Learning the Indicators of Energy Burden for Knowledge Informed Policy

Jasmine Garland[a*] **Rajagopalan Balaji**[a] **Kyri Baker**[a] **Ben Livneh**[a]

The University of Colorado at Boulder, USA[a]
jasmine.garland@colorado.edu*

Abstract

The United States is one of the largest energy consumers per capita, which puts an expectation on households to have adequate energy expenditures to keep up with modern society. This adds additional stress on low-income households that may need to limit energy use due to financial constraints. This paper investigates energy burden, the ratio of household energy bills to household income, within the United States West. Self-Organizing Maps, an unsupervised neural network, is used to learn the indicators attributed to energy burden to inform public policy. This is one of the first studies to consider environmental justice indicators, which include outdoor air quality metrics and health disparities as energy burden indicators. The results show significant ($p < 0.05$) differences among high energy burden areas and those with no energy burden for the environmental justice indicators. Thus, beyond the socioeconomic hardships of marginalized communities, counties with high energy burden suffer from environmental and health hazards, which will be amplified under a changing climate.

1 Introduction

Governing bodies at the local, state, and federal levels have recognized the challenges climate change brings for traditionally marginalized communities. However, how to create equitable and just public policies at the confluence of climate change, energy, and disadvantaged communities remains a complex question [1] with a dearth of research [2]. When creating public policy, meaningful indicators, including social indicators, are vital to well-informed policies. Well-informed policies are knowledge-influenced, meaning indicators have been thoroughly understood before creating the policy. Otherwise, it is purely a political policy [3]. Since energy impacts almost every dimension of modern society, when access to energy is limited, these impacts are compounded through housing, mobility, health, work, education, and other facets of life [4]. Energy resource access plays a vital role as compounded climate and electric infrastructure events occur. For example, the 2021 winter blackout in Texas left approximately 10 million people without electricity for as long as multiple days. Many news outlets reported that minority neighborhoods were disproportionately impacted [5]. While there are many constructs regarding inequality in the energy ecosystem, reviewed and compared in [4] and [6], this paper will focus on Energy Burden provided in Eq. 1.

$$\text{Energy Burden (\%)} = \frac{\text{Energy Bills (\$)}}{\text{Income (\$)}} \tag{1}$$

Submitted to Tackling Climate Change with Machine Learning: workshop at NeurIPS 2024.

The United States (U.S.) Department of Energy (DOE) states that households experiencing an energy burden of 6% or greater are considered to have a high energy burden, and households with an energy burden of 10% or higher have a severe energy burden. These metrics were created based on the notion that a household should not spend more than 30% of the income on housing expenses, and utility costs should not exceed 20% of housing expenses, although this often does not include transportation energy or water use [7, 8]. A high energy burden can result in shutoffs and "bundled burdens" such that economic trade-offs create a cumulative household risk. Trade-offs include living in comprised homes and the "heat or eat" phenomena, which refers to the co-occurrence of food and energy insecurity [9]. However, co-occurrences are not limited to energy and food but include medical care, proper shelter, and other life necessities. As global air temperatures continue to rise due to climate change, there is a precedent to predict how household energy use will change. An example is the 2022 California Heatwave, which led to the California Independent System Operator setting a new peak demand. The Western Interconnection also set a new peak demand in July 2024 [10]. Findings from [11] suggest that many areas will experience an increase in air conditioning (AC) expenditures but a decrease in heating expenditures as year-round temperatures are projected to increase. This creates a continuous need for increased use and adoption of AC in residential buildings [12]. However, previous studies have often concentrated on the interconnections of heating energy consumption, poverty, and residential dwellings. As such, this study focuses on the summer season, using summer temperature data and investigating AC types in residential buildings in the U.S. Western Census Region. The contributions of this study include:

- Understanding the implications of energy burden beyond energy use, social demographics, and income. These metrics are included in a majority of previous studies. While this paper does include those metrics, it is one of the first studies to include environmental justice indicators and resiliency measures.
- The development of a data-driven framework to identify energy burden indicators that are flexible to new inputs or different geographic regions.
- Poses to answer the question "What indicators are important in developing well-informed policies for energy burdened areas?"

2 Methods

The purpose of using unsupervised learning is to form groupings using general patterns and distributions related to indicators attributed to energy burden. These groupings could be used to better understand the characteristics linked to energy burden and the individuals most likely to experience a high energy burden. The intent is that through the development of such groups or profiles, policymakers and energy assistance programs can be better informed of the demographics and general characteristics of the individuals who will be impacted. A self-organizing map (SOM) is a form of unsupervised neural network based on competitive learning that has the capability to non-linearly map multi-dimensional data into two dimensions (2D). SOMs partition the data into nodes, which can be thought of as clusters, that are arranged in a 2D rectangular grid. This allows a SOM to embed a topological 2D manifold within a high-dimensional space. With each node, there is an associated weight vector; this corresponds to the node's position in the input space. Through an iterative process, for each data point, x_i, in the input data, the Euclidean distance to each node's weight vector is computed. The node whose weight vector is most similar to x_i is called the Best Matching Unit (BMU). The weights of the BMU and nodes within its neighborhood radius are then adjusted toward the input data while the SOM nodes stay in a fixed position. This naturally preserves and embeds a topological structure. Other clustering algorithms, such as k-means, allow the data to move in clusters with no direct relationship to one another [13]. The topological structure of SOMs is important, as counties near each other may have similar characteristics due to local and regional policy implications.

To date, SOM's have been implemented in energy related studies [14, 15] as well as studies investigating social-demographic disparities [16]. This paper aims to integrate both energy and social

Table 1: Data Categories Overview

Data Source	Data Description
Low-Income Energy Affordability Data Tool [18]	Average county energy burden (%).
Environmental Justice Screening Tool (EJ Screen) [19]	Contains 11 environmental indicators and six demographic indicators.
Community Resilience for Equity and Disasters Tool [20]	Measure of community resilience: zero risk, one-two risk, or three risk.
ResStock [21]	Housing stock metrics (type of cooling, energy expenditures, and the building type).
NOAA [22]	Mean temperature for the month of July.
Power Outage Data [23]	Average Power Outage information (outage duration, frequency, and customers impacted).

aspects to better understand energy burden for the first time to the author's knowledge. For this study, energy burden, latitude, and longitude were the inputs to the SOM, and the corresponding indicators were extracted using the latitude and longitude. The SOM used a 3x4 grid, and nodes were then grouped based on low, medium, and high rankings. Nodes with low energy burden are nodes that have an average energy burden of 4% or lower. Medium rankings have an average of 5% and/or contain a county with a high energy burden, while the high energy burden ranking nodes have an average of 6% or higher. Here, energy burden is rounded to the nearest integer percentage. Each indicator and the partitioned SOM nodes are then subjected to a Kruskal-Wallis test to determine whether that indicator has the same distribution across all SOM nodes. The Kruskal-Wallis test is a nonparametric equivalent to a one-way Analysis of Variance and is used in this study since it does not assume a normal distribution. To test for significance, the resulting p-values are obtained.

3 Results

Data For a better understanding of indicators for energy burden in the U.S. West, states within the West census division are considered. Census regions provide geographic frameworks at larger scales to perform statistical analysis, summarize data, and offer varying physical and cultural geography [17]. This study uses 45 indicators to understand influential indicators regarding energy. A full summary of each indicator may be found in the Supplemental Material. A brief overview of the data categories used is provided in Table 1. The data used in this study is at the county level, given the constraints of obtaining secure, publicly available data regarding socioeconomic status, residential dwellings, and power outages.

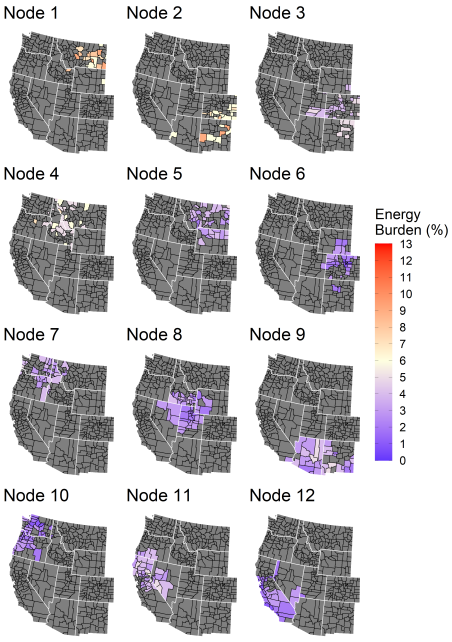

Figure 1: U.S. West Self Organizing Map

Self-Organizing Maps The results for the SOM are spatially shown in Figure 1. Here, nodes 1 and 2 are the high energy burden nodes while nodes 4 and 11 are the medium nodes. Node 1 contains 18 counties from Montana, Idaho, and Wyoming. Node 2 contains 25 counties from Colorado, New Mexico, and Arizona. Node 4 contains 31 counties from Montana, Idaho, Washington, Oregon, and Utah. Node 11 contains 40 counties and is largely made up of Northern California, Nevada, and Southern Oregon. The nodes provide

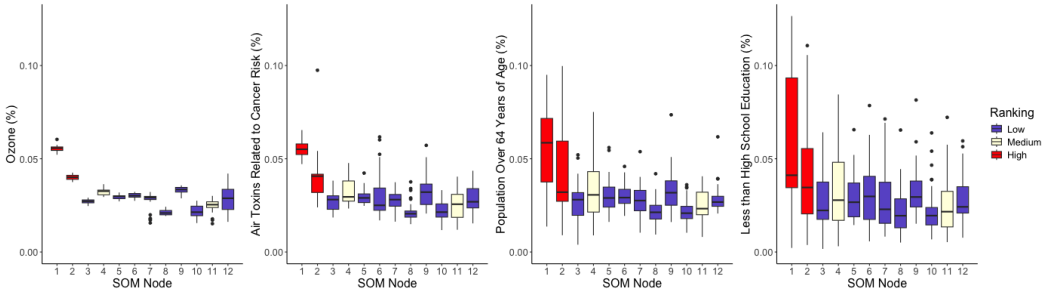

Figure 2: Distribution of Select Energy Burden Indicators

insight into similar counties being near each other within states, as clusters are typically formed. The spatial patterns are similar to findings in [24], which predicts areas in similar regions may experience similar burdens. Overall, the metrics relating to air quality, age, education level, and minority status are disproportionately higher in the nodes that correspond to high energy burden. Figure 2 provides select results while the distribution for all indicators are provided in the Supplemental Material along with the Kruskal-Wallis p-values.

Based on the Kruskal-Wallis results, 41 out of 45 indicators showed statistical significance ($p < 0.05$) among the SOM nodes, which means there are significant differences in the medians between the SOM nodes. This is important as it could inform policymakers in terms of energy burdened areas and also provides insight into geographic regions that could benefit from policy or aid as household energy demand increases and energy remains an indispensable aspect of life. However, the nexus between environmental justice, socioeconomic factors, housing, community resilience, and power outages are not always independently associated, and results from this study show they disproportionately impact counties that are already experiencing a high energy burden. Thus, results are advisable for creating knowledge-informed policy as shown in Figure 3. Such policies and the knowledge acquired through learning the indicators of energy burden create a path to climate action to avail and eliminate the compounded burdens of climate change on marginalized communities.

For instance, indicators related to air quality had higher values in high energy burden nodes, meaning particles related to ozone, cancer risk, respiratory hazard, and diesel particulate matter were consistently higher. Thus, identifying emissions sources, such as fossil fuel power plants, and replacing those sources with sustainable solutions, such as renewable energy, could improve air quality and aide in solving other environmental injustices such as polluted water from power sources. Additionally, these nodes also show higher percentages of the population in which poverty status is determined, a high population of individuals over 64 years of age, a higher percentage of adults with less than a high school education, and higher percentages of homes built prior to 1960. This further provides insight into the types of policies and programs, such as energy assistance programs, alternative energy rates, and weatherization programs that policy makers should consider. However, they must also consider the influence of age, race, and the traits of those living in such households to maintain equitably. Future studies should consider different methods to determine the SOM grid topology. This framework could additionally be used on different census regions or different geographic regions

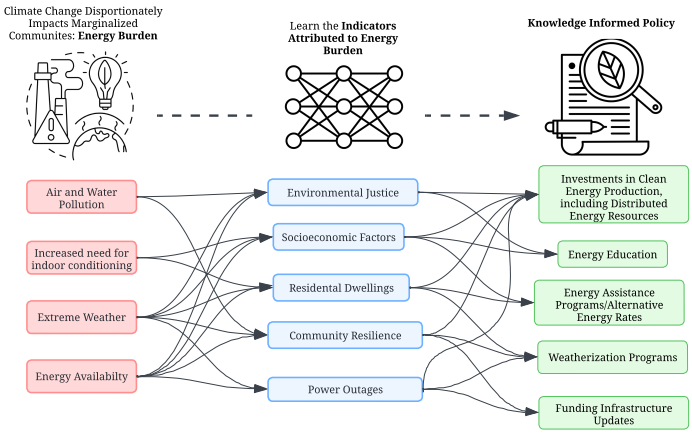

Figure 3: Path to Knowledge Informed Policy

such as states, voting districts, or census regions, which would be complementary to the development of publicly available data at more granular geographic scales.

References

[1] Oleg Golubchikov and Paola Deda. Governance, technology, and equity: An integrated policy framework for energy efficient housing. *Energy Policy*, 41:733–741, February 2012.

[2] Sonal Jessel, Samantha Sawyer, and Diana Hernández. Energy, poverty, and health in climate change: A comprehensive review of an emerging literature. *Frontiers in Public Health*, 7, December 2019.

[3] Judith Eleanor Innes. *Knowledge and Public Policy: The Search for Meaningful Indicators*. Transaction Publishers. Google-Books-ID: Qyd97IigMvEC.

[4] Marilyn A. Brown, Anmol Soni, Melissa V. Lapsa, Katie Southworth, and Matt Cox. High energy burden and low-income energy affordability: conclusions from a literature review. *Progress in Energy*, 2(4), 10 2020.

[5] James Dobbins and Hiroko Tabuchi. Texas blackouts hit minority neighborhoods especially hard. Available at `https://www.nytimes.com/2021/02/16/climate/texas-blackout-storm-minorities.html` (2021/11/26).

[6] Bethel Tarekegne, Barbara Pennell, Danielle Preziuso, and Rebecca O'Neil. *Review of Energy Equity Metrics*. Number PNNL-32179, 1830804. Oct 2021.

[7] Joel F. Eisenberg. Weatherization assistance program technical memorandum background data and statistics on low-income energy use and burdens. Available at `https://info.ornl.gov/sites/publications/Files/Pub49042.pdf`.

[8] Ariel Drehobl, Lauren Ross, and Roxana Ayala. How high are household energy burdens? *American Council for an Energy-Efficient Economy*, 2020.

[9] Jamal Lewis, Diana Hernández, and Arline T. Geronimus. Energy efficiency as energy justice: addressing racial inequities through investments in people and places. *Energy Efficiency 2019 13:3*, 13:419–432, 11 2019.

[10] Dede Subakti. Managing the july 2024 heat wave with our partners in california and the west, July 2024.

[11] Haojie Wang and Qingyan Chen. Impact of climate change heating and cooling energy use in buildings in the united states. *Energy and Buildings*, 82:428–436, 10 2014.

[12] L. Ortiz, H. Gamarro, J. E. Gonzalez, and T. McPhearson. Energy burden and air conditioning adoption in new york city under a warming climate. *Sustainable Cities and Society*, 76:103465, 1 2022.

[13] T. Kohonen. The self-organizing map. *Proceedings of the IEEE*, 78(9):1464–1480, 1990.

[14] Teemu Räsänen, Juhani Ruuskanen, and Mikko Kolehmainen. Reducing energy consumption by using self-organizing maps to create more personalized electricity use information. *Applied Energy*, 85(9):830–840, September 2008.

[15] Cristian Bucur, Bogdan George Tudorică, Simona-Vasilica Oprea, Dumitru Nancu, and Dorel Mihail Duşmănescu. Insights into energy indicators analytics towards european green energy transition using statistics and self-organizing maps. *IEEE Access*, 9:64427–64444, 2021.

[16] Eric Vaz, Fernando Bação, Bruno Damásio, Malik Haynes, and Elissa Penfound. Machine learning for analysis of wealth in cities: A spatial-empirical examination of wealth in toronto. *Habitat International*, 108:102319, February 2021.

[17] US Census Bureau. Geographic areas reference manual.

[18] Ookie Ma, Ricardo Oliveira Jon Weers Krystal Laymon, Megan Day, and Aaron Vimont. Low-income energy affordability data (lead) tool methodology. Available at `https://lead.openei.org/assets/docs/LEAD-Tool-Methodology.pdf` (2021/12/08).

[19] United States Environmental Projection Agency. Ejscreen: Environmental justice screening and mapping tool. Available at `https://www.epa.gov/ejscreen` (2021/11/26).

[20] United States Census Bureau. About community resilience estimates. Available at `https://www.census.gov/programs-surveys/community-resilience-estimates/about.html` (2021/11/26).

[21] National Renewable Energy Laboratory. Resstock.

[22] National Oceanic and Atmospheric Administration National Centers for Environmental Information. Climate modeling - monthly report.

[23] Christa Brelsford, Sarah Tennille, Aaron Myers, Supriya Chinthavali, Varisara Tansakul, Matthew Denman, Mark Coletti, Joshua Grant, Sangkeun Lee, Karl Allen, Evelyn Johnson, Jonathan Huihui, Alec Hamaker, Scott Newby, Kyle Medlen, Dakotah Maguire, Chelsey Dunivan Stahl, Jessica Moehl, Daniel Redmon, Jibonananda Sanyal, and Budhendra Bhaduri. A dataset of recorded electricity outages by united states county 2014–2022. *Scientific Data*, 11(1):271, March 2024.

[24] Chien-fei Chen, Jimmy Feng, Nikki Luke, Cheng-Pin Kuo, and Joshua S. Fu. Localized energy burden, concentrated disadvantage, and the feminization of energy poverty. *iScience*, 25(4):104139, Apr 2022.