

HUE: The Hourly Usage of Energy Dataset for Buildings in British Columbia

Stephen Makonin

Engineering Science, Simon Fraser University, Canada

1. Introduction

The HUE dataset contains donated data from residential customers of BCHydro, a provincial power utility. There are currently twenty-two houses contain within the dataset with most houses having three years of consumption history. Data is downloaded from BCHydro's customer web porthole by each customer how donated the data. The porthole only allows customers to download a maximum of three years worth of data. Only BCHydro customers were asked to donate to keep the data quality consistent. Weather data from the nearest weather station is also included.

2. Dataset Structure

Consumption history data and weather data is stored in simple CSV files. The summary data is stored in a fixed-length format to may it easy to read. The below table describes the files within the HUE dataset and the columns within each file type. Data frequency for all files is hourly (in local Pacific timezone). Due to seasonal time changes, there may be duplicated or missing year-hour rows in the fall and spring.

Filename	Description
All_Residential.txt	Summary data for each house in listed in a table by house ID.
Holidays.csv	
Residential_<#>.csv	Hourly consumption history for each house where <#> is the ID of each house.
Solar.csv	One years worth of hourly simulated solar production data generated from the PVWatts online tool .
Weather_<ID>.csv	Hourly weather station data where <ID> is the three-letter weather station ID listed in the summary data table.

Residential House 1 is the same house used in the [AMPds dataset](#) which has 2-years of per-minute data including appliance-level consumption data; and is House 1 in the [RAE dataset](#) with approximately 60-days of 1Hz including appliance-level consumption data. Residential House 2 is House 2 in the [RAE dataset](#) with approximately one-year of 1Hz including appliance-level consumption data.

Column	Description
ac_output	Solar AC energy produced in kilo-Watt-hours (kWh) after DC conversion.
date	Date of the recording in YYYY-MM-DD.
day	Day of the week; e.g., Monday.
dc_output	Solar DC energy produced in kilo-Watt-hours (kWh).
dst	Day light savings time indicator (-1 or +1 for hour adjustment).
energy_kWh	Energy consumed in kilo-Watt-hours (kWh).
holiday	Textual name of the holiday (indicates a working day off).
hour	Hour of the recording from 01-24.
humidity	Outside humidity in percentage (%).
pressure	Atmospheric pressure in kilopascals (kPa).
temperature	Outside ambient temperature in degrees Celsius (°C).
weather	A textual description the the type of weather; e.g, Mostly Cloudy.
weekend	Boolean value to indicate weekend.

3. Access & Citation

All person using this dataset my cite this paper. The dataset can be downloaded from Harvard Dataverse at the below URL.

Dataset URL: <https://doi.org/10.7910/DVN/N3HGRN>

4. Future Direction

The goal is to grow the dataset by adding more residential house data along with commercial and industrial data. If you are a BCHydro customer and would like to donate your consumption data follow the [these instructions](#) and send your data to smakonin@sfu.ca.