
SciGRID_gas: The raw LKD data set

Release 1.1

J.C. Diettrich & A. Pluta & W. Medjroubi

Apr 13, 2021

CONTENTS

1	Introduction	7
2	Data structure	9
2.1	Data structure description	9
2.1.1	Terminology	9
2.2	Summary	14
3	Data sources	15
3.1	Non-OSM data	16
3.2	The Long-term Planning and Short-term Optimization (LKD) data set	18
3.2.1	Pre-requirements for accessing the LKD data set	19
3.2.2	Data processing of the LKD data	19
3.2.3	Further alterations to the LKD data set	20
3.2.4	LKD data density	20
3.2.5	Copyright and disclaimer for the LKD data set	23
3.2.6	Summary LKD data	23
3.3	Data summary	25
3.4	Summary	25
4	Conclusion	27
5	Appendix	29
5.1	Glossary	29
5.2	Unit conversions	31
5.3	Attribute <i>exact</i>	31
5.4	Location name alterations	32
5.5	Country name abbreviations	32
5.6	Acknowledgement	33
	Bibliography	35

How to cite

J.C. Diettrich, A. Pluta, W. Medrjoubi
SciGRID_gas: The raw LKD data set
DLR Institute for Networked Energy Systems
Germany
doi: 10.5281/zenodo.3980984

Impressum

DLR Institute for Networked Energy Systems
Carl-von-Ossietzky-Str. 15
26129 Oldenburg
Germany
Tel.: +49 (441) 999 060



LIST OF FIGURES

2.1	Data structure for the SciGRID_gas data set.	10
3.1	Map of components of the LKD data set.	24

LIST OF TABLES

3.1	LKD <i>PipeSegments</i> data summary	21
3.2	LKD <i>Compressors</i> data summary	21
3.3	LKD <i>Storages</i> data summary	22
3.4	LKD <i>Productions</i> data summary	22
3.5	LKD <i>Nodes</i> data summary	22
3.6	LKD component summary	23
5.1	Dataset abbreviations	29
5.2	Glossary	30
5.3	Unit conversions	31
5.4	Unit conversions	31
5.5	Country codes	32

Summary

The goal of SciGRID_gas is to develop methods to create an automated process that can generate a gas transmission network data set for Europe. Gas transmission networks are fundamental for simulations by the gas transmission modelling community, to derive major dynamic characteristics. Such simulations have a large scope of application, for example, they can be used to perform case scenarios, to model the gas consumption, to minimize leakages and to optimize overall gas distribution strategies. The focus of SciGRID_gas will be on the European transmission gas network, but the principal methods will also be applicable to other geographic regions.

Data required for such models are the gas facilities, such as compressor stations, LNG terminals, pipelines, etc. One needs to know their locations, in addition to a large range of attributes, such as pipeline diameter and capacity, compressor capacity, configuration, etc. Most of this data is not freely available. However throughout the SciGRID_gas project it was determined, that data can be found and grouped into two fundamental different groups: a) OSM data, and b) non-OSM data. The OSM data consists of geo-referenced facility data that is stored in the OpenStreetMap (OSM) data base, and is freely available. However, the OSM data set currently contains hardly any other information than the location of the facilities. The Non-OSM data set can fill some of those gaps, by supplying information such as pipeline diameter, compressor capacity and more. Part of the SciGRID_gas project is to mine and collate such data, and combine it with the OSM data set. In addition heuristic tools are required to fill data gaps, so that a complete gas network data set can be generated.

Here, this document describes one of the non-OSM data set, called the “LKD” data set, which originated from “Electricity, Heat, and Gas Sector Data for Modeling the German System” [KKS+17]. This document here explains the origin and structure of this single data sets.

In this document, the chapter “Introduction” will supply some background information on the SciGRID_gas project, followed by the chapter “Data structure”, that gives a detailed description of the data structure that is being used in the SciGRID_gas project. Chapter “Data sources” describes the LKD data set.

The appendix contains a glossary, references, location name alterations convention and finishes with the table of country abbreviation.

INTRODUCTION

DATA STRUCTURE

A well designed and documented data structure is fundamental in any large-scale project. Good data structure in combination with tools, based on algorithms, improve the performance of any project output.

This structure needs to represent the gas flow facilities as good as possible. Hence, it needs to include components, such as pipelines, compressors etc. A finite number of components have been identified that are required as building blocks of a gas network. In addition, each component will contain attributes, such as pipeline diameter, maximal operating pressure, maximal capacity, number of turbines etc.

It is anticipated that the adopted data structure can be implemented in different types of gas flow models and will be used by the research community for topics, such as sector coupling or identifying gas transmission bottlenecks.

Within the SciGRID_gas project, the structure of the data model is part of classes defined within the Python code. Alterations may occur over the duration of the project, but it is envisaged that those will be small, and that compatibility will be assured.

The goal of this section is to describe in details the data structure that has been adopted and implemented into the Python code. This will be important in understanding other aspects of this document, such as exporting the data into CSV files or generating missing values.

2.1 Data structure description

This section contains information on the SciGRID_gas data structure, the format, and the code that can be used to import publicly available data into the project, so that it can be used in subsequent steps. Paramount for an understanding of the data structure is a good understanding of the terminology used throughout this section and the document in general. Hence, terminology will be introduced in the following sub-section.

2.1.1 Terminology

Throughout this document certain terms will be used, which will be described below and have been summarized in [Figure 2.1](#).

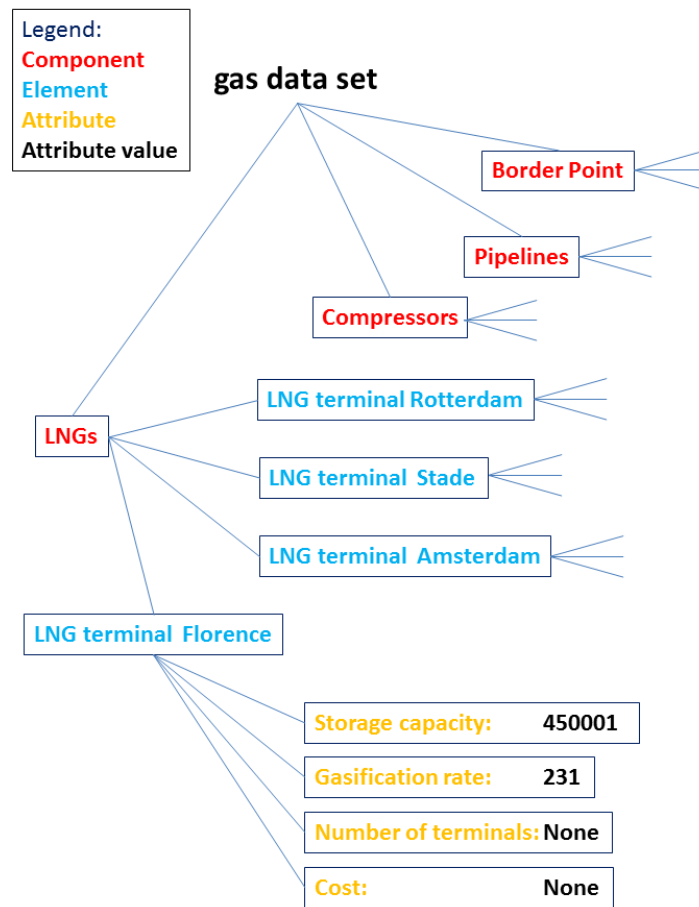


Figure 2.1: Data structure for the SciGRID_gas data set.

Gas transmission network

The term “gas transmission network” describes the physical gas transmission grid. This does not include the distribution of gas through gas distribution companies, but includes the long distance transmission of gas from producer countries to consumer countries, as carried out by the Transmission System Operators (TSO) [Wik20g]. In addition, throughout this document, the terms “transportation” and “transmission” are seen as interchangeable, and hence, will both be used describing the same.

Gas component data set

The term “gas component data set” is used for all raw data of objects/facilities that have been loaded using SciGRID_gas tools into a Python environment. Gas component data sets are used as input into our SciGRID_gas project. Several data sources can be loaded as gas component data sets, and then combined into a single gas component data set. However, not all elements (e.g. compressors) must be connected to pipelines. Hence, such a data set is referred to as a “gas component data set”.

Gas network data set

A “gas component data set” can be converted into a “gas network data set”, by connecting all non-pipeline elements to nodes and all nodes are connected to pipelines, and as part of the process all network islands have been connected or removed, resulting in a single network. Therefore, the network contains nodes and edges which are coherently connected, and all objects with the exception of pipelines are associated with nodes in this network, whereas pipelines are associated with edges.

Component

There are several component types in a gas transmission network, such as compressors, LNG terminals, or pipelines. In Figure 2.1 they are coloured red. Hence, whenever the word “component” is mentioned, it refers to one of these components. There are roughly a dozen different components that will form a gas network data set. They will be briefly explained below.

Element

The term “element” refers to individual facilities, e.g. the LNG terminal in Rotterdam, or the compressor in Radeland. In Figure 2.1 they are coloured blue. The first one is an element of the component LNG terminals, whereas the second one is an element of the component compressors. Hence, many elements make up a component. However, all elements are referring to different facilities by default. This means in a single network, one cannot have two elements of a component describing the same facility. The structure of elements is described below.

Attribute

“Attribute” is a term that is being used for the individual parameters that are associated with the elements. Examples of this term are gas “pipeline diameter”, “maximum capacity”, “max gas pipeline pressure”, to name just a few and in Figure 2.1 they are coloured yellow. Overall there will be several hundred attributes in the SciGRID_gas project. However, the same attributes can occur in more than one component, e.g. “max flow capacity” exist for pipelines and also for compressors. Throughout the project, we have tried to keep the units of such attributes the same, so that there is no unit conversion required.

Attribute value

Each attribute has a value, most likely a number or a string. In [Figure 2.1](#) they are coloured black. While booleans (*True/False*) are also allowed, more likely a “1” will stand for *True* and “0” for *False*. However, not all attribute values are given. Therefore, a no value for attribute values needs to be specified. In the SciGRID_gas Python code it is *None*.

The [Figure 2.1](#) depicts the relationships between the terms “gas data set”, “component”, “element”, “attribute”, and “attribute value”. As can be seen, a single gas data set consists of several components. On the next level, each component contains several elements. Further, each element has several attributes, where each attribute has a single or several values. The heuristic processes described in this document at a later stage will fill all missing values with heuristically generated values.

Gas component types

A gas transmission network consists of different components, such as pipelines, compressors etc. For the SciGRID_gas project a hand-full of components have been implemented, and will be described here briefly:

- *Nodes*: In a gas network, gas flows from one point to another point, which are given through their coordinates. All elements of all other components (such as compressor stations and power plants) have an associated node, which allows for the geo-referencing of each element. Overall the term *Nodes* will be used throughout this document, as it aligns with graph theory aspects.
- *PipeLines*: *PipeLines* allow for the transmission of the gas from one node to another. *PipeLines* are georeferenced by an ordered list of nodes.
- *PipeSegments*: *PipeSegments* are almost identical to *PipeLines*. However, are only allowed to connect two nodes. Hence, any *PipeLines* element (with 3 or more nodes) can easily be converted into multiple *PipeSegments* elements.
- *Compressors*: *Compressors* represent compressor stations, which increases the pressure of the gas, and hence, allows the gas to flow from one node to another node. A gas compressor station contains several gas compressors units (turbines).
- *LNGs*: *LNGs* is the acronym for the LNG terminals and LNG storages, which there are several in Europe, as some gas gets transported to Europe via ships.
- *Storages*: *Storages* are a further network component. Surplus gas can be stored underground (e.g. in old gas fields or salt caverns), and used during low supply or high demand periods.
- *Consumers*: *Consumers* is the term used for gas users, which can be households, industry, power plants or others.
- *Productions*: These can be wells inside a country where gas is pumped out of the ground. Most of the gas used in Europe comes from outside of the EU. However, there are several smaller gas production sites scattered throughout Europe.
- *BorderPoints*: *BorderPoints* are facilities at borders between countries, which are mainly used to meter the gas flow from one country to another.

Element structure

As described above, elements are describing individual facilities, such as compressors or LNG terminals. However, the overall structure of those elements is the same for all elements of all components, and is described as follows:

- *id*: A string that is the ID of the element, and must be unique.
- *name*: A string that is the name of the facility, such as “Compressor Radeland”. In most cases this is not supplied.
- *source_id*: A list of strings that are the data sources of the element. As several elements from different sources could have been combined into a single element, one might need to know the original data sources.
- *node_id*: The ID of a geo-referenced node to which an element of the network is associated to. For a compressor, this will be just a single *node_id*. However, for a gas pipeline this entry would be a list of at least two *node_id* values: the starts node id and the end node id.
- *lat*: The latitude value of an element. For elements of type *PipeLines* and *PipeSegments*, *lat* is a list of latitude values. Throughout the SciGRID_gas project the projection World Geodetic system 1984 (epsg:4326) will be used.
- *long*: The longitude, analogue to lat.
- *country_code*: A string indicating the 2-digit ISO country code (Alpha-2 code, see [Chapter 5.5](#) for list of countries and their codes) of the associated node of elements or list of nodes in case of *PipeLines* or *PipeSegments*.
- *comment*: An arbitrary comment that is associated with the element. In most cases this is not supplied.
- *tags*: This dictionary is reserved for OpenStreetMap data. It contains all associated key:value-pairs of an OpenStreetMap item.

In addition, there are three further groups of attributes to each element, which have been coded as Python “dictionaries”, named:

- *param*
- *method*
- *uncertainty*.

The structure within each dictionary is the same. The dictionary *param* (short for “parameter”) contains a list of attributes and their values. This list of attributes will be different for each component. For the component *PipeLines* they might be pipeline diameter, max pipeline pressure, and max pipeline capacity. For the component *Compressors* they might be, a number of turbines, overall turbine power, energy source of turbine or other.

The other two attribute dictionaries are *method* and *uncertainty*. Each of those two dictionaries contains exactly the same list of attributes as the *param* dictionary. However, their attribute values reflect the name of the dictionary. E.g. the attributes in the dictionary *method* contain the information on the method used to derive the attribute value that is stored in the *param* dictionary. Here methods of value generation can include heuristic methods names (in form of strings) that have been implemented in the SciGRID_gas project. However, if attribute values are not being generated by the SciGRID_gas project, but originate from one of the input data sources, then the attribute values in the *method* dictionary is set to “raw”.

See example below, for an *LNGs* element with the following entries:

- “make_Attrib(const)”: the attributes *end_year*, and *is_H_gas* have been set to a constant value
- “raw”, indicating that the two attributes *max_cap_store2pipe_M_m3_per_d* and *start_year* contain original values
- “Lasso(max_cap_store2pipe_M_m3_per_d)”, here for the attribute *median_cap_store2pipe_M_m3_per_d* a method was used that is based on the lasso method and uses the attribute *max_cap_store2pipe_M_m3_per_d* as input.

Similar is the content of the *uncertainty* dictionary. It contains information on the uncertainty of the attributes from the *param* dictionary of that component. Again, all attributes listed in the *param* dictionary are also present in the *uncertainty* dictionary. The attribute values here reflect the uncertainty of the attribute. Here, it is assumed that attributes with a method of “raw” have an uncertainty of zero. Only for those attributes, which were generated during heuristic SciGRID_gas methods an uncertainty larger than zero will be specified.

2.2 Summary

The SciGRID_gas software is designed to construct a gas transmission network data set from different open and non-open source gas component data sets. The gas transmission data set needs to be available and stored in a precise and predefined way, which was described in this section. We have identified several *component*-types of a gas transmission network grid, like pipelines, compressor stations, LNG-terminals etc. Each specific facility that falls under such a component is considered an *element* of that component. Each element is described by a list of *attributes* and correspondent *attribute values*, including information on the uncertainty of the attribute value and the way the attribute value was generated.

DATA SOURCES

Original data sets describing gas transmission networks are the property of the transmission system operators (TSOs) and are generally not freely available in the form and depth that is required for modelling purposes. The major reason for the difficulty of obtaining of such data is that most of the gas network infrastructure, namely pipelines, is buried underground. Thus, a pipeline diameter is hard to estimate locally. In addition, almost all of the data is commercially sensitive.

Nevertheless, some data is made available by gas transmission network operators, through different channels. E.g. information on the size and number of compressors could be made public through a press release, as part of a refurbishment. An example is given below (<https://www.maz-online.de/Lokales/Teltow-Flaeming/Neue-Verdichterstation-entsteht-in-Radeland>):

“Die Eugal-Pipeline dient dazu, Gas aus der neuen Ostseepipeline Nord Stream 2 bis zur tschechischen Grenze zu leiten. 275 Kilometer von ihr verlaufen in Brandenburg. Grundsätzlich soll die neue Leitung parallel zur bestehenden Opal-Pipeline gebaut werden.”

In addition, some information can be found on company web pages, (<https://www.open-grid-europe.com/cps/rde/SID-752BB6B5-E0A975F2/oge-internet-preview/hs.xsl/NewsDetail.htm?rdeLocaleAttr=en&newsId=50190C3B-E14F-4685-9E64-E40EEAB57A28>):

“Open Grid Europe (OGE) is investing roughly EUR 150 million at its compressor station in Werne to improve the security and flexibility of energy supply for North Rhine-Westphalia and Germany. The upgrade of the station, which is one of the hubs of the pipeline network, will allow gas flows to be switched (reversed) from north to south and south to north. In addition, OGE is preparing the station for the upcoming transition from L- to H-gas. Through this fitness programme, the station’s transmission capacity will increase by about 500,000 to 6.5 million m³/h, which is equivalent to the annual consumption of more than 2,100 single-family homes. The project, which is due for completion at the end of 2018, is fully on track.”

However, there is a public drive to gather such data and subsequently make it available. The major platform through which this is occurring is the Open Street Map database [Hel18]. OSM is a geo-referenced database through which people can supply geo-referenced information on all man-made and natural structures, ranging from mountains to buildings. To achieve this, people throughout the world wander the globe and geo-reference everything that they can find. This also includes gas-pipeline markers, compressor stations or LNG terminals. However, the major problem remains that one cannot measure or estimate the diameter of the underground pipelines, or the number and size of the compressor turbines, as compressors are within buildings, which are fenced off. Hence, such information is hardly supplied to the OSM platform.

For the reasons mentioned above, the available data can be separated into two different groups:

- OSM data: Data can be found in the OSM data base. OSM data is well geo-referenced, but contains little meta-information (information on the facility attributes, such as pipeline diameter or pipeline capacity). OSM data is very helpful to obtain accurate routes of pipelines.
- Non-OSM data: Non-OSM data have in general lower geographical accuracy but contain a lot of meta-information. Unfortunately, such information is only known for a few facilities. One exception to this rule

are shapefiles from TSOs. They are rare, but well geo-referenced. However, the resolution of the meta information can vary from TSO to TSO.

The following section will introduce non-OSM data sets, and at a later stage, this will be followed by a section on the OSM data.

3.1 Non-OSM data

Non-OSM data includes data from internet research, TSO press releases, TSO transparency platform, TSO public data, national open-source gas network data sets¹ etc.

Some of the TSO information had to be made available due to EU-regulations. Other information has been made public as part of a company's self-presentation and advertisement. The information used by the SciGRID_gas project focuses on:

- the quality of the data
- the format of the data
- the level of representation of the data
- and the copyright restrictions on the data.

In addition, each data source is unique. Source specific tools need to be developed, so that all data sources can be made accessible for the SciGRID_gas project.

A significant portion of the project was spent on finding non-OSM data sets. Further data sources might be available, but unknown to the authors. If the authors are made aware of additional sources, the project will try to incorporate those, as this would only increase the depth of the data available and increase the applicability of the gas network data set and model.

Non-OSM data sources are very specific, addressing only certain aspects of the entire gas infrastructure. E.g. the GIE [GasIEurope20] data set supplies information on the daily gas flow in and out of gas storages in LNG terminals. However, they fall short on specifying the fundamental information of the actual physical location. Other data sets, such as the LKD [KKS+17] data set is quite detailed in respect of pipelines, compressors and consumptions, however, only available for Germany.

Hence, the main task is to look closely at each data source, distil which data attribute values can be used, how it can be downloaded and incorporated into the SciGRID_gas model, and identify the copyright restrictions on the data source.

Due to copyright regulations, there are roughly two groups of data:

- Non-copyright restrictive data (N-CRRD): Here the copyright does not restrict the download, use and distribution of the data.
- Copyright restrictive data (CRRD): Here the data can be downloaded and used internally, but not re-distributed to others.

The following is a list of the data sources that will be used throughout the project and an identification into which group of copyright restriction they fall:

- **OSM** (<https://www.openstreetmap.org>) (N-CRRD)
- **GB** (<https://www.nationalgridgas.com/land-and-assets/network-route-maps>) (CRRD)
- **NO** (<https://www.npd.no/en/about-us/information-services/available-data/map-services/>) (N-CRRD)
- **LKD** (<https://tu-dresden.de/bu/wirtschaft/ee2/forschung/projekte/lkd-eu>) (N-CRRD)
- **ENTSOG** (<https://transparency.entsog.eu/>) (CRRD)

¹ An entire gas network data set is only available from the UK, see <https://www.nationalgridgas.com/land-and-assets/network-route-maps>.

- **EMAP** (https://www.entsog.eu/sites/default/files/2020-01/ENTSOG_CAP_2019_A0_1189x841_FULL_401.pdf) (CRRD)
- **GIE** (<https://www.gie.eu/>) (N-CRRD)
- **GSE** (<https://www.gie.eu/index.php/gie-publications/databases/storage-database>) (N-CRRD)
- **IGU** (<https://www.igu.org/>) (CRRD)
- **INET** (see `RefsInternetData`) (N-CRRD)
- **CONS** (N-CRRD).

Each data set and source comes with different copyright regulations. The copyright can be rather non-restrictive (e.g. INET) or can be restrictive (IGU). It is attempted to use only freely available data, so that such data can be re-distributed. In more restrictive data cases (IGU, GB), it is not allowed to download the data and distribute it to others. However, it is allowed to let other potential users know of the location of such data and supply them with tools that allow them to carry out the same data download and subsequent incorporation of the data into a gas network data set.

Note:

In case that other users are aware of other data sources that might be useful to this project, please get in touch and supply us with a brief description of the data and the location of such data, so that additional tools can be developed to incorporate the data in this project. Please use the following email address: `developers.gas(at)scigrid.de`

3.2 The Long-term Planning and Short-term Optimization (LKD) data set

The **Long-term planning and short-term optimization** data set (**lk-DEU**) is the second of three non-OSM data sets that contain geo-referenced gas facilities. It was generated by several German research institutes and funded through the German government grants. It was part of a much larger research project (see link below). Here the gas facilities from the lk-DEU data set were used and incorporated into SciGRID_gas data model as the **LKD** data set. It contains information on gas pipelines, gas production sites, gas storages, compressor locations, and nodes.

As this data set is extremely well geo-referenced, it is of particular interest to the SciGRID_gas data project. The LKD data set can be used in conjunction with the OSM data set for training purposes, and as a data source for the heuristic processes, as a lot of attributes are available for a lot of elements. In addition, pipelines from the LKD data set can be copied into the final SciGRID_gas data set.

The LKD facilities data set came in form of a shapefiles, and consisted of polylines with some attributes, such as pipe diameter, max gas flow capacity and more. In addition, parts of the shapefiles were tables of facilities, with information on storages, production, and industrial demand. Great care was taken from the original data set producers, to create a data set with a vast number of attributes, which will be used throughout the SciGRID_gas project. Overall, the topological quality of the data set is good, as was verified by some sample checks. Gas sites could be found on satellite images within a few hundred meters. Due to the large number of elements, with a good selection of attributes and good topological information, the entire LKD data set has been incorporated into the SciGRID_gas project.

Further external information on the lk-DEU data set

More information on the data can be found under the following URL:

<https://www.ewl.wiwi.uni-due.de/nl/forschung/forschungsprojekte-ewl/lkd-eu-langfristige-planung-und-kurzfristige-optimierung-des-elektrizitaetssystems-in-deutschland-im-europaeischen-kontext/>.

This link describes the (Long-term planning and short-term optimization data sets of the German electricity system within the European) data set [KKS+17]. The project was a joined effort by:

- German Institute for Economic Research (DIW Berlin)
- Working group for Infrastructure Policy (WIP) at Technische Universitaet Berlin (TUB)
- Chair of Energy Economics (EE2) at Technische Universitaet Dresden (TUD)
- House of Energy Markets & Finance at University of Duisburg-Essen.

This project was funded by the German Federal Ministry for Economic Affairs and Energy through the grant “LKD-EU”, FKZ 03ET4028A, with the aim of presenting a status quo of the German energy sector. The following three energy media were part of their project:

- electricity
- heat
- natural gas.

Here only the gas components are being used.

3.2.1 Pre-requirements for accessing the LKD data set

The SciGRID_gas project has received the right to use, change and redistribute the LKD data under an open license agreement. However, if you use this data or any data set which incorporates this data, you are also required to cite the original authors of the LKD data as follows:

Kunz et al. 2017, Data Documentation: Electricity, Heat, and Gas Sector Data for Modeling the German System

In addition, the data set can be downloaded from the following location:

<https://zenodo.org/record/1044463#.Xah7i2ZCSUk>

Please put a copy into the following location:

/SciGRID_gas/Eingabe/LKD/

In addition, some data changes needed to be carried out due to small mistakes in the LKD data. However, these have been carried out by the SciGRID_gas project and corrected data has been written to the CSV LKD output data set. The SciGRID_gas tools that carry out those changes will be supplied as part of the SciGRID_gas project.

3.2.2 Data processing of the LKD data

The LKD gas facilities data set came in the form of several shapefiles. It contained several tables, which were read in with tools and dissected to fit into the data structure of the SciGRID_gas project. The components that were read in are:

- *PipeSegments*, from the 'pipelines_utf8.shp' shapefile
- *Nodes*, from the 'nodes_utf8.shp' shapefile
- *Productions*, from the 'productions_utf8.shp' shapefile
- *Storages*, from the 'storages_utf8.shp' shapefile.

Subsequent to reading the data from the shapefiles, it was necessary to convert the data so that it adheres to the SciGRID_gas data structure. Some inconsistencies were found with the data set. The following fixes of the LKD data set had to be carried out:

- Some node ids were found more than once in the original data set for different nodes. Hence, this was manually rectified by changing node ids for 29 nodes.
- Some nodes had a wrong country code setting. For 10 nodes the country code attribute needed to be changed.

In an additional step, the elements of type *Compressors* were generated by using information that was supplied with the *Nodes*. The Node elements contained an attribute "comp_units", which stands for "number of compressor units". Hence, if this value was larger than 0, then it was assumed that the node contained a single compressor element at that location. In addition, the attribute "comp_units" was then used as the value for the number of compressor turbines at the compressor location. E.g. if the value was two, then the compressor element's attribute *num_turb* was set to two.

At this stage of the LKD data process, there were more than 1800 pipe-segments with more than 1400 nodes. It was not apparent why there were so many nodes and pipe-segments. For many pipe-segments, two individual pipe-segments that connect with the same node contained the same attributes with the same values, and the node in question only connected two pipe-segments, not forming a T-section. Hence, pipe-segments were joined and nodes removed if the following rules applied:

- The node in question connects only two pipe-segments.
- The attributes values for *max_pressure_bar*, *is_H_gas*, *diameter_mm* and *pipe_class_LKD* needed to be identical for both pipe-segments. An exception is made for the node "Haidach" and "N_805129", where no pipe-segment joining took place.

In addition, the following simplifications of the network were carried out:

- Nodes that were closer than 3 km were merged, removing some pipelines
- Pipelines that were connected to only one other element (pipeline or non-pipe component) were removed, if they were shorter than 5 km.

These processes reduced the number of segments to 1085, and the overall number of nodes to 721.

For some of the attribute values, the unit of the attribute value did not “agree” with the units used within the Sci-GRID_gas data project. Hence, unit transformation (see [Chapter 5.2](#)) had to be carried out for the following attributes of the following components:

- “Storages”, attribute converted from *max_cap_pipe2store_GWh_per_d* to *max_cap_pipe2store_M_m3_per_d*
- “Storages”, attribute converted from *max_cap_store2pipe_GWh_per_d* to *max_cap_store2pipe_M_m3_per_d*.

Subsequently, the old attributes with the “wrong” units were removed from the component data set.

Further attributes were added to the component:

- The length of the pipe-segment was derived using the polylines of each pipe-segment.
- The average latitude and average longitude were calculated by using the polylines of each pipe-segment.

In addition, the attribute “exact” was added to each *Nodes* element and a value of one was given.

In addition, for each *Nodes* element the following attributes were removed:

- ‘compressor’
- ‘ugs’
- ‘production’
- ‘comp_units’.

3.2.3 Further alterations to the LKD data set

Estimation of the attribute *max_cap_GWh_per_d*

The original data set contained for some pipelines the attribute *max_cap_GWh_per_d*. However, it was found that this value was incorrect. Therefore the attribute value *max_cap_M_m3_per_d* was generated, as described in [KKS+17], Chapter 4.2.2. To achieve this, a heuristic relationship was formed where the following were the independent variables: *max_pressure_bar*, *diameter_mm*, and *pipe_class_LKD*, and the attribute *max_cap_GWh_per_d* is the dependent attribute variable. The backbone of the heuristic relationship is the information from Table 25 from [KKS+17]. Here it is assumed that all values given in the original data set of the attributes *max_pressure_bar*, *diameter_mm*, and *pipe_class_LKD* have the same quality, no matter if found or estimated, and can be used in this heuristic process. This process generated a *max_cap_GWh_per_d* value for 937 pipesegments.

3.2.4 LKD data density

The data of the LKD data set contains elements from the following components:

- *PipeSegments*
- *Compressors*
- *Productions*
- *Nodes*
- *Storages*.

Each of those components and their attributes will be described below.

As all components have the following attributes, they are presented here once:

- *id*: unique identifier
- *name*: name of the pipe-segment
- *source_id*: a source id
- *node_id*: the id of the start and the end node of the pipe-segment
- *lat*: a list of latitude values
- *longitude*: a list of longitude values
- *country_code*: a string pair indicating the country code of the start and the end point
- *comment*: a user comment.

PipeSegments elements

Overall, there are 1085 *PipeSegments* elements in the LKD data set. In addition to the default attributes, the following non-standard attributes (see [Table 3.1](#)) are supplied. The number of attribute values supplied for each attribute is given by the column ‘Data density [%]’:

Table 3.1: LKD *PipeSegments* data summary

Attribute name	Description	Units	Data density [%]
<i>diameter_mm</i>	a pipe diameter	mm	88
<i>is_H_gas</i>	the gas type identifier	1 or 0	100
<i>length_km</i>	the total distance of the pipe-segment	km	100
<i>max_cap_M_m3_per_d</i>	maximum gas flow capacity	Mm ³ d ⁻¹	86
<i>max_pressure_bar</i>	maximum allowed pressure in the gas pipe	bar	83
<i>operator_name</i>	operator name		99
<i>pipe_class_LKD</i>	gas pipe-segment class type	1 to 6	87
<i>lat_mean</i>	calculated mean latitude value	degree	100
<i>long_mean</i>	calculated mean longitude value	degree	100

pipe_class_LKD

For reasons of attribute generation at a later stage, the values for *pipe_class_LKD* have been converted from A, B, C,... to 1, 2, 3,...

Compressors elements

Overall, there are 13 *Compressors* elements in the LKD data set. In addition to the default attributes, the following non-standard attributes were supplied (see [Table 3.2](#)) and partially populated for the component *Compressors*.

Table 3.2: LKD *Compressors* data summary

Attribute name	Description	Units	Data density [%]
<i>entsog_key</i>	key associated with EntsoG facility		38
<i>license</i>	indicator of the license		100
<i>num_turb</i>	the number of compressor turbines		100
<i>operator_name</i>	name of the operator		100

Storages elements

Overall, there are 14 *Storages* elements in the LKD data set. In addition to the default attributes, the following non-standard attributes (see Table 3.3) are supplied and populated for the component *Storages*.

Table 3.3: LKD *Storages* data summary

Attribute name	Description	Units	Data density [%]
<i>entsog_key</i>	key associated with EntsoG facility		100
<i>max_cap_pipe2store_M_m3_per_d</i>	maximum gas flow from the network into the storage unit	Mm^3d^{-1}	100
<i>max_cap_store2pipe_M_m3_per_d</i>	maximum gas flow from the storage unit into the network	Mm^3d^{-1}	100
<i>operator_name</i>	name of the operator		100

Productions elements

Overall, there are 6 *Productions* elements in the LKD data set. In addition to the default attributes, the following non-standard attributes (see Table 3.4) were supplied and populated for the component *Productions*.

Table 3.4: LKD *Productions* data summary

Attribute name	Description	Units	Data density [%]
<i>entsog_key</i>	key associated with EntsoG facility		100
<i>max_supply_M_m3_per_d</i>	maximum gas production	Mm^3d^{-1}	100
<i>is_H_gas</i>	boolean indicating that H gas type	1 or 0	100
<i>operator_name</i>	name of the operator		100

Nodes elements

Overall, there are 721 *Nodes* elements in the LKD data set. In addition to the default attributes, the following non-standard attributes (see Table 3.5) are supplied and partially populated for the component *Nodes*.

Table 3.5: LKD *Nodes* data summary

Attribute name	Description	Units	Data density [%]
<i>crossborder</i>	boolean indicating that node is a gas cross border point		98
<i>entry</i>	boolean indicating that node is a gas entry point		98
<i>entsog_key</i>	key associated with EntsoG facility		9
<i>exact</i>	value indicating the accuracy in geo-referencing	1 to 5	100
<i>exit</i>	boolean indicating that node is a gas exit point		98
<i>H_L_conver</i>	boolean indicating if converter between H & L gas		98
<i>license</i>	license key		98
<i>operator_Z</i>	additional operator name		64
<i>operator_name</i>	name of the operator		96

In addition the NUTS values (Nomenclature des unités territoriales statistiques) [Wik21] have been added to the nodes for NUTS-1, NUTS-2 and NUTS-3. For this the latitude and longitude of each node was used in a look-up in a spatial shape file. Hence for those values, the data density is 100 %.

Additional data from the LKD data set

In addition, there is data on gas demand on a spatial level of NUTS-3, leading to 402 elements. However, currently this is not being used, but might be used at a later stage.

3.2.5 Copyright and disclaimer for the LKD data set

The lk-DEU data set has been published under the Creative Commons Attribution 4.0 International Public License. This allows us to use the data in this project and re-distribute the data as well.

Disclaimer

The LKD data set is supplied on a best-effort basis only. While every effort is made to make sure the information is accurate and up-to-date, we do not accept any liability for any direct, indirect, or consequential loss or damage of any nature, however caused, which may be sustained as a result of reliance upon such information.

Acknowledgement

We would like to acknowledge the “Deutsches Institut für Wirtschaftsforschung” (Mohrenstr. 58, 10117 Berlin, Germany) for allowing the SciGRID_gas project to use their data.

3.2.6 Summary LKD data

The gas pipeline and gas facilities from the LKD data set is of great importance to the SciGRID_gas project. It is one of only three non-OSM data sets that contain gas facilities that are geo-referenced, and hence, can be used for validation processes covering all of Germany. In addition, it contains some attribute values in respect of gas pipelines that are fundamental for the gas data model. This data set was made available through a German research project and downloadable from the project’s web page. Tools have been written to load the LKD shapefiles and make them accessible for the SciGRID_gas project.

Below Table 3.6 summarises the number of elements for each component found:

Table 3.6: LKD component summary

Component Name	Count
<i>Compressors</i>	13
<i>Nodes</i>	721
<i>PipeSegments</i>	1085
<i>Productions</i>	6
<i>Storages</i>	14

In addition, the map in Figure 3.1 visualizes the data for Germany, of its more than 27,000 km transmission network.

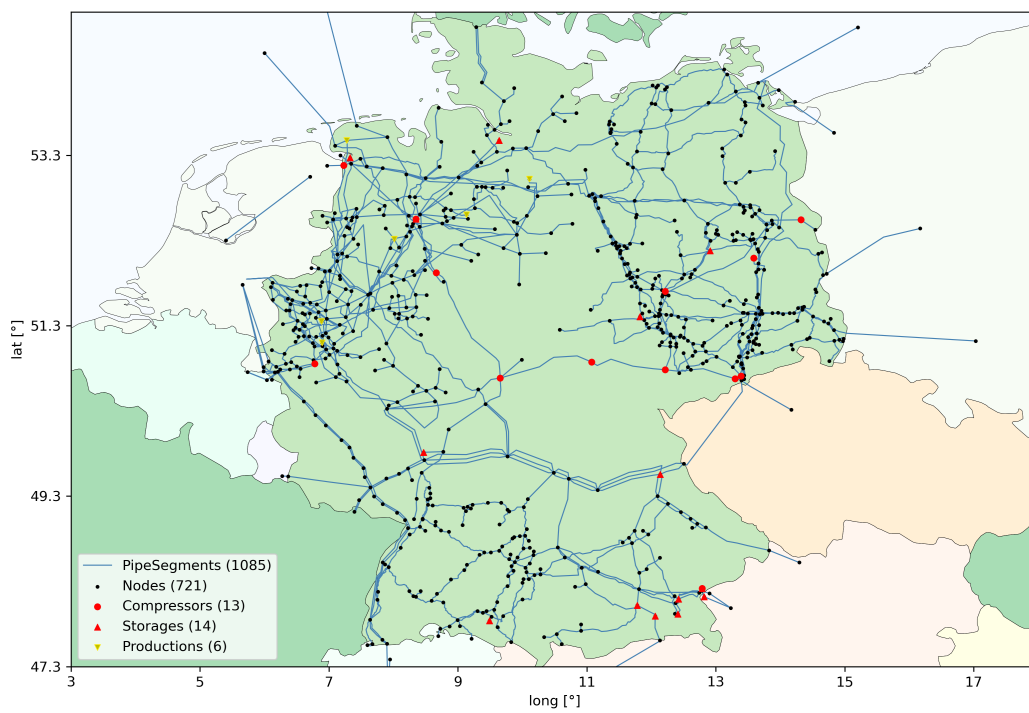


Figure 3.1: Map of components of the LKD data set.

3.3 Data summary

SciGRID_gas is based on open source data. To generate a gas pipeline network data set, one needs to access different data sets that were found throughout the project and presented here. Emphasis was given to depict the number of elements per component and the data density for each data set.

3.4 Summary

Gas component data sets come in different forms, licenses, formats and detail. The SciGRID_gas project can process such data and combine them to a consistent and reliable network data set.

The underlying gas component data sets were categorized into two different groups:

- OSM data: This is data originating from the OSM data base, containing well geo-referenced locations of gas facilities, such as pipe locations or gas storage facilities. However, it comes with very few meta information.
- Non-OSM data: These are all other data sources, which can “supply” detailed information on some of the gas facilities attributes. However, this information is sparse, as published only for a few facilities. Here, the INET data set was introduced as an example of the non-OSM data set, and the pathway of converting the raw data from the www into SciGRID_gas project component structure.

Here detailed information on one or several data sources have been given, and should be used as a reference for later data processes.

CONCLUSION

This document here is the documentation of one of the data sets that is part of the SciGRID_gas project. This document here started off with the introduction of the SciGRID_gas project, such as funding, duration and goals. In a subsequent chapter the data structure within the SciGRID_gas project was described, such as components, elements, attributes and attribute values, so that the transmission network data set could be an input to certain gas flow model. The third chapter introduced the LKD data set, which is a data set that generated by a third party, and the SciGRID_gas project is fortunat enough, that it can use the LKD data set and redistribut it through this project. Here only the raw LKD data set is being published, hence it will have some missing values. Despite some missing values, the re-distributed LKD data set consists of 0 LNG terminals, 14 storage facilities, 13 compressors and 1261 pipe segments throughout Germany.

5.1 Glossary

Dataset abbreviations can be found in [Table 5.1](#).

Table 5.1: Dataset abbreviations

Name	Abbreviation	Description
Raw InternetDaten data set	INET	Label/name for the raw InternetDaten data set
Raw Gas Infrastructure Europe data set	GIE	Label/name for the raw Gas Infrastructure Europe data set
Raw Gas Storage Europe data set	GSE	Label/name of the raw Gas Storage Europe data set
Raw Norwegian data set	NO	Label/name for the raw Norwegian data set
Raw Long-term planning and short-term optimization data set	LKD	Label/name for the raw Long-term planning and short-term optimization data set
Raw International Gas Union data set	IGU	Label/name for the raw International Gas Union data set
Raw EntsoG-Map data set	EMAP	Label/name for the raw EntsoG-Map data set
Merged and filled IGG data set	IGG	Filled data sets, for which the INET , GIE and GSE data sets were merged
Merged and filled IGGI data set	IGGI	Filled data sets, for which the INET , GIE , GSE and IGU data sets were merged
Merged and filled IGGIN data set	IGGIN	Filled data sets, for which the INET , GIE , GSE , IGU and the NO data sets were merged
Merged and filled IGGINL data set	IGGINL	Filled data sets, for which the INET , GIE , GSE , IGU , NO and the LKD data sets were merged
Merged and filled IGGIELGN data set	IGGIELGN	Filled data sets, for which the INET , GIE , GSE , IGU , EMAP , LKD , GB , and the NO data sets were merged

The glossary terms can be found in [Table 5.2](#).

Table 5.2: Glossary

Name	Abbreviation	Description
component		A gas network consists of different components, such as: pipelines, compressors, LNG terminals, storages, entry points and production sites
element		Elements are instances of components. Hence, “10 compressor elements” refers to a data set that contains information for 10 compressor stations
attribute		Gas facilities, such as pipelines or compressors, can be described with a large set of parameters, such as pipeline diameter, or compressor capacity. Those parameters are referred to as attributes
facility		General term used for a gas appliance, such as a single compressor station, or a single LNG terminal
PipeLine		A gas pipeline entity, which has one start and one end point, however, can run via many nodes
PipeSegment		A gas pipeline that has only one start and one end point, but no nodes in-between
LNG	LNG	Liquefied natural gas
CNG	CNG	Compressed natural gas
flow duration curve	FDC	It is the cumulative frequency curve that shows the percentage of time specified flow where equal or exceeded during a given period. The temporal information, when certain events occur, is lost
Energiewende		German term for the change in using primary energies, the move away from coal to renewable energies, such as wind or solar
gas component data set		Raw input data, associated with components of the gas transmission grid
gas network data set		Output data, a coherent network of gas transmission components
OSM	OSM	Data that is available from openstreetmap.org
non-OSM	Non-OSM	Data that is not part of the OSM data set
gas type		There are two types of gas: High (H) and Low (L) calorific gas
mean absolute error	MAE	mean difference between input values and estimated values
data density		The ratio of the number of usable (not missing) attribute values over number elements of the component, in units of [%]
Transmission System Operator	TSO	An entity entrusted with the transportation of natural gas/electricity, as defined by the European Union
gas transmission network		This describes the physical gas transmission grid, however, it excludes any facilities/components that would be part of a distribution network and their facilities
gas component data set		The term “gas component data set” is used for raw data sets of gas network facilities. However, not all elements (e.g. compressors) need to be connected to pipelines, where the emphasis is on the term component
gas network data set		A “gas component data set” can be converted into a “gas network data set”, by connecting all non-pipeline elements to nodes and all nodes are connected to pipelines. Hence, the emphasis here is on the term network

5.2 Unit conversions

Table 5.3: Unit conversions

From Unit	To Unit	MultiVal
LNG Mt	LNG Mm ³	2.47
gas tm ³ h ⁻¹	gas Mm ³ d ⁻¹	24/1000
LNG Mm ³	gas Mm ³	584
LNG t	gas Mm ³	1442.48
GWh (H)	gas Mm ³	0.0879757777
GWh (L)	gas Mm ³	0.1023541453

For some elements of some components, the calorific value was given through the references. Hence during the conversion process from GWh to M m³, the elements calorific value was used, however, wherever the element specific calorific value was not known, the default values from Table 5.3 was used in dependence of the gas type of the element. If no gas type was known, then high calorific gas is assumed.

5.3 Attribute *exact*

Each element of type *Nodes* has an attribute *exact*. With this, the SciGRID_gas project is trying to let the user know, how well the actual location of the *Nodes* elements are known. The actual location (latitude-longitude pair) can be spot on (verifiable through satellite imagery) or can be unknown by 10's or 100's of km, where city names or country names are known only. Here the attribute value for *exact* is being given, ranging from “1” to “5” as listed in Table 5.4 below.

Table 5.4: Unit conversions

Exact value	Description	Uncertainty [km]
1	The exact location of this node is known, as one was able to verify the facility through satellite data.	0
2	Here the lat/long is not known exactly. However, one assumes that the location is within a small region (e.g. Krummhörn). Hence, not being much larger than 10 km	10
3	Here so little is known about the exact location, and one only knows that the location is within a large region (e.g. Hamburg). Hence, the actual location could be out by 10 km or more but less than 100 km	100
4	Here so little is known about the exact location, and one only knows that the location is within a state (e.g. Niedersachsen). Hence, the actual location could be out by 100 km or more but less than 1000 km	1000
5	Here so little is known about the exact location, and one only knows that the location is within a country (e.g. Ukraine). Hence, the actual location could be out by 1000 km or more.	> 1000

5.4 Location name alterations

Location names should be changed into the 26 letters used in the English language.

For names from the individual countries please follow the suggested approach:

- Germany/Austria: *Umlaute* to be replaced with the letter followed by an ‘e’, e.g.: ü = ue.
- France/Belgium: Omit accent de gues and accent de graphs, e.g.: ó = o.
- Sweden: Please change the last three letters of the Swedish alphabet and replace e.g.: ä = a.
- Poland: Please change any letter that cannot be found in the English alphabet, knowing that for some letters that one can only use a single letter instead of the three different letters used in the Polish alphabet, e.g.: z = z.
- Spain/Portugal: Please change any letter that cannot be found in the English alphabet, e.g.: ñ = n.
- Greece: Please do not use Greek letters. Please try to write the Greek words with Latin letters.
- Denmark: Please change any letter that contains non-English letters, e.g.: “å” with ”aa”.
- Slovakia, Czech Republic, Hungary, Rumania, Latvia, Lithuania, Estonia, Bulgaria, Slovenia, Croatia: PLEASE use your common sense, based on the examples from the other countries above.

5.5 Country name abbreviations

For convenience we provide a short list of names and two-digit codes (see [Table 5.5](#)) for the probably most important countries associated with the European Transmission Grid.

Table 5.5: Country codes

Country name	Country code	Country name	Country code
Albania	AL	Kosovo	XK
Armenia	AM	Latvia	LV
Austria	AT	Liechtenstein	LI
Azerbaijan	AZ	Lithuania	LT
Belarus	BY	Luxembourg	LU
Belgium	BE	Malta	MT
Bosnia and Herzegovina	BA	Moldova	MD
Bulgaria	BG	Montenegro	ME
Croatia	HR	Netherlands	NL
Cyprus	CY	Norway	NO
Czech	CZ	Poland	PL
Denmark	DK	Portugal	PT
Estonia	EE	Romania	RO
Finland	FI	Serbia	RS
France	FR	Slovakia	SK
Georgia	GE	Slovenia	SI
Germany	DE	Spain	ES
Greece	GR	Sweden	SE
Hungary	HU	Switzerland	CH
Iceland	IS	Turkey	TR
Ireland and Northern Ireland	IE	Belarus	UA
Italy	IT	Great Britain	GB
Russia Federation	RU	Europe	EU
Ukraine	UA		

5.6 Acknowledgement

We acknowledge the contribution of Dr. Ontje Luensdorf from the DLR Institute of Networked Energy System to the SciGRID_gas project.

BIBLIOGRAPHY

- [AFW14] M. Ahmed, B.T. Fasy, and C. Wenk. *New Techniques in Road Network Comparison*. Penguin Random House, New York, NY, 2014.
- [AG99] H. Alt and L. Guibas. *Discrete geometric shapes: matching, interpolation, and approximation-a survey*. Sack JR, Urrutia J, Handbook of Computational Geometry, Elsevier, New York, NY, 1999.
- [DPDi20] J.C. Diettrich, A. Pluta, J. Dasenbrock, and W. i. *SciGRID_gas: The combined IGGIN gas transmission network data set*. DLR-Institut für Vernetzte Energiesysteme e.V., Oldenburg, Germany, 2020. URL: <https://zenodo.org/record/4288459#.YG7aFj9CSUk>, doi:10.5281/zenodo.4288458.
- [DPi20] J.C. Diettrich, A. Pluta, and W. i. *SciGRID_gas: The INET gas transmission network data set*. DLR-Institut für Vernetzte Energiesysteme e.V., Oldenburg, Germany, Aug 2020. URL: <https://doi.org/10.5281/zenodo.4008975>, doi:10.5281/zenodo.4008975.
- [DPM20a] J.C. Diettrich, A. Pluta, and W. Medjroubi. *SciGRID_gas: The combined IGG gas transmission network data set*. DLR-Institut für Vernetzte Energiesysteme e.V., Oldenburg, Germany, Aug 2020. URL: <https://doi.org/10.5281/zenodo.4009129>, doi:10.5281/zenodo.4009129.
- [DPM20b] J.C. Diettrich, A. Pluta, and W. Medjroubi. *SciGRID_gas: The combined IGGI gas transmission network data set*. DLR-Institut für Vernetzte Energiesysteme e.V., Oldenburg, Germany, Aug 2020. URL: <https://zenodo.org/record/4288468#.YG7aBz9CSUk>, doi:10.5281/zenodo.4288467.
- [DPM20c] J.C. Diettrich, A. Pluta, and W. Medjroubi. *SciGRID_gas: The combined IGGIELGN gas transmission network data set*. DLR-Institut für Vernetzte Energiesysteme e.V., Oldenburg, Germany, 2020. URL: <https://zenodo.org/record/4642569#.YG7ZhT9CSUk>, doi:10.5281/zenodo.4642568.
- [DPM20d] J.C. Diettrich, A. Pluta, and W. Medjroubi. *SciGRID_gas: The combined IGGINL gas transmission network data set*. DLR-Institut für Vernetzte Energiesysteme e.V., Oldenburg, Germany, 2020. URL: <https://zenodo.org/record/4288440#.YG7aHz9CSUk>, doi:10.5281/zenodo.4288439.
- [DPM20e] J.C. Diettrich, A. Pluta, and W. Medjroubi. *SciGRID_gas: The raw INET data set*. DLR-Institut für Vernetzte Energiesysteme e.V., Oldenburg, Germany, Aug 2020. URL: <https://doi.org/10.5281/zenodo.3985249>, doi:10.5281/zenodo.3985249.
- [DPM20f] J.C. Diettrich, A. Pluta, and W. Medjroubi. *SciGRID_gas: The raw LKD data set*. DLR-Institut für Vernetzte Energiesysteme e.V., Oldenburg, Germany, Aug 2020. URL: <https://doi.org/10.5281/zenodo.3985271>, doi:10.5281/zenodo.3985271.
- [Ent17] EntsoG. North West GRIP, Min Report. https://www.entsog.eu/sites/default/files/files-old-website/publications/GRIPs/2017/entsog_GRIP_NW_2017_main_xs.pdf, 2017. Accessed: 2021-03-17.
- [Gas19] GasTerra. Aardgas in Nederland. <https://www.gasterra.nl/uploads/fckconnector/20d5b75a-2ad6-529e-b024-1f5bd43bbd5e>, 2019. Accessed: 2021-03-17.
- [Hel18] D. Helle. OpenStreetMap - Deutschland. <https://www.openstreetmap.de/>, 2018. Accessed: 2019-12-12.

- [Kha13] Y. Khalid. What is Pickle in python? <https://pythontips.com/2013/08/02/what-is-pickle-in-python/>, 2013. Accessed: 2019-10-10.
- [KingSpalding18] King&Spalding. LNG in Europe 2018: An Overview of LNG Import Terminals in Europe. <https://globalnghub.com/wp-content/uploads/2018/09/King.pdf>, 2018. Accessed: 2018-09-01.
- [KKS+17] F. Kunz, M. Kendziorowski, W.-P. Schill, J. Weibezahn, J. Zepter, C. von Hirschhausen, and P. Hauser. *Electricity, Heat, and Gas Sector Data for Modeling the German System*. Deutsches Institut für Wirtschaftsforschung, Daten Dokumentation 92, Berlin, 2017.
- [LSS+19] P. Lustenberger, F. Schumacher, M. Spada, P. Burgherr, and B. Stojadinovic. Assessing the performance of the european natural gas network for selected supply disruption scenarios using open-source information. *Energies*, 12(4685):1–28, 2019. doi:{ 10.3390/en12244685 }.
- [MMK16] C. Matke, W. Medjroubi, and D. Kleinhans. SciGRID - An Open Source Reference Model for the European Transmission Network (v0.2). <https://power.scigrid.de>, 2016. Accessed: 2019-09-09.
- [Mic20] Microsoft. Bing maps api. <https://www.microsoft.com/en-us/maps/licensing>, 2020. Accessed: 2018 to 2021.
- [Nis20] A. Nisbet. Open topo data api. <https://www.opentopodata.org>, 2020. Accessed: 2020 to 2021.
- [San19] B. Sandvik. World Borders. http://thematicmapping.org/downloads/world_borders.php, 2019. Accessed: 2019-07-07.
- [SAB+17] M. Schmidt, D. Aßmann, R. Burlacu, J. Humpola, I. Joormann, N. Kanelakis, T. Koch, D. Oucherif, M.E. Pfetsch, L. Schewe, R. Schwarz, and M. Sirvent. *GasLib—A Library of Gas Network Instances*. 2017. doi:{ 10.3390/data2040040 }.
- [sl19] scikit-learn. 1.1. Linear Models (scikit learn). https://scikit-learn.org/stable/modules/linear_model.html, 2019. Accessed: 2019-08-08.
- [UoO14] USA University of Oregon. Comparing distributions: Z Test. <http://homework.uoregon.edu/pub/class/es202/ztest.html>, 2014. Accessed: 2020-07-07.
- [Wik20a] Wikipedia. Bootstrapping (statistics). [https://en.wikipedia.org/wiki/Bootstrapping_\(statistics\)](https://en.wikipedia.org/wiki/Bootstrapping_(statistics)), 2020. Accessed: 2019-06-06.
- [Wik20b] Wikipedia. Cross-validation (statistics). [https://en.wikipedia.org/wiki/Cross-validation_\(statistics\)#Exhaustive_cross-validation](https://en.wikipedia.org/wiki/Cross-validation_(statistics)#Exhaustive_cross-validation), 2020. Accessed: 2019-07-07.
- [Wik20c] Wikipedia. Jackknife resampling. https://en.wikipedia.org/wiki/Jackknife_resampling, 2020. Accessed: 2019-08-08.
- [Wik20d] Wikipedia. Lasso (statistics). [https://en.wikipedia.org/wiki/Lasso_\(statistics\)](https://en.wikipedia.org/wiki/Lasso_(statistics)), 2020. Accessed: 2020-04-04.
- [Wik20e] Wikipedia. Limited-memory BFGS. https://en.wikipedia.org/wiki/Limited-memory_BFGS, 2020. Accessed: 2020-06-06.
- [Wik20f] Wikipedia. Out-of-bag error. https://en.wikipedia.org/wiki/Out-of-bag_error, 2020. Accessed: 2019-07-07.
- [Wik20g] Wikipedia. Transmission system operator. https://en.wikipedia.org/wiki/Transmission_system_operator/, 2020. Accessed: 2019-09-09.
- [Wik20h] Wikipedia. JAGAL. <https://en.wikipedia.org/wiki/JAGAL>, 2020. Accessed: 2020-01-01.
- [Wik21] Wikipedia. Nomenclature of territorial units for statistics. https://en.wikipedia.org/wiki/Nomenclature_of_Territorial_Units_for_Statistics, 2021. Accessed: 2021-04-10.
- [BMW11] BMWi. Forschung für eine umweltschonende, zuverlässige und bezahlbare Energieversorgung. https://www.bmwi.de/Redaktion/DE/Publikationen/Energie/6-energieforschungsprogramm-der-bundesregierung.pdf?__blob=publicationFile&v=12, 2011. Accessed: 2019-02-02.

- [BMWi20] BMWi. Home page of BMWi. <https://www.bmwi.de/Navigation/DE/Home/home.html>, 2020. Accessed: 2020-03-03.
- [BundesregierungDeutschland20] Bundesregierung Deutschland. Home page of Bundesregierung Deutschland. https://www.bundesregierung.de/Webs/Breg/DE/Themen/Energiewende/_node.html, 2020. Accessed: 2020-01-01.
- [EntsoG20] EntsoG. Home page of EntsoG. <https://www.entsog.eu/>, 2020. Accessed: 2020-03-03.
- [GasIEurope20] Gas Infrastructure Europe. Home page of Gas Infrastructure Europe. <https://agsi.gie.eu>, 2020. Accessed: 2020-01-01.
- [GasSEurope20] Gas Storages Europe. Home page of Gas Storages Europe. <https://www.gie.eu/index.php/transparency/gse-transparency-template>, 2020. Accessed: 2020-01-01.
- [Gassco20a] Gassco. Data page of facilities from Gassco. <https://www.npd.no/en/about-us/information-services/available-data/map-services/>, 2020. Accessed: 2020-01-01.
- [Gassco20b] Gassco. Home page of Gassco Norway. <https://www.gassco.no/en/>, 2020. Accessed: 2020-01-01.
- [IGU20] IGU. Home page of International Gas Union. <https://www.igu.org/>, 2020. Accessed: 2018-10-01.
- [nationalGrid20] nationalGrid. Home page of National Grid UK. <https://www.nationalgrid.com/uk/>, 2020. Accessed: 2018-10-01.