Institute of the Environment & Sustainability

Energy Data in Context

Building the UCLA Energy Atlas

Cal TF Data Charette

June 21, 2023

Eric Daniel Fournier, PhD

Research Director, Assistant Researcher

California Center for Sustainable Communities (CCSC)

Institute of the Environment and Sustainability (IoES)

University of California Los Angeles (UCLA)

CCSC at UCLA

UCLA

menu

Residential Income

Click on one of the tiles to see the map visualization—or read more about the project here.

Building Vintage

UCLA Energy Atlas

This project is developed by the **UCLA California Center for Sustainable Communities (CCSC)** at UCLA, in the Institute of the Environment and Sustainability (IOES).

Building Type

Building Size

Historical Evolution of the Energy Atlas(es)

- ~2011 – Initial LADWP Energy Atlas
  - Funding support provided from LA County Internal Service Department. Initial batch of energy and water usage data Accessed through an NDA with LADWP.
- ~2014 – Los Angeles County Expansion
  - Increased efforts to collect data from other smaller municipal utilities throughout the region (Glendale Water and Power, Long Beach Utilities, Azusa Light and Water, etc.)
- ~2016 – SCE / SCAG Territory Expansion
  - Data sharing agreement with CPUC Energy Division provided bulk access to data for So-Cal IOUs. Access to standardized parcel data from SCAG was limiting factor in the expansion.
- ~2019 – BayREN Territory Expansion
  - Data for 9 counties in PG&E territory to support BayREN program development and evaluation. Standardized parcel data obtained through an NDA with MTC.

CATALENA

- CPUC Decision requiring that a "locationally aware" energy use reporting tool ”similar to the Energy Atlas" be developed. Strongly implied that the tool would be based on a new backend database of disaggregated customer usage data being assembled by CEC.
- The design of the RFP was originally in the hands of IOUs, lead by SCE. Preliminary scoping meetings were held with stakeholders – then COVID and hiatus.
- It appears that administration of the RFP has now been transferred to CEC but the development timeline is uncertain (as far as I know).
- CEC now has 15-minute interval account level customer usage data for all the IOUs, reported quarterly using an automated ETL process, that is being stored in Snowflake cloud database.

Key Challenges in Developing these Tools

- Spatial Data Issues
  - Instances of Poor-Quality Customer Address Information
  - Parcel Attribute Inaccuracies and Incompleteness
  - Parcel Boundary Overlaps and Inconsistencies
- Temporal Data Issues
  - Billing Interval Calendarization
  - Addressing Long-Duration Billing Intervals
- Customer Privacy Issues
  - Navigating Privacy Aggregation Guidelines

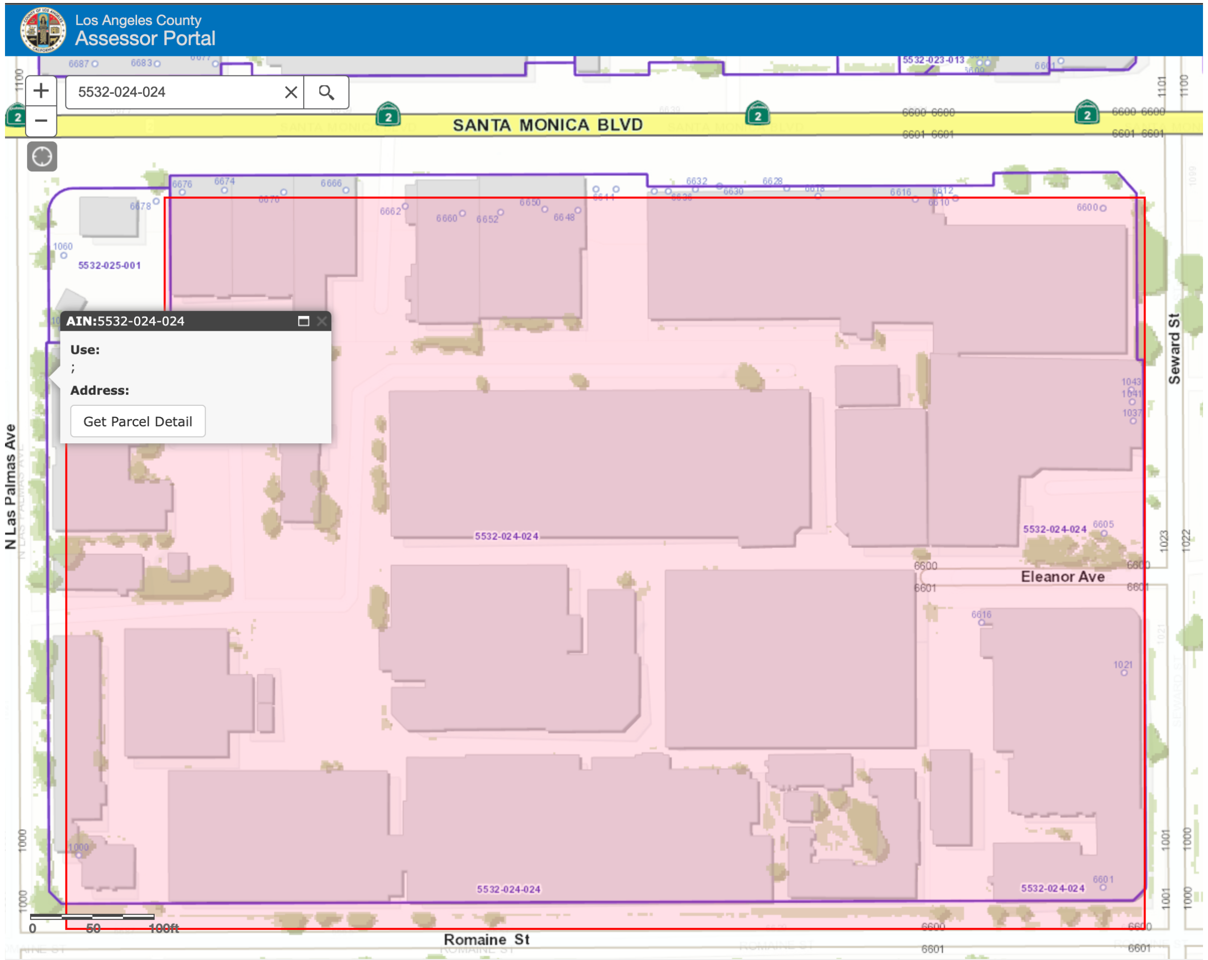
Spatial Challenge: Poor Quality Customer Address Information

(The following are fictionalized addresses – but represent actual quality issues that have been observed in real data)

ADDRESS	CITY	ZIP	COUNTY	STATE
1756 W 125 TH ST	[NULL]	[NULL]	[NULL]	[NULL]
NULL	MARTINEZ	94553	CONTRA COSTA	CA
3 MILES WEST OF MILE MARKER 27 ON HIGHWAY 101	VENTURA	93001	VENTURA	CA
!!!DO NOT USE!!! 5812 WELLINGTON COURT DRIVE	LOS ANGELES	90001	LOS ANGELES	CA
6625 SAN ANDREAS CT	DAVIS	-99999	YOLO	[NULL]

There is no standardized way in which customer addresses are represented in different utility databases. Sometimes they are concatenated into one string field, other times address components are separated into separate fields. There are also inconsistencies in terms of the specificity in which addresses are recorded (i.e., down to the unit level for multi-tenant properties, for example) and the ways in which street address prefixes and suffixes may or may not be abbreviated. There are also issues with incomplete fields, illegal values, and entries which are otherwise non-sensical. Methods must be developed for parsing these addresses to maximize the accuracy of geocoding results as well as handling customers whose addresses cannot be resolved down to the parcel level – but can be associated with some other level of the geographic hierarchy.

Spatial Challenge: Parcel Boundary Overlaps and Inconsistencies



Example: Note here the discrepancy between the parcel boundaries rendered on the base map (in purple) versus those that are shown when a selection is made (shaded in red). This is illustrative of the challenges associated with associated customer account addresses with parcel attributes. This task is very much analogous to creating a selection by clicking on a map.

Sometimes a single point (or selection) returns multiple parcel records due to the existence of overlapping polygons. Each of these may have their own attribute values - or none at all, as in this case of the “lost parcel” shown here. Methods must be developed for cleaning the parcel dataset and/or resolving these types of one-to-many conditions.

Spatial Challenge: Parcel Attribute Inaccuracies or Incompleteness

Parcel attribute quality and completeness are strongly correlated with taxable status. Tax-exempt properties (educational, institutional, religious, etc.) tend to have incomplete or otherwise very poor-quality attribute info.

Example: UCLA's entire campus is represented by a total of 2 parcels. Their combined building square footage area is listed as 0. And there are no building construction vintage attributes whatsoever.

Summary

AIN: 4363-001-903

Situs Address:
405 HILGARD AVE
LOS ANGELES CA 90095-9000

Use Type:
Other Property Type
Parcel Type:
Government Owned, Exempt
Tax Rate Area:
00067

Parcel Status: ACTIVE
Create Date:
Delete Date:
Tax Status: EXEMPT
Year Defaulted:
Exemption: None

Building & Land Overview
Use Code: 8831
Design Type:
Quality Class:

# of Units:
Beds/Baths: /
Building SqFt: 0

Year Built:
Effective Year:
Land SqFt:

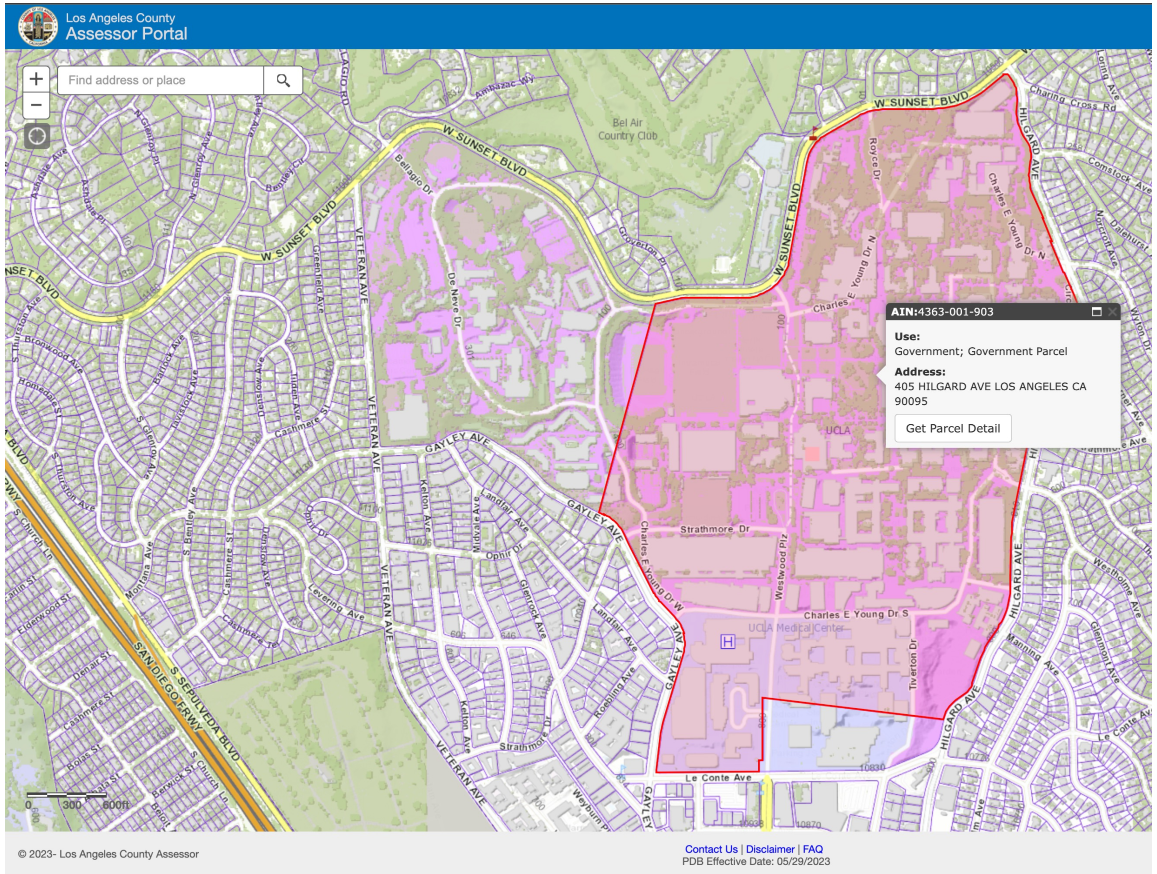
Parcel Map / Map Index

		2023 Roll Preparation		2022 Current Roll	RC	Year		1975 Base Value
Land	$	52,517,492	$	0	0	0	$	41,000,000
Improvements	$	0	$	0	0	0	$	0
Total	$	52,517,492	$	0			$	41,000,000

Assessor's Responsible Division
District: Hall of Administration
Region: 22
Cluster: 22385

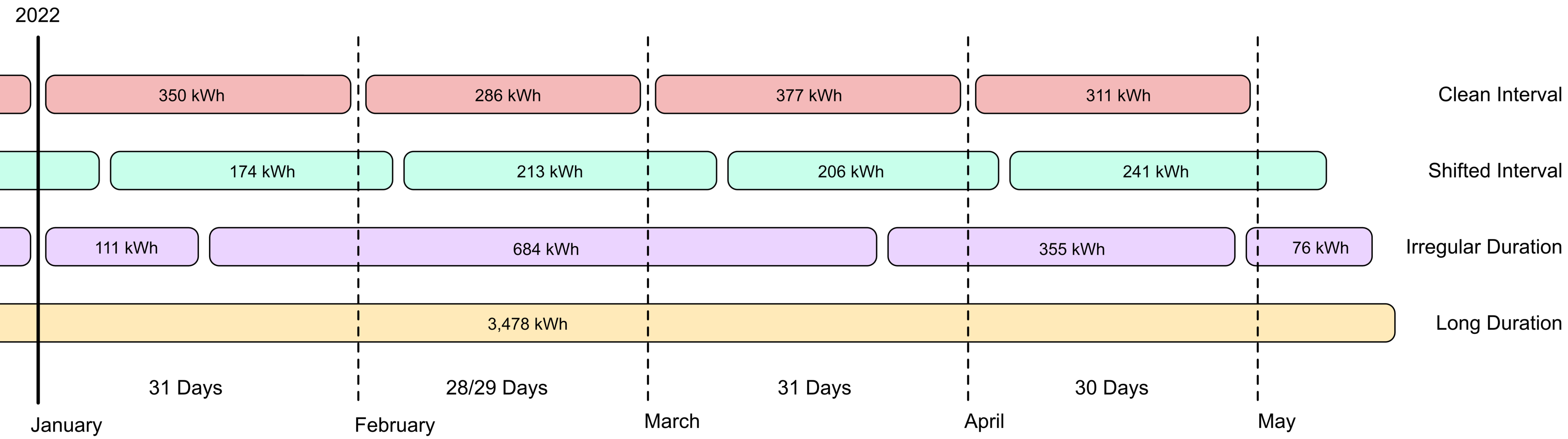
Hall of Administration
500 W. Temple St. Room 180
Los Angeles, CA 90012-2770

Phone: (213) 974-3108
Toll Free: 1 (888) 807-2111
M-F 7:30 am to 5:00 pm



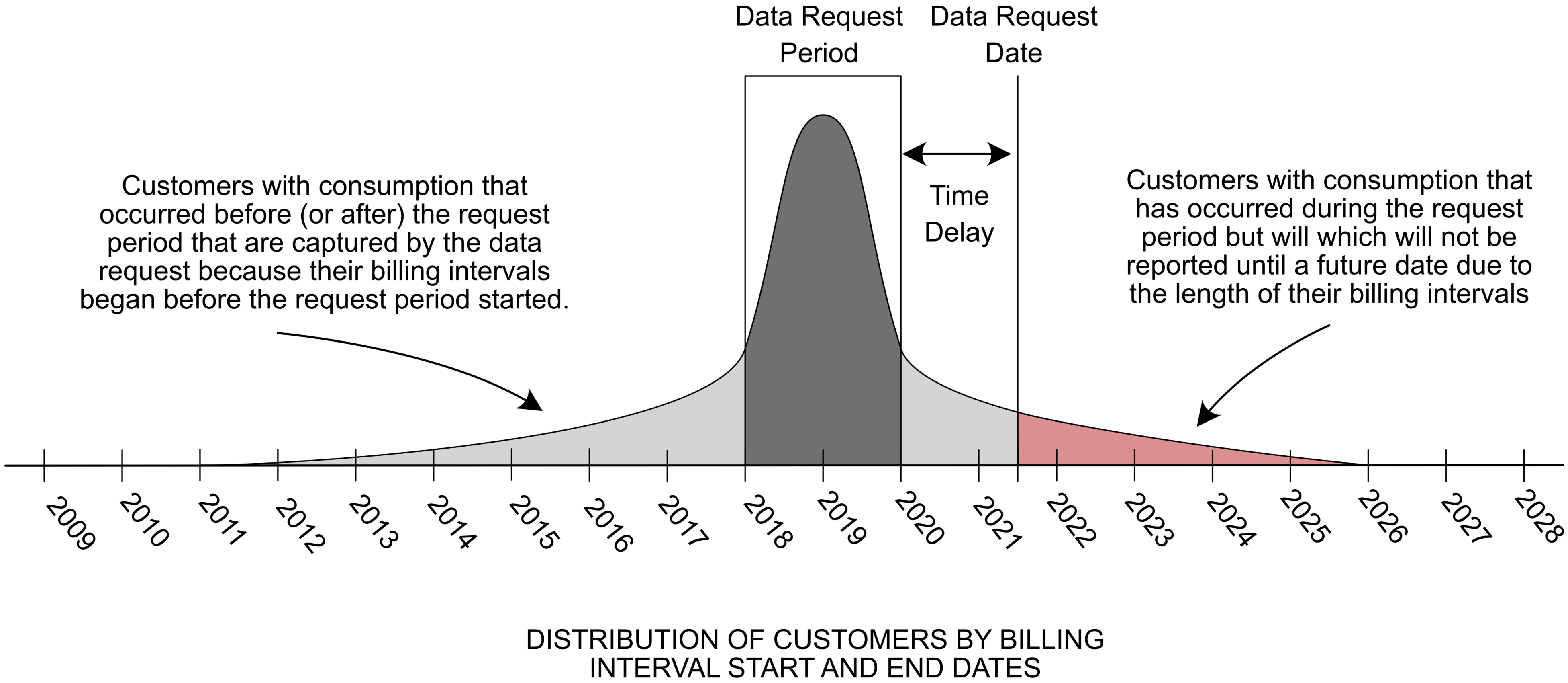
Temporal Challenge: Billing Interval Calendarization

(The following are fictionalized consumption records – but represent actual quality issues that have been observed in real data)



It is common to want to produce reporting / analytical products on regular calendar intervals (i.e. per month, year, etc.) Unfortunately, for many customers, especially those lacking advanced metering infrastructure (AMI), billing intervals can be quite irregular. Methods must be developed to apportion the consumption which occurs over non-conforming billing intervals into regular "calendarized" periods.

Temporal Challenge: Addressing Long Duration Billing Intervals



Every data request must be translated into a database query. What records are returned depends upon the specifics of how these queries are written. Depending upon the timing of the data request period, the timing of request query execution, and the timing of long duration billing interval closures, it is possible that many consumption records may not be captured in the request query results. Frustratingly, these tend to be associated with the largest customer accounts.

Customer Privacy Challenge: Navigating Privacy Aggregation Guidelines

California Public Utilities Commission (CPUC), 2014.
Decision [D. 14-05-016]
***Adopting Rules to Provide Access to Energy Usage and Usage-Related Data while Protecting Privacy of Personal Data*.**

Utility customer consumption data cannot be made public unless in aggregated form and only if the following conditions are met:

(1) The aggregated reporting group must contain at least 15 customers, for non-residential accounts, or 100 customers, for residential.

(2) No individual customer can comprise more than 15% of the reporting group's total consumption.

Customer Privacy Challenge: Navigating Privacy Aggregation Guidelines

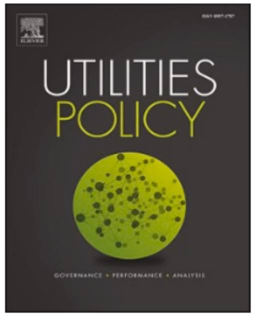
Utilities Policy 67 (2020) 101106

ELSEVIER

Contents lists available at ScienceDirect

Utilities Policy

journal homepage: http://www.elsevier.com/locate/jup



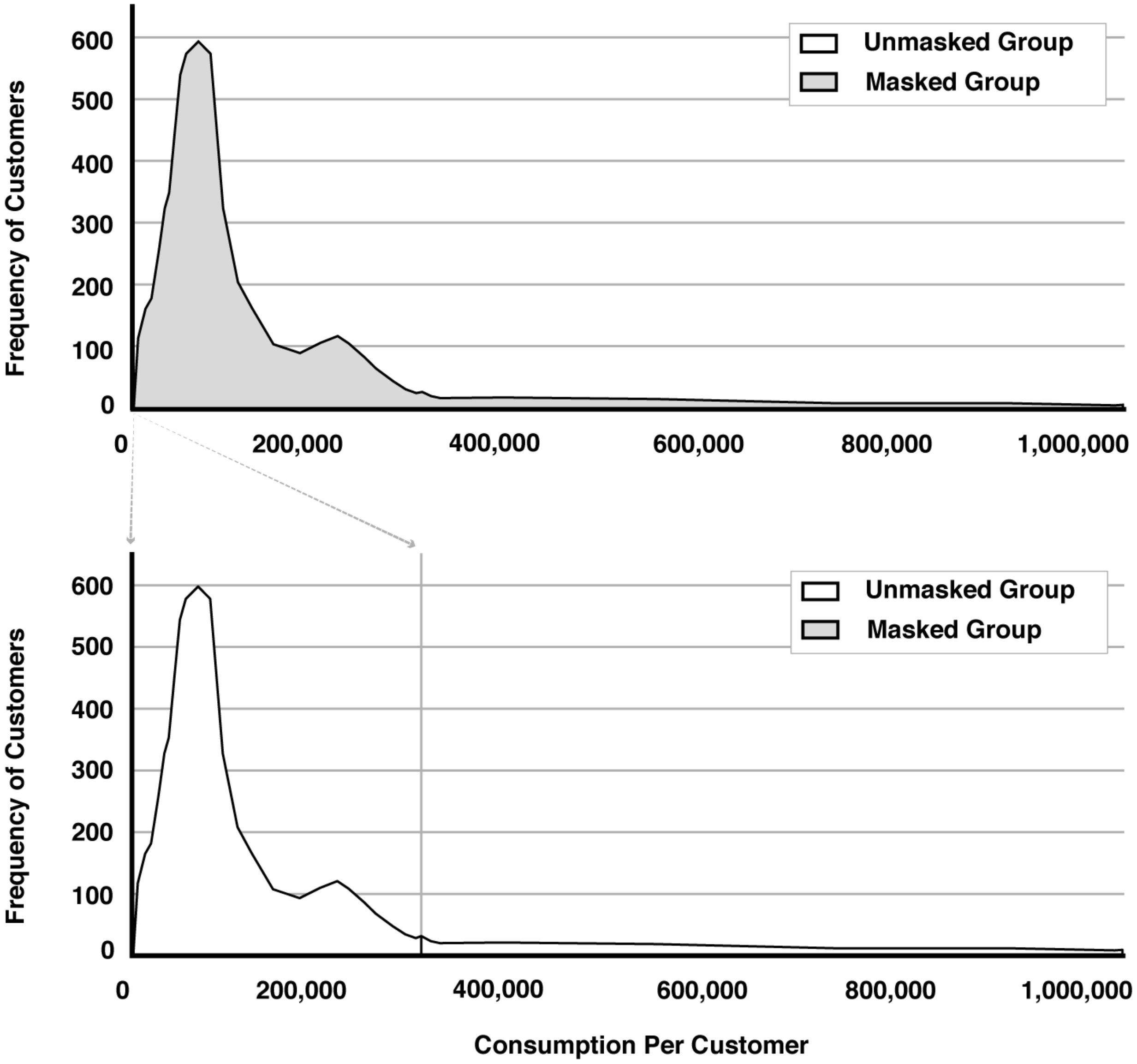
Check for updates

Guidance on the usability-privacy tradeoff for utility customer data aggregation

Benjamin L. Ruddell [a,*], Dan Cheng [b], Eric Daniel Fournier [b], Stephanie Pincetl [b], Caryn Potter [a], Richard Rushforth [a]

[a] Northern Arizona University, School of Informatics Computing and Cyber Systems, Flagstaff, AZ, USA
[b] University of California Los Angeles, California Center for Sustainable Communities at the Institute of the Environment and Sustainability, Los Angeles, CA, USA

Customer Privacy Challenge: Navigating Privacy Aggregation Guidelines

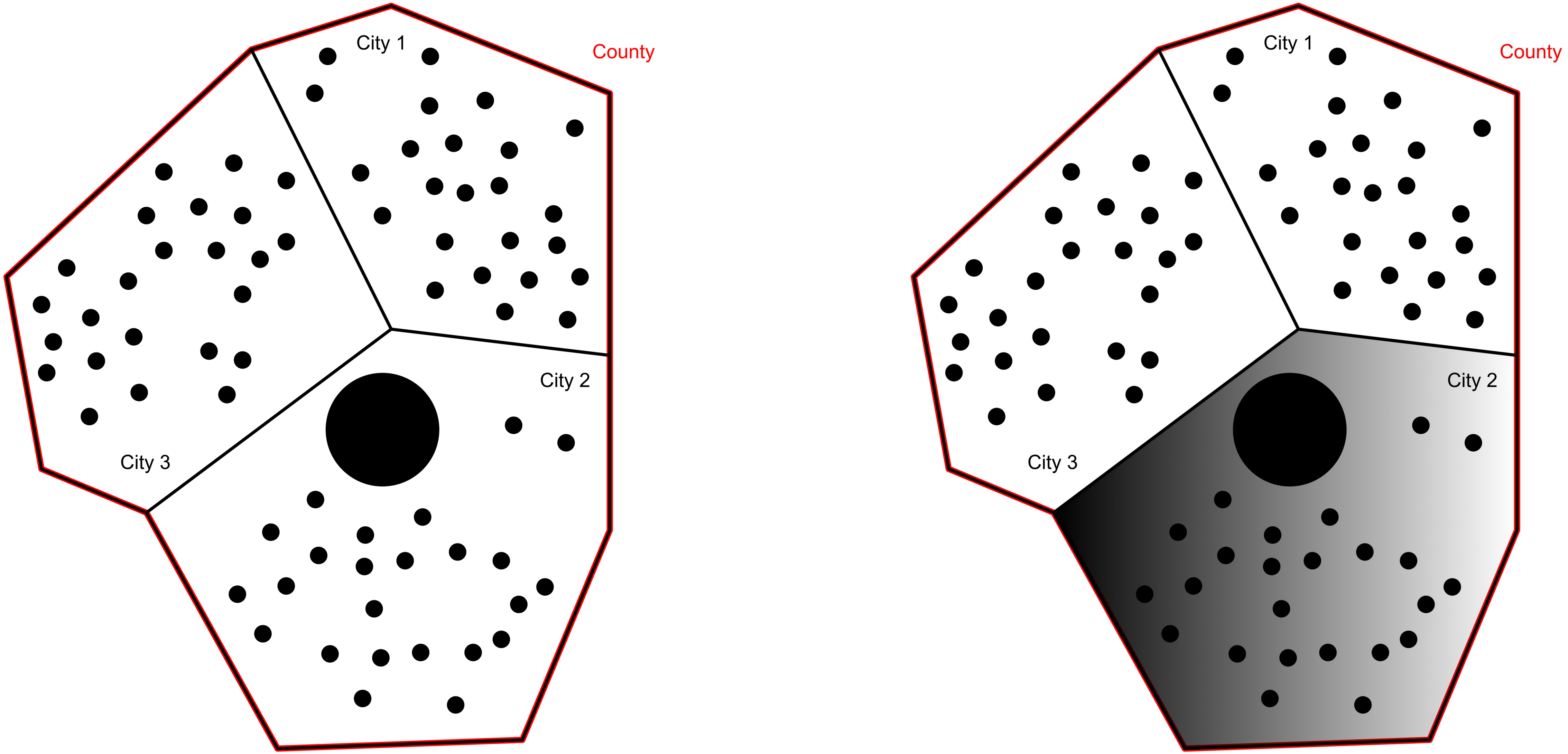


Within most large, urbanized service territories, if you plot the frequency distribution customers by their total usage you will probably see something that looks like this (top left).

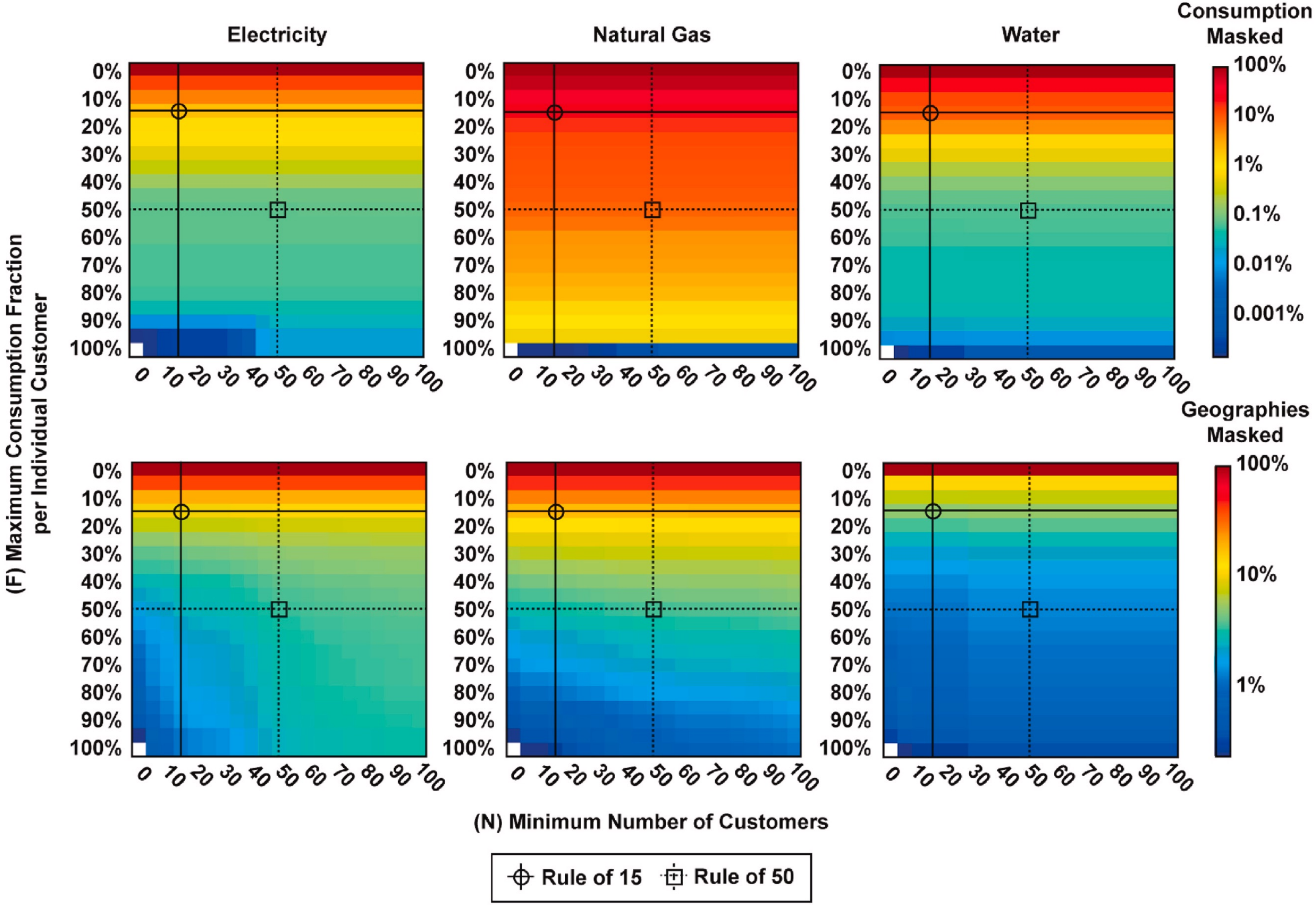
In our experience the distribution of customer level energy consumption data is “log-normal” – meaning that there are a small number of customers whose total usage is several orders of magnitude greater than the mean.

The, often sparse, geographic distribution of these “whale” customers, as we call them, can create significant challenges for energy usage data disclosure within the context of 15/15 or 15/100 style aggregation rules.

This is because it is highly likely that an arbitrary polygon will contain one customer whose usage is > 15% of the total for all customers in the group.



In this example, customer usage data would be publishable when aggregated to the county level (left) but not at the city level (right). The data for City #2 would have to be masked. This would potentially create a problem if trying to publish data across multiple levels of spatial aggregation simultaneously, as you must defend against N-1 deanonymization attacks.



The are other privacy preserving data anonymization schemes out there, such as "differential privacy," which are not based upon aggregation techniques, but they can be difficult to understand and complex to implement.

Our analysis found that changing the rule from 15/15 to 50/50 would provide a stronger guarantee of privacy to the average customer and significantly reduce the amount of masking that must take place.

Fig. 1. Empirical results from the application of a large number of masking rules to a large sample of real-world UCD taken from the UCLA Energy Atlas and Water Hub. Each rule represents a different pairwise combination of incremental values of N, the minimum number of accounts (horizontal axis) and F, the maximum fraction of consumption associated with a single account (vertical axis) The masking rate associated with each rule is plotted using a gradient colormap: blue = low, red = high. Masking is assessed both in terms of the percentage of consumption masked (top row) and the percentage of geographies masked (bottom row).

Thank You!

efournier@ioes.ucla.edu