APPENDICES

Urban Heat, Health, and Vulnerability in South African Cities:

A Mixed-Methods Approach to Predictive Modeling

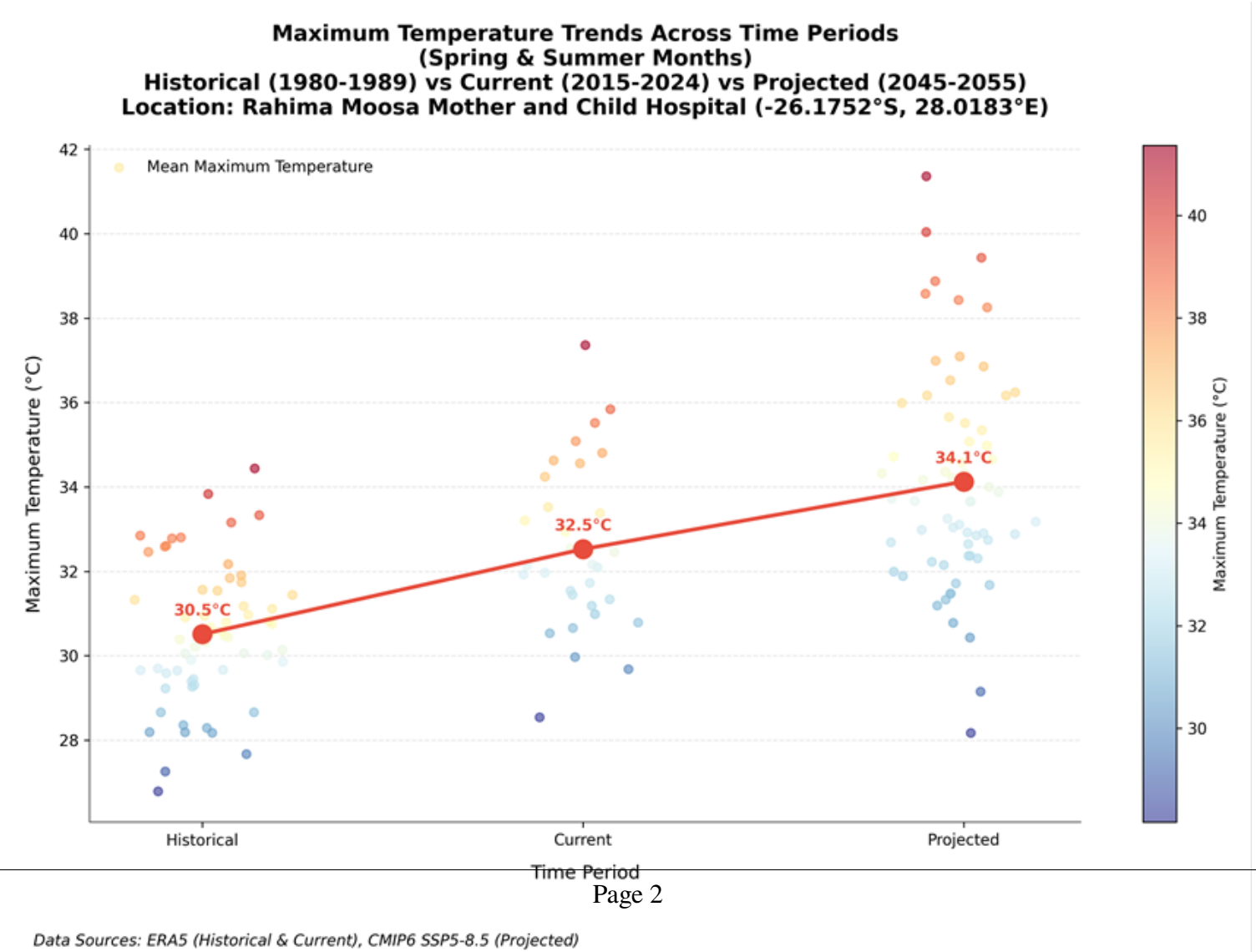
Craig Parker

Wits School of Public Health

April 2025

Appendix A: Supplementary Figures

A.1 Temperature Trends in Johannesburg



A.2 Seasonal Heat Patterns

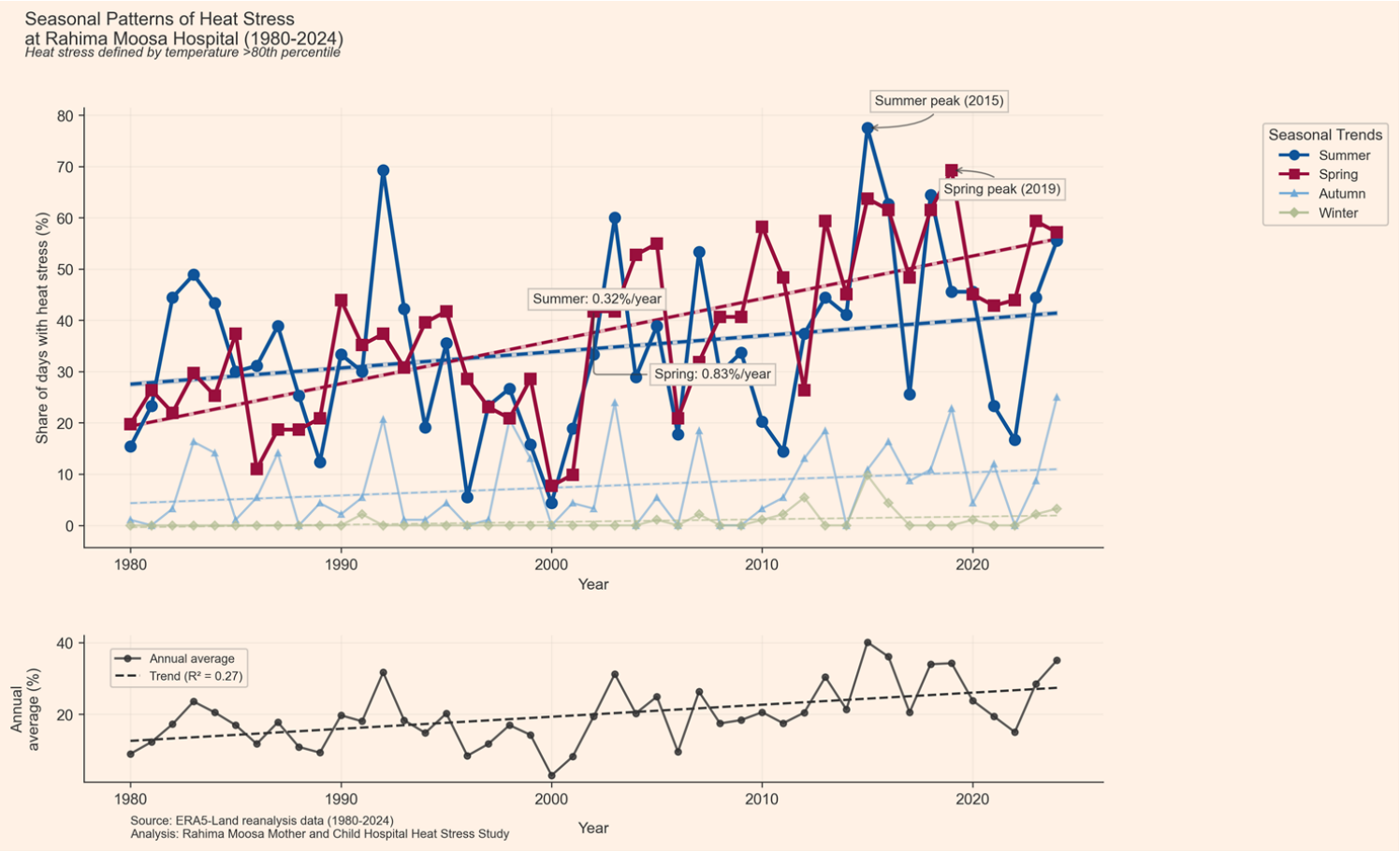


Figure 2: Seasonal heat patterns in the Johannesburg region from 1980-2024, highlighting increasing summer temperature extremes.

A.3 Global Temperature Comparison

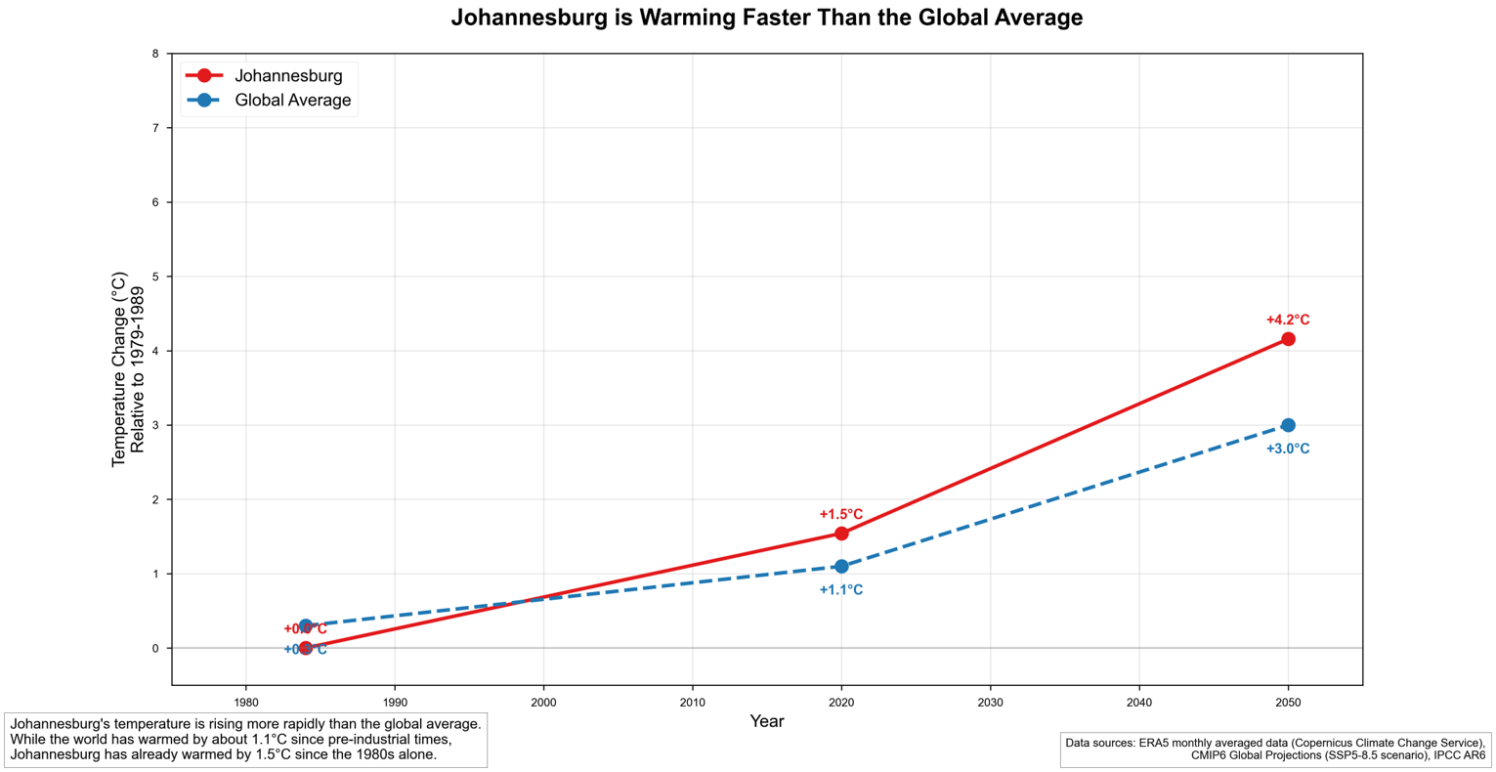


Figure 3: Comparison of global temperature anomalies versus Johannesburg-specific trends, showing amplified urban warming.

A.4 Urban Heat Island Analysis

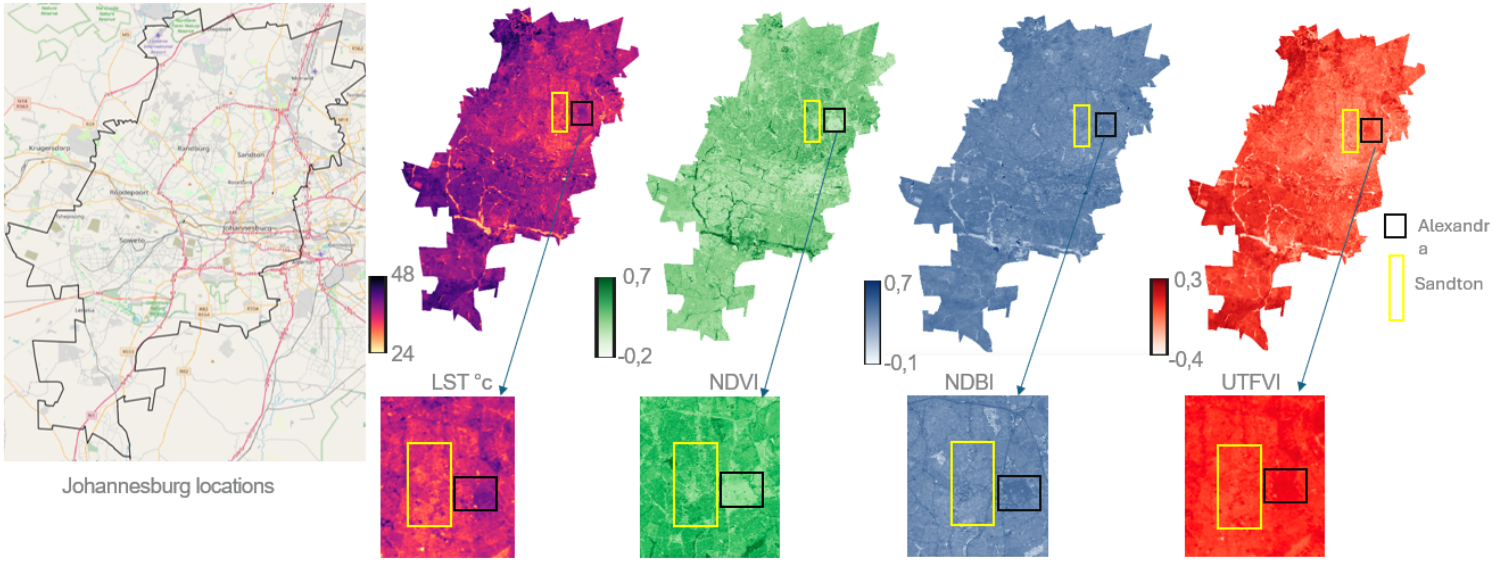


Figure 4: Urban heat island analysis using land surface temperature (LST), vegetation index (NDVI), built-up index (NDBI), and urban thermal field variance index (UTFVI) for Johannesburg.

A.5 Preliminary Heat Vulnerability Index

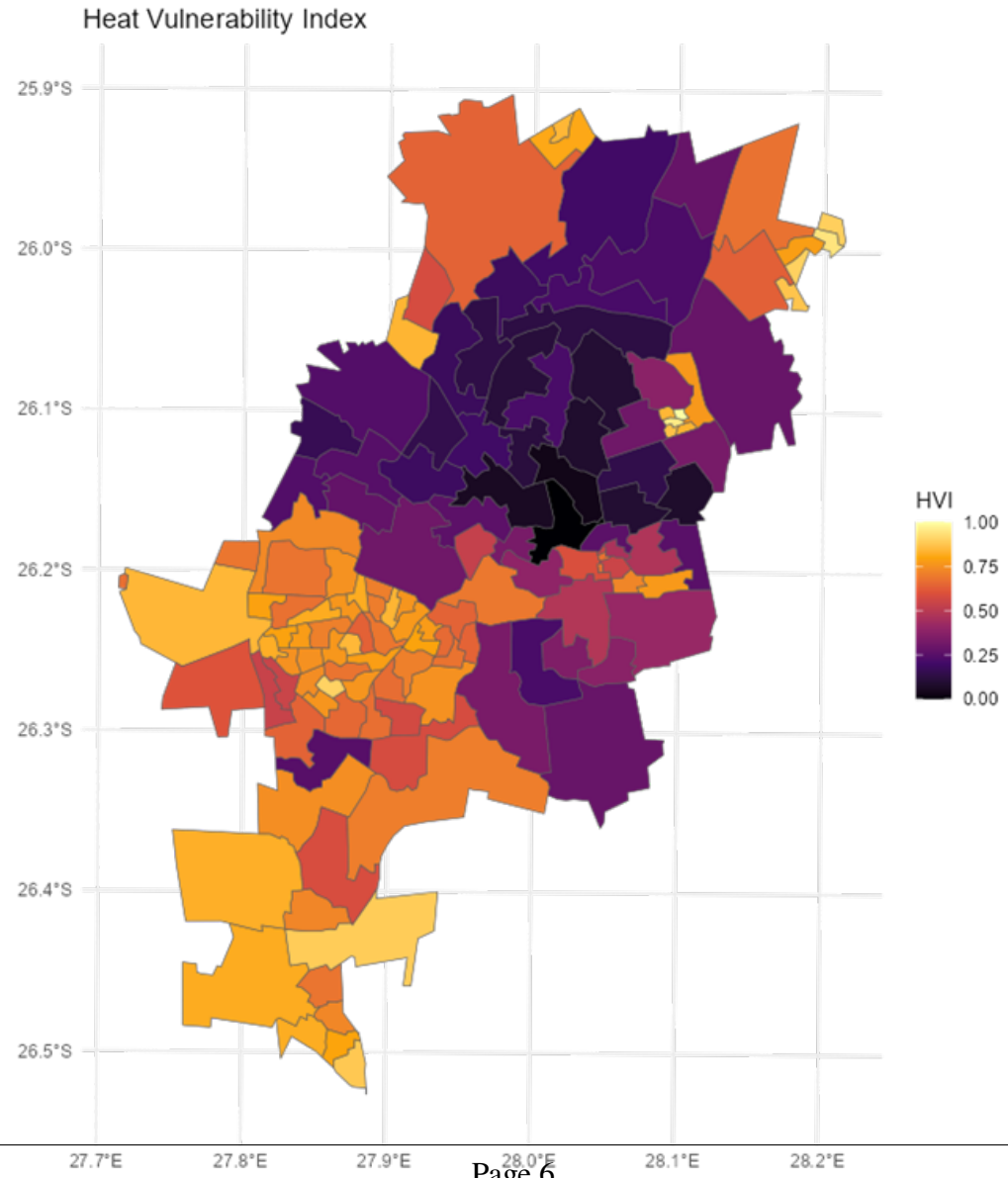


Figure 5: Preliminary heat vulnerability index (HVI) map for Johannesburg, integrating socioeconomic, demographic, and environmental factors.

A.6 Analytical Framework

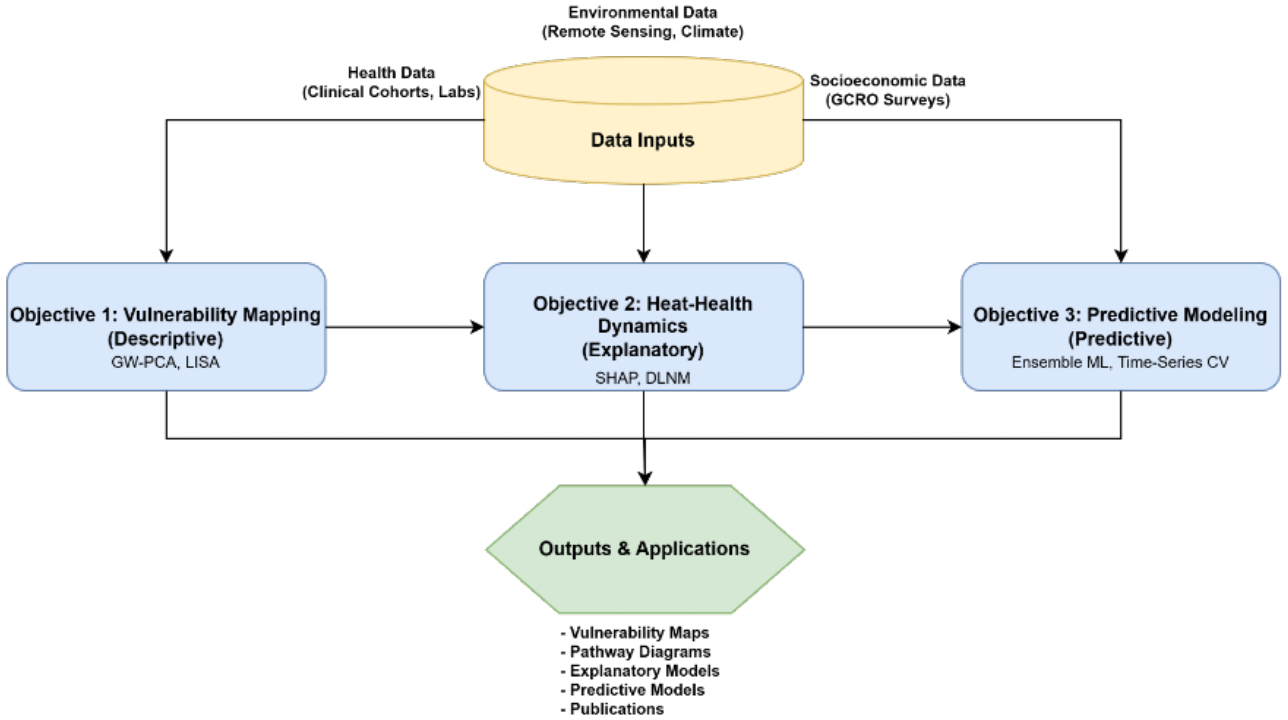
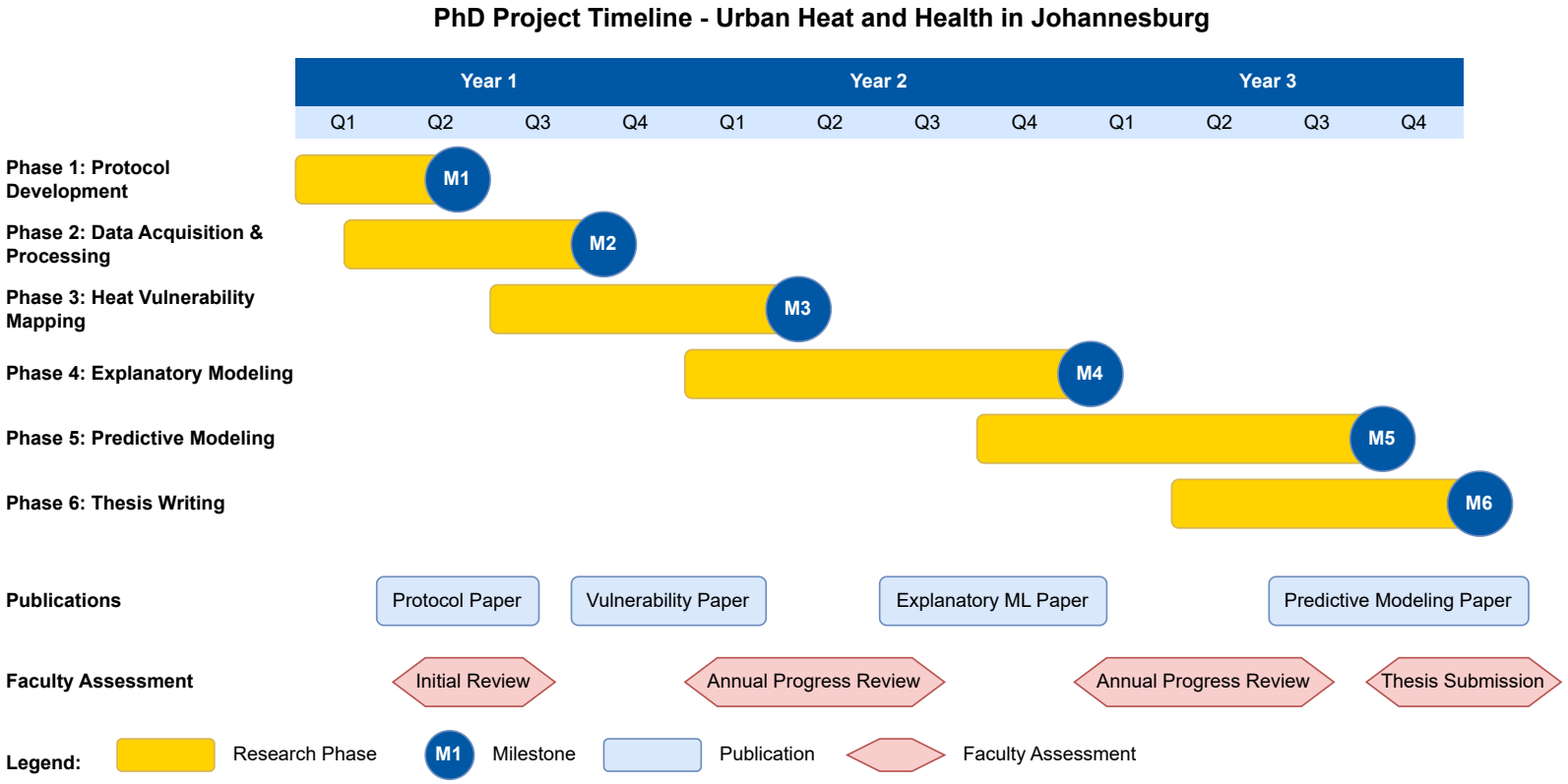


Figure 6: Detailed analytical framework showing the integration of vulnerability mapping, heat-health dynamics, and predictive modeling components.

F.9 Project Timeline (GANTT Chart)



Milestone Descriptions:

M1: Protocol Finalization (Month 3)
M2: Data Acquisition Completion (Month 9)
M3: Vulnerability Mapping Completion (Month 18)
M4: Model Validation (Month 30)
M5: Predictive Model Deployment (Month 33)
M6: Thesis Submission (Month 36)

Publication Plan:

1. Protocol Paper (Month 4): 'Leveraging data science and machine learning for urban climate adaptation in two major African cities: a HE2AT Center study protocol'
2. Vulnerability Paper (Month 9): 'Quantifying intra-urban socio-economic and environmental vulnerability to extreme heat events in Johannesburg, South Africa'
3. Explanatory ML Paper (Month 19): 'Uncovering Heat-Health Pathways in Johannesburg: An Explanatory Machine Learning Approach Using Harmonized Urban Cohort Data'
4. Predictive Modeling Paper (Month 33): 'Developing Stratified Heat-Health Early Warning Systems for Johannesburg: Integration of Vulnerability Mapping and Predictive Modeling'

Primary risks include data accessibility challenges, computational constraints, and model performance issues. Detailed timelines, activities, and risk mitigation strategies are provided in Appendix C.

Figure 7: Detailed project timeline showing research phases, milestones, and deliverables across the PhD duration.

APPENDIX DOCUMENT

The following tables and appendices are separated from the main document to support word count requirements.

Data Management and Methodological Tables for HE^{2}AT Center Research

Table 1: Key Data Sources for Heat-Health Research

Data Category	Data Source	Description	Key Variables	Relevance
Biomedical Data	Individual Participant Data Platform	Collation of prospectively collected high-quality data from pregnant women & neonates	Maternal health indicators, birth outcomes, demographic data	Study population with high rates of co-morbidities and adverse health outcomes
Climate/Weather Data	ECMWF Forecasts	Outputs from numerical weather prediction system	Temperature, solar irradiance, wind speed, precipitation, pressure	Determination of heat hazard and thermal comfort metrics

Data Category	Data Source	Description	Key Variables	Relevance
	ERA5 Reanalysis	Global reanalysis dataset combining observed data with meteorological models	Temperature, wind speed, precipitation, atmospheric water content	Historical climate exposures assessment
	ERA5-Land	High-resolution land component of the ERA5 climate reanalysis	Surface temperature, precipitation, near-surface winds	Detailed land surface parameter analysis
Remote Sensing Data	SRTM Elevation	Global elevation data (30m resolution)	Elevation	Urban heat island effect assessment
	Sentinel-2 Imagery	High-resolution multispectral satellite imagery	Vegetation coverage, land use classification, NDVI	Land cover and urban morphology analysis
	MODIS Land Surface Temperature	Daily global land surface temperature	Day/night land surface temperature	Heat exposure assessment
Socio-Economic Data	Gauteng City-Region Observatory	GIS data for Gauteng City-Region	Demographics, economics, environmental factors	Socio-economic context for urban areas

Data Category	Data Source	Description	Key Variables	Relevance
	South African Census	National population and housing census	Population density, housing quality, service access	Demographic vulnerability factors
	GCRO Quality of Life Survey	Biennial survey of Gauteng residents	Socioeconomic status, health status, service satisfaction	Subjective vulnerability factors

Table 1: Key Data Sources for Heat-Health Research

Table 2: Data Processing and Integration Workflow

Processing Stage	Key Activities	Responsible Team	Outputs
Pre-processing	Data reformatting, extraction of key variables, alignment with ontologies	Core Data Team	Standardized data formats ready for harmonization
Variable Mapping	Mapping variables to standardized ontologies, using AI tools for suggestions	Harmonization Team	Consistent variable naming and definitions across datasets
Mapping Validation	Cross-checking with original data, expert health review	Core Data Team, Health Experts	Validated variable mappings
Database Population	Application of mappings, transformation of data, de-identification	Core Data Team	Integrated consortium-shared dataset
Climate Data Integration	Automated retrieval of climate variables, spatial and temporal alignment	CSAG/UCT Team	Climate-integrated health dataset
De-identification	Safe Harbor method application, expert determination, geographic aggregation	Core Data Team	RP2 De-identified datasets
Data Analysis	Statistical analysis, machine learning applications	HE^2AT Consortium	Research outputs and inferential data

Processing Stage	Key Activities	Responsible Team	Outputs

Table 2: Data Processing and Integration Workflow

Table 3: Data Access Levels and Security Measures

Data Level	Description	Access Permissions	Security Measures	Retention Period
Original Study Data	Raw, unprocessed health data collected directly from studies	Core Data Team only	Encryption (AES-256), secure UCT servers, restricted access	5 years after project completion
Consortium Shared Data	Processed, harmonized data with limited indirect identifiers	HE^{2}AT Center Consortium partners	TLS protocols, access controls, authentication protocols	Indefinite (unless specified in DTA)
RP2 De-identified Data	Further de-identified data with aggregated geographic information	External Bone Fide Researchers (DAC-approved)	Data Transfer Agreements, ethical approval requirements	Indefinite (unless specified in DTA)
Inferential Data	Aggregated and anonymized data derived from analyses	Open access	NA - No identifying information	Indefinite

Table 3: Data Access Levels and Security Measures

Table 4: Geographic De-identification Techniques

Technique	Description	Application Scenario	Privacy Protection Level
Geographic Aggregation	Aggregating addresses into larger regions	Areas with adequate population density	Medium - Depends on aggregation level
Random Direction and Fixed Radius	Points displaced randomly within fixed radius	High spatial granularity needs	Medium - Predictable displacement bounds
Gaussian Displacement	Random direction with distances following Gaussian distribution	Detailed spatial analysis requirements	High - Variable displacement with population density
Donut Masking	Setting minimum and maximum displacement levels	Preventing both near and far relocations	High - Controls minimum displacement
K-anonymity Assessment	Ensuring each location is indistinguishable from k-1 others	Validation of masking effectiveness	Verification mechanism

Table 4: Geographic De-identification Techniques

Table 5: Data Request and Approval Process

Stage	Description	Responsible Party	Key Considerations
Data Request Submission	Completion of standardized form with research details	External Researcher	Research purposes, required datasets, ethics approval
Preliminary Screening	Confirmation of form completeness and basic compliance	HE^{2}AT Center SteerCo	Resource availability, completeness of application
DAC Review	Evaluation based on scientific, ethical, and feasibility criteria	Data Access Committee	Scientific merit, potential privacy risks, overlap with ongoing research
Decision Communication	Documentation and notification of approval or rejection	DAC	Transparent reasoning, conditions for approval
Data Transfer	Execution of DTA and data provision	Core Data Team	Secure transfer protocols, encryption
Ongoing Monitoring	Regular reviews of compliance with agreement terms	DAC	Audits if necessary

Table 5: Data Request and Approval Process

Table 6: Methodological Approaches for Research Objectives

Table 6: Methodological Approaches for Research Objectives

Objective	Methodological Approaches
Vulnerability Mapping	• Principal Component Analysis • Geospatial analysis using GIS • Satellite imagery integration • Local Indicators of Spatial Association (LISA)
Heat-Health Dynamics	• Random Forests for feature importance • XGBoost with SHAP value interpretation • Multi-scale time-series analysis • Physiological pathway investigation
Predictive Modeling	• Ensemble machine learning approaches • Demographic stratification • Advanced feature selection • Time-series cross-validation

Table 7: Time-Lag Analysis of Heat-Health Effects

Table 7: Time-Lag Analysis of Heat-Health Effects

Temporal Scale	Health Outcomes	Analytical Approach
Immediate		
(0-24 hours)	• Cardiovascular responses	• High-resolution temporal analysis
	• Renal function markers	• Threshold identification
	• Cognitive performance	• Diurnal pattern examination
Short-term		
(1-7 days)	• Inflammatory markers	• Moving average analysis
	• Metabolic changes	• Cumulative exposure modeling
	• Sleep quality metrics	• Change-point detection
Medium-term		
(7-30 days)	• Chronic condition exacerbation	• Distributed lag models
	• Adaptation responses	• Trend analysis
	• Cumulative health impacts	• Physiological pathway assessment

Appendix A: Roles and Responsibilities in Data Management

Role	Responsibilities	Current Personnel	Contact
DMAC PIs	Oversight of data management, plan assessment and updates	Christopher Jack (UCT), Sibusisiwe Makhanya (IBM)	cjack@csag.uct.ac.za, sibusisiwe.makhanya@ibm.com
Health Data Acquisition	Identification of health datasets, DTA development	Craig Parker (RP2)	Craig.parker@witsphr.org
Core Data Team	Data processing, de-identification, quality control	Lisa van Aardenne, Pierre Kloppers, Piotr Wolski, Peter Marsh, Nicholas Brink, Craig Parker	Contact DMAC PIs
Data Access Committee	Review of data access requests, oversight of data sharing	Caradee Wright (SAMRC), Sibusisiwe Makhanya (IBM), Christopher Jack (UCT)	caradee.wright@mrc.ac.za
Information Officer	POPIA compliance, data protection oversight	Sibusisiwe Makhanya (IBM)	sibusisiwe.makhanya@ibm.com

Table 8: Roles and Responsibilities in Data Management

Appendix B: POPIA Compliance Framework

1. **Lawful Processing**: All data processing is conducted exclusively for legitimate research purposes under POPIA Section 27.

2. **Purpose Limitation**: Data is collected and processed solely for examining heat-health relationships and vulnerability patterns.

3. **Data Minimization**: Only essential data elements required for valid scientific analysis are retained.

4. **Information Officer Designation**: A designated Information Officer oversees POPIA compliance with quarterly audits.

5. **Security Safeguards**: Technical and organizational measures including encryption, access controls, and secure transfer protocols are implemented.

6. **Data Subject Participation**: Where applicable, research participants are informed of data usage through ethics committee-approved processes.

7. **Processing Records**: Comprehensive documentation of all data processing activities is maintained.

8. **Impact Assessments**: Regular privacy impact assessments are conducted for high-risk processing activities.

Appendix C: Risk Assessment and Contingency Planning

Risk Category	Potential Issues	Mitigation Strategies
Data Access	• Delays in data acquisition • Restricted access to key datasets • Changes in data provider policies	• Early engagement with data providers • Development of alternative data sources • Flexible research design to accommodate data limitations
Data Quality	• Missing values in critical variables • Inconsistent data collection methods • Temporal or spatial gaps	• Robust imputation methods • Quality assessment protocols • Integration of multiple data sources

Risk Category	Potential Issues	Mitigation Strategies
Computational		
	• Processing limitations for large datasets	• Cloud computing resources
	• Software compatibility issues	• Regular code testing and validation
	• Model convergence failures	• Modular analytical approach
Ethical/Legal		
	• Changes in data protection regulations	• Regular compliance reviews
	• Challenges in maintaining anonymization	• Conservative de-identification approaches
	• Stakeholder concerns about findings	• Stakeholder engagement throughout research

Table 9: Risk Assessment and Mitigation Strategies