

# Annual Half-Hourly Real and Reactive Power Flows for 171 Primary Electrical Distribution Circuits in the North of England

Matthew Deakin, Newcastle University

February 13<sup>th</sup> 2023

Email: [matthew.deakin@newcastle.ac.uk](mailto:matthew.deakin@newcastle.ac.uk)

## Summary

This dataset provides half-hourly measurement data for 171 circuits in the English regions of the North East and Yorkshire and the Humber for one full calendar year (Jan-Dec 2022 inclusive). The data is derived from Primary Operational Metering data provided in Northern Powergrid's (NPG's) Open Data portal. The 171 circuits have been selected as they have high levels of data quality, with all circuits having less than one missing measurement per week on average. A visual inspection of all of the profiles has been carried out to ensure there are no obvious errors, although the data are reported from transducers with unspecified accuracy.

A clean version of the data is provided with no missing values. This is created using a heuristic data imputation step, replacing missing values with the average weekday or weekend values at the given half-hourly time period.

The raw data is provided as a csv file for easy re-use. It is envisioned that the data might be used to support a range of analysis use-cases, such as scheduling of flexible assets (e.g., EVs, heat pumps, batteries) to manage distribution network congestion; as input data to support risk- or uncertainty-based network management methods; or, as a pedagogical resource to support the teaching of distribution grid fundamentals.

## Acknowledgement

This dataset was supported by Northern Powergrid Open Data under the Northern Powergrid Open Data Licence v1.0 [2]. This work to create this dataset was supported by the Royal Academy of Engineering under the Research Fellowship programme.

## Attribution

This dataset can be cited as follows.

```
@Misc{deakin2023annual,  
author = {Deakin, Matthew},  
title = {Annual Half-Hourly Real and Reactive Power Flows for 171 Primary Electrical Distribution  
Circuits in the North of England},  
howpublished = {DOI: 10.25405/data.ncl.22047758},  
year = {2023}  
}
```

## Background

In 2022 the distribution network operator Northern Powergrid (NPG) started an Open Data portal, in line with the Energy Network Association's "Open Networks" initiative [3], and is (at the time of writing) periodically uploading data such as Primary Operational Metering, Grid Supply Point Operational Metering and the Embedded Capacity Register [1]. This data is provided with the "Northern Powergrid Open Data Licence v1.0", which is listed as being compatible with CC-BY 4.0, having been based on the Open Government Licence v3.0 [2].

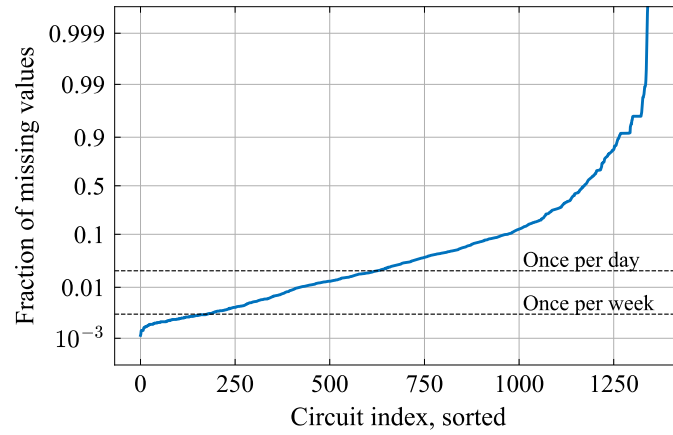


Figure 1. The number of missing values per circuit plotted on a logit scale. There are 171 circuits which have fewer than one missing value per week across 2022 ( $(7 \times 48)^{-1} = 0.298\%$ ).

This dataset is derived from the Primary Operational Metering data from [1]. That NPg data is provided as zip files of approximately 80 MB each, or 1 GB unzipped, with 1376 circuits having some reported real and/or reactive power flows in 2022. The datastreams having varying quality – Figure 1 shows that approximately half of the circuits have (on average) more than one missing value per day. This point is noted also by NPg, pointing out in the readme that the measurement transducers “...have been installed for indicative purposes only and take data from instrumentation circuits that were not intended for accurate measurement” [4].

#### Validation

NPg carry out some basic quality checks to remove datastreams which are flat [4]. A further visual inspection of the measurement streams has also been carried out to exclude a number of circuits which have low levels of missing data, but had other data quality issues. This manual inspection is used as outlier detection is tricky due to a very heavy tail in the distribution of powers (e.g., a change in network topology can easily double or triple demand). Some of these exclusions were due to data points which were clearly poor (see, e.g., Figure 2), but in other cases data streams for circuits were conservatively removed even when there was only one or two apparently erroneous values. The full list of excluded circuits are given in the appendix. In total, this table reports 19 circuits that had less than one missing value per week but failed this visual inspection.

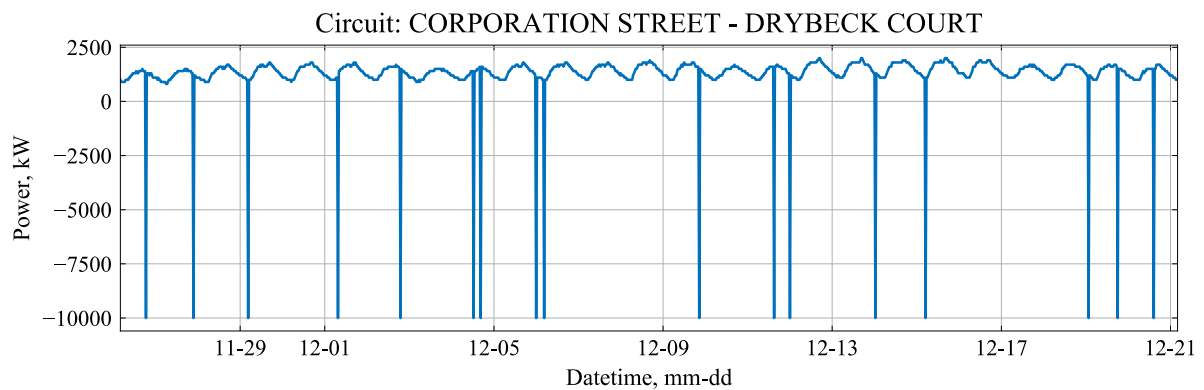


Figure 2: Outliers are observed throughout the year for Corporation Street - Drybeck Court, so this circuit’s measurements are excluded from the data streams, even though there is less than one missing value per week.

## The dataset

171 circuits are chosen which have, on average, fewer than one missing value per week (i.e., less than 0.298% missing values for a time series with half hourly temporal resolution). Two versions of the data are provided. Firstly, the data in “ts\_profiles.csv” is almost identical to the concatenation of the real and reactive powers in the NPg monthly zip (json) files. The only change is that there can only be one measured value per timestamp in the csv. Where there are two or more values in the zip file from [1] that correspond to just a single timestamp, the measurement that is recorded in “ts\_profiles.csv” is the final value is reported in the zip (json) file (the file is read sequentially). The profiles are given in the csv files in units of kVA.

A second version of the data is also provided, “ts\_profiles\_clean.csv”. This clean version has been made continuous using a data imputation heuristic: if there is missing data, then the data at that time instant is chosen to be the mean value of all other data at that time period and day type (weekday or weekend).

For example, Figure 3 shows the mean half-hourly real powers for Goole TRX2 circuit, showing a clear difference between weekday and weekend profiles. The difference between missing value replacement can be seen in the example shown in Figure 4, where the mid-morning weekday imputation on 14/1 (a Friday) has value over 4000 kW, where the midday imputation on 15/1 (a Saturday) is below 4000 kW. This can be expected given the weekend and weekday imputation profiles plotted in Figure 3.

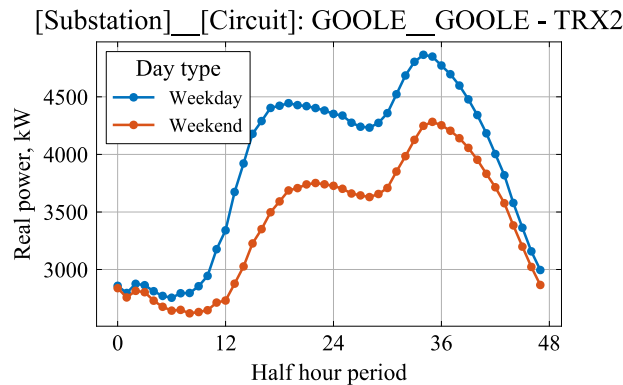


Figure 3: Mean of all non-missing time periods for the Goole substation circuit TRX2 when split into weekend and weekdays, as used for data imputation.

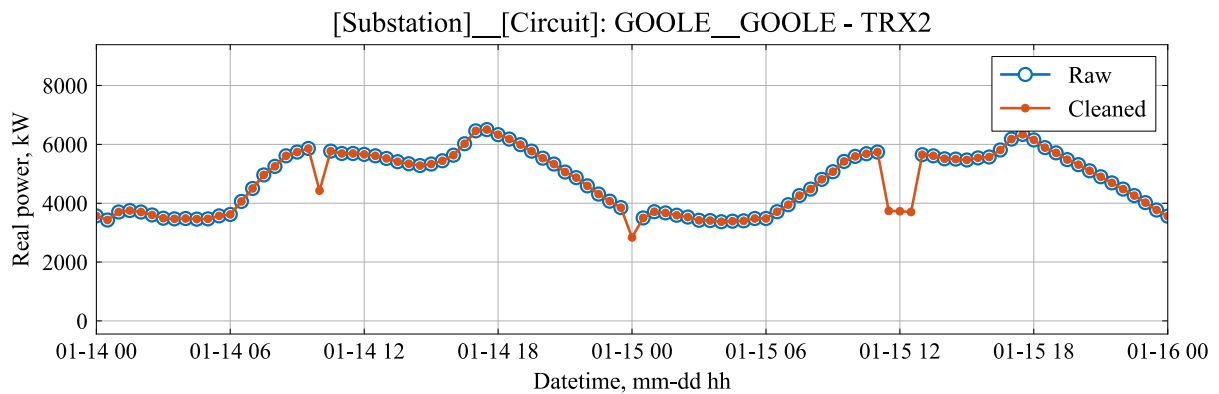


Figure 4: Real power for the raw data (with missing values, ‘Raw’) and the cleaned data (following data imputation, ‘Cleaned’) for 14/1 (a Friday) and 15/1 (a Saturday).

## Appendix: Circuits Excluded from Cleaning Following Visual Inspection

Table 1 reports the 19 circuits had fewer than one missing value per week on average, but were removed following a visual inspection.

Substation ID	Circuit ID(s)
RAWSON SPRING ROAD	TRX2
CORPORATION STREET	DRYBECK COURT, STOWELL HOTEL, TRX1, TRX2, ST JAMES PARK, ST JAMES BOULEVARD, COTTENHAM STREET, PLAZA ACCOMMODATION, WELLINGTON SCIENCE, DISPENSARY LANE TD AND YELLOW SECTION.
HARDEN	TRX2
BINGLEY 132/33	TRX2
BLUCHER	TRX2
BARRACK ROAD	TRX2
NATIONAL AVENUE	TRX2
OUTWOOD	TRX2
EDUCATIONAL PRECINCT	TRX2
HEXHAM	TRX1

*Table 1: Circuit IDs that are excluded from the dataset due to spurious measurements, even when they nominally had small numbers of missing values (less than one per week).*

## References

- [1] “Northern Powergrid Open Data Portal”. Accessed 11/2/23.  
<https://northernpowergrid.opendatasoft.com/pages/home/>
- [2] “Northern Powergrid Open Data Licence”. Accessed 11/2/23.  
<https://northernpowergrid.opendatasoft.com/p/opendatalicence/>
- [3] “Open Networks”, Energy Networks Association. Accessed 11/2/23.  
<https://www.energynetworks.org/creating-tomorrows-networks/open-networks/>
- [4] [Untitled: readme for primary operational metering]. Accessed 11/2/23.  
[https://northernpowergrid.opendatasoft.com/api/datasets/1.0/primary-operational-metering/attachments/readme\\_primary\\_operational\\_metering\\_pdf/](https://northernpowergrid.opendatasoft.com/api/datasets/1.0/primary-operational-metering/attachments/readme_primary_operational_metering_pdf/)