
Offene Verwaltungsdaten zur Analyse des Befallspotenzials von Grünlandbeständen mit Schadpflanzen am Beispiel von Kreuzkräutern

Autor:

M. Sc. Philipp Zacharias
Universität Rostock
Agrar- und Umweltwissenschaftliche Fakultät
Geodäsie und Geoinformatik (GG)
Justus-von-Liebig Weg 6
D-18059 Rostock, Deutschland
E-Mail: philipp.zacharias@uni-rostock.de

Zusammenfassung

Für die seit einigen Jahren vermehrt im Grünland auftretenden, giftigen Kreuzkrautarten Jakobs-Kreuzkraut (*Senecio jacobaea* L.) und Wasser-Kreuzkraut (*Senecio aquaticus* Hill s. str.) wurde das Befallspotenzial der Grünlandflächen im Stadtgebiet Hamburg näher untersucht. Eine Gruppe von Flurstücken, deren Befallszustand durch vegetationskundliche Kartierungen festgehalten wurde, diente als Ground-Truth Stichproben. Die Datenbasis wurde aus offenen, digitalen Geodaten des Hamburger Transparenzportals erstellt (z. B. digitales Höhenmodell, amtliche Bodenschätzungskarte), welche für diese Analyse in ein gemeinsames Untersuchungs-gitter aufgenommen wurden. Anschließend wurde der Datensatz durch Machine Learning Algorithmen hinsichtlich des Befalls ausgewertet. Es konnte gezeigt werden, dass das Höhenmodell, die Bodenart sowie die Grünlandzahl einen wesentlichen Einfluss auf die Ausbreitung von Jakobs-Kreuzkraut haben. Das Vorhandensein von Wasser-Kreuzkraut wird vor allem durch den hydrologischen Profiltyp, die Bodengesellschaft, die Genese sowie das Höhenmodell definiert. Die Ergebnisse der Klassifizierung wurden in Kartenform visualisiert, um die potenziell gefährdeten Gebiete besser identifizieren zu können. Mit dem Fallbeispiel wird unterstrichen, dass offene Verwaltungsdaten gut in wissenschaftlichen Fragestellungen Verwendung finden können.

Schlüsselworte: Open Data, Transparenzportal Hamburg, Kreuzkräuter, Maschinelle Lernverfahren

Abstract

The infestation potential of grassland areas in the urban area of Hamburg has been further investigated for the poisonous ragweed species *Senecio jacobaea* L. and

Senecio aquaticus Hill s. Str., which have been increasingly found in grassland for several years. A group of land parcels whose infested state was recorded by vegetation mapping served as ground truth samples. The data base was compiled from open, digital geodata taken from the Hamburg Transparency Portal (e.g. digital elevation model, official soil estimation map), which for this analysis were arranged in a joint data grid. Subsequently, the dataset was evaluated by machine learning algorithms with regard to the infestation. It could be shown that the elevation model, the soil type as well as the grassland number have a significant influence on the spread of tansy ragwort. The presence of aquatic ragwort is mainly defined by the hydrological profile type, soil composition, soil genesis and elevation model. The results of the classification were visualized in map form in order to better identify the potentially endangered areas. The case study underlines that open administrative data can be used well in scientific research.

Keywords: open data, Hamburg transparency portal, *Senecio jacobaea*, *Senecio aquaticus*, machine learning

1 Offenheit in Wirtschaft, Wissenschaft und Verwaltung

Überall entstehen heute große Datenmengen. Bundes- und Landesministerien sowie deren nachgeordnete Behörden verfügen über umfangreiche und stetig anwachsende Datenbestände. Jedoch können diese häufig nicht sinnvoll in anderen Behörden, in der Wirtschaft, der Wissenschaft und in der Zivilgesellschaft genutzt werden. Legitime Gründe hierfür können der Datenschutz oder die öffentliche Sicherheit sein. Zu oft allerdings scheitert die weitere Nutzung der Daten daran, dass diese nicht konsequent veröffentlicht werden.

Mit *Open Data* ändert sich diese Situation seit einigen Jahren. Das Thema ist innerhalb der fortschreitenden Entwicklung der deutschen und europäischen Wirtschaft hin zu einer digitalen Produktions- und Plattformökonomie von großer Bedeutung. Das Pendant zu Open Data für den öffentlichen Sektor heißt *Open Government Data*. Ziel ist es hier, Daten aus der Verwaltung auch der Öffentlichkeit zugänglich zu machen (KUNISCH 2016).

„Offene Daten sind nicht nur für einen transparenten Staat und eine effizientere Verwaltung wichtig. Sie haben auch ein bisher nicht ausgeschöpftes volkswirtschaftliches Potenzial.“ (Brigitte Zypries, Eröffnungsrede des Kongresses „Open Data Potenziale für die Wirtschaft“ 2016 im Bundesministerium für Wirtschaft und Energie, Berlin). Hierfür müssen die veröffentlichenden Stellen Open Data Strategien entwickeln. Im Optimalfall müssten die Daten in einer umfassenden Dateninfrastruktur gebündelt werden, um offene Daten für Wirtschaft und Wissenschaft noch besser nutzbar zu machen.

Schon im Dezember 2010 forderte die Dresdner Vereinbarung zum 5. Nationalen IT-Gipfel den Aufbau einer *„zentral zugänglichen, den Interessen der Nutzer an einem einheitlichen, leichten und benutzerfreundlichen Zugriff gerecht werdenden Open-Government-Plattform [...]“*. Schwerpunkte lagen dabei außerdem auf der medienbruchfreien und prozessorientierten Gestaltung des zukünftigen E-Governments, sowie die Erfüllung von fachlichen Qualitätserwartungen der Nutzer solcher Plattformen (BMWl 2010).

Ein Vorreiter war hier die Stadt Hamburg, wo die Bürgerschaft bereits im Jahr 2012 ein Transparenzgesetz verabschiedete (HMBTG 2012), und so die umfassende Veröffentlichungspflicht bislang verwaltungsinterner Informationen etablierte. Das daraus entstandene Transparenzportal (<https://transparenz.hamburg.de>) wird intensiv von der städtischen Verwaltung genutzt, um Bürgern intern erhobene Daten verfügbar zu machen.

In den letzten Jahren sind zahlreiche solcher Plattformen auf nationaler, föderaler wie auch kommunaler Ebene entstanden. GovData (<https://www.govdata.de>) ist das offizielle, deutsche Metadatenportal für Bund, Länder und Gemeinden. Hier beschäftigt man sich zudem auch mit der Entwicklung des Metadatenmodells OGD 2.0, als Standard für Deutschland, um künftig einfacher und umfassender Daten austauschen zu können. Ein weiteres Vorhaben von Bund, Ländern und Kommunen ist die Geodateninfrastruktur Deutschland (GDI-DE, <http://www.geoportal.de>), welche Teil der europäischen Geodateninfrastruktur (INSPIRE) ist. Die Plattform soll Geodaten vernetzen und Barrieren bei deren Nutzung abbauen. Entscheidungsträgern wird die Möglichkeit gegeben, sich aktuell und präzise zu informieren. Ein weiteres, länderübergreifendes Datenportal ist der Metadaten Verbund (<https://metaver.de>), der zentrale Zugangspunkt zu den Metadaten von Brandenburg, Bremen, Hamburg, Mecklenburg-Vorpommern und Sachsen-Anhalt. Als kommunales Beispiel sei das Portal OpenData.HRO (<https://www.opendata-hro.de>) genannt, welches als zentrale Plattform zur Veröffentlichung offener Daten der Hanse- und Universitätsstadt Rostock dient.

Inmitten all dieser Datenmengen finden sich unter anderem auch viele raumbezogene Geofachdaten. Hier bilden sich Einstiegsportale, welche die verfügbaren Geodaten beziehungsweise deren Metadaten bündeln, um die Vielzahl der Angebote an offenen Datenportalen sowie auch Geodateninfrastrukturen auf einen Blick erfassen zu können. Eines dieser Angebote ist die im Rahmen des BMVI-Projekts OpenGeoEdu entstandene Seite zu offenen Datenportalen im deutschsprachigen Bereich (<https://portal.opengeoedu.de>). Das Projekt will die breitere Nutzung und intelligentere Vernetzung offener Daten durch die Entwicklung innovativer und fortschrittlicher (Web-) Anwendungen vorantreiben (BILL et al. 2018).

2 Das Hamburger Transparenzportal

Durch das Hamburgische Transparenzgesetz (HMBTG 2012) erklärte Hamburg im Oktober 2012, als erstes Bundesland in Deutschland, einen Großteil seiner Verwaltungsdaten generell zu Open Data. Daraufhin wurde ein umfassendes Transparenzportal aufgesetzt (<https://transparenz.hamburg.de/>), welches der Öffentlichkeit ein breites Spektrum an Daten und Dokumenten unmittelbar über das Internet zugänglich macht. Das Portal ist seit Oktober 2014 online und wird von der fachlichen Leitstelle Transparenzportal betreut.

Mehr als 70 Behörden sowie natürliche oder juristische Personen des Privatrechts veröffentlichen auf dieser Plattform eine Vielzahl an Dokumententypen. Im November 2017 befanden sich über 70 000 Datensätze im Angebot. Diese sind dort anonym und

gebührenfrei abrufbar. Sie können entweder geordnet nach Kategorien (Gesundheit, Politik, Verkehr, etc.) oder Informationsgegenstand (Senatsbeschlüsse, Verwaltungspläne, Geodaten, etc.) durchsucht werden. Eine Volltextsuche oder die Suche nach bestimmten Dateitypen (PDF, XLSX, GML, SHP, WxS, etc.) wird ebenfalls ermöglicht.

Die Geodaten, welche sich durch ihren Raumbezug auszeichnen, machen einen wesentlichen Anteil des Angebotes aus. Zu finden sind Geobasisdaten, wie das Amtliche Liegenschaftskataster-Informationssystem (ALKIS), das Amtliche Topographisch-Kartographische Informationssystem (ATKIS) mit dem digitalen Basis-Landschaftsmodell (DLM), mehreren digitalen Höhenmodellen (DGM 1 bis 25), mehreren DTK-Skalen und dem digitalen Orthophoto (DOP 20). Daneben finden sich viele Geofachdaten zu verschiedenen Themengebieten, wie etwa Hydrologie, Umwelt und Landwirtschaft. Die Geodaten lassen sich durch interaktive Web-Karten darstellen. Auf diese Weise werden die Daten dem Nutzer auf verschiedensten, internetfähigen Endgeräten zugänglich gemacht (z. B. über ein Smartphone).

Für die Forschung können vor allem die zum Download angebotenen Geodaten von hohem Wert sein. Dabei handelt es sich um maschinenlesbare Rohdaten in gängigen Datenformaten wie TIF, CSV oder GML (Raster- und Vektordaten). Die meisten Daten lassen sich auch über OGC-konforme Web-Services (WMS und WFS) abrufen (Abb. 1). Diese lassen sich in Geo-Informationssysteme (GIS) importieren und in bestehende Auswerteworkflows integrieren. So können sie zu neuen Erkenntnissen führen oder die Grundlage für ganz neue Betrachtungen bilden.

Die Lizenzierung der Daten des Hamburger Transparenzportals wird fast ausschließlich mit der *Datenlizenz Deutschland – Namensnennung – Version 2.0* realisiert. Diese erlaubt es, Daten für kommerzielle und nicht kommerzielle Zwecke zu vervielfältigen, zu drucken, zu präsentieren, zu verarbeiten, zu bearbeiten sowie an Dritte zu übermitteln. Die Lizenz fordert einen Quellenvermerk. Dieser umfasst zum einen die Bezeichnung des Bereitstellers der Daten (nach dessen Maßgabe) sowie einen Hinweis auf die Lizenz selbst, inklusive eines Verweises zu den Lizenzbedingungen in Form der URI (GOVDATA 2017). Letzteres erfolgt häufig mit der Abkürzung *dl-de/by-2-0*, welche entsprechend verlinkt wird. Bei Medienformaten, die keine anklickbaren Links zulassen, muss die URI zusätzlich angegeben werden.

Das allgemeine Interesse an räumlichen Daten auf der Plattform ist hoch. Der Begriff „Geodaten“ lag im Februar 2017 in den Top 10 der häufigsten Suchbegriffe des Transparenzportals Hamburg (HERR et al. 2017).

Die Daten des Hamburger Transparenzportals können potenziell einen wichtigen Bestandteil für verschiedenste Analysen in Forschungsarbeiten bilden oder diese un-

Digitale Orthophotos 20cm (belaubt) Hamburg

Veröffentlichende Stelle: Landesbetrieb Geoinformation und Vermessung
Veröffentlichungsdatum: 18.04.2017



Digitale Orthophotos: Für die Herstellung von digitalen Orthophotos (DOP) und der daraus abgeleiteten Luftbildkarte 1:5000 erfolgt im 2. Jahrgang

Ressourcen





	1. Name: Dienst "WMS Digitale Orthophotos 20cm belaubt Hamburg" (GetCapabilities) Format: WMS (Größe: 0 KB)
	2.
	3. Name: Download DOP 20 , Sommer 2016, (JPG) Format: JPG (Größe: 4651.5 MB)
	Name: URL zum Geoportal Hamburg Format: HTML (Größe: 4.0 KB)

Abbildung 1: Beispiel für einen Geodatensatz auf dem Transparenzportal Hamburg. Die digitalen Orthophotos lassen sich entweder über OGC-Web-Services abrufen (1) oder direkt im Browser anzeigen (2). Außerdem können die Rohdaten, hier im JPG-Format, heruntergeladen werden (3).

terstützen. Darüber hinaus lassen sie sich zur Visualisierung der Forschungsergebnisse auf Basis von Hintergrundkarten (Verwaltungsgrenzen, topographische Karten) nutzen. Dies ist auch Gegenstand dieses Beitrags, in dem offene Geodaten genutzt wurden, um das Befalls-Potenzial von Schadpflanzen in Hamburg zu untersuchen.

3 Potenzialflächenanalyse für Kreuzkräuter im Grünland

Grünland stellt heute schon den größten Anteil an der landwirtschaftlichen Fläche Hamburgs dar. Dieser wird in Zukunft durch politische Vorgaben und die spezielle Standortbeschaffenheit Hamburgs weiter anwachsen, was einer nachhaltigen Bewirtschaftung von Grünland zukünftig eine noch höhere Bedeutung beimisst. Grünland befindet sich in Hamburg auch im Besitz und in der Bewirtschaftung der öffentlichen Hand. Zudem sind bisher fast alle Ausgleichsflächen der Freien und Hansestadt Hamburg (FHH) Grünlandflächen.

Verschiedene Vertreter aus der Pflanzengattung der Kreuzkräuter treten seit einigen Jahren vermehrt im Grünland auf. Durch ihre Toxizität stellen die Kreuzkräuter neue Herausforderungen an die extensive Grünlandnutzung. Die Pflanzen enthalten giftige Pyrrolizidinalkaloide (PA), auf die besonders Pferde und Rinder empfindlich reagieren. Auch für den Menschen stellen PA eine Gefahr dar (PRAKASH et al. 1999).

Die Pflanzen verhindern somit eine optimale Nutzung der befallenen Flächen, was bei der Tierhaltung und im Futtermittelbau zu erheblichen wirtschaftlichen Verlusten führen kann (PANTER et al. 2007).

Bei den Arten Jakobs-Kreuzkraut (*Senecio jacobaea* L.) und Wasser-Kreuzkraut (*Senecio aquaticus* Hill s. str.) handelt es sich um krautige Pflanzen, die Wuchshöhen bis zu einem Meter erreichen können. Sie sind zweijährig, können aber ausdauernd werden, wenn sie in ihrer Entwicklung, zum Beispiel durch Schnitt, gestört werden. Junge Pflanzen bilden in der vegetativen Phase eine grundständige Blattrosette. Erst im zweiten Jahr kommt es zum Schossen und im Hochsommer bildet sich ein gelber, schirmrispiger Blütenstand. Es handelt sich hierbei nicht um Neophyten, sondern um heimische Arten, deren Auftreten nicht grundsätzlich neu ist. Dennoch verhält sich die Pflanze invasiv.

Aus den Beobachtungen der letzten Jahre, speziell auch in Hamburg (BRANDT 2012, STROTDREES 2012, STROTDREES 2015), wird ersichtlich, unter welchen Bedingungen die Pflanzen sich bevorzugt ausbreiten. Das Jakobskreuzkraut besiedelt bevorzugt offene Böden mit geringem Konkurrenzbewuchs. Verbreitet ist es besonders auf Stilllegungsflächen, Straßen- und Wegrändern, Böschungen, aber auch auf extensiv genutzten Weiden, insbesondere Pferdeweiden und Extensivgrünland. Auf konventionell und damit vergleichsweise intensiv bewirtschaftetem Grünland ist das Ansiedlungspotenzial gering. Jedoch ist das Vorkommen auch hier in den letzten Jahren gestiegen (SUTER et al. 2007). Es wächst vor allem an Kahlstellen in der Narbe, wie etwa Trittschäden und Schäden durch Befahren. Warme und sonnige Standorte mit mäßigem Nährstoffgehalt werden bevorzugt. Während das Jakobs-Kreuzkraut eher die trockenen Standorte mit einem ausgeprägten Sandanteil besiedelt, ist das Wasser-Kreuzkraut ein ausgeprägter Feuchte- bis Nässezeiger. Es tritt daher typischerweise auf staunassen Wiesen und entlang von Entwässerungsgräben auf.

Für Hamburg ließe sich daher vermuten, dass das Jakobs-Kreuzkraut tendenziell eher auf den höher gelegenen, trockenen Geesten zu finden sein wird. Dem gegenüber steht das Wasser-Kreuzkraut, welches sich bevorzugt in den tiefer gelegenen, feuchten Marschen ansiedeln könnte. Während das Wasserkreuzkraut tatsächlich relativ selten auf der Geest wächst, steht jedoch das Jakobs-Kreuzkraut häufig auch in den Marschenlanden (STROTDREES 2015). Dies zeigt, dass die Verbreitung der beiden Pflanzen höchstwahrscheinlich von weiteren Faktoren abhängt, die es weiter zu untersuchen gilt. Vegetationskundliche Kartierungen von Schadpflanzen sind sehr aufwändig und können immer nur einen geringen Teil des Grünlandes beobachten. Es wurde daher nach einem Weg gesucht, schnell einen Überblick der befallsgefährdeten Bereiche des Stadtgebietes zu erhalten. In der vorliegenden Arbeit wurden offene Geodaten des Hamburger Transparenzportals verwendet, um das Verbreitungspoten-

zial der zwei Kreuzkrautarten und die darauf Einfluss nehmenden Faktoren näher zu untersuchen.

3.1 Forschungsdesign

Ziel war das Erkennen von Dauergrünlandflächen im Stadtgebiet Hamburgs, welche potentiell mit Jakobs-Kreuzkraut oder Wasser-Kreuzkraut befallen sein könnten. Es soll auch erörtert werden, ob Geodaten mit hydrologischem Hintergrund geeignet sind, um das Befallspotenzial von Wasser-Kreuzkraut zu modellieren, da dieses feuchtere Standorte bevorzugt.

Es wurde zunächst eine Stichprobe befallener Grünlandflächen ausgewählt. In Verbindung mit offenen, räumlich verzeichneten Geofachdaten des Hamburger Transparenzportals dienen diese als Grundlage für eine überwachte Klassifikation mit Hilfe von Machine Learning Algorithmen. Die Daten wurden im Shape- und Rasterformat, wie auch als Webdienste (WMS, WFS) in verschiedene GIS eingebunden und verarbeitet. Auf diese Weise wurde ermittelt, welche Flächeninformationen besonders stark auf Vorhandensein von Kreuzkrautarten hinweisen. Aus den so gewonnenen Informationen werden anschließend die Grünlandflächen des gesamten Stadtgebietes hinsichtlich ihrer Befallswahrscheinlichkeit klassifiziert. Die Ergebnisse werden dann in Kartenform visualisiert.

3.2 Aufarbeitung der Stichprobe-Flächen

Als Stichprobe wurden verschiedene pflanzenkundliche Kartierungen herangezogen, in denen für 73 Flurstücke im Stadtgebiet Hamburg das Vorkommen von Kreuzkrautarten untersucht wurde (BRANDT 2012, STROTDREES 2012, STROTDREES 2015). Ergänzt werden diese Daten durch Flächenbegehungen, die im Oktober 2016 stattfanden. Die Kartierungsarbeiten referenzieren begutachtete Flurstücke mit deren Gemarkung und Flurstücksnummer. Anhand dieser wurden die verschiedenen Datentabellen vereinigt. Da nicht in jedem Fall die gesamte Fläche bonitiert wurde, liegen nur sporadisch genaue Befallszahlen vor. Die Analyse betrachtet daher eine Fläche als befallen, wenn mindestens ein bestätigtes Auftreten von Jakobs- oder Wasser-Kreuzkraut im Zeitraum der Kartierungen vorliegt. Ist dies nicht der Fall, gilt die Fläche als nicht befallen. Für die spätere Analyse ergeben sich so die zwei dichotomen, abhängigen Variablen „Befall mit Jakobs-Kreuzkraut“ und „Befall mit Wasser-Kreuzkraut“.

Um den Bezug zum Flurstücksbestand herzustellen, wurden alle Flächen, für die Stichprobendaten vorliegen, im Polygonformat digitalisiert. Als Grundlage dienen

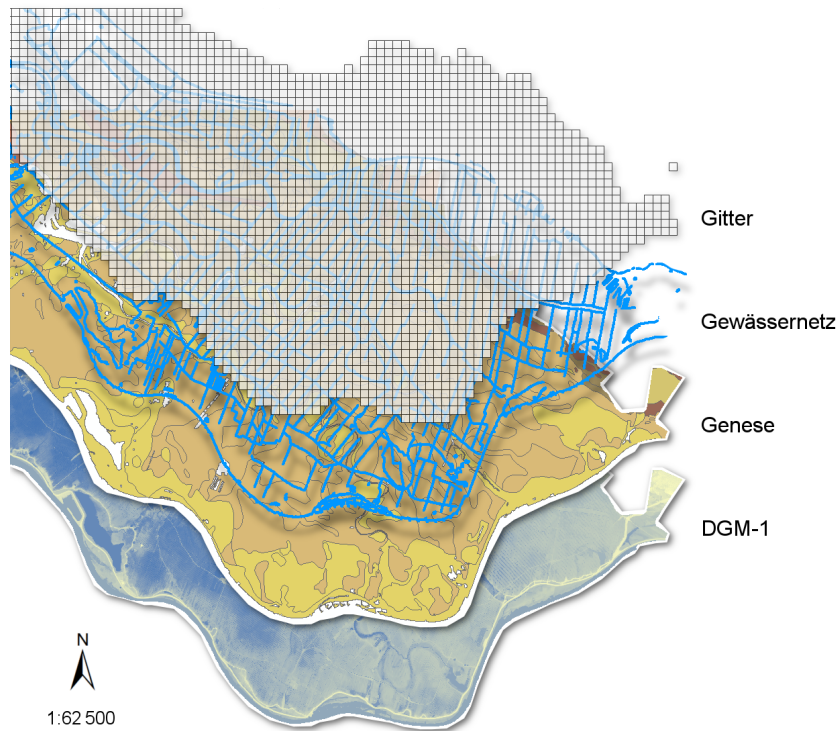


Abbildung 2: Beispiele für verwendete Geodaten der Stadt Hamburg im Raster-, Polygon- und Linienformat. Das Gitter (200x200 Meter Zellgröße) dient der einheitlichen Verarbeitung und Auswertung der Geodaten.

die ALKIS-Flurstücksumringe (LGV HAMBURG 2017), welche über einen WMS-Dienst eingebunden wurden. Die dazugehörige Feature-Tabelle fasst die vorliegenden Daten aller Stichprobe-Flächen zusammen. Außerdem erhält hier jedes Flurstück eine eindeutige Objekt-ID, welche für die späteren Prozesse wichtig ist.

3.3 Aufarbeitung der Geofachdaten für die Klassifizierung

Für die Analyse wurde das Hamburger Transparenzportal nach geeigneten Geofachdaten durchsucht, welche die geplanten Untersuchungen sinnvoll unterstützen können (Tab. 1). Dabei wurde auch das vorhandene Fachwissen bezüglich der Wachstumsbedingungen der beiden Pflanzenarten genutzt, um aus den vielen vorhandenen Datensätzen eine Auswahl zu treffen. Die Daten lassen sich grob in die Kategorien Landwirtschaft, Hydrologie, Boden und Gelände gliedern. Dabei handelt es sich beispielsweise um die Amtliche Bodenschätzungskarte, das Digitale Geländemodell, die Geologische Karte oder das Gewässernetz (Abb. 2). Erforderlich war außerdem, dass die Daten flächendeckend für das gesamte Stadtgebiet vorhanden sind. Damit fallen beispielsweise bodenphysikalische Untersuchungen heraus, da diese nur punktuell stattfinden.

Tabelle 1: Tabelle der Geofachdaten des Transparenzportals Hamburg, welche für die Untersuchungen herangezogen wurden. Geofachdaten, die letztendlich in die Analyse einbezogen wurden, sind fett dargestellt.

Datensatz	Art*	Format	enthaltene Geofachdaten
Amtliche Bodenschätzungskarte	P	GML	Kultur, Bodenart, Zustandsstufe, Wasserverhältnisse, Bodenzahl, Grünlandzahl
ATKIS Digitales Basis-Landschaftsmodell	P	GML	Landwirtschaft (Vegetationsmerkmal)
Ausgleichsflächen gemäß BNatSchG	P	GML	Ausgleichsfläche
Biotopkataster	P	GML	Biotoptyp, Gesamtbewertung
Bodenformengesellschaften	P	GML	Bodengesellschaft, Bodensubstrat
Bodenversiegelung	P	GML	Biotopgruppe , Versiegelungsgrad
Digitales Höhenmodell DGM-1	R	ASCII	Geländehöhe
Fachplan Schutzwürdige Böden	P	GML	Bodenschutzklasse
Flächen des Arten- und Biotopschutzes	P	GML	Flächengruppe
Flächennutzungsplan	P	GML	Nutzung
Geologische Karte 1:50000	P	GML	Genese , Geologisches Kürzel, Kaltzeit , Petrographie
Hydrogeologie	L	GML	Grundwassergleichen, hydrologischer Profiltyp
Landschaftsprogramm	L, P	GML	Gewässernetz , Nutzung, Trasse
Naturräume Geest und Marsch	P	GML	Naturraum
Verdunstungspotenzial	P	GML	Verdunstungsklasse
Versickerungspotenzial	P	GML	Versickerungstiefe

* R=Raster, P=Polygon, L=Linie

Nach dem Download liegen Vektordaten zumeist im GML-Format und Rasterdaten im ASCII-Format vor. Um eine effizientere Verarbeitung zu ermöglichen, wurden Vektordaten in Shapedateien und Raster in GeoTIFF-Dateien umgewandelt. Datensätze, die nach dem Importieren in das GIS in mehreren Teil-Layern vorlagen (bspw. DGM-1), wurden zu einem Layer zusammengefügt. Ein Großteil der Attribute (Spalten) in den Datentabellen sind für die Analyse nicht von Bedeutung. Dazu zählen etwa verwaltungsinterne Nummerierungen oder Kommentarspalten der Bearbeiter. Diese wurden vor der statistischen Auswertung gelöscht.

3.4 Geodatenverarbeitung

Um das Stadtgebiet komplett und einheitlich auszuwerten, wurde ein spezielles Python-Skript entwickelt. Es verwendet das ESRI Erweiterungspaket ArcPy für Python, um die Geodatenverarbeitungs-Werkzeuge von ArcGIS anzusteuern. Das Skript automatisiert den Prozess der Einordnung der Geofachdaten in ein regelmäßiges Gitter. Weiterhin ist das Skript in der Lage, nicht nur Geodaten im Polygonformat zu verarbeiten, sondern auch Raster- und Liniendaten (Abb. 3). Dadurch kann ein breites Spektrum an offenen Geodaten in die Analyse einbezogen werden.

Das Skript erstellt ein Gitter über der landwirtschaftlich genutzten Fläche im Untersuchungsgebiet (Abb. 2). Die einzelnen quadratischen Gitterzellen besitzen ei-

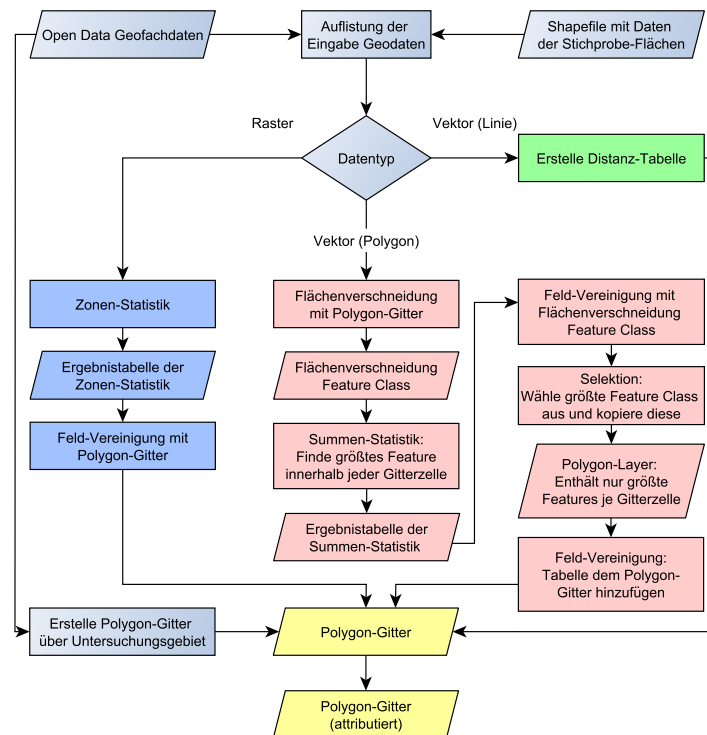


Abbildung 3: Modell zur Verarbeitung von verschiedenen Geodaten (Raster-, Vektordaten) zur Gewinnung der Basisdaten für eine statistische Flächenanalyse. Endergebnis ist das attribuierte Polygongitter, dass die Daten aggregiert.

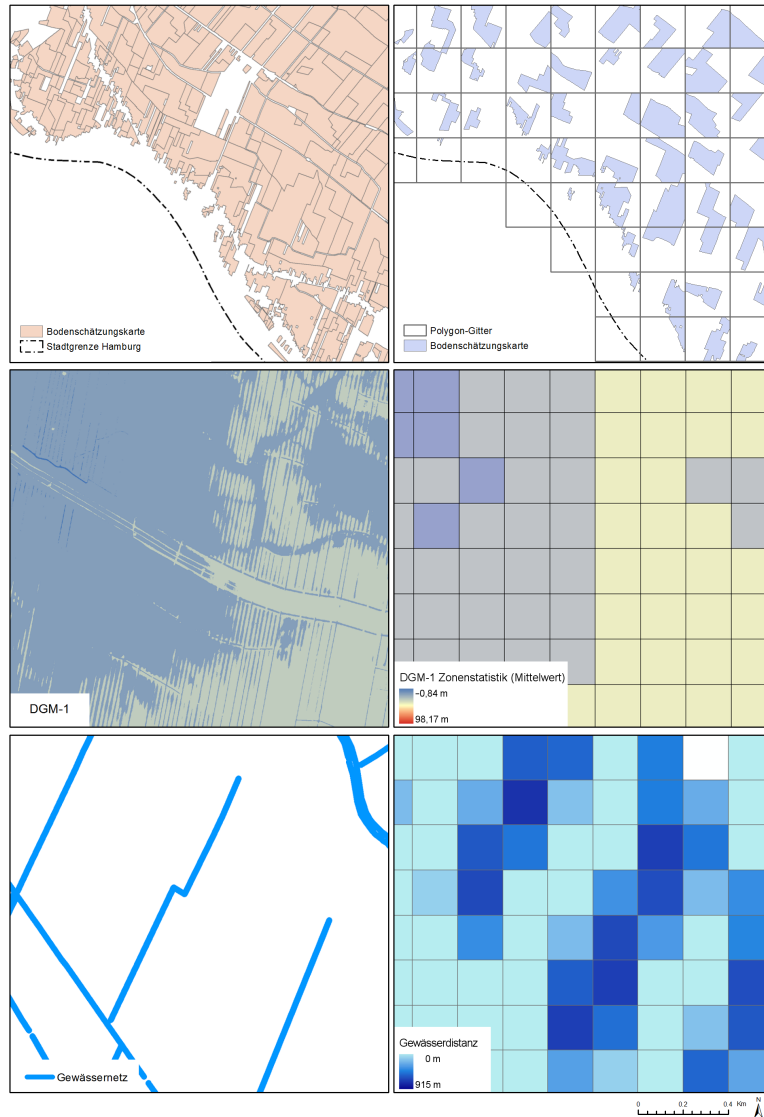


Abbildung 4: Beispiele für die Verarbeitung verschiedener Geofachdaten. Oben Polygone, in der Mitte Raster und unten Liniendaten.

ne Kantenlänge von 200 Metern. Die Gesamtmenge der Gitterzellen beträgt somit 7190. Die Länge wurde gewählt, um die Redundanz durch zu starke Unterteilung der Geofachdaten gering zu halten, jedoch relativ kleinräumig vorhandene Bereiche noch abtasten zu können. In den meisten Geofachdaten ist der durchschnittliche Durchmesser der Vektorobjekte erheblich größer als die gewählte Kantenlänge. Die Eingabedaten werden danach je Datentyp unterschiedlich weiter verarbeitet.

Bei Polygon-Daten wird eine Flächenverschneidung zwischen dem Gitter und jedem Eingabe-Polygonlayer vorgenommen. Anschließend wird pro Gitterzelle das Polygon mit dem größten Flächenanteil ausgewählt und dessen Attribute in die Datentabelle des Gitters kopiert (Abb. 4, oben). Später werden diese Attribute die gesamte Gitterzelle repräsentieren, da die Attribute der kleineren Polygone an dieser Stelle verworfen werden. Die Rasterdaten des DGM-1 werden über eine Zonen-Statistik ausgewertet. Die Werte werden je Gitterzelle subsumiert, der Mittelwert und die Standardabweichung berechnet und die Ergebnisse dem Gitter beigelegt (Abb. 4, mittig). Um auch Liniendaten, wie etwa die Gewässerlinien des Landschaftsprogramms, auswerten zu können, wird für jede Gitterzelle der Abstand zum nächstgelegenen Linienobjekt berechnet. Dieser Wert wird dann in das Gitter übernommen (Abb. 4, unten). Es wird angenommen, dass die Befalls-Wahrscheinlichkeit unter anderem von der Entfernung jeder Gitterzelle zu bestimmten Landschaftselementen (z.B. Fließgewässern) abhängt.

Außer den offenen Geofachdaten des Transparenzportals befindet sich außerdem der Polygonlayer mit den bereits bekannten Befallsdaten unter den Eingabedaten des Skriptes. Auf diese Weise werden auch sie in das Gitter übernommen und ermöglichen so die statistische Analyse auf Grundlage der Gitterzellen. Nach einer Verschneidung mit dem Gitter ergibt sich eine Stichprobengröße von 144 Gitterzellen, die aus 73 Flurstücken hervorgegangen sind. Die Stichprobe wächst leicht, da sich die Flurstücke über mehrere Gitterzellen erstrecken können. Da sich die relevanten Daten nun vollständig im Gitter befinden, werden die ursprünglichen Geofachdaten nicht mehr benötigt. Die Attributtabelle des Gitters wurde nun weiter ausgewertet.

3.5 Statistische Analyse

Viele der Informationen in den Datensätzen liegen als alphanumerische, kategoriale Variablen vor. Für die statistische Auswertung wurden diese numerisch kodiert. Einige der kategorialen Variablen besaßen sehr viele mögliche Ausprägungen. Da ein Großteil davon im Bezug auf die abhängigen Variablen unterbelegt waren, wurde eine

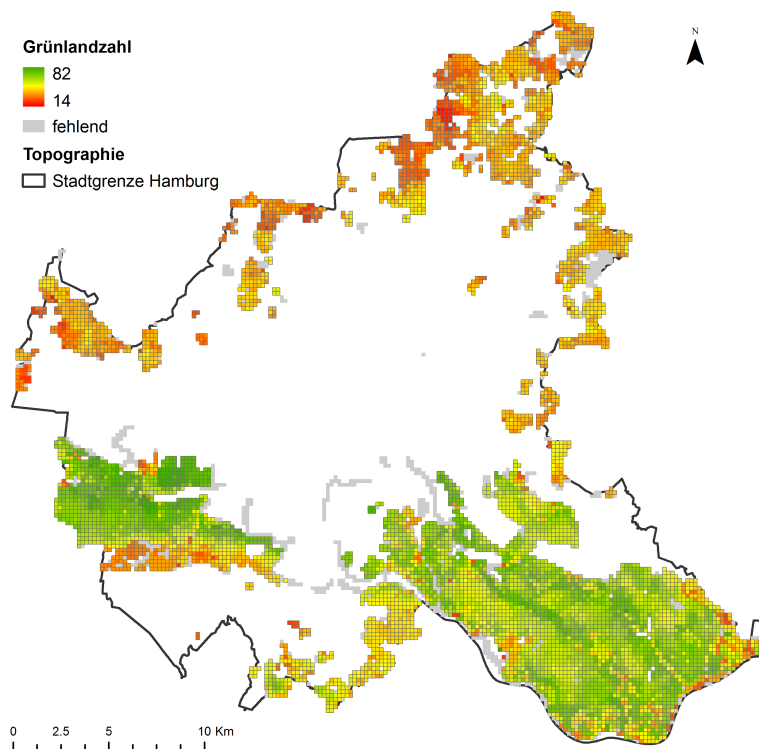


Abbildung 5: Die Grünlandzahlen aus der Amtlichen Bodenschätzungskarte nach der Verarbeitung durch das Skript.

Zusammenlegung von Ausprägungen erforderlich. Um den Einfluss auf die Ergebnisse zu minimieren, wurden die Ausprägungen dabei thematisch gruppiert. Bei der Variable Biotoptyp, welche besonders viele Ausprägungen besitzt, entstanden aus den ursprünglichen 426 Ausprägungen drei neue, welche für „trockenes Grünland“, „feuchtes Grünland“ sowie „alle weiteren Biotoptypen“ stehen. Häufig enthielten die Geofachdaten Variablen, welche untereinander stark korreliert sind und daher thematisch nahezu identische Aussagekraft besitzen (z. B. Petrographie und Genese). Diese wurden über eine Korrelationstabelle ermittelt und gelöscht. Variablen, bei denen sehr viele Werte fehlten, mussten ebenfalls von der Analyse ausgeschlossen werden (z. B. Bodenschutzklasse und Zustandsstufe). Tabelle 1 zeigt die letztlich als Modellvariablen genutzten Daten in fester Schriftstärke.

Die zuvor gesammelten Daten wurden nun mit den abhängigen Variablen „Befall mit Jakobs-Kreuzkraut“ und „Befall mit Wasser-Kreuzkraut“ durch unterschiedliche Verfahren analysiert. Es wurde das Methodenspektrum der R-Erweiterung Rattle verwendet. Dabei handelt es sich um Random Forest (BREIMAN 2001), Decision Tree (BREIMAN et al. 1998), Support Vector Machine (CORTES et al. 1995), sowie Neuronales Netz und die Logit-Regression, welche unter anderem in KOTSIANTIS (2007) beschrieben sind. Dies sind vorrangig Modelle der prädiktiven, multivariaten

Statistik, welche im Umfeld des Machine Learning sowie der Fernerkundung häufig zur Anwendung kommen (LARY et al. 2016). Der Begriff des überwachten Lernens spielt dabei eine zentrale Rolle. Ziel ist es, ein einfaches aber präzises Modell der Verteilung von Klassenbezeichnungen in Form von Prädiktormerkmalen zu erstellen. Der resultierende Klassifikator wird dann verwendet, um den Testfällen, deren Klassenbezeichnung unbekannt ist, eine solche zuzuordnen (KOTSIANTIS 2007).

Um die Modelle später testen zu können, wurden die Stichprobenfälle jeweils in eine zufällige Trainings- und Teststichprobe aufgeteilt. Dabei wurde das Verhältnis 7:3 verwendet, um ein robusteres Training zu erreichen. Die Modelle wurden jeweils mit Jakobs-Kreuzkraut und Wasser-Kreuzkraut als Zielvariable berechnet. Dabei schwankt die Größe der tatsächlich je Modell verarbeiteten Fälle leicht, da die Verfahren teils in der Lage sind, auch Fälle mit fehlenden Werten zu verarbeiten.

Das Random Forest Modell wurde zunächst mit dem Standardwert von 500 Bäumen berechnet. Die Anzahl wurde dann schrittweise bis auf 800 Bäume erhöht, bis keine Verbesserung des Trainingsfehlers mehr eintrat. Die einzelnen Bäume bestehen je aus zufälligen Teilstichproben der Trainingsstichprobe. Die Anzahl der für die Entscheidung an den Pfad-Verzweigungen berücksichtigten Variablen wurde iterativ, durch Optimierung des Out-of-Bag Fehlers (OOB), ermittelt. Der OOB ist der mittlere Vorhersagefehler bei jedem Trainingsfall x_i , wobei zur Berechnung jeweils nur die Bäume verwendet werden, deren Teilstichprobe x_i nicht enthält. Bei einer Anzahl von sieben Variablen ergab sich der geringste OOB von 12,9%. Die Auswahl dieser sieben Variablen erfolgte bei jedem Baum zufällig. Die Mean Decrease Accuracy gibt an, wie stark die Anpassung des Modells durch das Herauslassen der Variable sinken würde. Treten hier negative Werte auf, sollte das Modell durch Herauslassen dieser Variablen also eine Verbesserung erfahren. Dies wurde iterativ durchgeführt, wodurch die Variablen Biotopgruppe, Biotoptyp, Gewässerdistanz, Grundwassergleichen und Versickerungstiefe bei Jakobs-Kreuzkraut, sowie Biotoptyp, Gesamtbewertung und Grundwassergleichen bei Wasser-Kreuzkraut zusätzlich entfielen. Das so errechnete Modell wurde dann zur Klassifikation aller Fälle des Gitters verwendet. Schließlich wurden die Ergebnisse in Kartenform visualisiert.

3.6 Ergebnisse und Diskussion

Zum Vergleich der Verfahren wird die Korrektklassifikationsrate herangezogen. Sie setzt alle korrekt zugeordneten Fälle mit der Gesamtzahl der Testfälle ins Verhältnis. Für die Analyse des Jakobs-Kreuzkrautes lieferte das Random Forest Modell das beste Ergebnis. Es konnte im Test etwa 98% der Fälle korrekt klassifizieren (Tab. 2).

Bei Wasser-Kreuzkraut konnte das Random Forest Modell in rund 93 % der Testfälle den Befallszustand korrekt zuordnen.

Die anderen Modelle waren deutlich unterlegen, was sich besonders an den höheren Falsch-Negativ-Raten bemerkbar macht, also der fälschlicherweise als befallsfrei klassifizierten Fälle. An der Falsch-Positiv-Rate hingegen zeigt sich, dass gerade bei Jakobs-Kreuzkraut von einer Art Worst-Case Szenario auszugehen ist. Hier ergab der Test, dass zu viele der Fälle als befallsgefährdet eingestuft wurden. Dagegen zeigen die Ergebnisse der Analyse des Wasser-Kreuzkrautes hohe Falsch-Negativ-Raten. Hier werden also eher tatsächlich befallene Zellen als nicht befallen eingestuft. Eine Ursache dafür könnte sein, dass die Stichprobe des Wasser-Kreuzkrautes, im Vergleich zum Jakobs-Kreuzkraut, verhältnismäßig wenig positive Fälle enthält. Der Decision-Tree fiel in beiden Fällen nicht plausibel aus, da für die Klassifikation fast ausschließlich das DGM und die Grünlandzahl herangezogen wurden.

Tabelle 2: Test-Ergebnisse der Klassifikation durch die verschiedenen Verfahren.

Modell	Korrektklassifikationsrate [%]	Falsch-Positiv-Rate [%]	Falsch-Negativ-Rate [%]
Jakobs-Kreuzkraut			
Random Forest	97,6	8,0	0,0
SVM	90,5	33,3	3,0
Neuronales Netz	85,7	44,4	6,1
Decision-Tree	81,0	11,1	21,2
Lineares Modell	71,4	44,4	24,2
Wasser-Kreuzkraut			
Random Forest	92,7	0,0	37,5
Neuronales Netz	90,2	6,5	20,0
Decision-Tree	83,3	0,0	70,0
SVM	82,9	0,0	70,0
Lineares Modell	78,0	6,5	70,0

Die Wichtigkeit der Variablen kann für beide Betrachtungen mit Hilfe des Random Forest Modells ermittelt werden. Die Variablen lassen sich lose in die drei Gruppen Boden, Pflanze und Wasser einteilen (Abb. 6). Das Höhenmodell, die Bodenart, sowie die Grünlandzahl stellten sich als wichtigste Unterscheidungskriterien für das Vorkommen von Jakobs-Kreuzkraut heraus. Das Vorhandensein von Wasser-Kreuzkraut wird vor allem durch den hydrologischen Profiltyp, die Bodengesellschaft, die Genese, sowie das Höhenmodell definiert. Die eher dem Boden zugehörigen Variablen haben bei beiden Pflanzenarten eine vergleichsweise hohe Bedeutung für die Vorhersage. Die Bedeutung der Variablen der Gruppe Pflanze variiert stark. Während Grünlandzahl und Naturraum einen hohen Stellenwert einnehmen, ist die Bedeutung der Variablen Kultur, Nutzung sowie Biotopgruppe gering. Ebenso stark verteilt sind die Variablen, welche auf die hydrologischen Bedingungen hindeuten. Der hydrologische Profiltyp bildet eine wichtige, diskriminierende Größe, da hier die Wichtigkeit für beide Pflanzenarten relativ hoch ist. Die weiteren hydrologischen Werte, wie Gewässerdistanz oder Versickerungstiefe, stellten sich nur in begrenztem

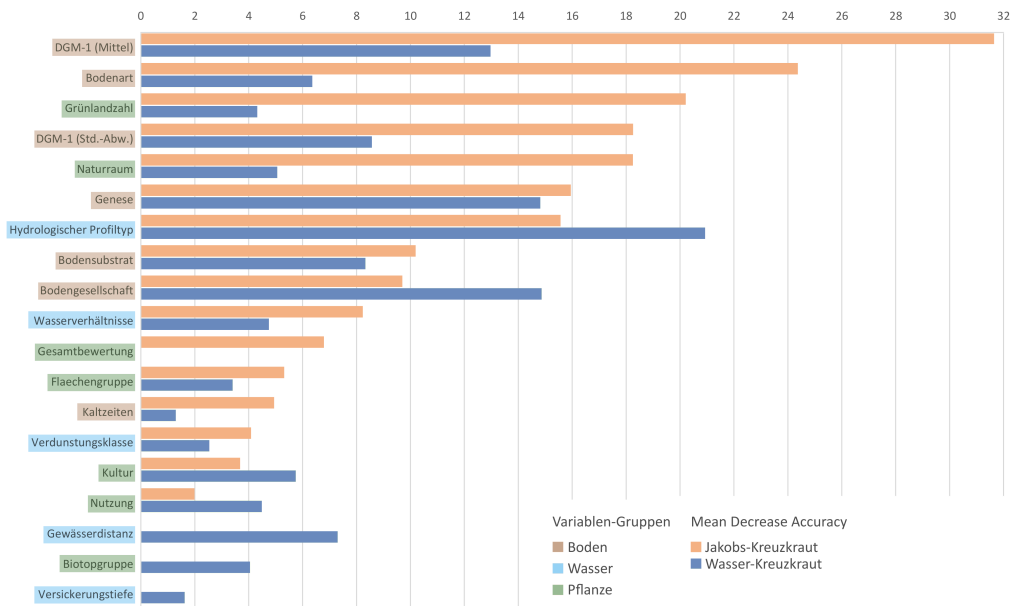


Abbildung 6: Wichtigkeiten der Variablen je Pflanzenart (Variablen vgl. Tab. 1).

Maße als relevante, diskriminierende Kriterien für das Wasser-Kreuzkraut heraus. Die Größen Biotopgruppe, Gesamtbewertung (der Biotope), Gewässerdistanz und Versickerungstiefe tragen nur zur Unterscheidung je einer Pflanzenart bei. Die Variablen Biotoptyp und Grundwassergleichen entfielen, wegen zu geringer Bedeutung für beide Spezies, komplett aus dem Modell.

Aus den errechneten Werten der besten Modelle wurden Kartenansichten erstellt, welche das Befallspotenzial im gesamten Untersuchungsraum (Stadtgebiet Hamburg) farblich anzeigen (Abb. 7 u. 8). Durch die Vielzahl der Variablen enthalten viele Gitterzellen auch fehlende Werte. Diese entstehen dadurch, dass Merkmale an der entsprechenden Stelle schlicht nicht existieren oder weil sie für bestimmte Bereiche im Rahmen des Hamburgischen Transparenzgesetzes nicht veröffentlicht werden. Daher sind auf der Karte die Bereiche, in denen keine Analyse möglich war, grau dargestellt. Die Karte für Jakobs-Kreuzkraut zeigt, dass fast das gesamte Grünland der Stadt potenziell befallgefährdet ist. Ausgenommen sind hiervon lediglich die elbferneren Bereiche in der Marsch. Die Wasser-Kreuzkraut-Karte deutet auf einen deutlich geringeren Befall hin. Es werden vor allem Bereiche nahe der Gewässerläufe und in Moorarealen als befallgefährdet ausgezeichnet. Auf der Geest sind nur vereinzelt Vorkommen zu erwarten.

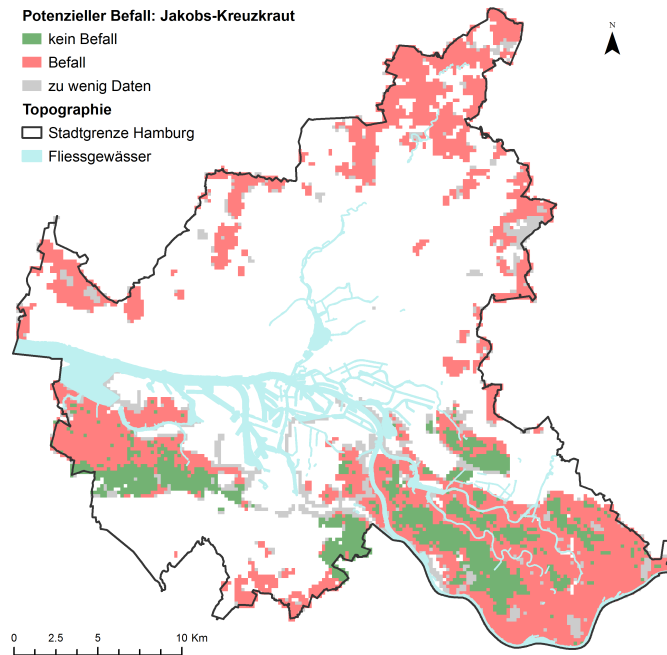


Abbildung 7: Karte des potenziellen Befalls durch Jakobs-Kreuzkraut.

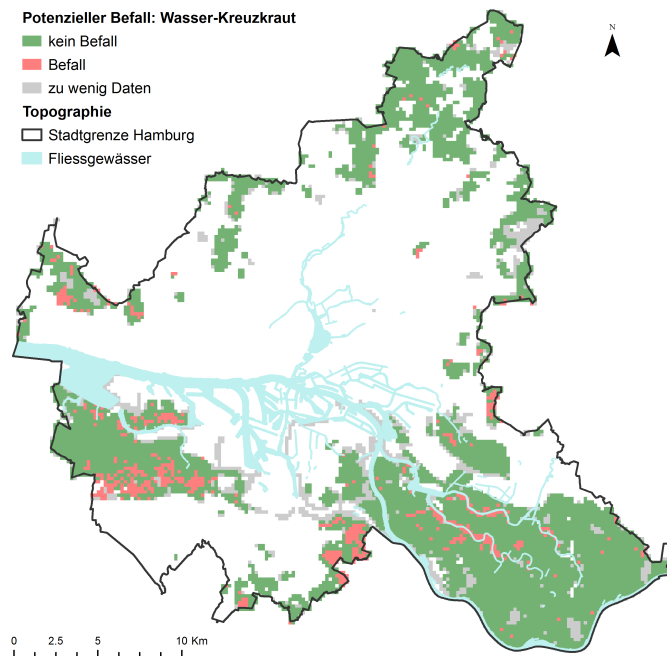


Abbildung 8: Karte des potenziellen Befalls durch Wasser-Kreuzkraut.

3.7 Ausblick

Mit der vorgestellten Fallstudie wird der Nutzen von offenen Daten für wissenschaftliche Fragestellungen belegt. Das konkrete Fallbeispiel wäre durch Hinzuziehung weiterer offener Daten erweiterbar. Auch für die Verrechnung der Daten wären andere Algorithmen und Annahmen denkbar. Im von der Freien und Hansestadt Hamburg geförderten Forschungsprojekt ist eine derartige Potenzialflächenanalyse ein erster Schritt zur Einschränkung möglicher Befallsgebiete in der Stadt. Darüber hinaus soll eine automatisierte Bestandscharakterisierung der Grünlandflächen hinsichtlich Bewuchshöhen, Weidequalität und Flächenmanagement durch raum-zeitliche GIS-Analysen und Data Mining erfolgen. Die bildbasierte Einzelpflanzenenerkennung mittels automatisierter Bildanalyse und Mustererkennung aus UAV-Daten wird am Beispiel des Jakobs-Kreuzkrautes untersucht. Mit einer App-Entwicklung für den mobilen Einsatz auf Smartphones oder Tablets soll die Unkrautbekämpfung ermöglicht werden. Dies geschieht auf ausgewählten Grünlandstandorten der Stadt Hamburg.

4 Danksagung

Der Autor bedankt sich bei der Behörde für Wirtschaft, Verkehr und Innovation der Freien und Hansestadt Hamburg für die Förderung des Forschungsprojekts „UAV-basiertes Grünlandmonitoring auf Bestands- und Einzelpflanzenebene“, in dessen Rahmen diese Untersuchung stattfinden konnte.

5 Lizenzhinweise

Die Darstellung und Bearbeitung der Abbildungen erfolgte nach der Datenlizenz „Deutschland - Namensnennung - Version 2.0“ (<https://www.govdata.de/dl-de/by-2-0>).

Die Daten wurden bereitgestellt durch die Freie und Hansestadt Hamburg, Landesbetrieb Geoinformation und Vermessung (<http://suche.transparenz.hamburg.de/dataset/digitales-hohenmodell-hamburg-dgm-11>, <http://suche.transparenz.hamburg.de/dataset/amtliche-bodenschatzungskarte19>, <http://suche.transparenz.hamburg.de/dataset/alkis-verwaltungsgrenzen-hamburg5>), sowie der Behörde für Umwelt und Energie (<http://suche.transparenz.hamburg.de/dataset/landschaftsprogramm-hamburg3>, <http://suche.transparenz.hamburg.de/dataset/geologische-karte-1-50-0004>).

Literatur

- BILL, R., LORENZEN-ZABEL, A., HINZ, M. (2018): Offene Daten für Lehre und Forschung in raumbezogenen Studiengängen – OpenGeoEdu. In: *Gis.science* (Heft 1), x–y.
- BRANDT, I. (2012): Vegetationskundlich orientierte Kartierung von Grünlandflächen in Hamburg 2012 als Grundlage zur Beurteilung von Vorkommen von Jakobskreuzkraut (*Senecio jacobaea*). Im Auftrag der Behörde für Stadtentwicklung und Umwelt, Amt für Natur- und Ressourcenschutz, Hamburg.
- BREIMAN, L. (2001): Random Forests. In: *Machine Learning* 45 (1), S. 5–32.
- BREIMAN, L., FRIEDMAN, J., STONE, C. J., OLSHEN, R. A. (1998): Classification and regression trees. Reprint. The Wadsworth statistics probability series. Boca Raton, FLA: Chapman & Hall CRC. 358 S.
- BUNDESMINISTERIUM FÜR WIRTSCHAFT UND TECHNOLOGIE, BMWI (2010): Dresdener Vereinbarung. Fünfter Nationaler IT-Gipfel Dresden.
- CORTES, C., VAPNIK, V. (1995): Support-vector networks. In: *Machine Learning* 20 (3), S. 273–297.
- GOVDATA (2017): Datenlizenz Deutschland – Namensnennung – Version 2.0. URL: <https://www.govdata.de/dl-de/by-2-0> (abgerufen am 17.05.2017).
- HAMBURGISCHES TRANSPARENZGESETZ (HMBTG) (2012): Gesetzestext. URL: <http://transparenz.hamburg.de/das-hmbtg> (abgerufen am 20.09.2017).
- HERR, M., MÜLLER, C., ENGEWALD, B., PIESKER, A., ZIEKOW, J. (2017): Abschlussbericht zur Evaluation des Hamburgischen Transparenzgesetzes. Institut für Gesetzesfolgenabschätzung und Evaluation. Speyer.
- KOTSIANTIS, S. B. (2007): Supervised Machine Learning: A Review of Classification Techniques. In: *Informatica* (31), S. 249–268.
- KUNISCH, M. (2016): Big Data in der Landwirtschaft – Perspektiven eines Datendienstleisters. In: *Landtechnik* 2016 (71(1)), S. 1–3.
- LANDESBETRIEB GEOINFORMATION UND VERMESSUNG HAMBURG (2017): ALKIS Liegenschaftskarte. URL: <http://metaver.de/trefferanzeige?docuuiid=F691CFB0-D38F-4308-B12F-1671166FF181> (abgerufen am 02.05.2017).
- LARY, D. J., ALAVI, A. H., GANDOMI, A. H., WALKER, A. L. (2016): Machine learning in geosciences and remote sensing. In: *Geoscience Frontiers* 7 (1), S. 3–10.
- PANTER, K., GARDNER, D., LEE, S., PFISTER, J., RALPHS, M., STEGELMEIER, B., JAMES, L. (2007): Important poisonous plants of the United States, S. 825–872.
- PRAKASH, A. S., PEREIRA, T. N., REILLY, P. E. B., SEAWRIGHT, A. A. (1999): Pyrrolizidine alkaloids in human diet. In: *Mutation Research/Genetic Toxicology and Environmental Mutagenesis* (Volume 443, Issues 1–2), S. 53–67.

- STROTDREES, J. (2012): Empfehlungen für die Bewirtschaftung von Grünland zur Regulierung von Kreuzkraut-Vorkommen in Hamburg. Im Auftrag der Behörde für Stadtentwicklung und Umwelt, Amt für Natur- und Ressourcenschutz, Hamburg.
- STROTDREES, J. (2015): Status von Wasser- und Jakobs-Kreuzkraut auf ausgewählten Grünflächen in Hamburg. Im Auftrag der Behörde für Umwelt und Energie Amt für Natur- und Ressourcenschutz, Hamburg.
- SUTER, M, SIEGRIST-MAAG, S, CONNOLLY, J, LÜSCHER, A (2007): Can the occurrence of *Senecio jacobaea* be influenced by management practice? In: *Weed Research* 47 (3), S. 262–269.