Monitoring deforestation and forest degradation linking high-resolution satellite data and field data in the context of REDD+ A case of Tanzania

PhD candidate: Lorena Hojas Gascón

PhD directors: Hugh Eva and Francisco Javier García Haro

VNIVERSITAT DE VALÈNCIA

European Commission, Joint Research Centre

Doctorat en Teledetecció

Maig de 2017

PHD THESIS IN REMOTE SENSING

BY

LORENA HOJAS GASCÓN

DEPARTAMENT DE FÍSICA DE LA TERRA I TERMODINÀMICA. FACULTAT DE FÍSICA

European Commission, Joint Research Centre

VNIVERSITAT DE VALÈNCIA

Monitoring deforestation and forest degradation linking high-resolution satellite data and field data in the context of REDD+ A case of Tanzania

DR. HUGH EVA, Doctor en Geografía, Investigador del Joint Research Centre (JRC), European Commission, Ispra, Italy

y

DR. FRANCISCO JAVIER GARCÍA HARO, Doctor en Física, Profesor Titular de Universitat del Departament de Física de la Terra i Termodinàmica de la Facultat de Física de la Universitat de València

HACEN CONSTAR QUE:

La Licenciada en Ciencias Ambientales LORENA HOJAS GASCÓN ha realizado bajo su dirección el trabajo titulado *Monitoring deforestation and forest degradation linking high-resolution satellite data and field data in the context of REDD+ A case of Tanzania*, que se presenta en esta memoria para optar al grado de Doctor por la Universitat de València.

Y para que así conste a los efectos oportunos, dando el visto bueno a la presentación de este trabajo delante del tribunal de tesis que corresponda, firmamos el presente certificado a 20 de Mayo de 2017.

Hugh Eva

Francisco Javier García Haro

Abstract

The main objective of this PhD is to support the development a forest monitoring system in Tanzania so as to report on current and historical emissions which derive from deforestation and forest degradation. The framework of the thesis is specifically focused on the emerging international context of the REDD+ (Reducing Emissions from Deforestation and Degradation) initiative, under which countries may obtain financial grants for demonstrating that they are reducing their carbon emissions from forest lands with respect to their recent historical practice.

In view of the complexity of this goal, the research focused on five focal areas of research:

Part (1) reviews the policy background to REDD+. It outlines the choices to be addressed by participatory countries and demonstrates some of the technical problems and options that they can face and adopt in the remote sensing technology. Research published in scientific literature, on institutional websites and national agencies were reviewed and assessed.

Part (2) presents the results from the PhD field work in Tanzania. This included the set-up of rapid field data collection, guidelines on execution and protocols to link the field data to the remote sensing data, so as to produce maps on above ground biomass, tree height and vegetation cover using very high resolution images.

Part (3) demonstrates the improvement to map forests at a fine spatial resolution and with high frequency of acquisitions with the arrival of the new Sentinel-2 satellites. This potential has been tested on an area of dry forest in Central Tanzania.

Part (4) tests a full scale estimate of above ground biomass for the whole of Tanzania, using a combination of remote sensing and field data. The predictive capability was investigated by comparing the results against ground measurements undertaken by the Tanzania forest service's national inventory.

Part (5) investigates the dynamics of deforestation around Dar es Salaam, along with a model to infer future probability of deforestation at the national level. The ability of the model to replicate spatial patterns of deforestation was assessed through ground truthing.

Among the main outcome of this PhD is that estimates of forest change from different sources have wide variance at national level and emissions estimates for the REDD process remain unreliable. The thesis has demonstrated there are a large number of choices facing a forest service, in terms of forest definitions and methods, and all may have an impact of the feasibility of implementation and on results. The difficulty of linking remote sensing data to the forest parameter collected by national surveys has been shown – with recommendations as how to improve future remote sensing – field survey data collection.

The use of image segmentation, and texture indices, was found to be useful the production of local maps of forest biomass from field data in conjunction with very high resolution satellite imagery Additionally, the arrival of Sentinel-2 data provides the opportunity to analyse medium high resolution data (<20m) in time series – only possible beforehand with low resolution data.

Lastly, from a technical perspective, the synergistic use of remote sensing and national field survey data can effectively contribute to the Tanzania forest service's needs and requirements for mapping and monitoring forest changes – notably forest degradation. Further, it reduces the (costly) field component and it shows a large potential as an operational pan-African monitoring system.

Spanish summary

La deforestación y la degradación forestal son la segunda fuente antrópica de gases invernadero emitidos a la atmosfera, después de los combustibles fósiles y antes incluso que el sector del transporte. Se estima que la deforestación y la degradación de los bosques tropicales suponen hasta el 14% de las emisiones globales de gases invernadero provocadas por el hombre y el 20% del total. Por tanto, reducir estas emisiones es una forma efectiva de combatir el cambio climático.

El área forestal global disminuyó un 3%, de 4128 M ha en 1990 a 3999 M ha en 2015, aunque la tasa anual de deforestación de 1990 se redujo a la mitad a 3.3 M ha entre 2010 y 2015. Los bosques tropicales disminuyeron con una tasa de 5.5 M ha al año desde 2010 a 2015. Entre estos años, África redujo su área forestal, siendo Tanzania el segundo país de mayor pérdida, seguido de Nigeria, causada por los porcentajes más altos de crecimiento poblacional a nivel mundial.

El programa de reducción de las emisiones de deforestación (RED) de la Convención Marco de las Naciones Unidas sobre el Cambio Climático (UNFCCC) apareció en 2005 con el objetivo de recompensar económicamente a los países en desarrollo por reducir sus emisiones de gases invernadero de los bosques. Este más tarde evolucionó en REDD+ (reducción de las emisiones de deforestación y degradación forestal), incluyendo también la reducción de emisiones provocadas por degradación forestal. Desde entonces el número de iniciativas nacionales relacionadas con REDD+ ha aumentado en los últimos años, siendo Tanzania uno de los dieciséis países piloto que recibieron el apoyo directo de las Naciones Unidas.

El concepto propone que los países puedan reclamar incentivos financieros reportando los cambios de área en la cubierta forestal (datos de actividad) y la densidad de las reservas de carbono forestal (factores de emisión) para estimar las emisiones antropogénicas de gases invernadero relacionadas con los bosques. Para ello, deberán establecer sistemas nacionales de monitoreo que usen una combinación apropiada de métodos de teledetección y de inventarios de campo de carbono forestal. Los países tendrán además que evaluar los cambios históricos en la cubierta forestal y establecer niveles de emisión de referencia forestal (FREL).

La definición de "bosque" tiene implicaciones importantes en el proceso. El establecimiento de una definición nacional de "bosque" es esencial para monitorear los cambios en el área forestal y un requisito previo para desarrollar un sistema de monitoreo de FREL consistente. La UNFCCC propuso que un bosque es un área de tierra con al menos 0,05-1 ha y una cobertura arbórea mínima del 10 al 30%, con árboles que han alcanzado o podrían alcanzar una altura mínima de 2-5 m en su madurez. Dentro de estos límites, los países pueden seleccionar su definición de bosque nacional.

De acuerdo a esta definición, el reporte de los datos de actividad y de los factores de emisión en el marco de REDD+ planteó nuevos requisitos para el monitoreo de la deforestación y la degradación de los bosques a nivel nacional. La deforestación fue definida como una disminución de la cobertura de la copa de los árboles inferior al 10-30%, y la degradación como una pérdida de reservas de carbono con una disminución en la cobertura de la copa de los árboles no inferior al 10-30%, de áreas forestales con un tamaño mínimo de 0,05-1 ha inducidas directamente por el hombre.

La deforestación ha sido monitoreada con éxito en los trópicos usando datos de satélite de resolución espacial moderada, predominantemente a partir del sensor Landsat de resolución espacial de 30 m. El Centro Común de Investigación (JRC) de la Comisión Europea, en el marco del proyecto TREES-3, ha llevado a cabo una evaluación del cambio de la cubierta forestal en los trópicos para los períodos 1990-2000, 2000-2005 y 2005-2010 utilizando un muestreo sistemático de imágenes Landsat , apoyando el estudio de teledetección de la Organización de las Naciones Unidas para la Agricultura y la Alimentación (FAO) sobre valoración a nivel mundial de los recursos forestales (FRA) cada cinco años basada en los informes de los servicios forestales nacionales. Proyectos similares se han llevado a cabo a escala nacional y regional, alcanzando unas precisiones de aproximadamente 90% con diferentes unidades cartográficas mínimas de hasta 5 ha utilizando datos de Landsat.

Más recientemente, la Universidad de Maryland y Google han producido los Mapas Globales de Alta Resolución del Cambio de Cubierta Forestal del siglo XXI (GFM), basados en una síntesis del archivo Landsat de 2000 a 2012, donde a cada píxel se le ha dado un porcentaje de cobertura arbórea, además de datos anuales de pérdidas forestales. Estos y otros intentos de mapear la cobertura forestal mundial y los cambio de cobertura forestal basados en Landsat han sido considerados por los países como una oportunidad para evaluar los cambios históricos en la cubierta forestal y establecer los niveles de emisión de referencia forestal (FREL), especialmente los GFM de cobertura continua.

Para mapear la degradación forestal tradicionalmente se han empleado varias técnicas y enfoques diferentes. En el contexto de REDD+ la degradación forestal es considerada como una reducción de la cubierta forestal a pequeña escala que conlleva una pérdida de carbono y de productividad a largo término. La densidad de carbono de un bosque se puede estimar a partir de datos de campo de altura de los árboles y área basal y ecuaciones barométricas.

La cartografía y el monitoreo de las reservas de carbono forestales de grandes áreas en los trópicos se ha basado cada vez más en enfoques de teledetección, que a su vez dependen de mediciones de

campo de biomasa para la calibración y la validación. Las estimaciones directas de biomasa basadas únicamente en datos satelitales no son todavía factibles.

Los sistemas en evolución como el nuevo RADAR de banda P y LIDAR no se revisan en la tesis, ya que requieren un tratamiento completo fuera del contexto de este trabajo. Los sistemas ópticos claramente no son óptimos para monitorear la biomasa, ya que sólo representan la superficie de la vegetación. Sin embargo, dada la disponibilidad de estos datos (RADAR de banda P y LIDAR aún no se han distribuido ampliamente), la tesis trata de ver si se pueden utilizar, junto con datos de campo, para mejorar (en términos de eficiencia) estimaciones de biomasa a nivel nacional en Tanzania.

La degradación de los bosques es una dinámica importante en Tanzania. La extracción de madera es el principal motor de la degradación forestal, principalmente para el consumo de energía en forma de leña y carbón vegetal. Estos representan más del 75% del consumo total de energía, principalmente para la cocina doméstica. Más del 80% de la población urbana consume carbón vegetal y se prevé un incremento de la presión sobre los recursos madereros debido al crecimiento de la población urbana; la población actual de 45M se duplicará en 2050. La zona más afectada del país es el oeste y norte de Dar es Salaam, donde la producción de carbón vegetal para suministrar a la ciudad es responsable de la degradación de 29 ha (24.6%) y la deforestación de 23ha (19.58%) de bosques cerrados y 93 ha (50.8%) de bosques.

Las bases para implementar los sistemas naciones de monitoreo para REDD+

La implementación de sistemas nacionales de monitoreo para estimar los cambios en la cubierta forestal y las reservas de carbono forestal implica la selección de fuentes de información, variables de monitoreo y técnicas de medición.

El capítulo 4 describe las reglas y opciones que deben ser abordadas por los países participantes en REDD + de acuerdo con las directrices de la UNFCCC en la implementación de los sistemas de monitoreo para estimar los cambios en la cubierta forestal y las reservas de carbono forestal, y demuestra algunos de los problemas técnicos y opciones que pueden enfrentar y adoptar en los ejercicios de teledetección. En particular revisa dos aspectos: 1) definición nacional de 'bosque' para la cartografía de la cubierta forestal, 2) datos globales para evaluar los cambios históricos en la cobertura de la tierra y establecer los FREL, como los GFM. Para ello se siguieron los siguientes pasos:

1. Análisis teórico de los impactos potenciales de la selección de la unidad de cartografía mínima (UCM), de la cobertura forestal mínima y de la altura mínima de los árboles de la definición forestal nacional.

2. Comparación de los GFMs con los datos del proyecto TREES-3 del JRC y con las estadísticas nacionales de la FAO FRA derivadas del Servicio Forestal de Tanzania, del año 2000 y del cambio de la cubierta forestal entre 2000 y 2010 para Tanzania.

3. Para analizar las causas de las diferencias de los datos anteriores y de la consistencia de los GFMs, comparación de la cobertura forestal del GFM de 2010 con las imágenes de resolución espacial muy alta del mismo año y datos de campo en cuatro sitios de estudio de las principales ecorregiones de Tanzania.

4. Evaluación del rendimiento de los datos de Landsat y RapidEye en la cartografía de bosques fragmentados (con una UCM de 0,5 ha) para un sitio de estudio, tomando datos de resolución espacial muy alta (<1 m de resolución espacial) como referencia de la "verdadera" área forestal.

Los principales hallazgos fueron:

Los valores elegidos en la definición nacional de bosque tendrán consecuencias técnicas, económicas y políticas que serán más o menos adecuadas para los países dependiendo de sus características y objetivos particulares.

Se encontraron importantes diferencias en la extensión real y la magnitud de los cambios en los conjuntos de datos globales existentes basados en Landsat sobre la cobertura forestal y el cambio de cobertura forestal (GFM, JRC, FAO) a nivel nacional, procedentes en parte de las diferentes técnicas empleadas, y en parte de las diferentes definiciones de bosque consideradas. Esas diferencias deben tenerse en cuenta cuando se utilizan como datos de referencia para establecer los FREL y/o resolverse para dar confianza a los responsables políticos y a las agencias potenciales de financiación.

En dos áreas de estudio en Tanzania se encontraron "errores" importantes en la distribución del porcentaje de cobertura forestal de los GFM en comparación con datos satelitales de resolución espacial muy alta. GFM sobrestimó cobertura forestal en el bosque húmedo tropical, desestimo cobertura forestal en el bosque seco tropical, y falló en detectar grandes áreas de cambio. Esto es debido a la resolución espacial de los datos del Landsat y al método de calibración empleado con respecto a la estacionalidad y la diversidad de ecorregiones.

Se demostró que la resolución espacial del sensor Landsat es demasiado baja para monitorear la deforestación para REDD + en comparación con los datos satelitales de alta resolución espacial (5 m RapidEye) en un área de estudio del bosque húmedo. Otro problema encontrado del uso de GFMs para establecer los FREL es la gran variación en el número de adquisiciones de Landsat en las últimas décadas (muy bajo antes de 2013 sobre áreas tropicales), lo que puede resultar en un sesgo en el

número de eventos de deforestación detectados y comprometer las estimaciones de deforestación entre períodos históricos y recientes.

La estimación de la degradación forestal es aún más difícil de medir que la deforestación con los datos Landsat de resolución espacial media, ya que requiere detectar cambios menos perceptibles en la cobertura forestal, cambios de menos del 30% en áreas de UCM, es decir cambios del tamaño de un píxel. Además, consiste no sólo en medir cambios en la cubierta forestal, sino también en las reservas de carbono, para lo cual se requieren datos de campo.

Integración de datos de campo y teledetección

En general, no es fácil obtener datos de campo para estimar los cambios en la cubierta forestal y las reservas de carbono forestal para estudios a escala nacional. Los estudios de campo a nivel nacional son muy costosos y requieren mucho tiempo debido a la cantidad de datos requeridos. Por lo tanto, muchos países carecen de inventarios forestales o cuando los tienen, pueden ser anticuados o incomprensibles, no estar disponibles para organismos externos y/o no ser apropiados para la aplicación deseada, en nuestro caso, la integración y explotación de datos satelitales.

El objetivo del capítulo 5 es dar orientación en la planificación, ejecución y explotación de un estudio de campo rápido y sencillo que pueda servir en la interpretación, calibración y validación de datos de teledetección de resolución espacial muy alta (<1m), en particular, para producir mapas de cobertura forestal y clasificaciones de biomasa forestal. Estos pueden servir para mejorar la ordenación forestal sostenible a nivel local, y en el futuro para el monitoreo de la deforestación y de la degradación forestal para la estrategia nacional de REDD+.

Durante la tesis se llevaron a cabo tres campañas de campo en 2012, 2013 y 2014 conjuntamente con el Servicio Forestal de Tanzania en varios sitios de estudio de bosque intacto y no intacto. Por limitaciones de tiempo, se requirió una estrategia de campaña eficiente con un método de muestreo selectivo para recolectar una muestra específica de datos de campo (ej. cubierta de copas, diámetro de los troncos, altura de los árboles y distribución de especies).

Los datos de campo recogidos se utilizaron para generar modelos en conjunto con datos satelitales de resolución espacial muy alta para producir mapas de cobertura forestal, área basal y biomasa de la reserva forestal de Pugu Hills, en el bosque tropical húmedo. Además se revisaron las diferencias en la composición de especies entre sitios ubicados en diferentes ecozonas y entre bosque intacto y no intacto

Para ello 1) se seleccionaron los parámetros de teledetección espectrales y texturales (características texturales de Haralick); 2) se calcularon las correlaciones y modelos de regresión

múltiple entre los parámetros de teledetección y los parámetros extraídos de los datos de campo; y 3) se extrapolaron los mejores modelos de regresión para clasificar las unidades forestales en clases de área basal, altura media de las copas y biomasa aérea.

Los modelos de regresión múltiple para la predicción de las variables relacionadas con la biomasa forestal a partir de los parámetros de teledetección mejoraron significativamente mediante la adición de medidas texturales, lo que sugiere que la estructura de la cobertura forestal está bien representada por la textura de las imágenes satelitales de resolución espacial muy alta.

La cubierta media de las copas de los árboles tuvo una menor correlación con los parámetros de teledetección, posiblemente debido a que las mediciones de campo se realizaron a 1,6 m de altura, incluyendo capas de árboles y arbustos cubiertos por capas superiores no registradas por las imágenes satelitales superiores del pabellón vistas por el satélite.

Los mapas obtuvieron resultados comparables a otros estudios y demostraron el alto nivel de degradación de la región húmeda en comparación con datos nacionales.

Para los estudios nacionales las imágenes satelitales de altísima resolución espacial tienen el inconveniente de ser difíciles de encontrar para ciertas zonas, muy caras y de baja cobertura, por lo que es difícil cubrir un país completo. En estos casos, pueden ser útiles para calibrar imágenes satelitales de menor resolución espacial y ayudar en la planificación e implementación de los trabajos de campo.

Estimaciones nacionales de biomasa

En Tanzania el programa nacional de monitoreo y evaluación de los bosques (NAFORMA) llevó a cabo un extenso inventario nacional de campo de recursos forestales durante un período de tres años (2010-2013). El inventario siguió un muestreo estratificado de 32.660 parcelas de 15m de radio por todo el país, diseñado para estimar el área forestal y la biomasa a nivel nacional.

Si bien el mapeo de las área forestales en grandes áreas es relativamente fácil utilizando la teledetección (aunque vemos una alta varianza en las estimaciones de los diferentes estudios), la estimación de la biomasa es mucho más problemática, ya que la información de las mediciones de campo tiene que ser extrapoladas a las propiedades espaciales de los datos satelitales.

En el capítulo 7 se hace una estimación de la biomasa forestal por encima del suelo a nivel nacional combinando datos del satélite RapidEye, que tiene un buen equilibrio de resolución espacial (5 m), cobertura de barrido y frecuencia de revisita, con datos de campo de NAFORMA. Para ello se adquirieron imágenes de RapidEye del año 2010 distribuidas sistemáticamente sobre el territorio de

Tanzania en 76 sitios de muestreo, cada uno de 20 km por 20 km, y la información de las parcelas correspondientes a las imágenes extraídas del inventario nacional de NAFORMA. La ubicación de los sitios en el punto de confluencia de la red geográfica se eligió para coincidir con el esquema de muestreo del estudio JRC-FAO para el cambio forestal y así poder detectar cambios históricos.

Las imágenes RapidEye (419) se adquirieron en el marco del programa GMES (Monitoreo Global para el Medio Ambiente y la Seguridad) de la Unión Europea. Para tener acceso a una muestra de los datos del inventario de campo nacional el JRC hizo un acuerdo con el Servicio Nacional de Tanzania en 2015.

La estimación se basó en el desarrollo de un modelo para relacionar los parámetros de teledetección extraídos de las imágenes (basados en reflectancia y textura) con los datos de campo (biomasa por encima del suelo, área basal y altura de los arboles) para los sitios correspondientes. El modelo se aplicó después a toda la extensión de cada una de las 76 imágenes para obtener el promedio de biomasa por sitio de muestreo, y luego, por expansión directa, del país.

Para ello 1) las imágenes se procesaron para tener objetos de estudio con una UCM (segmentación), que fue elegida de acuerdo con la definición de bosque del servicio nacional forestal, con la guía internacional para cumplir con los requisitos de presentación de informes para REDD y con metodología empleada en este estudio; 2) se extrajeron una gama muy extensa de índices de reflectancia y de textura de la imagen de los objetos que correspondieron a los datos del campo (localización geométrica); 3) se revisaron y "limpiaron" los datos de campo con respecto a las imágenes de satélite para eliminar datos ambiguos o erróneos; y 4) se probaron modelos para relacionar los datos de campo con los parámetros extraídos de las imágenes. Se probaron dos modelos de estadística inferencial: un modelo lineal generalizado y un modelo exponencial generalizado, y dos modelos de aprendizaje automático: 'Random Forest' y Support Vector Machine (SVM).

Los parámetros de teledetección más importantes en la correlación fueron el índice de sombra seguido de una medida de textura. El mejor ajuste fue proporcionado por el modelo 'Random Forest'.

A pesar de los grandes esfuerzos y costes del estudio nacional campo, la base de datos resultante de los parámetros de los bosques espacialmente explícitos fue difícil de explotar con datos de satélite. Se entraron problemas de desajuste geográfico entre los datos de campo y de satélite, un tamaño de las parcelas pequeño en comparación con la resolución espacial del satélite, cambios en la cobertura del suelo entre los dos conjuntos de datos (ej. por estacionalidad, inundaciones o incendios) e información de campo no adecuada para la calibración o validación de las imágenes. En al capítulo 7

se describen estos problemas y las soluciones aplicadas. Se requirió entonces un gran esfuerzo para "limpiar" los datos de campo, concluyendo que si los datos de campo deben ser comparados y explotados con imágenes de teledetección, las campañas de campo deben ser diseñadas específicamente para ello.

La mejora potencial de la estimación de la biomas que resultó de la combinación de un segundo conjunto de datos con los datos de campo, en este caso los datos de teledetección, se midió mediante la Eficiencia Relativa (ER). Se encontró un ER de 2,9, lo que significa que a pesar de los problemas, para obtener los mismos niveles de precisión, se necesitarían tres veces más parcelas de datos de campo que si se emplea datos RapidEye combinados con datos de campo. Esto implica que la recopilación de datos de campo puede reducirse, con significativos ahorros de costes.

Se concluyó que una combinación de 11.000 parcelas de campo (en lugar de 32.000 utilizadas por el estudio nacional) podría utilizarse con los datos de RapidEye para obtener el mismo nivel de precisión. Esto daría como resultado un gran ahorro de presupuesto y de tiempo; el trabajo de campo de tres años podría haber sido completado en un año con un tercio del coste de mano de obra (las imágenes de satélite tendrían que ser compradas).

También se muestra que la resolución de 5m de los datos de RapidEye está probablemente en el límite de la explotación; datos de 30m Landsat y 10m Sentinel produjeron niveles mucho más bajos de precisión).

Análisis multitemporales de imágenes satelitales de alta resolución

Períodos cortos de revisita de los satélites son potencialmente importantes para monitorear el bosque a escala nacional y regional. En primer lugar, el aumento de la cobertura ofrece más oportunidades de adquirir imágenes libres de nubes, particularmente importantes en las regiones tropicales húmedas. En segundo lugar, porque nos permitirá explotar las diferencias estacionales de las características de la reflectancia de la cubierta forestal como un medio de discriminar entre tipos de cobertura forestal y diferentes condiciones forestales (por ejemplo, bosques cerrados y abiertos o bosques caducifolios y degradados), lo que es especialmente importante para el bosque seco. El capítulo 6 se centrará en las necesidades especiales de las zonas secas, en el análisis de imágenes multitemporales de satelitales de alta resolución.

Los satélites Sentinel-2A y Sentinel-2B se lanzaron en junio de 2015 y marzo de 2017, respectivamente. Tienen frecuencias de revisión más altas (en conjunto 5 días), más bandas espectrales con anchuras de banda más estrechas (incluyendo tres bandas en el borde rojo, que ha demostrado ser útil para la evaluación cuantitativa del estado de la vegetación), resolución espacial

más fina (tres bandas visibles y una cerca del infrarrojo a 10 m) y mejor resolución radiométrica que los satélites SPOT y Landsat. La disponibilidad de datos sobre áreas objetivo aumentará, permitiendo el análisis de series temporales, generalmente restringido a datos de satélite de resolución espacial moderada (> 100m), en resoluciones espaciales más finas (10m), lo que permitirá una evaluación más precisa de la deforestación y degradación forestal a las escalas espaciales definidas por la UNFCCC. Además, las imágenes Sentinel tienen un acceso gratuito.

Mientras que en el momento del estudio, Sentinel-2 aún no había sido lanzado, y actualmente todavía hay pocas imágenes de Sentinel-2B, la agencia espacial francesa CNES bajó la órbita de SPOT4 para simular la frecuencia de revisita de Sentinel-2 durante el 29 de enero 2013 hasta el 19 de junio (programa 'SPOT4 take 5') con el fin de prepararse para el uso de sus datos. Para nuestro sitio de estudio ubicado en las tierras altas secas de Tanzania Central, obtuvimos 23 imágenes SPOT que van desde el final de la temporada húmeda hasta el final de la estación seca.

El objetivo es evaluar el potencial de la serie temporal de SPOT (simulando Sentinel-2) para discriminar entre diferentes clases de bosque, con el objetivo de examinar si la mejora del muestreo temporal con datos satelitales de alta resolución espacial (20m) mejora realmente nuestro conocimiento sobre la deforestación y la degradación forestal en los ecosistemas de bosques secos.

Para ello 1) se estimó el incremento de la disponibilidad de datos, comparando el área de imagen libre de nubes de SPOT4 Take 5 con la de Landsat; 2) se evaluó la mejora de la resolución temporal, comparando la serie temporal de SPOT4 Take 5 con la de MODIS; y 3) se estimó la mejora de la precisión de la clasificación forestal, calculando la capacidad de separar las clases de bosques utilizando Sentinel-2 A y B y sólo Sentinel-2 A en relación con la adquisición actual.

Se encontró un incremento importante de la frecuencia de datos sin nubes con Sentinel-2 (SPOT) con respecto a Landsat-8, y que la serie temporal SPOT NDVI compuesta por valores de adquisición únicos, se corresponde bien con la serie temporal MODIS compuesta por los valores medios de periodos de 16 días. El NDVI de las unidades forestales cae desde un pico al comienzo del período de observación (final de la estación húmeda) hasta el final del período (estación seca). El NDVI medio de las diferentes clases pasa de ser muy divergente al final de la estación húmeda para ser muy similar en la estación seca.

Hay una mejora general de las capacidades de separación de clases aumentando la frecuencia de las observaciones. Se observó un incremento positivo del 5,7% en la relación media de la distancia Jeffries-Matusita Distance (JM) con todas las observaciones (es decir, Sentinel A + B) respecto a la mitad de las observaciones (sólo un Sentinel) y una mejora del 14,5% al utilizar Sentinel A y B con respecto a un tercio de las observaciones (frecuencia de adquisición actual).

Utilizando ambos Sentinels, la distancia JM es suficiente para facilitar la discriminación entre las clases, cuando la cubierta forestal cae desde el 40%. Random forest sostiene los mismos hallazgos las clases cerradas (diferencia de la cubierta más pequeña que el 20%), son difíciles de discriminar mientras que cuando la diferencia de la cubierta forestal entre dos clases es mayor o igual al 40%, la separabilidad es fácil con una probabilidad del aproximadamente 95% .

Perspectivas futuras de los bosques en Tanzania

Tanzania está perdiendo entre 200.000 y 400.000 ha de bosques al año. Se ha estimado que el 90% de esta pérdida se debe al consumo de carbón y leña y que Dar es Salaam representa el 30% del consumo nacional. Con la población de Tanzania se duplicará para 2050, claramente el impacto en la cobertura forestal del país será enorme a menos que se encuentren fuentes alternativas de energía y se ofrezcan a la población a precios asequibles.

El impacto de la aglomeración urbana de Dar es Salaam se revisa en el capítulo 8. Su influencia sobre la cubierta forestal muestra una serie de dinámicas. En primer lugar, una ola de deforestación está saliendo de la ciudad, alimentando la demanda de construcción y, más importante aún, de leña. En segundo lugar, la integridad de las áreas protegidas se ve seriamente comprometida por la necesidad de abastecer a la capital con madera y carbón para cocinar. Las visitas de campo realizadas durante la tesis en la reserva de Pugu Hills, confirmaron esta dinámica, encontrando 5 hornos de carbón ilegal en 10 días.

El análisis de la deforestación histórica con un conjunto de parámetros geográficos a nivel nacional muestra que cuanto más cerca de las ciudades, carreteras, áreas deforestadas y bordes forestales, mayor es la probabilidad de deforestación. También la pendiente y la altitud tienen una fuerte relación con la deforestación: la probabilidad de deforestación disminuye cuando la pendiente y la altitud crecen. Tal vez no es sorprendente, pero cuando se mapea espacialmente, tal información puede ayudar a los tomar de decisiones.

Conclusiones generales

- Las estimaciones del área forestal de Tanzania y de los cambios son variadas y necesitan una revisión completa

- La teledetección puede ayudar en el mapeo local para apoyar el manejo forestal

- Es necesario modificar las metodologías tradicionales de los inventarios forestales de campo para que sean útiles para la explotación de imágenes de teledetección

- El uso de datos ópticos de resolución espacial más fina puede dar grandes ahorros en los inventarios nacionales de campo

- La explotación de la resolución temporal mejorada de datos de resolución espacial más fina, como la de Sentinel-2, puede traer mejoras importantes en la discriminación de tipos de bosque y de coberturas forestales

- La revisión de los mapas espaciales históricos de deforestación, junto con otros conjuntos de datos geoespaciales, puede ayudar a tomar medidas para reducir la deforestación y la degradación forestal

Acknowledgements

This PhD was funded by a JRC doctoral grant and hosted by the University of Valencia. It was also supported by CIFOR, the Center for International Forestry Research - Bogor, under a Collaboration Agreement with the JRC. Without the cooperation of the Tanzania Forest Services (TFS) and the Food and Agriculture Organization (FAO) of the United Nations the collection of the field data would not have been possible. The EU Delegation in Dar es Salaam was key to effecting a collaboration agreement with the Tanzanian authorities. The UN-REDD Tanzania provided the permits required to carry out work in the country.

I would specifically like to thank the following for their support:

The Tanzania Forest Service
Mr. Juma Mgoo, Former Chief Executive Officer
Mr. Nurdin Chamuya
Dr. Rogers Malingwi
Dr. Jared Otieno
Dr. Boniface Mbilinyi
Mr. Mathew Mwamuo
Mr. Jonathan Tangwa
Mr. Robert Nyali
Mr. Iddy Beya
Mr. Ally J. Mtosa
Mr. Yoab Kusimula
Mr. Moses Tenisula
Mr. Msalika Pastory

FAO
Mr. Søren Dalsgaard, Former Chief Technical Advisor Tanzania
Dr. Anssi Pekkarinen, FIN-FAO
Mr. Lauri Tamminen, Former FIN-FAO
Mr. Amana, FAO
Mr. Philipe Crete, FAO-REDD

EU Delegation in Tanzania
Mr. Filiberto Ceriani Sebregondi, Head of Delegation in Tanzania
Mr. Gianluca Azzoni, Head of Natural Resources Section
Ms. Maria Iarrera

CIFOR
Dr. Robert Nasi
Dr. Paolo Cerutti

JRC Ispra
Dr. Frédéric Achard
Ms. Silvia Carboni
Dr. Guido Ceccherini
Ing. Roman Seliger
Ing. Dario Simonetti
Dr. Ing. Hans Jürgen Stibig

Acronyms

COP Conference of the Parties

DBH Diameter at Breast Height

EC European Commission

FAO Food and Agriculture Organization of the United Nations

FRA Forest Resources Assessment

FREL Forest Reference Emission Levels

GHGs Greenhouse Gasses

IPCC Intergovernmental Panel on Climate Change

JRC Joint Research Centre

MMU Minimum Mapping Unit

MNRT Ministry of Natural Resources of Tanzania

NAFORMA National Forest Monitoring and Assessment

UNFCCC United Nations Framework Convention on Climate Change

REDD Reducing Emissions from Deforestation and Forest Degradation

REDD+ Reducing Emissions from Deforestation and Forest Degradation and conservation, sustainable management of forests and enhancement of forest carbon stocks

TREES-3 Tropical Environmental Monitoring by Satellite

VHR Very High Resolution

Table of Contents

Figures

Tables

1. Introduction

The driving issues behind this research came from supporting a number of ongoing research areas and priorities in the European Commission's Joint Research Centre, where the TREES project has been mapping and monitoring tropical forests since 1992. This study sort to build on that heritage and contribute to new areas of work, notably the REDD (Reducing Emissions from Deforestation and forest Degradation) initiative. Key factors in the choice of this area of work are listed.

1. European Union cooperation in Tanzania in climate change adaptation and fuel wood service for poverty reduction and environmental conservation.

2. Supporting the TREES project in the context of developing remote sensing methods to support countries in reporting for the REDD initiative.

3. Looking to explore the applications of the high resolution data from Sentinel-2 satellites (funded by the EC).

4. Previous collaboration with FRA (Forest Resource Assessment) of the Food and Agricultural Organization (FAO).

The information provided by this study can on one hand, increase effectiveness of suitable forest management (e.g. to control the fuel wood production capacity and detect areas of illegal harvesting) for national objectives, and in the future, improve forest-based biomass data and national carbon budget estimates by repeatable and verifiable methods, which can be used as an input for REDD (to estimate carbon emissions).

1.1 Background

1.1.1 CO_2 emissions from deforestation and forest degradation

Global forest area decreased 3% from 4,128 Mha in 1990 to 3,999 Mha in 2015, according to the Global Forest Resources Assessment 2015 (FRA 2015) of the Food and Agriculture Organization of the United Nations (FAO, 2015a). Although the forest area expanded in Europe, North America, the Caribbean, East Asia and Western-Central Asia, it declined in Central America, South America, South and Southeast Asia and all Africa. Tropical forest area declined at a rate of 5.5 M ha y^{-1} from 2010 to 2015 (at a lower rate though than in the 1990s). At national level, some tropical countries declined the net rate of forest loss of the 1990s significantly between 2010 and 2015 (Brazil, with the highest net loss rate, declined it 60%, Indonesia's dropped by two thirds and Mexico's halved), other

thirteen tropical countries either passed transitions from net forest loss to net forest expansion, or continued with the forest expansion that follows these transitions between 1990 and 2015 (Philippines, Laos, India or Vietnam) and others had net rise in forest area between 2010 and 2015 (China, with the highest rate of expansion, 1.5 M ha y^{-1}, though only 63% of the corresponding rate in the 2000s and Chile, 301 K ha y^{-1}).

In Africa, the greatest net losses in forest area between 2010 and 2015 were in Nigeria (410 $khay^{-1}$), Tanzania (372 $khay^{-1}$), Zimbabwe (312 $khay^{-1}$) and democratic Republic of Congo (311 $khay^{-1}$). Nigeria is the second country in the world with the highest rate of population growth between 2000-10 (3174 K persons y^{-1}), and the highest in Africa followed by Tanzania (1.095 K persons y^{-1}), which has caused high rates in forest area loss during this period, the third and the forth in the world (with 625 and 595 K ha y^{-1} respectively) and the first and the second in percentage (3.7% and 0.8%) (Keenan et al., 2015).

Deforestation and forest degradation contribute to atmospheric greenhouse-gas emissions (GHGs) through combustion of forest biomass and decomposition of remaining plant material and soil carbon. Deforestation is the second largest anthropogenic source of carbon dioxide to the atmosphere, after fossil fuel combustion. Emissions from tropical deforestation and forest degradation are estimated to account for up to 14% (17% when including peat degradation) of the total anthropogenic CO_2 emissions or 20% of the total CO_2 emissions (23% with peatlands) (van der Werf et al., 2009). There are large uncertainties in these emission estimates, which arise from inadequate data on the regional rates of deforestation (Haughton et al. 2012) and the carbon density of forests (Houghton et al., 2009). Deforestation and forest degradation also contribute to increase the CO_2 atmospheric levels by reducing important carbon sinks (Pan et al. 2011). Therefore reducing emissions from deforestation and forest degradation is a cost-effective way to mitigate anthropogenic greenhouse-gas emissions and combat climate change, especially in the tropics, which hold 44% of the global forest (FAO 2011) but also the highest rate of deforestation (FAO 2016). Moreover, tropical forests are also home to 90% of the world's biodiversity and provide a livelihood for millions of people. Recent efforts have been made to improve data sets that characterize the global distribution of forest biomass (e.g. Baccini et al., 2012; Saatchi et al., 2011). However, more investment in national forest monitoring is needed to provide more accurate information and better support for international initiatives to increase sustainable forest management and reduce forest loss and carbon emissions, particularly in tropical countries.

1.1.2 The REDD+ programme

The reducing emissions from deforestation program (RED) from the United Nations Framework Convention on Climate Change (UNFCCC) appeared at the Montreal eleventh session of the Conference of the Parties (COP-11) in 2005 (UNFCC 2006) with the aim to financially reward developing countries for reducing GHGs emissions from conversion of forest lands. Initially the program was focused on deforestation.

RED concept was broadened to REDD (reducing emissions from deforestation and forest degradation) at the Bali Action Plan, formulated at the COP 13 in 2007 (UNFCCC 2008), as a response to reports in some high-level publications (e.g. Asner et al. 2005). Such research showed that large areas of forest, notably in the Amazon, were severely degraded mainly due to selective logging, and nevertheless were still classified as "forest". These logged areas were estimated to be equivalent to 60% to 123% of previously reported deforestation area for the 1999 and 2002 period. The carbon losses in these areas were not reported in the deforestation statistics compiled by national forest monitoring programs. Demonstration activities were to be reported upon two years later along with an assessment of drivers of deforestation.

The activities proposed under the REDD framework brought new requirements for monitoring deforestation and forest degradation at national levels (UNFCCC, 2011). Deforestation is defined as a direct human-induced decrease in tree crown cover below 10-30% of forest areas with a minimum size of 0.05-1 ha (UNFCC COP-7, UNFCCC, 2001), and degradation as a loss of carbon stock of forest areas with a decrease in the tree crown cover not below the 10-30% (UNFCC COP-9, PCC, 2003).

For methodological guidance to the countries, the UNFCCC stated at the Copenhagen COP-15 that REDD+ should be implemented by establishing monitoring systems which could use an appropriate combination of remote sensing and ground-based forest carbon inventory approaches, with a focus on estimating forest area changes, forest carbon stocks and anthropogenic forest-related greenhouse gas emissions by sources and removals by sink (UNFCCC, 2009, p.12).

Finally, in 2010, at COP-16 (15) the Cancun Agreements included policy approaches and positive incentives on issues relating to reducing emissions from deforestation and forest degradation in developing countries, and REDD became REDD-plus (REDD+), to reflect new components to obtain a comprehensive approach to mitigating climate change: the role of conservation of forest carbon stocks, sustainable management of forests and enhancement of forest carbon stocks.

Countries may claim for incentives submitting reports/after reporting the area changes in forest cover (activity data) and density of forest carbon stock (emission factors) from national GHG

inventories to estimate forest carbon stock change (UNFCCC 2014). To qualify for future REDD+ financial incentives, countries need moreover to assess historical forest cover changes and establish Forest Reference Emission Levels (FREL).

The establishment of a national definition of "forest" is essential to monitor changes in forest area and a pre-requisite to develop a consistent FREL monitoring system. The UNFCCC proposed that forest is an area of land with at least 0.05-1 ha and a minimum tree-crown cover of 10%-30%, with trees that have reached, or could reach, a minimum height of 2-5 m at maturity (UNFCC COP-7, UNFCCC, 2001). Inside these limits, countries can select their national forest definition for reporting purposes.

At the COP-17 of the UNFCCC (UNFCCC 2012) it was decided that to estimate GHG emissions and removals countries should use the most recent International Panel on Climate Change (IPCC) guidelines and guidance as adopted - *i.e*. (IPCC, 1996) - or encouraged - *i.e*. Good Practice Guidance for Land Use, Land-Use Change and Forestry (GPG-LULUCF) (IPCC, 2003) - by the COP (UNFCCCC, 2014). Moreover, they can also rely on the updated information from the IPCC guidelines and supplementary Methods and Good Practice Guidance from the Kyoto Protocol (IPCC, 2013).

According to such guidelines there are three approaches for assessing forest area changes and three tiers for assessing emissions factors, with increasing accuracy and precision.

For forest area changes approach 1 only describes the total area as well as change for each land use category, approach 2 additionally tracks the conversions between the single land use categories, and approach 3 finally adds the spatial component of land use conversions (IPCC 2006, GOFC-GOLD et al., 2014).

The forest area changes in approach 3 must be assessed with satellite data at minimum mapping unit of 0.05-0.5 ha, as area changes of forest to non-forest, taking into account that a forest has a minimum tree-crown cover of 10%-30% of trees that have reached, or could reach, a minimum height of 2-5 m at maturity in the same location (*i.e.* deforestation), and also area changes within the same forest class when there is a reduction of the carbon stock (*i.e*. forest degradation) .

For assessing emission factors, tier 1 approach uses representative carbon stock values per ecological zone or biome (provided by the IPCC guidelines), tier 2 approach uses country-specific field data (directly or indirectly based on repeated field sample plots), and tier 3 uses higher spatial resolution data derived from country-specific remote sensing estimates and/or detailed and repeated field sample plots, possibly associated with modelling (IPCC 2003 (chapter 3.2), IPCC 2006 (volume 4.2)).

While tier 1 data have large uncertainty ranges, tier 2 and tier 3 data are expected to be more accurate and precise (Bucki et al., 2012). One main challenge of implementing REDD+ is the trade-off between the requirement for high quality forest monitoring systems and the associated costs. At the early stages of REDD+, most developing countries will lack the financial and technical capital required to implement country-level forest monitoring (UNFCCC 2009). Therefore, a number of countries will have to rely on the IPCC tier 1 default values instead of deriving higher tier data (Romijn et al 2012). However, through the phased approach of REDD+, countries will be able to improve this data over time through an iterative process under the UNFCCC.

1.1.3 Monitoring deforestation in the tropics

Methods to assess tropical deforestation are described in a range of reviews (DeFries et al., 2002; Herold et al., 2012; Kissinger and Herold, 2012; Olander et al., 2008; GOFC-GOLD et al., 2014). Deforestation has been successfully monitored at regional and national levels using moderate spatial resolution satellite data, predominantly from the 30 m spatial resolution Landsat sensor. The Joint Research Centre (JRC) of the European Commission, under its TREES-3 project, has carried out forest cover change assessment in the tropics for the periods 1990-2000, 2000-2005 and 2005-2010 using a systematic sampling of Landsat image subsets, supporting the FAO Forest Resource Assessment (FRA) remote sensing survey (FAO et al., 2009). The sampling units of 20 km x 20 km are centred at the confluence of the geographical latitude and longitude integer degrees and cover the full tropical forest belt. They are classified at a minimum mapping unit of 5 ha in four main classes: tree cover (more than 70% of forest cover), tree cover mosaic (between 40 and 70% of forest cover), woody vegetation (shrubs and forest regrowth) and other land cover (Achard et al., 2009). The Food and Agricultural Organization (FAO) of the United Nations carried out a remote sensing exercise for supporting its five year global Forest Resource Assessment (FRA) survey based on country reports from the national forest services. The most recent FRA survey of 2010 was executed in the tropics by the JRC (FAO, 2016). Similar projects have been carried out at national and regional scales, reaching accuracies of ca. 90% with different minimum mapping units up to 5 ha using Landsat data. The Brazilian National Space Agency (INPE) has carried out an annual wall-to-wall forest mapping of the Amazon based on remote sensing since 1989 (INPE, 2015).

More recently, the University of Maryland and Google have produced global forest change maps - the High-Resolution Global Maps of 21st-century Forest Cover Change (GFMs) - based on a synthesis

of the Landsat archive for 2000 to 2012, where each pixel cell is given a tree cover percentage (Hansen et al., 2013a). Data are also provided on a yearly basis for "forest losses".

These and other attempts at measuring global forest cover and forest cover change based on Landsat, have been seen by countries as an opportunity for assessing historical forest cover changes and establishing Forest Reference Emission Levels (FREL), especially the wall-to-wall GFMs. However, these datasets remain generally too coarse for deforestation measurements according to national scale definitions and REDD+ purposes. Reporting deforestation at 0.05 ha would require analysis at the half pixel level for Landsat data – clearly a challenge. Moreover the techniques used to produce them must be adapted to country forest characteristics. In chapter 4 it is demonstrated that these forest maps produced with a global aim must be integrated with more locally relevant and appropriate data sets.

1.1.4 Defining and estimating forest degradation

Already before the introduction of forest degradation in the REDD+ mechanism, international organizations such as FAO, IPCC, International Tropical Timber Organization (ITTO), Convention on Biological Diversity (CBD) or International Union of Forest Research Organizations (IUFRO) defined forest degradation with more or less level of detail (FAO, 2007). There is not a general consensus in what forest degradation is; however, there are certain parameters which are frequently found in the different definitions. Table 1 collects these parameters and their occurrence in the definitions of forest degradation from some international organizations.

Table 1: Parameters used in the definitions of forest degradation from international organizations.

	FAO (2000)	FAO * (2001, 2002)	ITTO (2002, 2005)	IPCC (2003)	CDB (2001, 2005)	IUFRO (2000)
Structure change		✓	✓		✓	
Canopy reduction	✓ <10%			✓ <30%	✓ (fragment)	
Carbon stock reduction	✓			✓ ≥y%		
Productivity loss			✓		✓	✓
Loss of function of the stand		✓	✓		✓	
Species composition			✓		✓	

loss								
Capacity reduction to provide products/goods				✓	✓	✓	✓	
Capacity reduction to provide services				✓	✓		✓	
Biodiversity reduction					✓		✓	
Soil (fertility, compaction, salinization)							✓	✓
Cause	Human-induced	✓			✓	✓	✓	
	Natural	✓						
Time scale	Long					✓ ≥x years		
	Specified duration					✓ from time T		
Activities excluded	Deforestation	✓				✓		
	Forest management					✓		

* From the Global Forest Resources Assessment (FAO 2001) and the Second Expert Meeting on Harmonizing Forest-related Definitions for Use by Various Stakeholders (FAO 2002)

Taking into account these parameters, forest degradation could be considered in general terms as *a long-term loss of productivity in a forest, this is, the capacity to provide goods and services, which occurs with a negative change in the structure (canopy and carbon stock reduction) induced by human activities*.

Estimating forest degradation is even more challenging to measure than deforestation with the medium resolution Landsat data, as it requires detecting less perceptible changes in the forest canopy (i.e. at least changes of one pixel). Moreover, it consists not only of measuring forest cover changes, but also carbon stock changes within a forest. Therefore, degradation needs to be assessed in term of changing forest parameters at different time periods and relates only to 'within' class changes (Figure 1).

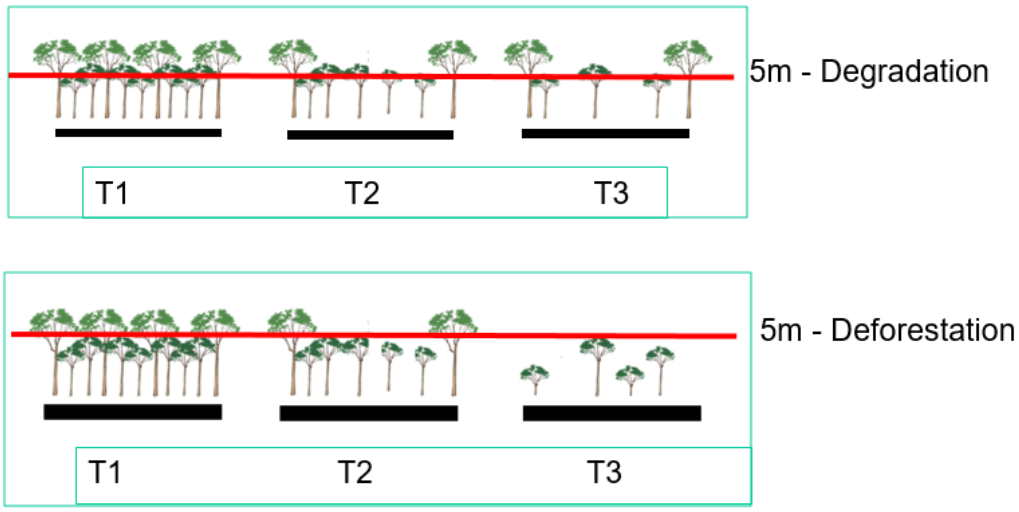

Figure 1: Example of forest degradation and deforestation process along three time periods.

The carbon stock in forest ecosystems is found in the living biomass of trees, understory vegetation, woody debris, dead mass litter and soil organic matter. The carbon stored in the aboveground living biomass of trees is typically the largest pool (except for peat swamp areas in South East Asia) and the most directly affected by deforestation and forest degradation. The aboveground living biomass is composed of three sub-pools: bole, branches and crown. The bole represents the greatest volume and has the highest carbon density, closely followed by the branches. In tropical rainforest it has been calculated that the total aboveground biomass of a tree can be estimated with high accuracy as an allometric equation with the bole volume as single variable (Eckert, 2012). In turn, the bole volume can be estimated with allometric equations with field data of tree height and basal area (essentially the cross section of a tree at a height of 1.3 m, approximately the height of an adult's breast) for forest types, or more generally, with the 50% of the cylinder which results from multiplying the basal area and the total height (Henry et al., 2011). Belowground carbon stock can be estimated as a proportion (often 20%) of the aboveground stock (Santantonio et al., 1977). Therefore, to estimate the aboveground biomass of a forest, apart from the area covered by woody vegetation, one needs to know the volume of the trunks, which can be estimated from the diameter at breast height (DBH) and the height of the trees. These two variables are generally correlated for a single type of forest, so sometimes one of them can be enough. After that, the carbon content of vegetation biomass is quite constant across a wide variety of species; it is almost always found to be between 45 and 50% by oven-dry mass (Magnussen and Reed, 2004).

1.1.5 Monitoring forest degradation

Forest degradation can be caused by human activities (timber extraction (cutting), selective logging, debarking, charcoal production, fuelwood collection, agriculture, grazing) or natural factors (biotic agents -as insects, fungus, diseases- and abiotic agents -as wind, fire).

A number of different techniques can be employed to map forest degradation and have been reviewed by (Miettinen et al., 2014) (Mitchell et al., 2017);

- Spectral unmixing of medium resolutions satellite data (Asner, 2005; Souza, 2003)
- Mapping of intact / non intact forests (Potapov et al., 2008)
- Analysis of a sample of very high resolution satellite images (Bey et al., 2016)
- A calibration of medium resolution satellite data using a sample of very high resolution data (Tyukavina et al., 2013)
- Field survey – a statistical sample of field plots that are revisited at specific periods (MNRT, 2015)

1.2 The Tanzania REDD+ initiative

The number of REDD-related country-led initiatives has increased in the last years. Tanzania was one of the sixteen pilot countries which received direct support by the UN to carry out a national REDD+ initiative (UN-REDD et al., 2009), with a number of strategies developed for the in-country implementation (UN-REDD, 2012, 2013a; Zahabu et al., 2012).

1.2.1 Forest in Tanzania

Tanzania has a total area of 945 090 km^2. The country is covered by four main ecosystems (White, 1983) -the Zambezian ecoregion, where woodland (miombo) is the predominant and characteristic vegetation type, followed by the Zanzibar-Inhambane ecoregion, where denser coastal forests were predominant; the Somali Masai region with more arid shrublands, and the Afromontane region, where again, denser forest types prevail (Figure 2).

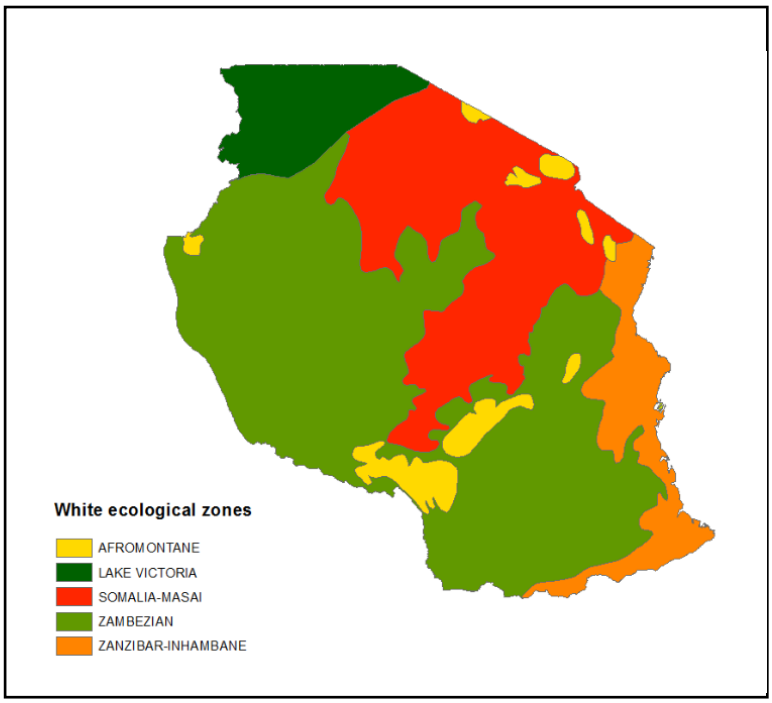

Figure 2: The ecological zone map of Tanzania [White 1983].

Forests and woodlands are key elements in the livelihoods of hundreds of thousands of people across Tanzania. Forest covers an estimated 33.4 million hectares (ha), around 38% of the country. Almost all of these forests are naturally regenerated, with only 1% of the forest cover being considered as ´primary´. Shrub lands and young forest regrowth account for over 11 million ha (13% of the country) (MNRT, 2015).

All forests in Tanzania come under public ownership. However the public administration directly manages only 37% of these forests. The majority (85%) of Tanzania´s forests have a management plan, 6% are within protected areas, and 39% are classed as permanent forest estates. Forest plantations increased from 150 000 ha in 1990 to 240 000 ha in 2010, but remain a small proportion (1%) of the total forest cover (MNRT, 2015).

Forests provide a wide range of benefits, both directly in the form of timber production (70%) (Burgess and Mbwana, 2000) and multiple uses (24%) such as forage, fruits, honey and beeswax, charcoal, traditional medicines, gums and resins, and indirectly, contributing to biodiversity preservation through protected areas, regulation of water catchments, control of erosion, nutrient cycling and maintenance of local climates.

1.2.2 Deforestation and forest degradation in Tanzania

Deforestation statistics for the country have been produced from a series of studies. The TREES-3 study on measuring deforestation in the tropics (1990-2000-2010) has a statistical sampling designed for estimation of forest changes (deforestation, fragmentation, loss of tree cover and loss of woody vegetation) at continental level (Achard et al., 2014). The 2010-2015 period was assessed for Tanzania for this study in conjunction with the Tanzania Forest Service under the MESA (Monitoring of Environment and Security for Africa) program (Eva et al., 2016). While these studies were designed for regional studies, for large countries such as Tanzania, they can give an indication of the forest dynamics in relation to loss of carbon stock in forest and woody vegetation areas. The Tanzania Forest Service has been involved in a number of different exercises to determine deforestation rates. However, there is wide variance in the results (Table 2) – partly coming from the different techniques employed, and partly from the different definitions of forest (Hojas Gascón et al., 2015). Such differences need to be resolved, so as to give confidence to policy makers and potential funding agencies as to the real extend and magnitude of forest changes. This is especially true for the much publicized Global Forest Change maps (Hansen et al., 2013a), which since being released have been shown to have major limitations for quantitative studies at the national level and specifically for Tanzania (Hojas Gascón et al., 2015, Tropek et al 2014)

Table 2: Estimates of forest change for Tanzania from different sources.

Source	Method	Forest change ha/year
Tanzania's Forest Reference Emission Level Submission to the UNFCCC (The United Republic of Tanzania, 2016)	Wall to wall Remote sensing based,	582,427
NAFORMA Main Report (MNRT, 2015)	Difference between two different land cover maps	621,687
NAFORMA Remote Sensing (Ortmann et al., 2015)	Landsat images for 860 sample sites	81,000
Global Forest Map adjusted (Ortmann et al., 2015)	Wall to wall annual Landsat data	122,000
Global Forest Map non-adjusted (Hansen et al., 2013a; Hojas Gascón et al., 2015)	Tree Cover (>75%) (>40%) (>10%)	 9,770 78,530 155,720
Joint Research Centre (Eva et al., 2016)	Sample of 76 Landsat images	190,000
Rep. of Tanzania to FRA 2015 (FAO, 2015b)		504,000

Analysis carried out in this study of the data produced by the TREES-3 project (Achard et al., 2014) shows that in Tanzania between 1990 and 2000 the main pattern of tree cover loss was forest fragmentation, this is, the conversion of tree cover into tree cover mosaic. In practical terms this means that forest areas of a minimum size of 5 ha with more than 70% of tree cover have been converted to areas with 40-70% of tree cover. Figure 3 depicts the percentages of forest fragmented area in the TREES-3 sites over the White vegetation map. The Zambezian ecoregion, where woodland is the predominant and characteristic vegetation type, is the most affected by forest fragmentation, followed by the Zanzibar-Inhambane ecoregion. Together with forest fragmentation, loss of woody vegetation was the major change during the decade.

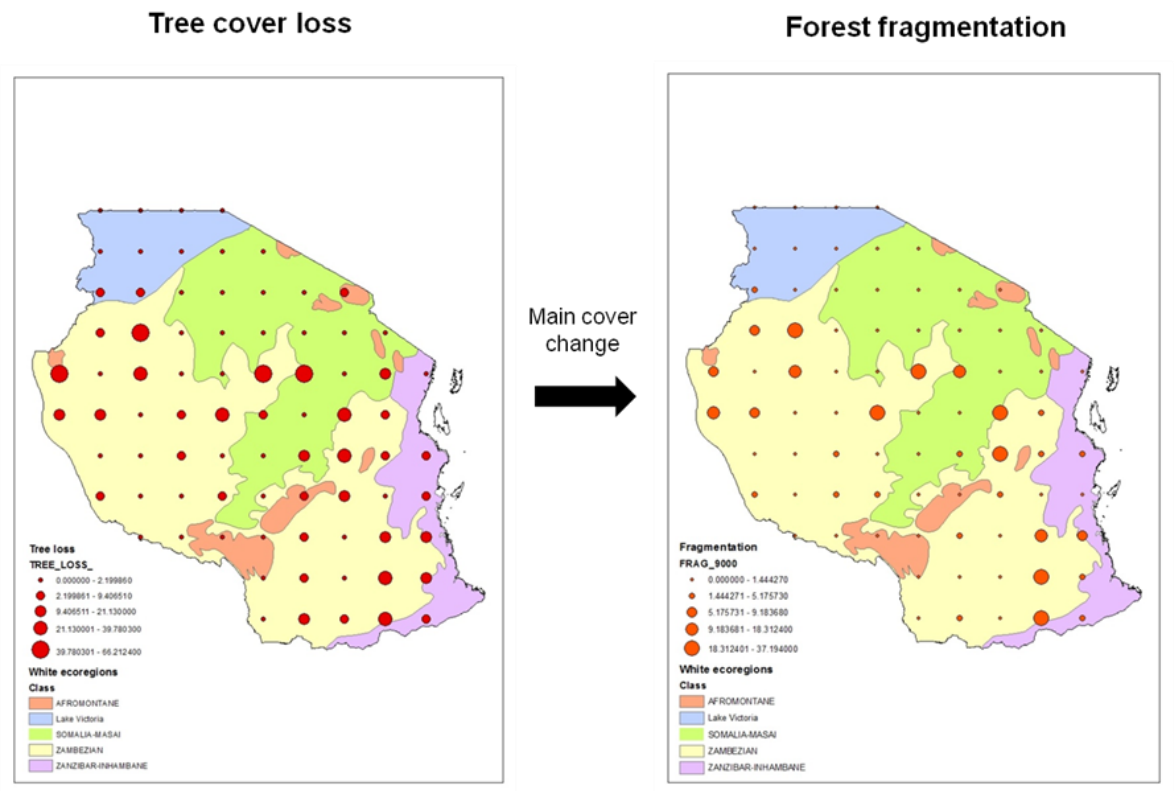

Figure 3: Estimation of tree cover loss (left) and forest fragmentation (right) from 1990 to 2000 from the TREES-3 project overlaid on the White vegetation map. The size of the circles represents the percentage of tree cover area loss and tree cover area converted into tree mosaic.

Wood extraction is the main driver of forest degradation in Tanzania, mainly for energy consumption, in the forms of fuel wood and charcoal. Fuelwood and charcoal account for over 75% of the total energy use, mainly for household cooking. More than 80% of urban population consumes charcoal and an increment of pressure on wood resources is predicted due to urban population growth; the current population of 45M is set to double by 2050 (Mwampamba, 2007).

Charcoal production is thought to be responsible for the degradation of 29,268 hectares (24.6 %) of closed woodland and deforestation of 23,308 hectares (19.58 %) of closed woodland and 92,761 hectares (50.8%) of open woodland in the catchment area to the west and north of Dar es Salaam that supplied charcoal to Dar es Salaam City (Malimbwi et al., 2005).

Charcoal dealers and transporters are categorized mainly in two groups: the commercial dealers who include the cyclists who operate within a radius of 60 kms from the city and those who use vehicles to transport more than 10 bags; and non-commercial transporters who use vehicles to transport less than 10 bags for private uses. Of the estimated amount of over 6500 bags of charcoal entering Dar es Salaam city every day, 84% were transported by commercial vehicles, 11% by bicycles and only 5% by non-commercial vehicles. Although vehicles transport most of the charcoal,

transportation by bicycles employs the largest number of people (Malimbwi et al., 2005). Chapter 8 analyses the pressure of urban population growth on wood resources around the city, and the impacts on the integrity of protected areas.

1.2.3 The National Forest Monitoring and Assessment program (NAFORMA)

In 2010, the National Forest Monitoring and Assessment program (NAFORMA), set up by the Ministry of Natural Resources of Tanzania (MNRT) and UN-REDD, with the technical support of FAO-Finland, was created with the aim of becoming the national framework for assessing and monitoring forest carbon pools to support the national carbon accounting system compatible with REDD+ requirements (MNRT, 2010).

The program has three components represented in Figure 4:

1) **Forest inventory and forest cover mapping**. NAFORMA produced a national forest inventory, combining field survey with remotely sensed data of medium spatial resolution, to better understand the current status of forest and Trees Outside Forests (TOF) of Tanzania. With the field information a wall-to-wall forest cover map of 2010 from Landsat data was also produced.
2) **Historical forest cover change assessments**. Forest area change assessments for the periods 1990-2000-2005-2010 have been done at regional level using Landsat satellite data by JRC, in collaborative partnership with the Remote Sensing Survey (RSS) of the FAO Forest Resources Assessment (FRA) programme. The cover changes were estimated and validated with the JRC Validation Tool and based on a systematic sampling of 10x10km boxes situated at the UTM confluence grid. For a national estimation, NAFORMA decided to estimate forest area and carbon stock changes for the periods 1980-1990-2000 repeating the JRC-FRA operational system (processing, classification, change assessment and validation) on some 860 5 km x 5 km boxes covering the field sample clusters of the national forest inventory (Ortmann et al., 2015).
3) **Future forest cover and forest-based biomass change assessments**. Biomass changes from deforestation and forest degradation will be estimated through field re-measurements of permanent sample plots and forest cover change re-assessments of the 5 km x 5 km boxes. The field data have been used by JRC also for calibration and validation of methods for evaluating forest degradation with satellite data of higher spatial resolution. The forest area

and carbon pool change estimates will be used to support both national sustainable forest management and monitoring for the future carbon market.

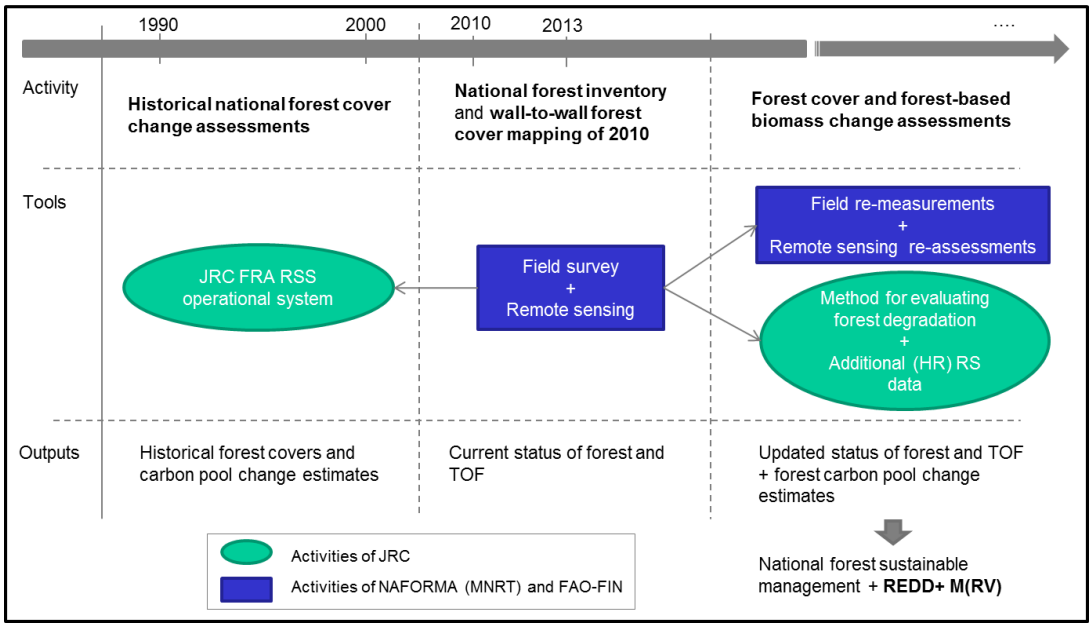

Figure 4: Overview of the activities of NAFORMA carried out by JRC, MNRT and FAO-FIN.

1.2.4 NAFORMA's field survey

The NAFORMA field inventory employed a stratified sampling based on biomass density (predicted growing stock) and work efficiency (distance and time planning according to roads, foot paths and topography) (Tomppo et al., 2010; Tomppo et al., 2014). It counts for 3419 clusters, each cluster has between 6 and 10 plots in an L shape, and in total there are 32,660 plots of 15m radius across the country (see Figure 5). From each plot very extensive data on tree biophysical information was collected between 2010 and 2013, among others, data on canopy cover, tree height, trunk diameter, species, dead wood and soil. It was designed to estimate forest area and biomass at national level. Around 25% of the plots were selected as 'permanent plots' and are planned to be re-measured in the future every five years (NAFORMA, 2010a).

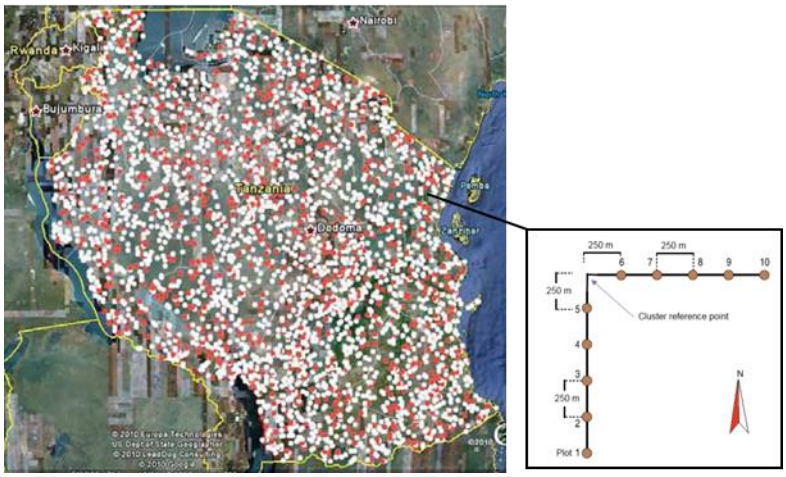

Figure 5: NAFORMA cluster distribution (red are selected as 'permanents') and example of plot distribution in a cluster.

1.2.5 Tanzania forest definition

The establishment of a national forest definition was essential to produce the wall-to-wall land cover map of 2010 from the field data and for assessments of forest area changes with a consistent monitoring system. Moreover it is necessary that it meets the thresholds establish by the UNFCC for Tanzania to participate in the REDD+ mechanism. There are several existing forest definitions that have been used in Tanzania described a continuation.

Initially, the national definition proposed by the MNRT to the UNFCC was 'an area of land with at least 0.05 hectares, with a minimum tree crown cover of 10% (or with existing tree species planted or natural having the potential of attaining more than 10% of crown cover), and with trees which have the potential or have reached a minimum height of 2 meters at maturity in situ' (UN-REDD, 2013b). The following points were considered in setting out the thresholds in this definition:

- Tanzania has vast land with less than 10% tree crown cover suitable for forest development under Clean Development Mechanisms (CDM). These are emission-reduction projects in developing countries implemented by countries with emission-reduction or emission-limitation commitment under the Kyoto Protocol.
- Large area of reserved forests with low cover is degraded with potential for regeneration. If Tanzania adopts a more than 10% forest cover, most of the degraded forest reserves will be re-categorized into other land uses, with no potential for forest enhancement.
- There is a high proportion of dry forest with short trees (e.g. Itigi thicket).

- Most household in Tanzania own small portions of land from 0 to 1 ha. A 10% forest cover in the definition will enable most of the small holdings eligible for REDD+ and improve forest carbon stock, biodiversity and soil conservation.
- It considers the interest of many stakeholders, the potential to involve many people, especially local communities and private sector.

The forest definition adopted by NAFORMA and the Tanzanian national REDD+ strategy was an 'area of land with at least 0.5 hectares, a minimum tree crown cover of 10% (or with tree species, planted or natural, having the potential of attaining more than 10% of crown cover) with trees which have the potential or have reached a minimum height of 5 meters at maturity in situ' (The United Republic of Tanzania, 2016).

The definition used by JRC for forest degradation studies, was adapted to the remote sensing methods (see chapter 7), and had a minimum land unit of 1ha, minimum crown cover of 30% of trees with a minimum height of 3 meters (dry miombo and semi-arid steppe woodlands hardly reach 3 m). This definition affects the scale at which forest degradation should be measured and the forest area and carbon change estimations. According to it, degradation occurs when there is a loss of carbon stock with a decrease in the canopy cover not below the 10-30% of forest areas with a minimum size of 1ha.

1.3 Monitoring deforestation and forests degradation with remote sensing in Tanzania

Field surveys to estimate carbon stocks at the national level, like the one done by NAFORMA, are very expensive and time consuming due to the amount of data required. Moreover, most of the forest inventories done in combination with remote sensing data used 30m spatial resolution Landsat imagery, which is too coarse to measure forest degradation at the scale required for REDD+ (Hojas Gascón et al., 2015). Fortunately, it has been demonstrated that carbon density can be estimated as a function of remote sensing parameters derived from very high spatial resolution satellite data. For example, tree biomass variables from field data showed significant correlations with crown diameter derived from Quickbird (Gonzalez et al., 2010), or with spectral and textural parameters, as shown in Chapter 4 (Hojas Gascón et al., 2014).

1.3.1 Satellite data to assess deforestation and forest degradation

Very high spatial resolution (approx. 1m) data, such as Quickbird, IKONOS, WordView-2, Geoeye, can detect and quantify forest degradation at small scale assessments, although they need limited field data to estimate biomass accurately (see Chapter 5). As they are expensive and cover small areas, they cannot be considered for a countrywide survey, although they are good for calibration and validation of large scale assessments. For this purpose JRC acquired 12 targeted images of 7km by 7km across Tanzania.

Satellites with high spatial resolution, like SPOT (20m), AVNIR (10m) or RapidEye (5m) can detect forest clearing of less than 0.5ha and identify some types of forest degradation, although they need supplement data or field inventories to estimate biomass. They have still small area coverage and high cost to cover a country. The coverage of the full country of Tanzania with RapidEye data would be too expensive (around 650.000 €). Therefore this study will adopted a systematic sampling at country level with RapidEye. RapidEye satellite constellation offers a good combination of large-area coverage, frequent revisit intervals, and multispectral capabilities at high spatial resolution.

To meet REDD+ monitoring requirements a multi-spatial resolution approach will be required. RapidEye need to be complemented with other remote sensing sources. Landsat data (30m) cannot detect small forest clearings and forest degradation according to REDD+. However, they are freely available for the years 1975, 1990 and 2000 globally orthorectified (with an automatic pre-processing) and so they can served for country scale historical assessments to identify forest

clearings of at least 5 ha. Low spatial resolution (>250m) satellite data, like MERIS or MODIS, with higher temporal acquisitions, allow analysis of time series to monitor inter-annual variations and seasonality. These can improve vegetation classifications, especially in dry areas, and identify time, cause and extent of regional forest degradation. RapidEye and Landsat have cloud coverage problem in humid areas, radar data (mosaics of 100m), like ALOS PALSAR, can penetrate clouds and estimate biomass. Their drawback are a limited spectral radiometry, difficult processing, layover in terrain and a limitation of 100 biomass tonnes per ha.

In the future the new Sentinels satellites, with a good balance of high spatial and radiometric resolution (10m), wide swath, high revisit frequency, will allow improving forest degradation at national level, allowing temporal analysis and with an open data access policy.

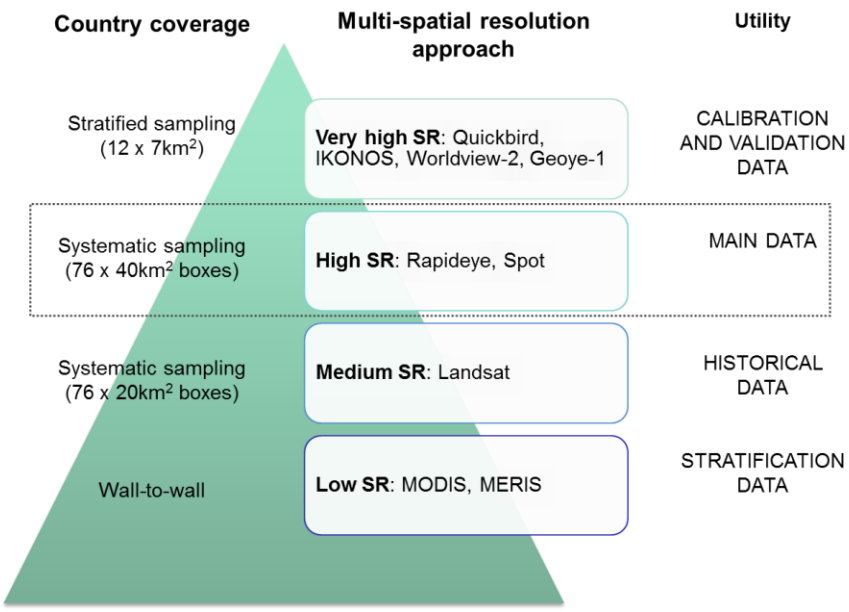

Figure 6: Multi-spatial resolution satellite data set for forest degradation study in Tanzania.

RapidEye

Under the European Union's GMES (Global Monitoring for Environment and Security) program, 419 RapidEye 5 m spatial resolution images were obtained over Tanzania acquired during the year 2010. These images were selected to cover a 40 km by 40 km box around each of the 76 confluence points of the latitude-longitude grid in Tanzania. This sampling system formed the basis of the FAO FRA JRC 1990-2000-2010 forest change analysis (FAO et al., 2009). RapidEye data come from a constellation of radiometrically intercalibrated satellites, each with a swath width of approximately 500 km, imaging data in 5 spectral channels, from blue to near infra-red, including a 'red edge' channel

(Table 3). The data revisit period is a nominal of 1 day on request (Tyc et al., 2005), however, the effective acquisition is limited due to recording and receiving capacities along with priority demands for commercial acquisitions. This is especially true for Africa, where data limitations from the RapidEye system can occur due to the demand for images over European agricultural zone, making temporal analysis difficult.

Table 3: Spectral channels of the RapidEye sensor

Channel	Spectral band	Spectral range (nm)
1	Blue	440–510
2	Green	520–590
3	Red	630–685
4	Red Edge	690–730
5	Near infra-red	760–850

The Red Edge band is spectrally located between the Red band and the Near-Infrared (NIR) band, covering the portion of the spectrum where reflectance of green vegetation drastically increases from the red portion (where chlorophyll strongly absorbs light) towards the NIR plateau (where the leaf cell structure strongly reflect it) (green curve in Figure 7). Therefore, variations in both the chlorophyll content and the leaf structure are often reflected in the inflection point of the Red Edge band. Several studies have suggested that the transition between the red absorbance and the NIR reflection is able to provide additional information about vegetation and its characteristics, in order to identify plant types, nutrition and health status, and characterize plant cover and abundance, among other feature. The Red Edge band has been typical used in very narrow spectral bands with success in determining green biomass and Chl content. Broad Red Edge band, as used in the RapidEye sensors, is also suitable for obtaining information about the Chl and N content of plants (Eitel et al. 2007), but with the possibility of monitoring extensive areas at high temporal frequency (Weichelt et al., n.d.).

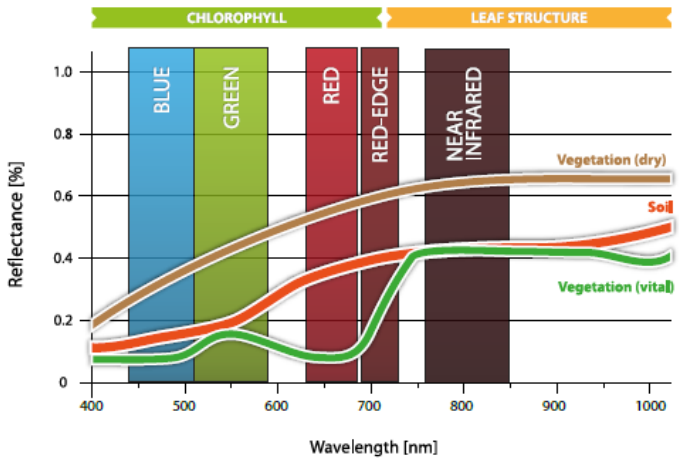

Figure 7: Typical spectral reflectance curves of dry vegetation, soil and vital vegetation in relation to the RapidEye spectral bands (Image from (Weichelt et al., n.d.).

The best (cloud and artifact free) images from RapidEye satellites 3, 4 and 5 were selected and mosaicked using all channels to cover 20 km by 20 km boxes centred at each confluence point (see distribution in Figure 8). The reduction from the 40 km by 40 km database available was undertaken to ensure 100% coverage of all boxes.

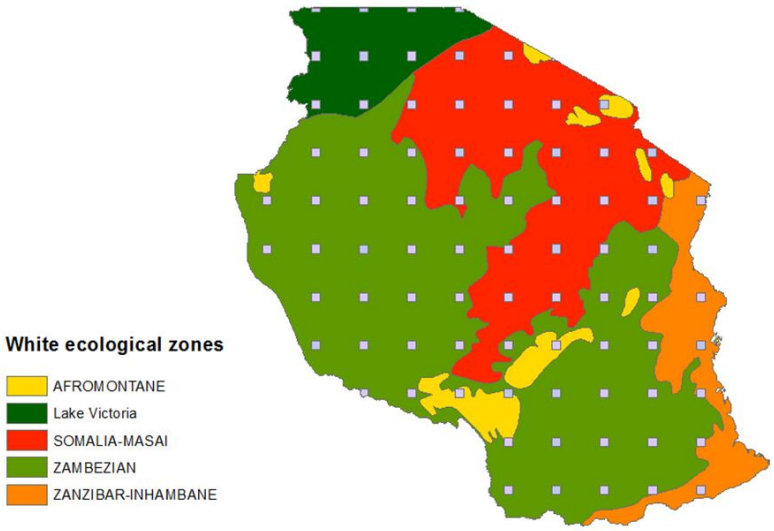

Figure 8 Location of the 20 km by 20 km sites for which field plots and RapidEye data were obtained over the ecological zone map of Tanzania

The Sentinel-2 program

The European Union's first Earth Observation programme, Copernicus, is building a series of technologically advanced satellites (the Sentinels), which includes the Sentinel-2 satellites. Sentinel-

2A and B aim to provide inputs for services relying on multi-spectral high-resolution optical observations over global land surfaces, building on the heritage of the SPOT and Landsat satellites, whilst adding new components to cover current limitations, such as higher revisit frequencies, more spectral bands with narrower bandwidths, finer spatial resolution and better radiometric resolution, in order to improve services as vegetation monitoring (ESA, 2010).

The design of the Sentinel-2 platform benefited from the experience and lessons learned from other satellites building on their technology. The selection of the spectral bands has been guided by the Landsat, SPOT-5, MERIS and MODIS heritage (ESA, 2010). The Multi Spectral Instrument (MSI)'s 13 spectral bands' centre ranges from 0.433 to 2.19µm. There are four visible and near-infrared bands at 10 m spatial resolution, three red edge, one near-infrared and two SWIR at 20 m, and three channels to help in atmospheric correction and cloud screening at 60 m (Drusch et al., 2012). Sentinel-2A was launched in June 2015 and Sentinel-2B in March 2017. The two satellites offset in orbit operating simultaneously on opposite sides, each carrying the same instruments, providing coverage every five days at the equator with a 290 km field of view (ESA, 2010).

Forest monitoring is one of the priority services of the Copernicus programme for which Sentinel-2 has been tailored. In fact, the revisit requirements were driven by vegetation monitoring, for which those of Landsat and SPOT were not enough. Sentinel-2 observations are explicitly intended to develop key inputs required for REDD+ reporting; potentially, they could contribute to the Baseline Mapping Service (ESA, 2010). The Copernicus plans also aim for multiple global acquisitions, and like the Landsat programme will adopt a free and open access data policy (European Commission, 2013).

More access to satellite data of high spatial resolution are needed in order to use remote sensing techniques for monitoring deforestation and forest degradation for REDD+ requirements. In this line, the Sentinel-2 program from ESA and EU can provide 10-m spatial resolution satellite data every five days with an open data access policy. The potential of Sentinel-2 like data for forest degradation mapping is assessed in Chapter 6.

1.3.2 Satellite-based indices for monitoring forest degradation

In a forest degradation process, there are some visible changes which are appreciated on satellite images. The reduction of the carbon stock normally is linked to a reduction of the crown cover or canopy density (fragmentation or forest thinning, but the latter is not detectable by remote sensing). Usually it starts with a loss of the biggest and oldest trees and most valuable species for fuel wood and biodiversity. This implies a consequent change in the forest layer composition (i.e. conversion into a single layer of small trees), species composition (i.e. loss of primary species) and size of the canopies. These changes in the structure and productivity can be captured by satellite images as changes in the spectral properties (e.g. reduction of dark green vegetation), and if they have enough spatial resolution, also in the texture (usually homogenization).

Figure 9 shows an example of the steps in a common degradation process in a humid forest as seen on a satellite image of very high spatial resolution. Firstly a canopy opening occurs leaving an open forest (crown shadows from big trees are replaced by holds), secondly canopy homogenization into closed shrub (vegetation regenerates after medium and small trees are taken out), and if degradation continuous, shrub opening with herbaceous layer.

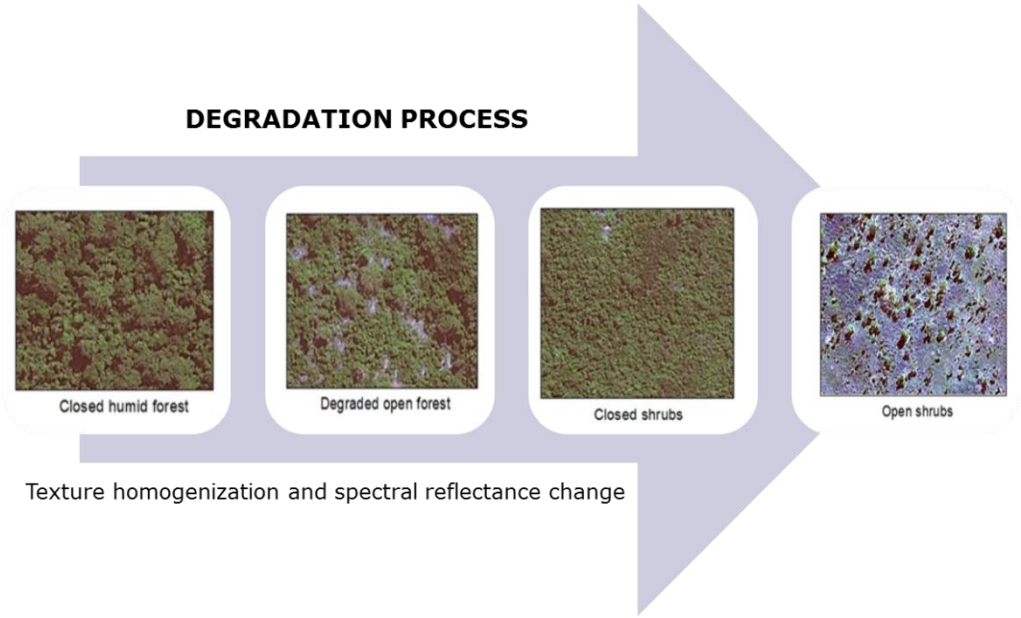

Figure 9: Example of the steps in degradation processes in a humid forest captured by a Worldview-2 satellite image.

Image reflectance information has been usually used to detect vegetation and more specifically forest parameters from satellite data. Either indices based on single spectral band information (reflectance, standard deviation), simple spectral band relations (reflectance to standard deviation ratio) and/or more complex vegetation indices calculated at pixel or pixel cluster level have been

employed. This type of information have achieved moderate success in tropical and subtropical regions for biomass estimation, where biomass levels are high, forest canopies are closed, with multiple layers, and there is a great diversity of species (Lu, 2005). According to some studies with high spatial resolution satellite data (Sarker and Nichol, 2011), texture measures correlate better with biomass than reflectance parameters, especially for degraded forests. A combination of both can improve the results (Eckert, 2012). Table 4 collects the list of reflectance and textural parameters and indices taken into account in this study.

Table 4: Range of remote sensing spectral parameters and derived indices that can be employed in remote sensing

Spectral parameters and indices	Formula or measure
Single band or band relations	
Band mean reflectance	Band # mean (DN)/6400
Band standard deviation	Band # SD
Band ratio of mean to standard deviation	[Band # mean]/[Band # SD]
Vegetation indices	
Intrinsic	
Red Vegetation Index (RVI) or Simple Ratio (SR)	NIR/Red
Green Vegetation Index (GVI)	NIR/Green
Green Red Vegetation Index (GRVI) or Greeness Index (GI)	Green/Red
Simple ratio (SRre)	NIR/Redge
Normalised Difference Vegetation Index (NDVI)	$(\mathrm{NIR}-\mathrm{Red})/(\mathrm{NIR}+\mathrm{Red})$
Green NDVI (GNDVI)	$(\mathrm{NIR}-\mathrm{Green})/(\mathrm{NIR}+\mathrm{Green})$
NDVI-RE	$(\mathrm{Redge}-\mathrm{Red})/(\mathrm{Redge}+\mathrm{Red})$
NDVIre	$(\mathrm{NIR}-\mathrm{Redge})/(\mathrm{NIR}+\mathrm{Redge})$
Modified Chlorophyll Absorption Ratio Index (MCARI)	(([Redge]-[Red])-0.2*([Redge]-[Green]))*([Redge]/[Red])
Modified Triangular Vegetation Index (MTVI2)	(1.5*(1.2*([NIR]-[Green])-2.5*(Red]-[Green])))/((2*[NIR]+1)^2-(6*[NIR]-5*([Red])^0.5)-0.5)^0.5
Soil-line related	
Soil Adjusted Vegetation Index (SAVI)	$(\mathrm{NIR}-\mathrm{Red})/((\mathrm{NIR}+\mathrm{Red}+\mathrm{L}))*(1+\mathrm{L})$
Optimized SAVI (OSAVI)	(1+5)*(NIR-Red)/(NIR+Red+0.16)
Atmospherically adjusted	
Enhance Vegetation Index (EVI)	G*((NIR-Red))/((NIR+C1*Red-C2*Blue+L))
Bare Soil RE	((([Redge]+[Red])-([NIR]+[Blue]))/(([Redge]+[Red])+([NIR]+[Blue])))+1

Shadow R	((1-[Blue])*(1-[Green])*(1-[Red]))^.33
GLCM texture features	1) In all directions, 0°, 45°, 90°, 135° 2) For all bands, band 1, band 2, band 3, band 4, band 5
Homogeneity	Local homogeneity
Contrast	Local variation
Dissimilarity	Local contrast
Angular 2nd moment	Local homogeneity
Entropy	Local variation
Mean	Average in terms of GLCM
Variance	Dispersion of values around the mean
Correlation	Linear dependency of grey levels of neighbouring pixels
GLDV texture	Angular 2nd moment, Entropy, Mean, Contrast

EVI: L is the canopy background adjustment that addresses non-linear differential NIR and red radiant transfer through a canopy; C1 and C2 are the coefficients of the aerosol resistance term, which uses the blue band to correct for aerosol influences in the red band. The coefficients adopted in the MODIS-EVI algorithm are L=1, C1=6, C2=7.5, and G (gain factor) = 2.5.

SAVI: L is the soil brightness correction factor. The value of L varies by the amount or cover of green vegetation: in very high vegetated regions L=0 and in areas with no green vegetation L=1. Generally, an L=0.5 works well in most situations and is the default used value. When L=0, then SAVI = NDVI.

Vegetation indices

Numerous studies have focused on developing and evaluating spectral indices in terms of their sensitivity to vegetation biophysical parameters, as well as to external factors affecting canopy reflectance. The vegetation indices can be broadly grouped into three categories: 1) intrinsic indices, 2) soil-line related indices, and 3) atmospherically adjusted indices (Rondeaux et al., 1996; Baret and Guyot, 1991).

Intrinsic indices rely solely on the measured spectral reflectance data and include ratios of two or more bands in the visible and near-infrared wavelengths. They are often very sensitive to background reflectance properties and difficult to interpret at low leaf area indices (Rondeaux et al., 1996). Soil-line vegetation indices use coefficients based on the generally linear relationship between near-infrared and visible reflectance for bare soil to reduce the influence of the soil on the canopy reflectance. Atmospherically adjusted indices generally incorporate the blue band into the equation to normalize temporal and spatial variations in atmospheric aerosol content to reduce atmospheric effects.

Simplest intrinsic vegetation indices use ratios of red, green and NIR spectral bands, like the Red Vegetation Index (RVI), which minimized background reflectance contributions, Green Vegetation Index (GVI), which responded more to leaf chlorophyll concentrations (Daughtry et al., 2000), Green Red Vegetation Index (GRVI) or Greeness Index (GI) (Eitel 2007), or rations of band differentials, such

the Normalized Differential Vegetation index (NDVI) (Rouse et al. 1974 in Kross 2014) or the Green NDVI (GNDVI) (Gitelson et al. 1996 in Kross 2014).

The utility of the **Red-Edge** band (sensitive to both Chl content and leaf and canopy structure) for estimating foliar chlorophyll concentrations has been demonstrated in many studies, mostly using field spectroscopy (Ramoelo 2010). Some studies have used the Red Edge band of RapidEye in vegetation indices to improve vegetation classifications, such as SRre (Simple Ratio Red Edge) (in Kross 2014, from Gitelson and Merzlyak 1994, and Ramoelo 2012), NDVIre (Red Edge Normalized Differential Vegetation index) (in Tapsall 2010 from Wu et al., 2009, and Kross 2014 from Gitelson 94) and NDVI-RE (Schuster 2012). The Red Edge NDVI (which uses the Red Edge band instead of the Red band) reported lesser saturation comparing to the traditional NDVI over highly vegetated (forested) regions (Haboudane et al. 2002, Vinal and Gitelson 2005, Tapsall 2010). A combination of NDVI and NDRE was able to predict Chl concentration in wheat with a coefficient of determination of R^2=0.77 (Schelling 2010) and NDRE was able to detect stress symptoms at an earlier stage than the other vegetation indices Eitel et al. (2011). Given the sensitivity of the NDRE, this index can be used for a variety of forest health applications.

The **atmospheric adjusted** Enhance Vegetation Index (EVI) has proved to be sensitive to canopy variations in the tropics. It is an 'optimized' index designed to enhance the vegetation signal in regions with high biomass and improve vegetation monitoring through a decoupling of the canopy background signal and a reduction in atmosphere influences. Whereas the NDVI (Normalized Difference Vegetation Index) is chlorophyll sensitive, the EVI is more responsive to canopy structural variations, including leaf area index (LAI), canopy type, plant physiognomy and canopy architecture. The two indices complement each other in global vegetation studies and improve the detection of vegetation changes and extraction of canopy biophysical parameters (Huete 1997, Solano et al., 2010).

In areas where vegetation cover is low (i.e. <40%) and the soil surface is exposed, the reflectance of light in the red and near infrared spectrum can influence the vegetation index values. This is especially problematic when comparisons are being made across different soil types that may reflect different amounts of light in the red and near infrared wavelengths (i.e. soils with different brightness values). The **soil-line related** SAVI (Soil Adjusted Vegetation Index) was developed as a modification of the NDVI to correct for the influence of soil brightness when vegetation cover is low. The SAVI is structured similar to the NDVI but with the addition of a 'soil brightness correction factor', is less sensitive to soil reflectance at low LAI than NDVI (Huete, 1988).

The use of the **both types of indices** can improve vegetation classification and vegetation change monitoring in large scale studies which include diverse biomes and vegetation conditions (e.g. EVI useful for the non-degraded and humid areas and SAVI for the degraded or dry areas), as we will see in Chapter 4.

In fact some studies demonstrated that pairs of spectral vegetation indices can estimate leaf chlorophyll concentrations with minimal confounding effects due to LAI and background reflectance, such as MCARI (Modified Chlorophyll Absorption Ratio Index), sensitive to chlorophyll concentration and OSAVI (Optimized SAVI) to minimize background reflectance contribution (Daughtry et al., 2000). Others modified existing indices to create new indices, like MTVI2 (second Modified Triangular Vegetation Index) from MTVI, both less sensitive to chlorophyll content variations and to green LAI, proving to be the best predictors of green LAI (Haboudane et al., 2004) . These indexes have been evaluated using RapidEye bands (Eitel et al., 2007), benefiting from the Red Edge band, where a new combined index derived from MCARI and MTVI2 in ratio, the MCARI-MTVI2, obtained the best results to estimate chlorophyll in dryland agriculture.

Haralick texture features

The texture of a forest can be appreciated in satellite data of high spatial resolution (already in 5 meters). To evaluate the texture of image land units there are two types of texture measures: texture features based on analysis of sub-objects, useful for highly textured data, and texture features based on the grey level co-occurrence matrix after Haralick (Haralick et al., 1973a). Due to the small area of land unit required for this study, the texture features after Haralick were employed.

The grey level co-occurrence matrix (GLCM), is a tabulation of how often different combinations of pixel grey levels occur in an image object. A different co-occurrence matrix exists for each spatial relationship. The feature depends on the direction of concern (all directions or 0°, 45°, 90° and 135° directions) and the spectral bands taken into account (all or each one of the spectral bands).

There are eight types of Haralick features, which measure similar but varying aspects of texture, called GLCM Homogeneity, Contrast, Dissimilarity, Angular 2nd Moment, Entropy, Mean, Standard deviation and Correlation. There are over 360 textural parameters based on Haralick that can be produced for a 5 spectral band satellite although many of these are highly inter-correlated.

1.4 Other studies to monitor forest degradation in Tanzania

There are other studies for monitoring forest cover and biomass currently being carried out in Tanzania. They use LIDAR data and/or a multi-source approach with medium and very high resolution optical data.

1) The LIDAR study is carried out by the Sokoine University of Agriculture with close collaboration with NAFORMA and FAO. The study will take place in two pre-selected study areas: Liwale district (LIDAR sampling) and Amani Nature Reserve (wall-to-wall scanning).

2) A multi-source approach is undertaken by Tuomas Hämes from VTT (Technical Research Centre of Finland) as part of the Recover project. Summary of his technique: Samples of very high resolution image data, better than one meter spatial resolution (Kompsat and Quickbird-2), are combined with wall-to-wall AVNIR and ALOS/PalSAR data to measure forest cover and biomass. It used a statistical sampling approach. Conclusion: it is possible to separate degraded forest from intact forest on the plots with VHR, and degraded forest from natural forest. Also closure estimation was good. Additional radar data is suggested for the cloudy areas.

2. Objectives

This thesis looks at how remote sensing technology, images acquired from earth orbiting satellites, can support countries in their efforts to map and monitor their forests, specifically in the context of meeting international requirements and opportunities such as the REDD (Reducing Emissions from Deforestation and Degradation) initiative.

The exploitation of satellite images can help in two areas, key to provide the information required; activity data; that is the area deforested or degraded; and emissions factors; the amount of biomass lost in the transition process.

The relatively new technology also opens an opportunity for civil society (NGOs etc.) to monitor forest exploitation, bringing a new tool to enforce governance and transparency to the sector.

Remote sensing technology

The imaging of large tracts of the Earth's surface brings many opportunities. Among these is the observation of forest resources. For tropical countries, the capacity to exploit this technology opens the opportunity for national agencies for the first time to monitor extensive areas of often remote forest resources at relatively low cost and in near real time. In contrast to this, traditional methods, based on field survey, are high cost, time consuming and point based, that is not extensive.

This thesis reviews the potential benefits that remote sensing technology can bring to tropical countries in their efforts to monitor and map their forests, in the context of reporting current and historical emissions which derive from deforestation and forest degradation. The framework of the thesis is specifically focused on the emerging international context of the REDD (Reducing Emissions from Deforestation and Degradation) initiative, under which countries may obtain financial grants for demonstrating that they are reducing their carbon emissions with respect to their recent historical practice.

To this end the thesis is structured in the following manner.

i) A review of REDD; what are the rules and what decisions must a country take to implement it.

ii) What Remote sensing data are available and how can these be used to report on REDD

iii) What considerations are needed so that traditional (field based) data be integrated with the new remotely sensed data so as to have an efficient and robust monitoring system

iv) What are the potential improvements in forest monitoring that will come about due to the arrival of the new Sentinel-2 satellite system

v) Does remote sensing data allow us to improve our efficiency in the estimation of biomass over large areas

vi) What can we predict as future areas of forest change in Tanzania

The thesis background

The thesis was undertaken in the context of the ReCaREDD project, funded by the European Commission's overseas aid directorate, DG DEVCO. ReCaREDD (Reinforcing Capacities for REDD) is implemented by the Commission's research directorate, the Joint Research Centre, at its Ispra site in Italy. In the duration of the thesis, two field missions were undertake in Tanzania (Hojas and Eva), and one further mission was undertaken so as to finalise forest change estimates (Eva et al). Work on the implications of REDD for tropical countries (the subject of chapter 4) was financed and supervised by CIFOR, the International Center for Forest Research.

The majority of the work presented in the thesis has already been published in peer reviewed journals or European Commission Joint Research Centre Technical Reports. The exception to this is chapter 7, which is, at time of writing, under review for the journal Remote Sensing. The list of publications is given in Annex 11.

The thesis examines the capacity of the 30-m spatial resolution satellite data Landsat senor for mapping fragmented forest cover compared to 5 m spatial resolution RapidEye data. While the former, medium resolution data give poor results, the transition to 5m data provides a fairly consistent measure of degradation.

Key conclusions from this part of the work are important for monitoring REDD.

Linking field data and remote sensing

Forest agencies in almost all tropical countries have a traditional approach to resource management and forest inventory. This methodology, while appropriate to the goals required, is not always adapted to the integration and exploitation of satellite data. The Tanzania forest service carried out an extensive national field survey for forest resources over a three year (2010-2013) period. Despite the extensive effort and costs, the thesis showed that the resulting database of spatially explicit forest parameters was difficult to exploit with satellite data. In chapter 7 the problems encountered and possible solutions are shown. The conclusions are based on field survey designed and carried out during the thesis as one of the goals of the work is to employ remote sensing data to support

reporting for REDD+. Therefore one aims to use field data to calibrate remotely sensed data, so that the point information of the former could be extrapolated using the spatially extensive properties of the latter. To effect this, field data was required for the work. At the start of the program no agreement was in place so as to be able to exploit the field data collected by the national agency NAFORMA. Time constraints on such an activity were severe, so an efficient campaign strategy was required to collect field data. As a result it was essential to collect a targeted sample of field data so as to be in a position to proceed with the thesis. The goals of this field data collection were:

- To gain experience in a targeted and rapid field data collection to link with remote sensing data, and provided guidelines on appropriate protocols
- To link the field data collected to the remote sensing data, so as to produce maps on above ground biomass, tree height and vegetation cover using Very High Resolution images
- To review differences in species composition between sites located in different ecozones
- To demonstrate the differences in composition and forest parameters between intact and non-intact forests

An important part of remote sensing science is the calibration and validation of maps and products derived from satellite images using 'ground truth', that is, in situ measurements. For statistical assessments, large quantities of reference points are required. Extensive field data are difficult to obtain for reasons of access, manpower, timing and costs and usually can only be undertaken by national forest agencies. These latter data are not always available to outside institutions and may not always target the application envisaged by other users.

Field survey was carried out in conjunction with the Tanzania Forest Service in 2012, 2013 and 2014. The field data collected in this work were used to generate models in conjunction with very high resolution satellite data to produce maps of forest cover, basal area and biomass for the Pugu Hills forest reserve.

New technologies – especially for dry ecosystems

Short revisit periods are potentially important to monitor forest at national/regional scales. Firstly, the increased coverage provides more opportunity for acquiring cloud-free images, particularly important in tropical regions (Beuchle et al., 2011). Secondly, because they should allow us to exploit seasonal differences in canopy reflectance characteristics as a means of discriminating between forest cover types and different forest conditions (e.g. closed and open forests, or deciduous and degraded forests), which is especially important for the dry forest.

With the advent of the Sentinel-2 program, data availability over target areas is increasing, allowing temporal analysis previously restricted to moderate (>100m) spatial resolution satellite data (such as MODIS), to be employed in the monitoring of forests at finer spatial resolutions. Sentinel-2 brings an improvement in the spatial resolution (with the three visible and a near infrared bands at 10m), which allows a more accurate assessment of deforestation and forest degradation areas taking the minimum scales defined by the UNFCC, and in the spectral sampling (i.e. higher amount of bands with narrower width), with the inclusion of three bands in the red edge, which has shown to be useful for quantitative assessment of vegetation status (Frampton et al., 2013).

The thesis aimed to see how the new remote sensing data can help in the monitoring of forest changes and forest degradation.

National Biomass Estimates

While the documentation of activity data (i.e. forest change) should be relatively easy using remote sensing (although we see high variance in estimates), the estimation of biomass is far more problematic. Evolving systems such as new P-band RADAR and LIDAR are not reviewed in the thesis – they require a full treatment out of the context of this work.

Optical systems clearly are not optimal for monitoring biomass, as they image only the surface of the vegetation. Nevertheless, given the availability of such data (the LIDAR and P band RADAR are yet to be widely distributed) the thesis attempts to see if they can be used, in conjunction with field data, to improve (in terms of efficiency) large area biomass estimates. In the context of the thesis a test was carried out over Tanzania.

The goal of the thesis is to see how the combination of remotely sensed data with field data can improve the estimates of biomass at the national scale.

The potential improvement was measured by the Relative Efficiency (RE) – showing the reduction (or otherwise) in variance that can result from the combination of a second data set, in this case the remote sensing data. This implies that the field data collection can be reduced, with significant cost savings.

Future prospects for forests in Tanzania

The final objective of the thesis is to look at explanatory spatially explicit models on the probability of deforestation. An analysis of historical deforestation with a set of geographical parameters is carried out. Variables such as proximity to towns, roads, deforested areas and forest edges are used.

3. Structure

Besides the introduction and synthesis, this thesis is organised into five chapters, following the focal areas of research described in the objectives. Publications were developed during this research project, each of which is presented as a separate chapter. Four of the papers are already published and hence available in the scientific domain.

The organisation of the chapters does not follow a chronological order; instead, the publications presented were arranged according to the topic they are dedicated to, and following a line of increasing complexity and added novelty of the approaches.

Chapter 4 reviews the policy background to REDD+ and outlines the rules and choices to be addressed by participatory countries according to the UNFCCC guidance in the implementation of monitoring systems to estimate forest cover changes and forest carbon stocks, and demonstrate some of the technical problems and options that they can face and adopt in the remote sensing exercises.

Chapter 5 addresses field survey, which is a major part of forest work required in reporting for REDD+ includes effective field survey. These data, which in the past formed part of 'stand alone' data for forest inventory, are more and more linked to remotely sensed images. For Chapter 5 the methods and opportunities brought about by this development are reviewed, and results from the PhD field work in Tanzania are presented.

The arrival of the new Sentine-2 satellites has provided an opportunity to map forests at a finer resolution with a high frequency of acquisitions***. Chapter 6*** demonstrates this potential using a simulation. The study focuses on an area of dry forest in Central Tanzania.

In ***Chapter 7*** presents a full scale estimate of above ground biomass for the whole of Tanzania, using a combination of remote sensing and field data.

The dynamics of deforestation around Dar es Salaam are demonstrated in ***Chapter 8***, along with a model to infer future probability of deforestation at the national level. The model's potential as an operational pan-African forest monitoring system was assessed, while issues of future research were specified.

4. Monitoring deforestation and forest degradation in the context of REDD+. Lessons from test studies in Tanzania

This chapter contains material published in:

Lorena Hojas- Gascón, Paolo Omar Cerutti, Hugh Eva, Robert Nasi and Christopher Martius , 2015, Monitoring deforestation and forest degradation in the context of REDD+. Lessons from Tanzania. CIFOR Infobrief. No. 124, May 2015. doi:10.17528/cifor/005642. CIFOR, Bogor.

4.1 Introduction and objectives

The implementation of national monitoring systems to estimate forest cover changes and forest carbon stocks for REDD+ purposes implies the selection of monitoring variables, measuring techniques and information sources.

There are a number of rules and choices to be addressed by REDD+ participatory countries according to the UNFCCC guidance in the implementation of such monitoring systems, and demonstrate some of the technical problems and options that they can face and adopt in the remote sensing exercises. In particular it covers three aspects:

- ☑ The national forest definition for forest cover mapping
- ☑ Baseline data (to assess historical land cover changes and establish Forest Reference Emission Levels)
- ☑ Biomass estimation (linking remote sensing data and forest carbon stock field data)

These issues raise a number of research and practical questions, which need to be addressed.

- *Which are the minimum mapping unit, the minimum canopy cover and the minimum tree height countries should choose for the national forest definition?*
- *Are the widely available Landsat data (30 m spatial resolution) adequate to monitor forest cover and forest cover changes or are finer spatial resolution data needed, e.g. RapidEye (5m)?*
- *Can data from the widely publicised Global Forest Maps (*(Hansen et al., 2013b) *adequate for use by countries as baseline data to establish Forest Reference Emission Levels? If not, why (course resolution, classification methodology, etc.)*
- *Is it possible to correlate information extracted from satellite data to forest biomass from field data? If so, how much of the forest biomass could be estimated by satellite data?*

4.2 Methodology

Within this chapter the following aspects are reviewed;

1. What are the potential impacts of the selection of the three main properties of the national forest definition: minimum mapping unit (MMU), minimum canopy cover and minimum tree height?
2. To examine the cause of the differences in the previous datasets and the consistency of the GFMs, comparison of the forest cover from the GFM of 2010 with VHR images of the same year (classified at 0.5 ha MMU) and field data for four study sites of the main ecoregions of Tanzania.
3. Assessing the performance of Landsat and RapidEye data in mapping fragmented forests (at a MMU of 0.5 ha) for one study site, taking very high spatial resolution data (<1 m spatial resolution) as a reference of the "true" forest area.

4.3 Forest definitions

What are the potential impacts of the selection of the three main properties of the national forest definition: minimum mapping unit (MMU), minimum canopy cover and minimum tree height?

In the framework proposed by the UNFCCC, forest is an area of land with at least 0.05-0.5 ha and a minimum tree-crown cover of 10%-30%, with trees that have reached, or could reach, a minimum height of 2-5 m at maturity in the same location (OECD and IEA 2007). Inside these limits, the national forest definition selected by countries will have its own technical and political consequences; for instance, the definition will affect the forest baseline area and change estimates. However, it is not easy for both practitioners and policy makers to determine *a priori* the exact future repercussions of the chosen forest definition; repercussions will depend on the proportions of different forest types and changes, as well as on the spatial and temporal scales at which they take place. At least three major concepts need clarification as they may foster better understanding of the potential future impacts of chosen definitions: the *minimum mapping unit*, the *minimum canopy cover* and the *minimum tree height*.

1. The minimum mapping unit (MMU) is the smallest land unit that is mapped. The adopted MMU will affect the most suitable type of satellite data to use, and therefore the type and amount of work required for monitoring a country. For example, a larger MMU facilitates

visual interpretation by expert analysts, but also means that one land unit will most likely contain mixed land-cover types (e.g. forests in different states of degradation). Therefore, clear protocols on land-cover legends will be needed to deal with such mixed types. In contrast, a smaller MMU implies more work involved in quality control because the number of land units to check will be larger; digital image processing will be more reliable by avoiding mixed classes (i.e. improved spectral purity).

2. The minimum canopy cover is the minimum area of a land unit that has to be covered by trees' canopies to be classified as forest. The selected minimum canopy cover will influence the classification of the land units (i.e. forest or non-forest) and therefore affect estimates of changes occurring in the carbon pools. For example, setting the canopy cover threshold at 30% will place degraded forests with a canopy cover between 10%-30% into the non-forest land use type, and therefore would not be included in any reporting system. The opposite will be true with a lower threshold (i.e. 10%), allowing those degraded forests to be included as forest in the baseline and change estimates.
3. The chosen minimum height of trees would similarly impact the land-cover transitions on which carbon pools changes are estimated. For example, if the minimum height is set at 2 m, an area that shifts from a situation with trees higher than 5 m to one with trees between 2-5 m would be classed as degradation. Instead, if the minimum height is set at 5 m, the same change would be classed as deforestation. Similarly, a minimum height set at 2 m would result in deforestation if trees of 2-5 m were to become lower than 2 m, while the same change would be classed as "no-change" if the minimum height were set at 5 m. This will have important consequences for forest management, as areas such as regenerating forest or logged-over forest will fall under different legislation depending on the land category.

TREES are from 2 - 5m

		Current		
		WV>5	*WV<5*	*NF*
	WV>5	NC	Deg	Def
Historical	*WV<5*	Regen	NC	Def
	NF	Aff.	Aff.	NC

Trees are above 5m

		Current		
		WV>5	*WV<5*	*NF*
	WV>5	NC	Def	Def
Historical	*WV<5*	Aff	NC	NC
	NF	Aff	NC	NC

WV>5 = woody vegetation greater than 5m	
WV<5 = woody vegetation from 2 to 5m	
NF = non forest	
Def	Deforestation
Deg	Degradation
NC	No change
Regen	Regeneration
Aff.	Afforestation

Figure 10: Changes in the minimum tree height has consequences on the resulting class of changes (Source: Eva and Hojas Gascon 2017)

In addition to the technical matters that play in setting the parameter's levels, eminently political choices will have to be made. In the best-case scenarios, such choices will have to be discussed and agreed upon in broad national consultations. For example, let us assume that a country with large areas of already degraded forest (e.g. low canopy cover and few tall trees) decides to adopt low thresholds in its definition of forest. It could decide for technical reasons: the baseline area of forest (on which present emissions and future changes will be reported) will be larger than with higher thresholds. But it could make the same decision for more political reasons: some degraded areas (for instance, those from where rural communities generally derive their livelihoods) could be included in the forest class to encourage community participation in the REDD+ process and foster community engagement in nation-wide environmental processes. Certainly, financial reasons can steer these political decisions from one definition to another; for example, emission levels could give potential financial credits, if levels are reduced.

4.4 Consistency of forest area and forest area change maps from the Global Forest Maps

The Global Forest Maps were published in 2013(Hansen et al., 2013b), just before the UNFCC Convention of the Parties meeting. As such, they were seen by some as a solution for assessing historical land-cover changes and establishing reference emissions levels for REDD+ purposes. These are wall to wall maps, and present not only the forest areas with different percentage tree cover, but also annual deforestation from the year 2000 onwards – available at the pixel level, and hence potentially allowing countries to report on historical emissions. Reviewing different sources of data on forest area and forest change (see Chapter 1 Table 2) it is clear that there are large discrepancies. The estimates from the Global Forest Maps are clearly at odds with other estimates of forest cover and forest cover change.

To examine the consistency of the GFM data, the maps were compared with very high spatial resolution (VHR) images[1] for four sites (7x7 km) in the main ecoregions of Tanzania. The sample sites are close to Kisarawe (tropical rainforest), Ikwiriri (tropical moist deciduous forest), Mtera (tropical dry forest) and Mgongila (tropical shrubland) (Figure 11). These were chosen because field data were available for three of the sites from previous fieldwork carried out by the JRC in 2012, 2013 and 2014. The field survey comprised of a targeted sampling aimed to characterize the main forest types and conditions along with the collection of biomass-related variables such as height and diameter at breast height (Hojas Gascón and Eva 2014). The VHR data were classified into tree, shrub and non-woody classes in land units of 0.5 ha. Tree height was estimated from the image texture information calibrated with the field data. The classifications were overlaid with the GFMs and the tree percentage for the year 2010 was compared by a cross tabulation (Figure 12).

[1] From the WorldView-2, Geoeye-1 and IKONOS-2 satellites acquired in the year 2010.

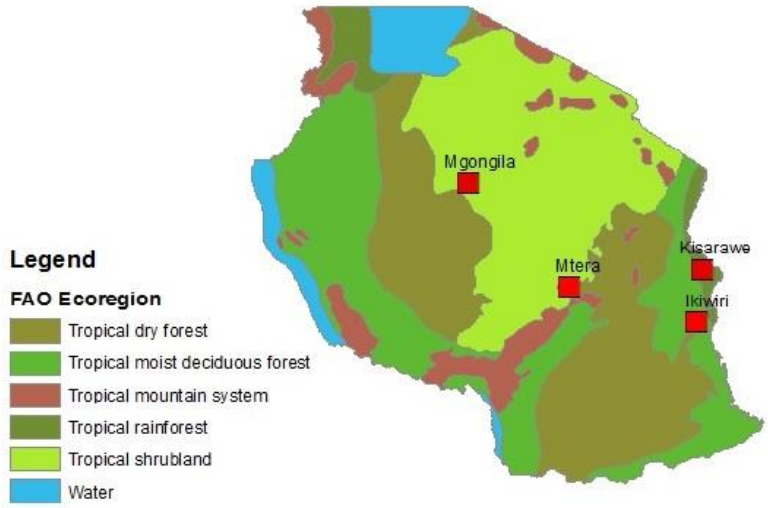

Figure 11: Location of the sample sites for the comparison study of the Global Forest Maps with very high spatial resolution data in Tanzania

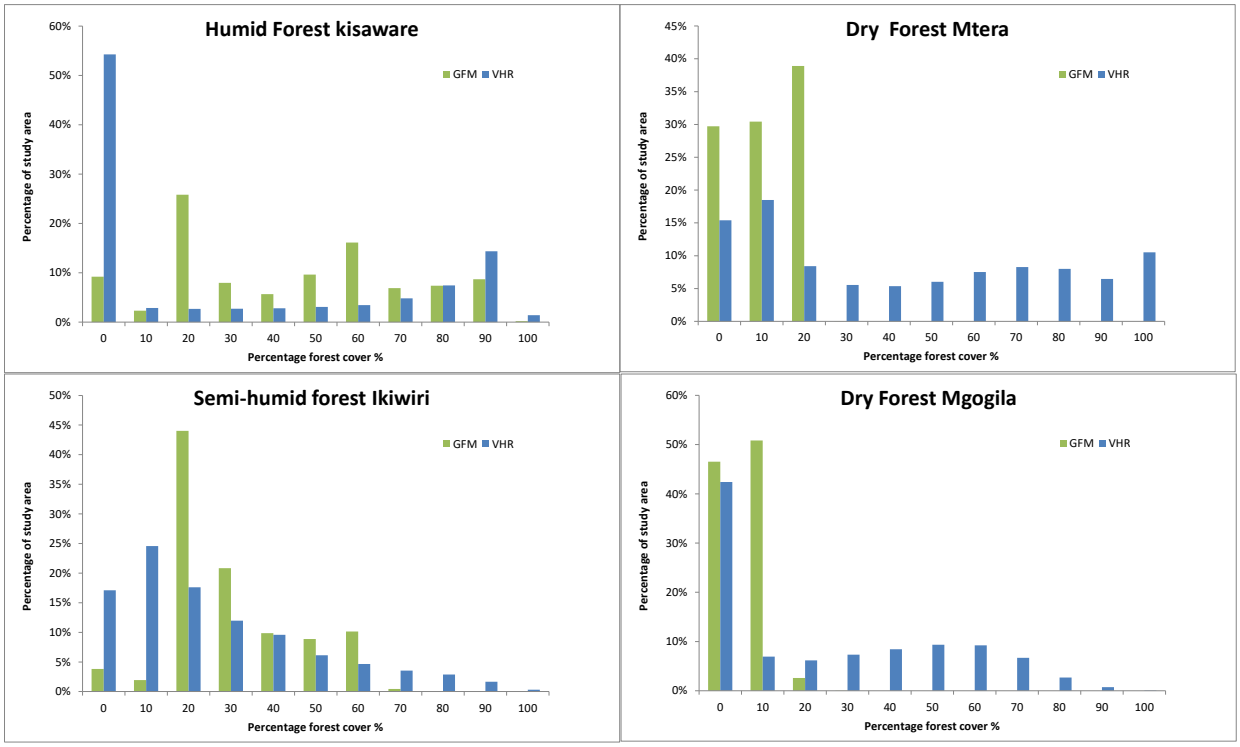

Figure 12: The distribution of forest cover percentage in the four study areas as mapped by VHR data (blue) and the Global Forest Maps (green)

Results indicate important differences between the two data sets. In the dry ecoregions (Mtera and Mgogila sites), the GFMs underestimate tree cover. The maximum tree cover percentage recorded on the Landsat data used by the GFMs was 30% (Figure 12: top and bottom right), while in the VHR data, supported by the JRC and NAFORMA field survey, tree cover percentages go up to 100%. On

the other hand, in the humid and semi-humid ecoregions (Kisarawe and Ikiwiri) (Figure 12: top and bottom left), the GFMs tend to overestimate forest cover, especially in the humid ecoregion. Areas classified as shrub (tree height lower than 5 m) from fieldwork and through examination of the VHR data were classified as tree cover in the GFMs. Also, significant areas where tree-cover loss had occurred in both regions went undetected in the GFMs. In the humid forest area of Kisarawe, the GFMs failed to detect an area of about 100 ha (over 1000 Landsat pixels) that was deforested in 2008 (Area A, Figure 13). At the same time, shrub cover was classified as 50%-60% of tree cover (Area C, Figure 13).

Conversely, in the dry forest area of Mtera, the GFMs map the full tree cover on the hills as only 15%-20% of tree cover (Figure 14). Tree height measured on the hills in the field survey was 8-10 m, with 100% cover.

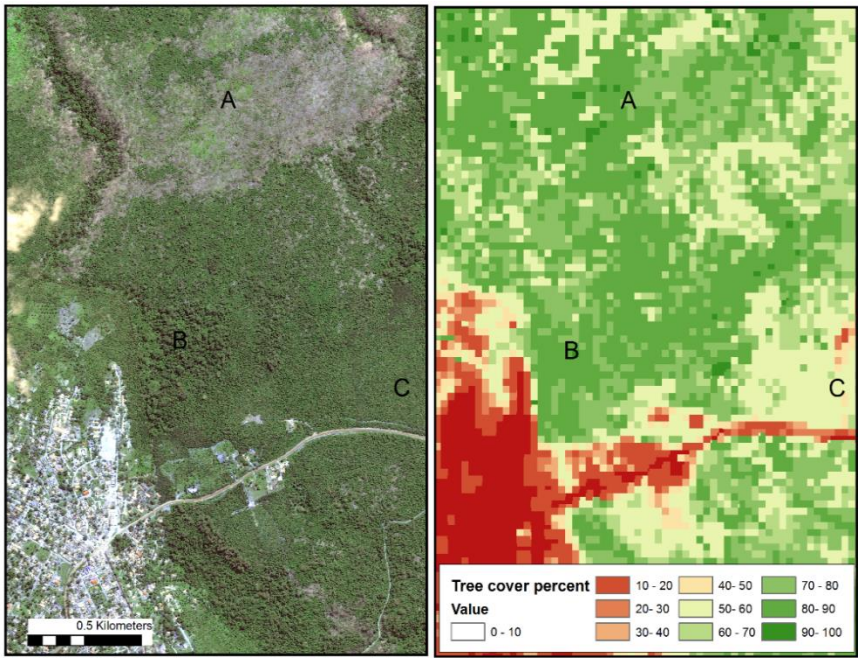

Figure 13: WorldView-2 image subset (true color composite) of 2010 of The Pugu Hills forest reserve around Kisarawe (left) and tree cover percentage from the Global Forest Map for the same area and date (right). Area A, deforested in 2008, is assessed as over 70% of tree cover. Area B is correctly assigned as over 80% of tree cover (field survey at this point found the average tree height to be 20 m). Area C is wrongly classed as 50%-60% of tree cover (field survey measured full canopy of shrubs with an average height of 3.6 m).

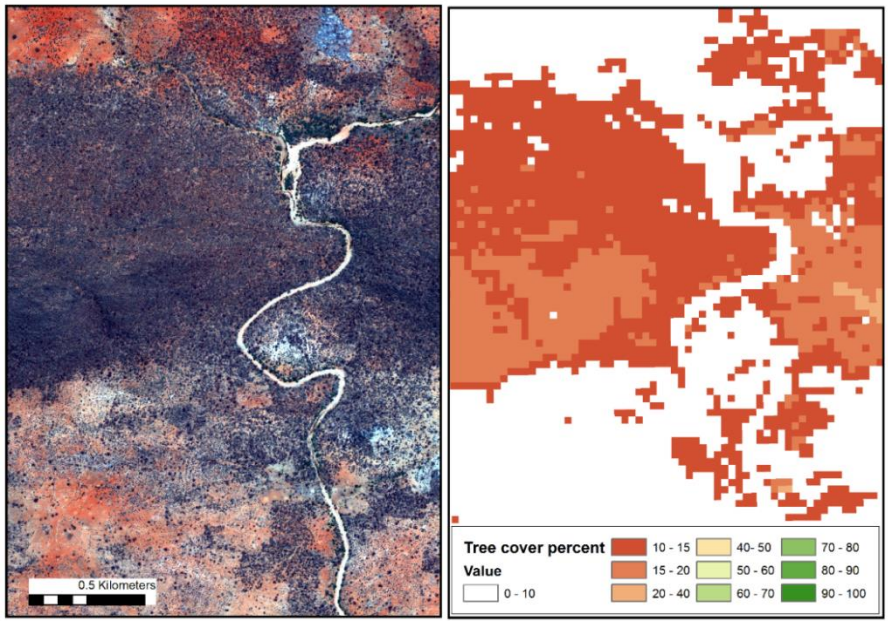

Figure 14: Worldview-2 image subset (true color composite) of 2010 of the study area at Mtera (left) and the corresponding Global Forest Map for the same year (right)

These errors of omission and commission arise due to the resolution of the Landsat data used to produce the GFMs and to the calibration method employed with respect to the seasonality and the diversity of ecoregions in Tanzania. For example, images acquired during the dry season over dry areas can be misclassified in low tree-cover classes as background soil, and will have a strong influence on the reflectance response; images with disturbed forests over the humid domain can be misclassified as high tree-cover as they can regenerate freely and appear in pristine conditions; also in the dry forest, minimum tree height should be lower (e.g. 3 m) as primary forest is shorter than in the humid ecoregion. All in all, results corroborate previous assessments of the possible misinterpretations present in the GFMs (e.g. Tropek et al. 2014). More importantly, and as also recommended by the GFMs' authors, results indicate that forest maps produced with a global aim must be integrated with more locally relevant and appropriate data sets (Hansen et al. 2014). This is especially true when countries plan to use the resulting information — in the national REDD+ debates or elsewhere — to devise strategies that could have serious impacts on the livelihoods of their citizens and beyond.

Another barrier to using the GFM (or other similar products derived from the Landsat archive) for historical forest changes for reference emissions levels is the major bias in the number of acquisitions that have been made throughout the life-period of the Landsat series. Reviewing the number of images available over Dar es Salaam as an example, one can see from Figure 15 the number of available scenes varies markedly throughout the period. Generally in the late 80's and

early 90's, only one image was acquired per year. This rises in the 2000's to around 12 images per year, and then, with the improved acquisition strategy to 40 a year in the current period. Clearly the chances of detecting a deforestation event will be linked to the number of acquisitions, especially for more subtle events, such as shifting cultivation, or small scale interventions, where the vegetation may grow back quickly.

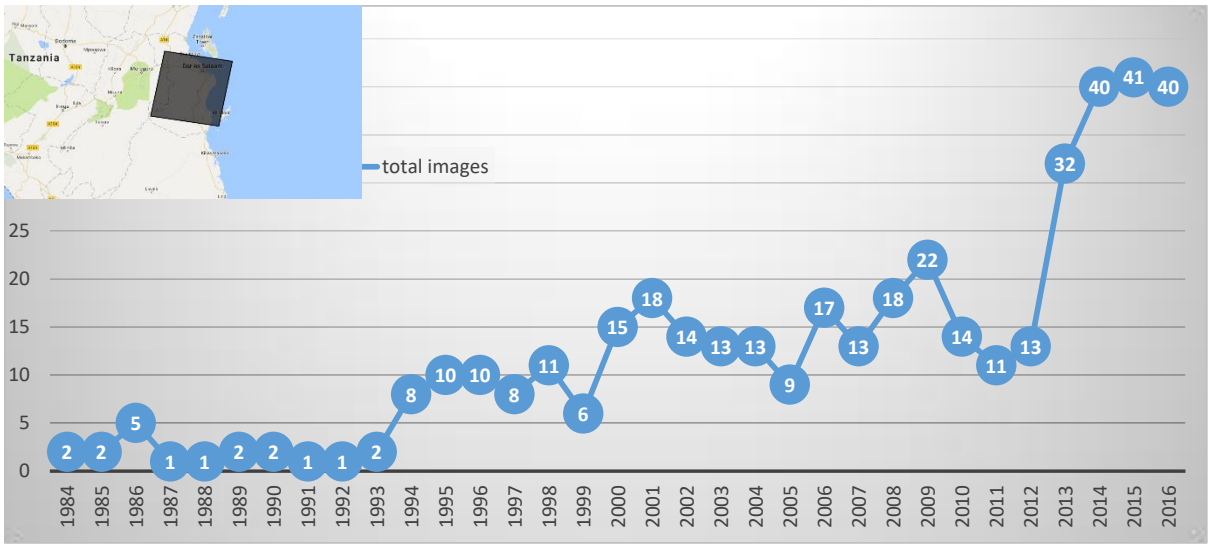

Figure 15: The number of Landsat acquisitions for Path Row 166-65 from 1984 to 2016– corresponding to Dar es Salaam (inset)

So, while large scale deforestation may be picked up with one image a year – or even at a letar date, finer dynamics, especially forest degradation, may be missed.

4.5 Assessing the performance of Landsat and RapidEye data in mapping fragmented forests

Large areas of fragmented and degraded forest area known to cover the tropics. Within a REDD+ monitoring scenario, the need to measure and to monitoring changes in these areas will be required. Clearly, monitoring of forest degradation is more challenging than monitoring deforestation, especially when considering regional and national levels. Indeed, the recent Brazilian FREL submission does not cover this issue, as it was believed that data were lacking to estimate it (Government of Brazil, 2014).

Since its introduction in 1984, the Landsat Thematic Mapper sensor, on board a sequence of satellites has been the workhorse of forest (amongst other) monitoring. While adapted and used extensively and successfully for deforestation (with major projects carried out by the UN FAO, the EU JRC TREES, Brazil's INPE and the GFM), monitoring of fragmented forests may be beyond the technical capacities of the 30 m spatial resolution of the sensor. Even for deforestation, the choice of a very fine MMU, (e.g. 0.05 ha) would require sub-pixel detection, as mentioned in Chapter 1.

For the Kisarawe site in Tanzania, the performance of Landsat and RapidEye (5 m) in mapping fragmented forests was compared using a MMU of 0.5 ha. VHR data from the WV2 sensor from the 11th June 2010 were employed as a reference of the "true" forest area. Results show that, in fragmented landscapes, Landsat fails to detect forest in many areas with the 20%-90% forest-cover range (red area, Figure 16 right), and has high variance in areas with forest cover higher than 90% (blue area, Figure 16 right), while RapidEye data provide a far better estimate (Figure 16, left). This indicates the spatial resolution of the Landsat sensor is too crude for monitoring forest degradation in the context of REDD+.

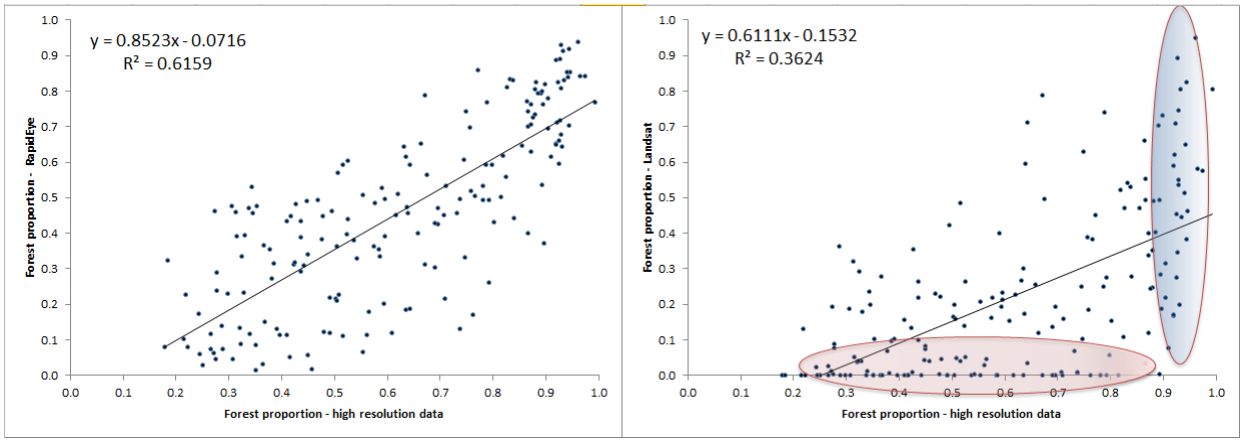

Figure 16: Forest proportion estimates from RapidEye (lef) and Landsat (right) compared to estimates from VHR WV-2 data

The differences in forest-cover estimates that arise from the selection of different MMU are difficult to predict (Hojas Gascón and Eva, 2017). They may also vary from site to site depending on the proportions of changes and the scales at which they take place. In this test, at an MMU of 0.5 ha (approximately the middle of range values proposed for REDD+), 30 m spatial resolution satellite data are too crude for mapping fragmented forest cover. One notes that 5 m spatial resolution data give considerable improvements, despite a relatively a relatively high variance. The selection of a

lower MMU could result in more accurate results, but also in too many land units to analyze and validate.

4.6 Conclusions

1. The values chosen for the MMU, the minimum canopy cover and the minimum tree height of the national forest definition will have technical, economic and political consequences (e.g. on the type of satellite data needed, appropriate remote sensing method and classification legend, amount of work to analyse data, forest cover change estimates, forest management and preservation implications…), that will be more or less suitable for countries depending on their particular characteristics and objectives (e.g. proportion of different forest types, spatial and temporal scales at which forest cover changes take place, willing to encourage community participation, obtaining high financial rewards…).
2. From Table 2, it is clear that there are major differences in magnitude in the existing global data sets based on Landsat on forest cover and forest cover change (GFMs, JRC, FAO) at national level, which have to be considered when using baseline data to establish Forest Reference Emission Levels.
3. Important 'errors' in the distribution of forest cover percentage of the GFMs were found in the four study areas (tropical rainforest, tropical moist deciduous forest, tropical dry forest and tropical shrubland) in comparison with very high resolution data, due to the spatial resolution of the Landsat data and to the calibration method employed with respect to the seasonality and the diversity of ecoregions.
4. The large variations in the number of Landsat acquisitions over the last few decades can result in a bias in the number of deforestation events detected, meaning that annual deforestation counts from such a data source are compromised.
5. The spatial resolution of the Landsat sensor was demonstrated too crude for monitoring deforestation for REDD+ purposes in comparison to high spatial resolution satellite data (i.e. 5 m RapidEye) in the study area of the tropical rainforest ecoregion.

In general, results indicate that estimates of forest cover changes and carbon stock at national levels valid for REDD+ reporting could be feasible using a combination of optical satellite data and field data. However, to obtain good results the national field survey ideally should be designed for use in combination with satellite images and these images should have higher spatial resolution than the most commonly used 30 m available on the Landsat sensor.

In the humid forest domain, biomass measurements from field survey were directly linked to remote sensing parameters with acceptable results. These could be improved with a better co-location of the field data already collected and the satellite images. Other improvements of this approach, especially for the other domains with higher variances in cover, stand height and condition (wet, dry or burnt), could include cleaning of plots with heterogeneous forest cover, different seasonality than image acquisition, or burned after image acquisition.

5. Field survey for forest mapping with high resolution satellite data.

This chapter contains material published in:

Lorena Hojas Gascón and Hugh Eva, 2014, *Field guide for forest mapping with high resolution satellite data.* Technical Report by the Joint Research Centre of the European Commission (JRC 92600, EUR 26922 EN). Publications Office of the European Union, Luxembourg, 2014.

5.1 Introduction

One of the goals of this thesis (Chapter 1) is to employ remote sensing data to support reporting for REDD+. The aim is to use field data to calibrate remotely sensed data, so that the point (high quality) information of the former could be extrapolated using the spatially extensive properties of the latter. To effect this, field data was required for the thesis work. At the start of the program no agreement was in place so as to be able to exploit the field data collected by the national agency NAFORMA. National agencies are quite understandably reluctant to share field data without an agreement on conditions of use. Time constraints on such an activity were severe, so an efficient campaign strategy was required to collect field data. As a result it was essential to collect a targeted sample of field data so as to be in a position to proceed with the thesis. The goals of this field data collection were:

- To gain experience in a targeted and rapid field data collection to link with remote sensing data, and provided guidelines on appropriate protocols
- To link the field data collected to the remote sensing data, so as to produce maps on above ground biomass, tree height and vegetation cover using Very High Resolution images
- To review differences in species composition between sites located in different ecozones
- To demonstrate the differences in composition and forest parameters between intact and non-intact forests

5.2 Background

An important part of remote sensing science is the calibration and validation of maps and products derived from satellite images using 'ground truth', that is, *in situ* measurements. For statistical assessments, large quantities of reference points are required. Extensive field data are difficult to obtain for reasons of access, manpower, timing and costs and usually can only be undertaken by

national forest agencies. These latter data are not always available to outside institutions and may not always target the application envisaged by other users.

The manual produced during the thesis (Hojas Gascón and Eva, 2014a) was targeted at guiding the collection of adequate field information in a relatively short time, for calibrating and validating classifications and biophysical parameters derived from satellite images of high spatial resolution for forest monitoring. These products have been shown to be useful to increase the effectiveness of suitable forest management at local level, e.g. to control the fuel wood production capacity and detect areas of illegal harvesting. In the future, they will allow us to improve the assessment of forest biomass and carbon budgets changes at national level, with the aim of monitoring deforestation and forest degradation in the context of the UN-REDD programme.

Field survey was carried out in conjunction with the Tanzania Forest Service in 2012, 2013 and 2014. The field data collected in this work were used to produce maps of forest cover, basal area and biomass. These maps were transferred to the Tanzania Forest Service.

The field survey was executed using basic, readily available and inexpensive methods (standard GPS, digital camera and paper data forms) so as to ensure that all services can carry out a similar exercise. More sophisticated equipment combining GPS, camera, data logger and GIS are available, but remain relatively expensive. As such equipment become available at lower prices, it will become feasible to collect more data in less time.

The final goal of the field survey was to calibrate satellite images for vegetation classification, and within forest areas, discriminate degradation levels. Therefore, two objectives were set out for the choice of the location of the field sites:

1. To obtain an overview of the different types of vegetation formations and degradation processes existing in Tanzania. For this a set of ***general field sites*** was required where rapid visits (with limited field work) were done, so as to gain first-hand experience of forest types and degradation dynamics throughout the country.
2. To get detailed information of one of the forest types in Tanzania. From the general sites, one was selected as the ***primary field site*** where intensive field work was carry out, in order to develop remote sensing methods for forest degradation monitoring.

5.3 General field sites

The goal was to locate a maximum of four representative sites of the main ecoregions of the country. The sites needed to be located in areas i) with evidence of forest degradation processes, ii) where supporting spatial and remote sensing data were available, and iii) there was feasible accessibility. To determine the potential sites which complied with these characteristics, a geographic database of the country was assembled, the historical changes in forest areas were reviewed, and the availability of satellite data and the accessibility through paths and roads were checked.

5.4 Satellite image screening

Satellite images of very high spatial resolution were sought for those TREES-3 sample sites with evidence of forest degradation. Such information was gleaned from the TREES study results, more updated sources (*i.e.* Google Earth) or areas suggested by national experts (most recent areas of degradation). The availability of these types of high resolution images is low in areas of high probability of cloud cover or low economic potential, and therefore several satellite data sources were used, in particular, images from Quickbird, Worldview-2 IKONOS and Geoeye-1.

These satellite data with a spatial resolution less than or equal to a metre give a highly detailed view of the target areas, but they are relatively expensive and cover small areas. Fourteen images of 7x7 km size were selected, located in different biomes and distributed across the country for the year 2010. They belong to the same season as the corresponding RapidEye images, always choosing the closest date available, and when possible, a second image was obtained for the year 2011, so as to be able to review evidence of forest degradation processes (Table 5).

5.5 Selection of general field sites

From the above mentioned fourteen sites covered by high spatial resolution satellite data ('reference sites'), four were selected for field visits ('general and primary sites') according to the accessibility by roads and paths. They were chosen so as to characterise the main forest domains of Tanzania: tropical rainforest, tropical moist deciduous forest, tropical shrubland and tropical dry forest, and correspond to the TREE boxes (LAT-LONG) S7E39, S8E39, S7E36 and S6E34 respectively.

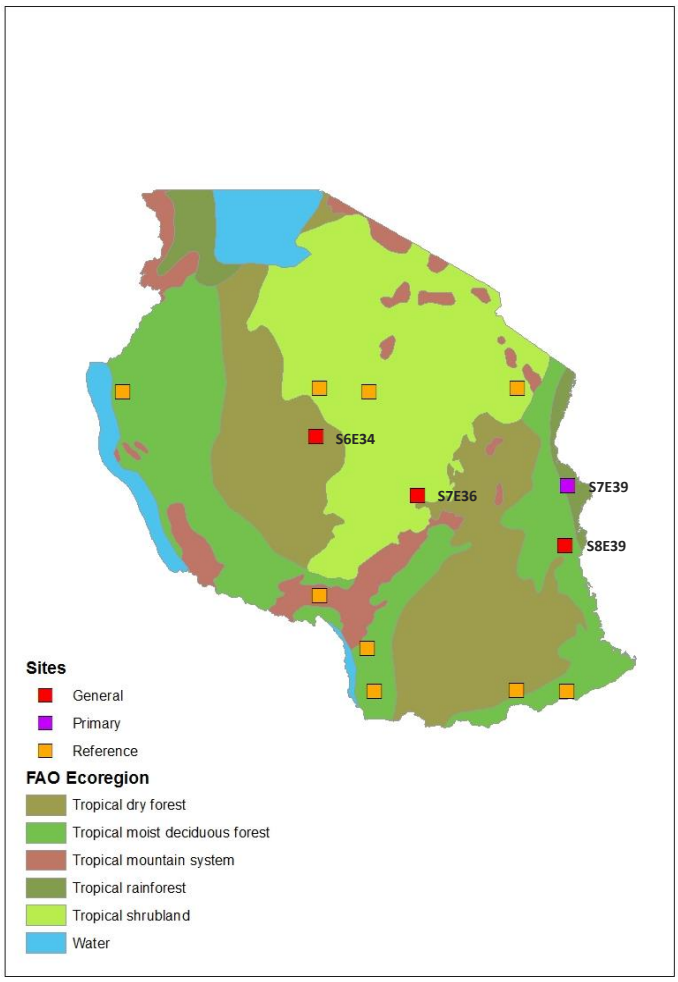

Figure 17: The location of all the VHR sites. Note that very few satellite images were available for the northern part of the country due to frequent cloud cover and low image acquisition. In yellow all the 'reference sites' covered by very high spatial resolution.

5.6 Primary field site

To select the primary field site a location was required where logistical support could be employed from the NAFORMA field team. The site corresponding to the TREES box S7E39, in the forest of the Pugu hills, near the town of Kisarawe and the closest to Dar es Salaam, fulfilled this condition. Initial examination of available remote sensing data (Landsat, RapidEye and WorldView-2) showed forests in different stages of maturity and degradation.

Table 5: Characteristics of the very high spatial resolution images acquired for the thesis: WV-2: Worldview 2, GE-1: Geoeye-1, IK: IKONOS, QB: Quickbird, - : no image for that date.

TREES box	Location	Biome	2010 date	2010 satellite	2011 date	2011 satellite
S11E39	Luatala	Tropical moist deciduous forest	5-Aug-2010	WV-2	2-Aug-2011	WV-2
S7E39	Kisarawi	Tropical rainforest	11-June-2010	WV-2	-	-
S8E39	Ikiwiri	Tropical moist deciduous forest	9-May-2010	GE-1	11-Jul-2011	GE-1
S7E36	Mtera	Tropical shrubland	4-Sep-2010	WV-2	6-Sep-2011	IKONOS-2
S5E38	Mkokoni	Tropical shrubland	27-Nov-2010	IK	-	-
S6E34	Mgandu (west of)	Tropical dry forest	12-April-2010	GE-1	-	-
S5E30	Kazuramimba	Tropical moist deciduous forest	15-Jul-2010	GE-1	1-Jun-2011	GE-1
S9E34	Ikuwo (south of)	Tropical mountain system	17-Dec-2010	WV-2	2-Sep-2011	GE-1
S8E36	Mlolo	Tropical mountain system	-	-	31-Oct-2011	IK
S11E38	Ruanda	Tropical dry forest	19-Jun-2010	GE-1	26-Oct-2011	GE-1
S5E34	Mgongila	Tropical shrubland	3-Jul-2010	QB	12-Aug-2011	WV-2
S10E35	Mahanje (west of)	Tropical moist deciduous forest	13-Oct-2010	GE-1	3-Oct-2011	WV-2
S11E35	Mbinga (east of)	Tropical moist deciduous forest	3-Nov-2010	WV-2	25-Sep-2011	WV-2
S5E35	Misughaa	Tropical shrubland	21-Aug-2010	QB	-	-

These images were used to provide information on the structure of different forest types across the country in different ecoregions to aid in image interpretation.

5.7 Site description

The TREES sample site (20 km x 20 km) centred at 7°S 39°E, suffered a forest fragmentation of 3.4% of the area between 1990 and 2000 (Bodart et al., 2011). More precisely the study was concentrated on a smaller 7x7 km window centred at 6.91°S 39.06°E, covered by a Worldview-2 satellite subset image. The area is located in the administrative region of Pwani, district of Kisarawe. At its centre is located the town of Kisarawe, at less than 29 km from the city of Dar es Salaam. The Dar es Salaam-

Maneromango road that passes through Kisarawe town is the main route that links the settlements in Kisarawe district with Dar es Salaam.

The site belongs to the Zanzibar-Inhambane ecoregion, which covers the coastal belt of East Tanzania, Figure 2 & 17, in the moist and dry lowland area of the country, and its vegetation is characterized by humid tropical rainforest. Most of the land is below 200 m and the rainfall is between 800 and 1200 mm per year[2] with a well-defined dry season which runs from mid-June until end of September[3]. The mean annual temperature is 26 °C. Forest is the most widespread climax vegetation, but has been largely replaced by secondary wooded grassland and cultivation. There are also extensive areas of scrub forest and edaphic grassland, and smaller areas of transition woodland, bushland and thicket. The forest is rich in species and difficult to classify (White 1983).

The North-Eastern part of the study area includes a section of the Pugu Forest Reserve, and the South-Western part a section of the Kazimzumbwe Forest Reserve. The rest of the area is composed by bushland with emergent trees mixed with croplands (N-W), mixed crops (S-W) and woodland with scattered crops and tree plantations (S-E) (see Figure 19). The forest reserves contain two main vegetation types: moist forest, remaining primarily only on hillsides, and impenetrable thickets. According to the Huntings vegetation map (1996) there have been some changes in the land use, mainly the intrusion of agriculture into the Pugu Forest Reserve and Kazimzumbwe Forest Reserve on their western boundaries. The dynamics in these forest reserves are the result of the population growth in the surrounding communities and the expansion of the city of Dar es Salaam. This has brought on the conversion of forest areas into settlements and agriculture in the surrounding areas and to the unsustainable harvesting of forest product by illegal activities within the forest reserves (logging, charcoal burning, timber sawing) to meet the demands of the population (Mdemu *et al*. 2012).

In both forest reserves there is evidence of forest degradation. The main cause of forest degradation is the extraction of fuel wood and charcoal production. In Tanzania firewood and charcoal account for over 75% of the total energy use, mainly for household cooking (Mwampamba 2007). This seems to be the main driver that has converted former lowland forest into shrub formations. During the field survey the presence of charcoal kilns, and charcoal and fuel wood transport (in both heavy vehicles and bicycles) and markets were frequently observed (Figure 18).

[2] Derived from Climatic zones from FAO forestry country information (Ref. page 70).
[3] Derived from Climatic diagrams from Cisgrasp (Ref. page 70).

Figure 18: Example of a charcoal kiln (left) and charcoal transport (right) in Kisarawe.

5.8 Worldview-2 image specifications

Satellite imagery for the primary field site was acquired by the Worldview-2 optical satellite sensor on the 11th of June 2010. The satellite was launched in October 2009, has a swath width of 16.4 km and a revisit frequency of 1.1 days. The sensor records data in the blue (0.45-0.51 µm), green (0.51-0.58 µm), red (0.63-0.69 µm) and near-infrared (0.77-0.89 µm) spectral bands at a panchromatic spatial resolution of 0.46 m.

The level 1 image data were geometrically corrected in ERDAS IMAGINE (ERDAS 2002) using Rational Polynomial Coefficients (RPCs) generated from the RPB file supplied with the data. GPS track data collected from a preliminary field visit in 2011 were overlaid on the corrected image to assess the geolocation accuracy. The GPS track and the roads and paths in the image were found to match within 3 m.

Figure 19 shows a true colour composite of the Worldview-2 image covering the primary field site with the approximate limits of the forest reserves overlaid. Three main areas of degraded forest can be discerned in the Pugu Reserve, north (1) and south (2) of the main highway, and in the Kazimzumbwi forest reserve (3).

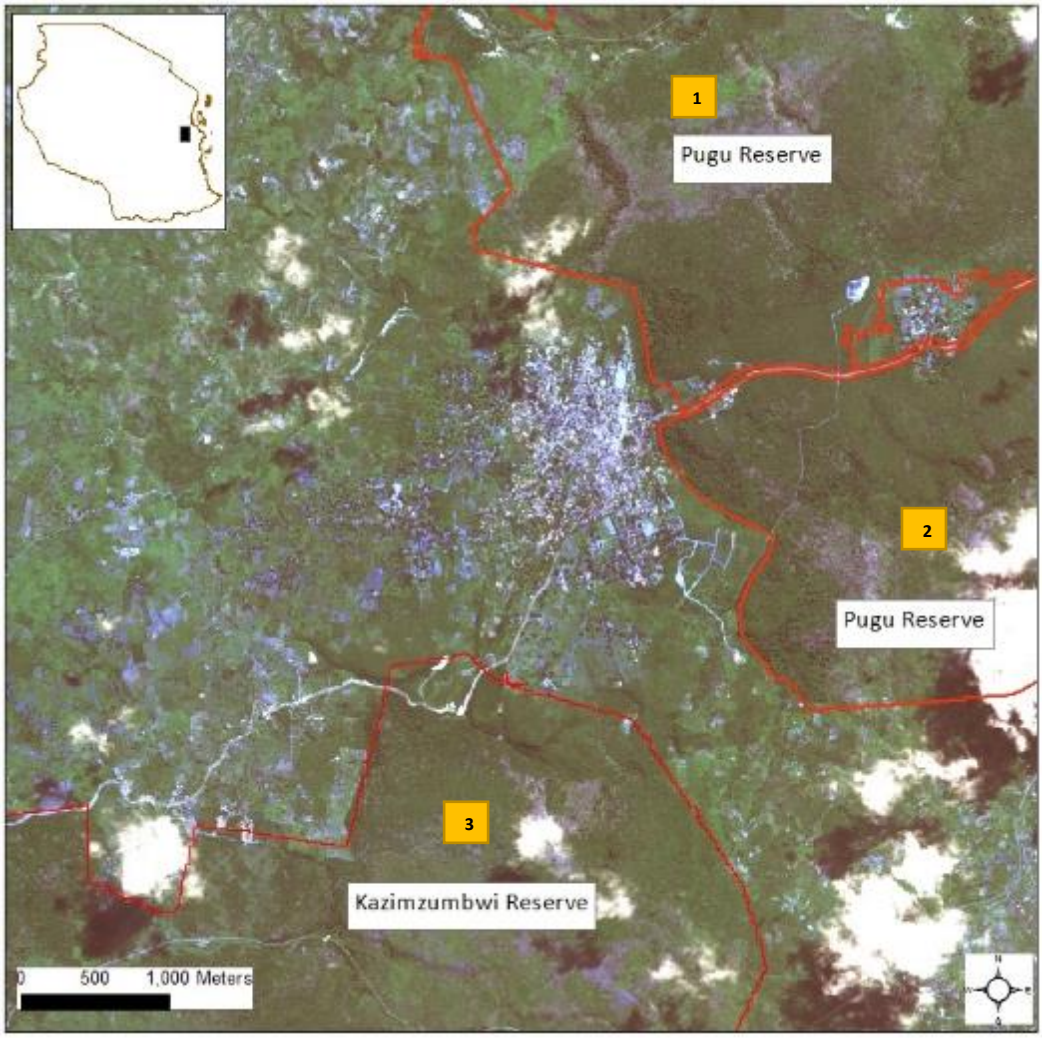

Figure 19: Worldview-2 image (7x7 km) of the primary filed site. In the centre there is the town of Kisarawe, at N-E the Pugu Forest Reserve and at S-W the Kazimzumbwe Forest Reserve. The limits of the forest reserves are approximated. The purple squares indicate three main areas of forest degradation.

5.9 Field measurements

Vegetation type

A vegetation classification scheme was developed to take into account the national (NAFORMA) classification. The classification is purely physiognomic, which means that discrimination between vegetation types is based on relative appearance, in terms of strata, canopy closure, stature and relative composition of trees, bushes and grass (NAFORMA 2010a), and therefore is appropriate for mapping with remote sensing. The main four classes: forest, woodland, bushland and 'other' are characterised by:

- Forest: three canopy layers (emergent, main and regenerative) and more than 80% of the canopy is covered by trees higher than 5 m. The tropical humid forest is characterised by a

full crown cover. In this type of forest, if the canopy cover is higher than 90% it will be considered closed, if less than 90% probably degraded.

- Woodland: two strata (one has trees with normally single stems, from which timber products can be extracted, and the other usually grassland) and the canopy cover is between 20 and 80%. If the canopy cover is higher than 40% it will be considered closed, if less than 40% open.
- Bushland and shrubs: a wide range of canopy densities and a height of less than 5 m. Multi-stemmed plants from a single root base are predominant, with the possibility of emergent trees.
- Other classes: include bare soil, grassland, urban area, water and cultivated land.

The main difference between this scheme and the NAFORMA classification is the minimum canopy cover of forest, which for the work needed to be consistent with remote sensing data for the humid forest. This has been defined here as 80% cover. Table 6 shows the main characteristics of the vegetation classes.

Table 6: Characterization of the main vegetation classes (forest, woodland and bushland) according to number of layers, canopy cover, height and number of stems.

	Layers	Canopy cover		Height	N°. stems	Other characteristics
Forest	3	80-100%	>90% intact	>5 m	Single	*The tropical humid forest has full crown cover
			<90% degraded*			
Woodland	2	20-80%	>40% closed	>5 m	Single	One strata composed by trees with single stems
			<40% open			
Bushland	1	Wide range	dense	<5 m	Multi	Possible presence of emergent trees
			sparse			

Figure 20 shows some examples of vegetation types as imaged by the Worldview-2 satellite. The different classes have not only different percentages of vegetation cover, but also different reflectance (colour) and spatial distribution (texture) properties that can be observed at this high spatial resolution. The latter properties are related with the structure, composition and maturity of the vegetation. For example, in a mature humid forest (top left picture, Figure 20 the crowns of the emergent trees can be well differentiated and project a shadow over the main canopy layer. In the study area degradation usually starts when these larger trees are extracted for fuel wood or timber,

leaving holes in the forest canopy (top right picture, Figure 20). If degradation continues, the layer of larger trees is further depleted (*e.g.* for fuel wood or charcoal production) and in a short time a homogeneous layer of shrubland will cover the area (bottom left picture, Figure 20).To aid in the vegetation classification of the image, the undergrowth vegetation type was classified with the following classes: no vegetation, bushes, grass, herbs and new tree regeneration (regrowth).

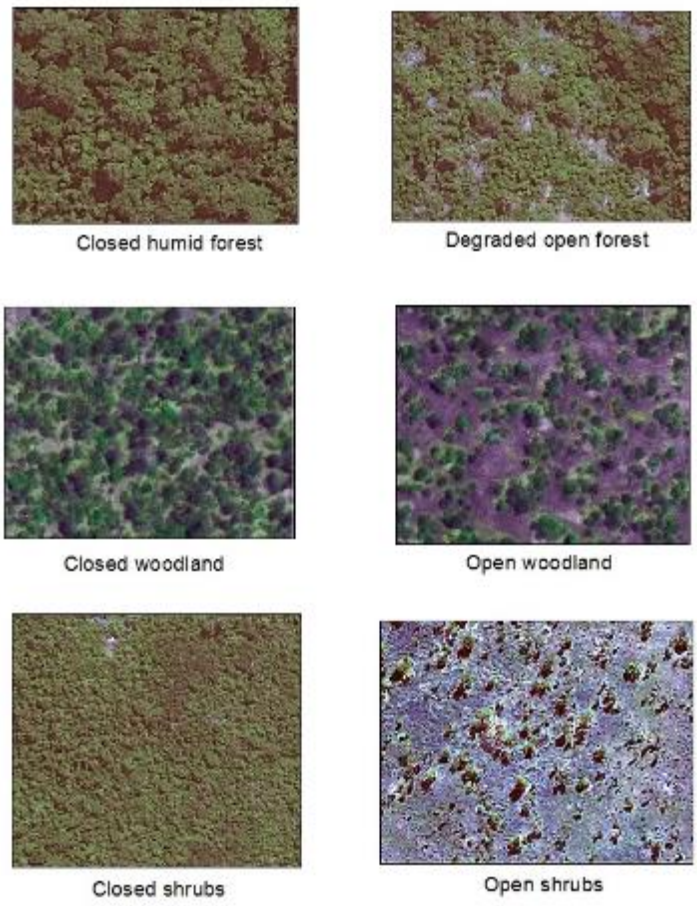

Figure 20: Examples of vegetation classes extracted from Worldview-2 images.

Canopy cover

The vegetation classes with woody biomass (forest and woodland) were sub-divided in classes according to the percentage of canopy cover; <20%, 20-40%, 40-80%, 80-90% and >90%.

Carbon stock

Therefore, to estimate the aboveground biomass of a forest, apart from the area covered by woody vegetation, one needs to know the volume of the trunks, which can be estimated from the diameter at breast height and the height of the trees.

These two variables are generally correlated for a single type of forest, so sometimes one of them can be enough. In the classification adopted, woody vegetation was then sub-divided according to diameter at breast height (DBH), height and the resulting aboveground biomass (AGB) in the following ranges shown in Table 7. In an attempt to be more precise in the estimation of the biomass, the crown diameter and tree height were also measured.

Table 7: Classes of woody vegetation according to height, diameter at breast height (DBH) and aboveground biomass (AGB).

Height (m)		DBH (cm)	AGB tC/ha
<2	Shrub	<5	0-5
>2		5-10	6-20
5-10	Low forest	10-20	21-50
10-15	Medium forest	20-40	51-100
>15	High forest	>40	>100

5.10 Sampling method

In the field work carried out in the study, the plot sample design was aimed to be compatible with that of NAFORMA. For this reason the sampling units had the same shape, circular plots, so the computation of the field measurements was similar.

Plot selection

The location of the field plots were visually pre-selected in ArcGIS using the very high resolution satellite image according to different texture and colour. A preliminary segmentation of the image was carried out in eCognition software to delimitate the image into landscape units (Figure 21). A multi-resolution segmentation algorithm was used, which minimizes the average heterogeneity of the image objects according to its spectral information and spatial homogeneity. An importance weight of 20% to the first and 80% to the second criteria was given, to avoid highly fractured image object results which can result from a strongly textured data (Trimble 2011), as with the pan-sharpened 0.46 m spatial resolution Worldview-2 image.

The resulting segments had different size, according to homogeneity, and a mean mapping unit of 0.5 ha (with a standard deviation of 0.4 ha), in accordance with the national forest area definition. Potential plots had to fall inside segments which fulfilled three conditions: 1) to have a

homogeneous texture and colour in all its area, 2) to be close to and accessible from a road or foot path and 3) to have no cloud cover.

The potential plots were classified into different strata of vegetation type according to Table 6. The stratification by forest type increases survey efficiency by reducing unnecessary sampling while ensuring that major vegetation variations of vegetation type are captured. When implementing the field survey, not all the potential plots can be measured for many reasons (accessibility, land cover change since image acquisition …), but it is important to maintain a well-balanced number of plots per strata to obtain a good diversity of vegetation types. Orthomaps were prepared for each site (Figure 22).

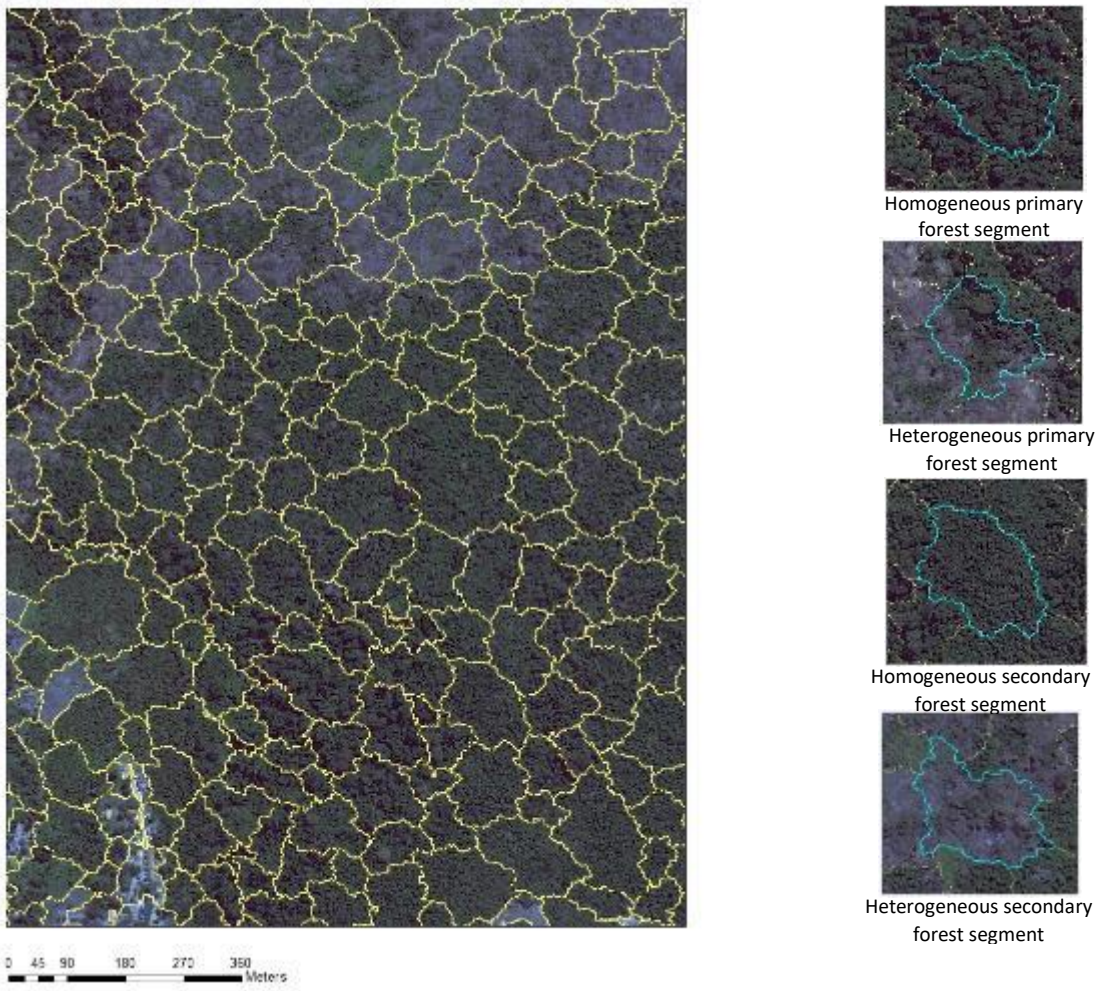

Figure 21: Segmentation of a Worldview-2 satellite image subset (1.0 x 1.2 km) for the selection of potential plot locations. On the right, examples of homogeneous segments with primary and secondary forests (suitable for plot selection) and heterogeneous segments with primary and secondary forests (not preferred for plot selection).

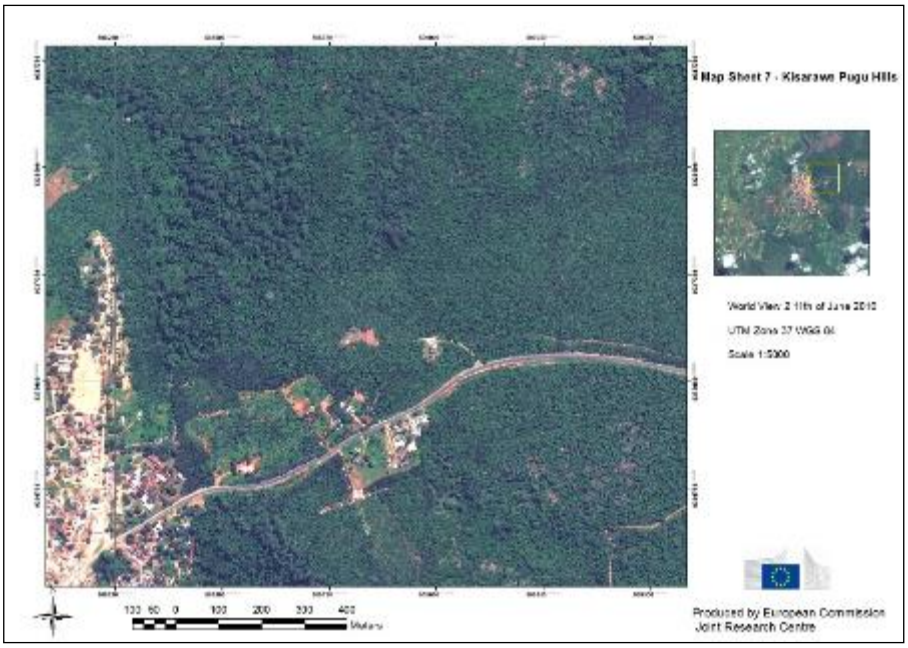

Figure 22: Example of an orthomap sheet of the Kisarawe study area, covering part of the Pugu Forest Reserve (area in the yellow square on the full map at the right), based on a Worldview-2 satellite image.

Figure 23: Example of photos taken at a plot centre, with the plot identification number, and in the cardinal directions (left) and a scheme of a plot (right). The blue circles indicate the points where the canopy cover was measured.

5.11 Analysis of field data

Waypoints and plots

A total of 184 waypoints were collected and 22 field plots measured during the field mission in 2012 (Figure 24); 14 in the primary field site in the humid forest (Figure 25), 4 in the general field sites in the dry and Miombo forests, and 5 in the general field site in the semi-humid forest.

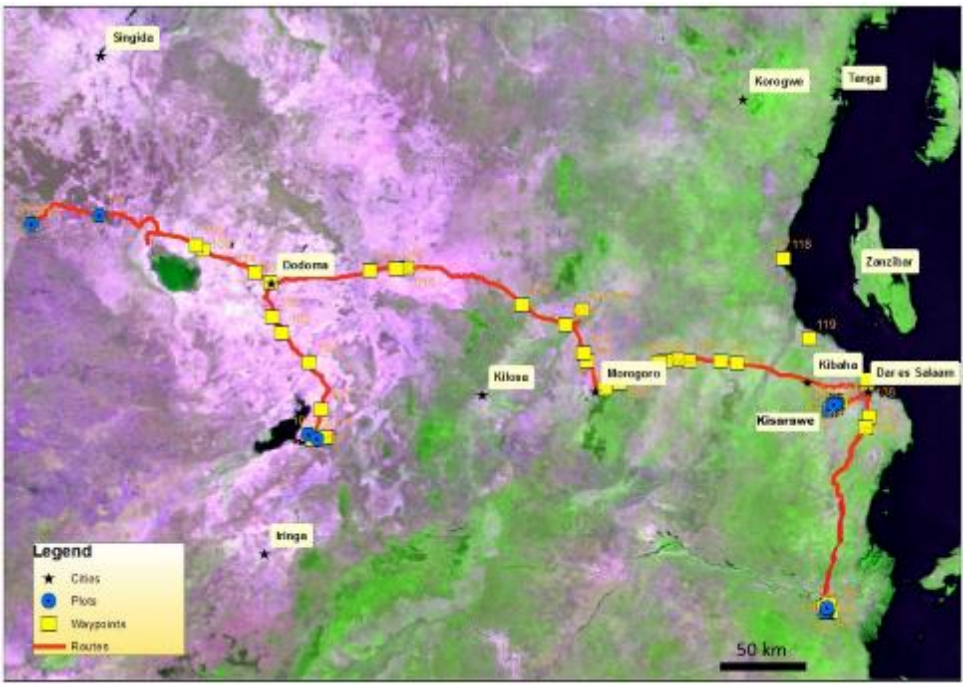

Figure 24: Route followed in the field work of 2012 with waypoints and plots. The two groups of plots on the East belong to the field sites in the humid (North) and the semi-humid (South) forests. The two groups on the West to the field sites in the dry and Miombo forests (respectively from East to West).

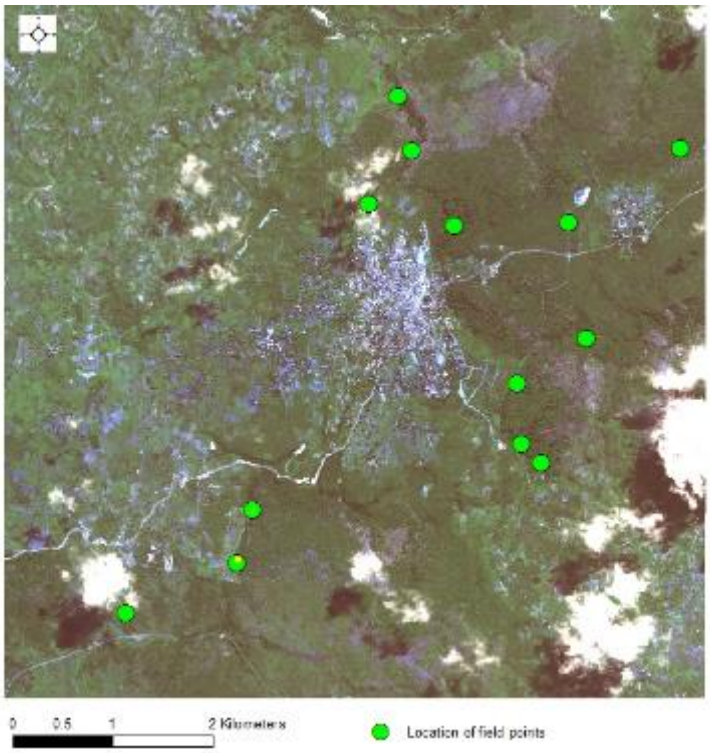

Figure 25: Location of the plots in the primary field site (in Kisarawe).

Plot parameter calculations

The data from the plots were used to calculate a range of forest parameters as shown in Table 8.

Table 8: Plot parameters derived from the field measurements.

Field measurement (units)	Plot parameter (units)	Formula / calculation	
Slope (%)	Plot corrected area (m^2)	$\pi * R^2$	R (m) is the corrected radius for slope >5%
Canopy cover (%)	Average canopy cover (%)	Mean of five measurements (C, N, E, S, W)	
Woody plants count	N° woody plants	Stems with DBH >5cm	
	N° trees	Stems with DBH >5cm and height >5m	
DBH (cm)	Average DBH (cm)	DBH >5cm	
	Standard deviation DBH (cm)	DBH >5cm	
	Basal area per ha (m^2)	$\sum_{i=1}^{N} \frac{0.785 * dbh_i^2}{\pi * R^2}$	dbh_i is the DBH (cm) of plant i, N the total number of plants and R (m) the radius
Height (m)	Average height (m)	DBH >5cm	
	Standard deviation height (m)	DBH >5cm	
DBH and height	Wood volume per ha (m^3)	$\sum_{i=1}^{N} \frac{0.392 * H_i * dbh_{ix}^2}{\pi * R^2}$	H_i is the height (m) of plant i, dbh_{ix} the DBH (cm) of stem x of plant i and R (m) the radius
	Aboveground biomass (t/ha)	Wood volume (m^3/ha) * wood density (0.5-0.58 t/m^3)	
	Aboveground carbon content (tC/ha)	Aboveground biomass (t/ha) * carbon content (0.5 tC/t)	
Species	N° species (S)	N° of different species	
	Species diversity index	$\frac{S}{N}$	S is the total number of species and N the total number of plants
	Margalef index	$\frac{S-1}{\ln N}$	
	Shannon index	$-\sum_{i=1}^{S} pi * \ln_2 pi$	p_i is the relative quantity of specie i, $p_i=(s_i/N)$ where s_i is the number of individuals of species i

5.12 Plot data results

Vegetation structure and biomass

For each plot a synthesis of the field data calculations and observations was produced, including field photos Figure 23. Table 9 shows the synthesis of all the plots, from the primary field (humid forest) and from the general field sites (dry, semi-humid and Miombo forest). In total they sum up 23 plots, covering a total area of 7.225m$^{2.}$

Allometric equations

The number of total stems measured in the humid, dry and semi-humid biomes was 172, 76 and 27 respectively. The relationship between height, H, (m) and DBH (cm) in the humid biome was:

$$H = 11.182 * \ln dbh - 17.323 \ (R^2 = 0.75)$$

This relationship was used to estimate tree height for sites with stumps from trees cut after the satellite image acquisition date (as in plots n°7 and n°8).

Crown diameter was difficult to measure with the laser rangefinder. Instead, allometric equations were obtained from the literature and adapted to Tanzania (after O'Brien *et al*. 1995).

$$\log(CA) = 1.3 * \log(dbh * 10) - 2.0$$

Ideally by measuring the crown area (CA) directly from the image, one can obtain the DBH, which is related to the height for a same type of forest, and therefore the volume.

5.13 Species distributions

The distribution of the species changes by biome. From the humid to the semi-humid and dry there is a dominance of fewer species of the 50% of the total number of tress (Figure 26, Figure 27 and Figure 28).

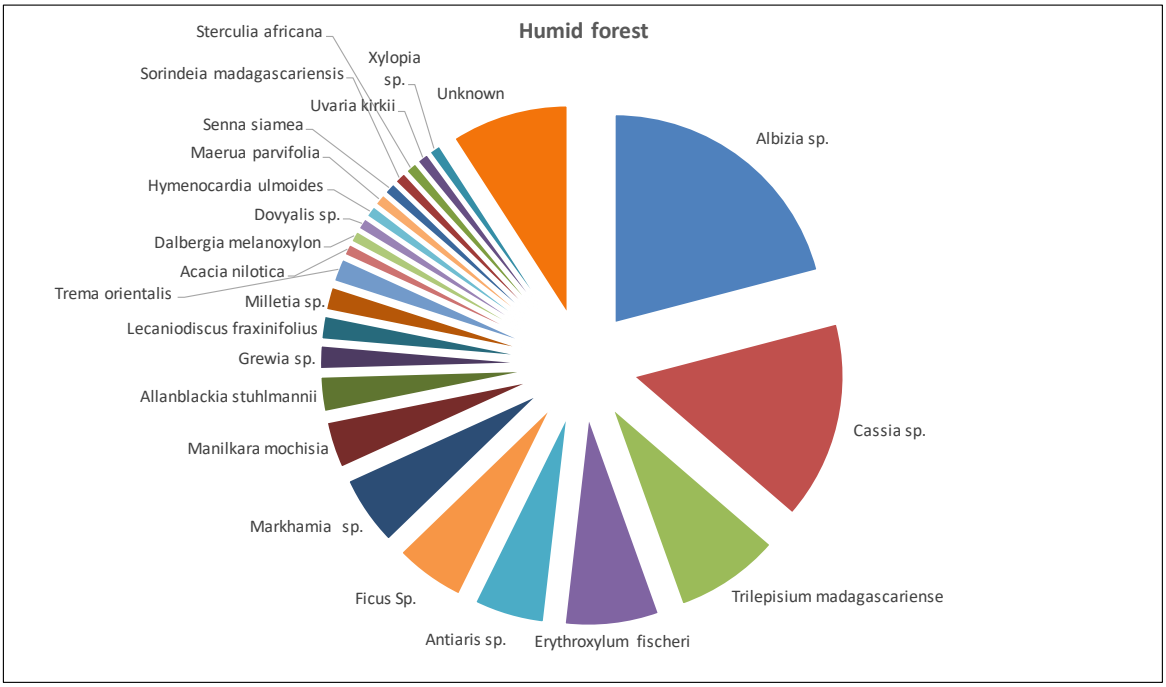

Figure 26: Species distribution within plots sampled in the humid forest.

Table 9: Synthesis of the field data calculations for all the plots.

Plot no.	Forest type	Location	Vegetation (and degradation)	Longitude DMS	Latitude DMS	Canopy cover (%)	No. of woody species	No. of tree species	Average DBH (cm)	SD DBH (cm)	Basal area per ha (m^2)	Average height (m)	SD height (m)	Volume per ha (m^3)	AGB (t/ha)	AGC (tC/ha)	Margalef index**	Disturbance
1	Humid	Kisarawe	Degraded forest, shrubs and emergent trees	39° 5' 51.6" E	6° 53' 36.9" S	55	14	14	12.1	3.8	6	8.1	2.7	23	13	7	2.27	Fire
2	Humid	Kisarawe	Lowland forest	39° 4' 37.7" E	6° 54' 2.1" S	100	12	12	33.2	23.0	38	20.8	12.6	651	377	189	0.40	No damage
3	Humid	Kisarawe	Degraded forest, shrubs and emergent trees	39° 6' 1.9" E	6° 52' 52.0" S	25	12	3	6.9	2.0	2	4.6	1.4	4	2	1	1.21	Wind and clear felling
4	Humid	Kisarawe	Lowland forest	39° 5' 20.6" E	6° 54' 39.0" S	95	21	21	20.9	7.4	26	15.5	7.1	234	136	68	0.66	Flood Illegal cutting
5	Humid	Kisarawe	Bushland dense	39° 5' 14.0" E	6° 54' 0.7" S	100	15	0	5.9	0.7	7	3.9	0.4	13	8	4	0.62	No damage
6	Humid	Kisarawe	Degraded forest, shrubs and emergent trees	39° 4' 7.6" E	6° 53' 54.6" S	40	12	6	7.6	2.4	2	4.7	1.0	5	3	1	0.40	Grazing
7	Humid	Kisarawe	Clearing in forest*	39° 4' 23.6" E	6° 53' 37.2" S	100	7	6	21.4	11.7	11	17.1	7.8	111	65	32	1.03	Clear felling and charcoal production
8	Humid	Kisarawe	Clearing in forest*	39° 4' 18.7" E	6° 53' 19.7" S	100	9	4	28.1	14.7	20	18.7	6.8	234	136	68	1.82	Clear felling, selective cutting
9	Dry	Mtera	Woodland closed	35° 57' 12.9" E	7° 4' 9.8" S	50	9	8	21.5	1.7	28	6.8	1.7	115	58	29	2.17	Grazing
10	Dry	Mtera	Woodland open	36° 0' 30.7" E	7° 5' 32.6" S	55	18	8	8.4	4.2	9	4.4	1.1	24	12	6	1.70	Selective cutting
11	Miombo	Itigi	Woodland closed	34° 43' 26.1" E	5° 45' 43.4" S	95	32	25	10.6	4.5	16	6.4	1.5	53	27	13	2.31	No damage
12	Miombo	Itigi	Woodland open	34° 19' 23.5" E	5° 48' 40.6" S	60	17	8	20.4	17.7	37	7.0	3.6	204	102	51	1.80	No damage
13	Semi-humid	Ikiwiri	Lowland forest	39° 2' 17.8" E	8° 7' 52.3" S	100	5	5	28.7	26.4	32	20.6	6.2	326	163	82	1.86	No damage
14	Semi-humid	Ikiwiri	Clearing in woodland	39° 2' 19.1" E	8° 7' 51.5" S	40	2	0	7.1	2.3	1	4.0	0.0	2	1	1	0.00	Tree cutting
15	Semi-humid	Ikiwiri	Woodland closed	39° 1' 41.9" E	8° 7' 42.2" S	50	12	10	20.7	14.7	24	6.9	3.2	113	56	28	2.82	No damage
16	Semi-humid	Ikiwiri	Woodland open	39° 2' 37.8" E	8° 6' 29.4" S	40	3	3	28.3	2.9	8	14.0	1.4	57	29	14	0.91	Grazing
17	Semi-humid	Ikiwiri	Woodland open	39° 2' 10.0" E	8° 6' 24.0" S	30	5	5	25.5	9.3	9	9.6	0.0	47	23	12	0.62	
18	Humid	Kisarawe	Lowland forest	39° 3' 31.0" E	6° 55' 34.9" S	100	17	17	28.6	19.7	49	23.3	12.8	840	487	244	1.41	No damage
19	Humid	Kisarawe	Bushland open with emergent trees	39° 3' 26.1" E	6° 55' 52.2" S	100	6	0	7.0	2.0	2	4.0	2.0	4	2	1		
20	Humid	Kisarawe	Degraded forest, shrubs and emergent trees	39° 2' 53.7" E	6° 56' 8.4" S	90	5	5	23.0	23.6	12	5.5	8.7	137	79	40	1.86	Damage unknown
21	Humid	Kisarawe	Lowland forest	39° 4' 59.6" E	6° 55' 13.3" S	100	20	20	25.6	17.7	39	17.2	11.8	570	331	165		

22	Humid	Kisarawe	Degraded forest, shrubs and emergent trees	39° 5' 6.6" E	6° 55' 19.3" S	90	20	5	10.5	5.4	7	6.8	3.2	32	18	9		
23	Humid	Kisarawe	Degraded forest, shrubs and emergent trees	39° 4' 58.3" E	6° 54' 53.6" S	95	12	4	10.3	4.8	6	5.4	2.6	21	12	6		

*Plots n°7 and n°8 values were simulated on the basis of the stumps. In plot n°8 3 out of 9 trees were still present. ** Only the biodiversity index which gave us better results, Margalef index, is shown in the table.

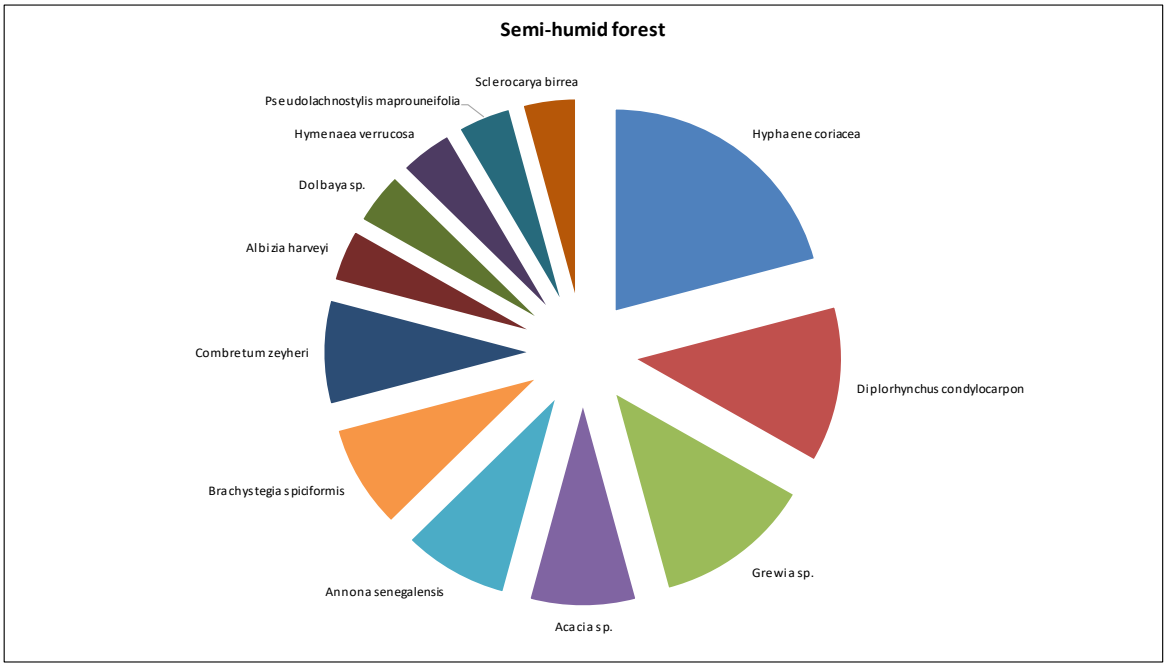

Figure 27: Species distribution within plots sampled in the semi-humid forest.

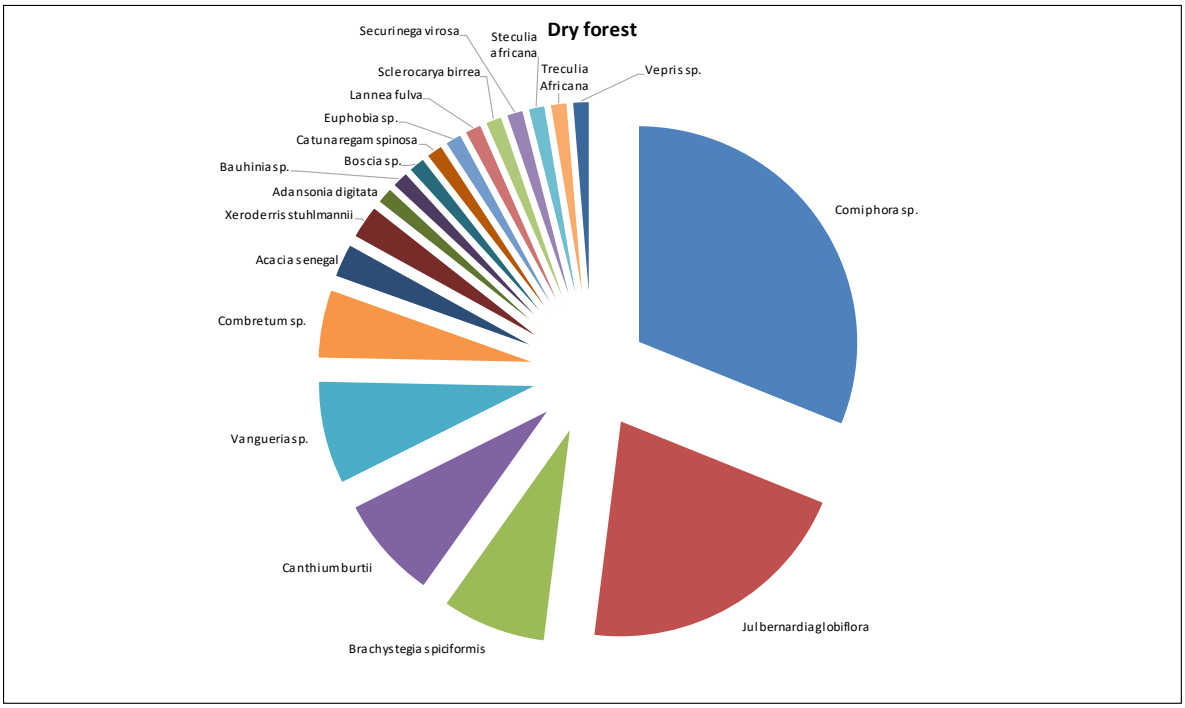

Figure 28: Species distribution within plots sampled in the dry and Miombo forests.

Biodiversity indices were calculated also at biome level, so that changes in biodiversity due to degradation at country level could be monitored in relation to the type of forest (Table 10). The results are just an indication as there are not enough plots to characterize the biome's biodiversity. The humid forest had the

highest values, followed by the dry forest. These values can also be affected by the fact that the species were recorded by different persons with different level of experience in each of the three biomes. In the dry forest all the species were identified, while in the humid and semi-humid forests 9.2% and 11% of the trees respectively were not recognised. In these cases, all the unknown trees were grouped in a same class, which decreased the biodiversity values.

Table 10: Biodiversity by biome.

	N° trees	N° species	Margalef index	Shannon index
Humid forest	109	23	5.11	1.15
Dry forest	77	19	4.14	0.97
Semi-humid forest	24	12	3.33	0.94

5.14 Vegetation and forest maps production

Vegetation map

A basic classification vegetation map with a minimum mapping unit of 0.5 ha was created with the support of the waypoints, photos, vegetation type and undergrowth data from the field survey and visual interpretation of the satellite image (Figure 29).

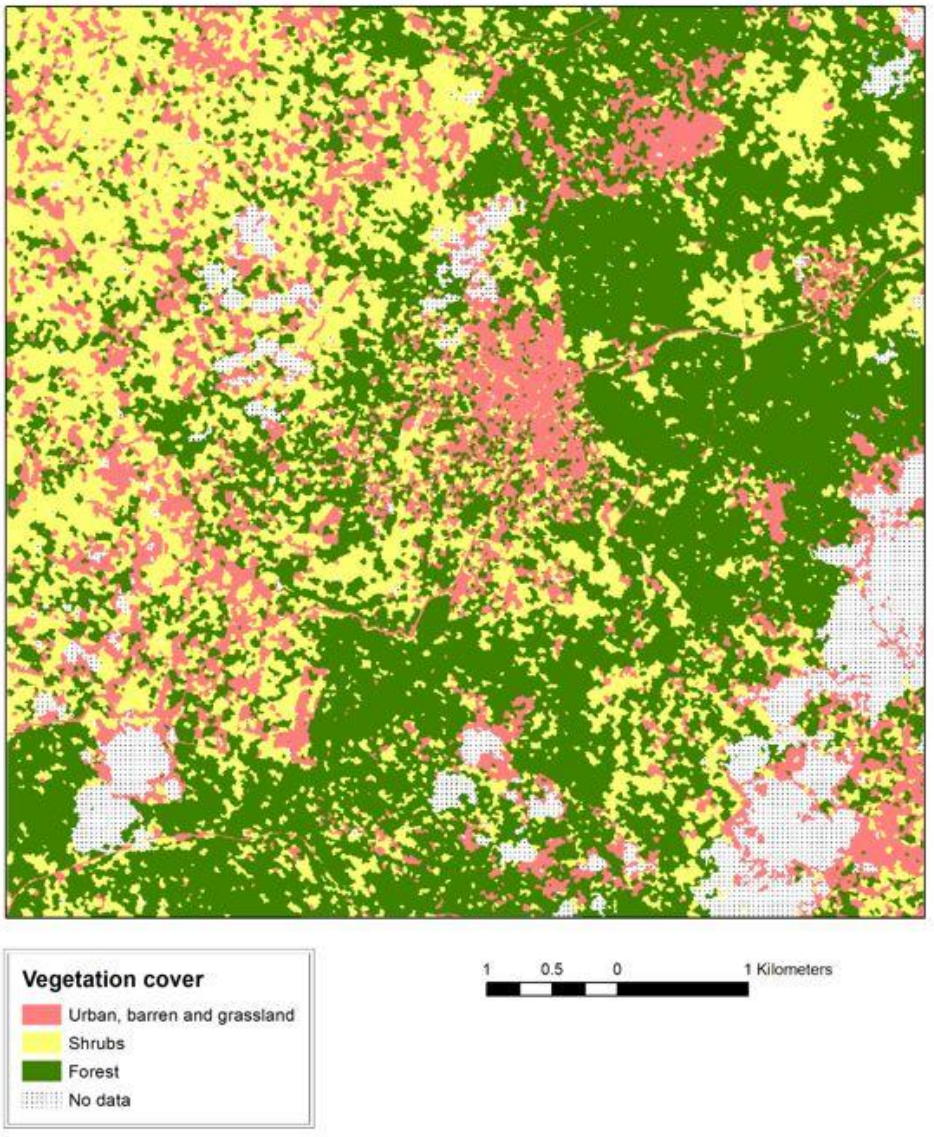

Figure 29: Vegetation map.

5.15 Forest maps

Forest areas as identified on the previous vegetation map were classified by biomass related parameters (basal area, height and aboveground biomass) by extrapolating the best regression models obtained from the reflectance and texture information of the image and the field plot data. The changes in these maps are aimed to indicate forest degradation processes.

5.16 Selection of remote sensing parameters

Image reflectance information is normally used to produce forest classification maps from satellite data. Either indices based on single spectral band information, simple spectral band relations and/or more complex vegetation indices are employed. This type of information have achieved moderate success in tropical and subtropical regions for biomass estimation, where biomass levels are high, forest canopies are

closed, with multiple layers, and there is a great diversity of species (Lu 2005). According to some studies (Sarker 2011), texture measures correlate better with biomass than spectral parameters, especially for degraded forests. A combination of both can improve the results (Eckert 2012).

To evaluate the texture of image objects there are two types of texture features: 1) texture features based on analysis of sub-objects, useful for highly textured data, and 2) texture features based on the grey level co-occurrence matrix after Haralick *et al*. (1973). Due to the high level of segmentation of the image, which resulted in small and quite homogeneous objects, the texture features of Haralick were employed. Overlaying the field plots on the segmented image, the image objects containing field plots were selected and a set of pre-selected spectral and textural parameters listed in Table 4 were extracted using the image processing eCognition software. Potential correlations between these parameters and the field plot parameters were then examined.

Table 11: Best Pearson correlation coefficients (r) between the remote sensing parameters and the field parameters with the significance level (p-value).

	Average DBH		Average height		Basal area		AGB		Volume		SD DBH		SD Height	
	r	p-value	r	p-value	r	p-value	r	p-value	r	p-value	r	p-value	r	p-value
SD NIR/SD Red	0.92	1.83E-5	0.88	7.55E-5	0.79	4.72E-3	0.73	1.83E-5	0.73	4.47E-3	0.77	1.86E-3	0.73	4.71E-3
GLCM Angular 2nd moment	0.91	3.45E-5	0.88	1.10E-4	0.74	6.71E-3	0.71	3.45E-5	0.71	6.43E-3	0.69	9.50E-3	0.65	1.58E-2
GLCM Homogeneity all directions	0.90	3.06E-5	0.89	4.5E-05	0.70	1.22E-2	0.66	3.06E-5	0.66	1.39E-2	0.68	1.04E-2	0.62	2.46E-2
GLCM Entropy	-0.9	9.67E-6	-0.87	8.79E-5	-0.72	4.07E-3	-0.69	9.67E-6	-0.69	8.63E-3	-0.70	7.14E-3	-0.63	1.97E-2

Multiple regressions

The variability in the field measurements that could be explained by the remote sensing parameters was estimated by multiple regressions in the statistical software R (R Core Team 2012).

Multiple regression models were calculated for only those field parameters which showed the best correlations, *i.e.* DBH, tree height, basal area and biomass. To select the 'best' model, the Akaike Information Criterion (AIC) was used, which is a measure of the relative quality of a statistical model for a given data set. It is based on the trade-off between the fit of the model and the complexity of the model. That is, it rewards the goodness of fit, but includes a penalty which is an increasing function of the number of independently adjusted parameters, so it discourages over-fitting. It is calculated with the following function: AIC = 2k - 2ln(L), where k is the number of parameters in the model and L the maximum value of

the likelihood function of the model (Akaike, 1974). Table 12 shows the best models for each of the field parameters (linear for DBH and height and exponential for basal area and biomass).

Table 12: Best fitting equations for the multiple regressions models for DHB, height, basal area and biomass with the remote sensing parameters.

	Best fitting equation	R^2	p-value
DBH	41.16 +230300 GLCM Angular2 -0.032 Mean 4 -0.63 SD2	0.95	5.76e-7
Height	76.56 +186000 GLCM Angular2 -0.026 Mean4 -0.6 SD2-53.47 NDVI	0.93	3.77e-5
Basal Area	EXP(-111.8 +37.81 GLCM H45 +2.046 GLCM SD2 -0.0077 Mean4 +35.74 NDVI)	0.86	3.00e-4
Biomass	EXP(-179+ 70.5 GLCM H45 +3.201 GLCM SD2 -0.012 Mean4 +59.01 NDVI)	0.84	5.56e-4

Calibration of the satellite image

The equations derived from the multiple regressions were then applied to all the forest polygons from the vegetation map to produce maps of tree height, basal area and biomass. Previously, the four remote sensing parameters included in the models had to be retrieved from all the polygons. The regressions are not fully robust, in that the variance of the models estimations can be high. The maps were then visually inspected to remove obvious misclassifications.

Map of tree height

After masking non-woody areas (grass, barren and urban) and no-data (clouds and cloud shadow) the height parameter was used to classify the polygons into shrubs (<5 m) and trees (>5 m). Within these two categories, polygons were further divided into two classes of shrubs (<2 m and >2 m) and three of tree height (5-10 m, 10-15 m and >15m) (Figure 30).

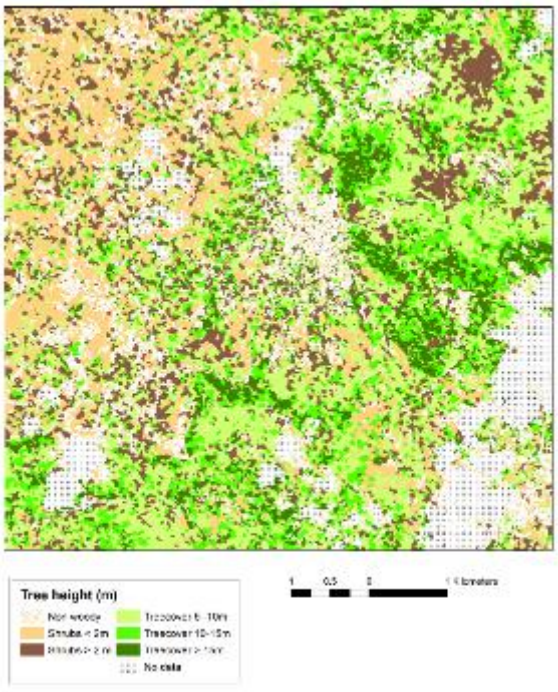

Figure 30: Map of tree height

Map of basal area

A map of basal area was produced, rather than a map of DBH, as it is closely related to forest management.

Within the forest class four classes of basal area are mapped (Figure 31).

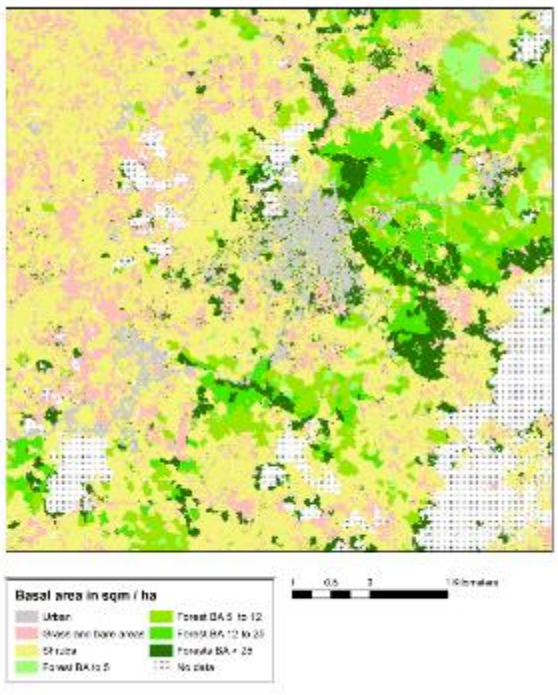

Figure 31: Map of basal area (BA).

Map of aboveground biomass

Finally a map of aboveground biomass was produced for the study area (Figure 32). Note that there is a low representation of the highest biomass class. This is because few areas of pristine forest remain, mainly on steep inclines, with maximum biomass values of 250 tC/ha. This is in line with other studies, such as Lewis *et al.* 2013, where closed-canopy tropical forest across 12 African countries had on average 197tC/ha. In this case however, most of the sample sites were located in Central Africa.

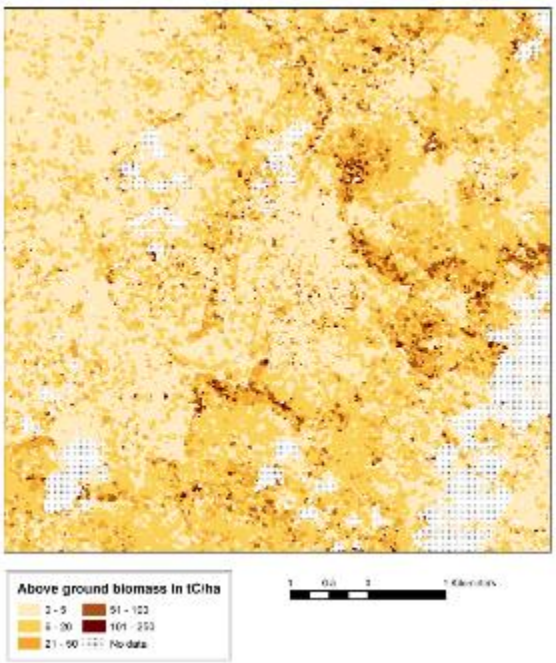

Figure 32: Map of aboveground biomass in tC per hectare.

5.17 Validation

Validation data were collected in a short field mission of two days. To obtain a rapid validation data set, rather than repeating the full field data collection, which has considerable time requirements, only the average canopy height was measured at 31 random locations across the primary field site in Kisarawe (Figure 33). Ideally, average tree DBH and average tree height should have been calculated to validate the maps of tree basal area, tree cover height and aboveground biomass, but these parameters require a lot of time to take many single tree measurements. As an alternative, the average canopy height was estimated once per location from the outside of the plot with the laser rangefinder. Highly significance inter-correlation between mean tree height and estimated canopy height was found in a data analysis of 50 plots between 0.012-0.25 ha along the coast of Tanzania (covering a total area of over 4.4 ha), primarily used to produce transect diagrams (Burgess, 2000). There were some difficulties to measure the average canopy height, for example when forests had multiple tree layers or the centre of the crown was obscured.

Nevertheless, the estimated average canopy height data were used to assess the general validity of the map production approach.

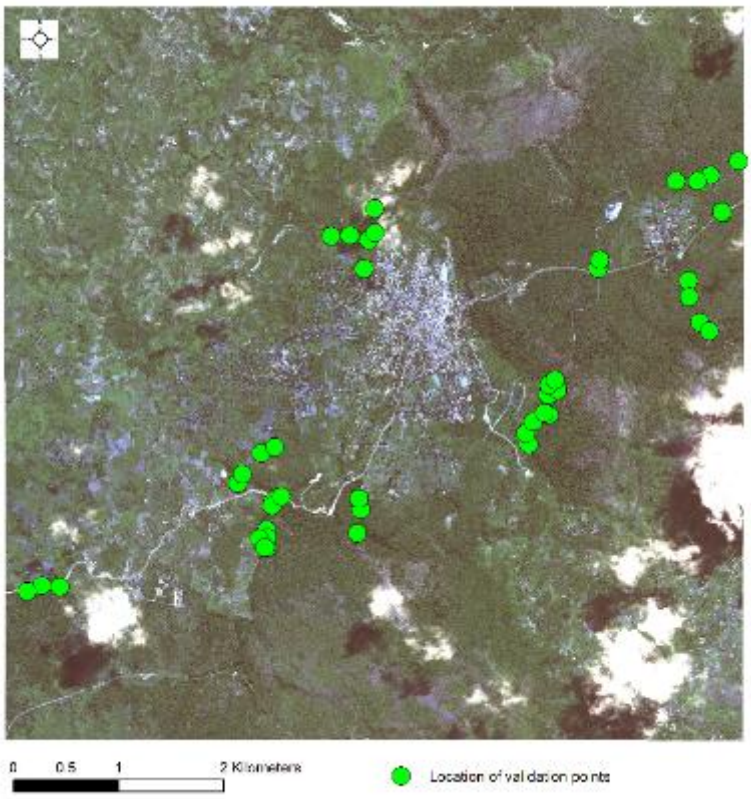

Figure 33: Location of the validation data on average canopy height.

For 31 validation locations, the average tree height was calculated from the model of tree height (Table 12). Other validation locations were not used because on the image either they were in mixed cover polygons (*e.g.* trees and grass or stands of different height), or had atmospheric effects (*e.g.* fine haze or shadow), that had a strong influence on the remote sensing parameters extracted.

A confusion matrix was made for the correspondence of measured average canopy height (validation) and modelled average tree height (map) (Table 13), using the four tree cover height classes (≤5 m, >5 and ≤10m, >10 and ≤15m and >15m). It resulted that 55% of the points had been correctly classified, 6.4% overestimated (error of commission) and 39% underestimated (error of omission) in the map regarding the validation data. The model tended to underestimate canopy height from class 5-10 m onwards.

Table 13: Contingency table between the field measurements on canopy height (validation) and the modelled results on tree cover height (map).

	Validation					
Map	**n=31**	**≤5 m**	**5-10 m**	**10-15 m**	**>15 m**	**Sum**
	≤5 m	8	3	0	1	12
	5-10 m	2	5	4	1	12
	10-15 m	0	0	1	3	4
	>15	0	0	0	3	3
	Sum	10	8	5	8	31

The possible errors in the classification could also be due to the way the validation data were obtained. Measuring the average canopy height from the outside of the plot can lead to an overestimation of the real average tree height, especially when forests reach certain height (>5 m) and form several layers. In these cases, if the canopy is too close, the low layers are covered by the top layers and not taken into account in the estimation. On the other hand, these shorter trees were measured in the field survey, decreasing the tree height average of the plots. This overestimation was also observed in the previously mentioned study in (Burgess and Mbwana, 2000). Also, obtaining an 'average' canopy height is problematic when the mapping unit is large, as there may be a large variance within the land units.

5.18 Conclusions

Utility of the study

To monitor forest degradation for the REDD programme, we need to monitor changes in the forest canopy cover, and also estimate changes in the forest carbon stock caused by human influences. The utility of very high resolution satellite data for deforestation monitoring has been demonstrated with respect to Landsat data. In this study, forest units were mapped at 0.5 ha, with a series of forest parameters (tree height, basal area and above ground biomass) which can be used in the management and carbon stock reporting of the Pugu Hills forests. These characteristics cannot be mapped with confidence from the widely available Landsat sensor.

This forest mapping approach is based mainly on vegetation structural parameters. Environmental factors affect the vegetation structure in the forest, but human disturbances also have an important influence on it, reducing parameters such as basal area and stem density. For instance, those of our plots where some type of human disturbance was recorded exhibited lower basal area (Table 8). The study of Burgess (2000)

suggests that human disturbances have a major influence in the structure of the costal forest of eastern Africa, and moreover that these disturbances remain detectable in the forest structure for many years. Disturbed forests have a low basal area and stem density even after 30-80 years the disturbance took place. Therefore, measurement of structural parameters is a good approach for detecting forest degradation by human disturbances, but changes of these parameters have to be monitored at long time scales.

In addition to this, structural parameters have the advantage that they are detectable and measurable using remote sensing techniques. In satellite imagery of very high spatial resolution, such as the Worldview-2 data, the textural properties of the forest can be well characterised and even tree crowns identified. As tree crown is related to DBH, the volume of the biomass could be directly retrieved using allometric equations. However, in close forests with overlapping tree crowns, only the top layer is recorded by the satellite images. Therefore, we followed an indirect approach, where the reflectance and texture parameters of the image were calibrated to biophysical variables related to forest biomass.

To calibrate the satellite images field survey is necessary, especially in the humid forest. Heavily disturbed forests that have regenerated freely may appear in pristine conditions, but they still retain evidences of such disturbance in the structural parameters. From field measurements we built the biophysical variables (i.e. basal area, height, volume and biomass) with general allometric equations grouping all species together. Important variations in volume geometry, wood density and carbon content between and within ecological zones and tree species have been reported in the Sub-Saharan African forest allometric equations (Henry 2011). However according to some authors, grouping all species together and using generalized allometric relationships, stratified by broad forest types or ecological zones, is highly effective for the tropics. This is because DBH alone explains more than 95% of the variation in aboveground carbon stocks in tropical forest, even in highly diverse regions (Brown 2002, Gibbs et al. 2007).

Limitations

In the field survey only the variables related to aboveground biomass were measured, which is usually the largest carbon pool in the forest and the most directly affected by deforestation and degradation. However, belowground biomass can represent an important portion of the total forest biomass, especially in dry areas. Measuring the biomass of the rooting systems is very costly and time consuming. Fractal geometry can be a good proxy to overcome the problems arising from the sampling of belowground tree biomass (Hairiah et al. 2001). For example, there is an inventory of available root-to-shoot ratios for the entire tropical domain created by the IPCC (2006). Apart from this, soil properties (e.g. fertility, compactness and salinization) could complement the study, as they can indicate an early stage of degradation which is not visible in the vegetation structure.

When using single date satellite images for mapping forest structural parameters, values cannot be compared to historical references. This is often unavoidable in the case of high spatial resolution satellite data, which became available only in the recent years (e.g. RapidEye satellite was launched in 2009) or which are difficult to find over certain areas (e.g. very high spatial resolution Worldview-2 data). For such cases, environmental factors should be taken into account in order to relate low values of basal area, height or biomass to degradation processes induced by human activities. This is because vegetation structure in the forest is also affected by factors such as altitude, water availability and edaphic properties. For example, Burguess (2000) found that areas with groundwater and deep soil had large and well-spaced trees, areas with rocky outcroppings and shallow soil had low trees, and areas in higher altitude (with higher rainfall), tall and large trees and intermediate or low density.

Satellite images of very high spatial resolution have the drawback in that they are difficult to find for certain areas, very expensive and cover small areas. Therefore, it is difficult to cover a full country with such images for national studies, such as in the Tanzania national REDD initiative. They can be useful however to calibrate satellite images of lower spatial resolution, but higher acquisition frequency and broader coverage, as the 10 m spatial resolution RapidEye data.

Due to time restrictions the total number of plots measured in the field survey was low. The reduced calibration dataset limited the significance of the correlations between the field measurements and the remote sensing parameters, and the robustness of the models to construct the forest structural maps. Moreover, the lack of sufficient data on species compositions precluded the analysis of the biodiversity indices with field measurements and remote sensing parameters.

When carrying out the field survey it is difficult to avoid an over influence of human activities, as the most accessible areas for sampling are close to roads or paths and therefore more likely to be used for fuel wood collection or charcoal production. However, this is not a main problem in this study, as the field survey was not aim at supporting statistical national studies, but characterising the main forest types as seen on the satellite image.

We found that a more structured approach to validation is needed. An ad hoc validation scheme, relying on improvisation in the field results in unreliable data. We found that many of our field measurements (single tree height) did not represent the average tree height in the land unit measured. The validation approach should therefore be reconsidered.

6. Potential improvement for forest cover and forest degradation mapping with the Sentinel-2 program

This chapter contains material of an article published in:

Lorena Hojas Gascón, Alan Belward, Hugh Eva, Olivier Hagolle, Javier García Haro and Paolo Cerutti, 2015, The potential improvement for forest cover and forest degradation mapping with the forthcoming SENTINEL-2 program. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XL-7/W3, 2015.*

This chapter examines whether improved temporal sampling at high spatial resolution (10 to 20 m) promised by the Sentine-2 series can actually improves our knowledge of deforestation and forest degradation in dry forest ecosystems, such as those found in Tanzania. The following were undertaken: 1) estimate the increase of data availability, by comparing the cloud free image area of SPOT4 Take 5 (simulating Sentine-2 acquisitions) with that of Landsat for the same period; 2) evaluate the improvement of the temporal resolution, by comparing the time series of SPOT4 Take 5 with that from MODIS; and 3) estimate the improvement of forest classification accuracy, which results from the different acquisition scenarios with images at every 5 days, 10 days and 16 days (Sentine-2A and 2B; Sentine-2 A; Landsat 8).

6.1 Monitoring deforestation and forest degradation

The addition of forest degradation in the REDD+ program implies that the estimations of forest carbon stock changes need to be based, not only on monitoring transitions of land cover classes (e.g. forest to non-forest), but also on transitions within the forest class when there is a loss of carbon sequestration (e.g. forest with more than 30% crown cover into forest with less than 10% crown cover). In this chapter forest has been considered as an area of land with at least 1 ha and a minimum tree crown cover of 10%, with trees which have, or have the potential, to reach a minimum height of 5 meters at maturity *in situ*, according to the definition adopted by the Tanzanian national REDD+ strategy (UN-REDD 2013).

6.2 The Sentinel-2 program

The European Union's first Earth Observation programme, Copernicus, is building a series of technologically advanced satellites (the Sentinels), which includes the Sentinel-2 satellites. Sentinel-2 aim to contribute providing inputs for services relying on multi-spectral high-resolution optical observations over global land

surfaces, like SPOT and Landsat satellites, but also attempt to cover current limitations with the addition of the technical needs for new requirements. These include higher revisit frequencies, more spectral bands with narrower bandwidths and finer spatial resolutions, in order to improve services as vegetation monitoring (ESA, 2010).

The design of the Sentinel-2 platform benefited from the experience and lessons learned from other satellites building on their technology. The selection of the spectral bands has been guided by the Landsat, SPOT-5, MERIS and MODIS heritage (ESA, 2010). The Multi Spectral Instrument (MSI)'s 13 spectral bands' centre range from 0.433 to 2.19µm. There are four visible and near-infrared bands at 10 m spatial resolution, three red edge, one near-infrared and two SWIR at 20 m, and three channels to help in atmospheric correction and cloud screening at 60 m (Drusch et al., 2012). When in full operation, the two Sentinel-2 satellites offset in orbit operating simultaneously on opposite sides (Sentinel-2A and Sentinel-2B), each carrying the same instruments. Sentinel-2A was launched in June 2015 and Sentinel-2B in early 2017, but is not yet operational. Together these two satellites can provide coverage every five days at the equator with a 290 km field of view (ESA, 2010).

Forest monitoring is one of the priority services of the Global Monitoring for Environment and Security (GMES) programme for which Sentinel-2 has been tailored. In fact, the revisit requirements were driven by vegetation monitoring, for which those of Landsat and SPOT were not enough. Sentinel-2 observations are explicitly intended to develop key inputs required for Kyoto protocol reporting. Potentially, they could contribute to the Baseline Mapping Service for the REDD+ programme (ESA, 2010). The Copernicus plans also aim for multiple global acquisitions and a free and open access data policy (European Commission, 2013), similarly to Landsat.

6.3 Time series for forest cover mapping

Short revisit periods are potentially important to monitor forest at national/regional scales. Firstly, the increased coverage provides more opportunity for acquiring cloud-free images, particularly important in tropical regions (Beuchle et al., 2011). Secondly, because they should allow us to exploit seasonal differences in canopy reflectance characteristics as a means of discriminating between forest cover types and different forest conditions (e.g. closed and open forests, or deciduous and degraded forests), which is especially important for the dry forest.

With the advent of the Sentinel-2 program, data availability over target areas is increasing, allowing temporal analysis previously restricted to moderate (>100m) spatial resolution satellite data (such as MODIS), to be employed in the monitoring of forests at finer spatial resolutions. Sentinel-2 brings an improvement in the spatial resolution (with the three visible and a near infrared bands at 10m), which will allow a more accurate assessment of deforestation and forest degradation areas taking the minimum scales defined by the UNFCC, and in the spectral sampling (i.e. higher amount of bands with narrower width), with the inclusion of three bands in the red edge, which has shown to be useful for quantitative assessment of vegetation status (Frampton et al., 2013).

6.4 SPOT4 Take 5

In January 2013 in order to prepare for the use of Sentine-2 data, the French space agency CNES lowered the orbit of SPOT4 to put it on the same repeat cycle of Sentinel-2 until late June of the same year. During this period, SPOT passed over the same 45 selected test sites every 5 days. One of these test sites, selected by the author was in the dry forest in Tanzania. SPOT4 records in 5 spectral bands: three visible, one near-infrared and one SWIR at 20 m spatial resolution (Hagolle et al., 2013). This experiment, SPOT4 Take 5, does not simulate the full spectral and radiometric capabilities of Sentinel-2, but does simulate its revisit frequency and the spatial resolution.

For the test site in Tanzania 23 SPOT images were obtained, being acquired from 6th Feb to 19th June 2013. These were supplied in level 2A format (ortho-rectified surface reflectance data provided with a cloud mask). They covered an area of 360,000 ha and the acquisition period ranged from the end of the wet season to deep into the dry season. The advantage of the SPOT4 Take 5 data is that the 5 day repeat cycle is available, not yet available with Sentine-2 at the time of writing and that the same spectral characteristics are available for all acquisitions.

6.5 Study area and reference data

The study area is located in the Somalia-Masai ecoregion, in the dry highlands of Central Tanzania. The climate is semiarid; the rainfall is less than 500mm per year with high interannual variation, the mean monthly temperature between 20 and 25 ºC, and it has a well-defined arid season from beginning of May to end of November. Most of the region is covered with deciduous bushland and thicket (Acacia-Commiphora is the climax vegetation), which grade into evergreen and semi-evergreen bushland and thicket on the lower slopes of the mountains. At higher altitude in the mountains dry forests dominate.

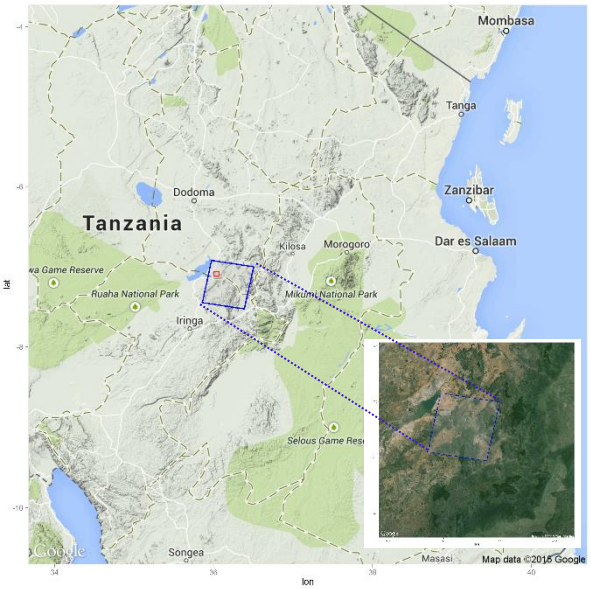

Figure 34: Study area

The area covered by the SPOT images (blue polygon, Figure 34) is centred at Lat Long (S 7.226 E 36.182), between the Dodoma and the Iringa regions. To the north there are agriculture fields and to the South a mosaic of degraded forest, which was fully covered by forest in 1980. From the Landsat archive one can see that severe deforestation and degradation took place between 1994 and 2011. Of interest, is how a multi-temporal remote sensing data can discriminate between different levels of forest cover, and perhaps degradation.

Figure 35: Landsat data (SWIR;NIR:RED) from 1994, 2008 and 2011. Note the changes in dry forest (brown) that occur in the south east corner of the image.

A pan-sharpened multispectral image acquired by the WorldView-2 satellite (0.5 m spatial resolution) on the 4th September 2010 was used as reference data, along with field data collected in 2012 (Hojas Gascón and Eva, 2014b). It covers an area of 5.000 ha centred at Lat Long (S 7.092, E 36.035) (red polygon, Figure 36.)

6.6 Data availability

From the cloud occurrence maps provided with the data the cloud free data frequency was summed up for each pixel in the SPOT scene area simulating 10 day (Sentinel-2A) and 5 day (Sentinel-2A and 2B) frequency acquisition.

Landsat-8 images (acquired every 16 days) were acquired for the same period as the SPOT4 Take5 data and cloud free data frequency maps were produced for comparison.

6.7 SPOT image segmentation

Two 'seasonal' mosaics were created, using SPOT images from the wet and dry season respectively. The image composition was achieved using the image with lowest cloud cover from the respective season, and replacing cloud contaminated pixels from the nearest date image sequentially. These were then segmented in combination to create polygons of a minimum mapping unit of both 1 ha and 0.5 ha size. This method was employed so as to retain and discriminate land features that may be distinct in either season.

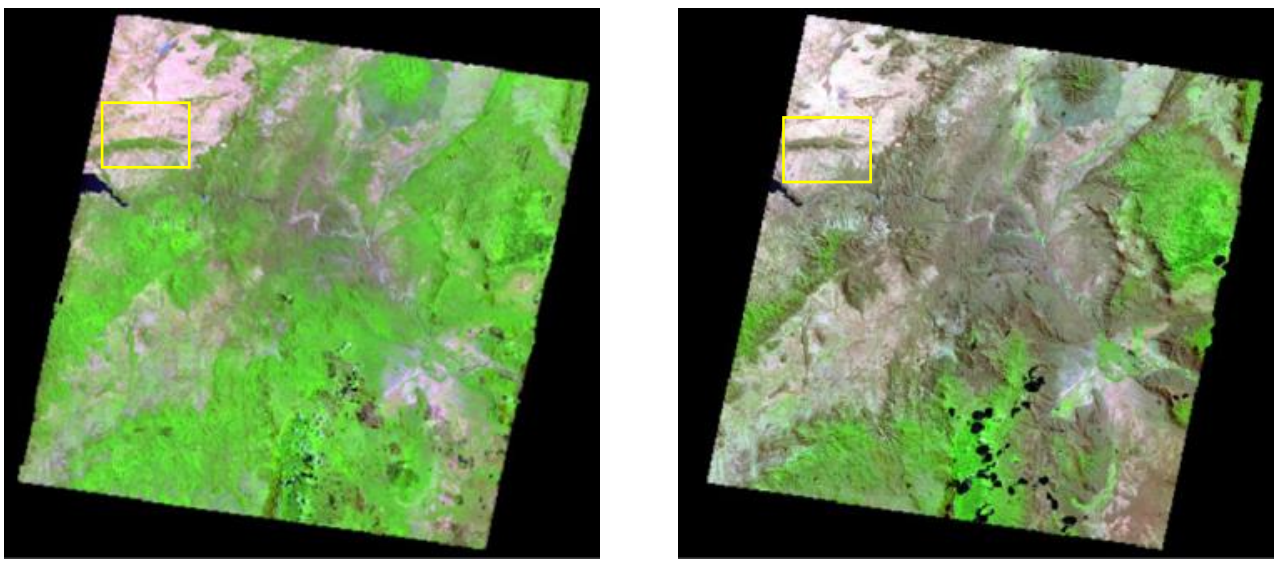

Figure 36: SPOT data (SWIR; RED; GREEN) from the wet (left) and dry (right) season. Note the dry forest (brown on the right) in the centre of the image. The evergreen forests are on the highlands.

NDVI, SAVI and MSAVI indices were calculated for the polygons. Analysis showed no significant difference between NDVI and SAVI trends, and the he MSAVI was not found to be effective at discriminating between woody and non-woody vegetation. For easy of processing the data were reduced to the NDVI series.

The NDVI was calculated for each single data image and layer stacked (Figure 37). From the layer stack the average NDVI was extracted for each of the segments created from the seasonal mosaic.

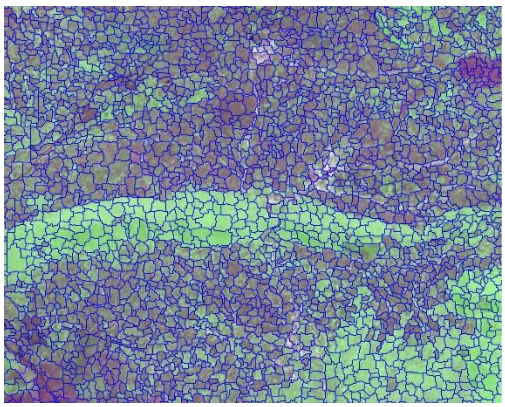

Figure 37: SPOT NDVI composite with 1 ha MMU corresponding to the box outlined in Figure 36.

6.8 Object vegetation classification

Reference data from VHR image

The WorldView-2 image was segmented so as to obtain polygons with a mean of 0.1 ha. Areas of bare soil and grass were identified using a 5% reflectance threshold in the red channel. Woody vegetation was then divided into tree cover (woody vegetation higher than 5 m) and shrub cover (woody vegetation lower than 5 m) on the basis of areal cover. Field data (Chapter 5) provided information on the ratio of woody vegetation height to crown width, which was found to be around 1. This was use to classify woody objects with crown width less than 5 m as shrub formations (Figure 37).

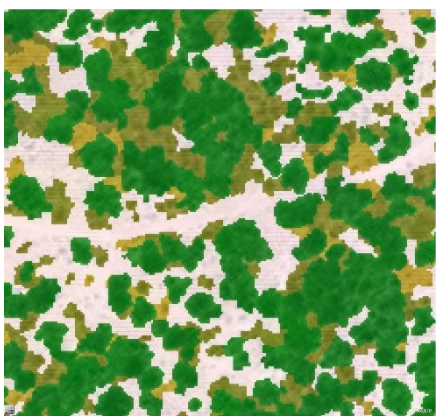

Figure 38: The WV-2 image (left) with a zoom on the box (centre) showing the trees and shrubs and the classification (right) – trees (green), shrubs (brown) and bare/grass (pink).

The segments from the SPOT data containing the NDVI profiles were then cross tabulated with the very high resolution (VHR) reference data so as to determine the proportion of tree cover, shrub cover and non-woody land cover (grass or bare soil) in each. The image objects were then classified using the proportions of the three elements with 6 category levels (0-10, 10-20, 20-40, 40-60, 60-80 and 80-100 cover percentages). For easy of nomenclature the woody and tree proportions were combined in a simple concatenation, for example, class W100F100 is all woody vegetation, of which trees (forest) is the only component, class W40F20 has 40% of woody vegetation, 20% of the polygon is trees and 20% shrubs, hence 60% non-woody vegetation.

Extraction of NDVI profiles

The NDVI profiles were examined by vegetation classes. Cloud-affected dates were removed from the series by averaging proximate date values.

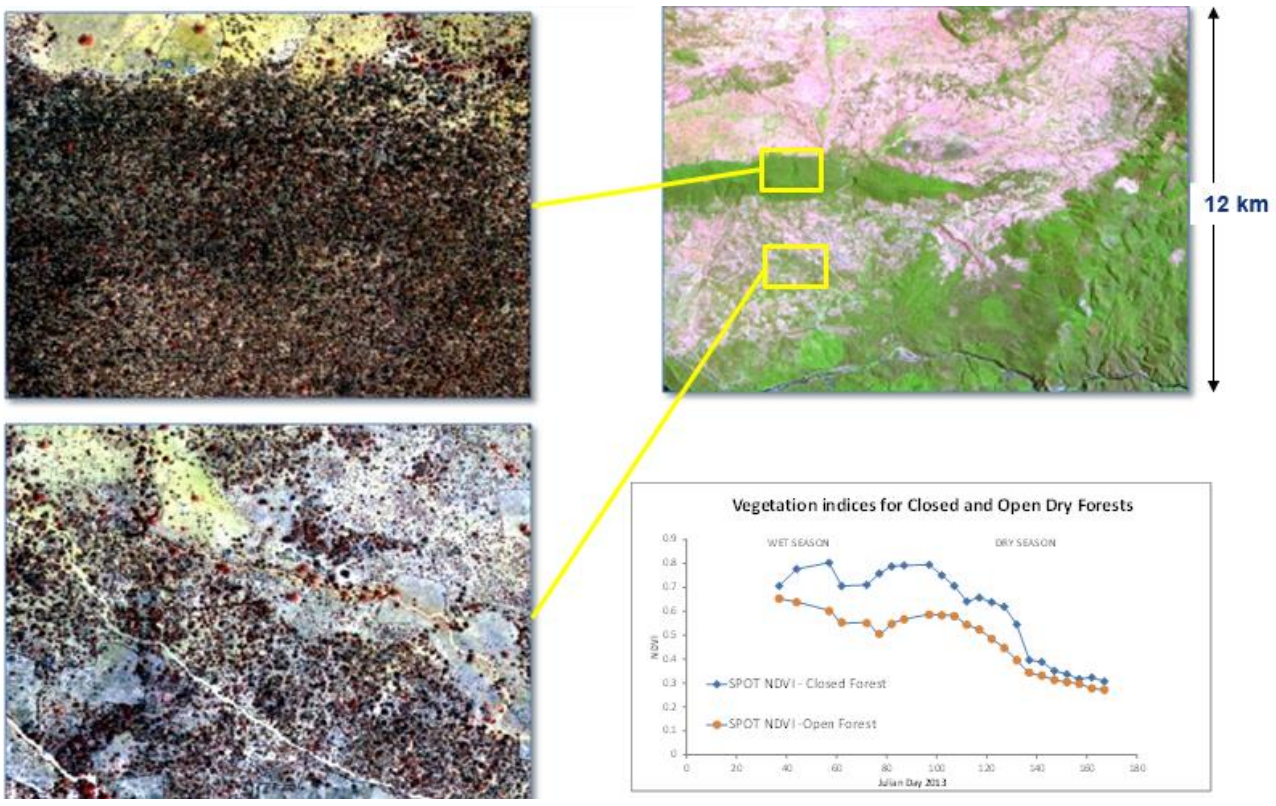

Figure 39: The WV-2 image (top left) showing the closed forest and (bottom left) the degraded areas. The equivalent areas on the SPOT composite and the NDVI profiles are shown top right and bottom right respectively. Data values for Julian day 60 were averaged due to cloud contamination.

A preliminary review of the NDVI profiles was made using the two field points – one in intact forest, the other the degraded forest area Figure 39. The profiles show a clear trend, whereby the NDVI difference becomes important during the development of the dry season – not at the start, neither at the end. Note that on the fifth date, a cloud contaminates the area over the closed forest.

Figure 39 shows that generally the NDVI of the land units falls from a peak at the start of the observation period (end of the wet season) until the end of the period (deep into the dry season). The closed forest formation remains greener for longer than the open (degraded) formations where shrubs and grass layers dry out more quickly than the trees which have deeper roots.

Random Forest

Random Forest is a learning algorithm widely used in the statistical community to cluster data in different classes (among other analysis), constructing a multitude of decision trees at training time (Breiman, 2001). It has been shown to be effective at land cover classification (Rodriguez-Galiano et al., 2012). The data were examined in Random Forest with a training of 150 of the 832 sample sites for the three acquisition modes (5 day; 10 day and 16 day). Confusion matrices were generated between the simulations and the WV2 reference data for each mode between the simulations along with overall accuracy and kappa statistics (Congalton and Green, 2009).

6.9 Results

Cloud free image area

For the full SPOT scene 90% of the area is acquired cloud free at least once during the five month period of the Spot4 Take 5 experiment simulating one Sentinel-2 (an image every 10 days). This rises to 99% with Sentinel-2A and 2B (acquiring images every 5 days). Figure 40 shows the increment of cloud free data frequency with both Sentinel-2 (SPOT) with respect to Landsat-8 during the same period. Combining both satellites (image not shown) the scene coverage is of 100%. During the period, only 5 Landsat acquisitions were made.

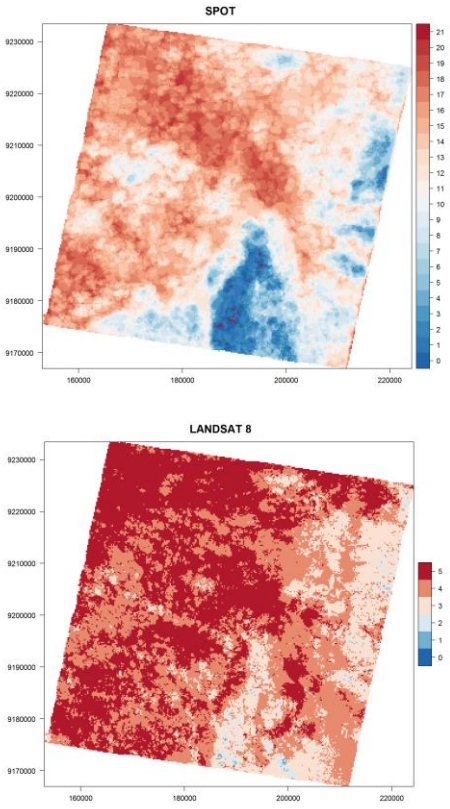

Figure 40: Cloud free data frequency maps with SPOT (top) and Landsat-8 (bottom).

NDVI time series

The trends in mean NDVI for the different classes of woody vegetation are shown in Figure 41. Classes with higher percentages of trees have higher NDVIs. The variance of the trajectories (not shown) however is high. It was noted that the mean NDVI from the different classes pass from being very divergent at the end of the wet season to be very similar in the dry season.

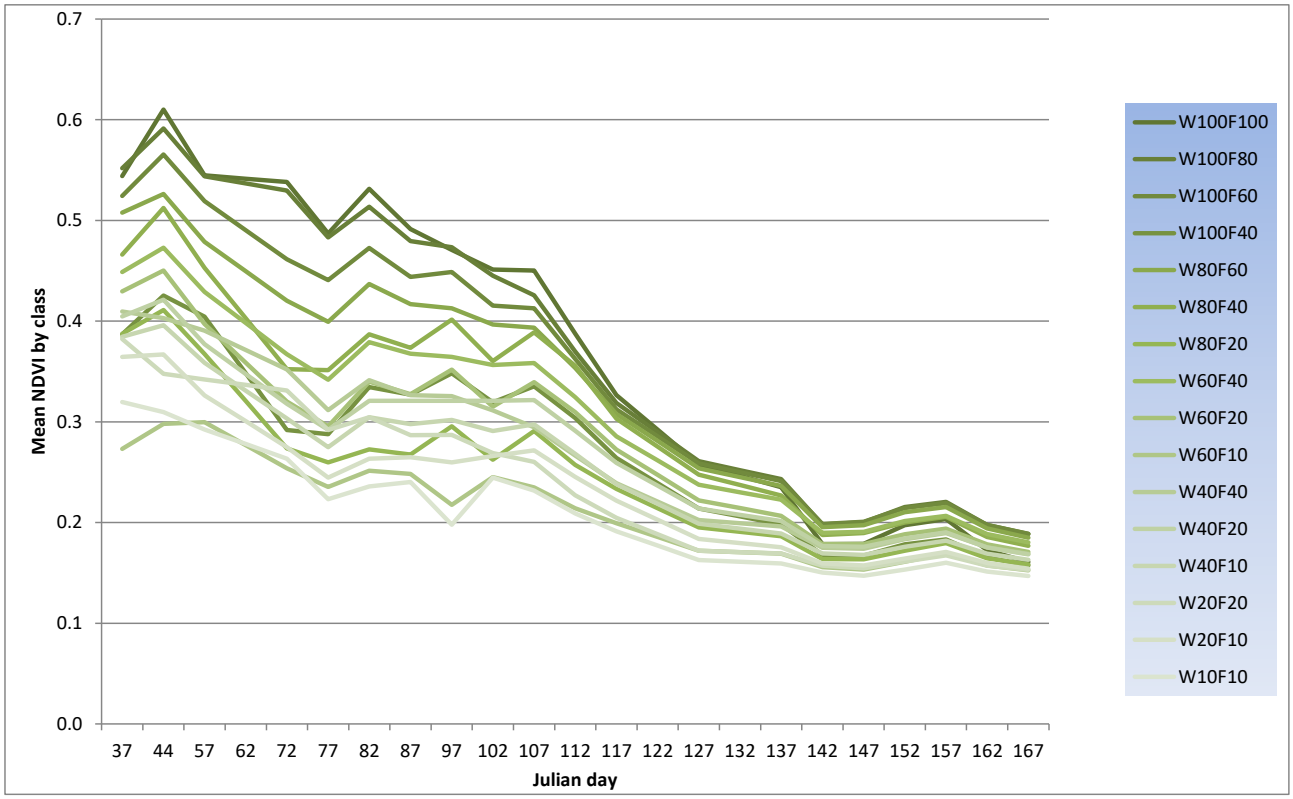

Figure 41: NDVI profiles from by vegetation classes.

The averaged SPOT NDVI time series was also compared to that obtained with the 250 m spatial resolution MODIS sensor (Figure 42). The latter has been produced from mean NDVI records from 2000 to 2012 in 16-days periods across the study area. Taking into account that the SPOT time series is a one single acquisition composite, it can be seen that it corresponds well with the smoothed MODIS product, except in the transitional period between the wet and the dry season. The dip in the profile at this period could be caused by an anomaly in temperature or rainfall regime or to the low quality of the cloud flagging.

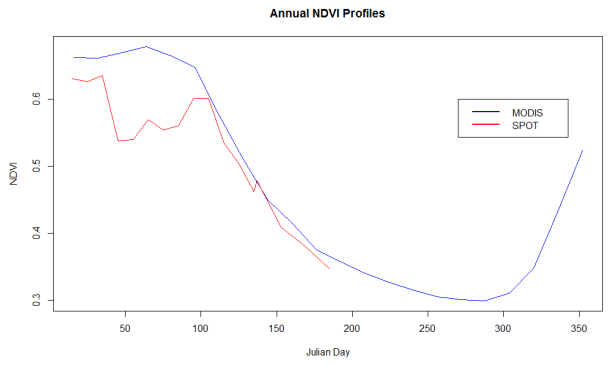

Figure 42: NDVI profile from SPOT and MODIS.

Classification of NDVI profiles

For each acquisition simulation the confusion matrix was produced, along with accuracy assessement, Kappa statistics and Z tests of significance.

Simulation of Sentine-2A and 2B acquisition – 5 day repeat cycle

Table 14: Confusion matrix of classification with SPOT simulating Sentinel-2 A and B in comparison to Worldview-2.

		S2A – S2B simulation					
		F100	F80	F60	F40	F20	F10
WV2	F100	242	151	40	10	6	5
	F80	77	567	176	25	3	16
	F60	16	209	442	147	19	44
	F40	4	17	158	342	119	72
	F20	2	0	18	138	232	190
	F10	2	1	7	7	121	670

From the confusion matrix (Table 14) the class separability is shown and hence one can assess by what percentage forest cover would have to fall to be correctly identified. It is seen that a change of two classes (i.e. 40% fall in forest cover) is required to ensure correct identification (e.g. from F100 to F60). The detection of 20% forest cover loss has a risk of misclassification by 20% approximately. The overall accuracy (Congalton and Green, 2009) is not that high at 58%, however, note that much of the misclassification is around the diagonal. Regrouping the classes into ranges of 40% - i.e. F100 & F80 combined; F60 & F40 combined and F20 & F10 combined, results in an overall accuracy of 77%.

Simulation of Sentine-2A acquisition – 10 day repeat cycle:

Table 15: Confusion matrix of classification with SPOT4 Take 5 (simulating Sentinel-2A) in comparison to Worldview-2 (reference data).

		S2A simulation					
		F100	F80	F60	F40	F20	F10
WV2	F100	240	151	41	10	7	5
	F80	78	556	183	27	5	15
	F60	15	217	442	144	21	38
	F40	5	19	158	337	116	77
	F20	2	1	18	141	233	185
	F10	1	1	8	8	129	664

The confusion matrix from the SPOT data with a 10 days acquisition frequency does not give very different results, with the same overall accuracy of 58%. The detection of 20% forest cover loss has a risk of misclassification of 25% approximately, while the detection of 40% forest cover loss of only 5%.

Normal Landsat 8 acquisition – 16 day repeat cycle

Table 16: Confusion matrix of classification with 16 day acquisition in comparison to Worldview-2 (reference data).

		L8 Simulation					
		F100	F80	F60	F40	F20	F10
WV2	F100	225	151	39	23	12	4
	F80	81	513	214	40	3	13
	F60	10	229	405	150	34	49
	F40	3	24	170	293	114	108
	F20	2	1	26	130	207	214
	F10	2	2	16	16	156	631

For the 16 day acquisition cycle, a drop of overall accuracy to 52% is found.

6.10 Testing the improvement between the confusion matrices

Cohen's kappa coefficient (hereafter kappa) is a statistic that measures inter-rater agreement for qualitative (categorical) items. This measure of agreement is based on the difference between the actual agreement in the error matrix (i.e., the agreement between two classifications) and the chance agreement that is indicated by the row and column totals (i.e., marginals). For further details see (Congalton and Green, 2009) . A Kappa value has been computed for each error matrix and is a measure of how well the remotely sensed classifications agree. Confidence intervals around the Kappa value can be computed using the approximate large sample variance and the fact that the Kappa statistic is asymptotically normally distributed. This fact also provides a means for testing the significance of the Kappa statistic for a single error matrix to determine if the agreement is significantly greater than 0 (i.e., better than a random classification).

Table 17 presents the results of the Kappa analysis on the individual error matrices. The Kappa values are a measure of agreement or accuracy. The values can range from +1 to −1. However, since there should be a positive correlation between the remotely sensed classification and the reference data, positive Kappa

values are expected (Landis and Koch, 1977) characterized the possible ranges for Kappa into three groupings: a value greater than 0.80 (i.e., >80%) represents strong agreement; a value between 0.40 and 0.80 (i.e., 40–80%) represents moderate agreement; and a value below 0.40 (i.e., < 40%) represents poor agreement.

Table 17: Kappa statistics for the three acquisition scenarios

Repeat cycle / Satellites	Kappa
5 day / S2A & S2B	0.4832
10 day / S2A	0.4902
16 day / Landsat 8	0.4251

Table 18 presents the results of the accuracy assessment, along with the statistics by class carried out with the R "caret" package. Overall accuracy is calculated at just over 57% for the first two cases, and 52% for the third one, with a p-value of 2 x 10^{-16}. The classifier seems to be doing a reasonable job of classifying items.

Table 18: Metrics for each of the acquisition scenarios

1st model (5 day)

Accuracy : 0.5752
95% CI : (0.5602, 0.59)
Statistics by Class:

	Class: A	Class: B	Class: C	Class: D	Class: E	Class: F
Sensitivity	0.70381	0.5884	0.5200	0.50525	0.45597	0.6748
Specificity	0.94592	0.9081	0.8738	0.89672	0.90837	0.9556

2nd model (10 day)

Accuracy : 0.5809

95% CI : (0.566, 0.5957)
Statistics by Class:

	Class: A	Class: B	Class: C	Class: D	Class: E	Class: F
Sensitivity	0.70554	0.6000	0.5256	0.51121	0.46400	0.6720
Specificity	0.94636	0.9113	0.8741	0.89796	0.90830	0.9582

3rd model (16 day)

Accuracy : 0.5276
95% CI : (0.5126, 0.5426)
Statistics by Class:

	Class: A	Class: B	Class: C	Class: D	Class: E	Class: F
Sensitivity	0.69659	0.5576	0.46552	0.44939	0.39354	0.6192
Specificity	0.94256	0.8965	0.86279	0.88546	0.90143	0.9417

Table 19 presents the results of the Kappa analysis that compares the error matrices, two at a time, to determine if they are significantly different. This test is based on the standard normal deviate and the fact that although remotely sensed data are discrete, the Kappa statistic is asymptotically normally distributed. At the 95% confidence level, the critical value would be 1.96. Therefore, if the absolute value of the test Z statistic is greater than 1.96, the result is significant and you would conclude that there are statistically significance differences between the models.

Table 19: Z tests to determine the significance of the improvements in classifications

Model 1 (5 day) vs Model 2 (10 day)
z-test = 0.66

Model 1 (5 day) vs Model 3 (16 day)
z-test = 6.18

Model 2 (10 day) vs Model 3 (16 day)
z-test = 5.49

The results of this pairwise test for significance between the first two confusion matrices reveals that they are not significantly different, while the third one is statistically different – that is to say that the improvement in accuracy of the first two (5 and 10 days) over the 16 day classification is not random.
This is not surprising since the overall accuracies were 57 and 58% and the Kappa values were 0.48 and 0.49, respectively.

6.11 Conclusions

The increased frequency (5 or 10 days) of Sentine-2 (~10m) resolution data acquisition compared to the current 16 day Landsat 8 (30 m resolution)) promises to bring higher potential for detecting and quantifying forest degradation. Using the 20m resolution SPOT4 Take 5 data, processed to a simple vegetation index (NDVI) demonstrates that forest degradation can be detected when a reduction of 40% canopy cover or more occurs in 0.5 ha land units. This is valid for both 5 and 10 day acquisitions. Lower reductions in canopy cover are also detectable, however, with a higher (~ 20-5%) chance of misclassification.

Deforestation and forest degradation monitoring in the context of REDD+ require change detections of 10% and less than 10% respectively in the forest cover of land units of 0.5 ha. Here one could only discriminate classes with forest cover with more than 40% difference. However, the results should underestimate the

potential of Sentinel, which has a finer (10m) resolution and finer band widths. At the same time only a limited (4 months) of the season were available.

The development of better indices, deviations from historical data and greater use of texture measures could help to characterize the vegetation changes over the full growing season, and allow us to better diagnose and discriminate degradation processes in forest ecosystems.

7. Estimating woody biomass to detect forest degradation in Tanzania using a combined approach of field data and remotely sensed images

This chapter contains material of a publication under review:

Lorena Hojas Gascón, Guido Ceccherini, F. Javier García-Haro, Valerio Avitabile and Hugh Eva, 2017, The potential of high resolution (5 m) RapidEye optical data to improve above ground biomass estimates over Tanzania, *Remote Sensing*, under review.

7.1 Use of Remote Sensing data to map biomass

Mapping and monitoring of forest carbon stocks across large areas in the tropics has become increasingly based on remote sensing approaches (Timothy et al., 2016), which in turn depend on field measurements of biomass for calibration and validation purposes. Direct biomass estimations based solely on satellite data are not yet feasible, satellite data have to be correlated with field-based measurements.

One can discern three general approaches of remote sensing methods for mapping and monitoring of carbon stocks (Goetz et al 2009). Approach 1 assigns a single carbon density value to each biome (i.e. tier 1 values). Approach 2 extends the first approach by adding various ancillary spatial data layers (e.g. Ruesch and Gibbs 2008). Approach 3 aims to derive carbon stock estimates from machine learning algorithms based on satellite data and other detailed spatial data coupled with field measurements (i.e. tier 3 values). There are a range remote sensing technologies and options that can be employed for forest carbon stock monitoring from global to local levels (Goetz and Dubayah, 2011; Timothy, 2015).

In approach 1 and 2 ('Biome-average' and 'combine and assigned') large area maps even up to global levels [Ruesch and Gibbs 2008] can be derived by using aboveground biomass (AGB) values from IPCC 2006 guidelines per ecological zone and continent, modified according to different rule-sets and 'best guesses'. These have the advantage of using ready-made products e.g. the Global Land Cover 2000 map (Batholomé and Belward, 2005) with each land cover class being given an assigned biomass value.

The Approach 3 ('direct remote sensing') requires the acquisition and processing of large volumes of data to produce global or national-scale carbon stock estimates. Moreover, in many tropical developing countries national forest inventories are either non-existent or out of date. Alternatively, biomass carbon content can be mapped using LIDAR and radar data that are capable to capture vertical tree canopy structure (height) which can

be used to estimate aboveground carbon stock (Goetz and Dubayah, 2011; Treuhaft et al. 2009) . For this, spaceborne LIDAR/radar sensors, which collect data along transects, are combined with other spatial data (e.g. optical satellite data) to expand/extrapolate the estimations to the horizontal coverage. Global forest canopy height map has been produced from the segmentation of the 500 m Moderate Resolution Imaging Spectroradiometer (MODIS) with a mean mapping unit of 25 km^2 (Lefsky, 2010). Height was calculated directly from the Geoscience Laser Altimeter System (GLAS) observations and then extrapolated with regression, a mean RMSE of 5.9 m and mean correlation (r2) of 0.67.

Approaches using coarse spatial resolution satellite imagery in combination with field plot measurements and space-borne LIDAR (light detection and ranging) data have been executed to derive wall-to-wall pan-tropical biomass maps at 1km resolution for the year 2000 (Saatchi et al. 2011) and 500 m resolution for 2007 (Baccini et al 2012). Both studies use a similar approach, correlating tree height from Geoscience Laser Altimeter System (GLAS) data points from/onboard the space-born Ice, Cloud, and land Elevation Satellite (ICESat) to AGB values of (overlapping) field inventory plots to deduce the AGB values for all non-overlapping GLAS data points. Then data fusion of multiple satellites (the moderate resolution imaging spectroradiometer (MODIS) sensor, shuttle radar topography mission (SRTM) and quick scatterometer (QSCAT) (the latter only for the Saatchi) were used to extrapolate spatially the AGB GLAS point values to a wall-to-wall coverage comprising almost all countries of the pantropical belt.

Analysis has been carried [Langer et al 2014] out on these two pantropical biomass maps (Saatchi et al (2011), Baccini et al (2012)) and a combination of both (combined dataset) in comparison with the IPPC values. IPCC Tier 1 default values may lead to an overestimation of AGB and the use of these datasets seems to be more appropriate as the default values in IPCC often refer to intact forest sites (Gibbs 2007). The mean AGB values per pan-tropical ecological zone show a good consistency between the two datasets with a correlation coefficient (R^2) of 0.87. When restricting the regression to intact forest areas (IFA) (Potapov et al 2008) R^2 is even higher: 0.97. For non-IFA R2 is lower: 0.80. A comparison between both pan-tropical maps generally shows higher AGB values from 500 m map (Baccini et al. 2012) than 1 km map Saatchi et al (2011). However, a detailed look at the data shows significant differences between the maps, 50% difference for tropical dry Africa.

Another comparison of these datasets was undertaken both globally (Mitchard et al., 2013) and specifically over the Amazon basin (Mitchard et al., 2014) and concluded that no single map was generally superior to the other

even despite the substantial differences. However, it has to be taken into account that the datasets are from different epochs 2000–2001 (Saatchi et al., 2011) and 2007–2008 (Bacchini et al., 2012).

Despite their differences it has been suggested (Mitchard et al., 2013) that these two datasets, with their transparent methodology, could be used to derive alternatives to the IPCC Tier 1 value. In particular, by using the average composite of the datasets analyzed by continental ecological zone and wherever feasible, using country-based estimates in order to best reduce the uncertainty of the estimates. Such an approach has been used with data fusion to produce a new pantropical biomass map with high accuracy (Avitabile, 2016) .

At the national level, AG Carbon loss estimates for the Democratic Republic of Congo (DRC) for the last decade (2000–2010), have been produced (Tyukavina, 2013) by relating forest type and loss from the FACET (Fôrets d'Afrique Centrale Evaluées par Télédétection) (FACET, 2010) product (a national-scale land cover change dataset), created using Landsat data, to carbon data derived from the Geoscience Laser Altimeter System (GLAS). Lidar data was calibrated using co-located field measurements. Biomass loss was estimated at pixel and subpixel level with error-adjusted area of forest cover loss between 2000 and 2010, gross AGC loss, and associated uncertainties.

While the Landsat sensor has been and is, essentially the 'workhorse' for forest monitoring at all scales, its spatial resolution been demonstrated to be too low for monitoring degradation for REDD+ reporting requirements, i.e. changes in forest cover and carbon stock in land units as small as 0.05-1 ha (Hojas Gascón et al., 2015), however combining Landsat with GLAS data has been successfully employed for AGB mapping in China (Liu, 2017). For sub-national studies, higher resolution spatial data has been used, such as Light Detection and Ranging (Lidar) data combined with QuickBird high-resolution satellite images, calibrated and validated by field measurements [35][Gonzalez 2010]. Despite good results, such approaches are only currently applicable to small area sites, due to the costs and availability of very high resolution data. The combination of PALSAR (Phased Array type L-band Synthetic Aperture Radar) and WorldView-2 data has been shown to be a promising approach to improve biomass estimation over a large area in Chinese forest (Deng, 2014) with biomass estimate accuracies improved to between 65% and 71%.

In conclusion, the major sources of uncertainty identified in IPCC guidelines (Aalde et al., 2006) (including field measurement error, remote sensing accuracy, biomass regression equations, and spatial autocorrelation) still need to be addressed.

7.2 Materials and Methods

Fine spatial (5 m) resolution optical data systematically distributed across Tanzania in 76 sample sites each of 20 km by 20 km were used in conjunction with a corresponding subset of field plots extracted from the national survey to obtain an estimate of above ground biomass for the country. The placing of the sites at confluence point of the geographic grid (e.g. S7 E35), was chosen to coincide with the FAO-JRC sampling scheme for forest change FRA 2010 assessment (Mayaux et al., 2005). The estimate was based on developing a model to relate remote sensing parameters extracted from the images (reflectance and texture) to the field data (AGB, basal area and tree height) for the correspond locations. The model was then to be applied to the full extent of each of the 76 images to obtain the average AGB per sample site, and then, by direct expansion, the country.

The potential improvement can be measured by the Relative Efficiency – showing the reduction (or otherwise) in variance that can result from the combination of a second data set, in this case the remote sensing data. This implies that the field data collection can be reduced, with significant cost savings.

The methodology was applied in the following steps:

1. Determination of the minimum mapping unit (MMU) to employ when processing the remote sensing data
2. Detection and removal of cloud and cloud shadow from the RE images
3. Processing of the images to obtain objects at the MMU
4. Extraction of a range of reflectance, texture and indices from those image objects which corresponded to the field data (geometric location)
5. Review and 'cleaning' of the field data to remove ambiguous or erroneous data
6. Testing of models to relate field data on above ground biomass (AGB), tree height (TH) and basal area (BA) with the spectral and textural parameters extracted from RS data
7. Application of the best model to the image objects in the full Remote Sensing data set
8. Calculation of AGB per unit area for each of the 76 image

9. Extrapolation to the national level and compare with the results from the full field survey

10. Compare the model results for RapidEye to those obtained by using data from Landsat 8 and Sentine-2 data

11. Review the potential for detecting changes in biomass over time

7.3 Study Area

The study area covers the full land surface of mainland Tanzania, which is described in Chapter 1. The actual analysis takes place of the set of 76 20km by 20 km satellite scenes acquired across Tanzania at the confluence points of the lat-long grid, see Figure 8Figure 1 They occur across the full range of the Tanzanian ecosystems, Table 22.

7.4 Field data

The national field data set collected by the NAFORMA project is described in section 1.2.4 It was designed to provide a national estimate of forest area and change, wood volume and growing stock (NAFORMA, 2010a), a set of social indicators (NAFORMA, 2010b) and a soil database. The survey design was stratified using potential biomass and accessibility to facilitate the field visits (Tomppo et al., 2014, 2010). Some 32,000 plots were visited over a period of 3 years, collecting the biophysical and social variables. For each plot (a 15 m radius circle) data were collected on canopy cover, tree and shrub height, trunk diameter (DBH), species, dead wood and soil. The final results were released in 2015 (MNRT, 2015). The biophysical data on basal area and tree height were then used by the project in a model to calculate the biomass carbon for each site, from which national level estimates were made. For tree volume, a set of allometric equations were applied to the dataset, with a set of different models used, four for specific tree species with distinct properties and two generic model for the remaining species. A more elaborate list of allometric equations for Tanzania has since been provided (Tanzania UNFCC REL 2016).

The main parameters for calculation biomass were tree height and basal area. Tree species was also used to select an appropriate allometric model and wood density.

Basal area in m^2 per ha $=$ $\sum_{i=1}^{N} \frac{0.785 * dbh_i^2}{\pi * R^2}$

Where: dbh is the diameter at breast height (cm) of woody species i, N the total number of plants and R (m) the radius of the sample circle.

Table 20: Tree volumes (V), using height (H) and diameter at breast height (dbh) [NAFORMA 015](MNRT, 2015)

Species	Equation
Eucalyptus grandis:	$V = 0.000065 * dbh^{1.633} * H^{1.137}$
Pinus patula:	$V = 0.00002117 * dbh^{1.8644} * H^{1.3246}$
Tectona grandis:	$V = 0.0001* dbh^{1.91} * H^{0.75}$
Dalbergia melanoxylon	$V = 0.00023* dbh^{2.231}$
Woodlands	$V = 0.0001 * dbh^{2.032} * H^{0.66}$
All other species	$V = 0.5 *\Pi * (0.01 * dbh / 2)^2 * H$

To obtain AGB the tree volume is multiplied by the wood density which was fixed to a value of 0.580(no units) for forest species and 0.5 for woodlands.

Through a collaborative agreement with the Tanzania Forest Service, the JRC was given access to a sample of the NAFORMA field data. These data consisted of the plots falling in 40 km by 40 km boxes centred on the 76 latitude-longitude confluence points in Tanzania, a total of around 5000 plots. These sites were chosen as they corresponded to the sample sites used for the JRC-FAO FRA 2010 survey (Mayaux et al., 2005).

The NAFORMA methods for basal area and tree height measurement were replicated for work in this thesis for some 20 sites in Tanzania in 2011 and 2012 and reported in Hojas and Eva 2014 (Hojas Gascón and Eva, 2014b). These results were successfully correlated with spectral and textural data from very high resolution (World View 2) data, to produce a biomass map of the Pugu hills forest reserve around the town of Kisarawe. The sites were separated into intact forests – with little or no sign of disturbance, and non-intact forests. The comparison of the intact sites, with the full national plot data for the same biome 'coastal forests', is shown in Chapter 5.

7.5 RapidEye satellite data

RapidEye images over Tanzania were acquired for the year 2010 as described in the Chapter 1, Figure 8. These images were selected to cover 20 km by 20 km boxes around each of the 76 confluence points of the latitude-longitude grid in Tanzania and coincide with the FAO FRA JRC 1990-2000-2010 forest change analysis (FAO et al., 2009). The spectral characteristics of the RapidEye sensors are given in Table 3.

7.6 RapidEye preprosessing

Radiometric calibration

RapidEye level 3A data are provided as 5 channel layer stacked Geo-Tiff files [Table 3], with a nominal 5 m pixel size stored as 16bit Digital Number (DN). To convert to at-sensor radiance in watts per steradian per square metre (W/m^2 sr µm), a scale factor is applied:

$L\lambda = DN\lambda * ScaleFactor(\lambda)$

where $ScaleFactor(\lambda) = 0.01$

The Top of Atmosphere reflectance is calculated by:

$\rho_\lambda = \pi * L_\lambda * d^2 / ESUN_\lambda * \cos \theta_{SZ}$

Where:
ρ_λ= TOA reflectance for band λ
L_λ = Radiance for band λ
θ_{SZ}= Local solar zenith angle
d = (1-0.01672 * cosine (0.01745 * (0.9856 * (Julian Day Image - 4)))

The mean exo-atmospheric solar irradiance $ESUN_\lambda$ in W/m_2 / µm) for each channel is respectively:
$ESUN_\lambda 1\text{-}5$ = [1997.8: 1863.5: 1560.4: 1395.0: 1124.4]

Formulas and parameter are derived from [RapidEye 2011][61]. To reduce noise in reflectance between images acquired at different dates and locations an 'evergreen forest normalisation' was implemented (Hojas Gascón et al., 2012; Bodart et al 2011), based on the theory of dark object subtraction (Hansen et al., 2008).

RapidEye satellite image segmentation and classification

Image segmentation was chosen so as to allow the mapping of landscape units at a cartographic minimum mapping unit (MMU), which corresponds as closely as possible that set by the national authority responsible for forests, in this case the Tanzania Forest Service and international guidelines (IPCC, 2003). Almost all national forest definitions are based on tree height, tree cover with a minimum mapping unit, whereas remote sensing data are imaged at pixel level. Segmentation clusters pixels of similar radiometric characteristics in conjunction

with textural information, into landscape units (e.g. 'a forest') (Blaschke 2010)

Choice of a minimum mapping unit

The minimum mapping unit should relate to the definition of the national forest service and international guidance for filling reporting requirements (e.g. forest cover changes for REDD+). The UNFCC has proposed that countries should choose a MMU from 0.05 to 1 ha (COP 7 2001). Initially, the Tanzanian authorities were in favour of a 0.05 ha mapping unit. The driving force behind this decision came from a political will to engage participative involvement in REDD at the community level – every small holder would able to estimate his or her plot's contribution to the REDD process, and forests held in village ownership would also be accounted for. Although the inclusive nature of the decision was attractive, the pragmatic consequences were quite daunting.

Before deciding on an appropriate MMU, the segmentation parameters need to be fixed. These control the rules that determine the aggregation of pixels to obtain polygons which reflect the landscape units. A series of tests effectuated changing the compactness and reflectance weights of the segmentation process (carried out in the eCognition software). It was found that a high compactness (more rounded polygons) provided results that were more in line with traditional photo interpretation results. Results with lower compactness gave highly fragmented landscapes, especially considering the finer spatial resolution of the RapidEye data.

For detecting levels of biomass from remote sensing, texture can be an effective parameter (Eckert, 2012; Gong et al., 1992). To have a consistent magnitude of texture the scale of analysis (kernel) needs to be large enough to cover the variation in the target landcover with respect to the resolution of the satellite data. One can imagine that with a 1 m resolution sensor the texture of tree crowns will show high variance until the unit of observation, the kernel, is larger than a single tree. To determine the minimum mapping unit that would give stable and consistent measures, a series of tests was carried out with incrementally increasing circular units in homogeneous landscapes from 0.25 ha to 6 ha. For these objects a set of texture measures were extracted so as to review the changes in these measures that related only to the size of the measurement unit, and not to the land cover. These tests showed that some texture measurements were highly sensitive to the size of the image object, however, around 1 ha the results stabilized

Figure 43). As a result a 1 ha as the MMU was therefore selected. This is larger than the national specification (0.5 ha) but is in line with international guidelines.

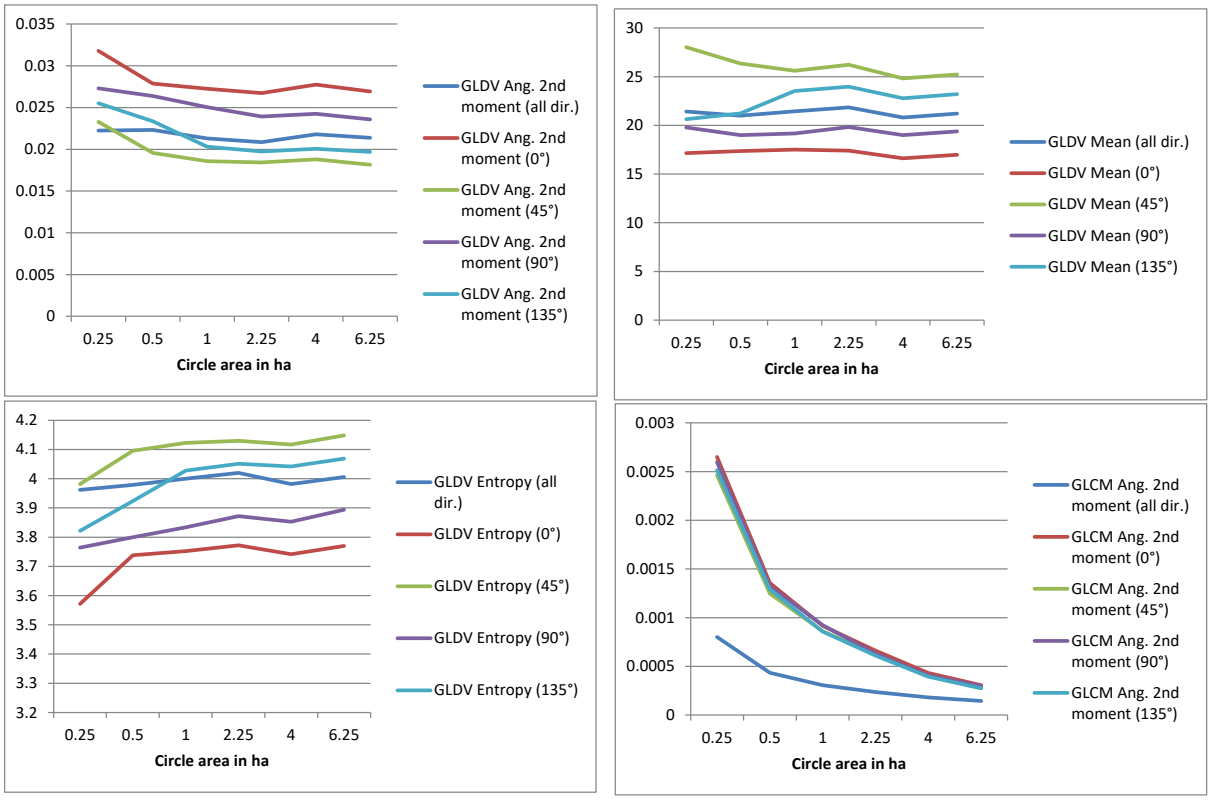

Figure 43: Scale dependence of different texture measures (see appendix for other measures).

The shape factor is the parameter that controls the size of the segments (objects) in the chosen software (Raši et al., 2011; Baatz et al., 2000) (Bodart et al., 2011), (Ma et al., 2016). However the results (i.e. the size of objects) are dependent on the image itself, that is, the image contrast, variance and landscape fragmentation. To arrive at the required MMU an iterative routine was applied to each image starting with initial, small, landscape units, and then in a stepwise process increasing the scape factor until the median of the image objects corresponded to the MMU. Polygons smaller than the MMU can then be dissolved and added to the adjacent polygon with the closest spectral signature (Raši et al., 2011).

As the median must be calculated only on land cover features, clouds and cloud shadows had to be detected, masked out and omitted from the routine.

Cloud and cloud shadow masking

The RapidEye data exhibited two artifacts that needed to be addressed for effective cloud and cloud shadow masking can be implemented; i) cloud and cloud shadow displayed a locational shift of several pixels between channels, and ii) the sun azimuth angle was missing in the metadata.

The locational shift (i) between channels is a parallax effect depending on the height of the cloud and is described in RapidEye documentation on line (Konstanski, 2012). This means that the cloud and cloud shadow masks cannot be set in a single channel. Lacking a thermal channel, thresholds have to be set in the VNIR channels. To initiate the masking processes, segmentation on the image was carried out producing fine (i.e. small) spatial image objects of approximately up to 10 pixels. For RapidEye this is obtained using a segmentation scale factor of 20.

An incremental approach based on (Szantoi and Simonetti, 2013), was employed whereby initial (high) thresholds applied to detect core cloud areas in each channel. The fine objects classified as core cloud (CC) are then merged and a lower (10%) set of thresholds are then applied only to those areas adjacent to the core areas, so as to expand the area classified as cloud. In this way edge of cloud (EOC) areas were detected. A final buffer of 30 m was added to the cloud so as to remove land cover areas affect by partial cloud, too fine to be easily classed but which modify the reflectance of the affected area. An initial review of areas detected as cloud identified a significant number of small areas which were in fact bright land cover types (bright soil, urban, calcium rock formations). A spatial filter was run on all targets labeled as cloud and declassified those small than 50 m^2, permitting them to be re-classified as an appropriate land cover class at a later stage.

To address (ii), the missing the sun azimuth angle, an interface was developed to visually approximate the sun azimuth angle using the position of clouds and their respective shadows. This was then used in the classification process to help in the identification of areas affected by cloud shadow, by identify segments with reflectance lower than 1% that are up to a maximum of 100 meters from the CC-EOC segments and at 180 degrees to the sun azimuth as cloud shadow. Adjacent cloud shadow polygons are in turn merged.

Processing RapidEye images to objects to the MMU

The remaining, unclassified segments, were then processed in an iterative loop, whereby the shape factor was increased progressively until the medium segment area was equal to 1 ha. Segments smaller than 1 ha were merged to the adjacent segment with the closest Euclidian spectral distance. These (level 1 ha) objects were then used as the basis for the biomass, tree height and basal area models and estimates.

7.7 Extracting parameters for developing models for above ground biomass, tree height and basal area

As with the work carried out with VHR data in Chapter 5, the goal was to review the possibility of relating the remote sensing data from the RapidEye sensor to field data an extensive set of parameters was extracted from the image objects which corresponded to the field plots. The parameters are given in Table 4. These parameters fall into four categories: simple reflectance means and standard deviations for the objects; derived indices designed to highlight various aspects of land cover (e.g. vegetation indices), advanced texture measures based on Grey Level Co-occurrence Matrices (GLCM) (Haralick et al., 1973b) (Haralick, 1979) and finally a categorical class was produced giving the percentages of the polygon effected to belong either to bare soil, woody vegetation or photosynthetically active non-woody vegetation.

Spectral indices have been used for detecting vegetation and more specifically forest parameters in a number of studies (e.g., Eitel 2007; Daughtry, 2000; Haboudane, 2004; Lu, 2011; Rikimaru, 2002; Akike, 2016). The short wave infra-red channels (1.6 and 2.7 µm) are found to be highly correlated with forest parameters and canopy cover (e.g., Halperin 2016). Unfortunately these channels are absent from the RapidEye sensor. Indices such as the shadow index have also been successful (Lu, 2011; Akikie, 2016). Improvements in assessing biomass have been achieved using temporal series to monitor phenological changes (Gwenzi, 2017), however, in this study data were limited to single data imagery.

As satellite data of finer spatial resolution has become more readily available, the use of the texture measures has become more common (Eckert, 2012; Gong, 1992). However, given that the list of texture indices available is extensive (around 80) and that they are computationally time consuming, in a first step the inter-correlation between the set of texture indices available was reviewed, with the objective of reducing them, since, while calculating them for a limited number of field plots is feasible, processing the full set polygons for each image (there may be up to 40,000 polygons in each image) would be challenging, and time consuming (Gong, 1992). Not only that, too many variables make the derivation of a model computationally difficult. A pairwise correlation tests on the parameters was run, so as to reduce the number of variables used. As a result the number of potential texture indices was reduced to 15.

For the percentages of bare soil, woody vegetation and photosynthetically active non-woody vegetation the Shadow Index (SI) [Rikimaru 2002] , the Bare Soil Index (BSI) (Rikimaru, 2002) and the Modified Chlorophyll Absorption Ratio Index (MCARI) (Eitel, 2007; Haboudane, 2004) were used. These indices applied to a sub-object

chessboard (i.e. pixel level segmentation) within each object. For each of these sub-object classes thresholds were established using visual examination. The average threshold for the shadow used was 0.889 (SD 0.02) for the MCARI and for the bare soil index 0.903 (SD 0.07). In consequence, a part from the biophysical reflectance and texture data, each polygon, whether for the object corresponding to the field data or for the rest of the objects, was assigned a percentage of woody, bare soil or non-woody (green) vegetation.

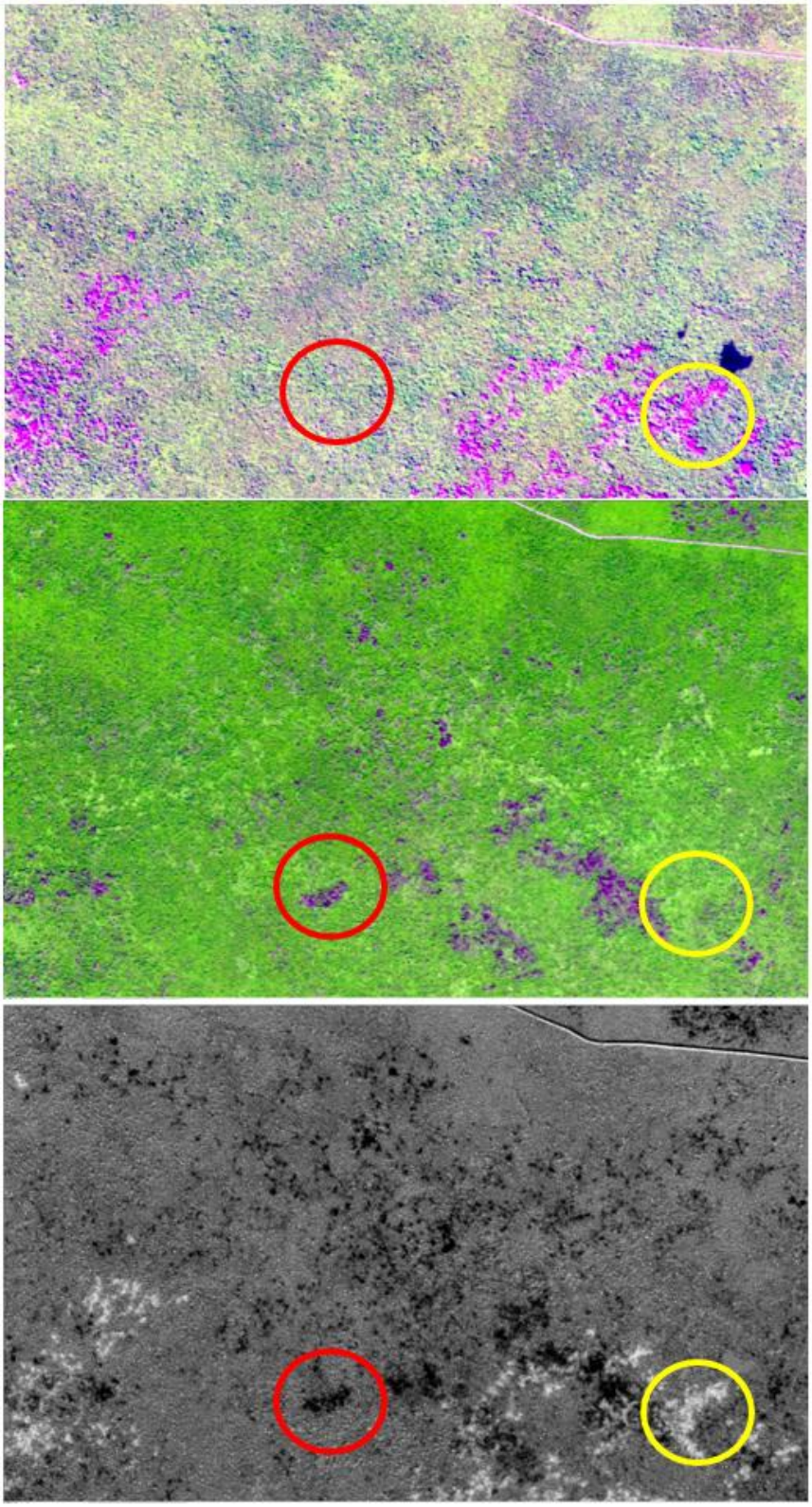

Figure 44: RapidEye data (NIR,RE, RED) from the Pugo hills showing 2010, 2013 and the change in shadow index. Yellow circles show an area that was disturbed in 2010 and regrew in 2013. Red circles show new disturbances in 2013.

7.8 Reviewing the field data with respect to the RapidEye data

To review and assess the correspondence between field data and image data, a graphic interface was developed so as to display plot by plot the field site, the image data, the plot location and the plot information.

This initial review raised a number of questions as to how to correlate the field data to the remote sensing data. Firstly the field plots are spatially small compared to the satellite data. Despite having relatively fine spatial resolution RapidEye data, the field data (15 m circle plots) cover only 27 RapidEye pixels. Secondly, preliminary analysis of the remote sensing data, forest management and UNFCC rules combined to suggest that a 1 ha (200 pixels) mapping unit is more practical. Thirdly, a number of issues relating to the accurate geolocation of the field data arose– especially given the accuracy +/- 7m of the GPS system used (Garmin C60) in the field and know problems changing between datum Arc60, used in topographic maps of East Africa - and hence in the field survey, and the satellite reference datum (Moses, 2012).

Fourthly, the visual inspection of the data using the interface found a number of discrepancies in the geolocation of various RapidEye scenes. Fifthly, a number of major issues arise from trying to relate field data from one specific date to remote sensing data acquired at a different date, allowing for possible changes in the land cover between the two data sets. Given the magnitude of the exercise the field data collection took three years; satellite data were only available for one of these years. Sixthly: Data collected in the field did not adequately provide information for calibration or comparison remote sensing. Within the 15 m radius no systematic estimation of the respective cover of trees, shrubs or other class was available. On a number of occasions, when given, the canopy cover did not correspond to that seen from the satellite image- perhaps due to georeferencing problems, or differences in the time between the field visit and the image acquisition, and canopy density is known to be difficult to measure with accuracy from the ground (Jennings, 1999). The land cover classification given to the field teams was not adapted to providing adequate field data for calibrating remote sensing data. The vast majority of plots were classified as ‘woodland’, without an indication as to the percentage of tree cover. The average tree height provided in the data analysis, could relate to one tree in a field or a full cover forest. Seventh: while land cover may not change throughout the year, its aspect does, especially in the tropics, predominantly due to seasonality. It may be in a lush green phase, drying out, exceptionally dry, burnt or flooded. All these present major different spectral signatures. Finally, in using a 1 ha MMU, it was found that the field plot was not always representative of the 1 ha image object on the remote sensing data.

These combined factors pose three major problems:

1. The geographic matching of the data

2. The consistence of the aspect of the land cover – i.e. the same land cover may present a different reflectance depending the state and condition of the vegetation (dry, burnt flooded)

3. Different combinations of woody biomass within the sample plots can give the same above ground biomass – e.g. one single large tree in the plot, surrounded by bare soil might give have the same biomass of a plot with 100% cover of shrubs.

7.9 Developing models for predicting AGB, BA and TH

Four predictive models were tested for biomass, basal area (BA) and tree height (TH) and evaluated their accuracies. The input data of the models are image bands and their textures, while the output are biomass, basal area and tree height. The four models include two simple parametric models from inferential statistics, i.e. a linear model and an exponential model, and two nonparametric machine learning models, i.e. a Random Forest (RF) and a Support Vector Machine (SVM).

In the linear model, the dependent variables (y) is a linear function of several independent variables (x), where each of them has a weight (i.e. a regression coefficient β). Thus, the model takes the form:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_n x_n$$

Results are usually interpretable. However, linear model requires a linear relation between the dependents and independent variables, while allometric equations describing tree parameters, are generally non-linear.

For this reason of non-linearity, another model was used carrying out the logarithmic transformation to the dependent variables. This model is called exponential regression. This model takes the form:

$$\log(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_n x_n$$

The third model tested is the Random Forest (RF), a tree-based method. The inner nodes of the tree hold decision rules according to explanatory variables (e.g. less/greater than a variable *x*), recursively splitting the data into sub-spaces. The leaf nodes at the end of the decision tree contain models for the response variable. Because a single tree is generally not effective enough to cope with strong non-linear multivariate relationships, a tree ensemble method was applied: Random Forests that combines regression trees grown from different bootstrap samples and randomly selected features at each split node (Breiman, 2001). The RF is thus an improved ensemble method that builds several decision trees (weak learners) and outputs the mean prediction of the individual tree models. The decorrelation of trees is achieved through the random selection of the input

explanatory variables (Hastie et al., 2001) by bootstrap methods. Typically, 63% of the data is used for training (in-bag data) and the remaining 37% (out-of-bag data) for validation.

The choice of Random Forest was mainly motivated by its advantageous features: (i) RF runs efficiently on large databases (i.e. it is relatively fast to build and even faster to predict), (ii) RF is resistant to outliers and over-training, (iii) RF does not require cross-validation for model selection, (iv) RF provides further information about the most relevant variables inputted in the model, (v) RF is fully parallelizable. The main limitation of RF is the impossibility to predict beyond the range of the response values in the training data.

The fourth model tested was the Support Vector Machines (SVM). Support Vector Machine (SVM) methods (Vapnik, 1998; Smola and Schölkopf, 2004) define a linear prediction model over mapped samples to a much higher dimensional space, which is nonlinearly related to the original input space. The Support Vector Regression (SVR) is the SVM implementation for regression, which has yielded good results in modelling some biophysical parameters (Camps-Valls et al. 2006). The main limitation of the SVR is that the free parameters must be tuned using cross-validation methods, which is a time-consuming task.

Machine learning methods generally improve accuracy, but it comes at a cost: interpretability. Random forest and SVM are akin to black box models, i.e. without any knowledge of its internal workings. Anyway, with Random Forest one still has an estimate for variable importance.

For the phases of training and testing of the models, the ground truth dataset is employed, i.e. the points where there are actual biomass, BA/TH measurements. Models have been built using calibration data, and evaluated on validation data. Validation is required since over fitting the calibration data results in poor performance in the validation. To that end, a random 70:30 split in the data set (i.e. 70% calibration, 30% validation) was used.

Figure 45, Figure 46 and Figure 47 show the scatterplots of the difference between observed and modeled Biomass, TH and BA of the four different models: linear regression, exponential regression, Random Forest and SVM.

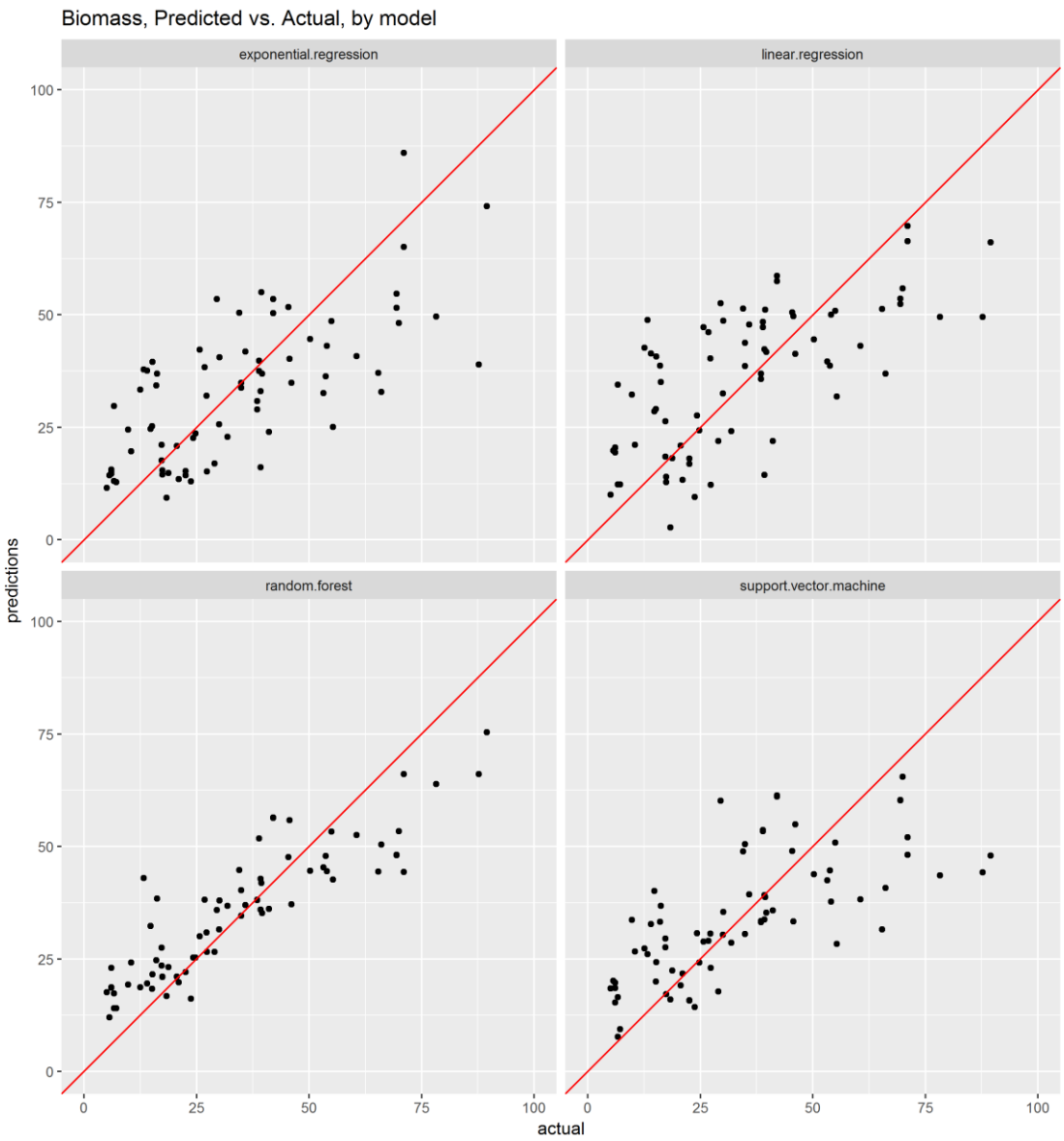

Figure 45: Scatterplot of Predicted vs. Modelled Biomass

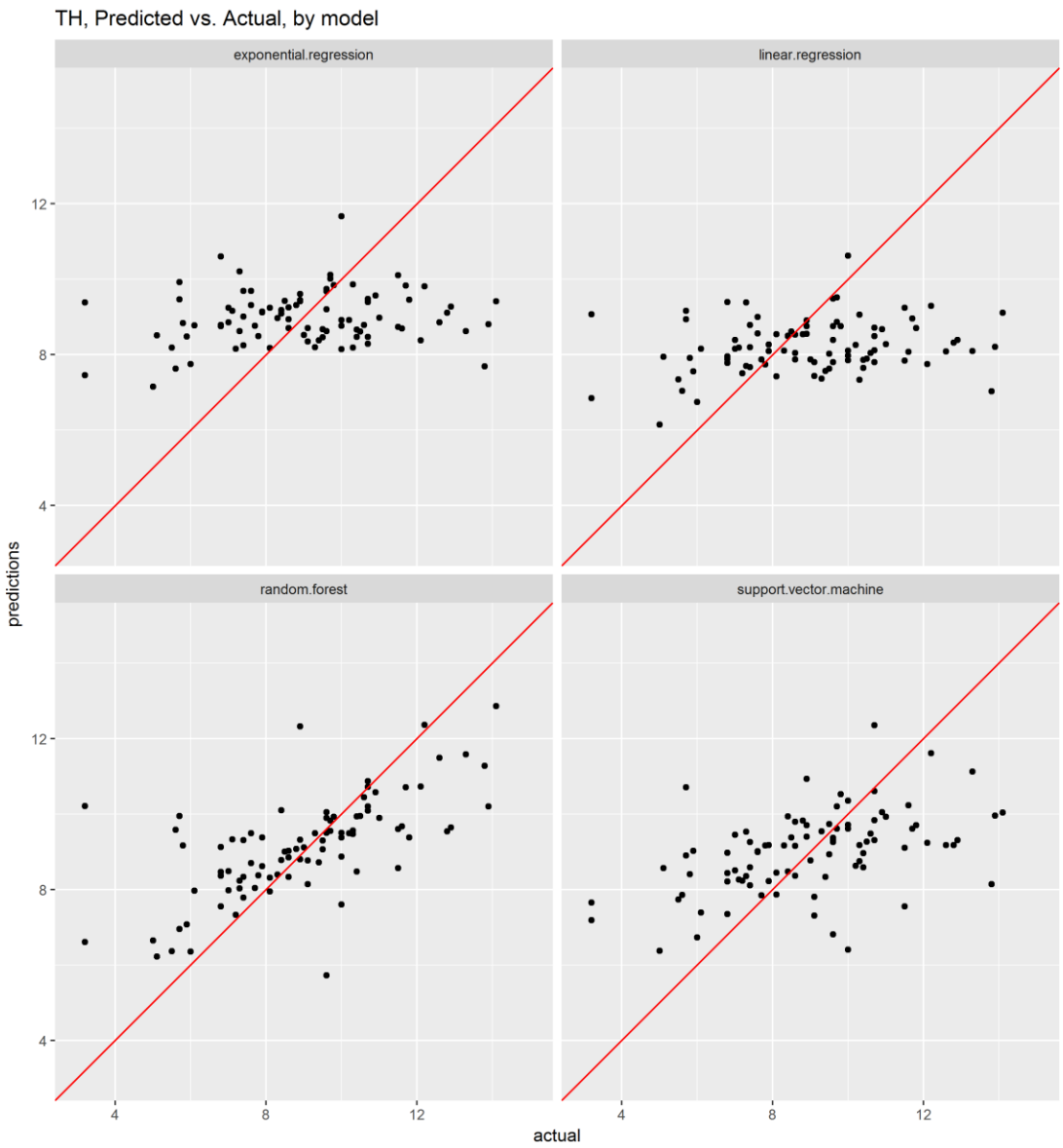

Figure 46: Scatterplot of Predicted vs. Modelled Tree Height.

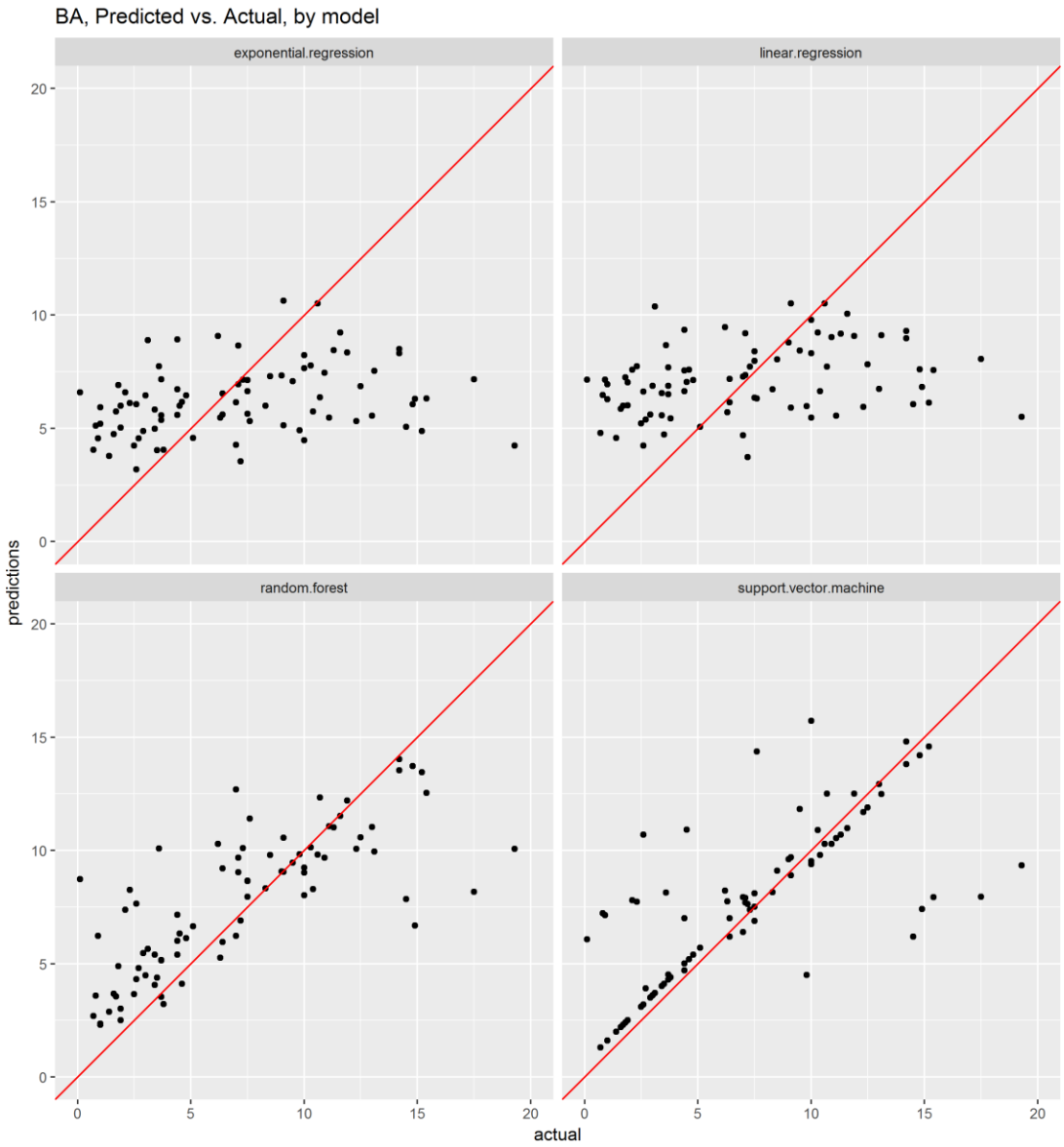

Figure 47: Scatterplot of Predicted vs. Modelled Basal Area.

The four methods have been compared in terms of accuracy and goodness of fit. Table 21 summarizes the main statistical indicators, i.e. the RMSE, the Mean Absolute Error (MAE) and the coefficient of determination (R^2) for Biomass. R^2 is indicative of the proportion of the variance in the predictive variable (e.g. biomass) explained by the model. TF followed by the SVM provided the most accurate results for biomass and Tree Height. SVM outperforms RF for Basal area. The two simple parametric methods, the linear and the generalized regression, clearly presented the worse results for all predictive variables.

Table 21: Statistical indicators of modeled data versus observed Biomass across validation dataset

Model	RMSE	MAE	R^2
Generalized Exponential Regression	19.22	14.20	0.11
Generalized Linear Regression	19.38	14.29	0.11
Random Forest	16.86	11.16	0.56
SVM	20.09	12.92	0.34

Random Forest provides reliable predictions for the forest variables. Specifically, the RF of Biomass is able to explain ~56% of the variance. The RF provides also a helpful estimate for the variable importance (figure 13). We can observe as the most relevant variable inputted in the biomass model is the relative shadow, followed by the textures.

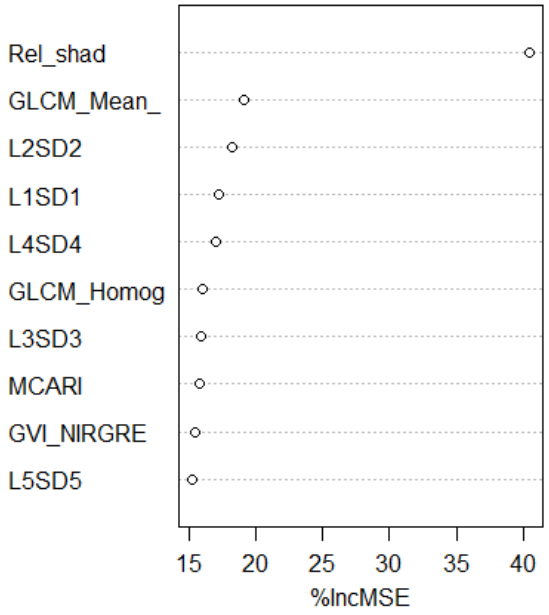

Figure 48: The importance of variables for AGB in the Random forest model, showing the increase in mean standard error (MSE) when removing a given variable.

7.10 Processing of Landsat 8 and Sentine-2 Mosaics to AGB maps

While the statistical extrapolation of field plots (combined with remote sensing) provides valid estimates for national AGB, wall to wall maps allow the potential not only for improved planning and management, but also when combined with mapped activity data (i.e. forest change) give a better correspond of AGB changes for emissions. Despite their lower spatial resolutions (10-20m S2 and 30m L8) their free access, large area coverage and higher temporal acquisitions make them attractive as a potential source of information for AGB mapping. Using the Google Earth Engine facility, full resolution mosaics from S2 and L8 for Tanzania were created. For L8 all reflective channels (6) were used, while for S2, the data volume limited us to using three channels, RED NIR and SWIR.

The L8 mosaic has been created using images from May 2013 to April 2014, whereas the S2 mosaic from Jan 2016 to Dec 2016. Note that the lack of temporal overlap between the two mosaics is due to the cloud cover of L8 during 2016 that hindered the creation of a cloud-free mosaic.

The AGB computation scheme is organized through the following steps:

1) The L8 and S2 reflectances and vegetation indices corresponding to the ground truth data (i.e. the AGB dataset) were extracted from the mosaics.

2) A Random Forest Regression Model is performed to establish a relationship among reflectances, vegetation indices and AGB. The proposed methodology supplies a quality flag. The flag documents the goodness of fit of the regression and the importance of the variables employed in the regression.

3) The Random Forest model using the reflectances and vegetation indices was applied at the original spatial resolution of 30 and 10 m for L8 and S2, respectively.

It is worth noting that the AGB computation scheme employed for L8 and S2 is different from RapidEye: with pixels and not segments being used.

The results gave a correlation coefficient R^2 of 0.17 and 0.21 and a mean AGB of 22.4 and 22.6 ha^{-1} for Sentine-2 and Landsat 8 respectively. Figure 49 shows the AGB using both Sentine-2 and Landsat 8 mosaics, and Figure 50 shows the differences between them (S2 – L8). Both AGB maps present a similar spatial pattern, revealing a reasonably high consistency and small differences (RMSE=8.8, MAE=0.2) between L8 and S2 estimates.

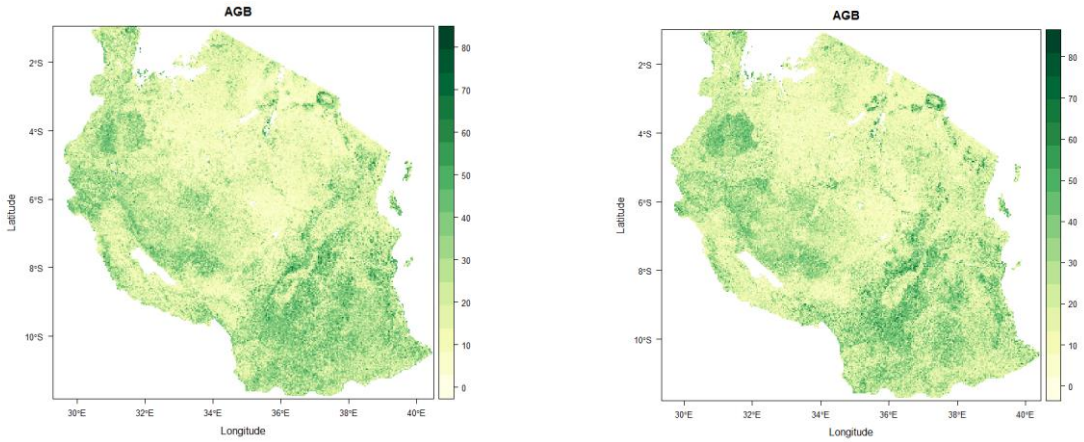

Figure 49: Aboveground Biomass (AGB) using the Landsat 8 mosaic left and Sentine-2.

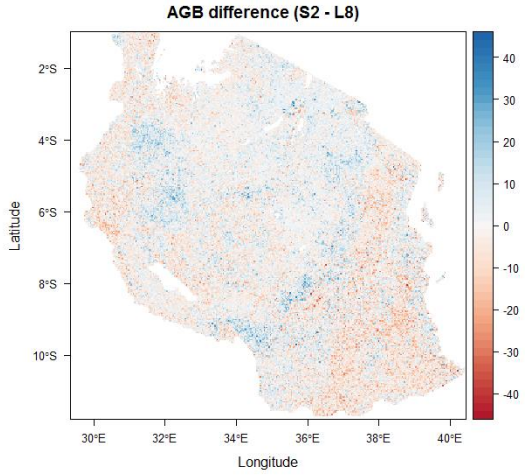

Figure 50: Difference of Aboveground Biomass obtained from Landsat 8 and Sentine-2.

7.11 Results

The national level

For the 76 sample sites, covering 3.3 % of the country, an average of 22.1 tons per ha of above ground biomass (ABG) per hectare is found – bearing in mind this is only for those landscape units that are classed as being 'woody biomass'. Using the areas provided by NAFORMA for forests and woodlands 54,534,500 ha, direct expansion was used to extrapolate to the country level using the Horvitz-Thompson Direct Expansion Estimator

(see Särndal[4] et al. 1992 for a general discussion of Horvitz-Thompson Estimators). The application of the direct expansion method has been used in various studies (Gallego, 2004; Achard et al., 2004; Eva et al., 2010)

$$AGB = D.\bar{y}$$

where D is the total area of the study region class and $\bar{y}$ is the average AGBha^{-1} of the sample areas. This results in a total of 602,474 MtC for the country and compares to the national estimate from the 32,000 field plots of 624,984 MtC, a difference of 4%. The standard error of the estimate is 5.9 %, Land cover, and hence biomass, is not generally distributed in a homogeneous manner, especially in Tanzania which exhibits land covers of highly different levels of biomass. From Figure 45, however, it can be seen that the variation in the estimated AG biomass against the actual one is high.

Analysis by ecoregion

Using the location of the individual sample sites the average above ground biomass was calculated for the respective ecoregions (Table 22). Due to the limited number of samples in the other ecoregions, the most informative results are Somalia-Masai and Zambezian.

Table 22. Estimates of average AGB (tha^{-1}) by White ecoregion

	Afromontane	Zanzibar-Inhambane	Somalia-Masai	Zambezian Lake Victoria	
Count	5	6	19	41	4
Minimum	17.3	15.1	9.8	11.6	17.8
Maximum	27.7	30.4	22.4	34.3	25.9
Mean	20.9	21.5	16.7	25.2	22.3
SD	4.9	5.6	4.1	4.6	3.3

[4] Särndal, C. E., Swensson, B., and Wretman, J., 1992. Model Assisted Survey Sampling, Springer-Verlag, New York.

Relative Efficiency

To assess the improvement brought by the use of the satellite data the relative efficiency is calculated. This is obtained by the ratio of the variance (VAR) of one estimate, the estimate from RapidEye combined with field data (VAR_{REFD}) to the other estimate (field data alone VAR_{FD}), to see what advantage the introduction of the remote sensing data brings. A relative efficiency of X would mean that the improvement (i.e. the accuracy as measured by the variance of the estimate) brought about by the introduction of the remotely sensed data could be achieved by using X times as many field plots (Gallego 2004; Hansen et al. 2015). In this case a RE of 2.9 was found, meaning that to obtain the same levels of accuracy, three times as many field data plots would be needed as could be used with employing RapidEye data combined with field data.

Relative Efficiency = VAR_{REFD} / VAR_{FD}

It is difficult to quantify the potential financial savings the inclusion of remote sensing data would bring to the project, without knowing the direct costs of field survey. In the NAFORMA project the total spent on salaries for the field survey component alone was around $5 million (FAO, 2009). However, this includes household surveys (3500) and soil data collections (4000). The costs of overheads, vehicles and fuel are not included. The costs of RapidEye data depends on the area covered, with volume discounts available. However, full country coverage for these type of data can be expected to be up to $ 0.90 million (approx. $1.1 / sqkm).

With a relative efficiency of 2.9, one can conclude that a combination of 11,000 field plots (instead of 32,000) could be used with RapidEye data to obtain the same level of accuracy. This result in a large saving of the budget and in time – the three year survey could have been completed in one year.

Sentine-2 and Landsat mapping

The recent availability of Sentine-2 data, combined with the processing capacity afforded by the Google Earth Engine facility, is an attractive prospect for wall to wall mapping. Here, however, the correlation between these decametric resolution sensors and the ground data is weak.

7.12 Discussion

In preparing data for testing models to relate RS parameters to field data (AGB BA TH) a major effort was required in ‘cleaning’ the field data. Not only that, the quality of the information and of the geo-location

between the two data sets was not as optimal as was desired. The benefits derived from linking fine (5 m) resolution RS data with field data are demonstrated in that, regardless of the shortcomings, a relative efficiency of 2.9 was obtained by integrating the RapidEye data with the field data.

Lower resolution (10-30m) data were too coarse to provide robust estimates of AGB, however, spatially they have the potential to give general distribution of biomass at the national level, and should be considered for providing ancillary temporal information on land cover class and condition. A number of key issues arise from the work linking the RS data to the extensive field data set.

i) Co-location of the data sets.

The strategy for the field site visits for field data plots was organized by the national project (NFAORMA) independently of remote sensing considerations in mind. Field teams, using the national topographic maps (in Arc60 datum) navigated to the pre-selected sites. Such field visits should be supported by cartographic and digital extracts of the field locations. Both high cost (high precision GPS with inbuilt satellite and/or map visualisation) and low cost (smart phone / pad and PDF maps) exist to facilitate both navigation and on site location. When local datum are used, so as relate to the national mapping datum, dual GPS locations with WGS 1984 should be employed.

ii) Data collected

The data collected for the national assessment were suited for the statistical assessment of biomass and for other parameters. However, for this study the data collected were found to be insufficient for linking them to RS data. RS optical data essentially records the vegetation canopy cover, its condition (dry / green) and its structure, exploiting a set of direct or derived parameters from the multispectral data. Field data collected need to correspond to these potential indicators. This means that vegetation cover, height and repartition within a sample site need to be recorded. Ideally, a vector map of the field sample site should be produced in conjunction with satellite imagery.

iii) Data cleaning

Apart from the geolocation issues, from the RS point of view a major 'disturbance' (i.e. high variation in the reflectance) in the spectral signature of the field plots was the inclusion of non-woody, low biomass sites. These sites exhibited an exceptionally high variation in land cover and land cover condition (and hence reflectance and texture) – barren surfaces, agriculture exploitations, grasslands and park savannah, each of which manifested different states – burnt, flooded, green flush, senescence, bare soil. Either an intelligent classifier needs to be employed (decision tree – see below) or they should be removed from the data set, as was employed here.

iv) Sampling sites

Within this study the RapidEye data were provided free of charge through GMES. These location of the sites was selected to contribute to the FAO FRA / JRC global forest resource estimates [JRC]. As such, the sample locations, at the confluence of the geographic lat/long grid, were designed to provide regional to continental estimates, and not suited to national estimates. These sample sites covered 3.3% of the land surface of Tanzania. Two major changes would bring more reliable results; far more, smaller sample sites and a stratification: in terms of number of sites, rather than use 76 samples of 20 km by 20 km each it would be more appropriate to expand to approximately 1200 sample sites of 5 km by 5 km each. The increase in sites allows a far better estimation of the variance of the land cover (Olofsson et al., 2014), and quality control is facilitated.

With such an increase in the number of sites reviewed, a better repartition of sites can be envisaged, using a stratification of sites by ecozones or another stratum.

v) Ancillary data

The parameters entered into the model came exclusively from the single date RapidEye data. While these (RapidEye) data are superior in terms of resolution, they are limited for temporal analysis. With the arrival of the Sentine-2b sensor to join Sentine-2a, 10m resolution data will some be available at a 5 day frequency. Simulation of this scenario with the SPOT4take5 program (Hagolle et al., 2013) has already shown that differences between intact and degradation dry forest formations can be monitored (Chapter 6). These data could supplement the finer spatial resolution data as an ancillary input into the model.

At the same time, clearly, the potential biomass for a given site, depends largely on environmental conditions; precipitation, altitude, soil and temperature variations. These data were not taken into account in the models as the actual biomass is determined by the anthropogenic impact.

vi) Production of Landsat and of Sentine-2 data

It is clear that neither the Landsat 8 nor Sentine-2 data processed here can provide accurate quantitative data on woody biomass. Nevertheless their respective capacities to provide multispectral, synthetic information on the land cover of the full country at 10-30 m resolution demands attention.

vii) Direct results

The estimates from the combined RS and field data at the national level are similar to that of the estimate achieved with the full 32,000 field plots. This is to be expected, as the model fit provides a replica of the input

training set. In terms of the results by ecosystem, there is a clear difference between the average biomass of the Somalia-Masai and the Zambezian regions. The other zones too are too small to consider, however the Zanzibar-Inhambane, home of the Coastal forests should have a high average AGB, however it is clearly impacted by anthropogenic impact of Dar es Salaam and Morogoro, the two main urban centres in the country (Hojas Gascón et al. 2016).

7.13 Conclusions

The use of finer spatial resolution optical data (5 m) opens the door to better mapping and estimation of AGB when combined with reliable field data. A major recommendation arising from the work is that adequate precautions are taken in the design, planning and execution of field data, in that they can be easily matched to image data and that they collect the type of information that relates to the biophysical data collected by satellite imagery. Textural measures, information on woody vegetation cover (via a shadow index) combined with a Random Forest model are effective in providing a prediction of AGB.

In terms of supporting REDD initiatives, this means that national inventories can be carried out in a more rapid and cost effective way. The results obtained, however, do not allow the method to report on changes in biomass levels, unless they are significant. This 'activity' data would have to come from direct satellite observations of forest cover change.

8. Urbanization and forest degradation in East Africa - A case study around Dar es Salaam, Tanzania

A shortened version of this chapter has been published in:

Hojas-Gascón, L, H.D. Eva, D. Ehrlich, M. Pesaresi, F. Achard and J. Garcia-Haro, 2016. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing Special Issue on IEEE 2016 International Geoscience and Remote Sensing Symposium (IGARSS 2016).

8.1 Introduction

Dar es Salaam is one of the fastest growing cities in Africa, with an average annual population increase of 4.1% (compared to the national rate of 2.9%) over the last twenty years (Kombe, 2005). This has resulted in a peri-urban sprawl which when combined with economic stagnation has produced an increase in urban poverty. The majority of households rely on fuelwood for cooking (fossil fuels such as kerosene being too expensive and used only for lighting). As a result of this increasing and highly concentrated demand for fuel wood, the forests around the city have been rapidly depleted, and the search for wood source has been extended further away and even into protected areas. The Global Human Settlement Layer (GHSL) for 2000 and 2014 (Freire and Pesaresi, 2016) in conjunction with the Global Forest Change database (Hansen et al., 2013b) is used to document the patterns of deforestation from 2000 to 2014 along with the spatial expansion of the city, and the influence of certain environmental and urban variables.

8.2 Background

Tanzania is losing between 200,000 and 400,000 ha of forests a year (see Chapter 1). It has been estimated (Makundi, 2001) that 90% of this loss is due to charcoal and fuelwood consumption and that Dar es Salaam accounts for 30% of the national consumption. With the population of Tanzania set to double by 2050, clearly the impact on the country's forest cover will be enormous unless alternate sources of energy are found and offered to the population at affordable prices.

8.3 Materials and Methods

Data

The growth of Dar es Salaam is mapped out using the recently available Global Human Settlement Layer, which map the urbanization extent for four epochs 1975, 1990, 2000 and 2014. These data are derived from the extensive Landsat archive of 80m and 30m spatial resolution optical images using image texture feature data sequencing and Symbolic Machine Learning. Data on the spatial patterns of deforestation are taken from the Global Forest Change dataset, again derived from Landsat Thematic and Enhanced Thematic Mappers. This dataset provides two products: a 'percentage tree cover' map for the year 2000 at 30m spatial resolution, and deforestation events mapped annually from 2000 to 2014, data at the pixel level with either 0 (no loss) or else a value in the range 1–14, representing loss detected primarily in the year 2001–2014, respectively. In the absence of earlier data on the spatial dynamics of deforestation, we confine our analysis to this period.

The protected areas are extracted from the IUCN World Database of Protected Areas (UNEP-WCMC, 2015).

The environmental and urban variables which have considered in this study to influence the spatial distribution of deforestation due to population growth are related to *topography* (altitude, slope, and aspect), *accessibility* (distance to roads, towns, and forest edge), *deforestation history* (distance to previous deforestation, i.e. deforestation of the 2000-2005 period). In our study, we use the variables shown in Table 23.

Table 23: Potential explicative variables of deforestation and forest degradation spatial distribution.

Product	Source	Variable derived	Unit	Resolution (m)
Deforestation maps (2000-2005-2010-2014)	Hansen et al. 2013 (1)	forest/non-forest	-	30
		distance to forest edge	m	30
		distance to previous deforestation	m	30
Deforestation maps (2000-2005-2010-2014)	Hansen et al. 2013 (1)	forest/non-forest	-	30
		distance to forest edge	m	30
		distance to previous deforestation	m	30

Digital Elevation Model	SRTM v4.1 CSI-CGIAR (2)	altitude	m	90
		slope	°	90

1) http://earthenginepartners.appspot.com/science-2013-global-forest/download_v1.2.html, Hansen Global Forest Change v1.2 (2000-2014), [3]

2) http://srtm.csi.cgiar.org, SRTM 90m Digital Elevation Database v4.1

3) http://www.geofabrik.de, data extracts from the OpenStreetMap project

8.4 Spatial Analysis

The goal of the study is to see how the supply of fuelwood for Dar es Salaam has developed over the last 15 years. As resources are depleted close to the city, supply must come from further away. In particular it is important to see if the demand is impacting on the preservation of protected areas.

Extracting human settlements areas for the city and the surrounding areas from the GHS layer, one can compare the population density using this layer's urban extent with that from official data sources, which use the administrative boundaries of the city.

From the percentage tree cover map for the year 2000 only pixels with more than 20% tree cover were used, in the assumption that pixels with lower than this threshold are unreliable due to resolution. The percentage of forest loss occurring at different years was analysed in a series of concentric buffers of 10km width spreading out to 100km from the city (see Figure 52). The analysis was limited to 100km radius from the city so as not to include wood supply to the next closest city, Morogoro, the second largest conglomerate in Tanzania.

The protected areas up to 100km around the city were used, extracted from the IUCN World Database of Protected Areas (UNEP-WCMC, 2015). A total of 24 protected areas were listed. For these the percentage of forest (tree cover > 20%) lost between 2000 and 2014 was calculated.

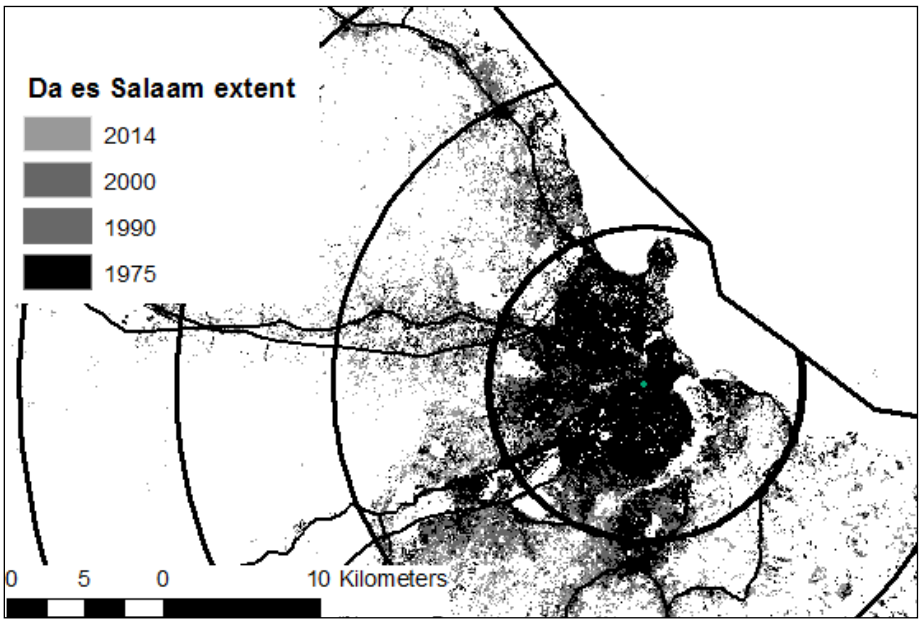

Figure 51: Urbanization expansion in the first 10km buffers around the city of Dar es Salaam from 1975 to 2014.

To obtain the current forest cover map and the total deforestation distribution the forest cover map of 2000 and the forest-cover change raster between 2000 and 2014 from the Global Forest Change database (shown in Figure 52) was used. The map on the left represents forest cover and deforestation patterns for the entire Tanzania, while the map on the right shows the area of Dar es Salaam.

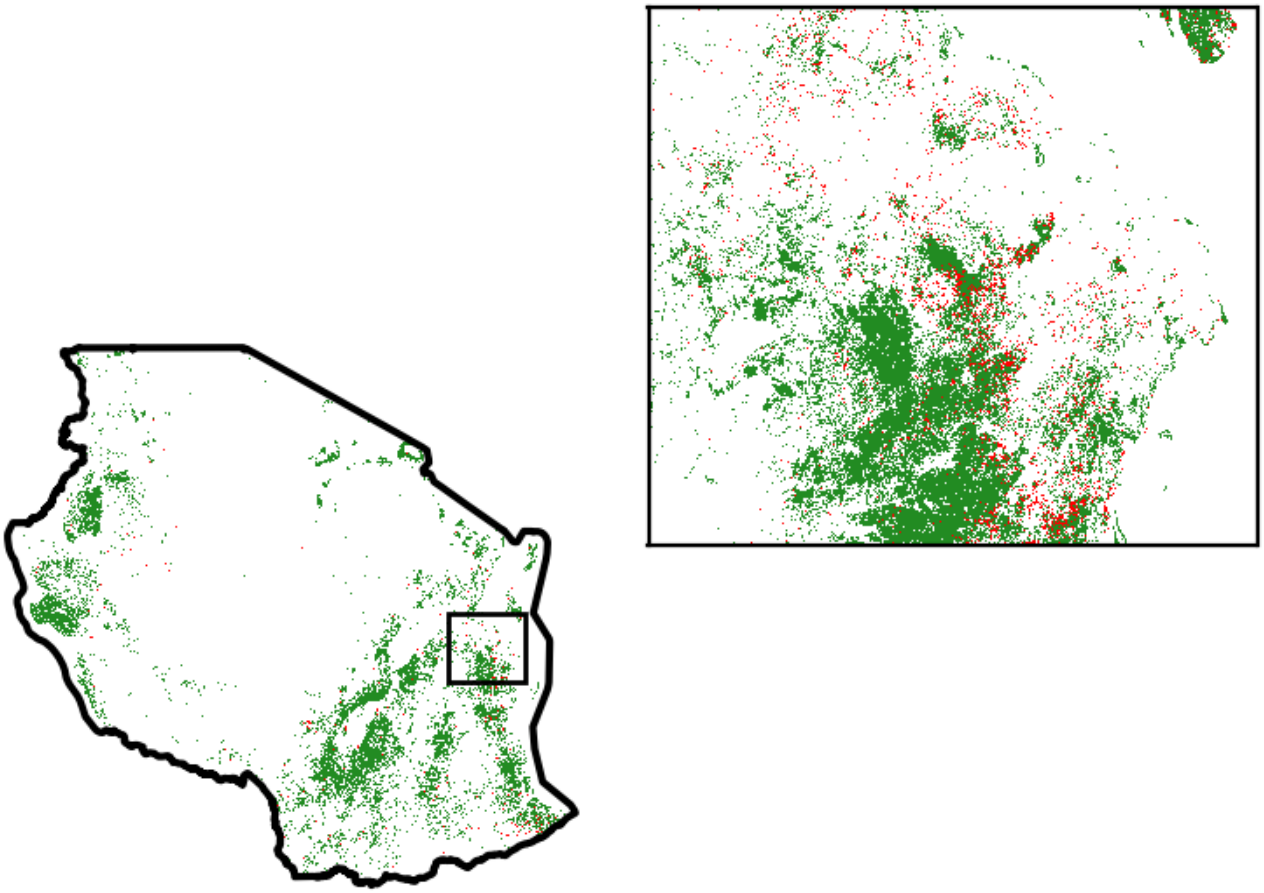

Figure 52: Remaining forest cover of 2014 (in green) and deforested area between 2000-2014 (in red) for the entire Tanzania (left) and for the area of Dar es Salaam (right).

The spatial aspect of the environmental and urban variables are shown in Figure 53. Note that the maps refer to the entire Tanzania, although the study area is the box of Dar es Salaam, as shown in Figure 52. Altitude, aspect and slope are derived from the Digital Elevation Model., *dist_road* is the distance to road (from the Open Stree Map archive), *dist_defor* refers to the distance to previous deforestation, *dist_edge* refers to the distance from the forest and *dist_town* refers to the distance from towns (for the entire Tanzania). Previous deforestation is the deforestation that occurred in the period 2000-2005.

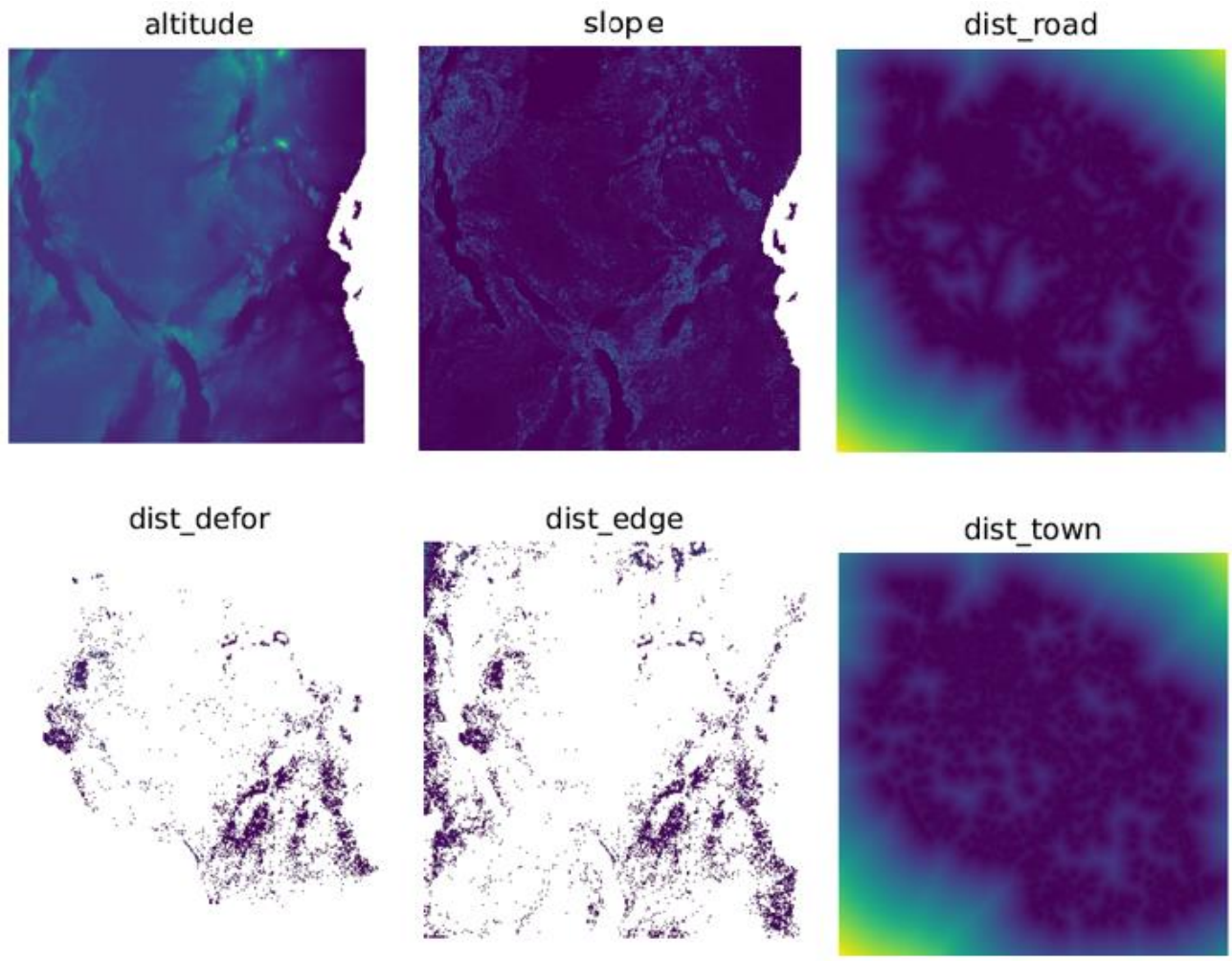

Figure 53: Environmental and urban variables considered in the study.

To find the relation between deforestation and these environmental and urban variables, 10,000 points (pixel centers) were sampled in the deforested areas and in the remaining forest (20,000 points in total) for the entire Tanzania. Sampling is necessary to reduce the computation burden. Then the values for each of the variables was extracted for each sampled point. Figure 54 shows sampled observations in the study area. Dark red dots indicate deforestation observations.

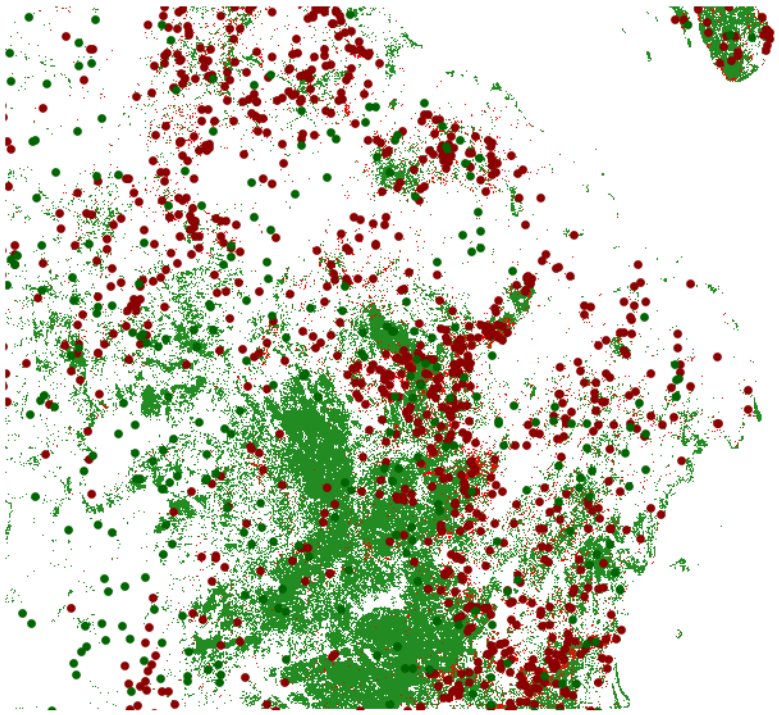

Figure 54: Red (*green*) dots refer to the sampling points in the deforested (*forested*) areas around Dar es Salaam.

The Binomial logistic regression model was used to estimate the deforestation probability of a pixel given the set of environmental variables. Logistic regression is a method for fitting a regression curve, *y = f(x)*, when *y* is a categorical variable. The typical use of this model is predicting *y* given a set of predictors *x*. The predictors can be continuous, categorical or a mix of both. The categorical variable y, in general, can assume different values. *y* is binary meaning that it can assume either the value 1 or 0 (i.e. forest/non forest).

8.5 Results

It was found that the city of Dar es Salaam grew from 229 km^2 in 1975 to 561 km^2 in 2014 (Figure 51). During the same period the population rose from 760,000 to 4.2 million inhabitants. The population density of Dar es Salaam is now (2017) approximately 7500 inhabitants per km^2, double that of 1975. Official population density statistics, based on the full area of the municipalities give half this rate for comparable years (United Republic of Tanzania, 2014). The main expansion has been along the main roads leading into the city.

Analysis of forest loss show a number of dynamics. Initially, forest loss increased when moving further from the city until a certain distance, suggesting that as forest resources close to the city are depleted, supply has been sought from further afield (Figure 55). In the year 2000 wood resources were harvested mainly at 50-70 km from

the city. Wood resources at more than 70 km were too costly for transportation. In 2014, nearly half the supply was coming from over 80 km away, compared to 30% fifteen years earlier. However closer to the city (up to 30km) forest loss increased slightly in 2014. This may indicate differences in the type and quality of wood being harvested; low value wood from already degraded forests close to the city for wood fuel or charcoal production for consumption of the growing population, and higher value timber from further away, for construction, exportation and furniture (Ahrends et al., 2010).

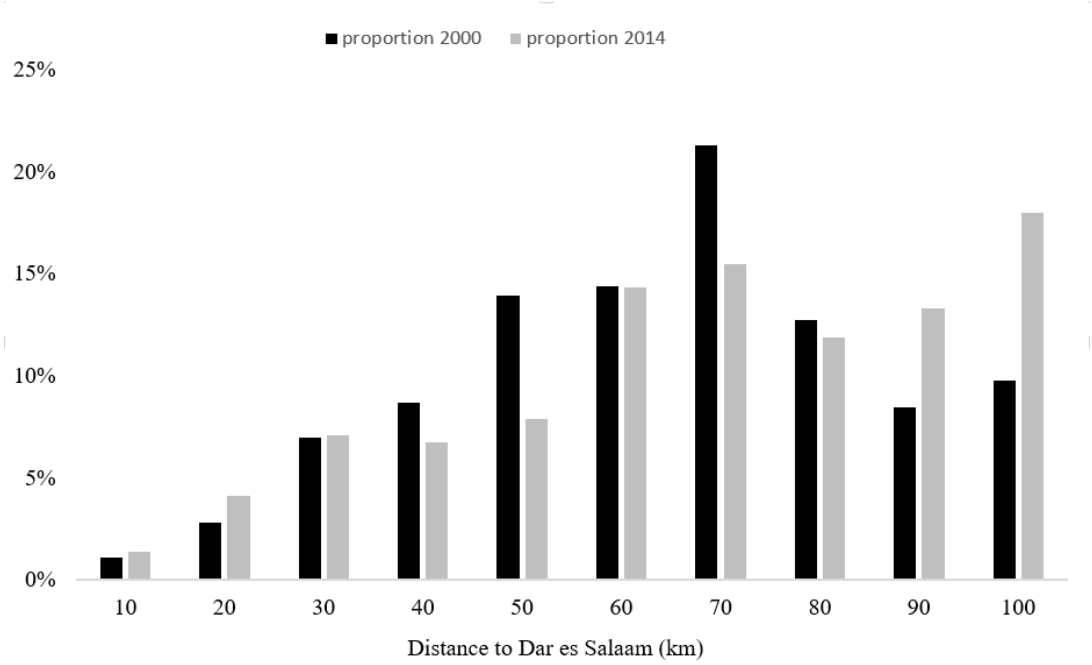

Figure 55: The proportion of forest loss at increasing distances from Dar es Salaam, comparing the year 2000 with that of 2014.

Incursions into protected areas increased over the period as the population density rose and the demand for fuelwood and construction materials increased (Table 24). Of the 24 protected areas situated within a 100 km radius of Dar es Salaam, nine have lost more than 15% of their forest cover (tree percentage higher than 20%) in the last fifteen years, 5 more than 25% and two parks, Masangan and Kingoma over 45%. To a large the analysis of the forest dynamics using the Global Forest Change database hides forest degradation, which is usually too 'subtle' to be detected by the Landsat sensor's 30m resolution. As shown in Chapter 4 (Hojas Gascón et al., 2015) the Landsat sensor, while adequate for mapping deforestation, cannot pick out forest degradation. The results presented, will therefore underestimate the impact of the growing city.

Table 24: Percentage of forest lost in protected areas within 100 km of Dar es Salaam.

Masangan	48%	Ruvu North	23%
Kingoma	46%	Pugu	17%
Kazimzumbwe	34%	Marenda	18%
Pugu North	25%	Simbo 2	20%
Kikoka	27%	Mtita	13%

In Figure 567 and Figure 57, the relationships between the sampled environmental and urban variables and the probability of deforestation are shown. Histograms on the left show the occurrence of deforestation (from the sampled points) of different classes. Distances (i.e. dist_*) and altitude are expressed in meters, whereas slope is expressed in degrees. The lines on the right show the deforestation probability's behavior according to each variable. It is clear that there is an abrupt drop in deforestation probability when the distance from road is greater than 10 km. Specifically, the deforestation probability drops to 50% when the distance is greater than 10 km. Also, distance to settlements, deforested areas and forest edge have an important influence on the occurrence of deforestation. The closer to towns, roads, deforested areas and forest edges, the higher the probability of deforestation. Also slope and altitude have strong relationship with deforestation: the probability of deforestation declines when the slope and the altitude grow.

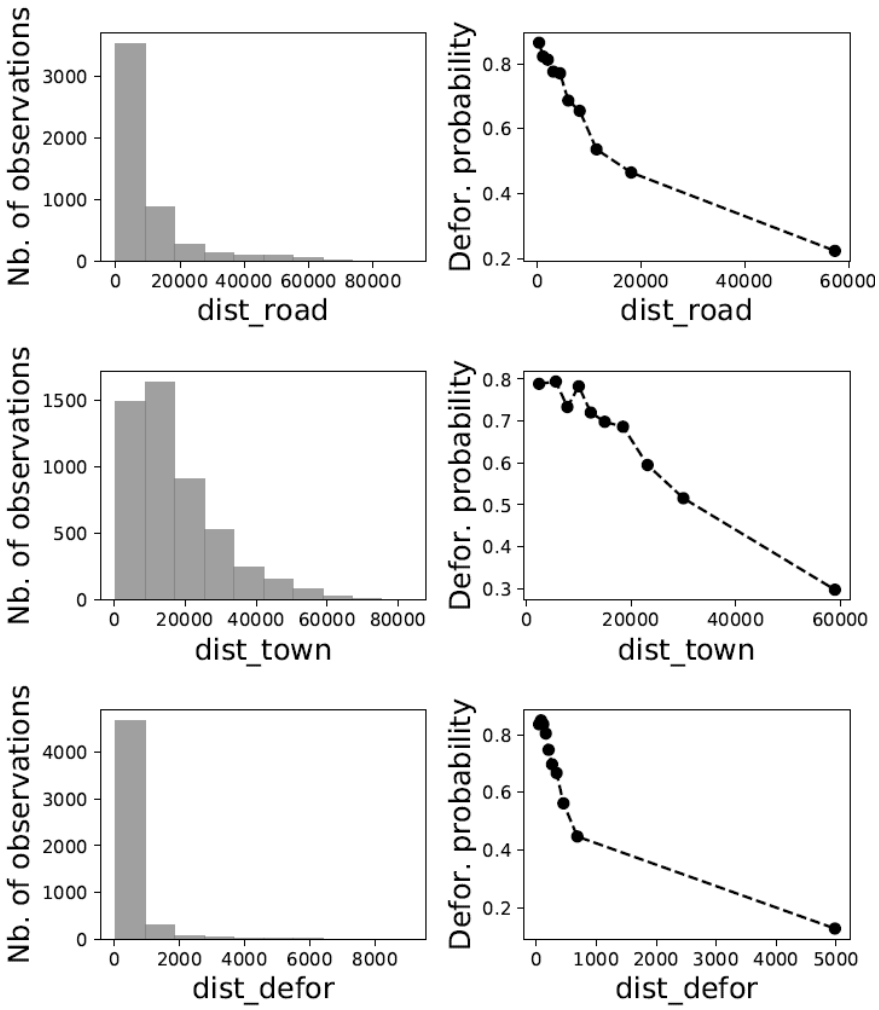

Figure 56: Relationships between the sampled environmental and urban variables and the probability of deforestation (part I). Distances are expressed in meters.

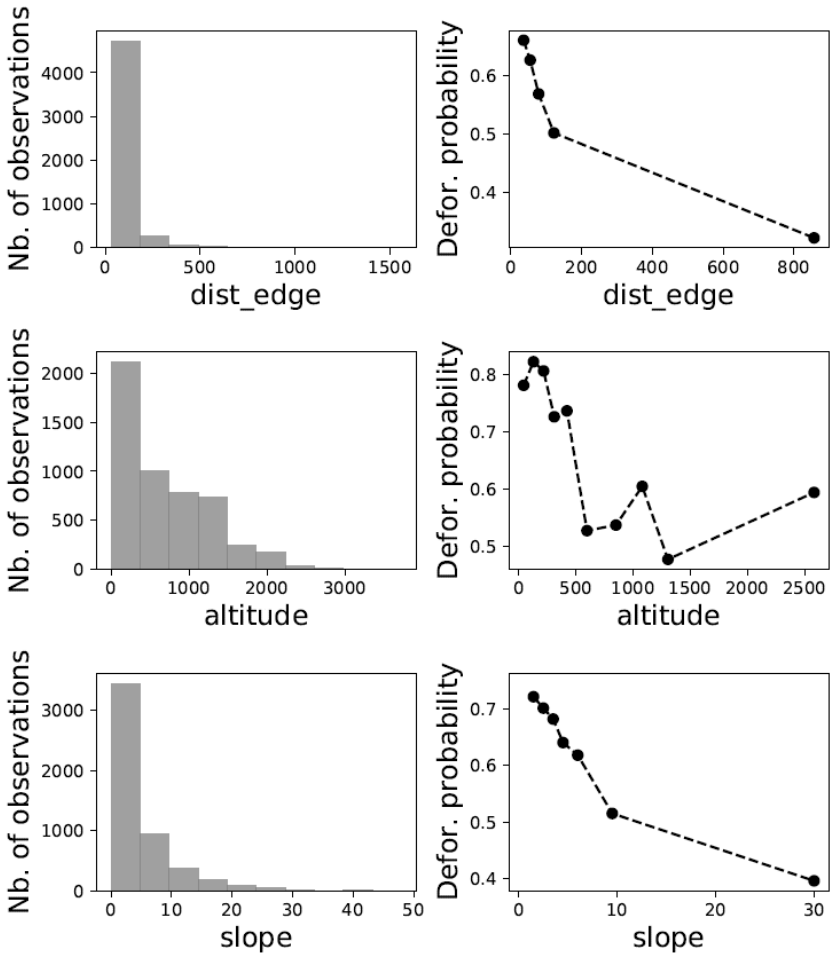

Figure 57: Relationships between the sampled environmental and urban variables and the probability of deforestation (part II). Distances and altitude are expressed in meters, slope is expressed in degrees.

It was necessary to create a model that predicts the probability of deforestation given a set of environmental variables. First, a binomial logistic regression model was fitted. Then, a summary of the model showing the parameter estimates can be reviewed. The 95% credible intervals obtained from the distribution of each parameter, except distance to nearest town *(dist_town*), do not include zero, indicating that parameters are significantly different from zero. Looking at the parameter estimates, one can see that deforestation probability decreases with altitude, slope, distance to past deforestation, forest edge, roads and towns. Parameter values are then coherent regarding the deforestation process and easy to interpret.

Table 25: Confidence Intervals of the binomial logistic regression model

	Mean	Std	CI_low	CI_high
Altitude	-0.0876	0.0418	-0.186	-0.0141
Slope	-0.525	0.0398	-0.614	-0.443
Distance to deforestation	-1.6	0.0758	-1.75	-1.44
Distance to edge	-0.532	0.0538	-0.635	-0.433
Distance to road	-0.585	0.0529	-0.686	-0.478
Distance to settlement	-0.204	0.0514	-0.305	-0.103

The spatial probability of deforestation for the entire country can then be mapped.

By applying the binomial logistic regression model, a probability map representing the probability of future deforestation was modelled for the entire country (Figure 58).

Interesting spatial patterns of deforestation at the national scale for Tanzania are obtained. In particular, five deforestation hotspots are visible in the map:

1) The highest risk is found close to the Burundi-Tanzania border in north-west Tanzania (Moyowosi and Kigosi game reserves);
2) Forests close to Lake Tanganyka (Mahale mountain national park);
3) Forests close to Arusha (Kilimanjaro and others national parks);
4) Zanzibar and the forest on the coast are at risk but considerably lower than for other zones;
5) Uluguru Mountains in eastern Tanzania in the outskirts of Morogoro
6) Eastern Arc Mountains (Udzungwa Mountains) in the outskirts of Dodoma

Moreover, the effect of climatic and altitudinal gradients on the deforestation patterns is clearly visible. In particular, it is possible to observe a drop of the deforestation risk with a circular shape close to the Kilimanjaro.

The spatial probability of deforestation is paramount to predict the future forest cover.

Given the spatial probability of deforestation and the number of hectares that should be deforested, we can predict the future forest cover. There are many studies on future deforestation (Vieilledent and Achard, In preparation). For each country, there are different future projection of deforestation considering various historical periods and deforestation scenarios (demographic growth, economic development, etc.). The mechanism is quite simple: the number of hectares provided in the studies of future deforestation could be converted into number of pixels to be deforested. Pixels with the highest probability of deforestation are

deforested first. Having this information, it will be possible to plot the predicted future forest cover (e.g. in 2050) and carbon emissions associated to future deforestation can be computed thereof.

It is also important to model deforestation patterns, predict the availability of wood supply, and develop urbanization and energy planning accordingly.

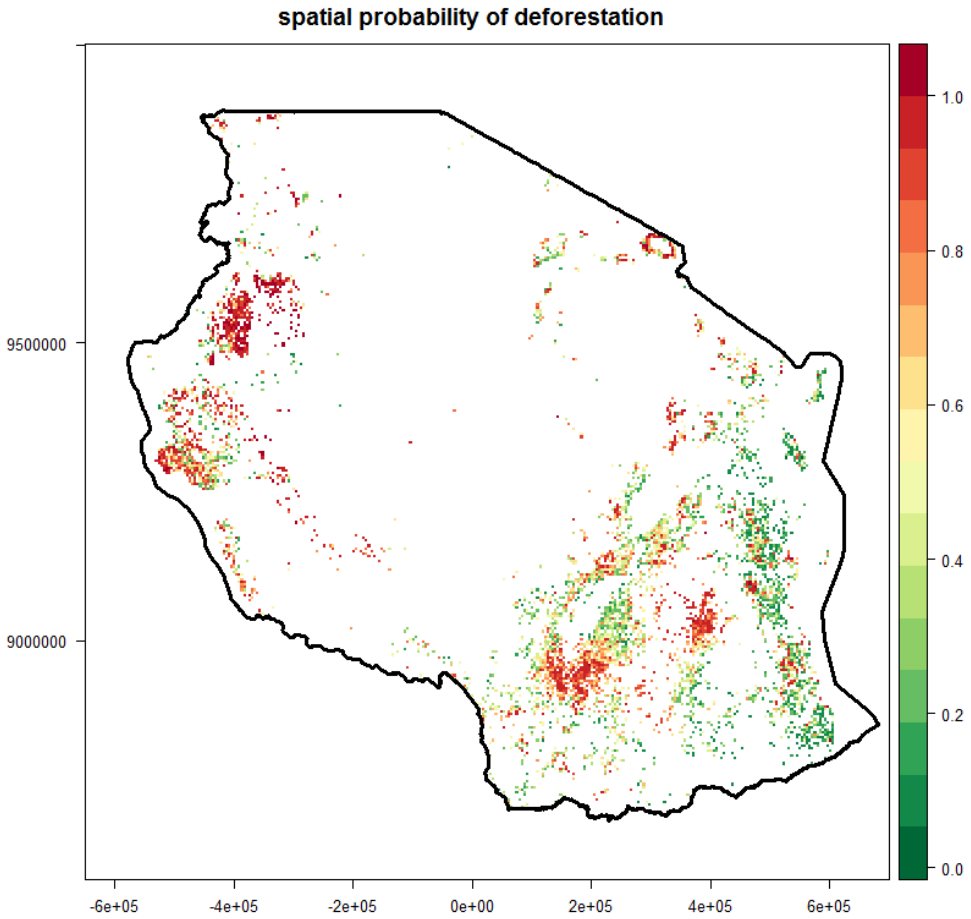

Figure 58: Spatial probability of deforestation across Tanzania.

8.6 Conclusions

Public planning can benefit from the combined use of data sets such as the Global Human Settlement Layers and the Global Forest Change. The GHSL data demonstrate that 'real' population density is far higher than public statistics, which are based on the administrative area, not on the actual extent of settlements. Is the urbanization extent which influences the spatial distribution of deforestation according to certain dynamics depending on the type or purpose of the wood collected. Other factors, related to urbanization but also environmental, affects the spatial distribution of forest loss, such as distance to roads, forest edges, deforestation areas, slope and altitude. Putting into evidence and analysing these relations is essential for planning a full range of public services, from transport to health, water and waste control. For economies of scale centres of demand for fuel and construction wood such as Dar es Salaam demand local sources of supply, and in the absence of *bona fide* sources, result in the illegal exploitation of protected areas.

9. General conclusions

The overall driving motive in the thesis has been to see how remotely sensed data can effectively contribute to the Tanzania forest service's needs and requirements for mapping and monitoring forest changes, notably forest degradation. It was shown in Chapter 1 Table 2 that forest area and area change estimates have wide variance. This means that emissions estimates for the REDD process remain unreliable.

This situation arises from i) lack of direct guidelines, ii) adequate control/monitoring, iii) political pressure to arrive at a positive funding situation, iv) complicity from consultants, and v) lack of 'best practice'.

The thesis has demonstrated the number of choices facing a forest service, in terms of forest definitions and methods are quite high, and all may have an impact on the feasibility of implementation and on results.

The difficulty has been shown of linking remote sensing data to the forest parameter collected by the TFS in their national survey, with recommendations as how to improve future remote sensing – field survey data collection.

Field data collected during the thesis was used in conjunction with VHR satellite imagery to produce local maps of forest parameters (basal area, tree height and above ground biomass). The use of image segmentation, and texture indices, was found to be useful in this.

The field data collected in targeted intact forests was compared to the data collect for the full coastal zone and showed that in general, this ecozone is heavily degraded.

While regional and global maps of historical and recent forest changes are of interest, for Tanzania they demonstrate a bias, and cannot be employed as a source of activity data for UNFCC reporting.

Within the thesis, the use of 5 m resolution RapidEye data was used, in conjunction with the national field survey to demonstrate the relative efficiency of using remotely sensed data to reduce the (costly) field component. In this exercise the shadow index proved to be more important in calibrating the above ground estimates, texture measures taking second place. A similar attempt with lower resolution Sentine-2 and Landsat 8 data was not successful.

The arrival of Sentine-2 data provides the opportunity to analyse medium-high resolution data (<20m) in time series, only possible beforehand with low resolution (MODIS, VGT) data.
The analysis over Mtera with data from the SPOT 4 Take 5 (20m) experiment shows that the use of vegetation indices can provide a discrimination between closed and degraded (or open) forests, as the latter dry out earlier (shallow roots) and have a grass or shrub layer that dries out. The area of forest cover correlates directly with the NDVI. Tree cover needs to fall by 40% within an object to be sure of detection.

When reviewing the overall dynamic the last part of the work shows the main deforestation driver is the supply of charcoal for an increasing population. As urban population centres are growing faster than the rural communities, the forests close to these centres have been depleted, so the deforestation frontier moves further away. As a result forest reserves have become an illegal source of fuel and construction wood. From this initial analysis, a probability map of future deforestation can be modelled.

9.1 Future research

Additional research is needed to enhance deforestation and forest degradation monitoring at country level.

Future work includes the following:

Improving the wall to wall by exploiting muti-sensor decametric resolution (Sentine-2/MSI and Landsat 8/OLI) free-of-charge data

Usage of Sentine-2A and 2B to obtain high frequent time series of satellite products (reflectance, texture, vegetation indices) to more reliably detecting/outlining changes in biomass over time and without typical gaps of single sensor time series.

10. References

Achard, F., Beuchle, R., Bodart, C., Brink, A., Carboni, S., Eva, H.D., Mayaux, P., Raši, R., Simonetti, D., Stibig, H.J., 2009. Monitoring forest cover at global scale: the JRC approach, in: 33rd International Symposium on Remote Sensing of Environment. Stresa, Italy, pp. 1–4.

Achard, F., Beuchle, R., Mayaux, P., Jurgen, H., 2014. Determination of tropical deforestation rates and related carbon losses from 1990 to 2010. Global Change Biology 20, 2540–2554. doi:10.1111/gcb.12605

Ahrends, A., Burgess, N.D., Milledge, S.A.H., Bulling, M.T., Fisher, B., Smart, J.C.R., Clarke, G.P., Mhoro, B.E., Lewis, S.L., 2010. Predictable waves of sequential forest degradation and biodiversity loss spreading from an African city. Proc. Natl. Acad. Sci. 107, 14556–14561. doi:10.1073/pnas.0914471107

Asner, G.P., 2005. Selective Logging in the Brazilian Amazon. Science 310, 480–482. doi:10.1126/science.1118051

Baatz, M., Schäpe, A., others, 2000. Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation. Angew. Geogr. Informationsverarbeitung XII 58, 12–23.

Baccini, A., Goetz, S.J., Walker, W.S., Laporte, N.T., Sun, M., Sulla-Menashe, D., Hackler, J., Beck, P.S.A., Dubayah, R., Friedl, M.A., Samanta, S., Houghton, R.A., 2012. Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps. Nat. Clim. Change 2, 182–185. doi:10.1038/nclimate1354

Beuchle, R., Eva, H.D., Stibig, H.-J., Bodart, C., Brink, A., Mayaux, P., Johansson, D., Achard, F., Belward, A., 2011. A satellite data set for tropical forest area change assessment. Int. J. Remote Sens. 32, 7009–7031. doi:10.1080/01431161.2011.611186

Bey, A., Sánchez-Paus Díaz, A., Maniatis, D., Marchi, G., Mollicone, D., Ricci, S., Bastin, J.-F., Moore, R., Federici, S., Rezende, M., Patriarca, C., Turia, R., Gamoga, G., Abe, H., Kaidong, E., Miceli, G., 2016. Collect Earth: Land Use and Land Cover Assessment through Augmented Visual Interpretation. Remote Sens. 8, 807. doi:10.3390/rs8100807

Bodart, C., Eva, H., Beuchle, R., Raši, R., Simonetti, D., Stibig, H.-J., Brink, A., Lindquist, E., Achard, F., 2011. Pre-processing of a sample of multi-scene and multi-date Landsat imagery used to monitor forest cover changes over the tropics. ISPRS J. Photogramm. Remote Sens. 66, 555–563. doi:10.1016/j.isprsjprs.2011.03.003

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Brown, S., 1997. Estimating Biomass and Biomass Change of Tropical Forests: A Primer. Food & Agriculture Org.

Burgess, N.D., Mbwana, S.B., 2000. Forestry, in: Coastal Forests of East Africa. IUCN, pp. 251–262.

Congalton, R.G., Green, K., 2009. Assessing the accuracy of remotely sensed data: principles and practices, 2nd ed. ed. CRC Press/Taylor & Francis, Boca Raton.

Daughtry, C.S.., Walthall, C.., Kim, M.S., Brown de Colstoun, E., McMurtrey, J.E., 2000. Estimating Corn Leaf Chlorophyll Concentration from Leaf and Canopy Reflectance. Remote Sens. Environ. 74, 229–239.

DeFries, R.S., Houghton, R.A., Hansen, M.C., Field, C.B., Skole, D., Townshend, J., 2002. Carbon emissions from tropical deforestation and regrowth based on satellite observations for the 1980s and 1990s. Proc. Natl. Acad. Sci. 99, 14256–14261.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., Bargellini, P., 2012. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. Remote Sens. Environ. 120, 25–36. doi:10.1016/j.rse.2011.11.026

Eckert, S., 2012. Improved Forest Biomass and Carbon Estimations Using Texture Measures from WorldView-2 Satellite Data. Remote Sens. 4, 810–829. doi:10.3390/rs4040810

Eitel, J.U.H., Long, D.S., Gessler, P.E., Smith, A.M.S., 2007. Using in-situ measurements to evaluate the new RapidEye^{TM} satellite series for prediction of wheat nitrogen status. Int. J. Remote Sens. 28, 4183–4190.

ESA, 2010. GMES Sentinel-2 Mission Requirements Document.

European Commission, 2013. Commission Delegated Regulations (EU) No 1159/2013 of 12 July 2013. Off. J. Eur. Union.

Eva, H.D., Benda, F., Carboni, S., Hojas Gascón, L., Simonetti, D., Velasco, M., 2016. The ReCaREDD project - Monitoring Forest Degradation In East Africa - Status and Results (No. EUR 28114 EN), JRC Technical Reports. European Commission.

FAO, 2016. Global Forest Resources Assessment. How are the world's forests changing? Food & Agriculture Org.

FAO, 2015a. Global forest resources assessment 2015. Desk reference.

FAO, 2015b. United Republic of Tanzania - Global Forest Resources Assessment 2015 – Country Report (No. AZ366), Global Forest Resources Assessment Country Reports. Rome.

FAO, 2007. Definitional issues related to reducing emissions from deforestation in developing countries, Eds: Schoene D., Killmann W., von Lüpke H., LoycheWilkie M. ed, Forests and Climate Change Working Paper 5. Rome, Italy.

FAO, JRC, SDSU, UCL, 2009. The 2010 Global Forest Resources Assessment Remote Sensing Survey: an outline of the objectives, data, methods and approach. Forest Resources Assessment Working Paper 155.

Frampton, W.J., Dash, J., Watmough, G., Milton, E.J., 2013. Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. ISPRS J. Photogramm. Remote Sens. 82, 83–92. doi:10.1016/j.isprsjprs.2013.04.007

Freire, S., Pesaresi, M., 2016. GHS population grid, derived from GPW4, multitemporal (1975, 1990, 2000, 2015) - CKAN [WWW Document]. URL http://data.jrc.ec.europa.eu/dataset/jrc-ghsl-ghs_pop_gpw4_globe_r2015a (accessed 12.2.16).

GOFC-GOLD, Achard, F., Boschetti, L., Brown, S., Brady, M., DeFries, R., Grassi, G., Herold, M., Mollicone, D., Pandey, D., Mora, B., others, 2014. A sourcebook of methods and procedures for monitoring and reporting anthropogenic greenhouse gas emissions and removals associated with deforestation, gains and losses of carbon stocks in forests remaining forests, and forestation. GOFC-GOLD.

Gonzalez, P., Asner, G.P., Battles, J.J., Lefsky, M.A., Waring, K.M., Palace, M., 2010. Forest carbon densities and uncertainties from Lidar, QuickBird, and field measurements in California. Remote Sens. Environ. 114, 1561–1575. doi:10.1016/j.rse.2010.02.011

Government of Brazil, 2014. Brazil's submission of a Forest Reference Emission Level (FREL) for reducing emissions from deforestation in the Amazonia biome for REDD+ results-based payments under the UNFCCC.

Hagolle, O., Huc, M., Dedieu, G., Sylvander, S., Houpert, L., Leroy, M., Clesse, D., Daniaud, F., Arino, O., Koetz, B., others, 2013. SPOT4 (Take5): Time series over 45 sites to prepare Sentinel-2 applications and methods, in: Proceedings of the ESA's Living Planet Symposium, Edinburgh, UK.

Hansen, M., Potapov, P., Margono, B., Stehman, S., Turubanova, S., Tyukavina, A., 2014. Response to Comment on "High-resolution global maps of 21st-century forest cover change." Science 344, 981–981. doi:10.1126/science.1248817

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O.,

Townshend, J.R.G., 2013a. High-Resolution Global Maps of 21st-Century Forest Cover Change. Science 342, 850–853. doi:10.1126/science.1244693
Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., Townshend, J.R.G., 2013b. High-Resolution Global Maps of 21st-Century Forest Cover Change. Science 342, 850–853. doi:10.1126/science.1244693
Haralick, R.M., 1979. Statistical and structural approaches to texture. Proc. IEEE 67, 786–804.
Haralick, R.M., Shanmugam, K., Dinstein, I. 'hak, 1973a. Textural features for image classification. IEEE Transactions on Systems, Man and Cybernetics SMC-3, 610–621.
Haralick, R.M., Shanmugam, K., Dinstein, I. 'hak, 1973b. Textural features for image classification. IEEE Transactions on Systems, Man and Cybernetics SMC-3, 610–621.
Henry, M., Picard, N., Trotta, C., Manlay, R.J., Valentini, R., Bernoux, M., Saint-André, L., 2011. Estimating tree biomass of sub-Saharan African forests: a review of available allometric equations. Silva Fennica 45, 477–569.
Herold, M., Verchot, L., Maniatis, D., Bauch, S., 2012. A step-wise framework for setting REDD+ forest reference emission levels and forest reference levels.
Hojas Gascón, L., Cerutti, P.O., Eva, H., Nasi, R., Martius, C., 2015. Monitoring deforestation and forest degradation in the context of REDD+: Lessons from Tanzania. CIFOR InfoBrief 124.
Hojas Gascón, L., Eva, H., 2014a. Field guide for forest mapping with high resolution satellite data. Publications Office, Luxembourg.
Hojas Gascón, L., Eva, H., 2014b. Field guide for forest mapping with high resolution satellite data. Publications Office of the European Union, Luxembourg.
Hojas Gascón, L., Eva, H., European Commission, Joint Research Centre, Institute for Environment and Sustainability, 2014. Field guide for forest mapping with high resolution satellite data. Publications Office, Luxembourg.
Hojas Gascón, L., Eva, H., Laporte, N., Simonetti, D., Fritz, S., 2012. The Application of Medium-Resolution MERIS Satellite Data for Continental Land-Cover Mapping over South America: Results and Caveats, in: Remote Sensing of Land Use and Land Cover: Principles and Applications. CRC Press/Taylor & Francis.
Houghton, R.A., Hall, F., Goetz, S.J., 2009. Importance of biomass in the global carbon cycle. J. Geophys. Res. Biogeosciences 114, G00E03. doi:10.1029/2009JG000935
Huete, A.., 1988. A soil-adjusted vegetation index (SAVI). Remote Sens. Environ. 25, 295–309. doi:10.1016/0034-4257(88)90106-X
INPE, 2015. Projeto Prodes - Monitoramento Da Floresta Amazônica Brasileira Por Satélite [WWW Document]. URL http://www.obt.inpe.br/prodes/index.php (accessed 3.30.15).
IPCC, 2013. 2013 Revised Supplementary Methods and Good Practice Guidance Arising from the Kyoto Protocol.
IPCC, 2003. Definitions and methodological options to inventory emissions from direct human-induced degradation forest and vegetation of other vegetation types. Ninth session of the Conference of the Parties (COP 9). Milan, Italy.
IPCC, 1996. Revised 1996 IPCC Guidelines for National Greenhouse Gas Inventories.
Keenan, R.J., Reams, G.A., Achard, F., de Freitas, J.V., Grainger, A., Lindquist, E., 2015. Dynamics of global forest area: Results from the FAO Global Forest Resources Assessment 2015. Forest Ecology and Management 352, 9–20.
Kissinger, G., Herold, M., 2012. Drivers of deforestation and forest degradation. Synth. Rep. REDD Policymakers.

Kombe, W.J., 2005. Land use dynamics in peri-urban areas and their implications on the urban growth and form: the case of Dar es Salaam, Tanzania. Habitat Int. 29, 113–135. doi:10.1016/S0197-3975(03)00076-6
Konstanski, H., 2012. Apparent Cloud Shift in RapidEye Image Data (No. RE-ENG-12-212). RapidEye.
Landis, J.R., Koch, G.G., 1977. The Measurement of Observer Agreement for Categorical Data. Biometrics 33, 159. doi:10.2307/2529310
Lu, D., 2005. Aboveground biomass estimation using Landsat TM data in the Brazilian Amazon. Int. J. Remote Sens. 26, 2509–2525. doi:10.1080/01431160500142145
Ma, L., Li, M., Blaschke, T., Ma, X., Tiede, D., Cheng, L., Chen, Z., Chen, D., 2016. Object-Based Change Detection in Urban Areas: The Effects of Segmentation Strategy, Scale, and Feature Space on Unsupervised Methods. Remote Sens. 8, 761. doi:10.3390/rs8090761
Magnussen, S., Reed, D., 2004. Knowledge reference for national forest assessments. FAO, Rome, Italy.
Makundi, W.R., 2001. Potential and Cost of Carbon Sequestration in the Tanzanian Forest Sector. Mitig. Adapt. Strateg. Glob. Change 6, 335–353. doi:10.1023/A:1013359415718
Malimbwi, R.E., Zahabu, E., Monela, G.C., Misana, S., Jambiya, G.C., Mchome, B., 2005. Charcoal potential of miombo woodlands at Kitulangalo, Tanzania. J. Trop. For. Sci. 197–210.
Mayaux, P., Holmgren, P., Achard, F., Eva, H., Stibig, H.-J., Branthomme, A., 2005. Tropical forest cover change in the 1990s and options for future monitoring. Philos. Trans. R. Soc. B Biol. Sci. 360, 373–384. doi:10.1098/rstb.2004.1590
Miettinen, J., Stibig, H.-J., Achard, F., 2014. Remote sensing of forest degradation in Southeast Asia—Aiming for a regional view through 5–30 m satellite data. Glob. Ecol. Conserv. 2, 24–36. doi:10.1016/j.gecco.2014.07.007
Ministry of Natural Resources & Tourism; Tanzania Forest Services Agency, 2015. NAFORMA National Forest Resources Monitoring and Assessment of Tanzania - Main results.
Mitchell, A.L., Rosenqvist, A., Mora, B., 2017. Current remote sensing approaches to monitoring forest degradation in support of countries measurement, reporting and verification (MRV) systems for REDD+. Carbon Balance Manag. 12. doi:10.1186/s13021-017-0078-9
MNRT, 2015. NAFORMA Main Results. Ministry of Natural Resources & Tourism, United Republic of Tanzania, Dar es Salaam, Tanzania.
MNRT, 2010. Tanzania Readiness Preparation Proposal to the Forest Carbon Partnership Facility.
Mwampamba, T.H., 2007. Has the woodfuel crisis returned? Urban charcoal consumption in Tanzania and its implications to present and future forest availability. Energy Policy 35, 4221–4234. doi:10.1016/j.enpol.2007.02.010
NAFORMA, 2010a. Field Manual, Biophysical survey. Ministry of Natural Resources and Tourism, Forestry and Beekeeping Division, Dar es Salaam, Tanzania.
NAFORMA, 2010b. Field Manual,Socioeconomic survey. Ministry of Natural Resources and Tourism, Forestry and Beekeeping Division, Dar es Salaam, Tanzania.
OECD, IEA, 2007. Financing mechanisms to reduce emissions from deforestation: issues in design and implementation.
Olander, L.P., Gibbs, H.K., Steininger, M., Swenson, J.J., Murray, B.C., 2008. Reference scenarios for deforestation and forest degradation in support of REDD: a review of data and methods. Environ. Res. Lett. 3, 025011. doi:10.1088/1748-9326/3/2/025011
Olofsson, P., Foody, G.M., Herold, M., Stehman, S.V., Woodcock, C.E., Wulder, M.A., 2014. Good practices for estimating area and assessing accuracy of land change. Remote Sens. Environ. 148, 42–57. doi:10.1016/j.rse.2014.02.015
Ortmann, A., Pekkarinen, A., Lindquist, E., Mponzi, B., Simon, O., Crete, P., 2015. Mapping historical changes in land cover using Landsat data and combination of automatic classification and visual

interpretation, in: National Forest Resources Monitoring and Assessment of Tanzania. Main Results. pp. 98–104.
Potapov, P., Yaroshenko, A., Turubanova, S., Dubinin, M., Laestadius, L., Thies, C., Aksenov, D., Egorov, A., Yesipova, Y., Glushkov, I., Karpachevskiy, M., Kostikova, A., Manisha, A., Tsybikova, E., Zhuravleva, I., 2008. Mapping the World's Intact Forest Landscapes by Remote Sensing. Ecol. Soc. 13. doi:10.5751/ES-02670-130251
Raši, R., Bodart, C., Stibig, H.-J., Eva, H., Beuchle, R., Carboni, S., Simonetti, D., Achard, F., 2011. An automated approach for segmenting and classifying a large sample of multi-date Landsat imagery for pan-tropical forest monitoring. Remote Sens. Environ. 115, 3659–3669. doi:10.1016/j.rse.2011.09.004
Rodriguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-Olmo, M., Rigol-Sanchez, J.P., 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS J. Photogramm. Remote Sens. 67, 93–104. doi:10.1016/j.isprsjprs.2011.11.002
Saatchi, S.S., Harris, N.L., Brown, S., Lefsky, M., Mitchard, E.T., Salas, W., Zutta, B.R., Buermann, W., Lewis, S.L., Hagen, S., others, 2011. Benchmark map of forest carbon stocks in tropical regions across three continents. Proc. Natl. Acad. Sci. 108, 9899–9904.
Santantonio, D., Hermann, R.K., Overton, W.S., 1977. Root biomass studies in forest ecosystems 17, 1–31.
Sarker, L.R., Nichol, J.E., 2011. Improved forest biomass estimates using ALOS AVNIR-2 texture indices. Remote Sens. Environ. 115, 968–977. doi:10.1016/j.rse.2010.11.010
Solano, R., Didan, K., Jacobson, A., Huete, A., 2010. MODIS vegetation index user's guide (MOD13 series). Veg. Index Phenol. Lab Univ. Ariz. 1–38.
Souza, C., 2003. Mapping forest degradation in the Eastern Amazon from SPOT 4 through spectral mixture models. Remote Sens. Environ. 87, 494–506. doi:10.1016/j.rse.2002.08.002
Szantoi, Z., Simonetti, D., 2013. Fast and robust topographic correction method for medium resolution satellite imagery using a stratified approach. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 6, 1921–1933.
The United Republic of Tanzania, 2016. Tanzania's forest reference emission level submission to the UNFCCC.
Timothy, D., Onisimo, M., Riyad, I., 2016. Quantifying aboveground biomass in African environments: A review of the trade-offs between sensor estimation accuracy and costs. Trop. Ecol. 57, 393–405.
Tomppo, E., Katila, M., Peräsaari, J., Malimbwi, R., Chamuya, N., Otieno, J., Dalsgaard, S., Leppänen, M., 2010. A Report to the Food and Agriculture Organization of the United Nations (FAO) in support of Sampling Study for National Forestry Resources Monitoring and Assessment (NAFORMA) in Tanzania.
Tomppo, E., Malimbwi, R., Katila, M., Mäkisara, K., Henttonen, H.M., Chamuya, N., Zahabu, E., Otieno, J., 2014. A sampling design for a large area forest inventory: case Tanzania. Can. J. For. Res. 44, 931–948. doi:10.1139/cjfr-2013-0490
Tropek, R., Sedláček, O., Beck, J., Keil, P., Musilová, Z., Šímová, I., Storch, D., 2014. Comment on "High-resolution global maps of 21st-century forest cover change." science 344, 981–981.
Tyc, G., Tulip, J., Schulten, D., Krischke, M., Oxfort, M., 2005. The RapidEye mission design. Acta Astronaut. 56, 213–219. doi:10.1016/j.actaastro.2004.09.029
Tyukavina, A., Stehman, S.V., Potapov, P.V., Turubanova, S.A., Baccini, A., Goetz, S.J., Laporte, N.T., Houghton, R.A., Hansen, M.C., 2013. National-scale estimation of gross forest aboveground carbon loss: a case study of the Democratic Republic of the Congo. Environ. Res. Lett. 8, 044039. doi:10.1088/1748-9326/8/4/044039
UNEP-WCMC, 2015. World Database on Protected Areas User Manual 1.0. Cambridge, UK.

UNFCCCC, 2014. Report of the Conference of the Parties on its nineteenth session, held in Warsaw from 11 to 23 November 2013. Part one: Proceedings.
United Republic of Tanzania, 2014. City Profile for Dar Es Salaam, Dar es Salaam City Council. Dar es Salaam.
UN-REDD, 2013a. Roadmap for Development of a Reference Emission Level / Reference Level for the United Republic of Tanzania. Discussion draft – v1.
UN-REDD, 2013b. Roadmap for Development of a Reference Emission Level / Reference Level for the United Republic of Tanzania. Discussion draft – v1.
UN-REDD, 2012. Draft Action Plan for implementation of National Strategy for REDD+.
van der Werf, G.R., Morton, D.C., DeFries, R.S., Olivier, J.G.J., Kasibhatla, P.S., Jackson, R.B., Collatz, G.J., Randerson, J.T., 2009. CO2 emissions from forest loss. Nat. Geosci. 2, 737–738. doi:10.1038/ngeo671
Vieilledent, G., Achard, F., In preparation. Including spatial-autocorrelation in deforestation model to obtain realistic deforestation projections at national or continental scales.
Weichelt, H., Rosso, P., Marx, A., Kim, S., Douglass, K., Heynen, M., n.d. The RapidEye Red Edge Band.
White, F., 1983. The vegetation of Africa, a descriptive memoir to accompany the UNESCO/AETFAT/UNSO vegetation map of Africa (3 Plates, Northwestern Africa, Northeastern Africa, and Southern Africa, 1: 5,000,000). Unesco, Paris.
Zahabu, E., Skutsch, M., Malimbwi, R.E., Nordholt, N.G., others, 2012. The likely mechanism for implementing REDD policy in Tanzania.

11. Annex – Publications

The potential of high resolution (5 m) RapidEye optical data to improve above ground biomass estimates over Tanzania. L. Hojas-Gascón, G. Ceccherini , J.García-Haro, Valerio Avitabile and H. Eva. Under review for Remote Sensing journal.

Urbanization and forest degradation in East Africa - A case study around Dar es Salaam, Tanzania. L. Hojas-Gascón, H.D. Eva, D. Ehrlich, M. Pesaresi, F. Achard and J. Garcia, 2016. To be publish in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing Special Issue on IEEE 2016 International Geoscience and Remote Sensing Symposium (IGARSS 2016).

Monitoring deforestation and forest degradation in the context of REDD+. Lessons from Tanzania. Lorena Hojas-Gascon, Paolo Omar Cerutti, Hugh Eva, Robert Nasi and Christopher Martius, 2015. CIFOR (Center for International Forestry Research) Infobrief. No. 124, May 2015. doi:10.17528/cifor/005642, http://www.cifor.org/publications/pdf_files/infobrief/5642-infobrief.pdf.

Potential improvement for forest cover and forest degradation mapping with the forthcoming Sentinel-2 program. L. Hojas-Gascón, A. Belward, H. Eva, G. Ceccherini, O. Hagolle, J. Garcia, P. Cerutti, 2015. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XL-7/W3, 2015. 36th International Symposium on Remote Sensing of Environment, 11–15 May 2015, Berlin, Germany. doi:10.5194/isprsarchives-XL-7-W3-417-2015, http://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XL-7-W3/417/2015/isprsarchives-XL-7-W3-417-2015.html, http://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XL-7-W3/417/2015/isprsarchives-XL-7-W3-417-2015.pdf

Field guide for forest mapping with high resolution satellite data - Monitoring deforestation and forest degradation in the context of the UN-REDD+ programme; The Tanzania REDD+ initiative. L. Hojas-Gascon and H.D Eva, 2014. Technical Report by the Joint Research Centre of the European Commission (JRC 92600, EUR 26922 EN). Publications Office of the European Union, Luxembourg, 2014. ISBN 978-92-79-44012-0 (PDF), ISSN 1831-9424 (online), doi:10.2788/657954, http://publications.jrc.ec.europa.eu/repository/bitstream/JRC92600/lbna26922enn.pdf.

Developing methods for monitoring forest degradation in Tanzania using fine spatial resolution RapidEye data. L. Hojas-Gascon and H.D Eva, 2013. NAFORMA derived studies. Publisher: NAFORMA, Ministry of Natural Resources and Tourism of Tanzania, in collaboration with FAO, Ministry of Foreign Affairs of Finland and European Commission, http://naforma.mnrt.go.tz/naforma, http://naforma.mnrt.go.tz/uploads/Derived%20studies/Appendix%201_5%20JRC-%20Derived%20study%20Monitoring%20Forest%20degradation%20using%20RapidEye%20Data__2014_05_25.pdf.