

Geodaten für prädiktive und präskriptive Analysen – Ergebnisse aus dem Projekt BigGIS

Andreas Abecker¹, Johannes Kutterer¹

Abstract: Im April 2018 wurde das dreijährige BMBF Big-Data-Projekt BigGIS abgeschlossen. Anhand dreier konkreter Anwendungsszenarien wurden Methoden und Technologien genutzt, erprobt, weiterentwickelt und zusammengeführt, um Geodaten und Geodatenanalysen für Vorhersagen und Entscheidungsunterstützung im Sinne von “Big Data” zu verwenden. Die drei Anwendungsszenarien waren Umweltmanagement am Beispiel invasiver Spezies, Katastrophenschutz am Beispiel einer Schadgaswolke und Smart City mit Blick auf das Stadtklima und insbesondere urbane Hitzeinseln. In diesem Beitrag skizzieren wir einige wesentliche Ergebnisse aus BigGIS sowie denkbare zukünftige Forschungsfragen.

Keywords: Spatial Big Data, Big Geo Data, Geodatenverarbeitung, Geodatenanalyse, Stadtklima, Urbane Hitzeinseln, Notfallkartierung, Micro Rapid Mapping, Invasive Spezies.

1 Einleitung und Überblick

Ziel des BMBF-Forschungsvorhabens BigGIS („Prädiktive und präskriptive Geoinformationssysteme basierend auf hochdimensionalen geo-temporalen Datenstrukturen“) war die Erforschung, prototypische Umsetzung und Evaluierung von Techniken, Modellen und Methoden, die in vielfältigen Anwendungsfällen Entscheidungen auf Basis von großen Mengen an *zeitlich-strukturierten Geodaten aus unterschiedlichen Quellen* (wie insbesondere Fernerkundung und Crowdsourcing, aber auch aus Legacy-Systemen zur Geodatenverarbeitung) unterstützen. Auf Basis dieser Daten sollten *deskriptive, prädiktive und präskriptive Analysen* unterstützt werden. Die Verarbeitung der Daten sollte dabei schnell genug erfolgen, um z.B. auch die Gesundheits- und Umweltsicherung im Katastrophenfall zu unterstützen. Die FuE-Arbeiten wurden von den praktischen Erfordernissen dreier Anwendungsszenarien getrieben:

Smart City und Gesundheit: Die Förderung der Gesundheit von Menschen in Städten am Beispiel der Umwelteinflüsse Feinstaub und Temperatur.

Katastrophenschutz: Entscheidungsunterstützung bei komplexen Schadenslagen am Beispiel von Schadgas-Situationen wie z.B. bei Feuer in einer Chemiefabrik.

Umweltmonitoring: Umweltmanagement am Beispiel invasiver Tier- und Pflanzenarten, konkret betrachtet für die Kirschesigfliege.

¹ Disy Informationssysteme GmbH, Ludwig-Erhard-Allee 6, 76131 Karlsruhe, andreas.abecker@disy.net

Im BigGIS-Projektkonsortium arbeiteten folgende Organisationen zusammen:

FZI Forschungszentrum Informatik am Karlsruher Institut für Technologie mit den Gruppen:

- Shared Research Group „Corporate Services and Systems“ (Prof. Thomas Setzer): Projektkoordination, deskriptive und präskriptive Analysen
- Abteilung „Logistics and Supply Chain Optimization“ (Prof. Stefan Nickel): Präskriptive Analysen, Anwendungsfokus auf dem Szenario Smart City
- Abteilung „Wissensmanagement“ (Prof. Rudi Studer, Prof. York Sure-Vetter): Metadaten, semantische Technologien, nutzergenerierte Inhalte

Universität Konstanz, Lehrstuhl „Datenanalyse und Visualisierung“ (Prof. Daniel Keim): Visual Analytics für räumliche und zeitliche Big Data

Hochschule Karlsruhe Technik und Wirtschaft, Fachgebiet Datenbanksysteme und Webtechnologien (Prof. Jens Nimis): Architektur, Infrastrukturvirtualisierung

Disy Informationssysteme GmbH, Karlsruhe: Experten für GIS und räumliche Datenanalyse, GDI, Software-Integration, Experten für Anwendungen der Umweltverwaltung; Anwendungsfokus auf dem Szenario Umweltmanagement

Exasol AG, Nürnberg: Datenbankexperten, In-Memory-Datenbanken, Cluster Computing, Kompressionsalgorithmen

EFTAS Fernerkundung Technologietransfer GmbH, Münster: Erzeugung und Analyse von Fernerkundungsdaten von Satelliten und UAS; Anwendungsfokus auf dem Szenario Katastrophenschutz

Landesanstalt für Umwelt, Messungen und Naturschutz Baden-Württemberg (LUBW), Karlsruhe: Datenbereitstellung, Anforderungsgeber und Evaluationspartner, Anwendungsfokus auf dem Szenario Umweltmanagement

Assoziierte Partner für die Szenarien Katastrophenschutz und Smart City waren das THW Karlsruhe und die Stadtverwaltung Karlsruhe.

Im vorliegenden Beitrag diskutieren wir, nach einer kurzen Übersicht zu den verwendeten Basistechnologien, pro bearbeitetem Anwendungsszenario kurz (i) zentrale Forschungsfragen, (ii) im Projekt erarbeitete Lösungsansätze und Ergebnisse und (iii) die Übertragbarkeit auf aktuelle und mögliche zukünftige Szenarien der Umweltinformatik.

2 Lösungsarchitektur und verwendete Softwarekomponenten

BigGIS-Anwendungslösungen werden aus einem Werkzeugkasten aufeinander abgestimmter Softwarekomponenten zusammengestellt. Die wichtigsten Infrastruktur- und Lösungskomponenten sind in der BigGIS-Lösungsarchitektur in Abb. 1 eingeordnet.

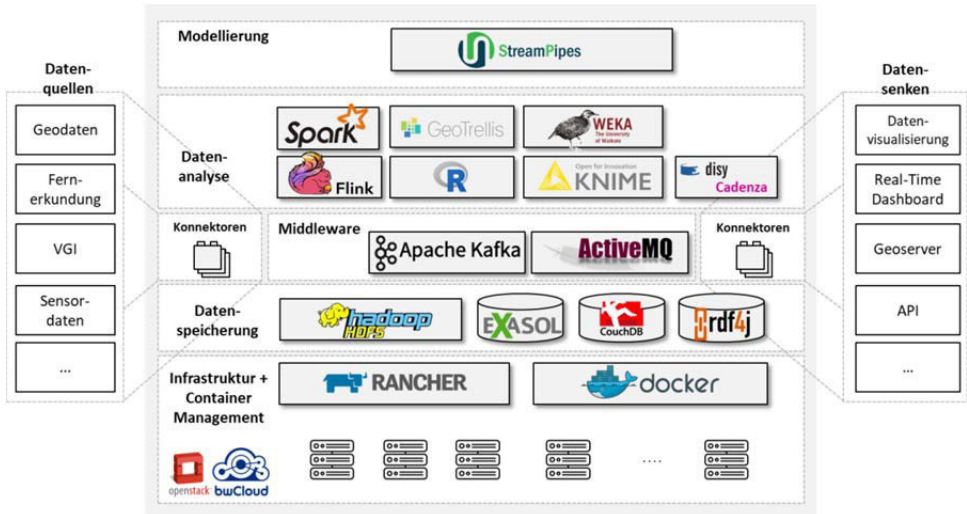


Abb. 1: Ausgewählte Softwarekomponenten für BigGIS-Anwendungslösungen

Auf der *Infrastrukturebene* finden sich Werkzeuge wie Docker und Rancher für die Bereitstellung der erforderlichen Speicher- und Rechenressourcen und für das automatisierte Management dieser Virtualisierungsschicht. Dadurch kann eine BigGIS-Lösung container-basiert verhältnismäßig einfach installiert und im Fall starker Schwankungen im Datenanfall und in den Rechenanforderungen leicht automatisch skaliert werden.

Zur *Datenspeicherung* kommen verschiedene moderne Datenbanktechnologien im Projekt zum Einsatz. Disy führte Benchmarks von In-Memory-Datenbanken wie SAP HANA im Vergleich mit etablierten Geodatenbanken wie Oracle Spatial und Postgres/PostGIS durch. Der BigGIS-Projektpartner Exasol erweiterte die Geodatenfähigkeiten seiner bereits führenden In-Memory Datenbank erheblich. Dokumentorientierte NoSQL-Ansätze für die hochverfügbare Datenspeicherung (Hadoop HDFS, CouchDB) werden insbesondere für sehr große Datenvolumina wie Rohdaten aus der Fernerkundung verwendet. RDF4J (früher Sesame) dient zur Speicherung semantischer Metadaten.

Auf der *Middleware-Ebene* geht es primär um das Message Brokering, also die Entkopplung der direkten synchronen Kommunikation von eingehenden Datenströmen mit Weiterverarbeitungsprozessen. Hier werden Kafka und ActiveMQ genutzt.

Die Ebene der *Data Analytics* war der Projektschwerpunkt verschiedener Partner. Vielfältige Tools und Frameworks stehen zur Verfügung, wie Flink, R und Spark als sehr generische Werkzeuge für (unter anderem) Datenanalyse und Maschinelles Lernen, oder auch GeoTrellis als Spezialwerkzeug für große Mengen von Raster-Geodaten. Die Universität Konstanz nutzte Maschinelle-Lern-Algorithmen aus KNIME und WEKA und entwickelte neue Visualisierungsmöglichkeiten zur interaktiven Datenanalyse. Disy inte-

grierte die Teillösungen prototypisch und koppelte sie mit Cadenza,² der eigenen Plattform für Data Analytics, Reporting und GIS.

Die *Ebene der Modellierung* unterstützt die Konfiguration komplexer Big-Data-Verarbeitungsworkflows mithilfe des FZI-Werkzeugs StreamPipes.³

3 Anwendungsszenario Smart City / Stadtklima

Die zunehmende Urbanisierung in Deutschland und weltweit, verstärkt durch die Folgen des Klimawandels, machen das Stadtklima (vgl. [Ku04], [Ku04b]) zu einem Thema mit wachsender Bedeutung. Das Stadtklima als Lokalklima auf der Meso- und Mikroebene hat Auswirkungen auf menschliches Wohlbefinden und Gesundheit, aber auch auf innerstädtische Flora und Fauna oder sogar auf die Standorteignung für bestimmte Wirtschaftstätigkeiten. Das Stadtklima wird durch Faktoren wie Temperatur, direkte Sonneneinstrahlung, Luftfeuchtigkeit, Windgeschwindigkeit, Luftqualität oder auch Pollenflug und Lärmbelastung bestimmt und kann z.B. durch städtebauliche Maßnahmen, Stadtbegrünung oder Maßnahmen der Verkehrssteuerung beeinflusst werden.

In BigGIS wurde federführend vom FZI in Zusammenarbeit mit der Stadt Karlsruhe und EFTAS das Phänomen der innerstädtischen Hitzeinseln (*Intra Urban Heat Islands*, IUHI) untersucht [BS17]. Ziel war es, IUHI genauer identifizieren und vorhersagen zu können. Auf dieser Basis würden dann z.B. Fußgängerrouting mit minimaler Hitzebelastung oder zielgerichtete Warnmeldungen für Risikogruppen (Kindergärten, Altenheime, etc.) möglich sein [RB17].

Schon für das kleinräumige Betrachten der Temperatur sind die amtlichen Messnetze bei Weitem zu weitmaschig. Deshalb war ein Hauptthema hier die gemeinsame Nutzung und Fusion von Sensordaten der Verwaltung, von privaten Nutzern, von Satelliten und aus mobiler Sensorik (Befliegung mit Thermalkamera, Sensorik auf Straßenbahn – betrifft nur Luftqualität⁴). Weitere Datenquellen waren ATKIS/ALKIS, NDVI und LoD2-3D für Karlsruhe.

Durch die Vielfalt der Sensorik wurden Verfahren zum Berücksichtigen, Verarbeiten und Darstellen von Unsicherheiten (z.B. mit Hierarchischen Bayes-Modellen, BHM) in den Eingangsdaten und Ergebnissen statistischer Analysen und Vorhersagen wichtig. Wegen häufig unzureichender Messpunktdichte wurden Interpolationsverfahren mit Verwendung von Zusatzinformation (Landnutzung, Uhrzeit) erprobt. Hieraus wurde am FZI ein landnutzungsbasiertes Temperaturprediktionsmodell erstellt, das stabile Vorhersage von Temperaturen auf Stundenbasis und im Bereich weniger Quadratmeter ermöglicht, angepasst auf innerstädtische Vorhersagen, implementiert auf Basis von DWD- und

² <http://www.disy.net/de/produkte/cadenza/>

³ <https://www.streampipes.org/de/>

⁴ <http://www.aero-tram.kit.edu/index.php>

LUBW-Wetterstationen. Außerdem wurden ein automatisches Qualitätsassessment und darauf basierendes Vorhersagemodell mit Crowd-basierten Wetterstationen erstellt und ein Optimierungsverfahren entwickelt, das unter Berücksichtigung der Sensorunsicherheiten die besten Standorte für zusätzliche Messstationen zum Erreichen einer gewünschten Vorhersagequalität bestimmt. Die Getis-Ord G^* Statistik zur Bestimmung von Hot Spots wurde zur Variante “Focal G^* ” erweitert, die stabilere Hot Spots identifiziert, d.h. weniger abhängig von Verfahrensparametern wie der Rasterauflösung [BS17]. Um Stadtplanern ein Instrument zum besseren Verständnis der IUHI-Entstehung zu geben, wurde von FZI und Uni Konstanz ein interaktives Simulationssystem erstellt, das auf Basis historischer Messwerte für einstellbare Parameter wie Lufttemperatur, Bodentemperatur, Luftdruck usw. die erwarteten IUHI im Zeitverlauf und in Abhängigkeit von Landnutzung etc. darstellt (Abb. 2).

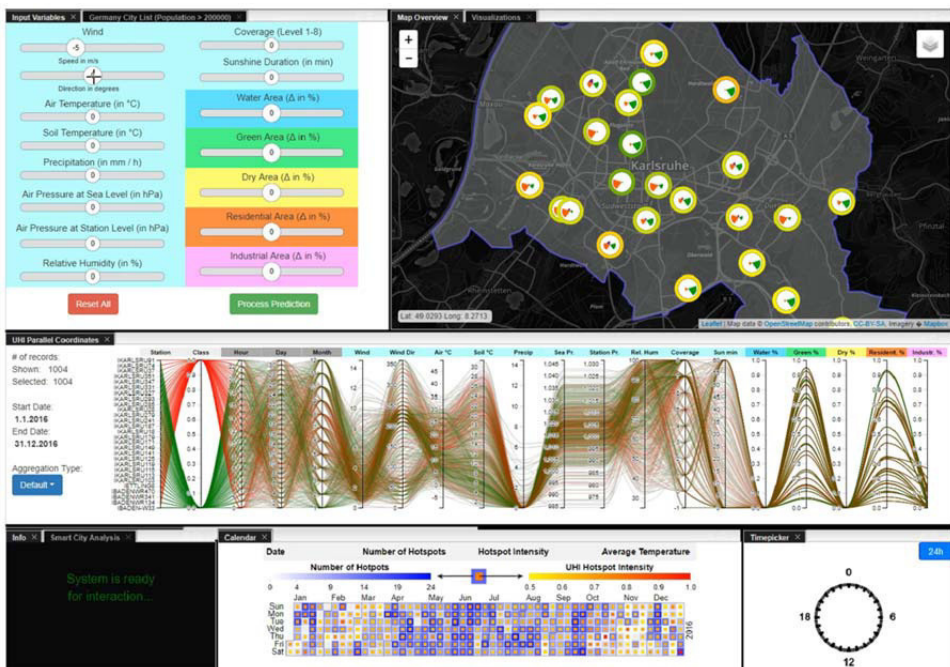


Abb. 2: Interaktives Simulationssystem für IUHI. Quelle: FZI, Uni Konstanz

Bewertung, Praxistauglichkeit, Folgeforschung: Die eingesetzten Verfahren lieferten aus Sicht des Machine Learning insgesamt sehr vielversprechende Ergebnisse. Die entwickelten Methoden zum Umgang mit Unsicherheiten und mit dünnen und heterogenen Messnetzen können als generisch für die Umweltinformatik betrachtet werden. Im konkreten Anwendungsbeispiel wurde nur ein kleiner Teil der denkbaren Eingangsparameter genutzt. Auch weitere Aspekte des Stadtklimas (wie Luftqualität mit Faktoren wie Feinstaub, Ozon, Schwefeloxide und Stickoxide) und der Einflussfaktoren zu beobachten und daraus Entscheidungsunterstützung für planerische oder auch operative

Maßnahmen abzuleiten, birgt noch vielfältige interdisziplinäre Forschungsfragen. Die Kombination amtlicher Messnetze (qualitätsgesichert, aber sehr dünn) mit mobilen oder stationären privaten Messstationen (im Sinne der Citizen Science) gewinnt zunehmende Bedeutung in Umweltsanwendungen.

4 Anwendungsszenario Katastrophenschutz

Das Szenario Katastrophenschutz wurde primär von EFTAS und FZI bearbeitet [BBG17]. Es ist bekannt, dass Feuerwehren bei Unfällen, insbesondere bei Gefahrstoffunfällen vom zeitnahen, einfachen und integrierten Zugriff auf vielfache Sach- und Geodaten profitieren könnten, wie z.B. Basisinformationen zu Natur- und Wasserschutzgebieten, kritischen Infrastrukturen, Verkehrs- und Evakuierungswegen, Chemikalieninformation u.v.m. Hier gibt es bereits vielfältige Lösungsansätze, die sich jedoch erst sehr langsam in der Praxis durchsetzen. Darüber hinaus ist es gerade in unübersichtlichen Katastrophensituationen wichtig, vor Ort sehr schnell ein integriertes aktuelles Lagebild zu Zerstörungen, Gefährdungen, Rettungsmitteln im Einsatz usw. zu gewinnen.

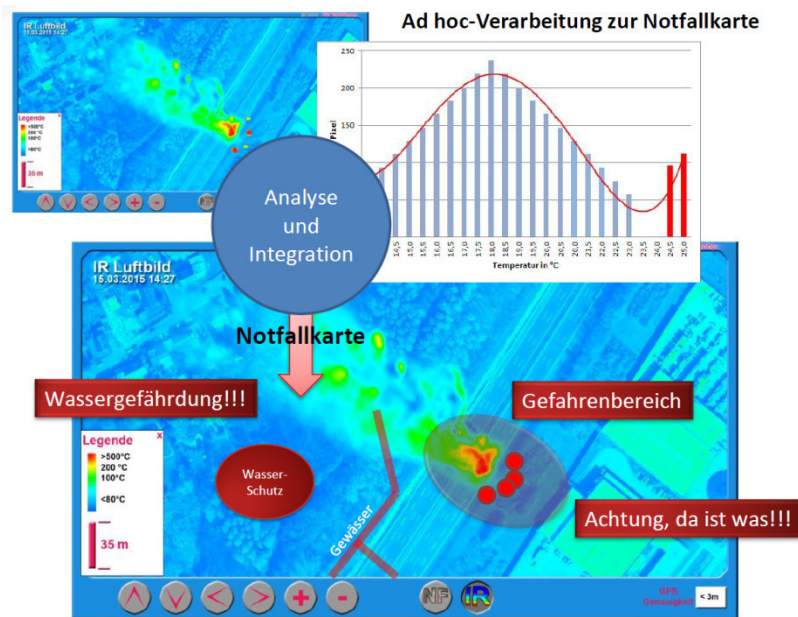


Abb. 3: Illustration der Idee des MRM - aktuelle Luftbilder werden mit Hintergrundkarten und bekannten Umweltdaten verschnitten und einsatzspezifisch aufbereitet. Aus: [BBG17]

EFTAS hat für diese Fragestellung den Begriff des „Micro Rapid Mapping“ (MRM) geprägt [BW14] (vgl. Abb. 3). Hier geht es darum, durch Informationsfusion von vorab bekannten Geobasisdaten, ggf. hinzuziehbarer aktuellen Satellitendaten, ggf. vorhandenen

Social Media Informationen und insbesondere von vor Ort mithilfe von Flugrobotern (Drohnen, Kopter, allgemein: UAV) gemachten (Hyperspektral-)Kamera-Aufnahmen innerhalb von max. 10 Minuten ein integriertes Lagebild mit den einsatzrelevanten Informationen in handhabbaren Visualisierungen zu erstellen und aktuell zu halten.

Die Herausforderungen sind vielfältig. Sie beginnen bei der Flugfähigkeit der Trägersysteme bei Wind, Regen oder in Feuernähe, über die Missionsplanung für die Befliegung, die Georeferenzierung der Luftbilder, die semantische Analyse der Aufnahmen, die Datenübertragung vom UAV zur Zentrale, bis zur Darstellung komplexer Zusammenhänge für Nicht-Geo-IT-Spezialisten unter Stress und Zeitdruck. Etliche dieser Probleme werden zurzeit an vielen Stellen in der Geoinformatik bearbeitet.

In BigGIS wurde ein prototypisches Gesamtsystem mit dem AiD FD8ex Octocopter, verschiedenen Kameras (IR / RGB / Hyperspektral), einem RTK GPS und einem Inertialsystem aufgesetzt und experimentell getestet. Es wurden Methoden entwickelt, um aus Tausenden von Bildern diejenigen auszuwählen, die am besten zusammen ein Orthophoto für den gesamten Unfallort ergeben. Dazu muss die Vorverarbeitung schon auf dem UAV geschehen. Die Erkennbarkeit von Rauchgaswolken und von ausströmendem Methan ("unsichtbares" Stadtgas) wurden untersucht und einfache Ausbreitungsprognosen einer Schadgaswolke erstellt [BBG17]. Schließlich wurden Methoden der diskreten Optimierung für die Flugplanung von UAV im Notfalleinsatz angewandt, um mit begrenzten Ressourcen den optimalen Informationsgewinn zu erzielen [GMT17].

Bewertung, Praxistauglichkeit, Folgeforschung: Zu allen oben angesprochenen Themen wurden Anfangserfolge erzielt. Aber all diese Felder erfordern auch noch weitere Forschung, um sie zu praxistauglichen Lösungen zu führen, die nicht nur Menschen und Anlagen durch bessere Kenntnisse von Umweltfaktoren besser schützen, sondern auch die Umwelt vor den Auswirkungen der Unfälle schützen. Dennoch schätzen wir dieses Szenario als dasjenige ein, das voraussichtlich am schnellsten zu operativen Lösungen führen wird. Dies liegt einerseits daran, dass diese Technologiekombination (UAV + Fernerkundung + Datenintegration in Lagebild oder anderes Backend System) vielfältigste Einsatzmöglichkeiten besitzt, bei Notfällen wie Unfällen oder Bränden, bei großräumigen Naturkatastrophen wie Erdbeben oder Hochwasser oder auch bei der Lageinformation für große Polizeieinsätze (Großdemonstrationen o.ä.). Zweitens daran, dass die diversen eingesetzten Technologien große Attraktivität für Forscher und Firmen besitzen. Drittens aber auch daran, dass für verschiedene schwierigere Aufgaben auch schon einfachere manuelle Lösungen anwendbar sind (wie manuelle Drohnensteuerung oder manuelle Bildinterpretation).

5 Anwendungsszenario Umweltmonitoring / Invasive Spezies

Hier wurden Tiere und Pflanzen mit negativen Auswirkungen auf die menschliche Gesundheit, das ökologische Gleichgewicht oder ökonomische Interessen, wie landwirtschaftlicher Ertrag, untersucht. Dies umfasst z.B. Pflanzenkrankheiten und -schädlinge

oder auch Spezies, die starke allergische Reaktionen bei Menschen hervorrufen, wie der Eichen-Prozessionsspinner oder die Beifußblättrige Ambrosie, oder die Krankheiten übertragen können, wie die Asiatische Tigermücke. Viele dieser Spezies sind aufgrund menschlicher Aktivitäten oder durch den Klimawandel begünstigt, erst in der jüngeren Vergangenheit eingewandert. Ein wichtiges Beispiel für solche „invasive Spezies“ ist die Kirschessigfliege, KEF, die in BigGIS federführend von der Universität Konstanz, Disy und der LUBW für die Weinbaugebiete in Baden-Württemberg untersucht wurde [Mu18], [Se17]. Sie befällt gesunde Früchte wie Birnen, Pflaumen, Kirschen, Trauben, Nektarinen, Pfirsiche, Blaubeeren, Himbeeren und Erdbeeren und stellt daher für Obstbauern und Winzer eine erhebliche wirtschaftliche Bedrohung dar. In den USA wurden in einigen Kulturen Ertragsverluste von bis zu 80% beobachtet [BGZ09].

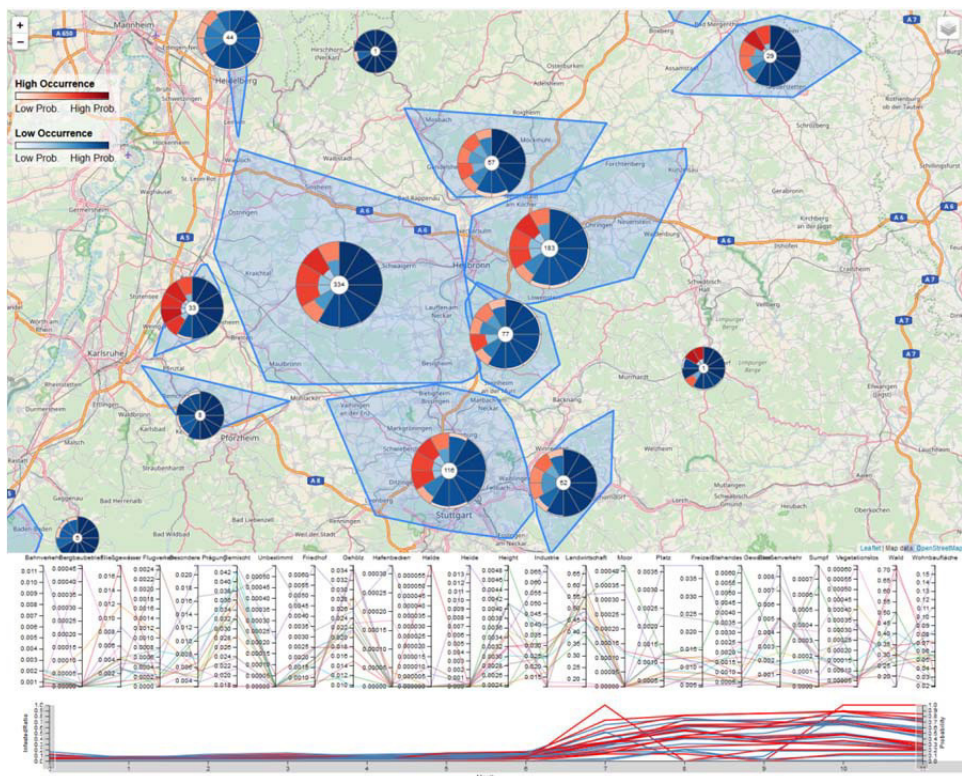


Abb. 4: Interaktives Tool zur Analyse der Befallswahrscheinlichkeit der Kirschessigfliege mit Karte, Glyphendarstellung, gereihten Koordinatenachsen und Liniendiagramm. Aus: [Se17]

In BigGIS wurde Ensemble-basierte Klassifikation auf Basis bekannter Lernalgorithmen (Entscheidungsbaum, k-NN, Random Forest, K*, SVM, ...) aus KNIME und WEKA angewandt, um räumlich-zeitliche Befallsprognosen zu erstellen [Se17]. Die Datengrundlage umfasste neben Geobasisdaten (Landnutzung, Geländemodell) und Wetterdaten des

DWD insbesondere Beobachtungsdaten von VitiMeteo⁵, dem Informationsdienst des Staatlichen Weinbauinstituts Freiburg (WBI), nämlich: (i) Daten zu Fallenfängen, (ii) Daten zu Eifunden von Mitarbeitern des WBI und der LVWO Weinsberg sowie (iii) Befallsbeobachtungen, die vor Ort von Rebschutzwarten, Weinbauberatern und Mitarbeitern von WBI und LVWO gemacht wurden. Die Daten wurden zunächst in eine vergleichbare, georeferenzierte Form überführt sowie typische Vorverarbeitungsschritte des Data Mining ausgeführt [Se17]. Danach wurde das Befallsrisiko für jedes Weinbaug Gebiet und jeden Monat des Jahres vorhergesagt. Zur Darstellung und Inspektion der gelernten Prognosen hat die Universität Konstanz verschiedene kombinierbare interaktive Visualisierungen für Fachexperten entwickelt (Abb. 4).

Bewertung, Praxistauglichkeit, Folgeforschung: Die Lernergebnisse und die Visualisierungsansätze haben bei Fachexperten für die KEF bereits großen Anklang gefunden. Man kann davon ausgehen, dass für das Verständnis des Ausbreitungsverhaltens durch Hinzuziehen weiterer Eingangsdaten und durch flexiblere räumliche und zeitliche Datenstrukturierungen noch bessere Vorhersageergebnisse möglich sind. Langfristig sind natürlich sowohl die Nutzbarkeit für die Auftrittsanalyse anderer invasiver, bzw. allgemein schädlicher Spezies von Interesse als auch die Nutzung der verwendeten Methoden für das bessere Verständnis anderer lokaler / regionaler ökologischer Phänomene wie z.B. regionales Bienensterben. Aus Big-Data-Sicht wird der Anwendungsfall dann noch sehr viel interessanter, wenn man Prognosemodelle in Echtzeit mit Beobachtungsdaten und Wetterprognosen koppelt, um dadurch aktuelle Handlungsempfehlungen für die Schädlingsbekämpfung zu erhalten. Solche Ansätze könnten allgemein in der Landwirtschaft für besseren Pflanzenschutz bei geringerem Einsatz von Pflanzenschutzmitteln beitragen.

6 Synergien und übergreifende Aspekte

Die vielfältigen Projektergebnisse aus BigGIS und auch der in diesem Beitrag zusammengefasste Teil davon sind teilweise etwas unzusammenhängend, weil das Projekt bewusst sehr breit angelegt war, um in den verschiedensten Teilen im Bereich Spatial Big Data Beiträge zu liefern und Technologielücken zu schließen, wie auch die Vielfalt des Themas aufzuzeigen (vgl. auch [Ab16]). Das ist auch darin begründet, dass die Anwendungsszenarien selber eine gewisse Breite und Generizität besitzen und bewusst komplementär angelegt waren, um die wesentlichen „vier V“ der Big Data Charakterisierung (Volume, Velocity, Variety, Veracity) abzudecken.

Dennoch gibt es natürlich einige Querschnitts- und generische Technologien, die in mehr als einem Anwendungsszenario relevant bzw. von allgemeinerem Interesse sind:

⁵ <http://www.vitimeteo.de/>

- Die Docker-basierte Ressourcenbereitstellung und Basisinfrastruktur ist generisch und wurde von allen Projektteilen verwendet.
- Aus Sicht der Gesamtarchitektur war es eine wesentliche Erkenntnis des Projekts, dass eine einheitliche und allumfassende „Spatial Big Data“ Architektur bei der Unterschiedlichkeit der Anwendungsfälle zurzeit nicht sehr viel Sinn macht. Stattdessen haben wir die grobe Struktur von Lösungen skizziert, erfahrungsgemäß gute Komponenten für bestimmte Aufgabentypen identifiziert (wie in Abb. 1 dargestellt) und beispielhafte Teile von Verarbeitungspipelines realisiert, die auch in unterschiedlichen Anwendungsfällen nutzbar sind bzw. wären. So kann z.B. die „Rasterdaten-Pipeline“ für die Einbindung von Fernerkundungsergebnissen von Satelliten oder von Drohnen sowohl im Szenario „Smart City“ als auch im Szenario „Katastrophenschutz“ genutzt werden. Genauso wie die „Vektordaten-Pipeline“ in ähnlicher Weise zum Integrieren von Beobachtungsdaten zu „Invasiven Spezies“ und zum Integrieren von Wetterstationsdaten aus der Citizen Science.
- Verschiedene Werkzeuge, die im Projekt entwickelt oder weiterentwickelt wurden, können allgemein verwendet werden. Die NoSQL-Datenbanktechnologie von Exasol ist natürlich generisch anwendbar, ebenso wie die Visualisierungsansätze der Universität Konstanz zur interaktiven Analyse komplexer Sachverhalte, die für Urbane Hitzeinseln und für Invasive Spezies genutzt wurden (vgl. Abb. 2 und Abb. 4). Allgemeine geostatistische Methoden zur Belastbarkeit von Aussagen in Abhängigkeit von der Dichte des Messnetzes sind die Basis der Flugplanung im Szenario „Katastrophenschutz“, aber auch für die Planung und Bewertung der Messwerterhebung im Bereich „Stadtklima“.

7 Zusammenfassung und Ausblick

Wir haben Ergebnisse des Projekts BigGIS skizziert. Auf der einen Seite sind dies *domänenunabhängige Technologien*, die jedoch *auch* in der Umweltinformatik anwendbar sind [Ab16]. Besonders überzeugend für den praktischen Einsatz waren die containerbasierten Virtualisierungsansätze, das Message Brokering mit Kafka und die verteilte Datenprozessierung mit Flink, womit aufwändige Verfahren der Geostatistik parallelisiert wurden. Wenn man in zukünftigen Erweiterungen feinkörnigere lokale Wetterdaten oder Echtzeitbeobachtungen für die Online-Analyse verwendet, werden Aspekte wie die leistungsfähige Datenstromverarbeitung noch deutlich relevanter. Allgemein weisen alle Verarbeitungspipelines in BigGIS eine hohe Komplexität auf. Wir sehen daher auch in der (teil-)automatisierten Konfiguration von Big-Data Workflows, z.B. mithilfe semantischer Metadaten für Daten und Verarbeitungsverfahren interessante zukünftige Arbeiten. Konzeptionen und erste Schritte in dieser Richtung wurden im Projektrahmen bereits unternommen [WSN17]. Um aus Spatial Big Data Frameworks einfach bedienbare und universell anwendbare Werkzeugkästen – auch für die Umweltinformatik – zu machen, besitzen einige der betrachteten Querschnittsthemen, wie Umgang mit sehr heterogenen

Daten, Umgang mit Unsicherheit und Unvollständigkeit in Daten, Nutzung semantischer Metadaten für die Datenintegration und die assistierte Pipeline-Konfiguration, Interaktivität bei der Datenanalyse, aber auch effiziente Einbindung von Fernerkundungsergebnissen und benutzerfreundliche Bedienoberflächen für komplexe Fragestellungen und/oder komplexe Fragekontexte (wie Lagezentralen im Katastrophenschutz) auch noch hohe Relevanz für Forschungs- und Entwicklungsarbeiten der kommenden Jahre.

Auf der anderen Seite wurden *szenariospezifische Entwicklungen* gemacht. Im Bereich *Smart City* wurde insbesondere die Thematik untersucht, aus räumlich und technisch heterogenen Sensornetzwerken zuverlässige Vorhersagen zum Stadtklima zu machen. Die entwickelten Methoden zum Umgang mit Unsicherheiten und mit heterogenen Messnetzen besitzen allgemeine Relevanz in zukünftigen UI-Anwendungen im Zusammenspiel von Verwaltung, allgemeiner Öffentlichkeit, Forschern und Firmen. Im Bereich *Umweltmonitoring* standen Ausbreitungsprognosen und Visualisierungen umfangreicher Lernergebnisse im Fokus. Für das Verständnis komplexer Analyseergebnisse wird auch dies noch lange relevant bleiben. Die Kopplung der Analyse mit *Datenströmen* zur Entscheidungsunterstützung in (Nah-)Echtzeit ist ein folgerichtiger nächster Schritt. Im Bereich *Katastrophenschutz* waren die Erkennbarkeit bestimmter Gaswolken aus Luftbildern und die Missionsplanung für UAV zentral. Die Idee des *Micro Rapid Mapping* für Notfallkartierungen birgt darüber hinaus noch etliche FuE-Bedarfe. Auch die Ansätze zur effektiven Nutzung von UAV zur Informationsgewinnung (optimierte Routenplanung, Dimensionsreduktion und Orthorektifizierung auf dem Gerät etc.) spielen über das Szenario Katastrophenschutz hinaus eine Rolle.

Danksagung: *Die vorgestellten Arbeiten wurden mit Unterstützung des Bundesministeriums für Bildung und Forschung (BMBF) im Rahmen des Projekts „BigGIS: Prädiktive und präskriptive Geoinformationssysteme basierend auf hochdimensionalen geo-temporalen Datenstrukturen“ (FKZ: 01IS14012A-G) durchgeführt. Die vorgestellten Ergebnisse wurden vom gesamten Projektkonsortium erarbeitet, federführende Partner und Quellen sind, wo sinnvoll möglich, angegeben. Weitere Informationen finden sich unter <http://biggis-project.eu/biggis-docs/>).*

Bibliographie

- [Ab16] Abecker, A. et al.: Überlegungen zu einer Spatial Big Data Architektur im BigGIS Projekt. In: (Freitag, U. et al. Hrsg.): 23. Workshop "Umweltinformationssysteme", Leipzig, Deutschland, 2016. CEUR Workshop Proceedings 1781.
- [BBG17] Bernsdorf, B.; Bruns, J.; Glock, K.: BigGIS - Einsatz von Flugrobotern zur Entscheidungsunterstützung bei Bränden. In: 2nd BMBF Big Data All Hands Meeting and 2nd Smart Data Innovation Conference, Karlsruhe, Deutschland, 2017.

- [BGZ09] Bolda, M.P.; Goodhue, R.E.; Zalom, F.G.: Spotted Wing Drosophila: Potential Economic Impact of Newly Established Pest. Giannini Foundation of Agricultural Economics, University of California. 2009.
- [BS17] Bruns, J.; Simko, V.: Stable Hot Spot Analysis for Intra Urban Heat Islands. In: GI-Forum 2017, Salzburg, Österreich, 2017.
- [BW14] Bernsdorf, B.; Woditsch, S.: Micro Rapid Mapping - Anforderungen räumlich begrenzter Notfallkartierungen an moderne GeoIT in Feuer- und Katastrophenschutz. In: (Pottebaum, J.; Erfurth, C.; Reuter, C. Hrsg.): Workshop IT-Unterstützung in Emergency Management & Response, 2014.
- [GMT17] Glock, K.; Meyer, A.; Tack, G.: Correlated Orienteering for Planning Emergency Surveillance Flights of Unmanned Aerial Vehicles. In: VeRoLog - Annual WS EURO WG Vehicle Routing & Logistics optimization, Amsterdam, Niederlande, 2017.
- [Ku04] Kuttler, W.: Stadtklima, Teil 1: Grundzüge und Ursachen. Umweltwissenschaften und Schadstoffforschung – Zeitschrift f. Umweltchemie u. Ökotoxikologie 16/3, 2004.
- [Ku04b] Kuttler, W.: Stadtklima, Teil 2: Phänomene und Wirkungen. Umweltwissenschaften und Schadstoffforschung – Zeitschrift f. Umweltchemie u. Ökotoxikologie 16/4, 2004.
- [Mu18] Müller, H. et al.: Nutzung von Big Data Technologien für den Umgang mit invasiven Spezies. In: (Weissenbach, K. et al. Hrsg.): Gesamtbericht INOVUM Phase II – Innovative Umweltinformationssysteme. KIT Scientific Reports, Karlsruhe, 2018.
- [RB17] Rußig, J.; Bruns, J.: Reducing Individual Heat Stress Through Path Planning. In: GI-Forum 2017, Salzburg, Austria, 2017.
- [Se17] Seebacher, D. et al.: Visual Analysis of Spatio-Temporal Event Predictions – Investigating the Spread Dynamics of Invasive Species. In: Symposium on Visualization in Data Science (VDS) at IEEE VIS 2017, Phoenix, Arizona, USA, 2017.
- [WSN17] Wiener, P.; Simko, V.; Nimis, J.: Taming the Evolution of Big Data and its Technologies in BigGIS: A Conceptual Architectural Framework for Spatio-Temporal Analytics at Scale. In: GISTAM 2017 – International Conference on Geographical Information Systems Theory, Applications and Management, Porto, Portugal, 2017.