Proceedings of the 35th European Safety and Reliability & the 33rd Society for Risk Analysis Europe Conference
Edited by Eirik Bjorheim Abrahamsen, Terje Aven, Frederic Bouder, Roger Flage, Marja Ylönen

doi: 10.3850/978-981-94-3281-3_ESREL-SRA-E2025-P9630-cd

Burning Inequities: Comparative Analysis of Socio-economic Drivers for Post-wildfire Resource Allocation in the Southwestern US States

Poulomee Roy

Industrial and Systems Engineering, State University of New York at Buffalo/Buffalo, New York, United States.
E-mail: poulomee@buffalo.edu.

Sayanti Mukherjee

Industrial and Systems Engineering, State University of New York at Buffalo/Buffalo, New York, United States.
E-mail: sayantim@buffalo.edu.

Wildfires are increasingly threatening the southwestern US states because of climatic extremes, heatwaves, dried vegetation, and anthropogenic interferences. While wildfire-prone regions in the US are more likely to be populated by higher-income groups, this fact overshadows the existence of thousands of low-income, underrepresented individuals, lacking resources to prepare for and recover from wildfires. However, state-level and local policies for wildfire management significantly differ across the states, driven by wildfire exposures, demographics, budgetary priorities, and political scenarios. Although there is a growing literature on wildfire management, there are limited studies analyzing state-level similarities and differences related to equitable wildfire resource allocation. To address this gap, this study aims to investigate the key socio-demographic and economic factors associated with post-wildfire resource allocation for the three southwestern US states (California, Arizona, Colorado), and compare/contrast the underlying inequities. Data on wildfire incidents and socio-demographic information is collected from multiple sources from 2015-2022, and interpretable machine learning models are implemented to evaluate the county-level social inequities in post-wildfire resource allocation across the states. Our preliminary results highlight that the disadvantaged Wildland Urban Interface (WUI) communities (higher proportions of low income, less education, Black and Hispanic populations) are disproportionately impacted by wildfires as opposed to their wealthier counterparts, which further worsened due to inadequate and inefficient post-wildfire resource allocations. The outcomes of this study will better inform strategic decisions and policymaking for equitable wildfire management.

Keywords: Wildfire management, Equitable resource allocation, Machine learning, Marginalized Community

1. Introduction

There is no denying that, due to climatic extremes, the increased rate of severe wildfires has posed significant threats to communities and critical infrastructures across the southwestern United States (US) [1, 2]. On this account, California (CA) and Arizona (AZ) witnessed the highest proportion of burnt areas from 2015-2022 [3], while in Colorado (CO), the increasing frequency of severe fires has devastating impacts on the wildland-urban interface (WUI) [4]. The menaces of these wildfires unveiled that inequality persists in post-wildfire recovery efforts and resource allocations. Often, the impoverished communities suffer the most owing to the scarcity of resources available to cope with the destruction [5]. For example, during the 2025 Pacific Palisade fire in Los Angeles CA, the City of Malibu, where residences of high-profile Oscar-winning celebrities like Jeff Bridges and Mel Gibson are situated, received substantial firefighting resources and substantial national and social media attention [6]. On the contrary, vulnerable communities significantly suffered [7]. A similar story unfolded during the 2011 Wallow fire in AZ. The remote communities had limited personnel and equipment to confront the inferno, while a resort community promptly received top-notch resources and infrastructure to manage the wildfire [8]. Similarly, in the aftermath of the 2020 East Troublesome Fire in CO, many communities with no/inadequate wildfire insurance experienced disparities as they were stuck in the administrative process of obtaining relief funds, which delayed rebuilding the community [9].

Though all three states face challenges in efficient wildfire management and resource allocation, the nature of the challenges is unique. For example, CA has expansive WUI communities, historically devastated by some of the deadliest wildfires, where significant disparities are observed in post-wildfire resource allocation [10]. Conversely, AZ's inadequate fire-suppressing infrastructures, lack of disaster preparedness, and CO's disproportionate post-wildfire financial allocation posed challenges to efficient wildfire management.

Although several studies focused on wildfire prediction and susceptibility to forest fires using remote sensing, geospatial risk analysis, and machine learning [11, 12], not too many studies focused on the challenges to post-wildfire recovery and mitigation strategies. For example, Auer et al. and [13] addressed the risk of wildfire associated with expanding WUI communities, emphasizing the need for poly-centric governance for wildfire emergency management across jurisdictions and stakeholders. Studies by Kolden et al. [14] suggest community partnerships and pre-fire mitigation activities are crucial in contingency planning. However, most of the studies overlooked the two significant aspects of post-wildfire resource management—financial allocation and recovery efforts. Moreover, the previous studies did not consider the interactions between the socio-demographic factors and the resource allocation/mitigation efforts. Finally, no state-level comparative assessment of post-wildfire resource management exists to understand how the inequities vary across different jurisdictions, political environments, and past experiences.

Therefore, this study aims to investigate and compare/contrast inequities in post-wildfire resource allocations across the three states to generate insights for comprehensive resource allocation and recovery strategy to advance equitable wildfire management leveraging a data-driven interpretable machine learning modeling approach.

2. Methodology

In this study, we propose a novel data-driven framework to investigate inequity in post-wildfire resource allocation. Figure 1 presents the approach of this study, including data collection, exploratory data analysis, development of the model and inference analysis, details of which are discussed in the following subsections.

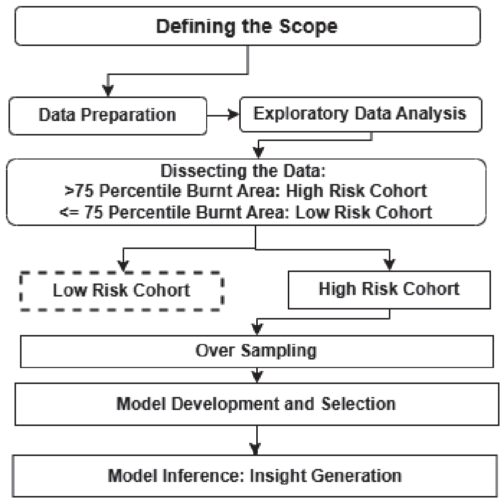

Fig. 1.: Study Approach

2.1. *Data Preparation*

2.1.1. *Data Collection*

We collected data on wildfire incidents, their impacts, and socio-demographic information.

SIT209 Dataset – This dataset comprises information regarding wildfire incidents, their location, spread, burnt area, impacts on structures, and fatalities. The data contains 31 relevant features for 504 incidents in CA, 128 in AZ, 104 in CO [3].

US Census Bureau- From this data source, we collected 26 socio-demographic, economic, and household-related factors like total population, educational attainment, elderly population, disabled population, race-ethnicity, poverty, crowded households, no vehicle households, etc. [15].

National Interagency Fire Center (NIFC) data - Information on personnel deployed in rescue operations and recovery costs are obtained from this data source [16].

2.1.2. *Data Pre-processing*

Various steps used in data pre-processing are discussed below.

Data Cleaning: After consolidating the dataset,

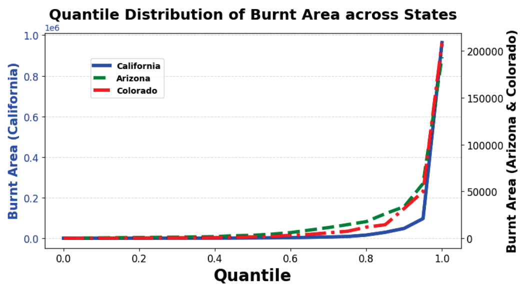

Fig. 2.: Quantile Distribution of *Burnt Area* for California, Arizona and Colorado

different data cleaning techniques were used. For imputing missing values, we used statistical imputing methods like linear interpolation, moving average, and machine learning algorithms[a].

Determining Response Variables and Features: Variables related to post-wildfire resource allocation—*Cost for Recovery* and *Resource Personnel for Rescue Operations*—are considered as the response variables. Since our response variables are available at county-level, we conducted out study as county-level. The final dataset comprises a total of 63 predictors depicting socio-demographics and impacts of wildfires from 2015-2022. We adopted the correlation screening method [17] and combined relevant variables to reduce the data dimensionality.

2.1.3. ***Exploratory Data Analysis***

The quantile distributions of the wildfire impact variable, e.g., burnt area (see Figure 2) show a sharp steep slope in the the third quantile (≥ 75 percentile), indicating a heavy flat-tailed distribution. This trend highlights the underlying different characteristics of wildfire damage and consequent resource allocation. Thus, this study focuses on the extreme wildfire incidents (falling in the fourth quantile), hereafter referred to as the "High-risk Cohort"; containing a total of 182 observations in different states (CA: 123; AZ: 37; CO: 22).

2.1.4. ***Oversampling Method:***

As we had limited samples across AZ and CO, we have used oversampling technique like Synthetic Minority Over-sampling Technique for Regression. This interpolates data points in between the existing data points to generate synthetic data for the model. We have generated 100 samples across high-risk cohorts of AZ and CO [18].

2.1.5. ***Modeling Approach***

Overview of Statistical Learning

Statistical learning emerged due to a multidisciplinary approach to problem-solving using mathematics, statistics, and computer science [19]. Supervised statistical learning aims to determine the relation between the variable of interest (response variable) y and n-dimensional vector x by determining function f, which we can write as: $y = f(x) + \epsilon$, where ϵ is irreducible error. When there is only one response variable y, the technique is named univariate supervised learning. In contrast, the method is called multivariate learning if there is more than one response variable(y). In this study, we implemented both univariate and multivariate supervised modeling techniques to investigate if the covariance between the two different response variables representing resource allocation contributes to the overall accuracy of the population health inequity assessment and the prediction models [20]. We trained a library of univariate and multivariate models—linear regression, decision tree, random forest, gradient boosting, AdaBoost, XGBoost, and Elastic Net. We would illustrate in the result section that the model gradient boost imparts statistically significant performance improvement compared to other models for both response variables for the high-risk cohorts of California and Arizona (except Colorado, where XGBoost has comparable results for *Cost for Recovery* variable). For comprehensive performance on this dataset, we have chosen gradient boost for insight generation.

Gradient Boosting: Gradient Boosting is a fundamental boosting [b] algorithm, starting with an initial prediction, which is the mean of the target followed by building the tree iteratively, computing the gradients in each round, followed by training the decision tree on these gradients, up-

[a]In ML-based imputation, all variables except the missing are treated as independent variables to predict the missing value

[b]multiple sequential weak learner trees to combine into a strong learner with higher predictive accuracy [21]

dating the model by adding the new tree's predictions(adjusted by learning rate). The final model would be the additive combination of all the trees [22, 23].

Multivariate and univariate regression models are developed after implementing oversampling methods like SMOTER and resampling techniques like bootstrapping. For Arizona and Colorado, we used this technique to create the model, which optimized the model performance by 32% and 60 % compared to the smaller original sample size. We selected the best model using bias-variance trade-off and cross-validation techniques like leave-one-out cross-validation (LOOCV)[c]. All the models were compared with the mean-only model, which is used a baseline for statistical analysis [20].

3. Results

3.1. *Inequities in Resource Allocation*

We investigated each state's resource allocation for high-risk cohorts to identify the key factors indicating underlying inequities.

3.1.1. *California*

We analyzed the high-risk cohort using univariate and multivariate predictive analyses to understand the associations between resource variables and socioeconomic factors. Our study shows that the multivariate gradient boosting algorithm outperforms other models for both the *Resource Personnel for Rescue Operations* (see Figure 3) and *Cost for Recovery* (see Figure 4).

Key predictors of resource allocation: Variables like crowded households and educational attainment are critical socio-economic factors associated with *resource personnel for rescue operations*. Similarly, economic and educational status are found to be correlated with *cost for recovery*.

Resource Personnel for Rescue Operations: The partial dependence plots (PDPs) (see Figure 5) show that counties with higher populations living in *crowded households* and with lower *educational attainment* receive lower *resource personnel for rescue operations*. This trend (with minor fluctuations) highlights the inequities in post-wildfire resource allocation where marginal communities (living in crowded households or educationally backward) receive fewer rescue personnel than their wealthier counterparts.

Cost for Recovery: The PDPs (Fig. 6) show that counties having higher percentages of *below poverty level* population and with *lower educational attainment* observe lower *cost for recovery*. This illustrates that the economically and educationally disadvantaged communities, lacking financial resources, experience inadequate resources in the aftermath of severe wildfire incidents.

3.1.2. *Arizona*

From Figs. 7 and 8, we observe that multivariate gradient boost outperforms other models of *resource personnel for rescue operations* and *cost for recovery* respectively and selected for inference analysis.

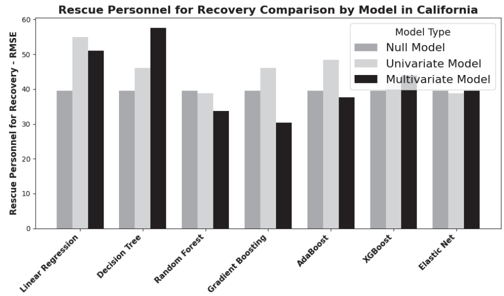

Fig. 3.: *Resource personnel for Rescue Operation* in CA: Model performance comparisons

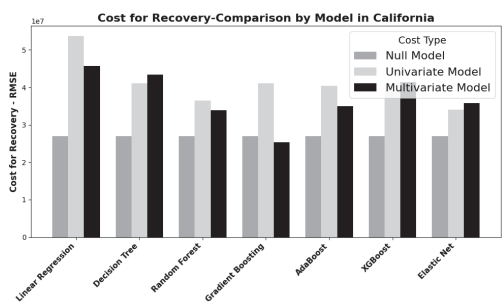

Fig. 4.: *Cost for Recovery* in CA: Model performance comparisons

[c] for smaller sample size one data point for validation and rest for training the dataset

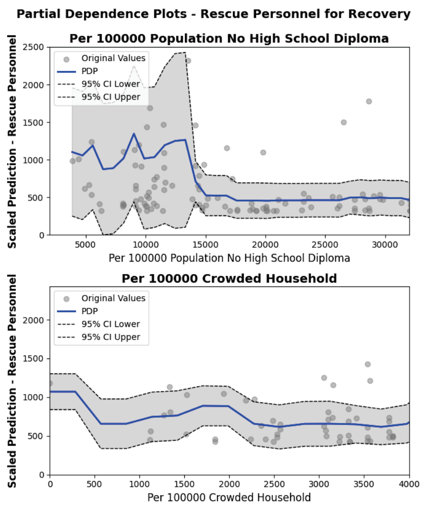

Fig. 5.: Relationship of key predictors with *resource personnel for rescue operations* in CA (the grey lines indicates 95% confidence interval and the dots show the concentration of data points for all the PDP plots)

Key predictors of resource allocation: Racial composition and educational attainment are found to be the key predictors of resource allocations in the aftermath of severe wildfire events in AZ.
Resource personnel for rescue operations: Fig. 9)

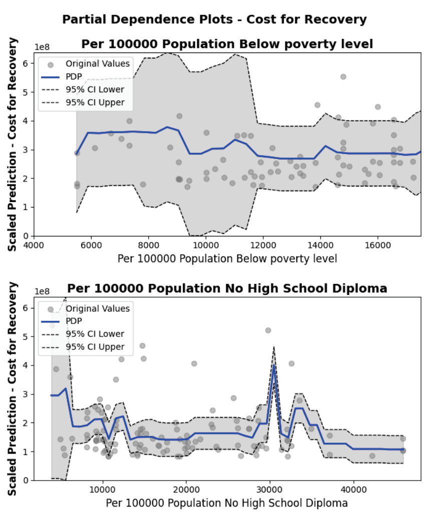

Fig. 6.: Relationship of key predictors with *cost for recovery* in CA

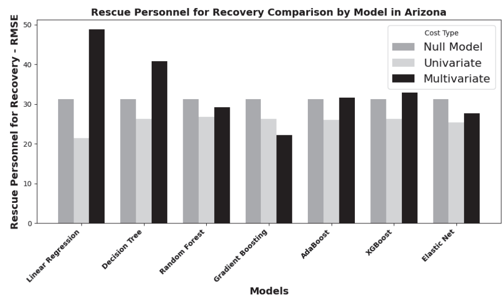

Fig. 7.: *Resource personnel for rescue operations* in AZ: Model performance comparison

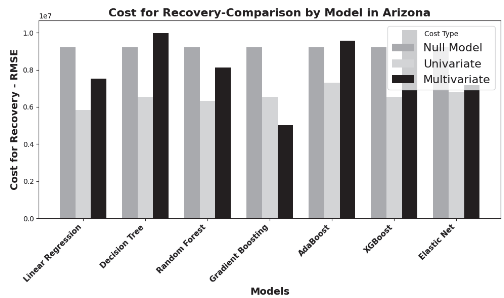

Fig. 8.: *Cost for Recovery* in AZ: Model performance comparison

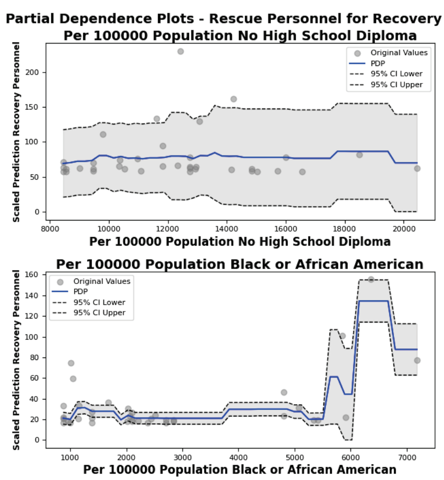

Fig. 9.: Relationship of key predictors with *resource personnel for rescue operations* in AZ

shows that in AZ, counties with higher percentages of *population with lower educational attainment* and *Black or African American population*

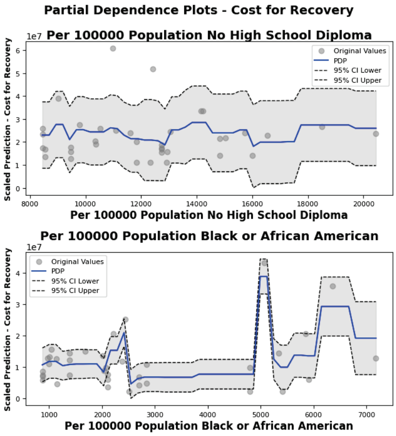

Fig. 10.: Relationship of key predictors with *cost for recovery* in AZ

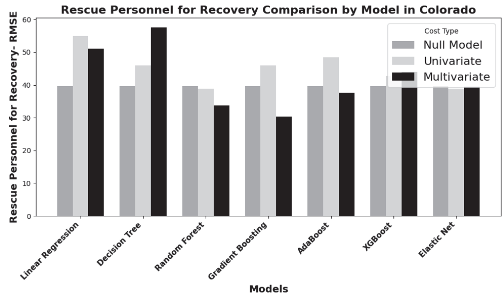

Fig. 11.: *Resource personnel for rescue operations* in CO: Model performance comparison

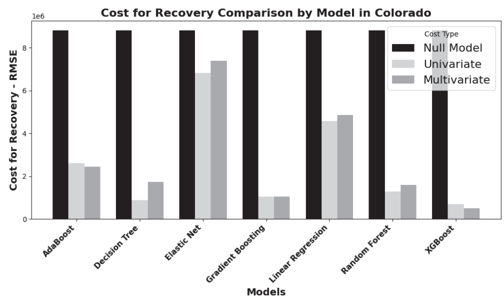

Fig. 12.: *Cost for Recovery* in CO: Model performance comparison

receive a higher number of resources personnel, indicating the prevalence of adequate recovery efforts. *Cost for recovery:* As observed from Fig. 10, significant uncertainty persists in the relationship of the key predictor, *percentage of population with no high school diploma* and the *cost for recovery*. The ambiguity and randomness in the trend fail to conclude a distinct relationship. However, we see a slight positive trend in the PDP of *Black or African American population* with the *Cost for recovery* that again reemphasizes our previous finding that AZ provides adequate resources to its marginalized communities in the aftermath of a severe wildfire. Arizona's hazard mitigation policies ensure equity in resource allocation by developing targeted recovery plans for marginal, tribal, and educationally disadvantaged communities [24].

3.1.3. *Colorado*

Figs. 11 and 12 show that the multivariate gradient boosting outperforms all other models for both variables. Though, for the variable, *Cost for Recovery*, XGboost also shows a slightly improved outcome than gradient boosting, the difference is not statistically significant. Therefore, gradient boost has been chosen for generating insights into its robust performance across all the variables.

Key predictors of resource allocation: The post-wildfire resource allocation in CO strongly correlates with the racial composition, percentage of disabled population, and household variables like living arrangements.
Resource personnel for rescue operations: From Fig. 13, both critical variables, i.e., *Per 100000 Population Black or African American*, *Per 100,000 Population with No High School Diploma* show an ambiguous trend for this response variable pertaining to uncertainty failing to show any distinguished patter. However, the overall trend inclines to a positive association.

Cost for Recovery: Fig. 14 shows that the driving variables, like the proportion of Black or African American Population, show a steep declining trend with cost allocation while the multiple household apartment shows an overall positive correlation with randomness due to uncertainty.

3.2. *Comparative Inequity Analysis across the US States*

Our study identified that counties with higher population of less educated people are at a high risk of

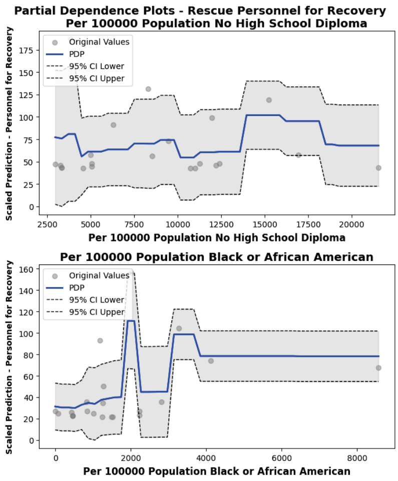

Fig. 13.: Relationship of key predictors with *Resource personnel for rescue operations* in CO

receiving less post wildfire resources in terms of both rescue personnel and funding in CA, whereas counties with similar demographic composition in AZ and CO are at a higher risk of receiving less recovery funding, but overall received adequate rescue personnel. Similarly, counties with higher population of Black or African American origin people are at a higher risk of receiving less resources for post wildfire recovery in terms of both rescue personnel and funding in CA, where such

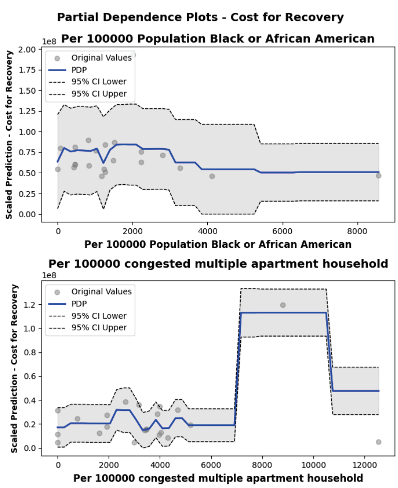

Fig. 14.: Relationship of key predictors with *Cost for Recovery* in CO

counties in CO are only at a higher risk of receiving less funding for recovery. On the contrary, in AZ such counties with higher population of Black or African American origin people is observed to receive adequate recovery resources. Counties with higher population of crowded households in CA are at a higher risk of receiving less rescue personnel, whereas this variable was not found to be an important predictor for AZ and CO. Overall, we found that the inequities in wildfire resource allocation is more prominent in CA compared to AZ and CO. One of the underlying reasons might be the sprawling of communities into the WUI regions in CA due to urban housing prices, exposing a higher population to wildfire risk in CA compared to the other states.

4. Conclusion

The study analyzes the inequities present in wildfire resource management across the southwestern US states—CA, AZ, and CO. Our data-driven analysis highlights the stark disparity in rescue resource allocation in the aftermath of devastating wildfires. With the increasing intensity and frequency of wildfires, the authorities should develop a policy framework that is not solely reliant on economic priorities but also includes social equity in the decision making as marginalized communities often need more attention to cope with the disasters. Our study addresses the need for targeted and equitable post-wildfire resource allocation to streamline data-informed decision-making.

Acknowledgment

The research is supported by the National Science Foundation under award number 2324616.

References

1. Nima Masoudvaziri, Prasangsha Ganguly, Sayanti Mukherjee, and Kang Sun. Impact of geophysical and anthropogenic factors on wildfire size: A spatiotemporal data-driven risk assessment approach using statistical learning. *Stochastic Environmental Research and Risk Assessment*, pages 1103–1129, 2021.
2. Prasangsha Ganguly and Sayanti Mukherjee. Understanding wildfire induced risk on in-

terconnected infrastructure systems using a bow-tie model with bayesian network and self organizing maps. In *31st European Safety and Reliability Conference*, number doi: 10.3850/978-981-18-2016-8_567, 2021.

3. Dung Nguyen, Erin J Belval, Yu Wei, Karen C Short, and David E Calkin. Dataset of united states incident management situation reports from 2007 to 2021. *Scientific data*, 11(1):23, 2024.
4. Land Life Company. Post-wildfire land restoration in colorado, 2024. Accessed: 2025-01-12.
5. Stephen D Wong, Jacquelyn C Broader, and Susan A Shaheen. Can sharing economy platforms increase social equity for vulnerable populations in disaster response and relief? a case study of the 2017 and 2018 california wildfires. *Transportation research interdisciplinary perspectives*, 5:100131, 2020.
6. Imogen James and Ian Youngs and Steven McIntosh. Mel gibson, jeff bridges and other la celebrities lose homes to wildfires, 2025. Accessed: 2025-01-12.
7. Times of India. Los angeles wildfire live updates: Hundreds of thousands without power as wildfires tear across california, 2025. Accessed: 2025-01-12.
8. Tracey Lou Dodrill and Charlynne Herman. Wallow fire, 2011. Poster presented at Earth Camp for Educators 2011, Mountainside Middle School, Scottsdale, AZ.
9. Destination Granby. East troublesome fire info & recovery, 2020. Accessed: 2025-01-14.
10. Sayanti Mukherjee Poulomee Roy. The inequity of wildfire rescue resources in california, December 2024. Accessed: 2025-01-14.
11. Younes Oulad Sayad, Hajar Mousannif, and Hassan Al Moatassime. Predictive modeling of wildfires: A new dataset and machine learning approach. *Fire safety journal*, 104:130–146, 2019.
12. Soumik Saha, Biswajit Bera, Pravat Kumar Shit, Sumana Bhattacharjee, and Nairita Sengupta. Prediction of forest fire susceptibility applying machine and deep learning algorithms for conservation priorities of forest resources. *Remote Sensing Applications: Society and Environment*, 29:100917, 2023.
13. Matthew R Auer. Considering equity in wildfire protection. *Sustainability science*, 16(6):2163–2169, 2021.
14. Crystal A Kolden and Carol Henson. A socio-ecological approach to mitigating wildfire vulnerability in the wildland urban interface: A case study from the 2017 thomas fire. *Fire*, 2(1):9, 2019.
15. census.gov. census.gov, 2024. Accessed: 2024-07-30.
16. National Interagency Fire Center. Fire information statistics, 2024. Accessed: 2024-07-30.
17. Sayanti Mukherjee and Roshanak Nateghi. Climate sensitivity of end-use electricity consumption in the built environment: an application to the state of florida, united states. *Energy*, 128:688–700, 2017.
18. Luís Torgo, Rita P Ribeiro, Bernhard Pfahringer, and Paula Branco. Smote for regression. In *Portuguese conference on artificial intelligence*, pages 378–389. Springer, 2013.
19. Ronald Aylmer Fisher et al. Statistical inference and analysis: Selected correspondence of ra fisher, edited by jh bennett, 1990.
20. Zhiyuan Wei, Adil Baran Narin, and Sayanti Mukherjee. Multidimensional population health modeling: A data-driven multivariate statistical learning approach. *IEEE Access*, 10:22737–22755, 2022.
21. Robert E Schapire et al. A brief introduction to boosting. In *Ijcai*, volume 99, pages 1401–1406. Citeseer, 1999.
22. Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.
23. Leo Breiman. Population theory for boosting ensembles. *The Annals of Statistics*, 32(1):1–11, 2004.
24. Arizona Department of Emergency and Military Affairs. Hazard mitigation planning. 2025. Accessed: 2025-02-23.