

# **Automatisierte Analyse der Versiegelungsentwicklung in urbanen Gebieten mittels Sentinel-2 Daten und Machine Learning am Beispiel München**

Bachelorarbeit im Studium Digitalisierung – Innovation – Gesellschaft

an der Fakultät für Digitale und Analytische Wissenschaften  
der Universität Salzburg  
Fachbereich Geoinformatik

eingereicht von

**Lennart Hammer**  
12102806

Gutachter: Assoz. Prof. Dr. Dirk Tiede  
Fachbereich: Geoinformatik Z\_GIS

Salzburg, Juni 2024

# Inhaltsverzeichnis

1	Einleitung.....	1
2	Methodik.....	3
2.1	Untersuchungsgebiet.....	3
2.2	Daten.....	5
2.2.1	Stadtgrenze und Gesamtgebiet.....	5
2.2.2	Sentinel-2.....	5
2.2.3	OpenStreetMap.....	7
2.2.4	Dynamic World.....	8
2.2.5	LUCAS.....	9
2.3	Validierung.....	11
2.4	Random Forest Klassifikator.....	13
2.5	Automatisierter Prozess.....	15
2.5.1	Preprocess.....	15
2.5.2	Train & Predict.....	19
2.5.3	Postprocess.....	20
3	Ergebnisse.....	21
4	Diskussion.....	28
5	Zusammenfassung.....	30
6	Literaturverzeichnis.....	31

## Abbildungsverzeichnis

Abbildung 1: Visualisierung des Untersuchungsgebiets.....	4
Abbildung 2: Räumliche Verteilung der 165 zufällig ausgewählten Kontrollpunkte .....	11
Abbildung 3: Detailansicht der generierten Trainingsdaten aus OSM und DW.....	12
Abbildung 4: Flowchart zum gesamten Prozess.....	15
Abbildung 5: Konfusionsmatrix des Random Forest Modells aus August 2020.....	19
Abbildung 6: Genauigkeiten je Klasse aller Modelle.....	22
Abbildung 7: Konfusionsmatrix zur Validierung der Klassifikation vs. LUCAS .....	23
Abbildung 8: Entwicklung der urbanen und vegetativen Fläche.....	24
Abbildung 9: Deltas der urbanen und vegetativen Fläche innerhalb des Zeitraums .....	25
Abbildung 10: Visualisierung der Landnutzungsänderung in München von 2016 bis 2023.....	26
Abbildung 11: Close-Up zu urbaner Expansion im Raum Freiam zwischen 2016 und 2023.....	27

## Tabellenverzeichnis

Tabelle 1: Zuordnung der Klassenwerte und zugehörige Sentinel-2 Reflexionen .....	18
---	----

# 1 Einleitung

Seit 2008 lebt die Mehrheit der Weltbevölkerung in Städten (Birch & Wachter, 2011). Fakten wie diese unterstreichen den Trend der Urbanisierung, definiert als Ausbreitung von urbanen Lebens- und Wirtschaftsformen, welcher das Gesellschaftsbild innerhalb der letzten Jahrzehnte stark geprägt hat (Kopfmüller, 2016). Obwohl die durchschnittliche Wachstumsrate von Städten weltweit seit 1960 abnimmt, ist ein Anstieg der absoluten Bevölkerungszahl nicht von der Hand zu weisen (McGranahan & Satterthwaite, 2014). Die Vereinten Nationen prognostizieren für das Jahr 2050 einen Zuwachs von 941 Millionen Stadtbewohnern in Afrika und eine Zunahme von 1213 Millionen Individuen in Asien. Die weltweite Urbanisierungsrate soll von derzeitigen 55.3% auf 68.4%, in weiterentwickelten Regionen sogar auf 86.6% steigen (United Nations Population Division, 2019). In Europa führte Belgien im Jahr 2022 mit einem Urbanisierungsgrad von 98.15% die Statistiken an, während Deutschland mit einem Wert von 77.65% knapp unter dem europäischen Durchschnitt von 77.84% lag (World Bank, 2022).

Eine unausweichliche Konsequenz der Urbanisierung ist die vermehrte Flächeninanspruchnahme, um den Bedarf an Wohnraum und die höheren Anforderungen an die regionale Infrastruktur gewährleisten zu können (Naumann et al., 2019). Dies resultiert in einer Zunahme der Bodenversiegelung, welche die intensivste Form der Flächeninanspruchnahme ist und die Zerstörung oder Bedeckung des Bodens durch vollständig oder teilweise undurchlässige Materialien, wie Asphalt oder Beton, beschreibt (Prokop et al., 2011). Laut Umweltbundesamt hat die Bodenversiegelung in Deutschland innerhalb der letzten 30 Jahre um 4.943 km<sup>2</sup> zugenommen, wobei neue Verkehrsflächen einen signifikanten Teil zu dieser Zunahme beitragen. Nachdem die jährliche Versiegelung seit den 1990er Jahren stetig abgenommen hat, ist auch durch die Erschließung neuer Wohngebiete in den letzten Jahren wieder ein leichter Anstieg zu vernehmen (Umweltbundesamt, 2024).

Die Versiegelung von Böden geht einher mit einigen Folgen. Zum einen wird der natürliche Wasserkreislauf gestört, indem das Eindringen von Wasser in den Boden vermindert wird. Dies wirkt sich negativ auf die Bildung von Grundwasser aus und kann zudem zu erhöhter Überschwemmungsgefahr führen (Lozán & Breckle, 2021). Zum anderen ist ein hoher Grad an Bodenversiegelung ein ausschlaggebender

Faktor für das Phänomen der städtischen Wärmeinseln. Ohne die kühlende Verdunstung durch die Vegetation und fehlende Winde auf Bodenhöhe durch starke Bebauung, kann es zu regionalen Klimaveränderungen, Luftverschmutzung und Verschlechterung der Sichtverhältnisse kommen (Fokaides et al., 2016). Des Weiteren verhindert der Verschluss der Erdoberfläche einen effektiven Gasaustausch zwischen der Atmosphäre und dem Boden, resultierend in einem negativen Einfluss auf die Bodenqualität und die biologischen Prozesse in der Erde. Zusammen mit der Veränderung von Lebensräumen, zieht dies einen Verlust von Biodiversität und somit ökologische Folgen für Flora und Fauna nach sich (Scalenghe & Marsan, 2009). Jegliche Versiegelung sollte daher mit Bedacht durchgeführt werden – nicht zuletzt, da Boden in der Nationalökonomie neben Arbeit und Kapital als essenzieller Wirtschaftsfaktor gesehen wird und eine Entsiegelung teuer und häufig unmöglich ist (Röpke & Lippelt, 2014).

Um die Ressource Boden zu schonen und die negativen Auswirkungen der Bodenversiegelung zu begrenzen, sind Analysen zur Veränderung der Landnutzung sowie Informationen über den Grad der Versiegelung von entscheidender Bedeutung. Bisher wird der Grad der Bodenversiegelung in Deutschland hauptsächlich durch Schätzverfahren mithilfe von Luftbildern und Rechenmodellen bestimmt, welche mit topografischen Karten, Katasterdaten und Bebauungsplänen abgeglichen werden. Diese Methoden ermöglichen eine grobe Abschätzung der Versiegelung. Dennoch sind die Ergebnisse mit Unsicherheiten behaftet und können lediglich einen räumlichen Vergleich zwischen verschiedenen Gebieten, jedoch keinen zeitlichen Vergleich zwischen mehreren Zeitpunkten ermöglichen. Eine regelmäßige amtliche Erfassung der Bodenversiegelung findet bisher nicht statt (Umweltbundesamt 2024).

Da Daten von anderen Quellen, wie Städten und Ländern, nur selten veröffentlicht und nahezu alle Produkte mit moderater Auflösung unregelmäßig erstellt werden oder unter Verschluss sind (Brown et al., 2022), stellt sich die Frage, auf welche Weise räumlich und zeitlich hochaufgelöste Informationen über die Landbedeckung generiert werden können. Die Nutzung von Satellitendaten ist unter anderem wegen der Abdeckung großer Gebiete, dem freien Zugang zu großen Datenmengen über Produkte wie Sentinel-2 und der Möglichkeit für multitemporale Analysen, ein beliebter Ansatz, um Rückschlüsse auf Vorgänge und Veränderungen auf der Erdoberfläche zu ziehen (Peroni et al., 2022). Dabei ist das Generieren von

Informationen über die Landbedeckung und Nutzung anhand von Bildklassifikationen eine der am häufigsten angewandten Methoden (Foody, 2002). Beispielsweise konnte basierend auf der Berechnung von Vegetations- und Bebauungsindizes ein neunprozentiger Anstieg der bebauten Fläche Madrids innerhalb der letzten 30 Jahre festgestellt und somit die urbane Expansion analysiert werden (Garcia & Perez, 2016). In Bezug auf die Bodenversiegelung konnten Fernerkundungsdaten mithilfe von räumlicher Regression genutzt werden, um den Einfluss der Urbanisierung auf die Landnutzung und Agrarfläche zu bewerten (Xiao et al., 2013) oder den Grad der Versiegelung zu messen (Kampouraki et al., 2006).

Aufbauend auf den Erkenntnissen von vorherigen Studien, untersucht diese Arbeit, inwiefern qualitativ hochwertige Flächenstatistiken zur Landbedeckung und Bodenversiegelung anhand von Sentinel-2 Bildern und frei zugänglichen Trainingsdaten wie OpenStreetMap (OSM) und Dynamic World in einem automatisierten Prozess erfasst werden können. Es wird ein Ansatz vorgestellt, welcher das Stadtgebiet Münchens mit nahezu monatlicher Auflösung in fünf Kategorien klassifiziert und die Veränderung der urbanen und vegetativen Fläche über einen Zeitraum von 8 Jahren misst. Mithilfe eines Random Forest Machine Learning Verfahrens ist der Ablauf über die Programmiersprache Python automatisiert und leicht skalierbar auf andere Städte und Regionen anwendbar. Es soll gezeigt werden, ob durch den Ansatz Daten zur Expansion von Städten ermittelt und Rückschlüsse auf die Ausmaße der Versiegelung gezogen werden können, um informationsgetriebene Entscheidungen zugunsten des Menschen und seiner Umwelt treffen zu können.

## **2 Methodik**

### **2.1 Untersuchungsgebiet**

Das gesamte Untersuchungsgebiet umfasst die bayerische Landeshauptstadt München im Süden von Deutschland. Das Areal erstreckt sich von den Koordinaten 48°14'58N, 11°21'47E und 48°03'31N, 11°43'17E mit einer Fläche von 559 km<sup>2</sup> und tangiert an jeder Kante die administrative Stadtgrenze Münchens, welche den Beobachtungsbereich nochmals unterteilt. Die Landklassifikation wird für das Gesamtgebiet erstellt, während die Flächenstatistiken nur innerhalb der Stadtgrenzen erfasst werden.

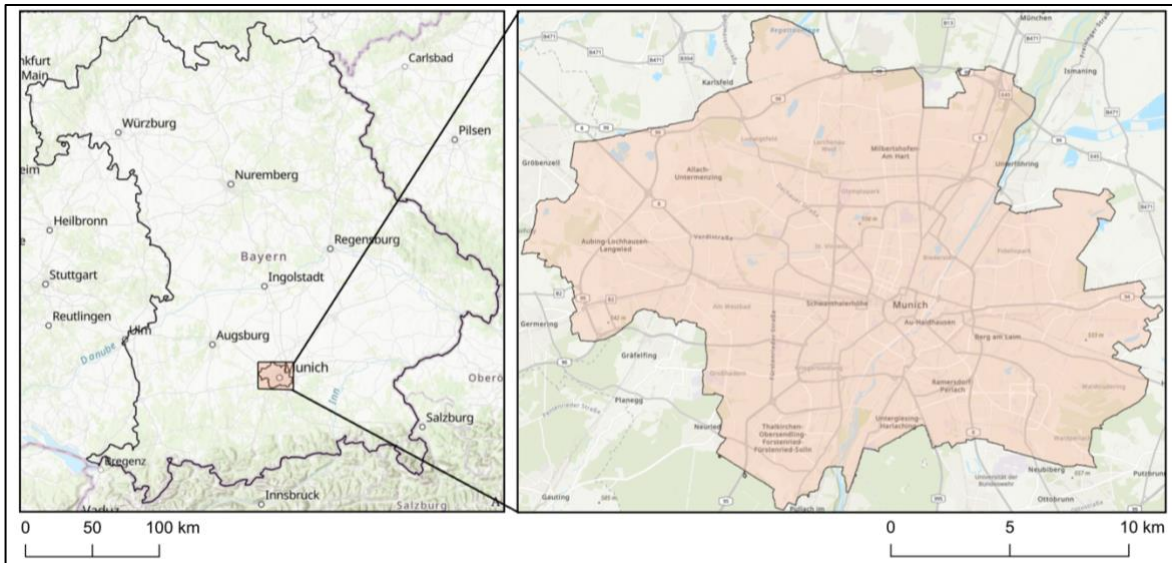


Abbildung 1: Visualisierung des Untersuchungsgebiets.  
 Links: Bereich innerhalb von Bayern; Rechts: Untersuchungsgebiet inkl. administrativer Stadtgrenze

Die Stadt München ist seit Jahren wirtschaftlich gut aufgestellt und wies im Jahr 2020 hinter Berlin das zweithöchste Bruttoinlandsprodukt deutscher Städte auf (Statistische Ämter des Bundes und der Länder, zitiert nach de.statista.com, 2022). Mit seiner Mischung aus globalen Unternehmen, einem wohlhabenden Mittelstand und einer Vielfalt an Branchen attraktiv für junge Erwachsene. Vor allem durch die Zuwanderung aufgrund der aussichtsreichen Ausbildungs- und Arbeitsplatzmöglichkeiten, verzeichnet München seit den 2000er Jahren einen signifikanten Bevölkerungszuwachs. Dieser Anstieg kombiniert mit Flächenknappheit und dem höchsten Mietpreinsniveau unter Deutschlands Großstädten, resultiert in einem angespannten Wohnungsmarkt. Um dieser Herausforderung gerecht zu werden, wurden Programme wie „Wohnen in München“ ins Leben gerufen, um bezahlbaren Wohnraum in der Stadt nachhaltig zu sichern (Landeshauptstadt München, 2022). So wurde beispielsweise in Freiham ein neuer Stadtteil in Größe einer Kleinstadt konzipiert. Auf dem 350 Hektar großen Areal wird die Infrastruktur geschaffen, um Wohn- und Lebensraum für mehr als 25.000 Menschen bereitzustellen (Landeshauptstadt München, 2023). Durch die hohe Dynamik der Stadtentwicklung, ist München ideal geeignet, um die Veränderung der Landnutzung zu analysieren.

## 2.2 Daten

### 2.2.1 Stadtgrenze und Gesamtgebiet

Während eine Stadt als eng besiedelter Bereich mit hoher Bebauungsdichte, Bevölkerungsanzahl und Sozialstruktur klar definiert werden kann (Schubert & Klein, 2020), gibt es für die Beschreibung ihrer Grenzen verschiedene Ansätze. Gerade dynamisch expandierende Städte können über ihre Grenzen hinauswachsen und für eine Abweichung zwischen der physikalischen und administrativen Ausdehnung sorgen (Taubenböck et al., 2019), weshalb es für geografische Analysen sinnvoll sein kann, die Grenzen anhand der Bevölkerungsdichte oder der Landbedeckung zu ziehen. Beispielsweise definiert die Europäische Union mittels Satellitendaten des Copernicus Programms urbane Zentren nach Einwohneranzahl im „*Global Human Settlement Layer*“ (Pesaresi et al., 2016) – ähnlich könnte man die Stadtgrenzen aus Daten der undurchlässigen und bebauten Landbedeckung des „*Copernicus Land Monitoring Services*“ extrahieren. Da jedoch sowohl Informationen zum Versiegelungsgrad als auch Kataster- und Bauungsdaten auf aggregierter Gemeindeebene vorliegen, wurde die administrative Stadtgrenze Münchens verwendet, um eine Vergleichbarkeit der Ergebnisse zu gewährleisten.

Die administrative Stadtgrenze der bayerischen Landeshauptstadt wurde vom „*GeoPortal München*“ akquiriert. Die Flächen der prognostizierten Landnutzungsklassen werden innerhalb dieser Grenzen berechnet. Außerdem definiert die Stadtgrenze gleichzeitig auch die Ausmaße des rechteckigen Gesamtgebiets. Alle verwendeten Daten werden auf diesen Bereich zugeschnitten, um sie daraufhin für das Trainieren und Vorhersagen des Machine Learning Modells zu verwenden.

### 2.2.2 Sentinel-2

Als Teil des Copernicus Programms bringt die Sentinel-2 Mission, unter Leitung der ESA, seit 2015 frei verfügbare optische Satellitendaten der Erdoberfläche hervor, mit dem Ziel, Informationen über die Umwelt bereitzustellen. Der Satellit bietet mit seiner vergleichsweise hohen räumlichen Auflösung von 10 bis 60 Metern je nach Band, seiner breiten spektralen Abdeckung von sichtbarem Licht bis hin zu nahem und kurzwelligem Infrarot sowie einer Wiederbesuchszeit von 5 Tagen sehr gute Ausgangsdaten für georäumliche Analysen (Gascon et al., 2014).

Die Daten werden weltweit erfolgreich in vielen Bereichen, insbesondere aufgrund



ihrer hohen räumlichen Auflösung und vielversprechenden Ergebnisse, eingesetzt. Die Kartierung von Wäldern, die Überwachung von Wasserressourcen, Naturkatastrophen und landwirtschaftlichen Flächen sind nur ein Teil der möglichen Anwendungen. Durch die hohe zeitliche Auflösung wird das Monitoring von Naturereignissen wie Fluten oder Vulkanausbrüchen ermöglicht. Neben diesen Applikationen bietet Sentinel-2 ebenfalls ein verlässliches Produkt für die Überwachung urbaner Regionen. So werden beispielsweise Informationen über Phänomene wie städtische Wärmeinseln und urbane Grünflächen generiert, sowie die Expansion von Städten überwacht (Phiri et al., 2020).

Trotz der vielversprechenden Eigenschaften muss auf die Limitationen der Sentinel-2 Daten in Bezug auf Landklassifikationen hingewiesen werden. Das Fehlen des panchromatischen und des thermischen Bandes erschwert jeweils die Prozesse des Pan-Sharpings und der Wolkenerkennung. Zudem erfordert die unterschiedliche räumliche Auflösung der einzelnen Bänder eine Skalierung oder Neuabtastung, um homogene Pixelgrößen von 10x10 Metern zu erreichen (Phiri et al., 2020). Obwohl Sentinel-2, wie auch Landsat, zu den hochauflösenden Satelliten gehört, ist die räumliche Auflösung von 10 Metern im Vergleich zu Produkten wie QuickBird oder RapidEye deutlich geringer. Die gröbere Auflösung macht es nahezu unmöglich, Informationen über kleine städtische Details zu erfassen (Gombe et al., 2017). Bei minimalen Veränderungen in Grünflächen oder der Klassifikation von Wegen und Pfaden stoßen Sentinel-2 Daten an ihre Grenzen (Haas & Ban, 2018). Die Unterscheidung zwischen nur leicht und vollkommen versiegelten Flächen kann nicht gewährleistet werden.

Um Satellitendaten effektiv nutzen zu können, müssen diese zuerst vorprozessiert werden. Da die nötige Aufbereitung für viele Anwendungen gleich ist, gibt es Produkte, welche „*Analysis Ready Data*“ zur Verfügung stellen. Das „*Centre National d’Etudes Spatiales*“ und „*Deutsches Zentrum für Luft- und Raumfahrt*“ stellen durch den „*MAJA*“ und „*WASP*“ Prozessor einen solchen Datensatz bereit. Im Wesentlichen führt der „*MAJA*“ Prozessor mit den Ausgangsdaten einige Korrekturen durch, um deren Qualität zu verbessern. Zu diesen gehören die Berichtigung der molekularen Absorption und Streuung, die Korrektur von Zirruswolken und die Verbesserung von atmosphärischen Effekten. Außerdem werden Wolken und deren Schatten erkannt und Fehler bedingt durch das Gelände und dessen Beleuchtung eliminiert. Das Verfahren fusioniert zudem multi-temporale

Daten, um Änderungen der Reflektionen auf der Erde zu erkennen und so die Ergebnisse zu optimieren. Die daraus resultierenden Level-2A Daten werden mithilfe des „WASP“ Prozessors zu einem gewichteten Durchschnitt über 45 Tage für jeden Monat kombiniert und so in wolkenfreie Level-3A Daten umgewandelt (Hagolle et al., 2021).

Die durch „WASP“ vorverarbeiteten Daten stehen nach Sentinel-2 Granule, die jeweils eine spezifische geografische Region abbilden, und Datum sortiert zur Verfügung und können mit monatlicher Auflösung über den Geoservice des DLR heruntergeladen werden. In diesem Verfahren werden die Level-3A Daten jeweils von März bis Oktober verwendet. Da die Qualität mancher Bilder durch Wolken oder Pixelfehler verfälscht ist, kommen manche Monate nicht für die Nutzung in Betracht. Insgesamt wurden 59 Szenen im Zeitraum von 2016 bis 2024 der UTM-Zone 32 UPU verwendet. Es wurde bewusst auf die Wintermonate verzichtet, da Schnee und häufiges schlechtes Wetter die Qualität der Daten und Ergebnisse negativ beeinträchtigen würden. Da im Winter tendenziell auch weniger Bauaktivität herrscht, ist diese Entscheidung für das Projekt akzeptabel.

### **2.2.3 OpenStreetMap**

Als Folge der technischen Weiterentwicklung in Kombination mit der breiten Verfügbarkeit von GPS-Geräten und dem Internet, hat sich Crowdsourcing im Bereich der Kartographie im Laufe der letzten Jahre etabliert (Haklay & Weber, 2008). OpenStreetMap (OSM) gehört mit 10.5 Millionen Nutzern (OpenStreetMap Wiki, 2024) zu dem meist genutzten und vielversprechendsten Projekten im Bereich „Volunteered Geographic Information“. Ziel ist es, jedem Menschen freien Zugang zu geografischen Informationen zu ermöglichen. Diese Informationen liegen strukturiert als Knoten (Punkte mit geografischer Position), Wege (Linien oder Polygone aus einer Liste von Knoten), Beziehungen (Relation zwischen vorhandenen Knoten und Wegen) und Tags (Key-Value Paare zur Speicherung von Metadaten) vor.

Aufgrund der großen Anzahl an Beitragenden sind jedoch Herausforderungen hinsichtlich der Vollständigkeit, Konsistenz und Genauigkeit vorhersehbar (Haklay & Weber, 2008). Gerade die räumliche Abdeckung ist je nach Region variabel – städtische Gebiete sind meist besser kartiert als ländliche, wodurch es dort teilweise zu fehlenden Straßen oder ungenauen Grundrissen kommen kann (Vargas-Munoz et al., 2020). Es muss somit evaluiert werden, inwiefern auch Unternehmen und die

Forschung von diesen Daten profitieren können.

Trotz der aufgeführten Unsicherheiten in Bezug auf Qualität basieren etliche Studien auf den von der Masse erhobenen Daten. Die Aussicht auf große, frei zugängliche Datenmengen für das Trainieren von KI-Modellen macht das Projekt vor allem für den Sektor des maschinellen Lernens interessant (Wu et al., 2020). Auch die Klassifizierung der Landbedeckung kann von OSM profitieren. So wurden die Informationen zusammen mit Satellitendaten beispielsweise erfolgreich genutzt, um eine Landklassifikation mit hohen Genauigkeiten der Stadt Wuhan zu erstellen (Luo et al., 2019) oder die urbane Expansion in Kathmandu zu dokumentieren (Khanal et al., 2019). Die Verfügbarkeit von historischen Daten eröffnet außerdem Möglichkeiten für die Analyse und Erkennung von regionalen Veränderungen der Erdoberfläche innerhalb eines Zeitraums und dient als valide Datenquelle für die Erstellung von multi-temporalen Fernerkundungsklassifizierungen (Viana et al., 2019).

Da es sich bei München um eine Großstadt mit einer aktiven Community handelt (OpenStreetMap Wiki, 2024), kann von einer vergleichbar hohen Qualität der Daten ausgegangen werden. Die Einträge von Gewässern, Straßen und Gebäuden sind in urbanen Gebieten meist sehr genau (Vargas-Munoz et al., 2020) und werden gewissermaßen in Echtzeit von den Nutzern kontrolliert, wodurch die Informationen geeignet für diese Studie sind. Sowohl die historischen als auch die aktuellen europaweiten OSM-Daten wurden über den Downloadserver von „Geofabrik“ heruntergeladen. Die Datensätze für Oberbayern wurden aus den Jahren 2016 bis 2024 akquiriert. Von den insgesamt 18 Shapefiles pro Jahr wurden Informationen zu Gebäuden, Parkplätzen, Straßen, Bahngleisen und Gewässern extrahiert. Andere Daten zu der Landnutzung, Parks oder Wiesen wurden ignoriert, da sich bebauten und unversiegelte Flächen innerhalb der Polygone nach eigenen Beobachtungen oft überschneiden und somit meist nicht konkret genug sind, um als Trainings- und Validierungsdaten zu fungieren.

#### **2.2.4 Dynamic World**

Um die Lücke einer zeitlich hochaufgelösten Land Use/Land Cover (LULC) Karte zu schließen, die im Gegensatz zu den meisten anderen Produkten nahezu ohne Verzögerung aktualisiert wird, wurde Dynamic World entwickelt. Die homogene, hochauflösende Klassifikation basiert auf Sentinel-2-Bildern, deren Pixel mithilfe eines Deep Learning-Verfahrens in insgesamt neun Landnutzungsklassen eingeteilt

werden. Das Fully Convolutional Neural Network (FCNN) wurde mit 24.000 manuell gelabelten Kacheln aus allen Biomen trainiert und validiert. Die Verarbeitung der großen Datenmengen erfolgt über die Google Earth Engine und die Cloud AI Platform, wodurch der Datensatz kontinuierlich aktualisiert wird, sobald neue Sentinel-2-Daten verfügbar sind. Die Klassifikation zeigte ähnliche oder bessere Ergebnisse im Vergleich zu anderen regionalen oder globalen Produkten und brachte laut Brown et al. in gemäßigten und baumreichen Biomen die besten Resultate hervor. Die Kombination dieser Eigenschaften macht das Produkt Dynamic World für viele Anwendungen und Nutzer interessant (Brown et al., 2022). Trotz der vielversprechenden Eigenschaften muss auf die Limitationen des Datensatzes eingegangen werden. Während bebaute Flächen mit 83%, Bäume mit 81% und Felder mit 78% gute Genauigkeiten aufzeigen, wurden die Klassen offener Boden nur zu 57% und Gras nur zu 34% richtig klassifiziert. Das Zusammenfassen von Klassen kann hilfreich sein, um Gesamtgenauigkeiten zu verbessern. Nichtsdestotrotz sollte eine Validierung mittels zufälliger Stichproben durchgeführt werden, um Genauigkeiten und Bias zu prüfen (Venter et al., 2022).

Innerhalb des Verfahrens werden die Klassen Bäume, Gras, offener Boden und Felder der Jahre 2016 bis 2024 genutzt. Um die Nutzung von Dynamic World als Trainingsdatensatz zu legitimieren, wurden die Klassen Bäume und Gras zu der Kategorie Vegetation kombiniert. Diese ist in Dynamic World eher unterklassifiziert, womit davon ausgegangen werden kann, dass Pixel dieser Klasse auch in Wirklichkeit Vegetation entsprechen. Außerdem wird nach dem Klassifizierungsprozess eine zusätzliche Validierung mit echten Ground Truth Daten durchgeführt, um die Genauigkeiten zu prüfen. Die Daten wurden über die Google Earth Engine für die jeweiligen Jahre und das Untersuchungsgebiet akquiriert.

### **2.2.5 LUCAS**

Die Flächenstichprobenerhebung zur Landnutzung und Bodenbedeckung (LUCAS) wird seit 2016 von Eurostat durchgeführt, um flächendeckende Statistiken zur Bodenbedeckung zu erfassen. Alle drei bis vier Jahre werden in Zusammenarbeit mit den EU-Ländern europaweite Bodenstichproben abgenommen, um Informationen für die Land- und Forstwirtschaft, den Natur- und Bodenschutz und die Politik zu sammeln und biologische Prozesse zu überwachen. Im Rahmen des Projekts wurden im Jahr 2018 über 330.000 Punkte von jeglichen Bodenarten

erfasst – die jüngsten vorläufigen Daten liegen für das Jahr 2022 vor, welche durch die einheitliche und konstante Methodik der Erfassung vollkommen homogen sind. Innerhalb der Stichproben werden einige Parameter des Bodens, wie dem Gehalt an groben Fragmenten, der Korngrößenverteilung, dem pH-Wert oder dem Gehalt an organischem Kohlenstoff analysiert. Anhand dieser Eigenschaften werden die Böden in acht Kategorien mit 29 Klassen eingeteilt, welche die Bodenbedeckung wie bebautes Land, Vegetation, Ackerflächen oder Wasser repräsentieren (Eurostat, 2024).

Die standardisierten und frei zugänglichen Daten bieten vielfältige Anwendungsmöglichkeiten sowohl in der Wissenschaft als auch in der politischen Entscheidungsfindung. Neben der Bewertung des Bodens in Hinblick auf wirtschaftliche Aktivitäten wie der Landwirtschaft oder dem Vergleich von regionaler, nationaler und internationaler Bodenqualität, können die Daten ebenfalls, wie in dieser Arbeit, zur Validierung von Karten genutzt werden (Orgiazzi et al., 2018).

Die Daten können über die Datenbank von Eurostat als CSV-Datei bezogen werden, welche anschließend geocodiert werden muss. Innerhalb des beschriebenen Untersuchungsgebiets um München sind für das Jahr 2018 insgesamt 37 und für das Jahr 2022 48 Untersuchungspunkte im LUCAS Datensatz vorhanden. Diese wurden durch manuell gelabelte, zufällig ausgewählte Punkte ergänzt. Für das Jahr 2018 wurden 128 und für 2022 wurden 117 zusätzliche Positionen klassifiziert, sodass für beide Zeitpunkte ein Validierungsdatensatz von jeweils 165 Punkten verfügbar ist. Im Jahr 2018 stehen 24 Kontrollpunkte für die Klasse Felder, 75 für Vegetation, 52 für bebaut, 6 für offener Boden und 8 für Wasser zur Verfügung. Für das Jahr 2022 sind 23 Punkte für die Kategorie Felder, 71 für Vegetation, 52 für bebaut, 9 für offener Boden und 10 für Wasser klassifiziert worden. Diese Daten dienen der Validierung der finalen, durch den Random Forest erzeugten Landklassifikation.

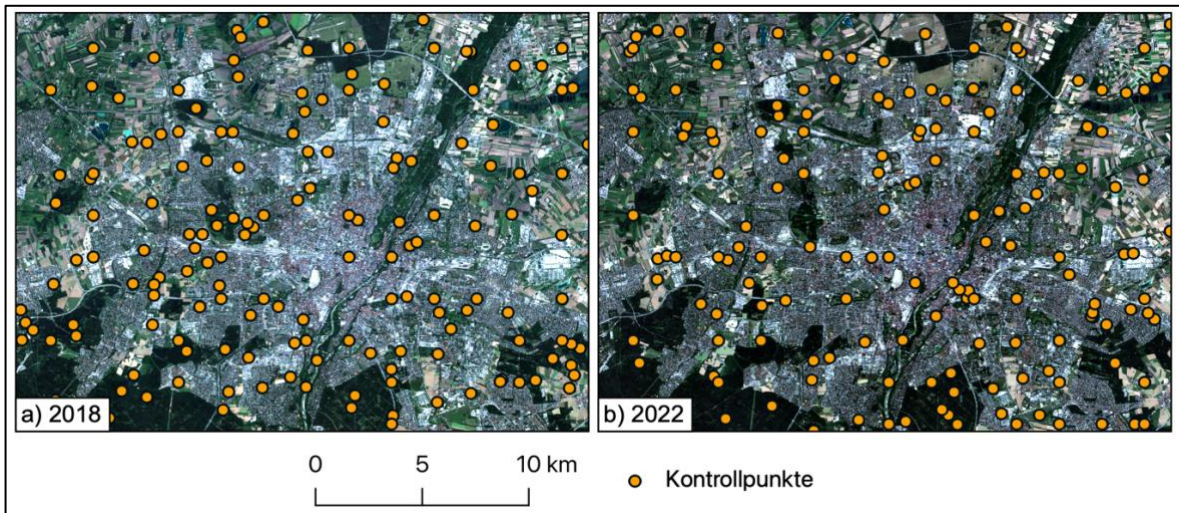


Abbildung 2: Räumliche Verteilung der 165 zufällig ausgewählten Kontrollpunkte für 2018 (links) und 2022 (rechts)

### 2.3 Validierung

Bei einer Landklassifikation über 59 Zeitpunkte verteilt über acht Jahre und einem dynamischen Prozess, wie der Expansion einer Stadt, stellt sich die Frage, wie die Leistung des Random Forests sinnvoll überprüft werden kann. Für die Validierung werden Ground Truth Daten der Erdoberfläche benötigt, um die Vorhersage mit der wahren Klasse des Pixels zu vergleichen und so Rückschlüsse über die Genauigkeit ziehen zu können. Wie schon erwähnt, sind öffentliche Produkte zeitlich oder räumlich meist zu schlecht aufgelöst, um für eine regelmäßige Validierung in Betracht zu kommen. Die herkömmliche Lösung für diese Problematik, welche in vielen Landklassifikationen Einsatz findet, ist das händische Labeln von Pixeln mithilfe von hochauflösten Fernerkundungsbildern oder durch In-Situ Bodendaten (Fonte & Martinho, 2017). Die manuelle Klassifizierung von Pixeln ist, gerade bei 59 verschiedenen Sentinel-2 Szenen, mit hohem zeitlichem Aufwand verbunden, weshalb diese Studie einen automatisierten und skalierbaren Ansatz zur Erstellung von simulierten Ground Truth Daten verfolgt. Durch die Kombination aus Stadt- und Gewässerdaten von OSM und Informationen über Vegetation und offenen Boden aus Dynamic World können jährliche, hochauflöste Referenzdaten dieser vier Klassen produziert werden. Es muss hervorgehoben werden, dass es sich bei den produzierten Trainings- und Validierungsdaten nicht um echte Ground Truth Daten handelt und keine hundertprozentige Garantie über die Genauigkeit gegeben werden kann. Studien zeigen jedoch, dass OSM gerade für Gebäude, Straßen und Bahnstrecken großes Potential hat, bei der Validierung von LULC-Produkten eingesetzt zu werden und als Datenquelle nicht unterschätzt

werden sollte (Fonte & Martinho, 2017). Auch Dynamic World zeigt für die Klassen Bäume und Ackerflächen hohe Genauigkeiten und kann unter Berücksichtigung der zuvor aufgezeigten Schwächen als Validierung von Landklassifikationen geeignet sein (Venter et al., 2022). Deshalb wird die Nutzung des simulierten Ground Truths, als Kombination aus OSM und Dynamic World, in diesem Verfahren als sinnvoll und gerechtfertigt erachtet.

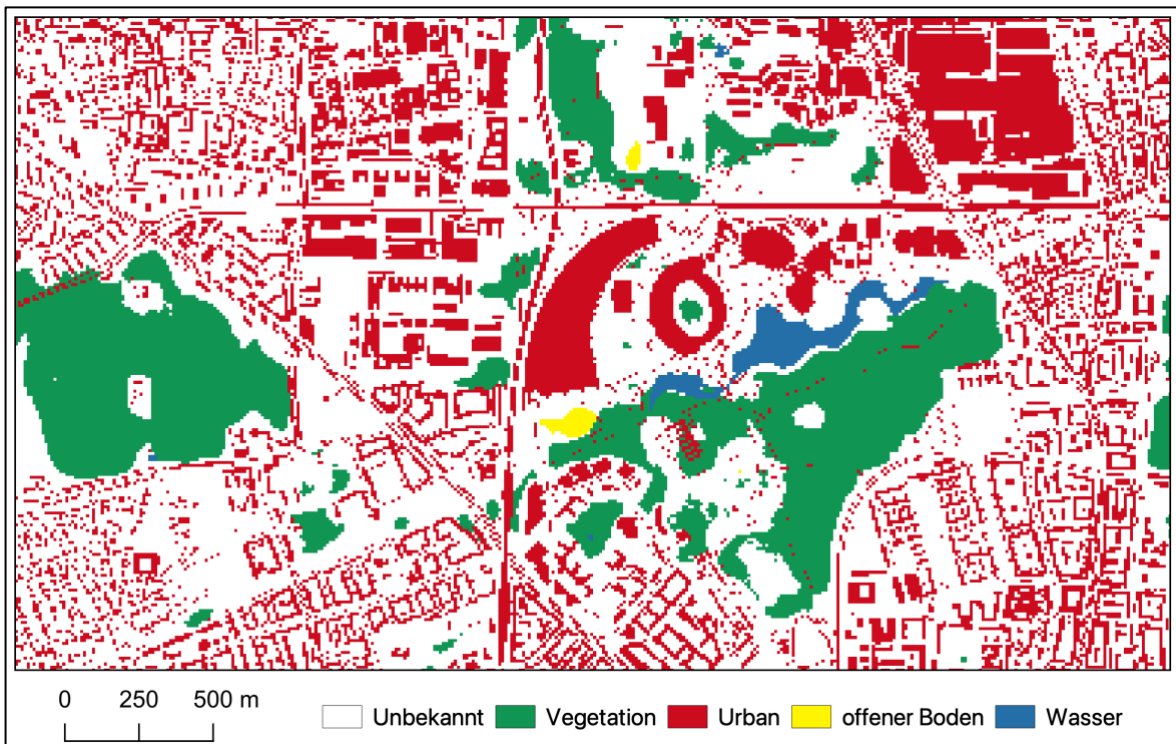


Abbildung 3: Detailansicht der generierten Trainingsdaten aus OSM und DW in der Umgebung des Olympiaparks München

Um den Random Forest zu validieren, werden dem Modell jeweils 20 % der Trainingsdaten vorbehalten, mit denen dann die Genauigkeit der Vorhersage beurteilt wird. Versuche mit einem Trainings- und Testsplit von 70 – 30 und 60 – 40 zeigten keinen Effekt auf die Leistung der Modelle.

Um die finalen, nachbearbeiteten Klassifikationen einer zusätzlichen Prüfung zu unterziehen, wird unter anderem ein Abgleich mit den zuvor beschriebenen LUCAS Daten und den manuell hinzugefügten Kontrollpunkten für die Jahre 2018 und 2022 durchgeführt. So kann die Genauigkeit der Vorhersagen zu den zwei Zeitpunkten mit echten Ground Truth Daten bestimmt werden. Sowohl für jedes Random Forest Modell als auch für die Ground Truth Vergleiche wird eine Gesamtgenauigkeit und eine Konfusionsmatrix, aus welcher man die Genauigkeiten der einzelnen Klassen und die Verteilung der Pixel innerhalb dieser ablesen kann, berechnet.

## 2.4 Random Forest Klassifikator

Der Random-Forest-Algorithmus, eingeführt von Breiman (2001), nutzt binäre Entscheidungsbäume, um Beobachtungen anhand zufällig ausgewählter Variablen zu klassifizieren (Luti et al., 2020).

Die Nutzung des Machine Learning Verfahrens bietet viele Vorteile im Gebiet der Fernerkundung. Diese reichen von einer effizienten Verarbeitung von großen Datensätzen und tausenden Eingangsvariablen, über eine interne Schätzung des Generalisierungsfehlers bis hin zu einer Bewertung der Bedeutung von Variablen, was eine Gewichtung dieser ermöglicht (Rodriguez-Galiano, 2012). Diese Eigenschaften, kombiniert mit den hohen Genauigkeiten in Verbindung mit Sentinel-2 Daten, macht den Random Forest zu dem weitverbreitetsten Machine Learning Modell für Pixel basierte Klassifikationen (Phiri et al., 2020).

Die Wahl zugunsten des Random Forest Algorithmus fiel auch aufgrund von vielversprechenden Ergebnissen in Testläufen, kombiniert mit einer akzeptablen Laufzeit. Als Alternativen wurden der KNeighbours- und der Gaussian Naive Bayes Klassifikator betrachtet. Ersterer überzeugte mit ähnlichen Genauigkeiten wie der Random Forest, schied jedoch aufgrund der sehr langen Dauer aus. Der GaussianNB Algorithmus war den zwei Konkurrenten zwar in Sachen Laufzeit stark überlegen, wurde jedoch durch eine rund 20% schlechtere Genauigkeit als ungeeignet für das Projekt bewertet.

In den ersten Ansätzen wurde versucht ein einziges Random Forest Modell für alle 59 Sentinel-2 Szenen innerhalb des gesamten Zeitraums zu trainieren. Da sich jedoch die Homogenität des vorverarbeiteten „WASP“ Datensatzes, entgegen den Erwartungen, als zu inkonsistent herausgestellt hat, wurde für jeden Zeitpunkt ein neues Modell berechnet. Diese Inkonsistenz kann durch große phänologische Unterschiede im Laufe der Jahreszeiten, variable Sonnenstände und durch die atmosphärischen Korrekturen erklärt werden. Um die beste Leistung der Modelle zu garantieren, wurden im Vorfeld einige Versuche mit unterschiedlichen Parametern und Daten durchgeführt und die Genauigkeiten verglichen. Zum einen wurden verschiedene Versionen der simulierten Ground Truth Daten für das Training und die Validierung des Modells evaluiert, wobei Dynamic World kombiniert mit OSM und Nullwerte für unbekannte Pixel, wie in Abbildung 3 zu sehen ist, die besten Resultate zeigte.

Außerdem wurde der Gebrauch eines Isolation Forest in Betracht gezogen, um



Anomalien in den Daten zu erkennen (Liu et al., 2008) und diese vom Trainieren des Modells auszuschließen. Trotz dieser vielversprechenden Methode stellten Tests fest, dass hauptsächlich Pixel der Klassen offener Boden und Wasser als Ausreißer identifiziert und fälschlicherweise als Urban oder Vegetation klassifiziert wurden. Dies führte zu suboptimalen visuellen Ergebnissen und hatte keine Verbesserung der berechneten Gesamtgenauigkeit zur Folge.

Neben dem Isolation Forest wurden Experimente mit verschiedenen Parametern durchgeführt, darunter die Anzahl der Entscheidungsbäume, die maximale Tiefe und die minimale Anzahl von Datenpunkten, um einen Baum aufzuteilen. Es konnte festgestellt werden, dass eine erhöhte Anzahl von Bäumen und eine größere maximale Tiefe die Genauigkeit etwas steigerten. Die minimale Anzahl von Datenpunkten hatte keinen signifikanten Einfluss auf die Ergebnisse. Basierend auf diesen Erkenntnissen wurde der beste Ansatz ausgewählt, um die Leistung der Klassifikationsmodelle zu optimieren.

Da es sich bei dem Verfahren um eine Machbarkeitsstudie handelt, wurden die Parameter nicht nur nach Genauigkeit der Ergebnisse ausgewählt, sondern der beste Kompromiss aus Performance und Laufzeit genutzt. Das finale Modell basiert wegen seiner hohen Genauigkeit und der vergleichbar kurzen Laufzeit auf einem Random Forest Klassifikator. Durch die hohe Rechengeschwindigkeit und der gleichzeitig nicht merklich schlechteren Ergebnisse, wurde mit 20 eine recht kleine Anzahl an Entscheidungsbäumen gewählt. Mit einer Aufteilung des Datensatzes in 80% Trainings- und 20% Testdaten stehen dem Modell etwas mehr Trainingsdaten zur Verfügung, ohne dass die Leistung darunter leiden muss. Die Trainingsdaten setzen sich zusammen aus OSM, Dynamic World und Nullwerten für unbekannte Pixel. Die Entscheidung fiel gegen einen Isolation Forest und für die besten Werte für die Parameter der maximalen Tiefe von null und der minimalen Samplegröße von zwei.

## 2.5 Automatisierter Prozess

Die zuvor beschriebenen Daten werden in einem automatisierten Python Prozessor aufbereitet und genutzt, um Informationen über die Entwicklung der Landbedeckung in München zu erhalten. Abbildung 4 stellt den gesamten Arbeitsablauf, von den Ausgangsdaten, über die Vorverarbeitung und das Trainieren des Modells bis hin zur Auswertung der Ergebnisse visuell dar.

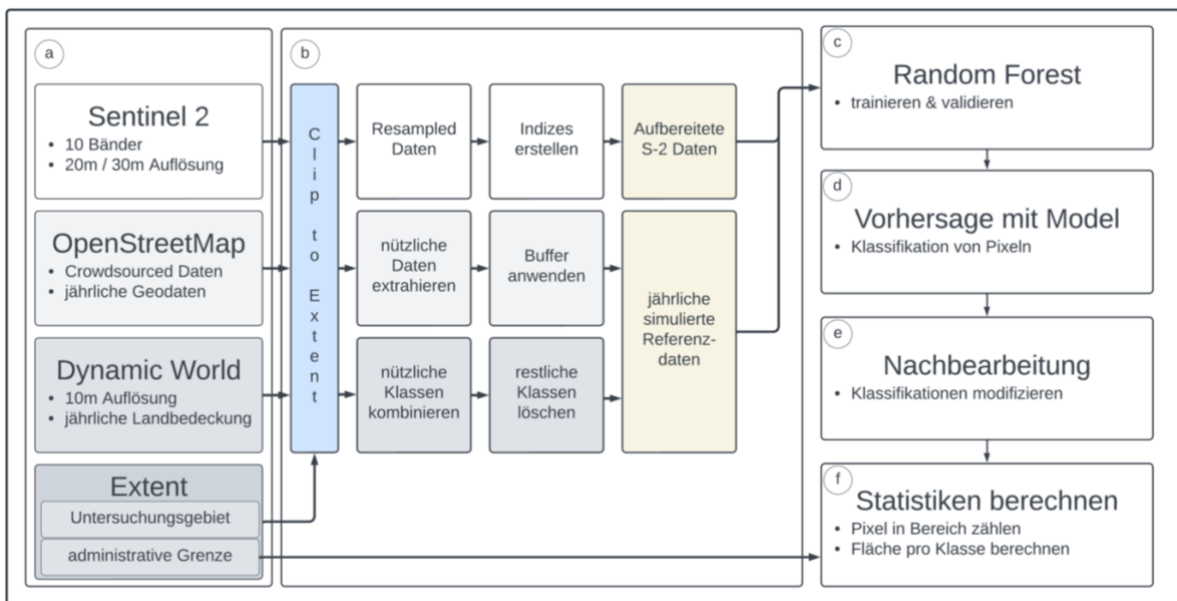


Abbildung 4: Flowchart zum gesamten Prozess.

a) Daten; b) Vorprozessierung; c) Training; d) Vorhersage; e) Nachprozessierung; f) Auswertung der Ergebnisse

### 2.5.1 Preprocess

Um als Trainingsdaten für den Random Forest zu fungieren, müssen die Ausgangsdaten zuerst vorverarbeitet und aufbereitet werden. Sowohl die einzelnen Bänder der Satellitenbilder als auch die Daten für den simulierten Ground Truth durchlaufen einige Schritte, welche im folgenden Abschnitt erläutert werden.

Die erste wichtige Modifikation der Daten ist die Reprojizierung auf das Koordinatenreferenzsystem EPSG:3035 und der Zuschnitt auf das rechteckige Untersuchungsgebiet um München. Dies wird vor allem durchgeführt, sodass später die Pixel der Sentinel-2 Szenen und der Trainingsdaten exakt übereinander liegen. Durch die zuvor angesprochene Variabilität in der räumlichen Auflösung der Satellitendaten, müssen die Bänder mit einer Resolution von 20 Metern auf 10 Meter gebracht werden. Hierzu wird der Cubic-Resampling-Algorithmus verwendet, welcher den Wert des neuen, hochauflösteren Pixels basierend auf den umliegenden Pixeln anhand einer kubischen Interpolation berechnet. Die bearbeiteten Bänder mit einer Auflösung von 10 Metern werden anschließend zu

insgesamt acht Indizes kombiniert, welche dem Machine Learning Algorithmus als weitere Informationen dienen. Die folgenden Indizes wurden aufgrund ihrer wertvollen Eigenschaften im Bereich Vegetation, versiegelte Oberflächen, offener Boden, Wasser und Felder für das Projekt berechnet (Sapena et al., 2024):

- $NDVI = \frac{NIR - RED}{NIR + RED} = \frac{B08 - B04}{B08 + B04}$
- $NDWI = \frac{Green - NIR}{Green + NIR} = \frac{B03 - B08}{B03 + B08}$
- $NDBI = \frac{SWIR1 - NIR}{SWIR1 + NIR} = \frac{B11 - B08}{B11 + B08}$
- $NDTI = \frac{SWIR1 - SWIR2}{SWIR1 + SWIR2} = \frac{B11 - B12}{B11 + B12}$
- $NBAI = \frac{\frac{(SWIR2 - SWIR1)}{Green}}{\frac{(SWIR2 + SWIR1)}{Green}} = \frac{\frac{(B12 - B11)}{B03}}{\frac{(B12 + B11)}{B03}}$
- $NBI = \frac{RED * SWIR2}{NIR} = \frac{B04 * B12}{B08}$
- $BSI = \frac{(SWIR1 + RED) - (NIR + BLUE)}{(SWIR1 + RED) + (NIR + BLUE)} = \frac{(B11 + B04) - (B08 + B02)}{(B11 + B04) + (B08 + B02)}$
- $DBSI = \frac{SWIR1 - GREEN}{SWIR1 + GREEN} - NDVI = \frac{B11 - B03}{B11 + B03} - NDVI$


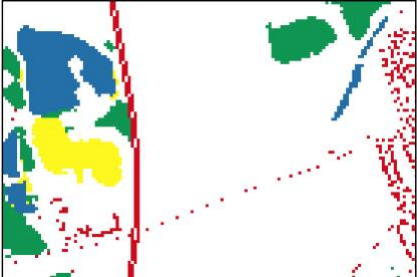

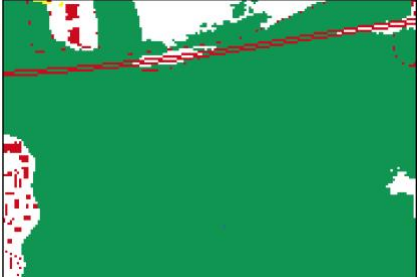



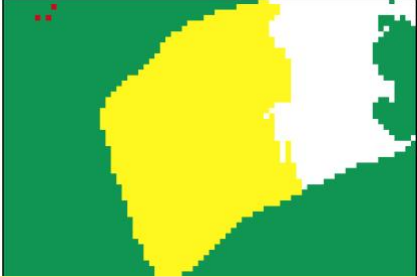


Die berechneten Indizes werden gemeinsam mit den Sentinel-2 Bändern, zu einer einzigen TIF-Datei mit insgesamt 18 Ebenen zusammengeführt. Dieser Prozess wird für alle 59 Zeitpunkte durchgeführt.

Auch die OSM-Daten müssen für den weiteren Verlauf bearbeitet werden. Wie bereits erwähnt, werden die Daten von Gebäuden, Straßen, Bahngleisen, Parkplätzen und Gewässer für die Weiterverarbeitung extrahiert. Von diesen werden einige Arten ausgenommen, welche für das Trainieren des Random Forests nicht sinnvoll sind. Zum einen werden Feuchtgebiete, also z.B. Sümpfe oder Moore, aus den Gewässerdaten gelöscht, da diese häufig fälschlicherweise als Vegetation klassifiziert werden und meist nicht eindeutig als Wasser identifiziert werden können. Außerdem werden U-Bahnen und Schmalspurbahnen aus dem Bahnnetz entfernt. Erstere, da diese auf einem Satellitenbild offensichtlich nicht erkennbar sind – letztere sind oft zu unscheinbar und klein, um als Versiegelungsquelle erkennbar zu sein. Aus dem Straßennetz werden Tunnel, aus dem gleichen Grund wie U-Bahnen, und kleine Pfade und Wege ausgeschlossen. Diese sind häufig durch Baumkronen verdeckt und würden die Genauigkeit der Trainingsdaten

verschlechtern. Da die Straßen- und Bahndaten als Linien vorliegen, müssen sie in Polygone umgewandelt werden. Dies wird erreicht, indem um die Linien ein Puffer angewendet wird, der eine Fläche in einer festgelegten Breite erzeugt. Die Größe der Puffer ist sehr konservativ gewählt worden, um keine falsch klassifizierten oder gemischten Pixel zu riskieren. So wird für Bahngleise beispielsweise ein Puffer von insgesamt einem Meter genutzt, wobei die Breite von Gleisen in Deutschland 1.4 Meter beträgt. Straßen erhalten je nach Typ einen variablen Puffer, sodass Autobahnen größere Flächen als eine Straße innerhalb eines Wohngebiets aufweisen. Abschließend werden die vorverarbeiteten Gewässerdaten gespeichert und die restlichen Daten zu einem urbanen Datensatz zusammengeführt.

Um aus den modifizierten OSM-Vektoren und Dynamic World einen simulierten Ground Truth zu generieren, werden zuerst die benötigten Klassen von Dynamic World kombiniert und mit neuen Werten versehen. Die zwei Vegetationsdaten von Bäumen und Gras werden in einer Klasse mit dem Wert 1 zusammengefasst. Die Klasse offener Boden erhält den Wert 3. Alle anderen, nicht benötigten Landbedeckungsarten, namentlich Wasser, überschwemmte Vegetation, Felder, Gestrüpp, Bebauung, Schnee und Eis, werden mit dem Wert 0 versehen und für den folgenden Trainingsprozess ignoriert. Im Anschluss werden die Gewässer und bebauten Gebiete, welche zuvor aus OSM extrahiert wurden, gerastert und über die Dynamic World Ebene gelegt. Die Pixel, bei denen ein Vektor von OSM vorhanden ist, werden je nach Klasse dem Wert 2 für bebaut oder dem Wert 4 für Wasser gleichgesetzt. Mit den kombinierten Datensätzen erhält man für jedes Jahr einen Trainings- und Validierungsdatensatz mit jeweils 5 Klassen, welche anschließend gemeinsam mit den Sentinel-2 Bildern für das Trainieren des Random Forest genutzt werden können.

Tabelle 1: Zuordnung der Klassenwerte in Trainingsdaten und zugehörige Sentinel-2 Reflexionen

Klasse	Wert	Sentinel-2 vs. Trainingsdaten	
Irrelevante Klassen	0		
Vegetation	1		
Urban	2		
Offener Boden	3		
Wasser	4		

## 2.5.2 Train & Predict

Alle vorprozessierten Sentinel-2 Szenen werden in einer Schleife iteriert und gemeinsam mit den Referenzdaten des jeweiligen Jahres eingelesen. Damit das Machine Learning Modell nur mit validen Daten der vier Klassen Vegetation, Urban, offener Boden und Wasser trainiert wird, müssen zuerst alle Pixel des vorliegenden Satellitenbildes herausgefiltert werden, bei dem der Ground Truth gleich null ist. Nachdem alle Daten im richtigen Format vorliegen, wird der Random Forest mit den zuvor beschriebenen Parametern und einem Trainings- und Testsplit von 80 – 20 trainiert. Die Leistung des trainierten Modells wird anschließend mit den zurückgehaltenen Testdaten bewertet. Für die Validierung wird für jedes Modell die Gesamtgenauigkeit und eine Konfusionsmatrix berechnet. Aus dieser kann die Genauigkeit jeder einzelnen Klasse berechnet werden, indem die Anzahl der korrekten Vorhersagen durch die Gesamtanzahl geteilt wird.

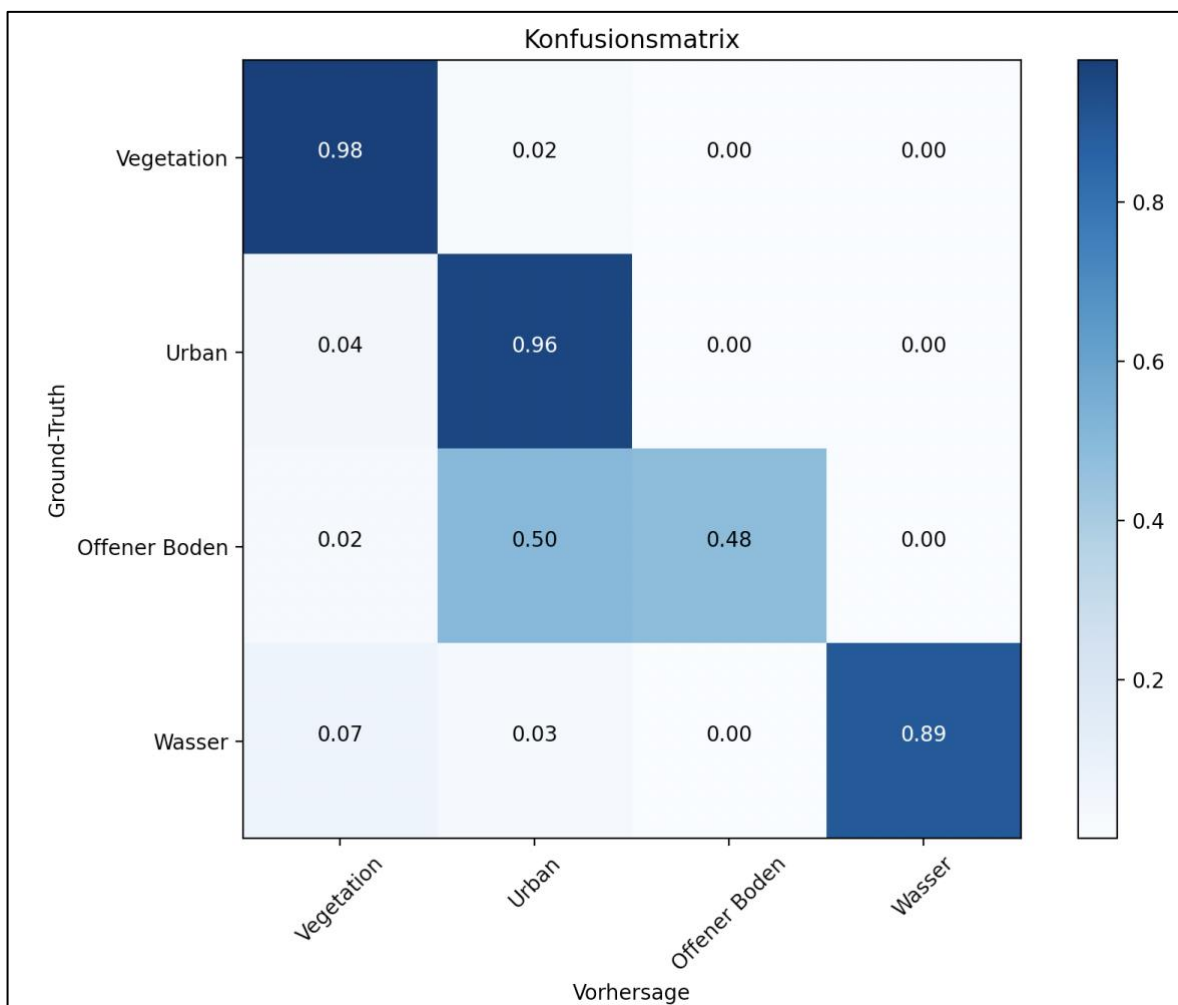


Abbildung 5: Konfusionsmatrix des Random Forest Modells aus August 2020

Die Ergebnisse werden automatisiert in einer Tabelle festgehalten. Abschließend wird jedes Modell lokal gespeichert, wobei die Datei komprimiert wird, um den benötigten Speicherplatz zu minimieren.

Die lokal gespeicherten Modelle werden im Anschluss auf die jeweiligen Sentinel-2 Daten angewendet. Der Random Forest bestimmt für jeden Pixel anhand der Reflexionswerte die wahrscheinlichste Klasse. Die erstellte Klassifikation wird als TIF-Datei exportiert und lokal abgelegt.

### **2.5.3 Postprocess**

Um die Ergebnisse weiter zu verbessern, findet mit den klassifizierten Szenen eine Nachbearbeitung statt. Basierend auf Regeln und der Klassifikation des vorherigen und der drei zukünftigen Szenen werden die vorhergesagten Klassen entweder beibehalten oder überschrieben.

Die Klasse Wasser wird als statisch angesehen, weshalb ein Pixel als Wasser klassifiziert wird, wenn er entweder in der Vergangenheit oder in den simulierten Ground Truth Daten als Wasser definiert wurde. Pixel, die in den aktuellen Daten Wasser sind, aber in der vorherigen eine andere Klasse waren, werden mit der vorherigen Klasse überschrieben, um fälschlich als Wasser klassifizierte Pixel zu korrigieren. Vegetation und offener Boden sind variable Klassen. Ein Pixel, der als Vegetation oder offener Boden bestimmt wurde, kann im nächsten Zeitpunkt eine der drei anderen Klassen annehmen. Ein Pixel wird nur als urban klassifiziert, wenn er entweder schon zuvor als bebaut beurteilt wurde oder er in der aktuellen und in den drei zukünftigen Szenen urban ist. Somit wird die Wahrscheinlichkeit für falsche urbane Pixel deutlich minimiert. Dies ist auch der Grund, weshalb die ersten und letzten drei Beobachtungen aus der Auswertung ausgeschlossen werden, da der Anstieg der Fläche bei diesen deutlich höher ausfällt. Wenn ein Pixel einmal als Urban klassifiziert wurde, kann dieser keinen anderen Wert mehr annehmen, durch die Annahme, dass versiegelte Flächen nicht wieder unversiegelt werden. Außerdem wird die Klasse Felder aus den Dynamic World Daten übernommen, da Felder häufig als bebaut klassifiziert wurden und dies die Ergebnisse deutlich verfälschen würde.

Um aus den finalen Klassifikationen einen Mehrwert zu generieren, muss die Anzahl der Pixel ausgewertet werden. Hierfür werden die Pixel je Klasse innerhalb der administrativen Stadtgrenzen Münchens gezählt. Da ein Pixel die Größe von 10 mal 10 Metern repräsentiert, muss die jeweilige Summe durch 10.000 geteilt werden,

um die Fläche in Quadratkilometern zu erhalten. Die Flächen pro Klasse und Zeitpunkt werden gemeinsam mit der Flächendifferenz zwischen den Szenen und den zuvor berechneten Genauigkeiten der Modelle in einer Tabelle festgehalten. Der abschließende Abgleich der nachbearbeiteten, finalen Klassifikationen mit den echten Ground Truth Daten von LUCAS und der händisch gelabelten Samples gibt Aufschluss über die Genauigkeit der Ergebnisse.

### **3 Ergebnisse**

Es wurden die Ergebnisse der insgesamt 59 Zeitpunkte zusammengetragen. Aufgrund der beschriebenen Methodik und um homogene und konstant generierte Daten auszuwerten, wurden die ersten und letzten drei Szenen aus der Analyse der Resultate ausgeschlossen. Deshalb wird im Folgenden Abschnitt die Genauigkeiten und die Flächenentwicklung in München zu 53 Beobachtungen von Juli 2016 bis August 2023 erläutert.

Abbildung 6 zeigt die Gesamtgenauigkeit, also wie viel Prozent der Pixel richtig vorhergesagt wurden, und die Leistung des Modells je Klasse über den eben genannten Zeitraum. Die durchschnittliche allgemeine Genauigkeit beträgt 95.6%, mit einem Höchstwert von 96.5% und einem niedrigsten Wert von 93.3%. Noch etwas besser wurde Vegetation klassifiziert. Hier betrug der Durchschnitt 97.2% mit einer Spanne von 96% bis 97.9%. Mit einer maximalen Genauigkeit von 95.8%, einem Tiefstwert von 91.4% und einer durchschnittlichen Prozentanzahl von 94.6 fielen die Ergebnisse der urbanen Klasse minimal schlechter aus. Die Klasse Wasser lässt sich im Bereich zwischen 72.2 und 89.9-prozentiger Genauigkeit wiederfinden. Der Durchschnitt für diese Klasse beträgt 85.3%. Offener Boden ist deutlich abgeschlagen mit Werten von 34.3% bis 59.6% und einem Durchschnitt von 49.7 %.



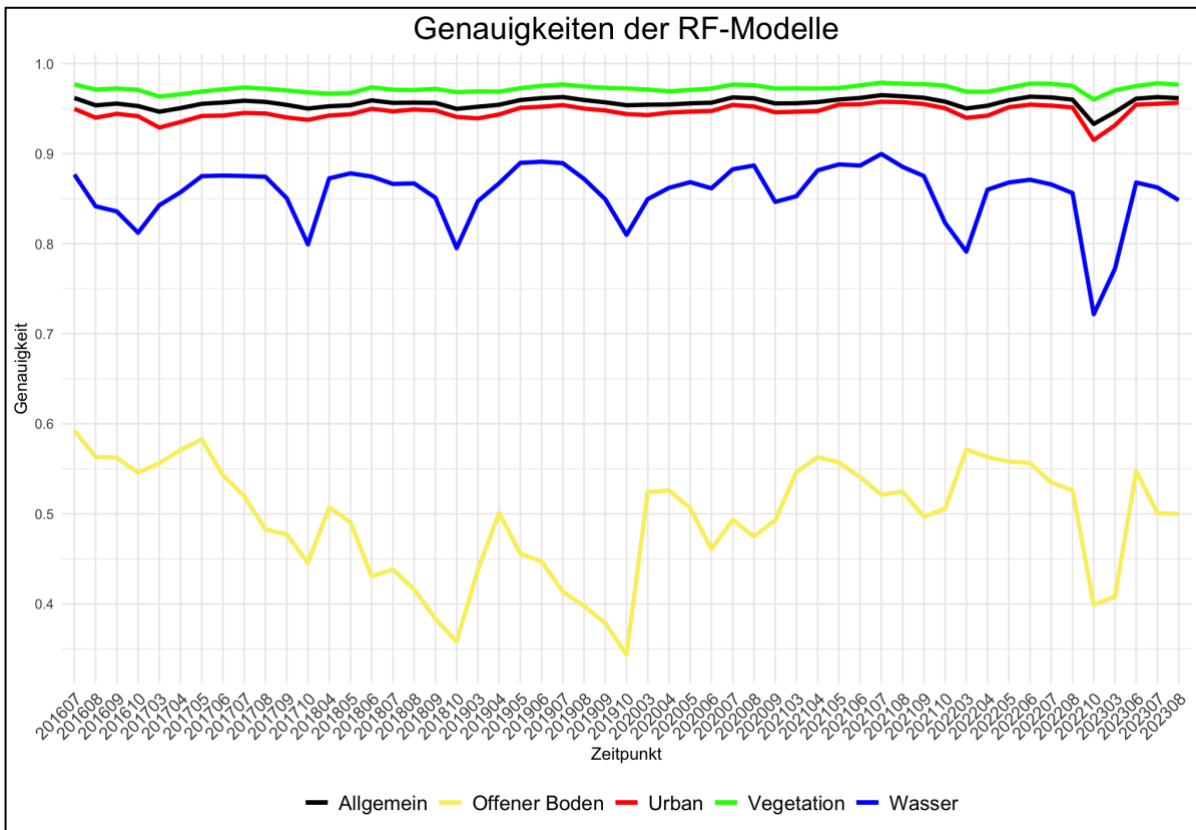


Abbildung 6: Genauigkeiten je Klasse aller Modelle

Innerhalb des Beobachtungszeitraums lässt sich für die Klassen Vegetation, Urban und Wasser ein klarer Trend erkennen. Tendenziell funktioniert das Modell in den Sommermonaten besser als im Herbst und Frühling. Gerade die Klasse Wasser zeigt diesen Verlauf deutlich. Dies ist aufgrund der zu- und abnehmenden Vegetation und der damit einhergehenden Änderung der Reflexionswerte zu erklären. Für die Klasse offener Boden ist kein klarer Trend erkennbar. Die schlechtere Leistung des Modells für diese Klasse kann zum einen auf die wenigen Trainingsdaten zurückgeführt werden, weshalb offener Boden unterrepräsentiert ist. Die Ungenauigkeiten der Dynamic World Daten für diese Klasse spiegeln sich in den Ergebnissen wider. Gleichzeitig ist diese Kategorie auch die variabelste und dynamischste, beispielsweise aufgrund von neuen oder fertiggestellten Baustellen, was eine konstante Vorhersage erschwert.

Genauigkeiten über 90% sind vielversprechend und für eine Machbarkeitsstudie als sehr gut zu bewerten. Gerade die gute Performance für die zwei bedeutendsten Klassen dieser Studie, Urban und Vegetation, ist sehr positiv und spricht für die Methodik.

Die Genauigkeiten der finalen, nachbearbeiteten Klassifikationen, überprüft mithilfe der angereicherten LUCAS Daten, zeigten etwas kleinere Werte. Der Vergleich

zwischen den 2018 Ground Truth Daten mit der Klassifikation von August 2018 ergab eine Gesamtgenauigkeit von 85%. Die Klasse Felder und Urban wurden zu 92% richtig klassifiziert. Die Genauigkeit von Vegetation beträgt 84% und von Wasser 88%. Von den 6 Kontrollpunkte der Kategorie offener Boden wurde keiner richtig bestimmt. Für 2022 beträgt die allgemeine Genauigkeit 83%, während Felder zu 91%, Vegetation zu 82%, Urban und Wasser zu 90% und offener Boden zu 22% richtig bestimmt wurden.

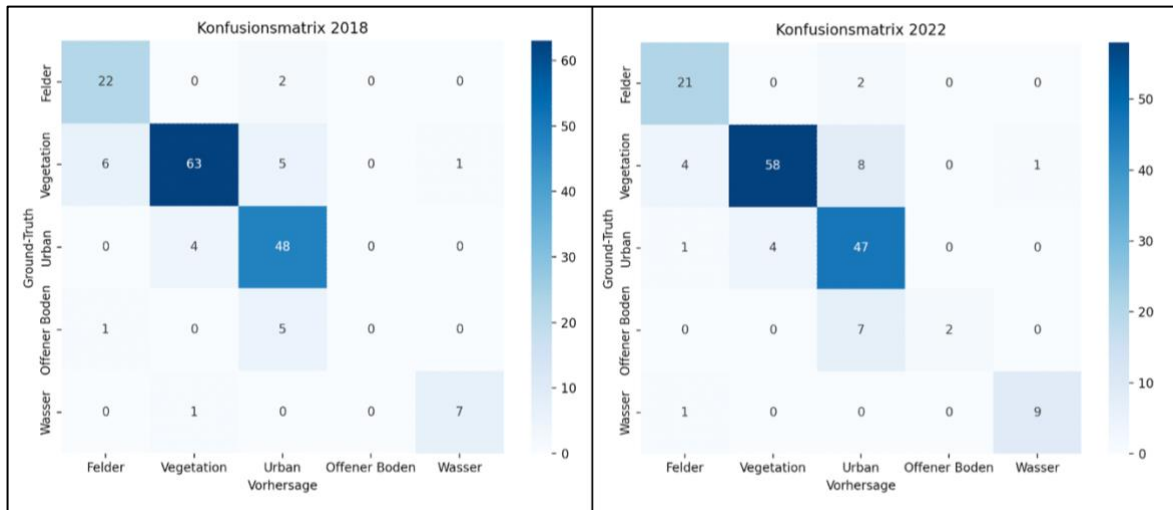


Abbildung 7: Konfusionsmatrix zur Validierung der Klassifikation aus August 2018 & 2022 gegenüber den LUCAS Ground Truth Daten von 2018 (links) und 2022 (rechts)

Das eigentliche Ziel der Studie, die Flächenentwicklung und den Anstieg von versiegelten Gebieten innerhalb Münchens aufzuzeigen, wird in Abbildung 8 visualisiert. In der Grafik sieht man einen recht konstanten Anstieg der urbanen Fläche, während die Vegetation korrelierend dazu abnimmt. Für Juli 2016 wurde eine urbane Fläche von 151.9 km<sup>2</sup> berechnet, welche bis August 2023 auf 173.4 km<sup>2</sup> anstieg. Die Grünflächen fielen im gleichen Zeitraum von 113.6 km<sup>2</sup> auf 89.5 km<sup>2</sup>. Dies entspricht einem Zuwachs von 21.5 km<sup>2</sup> urbaner und einem Verlust von 24.1 km<sup>2</sup> vegetativer Fläche innerhalb der acht Jahre. Die Klassen Wasser und Felder werden nicht im Diagramm aufgeführt, da sich die Fläche innerhalb eines Jahres nicht verändert und der größte Mehrwert aus den beiden anderen Kategorien gezogen werden kann. Durch die Unterrepräsentation sind die Veränderungen im offenen Boden zu marginal und verlaufen mit der Zeit gleich null, weshalb auch diese nicht im Diagramm zu sehen sind. Leider zeigen die Daten nicht, dass einem Anstieg in offenem Boden ein Anstieg der urbanen Fläche folgt.

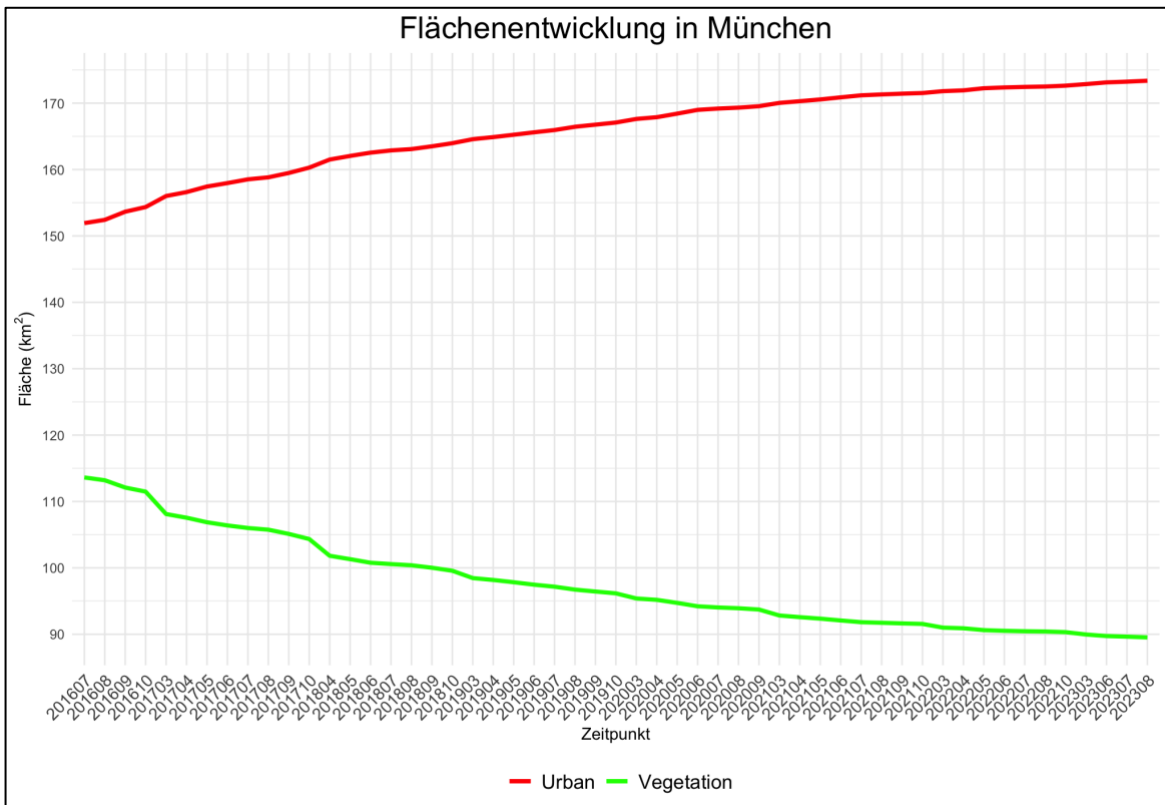


Abbildung 8: Entwicklung der urbanen und vegetativen Fläche

Ein weiterer interessanter Punkt ist die Betrachtung des Deltas, also der Veränderung der Fläche zwischen zwei Zeitpunkten. Der durchschnittliche Zuwachs an urbaner Fläche zwischen zwei Zeitpunkten beträgt  $0.72 \text{ km}^2$ , während die Vegetation um  $-0.73 \text{ km}^2$  schrumpft. Wie in Abbildung 9 zu erkennen ist, kann immer am Jahreswechsel ein größerer Anstieg in der urbanen Fläche beobachtet werden. So betrug beispielsweise im Wechsel von 2016 auf 2017 das Delta der urbanen Klasse  $1.66 \text{ km}^2$  und  $-3.37 \text{ km}^2$  der Vegetation. Durch die fehlenden Wintermonate ist der zeitliche Abstand zwischen diesen Beobachtungen größer, was ein größeres Delta wahrscheinlicher macht. Außerdem nimmt das Delta im Laufe der Zeit immer weiter ab. Dies ist jedoch eher auf die Methodik zurückzuführen als auf eine abfallende Neubauaktivität im Raum München.

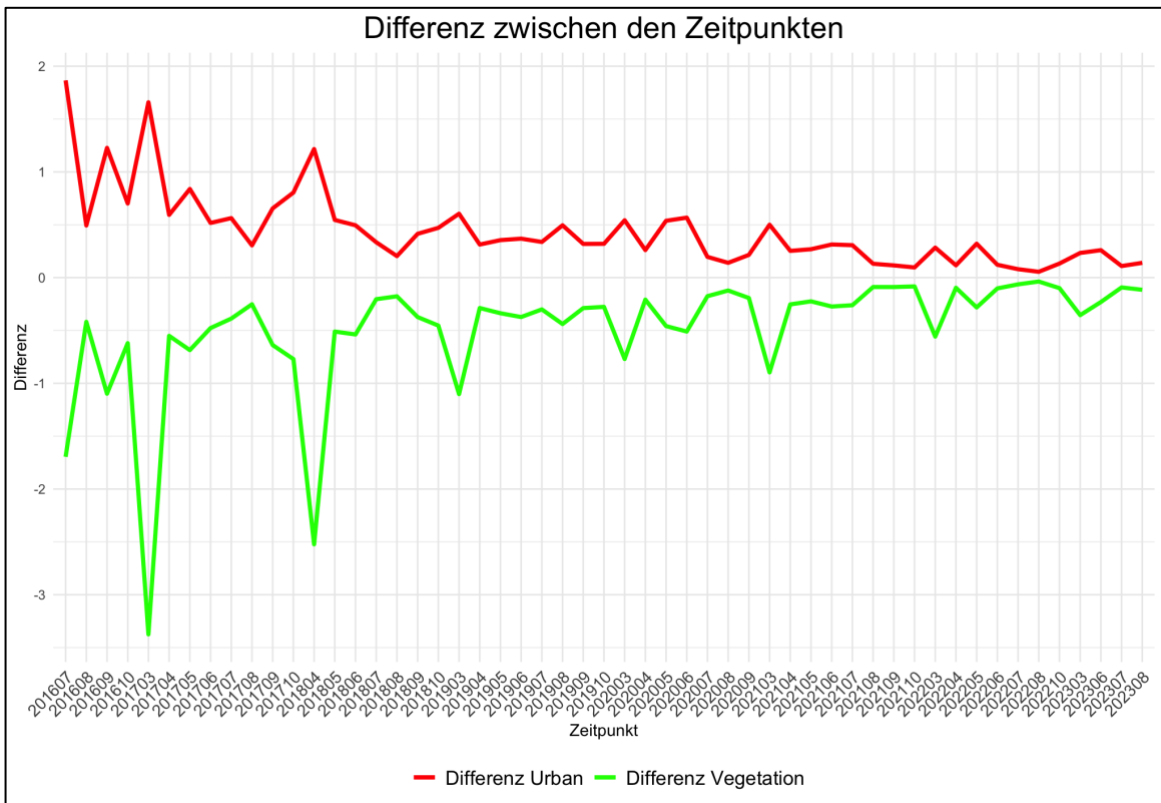


Abbildung 9: Deltas der urbanen und vegetativen Fläche innerhalb des Zeitraums

Offizielle Zahlen der Stadt München besagen, dass im Jahr 2019 44 % des Stadtgebiets versiegelt war (Landeshauptstadt München, 2020). Im Juli 2019 beträgt das Ergebnis unserer Berechnungen eine versiegelte Fläche von 165.9 km<sup>2</sup>, was 53 % der Gesamtfläche Münchens ausmacht. Generell konnte beobachtet werden, dass die Ergebnisse der Methodik knapp 10 % über offiziellen Werten liegen.

Der visuelle Vergleich in Abbildung 10 gibt weitere Aufschlüsse darüber, wie und in welchen Regionen sich die Flächennutzung über die Zeit am meisten ändert.

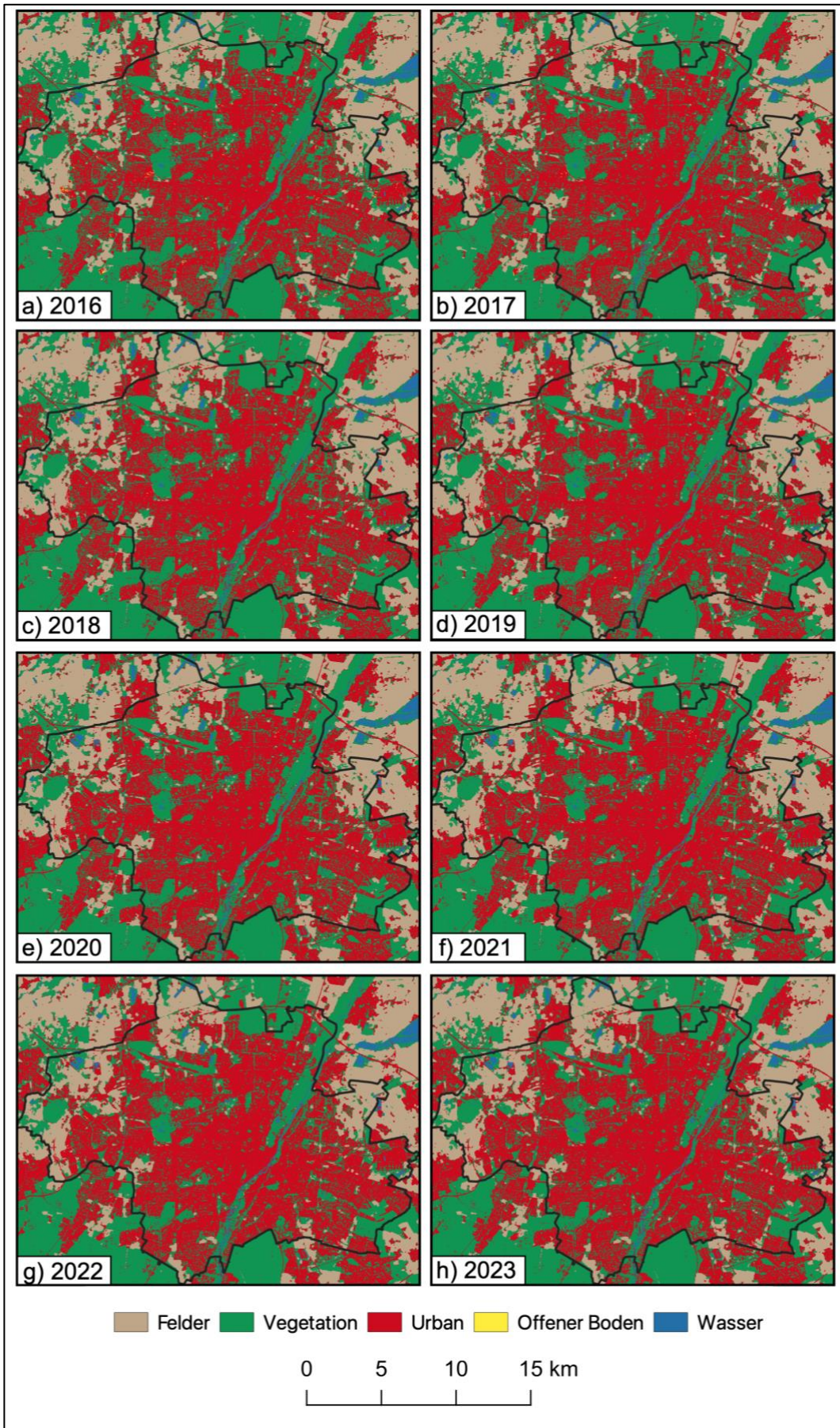


Abbildung 10: Visualisierung der Landnutzungsänderung in München von 2016 bis 2023

Die Veränderungen werden in Abbildung 11 noch klarer deutlich. Die Karte zeigt die Veränderung der urbanen Fläche zwischen 2016 und 2023 im Gesamtgebiet und in einer Detailansicht. Der vergrößerte Ausschnitt zeigt unter anderem den im Untersuchungsgebiet beschriebenen Stadtteil Freiham. Die Expansion im Westen der Stadt kann klar nachvollzogen werden.

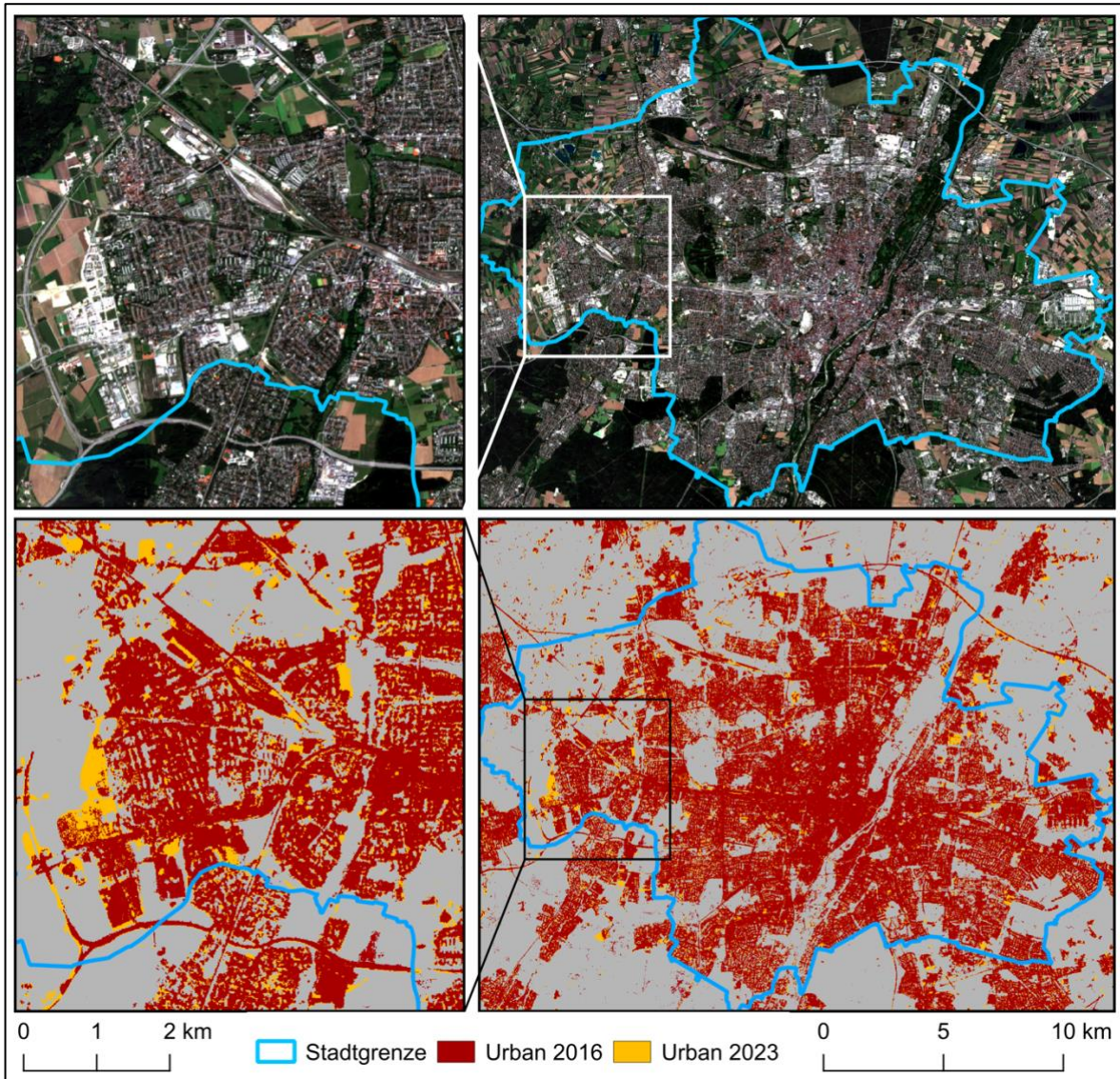


Abbildung 11: Close-Up zu urbaner Expansion im Raum Freiham zwischen 2016 und 2023

Laut offiziellen Zahlen der Stadt, entsteht der erste Realisierungsabschnitt des neuen Wohn- und Arbeitsraums in Freiham auf einem Areal von 85 Hektar (Landeshauptstadt München, 2023), was 0.85 km<sup>2</sup> entspricht. Die Umsetzung von 4.400 Wohnungen, einem Bildungscampus und weiterer Infrastruktur reicht von 2016 bis voraussichtlich 2026. Unser Modell zeigt im Stadtteil Freiham einen Zuwachs der urbanen Fläche im Zeitraum von 2016 bis 2023 von ca. 0.66 km<sup>2</sup>.

Beispiele wie diese unterstreichen die Effektivität der Methode und den Wert der Ergebnisse, um die Expansion von Städten zu analysieren.

#### **4 Diskussion**

Trotz der hohen Genauigkeiten und der vielversprechenden Ergebnisse sind im Laufe der Umsetzung einige Diskussionspunkte angefallen, welche nicht unerwähnt bleiben sollten.

Es ist wichtig anzumerken, dass trotz der bereits erzielten Ergebnisse durch die Verwendung von Sentinel-2-Daten mit einer Auflösung von 10 Metern, die Erfassung feinerer Details auf der Erdoberfläche nahezu unmöglich ist. So können beispielsweise schmale versiegelte Pfade oder Bäume und kleine Grünflächen innerhalb einer urbanen Umgebung nicht erkannt werden. Dies wirkt sich negativ auf die Genauigkeit der Klassifizierung aus, insbesondere bei der Unterscheidung von Landbedeckungen mit ähnlichen spektralen Signaturen. Daher wäre es interessant, die Studie erneut mit hochauflösten Satellitendaten durchzuführen und die Ergebnisse zu vergleichen. Durch die höhere Resolution wären feinere Details erkennbar und mehr Trainingsdaten durch die größere Anzahl an Pixeln verfügbar, was vermutlich zu einer besseren Leistung des Modells und einer genaueren Klassifizierung führen würde.

Auch die präzisere Identifizierung und Unterscheidung innerhalb von Flächentypen wie Gebäuden, Verkehrsflächen oder Erholungsgebieten könnte durch eine höhere Auflösung ermöglicht werden. Eine geteerte Straße hat beispielsweise einen höheren Versiegelungsanteil als ein Schotterweg, weshalb eine stärkere Differenzierung sinnvoll wäre. So könnten Flächen weiter in ihre prozentuale Versiegelung unterteilt und aussagekräftigere Ergebnisse generiert werden, wie die Analyse der Bodenversiegelung in Sachsen anhand von Luftbildern mit einer Bodenauflösung von 20 cm zeigte (Siemer et al., 2022).

Der Einsatz eines hochauflösenden Sensors könnte darüber hinaus die Anwendung einer objektbasierten Bildanalyse (OBIA) möglich machen. Die Identifikation von Objekten würde weitere Informationen über die Bodenbedeckung bereitstellen und somit eine bessere Klassifikation möglich machen. So wurde das Problem der Landnahme in einer periurbanen Umgebung Venedigs im Zeitraum von 2005 und 2010 anhand einer objektbasierten Methode und hochauflösenden Fernerkundungsdaten mit vielversprechenden Ergebnissen analysiert (Pugliese & Scarpetta, 2014).

Neben der Verwendung eines OBIA-Ansatzes wäre ein Vergleich zwischen dem in dieser Studie genutzte Random Forest Klassifikator und einem Random Forest Regressor, als alternative Machine Learning Methode, von Interesse. Die Kartierung von urbanen, undurchlässigen Flächen auf Subpixelebene mittels RF-Regressor und Landsat Bildern kann aussichtsreiche Ergebnisse und hohe Genauigkeiten für die Überwachung von versiegelten Flächen liefern (Deng & Zhu, 2020).

Ein weiterer Diskussionspunkt ist die schlechte Genauigkeit der Klasse offener Boden. Wie zuvor angesprochen resultiert diese vermutlich aus der Unterrepräsentation und Ungenauigkeit in den Trainingsdaten von Dynamic World und der hohen Variabilität der Klasse. Da offener Boden in der Theorie jedoch als ein starker Prädiktor für Bebauung in zukünftigen Monaten und Jahren dienen kann, wäre eine gute Leistung des Modells erstrebenswert. Neben der Nutzung von hochaufgelösten Satellitendaten könnten zusätzlich Datenquellen und manuell gelabelte Polygone herangezogen werden, um dem Random Forest Algorithmus bessere und mehr Trainingsdaten für die Klasse offener Boden zur Verfügung zu stellen, um so höhere Genauigkeiten zu erreichen.

Zuletzt muss diskutiert werden, inwiefern OSM und Machine Learning Produkte wie Dynamic World wirklich geeignete Trainingsdaten darstellen und für die Generierung eines „Ground Truths“ genutzt werden können. Wie zuvor aufgezeigt, haben sowohl OSM als auch Dynamic World viele positive Eigenschaften, was sie gerade für skalierbare Ansätze zu wertvollen Datenquellen macht. Trotzdem können ein gewisser Bias und fehlerhafte Informationen nicht ausgeschlossen und eliminiert werden. Ein noch ausführlicherer Vergleich zwischen den generierten und echten Ground Truth Daten wäre nötig, um fundierte Aussagen über die Qualität der simulierten Referenzdaten treffen zu können. Dies wurde durch die Validierung mit den LUCAS Daten für zwei Zeitpunkten angedeutet und sollte für eine höhere Gewissheit noch ausgeweitet werden. Die Integration der Klassenwahrscheinlichkeiten von Dynamic World-Pixeln (Brown et al., 2020) könnte ebenfalls dazu beitragen, die Leistung des Modells zu verbessern.

Durch die aufgeführten Punkte wird deutlich, dass trotz der erzielten Ergebnisse innerhalb dieser Machbarkeitsstudie noch Raum für Optimierung der Methodik besteht, um die Genauigkeit weiter zu verbessern.



## 5 Zusammenfassung

Diese Studie untersuchte die Entwicklung der Landbedeckung und Bodenversiegelung in München von Juli 2016 bis August des Jahres 2023 mithilfe von Sentinel-2-Satellitendaten und eines Random Forest Machine Learning Modells. Durch die automatisierte Klassifizierung der Landbedeckung konnte der ansteigende Trend der Urbanisierung und die Veränderungen in Vegetationsflächen aufgezeigt werden.

Die Endergebnisse der Analyse basieren auf 53 Zeitpunkten innerhalb der acht Jahre, welche eine Gesamtgenauigkeit des Modells von 95.6% zeigte. Die Klassen Vegetation und Urban, auf welchen das Hauptaugenmerk des Verfahrens lag, überzeugten mit Genauigkeiten von 97.2% bzw. 94.6%. Im Vergleich der Klassifikationen mit den editierten LUCAS Daten wurden für 2018 und 2022 allgemeine Genauigkeiten von 85% und 83% erreicht. Die urbane Fläche wurde dabei zu 92% und 90% richtig bestimmt. Dabei war die Leistung des Modells in den Sommermonaten besser als im Frühling und Herbst.

Es konnte gezeigt werden, dass innerhalb der letzten Jahre ein Anstieg der versiegelten Fläche stattgefunden hat, welcher mit dem Rückgang der vegetativen Flächen korreliert. Im Detail nahm die urbane Fläche laut den berechneten Daten von 151,9 km<sup>2</sup> im Juli 2016 auf 173,4 km<sup>2</sup> im August 2023 zu, während die Grünflächen im gleichen Zeitraum von 113,6 km<sup>2</sup> auf 89,5 km<sup>2</sup> zurückgingen.

Trotz der angesprochenen Diskussionspunkte im Hinblick auf die Ausgangsdaten konnte gezeigt werden, dass die Methodik valide Trainingsdaten für einen skalierbaren Ansatz mit aussichtreichen Ergebnissen liefern kann. Dadurch konnten bereits für andere Städte wie Hamburg, Frankfurt oder Würzburg erfolgreich Daten über die Landentwicklung mit geringem Aufwand generiert werden.

Gerade der Flächenvergleich der Klassifikationsergebnisse mit offiziellen Zahlen im Stadtteil Freiham zeigt, dass die beschriebene Methodik sehr vielversprechend ist, um Statistiken über die Landbedeckung und Bodenversiegelung in und um Städte zu erlangen. Die generierten Informationen bieten einen wertvollen Einblick in die Flächenentwicklung und können dazu beitragen, städtische Planungsmaßnahmen zu unterstützen, um den genannten negativen Auswirkungen der Versiegelung entgegenzuwirken und nachhaltige Lösungen für die Erhaltung von Grünflächen und Biodiversität zu fördern.

## 6 Literaturverzeichnis

- Birch, E. L., & Wachter, S. M. (Eds.). (2011). *Global urbanization*. University of Pennsylvania Press.
- Breiman, L. (2001). Random forests. *Machine learning*, 45, 5-32.
- Brown, C. F., Brumby, S. P., Guzder-Williams, B., Birch, T., Hyde, S. B., Mazzariello, J., ... & Tait, A. M. (2022). Dynamic World, Near real-time global 10 m land use land cover mapping. *Scientific Data*, 9(1), 251.
- Deng, C., & Zhu, Z. (2020). Continuous subpixel monitoring of urban impervious surface using Landsat time series. *Remote Sensing of Environment*, 238, 110929.
- Eurostat (Statistical office of the European Union). (2024). Land Cover/Use Statistics (LUCAS) Database. Eurostat, Luxemburg. Zugriff am 02.05.24, unter <https://ec.europa.eu/eurostat/web/lucas>
- Fokaides, P. A., Kylili, A., Nicolaou, L., & Ioannou, B. (2016). The effect of soil sealing on the urban heat island phenomenon. *Indoor and Built Environment*, 25(7), 1136-1147.
- Fonte, C. C., & Martinho, N. (2017). Assessing the applicability of OpenStreetMap data to assist the validation of land use/land cover maps. *International Journal of Geographical Information Science*, 31(12), 2382-2400.
- Foody, G. M. (2002). Status of land cover classification accuracy assessment. *Remote sensing of environment*, 80(1), 185-201.
- Garcia, P., & Perez, E. (2016). Mapping of soil sealing by vegetation indexes and built-up index: A case study in Madrid (Spain). *Geoderma*, 268, 100-107.
- Gascon, F., Cadau, E., Colin, O., Hoersch, B., Isola, C., Fernández, B. L., & Martimort, P. (2014, September). Copernicus Sentinel-2 mission: products, algorithms and Cal/Val. In *Earth observing systems XIX* (Vol. 9218, pp. 455-463). SPIE.
- Gombe, K. E., Asanuma, I., & Park, J. G. (2017). Quantification of annual urban growth of Dar es Salaam Tanzania from Landsat time Series data. *Advances in Remote Sensing*, 6(3), 175-191.
- Haas, J., & Ban, Y. (2018). Urban land cover and ecosystem service changes based on Sentinel-2A MSI and Landsat TM Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(2), 485-497.
- Hagolle, O., Colin, J., Coustance, S., Kettig, P., d'Angelo, P., Auer, S., ... & Desjardins, C. (2021). Sentinel-2 surface reflectance products generated by CNES and DLR: methods, validation and applications. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 51, 9-15.
- Haklay, M., & Weber, P. (2008). Openstreetmap: User-generated street maps. *IEEE Pervasive computing*, 7(4), 12-18.
- Kampouraki, M., Wood, G. A., & Brewer, T. (2006, July). The application of remote sensing to identify and measure sealed areas in urban environments. In *1st International Conference on Object-based Image Analysis (OBIA 2006)*.
- Khanal, N., Uddin, K., Matin, M. A., & Tenneson, K. (2019). Automatic detection of spatiotemporal urban expansion patterns by fusing OSM and landsat data in Kathmandu. *Remote Sensing*, 11(19), 2296.

- Kopfmüller, J. (2016). Urbanisierung. *Handbuch Umweltethik*, 312-320.
- Landeshauptstadt München, Referat für Klima- und Umweltschutz (2020). Bodenversiegelung in München. Zugriff am 29.04.24, unter [https://stadt.muenchen.de/infos/bodenversiegelung\\_muenchen.html](https://stadt.muenchen.de/infos/bodenversiegelung_muenchen.html)
- Landeshauptstadt München, Referat für Stadtplanung und Bauordnung (2022). Stadtentwicklungskonzeption „Perspektive München“. Zugriff am 23.04.24, unter [https://lhm.muenchen.swm.de/dam/jcr:b129023c-5202-4788-9f5e-a38b47b83776/2021\\_Fortschreibung\\_PerspektiveMuenchen.pdf](https://lhm.muenchen.swm.de/dam/jcr:b129023c-5202-4788-9f5e-a38b47b83776/2021_Fortschreibung_PerspektiveMuenchen.pdf)
- Landeshauptstadt München, Referat für Stadtplanung und Bauordnung (2023). Freiham – ein neues Stadtquartier. Zugriff am 23.04.24, unter <https://stadt.muenchen.de/infos/neuer-stadtteil-freiham.html>
- Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008, December). Isolation forest. In *2008 eighth IEEE international conference on data mining* (pp. 413-422). IEEE.
- Lozán, J. L., & Breckle, S. W. (2021). 5.2 Bodenversiegelung und Änderung des natürlichen Wasserkreislaufs.
- Luo, N., Wan, T., Hao, H., & Lu, Q. (2019). Fusing high-spatial-resolution remotely sensed imagery and OpenStreetMap data for land cover classification over urban areas. *Remote Sensing*, *11*(1), 88.
- Luti, T., Segoni, S., Catani, F., Munafò, M., & Casagli, N. (2020). Integration of remotely sensed soil sealing data in landslide susceptibility mapping. *Remote Sensing*, *12*(9), 1486.
- McGranahan, G., & Satterthwaite, D. (2014). *Urbanisation concepts and trends* (Vol. 220). International Institute for Environment and Development.
- Naumann, S., Frelüh-Larsen, A., Prokop, G., Ittner, S., Reed, M., Mills, J., ... & Miturski, T. (2019). Land take and soil sealing—drivers, trends and policy (legal) instruments: insights from European cities. *International Yearbook of Soil Law and Policy 2018*, 83-112.
- OpenStreetMap Wiki (2024). Zugriff am 24.04.2024, unter <https://wiki.openstreetmap.org/wiki>
- Orgiazzi, A., Ballabio, C., Panagos, P., Jones, A., & Fernández-Ugalde, O. (2018). LUCAS Soil, the largest expandable soil dataset for Europe: a review. *European Journal of Soil Science*, *69*(1), 140-153.
- Peroni, F., Pappalardo, S. E., Facchinelli, F., Crescini, E., Munafò, M., Hodgson, M. E., & De Marchi, M. (2022). How to map soil sealing, land take and impervious surfaces? A systematic review. *Environmental Research Letters*, *17*(5), 053005.
- Pesaresi, M., Ehrlich, D., Ferri, S., Florczyk, A. J., Freire, S., Halkia, M., ... & Syrris, V. (2016). *Operating procedure for the production of the Global Human Settlement Layer from Landsat data of the epochs 1975, 1990, 2000, and 2014* (pp. 1-62). Luxembourg: Publications Office of the European Union.
- Phiri, D., Simwanda, M., Salekin, S., Nyirenda, V. R., Murayama, Y., & Ranagalage, M. (2020). Sentinel-2 data for land cover/use mapping: A review. *Remote Sensing*, *12*(14), 2291.

- Prokop, G., Jobstmann, H., & Schönbauer, A. (2011). Overview of best practices for limiting soil sealing or mitigating its effects in EU-27. *European Communities*, 227, 24.
- Pugliese, L., & Scarpetta, S. (2014). An object based analysis applied to very high resolution remote sensing data for the change detection of soil sealing at urban scale. In *Recent Advances of Neural Network Models and Applications: Proceedings of the 23rd Workshop of the Italian Neural Networks Society (SIREN), May 23-25, Vietri sul Mare, Salerno, Italy* (pp. 155-162). Springer International Publishing.
- Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS journal of photogrammetry and remote sensing*, 67, 93-104.
- Röpke, L., & Lippelt, J. (2014). Kurz zum Klima: Bodenversiegelung in Deutschland und Europa. *ifo Schnelldienst*, 67(03), 60-63.
- Sapena, M., Mast, J., Schoepfer, E., & Taubenböck, H. Do-it-Yourself Built-Up Mapping Tool: A Practical Cloud-Based Solution Using Sentinel Imagery for Mapping Urban Expansion in Africa. Available at SSRN 4762416.
- Scalenghe, R., & Marsan, F. A. (2009). The anthropogenic sealing of soils in urban areas. *Landscape and urban planning*, 90(1-2), 1-10.
- Schubert, K., & Klein, M. (2020). Das Politiklexikon (7., aktual. u. erw. Aufl.). Bonn: Dietz. (Lizenzausgabe Bonn: Bundeszentrale für politische Bildung)
- Siemer, B., Ferber, U., & Eckert, K. (2022). Erfassung der Bodenversiegelung im Freistaat Sachsen. In *Dresdner Flächennutzungssymposium* (Vol. 80, pp. 267-275). Rhombos-Verlag.
- Statistische Ämter des Bundes und der Länder (2022). Städte mit dem höchsten Bruttoinlandsprodukt (BIP) in Deutschland im Jahr 2020 [Data set]. Zitiert nach de.statista.com. Zugriff am 13.05.24, unter <https://de.statista.com/statistik/daten/studie/1211712/umfrage/staedte-hoechstes-bruttoinlandsprodukt/>
- Taubenböck, H., Weigand, M., Esch, T., Staab, J., Wurm, M., Mast, J., & Dech, S. (2019). A new ranking of the world's largest cities—Do administrative units obscure morphological realities?. *Remote Sensing of Environment*, 232, 111353.
- The World Bank. World Development Indicators (2022). *Urban population (% of total population)*. Zugriff am 22.04.24 unter <https://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS>
- UBA – Umweltbundesamt (2024). Bodenversiegelung. Zugriff am 22.04.24, unter <https://www.umweltbundesamt.de/daten/flaeche-boden-land-oekosysteme/boden/bodenversiegelung#bodenversiegelung-in-deutschland>
- United Nations, Department of Economic and Social Affairs, Population Division (2019). *World Urbanization Prospects: The 2018 Revision (ST/ESA/SER.A/420)*. New York: United Nations.
- Vargas-Munoz, J. E., Srivastava, S., Tuia, D., & Falcao, A. X. (2020). OpenStreetMap: Challenges and opportunities in machine learning and remote sensing. *IEEE Geoscience and Remote Sensing Magazine*, 9(1), 184-199.

- Venter, Z. S., Barton, D. N., Chakraborty, T., Simensen, T., & Singh, G. (2022). Global 10 m land use land cover datasets: A comparison of dynamic world, world cover and esri land cover. *Remote Sensing*, 14(16), 4101.
- Viana, C. M., Encalada, L., & Rocha, J. (2019). The value of OpenStreetMap historical contributions as a source of sampling data for multi-temporal land use/cover maps. *ISPRS international journal of geo-information*, 8(3), 116.
- Wu, Z., Li, H., & Zipf, A. (2020, July). From Historical OpenStreetMap data to customized training samples for geospatial machine learning. In *Proceedings of the Academic Track at the State of the Map 2020 Online Conference*.
- Xiao, R., Su, S., Zhang, Z., Qi, J., Jiang, D., & Wu, J. (2013). Dynamics of soil sealing and soil landscape patterns under rapid urbanization. *Catena*, 109, 1-12.

## **Eidesstattliche Erklärung**

Ich erkläre hiermit eidesstattlich, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe. Alle Stellen, die wörtlich oder inhaltlich den angegebenen Quellen entnommen wurden, sind als solche kenntlich gemacht.

Die vorliegende Arbeit wurde bisher in gleicher oder ähnlicher Form noch nicht als Seminar-Arbeit, Bachelor-Arbeit oder Master-Arbeit eingereicht.

Salzburg, 03.06.2024

---

*Ort, Datum*

---

*Unterschrift*