

Landesweite Vegetationshöhenmodelle mit Deep Learning und Sentinel-2

Basierend auf multispektralen Sentinel-2 Bildern werden landesweite Vegetationshöhenmodelle für die Schweiz und Gabun erstellt. Dafür wird ein Convolutional Neural Network (CNN) mittels bestehenden Referenzdaten trainiert, um charakteristische Merkmale aus der Textur und der spektralen Signatur zu extrahieren und die Vegetationshöhe mittels Regression zu schätzen. Während in Gabun LiDAR Daten aus fünf Regionen als Referenz dienen, benutzen wir in der Schweiz das Vegetationshöhenmodell des Landesforstinventars (Ginzler und Hobi, 2015), welches aus photogrammetrischen Stereomessungen abgeleitet wurde. Die erstellten Karten mit einer 10 m Auflösung haben einen mittleren absoluten Fehler von 1.7 m in der Schweiz und 4.3 m in Gabun, wobei Höhen bis zu 55 m korrekt mit einem relativen Fehler von etwa 20% reproduziert werden.

Nous avons développé des modèles de la hauteur de la végétation pour tout le territoire national de la Suisse ainsi que du Gabon, ceux-ci sont basés sur les images multispectrales Sentinel-2. Pour les réaliser, nous entraînons un Convolutional Neural Network (CNN) à l'aide de données de référence existantes afin d'apprendre les signes caractéristiques, à partir de la texture et de la signature spectrale, qui permettent d'estimer la hauteur de la végétation. Pour le Gabon, les données de référence proviennent de cinq régions où ont été menées des campagnes LiDAR, alors que pour la Suisse nous utilisons le modèle de la hauteur de la végétation de l'Inventaire Forestier National (Ginzler et Hobi, 2015) qui se base sur des photos aériennes stéréoscopiques. Les cartes que nous avons créées avec une résolution de 10 m ont une erreur moyenne absolue de 1,7 m en Suisse et de 4,3 m au Gabon tout en reproduisant correctement des hauteurs jusqu'à 55 m avec une erreur relative d'environ 20%.

Abbiamo sviluppato dei modelli altimetrici della vegetazione per tutto il territorio nazionale della Svizzera e del Gabon, quest'ultimi sono basati sulle immagini multispettrali Sentinel-2. Per costruirli abbiamo allenato un Convolutional Neural Network (CNN) con dei dati di riferimento esistenti per estrarre le caratteristiche specifiche, basate sulla struttura spaziale e la firma spettrale, che ci permettono di predire l'altezza della vegetazione. Per il Gabon, i dati di riferimento provengono da cinque regioni dove sono state condotte delle campagne LiDAR, mentre che per la Svizzera abbiamo usato il modello altimetrico della vegetazione dell'Inventario Forestale Nazionale (Ginzler et Hobi, 2015) derivato da stereomisurazioni fotogrammetriche. Le carte che abbiamo creato con una risoluzione di 10 m presentano un errore medio assoluto di 1,7 m in Svizzera e di 4,3 m in Gabon, in cui le altezze fino a 55 m sono riprodotte correttamente con un errore relativo del 20% circa.

*N. Lang, Chr. Ginzler, K. Schindler,
J. D. Wegner*

Einleitung

Wälder erbringen wichtige Dienstleistungen, auf lokaler aber auch globaler Ebene.

Sie dienen als Naherholungsgebiet, Schutzwald vor Naturgefahren, stabilisieren den Boden, bieten Wohnraum für Tiere und Pflanzen und speichern Kohlenstoff, den sie als CO₂ der Atmosphäre entnehmen. Die Vegetationshöhe ist eine der wichtigsten biophysikalischen Varia-

blen, um Waldstrukturen zu charakterisieren. Auch zur grossflächigen Schätzung und Modellierung von Biomasse oder Biodiversität wird diese oft als erklärende Grösse verwendet.

Eine Messung der Vegetationshöhe ist mittels Airborne LiDAR möglich, wobei die Bodenhöhe sowie auch die Höhe des Blätterdaches aus der Punktwolke abgeleitet wird, woraus sich direkt die Vegetationshöhe bestimmen lässt. Die Datenerhebung ist aufwendig und eine landesweite Abdeckung wird oft erst über mehrere Jahre erreicht. Alternativ können digitale Oberflächenmodelle (DOM) mittels photogrammetrischen Stereomessungen aus überlappenden Luftbildern berechnet werden. Diese DOM können mit bestehenden digitalen Terrainmodellen (DTM) kombiniert werden, um die Vegetationshöhen abzuleiten. Dieses Verfahren wurde beispielsweise von Ginzler und Hobi (2015) für die Schweiz verwendet, um ein landesweites Vegetationshöhenmodell (VHM) zu generieren. Dazu wurden die routinemässig vom Bundesamt für Landestopografie (swisstopo) aufgenommenen ADS Luftbildern aus mehreren Jahren verwendet.

Der von uns entwickelte Ansatz kombiniert den jüngsten Erfolg von Machine Learning (genauer Deep Learning) Methoden mit optischen Sentinel-2 Satellitenbildern, um ein landesweites Vegetationshöhenmodell mit einer 10 m × 10 m Auflösung für die Schweiz und Gabun zu berechnen (Lang et al., 2019). Während bestehende Ansätze zur Modellierung der Vegetationshöhen aus multispektralen Satelliten Bildern (e.g. Hansen et al., 2016) bereits bei Höhen von 25 m gesättigt sind, ermöglicht unsere Methode eine korrekte Schätzung bis zu einer Höhe von 55 m. Für diese Regression wird ein Convolutional Neural Network (CNN) trainiert und evaluiert, welches die 13 Spektralkanäle von Sentinel-2 prozessiert und die Vegetationshöhe schätzt. Unsere Experimente zeigen, dass die räumliche Textur eine besonders wichtige Charakteristik für die Vegetationshöhe in optischen Satellitenbildern darstellt.

Daten

Sentinel-2 ist eine optische Satellitenmission des Copernicus Programms der Europäischen Weltraumorganisation ESA, welche aus zwei Satelliten besteht, die 2015 und 2017 gestartet wurden. Seit dem erfolgreichen Start des zweiten Satelliten liefert die Mission mindestens alle fünf Tage ein neues Bild der gleichen Region. Die identischen Sensoren besitzen 13 spektrale Kanäle mit unterschiedlicher räumlicher Auflösung von 10 m, 20 m und 60 m. Die Kanäle Grün, Blau, Rot und nahes Infrarot (NIR) haben eine Auflösung von 10 m. Mit dem Spektrum im nahen und kurzwellen Infrarot ist Sentinel-2 sehr gut geeignet für Anwendungen im Wald und Vegetation. Die Daten sind öffentlich zugänglich und sind als Level 1C Produkt (top-of-atmosphere reflectance) in 100 km × 100 km Tiles mit UTM WGS84 Projektion organisiert. Für die Schweiz und Gabun stammen die Referenzdaten jeweils aus dem Jahre 2016. Für den Trainingsdatensatz in Gabun werden Bilder mit maximal +/- 9 Monaten Differenz zum Referenz Datum verwendet. Um in der Schweiz nur Bilder in der belaubten Vegetationszeit zu verwenden, wird der Zeitraum von Mai bis September beschränkt. Sentinel-2 Bilder mit mehr als 70% Wolkenbedeckungen werden nicht verwendet.

Methode

Preprocessing

Mit der Standard ESA Toolbox sen2cor werden atmosphärische Effekte korrigiert

und die Level 2A Produkte erstellt. Zudem werden die Pixelwerte pro Kanal normalisiert, sodass der Mittelwert 0 und die Standardabweichung 1 beträgt. Diese Homogenisierung der Daten sollte die Regression vereinfachen und zu einer besseren Generalisierung des Modells führen. Zuletzt werden alle Spektralkanäle auf 10 m Auflösung gesampelt.

Convolutional Neural Networks (CNNs)

CNNs sind state-of-the-art in vielen Disziplinen der automatisierten Bildinterpretation (Klassifikation, Segmentierung, Objekterkennung). Dank ihrer Eigenschaft, wertvolle Merkmale und Muster in Textur und Spektrum direkt aus den Daten zu lernen, sind sie auch sehr gut für die Analyse von Satellitenbildern geeignet. Hier nutzen wir dieses Potenzial, um Vegetationshöhen aus multispektralen Sentinel-2 Bildern zu schätzen und die dafür notwendigen Merkmale (so genannte Features) direkt aus den Referenzdaten zu lernen.

Das Grundprinzip von CNNs lässt sich folgendermassen beschreiben. Ein Input Array (z.B. ein Satellitenbild) wird mit mehreren Filtern transformiert, so dass ein Set von Features entsteht, welches für die Lösung der eigentlichen Aufgabe (hier z.B. die Regression der Vegetationshöhe) geeignet ist. Die Bildfilter werden mittels einer Faltung auf das Bild angewendet (Skalarprodukt zwischen Input und Filter an jeder Position im Bild). Dieser linearen Operation folgt eine elementweise nicht-lineare Aktivierungsfunktion. Indem eine

lange Sequenz solcher Operationen hintereinander durchgeführt wird, wird das Netzwerk tiefer und erhält somit mehr Kapazität (mehr Parameter, stärkere Nichtlinearität), um die gesuchte Funktion zu approximieren. Daher der Begriff «Deep Learning». Die Parameter dieser Filter werden allerdings nicht manuell bestimmt, sondern direkt aus den zur Verfügung stehenden Referenzdaten gelernt.

Das verwendete CNN (Abb. 1) basiert auf der Xception Architektur von Collet (2017) und benutzt «depthwise separable convolutions». Insgesamt besteht das CNN aus einem Startblock mit 1×1 Filtern (Spektrum) und 18 identischen SepConv Blöcken (Textur und Spektrum). In der separable convolution werden spektrale Merkmale und räumliche Textur-Merkmale getrennt extrahiert, womit die Anzahl der Parameter in unserem Fall um Faktor 9 reduziert wird.

Wie werden nun diese Parameter gelernt?

Zu Beginn werden alle Parameter zufällig initialisiert. Die idealen Parameterwerte werden dann mit einem iterativen Optimierungsprozess gesucht, indem eine Kostenfunktion minimiert wird. Wir verwenden hier die L2-Norm. Dabei werden in jeder Iteration zufällig 36 Patches mit 15 × 15 Pixeln aus dem Trainingsdatensatz gezogen. Für diese Patches schätzt das Netzwerk die Vegetationshöhe, welche mit der Kostenfunktion auf den Referenzdaten evaluiert wird. Mit der zufälligen Initialisierung ist diese Schätzung

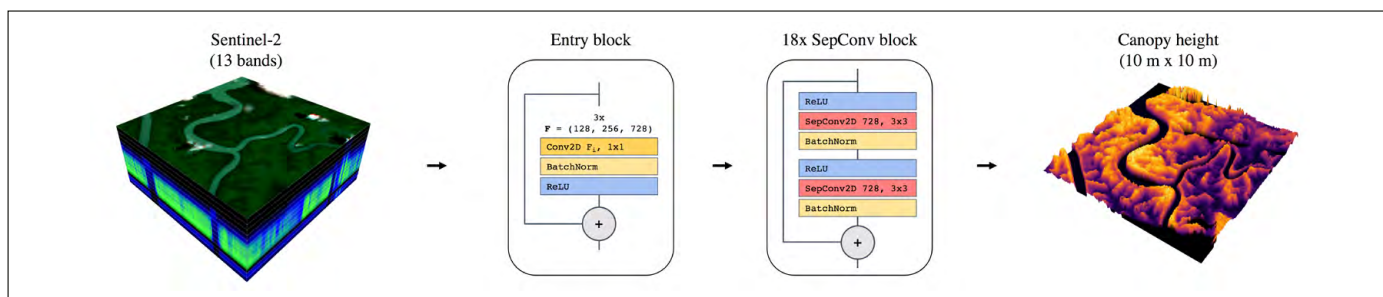


Abb. 1: Visualisierung des CNNs: Ein Sentinel-2 Bild mit 13 Spektralkanälen wird prozessiert, um die Vegetationshöhe mit einer 10 m Auflösung zu schätzen. Der Startblock erhöht die Kanaltiefe kontinuierlich auf 728 Features. Die 18 identischen SepConv Blöcke lernen nebst den spektralen Features auch räumliche Textur und Kontext Features, die mit der Vegetationshöhe korrelieren.

anfängs unbrauchbar, liefert aber das erste Signal, welches zur Optimierung der Parameter verwendet wird. Dazu wird die Kostenfunktion nach den zu schätzenden Parametern abgeleitet, wodurch wir einen Gradienten erhalten. Wir können uns diesen Prozess als «Kosten Landschaft» mit Hügeln und Tälern vorstellen, wobei das mehrdimensionale Koordinatensystem durch die zu lernenden Parameter aufgespannt wird. Folgen wir dem negativen Gradienten in kleinen Schritten, werden wir nach mehreren Iterationen ein Tal erreichen und damit ein besseres Resultat erzielen. Deshalb wird dieser Gradient mittels «Backpropagation» durch das Netzwerk propagiert und die Parameter werden so verbessert, dass die Kosten dementsprechend sinken. Die Schrittweite dieser Verbesserungen wird Lernrate genannt. Dieser Prozess ist als Stochastic Gradient Descent (SGD) bekannt, wobei wir hier eine übliche Variante von SGD namens ADAM (Kingma und Ba, 2014) verwenden. ADAM passt die Lernrate für jeden Parameter adaptiv an, was die Wahl der Lernrate vereinfacht. Insgesamt optimieren wir das Netzwerk mit 250 000 Iterationen bis es konvergiert und sich die Kosten nicht mehr ändern.

Experimente und Ergebnisse

Wir evaluieren unsere Methode auf insgesamt sieben Regionen, davon zwei in der Schweiz und fünf in Gabun. Mit diesen Testregionen wird die Genauigkeit mit verschiedenen Vegetationstypen untersucht, wobei in der Schweiz eine maximale Vegetationshöhe von 40 m angenommen wird und die Bäume in Gabun eine Höhe von über 85 m erreichen können. In der Schweiz verwenden wir die Referenz von Ginzler und Hobi (2015), wobei das photogrammetrisch abgeleitete VHM mit den Bilddaten aus dem Jahr 2016 aktualisiert wurde. Zudem wurde die Auflösung von 1 m auf die Auflösung von Sentinel-2 (10 m) reduziert. In Gabun dienen LiDAR Daten der LVIS Messkampagne aus dem Jahr 2016 als Referenz (Blair und Hofton, 2018). Somit ist es

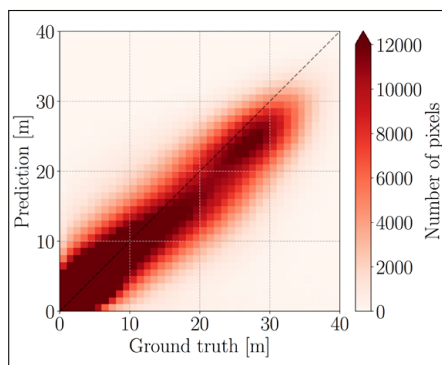


Abb. 2: Referenz vs. Schätzung in der Schweiz. Visualisiert als 2D-Histogramm mit 1-m-Intervallen.

möglich, die zeitlich korrespondierenden Sentinel-2 Bilder zu verwenden.

Jede Region wird in einen Trainings-, Validierungs- und Testbereich unterteilt. Um ein gegen radiometrische Variationen robustes Modell zu lernen, wird das CNN auf Bildern von mehreren Zeitpunkten im Jahr trainiert, mit der Annahme, dass sich die Vegetationshöhe in dem verwendeten Zeitraum nicht verändert hat. Bei der Evaluierung wird für jedes Satellitenbild im Testbereich die Vegetationshöhe an jedem Pixel geschätzt. Schlussendlich liefert der Median über den verwendeten Zeitraum ein Vegetationshöhenmodell mit einer 10 m × 10 m Auflösung.

Der mittlere absolute Fehler liegt in der Schweiz bei 1.7 m und in Gabun bei 4.3 m. Abbildung 3 zeigt den mittleren absoluten Fehler für 10 m Vegetationshöhenintervalle. In der Schweiz nimmt der Fehler mit zunehmender Höhe linear zu, wobei der relative Fehler bei etwa 20% liegt. In Gabun ist zu beobachten, dass bei einer Höhe von 30–50 m der relative Fehler signifikant besser ist. In Abbildung 2 wird ersichtlich, dass die geschätzten Werte ab einer Höhe

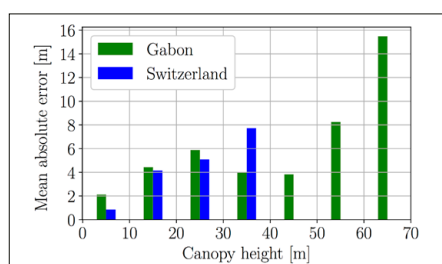


Abb. 3: Mittlerer absoluter Fehler pro 10 m Vegetationshöhenintervall.

von 15 m einen konstanten Offset von ca. -2.5 m im Vergleich zur Referenz haben. Die qualitative Evaluierung (Abb. 4) zeigt, dass die räumliche Verteilung der Vegetationshöhe korrekt reproduziert werden kann. Es ist zu sehen, dass die Schätzung im Vergleich zur Ground-Truth etwas geglättet sind. Diese Glättung lässt sich hauptsächlich durch die Aggregation von räumlichen Textur-Merkmalen über lokale Pixel Nachbarschaften erklären und ist ein bekannter Effekt bei der Verwendung von CNNs. In Abbildung 4 und 5 sind fehlende Daten in weiss dargestellt. Gründe dafür sind entweder Wolken, Wasser oder Schnee. Diese Pixel wurden nachträglich basierend auf der Sentinel-2 Level 2A Bodenklassifizierung gefiltert.

Landesweite Vegetationshöhenmodelle

Um landesweite Vegetationshöhenmodelle mit einer 10 m Auflösung für die Schweiz (Abb. 5) und Gabun zu erstellen, werden für jedes Tile (100 km × 100 km) jeweils die 10 Bilder mit der geringsten Bewölkung prozessiert. Aus den individuellen Schätzungen pro Bild wird der Median berechnet. Die Prozessierung dauerte für die Schweiz 76 Graphical Processing Unit (GPU)-Stunden (130 Tiles) und 274 GPU-Stunden für Gabun (470 Tiles), was allerdings auf einem Computer Cluster mit mehreren GPUs einfach parallelisierbar und damit realisierbar ist.

Fazit

Die von uns vorgeschlagene Methode erlaubt die Kartierung von Vegetationshöhen mit einer Auflösung von 10 m basierend auf Sentinel-2 Bildern. Bei einer angemessenen Menge von Referenzdaten ermöglicht die gute Generalisierung des Modells und die Parallelisierung der Berechnungen die Erstellung landesweiter Kartenprodukte. Diese Produkte mit einer 10 m Auflösung haben einen relativen Fehler von ≈20%. Die Vegetationshöhen sind ab einer Höhe von 55 m gesättigt, was in mitteleuropäischen Wäldern in der Regel unproblematisch ist. Der nächste Schritt wäre die Erstellung eines Vegetationshöhenmodells

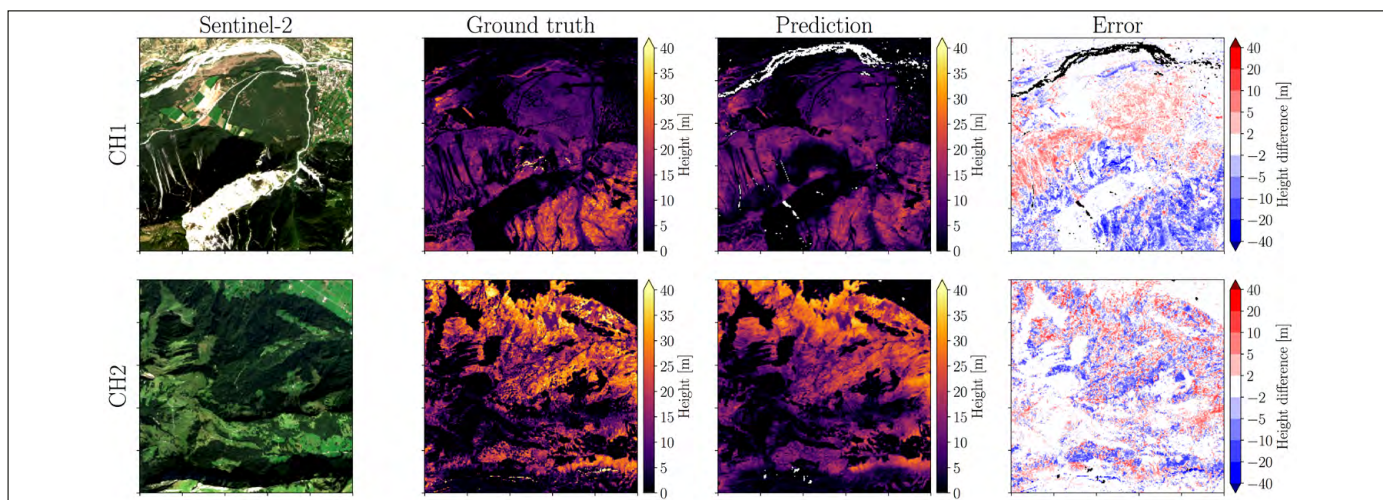


Abb. 4: Qualitative Resultate für zwei Beispiele von 5 km × 5 km Bereiche in den Testregionen.

mit einer 10 m Auflösung für ganz Europa. Wie weit das bestehende CNN innerhalb Europas anwendbar ist, muss evaluiert werden. Dazu müssen noch weitere Referenzdaten gesammelt und das CNN muss dementsprechend erneut trainiert werden.

Referenzen:

Blair, J.B., Hofton, M., 2018. AfriSAR LVIS L2 geolocated surface elevation product, version 1. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active

Archive Center. Online: <https://doi.org/10.5067/A0PMUXVUYNH>, accessed 1 March 2019.

Chollet, F., 2017. Xception: Deep learning with depthwise separable convolutions. Proc. IEEE conference on computer vision and pattern recognition, pp. 1251–1258.

Ginzler, C., Hobi, M., 2015. Countrywide stereo-image matching for updating digital surface models in the framework of the swiss national forest inventory. Remote Sensing, 7: 4343–4370.

Hansen, M.C., Potapov, P.V., Goetz, S.J., Turubanova, S., Tyukavina, A., Krylov, A., Kommareddy, A., Egorov, A., 2016. Mapping tree height distributions in sub-saharan Africa using Landsat 7 and 8 data. Remote Sensing of Environment, 185: 221–232.

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint, arXiv:1412.6980.

Lang, N., Schindler, K., Wegner, J. D., 2019. Country-wide high-resolution vegetation height mapping with Sentinel-2. arXiv preprint, arXiv:1904.13270. (accepted for publication in the journal Remote Sensing of Environment)

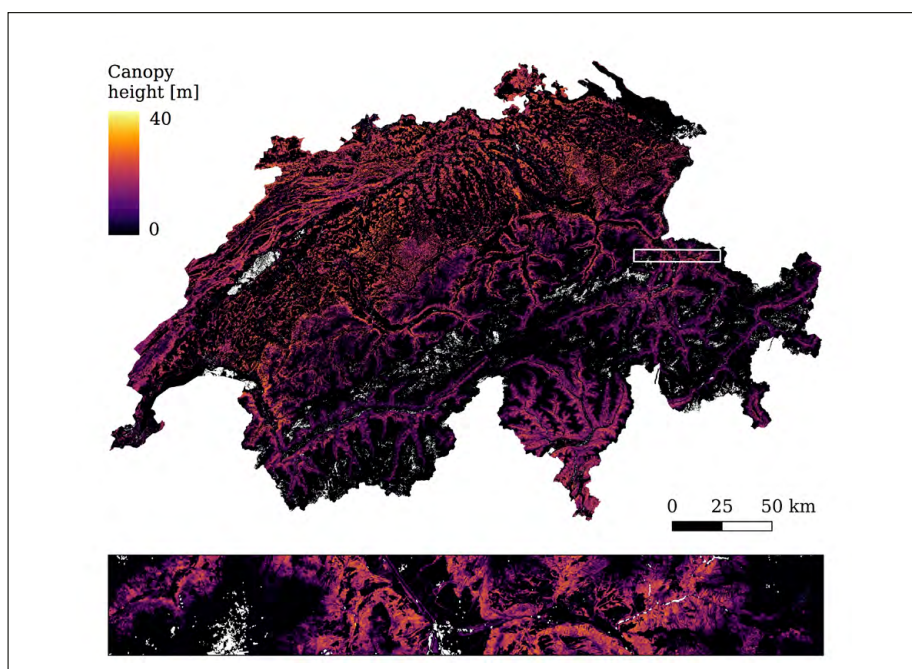


Abb. 5: Vegetationshöhenmodell der Schweiz für das Jahr 2017 mit einer Auflösung von 10 m, abgeleitet von Sentinel-2 auf Basis des Modells, welches mit Daten von 2016 erstellt wurde.

Nico Lang
Konrad Schindler
Jan Dirk Wegner
Institut für Geodäsie und Photogrammetrie, ETH Zürich
Stefano-Francini-Platz 5
CH-8083 Zürich
nico.lang@geod.baug.ethz.ch
schindler@geod.baug.ethz.ch
jan.wegner@geod.baug.ethz.ch

Christian Ginzler
Eidgenössische Forschungsanstalt für Wald, Schnee und Landschaft (WSL)
Zürcherstrasse 111
CH-8903 Birmensdorf
christian.ginzler@wsl.ch