Student thesis series INES nr 460

The road to food security: Evaluating and improving open-access road data in Tanzania with satellite imagery

Corey Nicholas Ragosnig

2018
Department of
Physical Geography and Ecosystem Science
Lund University
Sölvegatan 12
SE-223 62 Lund
Sweden

Seminar series nr 460

Corey Nicholas Ragosnig (2018).
The road to food security: Evaluating and improving open-access road data in Tanzania with satellite imagery
På väg mot en trygg livsmedelsförsörjning: En utvärdering och förbättring av öppen tillgång vägdatabaser i Tanzania med satellitbilder
Master degree thesis, 30 credits in *Geomatics*
Department of Physical Geography and Ecosystem Science, Lund University

Level: Master of Science (MSc)

Course duration: *January* 2018 until *June* 2018

Disclaimer

The road to food security: Evaluating and improving open-access road data in Tanzania with satellite imagery

---

Corey Nicholas Ragosnig

Master thesis, 30 credits, in *Geomatics*

Ola Hall
Lund University, Department of Human Geography

Exam committee:
Jonathan Seaquist and Abdulhakim Abdi
Lund University, Department of Physical Geography and Ecosystem Science

Abstract

Contemporary approaches to improving food security amongst the world's vulnerable populations have shifted from attempts to improve crop yield to improving accessibility to markets. Analyses show that there is a correlation between income, distance from cities, and the ratio of potential agricultural production to realised agricultural production.

Complete, up-to-date, and reliable road network data are essential to quantify accessibility. However, official road network data published by the relevant governing bodies is unavailable in many parts of the world. Thus, open-access datasets such as OpenStreetMap (OSM) and the Global Road Open Access Dataset (gROADS) have been used as alternatives due to their global coverage and easy-of-access. Nevertheless, the crowd-sourced nature of these datasets means that the spatial accuracy and extent of these datasets varies on both local and global scales.

This thesis project includes an evaluation of the existing open-access road network datasets and examines the possibility of extracting road network data from open-access Google Earth and Sentinel-2 satellite imagery where the existing data is lacking. Tanzania serves as the study area for this thesis project due to its comparability with other sub-Saharan African countries suffering from crop deficits and food insecurity despite good agricultural potential. The results show that the OSM data are in general quite accurate and provide a good coverage with some missing road segments, while the gROADS data are insufficient for analyses at local scales or those which require high spatial accuracy. The results also show that road networks featuring narrow unpaved roads can be successfully extracted from high resolution Google Earth imagery, though further refinements are necessary to reduce noise. The method applied to the Sentinel-2 imagery proved unsuccessful in this particular study.

Sammanfattning

Metoderna som används för att göra den globala livsmedelsförsörjningen tryggare har bytt fokus från själva produktionen av jordbruksgrödor till att öka människors tillgänglighet till marknader. Analyser visar att det finns ett samband mellan inkomst, avlägsenhet, och hur stor andel av god jordbruksmark faktiskt används för jordbruk.

För att beräkna tillgänglighet krävs fullständig, uppdaterad, och pålitlig vägdata; men sådan data saknas ofta i många länder. Trots detta kan öppen tillgång vägdatabaser, till exempel OpenStreetMap (OSM) och den Global Road Open Access Dataset (gROADS), används som alternativa datakällor eftersom de erbjuder en omfattande global täckning och kan laddas ner helt kostnadsfritt. Däremot har dess kvalité ifrågasatts eftersom vägdata samlas in i databaserna genom crowdsourcing, vilket påverkar rumslig noggrannhet och fullständighet på olika sätt i olika delar av världen beroende på vem som hade samlat in data och hur den hade samlats in.

Det här examensarbetet syftar till att utvärdera dessa befintliga öppen tillgång vägdatabaser samt försöka extrahera ett mer omfattande vägnät ur öppen tillgång Google Earth och Sentinel-2 satellitbilder där befintliga vägdatabaser har brister. Tanzania har valts ut som testområde därför att landet har en otrygg livsmedelsförsörjning trots att andelen jordbruksmark i landet räcker till att tillgodose befolkningens livsmedelsbehov. Resultaten visar att OSM-vägdatabasen har en god geografisk noggrannhet och fullständighet fastän några mindre vägar saknas, medan gROADS-vägdatabasen är olämplig för småskaliga analyser där hög geografisk noggrannhet krävs. De visar även att det är möjligt att extrahera ett vägnät som består av små grusvägar ur Google Earth:s satellitbilder med hög rumslig upplösning, dock metoden som har använts på de Sentinel-2 satellitbilderna misslyckades.

Table of Contents

1. Introduction

1.1 Background

Access to an adequate supply of nutritious food is vital for people to maintain a healthy and active life. However, food insecurity continues to pose a significant threat to many people around the world, in particular in the developing regions of sub-Saharan Africa and southern Asia (van Ittersum et al. 2016; United Nations 2018). According to the United Nations (2018), about 815 million people around the world, or approximately 11% of the world's population, are undernourished. In sub-Saharan Africa alone, the rate of undernourishment rises to almost 23% (United Nations 2018). Hunger and undernourishment pose significant humanitarian and economic problems for these regions; half of all deaths in children under age five and the stunted growth of a third of the world's children can be attributed to hunger and undernourishment (United Nations 2018). If the developing nations of sub-Saharan Africa are to provide an adequate standard of living for their inhabitants, then access to an adequate supply of nutritious food must be guaranteed to ensure that each person has the opportunity to lead active and healthy lives.

Historically, efforts to secure an adequate supply of nutritious food in developing nations have focused on increasing agricultural production and yields (Erikson et al. 2009). Such food security policies have had limited effects on food security, which is especially evident in areas that are currently facing food insecurity despite experiencing a crop production surplus (Dorosh et al. 2011; Erikson et al. 2009; Frelat et al. 2016). Frelat et al. (2016) suggest that improving access to markets could have a larger impact on food security than increasing agricultural production and crop yields. In a study of over 13 000 farming households across 17 countries in sub-Saharan Africa, income from selling crops and other off-farm sources of income contribute to 12% of a farming household's total food availability amongst farming households with insufficient access to food (Frelat et al. 2016). For farming households that do have sufficient access to food, off-farm sources of income account for 27% of their total food availability (Frelat et al. 2016). A majority of farming households also continue to sell part of their food crops as a source of household income, even when the total amount of food crops that they produce is insufficient to meet their own needs (Frelat et al. 2016). Considered together, this indicates that access to markets is vital for farming households to be able to sell their produce as a source of income in order to adequately contribute to the household's food availability.

Increasing access to markets as an alternate means to facilitate food security in sub-Saharan African nations rather than increasing crop production or improving agricultural yields is

further supported by Dorosh et al. (2011). Dorosh et al. (2011) show that agricultural production is strongly correlated with travel times to the nearest market, where *market* is defined as the nearest population centre with 100 000 or more inhabitants. Dorosh et al. (2011) suggest that production in such remote areas could be constrained by a limited local labour pool and relatively high transportation costs. With improved transportation links, farmers would have access to an expanded labour pool as well as possible reductions in the total time and cost of transportation that would make selling their produce at more distant markets more feasible; both of which are preconditions to increased levels of production.

Roads are the fundamental component of a transportation network that connects remote areas to large population centres and markets. Knowing how well the roads are able to connect these remote communities to a market is essential to identifying possible areas where road access could be improved so as to facilitate better agricultural outcomes. A complete, up-to-date, and reliable road network dataset is therefore required. Despite their importance in such analyses, existing spatial databases that describe the global road networks are often insufficient, being characterised by limited spatial extents, out-dated or missing attribute data, and geographic inaccuracies, among others (Haklay 2010; Ibisch et al. 2016). Nevertheless, Ibisch et al. (2016) used the available global road datasets, such as OpenStreetMap (OSM) and the Global Roads Open Access Data Set (gROADS), to map the world's roadless areas, acknowledging that their analysis is likely to have overestimated the size of the world's roadless areas due to the deficiencies of the global datasets in comparison to those possibly available within local databases. Of particular interest to this paper is sub-Saharan Africa, where large swaths of land in most countries are shown to be roadless.

There is thus a clear discord between improving market accessibility for farmers and the availability of road network datasets in developing regions. There are gaps in the coverage and quality of the existing global road network datasets, especially in developing regions (Ibisch et al. 2016; Hughes 2017). This is problematic for researchers and policy makers; if accessibility cannot be modelled to a reasonable degree of accuracy due to concerns about the quality and completeness of the datasets, then it becomes extremely difficult to implement policies that target food security through improving access to markets. The most recent models of accessibility (for example: Dorosh et al. 2011; Linard et al. 2012; Weiss et al. 2018) have been built upon these aforementioned openly available global road network datasets due to a lack of easily accessible higher quality data. If the quality of existing road network datasets were to be improved, then it is likely that market accessibility could be modelled with a higher degree of accuracy.

1.2 Objective

The main aim of this thesis project is to contribute towards improving the road network datasets that form the basis for accessibility analyses by addressing the following:

1. To evaluate the existing open-access global road network datasets, OpenStreetMap (OSM) and Global Roads Open Access Dataset (gROADS), in terms of their completeness and their spatial accuracy.
2. To apply and evaluate two proposed methods to extract unpaved road network data from open-access satellite imagery provided by Google Earth and the European Space Agency (ESA). The resulting road network from the two methods is to be compared in terms of completeness and spatial accuracy relative to OSM and gROADS.
3. To outline some of the difficulties and considerations when using either the existing road network data or using satellite imagery. This includes: determining in which contexts OSM and gROADS are most suitable; and how characteristics of the satellite imagery and the road features impact the choice of road extraction methods.

2. Literature Review

2.1. Food Security and Accessibility in Sub-Saharan Africa

A survey of recent literature suggests that traditional crop yield-based approaches to improving food security amongst the world's vulnerable populations are inadequate (Dorosh et al. 2011; Erikson et al. 2009; Frelat et al. 2016). Case studies in Ethiopia (Stifel and Minen 2008), Madagascar (Chamberlin et al. 2007), and China and India (Fan and Hazel 2001) have demonstrated that investments in public services and infrastructure rather than technologies to improve crop yield have had a positive impact on agricultural production and food security. Building upon the results presented in the Ethiopia and Madagascar case studies, Dorosh et al. (2011) show that remoteness has a statistically significant correlation with decreased agricultural productivity throughout all of sub-Saharan Africa. Both population and agricultural production are concentrated near cities, with 41.4% of the population and 23.6% of the total agricultural production located within 2.5 hours of a city with a population of at least 100 000 people (Dorosh et al. 2011). Furthermore, as travel time to the nearest city increases to or exceeds four hours, the ratio of actual agricultural production to potential agricultural production decreases by approximately 90% (Dorosh et al. 2011).

Dorosh et al. (2011) also note that, in general, west African countries tend to have a crop surplus whereas eastern sub-Saharan African countries tend to have a crop deficit. Simultaneously, the eastern sub-Saharan African countries tend to have lower population densities, smaller markets, and poorer road connections than their west African counterparts (Dorosh et al., 2011). Given that the proportion of suitable agricultural area is not significantly different between the two regions, Dorosh et al. (2011) suggest that the lower agricultural production in eastern sub-Saharan African countries, and subsequently their crop deficits, is a function of the population's remoteness relative to their counterparts in western Africa.

The aforementioned studies rely on high quality population and transportation data to draw accurate conclusions; however, this is generally a problem throughout many of the world's developing regions due to poor resolution data, the lack of consistent data collection with reliable methodologies, and insufficient detail (Barrington-Leigh and Millard-Ball 2017; Dorosh et al. 2011; Ibisch et al. 2016; Linard et al. 2012). To better understand the spatial distribution of population in Africa, Linard et al. (2012) were able to combine available lower resolution population density datasets with higher resolution satellite imagery by matching the population density data to patterns of urban and rural settlements seen in the satellite imagery, resulting in a population density dataset with a 100m spatial resolution. The remoteness of

any given settlement could then be estimated by modelled the quickest travel time to the nearest city with a population greater than 50 000 (Linard et al. 2012). On a global scale, Weiss et al. (2018) were able to estimate the remoteness of populations at a 1km spatial resolution by estimating the shortest travel time to the nearest city with a population greater than 100 000. Travel time was modelled by taking into account the effects of land cover, slope and elevation, distance to the nearest road, and road quality on travel speed (Weiss et al. 2018). Low-income individuals, mostly in sub-Saharan Africa, are disproportionately affected by remoteness; half of these individuals reside in locations more than one hour away from the nearest city in comparison to just 10% of high-income individuals (Weiss et al. 2018).

2.2 Quality of Existing Road Network Datasets

The estimates of remoteness in Linard et al. (2012) and Weiss et al. (2018) are highly dependent upon the road network dataset used in the model, as it is the road network that forms the backbone of the world's land-based transportation network. Unfortunately, sub-Saharan Africa and other developing regions of the world are generally poorly represented in the available road network datasets; both those that are published by the relevant official agencies of government and crowd-sourced alternatives such as OSM and gROADS (Barrington-Leigh and Millard-Ball 2017; Ibisch et al. 2016). Ibisch et al. (2016)'s global map of roadless areas, created from OSM and gROADS road data, posits that Tanzania itself is home to roadless areas that exceed 25 000 square kilometres in size. This is thought to be a gross overestimation, however, as supplementary data sources such as the National Oceanic and Atmospheric Administration (NOAA) night-time lights dataset and the Global Forest Watch deforestation dataset allude to the presence of roads within these purported roadless areas (Hughes 2017).

Night-time lights reveal population centres that appear to be completely isolated from their neighbours according to the road networks used in Ibisch et al. (2016)'s analysis. Although this supposition could hold true in some cases, the rather large number of isolated settlements inside of the rather large roadless areas of developing nations together with the existence of more detailed local databases suggests that roads do exist in these areas (Ibisch et al. 2016; Hughes 2017). Similarly, linear patterns in the deforestation dataset could indicate the existence of roads (Hughes 2017). Several cases of suspected missing roads in these so-called roadless areas of the world are confirmed by Hughes (2017) in many of the world's developing regions such as Myanmar, south China, Borneo, and east Madagascar. It is

therefore also likely that there are roads in Tanzania that are missing from the existing global road network datasets.

Knowing that the global road datasets are incomplete means that it is important to know whether the data are accurate enough to be used in a given study area as input to an analysis. Some people question the reliability of the OSM and gROADS road networks as they are examples of crowd-sourced data; it is often assumed that such data are reliable only in areas where a large concentration of users exists to collect, publish, and verify the data in the database (Barrington-Leigh and Millard-Ball 2017; Haklay 2010). A recent article by Barrington-Leigh and Millard-Ball (2017) suggests that OSM data are generally accurate and complete on a global scale. The accuracy of the dataset was determined by visually comparing the mapped road network with satellite imagery. A stratified random sample based on population density was conducted to choose 45 sampling areas in each country (Barrington-Leigh and Millard-Ball 2017). In each sampling area, the number of road segments missing from the OSM database that are visible in the satellite imagery at a scale of 1 : 5 000 were counted (Barrington-Leigh and Millard-Ball 2017). On a global scale, they estimate that, as of 2016, the OSM road network is about 83% complete, rising to 95% complete in 42% of the world's countries. Sub-Saharan Africa and Asia are notable regions where OSM completeness is lacking (Barrington-Leigh and Millard-Ball 2017). They also suggest that missing road segments are most likely to be found in regions with medium-low population density because the database is most robust in urban areas as well as in remote areas along regional routes between settlements (Barrington-Leigh and Millard-Ball 2017).

2.3. Methods to Extract Road Networks from Satellite Imagery

With the advent of publicly-accessible satellite imagery with spatial resolutions fine enough to be able to distinguish roads, it has become possible to update road network datasets in regions where limited data exists or where surveying proves to be costly, time-consuming, and impractical. There are several defining characteristics of roads that can be used to distinguish and extract road-features, including: surface properties; geometry; topology; and context (Auclair-Fortier et al. 1999; Wang et al. 2016). Traditional remote sensing techniques focus on detecting roads from their surface properties, whereby they are identified based on their spectral signature. This is often insufficient, as roads often share a similar spectral signature with other features in the imagery (Auclair-Fortier et al. 1999; Wang et al. 2016). Geometric, topological, and contextual characteristics are therefore necessary to distinguish roads from features in the image with similar spectral properties (Auclair-Fortier et al. 1999; Wang et al. 2016). This means identifying linear features with relatively consistent direction and uniform

widths, that form a network with other similar linear features, and which serve a real-world purpose.

The methods to extract road networks from satellite imagery are numerous, mostly because of the large variation in imagery (Auclair-Fortier et al. 1999; Wang et al. 2016). The method used depends not only on the spatial resolution and the spectral bands that are measured as defined by the technical specifications by the satellite, but also on the physical characteristics of the roads themselves, weather and other atmospheric disturbances, variations in pixel response due to influence from surrounding features, and fictitious discontinuities in the network caused by terrain, shadows, or overlapping features in the imagery (Wang et al. 2016). Despite the variation, road network extraction methods can generally be divided into four main components: preprocessing of the image to remove excess noise and enhance relevant details; detecting edges that define the road features; eliminating road-like pixels that represent other features by tracing the road network from a given starting point; and regrouping of road segments into a continuous network (Auclair-Fortier et al. 1999). Examples of methods described in Wang et al. (2016)'s review of road extraction methods from remote sensing images include: supervised classification, neural networking, graph theory division, mathematical morphology, and active contours.

The vast majority of the methods described in Wang et al. (2016)'s review focus on the extraction of paved road networks. In this case, the extraction of a road network from the image is relatively straightforward: a set of distinguishing characteristics for the road network in terms of its spectral signature, shape, and topology are established, and several sets of algorithms are used to identify pixels that conform to the set of these characteristics. For example, Wang et al. (1992) use SPOT images to extract roads by finding lines where there is a maximum change in gradient. Baumgartner et al. (1999) propose a method to extract roads in both an urban and rural context by using a multi-scale and contextual approach. At a coarse scale, the overall pattern of the road network is distinguished from the rest of the image as a set of line edges. This scale also establishes the urban-rural contexts that define the types of patterns that should be searched for. At a fine scale, details such as the spectral signature, road markings, width, and surroundings inform the algorithm's decision as to whether a pixel represents a road or non-road depending on the established context.

2.4. Considerations for Extracting Unpaved Roads

Unlike paved road extraction, the extraction of unpaved dirt roads poses several unique challenges. The first and most difficult of these is the lack of spectral information that can

differentiate dirt roads from the surrounding land cover. Figure 1 below is an example of typical spectral responses that could be measured for various materials for various spectral wavelengths. Asphalt paved roads have a relatively low reflectivity at all wavelengths in the visible and infrared spectra, whereas soil and green vegetation have high reflectivity responses in the short wave infrared and near-infrared spectra respectively. This makes using spectral information feasible to extract paved road networks. Additionally, the appearance of dirt roads may vary along the length of the road. For example, the road surface may be made up of different dirt materials in different locations in the image, or the unevenness of the road surface caused by different levels of use may cause light to reflect at different angles, affecting the measured spectral response.

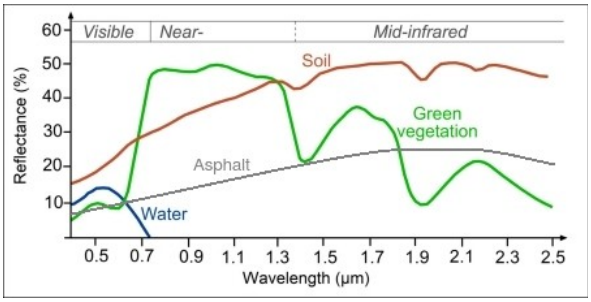


Figure 1. Spectral signatures for asphalt, green vegetation, soil, and water. Adapted from SEOS (no date) and Riebeek (2014).

There seems to be a general lack of related literature describing the extraction of unpaved roads, and in particular, the extraction of unpaved roads from low resolution satellite imagery. This is somewhat problematic for two reasons: firstly, unpaved roads constitute a sizeable portion of road networks in developing countries; and secondly, publicly accessible and free-to-download satellite imagery that offers a global coverage is limited to low-to-medium resolution satellites, such as the 10m resolution Sentinel-2 imagery, or to higher resolution imagery with more restrictive download capabilities or licensing, such as Google Earth (European Space Agency 2018; Google 2015). A database search for unpaved road extraction techniques reveals two papers: Gomes et al. (2004) and Ojo et al. (2016); both of which also mention a lack of research into unpaved road extraction techniques. The database search also reveals a masters thesis, where aerial imagery and object-oriented classification are used to classify an unpaved road network in rural Kenya with mixed results (Keskinen 2007).

The spatial resolution of the satellite imagery used differs between the two papers. For this reason, the problem of extracting the unpaved roads cannot be solved using the same approach. Gomes et al. (2004) make use of low-resolution Landsat TM imagery. Landsat TM

bands have a 30m spatial resolution, which is relatively low compared to the size of the road features being extracted (Gomes et al. 2004; United States Geological Survey 2018). This means that the spectral information of the surrounding land cover interferes with the spectral information that characterises unpaved roads in any given pixel that represents a road. The spectral signature of the road pixels in the image is therefore heterogeneous; varying with changes in the surrounding land cover type and material composition of the road. To eliminate the heterogeneity, Gomes et al. (2004) focus on identifying lines which correspond to a consistent relative difference between the potential road pixel and the pixels that represent the surrounding land cover.

In contrast, Ojo et al. (2016) make use of high resolution satellite imagery downloaded from Google Earth. With high resolution imagery, the problem becomes distinguishing regions that represent unpaved roads from other regions with similar spectral characteristics, such as agricultural fields. Unlike in the satellite imagery used by Gomes et al. (2004), Google Earth imagery allows for a rather accurate determination of the spectral signatures of the various materials that compose unpaved roads. First, pixels that represent unpaved roads are identified using their spectral signature. Pixels which are not bounded by two relatively parallel lines are then discarded, as unpaved roads are narrow, elongated, and have defined parallel edges.

3. Study Area, Data, and Methods

This thesis project is divided into three main components: an evaluation of the existing road network datasets; the development of a new road network dataset that improves upon the information provided in the existing datasets using algorithms to extract road network data from satellite imagery; and a comparison of the existing and new road network datasets in terms of how accurately they represent the actual road network in the study area. QGIS 2.18 is used to conduct the accuracy evaluation and Matlab 2018a is used to implement the road extraction algorithms. An illustration of the entire method can be seen in figure 2 below.

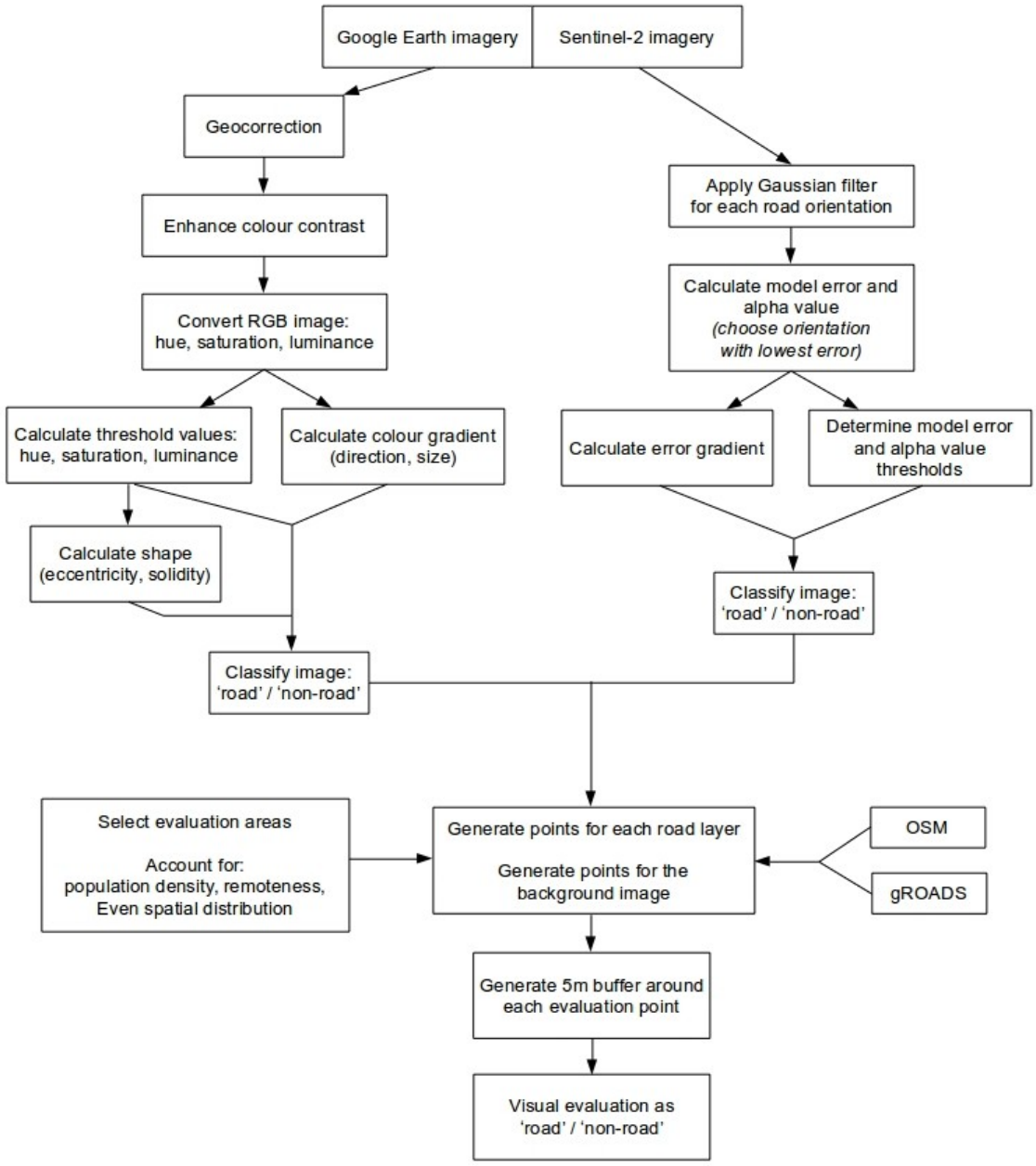


Figure 2. Flowchart illustrating the road extraction process and the road network evaluation.

3.1 Study Area

Tanzania is chosen as the study area for this thesis project because it representative in the context of improving food security amongst the world's vulnerable populations. Tanzania, like other eastern sub-Saharan African countries, is experiencing a crop deficit despite good agricultural potential (Dorosh et al. 2011). There are also active efforts to increase the country's agricultural productivity and the accessibility of its remote settlements through the implementation of an agricultural growth corridor in its planning policy (Smalley 2017; SAGCOT 2018). The OSM and gROADS data are evaluated on the national scale. Three smaller subregions approximately 2km by 2km in extent, shown in figure 3, are used to test the two road extraction methods from the satellite imagery provided by Google Earth and the ESA. These study areas were primarily selected due to the lack of road network data in either the OSM and gROADS data sets despite evidence from satellite imagery. These three study areas in particular were also selected to account for varying degrees of remoteness and urbanisation throughout Tanzania. Study areas A and B are located within the SAGCOT corridor whereas study area C is relatively remote in comparison to Tanzania's population centres. Study area A is more characteristic of a rural landscape whereas study areas B and C show signs of higher population density and degree of urbanisation.

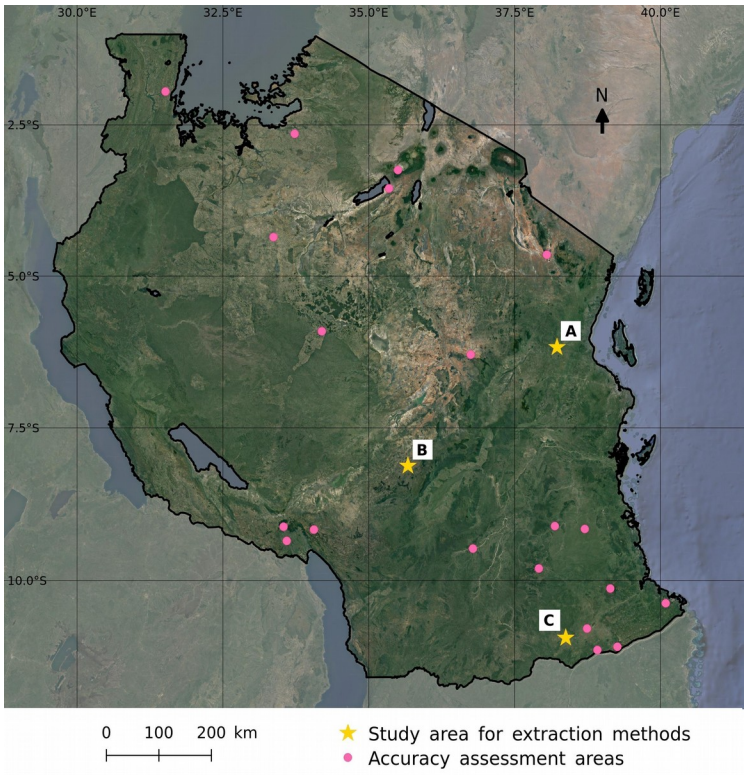


Figure 3. Overview map of Tanzania, showing the three study areas used to test the road extraction methods as well as the twenty other sample areas used to assess the accuracy of the gROADS and OSM data sets.

3.2 Data

Two different sets of satellite imagery are used in this thesis project to extract the updated road network: Google Earth imagery from the Airbus DigitalGlobe platform, and the ESA's Sentinel-2 satellite imagery. The main reason for using these two particular sets of satellite imagery are their high level of detail and ease of access compared to alternative satellite image products. Both of these image sets are openly accessible to the general public and may be downloaded for free. However, unlike the relatively permissive license that the ESA provides for the use of its Sentinel-2 imagery, Google Earth imagery is restricted to non-commercial applications and downloads are limited to a maximum image size of 4800 pixels by 4800 pixels, meaning that downloading Google Earth imagery with a very high spatial resolution means that its spatial extent is reduced (European Space Agency 2018; Google 2015). The Sentinel-2 imagery offers one of the highest spatial resolution datasets currently openly available to the general public at no cost, a global coverage, as well as a high revisit frequency that would allow for frequent updates to road network data and change tracking (European Space Agency 2018). The most recently available images with minimal cloud cover for the study areas were downloaded through Google Earth and through the ESA's Copernicus Open Data Portal. Table 1 below details the characteristics of each satellite image used.

Table 1. Metadata for the satellite imagery used to extract the updated road network.

Area	Platform	Date	Resolution (m)	Spectral bands	Spatial extent (WGS84)
A	DigitalGlobe (Google Earth)	2012-02-02	0.4	Blue, Green, Red	38.21919°E, 6.16732°S 38.23417°E, 6.17612°S
A	Sentinel-2	2016-10-08	10	Red, NIR, SWIR	38.20588°E, 6.16559°S 38.26329°E, 6.19940°S
B	DigitalGlobe (Google Earth)	2013-09-01	0.4	Blue, Green, Red	35.66688°E, 8.11694°S 35.68429°E, 8.12711°S
B	Sentinel-2	2017-09-16	10	Red, NIR, SWIR	35.64261°E, 8.10211°S 35.70050°E, 8.13644°S
C	DigitalGlobe (Google Earth)	2013-11-19	0.4	Blue, Green, Red	38.36944°E, 10.92744°S 38.38659°E, 10.93737°S
C	Sentinel-2	2017-09-08	10	Red, NIR, SWIR	38.35395°E, 10.91467°S 38.41271°E, 10.94889°S

The Google Earth imagery also needs to be georeferenced, as only the pixel location expressed in row-column coordinates as well as the red-green-blue values for each pixel are provided in the downloaded file. Six ground control points are selected in each image. Each

point's row-column coordinates in the Google Earth imagery are linked to its corresponding latitude-longitude coordinates as defined by the WGS84 coordinate system. The ground control points are selected such that they align with features that are easy to identify and have a static position, for example, corners of buildings or road junctions. A linear transformation is applied to each image to map the row-column coordinates to WGS84 latitude-longitude coordinates, using the ground control points to determine the required parameters for the transformation. A linear transformation is chosen because the images are small in spatial extent and have already been preprocessed for publishing on Google Earth, meaning that distortions in the image due to the Earth's curvature and satellite irregularities are limited. The root mean square error for the exact linear transformations used does not exceed two pixels, equivalent to an error of less than one metre. The full geocorrection statistics are available in appendix I. The Sentinel-2 imagery is provided with geometric correction already applied by the ESA (European Space Agency 2018).

3.3. Evaluation of OpenStreetMap and Global Roads Open Access Dataset

The road networks published by OSM and gROADS are the two most readily available and easily obtainable road network datasets that provide a global coverage. These two datasets are also the foundation for the latest models of global and local accessibility; for example, Dorosh et al. (2011)'s market accessibility analysis for farming households in sub-Saharan Africa, Linard et al. (2012)'s population density dataset, and Weiss et al. (2018)'s global map of travel times to cities. Therefore, these two road networks have been selected as a base case to which possible improvements are compared.

The most recently available road networks at the time of writing this report are used for evaluation. The OSM dataset was last updated 24 March 2018 while the gROADS dataset was last updated 07 February 2013 (CIESIN 2013; OpenStreetMap 2015). Although both datasets are available with a global coverage, the two road networks are clipped so that only roads within the political boundaries of Tanzania are used for the evaluation.

The accuracy of each of the road networks is evaluated by comparing the roads in each dataset with a set of ground truth values. Ideally, the ground truth values are established through field work, whereby a set of randomly sampled points are observed throughout the country and their status as either 'road' or 'non-road' recorded. However, since conducting a field survey is out of the scope of this thesis project, the most recently obtainable high resolution satellite images and/or aerial photographs are used to determine whether a sampled point constitutes a 'road' or 'non-road'. For this purpose, the satellite images and aerial

photography published via Google's mapping services are used because the images collected and published via Google's mapping services are often the highest resolution images available to the general public.

A preliminary set of evaluation areas, approximately one square kilometre in extent, are randomly selected to account for varying degrees of urbanisation and population density. The country is divided into remote and non-remote areas as well as sparsely populated and densely populated areas in accordance to Linard et al. (2012)'s population density dataset and Weiss et al. (2018)'s travel time to cities dataset. This preliminary selection is reduced to a total of twenty sampling areas such that remoteness and degree of urbanisation are fairly represented because of the time-consuming nature of the manual visual evaluation process described below. The three study areas used to test the satellite imagery extraction methods are also included in the accuracy evaluation, for a total of twenty three sampling areas used for the accuracy evaluation.

Within each sampling area, a stratified random sampling strategy is employed to select the evaluation points. Three strata are used: the OSM dataset, the gROADS dataset, and the background image. Fifty points in each strata are randomly generated; this ensures that some of the evaluation points will lie on the OSM dataset and gROADS dataset to evaluate the accuracy of the existing road data itself, as well as in the background image to evaluate whether there exist any roads that are not represented in the OSM and gROADS datasets. This means that the total number of points sampled in each sampling area varies between 50 and 150 points, depending on if all, one, or none of the road networks are present within the sampling area. A sampling point is considered to lie on a road if the road passes within 5 metres of the sampling point. This evaluation is done by visually comparing the location of the sampling points plus a five metre radius buffer with the ground truth values as according to Google's most recent satellite image of the corresponding sampling area. A scale of 1 : 4 000 is used for the visual evaluation; roads and tracks visible at this scale are considered to be real roads. Figure 4 below illustrates the sampling and accuracy evaluation method.

The accuracy of each road network is then determined by calculating the producer's accuracy, user's accuracy, and overall accuracy for each of the sampling areas and for all of the sampling areas combined. Producer's accuracy refers to the ratio of the number of mapped road points that lie on a road to the number of total ground truth road points, and gives a sense of how likely a real road segment is to be represented in the map. User's accuracy refers to the ratio of the number of mapped points that lie of a road to total number of points that lie

along the mapped network, and gives a sense of how likely any given point along the mapped road network actually corresponds to a point on a real road.

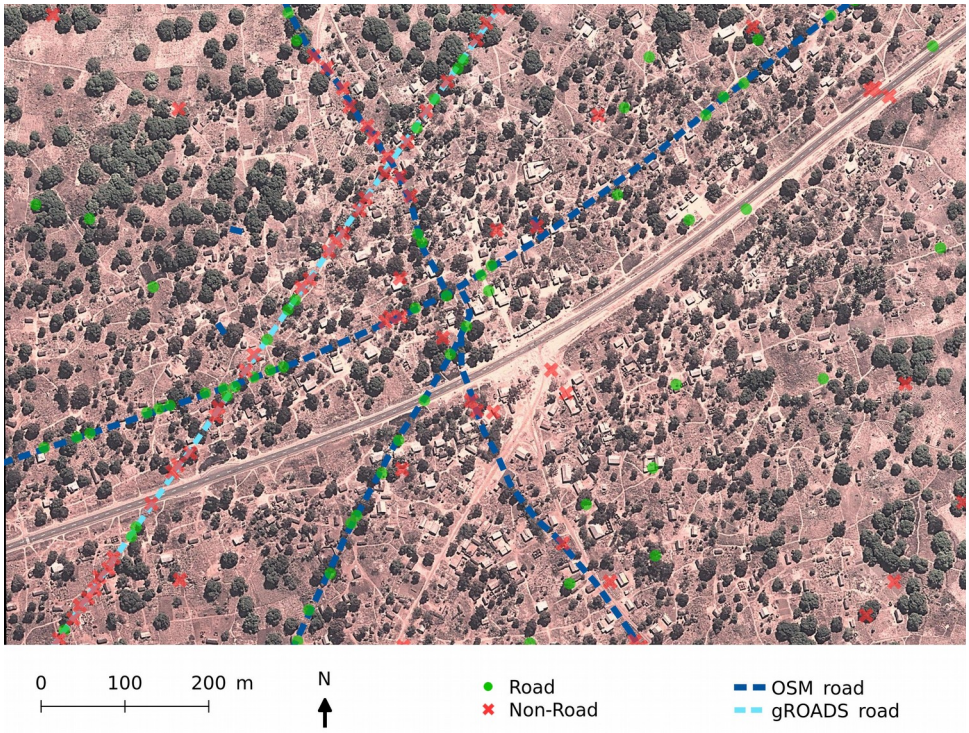


Figure 4. An example of the sampling method employed to evaluate the accuracy of the OSM and gROADS datasets. Green circles are randomly selected points that lie on a real road according the Google satellite imagery when observed at a scale of 1 : 4 000. Likewise, red crosses are those that do not lie on a real road. Satellite imagery provided by Google, Airbus DigitalGlobe.

3.4. Road Extraction from Satellite Imagery

The second component of this thesis project is to improve upon the existing global road network datasets that are evaluated in component one, namely OSM and gROADS. Until recently, such improvements would require extensive, costly, and time-consuming field work often in remote areas; however, recent developments in satellite technology together with a trend towards open-access data has provided the necessary framework that make remote sensing a viable option.

Because the Google Earth imagery has a high spatial resolution while the Sentinel-2 imagery has a much lower spatial resolution, a single method cannot be applied to both datasets to extract the updated road network. Instead, the method developed by Ojo et al. (2016) is applied to the Google Earth imagery while the Gomes et al. (2004) method is applied to the Sentinel-2 imagery.

3.4.1. Method for Google Earth Imagery

Because much of the high resolution satellite imagery is provided as three-colour bands (red-green-blue bands), the Ojo et al. (2016) method relies on being able to distinguish roads from their surroundings using the differences in colour. Adaptive Histogram Equalisation is applied to the images to enhance the colour contrast.

The contrast-enhanced images are then converted into the HSV (*hue, saturation, value*) and YC_BC_R (*luminance, blue-difference, red-difference*) colour spaces. Only the hue, saturation, and luminance values are needed for the analysis using this method. Equations 1a-b, 2, and 3 describe how hue, saturation, and luminance respectively are calculated from the red (*R*), green (*G*), and blue (*B*) values.

$$H = \begin{cases} \theta, B \le G \\ 360 - \theta, otherwise \end{cases} \tag{1a}$$

$$\theta = \cos^{-1}\left(\frac{1}{2}\frac{(R-G)+(R-B)}{\sqrt{(R-G)^2+(R-B)(G-B)}}\right) \tag{1b}$$

$$S = 1 - \frac{3 \cdot min(R,G,B)}{R+G+B} \tag{2}$$

$$Y = \frac{R+G+B}{3} \tag{3}$$

where: H is the hue of the colour
θ is the hue angle
S is the saturation of the colour
Y is the luminance of the colour
R is the red value
G is the green value
B is the blue value

Representative values for hue, saturation, and luminance for the roads are determined by sampling points in the image that lie along known roads or roads that can be easily visually identified. This is done by digitising several line segments along these roads and extracting the hue, saturation, and luminance values of the raster cells that coincide with a node along the digitised line segment. The number of nodes along each digitised line segment is increased by densifying the line geometry such that a new node is placed every 10 metres along each line. Increasing the number of nodes along the line geometry ensures that there

are a sufficient number of nodes that can be used as sample points along each road in the image.

The minimum and maximum values for H, S, and Y are determined by calculating the the mean (μ) and the standard deviation (σ) of the sampled points. Since the values can be shown to be normally distributed, the minimum and maximum values for each of H, S, and Y are initially set to be $\mu \pm 2\sigma$ to eliminate any outlier sample points.

Each pixel whose H, S, and Y values fall within the threshold values is classified as a potential road. This is determined by evaluating the logical statement given in Equation 4, where ∧ represents logical *AND,* $t_{min,H}$, $t_{max,H}$, $t_{min,S}$, $t_{max,S}$, $t_{min,Y}$, $t_{max,Y}$ represent the minimum and maximum threshold values of H, S, and Y respectively, and H_{pixel}, S_{pixel}, and Y_{pixel} represent the hue, saturation, and luminance of the candidate pixel. When all of H_{pixel}, S_{pixel}, and Y_{pixel} fall within the threshold values, then Equation 4 evaluates as true; if just one of the values is outside of the given threshold, then Equation 4 evaluates as false and the pixel is not considered to be a potential road.

$$\left(t_{min,H} \le H_{pixel} \le t_{max,H}\right) \wedge \left(t_{min,S} \le S_{pixel} \le t_{max,S}\right) \wedge \left(t_{min,Y} \le Y_{pixel} \le t_{max,Y}\right) \quad (4)$$

where: t_{min} is the minimum threshold value
t_{max} is the maximum threshold value
H_{pixel} is the hue of the pixel
S_{pixel} is the saturation of the pixel
Y_{pixel} is the luminance of the pixel

However, not all pixels in the image that fall within these thresholds are actual roads; because the roads being extracted from the image are unpaved, they share similar spectral and colour characteristics with other dirt surfaces such as agricultural fields. The classification needs to therefore be further refined by taking into account other defining characteristics of roads and surroundings, such as shape.

Since roads are long and narrow, it is expected that the spectral and colour characteristics will be relatively uniform along the length of the road while changing rapidly in the direction perpendicular to the road. This implies that the gradient, or rate of change, of the spectral characteristics along the edge of roads will be high, while the rate of change of the direction of the rate of change will be low. Pixels that represent other unpaved areas are expected to have a smaller gradient in varying directions.

Large rectangular or circle regions are further filtered out by calculating the eccentricity and solidity of the connected components of the road network. Eccentricity is a measure of how well a shape conforms to a circle. Solidity refers to the ratio between a region's area and the area of its bounding box. Regions with high eccentricity are more likely to represent roads because their shape more closely conforms to that that of a line or narrow ellipse. Regions with low eccentricity may also represent the road network if their solidity is also low. The final road network is the result after having filtered out these pixels.

3.4.2. Method for Sentinel-2 Imagery

The Sentinel-2 imagery have a 10m spatial resolution; much lower than the spatial resolution available for images downloaded from Google Earth. Therefore the spectral response for pixels containing roads contains interference from the spectral response of its surroundings because in many cases the width of the road is less than the width of a pixel. Because the relatively low spatial resolution of the pixel causes inconsistencies in the spectral response of the road pixels, Gomes et al. (2004) propose measuring the relative difference between the response of a road pixel with those of the surrounding pixels.

First, several Gaussian filters are applied to the image to reduce noise. These filters are oriented at 0°, 45°, 90°, and 135° measured counter-clockwise from North and are asymmetric, meaning that the filter assigns a greater weight to pixels in the same direction as the filter orientation than those that are perpendicular to the direction of the filter orientation. Equations 5a through 5c define the Gaussian filter weight H at pixel (x,y) with filter orientation θ. k is a constant that ensures that the weight of all pixels in the filter sum to 1, and σ_x and σ_y are constants that control the shape and size of the filter in the directions perpendicular and parallel to the filter orientation respectively. Since the filter orientation should be longer parallel to the road, σ_y must be greater than σ_x. A filter size of 5 pixels by 5 pixels is used to filter the images. A filter size of 5 is chosen through experimentation; a size 3 filter is too small to take into account significant values along the length of the road whereas a size 7 filter includes redundant pixels at its edges with little influence on the result.

$$H(x, y, \theta) = k \exp\left(-0.5\left(\frac{x_\theta^2}{\sigma_x^2} + \frac{y_\theta^2}{\sigma_y^2}\right)\right) \quad (5a)$$

$$x_\theta^2 = x\sin(\theta) + y\cos(\theta) \quad (5b)$$

$$y_\theta^2 = x\cos(\theta) - y\sin(\theta) \quad (5c)$$

where: H(x, y, θ) is the Gaussian filter weight
(x, y) is the coordinate of the pixel
θ is the filter orientation
(x_θ, y_θ) is the coordinate of the pixel after rotation by θ
k is a constant to ensure all weights sum to 1
σ_x is the distance from the pixel in the x-direction
σ_y is the distance from the pixel in the y-direction

After filtering the image, the relative difference between the spectral response of a potential road pixel and the surrounding pixels is calculated. Because each pixel contains interference from its surroundings, the spectral response of a pixel representing a road can be modelled as a combination of the spectral response of a dirt surface and the spectral response of the surrounding land cover, plus a margin of error (Gomes et al. 2004). This model is given by Equation 6 below, adapted from Gomes et al. (2004), where *r* represents the spectral response of the pixels, ε is the model error, and α is a constant that governs the proportion of the spectral response interference caused by the spectral response of the surrounding land cover. The more closely that any given pixel fits this model, the more likely that the pixel represents an unpaved road (Gomes et al., 2004).

$$r_{pixel} = \alpha r_{surroundings} + (1 - \alpha) r_{dirt} + \varepsilon \quad (6)$$

where: r_{pixel} is the spectral response of the pixel
$r_{surroundings}$ is the spectral response of the surrounding land cover
r_{dirt} is the spectral response of dirt
α is the proportion of spectral response interference from the surroundings
ε is the model error

The following relationship is obtained after rearranging Equation 6:

$$r_{pixel} - r_{dirt} = \alpha(r_{surroundings} - r_{dirt}) + \varepsilon \quad (7)$$

This relationship implies that the difference in the spectral response of a pixel and the spectral response of a dirt surface is a linear combination of the difference in the spectral response of the surrounding land cover and the model error. Figure 5 below shows a geometric interpretation of this relationship.

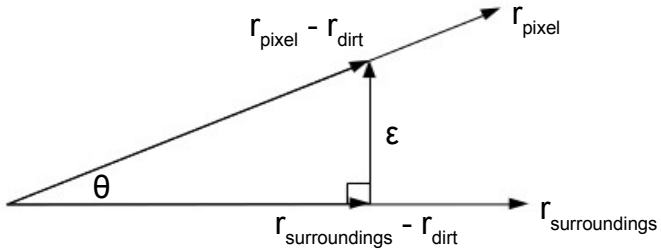


Figure 5. A geometric interpretation of the relationship between the pixel response of the road, the surrounding land cover, and dirt. Adapted from Gomes et al. (2004).

It can be concluded from figure 5 or through mathematical derivation that the smallest error value for a given pixel is the length of the vector perpendicular to the vector defined by $v = r_{surroundings} - r_{dirt}$. Likewise, α can be derived from the projection of the vector defined by $u = r_{pixel} - r_{dirt}$ on v. Therefore:

$$\varepsilon = \frac{|u|}{|v|} \sin\theta \tag{8}$$

$$\alpha v = proj_v u = \frac{u \cdot v}{|v|^2} v \Rightarrow \alpha = \frac{u \cdot v}{|v|^2} \tag{9}$$

where: α is the proportion of spectral response interference from the surroundings
ε is the model error
v is the vector difference between the spectral response of the surroundings and dirt based on the HSY-colour space
u is the vector difference between the spectral response of the pixel and dirt
θ is the angle between u and v

Values for α and ε are calculated for each pixel in the image, for every potential orientation of the road. The spectral response for the surrounding land cover is given by two pixels on either side of the road, that is to say, two pixels that lie on a line perpendicular to that of the road. A distance of five pixels on either side of the road is used to model the spectral response of the surroundings. The orientation with the lowest resulting ε value and the corresponding value for α is used to represent the pixel.

Road pixels are selected based on how well they fit the model. Because any given road pixel is a linear combination of the spectral response of dirt and the spectral response of the surrounding land cover, pixels that have a low model error are more likely to be roads. Secondly, a low α value indicates that a greater proportion of the pixel spectral response is due to dirt rather than the surrounding land cover; therefore, roads are also likely to occur on pixels with low α values. When these two factors are considered in tandem with the general narrow linear shapes that road features form in an image, this implies that pixels along lines of local minima in the ε and α images are most likely to be roads.

3.4.3. Evaluation of the Extracted Road Networks

The road networks created in component two of this thesis project are then compared to the existing road network datasets, OSM and gROADS. The same evaluation technique is used as for the OSM and gROADS data. A stratified random sample of points within each of the study areas is chosen such that they are distributed both along the mapped road network as well as randomly throughout the entire extent of the study area. The accuracy of the extracted road networks is determined by counting the number of points that correspond to real roads, as well as the number of points that lie along real roads that are not represented in the mapped road network. Similarly the number of points that lie along the mapped road network that do not correspond to a real road are counted.

4. Results

4.1. Evaluation of OpenStreetMap and Global Roads Open Access Dataset

A preliminary visualisation of the existing global road network datasets, namely OSM and gROADS, indicates that there is a significant difference in quality between the two datasets, as well as between the two datasets and the actual road network. This is in agreement with Hughes (2017), who suggests that the roadless areas identified by Ibisch et al. (2016) are likely to contain roads that are not represented in existing global road network datasets.

Figure 6 shows an overview of the OSM road network and the gROADS road network for the entirety of Tanzania. At this scale, it can be seen that the OSM road network is much more robust than the gROADS road network, containing many more roads than its counterpart. The difference between the two road networks is most evident within large urban centres and their immediate surroundings, as can be partially seen at the national scale in figure 6 and more clearly seen at a local scale in figure 7. The OSM dataset offers a much more extensive catalogue of roads within these areas that, aside from the main arterial roads, are completely absent from the gROADS road network. Conversely, figure 6 shows that the gROADS road network may offer a more complete regional road network. There are several regions in the map, particularly in the remote south-eastern and western regions of Tanzania, where several seemingly regionally important links are found in the gROADS data but not in the OSM data.

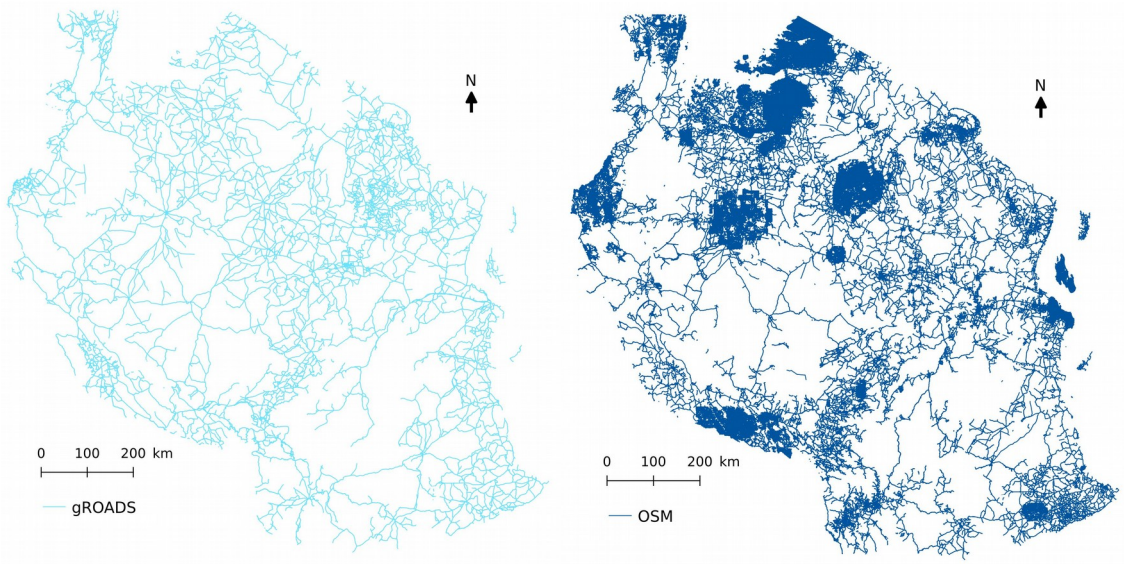


Figure 6. Overview map of Tanzania showing the extent of the gROADS (left) and OSM (right) road network data. UTM Zone 37S.

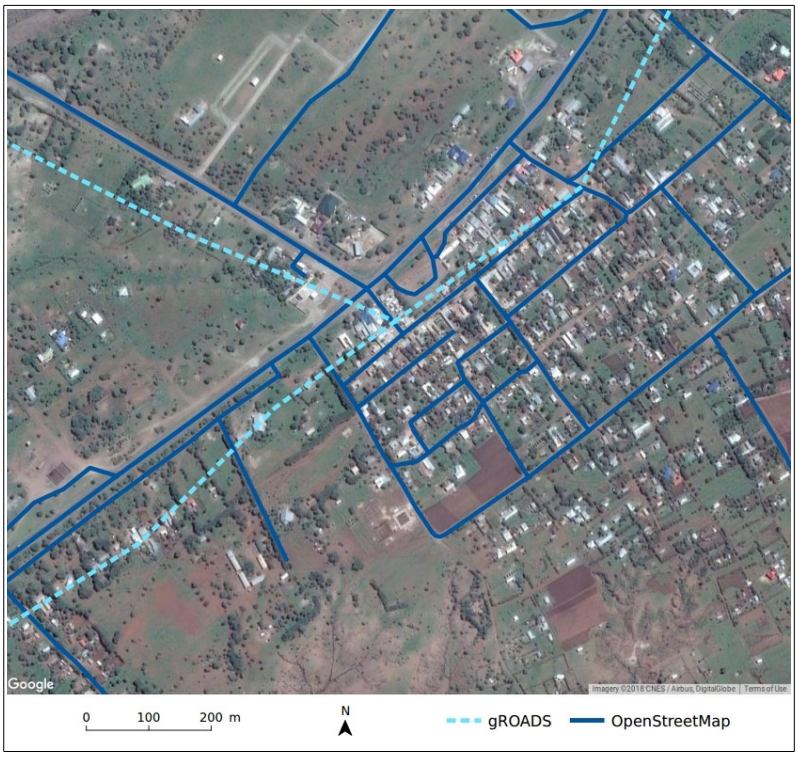


Figure 7. Local scale image showing the extent of the gROADS and OSM data. OSM data more completely maps the local streets with a high degree of spatial accuracy. The gROADS data exhibits a lower spatial accuracy and maps the fundamental regional routes. Background map data provided by Google, Airbus DigitalGlobe.

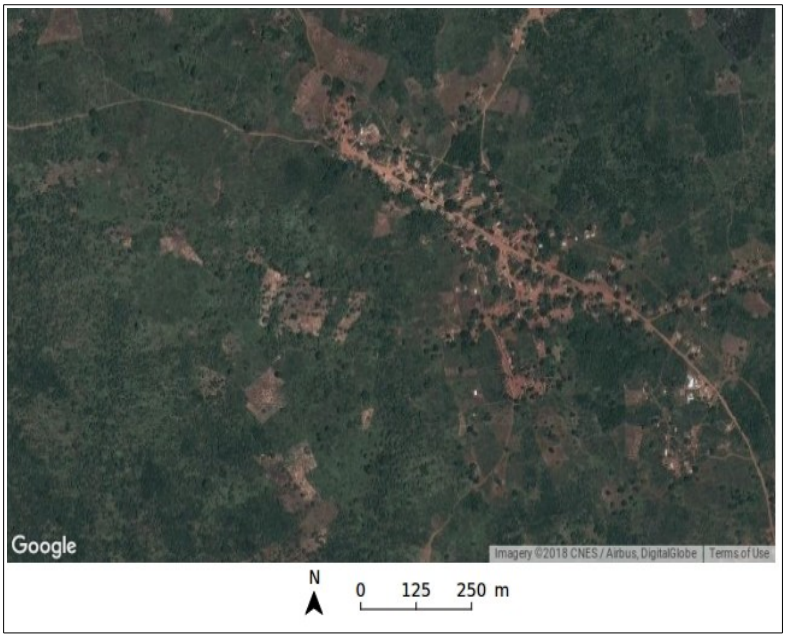


Figure 8. A satellite image of study area A. Satellite imagery shows clear evidence of a well-formed road network despite the lack of OSM and gROADS data. Satellite image provided by Google, Airbus DigitalGlobe.

As Ibisch et al. (2016) have noted, roads in many large areas of Tanzania are also missing from both the OSM and gROADS road networks. Figure 8 shows study area A, where Google satellite imagery shows clear indications that roads exist in the area. Although this problem is most apparent in the most remote regions of the country, the area shown in figure 8 is an example of the lack of available road network data in populated areas that are relatively close to urban centres.

Tables 2 and 3 below summarise the results of sample testing to determine the accuracy of both the OSM and gROADS datasets. A total of 2989 points across the 23 sample areas were visually evaluated; an average of 130 sampling points per area. The OSM dataset has a much higher producer's accuracy, user's accuracy, and overall accuracy compared to gROADS. When all of the sample areas and sampling points are considered together, the OSM road network has a producer accuracy, user accuracy, and overall accuracy of 71%, 86%, and 82% respectively. In contrast, the corresponding accuracy values for the gROADS dataset are 11%, 18%, and 39% respectively. OSM data are missing in two of the 23 sampled areas while gROADS data are missing in four of the sampled areas. The matrices used to calculate the producer and user accuracies are found in appendix II.

Table 2. Summary of the producer and user accuracy of the OSM and gROADS datasets for all sample areas.

	Producer accuracy (%)		User accuracy (%)	
	OSM	gROADS	OSM	gROADS
Missing areas	2 areas	4 areas	2 areas	4 areas
Minimum (Non-zero)	52	2	62	2
Maximum	100	28	100	47
Average	68	9	78	15

Table 3. Summary of the producer and user accuracy of the OSM and gROADS datasets for all sample areas combined.

	OSM	gROADS
Overall (%)	82	38
Producer (%)	71	11
User (%)	86	18

4.2. Method for Google Earth Imagery

Updated road networks for three study areas were extracted from the high resolution Google Earth imagery. Figures 9a-c below shows the three road networks as binary (road/non-road) raster layers.

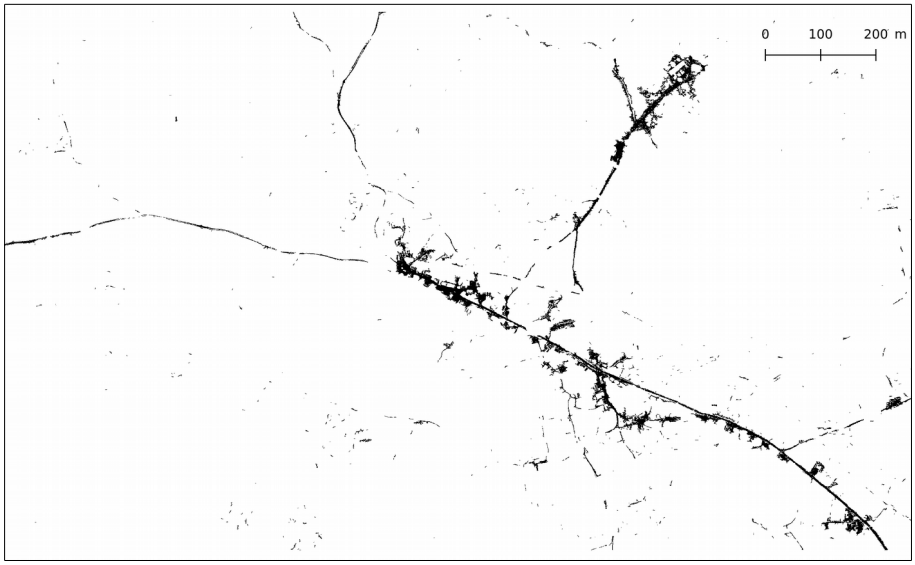


Figure 9a. Binary visualisations of the road network extracted in study area A in Tanzania from high-resolution Google Earth imagery using a method adapted from Ojo et al. (2016). Black pixels represent roads; white pixels represent non-roads.

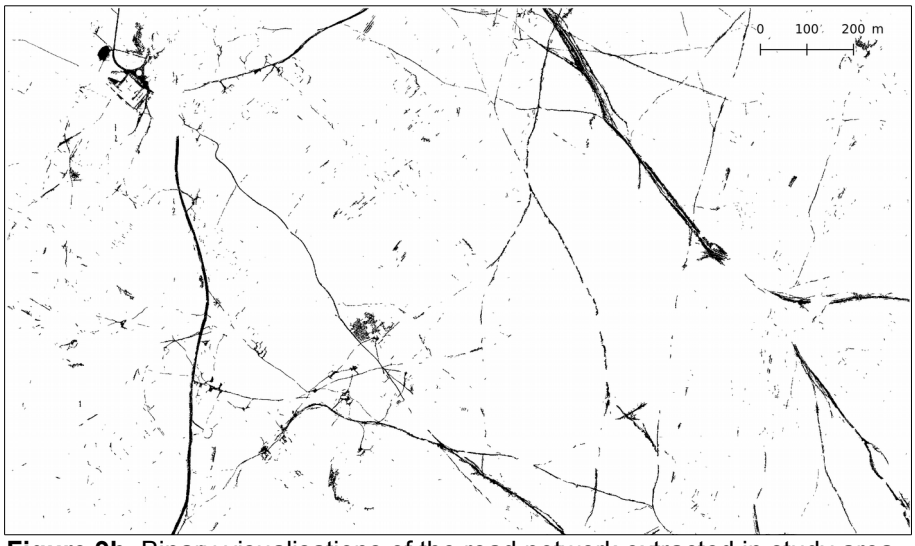


Figure 9b. Binary visualisations of the road network extracted in study area B in Tanzania from high-resolution Google Earth imagery using a method adapted from Ojo et al. (2016). Black pixels represent roads; white pixels represent non-roads.

Figure 9c. Binary visualisations of the road networks extracted in study area C in Tanzania from high-resolution Google Earth imagery using a method described in Ojo et al. (2016). Black pixels represent roads; white pixels represent non-roads.

The accuracy of each of the three updated road networks was evaluated using the same methodology used to evaluate the OSM and gROADS road networks. Sampling points were selected using a random stratified sampling strategy whereby half of the total points were selected in the regions representing roads, shown in figures 9a-c in black, and the other half in regions representing non-roads, shown in figures 9a-c in white. The producer accuracy, user accuracy, and overall accuracy for each road network, alongside those for the OSM and gROADS road networks in the same region, is summarised below in tables 4a-c.

Table 4a. Overall, producer, and user accuracy for each of the following road networks: Extracted network from high-resolution Google Earth imagery; OSM; gROADS. Study area A.

	Extracted network	OSM	gROADS
Overall (%)	88	0	0
Producer (%)	92	0	0
User (%)	83	0	0

Table 4b. Overall, producer, and user accuracy for each of the following road networks: Extracted network from high-resolution Google Earth imagery; OSM; gROADS. Study area B.

	Extracted network	OSM	gROADS
Overall (%)	77	75	40
Producer (%)	89	64	13
User (%)	62	82	29

Table 4c. Overall, producer, and user accuracy for each of the following road networks: Extracted network from high-resolution Google Earth imagery; OSM; gROADS. Study area C.

	Extracted network	OSM	gROADS
Overall (%)	74	81	29
Producer (%)	73	69	6
User (%)	76	83	8

Of the three study areas, only study area A had no OSM or gROADS road data available. Study area A also has the highest accuracies amongst the three study areas, with overall, producer, and user accuracies of 88%, 92%, and 83% respectively.

Figures 10a-c below shows the extracted road network in each study area in comparison to the OSM and gROADS road networks as well as the satellite image used for extraction and validation.

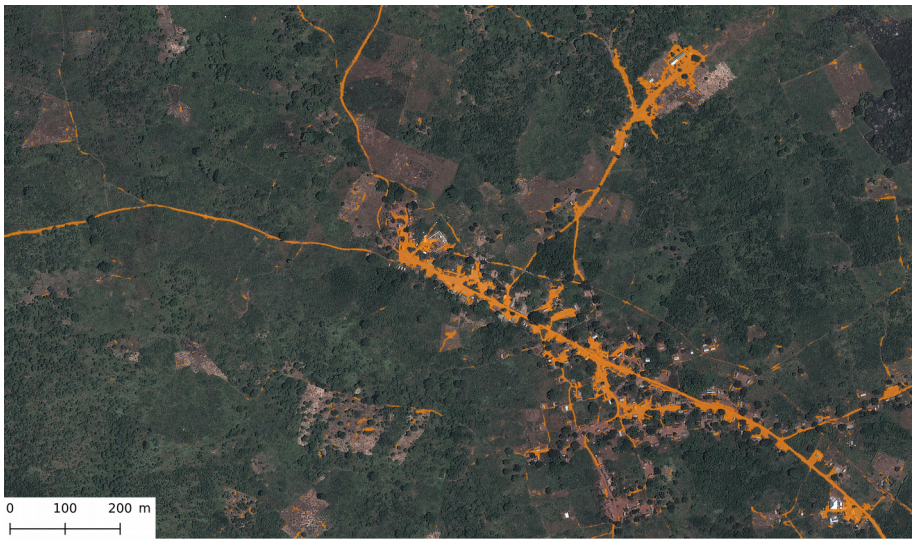


Figure 10a. Visualisation of the road network extracted in study area A from high-resolution Google Earth imagery using the method described in Ojo et al. (2016). In orange/red are the extracted road network. Both gROADS and OSM data are missing from this area. The satellite imagery used for extraction and validation is set as the background and is provided by Google, Airbus DigitalGlobe.

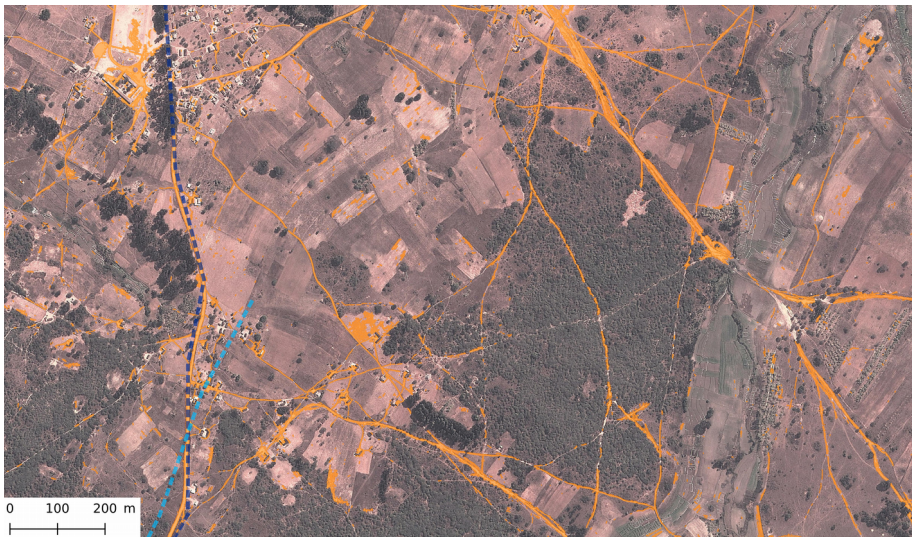


Figure 10b. Visualisation of the road network extracted in study area B from high-resolution Google Earth imagery using the method described in Ojo et al. (2016). In orange/red are the extracted road network. The gROADS road network is shown in light blue and the OSM road network is shown in dark blue. The satellite imagery used for extraction and validation is set as the background and is provided by Google, Airbus DigitalGlobe.

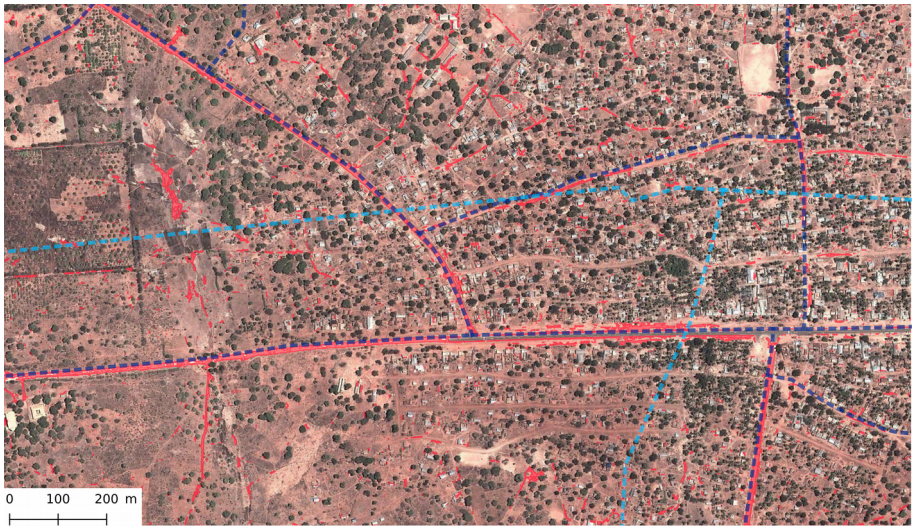


Figure 10c. Visualisation of the road network extracted in study area C from high-resolution Google Earth imagery using the method described in Ojo et al. (2016). In orange/red are the extracted road network. The gROADS road network is shown in light blue and the OSM road network is shown in dark blue. The satellite imagery used for extraction and validation is set as the background and is provided by Google, Airbus DigitalGlobe.

4.3 Method for Sentinel-2 Imagery

A second set of lower resolution Sentinel-2 imagery was used to test the road network extraction method outlined in Gomes et al. (2004). Because the Sentinel-2 imagery has a 10m spatial resolution, significantly lower than the half-metre resolution of the Google Earth imagery, the road network is extracted by calculating the ratio between the response of a potential road pixel with the response of the surrounding land cover pixels that interfere with the spectral response in the road pixel. Equation 6 describes how the ratio between road pixels and the surrounding land cover pixels is calculated.

Two values for every pixel in each study area are calculated: ε and α. ε describes how well Equation 6 models the ratio between the road pixel and its surroundings, while α describes the strength of the interference of the surrounding pixels on the spectral response of the road pixel. Figure 11 shows the resulting α image and ε image for study area B as an example.

The skeleton of a road network can be seen in the α image. Dark lines in the top and centre of the images correspond with roads visible in the Sentinel-2 imagery. Other traces of the road structure are also visible as lines that vary in intensity along their length, but have a smaller α value than the surrounding pixels. See examples I and II circled in figure 11. This observation follows from the definition of α as the proportion of the spectral response of the surrounding land cover that interferes with the spectral response of the dirt surface on the road. Roads are represented smaller α values than their surroundings because a larger proportion of the pixel is covered with the dirt road surface. α varies along the length of the road with variations in the surrounding land cover.

The model error values should therefore help to reduce the number of candidate road pixels. However, as can be seen in figure 11, the ε image is quite noisy and there is no discernible overarching road network structure. Upon closer inspection along the lines shown in the α image that correspond with real roads, it seems that the ε values vary significantly along their length. There is little indication that the real roads correspond with lines of local minima in the ε image. Therefore, no road network was extracted from the Sentinel-2 imagery using this method.

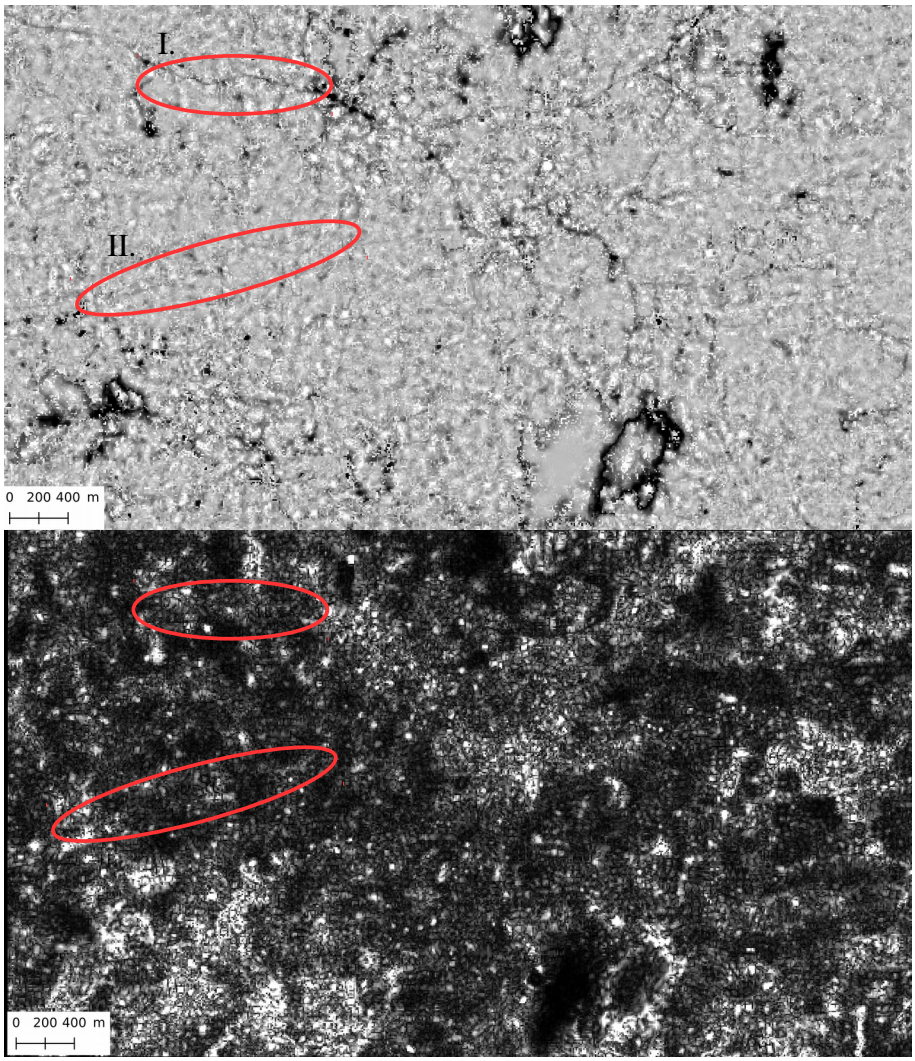


Figure 11. Images showing the calculated α (top) and error (bottom) values for study area B. Values for α range from 0 to 1.4 and are shown on a gradient from black to white. Likewise, values for ε range from 0 to 0.24 and are shown on a gradient from black to white. Circled regions show that traces of lines that are visible in the α image but are difficult to identify in the ε image.

5. Discussion

5.1. Evaluation of OpenStreetMap and Global Road Open Access Dataset

To evaluate the quality of the results for analyses undertaken using these datasets as an input, one must therefore first also know the quality of these datasets. OSM and gROADS differ in terms of their aims and data collection methods. OSM intends to offer a more comprehensive road network that can reliably used as an alternative to Google's datasets or officially publish road network data. The data are crowd-sourced, meaning that it is collected and published by the users of OSM themselves rather than centrally collected and published by the organisation. The gROADS dataset does not aim to be as comprehensive as the OSM dataset; instead, gROADS aims to catalogue regional roads between settlements rather than local roads within settlements (CIESIN 2013). The regional scale of the gROADS dataset is most suitable for mapping at scales of approximately 1 : 250 000 and spatial inaccuracies of up to 50 metres are tolerated (CIESIN 2013). Consequently, OSM data provides more comprehensive data that is required for analyses at local scales, while gROADS data are only sufficient for analyses at regional, national, or global scales.

The results of the accuracy testing performed in this thesis project are indicative of the aims and data collection methods used by OSM and gROADS. Over the 23 sampled areas throughout Tanzania, the OSM dataset is estimated to be 82% accurate overall while the gROADS dataset is estimated to be 39% accurate overall. The low accuracy of the gROADS dataset relative to that of OSM can be attributed to both the intended scale of use of the dataset as well as the methodology used to evaluate its accuracy.

The spatial inaccuracies in the gROADS dataset can easily be seen at the 1 : 4 000 scale used for the accuracy evaluation. In many of the sample cases, some of the points sampled along the gROADS dataset that were considered to be incorrectly mapped do actually correspond to a real road. However, because the accuracy evaluation was conducted at a local scale rather than at a regional scale, the spatial accuracy threshold used for the gROADS road network geometries means that the roads seem to be far off from their true positions in the real world as defined by the satellite imagery. Figure 12 below shows how the general pattern of the gROADS road network matches the real road position in the satellite imagery. The geometry of the gROADS road network seems to be very inaccurate at this scale, but the general shape and direction of the geometries match those of the underlying satellite image. In some cases, the distance between the gROADS road network and the real road network is also within the 50m spatial inaccuracy threshold. If the accuracy evaluation were to have been conducted

with a 50m tolerance for the gROADS dataset using a scale of 1 : 250 000, then it is probable that the accuracy of the dataset would increase.

The scale used for the accuracy evaluation also negatively affects the overall accuracy of the gROADS dataset because the local roads within settlements that are not intended to be mapped in the gROADS dataset are visible. Sampling points within each study area are considered to lie on a road if they are within 5 metres of a road visible in the satellite image at a scale of 1 : 4 000. Many of these roads are not visible at the 1 : 250 000 scale intended for viewing the gROADS road network. The overall accuracy of the gROADS network would increase if points lying on minor local roads were omitted from the evaluation.

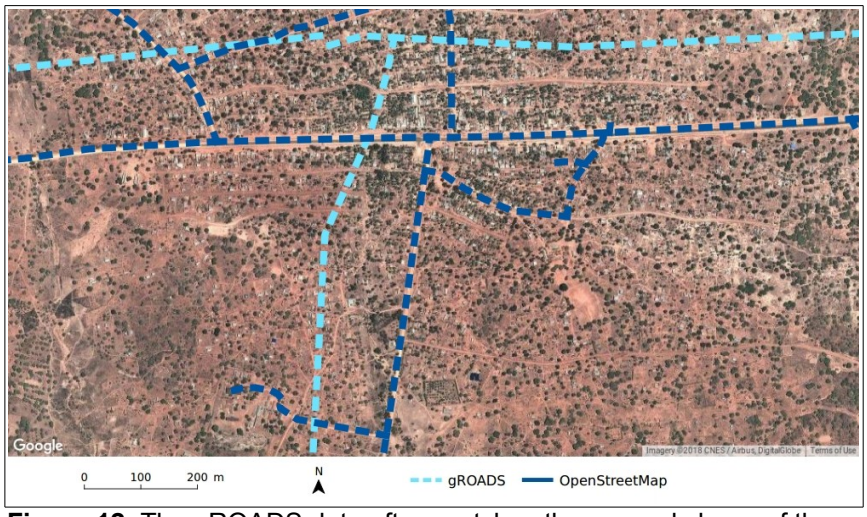


Figure 12. The gROADS data often matches the general shape of the real roads shown in the satellite imagery. However, gROADS is intended to be viewed at a regional scale of about 1 : 250 000. Viewing the gROADS data at this scale means that the tolerated spatial inaccuracies at the regional scale become greatly exaggerated. Satellite image provided by Google, Airbus DigitalGlobe.

Both the OSM and gROADS road networks are estimated to have a higher user accuracy than producer accuracy. Therefore, that the user accuracy for both OSM and gROADS is higher than the producer accuracy means that it is more common for real roads to be missing from the mapped road network datasets than for the mapped segments to be geometrically inaccurate. This mirrors the claims by Barrington-Leigh and Millard-Ball (2017), Ibisch et al. (2016), and Hughes (2017) that there are still substantial gaps in the OSM and gROADS datasets despite the data that does exist within each dataset being fairly accurate. As Barrington-Leigh and Millard-Ball (2017) suggest, some of the omissions in the OSM dataset correspond with regions with low-to-medium population density. Regional road segments between settlements in very low density areas are often well catalogued along with the road

networks in urban areas where the greatest number of OSM contributors reside. Local roads within small settlements and the outskirts of large urban regions are least represented in the OSM road network amongst the sampled areas.

The methodology employed to evaluate the accuracy of each road network dataset as well as the methodology used to collect each dataset creates some uncertainty in the evaluation beyond the sources of error already described above. Firstly, the accuracy estimates are not globally transferable. This is because OSM and gROADS rely input data collected by its users and external agencies, each with their own data collection methodologies and accuracy tolerances (CIESIN 2013; OpenStreetMap Contributors 2015). The road networks in each of the datasets are most accurate and complete in areas with the largest concentration of users and for countries with stable systems of governance and high Internet penetration (Barrington-Leigh and Millard-Ball 2017; Haklay 2010). Secondly, the satellite images used may not be up-to-date or consistent with the date of data collection in OSM and gROADS. Thus, changes in the real road network between the date the image was taken and the date the road data was collected may not be visible in the imagery. Thirdly, the evaluation relies on manual visual identification of real roads from the satellite imagery. It is possible that several points are misclassified due to obstructions, distortions, or ambiguities in the satellite image, or because the point was skipped over. Fourthly, the limited number of sample areas and the distribution of sampling points within each sample area leads to uncertainty in the evaluation. A more robust set of sample areas would be ideal to better capture the variation in the road network datasets due to population density, land cover types, and other possible factors. Lastly, there are also a large number of random sampling points that are classified as non-road by OSM, gROADS, and the satellite image, which inflates the overall accuracy of each dataset.

5.2. Methods for Road Extraction from Google Earth and Sentinel-2 Imagery

Figures 9a-c shows the final extracted road networks in each of the three study areas. The accuracy of the road networks is estimated to be quite high using the same methodology applied to the OSM and gROADS datasets; with overall accuracies of 88%, 77%, and 74% for study area A through C respectively. These estimates are comparable to that of the OSM dataset in study area B and C, where its accuracy is estimated to be 75% and 81% respectively. The estimated overall accuracy of the gROADS dataset is much lower than that of the extracted road network; even considering that the gROADS dataset is intended to map regional roads at a regional scale and the limitations of the methodology used to evaluate its accuracy, this method successfully extracts several regional roads not present in the gROADS

dataset. No OSM data or gROADS data were available for study area A; thus, these results show that there is good potential to augment the existing datasets using Google Earth imagery.

That the producer accuracy is higher than the user accuracy for all three study areas is opposite of the pattern that characterises the OSM and gROADS datasets. This indicates that the extent of the road network is greatly overestimated despite the algorithm correctly identifying many of the missing road segments. The OSM and gROADS datasets have a higher user accuracy than producer accuracy, which means that although the data may be incomplete, the roads that are mapped within each database are fairly accurate.

This difference can be partially attributed to the method itself and the difficulties of extracting dirt roads from satellite imagery. The algorithm detects roads by identifying pixels that fall within a given threshold range that characterises the colour of the road's dirt surface. It is often the case that other non-road features in the image, such as bare agricultural fields, are also characterised by the same spectral colour pattern. This problem is compounded by the fact that the method is being applied to high spatial resolution imagery. Excess noise due to the increased variation between the small pixels means that the spectral colour range that defines the dirt surface needs to be quite broad. Additionally, road features appear as elongated polygons at this resolution rather than edges, meaning that there is some difficulty for the algorithm to distinguish between a short road and narrow agricultural fields by their shape. Discontinuities in the road network and clusters of noise in the background of the results are manifestations of these challenges.

5.3. Comparisons to Previous Works

These results are comparable to those in Ojo et al. (2016). Images of the extracted road network from their study show that the algorithm successfully identifies road segments but has difficulty distinguishing roads from other similar features. Ojo et al. (2016) estimate that their road network is fundamentally continuous and complete, but their evaluation of the accuracy of the extraction process ignores that there are many noise pixels classified as roads. The road networks extracted in this thesis project contain more and larger discontinuities than in Ojo et al. (2016), though some of this may be attributable to the apparent difference in scope of the projects. The satellite images and results shown in their paper are approximately 350 pixels by 300 pixels in size whereas those used for this thesis project are approximately 4800 pixels by 2800 pixels in size, with similar spatial resolution (Ojo et al. 2016). The smaller spatial extent of the study areas used by Ojo et al. (2016) means that there is likely to be less variation between pixels representing the same road features than in an image that

covers a much larger extent. Consequently, the spectral response thresholds that characterise the dirt surface in a larger image need to be broader to take into account the larger variation between pixels in different areas within the image.

Results from Gomes et al. (2004)'s paper show that their method can successfully extract unpaved dirt roads using Landsat TM satellite imagery with a 30m spatial resolution using the red, near-infrared, and short wave infrared bands. Because the Sentinel-2 imagery has a spatial resolution three times finer than that of the Landsat TM imagery used by Gomes et al. (2004), then it follows that some of the roads not visible in the 30m resolution imagery are visible in the 10m resolution imagery. Indeed, tests to extract road network data conducted by Radoux et al. (2016) using Sentinel-2 imagery indicate that roads with widths as narrow as three metres can be identified and extracted. However, this method is limited to paved roads, where problems due to spectral similarity with the surroundings and spectral variation along the length of the road do not arise (Radoux et al. 2016).

Figure 11 shows the values for ε and α calculated for each pixel for study area B. Gomes et al. (2004) state that pixels in the Sentinel-2 image that are most likely to represent read roads occur at linear minima in the ε image because these pixels most closely fit the model described by Equation 6 than the surrounding pixels. Nevertheless, the ε values calculated for each of the study areas in this thesis project do not revel any linear patterns in the local minima. In contrast, the values for α do show clear linear patterns along the main roads. This follows from the definition of α, which is the proportion of the measured spectral response for any given pixel that is caused by interference from the spectral response of the surrounding pixels. Since more of the area covered by a pixel representing a road is a dirt surface, then it follows that α is lower for the pixel. However, the relationship does not seem to hold when measuring the model error; therefore a road network could not be extracted from the Sentinel-2 imagery by applying this method in this particular case. The linear patterns in the α value alone are not sufficient to extract the road network because of the large variation in α along each of the lines.

5.4. Spatial Resolution

The spatial resolution of the Sentinel-2 image offers a potential explanation for the lack of linear patterns in the ε image. As is the case with Google Earth imagery relative to Sentinel-2 imagery, the spatial resolution of the Sentinel-2 imagery is considerably higher than the spatial resolution of the Landsat TM imagery originally used by Gomes et al. (2004). For every pixel in the Landsat TM image, there exists nine pixels in the Sentinel-2 image that

cover the same spatial extent. Therefore, variations in land cover and the features within each Landsat TM pixel have the potential to introduce noise and variation within the nine Sentinel-2 pixels. With a coarser spatial resolution, noise and variation between adjacent pixels is reduced because the differences in spectral response are averaged out.
Because the method proposed by Gomes et al. (2004) relies on measuring the difference in spectral response relative to the surrounding pixels, the method is sensitive to noise and variation in the spectral response. Noise and variation in the spectral response could be reduced by resampling the image to a coarser spatial resolution, but this reduces the benefits of using Sentinel-2 imagery to detect narrower roads than would be possible with older satellite platforms with coarse resolutions. Figure 13 below shows how large variations in a set of 10m resolution pixels can be averaged out in a set of 30m resolution pixels. Figure 14 shows the level of detail visible in the Sentinel-2 imagery and examples of the large variation in pixel response along roads.

The spatial resolution of the Google Earth imagery is much higher than the Sentinel-2 imagery and thus is subject to larger variations in the measured spectral response between pixels. Crucially different, however, is that the Google Earth image pixels are smaller than the objects being extracted from the image. This means that the spectral response of the land cover adjacent to the road has little to no effect on the spectral response of the road pixel. Therefore, the higher spatial resolution of the Google Earth imagery allows for the calculation of more accurate and representative threshold values to classify road pixels than is the case for the Sentinel-2 imagery. Furthermore, the variation between pixels of the same type, that is road pixels and land cover pixels, can be more easily averaged out using filters; the higher spatial resolution implies that nearby pixels in a filter of any given size are more likely to be alike than in an image with a lower spatial resolution because the spatial extent that the filter covers is smaller.

67	76	74	65	78	73
68	72	71	67	74	71
70	70	67	66	70	71

70	70

Figure 13. An example comparison between a high resolution image on the left and a low resolution image on the right. The high resolution image exhibits a higher degree of variation in the pixel response because a smaller area is being measure by each pixel sensor. In the low resolution image, the variation in the larger area for each pixel is averaged out and produces uniformity.

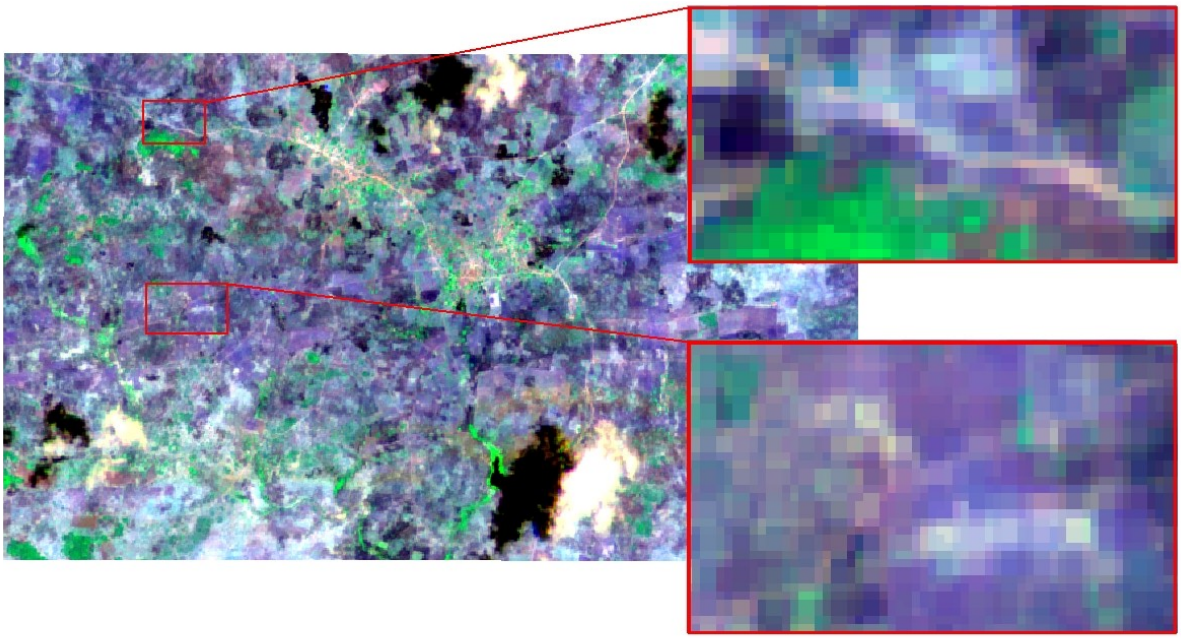


Figure 14. A false-colour composite of the Sentinel-2 imagery in study area A with the red, near-infrared, and short wave infrared bands represented as red, green, and blue respectively. Some lines representing the road network are visible in the imagery. Zooming in on some of these lines shows that the pixel response varies along the line and in the surroundings, making it difficult to model the ratio between the spectral response of the dirt surface and the surrounding land cover for candidate road pixels.

Conclusions

Access to complete, up-to-date, and reliable datasets is critical for researchers and policy makers to make informed decisions. In particular, access to such road network data is critical for reducing food insecurity amongst the world's most food insecure populations, as contemporary methods focus on improving accessibility to markets rather than attempting to increase crop yields (Dorosh et al. 2011; Erikson et al. 2009; Frelat et al. 2016). However, an overview of relevant literature has shown that the existing global road network datasets OSM and gROADS are often incomplete in regions where official data are infrequently updated or lacking altogether (Barrington-Leigh and Millard-Ball 2017; Ibisch et al. 2016; Hughes 2017). Existing datasets are adequate to measure accessibility on regional and national scales but lack the completeness necessary to perform local analyses.

Advances in the technical capabilities of satellite platforms and a substantial body of research into road extraction methods make it possible to extract road network data using satellite imagery where existing datasets lack sufficient quality. Google Earth and the ESA's Sentinel-2 satellite platform have a strong potential to contribute to updating road network data because of their global coverage, high spatial resolution compared to older satellite platforms, and most importantly, their ease of access by researchers, policy makers, and the general public. The high resolution data offered by Google Earth with a pixel spatial resolution of less than one metre means that even the smallest roads are visible; however, these data are prone to greater spectral variation and noise, offered only with red, green, and blue spectral bands, and is limited by more restrictive licensing and download capabilities (Google 2015). In contrast, Sentinel-2 imagery offers thirteen spectral bands and has a less limiting license, but suffers from lower 10m, 20m, and 60m resolution pixels which limits the visible detail (European Space Agency 2018).

Two methods to extract unpaved dirt road networks from Google Earth imagery and Sentinel-2 imagery were applied with mixed results. The method adapted from Ojo et al. (2016) applied to high resolution Google Earth imagery produced a road network that is more complete than both the OSM and the gROADS data in the three study areas tested in this thesis. This method is promising because it produces a rather robust road network using only colour and shape information; however, the resulting road network contains a significant amount of noise that reduces its user accuracy in comparison to OSM and gROADS data. The Gomes et al. (2004) method was applied unsuccessfully to the Sentinel-2 imagery. This is likely because this method relies heavily on the relationship between pixels that represent potential roads and the surrounding pixels. With the 10m spatial resolution of the Sentinel-2

pixels, the noise and variation in spectral response between the road pixel and its surroundings is too large to detect linear patterns that are clearer in the lower spatial resolution Landsat TM imagery originally used by Gomes et al. (2004).

Though the preliminary results are promising, much work is still needed to refine the methods to be suitable for where they are needed most. Sub-Saharan Africa in particular is home to some of the most food insecure populations and remote settlements while simultaneously having good agricultural potential and some of the largest gaps in the existing global road network datasets (Barrington-Leigh and Millard-Ball 2017; Dorosh et al. 2011; Frelat et al. 2016; Ibisch et al. 2016; Linard et al. 2012; Weiss et al. 2018). The methods tested in this thesis project need to be further refined so that unpaved roads can be more reliably differentiated from their surroundings. In addition, a limited number of evaluation sites were tested in this thesis; a more extensive evaluation of the OSM and gROADS datasets based on remoteness, population density, and agricultural potential can help to identify areas where satellite imagery is most needed to extract accurate road network data so that investments in agricultural development are effective and efficient.

7. References

Auclair-Fortier, M.-F., D. Ziou, C. Armenakis, and S. Wang. 2002. Survey of work on road extraction in aerial and satellite images. University of Sherbrooke, Technical Report No. 241, Sherbrooke, Canada, 14 pp.

Barrington-Leigh, C. and A. Millard-Ball. 2017. The world's user-generated road map is more than 80% complete. *PLoS ONE* 12(8): e0180698.

Baumgartner, A., C. Steger, H. Mayer, W. Eckstein, and H. Ebner. 1999. Automatic road extraction based on multi-scale, grouping, and context. *Photogrammetric Engineering & Remote Sensing* 65(7): 777-785.

Centre for International Earth Science Information Network (CIESIN). 2013. Global Roads Open Access Data Set, Version 1 (gROADSv1). Retrieved 07 April, 2018, from: http://dx.doi.org/10.7927/H4VD6WCT

Chamberlin, J., M. Tadesse, T. Bensson, and S. Zakaria. 2007. An atlas of the Ethiopian rural economy: Expanding the range of available information for development planning. *Information Development* 23(2): 181-192.

Dorosh, P., H.G. Wang, L. You, and E. Schmidt. 2011. Crop production and road connectivity in sub-Saharan Africa: A spatial analysis. *Agricultural Economics* 43(2012): 89-103.

Erikson, P.J., J.S.I. Ingram, and D.M. Liverman. 2009. Food security and global environmental change: Emerging challenges. *Environmental Science and Policy* 12(4): 373-377.

European Space Agency. Sentinel-2 MSI Introduction. Retrieved 24 March 2018, from https://earth.esa.int/web/sentinel/user-guides/sentinel-2-msi.

Fan, S. and P. Hazel. 2001. Returns to public investments in the less-favored areas of India and China. *American Journal of Agricultural Economics* 83(5): 1217-1222.

Frelat, R., S. Lopez-Ridaura, K.E. Giller, M. Herrero, S. Douxchamps, A.A. Djurfeldt, O. Erenstein, B. Henderson, et al. 2016. Drivers of household food availability in sub-Saharan Africa based on big data from small farms. *Proceedings of the National Academy of Sciences* 113(2): 458-463.

Gomes, O.F. M., R.Q. Feitosa, and H.L.C. Coutinho. 2004. Sub-pixel unpaved roads detection in Landsat images. In The *International Archive of Photogrammetry, Remote Sensing and Spatial Information Sciences XXXV, Part B3*, ed. O. Altan, pp. 1201-1206. XXth International Society for Photogrammetry and Remote Sensing Congress. Istanbul: ISPRS.

Google. 2015. Using Google Maps, Google Earth and Street View. Retrieved 15 April, 2018, from: https://www.google.com/permissions/geoguidelines.html.

Haklay, M. 2010. How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning B: Planning and Design* 37: 682-703.

Hughes, A.C. 2017. Global roadless areas: Hidden roads. *Science* 355(6332): 1381.

Ibisch, P.L., M.T. Hoffmann, S. Kreft, G. Pe'er, V. Kati, L. Biber-Freudenberger, D.A. DellaSala, M.M. Vale, et al. 2016. A global map of roadless areas and their conservation status. *Science* 354(6318): 1423-1427.

Keskinen, A. 2007. Mapping road infrastructure in developing countries: Applying remote sensing and GIS – The case of the Taita Hills, Kenya. Masters Thesis. Helsinki, Finland: University of Helsinki.

Linard, C., M. Gilbert, R.W. Snow, A.M. Noor, and A.J. Tatem. 2012. Population distribution, settlement patterns and accessibility across Africa in 2010. *PLoS ONE* 7(2): e31743.

Ojo, J.A., J.A. Akinlabi, and I.A. Adeyemo. 2016. Unpaved road extraction from satellite images. *European Journal of Computer Science and Information Technology* 4(5): 1-8.

OpenStreetMap Contributors. 2015. Planet dump. Retrieved 24 March, 2018, from https://planet.openstreetmap.org/.

Radoux, J., G. Chomé, D.C. Jacques, F. Waldner, N. Bellemans, N. Matton, C. Lamarche, R. d'Andrimont, et al. 2016. Sentinel-2's potential for sub-pixel landscape feature detection. *Remote Sensing* 8(6): 488-515.

Riebeek, H. 2014. Why is that forest red and that cloud blue? How to interpret a false-color satellite image. Retrieved 15 April, 2018, from: https://earthobservatory.nasa.gov/Features/FalseColor/page1.php.

Science Education through Earth Observation. n.d. Classification Algorithms and Methods. Retrieved 15 April, 2018, from: http://www.seos-project.eu/modules/classification/classification-c01-p05.html.

Smalley, R. 2017. Agricultural growth corridors on the eastern seaboard of Africa: An overview. Agricultural Policy Research in Africa, Future Agricultures Consortium, Working Paper 01, Brighton, UK, 41 pp.

Southern Agricultural Growth Corridor of Tanzania (SAGCOT). Retrieved 31 January, 2018, from http://www.sagcot.com/.

Stifel, D. and B. Minten. 2008. Isolation and agricultural productivity. *Agricultural Economics* 39(1): 1-15.

Tatem, A.J. 2010. Tanzania AfriPop Data 2010 (alpha version). Southampton, United Kingdom: University of Southampton (United Kingdom). Retrieved 12 April 2018 from http://ghdx.healthdata.org/record/tanzania-afripop-data-2010-alpha-version.

United Nations. Goal 2: End hunger, achieve food security and improved nutrition and promote sustainable agriculture. Retrieved 24 March 2018, from http://www.un.org/sustainabledevelopment/hunger/.

United States Geological Survey. 2018. About Landsat. Retrieved 24 March, 2018, from https://landsat.usgs.gov/about-landsat.

van Ittersum, M.K., L.G.J. van Bussel, J. Wolf, P. Grassini, J. van Wart, N. Guilpart, L. Claessens, H. de Groot, et al. 2016. Can sub-Saharan Africa feed itself?. *Proceedings of the National Academy of Sciences* 113(52): 14964-14969.

Wang, J., P.M. Treitz, and P.J. Howarth. 1992. Road network detection from SPOT imagery for updating geographical information systems in the rural-urban fringe. *International Journal for Geographical Information Systems* 6(2): 141-157.

Wang, W., N. Yang, Y. Zhang, F. Wang, T. Cao, and P. Eklund. 2016. A review of road extraction from remote sensing images. *Journal of Traffic and Transportation Engineering* 3(3): 271-282.

Weiss, D.J., A. Nelson, H.S. Gibson, W. Temperley, S. Peedell, A. Lieber, M. Hancher, E. Poyart, et al. 2018. A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature* 553: 333-336.

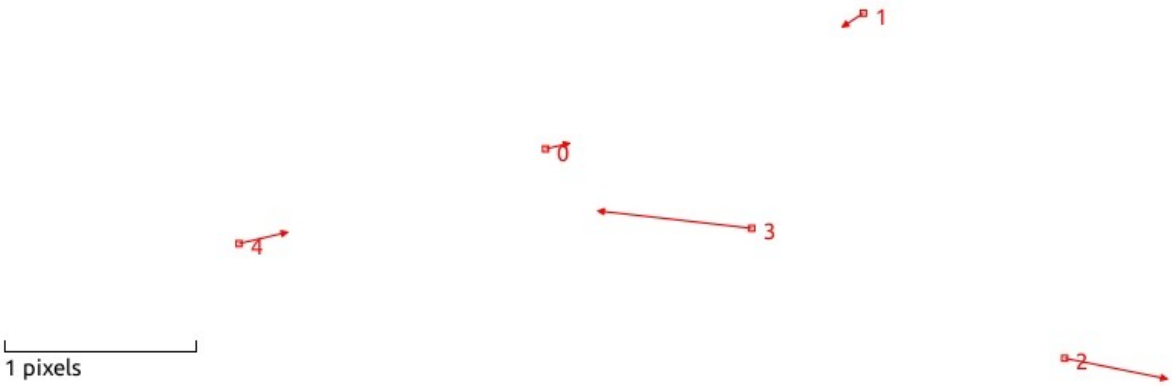
8. Appendices

Appendix I. Geocorrection Statistics for Google Earth Imagery

Transformation parameters (Linear)

Translation x	Translation y	Scale x	Scale y	Rotation [degrees]	Mean error [pixels]
38.219	-6.167	3.12089e-06	3.10288e-06	0	0.589996

Residuals



ID	Enabled	Pixel X	Pixel Y	Map X	Map Y	Res X (pixels)	Res Y (pixels)	Res Total (pixels)
0	yes	2296	-1210	38.226	-6.171	0.124224	-0.0291923	0.127608
1	yes	3577	-254	38.230	-6.168	-0.108443	0.0714589	0.12987
2	yes	4388	-2682	38.233	-6.176	0.532986	0.107494	0.543718
3	yes	3127	-1768	38.229	-6.173	-0.800323	-0.0906163	0.805436
4	yes	1060	-1876	38.223	-6.173	0.251556	-0.0591439	0.258415

Figure A1. Output statistics from the geocorrection procedure for the Google Earth satellite imagery used for study area A.

Transformation parameters (Linear)

Translation x	Translation y	Scale x	Scale y	Rotation [degrees]	Mean error [pixels]
35.667	-8.117	3.62618e-06	3.58718e-06	0	0.814185

Residuals

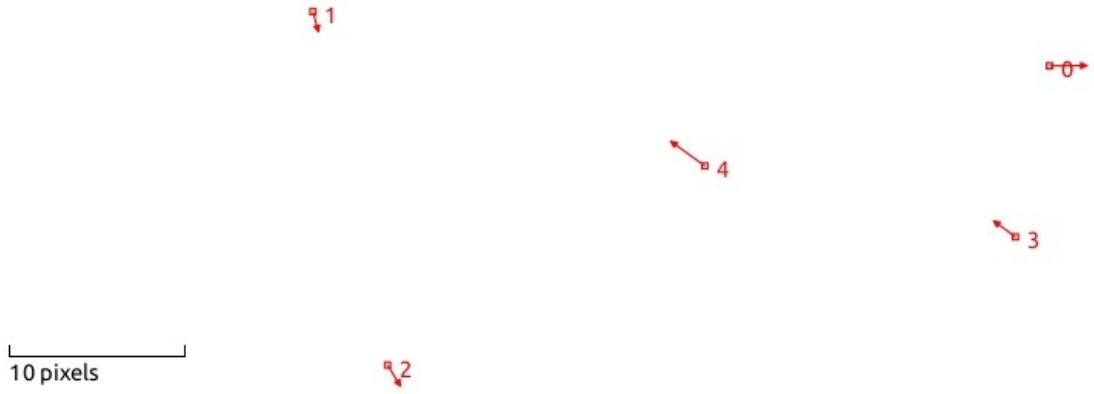
ID	Enabled	Pixel X	Pixel Y	Map X	Map Y	Res X (pixels)	Res Y (pixels)	Res Total (pixels)
0	yes	629	-380	35.669	-8.118	0.481804	-0.887512	1.00986
1	yes	1138	-1939	35.671	-8.124	-0.640835	-0.0341639	0.641745
2	yes	2745	-29	35.677	-8.117	-0.130859	0.679208	0.691699
3	yes	3922	-1356	35.681	-8.122	0.0930604	0.151199	0.177543
4	yes	2136	-1957	35.675	-8.124	0.196829	0.0912683	0.21696

Figure A2. Output statistics from the geocorrection procedure for the Google Earth satellite imagery used for study area B.

Transformation parameters (Linear)

Translation x	Translation y	Scale x	Scale y	Rotation [degrees]	Mean error [pixels]
38.369	-10.927	3.57132e-06	3.50132e-06	0	2.29261

Residuals



ID	Enabled	Pixel X	Pixel Y	Map X	Map Y	Res X (pixels)	Res Y (pixels)	Res Total (pixels)
0	yes	4287	-603	38.385	-10.930	2.12185	-0.0165408	2.12192
1	yes	1334	-224	38.374	-10.928	0.30697	1.12621	1.16729
2	yes	1633	-2696	38.375	-10.937	0.725345	1.18423	1.38872
3	yes	4150	-1800	38.384	-10.934	-1.22655	-0.898623	1.52051
4	yes	2905	-1304	38.380	-10.932	-1.92762	-1.39528	2.37961

Figure A3. Output statistics from the geocorrection procedure for the Google Earth satellite imagery used for study area C.

Appendix II. Error Matrices for the Road Network Accuracy Assessment

Table A1. Error matrix for the accuracy assessment of gROADS and OSM in sample area 1.

OSM	Satellite imagery		
	Road	**Non-road**	
Road	49	1	98%
Non-road	3	48	
	94%		96%

gROADS	Satellite imagery		
	Road	**Non-road**	
Road	1	2	33%
Non-road	51	47	
	2%		48%

Table A2. Error matrix for the accuracy assessment of gROADS and OSM in sample area 2.

OSM	Satellite imagery		
	Road	**Non-road**	
Road	50	1	98%
Non-road	2	97	
	96%		98%

gROADS	Satellite imagery		
	Road	**Non-road**	
Road	1	50	2%
Non-road	51	48	
	2%		33%

Table A3. Error matrix for the accuracy assessment of gROADS and OSM in sample area 3.

OSM	Satellite imagery		
	Road	**Non-road**	
Road	50	0	100%
Non-road	41	58	
	55%		72%

gROADS	Satellite imagery		
	Road	**Non-road**	
Road	23	27	46%
Non-road	68	31	
	25%		36%

Table A4. Error matrix for the accuracy assessment of gROADS and OSM in sample area 4.

OSM	Satellite imagery		
	Road	**Non-road**	
Road	35	16	69%
Non-road	13	85	
	73%		81%

gROADS	Satellite imagery		
	Road	**Non-road**	
Road	8	41	16%
Non-road	40	60	
	17%		46%

Table A5. Error matrix for the accuracy assessment of gROADS and OSM in sample area 5.

OSM	Satellite imagery		
	Road	**Non-road**	
Road	52	2	96%
Non-road	33	62	
	61%		77%

gROADS	Satellite imagery		
	Road	**Non-road**	
Road	24	27	47%
Non-road	61	37	
	28%		41%

Table A6. Error matrix for the accuracy assessment of gROADS and OSM in sample area 6.

OSM	Satellite imagery		
	Road	**Non-road**	
Road	32	20	62%
Non-road	10	38	
	76%		70%

gROADS	Satellite imagery		
	Road	**Non-road**	
Road	0	0	0%
Non-road	42	58	
	0%		0%

Table A7. Error matrix for the accuracy assessment of gROADS and OSM in sample area 7.

OSM	Satellite imagery		
	Road	**Non-road**	
Road	38	13	75%
Non-road	9	72	
	81%		83%

gROADS	Satellite imagery		
	Road	**Non-road**	
Road	0	32	0%
Non-road	47	53	
	0%		40%

Table A8. Error matrix for the accuracy assessment of gROADS and OSM in sample area 8.

OSM	Satellite imagery		
	Road	**Non-road**	
Road	49	1	98%
Non-road	8	89	
	86%		94%

gROADS	Satellite imagery		
	Road	**Non-road**	
Road	7	42	14%
Non-road	50	48	
	12%		37%

Table A9. Error matrix for the accuracy assessment of gROADS and OSM in sample area 9.

OSM	Satellite imagery			
		Road	**Non-road**	
	Road	51	0	100%
	Non-road	0	98	
		100%		100%

gROADS	Satellite imagery			
		Road	**Non-road**	
	Road	2	50	4%
	Non-road	49	48	
		4%		34%

Table A10. Error matrix for the accuracy assessment of gROADS and OSM in sample area 10.

OSM	Satellite imagery			
		Road	**Non-road**	
	Road	34	17	67%
	Non-road	32	66	
		52%		67%

gROADS	Satellite imagery			
		Road	**Non-road**	
	Road	12	37	24%
	Non-road	54	46	
		18%		39%

Table A11. Error matrix for the accuracy assessment of gROADS and OSM in sample area 11.

OSM	Satellite imagery			
		Road	**Non-road**	
	Road	40	10	80%
	Non-road	12	86	
		77%		85%

gROADS	Satellite imagery			
		Road	**Non-road**	
	Road	4	44	8%
	Non-road	48	52	
		8%		38%

Table A12. Error matrix for the accuracy assessment of gROADS and OSM in sample area 12.

OSM	Satellite imagery			
		Road	**Non-road**	
	Road	34	17	67%
	Non-road	18	74	
		65%		76%

gROADS	Satellite imagery			
		Road	**Non-road**	
	Road	2	41	5%
	Non-road	50	50	
		4%		36%

Table A13. Error matrix for the accuracy assessment of gROADS and OSM in sample area 13.

OSM	Satellite imagery			
		Road	**Non-road**	
	Road	0	0	0%
	Non-road	10	40	
		0%		0%

gROADS	Satellite imagery			
		Road	**Non-road**	
	Road	0	0	0%
	Non-road	10	40	
		0%		0%

Table A14. Error matrix for the accuracy assessment of gROADS and OSM in sample area 14.

OSM	Satellite imagery			
		Road	**Non-road**	
	Road	47	8	85%
	Non-road	17	78	
		73%		83%

gROADS	Satellite imagery			
		Road	**Non-road**	
	Road	11	45	20%
	Non-road	53	41	
		17%		35%

Table A15. Error matrix for the accuracy assessment of gROADS and OSM in sample area 15.

OSM	Satellite imagery			
		Road	**Non-road**	
	Road	42	8	84%
	Non-road	9	91	
		82%		89%

gROADS	Satellite imagery			
		Road	**Non-road**	
	Road	3	48	6%
	Non-road	48	51	
		6%		36%

Table A16. Error matrix for the accuracy assessment of gROADS and OSM in sample area 16.

OSM	Satellite imagery			
		Road	**Non-road**	
	Road	45	4	92%
	Non-road	18	32	
		71%		78%

gROADS	Satellite imagery			
		Road	**Non-road**	
	Road	0	0	0%
	Non-road	63	36	
		0%		0%

Table A17. Error matrix for the accuracy assessment of gROADS and OSM in sample area 17.

OSM	Satellite imagery			
		Road	**Non-road**	
	Road	53	3	95%
	Non-road	21	73	
		72%		84%

gROADS	Satellite imagery			
		Road	**Non-road**	
	Road	14	38	27%
	Non-road	60	38	
		19%		35%

Table A18. Error matrix for the accuracy assessment of gROADS and OSM in sample area 18.

OSM	Satellite imagery			
		Road	**Non-road**	
	Road	44	14	76%
	Non-road	20	71	
		69%		77%

gROADS	Satellite imagery			
		Road	**Non-road**	
	Road	13	38	25%
	Non-road	51	47	
		20%		40%

Table A19. Error matrix for the accuracy assessment of gROADS and OSM in sample area 19.

OSM	Satellite imagery			
		Road	**Non-road**	
	Road	52	0	100%
	Non-road	2	45	
		96%		98%

gROADS	Satellite imagery			
		Road	**Non-road**	
	Road	0	0	0%
	Non-road	54	45	
		0%		0%

Table A20. Error matrix for the accuracy assessment of gROADS and OSM in sample area 20.

OSM	Satellite imagery			
		Road	**Non-road**	
	Road	48	2	96%
	Non-road	31	69	
		61%		78%

gROADS	Satellite imagery			
		Road	**Non-road**	
	Road	12	41	23%
	Non-road	67	30	
		15%		28%

Table A21. Error matrix for the accuracy assessment of gROADS and OSM in study area A.

OSM	Satellite imagery			
		Road	**Non-road**	
	Road	0	0	0%
	Non-road	24	26	
		0%		0%

gROADS	Satellite imagery			
		Road	**Non-road**	
	Road	0	0	0%
	Non-road	24	26	
		0%		0%

Table A22. Error matrix for the accuracy assessment of gROADS and OSM in study area B,

OSM	Satellite imagery			
		Road	**Non-road**	
	Road	41	9	82%
	Non-road	23	53	
		64%		75%

gROADS	Satellite imagery			
		Road	**Non-road**	
	Road	8	20	29%
	Non-road	56	42	
		13%		40%

Table A23. Error matrix for the accuracy assessment of gROADS and OSM in study area C,

OSM	Satellite imagery			
		Road	**Non-road**	
	Road	44	9	83%
	Non-road	20	77	
		69%		81%

gROADS	Satellite imagery			
		Road	**Non-road**	
	Road	4	46	8%
	Non-road	60	40	
		6%		29%

Table A24. Overall error matrix for the accuracy assessment of gROADS and OSM in all areas.

OSM	Satellite imagery			
		Road	**Non-road**	
	Road	930	155	86%
	Non-road	376	1528	
		71%		82%

gROADS	Satellite imagery			
		Road	**Non-road**	
	Road	149	669	18%
	Non-road	1157	1014	
		11%		39%

Appendix III. Matlab Code for Road Extraction from Google Earth Imagery

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Title: Unpaved Road Extraction from Satellite Imagery I
% Description: Extracts an unpaved road network from three Google Earth
% satellite images.
% Author: Corey Nicholas Ragosnig
% Last modified: 2018/05/03



%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Import and Set-Up
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%Import Google Earth image data. Note these are multiband images so they
%are imported into MatLab as a 3D array. Order: 123 = RGB.
[sa2, R_sa2] = geotiffread('ge_sa2_2_georef.tif');
[sa3, R_sa3] = geotiffread('ge_sa3_2_georef.tif');
[sa4, R_sa4] = geotiffread('ge_sa4_2_georef.tif');

%Extract array size.
[rows2, cols2, ~] = size(sa2);
[rows3, cols3, ~] = size(sa3);
[rows4, cols4, ~] = size(sa4);

%Create empty arrays to store HSV and YCbCr information.
hsy2 = zeros(rows2, cols2, 3);
hsy3 = zeros(rows3, cols3, 3);
hsy4 = zeros(rows4, cols4, 3);

%Create empty arrays for binary road mask.
roads2 = uint8(zeros(rows2, cols2));
roads3 = uint8(zeros(rows3, cols3));
roads4 = uint8(zeros(rows4, cols4));



%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Part 1: Image Processing and Conversion
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%Perform adaptive histogram equalisation.
lab2 = rgb2lab(sa2);
l2 = lab2(:,:,1)/100;
l2 = adapthisteq(l2);
lab2(:,:,1) = l2*100;
sa2_eq = lab2rgb(lab2);

lab3 = rgb2lab(sa3);
l3 = lab3(:,:,1)/100;
l3 = adapthisteq(l3);
lab3(:,:,1) = l3*100;
sa3_eq = lab2rgb(lab3);

lab4 = rgb2lab(sa4);
l4 = lab4(:,:,1)/100;
l4 = adapthisteq(l4);
lab4(:,:,1) = l4*100;
sa4_eq = lab2rgb(lab4);

%Export images after adapted histogram equalisation.
geotiffwrite('sa2_eq.tif', sa2_eq, R_sa2,'CoordRefSysCode', 4326)
geotiffwrite('sa3_eq.tif', sa3_eq, R_sa3,'CoordRefSysCode', 4326)
geotiffwrite('sa4_eq.tif', sa4_eq, R_sa4,'CoordRefSysCode', 4326)

%Calculate values for HSV and YCbCr.
%HSV:
hsv2 = rgb2hsv(sa2_eq);
```

```
hsv3 = rgb2hsv(sa3_eq);
hsv4 = rgb2hsv(sa4_eq);

%YCbCr:
ycbcr2 = rgb2ycbcr(sa2_eq);
ycbcr3 = rgb2ycbcr(sa3_eq);
ycbcr4 = rgb2ycbcr(sa4_eq);

%Combine HS and Y values into one stacked array for each study area.
for r = 1:rows2
    for c = 1:cols2

        hsy2(r,c,1) = hsv2(r,c,1);
        hsy2(r,c,2) = hsv2(r,c,2);
        hsy2(r,c,3) = ycbcr2(r,c,1);

    end
end

for r = 1:rows3
    for c = 1:cols3

        hsy3(r,c,1) = hsv3(r,c,1);
        hsy3(r,c,2) = hsv3(r,c,2);
        hsy3(r,c,3) = ycbcr3(r,c,1);

    end
end

for r = 1:rows4
    for c = 1:cols4

        hsy4(r,c,1) = hsv4(r,c,1);
        hsy4(r,c,2) = hsv4(r,c,2);
        hsy4(r,c,3) = ycbcr4(r,c,1);

    end
end

%Export HSY images for threshold analysis.
geotiffwrite('hsy2.tif', hsy2, R_sa2,'CoordRefSysCode', 4326)
geotiffwrite('hsy3.tif', hsy3, R_sa3,'CoordRefSysCode', 4326)
geotiffwrite('hsy4.tif', hsy4, R_sa4,'CoordRefSysCode', 4326)

geotiffwrite('hsv2.tif', hsv2, R_sa2,'CoordRefSysCode', 4326)
geotiffwrite('hsv3.tif', hsv3, R_sa3,'CoordRefSysCode', 4326)
geotiffwrite('hsv4.tif', hsv4, R_sa4,'CoordRefSysCode', 4326)

geotiffwrite('ycbcr2.tif', ycbcr2, R_sa2,'CoordRefSysCode', 4326)
geotiffwrite('ycbcr3.tif', ycbcr3, R_sa3,'CoordRefSysCode', 4326)
geotiffwrite('ycbcr4.tif', ycbcr4, R_sa4,'CoordRefSysCode', 4326)



%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Part 2: Selection of Candidate Road Pixels
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%Set required threshold values.
tminH2 = 0.00839;
tminH3 = 0.00026;   %Minimum hue
tminH4 = 0.00010;

tmaxH2 = 0.17647;
tmaxH3 = 0.13081;   %Maximum hue
tmaxH4 = 0.13717;

tminS2 = 0.03716;
tminS3 = 0.08000;   %Minimum saturation
tminS4 = 0.11744;
```

```
tmaxS2 = 0.28097;
tmaxS3 = 0.28000;   %Maximum saturation
tmaxS4 = 0.41919;

tminY2 = 0.77000;
tminY3 = 0.85000;   %Minimum luminance
tminY4 = 0.85000;

tmaxY2 = 1.00000;
tmaxY3 = 1.00000;   %Maximum luminance
tmaxY4 = 1.00000;

%Logical combination of thresholds to create road mask.
%Study Area 1
for r = 1:rows2
    for c = 1:cols2

        if hsy2(r,c,1) >= tminH2 && hsy2(r,c,1) <= tmaxH2
            hue = 1;
        else
            hue = 0;
        end

        if hsy2(r,c,2) >= tminS2 && hsy2(r,c,2) <= tmaxS2
            sat = 1;
        else
            sat = 0;
        end

        if hsy2(r,c,3) >= tminY2 && hsy2(r,c,3) <= tmaxY2
            lum = 1;
        else
            lum = 0;
        end

        if hue + sat + lum == 3
            roads2(r,c) = 1;
        else
            roads2(r,c) = 0;
        end

    end
end

%Study Area 2
for r = 1:rows3
    for c = 1:cols3

        if hsy3(r,c,1) >= tminH3 && hsy3(r,c,1) <= tmaxH3
            hue = 1;
        else
            hue = 0;
        end

        if hsy3(r,c,2) >= tminS3 && hsy3(r,c,2) <= tmaxS3
            sat = 1;
        else
            sat = 0;
        end

        if hsy3(r,c,3) >= tminY3 && hsy3(r,c,3) <= tmaxY3
            lum = 1;
        else
            lum = 0;
        end

        if hue + sat + lum == 3
            roads3(r,c) = 1;
        else
```

```
            roads3(r,c) = 0;
        end

    end
end

%Study Area 3
for r = 1:rows4
    for c = 1:cols4

        if hsy4(r,c,1) >= tminH4 && hsy4(r,c,1) <= tmaxH4
            hue = 1;
        else
            hue = 0;
        end

        if hsy4(r,c,2) >= tminS4 && hsy4(r,c,2) <= tmaxS4
            sat = 1;
        else
            sat = 0;
        end

        if hsy4(r,c,3) >= tminY4 && hsy4(r,c,3) <= tmaxY4
            lum = 1;
        else
            lum = 0;
        end

        if hue + sat + lum == 3
            roads4(r,c) = 1;
        else
            roads4(r,c) = 0;
        end

    end
end



%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Part 3: Post-Selection Processing
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%Filter roads layer using shape of objects.
roads2_a = bwpropfilt(logical(roads2), 'Eccentricity', [0.95 1], 4);
roads2_b = bwpropfilt(logical(roads2), 'Solidity', [0 0.33], 4);
roads2_1 = or(roads2_a, roads2_b);
geotiffwrite('roads2_1.tif', roads2_1, R_sa2,'CoordRefSysCode', 4326)

roads3_a = bwpropfilt(logical(roads3), 'Eccentricity', [0.95 1], 4);
roads3_b = bwpropfilt(logical(roads3), 'Solidity', [0 0.33], 4);
roads3_1 = or(roads3_a, roads3_b);
geotiffwrite('roads3_1.tif', roads3_1, R_sa3,'CoordRefSysCode', 4326)

roads4_a = bwpropfilt(logical(roads4), 'Eccentricity', [0.95 1], 4);
roads4_b = bwpropfilt(logical(roads4), 'Solidity', [0 0.33], 4);
roads4_1 = or(roads4_a, roads4_b);
geotiffwrite('roads4_1.tif', roads4_1, R_sa4,'CoordRefSysCode', 4326)


%Export road masks.
geotiffwrite('roads2sk.tif', roads2sk, R_sa2,'CoordRefSysCode', 4326)
geotiffwrite('roads3sk.tif', roads3sk, R_sa3,'CoordRefSysCode', 4326)
geotiffwrite('roads4sk.tif', roads4sk, R_sa4,'CoordRefSysCode', 4326)
```

Appendix IV. Matlab Code for Road Extraction from Sentinel-2 Imagery

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Title: Unpaved Road Extraction from Satellite Imagery II
% Description: Extracts an unpaved road network from three Sentinel-2
% satellite images.
% Author: Corey Nicholas Ragosnig
% Last modified: 2018/05/03



%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Preprocessing
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%Import Sentinel band data
[sa2, R_sa2] = geotiffread('se_sa2.tif');
[sa3, R_sa3] = geotiffread('se_sa3.tif');
[sa4, R_sa4] = geotiffread('se_sa4.tif');

%Convert to double data type.
sa2 = double(sa2);
sa3 = double(sa3);
sa4 = double(sa4);

%Set values for spectral response of roads:
meanR2 = [1735, 2177, 3483];
meanR3 = [1900, 2850, 4600];
meanR4 = [2138, 2768, 3505];

%Set value for the size of the image:
[rows2, cols2, ~] = size(sa2);
[rows3, cols3, ~] = size(sa3);
[rows4, cols4, ~] = size(sa4);



%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Part 1: Oriented Gaussian Filters and Error Calculations
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%Input to define filters:
    % n = size of filter (5x5)
    % a = angle of filter orientation in radians (0, pi/4, pi/2, 3*pi/4)
    % sX = standard deviation in x-direction
    % sY = standard deviation in y-direction

fsize = 5; %filter size
m = 5; %distance from road to surroundings

%Calculate Oriented Asymmetric Gaussian filter.
h0 = calculateOAG(fsize,0,1/4,1);
h45 = calculateOAG(fsize,pi/4,1/4,1);
h90 = calculateOAG(fsize,pi/2,1/4,1);
h135 = calculateOAG(fsize,3*pi/4,1/4,1);

%Study area 1
    %Apply filter to each band.
    red2 = sa2(:,:,1);
    nir2 = sa2(:,:,2);
    swir2 = sa2(:,:,3);

        %Band 4:
        red2_0 = applyOAG(red2,h0,fsize);
        red2_45 = applyOAG(red2,h45,fsize);
        red2_90 = applyOAG(red2,h90,fsize);
        red2_135 = applyOAG(red2,h135,fsize);

        %Band 8:
```

```
        nir2_0 = applyOAG(nir2,h0,fsize);
        nir2_45 = applyOAG(nir2,h45,fsize);
        nir2_90 = applyOAG(nir2,h90,fsize);
        nir2_135 = applyOAG(nir2,h135,fsize);

        %Band 11:
        swir2_0 = applyOAG(swir2,h0,fsize);
        swir2_45 = applyOAG(swir2,h45,fsize);
        swir2_90 = applyOAG(swir2,h90,fsize);
        swir2_135 = applyOAG(swir2,h135,fsize);

    %Create error matrices for each orientation.
        %For orientation 0:
        %Initialise error matrix.
        error2_0 = zeros(rows2,cols2);
        alpha2_0 = zeros(rows2,cols2);
        pR2_0 = zeros(rows2,cols2);
        pS2_0 = zeros(rows2,cols2);

        for r = (m+1):rows2-(m+1)
            for c = (m+1):cols2-(m+1)

                %Initialise vectors.
                pR = [red2_0(r,c), nir2_0(r,c), swir2_0(r,c)];
                pSa = [red2_0(r,c-m), nir2_0(r,c-m), swir2_0(r,c-m)];
                pSb = [red2_0(r,c+m), nir2_0(r,c+m), swir2_0(r,c+m)];

                %Calculate difference with dirt roads.
                pR_dif = pR - meanR2;
                pSa_dif = pSa - meanR2;
                pSb_dif = pSb - meanR2;

                %Calculate size of pixel response.
                pR2_0(r,c) = norm(pR);

                %Calculate error values for each side of the road.
                [error2_0(r,c), alpha2_0(r,c), pS2_0(r,c)] = calculateError(pR_dif, pSa_dif,
pSb_dif);

            end
        end

        %For orientation 45:
        %Initialise error matrix.
        error2_45 = zeros(rows2,cols2);
        alpha2_45 = zeros(rows2,cols2);
        pR2_45 = zeros(rows2,cols2);
        pS2_45 = zeros(rows2,cols2);

        for r = (m+1):rows2-(m+1)
            for c = (m+1):cols2-(m+1)

                %Initialise vectors.
                pR = [red2_45(r,c), nir2_45(r,c), swir2_45(r,c)];
                pSa = [red2_45(r-m,c-m), nir2_45(r-m,c-m), swir2_45(r-m,c-m)];
                pSb = [red2_45(r+m,c+m), nir2_45(r+m,c+m), swir2_45(r+m,c+m)];

                %Calculate difference with dirt roads.
                pR_dif = pR - meanR2;
                pSa_dif = pSa - meanR2;
                pSb_dif = pSb - meanR2;

                %Calculate size of pixel response.
                pR2_45(r,c) = norm(pR);

                %Calculate error values for each side of the road.
                [error2_45(r,c), alpha2_45(r,c), pS2_45(r,c)] = calculateError(pR_dif,
pSa_dif, pSb_dif);

            end
```

```
        end

        %For orientation 90:
        %Initialise error matrix.
        error2_90 = zeros(rows2,cols2);
        alpha2_90 = zeros(rows2,cols2);
        pR2_90 = zeros(rows2,cols2);
        pS2_90 = zeros(rows2,cols2);

        for r = (m+1):rows2-(m+1)
            for c = (m+1):cols2-(m+1)

                %Initialise vectors.
                pR = [red2_90(r,c), nir2_90(r,c), swir2_90(r,c)];
                pSa = [red2_90(r-m,c), nir2_90(r-m,c), swir2_90(r-m,c)];
                pSb = [red2_90(r+m,c), nir2_90(r+m,c), swir2_90(r+m,c)];

                %Calculate difference with dirt roads.
                pR_dif = pR - meanR2;
                pSa_dif = pSa - meanR2;
                pSb_dif = pSb - meanR2;

                %Calculate size of pixel response.
                pR2_90(r,c) = norm(pR);

                %Calculate error values for each side of the road.
                [error2_90(r,c), alpha2_90(r,c), pS2_90(r,c)] = calculateError(pR_dif,
pSa_dif, pSb_dif);

            end
        end

        %For orientation 135:
        %Initialise error matrix.
        error2_135 = zeros(rows2,cols2);
        alpha2_135 = zeros(rows2,cols2);
        pR2_135 = zeros(rows2,cols2);
        pS2_135 = zeros(rows2,cols2);

        for r = (m+1):rows2-(m+1)
            for c = (m+1):cols2-(m+1)

                %Initialise vectors.
                pR = [red2_135(r,c), nir2_135(r,c), swir2_135(r,c)];
                pSa = [red2_135(r-m,c+m), nir2_135(r-m,c+m), swir2_135(r-m,c+m)];
                pSb = [red2_135(r+m,c-m), nir2_135(r+m,c-m), swir2_135(r+m,c-m)];

                %Calculate difference with dirt roads.
                pR_dif = pR - meanR2;
                pSa_dif = pSa - meanR2;
                pSb_dif = pSb - meanR2;

                %Calculate size of pixel response.
                pR2_135(r,c) = norm(pR);

                %Calculate error values for each side of the road.
                [error2_135(r,c), alpha2_135(r,c), pS2_135(r,c)] = calculateError(pR_dif,
pSa_dif, pSb_dif);

            end
        end

    errorM2 = zeros(rows2,cols2);
    alphaM2 = zeros(rows2,cols2);
    pRM2 = zeros(rows2,cols2);
    pSM2 = zeros(rows2,cols2);

    for r = 1:rows2
        for c = 1:cols2
```

```
            [errorM2(r,c), index] = min([error2_0(r,c), error2_45(r,c), error2_90(r,c),
error2_135(r,c)]);
            alist = [alpha2_0(r,c), alpha2_45(r,c), alpha2_90(r,c), alpha2_135(r,c)];
            pSlist = [pS2_0(r,c), pS2_45(r,c), pS2_90(r,c), pS2_135(r,c)];
            pRlist = [pR2_0(r,c), pR2_45(r,c), pR2_90(r,c), pR2_135(r,c)];
            alphaM2(r,c) = alist(index);
            pSM2(r,c) = pSlist(index);
            pRM2(r,c) = pRlist(index);
        end
    end

    %Save error matrix as a GeoTiff.
    geotiffwrite('errorM2.tif', errorM2, R_sa2,'CoordRefSysCode', 32737)
    geotiffwrite('alphaM2.tif', alphaM2, R_sa2,'CoordRefSysCode', 32737)


%Study area 2
    %Apply filter to each band.
    red3 = sa3(:,:,1);
    nir3 = sa3(:,:,2);
    swir3 = sa3(:,:,3);

        %Band 4:
        red3_0 = applyOAG(red3,h0,fsize);
        red3_45 = applyOAG(red3,h45,fsize);
        red3_90 = applyOAG(red3,h90,fsize);
        red3_135 = applyOAG(red3,h135,fsize);

        %Band 8:
        nir3_0 = applyOAG(nir3,h0,fsize);
        nir3_45 = applyOAG(nir3,h45,fsize);
        nir3_90 = applyOAG(nir3,h90,fsize);
        nir3_135 = applyOAG(nir3,h135,fsize);

        %Band 11:
        swir3_0 = applyOAG(swir3,h0,fsize);
        swir3_45 = applyOAG(swir3,h45,fsize);
        swir3_90 = applyOAG(swir3,h90,fsize);
        swir3_135 = applyOAG(swir3,h135,fsize);

    %Create error matrices for each orientation.
        %For orientation 0:
        %Initialise error matrix.
        error3_0 = zeros(rows3,cols3);
        alpha3_0 = zeros(rows3,cols3);
        pR3_0 = zeros(rows3,cols3);
        pS3_0 = zeros(rows3,cols3);

        for r = (m+1):rows3-(m+1)
            for c = (m+1):cols3-(m+1)

                %Initialise vectors.
                pR = [red3_0(r,c), nir3_0(r,c), swir3_0(r,c)];
                pSa = [red3_0(r,c-m), nir3_0(r,c-m), swir3_0(r,c-m)];
                pSb = [red3_0(r,c+m), nir3_0(r,c+m), swir3_0(r,c+m)];

                %Calculate difference with dirt roads.
                pR_dif = pR - meanR3;
                pSa_dif = pSa - meanR3;
                pSb_dif = pSb - meanR3;

                %Calculate size of pixel response.
                pR3_0(r,c) = sqrt(dot(pR,pR));

                %Calculate error values for each side of the road.
                [error3_0(r,c), alpha3_0(r,c), pS3_0(r,c)] = calculateError(pR_dif, pSa_dif,
pSb_dif);

            end
        end
```

```
        %For orientation 45:
        %Initialise error matrix.
        error3_45 = zeros(rows3,cols3);
        alpha3_45 = zeros(rows3,cols3);
        pR3_45 = zeros(rows3,cols3);
        pS3_45 = zeros(rows3,cols3);

        for r = (m+1):rows3-(m+1)
            for c = (m+1):cols3-(m+1)

                %Initialise vectors.
                pR = [red3_45(r,c), nir3_45(r,c), swir3_45(r,c)];
                pSa = [red3_45(r-m,c-m), nir3_45(r-m,c-m), swir3_45(r-m,c-m)];
                pSb = [red3_45(r+m,c+m), nir3_45(r+m,c+m), swir3_45(r+m,c+m)];

                %Calculate difference with dirt roads.
                pR_dif = pR - meanR3;
                pSa_dif = pSa - meanR3;
                pSb_dif = pSb - meanR3;

                %Calculate size of pixel response.
                pR3_45(r,c) = sqrt(dot(pR,pR));

                %Calculate error values for each side of the road.
                [error3_45(r,c), alpha3_45(r,c), pS3_45(r,c)] = calculateError(pR_dif,
pSa_dif, pSb_dif);

            end
        end

        %For orientation 90:
        %Initialise error matrix.
        error3_90 = zeros(rows3,cols3);
        alpha3_90 = zeros(rows3,cols3);
        pR3_90 = zeros(rows3,cols3);
        pS3_90 = zeros(rows3,cols3);

        for r = (m+1):rows3-(m+1)
            for c = (m+1):cols3-(m+1)

                %Initialise vectors.
                pR = [red3_90(r,c), nir3_90(r,c), swir3_90(r,c)];
                pSa = [red3_90(r-m,c), nir3_90(r-m,c), swir3_90(r-m,c)];
                pSb = [red3_90(r+m,c), nir3_90(r+m,c), swir3_90(r+m,c)];

                %Calculate difference with dirt roads.
                pR_dif = pR - meanR3;
                pSa_dif = pSa - meanR3;
                pSb_dif = pSb - meanR3;

                %Calculate size of pixel response.
                pR3_90(r,c) = sqrt(dot(pR,pR));

                %Calculate error values for each side of the road.
                [error3_90(r,c), alpha3_90(r,c), pS3_90(r,c)] = calculateError(pR_dif,
pSa_dif, pSb_dif);

            end
        end

        %For orientation 135:
        %Initialise error matrix.
        error3_135 = zeros(rows3,cols3);
        alpha3_135 = zeros(rows3,cols3);
        pR3_135 = zeros(rows3,cols3);
        pS3_135 = zeros(rows3,cols3);

        for r = (m+1):rows3-(m+1)
            for c = (m+1):cols3-(m+1)
```

```
                %Initialise vectors.
                pR = [red3_135(r,c), nir3_135(r,c), swir3_135(r,c)];
                pSa = [red3_135(r-m,c+m), nir3_135(r-m,c+m), swir3_135(r-m,c+m)];
                pSb = [red3_135(r+m,c-m), nir3_135(r+m,c-m), swir3_135(r+m,c-m)];

                %Calculate difference with dirt roads.
                pR_dif = pR - meanR3;
                pSa_dif = pSa - meanR3;
                pSb_dif = pSb - meanR3;

                %Calculate size of pixel response.
                pR3_135(r,c) = sqrt(dot(pR,pR));

                %Calculate error values for each side of the road.
                [error3_135(r,c), alpha3_135(r,c), pS3_135(r,c)] = calculateError(pR_dif,
pSa_dif, pSb_dif);

            end
        end

    errorM3 = zeros(rows3,cols3);
    alphaM3 = zeros(rows3,cols3);
    pRM3 = zeros(rows3,cols3);
    pSM3 = zeros(rows3,cols3);

    for r = 1:rows3
        for c = 1:cols3
            [errorM3(r,c), index] = min([error3_0(r,c), error3_45(r,c), error3_90(r,c),
error3_135(r,c)]);
            alist = [alpha3_0(r,c), alpha3_45(r,c), alpha3_90(r,c), alpha3_135(r,c)];
            pSlist = [pS3_0(r,c), pS3_45(r,c), pS3_90(r,c), pS3_135(r,c)];
            pRlist = [pR3_0(r,c), pR3_45(r,c), pR3_90(r,c), pR3_135(r,c)];
            alphaM3(r,c) = alist(index);
            pSM3(r,c) = pSlist(index);
            pRM3(r,c) = pRlist(index);
        end
    end

%Save error matrix as a GeoTiff.
geotiffwrite('errorM3.tif', errorM3, R_sa3,'CoordRefSysCode', 32736)
geotiffwrite('alphaM3.tif', alphaM3, R_sa3,'CoordRefSysCode', 32736)



%Study area 3
    %Apply filter to each band.
    red4 = sa4(:,:,1);
    nir4 = sa4(:,:,2);
    swir4 = sa4(:,:,3);

        %Band 4:
        red4_0 = applyOAG(red4,h0,fsize);
        red4_45 = applyOAG(red4,h45,fsize);
        red4_90 = applyOAG(red4,h90,fsize);
        red4_135 = applyOAG(red4,h135,fsize);

        %Band 8:
        nir4_0 = applyOAG(nir4,h0,fsize);
        nir4_45 = applyOAG(nir4,h45,fsize);
        nir4_90 = applyOAG(nir4,h90,fsize);
        nir4_135 = applyOAG(nir4,h135,fsize);

        %Band 11:
        swir4_0 = applyOAG(swir4,h0,fsize);
        swir4_45 = applyOAG(swir4,h45,fsize);
        swir4_90 = applyOAG(swir4,h90,fsize);
        swir4_135 = applyOAG(swir4,h135,fsize);

    %Create error matrices for each orientation.
```

```
        %For orientation 0:
        %Initialise error matrix.
        error4_0 = zeros(rows4,cols4);
        alpha4_0 = zeros(rows4,cols4);
        pR4_0 = zeros(rows4,cols4);
        pS4_0 = zeros(rows4,cols4);

        for r = (m+1):rows4-(m+1)
            for c = (m+1):cols4-(m+1)

                %Initialise vectors.
                pR = [red4_0(r,c), nir4_0(r,c), swir4_0(r,c)];
                pSa = [red4_0(r,c-m), nir4_0(r,c-m), swir4_0(r,c-m)];
                pSb = [red4_0(r,c+m), nir4_0(r,c+m), swir4_0(r,c+m)];

                %Calculate difference with dirt roads.
                pR_dif = pR - meanR4;
                pSa_dif = pSa - meanR4;
                pSb_dif = pSb - meanR4;

                %Calculate size of pixel response.
                pR4_0(r,c) = norm(pR);

                %Calculate error values for each side of the road.
                [error4_0(r,c), alpha4_0(r,c), pS4_0(r,c)] = calculateError(pR_dif, pSa_dif,
pSb_dif);

            end
        end

        %For orientation 45:
        %Initialise error matrix.
        error4_45 = zeros(rows4,cols4);
        alpha4_45 = zeros(rows4,cols4);
        pR4_45 = zeros(rows4,cols4);
        pS4_45 = zeros(rows4,cols4);

        for r = (m+1):rows4-(m+1)
            for c = (m+1):cols4-(m+1)

                %Initialise vectors.
                pR = [red4_45(r,c), nir4_45(r,c), swir4_45(r,c)];
                pSa = [red4_45(r-m,c-m), nir4_45(r-m,c-m), swir4_45(r-m,c-m)];
                pSb = [red4_45(r+m,c+m), nir4_45(r+m,c+m), swir4_45(r+m,c+m)];

                %Calculate difference with dirt roads.
                pR_dif = pR - meanR4;
                pSa_dif = pSa - meanR4;
                pSb_dif = pSb - meanR4;

                %Calculate size of pixel response.
                pR4_45(r,c) = norm(pR);

                %Calculate error values for each side of the road.
                [error4_45(r,c), alpha4_45(r,c), pS4_45(r,c)] = calculateError(pR_dif,
pSa_dif, pSb_dif);

            end
        end

        %For orientation 90:
        %Initialise error matrix.
        error4_90 = zeros(rows4,cols4);
        alpha4_90 = zeros(rows4,cols4);
        pR4_90 = zeros(rows4,cols4);
        pS4_90 = zeros(rows4,cols4);

        for r = (m+1):rows4-(m+1)
            for c = (m+1):cols4-(m+1)
```

```
                %Initialise vectors.
                pR = [red4_90(r,c), nir4_90(r,c), swir4_90(r,c)];
                pSa = [red4_90(r-m,c), nir4_90(r-m,c), swir4_90(r-m,c)];
                pSb = [red4_90(r+m,c), nir4_90(r+m,c), swir4_90(r+m,c)];

                %Calculate difference with dirt roads.
                pR_dif = pR - meanR4;
                pSa_dif = pSa - meanR4;
                pSb_dif = pSb - meanR4;

                %Calculate size of pixel response.
                pR4_90(r,c) = norm(pR);

                %Calculate error values for each side of the road.
                [error4_90(r,c), alpha4_90(r,c), pS4_90(r,c)] = calculateError(pR_dif,
pSa_dif, pSb_dif);

            end
        end

        %For orientation 135:
        %Initialise error matrix.
        error4_135 = zeros(rows4,cols4);
        alpha4_135 = zeros(rows4,cols4);
        pR4_135 = zeros(rows4,cols4);
        pS4_135 = zeros(rows4,cols4);

        for r = (m+1):rows4-(m+1)
            for c = (m+1):cols4-(m+1)

                %Initialise vectors.
                pR = [red4_135(r,c), nir4_135(r,c), swir4_135(r,c)];
                pSa = [red4_135(r-m,c+m), nir4_135(r-m,c+m), swir4_135(r-m,c+m)];
                pSb = [red4_135(r+m,c-m), nir4_135(r+m,c-m), swir4_135(r+m,c-m)];

                %Calculate difference with dirt roads.
                pR_dif = pR - meanR4;
                pSa_dif = pSa - meanR4;
                pSb_dif = pSb - meanR4;

                %Calculate size of pixel response.
                pR4_135(r,c) = norm(pR);

                %Calculate error values for each side of the road.
                [error4_135(r,c), alpha4_135(r,c), pS4_135(r,c)] = calculateError(pR_dif,
pSa_dif, pSb_dif);

            end
        end

    errorM4 = zeros(rows4,cols4);
    alphaM4 = zeros(rows4,cols4);
    pRM4 = zeros(rows4,cols4);
    pSM4 = zeros(rows4,cols4);

    for r = 1:rows4
        for c = 1:cols4
            [errorM4(r,c), index] = min([error4_0(r,c), error4_45(r,c), error4_90(r,c),
error4_135(r,c)]);
            alist = [alpha4_0(r,c), alpha4_45(r,c), alpha4_90(r,c), alpha4_135(r,c)];
            pSlist = [pS4_0(r,c), pS4_45(r,c), pS4_90(r,c), pS4_135(r,c)];
            pRlist = [pR4_0(r,c), pR4_45(r,c), pR4_90(r,c), pR4_135(r,c)];
            alphaM4(r,c) = alist(index);
            pSM4(r,c) = pSlist(index);
            pRM4(r,c) = pRlist(index);
        end
    end

    %Save error matrix as a GeoTiff.
    geotiffwrite('errorM4.tif', errorM4, R_sa4,'CoordRefSysCode', 32737)
```

```
    geotiffwrite('alphaM4.tif', alphaM4, R_sa4,'CoordRefSysCode', 32737)



%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Part 2:
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%Study area 1.
    %Calculate gradient values.
    [Gmag2, Gdir2] = imgradient(errorM2);

    %Filter based on gradient and gradient direction.
    for r = 2:rows2-1
        for c = 2:cols2-1

            g = Gdir2(r,c);
            e0 = errorM2(r,c);
            a = alphaM2(r,c);

            %Determine if error is a minimum, dependent on the direction of
            %gradient of the pixel.
            if (g > 22.5 && g < 67.5) || (g > -157.5 && g < -112.5) %NE/SW diagonal

                %Extract adjacent error values.
                e1 = errorM2(r-1,c+1);
                e2 = errorM2(r+1,c-1);

            elseif (g >= 67.5 && g <= 112.5) || (g >= -112.5 && g <= -67.5) %NS vertical

                %Extract adjacent error values.
                e1 = errorM2(r-1,c);
                e2 = errorM2(r+1,c);

            elseif (g > 112.5 && g < 157.5) || (g > -67.5 && g < -22.5) %NW/SE diagonal

                %Extract adjacent error values.
                e1 = errorM2(r-1,c-1);
                e2 = errorM2(r+1,c+1);

            elseif (g >= 157.5 || g <= -157.5) || (g >= -22.5 && g <= 22.5) %EW horizontal

                %Extract adjacent error values.
                e1 = errorM2(r,c-1);
                e2 = errorM2(r,c+1);
            end

            %Pixel is likely to be a road only if it is a local minimum.
            if (e0 <= e1 && e0 <= e2) && (a >=0 && a < 1)
                roadCandidates2(r,c) = 1;
            else
                roadCandidates2(r,c) = 0;
            end

        end
    end


%Study area 2.
    %Calculate gradient values.
    [Gmag3, Gdir3] = imgradient(errorM3);

    %Filter based on gradient and gradient direction.
    for r = 2:rows3-1
        for c = 2:cols3-1

            g = Gdir3(r,c);
            e0 = errorM3(r,c);
            a = alphaM3(r,c);
```

```
            %Determine if error is a minimum, dependent on the direction of
            %gradient of the pixel.
            if (g > 22.5 && g < 67.5) || (g > -157.5 && g < -112.5) %NE/SW diagonal

                %Extract adjacent error values.
                e1 = errorM3(r-1,c+1);
                e2 = errorM3(r+1,c-1);

            elseif (g >= 67.5 && g <= 112.5) || (g >= -112.5 && g <= -67.5) %NS vertical

                %Extract adjacent error values.
                e1 = errorM3(r-1,c);
                e2 = errorM3(r+1,c);

            elseif (g > 112.5 && g < 157.5) || (g > -67.5 && g < -22.5) %NW/SE diagonal

                %Extract adjacent error values.
                e1 = errorM3(r-1,c-1);
                e2 = errorM3(r+1,c+1);

            elseif (g >= 157.5 || g <= -157.5) || (g >= -22.5 && g <= 22.5) %EW horizontal

                %Extract adjacent error values.
                e1 = errorM3(r,c-1);
                e2 = errorM3(r,c+1);
            end

            %Pixel is likely to be a road only if it is a local minimum.
            if (e0 <= e1 && e0 <= e2) && (a >=0 && a < 1)
                roadCandidates3(r,c) = 1;
            else
                roadCandidates3(r,c) = 0;
            end

        end
    end

%Study area 3.
    %Calculate gradient values.
    [Gmag4, Gdir4] = imgradient(errorM3);

    %Filter based on gradient and gradient direction.
    for r = 2:rows4-1
        for c = 2:cols4-1

            g = Gdir4(r,c);
            e0 = errorM4(r,c);
            a = alphaM4(r,c);

            %Determine if error is a minimum, dependent on the direction of
            %gradient of the pixel.
            if (g > 22.5 && g < 67.5) || (g > -157.5 && g < -112.5) %NE/SW diagonal

                %Extract adjacent error values.
                e1 = errorM4(r-1,c+1);
                e2 = errorM4(r+1,c-1);

            elseif (g >= 67.5 && g <= 112.5) || (g >= -112.5 && g <= -67.5) %NS vertical

                %Extract adjacent error values.
                e1 = errorM4(r-1,c);
                e2 = errorM4(r+1,c);

            elseif (g > 112.5 && g < 157.5) || (g > -67.5 && g < -22.5) %NW/SE diagonal

                %Extract adjacent error values.
                e1 = errorM4(r-1,c-1);
                e2 = errorM4(r+1,c+1);
```

```
            elseif (g >= 157.5 || g <= -157.5) || (g >= -22.5 && g <= 22.5) %EW horizontal

                %Extract adjacent error values.
                e1 = errorM4(r,c-1);
                e2 = errorM4(r,c+1);
            end

            %Pixel is likely to be a road only if it is a local minimum.
            if (e0 <= e1 && e0 <= e2) && (a >=0 && a < 1)
                roadCandidates4(r,c) = 1;
            else
                roadCandidates4(r,c) = 0;
            end

        end
    end


%Export road networks.
geotiffwrite('rc2.tif', roadCandidates2, R_sa2,'CoordRefSysCode', 32737)
geotiffwrite('rc3.tif', roadCandidates3, R_sa3,'CoordRefSysCode', 32737)
geotiffwrite('rc4.tif', roadCandidates4, R_sa4,'CoordRefSysCode', 32737)


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Functions
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function H = calculateOAG(n,a,sX,sY)
% Inputs:
% n = size of the filter.
% a = angle of the filter orientation (measured in radians clockwise
% from north.
% sX = standard deviation in the x-direction after rotation.
% sY = standard deviation in the y-direction after rotation.
%
% Description:
% A function which calculations the coefficients for each cell for a nxn
% Gaussian filter. The filter can be oriented at a degrees measured
% clockwise from north, and can be asymmetric.
%

ind = -floor(n/2):floor(n/2); %Create an n x n grid for the filter.
[X, Y] = meshgrid(ind,ind);
a = -(a+(pi/2)); %Correct for Matlab orientation of grid, measuring from North.

%Rotate the x and y coordinates according to angle a. Negative to account
%for axes starting in the top-left instead of bottom-left.
X_0 = X*sin(a) + Y*cos(a);
Y_0 = X*cos(a) - Y*sin(a);

H_0 = exp((-1/2)*(((X_0.^2)/(sX)^2) + (((Y_0).^2)/(sY)^2))); %Calculate Gaussian.
H = H_0/sum(H_0(:)); %Normalize.

end


function output = applyOAG(im,H,n)
% Inputs:
% im = image to be filtered.
% H = filter to be used.
% n = size of the filter to be used.
%
% Description:
% Applies a filter H of size nxn to image im.
%

H = H(:); %Create column vector from H filter.
I_padded = padarray(im, [floor(n/2), floor(n/2)]); %Pad image so that entire image is
filtered.
```

```
C = im2col(I_padded, [n n], 'sliding'); %Split image into overlapping columns of the filter
size.
C_filter = sum(bsxfun(@times, C, H), 1); %Filter image using H filter.
output = col2im(C_filter, [n n], size(I_padded), 'sliding'); %Reassemble image from columns.

end


function [error, alpha_value, v] = calculateError(p, v1, v2)
% Inputs:
% p = vector representing the candidate pixel
% v1 = vector representing adjacent pixel 1
% v2 = vector representing adjacent pixel 2
%
% Description:
% Calculates the model error for the candidate pixel.
%

e1 = (norm(p - ( (dot(p,v1)/dot(v1,v1))*v1 )))/norm(v1);
e2 = (norm(p - ( (dot(p,v2)/dot(v2,v2))*v2 )))/norm(v2);

if e1 < e2
    error = e1;
    alpha_value = dot(p,v1)/dot(v1,v1);
    v = norm(v1);
elseif e1 > e2
    error = e2;
    alpha_value = dot(p,v2)/dot(v2,v2);
    v = norm(v2);
else
    error = e1;
    a1 = dot(p,v1)/dot(v1,v1);
    a2 = dot(p,v2)/dot(v2,v2);
    alpha_value = min([a1, a2]);
    v = norm(v1);
end

end
```