Mariana I. Fernández Correa | CEE 501.78 – Hazard Analysis

Abstract:

Hurricane Maria's landfall in Puerto Rico in 2017 triggered the longest blackout in United States history, resulting in 2,975 deaths and an estimated $90 billion in damages across PR and the US Virgin Islands (National Oceanic and Atmospheric Administration [NOAA], 2018; Natural Hazards Center, 2018). While post-disaster recovery strategies prioritized critical infrastructure, such as hospitals and economic centers, these efforts were significantly complicated by highly variable wind speeds across the island. Although previous research has used regression models to assess recovery predictability (Azad, S., & Ghandehari, M., 2021), there remains a notable gap in the literature on the creation of composite indices that can be broadly applied to power recovery prioritization as they relate to power recovery time.

Using data extracted from publicly available government documents and official correspondence (Central Office for Recovery, Reconstruction, and Resiliency [COR3], 2019), this study employs the methodology provided in the Handbook on Constructing Composite Indicators (OECD/European Commission Joint Research Center [JRC], 2008) to analyze these "non-wind-driven" variability gaps. By isolating these instances from the broader spectrum of recovery challenges, this research evaluates the impact of established prioritization protocols. The findings offer a critical assessment of the logistical and structural vulnerabilities that hindered power restoration following Hurricane María. Finding that critical infrastructure and socioeconomic variables overlap with Hurricane María's power recovery after wind effects are removed.

1.0 Introduction:

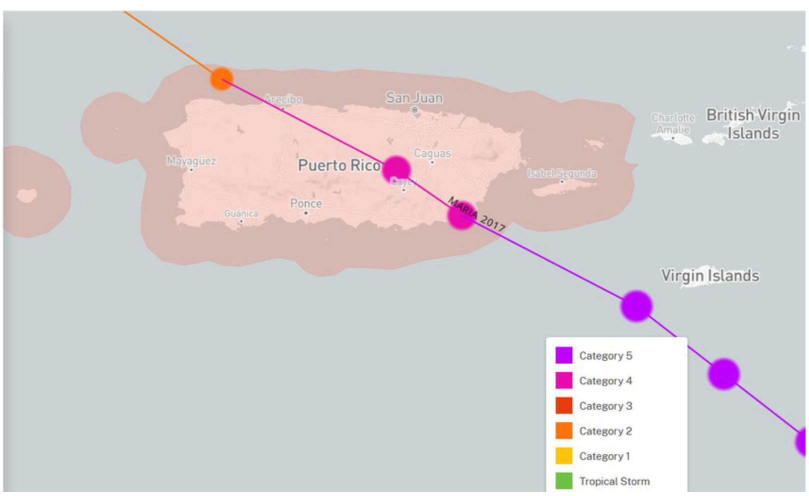


Figure 1: Speed and cetogrization of Hurricane Maria, National Weatehr Service

Puerto Rico's power systems have long been a focus of concern, especially following the events of 2017. That year, the island's power grid, comprising six major power plants and several smaller generation facilities, was devastated by Category 4 Hurricane María. The hurricane led to 2,975 direct and indirect deaths (Natural Hazards Center, 2018) and inflicted approximately $90 billion in damages on Puerto Rico and the Virgin Islands (NOAA, 2018). Most areas experienced 15-20 inches of rainfall and maximum sustained winds of 155 mph (National Weather Service [NWS], 2017). Figure 1 illustrates the variability in wind speeds and their categorization along the hurricane's path.

Hurricane María resulted in the longest blackout in U.S. history (NOAA, 2018). Due to the unprecedented duration of the outage, restoration efforts focused on high-priority locations, such as hospitals, economic centers, and critical public infrastructure (COR3, 2019). Nevertheless, prioritization was only one factor influencing recovery. As Figure 1 reveals, wind speeds varied greatly across the island, further complicating efforts to restore power.
Wind was a major cause of damage, significantly hindering recovery efforts. This paper investigates how non-wind-driven gaps in recovery align with the priority infrastructure identified by management agencies. While previous research has used regression analyses to assess future predictability (Azad, S., & Ghandehari, M., 2021), few studies have sought to develop an index to address these gaps.

To do this, I used publicly available government letters and documents to identify priority infrastructure and correlate them with blackout days not attributable to wind.

2. Literature Review

2.1 Handbook on Constructing Composite Indicators: Methodology and User Guide

The Handbook on Constructing Composite Indicators: Methodology and User Guide (published by the OECD and the European Commission) is considered the "gold standard" reference for statisticians, researchers, and policymakers tasked with creating composite indicators. More details on the application in this report are in the methods section.

2.2 A Study on the Association of Socioeconomic and Physical Cofactors Contributing to Power Restoration After Hurricane Maria

This paper uses NASA's Black Marble data (National Aeronautics and Space Administration [NASA], n.d.) and fits a Quasi-Poisson regression to best explain the differences in outages and draw conclusions, shedding light on the impacts for residents with different socioeconomic backgrounds. Their variables were:

- Elevations across the island
- Outage duration – same one used in this report
- Distance from landfall
- Road connectivity
- Landslide risk
- Poverty Rate

They found that proximity to landfall had the largest impact on recovery times; areas with direct access to road infrastructure had faster connectivity to the power grid; landslides played a large role in estimating recovery days; and even after controlling for physical damage, poverty remained a significant predictor of delay. For every 10% increase in a population living below the poverty line, recovery time increased by 2%.
This paper was used to draw references and conclusions *after* *the analysis shown below was completed.* So, the methods and variable selection here are independent.

2.3 A Study on the Association of Socioeconomic and Physical Cofactors Contributing to Power Restoration After Hurricane María

This paper serves as a reference on constructing and analyzing a composite index in Puerto Rico. It walks through the nuances of Puerto Rico's census data and finds that variables such as English proficiency are less relevant than others. The paper finds that 613 of 884 tracts saw a decline in population as people migrated out of Puerto Rico following the 2017 hurricane María. Previous data has been collected to measure the days of outage using the formula shown below:

Equation 1: Days without power averaged through the entire test period

$$\boldsymbol{DayswithoutPower = (1 - A) * B}$$

Where:

A: Mean available percent radiance between 20th September 2017 and 20th May 2017.

B: Number of days between 20th September 2017 and 20th May 2017.

3. Data Description

The data used for this project were focused on highlighting important infrastructure (Energy, Communications, Health, Transportation, and water) and two socioeconomic factors: population density and economic output. There are broader manuals calling for the use of the variables used here. Although no direct source structured what they considered critical infrastructure explicitly and comprehensively, publications relating to allocation of money regarding power recovery had lines such as:

"To allocate distribution investments (cost estimates) at the district level, the Working Group used the following factors: Initial post-hurricane damage assessment, The number of critical loads (hospitals, airports, sanitation facilities, etc.) and priority loads (urban centers, etc.) that need underground or redundant supply, Customer density, The total number of distribution poles that need to be upgraded" (COR3, 2019)

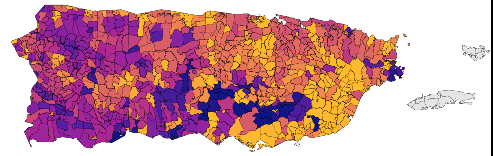
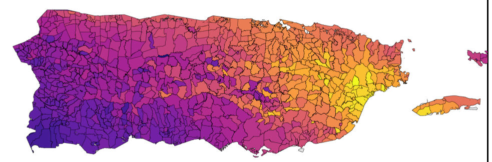
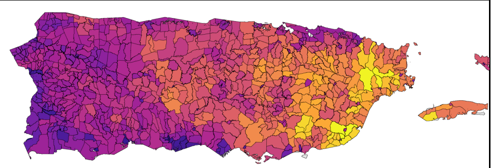
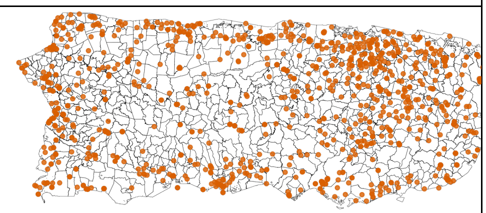
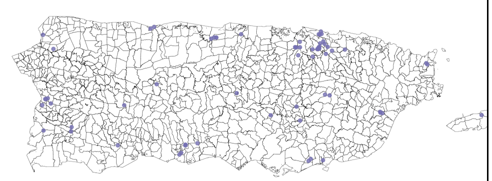
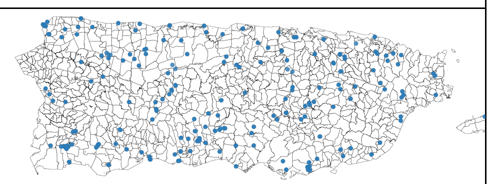
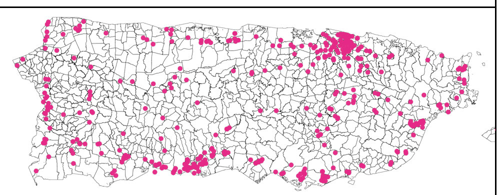
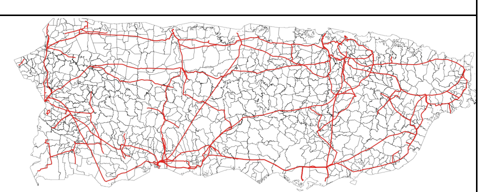
Additionally, post-recovery, FEMA's Community Lifelines serves as a framework for what the dataset of critical infrastructure is; it is a standardized, designed to help emergency managers and responders prioritize and stabilize the essential services required to keep a community functioning during a disaster (Federal Emergency Management Agency [FEMA], 2019), restoring normal societal functions as quickly as possible.

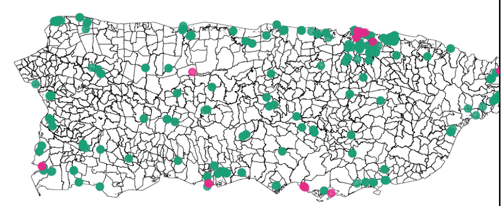
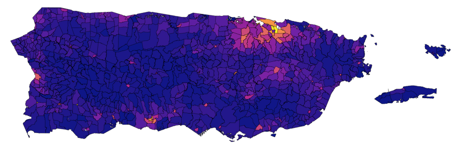
The 8 Community Lifelines:

- **Safety and Security:** Law enforcement, fire services, and search and rescue.
- **Food, Hydration, Shelter:** Supply chains and distribution for basic survival needs.
- **Health and Medical:** Medical care, public health, patient movement, and medical supplies.
- **Energy:** Power grid systems, fuel, and grid stability.
- **Communications:** Infrastructure, 911 dispatch, responder comms, and financial systems.

- **Transportation:** Highways, mass transit, aviation, and maritime assets.
- **Hazardous Materials:** Management of waste, pollutants, and contaminants.
- **Water Systems:** Potable water distribution and wastewater management.

Using definitions of critical infrastructure and communications from the government about prioritization with data availability, the following data has been collected to define the dataset for this project.

Hierarchy	**Variable**	**Formula/Processing**	**Source**	**Image (Spatial Component)**
Consequence	Recovery Days	% of recovery or Max Data days	Nasa Nightlight	
Hazard	Mean Windspeed	Area Averaged	NIST.gov	
	Max Windspeed	Area Averaged	NIST.gov	
Vulnerability	Telecom infrastructure	Count	OpenCellID	
	Hospitals	Count	HIFLD	
	Water treatment and pumping infrastructure	Count	OpenStreetMap	
	Power plants and Substation	Count	OpenStreetMap	
	Transmission lines	Sum Miles	OpenStreetMap	

	Ports (green) & Airports (Pink)	Count	HIFLD	
Demographic	Population Density	Total Pop / area Barrio (area weighted)	2017 ACS 5-year Census	

4. Methods

The methods of this project follow the steps outlined in the Composite Index Handbook, as mentioned in the literature review. The spatial unit used was Barrios. Puerto Rico has 920 barrios (neighborhoods) and 78 Municipios (cities) (U.S. Census Bureau, [2017]). A comparison of the two can be seen below

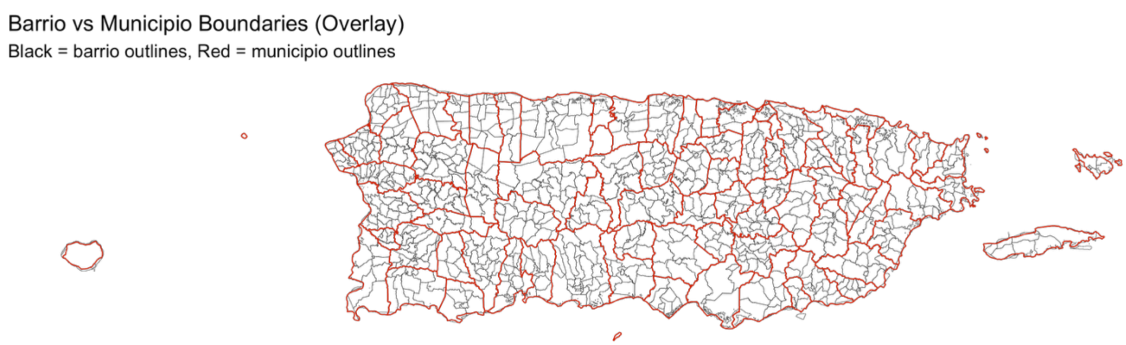


Figure 2: Outline overlay of TIGER/Line files for Barrios and Municipios

4.1. Theoretical Framework - This step defines the conceptual foundation of the index. This grounded the methodology in the FEMA Community Lifelines framework and Puerto Rico's energy policy mandates. It establishes the rationale for why specific infrastructure and demographic variables are critical to measuring resilience.

4. 2. Data Selection - This phase involves the systematic identification and acquisition of data. Variables are selected based on their analytical soundness. In this stage, most of the literature review was completed, and I chose to drop most socioeconomic variables, as I couldn't find strong evidence that anything other than economic output and high population density was prioritized; therefore, I excluded all variables in the 10-SVI (West, 2023) listed below. Since they focused on different outcomes from the hurricane.

Poverty Rate: Population under poverty / Total population

- **Unemployment:** Unemployed / Labor force
- **No HS Diploma:** Population without HS diploma / Population 25+
- **Age 17 or below:** (Population 17 and under) / (Total population)
- **Disability Rate:** Population 5+ with a disability / Population 5+
- **Crowding:** (Total number of people in household) / (Total number of rooms)
- **Single Parent:** Family households with no spouse present, with own children under 18 / Total family households
- **Age 65+:** (Population 65+) / (Total population)
- **Median Household Income**

4.3. Imputation of Missing Data - This step is about creating the rules of what happens for missing data. In our case, no data was missing that needed to be accounted for (other than possible incomplete databases, which are not accounted for).

4.4. Multivariate Analysis - A statistical review was conducted to explore the relationships between variables. This ensures that the variables chosen are not redundant (i.e., they do not measure the same underlying trend twice), thereby maintaining the index's efficiency and clarity. This report uses Principal Component Analysis (PCA) to analyze the data. Figure 4 and Figure 5.

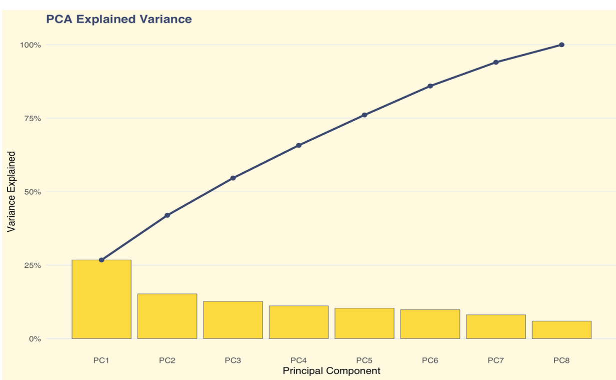


Figure 3 : PCA Analysis

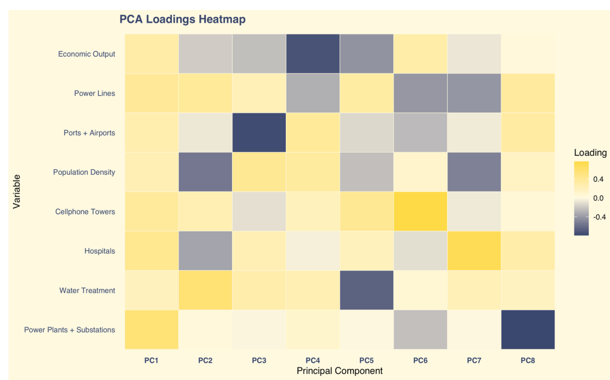


Figure 4: Loadings for each variable

4.5. Normalization - Since raw data exists in different units, such as miles of transmission lines versus population counts, this step transforms all indicators onto a common, standardized scale, using z-score standardization. This enables the mathematical comparison and combination of disparate data types.

$$z = \frac{x - \mu}{\sigma}$$

4.6. Weighting and Aggregation - This step involves assigning specific importance to each indicator and synthesizing them into a single score. Weights are determined by their policy-driven priorities, ensuring that "lifeline" infrastructure (such as hospitals and water treatment) exerts a proportional influence on the final resilience score. Based on the literature review, a weighting of 0.7 for infrastructure and 0.3 for socio-economic was used, using a multi-criteria approach, the following formula was used:

Equation 2: Multi-criteria analysis

$$index\ score = \sum w\_i * z\_i$$

Table 1: Importance Factor of each variable using a 30/70

Column 1	Tier Classification	Importance factor
Economic Output	Demographic/Economic	0.15
Power Lines	Infrastructure (Critical)	0.1166666667
Ports + Airports	Infrastructure (Critical)	0.1166666667
Population Density	Demographic/Economic	0.15
Cellphone Towers	Infrastructure (Critical)	0.1166666667
Hospitals	Infrastructure (Critical)	0.1166666667
Water Treatment	Infrastructure (Critical)	0.1166666667
Power Plants + Substations	Infrastructure (Critical)	0.1166666667

Where w_i is the weight for each normalized variable (defined in the data section above), z_i for the barrio (using index I for each barrio), after the index score for the barrio was calculated, I used 0-1 (min-max) normalization to make the system legible on the same scale.

Equation 3: Min-Max normalization

$$normlized\ index\ score = \frac{index\ score\ - min(index\ scores)}{Max(index\ scores) - Min(index\ scores)}$$

4.7. Robustness and Sensitivity Analysis - By varying the weighting parameters, the analysis confirms that the index's conclusions remain stable and are not unduly influenced by any single methodological assumption.

4.8. Back to the Real World - The final validation phase correlates the constructed index against observable, historical data, specifically recovery performance following Hurricane Maria through time. One thing is looking at averages; another is zooming in to see a single "barrio". The paper looks at Old San Juan, known for its colonial Spanish architecture and as a famous tourism spot, and Orocovis Downtown, considered the heart of Puerto Rico for its central geographic location on the island.

5. Results

The results (shown in Figure 6) indicate that the metro area and two areas in the south are designated as priorities for power recovery.

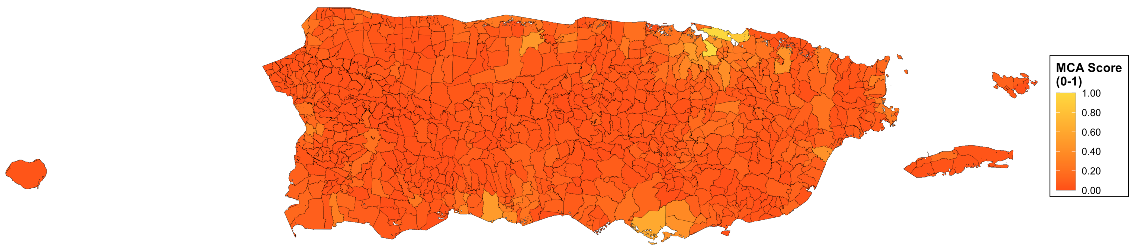


Figure 5: Results of a Multi-Criteria Approach with a 70/30 split weight between infrastructure and socioeconomic variables

6. Discussion: Interpretation of results, comparison to prior work, limitations

6.1 Interpretation of PCA results

The Primary Driver of Resilience is Critical Infrastructure (PC1). The highest values in the first column are for Powerplants/Substations (0.5481) and Hospitals (0.4218). Considering that PC1 explains the largest proportion of the variability, it makes sense that power generation and hospitals have the highest weights.

Water Infrastructure and Population Density are Distinct (PC2). There is a clear trade-off here. Water treatment/pumping (0.5703) and Population density (-0.5575) move in opposite directions.

Lastly, the scree plot shown in the methods indicates that no PC is overly explanatory of variability, so overall, the dataset is balanced.

Table 2: Loadings with numerical values of PCA

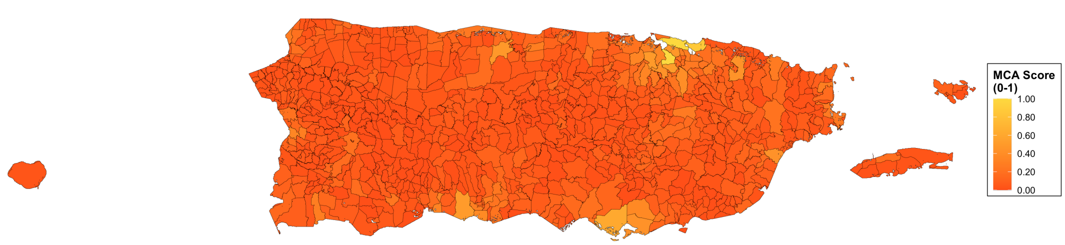
variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
economic output	0.3014	-0.1838	-0.2423	-0.7331	-0.4312	0.2945	-0.0771	0.0242
power lines	0.3774	0.3705	0.2158	-0.3098	0.3234	-0.4140	-0.4197	0.3544
ports airports count	0.2666	-0.0683	-0.7675	0.3712	-0.1339	-0.2611	-0.0573	0.3286
population density	0.2291	-0.5575	0.4021	0.3477	-0.2429	0.1126	-0.5109	0.1461
cellphone towers count	0.3597	0.2428	-0.1058	0.1986	0.4077	0.7670	-0.0618	0.0518
hospitals count	0.4218	-0.3580	0.2359	-0.0355	0.1915	-0.1070	0.7100	0.2913
water treatment pumping count	0.1880	0.5703	0.2770	0.2364	-0.6554	0.0627	0.2138	0.1660
powerplants substations count	0.5481	0.0327	-0.0172	0.1064	-0.0091	-0.2394	-0.0063	-0.7934

6.2 Sensitivity and Uncertainty

According to the handbook, a robust Composite Index (CI) should include an uncertainty and sensitivity analysis. The sources of uncertainty addressed in this report are *only* related to the weights, "Importance factors" of the report; however, there is still uncertainty regarding other factors:

- *The impact or lack thereof of volume:* In this report, all structures were considered equals, but this is an oversimplification. The most notable example is the variable of ports and Airports. Puerto Rico has 3 international airports: Luis Muñoz Marín International Airport (SJU) in San Juan, Rafael Hernández International Airport (BQN) in Aguadilla, and Mercedita International Airport (PSE) in Ponce. However, there are many more local airports for recreation or associated with military bases. Among the top three airports, SJU, the largest, is visited by ~13.2 million passengers per year (De Último Minuto, 2026), while the third largest has only around 284 thousand visitors each year (Instituto de Estadísticas de Puerto Rico, n.d.). For the purposes of this study, *counts* ignore SJU's large societal impact over PSE.
- Compensation, using a linear aggregation method such as multi-criteria aggregation (used here), implicitly makes decisions for the aggregation that may not uphold the actual prioritization process. Approaches such as a non-compensatory multi-criteria aggregation, where it's based on ranks, and not a summation across all factors, are a way to remove those compensation factors, where areas with lots of cellphone towers and no hospitals are treated the same as an area with a lot of hospitals (which is arguably a higher priority).

To assess uncertainty, the importance factors were assigned equal weights. This resulted in very minimal changes to the final index score, shown in Figure 7. A comparison can be seen in Figure 8.



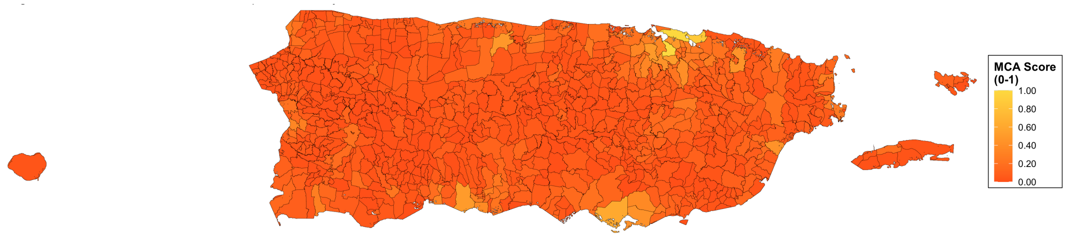


Figure 6: Original weighting top image, and equal weighting (bottom weighting) yielded little visual change

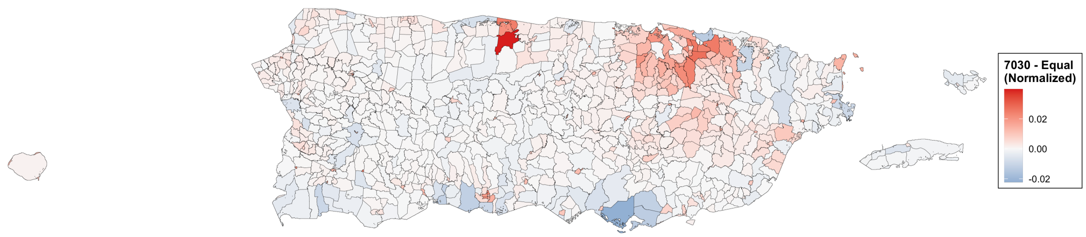


Figure 7: original weighting creates a higher power recovery index in the metro area and devalues the south, where most power generation takes place

6.3 Back to the Real world

The motivation for this composite index stems from prioritization and the large variability in power recovery following the 2017 Hurricane María. Using the formula and data depicted in equation , we get the average number of outage days, which is calculated using NASA's Black Marble for nighttime lights. A map of the average number of days without power by barrio can be seen here, in figure 8.

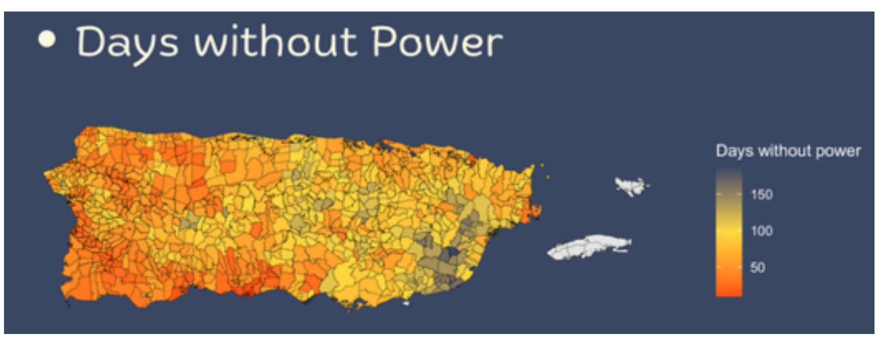
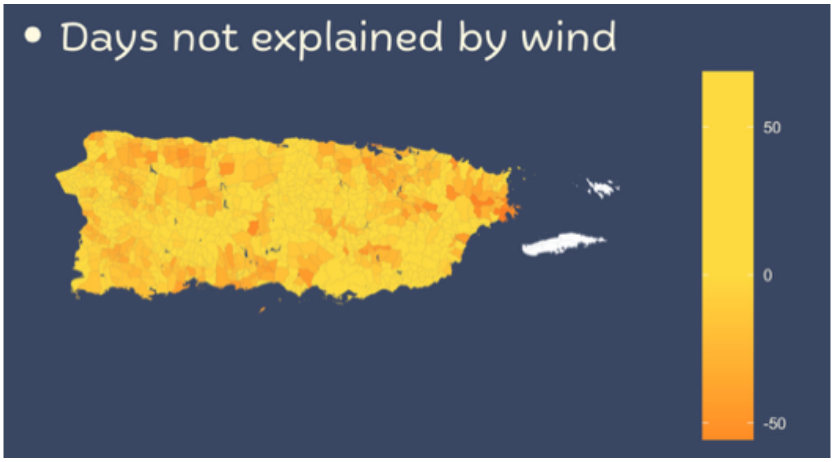


Figure 8: Replica of Average time without power by barrio from (PAPER)

To compare their recovery with a power recovery index, it's important to remove the effects of power line damage. To compare power recovery not explained by damage, the mean wind speed and the maximum wind speed were used in a linear regression, then subtracted from the days-without-power formula used in (PAPER). This transformed power recovery is seen in Figure 9



Compared to Figure 5, some overlap exists between the early recovery of some areas in Puerto Rico and the Composite index delineating priority, particularly in the metro area and the south of the island.

Figure 9: Days of Power not explained by wind

7 Conclusion

Creating relations from critical infrastructure and socioeconomic data isn't new (Azad, S., 2021); however, previous attempts have focused on a retroactive recovery approach. Using an index will make it less accurate than regressions on past events, but it can be referenced in the future because the weightings are based on prioritizing officials rather than being designed solely to accurately fit previous data.

Ultimately, recovery is nonlinear and highly varied across areas, so there are also limitations to using a composite index, such as how power recovery affects quality of life within households and power in people's jobs and communities. As a simplified example, Figure 10 compares Orocovis downtown and Old San Juan in their recovery over time. The graph shows relapses because of an overall fragile grid as it was being rebuilt.

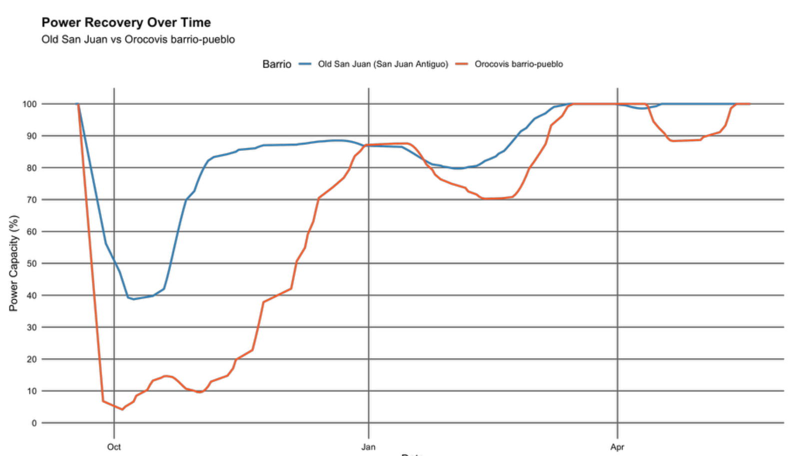


Figure 10: Recovery through time of Old San Juan and Orocovis.

Placing a single number on an area is a generalization of the area's residents, but can't help individuals much, since, like other aggregate measures, it only works at a large scale (island- or city-wide). But in terms of power recovery, there is also variability in how you define power recovery. Is power recovered when it's back to pre-disaster levels? Or is it considered recovered enough when essential services are powered? This index, in the future, can consider: Recovery as a threshold % rather than an average over a time period, using counts instead of volume. The index performed like the Azad paper, but future work addressing uncertainty (especially counts vs. volume) is probably best and remains to be done.

REFRENCES

Azad, S., & Ghandehari, M. (2021). A study on the association of socioeconomic and physical cofactors contributing to power restoration after Hurricane Maria. IEEE Access, 9, 98654–98664. https://doi.org/10.1109/ACCESS.2021.3093547

Central Office for Recovery, Reconstruction and Resiliency. (2019). Grid modernization plan for Puerto Rico. Government of Puerto Rico. https://recovery.pr.gov/es/documents/Grid%20Modernization%20Plan_20191213%20(2).pdf

De Último Minuto. (2026, January 9). Luis Muñoz Marín Airport in Puerto Rico breaks passenger records in 2025. https://deultimominuto.net/en/uncategorized/luis-munoz-marin-airport-in-puerto-rico-breaks-passenger-records-in-2025/

Federal Emergency Management Agency. (2019). Community lifelines implementation toolkit (Version 2.0). U.S. Department of Homeland Security. https://www.fema.gov/sites/default/files/2020-05/fema_community-lifelines_toolkit-2.0.pdf

Instituto de Estadísticas de Puerto Rico. (n.d.). Carga de pasajeros aéreos y marítimos [Data set]. https://www.estadisticas.pr.gov/productos/carga-y-pasajeros-aereos-y-maritimos

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., & Giovannini, E. (2008). Handbook on constructing composite indicators: Methodology and user guide. OECD Publishing. https://doi.org/10.1787/9789264043466-en

National Aeronautics and Space Administration. (n.d.). Black Marble. NASA Earthdata. https://www.earthdata.nasa.gov/data/projects/black-marble

National Oceanic and Atmospheric Administration. (2018, August 16). Hurricane Maria's devastation of Puerto Rico. Climate.gov. https://www.climate.gov/news-features/understanding-climate/hurricane-marias-devastation-puerto-rico

National Weather Service. (2017). Hurricane Maria - September 2017. https://www.weather.gov/sju/maria2017

Natural Hazards Center. (2018, August 28). Excess mortality and associated risk factors related to Hurricane Maria in Puerto Rico. University of Colorado Boulder. https://hazards.colorado.edu/public-health-disaster-research/excess-mortality-and-associated-risk-factors-related-to-hurricane-maria-in-puerto-rico

West, J. (2023). Social vulnerability and population loss in Puerto Rico after Hurricane Maria. Population and Environment, 45(2), Article 18. https://doi.org/10.1007/s11111-023-00418-3