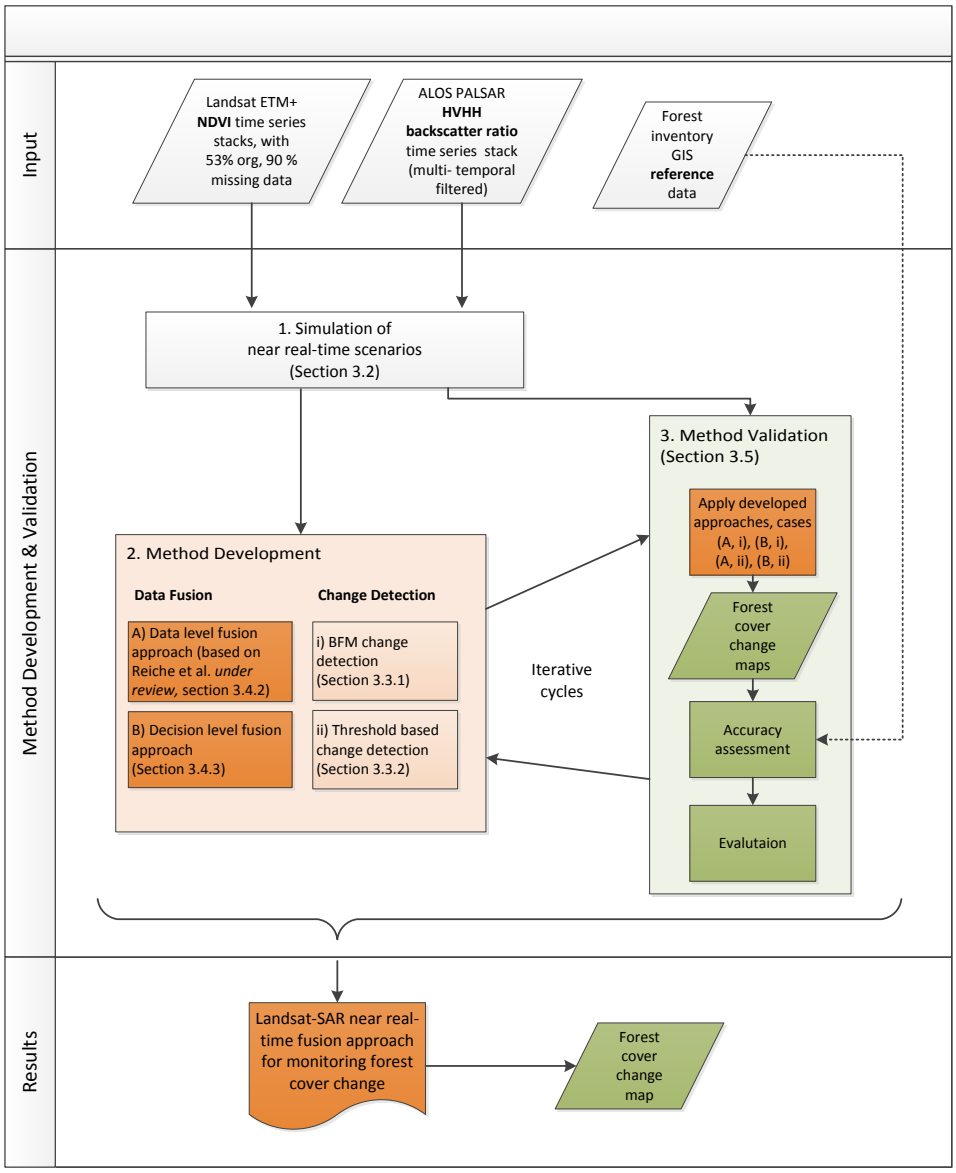



*Figure 2: Schematic representation of research methods. The steps are described in more detail in the indicated sections.*

## 3.2 Near Real-time Scenarios

NRT was simulated by starting with only the RS-data acquired before 01/2008. Proceeding chronologically, the respective next acquired RS-image, regardless if it was NDVI or HVHH, was added to the respective time series stack. With each new image, fusion and change detection was applied per pixel. This was repeated until the end of the study period 09/2010.

## 3.3 Time Series Change Detection

Two multi-temporal change detection methods were explored: i) change detection using BFM and ii) data-driven thresholds. Both take advantage of the large drop in the NDVI and SAR backscatter ratio signal commonly associated with deforestation or clear-cut logging as in this case (Reiche et al., 2013). Appendix Table 1 lists the parameterization of the different change detection algorithms.

| Method | Data driven threshold | BFM |
|---|---|---|
| **NDVI threshold(s)** | Threshold calculated with on eq. (1), k ∈ {1.5, 2, 2,5, 3, 4.5, 4, 4.5, 5} | -0.05 |
| **HVHH threshold(s)** | k ∈ {1.2, 1.4, 1.6, 1.8, 2, 2.2, 2.4} | -1 |
| **Minimal no. historical observations** | {2, 3} | 2 |
| **Maximal no. confirming observations n** | 6 | 6 |
| **Additional parameters** | - | h=0.25 – 1, depending on hist. length;<br>order=1;<br>formula= response ~ 1<br>history= {all observations,<br>the previous 10 observations,<br>all observations after 01-2005} |

*Table 1: Parameter settings for the change detection methods*

### 3.3.1 BFAST Monitor (BFM) Change Detection

BFM is an automated time series change detection method with NRT detection capabilities developed by Verbesselt et al. (2012). It is based on the more generic change detection approach Break For Additive Season and Trend (Verbesselt, Hyndman, Newnham, et al. 2010; Verbesselt, Hyndman, Zeileis, et al. 2010; de Jong et al. 2011). BFM estimates a seasonal-trend model of a time series based on a stable historical period and detects abrupt changes in a subsequent monitoring period for observations differing significantly from the model.

Modelling a seasonal trend was not possible, since the time series are too scarce and the NDVI and HVHH time series show no clear seasonal variation (see Figure 3 for an example from both sensors). The simple model was shown to be suitable for the study area by Reiche et al (accepted). For detected breaks in the time series, BFM provides the time and magnitude of change. To extract clear-cut harvesting, I selected breaks with a negative magnitude below -0.05 for NDVI and below -1 for HVHH, based on values derived in Reiche et al. (accepted).

### 3.3.2 Threshold-Based Change Detection

Data driven thresholds were statistically derived per pixel time series $X_i$ and time point $t_c$ from the historical time series and the stable forest data, using equation (1). A potential change was detected if an observation fell below a sensor-specific threshold relative to the time-series median.

Comparable methods with a more complex time series model have been used for multi-temporal forest cover change detection (see e.g. Zhu and Woodcock 2014a).

Since the data is not normally distributed and has many outliers, I used median and median absolute derivation (MAD, see equation 2) as robust measures of centre and variability. Both measures are more robust towards outliers than the arithmetic mean and the standard derivation respectively and do not require the assumption of normality (Leys et al. 2013).

The variability of stable was derived study-area wide, based on the area-wide observations of stable forest until the start of the monitoring period (historical stable forest). For NDVI, the 25 percent of the pixels with the highest standard derivation were excluded to reduce the effect of remaining clouds on the derived threshold.

$$Threshold_{X_i(t_c)} = median_{X_i(t<t_c)} - \mathrm{k} \times MAD_{hist.stable\ forest} \quad (1)$$

With threshold factor k

$$MAD = \mathrm{b} \times\ Median(\left|X_i(t) - Median\ \left(X_i(t)\right)\right|) \quad (2)$$

Constant *b* was set to 1.4826, resulting in $\mathrm{MAD}\ \approx\ \sigma$ for normally distributed data (see Leys et al. 2013).

Leys et al. (2013) propose k=2.5 as a rule of thumb to detect outliers. In this study, recommendations for the value of k are derived empirically based on the validation results. See Table 1 for the used factors.

## 3.4 Single-sensor Monitoring and Data Fusion

The development of a SAR – optical data integration method covered two general, predefined approaches: A) data-level fusion and B) decision-level fusion. For comparison, monitoring was also performed with both single-sensor time series.

### 3.4.1 Single-Sensor Monitoring

Both change detection methods were applied to the NDVI and HVHH time series stacks separately. Each detected break was verified by applying change detection to the n following observations after removing the preceding observation(s) where breaks were detected. Change was confirmed if all n observations were also detected as breaks in the modified time series. In this way false detection caused by remaining clouds and other artefacts could be reduced, based on the same principle as in recently proposed multi-temporal cloud masking approaches (Zhu and Woodcock 2014b). However, with the numbers of confirming breaks n also the time-lag of change confirmation inevitably increases.

### 3.4.2 Data-Level Fusion

For data-level fusion the MulTiFuse method by Reiche et al. (accepted) was adapted for NRT monitoring:

1. For each incoming image, MulTiFuse was applied per pixel to determine the correlation of SAR backscatter and NDVI time series. When a statistically significant correlation was found, the

time series were fused using the MulTiFuse method resulting in a single time series for the overlapping period.
2. Change detection was applied per pixel on the fused time series if available. If not available, the single-sensor time series with the last available observation was used for change detection. Each detected break was flagged as potential change event.
3. Detected breaks were confirmed with the subsequent n following observations, as described above for the single-sensor monitoring.
4. The results of the method validation were used to empirically evaluate the performance of the approach with different threshold parameter n.

The advantage of data-level fusion is that no decision rules are needed, if the data is fused successfully at data level.

### 3.4.3 Decision-level Fusion

For decision-level fusion a novel method was developed, based on integration of the change detection results from the separate sensors:

1. For each incoming image, change detection was applied per pixel on the respective SAR backscatter ratio or NDVI time series. For threshold-based change detection, different thresholds k were used depending on the sensor.
2. To confirm detected changes, the next incoming observations *of either time series* were used. This results in potential change events with a list of properties:
    1) The detecting sensor(s).
    2) The number and sensor type(s) of confirming observations (see above).
    3) The time between the first detection and the confirming observation(s) in the different time series.
3. In order to fuse the results of the change detection, I developed decision tree classifiers based on the first two properties above:

    Decision tree 1): Changes were confirmed if the number of confirming observations exceeded threshold n.

    Decision tree 2): As decision tree 1) but with different thresholds $n_{TOTAL}$ for the total number of confirming observations and $n_{HVHH}$ in the HVHH time-series. Detected changes events were confirmed when either of the thresholds was reached.

    With these classifiers the detection results from both time series were integrated, resulting in single fused change detection per pixel.
4. The different decision trees with different combinations of parameters $k_{NDVI}$, $k_{HVHH}$, $n_{TOTAL}$, and $n_{HVHH}$ were empirically evaluated with the results of the method validation.

## 3.5 Method Validation

Since only pixels with available reference data were used for the method development (wall-to-wall reference data), the validation became a map-comparison and no uncertainty was introduced by sampling (Stehman 2009). To asses spatial accuracy, classical accuracy measures were calculated: the confusion matrix, overall accuracy (OA), change omission error (OE) and commission errors (CE) (Congalton 1991; Foody 2002; Olofsson et al. 2014). With 50 percent of the monitored area being stable forest, the OA was influenced about equally by change omission and commission.

Crucial for NRT monitoring is the time-lag of confirming detected changes. This was assessed by calculating the mean time-lag (MTL) between time point of confirming change and reference data change for pixels with correct change detection (see Reiche et al. accepted). For some application it the time-lag of first detection could also be relevant, it is equal to the calculated time-lag of confirmation with n=1.

Pixels for which a change was detected before or long after the reference data change were classified as change commission error[1]. This corresponds to a NRT monitoring approach where changed pixels are labelled as "non-forest" and not further monitored. For these pixels the detection of the real change is not possible anymore. To take into account the sparse time series and minor errors in the reference data, changes detected within a tolerance window of 6 month before and 24 month after the referenced change time were counted as correct. Since the early-detected pixels are correctly labelled "non-forest" at the end of the monitoring period, one could also classify them as correctly detected. This gives a better comparison to bi-temporal change detection which never takes into account the time-point of change, but neglects the temporal errors of over-sensitive change detection. Since the temporal aspects are crucial in NRT monitoring, this second validation approach was only calculated to show the effect of both approaches.

[1] This means that OA may be influenced by CE stronger than by OE, since more pixels are treated as if their reference status would be stable forest.

# 4 Results & Discussion

## 4.1 Change Detection

### 4.1.1 BFAST Monitor Change Detection

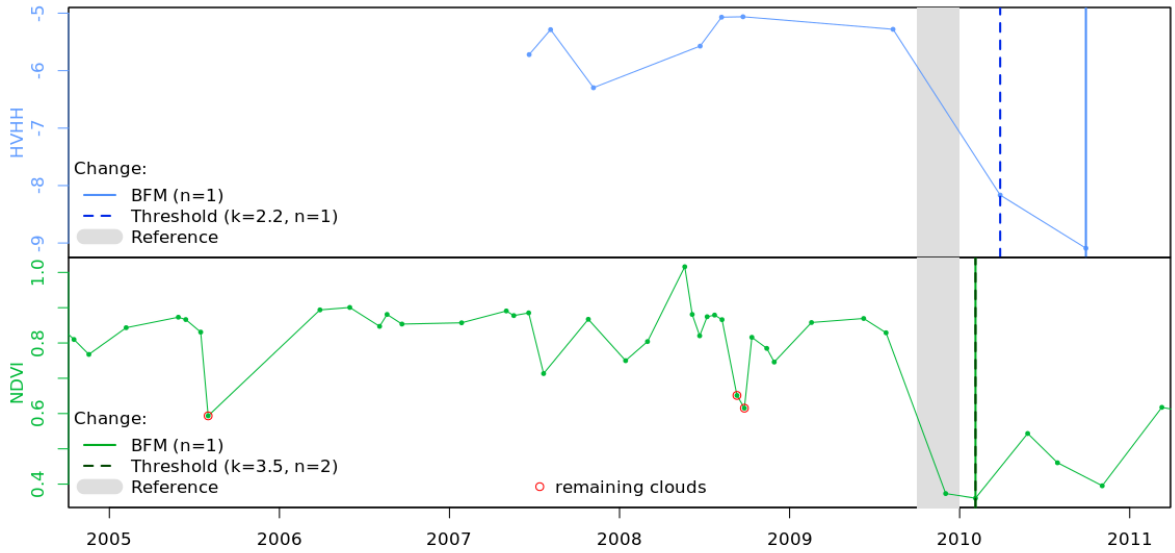



*Figure 3: Example for BFM vs threshold-based change detection , using the single sensor time series (top: HVHH, bottom: $NDVI_{org}$). In both BFM detected the second sig. lower observation as break. Threshold detection with the respective optimal settings (see Section 4.2.1) shows faster detection with HVHH, and the same result as with BFM $NDVI_{org,}$, mind the higher n. Clearly visible are artefacts in the NDVI from remaining clouds (red circles).*

Figure 3 shows an example of BFM and threshold-based change detection using a pixel with a change in 02/2009 and single sensor $NDVI_{org}$ and HVHH time series. As typically for BFM, the first observation with a significantly lower value than the history was not detected as a break in both time-series. This prevents early detection in the NDVI time series with artefacts from remaining clouds (red circles).

Resulting from these characteristics of the algorithm, the validation for BFM shows a relative conservative detection with low change commission and high MTL (see Figures 6, 8 & 13 for validation results with different methods). One confirming break (n = 1) gives the highest OA and smallest MTL for HVHH and $NDVI_{org/90MD}$.

### 4.1.2 Threshold-Based Change Detection

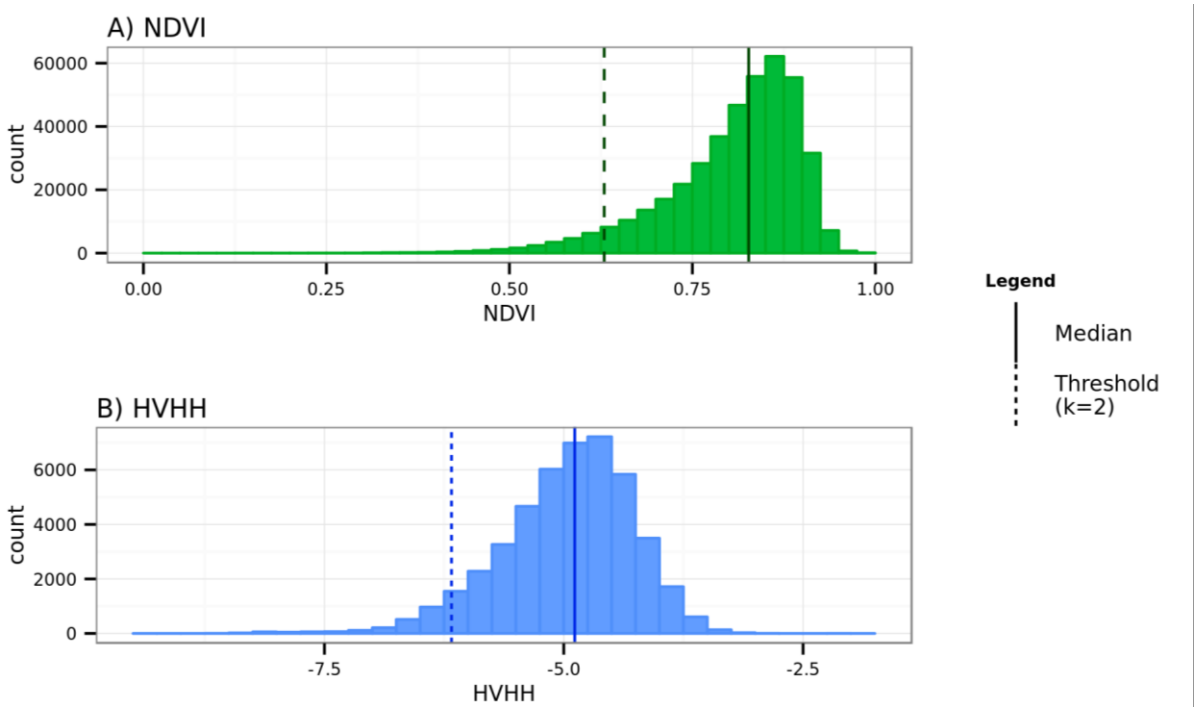



*Figure 4: Distribution of historical stable forest observations for A) NDVI and B) HVHH. The median (solid lines) and an example threshold calculated with k=2 and the area wide historical stable forest data (dashed line) are shown.*

Figure 4 shows the distribution of NDVI and HVHH values of the area-wide historical stable forest areas. Even after filtering for clouds by excluding pixels with high variation, distribution of NDVI values showed negative skew, indicating remaining cloud artefacts.

Thresholds MAD from this data 0.644 for HVHH and 0.086 for NDVI.

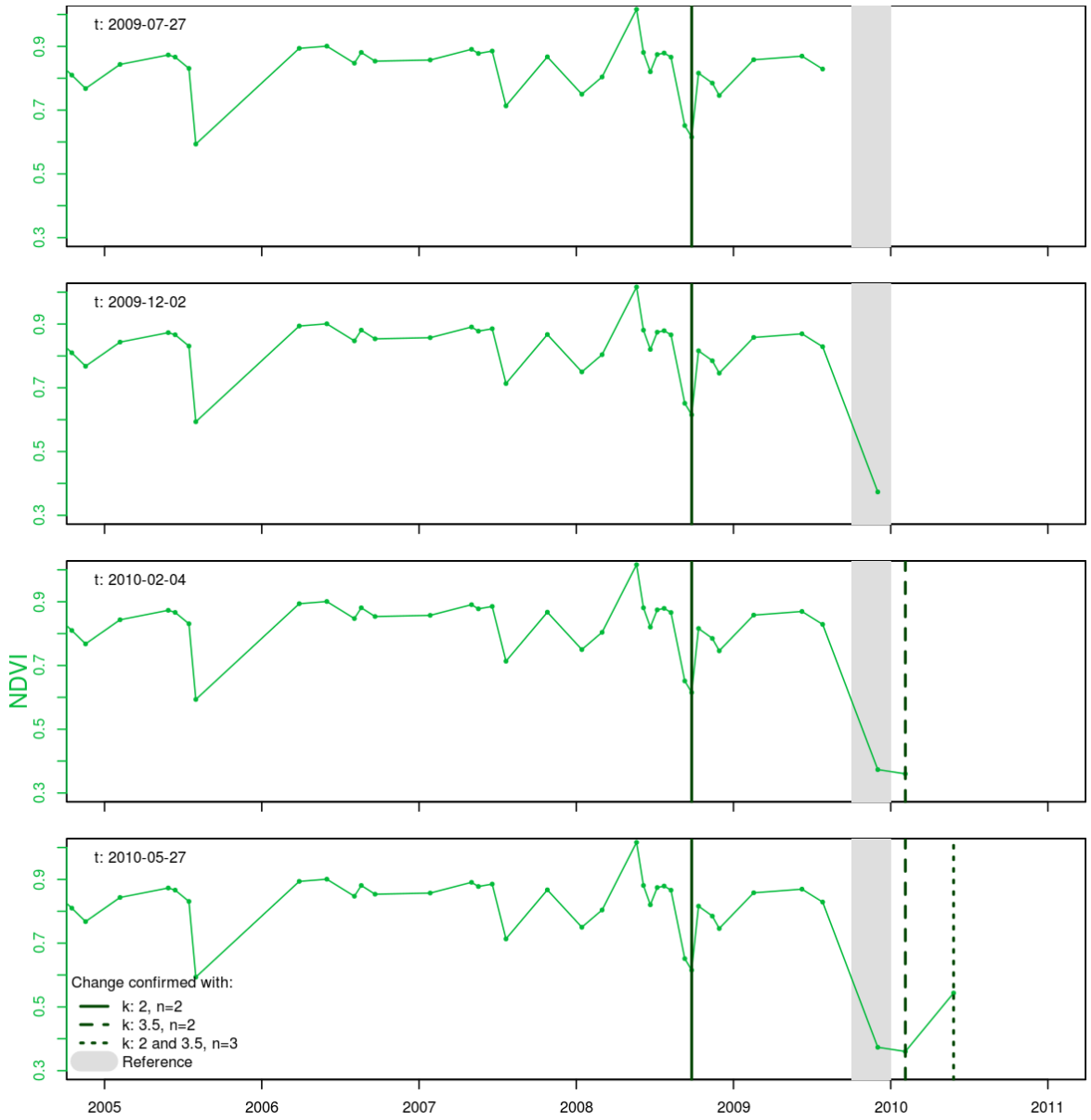



*Figure 5: Stepwise threshold-based monitoring of a $NDVI_{org}$ time series with different values for threshold parameters k and n. Higher thresholds parameters increase the detection lag, but avoid early detection.*

Figure 5 depicts an example of stepwise NRT monitoring of a $NDVI_{org}$ time series. Clearly visible are remaining clouds. Low thresholds k and n lead to too early detection (solid line), medium thresholds to correct detection (dashed line) and high thresholds to slightly late detection (dotted line).

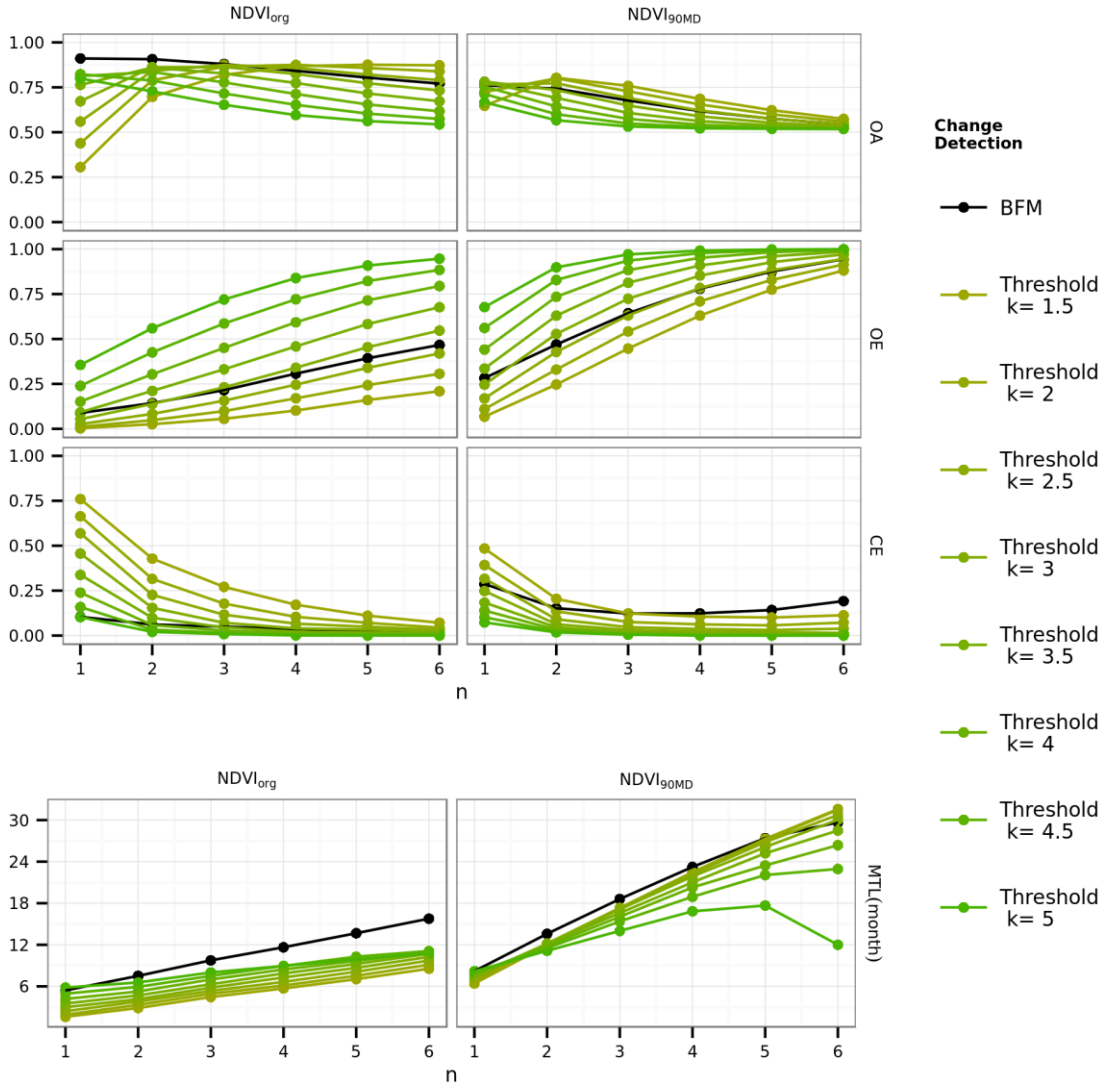



*Figure 6: OA, OE, CE and MTL (month) for $NDVI_{org}$ and $NDVI_{90MD}$ and all tested different change thresholds k and n. Black: BFM change detection results for comparison.*

Figure 6 shows the validation results (OA, CE and OE and MTL) of threshold-based change detection using single-sensor NDVI at all tested thresholds, BFM results are added for comparison. Clearly visible is a trade-off between change CE and OE for both thresholds k and n: For small k and small n, CE is relatively high and OE low, resulting in sensitive detection. For high k and nm CE is relatively low and change omission high, resulting in more conservative detection.

Resulting from this trade-off, OA first increases and then decreases again with higher thresholds k and n. The maximal OA is reached with higher thresholds for $NDVI_{org}$ than for $NDVI_{90MD}$ (see Figure 6, top Panel). As expected, the MTL increases with both n and k (Figure 6, bottom panel). Therefore, there is a trade-off between high OA (with low CE) and fast detection (with low OE) when trying to maximize OA and minimize MTL.

The same trade-offs (between CE and OE and between OA and MTL) also occurred with HVHH data and data fusion (see Figure 8, Figure 13 and Appendix Table 1).

For time-series with few observations ($NDVI_{90MD}$ and HVHH), the described relationships don't hold for higher n, since not enough confirming breaks can be found (see Figure 6, right panel and Appendix Table 1).

### 4.1.3 Discussion of Change Detection

For HVHH and $NDVI_{90MD}$ the results of threshold-based change detection were better than those of BFM; especially MTL is lower (see Figure 13 6, 8n and 13). The more conservative BFM detection is better suited for the outliers in the dense time-series: For $NDVI_{org}$ single-sensor monitoring BFM gives the highest OA, and the overall highest OA is achieved with decision-level fusion using BFM for $NDVI_{org}$ and threshold-based detection for HVHH (see Section 4.2.3). However, in both cases this comes at the cost of a higher MTL.

In the current BFM implementation, the width of the moving window for break detection depends on the number of historical observations (for details see Verbesselt, Zeileis, and Herold 2012). If the monitoring window is too wide, one significantly lower observation in the monitoring window is often not enough to exceed the detection threshold. This explains the late detection. Shortening the history period can help to reduce the window size, however thereby valuable information is discarded and false detection rates increase, leading to a lower spatial accuracy (OA < 0.8 when $NDVI_{org}$ starting only in 2005). If BFM was adjusted to allow flexible monitoring window sizes independent of the number of historical observations, the speed of detection could possibly be improved.

The threshold-based detection on the other hand was relatively fast but sensitive to artefacts in the NDVI data. Best results for $NDVI_{org}$ were achieved with 2-3 confirming observations which is in line results of previous studies that used subsequent observations to confirm detected changes in NDVI time-series (Reiche et al. accepted; Zhu and Woodcock 2014a). Since I use a very simple model (i. e. the median) for the stable forest fewer historical observations are required than with approaches more complex models such as BFM or the Continuous Change Detection presented by Zhu and Woodcock (2014a). However, the simple model may be less suited for environments with strong seasonality or trends, more subtle changes or different harvesting methods or when using sensors less sensitive to forest cover-change, e.g. C-band SAR.

The thresholds are derived from area-wide historical data, differing from BFM and the approache of Zhu and Woodcocks (2014a) which use only the historical data of each pixel. Therefore training areas are necessary to apply the proposed threshold-method method to other sites. The required historical data of stable forest should typically be available in NRT monitoring applications. With more historical observations available than in this study, pixel-based threshold calculation could also be possible. Moreover, the threshold factors k and n requires recalibration for different sites. Contrasting that, in Zhu and Woodcock (2014a) the detection accuracy was reported to be robust to different thresholds for NDVI and n > 2.

Since I focused on NRT application where the detection speed is crucial, I selected change detection based on data-driven thresholds for further development. For the two sensors and the fused data I selected the best thresholds parameters *k* and n using the validation. BFM was used for $NDVI_{org}$ and in combination with threshold-based detection in the decision-level fusion.

## 4.2 Single-sensor Monitoring and Data Fusion

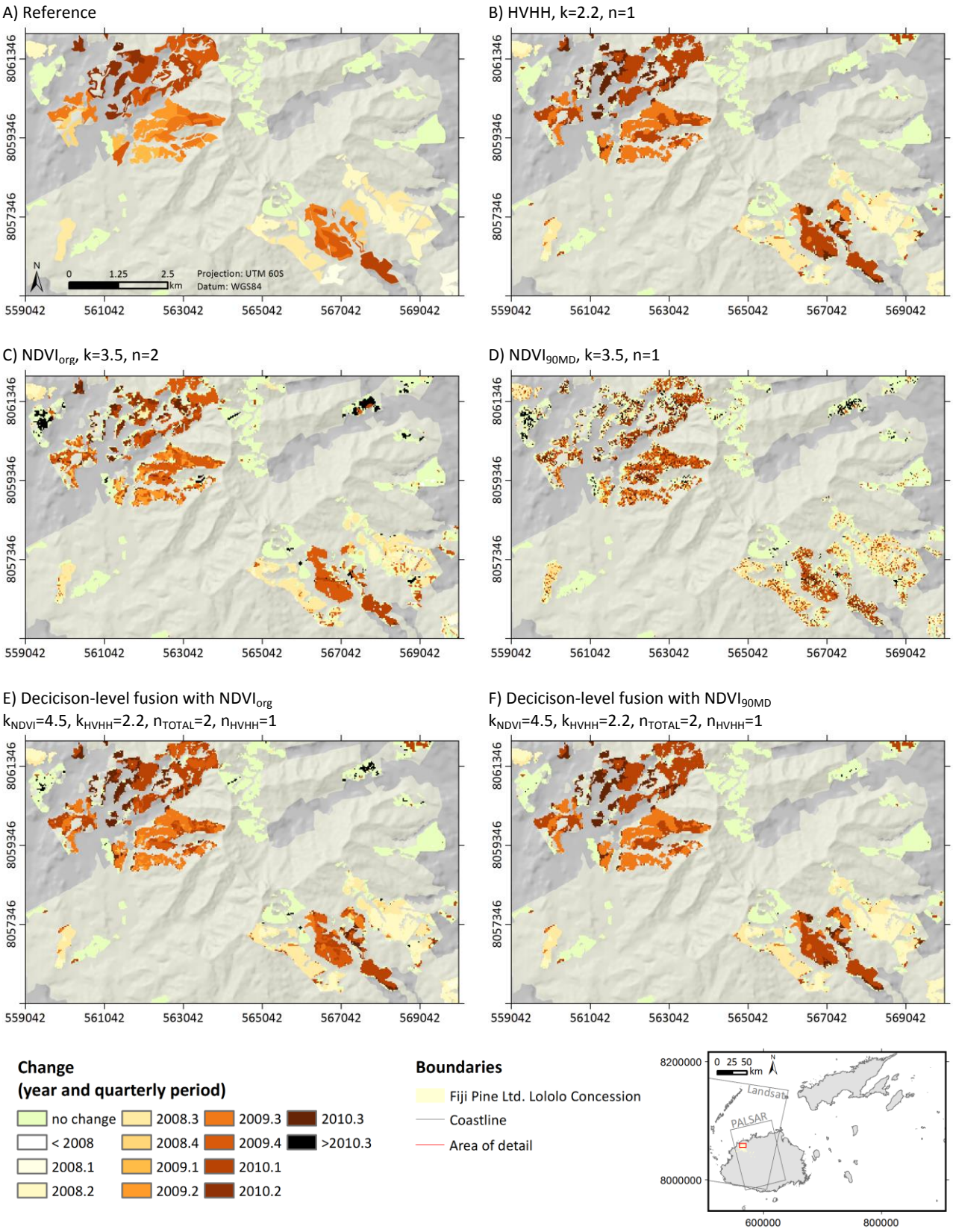



*Figure 7: Result maps (detail from the North of the study area) , showing A) the reference data harvesting periods and change detection time of confirmed changes with threshold-based change detection using B) HVHH, C) NDVI$_{org}$ and D) NDVI$_{90md}$; E) Decision-level fusion with NDVI$_{org}$ and F) NDVI$_{90MD}$. For HVHH, the change dates are quite uniform due to the few observations. NDVI$_{org}$ shows more CE than the HVHH and fusion results, NDVI$_{90MD}$ also has a lot of OE. For dec. level fusion with NDVI$_{org}$ the relatively fast detection is visible, with NDVI$_{90MD}$ results are very similar to the HVHH results*

### 4.2.1 Single-Sensor Monitoring

The NDVI data showed remaining cloud artefacts (see e.g. Figure 5), higher variance in the magnitude of the change than the HVHH time series and quicker recovery after harvest. Therefore, when using only NDVI data, there is a lot of CE (see Figure 7 C) and the trade-off between OA and MTL is strongly expressed as can be seen in Figure 6. When using more conservative detection (high thresholds, BFM), the detection lag and change omission are considerably higher than the results of using only HVHH, even at 60% MD where the NDVI series is denser (see Figure 13 A for comparison of HVHH and NDVI results). Figure 5 illustrates this trade-off with an example, were the sensitive settings lead to early detection and the conservative settings to late detection.

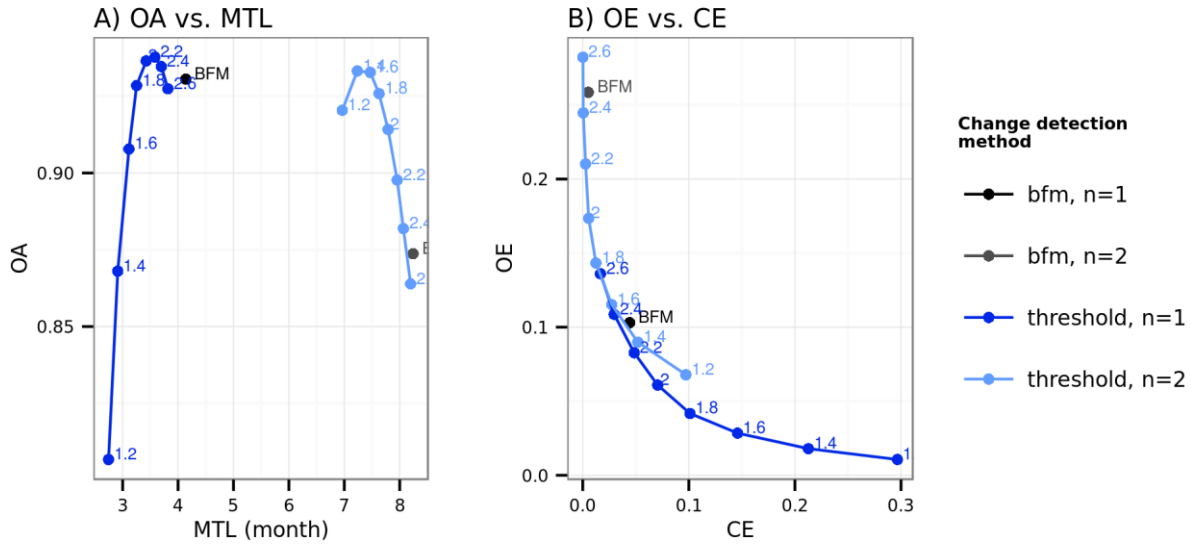



*Figure 8: Trade-off between A) OA and MTL B) OE and CE for HVHH single-sensor monitoring with threshold-based change detection at different k (labels in the plot) and n=1 and 2. The result of BFM change detection are shown for comaprison (black).*

Detection with HVHH was not impaired by remaining clouds, therefore CE is low (see result map in Figure 7 B).

Figure 8 A) depicts the OA and MTL validation results for single-sensor monitoring using HVHH with n=1 and 2, and all tested k values. The trade-off between OA and MTL is clearly visible. MTL increases steeply with higher n since the time series is scarce. The highest OA (0.938) was achieved with threshold-based change detection using n = 1 and k = 2.2.

With higher n, there were insufficient observations for multiple confirmations, which increased the change omission (See Figure 8 B, for n=3 OE =0.49, vs 0.08 for n=1, k=2.2). If only forest-stands harvested in 2008 are used, this effect is strongly reduced (OE of 0.17 for n=3).

### 4.2.2 Data-Level Fusion

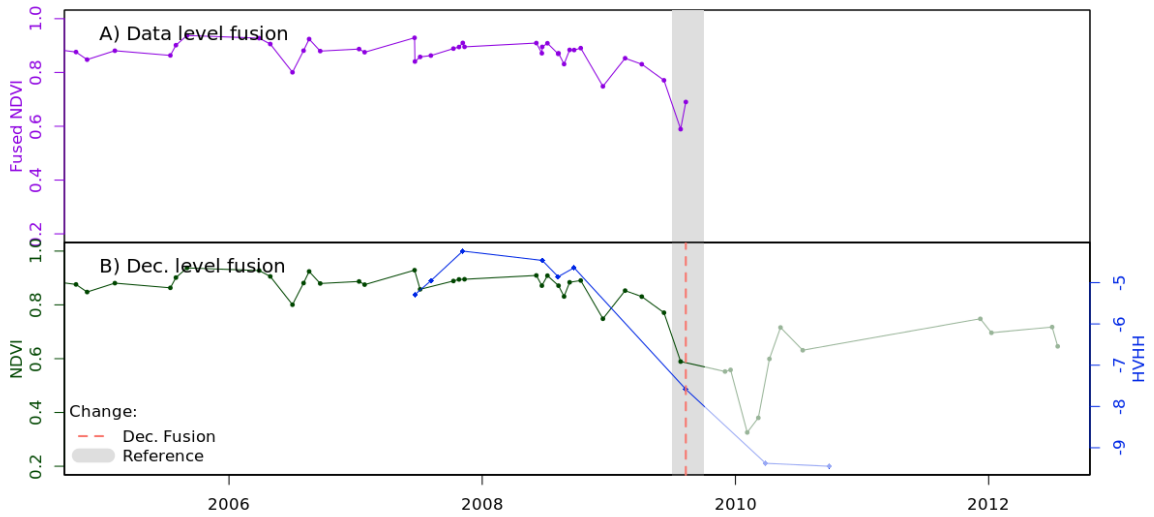



*Figure 9: Data vs decision-level fusion using $NDVI_{org}$. Change is omitted when monitoring the time series fused at dat level (dark red), since the correlation with gradually changing values does not represent the change. (The fused time-series shonw here is the result of data-level fusion with all data avilable until 2009-11-08, the date of the change detection of dec. level fusion (dashed line). HVHH and NDVI correlation p-value < 0.001, $R^2$=0.638.)*

The adaption of data-level fusion with MuTiFuse for NRT monitoring showed three main problems:

Data-level fusion with MultTiFuse doesn't take into account the higher reliability of the SAR data. The same thresholds are used irrespective of the original source for each observation. Therefore artefacts in the NDVI data could increase change commission rates. This could have been adapted; however this would make the method almost equivalent to decision-level fusion.

For more gradual changes, a significant correlation between the time-series was found based on mixed observations between forest and no-forest, resulting in a fused time-series that does not clearly show the drop in the signal. This leads to a late detection compared to using the HVHH time-series. Figure 9 shows an example of this effect, where the fused time-series doesn't represent the drop in HVHH signal. For comparison the detection with decision-level fusion is shown.

Unexpectedly, for about 30% of the pixels significant and high correlation between the HVHH and NDVI time series was found before the referenced change. This is caused by artefacts in short and scarce time series, and leads to similar problems as the cases with gradual change described above.

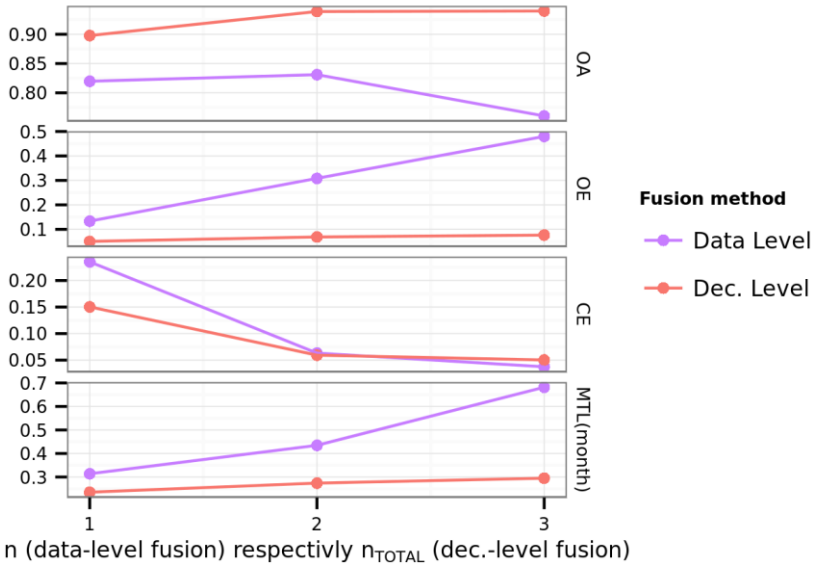



*Figure 10: Validation results for data-level vs decision level fusion with threshold-based change detection with different $n_{TOTAL}$ (NDVIorg/HVHH, k=4.5/22, $n_{HVHH}$=1). Data-level fusion shows higher errors and MTL.*

As a result, data-level fusion using MulTiFuse gave worse results than the decision-level fusion, with larger MTL lower OA, mainly due to higher CE. Figure 10 shows a comparison of the validation results of data-level-fusion and decision-level fusion. CE , OE and MTL increase with n, because of the scarce HVHH time-series. Decision level fusion can avoid this by using different $n_{TOTAL}$ and $n_{HVHH}$ (see the following section).

### 4.2.3 Decision-Level Fusion

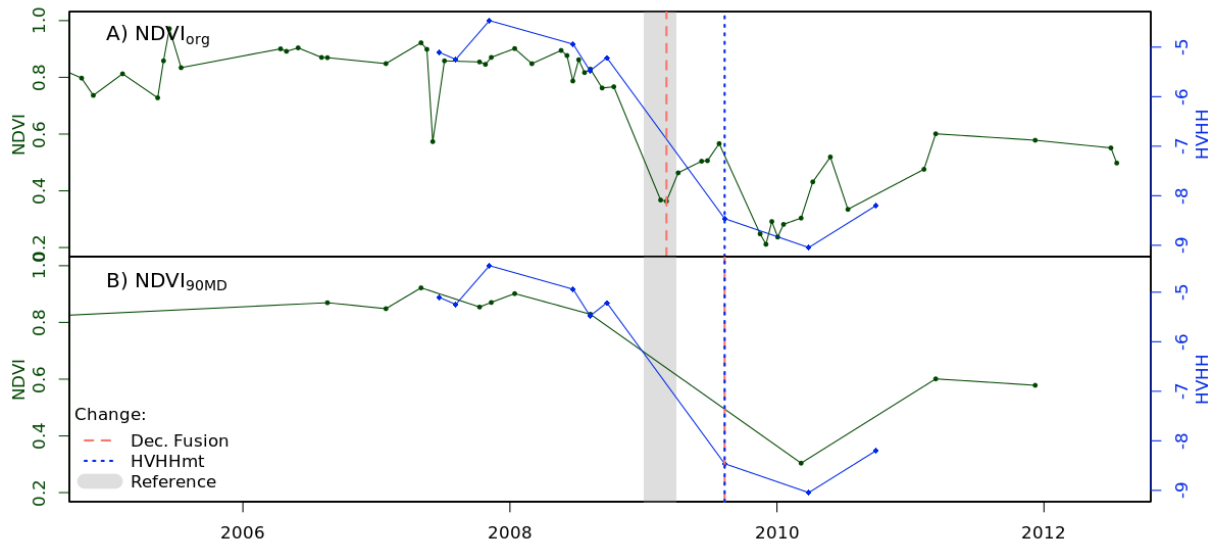



*Figure 11: Example of earlier detection with decision-level fusion.Comparison of decision-level fusion compared to single-sensor HVHH monitoring with A) $NDVI_{org}$ and B)$NDVI_{90MD}$. The change is detected at the second sig. lower NDVI observation in A) and at the first HVHH break in B). (Threshold-based change detection with k=4/2.2, n=2/1 NDVI/HVHH for fusion, and k=2.2, n=1 for single-sensor HVHH. NDVI and HVHH scaled by range for plotting, mind the different axis.*

Figure 7 E) shows a result map for decision-level fusion with threshold-based change detection using $NDVI_{org}$ and $k_{NDVI}$=4.5, $k_{HVHH}$=2.2, $n_{TOTAL}$=2, $n_{HVHH}$=1. Comparison with single sensor monitoring

(Figure 7 B: HVHH and C: $NDVI_{org}$) shows an earlier detection, lower CE than with $NDVI_{org}$ and lower OE than the HVHH results. These effects can be explained by the characteristics of the time series: Figure 11 A) shows an example pixel comparing decision-level to HVHH monitoring (same settings as above), the change is detected through the denser NDVI time-series, which decreased the detection-lag compared to using HVHH only. Figure 12 A) illustrates a different effect which lead to lower OE with decision-level fusion compared to using only HVHH: change was detected through the NDVI, while the HVHH change magnitude was too small for detection. The time-lag is high, since the high HVHH observations prevent change confirmation with 2 NDVI observations.

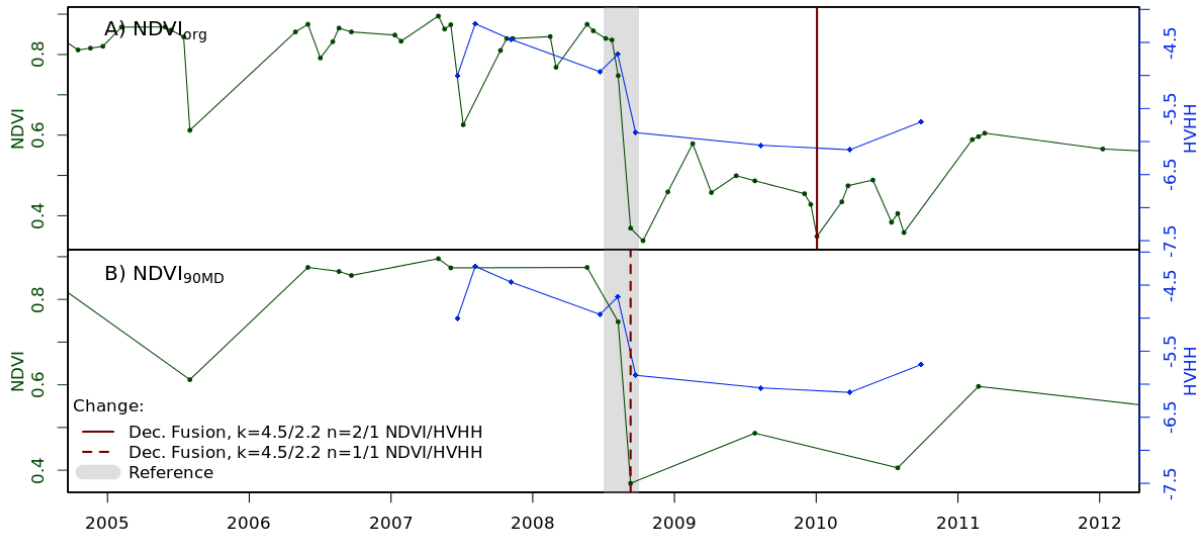



*Figure 12: Example for lower reduced OE with decision-level fusion at different MD levels: A) $NDVI_{org}$ and B) $NDVI_{90MD}$. Change is omitted when using only HVHH data, since the change magnitude is unusually small. (Threshold-based change detection, parameters see legend).*

Both effects are reflected in the validation results: The left panel of Figure 13 A) shows a comparison of OA and MTL of decision-level fusion vs. single-sensor monitoring for selected change detection methods at $NDVI_{org}$. Decision-level fusion gave higher OA and smaller MTL compared to using HVHH only, and much better results than using only $NDVI_{org}$. Again there was a trade-off between a low MTL and high OA for different change detection parameters. The left panel of Figure 13 B) shows the trade-off between OE and CE for $NDVI_{org}$. Decision-level fusion shows a higher CE but lower OE than HVHH single-sensor monitoring.

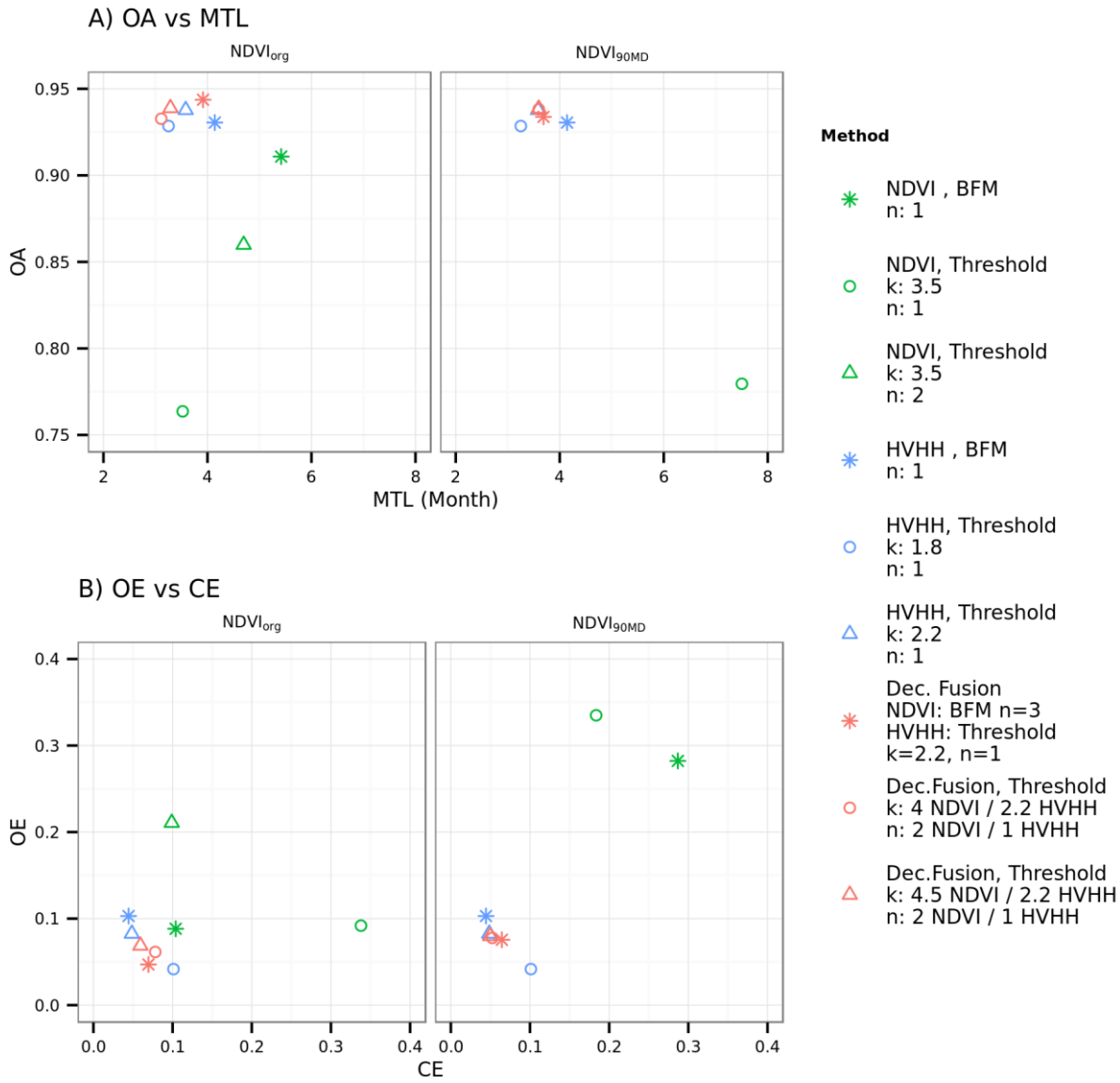



*Figure 13: A) OA vs. MTL and B) OE vs. CE for single-sensor and decision-level fusion, threshold and BFM change detection with selected parameters with the best results at different NDVI MD levels. Decision-level fusion gave the highest OA at lower MTL with $NDVI_{org}$, with $NDVI_{90MD}$ the differences between fusion and using only HVHH are very small. Fusion shows a higher CE but lower OE than the HVHH results at both MD levels. Some $NDVI_{90MD}$ results lie outside the plot limits, which were chosen in order to show more detail.*

The highest overall OA with threshold-based change detection was achieved decision-level fusion with $k_{HVHH}$=2.2, $k_{NDVI}$=4.5 and $n_{HVHH}$=1, $n_{NDVI}$=2 (decision tree 2). Both threshold parameters were higher for the NDVI than for HVHH. This prevented errors from noise in the NDVI to increase CE. Consequently SAR data dominated the detection, 63.8 percent of the changes were detected by one conforming break in the HVHH time series.

Decision-level fusion with BFM for $NDVI_{org}$ (n=3) and threshold-based detection for HVHH (k=2.2, n=1) gave the overall highest OA of 94.4%. The more conservative change detection for NDVI filtered out change commission, however as expected at the cost of a higher time-lag.

When using decision tree 1) with the same threshold n for both time series , the results are slightly worse, since more errors from the NDVI time series are in cooperated in the fused result. With $NDVI_{org}$, the highest OA for this decision tree is 0.922 with an MTL of 3.45 Month, achieved with BFM and n = 1.

### 4.2.4 Effects of Increased Missing Data in the NDVI Time Series

Increased md had a strong effect on change detection when using the optical data only: comparison of result maps C) and D) in Figure 7 reveals a much higher change omission with $NDVI_{90MD}$. The validation shows distinctly lower OA accuracy (0.762 vs. 0.911) and higher MTL (8.2 vs 5.4 Month) at 90% MD compared to 60% MD (using the change detection with the respective highest OA). Figure 6 shows a comparison of OA, OE, CE, and MTL for single-sensor monitoring using NDVI at different MD levels. At 90% MD, threshold detection performs better than BFM, since the sparser time series also contains less remaining clouds. CE stays lower at more sensitive threshold factors.

The main effect of increased MD on decision-level fusion is illustrated by an example in Figure 11 where A) shows earlier detection with two confirming NDVI observations at 60% MD. In Figure 11 B) with 90% MD the change is detected later through one HVHH break, because the thinned out NDVI time series reveals the change even later than the HVHH time-series.

In some cases where change would be omitted with using HVHH only, the fusion gave correct but detection with NDVI observations (see Figure 12 B for an example).

Figure 13 A) shows how both effects were reflected in the validation results. It depicts a comparison of OA and MTL of decision-level fusion and single-sensor monitoring at different MD levels for selected change detection parameters. With $NDVI_{90M}$, the lag can't be reduced compared to HVHH single-sensor monitoring. The maximal achieved OA of decision-level fusion is still higher than with HVHH single-sensor monitoring (0.9384 vs 0.9378), however this difference is hardly significant. The trade-off between OE and CE is visible in in Figure 13 B).

The share of changes detected by one HVHH observation increased to 79.4 percent.

### 4.2.5 Discussion of Sensor Characteristics and Data Fusion

At low levels of missing data, the optical and SAR data have complementary characteristics: $NDVI_{org}$ has a finer temporal resolution but yielded a lower spatial accuracy mostly due to remaining clouds. The HVHH data yielded higher OA, but has fewer observations per year. The higher accuracy achieved with PALSAR HVHH compared to $NDVI_{org}$ confirms findings in Reiche (accepted). However, previous studies using both sensors for forest/non-forest discrimination achieved contrary results (Lehmann et al. 2012; Reiche et al. 2013; Walker et al. 2010). This could be attributed to the use temporal mosaics of multiple Landsat images to reduce missing data, while the NRT approach of this study uses each single image and has to deal with missing data internally, which causes most of the change commission. The same problem is encountered by of Zhu and Woodcock (2014a), they achieve an OA of 92.8% using NDVI with low cloud cover (< 20%) over a non-tropical site, with higher CE than OE, caused partially by remaining clouds .

Because of the remaining clouds, the finer temporal density of the $NDVI_{org}$ time series couldn't yield the expected advantage in detection speed compared to HVHH. At least two observations have to be used for change confirmation to avoid commission of remaining clouds. Therefore, my results

indicate a clear advantage of well filtered PALSAR backscatter-ratio over Landsat NDVI for NRT monitoring of deforestation under cloudy conditions.

Data-level fusion using MulTiFuse showed results comparable to Reiche et al. (accepted), however contrary to their findings, in my results MulTIFuse does not outperform single-sensor monitoring with HVHH. This is partly caused by different validation approaches: MulTiFuse doesn't take into account the different reliability of the time-series, leading to early-detection through remaining clouds. This was counted as change commission in this work, but as correct in Reiche et al. (accepted, see Section 4.3). Moreover, MulTiFuse had inherent problems when applied to NRT monitoring, where forest and non-forest observations of both time series were not always available.

Decision-level fusion did improve OA and detection speed when using $NDVI_{org}$ since the complementing characteristics of optical and SAR data could be exploited. The best performing decision tree takes into account these different characteristics by using higher thresholds parameters $k_{NDVI}$ and $n_{TOTAL}$ for the less reliable NDVI time series. While some additional CE was introduced by using the NDVI, OE could be reduced compared to single-sensor HVHH monitoring. The effects were however smaller than expected, owing to the low reliability of the NDVI data. HVHH– $NDVI_{90MD}$ fusion results were almost identical to those of single-sensor HVHH monitoring. The sparser NDVI time series have no advantage over the HVHH data neither in density nor reliability. Reiche et al. (accepted) achieve better results with fusion even with $NDVI_{90MD}$, again this is largely due to the different validation approaches.

Because of the trade-off between fast detection and low change commission, there is not a single optimal parameter combination. Depending on the user requirements, faster, less accurate detection with low thresholds or more reliable yet slower detection with higher thresholds can be chosen. The results of both can also be combined to yboth early alerts and more reliable change detection. The approximate reliability of either method can be extrapolated from validation results.

With a minimal MTL of around 3.5 month at OA above 90%, monitoring with decision-level fusion and single-sensor HVHH data gave updates much faster than the yearly updates of most previous change detection studies with Landsat-like resolution (Hansen and Loveland 2012) Zhu and Woodcock (2014a) unfortunately don't report the detection lag. Even when using only $NDVI_{org}$ data, MTL is below 6 month. Numbers in Reiche et al. (accepted) are much lower because of their validation approach. My MTL results can't match the fortnightly updates of current operational NRT monitoring systems at coarser spatial resolution (Wheeler et al. 2014). Upcoming medium-resolution SAR systems with finer temporal resolution (ALOSS-2, SAOCOM) should be able to decrease the time-lag significantly.

## 4.3 Validation data and Methods

The validation showed some problems in the reference data: along the borders of some polygons OE was especially high (see e.g. Figure 7 B, in the South-East corner). This indicates a systematic spatial error in the co-registration of either RS- or reference data. Further some pixels in the North-West of the study area show a strong seasonal pattern, indicating a different land-cover than the expected evergreen forest.

Given the relative high OA when using the optimal methods, these problems appear to have low impact. Since the same reference was used for all methods, relative comparison is meaningful in any case. Therefore no further actions such as spatial buffering or comparison to higher resolution data were deemed necessary.

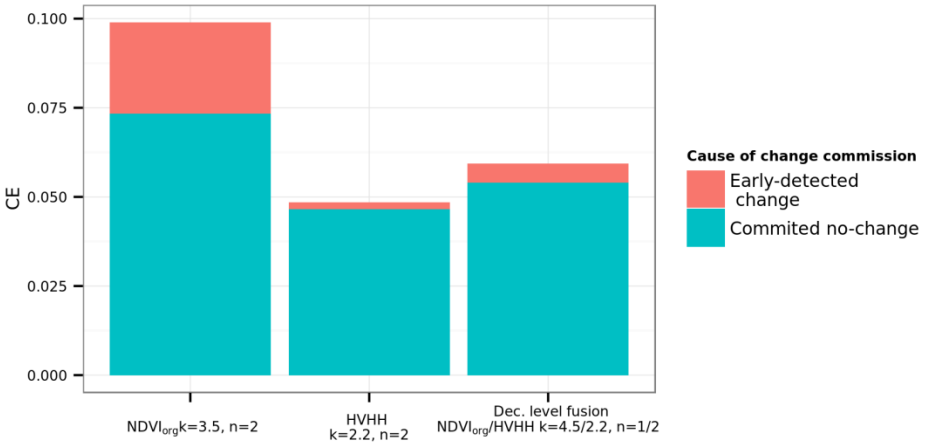



*Figure 14: Contribution of early-detected change to CE for the best performing threshold-based change detection parameters per data source using $NDVI_{org}$.*

Figure 14 shows the contribution of early-detected change to the CE for the different data sources, using the best performing threshold detection parameters. Especially when using NDVI only, early detection accounted for high shares of the commission error. When early-detected change was counted as correct as in Reiche et al. (accepted), OA increased (by 2-3 %) while MTL showed sometimes even negative values, especially for NDVI single-sensor monitoring. Figure 15 shows a a map detail displaying the lag of first detection in month for monitoring with $NDVI_{org}$.

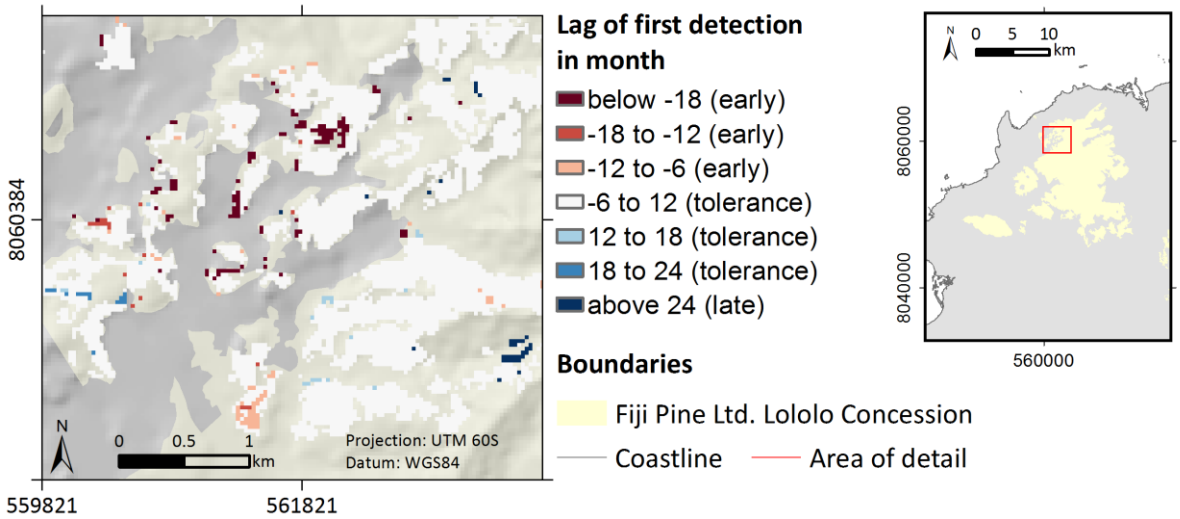



*Figure 15: Map detail of lag of first detection.Negative lags (early detection) are displayed in red colours. Single-sensor $NDVI_{org}$ monitoring, k=3.5, n=2*

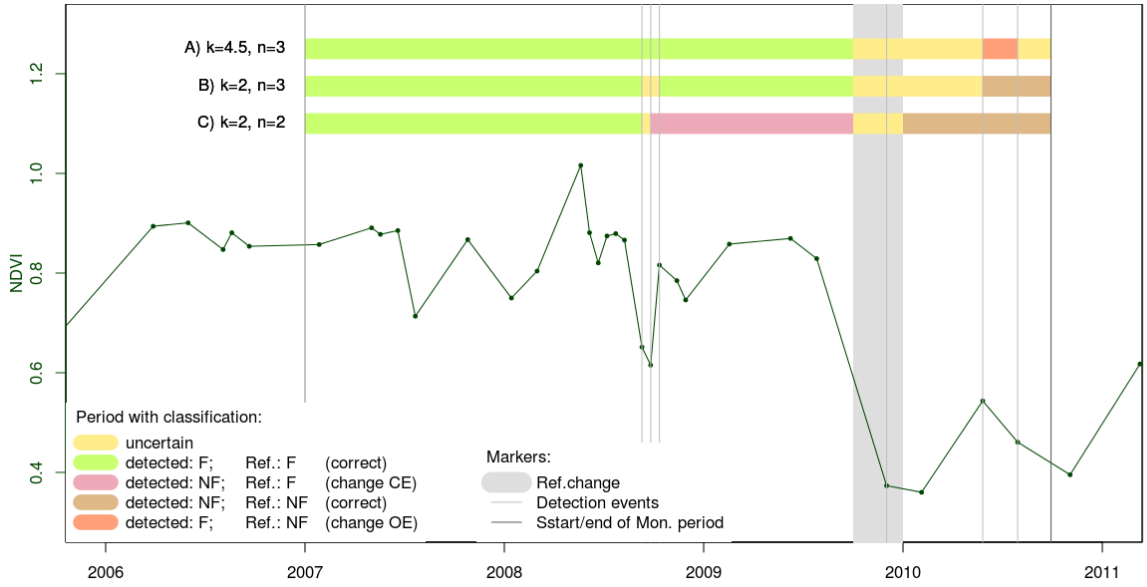



*Figure 16: Example for time based validation of change detection for a $NDVI_{org}$ time series and different parameter combinations A) k=4.5, n=3, B) k=2, n=3, C) k=2, n=2.F: Forest, NF: Non-Forest.*

Figure 16 illustrates how the validation status of a single pixel can change over time, depending on the change detection parameters. This temporal detail of the monitoring results is not represented by either of the validation approaches described above. The sparse literature on sub-yearly change detection doesn't provide more adequate methods, the reported the lag-of detection (Reiche et al. accepted) or percentage of pixels with detection after the reference time (Zhu, Woodcock, and Olofsson 2012) doesn't take into account early detection. This calls for development of specialized spatio-temporal validation measures for multi-temporal and especially NRT change detection.

A straightforward solution would be to treat monitoring results and referenced data as time series stacks of pixel classification states. Then one could derive the spatial accuracy measures for each point in time within the monitoring period, by comparing the corresponding layers of the result- and the reference-stack. The temporal resolution could be chose in accordance with the resolution of the reference data, for the data used herein this is 3 month.

The results per point in time can then be averaged over the monitoring-period, to derive global validation results. The ratio in which change CE and OE influence the global OA depend on the average share of forest-and non-forest over time, rather than on ratio of stable-forest vs. change pixels.

The same global results could also be derived by directly comparing all time-space-points of the result- and reference-stack in a global spatio-temporal confusion matrix. A third way to arrive at the same global results would be to first derive the time-based confusion matrix per pixel and then averaging over space (for a demonstration of all three ways of calculating global accuracy see Appendix II: Calculating Spatio-Temporal Accuracy Measures).

# 5 Conclusions

This thesis presents methods for NRT deforestation monitoring with integrated multi-temporal Landsat NDVI and PALSAR HVHH data. NDVI – HVHH data fusion was found to improve the detection accuracy and speed if the time series have complementing characteristics. This is the case for $NDVI_{org}$ with about 50 percent missing data. However, the improvements compared to using only HVHH data are rather small, and limited mostly by remaining clouds in the NDVI time series. With 90% MD the NDVI time-series become too sparse for data fusion to show an advantage over using only SAR data. Nevertheless, these results confirm previous findings on the potential for SAR-optical data fusion for multi-temporal forest change detection in the tropics (Lehmann et al., 2012; Reiche et al., accepted, 2013). For the first time a NRT monitoring approach is presented that highlights the higher detection speed achievable with data-fusion. Whether the advantages will justify data-fusion in operational monitoring systems depends mainly on the available data and cloud-handling methods.

The overall better outcomes of monitoring with HVHH compared to NDVI emphasize the limitations of cloud-contaminated optical data for NRT monitoring of tropical forest and the big potential of upcoming L-band data-streams (ALOS-2 PALSAR-2, SAOCOM) for this purpose.

Of the tested change detection methods, BFM gave better results with dense and noisy time series ($NDVI_{org}$) while threshold-based detection performed better with scarce time-series ($NDVI_{90MD}$, HVHH). Regarding the relevance of the results, the threshold-based change detection was shown to perform well on the tested site, but may not be suited for sites with higher seasonality or more gradual forest cover changes. The decision-level fusion can also be combined with other change detection methods, even different detection methods per sensor are possible, as demonstrated in this work. Modelling seasonality would require denser time series those used herein.

Decision level fusion performed well at low and high MD, MulTIFuse data-level fusion showed inherent problems with NRT application.

Since there is a trade-off between OA and MTL and OE and CE, the optimal method and parameter combination depends on the user requirements: If the user is interested in a short time-lag and minimal error of omission (*early alert scenario*) a more sensitive approach is suitable, threshold-based detection with relatively lower thresholds. This results in a smaller time-lag and lower overall accuracy. In a *low risk scenario*, where for example ground control is very resource intensive or there are negative consequences of wrong alerts, priority should lie on reducing change commission. Therefore a conservative detection with BFM and higher thresholds should be used. The different settings can however also be used in combination, labelling the detected changes accordingly. The *early-alert* labelled changes could be used as a first warning, while the better confirmed changes could trigger resource intensive responses.

The presented decision-level fusion approach can be used to integrate a variety of RS-based time series with comparable spatial resolution. Whenever the data shows complementing characteristics (e.g. in temporal resolution and reliability), the decision-level fusion is likely to improve the results. Further research should include time series of different sensitivity e.g. C-band SAR (Mitchell et al.

2014), which is becoming increasingly relevant with the upcoming freely-available data from ESA Sentinel-1 featuring global coverage and about fortnightly revisits (Torres et al. 2012).

Furthermore the decision-fusion approach presented here is not restricted to two time series but could be adapted to use an arbitrary number of data sources of comparable spatial resolution. It even allows integrating additional data sources into an operating monitoring system, and could withstand failure of one system. In this way it could set the base for a flexible near real-time monitoring framework that makes use of existing and upcoming medium resolution SAR and optical data streams.

The question of spatially correct but temporally incorrect detections should be reported reveals a short-coming of the commonly used spatial accuracy measures for multi-temporal change detection and especially NRT monitoring. With multi-temporal change detection becoming more relevant, there is a need for more accurate spatio-temporal accuracy measures. I therefore propose an approach that validates each pixel not based on a single change detection result, but by comparing the pixels status at each time-point in the monitoring period to the reference data at the same point of time.

# References


Almeida-Filho, R., Y. E. Shimabukuro, A. Rosenqvist, and G. A. Sánchez. 2009. "Using Dual-polarized ALOS PALSAR Data for Detecting New Fronts of Deforestation in the Brazilian Amazônia." *International Journal of Remote Sensing* 30 (14): 3735–43. doi:10.1080/01431160902777175.

Anderson, L.O., Y.E. Shimabukuro, R.S. DeFries, and D. Morton. 2005. "Assessment of Deforestation in near Real Time over the Brazilian Amazon Using Multitemporal Fraction Images Derived from Terra MODIS." *IEEE Geoscience and Remote Sensing Letters* 2 (3): 315–18. doi:10.1109/LGRS.2005.850364.

Assunção, Juliano, Clarissa Gandour, and Romero Rocha. 2013. *DETERring Deforestation in the Brazilian Amazon: Environmental Monitoring and Law Enforcement « CPI*. Climate Policy Initiative. http://climatepolicyinitiative.org/publication/deterring-deforestation-in-the-brazilian-amazon-environmental-monitoring-and-law-enforcement/.

Attarchi, Sara, and Richard Gloaguen. 2014. "Classifying Complex Mountainous Forests with L-Band SAR and Landsat Data Integration: A Comparison among Different Machine Learning Methods in the Hyrcanian Forest." *Remote Sensing* 6 (5): 3624–47.

Butler, Rhett A. 2010. "Brazil to Launch New Deforestation Monitoring System That 'Sees' through Clouds." *Mongabay*. October 11. http://news.mongabay.com/2010/1011-indicar_brazil.html.

Congalton, Russell G. 1991. "A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data." *Remote Sensing of Environment* 37 (1): 35–46. doi:10.1016/0034-4257(91)90048-B.

De Jong, Rogier, Sytze de Bruin, Allard de Wit, Michael E. Schaepman, and David L. Dent. 2011. "Analysis of Monotonic Greening and Browning Trends from Global NDVI Time-Series." *Remote Sensing of Environment* 115 (2). Elsevier Inc.: 692–702. doi:10.1016/j.rse.2010.10.011.

De Sy, Veronique, Martin Herold, Frédéric Achard, Gregory P Asner, Alex Held, Josef Kellndorfer, and Jan Verbesselt. 2012. "Synergies of Multiple Remote Sensing Data Sources for REDD+ Monitoring." *Current Opinion in Environmental Sustainability* 4 (6): 696–706. doi:10.1016/j.cosust.2012.09.013.

Erasmi, S., and A. Twele. 2009. "Regional Land Cover Mapping in the Humid Tropics Using Combined Optical and SAR Satellite Data—a Case Study from Central Sulawesi, Indonesia." *International Journal of Remote Sensing* 30 (10): 2465–78. doi:10.1080/01431160802552728.

Foody, Giles M. 2002. "Status of Land Cover Classification Accuracy Assessment." *Remote Sensing of Environment* 80 (1): 185–201. doi:10.1016/S0034-4257(01)00295-4.

Hammer, Dan, Robin Kraft, and David Wheeler. 2009. *Forma: Forest Monitoring for Action - Rapid Identification of Pan-Tropical Deforestation Using Moderate-Resolution Remotely Sensed Data*. SSRN Scholarly Paper ID 1517934. Rochester, NY: Social Science Research Network. http://papers.ssrn.com/abstract=1517934.

Hansen, Matthew C., and Thomas R. Loveland. 2012. "A Review of Large Area Monitoring of Land Cover Change Using Landsat Data." *Remote Sensing of Environment* 122 (July): 66–74. doi:10.1016/j.rse.2011.08.024.

Herold, Martin. 2009. *An Assessment of National Forest Monitoring Capabilities in Tropical Non-Annex I Countries: Recommendations for Capacity Building*. Jena: GOFC-GOLD Land Cover Project. http://dup.esrin.esa.int/prjs/Results/131-176-149-30_2009925141824.pdf.

Hussain, Masroor, Dongmei Chen, Angela Cheng, Hui Wei, and David Stanley. 2013. "Change Detection from Remotely Sensed Images: From Pixel-Based to Object-Based Approaches." *ISPRS Journal of Photogrammetry and Remote Sensing* 80 (June): 91–106. doi:10.1016/j.isprsjprs.2013.03.006.

Jobs, Steve. 2005. “Text of Steve Jobs’ Commencement Address (2005).” *Stanford Report*. June 14. http://news.stanford.edu/news/2005/june15/jobs-061505.html.

Kuplich, Tatiana M. 2006. “Classifying Regenerating Forest Stages in Amazônia Using Remotely Sensed Images and a Neural Network.” *Forest Ecology and Management* 234 (1–3): 1–9. doi:10.1016/j.foreco.2006.05.066.

Lehmann, Eric A., Peter A. Caccetta, Zheng-Shu Zhou, Stephen J. McNeill, Xiaoliang Wu, and Anthea L. Mitchell. 2012. “Joint Processing of Landsat and ALOS-PALSAR Data for Forest Mapping and Monitoring.” *Geoscience and Remote Sensing, IEEE Transactions on* 50 (1): 55–67.

Leys, Christophe, Christophe Ley, Olivier Klein, Philippe Bernard, and Laurent Licata. 2013. “Detecting Outliers: Do Not Use Standard Deviation around the Mean, Use Absolute Deviation around the Median.” *Journal of Experimental Social Psychology* 49 (4): 764–66. doi:10.1016/j.jesp.2013.03.013.

Luckman, Adrian, John Baker, Tatiana Mora Kuplich, Corina da Costa Freitas Yanasse, and Alejandro C. Frery. 1997. “A Study of the Relationship between Radar Backscatter and Regenerating Tropical Forest Biomass for Spaceborne SAR Instruments.” *Remote Sensing of Environment* 60 (1): 1–13. doi:10.1016/S0034-4257(96)00121-6.

Lu, Dengsheng, Guiying Li, and Emilio Moran. 2014. “Current Situation and Needs of Change Detection Techniques.” *International Journal of Image and Data Fusion* 5 (1): 13–38. doi:10.1080/19479832.2013.868372.

Lynch, Jim, Mark Maslin, Heiko Balzter, and Martin Sweeting. 2013. “Sustainability: Choose Satellites to Monitor Deforestation.” *Nature* 496 (7445): 293–94. doi:10.1038/496293a.

Masek, J.G., Vermote, E.F., Saleous, N., Wolfe, R., Hall, F.G., Huemmerich , K.F., Gao, F., Kutler , J., and Lim, T.K. 2013. “LEDAPS Calibration, Reflectance, Atmospheric Correction Preprocessing Code, Version 2.” *Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, U.S.A*. doi:10.3334/ORNLDAAC/1146.

Mitchell, Anthea L., Ian Tapley, Anthony K. Milne, Mark L. Williams, Zheng-Shu Zhou, Eric Lehmann, Peter Caccetta, Kim Lowell, and Alex Held. 2014. “C- and L-Band SAR Interoperability: Filling the Gaps in Continuous Forest Cover Mapping in Tasmania.” *Remote Sensing of Environment*. Accessed September 26. doi:10.1016/j.rse.2014.02.020.

Olofsson, Pontus, Giles M. Foody, Martin Herold, Stephen V. Stehman, Curtis E. Woodcock, and Michael A. Wulder. 2014. “Good Practices for Estimating Area and Assessing Accuracy of Land Change.” *Remote Sensing of Environment* 148 (May): 42–57. doi:10.1016/j.rse.2014.02.015.

Pohl, Cle, and J. L. Van Genderen. 1998. “Review Article Multisensor Image Fusion in Remote Sensing: Concepts, Methods and Applications.” *International Journal of Remote Sensing* 19 (5): 823–54.

Quegan, Shaun, and Jiong Jiong Yu. 2001. “Filtering of Multichannel SAR Images.” *Geoscience and Remote Sensing, IEEE Transactions on* 39 (11): 2373–79.

R Development Core Team. 2013. *R: A Language and Environment for Statistical Computing* (version 3.02). R Development Core Team. http://www.R-project.org/.

Reiche, Johannes, Carlos M. Souza Jr, Dirk H. Hoekman, Jan Verbesselt, Haimwant Persaud, and Martin Herold. 2013. “Feature Level Fusion of Multi-Temporal ALOS PALSAR and Landsat Data for Mapping and Monitoring of Tropical Deforestation and Forest Degradation.” *IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING* 6 (5): 2159–73.

Reiche, Johannes, Jan Verbesselt, Dirk Hoekman, and Martin Herold. accepted. “Time-Series Fusion of Landsat and (L-Band) SAR TS Data for Detecting Forest Change in Tropical Environment.” *Remote Sensing of Environment*

Rosenqvist, A., M. Shimada, S. Suzuki, F. Ohgushi, T. Tadono, M. Watanabe, K. Tsuzuku, T. Watanabe, S. Kamijo, and E. Aoki. 2014. “Operational Performance of the ALOS Global Systematic Acquisition Strategy and Observation Plans for ALOS-2 PALSAR-2.” *Remote Sensing of Environment* Corrected Proof. doi:10.1016/j.rse.2014.04.011.

Stehman, Stephen V. 2009. “Sampling Designs for Accuracy Assessment of Land Cover.” *International Journal of Remote Sensing* 30 (20): 5243–72. doi:10.1080/01431160903131000.

Torres, Ramon, Paul Snoeij, Dirk Geudtner, David Bibby, Malcolm Davidson, Evert Attema, Pierre Potin, et al. 2012. “GMES Sentinel-1 Mission.” *Remote Sensing of Environment*, The Sentinel Missions - New Opportunities for Science, 120 (May): 9–24. doi:10.1016/j.rse.2011.05.028.

Vaglio Laurin, Gaia, Veraldo Liesenberg, Qi Chen, Leila Guerriero, Fabio Del Frate, Antonio Bartolini, David Coomes, Beccy Wilebore, Jeremy Lindsell, and Riccardo Valentini. 2013. “Optical and SAR Sensor Synergies for Forest and Land Cover Mapping in a Tropical Site in West Africa.” *International Journal of Applied Earth Observation and Geoinformation* 21 (April): 7–16. doi:10.1016/j.jag.2012.08.002.

Verbesselt, Jan, Rob Hyndman, Glenn Newnham, and Darius Culvenor. 2010. “Detecting Trend and Seasonal Changes in Satellite Image Time Series.” *Remote Sensing of Environment* 114 (1). Elsevier B.V.: 106–15. doi:10.1016/j.rse.2009.08.014.

Verbesselt, Jan, Rob Hyndman, Achim Zeileis, and Darius Culvenor. 2010. “Phenological Change Detection While Accounting for Abrupt and Gradual Trends in Satellite Image Time Series.” *Remote Sensing of Environment* 114 (12). Elsevier B.V.: 2970–80. doi:10.1016/j.rse.2010.08.003.

Verbesselt, Jan, Achim Zeileis, and Martin Herold. 2012. “Near Real-Time Disturbance Detection Using Satellite Image Time Series.” *Remote Sensing of Environment* 123 (August): 98–108. doi:10.1016/j.rse.2012.02.022.

Walker, W.S., C.M. Stickler, J.M. Kellndorfer, K.M. Kirsch, and D.C. Nepstad. 2010. “Large-Area Classification and Mapping of Forest and Land Cover in the Brazilian Amazon: A Comparative Analysis of ALOS/PALSAR and Landsat Data Sources.” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 3 (4): 594–604. doi:10.1109/JSTARS.2010.2076398.

Wheeler, David, Dan Hammer, Robin Kraft, and Aaron Steele. 2014. *Satellite-Based Forest Clearing Detection in the Brazilian Amazon: FORMA, DETER and PRODES*. Full Brief. WRI Working Paper. Washington DC, USA: World Resources Institute. http://www.wri.org/publication/satellite-based-forest-clearing-detection-brazilian-amazon.

Wijaya, A., and R. Gloaguen. 2009. “Fusion of ALOS Palsar and Landsat ETM Data for Land Cover Classification and Biomass Modeling Using Non-Linear Methods.” In *Geoscience and Remote Sensing Symposium,2009 IEEE International,IGARSS 2009*, 3:III – 581 – III – 584. doi:10.1109/IGARSS.2009.5417824.

Xin, Qinchuan, Pontus Olofsson, Zhe Zhu, Bin Tan, and Curtis E. Woodcock. 2013. “Toward near Real-Time Monitoring of Forest Disturbance by Fusion of MODIS and Landsat Data.” *Remote Sensing of Environment* 135 (August): 234–47. doi:10.1016/j.rse.2013.04.002.

Zhang, Jixian. 2010. “Multi-Source Remote Sensing Data Fusion: Status and Trends.” *International Journal of Image and Data Fusion* 1 (1): 5–24. doi:10.1080/19479830903561035.

Zhu, Zhe, and Curtis E. Woodcock. 2012. “Object-Based Cloud and Cloud Shadow Detection in Landsat Imagery.” *Remote Sensing of Environment* 118: 83–94.

———. 2014a. “Continuous Change Detection and Classification of Land Cover Using All Available Landsat Data.” *Remote Sensing of Environment* 144: 152–71. doi:10.1016/j.rse.2014.01.011.

———. 2014b. “Automated Cloud, Cloud Shadow, and Snow Detection in Multitemporal Landsat Data: An Algorithm Designed Specifically for Monitoring Land Cover Change.” *Remote Sensing of Environment* 152 (September): 217–34. doi:10.1016/j.rse.2014.06.012.

Zhu, Zhe, Curtis E. Woodcock, and Pontus Olofsson. 2012. “Continuous Monitoring of Forest Disturbance Using All Available Landsat Imagery.” *Remote Sensing of Environment*, Landsat Legacy Special Issue, 122 (July): 75–91. doi:10.1016/j.rse.2011.10.030.

# Appendix I: Result Table

| Data | NDVI MD (%) | Change dect. | $K_{NDVI}$ | $k_{HVHH}$ | $n_{TOTAL}$ | $n_{HVHH}$ | OA | CE | OE | MTL (Month) |
|---|---|---|---|---|---|---|---|---|---|---|
| **NDVI** | 60 | bfm | NA | NA | 1 | NA | 0.9109 | 0.1039 | 0.0883 | 5.42 |
| **NDVI** | 60 | bfm | NA | NA | 2 | NA | 0.9068 | 0.0616 | 0.1428 | 7.52 |
| **NDVI** | 60 | bfm | NA | NA | 3 | NA | 0.8794 | 0.0493 | 0.2151 | 9.75 |
| **NDVI** | 60 | threshold | 1.5 | NA | 1 | NA | 0.3060 | 0.7597 | 0.0029 | 1.60 |
| **NDVI** | 60 | threshold | 1.5 | NA | 2 | NA | 0.6966 | 0.4284 | 0.0258 | 2.88 |
| **NDVI** | 60 | threshold | 1.5 | NA | 3 | NA | 0.8198 | 0.2711 | 0.0558 | 4.46 |
| **NDVI** | 60 | threshold | 2 | NA | 1 | NA | 0.4386 | 0.6640 | 0.0107 | 1.91 |
| **NDVI** | 60 | threshold | 2 | NA | 2 | NA | 0.7890 | 0.3155 | 0.0482 | 3.34 |
| **NDVI** | 60 | threshold | 2 | NA | 3 | NA | 0.8653 | 0.1778 | 0.0983 | 4.93 |
| **NDVI** | 60 | threshold | 2.5 | NA | 1 | NA | 0.5596 | 0.5691 | 0.0262 | 2.41 |
| **NDVI** | 60 | threshold | 2.5 | NA | 2 | NA | 0.8410 | 0.2266 | 0.0828 | 3.81 |
| **NDVI** | 60 | threshold | 2.5 | NA | 3 | NA | 0.8749 | 0.1161 | 0.1565 | 5.35 |
| **NDVI** | 60 | threshold | 3 | NA | 1 | NA | 0.6726 | 0.4567 | 0.0533 | 2.99 |
| **NDVI** | 60 | threshold | 3 | NA | 2 | NA | 0.8624 | 0.1546 | 0.1385 | 4.15 |
| **NDVI** | 60 | threshold | 3 | NA | 3 | NA | 0.8630 | 0.0709 | 0.2303 | 5.82 |
| **NDVI** | 60 | threshold | 3.5 | NA | 1 | NA | 0.7637 | 0.3379 | 0.0919 | 3.52 |
| **NDVI** | 60 | threshold | 3.5 | NA | 2 | NA | 0.8599 | 0.0989 | 0.2107 | 4.70 |
| **NDVI** | 60 | threshold | 3.5 | NA | 3 | NA | 0.8270 | 0.0452 | 0.3311 | 6.31 |
| **NDVI** | 60 | threshold | 4 | NA | 1 | NA | 0.8125 | 0.2389 | 0.1514 | 4.18 |
| **NDVI** | 60 | threshold | 4 | NA | 2 | NA | 0.8355 | 0.0582 | 0.3031 | 5.27 |
| **NDVI** | 60 | threshold | 4 | NA | 3 | NA | 0.7774 | 0.0275 | 0.4501 | 6.99 |
| **NDVI** | 60 | threshold | 4.5 | NA | 1 | NA | 0.8233 | 0.1587 | 0.2389 | 4.94 |
| **NDVI** | 60 | threshold | 4.5 | NA | 2 | NA | 0.7877 | 0.0331 | 0.4253 | 5.93 |
| **NDVI** | 60 | threshold | 4.5 | NA | 3 | NA | 0.7159 | 0.0144 | 0.5859 | 7.51 |
| **NDVI** | 60 | threshold | 5 | NA | 1 | NA | 0.7992 | 0.1025 | 0.3554 | 5.83 |
| **NDVI** | 60 | threshold | 5 | NA | 2 | NA | 0.7278 | 0.0208 | 0.5597 | 6.56 |
| **NDVI** | 60 | threshold | 5 | NA | 3 | NA | 0.6530 | 0.0080 | 0.7196 | 8.01 |
| **HVHH** | | bfm | NA | NA | 1 | NA | 0.931 | 0.044 | 0.103 | 4.14 |
| **HVHH** | | bfm | NA | NA | 2 | NA | 0.874 | 0.005 | 0.259 | 8.24 |
| **HVHH** | | bfm | NA | NA | 3 | NA | 0.741 | 0.004 | 0.537 | 11.77 |
| **HVHH** | | threshold | NA | 1.2 | 1 | NA | 0.807 | 0.297 | 0.011 | 2.75 |
| **HVHH** | | threshold | NA | 1.2 | 2 | NA | 0.920 | 0.097 | 0.068 | 6.96 |
| **HVHH** | | threshold | NA | 1.2 | 3 | NA | 0.833 | 0.032 | 0.330 | 10.37 |
| **HVHH** | | threshold | NA | 1.4 | 1 | NA | 0.8680 | 0.2127 | 0.0180 | 2.91 |
| **HVHH** | | threshold | NA | 1.6 | 1 | NA | 0.9078 | 0.1459 | 0.0284 | 3.11 |
| **HVHH** | | threshold | NA | 1.8 | 1 | NA | 0.9285 | 0.1011 | 0.0417 | 3.25 |
| **HVHH** | | threshold | NA | 2 | 1 | NA | 0.9366 | 0.0706 | 0.0608 | 3.43 |
| **HVHH** | | threshold | NA | 2.2 | 1 | NA | 0.9378 | 0.049 | 0.083 | 3.58 |
| **HVHH** | | threshold | NA | 2.2 | 2 | NA | 0.8980 | 0.003 | 0.210 | 7.95 |
| **HVHH** | | threshold | NA | 2.2 | 3 | NA | 0.7630 | 0.000 | 0.492 | 11.03 |
| **HVHH** | | threshold | NA | 2.4 | 1 | NA | 0.9347 | 0.0296 | 0.1085 | 3.70 |
| **HVHH** | | threshold | NA | 2.6 | 1 | NA | 0.9275 | 0.0167 | 0.1360 | 3.82 |
| **Dec. level fusion** | 60 | bfm/ threshold | NA | NA | 1 | 1 | 0.9235 | 0.1238 | 0.0274 | 2.98 |
| **Dec. level fusion** | 60 | bfm/ threshold | NA | NA | 2 | 1 | 0.9422 | 0.0795 | 0.0386 | 3.51 |

| Data | NDVI MD (%) | Change dect. | $K_{NDVI}$ | $k_{HVHH}$ | $n_{TOTAL}$ | $n_{HVHH}$ | OA | CE | OE | MTL (Month) |
|---|---|---|---|---|---|---|---|---|---|---|
| **Dec. level fusion** | 60 | bfm/ threshold | NA | NA | 3 | 1 | 0.9437 | 0.0695 | 0.0470 | 3.91 |
| **Dec. level fusion** | 60 | threshold | 3 | 2.2 | 1 | 1 | 0.6802 | 0.4502 | 0.0235 | 1.97 |
| **Dec. level fusion** | 60 | threshold | 3 | 2.2 | 2 | 1 | 0.8951 | 0.1583 | 0.0443 | 2.80 |
| **Dec. level fusion** | 60 | threshold | 3 | 2.2 | 3 | 1 | 0.9300 | 0.0859 | 0.0583 | 3.48 |
| **Dec. level fusion** | 60 | threshold | 3.5 | 2.2 | 1 | 1 | 0.7825 | 0.3290 | 0.0320 | 2.21 |
| **Dec. level fusion** | 60 | threshold | 3.5 | 2.2 | 2 | 1 | 0.9192 | 0.1096 | 0.0540 | 2.96 |
| **Dec. level fusion** | 60 | threshold | 3.5 | 2.2 | 3 | 1 | 0.9365 | 0.0666 | 0.0660 | 3.52 |
| **Dec. level fusion** | 60 | threshold | 4 | 2.2 | 1 | 1 | 0.8528 | 0.2282 | 0.0409 | 2.50 |
| **Dec. level fusion** | 60 | threshold | 4 | 2.2 | 2 | 1 | 0.9327 | 0.0781 | 0.0615 | 3.11 |
| **Dec. level fusion** | 60 | threshold | 4 | 2.2 | 3 | 1 | 0.9396 | 0.0550 | 0.0718 | 3.54 |
| **Dec. level fusion** | 60 | threshold | 4.5 | 2.2 | 1 | 1 | 0.8978 | 0.1503 | 0.0507 | 2.82 |
| **Dec. level fusion** | 60 | threshold | 4.5 | 2.2 | 2 | 1 | 0.9389 | 0.0593 | 0.0688 | 3.29 |
| **Dec. level fusion** | 60 | threshold | 4.5 | 2.2 | 3 | 1 | 0.9397 | 0.0503 | 0.0765 | 3.54 |
| **Data-level fusion** | 60 | threshold | 4.5 | NA | 1 | NA | 0.8197 | 0.1335 | 0.2352 | 3.76 |
| **Data-level fusion** | 60 | threshold | 4.5 | NA | 2 | NA | 0.8310 | 0.3083 | 0.0633 | 5.22 |
| **Data-level fusion** | 60 | threshold | 4.5 | NA | 3 | NA | 0.7605 | 0.4798 | 0.0374 | 8.18 |
| **NDVI** | 90 | bfm | NA | NA | 1 | NA | 0.7618 | 0.2868 | 0.2822 | 8.19 |
| **NDVI** | 90 | bfm | NA | NA | 2 | NA | 0.7429 | 0.1520 | 0.4688 | 13.58 |
| **NDVI** | 90 | threshold | 1.5 | NA | 1 | NA | 0.6477 | 0.4845 | 0.0677 | 6.41 |
| **NDVI** | 90 | threshold | 1.5 | NA | 2 | NA | 0.8031 | 0.2042 | 0.2464 | 11.91 |
| **NDVI** | 90 | threshold | 2 | NA | 1 | NA | 0.7236 | 0.3924 | 0.1100 | 6.72 |
| **NDVI** | 90 | threshold | 2 | NA | 2 | NA | 0.8000 | 0.1351 | 0.3298 | 12.18 |
| **NDVI** | 90 | threshold | 2.5 | NA | 1 | NA | 0.7654 | 0.3165 | 0.1692 | 7.04 |
| **NDVI** | 90 | threshold | 2.5 | NA | 2 | NA | 0.7727 | 0.0903 | 0.4277 | 12.11 |
| **NDVI** | 90 | threshold | 3 | NA | 1 | NA | 0.7818 | 0.2480 | 0.2463 | 7.32 |
| **NDVI** | 90 | threshold | 3 | NA | 2 | NA | 0.7348 | 0.0626 | 0.5279 | 12.15 |
| **NDVI** | 90 | threshold | 3.5 | NA | 1 | NA | 0.7796 | 0.1836 | 0.3349 | 7.50 |
| **NDVI** | 90 | threshold | 3.5 | NA | 2 | NA | 0.6912 | 0.0454 | 0.6295 | 12.13 |
| **NDVI** | 90 | threshold | 4 | NA | 1 | NA | 0.7538 | 0.1391 | 0.4410 | 7.83 |
| **NDVI** | 90 | threshold | 4 | NA | 2 | NA | 0.6438 | 0.0323 | 0.7348 | 11.82 |
| **NDVI** | 90 | threshold | 4.5 | NA | 1 | NA | 0.7128 | 0.1019 | 0.5611 | 7.99 |
| **NDVI** | 90 | threshold | 4.5 | NA | 2 | NA | 0.6003 | 0.0225 | 0.8285 | 11.53 |
| **NDVI** | 90 | threshold | 5 | NA | 1 | NA | 0.6664 | 0.0730 | 0.6773 | 8.19 |
| **NDVI** | 90 | threshold | 5 | NA | 2 | NA | 0.5669 | 0.0178 | 0.8988 | 11.14 |
| **Dec.Fusion** | 90 | bfm/ threshold | NA | NA | 1 | NA | 0.8575 | 0.2312 | 0.0424 | 3.28 |
| **Dec.Fusion** | 90 | bfm/ threshold | NA | NA | 3 | NA | 0.9338 | 0.0755 | 0.0645 | 3.69 |
| **Dec.Fusion** | 90 | threshold | 3 | 2.2 | 1 | 1 | 0.8652 | 0.2075 | 0.0478 | 3.12 |
| **Dec.Fusion** | 90 | threshold | 3 | 2.2 | 2 | 1 | 0.9352 | 0.0638 | 0.0724 | 3.62 |
| **Dec.Fusion** | 90 | threshold | 3.5 | 2.2 | 1 | 1 | 0.8983 | 0.1458 | 0.0572 | 3.19 |
| **Dec.Fusion** | 90 | threshold | 3.5 | 2.2 | 2 | 1 | 0.9374 | 0.0566 | 0.0750 | 3.61 |
| **Dec.Fusion** | 90 | threshold | 4 | 2.2 | 1 | 1 | 0.9171 | 0.1052 | 0.0655 | 3.28 |
| **Dec.Fusion** | 90 | threshold | 4 | 2.2 | 2 | 1 | 0.9382 | 0.0524 | 0.0777 | 3.60 |
| **Dec.Fusion** | 90 | threshold | 4.5 | 2.2 | 1 | 1 | 0.9293 | 0.0769 | 0.0707 | 3.38 |
| **Dec.Fusion** | 90 | threshold | 4.5 | 2.2 | 2 | 1 | 0.9385 | 0.0497 | 0.0798 | 3.60 |

*Table 2: Validation results for selected settings which represent the general trends. Results plotted in Figure 13 are highlighted.*

# Appendix II: Calculating Spatio-Temporal Accuracy Measures

Be $x_i$ the number of correctly classified space points at time $i$. Spatial overall accuracy at time $i$, is then calculated as $OA_i = \frac{x_i}{m}$, with m being the total number of space-points (see e.g. Foody 2002 for calculation of space-based accuracy measures).

Similarly, for $y_j$ being the number of correctly classified time points at location j, temporal overall accuracy for location $j$, is calculated as $OA_j = \frac{y_j}{n}$, with n being the total number of time-points.

Global OA can then either be calculated as average of all $OA_j$ over space or as average of all $OA_i$ over time: $OA_{global} = \frac{1}{n}\sum_{i=1}^{n} OA_i = \frac{1}{n}\sum_{i=1}^{n} \frac{x_i}{m} = \frac{1}{m \times n}\sum_{i=1}^{m} x_i = \frac{1}{N}\sum_{i=1}^{m} x_i$ or similarly $OA_{global} = \frac{1}{N}\sum_{j=1}^{m} y_j$. Therefore $\sum_{j=1}^{m} y_j = \sum_{i=1}^{m} x_i$ = the total sum of time-space-points classified correctly.

The time-based, space-based and global CE and OE or producer's- and user's-accuracies per class can be derived in similar ways, using the respective correct or incorrect pixels per map or reference class.