Causal inference of urban flood impacts on mobility using difference-in-differences and synthetic control in Lagos, Nigeria

Le-Le Zhang[1], Xin (Bruce)- Wu[1], Kai-Lun Liu[1], Md Abdullah Al, Mehedi[2], Jia-Shu Zhou[3], Virginia Smith[1], Chen-Feng Xiong[1]*

[1] *Department of Civil and Environmental Engineering, Villanova University, Villanova, PA 19085, USA*

[2] *First Street Foundation, Brooklyn, NY 11201, USA*

[3]*Stanford University, Stanford, CA 94305, USA*

Abstract: Urban flooding presents increasing risks to human mobility, service accessibility, and urban resilience, particularly in low- and middle-income countries (LMICs) where behavioral data and institutional infrastructures are often constrained. Despite growing availability of high-resolution geospatial and mobility datasets, robust causal assessment of flood induced behavioral disruptions remains challenging due to the absence of clearly defined treatment periods and suitable control groups. To address these challenges, this study proposes a framework that integrates Difference-in-Differences (DiD) with a hybrid Synthetic Control Method (SCM-DiD) supported by a dual diagnostic procedure based on the Maximal Information Coefficient (MIC) and the Coefficient of Variation (CV). These metrics enable the identification of Points of interest (POI) categories that exhibit minimal correlation with precipitation intensity and high temporal stability, thereby enhancing the credibility of counterfactual estimates in the absence of randomized exposure. By leveraging four data sources (1) mobile location-based services data (LBS), (2) ERA5 reanalysis precipitation, (3) Sentinel-1 SAR flood inundation imagery, and (4) OpenStreetMap infrastructure the framework constructs a daily POI-level panel across 12 urban sectors in Lagos, Nigeria, spanning the June and July 2020 flood season. Temporal segmentation into benchmark, preparation, flood, and recovery periods enables phase specific impact estimation.

Findings reveal sector-specific disruptions, including +39.6% and +50.6% increases in healthcare visitation during flood and recovery phases, and a −19.9% decline in transportation activity with no observed rebound. This study contributes a transferable and scalable methodology for evaluating climate induced disruptions to urban systems and supports more targeted resilience planning in data limited LMIC environments.

Keywords: Urban flooding, Location-Based Services, Points of Interest, Satellite imagery, Low and middle-income countries, Causal inference

*Corresponding author.

E-mail address: chenfeng.xiong@villanova.edu

1. Introduction

The convergence of climate change, accelerated sea-level rise, and rapid urban expansion has amplified the frequency and severity of hydro-meteorological hazards worldwide[1, 2]. Among these, urban flooding stands out as one of the most disruptive and inequitable threats, particularly in rapidly growing cities of the Global South. Beyond its immediate physical impacts—damaged infrastructure, displaced populations, and destroyed property, urban flooding disrupts critical human systems. These disruptions include interruptions to daily mobility, restricted access to essential services such as healthcare and education, and the exacerbation of pre-existing socioeconomic inequalities. Current estimates indicate that approximately 1.81 billion people, or 23% of the global population, are directly exposed to 1-in-100-year flood events, highlighting the urgent need for improved strategies to understand and mitigate flood risks [3].

Despite growing recognition of flooding as a systemic urban hazard, research has predominantly focused on assessing physical damages[4-8]. In contrast, the behavioral and cascading consequences—such as altered mobility patterns, disrupted commerce, and constrained service access—remain poorly understood, especially in low- and middle-income countries (LMICs) where institutional capacity and data infrastructure are limited [9, 10]. Due to its low-lying coastal geography, chronic infrastructure deficits, and exposure to extreme weather, Lagos is increasingly identified as a global flood risk hotspot. It currently ranks among the top 50 cities most exposed to sea-level rise, with projections indicating a shift from 30th to 15th in global population exposure by the 2070s [11, 12]. Significant flood events, such as the multi-week inundation commencing on June 17, 2020, frequently immobilize substantial portions of the city,

impeding mobility and access to critical services. Such occurrences thus provide crucial opportunities and underscore the pressing need for granular behavioral impact studies.

Quantifying the behavioral impacts of such events in LMIC contexts presents persistent methodological challenges. Traditional mobility surveys are infrequent[13], traffic monitoring systems are underdeveloped [14, 15], and real-time data on access to key facilities, such as hospitals, markets are typically unavailable [16]. These limitations hinder both immediate disaster response and the development of robust, data-driven analyses of behavioral adaptation under extreme conditions. Structural deficiencies in sanitation and public health systems further compound the societal impacts of flooding in such settings[17-19].

Recent advances in digital sensing technologies offer promising tools to address these gaps in understanding flood impacts. Anonymized Location-Based Services (LBS) data, derived from mobile devices, now enable high-frequency tracking of population mobility at unprecedented scales [20-22], facilitating the detection of disruptions in routine visitation patterns during flood events [23, 24]. Concurrently, satellite-based systems, such as Synthetic Aperture Radar (SAR) imagery from Sentinel-1, provide near real-time flood mapping capabilities, even in challenging conditions like cloud cover or inaccessible regions [25-28]. However, despite these significant technological advancements and the proliferation of rich spatiotemporal datasets, a critical gap persists in the rigorous integration of such data into robust causal inference frameworks [29]. A predominant limitation of existing studies, particularly those focused on urban flooding, is their reliance on descriptive or correlational approaches, such as visual overlays or simple trend comparisons [30, 31]**.** These methodologies, while useful for initial characterization, frequently fall short in adequately accounting for complex confounding factors like inherent seasonal

variations, pre-existing trends, or concurrent, unrelated policy interventions that can independently influence mobility patterns[32]. Consequently, it remains challenging to definitively attribute observed behavioral changes solely to flood events. Moreover, the inherent characteristics of flood occurrences often lacking sharply defined onset and cessation and affecting populations non-randomly further complicate the identification of clear 'treatment' periods and statistically comparable 'unaffected' control groups, thereby hindering the establishment of strong causal links between flooding and behavioral impact. This methodological gap underscores the pressing need for more sophisticated analytical approaches that can isolate the precise effects of hydro-meteorological hazards on human systems.

Importantly, urban flood impacts should not be viewed solely through the lens of isolated, catastrophic events. In many cities—particularly within rapidly urbanizing contexts—flooding increasingly manifests as a cumulative and recurring stressor, unfolding over entire rainy seasons and interacting with existing infrastructural and institutional vulnerabilities. These events are rarely discrete; rather, they emerge through overlapping episodes that exhibit diffuse spatial patterns and ambiguous temporal boundaries. Unlike sudden-onset hazards such as hurricanes or earthquakes, which typically have well-defined initiation points and geographically bounded impacts, urban floods often lack precise start and end times. This is especially true in flood-prone areas where inundation events frequently overlap, reoccur within short intervals, or persist intermittently across sectors, leading to duplicate or blurred exposure periods. Such complexity poses major analytical challenges. Standard impact evaluation approaches often assume sharply defined treatment phases and the existence of unaffected control groups—assumptions that are frequently violated in contexts of widespread, temporally fluid flood exposure.

This study addresses these challenges by developing a data-driven, quasi-experimental framework for causal inference in urban flood impact analysis. Leveraging the June–July 2020 flood events in Lagos, Nigeria as a natural experiment, the proposed methodology integrates high-resolution behavioral, geospatial, and environmental data within a unified spatiotemporal model. The approach combines a Difference-in-Differences (DiD) framework[33] with a hybrid Synthetic Control Method (SCM-DiD)[34], designed to capture both sector-specific heterogeneity and non-parallel pre-treatment dynamics. Key data sources include anonymized LBS data, satellite-derived flood inundation maps from Sentinel-1 SAR imagery, and ERA5 reanalysis precipitation records.

To strengthen causal identification under conditions of non-random exposure and limited observational controls, the framework introduces two diagnostic tools. The Maximal Information Coefficient (MIC) [35] is employed to assess the statistical independence of POI visitation patterns from precipitation intensity, thereby guiding the exclusion of precipitation-sensitive categories from the control pool. Concurrently, the Coefficient of Variation (CV) is used to evaluate the temporal stability of visitation trends across flood phases, ensuring that selected control groups exhibit consistent pre-treatment behavior. Together, these diagnostics support the empirical selection of statistically valid comparison groups, mitigating common sources of bias in observational disaster studies. The overall research workflow—from data integration to impact estimation is summarized in Figure 1.

More broadly, this study responds to persistent empirical and methodological constraints in evaluating the behavioral consequences of flooding in data-scarce settings. First, to address the lack of integrated behavioral and environmental datasets LMICs, it constructs a high-resolution panel dataset that captures daily POI visitation dynamics across diverse urban sectors. Second, to

overcome the absence of unaffected control units in highly exposed cities, it proposes a reproducible control selection procedure grounded in MIC and CV diagnostics. Third, to accommodate the complex temporal structure of urban flood impacts, it applies a sector-level SCM-DiD framework capable of capturing both anticipatory and lagged behavioral responses. Collectively, these innovations provide a scalable and transferable framework for causal inference in climate impact research, particularly in urban contexts where traditional experimental designs and administrative baselines are not feasible.

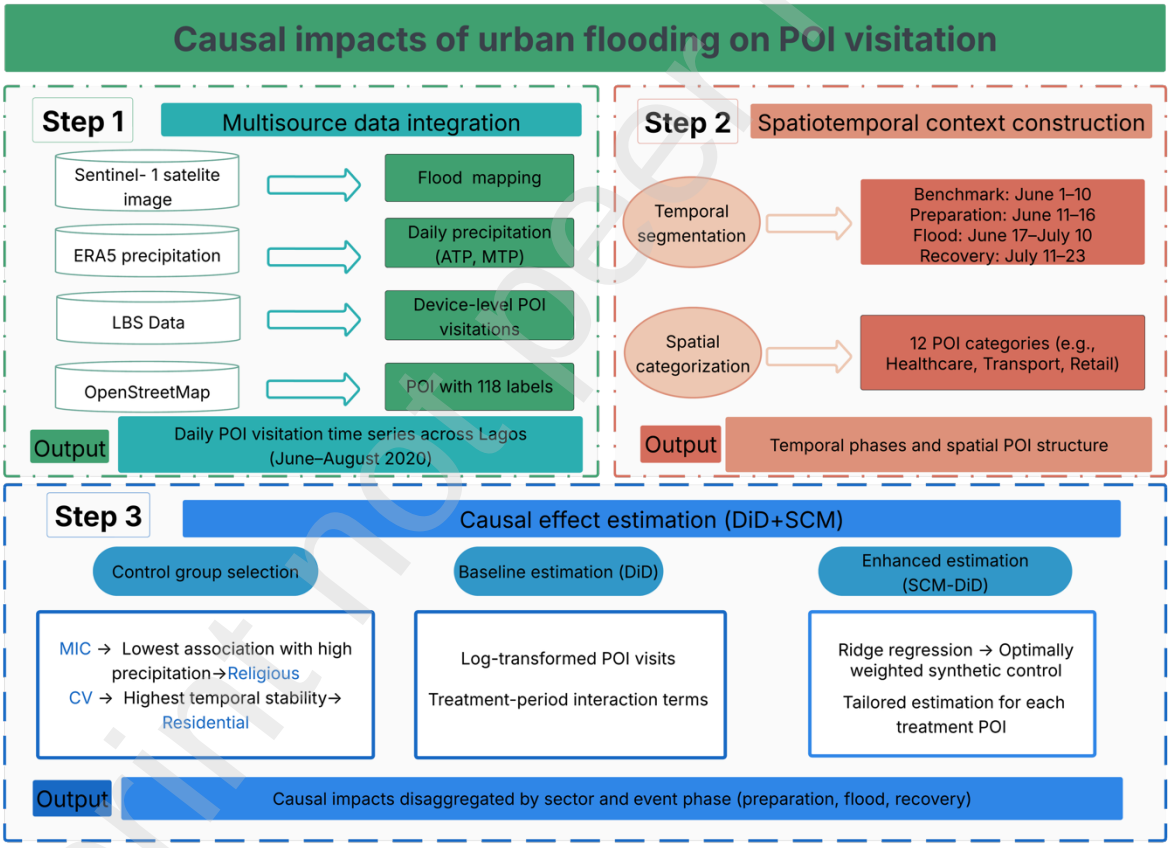

Figure 1. Framework for assessing urban flooding impacts on human mobility

2. Datasets and processing

This study leverages a suite of datasets spanning behavioral, geospatial, and environmental domains to facilitate robust causal inference on the effects of urban flooding on human mobility in Lagos. Situated within a DiD framework, these data sources serve three core functions:

(1) Flood exposure characterization: Precisely delineating the temporal onset of flood events using satellite-derived and remote-sensed environmental data.
(2) Spatial unit definition: Structuring the analysis around geolocated POI, representing distinct categories of urban activity.
(3) Outcome measurement: Capturing population mobility dynamics via aggregated human activity signals at POI, reflecting behavioral responses to flood exposure.

These datasets enable the assessment of spatially and temporally disaggregated treatment effects across multiple urban sectors. The subsections below describe the provenance, characteristics, and preprocessing protocols applied to each data component within the analytical pipeline.

2.1 Characterizing flood exposure

Accurate delineation of flood timing is essential for defining treatment status within the causal inference framework. To this end, the study integrates satellite-based radar imagery with ground-validated precipitation reanalysis data, enabling the temporal characterization of flood events in Lagos.

2.1.1 Satellite-derived flood inundation

To identify the spatial extent and temporal progression of flood inundation, this study employs synthetic aperture radar imagery from the Sentinel-1 satellite constellation, accessed via the Google Earth Engine platform. The European Space Agency's Sentinel-1A and Sentinel-1B satellites provide C-band SAR data and all-weather imaging capabilities, particularly suited for flood monitoring due to their ability to penetrate cloud cover[36].

Level 1 Ground Range Detected products were processed to detect changes in surface reflectivity associated with flood events. A standardized change detection approach was applied, comparing radar backscatter values between a pre-flood reference period and flood-period acquisitions. Prior to change detection, imagery was preprocessed to reduce speckle noise (using a Lee filter) and converted to decibel (dB) units. Pixels exhibiting a substantial increase in backscatter intensity indicative of open water surfaces were classified as inundated based on empirically derived thresholds [37, 38].

Flood extent maps were generated for each acquisition date between June and August 2020, as detailed in Table 1. Two major flood events were identified, peaking around June 23 and July 5, 2020. Figure 2a situates Lagos State geographically within Nigeria, highlighting its vulnerability

to seasonal flooding. Figure 2b presents a non-flood baseline image, while Figures 2c and 2d illustrate the progressive spatial expansion of floodwaters across low-lying urban areas during the two peak events. Notably, the July 5 inundation demonstrates heightened severity, particularly affecting densely populated coastal zones. These satellite-derived products provide the foundation for defining flood exposure at temporal resolution.

Table 1 Sentinel-1 data acquisition dates for Lagos State (June–August 2020; bold dates indicate flooded areas detected in inundation maps)

Months	Dates
June	6-6-2020, 6-11-2020, **6-23-2020,** 6-30-2020
July	**7-5-2020,** 7-12-2020, 7-17-2020, 7-24-2020
August	8-5-2020, 8-10-2020, 8-17-2020, 8-22-2020, 8-30-2020

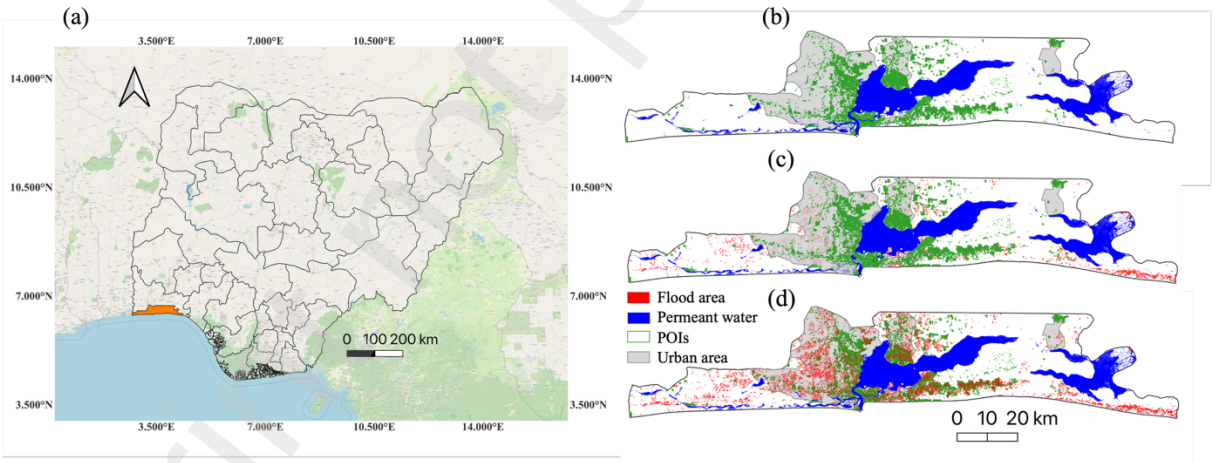

Figure 2: Spatial and temporal analysis of flood events in Lagos State, Nigeria, 2020: (a) Map of Nigeria with Lagos State highlighted in orange, (b) Sentinel-1 image during non-flood period, (c) Flood extent on June 23, and (d) Flood extent on July 5.

While SAR imagery provides high spatial resolution with 10 meters, its six-day revisit cycle presents temporal limitations, potentially missing brief flood events or mischaracterizing urban flooding under dense canopies. To overcome these constraints and enhance the temporal accuracy of flood onset and duration, integrated high-frequency precipitation data.

2.1.2 High-frequency precipitation dynamics

To complement the satellite-derived flood inundation maps and enhance the temporal resolution of flood exposure delineation, this study incorporates high-frequency meteorological data from the ERA5 reanalysis product, developed by the European Centre for Medium-Range Weather Forecasts. ERA5 provides globally consistent, gridded atmospheric variables by assimilating a wide range of observational inputs into a state-of-the-art numerical weather prediction model, thereby offering reliable retrospective climate estimates[39].

The analysis focuses on daily total precipitation (TP) over Lagos State during the 2020 rainy season. Two precipitation metrics were derived:

(1) Average Total Precipitation (ATP), defined as the spatial mean of daily TP across all ERA5 grid cells covering Lagos State.
(2) Maximum Total Precipitation (MTP), representing the highest daily TP value recorded at any individual grid cell within the state.

To address potential biases in reanalysis-derived precipitation magnitudes—while preserving critical temporal dynamics—both ATP and MTP were bias-corrected using a linear

scaling method calibrated against NASA's Integrated Multi-satellite Retrievals for GPM product [40, 41]. Detailed procedures are provided in Appendix 1.

As shown in Figures 3a and 3b, the ERA5-based precipitation patterns confirm June and July 2020 as the core of the wet season, aligning with both local meteorological reports and flood events detected via Sentinel-1 imagery. Figure 3c presents daily precipitation trends, revealing prominent rainfall peaks on June 18 and June 29 both temporally proximate to the major flood events identified in Section 2.1.1. The integration of satellite-based flood detection with high-resolution rainfall dynamics facilitates a robust and temporally precise characterization of flood exposure.

This multi-source environmental data synthesis supports the demarcation of analytically distinct event phases, including benchmark, preparation, flood, and recovery as formally defined in Section 3.1. Figure 3d summarizes average precipitation levels associated with each phase, providing empirical justification for the phase boundaries used in subsequent causal analysis.

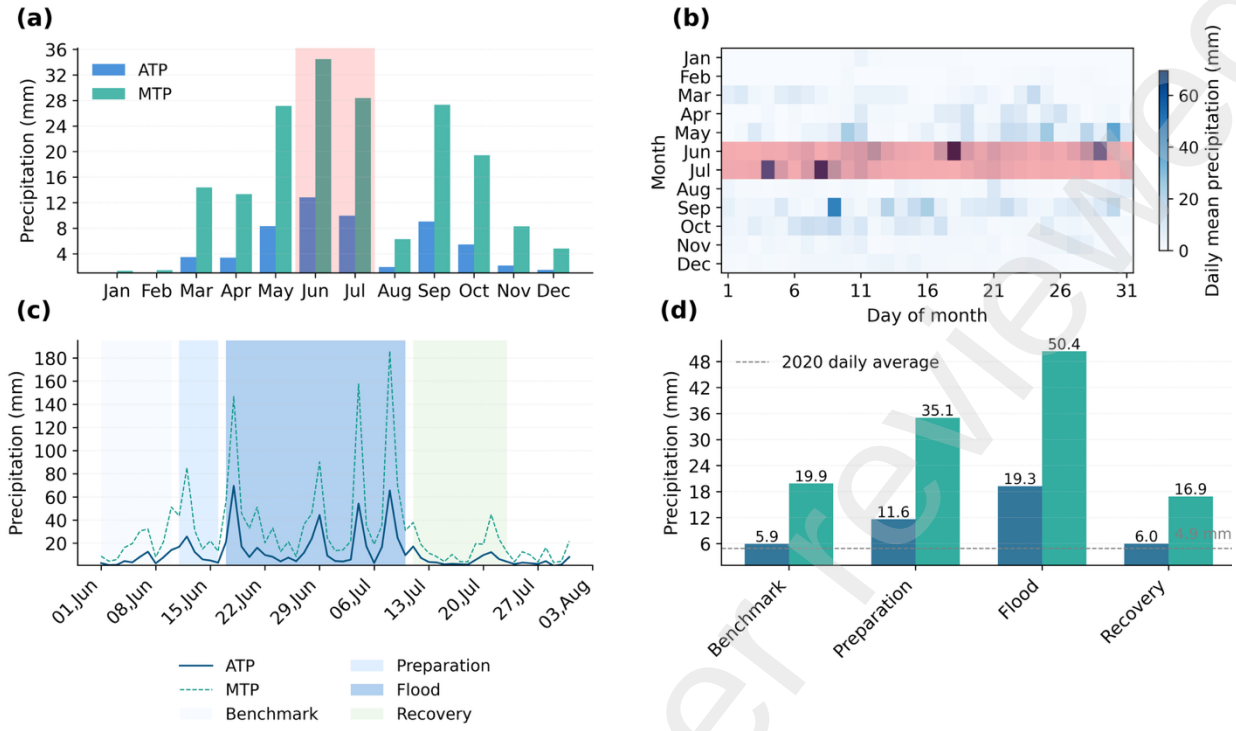

Figure 3 ERA5 Precipitation Analysis for Lagos State (June-August 2020). (a) Monthly mean ATP and MTP in 2020. (b) Heatmap of daily mean precipitation in 2020. (c) Daily ATP and MTP, June 1 – July 31, 2020, with defined flood-related periods. (d) Average precipitation levels during defined periods.

2.2 POI from OpenStreetMap

POI serve as the fundamental spatial units for evaluating flood-induced disruptions to urban mobility. A total of 169,071 POI polygons were extracted from OpenStreetMap (OSM) using the `osm2gmns` Python package[42]. These features encompass a diverse array of urban facilities, including healthcare centers, educational institutions, commercial zones, religious sites, and public services.

Given substantial heterogeneity across the 118 original OSM labels, a systematic classification procedure was implemented to consolidate POI into 12 categories (e.g., Healthcare, Retail, Education). Ambiguities and missing classifications were cross validated using metadata from Google Maps to enhance semantic consistency. The complete mapping schema, detailing the correspondence between raw OSM labels and final POI categories, is provided in Appendix 2.

This refined POI typology facilitates sector-specific analysis of mobility responses to flooding, ensuring spatial consistency across the exposure and outcome measures.

2.3 Human mobility via LBS data

To capture spatiotemporal patterns of human mobility constituting the primary outcome variable in the DiD framework, this study employs anonymized, passively collected LBS data. These data are derived from mobile devices that emit geolocation signals when applications trigger positioning updates via GPS, Wi-Fi, Bluetooth, or cellular networks. Each record includes a hashed device identifier, geographic coordinates, a timestamp, and an associated accuracy estimate[20].

The LBS dataset used in this analysis was obtained from a state-level mobility roster for Lagos, spanning June to August 2020, and compiled by the Villanova Mobility Lab [43]. It contains a total of 4,626,086 individual trip records during the research period. Preprocessing routines, applied prior to analysis, included the reconstruction of trip trajectories and inference of origin-destination pairs, departure and arrival times, and movement durations. Data quality was rigorously assessed along key dimensions, including sample size adequacy, spatial and temporal representativeness, geolocation precision, and internal consistency. This validation ensures the suitability of dataset for evaluating behavioral responses to environmental disruptions at high spatial and temporal resolution.

To quantify mobility activity at the sectoral level, LBS-derived sighting points were spatially mapped onto the POI polygons. This process enabled the construction of a temporally indexed panel of POI-level activity across the twelve functional categories. The resulting time series forms the empirical basis for assessing flood-induced disruptions in access to essential services and urban functions.

Figure 4 illustrates the evolution of normalized POI activity over the study period, with shaded bands indicating distinct flood-related phases as defined in Section 3.1. These trajectories provide preliminary evidence of sector-specific temporal dynamics in response to the 2020 flood events, which are explored more formally in the subsequent causal analysis.

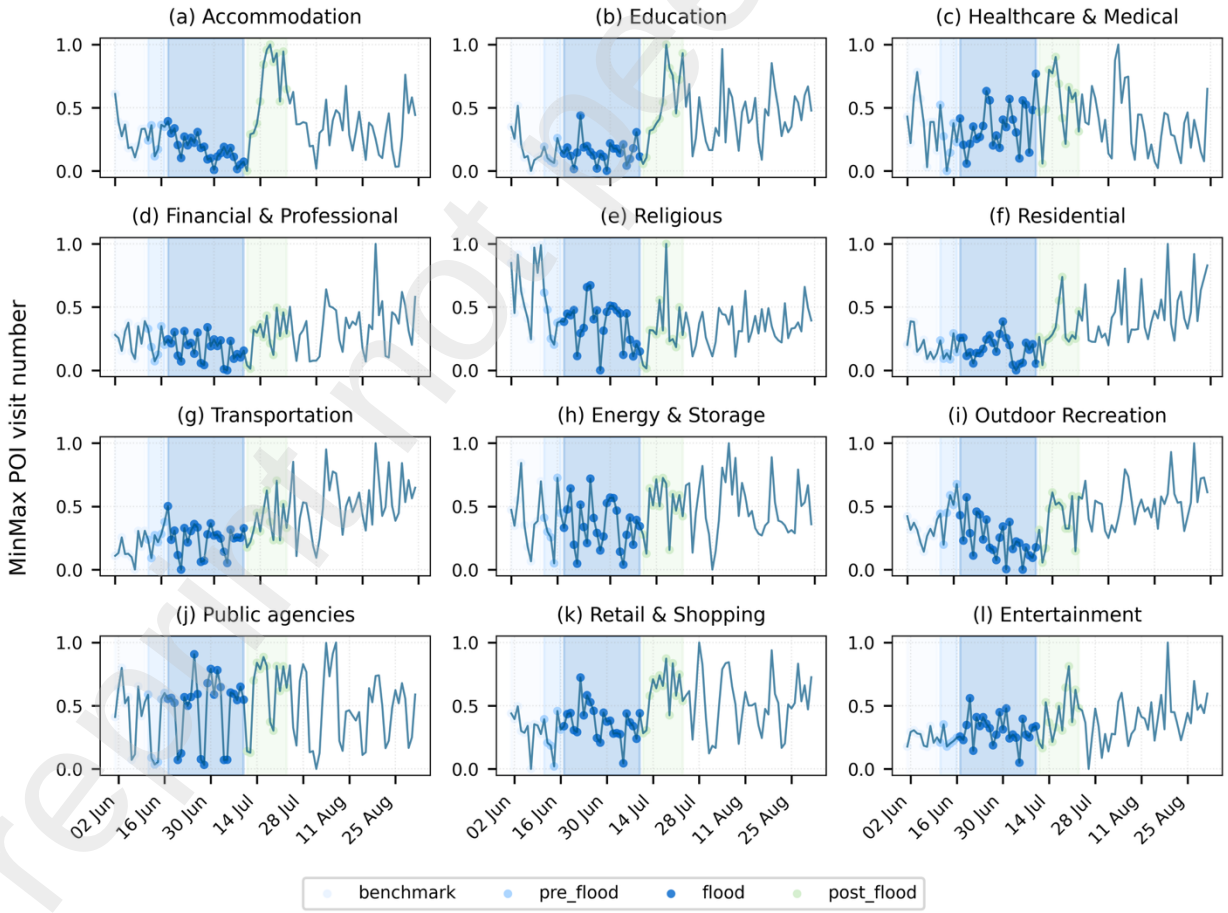

Figure 4 Temporal evolution of Min-Max normalized sighting counts mapped within different categories of POI polygons in Lagos State, Nigeria (June – August 2020). The shaded regions denote key flood-related periods: preparation is light blue, flood is blue, and recovery is yellow.

3. Method

This section outlines the strategy employed to estimate the causal effects of recurrent urban flooding on human mobility, as proxied by activity patterns at POI across Lagos. Rather than focusing on an isolated extreme event, this study adopts a broader behavioral lens by treating seasonally recurring flood episodes as an experiment. This approach facilitates the investigation of population-level mobility adaptations to persistent and anticipated flood conditions characteristic in Lagos state.

By leveraging high-resolution, temporally resolved datasets within a DiD framework, the study captures the heterogeneous and temporally dynamic behavioral responses to flooding across multiple urban sectors. This methodological orientation emphasizes cumulative exposure and adaptation, rather than single-event shock responses in flood-prone urban contexts.

3.1 Temporal delineation of representative flood episodes

To operationalize this causal framework, the analysis centers on a representative and well-documented flood episodes in Lagos during the summer of 2020. The temporal segmentation of these events is grounded in a multi-source empirical basis, including satellite-derived flood inundation maps, high-frequency ERA5 precipitation data, and triangulating evidence from social media, local news coverage, and Twitter data. This integrative approach enables a fine-grained

characterization of flood lifecycle dynamics and minimizes arbitrary or exogenous timing assumptions.

Four analytically distinct periods are defined to reflect the onset, peak, and resolution of the flood episodes, providing a structured temporal framework for causal estimation:

Benchmark Period (10 days: June 1 – June 10, 2020): Serves as the pre-intervention baseline, capturing mobility conditions under non-flooded circumstances. This period immediately precedes the seasonal rainfall onset, allowing for a stable estimation of baseline trends while minimizing confounding from seasonal variability. Satellite imagery and precipitation data confirm the absence of significant flood activity during this interval.

Preparation Period (6 days: June 11 – June 16, 2020): Represents the anticipatory phase during which no widespread inundation occurred. The start date aligns with typical operational lead times provided by medium-range weather forecasts (5–7 days), offering empirical justification for its designation as a behavioral preparation window. This period reflects adaptive or precautionary changes in mobility in response to elevated flood risk signals [44] .

Flood Period (24 days: June 17 – July 10, 2020): Denotes the primary treatment window, encompassing multiple high-impact flood events rather than a single continuous shock. This interval begins with the first widely reported impacts on June 17 and includes several peak rainfall events (e.g., June 18, June 29, July 4, July 8), during which ATP frequently exceeded 40 mm and MTP surpassed 100 mm. Notably, two major flood dates, June 23 and July 5 fall within this period, both of which were associated with widespread inundation. Consistent satellite and media reports

confirm sustained inundation across the city, reflecting cumulative environmental stress and mobility disruption.

Recovery Period (13 days: July 11 – July 23, 2020): Captures the initial post-flood normalization trajectory. The start date coincides with satellite-detected recession of floodwaters (observed on July 12), marking the onset of physical recovery. The two-week duration permits observation of mobility stabilization, acknowledging the possibility of lagged or incomplete behavioral recovery across sectors.

This temporal segmentation serves as the foundation for constructing treatment and control contrasts within the DiD framework, allowing for the estimation of both aggregate and POI category flood impacts on urban mobility patterns.

3.2 Identification of treatment and control groups

This section outlines the procedures for selecting control and treatment POI categories to enable credible causal inference. Control group identification is grounded in data-driven screening to ensure minimal confounding and temporal stability, while treatment groups comprise the remaining POI categories subjected to comparative analysis.

3.2.1 Control group selection

To construct counterfactuals for flood impact estimation, POI categories suitable as control units for standard DiD or as donor units for SCM-DiD were empirically selected based on two principal criteria: (1) minimal statistical dependence on precipitation, and (2) high temporal stability across defined analytical periods. These properties were quantified using the MIC and the CV, respectively.

(a) Minimal precipitation association via MIC

The MIC was employed to identify POI categories whose daily visitation patterns exhibit weak dependence on weather variability. MIC is a non-parametric statistic that captures a broad range of associations between two variables, in this case, MTP, P_t, and POI visitation $V_{i,t}$. Unlike linear correlation metrics, MIC does not impose structural assumptions, enabling robust identification of nonlinear relationships[35, 45].

For each POI category i. MIC was calculated between P_t and $V_{i,t}$ over the preparation and flood periods, which capture relevant variability in rainfall. The dimensions of the grids, G_P and G_{V_i} are bound by:

$$|G_P|,|G_{V_i}| \leq cT^{\alpha} \quad \text{(Eq.1)}$$

Where T represents the total number of daily observations, use precipitation and POI visit data from both the preparation period and the flooding period (6 days + 24 days = 30 days) encompassing a range of rainfall conditions relevant to the flood event. he grid resolution parameter B was set to 15, and the scaling exponent α to 0.6, consistent with best practices and the *minepy* implementation. The MIC statistic is defined as:

$$\mathrm{MIC}(P,V_i) = \max_{G_p,G_{V_i}} \frac{I^*(P,V_i)}{\log_2 \min(|G_P|,|G_{V_i}|)} \quad \text{(Eq.2)}$$

Where $I^*(P,V_i)$ is the maximum mutual information over the optimized grids. The log_2 denominator normalizes MIC values to the range [0,1], POI categories i exhibiting lower MIC (P,V_i) values were considered preferable, indicating weaker dependence of their visitation patterns on MTP.

Lower MIC values indicate weaker dependence between visitation and precipitation, suggesting greater resilience to direct weather effects.

(b) Temporal stability via $CV_{between}$

To assess the robustness of visitation patterns across the flood timeline, temporal stability is evaluated using the between-period coefficient of variation $CV_{between}$,This metric quantifies the extent to which average mobility levels fluctuate across distinct analytical phases, enabling standardized comparisons across POI categories with differing baseline visit volumes.

The $CV_{between}$ is particularly well-suited for identifying systematic variations attributable to flood-related disruptions, as it captures dispersion in period-level means rather than random intra-phase fluctuations. Daily visit counts are first log-transformed to stabilize variance and accommodate zero observations, using the formulation:

$$L_{i,t} = \ln(V_{i,t} + 1) \quad \text{(Eq.3)}$$

Where $V_{i,t}$ denotes the raw visit count to POI category i on day t. Let $\bar{L}_{i,p}$ denote represent the mean of $L_{i,t}$ for POI category i during analytical period. Where

$$p \in \{Benchmark, preparation, flooding, recovery\}$$

The between-period coefficient of variation is then defined as:

$$CV_{between}(i) = \frac{\mathrm{std}(\bar{L}_{i,benchmark}, \bar{L}_{i,preparation}, \bar{L}_{i,flooding}, \bar{L}_{i,recovery})}{\mathrm{mean}(\bar{L}_{i,benchmark} + \bar{L}_{i,preparation} + \bar{L}_{i,flooding} + \bar{L}_{i,recovery})} \quad \text{(Eq.4)}$$

Lower values of $CV_{between}(i)$ indicate higher temporal stability in log-transformed visitation levels across the four defined phases. Such stability is indicative of resilience to flood-

related and contemporaneous shocks, and thus marks POI categories as more suitable for serving as comparator groups in DiD and SCM-DiD estimations.

(c) Selected control POI categories:

Upon evaluating all 12 POI categories against both metrics, two categories emerged as optimal control units exhibiting low MIC and $CV_{between}$ scores. These were selected to form the pooled control group for the DiD analysis (Section 3.3) and to constitute the donor pool for SCM-DiD (Section 3.4).

3.2.2 Treatment group selection

The remaining 10 POI categories, not included in the control group, were designated as treatment groups. Each is analyzed independently to enable sector-specific estimation of flood-related impacts. This approach supports disaggregated causal analysis and highlights heterogeneity in mobility responses across different urban service domains.

3.3 DiD estimation

As a primary analytical strategy, this study employs a standard DiD regression model to estimate the causal impact of the flood event on visitation patterns across various POI categories[46]. This quasi-experimental approach compares the temporal evolution of visitation for a treatment POI category (i.e., a category whose visitation is potentially affected by the flood) against that of an empirically selected pooled control group. The control group, comprising two specific POI categories identified as being minimally associated with precipitation and exhibiting

high temporal stability, serves as a proxy for the counterfactual trend that the treatment POI category would have followed in the absence of the flood.

The DiD model is estimated using daily panel data, where each observation represents the visit count for a specific POI category. For each POI category j designated as a treatment, the following specification is estimated:

$$\ln(V_{k,t}+1) = \gamma_0 + \gamma_1 \cdot Treat_k + \sum_{p \epsilon P} \lambda_p \cdot Period_{p,t} + \sum_{p \epsilon P} \beta_p \cdot Period_{p,t} \times Treat_k) + \varepsilon_{k,t} \quad \text{(Eq.5)}$$

Where:

- $V_{k,t}$ is the visit count for POI category k on day t.
- $Treat_k$ is a binary indicator for treatment status.
- $Period_{p,t}$ are dummies for each temporal phase $p \epsilon \{preparation, flood, recovery\}$ with the benchmark period omitted as the reference.
- γ_0 is the intercept, representing the average log-transformed visitation for the control group during the benchmark period.
- γ_1 captures baseline differences between treatment and control groups.
- λ_p represent period effects for the control group.
- β_p are the DiD estimators of interest, capturing differential changes for the treatment group relative to the control group.
- $\varepsilon_{k,t}$ is the idiosyncratic error term.

Heteroscedasticity-consistent (HC3) standard errors are employed. The model is estimated separately for each of the 10 treatment POI categories. Statistically significant β_p values are interpreted as evidence of flood-induced behavioral shifts in visitation.

3.4 SCM-DiD for enhanced causal inference

To strengthen the credibility of causal estimates, this study implements a hybrid SCM-DiD framework that integrates the strengths of both the SCM and the DiD approach. While DiD relies on the assumption of a uniform counterfactual trajectory for all treated units, SCM-DiD relaxes this assumption by constructing individualized synthetic controls that replicate the pre-treatment behavior of each treated POI category. This is particularly valuable in contexts where treatment groups are heterogeneous and may violate the parallel trends assumption inherent in standard DiD models.

By generating treatment-specific synthetic trajectories using a weighted combination of donor units[47]. The SCM-DiD framework ensures that each counterfactual is empirically grounded and tailored to the unique pre-flood dynamics of the corresponding treatment unit. This integration significantly enhances the robustness and interpretability of causal inference, especially in urban systems characterized by complex and uneven exposure to environmental hazards such as flooding.

For each treatment POI category j , a set of weights, $W_j = (w_{j1,}w_{j2})^T$, is computed. These weights are optimized to minimize the squared deviation between the standardized, log-transformed visitation trajectory of the treatment POI and a convex combination of donor POI during the benchmark period. Ridge regression with a regularization parameter $\lambda = 1.0$, subject to non-negativity constraints, is used to obtain initial weights. These are subsequently normalized such that. $w_{ji} \geq 0$ and $\sum_i w_{ij} = 1$, The synthetic trajectory for treatment unit j is constructed as:

$$\tau_{jt} = \ln(V_{j,t} + 1) - \ln\left(V_{j,t}^{syn} + 1\right) \quad \text{(Eq.6)}$$

Where τ_{jt} represents the estimated treatment effect at time t, and $V_{j,t}^{syn}$ is the synthetic counterfactual visitation. To estimate average treatment effects across periods, τ_{jt} is regressed on temporal indicators:

$$\tau_{jt} = \delta_0 + \delta_1 Pre_t + \delta_2 Flood_t + \delta_3 Roc_t + v_{jt} \quad \text{(Eq.7)}$$

where δ_0 captures pre-treatment fit (ideally near zero), and δ_1,δ_2, δ_3 represent the average treatment effects during the preparation, flood, and recovery periods, respectively. HC3 robust standard errors are applied for statistical inference.

In sum, the SCM-DiD approach enhances causal identification by tailoring control trajectories to each treated unit, thereby addressing limitations of pooled counterfactual assumptions and improving the reliability of estimated flood impacts across heterogeneous POI categories.

4. Results

This section presents the empirical findings on the causal impacts of seasonal urban flooding on human mobility in Lagos. It begins by detailing the selection of robust POI categories that served as control or donor units for the DiD and SCM-DiD frameworks and then reports the estimated causal effects derived from both approaches, focusing on variation in POI visitation across flood-affected and control categories during distinct phases of the 2020 rainy season.

4.1 Identifying optimal control POI for causal inference

A prerequisite for credible causal analysis is the construction of reliable counterfactuals. To this end, a data-driven screening procedure was applied to identify POI categories that exhibit both low sensitivity to precipitation and high temporal stability in visitation patterns. These POI types were designated as either control units (for DiD estimation) or donor units (for SCM-DiD synthesis). Two complementary metrics were employed:

(1) Sensitivity to precipitation via MIC

To evaluate the extent to which visitation patterns were directly influenced by weather conditions, the MIC was computed between daily POI visitation counts and MTP from June 10 to July 13, 2020. This 30-day window captures both dry and flood-affected conditions.

As shown in Figure 5a, POI categories demonstrated substantial heterogeneity in precipitation sensitivity. Transportation exhibited the strongest dependence with MIC is 0.47, followed by Public Agencies, Retail & Shopping, and Healthcare & Medical with MIC larger than 0.35. In contrast, Religious and Accommodation categories had the lowest MIC scores with 0.19 and 0.27 respectively, suggesting reduced vulnerability to direct precipitation impacts.

(2) Temporal stability via $CV_{between}$

To assess the consistency of POI visitation across event phases, the coefficient of variation across the four analytical periods, benchmark, preparation, flood, and recovery was calculated for each category using log-transformed visitation data. This metric isolates structural variability attributable to flood timing from baseline fluctuations. Figure 5b reveals that Residential and Energy & Storage categories exhibited the lowest temporal variability with CV is close to 0.01, reflecting stable visitation throughout the study period. Financial & Professional Services and

Healthcare & Medical categories also demonstrated moderate stability with CV nearly 0.02. While Religious POI had low MIC values, their higher CV with 0.05 reflected routine weekly fluctuations, making them less ideal as donor units in SCM-DiD despite their low weather sensitivity.

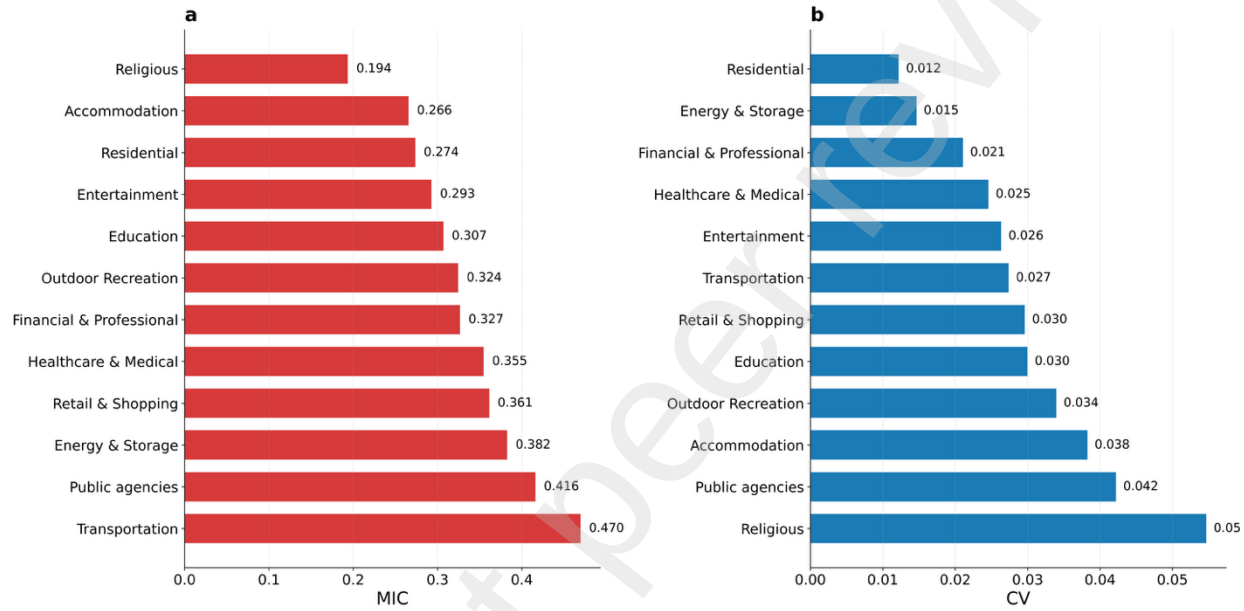

Figure 5 Comparative analysis of MIC and CV across POI categories

Based on this dual screening, low MIC and$CV_{between}$, two POI categories, Religious and Residential were selected as optimal comparators for causal inference. These categories formed the pooled control group for DiD estimation and the donor pool for SCM-DiD. Their low susceptibility to exogenous weather shocks and internal temporal volatility makes them especially well-suited to serve as stable baselines in estimating flood-related disruptions.

4.2 Estimation of flood impacts via DiD and SCM-DiD approaches

This section presents the estimated causal impacts of the 2020 Lagos flood episodes on visitation patterns to POI across a diverse range of urban sectors. Leveraging both DiD and SCM-DiD frameworks, quantify behavioral responses to flooding through comparisons between treatment and control groups across defined temporal phases, preparation, flood, and recovery.

For interpretability, log-point coefficients from the regression models are exponentiated and converted into percentage changes using the transformation $(e^{\beta} - 1) \times 100$. All models incorporate period-specific fixed effects, and heteroscedasticity-robust (HC3) standard errors are used to ensure valid inference. Three control strategies are employed:(1) Religious POI as the control group. (2) Residential POI as the control group. (3) An optimized synthetic control combining both categories. The complete DiD and SCM-DiD results are presented in Table 2, which summarizes log-point coefficients across flood phases and model specifications.

4.2.1 Sectoral impacts: disruption, resilience, and recovery

The results reveal a heterogeneous landscape of flood impacts, with sectoral variation driven by differences in infrastructure exposure, behavioral responses, and institutional recovery capacity.

(a) Healthcare: systemic stress and prolonged demand

The Healthcare & Medical sector consistently exhibited robust increases in visitation across all phases and model specifications. Under the SCM-DiD model, visitation increased by +39.6% during the flood period ($p < 0.01$) and remained elevated by +50.6% during the recovery phase ($p < 0.01$). This pattern reflects multiple compounding stressors: waterborne disease outbreaks, flood-related injuries, and heightened mental health needs[48-50]. The persistence of elevated demand into the recovery period signals a prolonged burden on health infrastructure. underscoring the urgent need for enhanced public health preparedness and resource allocation for sustained post-flood medical services in LMICs.

These sector-specific effects are visualized in Figure 6, which presents estimated log-point changes in POI visitation across each control strategy. The healthcare sector stands out for its statistically significant and sustained increases in visitation during both the flood and recovery periods, highlighting its dual role as an emergency service provider and a pressure point in post-disaster recovery systems. To enhance interpretability, Figure 7 translates the model-estimated log-point coefficients into percentage changes, providing a more intuitive comparison of visitation dynamics across POI categories and flood phases. The figure highlights sectors with contrasting trajectories—those exhibiting resilience (e.g., Retail, Accommodation) versus those with persistent disruption (e.g., Transportation, Outdoor Recreation).

Table 2: Estimated causal effects of flooding on POI visitation using DiD and SCM-DiD

Sector	Log-Point Coefficients			N	R2	Weights (Rel./Res.)
	Preparation	Flood	Recovery			
Panel A: Religious Control						
Healthcare & Medical	-0.096(0.224)	0.189*(0.132)	0.552***(0.205)	244	0.836	
Public agencies	-0.452(0.420)	0.105(0.192)	0.486**(0.243)	244	0.770	
Retail & Shopping	-0.307(0.213)	-0.001(0.125)	0.446**(0.190)	244	0.919	
Entertainment	-0.297*(0.168)	-0.024(0.131)	0.320(0.207)	244	0.812	
Energy & Storage	-0.142(0.232)	-0.045(0.134)	0.293(0.199)	244	0.826	
Accommodation	-0.065(0.173)	-0.145(0.122)	0.600***(0.214)	244	0.830	
Education	-0.316**(0.161)	-0.208*(0.118)	0.381*(0.202)	244	0.926	
Outdoor Recreation	0.064(0.197)	-0.329**(0.139)	0.126(0.209)	244	0.648	
Transportation	-0.397**(0.182)	-0.319**(0.133)	0.075(0.195)	244	0.856	
Financial & Professional	-0.252(0.221)	-0.292**(0.145)	0.118(0.210)	244	0.720	
Panel B: Residential Control						
Healthcare & Medical	0.170(0.180)	0.351***(0.084)	0.392***(0.101)	244	0.690	
Public agencies	-0.186(0.398)	0.267(0.163)	0.326*(0.182)	244	0.394	
Retail & Shopping	-0.041(0.166)	0.162**(0.074)	0.286***(0.100)	244	0.166	
Entertainment	-0.031(0.103)	0.138*(0.084)	0.160(0.130)	244	0.768	
Energy & Storage	0.124(0.189)	0.117(0.088)	0.133(0.117)	244	0.743	
Accommodation	0.201*(0.111)	0.017(0.068)	0.439***(0.141)	244	0.816	
Education	-0.050(0.090)	-0.045(0.061)	0.221*(0.121)	244	0.267	
Outdoor Recreation	0.330**(0.146)	-0.166*(0.095)	-0.034(0.130)	244	0.896	
Transportation	-0.131(0.124)	-0.157*(0.086)	-0.085(0.110	244	0.621	
Financial & Professional	0.014(0.176)	-0.130(0.103)	-0.042(0.145)	244	0.781	
Panel C: Synthetic Control						
Healthcare & Medical	0.146(0.134)	0.334***(0.076)	0.410***(0.106)	122	0.189	0.603 / 0.397
Public agencies	-0.208(0.350)	0.251*(0.145)	0.342**(0.159)	122	0.062	0.575 / 0.425
Retail & Shopping	-0.057(0.108)	0.151***(0.057)	0.297***(0.068)	122	0.130	0.478 / 0.522
Entertainment	-0.050(0.104)	0.125*(0.067)	0.173(0.118)	122	0.049	0.524 / 0.476
Energy & Storage	-0.014(0.131)	0.024(0.065)	0.215**(0.085)	122	0.049	0.940 / 0.060
Accommodation	0.168**(0.084)	-0.006(0.057)	0.462***(0.105)	122	0.240	0.678 / 0.322
Education	-0.059*(0.034)	-0.052(0.054)	0.228***(0.080)	122	0.110	0.351 / 0.649
Outdoor Recreation	0.260*(0.160)	-0.216***(0.083)	0.010(0.098)	122	0.116	0.839 / 0.161
Transportation	-0.226*(0.136)	-0.222***(0.068)	-0.027(0.095)	122	0.073	0.890 / 0.110
Financial & Professional	-0.252**(0.124)	-0.292***(0.113)	0.118(0.143)	122	0.069	1.000 / 0.000

Coefficients represent log-point changes in visitation relative to control groups. Standard errors in parentheses are heteroskedasticity-robust (HC3). Statistical significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Weights in Panel C reflect the contribution of Religious and Residential controls. All models include period fixed effects.

(b) Retail: deferred consumption and rapid recovery

Retail & Shopping POI experienced minimal disruption during the flood itself but rebounded sharply in the recovery phase. Under SCM-DiD, visitation rose by +34.6% ($p < 0.01$), likely driven by pent-up demand, restocking, and normalization of supply chains. Unlike the Healthcare sector, which absorbs systemic shock, Retail functions as a barometer of economic reactivation and consumer confidence.

(c) Transportation: acute disruption and structural fragility

Transportation POI were among the most negatively affected. SCM-DiD estimates indicate a −19.9% decline during flooding ($p < 0.01$), with no significant recovery thereafter. These patterns reflect direct infrastructure damage (e.g., inundated roads, disrupted transit)[51, 52], safety concerns, and altered commuting behaviors. The lack of post-flood rebound highlights vulnerabilities in Lagos's mobility systems and the absence of short-term recovery mechanisms.

(d) Outdoor recreation: discretionary contraction and behavioral sensitivity

Outdoor Recreation POI experienced a −19.4% decline in visitation during flooding ($p < 0.01$), with no statistically significant rebound during recovery. These effects underscore the discretionary nature of recreational mobility, which is highly sensitive to environmental stress and often deprioritized in post-disaster contexts.

(e) Accommodation: delayed demand and shelter-driven rebound

Accommodation POI showed muted impacts during flooding but strong post-event rebounds. Visitation increased by +58.8% in the recovery phase ($p < 0.01$) under SCM-DiD, likely

reflecting increased demand for temporary housing by displaced populations. These dynamics mirror findings in post-disaster housing, where private sector lodging often compensates for gaps in public emergency shelter provision[53].

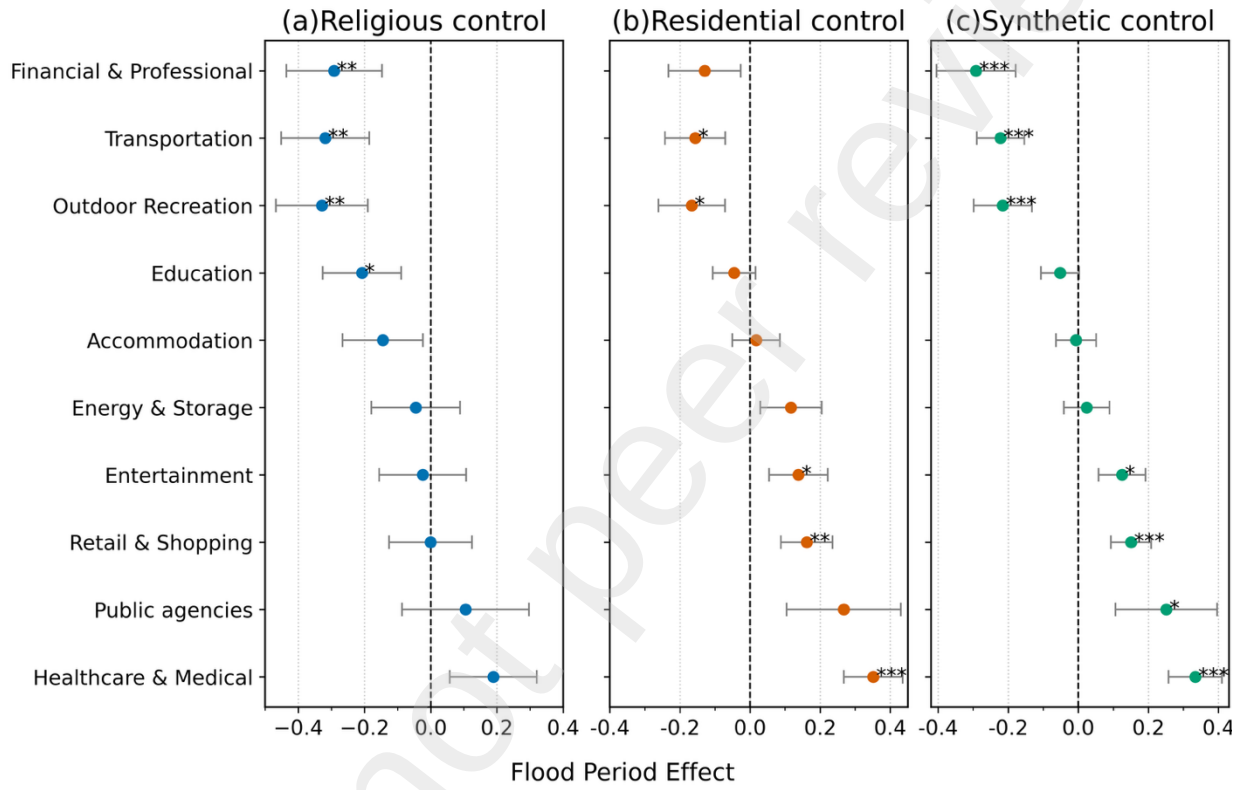

Figure 6. Estimated flood period effects on POI visitation across control strategies. (Statistical significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$)

4.2.2 Comparative evaluation of estimation strategies

The comparative analysis of control strategies underscores the methodological advantages of SCM-DiD in complex, heterogeneous environments. Religious-only controls exhibit inherent cyclical noise (e.g., weekly worship patterns), introducing volatility into effect estimation. Residential-only controls improve baseline stability but are susceptible to displacement effects that may contaminate control trends.

SCM-DiD, by combining Religious and Residential POI through optimized weighting, produces superior pre-treatment fits and more stable post-treatment estimates. It performs particularly well in sectors with subtle or non-linear treatment effects (e.g., Entertainment, Public Agencies), where traditional DiD is prone to over- or underestimation. The results affirm the methodological value of SCM-DiD in capturing sector-specific disruptions without relying on strong parallel trend assumptions.

4.2.3 Synthesis of sectoral heterogeneity and urban resilience implications

The analysis reveals that flood impacts are highly sector-dependent, driven by a complex interplay of infrastructure exposure, behavioral adaptation, and institutional responsiveness. Sectors like Healthcare absorb systemic shocks due to emergency-driven demand, while sectors like Transportation collapse under infrastructural fragility and behavioral avoidance. Meanwhile, sectors such as Retail and Accommodation exhibit rebound dynamics consistent with market-driven recovery trajectories.

Anticipatory behavioral changes, particularly in Education and Transportation, signal the importance of incorporating the preparation phase in future disaster models. Ignoring pre-event adaptation may result in misattributing these changes to flood onset, leading to flawed policy implications.

From a methodological perspective, SCM-DiD offers a critical enhancement for disaster impact research. By constructing unit-specific synthetic counterfactuals, it enables granular estimation of causal effects without rigid parallel trends assumptions. This approach is particularly powerful in dynamic, data-scarce urban environments typical of LMICs, where standard methods may falter.

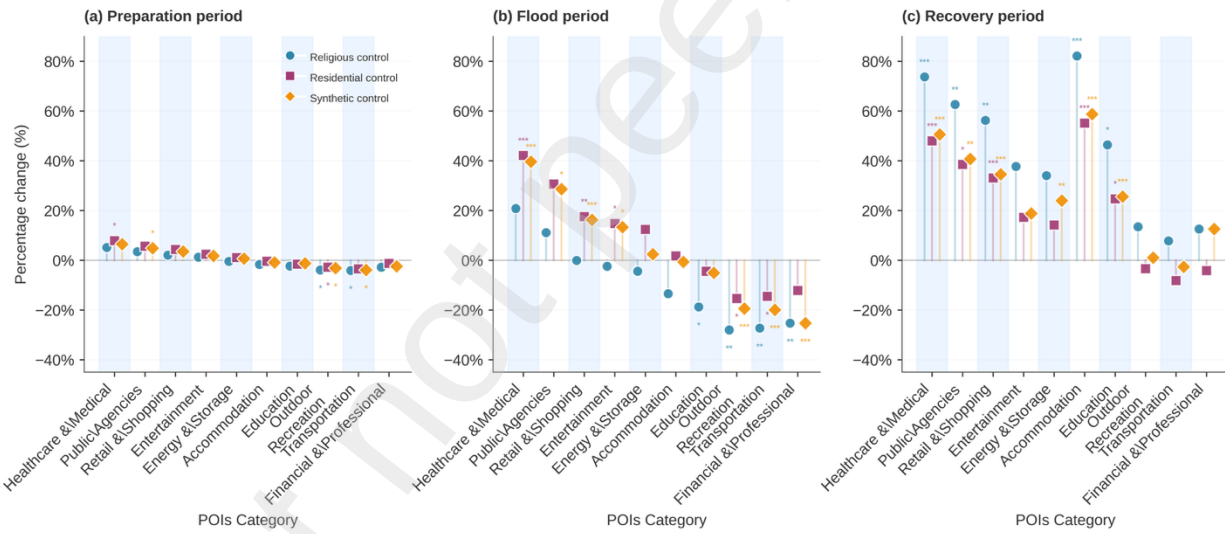

Figure 7. Percentage change in POI visitation across flood phases by control strategy.

5. Conclusion

This study presents a methodological framework for evaluating the causal impacts of urban flooding on human mobility, integrating a DiD approach with a hybrid SCM-DiD and guided by two diagnostic metrics: the MIC and the CV. These innovations directly address persistent

challenges in observational disaster research, particularly within LMICs, where behavioral datasets are limited and exposure events rarely align with natural experiment conditions. Causal impact estimation in flood research is notably complex. Unlike discrete shocks such as earthquakes or hurricanes, which often have well-defined temporal onsets and geographic boundaries, flood events typically unfold gradually, recur across overlapping episodes, and exhibit uneven spatial footprints. This temporal ambiguity complicates the definition of treatment periods and increases the risk of exposure misclassification. Moreover, in highly exposed urban settings, where entire metropolitan regions may be affected simultaneously, the selection of unaffected control areas or even unaffected POI categories becomes inherently problematic. POI categories themselves may be structurally interdependent or exhibit latent behavioral coupling with environmental variables such as precipitation, rendering naïve comparisons vulnerable to bias.

To overcome these limitations, this study introduces a diagnostic control selection procedure that empirically screens candidate POI categories based on their statistical independence from precipitation via MIC and temporal stability across event phases via CV. This dual-screening process enables the identification of suitable counterfactual units even in the absence of spatially distinct control zones. The subsequent application of SCM-DiD further strengthens inference by tailoring synthetic baselines for each treatment group, accounting for category specific pre-treatment dynamics and allowing for non-parallel trends.

Demonstrated through a case study of seasonal flooding in Lagos, Nigeria, the framework successfully identifies heterogeneous mobility responses across urban sectors. The analysis reveals persistent increases in healthcare related visitation, likely driven by waterborne disease and injury, and protracted declines in transportation and recreational activity suggesting infrastructure fragility and behavioral aversion. Sectors such as retail and accommodation exhibit rebound

dynamics, reflecting market-based recovery mechanisms. These findings highlight the sector-specific vulnerabilities and adaptive behaviors that shape post-flood urban dynamics. While the empirical case contributes to understanding behavioral disruptions in one of the most flood-prone areas of word, the broader value of this study lies in its methodological generalizability. The proposed framework provides a scalable and transferable tool for causal inference in environmental impact research, particularly suited to data-scarce urban contexts where randomized exposure is infeasible and administrative records are incomplete or unreliable.

While this study introduces a robust causal framework for assessing flood-induced disruptions to urban mobility, certain analytical constraints remain. The use of aggregated POI categories, while effective for sectoral inference, may overlook finer intra-sectoral dynamics or localized displacement patterns. Moreover, although the SCM-DiD approach mitigates key biases, unobserved factors such as infrastructure failures or policy shifts may still confound mobility responses. Future research should build on this framework by incorporating longitudinal behavioral trajectories, linking mobility changes to health or economic outcomes, and conducting cross-city comparisons to test external validity across varying infrastructural and institutional settings. These extensions will enhance the relevance of the framework for evaluating climate-induced disruptions in complex urban systems and support the design of targeted resilience strategies in flood-prone, data-constrained LMIC settings.

Credit authorship contribution statement

Lele Zhang: Conceptualization, Methodology, Data curation, Writing – original draft. Xin (Bruce) Wu: Conceptualization, Methodology, Writing – review & editing. Kailun Liu: Conceptualization, Writing – review & editing. Mehedi, Md Abdullah Al: Writing – review & editing. Jiashu Zhou:

Writing – review & editing. Virginia Smith: Methodology, Writing – review & editing. Chenfeng Xiong: Conceptualization, Methodology, Writing – review & editing, Funding acquisition.

Declaration of competing interest

The authors declare no conflicts of interest related to this work.

Data availability

Data will be made available on request.

Acknowledgments

The authors express our sincere gratitude to all those who supported us throughout this research. Special thanks to the US National Institutes of Health (NIH) for their financial support through grant number U54TW012041-02 entitled "Role of Data Streams in Informing Infection Dynamics in Africa – INFORM-Africa."

Reference

1. Rogers, J.S., et al., *The role of climate and population change in global flood exposure and vulnerability.* Nature Communications, 2025. **16**(1): p. 1287.
2. Tellman, B., et al., *Satellite imaging reveals increased proportion of population exposed to floods.* Nature, 2021. **596**(7870): p. 80-86.
3. Rentschler, J., M. Salhab, and B.A. Jafino, *Flood exposure and poverty in 188 countries.* Nature communications, 2022. **13**(1): p. 3527.
4. Park, K. and E.H. Lee, *Urban flood vulnerability analysis and prediction based on the land use using Deep Neural Network.* International Journal of Disaster Risk Reduction, 2024. **101**: p. 104231.
5. Chang, H., et al., *Assessment of urban flood vulnerability using the social-ecological-technological systems framework in six US cities.* Sustainable Cities and Society, 2021. **68**: p. 102786.

6. Hu, S., et al., *Multi-crowdsourced data fusion for modeling link-level traffic resilience to adverse weather events.* International Journal of Disaster Risk Reduction, 2024. **112**: p. 104754.
7. Serre, D. and C. Heinzlef, *Assessing and mapping urban resilience to floods with respect to cascading effects through critical infrastructure networks.* International Journal of Disaster Risk Reduction, 2018. **30**: p. 235-243.
8. Zhu, Y., et al., *Burden of infant mortality associated with flood in 37 African countries.* Nature Communications, 2024. **15**(1): p. 10171.
9. Kakinuma, K., et al., *Flood-induced population displacements in the world.* Environmental Research Letters, 2020. **15**(12): p. 124029.
10. Haque, A.N., *Climate risk responses and the urban poor in the global South: the case of Dhaka's flood risk in the low-income settlements.* International Journal of Disaster Risk Reduction, 2021. **64**: p. 102534.
11. Shuaibu, A., et al., *Flood risk assessment and mapping in the Hadejia River Basin, Nigeria, using hydro-geomorphic approach and multi-criterion decision-making method.* Water, 2022. **14**(22): p. 3709.
12. Ndimele, P.E., et al., *Vulnerability, resilience and adaptation of Lagos coastal communities to flooding.* Earth Science, Systems and Society, 2024. **4**(1): p. 10087.
13. Asah, F.N., P. Nielsen, and J.I. Sæbø. *Challenges for health indicators in developing countries: misconceptions and lack of population data*. in *Information and Communication Technologies for Development: 14th IFIP WG 9.4 International Conference on Social Implications of Computers in Developing Countries, ICT4D 2017, Yogyakarta, Indonesia, May 22-24, 2017, Proceedings 14*. 2017. Springer.
14. Olagunju, K., *Evaluating traffic congestion in developing countries. A case study of Nigeria.* Journal of the Chartered Institute of Logistics and Transport-Nigeria, 2015. **2**(3): p. 23-26.
15. Kaiser, N. and C.K. Barstow, *Rural transportation infrastructure in low-and middle-income countries: a review of impacts, implications, and interventions.* Sustainability, 2022. **14**(4): p. 2149.
16. Macharia, P.M., et al., *A geospatial database of close-to-reality travel times to obstetric emergency care in 15 Nigerian conurbations.* Scientific data, 2023. **10**(1): p. 736.
17. Nashwan, A.J., et al., *Impact of natural disasters on health disparities in low-to middle-income countries.* Discover Health Systems, 2023. **2**(1): p. 23.
18. Khatri, R.B., et al., *Preparedness, impacts, and responses of public health emergencies towards health security: qualitative synthesis of evidence.* Archives of public health, 2023. **81**(1): p. 208.
19. Batra, M. and B. Erbas, *Extreme Weather, Vulnerable Populations, and Mental Health: The Timely Role of AI Interventions.* International Journal of Environmental Research and Public Health, 2025. **22**(4): p. 602.
20. Zhang, L., et al., *Next generation National Household Travel Survey National Origin Destination Data Passenger Origin-Destination Data Methodology Documentation.* Federal Highway Administration.[Google Scholar], 2021.

21. Liu, X., et al., *A new approach to estimating flood-affected populations by combining mobility patterns with multi-source data: A case study of Wuhan, China.* International Journal of Disaster Risk Reduction, 2021. **55**: p. 102106.
22. Li, B., et al., *Mobility behaviors shift disparity in flood exposure in US population groups.* International Journal of Disaster Risk Reduction, 2024. **108**: p. 104545.
23. Xiong, C., et al., *Mobile device data reveal the dynamics in a positive relationship between human mobility and COVID-19 infections.* Proceedings of the National Academy of Sciences, 2020. **117**(44): p. 27087-27089.
24. Hu, S., et al., *A big-data driven approach to analyzing and modeling human mobility trend under non-pharmaceutical interventions during COVID-19 pandemic.* Transportation Research Part C: Emerging Technologies, 2021. **124**: p. 102955.
25. Garg, S., et al., *Unlocking the full potential of Sentinel-1 for flood detection in arid regions.* Remote Sensing of Environment, 2024. **315**: p. 114417.
26. Adedeji, O., et al., *An assessment of flood event along Lower Niger using Sentinel-1 imagery.* Environmental monitoring and assessment, 2021. **193**: p. 1-17.
27. OKODUWA, K., *ADOPTING SENTINEL-1 SAR DATA FOR FLOOD MAPPING: A CASE STUDY OF BORNO STATE, NORTHEASTERN NIGERIA* OKODUWA, KA, AMAECHI, CF, BIOSE, E., OMORUYI, CI 2 AND HENRY, B.* Ethiopian Journal of Environmental Studies & Management, 2024. **17**(4): p. 524-538.
28. Faisal Koko, A., et al., *Analyzing urban growth and land cover change scenario in Lagos, Nigeria using multi-temporal remote sensing data and GIS to mitigate flooding.* Geomatics, Natural Hazards and Risk, 2021. **12**(1): p. 631-652.
29. Shiba, K., et al., *Causal inference in studying the long-term health effects of disasters: challenges and potential solutions.* American journal of epidemiology, 2021. **190**(9): p. 1867-1881.
30. Zhang, C., et al., *Spatial-temporal evolution of influencing mechanism of urban flooding in the Guangdong Hong Kong Macao greater bay area, China.* Frontiers in Earth Science, 2023. **10**: p. 1113997.
31. Jia, R., X. Ma, and V.W. Xie, *Expecting floods: Firm entry, employment, and aggregate implications*. 2022, National Bureau of Economic Research.
32. Risser, M.D., M. Ombadi, and M.F. Wehner, *Granger causal inference for climate change attribution.* Environmental Research: Climate, 2025. **4**(2): p. 022001.
33. Goodman-Bacon, A., *Difference-in-differences with variation in treatment timing.* Journal of econometrics, 2021. **225**(2): p. 254-277.
34. Doudchenko, N. and G.W. Imbens, *Balancing, regression, difference-in-differences and synthetic control methods: A synthesis*. 2016, National Bureau of Economic Research.
35. Reshef, D.N., et al., *Detecting novel associations in large data sets.* science, 2011. **334**(6062): p. 1518-1524.
36. Torres, R., et al., *GMES Sentinel-1 mission.* Remote sensing of environment, 2012. **120**: p. 9-24.
37. Cotugno, A., et al., *A framework for calculating peak discharge and flood inundation in ungauged urban watersheds using remotely sensed precipitation*

data: a case study in freetown, sierra leone. Remote Sensing, 2021. **13**(19): p. 3806.
38. Mehedi, M.A.A., V. Smith, and P. Kremer. *Comparing Urban Flood Dynamics Using SAR Imagery and Google Earth Engine: Case Studies In Dhaka And Houston*. in *AGU Fall Meeting Abstracts*. 2022.
39. Hersbach, H., et al., *The ERA5 global reanalysis, quarterly journal of the royal meteorological society.* 2020.
40. Teutschbein, C. and J. Seibert, *Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods.* Journal of hydrology, 2012. **456**: p. 12-29.
41. Jiang, Q., et al., *Evaluation of the ERA5 reanalysis precipitation dataset over Chinese Mainland.* Journal of hydrology, 2021. **595**: p. 125660.
42. Lu, J. and X.S. Zhou, *Virtual track networks: A hierarchical modeling framework and open-source tools for simplified and efficient connected and automated mobility (CAM) system design based on general modeling network specification (GMNS).* Transportation Research Part C: Emerging Technologies, 2023. **153**: p. 104223.
43. Luo, W., et al., *Revealing human mobility trends during the SARS-CoV-2 pandemic in Nigeria via a data-driven approach.* South African Journal of Science, 2023. **119**(5-6): p. 1-9.
44. Woods, A., *Medium-range weather prediction: The European approach*. 2005: Springer Science & Business Media.
45. Reshef, Y.A., et al., *Measuring dependence powerfully and equitably.* Journal of Machine Learning Research, 2016. **17**(211): p. 1-63.
46. Callaway, B. and P.H. Sant'Anna, *Difference-in-differences with multiple time periods.* Journal of econometrics, 2021. **225**(2): p. 200-230.
47. Arkhangelsky, D., et al., *Synthetic difference-in-differences.* American Economic Review, 2021. **111**(12): p. 4088-4118.
48. Jang, M., et al., *Changes in healthcare utilization after the 2022 Seoul metropolitan flood: Applying a generalized synthetic control approach.* GeoHealth, 2024. **8**(10): p. e2024GH001084.
49. Ahmed, S.H., et al., *Water-related diseases following flooding in South Asian countries–a healthcare crisis.* European Journal of Clinical and Experimental Medicine, 2024. **22**(1): p. 232-242.
50. Cross, B.R., *Risk of waterborne disease after Pakistan floods*. 2024, British Red Cross. Retrieved April.
51. Akanmu, A.A. and U.O. Salisu, *The effects of environmental peculiarities on transportation infrastructure performance in Lagos Metropolis, Nigeria: residents' experiences.* Zeszyty Naukowe. Transport/Politechnika Śląska, 2024.
52. Beitelmal, W.H., et al., *Exploring adaptation strategies to mitigate climate threats to transportation infrastructure in Nigeria: Lagos City, as a case study.* Climate, 2024. **12**(8): p. 117.
53. Leite, F. and J. O'Connor, *Understanding Hosting Communities as a Stakeholder in the Provision of Urban Infrastructure to Displaced Populations.*

Appendix

Appendix 1. Bias correction method for ERA5 precipitation estimates

ERA5 precipitation data are known to exhibit systematic biases under certain conditions[41]. Since the primary objective of this study is to analyze temporal patterns and variability—rather than absolute magnitudes—of rainfall, a simple linear scaling approach is employed to correct both the daily accumulated total precipitation and daily maximum total precipitation values [40]. This correction preserves the temporal dynamics inherent in the original ERA5 dataset.

$$P^{*}_{ERA5}(d) = P_{ERA5}(d)\left\lfloor \frac{\mu_{june-july}(P_{ERA5}(d))}{\mu_{june-july}(P_{IMERG}(d))} \right\rfloor \quad \textbf{(a1)}$$

Where:

$P^{*}_{ERA5}(d)$ is the bias-corrected ERA5 ATP at day d

$P_{ERA5}(d)$ is the original ERA5 ATP at day d

$P_{IMERG}(d)$ is the ATP from NASA's Earth IMERG dataset at day d

$\mu_{june-july}(\cdot)$denotes the mean-value of ATP across June and July 2020

The scaling factor, defined as the ratio between the mean IMERG and ERA5 ATP over June–July 2020, is also applied to correct MTP values. These two months are selected due to their anomalously high precipitation levels in the study area during the 2020 summer season.

Appendix 2. POI labels and categories

Table a1. POI category consolidation from OSM/Google Map Labels

Number	POI Categories	POI labels in OSM and Google Map
1	Accommodation	Hotel, lodging
2	Education	University, secondary school, primary school, school, college, childcare, library

3	Healthcare & Medical	Hospital, dentist, doctor, veterinary care, physiotherapist, pharmacy, drugstore, clinic
4	Financial & Professional	Accounting, lawyer, electrician, plumber, roofing contractor, moving company, car dealer, car repair, locksmith, carwash, laundry, construction, industrial, manufacture, farm auxiliary studio, office, bank, ATM, insurance agency, real estate agency
6	Religious	Mosque, church, Hindu temple, place of worship
5	Residential	Apartments, detached, house
7	Transportation	Airport, bus station, taxi stand, train station, transit station, hangar, parking, ferry terminal, car rental, travel agency, bicycle store, transportation
8	Energy & Storage	Gas station, storage, warehouse, fuel
9	Outdoor recreation	Nature reserve, track, pitch, amusement park, campground, golf course, playground, stadium, park
10	Public agencies	City hall, fire station, police, local government office, courthouse, embassy, civic, community center, post office, service
11	Retail & Shopping	Bookstore, clothing store, convenience store, department store, electronics store, furniture store, hardware store, home goods store, jewelry store, liquor store, supermarket, commercial, common, marketplace, pet store, shoe store, shopping mall, store
12	Entertainment	Art gallery, movie theater, nightclub, casino, bowling alley, Sports center, museum, cinema, arts center, events venue, movie rental, nightclub, fitness center, tourist attraction, spa, beauty salon, hair care, gym, swimming pool, painter, horse-riding

1

2

Causal inference of urban flood impacts on mobility using difference-in-differences and synthetic control in Lagos, Nigeria

Le-Le Zhang[1], Xin (Bruce) Wu[1], Kai-Lun Liu[1], Md Abdullah Al, Mehedi[2], Jia-Shu Zhou[3], Virginia Smith[1], Chenfeng Xiong[1]*

[1] Department of Civil and Environmental Engineering, Villanova University, Villanova, PA 19085, USA

[2] First Street Foundation, Brooklyn, NY 11201, USA

[3]Stanford University, Stanford, CA 94305, USA

* Corresponding Author, 800 Lancaster Ave, Villanova, PA, 19085 USA
Chenfeng.xiong@villanova.edu