Beware the Rise of Models When They Are Wrong: A Look at Heat Vulnerability Modeling Through the Lens of Sensitivity

Seda Şalap-Ayça[1]
Department of Earth, Environmental, and Planetary Sciences, Brown University, Providence, RI, USA
Institute at Brown for Environment and Society, Brown University, Providence, RI, USA
Department of Earth, Geographic, and Climate Sciences, University of Massachusetts, Amherst, MA, USA

Erica Akemi Goto[1]
Arizona Institute for Resilience, University of Arizona, Tucson, AZ, USA

Abstract

Extreme heat affects communities across the globe and is likely to increase as the climate changes; however, its consequences are not uniform. Geographically weighted regression is a useful modeling effort to understand the spatial linkage between various factors to heat-related casualty and supports decision-making in the spatial context. Still, as every complex spatial modeling approach, it is also bounded by uncertainty. Understanding model uncertainty and how this uncertainty is related to model input can be revealed by sensitivity analysis. In this study, we applied a spatial global sensitivity analysis to assess the model dynamics to address which input factors need to be prioritized in decision-making. A visual representation of the model's sensitivity and the spatial pattern of factor influence is an important step toward establishing a robust confidence mechanism for understanding heat vulnerability and supporting policy-making.

2012 ACM Subject Classification Information systems → Geographic information systems

Keywords and phrases heat vulnerability, uncertainty, sensitivity analysis

Digital Object Identifier 10.4230/LIPIcs.GIScience.2023.64

Category Short Paper

Acknowledgements Seda Şalap-Ayça wants to thank Aykut Ayça for his discussion on selecting probability distribution functions.

1 Introduction and Background

Extreme heat causes injuries and fatalities in many regions of the U.S. Southwest region, especially during the summer months [12] [13]. Furthermore, extreme weather events like heat waves will likely increase with climate change [21]. However, studies have found that communities are not impacted the same by these extreme events [2] [9] as some communities are more vulnerable than others [9] [10].

There is a combination of factors that influence heat vulnerability, such as social (e.g., age and isolation[2][14][19]), economic (e.g., income and poverty[15]), health (e.g., pre-existing or chronic health conditions [15]), and environmental (e.g., lack of tree canopies or temperature [15]) factors. Scholars used these factors and applied various methods to measure heat vulnerability. Some of these methods were composite where the contributing factors are combined into an index [4][11][24] or regression which explains the relationship between

[1] corresponding author

independent and dependent variables [3],[16], or both [8]. Regression analysis draws a more reliable picture in terms of variables' influence on heat vulnerability compared to composite methods; however, the traditional regression methods do not address the spatial configuration of the factors, which can be solved by employing geographically weighted regression.

Despite the effort to address heat vulnerability, none of these methods are immune to uncertainty; as the presence and importance of uncertainty in spatial data is not new in data collection or GIS. Moreover, how the uncertainty is intertwined with vulnerability representation and its disproportionate impact on marginalized populations is an important dimension that is not addressed enough[6]. In the realm of policy decision-making models, when we contemplate the renowned aphorism of George Box, "all models are wrong but some are useful" in conjunction with Franklin's [6] observation "poor data often disadvantages the disadvantaged"' understanding uncertainty becomes more crucial in informed decisions and resource allocation. Therefore, in this study, we are interested in identifying how uncertainty in heat vulnerability related factors influences the prediction of health casualty. This research aims to advance the field of vulnerability to natural hazards and GIS by employing geographic weighted regression analysis coupled with sensitivity analysis.

2 Methodology

2.1 Data Acquisition and Variable Reduction

Our analysis area was the U.S. Southwest region, including Arizona, New Mexico, Texas, and Oklahoma states. The unit of analysis was chosen as the county level due to the availability of the data.

Our dataset was selected based on a thorough literature review of previous studies in the field and data availability. Our study combined social, economic, health, and environmental data as independent variables, county population (population 2015) and number of heat events as control variables, and casualty (fatalities and injuries) as the dependent variable. We used social and economic data from the 2015 American Community Survey 5-Year (ACS5), health data from both the ACS5 (2015) and the Global Health Exchange Data (GHDx) from 2014 and 2015, and environmental data from the National Oceanic and Atmospheric Administration (NOAA) from 2016 to 2020 and the Multi-Resolution Land Characteristics Consortium (MRLC) from 2016. Mortality and injury data were obtained from SHELDUS from 2016 to 2020. The initial dataset included 24 social, economic, health, and environmental variables.

We first tested the correlation between independent and dependent variables to reduce the number of independent variables. We then dropped independent variables whose relationship with the dependent variable was not statistically significant ($p - values > 0.05$). We also conducted a correlation matrix including all remaining independent variables and dropped one of the independent variables with a high correlation (> 0.7). Finally, we removed variables that had high spatial correlation. From our initial 24 independent variables, we ended up with 9 independent variables, which are elderly population, disabled population, black population, population with no car, unemployed population, number of months with a temperature higher than 38°C, and impervious surface and two control variables.

2.2 Geographically Weighted Regression for Heat Vulnerability

To understand the spatial relation between independent variables to dependent variables, we applied geographically weighted regression (GWR). GWR has been widely used over a decade to model the potential spatially varying relationships [1][5][23]. The model outcome

y_i(health-related casualty) can be expressed by

$$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i)x_{ik} + \epsilon_i \tag{1}$$

where y_i represents the dependent variable, k are independent variables, β is the parameter to be estimated, ϵ is the error term. (u_i, v_i) denotes the coordinates of the i^{th} feature and k(u_i, v_i) is a realization of the continuous function k(u, v) at feature i.

2.3 Global Sensitivity Analysis for Heat Vulnerability GWR

Global Sensitivity analysis (GSA) is a forward looking approach to modeling to understand the linear (individual) and nonlinear (interaction) relationship between input variables and the output of the model [18]. In this study, since we are focusing on GWR's sensitivity, our focus is on which independent variable(s) influence most the prediction of causality. GSA starts with generating random samples which are used to replicate the behavior of the input set when the model is run multiple times. These sample sets mimic the original probability distribution function (pdf) of the input variables. Therefore, we conducted a systematic analysis to acquire the pdf of each variable before generating the samples. Due to limited information about some factors' priori distribution, GSA is only conducted for 7 variables (disability, elderly, impervious surface, black population, population with no car, unemployed, 2015 county population), whereas all 9 are included in the GWR. Once the pdfs are determined, the following framework is applied:

1. For each parameter, we generated 2048 random variables. This number is based on the experimental example set N (2^7) and number of model inputs $D(7)$, which yields $N(2D + 2) = 2048$ samples.
2. Prepare the sample set for GWR input using python based pandas library [22]
3. Run GWR model 2048 times on county scale with randomly generated sample set of independent variables
4. Exporting GWR output for GSA
5. Implementing GSA using SAlib package in python [20] [7]

3 Results and Discussion

The result of the GSA has been visualized as the most influential variable for the individual (Figure 1) and interaction effects' (Figure 2) influences in terms of the model's explanatory power for each county. These maps represent variables contributing the most to the model's variability or uncertainty. For each input variable that is fed into GSA, the analysis produces a unique GSA map. In order to reduce the visual load of the GSA output (7 individual and 7 interaction effects map), self-organizing map-based exploratory analysis [17] has been used where the neural networks evaluate the similarities among indices per future and results in clusters where patterns are dominant. While Figure 1 shows us where individual variances of each variable affect the heat-related casualty uncertainty, Figure 2 depicts the interaction effect influence on the model uncertainty. For example, when we look at Arizona State, GWR output is most sensitive to any small variation in no car variable (observed in 3 counties) when each independent variable is singly treated (Figure 3). However, as the spatial complex nature of these variables plays an important role in GWR prediction, we can see an increase in disability and elderly variables when the interaction among parameters is considered. This means individual effects will not be enough to see the whole picture when we try to understand model dependencies. Also, as we can see, the influential variables vary among the four states. When heat vulnerability modeling efforts are in action, each state might prioritize its resources depending on how these variables are distributed.

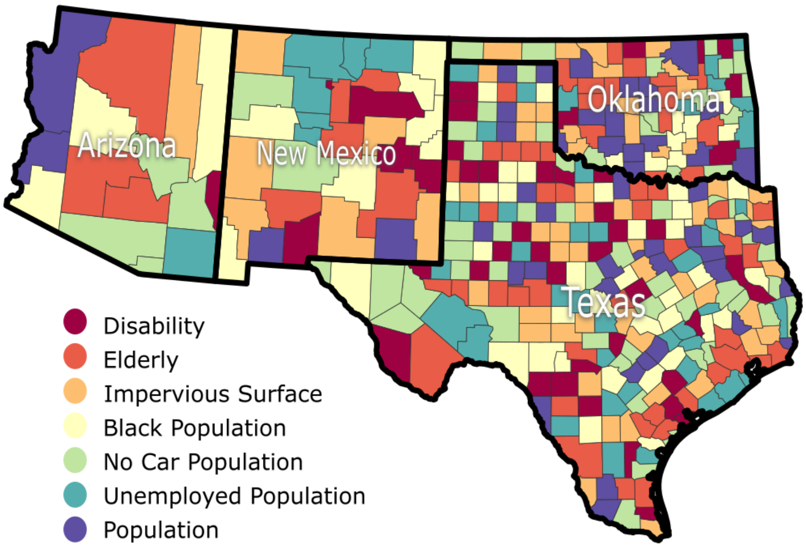

Figure 1 Most influential factors to the model uncertainty based on Individual Effects of Input Variables to Predicted Heat-related Casualty.

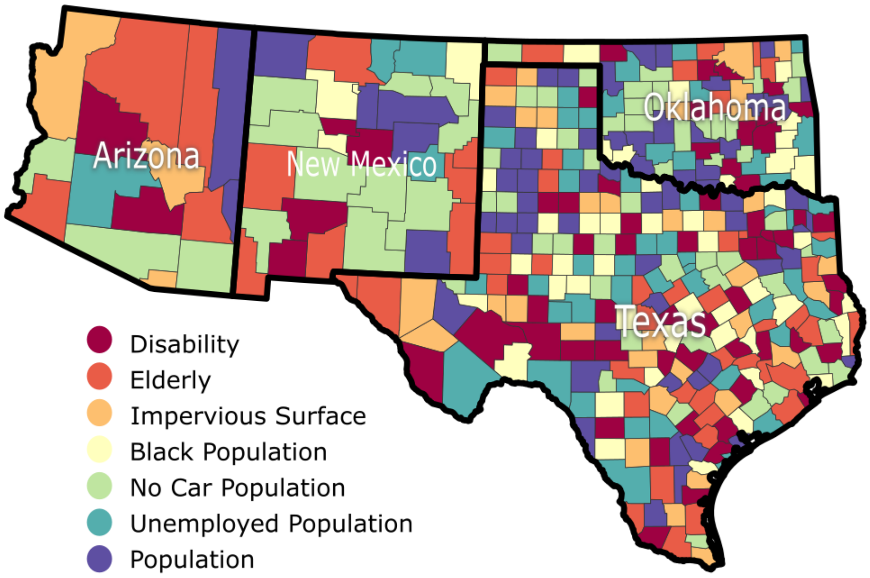

Figure 2 Most influential factors to the model uncertainty based on Interaction Effects of Input Variables to Predicted Heat-related Casualty.

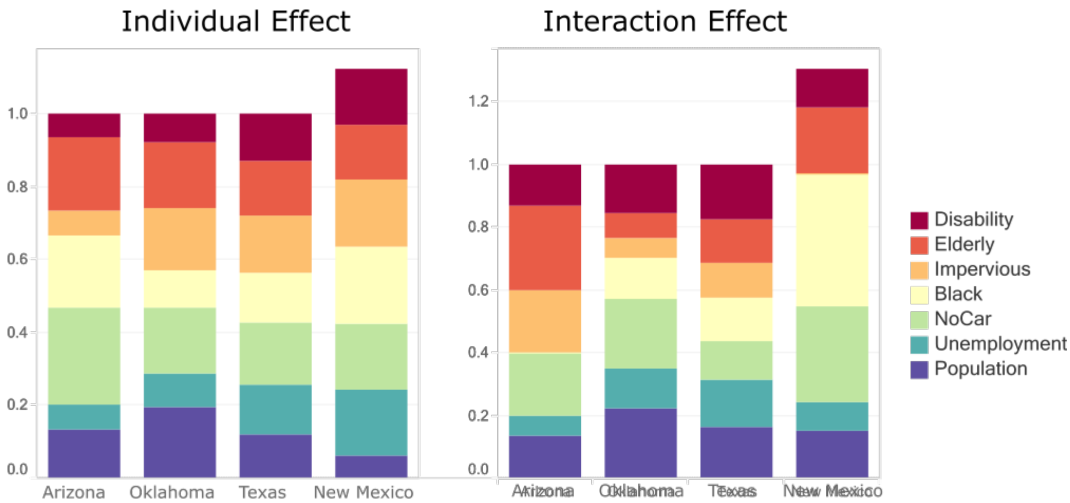

Figure 3 Frequency distribution of individual effect dominant factors per state.

4 Conclusion and Future Work

The geographic focus is crucial for equitable risk planning, resilience strategies, and response to heat risk. Moreover, it can be used to communicate results and support decision-makers. Data acquisition is the most time and effort-consuming part of the spatial decision-making process; but crucial as the interaction of variables produces different results. Considering the unavoidable uncertainty, it is important to know the models' weaknesses and strengths and the spatial variability of the results so that the resource allocation can be optimum. Moreover, heat vulnerabilities indicated by dominant factors depicted in Figures 1 and 2, can help decision makers and modelers to prioritizing resources. This effort will help us identify the influential variables and where they cluster as an initial step and can be followed by the involvement and insight of the communities which need to be a part of the solution.

References

1 Chris Brunsdon, Stewart Fotheringham, and Martin Charlton. Geographically weighted regression-modelling spatial non-stationarity. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 47(3):431–443, 1998. URL: `http://www.jstor.org/stable/2988625`.

2 Susan L Cutter, Bryan J Boruff, and WL Shirly. Social science quarterly. *Soc. Vulnerability Environ. Hazards*, 84:242–261, 2003.

3 Yaella Depietri, Torsten Welle, and Fabrice G. Renaud. Social vulnerability assessment of the cologne urban area (germany) to heat waves: links to ecosystem services. *International Journal of Disaster Risk Reduction*, 6:98–117, 2013.

4 Abbas El-Zein and Fahim N Tonmoy. Assessment of vulnerability to climate change using a multi-criteria outranking approach with application to heat stress in sydney. *Ecological Indicators*, 48:207–2017, 2015.

5 A. Fotheringham, Chris Brunsdon, and Martin Charlton. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Wiley, 2002.

6 Rachel Franklin. Quantitative methods i: Reckoning with uncertainty. *Progress in Human Geography*, 46(2):689–697, 2022.

7 Jon Herman and Will Usher. SALib: An open-source python library for sensitivity analysis. *The Journal of Open Source Software*, 2(9), January 2017.

8 Daniel P. Johnson, Austin Stanforth, Vijay Lulla, and George Luber. Developing an applied extreme heat vulnerability index utilizing socioeconomic and environmental data. *Applied Geography*, 35(1):23–31, 2012.

9 Shirley Laska and Betty Hearn Morrow. Social vulnerabilities and hurricane katrina: an unnatural disaster in new orleans. *Marine technology society journal*, 40(4), 2006.

10 K. Lieberknecht, D. Zoll, and K. Castles. Hurricane harvey: equal opportunity storm or disparate disaster? *Local Environment*, 2(26), 2021.

11 Francisco de la Barrera Luis Inostroza, Massimo Palme. A heat vulnerability index: Spatial patterns of exposure, sensitivity and adaptive capacity for santiago de chile. *PLOS ONE*, 2016.

12 Alex Nguyen and Erin Douglas. Texas heat-related deaths reached a two-decade high in 2022 amid extreme temperatures. *CHRON*, 2012.

13 AZHS Arizona Department of Health Services. Extreme weather & public health. URL: `https://www.azdhs.gov/preparedness/epidemiology-disease-control/extreme-weather/heat-safety/index.php#heat-home`.

14 D. Reckien. What is in an index? construction method, data metric, and weighting scheme determine the outcome of composite social vulnerability indices in new york city. *Regional Environmental Change*, 18:1439–1451, 2018.

15 C. Rinner, D. Patychuk, K. Bassil, S. Nasr, S. Gower, and M. Campbell. The role of maps in neighborhood-level heat vulnerability assessment for the city of toronto. *Cartography and Geographic Information Science*, 1(37), 2010.

16 Samain Sabrin, Maryam Karimi, Md Golam Rabbani Fahad, and Rouzbeh Nazari. Quantifying environmental and social vulnerability: Role of urban heat island and air quality, a case study of camden, nj. *Urban Climate*, 34, 2020.

17 Seda Şalap-Ayça. Self-organizing maps as a dimension reduction approach for spatial global sensitivity analysis visualization. *Transactions in GIS*, 26(4):1718–1734, 2022.

18 Seda Şalap-Ayça, Piotr Jankowski, Keith C Clarke, Phaedon C Kyriakidis, and Atsushi Nara. A meta-modeling approach for spatio-temporal uncertainty and sensitivity analysis: an application for a cellular automata-based urban growth and land-use change model. *International Journal of Geographical Information Science*, 32(4):637–662, 2018.

19 S. Sheridan T. Dolney. The relationship between extreme heat and ambulance response calls for the city of toronto, ontario, canada. *Environmental Research*, 100:94–103, 2006.

20 Iwanaga Takuya, William Usher, and Jonathan Herman. Toward SALib 2.0: Advancing the accessibility and interpretability of global sensitivity analyses. *Socio-Environmental Systems Modelling*, 4:18155, May 2022. `doi:10.18174/sesmo.18155`.

21 Michael Wehner, Sonia Seneviratne, Xuebin Zhang, Muhammad Adnan, Wafae Badi, Claudine Dereczynski, Alejandro Di Luca, Subimal Ghosh, Iskhaq Iskandar, James Kossin, et al. Weather and climate extreme events in a changing climate. In *AGU Fall Meeting Abstracts*, volume 2021, pages U13B–11, 2021.

22 Wes McKinney. Data Structures for Statistical Computing in Python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 56–61, 2010.

23 David C. Wheeler and Antonio Páez. Geographically weighted regression. In Manfred M. Fischer and Arthur Getis, editors, *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications*, pages 461–486. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

24 Tanja Wolf and Glenn McGregor. The development of a heat wave vulnerability index for london, united kingdom. *Weather and Climate Extremes*, 1:59–68, 2013.