SciGRID_gas - Data Model of the European Gas Transport Network

Adam Pluta, Wided Medjroubi, Jan C. Diettrich, Jan Dasenbrock, Hendrik-Pieter Tetens Javier E. Sandoval and Ontje Lünsdorf
Institute of Networked Energy Systems
German Aerospace Center (DLR)
Oldenburg, Germany
Adam.Pluta@dlr.de

***Abstract*—The current transition in the European energy sector towards climate neutrality requires detailed and reliable energy system modeling. The quality and relevance of the energy system modeling highly depend on the availability and quality of model input datasets. However, detailed and reliable datasets are still missing, especially for the gas infrastructure. In this contribution, we present our approach for developing an open-source model of the gas transport network in Europe. Various freely available data sources were used to collect gas transport data. The datasets from multiple sources were merged, and further, statistical methods were used to generate missing data.**

As a result, we successfully created a gas transport network model only using open-source data. The SciGRID_gas model contains 206,000 km of pipeline data which is roughly in accordance to former estimations. In addition, datasets of compressor stations, LNG terminals, storages, production sites, gas power plants, border points, and demand time series are provided. Finally, we have discussed data gaps and how they can potentially be closed.

***Index Terms*—gas transport network, openstreetmap, data model, gas pipelines**

I. Introduction

The most significant shares of gross energy consumption in Europe in 2019 were held by oil and petroleum products (34.5 %), followed by natural gas (23.1 %) representing almost 60 % of energy consumption [Eur21b]. The high share of natural gas in the energy mix reflects its essential role as an energy carrier and the need to decarbonize sectors where it is used as a fuel source. This will be achieved mainly by ramping up the integration of renewable energy sources (RES) in the energy system. In this context, new flexibility concepts are needed to integrate higher RES shares while maintaining energy supply stability and reliability. Such flexibilities are the Power-to-X (P2X) technologies combined with energy storage [SB15]. P2X refers to various processes to convert and store electricity using surplus RES electric power to seasonally and spatially balance energy. A promising P2X technology is power-to-gas (P2G) which has a vast potential in decarbonizing different energy sectors such as heating and transport. P2H uses surplus electric power to produce hydrogen via water electrolysis [SSZ21], [KKT+21], [SGR+15]. Hydrogen can then be used as a fuel directly or converted to LPG, syngas, or methane. The produced gas can then be transported using the current gas transport network.

Despite the importance of modeling and analysis of the gas sector and its interactions with other energy sectors, there exist no reliable open-source datasets for the European gas transport network. Examples of available datasets are limited to single countries and do not provide details of the grid components. Such examples are the LKD-EU dataset for Germany [KKS+17] and the National Grid dataset for the UK [nat20]. The lack of complete and coherent datasets motivated us to initiate the SciGRID_gas project [PDM18] at the DLR Institute of Networked Energy Systems. The goal of the SciGRID_gas project is to derive a reliable and detailed dataset for the European gas transport grid, which can be used for modeling and analysis purposes. In practice, the source code of the data model, the geo-referenced datasets describing the gas transport grid as well as the documentation, are made available under the CC-BY open source licences. In order to use SciGRID_gas in the simulation of energy systems, the integration of the datasets in existing energy system models such as open_eGo, PyPSA, and pandapipes is already in work. With the SciGRID_gas data model, we would like to answer the following research questions:

- Can we build a comprehensive data model for the European gas transport grid using only publicly available data?
- Is the amount of available parameter data sufficient to estimate missing parameter data via statistical methods?

This contribution is structured as follows: In Chapter II, we discuss the data sources used for constructing the open-source gas transport network model. This is followed by the discussion of the model architecture in Chapter III. Chapter IV gives a short overview of some suitable methods for creating the model. Due to the page number limitation, more detailed information is also available in the respective model documentation, which is accessible online [PDM18]. Chapter V presents the graphical and statistical results of our model. This is followed by the discussion in Chapter VI and the conclusion and outlook in Chapter VII.

II. Data sources

Obtaining reliable open-source data of the gas transport system is a challenging task. The grid data of gas Transmission System Operators (TSOs[1]) are commonly not standardized, nor are they freely accessible [ENT20]. Data are generally not geo-referenced and mainly available as PDF maps. Most individual TSOs are not willing to share their data due to competitive reasons. Within the SciGRID_gas project, we have gathered freely available data from different sources. The most relevant are presented below. In the name of the subsections, we first indicate the source of the dataset (e.g., web search) followed by the name (e.g., INET) we gave the dataset in the SciGRID_gas project. The data sources are split into the categories "CC-BY compatible" and "non-CC-BY compatible". The CC-BY license allows the user to redistribute and edit the data as long as an appropriate reference to the license and a notice that changes have been made are included. Data compatible with CC-BY could be easily included in the SciGRID_gas dataset. Non-CC-BY compatible data has only been used for statistical analysis in the attribute generation process but was not included directly.

A. CC-BY Compatible

1) Web search - INET: We carried out a web search on all gas network components and compiled the gathered data into the INET dataset. The data stems from TSO press releases, TSO transparency platforms, and TSO public data. Some TSO information was available due to mandatory EU regulations [The13]. Other information was made public as part of a company's self-presentation and advertisement. The collected information contains data on network components, including their positions and relevant energy modeling parameters, such as diameter, capacity, power, pressure, etc.

2) German gas model - LKD: The "long-term planning and short-term optimization" dataset (LKD) [KKS+17] contains geo-referenced data on gas facilities in Germany. It was created by several German research institutes and includes information on gas pipelines, production sites, storage, compressor locations, and nodes. The SciGRID_gas project was gratefully granted the right to use, change and redistribute the LKD data under an open license.

3) ENTSOG - EMAP: The development of the European gas transport network is facilitated and enhanced by the European Network of Transmission System Operators for Gas (ENTSOG)[2]. The ENTSOG is an association of 44 European TSOs, three associated partners, and nine observers. ENTSOG members are required to publish certain information according to EU directives. A significant amount of this information is incorporated into the freely available and regularly updated map of the gas pipelines, drilling platforms, and storage facilities. The SciGRID_gas project extracted the course and location of the depicted gas pipelines, storage, and production facilities from the ENTSOG map[3] of 2019 using Python and QGIS.

4) Eurostat - Cons: The European Statistical Office (Eurostat) collects and publishes data on energy supply, transformation, and consumption on a monthly and annual basis. Eurostat statistics [Eur21a] and others provided the data foundation for our study regarding the European gas demand. We derived the daily time series concerning the seasonal, geographical, and sector-specific variability of the gas demand in Europe [San21] on NUTS 3 level[4] for 27 European countries covering the years 2010 to 2019. To provide detailed information for modelers and dataset flexibility, the time series distinguishes between the sectors *households*, *commercial* and *industry*. Figure 1 provides an exemplary data plot for the annually-averaged residential gas demand in Europe disaggregated into NUTS 3 regions. The data shows good benchmarking results against three existing time series of gas demand in Germany originating from the DemandRegio project [Got20].

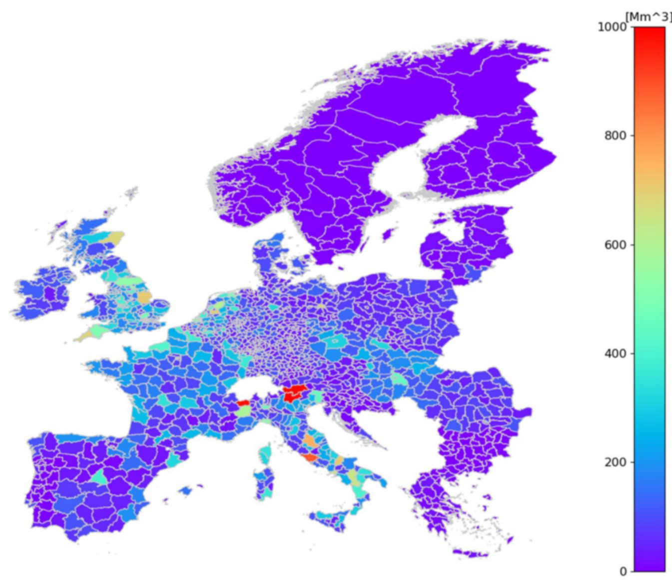

Fig. 1. Annually averaged gas demand in the residential sector between 2010-2019, disaggregated into NUTS 3 regions. [San21].

B. Non-CC-BY Compatible

1) OpenStreetMap - OSM-SciGRID: OpenStreetMap (OSM) [Hel18] is a freely adjustable and accessible geo-data database with steadily increasing data coverage and data quality. OSM data has the downside of being licensed under ODbL, which is not compatible with the CC-BY license, a hurdle that can be overcome using a collective database. In the past, OSM data has contributed to the field of energy

[1]The European operators of the gas transmission (transport) grid are associated in the European Network of Transmission System Operators for Gas (ENTSOG)

[2]https://www.entsog.eu/

[3]https://www.entsog.eu/maps

[4]NUTS is a geographical system to divide the EU territory into hierarchical levels. In Germany, NUTS 1 are federal States, NUTS 2 are governmental regions, and NUTS 3 are the districts.

system modeling, for example, in the creation of power grid models [MMS+17] or the optimization of flexibility options for urban areas [AMVA17a], [AMVA17b]. With the `esy-osmfilter` [PL20a] we used a Python library to easily access and filter data from OpenStreetMap. We have used this library for the creation of `osmscigrid` [PL20b], another library, which is capable of converting OSM pipeline data directly into SciGRID_gas data format for easier integration of OSM datasets. Pluta and Lünsdorf [PL20a] have stated that the European gas pipeline data content from OSM was rapidly growing between 2014-2019. However, we saw that this trend has stagnated in the last two years, possibly due to the current COVID-19 pandemic. In today's OSM gas pipeline data, we still see significant data gaps. Moreover, data relating to system-relevant components like compressors stations or storage is still missing. Thus, we have decided that the current data will not be used directly in our model but only to validate the topology of pipelines.

2) Norway - NO : Norway's primary national pipeline operator Gassco [Gas20b] publishes precise geo-referenced non-infield gas and oil facility data. Next to the location of the pipeline segments, the data includes additional information like the pipeline diameter. The data is published under an open-source license. However, Gassco needs to be cited for using this data.

3) Great Britain - GB: The British TSO provides shapefiles of its transport gas network on their website [nat20]. Their data is restricted to noncommercial purposes.

4) International Gas Union - IGU: The International Gas Union (IGU) [IGU20] is a worldwide operating non-profit organization that maintains a dataset, which is accessible through its public internet appearance. We developed a data scraping tool to automatically download project-relevant data from their HTML pages.

5) Gas Infrastructure Europe - GIE: Gas Infrastructure Europe (GIE) is an organization representing the interests of the gas infrastructure industry. As of today, it has 68 members in 25 countries in Europe. It is an umbrella organization for *Gas Storage Europe* representing the Storage System Operators, Gas LNG Europe representing the LNG Terminal Operators (TO), and Gas Transmission Europe representing the Transmission System Operators (TSO) [Gas20a]. The GIE API has been used to retrieve static data and time-series data of Gas Storages and LNG Terminals. Time-series data were used to determine the maximum working gas volume in the LNG storage tanks and the maximum and medium storage to gas pipeline gas flows.

III. Data Model Architecture

The SciGRID_gas data model network consists of several component classes, each representing a list of objects. The following component classes have been implemented: *PipeSegments* (PS), *BorderPoints* (BP), *Compressors* (CS), *LNGs* (LNG), *PowerPlants* (PP), *Productions* (PO), *Consumers* (CO), *Storages* (ST). Any object which is a member of a component class is defined as an element of that respective class and can therefore be described by a common component-specific set of attributes. Each attribute by itself consists of the value, the error, and the method used to generate the value. To make this data structure suitable for modeling gas networks, we need to restructure the data using nodes and edges. These are connected to dataset elements by a unique ID. In that way, all components except for *PipeSegments* are implemented as nodes. *PipeSegments* go from one node to another and are therefore implemented as edges. The exact geographical path of each *PipeSegment* is stored in a list of intermediate pipeline points and attached as an attribute of that *PipeSegment*. We have created and released datasets from the various data sources mentioned in Chapter II on the project website [PDM18] under the section *downloads*, which were converted from their original format to the SciGRID_gas format described in Chapter III. Table I gives an overview of the data sources constituting the final dataset, named IGGIELGNC-1.

TABLE I
OVERVIEW OF THE AVAILABLE DATA SOURCES FOR DIFFERENT GAS TRANSPORT COMPONENTS.

component	**data source**
PipeSegments	INET, EMAP, LKD, GB, NO
Nodes	INET, EMAP, LKD, GIE, CONS
LNGs	INET, GIE
Storages	GIE, GSE, LKD, EMAP, IGU
PowerPlants	INET, CONS
Productions	INET, EMAP, LKD
Compressors	INET, LKD
Borderpoints	INET
Consumers	INET, CONS

IV. Methodology

In this section, we describe our methodology to create a comprehensive gas network dataset. This addresses, in particular, the merging of datasets from various sources, the creation of a merged network dataset, the post-processing, and the visualization.

A. Data Merging Process

The task of merging the various datasets required identifying duplicate elements that may exist in more than one dataset. For this task, we relied on the criteria of spatial and name similarity. For the latter, we were using the `fuzzywuzzy` Python package [Inc14]. The Python algorithm will pairwise evaluate the identity of all objects with an identity score between 0 and 100, where 0 indicates no similarity and 100 indicates apparent duplicity. Elements are merged if they exceed a component-specific threshold between 80 and 95. In case that the likely duplicates do not share the same attributes, the attributes of the subjectively most trustworthy source from Chapter II are adopted.

This process works for all components which are implemented as nodes. The process of merging pipelines is more complex. For edges, the process is built around a similarity check of the start and end node positions, as well as comparisons of the diameter, pressure, capacity, and length values. The respective algorithm is described in more detail in the documentation of the final dataset on the website [PDM18]. We have used this algorithm to transfer pipeline attributes from INET, LKD, and NO datasets to their respective counterpart in the EMAP pipeline data, which we have used as a basis for the final dataset. Pipelines with no counterpart in EMAP have also been added to this data.

B. Attribute Generation

In general, all preexisting attribute values are classified as "raw", and the corresponding error is set to zero. Once a merged dataset has been compiled, we focus on estimating missing data on the attribute level. Depending on the specific attribute, various approaches produced different results to estimate missing values. Our approach was to exploit linear relations between different attribute type pairs. For each pair of attributes, we have simulated all possible correlation functions (linear, squared, log, mean, min, max, median) and then selected the one with the smallest error. For this purpose, we have used the Lasso-linear regression method from `scikit-learn` [sl19]. In the end, we did a Z-score value comparison between the input data and the estimated values, validating the similarity between the input distribution and distribution of estimated values. However, meaningful linear correlations were often not identifiable or suffered from insufficient data density. Thus, the method applied the mean or median in such cases.

C. Post-Processing and Visualization

Finally, we have added artificial *Consumer* nodes to the network and connected them to the nearest pipeline. We have compiled the final dataset for the three different consumer aggregation levels: NUTS 1, NUTS 2, and NUTS 3, which resulted in the datasets: IGGIELGNC-1 [DPS+21a], IGGIELGNC-2 [DPS+21b] and IGGIELGNC-3 [DPS+21c], respectively. Further, some cleanup routines have been implemented, e.g., removing or connecting unconnected isolated elements to create a coherent network. Also, during post-processing, the elevation of each node was determined with the help of *Bing Maps Elevation API* [Mic20]. Additionally, we have released `qplot` [Plu20] library for the visualization SciGRID_gas data. The library was used for the creation of Figures 2–4.

V. Results

We have released our final dataset under the name IGGIELGNC-1 [DPS+21a]. The data is licensed under CC-BY and available in CSV and GeoJSON formats, and accompanied by methodical documentation. We want to emphasize that all results stem from version 1.1 of our dataset and that the data is subject to changes with future updates.

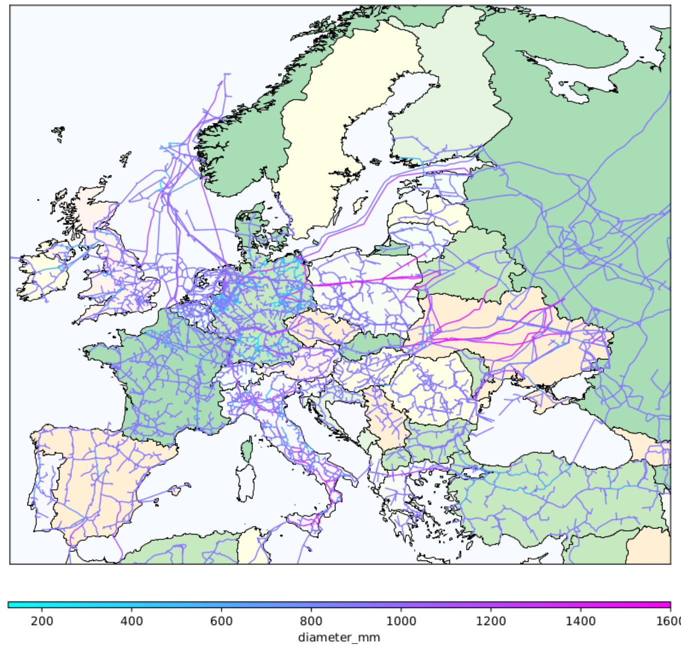

Fig. 2. The pipeline system of the European gas transport grid of the IGGIELGNC-1 dataset. Pipelines are colored according to their diameter values.

First, we have used pipeline routes from EMAP with a total length of about 207,000 km and added the attribute information from the INET pipelines (60,000 km) and LKD (27,000 km), NO (9,000 km) and GB (8,000 km) to this data. The missing pipeline information has been then generated by our statistical attribution generation process. This pipeline network is plotted together with the respective diameters in Fig. 2. For comparison, the extrapolated pipeline validation dataset from OSM in 2020 only contains a total length of 108,000 km, which is only about 52 % of the total pipeline length of the IGGIELGNC-1 dataset. In Fig.3 we present our final grid, containing 109 BP, 248 CS, 32 LNG, 314 PP, 102 PD, 108 CO, and 294 ST (see Section III for the nomenclature). A country-wise overview of all components for some EU countries is shown in Table II.

TABLE II
TOTAL PIPELINE LENGTH IN KM AND THE COUNT OF ELEMENTS PER COMPONENT CLASS FOR SOME EUROPEAN COUNTRIES. DATA SOURCE: IGGIELGNC-1

country code	PS length	CS count	LNG count	BP count	ST count	CO count
AT	2,451	7	0	4	15	3
BE	2,312	6	1	6	1	3
CH	1,012	1	0	2	0	1
CZ	2,159	6	0	3	10	1
DE	27,708	35	0	15	68	15
DK	841	1	0	1	3	1
ES	8,389	18	7	4	8	5
FR	15,424	40	4	7	23	12
GB	6,836	28	4	4	22	11
IT	12,053	14	4	6	20	4

In Tab. III we address the attribute data density for different component classes. We have chosen up to three of the most

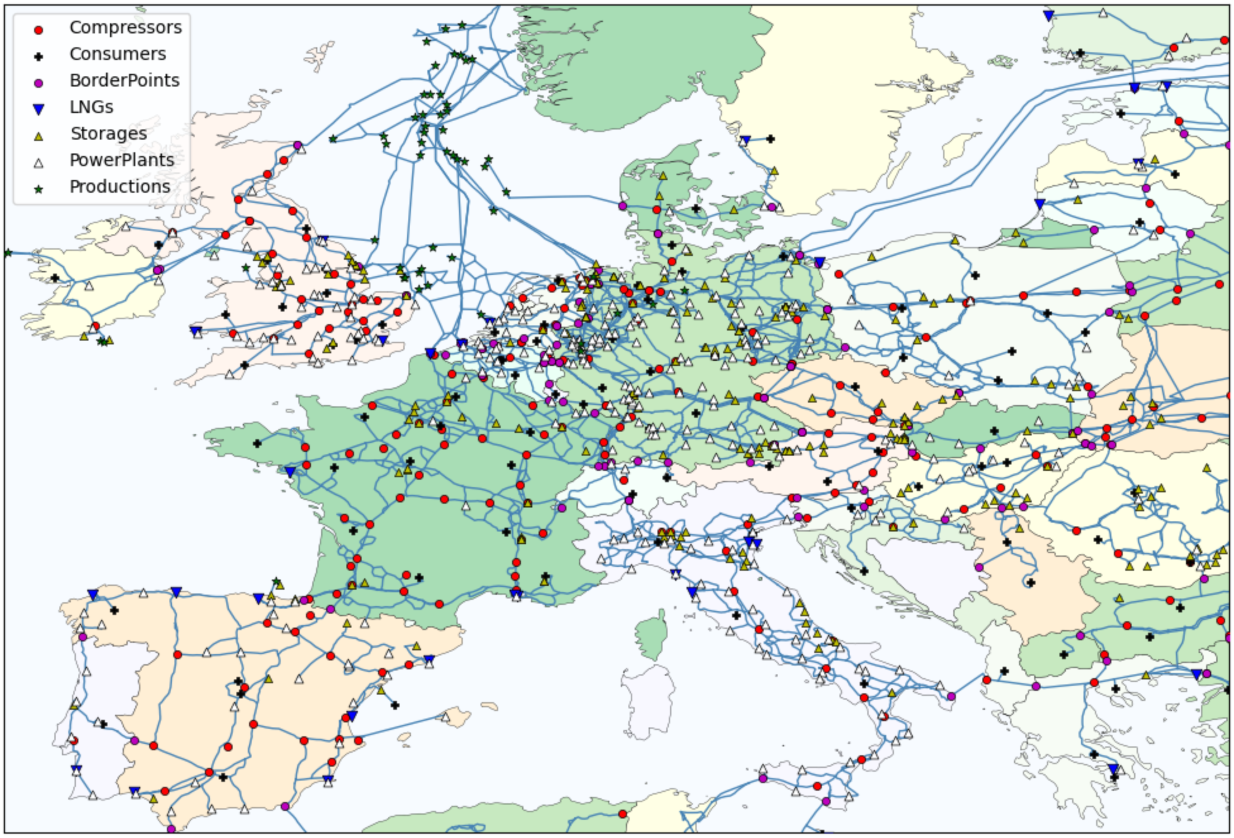

Fig. 3. Extract of the European gas network of the IGGIELGNC-1 dataset with all its components.

relevant attributes and determined the respective percentage of raw data availability. We have not considered *BorderPoints* or *Consumers* for this analysis as their attributes resulted from a single dataset and were mostly complete or fully derived.

TABLE III
RAW PARAMETER DENSITY OF THE IGGIELGNC-1 DATASET.

component	**density**	**density**	**density**
PipeSegments	capacity 13%	diameter 32%	pressure 19%
LNGs	capacity 94%	size 69%	-
Storages	capacity 63%	power 28%	pressure 35%
Compressors	capacity 7%	power 15%	pressure 7%
PowerPlants	power 100%	-	-
Productions	supply 5%	-	-

Due to a current lack of a validation data, a profound validation of our attribute data estimation techniques is still pending. However, we have at least partially validated the topology of our pipeline network topology against the assumed complete OSM pipeline data for the region of Spain, illustrated in Fig. 4. After removing all pipelines outside of Spain from both datasets, we calculate the "Hänsel and Gretel" distance [AFW14] with the result that 88 % of all pipelines in IGGIELGNC-1 are within a 15 km radius of the pipelines in OSM.

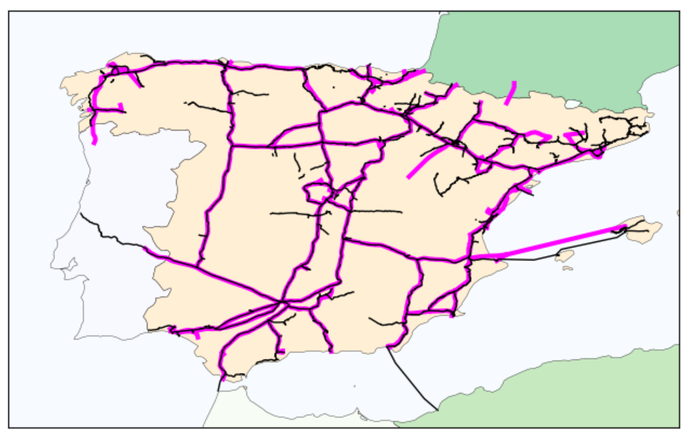

Fig. 4. Comparison of pipelines from OSM-SciGRID (black, 10,658 km) and IGGIELGNC-1 (magenta, 8,369 km).

VI. DISCUSSION

We presented our approach of creating a gas transport network data model of Europe from open-source data. In terms of pipelines, our data model has a total length of about 206,000 km, which is roughly in accordance with the commonly assumed length of 200,000 km [CLM+14]. The slight overestimation is probably attributed to our broader definition of the European gas network, which also incorporates the Western part of Russia, some Northern African States, and Turkey. However, the total length is a good indicator for the overall success of modeling the European gas transport network from open-source data. Further, we have used OpenStreetMap data to validate our network's topology. Our data show good accordance in this regard. The examination of other components like *PowerPlants* and *Productions* has shown data gaps in some regions stemming from incomplete data sources. However, the validation of our statistical attribute generation methods, which are described in more detail in the dataset documentation, is currently not feasible due to a lack of accessible validation datasets. For that reason, we instead want to discuss the potential accuracy of such methods. The results of any attribute generation method, designed to predict partially unknown data, will scale in accuracy with the percentage of known data. We have analyzed our data regarding important parameters of different network components. Therefore, we can state that the generated attributes for *LNGs*, *Storages*, and *PowerPlants* are more trustworthy than for *PipeSegments*, *Compressors*, and *Productions*. It can be assumed that especially the use of heuristics could lead to more accurate results. For example, in the case of missing pipeline capacity values, one could use the capacity of an adjacent compressor station to derive this value under the consideration of all other incoming and outgoing pipelines.

From our perspective, more focus needs to be put into the data acquisition for these components. Such data might be provided by OpenStreetMap (OSM) soon, as we stated earlier. If this trend continues or will even be supported by TSOs, OSM might become a good source for this data. At some point, it might even be possible to create an entire network from OSM data like it was done for the open-source power transmission dataset SciGRID_power [MMK16]. The reason why this is currently not possible for the gas grid is mainly rooted in the fact that gas transport pipelines are buried underground. This makes the direct identification of their position and additional properties difficult for OpenStreetMap mappers. In order to address missing topological data in datasets like OSM, one can use the approach suggested by Dasenbrock et al. [DPZM21]. There it was demonstrated that by using satellite imagery taken during gas pipeline construction phase and AI, the topological pathway of gas pipelines could be determined.

VII. Conclusion and Outlook

This contribution shows that the creation of a comprehensive European open-source gas transport data and network model from publicly available data sources is possible. However, the quality of estimating missing data with statistical methods strongly depends on the original data density, which varies for different component classes and for different geographical regions. We believe that our data model is a valid approximation of the European gas transport network. Nonetheless, a definite quality assessment was not feasible during the project due to a lack of verification data. Such validation and quantification of uncertainty are necessary to make our model a reliable basis for energy system design. Therefore, we are looking for ways to advance the verification of this data. This should also include a manual process to inspect or even merge data as in principle the number of nodes in the European gas transport grid would make this approach feasible.

We believe that the creation of such an open-source model can potentially encourage TSOs to make their data open-source, which in the long term will result in a more precise representation of the gas grid and more suited energy scenarios. Our analysis of the underlying component attribute data showed that gas transport pipelines and compressor station data show low density in terms of their main attributes. Since both components are critical in modeling gas flows, future work will hopefully diminish these data gaps. We have discussed how some data gaps can potentially be closed either by a steady growth of OSM data or by remote sensing methods. Yet, some data, especially on the pipeline materials and roughness values, are not accessible without the assistance of TSOs.

After all, we have to rely on the modeling community feedback. At this point, we are internally discussing a way forward to create a sustainable process for data correction.

Funding

This research was funded under the “SciGRID_gas” project by the German Federal Ministry for Economic Affairs and Energy (BMWi) within the funding of the 6. Energieforschungsprogramm der Bundesregierung. Funding Code: 03ET4063.

References

[AFW14] Mahmuda Ahmed, Brittany Terese Fasy, and Carola Wenk. Local persistent homology based distance between maps. In *Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 43–52, 2014.

[AMVA17a] A. Alhamwi, W. Medjroubi, T. Vogt, and C. Agert. GIS-based urban energy systems models and tools: Introducing a model for the optimisation of flexibilisation technologies in urban areas. *Applied Energy*, 191:1–9, apr 2017.

[AMVA17b] A. Alhamwi, W. Medjroubi, T. Vogt, and C. Agert. Openstreetmap data in modelling the urban energy infrastructure: a first assessment and analysis. In *Proceedings of the 9th International Conference on Applied Energy*, volume 142, page 1968–1976. Elsevier, 2017.

[CLM+14] D. Connolly, H. Lund, B.V. Mathiesen, S. Werner, B. Möller, U. Persson, T. Boermans, D. Trier, P.A. Østergaard, and S. Nielsen. Heat Roadmap Europe: Combining district heating with heat savings to decarbonise the EU energy system. *Energy Policy*, 65:475–489, 2014.

[DPS+21a] J.C. Diettrich, A. Pluta, J.E. Sandoval, J. Dasenbrock, and W. Medjroubi. SciGRID_gas: The final IGGIELGNC-1 gas transmission network data set. https://doi.org/10.5281/zenodo.5017621, July, 2021.

[DPS+21b] J.C. Diettrich, A. Pluta, J.E. Sandoval, J. Dasenbrock, and W. Medjroubi. Scigrid_gas: The final IGGIELGNC-2 gas transmission network data set. https://doi.org/10.5281/zenodo.5017641, July, 2021.

[DPS+21c] J.C. Diettrich, A. Pluta, J.E. Sandoval, J. Dasenbrock, and W. Medjroubi. SciGRID_gas: The final IGGIELGNC-3 gas transmission network data set. https://doi.org/10.5281/zenodo.4922529, July, 2021.

[DPZM21] J. Dasenbrock, A. Pluta, M. Zech, and W. Medjroubi. Detecting pipeline pathways in landsat 5 satellite images with deep learning. *Energies*, 14(18), 2021.

[ENT20] ENTSOG. Home page of ENTSOG. https://www.entsog.eu/, 2020. Accessed: 2020-03-03.

[Eur21a] EuroStat. Complete energy balances. https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=nrg_bal_c&lang=en, 2021. Accessed: 2021-02-08.

[Eur21b] Eurostat. Complete energy balances: Gross inland energy consumption. https://ec.europa.eu/eurostat/databrowser/view/nrg_bal_c/default/table?lang=en, 2021. Online data code: NRG_BAL_C. Last update: 06/06/2021-23:00. Accessed: 2020-01-01.

[Gas20a] Gas Infrastructure Europe. Home page of Gas Infrastructure Europe. https://agsi.gie.eu, 2020. Accessed: 2020-01-01.

[Gas20b] Gassco. Home page of Gassco Norway. https://www.gassco.no/en/, 2020. Accessed: 2020-01-01.

[Got20] Gotzens, F. and Gillessen, B. and Burges, S. and Hennings,W. and Müller-Kirchenbauer, J. and Seim, S. and Verwiebe, P.and Schmid, T. and Jetter, F. and Limmer, T. DemandRegio: Harmonisierung und Entwicklung von Verfahren zur regionalen und zeitlichen Auflösung von Energienachfragen, 2020.

[Hel18] D. Helle. OpenStreetMap - Deutschland. https://www.openstreetmap.de/, 2018. Accessed: 2019-12-12.

[IGU20] IGU. Home page of International Gas Union - About. https://www.igu.org/about/, 2020. Accessed: 2021-10-21.

[Inc14] SeatGeek Inc. *fuzzywuzzy: Fuzzy String Matching in Python*, 2014.

[KKS+17] F. Kunz, M. Kendziorski, W.-P. Schill, J. Weibezahn, J. Zepter, C. R. von Hirschhausen, P. Hauser, M. Zech, D. Möst, S. Heidari, et al. Electricity, heat, and gas sector data for modeling the german system. Technical report, DIW Data Documentation, 2017.

[KKT+21] G. Kakoulaki, I. Kougias, N. Taylor, F. Dolci, J. Moya, and A. Jäger-Waldau. Green hydrogen in europe–a regional assessment: Substituting existing production with electrolysis powered by renewables. *Energy Conversion and Management*, 228:113649, 2021.

[Mic20] Microsoft. Bing maps api. https://www.microsoft.com/en-us/maps/licensing, 2020. Accessed: 2018 to 2021.

[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.

[MMS+17] W. Medjroubi, U. P. Müller, M. Scharf, C. Matke, and D. Kleinhans. Open data in power grid modelling: New approaches towards transparent grid models. *Energy Reports*, 3:14 – 21, 2017.

[nat20] nationalGrid. Home page of National Grid UK. https://www.nationalgrid.com/uk/, 2020. Accessed: 2018-10-01.

[PDM18] A. Pluta, J.C. Diettrich, and W. Medjroubi. SciGRID_gas - open source reference model of european gas transport networks for scientific studies on sector coupling. https://gas.scigrid.de, 2018. Accessed: 2019-09-09.

[PL20a] A. Pluta and O. Lünsdorf. esy-osmfilter: A python library to efficiently extract openstreetmap data. *Journal of Open Research Software*, 8, 2020.

[PL20b] A. Pluta and O. Lünsdorf. *geopy: Python library to create a SciGRID_gas Pipeline dataset from OpenStreetMap*, 2020.

[Plu20] A. Pluta. qplot - SciGRID_gas Visualisation Library. https://gitlab.com/dlr-ve-esy/qplot, 2020.

[San21] J.E. Sandoval. Estimation and simulation of a gas demand time series for the european nuts 3 regions. Master's thesis, Carl von Ossietzky Universität Oldenburg, Germany, Fak. 5, Institute of Physics (PPRE) D-26111 Oldenburg, Germany, 6 2021. Supervised by Prof. Dr. Carsten Agert, Dr. Herena Torio and Dr. Wided Medjroubi.

[SB15] A. Sternberg and A. Bardow. Power-to-what?–environmental assessment of energy storage systems. *Energy & Environmental Science*, 8(2):389–400, 2015.

[SGR+15] S. Schiebahn, T. Grube, M. Robinius, V. Tietze, B. Kumar, and D. Stolten. Power to gas: Technological overview, systems analysis and economic assessment for a case study in Germany. *International Journal of Hydrogen Energy*, 40(12):4285–4294, 2015.

[sl19] scikit learn. 1.1. Linear Models (scikit learn). https://scikit-learn.org/stable/modules/linear_model.html, 2019. Accessed: 2019-08-08.

[SSZ21] F. Stöckl, W.P. Schill, and A. Zerrahn. Optimal supply chains and power sector benefits of green hydrogen. *Scientific Reports*, 11(1):1–14, 2021.

[The13] The European Commission. Commission Regulation (EU) No 543/2013 of 14 June 2013 on submission and publication of data in electricity markets and amending Annex I to Regulation (EC) No 714/2009 of the European Parliament and of the Council. https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:163:0001:0012:EN:PDF, 2013.