**Title**
Urbanization, its Exposure to Environmental Hazards, and Equity in Adaptation Programs



**Author**
Ju, Yang

**Publication Date**
2019

Peer reviewed|Thesis/dissertation

# Urbanization, its Exposure to Environmental Hazards, and Equity in Adaptation Programs

By

Yang Ju

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy
in
Landscape Architecture and Environmental Planning
in the
Graduate Division
of the
University of California, Berkeley

Committee in charge:

Professor John D. Radke, Chair
Professor Iryna Dronova
Professor Rachel Morello-Frosch

Summer 2019

Abstract

Urbanization, its Exposure to Environmental Hazards, and Equity in Adaptation Programs

by

Yang Ju

Doctor of Philosophy in Landscape Architecture and Environmental Planning

University of California, Berkeley

Professor John D. Radke, Chair

Urban areas are increasingly exposed to weather-related environmental hazards under climate change, and equity concerns exist in adaptation and mitigation programs for these exposures and their consequences. To address these issues above, we used data-driven approaches to investigate three unique cases. We first developed a novel, cost-effective, and spatially-explicit classification of urbanization trends over time, using time-series nighttime light remote sensing images and unsupervised classification, in mainland China between 1992 and 2013. We identified five temporal typologies of urbanization, namely stable urban activity, high-level steady growth, acceleration, low-level steady growth, and fluctuation. Our classification characterizes distinct urbanization patterns over time and can be applied to environmentally sensitive and hazard-prone areas where monitoring of urbanization is critical. Next, we perform a multitemporal and multi-scenario projection of exposure to flooding, caused by sea level rise and storm surge, in the highly urbanized San Francisco Bay Area. We found increased uncertainty in exposure over time and in scenarios with higher greenhouse gas concentrations. Such elevated uncertainty suggests that stakeholders should employ adaptation strategies that are no-regret, reversible, and flexible, and that regulators may explicitly require a long-term planning horizon for adaptation programs and new developments. Finally, we investigated the allocation of a widely-adopted adaptation and mitigation program, clean vehicle rebates, from two major policy programs in California. We evaluated rebate allocation rates with respect to community characteristics including socioeconomic and environmental disadvantages, household income, racial and ethnical composition, and ambient air pollution. We found that when rebate assignment and amount were only based on vehicle technology and did not consider the varied socioeconomic backgrounds of potential applicants, rebate allocation rates were higher in advantaged, wealthier communities, and communities with intermediate levels of nitrogen dioxide concentration, but the rates were lower in communities with higher percentages of Hispanics and Non-Hispanic Blacks. After introducing an income cap, expanded vehicle eligibility, and income- and geography-tiered rebate amounts, rebate allocation rates increased in communities with lower-household income, higher percentages of Hispanics, and slightly higher nitrogen dioxide concentration. The findings of this study implied the need for and effectiveness of equity-related policy designs to spread the benefits of adaptation and mitigation programs to more diverse populations. In all, these studies seek to engage a broader conversation about environmental and societal challenges in urban

areas under climate change, as well as how data-driven approaches can reveal the underlying processes of these challenges and inform better policy and decision-making.

# Acknowledgements

I would like to first thank my advisor, Prof. John Radke, for his guidance and support during my Ph.D. and for this dissertation. Prof. Radke has always impressed me with his passion about the environment and society, his determination to make real changes to the world, his engagement with stakeholders, his support to students, and his dedication to teaching. We have worked together on two major research projects, which builds the backbone of this dissertation and my Ph.D. I believe what I have learned from him will guide me profoundly in my career and life.

I am also grateful to have two other committee members, Prof. Rachel Morello-Frosch and Prof. Iryna Dronova, whom I collaborate with on two studies in my dissertation. I appreciate the opportunity to integrate both professors' expertise into my dissertation, which in turn makes my Ph.D. and future career more interdisciplinary.

Next, I want to acknowledge other professors, colleagues, and friends that I have met at different stages of my Ph.D. study. I would like to thank Prof. Lara Cushing, Prof. Greg Bigging, and Prof. Peng Gong, who have also guided my research. I am fortunate to work with colleagues including Martine Schmidt-Poolman, Howard Forster, Vladimir Ulyashin, Emery Roe, Sarah Lindbergh, Yiyi He, Peter Norton, Liam Maier, Matthew Ashenfarb, Michelle Wray, Amna Alruheil, Tessa Beach, and Sophie Taddeo. I also want to thank other fellow Ph.D. students and post-docs including Ruoyin Xu, Xize Wang, Amir Gohar, and Qi Yuan, for the time we talk about Ph.D. life and having meals together.

Family has always been my strongest support. My wife, Qin Ma, who is also a researcher herself, has been my partner for both life and research. My father, Weimin Ju, who also happens to be a professor, and mother Bo Yang, have always reminded me to progress in my research and to take care of my non-research life. Although I am not by my parents' side and serve my duty as a son, I always feel they have covered my back.

Finally, my Ph.D. would not be possible without generous funding supports. Research in the dissertation is funded by California Energy Commission, California Office of Environmental Health Hazard Assessment, and Wellcome Trust. I am also grateful for China Scholarship Council's financial support that compensates for the expensive living cost in the San Francisco Bay Area.

Table of Contents

List of Figures

List of Tables

# Chapter 1. Introduction

Urban areas are the home to most of the world's population, development, and economic activities, which are exposed to weather-related environmental hazards including flooding, heat waves, and drought. Such exposure is likely to increase due to climate change that elevates the extent, frequency, and severity of those hazards. At the same time, socioeconomically disadvantaged populations likely undergo more exposure to these hazards and have limited capacity to cope with and recover from them. Adaptation and mitigation programs have been established to cope with anticipated impacts of environmental hazards. However, if a program is 'one size fits all' or does not consider the varied socioeconomic status of its potential participants, the program may disproportionally favor advantaged participants. Equity-related policy designs should and often can allocate the program's benefits to more diverse populations.

Research challenges remain in: (1) cost-effective monitoring of urbanization over time with fine resolution data, so that planners and decision-makers can assess the growing environmental footprint of urban areas and measure its potential exposure to environmental hazards; (2) projecting urban exposure to weather-related environmental hazards under distinct climate scenarios over the long term to facilitate long-term planning, and understanding the implications for public and private-sector stakeholders with respect to uncertainty and long-term projection; and (3) evaluating how specific policy design elements can improve equity of policy programs, which may often be 'one size fits all' and likely favor socioeconomically advantaged population groups. This dissertation seeks to address these challenges. The dissertation also discuss how data-driven approach improve our understandings about the underlying processes of these challenges and these understandings can inform decision-making.

## 1.1 Urbanization and weather-related environmental hazards

Urbanization, a process where population, economic activities, and land use shift from rural to urban areas, is an essential component of global environmental change (Buhaug and Urdal 2013; Qian Zhang and Seto 2013).The proportion of the global population living in urban areas has increased from 30% (~0.7 billon) in 1950 to 54% (~3.9 billon) in 2014, and is estimated to reach 66% (~6.3 billon) in 2050 (United Nations, 2014). To some extent, while urbanization facilitates economic growth and can improve quality of life (Turok & McGranahan, 2013), yet this process is also associated with undesired environmental consequences including pollution, degraded ecosystem functions, increased exposure to environmental hazards (e.g. flooding, wildfire, and heatwaves), and climate change (Doygun & Alphan, 2006; Hollis, 1975; Kalnay & Cai, 2003; Wu, 2008).

Weather-related environmental hazards including flooding, heat waves, and drought affect urban areas (Guerreiro, Dawson, Kilsby, Lewis, & Ford, 2018; Hunt & Watkiss, 2010). Exposure to these hazards is because of: (1) urbanization often originates and re-iterates in hazard-prone regions (e.g. coastal areas and flood plains); (2) the concentration of people, economic activities, and physical infrastructures; (3) environmental degradation often accompanies urbanization; and (4) urban systems being less self-sufficient than rural areas, as urban areas are reliant on rural areas for input of resources (Pelling, 2003). Many cities originate in flood plains and rebuild themselves in the same areas after major floods, to take advantage of water-related benefits including low-cost transportation, provision of natural resources, and amenities (De Sherbinin, Schiller, & Pulsipher, 2007; McDermott, Michaels, & Rauch, 2015; UNU-IHDP, 2015). Heat waves in urban areas are exacerbated by urban heat islands, a

phenomenon where urban areas have higher temperatures than adjacent rural areas (Ziter, Pedersen, Kucharik, & Turner, 2019). Urban heat islands result from: (1) the prevalence of building materials that absorb solar radiation during the day and re-radiation at night (Rizwan, Dennis, & Liu, 2008; Yamaguchi & Ihara, 2020); (2) lack of greenspaces providing evapotranspiration and shading, which lead to cooler temperature than the surrounding built areas (Li & Zhou, 2019; Ziter et al., 2019); and (3) presence of anthropogenic heat sources including industrial land use, air conditioning, and vehicle exhaust (Yue, Qiu, Xu, Xu, & Zhang, 2019). Drought affects urban areas by disrupting transportation routes, reducing water and power supplies, which can adversely affect local economies and increase poverty (Güneralp, Güneralp, & Liu, 2015; Rodriguez-Oreggia, Fuente, Torre, & Moreno, 2013).

## 1.2 Increased urban exposure to environmental hazards under climate change

Urban areas are more likely be exposed to weather-related environmental hazards in the future, due to both anthropogenic climate change and urban expansion. Climate change increases the spatial extent, temporal frequency, and intensity of present-day weather-related environmental hazards. A European-wide study of 571 cities found that the likelihood of increased exposure to heat waves, flooding and drought would be elevated for nearly all cities, under climate scenarios characterized by high greenhouse gas (GHG) emissions (Representative Concentration Pathway (RCP) 8.5) (Guerreiro et al., 2018). At the same time, urban areas are projected to increase in currently hazard-prone areas: from 2000 to 2030, global urban land in current high-frequency flood zones has been projected to increase from 195,493 km2 (30% of global urban land) to 723,167 km2 (39%), and in current drylands will increase from 173,200 km2 (27%) to 497,825 km2 (27%) (Güneralp et al., 2015).

Researchers and governments have been working to characterize and project urban areas' increased exposure to weather-related environmental hazards under different climate change scenarios. Studies have evaluated climate change's impact on urban areas with respect to different populations (Kaźmierczak & Cavan, 2011; Nutters, 2012; Martinich, Neumann, Ludwig, & Jantarasami, 2013; KC, Shepherd, & Gaither, 2015; Bickers, 2014), physical infrastructure (Biging, Radke, & Lee, 2012; Demirel, Kompil, & Nemry, 2015; Radke et al., 2018, 2017), emergency service (Lang, Radke, Chen, & Chan, 2016), and the natural environment (Schile et al., 2014; Zhu, Xi, Hoctor, & Volk, 2015). Cities have been updating their master, strategic, and action plans to adapt to climate change and mitigate the impact of weather-related environmental hazards (Jabareen, 2015). Governments have also established various policy programs to reduce greenhouse gas (GHG) and air pollutants emissions, by promoting transit-oriented and mixed land use, and to aid socioeconomically and environmentally disadvantaged populations. For example, through the auspices of the Global Warming Solutions Act (AB 32), the California's Air Resource Board implements programs to reduce GHG emission, improve fuel quality, and promote clean vehicle adoption (California Air Resource Board, 2019).

## 1.3 Equity in climate change mitigation and adaptation programs

Equity concerns exist in exposure to environmental hazards, and these concerns are likely amplified and reinforced under climate change and in adaptation and mitigation programs (Department of Economic and Social Affairs, 2016; Reckien et al., 2017; Winsemius et al.,

2018). Studies show that socioeconomically disadvantaged populations are generally more exposed to weather-related environmental hazards including heat waves (Klinenberg, 2015), drought (Winsemius et al., 2018), and flooding (Kaźmierczak & Cavan, 2011; Martinich et al., 2013). Such disproportional exposure pattern is due to that disadvantaged populations are more likely to reside in high risk or marginalized areas with lower-quality infrastructure, and limited access to social capital and mobilities in anticipation of environmental hazardous events (Pelling, 2003).

In addition, marginalized populations may be less likely to benefit from mitigation and adaptation programs, particularly when program designs are "one size fits all". The 'one size fits all' design of many such programs does not consider the varied socioeconomic status of their potential participants, therefore the programs may not adequately serve vulnerable groups, or can disproportionally favor advantaged participants. Equity-related policy designs can ensure that sustainability, resilience and equity goals are met by ensuring a program's benefits are allocated to more diverse populations. For example, the first iteration of a California-wide Clean Vehicle Rebate Project assigns a rebate amount based only on vehicle technology, without additionally considering potential applicants' socioeconomic characteristics such as income. One study found that higher income communities and communities with fewer Hispanic and non-Hispanic African-American populations had higher rebate allocation rates from this Project (Rubin & St-Louis, 2016). It is both regional and global consensus that equity should be an inherent concern in climate change adaptation (California Environmental Protection Agency, 2017; Thomas & Twyman, 2005). Specific policy design elements can and should be applied to integrate sustainability and equity goals of these programs and promote program benefits to more diverse populations.

## 1.4 Objectives of this dissertation

This dissertation engages increased urban exposure to weather-related environmental hazards under climate change, and equity challenges in adaptation and mitigation programs. Accordingly, we conduct the following three studies using a data-driven and model-based empirical approach.

In study 1 we undertake a novel, cost-effect, and reproducible classification of temporal typologies of urbanization in recent history and interpret these typologies with respect to economic changes and other relevant transitions. Understanding these temporal typologies is critical for informing land use planning in the face of growing urban population and increasing environmental footprint of expanding urban regions. Focusing on the rapidly urbanizing mainland China, where urban expansion and shrinkage coexist, we classify its large number of nighttime light time-series (n=384,849) between 1992-2013 as an example into a small set of typologies (n=5), each representing a distinct urbanization trend over time. Nighttime light data is retrieved from the DMSP-OLS nighttime light dataset (National Geophysical Data Center, 2015). We employ K-Means, a widely-adopted unsupervised classification algorithm, to group the time-series based on similarity in their main descriptive statistical parameters. The resulting groups represent stable urban activity, high-level steady growth, acceleration, low-level steady growth, and fluctuation. We specially analyse cities dominated by acceleration and find several spatial clusters of this typology at prefecture city and county levels. The formation of these clusters is linked to their underlying socioeconomic characteristics and developmental history. Using China as an example, this research provides an innovative approach for characterizing distinct patterns of urbanization over time in our recent history, particularly in environmentally

sensitive and hazard-prone areas where monitoring of urban expansion is critical to avoid adverse environmental consequences and potential losses to environmental hazards (Du, He, Huang, & Shi, 2018).

In study 2 we project urban exposure to coastal flooding under various climate change scenarios and discuss the implications for stakeholders. In the San Francisco Bay Area, USA, we model this region's exposure of population, land uses, infrastructure (e.g. roads and utilities), and emergency responders (e.g. fire stations and hospitals) to worst-case-scenario flooding (as a combined effect of storm surge, sea level rise, and tides) projected under 24 distinct climate scenarios over five 20-year periods between 2000 and 2100. Uncertainty in the projected exposures are analysed and communicated with some infrastructure stakeholders. Our analysis shows substantial uncertainty in scenarios with higher greenhouse gas atmospheric concentrations and/or over a longer time horizon. We observe an uncertainty-avoidance behaviour where stakeholders prefer short-term and more probable results, that are suboptimal for long-term land use and infrastructure planning. This implies the need for coastal urban areas to cope with climate-related uncertainties and to focus on the long term when developing planning strategies and policies for climate change mitigation and adaptation.

In study 3 we undertake an equity analysis of clean vehicle rebate programs in California. To improve air quality and mitigate climate change by reducing greenhouse gas emission, these programs are typically funded by governments to incentivize purchase of lower-emission vehicles. However, these programs may disproportionally allocate rebates to wealthier households, since upfront capital is often required to acquire a vehicle. If these programs are "one size fits all" and do not consider the potential socioeconomic barriers to low emission vehicle purchases, the programs will likely further exclude lower income households. Using two clean vehicle rebate programs in California with varied emphasis on equity, we examine their rebate allocation rates with respect to community characteristics, including disadvantage, household income, racial and ethnic composition, and concentrations of air pollution. Our findings indicate that when rebate assignment and amount is only based on vehicle technology, rebates are more allocated to those that are advantaged, the wealthier communities with intermediate levels of nitrogen dioxide ($NO_2$) concentration. We find rebates are less allocated to those with higher percentages of Hispanics and Non-Hispanic Blacks. Introducing an income cap, expands vehicle eligibility, and income- and geography- tiered rebate amounts, direct rebates to communities with lower-household income, higher percentages of Hispanics, and slightly higher $NO_2$ pollution. Our findings imply the need and effectiveness of equity-related designs to spread benefits of policy programs to more diverse populations.

Taken together, these studies demonstrate how we can use data-driven approaches to: (1) measure urbanization, a process with significant environmental consequences and therefore requires close monitoring; (2) project long-term impact of coastal flooding under different climate scenarios with substantial uncertainty, which in term posts challenges for urban development, mitigation, and adaptation; and (3) demonstrate how we can assess and fine tune climate change adaptation and mitigation programs as they unfold, so that the program benefits are accessible by diverse groups of population, particularly those are marginalized and disadvantaged.

In conclusion, given rapid global trends in urbanization, as well as future environmental and societal challenges under climate change, solutions to a sustainable and socially-just society lie in our cities. The studies in this dissertation seek to inform the broader conversation about these challenges and solutions. This dissertation also propose that the use of scientific data and

models can help us better understand the underlying process of the challenges and solutions above, and form the basis for more informed decision-making.

**References**


Bickers, K. M. (2014). *Vulnerable Populations to Climate Change in New Jersey*. Edward J. Bloustein School of Planning and Public Policy Rutgers, The State University of New Jersey Rutgers University.

Biging, G., Radke, J., & Lee, J. H. (2012). *Impacts of predicted sea-level rise and extreme storm events on the transportation in the San Francisco Bay Region* (No. CEC - 500 - 2012 - 040). California Energy Commission.

California Air Resource Board. (2019). Climate Change Programs. Retrieved June 2, 2019, from https://ww2.arb.ca.gov/our-work/programs/climate-change-programs

California Environmental Protection Agency. (2017). *Designation of Disadvantaged Communities Pursuant to Senate Bill 525 (de León)*. Retrieved from https://calepa.ca.gov/wp-content/uploads/sites/6/2017/04/SB-535-Designation-Final.pdf

De Sherbinin, A., Schiller, A., & Pulsipher, A. (2007). The vulnerability of global cities to climate hazards. *Environment and Urbanization*, *19*(1), 39–64. https://doi.org/10.1177/0956247807076725

Demirel, H., Kompil, M., & Nemry, F. (2015). A framework to analyze the vulnerability of European road networks due to Sea-Level Rise (SLR) and sea storm surges. *Transportation Research Part A: Policy and Practice*, *81*, 62–76. https://doi.org/10.1016/j.tra.2015.05.002

Department of Economic and Social Affairs. (2016). *World Economic and Social Survey (WESS) 2016: Climate Change Resilience: An Opportunity for Reducing Inequalities*. Retrieved from United Nations website: https://www.un.org/development/desa/publications/wess-2016.html

Doygun, H., & Alphan, H. (2006). Monitoring Urbanization of Iskenderun, Turkey, and its Negative Implications. *Environmental Monitoring and Assessment*, *114*(1–3), 145–155. https://doi.org/10.1007/s10661-006-2524-0

Du, S., He, C., Huang, Q., & Shi, P. (2018). How did the urban land in floodplains distribute and expand in China from 1992–2015? *Environmental Research Letters*. https://doi.org/10.1088/1748-9326/aaac07

Guerreiro, S. B., Dawson, R. J., Kilsby, C., Lewis, E., & Ford, A. (2018). Future heat-waves, droughts and floods in 571 European cities. *Environmental Research Letters*, *13*(3), 034009. https://doi.org/10.1088/1748-9326/aaaad3

Güneralp, B., Güneralp, İ., & Liu, Y. (2015). Changing global patterns of urban exposure to flood and drought hazards. *Global Environmental Change*, *31*, 217–225. https://doi.org/10.1016/j.gloenvcha.2015.01.002

Hollis, G. E. (1975). The effect of urbanization on floods of different recurrence interval. *Water Resources Research*, *11*(3), 431–435. https://doi.org/10.1029/WR011i003p00431

Hunt, A., & Watkiss, P. (2010). Climate change impacts and adaptation in cities: A review of the literature. *Climatic Change*, *104*(1), 13–49. https://doi.org/10.1007/s10584-010-9975-6

Jabareen, Y. (2015). City planning deficiencies & climate change – The situation in developed and developing cities. *Geoforum*, *63*, 40–43. https://doi.org/10.1016/j.geoforum.2015.05.017

Kalnay, E., & Cai, M. (2003). Impact of urbanization and land-use change on climate. *Nature*, *423*(6939), 528–531. https://doi.org/10.1038/nature01675

Kaźmierczak, A., & Cavan, G. (2011). Surface water flooding risk to urban communities: Analysis of vulnerability, hazard and exposure. *Landscape and Urban Planning*, *103*(2), 185–197. https://doi.org/10.1016/j.landurbplan.2011.07.008

KC, B., Shepherd, J. M., & Gaither, C. J. (2015). Climate change vulnerability assessment in Georgia. *Applied Geography*, *62*, 62–74. https://doi.org/10.1016/j.apgeog.2015.04.007

Klinenberg, E. (2015). *Heat wave: A social autopsy of disaster in Chicago*. University of Chicago Press.

Lang, W., Radke, J. D., Chen, T., & Chan, E. H. W. (2016). Will affordability policy transcend climate change? A new lens to re-examine equitable access to healthcare in the San Francisco Bay Area. *Cities*, *58*, 124–136. https://doi.org/10.1016/j.cities.2016.05.014

Li, X., & Zhou, W. (2019). Optimizing urban greenspace spatial pattern to mitigate urban heat island effects: Extending understanding from local to the city scale. *Urban Forestry & Urban Greening*. https://doi.org/10.1016/j.ufug.2019.04.008

Martinich, J., Neumann, J., Ludwig, L., & Jantarasami, L. (2013). Risks of sea level rise to disadvantaged communities in the United States. *Mitigation and Adaptation Strategies for Global Change*, *18*(2), 169–185. https://doi.org/10.1007/s11027-011-9356-0

McDermott, T. K., Michaels, G., & Rauch, F. (2015). *CEP Discussion Paper No 1398 December 2015 Flooded Cities Adriana Kocornik-Mina*.

National Geophysical Data Center. (2015, December 10). Version 4 DMSP-OLS Nighttime Lights Time Series. Retrieved December 10, 2015, from http://ngdc.noaa.gov/eog/gcv4_readme.txt

Nutters, H. (2012). *Addressing social vulnerability and equity in climate change adaptation planning*. San Francisco Bay Conservation and Development Commission.

Pelling, M. (2003). *The vulnerability of cities. [Electronic resource]: Natural disasters and social resilience.* Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=cat04202a&AN=ucb.b13633897&site=eds-live

Radke, J., Biging, G., Roberts, K., Foster, H., Roe, E., Ju, Y., … Dalal, A. (2018). *Assessing Extreme Weather-Related Vulnerability and Identifying Resilience Options for California's Interdependent Transportation Fuel Sector* (No. CCCA4-CEC-2018–012). Retrieved from California Energy Commission website: http://www.climateassessment.ca.gov/techreports/docs/20180827-Energy_CCCA4-CEC-2018-012.pdf

Radke, J., Biging, G., Schimidt-Poolman, M., Foster, H., Roe, E., Ju, Y., … Reeves, I. (2017). *Assessment of Bay Area Gas Pipeline Vulnerability to Climate Change* (No. CEC-500-2017-008). Retrieved from California Energy Commission website: https://www.energy.ca.gov/2017publications/CEC-500-2017-008/CEC-500-2017-008.pdf

Reckien, D., Creutzig, F., Fernandez, B., Lwasa, S., Tovar-Restrepo, M., Mcevoy, D., & Satterthwaite, D. (2017). Climate change, equity and the Sustainable Development Goals: An urban perspective. *Environment and Urbanization*, *29*(1), 159–182. https://doi.org/10.1177/0956247816677778

Rizwan, A. M., Dennis, L. Y. C., & Liu, C. (2008). A review on the generation, determination and mitigation of Urban Heat Island. *Journal of Environmental Sciences*, *20*(1), 120–128. https://doi.org/10.1016/S1001-0742(08)60019-4

Rodriguez-Oreggia, E., Fuente, A. D. L., Torre, R. D. L., & Moreno, H. A. (2013). Natural Disasters, Human Development and Poverty at the Municipal Level in Mexico. *The*

*Journal of Development Studies*, *49*(3), 442–455. https://doi.org/10.1080/00220388.2012.700398

Rubin, D., & St-Louis, E. (2016). Evaluating the Economic and Social Implications of Participation in Clean Vehicle Rebate Programs. *Transportation Research Record: Journal of the Transportation Research Board*, *2598*, 67–74. https://doi.org/10.3141/2598-08

Schile, L. M., Callaway, J. C., Morris, J. T., Stralberg, D., Parker, V. T., & Kelly, M. (2014). Modeling Tidal Marsh Distribution with Sea-Level Rise: Evaluating the Role of Vegetation, Sediment, and Upland Habitat in Marsh Resiliency. *PLoS ONE*, *9*(2), e88760. https://doi.org/10.1371/journal.pone.0088760

Thomas, D. S. G., & Twyman, C. (2005). Equity and justice in climate change adaptation amongst natural-resource-dependent societies. *Global Environmental Change*, *15*(2), 115–124. https://doi.org/10.1016/j.gloenvcha.2004.10.001

Turok, I., & McGranahan, G. (2013). Urbanization and economic growth: The arguments and evidence for Africa and Asia. *Environment and Urbanization*, *25*(2), 465–482. https://doi.org/10.1177/0956247813490908

United Nations. (2014). World urbanization prospects: The 2014 revision. *United Nations, Department of Economic and Social Affairs (DESA), Population Division, New York*.

UNU-IHDP. (2015). *Coastal Zones and Urbanization. Summary for Decision-Makers.* Bonn: UNU-IHDP.

Winsemius, H. C., Jongman, B., Veldkamp, T. I. E., Hallegatte, S., Bangalore, M., & Ward, P. J. (2018). Disaster risk, climate change, and poverty: Assessing the global exposure of poor people to floods and droughts. *Environment and Development Economics*, *23*(3), 328–348. https://doi.org/10.1017/S1355770X17000444

Wu, J. (2008). Toward a Landscape Ecology of Cities: Beyond Buildings, Trees, and Urban Forests. In M. M. Carreiro, Y.-C. Song, & J. Wu (Eds.), *Ecology, Planning, and Management of Urban Forests* (pp. 10–28). https://doi.org/10.1007/978-0-387-71425-7_2

Yamaguchi, K., & Ihara, T. (2020). Countermeasures to Urban Heat Island Considering Urban Energy Usage. In N. Enteria, H. Awbi, & M. Santamouris (Eds.), *Building in Hot and Humid Regions: Historical Perspective and Technological Advances* (pp. 15–57). https://doi.org/10.1007/978-981-13-7519-4_2

Yue, W., Qiu, S., Xu, H., Xu, L., & Zhang, L. (2019). Polycentric urban development and urban thermal environment: A case of Hangzhou, China. *Landscape and Urban Planning*, *189*, 58–70. https://doi.org/10.1016/j.landurbplan.2019.04.008

Zhu, M., Xi, X., Hoctor, T. S., & Volk, M. (2015). Integrating conservation costs into sea level rise adaptive conservation prioritization. *Global Ecology and Conservation*, *4*, 48–62. https://doi.org/10.1016/j.gecco.2015.05.007

Ziter, C. D., Pedersen, E. J., Kucharik, C. J., & Turner, M. G. (2019). Scale-dependent interactions between tree canopy cover and impervious surfaces reduce daytime urban heat during summer. *Proceedings of the National Academy of Sciences*, *116*(15), 7575–7580. https://doi.org/10.1073/pnas.1817561116

# Chapter 2. Analysis of urbanization dynamics in mainland China using pixel-based nighttime light trajectories from 1992 to 2013[1]

**Abstract**: Understanding urbanization dynamics, or how intensity of urbanization changes over time, is an important basis for urban planning and management, which has been investigated using various data-driven approaches. Considering the advantages and constraints of different data sources, we use pixel-based, time-series nighttime light (NTL) trajectories to characterize urbanization dynamics in mainland China where massive urban development has been occurring in recent decades. After pre-processing the data, we extracted time-series NTL trajectories for each 1 km by 1 km pixel between 1992 and 2013 and used the unsupervised *k*-means classification to identify the major typologies of these trajectories as urbanization dynamics based on their main statistical parameters. The classification identified five urbanization dynamics, namely, stable urban activity, high-level steady growth, acceleration, low-level steady growth, and fluctuation. Their distributions and spatial patterns were further summarized and compared among different Chinese administrative divisions. We specifically analysed the acceleration trajectories that showed rapid transitions from rural to urban, as we considered these trajectories as potential indicators for aggressive urbanization. We found several clusters at prefecture city and county levels with high proportion of the acceleration and referred to the underlying socioeconomic characteristics and developmental history to understand how these clusters could had been formed. Through this study, we revealed the dominant tendencies of urbanization in China over space and time and developed an analysis framework that could be extended to other regions.
**Keywords**: urbanization dynamics, nighttime light, time-series trajectory, unsupervised classification, spatial pattern

## 2.1 Introduction

Urbanization, a process where population and economy transform from rural to urban and land cover changes from natural to predominantly built-up (Buhaug & Urdal, 2013; Qian Zhang & Seto, 2013), continues to be an important component of global environmental change. The United Nations' 2014 projection showed that 30% of the world's population was urban in 1950, and this number would increase to 66% by 2050 (United Nations, 2014). Over the next several decades, developing countries and countries in Asian and Africa are expected to be the hotspots for population growth and urbanization (United Nations, 2014; L. Jiang & O'Neill, 2015). With its significant land cover change and high concentration of population and development, urbanization could cause unfavourable impacts, including pollution (Doygun & Alphan, 2006; Schetke & Haase, 2008), increased flooding (Hollis, 1975; Kalnay & Cai, 2003), changes in climate (Kalnay & Cai, 2003), and degraded ecosystem function (Jianguo Wu, 2008; Y. Li, Zhu, Sun, & Wang, 2010). Well-informed planning and management should be applied to

[1] This chapter is from the journal paper: Ju, Y., Dronova, I., Ma, Q., & Zhang, X. (2017). Analysis of urbanization dynamics in mainland China using pixel-based night-time light trajectories from 1992 to 2013. International Journal of Remote Sensing, 38(21), 6047–6072. https://doi.org/10.1080/01431161.2017.1302114

urban areas to mitigate these consequences. We argue that to inform the decisions for the future, an important step is to understand different forms of urbanization dynamics, or how intensity of urbanization changes over time (Qingling Zhang & Seto, 2011), in the recent history. Planners and decision-makers can correlate the dynamics with socioeconomic factors to find urbanization drivers and areas undergoing non-typical transitions and make subsequent planning and management interventions.

China is playing an important role in global urbanization due to its rapid urban development and large urban population. The country has seen a massive transition from rural to urban since its economic reform in 1978. At the national level, China's fraction of urban population increased from 17.9% to 52.6% between 1978 and 2012 (Bai, Shi, & Liu, 2014), and its urban built-up area expanded from 7438 $km^2$ to 45566 $km^2$ between 1981 and 2012 (M. Chen, Liu, & Lu, 2015). Even in its present, advanced urbanization stage, China was still expected to add 292 million in urban population between 2014 and 2050, making the country one of the three major sources for the world's urban population growth (United Nations, 2014). At the local level, urbanization rate varies spatially and shows a gradient from the east to the west, where the east, particularly the coastal regions, are more urbanized (Bai et al., 2014). Because of such variation and different underlying socioeconomic drivers, China is expected to manifest different types of urbanization dynamics across the country.

A particular phenomenon, ghost city, has attracted the attention from researchers, decision makers, and the public (M. Chen et al., 2015; Woodworth, 2015). Characterized by fast development beyond the actual population needs and business speculation, ghost city potentially leads to several issues including financial risk and social injustice (M. Chen et al., 2015; X.-R. Wang, Hui, Choguill, & Jia, 2015). To our knowledge, there is no study to date focusing on measuring the urbanization dynamics during the formation of a ghost city. However, since such a city tends to be under fast development within a limited period, we assume it is likely to be dominated by urbanization dynamics showing rural to urban transformations with a rapid and accelerating speed. Ghost city, along with many other urbanization phenomena, need to be better understood by researchers and decision makers to make any necessary interventions to the current urbanization process. Studying urbanization dynamics provides a basis to understand those phenomena and to inform decision making.

To study urbanization dynamics and urban systems, researchers have relied on different types of information, particularly (1) socioeconomic data on population and economy, (2) remote sensing-derived data such as land cover and nighttime lights, and (3) user generated, location-based social network data such as Twitter and Flickr Photos. Socioeconomic data directly reflect urbanization as population and economic change and are often available in time-series forms. However, the main limitation is that such data are aggregated into coarse administrative divisions, such as city, county, and province, which eliminates the finer-scale heterogeneity within the divisions. Urban land cover data interpreted from remote sensing images preserve spatial heterogeneity in the form of urbanized pixels at finer spatial resolutions, such as 30 meter for a global coverage (Gong et al., 2013). With such data from multiple years, it is possible to analyze how individual cities expand spatially (Weng, 2002; Xiao et al., 2006; Jat, Garg, & Khare, 2008; Z. Liu, He, Zhang, Huang, & Yang, 2012) and how urbanization unfolds at the national level (L. Wang et al., 2012; Sleeter et al., 2013). However, as binary categories of "urban" and

"non-urban", land cover does not reflect how population and economy change within urbanized areas. Finally, location-based social network data are generated by human activities and thus are often used to locate people's position (Chi, Liu, Wu, & Wu, 2015) and activity centres (Hu et al., 2015), and to delineate urban boundaries (B. Jiang & Miao, 2015). However, such data can be biased as they only represent certain population groups using the social media (Chi et al., 2015; Hu et al., 2015). Furthermore, social network data have a fine temporal resolution due to their near real-time collection, but only cover a limited time span due to their short history, reducing the possibility to establish a long-term history (e.g. decadal) of urbanization dynamics.

DMSP-OLS nighttime light (NTL) (National Geophysical Data Center, 2015), collected by the U.S. Air Force Defence Meteorological Satellite Programs (DMSP) Operational Linescan System (OLS), offer the research community a new perspective to understand long-term urbanization dynamics. These data have a relatively long time span, an annual temporal resolution, a nearly global coverage, a moderate spatial resolution, and a capability to reflect the intensity of human activities, as demonstrated by various studies (Z. Chen et al., 2015; Elvidge et al., 2001; Ghosh et al., 2010; T. Ma, Zhou, Pei, Haynie, & Fan, 2012; Shi, Yu, et al., 2014). In mapping urbanization, researchers have mainly applied thresholds on NTL intensity to identify urban extent (Z. Liu et al., 2012; Q. Ma et al., 2014; Shi, Huang, et al., 2014; Tan, 2016; Yu et al., 2014; Zhou et al., 2014, 2015). In addition, researchers have recently started to extract time-series NTL trajectories to characterize urbanization dynamics at administrative division level and pixel level. At administrative division level, Stathakis, Tselios, & Faraslis (2015) calculated the Sum of Lights (SoL) index for European regions, and used linear regression to calculate overall and decadal slope of the SoL trend to represent urbanization process in each region. At pixel level, Qingling Zhang & Seto (2011) proposed five archetypes of urbanization dynamics, extracted pixel-based, time-series NTL trajectories, and classified those trajectories based on the archetypes. Jiansheng Wu, Ma, Li, Peng, & Liu (2014) classified the time-series NTL trajectories in China to calculate the composition of the classes for each administrative division, grouped the divisions based the compositions, and finally compared the average composition of each group with a baseline composition to evaluate China's urbanization.

Although these studies (Jiansheng Wu et al., 2014; Qingling Zhang & Seto, 2011) established an important foundation for NTL-based urbanization analysis, they have not yet extensively discussed the detected trajectories nor explicitly related them to socioeconomic data and underlying urbanization context. Furthermore, the number of classes in the unsupervised classifications was determined mainly through a subjective process, which might not produce an optimal setting to maximize inter-class differences and improve intra-class similarities. Therefore, our study had two primary objectives. First, we aimed to use the dynamics to investigate urbanization in mainland China. We mapped the areas under different urbanization dynamics, calculated the proportion of each type of dynamics in different administrative divisions, and quantified spatial patterns of these proportions to reveal the underlying drivers. In particular, we focused on the dynamics showing accelerated growth in NTL between 1992 and 2013, as it might be an indicator for rapid and even aggressive urbanization, which could lead to ghost cities in some cases. Second, we aimed to develop a framework to classify urbanization dynamics at pixel level using the time-series NTL trajectories. Based on previous

methods by Qingling Zhang & Seto (2011) and Jiansheng Wu et al. (2014), our classification extracted a set of eight trajectory parameters to conduct a *k*-means unsupervised classification, and used a silhouette analysis (Rousseeuw, 1987) as the basis to determine the number of classes. In all, we hypothesized that administrative divisions with high proportion of accelerated growth tended to be clustered, and the formation of the clusters could be explained by their unique socioeconomic characteristics and developmental history.

## 2.2 Materials and methods

### 2.2.1 Nighttime light data and data pre-processing

We used the stable light composite from the version 4 DMSP-OLS NTL time-series in this study. Version 4 DMSP-OLS NTL time-series provides three annual products: cloud-free composite, average visible light composite, and stable light composite, from 1992 to 2013 (National Geophysical Data Center, 2015). DMSP-OLS sensors collect a set of NTL images of the Earth with 30 arc second (approximately 1 km) spatial resolution twice a day (Elvidge et al., 2009), and the images together cover -180° to 180° longitude and -65° to 75° latitude (National Geophysical Data Center, 2015). Each pixel has a digital number (DN) for NTL intensity. An annual composite is made from the selected, highest quality images that exclude sunlit data, glare, moonlit data, observations with clouds, and lighting features from the aurora (National Geophysical Data Center, 2015). We used the stable light composite, as it provided persistent lights from cities and towns, and excluded ephemeral events, such as fire and explosion. In this composite, the background noise is replaced with 0, and the DNs range from 1 to 63 (National Geophysical Data Center, 2015). This composite provides 34 annual images from six satellite missions (i.e. F10, F12, F14, F15, F16, and F18). Some years may have two images as two satellite missions were operating at the same times. For such years, we used the average of the two images to represent the NTL.

There are three major issues with NTL data, namely, the "overglow" effect, the saturation in urban centres, and the inconsistency between satellite missions (X. Li & Zhou, 2017). The "overglow" effect results from the radiation of non-coherent light in all directions from its source (Townsend & Bruce, 2010), and causes an overestimation of urban extent (T. Ma et al., 2012). Both the "overglow" removal model (ORM) by Townsend & Bruce (2010) and the thresholding method (Amaral, Monteiro, Camara, & Quintanilha, 2006; Imhoff, Lawrence, Stutzer, & Elvidge, 1997; Jiansheng Wu et al., 2014) can reduce this effect. We applied the thresholding method mainly for its simplicity, and set the threshold as DN value of 12, following Jiansheng Wu et al. (2014). Pixels with an average DN less than 12 during the study period were considered as non-urban and were excluded from the analysis.

The saturation occurs due to NTL sensor's limited detection range between 0 and 63, which is insufficient to measure urban centers where the actual NTL intensity might exceed the upper limitation. However, the proportion of saturated pixels in China is less than 1.2% for each NTL image in the study time period (J. Liu & Li, 2014), making saturation a minor and acceptable issue. In those saturated areas, we acknowledge that the dataset is not able to show further variations in NTL intensity.

The third and particularly important issue is the inconsistency between NTL satellite missions. Due to the differences between the sensors and a lack of on-board calibration, NTL annual composites of the same years from different satellite missions may have different DNs for the same locations (Jiansheng Wu, He, Peng, Li, & Zhong, 2013). As Figure 2-1(a) shows, before the calibration, China's sum of DNs in the overlapping years were different between the missions. Such inconsistency is an obstacle for a time-series analysis; therefore, we calibrated the annual images to restore the consistency before further analysis using a quadratic regression model (equation (1)) by Elvidge et al. (2009). We obtained the model parameters $a_t$, $b_t$, and $c_t$ for each year $t$ from a study by Liu & Li (2014), which used 2007's NTL image from mission F16 as the baseline and conducted the quadratic regressions over the invariant areas that should have little change of actual DNs over time. In addition, we excluded the pixels with 0 pre-calibration value from the calibration, as 0 indicated no light in the original data. Figure 2-1($b$) shows the sum of DNs in our study area from different missions after the calibration, where the inconsistency at certain years, such as 2000 and 2010, were bridged, providing a continuous change of DNs over time and indicating the calibrated data can be used for the subsequent analysis.

$$L'_t = a_t \times {L_t}^2 + b_t \times L_t + c_t \quad (1)$$

where $L'_t$ is a pixel's DN after calibration in year $t$, $L_t$ is the pixel's original DN in that year, and $a_t$, $b_t$, and $c_t$ are the corresponding model parameters.

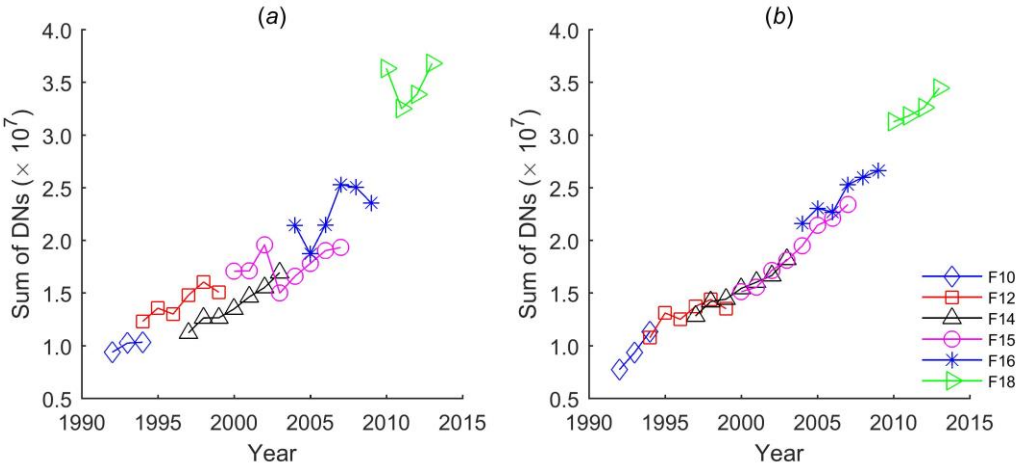



Figure 2-1 Sum of DNs in the study area from 1993 to 2013 across all NTL satellite missions, (a): before calibration; (b): after calibration. F10, F12, F14, F15, F16, and F18 are the six NTL satellite missions included in the dataset.

### 2.2.2 Trajectory extraction, noise removal, and Principal Component Analysis on the parameters

We extracted and smoothed pixel-based, time-series NTL trajectories from the calibrated NTL images. First, for any year with two NTL images, we used the average of the two for that year. Second, we extracted time-series NTL trajectories for each 1 km by 1 km pixel. Finally, we applied the Savitzky-Golay filter (Savitzky & Golay, 1964) to

remove the noise and to smooth each trajectory, as we assumed the change of DNs in each pixel should be a continuous process.

We then extracted two sets of four parameters to represent each trajectory. The first set utilized the DNs of each trajectory, which included the maximum, minimum, mean, and standard deviation. The second set used the slope of each trajectory indicating how much the DNs changed between two adjacent years (i.e. year$_t t$ to $year_{t+1} t+1$)

(equation (2)). We included the maximum, minimum, and standard deviation of the slopes, with an additional mean slope calculated from a linear regression (equation (3)) of each trajectory's DNs. Figure 2-2 shows a conceptual diagram of the main parameters.

$$S_{t+1} = L'_{t+1} - L'_t \quad (2)$$

$$\boldsymbol{L'} = \beta_0 + \beta_1 \boldsymbol{T} + \boldsymbol{\varepsilon} \quad (3)$$

where $S_{t+1}$ is the slope at year $t+1$ in a pixel, $L'_{t+1}$ and $L'_t$ are the calibrated DNs. $\boldsymbol{L'}$ is a vector of the calibrated DNs for the years studied in a pixel, $\beta_0$ is the intercept, $\beta_1$ is the slope of a linear regression, or the mean slope in our study, $\boldsymbol{T}$ is a vector of the years, $\boldsymbol{\varepsilon}$ is the random error.

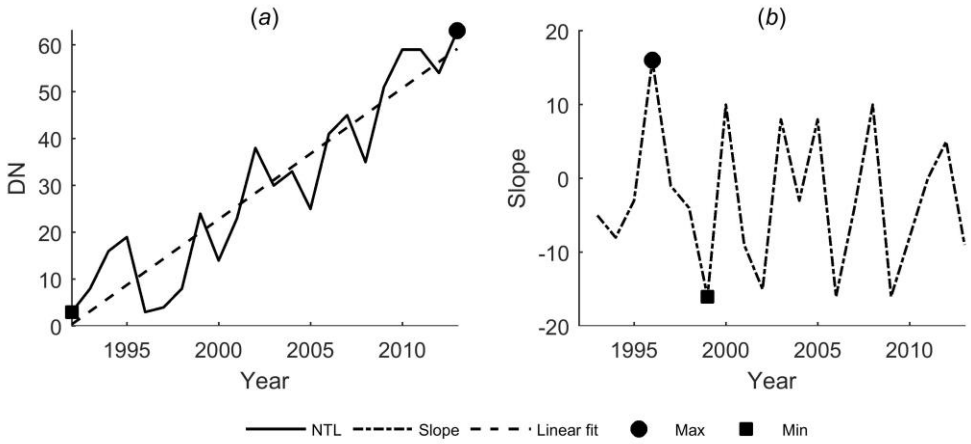



Figure 2-2 Conceptual diagram of main trajectory parameters, (a): parameters based on DN, including maximum and minimum DN, and a mean slope (β1) from a linear fit defined by equation (3); (b): parameters based on slope, including maximum and minimum slope. Slope is defined by equation (2).

Finally, we conducted the Principal Component Analysis (PCA) (Pohl & Genderen, 1998) over the eight parameters to reduce dimensionality of the data and computational intensity of the subsequent analysis. PCA is a widely-adopted dimension reduction technique that linearly combines the input data to get a set of Principal Components (PCs) that are more independent of each other. Each PC explains certain percentage of variation, or certain amount of information, in the original data. We used the first three PCs that together explained 97.97% of the variation, in the following classification.

### 2.2.3 K-means unsupervised classification and determining the optimal number of classes

We applied the *k*-means unsupervised classification (Lloyd, 1982) to classify the NTL trajectories into a limited number of classes, with the first three PCs as the inputs. Each class was expected to represent a type of urbanization dynamics. *K*-means iteratively partitions n samples in to *k* classes, and minimizes the within-class sum of square distance from each member to the mean center (Lloyd, 1982). The algorithm starts with allocating a set of random initial mean centers, and the result is dependent on the centers' locations. Therefore, it's possible to have locally optimal results, rather than globally optimal ones (Steinley, 2003). To reduce the dependence of results on the initial mean center allocation, we iterated the classification 10 times to generate initial mean centers and used the iteration with the optimal results in our analysis.

Another important consideration in this classification was the number of classes. We used the silhouette analysis (Rousseeuw, 1987) as a basis to determine this number, which was different from previous studies that were more relied on subjective decisions (Qingling Zhang & Seto, 2011; Jiansheng Wu et al., 2014). Given a classification with a certain number of classes, the silhouette analysis measures the closeness of each member in one class to members in the neighboring classes (Scikit-learn, 2015), and provides an average silhouette coefficient that ranges from -1 to 1 for all classes. 1 indicates a good classification where one class is far away from the neighboring classes, and -1 indicates a poor classification where one class is very close to the neighboring classes. After testing with a series numbers of classes, we chose the classifications with higher silhouette coefficients as the candidates. In addition, we also examined whether a candidate can properly represent urbanization dynamics in the study area, to find the optimal classification.

We chose five classes in the *k*-means unsupervised classification. The silhouette analysis showed that a 2-class classification had the highest coefficient (Figure 2-3). However, we considered that having only two classes would be insufficient to differentiate urbanization dynamics. Thus we instead used a 5-class classification with the second highest coefficient where the number of classes could sufficiently represent various urbanization dynamics.

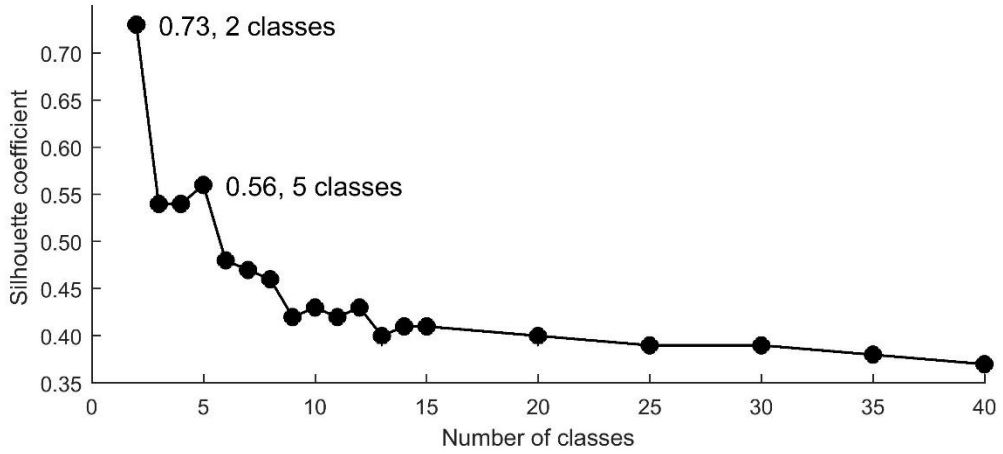



Figure 2-3 Change of silhouette coefficient value with different numbers of classes.

Finally, we labelled each class with a unique type of urbanization dynamics by examining the shape of each class's average trajectory, the range of the members, and the shapes of 1% randomly sampled members in this class. We also used the average value of the eight trajectory parameters in each class to help the labelling.

### 2.2.4 Validation

Validating the assignment of the unsupervised classes to specific urbanization dynamics was challenging, as the classification results were at a finer spatial resolution compared with candidate validation datasets, such as socioeconomic data typically reported at coarse administrative divisions. Several studies verified NTL-derived urbanization dynamics at pixel scale with fine resolution Google Earth images from different years (Q. Ma et al., 2014; Qian Zhang & Seto, 2013). In particular, using Google Earth images, Qian Zhang and Seto (2013) validated their urbanization dynamics assignment, such as rapid urbanization and slow urbanization, for their unsupervised classes of NTL trajectories.

To validate the urbanization dynamics assignment, we examined whether changes in NTL and impervious-surface-derived urbanization ratio within each class had similar patterns in a set of selected cities. We used 30 m resolution Landsat remote sensing images for every two years between 1992 and 2013 (i.e. 1992, 1994, …, 2010, 2011, 2013) to estimate the extent of impervious surface as described below, and calculated urbanization ratio for each unsupervised class $i$, or the fraction of impervious surface within the spatial extent of $i$, in a city $j$ in year $t$ (equation (4)). Impervious surface was used as an indicator for urbanization in many studies (Jat et al., 2008; Schneider & Woodcock, 2008; Bhatta, 2009; Bhatta, Saraswati, & Bandyopadhyay, 2010). If changes in the urbanization ratio and NTL were similar, we assumed that we had a reasonable assignment of the unsupervised classes to urbanization dynamics. We further measured the similarity as the correlation between a class's average NTL trajectory and the class's average urbanization ratio in the set of cities (equation (5)).

$$U_{i,j,t} = \frac{D_{i,j,t}}{C_{i,j}} \tag{4}$$

$$r_i = \frac{\sum\left(\overline{\boldsymbol{L}'}_{i,t} - \overline{\overline{\boldsymbol{L}'}}_{\iota}\right)\left(\overline{\boldsymbol{U}}_{i,t} - \overline{\overline{\boldsymbol{U}}}_{\iota}\right)}{\sqrt{\sum\left(\overline{\boldsymbol{L}'}_{i,t} - \overline{\overline{\boldsymbol{L}'}}_{\iota}\right)^2}\sqrt{\sum\left(\overline{\boldsymbol{U}}_{i,t} - \overline{\overline{\boldsymbol{U}}}_{\iota}\right)^2}} \tag{5}$$

where $U_{i,j,t}$ is the urbanization ratio for class $i$ in a city $j$ and year $t$ (i.e. 1992, 1994, …, 2010, 2011, 2013), $D_{i,j,t}$ is the impervious surface area within class $i$'s spatial extent in city $j$ and year $t$, $C_{i,j}$ is the area of class $i$'s spatial extent in city $j$. $r_i$ is the correlation between class $i$'s average NTL trajectory ($\overline{\boldsymbol{L}'}_{i,t}$) over the class members, and average urbanization ratio trajectory ($\overline{\boldsymbol{U}}_{i,t}$) over the selected cities, $\overline{\overline{\boldsymbol{L}'}}_{\iota}$ and $\overline{\overline{\boldsymbol{U}}}_{\iota}$ are average value for $\overline{\boldsymbol{L}'}_{i,t}$ and $\overline{\boldsymbol{U}}_{i,t}$ over the years ($\boldsymbol{t}$).

We obtained the impervious surface by a Normalized Difference Spectral Vector (NDSV) method (Patel et al., 2015) on Landsat remote sensing images for the selected years in five cities including Harbin in the north, Nanning in the south, Hangzhou and Nanchang in the east, and Hanzhong in the west. The cities were selected mainly based on representativeness for geographic region and urbanization dynamics, and availability of high-quality Landsat images with limited cloud cover. Two cities in the east were included as most Chinese cities were in this region. NDSV contains all the possible two-band normalized spectral indices from an image, such as the Normalized Difference Vegetation Index (NDVI), the Normalized Difference Water Index (NDWI), and the Normalized Built-up Index (NDBI), which can be used to differentiate various land cover types (Patel et al., 2015). We constructed a NDSV containing 15 indices from Landsat images, and conducted supervised classification with Support Vector Machine (Huang, Davis, & Townshend, 2002) to identify vegetation, water, and impervious surface in the five cities. We queried orthorectified Landsat Top of Atmosphere (TOA) images with cloud cover less than 10% in each year as the input. Depending on cloud cover, we typically found a collection of several images from different times in a year, and we reduced such a collection to one cloud-free image using the median of DNs from the different times at each pixel. Median helped to remove clouds that had a high DN, and shadows that had a low DN (Bian, Li, Liu, & Huang, 2016). Since a pixel was not always under cloud or shadow, this process would yield a cloud-free image. Training samples were collected where land cover was consistent across all the selected years. We then classified the images using the Support Vector Machine algorithm (Huang et al., 2002). Finally, we visually inspected the classification results and found them satisfactory. We accessed the images, collected the samples, and conducted the classification using the Google Earth Engine, an online environmental data monitoring platform that analyzed publicly available remote sensing images such as Landsat (Gorelick, 2013; Patel et al., 2015), to reduce processing time and enhance feasibility.

In this study, the impervious surface-derived urbanization ratio had to be used with caution, for two reasons. First, the impervious surface used in the calculation may contain some non-urban areas that were excluded in the analysis due to low NTL intensity. Second, changes in the impervious surface cover do not necessarily capture every dimension of urbanization. Urbanization is a combination of land cover change, population migration, and economic development. Only looking at the change of impervious surface or the derived urbanization ratio may neglect the other two dimensions of urbanization. However, NTL may capture all the three dimensions, as NTL is essentially an indicator for human activity. Given these two reasons, the change of impervious surface does not fully equal to the change of NTL. However, we still used impervious surface from Landsat images for validation as the surface was available in a finer resolution compared with the NTL-derived unsupervised classes. Other advantages of Landsat images included a broad temporal and spatial coverage, and being publicly accessible, making it possible to apply this validation method in other areas of interest.

### 2.2.5 Urbanization pattern recognition

Following the classification and validation, we calculated the proportion of each type of urbanization dynamics in different Chinese administrative divisions including province (primary administrative division), prefecture city (secondary administrative

division), and county (tertiary administrative division). Such proportions showed how different urbanization dynamics contributed to the overall urbanization between 1992 and 2013 in the administrative divisions.

To measure the spatial pattern, we conducted both global and local Moran's *I* analysis to quantify the spatial autocorrelations in those proportions. Moran's *I* is a commonly used measure for spatial autocorrelation (Su, Jiang, Zhang, & Zhang, 2011), which indicates how a variable correlates with itself over space. Calculated by equation (6) (ESRI, 2016), global Moran's *I* measures the overall spatial autocorrelation across all the administrative divisions in the study area. Its value ranges from -1 to 1, where 1 means positive autocorrelation or similar values neighboring with each other, 0 means spatial randomness, and -1 means negative autocorrelation or dissimilar values neighboring with each other (Su et al., 2011). Local Moran's *I* (Anselin, 1995), calculated by equations (7)-(8) (ESRI, 2016), measures spatial autocorrelation for each administrative division $i$, and further finds clusters and outliers in the divisions. Clusters are administrative divisions with similar values, which include high-high clusters where a high value is surrounded by high values, and low-low clusters where a low value is surrounded by low values. Outliers are administrative divisions with contrasting values, which further include high-low outliers where a high value is surrounded by low values, and low-high outliers where a low value is surrounded by high values.

Given the same set of data, different spatial weight ($w_{ij}$) may lead to different results in both global and local Moran's *I*. In this study, we performed both global and local Moran's *I* using ArcMap 10.4 (ESRI, 2016), and chose 'contiguity edges corners' to generate the spatial weight. In 'contiguity edges corners', administrative divisions that share an edge and/or a corner with the target administrative division have weights of 1 to be included in the Moran's *I* computation, and the rest have weights of 0 to be excluded from the computation.

$$I = \frac{n\sum_j \sum_k w_{j,k}(x_j - \bar{\boldsymbol{x}})(x_k - \bar{\boldsymbol{x}})}{(\sum_j \sum_k w_{j,k})(\sum_j (x_j - \bar{\boldsymbol{x}})^2)} \tag{6}$$

$$I_j = \frac{(x_j - \bar{\boldsymbol{x}})}{S_j^2} \sum_k w_{j,k}(x_k - \bar{\boldsymbol{x}}) \tag{7}$$

$$S_j^2 = \frac{\sum_k (x_k - \bar{\boldsymbol{x}})^2}{n-1} \tag{8}$$

where $I$ is the global Moran's *I*, $I_j$ is the local Moran's *I* for administrative division $j$, $x_j$ is the value of administrative division $j$, $x_k$ is the value of $j$'s neighbour $k$ ($k \neq j$), $\bar{\boldsymbol{x}}$ is the mean value of all administrative divisions, $w_{j,k}$ is a spatial weight between $j$ and $k$, $n$ is the number of administrative divisions.

We computed global Moran's *I* for the proportion of each type of urbanization dynamics at provincial, prefecture city, and county levels. For dynamics showing accelerated urbanization, we further employed local Moran's *I* to identify local clusters and outliers at prefecture city and county levels. The spatial patterns of the proportions, particularly the clusters, help to inform the underlying urbanization process. For example,

positive global spatial autocorrelation of a certain type of urbanization dynamics indicates the clustering of administrative divisions with similar proportions of the dynamics, which may further suggest similarity in other characteristics, such as policy, development history, and natural resources that form those clusters. Local clusters identify which administrative divisions belong to the same cluster, which allow synthesizing the commonalities within a cluster and differences between clusters. Those commonalities and differences may explain the observed pattern of the proportions.

## 2.3 Results

### 2.3.1 Five Types of Urbanization dynamics in China

We labelled the five classes as stable urban activity, high-level steady growth, acceleration, low-level steady growth, and fluctuation. Figure 2-4 shows each class's average trajectory, the 90% range (i.e. 5th percentile to 95th percentile), and 1% randomly sampled trajectories in the class. Table 2-1 shows average value of the eight trajectory parameters for each class. Class 1 represented stable urban activity, as its trajectories had relatively constant DNs that were close to the maximum value (Figure 2-4(*a*)) and the smallest mean slope indicating the least changes between 1992 and 2013 (Table 2-1). Both class 2 and class 4 represented steady growth as their average trajectories were close to straight lines (Figure 2-4(*b, d*)), and trajectories in both classes had smaller mean slopes with less variation (Table 2-1). However, class 2 on average had higher mean DN compared with class 4, therefore we considered class 2 as high-level steady growth and class 4 as low-level steady growth. Class 3 represented acceleration as its average trajectory showed a concave shape where the increase of DN started to accelerate between 2000 and 2005 (Figure 2-4(*c*)). Trajectories in this class also had the highest mean slopes, indicating the most rapid growth (Table 2-1). Finally, class 5 represented fluctuation as its sampled trajectories had a wide range, and were not stable compared with other classes (Figure 2-4(*e*)). Trajectories in this class on average also had the largest average standard deviations for slope (Table 2-1), indicating high variation in annual changes of NTL intensity.

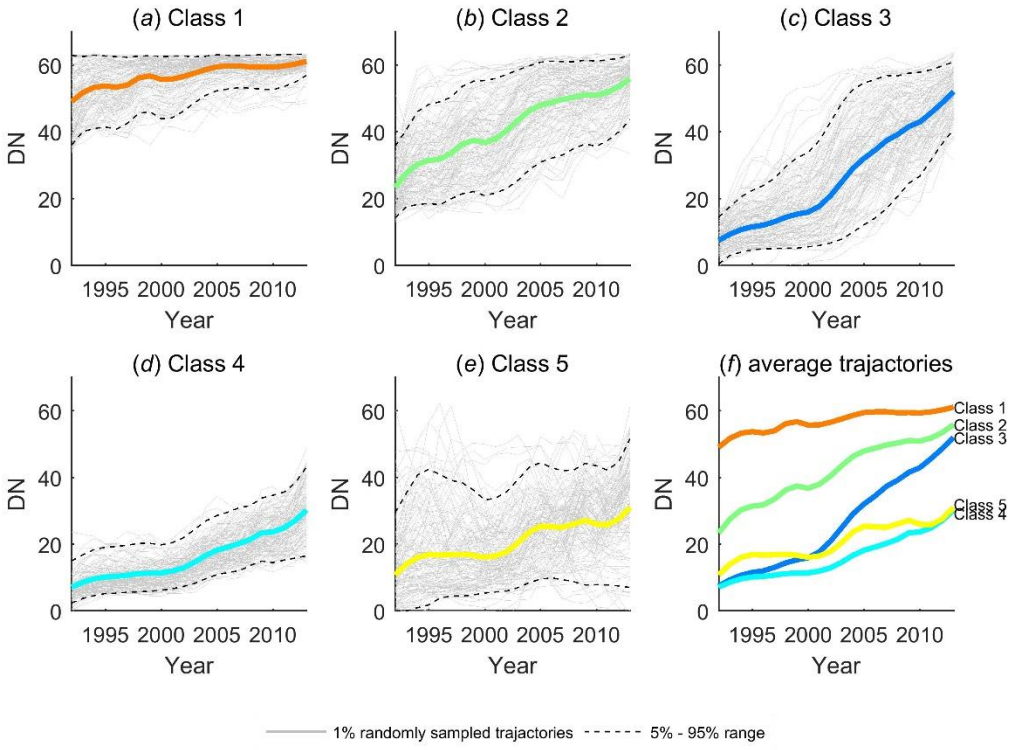



Figure 2-4 Average trajectories for each class from the k-means unsupervised classification with k=5. Trajectories from 1% randomly selected samples from each class and 90% range of all members are also plotted.

Table 2-1 Average value of the eight trajectory parameters in each class. Class 1: stable urban activity; class 2: high-level steady growth; class 3: acceleration; class 4: low-level steady growth; class 5: fluctuation.

| | Class 1 | Class 2 | Class 3 | Class 4 | Class 5 |
|---|---|---|---|---|---|
| Number of trajectories | 23503 | 54542 | 127887 | 150504 | 28413 |
| Average Maximum DN | 61.69 | 56.51 | 52.52 | 31.44 | 40.80 |
| Average Minimum DN | 47.76 | 22.03 | 7.16 | 6.53 | 7.46 |
| Average Mean DN | 56.88 | 41.71 | 26.23 | 16.16 | 20.87 |
| Average Std Dev DN | 3.86 | 10.49 | 15.27 | 7.44 | 10.00 |
| Average Maximum Slope | 4.69 | 7.97 | 9.16 | 5.96 | 10.89 |
| Average Minimum Slope | -2.48 | -3.64 | -2.47 | -2.64 | -8.05 |

| | | | | | |
|---|---|---|---|---|---|
| Average Mean Slope | 0.47 | 1.44 | 2.19 | 1.00 | 0.81 |
| Average Std Dev Slope | 1.72 | 2.95 | 3.00 | 2.13 | 4.42 |

### 2.3.2 Validation

Validation results were consistent with our interpretation of the unsupervised classes as the types of urbanization dynamics. The comparison between the NTL trajectories and the urbanization ratio trajectories showed certain level of agreement between the two: for each mean NTL trajectory, the corresponding mean urbanization ratio trajectory had a similar change pattern with correlation coefficient between 0.80 and 0.99 (Figure 2-5). We also noticed that from class 2 to class 5, the NTL trajectories tended to achieve relatively higher levels than the urbanization ratio trajectories did in years. Such differences were partially due to urbanization ratio only capturing changes in impervious surface while NTL capturing other dimensions of urbanization such as increased human and economic activities. Human and economic activities could continue to intensify while land transformation to impervious surface stopped. In addition, the 'overglow' effect could cause overestimation of NTL-derived urbanized areas (i.e. $Class_{i,j}$ in equation (4)), which in turn underestimated the urbanization ratio. In general, the high correlation between the two types of trajectories and their relatively similar growth pattern suggested that the interpreted NTL dynamics might be consistent with the temporal change of urbanization ratio.

In addition, spatial distribution of the classes in individual cities was consistent with our understanding about Chinese cities. For most cities, we found a concentric gradient of urbanization dynamics (Figure 2-6). From the urban core to the periphery, there was a transition from stable urban behavior (class 1) to high-level steady growth (class 2) to acceleration (class 3) to low-level stead growth (class 4) to fluctuation (class 5). This transition could be explained by the following: (1) Chinese urban cores (class 1 and class 2) were more stable and developed due to longer development and higher concentrations of population, jobs, amenities, and other resources, (2) the middle zone (class 3) tended to be the most active area for its abundant land resource, low land price, and close distance to urban cores and existing resources, and (3) the periphery areas were too far away from the cores and resources to have enough incentives for urbanization (class 4), and might become unstable (class 5). Both validation using the urbanization ratio and spatial distribution of classes indicated that our labels we were likely to match the dominant urbanization dynamics reasonably well.

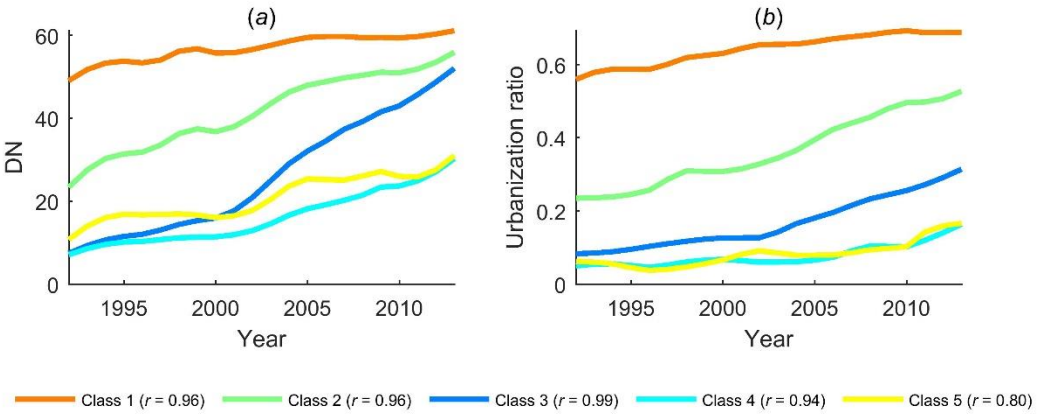



Figure 2-5 Comparison between the trajectories from NTL and urbanization ratio, (*a*): mean NTL trajectories, (*b*): mean urbanization ratio trajectories.

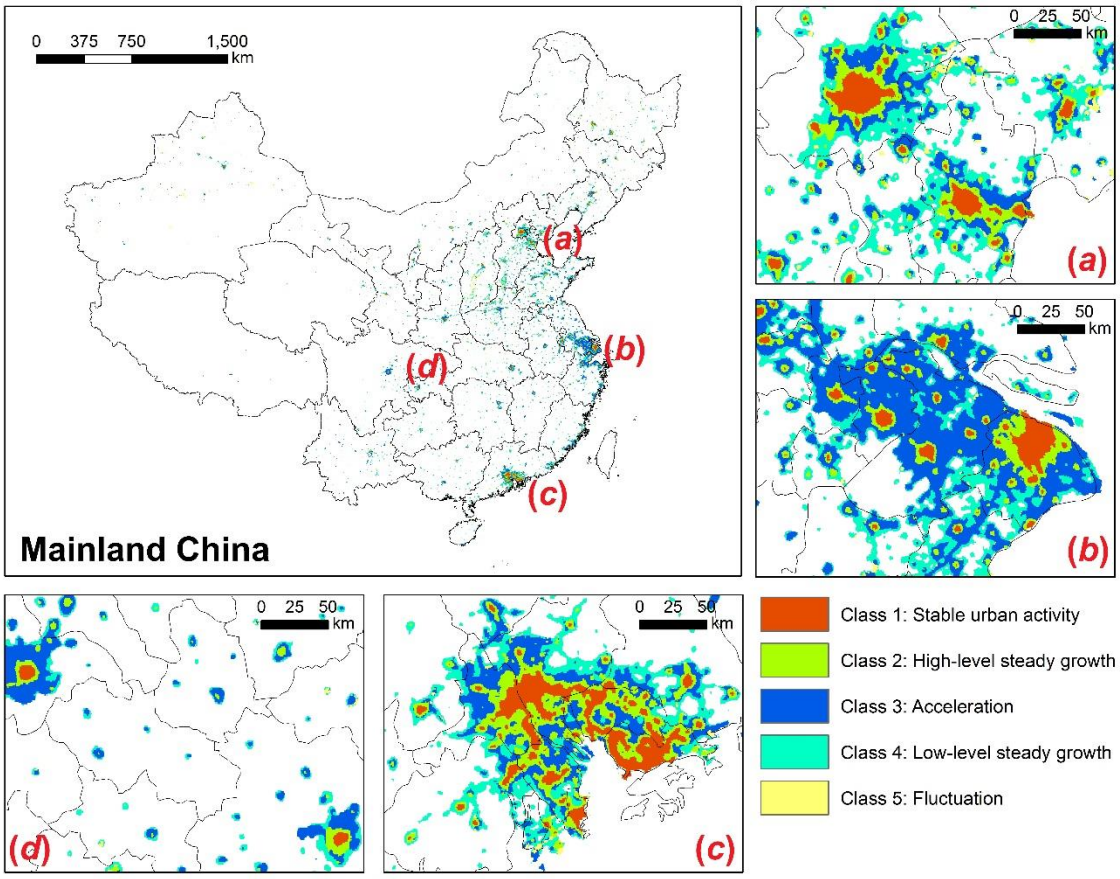



Figure 2-6 Map of different urbanization types from the unsupervised classification with five classes in four major urban agglomerations in China, (*a*): Beijing – Tianjin, (*b*): the Yangtze River Delta, (*c*): the Pearl River Delta, (*d*): Chendu - Chongqing.

### 2.3.3 Composition of the urbanization dynamics in Chinese administrative divisions

We summarized the proportion and the global Moran's *I* value for each type of urbanization dynamics at provincial level in Table 2-2, and plotted the composition of the urbanization dynamics in each province in Figure 2-7. The stable urban activity contributed to an average of 5.63% of urban area in a province. There was a positive but not significant spatial autocorrelation of stable urban activity's proportion at provincial level. Several provinces in the north, and Guangdong and Guangxi Province in the south tended to more stable urban activity (Figure 2-8 (*a*)). Shanghai, Beijing, Guangdong, Tianjing, and Liaoning were the five administrative divisions with the highest proportion. An average of 13.35% of a province's urban area was under the high-level steady growth. Guangdong, Shanghai, Beijing, Tianjing, and Heilongjiang had the highest proportion of this type of dynamics. The proportion of high-level steady growth also showed a positive but not significant spatial autocorrelation between the provinces. The acceleration urbanization dynamics on average occupied 33.90% of a province's urban area. Jiangsu, Xizang, Zhejiang, Ningxia, and Chongqing had the highest proportions of such dynamics. The proportion of acceleration had a significant ($p = 0.030$), positive spatial autocorrelation. The low-level steady growth urbanization dynamics on average accounted for 39.27% of a province's urban area, and its proportion had a significant ($p = 0.010$), positive spatial autocorrelation. Provinces with the highest proportion of low-level steady growth included Guizhou, Hebei, Guangxi, Hunan and Henan. Compare with the spatial patterns for stable urban activity and fluctuation (Figure 2-8 (*a*), (*e*)), those of the two steady growth and the acceleration (Figure 2-8 (*b*), (*c*), (*d*)) were more ambiguous: we found less obvious local clusters or geographic boundaries between high and low proportions for the latter three (i.e. high-level steady growth, low-level steady growth, and acceleration). An average of 7.84% of a province's urban area was under the fluctuation. The proportion of this type of dynamics also had a significant ($p = 0.010$), positive spatial autocorrelation. Provinces with lower proportion of fluctuation were concentrated along the eastern coast of China, and the ones with higher proportion were mainly clustered in the northwest (Figure 2-8 (*e*)). The provinces with the highest proportion of fluctuation were Xinjiang, Shanxi, Heilongjiang, Qinghai, and Shanxi.

Table 2-2 Summary statistics and global Moran's I of the composition of the five type of urbanization dynamics in provinces, prefecture cities and counties

| Administrative division | Urbanization dynamics pattern | Proportion in each administrative division | | | | Global Moran's *I* |
|---|---|---|---|---|---|---|
| | | Mean (%) | Minimum (%) | Maximum (%) | Std Dev | |
| Province | Stable urban activity | 5.63 | 1.82 | 17.82 | 4.18 | 0.1 |
| ($n = 31$) | High-level steady growth | 13.35 | 7.62 | 22.86 | 4.36 | 0.15 |
| | Acceleration | 33.90 | 13.49 | 56.17 | 10.60 | 0.21** |
| | Low-level steady growth | 39.27 | 19.31 | 55.90 | 8.06 | 0.25** |
| | Fluctuation | 7.84 | 0.24 | 27.70 | 6.35 | 0.23*** |
| Prefecture city | Stable urban activity | 3.55 | 0.00 | 49.93 | 5.17 | 0.21*** |

| | | | | | | |
|---|---|---|---|---|---|---|
| (*n*=361) | High-level steady growth | 12.13 | 0.00 | 43.15 | 7.15 | 0.35*** |
| | Acceleration | 30.87 | 0.00 | 85.71 | 16.72 | 0.33*** |
| | Low-level steady growth | 45.23 | 5.37 | 100.00 | 16.47 | 0.28*** |
| | Fluctuation | 8.22 | 0.00 | 68.28 | 9.03 | 0.34*** |
| County (*n*=2081) | Stable | 1.45 | 0.00 | 50.94 | 4.51 | 0.06*** |
| | High-level steady growth | 8.17 | 0.00 | 46.51 | 9.69 | 0.28*** |
| | Acceleration | 27.21 | 0.00 | 100.00 | 22.20 | 0.34*** |
| | Low-level steady growth | 54.04 | 0.00 | 100.00 | 24.75 | 0.31*** |
| | Fluctuation | 9.13 | 0.00 | 100.00 | 13.94 | 0.27*** |

**: significant at 95% confidence interval ($p \leq 0.05$); ***: significant at 99% confidence interval ($p \leq 0.01$).

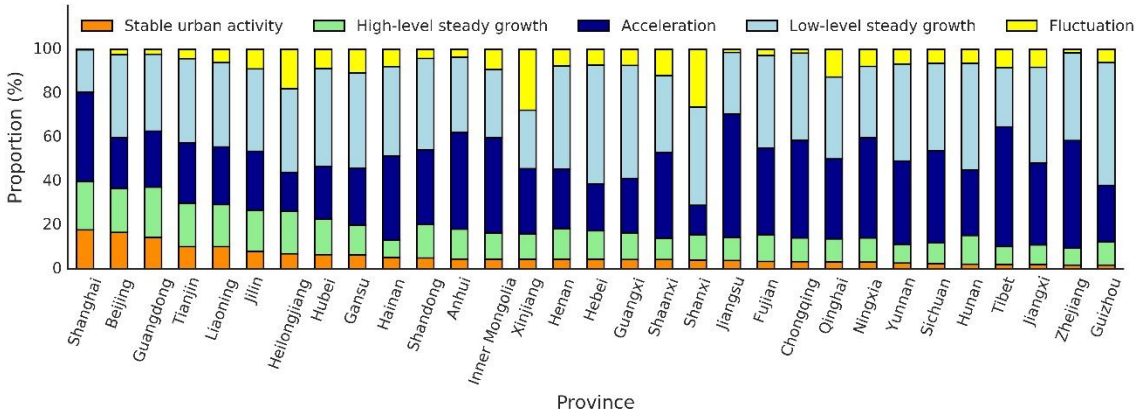



Figure 2-7 Composition of the five types of urbanization dynamics in Chinese provinces.

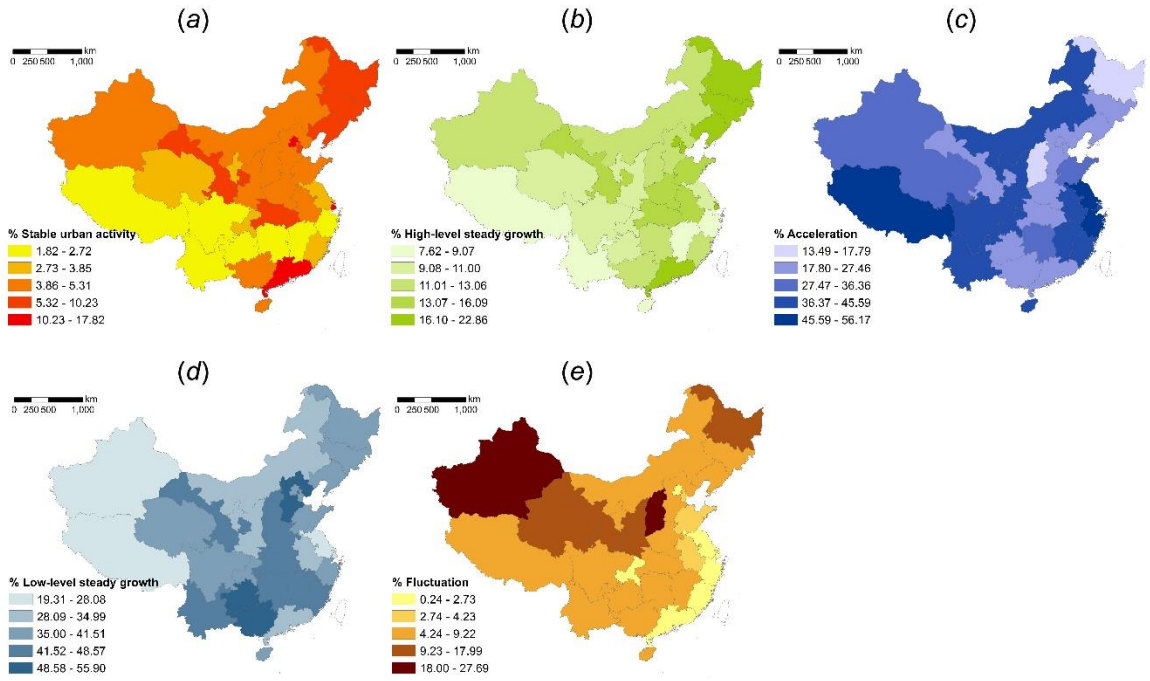



Figure 2-8 Proportion of the five types of urbanization dynamics at provincial level. (*a*): stable urban activity, (*b*): high-level steady growth, (*c*): acceleration, (*d*): low-level steady growth, (*e*): fluctuation.

We further summarized composition of the five urbanization dynamics for prefecture cities and counties in China (Table 2-2). On average, 3.55% of the urbanized area in a prefecture city was stable, 12.13% was under the high-level steady growth, 30.87% was under the acceleration, 45.23% was under the low-level steady growth, and 8.22% was under the fluctuation. In an average county, 1.45% of the urbanized area was stable, 8.17% was under the high-level steady growth, 27.21% was under the acceleration, 54.04% was under the low-level steady growth, and 9.13% was under the fluctuation. Global Moran's *I* showed that those proportions were all significantly and spatially auto-correlated.

The acceleration was used as the dynamics of interest to investigate its proportion in prefecture cities and counties, and the corresponding local clusters and outliers (Figure 2-9 and Figure 2-10). We chose this type of dynamics as it highlighted areas transforming from rural with low NTL intensity to urban with high NTL intensity at an accelerating speed (Figure 2-4 (*c*)), which might indicate an aggressive urbanization process and potentially lead to ghost cities. Acceleration also appeared to be a ubiquitous phenomenon in China, as on average 33.90% of a province's urban area, 30.87% of a prefecture city's urban area, 27.21% of a county's urban area were composed of such dynamics. In addition, the ambiguous spatial pattern of this dynamics at provincial level required further examination at the other two levels. Local Moran's *I* analysis showed that at both county and prefecture city levels, the clusters with high proportion of the accelerated growth were mainly located in the northern, northwestern, southwestern, and eastern part of the country, while those with the low proportion tended to be in the central and southern part of the country. There were also several outliers that had dissimilar values with their neighbors in the country (Figure 2-10).

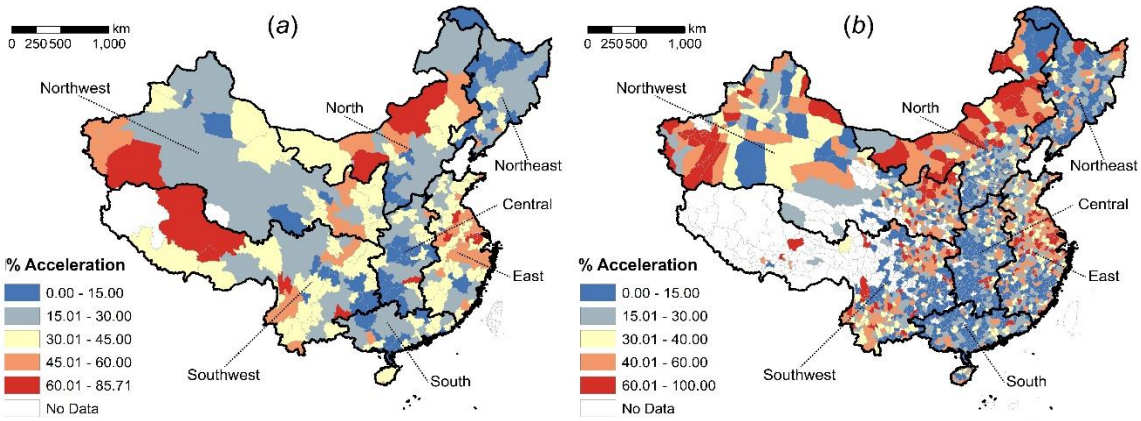



Figure 2-9 Proportion of class 3 (acceleration) for China's administrative divisions, (a): the proportion for prefecture cities, (b): the proportion for counties.

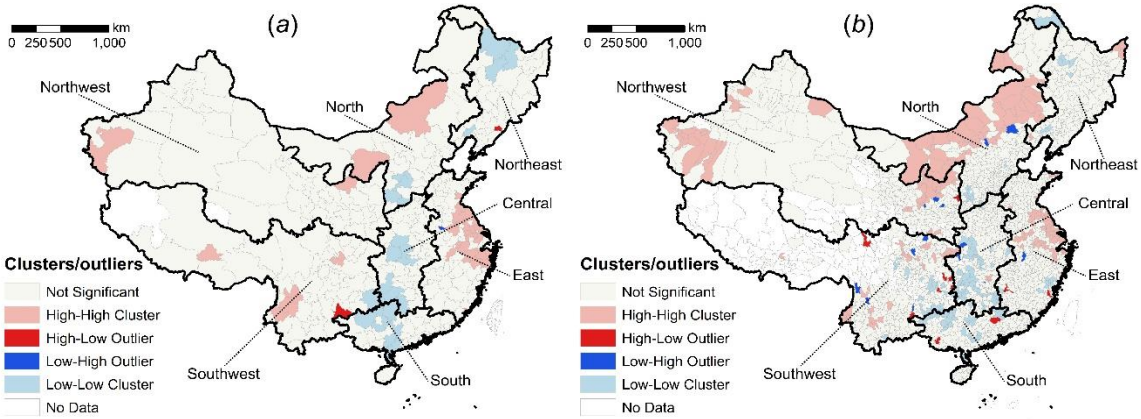



Figure 2-10 Clusters and outliers for proportion of class 3 (acceleration) for China's administrative divisions, (a): prefecture cities, (b): counties.

## 2.4 Discussion

### 2.4.1 Understanding urbanization through the new lens

Understanding different urbanization dynamics and their spatial patterns is an important prerequisite to inform planning in the face of growing urban population and increasing environmental footprint of expanding urban regions (Deng, Wang, Hong, & Qi, 2009; L. Jiang & O'Neill, 2015; Stathakis et al., 2015; Jiansheng Wu et al., 2014; Qingling Zhang & Seto, 2011). With our innovative approach, we identified several important trajectories describing urbanization dynamics in mainland China. The five types of urbanization dynamics provided a spatially explicit reconstruction of urban development as manifested in the intensity of version 4 DMSP-OLS NTL signal. When combined with other socioeconomic datasets and knowledge about local and national developmental history and policy, the urbanization dynamics may help to reveal the large-scale urbanization processes and confirm the underlying mechanisms, which provide more insights into the ongoing urbanization in China.

We found the acceleration urbanization dynamics to be an especially useful indicator of fast and even aggressive urbanization, as it showed an accelerating

transformation from rural to urban between 1992 and 2013. Spatial pattern of these dynamics revealed urbanization characteristics at different spatial scales. At the finer scale (i.e. 1 km by 1 km pixel), the acceleration outlined areas under rapid urbanization, which provided a base map for more formal and labor-intensive verification methods such as in situ land use survey. At the aggregated scales, administrative divisions with high proportion of such dynamics formed several clusters in different parts of China (Figure 10).

The underlying processes that formed these clusters varied among different geographic regions. The clusters in the eastern part of China largely overlapped with the Yangtze River Delta (YRD), where urbanization was partially caused by development of non-agricultural industries through rural collective accumulation (Gu, Hu, Zhang, Wang, & Guo, 2011). Since 1991, particularly in the northern YRD, such industries have started to concentrate, and stimulated urbanization and construction of small towns (Gu et al., 2011). These industries also attracted significant amount of labor and caused population growth. Other clusters likely resulted from different processes, such as aggressive urbanization due to financial speculation (Sorace & Hurst, 2016) and land-centered urbanization policy (M. Chen et al., 2015; Sorace & Hurst, 2016) that encouraged construction of urban areas beyond population needs. A typical example of this aggressive process is Ordos in the Inner Mongolia Autonomous Region. The city was the poorest one in its province before 2004 and then rapidly became one of China's wealthiest cities due to the discovery and production of significant amount of natural resources such as rare earth minerals and natural gas (Woodworth, 2015; Sorace & Hurst, 2016). Such a great shift in economy stimulated an unsustainable urbanization, which was characterized by massive land development and an inflated real estate market, while lacked a broader industrial base other than the extraction of natural resources (Woodworth, 2015). When the real estate market collapsed in 2011 and China's economy slowed down, the resulting vacant residential and commercial projects made the city known as a 'ghost town' in many major media outlets (Woodworth, 2015). Data from China city statistical survey (National Bureau of Statistics of the People's Republic of China, 2016) showed that urbanization in Ordos was more likely driven by its economy rather than population when compared with YRD's typical city, Suzhou. With 1% increase in urban area between 2001 and 2013, Suzhou had 0.2% increase in population while Ordos only had 0.06% increase, suggesting that Ordos' urbanization was beyond the actual population growth needs (Table 2-3). Qianwen Zhang & Su (2016) also found that population were the main urban expansion driver for Nanjing and Hangzhou in the YRD, while economy became important for Hohhot in Inner Mongolia, suggesting different urbanization forces between the two regions. In both Suzhou and Ordos, our analysis was able to detect the areas under rapid urbanization (Figure 11).

For cases like Ordos, policy interventions are required to control rapid urbanization and to ensure that urban development is compatible with population and economic growth. As a result, China's central government announced the "National New-type Urbanization Plan (2014-2020)" in 2014, which regulated aggressive urban development as the plan required urban construction land to be less than 100 $m^2$ per capita in 2020 (X.-R. Wang et al., 2015). However, local governments may still find other ways to get around this regulation (X.-R. Wang et al., 2015). A potential solution to enforce the plan is to develop a monitoring framework that includes large scale mapping

tools, such as the approach of our study, to feasibly achieve a global understanding about urbanization and to find hidden issues under the overall pattern. The framework should also incorporate fine scale methods, such as in situ land use survey, to accurately identify places under aggressive urbanization.

Table 2-3 Percentage increase in built-up area, population, and GDP between 2001 and 2013 in Suzhou and Ordos. Original percentages are listed, and normalized percentages by built-up area are listed in the brackets.

| Parameter | Percentage increase (%) between 2001 and 2013 – original value (normalized value by increase in built-up area) | |
|---|---|---|
| | Suzhou | Ordos |
| Built-up area | 3.02 (1.00) | 6.38 (1.00) |
| Population | 0.59 (0.20) | 0.36 (0.06) |
| GDP | 9.70 (3.21) | 21.19 (3.32) |

Urbanization dynamics found in this study indicated growth in general, and we were not able to identify shrinking areas. Urban shrinkage is a world-wide phenomenon that cities experience population loss in large areas and economic transformations indicating a structural crisis (Wiechmann & Pallagst, 2012). In the Chinese context, Long and Wu (2016) argued that urban shrinkage was due to migrant flow from rural areas and small cities to large and developed cities, and discovered 180 shrinking cities in China using Chinese population census in 2000 and 2010. Due to the reduction in active urban land use, population, and economic activities, we assumed that shrinking cities would tend to show declining NTL trajectories.

Contrary to the expectations, we were not able to detect declining NTL trajectories, or urban shrinkage, in this study. Such inability was likely due to the following reasons. First, most Chinese urban areas were still developing between 1992 and 2013, which led to generally increasing trends in NTL. We observed decreases in NTL in several time periods, such as years around the 2008 world financial crisis. However, since our method did not segment the trajectories between 1992 and 2013, we were not able to isolate the segments with decreased NTL and thus failed to find areas that were declining in specific time periods. Second, there could be time lags between urban shrinkage and NTL change: urban areas might be abandoned, but electricity was still used and delayed the pronounced decline in NTL. Since urban shrinkage is an emerging phenomenon in China, our study period (i.e. 1992 – 2013) might be too short to incorporate such time lags. Finally, as the country is still urbanizing, the shrinkage can be limited and occurred at spatial scales finer than the 1 km by 1 km NTL pixel. The scattered spatial distribution and fine scale may also prevent the shrinking areas from being detected by NTL trajectories.

### 2.4.2 An innovative approach to monitor urbanization dynamics and its challenges

Urbanization is an important global phenomenon that has both environmental and social implications. With the increased availability of data and analysis methods, there has been a tendency to study urban and regional issues through the data-driven, geospatial approach. Some examples include characterizing post-disaster population displacement pattern through 1.9 million mobile phone users in Haiti (Lu, Bengtsson, & Holme, 2012), and identifying “Ghost Cities” using larger volume positioning data and

points of interests (Chi et al., 2015). In general, those studies tried to use the pattern within the datasets to inform the underlying process. We consider our study as another example of such type of methods.

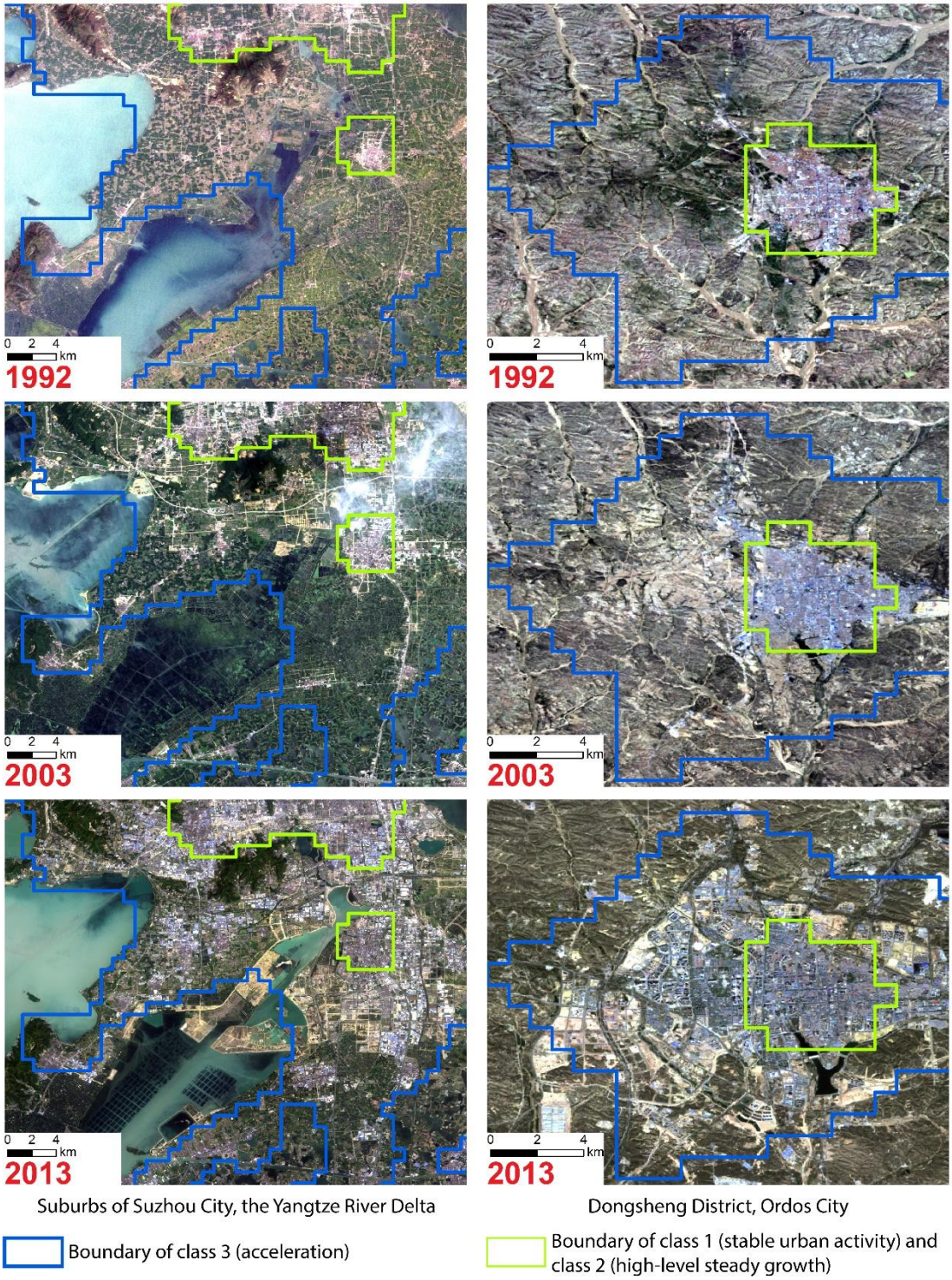



Figure 2-11 Urbanization outlined by class 3 (acceleration) in suburbs of Suzhou City, the Yangtze River Delta, and in Dongsheng District, Ordos City

Remote sensing allows monitoring large scale urbanization over relatively long time periods. However, previous remote sensing studies focused more on mapping the changing urban extent (Gao, Huang, He, Sun, & Zhang, 2016; Jat et al., 2008; L. Wang et al., 2012), which were insufficient to capture other dimensions of urbanization including changes in population and economic activities (Qingling Zhang & Seto, 2011). We provided an innovative and low-cost approach that used pixel-based, time-series NTL trajectories to monitor urbanization dynamics. Compare to other data, NTL provides a moderate spatial resolution, a long temporal coverage, and a low cost in computation. NTL's ability to characterize intensity of human activities also captures multiple dimensions in urbanization including urban extent, population and economic activities. Compare with previous studies that also employed time-series NTL trajectories (Qingling Zhang & Seto, 2011; Jiansheng Wu et al., 2014), we provided an alternative approach that addressed potential challenges in earlier studies. First, we applied the silhouette analysis to more objectively inform the number of classes based on the pattern of the data. Second, we validated our labelling of the unsupervised classes with time-series urbanization ratio estimated from impervious surface. With this approach, we could yield a reasonable set of urbanization dynamics for mainland China between 1992 and 2013. In addition, through visual inspection we found that the spatial distribution of our results was consistent with the one from (Qingling Zhang & Seto, 2011) for the Pearl River Delta (Figure 2-6(*c*)) and the Yangtze River Delta (Figure 6(*b*)), and with the one from Wu et al. (2014) for Beijing and Tianjin Area (Figure 2-6 (*a*)). Finally, we reduced the amount of input data for the classifier. To our knowledge, we were the first to use secondary descriptors for the trajectories, the eight trajectory parameters, in the classification. Furthermore, we could accelerate the classification by reducing the eight parameters to three input variables using the dimension reduction approach.

### 2.4.3 Limitations and future research directions

Data-driven methods to understand urbanization generally provide a broad picture about this phenomenon and may reveal some unexpected findings that cannot be uncovered with traditional methods such as survey, interview, and observation. However, data-driven methods have their own uncertainties, which mainly result from the properties of the data itself and the analysis method. For example, our study's outcome was affected by the NTL data quality and the *k*-means clustering method, even though we performed our analysis with the best of our knowledge. With an improved NTL dataset, such as the multi-month VIIRS (http://ngdc.noaa.gov/eog/viirs/download_monthly.html), and different sets of parameters for the *k*-means clustering, we might expect some potential differences in the analysis outcomes. Therefore, such data-driven studies should not only rely on a single dataset and a single method. Instead, multiple datasets (J. Liu, Li, Li, & Wu, 2016) and methods should be implemented in the future work to cross validate their results. When results from such data-driven studies are applied, they should be used as complimentary to those from the traditional methods (J. Liu et al., 2016).

Several specific challenges should be considered in interpreting our study results. First, identifying the number of classes for the unsupervised classification may have important effects on the derived urbanization dynamics, because too few classes may oversimplify the urbanization dynamics, while too many classes may be too difficult for interpretation. In addition, different number of classes may lead to different conclusions.

The underlying problem is the difficulty to draw distinct decision boundaries for the NTL trajectories, as they showed a broad and continuous variation of shapes and levels where the differences could be subtle (shown by the 1% randomly sampled trajectories in Figure 2-4). To overcome the problem, future studies should invite local planners to verify if the number of classes can properly identify different typologies of urbanization in the planners' communities and cities. Researchers could also select a few representative cities, and verify the urbanization dynamics from NTL with land cover change from high resolution remote sensing images (Qian Zhang & Seto, 2013). Furthermore, a sensitivity analysis should be carried out to understand how different numbers of classes may affect the analysis.

Second, the differences in timing of the NTL changes during the study period should be incorporated to characterize more specific urbanization dynamics and to reduce variation in each type of dynamics. The eight trajectory parameters we used did not show the timing of the changes, therefore our urbanization dynamics were relatively general with broad range in certain classes (for example, class 2, 3, and 5 in Figure 2-4), and required us to interpret the classes with caution when referring the temporal aspect of them. Such general urbanization dynamics were particularly insufficient to identify urban shrinkage as it might occur during specific time periods in mainland China. We suggest future researchers apply temporal segmentation methods that identify the timing of major changes to in this type of analysis, which would help to understand different phases in urbanization.

Finally, a more systematic validation of the NTL-derived urbanization dynamics in more cities and with multiple data sources is necessary. While our study only used five cities as validation sites, future researchers should include more representative cities in terms of urbanization type and level, spatial pattern of land use, economy, and geography. After stratifying the five validation cities into eastern cities including Hangzhou and Nanchang, and non-eastern cities including Harbin, Nanning, and Hanzhong, we found that non-eastern cities on average reached lower urbanization ratio in each NTL urbanization dynamics outlined zones (i.e. $Class_{ij}$ in equation (4)). We hypothesize that the lower urbanization ratios of the non-eastern cities were due to their stronger 'overglow' effect. The non-eastern cities are characterized by lower density development (F. Zhang et al., 2005), which could produce more scattered urban areas and light sources, and therefore stronger 'overglow' effect and overestimation of possible urbanizing areas from NTL-derived urbanization dynamics. Qian Zhang & Seto (2013) also found that NTL had high accuracy of characterizing urbanization dynamics in places where urbanization actually occurred, and lower accuracy in places that were not urbanized but influenced by 'overglow'. If possible, we recommend that future researchers identify NTL-derived urbanization dynamics within urbanized areas determined by land cover to improve accuracy. In addition, finer resolution socioeconomic data such as population and economy should be used together with land cover to verify if they show similar changes with NTL. With such more systematic validation and reduction in 'overglow' effect, NTL-derived dynamics can more accurately reflect the actual urbanization process.

## 2.5 Conclusions

Urbanization becomes one of the central drivers of global environmental change. Monitoring and understanding urbanization process becomes increasingly critical for planning and designing towards sustainable cities. With the increased availability of data and analytical power, we can utilize data-driven approaches along with traditional approaches to study urbanization, identify the patterns, and reveal the underlying process. By understanding the pattern and the process, planners, designers, and decision makers can apply proper strategies to enhance or intervene the ongoing urbanization. Our study, which used pixel-based, time-series NTL trajectories to identify different urbanization dynamics, demonstrates the power of such data-driven approach.

Building on existing knowledge and research approaches, we proposed an alternative method that better determined the types of urbanization dynamics and reduced the input information. Taking mainland China as a case study example, we could identify five major types of urbanization dynamics between 1992 and 2013. With a focus on accelerated growth, we found clusters with high proportion of this urbanization dynamics and interpreted the formation of the clusters by relating them to existing knowledge on policy and developmental history. This approach provides a low-cost, timely, and large-scale understanding about how urbanization happens over space and time, and should be used in combination with socioeconomic data, land cover mapping, and in situ land use survey to provide a solid basis for better urban planning, design, monitoring, and management.

**References**


Amaral, S., Monteiro, A. M. V., Camara, G., & Quintanilha, J. A. (2006). DMSP/OLS night-time light imagery for urban population estimates in the Brazilian Amazon. *International Journal of Remote Sensing*, *27*(5), 855–870. https://doi.org/10.1080/01431160500181861

Anselin, L. (1995). Local Indicators of Spatial Association—LISA. *Geographical Analysis*, *27*(2), 93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x

Bai, X., Shi, P., & Liu, Y. (2014). Society: Realizing China's urban dream. *Nature*, *509*(7499), 158–160. https://doi.org/10.1038/509158a

Bhatta, B. (2009). Analysis of urban growth pattern using remote sensing and GIS: A case study of Kolkata, India. *International Journal of Remote Sensing*, *30*(18), 4733–4746. https://doi.org/10.1080/01431160802651967

Bhatta, B., Saraswati, S., & Bandyopadhyay, D. (2010). Urban sprawl measurement from remote sensing data. *Applied Geography*, *30*(4), 731–740. https://doi.org/10.1016/j.apgeog.2010.02.002

Bian, J., Li, A., Liu, Q., & Huang, C. (2016). Cloud and Snow Discrimination for CCD Images of HJ-1A/B Constellation Based on Spectral Signature and Spatio-Temporal Context. *Remote Sensing*, *8*(1), 31. https://doi.org/10.3390/rs8010031

Buhaug, H., & Urdal, H. (2013). An urbanization bomb? Population growth and social disorder in cities. *Global Environmental Change*, *23*(1), 1–10. https://doi.org/10.1016/j.gloenvcha.2012.10.016

Chen, M., Liu, W., & Lu, D. (2015). Challenges and the way forward in China's new-type urbanization. *Land Use Policy*. https://doi.org/10.1016/j.landusepol.2015.07.025

Chen, Z., Yu, B., Hu, Y., Huang, C., Shi, K., & Wu, J. (2015). Estimating House Vacancy Rate in Metropolitan Areas Using NPP-VIIRS Nighttime Light Composite Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *8*(5), 2188–2197. https://doi.org/10.1109/JSTARS.2015.2418201

Chi, G., Liu, Y., Wu, Z., & Wu, H. (2015). Ghost Cities Analysis Based on Positioning Data in China. *ArXiv:1510.08505 [Cs]*. Retrieved from http://arxiv.org/abs/1510.08505

Deng, J. S., Wang, K., Hong, Y., & Qi, J. G. (2009). Spatio-temporal dynamics and evolution of land use change and landscape pattern in response to rapid urbanization. *Landscape and Urban Planning*, *92*(3–4), 187–198. https://doi.org/10.1016/j.landurbplan.2009.05.001

Doygun, H., & Alphan, H. (2006). Monitoring Urbanization of Iskenderun, Turkey, and its Negative Implications. *Environmental Monitoring and Assessment*, *114*(1–3), 145–155. https://doi.org/10.1007/s10661-006-2524-0

Elvidge, C. D., Imhoff, M. L., Baugh, K. E., Hobson, V. R., Nelson, I., Safran, J., … Tuttle, B. T. (2001). Night-time lights of the world: 1994–1995. *ISPRS Journal of Photogrammetry and Remote Sensing*, *56*(2), 81–99. https://doi.org/10.1016/S0924-2716(01)00040-5

Elvidge, C. D., Ziskin, D., Baugh, K. E., Tuttle, B. T., Ghosh, T., Pack, D. W., … Zhizhin, M. (2009). A Fifteen Year Record of Global Natural Gas Flaring Derived from Satellite Data. *Energies*, *2*(3), 595–622. https://doi.org/10.3390/en20300595

ESRI. (2016). ArcGIS Desktop (Version 10.4). ArcMap 10.4: ESRI.

Gao, B., Huang, Q., He, C., Sun, Z., & Zhang, D. (2016). How does sprawl differ across cities in China? A multi-scale investigation using nighttime light and census data. *Landscape and Urban Planning*, *148*, 89–98. https://doi.org/10.1016/j.landurbplan.2015.12.006

Ghosh, T., L Powell, R., D Elvidge, C., E Baugh, K., C Sutton, P., & Anderson, S. (2010). Shedding light on the global distribution of economic activity. *The Open Geography Journal*, *3*(1).

Gong, P., Wang, J., Yu, L., Zhao, Y., Zhao, Y., Liang, L., … Chen, J. (2013). Finer resolution observation and monitoring of global land cover: First mapping results with Landsat TM and ETM+ data. *International Journal of Remote Sensing*, *34*(7), 2607–2654. https://doi.org/10.1080/01431161.2012.748992

Gorelick, N. (2013). Google Earth Engine. *EGU General Assembly Conference Abstracts*, *15*, 11997. Retrieved from http://adsabs.harvard.edu/abs/2013EGUGA..1511997G

Gu, C., Hu, L., Zhang, X., Wang, X., & Guo, J. (2011). Climate change and urbanization in the Yangtze River Delta. *Habitat International*, *35*(4), 544–552. https://doi.org/10.1016/j.habitatint.2011.03.002

Hollis, G. E. (1975). The effect of urbanization on floods of different recurrence interval. *Water Resources Research*, *11*(3), 431–435. https://doi.org/10.1029/WR011i003p00431

Hu, Y., Gao, S., Janowicz, K., Yu, B., Li, W., & Prasad, S. (2015). Extracting and understanding urban areas of interest using geotagged photos. *Computers, Environment and Urban Systems*, *54*, 240–254. https://doi.org/10.1016/j.compenvurbsys.2015.09.001

Huang, C., Davis, L. S., & Townshend, J. R. G. (2002). An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, *23*(4), 725–749. https://doi.org/10.1080/01431160110040323

Imhoff, M. L., Lawrence, W. T., Stutzer, D. C., & Elvidge, C. D. (1997). A Technique for Using Composite DMSP/OLS "'City Lights'" Satellite Data to Map Urban Area. *Remote Sensing of Environment*, *61*(3), 361–370. https://doi.org/10.1016/S0034-4257(97)00046-1

Jat, M. K., Garg, P. K., & Khare, D. (2008). Monitoring and modelling of urban sprawl using remote sensing and GIS techniques. *International Journal of Applied Earth Observation and Geoinformation*, *10*(1), 26–43. https://doi.org/10.1016/j.jag.2007.04.002

Jiang, B., & Miao, Y. (2015). The Evolution of Natural Cities from the Perspective of Location-Based Social Media. *The Professional Geographer*, *67*(2), 295–306. https://doi.org/10.1080/00330124.2014.968886

Jiang, L., & O'Neill, B. C. (2015). Global urbanization projections for the Shared Socioeconomic Pathways. *Global Environmental Change*. https://doi.org/10.1016/j.gloenvcha.2015.03.008

Kalnay, E., & Cai, M. (2003). Impact of urbanization and land-use change on climate. *Nature*, *423*(6939), 528–531. https://doi.org/10.1038/nature01675

Li, X., & Zhou, Y. (2017). Urban mapping using DMSP/OLS stable night-time light: A review. *International Journal of Remote Sensing*, *0*(0), 1–17. https://doi.org/10.1080/01431161.2016.1274451

Li, Y., Zhu, X., Sun, X., & Wang, F. (2010). Landscape effects of environmental impact on bay-area wetlands under rapid urban expansion and development policy: A case study of Lianyungang, China. *Landscape and Urban Planning*, *94*(3–4), 218–227. https://doi.org/10.1016/j.landurbplan.2009.10.006

Liu, J., Li, J., Li, W., & Wu, J. (2016). Rethinking big data: A review on the data quality and usage issues. *ISPRS Journal of Photogrammetry and Remote Sensing*, *115*, 134–142. https://doi.org/10.1016/j.isprsjprs.2015.11.006

Liu, J., & Li, W. (2014). A nighttime light imagery estimation of ethnic disparity in economic well-being in mainland China and Taiwan (2001–2013). *Eurasian Geography and Economics*, *55*(6), 691–714. https://doi.org/10.1080/15387216.2015.1041147

Liu, Z., He, C., Zhang, Q., Huang, Q., & Yang, Y. (2012). Extracting the dynamics of urban expansion in China using DMSP-OLS nighttime light data from 1992 to 2008. *Landscape and Urban Planning*, *106*(1), 62–72. https://doi.org/10.1016/j.landurbplan.2012.02.013

Lloyd, S. P. (1982). Least squares quantization in PCM. *Information Theory, IEEE Transactions On*, *28*(2), 129–137.

Long, Y., & Wu, K. (2016). Shrinking cities in a rapidly urbanizing China. *Environment and Planning A*, *48*(2), 220–222. https://doi.org/10.1177/0308518X15621631

Lu, X., Bengtsson, L., & Holme, P. (2012). Predictability of population displacement after the 2010 Haiti earthquake. *Proceedings of the National Academy of Sciences*, *109*(29), 11576–11581. https://doi.org/10.1073/pnas.1203882109

Ma, Q., He, C., Wu, J., Liu, Z., Zhang, Q., & Sun, Z. (2014). Quantifying spatiotemporal patterns of urban impervious surfaces in China: An improved assessment using

nighttime light data. *Landscape and Urban Planning*, *130*, 36–49. https://doi.org/10.1016/j.landurbplan.2014.06.009

Ma, T., Zhou, C., Pei, T., Haynie, S., & Fan, J. (2012). Quantitative estimation of urbanization dynamics using time series of DMSP/OLS nighttime light data: A comparative case study from China's cities. *Remote Sensing of Environment*, *124*, 99–107. https://doi.org/10.1016/j.rse.2012.04.018

National Bureau of Statistics of the People's Republic of China. (2016). *China City Statistical Yearbook*. Retrieved from http://tongji.cnki.net/kns55/navi/HomePage.aspx?id=N2014050073&name=YZGCA&floor=1

National Geophysical Data Center. (2015, December 10). Version 4 DMSP-OLS Nighttime Lights Time Series. Retrieved December 10, 2015, from http://ngdc.noaa.gov/eog/gcv4_readme.txt

Patel, N. N., Angiuli, E., Gamba, P., Gaughan, A., Lisini, G., Stevens, F. R., … Trianni, G. (2015). Multitemporal settlement and population mapping from Landsat using Google Earth Engine. *International Journal of Applied Earth Observation and Geoinformation*, *35, Part B*, 199–208. https://doi.org/10.1016/j.jag.2014.09.005

Pohl, C., & Genderen, J. L. V. (1998). Review article Multisensor image fusion in remote sensing: Concepts, methods and applications. *International Journal of Remote Sensing*, *19*(5), 823–854. https://doi.org/10.1080/014311698215748

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Savitzky, Abraham., & Golay, M. J. E. (1964). Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, *36*(8), 1627–1639. https://doi.org/10.1021/ac60214a047

Schetke, S., & Haase, D. (2008). Multi-criteria assessment of socio-environmental aspects in shrinking cities. Experiences from eastern Germany. *Environmental Impact Assessment Review*, *28*(7), 483–503. https://doi.org/10.1016/j.eiar.2007.09.004

Schneider, A., & Woodcock, C. E. (2008). Compact, Dispersed, Fragmented, Extensive? A Comparison of Urban Growth in Twenty-five Global Cities using Remotely Sensed Data, Pattern Metrics and Census Information. *Urban Studies*, *45*(3), 659–692. https://doi.org/10.1177/0042098007087340

Scikit-learn. (2015, October 12). Selecting the number of clusters with silhouette analysis on KMeans clustering—Scikit-learn 0.17 documentation. Retrieved December 10, 2015, from http://scikit-learn.org/stable/auto_examples/cluster/plot_kmeans_silhouette_analysis.html

Shi, K., Huang, C., Yu, B., Yin, B., Huang, Y., & Wu, J. (2014). Evaluation of NPP-VIIRS night-time light composite data for extracting built-up urban areas. *Remote Sensing Letters*, *5*(4), 358–366. https://doi.org/10.1080/2150704X.2014.905728

Shi, K., Yu, B., Huang, Y., Hu, Y., Yin, B., Chen, Z., … Wu, J. (2014). Evaluating the Ability of NPP-VIIRS Nighttime Light Data to Estimate the Gross Domestic Product and the Electric Power Consumption of China at Multiple Scales: A Comparison with DMSP-OLS Data. *Remote Sensing*, *6*(2), 1705–1724. https://doi.org/10.3390/rs6021705

Sleeter, B. M., Sohl, T. L., Loveland, T. R., Auch, R. F., Acevedo, W., Drummond, M. A., … Stehman, S. V. (2013). Land-cover change in the conterminous United States from 1973 to 2000. *Global Environmental Change*, *23*(4), 733–748. https://doi.org/10.1016/j.gloenvcha.2013.03.006

Sorace, C., & Hurst, W. (2016). China's Phantom Urbanisation and the Pathology of Ghost Cities. *Journal of Contemporary Asia*, *46*(2), 304–322. https://doi.org/10.1080/00472336.2015.1115532

Stathakis, D., Tselios, V., & Faraslis, I. (2015). Urbanization in European regions based on night lights. *Remote Sensing Applications: Society and Environment*, *2*, 26–34. https://doi.org/10.1016/j.rsase.2015.10.001

Steinley, D. (2003). Local Optima in K-Means Clustering: What You Don't Know May Hurt You. *Psychological Methods*, *8*(3), 294–304. https://doi.org/10.1037/1082-989X.8.3.294

Su, S., Jiang, Z., Zhang, Q., & Zhang, Y. (2011). Transformation of agricultural landscapes under rapid urbanization: A threat to sustainability in Hang-Jia-Hu region, China. *Applied Geography*, *31*(2), 439–449. https://doi.org/10.1016/j.apgeog.2010.10.008

Tan, M. (2016). Use of an inside buffer method to extract the extent of urban areas from DMSP/OLS nighttime light data in North China. *GIScience & Remote Sensing*, *53*(4), 444–458. https://doi.org/10.1080/15481603.2016.1148832

Townsend, A. C., & Bruce, D. A. (2010). The use of night-time lights satellite imagery as a measure of Australia's regional electricity consumption and population distribution. *International Journal of Remote Sensing*, *31*(16), 4459–4480. https://doi.org/10.1080/01431160903261005

United Nations. (2014). World urbanization prospects: The 2014 revision. *United Nations, Department of Economic and Social Affairs (DESA), Population Division, New York*.

Wang, L., Li, C., Ying, Q., Cheng, X., Wang, X., Li, X., … Gong, P. (2012). China's urban expansion from 1990 to 2010 determined with satellite remote sensing. *Chinese Science Bulletin*, *57*(22), 2802–2812. https://doi.org/10.1007/s11434-012-5235-7

Wang, X.-R., Hui, E. C.-M., Choguill, C., & Jia, S.-H. (2015). The new urbanization policy in China: Which way forward? *Habitat International*, *47*, 279–284. https://doi.org/10.1016/j.habitatint.2015.02.001

Weng, Q. (2002). Land use change analysis in the Zhujiang Delta of China using satellite remote sensing, GIS and stochastic modelling. *Journal of Environmental Management*, *64*(3), 273–284. https://doi.org/10.1006/jema.2001.0509

Wiechmann, T., & Pallagst, K. M. (2012). Urban shrinkage in Germany and the USA: A Comparison of Transformation Patterns and Local Strategies. *International Journal of Urban and Regional Research*, *36*(2), 261–280. https://doi.org/10.1111/j.1468-2427.2011.01095.x

Woodworth, M. D. (2015). Ordos Municipality: A market-era resource boomtown. *Cities*, *43*, 115–132. https://doi.org/10.1016/j.cities.2014.11.017

Wu, Jianguo. (2008). Toward a Landscape Ecology of Cities: Beyond Buildings, Trees, and Urban Forests. In M. M. Carreiro, Y.-C. Song, & J. Wu (Eds.), *Ecology, Planning, and Management of Urban Forests* (pp. 10–28). https://doi.org/10.1007/978-0-387-71425-7_2

Wu, Jiansheng, He, S., Peng, J., Li, W., & Zhong, X. (2013). Intercalibration of DMSP-OLS night-time light data by the invariant region method. *International Journal of Remote Sensing*, *34*(20), 7356–7368. https://doi.org/10.1080/01431161.2013.820365

Wu, Jiansheng, Ma, L., Li, W., Peng, J., & Liu, H. (2014). Dynamics of Urban Density in China: Estimations Based on DMSP/OLS Nighttime Light Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *7*(10), 4266–4275. https://doi.org/10.1109/JSTARS.2014.2367131

Xiao, J., Shen, Y., Ge, J., Tateishi, R., Tang, C., Liang, Y., & Huang, Z. (2006). Evaluating urban expansion and land use change in Shijiazhuang, China, by using GIS and remote sensing. *Landscape and Urban Planning*, *75*(1–2), 69–80. https://doi.org/10.1016/j.landurbplan.2004.12.005

Yu, B., Shu, S., Liu, H., Song, W., Wu, J., Wang, L., & Chen, Z. (2014). Object-based spatial cluster analysis of urban landscape pattern using nighttime light satellite images: A case study of China. *International Journal of Geographical Information Science*, *28*(11), 2328–2355. https://doi.org/10.1080/13658816.2014.922186

Zhang, F., Hao, J., JIang, G., Ding, Z., Li, X., & Li, T. (2005). Spatial-temporal Variance of Urban Land Use Intensity [J]. *China Land Science*, *1*, 003.

Zhang, Qian, & Seto, K. C. (2013). Can Night-Time Light Data Identify Typologies of Urbanization? A Global Assessment of Successes and Failures. *Remote Sensing*, *5*(7), 3476–3494. https://doi.org/10.3390/rs5073476

Zhang, Qianwen, & Su, S. (2016). Determinants of urban expansion and their relative importance: A comparative analysis of 30 major metropolitans in China. *Habitat International*, *58*, 89–107. https://doi.org/10.1016/j.habitatint.2016.10.003

Zhang, Qingling, & Seto, K. C. (2011). Mapping urbanization dynamics at regional and global scales using multi-temporal DMSP/OLS nighttime light data. *Remote Sensing of Environment*, *115*(9), 2320–2329. https://doi.org/10.1016/j.rse.2011.04.032

Zhou, Y., Smith, S. J., Elvidge, C. D., Zhao, K., Thomson, A., & Imhoff, M. (2014). A cluster-based method to map urban area from DMSP/OLS nightlights. *Remote Sensing of Environment*, *147*, 173–185. https://doi.org/10.1016/j.rse.2014.03.004

Zhou, Y., Smith, S. J., Zhao, K., Imhoff, M., Thomson, A., Bond-Lamberty, B., … Elvidge, C. D. (2015). A global map of urban extent from nightlights. *Environmental Research Letters*, *10*(5), 054011. https://doi.org/10.1088/1748-9326/10/5/054011

# Chapter 3. Climate-related uncertainties in urban exposure to sea level rise and storm surge flooding: a multi-temporal and multi-scenario analysis[2]

**Abstract**: Climate change-induced sea level rise and intensified storms pose emerging flood threats to global coastal urban areas. While such threats have been mapped, their uncertainties from different climate scenarios and longer planning horizons have yet to be addressed from both an exposure assessment and a stakeholder outreach perspective. Therefore, we chose the highly urbanized San Francisco Bay Area as an example to project its flood areas every 20 years between 2000 and 2100, under 24 varied climate scenarios with two greenhouse gas (GHG) concentration levels. We then assessed flood exposure by intersecting the flood areas with demographic and socioeconomic distributions, developed areas, lifeline infrastructures, and emergency responders in low elevation (<10 m) coastal zones. Our median estimates under the low GHG scenarios indicated that 10-38% of the items assessed above are flood-exposed in 2000-2020, with this exposure increasing to 20-54% during 2080-2100. The median estimates under the high GHG scenarios for the same periods are 0-35% and 40-67%, respectively. The expected uncertainties, or standard deviations, of the exposures for a given item assessed above under the low and high GHG scenarios are 1-2% in 2000-2020 and 7-10% in 2080-2100. Despite our modelling capability for a range of climate scenarios over the long term, some stakeholders, particularly those in the private sector, prefer near-term results with lower uncertainties. This implies the need for coastal urban areas to cope with climate-related uncertainties and to focus on the long term when developing strategies and policies for climate change adaptation.
**Keywords**: Sea level rise, flood exposure, uncertainty, stakeholders, climate change

## 3.1 Introduction

The interaction between environmental hazards, urbanization, and climate change is likely to put more people and assets at risk. Environmental hazards affect urban areas and the well-being of their residents through storms, floods, heat waves, and drought (Hunt & Watkiss, 2010). At the same time, urbanization, particularly disorderly urbanization, is increasingly exposing urban areas to these hazards (Dawson et al., 2011; Pelling, 2003; Storch & Downes, 2011). This urban-hazard interaction alters under climate change which increases the extent, frequency, and severity of current-day environmental hazards, eventually affecting urban areas that were not previously exposed. To mitigate these undesirable consequences, cities have addressed environmental hazards and climate change issues in their master, strategic, and action plans (Jabareen, 2015). Researchers have also modeled the impact of climate change and environmental hazards on different population groups (Kaźmierczak & Cavan, 2011; Nutters, 2012; Martinich, Neumann, Ludwig, & Jantarasami, 2013; KC, Shepherd, & Gaither, 2015; Bickers, 2014), infrastructures (Biging, Radke, & Lee, 2012; Demirel, Kompil, & Nemry, 2015; Radke et al., 2018, 2017), and natural environments (Schile et al., 2014; Zhu, Xi, Hoctor, & Volk, 2015). The modeled impacts can also be used to facilitate public participation in adaptation planning (Wadey et al., 2015).

Coastal urban areas are likely to have intensified flood threats from storm surge and sea level rise (SLR) while simultaneously experiencing increasing development pressure, and thus

[2] This chapter is from the journal paper: Ju, Y., Lindbergh, S., He, Y., & Radke, J. D. (2019). Climate-related uncertainties in urban exposure to sea level rise and storm surge flooding: a multi-temporal and multi-scenario analysis. Cities, 92, 230–246. https://doi.org/10.1016/j.cities.2019.04.002

should be prioritized for climate change adaptation (Carter et al., 2015; Rosenzweig, Solecki, Hammer, & Mehrotra, 2010). Due to their access to water, maritime transportation, fertile soil, raw materials - including salt and sand, and tourist attractions, coastal zones have been hotspots for development and population growth (Creel, 2003; De Sherbinin, Schiller, & Pulsipher, 2007; UNU-IHDP, 2015). A 2007 assessment indicated that while low elevation (i.e. less than 10 m above sea level) coastal zones only covered 2% of the world's land area, they contained 10% of the world's population and 13% of the world's urban population (McGranahan, Balk, & Anderson, 2007). A global meta-analysis of 292 locations found that the average urban expansion rate in low elevation coastal zones between 1970 and 2000 was greater than 5.7%, higher than the rate of all other areas combined (Seto, Fragkias, Güneralp, & Reilly, 2011). Coastal zones are growth hotspots for the future: a conservative projection estimates that the global population in low elevation coastal zones will increase by 50% between 2000 and 2030, and by more than 160% between 2000 and 2060 (Neumann, Vafeidis, Zimmermann, & Nicholls, 2015). Such high concentration of population and assets explain why coastal urban areas are more vulnerable to flood hazards from SLR and storm surge than other areas (Hunt & Watkiss, 2010).

Studies have modeled flooding from SLR and storm surge and analyzed the exposure of coastal areas (Barnard et al., 2009, 2014; Biging et al., 2012; Dasgupta, Laplante, Meisner, Wheeler, & Yan, 2008; Knowles, 2009, 2010; Lang, Radke, Chen, & Chan, 2016; Marcy et al., 2011; Radke et al., 2017; Schile et al., 2014; Strauss, Ziemlinski, Weiss, & Overpeck, 2012). A typical approach in these studies is to produce and utilize flood maps under different storm recurrence intervals (e.g. 10-year and 100-year) combined with generic, incremental SLR values that are not closely associated with specific climate scenarios and planning horizons. While this generic approach provides the flexibility to adapt itself to various and continuously updating climate projections, it may lack the specificity to show what the flooding is like under specific climate scenarios and planning horizons that stakeholders are interested in, as well as the uncertainty introduced by different climate scenarios and how the uncertainty propagates over time.

To show the uncertainty in long-term flood projections under climate change, and to discuss the implications of such uncertainty in planning and management for climate change adaptation, we conducted flood simulations that incorporated SLR and storm surge to identify a coastal urban area's flood exposures under a range of 24 different climate scenarios at two greenhouse gas (GHG) concentration levels, every 20 years between 2000 and 2100. We measured flood exposures by intersecting the simulated flood with demographic and socioeconomic distributions, developed areas, lifeline infrastructures, and emergency responders. Uncertainties were further calculated as standard deviations of those flood exposures.

We focus on the San Francisco Bay Area (Bay Area), a major U.S. urban area in the State of California, in this study. The Bay Area is a prime candidate for this research as it is prone to flooding from SLR and intensified storms induced by climate change, and it has a large concentration of development in threatened low-lying coastal areas (Figure 3-1). The Bay Area contains major U.S. urban agglomerations including San Francisco, Oakland, and the Silicon Valley complex centered in San Jose. This area continues to rapidly expand due to a growing economy and employment opportunities. Population, jobs, households and housing units in the Bay Area are projected to increase by 24-30% between 2010 and 2040, introducing new demands for land development (The Association of Bay Area Governments (ABAG), Metropolitan Transportation Commission (MTC), 2013). Due to the demand for land

development and insufficient regulations, the Bay Area's coastal zones have been transforming rapidly from their natural state into industrial and residential use located at or below sea level (Pinto & Kondolf, 2016). As a result, the Bay Area's exposures to potential flooding from storm surge and SLR have increased. A 2012 study shows that a 1 m SLR combined with a 100-year storm will put 220,000 people at risk and cost 49 billion US dollars to replace the impacted property, and that a 1.4 m SLR would increase the numbers to 270,000 and 62 billion US dollars (Heberger, Cooley, Moore, & Herrera, 2012).

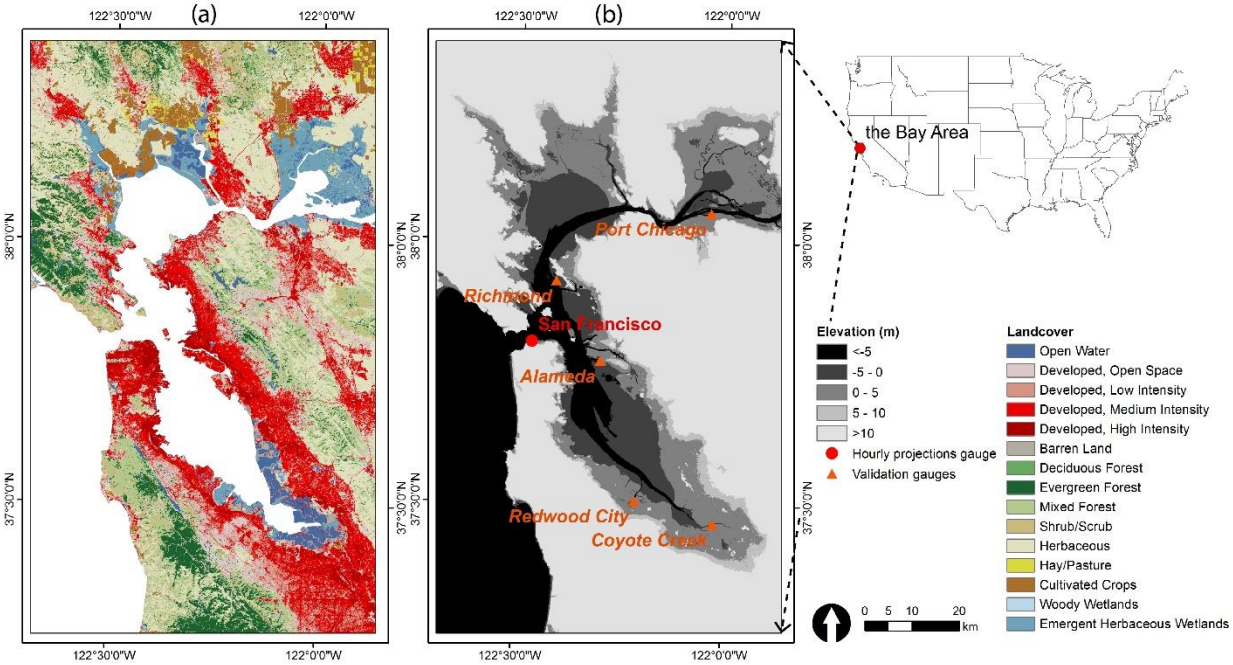



Figure 3-1. The San Francisco Bay Area. (a) shows the Bay Area's landcover types in 2011, and (b) shows its elevation with areas below 10 m highlighted in darker tones, the gauge with the hourly sea level projections, and the validation gauges used for flood model calibration.

Taken together, our primary goal is to illustrate the uncertainty in long-term projection of urban exposure to climate change-induced flooding. To achieve this goal, we first provide comprehensive mapping of flood hazard and exposures of urban areas under different climate scenarios and time horizons. Second, we highlight how the exposure's uncertainties change with time and vary by groups of the climate scenarios defined by their GHG levels. In addition, we discuss the uncertainties' implications in planning and management based on our outreach process with stakeholders. Through these efforts, we highlight the necessity of understanding climate-related uncertainties in coastal flood projections and exposure analysis. We also reveal some challenges for stakeholders in utilizing multi-scenario and long-term climate and environmental hazard projections in planning, management, and decision-making. While we focus on the Bay Area and coastal flooding, our findings are informative for other coastal urban areas faced with similar flood threats, and for adaptation to other climate change-induced environmental hazards (e.g. drought and heatwave) whose uncertainties are intrinsic and substantially larger in the longer term.

## 3.2 Data and Methods

### 3.2.1 Flooding model

Various models, both static and dynamic, have been used to simulate coastal flood. Earlier models tend to be static, identifying flood areas below a time-invariant water surface such as the projected mean sea level. These static models are computationally feasible, therefore can be applied to large areas with fine spatial resolutions (Biging et al., 2012; Dasgupta et al., 2008; Marcy et al., 2011; Strauss et al., 2012). However, these models do not incorporate the temporal and flow dynamics of water movement. More recent research has employed process-based, hydrodynamic, two-dimensional (2D) models to account for these dynamics. These 2D models are computationally intensive, and thus difficult to implement at regional scales and/or with fine spatial resolutions. Few studies, including the model by Knowles (2009, 2010), the Coastal Storm Modeling System (CoSMoS) (Barnard et al., 2009, 2014), and the CalFloD-3D model (Radke et al., 2017), adopted 2D hydrodynamic models for the Bay Area.

We employed a 3Di hydrodynamic model (Stelling, 2012) to simulate flood areas and depth at 50 m spatial resolution during extreme sea level events projected under a range of 24 climate scenarios defined by two greenhouse gas (GHG) concentration levels, every 20 years between 2000 and 2100. The 3Di model dynamically simulates the movement of water, by user-defined time steps, over a digital surface of topography and bathymetry. This model has been successfully implemented in urban and rural watersheds to simulate flood by inland rainfall-runoff (Dahm, Hsu, Lien, Chang, & Prinsen, 2014; Hsu, Prinsen, Bouaziz, Lin, & Dahm, 2016; Leicher, 2016), SLR, and storm surge (Ju et al., 2017; Radke et al., 2017). The 3Di model's primary inputs are time-series water levels as boundary forcing to generate waterflows, and digital surface data containing topography, bathymetry, and/or aboveground objects such as levees to direct the waterflows. For every extreme sea level event, the 3Di model produces a time-series of flood areas with water depths. A unique advantage of the 3Di model for this study is its ability to feasibly compute over large regions at fine spatial resolutions, which is enabled by the model's compression algorithm that simplifies the input digital surface while preserving significant topographic variations such as those from levees and buildings (Stelling, 2012). In this study, we modeled at 50 m spatial resolution to feasibly iterate through the various extreme sea level events.

### 3.2.2 Extreme sea level events

We used extreme sea level events to simulate the worst-case flood hazards with the 3Di model. These extreme sea level events were extracted from an hourly sea level projection (Cayan, Kalansky, Iacobellis, & Pierce, 2016) at a San Francisco gauge (Figure 3-1) and incorporated long-term SLR and short-term fluctuations from tide and storm surge. Each event is a 72-hour window starting with the highest sea level under a given climate scenario during a 20-year period between 2000 and 2100 (Figure 3-2 (a)). The events together represent a spectrum of 24 climate scenarios generated from a hierarchy of two Representative Concentration Pathways (RCPs) defining low and high atmospheric concentrations of GHG (i.e. RCP 4.5 versus 8.5), three probabilistic SLR values, and four Global Climate Models (GCMs) emphasizing different historical and future climate patterns in California (Pierce, Cayan, & Dehann, 2016) (Figure 3-2 (b)). For each RCP and a 20-year period, we ranked the corresponding twelve extreme sea level events by their peak sea levels to derive high, medium, and low estimates of flood hazard. We used the 20-year intervals to match the simulated flood and exposure analysis with typical

planning horizons and investment cycles that were often multidecadal. It is worth noting that the climate scenarios, our flood projections, and exposure analysis are intended to show a wide range of plausible futures and not to predict an exact one (Moss et al., 2010).

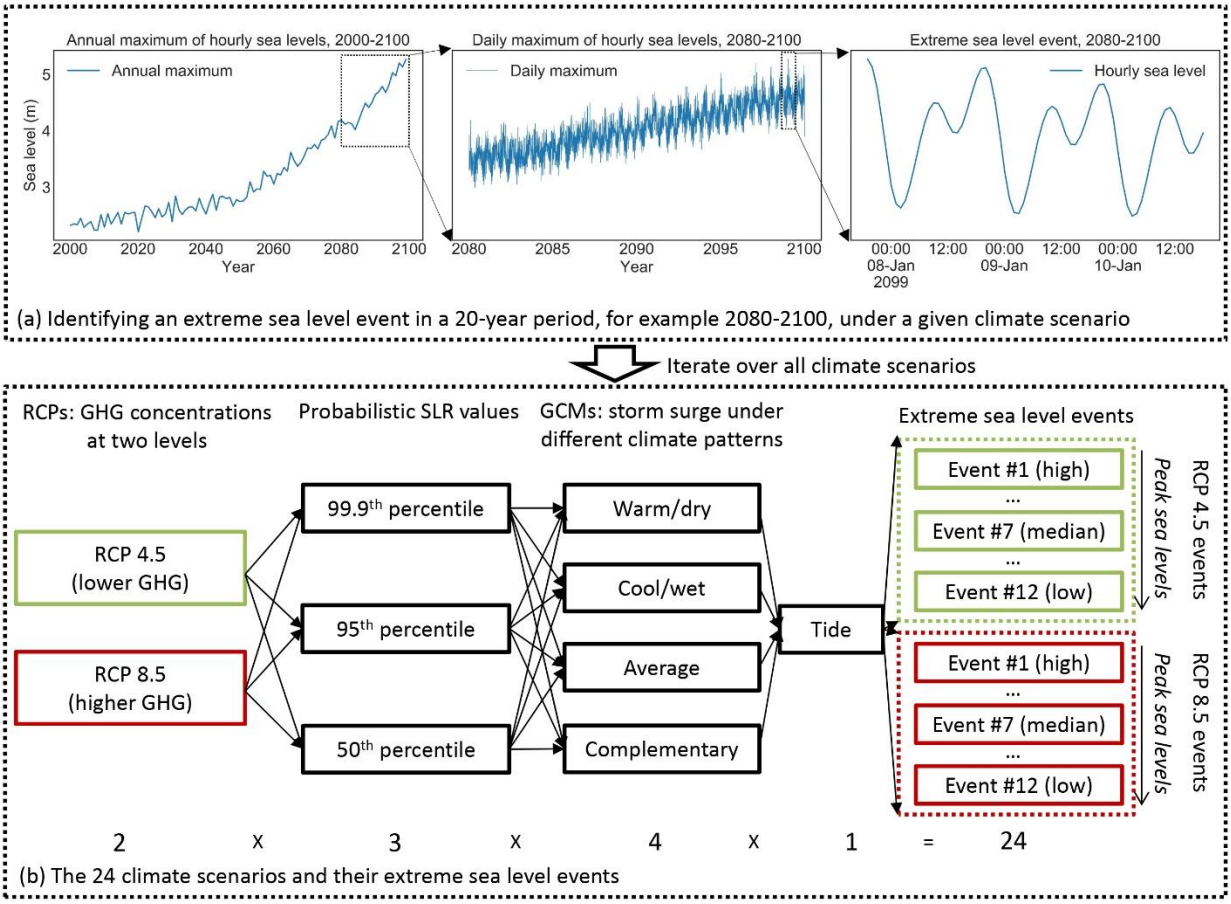



Figure 3-2. The climate scenarios and their extreme sea level events. (a) shows an example of extracting an extreme sea level event for a 20-year period and a given climate scenario from the hourly sea level projections. This extreme sea level event is a 72-hour period started with the highest sea level projected under this climate scenario and during this 20-year period. (b) shows the hierarchy to generate the climate scenarios and ranking of the extreme sea level events by their peak sea levels.

The extreme sea level events show higher and increasingly varied peak sea levels as the events move into the future (Figure 3-3). Under RCP 4.5 scenarios, the peak sea level reaches 2.61-2.94 m during the 2000-2020 period, and 3.08-4.40 m during the 2080-2100 period. Under RCP 8.5 scenarios, the numbers shift to 2.49-2.63 m during the 2000-2020 period, and 3.78-5.50 m during the 2080-2100 period. The peak sea levels under RCP 8.5 tend to be higher than those under RCP 4.5, which is likely due to RCP 8.5's higher GHG concentrations, stronger climate change, greater SLR, and more intensified storms (see Appendix A for a visualization of the hourly sea level projections, SLR, and storm surge). However, during the 2000-2020 period, RCP 4.5 scenarios project higher peak sea levels than RCP 8.5 scenarios, which is due to stronger storm surge projections by one GCM under RCP 4.5 scenarios (Appendix A). The variability of the peak sea levels exacerbates over time, for example, increasing from 0.14 m (i.e.

2.63 m versus 2.49 m) in the 2000-2020 period to 1.72 m (5.50 m versus 3.78 m) in the 2080-2100 period under RCP 8.5.

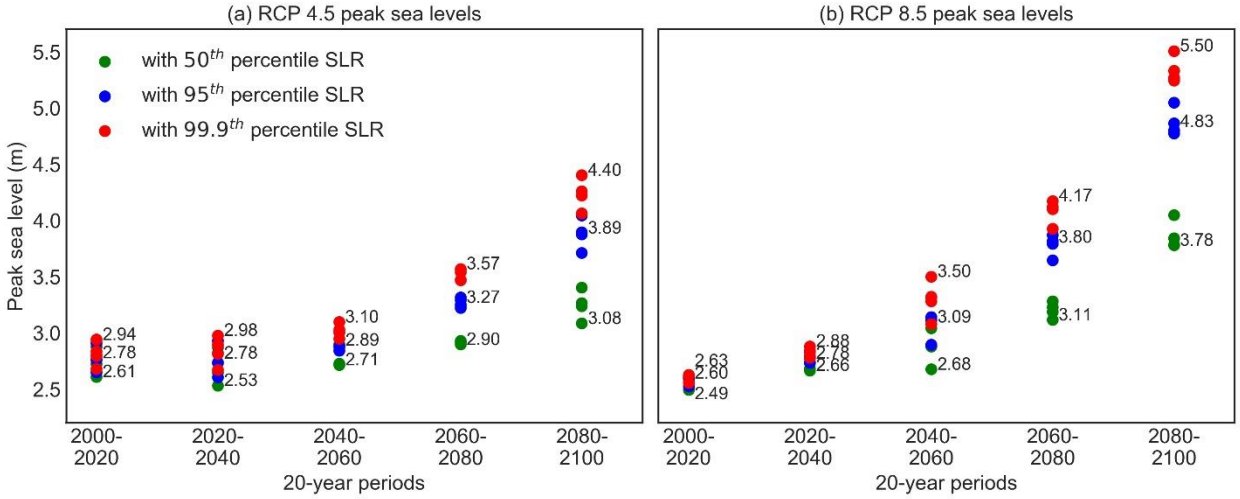



Figure 3-3. Projected peak sea levels during the extreme sea level events at the San Francisco gauge. Each dot represents the peak sea level under a climate scenario generated from permutations of two RCPs, three probablistic SLRs, and four GCMs during a 20-year period. The maximum, median, and minimum estimates of peak sea levels for an RCP and a 20-year period are labeled. The colors differentiate the three probabilistic SLRs generating the peak sea levels. (a) shows the peak sea levels under RCP 4.5 scenarios, and (b) shows the peak sea levels under RCP 8.5 scenarios.

### 3.2.3 Topography and bathymetry data

We generated a continuous, 50 m resolution surface from the best-available topography and bathymetry datasets (Table 3-1) of the Bay Area as the second input for the 3Di model. Our surface contains bare ground elevation, bathymetry, and levees. We excluded buildings as they were too granular for the 50 m resolution surface. Since the datasets are originally all finer than the targeted 50 m spatial resolution, we first conducted an average aggregation and then mosaicked the aggregated datasets to generate the 50 m resolution surface. Flood control structures, such as levees, have their elevation information preserved to some extent in the 50 m resolution surface. However, underestimated elevations are introduced when averaging levee segments with their lower elevation surroundings. Similarly, small channels protected by extensive levees can have overestimated elevations after the aggregation. Such underestimation for the levees and overestimation for the small channels can either generate non-existent water flow pathways or eliminate existing ones. Therefore, the 50 m resolution surface can over/or underestimate flooding. However, we used the average aggregation (versus maximum or minimum aggregation) as it showed an average condition of the topography. Interested researchers may further test the effects of other aggregation methods on the simulated floods.

Table 3-1. Topography and bathymetry datasets used to construct the 50 m surface

| Data | Spatial resolution | Spatial coverage | Use in the 50 m surface | Source |
|---|---|---|---|---|

| | | | | |
|---|---|---|---|---|
| USGS South Bay Lidar | 1 m | Low elevation coastal zones | Fine resolution topography | USGS California Coastal LiDAR Project (https://coast.noaa.gov/htdata/lidar1_z/geoid12a/data/1406/) |
| NOAA North Bay Lidar | | | | NOAA California Coastal LiDAR Project (https://coast.noaa.gov/htdata/lidar1_z/geoid12a/data/584/) |
| National Elevation Dataset (NED) | 10 m | The entire Bay Area | Supplementary topography to finer resolution datasets | USGS National Elevation Dataset (https://lta.cr.usgs.gov/NED) |
| DWR Bay-Delta bathymetry | 2-10 m | The entire Bay Area | Fine resolution bathymetry | California Department of Water Resources (https://www.sciencebase.gov/catalog/item/58599681e4b01224f329b484) |

### 3.2.4 Model validation and representation of flood hazard

We modeled a historical high sea level event on Jan 11$^{th}$, 2017 in the Bay Area (NOAA, 2019) to validate the model's settings before iteratively simulating through the extreme sea level events identified in section 3.2.2. We compared the simulated water levels against observations at five validation gauges in Figure 3-1. The results (Figure 3-4) show that the model simulates similar water levels to the historical event, with Person's correlation (equation (1)) between the simulations and the observations no less than 0.81 and root mean square error (RMSE, calculated with equation (2)) no greater than 0.69 m, when no time lag between the simulations and observations is applied to calculating correlation and RMSE. The difference between the observations and the simulations increases when moving inland, with it being smaller for the near coast Alameda gauge (r = 0.92, RMSE = 0.31 m) and greater for inland gauges such as Port Chicago (r = 0.81, RMSE=0.57 m) and Coyote Creek (r = 0.82, RMSE = 0.69 m). Overall, the simulations tend to underestimate the observations, which can be explained by the absence of river discharges that affect water levels in the simulations. We also notice time lags between the simulations and the observations, as a certain amount of time is required for the simulated waterflows to pass through the gauges and to alter the initial water surface of 0 m in elevation. When an arbitrary 30-min lag was added to comparing the observations and the simulations, the differences reduced for all gauges, and the lowest correlation increased from 0.81 (without lag) to 0.90 (with lag) and the highest RMSE reduced from 0.69 m (without lag) to 0.54 m (with lag). The existence of such time lags justifies the need for simulating the extreme sea level events over longer time windows (e.g. 72 hours), so that the simulated water peaks can pass through the study area.

$$r = \frac{\sum_{t=1}^{n}(x_t - \overline{X})(\hat{x}_t - \overline{\hat{X}})}{\sqrt{\sum_{t=1}^{n}(x_t - \overline{X})^2}\sqrt{\sum_{t=1}^{n}(\hat{x}_t - \overline{\hat{X}})^2}} \quad (1)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{n}(\hat{x}_t - x_t)^2}{n}} \quad (2)$$

where $x_t$ is the observed value at time step t, and $\hat{x}_t$ is the simulated value, $\overline{X}$ is the average value of the observations, $\overline{\hat{X}}$ is the average value of the simulations, n is the total number of time steps.

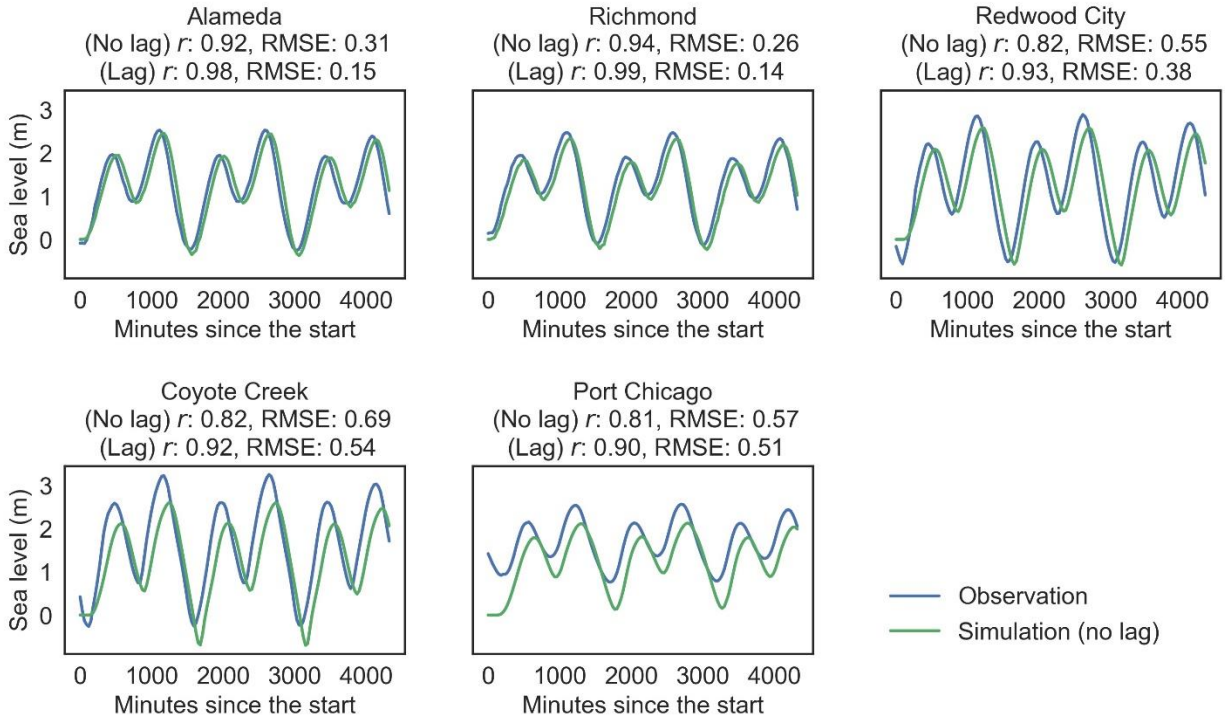



Figure 3-4. Observed and simulated water levels during a historical extreme sea level event in the Bay Area starting in Jan 11th, 2017. Observations and simulations were compared using correlation and root mean square error (RMSE), with and without an arbitrary 30-min time lag.

During each extreme sea level event our simulation produced hourly flood area and depths, which were compressed into a single map of the maximum flood extent and depth during the event. We also grouped the maps from different events by their respective RCPs and 20-year periods to show coastal flood hazard under different GHG concentration levels and over time.

### 3.2.5 Exposure analysis

We collected fourteen datasets representing demographic and socioeconomic distributions, developed areas, locations of lifeline infrastructures, and emergency responders (Table 3-2) to comprehensively assess the Bay Area's exposure to the projected flood. Demographic and socioeconomic data including population, number of households and housing

units were used to estimate the total amount of people, families, and housing structures exposed to flooding. Similarly, developed areas of different development intensities were used to approximate the exposure of the built-environment (i.e. man-made environment ranging from dense urban centers and parks) in general (Wood, 2009). We included lifeline infrastructures containing roads and co-located utilities (e.g. electric power and natural gas) as they support the recovery of places directly hit by environmental hazards, and the failure of these infrastructures can spread the hazards' impacts to a broader region (EERI, 2016; Oh Eun Ho, Deshmukh Abhijeet, & Hastak Makarand, 2013). Emergency responders, such as fire stations and hospitals, were also included in the exposure analysis due to their importance in disaster response and their abilities to assist needed citizens affected by environmental hazards (Biging et al., 2012; Coles, Yu, Wilby, Green, & Herring, 2017).

We used relative exposure, which was calculated as the percentage of a dataset's low-lying portion (i.e. less than 10 m in elevation) that was flooded, in the following analysis. Calculating exposure is a preliminary step towards understanding flooding's impact, as this measurement does not show outcomes such as monetary loss from flooded homes and infrastructures, reduced service area and capacity of emergency responders, or any cascading effects due to disruptions in lifeline infrastructures. However, this exposure metric can be easily iterated through the multiple extreme sea level events to provide an overview of the uncertainties driven by the climate scenarios and time. In addition, we focused on relative exposure to facilitate comparison across different datasets. We also limited the analysis to the low-lying portion of these datasets as coastal flooding from SLR and storm surge is a localized phenomenon that rarely expands to higher elevations. A detailed description of what relative exposure means to each dataset is contained in Table 3-2. In addition, we report absolute exposures of the datasets in Appendix B.

Table 3-2. Datasets used in the exposure analysis

| Datasets | | Source | Relative exposure metric |
|---|---|---|---|
| Demography and socioeconomics | Population | The U.S. Census Grids (http://sedac.ciesin.columbia.edu/data/collection/usgrid) | % people |
| | Number of households | | % households |
| | Number of housing units | | % housing units |
| Developed areas | Open space | National Land Cover Database (https://www.mrlc.gov/nlcd11_data.php) | % area |
| | Low intensity development | | |
| | Medium intensity development | | |
| | High intensity development | | |
| Lifeline infrastructures | Roads | ArcGIS Business Analyst 2016 (ESRI, 2016) | % length |
| | Electric transmission line | GIS Open Data, California Energy Commission (https://cecgis-caenergy.opendata.arcgis.com) | % length |
| | Electric substations | | % substations |
| | Natural gas pipelines | | % length |
| | Natural gas stations | | % stations |
| | Fire stations | (Biging et al., 2012) | % stations |

| | | | |
|---|---|---|---|
| Emergency responders | Hospitals | Licensed Healthcare Facility Listing, California Health and Human Services Open Data Portal (https://data.chhs.ca.gov/dataset/licensed-healthcare-facility-listing) | % hospitals |

%: percentage of a dataset's low-lying portion (i.e. less than 10 m in elevation) that is flooded

### 3.2.6 Analyzing uncertainties in the exposures

We calculated uncertainties in the exposures as standard deviation ($Std_{j,t,k}$) of a dataset's ($k$) relative exposures during all extreme sea level events for a given 20-year period ($t$) and an RCP ($j$) (equation (3)). We performed a simple linear regression to understand how these uncertainties changed with the RCPs and time (equation (4)). The fitted value from this regression represents expected uncertainty in relative exposures of a dataset for a RCP and a 20-year period. We calculated the uncertainties and performed the regression above for the overall exposure with the maximum flood depth greater than 0 m, as well as for five different levels of exposure stratified by the maximum flood depth: low (0-0.5 m), medium (0.5-1.0 m), high (1.0-1.5 m), very high (1.5-2.0 m), and extreme (>2.0 m).

$$Std_{j,t,k} = \sqrt{\frac{\sum_{i=1}^{N}(E_{i,j,t,k} - \overline{\boldsymbol{E_{j,t,k}}})^2}{N}} \quad (3)$$

$$Std_{j,t,k} = \beta_0 + \beta_1 RCP_j * Time_t + \varepsilon_{j,t,k} \quad (4)$$

where $E_{i,j,t,k}$ is a dataset $k$'s relative exposure during an extreme sea level event $i$ of RCP $j$ and a 20-year period $t$. $\overline{\boldsymbol{E_{j,t,k}}}$ is the average relative exposure across the extreme sea level events. $N$ is the total number of extreme sea level events of RCP $j$ and 20-year period $t$, which equals to 12 based on permutations of the four GCMs and three probabilistic SLR values. $Std_{j,t,k}$ is standard deviation of the dataset's relative exposures under RCP $j$ and during 20-year period $t$. $RCP_j$ is a dummy variable for either RCP 4.5 or RCP 8.5, and $Time_t$ is a dummy variable for one of the five 20-year periods.

## 3.3 Results

### 3.3.1 Flood hazard

Our simulations of the extreme sea level events indicate that flood area and depth increase with the projected SLR and intensified storms. The median estimates of RCP 8.5 scenarios and their respective extreme sea level events show that flood area increases from 733 km$^2$ during the 2000-2020 period, to 784 km$^2$ during the 2020-2040 period, and to 1,066 km$^2$ during the 2080-2100 period (Figure 3-5(a), solid red line). Compared with the median estimates of RCP 8.5, the ones of RCP 4.5 scenarios results in larger flood areas during the 2000-2020 period, and smaller flood areas during the remaining periods (Figure 3-5(a), solid green line). Such patterns are expected as RCP 4.5 scenarios have higher peak sea levels for the events during the 2000-2020 period but lower peak sea levels for the events during the other periods, as shown in section 3.2.2. Our results also indicate that the increased portions of the flood areas have greater depth (Figure 3-5(b-c)). For example, the median estimates of RCP 8.5 scenarios

showed 13% (97 km$^2$) of the total flood area with extreme depth (i.e. > 2.0 m) during the 2000-2020 period, 21% (165 km$^2$) during the 2020-2040 period, and 77% (822 km$^2$) during the 2080-2100 period (Figure 3-5(c)).

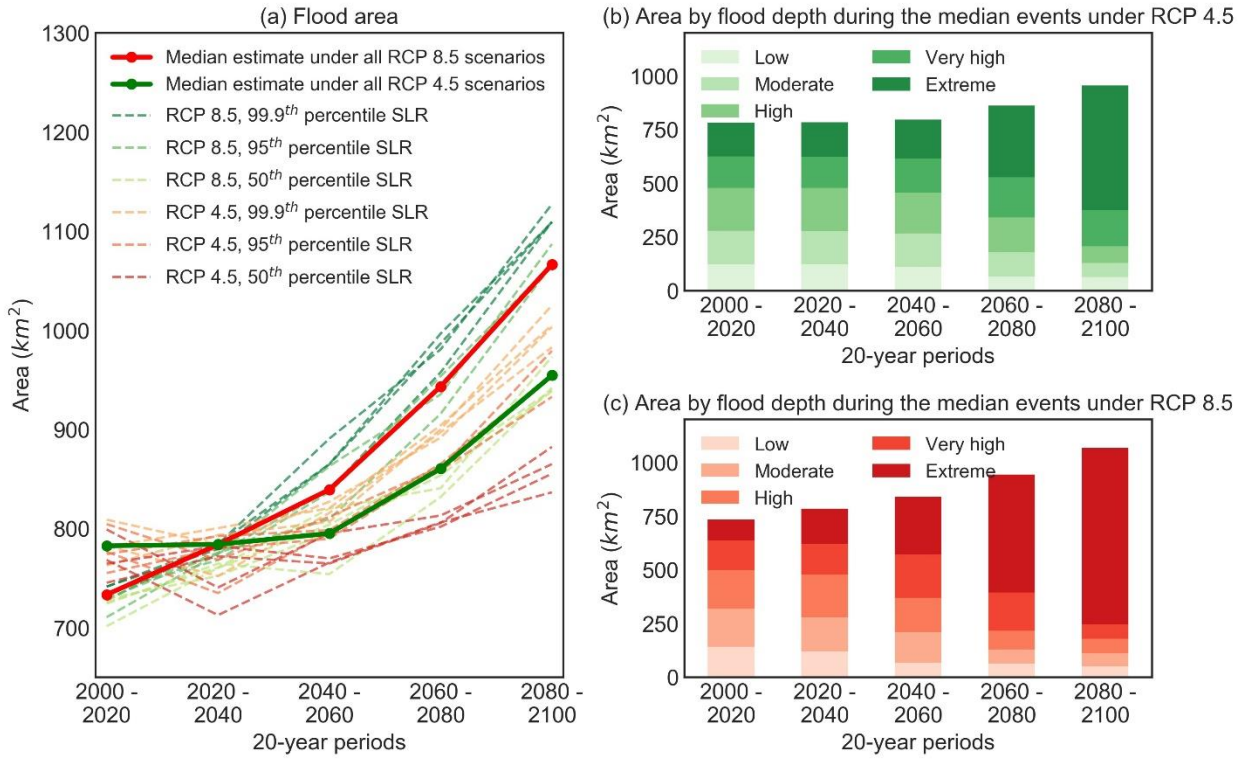



Figure 3-5. Flood area during the extreme sea level events identified every 20 years between 2000 and 2100 under the 24 climate scenarios. (a) Flood area of each event, shown as dashed lines colored by their RCPs and probabilistic SLR values. The median estimate of each RCP is shown as a solid line. (b) and (c) show area with different maximum flood depths (i.e. low: 0 - 0.5 m, moderate: 0.5 – 1.0 m, high: 1.0 – 1.5 m, very high: 1.5 – 2.0 m, extreme: > 2.0 m) based on the median estimates of each RCP.

The flood areas mainly include wetlands in the north and south, and developed areas in the central section of the Bay Area. While flooding over the wetlands moves further inland, in the developed areas it is constrained by steep slopes and hills (Figure 3-6 (a)). Due to these topographic constraints, the developed areas are likely to have limited expansions in flood area but greater increases in flood depth, which partially explains why areas with greater flood depth grew disproportionally in Figure 3-5 (b-c). We also visually identify several locations that will likely have disproportional increases in areas with deeper flooding in Figure 3-6 (b).

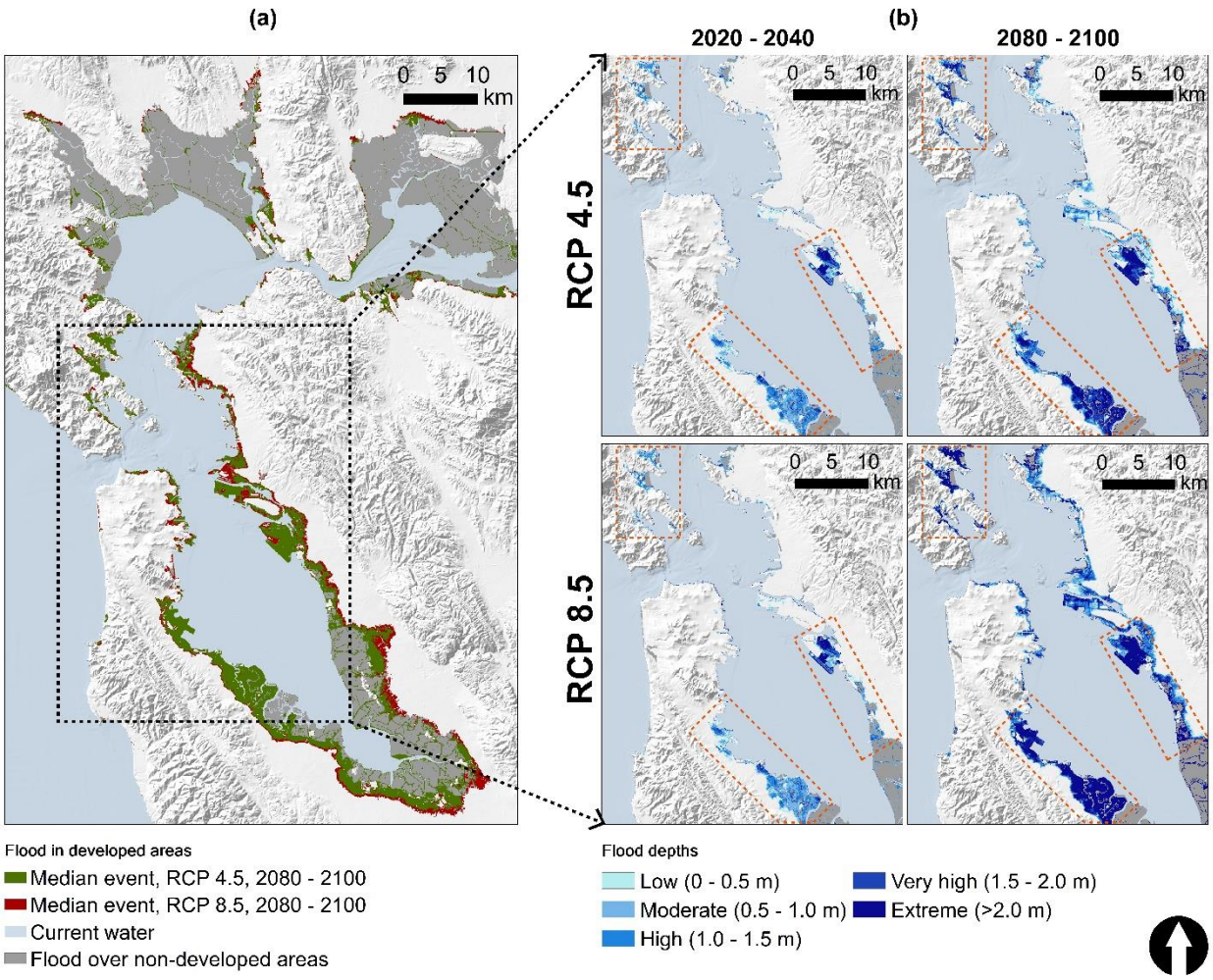



Figure 3-6 Flood in developed areas based on the median estimates of RCP 4.5 and 8.5 scenarios. (a) shows overall flood area during the 2080–2100 period. (b) compares flood depths from the median estimates of the two RCPs and two 20-year periods. (b) also outlines areas with limited expansion in flood area but disproportional increases in flood depth.

### 3.3.2 Urban exposure to flooding

Our results indicate that the fourteen datasets assessed are increasingly exposed to the simulated flooding and that RCP 8.5 scenarios generally cause greater exposures than do the RCP 4.5 scenarios (Figure 3-7). While we report relative exposures (i.e. percentage of a dataset's low-lying portion that is flooded, see Table 3-2 for a detailed description for each dataset) in this study, we include absolute exposures in Appendix B.

Under the median estimates of RCP 8.5 scenarios, the fourteen datasets' relative exposures increase from 0-35% (i.e. dataset dependent, similar below) during the 2000-2020 period to 40-67% during the 2080-2100 period. Under the median estimates of RCP 4.5 scenarios, the exposures are 10-38% during of the 2000-2020 period, and 20-54% during the 2080-2100 period. Compared with RCP 4.5, RCP 8.5 does not consistently lead to more exposures of the datasets. Over the five 20-year periods combined, the median estimates of RCP 8.5 scenarios result in significantly higher (p-value = 0.038) exposures of the fourteen datasets.

However, before 2040, RCP 4.5's median estimates lead to insignificantly higher exposures (p-value$_{2000\text{-}2020}$ = 0.394, p-value$_{2020\text{-}2040}$= 0.916). After 2040, RCP 8.5 scenarios' median estimates start to cause more exposures, and such pattern becomes increasingly significant (p-value$_{2040\text{-}2060}$ = 0.368, p-value$_{2060\text{-}2080}$= 0.036, p-value$_{2080\text{-}2100}$ = 0.000). This pattern is due to the differences in projected peak sea levels between the two RCPs, as illustrated earlier in section 3.2.2.

Developed open space and low intensity development are the most exposed categories on average over the 20-year periods, whereas hospitals, demographic and socioeconomic distributions are less exposed. Averaged over the median estimates during the five 20-year periods, 43% (RCP 4.5) to 47% (RCP 8.5) of developed open space and 41% (RCP 4.5) to 44% (RCP 8.5) of low intensity development within the low-lying coastal zones are exposed to flooding. The less exposed categories, such as hospitals and housing units, have exposures of 12% (RCP 4.5) to 18% (RCP 8.5) and 19% (RCP 4.5) to 22% (RCP 8.5), respectively. Development is the most exposed category we assess, followed by lifeline infrastructures, demographic and socioeconomic distributions, and emergency responders.

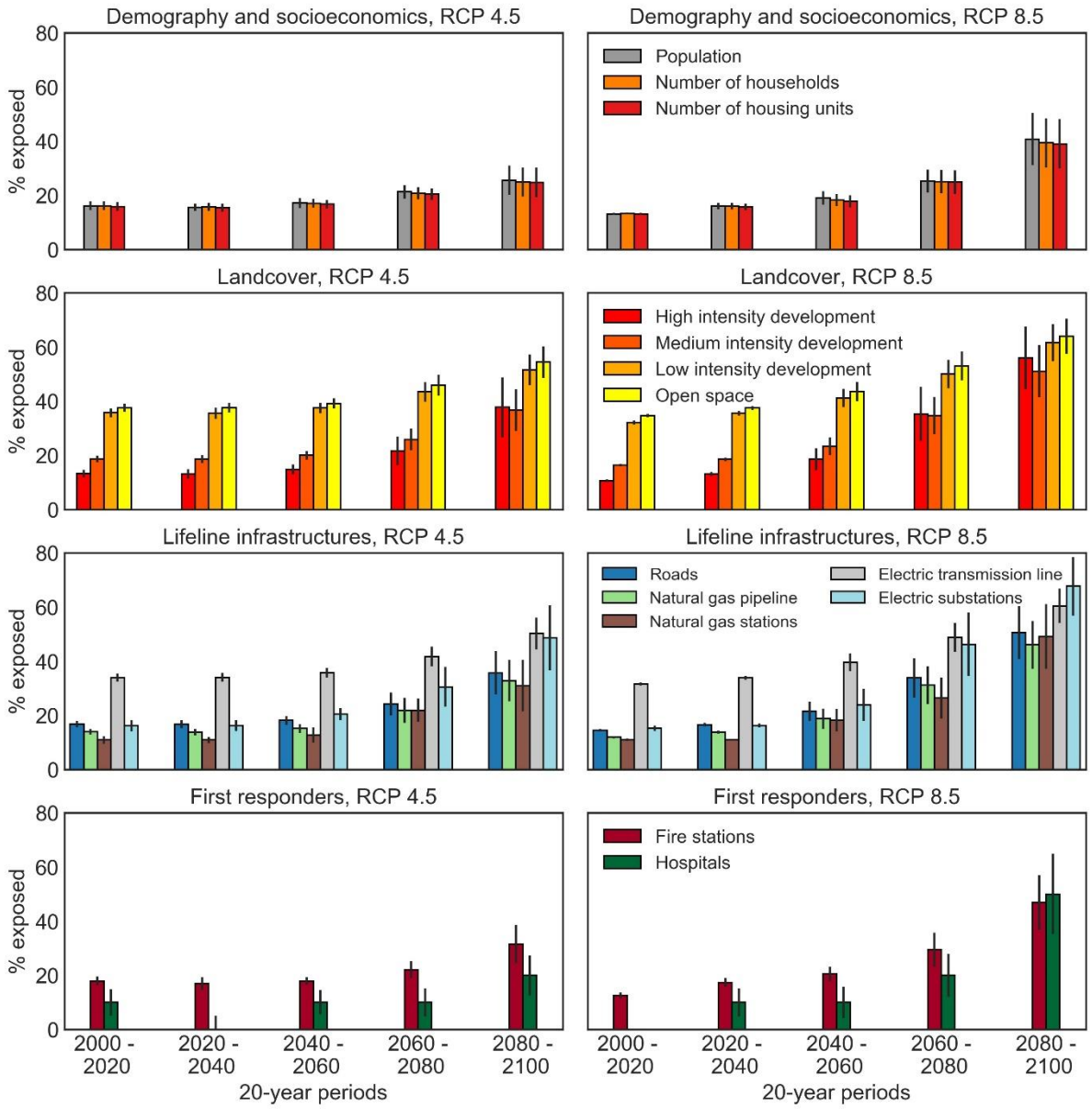



Figure 3-7. Relative flood exposure of demographic and socioeconomic distributions, developed areas, lifeline infrastructures, and first responders during the extreme sea level events under RCP 4.5 and 8.5 scenarios, every 20 years between 2000 and 2010. Relative exposure is calculated as percentage of a dataset's low-lying portion (i.e. less than 10 m in elevation) that is flooded. The solid color bars show the relative exposures during median estimates of the extreme sea level events. The black lines show standard deviations of the relative exposures during all extreme sea level events for a given RCP and a 20-year period.

### 3.3.3 Uncertainties driven by the climate scenarios and time

The uncertainties (i.e. standard deviations) of the datasets' relative exposures increase over time, and RCP 8.5 generally leads to more uncertainties compared with RCP 4.5 (Figure 3-8). Under RCP 4.5 scenarios, the 2000-2020 period has an expected uncertainty (i.e. based on the regression of equation 4) of 2% in relative exposure of a given dataset, and the uncertainty is

relatively unchanged until the 2060-2080 period when the value doubles to 4%. The final 2080-2100 period has the highest expected uncertainty of 7% (Figure 3-8 (a)). A similar trend is found for RCP 8.5's average uncertainties, despite an earlier rise in the 2020-2040 period (Figure 3-8 (b)). During the 2000-2020 and 2020-2040 periods, RCP 8.5 has smaller expected uncertainties than those of RCP 4.5, but this relationship inverts during the remaining periods.

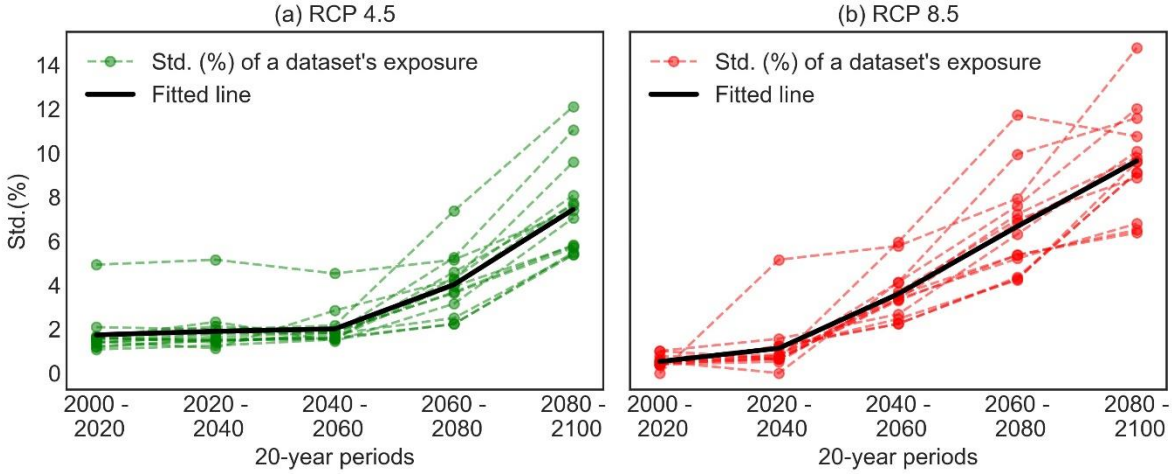



Figure 3-8. Temporal trends of the uncertainties in the relative flood exposures of the fourteen datasets. (a) shows the uncertainties under RCP 4.5 scenarios. (b) shows the uncertainties under RCP 8.5 scenarios. The color dashed line shows the uncertainties in relative exposures of a dataset under a given RCP over time, whereas the solid black line shows the expected uncertainties. For a dataset, uncertainty is the standard deviation (Std.) of its relative exposures under a RCP and during a 20-year period. Expected uncertainty is predicted with the regression of equation 4.

The expected uncertainties in the dataset's relative exposures to different flood depths show varied temporal trends (i.e. based on the regression of equation 4) (Figure 3-9). Low to very high exposures (i.e. the max. flood depth less than 2 m) show similar temporal trends with moderate increases in their uncertainties over time. However, extreme exposure (i.e. the max. flood depth greater than 2 m) has more distinct temporal trends with much larger uncertainties, particularly during the later periods such as those after 2080 for RCP 4.5 scenarios (Figure 3-9 (a)) and after 2060 for RCP 8.5 scenarios (Figure 3-9 (b)).

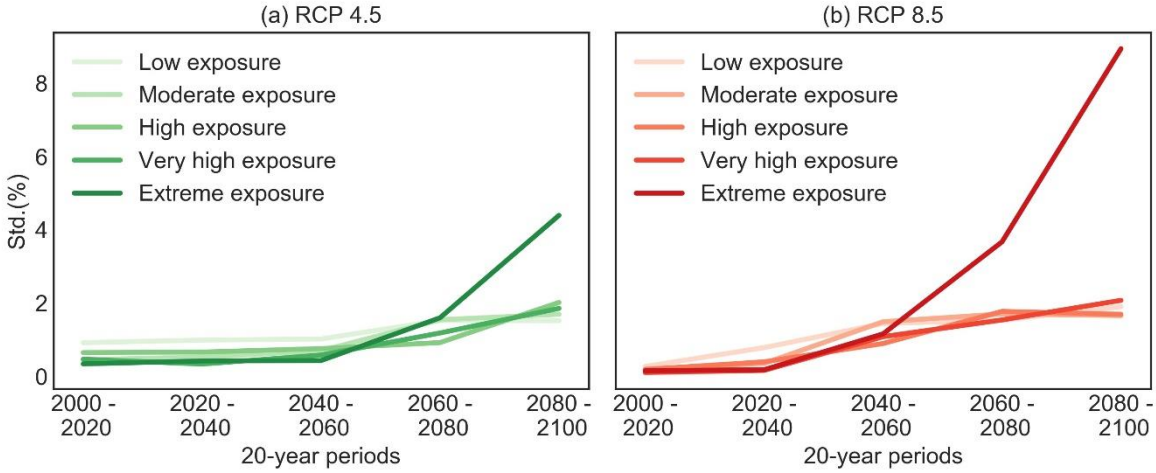


Figure 3-9. Temporal trends of the expected uncertainties in the fourteen datasets' relative exposures to different flood depths. (a) the temporal trends under RCP 4.5, (b) the temporal trends under RCP 8.5.

## 3.4 Discussion

### 3.4.1 Flood exposure under an uncertain future

Understanding flood exposure caused by SLR and storm surge is a prerequisite to obtain further information regarding risk, vulnerability, impact, resilience, and adaptation options in coastal urban areas under climate change. Projecting flood exposure involves a substantial amount of uncertainties that are propagated through a chain of different political, socioeconomic and technological assumptions, GHG emissions and concentrations (i.e. the RCPs in this study), and climate models (i.e. the GCMs and probabilistic SLR values in this study) (Wilby Robert L. & Dessai Suraje, 2010). Our results empirically show what the climate-related uncertainties would be like in a highly urbanized area, regarding the exposures of socioeconomic and demographic distributions, developed areas, lifeline infrastructures, and emergency responders to modeled flooding during extreme sea level events projected under two groups of climate scenarios (i.e. RCP 4.5 and 8.5), every 20 years between 2000 and 2100. We postulate that uncertainties should not be neglected, particularly for planning horizons beyond 2040 where our results indicate more salient uncertainties in relative exposures (Figure 3-8 and Figure 3-9).

Since flooding is sensitive to its underlying terrain conditions, coastal areas with flat terrain are more likely to experience uncertainties in projected flood exposures when compared with areas on steeper slopes. In flat areas, slight variations in projected sea levels can make the resulting flood areas advance or retreat over longer distances. For example, cities in the northeastern valleys of the Bay Area have more agreements in their flood areas between the low, median, and high estimates of the extreme sea level events (Figure 3-10 (a)). Alameda, a city built on flat landfill, is not only projected to have larger flood areas but also greater differences between the low, median and high estimates (Figure 3-10 (b)). Additional coastal developments are expected to emerge throughout the Bay Area due to the region's growing economy, projected increases in population and jobs, and amenities near the waterfront. A 2015 map found that 27 major proposed and ongoing real estate projects were located in flood-prone coastal zones (Wachtel, Ennamorato, & Burson, 2015) where the projected flood exposures were not only salient but also more uncertain, adding challenges to planning, management, and decision-making.

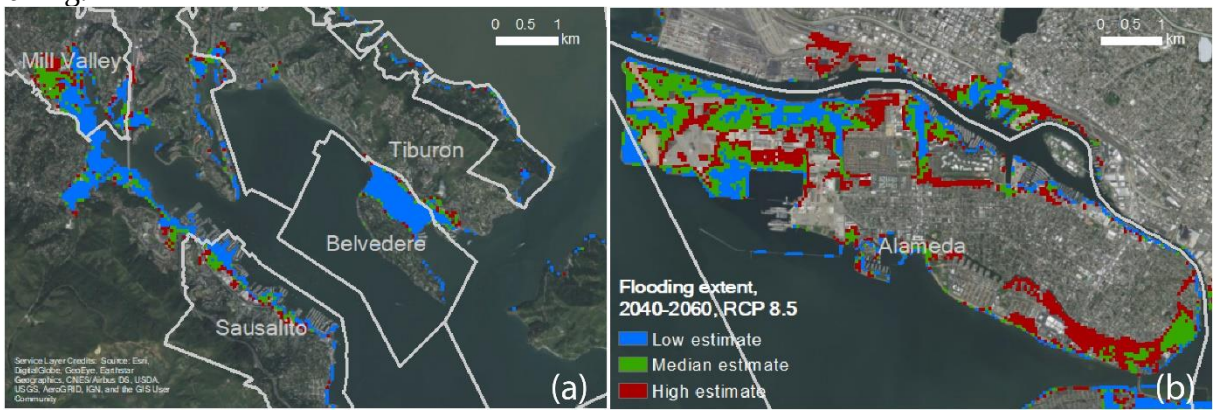


Figure 3-10. Comparing high, median, and low estimates of flood areas in places with steep (a) and flat (b) terrain. This example is illustrated using extreme sea level events under the RCP 8.5 scenarios during the 2040-2060 period.

While we only highlight climate-related uncertainties in this study, non-climatic factors can also affect the uncertainties in projected flood hazards and exposures. These factors include land subsidence and uplift, different population distribution projections, land cover and land use changes, and improvements or failures of flood control structures. A recent model at 2 m spatial resolution (Shirzaei & Bürgmann, 2018) indicates that land subsidence and uplift will make flood areas increase from 51-168 $km^2$ to 98-218 $km^2$ by 2100 in the southern Bay Area under a likely range (i.e. 67% probability) of SLR scenarios. These land subsidence and uplift projections slightly (i.e. statistically insignificant) alter our results at 50 m resolution, causing flood areas within the southern Bay Area (the extent of this area is in Appendix C) to increase by 0.23 $km^2$ and the datasets' relative exposures to increase by 0.62% on average during the extreme sea level events. In addition, the uncertainties in flood areas decrease by 0.07 $km^2$, and the uncertainties in a dataset's exposures decrease by 0.31% on average under the two RCPs and during the five 20-year periods. However, all these changes are small and statistically insignificant, for two reasons. First, Shirzaei et al. only consider SLR and exclude extreme storm surge, whereas we include both phenomena. Therefore, compared with Shirzaei et al., our projected flood areas are more extensive and likely cover coastal regions with salient land subsidence, even without including this effect. Adding land subsidence to our models should not change projected flood area in these regions or the broader study area. Second, our 50 m spatial resolution and average aggregation may attenuate the effects of land subsidence and uplift. Given the sensitivity of flood models to spatial resolution (Haile & Rientjes, 2005; Jakovljević & Govedarica, 2019; Ju et al., 2017), we assume that models at a finer spatial resolution such as 2 m may lead to significant changes in flood exposures and their uncertainties such as the ones observed by Shirzaei & Bürgmann (2018). These uncertainties may further increase when considering different land use and population projections, such as a recent Californian land use projection that contains four land use scenarios based on different population growth trajectories (Sleeter, Wilson, Sharygin, & Sherba, 2017). Levees are more relevant in low-lying and flat terrains, such as the adjacent Sacramento-San Joaquin River Delta whose islands are mostly 3-8 m below the current-day sea level (Ingebritsen, Ikehara, Galloway, & Jones, 2000) and protected by an extensive levee system (Mount & Twiss, 2005). These islands will likely be entirely flooded if the levees are overtopped or breached, and thus levee condition is a critical and uncertain factor for adaptation planning in this region. These non-climatic factors above, in addition to the climate drivers, serve to broaden the uncertainties in projected flood exposures.

### 3.4.2 Implications of the uncertainties for planning and management

To understand the implications of our results for planning and management in the Bay Area, we engaged with private and public stakeholders in the transportation and energy sectors. This outreach process is part of a broader project (Radke et al., 2018) that studies flood-related vulnerabilities of some lifeline infrastructures in our analysis. Two implications for planning and management are worth highlighting from this outreach process. First, the increased uncertainties over time pose obstacles for stakeholders to appreciate our long-term, multi-scenario flood projections and exposure analysis. Second, the obstacles above require adaptation strategies to cope with uncertainties and to promote long-term planning.

Concerns about uncertainties in long-term climate change projections and adaptations are not unique to the stakeholders we engaged with, they are also relevant for stakeholders in other sectors. Many of the stakeholders we interviewed are more interested in our near-term projections such as for 2000-2020 and 2020-2040, with insights from a few private-sector stakeholders suggesting that projections exceeding 10 years would be irrelevant in near-term decision making. This short-term focus is likely an uncertainty-avoidance organizational behavior (Slawinski, Pinkse, Busch, & Banerjee, 2017), that can be suboptimal in the long-term for infrastructures with longer investment, life, and planning cycles (Hallegatte, 2009). Concerns about uncertainty have also been reported in other sectors such as public land management (Peters, Schwartz, & Lubell, 2018), urban planning (Carter et al., 2015), business firms (Slawinski et al., 2017), and public outreach (Morton, Rabinovich, Marshall, & Bretschneider, 2011). Since the negative consequences of future climate are not certain, stakeholders worry about losing predictability and control, and are inclined to keep self-interested behaviors, which may reduce the likelihood of taking actions for adaptation (Morton et al., 2011; Slawinski et al., 2017). This could explain the preference of stakeholders for short-term projections and exposure analysis that have smaller uncertainties as discovered in our results.

Based on our findings and from the existing literature, planning and management strategies should incorporate the uncertain nature of future climate and promote long-term thinking (Oddo et al., 2017; Walker, Haasnoot, & Kwakkel, 2013; Woodward, Kapelan, & Gouldby, 2014). Hallegatte (2009) recommends no-regret (i.e. beneficial even without climate change) and reversible (i.e. low cost for being wrong about future climate) adaptation strategies to cope with uncertainties. Additionally, uncertainty-coping strategies are not only technical but also "soft" by incorporating financial, organizational and institutional measures. Such strategies should enhance redundancy of a given system and enable decision-making in a timely manner to keep pace with new situations. Our outreach process with the stakeholders reveals preferences for technical and irreversible adaptation strategies such as armoring infrastructures and building levees. We also find some interest in no-regret and "soft" strategies including building mutual aid and self-sufficient logistic groups for emergency preparedness, which can be promoted as these strategies are assumed to be lower-cost and more flexible than the technical and irreversible ones when adjusting to new climate projections. Another salient example of 'soft' strategy that has yet to be considered is the Sand Engine, a mega scale sand replenishment driven by waves and currents, implemented in the Netherlands to control for SLR-driven coastal recession while reducing project cost and disruptions to nature (Stive et al., 2013).

The long-term thinking is enhanced with the inclusion of certainties in climate change adaptation, such as adopting stable regulations over time that provide more predictability about future policy environment and reduce the risk of implementing certain adaptation actions (Slawinski et al., 2017). Furthermore, policies may explicitly require development projects to consider the long term. For example, a Bay Area's regulation has asked coastal projects to cope with flood projections for 2050 or 2100 depending on project life and existence of public safety risks (The San Francisco Bay Conservation and Development Commission, 2011). In realizing such long-term consideration, researchers, planners, and decision-makers may utilize the concept of 'adaption tipping point' (Kwadijk et al., 2010), which is carried out in the Netherlands to identify the time points when current strategies are no longer valid due to climate change and when new strategies are need.

### 3.4.3 Limitations and future research directions

While we focus on exposure in this study, it does not equal to flood impacts on urban areas. Future researchers may acquire additional information to transform exposure into impact related metrics such as disruptions in traffic flows, damages to homes, and loss of life through damage-depth curves (Huizinga, Moel, Szewczyk, European Commission, & Joint Research Centre, 2016; Pistrika, Tsakiris, & Nalbantis, 2014). In addition to studying direct impacts in exposed areas, researchers may further consider a network framework where flood impacts can propagate through intra- or inter-connected infrastructures, spreading impacts over a much broader region (Biging et al., 2012; Ge, Dou, & Zhang, 2017). Researchers suggest that this network perspective provides a more holistic picture on flood impacts and their related uncertainties (Balijepalli & Oppong, 2014; Eleutério, Hattemer, & Rozan, 2013; Haimes, 2009; Herrera, Flannery, & Krimmer, 2017; International Transport Forum (ITF), 2017; Lleras-Echeverri & Sanchez-Silva, 2001; Martinson, 2017; O'Rourke, 2007; Rodríguez-Núñez & García-Palomares, 2014).

Researchers can also benefit from collaborating with stakeholders during the flood modeling process. While we mainly presented our results to the stakeholders responsible for some lifeline infrastructures, future studies may expand to other relevant industries and sectors, collecting their opinions on flood modeling. The differences in the stakeholder's reactions (e.g. the preferences for short versus long term projections) may indicate their varied priorities and concerns about flooding and climate change adaptation. Such knowledge can inform more relevant and applicable regulations and policies. Furthermore, stakeholder engagement can help to identify appropriate modeling strategies in terms of spatial resolution, time horizons, areas of interest, and metrics for projecting flood hazard, which will hopefully lead to more stakeholder-specific results (Wadey et al., 2015).

Finally, a long-term, fine resolution, and consistent flood mapping inventory will benefit flood modeling, exposure and impact assessments, and consequently stakeholders. Such an inventory is becoming increasingly possible with the advances in remote sensing (Pekel, Cottam, Gorelick, & Belward, 2016), sensor networks (Chang & Guo, 2006), and volunteered geographic information (Poser & Dransch, 2010). As more floods are monitored and mapped, flood models can be calibrated with this growing historical inventory, instead of water level recordings from sparse gauges used by this study. With such an inventory in place, researchers can use flood maps to develop statistical models, which can be better applied to large regions and used to compare with process-based models; like the 3Di used in this study (Siahkamari, Haghizadeh, Zeinivand, Tahmasebipour, & Rahmati, 2017; Tien Bui et al., 2016; Wang et al., 2015). Additionally, historical maps are intuitive tools that increase public awareness about flood hazards and are baselines to better understand variabilities in flood exposure and impact under future climates.

## 3.5 Conclusions

SLR and storm surge under climate change pose additional flood threats to coastal urban areas. Successful adaptation requires pro-active planning that incorporates long-term flood projections and their intrinsic uncertainties. While these intrinsic uncertainties are previously understudied, in this study we utilize a scalable hydrodynamic model to simulate coastal flooding under multiple time horizons and climate scenarios to show the range of flood exposures in a highly urbanized coastal area. We assess flood exposure by intersecting simulated flood areas with fourteen datasets describing demographic and socioeconomic distributions, developed

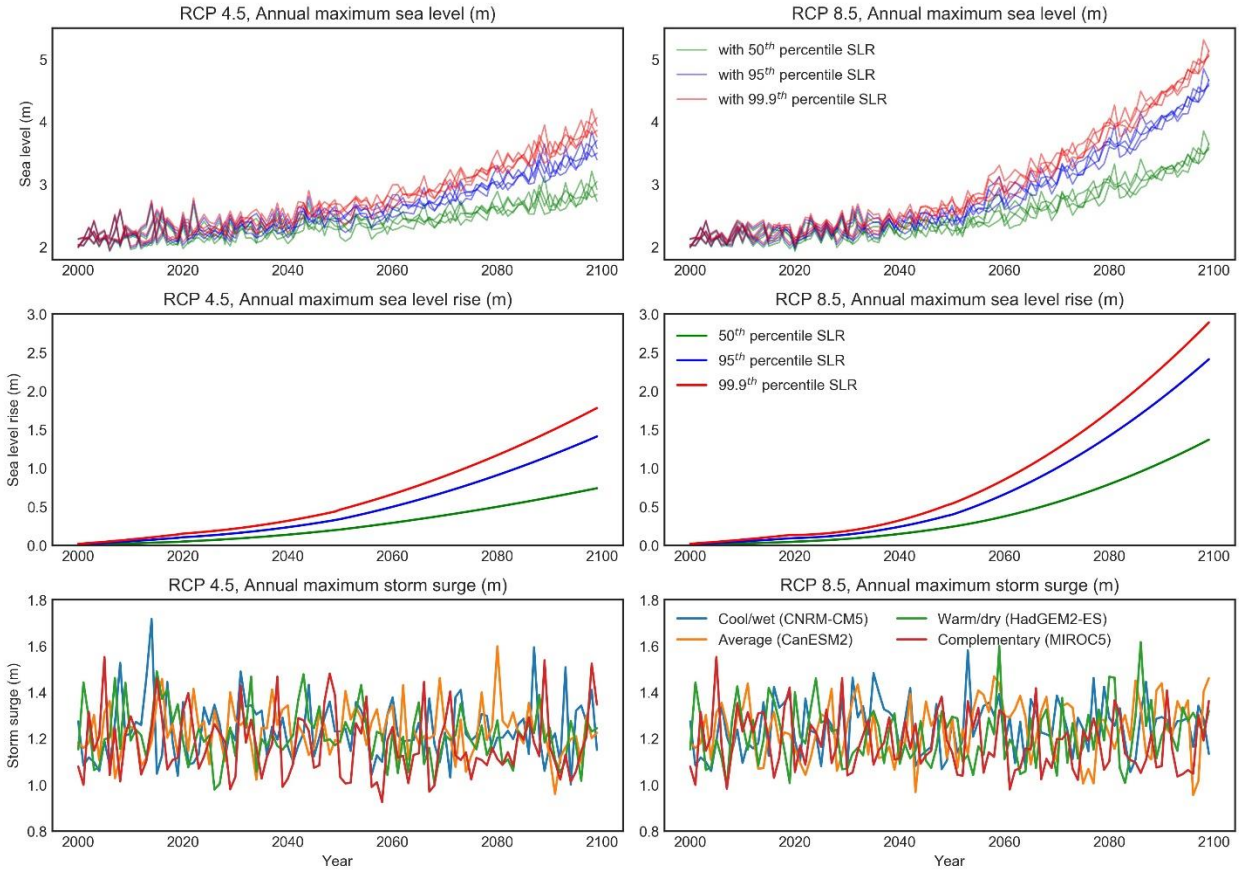
areas, lifeline infrastructures, and emergency responders. Our results are useful as a comprehensive base map for adaptation planning and to facilitate discussions with different stakeholders on adaptation options for future flood hazards while promoting transparency in the climate-related uncertainties associated. The approach we develop here can be used in coastal urban areas throughout the world.

Our results show increased flood exposures and broader uncertainties over time in the San Francisco Bay Area. The median estimates of lower GHG concentration scenarios (i.e. RCP 4.5) indicate 10-38% of the datasets' low-lying portions (i.e. <10 m in elevation) are exposed to flood in 2000-2020 and 20-54% exposed in 2080-2100. These numbers change to 0-35% and 40-67% respectively under the median estimates of higher GHG concentration scenarios (i.e. RCP 8.5). The expected uncertainties (i.e. standard deviations) in a given dataset's exposures is 1-2% (RCP 8.5 and 4.5) in the 2000-2020 period and 7-10% (RCP 4.5 and 8.5) in the 2080-2100 period. We find that these increased uncertainties are challenging for stakeholders when using long-term projections, particularly for private-sector stakeholders who tend to focus on short-term investment and planning cycles. Therefore, adaptation options should favor no-regret, reversible, and redundant strategies, and policies should be stable over time. With such efforts, stakeholders are more likely to engage with flood and climate change adaptation.

**Acknowledgements**

This work is funded by the Fourth Climate Change Assessment program of the California Energy Commission, grant FED-15-001 to the University of California, Berkeley. We thank Professor Manoochehr Shirzaei from Arizona State University for providing us the land subsidence and uplift projections in the southern Bay Area. We also thank Peter Norton from University of California Berkeley and Qin Ma from University of California Merced for their suggestions on the manuscript.

Appendix A. Visualization of annual maximum value of the hourly sea level projections, SLR, and storm surge from Cayan et al. (2016) .




Appendix B-1. Absolute exposures under RCP 4.5 and 8.5 scenarios, every 20 years between 2000 and 2100

| | RCP 4.5 | | | RCP 8.5 | | |
|---|---|---|---|---|---|---|
| Estimates | Min. | Median | Max. | Min. | Median | Max. |
| Period | 2000 - 2020 | | | | | |
| Population (1000 people) | 163 | 195 | 233 | 156 | 162 | 164 |
| Number of households (1000 households) | 60 | 73 | 85 | 58 | 60 | 60 |
| Number of housing units (1000 housing units) | 63 | 77 | 90 | 62 | 63 | 64 |
| High intensity development ($km^2$) | 17 | 21 | 24 | 15 | 17 | 18 |
| Medium intensity development ($km^2$) | 50 | 56 | 61 | 45 | 49 | 50 |
| Low intensity development ($km^2$) | 41 | 45 | 48 | 38 | 41 | 41 |
| Open space ($km^2$) | 24 | 26 | 28 | 23 | 24 | 24 |
| Roads (km) | 1215 | 1393 | 1540 | 1095 | 1197 | 1225 |
| Natural gas pipeline (km) | 66 | 75 | 85 | 59 | 65 | 66 |
| Natural gas stations (station) | 6 | 6 | 8 | 5 | 6 | 6 |
| Electric transmission line (km) | 312 | 332 | 355 | 294 | 308 | 311 |
| Electric substations (substation) | 17 | 18 | 22 | 14 | 16 | 17 |

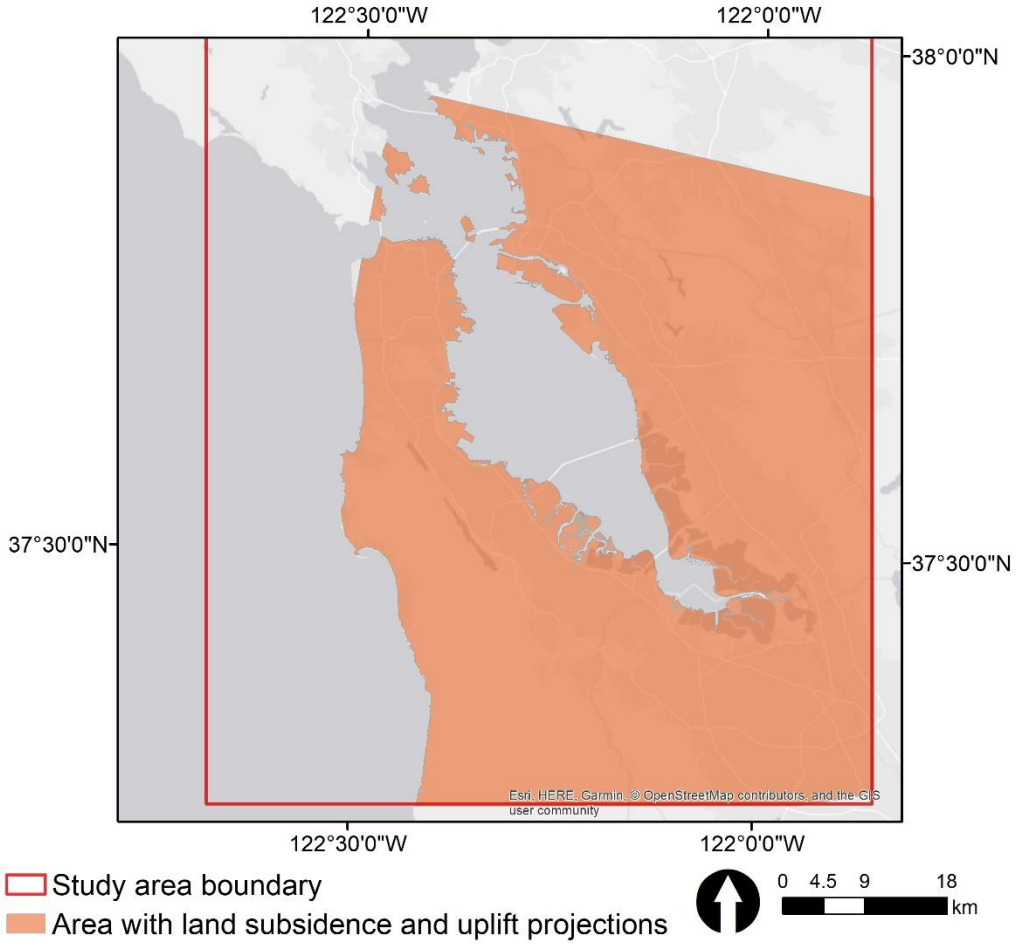
| | | | | | | |
|---|---|---|---|---|---|---|
| Fire stations (station) | 13 | 17 | 17 | 10 | 13 | 13 |
| Hospitals (hospital) | 0 | 1 | 1 | 0 | 0 | 0 |
| Period | 2020 - 2040 | | | | | |
| Population (1000 people) | 143 | 195 | 200 | 163 | 189 | 208 |
| Number of households (1000 households) | 52 | 73 | 75 | 59 | 70 | 77 |
| Number of housing units (1000 housing units) | 55 | 77 | 80 | 63 | 74 | 82 |
| High intensity development ($km^2$) | 15 | 21 | 24 | 18 | 21 | 22 |
| Medium intensity development ($km^2$) | 46 | 56 | 61 | 51 | 56 | 57 |
| Low intensity development ($km^2$) | 39 | 45 | 47 | 42 | 45 | 45 |
| Open space ($km^2$) | 23 | 26 | 27 | 25 | 26 | 26 |
| Roads (km) | 1121 | 1404 | 1530 | 1245 | 1381 | 1426 |
| Natural gas pipeline (km) | 60 | 76 | 82 | 68 | 75 | 76 |
| Natural gas stations (station) | 6 | 6 | 8 | 6 | 6 | 6 |
| Electric transmission line (km) | 296 | 333 | 350 | 315 | 332 | 335 |
| Electric substations (substation) | 14 | 18 | 22 | 17 | 17 | 18 |
| Fire stations (station) | 11 | 17 | 17 | 13 | 17 | 17 |
| Hospitals (hospital) | 0 | 1 | 1 | 0 | 1 | 1 |
| Period | 2040 - 2060 | | | | | |
| Population (1000 people) | 178 | 193 | 229 | 173 | 214 | 256 |
| Number of households (1000 households) | 66 | 73 | 83 | 63 | 75 | 90 |
| Number of housing units (1000 housing units) | 70 | 77 | 88 | 67 | 80 | 95 |
| High intensity development ($km^2$) | 20 | 23 | 28 | 18 | 29 | 41 |
| Medium intensity development ($km^2$) | 53 | 59 | 67 | 52 | 69 | 86 |
| Low intensity development ($km^2$) | 44 | 46 | 50 | 43 | 52 | 57 |
| Open space ($km^2$) | 25 | 27 | 29 | 25 | 30 | 33 |
| Roads (km) | 1314 | 1475 | 1703 | 1268 | 1762 | 2267 |
| Natural gas pipeline (km) | 71 | 79 | 94 | 68 | 101 | 136 |
| Natural gas stations (station) | 6 | 6 | 10 | 6 | 10 | 13 |
| Electric transmission line (km) | 323 | 343 | 372 | 317 | 383 | 429 |
| Electric substations (substation) | 17 | 21 | 23 | 17 | 26 | 38 |
| Fire stations (station) | 15 | 17 | 19 | 14 | 20 | 23 |
| Hospitals (hospital) | 0 | 1 | 1 | 0 | 1 | 2 |
| Period | 2060 - 2080 | | | | | |
| Population (1000 people) | 219 | 262 | 244 | 236 | 288 | 381 |
| Number of households (1000 households) | 79 | 92 | 89 | 83 | 107 | 138 |
| Number of housing units (1000 housing units) | 84 | 98 | 95 | 87 | 113 | 147 |
| High intensity development ($km^2$) | 24 | 34 | 45 | 28 | 56 | 70 |
| Medium intensity development ($km^2$) | 60 | 76 | 90 | 68 | 104 | 124 |
| Low intensity development ($km^2$) | 48 | 54 | 59 | 51 | 63 | 69 |
| Open space ($km^2$) | 28 | 32 | 34 | 30 | 37 | 40 |
| Roads (km) | 1509 | 1970 | 2393 | 1730 | 2793 | 3338 |
| Natural gas pipeline (km) | 83 | 115 | 144 | 97 | 168 | 199 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Natural gas stations (station) | 8 | 12 | 12 | 10 | 15 | 23 |
| Electric transmission line (km) | 353 | 402 | 441 | 376 | 475 | 519 |
| Electric substations (substation) | 22 | 32 | 41 | 24 | 49 | 58 |
| Fire stations (station) | 17 | 21 | 24 | 19 | 28 | 36 |
| Hospitals (hospital) | 1 | 1 | 2 | 1 | 2 | 3 |
| Period | | | 2080 - 2100 | | | |
| Population (1000 people) | 203 | 279 | 414 | 277 | 483 | 546 |
| Number of households (1000 households) | 72 | 101 | 151 | 103 | 172 | 199 |
| Number of housing units (1000 housing units) | 76 | 108 | 163 | 109 | 183 | 214 |
| High intensity development ($km^2$) | 28 | 59 | 77 | 54 | 87 | 100 |
| Medium intensity development ($km^2$) | 68 | 108 | 134 | 102 | 151 | 175 |
| Low intensity development ($km^2$) | 51 | 64 | 72 | 62 | 77 | 84 |
| Open space ($km^2$) | 30 | 37 | 42 | 36 | 44 | 48 |
| Roads (km) | 1736 | 2906 | 3634 | 2747 | 4129 | 4784 |
| Natural gas pipeline (km) | 98 | 174 | 218 | 165 | 243 | 286 |
| Natural gas stations (station) | 10 | 17 | 26 | 15 | 27 | 32 |
| Electric transmission line (km) | 380 | 485 | 544 | 470 | 584 | 624 |
| Electric substations (substation) | 25 | 51 | 63 | 49 | 70 | 77 |
| Fire stations (station) | 19 | 30 | 38 | 28 | 45 | 51 |
| Hospitals (hospital) | 1 | 2 | 3 | 2 | 5 | 5 |

Appendix B-2. Total amount of each dataset used in the exposure analysis

| Dataset | Total amount | Dataset | Total amount |
|---|---|---|---|
| Population (1000 people) | 1153 | Roads (km) | 8160 |
| Number of households (1000 households) | 430 | Natural gas pipeline (km) | 531 |
| Number of housing units (1000 housing units) | 464 | Natural gas stations (station) | 55 |
| High intensity development (km2) | 155 | Electric transmission line (km) | 967 |
| Medium intensity development (km2) | 296 | Electric substations (substation) | 105 |
| Low intensity development (km2) | 125 | Fire stations (station) | 95 |
| Open space (km2) | 69 | Hospitals (hospital) | 10 |

Appendix C. Area of the analysis with land subsidence and uplift data




**References**


Balijepalli, C., & Oppong, O. (2014). Measuring vulnerability of road network considering the extent of serviceability of critical road links in urban areas. *Journal of Transport Geography*, *39*, 145–155. https://doi.org/10.1016/j.jtrangeo.2014.06.025

Barnard, P. L., O'Reilly, B., van Ormondt, M., Elias, E., Ruggiero, P., Erikson, L. H., … Adams, P. N. (2009). *The Framework of a Coastal Hazards Model- A Tool for Predicting the Impact of Severe Storms*. U. S. Geological Survey.

Barnard, P. L., Ormondt, M. van, Erikson, L. H., Eshleman, J., Hapke, C., Ruggiero, P., … Foxgrover, A. C. (2014). Development of the Coastal Storm Modeling System (CoSMoS) for predicting the impact of storms on high-energy, active-margin coasts. *Natural Hazards*, *74*(2), 1095–1125. https://doi.org/10.1007/s11069-014-1236-y

Bickers, K. M. (2014). *Vulnerable Populations to Climate Change in New Jersey*. Edward J. Bloustein School of Planning and Public Policy Rutgers, The State University of New Jersey Rutgers University.

Biging, G., Radke, J., & Lee, J. H. (2012). *Impacts of predicted sea-level rise and extreme storm events on the transportation in the San Francisco Bay Region* (No. CEC - 500 - 2012 - 040). California Energy Commission.

Carter, J. G., Cavan, G., Connelly, A., Guy, S., Handley, J., & Kazmierczak, A. (2015). Climate change and the city: Building capacity for urban adaptation. *Progress in Planning*, *95*, 1–66. https://doi.org/10.1016/j.progress.2013.08.001

Cayan, D. R., Kalansky, J., Iacobellis, S., & Pierce, D. (2016). *Creating Probabilistic Sea Leve Rise Projections to support the 4th California Climate Assessment* (No. 16-IEPR-04). Retrieved from California Energy Commission website: http://docketpublic.energy.ca.gov/PublicDocuments/16-IEPR-04/TN211806_20160614T101823_Creating_Probabilistic_Sea_Leve_Rise_Projections.pdf

Chang, N.-B., & Guo, D.-H. (2006). Urban Flash Flood Monitoring, Mapping and Forecasting via a Tailored Sensor Network System. *2006 IEEE International Conference on Networking, Sensing and Control*, 757–761. https://doi.org/10.1109/ICNSC.2006.1673241

Coles, D., Yu, D., Wilby, R. L., Green, D., & Herring, Z. (2017). Beyond 'flood hotspots': Modelling emergency service accessibility during flooding in York, UK. *Journal of Hydrology*, *546*, 419–436. https://doi.org/10.1016/j.jhydrol.2016.12.013

Creel, L. (2003). *Ripple effects: Population and coastal regions*. Population Reference Bureau Washington, DC.

Dahm, R., Hsu, C.-T., Lien, H.-C., Chang, C.-H., & Prinsen, G. (2014). *Next Generation Flood Modelling using 3Di: A Case Study in Taiwan*. Presented at the DSD International Conference 2014, sustainable stormwater and waste water management, Hong Kong.

Dasgupta, S., Laplante, B., Meisner, C., Wheeler, D., & Yan, J. (2008). The impact of sea level rise on developing countries: A comparative analysis. *Climatic Change*, *93*(3–4), 379–388. https://doi.org/10.1007/s10584-008-9499-5

Dawson, R. J., Ball, T., Werritty, J., Werritty, A., Hall, J. W., & Roche, N. (2011). Assessing the effectiveness of non-structural flood management measures in the Thames Estuary under conditions of socio-economic and environmental change. *Global Environmental Change*, *21*(2), 628–646. https://doi.org/10.1016/j.gloenvcha.2011.01.013

De Sherbinin, A., Schiller, A., & Pulsipher, A. (2007). The vulnerability of global cities to climate hazards. *Environment and Urbanization*, *19*(1), 39–64. https://doi.org/10.1177/0956247807076725

Demirel, H., Kompil, M., & Nemry, F. (2015). A framework to analyze the vulnerability of European road networks due to Sea-Level Rise (SLR) and sea storm surges. *Transportation Research Part A: Policy and Practice*, *81*, 62–76. https://doi.org/10.1016/j.tra.2015.05.002

EERI. (2016). *Improve Reliability of Lifeline Infrastructure Systems* [Policy White Paper]. Retrieved from Earthquake Engineering Research Institute website: https://www.eeri.org/wp-content/uploads/eeri-policy-lifelines.pdf

Eleutério, J., Hattemer, C., & Rozan, A. (2013). A systemic method for evaluating the potential impacts of floods on network infrastructures. *Natural Hazards and Earth System Science*, *13*(4), 983–998. https://doi.org/10.5194/nhess-13-983-2013

ESRI. (2016). ArcGIS Desktop (Version 10.4). ArcMap 10.4: ESRI.

Ge, Y., Dou, W., & Zhang, H. (2017). A New Framework for Understanding Urban Social Vulnerability from a Network Perspective. *Sustainability*, *9*(10), 1723. https://doi.org/10.3390/su9101723

Haile, A. T., & Rientjes, T. H. M. (2005). Effects of LiDAR DEM resolution in flood modelling: A model sensitivity study for the city of Tegucigalpa, Honduras. *Isprs Wg Iii/3, Iii/4*, *3*, 12–14.

Haimes, Y. Y. (2009). On the Complex Definition of Risk: A Systems-Based Approach. *Risk Analysis*, *29*(12), 1647–1654. https://doi.org/10.1111/j.1539-6924.2009.01310.x

Hallegatte, S. (2009). Strategies to adapt to an uncertain climate change. *Global Environmental Change*, *19*(2), 240–247. https://doi.org/10.1016/j.gloenvcha.2008.12.003

Heberger, M., Cooley, H., Moore, E., & Herrera, P. (2012). *The impacts of Sea Level Rise on the San Francisco Bay* (No. CEC - 500 - 2012 - 014). California Energy Commission.

Herrera, E. K., Flannery, A., & Krimmer, M. (2017). Risk and Resilience Analysis for Highway Assets. *Transportation Research Record: Journal of the Transportation Research Board*, *2604*(1), 1–8. https://doi.org/10.3141/2604-01

Hsu, Y.-C., Prinsen, G., Bouaziz, L., Lin, Y.-J., & Dahm, R. (2016). An Investigation of DEM Resolution Influence on Flood Inundation Simulation. *Procedia Engineering*, *154*, 826–834. https://doi.org/10.1016/j.proeng.2016.07.435

Huizinga, J., Moel, H. de, Szewczyk, W., European Commission, & Joint Research Centre. (2016). *Global flood depth-damage functions methodology and the database with guidelines.* Luxembourg: Publications Office.

Hunt, A., & Watkiss, P. (2010). Climate change impacts and adaptation in cities: A review of the literature. *Climatic Change*, *104*(1), 13–49. https://doi.org/10.1007/s10584-010-9975-6

Ingebritsen, S. E., Ikehara, M. E., Galloway, D. L., & Jones, D. R. (2000). *Delta subsidence in California; the sinking heart of the state* (No. Fact Sheet 005-00). Retrieved from U.S. Geological Survey website: http://pubs.usgs.gov/fs/2000/fs00500/

International Transport Forum (ITF). (2017). *Adapting Transport to Climate Change and Extreme Weather: Implications for Infrastructure Owners and Network Managers*. Retrieved from http://dx.doi.org/10.1787/9789282108079-en

Jabareen, Y. (2015). City planning deficiencies & climate change – The situation in developed and developing cities. *Geoforum*, *63*, 40–43. https://doi.org/10.1016/j.geoforum.2015.05.017

Jakovljević, G., & Govedarica, M. (2019). Water Body Extraction and Flood Risk Assessment Using Lidar and Open Data. In W. Leal Filho, G. Trbic, & D. Filipovic (Eds.), *Climate Change Adaptation in Eastern Europe: Managing Risks and Building Resilience to Climate Change* (pp. 93–111). https://doi.org/10.1007/978-3-030-03383-5_7

Ju, Y., Hsu, W.-C., Radke, J., Fourt, W., Lang, W., Hoes, O., … Maier, W. (2017). Planning for the Change: Mapping Sea Level Rise and Storm Inundation in Sherman Island Using 3Di Hydrodynamic Model and LiDAR. In P. (Vonu) Thakuriah, N. Tilahun, & M. Zellner (Eds.), *Seeing Cities Through Big Data* (pp. 313–329). https://doi.org/10.1007/978-3-319-40902-3_18

Kaźmierczak, A., & Cavan, G. (2011). Surface water flooding risk to urban communities: Analysis of vulnerability, hazard and exposure. *Landscape and Urban Planning*, *103*(2), 185–197. https://doi.org/10.1016/j.landurbplan.2011.07.008

KC, B., Shepherd, J. M., & Gaither, C. J. (2015). Climate change vulnerability assessment in Georgia. *Applied Geography*, *62*, 62–74. https://doi.org/10.1016/j.apgeog.2015.04.007

Knowles, N. (2009). *Potential inundation due to rising sea levels in the San Francisco Bay Region* (No. CEC-500-2009-023-D). Retrieved from California Climate Change Center website: http://www.energy.ca.gov/2009publications/CEC-500-2009-023/CEC-500-2009-023-D.PDF

Knowles, N. (2010). Potential Inundation Due to Rising Sea Levels in the San Francisco Bay Region. *San Francisco Estuary and Watershed Science*, *8*(1). Retrieved from http://escholarship.org/uc/item/8ck5h3qn

Kwadijk, J. C. J., Haasnoot, M., Mulder, J. P. M., Hoogvliet, M. M. C., Jeuken, A. B. M., Krogt, R. A. A. van der, … Wit, M. J. M. de. (2010). Using adaptation tipping points to prepare for climate change and sea level rise: A case study in the Netherlands. *Wiley Interdisciplinary Reviews: Climate Change*, *1*(5), 729–740. https://doi.org/10.1002/wcc.64

Lang, W., Radke, J. D., Chen, T., & Chan, E. H. W. (2016). Will affordability policy transcend climate change? A new lens to re-examine equitable access to healthcare in the San Francisco Bay Area. *Cities*, *58*, 124–136. https://doi.org/10.1016/j.cities.2016.05.014

Leicher, A. M. (2016). *Urban flood analyses in sloping areas with 3Di modelling*.

Lleras-Echeverri, G., & Sanchez-Silva, M. (2001). Vulnerability analysis of highway networks, methodology and case study. *Proceedings of the Institution of Civil Engineers, Transport 147*(4), 223–236.

Marcy, D., Herold, N., Waters, K., Brooks, W., Hadley, B., Pendleton, M., … Ryan, S. (2011). *New mapping tool and techniques for visualizing Sea Level Rise and coastal flooding impacts*. NOAA Coastal Service Center.

Martinich, J., Neumann, J., Ludwig, L., & Jantarasami, L. (2013). Risks of sea level rise to disadvantaged communities in the United States. *Mitigation and Adaptation Strategies for Global Change*, *18*(2), 169–185. https://doi.org/10.1007/s11027-011-9356-0

Martinson, R. (2017). Resilience in a Transportation System: A Whole System Approach. *Transportation Research Circular*, (Circular E-226), 1–9.

McGranahan, G., Balk, D., & Anderson, B. (2007). The rising tide: Assessing the risks of climate change and human settlements in low elevation coastal zones. *Environment and Urbanization*, *19*(1), 17–37. https://doi.org/10.1177/0956247807076960

Morton, T. A., Rabinovich, A., Marshall, D., & Bretschneider, P. (2011). The future that may (or may not) come: How framing changes responses to uncertainty in climate change communications. *Global Environmental Change*, *21*(1), 103–109. https://doi.org/10.1016/j.gloenvcha.2010.09.013

Mount, J., & Twiss, R. (2005). Subsidence, Sea Level Rise, and Seismicity in the Sacramento–San Joaquin Delta. *San Francisco Estuary and Watershed Science*, *3*(1). Retrieved from http://escholarship.org/uc/item/4k44725p

Neumann, B., Vafeidis, A. T., Zimmermann, J., & Nicholls, R. J. (2015). Future Coastal Population Growth and Exposure to Sea-Level Rise and Coastal Flooding—A Global Assessment. *PLOS ONE*, *10*(3), e0118571. https://doi.org/10.1371/journal.pone.0118571

NOAA. (2019). Water Levels—NOAA Tides & Currents. Retrieved March 20, 2019, from https://tidesandcurrents.noaa.gov/waterlevels.html?id=9414290

Nutters, H. (2012). *Addressing social vulnerability and equity in climate change adaptation planning*. San Francisco Bay Conservation and Development Commission.

Oddo, P. C., Lee, B. S., Garner, G. G., Srikrishnan, V., Reed, P. M., Forest, C. E., & Keller, K. (2017). Deep Uncertainties in Sea-Level Rise and Storm Surge Projections: Implications for Coastal Flood Risk Management: Deep Uncertainties in Coastal Flood Risk Management. *Risk Analysis*. https://doi.org/10.1111/risa.12888

Oh Eun Ho, Deshmukh Abhijeet, & Hastak Makarand. (2013). Criticality Assessment of Lifeline Infrastructure for Enhancing Disaster Response. *Natural Hazards Review*, *14*(2), 98–107. https://doi.org/10.1061/(ASCE)NH.1527-6996.0000084

O'Rourke, T. D. (2007). *Critical Infrastructure, Interdependencies, and Resilience*. 8.

Pekel, J.-F., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. *Nature*, *540*(7633), 418–422. https://doi.org/10.1038/nature20584

Pelling, M. (2003). *The vulnerability of cities. [Electronic resource]: Natural disasters and social resilience.* Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=cat04202a&AN=ucb.b13633897&site=eds-live

Peters, C. B., Schwartz, M. W., & Lubell, M. N. (2018). Identifying climate risk perceptions, information needs, and barriers to information exchange among public land managers. *Science of The Total Environment*, *616–617*, 245–254. https://doi.org/10.1016/j.scitotenv.2017.11.015

Pierce, D., Cayan, D. R., & Dehann, L. (2016). *Creating Climate projections to support the 4th California Climate Assessment* (No. 16-IEPR-04). Retrieved from California Energy Commission website: http://www.energy.ca.gov/2016_energypolicy/documents/2016-06-21_workshop/2016-06-21_documents.php

Pinto, P. J., & Kondolf, G. M. (2016). Evolution of Two Urbanized Estuaries: Environmental Change, Legal Framework, and Implications for Sea-Level Rise Vulnerability. *Water*, *8*(11), 535. https://doi.org/10.3390/w8110535

Pistrika, A., Tsakiris, G., & Nalbantis, I. (2014). Flood Depth-Damage Functions for Built Environment. *Environmental Processes*, *1*(4), 553–572. https://doi.org/10.1007/s40710-014-0038-2

Poser, K., & Dransch, D. (2010). Volunteered Geographic Information for Disaster Management with Application to Rapid Flood Damage Estimation. *Geomatica*, *64*(1), 89–98.

Radke, J., Biging, G., Roberts, K., Foster, H., Roe, E., Ju, Y., … Dalal, A. (2018). *Assessing Extreme Weather-Related Vulnerability and Identifying Resilience Options for California's Interdependent Transportation Fuel Sector* (No. CCCA4-CEC-2018–012). Retrieved from California Energy Commission website: http://www.climateassessment.ca.gov/techreports/docs/20180827-Energy_CCCA4-CEC-2018-012.pdf

Radke, J., Biging, G., Schimidt-Poolman, M., Foster, H., Roe, E., Ju, Y., … Reeves, I. (2017). *Assessment of Bay Area Gas Pipeline Vulnerability to Climate Change* (No. CEC-500-2017-008). Retrieved from California Energy Commission website: https://www.energy.ca.gov/2017publications/CEC-500-2017-008/CEC-500-2017-008.pdf

Rodríguez-Núñez, E., & García-Palomares, J. C. (2014). Measuring the vulnerability of public transport networks. *Journal of Transport Geography*, *35*, 50–63. https://doi.org/10.1016/j.jtrangeo.2014.01.008

Rosenzweig, C., Solecki, W., Hammer, S. A., & Mehrotra, S. (2010). Cities lead the way in climate-change action. *Nature*, *467*(7318), 909–911. https://doi.org/10.1038/467909a

Schile, L. M., Callaway, J. C., Morris, J. T., Stralberg, D., Parker, V. T., & Kelly, M. (2014). Modeling Tidal Marsh Distribution with Sea-Level Rise: Evaluating the Role of Vegetation, Sediment, and Upland Habitat in Marsh Resiliency. *PLoS ONE*, *9*(2), e88760. https://doi.org/10.1371/journal.pone.0088760

Seto, K. C., Fragkias, M., Güneralp, B., & Reilly, M. K. (2011). A Meta-Analysis of Global Urban Land Expansion. *PLOS ONE*, *6*(8), e23777. https://doi.org/10.1371/journal.pone.0023777

Shirzaei, M., & Bürgmann, R. (2018). Global climate change and local land subsidence exacerbate inundation risk to the San Francisco Bay Area. *Science Advances*, *4*(3), eaap9234. https://doi.org/10.1126/sciadv.aap9234

Siahkamari, S., Haghizadeh, A., Zeinivand, H., Tahmasebipour, N., & Rahmati, O. (2017). Spatial prediction of flood-susceptible areas using frequency ratio and maximum entropy models. *Geocarto International*, *0*(0), 1–15. https://doi.org/10.1080/10106049.2017.1316780

Slawinski, N., Pinkse, J., Busch, T., & Banerjee, S. B. (2017). The Role of Short-Termism and Uncertainty Avoidance in Organizational Inaction on Climate Change: A Multi-Level Framework. *Business & Society*, *56*(2), 253–282. https://doi.org/10.1177/0007650315576136

Sleeter, B. M., Wilson, T. S., Sharygin, E., & Sherba, J. T. (2017). Future Scenarios of Land Change Based on Empirical Data and Demographic Trends: LAND USE PROJECTIONS FOR CALIFORNIA. *Earth's Future*. https://doi.org/10.1002/2017EF000560

Stelling, G. S. (2012). Quadtree flood simulations with sub-grid digital elevation models. *Proceedings of the ICE - Water Management*, *165*(10), 567–580. https://doi.org/10.1680/wama.12.00018

Stive, M. J. F., de Schipper, M. A., Luijendijk, A. P., Aarninkhof, S. G. J., van Gelder-Maas, C., van Thiel de Vries, J. S. M., … Ranasinghe, R. (2013). A New Alternative to Saving Our Beaches from Sea-Level Rise: The Sand Engine. *Journal of Coastal Research*, 1001–1008. https://doi.org/10.2112/JCOASTRES-D-13-00070.1

Storch, H., & Downes, N. K. (2011). A scenario-based approach to assess Ho Chi Minh City's urban development strategies against the impact of climate change. *Cities*, *28*(6), 517–526. https://doi.org/10.1016/j.cities.2011.07.002

Strauss, B. H., Ziemlinski, R., Weiss, J. L., & Overpeck, J. T. (2012). Tidally adjusted estimates of topographic vulnerability to sea level rise and flooding for the contiguous United States. *Environmental Research Letters*, *7*(1), 014033. https://doi.org/10.1088/1748-9326/7/1/014033

The Association of Bay Area Governments (ABAG), Metropolitan Transportation Commission (MTC). (2013). *Plan Bay Area: Strategy for a sustainable region*.

The San Francisco Bay Conservation and Development Commission. (2011). *SFBCDC - New Sea Level Rise Policies Fact Sheet*. Retrieved from http://www.bcdc.ca.gov/BPA/SLRfactSheet.html

Tien Bui, D., Pradhan, B., Nampak, H., Bui, Q.-T., Tran, Q.-A., & Nguyen, Q.-P. (2016). Hybrid artificial intelligence approach based on neural fuzzy inference model and metaheuristic optimization for flood susceptibilitgy modeling in a high-frequency tropical cyclone area using GIS. *Journal of Hydrology*, *540*, 317–330. https://doi.org/10.1016/j.jhydrol.2016.06.027

UNU-IHDP. (2015). *Coastal Zones and Urbanization. Summary for Decision-Makers.* Bonn: UNU-IHDP.

Wachtel, M., Ennamorato, M., & Burson, B. (2015). *A baywide building boom threatened by rising waters*. Retrieved from http://sfpublicpress.org/news/searise/2015-07/interactive-map-a-baywide-building-boom-threatened-by-rising-waters

Wadey, M. P., Cope, S. N., Nicholls, R. J., McHugh, K., Grewcock, G., & Mason, T. (2015). Coastal flood analysis and visualization for a small town. *Ocean & Coastal Management*, *116*, 237–247. https://doi.org/10.1016/j.ocecoaman.2015.07.028

Walker, W., Haasnoot, M., & Kwakkel, J. (2013). Adapt or Perish: A Review of Planning Approaches for Adaptation under Deep Uncertainty. *Sustainability*, *5*(3), 955–979. https://doi.org/10.3390/su5030955

Wang, Z., Lai, C., Chen, X., Yang, B., Zhao, S., & Bai, X. (2015). Flood hazard risk assessment model based on random forest. *Journal of Hydrology*, *527*(Supplement C), 1130–1141. https://doi.org/10.1016/j.jhydrol.2015.06.008

Wilby Robert L., & Dessai Suraje. (2010). Robust adaptation to climate change. *Weather*, *65*(7), 180–185. https://doi.org/10.1002/wea.543

Wood, N. (2009). Tsunami exposure estimation with land-cover data: Oregon and the Cascadia subduction zone. *Applied Geography*, *29*(2), 158–170. https://doi.org/10.1016/j.apgeog.2008.08.009

Woodward, M., Kapelan, Z., & Gouldby, B. (2014). Adaptive Flood Risk Management Under Climate Change Uncertainty Using Real Options and Optimization: Adaptive Flood Risk Management. *Risk Analysis*, *34*(1), 75–92. https://doi.org/10.1111/risa.12088

Zhu, M., Xi, X., Hoctor, T. S., & Volk, M. (2015). Integrating conservation costs into sea level rise adaptive conservation prioritization. *Global Ecology and Conservation*, *4*, 48–62. https://doi.org/10.1016/j.gecco.2015.05.007

# Chapter 4. An Equity Analysis of Clean Vehicle Rebate Programs in California[3]

**Abstract**: Rebates incentivize clean vehicle adoption to reduce mobile-source emissions of greenhouse gases and other air pollutants. However, rebates raise equity concerns because they require upfront capital to acquire a vehicle. Since wealthier communities also typically experience better air quality, rebates may also not incentivize clean vehicle acquisitions in more polluted regions where the air quality benefits would be greatest. We compare rebate allocation across California (CA) with respect to community socioeconomic and environmental disadvantage, household income, race and ethnicity, and ambient nitrogen dioxide and particulate matter concentrations using data from two rebate programs. We find that CA's Clean Vehicle Rebate Project issued more rebates per thousand households to advantaged, higher-income communities and less to those with higher percentages of Hispanics and Blacks. Implementing an income cap and additional rebate amount for lower-income consumers attenuated income's effect on rebate allocation. CA's Enhanced Fleet Modernization Program, which includes specific designs to address equity, allocated rebates to communities that were more disadvantaged, lower-income, and with higher percentages of Hispanics. These findings indicate that specific policy designs, including an income cap, expanded vehicle eligibility, and income- and geography- tiered rebate amounts, can help to ensure that rebates are accessible to diverse populations.

## 4.1 Introduction

Clean vehicles including plug-in hybrid electric vehicles (PHEVs), battery electric vehicles (BEVs), and fuel cell electric vehicles (FCEVs) reduce roadway emissions of greenhouse gases (GHGs) and other hazardous co-pollutants emitted from gasoline and diesel-powered internal combustion engines (ICEs) (Hardman, Chandan, Tal, & Turrentine, 2017). Government rebates, or monetary refunds after purchase or lease (defined as acquisition hereafter), are used to promote the adoption of clean vehicles in several U.S. states, including California (DeShazo, 2016), in order to meet clean air and climate change mitigation goals. However, rebates require consumers to acquire the vehicle upfront, presenting a barrier for lower-income consumers with limited financial assets or access to credit when compared with point-of-sale incentives such as sales tax exemption and government purchase discount (Hardman et al., 2017; Snelling, 2018). Rebate programs can be even less accessible to lower-income consumers if the rebate amount does not increase according to their income. Wealthier consumers are also more likely to take advantage of and benefit from rebate programs that lack income or vehicle price caps (DeShazo, 2016; Snelling, 2018). Together, these features raise equity concerns regarding the design and implementation of clean vehicle rebate programs, particularly if disadvantaged consumers are not as able to benefit.

Transportation is the largest source of GHG emissions and responsible for substantial portions of co-pollutants including $NO_x$, $SO_x$, CO, Ozone and PM in California (Anderson, Kissel, Field, & Mach, 2018; California Air Resource Board, 2018a). Clean vehicle rebate programs are established to help reduce emissions from the transportation sector and to support California's emission reduction goals required by the state's Global Warming Solutions Act (AB

[3] Reproduced in part with permission from Environmental Science & Technology, submitted for publication. Unpublished work copyright 2019 American Chemical Society.

32) passed in 2006 (California Air Resource Board, 2018b; Rubin & St-Louis, 2016). Cap-and-Trade auction proceeds partially fund these rebate programs, and state legislation SB 535 requires 25% of the proceeds to be used in projects that benefit socioeconomically and environmentally disadvantaged communities (California Environmental Protection Agency, 2017). Therefore, in addition to emission reduction, equity is an inherent goal of the rebate programs.

In this study, we analyze the equity implications of two major clean vehicle rebate programs in California with different policy design elements. The first is the statewide Clean Vehicle Rebate Project (CVRP), which issues rebates for new clean vehicles (PHEVs, BEVs, and FCEVs) and has limited considerations of equity. Implemented in 2010, the CVRP provides consumers with rebates after the acquisition of a qualified vehicle. During the first iteration (i.e. March 2010 to March 2016) of this program, the rebate amount ranged between $1500 and $5000, depending only on vehicle technology rather than other factors such as consumer's income. Since these additional factors were not considered, it is likely that wealthier consumers received more rebates during this first iteration.

Some researchers suggest that combining rebates with vehicle retirement incentives, implementing caps for income and vehicle prices, and adopting income-tiered rebate amounts can improve the equity and effectiveness of clean vehicle rebate programs in terms of covering more lower-income consumers, distributing more rebates and encouraging clean vehicle adoption, and reducing emission and air pollution from the retired vehicles (DeShazo, 2016; DeShazo, Sheldon, & Carson, 2017; Snelling, 2018). Beginning in April 2016, the CVRP implemented an income cap to exclude PHEV and BEV consumers with gross annual individual income greater than $150,000 and offered an additional $2000 to lower-income consumers with annual household income below 300% of the Federal Poverty Level (Center for Sustainable Energy, 2016).

The second clean vehicle rebate program in California, the Enhanced Fleet Modernization Program (EFMP), sought to further addresses the equity concerns presented in the CVRP by expanding eligible vehicles, targeting lower-income consumers, and focusing on socioeconomically and environmentally disadvantaged communities (California Air Resource Board, 2018b; Pierce & DeShazo, 2017). EFMP includes a retire and replacement program, launched in 2015, to subsidize the purchase of more efficient, used (less than eight years old) and new vehicles after the retirement of older, higher-emitting vehicles. A complementary Plus-up program provides an additional rebate amount for consumers living in socioeconomically and environmentally disadvantaged communities as identified using California Environmental Protection Agency (Cal-EPA) mapping tool, CalEnviroScreen (August, 2016; Faust et al., 2017). EFMP also includes a vehicle retirement only program, but we focus on the retire and replace and the Plus-up programs that are the subject of our analysis. Compared with CVRP, EFMP expands eligible vehicles to include efficient ICE vehicles and non-plug-in hybrid vehicles, which are presumably more affordable than PHEVs, BEVs, and FCEVs with similar characteristics. EFMP also limits eligible participants to those with annual household income below 400% of the Federal Poverty Level and provides a higher rebate amount for those with lower incomes and, through the Plus-up program, those living in disadvantaged communities. Furthermore, EFMP can be combined with CVRP when buying a new PHEV or BEV, further increasing the rebate amount (California Air Resource Board, 2018b). Thus, while total EFMP rebate ranges from $2,500 to $9,500 depending on vehicle type, household income, and neighborhood disadvantage, when combined with CVRP in acquiring a new PHEV or BEV, a

consumer can receive a combined rebate from $4,000 to $14,000 (see Table S1 for the amount of rebates than can be received by lower-income consumers from the two programs). EFMP is currently operating as a pilot program only in the South Coast and San Joaquin Valley air districts, two regions with the worst air quality concerns in the state (American Lung Association, 2018). The geography, timeline, and rebate allocation pattern of the two programs are in Figure 4-1.

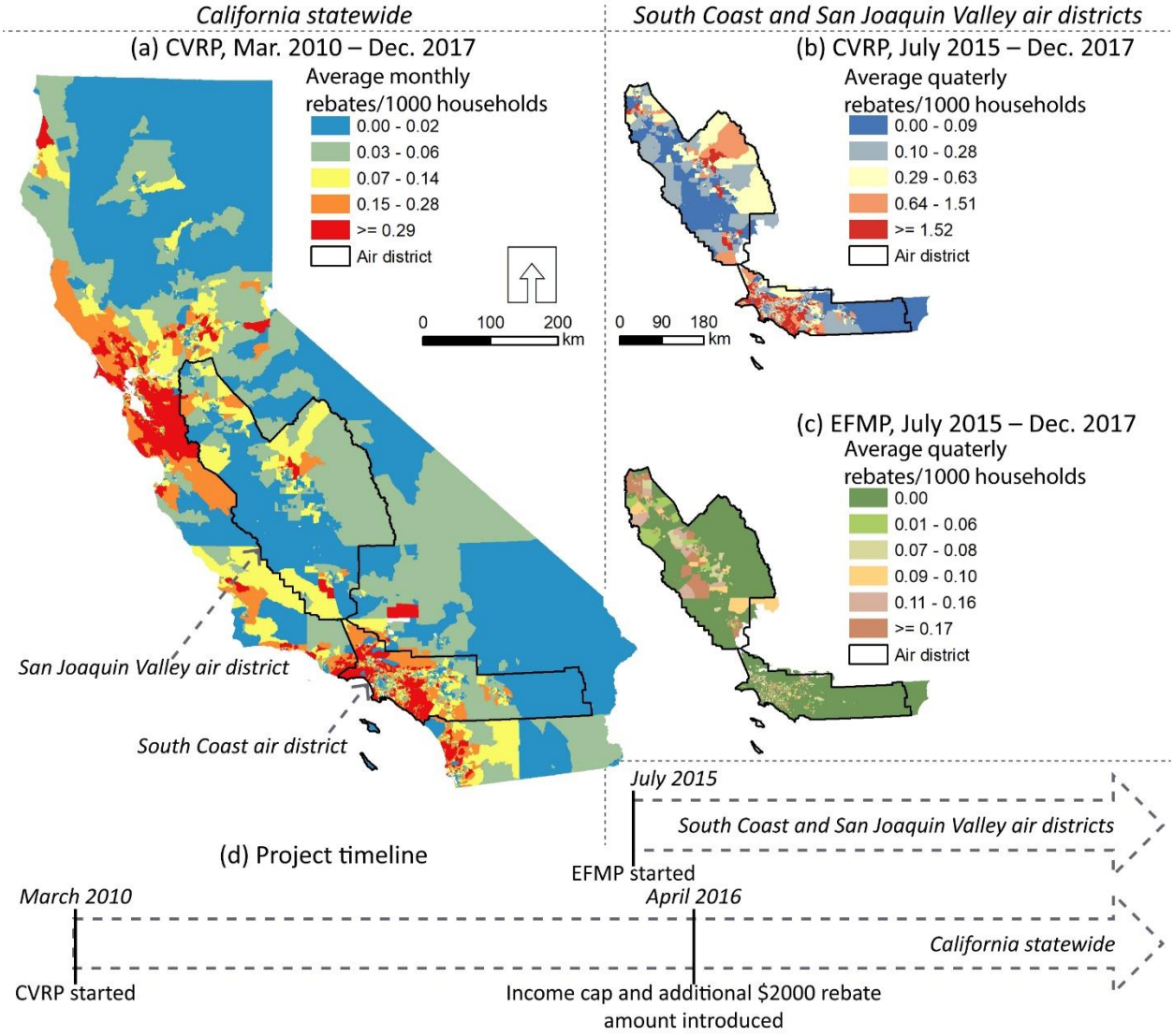



Figure 4-1. Geography, timeline, and spatial pattern of rebate allocation rates. (a) monthly Clean Vehicle Rebate Project (CVRP) rebates received per thousand households in California census tracts, averaged between March 2010 and December 2017. Average quarterly rebates received per thousand households between July 2015 and December 2017 in South Coast and San Joaquin Valley air districts census tracts, when and where Enhanced Fleet Modernization Program (EFMP) is currently implemented, are mapped for CVRP (b) and EFMP (c). (d) summarizes the timelines of the two programs, including an income cap and additional rebate amount for lower-income consumers introduced to CVRP in April 2016. Rebate allocation in (a) and (b) is classified by quintiles of all census tracts. Rebate allocation in (c) is classified by quintiles of the

census tracts receiving non-zero rebates, due to the existence of substantial amount (62%) of zero values.

A few studies have assessed California's clean vehicle rebate programs with respect to equity and find that while lower-income consumers and/or certain race and ethnicity groups tend to receive fewer rebates, such disparities can be reduced by specific policy design elements (DeShazo et al., 2017; Pierce & DeShazo, 2017; Rubin & St-Louis, 2016; Williams, 2018). A simulation-based study on CVRP shows that introducing increased rebate amount for lower-income consumers and/or caps for income and vehicle prices can increase the percentage of rebate allocated to households with incomes under $75,000 when compared with assigning rebate amount only based on vehicle technology (DeShazo et al., 2017). The remaining studies adopt an empirical approach using actual rebate allocation data. Rubin et al. comprehensively assess the relationship between CVRP rebate allocation and several community characteristics including income and race and ethnicity composition. The researchers find that rebates are disproportionately allocated to wealthier neighborhoods and that the number of rebates received per thousand households is inversely related to the proportion of Hispanic and African-American residents (Rubin & St-Louis, 2016). However, Rubin et al. focuses on the first iteration of CVRP, and therefore is not able to investigate how the income cap and additional rebate amount for lower-income consumers introduced later to the program would affect rebate allocation. Other more recent but less comprehensive studies reveal how certain program design elements improve equity in rebate allocation. Williams finds that the percentage of lower-income CVRP participants has increased since the introduction of an income cap and an additional rebate amount for lower-income consumers (Williams, 2018). The first-year operation of EFMP shows that rebates are largely distributed to consumers in the lowest income bracket of program eligibility and in areas containing disadvantaged communities (Pierce & DeShazo, 2017).

Despite those efforts above, a more comprehensive empirical study is still needed to understand how specific program design elements can improve equity in rebate allocation. Therefore, using publicly available data, we evaluate whether program design elements, including an income cap, additional rebate amounts for lower-income consumers and residents of disadvantaged communities, and expanded vehicle eligibility, can change the outcomes of CVRP and EFMP in terms of: (1) rebate allocation rates between disadvantaged and non-disadvantaged communities; (2) the association between rebate allocation rates and related community-level socioeconomic, demographic, and environmental factors. The results can inform future iterations of vehicle rebate programs to broaden participation by more diverse communities.

## 4.2 Materials and methods

We conducted two sets of analyses:

(1) a CVRP-only analysis, which investigated rebate allocation statewide between March 2010 and December 2017 and assessed whether distributional patterns of rebates changed after the introduction of an income cap and additional rebate amount for lower-income consumers in April 2016;

(2) a comparison between CVRP and EFMP, which was restricted to the South Coast and San Joaquin Valley air districts between July 2015 and December 2017, where and when the two programs overlapped.

### 4.2.1 Rebate allocation rate

We used the number of rebates issued to individual applicants per thousand households in a census tract monthly (CVRP-only analysis) or quarterly (CVRP-EFMP comparison) to measure rebate allocation rate. We downloaded publicly available participant level data between March 2010 and December 2017 for CVRP (Center for Sustainable Energy, 2019), and between July 2015 and December 2017 for EFMP (California Air Resource Board, 2019). While the CVRP and EFMP datasets have data reported till the more present time, we chose the time periods above as during these periods the other datasets used in the analysis were also available. The CVRP dataset includes information on applicant type, which we used to exclude those rebates issued to businesses, government entities, and non-profit organizations since these were not the focus of our analysis. We assumed that all EFMP rebates were assigned to individuals based on this program's design (San Joaquin Valley Air Pollution Control District, 2019; South Coast AQMD, 2019). Both CVRP and EFMP datasets provide rebate information at the census tract level. Therefore, we aggregated participant-level data and calculated the number of rebates issued to individuals per thousand households for each census tract. We retrieved household counts from a time-series of American Community Survey 5-year estimates ending at each year of our analysis (United States Census Bureau, 2019). Temporal aggregation was done by month or quarter depending on our analysis: for the CVRP-only analysis, we aggregated the number rebates by month; for the comparison between the CVRP and EFMP, we aggregated at a quarterly scale for which the EFMP rebates data were reported.

### 4.2.2 Community characteristics

Disadvantaged and non-disadvantaged communities were classified according to Cal-EPA's guidelines, which relied upon CalEnviroScreen 3.0, a mapping tool that applied an index combining measures of population vulnerability (including sensitive population and socioeconomic status) and pollution burden (including exposure to pollutants and proximity to hazardous sites) across California census tracts (August, 2016; Faust et al., 2017). Cal-EPA designates census tracts within the top $25^{th}$ percentile of CalEnviroScreen 3.0 scores, as well as 22 tracts in the top $5^{th}$ percentile of pollution burden but without reliable population vulnerability score (due to limited data or low population) as disadvantaged communities (California Environmental Protection Agency, 2017). These disadvantaged communities are prioritized in climate mitigation and adaptation projects. For example, only these communities are eligible for the additional rebate amount from the EFMP Plus-up program.

Covariates describing tract-level socioeconomic and demographic characteristics include median household income, racial/ethnic composition, percent of renter occupied housing units, average number of vehicles per household, and population density. These covariates are time-variant and collected from a time-series of American Community Survey 5-year estimates (United States Census Bureau, 2019) ending at each year of the analysis. We calculated population density by normalizing total population by developed area, rather than total area, of a census tract. Developed areas were derived from the National Land Cover Database 2011 percent developed imperviousness dataset (Multi-Resolution Land Characteristics Consortium, 2011). Developed areas provide a better characterization of population density, as most people reside and work in these areas, while census tract area, the alternative, may underestimate population density, particularly in rural areas containing a large amount of other land covers (e.g. forest and water bodies) that are not occupied by humans.

We also developed time-invariant socioeconomic covariates including the density of electric and hydrogen charging stations and urban-rural designation of the census tracts. Locations of electric and hydrogen charging stations were obtained from National Renewable Energy Laboratory (National Renewable Energy Laboratory, 2019). Each station has information on its ownership and operation status including open, planned and temporarily unavailable. We included both public and private stations as both types can facilitate clean vehicle adoption (Rubin & St-Louis, 2016). We used the stations that were currently open in the analysis for two reasons. First, we assumed that the availability of open stations could more effectively affect consumers' decisions about acquiring clean vehicles. Second, the number of open stations and all stations were highly corrected (r=0.996, measured by Spearman rank correlation, similar below), therefore using either set should produce similar results. Like population density, the density of open charging stations was calculated using developed area within a census tract to yield more accurate estimates. We treated the density of these open charging stations as a time-invariant covariate, as only 38% of stations had their opening date reported. We retrieved urban-rural designation of the census tracts from the 2010 Census (United States Census Bureau, 2010). Urban represents the census blocks (i.e. subdivisions of census tracts) that have high population density and/or a large amount of developed areas, whereas rural represents the rest of the census blocks (Ratcliffe, Burd, Holder, & Fields, 2016).

### 4.2.3 Air pollution

We used surface concentrations of nitrogen dioxide ($NO_2$) and particulate matter (PM2.5) as proxies for air pollution, to assess whether more polluted areas received more rebates. We chose $NO_2$ and PM2.5 because significant portions of these pollutants are emitted from mobile sources (Anderson et al., 2018), and clean vehicles can reduce these emissions locally. We acquired $NO_2$ concentrations from the Berkeley High-Resolution (BEHR) dataset (The Berkeley Satellite Group, 2019) and PM 2.5 concentrations from Donkelaar et al. (Atmospheric Composition Analysis Group, 2019; van Donkelaar, Martin, Li, & Burnett, 2019). The $NO_2$ dataset provides daily ambient concentration estimates gridded at $0.05° \times 0.05°$ ($\approx 5.56$ km $\times$ 5.56 km) resolution, and we first temporally aggregated these daily estimates to derive an annual mean for each grid cell. We then calculated area-weighted annual mean $NO_2$ concentrations for each census tract between 2010 and 2016, a time frame that overlapped with our study period. Finally, we calculated the average of the area-weighted annual means above as the covariate for $NO_2$ concertation. The PM 2.5 dataset provides the annual mean of daily estimates gridded at $0.01° \times 0.01°$ ($\approx 1.11$ km $\times$ 1.11 km) resolution, and a similar procedure was conducted to calculate the average of area-weighted annual mean PM 2.5 concentrations for each census tract between 2010 and 2016.

### 4.2.4 Analytical approach

We used mean comparisons and correlations to examine whether more disadvantaged communities, as defined by CalEnviroScreen 3.0, received fewer rebates. We also examined if this pattern was different between the CVRP and EFMP, and whether the pattern had changed after the introduction of an income cap and an additional rebate amount in CVRP. We conducted the mean comparisons using a permutation t-test (Millman, 2015) and measured the correlations by Spearman rank correlation coefficient. The null hypothesis in the mean comparisons was that rebate allocation rates were equal between advantaged and disadvantaged communities.

We further conducted a multivariate regression analysis to estimate associations between rebate allocation rate, our main outcome of interest, and key covariates including income, racial/ethnic composition, air pollution, and other relevant community characteristics. To do this, we built several negative binomial models, with their basic form specified as equation (1). We chose negative binomial models to account for that rebate allocation rate, measured as the number of rebates received per thousand households monthly or quarterly, was discrete and did not have values less than zero (Allison, 2009). While we did not transform most covariates, we added a quadratic term for air pollution to account for the environmental Kuznets curve that represented a non-linear relationship between pollution and income and consequently the likelihood of applying for rebates (Bechle, Millet, & Marshall, 2011; Pastor, Morello-Frosch, & Sadd, 2005). We also added an interaction term between population density and urban-rural designation of the census tracts, as previous studies have found the association between population density and participation in similar programs is different between rural and urban areas (Lachapelle, 2013).

$$logY_{i,j,m,n} = \beta X_{i,j,n} + \gamma Z_{i,j} + \delta t + c_j + d_m \quad (1)$$

where $Y_{i,j,m,n}$ is the rebate allocation rate, or number of rebates received per thousand households in census tract $i$ of county $j$ during month or quarter $m$ in year $n$, $X_{i,j,n}$ are the covariates varying by both census tract and year, $Z_{i,j}$ represents the covariates varying only by census tract, $t$ estimates a linear temporal trend measured as number of months or quarters since the start of the programs, $c_j$ denotes county fixed effects, $d_m$ is month- or quarter-of-year fixed effects.

In the CVRP-only analysis, we also sought to estimate the combined effect on rebate allocation of an income cap and additional $2000 rebate amount for lower-income consumers implemented since April 2016. In the CVRP-EFMP comparison, we assessed whether the two programs differ in the associations between rebate allocation and the covariates analyzed. To estimate the effect of the income cap and addition rebate amount for lower-income consumers, we used a dummy variable ($D_{m,n}$) indicating the presence (coded as 1) or absence (coded as of 0) of these policy design elements in the CVRP-only model (equation (2)). We interacted this dummy variable with the linear temporal trend to show any changes to this trend (Bernal, Cummins, & Gasparrini, 2017), and with median household income to estimate whether its relationship with rebate allocation was changed after introducing these policy design elements.

$$logY_{i,j,m,n} = \beta X_{i,j,n} + \gamma Z_{i,j} + \delta t + \theta D_{m,n} + \mu t D_{m,n} + \rho income_{i,j,n} D_{m,n} + c_j + d_m \quad (2)$$

where $D_{m,n}$ is a dummy variable indicating the presence (coded as 1) or absence (coded as 0) of the income cap and addition rebate amount for lower-income consumers, $income_{i,j,n}$ is median household income.

Furthermore, we accounted for the effects of place and time that were not reflected by the covariates above to reduce omitted-variable bias and to more accurately estimate model coefficients. We included county fixed effects ($c_j$) to control for unobserved county-level characteristics such as specific incentive programs and awareness of clean vehicles and the rebate programs that might differ between the 52 California counties. We included a linear temporal trend ($t$) measured as number of months or quarters since the start of the rebate

programs to control for the overall secular increase in market penetration of clean vehicles and awareness of the rebate programs, and month- or quarter-of-year fixed effects ($d_m$) to reflect monthly (CVRP-only analysis) or quarterly fluctuations (CVRP-EFMP comparison) in rebate applications and vehicle acquisitions.

We estimated the models using a pooled estimator with cluster-robust standard errors. The pooled estimator identifies coefficients for both time-variant ($X_{i,j,n}$) and -invariant ($Z_{i,j}$) covariates. We used cluster-robust standard errors to account for that our observations were clustered by census tracts within counties. Such error leads to larger but more accurate confidence intervals and consequently more conservative statistical significance for the coefficients (Allison, 2009; Colin Cameron & Miller, 2015).

We used the average marginal effect to evaluate the influence of certain covariates on the allocation of rebates. To obtain the average marginal effect we first calculated marginal effect at every observed value of a covariate and then averaged across the resulting effect estimates (Leeper, 2018).

## 4.3 Results and discussion

### 4.3.1 Rebate allocation rates and community disadvantage

Disadvantaged communities had significantly lower rebate allocation rates from the CVRP but higher rebate allocation rates from the EFMP compared to their more advantaged counterparts (Figure 4-2). Averaged between March 2010 and December 2017, disadvantaged communities on average received 0.05 (standard deviation (SD) = 0.07) CVRP rebates per thousand households monthly, whereas non-disadvantaged communities on average received 0.22 (SD = 0.25) CVRP rebates per thousand households monthly. The implementation of an income cap and additional $2000 rebate amount in April 2016 helped to reduce but did not close the gap between disadvantaged and non-disadvantaged communities in CVRP rebate allocation rates. Before April 2016, disadvantaged communities on average received 0.04 (SD = 0.05) rebates per thousand households monthly, which was 22% of the 0.18 (SD = 0.23) rebates per thousand households monthly received by non-disadvantaged communities; After April 2016, disadvantaged communities on average received 0.10 (SD = 0.14) rebates per thousand households monthly, 30% of the 0.33 (SD = 0.34) rebates per thousand households monthly in non-disadvantaged communities ((Figure 4-2(a)).

Averaged between July 2015 and December 2017, in the South Coast and San Joaquin Valley air districts where the EFMP and CVRP overlapped, CVRP on average issued 0.29 (SD = 0.36) rebates to disadvantaged communities and 1.10 (SD = 0.92) rebates to non-disadvantaged communities per thousand households quarterly, whereas EFMP on average issued 0.07 (SD=0.11) rebates to disadvantaged communities and 0.03 (SD = 0.07) rebates to non-disadvantaged communities per thousand households per quarter ((Figure 4-2(b)). In sum, CVRP differentially benefitted non-disadvantaged communities, even after implementing an income cap and an additional rebate amount for lower-income consumers, whereas EFMP differentially benefitted disadvantaged communities.

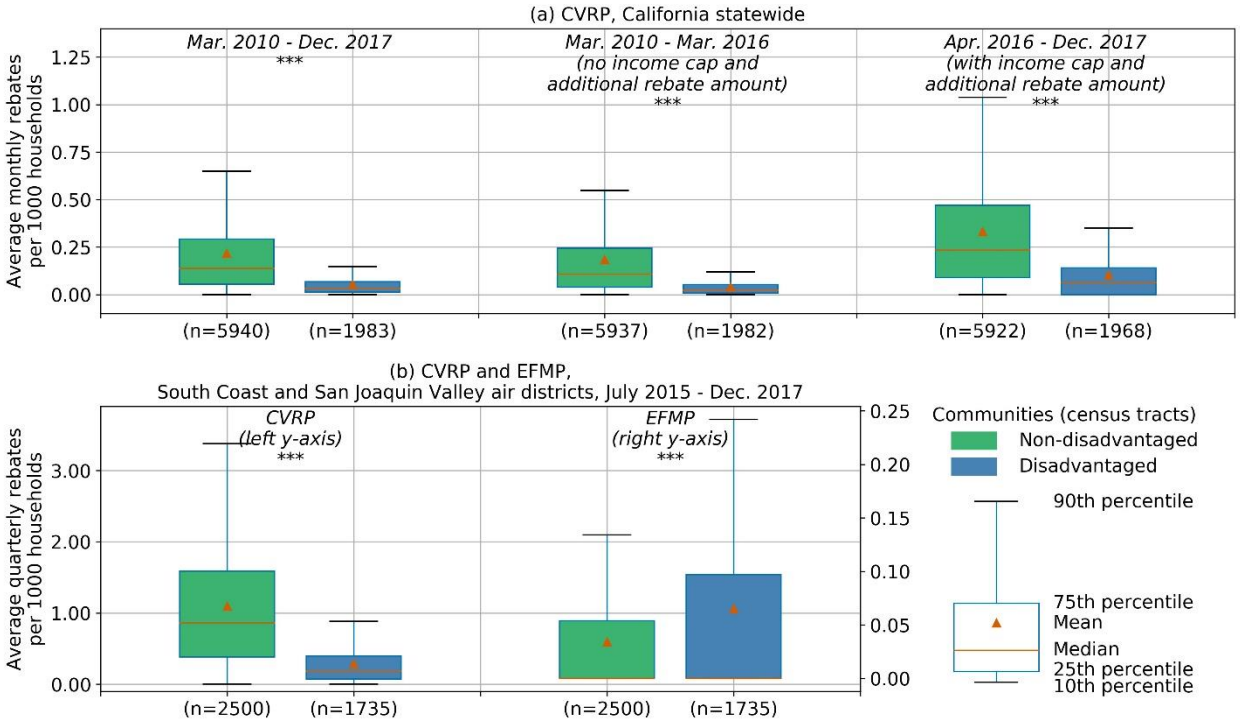



Figure 4-2. Rebate allocation rates between non-disadvantaged and disadvantaged communities. (a) California-wide Clean Vehicle Rebate Project (CVRP) between March 2010 and December 2017 and before and after implementation of an income cap and additional $2000 rebate amount for lower-income consumers in April 2016. (b) Enhanced Fleet Modernization Program (EFMP) in South Coast and San Joaquin Valley air districts between July 2015 and December 2017. CVRP rebate allocation rate in the same region and period is also shown. Rebate allocation rates is number of rebates received by individual applicants per thousand households monthly (a) or quarterly (b), averaged over its corresponding period. 25% of California census tracts are designated as disadvantaged using CalEnvrioScreen 3.0 (August, 2016; Faust et al., 2017). *** indicate significant mean difference between non-disadvantaged and advantaged communities at 99% confidence interval which was measured by a permutation t-test (Millman, 2015).

We also found correlations between CVRP and EFMP rebate allocation rates and CalEnviroScreen 3.0 score; correlations were stronger with population vulnerability score than with pollution burden score (Figure 4-3). Despite an income cap and additional $2000 rebate amount for lower-income consumers in the CVRP, the program's rebate allocation rates were consistently negatively correlated with CalEnviroScreen 3.0 scores, indicating that rebates were more often provided to less disadvantaged communities. In contrast, the EFMP rebate allocation rates were positively correlated with CalEnviroScreen 3.0 scores, likely due to the program's expanded vehicle eligibility, more income-tiered rebate amount, and an extra rebate amount offered to consumers of disadvantaged communities. In addition, rebate allocation correlated more strongly with population vulnerability than with environmental burden. This pattern likely reflects the rebate program's predominant focus on consumer's income rather than other socioeconomic and environmental characteristics. One exception is EFMP Plus-up which also considers the communities' environmental burden.

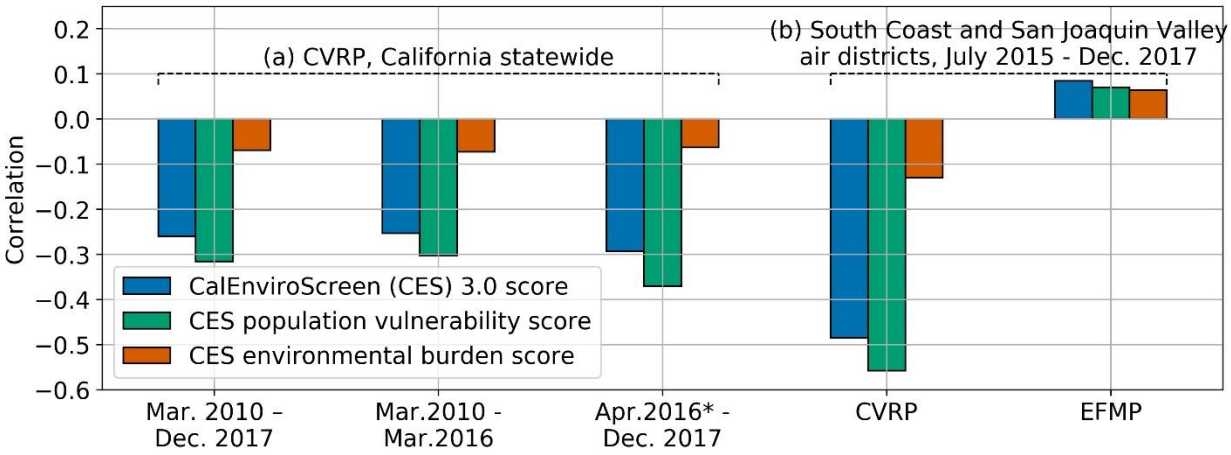



Figure 4-3. Correlations between CalEnviroScreen 3.0 scores and rebate allocation rates. Rebate allocation rate is number of rebates received by individual applicants per thousand household monthly (a) or quarterly (b), averaged over its corresponding period. CVRP: Clean Vehicle Rebate Project. EFMP: Enhanced Fleet Modernization Program. * detonates an income cap and additional $2000 rebate amount for lower-income consumers that was implemented in CVRP since April 2016. Correlations were measured by Spearman rank correlation.

### 4.3.2 Multivariate models for the Statewide CVRP program

Results for the statewide CVRP regression models are shown in Table 4-1. Model 1 estimated the associations between rebate allocation rate, community socioeconomic and demographic characteristics, and ambient air pollution concentrations, after controlling for a linear temporal trend, county and month-of-year fixed effects and clustering of our observations by counties. Model 2 additionally estimated the effect of the April 2016 implementation of an income cap and additional $2000 rebate amount for lower-income consumers in the CVRP by including a dummy variable for pre- or post-April 2016. An interaction term between this dummy variable and median household income was included to determine whether these changes in program design affected the association between rebate allocation rate and median household income. Both models assumed that the observations and model errors clustered by counties, resulting in broader confidence intervals and more conservative results than other model specifications assuming independent model errors or clustered errors by census tracts. All the covariates, other than Time and Income cap and increased rebate amount, were standardized to make their coefficients comparable with each other, to identify the covariates that were more influential on rebate allocation.

Table 4-1. Negative binominal regression models for CVRP rebate allocation rate, statewide, March 2010 – December 2017

| Covariates | Model 1 | Model 2 |
|---|---|---|
| Median household income | 0.318*** | 0.391*** |
| | (0.242, 0.393) | (0.320, 0.462) |
| Income cap and increased rebate | | 2.179*** |
| | | (1.710, 2.648) |
| Median household income × Income cap | | -0.179*** |
| and increased rebate | | (-0.198, -0.159) |

| | | |
|---|---|---|
| % non-Hispanic Black | -0.137*** | -0.130*** |
| | (-0.164, -0.110) | (-0.158, -0.102) |
| % Hispanic | -0.478*** | -0.467*** |
| | (-0.531, -0.425) | (-0.521, -0.413) |
| % non-Hispanic Asian/Pacific Islander | 0.040 | 0.046 |
| | (-0.023, 0.104) | (-0.015, 0.107) |
| Average NO2 concentration | 0.162 | 0.148 |
| | (-0.043, 0.367) | (-0.047, 0.343) |
| Average NO2 concentration squared | -0.153*** | -0.146*** |
| | (-0.255, -0.051) | (-0.243, -0.049) |
| Average PM2.5 concentration | 0.062 | 0.069 |
| | (-0.048, 0.173) | (-0.039, 0.176) |
| Average PM2.5 concentration squared | -0.021 | -0.019 |
| | (-0.070, 0.028) | (-0.068, 0.029) |
| Density of operating electric and hydrogen charging stations | 0.006 | 0.006 |
| | (-0.014, 0.025) | (-0.014, 0.025) |
| % Renter occupied housing units | -0.060*** | -0.050** |
| | (-0.101, -0.019) | (-0.089, -0.012) |
| Population density | -0.052 | -0.043 |
| | (-0.129, 0.026) | (-0.122, 0.035) |
| Urban | -0.215*** | -0.208*** |
| | (-0.279, -0.150) | (-0.273, -0.143) |
| Urban × Population density | -0.075* | -0.077** |
| | (-0.151, 0.001) | (-0.151, -0.003) |
| Average vehicles per household | 0.038 | 0.045 |
| | (-0.023, 0.100) | (-0.015, 0.106) |
| Time (number of months since the start) | 0.025*** | 0.043*** |
| | (0.024, 0.026) | (0.041, 0.044) |
| Time × Income cap and increased rebate | | -0.037*** |
| | | (-0.043, -0.031) |
| Pseudo R2 | 0.199 | 0.220 |
| Confidence interval | County-clustered | County-clustered |
| Spatial and temporal scale | Monthly, Mar. 2010 – Dec. 2017, all CA census tracts | |
| Number of tract-month observations | 740010 | 740010 |
| Month-of-year fixed effects | Yes | Yes |
| County fixed effects | Yes | Yes |

Note: Coefficients should be interpreted as the change in ln(monthly rebates received per thousand households) for one unit change in the corresponding covariate, when holding other covariates constant. All covariates other than Time and Income cap and increased rebate are standardized. Covariates on income, race and ethnicity, renter occupied housing units, and population density are time-varying on an annual basis. The covariate on the income cap and additional rebate amount for lower income consumers is time-varying on a monthly basis. Covariates on charging station density, urban status, and air pollution are time-invariant. *, **, and *** indicate significant at 90%, 95%, and 99% confidence intervals. 95% confidence intervals are in the parenthesis.

As we expected, median household income had a positive and statistically significant association with rebate allocation rate and was the most influential variable given its larger coefficient size in the models. This is consistent with CVRP's design that consumers need upfront capital to acquire a new eligible vehicle before receiving the rebate. Compared with ICE

vehicles, the additional cost of clean vehicles likely makes lower-income consumers less willing to pay for them (Erdem, Şentürk, & Şimşek, 2010; Poder & He, 2017; Potoglou & Kanaroglou, 2007). A similar phenomenon has also been observed in other rebate programs that encourage the adoption of environmental-friendly technologies, such as rooftop solar systems, which require a substantial upfront investment (Briguglio & Formosa, 2017; Soskin & Squires, 2013).

The models also show disparities in rebate allocation with respect to community racial/ethnic composition. Increases in the percentages of non-Hispanic Black and Hispanic populations were significantly associated with lower rebate allocation rates in the models. In contrast, increase in the percentage of non-Hispanic Asian/Pacific Islander population was associated with an increase in rebate allocation rates; however, this association was not statistically significant. A similar disparity in CVRP rebate allocation for Non-Hispanic Blacks and Hispanics has also been previously reported (Rubin & St-Louis, 2016). Racial disparities in clean technology access are not unique to CVRP. For example, Kwan finds that across the United States, communities with higher percentages of Black and Asian residents have lower rates of the residential solar PV installation, whereas the installation is higher for communities with more Hispanics (Kwan, 2012). We suggest that the racial and ethnic disparity in rebate allocation rate is partially resulted from differences in wealth, as in our analysis median household income is positively correlated with the percentage of non-Hispanic Asians/Pacific Islanders, but negatively correlated with the percentages of non-Hispanic Blacks and Hispanics. Other factors contributing to the disparity may include the racial/ethnical group's differed perceptions about clean vehicles and preference for ICE vehicles, and varied outreach levels of the rebate programs in different communities.

We observed a non-linear relationship between ambient air pollutant concentrations and CVRP rebate allocation rates. In census tracts with lower levels of average $NO_2$ and PM2.5 concentrations, increases in ambient concentrations of either pollutant were associated with an increase in rebate allocation rates. However, the positive associations between air pollutant concentrations and rebate allocation rates become negative as average $NO_2$ and PM2.5 concentrations increase (Figure S1). Associations between rebate allocation rates and air pollutants were statistically significant for $NO_2$ but mostly insignificant for PM2.5. Therefore, in general, it can be expected that CVRP rebate allocation rates are higher in areas with moderate levels of air pollution, but lower in areas with very high or low air pollution levels.

Other covariates generally exhibited relationships with rebate allocation in the expected direction. The density of electric and hydrogen charging stations showed a positive association with rebate allocation rates in our models. However, the associations are statistically insignificant. Prior research has found that charging station is one of the major concerns about using electric vehicles (Egbue & Long, 2012), and can even be the most influential factor for nationwide electric vehicle adoption (Sierzchula, Bakker, Maat, & van Wee, 2014). The insignificant associations observed in our models may have resulted from our ability to estimate the density of electric and hydrogen charging stations only at the census tract level rather than the individual level (i.e. at home and at work). Percent renter occupied housing units showed statistically significant negative associations with rebate allocation rates, which could be explained by that compared with homeowners, renters were less able or willing to install charging facilities at their homes, therefore felt less convenient to use electric vehicles. This also implies the positive role of charging facilities in clean vehicle adoption.

The association between rebate allocation and population density depended on whether a census tract was rural or urban (column 1-2 of Table S2). In rural census tracts, increases in

population density were insignificantly associated with greater rebate allocation. Among urban census tracts, higher population density was associated with a statistically significant decrease in rebate allocation. A similar trend was observed in an early vehicle retirement program in Quebec, in which the rate of participation was higher in low-density metropolitan areas (Lachapelle, 2013). In urban areas, higher population density may indicate more alternative transport, such as public transit, which reduces the demand for personal vehicles. Densely populated urban areas also have limited off-street parking for installing electric vehicle chargers, making operating PHEVs and BEVs more challenging.

The average number of vehicles per household showed a positive but statistically insignificant association with rebate allocation rate. This is in contrast to Rubin et al. (2016), who found that CVRP rebate allocation between 2010 and 2015 was negatively associated with the average number of vehicles per household (Rubin & St-Louis, 2016). However, our results are consistent with research suggesting that consumers tend view electric vehicles as secondary cars when concerned about their battery range (Skippon & Garwood, 2011; Tamor & Milačić, 2015).

Finally, we found a statistically significant and increasing temporal trend in rebate allocation, which could be explained by the growing market penetration of clean vehicles, likely due at least in part to the rebate programs themselves. Combined PHEV and BEV sales in the U.S. increased by 446% from 2011 (17,763 vehicles) to 2013 (97,102 vehicles), and incentives such as rebate likely had strong positive influences on this increase (Zhou et al., 2015). A 2013-2015 CVRP consumer survey shows that receiving a rebate is the most important factor in deciding to purchase a PHEV or BEV, with 41% and 50% of the respondents indicating they would not have purchased a PHEV or BEV without the rebate (Johnson, Williams, Hsu, & Anderson, 2017). We also hypothesize that the temporal increase in rebate allocation may be partially due to a growing awareness of CVRP over time.

### 4.3.3 The Effects of the Income Cap and Additional Rebate Amount on CVRP Rebate Allocation

Implementing equity-promoting policy design elements, i.e. an income cap and additional $2000 rebate amount for lower-income consumers, in April 2016 reduced the association between CVRP rebate allocation rate and neighborhood income. Based on Model 2, the average marginal effect of these equity-promoting policy design elements was to reduce the rebate allocation rate by 0.099 rebates (95% confidence interval: -0.137 to -0.061, p-value<0.01) per thousand households per month, suggesting that the income cap may have excluded high-income consumers that was not offset by an increase in low-income consumers attracted by the increased rebate amount. Additionally, our estimates showed that without the income cap and additional rebate amount, a one-standard-deviational increase in neighborhood median household income was associated with 0.126 more rebates per thousand households per month. With the income cap and additional rebate amount, the same increase in median household income was only associated with 0.048 more rebates per thousand households per month (Table 4-2). Therefore, the association between median household income and rebate allocation rate was attenuated significantly with the implementation of an income cap and an additional rebate amount for lower-income consumers. Our findings are consistent with simulation studies suggesting an income cap and income-base rebate amounts increase the proportion of rebates allocated to households with incomes under $75,000 (DeShazo et al., 2017).

Table 4-2 Average marginal effects of median household income on CVRP rebate allocation rate, with and without an income cap and additional $2000 rebate amount for lower-income consumers implemented since April 2016

| Covariate | Income cap and additional rebate amount for low income households | Average marginal effect |
|---|---|---|
| Median household income | If not implemented | 0.126*** (0.101, 0.151) |
| | If implemented | 0.048*** (0.030, 0.065) |
| | Difference (implemented – not implemented) | -0.079*** (-0.090, -0.067) |

Note: average marginal effect is the average change in rebate allocation rate for one-standard-deviational change in median household income. Rebate allocation rate is the number of rebates received per thousand households monthly. This average marginal effect is also conditioned on the implementation of an income cap and additional $2000 rebate amount for lower-income consumers since April 2016. All other covariates are held at their observed values in this calculation. The average marginal effect here was estimated based on model 2 of Table 4-1. *, **, and *** indicate significant at 90%, 95%, and 99% confidence intervals. 95% confidence intervals are in the parenthesis.

### 4.3.4 Comparison between CVRP and EFMP

The models for CVRP and EFMP rebate allocation rates in the South Coast and San Joaquin Valley air districts between July 2015 and December 2017 are shown in Table 4-3. The models control for county and quarter fixed effects, as well as county clustered standard errors. For CVRP, rebate allocation follows the statewide pattern: the equivalent model for the two air districts (Model 3 in Table 4-3) and the entire state (Model 1 in Table 4-1) show similar directions of the covariates' effects, and the models produce similar average marginal effects for population density (column 3 in Table S2) and air pollutants (Figure S2) on rebate allocation.

Unlike CVRP, median household income was associated with a reduction in EFMP rebate allocation rate, and an increase in the percentage of Hispanics was associated with an increase in rebate allocation rate (Model 2 in Table 4-3). In addition, there was a positive but statistically insignificant association between EFMP rebate allocation rate and percentage of non-Hispanic Black population in a census tract. These different results between CVRP and EFMP can likely be explained by their different designs: while CVRP offered an additional $2000 to lower-income consumers during the study period, EFMP offered higher rebate amounts to lower-income consumers, expanded the set of eligible vehicles, and gave additional rebates to consumers living in socioeconomically and environmentally disadvantaged communities.

Another difference between EFMP and CVRP in the two air districts is that, while there was a positive but statistically insignificant association between charging station density and rebate allocation for CVRP, the association was negative and statistically significant for EFMP. The reasons for this are unknown but may be partially explained by that non-plug-in vehicles, which are not dependent on charging stations, are eligible for EFMP, but not CVRP rebates.

The associations between other covariates and rebate allocation are similar between CVRP and EFMP. For both programs, moderate levels of air pollution were generally associated

with higher rebate allocation rates (Figure S2). By design EFMP Plus-up program should attract more consumers in disadvantaged communities that have higher pollution burden and population vulnerability. Our results indicated that this assumption likely held for $NO_2$: compared with CVRP, EFMP rebates were more likely to go to areas with higher $NO_2$ concentration, given the tipping point, where the average marginal effect of air pollution on rebate allocation rates shifts to negative, was in more polluted areas for EFMP in Figure S2(a). However, we could not conclude for PM2.5, as the estimates for its average marginal effect on EFMP rebate allocation rates were all statistically insignificant (Figure S2(b)). Both CVRP and EFMP also show positive associations between rebate allocation rates and vehicle ownership. This association is however only statistically significant for EFMP. In the case of EFMP, which includes a retire-and-replace program, this may be because households with more vehicles are more likely to own vehicles eligible for retirement. CVRP, however, may include first-time buyers that don't have any vehicles, therefore vehicle ownership had weak and statistically significant association with rebate allocation rates.

Table 4-3. Regression for CVRP and EFMP rebate allocation rate per thousand households in South Coast and San Joaquin air districts, quarterly between July 2015 and December 2017

| | Model 3 | Model 4 |
|---|---|---|
| Median household income | 0.273*** | -0.520*** |
| | (0.146, 0.400) | (-0.780, -0.260) |
| % non-Hispanic Black | -0.107*** | 0.032 |
| | (-0.146, -0.068) | (-0.054, 0.119) |
| % Hispanic | -0.467*** | 0.339** |
| | (-0.555, -0.380) | (0.038, 0.640) |
| % non-Hispanic Asian/Pacific | 0.051* | 0.393*** |
| Islander | (-0.001, 0.102) | (0.290, 0.497) |
| Average NO2 concentration | -0.027 | 0.305*** |
| | (-0.093, 0.039) | (0.123, 0.486) |
| Average NO2 concentration | -0.182*** | -0.399*** |
| squared | (-0.219, -0.146) | (-0.608, -0.191) |
| Average PM2.5 concentration | 0.018 | 0.036 |
| | (-0.041, 0.077) | (-0.145, 0.217) |
| Average PM2.5 concentration | -0.021* | 0.006 |
| squared | (-0.044, 0.001) | (-0.033, 0.046) |
| Density of operating electric and | 0.005 | -0.113*** |
| hydrogen charging stations | (-0.018, 0.028) | (-0.161, -0.065) |
| % Renter occupied housing units | -0.026 | 0.010 |
| | (-0.096, 0.045) | (-0.136, 0.156) |
| Population density | -0.269 | 0.121 |
| | (-0.686, 0.147) | (-0.260, 0.502) |
| Urban | -0.062 | -0.118 |
| | (-0.240, 0.116) | (-0.489, 0.253) |
| Urban × Population density | 0.185 | -0.287 |
| | (-0.217, 0.588) | (-0.668, 0.094) |
| Average vehicles per household | 0.081 | 0.263** |
| | (-0.024, 0.185) | (0.009, 0.517) |
| Time (number of quarters since | 0.011* | 0.001 |
| the start) | (-0.000, 0.021) | (-0.066, 0.069) |
| Pseudo $R^2$ | 0.176 | 0.090 |

| | | |
|---|---|---|
| Confidence interval | County-clustered | County-clustered |
| Scale | Quarterly, July 2015 – Dec. 2017, census tracts in the South Coast and San Joaquin Valley air districts | |
| Number of observations | 42260 | 42260 |
| Month fixed effects | Yes | Yes |
| County fixed effects | Yes | Yes |

Note: Coefficients should be interpreted as the change in ln(quarterly rebates received per thousand households) for one unit change in the corresponding covariate, when holding other covariates constant. All covariates other than Time are standardized. Covariates on income, race and ethnicity, renter occupied housing units, and population density are time-varying on an annual basis. Covariates on charging station density, urban status, and air pollution are time-invariant. *, **, and *** indicate significant at 90%, 95%, and 99% confidence intervals. 95% confidence intervals are in parenthesis.

To summarize, our analysis suggests that different vehicle rebate design elements affect equity-based outcomes in terms of the associations between rebate allocation rates and community socioeconomic status, racial/ethnic composition, and air pollution. The first iteration of CVRP assigns rebate amount only based on vehicle technology (i.e. PHEV, BEV, or FCEV, as shown in Table S1), therefore it can be assumed that wealthier consumers will more likely apply, which has been confirmed by our analysis and others (Rubin & St-Louis, 2016). Introducing an income cap and an addition $2000 rebate amount for lower-income consumers to CVRP has lowered the importance of income in rebate allocation. The inclusion of a retire-and-replace element, extending eligibility to more fuel-efficient used cars, increasing rebate amount for lower income households and residents of disadvantaged communities – all features of the EFMP program – appears to have had a stronger effect in extending participation to lower-income and Hispanic communities.

In addition to the policy designs described above, other policy designs and incentives may further improve the equity in acquiring clean vehicles. For potential clean vehicle consumers and rebate applicants with lower income, the issue is to find the upfront capital for acquiring the vehicle (Rubin & St-Louis, 2016). To resolve this issue, in 2018 CVRP started to preapprove rebate applications in San Diego County (Center for Sustainable Energy, 2017). These pre-approved rebates are likely to make it easier and faster for consumers to acquire clean vehicles.

While this study uses rebate allocations to evaluate the distribution of the benefits of clean vehicle rebate programs, this outcome does not capture a perhaps more essential aspect of the programs: the actual reductions in mobile-source emissions of GHG and the co-pollutants from the vehicles receiving rebates. To our knowledge, no information on driving behavior, such as vehicle mile traveled and typical routes, have been collected from those vehicles. Therefore, it is hardly possible to estimate the amount of reductions in mobile-source emissions due to the rebate programs, and how these reductions are distributed between disadvantaged and non-disadvantaged communities. Collection of participant-level data on driving behavior is critical in estimating these reductions. Combining existing data including rebate allocation and air pollution, however, may not be able to estimate the reduction in air pollution due to the rebate programs for two reasons. First, where the rebates are allocated is not a strong indicator of where and how the vehicles are operated. Second, observed reduction in air pollution may be largely resulted from other efforts, such as improved fuel quality. Research has also shown that only substantial replacement of conventional vehicles with clean vehicles (e.g. a replacement rate of

25%) can result in significant reduction in air pollution (Ferrero, Alessandrini, & Balanzino, 2016; Soret, Guevara, & Baldasano, 2014). Therefore, participant-level data including driving behavior is needed to better estimate air quality related benefits of these rebate programs and their equity implications.

**Acknowledgements**

This work was funded by the California Office of Environmental Health Hazard Assessment (https://oehha.ca.gov/; Contract #16-E0012-2).

**References**

Allison, P. (2009). *Fixed Effects Regression Models*. https://doi.org/10.4135/9781412993869

American Lung Association. (2018). *State of the Air 2018*. Retrieved from American Lung Association website: https://www.lung.org/local-content/california/our-initiatives/state-of-the-air/2018/state-of-the-air-2018.html

Anderson, C. M., Kissel, K. A., Field, C. B., & Mach, K. J. (2018). Climate Change Mitigation, Air Pollution, and Environmental Justice in California. *Environmental Science & Technology*, *52*(18), 10829–10838. https://doi.org/10.1021/acs.est.8b00908

Atmospheric Composition Analysis Group. (2019). Surface PM 2.5 North American Regional Estimates. Retrieved February 8, 2019, from http://fizz.phys.dal.ca/~atmos/martin/?page_id=140#V4.NA.02

August, L. (2016, December 29). CalEnviroScreen 3.0 [Text]. Retrieved February 11, 2019, from OEHHA website: https://oehha.ca.gov/calenviroscreen/report/calenviroscreen-30

Bernal, J. L., Cummins, S., & Gasparrini, A. (2017). Interrupted time series regression for the evaluation of public health interventions: A tutorial. *International Journal of Epidemiology*, *46*(1), 348–355. https://doi.org/10.1093/ije/dyw098

Briguglio, M., & Formosa, G. (2017). When households go solar: Determinants of uptake of a Photovoltaic Scheme and policy insights. *Energy Policy*, *108*, 154–162. https://doi.org/10.1016/j.enpol.2017.05.039

California Air Resource Board. (2018a). *California Greenhouse Gas Emissions for 2000 to 2016*. Retrieved from California Air Resource Board website: https://www.arb.ca.gov/cc/inventory/pubs/reports/2000_2016/ghg_inventory_trends_00-16.pdf

California Air Resource Board. (2018b, December 6). Enhanced Fleet Modernization Program. Retrieved January 15, 2019, from https://www.arb.ca.gov/msprog/aqip/efmp/efmp.htm

California Air Resource Board. (2019). EFMP Retire and Replace Participant-level data. Retrieved February 11, 2019, from https://www.arb.ca.gov/msprog/aqip/efmp/stats/participant_data_q3_2018.xlsx

California Environmental Protection Agency. (2017). *Designation of Disadvantaged Communities Pursuant to Senate Bill 525 (de León)*. Retrieved from https://calepa.ca.gov/wp-content/uploads/sites/6/2017/04/SB-535-Designation-Final.pdf

Center for Sustainable Energy. (2016, May 31). Income Eligibility. Retrieved January 14, 2019, from Clean Vehicle Rebate Project website: https://cleanvehiclerebate.org/eng/income-eligibility

Center for Sustainable Energy. (2017, November 8). Preapproval for CVRP Rebates. Retrieved March 18, 2019, from Clean Vehicle Rebate Project website: https://cleanvehiclerebate.org/eng/rebatenow

Center for Sustainable Energy. (2019, January 14). California Air Resources Board Clean Vehicle Rebate Project, Rebate Statistics. Retrieved January 21, 2019, from https://cleanvehiclerebate.org/eng/rebate-statistics

Colin Cameron, A., & Miller, D. L. (2015). A Practitioner's Guide to Cluster-Robust Inference. *Journal of Human Resources*, *50*(2), 317–372. https://doi.org/10.3368/jhr.50.2.317

DeShazo, J. R. (2016). Improving Incentives for Clean Vehicle Purchases in the United States: Challenges and Opportunities. *Review of Environmental Economics and Policy*, *10*(1), 149–165. https://doi.org/10.1093/reep/rev022

DeShazo, J. R., Sheldon, T. L., & Carson, R. T. (2017). Designing policy incentives for cleaner technologies: Lessons from California's plug-in electric vehicle rebate program. *Journal of Environmental Economics and Management*, *84*, 18–43. https://doi.org/10.1016/j.jeem.2017.01.002

Egbue, O., & Long, S. (2012). Barriers to widespread adoption of electric vehicles: An analysis of consumer attitudes and perceptions. *Energy Policy*, *48*, 717–729. https://doi.org/10.1016/j.enpol.2012.06.009

Erdem, C., Şentürk, İ., & Şimşek, T. (2010). Identifying the factors affecting the willingness to pay for fuel-efficient vehicles in Turkey: A case of hybrids. *Energy Policy*, *38*(6), 3038–3043. https://doi.org/10.1016/j.enpol.2010.01.043

Faust, J., August, L., Bangia, K., Galaviz, V., Leichty, J., Prasad, S., … Zeise, L. (2017). *CalEnviroScreen3*. Retrieved from Office of Environmental Health Hazard Assessment website: https://oehha.ca.gov/media/downloads/calenviroscreen/report/ces3report.pdf

Ferrero, E., Alessandrini, S., & Balanzino, A. (2016). Impact of the electric vehicles on the air pollution from a highway. *Applied Energy*, *169*, 450–459. https://doi.org/10.1016/j.apenergy.2016.01.098

Hardman, S., Chandan, A., Tal, G., & Turrentine, T. (2017). The effectiveness of financial purchase incentives for battery electric vehicles – A review of the evidence. *Renewable and Sustainable Energy Reviews*, *80*, 1100–1111. https://doi.org/10.1016/j.rser.2017.05.255

Johnson, C., Williams, B., Hsu, C., & Anderson, J. W. (2017). *The Clean Vehicle Rebate Project: Summary Documentation of the Electric Vehicle Consumer Survey, 2013–2015 Edition*. San Diego, CA: Center for Sustainable Energy.

Kwan, C. L. (2012). Influence of local environmental, social, economic and political variables on the spatial distribution of residential solar PV arrays across the United States. *Energy Policy*, *47*, 332–344. https://doi.org/10.1016/j.enpol.2012.04.074

Lachapelle, U. (2013). Participation and Incentive Choice of Participants in an Early Vehicle Retirement Program in Quebec, Canada. *Transportation Research Record: Journal of the Transportation Research Board*, *2375*(1), 8–17. https://doi.org/10.3141/2375-02

Leeper, T. J. (2018). *Interpreting Regression Results using Average Marginal Effects with R's margins*. 32.

Millman, K. J. (2015). *permute—A Python package for permutation tests and confidence sets* (University of California, Berkeley). Retrieved from http://www.jarrodmillman.com/publications/millman2015thesis.pdf

Multi-Resolution Land Characteristics Consortium. (2011). NLCD 2011 Percent Developed Imperviousness (CONUS). Retrieved January 25, 2019, from https://prd-tnm.s3.amazonaws.com/StagedProducts/NLCD2011/Urban_Imperviousness/CONUS/nlcd_2011_impervious_2011_edition_2014_10_10.zip

National Renewable Energy Laboratory. (2019, February 7). Alternative Fuel Stations. Retrieved February 7, 2019, from https://developer.nrel.gov/docs/transportation/alt-fuel-stations-v1/

Pierce, G., & DeShazo, J. R. (2017). *Design and Implementation of the Enhanced Fleet Modernization Plus-Up Pilot Program*. Retrieved from UCLA Luskin Center for Innovation website: http://innovation.luskin.ucla.edu/sites/default/files/Design%20and%20Implementation%20of%20the%20Enhanced%20Fleet%20Modernization%20Plus-Up%20Pilot%20Program.pdf

Poder, T. G., & He, J. (2017). Willingness to pay for a cleaner car: The case of car pollution in Quebec and France. *Energy*, *130*, 48–54. https://doi.org/10.1016/j.energy.2017.04.107

Potoglou, D., & Kanaroglou, P. S. (2007). Household demand and willingness to pay for clean vehicles. *Transportation Research Part D: Transport and Environment*, *12*(4), 264–274. https://doi.org/10.1016/j.trd.2007.03.001

Ratcliffe, M., Burd, C., Holder, K., & Fields, A. (2016). *Defining Rural at the U.S. Census Bureau*. 8.

Rubin, D., & St-Louis, E. (2016). Evaluating the Economic and Social Implications of Participation in Clean Vehicle Rebate Programs. *Transportation Research Record: Journal of the Transportation Research Board*, *2598*, 67–74. https://doi.org/10.3141/2598-08

San Joaquin Valley Air Pollution Control District. (2019). Drive Clean in the San Joaquin, Replace, Restrictions. Retrieved May 8, 2019, from Drive Clean in the San Joaquin | Replace website: http://valleyair.org/drivecleaninthesanjoaquin/replace

Sierzchula, W., Bakker, S., Maat, K., & van Wee, B. (2014). The influence of financial incentives and other socio-economic factors on electric vehicle adoption. *Energy Policy*, *68*, 183–194. https://doi.org/10.1016/j.enpol.2014.01.043

Skippon, S., & Garwood, M. (2011). Responses to battery electric vehicles: UK consumer attitudes and attributions of symbolic meaning following direct experience to reduce psychological distance. *Transportation Research Part D: Transport and Environment*, *16*(7), 525–531. https://doi.org/10.1016/j.trd.2011.05.005

Snelling, A. (2018). *Evaluation of State Plug-In Electric Vehicle Purchase Incentive Programs: What Drives Vehicle Uptake?* Retrieved from http://dspace-prod.mse.jhu.edu:8080/handle/1774.2/59880

Soret, A., Guevara, M., & Baldasano, J. M. (2014). The potential impacts of electric vehicles on air quality in the urban areas of Barcelona and Madrid (Spain). *Atmospheric Environment*, *99*, 51–63. https://doi.org/10.1016/j.atmosenv.2014.09.048

Soskin, M., & Squires, H. (2013). Homeowner willingness to pay for rooftop solar electricity generation. *Environmental Economics*, *4*(1), 10.

South Coast AQMD. (2019). Replace Your Ride, Eligibility Overview. Retrieved May 8, 2019, from https://xappprod.aqmd.gov/RYR/Home/Eligibility

Tamor, M. A., & Milačić, M. (2015). Electric vehicles in multi-vehicle households. *Transportation Research Part C: Emerging Technologies*, *56*, 52–60. https://doi.org/10.1016/j.trc.2015.02.023

The Berkeley Satellite Group. (2019). The Berkeley High-resolution NO2 Product. Retrieved February 8, 2019, from http://behr.cchem.berkeley.edu/TheBEHRProduct.aspx

United States Census Bureau. (2010). 2010 Census Urban and Rural Classification and Urban Area Criteria. Retrieved February 7, 2019, from https://www.census.gov/geo/reference/ua/urban-rural-2010.html

United States Census Bureau. (2019). American Community Survey. Retrieved January 25, 2019, from https://www.census.gov/developers/?#

van Donkelaar, A., Martin, R. V., Li, C., & Burnett, R. T. (2019). Regional Estimates of Chemical Composition of Fine Particulate Matter using a Combined Geoscience-Statistical Method with Information from Satellites, Models, and Monitors. *Environmental Science & Technology*. https://doi.org/10.1021/acs.est.8b06392

Williams, B. (2018, December). *CVRP: Data and Analysis Update*. Presented at the Public Workshop: Update to the 3 - Year Plan for LDV Investments, El Monte, California. Retrieved from https://cleanvehiclerebate.org/sites/default/files/attachments/CVRP_Analysis_Update-2018-12-04.pdf

Zhou, Y., Wang, M., Hao, H., Johnson, L., Wang, H., & Hao, H. (2015). Plug-in electric vehicle market penetration and incentives: A global review. *Mitigation and Adaptation Strategies for Global Change*, *20*(5), 777–795. https://doi.org/10.1007/s11027-014-9611-2

**Supporting Information**

Table S1. Rebate amounts available under California's CVRP and EFMP clean vehicle rebate programs. See main text for program descriptions.

| Annual household income | Program ^ | | Conventional Vehicle 20+ MPG | Hybrid 20+ MPG | Hybrid 35+ MPG | Plug-in Hybrid (PHEV) | Battery electric (BEV) | Fuel cell electric (FCEV) |
|---|---|---|---|---|---|---|---|---|
| <= 225% FPL | CVRP | | | | | $1500+$2000* | $2500+$2000* | $5000+$2000* |
| | EFMP | Retire and replace | $4000 | $4000 | $4500 | $4500 | $4500 | |
| | | Plus-up | | $2500 | $2500 | $5000 | $5000 | |
| 226%-300% FPL | CVRP | | | | | $1500+$2000* | $2500+$2000* | $5000+$2000* |
| | EFMP | Retire and replace | | | $3500 | $3500 | $3500 | |
| | | Plus-up | | | $1500 | $4000 | $4000 | |
| 301%-400% FPL | CVRP | | | | | $1500 | $2500 | $5000 |
| | EFMP | Retire and replace | | | | $2500 | $2500 | |
| | | Plus-up | | | | $3000 | $3000 | |
| EFMP combined with CVRP, new vehicles only | | | No | No | No | Yes | Yes | No |

Note: CVRP: Clean Vehicle Rebate Project; EFMP: Enhanced Fleet Moderation Program; FPL: Federal Poverty Level; MPG: miles per gallon; PHEVs: plug-in hybrid vehicle; BEV: battery electric vehicles; FCEV: fuel cell vehicles. * denotes the additional $2000 CVRP rebate issued to lower-income consumers (whose annual household income falls below 300% of the Federal Poverty Level) since March 29, 2016. ^ Plus-up rebates are additional to Retire and Replace rebates available to consumers living in disadvantaged communities as identified by CalEnviroScreen.

Table S2. Average marginal effect of population density on CVRP and EFMP rebate allocation, conditioned on rural and urban status of the census tracts. See description of model specification in Table 2 of the main text.

| | (1) CVRP, statewide, March 2010 and December 2017 | (2) CVRP, state wide, March 2010 and December 2017 | (3) CVRP, South Coast and San Joaquin Valley air districts, July 2015 – December 2017 | (4) EFMP, South Coast and San Joaquin Valley air districts, July 2015 – December 2017 |
|---|---|---|---|---|
| Rural | -0.012<br>(-0.029, 0.006) | -0.009<br>(-0.026, 0.008) | -0.234<br>(-0.593, 0.125) | 0.007<br>(-0.017, 0.031) |
| Urban | -0.023***<br>(-0.029, -0.017) | -0.022***<br>(-0.028, -0.016) | -0.066***<br>(-0.098, -0.034) | -0.008***<br>(-0.011, -0.004) |
| Difference (urban – rural) | -0.011<br>(-0.028, 0.005) | -0.012<br>(-0.028, 0.004) | 0.168<br>(-0.181, 0.518) | -0.015<br>(-0.038, 0.009) |

Note: The average marginal effect is the average change in the number of rebates per thousand households received in a month (for column 1 and 2) or a quarter (for column 3 and 4) per one standard deviation change in population density, conditioned on whether a census tract is rural or urban. Column 1 and 2 are estimated based on model 1 and 2 in Table 4-1. Column 3 and 4 are estimated based on model 3 and 4 in Table 4-3. All other covariates are held at their observed values in this estimation. *, **, and *** indicate significant at 90%, 95%, and 99% confidence intervals. 95% confidence intervals are in parenthesis.

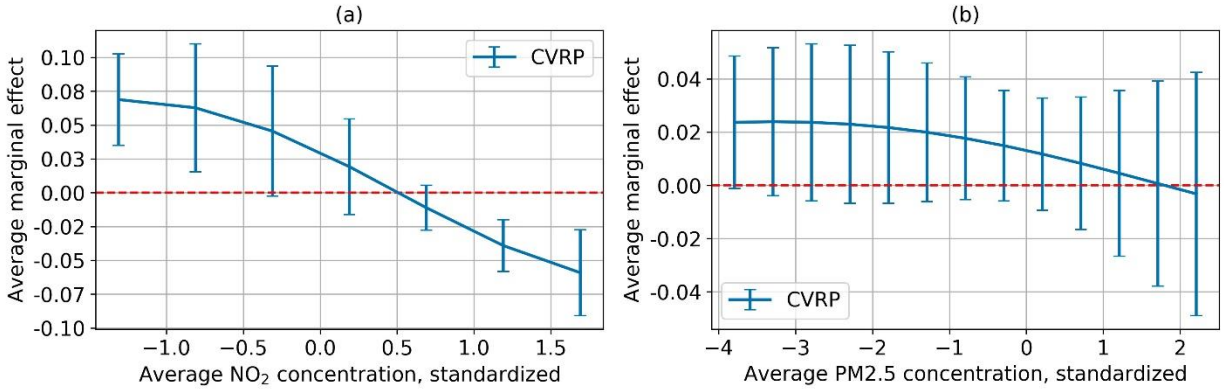



Figure S1. Average marginal effects of census tract-level average ambient NO2 and PM2.5 concentrations on CVRP rebate allocation statewide, March 2010 – December 2017. The average marginal effects are estimated based on model 2 in Table 4-1, and can be interpreted as the average change in the number of rebates per thousand households received in a month per one standard deviation change in the air pollutant concentration, conditional on initial pollution concentrations. All other covariates are held at their observed values in this estimation. Error bars are 95% confidence intervals of the estimates for average marginal effects.

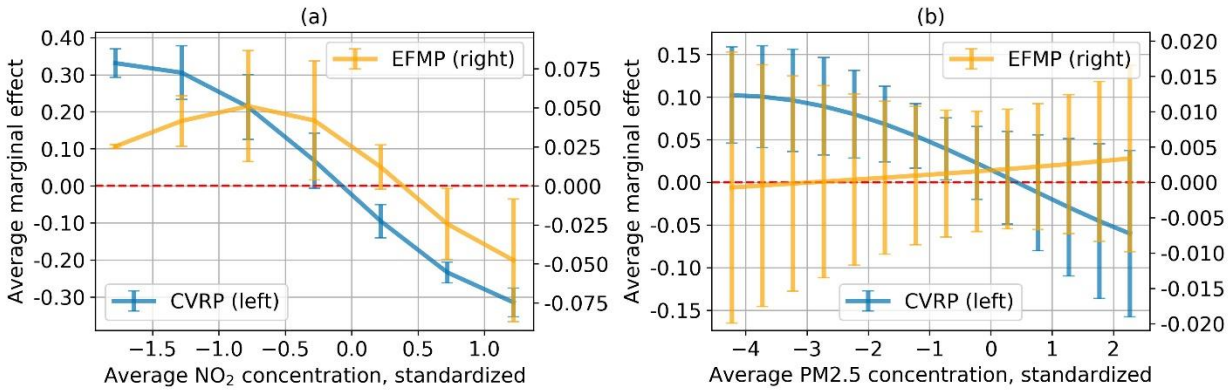



Figure S2. Average marginal effects of NO2 and PM2.5 concentration on CVRP rebate allocation in South Coast and San Joaquin air districts, Q3 2015 – Q4 2017. The average marginal effects are estimated based on Table 4-3. They represent average change in number of rebates per thousand households receive in a quarter for one standard deviation change in the air pollutants, conditioned on different pollution levels. All other covariates are held at their observed values in this estimation. Error bars are one standard deviations of the estimates for average marginal effects.

# Chapter 5. Conclusions

This dissertation focused on characterizing and projecting large-scale urbanization typologies, urban exposure to weather-related environmental hazards (sea level rise and storm surge) under climate change, and the equity implications of adaptation and mitigation programs that address climate change. Through three case studies, we (1) cost-effectively and reproducibly quantify, using time-series remote sensing to data, urban expansion in large regions; (2) project climate-related uncertainties in urban exposure to sea level rise and storm surge under different climate scenarios, and open a discussion with stakeholders on the implications of the uncertainties and long-term projection; (3) quantify the effectiveness of policy interventions in improving equity aspect of 'one size fits all' policy programs for climate change adaptation and mitigation. These studies together demonstrate how we can use models and data-driven approaches to inform decision-making that can better support and integrate sustainability and social equity goals in urban development and for urban population.

In chapter 2, we design an automated approach to classify temporal typologies of urbanization with a dense time-series stack of nighttime light remote sensing images and unsupervised classification. Taking mainland China as an example, we identify five temporal typologies of urbanization, namely stable urban activity, high-level steady growth, acceleration, low-level steady growth, and fluctuation, between 1992 and 2013. Our validation shows that the spatial distribution of the five temporal typologies in a city followed a concentric gradient, a typical urban development pattern particularly in areas with limited environmental constraints (e.g. river and topography). For each temporal typology, the change of its nighttime light intensity strongly correlated with the change of proportion of built-up area (with Pearson correlation between 0.80 and 0.99). These validations suggest that our approach is valid to understand temporal typologies of urbanization, or how urbanization propagated in the past. Comparing this to traditional methods including: socioeconomic data, land-cover mapping, and in-situ land use survey, our approach is low-cost, timely, and easier to scale up in measuring urbanization. Combing our approach with traditional methods can provide a comprehensive understanding about urbanization, in terms of how it progresses in space and time. Our approach addresses the temporal aspect of urbanization, by providing planners and decision-makers a map of where rapid urbanization occurs, and if exists, where urban shrinkage takes place. Overlaying these temporal patterns of urbanization with maps of environmental hazards can inform whether urban exposure to these hazards is likely to increase due to recent urban expansion (or drastic increases in nighttime light intensity) or decrease because of urban shrinkage (or decrease in nighttime light intensity). This forms a basis for more informed urban planning and management.

In chapter 3 we conduct a multi-temporal and multi-scenario projection of an urban area exposure to worst-case scenario flooding under climate change and measured how exposure uncertainties change by climate scenarios and projection time horizons. Focusing on the San Francisco Bay Area's low-lying coastal zones (< 10 m in elevation), we find that between low and high greenhouse gas emission scenarios, the expected uncertainties in flood exposures of population distribution, developed areas, lifeline infrastructures, and emergency responders are 1–2% (i.e. standard deviation of exposure, with exposure measured as the percentage of an item that is flooded) in 2000–2020 and 7–10% in 2080–2100. Such increased uncertainties over time post challenges for stakeholders, particularly for some private-sector stakeholders with short-term investment and planning cycles. The uncertainties and challenges may make stakeholders overlook long-term planning and limit themselves to short-term options such as emergency

management and band aid fixes. To address these issues, decision-makers should promote adaptation and mitigation strategies that are no-regret, reversible, and flexible. Decision-makers may also explicitly require a long-term focus for adaptation and mitigation actions and new development projects.

In chapter 4 we perform an equity analysis of a type of climate change adaptation and mitigation program that incentives the purchase of clean vehicles. Clean vehicle rebate programs are amongst major climate change adaptation and mitigation efforts in California, which aim to improve air quality and reduce greenhouse gas emissions, a large portion of which are produced by the transportation sector. Addressing the equity aspect of rebate programs may facilitate clean vehicle adoption in disadvantaged and low-income communities that are less able to afford clean vehicles. Using California's Clean Vehicle Rebate Project and Enhanced Fleet Modernization Program as examples, we investigate how equity-related policy designs shift rebate allocation to disadvantaged, lower-income communities, and communities of color. We find that under a generalized program design, where the rebate amount is only based on vehicle technology (during the first iteration of Clean Vehicle Rebate Project), rebates are more distributed to advantaged, wealthier communities with intermediate levels of nitrogen dioxide ($NO_2$) concentration, and less distributed to communities with higher percentages of Hispanics and Non-Hispanic Blacks. After introducing an income cap, expanding vehicle eligibility, and adding income- and geography- tiered rebate amounts (during the second iteration of Clean Vehicle Rebate Project and Enhanced Fleet Modernization Program), rebate allocation rates have increased in communities with lower-household income, higher percentages of Hispanics, and slightly higher $NO_2$ pollution. These findings confirm that equity concerns exist in clean vehicle rebate programs with generalized designs that have limited considerations on diverse socioeconomic characteristics of potential participants. Our analysis suggests that equity-related designs can distribute program benefits to populations with more diverse socioeconomic backgrounds.

Overall, modeling and data-driven approaches in this dissertation with different types of geospatial data provides insights into urbanization, its exposure to weather-related environmental hazards under climate change, and equity implications of adaptation and mitigation programs. The three studies in this dissertation are connected, as they all examine how data and models can outline the underlying processes of environmental and societal challenges (i.e. urbanization, coastal flooding under climate change, and equity in adaptation and mitigation programs). The studies also demonstrate how understandings of these processes can inform decision-making, by identifying areas under overly rapid urbanization, discussing uncertainty-related challenges in climate change adaptation, and evaluating equity improvements from interventions of clean vehicle rebate programs. Therefore, we view the studies in this dissertation as part of a broader effort in developing evidence- and process-based decisions and policies.

While the themes are connected, the three studies examine different geographies and cases driven by the best available data. A comprehensive investigation in a single study area, its exposure to one (multiple) environmental hazard(s), and adaptation to and mitigation of the environmental hazard(s) should be conducted, so that the findings can be more informative for this study area, its planners, decision-makers, and other stakeholders. Conducting such a comprehensive, place-based study becomes increasingly necessary and possible, with the growing impact of environmental hazards on urban areas, availability of adaptation and mitigation programs, and public awareness of the need to characterize, project, address, and prevent the adverse consequences of urbanization and climate change, particularly for

disadvantaged and marginalized population groups. For example, the approaches in this dissertation can be applied to the wildland-urban-interface fires in California and evaluate related adaptation and mitigation programs. The other future direction from this dissertation is about improving and refining spatial and temporal resolution of the modeling, in which the increased availability of non-traditional data (e.g. Google Street View, OpenStreetMap) and 'new' analytical methods (e.g. deep learning) increases our ability to characterize the built and socioeconomic environments at a fine scale and in new dimensions. These improvements in data and methods may give us new perspectives to better understand our interactions with the environment, environmental hazards, and the development of fine-tuned policy programs.