Data Article

A global inventory of electricity infrastructures from 1980 to 2017: Country-level data on power plants, grids and transformers

Gerald Kalt\textsuperscript{a,*}, Philipp Thunshirn\textsuperscript{a,b}, Helmut Haberl\textsuperscript{a}

\textsuperscript{a}Institute of Social Ecology, University of Natural Resources and Life Sciences, Vienna, Schottenfeldgasse 29, Vienna 1070, Austria
\textsuperscript{b}University of Applied Sciences Technikum Vienna, Giefinggasse 6, Vienna 1210, Austria

\begin{abstract}
Electricity infrastructures are crucial for economic prosperity and underpin fundamental energy services. This article provides global datasets on installed power plant capacities, transmission and distribution grid lengths as well as transformer capacities. A country-level dataset on installed electricity generation capacities during 1980 to 2017, comprising 14 types of power plants and technologies, is obtained by combining data from three different online databases. Transmission grid lengths are derived from georeferenced data available from OpenStreetMap, augmented with data from national and international statistics. Data gaps are filled and historical developments estimated by applying a linear regression model. Statistical data on distribution grids lengths are collected for 31 countries that make up almost 50\% of the global electricity consumption. Estimates for distribution grid lengths in the remaining countries are again obtained through linear regression. Data on installed transformer capacities are sparsely available from market intelligence reports and specialist journals. For most countries, they are estimated from typical transformer-to-generator ratios,
\end{abstract}

\begin{keywords}
Energy system  
Installed capacities  
Power grid  
Transmission lines  
High-voltage grid  
Distribution grid
\end{keywords}


\textsuperscript{*}Corresponding author.
\textsuperscript{E-mail address:} gerald.kalt@boku.ac.at (G. Kalt).

DOI of original article: 10.1016/j.resconrec.2021.105723
Social media: (G. Kalt)

https://doi.org/10.1016/j.dib.2021.107351
2352-3409/© 2021 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)
Specifications Table

| Subject                      | Energy        |
|------------------------------|---------------|
| Specific subject area        | Electricity   |
| Type of data                 | Table         |
| How data were acquired       | Desktop research (power plants: combination of different online databases; transmission grids: derived from GIS data available from OpenStreetMap, national and international databases; distribution grids and transformers: mainly national statistics and reports; data gaps are filled through linear regression models) |
| Data format                  | Analyzed      |
| Filtered                     |
| Parameters for data collection | Not applicable. |
| Description of data collection | Data on installed capacities from 1980 to 2017, broken down to 14 technology types, are derived from various online databases and publications. Grid lengths are derived from georeferenced data available from OpenStreetMap, national and international statistics. Data gaps are filled using linear regression models. Transformer capacities are available from market intelligence reports and specialist journals; for most countries, they are estimated from generation capacities, assuming typical transformer-to-generator ratios |
| Data source location | • EIA, Independent Statistics & Analysis. International, Website of the U.S. Energy Information Administration. (2020). https://www.eia.gov/international/data/world.  
• IRENA, Renewable capacity statistics 2019, International Renewable Energy Agency, Abu Dhabi, 2019.  
• UNSD, Energy Statistics Database. Electricity, net installed capacity of electric power plants, UN Data. Website of the United Nations Statistics Division. (2020). http://unstats.un.org/unsd/energy/edbase.htm. |
Value of the Data

- Electricity infrastructures, comprising generation units (power plants), transmission and distribution systems, are key to economic prosperity and modern energy services, and thus highly relevant for studies on economic and societal wellbeing.
- Data on the historical development of electricity infrastructures are used for various research fields, ranging from industrial ecology and energy economics to sustainability research and studies on human development.
- The provided data can be used in energy system models, econometric analyses, for national indices and many more purposes.
- Global country-level capacities of power plants are available from various sources, but timeframes and levels of detail vary between databases. The most detailed data have been identified and combined in a consistent database.
- Data on grids and transformers are less readily available. Grid data have been compiled from GIS data and numerous other sources to support estimates of global infrastructure stocks that we consider as a valuable basis for future research.

1. Data Description

This article provides data on global electricity infrastructures for 178 countries from 1980 to 2017. This includes generation capacities (broken down by 14 power plant types/technologies) transmission and distribution grid lengths and transformers (broken down by three types: transmission, distribution and generator step-up transformers). Since freely available databases on global power plant capacities vary in scope, detail and timeframe, the consolidated dataset provided here is considered a valuable resource for various research fields, such as industrial ecology or energy economics. To the best knowledge of the authors, global electricity grid lengths and transformer capacities are less well documented: only dispersed data on individual countries or regions are available from national authorities, country reports, etc. The data presented here are the outcome of a comprehensive literature review and data collation exercise that resulted in an as yet unavailable stock estimate of global electricity infrastructures.

The complete datasets are provided in the spreadsheet appendix. The following sub-sections give an overview and interpretation of the data. As in the related article published in Resources, Conservation and Recycling [1], a regional grouping system that is based on politico-economic and geographic criteria is used (Fig. 1) to analyse the data on an appropriate spatial resolution. The country grouping combines the grouping systems used in previous studies and distinguishes between two individual countries (China and India), four world regions (Latin America and the Caribbean, Middle East and Northern Africa, Sub-Saharan Africa, Rest of Asia) and three country
groups characterized by common socio-political histories (Former Soviet Union, Industrial-New World, Industrial-Old World). The correspondence table is provided in the spreadsheet appendix.

1.1. Power plants

Fig. 2 illustrates the development of global power plant capacities since 1980 in the nine country groups; (a) shows installed capacities in Gigawatt (GW) and (b) the relative shares of fuel types/energy sources. A comparison between country groups reveals large differences with regard to the dynamics as well as the relative shares of technologies and fuels. Especially conspicuous aspects include a large variety regarding the share of coal and oil power plants, the increase of natural gas especially since the 1990ies, and the regionally diverse and generally diminishing relevance of nuclear energy. Since the turn of the millennium, solar energy and wind power have been growing rapidly in industrialized countries and later also in the rest of the world, especially in China. The role of bioenergy in electricity generation is generally low, with the highest shares occurring in Latin America and the Caribbean (primarily based on sugar cane bagasse) and “Rest of Asia”. The share of hydropower is declining globally, and its role is relatively similar across world regions, except for Latin America and the Caribbean, where it accounts for about half of the total installed capacities and Middle East and Northern Africa, where it is of minor importance. The capacities of tidal and wave energy converters are negligible throughout the world, and concentrated solar power (CSP) as well as geothermal power plants are at present also of very little quantitative relevance.

Fig. 3 illustrates the installed capacities in 2017 on country level. With regard to total generation capacities (panel a), China ranks in the first place, followed by the United States. Other countries with major generation capacities include India (388 GW), Japan (302 GW), Russia (270 GW) and Germany (214 GW). Countries with generation capacities below 10 GW are primarily located in Africa and Central America as well as Eastern Europe and Asia. A more meaningful country comparison, illustrating installed capacities per capita is shown in panel b. This figure reveals small countries with exceptionally high generation capacities per inhabitant (e.g. Norway and Iceland with their vast hydropower capacities) as well as countries with underdeveloped power sectors (primarily in Africa, Southern Asia including India, Indonesia, North Korea as well as several countries in Central and South America). In panel c, illustrating generation capacities per square kilometre land area, countries with high population densities and well-developed power sectors become apparent (e.g. Germany, UK, Japan and Italy).

To give an indication of the composition and diversity of installed capacities at the country level, Fig. 3d shows the share of renewable energy technologies in total capacities. The power sectors of countries with close to 100% renewable generation capacities are all dominated by hy-
Fig. 2. Development of installed capacities in nine world regions from 1980 to 2017. Capacities in Gigawatt (GW) by fuel type/energy source (top) and relative shares of fuel types/energy source (bottom). Tidal/wave power plants are not included due to negligible capacities worldwide.

dropower. Apart from Norway and Iceland, this is true for Paraguay, Nepal, Laos and some Sub-Saharan countries, namely Ethiopia, the Democratic Republic of the Congo, Zambia and Malawi.

1.2. Electricity grids

Fig. 4 illustrates the transmission (a) and distribution grid lengths (b) in circuit kilometres (km)\(^1\) in 2017, broken down by overhead lines of different voltage levels and underground cables. Low-voltage distribution lines (below 1 kV) are not differentiated by cables/overhead lines due to data issues.

The total global transmission circuit length is estimated at 4.7 million km\(^2\), with China accounting for a share of 28%. Apart from the considerable uncertainties regarding transmission

\(^1\) Grid lengths are usually indicated in circuit kilometres rather than the combined distances covered by transmission lines (route length), to account for multiple-circuit lines. Hence, distances covered by lines with more than one circuit (corresponding to three conductors in three-phase AC systems and two conductors in DC systems) are multiplied with the number of circuits to obtain circuit lengths. Data on grid lengths stated in this article are always circuit lengths.

\(^2\) According to Ref. [2], the leading 215 transmission system operators worldwide owned 4.14 million km of transmission lines.
cable lengths, this estimate is regarded as being quite robust, so no uncertainty ranges are considered here.

Uncertainties are considerably larger with regard to distribution grids, as the results in Fig. 4b are largely based on a linear regression model. The error bars, derived from the 95% confidence interval of the regression, are an indication of these uncertainties. Our medium estimate for the circuit length of the global distribution grids is about 96 million km, with low-voltage lines accounting for a share of 63%. A comparison between Fig. 4a and b reveals considerable, yet plausible, differences regarding the contributions of individual country groups: Groups with comparatively low population densities, especially Industrial (New World) and Former Soviet Union, have relatively large transmission grids in relation to the size of their distribution grids, while distribution grids in densely populated regions (India, China, Rest of Asia and Industrial (Old World)) have relatively large distribution grids.

Fig. 4c illustrates differences in transmission grid densities on country level. The country with the largest transmission grids per km² land area is Singapore (3.16 km/km²). High transmission grid densities are prevalent in European countries (e.g. Germany, UK, Czechia, Austria) and small countries in the Middle East (e.g. United Arab Emirates, Kuwait, Israel). India and China are the largest countries with relatively high transmission grid densities; they rank on 19th and 25th place. Distribution grid densities on country level (Fig. 4d) show similar patterns as transmission grid densities with high values of about 2.5 to 6 km/km² in Central European countries, as compared to a global median value of 0.67 km/km².

1.3. Transformers

The total global capacity of transformers is estimated 40,600 ± 4500 Gigavolt-Ampere (GVA), with transmission and distribution transformers accounting for slightly above 40% each, and gen-
G. Kalt, P. Thunshirn and H. Haberl / Data in Brief 38 (2021) 107351

Fig. 4. Global circuit lengths of (a) transmission grids and (b) distribution grids in 2017 broken down by world regions, type of infrastructure (OH = overhead line; underground cable) and voltage levels. Error bars to distribution grid lengths are aggregated uncertainties for all categories. (c) Transmission circuit length relative to land area. (d) Distribution circuit length relative to land area. Note the non-linear colour gradient in panel c and d; bright yellow corresponds to 10% of the maximum values of the scales, i.e. 0.03 km/km² in c) and 0.5 km/km² in d); the maximum values of the scales are set to 0.3 km/km² and 5 km/km², respectively, despite significantly higher values in countries like Singapore or Bahrain.

Western industrialized countries account for about 40% and China for 26% of the globally installed transformer capacity (Fig. 5). Data for India as well as for transmission transformers in China are adopted from national statistics [3,4], are thus considered as robust and not displayed with error bars.3 For the remaining country groups, the data are largely based on generation capacities (see methods described below) with error bars representing ranges of up to +/-20%. (Statistical data available for individual countries reduce the uncertainty ranges of some country groups to less than the default assumption of +/-20%.)

2. Materials and Methods

Fig. 6 gives an overview of the considered infrastructure components, the respective main data sources and the applied methods. Detailed explanations are provided in the following subsections.

---

3 In the spreadsheet appendix, low and high estimates are identical to the medium estimate in these cases.
Comparing various statistical databases on power plant capacities, it was found that the highest possible level of detail can be achieved by combining information from three databases: The database provided by the US Energy Information Administration [5] includes power plant capacities for the time period 1980 to 2017, but data for renewable energy technologies except for hydropower are only available since 2000. The database of the International Renewable Energy Agency (IRENA) [6] is limited to renewable energy technologies and the timeframe 2000 to 2017, but has a higher level of detail. The UN database [7] is used to fill data gaps (e.g. renewable capacities during 1990 to 2000). Renewable energy capacities for the period 1980 to 1989 are derived from Ref. [8]. Additionally, data from Ref. [8] are used to break down capacities of fossil fuel powered plants to the categories coal, gas and oil, as well as hydropower to run-of-river and storage plants.

Fig. 5. Capacities of transmission and distribution transformer in the 9 world regions in 2017.

Fig. 6. Overview of the main data sources and methods.

2.1. Power plants
2.2. Transmission and distribution grids

Data on electricity grid lengths are less readily available than generation capacities. In absence of universal definitions of transmission and distribution grids, overhead lines and underground cables with a voltage level below 100 kV are here classified as distribution and such with 100 kV or more as transmission grid components. Transmission grid lengths on national scale are extracted from GIS data available from OpenStreetMap4 (OSM) using the open source Geographic Information System QGIS (Version 3.12.1). OSM data as of 2019 are used for deriving 2017-estimates of transmission grid lengths. This appears justified by presumed time delays between the construction of transmission lines and their inclusion in OSM datasets. Spatially explicit OSM data are translated to country-level spreadsheet data by merging the OSM data layer with the GIS country layer “Countries_WGS84” obtained from ESRI ArcGIS Hub5. This approach yielded grid lengths for 178 countries. Furthermore, statistical data on transmission grid lengths have been collected from international [9] and national institutions [3,4,10,11] and market intelligence reports [12]. Such statistical data are generally considered more reliable than data obtained from OSM, and are thus favoured over OSM-based data whenever available. The country-level database on current (2017) transmission grid lengths in 178 countries is thus composed of GIS-based as well as statistical data.

To account for multiple-circuit lines (transmission towers/pylons carrying more than one circuit, i.e. 3 conductors in three-phase AC systems and 2 conductors in DC systems), it is often differentiated between route lengths and circuit lengths (see footnote 1). Some statistics provide both route and circuit lengths, and OSM data also provide the number of circuits or conductors to most georeferenced transmission lines. After clearing obviously incorrect entries (e.g. number of circuits exceeding four, as there are no known tower designs carrying more than four circuits) and assuming one circuit as default in case of missing entries, circuit lengths are obtained by multiplying all distances covered by all transmission lines with the respective number of circuits.

Transmission cables are primarily used in Europe (mainly included in the group “Industrial (Old World)”), but there is some uncertainty regarding cable line lengths in the rest of the world, as no up-to-date statistical data are available on global scale. Data in Ref. [13] for 28 countries are found to be implausibly low, based on a comparison with Ref. [9], and line lengths extracted from GIS data are unreliable with regard to cables. Transmission cable lengths are thus likely to be underestimated. However, it is plausible that “Industrial (Old World)” accounts for the largest share of transmission cables globally [13], as illustrated in Fig. 4; their relatively low contribution to total circuit lengths in this group is an indication that the likely underestimation of cables does not cause a severe error in our results.

For India and China, data on grid development since 1980 are available from Refs. [3,4]. For all other countries, historical data on transmission line lengths are estimated through linear regression. Various independent variables have been considered, including, for example, land area, generation capacities, population and electricity consumption. Satisfying regression results for transmission grid lengths were obtained from a linear regression model derived from the country level data on transmission grid lengths in 2017, with land area in $10^6$ km$^2$, the population with access to electricity$^6$ and total power plant capacities in GW as independent variables. In contrast to the regression for distribution grids (see below), the product of circuit lengths (in km) and the respective voltage level (in kV) are here used as dependent variable, in order to account for the varying power transmission capacities of different voltage levels.

The regression has the form

$$TGS_i = \kappa + \delta \cdot PAE_i + \varepsilon \cdot LA_i + \zeta \cdot PPC_i$$

(1)

with $TGS_i$ representing the transmission grid stock (in km$^2$ kV) in region $i$, $PAE_i$ the population with access to electricity, $LA_i$ the land area in million km$^2$ and $PPC_i$ the installed power plant capacities.

4 Complete copies of the OpenStreetMap.org database are available from https://planet.openstreetmap.org.
5 See: https://hub.arcgis.com/.
6 See: https://data.worldbank.org/indicator/EG.ELC.ACCS.ZS?end=2018&start=1998.
capacity. The regression coefficients are determined as \( \kappa = -1.21 \cdot 10^6 \), \( \delta = 52,093 \), \( \varepsilon = 2.6 \cdot 10^6 \) and \( \xi = 112,728 \).

The adjusted \( R^2 \) of the regression is 0.97. Fig. 7 shows a comparison of the regression results derived on country level with the actual data for 2017, both aggregated to world regions. Despite notable deviations for Rest of Asia and Latin America and the Caribbean, the regression results are considered suitable for deriving estimates on historical grid lengths. The regression results are finally converted to circuit lengths and circuit lengths of different voltage levels estimated by assuming the average voltage level and distributions according to the data for 2017.

Representation of distribution grids in OSM is clearly less complete than in case of transmission grids, partly owing to a higher relevance of underground cables which cannot be identified from aerial images. Furthermore, vast regional differences in the completeness of medium (1 to 100 kV) and low voltage lines (< 1 kV) have been discovered by visual comparison of random areas with similar population densities and settlement structures. It was concluded that OSM data are too unreliable for distribution grids. Statistical data on distribution grids covering all voltage levels below 100 kV are also sparse. Data for 31 countries have been collected, including most of continental Europe [9], the United States [14], India [10], Indonesia [15] and Japan [11], that make up almost 50% of the global electricity consumption. Estimates on distribution grid lengths in the remaining countries are obtained from a linear regression model. Satisfying regression results for distribution grid lengths were obtained with population data and final electricity consumption. For plausibility reasons, the number of households with access to electricity (estimated from population data, the fraction with access to electricity according to World Bank data and average household sizes according to UNDESA\(^\text{7}\)) have been used instead of total population.

The regression for the distribution grid has the form

\[
DGL_i = \alpha \cdot HAE_i + \beta \cdot EC_i
\]

\[\text{Fig. 7. Comparison of statistical data on the total transmission grid stocks (measures in circuit length times voltage level; } 10^6 \text{ km}^2\text{kV}) \text{ and regression results for all world regions (data extracted from OSM and based on statistical data from various sources mentioned in the text).}\]

\[\text{7 See: https://www.un.org/development/desa/pd/data/household-size-and-composition.}\]
with $DGI_i$ representing the total distribution grid length in km in country $i$, $HAE_i$ the number of households with access to electricity and $EC_i$ the total final electricity consumption in TWh. The regression coefficients are determined as $\alpha = 0.03535$ km and $\beta = 1459.19$ km/TWh. As Fig. 8a shows, the regression results often deviate considerably from the actual grid lengths. Thus, the lengths derived for other countries via regression, as well as the historical developments, are significantly less certain than generation capacities and transmission grid lengths, and should be considered as rough estimates. The range of the 95% confidence interval of the regression have been used to derive upper and lower estimates to account for this uncertainty.

The share of low voltage grids is simplistically assumed to correspond to the average share in EURELECTRIC member states included in [9]. Fig. 8b shows that the average value is 63%, with a standard deviation of 5%.

2.3. Transformers

Data on transformers are similarly sparse, with data on installed capacities in individual countries available from specialist journals [16], market intelligence reports [2,12] and national energy institutions and authorities (e.g. [3,4]). The obtained dataset includes 23 countries and is considered insufficient for deriving global estimates through linear regression. However, following an industry rule of thumb, national level estimates can be obtained from power plant capacities by assuming typical transformer-to-generator ratios [16]. Three transformer types are differentiated: transmission, distribution and generator step-up transformers. If no country-specific data are available, transformer-to-generator ratios of 2.5 are assumed for transmission transformers as well as distribution transformers. (Ref. [16] provides data for 10 individual countries, with average ratios of 2.77 and 2.17, respectively, and mentions 2.5 as a typical value for transmission.) For generator step-up transformers, a uniform transformer-to-generator ratio of 1.1 is assumed. (This ratio applies to nine out of ten countries, according to Ref. [16]). That is, transformer capacities are estimated by multiplying total power plant capacities (in GW) with a factor of 2.5 GVA/GW for transmission, 2.5 GVA/GW for distribution and 1.1 GVA/GW for generator...
step-up transformers. For low and high estimates for transmission and distribution transformers, transformer-to-generator ratios are varied from 2 to 3 GVA/GW.

Ethics Statement

Not applicable. Our work does not involve any use of human subjects, animal experiments or data collected from social media platforms.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have or could be perceived to have influenced the work reported in this article.

CRediT Author Statement

Gerald Kalt: Conceptualization, Methodology, Data curation, Writing – original draft, Writing – review & editing; Philipp Thunshirn: Methodology, Data curation, Writing – original draft; Helmut Haberl: Writing – review & editing.

Acknowledgments

This research has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (MAT_STOCKS, Grant Agreement No. 741950). The authors acknowledge the assistance given by Markus Löw in connection with statistical analysis and express their gratitude to Detlev Reiners from 123map for providing grid data extracted from OSM.

Supplementary Materials

Supplementary material associated with this article can be found in the online version at doi: 10.1016/j.dib.2021.107351.

References

[1] G. Kalt, P. Thunshirn, D. Wiedenhofer, F. Krausmann, W. Haas, H. Haberl, Material stocks in global electricity infrastructures from 1980 to 2017—a socio-metabolic analysis of the world’s power sectors, Resour. Conserv. Recycl. 173 (2021) 105723, doi: 10.1016/j.resconrec.2021.105723.
[2] Global transmission, global TSOs: growth in transmission networks of leading TSOs, global transmission report: information and analysis on the global electricity transmission industry. (2019). https://www.globaltransmission.info/archive.php?id=36488#https://startpage.com/do/metasearch.pl. Accessed September 2, 2020.
[3] CEA, Grid operation & distribution, website of the central electricity authority (India). (2020). http://cea.nic.in/index.html. Accessed August 24, 2020.
[4] CEC, Data & Publications. Data/statistics, website of the China electricity council. (2020). https://english.cec.org.cn/menu/index.html?2251. Accessed September 3, 2020.
[5] EIA, Independent Statistics & Analysis. International, website of the U.S. energy information administration. (2020). https://www.eia.gov/international/data/world. Accessed August 24, 2020.
[6] IRENA, Renewable Capacity Statistics 2019, International Renewable Energy Agency, Abu Dhabi, 2019.
[7] UNSD, Energy Statistics Database. Electricity, net installed capacity of electric power plants, UN Data. Website of the United Nations Statistics Division. (2020). http://unstats.un.org/unsd/energy/edbase.htm.
[8] J. Farfan, C. Breyer, Structural changes of global power generation capacity towards sustainability and the risk of stranded investments supported by a sustainability indicator, J. Clean. Prod. 141 (2017) 370–384, doi: 10.1016/j.jclepro.2016.09.068.
[9] EURELECTRIC, Power Distribution in Europe. Facts and Figures, Union of the Electricity Industry—EUROELECTRIC Aisbl, Brussels, Belgium, 2013.

[10] WEC-IMC, Transmission and Distribution in India A report, World Energy Council-Indian Member Committee, Power Grid Corporation of India Limited, 2010.

[11] JEPIC, The Electric Power Industry in Japan 2019, Japan Electric Power Information Centre, Inc., Tokyo, Japan, 2019.

[12] Global Transmission, Existing Electricity Transmission Infrastructure in Middle East, Global Transmission Report: Information and Analysis on the Global Electricity Transmission Industry, 2018 https://www.globaltransmission.info/archive.php?id=35513. Accessed September 3, 2020.

[13] CIGRE, Statistics of AC Underground Cables in Power Networks, CIGRÉ, Paris, France, 2007.

[14] MIT, The Future of the Electric Grid: an Interdisciplinary MIT Study, Massachusetts Institute of Technology, 2011.

[15] A. Amiruddin, Indonesia Electricity Policy, Ministry of Energy and Mineral Resources, Directorate General of Electricity, Jakarta, Indonesia, 2016.

[16] S. Aubertin, Power and distribution transformer markets. Market values vs. Market Volumes, Transform. Mag. 3 (2016) 8.