



PBL Netherlands Environmental
Assessment Agency

DETECTING ROADS FROM SPACE

Testing the potential of Sentinel-1 SAR imagery and deep learning for
automated road mapping

Tamara Keijzer, Aafke Schipper, Johan Meijer, Wiebe Nijland
8 March 2022

PBL

Colophon

Detecting roads from space: testing the potential of Sentinel-1 SAR imagery and deep learning for automated road mapping

© PBL Netherlands Environmental Assessment Agency
The Hague, 2022
PBL publication number: 4417

Corresponding author
Tamara.keijzer@pbl.nl

Authors
Tamara Keijzer, Aafke Schipper, Johan Meijer, Wiebe Nijland (Utrecht University)

Production coordination
PBL Publishers

Parts of this publication may be reproduced, providing the source is stated, in the form:
Keijzer T. et al. (2022), Detecting roads from space: testing the potential of Sentinel-1 SAR imagery and deep learning for automated road mapping. PBL Netherlands Environmental Assessment Agency, The Hague.

PBL Netherlands Environmental Assessment Agency is the national institute for strategic policy analysis in the fields of the environment, nature and spatial planning. We contribute to improving the quality of political and administrative decision-making by conducting outlook studies, analyses and evaluations in which an integrated approach is considered paramount. Policy relevance is the prime concern in all of our studies. We conduct solicited and unsolicited research that is both independent and scientifically sound.

Contents

Colophon	2
Corresponding author	2
Authors	2
Production coordination	2
Summary	5
1 Introduction	7
1.1 Background information	7
1.2 Road extraction from remote sensing imagery	8
1.3 Study objectives	9
2 Methods	11
2.1 Study areas	11
2.1.1 Division No. 14, Alberta, Canada	11
2.1.2 Sangha Department, Republic of the Congo	11
2.2 Data and pre-processing	14
2.2.1 Sentinel-1 SAR	14
2.2.2 Training data	16
2.3 Road extraction	18
2.3.1 Pre-processing	18
2.3.2 U-Net model	19
2.3.3 Post-processing	19
2.4 Performance evaluation	20
2.5 Comparison to the GRIP data set	23
3 Results	24
3.1 Performance evaluation	24
3.1.1 Division No. 14, Alberta, Canada	25
3.1.2 Sangha Department, Republic of the Congo	27
3.2 Comparing model results with the GRIP data set	29
4 Discussion	32
5 Conclusion	35
References	36
Appendices	39
Appendix A: Set-up Azure	39
Appendix B: Examples of deep learning	40
Appendix C: Overview of the U-Net model from Stewart et al. (2020)	42
Appendix D: GRIP data on the study areas	43
Appendix E: Accuracy results per sub-set in Division No. 14 Alberta, Canada	44
Appendix F: Accuracy results per sub-set in the Sangha Department, Republic of the Congo	46

Summary

The global biodiversity model for policy support (GLOBIO) was developed by PBL Netherlands Environmental Assessment Agency to support the design and evaluation of policy measures to halt the global loss of biodiversity by human action. One of the human pressures taken into account are roads. Roads disturb biodiversity in multiple ways. They cause habitat fragmentation, edge effects, increased animal mortality due to road kills and increased accessibility for human activity (e.g. hunting, mining). In GLOBIO, roads are represented by the data set from the Global Roads Inventory Project (GRIP), which was created by manually merging existing, publicly available national and supranational road maps. Even though GRIP provided a significant improvement over existing global road data sets (Meijer et al., 2018), the data set is still spatially and temporally incomplete.

The aim of this study was to test the potential of automated road extraction from remote sensing imagery to create a more complete and up-to-date global road network data set. The method is based on Sentinel-1 SAR imagery and the recently developed deep learning model from Stewart et al. (2020). This is a deep learning model based on the U-Net architecture for image segmentation and was created to detect roads in desert areas. For this study, we tested the model using two study areas: Division No. 14 in Alberta, Canada, and the Sangha Department in the Republic of the Congo. These areas were chosen based on the availability of training data, the abundant presence of roads according to the training data, computational feasibility (based on the size of the test regions), and difference in relevant characteristics (e.g. weather, road appearance and surroundings) between the two regions. Existing publicly available road data were used for model training. Different Sentinel-1 SAR polarisations were tested as input to the model: vertically emitted and vertically received radiation (VV), vertically emitted and horizontally received radiation (VH), and a combination of VV&VH. The different polarisations provide different surface information. Model accuracy was assessed based on high-quality and as-complete-as-possible validation data obtained by manual mapping in multiple sub-sets.

With VH as input, the model showed the highest completeness (65% and 53% in Canada and Congo, respectively) and overall accuracy ($F1=0.48$ and 0.44 , $IoU=0.65$ and 0.61). Correctness and rank distance were highest with VV&VH as input (0.80 and 0.86 , and 0.68 and 0.67 , for Canada and Congo, respectively). Not all roads in the area were detected by the model (false negatives) and non-roads were occasionally classified as roads (false positives). Overall, it was easier for the model to detect large (i.e. primary) roads compared to small (i.e. local) roads. The causes of false positives and false negatives are similar to those mentioned in other studies on detecting roads. False positives were mainly caused by objects with reflectance and spatial characteristics similar to those of roads (e.g. grass strips, rivers, terrain edges), while false negatives were mainly due to complex surroundings (e.g. roads and surrounding fields having similar values on the SAR image) or vegetation covering roads. Compared to results from the GRIP project, our model results were positionally more accurate, complete and up to date. However, they do not distinguish between road types.

The method tested here is a prototype and further improvements are needed to increase accuracy. Improvements may include the use of optical Sentinel-2 imagery or more extensive post-processing (connecting road segments and removing false positives). Pending further improvements in accuracy, the methodology has the potential to provide a robust, cost-effective, and scalable approach for the mapping and monitoring of road networks.

1 Introduction

1.1 Background information

Biodiversity around the world shows a rapid decline due to the impact on nature caused by human activity. This not only causes a decrease in species abundance or even the disappearance of species, it also poses a threat to various ecosystem services as it disturbs local ecosystems (IPBES, 2019). To halt this impact, decision-making should be based on knowledge about the impact of current drivers, future drivers and possible conservation measures on biodiversity (IPBES, 2016). For this purpose, the global biodiversity model for policy support (GLOBIO) was developed by PBL Netherlands Environmental Assessment Agency, in collaboration with various partners. GLOBIO calculates the intactness of local terrestrial biodiversity on a regional to global level, as a function of various human pressures, such as climate change, atmospheric nitrogen deposition, hunting, land use, land fragmentation and road disturbance (Schipper et al., 2020). Roads disturb biodiversity in multiple ways. They cause habitat fragmentation, edge effects, increased animal mortality due to road kills, and reinforce other pressures due to the increased accessibility for human activity (e.g. hunting, mining). These effects are particularly evident in tropical regions (Laurance et al., 2009; Benítez-López et al., 2010).

In GLOBIO, the impact of road disturbance on biodiversity is quantified using a cause–effect relationship based on the meta-analysis from Benítez-Lopez et al. (2010) combined with a global roads map from the data set of the Global Roads Inventory Project (GRIP) (PBL, 2016; Meijer et al., 2018). This data set was created using publicly available national and supranational data sets from governments, research institutes, NGOs and crowd-sourcing initiatives. Over 50% of total road length in the GRIP data set is from data sources published in or after 2010. The total road length in the GRIP data set (21.6 million km) is greater than the road length covered by older global road data sets (VMAPo: 7.4 million km, gRoads: 9.1 million km). Differences between the GRIP data set and others consist of newly constructed roads as well as the better coverage of the GRIP data set. However, the total road length in the GRIP data set is shorter than indicated by the World Road Statistics (WRS) database, based on the sum of country data averaged over the 2005–2014 period (32 million km). Although this difference may partially be caused by the various methods used to measure road length for the country data on which the WRS data set is based, which limits the reliability of the WRS data, it also indicates that the coverage in the GRIP data set could be improved (NIMA, 1997; CIESIN, 2013; IRF, 2017; Meijer et al., 2018).

The GRIP data set divides roads into five types: highways, primary roads, secondary roads, tertiary roads and local roads. Local roads represent 23% of the total road length. The corresponding class ‘other roads’ in the WRS data set comprises 68% of the total roads (IRF, 2017; Meijer et al., 2018). Therefore, Meijer et al. (2018) assume local roads to be underrepresented in the current GRIP data set. This is likely due to geographical bias in data availability. The country-level source data sets provide limited coverage of local roads. Many local roads (60%) were obtained from OpenStreetMap, which has higher coverage in more developed regions and urban areas, compared to developing countries and rural areas (Barrington-Leigh and Millard-Ball, 2017). Furthermore, unofficial roads are likely underrepresented in the GRIP data set due to the use of data from government sources. Unofficial roads may make up a large part of the rapidly growing road

network, particularly in hitherto pristine areas in the Congo and Amazon Basins (Barber et al., 2014; Laurance et al., 2014; Laurance et al., 2017).

Information on the growth of the road network is important to estimate future impacts and create long-term policies. Meijer et al. (2018) predicted future road length on a country level using GRIP data and a regression model relating road length to country area, human population density, GDP and OECD membership. A regression model based on all road types combined was the most accurate (cross-validated mean R^2 of 0.90 for all roads combined; models based on subtypes had R^2 from 0.63 to 0.77). Projections using the overall model showed large increases of road length in some of the last wilderness areas in the world, such as the Amazon, Congo Basin and New Guinea. However, the predictor variables used do not necessarily represent all factors influencing road expansion. Secondly, the projections were based on space-for-time substitution because GRIP data and their underlying sources do not include information on the year of construction of the roads. Third, the estimates are at country level, which means that there is no indication where new roads are expected to be created at sub-country level. This inhibits the use in spatially explicit impact assessments (e.g. GLOBIO).

Spatially detailed and complete data of the past and current global road network would provide insights in road expansion trajectories and potential underlying explanatory factors (e.g. land use/land cover; Jiang, 2007). These factors can then be used to predict the location of future roads and improve the scenario projections made with the GLOBIO model. Complete and up-to-date road data could also be used when studying, for example, transport and economy, poverty and inequality by lack of mobility, disaster planning, and the impact of road development on the landscape (Nelson et al., 2006; Kearney et al., 2020). Hence, there is a clear need for complete and up-to-date road data.

1.2 Road extraction from remote sensing imagery

Extracting roads from remote sensing images may improve the spatial and temporal coverage of global road data. Road extraction from recent imagery has the potential to provide up-to-date and complete data, as it does not depend on occasional and possibly biased collection by governments or crowdsourcing. Additionally, mapping roads based on satellite imagery allows for the monitoring of road development over time. The extraction of roads from imagery is challenging because the appearance of roads on the images are influenced by the satellite used (i.e. sensor type, spectral and spatial resolution), variations in weather and light, characteristics of the road (i.e. width, material), and characteristics of the surrounding environment (Ubukawa et al., 2014; Wang et al., 2016). Over the past decades, various methods have been developed to extract roads from remote sensing imagery, all with their own advantages, disadvantages and varying degrees of success (see the review of Wang et al., 2016). Especially automatic extraction of roads using deep learning is gaining more attention recently (e.g. Panboonyuen et al., 2017; Shi et al., 2017; Xu et al., 2018; Kearney et al., 2020). This method is efficient in real-time updating and saves manual extraction costs.

In recent years, convolutional neural networks (CNNs) — a form of deep learning — have become a popular method for image segmentation. These CNN architectures are also becoming prominent in remote sensing image interpretation due to their ability to combine spectral and spatial information for feature extraction. Multiple levels of representation can be used to discover

features. The lowest level uses spectral properties of the individual pixel, while higher levels explore spatial patterns like edges and corners (Alshehhi et al., 2017; Gao et al., 2019). Studies using CNNs for road detection and monitoring using satellite imagery propose various architectures. Many are applied to extract roads from very high resolution (<5m) satellite imagery. For example, Cheng et al. (2017), Panboonyuen et al. (2017), Shi et al. (2017), Xu et al. (2018), Zhang et al. (2018), Gao et al. (2019), Yang et al. (2019), Abderrahim et al. (2020) and Kearney et al. (2020) used high-resolution optical imagery, and Sen et al. (2019) and Henry et al. (2020) used high-resolution synthetic aperture radar (SAR) imagery. However, the very high resolution imagery used in these studies is not ideal for global application due to the irregular and incomplete temporal and spatial coverage. Furthermore, these products are not free of charge. Landsat (30m resolution) and Sentinel (10m resolution) imagery are free of charge and have global coverage with a regular revisit time. Yet, detection of roads is expected to be more difficult as the resolution of the imagery may exceed the width of the road, resulting in pixel values that represent a mixture of the road surface and other land covers (Ubukawa et al., 2014). Therefore a proper pixel-to-pixel or end-to-end CNN architecture, which assigns a class to each individual pixel, is needed to classify roads on lower resolution imagery (Alshehhi et al., 2017).

A particularly effective end-to-end CNN model for image segmentation is the U-Net architecture, which has been frequently applied in recent road detection studies (Henry et al., 2018; Xu et al., 2018; Zhang et al., 2018; Yang et al., 2019; Zhang et al., 2019; Abderrahim et al., 2020; Stewart et al., 2020). This type of architecture can be trained with few images and is fast, which makes it suitable for locations lacking abundant and high-quality training data (Ronneberger, 2015). Zhang et al. (2019) and Stewart et al. (2020) used a U-Net architecture to extract roads from Sentinel-1 synthetic aperture radar (SAR) images. Compared to optical imagery, the usability of SAR is not affected by weather conditions like clouds. Zhang et al. (2019) concluded that their U-Net model, trained with a handmade training data set and with an F1 score of 89%–94%, was more accurate than other methods when mapping roads in Beijing. Stewart et al., 2020 used the U-Net architecture to successfully map roads in deserts (Intersection over Union > 84%). They used OpenStreetMap (OSM) data to train the model, which is freely available. OSM data is available at a global extent, but it is not always complete, precise or up to date. The results of Stewart et al. (2020) show the possibility of using freely available input (Sentinel-1 SAR) and imperfect training data with a global coverage to automatically extract roads. However, their U-Net architecture CNN model for road extraction was only applied to deserts, hence it is unknown how it would perform in other ecosystems.

1.3 Study objectives

Automatic extraction from remote sensing images could be a solution to create a more complete and up-to-date global road network data set compared to alternative mapping approaches. It may also enable identifying temporal changes in the road network and influencing factors. Using the U-Net architecture from Stewart et al. (2020), this study aimed to explore and evaluate the potential of automated road extraction from remote sensing imagery. In view of the eventual purpose of global application, the extraction was based on remote sensing imagery with a global coverage (Sentinel-1 SAR) and existing publicly available training data. The method was applied to and evaluated for two study areas: Division No. 14 in Alberta, Canada, and the Sangha Department in the Republic of the Congo. These areas were chosen based on the availability of training data, the abundant presence of roads according to the training data, computational feasibility (based on the size of the test regions), and difference in relevant characteristics (e.g. weather, road appearance

and surroundings) between the two regions. The latter is important to evaluate the model's ability to extract roads given differences in relevant environmental conditions and characteristics (i.e. appearance of roads, vegetation, climate, land use and human activities involving roads). Because of the uncertainty regarding the completeness of the local road network in the GRIP data set, the study areas cover rural landscapes (with villages and possibly small cities) where high presence of tertiary/local roads was suspected.

The specific objectives of this study were to:

- Extract roads in the study areas using the model from Stewart et al. (2020).
- Evaluate model performance in relation to different SAR inputs and for the two study areas.
- Analyse the differences between the extracted roads and the GRIP data set.

2 Methods

The model and data used in this research were chosen based on their potential for future global application. The methodology to extract roads can be divided in three steps. The first step is retrieving and pre-processing of the data needed as model inputs (Section 2.2), including different types of satellite imagery and a mask layer telling which pixels in the satellite images are roads. The second is the deep learning model with accompanying pre- and post-processing (Section 2.3). The third step concerns the evaluation of the model (Section 2.4). This was done based on a manually mapped road network, as this was more precise and accurate than the training data. The road extraction method was applied to two study areas, which will be described below (Section 2.1).

2.1 Study areas

Two complementary study areas were chosen to test the applicability of the model for automatic road extraction. The areas were chosen based on the availability of training data, the abundant presence of roads according to the training data, computational feasibility (based on the size of the test regions), and difference in relevant characteristics (e.g. weather, road appearance and surroundings) between the two regions. The latter is important to capture the model's ability to extract roads from different environments (i.e. appearance of roads, vegetation, climate, land use and human activities involving roads). Because of the uncertainty regarding the completeness of the local road network in the GRIP data set, the study areas cover rural landscapes (with villages and possibly small cities) where high presence of tertiary/local roads was suspected.

2.1.1 Division No. 14, Alberta, Canada

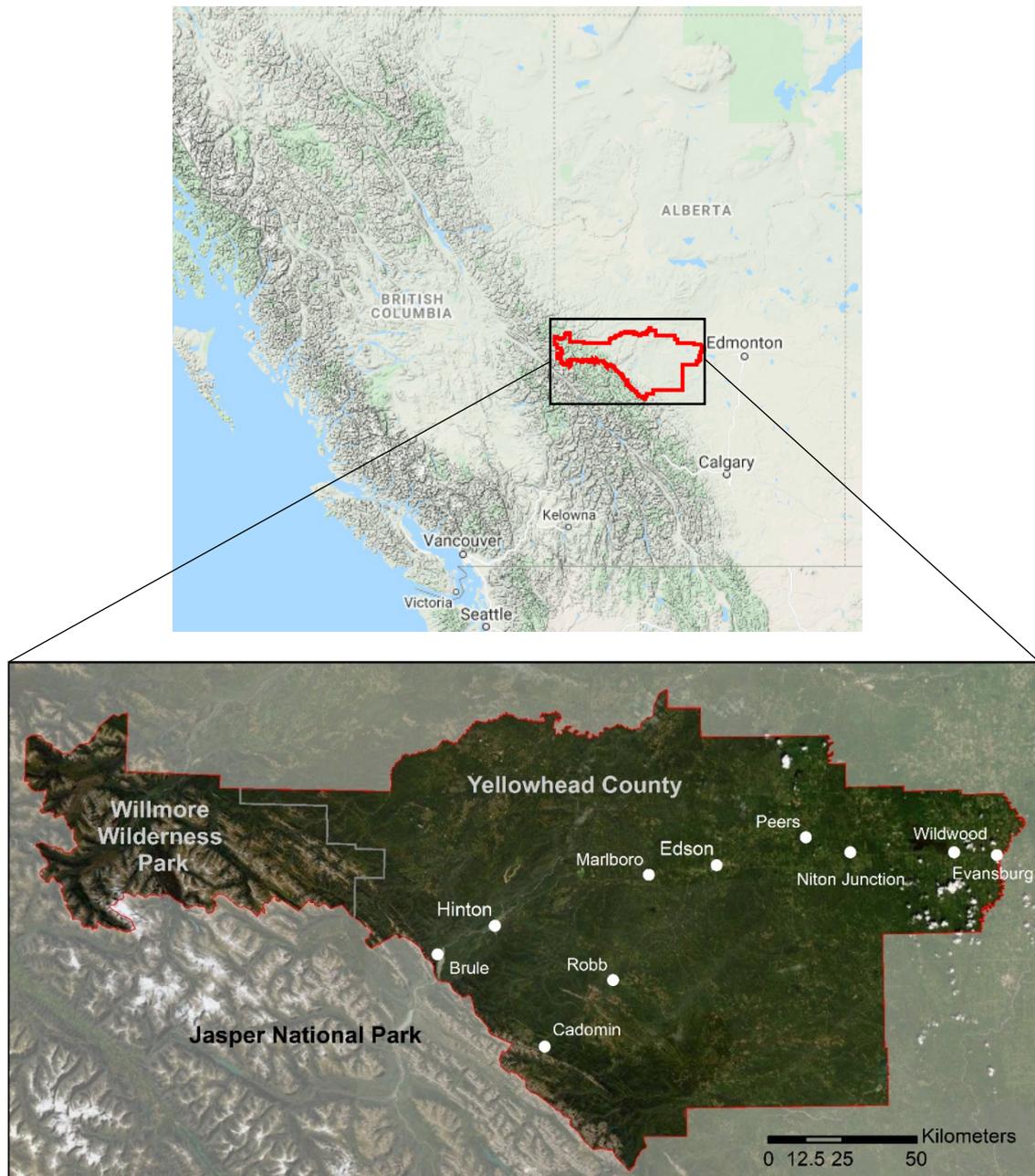
The first study area is census Division No. 14 in Alberta, Canada, with a size of 26,965 km². This study area, from here on referred to as Canada, is located west of Alberta's capital city Edmonton and northeast of Jasper National Park (Figure 1). It comprises the municipality of Yellowhead County and the Improvement District 25 (Willmore Wilderness Park). The area was selected because of the availability of high-quality road data (Statistics Canada, 2020) and because it covers a combination of different road types and landscapes, including forests, agricultural land, mountainous areas, rivers and lakes, and small cities and villages. The largest towns are Hinton (9,882 inhabitants) and Edson (8,414 inhabitants). The total population is 29,291, corresponding with a population density of 1.1 inhabitants per km² (Statistics Canada, 2016).

2.1.2 Sangha Department, Republic of the Congo

The second study area is the Sangha Department in the northern Republic of the Congo (Figure 2). This study area, from here on referred to as Congo, has a size of 55,800 km², which makes it more than twice as large as the first study area. In 2007, the Sangha region had a population of 85,738 (National Center for Statistics and Economic Studies, 2007). Its capital city is Ouéssou (28,179 inhabitants). Other cities/villages are Mokéko (30,172 inhabitants), Souanké (9,711 inhabitants), Sembé (9,550 inhabitants), Ngbala (4,613 inhabitants) and Pikounda (3,513 inhabitants) (National Center for Statistics and Economic Studies, 2007). The Sangha region lies within the Congo Basin, which is a widely known biodiversity hotspot. The lowland rainforests are among the last large mammal strongholds in the Congo Basin (WWF, 2020). There are several protected areas in the Sangha region, which are home to numerous bird, plant and tree species and populations of chimpanzees, forest elephants, western lowland gorillas and other endangered mammals. Outside

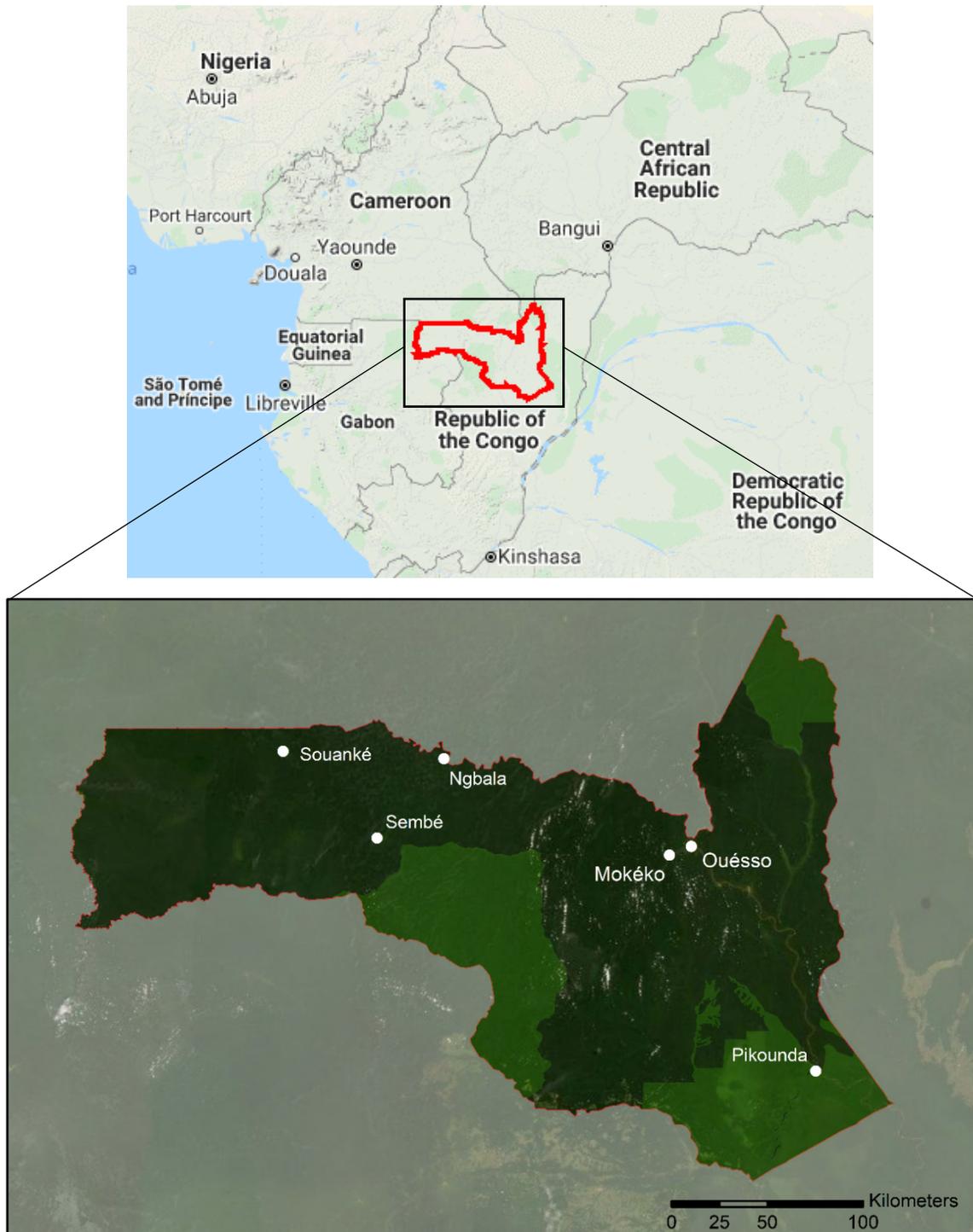
the national parks, mechanised logging has been developed across the region. This is associated with road construction in otherwise inaccessible areas. The rapidly expanding and improving road network also increases the trade in commercial bushmeat (IUCN, 2012; WWF, 2020).

Figure 1
Division No. 14, Alberta, Canada



Two maps: one showing the location of Division No. 14 in Alberta, Canada, and one showing the locations of villages, hamlets and regions. Base maps were retrieved from Google Earth Engine and ArcMap.

Figure 2
Sangha Department, Congo



Two maps: one showing the location of the Sangha region and one showing the location of cities/villages and protected areas (these are the light green areas). Base maps were retrieved from Google Earth Engine and ArcMap, the protected areas from the World Database of Protected Areas (2021).

2.2 Data and pre-processing

2.2.1 Sentinel-1 SAR

The ESA Copernicus Sentinel-1 archive was exploited for satellite imagery. As the imagery is available under free and open access policy and acquired at regular intervals worldwide, it has a high potential for global application. The Sentinel-1 mission consists of two C-Band synthetic aperture radar (SAR) satellites, together acquiring global coverage every 24 days from April 2014 and every 12 days from April 2016, after the launch of the second satellite. The retrieval and pre-processing of the Sentinel-1 SAR imagery was executed in Google Earth Engine (GEE). This is a freely available high-performance cloud platform which enables the processing and analysis of multi-petabyte remote sensing data without downloading. It provides global and publicly available satellite imagery and other global data sets (Gorelick et al., 2017).

Ground Range Detected (GRD) Sentinel-1 SAR images are available within Google Earth Engine, which means they are already radiometrically calibrated and ortho-corrected. The values in the images represent surface backscatter values. The amount of backscatter is affected by surface parameters as roughness, shape and dielectric properties (permittivity and conductivity, which are influenced by moisture content). Smoother surfaces tend to have a lower backscatter. For example, calm water or roads tend to have low backscatter, while urban areas have high backscatter. In this study dual-polarimetric (VV and VH) images from the interferometric wide (IW) swath model are used, following the road mapping applications by Zhang et al. (2019) and Stewart et al. (2020). VV means vertically emitted and vertically received radiation. VH means vertically emitted and horizontally received radiation. These different polarisations provide different surface information. The images have a pixel size of 10x10m with a swath width of 250km. Images of both ascending and descending orbital modes were used.

The number of SAR images available for the different polarisations varied per year (Table 1). Some images showed very low backscatter values at the edges, which is an artefact of pre-processing from raw data to GRD images (Ali et al., 2018; Filipponi, 2019). This noise can be removed by thresholding (Hajduch and Miranda, 2018). Based on visual inspection, pixels with a backscatter value lower than -30 (i.e. the threshold) were removed. This only removed pixels at image edges and did not affect the rest of the image. To reduce speckle, the median of all images within a time period was taken for each polarisation (Stewart et al., 2020; Figure 3). By taking the median, occasional erroneous values in images are circumvented. Because in Stewart et al. (2020) the input yielding the best results varied among study areas, multiple model inputs were tested in this study: VV, VH and a combination of VV and VH. For Canada, only images from May up to and including August were selected, as road recognition may be hindered because of snow cover during the other months. The VV 2015, VH 2017 and VV&VH 2017 median images were used as different SAR model inputs. These years were chosen based on the number of available images (Table 1 and Figure 4) and to have imagery temporally as close as possible to the training data (2012). For the study area in Congo the median VV and VH of September 2020 up to and including December 2020 were used as the training data is from January 2021. Coverage of the area by the images can be viewed in Figure 5. The period of 4 months was chosen to match the time period of the images used for Canada. Multiple years or longer time periods were not aggregated to limit possible noise due to the appearance or disappearance of roads on part of the images.

The pixels of different images are not all aligned due to shifts in image positions and orientation (ascending/descending; Figures 4 and 5), resulting in an altered pixel size in the median images. As the median images are dealt with as arrays in the modelling part, pixel size must be constant and the grid must be the same among the inputs. Therefore, all median SAR images were reprojected to the projection of the first image in the VV collection and the resolution was set to 10m by taking the arithmetic mean (Figure 3). This result was then clipped into a square around the study area and exported as tiff. These tiffs were input for the method described in Section 2.3.

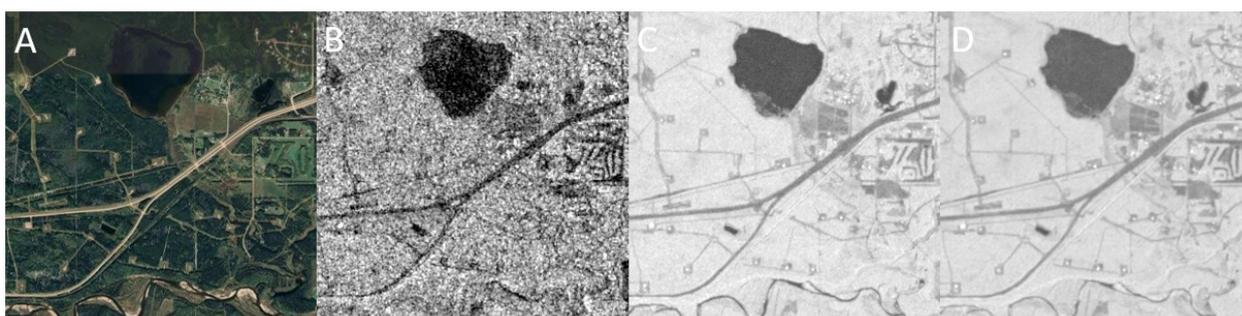
Data retrieval from Google Earth Engine and pre-processing of all data were performed using Google Colab. The Jupyter Notebooks can be found at: [this link to Google Colab Canada](#) for Division No. 14, Alberta, Canada and [this link to Google Colab Congo](#) for the Sangha Department, Congo.

Table 1
Sentinel-1 SAR image availability

	Year	VV	VH
Division No. 14 Alberta, Canada 1 May – 1 September	2015	45	0
	2016	64	12
	2017	115	106
	2018	119	119
	2019	147	147
	2020	137	137
Sangha Department, Congo 1 September – 1 January	2015	68	68
	2016	103	103
	2017	103	103
	2018	101	101
	2019	100	100
	2020	101	101

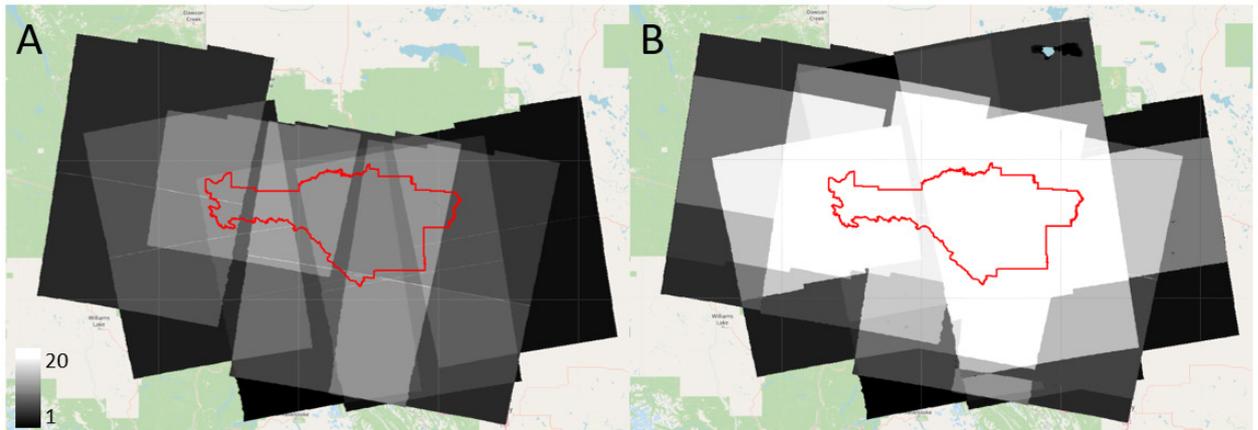
Number of images available per year and per polarisation, for the two study areas. The data used as model inputs are presented in **bold**.

Figure 3
Pre-processing of the imagery



Four images showing pre-processing formats, depicting the hamlet of Marlboro in Canada. A) Optical satellite base map GEE; B) Single speckled VV image; C) Median VV 2017; and D) Reprojected 10m resolution median VV 2017.

Figure 4
Image coverage of Division No. 14, Canada



Two figures showing the number of images covering Division No. 14 in Alberta, Canada, (the border of the area is outlined in red): A) VV 2015; and B) VV and VH 2017. Base map: OpenStreetMap via Google Colab.

Figure 5
Image coverage of the Sangha Department, Congo

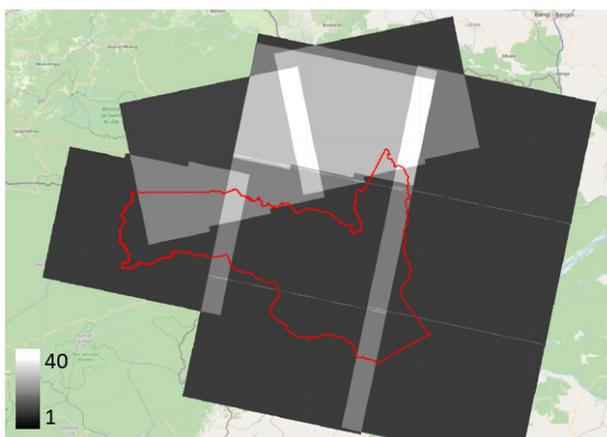


Figure showing the number of images covering the Sangha Department in Congo, (the border of the area is outlined in red) for both VV and VH 2020. Base map: OpenStreetMap via Google Colab.

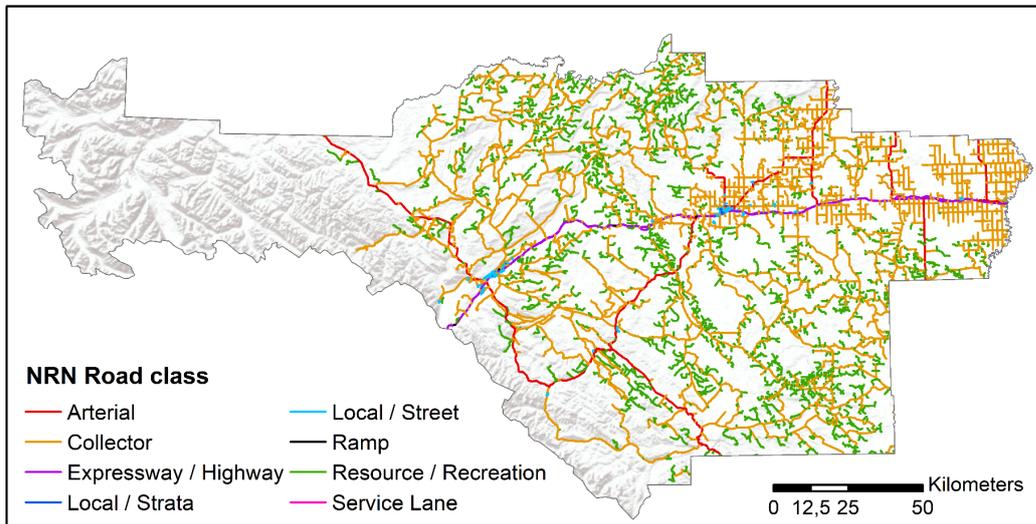
2.2.2 Training data

For the study area in Canada, the Canadian National Road Network (NRN) was used as training data (Figure 6). This data set was created by the GeoBase initiative (Johnson and Singh, 2003), including data providers from the federal, provincial, territorial and municipal governments. The NRN consists of centre line road features with additional information including creation date, revisit date, road class, paved/unpaved status, road name and place if known. Almost all road features of the NRN within the study area were revisited in 2012. A few roads were revisited more recently. The data set can be found on the [Federal Government's Open data portal](#). The version used in this study was downloaded in November 2020.

For the study area in Congo, the freely available OpenStreetMap (OSM) road data were used for model training (Figure 7). Country-level data were downloaded in January 2021 and sub-set to the study area. As OSM comprises crowdsourced data, it may not be complete, precise or up to date. However, OSM data is globally available and often also available for more remote areas, which makes the data potentially suitable for global application.

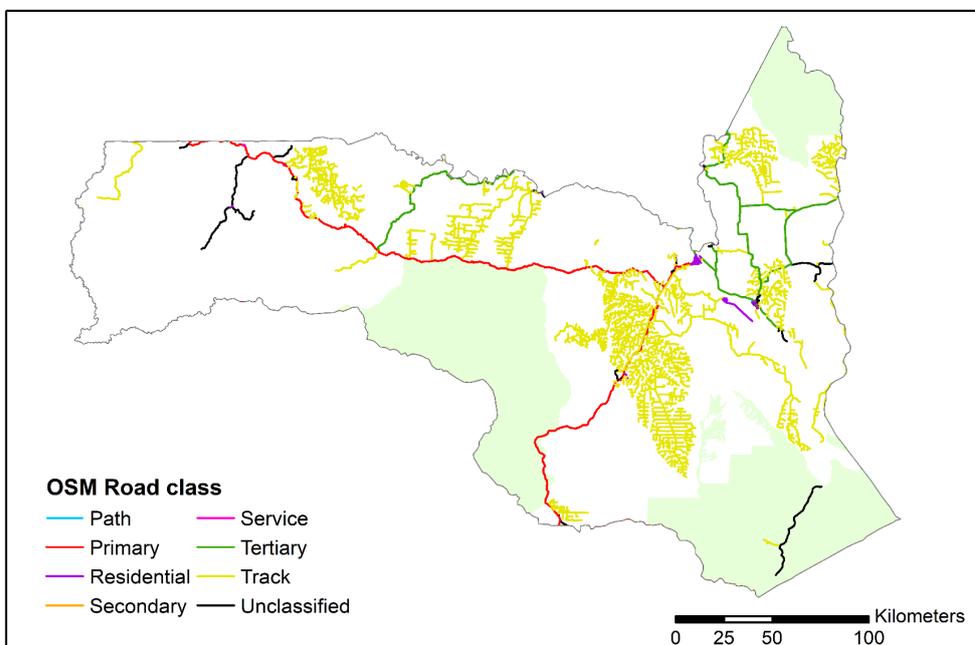
For each area, the training data was transformed to raster format with the same resolution and projection as the SAR model inputs. All lines were transformed to pixels representing roads with a width of 1 pixel. This created binary maps with 1=road and 0=no road, which were used as a model input for training.

Figure 6
National Road Network (NRN) in Division No. 14, Canada



This figure shows a map from the National Road Network (NRN) in Division No. 14 in Alberta, Canada. The various types of roads are distinguished by colour.

Figure 7
OpenStreetMap in the Sangha Department, Congo



This figure shows a map obtained from the OpenStreetMap for the Sangha region in Congo. The various types of roads are distinguished by colour. Protected areas are indicated in light green (WDPA, 2021).

2.3 Road extraction

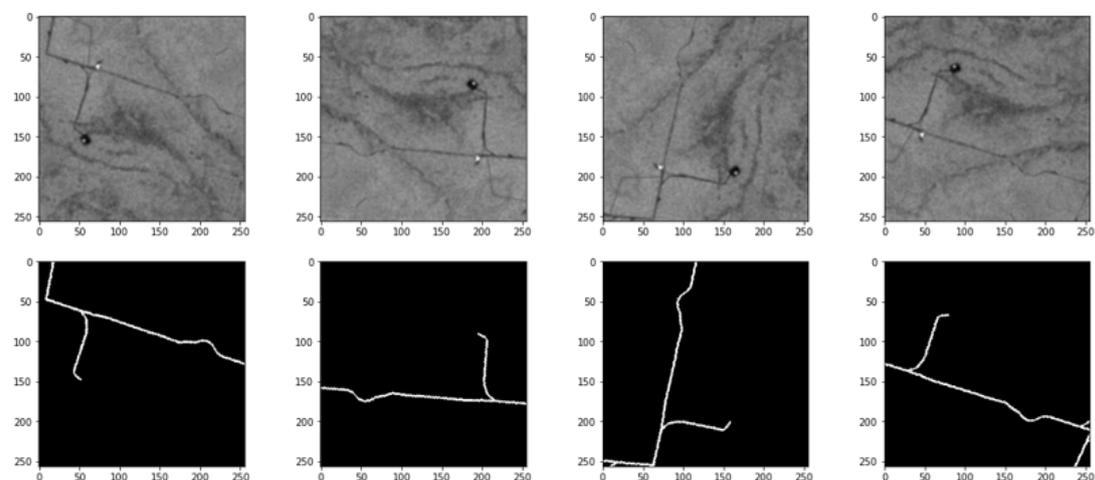
The inputs for the road classification model are a training data layer and one of three combinations of Sentinel-1 SAR layers: VV, VH or VV&VH. All layers have the same extent and spatial resolution. Modelling was performed in Python 3.7 with Keras and Tensorflow. Road extraction could not be performed in Google Colab due to memory limitations. Therefore, road extraction was executed on an Azure Virtual Machine with an Intel Xeon CPU E5-2667 v3 and a memory of 112 GB. See Appendix A for a description of the Azure set-up. The extraction of roads included pre-processing of the model inputs, model training and application, and post-processing of the model results.

2.3.1 Pre-processing

All inputs were first split into adjacent, non-overlapping patches of 256x256 pixels. The patches comprise the input data for training the U-Net model. Patches without roads according to the training data or containing no-value pixels were removed from model training. The SAR patches were normalised to contain pixel values from 0 to 1, which speeds up and eases model training (Huang and Nowack, 2020). For the study area in Canada, 2,396 patches were available for training. For the study area in Congo, 1,808 patches were available for training.

Data augmentation was applied to increase the diversity of the model training samples without collecting new data (Stewart et al., 2020). The model was trained by a learning algorithm working through all patches multiple times. Each time, the patches were a little different than the original due to data augmentation. Data augmentation helps to prevent overfitting in deep learning models (Shorten and Khoshgoftaar, 2019). The data augmentation included horizontal flips, vertical flips and random rotation (360° range; Figure 8). Reflect mode was selected to fill in gaps following transformations, such as corners of patches after rotation. This mode is the best fill option as it fills the gaps using the mirror image of the same number of non-blank pixels across the dividing line. In this way lines are continued, as would be expected with roads (Stewart et al., 2020). Data augmentation was performed using the image data generator of the Keras library.

Figure 8
Data augmentation



This figure shows two sets of four images, showing data augmentation for an input SAR VH 2017 patch (the top four images) and corresponding reference road data (the four images on the bottom). The far left set of images are the originals.

2.3.2 U-Net model

Roads were extracted with the U-Net model originally proposed by Paul (2020), who created the model to extract roads from high-resolution (1m) optical data, and modified by Stewart et al. (2020). The U-Net model consists of a down-sampling encoder and up-sampling decoder part, which are connected via cross-connections. Convolutional layers are used to extract information. These layers use filters to recognise the appearance of structures of interest, in this case roads. In the down-sampling part, resolution is decreased by pooling layers. By using different resolutions, different object information can be derived. For example, with high resolution object edges can be detected, while context as textures or objects parts is recognised using lower resolutions (Appendix B). Model training is an iterative process during which the trainable parameters in the model are optimised to recognise the roads on the SAR imagery. Locations of roads on the SAR input patches are indicated by the corresponding training data patches. Following Stewart et al. (2020), the model was trained separately for each area and for each of the three different inputs (VV, VH and a combination of VV&VH) using all patches for training. The model hyperparameters were adopted from Stewart et al. (2020), as adjustment of these would be a process of trial and error, which was beyond the scope of this study. For more information on the U-Net model, see Stewart et al. (2020) and Appendix C.

2.3.3 Post-processing

The trained model was applied to predict the presence of roads in all patches. The model returns a segmented image for each input patch, with pixel values ranging from 0 to 1 representing the probability of road presence. Following Stewart et al. (2020) and Henry et al. (2018), the pixels with a value equal or greater than 0.5 were set to 1 (road), while those with a value smaller than 0.5 were set to 0 (no road). The patches were then stitched back together to create a binary map of the area with the predicted roads. For each area this was done with VV, VH, and VV&VH inputs, creating three different predicted road maps.

The raster layers were converted to line segments in ArcMap using the centre line vectorisation method in the ArcScan toolbar. This generates vector line features at the centre of the raster cells. Gap setting was set to 10 pixels to connect small gaps (max 100m) between lines. This threshold was determined based on visual inspection, which revealed wrong connections when larger thresholds were adopted. Other vectorisation settings were kept at default. If groups of pixels with value 1 (roads) were not elongated (>20 pixels wide) and/or too small (detected as noise by the function) these pixels are not transformed to a line feature by the centre line vectorisation method (Figure 9). Average width (pixels) was added as attribute to the features.

Figure 9
Post-processing



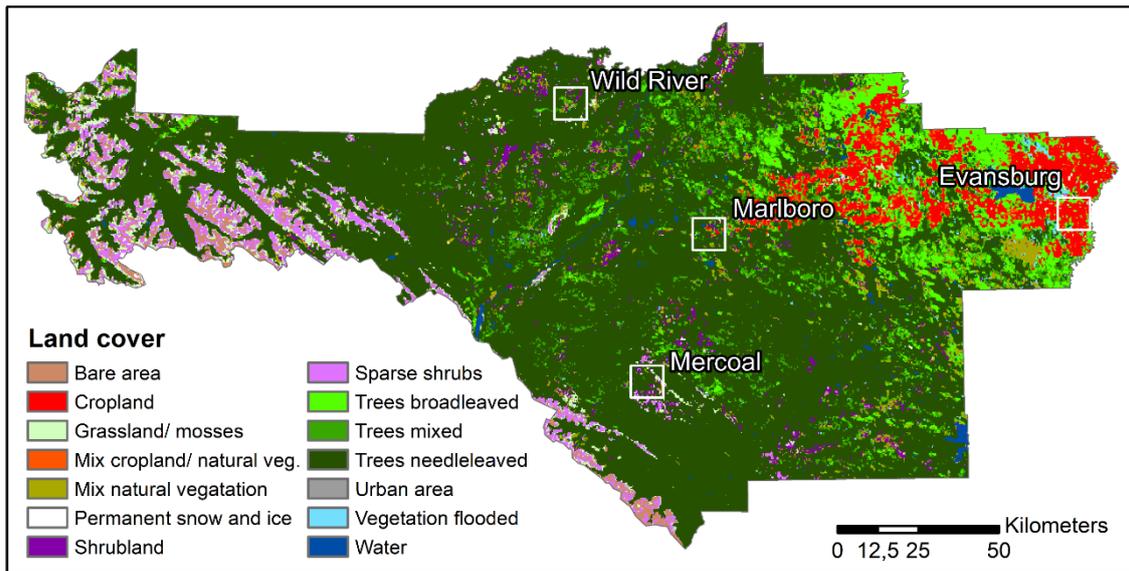
The figure shows two maps of post-processing, visualising the area around the hamlet of Marlboro in Canada. On the left: the model output in raster format. On the right: the model output after conversion into vector format with the centre line vectorisation method. Both maps have a small circle and a large circle showing the removal and connection of groups of pixels respectively.

2.4 Performance evaluation

The performance of the model was evaluated based on manually digitised road maps, because of omissions and inaccuracies in the training data. Manual editing of the training data was avoided in the interest of testing a globally applicable method. However, for a representative evaluation of model performance, it was preferred to obtain validation data as accurate as possible (Stewart et al., 2020).

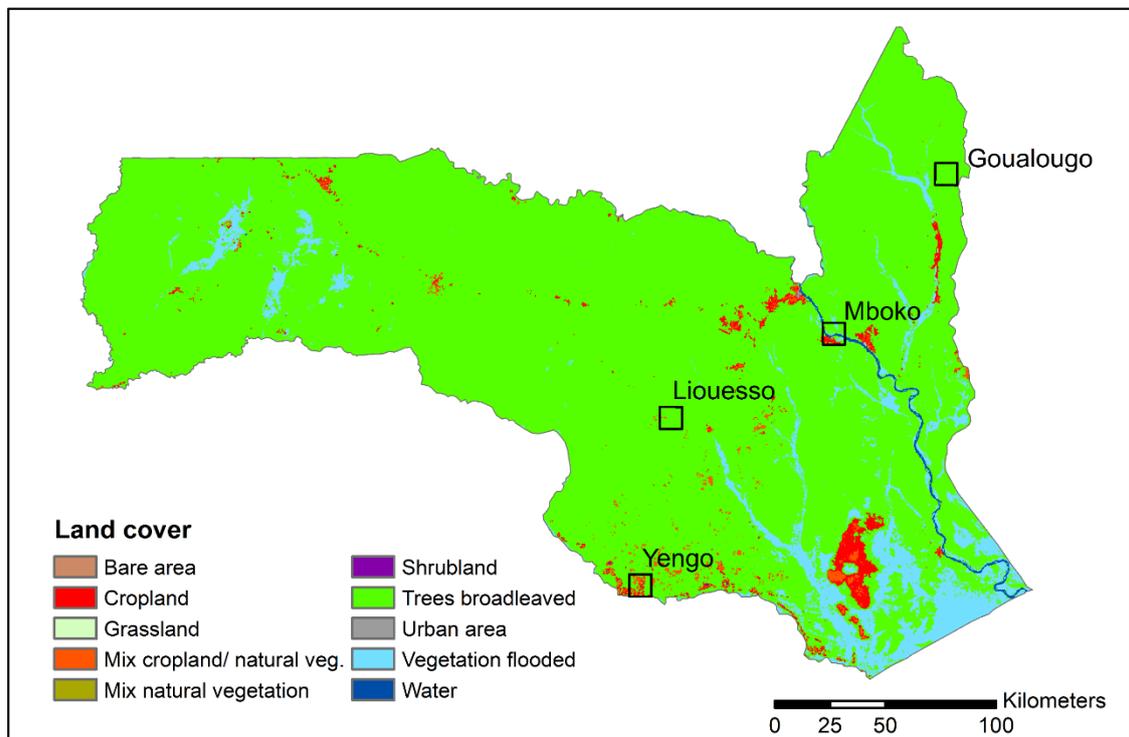
Manual digitisation was applied to four sub-sets of 10x10km in each area. Multiple sub-sets were chosen in order to cover differences in land cover (Figures 10 and 11), geographical location, appearance of roads, presence of different types of roads, and abundance of roads in the training data and the model predictions. Elongated lines visible in the SAR imagery were assigned to different classes, including different road types and linear non-road landscape elements (Table 2). The optical imagery base map available in ArcMap, Google Maps and Sentinel-2 imagery were utilised to determine the identity of the structures. The different road types were distinguished using the GRIP data set, the training data set and visual appearance, such as road width, colour and position within the network.

Figure 10
Location of the sub-sets in Division No. 14, Canada



Map showing the location of the sub-sets (indicated by white outlined squares) that were used in manual mapping and model evaluation in Division No. 14 Alberta, Canada. The base map shows the various types of land cover. Land-cover data were retrieved from ESA 300m resolution global land cover 2015.

Figure 11
Location of the sub-sets in the Sangha region, Congo



Map showing the location of the sub-sets (indicated by black outlined squares) that were used in manual mapping and model evaluation in the Sangha region in Congo. The base map shows the various types of land cover. Land-cover data were retrieved from ESA 300m resolution global land cover 2015.

Table 2
Classes used for validation

Road class	Number	Description
Highways	1	As indicated in the GRIP data set, but the study areas did not contain highways.
Primary roads	2	As indicated in the GRIP data set.
Secondary roads	3	Roads branching off from primary roads, often still relatively wide and clearly visible in optical imagery, compared to some other road types.
Tertiary roads	4	Roads branching off from primary and secondary roads. Less wide and less visible on optical imagery, compared to secondary roads.
Local roads	5	Roads branching off from secondary and tertiary roads, often narrow and less visible on optical imagery, compared to secondary/tertiary roads. Also dead-end roads.
Grass/dirt strips	6	Elongated strips of grass or bare ground according to imagery. Not clear if they are used as roads.
Rivers	7	Rivers or river edges visible in imagery.
Railroads	8	Railroads (and edges). Locations of railroads checked using OSM data.
Terrain edges	9	Edges mapped at elevation differences, elongated roughness differences, e.g. forest edges
Other	10	Unidentifiable model prediction. Often occurs in open or urban areas.

Name and description of the road classes used in the model validation and the class number as used in the validation data.

The model predictions were compared to the validation data using a buffer (flat end) of 40m around the lines of classes of the manual data set. The width of the buffer was chosen based on visual inspection. If the buffers of the different road types overlapped, the buffer of the lower class was removed. Using the buffers, model predictions were assigned to the different classes (Table 1). Model predictions outside the buffers were manually assigned to a road or non-road class.

Next, a confusion matrix was formed to calculate model accuracy. True positive (TP) is the total length of predicted roads assigned to road classes. False positive (FP) is the total length of predicted roads assigned to non-road classes. False negative (FN) is the total length of roads the model did not capture i.e. the length of roads in the manual data set minus the true positives. Based on this confusion matrix, accuracies were calculated according to different methods found in other papers mapping roads (Table 3). The metrics were calculated for the different model inputs (VV, VH and VV&VH) for all 10x10km sub-sets separately and for each study area as a whole.

Table 3
Accuracy indices used to evaluate the model outcomes

Index	Description	Formula	Used by
Completeness (Recall)	Completeness of the prediction compared to the reference data.	$\frac{TP}{TP + FN}$	Alshehhi et al. (2017) Panboonyuen et al. (2017) Shi et al. (2017) Xu et al. (2018) Gao et al. (2019) Zhang et al. (2019)* Kearney et al. (2020) Stewart et al. (2020)*
Correctness (Precision)	Fraction of the prediction that is correct according to the reference data.	$\frac{TP}{TP + FP}$	Alshehhi et al. (2017) Panboonyuen et al. (2017) Shi et al. (2017) Xu et al. (2018) Gao et al. (2019) Zhang et al. (2019)* Kearney et al. (2020) Stewart et al. (2020)*
Intersection over Union (IoU, Quality)	Overall goodness taking into account completeness and correctness.	$\frac{TP}{TP + FN + FP}$	Shi et al. (2017) Kearney et al. (2020) Stewart et al. (2020)*
F1	Combination of Completeness (Com) and Correctness (Cor).	$\frac{2 * Com * Cor}{Com + Cor}$	Panboonyuen et al. (2017) Gao et al. (2019) Zhang et al. (2019)*
Rank distance	Combination of Completeness (Com) and Correctness (Cor).	$\sqrt{\frac{Com^2 + Cor^2}{2}}$	Stewart et al. (2020)*

For each index the description and formula, and publications in which the index is used. All indices range from 0 to 1, with 1 being a perfect score. *indicates the use of Sentinel-1 SAR data in the publication. The other publications use very high resolution data.

2.5 Comparison to the GRIP data set

The model results were compared with the existing global road network data, in the GRIP data set (Appendix D). The GRIP data set contains the following road classes: highways (1), primary roads (2), secondary roads (3), tertiary roads (4) and local roads (5). For Canada, GRIP data are from the [Canadian Government's Atlas of Canada](#). For Congo the source is OpenStreetMap from 2012. The length of roads manually mapped and predicted by the model in the sub-set areas of the performance evaluation were compared with the road length according to the GRIP data set.

3 Results

3.1 Performance evaluation

The model was able to detect roads in both study areas with similar accuracy (Table 4), despite the differences in environments (i.e. appearance of roads, vegetation, climate, land use and human activities involving roads). The performance varied among the accuracy indices, which shows the importance of calculating a set of complementary indices. The VH input resulted in the highest completeness, while the VV&VH input resulted in the highest correctness. Combining completeness and correctness, the model gave the most accurate result with VH input according to the IoU and the F1 index. According to rank distance the model predicted roads most accurately with VV&VH inputs (and VV in Congo).

The high completeness with VH input reflects that the model classified a relatively large number of cells as roads based on this input. Overall, the model predicted a total road length of 17,741 km in Canada and 6,482 km in Congo with VH input. With VV&VH input a smaller total road length was predicted: 12,359 km in Canada and 2,784 km in Congo. This is perhaps caused by the smaller range in values in VH backscatter compared to VV backscatter, making it more difficult for the model to separate different features. The combination of VV&VH yields a more selective model that predicts less roads and results in a higher correctness. However, this also decreased the completeness of road mapping. In particular in Canada, VV input results in lower model performance compared to VH and VV&VH.

The model performance in each study area is described in more detail in the subsections below. The results per sub-set can be viewed in Appendix E and F.

Table 4
Accuracy results

Index	Canada			Congo		
	VV	VH	VV&VH	VV	VH	VV&VH
Completeness	0.55	0.65	0.54	0.48	0.53	0.40
Correctness	0.64	0.65	0.80	0.82	0.71	0.86
IoU	0.42	0.48	0.47	0.43	0.44	0.37
F1	0.59	0.65	0.64	0.60	0.61	0.54
Rank distance	0.60	0.65	0.68	0.67	0.63	0.67

The accuracy results per study area, per input. Calculations were based on manual mapping in four sub-set areas of 10x10 km. The best input according to the index is provided in **bold**.

3.1.1 Division No. 14, Alberta, Canada

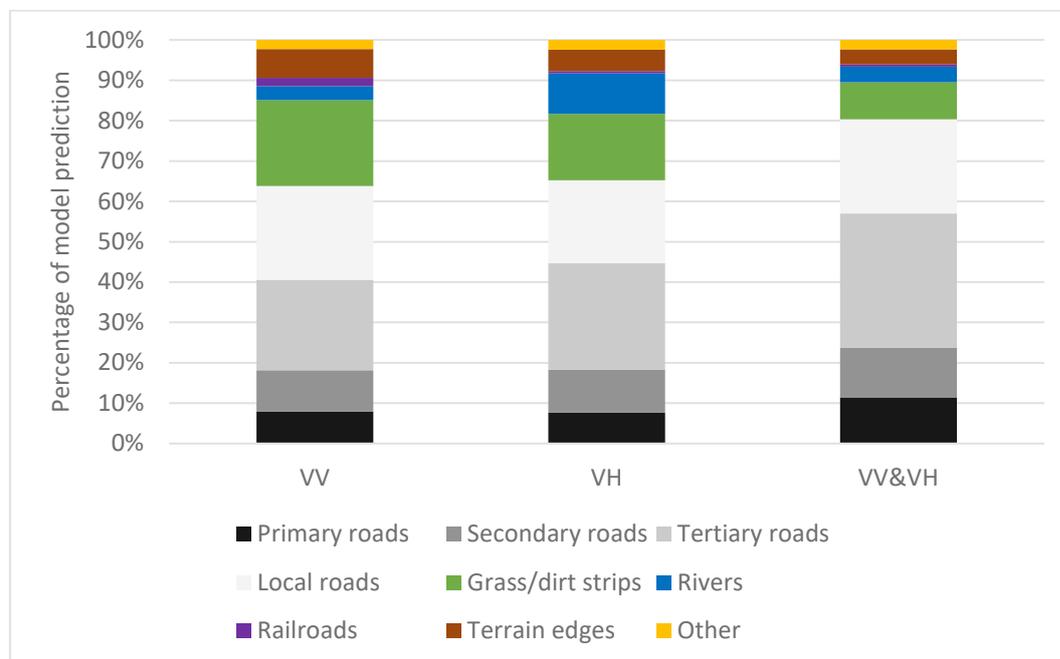
Overall, the model produced most road predictions with VH input (total of 435 km in the sub-sets) and the least with VV&VH input (290 km). The model showed the highest completeness with VH backscatter as input, followed by VV backscatter and VV&VH backscatter (Tables 4 and 5). Though VV&VH input showed the lowest completeness score, it had the least false positives, i.e. non-roads classified as roads, resulting in the highest correctness (Figure 12). VH input showed the highest false positive rate. Most of the false positives (47%–59%) were caused by elongated strips of grass or dirt in forested areas (Figures 12 and 13). These appear the same as roads on the radar imagery but are actually seismic lines (corridors used to transport and deploy geophysical survey equipment) created for oil and natural gas exploration or pipelines. Other elongated features as rivers (10%–29% of false positives) or terrain edges (15%–20%) were also often misinterpreted as roads. Rivers were confused with roads, relatively often, with VH input.

Table 5
Confusion matrices for Division No. 14, Canada

	VV		VH		VV&VH	
	Predicted roads	Predicted non-roads	Predicted roads	Predicted non-roads	Predicted roads	Predicted non-roads
Real roads	239	195	284	150	233	201
Non-roads	135		151		57	

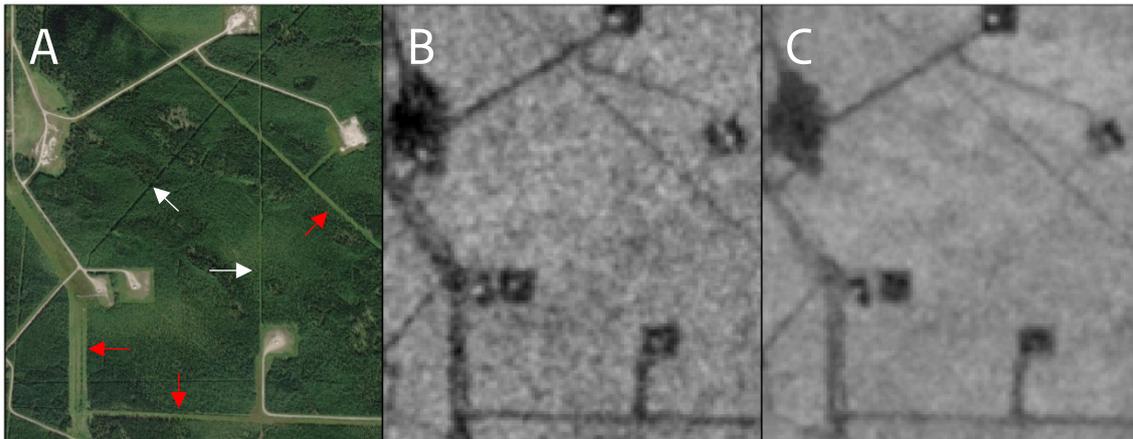
Confusion matrices for real and detected roads (total length in km), for various model inputs in the sub-sets of manual mapping.

Figure 12
Distribution of model predictions over the road and non-road classes for Division No. 14, Canada



This figure shows the distribution of model predictions over the road and non-road classes (in %). The share of all road classes (true positives) combined provides the correctness. Non-road classes show the origin of false positives in model predictions. Table 5 provides the total length of model predictions.

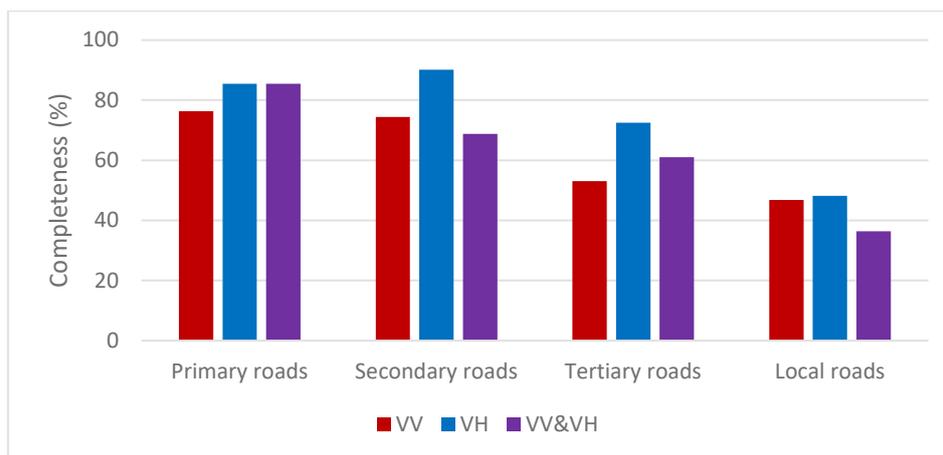
Figure 13
Example of elongated strips/corridors creating false positives in model predictions



Three figures showing examples of corridors created for oil and natural gas exploitation (i.e. seismic lines and pipelines), from left to right : A) shows optical imagery (base map ArcMap, 28 August 2019), B) shows the VV 2015 model input, and C) shows the VH 2017 model input. Arrows point to the features in image A. The white arrows indicate narrow corridors that are not clearly visible in the radar imagery, while the red arrows indicate wider corridors that can be mistaken for roads in model predictions.

In total 434 km of roads were manually mapped within the four sub-sets of 10x10km, of which 39 km primary roads (9.0%), 51 km secondary roads (11.8%), 158 km tertiary roads (36.5%) and 186 km local roads (42.8%). Overall, the model showed a higher ability to detect primary and secondary roads than tertiary and local roads (Figure 14). The latter were less distinctive in the radar imagery as they can be narrow, unpaved or partially covered by trees. Roads were furthermore missed due to surroundings with similar backscatter values, such as open fields or bare ground. With VH input the model showed the highest completeness score for each road type.

Figure 14
Completeness of the model predictions per road type for Division No. 14, Canada



This figure shows the completeness (%) of each road type in model predictions, i.e. the percentage of actual roads captured by the model, with different model inputs (VV, VH or VV&VH).

3.1.2 Sangha Department, Republic of the Congo

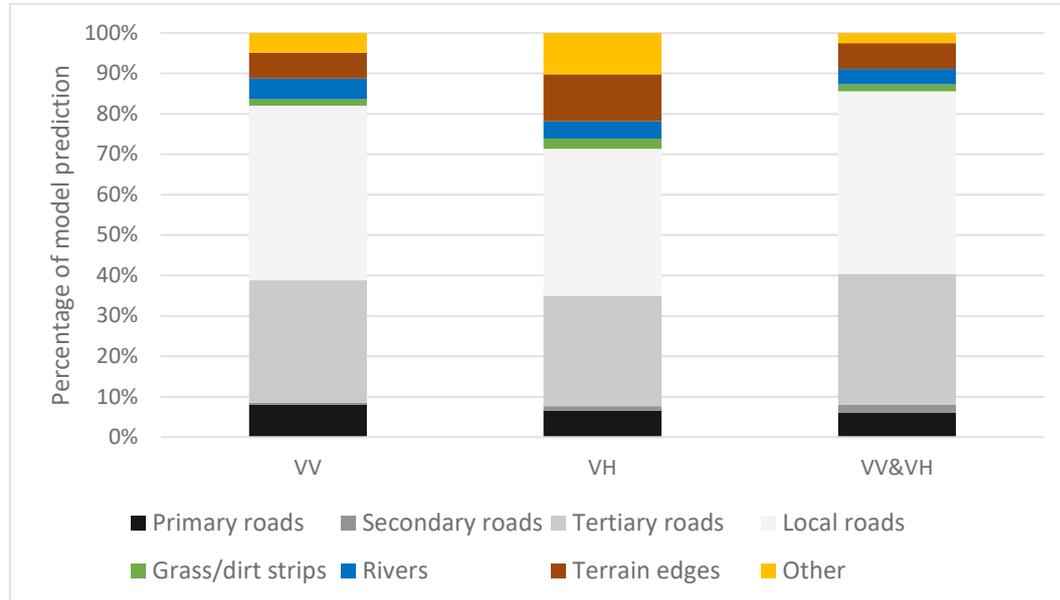
For the sub-sets in Congo, the model showed most road predictions with VH as input (182 km) and the least with VV&VH input (114 km). Overall, the prediction with VH input was most complete (53%), followed by VV (48%) and VV&VH (40%). Correctness scores were high in general being 0.82 for VV, 0.71 for VH and 0.86 for VV&VH model input (Tables 4 and 6, Figure 15). Model predictions with VV&VH had the least false positives both in total length and percentage. Predictions with only VH as input had the most false positives. Terrain edges, especially forest edges, were mistaken for roads the most often (35%–45% of the false positives; Figures 15 and 16). For urban areas, the model often predicted all pixels including buildings as being roads. The roads within these areas were not in the model output, as the conversion into lines was only applied to elongated pixel groups.

Table 6
Confusion matrices for the Sangha region, Congo

	VV		VH		VV&VH	
	Predicted roads	Predicted non-roads	Predicted roads	Predicted non-roads	Predicted roads	Predicted non-roads
Real roads	117	129	130	116	98	149
Non-roads	26		52		16	

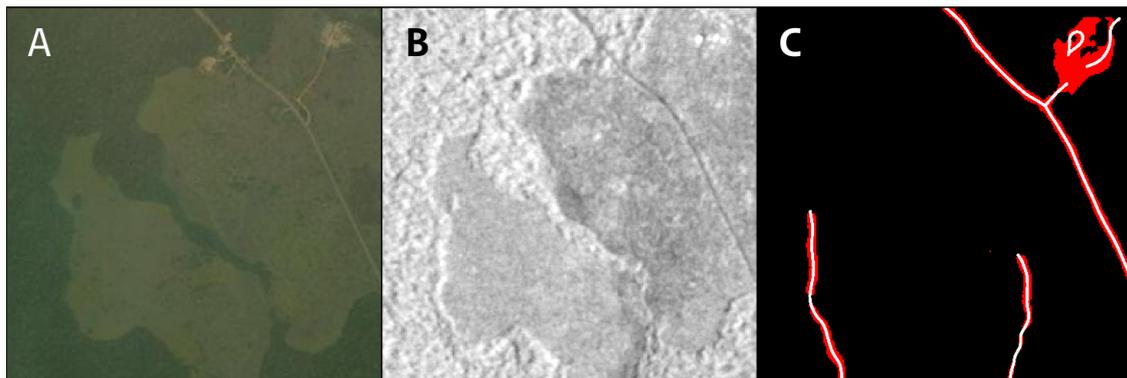
Confusion matrices for real and detected roads (total length in km), for various model inputs in the sub-sets of manual mapping.

Figure 15
Distribution of model predictions over the classes for the Sangha region, Congo



This figure shows the distribution of model predictions over the road and non-road classes (in %). The share of all road classes (true positives) combined provides the correctness. Non-road classes show the origin of false positives in model predictions. Table 6 provides the total length of model predictions.

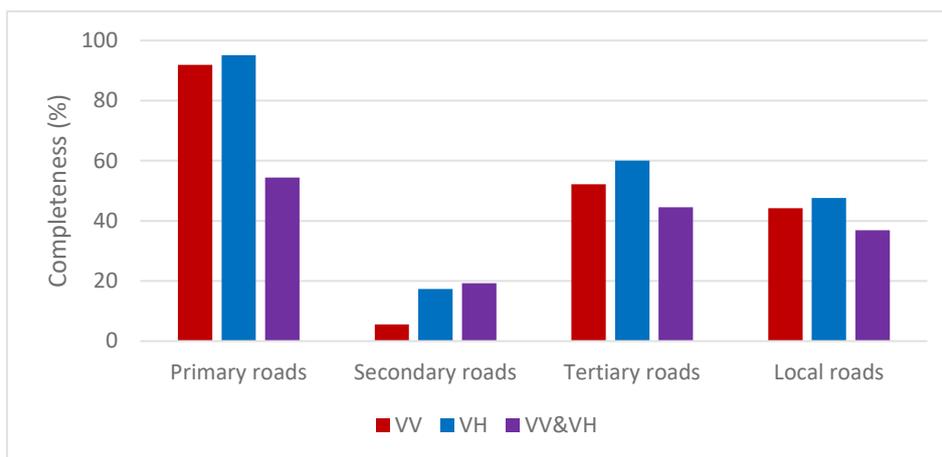
Figure 16
Example of forest edges and urban structures creating false positives in model predictions



Three figures showing examples of forest edges and urban structures detected as roads, from left to right: A) shows optical imagery (base map ArcMap, 31 January 2018), B) shows VV model input, and C) shows VV model output in pixels (in red) and centre lines (in white). The detection of roads in and around an urban structure is visible in the top right corner of the images. Forest edges indicated as roads are visible as the north–south oriented lines in image C.

In the study area in Congo, 246 km of roads were manually mapped, of which 13 km were primary roads (5.1%), 12 km secondary roads (4.7%), 83 km tertiary roads (33.6%) and 139 km local roads (56.6%). Road density is lower in this study area compared to the study area in Canada. Primary roads were relatively well detected in all models (completeness = 54%–95%; Figure 17). The prediction of secondary roads was the most incomplete (completeness = 5%–19%). The sub-sets contained only one secondary road, which was located alongside or within open fields and, therefore, was not clearly distinctive from its surroundings, in backscatter values. Tertiary and local roads were also challenging to detect (completeness = 37%–60%). These roads were often narrow, partly or fully covered by trees, and did not always show a clear difference with surrounding backscatter values. With VH input the model showed the highest completeness score for all road types except secondary roads.

Figure 17
Completeness of the model predictions per road type for the Sangha region, Congo



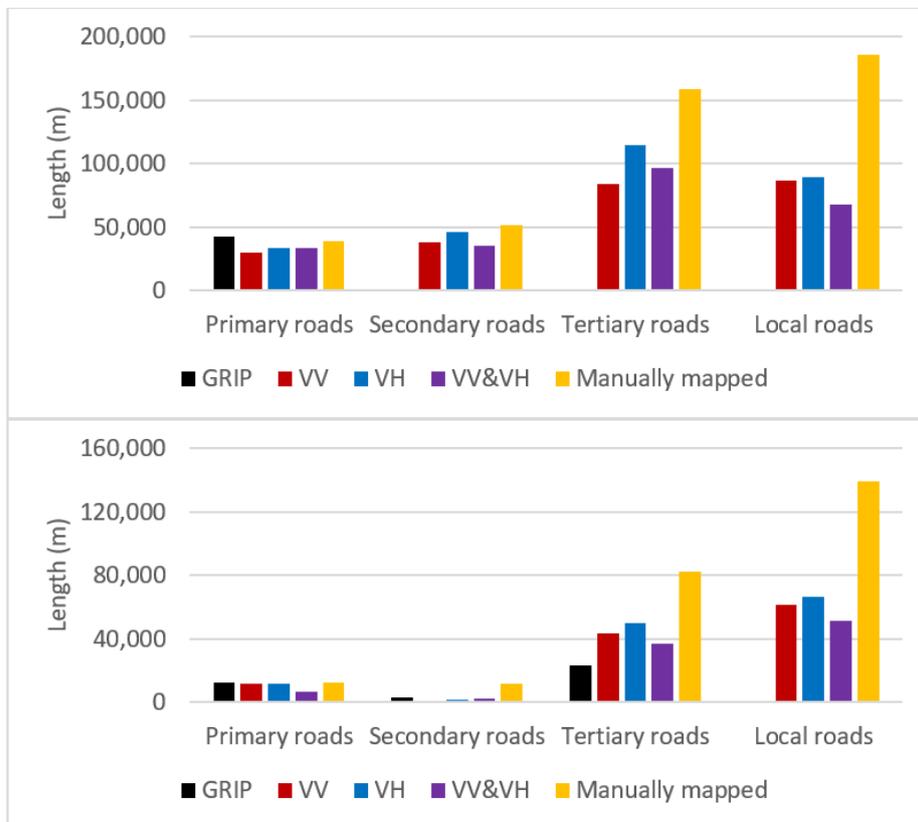
This figure shows the completeness (%) of each road type in model predictions, i.e. the percentage of actual roads captured by the model, with different model inputs (VV, VH or VV&VH).

3.2 Comparing model results with the GRIP data set

Two things stand out when comparing the model results with the GRIP data set. Firstly, overall, the model results include more roads than the GRIP data set (Figures 19 and 20). Although primary roads are well represented in the GRIP data set, this holds much less for tertiary or local roads. Remarkably, the GRIP data set does not include local roads in the study areas (Appendix D), while these are at least partly detected by the model.

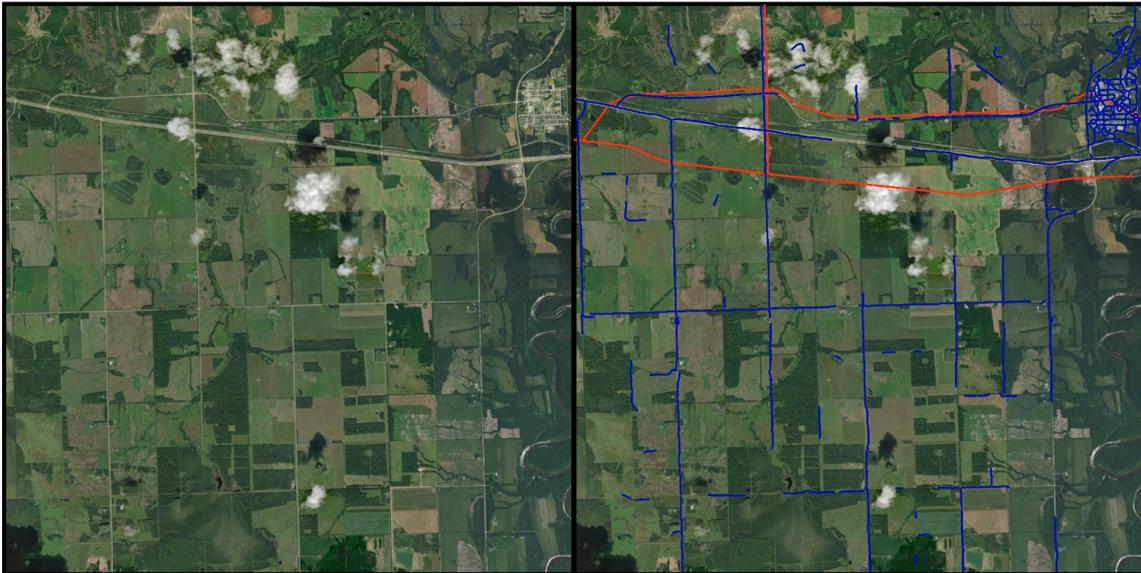
Further, the locations of the roads in the GRIP data set may deviate from the actual locations (Figures 20 and 21). This can be due to positional inaccuracy of the GRIP data (Figure 20) or the appearance and disappearance of roads (Figure 21). The latter is especially evident in Congo, where new roads are constructed for logging activities and old logging roads are abandoned. When zooming in on the area that contained roads according to the GRIP data set, but no roads according to the model, it became apparent that those roads were not there anymore, as can be seen on satellite imagery. The roads detected by the model, but which were not in the GRIP data set, were indeed visible on satellite imagery. These differences reflect that existing road maps, which formed the basis for the GRIP data set, have a high probability of being outdated particularly in areas subject to considerable dynamics in spatial planning or resource use.

Figure 19
Comparison of total road lengths



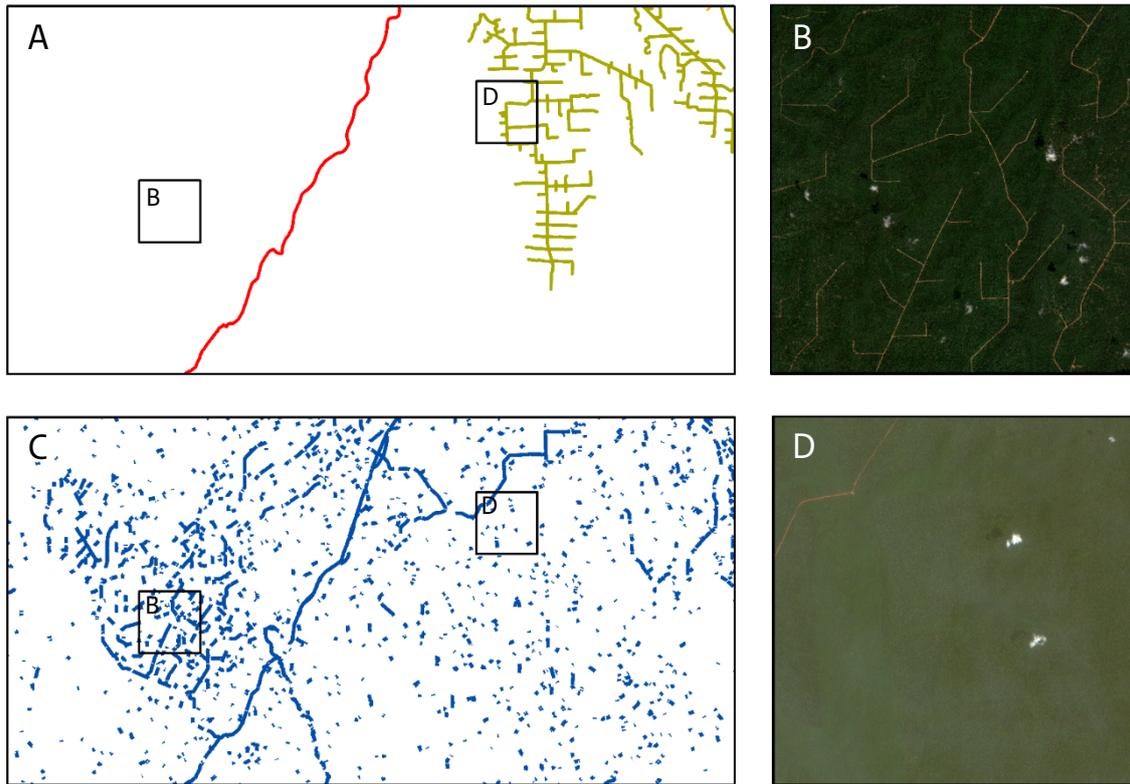
This figure shows two graphs on total road length per road class in the four sub-sets on Canada (top image) and Congo (bottom image), as predicted by the model, manually mapped, and according to the GRIP data.

Figure 20
Example of roads in the GRIP data set and in model predictions



This figure shows two maps of the sub-set Evansburg in Canada. The map on the left is the base map from ArcMap, and the one on the right has a GRIP data overlay (the red lines) and VH model results (the blue lines).

Figure 21
Examples of the appearance and disappearance of roads



This figure shows four maps as examples of differences between the GRIP data set (map A) and model result with VH input (map C). The primary road in the GRIP data set (shown as a red line) is also visible in the model results, while the tertiary roads in the GRIP data set (in green) are not (e.g. the black outlined square D). The model shows a high density of roads, whereas the GRIP data set does not show any roads (e.g. the black outlined square B, which is the sub-set Liouesso). Maps B and D show the presence or absence of roads as visible on Sentinel-2 satellite imagery in the corresponding squares. The image in map B was taken on 6 December 2020 and the image in map D was taken on 1 March 2021.

4 Discussion

The aim of this study was to explore the potential of automated road extraction from remote sensing imagery to create a more complete and up-to-date global road network data set, based on the deep learning model from Stewart et al. (2020) and Sentinel-1 SAR imagery. As the model was created for desert areas, its applicability to other environments was not yet known.

The model gave similar accuracy results for the study areas in Congo and Canada, despite the differences in environments, demonstrating the robustness of the methodology. The model showed the highest completeness and overall accuracy (F1, IoU) with VH as input. Correctness and rank distance were highest with VV&VH as input (Table 4). The accuracy results of this study are somewhat lower than those of Stewart et al. (2020), who used the same model to map roads in deserts. In two of the three desert regions their model was most accurate using VV as input, with completeness of 64%–71%, correctness of 81%–88%, IoU of 68%–89% and rank distance of 76%–80%. The higher model accuracy in Stewart et al. (2020) was to be expected as the model parameters were optimised for these desert regions. Optimisation of the parameters was avoided in this study, as this process of trial and error would deteriorate the global applicability of the method. Zhang et al. (2019) also detected roads on Sentinel-1 SAR imagery in Beijing using a U-Net model trained with handmade training data. They found VV&VH input to be the most accurate (completeness = 92%, correctness = 95%, F1 = 0.94). Hence, no specific SAR polarisation stands out as most suitable for road detection. The preferred polarisation may differ per model setting and study region, and with the quality of the layer used. The high accuracy achieved by Zhang et al. (2019) could be due to the high-quality manually collected training data. In this study, existing and publicly available training data were used, making the methodology more generic and cost-effective. Though the quality of the training data varied regarding positional accuracy and completeness/correctness, the model was able to discard old or positionally inaccurate roads in the training data and detect new roads on the SAR imagery.

The results show that both large (i.e. primary) and small (i.e. local) roads can be detected on Sentinel-1 SAR imagery, though in general it was easier for the model to detect large roads compared to small roads. The latter were less distinctive in the radar imagery as they can be narrow, unpaved, partially covered by trees, and did not always show a clear difference with surrounding backscatter values. Not all roads in the areas were detected by the model (false negatives) and non-roads were occasionally classified as roads (false positives). The causes of these errors are similar to those mentioned in other studies detecting roads. Objects with similar characteristics as roads (e.g. tree hedges, rivers), variability in the appearance of roads, complex surroundings or coverage of roads are problematic for detection of roads on remote sensing imagery in general (Wang et al., 2016; Xu et al., 2018). Stewart et al. (2020) also faced false negatives related to the context of the roads, for example if a road ran alongside other infrastructures or through areas with similar backscatter, such as fields or bare ground. Kearney et al. (2020) mapped roads in Alberta using a Convolutional Neural Network and 5m resolution optical data. Their model also misclassified rivers as roads and missed roads due to surrounding exposed soil and rock or occlusion by forest canopy. Henry et al. (2020) performed road detection on higher resolution (1m) SAR imagery with different types of Convolutional Neural Networks. In their study, even the models achieving the best results (completeness = 51%–75%, correctness = 51%–72%, IoU = 40%–46%) had difficulties, with false positives resulting from mounds and forest borders and false negatives due to less visible roads. The errors are thus not specific to the methods or areas used in this study, but are encountered with road mapping from satellite imagery in general.

Compared to the current GRIP data set, the model detected more roads in the study areas. The comparison showed that especially local roads are missing in the GRIP data set, which confirms the assumption of Meijer et al. (2018). Furthermore, GRIP data are not always positionally accurate, whereas the roads detected by the model are positionally accurate as they are derived from georeferenced remote sensing data. The comparison with the GRIP data further showed changes in the network of roads in Congo. GRIP data contain roads which no longer exist, while new roads are not represented. This example highlights that road detection from recent remote sensing imagery has the advantage of providing up-to-date information. This is especially important in hitherto pristine areas with fast developing road networks, as the Congo and Amazon Basins (Barber et al., 2014; Laurance et al., 2014; Laurance et al., 2017). However, although the modelling approach provides more temporally and spatially detailed data than the GRIP data set, it lacks the distinction of road type. This can be important for applications. For example in GLOBIO, road impact is assumed to be caused by highways, primary roads and secondary roads. Tertiary and local roads would induce much less avoidance behaviour by wildlife (Brehme et al., 2013).

The automated road extraction with remote sensing has shown promising results, as this proved able to capture roads that were not in the GRIP data set, leave out abandoned roads and have high positional accuracy. Therefore, it can be a valuable asset to mapping and monitoring road networks. However, the method proposed here is a prototype and further research is needed to improve the accuracy. Regarding inputs, the addition of optical Sentinel-2 imagery could be considered. Grass/dirt strips, rivers, or certain terrain edges, now often confused with roads and leading to false positives, may be better distinguishable from roads using optical imagery compared to the SAR imagery. Nonetheless, Sentinel-2 imagery often contains clouds, especially in tropical regions. Removal of clouds and mosaicking of images is necessary to create a cloud-free composite usable as input. Such a layer could be self-produced, or existing cloud-free composites could be considered. Another option to improve the model is applying more extensive post-processing. Post-processing in this study was limited to the creation of line segments in ArcMap, with automatic removal of small or not elongated features and connection of gaps smaller than 100 meters. A more refined post-processing method was for example created by Das et al. (2011) and applied by Kearney et al. (2020), based on both the binary prediction and probability output of the model. It involved removal and connection of pixel groups based on probability, eccentricity and size using thresholds and filters. As these thresholds may be site dependent, Kearney et al. (2020) proposed the use of cross-validated approaches to identify thresholds to preserve an automated approach. Their post-processing improved the overall quality (IoU) of the road extraction by 13%. Successful post-processing could also improve the correctness of topology in the final result as in general road networks are not suddenly interrupted. This is for example important when studying habitat fragmentation by roads. Additional post-processing could include the removal of false positives by using optical imagery and indices such as the Normalized Difference Vegetation Index or Normalized Difference Water Index. Compared to the GRIP data set, automated road extraction applied in this study does not differentiate between road types. No studies were found distinguishing road types based on remote sensing imagery. Though the road types cannot be distinguished, for example by number of lanes, the number of cars or the material of roads on Sentinel-1 SAR imagery, other characteristics may be useful to assign road types. Examples are average width or length of road segments, optical characteristics (i.e. road surface material) or topological characteristics within the road network.

Wang et al. (2016), comparing different methods to extract roads, concluded that the realisation of completely automatic road extraction from remote sensing imagery remains challenging and

stressed the difficulty of solving all the existing problems in road extraction. This is also evident from this study, as similar problems were faced. According to Wang et al. (2016), neural networks performed well compared to other automated and semi-automated methods. The use of deep learning and specifically Convolutional Neural Networks for image segmentation has expanded greatly within the past 5 years (Kearney et al., 2020). Further developments may provide more suitable model architectures for road extraction from remote sensing imagery and are important to keep an eye out for.

5 Conclusion

The aim of this study was to test the potential of automated road extraction from remote sensing imagery to create a more complete and up-to-date global road network data set, based on Sentinel-1 SAR imagery and the deep learning model from Stewart et al. (2020). This model was created for desert areas, and the applicability to study areas in Canada and the Republic of Congo was tested in this study. Model results were evaluated regarding different SAR inputs (VV, VH, VV&VH). The model gave similar accuracy results in the two study areas despite the differences in environments. The model showed the highest completeness and overall accuracy (F1, IoU) with VH as input. Correctness and rank distance were highest with VV&VH as input. Limitations of the methodology are indicated by the presence of false negatives and false positives. These were caused by objects with similar characteristics as roads (e.g. tree hedges, rivers), complex surroundings (i.e. fields) or coverage of roads, and reflect problems commonly encountered in the detection of roads on remote sensing imagery. Compared to the data in the GRIP roads data set, the model results were positionally more accurate and up to date. They also showed more roads in the study areas than indicated in the GRIP data set. The method proposed here is a prototype and further research is needed to improve the accuracy. Improvements may be the use of optical Sentinel-2 imagery or other post-processing methods. Further research may also look into the possibilities to differentiate between road types, as in the GRIP data set. If the challenges encountered can be solved, the methodology for automated road detection has the potential to provide a robust, cost-effective and globally applicable solution for the mapping and monitoring of road networks.

References

- Abderrahim NYQ, Abderrahim S and Rida A. (2020). Road Segmentation using U-Net architecture. In *2020 IEEE International conference of Moroccan Geomatics (Morgeo)* (pp. 1–4). IEEE.
- Ali I, Cao S, Naeimi V, Paulik C and Wagner W. (2018). Methods to remove the border noise from Sentinel-1 synthetic aperture radar data: implications and importance for time-series analysis. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(3), 777–786.
- Alshehhi R, Marpu PR, Woon WL and Dalla Mura M. (2017). Simultaneous extraction of roads and buildings in remote sensing imagery with convolutional neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 130, 139–149.
- Barber CP, Cochrane MA, Souza Jr CM and Laurance WF. (2014). Roads, deforestation, and the mitigating effect of protected areas in the Amazon. *Biological conservation*, 177, 203–209.
- Barrington-Leigh C and Millard-Ball A. (2017). The world’s user-generated road map is more than 80% complete. *PLoS one*, 12(8), e0180698.
- Benítez-López A, Alkemade R and Verweij PA. (2010). The impacts of roads and other infrastructure on mammal and bird populations: a meta-analysis. *Biological conservation*, 143(6), 1307–1316.
- Brehme CS, Tracey JA, McClenaghan LR and Fisher RN. (2013). Permeability of roads to movement of scrubland lizards and small mammals. *Conservation Biology*, 27(4), 710–720.
- Center For International Earth Science Information Network (Ciesin)–Columbia University, and Information Technology Outreach Services (Itos)–University of Georgia (2013). *Global Roads Open Access Data Set. Version 1 (gROADSv1)*. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC).
- Cheng G, Wang Y, Xu S, Wang H, Xiang S and Pan C. (2017). Automatic road detection and centerline extraction via cascaded end-to-end convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 55(6), 3322–3337.
- Das S, Mirnalinee TT and Varghese K. (2011). Use of salient features for the design of a multistage framework to extract roads from high-resolution multispectral satellite images. *IEEE transactions on Geoscience and Remote sensing*, 49(10), 3906–3931.
- Filipponi F. (2019). Sentinel-1 GRD preprocessing workflow. In *Multidisciplinary Digital Publishing Institute Proceedings*, 18(1), 11.
- Gao L, Song W, Dai J and Chen Y. (2019). Road extraction from high-resolution remote sensing imagery using refined deep residual convolutional neural network. *Remote Sensing*, 11(5), 552.
- Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D and Moore R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote sensing of Environment*, 202, 18–27.
- Hajduch G and Miranda N. (2018). Masking ‘No-Value’ Pixels on GRD Products Generated by the Sentinel-1 ESA IPF. *Document Reference MPC-0243*. Paris, France: Collecte Localisation Satellites; European Space Agency.
- Henry C, Azimi SM and Merkle N. (2018). Road segmentation in SAR satellite images with deep fully convolutional neural networks. *IEEE Geoscience and Remote Sensing Letters*, 15(12), 1867–1871.
- Huang J and Nowack RL. (2020). Machine Learning Using U-Net Convolutional Neural Networks for the Imaging of Sparse Seismic Data. *Pure and Applied Geophysics*, 1–16.

- IPBES (2016). *Assessment report on scenarios and models of biodiversity and ecosystem services*. Bonn, Germany: Secretariat of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services.
- IPBES (2019). *Global assessment report on biodiversity and ecosystem services*. Bonn, Germany: Secretariat of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services.
- IRF (2017). *World Road Statistics 2016*. Geneva: International Road Federation.
- IUCN (2012). World heritage nomination – IUCN technical evaluation Sangha trinational [online]. Retrieved from <https://whc.unesco.org/en/list/1380/> [accessed on 15 February 2021].
- Jiang Z. (2007). The road extension model in the land change modeler for ecological sustainability of IDRISI. In *Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems* (pp. 1–8).
- Johnson BD and Singh J. (2003). Building the National GeoBase for Canada. *Photogrammetric Engineering & Remote Sensing*, 69(10), 1169–1173.
- Jones N. (2014). Computer science: The learning machines [online]. Retrieved from <http://www.nature.com/news/505146a-i2-jpg-7.14689?article=1.14481> [accessed on 18 February 2021].
- Kearney SP, Coops NC, Sethi S and Stenhouse GB. (2020). Maintaining accurate, current, rural road network data: An extraction and updating routine using RapidEye, participatory GIS and deep learning. *International Journal of Applied Earth Observation and Geoinformation*, 87, 102031.
- Laurance WF, Campbell MJ, Alamgir M and Mahmoud MI. (2017). Road expansion and the fate of Africa's tropical forests. *Frontiers in Ecology and Evolution*, 5, 75.
- Laurance WF, Clements GR, Sloan S, O'connell CS, Mueller ND, Goosem M, Venter O, Edwards DP, Phalan B, Balmford A, Van Der Ree R and Arrea IB. (2014). A global strategy for road building. *Nature*, 513(7517), 229–232.
- Laurance WF, Goosem M and Laurance SG. (2009). Impacts of roads and linear clearings on tropical forests. *Trends in ecology & evolution*, 24(12), 659–669.
- Mahdi A, Qin J and Crosby G. (2019). DeepFeat: A bottom-up and top-down saliency model based on deep features of convolutional neural networks. *IEEE Transactions on Cognitive and Developmental Systems*, 12(1), 54–63.
- Meijer JR, Huijbregts MA, Schotten KC and Schipper AM. (2018). Global patterns of current and future road infrastructure. *Environmental Research Letters*, 13(6), 064006.
- National Center for Statistics and Economic Studies (2007). Population, People Republic of Congo [online]. Retrieved from <https://rcongo.opendataforafrica.org/mjyvlxe/population-people-republic-of-congo-2012> [accessed on 8 February 2021].
- Nelson A, de Sherbinin A and Pozzi F. (2006). Towards development of a high quality public domain global roads database. *Data Science Journal*, 5, 223–265.
- NIMA (1997). *Vector Map Level 0, 3rd edition of the Digital Chart of the World*. Fairfax, VA: National Imagery and Mapping Agency.
- Panboonyuen T, Jitkajornwanich K, Lawawirojwong S, Srestasathiern P and Vateekul P. (2017). Road segmentation of remotely-sensed images using deep convolutional neural networks with landscape metrics and conditional random fields. *Remote Sensing*, 9(7), 680.
- Paul J. (2020). Semantic Segmentation of Roads in Aerial Imagery [online]. Retrieved via Stewart et al., 2020 from <https://github.com/Polymorphous/skeyenet> (accessed by Stewart et al. (2020) on 20 May 2020).

- PBL (2016). *The GLOBIO Model. A technical description of version 3.5*. PBL Netherlands Environmental Assessment Agency.
- Ronneberger O, Fischer P and Brox T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234–241). Springer, Cham.
- Schipper AM, Hilbers JP, Meijer JR, Antão LH, Benítez-López A, de Jonge MMJ, Leemans LH, Scheper E, Alkemade R, Doelman JC, Mylius S, Stehfest E, Van Vuuren DP, Van Zeist W and Huijbregts MAJ. (2020). Projecting terrestrial biodiversity intactness with GLOBIO 4. *Global change biology*, 26(2), 760–771.
- Şen N, Olgun O and Ayhan Ö. (2019). Road and railway detection in SAR images using deep learning. In *Image and Signal Processing for Remote Sensing XXV* (Vol. 11155, p. 111550D). International Society for Optics and Photonics.
- Shi Q, Liu X and Li X. (2017). Road detection from remote sensing images by generative adversarial networks. *IEEE access*, 6, 25486–25494.
- Shorten C and Khoshgoftaar TM. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1), 1–48.
- Statistics Canada (2016). Population and Dwelling Count Highlight Tables, 2016 Census [online]. Retrieved from <https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/hlt-fst/pd-pl/Table.cfm?Lang=Eng&T=701&SR=1&S=3&O=D&RPP=25> [accessed on 4 February 2021].
- Statistics Canada (2020). National Road Network [online]. Retrieved from <https://www.statcan.gc.ca/eng/consultation/2019/nrn> [accessed on 8 February 2021].
- Ubukawa T, de Sherbinin A, Onsrud H, Nelson A, Payne K, Cottray O and Maron M. (2014). A review of roads data development methodologies. *Data Science Journal*, 14-001.
- Wang W, Yang N, Zhang Y, Wang F, Cao T and Eklund P. (2016). A review of road extraction from remote sensing images. *Journal of traffic and transportation engineering (English edition)*, 3(3), 271–282.
- WWF (2020). WWF in the Republic of Congo [online]. Retrieved from https://www.wwf-congobasin.org/where_we_work/republic_of_congo/ [accessed on 15 February 2021].
- Xu Y, Xie Z, Feng Y and Chen Z. (2018). Road extraction from high-resolution remote sensing imagery using deep learning. *Remote Sensing*, 10(9), 1461.
- Yang X, Li X, Ye Y, Lau RY, Zhang X and Huang X. (2019). Road detection and centerline extraction via deep recurrent convolutional neural network U-Net. *IEEE Transactions on Geoscience and Remote Sensing*, 57(9), 7209–7220.
- Zhang Z, Liu Q and Wang Y. (2018). Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters*, 15(5), 749–753.
- Zhang Q, Kong Q, Zhang C, You S, Wei H, Sun R and Li L. (2019). A new road extraction method using Sentinel-1 SAR images based on the deep fully convolutional neural network. *European Journal of Remote Sensing*, 52(1), 572–582.

Appendices

Appendix A: Set-up Azure

Model runs were performed in an Azure Virtual Machine with an Intel Xeon CPU E5-2667 v3 and memory of 112 GB. Anaconda Navigator version 1.10.0 was installed, and a Python 3.7.9 environment was created with the necessary packages (Table A). The Tensorflow and Keras versions optimal for CPU were used. Then Spyder 4.1.5 was installed with this environment. The model was run within Spyder. Model runs took approximately 16 hours. QGIS was also installed to check model outputs.

Table A
Packages and versions installed in the Python environment

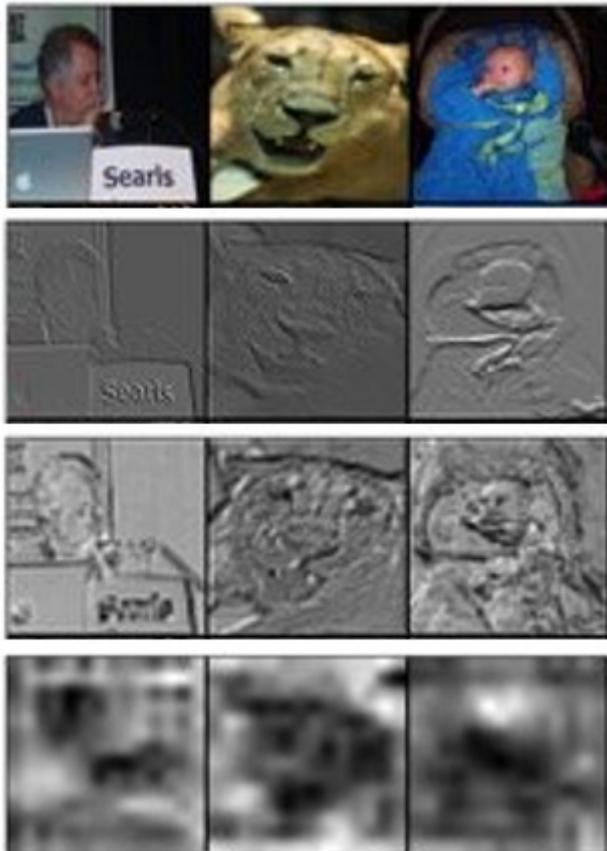
Package	Version
Gdal	3.0.2
Geotiff	1.5.1
Keras	1.0.8
Matplotlib	3.3.2
Rasterio	1.1.0
Tensorflow-mkl	2.3.0

Packages installed in addition to the standard available when creating a Python 3.7 environment in Anaconda.

Appendix B: Examples of deep learning

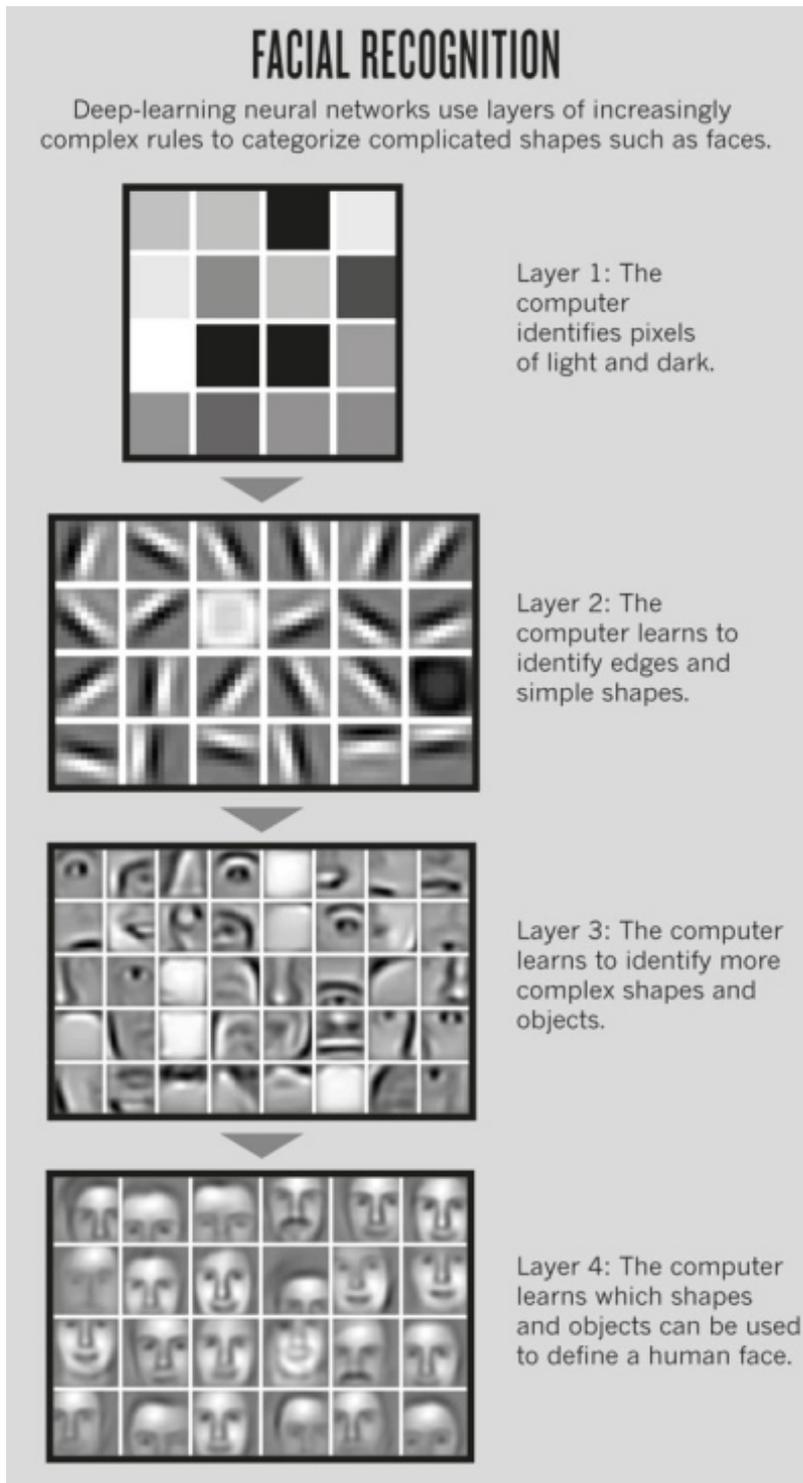
Figures B1 and B2 show examples of the workings of deep learning models. For more information on deep learning for remote sensing the reader is referred to the 2020 course Advanced Machine Learning for Remote Sensing by R. Roscher from the University of Bonn. This course is available on YouTube with tutorials explaining the basics of machine learning, neural networks, explanations of the possible settings and parameters, and how model training works.

Figure B1:
Example of features of different convolutional layers within a CNN



This figure shows four rows of images. The top row contains the originals, and the other rows show the layers deeper within the network. Border definitions can be seen in the images in the second row, while parts of objects are recognisable in the bottom row. Adapted from Mahdi et al. (2019).

Figure B2
Example of facial recognition using deep learning

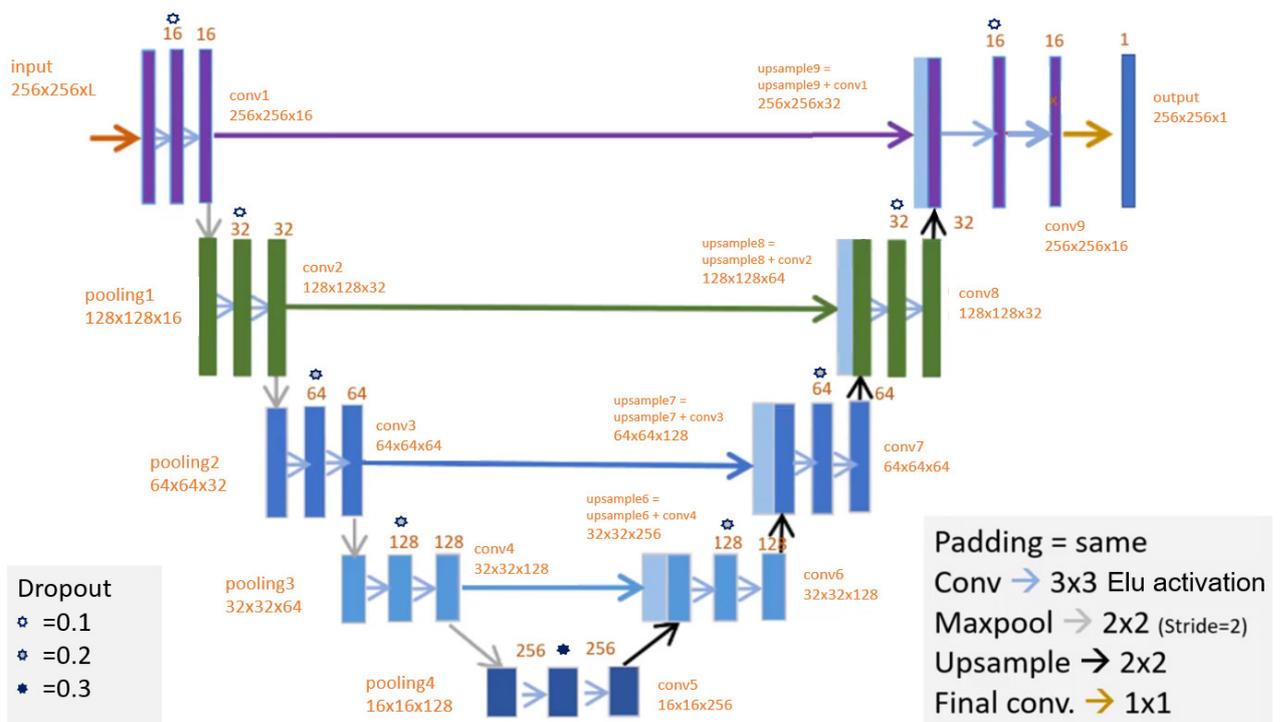


This figure contains four images showing examples of layers in facial recognition using deep-learning neural networks, from Jones (2014).

Appendix C: Overview of the U-Net model from Stewart et al. (2020)

An overview of the model used in this study and adapted from Stewart et al. (2020) is visible in Figure C. The convolutional layers all have 3x3 pixel windows, with exponential linear units as the activation and He normal initialiser. Only the last output layer is a 1x1 pixel convolutional layer with sigmoid activation. Batch normalisation occurred after each convolution. Dropout layers were also included to prevent overfitting. The total number of trainable parameters throughout the model were 1,943,761 using 1 input layer and 1,943,905 using 2 input layers. The models were initiated with random weights. These weights are adapted during model training. Model training is an iterative process during which the trainable parameters are changed. Optimisation is reached by moving toward the minimum value of the loss function. The loss function represents the difference between the estimate by the model and the reference. The loss function in the model is the soft Dice loss, which takes into account the class imbalance of road and non-road pixels. For the same reason, Intersection over Union (IoU) assesses the performance of the model during training.

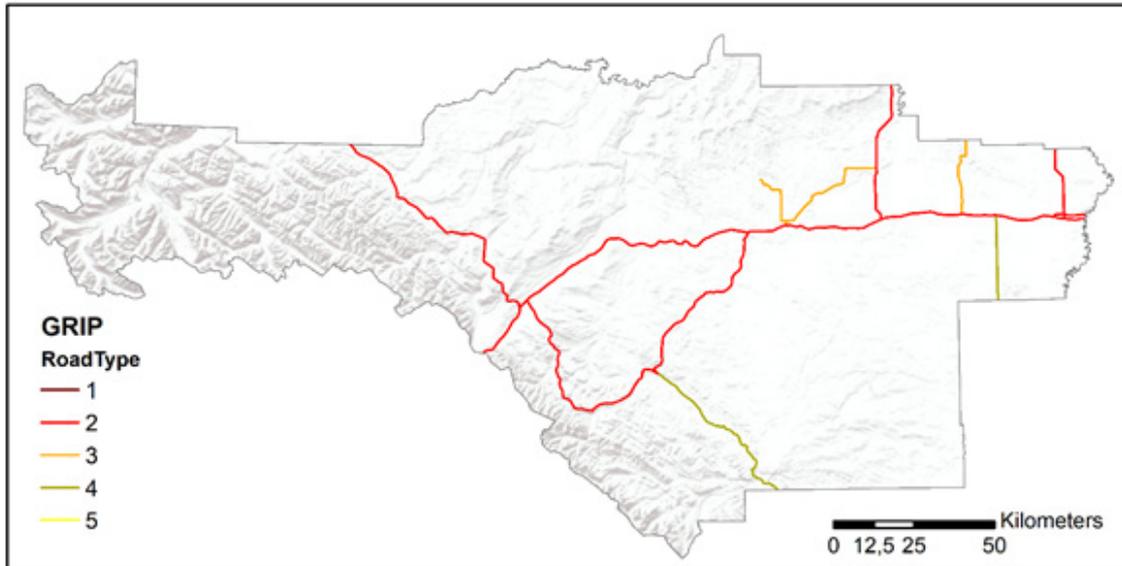
Figure C
U-Net architecture used in this study



The figure shows a U-shaped schematic, the left half of which is the down-sampling encoder part, and the right half is the up-sampling decoder part. Horizontal arrows on the inside of the U-shape are the cross-connections. Each bar corresponds to a multi-channel feature map. The number of channels is denoted on top of the bar. The size of inputs is provided to the left of the bars, the size of outputs is to the right of the bars. Light blue bars represent copied feature maps. The arrows denote the different operations. Stars indicate dropout in between the operations.

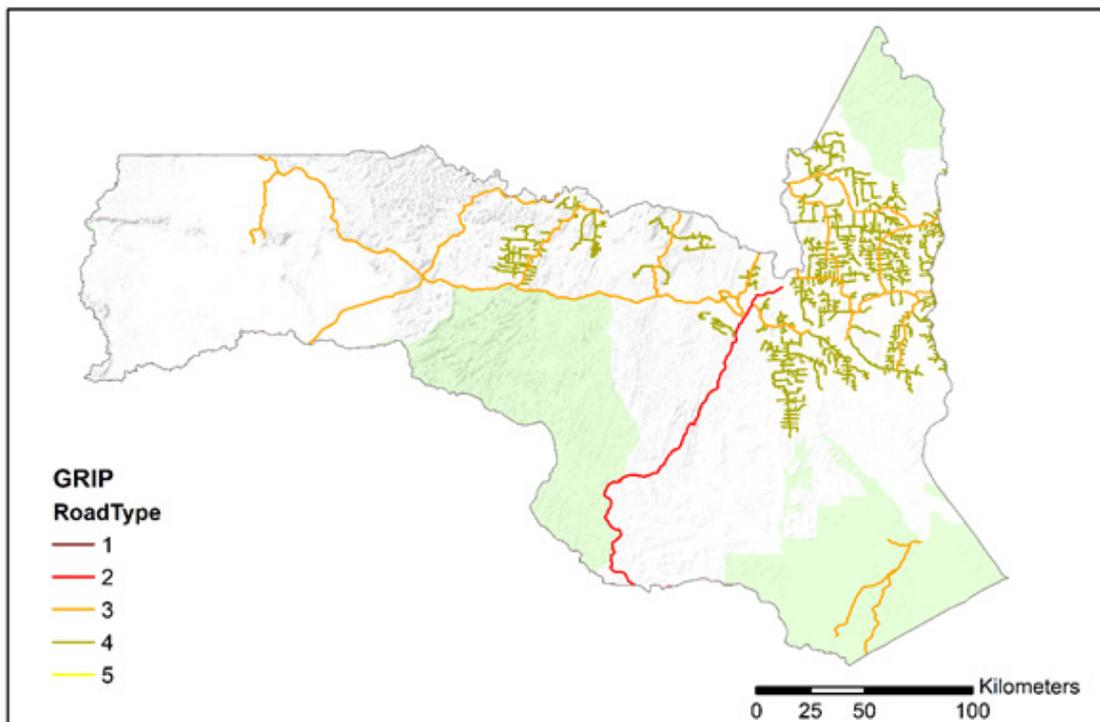
Appendix D: GRIP data on the study areas

Figure D1
GRIP data on Division No. 14 Alberta, Canada



This figure shows the location of various road types, indicated in colours, as in the GRIP data set for Division No. 14 Alberta, Canada.

Figure D2
GRIP data on the Sangha region, Congo



This figure shows the location of various road types, indicated in colours, as in the GRIP data set, for the Sangha region, Congo. Protected areas are shown in light green (WDPA, 2021).

Appendix E: Accuracy results per sub-set in Division No. 14 Alberta, Canada

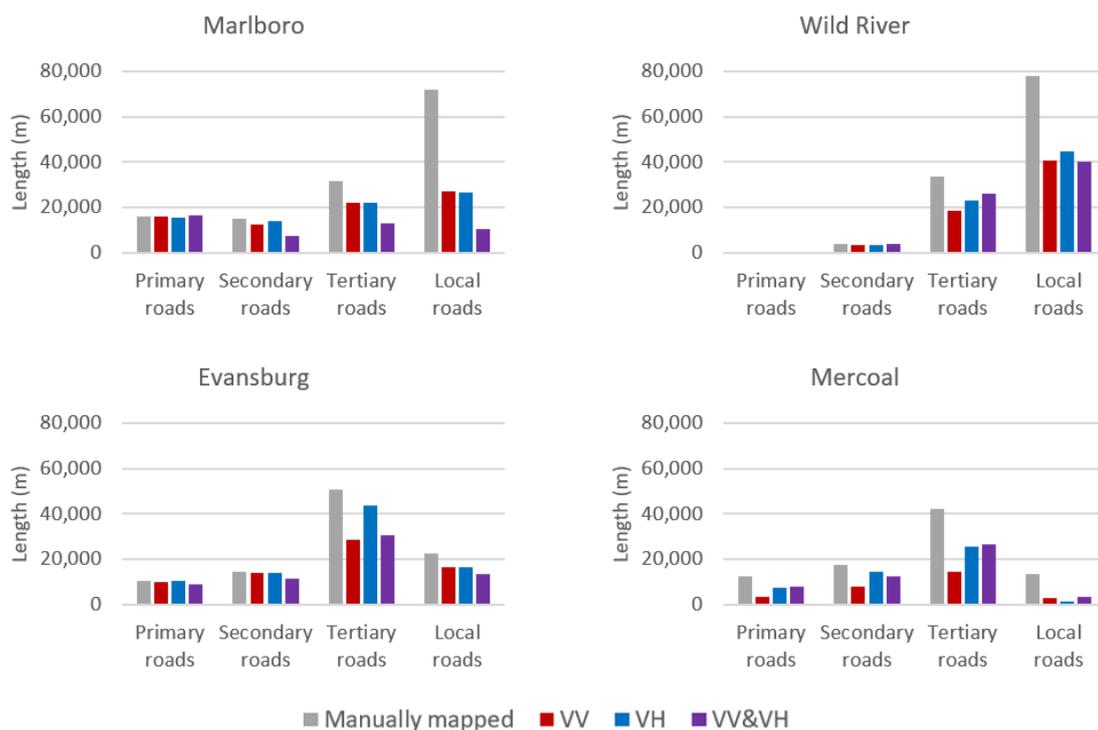
For the location of the sub-sets, see Figure 10.

Table E
Accuracy results per sub-set in Division No. 14, Canada

Index	Marlboro			Wild River			Evansburg			Mercoal		
	VV	VH	VV & VH	VV	VH	VV & VH	VV	VH	VV & VH	VV	VH	VV & VH
Completeness	0.58	0.58	0.35	0.54	0.62	0.61	0.70	0.86	0.66	0.34	0.57	0.59
Correctness	0.68	0.67	0.87	0.69	0.75	0.83	0.67	0.66	0.81	0.45	0.53	0.70
IoU	0.46	0.45	0.33	0.44	0.51	0.54	0.52	0.60	0.58	0.24	0.38	0.47
F1	0.63	0.62	0.50	0.61	0.67	0.70	0.68	0.75	0.73	0.39	0.55	0.64
Rank distance	0.63	0.63	0.66	0.62	0.68	0.73	0.68	0.77	0.74	0.40	0.55	0.65

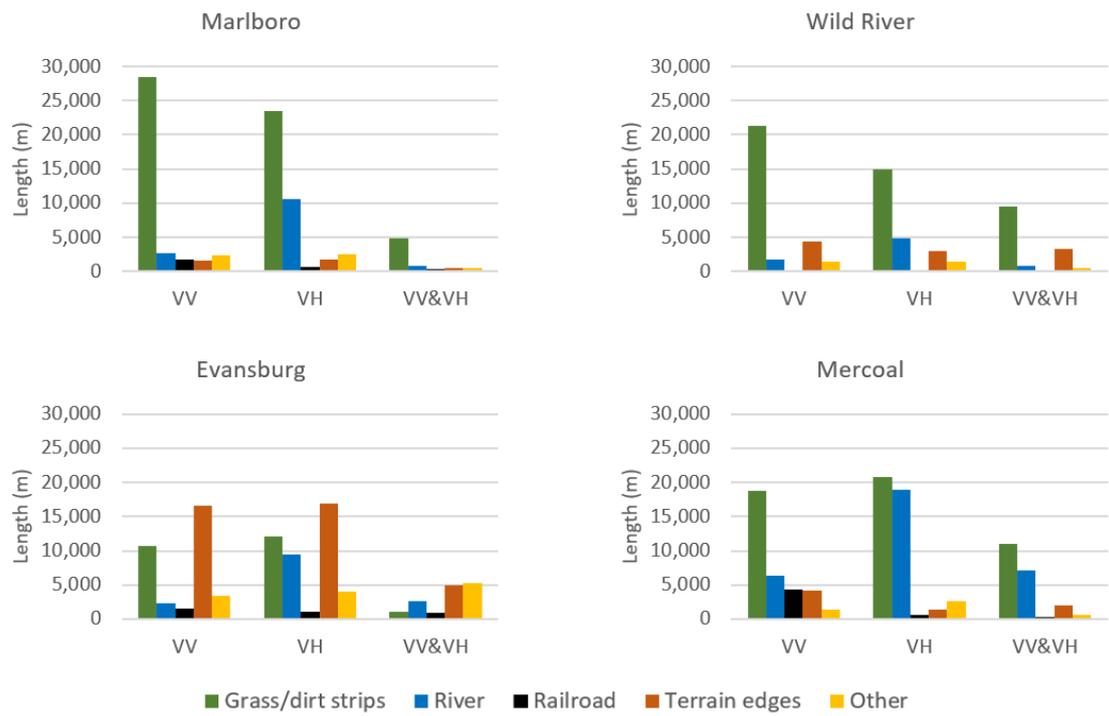
Accuracy results per sub-set, per input, for various accuracy indices. Colours indicate value (lower=red, higher=green).

Figure E1
Total road length per sub-set in Division No. 14, Canada



This figure shows the total length of roads, manually mapped, and per model prediction, for each road class in four sub-sets. It shows the completeness of road types in the model predictions.

Figure E2
False positives per sub-set in Division No. 14, Canada



This figure shows the total length of false positives, in model predictions, in four sub-sets.

Appendix F: Accuracy results per sub-set in the Sangha Department, Republic of the Congo

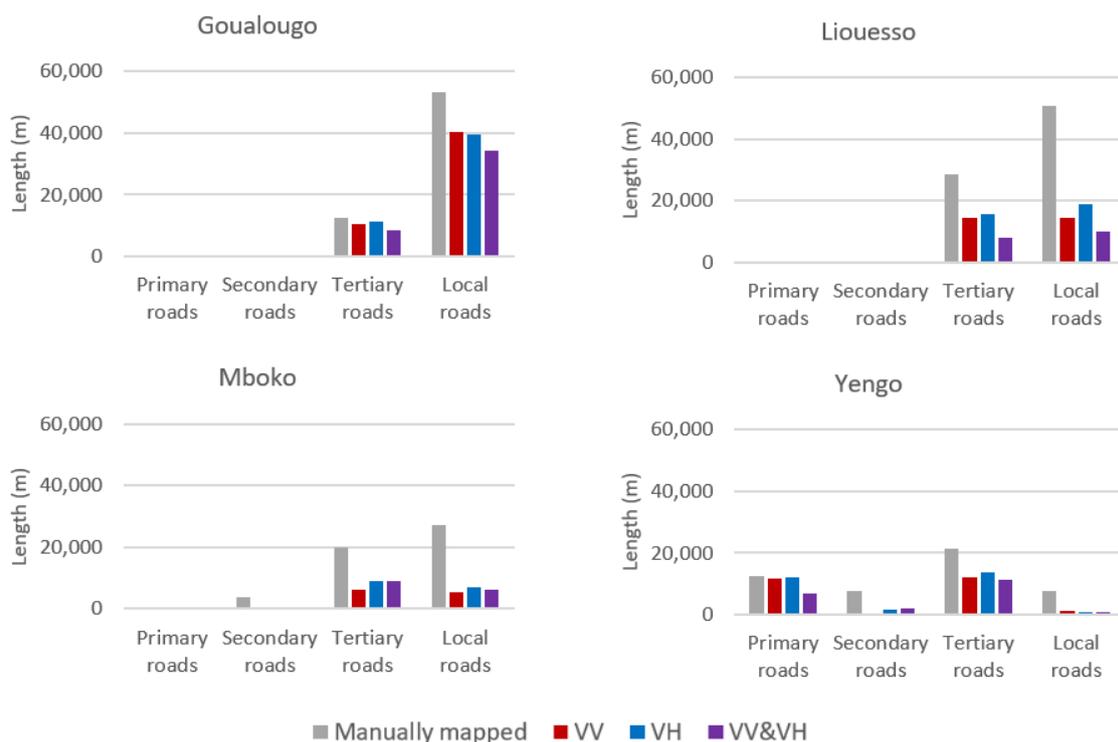
For the location of the sub-sets see Figure 11.

Table F
Accuracy results per sub-set in the Sangha region, Congo

Index	Goulougo			Liouesso			Mboko			Yengo		
	VV	VH	VV & VH	VV	VH	VV & VH	VV	VH	VV & VH	VV	VH	VV & VH
Completeness	0.77	0.77	0.65	0.36	0.43	0.23	0.23	0.32	0.29	0.52	0.57	0.43
Correctness	0.84	0.77	0.88	0.96	0.88	0.99	0.63	0.50	0.60	0.77	0.65	0.99
IoU	0.67	0.63	0.60	0.35	0.41	0.23	0.20	0.24	0.25	0.45	0.43	0.43
F1	0.80	0.77	0.75	0.52	0.58	0.37	0.34	0.39	0.39	0.62	0.60	0.60
Rank distance	0.81	0.77	0.77	0.73	0.69	0.72	0.48	0.42	0.47	0.65	0.61	0.76

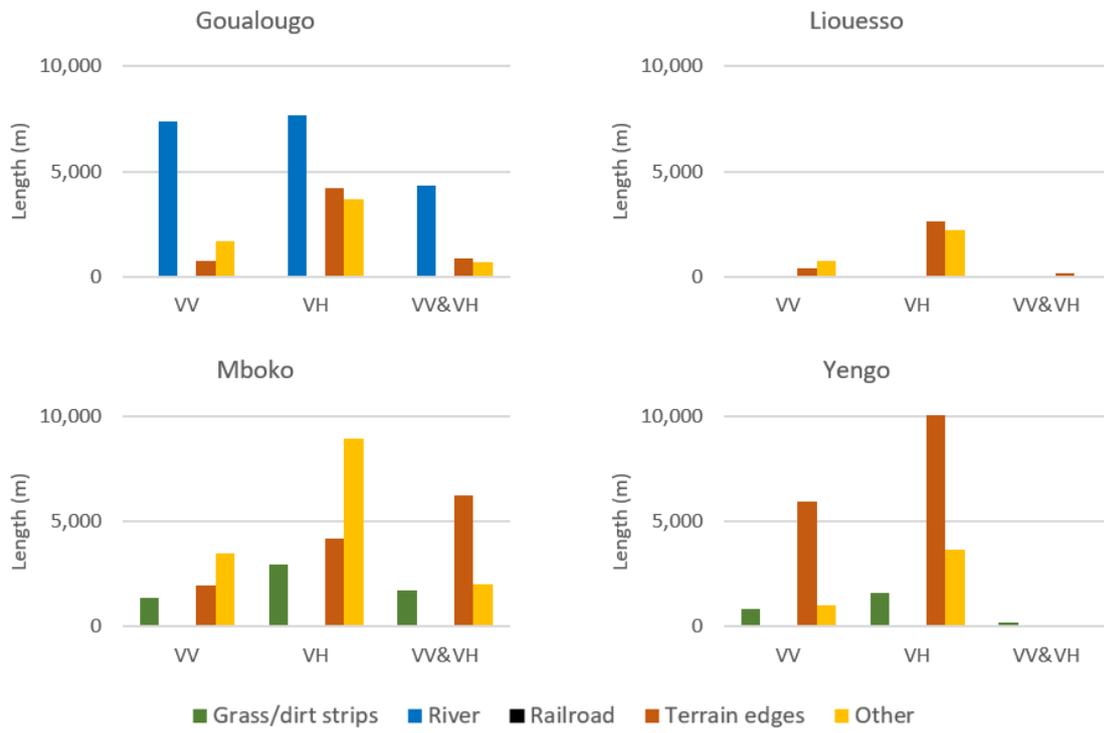
Accuracy results per sub-set, per input, for various accuracy indices. Colours indicate value (lower=red, higher=green).

Figure F1
Total road length per sub-set in the Sangha region, Congo



This figure shows the total length of roads, manually mapped, and per model prediction, for each road class in four sub-sets. It shows the completeness of road types in the model predictions.

Figure F2
False positives per sub-set in the Sangha region, Congo



This figure shows the total length of false positives, in model predictions, in four sub-sets.