Urban Climate Counterfactuals: A Causal Dataset for Street-Level Heat Mitigation Interventions

Ahanaf Ariq
Ideal School and College
ariqahanaf@gmail.com

Abstract

UrbanCLIM-Causal enables AI to answer: *Which $20k cooling intervention reduces heat deaths most in this district?* – by providing the first openly shareable, causally rich dataset linking urban interventions to street-level microclimate impacts. By combining satellite imagery, ground sensors and physics-based counterfactual simulations, it quantifies cooling effects of interventions (e.g., tree planting, cool roofs). The dataset aims to accelerate robust, policy-guiding AI tools for climate science, urban planning and public health, with equitable coverage and global applicability. We address reviewer concerns by detailing the simulation validation protocol, formalizing equity metrics and providing a systematic comparison to prior work.

1 Problem and Scientific Importance

Urban heat islands (UHIs) are a critical global challenge, escalating heat-related morbidity and energy demand and disproportionately impacting vulnerable populations [**?**]. Planning is hindered by a data bottleneck: the absence of high-resolution, causally informed datasets linking specific cooling interventions (e.g., tree planting, cool roofs, permeable surfaces) to localized (10–50 m) temperature/heat-index changes while controlling for confounders (weather variability, morphology, socioeconomics). Observational data are largely correlative; counterfactuals are missing. Recent GPU advances (e.g., NVIDIA H100) now enable city-scale ENVI-met simulations at 10–50m resolution, making this dataset feasible for the first time.

2 AI Task Definition

UrbanCLIM-Causal supports **causal effect estimation and counterfactual prediction**: (i) estimate Δ temperature/heat index attributable to specific interventions (e.g., +10% canopy, 50% cool roofs on a block); (ii) forecast what temperatures would have been if an intervention had (or had not) occurred; (iii) recommend intervention portfolios to meet cooling targets under budget/space/equity constraints; and (iv) generalize across cities/climates and intervention types.

3 Dataset Rationale (Bottleneck) and Novelty

The core bottleneck is moving beyond observational correlations to causal insights. Remote sensing provides coverage but not counterfactuals; ground sensors offer fidelity but are sparse and rarely aligned to intervention timing. Existing resources typically lack standardized, georeferenced intervention metadata paired with pre/post microclimate and validated counterfactuals. UrbanCLIM-Causal explicitly constructs *paired observations with simulated counterfactuals*, supplying the structure required for causal learning and interventional benchmarking in real urban environments.

Novelty and Comparison to Prior Work. While datasets like ClimateSet and AI2Earth provide valuable observational data, they do not contain the necessary causal ground truth. Our key distinction is the **physics-based counterfactual generation** and the **policy-driven intervention framing**. Existing tools for heat mitigation planning are often based on simple correlations or non-validated models; UrbanCLIM-Causal provides a principled benchmark for causal AI models to quantify the true effect of interventions. We also include a **pilot release** of data for one city (Phoenix) with full schema and sample visualizations to demonstrate feasibility.

4 Dataset Specification

Inputs: (i) Multispectral and thermal satellite imagery (Sentinel-2, Landsat; LST from TIRS/ECOSTRESS), (ii) urban geospatial layers (building footprints/heights/materials; canopy; roads; imperviousness), (iii) aggregated sociodemographics, (iv) meteorology (ERA5-Land), (v) intervention metadata (type, timing, intensity; polygons/points). The specific intervention types included are: **tree density change**, **albedo change (cool roofs/pavements)** and **permeable surface introduction**.
Labels: (i) Observed microclimate: air temperature/humidity from low-cost sensors (pre/post), aggregated to street/block level; (ii) Simulated counterfactual microclimate: ENVI-met/WRF-Urban outputs for the same spatiotemporal windows under a *no-intervention* scenario, validated to target RMSE $< 0.5\,^{\circ}\mathrm{C}$ in pilot cities.
Scale: Initial release covers 5–10 diverse cities (e.g., Dhaka, Barcelona, Phoenix); 10–50 m resolution; hourly/sub-hourly; multi-year (2015–2025).
License: CC BY 4.0.

5 Data-Creation Pathway and Validation Protocol

(1) Process open EO (Sentinel, Landsat) and reanalysis (ERA5-Land). (2) Integrate municipal GIS and OpenStreetMap layers. (3) Partner with city/citizen networks for targeted low-cost sensors in intervention/control areas, prioritizing heat-vulnerable neighborhoods (identified via CDC/WHO heat-risk indices) to correct historical under-monitoring. (4) Run ENVI-met/WRF-Urban to generate counterfactuals; validate simulations against sensors. (5) Curate intervention metadata from city reports and studies. (6) Harmonize under a documented schema; release code and loaders.

Simulation Validation Protocol. To achieve the target RMSE $< 0.5\,^{\circ}\mathrm{C}$, we employ a three-stage calibration process:

(i) **Morphology Calibration:** Use high-resolution GIS data (building height, material, land cover) to accurately model the urban geometry in the physics simulators.

(ii) **Meteorological Calibration:** Calibrate the model's boundary conditions (wind speed, solar radiation, air temperature) using local meteorological station data and ERA5-Land reanalysis.

(iii) **Microclimate Validation:** Compare simulated air temperature and relative humidity outputs against a network of ground-truth low-cost sensors in a control (no-intervention) area for a minimum of 7 consecutive clear-sky summer days. The model parameters are iteratively tuned until the mean RMSE for air temperature is below $0.5\,^{\circ}\mathrm{C}$. This validated model is then used to generate the counterfactuals.

6 Acceleration Potential and Benchmarks

UrbanCLIM-Causal enables: (i) benchmarking of causal inference and interventional generalization in complex geospatial settings; (ii) policy-precise guidance on *degrees cooled per dollar* and expected reduction in heat illness; (iii) rapid design of targeted, equitable adaptation strategies; (iv) a testbed for generative urban design models grounded in counterfactuals.

Formalized Equity Benchmarks. With standard performance metrics, we formalize equity-aware benchmarks:

- **Disaggregated Cooling Efficacy:** Measure the causal effect of interventions across different socio-economic strata (e.g., by census block income quartile) to ensure cooling benefits are not concentrated in affluent areas.
- **Heat Vulnerability Reduction:** Benchmark models on their ability to recommend interventions that maximize the reduction in heat-related health risk (using CDC/WHO indices) in the most vulnerable neighborhoods.
- **Bias Audit:** Provide a framework for auditing models for potential biases in intervention recommendation, ensuring that high risk areas are not systematically overlooked due to data sparsity or model limitations.

7 Cost, Scalability and Governance

Core Phase (pilot): 10k–20k CPU-hours for simulations ($5k–$10k) plus sensors ($2k–$5k). Scalability via open data, parallelizable simulations, citizen science and federated city contributions under a common schema and QC. The cost estimate is based on the initial pilot city (Phoenix) and is considered a minimum, with a contingency budget for local engagement and data curation.

Governance and Maintenance. The dataset will be stewarded by a dedicated, non-profit consortium.

- **Versioning:** We will adopt a semantic versioning scheme (e.g., v1.0, v1.1) with a clear changelog and persistent DOIs for each major release.
- **Update Cycle:** A bi-annual update cycle is planned to incorporate new sensor data, updated urban morphology and new city simulations.
- **Contributor Onboarding:** A formal process will be established for federated contributions from city partners and researchers, including a standardized data submission pipeline and quality control checks.

8 Ethics, Privacy, Governance

Sociodemographics aggregated (census-block or coarser). Sensor data anonymized and georeferenced to public segments only. Proactive bias audits (placement equity and algorithmic fairness); prioritized sensor deployment in underserved neighborhoods. Transparent governance, versioning and documented data quality.

9 Broader Impacts & Shareability

Global health benefits (reduced heat risk); resilient urban planning and efficient public spending; community engagement via citizen science; public hosting (e.g., Hugging Face/Google Cloud) with documentation, tutorials, open-source code, API and contribution pathways.

References

[1] IPCC (2021). Climate Change 2021: The Physical Science Basis.

[2] Oke, T. R. (1987). *Boundary Layer Climates*. Routledge.

[3] Georgescu, M., et al. (2014). Urbanization and thermal environment of Phoenix. *JAMC*, 53(1), 1–21.

[4] ClimateSet: A Global Dataset for Climate-Aware Machine Learning. (2023). *NeurIPS Datasets and Benchmarks*.

[5] AI2Earth: A Platform for AI-Powered Climate Research. (2022). *Nature Climate Change*.

[6] Microsoft Planetary Computer: A Platform for Geospatial Data Science. (2021). *Online Resource*.

[7] ENVI-met V4: A new generation of the microclimate model. (2018). *Journal of Environmental Modelling and Software*.

[8] WRF-Urban: Modeling the urban boundary layer. (2010). *Bulletin of the American Meteorological Society*.