

Gefördert durch:



Bundesministerium  
für Wirtschaft  
und Klimaschutz

aufgrund eines Beschlusses  
des Deutschen Bundestages



Deutsches Zentrum  
für Luft- und Raumfahrt

Institut für  
Vernetzte Energiesysteme

# Schlussbericht

des Einzelvorhabens

## SciGRID\_gas - Offenes Referenzmodell Europäischer Gastransportnetze für wissenschaftliche Untersuchungen zur Sektorenkopplung



**Förderkennzeichen:** 03ET4063

**Laufzeit:** 01.01.2018 bis 31.07.2021

**Berichtszeitraum:** 01.01.2018 bis 31.07.2021

**Antragssteller:**

Deutsches Zentrum für Luft- und Raumfahrt e.V.

Institut für Vernetzte Energiesysteme

Carl-von-Ossietzky-Str. 15

26129 Oldenburg

Projektleiterin: Dr. Wided Medjroubi

Telefon: +49 (0)441 99906 114

Fax: +49 (0)441 99906 109

Email: [wided.medjroubi@dlr.de](mailto:wided.medjroubi@dlr.de)

Oldenburg, den 23.11.2021

Das in diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Wirtschaft und Energie unter dem Förderkennzeichen 03ET4063 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autoren.

**Liste der Autoren:** Dr. Wided Medjroubi, Dr. Adam Pluta, Dr. Jan Diettrich, Jan Dasenbrock, Hendrik-Pieter Tetens, Javier Sandoval

## Inhalt

|   |   |    |
|---|---|----|
| 1 | Kurzdarstellung.....                                  | 3  |
| 2 | Eingehende Darstellung der Ergebnisse.....            | 6  |
| 3 | Verwertung der Ergebnisse.....                        | 34 |
| 4 | Fortschritte auf dem Gebiet bei anderen Stellen ..... | 34 |
| 5 | Veröffentlichungen .....                              | 34 |
|   | Literaturverzeichnis.....                             | 37 |

# 1 Kurzdarstellung

Die Modellierung von Energieinfrastrukturen spielt eine wichtige Rolle bei der Planung, Betrieb und Bewertung zukünftiger Energiesysteme. Die Modellierung erfordert den Zugriff auf verschiedene Datensätze auf unterschiedlichen Skalen und Aggregationsebenen. Solche Datensätze betreffen nicht nur die Stromerzeugung und -verbrauch sondern auch die Transport- und Verteilnetze für Strom und Gas (und Wärme) die für die Abbildung von Energiesysteme von hoher Bedeutung sind. Die Verfügbarkeit belastbare Infrastrukturdaten ist angesichts der zunehmenden Komplexität des Energiesystems und der Kopplung der verschiedenen Energiesektoren umso relevanter für die Modellierer sowie alle Akteure im Energiebereich. Angesichts der aktuellen deutschen und europäischen Debatte über die Rolle des Gastransportnetzes bei der Integration erneuerbarer Energiequellen sowie seiner Rolle als Transportsystem für (grünen) Wasserstoff besteht ein großer Bedarf nach Datensätzen zur Gastransportinfrastruktur in Europa.

Doch trotz ihrer großen Bedeutung sind Gastransportdatensätze für Europa nicht öffentlich zugänglich. Die Nichtverfügbarkeit der Datensätze führt zu negativen Aspekten, von denen einige im Folgenden aufgeführt werden.

Modellierungsarbeiten die eine detaillierte Abbildung des Gastransportnetzes benötigen, sind schwer umzusetzen da es an detaillierten Datensätze mangelt. Darüber hinaus ist eine Validierung von Simulationsergebnissen aus Gasnetz- sowie Energiesystemsimulationen sehr begrenzt bis unmöglich. Innovative Ansätze und Verfahren auf diesem Forschungsgebiet werden gehemmt, da die Modellierer begrenzten Zugang zu Netzdaten mit einer ausreichenden Auflösung sowie Detaillierungsgrad haben. Andere Akteure haben infolgedessen wenig Vertrauen in die Modellierungsergebnisse aufgrund der mangelnden Transparenz, was eine Einschätzung der Datenqualität sowie der Annahmen und Vereinfachungen bei deren Beschaffung und Nutzung erschwert. Diese Aspekte sind insbesondere vor dem Hintergrund der nationalen und internationalen Diskussionen um eine Umwidmung der Gastransportnetze für den Wasserstofftransport (technische und wirtschaftliche Umsetzung) und der damit verbundenen Kosten von sehr großer Bedeutung. In diesem Zusammenhang wurde im Rahmen des SciGRID\_gas Projekts ein Open Source und Open Data Modell des europäischen Gastransportnetz entwickelt.

## Aufgabenstellung

Das Hauptziel des SciGRID\_gas Projekts ist die Entwicklung eines Open Source und Open Data Modells des europäischen Gastransportnetzes, das für die Modellierung des Gastransportnetzes sowie für Untersuchungen zur Sektorenkopplung genutzt werden kann. Dies wurde erreicht durch (1) **Klassifizierung und Charakterisierung von Gastransportnetzmodellen**, (2) die **Datengenerierung und heuristische Verfahren**, (3) **Entwicklung eines Open-Source-Modells** des europäischen Gastransportnetzes und dessen Veröffentlichung sowie der dazugehörigen Datensätze unter Open Source und Open Data Richtlinien inklusive einer kompletten Dokumentation, (5) **Vergleich und Validierung** des Datensatzes und (4) durch eine **Schnittstellenbildung zu Softwaretools von Open-Source-Anbietern**.

Die Erfahrungen bei der Modellentwicklung bezüglich der Genese und Bearbeitung heterogener und großer georeferenzierter Datensätze sowie der Veröffentlichung und Verbreitung der Projektergebnissen sind sehr wertvoll und stellen einen Mehrwert sowohl für das DLR Institut für

Vernetzte Energiesysteme (DLR-VE) als auch für die Modellierer-Community und die Gas- und Energiesystemmodellierung im Allgemeinen dar.

### **1.1 Stand von Wissenschaft und Technik**

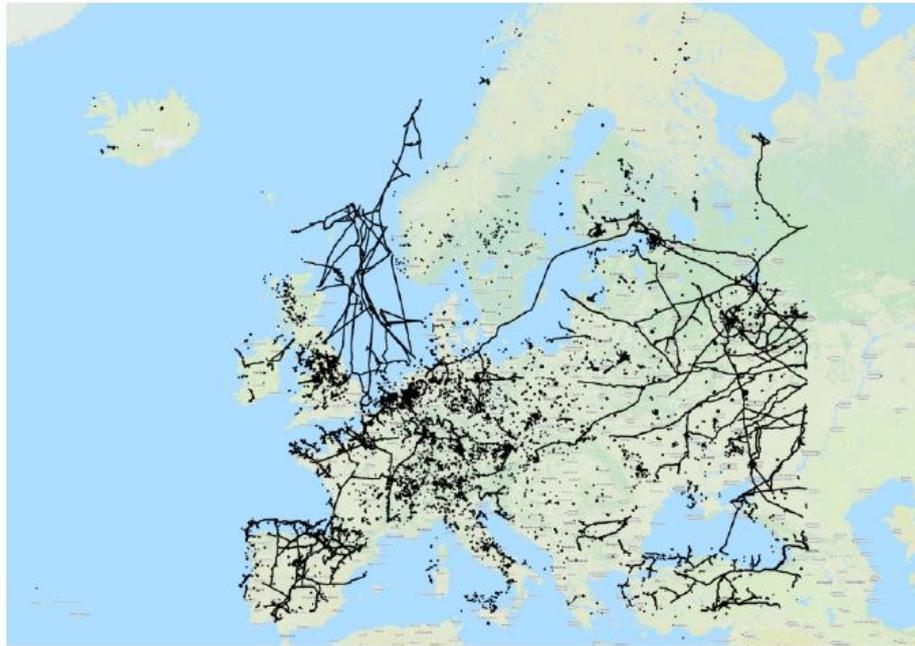
Der Fokus der projektrelevanten Forschungsarbeiten im Bereich der Gastransportnetzmodellierung lag zum Zeitpunkt der Antragsstellung (Jahr 2017) eher auf der technischen Ebene (Troost, 2016 ) (Götz, 2011) und beschäftigt sich vorrangig mit Fragen hinsichtlich der Wirtschaftlichkeit (Baumann, 2016) (Kirchmayr, 2014), der Bewertung der Systemflexibilität (Metz, 2014) sowie der Einsatzplanung und Versorgungssicherheit (Michaelis, 2013 ).

In verschiedenen Forschungsinstituten existieren hauseigene Gastransportnetzmodelle, die für Simulationen zum Gesamtsystem und für Marktsimulationen eingesetzt werden, wie z.B. Deutsches Institut für Wirtschaftsforschung (Abteilung Energie, Verkehr, Umwelt), FAU Erlangen-Nürnberg (Lehrstuhl für Wirtschaftsmathematik), TU Dresden (Fakultät Wirtschaftswissenschaften) und TU Clausthal (Institut für Erdöl- und Erdgastechnik). Diese Modelle sind jedoch öffentlich nicht zugänglich bzw. einsehbar, wodurch neue und innovative Ansätze durch andere Kompetenzträger aus Wissenschaft und Wirtschaft erschwert werden. Des Weiteren werden in den genannten Modellen vielfach vereinfachte topologische Abbildungen des Gastransportnetzes verwendet, die z.B. eine reduzierte Zahl der Gasleitungen und eine quasi-stationäre und verlustfreie Übertragung annehmen, die eine detailliertere Betrachtung sowie eine realistische Einschätzung der Potenziale und optimalen Synergieeffekte der Gas- und Stromnetze nicht erlauben. Zudem sind die Implikationen einer Vereinfachung der Netztopologie auf die Ergebnisgüte nicht bekannt. Ein Gasnetzreferenzmodell existiert ebenfalls nicht. Ferner sind nach unserem Kenntnisstand die Daten des deutschen oder europäischen Gastransportnetzes nicht frei zugänglich und müssen entweder käuflich erworben werden oder stehen unter restriktiven Copyright-Lizenzen. So bietet zum Beispiel die Gastransportnetz Karte der ENTSO-G einen detaillierten Überblick des europäischen Gastransportnetzes (ENTSO, 2009), jedoch stehen die Daten nur visuell als interaktive Karte zur Verfügung. Die zugrundeliegenden Daten der Pipelineverläufe und Verdichterstationen unterliegen einer Copyright-Lizenz und sind nicht georeferenziert. Hierdurch ist eine Verwertung für frei verfügbare wissenschaftliche oder technische Studien nicht gegeben. Im europäischen Vergleich bieten viele Netzbetreiber ihre Gasnetzdaten als PDF-Dateien ohne weitere Informationen, Downloadmöglichkeiten oder Lizensierungen an. Eine Ausnahme bildet der vom British National Grid<sup>1</sup> veröffentlichte Datensatz mit georeferenzierten Pipelines und Verdichterstationen auf der Transportebene. Zumeist stehen jedoch die Geokoordinaten der Gasnetzelemente nicht zur Verfügung und trotz einer Datenfreigabe bleiben Netzmodelle und Daten unzugänglich und die Details ihrer Generierung sind nicht transparent.

Zum Zeitpunkt der Antragsstellung standen Modelle des deutschen und europäischen Gastransportnetzes, die transparent sowohl die Modellabstrahierung als auch die Datenaufbereitung dokumentieren und für wissenschaftlich belastbare Analysen nutzbar sind, nicht frei verfügbar. Uns waren auch keine Bestrebungen anderer Arbeitsgruppen bekannt, ein Open Source Modell zu entwickeln und zur Verfügung zu stellen.

---

1 <https://www.nationalgrid.com/uk/gas-transmission/land-and-assets/network-route-maps>



**Abbildung 1:** Beispiel für verfügbare Daten auf „openstreetmap.org“ zu Gasfern- und verteilnetzen in Deutschland (Graphische Aufbereitung durch QGIS, Status 2018).

Die Hauptgründe gegen eine öffentliche Verfügbarkeit von den beschriebenen Gasnetzdaten sind Sicherheitsbedenken und wirtschaftliche Aspekte. Diese Daten gelten als sensibel, wobei insbesondere die Offenlegung des Standorts von Pipelines und Verdichterstationen dazu dienen könnte, deren Betrieb zu stören und Versorgungsausfälle zu verursachen (Otuoze, 2018) (Sanjab, 2016) (Onyeji, 2014) (Ten, 2010). Der wirtschaftliche Aspekt spiegelt wider, dass Netzdaten zur Berechnung der Erlöse von Gasnetzbetreibern und zur Spekulation über Energiepreise verwendet werden könnten (Kunz, 2017).

Im Rahmen von SciGRID\_gas wurde ein neuer innovativer Ansatz umgesetzt, mit dem auf Basis von frei verfügbaren Daten (wie zum Beispiel Zeitungsartikeln, Informationen auf Netzbetreiberwebseiten), Open Source Datenquellen (wie zum Beispiel OpenStreetMap (OSM)) sowie der Nutzung von frei verfügbaren GIS-Anwendungen und Methoden ein funktionales, frei verfügbares Datenmodell des europäischen Gastransportnetzmodells für wissenschaftliche Analysen entwickelt wurde. Ein Beispiel für in der OSM-Datenbank verfügbare Daten zum europäischen Gasübertragungsnetz ist in Abbildung 1 dargestellt. Wie aus Abbildung 1 ersichtlich ist, fehlen in OSM Gasnetzelemente in größeren Teilen Europas. Darüber hinaus sind Eigenschaften von Pipelines und Verdichterstationen (wie z.B. maximaler Druck, Kapazität, Zahl der Kompressoren), die wichtig für die Gasnetzmodellierung sind, nicht vorhanden. Da SciGRID\_gas Daten zu den Eigenschaften von Gasleitungen und Knoten den Anforderungen der Gastransportnetzmodellierung entsprechen sollten, wurden weitere Datenquellen herangezogen, um die fehlenden Informationen zu ergänzen. Die benutzten Datenquellen und deren Eigenschaften sind in der SciGRID\_gas Dokumentation detailliert beschrieben und werden kurz in Kapitel 2.3 vorgestellt.

Des Weiteren wurden SciGRID\_gas Datensätze in verschiedenen Modellen im Bereich Stromnetzsimulation und der Sektorenkopplung genutzt. Als Beispiele sind die Modelle eGon<sup>2</sup>, PyPSA<sup>3</sup> und REMix<sup>4</sup> zu erwähnen.

In den weiteren Teilen dieses Dokuments werden, sofern nicht anders angegeben, die Begriffe Gastransportnetze und Gasnetze, Erdgas und Gas sowie Gasleitung und Pipeline synonym verwendet.

## **1.2 Zusammenarbeit mit anderen Stellen**

Bei dem Projekt **SciGRID\_gas - Offenes Referenzmodell Europäischer Gastransportnetze für wissenschaftliche Untersuchungen zur Sektorenkopplung** handelt es sich um ein Einzelvorhaben.

## **2 Eingehende Darstellung der Ergebnisse**

### **2.1 Verwendung der Zuwendungen und Ergebnisse**

Im Folgenden werden die Ergebnisse der im Projekt durchgeführten Arbeiten nach den Gliederungen der Arbeitspakete dargestellt. Die Methoden und Techniken sowie die verwendete Software werden ebenfalls hervorgehoben.

### **2.2 Klassifizierung und Charakterisierung von Gastransportnetzmodellen**

#### **2.2.1 Ermittlung verschiedener Gastransportnetzmodellierungsansätze**

Analog zu elektrischen Übertragungsnetzen existieren verschiedene Methoden und Ansätze um Gastransportnetze zu modellieren und zu berechnen, wobei die Auswahl des verwendeten Ansatzes und der Methodik entsprechend der betrachteten Skala und des Ziels der Modellierung variiert. Gastransportnetzmodelle ermöglichen eine simulationsbasierte Vorhersage des Netzwerkverhaltens für verschiedenste Ausgangssituationen und Bedingungen und sind zudem wichtig bei der Gestaltung, Verwendung und Erweiterung bestehender, realer Gastransportnetzsysteme. Im Rahmen einer gründlichen Literaturrecherche wurde die Fachliteratur analysiert, die sich mit der Modellierung und Simulation von Gasnetzen befasst. Die Ergebnisse der Literaturrecherche und deren Analyse werden in den folgenden Unterabschnitten vorgestellt.

#### **2.2.2 Charakterisierung der Modellierungsansätze**

Für die Beschreibung des Gastransportnetzes wird ein mathematisches Modell eines solchen Systems erstellt, das eine gewisse Vereinfachung der Realität darstellt (Osiadacz, 1987). Das Modell des Gastransportsystems (oder -netzes) wird verwendet, um dessen Betrieb zu modellieren und zu

---

<sup>2</sup> Das [eGon](#) Projekt (FKZ: 03EI1002C) stellt die Erweiterung des [open\\_eGo](#) Modells (FKZ: 0325881D) auf die Sektoren Gas, Wärme und Mobilität dar. Open\_eGo ist ein integriertes und hoch aufgelöstes netzebenenübergreifendes Netzplanungssoftware.

<sup>3</sup> <https://pypsa.org/>

<sup>4</sup> [REMIX](#) wurde am Institut für Technische Thermodynamik des DLR entwickelt und ermöglicht eine detaillierte Darstellung der Stromerzeugung aus EE und ihrer Auswirkungen auf die Reservekapazitäten, die Übertragung und den Speicherbedarf.

optimieren. Die Simulation von Gasnetzen dient verschiedenen Zielsetzungen, die sich in vier Schwerpunktthemen zusammenfassen lassen: Engineering, Planung, Terminierung und Betrieb (PSIG, 2000). Im Folgenden werden einzelne Aspekte dieser Themen kurz beleuchtet.

- Die technischen Gasnetzsimulationen (Engineering) konzentrieren sich hauptsächlich auf die Auslegung von Gastransportleitungen und Gas-Druckregel- und Messanlagen und umfassen die Berechnung verschiedener Variablen wie z.B. Durchflussmengen, Druckanstieg und -abfall sowie die Auswirkungen von Wärme und Kälte auf die einzelnen Netzelemente.
- Die Planung (Planning) von Gasnetzen erfordert umfassendere Simulationen, wie z. B. eine Gasnetzanalyse für mehreren Szenarien oder eine Kosten- und Zuverlässigkeitsanalyse für die mittel- und langfristige Planung.
- Die Terminierung (Scheduling) umfasst Betriebsprognosen, Ausfallanalysen, die Ermittlung von Engpässen und die Systemoptimierung. Letztere erfordert die Minimierung der Betriebskosten und die Maximierung des Gasflusses unter Verwendung von Optimierungsmethoden unter Berücksichtigung von verschiedenen Szenarien.
- Der Betrieb von Gasnetzen verfolgt das Ziel, sowohl den Betrieb des Gasnetzes als auch einzelner Netzelemente zu modellieren, zu überwachen, zu prognostizieren und zu optimieren.

Vor der Liberalisierung der Gasmärkte hat sich die Modellierung von Gasnetzen hauptsächlich mit der Verbesserung bestehender Simulationen von Gastransportnetzen zur Bewertung der Gasnetzkapazitäten und zur Prognose von Gasverkäufen und -einkäufen beschäftigt (Cerbe, 2016). Nach der Liberalisierung wurden neben der Bewertung der Netzkapazitäten auch Modelle für die Modellierung von Gasmärkten sowie für die Netzplanung und den Netzbetrieb (Simulation und Optimierung) eingeführt. Dies führte dazu, dass die Betrachtung des wirtschaftlichen und politischen Rahmens für die Gasnetzanalyse erforderlich wurde, was die Art und Weise, wie die Netzplanung durchgeführt wird, grundlegend veränderte (Scheib, 2008).

Im Projekt SciGRID\_gas haben wir uns auf die Simulation und Optimierung des Gastransports (Analyse und Prognose der Gasversorgung und des -verbrauchs sowie die Entwicklung von Gasmodellen für große Netzgebiete) sowie auf den optimalen wirtschaftlichen Dispatch in Gasnetzen konzentriert.

Zur Beschreibung des Gasflusses in einem Rohrleitungsnetz werden die Grundsätze der Strömungsdynamik angewandt und die Gasflussgleichungen aus der Erhaltung von Masse, Impuls und Energie abgeleitet (Szoplik, 2016) (Szoplik, 2012) (Ke, 2000). Die Simulation des Gasflusses unterscheidet zwischen stationären und instationären (oder dynamischen) Modellen, die einen stationären oder ggf. instationären Zustand des Gasflusses in den Gasleitungen widerspiegeln. Der Betrieb des Gasnetzes ist zwar grundsätzlich dynamisch, doch die stationäre Näherung kann zur Vereinfachung verwendet werden. Die instationäre Gasströmung resultiert in partiellen Differentialgleichungen, während die stationäre Gasströmung zu nichtlinearen algebraischen Gleichungen führt (Osiadacz, 1987). Die numerische Lösung der partiellen Differentialgleichungen ist komplex und aufwändig. Wenn Änderungen des Gasdrucks und die Dynamik der Strömung vernachlässigt werden können, eignet sich daher die stationäre Simulation.

Der Gasfluss ist durch verschiedene Gleichungen, abhängig vom betrachteten Druckbereich, zu beschreiben. Diese Gleichungen sind wie folgend (wir verweisen auf (Qadrdan, 2020) (Osiaadacz, 1987) für detaillierte Beschreibungen der Gleichungen):

- Lacey's Gleichung: Diese gilt für den niedrigen Druckbereich zwischen 0 und 75 mbar.
- Polyflo Gleichung: Diese gilt für den mittleren Druckbereich 0.75 – 7 bar.
- Panhandle A Gleichung: Sie beschreibt den höheren Druckbereich über 7 bar.
- Weymouth Gleichung: Diese gilt für alle Druckbereiche durch die Anpassung der Gleichungsparameter (den so genannten Pipelineparameter).

### Die Graphentheorie als Beschreibungsmittel von Gasnetzen

Im Falle von Gasnetzen kann die Struktur des Netzes mit Hilfe der Graphentheorie dargestellt werden, die eine einfache Darstellung der Netzstruktur anhand der Eigenschaften der Inzidenz der Netzelemente ermöglicht (Abeysekera, 2016) (Rüdiger, 2009) (Szoplik, 2012) (Osiaadacz, 1987). Ein Graph besteht aus einer Menge von **Knoten** (Nodes) und **Kanten** (Edges). Die Knoten werden durch Punkte dargestellt, die mit durch Liniensegmenten dargestellten Kanten verbunden sind. Manchmal ist ein Knoten mit sich selbst verbunden, wobei dieser geschlossene Pfad von Knoten und Kanten als **Schleife** (loop) bezeichnet wird. Das Gas tritt in das Netz an einem sogenannten Versorgungsknoten ein und verlässt das Netz an einem Grenzknoten<sup>5</sup>. Wird ein Anfangs- und Endpunkt der Kante definiert, handelt es sich um einen gerichteten Graphen. Im anderen Fall, wenn die Kantenrichtung nicht definiert wird, so handelt es sich um einen ungerichteten Graphen. In der Gasnetzsimulation sind beide Graphen benötigt.

Eine Kante im Gasleitungsnetz ist eine Pipeline mit konstantem Durchmesser und konstanter Rauheit. Unter Berücksichtigung der Position einer Kante im Netz gibt es Grenzknoten sowie interne Knoten.

### Die mathematische Darstellung von Gasnetzen

Die Verflechtung der Pipelines im Netz kann durch zwei Inzidenzmatrizen dargestellt werden, nämlich durch die **Branch-Nodal-Inzidenzmatrix** (A) und die **Branch-Nodal-Inzidenzmatrix** (A). Die Inzidenzmatrix A hat  $n \times m$  Elemente, wobei  $n$  der Anzahl an Netzknoten entspricht und  $m$  der Anzahl an Kanten entspricht. Die Inzidenzmatrix B verfügt über  $k \times m$  Elemente, wobei  $k$  der Zahl an Loops entspricht. Jedes Element der Inzidenzmatrizen A und B ist entweder gleich 0, 1 oder -1 (Szoplik, 2012) (Osiaadacz, 1987). Darüber hinaus können Gasnetzelemente wie z.B. Verdichterstationen, Regler und Ventile vereinfacht durch die verursachten Änderungen im Gasdruck sowie -fluss modelliert werden.

### Die Berechnungsverfahren von Gasnetzen

Für die Simulation des stationären Flusses im Gasnetz wird die Analogie zwischen elektrischen Netzen und Gasnetzen genutzt und auf die Gasnetzberechnung adaptiert. Es ist zu beachten, dass die Analogie, wegen der Nichtlinearität der Gasströmung durch die Pipelinereibung, nur eingeschränkt anwendbar

---

<sup>5</sup> Im Falle des Gastransportnetzes sind die Grenzknoten die Punkte, an denen die Netzelemente mit dem gesamten Netz verbunden sind, und dies ist der Knoten, der zu einem einzelnen Netzelement gehört.

ist. Die Nichtlinearität resultiert aus der Druckverlustgleichung von Darcy, was dem Ohmschen Gesetz in elektrischen Netzen entspricht. Dies hat zur Folge, dass zahlreiche Berechnungsverfahren der Elektrotechnik nur begrenzt in der Gasnetzberechnung eine Anwendbarkeit finden. Im Folgenden werden Simulationsverfahren von Gasnetzen kurz dargestellt. Für eine ausführliche Beschreibung der Simulationsverfahren verweisen wir auf die folgenden Veröffentlichungen (Szoplik, 2012) (Rüdiger, 2009) (Osiađacz, 1987).

Ziel der Gasflusssimulation ist es die Gasflüsse in jeder Pipeline (Netzkanten) und den Druck an jedem Netzknoten zu berechnen. Die berechneten Gasströmungsraten und Drücke müssen der Strömungsgleichung entsprechen und zusammen mit den Gaslasten und -entnahmen dem ersten und zweiten Kirchhoffschen Gesetzen entsprechen. Laut des ersten Kirchhoff'schen Gesetzes soll die algebraische Summe der Flüsse an jedem Knoten gleich Null sein. Das bedeutet, dass die Summe der Flüsse, die in einen Knoten eintreten, gleich der Summe der Ströme sein muss, die diesen Knoten verlassen. Das zweite Kirchhoffsche Gesetz besagt, dass der Druckverlust jedem geschlossenen Loop gleich Null ist.

Zur Lösung des stationären Problems kommen mehrere Methoden in Frage; die am häufigsten verwendeten sind die Knotenmethode (*node method*), die Schleifenmethode (*loop method*) und die Schleifen-Knoten-Methode (*loop-node method*) (Brkić D. &, 2019). Die Lösung der stationären Strömungsgleichung für Gasnetze erfordert die Lösung eines Systems von nichtlinearen algebraischen Gleichungen. Die am häufigsten verwendeten numerischen Lösungsmethoden sind die Newton-Methode (Li, 2003) (Pernice, 1998) und die Hardy-Cross Methode (Brkić, 2009) (Lopes, 2004) (Gay, 1971), die im mehrdimensionalen Fall die Lösung eines Systems algebraischer Gleichungen für jede Iteration erfordert.

### **Gasnetzdaten**

Ein weiterer Aspekt liegt in der Verwendung von Daten zu Gastransportnetzstrukturen wie beispielsweise Gastransportleitungsverläufe und -längen, Druckbereiche oder die Lokalisierung von Verdichterstationen. Diese Daten sind firmeneigen und werden nicht kostenfrei oder öffentlich zugänglich bereitgestellt, was auch für die zugrundeliegenden Modellannahmen, Vereinfachungen und Aggregationsmethoden zur Datenakquisition gilt. Dieser Umstand trägt weiter zu einer steigenden Komplexität beim Vergleichen und Validieren von Gastransportnetzmodellen bei. Die Möglichkeit, Simulationsergebnisse zu diskutieren und Gastransportnetzmodelle zu validieren, ist daher stark limitiert und nahezu unmöglich.

### **2.2.3 Vergleich der Ansätze, Methoden und Modelle**

Die verschiedenen zuvor vorgestellten Methoden weisen Vor- und Nachteile sowie eine Sensitivität gegenüber einzelnen Parametern auf. Im Folgenden werden die Sensitivitäten der Knoten-, Schleifen- und Knoten-Schleifen-Methoden kurz vorgestellt.

#### **Sensibilität gegenüber dem Startwert**

Bei allen drei Methoden ist eine der Voraussetzungen, dass ein Startwert vorhanden ist. Für die Node-Methode bedeutet dies eine erste Schätzung der Drücke an den Netzknoten. Für die Loop-Methode und die Loop-Node-Methode bedeutet dies eine anfängliche Schätzung des Gasflusses durch die Pipelines. Ein guter Startwert kann zu einer schnellen Konvergenz der Lösung führen, während ein schlechter Startwert zu einer langsamen Konvergenz oder sogar zu einer Divergenz führen kann. In der

Regel ist die Node-Methode im Vergleich zur Loop- und Loop-Node-Methode empfindlicher gegenüber dem Startwert.

### **Speicher und Rechenzeit**

Die Rechenzeit ist die Zeit, die der Computer benötigt, um eine Simulation durchzuführen. Bei allen drei Lösungsmethoden hängt die Rechenzeit von mehreren Eigenschaften der Methoden ab.

Die erste Eigenschaft der Methoden, die sich auf die Rechenzeit auswirkt, ist wie Dünnbesetzt die Inzidenzmatrix ist. Dies führt dazu, dass die Loop-Methode den meisten Speicherplatz und damit mehr Rechenzeit benötigt als die beiden anderen Methoden.

Eine weitere Eigenschaft, die die Berechnungszeit beeinflusst, ist die Anzahl der Gleichungen, die in dem iterativen Prozess gelöst werden müssen. Bei der Loop-Methode und der Loop-Node-Methode, die beide auf den Loop-Gleichungen beruhen, ist die Anzahl der zu lösenden Gleichungen durch die Gesamtzahl der Schleifen vorgegeben, während die Anzahl der zu lösenden Gleichungen bei der Node-Methode die Gesamtzahl der Last-Knoten entspricht. Das bedeutet, dass die Anzahl der Gleichungen, die für die Node-Methode gelöst werden müssen, die größte unter den drei Verfahren ist. Andererseits erfordern sowohl die Loop-Methode als auch die Loop-Node-Methode eine Schleifengenerierungsmethode, die zusätzliche Berechnungen erfordert. Eine aufwändige oder vergleichsweise langsame Schleifengenerierungsmethode kann zu einer noch längeren Simulationszeit führen.

#### **2.2.4 Veröffentlichung der Ergebnisse**

Die gesammelten Ergebnisse aus dem Arbeitspaket „Klassifizierung und Charakterisierung von Gastransportnetzmodellen“ werden derzeit in einer wissenschaftlichen Veröffentlichung verfasst und in einem Open Access Journal eingereicht. Die Veröffentlichung der Ergebnisse verzögerte sich, da der Fokus auf die Fertigstellung des Datenmodells, die Validierung der resultierenden Datensätze sowie die Entwicklung und Implementierung der heuristischen Methoden gelegt wurde. Dennoch ist es gelungen das SciGRID\_gas-Modell und die daraus resultierenden Datensätze in einem wissenschaftlichen Journal zu veröffentlichen (Pluta, 2022).

### **2.3 Datengenerierung und heuristische Verfahren**

#### **2.3.1 Spezifizierung von Datenstrukturen und -modellen sowie Erstellung der relationalen Datenbank**

Aufgrund des Erfahrungsrückflusses aus dem Vorgängerprojekt SciGRID, wurde sich zur Speicherung der Daten gegen eine ursprünglich geplante Erstellung einer relationalen Datenbank und für die Erstellung eines reinen Datensatzes entschieden. Ursprünglich wurden Datenbanken genutzt, weil der Zugriff und die Datenverarbeitung von OSM Daten dies erforderlich machten. Innerhalb des SciGRID\_gas Projekts wurde eine Methode (esy-osmfilter) entwickelt, die diesen Prozess sehr effizient in Python ausführen kann. Die maßgeblichen Vorteile einer Datenbank, wie zum Beispiel die problemlose gleichzeitige Nutzung durch mehrere Benutzer oder die Verarbeitungsmöglichkeit von riesigen Datenmengen stellte damit keine projektrelevanten Vorteile mehr dar, weder für die Ersteller noch für die Nutzer. Stattdessen wurde die Vorteile einer rein Python-basierten Prozessierungskette

und die Möglichkeit Versionsarchivierungen der erstellten Datensätze auf open-access Repositories (z.B. Zenodo) sehr positiv bewertet.

Im folgenden Abschnitt werden die Auswahl und Erstellung des topologischen Datenmodell des Gastransportnetzes inklusive der Spezifizierung der Eingangsdatenstrukturen vorgestellt. Anschließend werden die Verarbeitung der OSM Daten sowie Gasnetzdaten aus anderen Quellen dargestellt. Im nächsten Schritt wird die Methodik zur Erstellung eines kombinierten Datensatzes kurz erläutert. Zu diesem Zweck wurden Routinen zur automatisierten Extrahierung und Auswertung von Daten entwickelt.

In den folgenden Abschnitt werden die Auswahl und Erstellung des topologischen Datenmodell des Gastransportnetzes inklusive der Spezifizierung der Eingangsdatenstrukturen vorgestellt. Anschließend wird die Verarbeitung der OSM XML-Daten sowie Gasnetzdaten aus anderen Quellen und die Erstellung der relationalen Datenbanken kurz dargestellt. Zu diesem Zweck wurden Routinen zur automatisierten Extrahierung und Auswertung von Daten entwickelt. Es wurde gegen eine Nutzung einer relationalen Datenbank für die Speicherung der Datensätze entschieden. WHY??

Die OSM-Daten bestehen aus georeferenzierten Objektdaten, die in der OSM Datenbank gespeichert sind und frei verfügbar sind. Allerdings enthält der OSM-Datensatz derzeit kaum andere Informationen als den Standort der Einrichtungen. Wir haben den Ansatz verfolgt Non-OSM-Datensätze zu sammeln um einige der Datenlücken zu füllen. Non-OSM Daten sollen Informationen wie z.B. Pipeline-Durchmesser oder Kompressorkapazitäten liefern.

Ein Teil des SciGRID\_gas-Projekts bestand darin, solche Daten zu ermitteln und zusammenzustellen, und sie mit dem OSM-Datensatz zu kombinieren. Darüber hinaus wurden heuristische Methoden genutzt, um Datenlücken zu schließen, so dass ein vollständiges Gasnetz erstellt werden kann.

Datenstruktur: Beschreibung, Auflistung der Netzkomponenten (Gaskomponententypen) Beschreibung der Attribute und deren Werte

Eine gut durchdachte und dokumentierte Datenstruktur ist von grundlegender Bedeutung und sollte das Gastransportnetz so gut wie möglich abbilden können. Es wurde eine Anzahl von Netzkomponenten ermittelt, die als Bausteine des Gasnetzes definiert worden sind. Darüber hinaus enthält jede Netzkomponente Attribute (auch Merkmale oder Eigenschaften genannt) wie z.B. Pipelinedurchmesser, maximaler Betriebsdruck, maximale Kapazität oder Anzahl der Turbinen einer Verdichterstation. Diese Attribute sind wichtige Informationen, die für die Beschreibung und Modellierung des Gasnetzes sehr wichtig sind. Im Projekt SciGRID\_gas wurde die Datenstruktur anhand von Python-Klassen definiert. Dieser Abschnitt enthält zusammengefasste Informationen über die SciGRID\_gas-Datenstruktur und das Datenformat. Entscheidend für das Verständnis der Datenstruktur ist ein gutes Verständnis der im SciGRID\_gas Projekt und Modell allgemein verwendeten Terminologie. Diese wird hier kurz eingeführt. Eine detaillierte Beschreibung der Terminologie kann aus der SciGRID\_gas-Dokumentation entnommen werden.

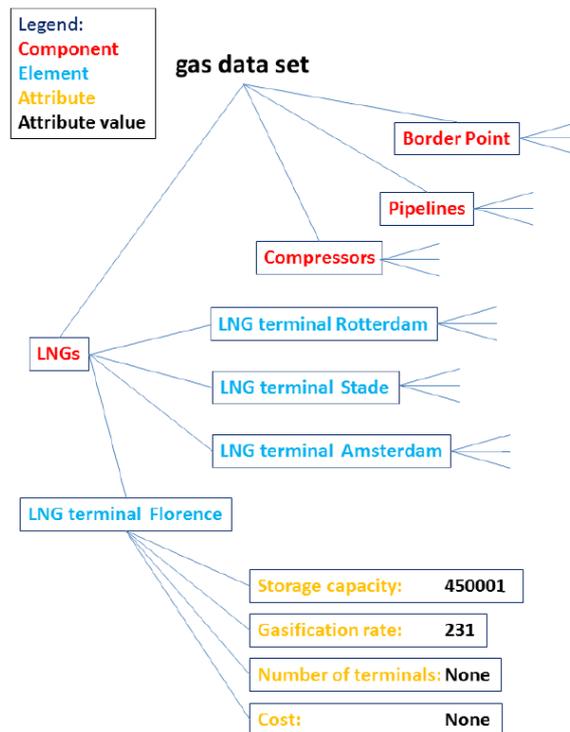
## **Modellterminologie**

Die Zusammenhänge zwischen den Begriffen "Netzdatensatz", "Komponente", "Element", "Attribut" und "Attributwert" sind in Abbildung 2 dargestellt. Wie zu erkennen ist, besteht ein einzelner Netzdatensatz aus mehreren Komponenten, wobei jede Komponente mehrere Elemente enthält und jedes Element mehrere Attribute hat, die jeweils mit einem Wert versehen sind, wobei "None" für nicht vorhandene Werte steht.

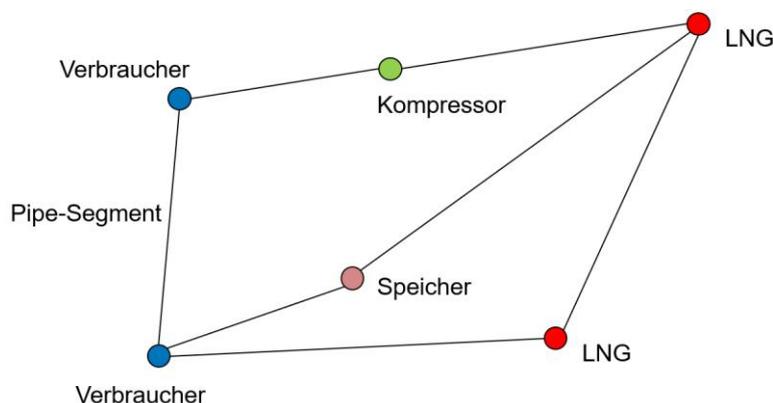
1. **Gastransportnetz:** Der Begriff „Gastransportnetz“ bezeichnet das physikalische Netz für den Gastransport aus Gasexportländer wie z.B. Russland zu den Verbraucherländern wie Deutschland. Es wird zwischen Gastransport- und Gasverteilungsnetz unterschieden. Im Projekt SciGRID\_gas wurde nur das Transportnetz, d.h. Gasnetze ab einem Druckbereich von 20 bar (auch Gasfernleitungsnetz genannt) berücksichtigt.
2. **(Gas)Netzkomponente:** Das Gasnetz besteht aus einer Reihe von Komponenten wie z.B. Kompressoren, Verdichterstationen, LNG-Terminals oder Pipelines. Die Anzahl der zu berücksichtigenden Komponenten hängt von der angestrebten Detailtiefe der Modellierung ab. Das SciGRID\_gas Datenmodell besteht aus etwa einem Dutzend verschiedenen Komponenten. Diese werden im Folgenden erläutert. Abbildung 2 zeigt 4 Netzkomponenten in Rot markiert.
3. **(Gas)Netzelement:** Der Begriff "Element" bezieht sich auf einzelne Anlagen (Komponenten), z. B. das LNG-Terminal in Rotterdam oder den Verdichter in Radeland. In Abbildung 2 sind Netzelemente in Blau markiert. Die erste Anlage (LNG-Terminal in Rotterdam) ist ein Element der Komponente LNG-Terminals, während die zweite Anlage (Verdichter in Radeland) ein Element der Komponente Kompressoren ist. Das bedeutet, dass in einem Gasnetz zwei Elemente jeweils zwei verschiedene Anlagen einer bestimmten Art von Komponenten beschreiben.
4. **(Gas)Netzdatensatz:** Dieser Netzdatensatz enthält Knoten und Kanten, die zusammenhängend miteinander verbunden sind. Alle Netzelemente, mit Ausnahme der Rohrleitungen, sind den Knoten zugeordnet, während die Rohrleitungen den Kanten zugeordnet sind.
5. **Attribute:** Attribute definieren die (technischen) Eigenschaften, die mit einzelnen Elementen verknüpft sind. Beispiele für Attribute sind "Pipeline-Durchmesser", "maximale Kapazität", und "maximaler Gasleitungsdruck". In Abbildung 2 sind Attribute in Gelb markiert. Im SciGRID\_gas Datenmodell ist eine große Zahl an Attributen definiert und dieselben Attribute können für verschiedenen Komponenten vorkommen, z.B. "max flow capacity" ist sowohl für Pipelines als auch für Verdichter definiert.
6. **Attributwert:** Jedes Attribut hat einen Wert, meist eine Zahl oder einen String, das einen Eigenschaftswert definiert. In Abbildung 2 sind Attributwerte in schwarz markiert. Es steht eine "1" für Wahr (True) und eine "0" für Falsch (False). Wenn keine Attributwerte vorhanden sind, weil der Wert fehlt, wird "None" als Attributwert genommen.

### SciGRID\_gas Modelldatenstruktur

Auf der Grundlage der Literaturrecherche und Untersuchung in AP2 (Klassifizierung und Charakterisierung von Gastransportnetzmodellen) wurde eine Datenstruktur für das Gastransportnetz festgelegt. Die Datenstruktur ist in Abbildung 3 dargestellt und umfasst die Bestandteile des Netzes und deren technischen Eigenschaften. Diese wird hier kurz eingeführt. Eine detaillierte Beschreibung der Datenstruktur sowie der Quellcode, mit dem öffentlich verfügbare Daten importiert werden können, sind in der SciGRID\_gas-Dokumentation die zusammen mit den jeweiligen Gasnetzdatensätze veröffentlicht wurde, enthalten<sup>2</sup>.



**Abbildung 2:** Darstellung der Terminologie des SciGRID\_gas Modells



**Abbildung 3:** Datenstruktur des SciGRID\_gas Modells

- 1) **Knoten:** Im Gasnetz fließt das Gas zwischen zwei Knoten (Punkten oder auch Ecken genannt) entlang einer Pipeline. Diese Punkte sind durch ihre Koordinaten definiert. Alle andere Gasnetzelemente (außer Pipelines) wie z.B. Verdichterstationen, Speicher, LNG, Kompressoren und Gaskraftwerke (Erzeuger) sind im Gasnetzmodell als Knoten dargestellt (siehe Abbildung 3). Der Begriff Knoten kommt aus der Graphentheorie und beschreibt zusammen mit den Kanten die Bestandteile eines Netzes.
- 2) **PIPELINE (Gastransportleitung):** PIPELINES oder auch Gastransportleitungen ermöglichen den Transport des Erdgases von einem Ort (Knoten) zu einem anderen Ort. Pipelines sind auch ebenfalls georeferenziert, durch die Auflistung der Koordinaten deren Start- und Endknoten.

- 3) **PIPESEGMENT (Pipeline Segment):** *PIPESEGMENTS* sind Pipelinesegmente, die nur zwei Knoten verbinden dürfen. Daher kann ein *PIPELINES*-Element (das 3 oder mehr Knoten verbindet) leicht in mehrere *PIPESEGMENTS*-Elemente umgewandelt werden.
- 4) **COMPRESSOR (Verdichterstation):** Erdgas, das durch die Gaspipeline fließt, verliert u.a. aufgrund der Reibung an den Pipelinewänden an Druck. Verdichterstationen dienen der Erhöhung des Gasdrucks und ermöglichen somit, dass das Erdgas von einem Knoten zum nächsten fließt. Eine Gasverdichterstation enthält mehrere Gasverdichter (auch Kompressoren oder Turbinen genannt).
- 5) **LNG (Liquified Natural Gas):** es gibt zwei LNG Komponenten im SciGRID\_gas Modell, LNG-Terminal und LNG-Storage (Speicher). Das Flüssigerdgasterminal (**LNG-Terminal**) ist der Knotenpunkt für die Entladung von LNG-Tankern, die Regasifizierung und Einspeisung in das Gasnetz oder die verflüssigte Einlagerung in Tanks (**LNG-Storage**) sowie für den Weitertransport von flüssigem Erdgas auf Schienen (Kesselwagen) oder auf Wasser (in bunkerschiffen).
- 6) **STORAGE (Speicher):** Erdgasspeicher sind große, größtenteils unterirdische Speicher (z. B. in alten Gasfeldern oder Salzkavernen), mit dem sich saisonale Nachfrageschwankungen oder Versorgungsengpässe bei Erdgas ausgleichen bzw. vorbeugen lassen. Die Speicher werden in der Regel in den Sommermonaten befüllt und während der Heizperiode im Winter teilweise entleert.
- 7) **CONSUMER (Verbraucher):** *CONSUMERS* bezeichnet die Erdgasverbraucher. Das können Haushalte, Industrie, Gewerbe und Kraftwerke sein
- 8) **PRODUCTION (Erzeugung):** Dabei handelt es sich um Erdgasbohrungen innerhalb eines Landes, bei denen das Gas aus dem Boden gefördert wird. Der größte Teil des in Europa verbrauchten Erdgases wird außerhalb Europas gefördert. Es gibt jedoch mehrere kleinere Gasproduktionsstätten, die über ganz Europa verteilt sind (z.B in Norwegen und in die Niederlande).
- 9) **BORDERPOINT (Grenzpunkt):** *BORDERPOINTS* sind grenzüberschreitende Knoten (zwischen verschiedenen Ländern), die meist der Gasverbrauchsabrechnung dienen. Die meisten Großgasleitungen haben grenzüberschreitende Verdichterstationen (z.B. *Ellund* (lat/long: 54.80181, 9.289079) an der Grenze zwischen Deutschland und Dänemark, mit Gasanlagen auf beiden Seiten der Grenze).

## Datenquellen

Aufgrund von Urheberrechtsbeschränkungen ist es dem SciGRID\_gas Projekt/Team nicht erlaubt, die tatsächlich heruntergeladenen Daten und Attributwerte aus machen Datenquellen, wie z.B. IGU und GIE, weiterzugeben. Die Informationen aus diesen Datensätzen konnten jedoch im Rahmen der heuristischen Datengenerierung verwendet werden, um fehlende Attribute zu korrigieren bzw. zu ergänzen.

## A) OpenStreetMap

OpenStreetMap (OSM)<sup>6</sup> ist eine frei verfügbare und modifizierbare Geodatenbank der Welt mit stetig wachsender Datenabdeckung und Datenqualität. In der Vergangenheit haben die OSM-Daten einen Beitrag zur Modellierung von Energiesystemen geleistet, beispielsweise bei der Erstellung von Stromnetzmodellen in dem SciGRID Projekt (BMW,FKZ: 03SF0471).

Im Allgemeinen können Objekte in der OSM-Datenbank durch relevante Schlüssel-Wert-Paare (key, values), so genannte Tags, identifiziert werden. Innerhalb des SciGRID\_gas Projektes wurden relevante Schlüssel-Wert-Paare <sup>7</sup> im Zusammenhang mit unterschiedlichsten Komponenten des Gastransportnetzwerkes identifiziert.

Im Folgenden wurde die Vollständigkeit der verfügbaren Daten für alle Gasnetzkomponenten untersucht. Dieser iterative Prozess erforderte die wiederholte Filterung der OSM-Datenbank nach verschiedensten Schlüssel-Werte-Paaren; ein Prozess der sich aufgrund der für Europa zugrunde liegenden Datenmenge von 20,6 GB (pbf-Dateigröße Europa 2019) mit bisherigen frei verfügbaren OSM-Tools nicht in einem befriedigendem Zeitrahmen realisieren ließ. Weiterhin erforderten bestehende Prozessierungsketten der Daten die Verwendung von verschiedenen UNIX basierten Tools, die unter anderem die Nutzung von Datenbanken erforderlich machten. Das SciGRID\_gas Projekt hat aus diesem Grund sein eigenes Python basiertes Tool zur effizienten Filterung von OSM-Daten entwickelt und der wissenschaftlichen Community zur Verfügung gestellt.

Der **esy-osmfilter**<sup>8</sup> wurde samt Dokumentation als PIPY Library veröffentlicht und sowohl in einer Journalveröffentlichung als auch im Rahmen eines Energy Modelling Workshops der wissenschaftlichen Community vorgestellt. Mit bisher über 36.000 Downloads (Stand 21.1.2022), kann der Erfolg und Impact dieses Tools sehr positiv bewertet werden und scheint weit über unsere eigene Forschungscommunity hinaus zu gehen. Insbesondere zeigen wir uns darüber erfreut, dass das Projekt „PyPSA meets Africa“ unsere Library nutzt und zitiert<sup>9</sup>.

Anstatt die erzeugten Daten in einem SciGRID\_gas Datensatz zu veröffentlichen, der zum Zeitpunkt seiner Veröffentlichung schon veraltet wäre, haben wir uns dazu entschlossen eine weitere Python Bibliothek `osmscigrid` zu veröffentlichen, welche im Stande ist, den aktuellsten OSM-Datensatz in ein SciGRID\_gas Datensatz umzuwandeln. Dieser Prozess beinhaltet unter anderem das automatische Identifizieren von nicht standardisierten Attributen (oder Parameterinformationen) wie z.B. Durchmesser und Drücke mittels Machine Learning (ML) Techniken und deren Konvertierung in einheitliche Einheiten. Die erzeugten Daten wurden hinsichtlich ihrer Eignung zur Erzeugung eines Gastransportnetzwerkes untersucht.

Im Hinblick auf die reine Vollständigkeit der in OSM hinterlegten Gas Transportleitungen zeigte sich, dass die Datenlage von Land zu Land noch immer stark variiert. Wir haben **esy-osmfilter** verwendet, um das Datenwachstum zu analysieren. Dieses Wachstum wurde mittels historischer Daten zwischen (2014-2019) in einem Video auf der Projekt Homepage visualisiert<sup>10</sup>. Unsere Analyse zeigte, dass die

---

<sup>6</sup> <https://www.openstreetmap.org/>

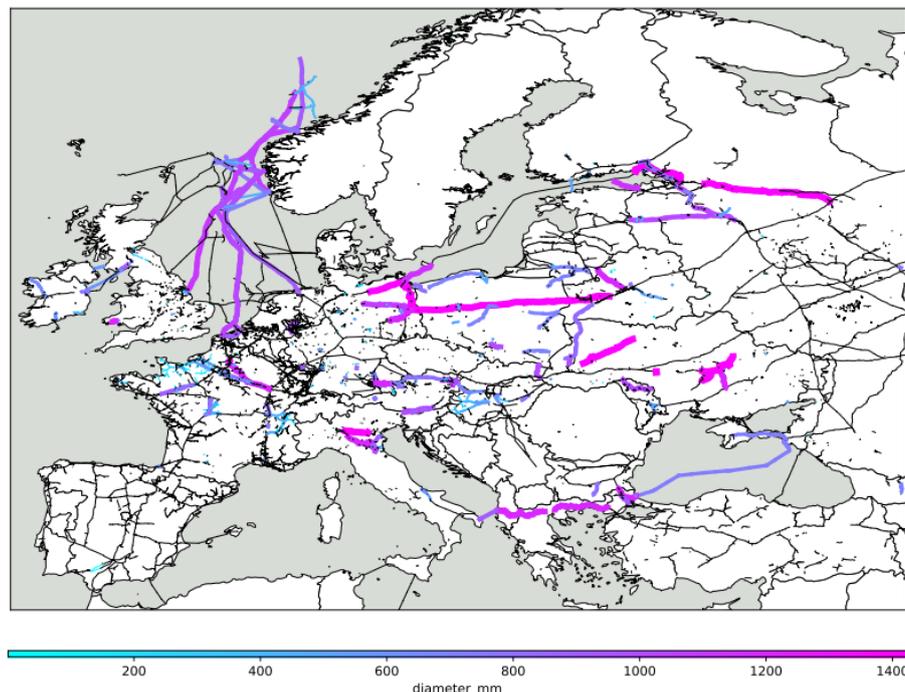
<sup>7</sup> z.B.: „man\_made“:“pipeline“, „substance“:“gas“

<sup>8</sup> <https://gitlab.com/dlr-ve-esy/esy-osmfilter>

<sup>9</sup> Kirli et al. (2021). PyPSA meets Africa: Developing an open source electricity network model of the African continent. 1-6. 10.1109/AFRICON51333.2021.9570911.

<sup>10</sup> [https://www.gas.scigrd.de/posts/2019-Nov-04\\_european-gas-pipelines-in-osm-2014-2019.html](https://www.gas.scigrd.de/posts/2019-Nov-04_european-gas-pipelines-in-osm-2014-2019.html)

Daten von Gastransportleitungen schnell wachsen, dies aber von Land zu Land mit sehr unterschiedlichen Geschwindigkeiten. Als besonders bemerkenswert empfinden wir die Tatsache, dass für Spanien, dem einzige große Flächenland mit vollständig hinterlegtem Pipelinenetz, fast alle Daten zwischen 2015-2016 hinterlegt wurden. Weitere Nachforschungen bestätigten die Vermutung, dass hier der Pipelinebetreiber selbst seine Daten offengelegt hatte. Dies führt allerdings auch zu dem Umkehrschluss, dass ohne das zu tun der Pipelinenetz Betreiber, die Vollständigkeit der Pipelinenetzinformationen noch lange auf sich warten lassen wird.



**Abbildung 4:** Gas Transport Pipelines mit Diameter in OSM für Europa, extrahiert mit osmscigrd (Stand 2020).

Des Weiteren haben wir die Datenlage auch im Hinblick auf Parameterinformationen der Pipelines analysiert. Hierbei zeigte sich, dass die Datenlage für Pipeline-Parameter sehr dünn ausfällt und selbst einer der physikalisch wichtigsten Parameter, wie der Diameter ist bisher nur sehr selten in den Daten hinterlegt ist (siehe Abbildung 4). Noch schlechter sieht die Datenlage bei den Maximaldrücken der Pipelines aus. Die Gründe hierfür liegen immer noch in der Tatsache begründet, dass europäische Pipelines zu weit über 90% unterirdisch verlegt sind.

### **B) Non-OSM-Datensätze**

Zu den non-OSM-Daten gehören Daten aus u.a. eigenen Internetrecherchen, Pressemitteilungen, Transparenzplattformen und öffentliche Daten der Übertragungsnetzbetreiber (ÜNBs oder TSO auf Englisch) sowie nationale Gasnetzdatensätze (wie z.B. Daten der britischen Gasnetzbetreiber *National Grid*<sup>11</sup> oder Daten des Norwegisches Erdöldirektorat<sup>11</sup>). Einige der Informationen mussten aufgrund entsprechender EU-Verordnungen von TSOs zu Verfügung gestellt werden. Andere Informationen

<sup>11</sup> <https://www.npd.no/en/about-us/information-services/open-data/>

wurden im Rahmen der Selbstdarstellung der Unternehmen veröffentlicht (z.B. Gas Infrastructure Europe<sup>12</sup>). Im Rahmen des SciGRID\_gas-Projekts wurden Daten verwendet die ausreichend Qualität und Detaillierungsgrad aufweisen, gängige und einheitliche Formate haben sowie unter offene Lizenzen bereitstehen. Wie bereits erwähnt waren die OSM-Datensätze des europäischen Gasnetzes unvollständig, sodass ein erheblicher Zeitaufwand für die Recherche und Beschaffung von non-OSM-Datensätzen erforderlich war.

Non-OSM-Datenquellen sind sehr spezifisch und beziehen sich nur auf bestimmte Erdgasinfrastrukturkomponenten. Der Datensatz GIE<sup>12</sup> beispielsweise liefert Informationen über den täglichen Gasfluss in und aus den Gasspeichern in LNG-Terminals, jedoch enthalten die Daten enthalten jedoch keine Informationen über deren georeferenzierten Standort. Andere Datenquellen, wie z. B. der LKD-Datensatz<sup>13</sup>, sind recht detailliert in Bezug auf Pipelines, Verdichterstationen und Gasverbräuche, jedoch sind die Daten nur für Deutschland verfügbar. Daher bestand die Hauptaufgabe darin, die non-OSM Datenquellen genau zu untersuchen und herauszufinden welche Attributwerte verwendet und in das SciGRID\_gas-Modell integriert werden können. Darüber hinaus wurden die urheberrechtlichen Einschränkungen der Datenquellen ermittelt und dokumentiert. Im Folgenden werden die wichtigsten non-OSM Datenquellen und deren Abkürzungen im SciGRID\_gas Datenmodell, die im Rahmen des Projekts verwendet werden, kurz vorgestellt. Die verschiedenen Datensätze wurden in verschiedenen Iterationen zusammengefügt, um ein zusammenhängendes Gasnetz zu erstellen. Für diese Iterationen wurden Datensätze zusammengeführt und als Datensatzpaket auf Zenodo unter CC-BY-Lizenz veröffentlicht.

- 1) **InternetDaten (INET raw)**: Dieser Datensatz ist eine Sammlung aus 224 Open Source Quellen (siehe die Datendokumentation auf Zenodo<sup>14</sup>) zum europäischen Gastransport die im SciGRID\_gas-Projekt erstellt wurde. Er enthält Daten über 60.000 km des europäischen Transportleitungsnetzes sowie über Gasspeicher, -erzeugung, LNG-Terminals, EntryPoints, InterConnectionPoints und Verdichterstationen.
- 2) **InternetDaten (INET filled)**<sup>15</sup>: Dieser Datensatz enthält den ursprünglichen SciGRID\_gas INET\_raw-Datensatz. Fehlende Werte in INET\_raw wurden mit heuristischen Methoden ermittelt und in den Datensatz ergänzt.
- 3) **GIE**: Die Daten stammen von der Transparenzplattform *Gas Infrastructure Europe*<sup>12</sup> (GIE) und beinhalten Daten zu den Speichern und LNG-Terminals in Europa. Die Speicher- und LNG-Daten können über eine von GIE bereitgestellte API abgerufen werden. Es war notwendig, ein Python-Tool zu erstellen, um die Standorte der Komponenten in georeferenzierte Koordinaten zu konvertieren, damit sie im SciGRID\_gas-Projekt eingesetzt werden können.
- 4) **GSE**: Der GSE-Datensatz (Gas Storage Europe) enthält nur Informationen über die Gasspeichereinheiten und unterscheidet sich leicht vom GIE-Datensatzes. Die Daten können als Excel-Tabelle von der GIE-Webseite heruntergeladen werden. Wie im Fall der GIE-Daten, wurde ein Konvertierungstool in Python geschrieben, um die georeferenzierten Koordinaten der Komponenten zu erhalten.

---

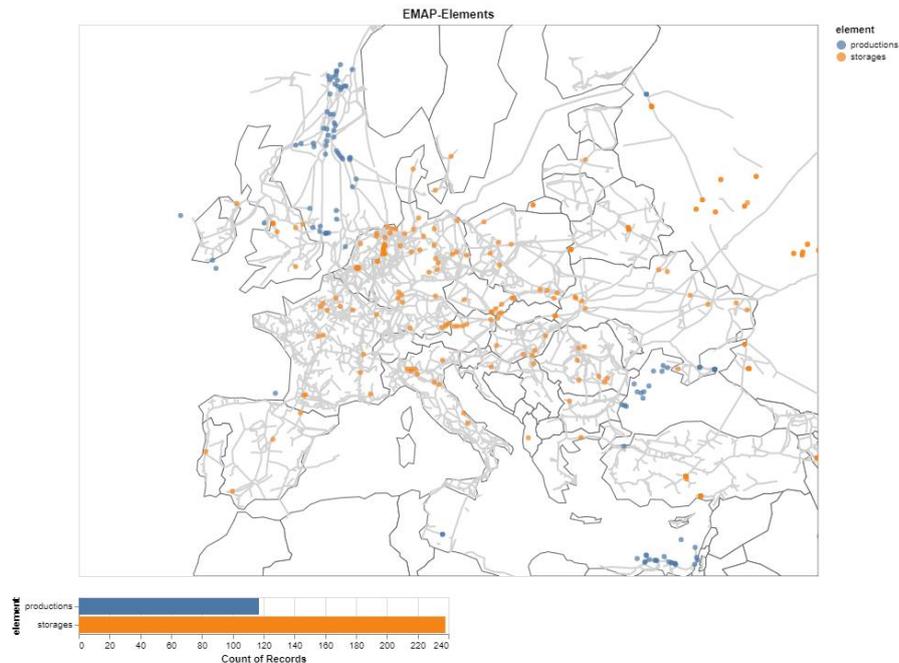
<sup>12</sup> <https://www.gie.eu/>

<sup>13</sup> <https://tu-dresden.de/bu/wirtschaft/ee2/forschung/projekte/lkd-eu>

<sup>14</sup> [https://zenodo.org/record/4616358#.Yd8p7\\_6ZOUl](https://zenodo.org/record/4616358#.Yd8p7_6ZOUl)

<sup>15</sup> [https://zenodo.org/record/4008975#.Yd8qq\\_6ZOUk](https://zenodo.org/record/4008975#.Yd8qq_6ZOUk)

- 5) **IGU:** Ein weiterer Datensatz für Gasspeicher stammt von der International Gas Union (IGU). Bei den Daten handelt es sich um webbasierte Tabellen für etwa 170 europäische Speicherstätten, die über ein Online-Portal abgerufen werden können. Die Tabellen beinhalten verschiedenen Informationen über u.a. die Spitzenentnahmekapazität und die Injektionskapazität der Gasspeicher. Ein Python-Tool wurde entwickelt, um die Daten der IGU-HTML-Webseite zu konvertieren und sie im SciGRID\_gas-Projekt zu verwenden.



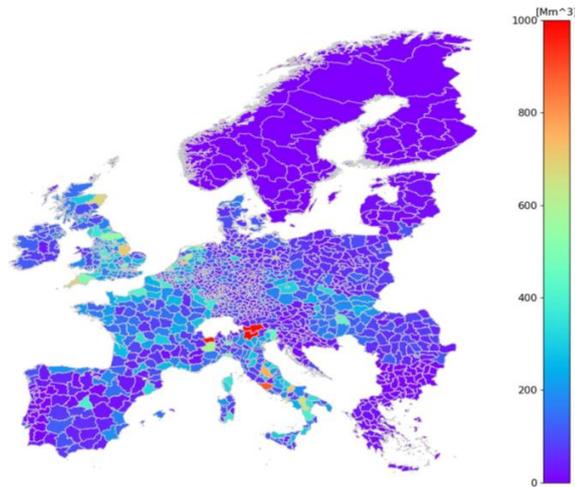
**Abbildung 5:** Visualisierung des EMAP-Datensatzes. Die interaktiv programmierte Karte kann nach Land und Komponententyp visualisiert werden.

- 6) **LKD:** Dieser Datensatz beinhaltet die georeferenzierten Daten des deutschen Gastransportnetzes und wurde im Rahmen des Forschungsprojekts LKD-EU (*Langfristige Planung und kurzfristige Optimierung des Elektrizitätssystems in Deutschland im europäischen Kontext*) erstellt. Der LKD-EU Datensatz umfasst die Koordinaten und (einige) Attributwerte für Pipelines, Erzeugung, Speicher, Verdichterstationen und den verschiedenen Knotenpunkten.
- 7) **GB:** Wie im Kapitel 1.1 ausgeführt, stellt der britische Strom- und Gasnetzbetreiber *National Grid* der Öffentlichkeit Informationen über sein Gastransportnetz unter einer offenen Lizenz zur Verfügung. Insgesamt ist die topologische Qualität des Datensatzes sehr hoch und umfasst detaillierte Informationen über die Standorte von Pipelines und Kompressoren in England, Schottland und Wales.
- 8) **NO:** Wie in Großbritannien stellt der norwegische Gastransportbetreiber *Gassco* georeferenzierte Datensätze für das Gastransportnetz zur Verfügung, die die Hoheitsgewässer von Norwegen, Frankreich, Großbritannien, Deutschland, Dänemark, den Niederlanden und Belgien umfassen. Der Anlagendatensatz liegt in Form eines Shapefiles vor und besteht aus Polylinen (Polygonzug) mit wichtigen Attributen, wie z. B. dem Rohrdurchmesser. Insgesamt

ist die topologische Qualität des Datensatzes sehr hoch. Der gesamte Geodatensatz wird automatisch in den endgültigen SciGRID\_gas-Gasdatensatz integriert.

- 9) **CONS:** Die Modellierung von Gas Konsumenten war ursprünglich nicht Teil des SciGRID\_gas Projekts. Zur Projektlaufzeit erreichte uns jedoch das Feedback von vielen potentiellen Anwendern, dass diese Daten einen sehr hohen Mehrwert für das SciGRID\_gas Projekt darstellen würden. Aus diesem Grund wurden die Daten im Zuge einer Masterthesis erstellt:

*“Estimation and simulation of a gas demand time series for the European NUTS 3 regions”*  
(Javier Enrique Sandoval, 2021)<sup>16</sup>



**Abbildung 6:** Jährlich gemittelte Gasnachfrage im Haushaltssektor zwischen 2010-2019, aufgeschlüsselt nach Nuts-3-Regionen.

Das Europäische Statistische Amt (Eurostat) sammelt und veröffentlicht monatlich und jährlich Daten über Energieversorgung, -umwandlung und -verbrauch. Eurostat-Statistiken, wie der EuroStat Datensatz „complete energy balances“<sup>17</sup> und andere, lieferten die Datengrundlage für eine unserer Studien über den europäischen Gasbedarf. Wir haben Zeitreihen für die Gasnachfrage mit täglicher Auflösung für die Jahre 2010-2019 mit einer räumlichen Auflösung von NUTS1, NUTS2 und NUTS 3<sup>18</sup> abgeleitet. Um detaillierte Informationen für die Modellierer und die Flexibilität des Datensatzes bereitzustellen, werden die Zeitreihen nach den Sektoren Haushalte, Gewerbe und Industrie unterschieden.

Die Datenableitungstechniken zeigten gute Benchmarking-Ergebnisse gegenüber drei bestehenden Zeitreihen der Gasnachfrage in Deutschland, die aus dem DemandRegio-Projekt<sup>19</sup> stammen. Wesentliche Erkenntnisse wurden bei der Analyse der Zeitreihen

---

<sup>16</sup> <https://elib.dlr.de/144187/>

<sup>17</sup> [https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=nrg\\_bal\\_c&lang=en](https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=nrg_bal_c&lang=en)

<sup>18</sup> NUTS ist ein geografisches System zur Unterteilung des EU-Gebiets in hierarchische Ebenen. In Deutschland sind NUTS 1 die Bundesländer, NUTS 2 die Regierungsbezirke und NUTS 3 die Kreise für 27 europäische Länder.

<sup>19</sup> Gotzens et al.: DemandRegio: Harmonisierung und Entwicklung von Verfahren zur regionalen und zeitlichen Auflösung von Energienachfragen, 2020

hinsichtlich der saisonalen, geographischen und sektorspezifischen Variabilität der Gasnachfrage in Europa gewonnen.

## 2.4 Beschaffung von Referenzdaten

Der Verband Europäischer Fernleitungsnetzbetreiber für Gas (kurz ENTSO-G) stellt eine PDF-Karte zur Verfügung, die Informationen über das Gastransportnetz in Europa einschließlich Nicht-EU-Staaten wie Russland, Ukraine, Weißrussland, Georgien und Aserbaidschan enthält. Die Karte wird in unregelmäßigen Abständen veröffentlicht. Die letzte aktuelle Karte wurde 2019 herausgegeben<sup>20</sup>. Die Beschaffung von ENTSO-G Daten als Referenzdatensatz dient der Validierung des SciGRID\_gas Gastransportnetzmodells. Zur Konvertierung der in der PDF-Karte bereitgestellten Daten in die SciGRID\_gas-Datenstruktur sind mehrere Schritte notwendig. Im Rahmen dieses Arbeitspakets wurden Python-basierte Skripte erstellt, um einen Datensatz aus der PDF-Karte zu extrahieren. Dieser Prozess lässt sich jedoch leider nicht vollständig automatisieren. Es wurden dennoch Schritte unternommen, um so viele Schritte wie möglich zu automatisieren, während einige Bereinigungen vom Benutzer manuell durchgeführt werden müssen. Die in der PDF-Karte verfügbaren Komponenten und deren Attributen sind wie folgt: Pipelines (Koordinaten, Länge, Dicke), Gasspeicher (Koordinaten), und Gaserzeugungspunkte (Koordinaten).

### 2.4.1 Prozessierung von Referenzdaten

Die Aufbereitung des ENTSO-G-Datensatzes aus der PDF-Karte ist eine ziemlich aufwändige Aufgabe, die hier nur kurz vorgestellt werden soll. Der gesamte Prozess ist in der SciGRID\_gas-Dokumentation ausführlich erläutert. Hier wird beschrieben, wie der ENTSO-G-Datensatz, hier **EMAP** genannt, aus der PDF-Karte generiert wurde und in ein SciGRID\_gas-Datenobjekt umgewandelt wurde. Lustenberger et al. (Lustenberger, 2019) haben einen ähnlichen Ansatz zur Aufbereitung desselben Datensatzes vorgestellt, allerdings unter Verwendung des nicht-offenen Tools ArcGIS. Hier wird das Open-Source-Tool QGIS verwendet. Nachfolgend wird ein allgemeiner Überblick über die Schritte zur Konvertierung der ENTSO-G -PDF-Karte in ein einzelnes SciGRID\_gas-Gasnetzdatenobjekt dargestellt.

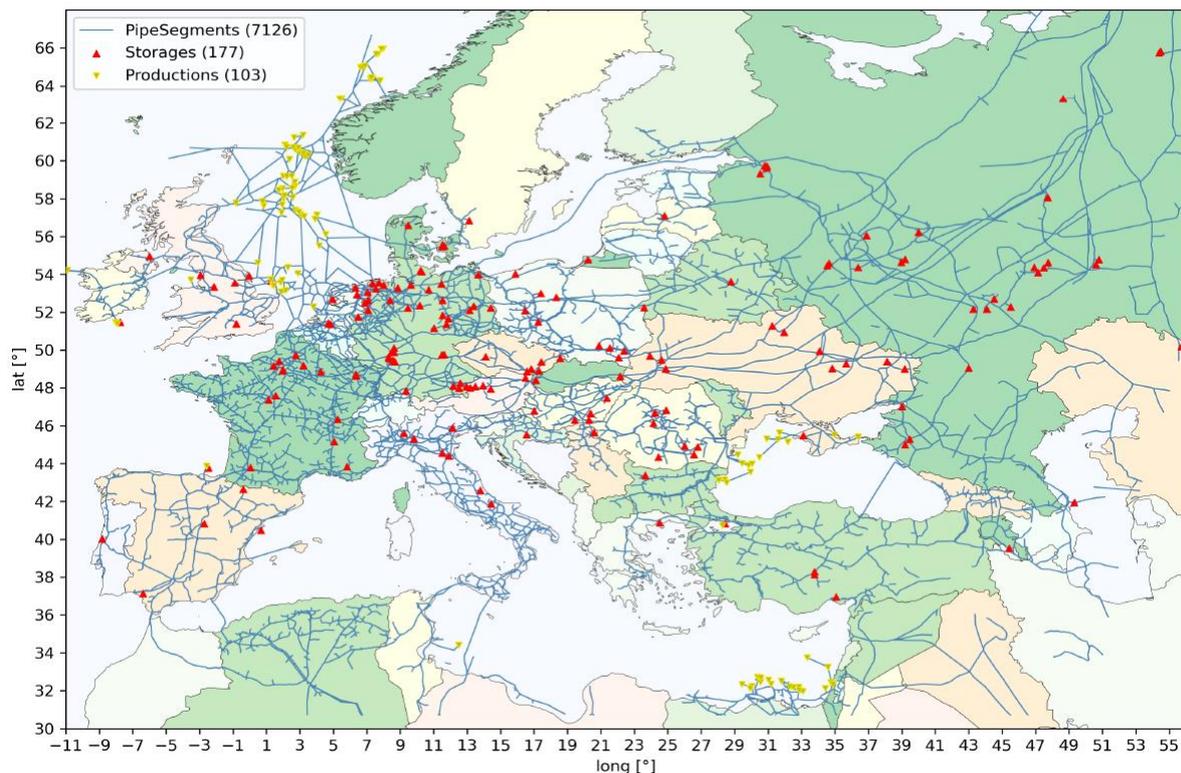
- 1- Trennung der einzelnen PDF-Ebenen (Layers) aus der ursprünglichen PDF-Karte in separate Dateien (**Erzeugung der PDF-Ebenen**).
- 2- Konvertierung der erzeugten PDF-Dateien in hochaufgelöste TIFF-Dateien (**PDF-zu-TIFF-Konvertierung**).
- 3- Georeferenzierung der TIFF-Dateien um einer Rasterebene (Raster layer) zu erzeugen (**Georeferenzierung der TIFF-Dateien**).
- 4- Konvertierung der Rasterebene in das SciGRID\_gas Format für PipeLines, Storages und Productions (**Generierung von SciGRID\_gas-Netzelementen**).
- 5- Entfernen kurzer Pipelines, von denen man annimmt, dass sie falsche Artefakte des PDF-zu-TIFF-Konvertierungsprozesses darstellen (**Entfernen von fehlerhaften Netzelementen**).
- 6- Zusammenführung von Netzelementen, die aus mehreren unverbundenen *PipeLines*, *Storages* und *Productions* bestehen (**Datenzusammenführung**).

---

<sup>20</sup> [https://www.entsog.eu/sites/default/files/2020-01/ENTSOG\\_CAP\\_2019\\_A0\\_1189x841\\_FULL\\_401.pdf](https://www.entsog.eu/sites/default/files/2020-01/ENTSOG_CAP_2019_A0_1189x841_FULL_401.pdf)

## 7- Verknüpfung von *PipeLines*, *Storages* und *Productions* zu einem einheitlichen SciGRID\_gas-Netz, dem **EMAP Datensatz (Erzeugung eines kohärenten SciGRID\_gas-Datensatzes)**.

Das Gesamtergebnis dieses Prozesses ist die Umwandlung der PDF-Karte in mehr als 3.000 *PipeSegments*-Elemente, mehr als 200 *Storages*-Elemente und mehr als 100 *Productions*-Elemente in ganz Europa, einschließlich Russland und anderen Nicht-EU-Staaten, was eine Gesamtlänge von mehr als 200.000 km an Leitungen ergibt (siehe Abbildung 7).



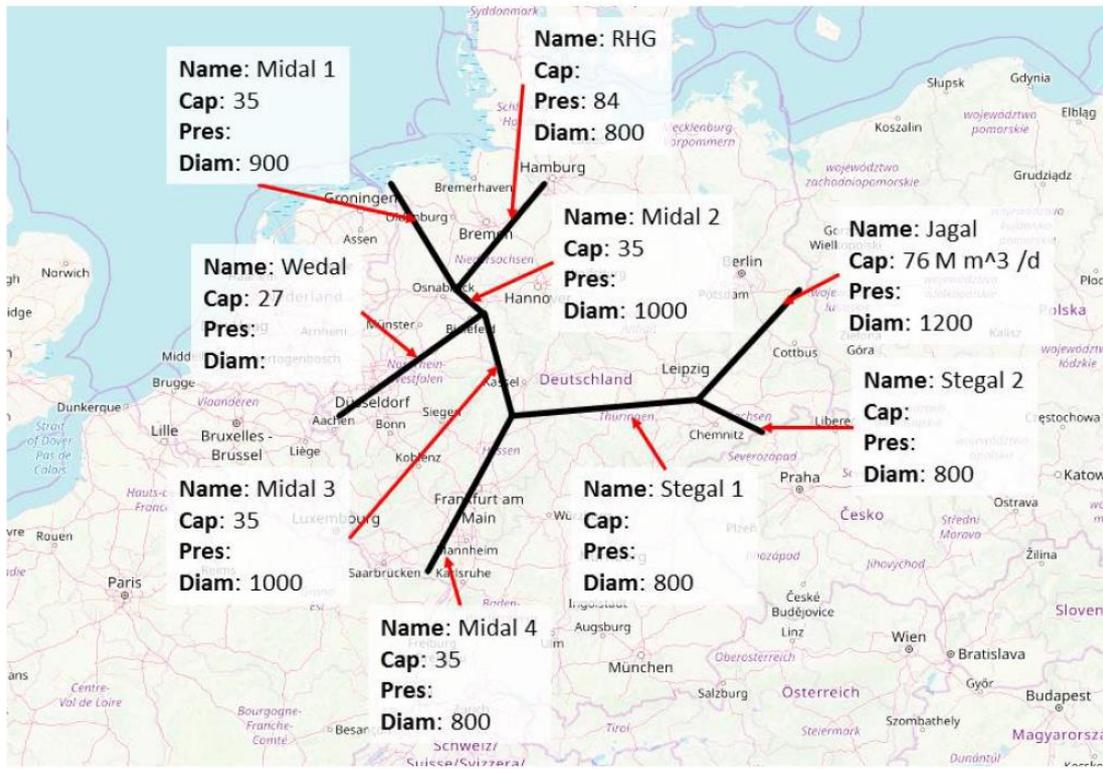
**Abbildung 7:** Visualisierung des ENTSO-G Datensatzes (EMAP). Die interaktiv programmierte Karte kann nach Land und Komponententyp visualisiert werden ([SciGRID\\_gas Webseite](#)).

### 2.4.2 Spezifizierung heuristischer Verfahren

Trotz der Zusammenführung mehrerer Datenquellen (siehe Kapitel 2.3.1) von Gasnetzelementen enthält der resultierende Datensatz eine große Anzahl fehlender Attributwerte. Das Problem der fehlenden Attributwerte wird anhand eines Beispieldatensatzes beschrieben, gefolgt von der Beschreibung der heuristischen Methoden, die zum Auffüllen fehlender Attributwerte verwendet werden. Ein Beispiel Gasnetzdatensatz mit fehlenden Attributwerten ist in Abbildung 8 und in Tabelle 1 dargestellt.

Wir definieren die Datendichte (*data density*) als das Verhältnis der Anzahl der verfügbaren Attributwerte zur Anzahl aller Elemente einer Komponente. Aus Tabelle 1 ergeben sich entsprechende Datendichten für die Attribute Kapazität, Druck und Durchmesser von 67 %, 22 % bzw. 89 %. Das Gesamtziel der Einsatz der heuristischen Methoden besteht darin, für alle Attribute eine Datendichte von 100 % zu erreichen.

In diesem Abschnitt wird beschrieben, wie die fehlenden Attributwerte durch verschiedene heuristische Methoden generiert werden könnten.



**Abbildung 8:** Beispiel Gasnetz mit fehlenden Werten für die Attribute Kapazität (Cap), Druck (Pres), und Durchmesser (Diam).

| Pipeline name | capacity [M m <sup>3</sup> d <sup>-1</sup> ] | pressure [bar] | diameter [mm] |
|---------------|--|----------------|---------------|
| Jagal         | 76   | 80             | 1200          |
| RHG           |  | 84             | 800           |
| Midal 1       | 40   |                | 900           |
| Midal 2       | 50   |                | 1000          |
| Midal 3       | 35   |                | 800           |
| Midal 4       | 34   |                | 800           |
| Stegal_1      |  |                | 800           |
| Stegal_2      |  |                | 900           |
| Wedal         | 27   |                |               |

**Tabelle 1:** Zusammenfassung der Elemente und Attributwerte des Gasnetzes in Abbildung 8.

### Mean/Median Methode

Für die Ermittlung der fehlenden Attributwerte gibt es mehrere Möglichkeiten. Eine einfache Lösung wäre die Verwendung des Durchschnitts oder Medians der Eingabewerte als Methode zur Schätzung eines fehlenden Wertes. Der Mittelwert und der Medianwert für den fehlenden Kapazitätswert des RHG-Pipeline-segments betragen hier 41 M m<sup>3</sup> d<sup>-1</sup> bzw. 35 M m<sup>3</sup> d<sup>-1</sup>. Die Auswahl des besten Ansatzes (Mittelwert oder Median) kann jedoch schwierig sein. Daher wird ein "Unsicherheitsbewertungsindikator" als Entscheidungskriterium verwendet, um den besten

heuristischen Ansatz zu bestimmen. Hier wird der Unsicherheitsbewertungsindikator als die Abweichung zwischen dem geschätzten und dem Eingangswert definiert. Bei der Verwendung heuristischer Methoden werden die verfügbaren Daten üblicherweise in einen Trainings- und einen Testdatensatz aufgeteilt. In der Regel wird einen 70/30-Ansatz gewählt, bei dem 70% der Daten in den Trainingsdatensatz und 30% in den Testdatensatz einfließen. Wir haben jedoch nur Zugang zu relativ kleinen Datensätzen, sodass eine Aufteilung des Datensatzes in einen Trainings- und einen Testdatensatz nicht zu einer verbesserten Attributwerteinschätzung führen würde. Daher wird im gesamten SciGRID\_gas-Projekt die "Leave-one-out"-Methode verwendet und der Fehler wird anhand des "mittleren absoluten Fehlers" (mean absolute error, MAE) berechnet, wobei der absolute Fehler die absolute Differenz zwischen einem einzelnen Eingabewert und der Schätzung dieses Wertes ist.

Wenn wir auf das Beispiel der RHG-Pipeline zurückgreifen, beträgt der MAE für die Mittelwert- und die Medianmethode  $25 \text{ m}^3 \text{ d}^{-1}$  bzw.  $16 \text{ m}^3 \text{ d}^{-1}$ . Auf der Grundlage des MAE wäre es daher am besten, die Median-Methode zu verwenden, und man könnte alle fehlenden Werte mit dem Wert von  $35 \text{ m}^3 \text{ d}^{-1}$  ergänzen, was einem MAE von  $16 \text{ m}^3 \text{ d}^{-1}$  entspricht. In diesem Fall aber ist die MAE allerdings sehr groß in Verhältnis zum tatsächlichen Kapazitätswert. Daher müssen auch andere Methodenansätze untersucht werden.

### **Lasso lineare Regression**

Ein Attribut kann auch mit anderen Attributen verbunden sein. So kann z. B. der maximale Druck der Pipeline mit dem Durchmesser der Pipeline in Verbindung gebracht werden.

Basierend auf der Datenstruktur aus AP 3.1 werden verschiedene heuristische Methoden und Verfahren auf deren Eignung für die Erstellung topologischer Gastransportnetzmodelle untersucht. Die heuristischen Methoden werden hinsichtlich ihrer zu implementierenden Rechenprozedur, benötigten Eingangsdaten und verwendeten Annahmen und Vereinfachungen getestet. Darüber hinaus werden das Wissen aus AP 2 und die aufgewendete Rechenzeit der Heuristiken zur Bereitstellung von Eingangsdaten Entscheidungsparameter bei der Auswahl einer geeigneten Methodik sein. Hier können Methoden wie die lineare Regression eingesetzt werden. Allerdings neigen lineare Regressionen dazu, die Daten der unabhängigen Variablen gleich zu gewichten. Dies kann zu einer schlechten Vorhersage der fehlenden Attributwerte führen, insbesondere dann, wenn die gewählten Attribute miteinander korreliert sind. Allerdings finden sich andere Regressionsmethoden, wie die Lasso-Lineare Regression und die Ridge-Regression, die dazu neigen, die unabhängigen Variablen ungleich zu gewichten. Wir verwenden hier die Lasso-Lineare Regression anstelle der einfachen linearen Regression, um fehlende Werte zu prognostizieren, wenn mehr als ein Attribut berücksichtigt wird.

Die Lasso-Lineare Regression wird auf das Attribut (Variable) Kapazität (auch "Prädiktor" oder "Regressionsinput" genannt) und das Attribut Durchmesser (die unabhängige Variable, auch "Feature"-Variable genannt) des vorherigen Beispiels der RHG-Pipeline angewandt. Im Beispiel fehlt der Kapazität-Attributwert, und es wird ein Durchmesserwert von 800 mm angegeben. Die lineare Lasso-Regression schätzt die Kapazität auf  $33,9 \text{ m}^3 \text{ d}^{-1}$  mit einem MAE von  $2,4 \text{ m}^3 \text{ d}^{-1}$ . Wie zu erkennen ist, ist der MAE der Lasso-Methode im Vergleich zum MAE der Mean- und der Median-Methode deutlich geringer.

Das oben beschriebene Verfahren wurde auf die in Tabelle 1 dargestellten Beispieldaten angewandt. In Tabelle 2 und Tabelle 3 sind die Eingabewerte, die geschätzten Attributwerte, die Prognosemethode und der entsprechende MAE-Wert auf der Grundlage des "Leave-one-out"-Ansatzes zusammengefasst.

| Pipeline name | Input <i>capacity</i> | Estimated <i>capacity</i> | Method | Uncertainty |
|---------------|-----------------------|---------------------------|--------|-------------|
| Jagal         | 76                    | 69.7                      | Lasso  | 2.4         |
| RHG           |                       | 33.9                      | Lasso  | 2.4         |
| Midal 1       | 40                    | 42                        | Lasso  | 2.4         |
| Midal 2       | 50                    | 51.8                      | Lasso  | 2.4         |
| Midal 3       | 35                    | 33.9                      | Lasso  | 2.4         |
| Midal 4       | 34                    | 33.9                      | Lasso  | 2.4         |
| Stegal_1      |                       | 33.9                      | Lasso  | 2.4         |
| Stegal_2      |                       | 42.8                      | Lasso  | 2.4         |
| Wedal         | 27                    | 43.7                      | Mean   | 12.9        |

**Tabelle 2:** Eingabewerte (input capacity), die geschätzten Attributwerte der Kapazitätsattribut (estimated capacity), die Prognosemethode und der entsprechende MAE-Wert (uncertainty) für das Beispiel in Tabelle 1. Die Attributwerte werden in der Einheit [m<sup>3</sup> d<sup>-1</sup>] angegeben.

| Pipeline name | Input <i>diameter</i> | Estimated <i>Diameter</i> | Method | Uncertainty |
|---------------|-----------------------|---------------------------|--------|-------------|
| Jagal         | 1200                  | 1233                      | Lasso  | 23          |
| RHG           | 800                   | 900                       | Mean   | 100         |
| Midal 1       | 900                   | 875                       | Lasso  | 23          |
| Midal 2       | 1000                  | 975                       | Lasso  | 23          |
| Midal 3       | 800                   | 826                       | Lasso  | 23          |
| Midal 4       | 800                   | 816                       | Lasso  | 23          |
| Stegal_1      | 800                   | 900                       | Mean   | 100         |
| Stegal_2      | 900                   | 900                       | Mean   | 100         |
| Wedal         |                       | 746                       | Lasso  | 23          |

**Tabelle 3:** Eingabewerte (input capacity), die geschätzten Attributwerte der Durchmesserattribute (estimated diameter), die Prognosemethode und der entsprechende MAE-Wert (uncertainty) für das Beispiel in Tabelle 1. Die Attributwerte werden in der Einheit [mm] angegeben.

Es ist deutlich, dass die geschätzten Werte bei der Lasso-Methode besser mit den Eingangsdaten übereinstimmen als bei der Mean-Methode. Allerdings konnten nicht alle Attributwerte mit der Lasso-Methode geschätzt werden, da Attributwerte des unabhängigen Attributs fehlten (z. B. der Durchmesser für die Pipeline *Stegal\_2*). Daher wurde für einige fehlende Werte stattdessen die Mean- oder die Median-Methode verwendet (z. B. Kapazität für die Leitung "*Wedal*").

### 2.4.3 Implementierung heuristischer Verfahren

In diesem Abschnitt wird der Arbeitsablauf der im Projekt SciGRID\_gas verwendeten und angewendeten heuristischen Methoden dargestellt. Die Implementierung der heuristischen Verfahren ist in der SciGRID\_gas Dokumentation detailliert beschrieben. Der Arbeitsablauf besteht aus fünf Schritten, die im Folgenden kurz beschrieben werden.

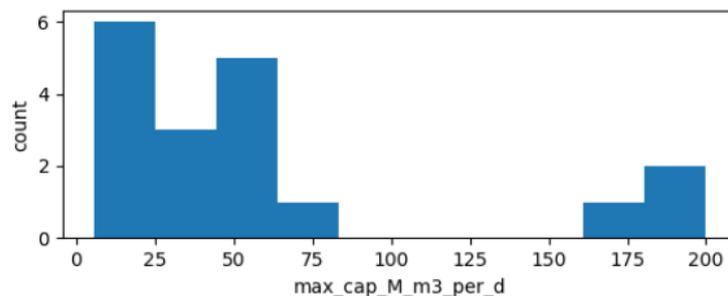
1. Der erste Schritt besteht darin, den verfügbaren Datensatz, der bearbeitet werden soll, hochzuladen.
2. **Anpassung der Konfiguration der erforderlichen Konfigurationsdateien:** In diesem Schritt werden die drei erforderlichen csv-Skripte vorbereitet.
  - I. Das erste Skript **Copy\_Attribs.csv** ermöglicht es dem Benutzer, Attributwerte von einem Element in ein anderes Element zu kopieren, wobei die Elemente

unterschiedliche Komponententypen haben. Das Skript ermöglicht die Übernahme von Attributwerten von einem Element in ein anderes (aus verschiedenen Komponenten), sofern die Elemente physisch miteinander verbunden sind. Zum Beispiel kann das Attribut *max\_power\_MW* des *Compressors* für ein *PipeSegment* übernommen werden, wenn die beiden mit einander verbunden sind (d.h. der *Compressor* und *PipeSegment* haben die gleiche *node\_id*).

- II. Das zweite Skript ***StatsMethodsSettings.csv*** enthält Metadaten für jede heuristische Methode (z. B. Mean, Median). Das Skript enthält auch die möglichen Parameter, die von den heuristischen Methoden verwendet werden. Zurzeit benötigt nur die LogisticReg (Logistische Regression) Methode die Angabe des zu verwendenden Solvers (z.B. Ibfgs). Alle anderen Methoden (Mean, Median, Ordinary Least Squares, Max, Min) benötigen keine zusätzlichen Parametereinstellungen.
- III. Das dritte Skript ***StatsAttribSettings.csv*** enthält eine Liste von Attributen für die gewünschte Komponente einschließlich attributspezifischer Metadaten. Der Benutzer wählt die Attribute aus (z. B. *max\_cap\_M\_m3\_per\_d*), die bei der Anwendung der heuristischen Methode aktualisiert werden sollen.

Die oben genannten Skripte enthalten alle Informationen, die für die Generierung von Attributwerten erforderlich sind. Der Benutzer kann somit unterschiedliche heuristische Methoden und Einstellungen ausprobieren, um verschiedene Optionen zu testen.

3. **Bewertung der Datenqualität:** Vor dem Einsatz heuristischer Methoden zur Generierung fehlender Attributwerte ist es notwendig, eine Vorstellung von der Qualität und Quantität der verfügbaren Daten zu haben. Daher müssen die Daten, die als Input für den heuristischen Prozess verwendet werden sollen, geprüft werden. Dies erfolgt durch die Python-Funktion ***M\_Stats.gen\_DataHists***, die speziell für die Überprüfung der Datenqualität entwickelt wurde. Die Funktion *M\_Stats.gen\_DataHist* benötigt die folgenden Eingaben: *Netz* (die Kopie des Gasnetzes mit den fehlenden Attributwerte), *CompNames* (die Liste der zu visualisierenden Komponenten), *AttribNames* (die Liste der zu visualisierenden Attribute), *StatsInputDirName* (der relative Verzeichnispfad zum Skript *StatsAttribsSettings.csv*) und *DataStatsOutput* (der relative Verzeichnispfad, in dem die Diagramme gespeichert werden sollen). Ein Beispiel für die von der *M\_Stats.gen\_DataHists* Funktion erzeugten Diagramme ist in Abbildung 9 dargestellt.

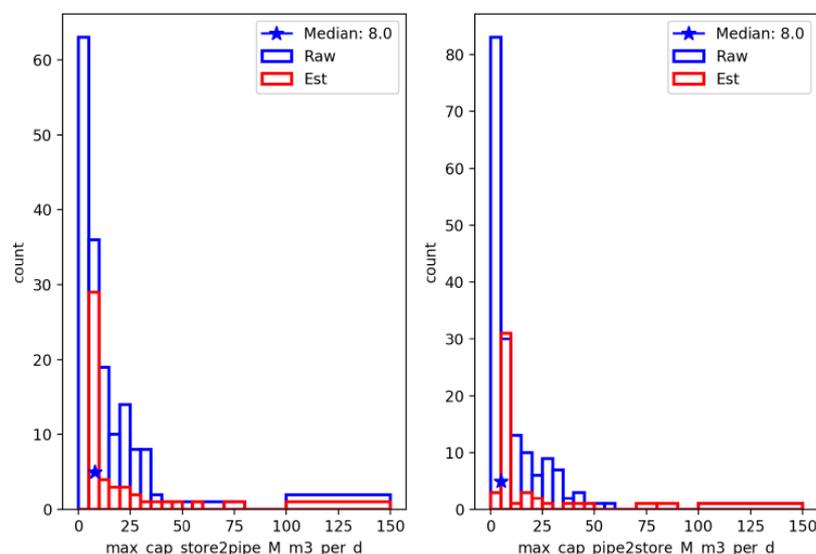


**Abbildung 9:** Beispiel für eine Histogramm-Darstellung des Kompressor-Attributs *max\_cap\_M\_m3\_per\_d*.

4. **Auswahl der heuristischen Methode:** Nach der Anwendung der Funktion `M_Stats.gen_StatsParam` mit verschiedenen heuristischen Methoden und Parametern und der Überprüfung der resultierenden Diagramme (als Prädiktor-Schritt bezeichnet) ist es erforderlich, eine geeignete heuristische Methode auszuwählen, um den fehlenden Attributwert zu ergänzen (als Korrektor-Schritt bezeichnet). Anhand der resultierenden Diagramme und Tabellen kann der Benutzer eine informierte Entscheidung über die geeignete Methode zur Erzeugung fehlender Attributwerte treffen. Es ist zu beachten, dass eine Kombination von verschiedenen Methoden verwendet werden können, um fehlende Attributwerte für verschiedene Komponententypen zu erzeugen.
5. **Generierung der fehlenden Attributwerte:** Die eigentliche Bestimmung und Generierung von Attributwerten wird mit der Funktion `M_Stats.pop_Attribs` durchgeführt. Die Ausgabe dieser Funktion ist ein Komponentendatensatz, in dem alle fehlenden Attributwerte generiert wurden.

#### 2.4.4 Aktualisierung der (Datenbank und) Datensätze

Die erfassten Datensätze (siehe Kapitel 2.3) werden auf fehlende Attributwerte überprüft. Die in Kapitel 2.4.2 vorgestellten Methoden werden verwendet, um die fehlenden Attributwerte zu ermitteln. Dieses Verfahren wird für alle Attributwerte angewandt, mit Ausnahme der Werte, die aufgrund fehlender Informationen oder zu hoher Unsicherheiten nicht aktualisiert werden können. Die resultierenden Datensätze sind aktualisierte Datensätze, die entweder keine oder eine minimale Anzahl fehlender Werte enthalten. Im Folgenden stellen wir ein Beispiel für einen aktualisierten Datensatz vor, der Attributwerte enthält, die mit heuristischen Methoden ermittelt wurden.



**Abbildung 10:** Histogramm der unverarbeiteten (blau) und geschätzten (rot) Werte für `max_cap_store2pipe_M_m3_per_d` (links) und `max_cap_pipe2store_M_m3_per_d` (rechts) der Komponente Storages. Beide Teilgrafiken zeigen auch die Lage des Medianwertes für die unverarbeiteten Werte (Stern).

Als Beispiel wird das Ergebnis der Generierung fehlender Werte für die Attribute `max_cap_pipe2store_M_m3_per_d` und `max_cap_store2store_M_m3_per_d` der Komponente

Speicher für den kombinierten IGG-Datensatz dargestellt. Insgesamt enthält der Datensatz 216 Stages-Elemente. Für beide Attribute fehlten 48 Werte. Nach der Schätzung der Attributwerte betrug der Mean (Mittelwert) der Attribute `max_cap_pipe2store_M_m3_per_d` und `max_cap_store2pipe_M_m3_per_d` 13,9 bzw. 14,6. Der Mean Absolute Error M (mittlere absolute Fehler) beträgt für `max_cap_pipe2store_M_m3_per_d` und `max_cap_store2pipe_M_m3_per_d` 10,9 bzw. 11,2. Diese erscheint groß im Vergleich zum Mean-Wert. Allerdings liegen die einzelnen Attributwerte für beide Attribute zwischen 0 und 100. Ein Histogramm der unverarbeiteten und der geschätzten Attributwerte, dargestellt in Abbildung 10, zeigt die Verteilung für beide Attribute.

Die geschätzten Attributwerte haben in etwa die gleiche Verteilung wie die unverarbeiteten Attributwerte. Die Ausnahmen sind die größeren Balken, die den Median des unverarbeiteten Datensatzes enthalten. Der Median wird verwendet, wenn es keine andere Möglichkeit gibt einen fehlenden Wert zu ermitteln.

#### **2.4.5 Zusammenführung von OSM und nicht-OSM Datensätze**

Im Gegensatz zur Ursprünglichen Projektbeschreibung wurden OSM Daten nur zu Datenverifikation benutzt und die Verifikationsdaten aus dem EMAP Datensatz mit den nicht-OSM Datensätzen zusammengeführt. Die Gründe die zu dieser Entscheidung führten waren die folgenden:

- 1) OSM-Daten haben den Nachteil, dass sie unter OdBL lizenziert sind. Diese ist nicht mit der CC-BY-Lizenz (für die Non-OSM-Datensätze) kompatibel<sup>21</sup>.
- 2) Die unvollständige Datenlage von OSM macht diese Daten zum gegenwärtigen Zeitpunkt höchsten als Grundlage für Pipeline Routen interessant.
- 3) Der EMAP Pipeline Datensatz welcher aus der ENTSO-G Karte erstellt wurde (siehe Kapitel 2.4), ist geographisch deutlich weniger exakt als die OSM Pipelines, allerdings ist er vollständig, weshalb dieser als Pipeline Routen Datenbasis für das Datenmodell herangezogen wurde. Insbesondere wird zur Erstellung von Datenmodellen in der Energiesystemanalyse die Exaktheit des Pipelineverlaufs als weniger entscheidend angesehen als die Vollständigkeit der vorliegenden Pipeline Routen.
- 4) Die juristische Prüfung der Extrahierten ENTSO-G Karten Daten zeigte, dass diese problemlos unter einer open-source Lizenz veröffentlicht und damit auch zur Zusammenführung verwendet werden dürfen.

Die entsprechende Zusammenführung der OSM-Daten wurde zusammenfassend nicht vollzogen, weil der EMAP Datensatz zum gegenwärtigen Zeitpunkt dieselbe Aufgabe besser erfüllt.

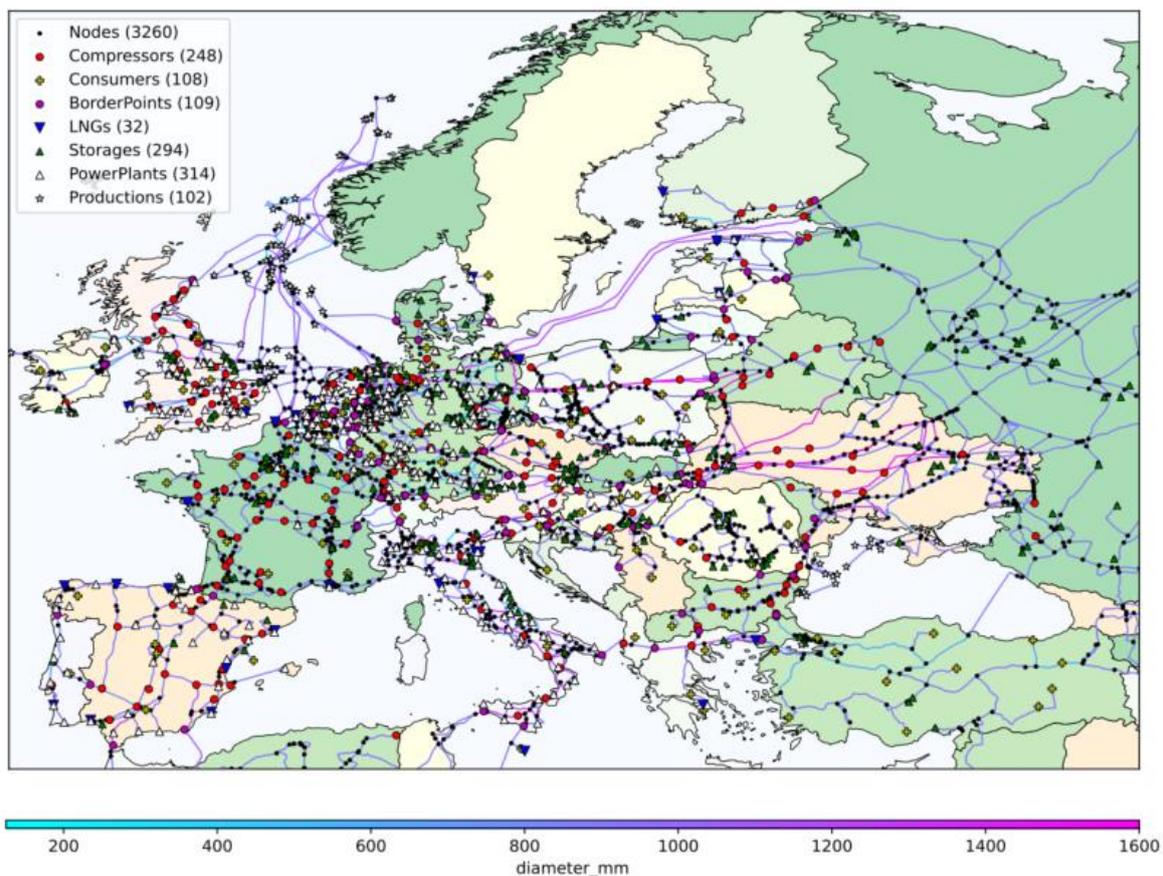
---

<sup>21</sup> Eine Vermischung der Daten ist nach ausgiebiger juristischer Recherchearbeit prinzipiell möglich, falls jedes einzelne Objekt in der Datenbank seine eigene Lizenz bekommt.

## Finaler SciGRID\_gas Datensatz: IGGIELGNC1

Der IGGIELGNC1-Datensatz<sup>22</sup> (siehe Abbildung 11) ist eine Sammlung von Open-Source-Daten zum europäischen Gastransportnetz, die im SciGRID\_gas-Projekt erstellt wurde. Der Datensatz enthält geografische sowie Metainformationen (Attribute und deren Werte) über das Gastransportnetz und dessen Komponenten. Der Datensatz resultiert aus der Zusammenführung der INET, GIE, GSE, IGU, EMAP (siehe Kapitel 2.4 und 2.4.1 für die Beschreibung der EMAP-Datensätze), LKD, GB und NO Datensätze. Es wurden heuristische Methoden zum Ergänzen fehlender Attributwerte auf die Rohdaten angewandt. IGGIELGN stellt 241.000 km des europäischen Transportleitungsnetzes dar und enthält die Komponenten Speicher (STORAGES), Erzeugung (PRODUKTIONS), LNG-Terminals, Verdichterstationen (COMPRESSORS) und Verbrauchern (CONSUMERS) auf NUTS1. Der Datensatz ist ebenfalls als IGGIELGNC2 und IGGIELGNC3, angepasst auf NUTS2 and NUTS3 Verbraucher, verfügbar.

Die einzelnen Prozessierungsschritte sind kompakt in einer Veröffentlichung zum SciGRID\_gas Datensatz beschrieben (Pluta, 2022) und in ausführlicher Form in der jedem Datensatz beigefügten Dokumentation.



**Abbildung 11:** Visualisierung des IGGIELGNC1-Datensatzes für das europäische Gastransportnetz.

<sup>22</sup> <https://zenodo.org/record/5509988>

## **2.5 Entwicklung eines Open-Source-Modells**

### **2.5.1 Definition des topologischen Gastransportnetzmodells**

Die Definition des Gastransportnetzmodells wurde in der SciGRID\_gas Journalveröffentlichung (Pluta, 2022) kompakt zusammengefasst. Das SciGRID\_gas Datenmodellnetzwerk besteht aus mehreren Komponentenklassen, die jeweils eine Liste von Objekten repräsentieren. Folgende Komponentenklassen wurden implementiert: PIPESEGMENTS, BORDERPOINTS, COMPRESSORS, LNGS, POWERPLANTS, PRODUCTIONS, CONSUMERS, STORAGES und in ausführlicher Form in der jedem SciGRID\_gas-Datensatz beigefügten Dokumentation). Jedes Objekt, das Mitglied einer Komponentenklasse ist, wird als Element der jeweiligen Klasse definiert und kann daher durch einen gemeinsamen komponentenspezifischen Satz von Attributen beschrieben werden. Jedes Element verfügt über drei sogenannte Dictionaries. Das Erste beinhaltet den vollständigen Satz an Parametern, das Zweite beinhaltet die zugehörigen Fehler und das Dritte die zugehörigen Quellen und Prozessierungsschritte für jeden Parameter.

Um den Überblick über die Datenverarbeitungsschritte zu behalten, haben wir auch Parameterinformationen für jedes Attribut eines bestimmten Elements aufgenommen. Um diese Datenstruktur für Gasnetze geeignet zu machen, müssen wir die Daten mit Hilfe von Knoten und Kanten umstrukturieren. Diese sind über eine eindeutige ID mit den Elementen in unserem Datensatz verbunden. Auf diese Weise werden alle Komponenten mit Ausnahme von PIPESEGMENTS als Knoten implementiert. PIPESEGMENTS sind jedoch auch an einen Start- und Endknoten gebunden, so dass sie als Kanten implementiert wurden. Pipeline-Zwischenpunkte, die den geografischen Verlauf von PIPESEGMENTS widerspiegeln, werden in separaten Listen gespeichert.

### **2.5.2 Erstellung des topologischen Gastransportnetzmodells**

In diesem Schritt wird das Gastransportnetzmodell implementiert. Zusätzlich wird der Abstraktionsprozess implementiert, sodass ausschließlich relevante Informationen zum Netzwerk und dessen Komponenten in dem resultierenden Datensatz vorhanden sind. Weitere Methoden zur Aktualisierung der relationalen Datenbank werden erarbeitet.

### **2.5.3 Aktualisierung des Gastransportnetzmodells**

Der auf den heuristischen Methoden basierende Gasnetzdatensatz wird in das SciGRID\_gas Modell integriert. Zu diesem Projektzeitpunkt kann das SciGRID\_gas Modell alle verfügbaren Daten der relationalen Datenbank verarbeiten. Auf der Projektwebseite werden die verwendeten Eingangsdaten, das Modell, das Gastransportnetz und die zugehörigen Dokumentationen zur Verfügung gestellt.

### **2.5.4 Sensitivitätsanalyse des Gastransportnetzmodells**

Im AP 4.4 wird der Einfluss der Annahmen und Vereinfachungen der heuristischen Methode (AP 3.5) auf den resultierenden Datensatz für das Gastransportnetz untersucht. Das Ziel ist dabei eine qualitative Einschätzung des Einflusses von verschiedenen Parametern der heuristischen Methode auf den resultierenden Datensatz zu ermitteln. Die Ergebnisse dieser Sensitivitätsanalyse werden analysiert, anschließend dokumentiert und auf der Projektwebseite zur Verfügung gestellt.

## 2.6 Vergleich und Validierung

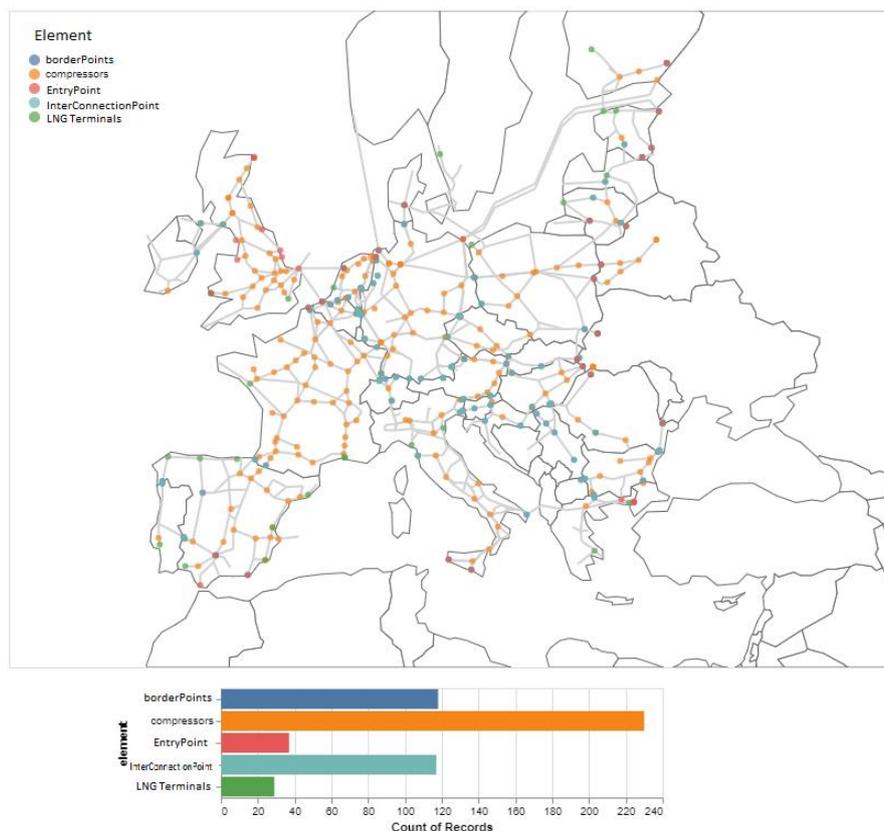
### 2.6.1 Entwicklung von mathematischen Methoden zur Charakterisierung und zum Vergleich von Gastransportnetzmodellen

Aufgrund des Mangels von geeigneten Validierungsdaten auf Parameterebene war eine Charakterisierung verschiedener Gastransportnetzmodelle in der hier angedachten Form mittels mathematischer Methoden nicht möglich. Stattdessen wurde der visuellen Charakterisierung der Daten eine höhere Wichtigkeit zugesprochen, was letztendlich zu der Entwicklung der wissenschaftlichen Analyse- und Visualisierungsbibliothek **qplot**<sup>23</sup> führte, deren Methodenumfang weit über das ursprüngliche Konzept hinaus ging.

### 2.6.2 Entwicklung von graphischen Werkzeugen zum Vergleich von Gastransportnetzmodellen

Die Visualisierung der SciGRID\_gas Daten wurde auf mehreren Ebenen realisiert:

- 1) Auf der Homepage wurden Interactive Plots erstellt und die Eingangsdatensätze erstellt, die es erlauben sich einen sehr schnellen Überblick über die Daten zu verschaffen und das Filtern nach verschiedenen Komponenten und Ländern zu ermöglichen (siehe Abbildung 12 für das Beispiel der INET-Filled-Datensatz)



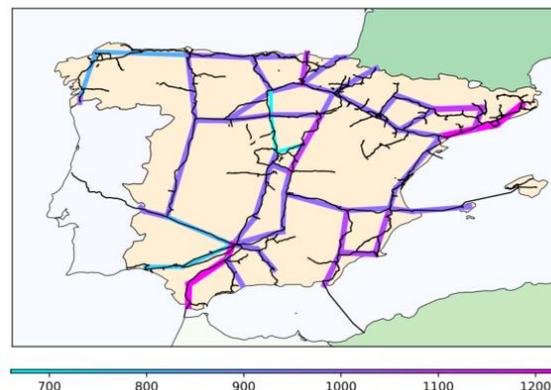
**Abbildung 12:** Visualisierung des INET-filled-Datensatzes auf der Projektwebseite.

<sup>23</sup> Qplot <https://gitlab.com/dlr-ve-esy/qplot>

- 2) Darüber hinaus lässt sich der Datensatz aufgrund seines hervorragend kompatiblen GeoJSON Formates per „Drag & Drop“ auf jeder GIS Anwendung und auf der Website [geojson.io](http://geojson.io) einfach und bequem darstellen.
- 3) Mit der `qplot`<sup>23</sup> Library wurde eine Python Visualisierungsbibliothek für SciGRID\_gas Daten auf PIPY<sup>24</sup> veröffentlicht. Das Repository enthält darüber hinaus eine ausführliche Dokumentation mit diversen Beispielen zur Visualisierung von SciGRID\_gas Daten für wissenschaftliche Auswertungen. Diese Beispiele beinhalten Optionen wie, eine Hintergrundkarte für Europa, das Übereinanderlegen von verschiedenen Datensätze, die Visualisierung von Paralleleitungen, die Visualisierung von einzelnen Parametern (Durchmesser, Druck, etc.) durch Farben und/oder Liniendicken und sogar die Möglichkeit einzelne Komponenten anzuklicken um sich ihre Parameter auflisten zu lassen. **Abbildung 11** zeigt ein Beispiel für die Darstellung von SciGRID\_gas Daten mit der `qplot` Library.

### 2.6.3 Validierung von SciGRID\_gas

Die Validierung der Daten erwies sich wie schon dargestellt als äußerst schwierig da geeignete Validierungsdaten fehlten. Jedoch konnten die nicht im Model verwendeten OSM Daten für einen topologischen Vergleich der Pipeline Routen herangezogen werden. Für solch einen Vergleich wurde jedoch ein visuelles Verfahren als geeignet bewertet. Hierfür wurden die beschriebene ***qplot*** Library verwendet, um die verschiedenen Pipeline Netzwerke aus unterschiedlichen Datensätzen miteinander zu vergleichen. Insbesondere konnte gezeigt werden, dass die geographische Lage der Pipelines aus dem EMAP Datensatz eine sehr gute Approximation der genauen Positionen aus dem OSM Datensatz darstellen (siehe **Abbildung 13**).



**Abbildung 13:** Topologischen Verifizierung von Pipelinedaten. In diesem Beispiel werden OSM Pipelines (schwarz) und INET Pipelines (koloriert nach Durchmesser in mm) visuell miteinander verglichen.

Hiermit konnte die Vermutung, dass der EMAP Datensatz alle europäischen Pipelines vollständig enthält mit den OSM Daten nicht widerlegt werden. Dennoch enthält der OSM-Datensatz aufgrund

<sup>24</sup> <https://pypi.org/project/qplot/>

erheblicher Datenlücken in einigen Regionen derzeit nur etwa 45 % aller Pipelines. Außerdem kann diese Methode nicht zur Validierung paralleler Pipelines verwendet werden, da diese in OSM nicht explizit angegeben sind.

Das SciGRID\_gas Datenmodell deckt eine Gesamtlänge von 237.000 km ab, was in etwa mit der angenommenen Länge von 200.000 km übereinstimmt (Connolly, 2014). Die leichte Überschätzung ist eine unmittelbare Folge unserer umfassenderen Definition des europäischen Gasnetzes, die Westrussland, die nordafrikanischen Länder und die Türkei einschließt. Die Gesamtlänge ist ein sehr guter Indikator für den Gesamterfolg der Modellierung des europäischen Gastransportnetzes unter ausschließlicher Verwendung frei verfügbarer Datensätze.

Da sich aufgrund der fehlenden Referenzdaten die Validierung der Komponentenparameter nicht validieren lässt, war es uns wichtig genau anzugeben wie viele der Parameter aus Originaldaten stammen. Aus diesem Grund haben wir es auch vermieden, unsere Methoden zur Attributgenerierung zu evaluieren, die in der Dokumentation des Datensatzes ausführlicher beschrieben sind. Wir können jedoch eine Aussage zur potentiellen Genauigkeit unsere Methoden machen. Da die Ergebnisse jeder Methode zur Generierung von Attributen, die für die Vorhersage von teilweise unbekanntem Daten entwickelt wurde, skalieren in ihrer Genauigkeit mit dem Prozentsatz der bekannten Daten. Wir haben unsere Daten hinsichtlich wichtiger Parameter verschiedener Netzwerkkomponenten analysiert. Wir können daher feststellen, dass die generierten Attribute für LNGS, STORAGES und POWERPLANTS vertrauenswürdiger sind, als für PIPESEGMENTS, COMPRESSORS und PRODUCTIONS (siehe Tabelle 4).

| <b>component</b> | <b>density</b>  | <b>density</b>  | <b>density</b>  |
|------------------|-----------------|-----------------|-----------------|
| PIPESEGMENTS     | capacity<br>13% | diameter<br>32% | pressure<br>19% |
| LNGS             | capacity<br>94% | size<br>69%     | -               |
| STORAGES         | capacity<br>63% | power<br>28%    | pressure<br>35% |
| COMPRESSORS      | capacity<br>7%  | power<br>15%    | pressure<br>7%  |
| POWERPLANTS      | energy<br>100%  | -               | -               |
| PRODUCTIONS      | supply<br>5%    | -               | -               |

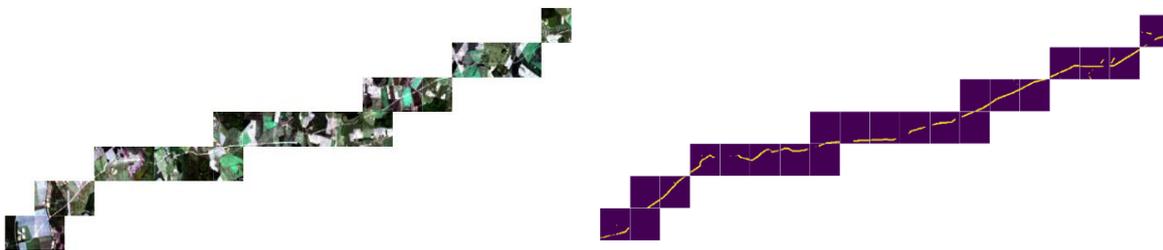
**Tabelle 4:** Original Datendichte für die wichtigsten Parameter von ausgewählten Komponenten, IGGIELGNC-1 Datensatz.

Aus unserer Sicht muss mehr Aufmerksamkeit auf die Datenerfassung für diese Komponenten gelegt werden.

Da die Validierungsmöglichkeiten der SciGRID\_gas Daten zurzeit begrenzt sehr sind, wurde dieses Arbeitspaket außerdem genutzt um ein Verfahren zu entwickeln, welches die topologische Validierung des Europäischen Gas Transportnetzes auch mit den bisher nicht in OSM eingetragenen Pipelines ermöglichen könnte. Zu diesem Zwecke wurde der Einsatz von Satellitendaten und die Verwendung von Machine-Learning Algorithmen erforscht. Dies geschah im Rahmen von zwei Masterarbeiten und einer wissenschaftlichen Veröffentlichung:

- “Pipeline Detection with Satellite Images Using Machine Learning” (Dasenbrock, 2020)<sup>25</sup>
- “Pipeline Detection Using Uncertainty-Driven Machine Learning” (Tetens, 2021)<sup>26</sup>
- “Detecting Pipeline Pathways in Landsat 5 Satellite Images with Deep Learning “(Dasenbrock, Pluta, Zech, Medjroubi,2021)<sup>27</sup>

Anhand von Testbildern aus Großbritannien und Deutschland konnte gezeigt werden, dass sich die Bauschneisen auf historischen Satellitendaten dazu eigenen Pipelineverläufe mit Hilfe von trainierten Neuronalen Netzen zu detektieren (siehe Abbildung 14). Ebenfalls konnte gezeigt werden, dass sich dieses Verfahren auch Prinzipiell auf andere Region von Europa anwenden lässt. Eine der vermutlich wenigen Ausnahmen stellt hierbei jedoch Spanien dar, das aufgrund seiner Topografie nur einen sehr geringen Kontrast zwischen Bauschneisen und Umgebung besitzt.



**Abbildung 14:** Das linke Bild zeigt einen Abschnitt der NEL Pipeline die mit historischen Satellitendaten aus 22 Bilder erstellt wurde. Das rechte Bild zeigt die Pipelineline die aus der Bildanalyse des trainierten Neuronalen Netzwerks entstanden ist. “ (Dasenbrock, Pluta, Zech, Medjroubi, 2021)

Die Anwendung des beschriebenen Verfahrens auf die Gesamtfläche Europas ließ sich aufgrund des hohen Zeitaufwandes hinsichtlich der Erstellung von manuellen Trainingsdaten nicht zur Laufzeit des Projektes realisieren. Es würde jedoch einen hohen Mehrwert zur Verifizierung von Pipelinelines darstellen und könnte auch genutzt werden um fehlende Pipelines in OSM zu hinterlegen. Die Veröffentlichten Arbeiten geben einen Ausblick welche weiteren Arbeitsschritte notwendig wären um das Verfahren weiter zu automatisieren und auf ganz Europa anzuwenden.

## 2.7 Schnittstellenbildung zu Softwaretools von Open-Source-Anbietern

Zur Schnittstellenbildung zwischen SciGRID\_gas Daten wurden die Entwickler von Open-Source-Gasmodellierungssoftware befragt. Die Nutzung von SciGRID\_gas Daten konnte hieraufhin durch die Konvertierung in ein Standard-Geodatenformat sichergestellt werden. Hierzu wurde ein Datenkonverter geschrieben, der es erlaubt die SciGRID\_gas Daten in das GeoJSON Format zu konvertieren. GeoJSON ist ein universell einsetzbares geographisches Daten Format, welches sich mühelos in jedes beliebige andere geographische Datenformat umwandeln lässt. Die Daten können zum

<sup>25</sup> <https://elib.dlr.de/137734/>

<sup>26</sup> <https://elib.dlr.de/148468/>

<sup>27</sup> *Energies* **2021**, 14(18), 5642; <https://doi.org/10.3390/en14185642>

Beispiel einfach per „Drag & Drop“ in jeder beliebigen GIS Applikation eingelesen und visualisiert werden. Das Einlesen und Konvertieren der Daten für jede beliebigen Gasmodellierungssoftware ist damit für jeden Anwender zu einer Routinearbeit geworden.

Unabhängig von uns haben Entwicklern und Anwendern von pandapipes<sup>28</sup> und PyPSA<sup>3</sup> bereits eigenständig für unsere ersten Testdatensätze ihren eigenen SciGRID\_gas Datenkonverter geschrieben. In der Folge lässt sich sagen, dass Entwicklung des GeoJSON-Konvertierungstools und die frühe Veröffentlichung von Testdatensätze ausreichend waren, um die Nutzung der Daten sicherzustellen und extrem zu vereinfachen. Alle folgenden Unterarbeitspakete können in diesem Zusammenhang als erfüllt angesehen werden:

### **2.7.1 Identifikation von geeigneten Schnittstellen**

### **2.7.2 Entwicklung von Konvertierungstools**

### **2.7.3 Implementierung der Konvertierungsmethoden**

## **3 Verwertung der Ergebnisse**

Die erzielten Ergebnisse innerhalb des Verbundprojekts wurden vom DLR-Institut für Vernetzte Energiesysteme durch diverse Veröffentlichungen publiziert. Eine Übersicht ist Kapitel 5 zu entnehmen.

## **4 Fortschritte auf dem Gebiet bei anderen Stellen**

Fortschritte auf dem Gebiet des Vorhabens bei anderen Stellen sind dem DLR-Institut für Vernetzte Energiesysteme in Bezug auf die Erstellung und Bereitstellung eines offenen europäischen Gastransportmodells zum aktuellen Zeitpunkt nicht bekannt.

## **5 Veröffentlichungen**

Die wissenschaftlichen Ergebnisse aus dem SciGRID\_gas Projekt konnten vom DLR-Institut für Vernetzte Energiesysteme bereits innerhalb der Projektlaufzeit veröffentlicht werden. Eine exemplarische Übersicht ist der folgenden Aufstellung zu entnehmen:

### **Journalveröffentlichungen:**

- Pluta, A., Medjroubi, W., Diettrich, Jan., Dasenbrock, J., Tetens, H.P., Sandoval, J.E. **SciGRID\_gas – Data Model for the European Gas Transport Network**. In IEEE eingereicht (2.12.2021).

---

<sup>28</sup> <https://www.pandapipes.org/>

- Dasenbrock, J., Pluta, A., Zech, M., and Medjroubi, W. (2021). **Detecting Pipeline Pathways in Landsat 5 Satellite Images With Deep Learning**. Energies. Multidisciplinary Digital Publishing Institute (MDPI). ISSN 1996-1073.
- Pluta, A. und Lündsdorf, O. (2020). **esy-osmfilter A Python Library to Efficiently Extract OpenStreetMap Data**. Journal of Open Research Software, 8 (19). Ubiquity Press. doi: 10.5334/jors.317. ISSN 2049-9647.

#### Studentische Arbeiten:

- Tetens, Hendrik-Pieter (2022). **Pipeline Detection Using Uncertainty-Driven Machine Learning**. Masterarbeit, Carl von Ossietzky Universität Oldenburg.
- Sandoval, Javier Enrique (2021). **Estimation and Simulation of a Gas Demand Time Series for the European NUTS 3 Regions**. Masterarbeit, Carl von Ossietzky Universität Oldenburg.
- Dasenbrock, Jan (2020). **Pipeline Detection with Satellite Images Using Machine Learning**. Masterarbeit, Carl von Ossietzky Universität Oldenburg.

#### Software:

- **Qplot** (2021), Gitlab repository, <https://gitlab.com/dlr-ve-esy/qplot>
- **osmscigrd** (2021), Gitlab repository, <https://gitlab.com/dlr-ve-esy/osmscigrd>
- **esy-osmfilter** (2020), Gitlab repository, <https://gitlab.com/dlr-ve-esy/esy-osmfilter>

#### Datensätze:

- Diettrich, J., Pluta, A., Medjroubi, W., Dasenbrock, J., und Sandoval, J.E. (2021). SciGRID\_gas IGGIELGNC-1 (0.2) Zenodo. <https://doi.org/10.5281/zenodo.5509988>
- Diettrich, J., Pluta, A., Medjroubi, W., Dasenbrock, J., und Sandoval, J.E. (2021). SciGRID\_gas IGGIELGNC-2 (0.2) Zenodo. <https://doi.org/10.5281/zenodo.5509990>
- Sandoval, J.E., Diettrich, J., Pluta, A., und Medjroubi, W. (2021). SciGRID\_gas Cons (0.4) Zenodo. <https://doi.org/10.5281/zenodo.5529710>
- Diettrich, J., Pluta, A., Medjroubi, W., Dasenbrock, J., und Sandoval, J.E. (2021). SciGRID\_gas IGGIELGNC (0.1). Zenodo. <https://doi.org/10.5281/zenodo.5079748>
- Diettrich, J., Pluta, A., und Medjroubi, W. (2021). SciGRID\_gas IGGIELGN (1.1.2) Zenodo. <https://doi.org/10.5281/zenodo.4767098>
- Diettrich, J., Medjroubi, W. und Pluta, A. (2021). SciGRID\_gas INET\_Filled (1.1.0) Zenodo. <https://doi.org/10.5281/zenodo.4642563>
- Diettrich, J., Pluta, A., und Medjroubi, W. (2021). SciGRID\_gas EMAP\_Raw (1.2.0) Zenodo. <https://doi.org/10.5281/zenodo.4751020>

- Dasenbrock, J., Diettrich, J., Pluta, A., und Medjroubi, W. (2021). SciGRID\_gas NO\_Raw (1.2.0) Zenodo. <https://doi.org/10.5281/zenodo.4751001>
- Diettrich, J., Pluta, A., und Medjroubi, W. (2021). SciGRID\_gas LKD\_Raw (1.2.0) Zenodo. <https://doi.org/10.5281/zenodo.4750988>
- Diettrich, J., Pluta, A., und Medjroubi, W. (2021). SciGRID\_gas GIE\_Raw (1.2.0) Zenodo. <https://doi.org/10.5281/zenodo.4750985>
- Diettrich, J., Medjroubi, W. und Pluta, A. (2020). SciGRID\_gas IGGI (1.0.0) Zenodo. <https://doi.org/10.5281/zenodo.4288468>
- Diettrich, J., Medjroubi, W. und Pluta, A. (2020). SciGRID\_gas IGGIN (1.0.0) Zenodo. <https://doi.org/10.5281/zenodo.4288459>
- Diettrich, J., Medjroubi, W. und Pluta, A. (2020). SciGRID\_gas IGGINL (1.0.0) Zenodo. <https://doi.org/10.5281/zenodo.4288440>
- Diettrich, J., Pluta, A., und Medjroubi, W. (2020). SciGRID\_gas IGG (1.0.0) Zenodo. <https://doi.org/10.5281/zenodo.4009129>

#### Poster und Vorträge:

- Pluta, A., Diettrich, J. (2021). **SciGRID\_gas: Open Source Data und Software für das Europäische Gas Transportnetz** (Vorträge). Jahrestreffen Forschungsnetzwerk Energiesystemanalyse, 18.-20. Mai 2021 (Digitalevent)
- Dasenbrock, J., Pluta, A., Diettrich, J., und Medjroubi, W. (2020). **Gas Pipeline Detection Using Satellite Images and Machine Learning Algorithms** (Poster). Open Energy Modelling Workshop, 15.-17. Januar 2020, Hertie School of Governance, Berlin (Digitalevent).
- Diettrich, J., und Pluta, A. (2020). **Mining and Processing of a European Gas Transport Network Dataset Within the SciGRID\_gas Project** (Vortrag). Open Energy Modelling Workshop, 15.-17. Januar 2020, Hertie School of Governance, Berlin (Digitalevent).
- Pluta, A., Lünsdorf, O. (2020). **Creating Energy System Datasets from OpenStreetMap** (Vortrag). Open Energy Modelling Workshop, 15.-17. Januar 2020, Hertie School of Governance, Berlin (Digitalevent).
- Pluta, A., Diettrich, J. (2019). **SciGRID\_gas Workshop: Einführung in Gasnetzmodelle und Open Data** (Vorträge). Jahrestreffen Forschungsnetzwerk Energiesystemanalyse, 23.-24. Mai 2019, Aachen.
- Pluta, A., Diettrich, J., Medjroubi, W., und Lünsdorf, O. (2019). **SciGRID\_gas – A Topological Open Source Model of the European Gas Transmission Network** (Poster). Jahrestreffen Forschungsnetzwerk Energiesystemanalyse, 23.-24. Mai 2019, Aachen.

## Literaturverzeichnis

- Abeyssekera, M. W. (2016). Steady state analysis of gas networks with distributed injection of alternative gas. *Applied Energy*, 164, S. 991-1002.
- Baumann, C. (2016). *Simulation der europäischen Märkte für Erdgas und Strom*. Aachen: Jahresbericht des Instituts für Elektrische Anlagen und Energiewirtschaft der RWTH Aachen .
- Brkić, D. &. (2019). An efficient iterative method for looped pipe network hydraulics free of flow-corrections. *Fluids*, 4(2), S. 73.
- Brkić, D. (2009). An improvement of Hardy Cross method applied on looped spatial natural gas distribution networks. *Applied Energy* 86(7-8), S. 1290-1300.
- Cerbe, G. a. (2016). *Grundlagen der Gastechnik: Gasbeschaffung-Gasverteilung-Gasverwendung*. Carl Hanser Verlag GmbH Co KG.
- Connolly, D. L. (2014). Heat Roadmap Europe: Combining district heating with heat savings to decarbonise the EU energy system. *Energy Policy*, S. 475-489.
- ENTSO-G. (2009). *Verband Europäischer Fernleitungsnetzbetreiber für Gas (ENTSO-G)*. Von <https://www.entsog.eu/> abgerufen
- Gay, B. &. (1971). The solution of pipe network problems. *Chemical engineering science*, 26(1), S. 109-123.
- Götz, M. R. (2011). Storage of volatile renewable energy in the gas grid applying 3-phase methanation. *Proceeding of International Gas Union Research Conference*. Seoul, République de Corée.
- Ke, H. a. (2000). Transient analysis of isothermal gas flow in pipeline network. *Chem. Eng. Jour.* 76, S. 169-177.
- Kirchmayr, M. (2014). *Power-to-Gas: Modellierung der Energieverwertungspfade und Einflussnahme einer veränderten Strommarktsituation*. Diplomica Verlag.
- Koch, T. a. (2015). *Evaluating gas network capacities*. SIAM.
- Kralik, J. S. (1988). *Dynamic Modeling of Large-Scale Networks with Application to Gas Distribution*. Amsterdam: Elsevier.
- Kunz, F. K. (2017). *Kunz, Friedrich, et al. Electricity, heat, and gas sector data for modeling the German system*. No. 92. DIW Data Documentation.
- Li, Q. A. (2003). *Solving natural gas loadflow problems using electric loadflow techniques*. Proc. of the North American power symposium.
- Lopes, A. M. (2004). Implementation of the Hardy-Cross method for the solution of piping networks. *Computer Applications in Engineering Education*, 12(2), S. 117-125.

- Lustenberger, P. S. (2019). Assessing the performance of the european natural gas network for selected supply disruption scenarios using open-source information. *Energies*, 12(24), 4685.
- Metz, M. (2014). *Flexible Energieversorgung: Modellierung der Last-und Erzeugungssituation dezentraler Versorgungsgebiete zur Bestimmung der Systemflexibilität*. Dortmund : Diss. Universitätsbibliothek.
- Michaelis, J. J. (2013 ). Eine bewertung der regelenergievermarktung im power-to-gas-konzept. *Zeitschrift für Energiewirtschaft*, S. (37.3) 161-175.
- Onyeji, I. B. (2014). Cyber security and critical energy infrastructure. *The Electricity Journal*, S. 27(2), 52-60.
- Osiadacz, A. (1987). *imulation and analysis of gas network*. London: E&FN Spon.
- Otuoze, A. O. (2018). Smart grids security challenges: Classification by sources of threats. *Journal of Electrical Systems and Information Technology*, S. 5(3), 468-483.
- P. Crucitti, V. L. (2004). A topological analysis of the Italian electric power grid. *Physica A: Statistical mechanics and its applications*, 338 (1), S. 92-97.
- P. Hines, S. B. (2010). The topological and electrical structure of power grids. *43rd Hawaii International Conference on System Sciences (HICSS)* (S. 1-10). Hawaii: IEEE.
- Pernice, M. a. (1998). NITSOL: A Newton iterative solver for nonlinear systems. *SIAM Journal on Scientific Computing*, 19, S. 302-318.
- Pluta, A. M. (24. January 2022). SciGRID\_gas - Data Model of the European Gas Transport Network. *arXiv 2201.08827*.
- PSIG. (2000). *Recommended Practice for Software Acquisition*. Pipeline Simulation Interest Group.
- Qadrdan, M. A. (2020). Fundamentals of Natural Gas Networks. *The Future of Gas Networks* (S. 5-22). Springer, Cham.
- Rüdiger, J. (2009). Gasnetzsimulation durch potentialanalyse. *Dissertation*. Hamburg, Universität der Bundeswehr : Helmut-Schmidt-Universität, Fakultät Elektrotechnik, Lehrstuhl: Regelungstechnik.
- Sanjab, A. S. (2016). Smart grid security: Threats, challenges, and solutions. *arXiv preprint*, S. arXiv:1606.06992.
- Scheib, P. (2008). *Netzzugang im liberalisierten deutschen Gasmarkt : eine modellgestützte Analyse zur Bildung von Marktgebieten*. Karlsruhe: Universität Karlsruhe.
- Szoplik, J. (2012). *Advances in Natural Gas Technology*. Saudi Arabia: IntechOpen.
- Szoplik, J. (2016). Improving the natural gas transporting based on the steady state simulation results. *Energy*, 109, S. 105-116.

Ten, C. W. (2010). Cybersecurity for critical infrastructures: Attack and defense modeling. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, S. 40(4), 853-865.

Trost, T. H. (2016 ). Erneuerbares Methan: Analyse der CO2-Potenziale für Power-to-Gas Anlagen in Deutschland. *Interdisziplinäre Aspekte der Energiewirtschaft*, S. 295-312.

## Abbildungsverzeichnis

|   |    |
|---|----|
| Abbildung 1: Beispiel für verfügbare Daten auf „openstreetmap.org“                            | 5  |
| Abbildung 2: Darstellung der Terminologie des SciGRID_gas Modells                             | 13 |
| Abbildung 3: Datenstruktur des SciGRID_gas Modells  | 13 |
| Abbildung 4: Gas Transport Pipelines mit Diameter in OSM für Europa (Stand 2020)              | 16 |
| Abbildung 5: Visualisierung des EMAP-Datensatzes  | 18 |
| Abbildung 6: Jährlich gemittelte Gasnachfrage im Haushaltssektor                              | 19 |
| Abbildung 7: Visualisierung des ENTSO-G Datensatzes (EMAP)                                    | 21 |
| Abbildung 8: Beispiel Gasnetz mit fehlenden Werten für ausgewählte Attribute                  | 22 |
| Abbildung 9: Histogramm-Darstellung des Kompressor-Attributs max_cap_M_m3_per_d.              | 25 |
| Abbildung 10: Histogramm der unverarbeiteten und geschätzten Werte der Komponente Storages    | 26 |
| Abbildung 11: Visualisierung des IGGIELGNC1-Datensatzes für das europäische Gastransportnetz. | 28 |
| Abbildung 12: Visualisierung des INET-filled-Datensatzes                                      | 30 |
| Abbildung 13: Topologischen Verifizierung von Pipelinedaten                                   | 31 |
| Abbildung 14: NEL Pipeline erstellt aus 22 historischen Satellitendaten Bilder                | 33 |

## Tabellenverzeichnis

|  |    |
|--|----|
| Tabelle 1: Zusammenfassung der Elemente und Attributwerte des Gasnetzes.                   | 22 |
| Tabelle 2: Eingabewerte (input capacity)   | 24 |
| Tabelle 3: Eingabewerte (input capacity)   | 24 |
| Tabelle 4: Original Datendichte für die wichtigsten Parameter von ausgewählten Komponenten | 32 |