



RUNDER TISCH GIS E.V.

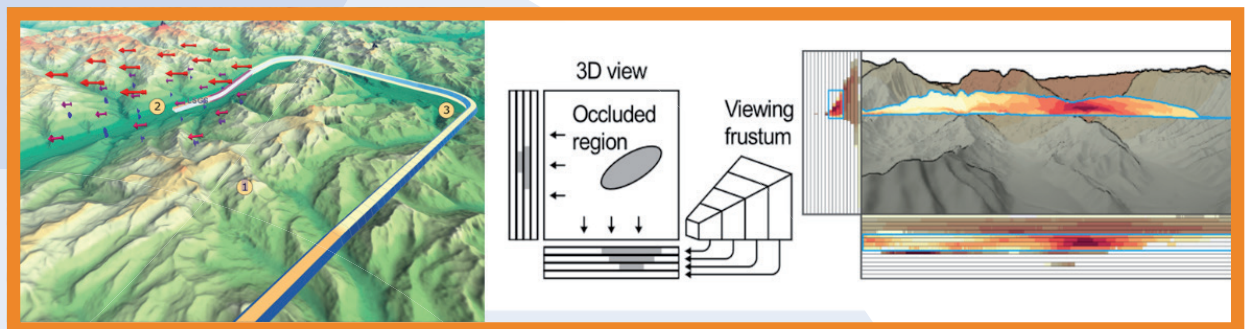


Technische Universität München

Kolbe / Bill / Donaubauer (Hrsg.)

# Geoinformations- systeme 2019

Beiträge zur 6. Münchner GI-Runde



München, März 2019

**Titelgrafik:**

siehe Beitrag "Visuelle Analyse von Daten mit 2- oder 3-dimensionalem räumlichen Bezug" (Abb. 3)

Alle in diesem Werk enthaltenen Angaben, Daten, Ergebnisse, usw. wurden von den Autoren nach bestem Wissen erstellt und mit größtmöglicher Sorgfalt geprüft. Dennoch sind inhaltliche Fehler nicht völlig auszuschließen. Daher erfolgen die Angaben usw. ohne jegliche Verpflichtung oder Garantie der Autoren. Sie übernehmen deshalb keinerlei Verantwortung und Haftung für etwa vorhandene inhaltliche Unrichtigkeiten.

© 2019 Runder Tisch GIS e.V.

**Erscheinungsweise**

Digital als PDF unter [www.rundertischgis.de/veranstaltungen/muENCHNER-GI-RUNDE.html](http://www.rundertischgis.de/veranstaltungen/muENCHNER-GI-RUNDE.html)

**Urheberrecht**

Dieses Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Runden Tisches GIS e.V. unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

# Inhaltsverzeichnis

<b>Vorwort</b> .....	5
<b>Förderpreis Geoinformatik</b>	
<i>Boates, I.:</i> Demonstrating Utility Network Interdependency Modelling Using the UtilityNetwork Application Domain Extension for CityGML.....	6
<i>Coduro, T.:</i> Straßenraummodellierung mittels Mobile Mapping in OpenDRIVE und CityGML sowie Entwicklung geeigneter Visualisierungsmethoden.....	13
<i>Hoedt, F.:</i> Einflussfaktoren auf die Lokalisationsgüte von Passive Audio Monitoring Systems (PAMS) .....	16
<i>Mollenhauer, T.:</i> Extracting Traffic Lanes from Floating Car Data .....	21
<i>Zimmer, S.:</i> An Efficient Algorithm for Computing Space-Time-Linguistics Similarities and Labelling Social Media Posts.....	23
<i>Meyer, H.:</i> Datengetriebene Modellentwicklung in der Umweltgeographie: Methodische Weiterentwicklungen und wissenschaftliche Anwendungen .....	26
<i>Westerholt, R.:</i> Statistische Räumliche Analyse in der Digitalen Transformation: das Beispiel Geosozialer Medien.....	29
<b>Straßenraummodellierung und Monitoring</b>	
<i>Schwab, B. and Kolbe, T.:</i> Requirements of road space models in the context of automated driving.....	36
<i>Radies, C.:</i> High Definition Maps – so entstehen hochgenaue Referenzkarten .....	38
<i>Schnettler, M.:</i> Erweiterung der Umfeldwahrnehmung automatisierter Fahrzeuge durch intelligente Infrastruktur .....	41
<b>Simulation und GIS</b>	
<i>Westermann, R.:</i> GPU-basierte Simulation – Stand der Technik und Anwendungspotenziale.....	44
<i>Grigoropoulos, G. and Keler, A.:</i> Using Geodata for Simulating Urban Traffic – Current Research in the Field of Traffic Engineering and Control .....	45
<i>Fischer, K. und Trometer, S.:</i> Verwundbarkeitsanalyse urbaner Gebiete, 3D Bewertung von disruptiven Ereignissen zur Steigerung der Resilienz ....	50
<b>Geo-Visualisierung</b>	
<i>Kreuziger, U.:</i> Visualisierung von Geobasisdaten mit georeferenzierender AR-Technologie (Praxisbeispiele).....	53
<i>Meng, L.:</i> Visualisierung von Geodaten – von der letzten Meile zur besten Meile.....	56
<i>Schumann, H. und Dübel, S.:</i> Visuelle Analyse von Daten mit 2- oder 3-dimensionalem Bezug .....	58
<b>Punktwolken</b>	
<i>van Oosterom, P., Meijers, M., Verbree, E., Liu, H. and Tijssen, T.:</i> Towards a relational database Space Filling Curve (SFC) interface specification for managing nD-PointClouds .....	61
<i>Richter, R., Discher, S. und Döllner, J.:</i> Verfahren zur Auswertung, Analyse und Nutzung von 3D-Punktwolken .....	72
<i>Martens, J. und Blankenbach, J.:</i> Scan2BIM... und warum die „2“ nicht nur eine Primzahl ist.....	75

**Geodateninfrastrukturen***Bill, R.:*

Hauptsache Open! – Der Zugang zu offenen Geodaten..... 77

*Beyel, S. und Fischer, M.:*

Visualisierung von Echtzeitgeodaten in der Urban Data Platform in Hamburg ..... 79

*Portele, C.:*

Next Generation GDI..... 82

**BIM / GIS / Smart Cities***Holweg, D., Willkomm, P. und Schulz, H.:*

Deep Space BIM - Digitaler Bauassistent der Zukunft..... 86

*Coors, V.:*

Integrierte Web-basierte Visualisierung von 3D-Stadt- und -Landschaftsmodellen und Sensordaten auf Basis von OGC Standards ..... 88

*Blessing, E., Bludovsky, S., Graf, G. und Vögele, M.:*

Smart Villages - Vernetzung von Geoinformationen und umwelt- relevanten Sensordaten ..... 90

**Förderpreis Geoinformatik***Nienaber, J.:*

GIS-basierte Straßenkontrolle der Freien und Hansestadt Hamburg ..... 92

## Vorwort

Der vorliegende Tagungsband ist eine Sammlung der erweiterten Abstracts zur *Münchener Geoinformatik-Runde 2019*, die vom *Runden Tisch GIS e.V.* am 14. und 15. März 2019 an der Technischen Universität München ausgerichtet wurde.

Die *Münchener GI-Runde 2019* setzte sich zusammen aus fünf Foren mit eingeladenen Vorträgen aus der Wissenschaft, drei Sitzungen mit Beiträgen aus der Praxis sowie dem *Innovationsforum* des *Runden Tisch GIS e.V.*, für das sich Unternehmen mit kreativen, neuartigen Anwendungen und Lösungen rund um die Verwendung von Geoinformation bewerben konnten.

Die Beiträge in diesem Tagungsband entstammen den Wissenschafts- und Praxisforen, für die hochkarätige Vortragende aus der Geoinformatik und benachbarten Disziplinen wie Computergrafik und Verkehrstechnik gewonnen werden konnten. Die erweiterten Abstracts geben einen Überblick über aktuelle Arbeiten und Projekte aus den Bereichen *Straßenraummodellierung und Monitoring, Simulation und GIS, Geodateninfrastrukturen, Geo-Visualisierung, BIM / GIS / Smart Cities, Punktwolken* sowie *Mobilität und GIS*.

Traditionell war eines der Wissenschaftsforen den Finalisten des *Münchener Förderpreises Geoinformatik* gewidmet, sodass wir Ihnen auch Kurzfassungen der ausgewählten Abschlussarbeiten und Dissertationen in diesem Tagungsband präsentieren können.

Wir wünschen Ihnen eine erkenntnisreiche Lektüre und hoffen, dass Sie zahlreiche Anknüpfungspunkte und Ideen für Ihre eigenen Arbeiten in den Beiträgen zur *Münchener GI-Runde* finden.

Unser Dank gilt *Frau Dr. Gabriele Aumann* für die Koordination und Erstellung dieses Tagungsbands, allen Autorinnen und Autoren, dem Gutachterteam des *Münchener Förderpreises Geoinformatik*, den Leitern der *Kompetenzpools* am *Runden Tisch GIS e.V.*, die für die Programmgestaltung der Praxisforen verantwortlich waren und nicht zuletzt den Firmenausstellern sowie allen Teilnehmerinnen und Teilnehmern der *Münchener GI-Runde 2019*.

München, im März 2019

Thomas H. Kolbe, Ralf Bill und Andreas Donaubaue

# Demonstrating Utility Network Interdependency Modelling Using the UtilityNetwork Application Domain Extension for CityGML

M.Sc. Isaac BOATES<sup>1,2</sup>

Supervisor: Prof Dr. Detlef Günther-Diringer<sup>1</sup>

Advisors: M.Sc. Alexandru NICHERSU, Dr. Giorgio AGUGIARO<sup>3</sup>

<sup>1</sup>Hochschule Karlsruhe Technik und Wirtschaft · Hoffstraße 3 · 76131 Karlsruhe, Germany

<sup>2</sup>European Institute for Energy Research · Emmy-Noether-Straße 11 · 76131 Karlsruhe, Germany

<sup>3</sup>Technische Universität Delft · Mekelweg 5 · 2628 CD Delft, Netherlands

E-Mail: Isaac.Boates@eifer.uni-karlsruhe.de

## 1 Introduction

### 1.1 CityGML and the UtilityNetwork ADE

The City Geography Markup Language (CityGML) data model is an open data model and eXtensible Markup Language (XML)-based format for the storage and exchange of virtual 3D city models (OPEN GEOSPATIAL CONSORTIUM 2019). It is a well-established data format that has been used in a multitude of applications for data storage and distribution (HERREHUELA, NAGEL, & KOLBE 2012), visualisation (KOLBE, BURGER, & CANTZLER 2015), as well as for scientific research on topics such as building solar irradiance estimation (SIMONS 2014) (PILI, DESOGUS, & MELIS 2018) (QUAN, LI, AUGENBROE, & YANG 2015) and energy demand modelling (EGUSQUIZA, PRIETO, IZKARA, & BÉJAR 2018) (EICKER, ZIRAK, BARTKE, RODRÍGUEZ, & COORS 2018) (AGUGIARO 2016a) (AGUGIARO 2016b).

The core data model of CityGML, however, lacks the capability to model utility networks and their constituent elements. A number of alternative data models already exist, such as the Infrastructure for Spatial Information in Europe (INSPIRE) utility network data model (EUROPEAN COMMISSION 2019), the Industry Foundation Classes (IFC) (buildingSMART, 2019), the ArcGIS utility network data model (ESRI 2019) and the Source for Environmental Data Representation (SEDRIS) (SEDRIS 2019). These data models have been assessed as being at least partially deficient in scientific literature (BECKER, NAGEL, & KOLBE 2013) (KUTZNER & KOLBE 2016). **Fehler! Verweisquelle konnte nicht gefunden wer-**

	INSPIRE Utility Networks	IFC	ArcGIS Utility Networks	SEDRIS	Pipeline ML	CityGML UtilityNetwork ADE
Representation of heterogeneous networks	+	•	•	+	•	++
Dual representation	+	++	+	++	-	++
Topographic/graphic aspects	++	++	++	++	++	++
3D geometries	-	++	-	+	-	+
Functional aspects	-	-	-	-	-	•
Hierarchical modelling						
• networks/ subnetworks	++	-	-	++	-	++
• components/ subcomponents	++	++	•	-	•	++
Interdependencies between						
• network features and city objects	-	•	-	•	-	++
• network features of different network types	-	++	-	-	-	++

- = no support, • = basic support, + = sophisticated support, ++ = comprehensive support

**Fig. 1:**

Assessment of available data models for utility networks, from (KUTZNER & KOLBE 2016).

den. shows the results of one such assessment.

The UtilityNetwork Application Domain Extension (UtilityNetwork ADE) scored highly in this assessment, due it having been developed as a response to the various shortcomings of the aforementioned alternatives. It is a modular extension, based upon the “Application Domain Extension” (ADE) mechanism of the CityGML data model (BILJECKI, KUMAR, & NAGEL 2018). It extends the standard schema by adding new data element definitions, as well as adding new property definitions to standard CityGML data elements which are specific to the utility network domain.

A mapping of the CityGML data model to a relational database management system (RDBMS) called “the 3-Dimensional City Database” (3DCityDB) also exists. This mapping allows for leveraging the computational efficiency and transactional security of an RDBMS while maintaining the familiar structure of the CityGML data model (TECHNISCHE UNIVERSITÄT MÜNCHEN 2019). This database schema can be extended to facilitate use of ADEs by adding new columns to existing tables and new tables with associated table relationships and constraints. Extensions for some ADEs, including

the UtilityNetwork ADE have been developed and can be installed on existing instances of the 3DCityDB (AGUGIARO 2019).

## 1.2 Previous and Related Work

While still relatively nascent, the UtilityNetwork ADE data model has been featured in recent publications and work. The first published data sample using the UtilityNetwork ADE was the conversion of an open source freshwater pipe network dataset provided by an open data catalogue (CITY OF NANAIMO 2019a) from the city of Nanaimo, British Columbia, Canada from ESRI Shapefile format using SAFE Software’s “Feature Manipulation Engine” (FME) software (BOATES 2019b) (BOATES, NICHERSU, & KUTZNER 2018). Further work using this data sample was conducted by importing it into an instance of the 3DCityDB extended with the UtilityNetwork ADE, as well as performing network routing analysis using the topological routing extension for PostgreSQL “pgRouting” (BOATES, AGUGIARO, & NICHERSU 2018).

Another data sample was produced using utility network data from Rotterdam, The Netherlands. This sample was also produced using FME and was imported into a UtilityNetwork ADE-extended 3DCityDB instance for use in topological routing with pgRouting (DEN DUIJN 2018a) (DEN DUIJN 2018b) (DEN DUIJN 2019) (DEN DUIJN, AGUGIARO, & ZLATANOVA 2018).

A Python application programming interface (API) was developed that is able to connect to an instance of a UtilityNetwork ADE-extended 3DCityDB instance (BOATES 2019a). Using this API, atomic read and write operations can be performed on the database in order to retrieve or set property values of network elements. pgRouting-based topological routing operations from the API can also be used to determine if any network element is able to be reached from another, and via which shortest path. In addition to performing routing operations, a filter can be applied to exclude specific elements from the topological network which are considered broken or otherwise out of service.

## 2 Methodology

### 2.1 Research Goals

It has been claimed that the UtilityNetwork ADE offers a novel means of modelling virtually any number of networks and their relationships and interdependencies, at virtually any scale (KUTZNER & KOLBE 2016) (BECKER, NAGEL, & KOLBE 2011) (BECKER, NAGEL, & KOLBE 2013). As a consequence of that same novelty, however, there are not many examples of the ADE in use, and virtually none showing the more advanced capabilities, such as modelling of heterogeneous networks and their relationships and hierarchies, explicit difference between topographical and topological representations, or modelling cascading effects on one network due to an interruption in another. Attempting to test every feature of the UtilityNetwork ADE is a monumental endeavour, and so the scope of this thesis work was constrained to the following goals, which were intended to assess the robustness of the UtilityNetwork ADE in contexts based upon the three tiers of a computer software system:

1. To create a data sample that models two heterogeneous networks and a dependency between them in both CityGML file format and in an instance of the 3DCityDB (data tier).
2. To develop an agent-based model capable of operating on this multi-network sample in order to model the status and elements in both networks over a predefined time period. Elements will adjust their behaviour based on the status of other elements in the networks in such a way that changes in one network can cascade to another (logic tier).
3. To create a dual representation of the multi-network by deriving an alternate representation of the geometry of the multi-network features to serve as a topological diagram of network connectivity (presentation tier).

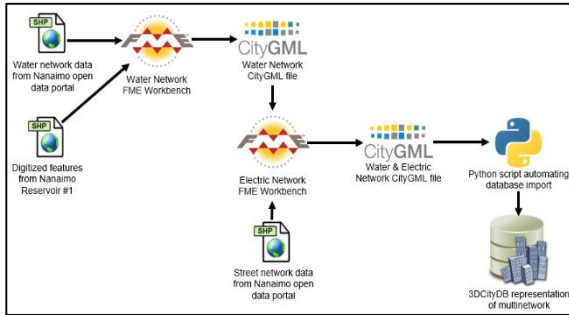
### 2.2 Source Data

This thesis expanded upon the previous work with the Nanaimo freshwater pipe network (BOATES 2019b) (BOATES, NICHERSU, & KUTZNER 2018). In addition to this data, a local Nanaimo project named “Nanaimo Reservoir No. 1” was found to be a sufficient real-world example of a coupling between a water network and an electrical network. In this project, a facility was constructed which harnesses the water pressure from the city’s freshwater delivery network to generate extra electricity for the local electrical grid (CITY OF NANAIMO 2019b). Since the water flow is variable, the electrical output of the facility is also variable, dependent at all times on the status of the water network. Technical drawings of the facility were digitized to supplement the water network data (CITY OF NANAIMO 2012), and actual facility water usage data was acquired from the city’s water resources department, along with other technical information about how the facility operates (BOATES 2018).

An electrical network data sample was not able to be acquired, due to it being property of a private corporation. Instead, the city’s road network centreline dataset was supplemented with extra connections to the city’s buildings to act as a surrogate electrical network dataset. Furthermore, a community initiative called “The Nanaimo Foodshare” has been collecting data on solar power generation potential in the city for several years (ENERGY SOLUTIONS FOR VANCOUVER ISLAND 2019), and the data was shared for use in the thesis work. Random buildings in the study area were selected to be

assumed as being equipped with solar panels to include an additional effect of solar power generation in the electrical network.

### 2.3 Building Multi-network Sample



**Fig. 2:**

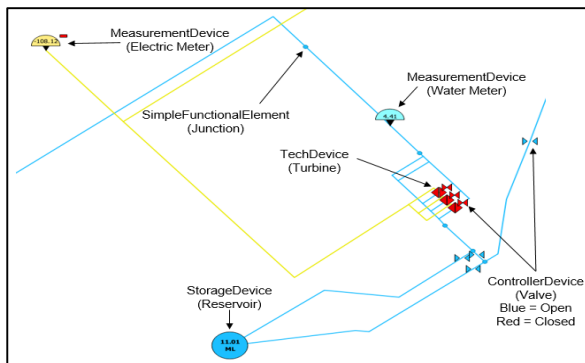
Process of creating the Nanaimo multi-network dataset from source data.

As depicted in **Fehler! Verweisquelle konnte nicht gefunden werden.**, FME was used to process and convert the aforementioned data source data from ESRI Shapefile format into a CityGML UtilityNetwork ADE file. With consultation of the UtilityNetwork ADE Unified Modelling Language (UML) diagrams (KUTZNER 2019) a single CityGML file containing the water and electrical multi-network was produced. It contained all the network features in a sample area encompassing the Reservoir No. 1 facility, as well as a nearby suburb containing houses to act as sources of water and electricity demand. It also contained elements from the “Supply” family of elements from the UtilityNetwork ADE data model schema, which are used to model properties such as medium flow rate and medium storage.

A custom Python script was then developed to read the resultant CityGML file and import all the features into an instance of the 3DCityDB that had been extended with the UtilityNetwork ADE schema. It made use of the “smart insert functions” that were developed by the UtilityNetwork ADE 3DCityDB extension (AGUGIARO 2019). These functions allow individual CityGML elements to be imported into a 3DCityDB instance with a single SQL statement, automatically taking care of validating the integrity of the resultant data with respect to the database constraints.

### 2.4 Developing Behavioural Model

Using the API described in section 1.2, a functional model was developed which models important elements in the Nanaimo multi-network as agents with behaviour that is driven either by external data, or by the status of other network elements and the topological connectivity between them. A 24-hour cycle is played out in which all the agents change their properties based on the aforementioned driving factors. **Fehler! Verweisquelle konnte nicht gefunden werden.** shows the layout of multi-network at the Nanaimo Reservoir No. 1 facility, where most of the following (simplified) procedure occurs during every half-hour time step in a 24 hour cycle:



**Fig. 3:**

Nanaimo multi-network at the Reservoir No. 1.

1. Water flow volume is read from an external file and set as the flow rate property of a water meter just before the facility.
2. As the meter’s water flow property increases, valve elements inside the facility change their status property from being closed to open, allowing turbine elements to become topologically connected to the water meter.
3. Any turbine elements which have a topological connection to the water meter change their electrical generation property to a fixed amount, others which such a lack topological connection change it to zero.
4. All connected buildings read a water demand and electrical demand value from an external file, but only if they have a topological connection to the reservoir element and the electrical meter element, respectively.
5. All solar panels affixed to the houses read an electrical production value from an external file and use it to offset the building’s electrical demand, or change it to an electrical surplus value if the generated electricity is greater than the house’s demand



6. If the total water volume input is greater than the total water demand, the fill level property of the reservoir element is increased accordingly. It is reduced accordingly if the total water demand is greater than the total volume input.
7. The total building electrical demand is subtracted from the production by the turbine elements and the solar panel elements, and the difference is used to set the electrical meter's flow rate property.

Additionally, it is possible to specify a “breakage schedule”, wherein network elements are indicated as having a time step at which they become “broken” and another at which they become “fixed”. Between these time steps, these elements are removed from all topological routing operations, simulating a service interruption in the network.

## 2.5 Dual Representation

The UtilityNetwork ADE's explicit difference between network features' topographical and topological representation was leveraged to create a dual representation of the Nanaimo multi-network. FME was used to export the network features to Esri Shapefile format, where they then had their geometry adjusted to be aligned to a regularly spaced grid. Features with high semantical importance (e.g. valves, reservoirs, meters, etc.) were placed in obvious locations with their topological connections clearly drawn and sparse areas were deflated. The manually-adjusted network features were then processed again by FME to assign their geometry as the source features' respective topological feature's geometry.

## 3 Results

### 3.1 Multi-network Sample

A CityGML UtilityNetwork ADE data sample was produced which contains all the properties and information from the source ESRI Shapefile data. It also contains a definition of the topological connections between elements in the networks and a means for modelling the dynamic state of network elements, such as flow rate and storage level. This file is easy to exchange and can be validated against the schema definition for the UtilityNetwork ADE.

The same dataset was also converted into a 3DCityDB. This format allows for display of network features as layers in GIS software and is validated by a series of defined database relationships and constraints. Furthermore, all the information present in the original CityGML file is still present in this format, but it is more cumbersome to facilitate exchange, requiring database dumps or a web-based feature service.

### 3.2 Behavioural model

The developed behavioural model shows the dynamic nature of electrical production and demand within a small closed network in which the electrical generation is dependent upon another network. The amount of electricity produced is partially dependent on the activation of turbines located at the power generation facility, these turbines are dependent on their access to an incoming water source, and the valves that control the water flow to the turbines are in turn dependent on the incoming water flow volume. It is therefore accurate to say that the electrical network is dependent on the status and behaviour of elements in the water network, and a network interdependency has been modelled using the UtilityNetwork ADE. Furthermore, by modifying the properties of specific elements in either network, they can be excluded from topological routing operations and simulate a service interruption which can affect the overall dynamics of the multi-network system. For example, marking an electrical cable element as being out of service will break the topological connection of all buildings beyond it, causing less of an overall demand for electricity as a result. These effects can cascade from one network to another, as in the case of a water pipe being marked as out of service which is critical for water flow to reach a turbine element. If this pipe was necessary for feeding a turbine element with water, a visible drop in electrical output is seen in the electrical production curve over the duration of a model run. This effect is depicted in **Fehler! Verweisquelle konnte nicht gefunden werden.**, wherein the blue line represents a model run with no such water pipe breakage, and the yellow line represents a model run with such a breakage.

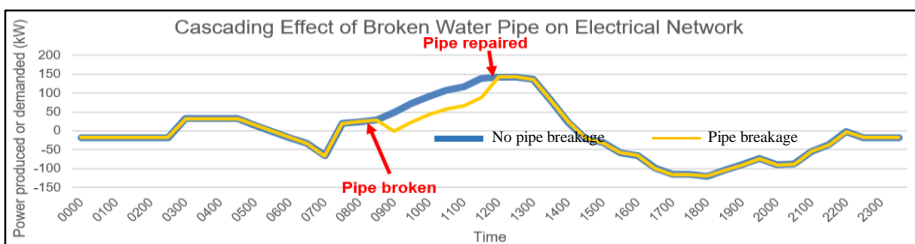


Fig. 4: Detectable changes in the operation of a network due to an interruption in another.

### 3.3 Dual Representation

The derived topological representation of the Nanaimo multi-network is able to communicate to the viewer the functional connectivity of elements while still maintaining the original geometry of individual network features. Since both the topographical and topological representation are inherently just properties of the network features themselves, they can be used as the geometry with which to drive a layer in a GIS software. **Fehler! Verweisquelle konnte nicht gefunden werden.** shows the two representations of the study area in QGIS. The topographical representation communicates the real-world position and extent of network features but it is difficult to understand the functional workings without gratuitous use of inset maps. The topological representation ignores spatial accuracy in favour of depicting an easily under-

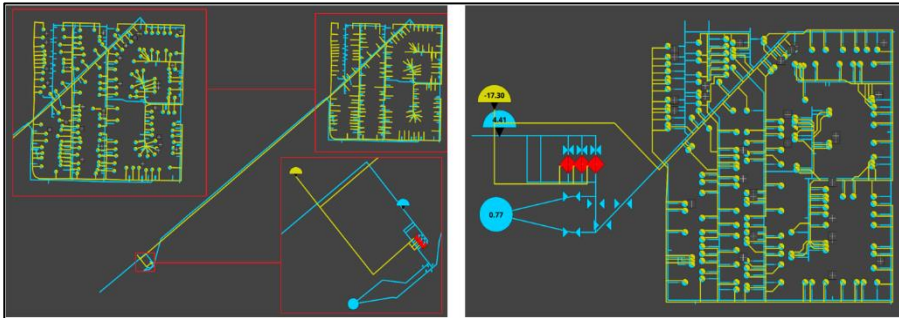


Fig. 5: Dual representation of the Nanaimo multi-network, topographical (left) and topological (right).

standable diagram of connectivity of network elements.

## 4 Conclusion

This thesis work intended to address the robustness of the UtilityNetwork ADE for CityGML by assessing it in the context of the three tiers of a computer software system with a real-world use case of a coupled water and electrical network. In the “data tier” context, it has been shown that the UtilityNetwork ADE is capable of modelling a real world multi-network system without losing significant source information in both a file-based and RDBMS-based format. With regard to the “logic tier context”, it has been shown that one network can be modelled as having inherent dependencies on the status and activity in another, and that an interruption in the normal activity in one network can cause detectable differences in the behaviour in another. Finally, in the “presentation tier” context, it has been shown that the UtilityNetwork ADE offers a novel method of explicit topographical and topological definition in the same dataset, which lends itself to effective presentation of network features for the purposes of real-world accuracy or for understanding of functional relationships between network elements. This supports the conclusion that the UtilityNetwork ADE is a robust solution for modelling utility networks which contain dependencies between networks.

## 5 Bibliography

- AGUGIARO, G. (2016a), Energy planning tools and CityGML-based 3D virtual city models: experiences from Trento (Italy). *Applied Geomatics*, 41-56.
- AGUGIARO, G. (2016b), Enabling "Energy Awareness" in the Semantic 3D City Model of Vienna. 1st International Conference on Smart Data and Smart Cities. Split, Croatia: ISPRS.
- AGUGIARO, G. (2019), 3DCityDB ADE. Retrieved 01 31, 2019, from Github: <https://www.3dcitydb.org/3dcitydb/3dcitydbhomepage/>.
- BECKER, T., NAGEL, C., & KOLBE, T. (2011), Integrated 3D Modeling of Multi-utility Networks and Their Interdependencies for Critical Infrastructure Analysis. *Advances in 3D Geo-Information Sciences, Lecture Notes in Geoinformation and Cartography*, 1-12.
- BECKER, T., NAGEL, C., & KOLBE, T. (2013), Semantic 3D Modeling of Multi-Utility Networks in Cities for Analysis and 3D Visualization. *Progress and New Trends in 3D Geoinformation Sciences, Lecture Notes in Geoinformation and Cartography*, 41-62.
- BILJECKI, F., KUMAR, K., & NAGEL, C. (2018), CityGML Application Domain Extension (ADE): overview of developments. *Open Geospatial Data, Software and Standards*.
- BOATES, I. (2018), Demonstrating Utility Network Interdependency Modelling Using the UtilityNetwork Application Domain Extension for CityGML. Karlsruhe: Hochschule Karlsruhe.

- BOATES, I. (2019a), 3DCityDB UtilityNetwork ADE Manager Repository. Retrieved 01 31, 2019, from Github: <https://github.com/iboates/3dcitydb-utilitynetwork-ade-manager>.
- BOATES, I. (2019b), CityGML UtilityNetwork ADE Nanaimo Water Network Sample Repository. Retrieved 01 31, 2019, from Github: <https://github.com/iboates/CityGML-UtilityNetwork-ADE-Nanaimo-Water-Network-Sample>.
- BOATES, I., AGUGIARO, G., & NICHERSU, A. (2018), Network Modelling and Semantic 3D City Models: Testing the Maturity of the Utility Network ADE for CityGML with a Water Network Test Case. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, IV-4 (pp. 13-20). Delft: ISPRS.
- BOATES, I., NICHERSU, A., & KUTZNER, T. (2018), Nanaimo Water Pipes Utility Network ADE Sample. Karlsruhe: EIFER. Retrieved 01 31, 2019, from [https://en.wiki.utilitynetworks.sig3d.org/images/upload/2017-12-08\\_UtilityNetworkADE\\_Karlsruhe.pdf](https://en.wiki.utilitynetworks.sig3d.org/images/upload/2017-12-08_UtilityNetworkADE_Karlsruhe.pdf).
- BUILDINGSMART (2019), IFC Overview. Retrieved 01 31, 2019, from <http://www.buildingsmart-tech.org/specifications/ifc-overview>.
- CITY OF NANAIMO (2012), Reservoir No. 1 and Energy Recovery Facility Project. Retrieved 01 31, 2019, from <https://www.nanaimo.ca/docs/your-government/Projects/2012-reservoir-no-1/754ea6391b316d6b9fc9ff00001037d2.pdf>.
- CITY OF NANAIMO (2019a), Nanaimo Data Catalogue. Retrieved 01 31, 2019, from <https://www.nanaimo.ca/open-data-catalogue>.
- CITY OF NANAIMO (2019b), Reservoir No. 1 and Energy Recovery Facility. Retrieved 1 31, 2019, from <https://www.nanaimo.ca/your-government/projects/projects-detail/reservoir-no-1>.
- DEN DUIJN, X. (2018a), A 3D data modeling approach for integrated management of below and above ground utility network features. Delft: TU Delft.
- DEN DUIJN, X. (2018b), A 3D data modelling approach for integrated management of below and above ground utility network features. TU Delft. Retrieved 01 31, 2019, from [https://en.wiki.utilitynetworks.sig3d.org/images/upload/UtilityNetworkADE\\_Karlsruhe\\_XanderdenDuijn.pdf](https://en.wiki.utilitynetworks.sig3d.org/images/upload/UtilityNetworkADE_Karlsruhe_XanderdenDuijn.pdf).
- DEN DUIJN, X. (2019), Rotterdam UtilityNetwork ADE Sample Repository. Retrieved 01 31, 2018, from Github: <https://github.com/XanderdenDuijn/CityGML-Utility-Network-ADE>.
- DEN DUIJN, X., AGUGIARO, G., & ZLATANOVA, S. (2018), Modelling Below- and Above-Ground Utility Network Features with the CityGML Utility Network ADE: Experiences from Rotterdam. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, IV-4/W7 (pp. 43-50). Delft: ISPRS.
- EGUSQUIZA, A., PRIETO, I., IZKARA, L. I., & BÉJAR, R. (2018), Multi-scale urban data models for early-stage suitability assessment of energy conservation measures in historic urban areas. *Energy and Buildings*, 87-98.
- EICKER, U., ZIRAK, M., BARTKE, N., RODRÍGUEZ, L. R., & COORS, V. (2018), New 3D model based urban energy simulation for climate protection concepts. *Energy and Buildings*, 79-91.
- ENERGY SOLUTIONS FOR VANCOUVER ISLAND (2019), Nanaimo Foodshare Solar Power Production. Retrieved 01 31, 2019, from Enlighten: [https://enlighten.enphaseenergy.com/pv/public\\_systems/6eHN1176825/overview?preview=1](https://enlighten.enphaseenergy.com/pv/public_systems/6eHN1176825/overview?preview=1).
- ESRI (2019), What is a utility network? Retrieved 01 31, 2019, from <https://pro.arcgis.com/en/pro-app/help/data/utility-network/what-is-a-utility-network-.htm>.
- EUROPEAN COMMISSION. (2019), Utility and governmental services. Retrieved 01 31, 2019, from INSPIRE: <http://inspire.ec.europa.eu/theme/us>.
- HERREHUELA, J., NAGEL, C., & KOLBE, T. (2012), Value-added Services for 3D City Models using Cloud Computing. *Geoinformatik 2012 "Mobilität und Umwelt"*.
- KOLBE, T., BURGER, B., & CANTZLER, B. (2015), CityGML goes to Broadway. *Photogrammetric Week*, 343-355.
- KUTZNER, T. (2019), UtilityNetwork ADE UML Diagrams. Retrieved 01 31, 2019, from Github: [https://github.com/TatjanaKutzner/CityGML-UtilityNetwork-ADE/blob/master/UML/UML\\_diagrams\\_CityGML\\_UtilityNetwork\\_ADE.pdf](https://github.com/TatjanaKutzner/CityGML-UtilityNetwork-ADE/blob/master/UML/UML_diagrams_CityGML_UtilityNetwork_ADE.pdf).
- KUTZNER, T., & KOLBE, T. (2016), Extending Semantic 3D City Models by Supply and Disposal Networks for Analyzing the Urban Supply Situation. In T. Kersten, *Publikationen der Deutschen Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation e.V.* (pp. 382-394).
- OPEN GEOSPATIAL CONSORTIUM (2019), CityGML. Retrieved 01 31, 2019, from <https://www.opengeospatial.org/standards/citygml>.

- PILI, S., DESOGUS, G., & MELIS, D. (2018), A GIS tool for the calculation of solar irradiation on buildings at the urban scale, based on Italian standards. *Energy and Buildings*, 629-646.
- QUAN, J., LI, Q., AUGENBROE, G., & YANG, P.-J. P. (2015), A GIS-based Energy Balance Modeling System for Urban Solar Building. *Energy Procedia*, 2946-2952.
- SEDRIS (2019), SEDRIS Standards. Retrieved 01 31, 2019, from <https://standards.sedris.org/>.
- SIMONS, A. (2014), Development of a CityGML infrastructure for the implementation of an energy demand method with different data sources. Karlsruhe: Hochschule Karlsruhe.
- TECHNISCHE UNIVERSITÄT MÜNCHEN (2019), 3DCityDB. Retrieved 01 31, 2019, from <https://www.3dcitydb.org/3dcitydb/3dcitydbhomepage/>.

# **Straßenraummodellierung mittels Mobile Mapping in OpenDRIVE und CityGML sowie Entwicklung geeigneter Visualisierungsmethoden**

Theresa CODURO

E-Mail: th.coduro@googlemail.com

## **1 Einführung ins Thema**

Dieser Vortrag beinhaltet Ergebnisse der gleichnamigen Masterarbeit, in der Methoden zur Überführung von Straßenraumdaten in ein semantisches 3D-Stadtmodell aufgezeigt werden. Die Methoden werden anhand einer Umwandlung von Mobile Mapping Daten, die für das OpenDRIVE-Format aufbereitet wurden, in das CityGML-Format exemplarisch untersucht.

Neben verwandten Arbeiten wird insbesondere auch auf Daten eingegangen, die für die Straßenraummodellierung verwendet werden können. Die Beurteilung der aufgezeigten Datenquellen erfolgt anhand von neun Kriterien und legt die Kombination unterschiedlicher Datenquellen als bestmögliche Grundlage zur Erzeugung von Straßenraummodellen nahe. Zusätzlich wird eine Übersicht über unterschiedliche Modellierungsstandards aus dem Bereich der Straßenraummodellierung gegeben. Hauptsächlich wird vergleichend auf die Standards OpenDRIVE und CityGML eingegangen. Aufbauend auf dem Datenerfassungsprozess des beteiligten Unternehmens 3D Mapping Solutions, welches sich mit der kinematischen Aufnahme von Mobile Mapping Daten befasst, wird eine Strategie für eine mögliche Visualisierung von Straßenraumdaten vorgestellt.

Das Ergebnis der praktischen Umsetzung ist ein semantisches Straßenraummodell im Level of Detail 3, welches neben einem texturierten Spurmodell zusätzlich verschiedene Straßenraumobjekte beinhaltet. Die abschließende Visualisierung erfolgte im, auf Cesium basierenden, 3D-Web-Map-Client der 3DCityDB.

## **2 Zusammenfassung des Vortragsinhalts**

3D-Stadtmodelle dienen heutzutage als Informations- und Analysegrundlage für unterschiedliche Aufgaben. Der umfassende Journal-Artikel von (BILJECKI et al. 2015) der TU Delft mit dem Titel „Applications of 3D City Models: State of the Art Review“ stellt eine umfangreiche Auflistung von 29 möglichen Anwendungsfeldern für 3D-Stadtmodelle vor. Die große Bedeutung von Straßenraummodellen wird auch in momentan sehr aktuellen Themen deutlich. Fahrassistenzsysteme und das autonome Fahren benötigen Straßenraummodelle als Grundlage. Die für solche Anwendungsbereiche erforderte Positionsgenauigkeit von wenigen Zentimetern kann vor allem in urbanen Gebieten nicht allein durch die Lokalisierung mittels GNSS erreicht werden. Mit zusätzlich eingesetzten Sensoren wie Stereokameras oder Laserscansystemen können Grundlagendaten gewonnen werden, durch die eine Positionierung in hochauflösenden Karten möglich ist (SCHREIBER et al. 2013, LEVINSON et al. 2007).

Die Bearbeitung der Masterarbeit erfolgte in Kooperation mit der Firma 3D Mapping Solutions GmbH. Das beteiligte Unternehmen hat sich auf die Kalibrierung von Multisensorsystemen und die kinematische Vermessung von Verkehrsnetzen spezialisiert. Der Schwerpunkt bei der Datenauswertung liegt dabei auf der Bereitstellung von Referenzinformationen und Referenzkarten von Straßennetzen im OpenDRIVE-Format.

Solche kinematisch erfassten hochgenauen Straßendaten können als Grundlage für genannte Anwendungsfälle eingesetzt werden. Dabei wird in den meisten Fällen eine flächenhafte, realitätsnahe Darstellung des Straßenraums vorausgesetzt. Mit der Darstellung von Straßenräumen hat sich auch (BEIL 2017) unter Verwendung des Standards CityGML beschäftigt. Dabei entstand ein CityGML-konformes anschauliches 3D-Straßenraummodell der Stadt New York City. Vor diesem Hintergrund wird in dieser Arbeit eine Strategie zur Visualisierung der Mobile Mapping Daten in CityGML erarbeitet und praktisch umgesetzt.

### **2.1 Theoretische Erkenntnisse**

Als Grundlage für die praktische Umsetzung wurden mit einer Studie relevanter den Straßenraum betreffender Arbeiten verschiedene Anwendungsbereiche, Methoden der Datenextraktion und Datenquellen im Bereich der Straßenraummodellierung aufgezeigt. Anhand verschiedener Kriterien wie Verfügbarkeit, Vollständigkeit und semantischer Informationsgehalt geht hervor, dass in den meisten Fällen eine Kombination verschiedener Datenquellen, wie sie bei Mobilien Mapping Systemen eingesetzt werden, verwendet werden. Dies bringt eine Verknüpfung verschiedenster Daten mit sich und führt zu hochgenauen und semantisch detailreichen Ergebnissen.

Insbesondere wird zudem vergleichend auf die Modellierungsstandards OpenDRIVE und CityGML eingegangen. Obwohl die XML-basierten Standards grundsätzlich beide für die Modellierung des Straßenraums verwendet werden können, weisen die Datenmodelle große Unterschiede auf. Grundlegender Unterschied ist die geometrische Beschreibung der Objekte. Während diese in OpenDRIVE mittels linearer Referenzierung und aufwendigen mathematischen Funktionen erfolgt, werden Geometrieobjekte in CityGML flächenhaft mit Hilfe absoluten Koordinaten und Primitiven beschrieben. Der Vergleich der beiden Datenmodelle in Inhalt, Struktur, Geometrie, Semantik, Topologie und visueller Erscheinung verdeutlicht die Unterschiede, zeigt aber auch dass eine Abbildung von OpenDRIVE nach CityGML grundsätzlich mit einigen Abwandlungen möglich ist.

## 2.2 Praktische Umsetzung

Aus diesen Erkenntnissen werden Methoden zur Überführung von Straßenraumdaten in ein semantisches 3D-Stadtmodell erarbeitet. Die Methoden werden anhand einer Umwandlung von Mobile Mapping Daten, die für das OpenDRIVE-Format aufbereitet wurden, in das CityGML-Format exemplarisch untersucht. Für die Durchführung von Experimenten wurden Datensätze durch das beteiligte Unternehmen 3D Mapping Solutions GmbH zur Verfügung gestellt. Aufbauend auf den Prozess der Straßenraummodellierung bei 3D Mapping Solutions und den erarbeiteten Grundlagen der beiden Modellierungsstandards wird eine Strategie für eine mögliche Visualisierung der kinematisch erfassten Vermessungsdaten des Unternehmens vorgestellt.

Im Zuge der Realisierung des entwickelten Ansatzes konnte anhand konkreter Testgebiete eine dreidimensionale Visualisierung der kinematisch erfassten Straßenraumdaten ermöglicht. Aus diesen Spur- und Objektdaten konnte ein virtuelles semantisches CityGML-konformes und texturiertes Straßenraummodell im Level of Detail 3 erzeugt werden, welches unter anderem aus Fahrspuren, Grünflächen, Parkflächen, Verkehrszeichen und Bäumen besteht (vgl. Abb. 1).



**Abb. 1:** Darstellung der Mobile Mapping Daten als texturiertes Straßenraummodell.

Die abschließende Visualisierung im, auf Cesium basierenden, Web-Map-Client der 3DCityDB ermöglicht die Verknüpfung von detaillierten, hochgenauen Straßendaten und einem virtuellen 3D-Gebäudemodell. Die Straßenraumdaten gewinnen durch die realitätsnahe Präsentation deutlich an Anschaulichkeit. Über die Visualisierung hinausgehend sind mithilfe des Clients Abfragen über einzelne Objekte des Modells möglich. Ein detailliertes, räumlich-semantisches Modell des Straßenraums eröffnet zahlreiche weitere Anwendungen jenseits der Fahrsimulation und Navigation und ist für die Präsentation der kinematisch aufgenommenen Vermessungsdaten von Vorteil.

## Literatur

BEIL, C. (2017), Detaillierte Repräsentation des Straßenraums in 3D-Stadtmodellen. Masterarbeit, Technische Universität München.

- BILJECKI, F., STOTER, J., LEDOUX, H., ZLATANOVA, S. & CÖLTEKIN, A. (2015), Applications of 3D city models: State of the art review. *ISPRS International Journal of Geo-Information*, 4(4):2842–2889.
- SCHREIBER, M., KNOEPEL, C. & FRANKE, U. (JUNE 2013), LaneLoc: Lane marking based localization using highly accurate maps. In *2013 IEEE Intelligent Vehicles Symposium (IV)*, pages 449–454.
- LEVINSON, J., MONTEMERLO, M. & THRUN, S. (2007), Map-based precision vehicle localization in urban environments. *Robotics Science and Systems*.

# Einflussfaktoren auf die Lokalisationsgüte von Passive Audio Monitoring Systems (PAMS)

Florian HOEDT

Zentrum Informationsmanagement · Thünen Institut · Bundesallee 44 · 38116 Braunschweig  
E-Mail: florian.hoedt@thuenen.de

## 1 Einführung

Weltweit zeigen sich Veränderungen der Vogelpopulationen mit zum Teil starken Abnahmen (HAYHOW et al. 2016). Als Folgen des Landnutzungs- und Klimawandel wird erwartet, dass die Vogelpopulationen auch weiterhin starken Veränderungen in ihrer Anzahl als auch Verteilung unterliegen (JOHNSTON et al. 2013, BAIRLEIN 1996). Entsprechend steigt der Bedarf Ökosysteme zu überwachen. Bisher werden diese Aufnahmen manuell und Vorort durchgeführt, was vor allem in schwer erreichbaren Gebieten mit hohem Aufwand und Störungen des Habitats verbunden ist. Um den zukünftig steigenden Bedarf an Umweltbeobachtung zu realisieren, können Methoden der Bioakustik mit autonomen Sensor-Systemen angewandt werden (SETHI et al. 2018, ALQUEZAR & MACHADO 2015, FROMMOLT 2017, FRISTRUP & MENITT 2012). Solche *Passive Audio Monitoring Systems* (PAMS) ermöglichen die Wiederholbarkeit der Analysen durch aufgezeichnete Soundsamples und eliminieren den Einfluss des Beobachters (AIDE et al. 2013). Jedoch können Vogeldichten oftmals nur unzureichend ermittelt werden (SHONFIELD & BAYNE 2017).

Eine vielversprechende Methode diese zu ermitteln ist die Lokalisation der jeweilig detektierten Vogellaute. Dabei stellt sich die Frage, inwiefern die Schallverschattung durch Vegetation und andere Objekte die Güte der Lokalisation beeinflusst. Bisherige Untersuchungen unter tropischem Regenwald von (MENNILL et al. 2006) zeigen keine signifikanten Auswirkungen. Des Weiteren konnten (LAMBERT & MCDONALD 2014) keine signifikanten Unterschiede in der Abnahme der Lautstärke zwischen den untersuchten Vegetationstypen ermitteln. Der dort dominante Unterwuchs bestand jedoch aus Stauden (*Lantana camara*) und es wurde lediglich die Lautstärkeabnahme, nicht jedoch der Klassifizierungs- oder Lokalisierungserfolg, betrachtet.

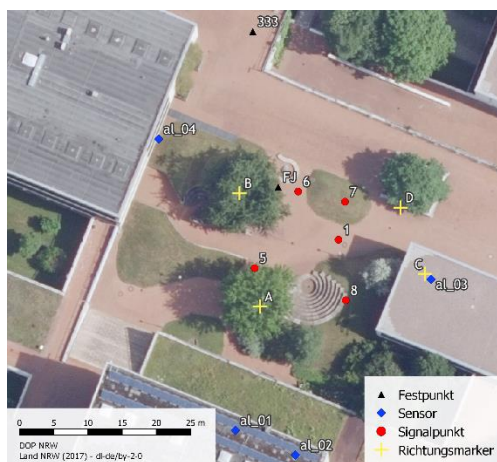
Daher verbleibt die Frage, welchen Einfluss mitteleuropäische Vegetation verschiedener Ausprägung auf die Klassifikation und Lokalisation hat. Daher wird folgende Hypothese geprüft:

Die Güte der Lokalisation ist abhängig von der Schallverschattung zwischen Schallgeber und Mikrofon.

## 2 Zusammenfassung des Vortragsinhalts

### 2.1 Methodik

Als Untersuchungsgebiet ist der Innenhof des Campus der Hochschule Ostwestfalen-Lippe (HS OWL) am Standort Höxter definiert (vgl. Abb. 1). Zwischen den Signalpositionen und den Sensoren des verwendeten Systems befinden sich verschiedene Bäume und Sträucher, welche schallverschattend wirken.



**Abb. 1:**

Karte des Versuchsaufbaus auf dem Innenhof der Hochschule Ostwestfalen-Lippe am Standort Höxter.

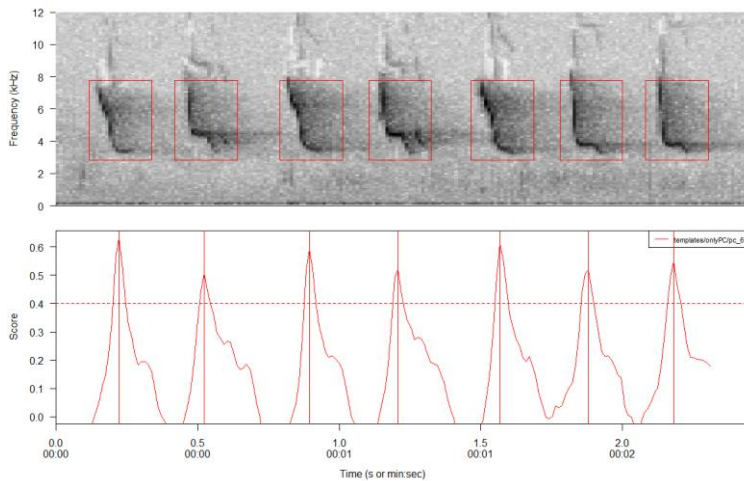
Die Positionen der Mikrofone wurden per Tachymeter (*Leica Builder 500, Leica Camera AG, Wetzlar, DE*) eingemessen. Im Innenhof wurde ein Lautsprecher positioniert sowie dessen Wiedergabeposition aufgenommen. Das eingesetzte



PAMS wurde aus vier *Raspberry Pi 3B* (*Farnell element14, Leeds, UK*) Microcomputern sowie günstigen *Foxnovo Portable USB 2.0-Kondensator-Mikrofonen* (*Typ SF-555B*) aufgebaut. Das Sensor-System verwendet *Debian Stretch* und eine in Python erstellte Aufnahmelogik. Über das *Network Time Protocol* (MILLS 1991) sind die Sensorsysteme synchronisiert und die Aufnahmen mit Zeitstempel markiert worden. Die verwendete Abtastrate der Aufnahme wurde auf 44.1 kHz gesetzt, wie für Aufnahmen von Vogellauten verwendet und empfohlen (SETHI et al. 2018, BROWNING et al. 2017, FROMMOLT & TAUCHERT 2014, AIDE et al. 2013).

Für die Klassifikation der aufgenommenen Daten wurde der *spectrogramm cross correlation* (*corMatch*) Algorithmus des R package *monitoR* (KATZ et al. 2016) genutzt. Dieser Klassifikator erstellt aus Trainingsdatensätzen eine Spektrogramm Vorlage. Diese wird in der Kreuzkorrelation verwendet, um den gesuchten Laut zu detektieren.

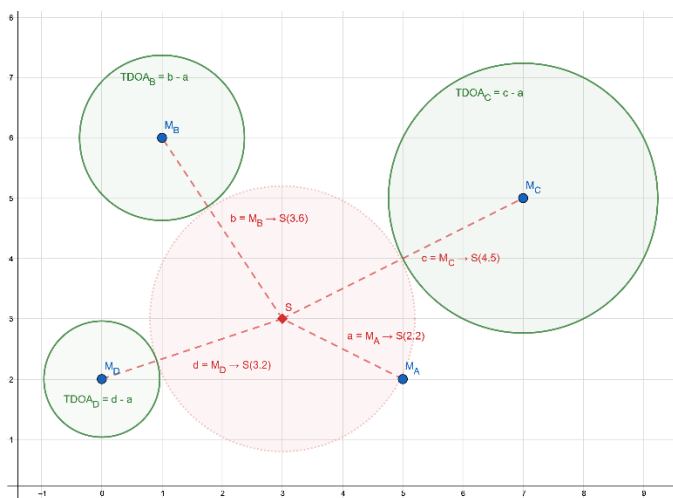
Die so erstellten Vorlagen werden über *monitoR* in den aufgenommenen Audiodateien zeitlich verortet, wie in Abb. 2 beispielhaft gezeigt. Die Detektionen wurden je Sensor und Art als CSV-Datei, mit Angabe der Audiodatei, des Zeitstempels sowie des Kreuzkorellationsergebnisses gespeichert.



**Abb. 2:** Über *monitoR* erstellte Kreuzkorrelation für eine *Phylloscopus collybita* Vorlage über Gesang - Spektrogramm mit einer *Hanning Window* Länge von 512 pt und 0 % Überschneidung. Die rote gestrichelte Linie gibt das zur Detektion notwendige Kreuzkorrelationsergebnis (*score cut-off*) an und wurde auf 0.4 gesetzt.

Für die Berechnung der Schallverschattung werden die Daten des Laserscans NRW (LAND NRW 2017) verwendet. Dabei wird von der Schallgeberposition eine 3D-Röhre mit 0.5 m Durchmesser zu den jeweiligen Mikrofonen konstruiert und ein 3D-Intersect mit der Punktwolke des Laserscans durchgeführt. Hierbei entspricht die Anzahl der überschrittenen Punkte dem Maß der Schallverschattung.

Der verwendete Lokalisierungsalgorithmus betrachtet die Positionsfindung als Optimierungsproblem. Der Algorithmus basiert auf dem *SciPy.minimize* Modul (THE SCIPY COMMUNITY 2017). Der Algorithmus streut zufällig Punkte ein und minimiert auf eine, im Folgenden beschriebene, Fitnessfunktion:



**Abb. 3:** Beispielaufbau des auf Standardabweichung ( $\sigma$ ) minimierenden Lokalisierungsalgorithmus. Gegeben seien die vier Sensoren  $M_a \dots M_d$  sowie die Signalposition  $S$ . Die grünen Umkreise entsprechen den berechneten TDOA.

Gegeben seien die Mikrofonpositionen  $M_n$  sowie die Signalposition  $S$ . Hierbei ist die Mikrofonposition  $M_a$  dem Signal  $S$  am nächsten, wie beispielhaft in Abb. 3 dargestellt. Für die Mikrofonpositionen  $M_{n'} \rightarrow n''$  können somit, wie in Gleichung 1 beschrieben, die Time Difference of Arrival (TDOA) berechnet werden.

Gleichung 1:

$$TDOA_{M_n} = \overline{S M_n} - \overline{S M_A}$$

Für die Lokalisation des Signalpunktes  $SP_{[x,y]}$  kann die Standardabweichung der TDOA zu den jeweiligen Mikrofonpositionen verwendet werden (vgl. Abb. 3). Hieraus ergibt sich die in Gleichung 2 gezeigte Fitness Funktion  $SP_{[x,y]fit}$ .

Gleichung 2:

$$SP_{[x,y]fit} = \sigma \left\| \begin{array}{c} \overline{SP_{[x,y]} M_n} \\ \vdots \\ TDOA_{M_n} \end{array} \right\|$$

Für die gesuchte Signalposition SP gilt hierbei Gleichung 3.

Gleichung 3:

$$S_{[x,y]} \mapsto SP_{[x,y]fit} = 0$$

Die hier beschriebene Abhängigkeit der  $SP_{[x,y]fit}$  basierten Fitnessfunktion wurde über eine Simulation getestet und ausgewertet. Der Zusammenhang konnte mit einer *Pearson's product-moment correlation* von 0.8 sowie einem p-value von  $< 2.2 \cdot 10^{-16}$  nachgewiesen werden.

Die zur Berechnung der Schallgeschwindigkeit genutzte Formel ist in Gleichung 4 gezeigt (WILSON ET AL. 2014).

Gleichung 4:

$$v = 331.3 * \sqrt{1 + Temperatur/273.15}$$

Die Formel ist zur Berechnung der TDOA basierten Distanzdifferenzen (vgl. grüne Umkreise in Abb. 3) verwendet worden. Die hierbei notwendigen Temperaturen in °C wurden an der Wetterstation Höxter ( $51^\circ 46' 3.36''N$ ,  $9^\circ 22' 9.86''E$ , 156m über NN), in unmittelbarer Nähe des Aufnahmeortes erfasst.

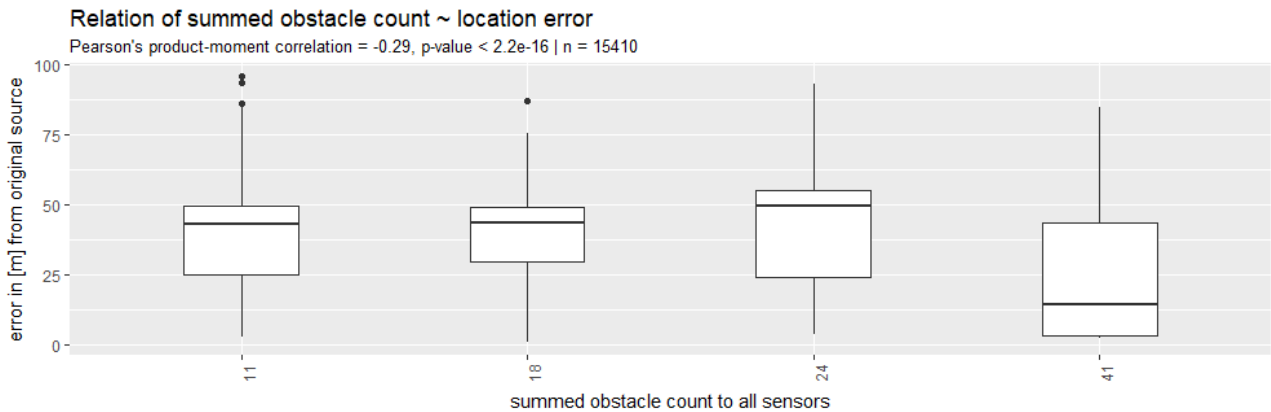
## 2.2 Ergebnisse

Das verwendete PAMS nimmt vierspürige Audiodaten auf. In jeder Spur werden Vogellaute detektiert und mit Zeitstempel versehen. Diese einzelnen Detektionen werden zu Audio-Ereignissen zusammengefasst. Ein Ereignis besteht aus mehreren miteinander zusammenhängenden Detektionen. Hierbei gilt ein Ereignis als valide, wenn:

1. alle vier Sensoren dasselbe Template,
2. innerhalb der festgelegten  $\Delta$ -Zeit detektiert haben.

Für eine erfolgreiche dreidimensionale Lokalisation mit der oben genannten Methode sind vier Detektionen für ein Ereignis notwendig. In den aufgenommenen Audiodateien wurden insgesamt 15.412 Ereignisse erfasst. Die Lokalisation zeigt Fehlerwerte von  $\bar{x} 34.96 \pm 19.49 m$  (min: 1.1, max: 95.8 m).

Die Hypothese, die Güte der Lokalisation ist abhängig von der Schallverschattung des Schallgebers zum Mikrofon, kann bestätigt werden. Es liegt eine signifikante negative Korrelation zwischen der Güte der Lokalisation und der Anzahl an schallverschattenden Objekten vor (vgl. Abb. 4).



**Abb. 4:** Zusammenhang zwischen Lokalisationsgenauigkeit und dem Maß an Schallverschattung. Es zeigt sich, mit einer Pearson's product-moment correlation von  $-0.29$  (p-value  $< 2.2 \cdot 10^{-16}$ ), eine signifikante negative Korrelation zwischen den untersuchten Größen.

## 2.3 Fazit

Die Lokalisierung der Vogellaute mit dem verwendeten PAMS ist sehr ungenau und verortet im Mittel die berechneten Punkte über den gesamten Versuchsaufbau. Hiermit ist das verwendete PAMS nicht für die Vogeldichteberechnung nutzbar.

Die in den erfassten Daten hohen Lokalisierungsfehler übersteigen die in WILSON ET AL. (2014) ( $\bar{x} = 4.3\text{ m}$ , 2D) und STEPANIAN ET AL. (2016) ( $\bar{x} < 10\text{ m}$ , 3D) genannten Werte um ein Vielfaches ( $\bar{x} = 34.96\text{ m}$ , 3D). Mögliche Ursachen für diese Ungenauigkeit sind die Zeitsynchronisation der Sensoren, der verwendete Klassifikations- und Lokalisationsalgorithmus, sowie der verwendete Schalldruck der Signale.

Die vorherrschende Vegetation des Untersuchungsgebietes entspricht nicht der typischen Vegetation von zu untersuchenden Ökosystemen (Wälder, Offenländer, Grünflächen). Zudem fand die Aufnahme nicht unter gleichen phänologischen Bedingungen wie der in der Arbeit verwendete Laserscan statt. Hierdurch ist die Rückgabeanzahl des Laserscans zu hoch, im Vergleich zum tatsächlich zur Aufnahme vorhandenen Blattwerk. Die Abfrage der Punktwolke hat einige Bodenpunkte miterfasst, welche nicht Schallverschattend wirken. Die genutzten Aufnahmepositionen haben letztlich eine Population von nur vier verschiedenen Datenpunkten an Schallverschattung ergeben und somit ist die verwendete Metrik für das Maß der Schallverschattung kritisch zu hinterfragen.

Es zeigt sich, dass die Verwendung von low-cost PAMS als Mittel zur Umweltüberwachung großes Potentiale besitzen, sie jedoch nicht ohne weiteres für die Vogeldichteberechnung einsatzbereit sind.

## Literatur

- AIDE, T. M., CORRADA-BRAVO, C., CAMPOS-CERQUEIRA, M., MILAN, C., VEGA, G., ALVAREZ, R. (2013), Real-time bioacoustics monitoring and automated species identification. In: PeerJ, 1: e103. URL <https://peerj.com/articles/103>.
- ALQUEZAR, R. D., & MACHADO, R. B. (2015), Comparisons Between Autonomous Acoustic Recordings and Avian Point Counts in Open Woodland Savanna. In: The Wilson Journal of Ornithology, 127(4), 712–723. <https://doi.org/10.1676/14-104.1>
- BAIRLEIN, F. (1996), Ökologie der Vögel: Physiologische Ökologie – Populationsbiologie – Vogelgemeinschaften - Naturschutz. 149 S. Gustav Fischer, Stuttgart.
- BROWNING, E., GIBB, R., GLOVER-KAPFER, P., JONES, K. (2017), Passive acoustic monitoring in ecology and conservation. URL [https://www.researchgate.net/publication/320323376\\_Passive\\_acoustic\\_monitoring\\_in\\_ecology\\_and\\_conservation](https://www.researchgate.net/publication/320323376_Passive_acoustic_monitoring_in_ecology_and_conservation), Zugriff 02/18.
- FRISTRUP, K. M., & MENNITT, D. (2012), Bioacoustical monitoring in terrestrial environments. In: Acoust. Today, 8(3), 16-24.
- FROMMOLT, K. H. (2017), Information obtained from long-term acoustic recordings: applying bioacoustic techniques for monitoring wetland birds during breeding season. In: Journal of Ornithology, 158(3), 659–668. <https://doi.org/10.1007/s10336-016-1426-3>.
- FROMMOLT, K. H., & TAUCHERT, K. H. (2014), Applying bioacoustic methods for long-term monitoring of a nocturnal wetland bird. In: Ecological Informatics, 21, 4–12. <http://doi.org/10.1016/j.ecoinf.2013.12.009>.
- HAYHOW D.B., BURNS F., EATON M.A., AL FULAIJ N., AUGUST T.A., BABEY L., BACON L., BINGHAM C., BOSWELL J., BOUGHEY K.L., BRERETON T., BROOKMAN E., BROOKS D.R., BULLOCK D.J., BURKE O., COLLIS M., CORBET L., CORNISH N., DE MASSIMI S., DENSHAM J., DUNN E., ELLIOTT S., GENT T., GODBER J., HAMILTON S., HAVERY S., HAWKINS S., HENNEY J., HOLMES K., HUTCHINSON N., ISAAC N.J.B., JOHNS D., MACADAM C.R., MATHEWS F., NICOLET P., NOBLE D.G., OUTHWAITE C.L., POWNEY G.D., RICHARDSON P., ROY D.B., SIMS D., SMART S., STEVENSON K., STROUD R.A., WALKER K.J., WEBB J.R., WEBB T.J., WYNDE R. AND GREGORY R.D. (2016), The State of Nature in the UK and its Overseas Territories. The State of Nature partnership.
- JOHNSTON, A., AUSDEN, M., DODD, A. M., BRADBURY, R. B., CHAMBERLAIN, D. E., JIGUET, F., ... PEARCE-HIGGINS, J. W. (2013), Observed and predicted effects of climate change on species abundance in protected areas. In: Nature Climate Change, 3(12), 1055–1061. <https://doi.org/10.1038/nclimate2035>.
- KATZ, J., HAFNER, S. D., DONOVAN, T. (2016), Tools for automated acoustic monitoring within the R package monitoR. In: Bioacoustics, 25, 2: 197–210. URL <http://www.tandfonline.com/doi/full/10.1080/09524622.2016.1138415>.
- LAMBERT, K. T. A., McDONALD, P. G. (2014), A low-cost, yet simple and highly repeatable system for acoustically surveying cryptic species. In: Austral Ecology, 39, 7:779–785. URL <http://onlinelibrary.wiley.com/doi/10.1111/aec.12143/abstract>.

- LAND NRW (2017), Digitales Oberflächenmodell mittlerer Punktabstand 1m. URL <https://www.opengeo-data.nrw.de/produkte/geobasis/dom/dom11/>, Zugriff 12/18.
- MENNILL, D. J., BURT, J. M., FRISTRUP, K. M., VEHRENCAMP, S. L. (2006), Accuracy of an acoustic location system for monitoring the position of duetting songbirds in tropical forest. In: *The Journal of the Acoustical Society of America*, 119, 5: 2832–2839. URL <http://asa.scitation.org/doi/10.1121/1.2184988>.
- SETHI, S. S., EWERS, R. M., JONES, N. S., ORME, C. D. L., & PICINALI, L. (2018), Robust, real-time and autonomous monitoring of ecosystems with an open, low-cost, networked device. In: *Methods in Ecology and Evolution*. <http://doi.org/10.1111/2041-210X.13089>.
- SHONFIELD, J., BAYNE, E. M. (2017), Autonomous recording units in avian ecological research: current use and future applications. In: *Avian Conservation and Ecology*, 12, 1:art14. URL <http://www.ace-eco.org/vol12/iss1/art14/>.
- STEPANIAN, P. M., HORTON, K. G., HILLE, D. C., WAINWRIGHT, C. E., CHILSON, P. B., KELLY, J. F. (2016), Extending bioacoustic monitoring of birds aloft through flight call localization with a three-dimensional microphone array. In: *Ecology and Evolution*, 6, 19: 7039–7046.
- THE SCIPY COMMUNITY (2017), `scipy.optimize.minimize` SciPy v0.19.0 Reference Guide. URL <https://docs.scipy.org/doc/scipy-0.19.0/reference/generated/scipy.optimize.minimize.html>.
- WILSON, D. R., BATTISTON, M., BRZUSTOWSKI, J., MENNILL, D. J. (2014), Sound Finder: a new software approach for localizing animals recorded with a microphone array. In: *Bioacoustics*, 23, 2: 99–112. URL <http://www.tandfonline.com/doi/abs/10.1080/09524622.2013.827588>.

# Extracting Traffic Lanes from Floating Car Data

Tanja Mollenhauer

E-Mail: tanja.mollenhauer@gmail.com

Masterthesis at the Professorship for Applied Geoinformatics - Institute of Geography - University of Augsburg

Corrector and supervisor: Prof. Jukka Krisp; External Supervisor: Dr. Wolfgang Buhl (Bertrandt Ingenieurbüro GmbH)

## 1 Introduction

Navigable maps have become a daily presence in our everyday life. With the evolution of smartphones and automotive equipment navigation systems accompany us wherever we travel. These digital maps are presented as road network graphs. Creating and updating these maps is expensive and labour intensive. To support ITS and autonomous driving, vehicles need to be able to locate their precise position on the road down to the centimetre. To address these challenges ultra-precise digital driving instructions are needed. Standard road network graphs are no longer sufficient for these new maps.

The commoditization of GPS technology and the low costs of fleet management have enabled the generation of large amounts of vehicle tracking data (AHMED et al. 2015a). This data is called Floating car data (FCD), or Probe Vehicle Data. FCD is positional data acquired from individual vehicles through GPS devices, mobile phones or Bluetooth devices. This data is often a byproduct of existing processes and already available in vast quantities (LIU et al. 2012). The benefits of FCD compared to other map creation solutions are: low cost, quantity, high coverage, actuality, and no dependence on the collector. FCD indirectly represents the underlying road network, as the vehicles should drive on existing roads. If we assume that people generally drive in the middle of the lane, we can expect FCD to also cluster in the middle of the lane. This enables us to extract road centrelines and, even more desirable, lane centrelines from FCD.

The objective of this master thesis is to utilize real-world Floating Car Data to extract road lane centrelines and lane-level intersections. We utilize existing map creation algorithms to extract our lanes and propose an approach to extract lane centrelines in intersection areas. The data set covers a 2.6 km long motorway segment and four intersections with inner-city highways in the city of Augsburg. The data was recorded from 6. to 12. November 2017.

## 2 Summary

### 2.1 Method

After preprocessing the data, we create a road network graph to assist the lane extraction process. We compare four selected map creation algorithms and apply them to our data. (AHMED & WENK 2012) produced an acceptable road network graph for our data.

To extract the road lane centres, we use the created graph. We create perpendicular lines in fixed intervals along the centrelines. The perpendiculars are intersected with our traces. For each direction, we handle the intersection points as a separate input dataset. To extract geometries from the FCD traces we use a method utilizing the Kernel Density Estimation (KDE). We chose the input values (X- or Y-coordinates) of the KDE dynamically, depending on the orientation of the perpendicular. The KDE produces a density function from which we derive the local maxima. The local maxima represent the detected centres of the lanes and are connected to lane centrelines.

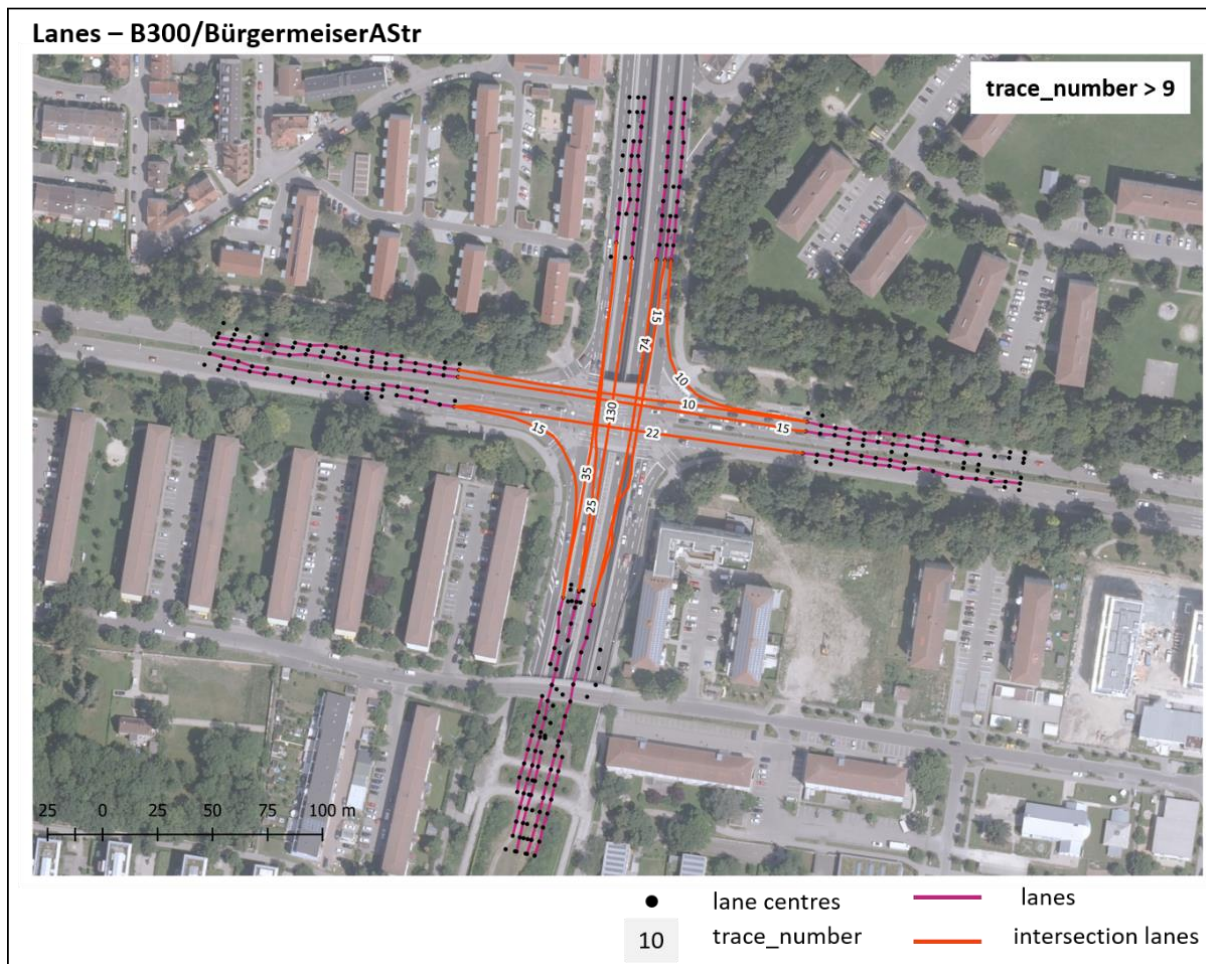
To model intersection areas, we utilise the methods used in the lane extraction process and apply them to traces in the intersection area. As input, we use the created basegraph, traces and lanes. We assign lanes to each intersection and map match the traces to these lanes, while considering the lanes and traces heading. After every lane is allocated its traces, we search for matches between different lanes. For every relation we create a lane centreline geometry.

### 2.2 Results

The resulting lane geometries possess inconsistent coverage and accuracy. On motorways, we have a higher correctness, than on the inner-city highway segments. Overall, we are able to extract the correct centre lane points with a 60%-80% probability. Reasons for the low extraction rate is the high positional error and the uneven distribution of the data. The resulting intersection geometries depict an accurate representation of the lane relations, if the amount of input traces is sufficient and the lane geometries are accurate. According to (NEUHOLD et al. 2017) the assumption that the highest density of traces corresponds to the lanes centre does not apply to input datasets with a significant accumulation of positional errors. Differences in density are evened out and lead to the detection of inaccurate lane centre points and the omission of accurate lane centre points. A characteristic of our analysis is that we survey not only straight motorway roads, but also lower level roads near intersections. Roads leading to and from intersections produce noisy data, which is less separated than on high velocity roads.

Contrary to data used in current literature, our data is not specifically created for the purpose of extracting road geometry. This leads to new challenges in the map creation process, such as an increase in preprocessing tasks, an unequal amount of available traces per road and lane, and the break out of traces to other areas outside of the road. Our lane extraction methods do not make assumptions on the underlying road structure, such as lane parallelism, number of lanes or type of intersection. As one of our use cases is the creation of digital maps in commercially unattractive regions, we cannot assume the existence of simple road network graphs to detect our lanes

When using vehicle traces for geometry creation, it is important to note that the traces do not represent the centreline of the lane, but the actual driving behaviour of the vehicles owner. If a lane geometry tailored towards the driving behaviour of humans is inconvenient or beneficial depends on the use case of the lane network graph.



**Abb. 1:** Results of lane extraction process: Intersection B300/Bürgermeister-Ackermann-Straße.  
Source: own representation (Background: Geobasisdaten: Bayerische Vermessungsverwaltung).

## Literatur

- AHMED, M. et al. (2015), A comparison and evaluation of map construction algorithms using vehicle tracking data. In *GeoInformatica* 19 (3), pp. 601–632.
- AHMED, M. & WENK, C. (2012), Constructing street networks from GPS trajectories. In: *European Symposium on Algorithms*. Springer, pp. 60–71.
- LIU, X. (2012), Mining large-scale, sparse GPS traces for map inference: comparison of approaches. In: *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, pp. 669–677.
- NEUHOLD, R. et al. (2017), Generating a Lane-Specific Transportation Network Based on Floating-Car Data. In: *Advances in Human Aspects of Transportation*: Springer, pp. 1025–1037.
- UDUWARAGODA, E.R.I.A.C. et al. (2013), Generating lane level road data from vehicle trajectories using kernel density estimation. In: *Intelligent Transportation Systems-(ITSC), 2013 16th International IEEE Conference on*. IEEE, pp. 384–391.

# An Efficient Algorithm for Computing Space-Time-Linguistics Similarities and Labelling Social Media Posts

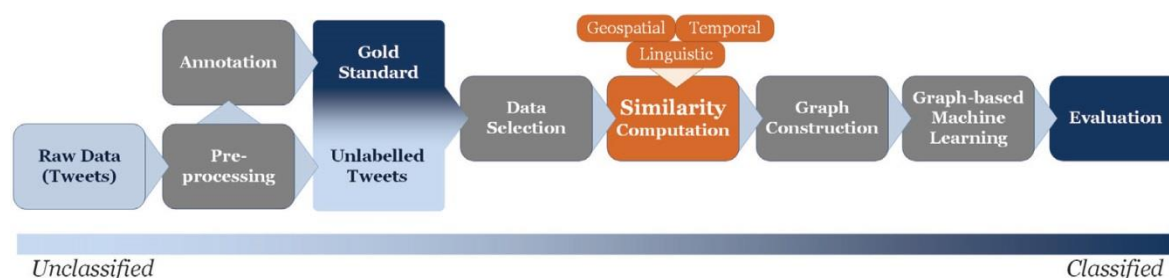
Stefan ZIMMER

Umweltfernerkundung und Geoinformatik · Universität Trier · Behringstraße 21 · 54296 Trier  
E-Mail: stefan.zimmer\_shb@t-online.de

## 1 Einführung ins Thema

Soziale Netzwerke wie Twitter und Facebook sind wachsende Online-Plattformen, auf denen Menschen persönliche Informationen als Reaktion auf aktuelle Ereignisse in Form von Beiträgen (Posts) teilen und auf bestehende Inhalte in nahezu Echtzeit reagieren können. Bisherige Analysen dieser nutzergenerierten Daten umfassten zumeist die Auswertung des textuellen Inhalts und ließen hierbei zwei wertvolle Metadaten außer Acht: die zeitliche und räumliche Komponente. Macht man sich den Informationsgehalt aller drei Dimensionen zunutze, ebnet dies den Weg für eine breite Palette von Anwendungen. So demonstrieren RESCH et al. (2016), wie sich mit Hilfe von geolokalisierten, in einem urbanen Raum veröffentlichten Twitter-Daten (Tweets) mit Methoden der Geoinformatik und Computerlinguistik eine Bürger-zentrierte Stadtplanung realisieren lässt. Das Produkt dieser Auswertung ist eine Karte von einer Stadt, die die Emotionen der Bürger reflektiert und als zusätzliche Informationsebene von Entscheidungsträgern und Stadtplaner zwecks Erkenntnisgewinn und zur Verbesserung des städtischen Lebens zurate gezogen werden kann. Eine Herausforderung ist hierbei die zunehmende Anzahl verwertbarer Posts, die mit einer wachsenden Zahl an mobilen Endgeräten einhergeht. Während diese Entwicklung auf der einen Seite die Zuverlässigkeit und Abdeckung der Analysen erhöht, erfordert sie auf der anderen Seite schnellere und skalierbare Algorithmen zur Bearbeitung der immer größer werdenden Datenmenge.

Der vorliegende Beitrag präsentiert eine Methode zur Extraktion von Emotionsinformationen aus Beiträgen, welche in sozialen Medien publiziert wurden. Grundlage dieses verbesserten Algorithmus ist das in Abb. 1 hervorgehobene Verfahren von (SUMMA et al. 2016), das Ähnlichkeiten im geographischen, zeitlichen und linguistischen Raum zwischen Posts berechnet. Auf Basis des erzeugten Ähnlichkeitsgraphen wird im Anschluss der semi-überwachter Maschine Learning Algorithmus Modified Adsorption von (TALUKDAR & CRAMMER, 2009) ausgeführt, um jeden Post mit seiner primär assoziierten Emotionskategorie – z. B. Fröhlichkeit, Wut oder Traurigkeit – zu versehen. Der vorliegende Beitrag demonstriert, wie sich eine Verbesserung der Skalierungseigenschaft des Graphen-Algorithmus von einer quadratischen auf eine quasi-lineare Komplexität bei gleichzeitig gesteigerter Klassifikationsleistung erreichen lässt.

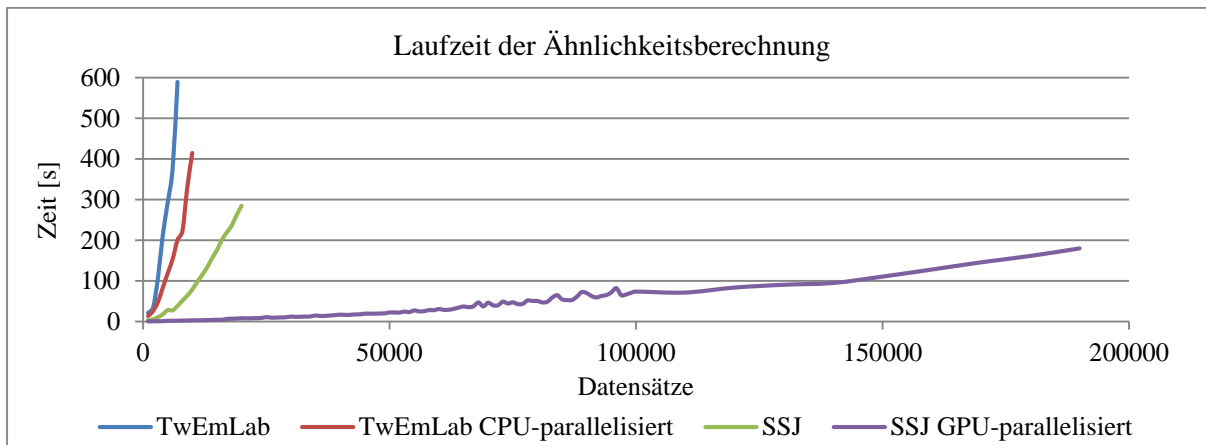


**Abb. 1:** Verarbeitungspipeline Twitter Emotion Labeller (TwEmLab) von (SUMMA et al. 2016) zur Klassifikation von Beiträgen aus sozialen Netzwerken nach ihrer prävalenten Emotion.

## 2 Zusammenfassung des Vortragsinhalts

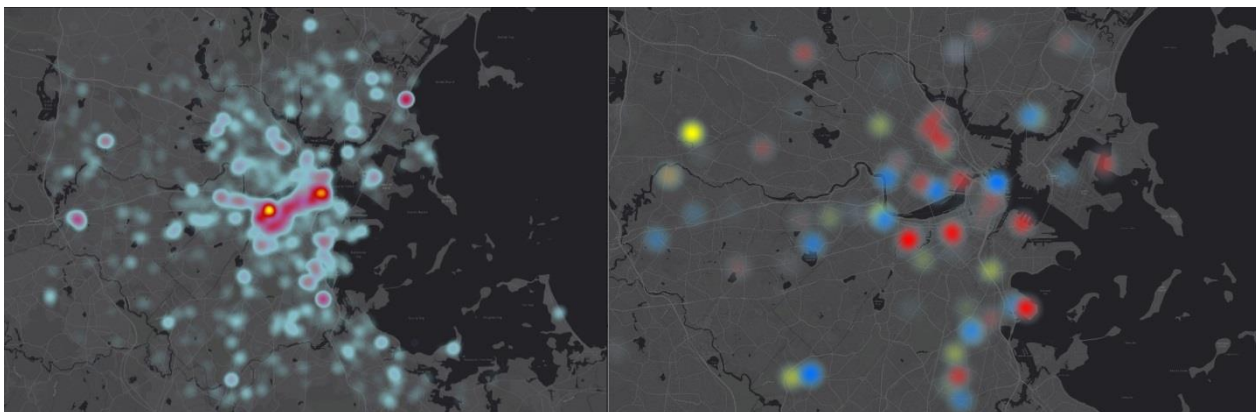
Um eine erhöhte Ausführungsgeschwindigkeit bei der Erstellung des Ähnlichkeitsgraphen mit TwEmLab zu erreichen, wurde das bestehende Verfahren zu Anfang CPU-parallelisiert. Da diese Herangehensweise die Skalierungseigenschaften unangetastet lässt und sich der Zeitaufwand lediglich umgekehrt proportional zur Anzahl der eingesetzten Prozessorkerne verbessert, brachte sie nicht den gewünschten Performancesprung bei der Bearbeitung großer Datenmengen (vgl. Abb. 2). Aus diesem Grund kam in einem zweiten Schritt die intelligentere Methode der Set Similarity Joins (SSJ) zum Einsatz. Dies erforderte zunächst die Entwicklung von Verfahren zur Entschlüsselung von Posts hinsichtlich ihrer emotionalen Botschaft: es wurden linguistische Features verwendet, die einzelne Wörter und Satzelemente sowie die schriftliche Ausdrucksweise eines Beitrags analysieren. So kamen neben Emoji-, Akronym- und Onomatopöie-Stimmungsvalenzen sowie der Art und Häufigkeit der Interpunktion auch N-Gramme zum Einsatz. Diese Features ermöglichen die semantische Charakterisierung von Posts in sozialen Medien in Bezug auf ihre Emotionslage und werden abseits der räumlichen und

zeitlichen Eigenschaften nach ihrer Kodierung in eine geeignete Token-Repräsentation bei der SSJ-basierten Ähnlichkeitsberechnung verwendet. Je näher zwei Beiträge hinsichtlich der genannten Eigenschaften beieinanderliegen – die Kanten im Ähnlichkeitsgraphen haben sodann ein hohes Gewicht – desto eher werden beide im Rahmen der semi-überwachten Klassifizierung mit derselben Emotionskategorie versehen. Abb. 2 verdeutlicht, dass ein top-k Set Similarity Join unter Verwendung des Jaccard-Koeffizienten der Ausgangsmethode TwEmLab bei der Verarbeitung großer Datenmengen rechentechnisch überlegen ist. Insbesondere die Weiterentwicklung des Algorithmus, die auf tausenden GPU-Kernen parallel ausgeführt wird und eine quasi-lineare Komplexität aufweist, zeigt eine fundamentale Performancesteigerung im Rahmen von zwei Größenordnungen bei der Ähnlichkeitsberechnung im Vergleich zur Erzeugung des Graphen mit TwEmLab. Eine weitere positive Eigenschaft ist, dass sich der Algorithmus auch bei schiefen Verteilungen der Token ausreichend robust verhält und demnach für ein größeres Spektrum an Eingangsdatensätzen geeignet ist.



**Abb. 2:** Laufzeit der Ähnlichkeitsberechnung mit TwEmLab und Set Similarity Joins in verschiedenen Parallelisierungskonfigurationen für einen Twitter-Referenzdatensatz, der die US-amerikanische Stadt Boston zum Zeitpunkt des Anschlags auf den Stadtmarathon im April 2013 abdeckt.

Die Evaluation der Klassifikationsleistung bestätigt, dass mit einem SSJ-basierten Graphen ähnliche Ergebnisse wie mit TwEmLab erreicht werden können. Durch Verwendung von neuen linguistischen Features wie z. B. der Emoji-Stimmungswalenz konnten die Resultate von (RESCH et al. 2016) – insbesondere bei den in sozialen Medien weniger dominanten Emotionsklassen wie Angst und Wut – noch übertroffen werden (vgl. Abb. 3).



**Abb. 3:** Räumliche Verteilung und Dichte von ca. 200000 Tweets im Raum Boston klassifiziert nach den Emotionen Fröhlichkeit (links), Angst (rechts, blau), Wut / Ekel (rechts, rot) und Traurigkeit (rechts, gelb).

## Literatur

- RESCH B, SUMMA A, ZEILE P, STRUBE M (2016), Citizen-Centric Urban Planning through Extracting Emotion Information from Twitter in an Interdisciplinary Space-Time-Linguistics Algorithm. UP 1:114. doi: 10.17645/up.v1i2.617.
- RIBEIRO-JÚNIOR S, QUIRINO RD, RIBEIRO LA, MARTINS WS (2016), gSSJoin: A GPU-based Set Similarity Join Algorithm. In: Machado JC (ed) 31<sup>o</sup> S<sup>o</sup> Brasileiro de Banco de Dados, SBBDD 2016, Salvador, Bahia, Brasil, October 4-7, 2016. SBC, pp 64–75.



- SUMMA A, RESCH B, STRUBE M (2016), Microblog emotion classification by computing similarity in text, time, and space. In: Proceedings of the Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media (PEOPLES), pp 153–162.
- TALUKDAR PP, CRAMMER K (2009), New Regularized Algorithms for Transductive Learning. In: Buntine W, Grobelnik M, Mladenić D, Shawe-Taylor J (eds) Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2009, Bled, Slovenia, September 7-11, 2009, Proceedings, Part II. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 442–457.

# Datengetriebene Modellentwicklung in der Umweltgeographie: Methodische Weiterentwicklungen und wissenschaftliche Anwendungen

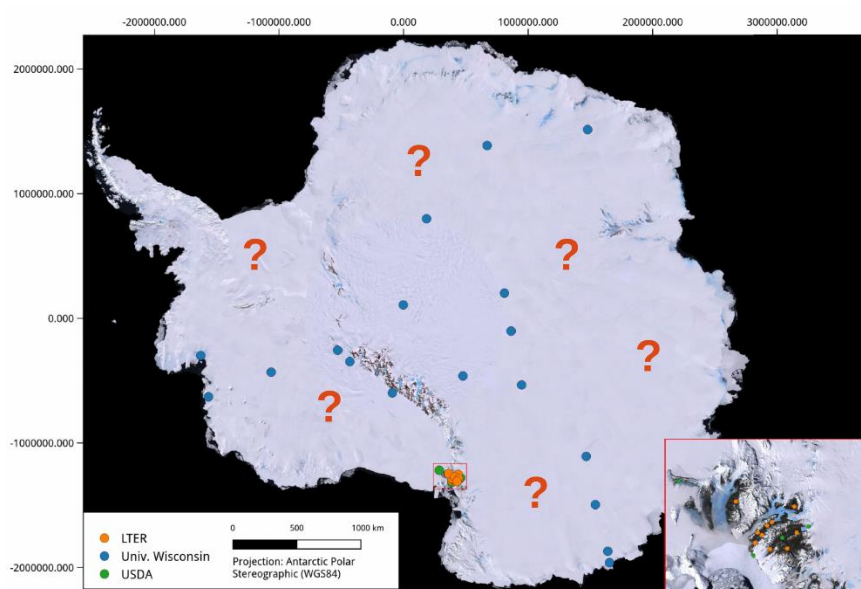
Hanna MEYER

Institut für Geoinformatik · Universität Münster · Heisenberg Str. 2 · 48149 Münster

E-Mail: hanna.meyer@uni-muenster.de

## 1 Einführung ins Thema

Die Erfassung räumlich kontinuierlicher Daten und raum-zeitlicher Dynamiken ist ein Forschungsschwerpunkt der Umweltgeographie. Zu diesem Ziel sind Modellierungsmethoden erforderlich, die es ermöglichen, aus limitierten Felddaten raum-zeitliche Informationen abzuleiten (Abb. 1). Die Komplexität von Umweltsystemen erfordert dabei die Verwendung von Modellierungsstrategien, die es erlauben, komplexe Zusammenhänge zwischen einer Vielzahl potentieller Prädiktoren zu berücksichtigen. In diesem Zusammenhang haben sich maschinelle Lernverfahren als ein wichtiges Werkzeug erwiesen, um Muster in nicht-linearen und komplexen Systemen zu erfassen. Auffallend ist jedoch, dass maschinelle Lernverfahren in der Geographie genauso angewendet werden wie in anderen Bereichen, beispielsweise der Medizin oder Wirtschaft. Geographische Daten haben aber besondere Merkmale, vor allem räumliche Abhängigkeiten, womit sie sich von "gewöhnlichen" Daten abheben. Diese Eigenschaften werden aber in maschinellen Lernanwendungen bisher weitestgehend ignoriert.



**Abb. 1:** Problemstellung im Beispiel: Klimastationen dienen als Grundlage zur Erfassung der Lufttemperatur. Gerade in abgelegenen Regionen wie der Antarktis reichen die Daten allerdings nicht aus um raum-zeitliche Muster zu beschreiben (d.h. es ist nichts über die Lufttemperatur zwischen den Klimastationen bekannt). Maschinelle Lernverfahren haben sich als vielversprechende Methode herausgestellt um mit Hilfe von Fernerkundungsdaten (z.B. MODIS Oberflächentemperaturen) flexible Modelle zu entwickeln die diese Lücken schließen. (Abbildung basierend auf MEYER et al. 2016).

## 2 Zusammenfassung des Vortragsinhalts

### 2.1 Ziele und Hypothesen

Im vorliegenden Beitrag wird das Potenzial, aber auch die Sensitivität von maschinellen Lernverfahren für geographische Daten analysiert. Grundlage dafür sind Fragestellungen aus unterschiedlichen Bereichen der Umweltwissenschaften, von Bodenkunde über Klima und Vegetationsdynamiken. Dabei beschäftigt sich der Beitrag mit der übergeordneten Hypothese, dass datengetriebene Modellierungsstrategien nur dann zu einem Informationsgewinn und zu robusten raum-zeitlichen Ergebnissen führen, wenn die Eigenschaften von geographischen Daten berücksichtigt werden. Konkret wird davon ausgegangen, dass räumliche maschinelle Lernanwendungen sowohl räumliche Validierungsmethoden erfordern,

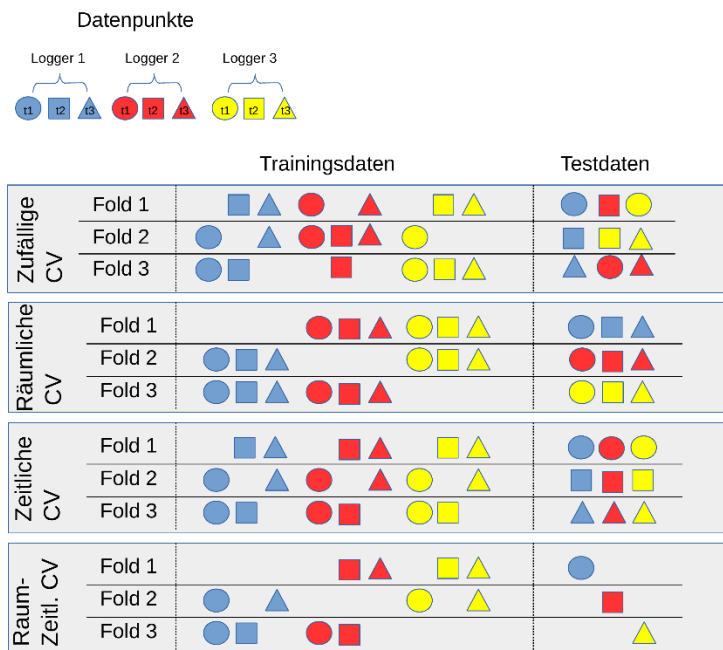
aber vor allem auch neu zu entwickelnde Strategien zur räumlichen Variablenselektion um räumliche Überfittung zu minimieren. Dieser Beitrag hat das Ziel, die Problematik mit räumlich autokorrelierten Daten in maschinellen Lernanwendungen zu verdeutlichen und neue Strategien aufzuzeigen.

### 2.2 Methoden

In diesem Beitrag wird der bekannte „Random Forest“ Algorithmus als maschinelles Lernverfahren eingesetzt. Verschiedene Fallstudien werden herangezogen um von punktuellen Umweltbeobachtungen mittels Fernerkundungsdaten und maschinellen Lernverfahren zu flächendeckenden und zeitlich quasi-kontinuierlichen Datensätzen zu kommen.

Dabei wird im Hinblick auf die Hypothesen der Effekt von verschiedenen Validierungsstrategien auf die Modellperformanz untersucht. Ausgangspunkt ist die gängige zufällige k-fache Kreuzvalidierung, bei der alle Datenpunkte zufällig in k Teile geteilt werden. Es werden k Modelle trainiert, wobei jeweils ein Teil der Daten herausgelassen wird und zur Validierung verwendet wird. Entsprechend der Hypothese werden zusätzlich zielorientierte Validierungsmethoden vorgestellt, die die Daten nicht zufällig, sondern auf Basis ihrer räumlichen und/oder zeitlichen Abhängigkeit in Teile teilt (Abb. 2). Die Ergebnisse dieser Validierungsmethoden dienen als Referenz um die Übertragbarkeit der Modelle auf neue Raum- und Zeiteinheiten abzuschätzen.

Die Ergebnisse zeigen, dass die zufällige Kreuzvalidierung im Vergleich zur räumlichen Validierung die Modelle zu optimistisch einschätzt. Es wurde vermutet, dass die räumliche Autokorrelation der Prädiktorvariablen für die geringe räumliche Performanz verantwortlich ist. Daher wurde eine neue Variablenselektion entwickelt, die die Bedeutung der Variablen im Hinblick auf ihre Relevanz für räumliche Übertragbarkeit untersucht und Variablen vom Modell ausschließt die keine räumliche Relevanz haben oder auf Grund von Überfittung sogar kontraproduktiv sind (Implementiert in das R Paket „CAST“, MEYER et al., 2018b).

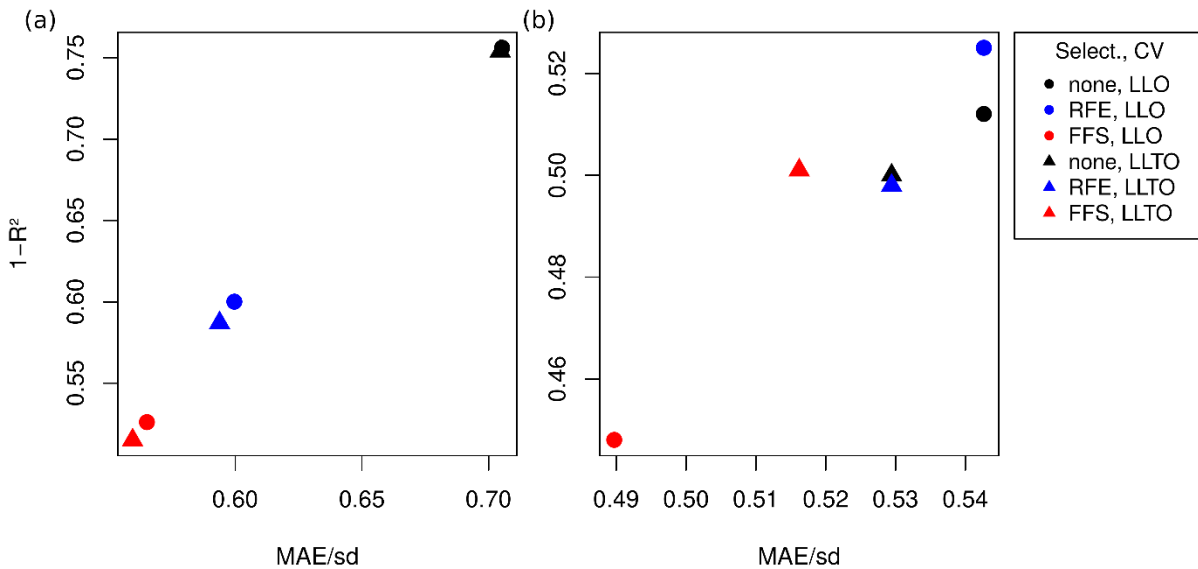


**Abb. 2:** Visualisierung der in der Arbeit verwendeten Kreuzvalidierungsmethoden (Cross-validation, CV). Zur Veranschaulichung dienen drei Datenlogger (z.B. an Klimastationen, durch Farbe dargestellt), die jeweils Daten zu drei Zeitpunkten (t1-t3) aufgenommen haben. Daten können sowohl zufällig (Standard) aber auch räumlich, zeitlich, oder raum-zeitlich geteilt werden um die Fähigkeit eines Modells zu testen, auf neue Räume oder neue Zeitpunkte vorherzusagen. (Abbildung nach MEYER et al., 2018)

### 2.2 Ergebnisse und Diskussion

Die Ergebnisse zeigen, dass räumliche Validierung zwingend notwendig ist, um die Fähigkeit der Modelle im Hinblick auf räumliche Vorhersage hin zu analysieren. Diese Beobachtung wird zunehmend in der Literatur unterstützt (z.B. ROBERTS et al., 2017). Zufällige Validierungsmethoden können bei räumlichen Abhängigkeiten nur verwendet werden um die Fähigkeit des Modells zu validieren die Daten zu reproduzieren. Es konnte gezeigt werden, dass Modelle in der

Regel deutlich weniger in der Lage sind auf neue Standorte vorherzusagen als von zufälligen Kreuzvalidierungsergebnissen erwartet werden würde. Die schwache räumliche Performanz konnte zu großen Teilen auf Überfittung zurückgeführt werden, die durch die Fehlinterpretation räumlich autokorrelierter Prädiktoren entsteht. Um diesem Problem entgegenzuwirken, wurde im Rahmen dieser Arbeit eine neue, an Geodaten angepasste Methode zur Modellentwicklung entwickelt, die räumliche Validierungsstrategien mit räumlicher Variablenselektion kombiniert wodurch deutlich verbesserte Ergebnisse erzielt werden können (Abb. 3).



**Abb. 3:** Effekt räumlicher („Leave-Location-Out“, LLO, dargestellt als Punkte) und raum-zeitlicher („Leave-Location-and-Time-Out“ (LLTO), dargestellt als Dreiecke) Variablenselektion auf die Validierungsergebnisse von Modellen zur raum-zeitlichen Vorhersage der Lufttemperatur in der Antarktis (a) sowie der Bodenfeuchte der „Cookfarm“, USA (GASCH et al., 2015) (b). Die Farben zeigen die verwendete Methode der Variablenselektion: keine Selektion in schwarz, in blau die häufig verwendete Rekursive Feature Elimination (RFE) und in rot die neu entwickelte räumliche Vorwärts-Variablenselektion (Forward Feature Selection, FFS). Die x-Achse gibt den normalisierten mittleren Fehler an und die y-Achse  $1-R^2$ . Auffallend sind die deutlich besseren Validierungsergebnisse die nach der FFS erzielt werden konnten im Vergleich zu keiner Selektion bzw. der RFE.

Dieser Beitrag ist abschließend als Appell zu verstehen, über die Standardanwendungen der maschinellen Lernverfahren hinauszudenken, da sie beweist, dass die Anwendung von Standardverfahren auf Geodaten zu starker Überfittung und Fehlinterpretation der Ergebnisse führt. Erst wenn Eigenschaften von geographischen Daten berücksichtigt werden, bietet das maschinelle Lernen ein leistungsstarkes Werkzeug, um wissenschaftlich verlässliche Ergebnisse für die Umweltgeographie zu liefern.

## Literatur

- GASCH, C., HENGL, T., GRÄLER, B., MEYER, H., MAGNEY, T., BROWN, D.J. (2015), Spatio-temporal interpolation of soil moisture, temperature, and electrical conductivity in 3D+T: the Cook Farm data set. In: *Spatial Statistics*, 14 (A), 70-90.
- MEYER, H., REUDENBACH, C., HENGL, T., KATURJI, M., NAUSS, T. (2018), Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. In: *Environmental Modelling & Software*, 101, 1-9.
- MEYER, H., REUDENBACH, C., NAUSS, T. (2018), CAST: 'caret' Applications for Spatial-Temporal Models. R package version 0.1.0.
- MEYER, H., KATURJI, M., APPELHANS, T., MÜLLER, M.U., NAUSS, T., ROUDIER, P., ZAWAR-REZA, P. (2016), Mapping daily air temperature for Antarctica based on MODIS LST. In: *Remote Sensing*, 8(9), 732.
- ROBERTS, D. R., BAHN, V., CIUTI, S., BOYCE, M. S., ELITH, J., GUILLERA-ARROITA, G., HAUENSTEIN, S., LAHOZ-MONFORT, J. J., SCHRÖDER, B., THUILLER, W., WARTON, D. I., WINTLE, B. A., HARTIG, F., & DORMANN, C. F. (2017), Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. In: *Ecography*, 40, 913-929.

# Statistische Räumliche Analyse in der Digitalen Transformation: das Beispiel Geosozialer Medien

René WESTERHOLT<sup>1,2</sup>

<sup>1</sup>Centre for Interdisciplinary Methodologies · University of Warwick · Gibbet Hill Road · CV4 7AL Coventry · UK  
E-Mail: Rene.Westerholt@warwick.ac.uk

<sup>2</sup>Geographisches Institut · Universität Heidelberg · Im Neuenheimer Feld 368 · 69120 Heidelberg · Deutschland  
E-Mail: westerholt@uni-heidelberg.de

## 1 Einführung

Unsere moderne Informationsgesellschaft produziert fortlaufend raumzeitliche Datensätze in einer nie dagewesenen Geschwindigkeit und Menge. Menschen hinterlassen ihre digitalen Spuren bei der Nutzung des öffentlichen Personennahverkehrs mit E-Tickets oder bei der Preisgabe ihrer Markenpräferenzen über den Einsatz von Kundenkarten. Diese Beispiele demonstrieren die flächendeckende und tiefe Verankerung digitaler Technologien im Alltag, welche zu einer Vermischung der Rollen des Datenproduzierenden und des Nutzers geführt hat (HAKLAY et al. 2008, COLEMAN 2009, RITZER et al. 2012). Geosoziale Medien, also soziale Medien wie Twitter oder Facebook deren Nachrichten georeferenziert werden können, haben dabei jüngst eine besondere Aufmerksamkeit in der akademischen Forschung erfahren. Nutzer hinterlassen hierbei zumeist proaktiv, jedoch oft unbewusst, subjektive Eindrücke und Meinungen über Orte und soziale Ereignisse. Die so gesammelten Informationen werden, in Kontrast zum Begriff der *Volunteered Geographic Information*, oft als *Ambient Geospatial Information* bezeichnet (STEFANIDIS et al. 2013). In der Geographie und der Geoinformatik erscheinen in jüngerer Zeit zahlreiche Arbeiten, die diese Daten für die Untersuchung menschlicher Mobilität (z.B. HASAN et al. 2013, STEIGER et al. 2016), urbaner Räume (z.B. CRANSHAW et al. 2012, SHELTON et al. 2015) und für weitere Zwecke nutzen. Auf diese Weise unterstützen diese neuartigen Datenquellen die Erlangung neuer Erkenntnisse über die räumliche Organisation des Alltagslebens unserer Gesellschaften.

Ein wichtiger Schritt in der Untersuchung der Inhalte geosozialer Medien besteht in deren statistischer räumlicher Analyse. Diese erlaubt etwa die Untersuchung räumlicher Verflechtungen und eine geographische Kontextualisierung der Daten. Dabei treten jedoch veritable methodische Herausforderungen auf: die Beziehung zwischen einer Nachricht und dem zugehörigen Ort ist oft nicht eindeutig bestimmbar. Ferner verkomplizieren Probleme wie Selbstselektivität oder die Ambiguität der semantischen Beiträge räumliche Analysen (SENGSTOCK & GERTZ 2012, TUFEKCI 2014). Ein besonders gravierendes Problem besteht jedoch in der räumlich und zeitlich koinzidenten Repräsentation verschiedenster Phänomene, sowie in deren erschwelter Trennbarkeit. Nutzer tragen Nachrichten in unterschiedlichen linguistischen, kulturellen und intentionalen Arten und Weisen über verschiedene Phänomene bei, in der Regel in Form sehr kurzer Beiträge. Selbst nach ausgeprägter Vorverarbeitung sind vorhandene räumliche Analysemethoden oft nicht hinreichend in der Lage, die Eigenschaften solcher Zufallsvariablen zu berücksichtigen. Statistische ausgedrückt können Daten aus geosozialen Medien demnach als *räumlich (und zeitlich) überlagerte Zufallsvariablen* bezeichnet werden, bei denen der räumlichen Struktur der Heterogenität als Indikator für die inhaltliche Durchmischung an Orten eine besondere Bedeutung zukommt. Erkenntnisse über die Auswirkungen der resultierenden Wechselwirkungen mit bekannten Methoden, sowie neue methodische Ansätze sind deshalb notwendig, um effektiv detaillierte Erkenntnisse über die räumliche Organisation von Orten mittels Daten sozialer Medien zu erzielen.

Die nachfolgend und im zugehörigen Vortrag präsentierte Forschung befasst sich mit der Schnittstelle zwischen räumlicher Analysemethodik und den Eigenschaften von Daten aus geosozialen Medien. Die Ergebnisse wurden im Rahmen einer an der Universität Heidelberg kumulativ erstellten Dissertation erlangt. Drei Arten wissenschaftlicher Beiträge werden dabei vorgestellt: es werden (i) empirische Erkenntnisse hinsichtlich der Interaktion zwischen Methodik und Dateneigenschaften; (ii) methodische Beiträge zur Berücksichtigung und weitergehenden Charakterisierung dieser Eigenschaften; und (iii) eine Einbettung dieser Erkenntnisse in einen breiteren Kontext, der über die Geographie und die Geoinformatik hinausgeht.

## 2 Räumliche Dateneigenschaften und Analysemethodik

Die Anwendung von Methoden wird maßgeblich von Dateneigenschaften beeinflusst. Angewandte Methoden sollten deshalb stets in Einklang mit den Eigenschaften der zu untersuchenden Daten sein. Die räumliche Analyse stellt hierbei keine Ausnahme dar. Viele der üblicherweise angewandten räumlichen Methoden stammen aus einer Zeit vor dem Aufkommen geosozialer Medien. Vor diesem Hintergrund beleuchten die nachfolgenden Unterkapitel die Auswirkungen verschiedener Dateneigenschaften geosozialer Mediendaten auf zwei exemplarische räumlich-statistische Methoden: den Schätzer räumlicher Autokorrelation Moran's  $I$ , sowie die Hot-Spot-Methode Getis-Ord  $G_i^*$ .

## 2.1 Überlagerte Maßstäbe

Räumliche Analysen mit geosozialen Mediendaten werden dadurch erschwert, dass Menschen selbstständig Beobachtungen ohne jegliche thematischen Einschränkungen einbringen. In der Folge werden mehrere Phänomene auf unterschiedlichen und räumlich durchmischten Skalen simultan und an gleichen Orten erhoben. Darüber hinaus verwenden Menschen unterschiedliche idiosynkratische Skalenkonzepte, wenn sie subjektive Eindrücke über den gleichen Aufenthaltsort beibringen, was zu vielfältigen räumlichen Konzeptualisierungen führt (WENDER et al. 2002, DANGSCHAT 2007). Dies steht im Gegensatz zu eigens für die geografische Forschung erhobenen Primärdaten, bei denen die Erfassungsskala an ein einziges kohärentes Phänomen angepasst wird und bei denen keine subjektiv wahrgenommenen räumlichen Skalen einbezogen werden. Ähnliches gilt, wenn auch in eingeschränkter Form, für professionell erhobene Sekundärdaten wie etwa jene aus dem Zensus.

Geosoziale Mediendaten enthalten unterschiedliche Maßstäbe in räumlich vermischter Form. Die Ergebnisse in (WESTERHOLT et al. 2015) zeigen, dass für räumliche Analysen gebildete Nachbarschaften bei geosozialen Mediendaten heterogen sind. Fünf verschiedene Maßstäbe wurden herangezogen und auf jedem dieser Maßstäbe interagieren 70 – 90% der eingeschlossenen Beobachtungen auf Maßstäben jenseits des jeweiligen Spektrums von Interesse. Zusätzlich dominieren kleine räumliche Maßstäbe die konstruierten räumlichen Nachbarschaften so, dass die kleinsten beiden der untersuchten Maßstäbe mehr als 40% aller enthaltenen Social-Media-Nachrichten repräsentieren. Unter Berücksichtigung der Attributwerte (hier: semantische Ähnlichkeiten) zeigt sich weiterhin, dass der Anteil der Attributwerte, die durch die kleinsten Beobachtungen beigetragen werden, ihren quantitativen Anteil um bis zu 80% übersteigt, was darauf hindeutet, dass kleine Maßstäbe stark überbewertet sind. Daher ist es schwierig, aus den untersuchten geosozialen Mediendaten signifikante Hot-Spot-Muster auf größeren Maßstäben zu erkennen, ohne von den kleinsten enthaltenen Maßstäben beeinflusst zu werden.

Die erzielten Ergebnisse zeigen ferner, dass Hot-Spots häufig falsch interpretiert werden oder unentdeckt bleiben, wenn Maßstäbe in räumlich überlagerter Form auftreten. Untersucht wurde der Hot-Spot-Schätzer  $G_i^*$  (GETIS & ORD 1992, ORD & GETIS 1995) und dessen Anwendung auf Twitter-Daten aus San Francisco. Die Ergebnisse zeigen, dass dieser Schätzer im Mittel einen positiven Trend über die untersuchten Maßstäbe aufweist. Die Erkennung von Hot-Spots aus geosozialen Medien ist demnach eine Funktion des Maßstabs, infolgedessen die Nullhypothese (d. h. kein Hot-Spot) auf groben Maßstäben zu oft abgelehnt wird. Dies geschieht, da durch die räumliche Durchmischung signifikante Hot-Spots aus kleineren Maßstäben auf größere übertragen werden. Auf kleinen Analysemaßstäben zeigt sich hingegen häufig ein gegenteiliges Verhalten, wenn sich niedrige und hohe Attributwerte maßstabsübergreifend vermischen. Hierdurch werden potenzielle Hot-Spot-Muster und somit interessante Effekte übersehen. Ferner bedeutet dies, dass Hot-Spots auf unterschiedlichen Maßstäben statistisch voneinander abhängig sind, auch wenn dies ggfs. nicht der Realität entspricht. Ähnlich den bekannten Problematiken bei multiplem Hypothesentesten müssen, bei der Verwendung von Daten aus geosozialen Medien, die beschriebenen Effekte kontrolliert werden, um realistische und interpretierbare Ergebnisse zu erzielen.

## 2.2 Topologisch-bedingte Effekte

Die Topologie konstruierter räumlicher Analysenachbarschaften hat großen Einfluss auf die Ergebnisse räumlicher Analysen. Der Schätzer Morans  $I$  (CLIFF & ORD 1969, GETIS 2010) etwa ist empfindlich gegenüber topologischen Ungleichgewichten im geographischen Layout einer Region (TIEFELSDORF & BOOTS 1997, TIEFELSDORF et al. 1999, SHORTRIDGE 2007). So ergeben sich Problematiken bei vielen Randeinheiten mit geringfügiger Konnektivität bei gleichzeitigem Vorhandensein von hochgradig verbundenen geographischen Einheiten. Die Durchmischung verschiedenster Phänomene in geosozialen Medien führt zu einer komplexen Anordnung mit teilweise artifizieller geographischer Nähe. Der Einfluss dieser Problematik auf die Analyse räumlicher Autokorrelation ist Gegenstand der Studie in (WESTERHOLT et al. 2016).

Stark interagierende räumliche Einheiten im Überlagerungsbereich verschiedener Prozesse beeinflussen die Verteilung der räumlichen Autokorrelationsstatistik Morans's  $I$ . Die Eigenwerte von räumlichen Gewichtungsmatrizen, welche das Interaktionspotenzial zwischen den Einheiten bestimmen, liefern ein detailliertes Verständnis dafür, wie die Verbundenheit in einem räumlichen Arrangement die räumliche Autokorrelation beeinflusst. Die erzielten Ergebnisse zeigen, dass überlagerte Eigenwertmuster sehr heterogen mit gleichzeitigen hohen und niedrigen Werten sind, wobei für diejenigen Einheiten, die über verschiedene überlagerte Muster hinweg interagieren, die stärksten Eigenwerte auftreten. Diese Einheiten erhöhen die Variabilität im Eigenwertspektrum, was wiederum den Bereich der theoretisch möglichen Morans's- $I$ -Werte streckt und auch die Form ihrer Verteilung verändert. Die Interaktion zwischen überlagerten Mustern hat somit einen starken Einfluss auf die Aufdeckung räumlicher Muster, da diese durch eine rein technisch bedingte Interaktion in besonderem Maße in die Analyse einfließen. Inferenzen, die über Morans  $I$  gezogen werden, sind dann verzerrt und haben eine geringere statistische Aussagekraft, insbesondere bei Annahme einer Normalverteilung der statistischen Werte.

Mehrere möglicherweise gegensätzliche räumliche Prozesse werden in überlagerten Daten falsch identifiziert. Moran Scatterplots erlauben die Zerlegung von Morans  $I$  in seine lokalen Bestandteile. Obwohl in den untersuchten überlagerten Mustern defacto nur ein räumliches Muster existiert, zeigt das Streudiagramm des Moran Scatterplot drei verschiedene räumliche Prozesse. Eines davon spiegelt die tatsächliche räumliche Interaktion in den Daten wider und zeigt eine positive Steigung in der Regressionslinie (d. h. räumliches Clustering). Darüber hinaus erscheinen eine weitere falsche Gerade

mit positiver, sowie eine weitere mit negativer Steigung. Diese bilden die Wechselwirkungen zwischen den beteiligten Prozessen ab. Jede dieser Geraden ist einer der Maßstäbe der beteiligten Prozesse zugeordnet. Wenn diese Problematik nun nicht explizit berücksichtigt wird, werden diese fälschlich identifizierten räumlichen Prozesse in die Charakterisierung des gesamten räumlichen Interaktionsverhaltens der analysierten Daten einbezogen. Wenn man diese Komponenten in Beziehung zu den oben skizzierten Eigenwerten setzt, zeigt sich weiter, dass sich die beiden Komponenten bei zunehmenden Eigenwerten gegenläufig verhalten. Ihre Einflüsse werden jedoch stärker, was die starke Wirkung der räumlichen Überlagerung auf die Interpretation räumlicher Muster verdeutlicht.

### 2.3 Effekte statistischer Heterogenität

Eine Vielzahl verschiedener Prozesse sind in geosozialen Mediendaten abgebildet. Statistische Datenmerkmale beeinflussen jedoch die Schätzung räumlicher Autokorrelationen stark. Es ist bekannt, dass Morans  $I$  schneller auf die Normalverteilung konvergiert, wenn die eingehenden Zufallsvariablen selbst nahezu normal sind, oder zumindest einer symmetrischen Verteilung folgen (GRIFFITH 2010). Ebenso setzt Moran's  $I$  eine minimale Variabilität innerhalb der untersuchten Zufallsvariablen voraus, um eine ausreichende statistische Aussagekraft zu gewährleisten (WALTER 1992a, 1992b). Gleichzeitig sollten die Varianzen der zugrundeliegenden datengenerierenden Prozesse gleichförmig sein, um zuverlässige Rückschlüsse auf Morans  $I$  ziehen zu können (ODEN 1995, WALDHÖR 1996, SHEN et al. 2016). Die Heterogenität statistischer Merkmale in geosozialen Medien hat somit einen großen Einfluss auf die Erkennung räumlicher Strukturen. Diese wurden in (WESTERHOLT 2018) näher beleuchtet.

Überlagerte Mittelwerte verschiedener Intensitäten verändern die Interpretation von Moran's- $I$ -Schätzungen. Im Hinblick auf statistische Unterschiede wird die Stärke der räumlichen Autokorrelation mit stark unterschiedlichen Mittelwerten unterschätzt, wenn die sich überlappenden Muster räumlich willkürlich sind, d. h., wenn diese für sich genommen nicht räumlich autokorreliert sind. Im Gegensatz dazu hängt der Grad der Unterschätzung von räumlich strukturierten überlagerten Mustern zusätzlich vom geometrischen Maßstab, der mit dem stärkeren der beteiligten Mittelwerte verbunden ist, ab. Daher führen dominante großmaßstäbige Muster zu einer stärkeren Unterschätzung. Die Geschwindigkeit, mit der diese Effekte wirksam werden, ist überdies schneller als bei dominanten kleinmaßstäbigen Mustern. Die Wirkung unterschiedlicher Mittelwerte führt daher zu einer Fehlinterpretation der Stärke der räumlichen Strukturierung in den Daten.

Unterschiedliche Attributvarianzen in überlagerten Zufallsvariablen erhöhen die Unsicherheit von räumlichen Autokorrelations-schätzungen. Die Unsicherheit bei der Schätzung von Moran's  $I$  steigt, wenn die Varianzen der überlagerten Attribute unterschiedlich sind. Dieser Effekt ist sowohl für räumlich zufällige als auch für räumlich strukturierte Überlappungsmuster ähnlich. Auch in Bezug auf die Maßstäbe der beteiligten Muster ist der Einfluss unterschiedlicher Varianzen symmetrisch. Dabei spielt es keine Rolle, ob das großmaßstäbigere oder das kleinmaßstäbigere Muster in Bezug auf die Streuung dominiert. Der Varianz-Effekt wird schnell wirksam, so dass bereits kleine Unterschiede zu einer recht starken Vergrößerung des theoretisch möglichen Bereichs der Moran's- $I$ -Werte führen. Bemerkenswert ist auch, dass Varianzabweichungen im Allgemeinen zu einer Prävalenz gegenüber größeren Moran's- $I$ -Schätzungen führen. Während also Mittelwertabweichungen zu Unterschätzungen führen, können Unterschiede in den beteiligten Varianzen die Auswirkungen der Mittel ausgleichen. Dieser Befund ergänzt die für die verschiedenen gemeinsam auftretenden räumlichen Prozesse im Moran Scatterplot ermittelten Ausgleichseffekte (gegenteilige Regressionslinien).

## 3 Erarbeitung neuer methodischer Ansätze

Die aufgezeigten Problematiken verlangen neuartige methodische Vorgehensweisen, um den räumlichen Spezifika von geosozialen Mediendaten zu begegnen. Nachfolgend werden zwei Ansätze vorgestellt, welche einige der oben ausgeführten Eigenschaften berücksichtigen.

### 3.1 Hot-Spot-Schätzung mit überlagerten Maßstäben

Die Übereinstimmung von Analyse- und Phänomen-Maßstab ist entscheidend für die Identifizierung sinnvoller Strukturen (GOODCHILD 2001). Hot-Spot-Schätzer sind besonders anfällig für Maßstabsfehler, da diese Attributwerte lokale Datenpunkte additiv zusammenfassen. Besonders wenn der Analysemaßstab grob ist, werden die Ergebnisse durch die Einbeziehung von nicht relevanten Datenpunkten beeinflusst. Da geosoziale Medienfeeds über diskrete geographische Einheiten indiziert werden, ist es erforderlich, eine Matrix von paarweisen Beziehungen zwischen den gesampelten Standorten zu adjustieren. Diese regelt auch den Maßstab einer Analyse. Es steht eine Vielzahl von verschiedenen Matrixdesigns zur Verfügung (vgl. ALDSTADT & GETIS 2006, MAWARNI & MACHDI 2016, ERMAGUN & LEVINSON 2017). Es existiert jedoch kein Ansatz, der die spezifischen räumlichen Eigenschaften von überlagerten Zufallsvariablen berücksichtigt. (WESTERHOLT et al. 2015) behandeln die Hot-Spot-Schätzung aus überlagerten Zufallsvariablen, indem zwei Beiträge vorgeschlagen werden: (i) Es wird ein neuartiger Ansatz zur räumlichen Gewichtung vorgestellt, der sich von den verfügbaren in stratifizierten Nachbarschaften in Bezug auf die lokal enthaltenen Interaktionsmaßstäbe unterscheidet. Darüber hinaus wird (ii) ein neuartiger Hot-Spot-Schätzer abgeleitet, der es ermöglicht, Hot-Spots auf verschiedenen Skalen getrennt voneinander offenzulegen.

Zur Ableitung von räumlichen Gewichten für überlagerte Zufallsvariablen wird ein zweistufiges Verfahren vorgeschlagen: Um jede Raumeinheit herum wird zunächst eine kreisförmige Grenze gezogen, deren Abstandsgrenze der geometrischen Reichweite des analysierten Prozesses entspricht. Alle paarweisen Beziehungen zwischen den darin enthaltenen Zufallsvariablen werden dann untersucht, um festzustellen, ob sie sich auf dem Maßstab befinden, auf dem der analysierte Prozess voraussichtlich abläuft. Das abgeleitete Gewichtungsschema bildet somit einen hybriden Ansatz, der geometrische (die kreisförmige Grenze) und topologische Prinzipien (die relative Platzierung von räumlichen Einheiten) beinhaltet, die es ermöglichen, lokale Nachbarschaften in verschiedene unterschiedliche, sich aber geometrisch überlappende Teile zu stratifizieren. Bestimmte Maßstäbe können ein- und ausgeschaltet werden, um sie anschließend separat auszuwerten. Während konventionelle verfügbare räumliche Gewichtungssysteme davon ausgehen, dass die Nachbarschaft in sich stimmig ist, ohne dass sie weiter untergliedert werden muss, stellt der vorgestellte Ansatz ein System für den Fall von räumlich überlagerten geographischen Zufallsvariablen dar.

Ferner wird eine modifizierte Version von  $G_i^*$  abgeleitet. In diesem neuen Maß berücksichtigt der normalisierende Nenner die Existenz verschiedener räumlicher Maßstäbe durch Auswertung eines binären Vektors, der die Maßstäblichkeit angibt. Die ursprüngliche Methode berücksichtigt keine unterschiedlichen Maßstäbe und verwendet einen konstanten Nenner, der auf allen verfügbaren Daten basiert. Darüber hinaus ist der Zähler des neuen Schätzers mit dem oben vorgeschlagenen Gewichtungsschema integriert. Dies ermöglicht es, irrelevante Informationen auszusortieren, erfordert jedoch eine Korrektur der Freiheitsgrade des Schätzers. Schließlich werden Ausdrücke für den Mittelwert und die Varianz des Maßes hergeleitet, die ebenfalls auf relevante Maßstäbe beschränkt sind. Das letztlich vorgeschlagene Maß wird als asymptotisch normaler z-Score präsentiert, der eine einfache Inferenz und Interpretation ermöglicht. Mit dem vorgestellten Verfahren ist es möglich, Hot-Spots auf verschiedenen Maßstäben isoliert zu bewerten und damit ansonsten nicht erkennbare Phänomene aufzudecken.

Die Anwendung von  $G_i^*$  und dem neuen, abgeleiteten Maß auf einen Twitter-Datensatz zeigt den Nutzen des letzteren für diese Art von Daten. Während der Mittelwert von  $G_i^*$  einen stark positiven Trend zu größeren Maßstäben zeigt (siehe Abschnitt 2.1), ist die Trendlinie für das neue Maß flach und bleibt nahe Null, was das erwartete Verhalten für z-Scores ist. Der neue Schätzer erlaubt ferner eine bessere Identifizierung und Trennung von räumlichen Hot- und Cold-Spots auf verschiedenen Maßstäben. Der heterogene Central Business District (CBD) von San Francisco erscheint als starker Cold-Spot, während das asiatische Viertel im nördlichen Teil einen markanten Hot-Spot aufweist, der chinesische Neujahrsfeiern darstellt. Darüber hinaus treten in zentralen Stadtteilen auf der größten analysierten Skala kleine Hot Spots auf (z. B. ein College-Campus), während andere nur auf kleineren Maßstäben vorhanden sind (z. B. eine High-School im Norden). Diese Phänomene sind entweder mit  $G_i^*$  nicht nachweisbar (Typ-II-Fehler) oder bleiben auf den meisten Analysemaßstäben fälschlicherweise präsent (Typ-I-Fehler).  $G_i^*$  zeigt somit eine hohe Anzahl von False Positives auf großen Analysemaßstäben, wobei 33,56% aller Zufallsvariablen signifikant gekennzeichnet sind. Dies ist bei dem abgeleiteten Maß nicht der Fall, da hier lediglich 3,77% aller Beobachtungen als signifikant bewertet werden, was der angegebenen Fehlerwahrscheinlichkeit von 5% nahe kommt.

### 3.2 Schätzer räumlich strukturierter Varianz

Die Untersuchung der räumlichen Heterogenität von überlagerten Zufallsvariablen ist für ein besseres Verständnis dieser Art von Daten nützlich. Räumliche Heterogenität ist ein Indikator für die räumliche Instabilität in Zufallsvariablen (DUTILLEUL & LEGENDRE 1993). In traditionellen Datensätzen bezieht sich dies auf Instabilität in statistischen Momenten über einen Beobachtungsbereich. Bei überlagerten Zufallsvariablen sollte zusätzlich die räumliche Heterogenität lokal innerhalb der Standorte untersucht werden, was detaillierte Charakterisierungen erlaubt, wie die lokale räumliche Anordnung von Zufallsvariablen deren Vielschichtigkeit beeinflusst. Dies wiederum hilft, die räumliche Organisation von Orten besser zu verstehen. Um räumliche Mischungen von Prozessen besser zu verstehen, wurde ein statistischer Test über den Zusammenhang zwischen räumlicher Anordnung und Varianz entwickelt (WESTERHOLT et al. 2018). Diese Teststatistik basiert auf zwei Prinzipien: (i) sie nutzt nur lokale Informationen, um räumliche Heterogenität zuzulassen; und (ii) es beinhaltet einen streng lokalen inferentiellen Bezugsrahmen. Diese Prinzipien ermöglichen es zu beurteilen, ob die Art und Weise, wie Zufallsvariablen räumlich angeordnet sind, die Varianz an einem bestimmten Ort reduziert oder erhöht, oder ob diese beiden Merkmale nicht miteinander in Beziehung stehen.

Das vorgeschlagene Maß namens Local Spatial Dispersion (LSD) ermöglicht es, den Einfluss eines räumlichen Musters auf die lokale Variabilität vor Ort zu testen, ohne von der globalen geographischen Varianzverteilung beeinflusst zu werden. Im Gegensatz dazu zeigt Local Spatial Heteroscedasticity (LOSH), eine kürzlich vorgeschlagene Methode (ORD & GETIS 2012), von der LSD abgeleitet ist, lediglich Varianz-Hot-Spots, die sich im globalen Vergleich abheben. Ausschließlich lokale Varianzstrukturen, die global nicht herausragen, werden hingegen nicht detektiert. Der wesentliche technische Unterschied zu LOSH besteht darin, dass LSD geschätzte Residuen über lokale, räumlich gewichtete Mittelwerte mit ihren eigenen lokalen Durchschnittswerten vergleicht, während LOSH den globalen Durchschnitt dieser Residuen in lokale Vergleiche einbezieht. Auf diese Weise bewertet LSD den gänzlich lokalen Einfluss der räumlichen Anordnung auf die Varianz, ohne auf das Streuverhalten an anderen Orten Bezug zu nehmen. Es ist somit möglich, zu erkennen und zu charakterisieren, wie sich das lokale räumliche Layout auf die Varianz auswirkt, auch an Orten, die von LOSH als global nicht signifikant identifiziert werden.



Das für LSD entwickelte Inferenz-Framework ermöglicht räumliche Heterogenität zweiter Ordnung durch lokale Randomisierung. Das Ziehen lokaler Schlüsse geht jedoch auf Kosten der zur Verfügung stehenden Informationen. Potentiell kleine lokale Nachbarschaften bieten oft nur spärliche Informationen, was zu unzuverlässigen Referenzverteilungen führen kann. Um diesem Problem zu begegnen, beinhaltet der vorgeschlagene Ansatz ein Bayes'sches Rahmenwerk zur Vorhersage zusätzlicher synthetischer lokaler Mittelwerte. Zusätzliche lokale Residuen können dann über diese Mittel geschätzt werden, so dass es möglich ist, einen lokalen Bootstrap aus einer beliebigen Anzahl von Monte-Carlo-Replikationen zu berechnen. In einem ersten Schritt werden die globalen statistischen Informationen aus allen verfügbaren lokalen, räumlich gewichteten Mittelwerten ausgewertet, um ihre gemittelte Priorverteilung abzuleiten. In diesem Schritt werden die lokalen Informationen aus der jeweiligen Nachbarschaft ausgeklammert, um eine doppelte Verwendung der Daten zu vermeiden. Der Prior wird dann mit lokalen Informationen kombiniert, um die anfänglich konstruierte Verteilung an die örtlichen Gegebenheiten anzupassen. Dieser Ansatz hat zwei Vorteile: Der globale Prior reduziert das Risiko einer lokalen Überanpassung an eventuell extreme Gegebenheiten, während die Verwendung lokaler Informationen eine starke globale Mittelung vermeidet und zu einer realistischeren Darstellung des lokalen Streuverhaltens führt.

## 4 Schlussfolgerungen und Ausblick

Geosoziale Medien und verwandte Daten sind im Hinblick auf räumliche Analysen herausfordernd. Sie entstammen einem größtenteils unkontrollierten Erhebungsprozess, was im wissenschaftlichen Kontext zu neuartigen Problemstellungen führt. Räumlich betrachtet finden sich zahlreiche unterschiedliche Phänomene simultan in den Daten. Deren Trennung ist komplex und oft unmöglich, da es sich oftmals um Nuancen handelt. Aus diesem Grund können georeferenzierte Social-Media-Daten als räumlich strukturierte, heterogene Mischung betrachtet werden.

Die hier dargelegten Arbeiten zeigen eine Reihe von Problemstellungen auf. Diese reichen von Maßstabsproblematiken über topologische Problemstellungen bis hin zu räumlich komplex durchmischter statistischer Variation. Ferner wurden erste Ansätze zur räumlich-statistischen, methodischen Behandlung dieser Problematiken aufgezeigt. Insgesamt deuten die Ergebnisse darauf hin, dass räumlich überlagerte und heterogene Zufallsvariablen räumlich-statistische Darstellungen von *örtlichen* (im Gegensatz zu *räumlichen*) Informationen sind. Die vorgelegten Ergebnisse unterstützen daher die jüngsten Diskussionen über geosoziale Mediendaten als weitgehend örtlich charakterisiert (GAO et al. 2014, QUESNOT & ROCHE 2015, JENKINS et al. 2016, ROCHE 2016). In diesem Sinne ist die räumliche Überlagerung ein Artefakt der Komplexität subjektiv wahrgenommener und erlebter Orte. Viele Menschen halten sich zwar an einem Standort auf, weisen diesem jedoch gänzlich unterschiedliche Bedeutungen zu. Die zukünftige Forschung sollte demnach zwei Richtungen einschlagen: Zum einen sollte die räumlich-statistische Forschung auf ein geeignetes Instrumentarium für räumlich überlagerte Zufallsvariablen hinarbeiten. Dies erscheint sinnvoll, da auch die räumliche Komponente als Teilausschnitt eines Ortes von wissenschaftlichem Interesse ist. Ferner sollte ein neues Feld zur statistisch-quantitativen Betrachtung subjektiver Ortswahrnehmungen eingeführt werden. Letzteres dürfte als langfristige Zielsetzung gelten, ist jedoch indirekt Gegenstand der aktuellen Debatte in der Geoinformatik (SCHEIDER & JANOWICZ 2014, GOODCHILD 2015). Ferner ist anzumerken, dass, in begrenztem Umfang, die dargelegten Problematiken auch auf andere, moderne Daten zutreffen. Ein Beispiel hierfür sind etwa georeferenzierte Umfragedaten, die über geographische Mechanismen ortsgesteuert aktiviert werden (z.B., wenn ein Proband einen bestimmten Bereich betritt). Auch hier wird die räumliche Kontrolle seitens des Datenerhebenden teilweise aufgegeben, was zu ähnlichen Problemstellungen führen kann (BLUEMKE et al. 2017).

## Literatur

- ALDSTADT, J. & GETIS, A., 2006, Using AMOEBA to Create a Spatial Weights Matrix und Identify Spatial Clusters. *Geographical Analysis*, 38 (4), 327–343.
- BLUEMKE, M., LECHNER, C., RESCH, B., WESTERHOLT, R. & KOLB, J., 2017, Integrating Geographic Information into Survey Research: Current Applications, Challenges, und Future Avenues. *Survey Research Methods*, 11 (3), 307–327.
- CLIFF, A. & ORD, J., 1969, The Problem of Spatial Autocorrelation. In: A. Scott, ed. *London Papers in Regional Science (1)*, *Studies in Regional Science*. London: Pion, 25–55.
- COLEMAN, D., 2009, Volunteered Geographic Information in Spatial Data Infrastructure: An Early Look at Opportunities und Constraints. In: A. Rajabifard, J. Crompvoets, M. Kanantari, und B. Kok, eds. *Spatially Enabling Society: Research, Emerging Trends und Critical Assessment*. Leuven: Leuven University Press, 131–148.
- CRANSHAW, J., SCHWARTZ, R., HONG, J. & SADEH, N., 2012, The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City. In: *Proceedings of the 6th International AAAI Conference on Weblogs und Social Media*. Dublin.
- DANGSCHAT, J., 2007, Raumkonzept zwischen struktureller Produktion und individueller Konstruktion. *Ethnologie und Raum*, 9 (1), 24–44.
- DUTILLEUL, P. & LEGENDRE, P. 1993, Spatial Heterogeneity Against Heteroscedasticity: An Ecological Paradigm Versus a Statistical Concept. *Oikos*, 66 (1), 152–171.
- ERMAGUN, A. & LEVINSON, D., 2017, An Introduction to the Network Weight Matrix. *Geographical Analysis*, forthcoming.

- GAO, S., YANG, J., YAN, B. & MCKENZIE, G., 2014, Detecting Origin-Destination Mobility Flows From Geotagged Tweets in Greater Los Angeles Area. *In: Proceedings of the Eighth International Conference on Geographic Information Science*. Vienna.
- GETIS, A., 2010, Spatial Autocorrelation. *In: M. Fischer und A. Getis, eds. Handbook of Applied Spatial Analysis*. Heidelberg: Springer, 255–278.
- GETIS, A. & ORD, J., 1992, The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis*, 24 (3), 189–206.
- GOODCHILD, M., 2001, Models of Scale und Scales of Modeling. *In: N. Tate und P. Atkinson, eds. Modelling Scale in Geographical Information Science*. Chichester, UK: John Wiley & Sons, 3–10.
- GOODCHILD, M., 2015, Space, Place und Health. *Annals of GIS*, 21 (2), 97–100.
- GRIFFITH, D., 2010, The Moran Coefficient for Non-Normal Data. *Journal of Statistical Planning und Inference*, 140 (11), 2980–2990.
- HAKLAY, M., SINGLETON, A. & PARKER, C., 2008, Web Mapping 2.0: The Neogeography of the GeoWeb. *Geography Compass*, 2 (6), 2011–2039.
- HASAN, S., ZHAN, X. & UKKUSURI, S., 2013, Understanding Urban Human Activity und Mobility Patterns Using Large-Scale Location-Based Data from Online Social Media. *In: Y. Zheng, ed. Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing*. New York, NY: ACM Press.
- JENKINS, A., CROITORU, A., CROOKS, A. & STEFANIDIS, A., 2016, Crowdsourcing a Collective Sense of Place. *PLOS ONE*, 11 (4), e0152932.
- MAWARNI, M. & MACHDI, I., 2016, Dynamic Nearest Neighbours for Generating Spatial Weight Matrix. *In: 2016 International Conference on Advanced Computer Science und Information Systems*. Malang, Indonesia: IEEE, 257–262.
- ODEN, N., 1995, Adjusting Moran's I for Population Density. *Statistics in Medicine*, 14 (1), 17–26.
- ORD, J. & GETIS, A., 1995, Local Spatial Autocorrelation Statistics: Distributional Issues und an Application. *Geographical Analysis*, 27 (4), 286–306.
- ORD, J. & GETIS, A., 2012, Local Spatial Heteroscedasticity (LOSH). *The Annals of Regional Science*, 48 (2), 529–539.
- QUESNOT, T. & ROCHE, S., 2015, Platial or Locational Data? Toward the Characterization of Social Location Sharing. *Proceedings of the Annual Hawaii International Conference on System Sciences*, 1973–1982.
- RITZER, G., DEAN, P. & JURGENSON, N., 2012, The Coming of Age of the Prosumer. *American Behavioral Scientist*, 56 (4), 379–398.
- ROCHE, S., 2016, Geographic Information Science II. *Progress in Human Geography*, 40 (4), 565–573.
- SCHEIDER, S. & JANOWICZ, K., 2014, Place Reference Systems. *Applied Ontology*, 9 (2), 97–127.
- Sengstock, C. und Gertz, M., 2012, Latent Geographic Feature Extraction from Social Media. *In: Proceedings of the 20th International Conference on Advances in Geographic Information Systems*. New York, NY: ACM Press, 149–158.
- SHELTON, T., POORTHUIS, A. & ZOOK, M., 2015, Social media und the city: Rethinking urban socio-spatial inequality using user-generated geographic information. *Landscape und Urban Planning*, 142, 198–211.
- SHEN, C., LI, C. & SI, Y., 2016, Spatio-Temporal Autocorrelation Measures for Nonstationary Series: A New Temporally Detrended Spatio-Temporal Moran's Index. *Physics Letters, Section A: General, Atomic und Solid State Physics*, 380 (1–2), 106–116.
- SHORTRIDGE, A., 2007, Practical Limits of Moran's Autocorrelation Index for Raster Class Maps. *Computers, Environment und Urban Systems*, 31 (3), 362–371.
- STEFANIDIS, A., CROOKS, A. & Radzikowski, J., 2013, Harvesting Ambient Geospatial Information from Social Media Feeds. *GeoJournal*, 78 (2), 319–338.
- STEIGER, E., RESCH, B., DE ALBUQUERQUE, J. & ZIPF, A., 2016, Mining und Correlating Traffic Events from Human Sensor Observations with Official Transport Data Using Self-Organizing-Maps. *Transportation Research Part C: Emerging Technologies*, 73, 91–104.
- TIEFELSDORF, M. & BOOTS, B., 1997, A Note on the Extremities of Local Moran's Iis und Their Impact on Global Moran's I. *Geographical Analysis*, 29 (3), 248–257.
- TIEFELSDORF, M., GRIFFITH, D. & BOOTS, B., 1999, A Variance-Stabilizing Coding Scheme for Spatial Link Matrices. *Environment und Planning A*, 31 (1), 165–180.
- TUFEKCI, Z., 2014, Big Questions for Social Media Big Data: Representativeness, Validity und Other Methodological Pitfalls. *In: E. Adar und P. Resnick, eds. Proceedings of the Eighth International AAAI Conference on Weblogs und Social Media*. Ann Arbor, MI: The AAAI Press, 505–514.
- WALDHÖR, T., 1996, The Spatial Autocorrelation Coefficient Moran's I Under Heteroscedasticity. *Statistics in Medicine*, 15 (7–9), 887–892.
- WALTER, S., 1992a, The Analysis of Regional Patterns in Health Data. I. Distributional Considerations. *American Journal of Epidemiology*, 136 (6), 730–741.
- WALTER, S., 1992b, The Analysis of Regional Patterns in Health Data. II. II. The Power to Detect Environmental Effects. *American Journal of Epidemiology*, 136 (6), 742–759.
- WENDER, K., HAUN, D., RASCH, B. & BLÜMKE, M., 2002, Context Effects in Memory for Routes. *In: C. Freksa, W. Brauer, C. Habel, und K. Wender, eds. Spatial Cognition III*. Tutzing: Springer, 209–231.
- WESTERHOLT, R., 2018, The Impact of Different Statistical Parameter Values between Point Based Datasets when

- Assessing Spatial Relationships. *In: Proceedings of the AGILE'2018 International Conference on Geographic Information Science*. Lund, Sweden: AGILE.
- WESTERHOLT, R., RESCH, B., MOCNIK, F.-B. & HOFFMEISTER, D., 2018, A statistical test on the local effects of spatially structured variance. *International Journal of Geographical Information Science*, 32 (3), 571–600.
- WESTERHOLT, R., RESCH, B. & ZIPF, A., 2015, A Local Scale-Sensitive Indicator of Spatial Autocorrelation for Assessing High- und Low-Value Clusters in Multiscale Datasets. *International Journal of Geographical Information Science*, 29 (5), 868–887.
- WESTERHOLT, R., STEIGER, E., RESCH, B. & ZIPF, A., 2016. Abundant Topological Outliers in Social Media Data and Their Effect on Spatial Analysis. *PLOS ONE*, 11 (9), e0162360.

# Requirements of road space models in the context of automated driving

Benedikt SCHWAB<sup>1</sup> and Prof. Dr. rer. nat. Thomas H. KOLBE<sup>2</sup>

<sup>1</sup> AUDI AG · 85045 Ingolstadt  
E-Mail: benedikt.schwab@audi.de

<sup>2</sup> Lehrstuhl für Geoinformatik · Technische Universität München · Arcisstraße 21 · 80333 München  
E-Mail: thomas.kolbe@tum.de

## 1 Introduction

The automation of the driver will have an impact on several aspects of society. This technology has the potential to reduce traffic accidents, shift the concept of privately-owned cars towards mobility as a service and consequently lower the total need of vehicles. However, the development challenges not only automobile manufacturers and mobility provider companies, but also legislators and certification services. One of the main questions is the verification of the correct functioning and the testing processes during the development phase.

As the number of different traffic scenarios is exceptionally high, the testing of incrementally new software versions cannot be based on real test drives. Exposing automated driving functions to a simulated environment enables reproducible, variable and efficient testing of the specific function or component. Although simulation is not new in the automotive development process, its application has been primarily focused on supporting the testing of controlling functions with a limited understanding of the vehicle's environment. However, realizing automated driving requires a much more comprehensive understanding of the environmental situation and thus a simulation thereof.

While the generation of fictional environments is suited for some test cases, there also exist conditions, in which the comparison to real environments is inevitable. This includes general questions, as finding the balance between testing in virtual or real environments, but also specific development tasks, such as the validation of sensor models. Thus, the development of high-precision georeferenced street space and city models is necessary. In order to build up georeferenced testbeds, already available and adequate geospatial data is combined with dedicated road surveys and then prepared for road space models. Depending on the input formats of the utilized simulators semantic and syntactic transformations may be required. Testing automated driving functions will pose requirements on each of those process steps.

## 2 Summary of the presentation content

### 2.1 Test concepts of automated driving functions

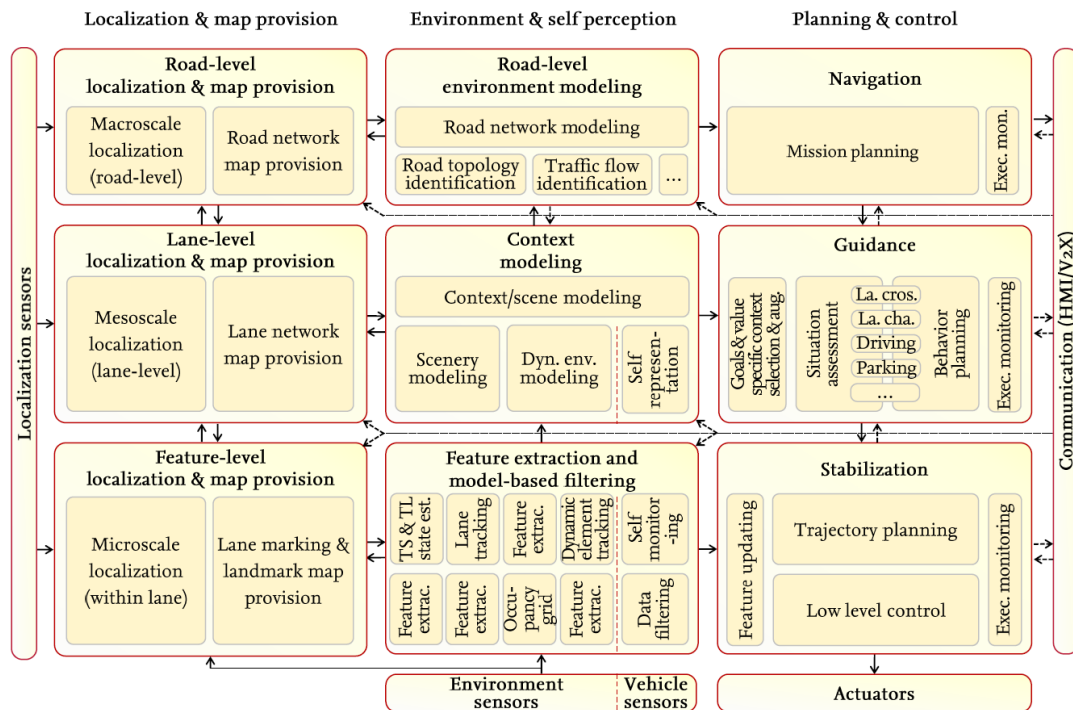
General requirements can be derived from the testing concepts for automated driving. This includes overall testing criteria, but also the different stages from testing pure software functions towards the testing of complete vehicles (SCHULDT 2017). To support these testing stages submicroscopic simulators are used for simulating driving behavior, sensors and vehicle dynamics (VON NEUMANN-COSEL et al. 2014). Figure 1 shows the submicroscopic simulator *Virtual Test Drive*, which provides interfaces to expose functions under development to the simulated environment.



**Fig. 1:** 3D-visualization of the submicroscopic traffic and vehicle simulator *Virtual Test Drive* (VTD) and a real-time electromagnetic wave propagation simulation of Car-to-Car-Communication within VTD (SCHILLER et al. 2014).

## 2.2 Use cases and requirements

Although there exist common requirements regarding road space models, some requirements are specific to the component or function under test. Therefore, Figure 2 shows a generic system architecture of an automated vehicle comprising the different automated driving functions. The architecture is structured in three layers. Navigational decisions are made on the strategic layer (top). Behavior planning takes place on the tactical layer (middle) and the operational layer (bottom) serves for speed control and course keeping. Furthermore, localization and perception related functions are structured in dedicated columns (ULBRICH et al. 2017).



**Fig. 2:** Generic structure of a functional system architecture for an automated vehicle (ULBRICH et al. 2017).

Requirements for road space models can be derived from the function, which needs to be tested. For example, testing microscale localization within the lane will require the representation of road markings, as the localization algorithms on the operational layer will use this information as input (MATTHAEI 2015). Another example constitutes the simulation of the different environment sensors, which are installed in an automated vehicle. This includes the modeling of LiDAR, camera, radar and ultrasound sensors.

## Literature

- MATTHAEI, R. (2015), Ein Wahrnehmungsgestützte Lokalisierung in fahstreifengenauen Karten für Assistenzsysteme und automatisches Fahren in urbaner Umgebung. In: Berichte aus der Fahrzeugtechnik.
- SCHILLER, M., KERN, A., KNOLL, A. (2014), Real-time Electromagnetic Wave Propagation using OptiX for Simulation of Car-to-Car-Communication. In: NVIDIA GTC 2014, San Jose.
- SCHULDT, F. (2017), Ein Beitrag für den methodischen Test von automatisierten Fahrfunktionen mit Hilfe von virtuellen Umgebungen. Universitätsbibliothek Braunschweig.
- ULBRICH, S., RESCHKA, A., RIEKEN, J., ERNST, S., BAGSCHIK, G., DIERKES, F., NOLTE, M. & MAURER, M. (2017), Towards a Functional System Architecture for Automated Vehicles. In: arXiv, 1703.08557v2 [cs.SY].
- VON NEUMANN-COSEL, K., DUPUIS, M., WEISS, C. (2009), Virtual test drive-provision of a consistent tool-set for [d, h, s, v]-in-the-loop. In: Proceedings of the Driving Simulation Conference Monaco.

# High Definition Maps – so entstehen hochgenaue Referenzkarten

Christiane RADIES

3D Mapping Solutions · Abteilung HD Maps · Raiffeisenstr. 16 · 83607 Holzkirchen

E-Mail: christiane.radies@3d-mapping.de

## 1 Einführung in das Thema

High Definition Maps sind hochaufgelöste hochgenaue Referenzkarten, die ein digitales Abbild der Realität darstellen. Grundlagendaten sind Laserscandaten und Kamerabilder von Autobahnen, Prüfgeländen oder Innenstädten aus der kinematischen Vermessung. 3D Mapping Solutions wertet diese Daten mit Hilfe einer dafür speziell entwickelten Prozesskette aus und führt die Ergebnisse in einem Bestandsplan zusammen. Anschließend wird die Verkehrslogik, also die topologische Struktur aller Spuren, Verbindungen, Gehwege, Radwege etc. anhand der Beschilderung oder der Straßenmarkierungen ermittelt und in den Bestandsplan eingearbeitet. Die zusätzlich enthaltene Verkehrslogik ist Voraussetzung, um auf Basis von HD Maps Anwendungen der Automobilindustrie simulieren zu können. Exportformate für HD Maps sind zum Beispiel OpenDRIVE® oder ROAD5 for CarMaker by IPG Automotive. Der nachfolgende Bericht beschreibt den Prozess von der homogenisierten Punktwolke bis zur HD Map, die Verkehrslogik enthält und Basis für verschiedene Simulationsanwendungen ist.

## 2 Der Prozess

### 2.1 Objektextraktion

Ausgangsdaten sind homogenisierte Punktwolken. Dabei werden die Punktwolken für die einzelnen Gebiete über eine Trajektorienverbesserung auf einander angepasst. Mittels Schwellwert Segmentationsverfahren werden Objekte, wie Fahrspuren, Bordsteine oder Häuserkanten möglichst automatisch extrahiert (vgl. Abb. 1 und Abb. 2).



**Abb. 1:** 3D Mapping Solutions Scannerdaten.



**Abb. 2:** 3D Mapping Solutions Scannerdaten mit Objektextraktion.

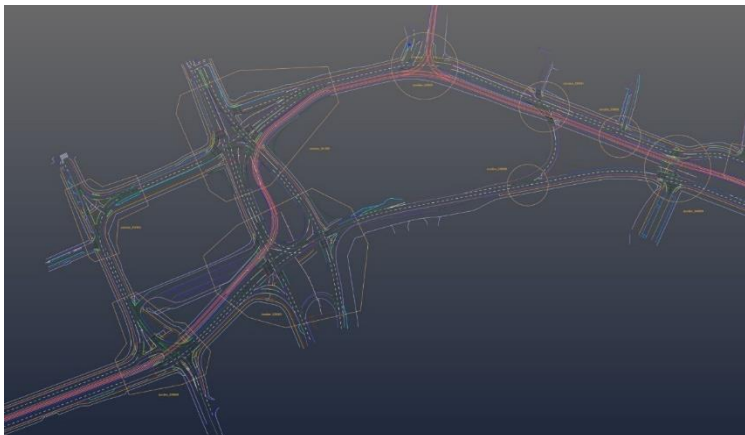
Verkehrsgebende Signale werden aus stereoskopisch aufgenommenen Kamerabildern mit der hauseigenen Software RoadView gewonnen. Dabei wird neben der Position der Signale, die Kategorie und der Richtungswinkel kartiert. (vgl. Abb. 3 und Abb. 4).



**Abb. 3 und Abb. 4:** Stereoskopische Kamerabilder.

## 2.2 Der Bestandsplan

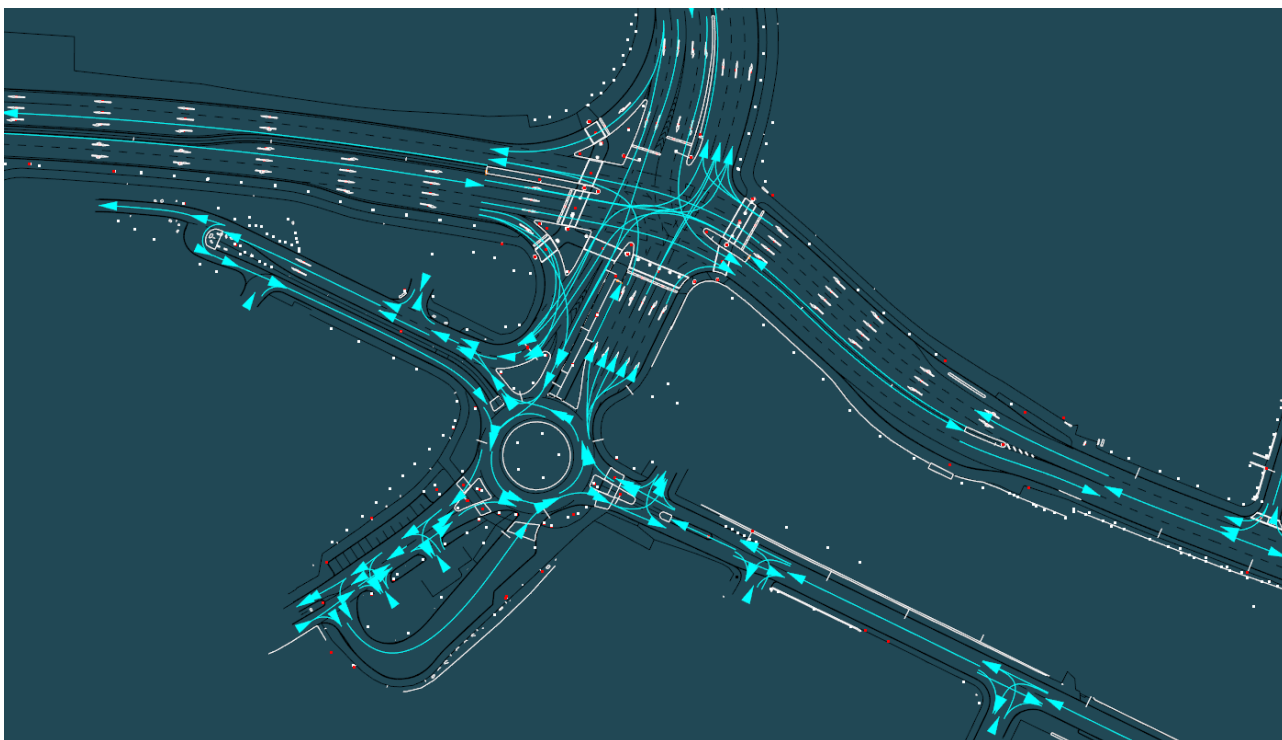
Die Zusammenführung der extrahierten Geometrien und Signale lässt den Bestandsplan Layer um Layer wachsen. Die nachfolgende Abbildung veranschaulicht die hohe Komplexität des Bestandsplans.



**Abb. 5:** Bestandsplan mit Spurgeometrien aus der Objektextraktion mit einer relativen Genauigkeit von 1-3 cm.

## 2.3 Die HD Map

Der Bestandsplan wird anschließend um ein logisches Verkehrsnetz mit Straßenabschnitten und Kreuzungen ergänzt, das semantische Information und topologische Information enthält.



**Abb. 6:** HD Map mit Verkehrslogik.

## Literatur

OpenDRIVE®, Open File Format. Standardisierung by ASAM <https://www.asam.net/standards/detail/opendrive/>  
Download Spezifikation OpenDRIVE 1.4 H <http://www.opendrive.org/>



---

# Erweiterung der Umfeldwahrnehmung automatisierter Fahrzeuge durch intelligente Infrastruktur

Maximilian SCHNETTLER

fortiss GmbH - Forschungsinstitut des Freistaats Bayern für softwareintensive Systeme und Services · Guerickestr. 25 ·  
80805 München  
E-Mail: schnettler@fortiss.org

## 1 Einführung

Staus und Unfälle gehören auf deutschen Autobahnen zum Alltag und bewirken einen enormen gesamtwirtschaftlichen Schaden, unabhängig der negativen Auswirkungen auf den individuellen Fahrer und dessen Sicherheit und Komfort im Straßenverkehr. Hochautomatisiertes Fahren und eine intelligente Verkehrsflusssteuerung sollen dem zunehmenden Individualverkehr und dessen Auswirkungen entgegenwirken. Zusätzlich versprechen sie generell eine Steigerung von Sicherheit und Komfort.

Automatisierte Fahrzeuge benötigen ein umfassendes und genaues Modell ihrer Umwelt, um erfolgreich die nächsten Fahrmanöver zu planen und durchzuführen. Auch eine intelligente Verkehrssteuerung, die möglichst exakt die empfohlenen Fahrmanöver für jedes Fahrzeug ermitteln soll, erfordert ein positionsgenaues Abbild aller Verkehrsteilnehmer. Dies umfasst nicht nur hochgenaues Kartenmaterial, sondern zusätzlich auch alle statischen und dynamischen Objekte in der Umgebung. Dieses Abbild des gesamten Umfeldes in einem Modell in Echtzeit wird allgemein als Digitaler Zwilling bezeichnet.

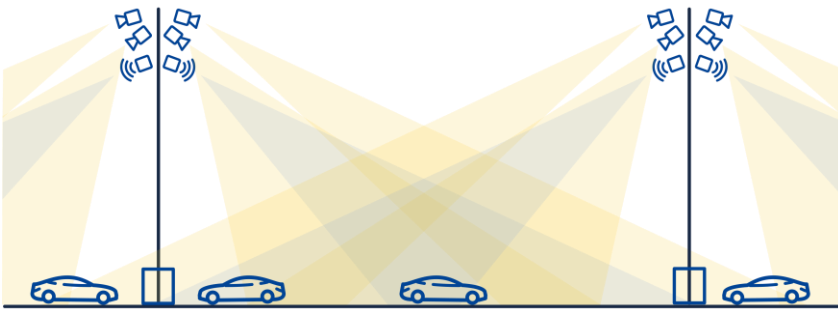
Jedes Fahrzeug mit bordeigenen Sensoren erzeugt ein eigenes Modell seines Umfeldes. Allerdings ist dieses durch die Reichweite der Sensoren und Verdeckungen durch andere Fahrzeuge oder statische Objekte begrenzt. Diese Limitierung ihres Wahrnehmungshorizontes müssen automatisierte Fahrzeuge in ihrer Manöverplanung berücksichtigen, wodurch sich die Fahrzeuge nur sehr eingeschränkt und passiv im Straßenverkehr verhalten können. Daher ist es insbesondere in Umgebungen mit hohen Geschwindigkeiten oder vielen Verdeckungen notwendig das Umgebungsmodell des einzelnen Fahrzeugs über dessen Sensorkapazitäten hinaus zu erweitern. Dies lässt sich zum einen durch die Vernetzung der Fahrzeuge und des Austausches der individuellen Umgebungsinformationen erreichen. Die Qualität und Zuverlässigkeit des daraus resultierenden digitalen Zwillings ist allerdings stark abhängig von der Sensorausstattung und der Verbreitung der vernetzten Fahrzeuge. Insbesondere in Szenarien mit Mischverkehr und einer geringen Verbreitung hochautomatisierter Fahrzeuge kann die Vollständigkeit des Umgebungsmodells nicht garantiert werden.

Abhilfe kann hier die Erfassung des Verkehrsgeschehens mittels Infrastruktursensorik und der Erstellung des digitalen Zwillings in einem zentralen Backend schaffen. Ein spezifisches Systemdesign und hochwertige Sensorik ermöglicht eine hohe Zuverlässigkeit und kontrollierte Genauigkeit des digitalen Zwillings, ohne Anforderungen an die Ausstattung aller Fahrzeuge zu stellen. Durch Redundanz von Sensorik und verringerte Verdeckungen durch eine erhöhte Beobachtungsperspektive kann ein robustes, vollumfängliches und genaues Umgebungsmodell erreicht werden. Die Erforschung der Möglichkeiten und Herausforderungen eines solchen Systems sowie der Aufbau eines funktionsfähigen Prototyps wird im Rahmen des Forschungsprojektes Providentia behandelt.

## 2 Zusammenfassung des Vortragsinhalts

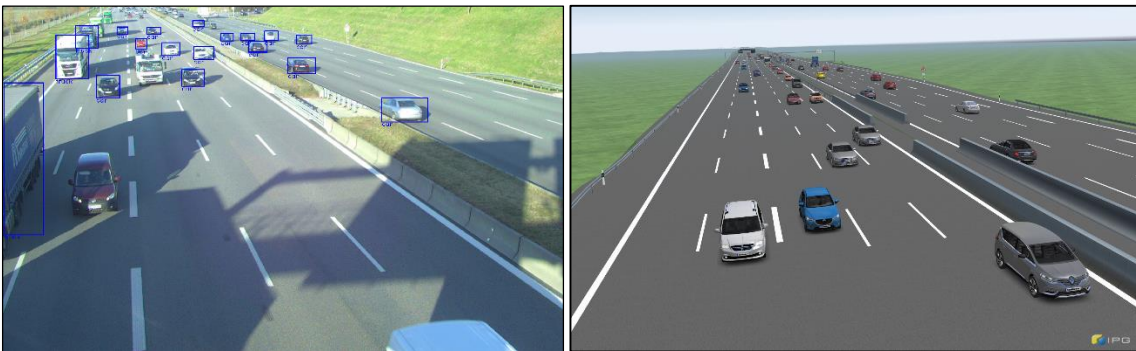
### 2.1 Das Forschungsprojekt Providentia

Ziel des Projektes Providentia ist die Umfeldwahrnehmung von automatisierten Fahrzeugen durch intelligente Infrastruktur zu erweitern und dadurch komfortables und sicheres Fahren auf Autobahnen zu ermöglichen. Dafür wird das Verkehrsgeschehen mit entlang der Autobahn installierten Kameras und Radarsensoren erfasst (siehe Abb. 1) und in Echtzeit ein anonymisierter digitaler Zwilling über ein zentrales Backend mittels 5G-Mobilfunktechnik den Fahrzeugen zur Verfügung gestellt (siehe Abb. 2). Dieser Digitale Zwilling bildet neben der Umfeldenerweiterung die Grundlage für diverse weitere Anwendungsmöglichkeiten, wie bspw. eine intelligente Verkehrsflusssteuerung.



**Abb. 1:** Schematischer Aufbau der Infrastruktursensorik: mit Kameras und Radarsensoren ausgestattete Messpunkte decken die gesamte Fahrbahn redundant aus verschiedenen Blickwinkeln ab.

Providentia ist ein vom Bundesministerium für Verkehr und digitale Infrastruktur im Rahmen der „Förderrichtlinie automatisiertes und vernetztes Fahren auf digitalen Testfeldern in Deutschland“ gefördertes Projekt und im Digitalen Testfeld Autobahn A9 angesiedelt. Neben fortiss als Konsortialführer sind technologisch führende Unternehmen aus der gesamten Entwicklungskette des automatisierten Fahrens und des 5G-Mobilfunks vertreten (BMW, Cognition Factory, Elektrobit, Huawei, IPG, Rohde & Schwarz, Deutsche Telekom). Erwartet werden schrittmachende Erkenntnisse sowohl für die Technik und anstehenden Normungen des Mobilfunks der nächsten Generation (5G), für die Erzeugung hochzuverlässiger Abbilder der Realität in der Infrastruktur durch Sensorfusion, für die Informationsversorgung vernetzter hochautomatisierter Fahrzeuge und für die Virtualisierung komplexer Verkehrsszenarien.



**Abb. 2:** Ziel des Projektes Providentia: Visualisierung des aktuellen Verkehrsgeschehens in Form eines digitalen Zwillingings in Echtzeit (links: Kamerabild mit Objektdetektionen, rechts: Illustration des digitalen Zwillingings im hochgenauen Straßenmodell).

Hierzu wird im Rahmen des Projektes ein intelligentes Infrastruktursystem entwickelt (HINZ 2017) und auf über zwei Kilometern Länge entlang der Autobahn A9 bei Garching-Süd aufgebaut. Ein erstes Teilstück von knapp einem Kilometer ist bereits aufgebaut und befindet sich im Testbetrieb. Ein solches Multi-Sensor-System erfordert die Erforschung und Entwicklung neuartiger Software und Algorithmen, um einen zuverlässigen digitalen Zwilling erzeugen zu können. Eine exakte Kalibrierung der hohen Anzahl an verteilten Sensoren bildet die Grundlage zur weiteren Verwendung der Sensordaten. Hierfür sind Algorithmen notwendig, die im laufenden Betrieb verwendet werden können und ohne spezielle Kalibrierungstargets auskommen (bspw. DUBSKÁ 2014). Mittels Neuronaler Netze werden in den Kamerabildern alle Verkehrsteilnehmer detektiert (siehe Abb. 2 links). Dies stellt besondere Anforderungen an die Qualität und Echtzeitfähigkeit dieser Objektdetektion, um ein genaues Abbild erzeugen zu können. Die detektierten Objekte aus den Kamera- und Radardaten werden anschließend in jedem Messpunkt fusioniert. Aufgrund der Vielzahl an Sensoren und erkannten Objekten sind spezielle Fusionsalgorithmen notwendig (siehe bspw. CLARK 2006). Im zentralen Backend werden dann die Daten aller Messpunkte im digitalen Zwilling kombiniert. Dabei werden keinerlei personenbezogene Daten verarbeitet oder gespeichert – der digitale Zwilling erlaubt keinerlei Rückschlüsse auf die erfassten Verkehrsteilnehmer. Dieser digitale Zwilling bildet dann die Grundlage zur Wahrnehmungserweiterung der Fahrzeuge und zur Prädiktion des zukünftigen Verhaltens der Verkehrsteilnehmer.

Zusätzlich wird die Möglichkeit der Integration von Sensordaten automatisierter Fahrzeuge in das Providentia System erforscht. Solange sich ein mit Sensoren ausgestattetes Fahrzeug im Testfeld befindet, kann dessen internes Umfeldmodell über die 5G Mobilfunkschnittstelle dem Backend bereitgestellt werden (siehe Abb. 3). Dadurch kann die Qualität und Genauigkeit des digitalen Zwillingings weiter verbessert werden.



**Abb. 3:** Integration der LiDAR-Messdaten des fortiss Forschungsfahrzeugs in das Sensorsystem von Providentia.

## 2.2 Integration hochgenauer Straßenmodelle

Eine hochgenaue Straßenkarte bildet die Grundlage für automatisiertes Fahren. Ein möglichst präzises Modell der Straße und Umgebung mit allen statischen Landmarken ermöglicht eine genaue Eigenlokalisierung der Fahrzeuge. Das Unternehmen HERE ist ein beispielhafter Anbieter von solchen hochgenauen Karten für das automatisierte Fahren und aktuelle Serienfahrzeuge. Das Providentia-System reichert ein solches statisches Umgebungsmodell mit den vorhandenen dynamischen Objekten an, in diesem Fall den Verkehrsteilnehmern. Dadurch können die detektierten Verkehrsobjekte sehr präzise im digitalen Zwilling verortet und den Fahrzeugen an der richtigen Position bereitgestellt werden.

Im Projekt wird eine hochgenaue Karte der Firma 3D-Mapping verwendet, die die Autobahn A9 von Garching bis München-Schwabing abdeckt. Dadurch werden Testfahrten mit automatisierten Fahrzeugen in Bereichen mit und ohne Infrastruktursensorik ermöglicht.

## 2.3 Potentielle Anwendungen und Ausblick

Primäres Ziel des Forschungsprojekts Providentia ist die Generierung eines digitalen Zwillings mittels Infrastruktursensorik und dessen Bereitstellung für automatisierte Fahrzeuge, um deren Umfeldwahrnehmung zu erweitern. Dadurch werden die limitierten Wahrnehmungsreichweiten der Fahrzeuge aufgelöst und ein prinzipiell unbegrenzter Sichtbereich ermöglicht. Die Fahrzeuge können mit dem digitalen Zwilling sowohl zeitlich als auch räumlich weiter vorausschauen und bessere Entscheidungen für zukünftige Fahrmanöver treffen.

Zusätzlich bietet das aufgebaute Providentia-System eine vielversprechende Basis für weitere Anwendungsfelder jenseits der Wahrnehmungserweiterung. Die hochgenaue Infrastruktursensorik kann beispielsweise als Referenzsystem zur Entwicklung und Validierung von Fahrerassistenzsystemen und automatisierten Fahrfunktionen verwendet werden. Der digitale Zwilling kann dabei als Vergleich für das vom Fahrzeug erstellte Umgebungsmodell dienen. Eine weitere Anwendung bildet die Erstellung eines wissenschaftlichen Datensatzes für eine Vielzahl von Forschungsbereichen. Zum Beispiel können die anonymisierten Bewegungsdaten der Verkehrsobjekte für die Parametrierung von Verhaltensmodellen menschlicher Fahrer und automatisierter Fahrzeuge verwendet werden. Diese Modelle sind notwendig, um das Verhalten der Verkehrsteilnehmer in der Zukunft zu schätzen und in die Planung der Fahrmanöver einzubeziehen. Außerdem können durch die dauerhafte Installation wichtige Rückschlüsse auf den gesamten Verkehrsfluss und dessen intelligente Beeinflussung getroffen werden. In Zukunft kann der digitale Zwilling dafür verwendet werden jedem Fahrzeug eine optimale, fahrspurgenaue Route vorzuschlagen, um eine optimale Ausnutzung der verfügbaren Infrastruktur zu erreichen. All diese Anwendungen bilden wichtige Schritte auf dem Weg zum sicheren und komfortablen automatisierten Fahren.

## Literatur

- HINZ, G. et al. (2017), Designing a far-reaching view for highway traffic scenarios with 5G-based intelligent infrastructure. In: 8. Tagung Fahrerassistenz.
- DUBSKÁ, M. et al. (2014), Automatic Camera Calibration for Traffic Understanding. In: Proceedings of the British Machine Vision Conference, 42.1-42.12.
- CLARK, D. (2006), Multiple Target Tracking with The Probability Hypothesis Density Filter.

# GPU-basierte Simulation – Stand der Technik und Anwendungspotenziale

Rüdiger Westermann

Fakultät für Informatik · Technische Universität München · Boltzmannstraße 3 · 85748 Garching  
E-Mail: westermann@tum.de

## 1 Einführung ins Thema

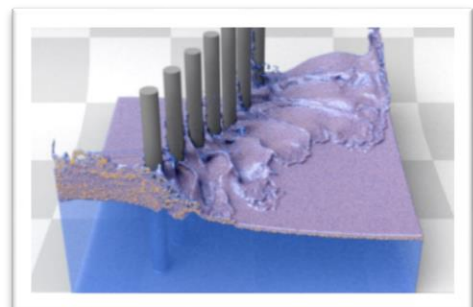
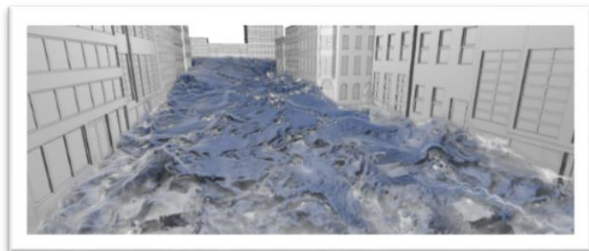
Moderne Grafikkarten (Graphics Processing Units – GPUs) zeichnen sich durch einen hohen Grad an Parallelismus und ein extrem schnelles Speicher-Interface aus. Mit bis zu 32 GB lokalem Speicher können selbst große Datenmenge effizient und ohne Zugriff auf den CPU-Speicher bearbeitet werden. Diese Entwicklung hat dazu geführt, dass GPUs schon lange nicht mehr ausschließlich für Grafik-Anwendungen genutzt werden sondern zunehmend auch als Hochleistungsrechner im Bereich der numerischen Simulationsrechnung zum Einsatz kommen. Sowohl im Bereich der Gitter- und Partikel-basierten numerischen Simulationsrechnung haben sich GPUs als extrem effiziente Rechnerarchitekturen positioniert, und haben hier insbesondere zu interaktiven Simulationen komplexer physikalische Phänomene geführt.

## 2 Zusammenfassung des Vortragsinhalts

Im vorliegenden Beitrag werde ich näher auf die Verwendung von GPUs im Bereich der Gitter- und Partikel-basierten Strömungssimulation eingehen. Ich werde insbesondere die effiziente Lösung der inkompressiblen Navier-Stokes-Gleichung auf regulären Gitterstrukturen demonstrieren und spezielle Möglichkeiten zur Handhabung von großen Teilchenmengen diskutieren. Hierbei werde ich sowohl auf Aspekt der Parallelisierung auf GPUs als auch auf algorithmische Aspekte zur Steigerung der Effizienz moderner numerischer Löser eingehen. Abschließend werde ich am Beispiel einer interaktiven Simulation von Erosionseffekten durch die mechanischen Einflüsse von Wasser das Potenzial von GPUs im Bereich der Geo-Simulation aufzeigen.

## Literatur

- DICK, C. & WESTERMANN, R. (2011), A Real-Time Multigrid Finite Hexahedra Method for Elasticity Simulation using CUDA. In: *Simulation Modelling Practice and Theory* 19(2).
- WEBER, D. & BENDER, J. & SCHNOES, M. & STORK, A. & FELLNER, D. (2013), Efficient GPU Data Structures and Methods to Solve Sparse Linear Systems in Dynamics Applications. In: *Computer Graphics Forum* 32(1).
- DICK, C. & ROGOWSKY, M. & WESTERMANN, R. (2016), Solving the Fluid Pressure Poisson Equation Using Multigrid—Evaluation and Improvements. In: *IEEE Transactions of Computer Graphics and Visualization* 22(11).
- WU, J. & DICK, C. & WESTERMANN, R. (2016), A System for High-Resolution Topology Optimization. In: *IEEE Transactions on Visualization and Computer Graphics* 22(3).
- FERSTL, F. & ANDO, R. & WOJTAN, C. & WESTERMANN, R. & THUREY, N. (2016), Narrow Band FLIP for Liquid Simulations. In: *Computer Graphics Forum* 35(2).



# Using Geodata for Simulating Urban Traffic – Current Research in the Field of Traffic Engineering and Control

Georgios GRIGOROPOULOS, Andreas KELER, Heather KATHS, Matthias SPANGLER and Fritz BUSCH

Chair of Traffic Engineering and Control · Technical University of Munich · Arcisstraße 21 · 80333 Munich

E-Mail: george.grigoropoulos@tum.de, andreas.keler@tum.de, heather.kaths@tum.de, matthias.spangler@tum.de, fritz.busch@tum.de

## 1 Introduction

Daily mobility has manifold appearances, especially in urban environments. Numerous modes of transport, different types of traffic participants and differing travel demands and supplies occur within often restricted urban infrastructures. One key for understanding these complex phenomena is the representation of not only space and time, but also of the interactions of different traffic participants with each other and the specific traffic control. We are able to represent recorded trajectories of traffic participants, traffic light signals and traffic measures, together with urban infrastructural elements. From this data input, we can conduct microscopic traffic flow simulations, which may deliver further insights into complex urban traffic. Modelling and simulating vulnerable road users is challenging due to frequent movements on not assigned paths and disobeying traffic rules (red light violations). Therefore, higher quality geodata is crucial for the traffic flow simulation quality and its possible visualization in physical vehicle and bicycle simulators.

## 2 The usage of geodata in the transportation domain – focus on mapping and modelling the urban environment

Starting from modelling the transport infrastructure, the respectively used geodata has crucial importance on analyses and simulation outcomes in various spatial and temporal scales. Besides this, dynamic spatiotemporal information such as trajectories of traffic participants or traffic light states serve for modelling and simulating microscopic traffic flow. As an extension, we can validate our simulation results within simulator studies, and, extend this approach towards a selection of tools for active mobility research, which implies the inclusion of simulator study test subjects and the resulting qualitative and quantitative data.

### 2.1 Modeling static and dynamic components of urban traffic – transport infrastructure and trajectories of different traffic participants

Modeling static and dynamic components of urban traffic can start with traffic observations. Knowledge on partitions of traffic situations can already serve as an input for calibrating a model of microscopic traffic flow. Most of the accessible traffic information should imply a spatial component, of single positions of inductive loops, sequences of positions from GNSS positioning devices or of fixed positions, angles and heights when using video devices.

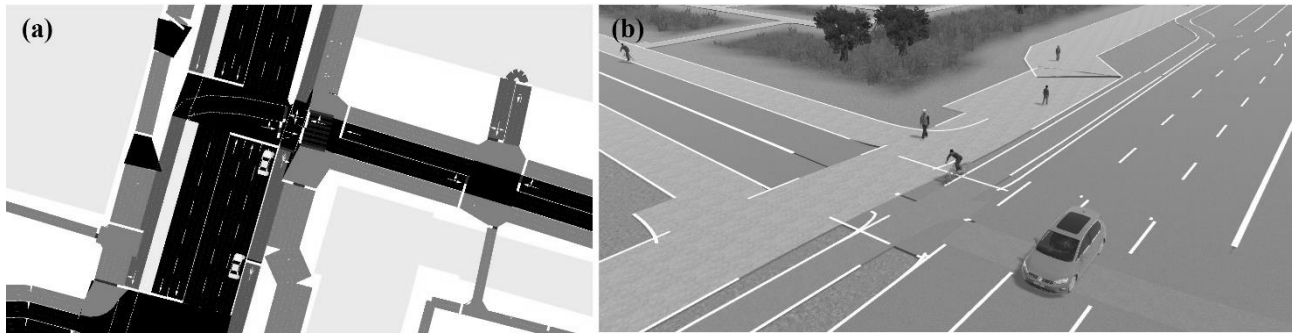
TWADDLE et al. (2014) use automated video analysis for the evaluation of bicycle movement and interaction. The sensors providing these videos might be RGB cameras installed at various fixed locations within the urban environment, favorably at buildings showing a visible road intersection. Via procedures coming from image processing and computer vision, we are able to extract every trajectory with the additional attributes speed, acceleration and road user type. Subsequently the data can be georeferenced and post-processed for enriching the trajectories with available infrastructural information as pictured in Figure 1.



**Fig. 1:** Trajectories of vehicle drivers, bicyclists and pedestrians at one road intersection in Munich.

Starting with analyzing the data in Figure 1, we can, at a certain state, distinguish between typical, frequent and unusual travel behaviors of every type of traffic participant. The modelling of these behavior comes together with compromises to the used software and to the achievable complexity of implemented models. This continues until modelling specific interactions between traffic participants, traffic signaling, and, even predictions resulting from them. TWADDLE & BUSCH (2018) for example are proposing promising models for predicting responses to red signals, together with predicting the type of left-hand turns. Nevertheless, predicting infrastructure selection and the direction of travel are difficult.

It is possible to integrate external models in SUMO (TWADDLE, GRIGOROPOULOS & BUSCH 2016). This guarantees the modelling of complex urban traffic more precisely. After extracting or modelling static transport infrastructure, we include traffic light signaling, flows and assigned road segments with this software before performing the first simulations (LOPEZ et al. 2018, BEHRISCH et al. 2011, KRAJZEWICZ et al. 2012).



**Fig. 2:** Modelling process of (a) the road network in SUMO, and, (b) its subsequent representation in 3D after converting into the OpenDRIVE format.

As pictured in Figure 2, we are not only converting the transport infrastructure (Figure 2a) into its 3D representation in the OpenDRIVE format (OROZCO IDROBO 2015) via the software DYNAAanimation<sup>1</sup> by TESIS (Fig. 2b), but also translating maneuvers of other traffic participants and states of traffic light signals. Translating infrastructure and traffic situations into representations resembling our perceived environment is the base for conducting our current bicycle simulator studies (KELER et al. 2018). The base for guaranteeing realistic traffic flows of vehicle drivers, bicyclists and pedestrians is applying a calibration and validation step that might use the most recent traffic information as coming from video observations (Fig. 1) or recorded traffic light signal, as it is for our cases coming from the city of Munich.

## 2.2 Simulating vulnerable road users – importance of bicycle traffic and pedestrian areas

Modeling and simulating microscopic traffic flow is not only a possibility for understanding traffic-related problems or evaluating the implemented traffic control strategies, it can also serve as a tool for planning the installation of additional and novel infrastructural elements, such as bicycle highways (FGSV 2014). By modifying widths and ordering of bicycle lanes, together with redesigning every affected road intersection, it is possible to test similar or varying flows of bicyclists together with variation in delays, congestion lengths and stop times. By varying bicycle type compositions, traffic conditions and traffic control strategies, we can evaluate the possible effect of installing bicycle highway infrastructure within complex built urban environments. Measures for evaluating the infrastructural modifications, as pictured with the two networks in Figure 3, can focus on traffic safety and traffic quality. One important component within the whole procedure is modeling the behavior bicyclists (TWADDLE, SCHENDZIELORZ & FAKLER 2014), together with respective speed, acceleration, and deceleration (TWADDLE & GRIGOROPOULOS 2016).

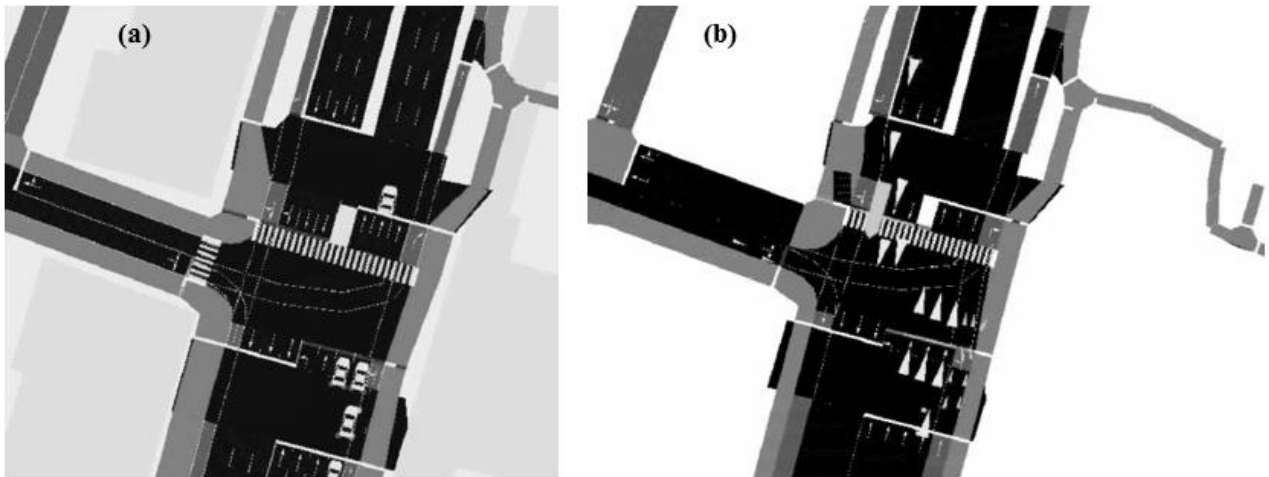
As pictured in Figure 4b, it is possible to visually investigate 3D representations of simulated traffic through a variety of sensors. The possibly 2 most common sensors of a Bicycle Simulator are speed and steering (KELER et al. 2018). Design and hardware selection for a bicycle simulator may imply many different sensors that guarantee different accuracies (DIALYNAS, HAPPEE & SCHWAB 2019), always depending on the implemented movement model for the bicycle simulator application (LEE et al. 2017). Depending on the latter we can guarantee more or less realism or accuracy in whether animation or usability. From the early 2000s until the late 2010s there has been numerous research on different bicycle simulators used as tool for analyzing various traffic-related aspects (KWON et al. 2001, HE, FAN & MA 2005, YIN & YIN 2007, SUN & QING 2018, SCHRAMKA et al. 2018).

Another approach in including realistic appearance of complex urban traffic is acquiring trajectories from video observations as pictured in Figure 1 and Figure 4a. This is apparently an ever growing popular application, when perceiving image processing and computer vision as the associated research fields. One example is the company DataFromSky<sup>2</sup>

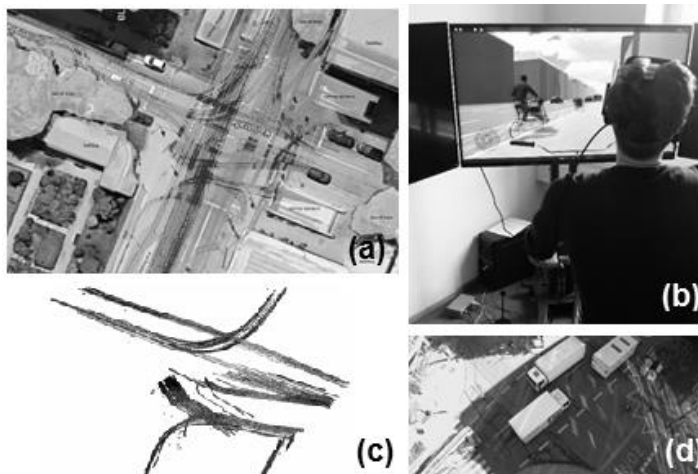
<sup>1</sup> <https://www.thesis.de/dynaanimation/>

<sup>2</sup> <http://datafromsky.com/>

serving numerous users with highly accurate trajectories from various differentiated (through classification) road users. Due to our current capacities of processing massive video data sets on the fly, we are able to link different observations through modern communication technologies. On the other hand, we can use this data for calibrating our modelled simulation networks.



**Fig. 3:** Modelled road intersection at Ludwigstraße-Theresienstraße with (a) recent state of the infrastructure, and, (b) with modified traffic control and bicycle infrastructure.

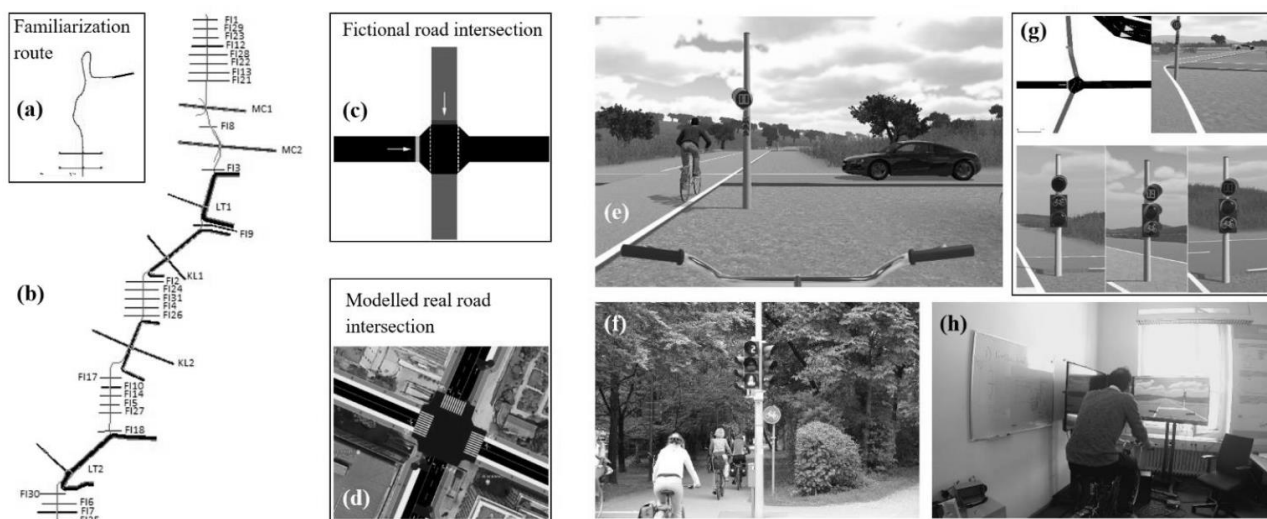


**Fig. 4:**

Tasks of trajectory extraction and video observation of modelled urban environments with (a) extracted conflict points between bicyclists and vehicle drivers from video trajectories of both traffic participants, (b) test subject following a designed bicycle highway at the modelled Ludwigstraße in Munich, (c) visualization of trajectory partitions of few frames of vehicle drivers and bicyclists at a mixed traffic road intersection, (d) video frame of heavy goods vehicles interacting with other traffic participants at a non-signalized intersection in Munich.

Resulting from the extracted trajectories, we are able to model the interaction between bicyclists and vehicle drivers as it is pictured in Figure 4c with the relation of vehicle and bicycle trajectory partitions of several frames. Besides the temporal selection, there is one important spatial component within every trajectory. Depending on the instantaneous velocities, movement directions and shapes of maneuvers, we are able to infer fields of view for every recorded trajectory. This is one of the most important attributes when inspecting interactions of traffic participants, and, the most important, when observing communication between every traffic participant. One option for evaluating the gained knowledge of timestamped visible spaces is the observation of specific maneuvers via VR glasses, as pictured in Figure 4b. For this option the field of view can be recorded directly and used for matching visible (moving) objects. This might also facilitate the, often expensive, computing of visibility polygons or volumes for every traffic participant. One bicycle rider or pedestrian might be influenced by numerous heavy-good vehicles that appear as fences at an urban intersection as pictured in Figure 4d. Assessing the visible parts in space might appear easier within a VR environment with this situation simulated at a bicycle simulator.

Besides assessing various traffic situations based on calibrated and validated simulation networks (with data on traffic flows and traffic light signaling), we can imply more usability into our case studies. This is due to evaluating novel traffic control signal displays via comparable test rides as for example the evaluation of different countdown timer displays for bicyclists as pictured in Figure 5. In this case study the comparability of results is more important than interacting with simulated traffic, as test subjects have to experience the same situations at specific positions of their trips. This is technically being realized via spatial trigger that do not only trigger the remaining countdown seconds and traffic light durations at specific locations, but also trigger predefined movements of other traffic participants.



**Fig. 5:** Specifications of the first experiment for evaluating countdown timer displays with (a) familiarization route, (b) whole route with 33 road intersections, (c) fictional road intersections, (d) modelled real road intersections, (e) resulting bike simulator visualization via DYNAanimation, (f) one installed countdown timer display in Munich, (g) respective visualizations of the ego vehicle via SumoConnect function between SUMO and DYNAanimation and available countdown timer displays and traffic lights, and, (h) test subject using the bike simulator.

### 2.3 Further applications – active mobility research and geodata for traffic emission modelling

The flexible usage of massive trajectory data from video observations, its usage for calibrating different network representations, and, the simulation of three different modes of traffic might serve as an extendable workflow for improving simulator studies. Moving away from inspecting user experience of triggered traffic signals and road users, we focus on experiencing simulated traffic and its perception by different test subjects. Besides novel approaches for simulator test data collection such as recording eye movements, gesture detection and field of view estimations, there are several possible extensions that result from varying the introduced workflows. One important variation is the inclusion of additional modes of transport that rely more on dynamic developments and trends, such as the users of Segway Personal Transporters, E-Scooters and Skateboards, together with respective changes of modes, which appear frequently within complex and busy urban environments of today's cities. This new perception of urban traffic is a challenge for classification and realistic simulation. Nevertheless, it is important to include the analyst as a participating part of physical simulators that allow additional qualitative evaluation of simulative results.

Besides these approaches that might support getting more insights into the connection between perceiving complex urban space and complex traffic situations, we are able to include further georeferenced data into our analyses. This can support knowledge discovery on unusual correlations as the connection of bicycling quality and traffic emissions. Traffic simulation can also be the starting point of traffic emission modelling. SO ET AL. (2017) propose an integrated simulation approach for estimating emissions based on more reliable vehicle performance measures. It consists of a microscopic traffic simulation model, a vehicle dynamics model, and, an emission estimation model.

Another field of applications consists of introducing location-based services for traffic participants that might increase traffic efficiency and traffic safety, as the Sitraffic SiBike application for bicyclists (GRIGOROPOULOS ET AL. 2018)

While selecting suitable data sources for relating further environmental aspects to traffic simulation results can be trivial, the semantic relations or possible matching strategies of this data to simulated traffic is often complex. This contributes to research on modelling complex urban relationships with various evaluation steps coming from different domains.

## References

- BEHRISCH, M., BIEKER, L., ERDMANN, J. & KRAJZEWICZ, D. (2011), SUMO–simulation of urban mobility: an overview. In: Proceedings of SIMUL 2011, The Third International Conference on Advances in System Simulation.
- DIALYNAS G., HAPPEE, R. & SCHWAB, A. L. (2019), Design and hardware selection for a bicycle simulator. *Mechanical Sciences*, 10, 1-10.
- FGSV (2014), Einsatz und Gestaltung von Radschnellverbindungen. Arbeitspapier, Arbeitsgruppe Straßenentwurf.
- GRIGOROPOULOS, G., TWADDLE, H., SPANGLER, M., HAGENBRING, M. & DÜSTERWALD, M. (2018), Evaluierung der dynamischen Grünen Welle für Radfahrer - Sitraffic SiBike - in Marburg. In: *Straßenverkehrstechnik*, 4.2018, 268-274.



- HE Q., FAN, X. & MA, D. (2005), Full bicycle dynamic model for interactive bicycle simulator. In: *Journal of Computing and Information Science in Engineering*, 5(4), 373-380.
- KELER, A., KATHS, J., CHUCHOŁOWSKI, F. E., CHUCHOŁOWSKI, M., GRIGOROPOULOS, G., SPANGLER, M., TWADDLE, H. A. & BUSCH, F. (2018), A bicycle simulator for experiencing microscopic traffic flow simulation in urban environments. In: *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 3020-3023.
- KRAJZEWICZ, D., ERDMANN, J., BEHRISCH M. & BIEKER, L. (2012), Recent development and applications of SUMO-Simulation of Urban Mobility. In: *International Journal on Advances in Systems and Measurements*, 5 (3&4), 128-138.
- KWON, D. S., YANG, G. H., LEE, C. W., SHIN, J. C., PARK, Y., JUNG, B., LEE, D. Y., LEE, K., HAN, S. H., YOO, B. H., WOHN, K. Y. & AHN, J. H. (2001), KAIST interactive bicycle simulator. In: *Proceedings 2001 ICRA - IEEE International Conference on Robotics and Automation (Cat. No.01CH37164)*, Seoul, South Korea, 2001, 3, 2313-2318.
- LEE, O., DIALYNAS, G., DEWINTER, J. C. F., HAPPEE, R. & SCHWAB A. L. (2017), Description of a model based bicycle simulator. In: *6th Annual International Cycling Safety Conference (ICSC'17)*, 21–22 September 2017, Davis, California, USA.
- LÓPEZ, P. A., BEHRISCH, M., BIEKER-WALZ, L., ERDMANN, J., FLÖTTERÖD, Y.-P., HILBRICH, R., LÜCKEN, L., RUMMEL, J., WAGNER, P. & WIEßNER, E. (2018). Microscopic Traffic Simulation using SUMO. In: *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 2575-2582.
- OROZCO IDROBO, A. M. (2015), Extension of the Geospatial Data Abstraction Library (GDAL/OGR) for OpenDRIVE Support in GIS Applications for Visualisation and Data Accumulation for Driving Simulators. Technical University of Munich.
- SCHRAMKA, F., ARISONA, S., JOOS, M. & ERATH, A. (2018), Development of a Virtual Reality Cycling Simulator. In: *Journal of Computers*, 13, 6, 603-615.
- SO, J., MOTAMEDIDEHKORDI, N., WU, Y., BUSCH, F. & CHOI, K. (2017), Estimating Emissions Based on the Integration of Microscopic Traffic Simulation and Vehicle Dynamics Model. In: *International Journal of Sustainable Transportation*, 12 (4), 286-298.
- SUN, C. & QING, Z. (2018), Design and Construction of a Virtual Bicycle Simulator for Evaluating Bicycle Traffic Control and Facilities Design. In: *Advances in Civil Engineering*, vol. 2018, Article ID 5735820.
- TWADDLE, H. & BUSCH, F. (2018), Binomial and multinomial regression models for predicting the tactical choices of bicyclists at signalised intersections. *Transportation Research Part F: Traffic Psychology and Behaviour*, 60, 47-57.
- TWADDLE, H. & GRIGOROPOULOS, G. (2016), Modeling the Speed, Acceleration, and Deceleration of Bicyclists for Microscopic Traffic Simulation. In: *Transportation Research Record: Journal of the Transportation Research Board*, (2587), 8-16.
- TWADDLE, H., GRIGOROPOULOS, G. & BUSCH, F. (2016), Integration of an external bicycle model in SUMO. In: *SUMO 2016 - Traffic, Mobility, and Logistics - Proceedings*, 93-102.
- TWADDLE, H., SCHENDZIELORZ, T., FAKLER, O. & AMINI, S. (2014), Use of automated video analysis for the evaluation of bicycle movement and interaction. In: *Video Surveillance and Transportation Imaging Applications 2014*, 9026, 1-13.
- TWADDLE, H., SCHENDZIELORZ T. & FAKLER, O. (2014), Bicycles in urban areas: Review of existing methods for modeling behavior. *Transportation Research Record: Journal of the Transportation Research Board*, 2434.1, 140-146.
- YIN S. & YIN Y. (2007), Implementation of the interactive bicycle simulator with its functional subsystems. In: *Journal of Computing and Information Science in Engineering*, 7(2), 160-166.

# Verwundbarkeitsanalyse urbaner Gebiete, 3D Bewertung von disruptiven Ereignissen zur Steigerung der Resilienz

Kai FISCHER<sup>1</sup>, Stefan TROMETER<sup>2</sup>

<sup>1</sup>Fraunhofer-Institut für Kurzzeitdynamik, Ernst-Mach-Institut · Am Klingelberg 1 · 79588 Efringen-Kirchen  
E-Mail: kai.fischer@emi.fraunhofer.de

<sup>2</sup>VirtualCitySYSTEMS GmbH · Marktplatz 2 · 85567 Grafing  
E-Mail: strometer@virtualcitysystems.de

## 1 Einführung ins Thema

Wachsende Urbanisierung, steigende Komplexität kritischer Infrastrukturen sowie das Auftreten von disruptiven Ereignissen stellen Städteplaner und Entscheidungsträger vor neue Herausforderungen. Ebenfalls können Naturkatastrophen oder von Menschen verursachte Ereignisse zu chaotischen und schwierigen Bedingungen für Rettungskräfte führen. Es müssen schnelle Entscheidungen getroffen werden um mögliche Konsequenzen zu mindern.

Im Rahmen dieses Beitrags wird eine Methodik vorgestellt, die es ermöglicht städtische Gebiete ganzheitlich zu bewerten und Schwachstellen zu identifizieren. Empirische Daten sowie physikalische und Ingenieurmodelle werden herangezogen um zu erwartende Schäden an Gebäuden und Infrastrukturelementen zu charakterisieren.

Das Konzept der Resilienz wird über die fünf Phasen Vorbereitung, Prävention, Schutz, Reaktionsschnelligkeit und Wiederherstellung definiert. Die vorgestellte Methodik ermöglicht eine Bewertung und einen Vergleich einzelner Resilienzphasen um somit die Resilienz eines urbanen Gebietes zu steigern. Mithilfe eines Webservices kann die Methodik an gängige 3D Stadtmodelle angebunden werden und die Resultate werden benutzerfreundlich visualisiert.

## 2 Zusammenfassung des Vortragsinhalts

Der Ansatz, das Risiko urbaner Räume aus Anfälligkeit und Verwundbarkeit zu quantifizieren, ermöglicht eine leichte Identifizierung potentieller Schwachstellen im urbanen Gebiet (FISCHER, HÄRING, RIEDEL, VOGELBACHER & HIERMAIER 2016). Somit kann mit quantitativen Ansätzen die klassischen Fragen der Risikoanalyse (KAPLAN & GARRICK 1981) beantwortet werden:

- Was kann passieren?
- Wie wahrscheinlich ist es, dass etwas passiert?
- Wenn etwas passiert, was sind die zu erwartenden Schäden?

Basierend auf wachsender Urbanisierung, steigender Komplexität kritischer Infrastruktur und neuen Bedrohungen, reichen Risikoanalysen allein jedoch nicht aus, um nachhaltigere und robustere Städte zu generieren (THOMA, SCHARTE, HILLER & LEISMANN 2016). Orientiert an den fünf Handlungsbereichen definiert sich Resilienz als die Fähigkeit:

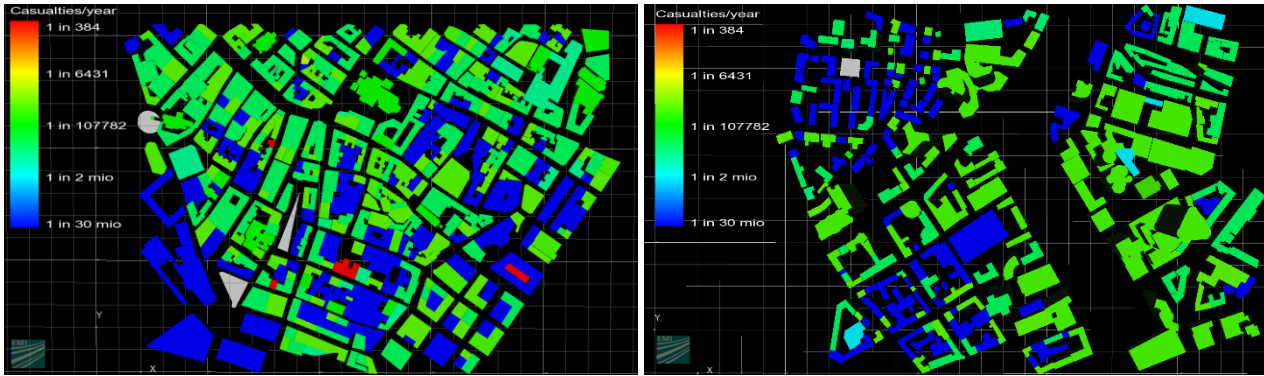
- sich auf widrige Verhältnissen vorzubereiten und aus vergangenen Ereignissen zu lernen,
- sie nach Möglichkeit zu verhindern,
- sich davor zu schützen,
- schnellstmöglich zu reagieren und
- sich zügig davon zu erholen.



**Abb. 1:**  
Phasen des Resilienz-Zyklus.

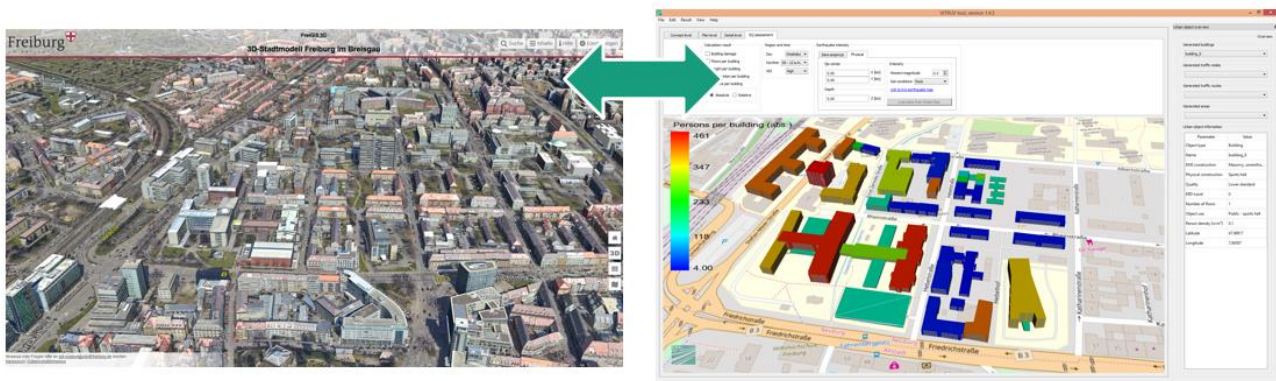
Im Rahmen einer Forschungsarbeit (FISCHER 2018) wurde eine Methodik entwickelt, die es ermöglicht einzelne Resilienzphasen für urbane Gebiete zu quantifizieren. Resultate sind in einem Software Demonstrator implementiert, welcher an 3D Stadtmodelle gekoppelt ist. Somit können benutzerfreundlich zu erwartende Schäden oder die Effektivität Resilienz steigernder Maßnahmen visualisiert werden.

Urbane Gebiete werden über die komplexe Interaktion verschiedener Zonierungen und Funktionen bestimmt. Gebäude, Verkehrswege, freie Fläche und Landschaft bestimmen die Morphologie. **Abb. 2** zeigt den Vergleich einer kompakten Stadt und eines Stadtmodells mit eindeutiger Nutzungszuordnung (FISCHER, HIERMAIER, RIEDEL & HÄRING 2018). Somit kann die ganzheitliche Bewertung und deren Visualisierung im 3D Stadtmodell Beiträge liefern, um das Wachstum urbaner Räume nachhaltig zu gestalten.



**Abb. 2:** Bewertung zu erwartender Schäden im urbanen Raum und deren Visualisierung in Abhängigkeit der Morphologie (FISCHER, HIERMAIER, RIEDEL & HÄRING 2018).

Die Methodik zur Bewertung von Resilienz erfordert die Charakterisierung einzelner urbaner Objekte. Informationen zur Gebäudenutzung liefern beispielsweise Angaben zur Personendichte oder zum Gebäudewert. Angaben zur Konstruktionsart, die Abmessungen und die Anzahl an Etagen ermöglichen die Analyse der Robustheit oder der Gebrauchstauglichkeit. Die Entwicklung einer technischen Spezifikation definiert die geforderten Attribute und zeigt auf, wie diese Informationen aus 3D Stadtmodellen extrahiert werden können.



**Abb. 3:** 3D Stadtmodell der Stadt Freiburg (links) und die Bewertung potentieller Risiken (rechts) basierend auf einem Austausch mittels Webservice.

**Abb. 3** zeigt im linken Bild das 3D Modell der Stadt Freiburg. Basierend auf diesem Modell wurde ein kleinerer Bereich extrahiert und mit der vorgestellten Methodik bewertet. Das rechte Bild zeigt beispielhaft die Abschätzung der Personen pro Gebäude in Abhängigkeit der Nutzung, der Tageszeit und des Wochentags. Dieses Resultat liefert eine wichtige Information für die Koordination eines Rettungseinsatzes, beispielsweise zur Evakuierung eines Bereichs.

## Literatur

- FISCHER, K., HÄRING, I., RIEDEL, W., VOGELBACHER, G. & HIERMAIER, S. (2016), "Susceptibility, vulnerability and averaged risk for resilience enhancement of urban areas," *International Journal of Protective Structures* 7(1), pp. 45-76.
- KAPLAN, S. & GARRICK, J. (1981), "On the Quantitative Definition of Risk," *Risk Analysis* 1(1), pp. 11-27.
- THOMA, K., SCHARTE, B., HILLER, D. & LEISMANN, T. (2016), "Resilience Engineering as Part of Security Research: Definitions, Concepts and Science Approaches," *European Journal for Security Research* 1(1), pp. 3-19.

- FISCHER, K. (2018), Resilience quantification of urban areas - An integrated statistical-empirical-physical approach for man-made and natural disruptive events (PhD thesis), Freiburg: Fraunhofer-Verlag.
- FISCHER, K., HIERMAIER, S., RIEDEL, W. & HÄRING, I. (2018), „Morphology Dependent Assessment of Resilience for Urban Areas,“ *Sustainability 10*.
- STADT FREIBURG, „3D-Stadtmodell Freiburg im Breisgau,“ [Online]. Available: <https://3d.freiburg.de/>.

# Visualisierung von Geobasisdaten mit georeferenzierender AR-Technologie (Praxisbeispiele)

Ulf KREUZIGER

Verband für Landentwicklung und Flurneuordnung Brandenburg · Friedrich-Engels-Straße 23 · 14473 Potsdam

## 1 Einführung ins Thema

Smartphones und Tablets gehören zum Alltag und somit ebenfalls zum Arbeitsalltag. Diese Multifunktionswerkzeuge sind zum ständigen Begleiter der Menschen im Privat- sowie Arbeitsleben geworden, somit allzeit massenhaft verfügbar und können in wenigen einfachen Arbeitsschritten auch für die Visualisierung von Geobasisdaten eingesetzt werden. Insbesondere kann mit derartigen Geräten Augmented Reality (AR) realisiert und dabei Nichtvorhandenes sichtbar gemacht werden. Seit vielen Jahren wird an derartigen Applikationen, die die Fähigkeit besitzen, die Realität und die Virtualität miteinander zu vermischen, gearbeitet und praktische Problemstellungen mit wissenschaftlichen Methoden gelöst.

Im Beitrag wird anhand drei realisierter Funktionalitäten im Kontext der Visualisierung von Geobasisdaten exemplarisch der praktische Entwicklungsstand aufgezeigt. Hierbei handelt es sich um die unterschiedlichen Formen perspektivischer AR-Darstellungen von Geobasisdaten unter Verwendung von Bewegungssensorik, die terrestrische AR-Aufnahme von Bodenpunkten nebst georeferenziertem Foto und um die AR-Sachdatenabfrage beliebiger Geobasisobjekte, z. B. Flurstücksgeometrien.

## 2 Zusammenfassung des Vortragsinhalts

### 2.2 AR-Visualisierung von Geobasisdaten

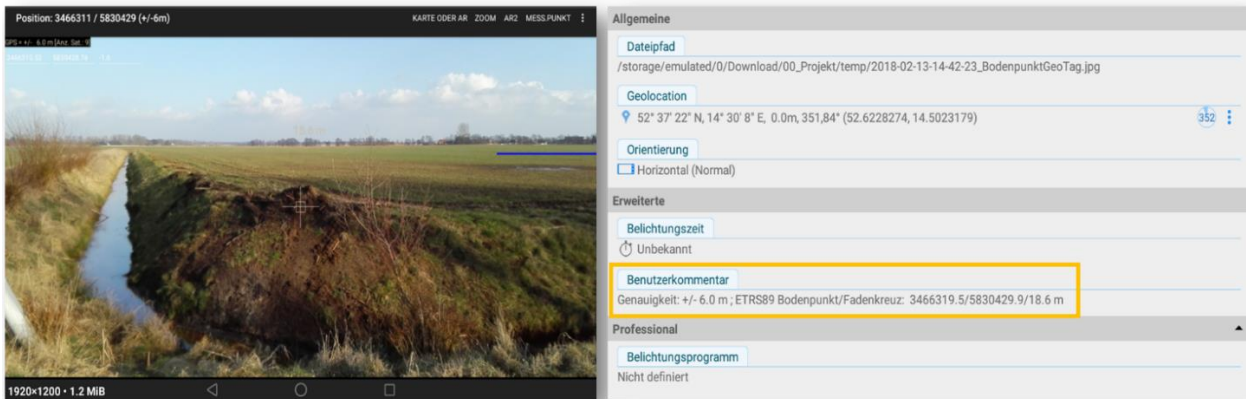
Geobasisdaten können auf einfache Weise per Web Map Service (WMS) oder Web Feature Service (WFS) in ein georeferenzierendes Augmented Reality System für Geoobjekte (GeoARS) geladen und verwendet werden. Herkömmliche Mobile Kartenanwendungen verwenden als Kartenansicht bzw. Grafikdatenansicht regelmäßig ausschließlich eine Draufsicht (vgl. Abb. 1 links). In Abbildung 1 werden Bildschirmfotos einer GeoARS-Applikation dargestellt, die Geobasisdaten und ihre Geoobjekte in unterschiedlicher Form präsentieren: Kartenansicht, Exozentrische Ansicht, Egozentrische Ansicht. Die Anwendung der AR-Technologie ermöglicht folglich, zusätzlich zur Draufsicht weitere Datenansichten. Die visualisierten Geoobjekte eines WFS haben hierbei eine besondere Bedeutung, da sie objektorientierte Sachdatenattribute tragen, die semantisch beauskunftet werden können und sie sich aufgrund Ihrer Vektorgrafikeigenschaften für die Darstellung in der AR-Egoperspektive besonders gut eignen.



**Abb. 1:** Formen der AR-Darstellung; hier: Bildschirmfotos einer GeoARS-Applikation mit Topographischer Karte WMS (links), Digitalem Orthophoto WMS (Mitte) und Live-Kamerabild (rechts), jeweils überlagert mit WFS-Flurstücksgeometrien.

## 2.2 Terrestrische AR-Aufnahme von Bodenpunkten mittels GeoARS

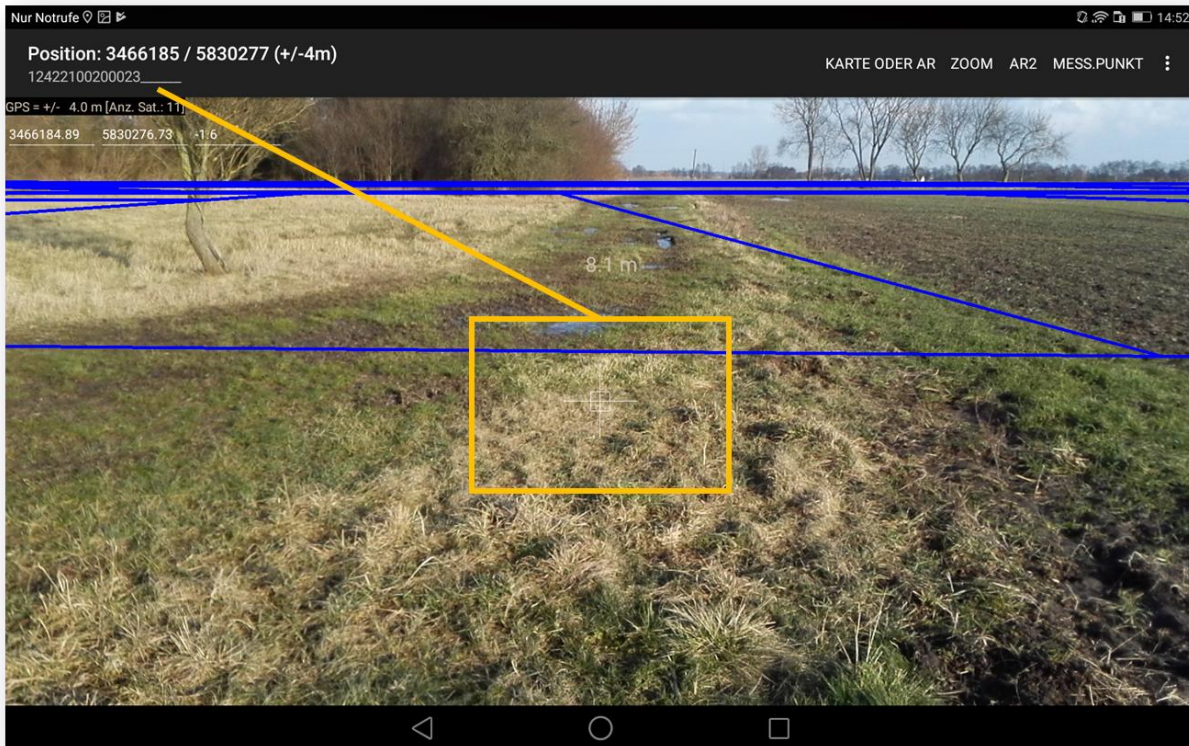
Durch konsequente Anwendung der AR-Technologie, insbesondere mit Hilfe der Bewegungssensorik, bieten Smartphone-GeoARS die Möglichkeit, zusätzlich zum Standpunkt (Geolocation), weitere Zielpunkte der näheren Umgebung auf der Erdoberfläche durch eine terrestrische Aufnahme zu erfassen. Bei diesem Messvorgang entstehen Paare geographischer Koordinaten aus Stand- und Zielpunkt. Der gemessene Zielpunkt (Bodenpunkt/Fadenkreuz) kann mit zusätzlichen Informationen versehen und persistent abgespeichert werden. Hierbei kann gleichzeitig zu jedem Zielpunkt ein georeferenziertes Foto (vgl. Abb. 2), inklusive abgebildetem Fadenkreuz zur eindeutigen Identifizierung des Zielpunktes und zusätzliche Bildinformationen (Metadaten), z. B.: Benutzerkommentare/ digitale Notizen, Genauigkeit der Koordinaten, gespeichert werden. Das georeferenzierte Foto nebst Metadaten und die Koordinatendatei können dann, z. B. in beliebigen GIS, weiter verwendet werden.



**Abb. 2:** Foto (links) mit Geo-Tag Informationen und weiteren Metadaten (rechts) zum aufgemessenen Stand- und Zielpunkt.

## 2.2 AR-Sachdatenabfrage von Geobasisobjekten

Werden die Geobasisdaten als Vektorgeometrien, z. B. per WFS in das GeoARS geladen, so sind regelmäßig weitere Sachdatenattribute und deren Werte in den Daten enthalten. Diese können auch weiterführend mit eigenen Sachdaten verknüpft (z. B. per Relationen zu eigenen Datentabellen) und im GeoARS verwendet werden. Hierbei bieten GeoARS die Möglichkeit die Umgebung auf besondere Weise thematisch abzufragen, indem in der Egozentrischen Ansicht (vgl. Abb. 1, rechts) unter Verwendung des Kamerabildes mit dem Fadenkreuz auf einen Bodenpunkt gezeigt wird. Zu diesem Punkt werden dann kontinuierlich und in Echtzeit die verfügbaren thematischen Sachdaten im Display des GeoARS angezeigt, z. B. das Flurstückskennzeichen (vgl. Abb. 3, oben links im Bild) oder beliebige weitere thematische Attribute, wie z. B. Grundstückseigentümer, Lagebezeichnung, Nutzungsart.



**Abb. 3:** AR-Sachdatenabfrage von Geobasisobjekten; hier: Flurstücksgeometrie mit Flurstückskennzeichen.

## Literatur

- ADV (2018), Dokumentation zur Modellierung der Geoinformationen des amtlichen Vermessungswesens (GeoInfoDok) Hauptdokument Version 7.0.3. Verfügbar unter: <http://www.adv-online.de/AAA-Modell/Dokumente-der-GeoInfoDok/GeoInfoDok-7.0/> . 2018, letzter Zugriff am: 2018-01-19. – Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV).
- AZUMA, R. (1997), A Survey of Augmented Reality. Presence, Ausgabe 04/1997.
- BILL, R. (2016), Grundlagen der Geo-Informationssysteme. 6. Auflage. Wichmann, Berlin.
- KREUZIGER, U. (2017), Entwicklung georeferenzierender Augmented Reality Systeme auf Mobilgeräten mit Mikrosensoren. Verlag der Bayerischen Akademie der Wissenschaften, München.

# Visualisierung von Geodaten – von der letzten Meile zur besten Meile

Liqui MENG

Lehrstuhl für Kartographie · Technische Universität München · Arcisstr. 21 · 80333 München

E-Mail: liqui.meng@tum

## 1 Geovisualisierung am Anfang und Ende der Wertschöpfungskette

Geovisualisierung in Form von Karten und kartenverwandten Darstellungen befindet sich oft am Ende einer Wertschöpfungskette aus Geodaten. Logischerweise wird sie in der Geodäsie und Geoinformatik als die letzte Meile wahrgenommen. Man hält diese letzte Meile für wichtig, denn ein Kartenbild sagt mehr als tausend Worte, indem es raumzeitlich komplexe Relationen veranschaulicht. Eine in archäologischen Fundstellen neu ausgegrabene Kartenskizze könnte den Anfang der Weltgeschichte umschreiben. Zugleich setzt sich die Weltgeschichte mit täglichen Nachrichtenkarten als das neue Kapitel fort. Geovisualisierung wird aber auch oft zur Verdeutlichung diverser raumbezogener Datenunterlagen von unterschiedlichen Qualitäten für eine bevorstehende Wertschöpfungskette eingesetzt und damit als die erste Meile wahrgenommen. Man hält diese erste Meile ebenfalls für wichtig, weil sie den Ist-Stand eines Entwicklungsprozesses vertritt, der mit dem Sollstand visuell am besten kontrastiert.

Die Auffassung der Geovisualisierung als die letzte oder die erste Meile bereitet allerdings zwei zusammenhängende Probleme: (1) die Arbeitsabfolge zwischen den beiden Enden wird in der Regel als ein sequentieller Prozess zur schrittweisen und rechenintensiven Informations- und Wissensgewinnung aus Geodaten betrachtet. Die methodische Vorgehensweise im Prozess einschließlich Fehlerfortpflanzung und Unsicherheit versteckt sich in einer Blackbox; (2) die Geovisualisierung am Rand der Arbeitsabfolge dient lediglich der effizienten Kommunikation der Eingangsdaten und der Endergebnisse. Eine derartige peripherisierte Rolle kann den Wertschöpfungsprozess kaum beeinflussen und daher keinen Mehrwert generieren.

## 2 Geovisualisierung in der offenen Wertschöpfungsplattform

Mit der Einführung der visuellen Analytik in die Geoinformatik vor einem Jahrzehnt im Zusammenhang mit der steigenden Verfügbarkeit der Volunteered Geographic Information (VGI) wächst der Bedarf an Entwicklung offener Datenaustausch- und Wertschöpfungsplattformen (Abb.1), um die Bürgerbeteiligung an den raumbezogenen Entscheidungsprozessen zu fördern bzw. mehr Bürger aus der Bedeutungslosigkeit in der digitalen Gesellschaft herauszuführen (HARARI 2018, S.42). Eine Vielzahl von neuen Geovisualisierungsverfahren ist entstanden und schnell verbreitet. Sie demonstriert das Potential der Visualisierung als die beste Meile zur Begleitung, der Transparenzerhöhung und der Qualitätssicherung der einzelnen Datenveredlungsprozeduren von der Integration bis hin zur Wissensentdeckung. Um diese beste Meile zu bewerkstelligen ist es jedoch erforderlich, eine wissenschaftliche Einsicht ins neuronale System des Menschen bei der visuellen Wahrnehmung und Kognition raumzeitlicher Relationen und Prozesse zu gewinnen.

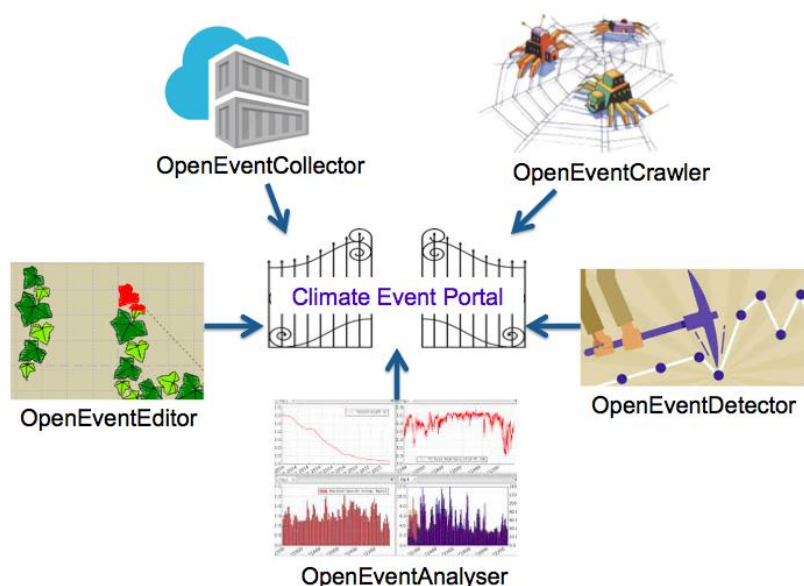


Abb. 1: Komponente einer offenen Plattform in Bayern zum Thema „Klima-Event“.



Im vorliegenden Beitrag werden die aktuellen Forschungsarbeiten am Lehrstuhl für Kartographie der TU München vorgestellt. Die Autorin gibt einen Überblick über die vielfältigen Geovisualisierungsmethoden und deren sinnvolle Zusammensetzung zur Unterstützung der Entdeckung und Interpretation natürlicher Vorgänge und menschlicher Aktivitäten im städtischen und ländlichen Raum. Die jüngste Entwicklung wird anhand von zwei Fallbeispielen demonstriert. Beim ersten Beispiel handelt es sich um eine auf Bayes-Netzen basierende visuell analytische Plattform, die eine Verschränkung der menschlichen Intuition in Form von Fachwissen und Expertenerfahrungen mit der rechnerischen Rationalität erlaubt, um den interaktiven und iterativen Reasoning-Prozess bei Unsicherheit durchzuführen und Zwischenergebnisse mit Unsicherheitsmaß kontinuierlich zu visualisieren (Abb. 2). Das zweite Beispiel widmet sich dem Deep-Learning des Begriffes „Standort“ als ein räumlicher Behälter der Mensch-Umwelt-Interaktionen, der zugleich eine semantisch bedeutsame Einheit ohne eine scharfe und statische Grenze bildet. Mit einer großen Zahl von Bildern aus unterschiedlichen Kamerapositionen auf der Straßen-Ebene werden die generativen Modelle „Variational Autoencoders“ von „Deep Convolutional Neural Networks“ trainiert und in Kombination mit Vergleichsmethoden getestet, um zu simulieren, wie sich die einzelnen voneinander differenzierbaren Standorte durch menschliche visuelle Kognition kollektiv generieren lassen (Abb. 3).

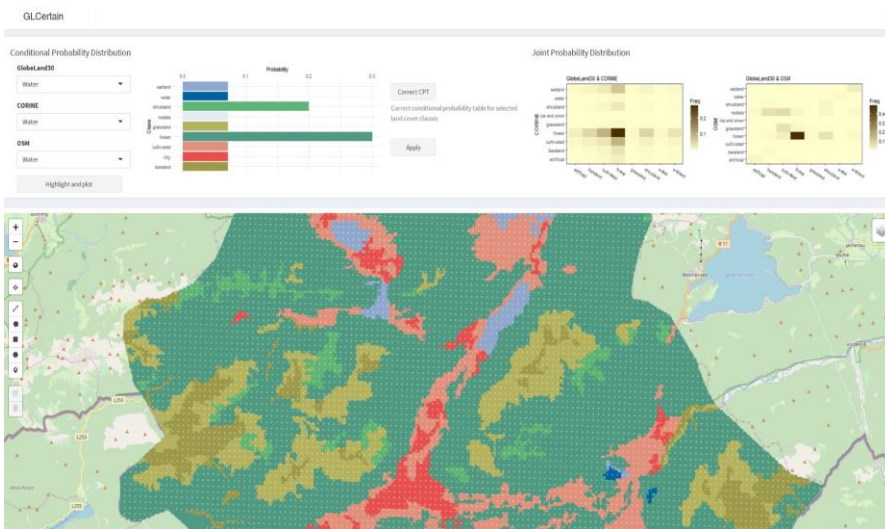


Abb. 2: Die Bedienungsfläche der auf Bayes-Netzen basierenden visuell analytischen Plattform für Reasoning bei Unsicherheit (CHUPRIKOVA 2019).

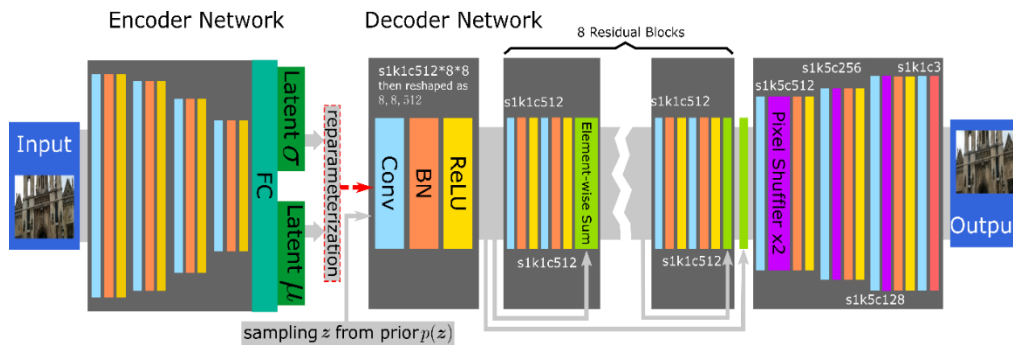


Abb. 3: Die Architektur der Deep Convolutional Neural Networks zum Daten-betriebenen Lernen des Standorts (LYU 2019).

## Literatur

CHUPRIKOVA, E. (2019), Visualizing Uncertainty in Reasoning – A Bayesian Network-enabled Visual Analytics Approach for Geospatial Data, Doktorarbeit, Lehrstuhl für Kartographie, TUM.  
 HARARI, Y.-N. (2018), 21 Lektionen für das 21. Jahrhundert, C.-H. Beck.  
 LYU, H. (2019), Approaching a collective place definition from street-level images using deep learning methods”, Doktorarbeit, Lehrstuhl für Kartographie, TUM.

# Visuelle Analyse von Daten mit 2- oder 3-dimensionalem Bezug

Heidrun SCHUMANN und Steve DÜBEL

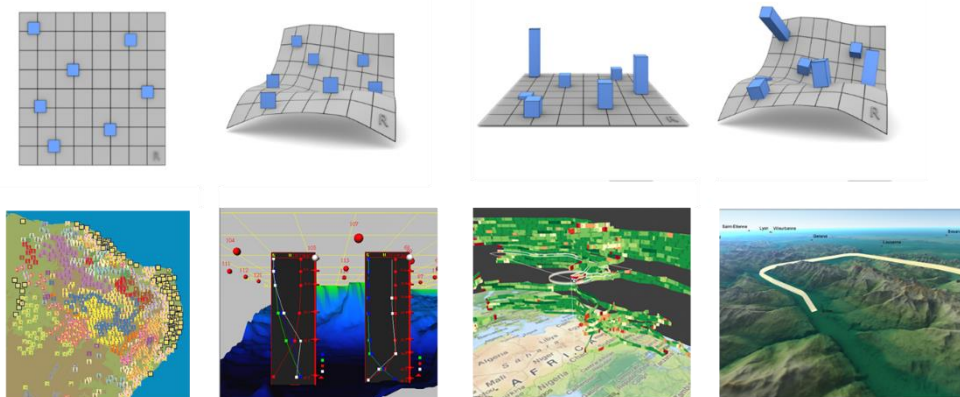
Institut für Informatik · Universität Rostock · Albert-Einstein Straße 22 · 18051 Rostock

E-Mail: heidrun.schumann@uni-rostock.de, steve.duebel@uni-rostock.de

## 1 Einführung ins Thema

Ziel der Visualisierung ist es, Daten graphisch so zu repräsentieren, dass strukturelle Zusammenhänge und relevante Eigenschaften analysiert werden können. Komplexität und Umfang heutiger Datenmengen erfordern hierfür jedoch spezielle Lösungsansätze. Dies gilt insbesondere, wenn die Daten in ihrem räumlichen Kontext untersucht werden sollen.

Dabei ist zu entscheiden, ob Daten und Bezugssystem jeweils in 2D oder 3D zu repräsentieren sind. Hieraus ergeben sich 4 Optionen, die in Abbildung 1 zusammengefasst sind.



**Abb. 1:** Kategorisierung der Darstellungsoptionen von Daten im räumlichen Kontext.

Obere Zeile: Schematische Illustration der Optionen, untere Zeile: Beispielbilder.

Von links nach rechts: 2D-Daten- und 2D-Raumdarstellung (NOCKE et al. 2005), 2D-Daten- und 3D-Raumdarstellung (KREUSELER 2000), 3D-Daten- und 2D-Raumdarstellung (TOMINSKI et al. 2014) sowie 3D-Daten- und 3D-Raumdarstellung (DÜBEL et al. 2017)

Im vorliegenden Beitrag werden für jede Kategorie Visualisierungsansätze vorgestellt. Dabei wird der Schwerpunkt auf die Visualisierung von Daten im 3D-Gelände gelegt. Es werden Methoden diskutiert, die hierbei die Bildkomplexität reduzieren und Verdeckungen auflösen.

## 2 Zusammenfassung des Vortragsinhalts

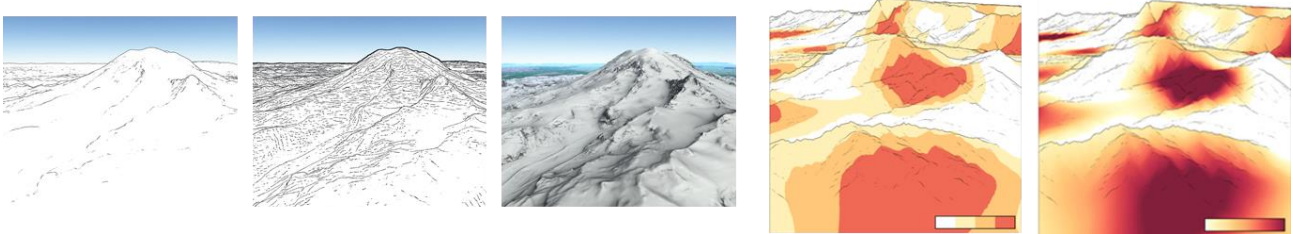
### 2.1 2- oder 3-dimensionale Repräsentation

Bei 2D Repräsentation werden 2-dimensionale graphische Primitive eingesetzt, um die Daten oder den Bezugsraum zu repräsentieren, bei 3D Repräsentationen sind es 3-dimensionale Primitive. Die Entscheidung, ob Daten oder Bezugsraum in 2- oder 3D zu repräsentieren sind, hängt von vielen Faktoren ab. Wenn z.B. der Zusammenhang von Daten und zugehöriger Geländestruktur wichtig ist, sollte das Gelände auch in 3D repräsentiert werden.

### 2.1 Reduzierung der Bildkomplexität

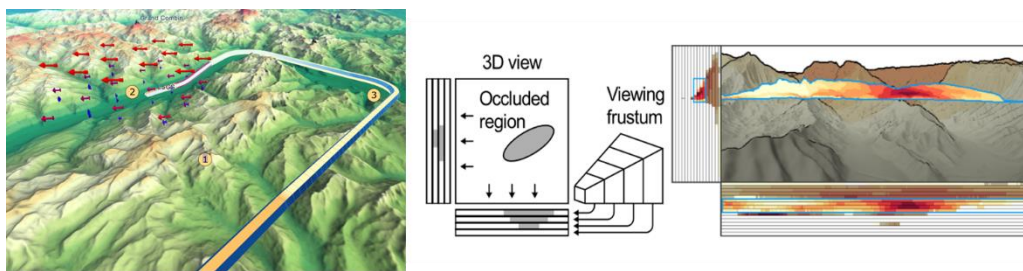
Komplexe 3-dimensionale Gelände und große Datenmengen lassen sich in der Regel nicht gleichzeitig mit einem Bild veranschaulichen ohne dass Visual Clutter entsteht. Deshalb ist es wichtig, Prioritäten zu setzen und die Darstellung entsprechend anzupassen. Die Anpassung kann sich sowohl auf den globalen als auch auf den lokalen Detaillierungsgrad von Gelände und Daten beziehen. Bei einer globalen Anpassung wird entweder die Gelände- oder die Datendarstellung priorisiert. Bei einer lokalen Anpassung werden Fokus & Kontext-Techniken eingesetzt oder Feature von besonderem Interesse hervorgehoben.

Die Priorisierung hängt vom gegebenen Kontext ab. Wenn die Geländestruktur sehr wichtig ist, muss der Geländedarstellung der Vorrang gegeben werden. Wenn es aber primär um die Analyse der zugehörigen Daten geht, z.B. um die Analyse von Wetterdaten, müssen vor allem die Daten mit allen Details dargestellt werden. Die Priorisierung kann sich ändern, wenn sich der Fokus der visuellen Analyse ändert. Das macht eine erneute globale Anpassung der Darstellung erforderlich. Abbildung 2 zeigt ausgewählte Repräsentationen von Gelände und Daten, die bei der globalen Anpassung miteinander je nach Bedarf kombiniert werden.



**Abb. 2:** Unterschiedliche Darstellungsarten von Gelände und Daten für die globale Anpassung. Beim Gelände variiert die Darstellung von der Repräsentation der Silhouetten bis zur Visualisierung schattierter Oberflächen. Bei den Daten werden entweder individuelle oder aggregierter Werte visualisiert.

Neben der globalen Anpassung kann auch eine lokale Anpassung erforderlich sein, um bestimmte Bereiche des Geländes oder der Daten in der Darstellung hervorzuheben; zum Beispiel das Gelände in der unmittelbaren Umgebung des Ziel-flughafens oder Daten, die einen zulässigen Schwellwert überschreiten. Abbildung 3 (links) illustriert so eine lokale Anpassung.



**Abb. 3:** Illustration der lokalen Anpassung (links): höhenkodierte Geländedarstellung (1) mit eingebetteter Flugzeugtrajektorie (3) und Winddaten, die nur für die Umgebung des Zielflughafens Sion angezeigt werden (2) und Illustration von Visibility Widgets (rechts): durch das Gelände verdeckte Daten werden zusätzlich angezeigt.

## 2.2 Behandlung von Verdeckungen

Bei 3-dimensionalen Darstellungen treten Verdeckungen auf. Visibility Widgets sind zusätzliche Ansichten, die verdeckte Informationen anzeigen. Zunächst müssen dafür die nicht sichtbaren Informationen identifiziert werden. Es gibt verschiedene Gründe, warum Informationen nicht sichtbar sind. Zum Beispiel wird entsprechend der gewählten Kameraeinstellungen nur ein Ausschnitt des Geländes und der Daten dargestellt. Zudem können Informationen auch von Geländestrukturen oder Datendarstellungen verdeckt werden, die weiter im Vordergrund liegen.

Nicht sichtbare Informationen lassen sich durch verschiedene Visibility Widgets anzeigen. In Abbildung 3 (rechts) werden z.B. Visibility Bands genutzt, um verdeckte Datenwerte zu visualisieren. Mit Panorama Bands können dagegen Informationen außerhalb des Sichtbereichs angezeigt werden. Unabhängig von ihrer konkreten Ausprägung, ermöglichen Visibility Widgets die Analyse und das Interagieren mit verdeckten Informationen.

## Literatur

- DÜBEL, S. & SCHUMANN, H. (2017), Visualization of Features in 3D Terrain. In: ISPRS International Journal of Geo-Information, 6 (11), 357.
- DÜBEL, S., RÖHLIG, M., SCHUMANN, H. & TOMINSKI, T. (2017), Visualizing 3D Terrain, Geo-spatial Data, and Uncertainty. In: Informatics, 4 (1), 6.
- DÜBEL, S., RÖHLIG, M., SCHUMANN, H. & TRAPP, M. (2014), 2D and 3D presentations of spatial data: A systematic review. In: Proceedings 3DVis@IEEEVIS2014: Does 3D really make sense for Data Visualization? (Paris).

RÖHLIG, M., LUBOSCHIK, M. & SCHUMANN, H. (2017), Visibility Widgets for Unveiling Occluded Data in 3D Terrain Visualization. *Journal of Visual Languages & Computing*, 42 (Supplement C).

# Towards a relational database Space Filling Curve (SFC) interface specification for managing nD-PointClouds

Peter VAN OOSTEROM, Martijn MEIJERS, Edward VERBREE, Haicheng LIU and Theo TIJSSEN

Section GIS technology, Faculty of Architecture and the Built Environment, TU Delft

E-Mails: P.J.M.vanOosterom@tudelft.nl, B.M.Meijers@tudelft.nl, E.Verbree@tudelft.nl, and H.Liu-6@tudelft.nl

## Abstract

In this paper we propose to treat point clouds as a first-class representation (similar to vector or raster representations), with the nD-PointCloud as the solution for this, offering deep integration of space, time and scale. For efficiency reasons spatial indexing and clustering of these large point clouds is extremely important and this is obtained based on a Space Filling Curve (SFC). In order to get beyond the current state of the art of storing/ managing point clouds in files, a DBMS solution is presented (with all benefits: integration with other data types, scalability, multi-user, transaction support, etc.). Finally, a DBMS SFC interface specification for point clouds is proposed.

## 1 Introduction

Gridded (raster) or object (vector) models are state of the art to represent spatio-temporal features as for instance in urban planning, geo-information or human and social geography. Point clouds are emerging as a data acquisition technique, but not used throughout the whole processing chain (model storage, spatial analysis and visualization). Typically, current models are organized in a fixed number of levels of importance (detail/scale), introducing limitations as fixed level choices and data density jumps between levels. We propose an added dimension to space and time that will facilitate continuous levels of importance (scale). The goal is to construct a novel nD-PointCloud model for handling massive multi-dimensional (nD) point cloud data sets, representing space, time, scale and added information (colour, material properties, velocity etc.). The nD-PointCloud as serious alternative to raster of vector models offers unique new possibilities: more natural representations (e.g. vegetation), more detail and no conversion loss enabling better spatial analysis (e.g. flow computation or routing). Based on a novel use of high-resolution nD space filling curves deep integration of space, time and scale as basis for data organization is realized and applied for big data (towards  $10^{15}$  points). By enabling operations directly on the raw point cloud data, nD-PointCloud largely avoids the extract-transform-load hurdle, which is an increasing problem in using big data. This enables major advances in domains requiring lossless spatio-temporal data of extremely high accuracy, such as applications of scanned surface models and moving object (trajectory) models, which are used as Proof-of-Principle. If successful, then nD-PointCloud will become the most used model (more than vector or raster), enabling progress in research fields like water management, land administration, vegetation monitoring, building modelling, transportation and mobility.

Deep integration allows analysis and visualization which would otherwise not be possible. Furthermore, this deep integration is key to efficient solutions; e.g. nD points represented via space filling curves, such as Morton or Hilbert curves (GUAN et al. 2018). Conceptually this deep integration was applied for the spatial and temporal dimensions in Hägerstrand's space-time cube (1970) to represent and visualize trajectories of moving objects (in space over time) and being able to decide if multiple trajectories meet in space and time (e.g. when tracking the attendance of people in a music concert when the individuals have positioning devices) or if trajectories represent people visiting the same place but at different moments in time (e.g. people shopping in the same market at various hours of the day). In a similar manner Meijers and van Oosterom (2012) propose the deep integration of spatial and scale (continuous Level of Detail) dimensions for vector data to provide solutions for efficient overviews, smooth zooms, and deriving uniform-scale views and mixed-scale (perspective or fish-eye) views: the space-scale cube. nD-PointCloud aims at being a generic approach, applicable to many domains.

In the majority of today's applications, point clouds are managed using file solutions with specific formats (e.g. ASPRS LAS). Typical file-based solutions include desktop applications (usually vendor-specific) and command-line executables, like Rapidlasso's LasTools (mixed-source) or the Point Cloud Abstraction Library (PDAL) (an open source project). The work-flow includes reading one or more files, processing the data and writing files back to the user. These traditional file-based solutions have major drawbacks, such as data isolation, data redundancy, and application dependency. DataBase Management Systems (DBMSs) do not have these drawbacks. The database community, commercial and open source, provides point cloud specific data structures; e.g. Oracle (Spatial and Graph) and PostgreSQL (PostGIS). The DBMS-based solutions can be categorized into two types (VAN OOSTEROM et al. 2015): the block and flat-table models. In the flat-table model, points are directly stored in a database table, one row per point, resulting in tables with many rows (WANG & SHAN 2005). This flat-table model is easy to implement and flexible for query and manipulation, but multi-dimensional clustering and indexing is a challenge (MARTINEZ-RUBI et al. 2014). For this purpose, the Space Filling

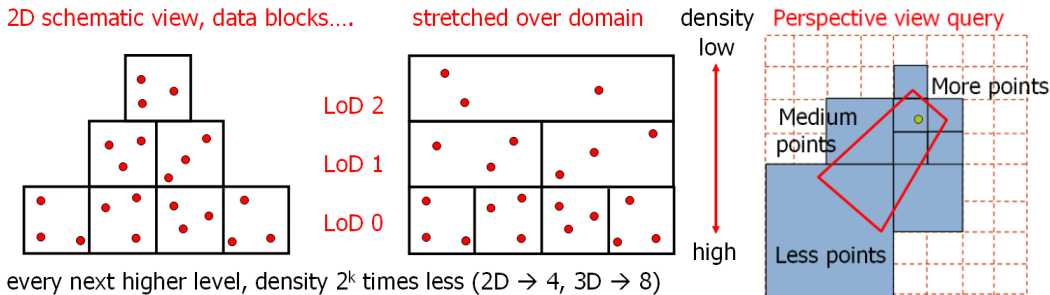
Curve (SFC) is introduced. This results in the nD-PointCloud as a new datatype, which is described in more detail in Section 2. A first proposal for the DBMS interface specification for nD-PointCloud is given in Section 3. Finally, conclusions and future work can be found in Section 4.

## 2 The nD-PointCloud as a new datatype

The nD-PointCloud approach (model with accompanying theory and technology) will be a radically new line, compared to today's gridded (raster, voxel) or object (vector, polyline/ polygon/ polyhedron) representations (Molenaar 1998) that work with discrete number of levels of importance (detail/scale). The continuous Level of Importance (cLoI) support is provided via its deep integration with space and time, i.e. their combination into a single representation. The  $n$  organizing dimensions, usually a subset of space, time and cLoI, define the location of a point in the  $nD$  Organizing Space. In addition, and depending on the specific domain and data set, there are  $m$  additional property dimensions (e.g. colour, flow direction, velocity, surface normal, different spectral channels, temperature, pressure, classification, object identity, light intensity, uncertainty, content descriptors, etc.), which together conceptually define points in a  $k=n+m$  dimensional space. Deep integration offers a range of benefits: spatio-temporal consistent models, efficient selection (on organizing dimensions; e.g. simultaneous selection based on space and time search predicates), new types of spatio-temporal computations, use of continuous importance, and integration of data from multiple sources.

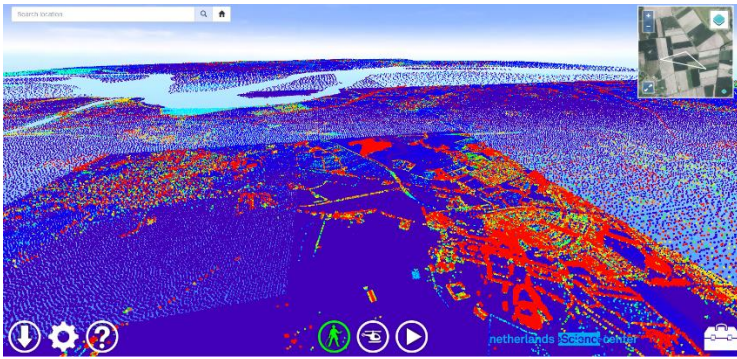
### 2.1 Discrete vs. continuous levels

There is a serious drawback in current models as their discrete number of levels will give density difference artefacts. The current state of the art is organizing point clouds in discrete LoD, or data pyramids (ARIKAN et al. 2014, VAN OOSTEROM et al. 2016) as illustrated in Figure 1 for the 2D case. This organization allows fast spatial searching including LoD selection: the further away from viewer the lesser points selected, i.e. the higher level blocks/points (Figure 3-right). These representations offer a discrete set of levels, i.e. multiple map-scales (JONES et al. 1996, KILPELAINEN 1997, FRIIS-CHRISTENSEN & JENSEN 2003). Discrete levels have well known restrictions; e.g. in perspective views (more detail nearby, less detail further away), the abrupt transition between various levels is often disturbing to the human eye, in spatio-temporal analysis/computation just the predefined discrete levels are available and 'optimally detailed' representation may not be available. There is a continuous point cloud approach proposed by Microsoft (2014), but this has severe drawbacks as it is not based on using cLoI as an organizing dimension and the solution is less general (focuses on visualization of 3D point clouds).

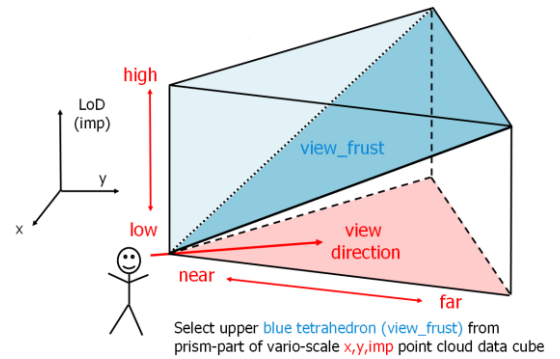


**Fig. 1:** Discrete LoD (scale/importance), left/middle: symbolic 'side view' of 2D point cloud data blocks with height now LoD; right: 'top view' for selecting point cloud data blocks for perspective query.

The human eye is very sensitive to this disturbing effect as can be noticed in Figure 2, but the same applies to 'suboptimal' analysis based on similar multi-level organization. More gain is possible by taking this strategy to the limit using an infinite number of levels, which is what the proposed cLoI approach addresses; see Figure 3. This is one of the key features of my nD-PointCloud model: cLoI becomes a true continuous dimension and it is used in the data organization: clustering, indexing (point cloud in space-time-importance data blocks) for fast retrieval and transmission. Regarding the time dimension, current solutions usually construct independent data structures for each time snapshot of the scanned surface. This disables efficient spatio-temporal analyses. The nD-PointCloud approach which deeply integrates space, time and importance aims at better enabling this kind of analyses. Changes between different moments in time can be more efficiently detected and analysed (up to its maximum available resolution), i.e. directly on point cloud data without down-sampling to grid or other representations.



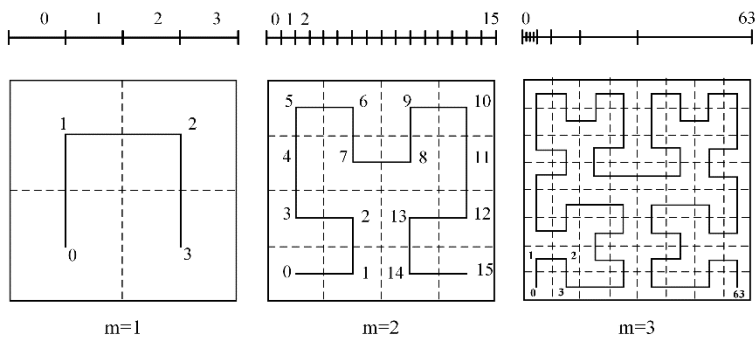
**Fig. 2:** Potree webviewer with AHN2 data (640 billion points) showing the discrete LoD (note the data density differences).



**Fig. 3:** Perspective view: further away from viewer, gradually only points with higher cLoI.

## 2.2 Space Filling Curves for organizing nD-PointClouds

Space filling curves (SFCs) such as Morton or Hilbert curves (BUTZ 1971, VAN OOSTEROM 1999, LAWDER & KING 2001, BADER 2012, GUAN et al. 2018) are proposed as a foundation for spatio-temporal representation. SFCs have the ability to cluster points close in reality, close on the curve. In mathematics, a space-filling curve is a continuous bijection from the hypercube in nD space to a 1D curve, i.e.  $C : \mathbb{R}^n \rightarrow \mathbb{R}$  (Jaffer 2014). The nD hypercube is of the order  $m$  if it has a uniform side length  $2^m$  (see Fig. 4). Analogously, the curve  $C$  also has an order  $m$  and its length equals to total number of  $2^{n \cdot m}$  cells.



**Fig. 4:** The illustration of 2D Hilbert curves with different orders.

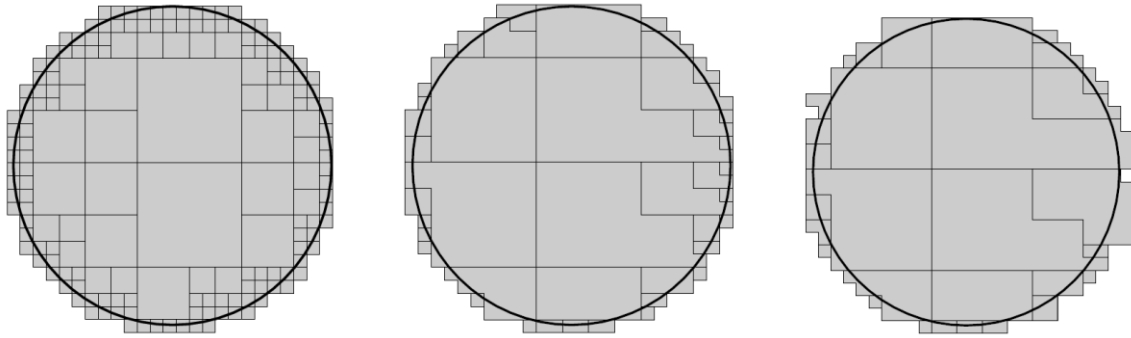
The method of SFCs will enable the integration of all the organizing dimensions into a single value. The design options that need to be researched and analysed for their impact, include: what are the organizing dimensions (used for SFC computation), what are the other attributes, what type of organization (relative scaling of organizing dimensions encoding, exact type of SFC, e.g. Morton or Hilbert curve, explicit data partitioning parameters such as clustering/ blocking size), what compression, etc.

The proposed nD-PointCloud approach supports efficient clustering, indexing, compression, data selection, transmission and use of huge point cloud data sets; towards  $10^{15}$  points. One might end-up in organizational spaces with five or more (generally  $n$ ) organizing dimensions with  $m$  attributes as property dimensions. The right balancing of organizing dimensions and other attributes when designing the data model is one of the research challenges. Given the huge volumes of data, parallel processing will be employed. The parallel approaches enable an nD-PointCloud analysis where serial approaches are not working (e.g. memory limits). Due to the 'quadrant' recursive partitioning of space, the SFCs may be used to subdivide a large problem into smaller parts suitable for parallel processing.

## 2.3 Using the SFC for multidimensional queries

Multidimensional selections using SFCs require a modified query algorithm that takes into account the SFC organization. This means that the query geometry needs to be translated into a number of continuous runs on the curve. Because all the above-mentioned queries correspond to a kind of multi-dimensional range query, we make use of the relationship between the SFC and the quadtree (VAN OOSTEROM & VIJLBRIEF 1996, GARGANTINI 1982) or  $2^n$  trees for higher dimensions. The maximum depth of the tree affects: (1) the number of SFC ranges that compose the query, and (2) the approximation of the query geometry. Only requesting

higher levels will give coarser  $2^n$  tree cells, resulting in additional points. The query procedure starts with filtering (and can be optionally followed by exact refinement computation). The filtering recursively checks the  $2^n$  tree cells that intersect with the query region, up to a specific depth, producing a mixture of big and small ranges returned, with the smaller ones located mostly near the boundary (Fig. 5 left). The cells are then translated into the equivalent SFC ranges and the neighbouring ranges on the curve are merged without adding any tree cell space (Fig. 5 middle). The direct neighbour merging can create non-rectangular ones. The ranges are further merged with nearby (but non-direct) neighbours in case they exceed a specified maximum amount (Fig. 5 right). Merging of non-direct neighbours will always result in additional tree cell space added to the original situation. The returned ranges are used for fetching the data. Despite merging of ranges, the result is an approximation which is very accurate, compared to the bounding box.



**Fig. 5:** The different steps in the preparation of the filter step: (left) Tree cell identification, 176 cells, (middle) direct neighbour merging, 42 cells and, (right) merging to maximum number of 30 cells.

### 3 DBMS interface specification for nD-PointClouds

This paper presents data structures and functions required to manage n-dimensional (nD) point cloud data in a DBMS. The system (interface) is designed to be flexible, easy to use and efficient at the same time. To some extent these are contradictory requirements, this makes it relatively difficult to implement the system in such a way that these requirements are well balanced (the optimal solution may be context dependent, but we propose a generic solution). The nD-PointClouds can be of quite different nature or background but are all within the scope of this specification. So, on one side we want to represent nD-PointClouds which are a result of scanning surfaces (LiDAR, MBES, dense matching images, etc.), but we also want to represent nD-PointClouds which are the result of tracing moving objects.

In this section the general ideas and interface will be described, many details can be found in earlier papers which describe the research and partial implementation related to management of nD-PointClouds in a DBMS. (VAN OOSTEROM et al. 2015) introduce the challenge to manage massive point cloud data in a database, present a benchmark, and propose to use a space filling curve (SFC) for efficient 3D point cloud data management with test case the AHN2 data. (PSOMADAKI et al. 2016) propose to use a SFC for the management of dynamic point cloud data and analyse organizations options: ranging from 2D (just xy in SFC key) to 4D (xyzt in key) with test case coastal monitoring data. The SFC computation was in both cases outside the Oracle database and implemented in Python (drawback a lot of data transfer between database and SFC programs). (GUAN et al. 2018) implemented the nD SFC Library in C++ (more efficient than Python) and tested up-to 5D by adding also the importance dimension to the SFC key (xyzt<sub>i</sub>), but expensive data transfer between database and external SFC software remained. In all of the above solutions the Oracle database with the Index Organized Table (IOT) was used, with several performance benefits: a. data is clustered (in order of key) -> faster, no memory jumping after range selection, b. single structure is more compact than 2 structures (table + index)-> faster, less reading, c. no combination of 2 structures ('join' of index and table needed) -> faster, less processing.

(DE VREEDE 2016) implemented using Python the SFC key technique in combination with the MongoDB database and tested this with moving ship trajectory data of the Automatic Identification System (system). Unfortunately, MongoDB does not support clustering of data (as in IOT). (MEIJERS et al. 2017) implemented using Python and Rust the SFC key technique inside the PostgreSQL database and also tested this with AIS data (2D space and 1D time). PostgreSQL does support clustering the table according to order on the index (SFC based), but index and table are different structures, so not completely having the benefits of an IOT. Finally (LIU et al. 2018) present the ambition towards  $10^{15}$ -level point clouds management based on a generic nD PointCloud structure with functionality inside the database (and benefiting from IOT). In addition to the cited publications most of the accompanying software is also available as open source; e.g. the Python code for the Dynamic Point Cloud is available at <https://github.com/stpsomad/DynamicPCDMS>, the C++ code for Morton/Hilbert encode/decode/range generation is available at <https://github.com/kwan2004/SFCLib> and the Python



and Rust code for running inside PostgreSQL is available at <https://bitbucket.org/bmmeijers/sfc-rs> and <https://bitbucket.org/bmmeijers/sfc-rs-ffi>.

In general, all the above approaches store one point in one row (record, tuple) in the database and organize the points well (according to the nD SFC). For several reasons the next step could be to group points in blocks. When starting from the sorted individual points (with their SFC key) this is now rather easy. Some benefits of blocks are: nice communication unit between server and web-client (allowing caching), compression well possible because nearby points will be in same block and their first part (most significant bits) are quite equal. However, in this document we limit ourselves to one point per row (and blocks are future work). Some of the lessons learnt from these experiences in various database environments are:

- Oracle IOT is elegant and efficient structure,
- full high-res key allows omitting storing the corresponding attributes (e.g. x or y as they can be decoded in full resolution from the SFC key),
- encoding/decoding, range generation outside database results in quite a bit of communication overhead,
- full high-res and higher dimensional keys results in large keys (for sure not fitting in 64 bits, so actual representation requires attention, e.g. use varchar or raw bytes),
- the PostgreSQL SFC key implementation within the database avoids massive data transfer in/out data when using the functions,
- the roles of organizing properties (e.g. x, y, time, importance called organizing dimensions) and other properties (non-organizing, call attributes) is rather arbitrary and context/use dependent,
- relative scaling of different organizing dimensions (space, time, importance) is needed to compute SFC (e.g. 1 meter = 1 second, but non-trivial as this influences the actual clustering). Finally,
- directly generating SFC ranges for different query geometry shapes has big benefits over simple bounding box approximation (more precise approximation will result in less false hits), but a high-resolution query geometry range generation in nD space may result in (too) many ranges especially when specifying just 2 dimensions of the key and the other open (implying they extend to infinite).

### 3.1 Handling nD points with a Space Filling Curve based organisation

Most databases cannot handle multi-dimensional (nD) point cloud data very well. Solutions are available for multi-dimensional spatial vector or raster data, but these are not very suitable for nD-PointClouds. Other databases exist that target multi-dimensional array data. The solution chosen here is to ‘map’ the multiple dimensions of the point to a single dimension. This takes the form of what will be called a SFC key. SFC is used because of optimal organisation of the dimensions when SFC encoding (Morton or Hilbert) is applied. The SFC is the key of the IOT (single structure).

The main idea for the point cloud management system described here is that fast access to sets of points can be achieved by means of a SFCkey. Each point of the point cloud is stored as a full resolution SFCkey, possibly followed by (other) point attributes. The SFCkey consists of a combination of dimensions. Dimensions are point attributes that play a role in the organisation (structuring) of the data, the SFCkey represents this organisation. Typical dimensions for point clouds are the spatial X, Y, Z dimensions, and time. Other potential dimensions are e.g. importance or level of detail. A point cloud is seen as any collection of points with spatial and other attributes, so in this sense a GPS track is also considered a point cloud. In this last case an ID could become a dimension. From these examples it is clear that also non-spatial attributes can be part of the SFCkey.

The dimensions are included in the SFCkey in such a way that points close together in n-dimensional ‘space’ (nD hypercube) also have a similar key. But obviously not all dimensions are equally prominent, depending on the relative scaling of the dimensions added to the SFCkey. So, the scaling, of dimensions have repercussions for the efficiency of the SFCkey for data retrieval. A consequence of this is that different applications might require different inclusion/scaling of dimensions, hence there exists a requirement for a flexible setup of dimensions and SFCkey.

A distinction can be made between full resolution keys and partial resolution keys. With full resolution keys the original attribute (= dimension) values can be calculated from the SFCkey without loss of information, in this case the values for the dimensions do not have to be stored explicitly. For partial resolution keys this is not possible, all values for dimensions and other attributes have to be stored in addition to the SFCkey (in this last case the SFCkey merely has the function of a cluster and index structure and implies group/blocks of points). The advantage of partial resolution keys is that less space is required for the SFCkey itself. However, when the key is used as primary key for the table in which the point data is stored, it must be unique (and not NULL). This is important when applying the SFCkey in combination with the IOT (as this needs a primary key). For partial resolution keys this is not the case and an option is to include another attribute to the key, or make the key high resolution. So, these are not considered (for the time being). Although not strictly necessary it can be an option to store all dimensions and other attributes explicitly, even if a full resolution SFCkey is used: This would make superfluous the use of a decoding function (less processing at the expense of more storage).

Queries on point clouds in the system described here use a two-stage query process similar to that often used for spatial data. Query geometries have to be translated to range queries on the SFCkey: the primary filter retrieves parts of the hypervolume by specifying a number of key ranges. If more fine ranges are specified the primary filter becomes more selective, but at the same time more ranges have to be processed. Less, but larger ranges, mean coarser resolution with as a result more points selected and as a consequence more work for the secondary filter. Somewhere an optimum will exist for the number of ranges to generate for a query. Range generation is an important function the system has to provide to achieve good query performance.

To a large extent the implementation of the management of point cloud data in a DBMS can make use of existing data types, functions, structures, database tools, and so on. This is mainly due to the reduction of multiple dimensions to a single dimension. But some functions and data structures are specific for the proposed solution and have to be added to the DBMS, these are:

- Metadata for the point cloud dataset and the dimensions
- SFC\_ENCODE to compute SFCkey from the multiple (n) dimensions using the metadata
- SFC\_DECODE to compute the multiple (n) dimensions from the SFC key using the metadata
- SFC\_RANGES to compute SFCkey ranges that cover the nD query geometry
- SFC\_CREATE to set-up and load point cloud data (using metadata/encode) it IOT
- SFC\_DROP to remove the point cloud and related IOT

The metadata is detailed below and is specifying the dimensions, their range and resolution, the actual encoding format of the SFCkey, etc. The encode, decode and range functions need the metadata to be able to work properly. In general, the SFCkey solution is rather generic and can be applied to different DBMSs. However, below we assume Oracle as platform amongst others because the availability of the IOT.

### 3.2 Summary of requirements

Three categories of requirements have been identified for the relational database SFC interface specification for managing nD-PointClouds: Flexibility requirements, Performance requirements, and Ease-of-use requirements. These will now be elaborated on below.

#### 1. Flexibility requirements

- support for multiple dimensions with flexible ordering, scaling
- support for various types of SFC (e.g. Morton, Hilbert)
- support for different encodings of SFCkey (e.g. number, character) for high-res large nD values needing more than 64 (or 128) bits
- support for partial and full resolution keys
- support for implicit (as part of SFCkey) or explicit (as additional attributes) storage of dimensions
- higher dimensional query geometries of various types (more than just an nD box)

#### 2. Performance requirements

- data that will be used together must be stored together
- processing related to storage and query of point cloud data should take place in the environment where the data is stored: in the database
- data movements should be minimised (and data encode/decode conversions with SFC software should be inside database to get access to attribute in full resolution SFCkey)

#### 3. Ease-of-use requirements

- functionality must be available via SQL functions
- despite all tuning options, use of functions should be straightforward (balance: one time tuning via the info in meta-data, which may be a bit harder compared to the actual use of encode/ decode/ range functions as the later are used more often)
- functionality should fit well with other spatial DBMS functionality (e.g. to perform second filter point in polygon test)
- specifying a nD query geometry may be done via nD simplicial complexes or nD regular polytopes, but users are most likely unfamiliar with this, so simpler alternatives needed (e.g. extruded polygon/ polyhedron)
- the number of SFC ranges is a crucial performance factor which user should be able to influence (comparable to resolution, depth in recursion, and involve glueing of ranges)

### 3.3 nD-PointCloud metadata and signatures of the SFC functions

The metadata contain the details of the nD space/hypercube as defined by the organizing dimension and their relative scaling, resolution and encoding. There are two types: one per dataset and the other per organizing dimension. In Oracle the metadata table per dataset could be called USER\_PC\_SFC\_METADATA and contain the following attributes:

#### Dataset metadata

PC_ID	number	system-generated, unique number (link to functions)
PC_NAME	varchar2(32)	name of point cloud dataset
PC_DESCRIPTION	varchar2(256)	description of point cloud
PC_TABLE	varchar2(32)	name final structured table IOT
STAGING_TABLE	varchar2(32)	name intermediate staging table (needed, may be virtual)
PC_DIMENSIONS	number	number of dimensions
PC_NUMBITS	number	number of bits per dimension with which to encode
SFC_TYPE	varchar2(32)	'Morton' or 'Hilbert'
SFC_ENCODING	varchar2(32)	'Number' or 'Character' or 'Base32' or 'Base64' or 'Raw'...
KEY_TYPE	varchar2(32)	'Partial' or 'Full'
DIM_PROPERTIES	varray(PC_DIMENSIONS) of DIM_SPEC	
SRID	number	spatial reference of points in point cloud (in 2D or 3D)
TIME_EPOCH	datetime	start encoded time value (UTC)
TIME_UNIT	varchar2	unit time: year, week, day, hour
TIME_UCT_OFFSET	number	UTC offset

Note: These last 4 items belong to some specific DIM\_SPEC. However, these items carry important meaning for whole point related to spatio-temporal reference. They might be specified per dimension (in DIM\_SPEC), but this would introduce a bit of redundancy (e.g. 3 times same SRID for spatial dims X, Y, and Z). Also, the possible temporal dimension has special meaning. However, not sure if the above is indeed the best decision. What about importance (scale) is this also not having specific predefined meaning, which is needed for perspective view queries.

#### Dimension specific metadata via custom data type DIM\_SPEC (one set of values for each dimension)

DIM.NAME	varchar2(32)	name of dimension
DIM.DESCRPTION	varchar2(256)	description of dimension
DIM.OFFSET	number	value to be added for normalized dimension value
DIM.SCALE	number	scale factor for normalized dimension value
DIM.MINVAL	number	minimal dim value in dataset
DIM.MAXVAL	number	maximum dim value in dataset
DIM.TOL	number	tolerance associated with dim

#### Encode function

```
SFC_ENCODE (PC_ID, VAL_D1, VAL_D2, ... , VAL_Dn)
```

Return: SFCkey

Note: The actual encoding depends on and must be in sync with the metadata (Morton/Hilbert, resolution, etc.)

#### Decode function

```
SFC_DECODE (PC_ID, SFCkey)
```

Return: VAL\_D1, VAL\_D2, ... , VAL\_Dn

#### Range generation function

```
SFC_RANGES (PC_ID, QUERY_GEOMETRY, [recursion depth,] [max_nr_ranges,] [type] ...)
```

Return: table of SFCkey\_LOW, SFCkey\_HIGH values [,range\_type]

Notes:

1. the SFC\_RANGES are used for the primary filter in the query process (if needed a secondary filter should be applied with query geometry to get exact response)
2. if specified, recursion depth indicates the actual resolution (cell size of SFC ranges), and user should be careful because if resolution too high then many ranges will be generated
3. if specified, max\_nr\_ranges then this number of ranges will be returned. In case during computation more ranges are created (due to specified recursion depth), then ranges with smallest gap between them are merged together.
4. some options for the query geometry are: a. nD-hyperbox, b. nD-hypersphere, c. 2D polygon+extrusion min/max other dimensions, d. 3D polyhedron+extrusion min/max other dimensions, e. intersection nD-halfspaces (regular polytope), etc.
5. important special case may be the perspective view (gradually less points when further away from viewer), note this assumes the presence of importance dimension.
6. Some SFC ranges may be complete inside query geometry, some may be on boundary. With the 'type' parameter user may specify how these will be treated in return table

We could have two types of ranges: 1. Completely inside query geometry and 2. On boundary of query geometry. For points with their SFCkey in a range which is completely inside, there is no need to perform a secondary filter (which most likely is expensive to calculate), for sure point is inside. Only in case point has it SFCkey in range which is on the boundary of the query geometry the second filter step with exact query geometry has to be applied. How to include this in the interface of the sfc\_ranges function? One option is to add overlap column to the result of sfc\_ranges function that has either 'full' or 'partial'. Based on the overlap type, the join to the original table can be expressed with or without secondary filter. The parameter 'type' can have 2 values: added or not\_added (last which is default. If it is requested to add the range type, then the return table has 3<sup>rd</sup> column indicating if range is completely inside or on boundary (alternative the type could be encode by a sign to ranges: ranges with + value are inside and ranges with - value are on boundary).

### 3.4 Loading data and example use of SFC functionality in queries

The typical data sources are CSV files or LAS/LAZ files, but many more types do exist depending on nature and background of point cloud (and associated attributes, such as colour (RGB), intensity, but also classification and object id as potential attributes). There may be several options to load the data:

1. Load direct from input files to final, structured table (this may not be the most efficient when the structure is an IOT, as then during adding data the Btree is continuously maintained).
2. Load from input files via staging table (heap) to final, structured table (this may be more efficient as structure is created in a bulk-manner, but an additional staging table is needed). It is to be investigated if this staging table can be virtual, i.e. the actual data is in the input file(s) which are mapped to a database table (without explicit storing/ copying but giving to the user the normal table interface/ feeling, e.g. by means of using a Foreign Data Wrapper, FDW).

Example use of the functionality for loading and assuming a staging table with the input data (option 2), after first filling metadata table (and during actual conversion to the IOT/SFC structure also sfc\_encode function is used):

1. Filling metadata:

```
Insert into user_pc_sfc_metadata values (1, 'my test', 'a test point cloud',
'my_pc', 'test_pc_staging', 3, 21, 'Hilbert', 'Number', 'Full', 28992, -1,
'na', 0)
```

2. Loading a table:

```
Select sfc_create(pc_id);
```

The query geometry is converted to the relevant SFC ranges. Below first example query to just get the SFC ranges for a specific query geometry:

```
select * from sfc_ranges(pc_id, query_geom);
  min | max
-----+-----
   502 | 503
   507 | 508
   608 | 609
(3 rows)
```

This can be combined in a query with the pc data table in join to obtain the selected points according to the SFC ranges:

```
select sfc_decode(pc_id, key)
from my_pc pc, sfc_ranges(pc_id, query_geom) range
where pc.key between range.min and range.max
```

Next, an example with secondary filter for ranges on boundary (partial outside and outside):

```
Select sfc_decode(my_pc, key)
from pc_table pc, sfc_ranges(my_pc, query_geom, 'added') range
where (range.type='full' and pc.key between range.min and range.max)
or (range.type = 'partial' and
pc.key between range.min and range.max and
overlaps(query_geom, geometry(sfc_decode(my_pc, key))))
```

Drop a point cloud:

```
Select drop_sfc_pc(pc_id);
```

## 4 Conclusions and some future work

In this paper we argued that point cloud should be treated as a first-class representation (similar to vector or raster representations). The nD-PointCloud was presented as the solution for this, offering deep integration of space, time and scale. For efficiency reasons spatial indexing and clustering of these large point clouds is extremely important and the Space Filling Curved (SFC) was introduced for this purpose. In order to get beyond the current state of the art of storing/managing point clouds in files, a DBMS solution was presented (with all benefits: integration with other data types, scalability, multi-user, transaction support, etc.). Finally, a DBMS SFC interface specification for point clouds was proposed. This proposal will be further discussed within standardization organizations, such as the OGC and more specifically the Domain Working Group (DWG) on point clouds.

There is still a significant amount of future work ahead. We will now focus of the challenge of having rather precise queries without too many SFC ranges. As indicated earlier generating high resolution ranges for a query geometry is good for precision (reporting right points) but the amount of ranges may get so high that we can lose reliable performance. An example is when we have high resolution nD keys (e.g. with n=5) and specifying in the query geometry just 2 dimensions and then leave the other 3 dimensions open (nothing specified). This implies the open dimensions extend to infinity and will generate a lot of SFC ranges over there (most likely without points). Some improvements are possible here. Simplest solution: if we know the actual data values of the points in all dimensions of the key, we can (implicitly) add this knowledge to the query geometry. In this case we do not generate SFC ranges where we do not have data. However, we do need to keep track of the actual data and maintain statistics (at least min and max values per dimension) when loading or adding data. Also, the range generation may become data dependent, meaning that when one more point is inserted stretching the current actual data range, then the sfc\_ranges function before and after the insert will return different ranges.

Taking this one step further than just keeping track of points or no points in a certain part of the nD space: even if there is 1 point (or a few points) in a certain part of the nD space, why generate many ranges if the data is so sparse over there. This is wasting ranges. So, the next level of refinement is having a histogram with the actual data distribution (density). How would such a histogram look like for the nD point cloud? It could be an nD array (corresponding to nD hypercube) of a certain resolution, but what resolution, if too coarse then it will be of little use when generating ranges, but if too detailed then it will become too heavy (a lot of overhead). A proposed, but still to be implemented and tested solution is to collect statistics at the  $2^n$ -tree internal nodes during the actual loading and computing of the SFC keys that corresponds to the recursion of the SFC computations. At top level a single value contains a total count of all points in the data set. At second level the points are counted per sub-cell (if n=3 then 8 children, each with their own count and the sum of the 8 counts same value as parent count). If count is 0, then no need to represent this branch of the tree (and also no need to refine for its children) and nothing wasted on empty part of this histogram. With this recursive nD-histogram we can decide to descend or not when generating ranges for a certain query geometry based on the estimated number of points. So, if this is a low number of point (even at high level in tree/ recursion) we can stop refining the range and be happy with course range (big) cell as it does not contain many points. If there is still a high number of points then we should further refine this large range (heavy cell) into smaller cells even if we are at already medium or lower levels (apparently there is a local very high density of points, which justifies refining the ranges).

## Acknowledgements

We would like to thank Mike Horhammer from Oracle and Xuefueng Guan from Wuhan University for their active participation in developing the presented ideas related to managing nD-PointClouds. We would further like to thank the TU Delft MSc Geomatics students doing point cloud projects, amongst others, Stella Psomadaki and Irene de Vreede.

## Literature

- ARIKAN, M., R. PREINER, C. SCHEIBLAUER, S. JESCHKE, M. WIMMER (2014), Large-Scale Point-Cloud Visualization through Localized Textured Surface Reconstruction. *IEEE Transactions on Visualization and Computer Graphics*, 20(9):1280-92.
- BADER, M. (2012), Space-filling curves: an introduction with applications in scientific computing. Springer Science & Business Media.
- BUTZ, A.R. (1971), Alternative algorithm for Hilbert's space-filling curve. *IEEE Tr. on Computers*, 424–426.
- FRIIS-CHRISTENSEN, C.S. & A. JENSEN (2003), Object-relational management of multiply represented geographic entities. Cambridge, MA, USA, July 9–11, 2003. *Proceedings of the Fifteenth International Conference on Scientific and Statistical Database Management*.
- GARGANTINI, I. (1982), An effective way to represent quadtrees. *Commun. ACM* 25(12), pp. 905–910.
- GUAN, X., P. VAN OOSTEROM & B. CHENG (2018), A Parallel n-dimensional Space-Filling-Curve Library and Application in Massive Point Cloud Management. In: *ISPRS Int. J. Geo-Inf.*, MDPI AG, 7(8), 19 pages.
- HÄGERSTRAND, T. (1970), What about people in regional science?. *Papers of the Regional Science Association* 24 (1): 6–21.
- JAFFER, A. (2014), Recurrence for Pandimensional Space-Filling Functions. arXiv preprint arXiv:1402.1807.
- JONES C.B., D.B. KIDNER, L.Q. LUO, G.L. BUNDY & J.M. WARE (1996), Database design for a multi-scale spatial information system. *International Journal Geographic Information Science*, Vol. 10, pp. 901–920.
- KILPELAINEN, T. (1997), Multiple representation and generalisation of geo-databases for topographic maps. PhD thesis. Finnish Geodetic Institute.
- LAWDER, J.K. & P.J.H. KING (2001), Querying multi-dimensional data indexed using the Hilbert space-filling curve. *SIGMOD Rec.* 30, pp. 19–24.
- LIU, H., P. VAN OOSTEROM, M. MEIJERS & E. VERBREE (2018), Towards 10<sup>15</sup>-level point clouds management - a nD PointCloud structure, In: *Proceedings of the 21th AGILE Conference on Geographic Information Science*, Lund, pp. 7.
- MARTINEZ-RUBI, O., M. KERSTEN, R. GONÇALVES, R. & M. IVANOVA (2014), A column-store meets the point clouds. FOSS4GEurope Academic Track.
- MARTINEZ-RUBI, O., S. VERHOEVEN, M. VAN MEERSBERGEN, M. SCHÜTZ, P. VAN OOSTEROM, R. GONÇALVES & TH. TIJSSSEN (2015), Taming the beast: Free and open-source massive point cloud web visualization, In: *Capturing Reality Forum 2015*, Salzburg, 12 pages.
- MEIJERS, M., W. QUAK & P. VAN OOSTEROM (2017), Archiving AIS messages in a Geo-DBMS, In: *Proceedings of the 20th AGILE Conference on Geographic Information Science (Arnold Bregt, Tapani Sarjakoski, Ron van Lammeren, Frans Rip, eds.)*, Wageningen University & Research, 3 pages.
- MOLENAAR, M. (1998), An introduction to the theory of spatial object modelling for GIS, Taylor and Francis (CRC Press), London, p. 200.
- PSOMADAKI, S., P. VAN OOSTEROM, TH. TIJSSSEN & F. BAART (2016), Using a Space Filling Curve Approach for the Management of Dynamic Point Clouds, Chapter in: *ISPRS Annals, IV-2/W1*, pp. 107-118.
- VAN OOSTEROM, P. & T. VIILBRIEF (1996), The spatial location code. In: *Proceedings of the 7th International Symposium on Spatial Data Handling*, Delft, the Netherlands.
- VAN OOSTEROM, P. (1999), Spatial Access Methods, Chapter T2.3 in *Geographical Information Systems Principles, Technical Issues, Management Issues, and Applications* (edited by Longley, Goodchild, Maguire en Rhind), Wiley, pages 385-400 (vol.1).
- VAN OOSTEROM, P. & M. MEIJERS (2012), Method and system for generating maps in an n-dimensional space. Technische Universiteit Delft, assignee. Patent NL2006630C / WO 2012144893 A2. 26 Oct. 2012. (<https://patents.google.com/patent/WO2012144893A2/en>).
- VAN OOSTEROM, P., O. MARTINEZ-RUBI, M. IVANOVA, M. HORHAMMER, D. GERINGER, S. RAVADA, T. TIJSSSEN, M. KODDE & R. GONÇALVES (2015), Massive point cloud data management: design, implementation and execution of a point cloud benchmark. *Computers & Graphics* 49, 92–125.
- VAN OOSTEROM, P., O. MARTINEZ-RUBI, TH. TIJSSSEN & R. GONÇALVES (2016), Realistic Benchmarks for Point Cloud Data Management Systems, Chapter in: *Advances in 3D Geoinformation*, pp. 1-30.
- DE VREEDE, I. (2016), Managing Historic Automatic Identification System data by using a proper Database Management System structure, Master's thesis, Delft University of Technology, pp. 90.

WANG, J. & J. SHAN (2005), Space filling curve based point clouds index, Proceedings of the 8th International Conference on GeoComputation, pp. 551–562.

# Verfahren zur Auswertung, Analyse und Nutzung von 3D-Punktwolken

Rico RICHTER, Sören DISCHER und Jürgen DÖLLNER

Hasso-Plattner-Institut · Universität Potsdam · Prof.-Dr.-Helmert-Str. 2-3 · 14482 Potsdam

E-Mail: rico.richter@hpi.de; soeren.discher@hpi.de; juergen.doellner@hpi.de

## 1 Einführung 3D-Punktwolken

Die digitale Erfassung von Bauwerken, Städten und Ländern ist mit LiDAR-Technologie und photogrammetrischen Verfahren effizient, automatisiert und kostengünstig möglich. Terrestrische Scanner, mobile Erfassungssysteme (z.B. Fahrzeuge) und luftgestützte Erfassungssysteme (z.B. UAVs, Flugzeuge) werden genutzt um 3D-Punktwolken zu erzeugen. 3D-Punktwolken haben keine innere Struktur oder Regularität und sind deshalb Ideal für die Repräsentation von beliebigen Strukturen und Objekten geeignet. Sie bilden die Grundlage für eine Vielzahl von Anwendungen aus unterschiedlichen Bereichen, wie z.B. der Stadtplanung, dem Bauwesen, dem Katastrophenmanagement oder der Umweltüberwachung (NEBIKER, BLEISCH & CHRISTEN 2010). All diese Anwendungen erfordern typischerweise nur eine Teilmenge der Daten (z.B. Gebäude-, Vegetations-, oder Bodenpunkte), so dass eine Analyse und Klassifizierung der Erfassungsdaten erforderlich ist. Bedingt durch die üblicherweise extrem hohen Datenmengen (beispielsweise im Petabyte-Bereich für ganze Länder) sind leistungsfähige Algorithmen und Softwaresysteme erforderlich, um das volle Potential von 3D-Punktwolken zu erschließen und Informationen bedarfsgerecht bereitzustellen.

In diesem Beitrag werden Verfahren für die Analyse und Visualisierung von 3D-Punktwolken vorgestellt. Neben einer Klassifizierung der Daten in unterschiedliche Oberflächenkategorien ist die Veränderungsanalyse wesentlich. Durch regelmäßige Erfassungen können so Aufnahmen von unterschiedlichen Zeitpunkten miteinander verglichen werden. Weitere Themen sind die flächendeckende Ableitung von Baumstandorten und die Web-Visualisierung, die es ermöglicht 3D-Punktwolken einer Vielzahl von Anwendern zugänglich zu machen.

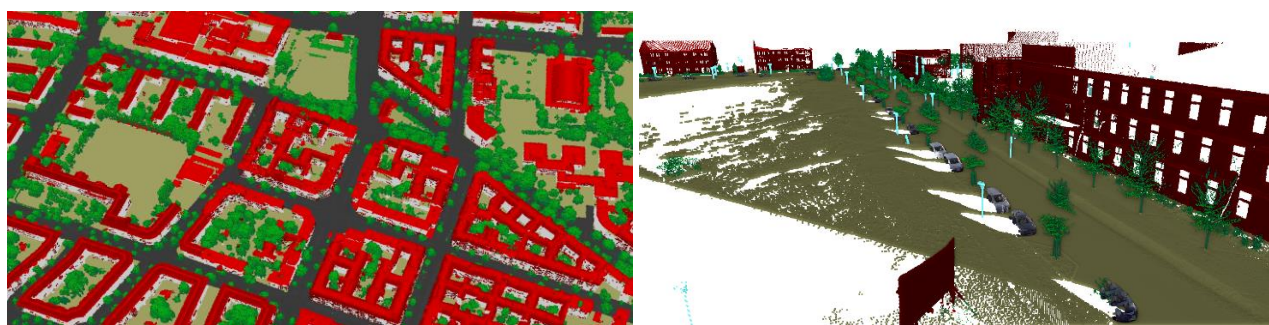
## 2 Zusammenfassung des Vortragsinhalts

### 2.1 Datenerfassung

Die Verfügbarkeit von 3D-Punktwolken hat in den letzten Jahren stark zugenommen. Das liegt vor allem daran, dass aus überlappenden Luftbildern, die viele Städte jährlich erheben, auch 3D-Punktwolken abgeleitet werden. Je nach Bildauflösung lassen sich mit photogrammetrischen Verfahren 3D-Punktwolken mit 25-100 Punkten pro m<sup>2</sup> erzeugen. Auch Landesvermessungsämter befliegen regelmäßig, so dass aktuelle 3D-Punktwolken und Luftbilder bezogen werden können. Neben den Befliegungen etabliert sich zunehmend die flächeneckende Erfassung des Straßenraumes mit Mobile-Mapping-Fahrzeugen. Hierbei kommt oft eine Kombination aus LiDAR und bildbasierter Erfassung zum Einsatz, so dass neben 3D-Punktwolken auch Panoramabilder des Straßenraums zur Verfügung stehen.

### 2.2 Klassifikation

Klassifikationsverfahren für 3D-Punktwolken werden eingesetzt, um Punkte mit semantischen Informationen anzureichern. Basierend auf einer Analyse der Punktwolkenstruktur wird jeder Punkt auf Basis der lokalen Punktumgebung charakterisiert, so dass er einer Objektklasse zugeordnet werden kann (GRILLI, MENNA & REMONDINO 2017). Abbildung 1 zeigt das Klassifizierungsergebnis für eine Punktwolke aus einer Befliegung (links) und einer Mobile-Mapping-Befahrung (rechts).

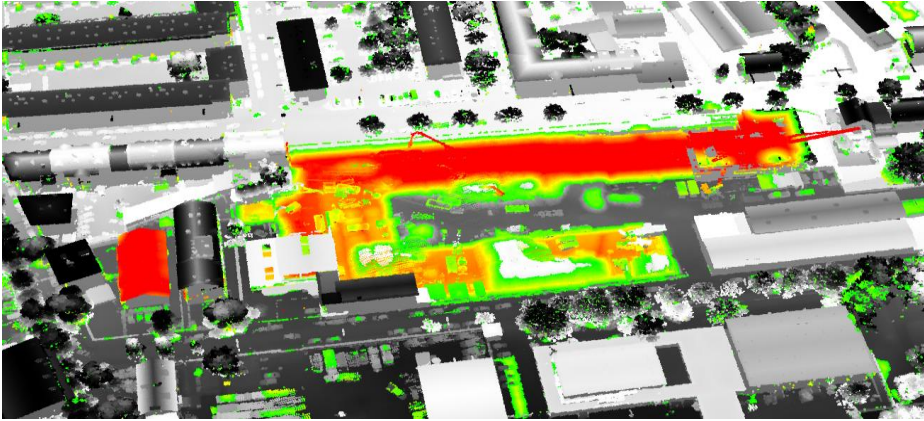


**Abb. 1:** Klassifizierung einer 3D-Punktwolke in Bebauung (rot), Vegetation (grün), Boden (braun) und Stadtmobiliar (blau).



## 2.2 Veränderungsanalyse

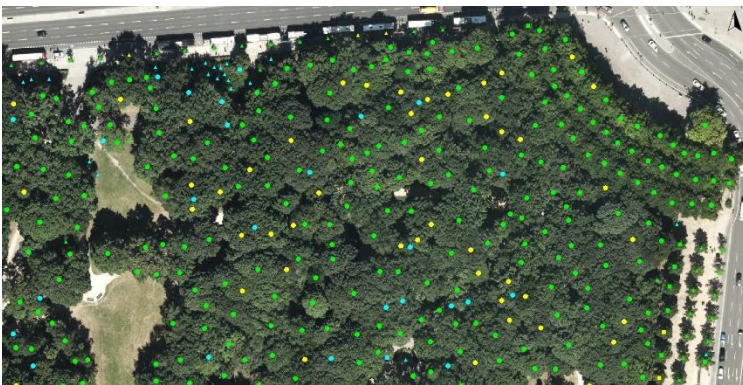
Der Vergleich von 3D-Punktwolken aus Erfassungen des gleichen Gebietes zu unterschiedlichen Zeitpunkten ermöglicht das Erkennen von Veränderungen. Veränderungen sind dabei Objekte und Strukturen, die hinzugekommen sind oder entfernt wurden. Bei regelmäßigen Erfassungen betreffen Veränderungen typischerweise nur einen kleinen Teil des Erfassungsgebiets, so dass durch eine Erkennung und Quantifizierung von Veränderungen eine effiziente Aktualisierung von bestehenden Datenbeständen (z.B. 3D-Stadtmodelle, Vegetationsbestände, Geländemodelle) möglich ist (RICHTER, KYPRIANIDIS & DÖLLNER 2013). Abbildung 2 zeigt die Ergebnisse einer Veränderungsanalyse für ein Gebäude und eine Baustelle.



**Abb. 2:** 3D-Punktwolke dargestellt mit Ergebnissen der Veränderungsanalyse. Ein Farbverlauf von grün (wenig) nach rot (stark) zeigt die Stärke der Veränderung.

## 2.3 Anwendungen

Auf Basis der Klassifizierung und Veränderungsanalyse lässt sich eine Vielzahl von Anwendungen realisieren. Ein Beispiel ist die flächendeckende Ableitung von Baumstandorten. Dazu werden aus der klassifizierten 3D-Punktwolke alle Vegetationspunkte ermittelt um danach durch eine Strukturanalyse Einzelbaumstandorte zu bestimmen. Neben der Lage können Höhe, Kronendurchmesser, Volumen und Farbe automatisiert bestimmt werden. Das Ergebnis kann dann beispielsweise mit einem bestehenden Kataster abgeglichen werden, um dieses zu ergänzen, zu aktualisieren oder auf Plausibilität zu prüfen (OEHLKE, RICHTER & DÖLLNER 2015). Abbildung 3 zeigt beispielhaft die ermittelten Baumstandorte auf Basis einer 3D-Punktwolke.



**Abb. 3:** Ergebnisse der Einzelbaumerkennung auf Basis von 3D-Punktwolken.

Auch 3D-Stadtmodelle lassen sich durch den Vergleich von Bestandsdaten mit klassifizierten 3D-Punktwolken aktualisieren. Dabei lassen sich Gebäude in unterschiedliche Kategorien einteilen: Unverändert, baulich verändert, neu oder nicht mehr vorhanden.

## 2.3 Web-Visualisierung

Die Visualisierung und Bereitstellung von 3D-Punktwolken ist für viele Anwender wesentlich. Das interaktive Explorieren, Messen von Strecken und Flächen, Anzeigen von Analyseergebnissen und Teilen mit anderen Nutzern sind dabei wesentliche Anforderungen. Aufgrund der enormen Datenmenge bieten sich hier web-basierte Lösungen an, die durch

die Verwendung von Streaming-Verfahren und Level-of-Detail-Konzepten die Hardwarevoraussetzungen für Nutzer signifikant reduzieren. Die fusionierte Visualisierung von 3D-Punktwolken aus unterschiedlichen Erfassungen in Kombination mit Analyseergebnissen ist eine wesentliche Voraussetzung, um die Daten und deren Zusammenhänge zu verstehen und eine Informationsgrundlage für eine Vielzahl von Anwendungen zu schaffen (DISCHER, RICHTER & DÖLLNER 2018).

## Literatur

- DISCHER, S., RICHTER, R. & DÖLLNER, J. (2018), A Scalable WebGL-based Approach for Visualizing Massive 3D Point Clouds using Semantics-Dependent Rendering Techniques. In: *Web3D '18: The 23rd International Conference on Web3D Technology*, 19. Poznan.
- GRILLI, E., MENNA, F. & REMONDINO, F. (2017), A review of point clouds segmentation and classification algorithms. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives* 42(2W3), 339–344.
- NEBIKER, S., BLEISCH, S. & CHRISTEN, M. (2010) Rich point clouds in virtual globes – A new paradigm in city modeling? In: *Computers, Environment and Urban Systems* 34(6), 508–517.
- OEHLKE, C., RICHTER, R. & DÖLLNER, J. (2015), Automatic Detection and Large-Scale Visualization of Trees for Digital Landscapes and City Models based on 3D Point Clouds. In: *16th Conference on Digital Landscape Architecture*, 151–160.
- RICHTER, R., KYPRIANIDIS, J. E., & DÖLLNER, J. (2013), Out-of-Core GPU based Change Detection in Massive 3D Point Clouds. In: *Transactions in GIS*, 17(5), 724–741.

# Scan2BIM... und warum die „2“ nicht nur eine Primzahl ist.

Jan MARTENS und Jörg BLANKENBACH

Geodätisches Institut und Lehrstuhl für Bauinformatik & Geoinformationssysteme · RWTH Aachen University · Mies-van-der-Rohe-Str. 1 · 52074 Aachen

E-Mail: jan.martens@gia.rwth-aachen.de, blankenbach@gia.rwth-aachen.de

## 1 Einführung

Die Erfassung und Dokumentation der bebauten Umwelt stellt für die Planung, Erstellung und den Betrieb baulicher Infrastruktur eine wichtige Grundlage dar. Die Innovationen in der Messtechnik haben in den letzten Jahren leistungsfähige Erfassungssysteme hervorgebracht, die eine effiziente und flächenhafte digitale Erfassung der bebauten Umwelt, insbesondere mit Hilfe von Kameras und/oder Laserscannern, erlauben. Gleichzeitig hat der digitale Wandel zum Aufkommen neuer Verfahrensweisen geführt. Im Bauwesen ist hier insbesondere das Building Information Modeling (BIM) zu nennen, bei dem dreidimensionale, bauteilorientierte Bauwerksmodelle eine wesentliche Grundlage darstellen. In der Praxis vor allem als kooperative Arbeitsmethode in der Neuplanung eingesetzt, gewinnt die Methode BIM auch für Bestandsbauwerke zunehmend an Bedeutung, z.B. für das Planen und Bauen im Bestand oder das Facility Management bzw. Erhaltungsmanagement. Für Bestandsbauwerke muss jedoch vielfach zunächst eine digitale as-built bzw. as-is Erfassung mit anschließender Erstellung des „digitalen Zwillings“ des realen Bauwerks erfolgen.

Punktwolken aus einer Laserscanning-Erfassung oder abgeleitet aus Bildverbänden (Bildmatching) stellen häufig die Datengrundlage für die Erstellung der digitalen Bauwerksmodelle dar („Scan2BIM“). In der Praxis erfolgt dieser Schritt überwiegend durch manuelles Modellieren der BIM-Objekte aus den Eingangsdaten. Dieses Vorgehen ist jedoch mit hohem Zeit- und Kostenaufwand verbunden, der den Erfassungsaufwand in der Regel deutlich übersteigt. Eine höhere Automatisierung in der gesamten Datenverarbeitungskette ist daher wünschenswert.

Im Beitrag soll ein Überblick über das Vorgehen zum automatisierten Modellieren gegeben und einzelne Verfahren exemplarisch beschrieben werden. Insbesondere die Punktwolkenerfassung sowie verschiedene Verarbeitungsschritte und deren Integration in State-of-the-Art BIM-Software werden beispielhaft vorgestellt und deren Beziehung zueinander untersucht. Neben der geometrischen Analyse und Segmentierung werden Ansätze zur Rekonstruktion und Klassifizierung von BIM-Komponenten diskutiert, welche den Sprung von Scan-*to*-BIM vervollständigen.

## 2 Von der Punktwolke zum Modell

Als grobe Unterteilung der Arbeitsschritte lassen sich für den automatisierten Scan2BIM-Vorgang vier Schritte definieren: Die Punktwolkenerfassung und Vorverarbeitung, die Punktwolkensegmentierung, die Klassifizierung der Segmente sowie das Erzeugen von BIM-Objekten aus den klassifizierten Segmenten.

### 2.1 Datenerfassung

Die Erfassung der Daten lässt sich mit verschiedenen Methoden realisieren, wobei einige davon in den letzten Jahren aufgrund neuer technischer Entwicklungen an Attraktivität gewonnen haben. Terrestrisches Laserscanning (TLS) hat sich mittlerweile für die präzise flächenhafte 3D-Erfassung etabliert; die Geräte sind darüber hinaus portabler und preisgünstiger geworden. Mobile Laserscanner stellen hingegen eine neuere Entwicklung dar und ermöglichen schnellere Bestandserfassung auf Kosten von geringerer Messgenauigkeit (RIISGAARD, S. & BLAS M. R. 2005). Time-of-Flight (TOF) Kameras bieten neben der bildbasierten Punktwolkenerzeugung kostengünstige Alternativen zur räumlichen Erfassung, leiden jedoch unter erhöhtem Rauschen in den Daten.

### 2.2 Vorverarbeitung

Die erfassten oder erzeugten Punktwolken bedürfen in der Regel einiger Vorverarbeitungsschritte, z.B. das Reduzieren von Rauschen („Denoising“) und Ausdünnen. Auch die Ausrichtung der Punktwolke entlang der lokalen Koordinatenachsen gehört zur Vorverarbeitung, um weitere Arbeitsschritte zu vereinfachen. Speziell vor dem manuellen Modellieren handelt es sich hierbei oft um den ersten Bearbeitungsschritt, aber auch für automatisierte Verfahren ist dies bei Verwendung von Datenstrukturen wie Voxelgittern und Octrees (WANG, CHO & KIM 2015) vorteilhaft. Automatisierte Verfahren profitieren bei ausgerichteten Punktwolken von einer höheren Genauigkeit, wenn Daten auf Voxel Ebene berechnet werden und Octrees fallen oft deutlich kompakter aus.

### 2.3 Segmentierung

Aufbauend auf den Vorverarbeitungsschritten lassen sich Streuungswerte berechnen, etwa um Kanten in der Punktwolke zu detektieren. Diese charakteristischen Werte eignen sich insbesondere für die Rekonstruktion von Bauwerksgeometrie und zum Verfeinern extrahierter Strukturen. Weitere Verfahren wie RANSAC (JUNG, HONG & JEONG 2014) eignen sich zum Extrahieren ebener Flächen. Durch Einbinden von Annahmen und Bedingungen können die detektierten Flächen insbesondere in ausgerichteten Punktwolken von Störobjekten unterschieden werden.



**Abb. 1:** Laserscan-Aufnahme eines Raums (links) und automatisch segmentierte Flächen in explodierter Ansicht (rechts).

### 2.4 Klassifizierung und Objektbildung

Die beiden letzten Schritte, die Klassifizierung und Objektbildung, sind eng miteinander verzahnt. Je nach Textur der detektierten, ebenen Flächen kann eine Klassifizierung Bauteilmaterialien bestimmen. Eine Anpassung bewährter Texturklassifizierungsverfahren aus dem Bereich der Bilderkennung erscheint hierbei in Kombination mit Techniken des maschinellen Lernens wie Clustering oder Support Vector Machines vielversprechend (MALISIEWICZ, GUPTA & ELFROS 2011).

## Fazit

Der aktuelle Stand in der Messtechnik erlaubt die dreidimensionale, flächenhafte Erfassung der bebauten Umwelt mit hoher räumlicher Auflösung und Genauigkeit. Die Ableitung von digitalen, objektbasierten Modellen stellt in der Praxis jedoch noch immer einen zeitaufwändigen Prozess mit vielen manuellen Arbeitsschritten dar. Für eine effiziente Modellableitung gilt es daher Verfahren zu entwickeln, welche den Ablauf höher automatisieren. Exemplarisch wurden im Beitrag Ansätze in der Prozesskette von der Vorverarbeitung über die Segmentierung bis hin zur Klassifizierung und Objektbildung aufgezeigt.

## Literatur

- RIISGAARD, S. & BLAS, M. R. (2005), SLAM for Dummies: A Tutorial Approach to Simultaneous Localization and Mapping., Online: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.208.6289>, Zugriff am 17.01.2019.
- WANG, C., CHO, Y.K. & KIM C. (2015), Automatic BIM component extraction from point clouds of existing buildings for sustainability application. In: Automation in Construction, Band 56, Seiten 1-13.
- JUNG, J., HONG, S. & JEONG, S. (2014), Productive modeling for development of as-built BIM of existing indoor structures. Automation in Construction, Band 42, Seiten 68-77.
- MALISIEWICZ, T., GUPTA A. & EFROS A. (2011), Ensemble of Exemplar-SVMs for Object Detection and Beyond. In: ICCV.

# Hauptsache Open! – Der Zugang zu offenen Geodaten

Ralf BILL

Professur für Geodäsie und Geoinformatik · Universität Rostock · Justus-von-Liebig-Weg 6 · 18056 Rostock  
E-Mail: ralf.bill@uni-rostock.de

## 1 Mehr Offenheit

„Aktuell findet ein Paradigmenwechsel zu mehr Offenheit in der Gesellschaft, der Verwaltung, der Wirtschaft und auch der Wissenschaft statt. Noch vor wenigen Jahren wurden Informationen sorgsam gehütet. Sie wurden in Unternehmen als Betriebsgeheimnis gesehen, Verwaltungen sammelten sie in Ordnern und Planschränken. Auskunftsbegehren wurden fast schon als Majestätsbeleidigung angesehen. Wissenschaftler publizierten zwar, jedoch ohne die genutzten Daten zur Verfügung zu stellen, womit die Nachvollziehbarkeit der Ergebnisse deutlich eingeschränkt war. Dies ändert sich gerade dramatisch. Informationsfreiheit, Bürgerbeteiligung, Offene Daten, Transparenzgesetze, Digitalisierung. Neue Schlagworte und Aktivitäten, die andeuten, dass Offenheit in den unterschiedlichsten Bereichen im Vormarsch ist.“[1]

Diese „Open-Bewegung“ ist mit Begriffen wie Open Source Software, Open Science, Open Data, Open Content, Open Access, Open Innovation, Open Government, Open Educational Resources u.v.a. verbunden. Hinter der Open-Bewegung stehen daher unterschiedliche Organisationen wie die Open Knowledge Foundation (OKF) [2], die Open Science Foundation [3] oder die Open Source Initiative (OSI) [4], die den Begriff Offenheit auch durchaus leicht unterschiedlich definieren. Und auch die Politik hat sich spätestens mit der Unterzeichnung der G8 Open Data Charter 2013 oder der International Open Data Charter 2015 dazu bekannt, so viele Daten wie möglich in einer Form, die für alle zugänglich ist, zu veröffentlichen. Allen diesen Begrifflichkeiten und Initiativen gemeinsam ist die Annahme, dass Offenheit von Daten/Wissen/Software/... die Entwicklung der Wissenschaft, Wirtschaft und Verwaltung befördert und zu mehr Transparenz, Partizipation und Kooperation, auch bis zum Bürger, führt.

Der Beitrag stellt kurz die Entwicklung hin zu mehr Offenheit in Gesellschaft, Verwaltung, Wirtschaft und Wissenschaft vor. Der Fokus des Artikels liegt dabei auf dem Angebot an offenen Geodaten/Geoinformationen aus der Verwaltung, also Open Government Data.

## 2 Offene (Geo-)Daten

### 2.1 Datenangebote – ein Überblick und eine Systematisierung

Im Beitrag betrachten wir offene Daten (Open Data), die im Open Data Handbook [5] wie folgt definiert sind:

Offene Daten sind Daten, die von jedermann frei benutzt, weiterverwendet und geteilt werden können – die einzige Einschränkung betrifft die Verpflichtung zur Nennung des Urhebers.

Mit Blick auf das Angebot an Geodaten/Geoinformationen wird im Beitrag nach der Herkunft resp. Anbietern und Ausprägungen – nicht immer dieser umfassenden Definition von Open streng gerecht werdend – unterschieden in Datenangebote von Verwaltungen, der Wirtschaft, der Wissenschaft und der Zivilgesellschaft. Fragen der Verfügbarkeit, der Formate und Lizenzen/Nutzungsbedingungen von Portalen auf unterschiedlichen Ebenen werden diskutiert. Ebenso werden Chancen und Potenziale offener Daten, aber auch aktuelle Hemmnisse angesprochen. (BILL, 2018)

### 2.2 Das Open Data Portal in OpenGeoEdu

Da klassische Suchmaschinen i.d.R. wenig geeignet sind, um Geodaten aufzufinden, wurde im Projekt OpenGeoEdu ein Portal der Datenportale [6] entwickelt, welches Nutzer bei der Suche nach Geodaten unterstützen und durch Nutzer komplettiert werden soll. Eine Beschreibung der Architektur und Inhalte des Portals findet sich in (HINZ & BILL, 2018a/b). Aktuell sind hier etwa 300 Datenportale in Deutschland, Österreich und der Schweiz gelistet, eine Übersicht nach Reichweite und Art des Portals gibt Tabelle 1.

**Tabelle 1:** Anzahl der Datenportale nach Portaltyp, Land und Reichweite

Reichweite		international	national	regional	kommunal
Typ	Land				
Open Data Portal (97)	Deutschland (68)	2	12	10	44
	Österreich (20)	1	8	8	3
	Schweiz (7)	0	3	3	1
	Sonstige (2)	2	0	0	0
GDI/Geoportal (93)	Deutschland (68)	1	4	32	31
	Österreich (12)	0	3	9	0
	Schweiz (10)	0	2	8	0
	Sonstige (3)	1	0	2	0
Statistikportal (66)	Deutschland (21)	0	6	13	2
	Österreich (11)	0	1	9	1
	Schweiz (31)	0	3	23	5
	Sonstige (3)	1	2	0	0
Umweltportal (25)	Deutschland (23)	0	6	17	0
	Österreich (1)	0	1	0	0
	Schweiz (0)	0	0	0	0
	Sonstige (1)	1	0	0	0
Forschungsdatenportal (9)	Deutschland (7)	4	0	3	0
	Österreich (0)	0	0	0	0
	Schweiz (1)	1	0	0	0
	Sonstige (1)	1	0	0	0
Citizen Science Projekt (5)	Deutschland (5)	3	0	0	2
	Österreich (0)	0	0	0	0
	Schweiz (0)	0	0	0	0
	Sonstige (0)	0	0	0	0
<b>Datenportale insgesamt: 295 (DE: 192, AU: 44, CH: 49, Sonst: 10)</b>					

## Literatur

- BILL, R. (2018): Offene Geodaten – ein Paradigmenwechsel. In: GIS-Report 2018/2019. Harzer Verlag.
- HINZ, M., BILL, R. (2018a): Offene Geodaten – Mehr Transparenz durch ein Portal der Open Data Portale. In: Bill, R.; Zehner, M.; Lerche, T.; Schröder, J. (Hrsg.): GeoForum MV 2018 - Geoinformation und Digitalisierung. Berlin: GITO mbH. S. 49 - 58.
- HINZ, M., BILL, R. (2018b): Ein zentraler Einstiegspunkt für die Suche nach offenen Geodaten im deutschsprachigen Raum. In: Universität Salzburg | IFFB Geoinformatik – Z\_GIS (Hrsg.): AGIT Journal für Angewandte Geoinformatik. Ausgabe 4-2018. S. 298-307. doi:10.14627/537647038

## Internetquellen:

- [1] OpenGeoEdu | Offener Onlinekurs OpenGeoEdu. Kurseinheit Open XX: <https://learn.opengeoedu.de/openx/vorlesung>
- [2] Open Knowledge Foundation Deutschland: <https://okfn.de/>
- [3] Open Science Foundation: [openseiencefoundation.eu/](https://openseiencefoundation.eu/)
- [4] Open Source Initiative: <https://opensource.org/>
- [5] Open Data Handbook: [opendatahandbook.org](https://opendatahandbook.org)
- [6] OpenGeoEdu | Open Data Portal: <https://portal.opengeoedu.de/>

## Danksagung

Der Verfasser dankt dem Bundesministerium für Verkehr und Digitale Infrastruktur (BMVI) für die Förderung im Rahmen des mFUND-Programms (FKZ: 19F2007A).

# Visualisierung von Echtzeitgeodaten in der Urban Data Platform in Hamburg

Sven BEYEL und Michael FISCHER

Landesbetrieb Geoinformation und Vermessung Hamburg · Neuenfelder Straße 19 · 21109 Hamburg  
E-Mail: sven.beyel@gv.hamburg.de; michael.fischer1@gv.hamburg.de

## 1 Einführung ins Thema

Die Freie und Hansestadt Hamburg hat die Strategie „Digitale Stadt“ und das Vorhaben „Digital First“ auf den Weg gebracht. Damit möchte Hamburg die Chancen der Digitalisierung zur Verbesserung seiner Lebensqualität und seiner wirtschaftlichen Attraktivität einsetzen. Eines von mehreren Digitalisierungsvorhaben ist die, in Abstimmung mit der Leitstelle Digitale Stadt der Senatskanzlei, erfolgte Einrichtung einer Organisationseinheit, dem Urban Data Hub und die dafür nötige technische Infrastruktur, die Urban Data Platform.

In fast jedem Bereich der Stadt nimmt die Digitalisierung weiter zu. Es sind bereits eine Vielzahl an Daten aus den Bereichen Ver- und Entsorgung, Verwaltung, Gesundheit, Bauwesen, Mobilität, Logistik, Transport und Wirtschaft vorhanden – aber bisher kaum miteinander vernetzt. Das ändert sich durch die Urban Data Platform. Die Integration und Vernetzung von städtischen Daten über standardisierte Schnittstellen wird durch diese zentrale Datendrehscheibe schnell und einfach realisiert.

Viele Städte und Kommunen stehen aktuell zusätzlich vor der Herausforderung, Sensordaten bzw. Echtzeitdaten aus den Bereichen Verkehr, Energie, Wasser und Umwelt miteinander zu verknüpfen und in Echtzeit auswertbar zu machen bzw. zur Verfügung zu stellen (EUROPÄISCHE KOMMISSION 2019). In Hamburg wird diese Aufgabe ebenfalls über die Urban Data Platform realisiert. Die im Kern auf der Geodateninfrastruktur und somit auf langjährigem Knowhow im Datenmanagement basierende Urban Data Platform wird um eine Echtzeitdatenkomponente erweitert, um den Anforderungen des ubiquitär präsenten Internet der Dinge gerecht zu werden. Neben den Herausforderungen im Backend nimmt die Visualisierung eine zentrale Rolle ein.

Im vorliegenden Beitrag liegt der Schwerpunkt auf der Anbindung der Echtzeitdaten vom Backend an das Frontend. Anhand eines Beispiels (Elektroladinfrastruktur) wird auf die dahinterliegende Technologie näher eingegangen.

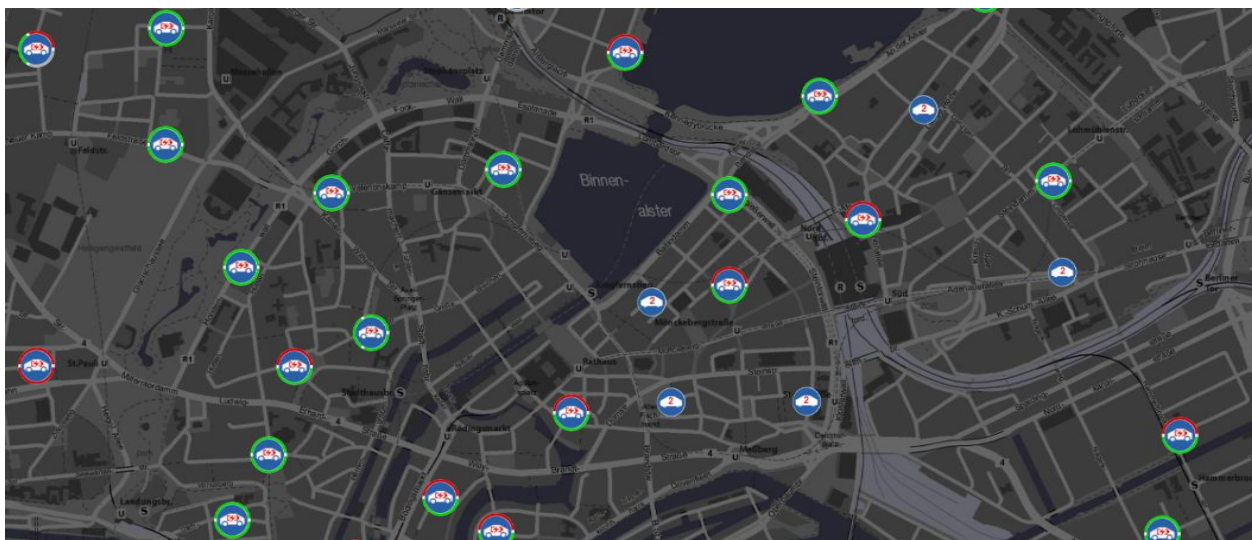


Abb. 4: Ausschnitt des Sensorportals. Hintergrundkarte: WMS – GEOBASISKARTEN HAMBURG (2019).

## 2 Zusammenfassung des Vortragsinhalts

In der Freien und Hansestadt Hamburg ist beim Landesbetrieb Geoinformation und Vermessung das Masterportal als Basis für zahlreiche Geoportale/WebGIS Anwendungen in der Urban Data Platform entstanden. Diese Open Source Anwendung wird kontinuierlich weiterentwickelt, um den aktuellen Ansprüchen gerecht zu werden. Bisher lassen sich



**Abb. 5:** Heatmap der belegten (links) und freien (rechts) Elektroladesäulen. Hintergrundkarte: WMS – GEOBASISKARTEN HAMBURG (2019).

Geo- sowie Metadaten über standardisierte, OGC-konforme Webdienste im Masterportal anzeigen. So werden zur Anzeige von Geodaten WMS und WFS verwendet. Als weiterer, innovativer Service wurde kürzlich, zur Visualisierung von Echtzeitdaten, die ebenfalls standardisierte OGC SensorThings-API (STA) integriert.

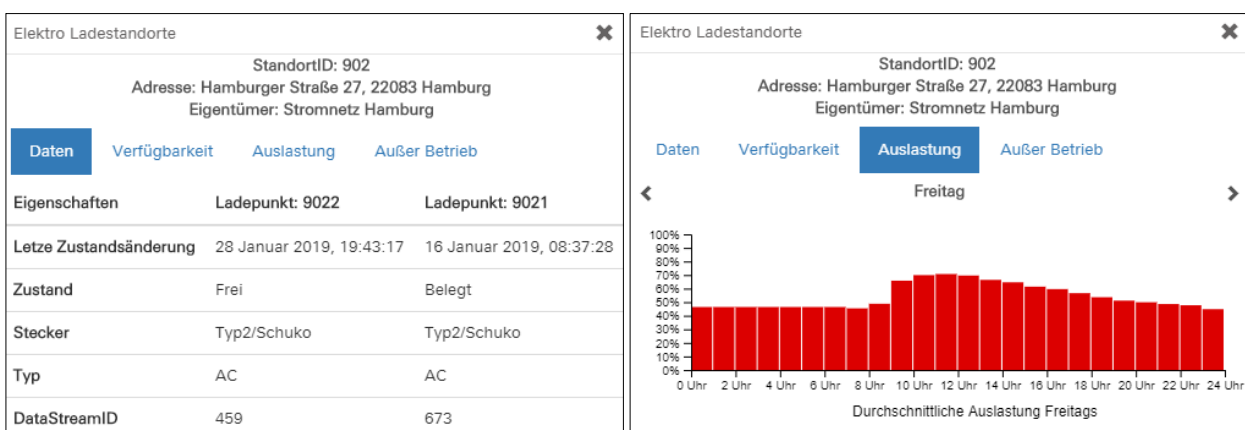
Zu diesem Zweck wird auf der Seite des Backends der FROST-Server des Fraunhofer Instituts eingesetzt (FRAUNHOFER INSTITUT 2019). Dieser ermöglicht das Persistieren und Bereitstellen von Sensordaten über eine REST-Schnittstelle. Darüber hinaus findet das Internetprotokoll MQTT Anwendung. Somit lassen sich serverseitig Daten eventbasiert, in Echtzeit an entsprechend angemeldete Clients übermitteln (one-to-many).

Auf der Frontendseite wurde im Masterportal ein Layertyp zur Verarbeitung der Sensordaten definiert. Zunächst werden hierbei die aktuellen Daten der Features zu einem Thema (Topic) vom FROST-Server entsprechend der STA angefordert und im Portal dargestellt. Im Anschluss wird eine MQTT-Verbindung über Websockets zum FROST-Server hergestellt und entsprechende Datenströme der ausgewählten Features abonniert. Beim Eintreffen neuer Datensätze wird der alte Datenbestand des entsprechenden Features überschrieben und automatisiert im Portal neu gerendert.

Als erster Datensatz wurden Sensordaten der Elektroladesäulen in Hamburg integriert. Diese repräsentieren die Belegung einer Ladesäule. Hierzu wurden eigene konfigurierbare Styles (Symbole) im Masterportal entwickelt, um die Zustände entsprechend darzustellen. **Fehler! Verweisquelle konnte nicht gefunden werden.** zeigt die im Masterportal dargestellten Elektroladesäulen. Dabei werden die Symbole, je nach Anzahl der Lademöglichkeiten und Zustand an einer Säule eventbasiert visualisiert.

Als weitere Darstellungsmethode wurde neben der Style-Definition eine Möglichkeit integriert, Features als Heatmaps anzuzeigen. (**Fehler! Verweisquelle konnte nicht gefunden werden.**)

Da die SensorThings-API auch das historisieren von Sensordaten ermöglicht, lassen sich Analysen zu den bisherigen Zuständen der Ladesäulen durchführen. Diese wurden ebenfalls in das Masterportal integriert und werden on-the-fly erzeugt. So können Auslastungen bzw. Verfügbarkeiten zu jedem Wochentag, anhand der durchschnittlichen Belegungszeit, ermittelt und grafisch dargestellt werden. Daneben lassen sich weitere Informationen in Tabellenform anzeigen.



**Abb. 1:** GetFeatureInfo zu einer Elektroladesäule mit zwei Steckplätzen (links). Durchschnittliche Tagesauslastung einer Elektroladesäule (rechts).



## Literatur

- EUROPÄISCHE KOMMISSION (2019), Digitaler Binnenmarkt: EU-Verhandlungsführer einigen sich auf neue Regeln für die gemeinsame Nutzung der Daten des öffentlichen Sektors- [http://europa.eu/rapid/press-release\\_IP-19-525\\_de.htm](http://europa.eu/rapid/press-release_IP-19-525_de.htm).
- FRAUNHOFER INSTITUT (2019), FROST-Server. A Server implementation of the OGC SensorThings-API <https://github.com/FraunhoferIOSB/FROST-Server>, abgerufen am 30.01.2019.
- URBAN DATA PLATFORM HAMBURG (2019), <http://www.urbandataplatform.hamburg/>, abgerufen am 30.01.2019.
- URBAN DATA PLATFORM (2019), <https://eu-smartcities.eu/initiatives/68/description>, abgerufen am: 24.01.2019.
- WMS – GEOBASISKARTEN HAMBURG (2019), [https://geodienste.hamburg.de/HH\\_WMS\\_Geobasisdaten\\_SG](https://geodienste.hamburg.de/HH_WMS_Geobasisdaten_SG), abgerufen am 30.01.2019.

# Next Generation GDI

Clemens PORTELE

interactive instruments GmbH · Trierer Straße 70-72 · 53115 Bonn

E-Mail: portele@interactive-instruments.de

## 1 Einführung ins Thema

Geodateninfrastrukturen wie die GDI-DE und INSPIRE setzen auf den Standards des Open Geospatial Consortiums (OGC) und des ISO/TC 211 „Geographic Information/Geomatics“ auf. Die Standards bilden ein solides Fundament für den produktunabhängigen Umgang mit geographischen Informationen und deren verteilte Bereitstellung / Nutzung über das Internet.

Die Verwendung dieser Daten und der zugehörigen Schnittstellen erfordert i.d.R. GIS-Werkzeuge sowie die Einarbeitung in die zugrundeliegenden Standards und Spezifikationen, was für Nicht-Experten aufgrund des reichhaltigen Materials und Funktionsumfangs oft eine erhebliche Hürde darstellt, wenn auch ein begrenzter Funktionsumfang ausreichen würde.

Darüber hinaus hat sich in den letzten Jahren das Web weiterentwickelt und mit ihm die Erwartung von Entwicklern, wie Daten im Web geteilt werden. Die über die letzten 20 Jahre entwickelten Standards unterstützen dabei die heute erwarteten Praktiken nur begrenzt.

Die beiden Standardisierungsorganisationen World Wide Web Consortium (W3C) und OGC haben vor diesem Hintergrund in den letzten Jahren zusammengearbeitet, um die derzeit empfohlenen Praktiken für die "web-freundliche" Veröffentlichung von Geodaten zu dokumentieren. Die Ergebnisse sind in den "Data on the Web Best Practices" (W3C) und "Spatial Data on the Web Best Practices" (W3C/OGC) dokumentiert, siehe insbesondere das Kapitel "Why are traditional Spatial Data Infrastructures not enough?".

Das Thema einer technischen Modernisierung der Geodateninfrastrukturen wird inzwischen auch in OGC, ISO/TC 211, INSPIRE und vielen anderen Stellen intensiv bearbeitet und diskutiert. Der vorliegende Beitrag fasst wichtige Aktivitäten und deren Stand bei dieser Weiterentwicklung zusammen.

## 2 Zusammenfassung des Vortragsinhalts

### 2.1 OGC Web Feature Service 3.0 und Folgeaktivitäten

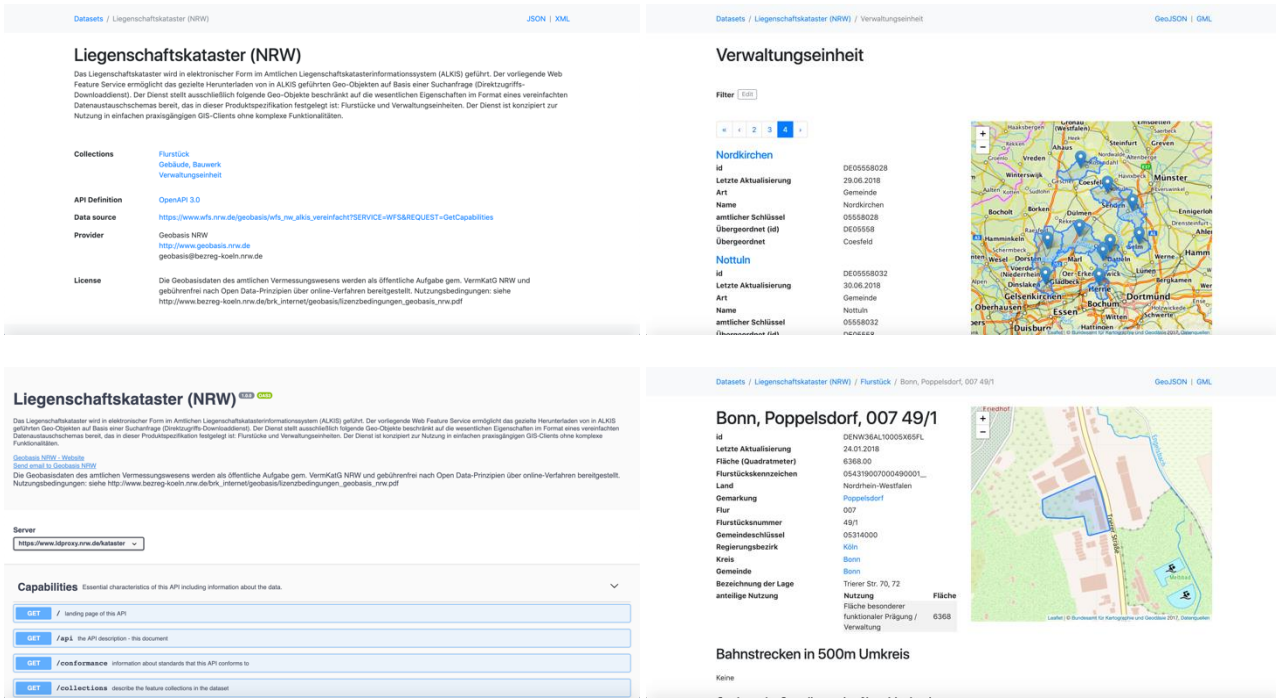
In 2017 haben OGC und ISO/TC 211 mit der Modernisierung des Web Feature Service Standards (WFS) auf Basis der Best Practices begonnen. Alle bisherigen Versionen von WFS (und WMS, CSW, usw.) verwenden eine Architektur, die vor 20 Jahren dem Stand der Technik entsprach, heute aber als veraltet angesehen wird. WFS 3.0 ist eine Überarbeitung des Standards, der

- eine zeitgemäße, mit der aktuellen Architektur des World Wide Web konsistente Architektur verwendet,
- auf modernen Technologien wie der OpenAPI-Spezifikation aufsetzt, die Schnittstellen für Entwickler verständlich machen können, ohne dass diese viele Spezifikationen verstehen müssen,
- nicht mehr zwingend die Definition von Schemata für die Geodaten fordert,
- die Verwendung von GeoJSON und HTML zur Kodierung empfiehlt (GML aber weiter unterstützt),
- einen Fokus auf den Software-Entwickler als „Kunden“ der Spezifikation setzt,
- Authentifizierung und Zugriffsmechanismen über die Verwendung von OpenAPI von Beginn an berücksichtigt und
- modulare Bausteine für den feingranularen Zugriff auf Geoobjekte für die Verwendung in Web-APIs definiert.

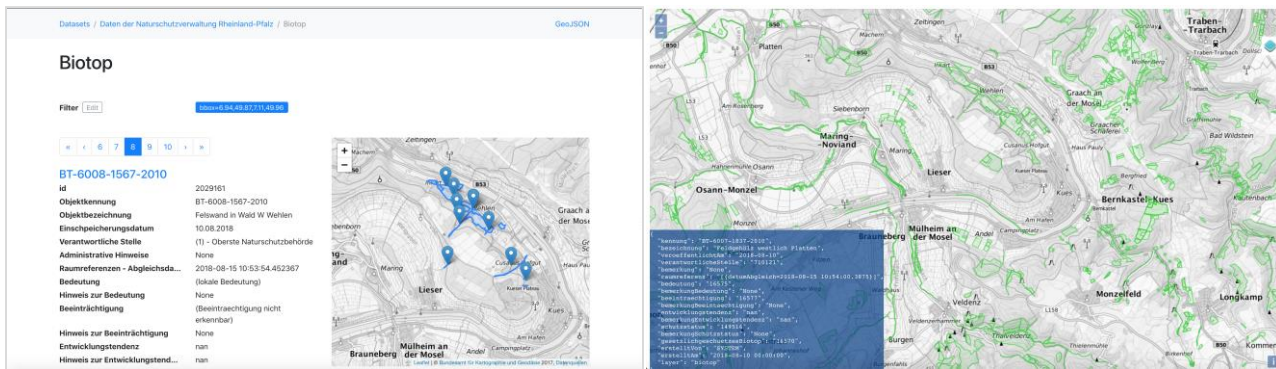
WFS 3.0 soll einfacher und moderner, aber dennoch eine Evolution der früheren Versionen und ihren Implementierungen sein.

In OGC wird auch über WFS 3.0 hinaus über die nächste Generation der OGC-Dienste und deren Architektur diskutiert. Dabei gibt es eine klare Tendenz, auf dem Ansatz von WFS 3.0 aufzusetzen und die dort skizzierte Architektur als Basis für die nächste Generation weiterer Ressourcen, wie serverseitig aufbereiteten Karten (WMS, WMTS), clientseitig aufbereiteten Karten (WMTS mit Vector Tiles), Datenprozessen (WPS), Katalogdiensten (CSW), zu verwenden.

Dabei geht es weniger um die Definition neuer Versionen der bestehenden OGC-Webdienste, als um die Definition von modularen Elementen für die Verwendung in modernen Web-APIs, die bisherige Trennung der Dienste verschwimmt und entfällt zunehmend. Dies ist gewollt und reflektiert einen Paradigmenwechsel, wie voraussichtlich zukünftig Geodaten über offene Standards bereitgestellt werden.



**Abb. 1:** HTML-Seiten von WFS-3.0-Servern mit Daten von Geobasis NRW. Von links oben, im Uhrzeigersinn: Startseite für den vereinfachten ALKIS-Datensatz, Seitenweiser Zugriff auf alle Verwaltungseinheiten, ein einzelnes Flurstück und der aus der OpenAPI-Definition generierte Web-Client für die API (Dokumentation der API und eine einfache Möglichkeit mit der API zu „spielen“).



**Abb. 2:** Experimentelle Erweiterung der WFS 3.0 API im Rahmen des OGC Vector Tiles Pilot. Links der normale Zugriff auf Biotope in Rheinland-Pfalz über die Standard-API, rechts der Zugriff aus einem einfachen OpenLayers-Client auf dieselben Geoobjekte, organisiert in Vector Tiles über dieselbe API mit Erweiterungen für gekachelte Daten.

## 2.2 Metadaten und das Finden von Daten

Wenn Daten im Web geteilt werden, dann geht es nicht nur um die Bereitstellung der Daten selbst, sondern auch darum, wie diese gefunden und untersucht werden können. Die derzeitigen GDI-Mechanismen hierfür (ISO 19115-Metadaten, Katalogdienste, Geoportale) leisten dies für einen eingeschränkten Expertenkreis, aber nicht darüber hinaus. Ein Effekt davon ist, dass die Geodaten aus GDIs und deren Angebot heute in der Regel nicht in den Suchmaschinen indiziert werden, oder wenn dann meist mit einem niedrigen Ranking. Und Suchmaschinen sind heutzutage in aller Regel das erste Werkzeug, das von Nicht-Experten zur Suche nach etwas verwendet wird, auch von Daten.

Mit der Empfehlung von WFS 3.0 für eine Bereitstellung aller Inhalte auch direkt über HTML wird ein erster Schritt getan, dies zu ändern. Die Datensätze selbst und auch die eigentlichen sind direkt zugänglich und werden von Suchmaschinen indiziert – ohne Umweg über separate Metadaten oder Geoportale.

Darüber hinaus gibt es mit schema.org ein von den Suchmaschinenbetreibern unterstütztes Vokabular, wie man Webseiten so annotieren kann, dass Software, auch die Crawler der Suchmaschinen, die Inhalte von Webseiten „verstehen“ kann. Unter anderem unterstützt schema.org auch Datensätze, dabei wird das von W3C entwickelte Data Catalog Vocabulary (DCAT) übernommen. Durch eine Unterstützung dieser Standards wird das Finden von Geodaten für alle erleichtert. Daher empfiehlt WFS 3.0 auch die Unterstützung von schema.org in den HTML-Seiten sowohl für Datensätze als auch für einzelne Geoobjekte.

Seit September 2018 bietet Google auch eine spezielle Suchmaschine für Datensätze an (Google Dataset Search). Eine Motivation für diese Aktivität war, dass es für Wissenschaftler und Journalisten immer schwieriger wird, für sie relevante Daten im Netz zu finden. Abb. 3 zeigt das Ergebnis einer Google-Dataset-Search-Suche.

The screenshot shows the Google Dataset Search interface. The search query is 'Liegenschaftskataster NRW'. The results page displays 54 results. The top result is 'Liegenschaftskataster (NRW)' from 'www.ldproxy.nrw.de'. Below this, there are several other results listed with their respective icons and URLs. The main result details are expanded, showing the following information:

- Liegenschaftskataster (NRW)**
- URL: [www.ldproxy.nrw.de](http://www.ldproxy.nrw.de)
- This dataset is cited by 8 scientific articles (see [Google Scholar](#)).
- Datensatz bereitgestellt von:** Geobasis NRW
- Lizenz:** Die Geobasisdaten des amtlichen Vermessungswesens werden als öffentliche Aufgabe gem. VermKatG NRW und gebührenfrei nach Open Data-Prinzipien über online-Verfahren bereitgestellt. Nutzungsbedingungen: siehe [http://www.bezreg-koeln.nrw.de/brk\\_internet/geobasis/lizenzbedingungen\\_geobasis\\_nrw.pdf](http://www.bezreg-koeln.nrw.de/brk_internet/geobasis/lizenzbedingungen_geobasis_nrw.pdf)
- Verfügbare Downloadformate der Anbieter:** GML-SF2\*, VERSION=3.2, GEO+JSON
- Abgedeckter Zeitraum:** 18.05.2018 - 26.01.2019
- Abgedecktes Gebiet:** Europa, Iserlohn
- Beschreibung:** Das Liegenschaftskataster wird in elektronischer Form im Amtlichen Liegenschaftskatasterinformationssystem (ALKIS) geführt. Der vorliegende Web Feature Service ermöglicht das gezielte Herunterladen von in ALKIS geführten Geo-Objekten auf Basis einer Suchanfrage (Direktzugriffs-Downloaddienst). Der Dienst stellt ausschließlich folgende Geo-Objekte beschränkt auf die wesentlichen Eigenschaften im Format eines vereinfachten Datenaustauschschemas bereit, das in dieser Produktspezifikation festgelegt ist: Flurstücke und Verwaltungseinheiten. Der Dienst ist konzipiert zur Nutzung in einfachen praxistauglichen GIS-Clients ohne komplexe Funktionalitäten.

**Abb. 3:** Google Dataset Search Ergebnis bei der Suche nach „Liegenschaftskataster NRW“. Die angezeigten Metadaten wurden von Google aus den schema.org-Annotationen der HTML-Startseite des WFS-3.0-Servers für den vereinfachten ALKIS-Datensatz aus Nordrhein-Westfalen extrahiert.

Gemäß der Architektur des Web sind in WFS 3.0 alle Geoobjekte über Links erreichbar, d.h. grundsätzlich können auch alle Geoobjekte von Suchmaschinen indiziert werden. Bei großen Datensätzen mit sehr vielen Geoobjekten findet dies derzeit allerdings noch nicht statt, sodass man Suchmaschinen derzeit vor allem für die Suche nach Datensätzen und (noch) nicht einzelnen Objekten aus indizierten Datensätzen verwenden kann.

### 2.3 Auswirkungen auf Geodateninfrastrukturen

Es ist absehbar, dass sich die Geo-Standards und die Geodateninfrastrukturen in den nächsten Jahren sukzessive in Richtung der in den (Spatial) Data on the Web Best Practices zusammengestellten Empfehlungen entwickeln werden. In Nordrhein-Westfalen befasst man sich zum Beispiel bereits seit 2017 aktiv mit dem Thema und erprobt die neuen Technologien. Auch in anderen Ländern (z.B. den Niederlanden, Großbritannien, Finnland, Australien, Kanada und USA) gibt es verschiedene nationale Aktivitäten, die sich mit diesen oder ähnlichen Themen befassen. In INSPIRE befasst sich seit 2018 eine Arbeitsgruppe mit dem Thema „Spatial Data on the Web“ mit den folgenden Schwerpunkten: Daten-Dienste-Koppelung, WFS 3.0, Auffinden von Daten über Suchmaschinen und Verlinkung von Geodaten / persistente Identifikatoren.

## Literatur

FARIAS LÓSCIO, B., BURLE, C., CALEGARI, N. (2017), Data on the Web Best Practices, W3C Recommendation. <https://www.w3.org/TR/dwbp/>.

- TANDY, J., VAN DEN BRINK, L., BARNAGHI, P. (2017), Spatial Data on the Web Best Practices, W3C Note / OGC Best Practice. <https://www.w3.org/TR/sdw-bp/>.
- PORTELE, C. & VRETANOS, P. (2018), OGC Web Feature Service 3.0: Part 1 – Core. Draft OGC Standard. [https://cdn.rawgit.com/opengeospatial/WFS\\_FES/3.0.0-draft.1/docs/17-069.html](https://cdn.rawgit.com/opengeospatial/WFS_FES/3.0.0-draft.1/docs/17-069.html).
- PORTELE, C., BÖHME, S., SCHLEGEL, B., CAFFIER, A., RATH, C. (2017), Geodaten für Jedermann. In: GDI-DE News, Dezember 2017, 13-14. Verfügbar unter [https://www.geoportal.de/SharedDocs/Newsletter/DE/Newsletter2017/12\\_2017\\_GDI\\_DE\\_NEWS.html](https://www.geoportal.de/SharedDocs/Newsletter/DE/Newsletter2017/12_2017_GDI_DE_NEWS.html).
- PORTELE, C. (2017), ELISE report: Spatial Data on the Web tools and guidance for data providers, EU Joint Research Centre. Verfügbar unter <https://inspire.ec.europa.eu/events/webinar-spatial-data-web-and-inspire>.
- BRICKLEY, D. & THE SCHEMA.ORG COMMUNITY, schema.org. <https://schema.org/>.

# Deep Space BIM - Digitaler Bauassistent der Zukunft

Daniel HOLWEG, Philipp WILLKOMM und Harald SCHULZ

M.O.S.S. Computer Grafik Systeme GmbH · Hohenbrunner Weg 13 · 84024 Taufkirchen

E-Mail: dholweg@moss.de; hschulz@moss.de

## 1 Einführung ins Thema

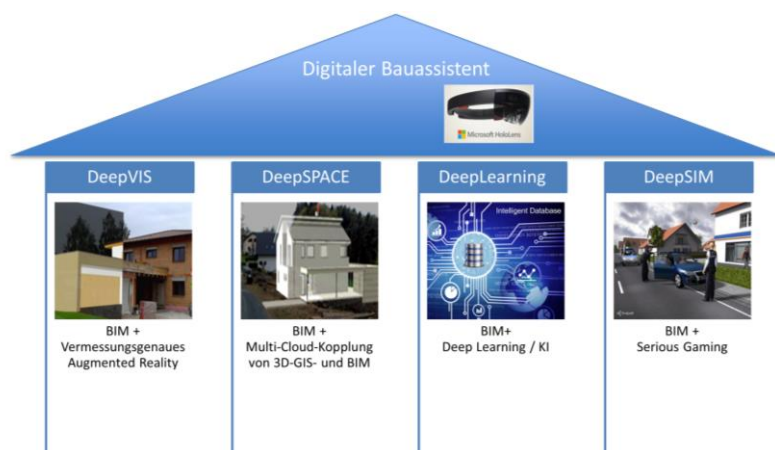
Großprojekte, egal ob in Deutschland oder international, leiden unter zunehmender Komplexität. Diese kann durch herkömmliche Managementmethoden kaum mehr bewältigt werden. Auf der anderen Seite nutzt die deutsche Bauwirtschaft Digitalisierungspotentiale längst nicht im vollen Umfang um diese Komplexität leichter zu bewältigen. Auch bei der Einführung von BIM (Building Information Modeling) als eine Kernkomponente der Digitalisierung der Bauwirtschaft hinkt Deutschland im internationalen Vergleich hinterher. Durch das Antragsvorhaben „DeepSpaceBIM 4.1“ soll in Form einer experimentellen Entwicklung ein „digitaler Bauassistent“ mit dem Gesamtziel geschaffen werden, die Komplexität von Großprojekten mittels moderner Informationstechnik wieder beherrschbarer zu machen.

Der „Digitale Bauassistent“ soll Verantwortlichen bei komplexen Entscheidungsszenarien helfen, die jeweils beste Entscheidung zu fällen. Diese Hilfestellung erfolgt durch mobil aus einer Multi-Cloud-Umgebung verfügbar gemachter, relevanter Statusinformation, vermessungsgenauen BIM-3D-Darstellungen und Entscheidungstools gestützt auf künstlicher Intelligenz. Der Assistent wird bei drei realen BIM-unterstützten Bauprojekten in unterschiedlichen Anwendungsszenarien exemplarisch evaluiert, um die Möglichkeiten der Handlungsverbesserung durch disruptive Informationstechnologie in der Bauwirtschaft erfahrbar zu machen.

## 2 Vortragsinhalt

### 2.1 Projektinhalt

Mit dem Projekt „DeepSpaceBIM 4.1“ sollen in Form einer experimentellen Entwicklung die technischen Umsetzungsmöglichkeiten von BIM durch „Deep Learning“, „Augmented Reality“ (AR) und „Multi-Cloud-Kopplung“ von BIM- und GIS-Daten (SPACE) entscheidend erweitert und auf einen neuen technischen Level gehoben sowie in Form eines „Digitalen Bauassistenten“ u.a. auch auf einer Microsoft HoloLens Applikation zusammengeführt werden.



**Abb. 1:** Projektschwerpunkte des Vorhabens DeepSpaceBIM 4.1 und deren Rollen im Projekt.

Der digitale Bauassistent wird im Rahmen des Projektes prototypisch umgesetzt und exemplarisch auf drei reale BIM-Bauvorhaben als Datenbasis abgestützt und im Rahmen wiederholter Reality Checks evaluiert. Die für die Reality Checks geplanten Bauvorhaben decken die Aspekte Hochbau, Infrastrukturbauvorhaben und Flächenvermarktung ab.

In den ersten Projektmonaten wurden die Kompatibilität und das Zusammenwirken der von den einzelnen Projektpartnern beigesteuerten Projektbausteine verifiziert. Mit Eintritt in die Umsetzungsphase ab April 2019 des Projektes erfolgt die ca. 2-jährige technische Erprobung anhand ausgewählter BIM Projekte.

## 2.2 Szenario DeepSpace

Im Kontext des Szenarios DeepSpace erfolgt die Kopplung von GIS und BIM anhand der Umsetzung eines Vermarktungsassistenten. Hierzu werden unterschiedliche Planungs- und Ausgestaltungsvarianten der Projektgebiete in einem Informationssystem auf Basis der WEGA Technologie von M.O.S.S. abgebildet. Konkretisierungen der Ausgestaltungsplanungen können auf Basis von BIM Modellen mittels novaFACTORY in die Umgebung übernommen werden und gemeinsam mit dem Geodatenbestand ausgewertet werden. Mittels WEGA 3D erfolgt die Darstellung in einer 3D Umgebung. Darüber hinaus wird die Augmented Reality Ausgabe z.B. mittels HoloLens unterstützt.



Abb. 2: Variantenmanagement mit WEGA VARIANT.

Die Umsetzung erfolgt auf Basis der M.O.S.S. Software WEGA und novaFACTORY, die bereits heute im Bereich der Planung (z.B. Windparkplanung, Sanierungsplanung) als zentrales Informationssystem und Datenquelle für CAD- und GIS-basierte Werkzeuge eingesetzt wird. BIM – verstanden als Digitalisierungsstrategie des Bauwesens – stellt dabei nicht nur die durchgängige Überführung in weitere räumliche (3D) oder zeitliche (4D) Dimensionen der Planungsdaten dar, sondern die Unterstützung der Prozesse durch ein Informationsmanagement und geeignete Informationsbereitstellung für die Bearbeitungswerkzeuge in den jeweiligen Prozessschritten.

## 2.3 Projektpartnerschaft und -rahmen



Für die Realisierung des Vorhabens haben sich unter der Gesamtkoordination von M.O.S.S. Computer Grafik Systeme GmbH die Partner Robotic Eyes GmbH, Graz, DMT GmbH & Co. KG, Essen, Steinmann Kauer Consulting GbR, Herrsching, Technische Universität Darmstadt, und Drees & Sommer Advanced Building Technologies GmbH, Stuttgart zusammengesetzt.

Das Projekt wird im Rahmen des Programms mFUND vom Bundesministerium für Verkehr und digitale Infrastruktur gefördert.

# Integrierte Web-basierte Visualisierung von 3D-Stadt- und -Landschaftsmodellen und Sensordaten auf Basis von OGC Standards

Volker COORS

Hochschule für Technik Stuttgart · Schellingstr. 24 · 70174 Stuttgart

E-Mail: volker.coors@hft-stuttgart.de

## 1 Einführung ins Thema

Seit der Gründung der ersten Städte in der Antike beschäftigen sich verschiedenste Fachdisziplinen von Stadtplanung, Bauingenieurwesen bis hin zu Soziologie mit Fragen des Zusammenlebens im urbanen Raum. Das Konzept einer „Smart City“ fügt mit der umfassenden Digitalisierung eine weitere Komponente hinzu – den digitalen Raum. Die britische Standardisierung definiert Smart Cities als „... the effective integration of physical, digital and human systems in the built environment to deliver a sustainable, prosperous and inclusive future for its citizens.“ (BSI 2014).

Sowohl der reale urbane Raum als auch die sozialen Beziehungen zwischen den Bürgern werden in der digitalen Welt abgebildet, beispielsweise durch die sog. „Sozialen Medien“. Hierzu werden Datenmodelle als Abstraktion der Realität entwickelt, die den für einen bestimmten Zweck relevanten Ausschnitt der realen Welt abbilden, dem „Universe of Discourse“. Für Smart City Anwendungen bietet es sich an, den urbanen Raum durch ein digitales 3D-Stadt- und -Landschaftsmodell zu repräsentieren. Mit dem OGC Standard CityGML lassen sich wesentliche Teile davon implementieren, beispielsweise 3D-Gebäudemodell, Straßenraum und Vegetation. Weitere Objektarten wie Ver- und Versorgungsnetze können durch fachspezifische Erweiterungen, sog. Application Domain Extensions hinzugefügt werden (BILJECKI et al. 2018).

Die Modellierung zeitlicher Veränderungen wird in CityGML jedoch nur sehr rudimentär unterstützt. Mit dem Dynamizer-Konzept (CHATURVEDI & KOLBE 2016) soll CityGML in der 3.0 ertüchtigt werden, auch dynamische Änderungen in hoher zeitlicher Auflösung abzubilden. In diesem Vortrag wird jedoch ein anderer Ansatz verfolgt, um bereits mit der aktuellen Version von CityGML mit Hilfe des OGC Standards Sensor Things (LIANG et al. 2016) Sensordaten in ein 3D-Stadt- und -Landschaftsmodell zu integrieren. Da die Daten in der Regel von verschiedenen Fachdisziplinen erhoben und bereitgestellt werden, sind standardisierte Schnittstellen sinnvoll, um den Aufwand der Vernetzung der Datenquellen zu reduzieren.

## 2 Zusammenfassung des Vortragsinhalts

### 2.1 Sensor Things

Das Datenmodell von SensorThings beschreibt ein Objekt der realen Welt als Ding / Thing, dem Eigenschaften und ein Raumbezug zugewiesen werden. Einem Ding können beliebig viele Datenströme zugeordnet werden, wobei jeder Datenstrom eine von einem Sensor gemessene Zeitreihe genau einer Eigenschaft des Dings ist. Die Vereinigung all dieser Informationen wird im Datenmodell als „Beobachtung“ eines „Features of Interest“ bezeichnet. Dieses Datenmodell erlaubt es, den Standort eines Dings ständig zu aktualisieren, d.h. es können zusätzlich zu Änderungen der Eigenschaften des Dings auch Bewegungsdaten abgebildet werden. Dies erlaubt eine zeit- und orts aufgelöste Beobachtung von mobilen Dingen. Bei Sensor Things wird nicht nur ein Datenmodell standardisiert, sondern auch eine Programmierschnittstelle (API), um auf die Daten zuzugreifen. Die Kommunikation folgt über eine REST Schnittstelle (REpresentational State Transfer) und lässt sich damit gut in eine moderne Web-basierte IT-Infrastruktur integrieren.

Im Sensor Things Datenmodell lassen sich einem Ding beliebige Eigenschaften zuordnen. Damit lässt sich ein Ding durch eine externe Referenz auch mit einem Objekt im 3D-Stadt- und -Landschaftsmodell verbinden. In CityGML hat jedes Objekt einen eindeutigen Identifier. Darüber wird ein Ding im SensorThings Datenmodell mit dem Objekt verbunden, das dieselbe Entität im 3D-Stadt- und -Landschaftsmodell repräsentiert (SANTHANAVANICH et al. 2018).

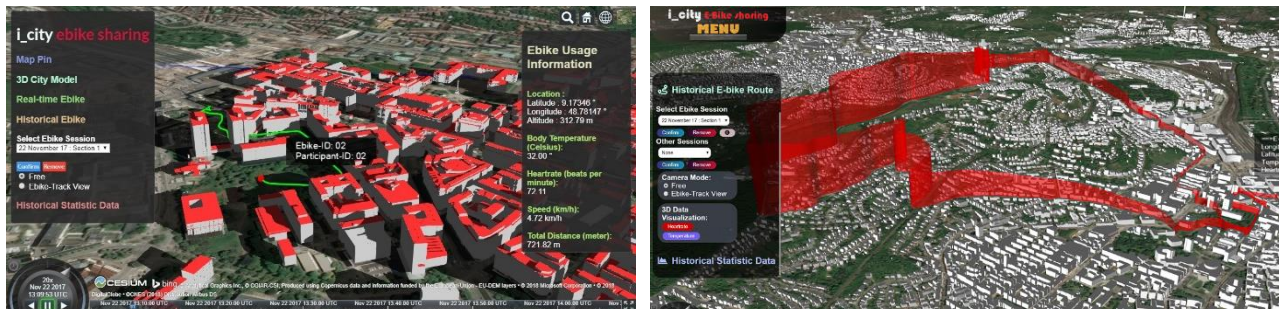
Um beispielsweise den laufenden Stromertrag einer PV-Anlage zu monitoren, kann dieses Modell wie folgt angewandt werden: in SensorThings wird die PV-Anlage als Ding modelliert. Ein Sensor misst den Stromertrag in einer zeitlichen Auflösung von etwa einer Minute. Diese Zeitreihe wird als Datenstrom über eine REST Schnittstelle bereitgestellt. Im 3D-Stadt- und -Landschaftsmodell wird das Gebäude, auf dessen Dach die PV-Anlage installiert ist, als CityGML Building modelliert. In LoD 2 ist auch die Dachfläche selbst als RoofSurface vorhanden. Die PV-Anlage selbst kann als Building Installation modelliert sein. Je nachdem, ob die PV-Anlage im Gebäudemodell vorhanden ist, verlinkt das Ding in SensorThings auf die BuildingInstallation, die die PV-Anlage repräsentiert, oder auf das RoofSurface, das die Dachfläche repräsentiert, auf dem die PV-Anlage installiert ist.



## 2.2 Anwendungsbeispiel eBike Sharing

Im Forschungsprojekt iCity wird u.a. die Integration von Sensordaten und 3D-Stadt- und -Landschaftsmodell in einem prototypischen System für ein free floating eBike Sharing System erprobt. Dazu werden die Standorte der Fahrräder, die nicht stationsgebunden sind, über die SensorThings API getrackt. Jedes Fahrrad verfügt über eine onboard Unit, über die permanent Datenströme zu Position und Geschwindigkeit des Rades, Ladezustand und auch Trittkraft des Fahrers bereitgestellt werden. Diese Daten werden in einer 3D-Webvisualisierung derzeit mit einem 3D-Gebäudemodell verknüpft (siehe **Abb. 2**).

Zu Forschungszwecken werden zusätzlich Probanden mit einer Garmin Fenix 5x Smartwatch und einem Movisens Brustgurt ausgestattet, die Vitaldaten wie Pulsfrequenz, Hauttemperatur und Herzratenvariabilität messen. Verschiedene Sensordaten können interaktiv korreliert werden, um z.B. Ladezustand des Akkus, Pulsfrequenz und Geländeprofil gegenüber zu stellen.



**Abb. 2:** Online Tracking eines eBikes über die SensorThings API (links), Pulsfrequenz einer Testperson entlang der gefahrenen Teststrecke (rechts).

Zukünftig sollen auch Vegetation und insbesondere der Straßenraum in das 3D-Modell integriert werden. Hierzu wurde bereits ein Workflow entwickelt, um ein 3D-DLM zu generieren (ROTHENGASS 2018). Der Workflow wurde im Stadtteil Stuttgart Stöckach bereits erfolgreich für einen anderen Anwendungszweck evaluiert.

## Danksagung

**Das diesem Bericht zugrundeliegende Vorhaben iCity wird mit Mitteln des Bundesministeriums für Bildung, und Forschung unter dem Förderkennzeichen 03FH9I01IA gefördert.** Das Teilprojekt 5 - Nachhaltige Mobilität erfolgt in Zusammenarbeit mit der Daimler TSS GmbH.

## Literatur

- BILJECKI, F., KUMAR, K., & NAGEL, C. (2018), CityGML Application Domain Extension (ADE): overview of developments, Open Geospatial Data, Software and Standards 2018 3:13, DOI 10.1186/s40965-018-0055-6.
- BSI (2014), Smart Cities Vocabulary, PAS 180:2014, The British Standards Institution, ISBN 978 0 580 81874, 2014
- CHATURVEDI, K., & KOLBE, T. (2016), Integrating Dynamic Data and Sensors with Semantic 3D City Models in the Context of Smart Cities. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume IV-2/W1, 2016 11th 3D Geoinfo Conference, 20–21 October 2016, Athens, Greece.
- LIANG S.H.L., HUANG, C.-Y., & KHALAFBEIGI, T. (2016), OGC SensorThings API Part I: Sensing, <http://docs.opengeo-spatial.org/is/15-078r6/15-078r6.html>, letzter Zugriff 1/2019.
- ROTHENGASS, L. (2018), Generierung und Evaluierung dreidimensionaler Landschaftsmodelle für eine CFD-Windsimulation, Master-Arbeit Vermessung, HFT Stuttgart, WS 2017/18.
- SANTHANAVANICH, T., SCHNEIDER, S., RODRIGUES, P., & COORS, V. (2018), Integration and Visualisation of Heterogeneous Sensor Data and Geospatial Information. In: ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. IV-4/W7, S. 115–122. DOI: 10.5194/isprs-annals-IV-4-W7-115-2018.

# Smart Villages - Vernetzung von Geoinformationen und umwelt-relevanten Sensordaten

Elke BLESSING<sup>1</sup>, Stephan BLUDOVSKY<sup>2</sup>, Gerald GRAF<sup>3</sup>, Monique VÖGELE<sup>4</sup>

Landesamt für Geoinformation und Landentwicklung Baden-Württemberg,  
Büchsenstraße 54, 70174 Stuttgart

<sup>1</sup>E-Mail: elke.blessing@lgl.bwl.de, <sup>2</sup>E-Mail: stephan.bludovsky@lgl.bwl.de,

<sup>3</sup>E-Mail: gerald.graf@lgl.bwl.de, <sup>4</sup>monique.voegele@lgl.bwl.de

## 1 Einführung

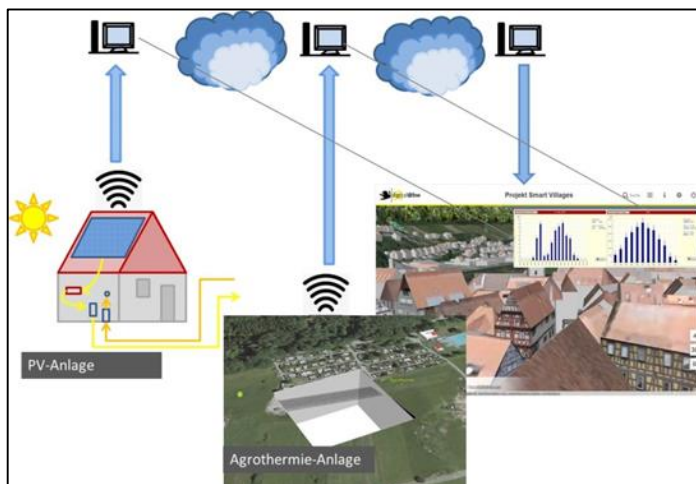
Das Landesamt für Geoinformation und Landentwicklung Baden-Württemberg (LGL) stellt mit dem Digitalen Geländemodell (DGM) und dem Digitalen Gebäudemodell (LoD2) dreidimensionale Geobasisdaten in hoher Qualität, Aktualität und landesweiter Verfügbarkeit bereit. Über die Funktion der Geobasisdaten hinausgehende besondere Wertschöpfungseffekte können erzielt werden, wenn standardisierte 3D-Modelle mit weiteren Geobasisdaten und Fachdaten verknüpft werden. Innovative 3D-Nutzungsszenarien findet man zwischenzeitlich z. B. in den Bereichen energetisch orientierter Projekte sowie in 3D-Stadtmodellen. In Architektur und Stadtplanung sind 3D-Modelle seit Langem eine gängige Form der Präsentation von Planungsmaßnahmen, die sich aber meist auf große Städte beschränken. Es sind nicht zuletzt die hohen Kosten der 3D-Modellierung im Vergleich zum planerischen Mehrwert, die für kleine und mittlere Kommunen im ländlichen Raum eine Hürde darstellen und sie daran hindern, die Vorteile dieses Instruments zu nutzen (GRAF, G. et al. 2016).

Im Rahmen der Digitalisierungsoffensive des Landes Baden-Württemberg digital@bw arbeitet das Landesamt für Geoinformation und Landentwicklung in Zusammenarbeit mit der Hochschule für Technik Stuttgart im Projekt „Smart Villages – attraktive Orte im ländlichen Raum“ zusammen. Für das Projekt wird eine 3D-Webplattform, die das 3D-Landschafts- bzw. 3D-Gebäudemodell enthält, bereitgestellt und Lösungsansätze entwickelt, welche die Integration von Echtzeit-Sensorfachdaten ermöglichen und somit einen klaren Mehrwert schaffen. Anwender dieser 3D-Webplattformen sollen Bürger und kommunale Entscheidungsträger kleiner und mittlerer Gemeinden sein. So ist in einem Teilprojekt in Kooperation mit der Gemeinde Wüstenrot, einer Gemeinde auf dem Weg zur Plusenergiegemeinde, vorgesehen, Maßnahmen der innovativen und nachhaltigen Energiegewinnung über die 3D-Webanwendung auch für Laien zugänglich zu machen.

## 2 Zusammenfassung des Vortragsinhalts

### 2.1 Implementierung basierend auf OGC-Standards

Eine Herausforderung bei der Web-Visualisierung von 3D-Geodaten und der Integration von Fachdaten ergibt sich aus der Vielzahl verschiedener Sensortypen mit ihren unterschiedlichsten Datenformaten. Die Konzeption des Teilprojekts sieht vor, dass das Hosting der Geodaten getrennt vom Hosting der Sensordaten erfolgt. Aufgabe des LGLs ist die Bereitstellung nicht nur der bloßen DGM- und LoD2-Daten, sondern einer Web-Plattform, welche diese Daten bereits modelliert für die Nutzer zur Verfügung stellt, sowie der Schnittstelle zur Integration von Fachdaten anderer Plattformen. Den Anwendern (Unternehmen, Dienstleistern, Verwaltungen) eröffnet dies die Möglichkeit, ihre Geschäftsmodelle mithilfe der 3D-Webanwendung als eine Dienstleistung des LGLs umzusetzen.



**Abb. 1:**

Verknüpfung von Fachdaten mit dem 3D-Basismodell. Eigene Darstellung mit Diagrammen der Solar Community (<https://solar-community.eu>).

Um einen breiten Nutzerkreis zu erreichen basiert die Implementierung auf den OGC-Standards CityGML und Sensor Things Services. SensorThings, seit Februar 2016 von der OGC bestätigt, bildet hierbei die Schnittstelle zu den Sensoren und Datenströmen, während CityGML – als urbanes Informationsmodell – für die Vernetzung von Sensor- und Geodaten sorgt (COORS & SCHNEIDER 2018). Zur performanten Visualisierung von 3D-Daten im Internet auf Basis des CESIUM-Globus werden die Softwareprodukte der Firmen virtualcitySYSTEMS und M.O.S.S. eingesetzt, die als Webanwendung die Schnittstelle zum Nutzer bilden (vgl. Abb. 1).

## 2.2 Erste Umsetzungen in der Gemeinde Wüstenrot

Im Fokus des Teilprojekts stehen derzeit eine agrothermische Anlage ebenso wie die Photovoltaikanlagen auf den öffentlichen Gebäuden der Gemeinde Wüstenrot. So werden im Rahmen einer Masterarbeit der aktuelle Energiefluss und die Energieflusserträge nach vorgegeben Zeitabschnitten über einen Sensordatenaustausch im 3D-Modell dargestellt. Der Mehrwert liegt darin, dass die aktuellen Energieflüsse bestehender Anlagen jederzeit abgerufen werden können. Damit werden verlässlichere Simulationen über zu erwartende Erträge künftiger weiterer Anlagen und somit die Festlegung optimaler Standorte möglich. Denn diese erfolgen auf der Basis von Messungen in der unmittelbaren Umgebung. Damit stellt die Anwendung ein wertvolles Planungswerkzeug für die Kommune dar.

Weiterhin befindet sich ein Energieerlebnispfad im Aufbau, der die „Wüstenroter Energiewende“ über voraussichtlich zehn beschilderte Stationen vor Ort, in Augmented Reality Technologie und als 3D-Web-Präsentation, darstellt (vgl. Abb. 2). Auch hier werden die 3D-Geodaten mit den Sensordaten der verschiedenen Energiegewinnungsanlagen, beispielsweise mit denen eines Windrads, verknüpft.



**Abb. 2:** Fachspezifische Daten wie umweltrelevante Sensordaten, z.B. der Stromertrag von Windkraftanlagen, werden zum Monitoring in die Plattform integriert und können für weitere Simulationen genutzt werden. Eigene Darstellung mit Diagramm des Windenergie-Branchenportals (<https://www.windbranche.de/wind/windstrom/windenergie-deutschland>).

## 3 Danksagung

Das in diesem Abstract zu Grunde liegende Vorhaben „Smart Villages – attraktive Orte im ländlichen Raum“ wird mit Mitteln der Landesregierung Baden-Württemberg im Rahmen der Initiative digital@bw gefördert und erfolgt in Zusammenarbeit mit der Hochschule für Technik Stuttgart. In diesem Zusammenhang sei auch den folgenden Projektbeteiligten, -mitarbeitern und Studenten gedankt. Prof. Dr. Volker Coors, Thunyathep Santhanavanich, Patrick Würstle, Sabo Kwada Sini. Die Verantwortung für den Inhalt des Abstracts liegt bei den Autoren.

## Literatur

- COORS, V. & SCHNEIDER, S. (2018), Smart Cities benötigen Sensoren und smarte Geodaten. In: Zeitschrift für das Öffentliche Vermessungswesen des Landes Sachsen Anhalt, 24. Jahrgang, 1/2018, 7-16, Magdeburg.
- GRAF, G., HOLUBA, K.-H., BLESSING, E., (2016), 3D-Webvisualisierung von Planungsdaten. In: Geoinformationssysteme, Beiträge zur 3. Münchner GI-Runde, 187-195, Berlin.
- <http://www.envisage-wuestenrot.de/fotos-termine-dateien/lehrpfad>

# GIS-basierte Straßenkontrolle der Freien und Hansestadt Hamburg

Jens Nienaber

Dataport · Hamburg · Billstraße 81 · 20539 Hamburg

E-Mail: jens.nienaber@dataport.de

## 1 Einführung in die Thematik

Die Freie und Hansestadt Hamburg (FHH) treibt die Digitalisierung und Vereinheitlichung ihrer Prozesse und Datenstrukturen im Zuge des zentralen urbanen Datenmanagements „Urban Data Hub“ und der Senatsstrategie „Digitale Stadt“ konkret voran.

Ziel des Projektes „GIS-basierte Straßenkontrolle“ war die digitale, revisionssichere Aufnahme der durch die Tiefbau Unterhaltung aufgenommenen Mängel. Im Laufe der Pilotphase des Projektes sind weitere Anforderungen wie die Erzeugung auftragsbegründender Unterlagen hinzugekommen. Durch die Einführung der Software verändern sich die Arbeitsbedingungen von mehr als 150 Mitarbeitern.

Das Projekt „GIS-basierte Straßenkontrolle“ hatte seinen operativen Kick Off am 12.09.2017 und ist eines der ersten Projekte, das die neuen Anforderungen des Urban Data Hub und der Senatsstrategie implementiert und umsetzt. Der Abschluss des Projektes ist für den 07.02.2019 geplant. Die öffentliche Ausschreibung konnte die Firma AED-SYNERGIS GmbH mit Ihrem Produktpaket Pro Office und Web Office für sich entscheiden.

## 2 Rahmenbedingungen des Projektes GIS-basierte Straßenkontrolle

In der FHH obliegt die Unterhaltung des Tiefbauvermögens den sieben Bezirken. Diese sind größtenteils eigenverantwortlich organisiert und in ihren Arbeitsprozessen, der technischen Ausstattung sowie den verfügbaren Personalressourcen heterogen. Aufgrund dieser Heterogenität wurde die zentrale Stelle der Bezirke für IT Angelegenheiten (N/ITB) mit der Umsetzung einer einheitlichen GIS-basierten Straßenkontrolle beauftragt. Aufgrund des zu erwartenden hohen Management Aufwandes wurde daraufhin ein Projektleiter bei Dataport akquiriert.

In der öffentlichen Verwaltung stehen neben der fachlich besten Software Aspekte wie Kosteneinsparung, Revisionsicherheit und gesicherte langfristige Service Level Agreements im Vordergrund. Daher verfolgt die FHH in der Strategie „Digitale Stadt“ den Ansatz, bei Verfügbarkeit Standard Software einzusetzen und diese gegebenenfalls an die Bedürfnisse der Fachlichkeit anzupassen.

## 3 Anforderungen an die Lösung „GIS-basierte Straßenkontrolle“

Die Anforderungen an die einzuführende Lösung setzen sich zusammen aus verwaltungsrechtlichen und organisatorischen Rahmenbedingungen in der FHH, den Vorgaben aus der aktuellen Senatsstrategie, technischen Vorschriften des Urban Data Hubs sowie den fachlichen Anforderungen aus den sieben, teils unterschiedlich agierenden Tiefbau Unterhaltungsabteilungen. Hinzu kommt, dass ein Großteil der Aufgaben eines Wegewartes im Außendienst wahrgenommen wird, so dass das Verfahren aus einer mobilen und einer Innendienst-Komponente besteht. Die folgenden Anforderungen mussten daher auf einem Tablet sowie einem Desktopsystem umgesetzt werden.

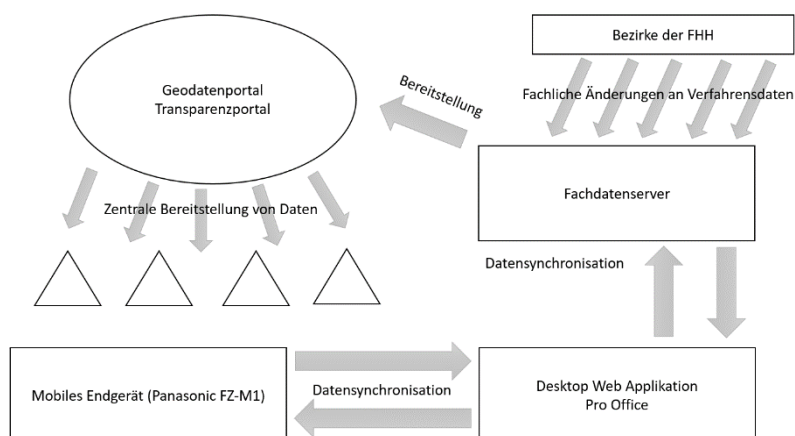


Abb. 1: Datenströme des Verfahrens.

### 3.1 Organisatorische Anforderungen an die Lösung „GIS-basierte Straßenkontrolle“

Die erfolgreiche Umsetzung eines komplexen Projektes wie der „GIS-basierten Straßenkontrolle“ erfordert die Mitarbeit und das Engagement aller beteiligten Akteure. Die folgende Tabelle (**Tabelle 1**) stellt die Akteure und ihre Rolle im Projekt kurz vor:

**Tabelle 1:** Beteiligte Institutionen und ihre Rolle im Verfahren.

Institution	Rolle
N/ITB	Mit Projektumsetzung beauftragt
Hamburger Bezirke	Fachliche Ansprechpartner; Nutzer der eingeführten Software
Landesbetrieb Geodaten und Vermessung (LGV)	Verantwortlich für den zentralen Fachdatenserver
Dataport	Projektleitung; Fachliches Verfahrensmanagement; Server Hosting
Personalrat	Mitbestimmungspflichtig bei Dienstanweisungen / Dienstvereinbarungen

Kommt es an einer Stelle zu Verzögerungen ist der gesamte Projektzeitplan gefährdet.

Aufgrund der Vielzahl von Playern mussten Termine und Aufgaben frühzeitig geplant, vergeben und umgesetzt werden. Kritische neue Aufgabenpakete können während des Projektes auftreten und müssen dann zeitnah in den Projektplan einfließen. Beispielhaft wurde gleich zu Beginn des Projektes erkannt, dass parallel zur Einführung der „GIS-basierten Straßenkontrolle“ auch ein Migrationsprojekt durchgeführt werden musste.

Ein solches Projekt erfordert daher konstantes Monitoring und Controlling (Anforderungen, Aufgabenpakete, etc.) und soweit möglich zeitliche Puffer.

### 3.2 Fachliche Anforderungen an die Lösung „GIS-basierte Straßenkontrolle“

Im Zuge der Einführung „GIS-basierte Straßenkontrolle“ war die Analyse der fachlichen Prozesse und Workflows sowie die Übersetzung dieser in ein digitales Verfahren zentraler Bestandteil.

Der Anforderungskatalog erstreckt sich dabei von einer leichten Usability über mandanten- und rollenspezifische Ansichten und Rechte. Hinzu kommen ein fachgerechter Objektkatalog, Revisionsicherheit, fachliche Berichterstellung, anwendungsgesteuerte Excel-Exporte, ein umfassendes Such- und Abfragesystem, die gesicherte Dokumentation von durchgeführten Begehungen sowie die Möglichkeit, auftragsbegründende Unterlagen für die externe Vergabe von Arbeitsaufträgen zu generieren.

Die fachlichen Anforderungen sind in bezirks- und fachbehördenübergreifenden Workshops definiert, dokumentiert und in Aufgabenpaketen zusammengefasst worden. Unterstützt wurde dies durch die Anforderungsanalyse einer pilotierenden Dienststelle. Daraus resultierte eine präzisere Umsetzung der fachlichen Anforderungen.

### 3.3 Technische Anforderungen an die Lösung „GIS-basierte Straßenkontrolle“

Der Aufbau des zentralen Auskunftssystems „Urban Data Hub“ in der FHH geht mit klar definierten Anforderungen einher. Dies beinhaltet unter anderem ein einheitliches Koordinatensystem, abgestimmte und klar definierte Geodatenstrukturen und Datenmodelle, vollständige Metadatenangaben sowie die ausschließliche Verwendung von Open Geospatial Consortium Diensten (OGC). Die Daten dieses Verfahren (vgl. Abb. 1) erfüllen diese Kriterien hatten aber u. A. neue Arbeitspakete zur Folge:

- Hintergrunddaten wie Luftbilder mussten über WMTS Dienste bezogen, in Mosaic Dataset überführt und dann als Tilepackage auf die mobilen Geräte gespielt werden.
- Straßenabschnittsegmente wurden von den Bezirken in Esri Formaten bereitgestellt über Scripte in einen WFS umgewandelt und auf den Mobilgeräten verfügbar gemacht.
- Mit dem Mobilgerät erfasste Mängelpunkte mussten mittels WFS-T Dienst in den Fachdatenserver des LGV gespielt werden.

#### Literatur

<https://www.hamburg.de/pressearchiv-fhh/11362996/2018-07-09-bsw-urban-data-hub/>

<https://www.hamburg.de/digitalisierung/>

<https://www.dataport.de/Seiten/L%C3%B6sungen/IT-Infrastruktur/Rechenzentrum/%C3%9Cb%C3%9Cbersicht.aspx>