A Distributed Computing Prototype for Climate Change Impact Simulation: Case of Nairobi, Kenya

By

Clarian Makungu

170490

Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Information Technology at Strathmore University

School of Computing & Engineering Sciences

Strathmore University

Nairobi, Kenya

June 2025

Declaration and Approval

Declaration

I declare that this work has not been previously submitted and approved for the award of a degree by this or any other University. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

Student's Name: **Clarian Makungu**

Sign: ______________________ Date: **May 22, 2025**

Approval

The thesis of **Clarian Makungu** was reviewed and approved for examination by the following:

Dr. Njeri Ngaruiya
School of Computing & Engineering Sciences,
Strathmore University

Dr. Julius Butime,
Dean, School of Computing & Engineering Sciences,
Strathmore University

Prof. Bernard Shibwabo,
Director of Graduate Studies,
Strathmore University

Abstract

Climate change is one of the most pressing issues facing the world in the 21st century, and Nairobi is no exception. Climate change in Nairobi impacts climate-sensitive variables, including unpredictable rainfall, rising temperatures, and increased frequencies of extreme temperature and precipitation events, significantly affecting current and future urban infrastructures, public health, transportation systems, and livelihood activities. This project sought to develop a distributed computing environment for detailed climate change modeling and effects assessment in Nairobi, Kenya. The prototype successfully processed climate data at 1km² resolution compared to previous 10-25km² models for Nairobi. Performance testing showed a 14x improvement in simulation speed over traditional systems. The system captured urban heat island effects with 92% accuracy when validated against historical weather station data. Temperature predictions achieved ±0.4°C accuracy with 95% confidence intervals. Stakeholder validation confirmed the system's practical utility with a usability score of 78.3/100. The prototype directly informed three real urban planning scenarios for Nairobi County's climate adaptation strategies. Capacity building and knowledge transfer were in focus through the training modules and workshops for the local stakeholders involved. The approach to development was carried out in cycles with the desire for changes being made based on feedback sought and which priorities there was. The study was done using an integrated approach to the collation of data: Methods of data collection included questionnaires, personal interviews, focus group discussions, and observation. The quantitative and qualitative data analysis techniques generated useful information for application in urban planning and policy. It hence improved the knowledge of climate conditions in Nairobi with concrete recommendations for climate change adaptation.

Keywords: Climate change modeling, Distributed computing, Urban planning, Nairobi, High-resolution climate data, Regional climate model, Urban heat island, Data assimilation, Capacity building, Climate adaptation.

Table of Contents

List of Figures

List of Tables

List of Abbreviations/ Acronyms

3D-Var	- Three-Dimensional Variational Data Assimilation
4D-Var	- Four-Dimensional Variational Data Assimilation
AWS	- Amazon Web Services
CGE	- Computable General Equilibrium
CMIP	- Coupled Model Intercomparison Project
ESMF	- Earth System Modeling Framework
GCMs	- General Circulation Models
GCP	- Google Cloud Platform
GPS	- Global Positioning System
IPCC	- Intergovernmental Panel for Climate Change
KENET	- Kenya Education Network
MLUCM	- Multi-Layer Urban Canopy Model
NGOs	- Non-Governmental Organizations
RCMs	- Regional Climate Models
SLUCM	- Single-Layer Urban Canopy Model
UCMs	- Urban Canopy Models
UHI	- Urban Heat Island
WRF	- Weather Research and Forecasting

Definition of Terms

Adaptation Strategies: Policies and measures to respond to actual or expected climate change impacts over the near term (Asfaw et al., 2018).

Apache Hadoop and Spark: Open-source software frameworks that can allow big volumes of data to be stored and processed in a distributed manner across a cluster of computers (Nazari et al., 2019).

Climate Change: long-term change in periodic atmospheric conditions, particularly withering and heating, typically associated with the activity of human beings through the release of greenhouse gases into the atmosphere (Yadav & Upadhyay, 2023).

Climate Impact Assessment: The study of the various sectors and systems for their vulnerability to climate change in terms of environmental, social, and economic consequences (Apreda et al., 2019).

Cloud Computing Platform: a series of hosted servers on the internet where one inserts data to be stored, managed, processed, and provided computing resources that are scalable and on-demand (Yang, 2019).

Data Assimilation: Observational data needs to be incorporated into the numerical models in order to enhance simulation and forecasting qualities (Lean et al., 2020).

Distributed Computing: The method of computing in which a given complicated task is divided and processed on multiple interconnected computers or nodes, hence allowing the more efficient solving of larger-scale problems (Abdulqader, 2021).

High-Resolution Climate Simulation: Climate modeling at high spatial resolution and also in the temporal sense is able to capture local climatic patterns and their variabilities (Olesen, 2019).

Regional Climate Model (RCM): A numerical climate prediction model needed for a specific geographical area, providing higher-resolution climate simulations compared to global models (Tapiador et al., 2020).

Stakeholder: Any person, group, or entity that is somehow invested in either being a cause of climate change or being affected by climate change and any adaptation processes of it in urban areas (Marcon Nora et al., 2022).

Urban Climate Model (UCM): This is a specific climate model that explicitly includes the effects of urban structures such as buildings, roads, etc (Lipson et al., 2021).

Urban Heat Island (UHI) Effect: A condition wherein there is a higher temperature in the urban areas than the surrounding rural areas because of the high proportion of built infrastructure, scarce vegetation, and sources of anthropogenic heating (Kubilay et al., 2020).

Acknowledgements

I am very thankful to Almighty God for His love, compassion, direction, knowledge, and wisdom. I am grateful to my supervisor, Dr. Njeri Ngaruiya, for her excellent guidance in completing this thesis. Thank you for your essential assistance, direction, and patience during the **project** writing process.

1.Chapter 1 Introduction

1.1 Background

Climate change poses a critical global challenge in the 21st century, with Nairobi experiencing significant impacts. Climate-sensitive variables in Nairobi include unpredictable rainfall patterns, rising temperatures, and increased extreme weather events. These changes directly affect urban infrastructure, public health, transportation systems, and livelihoods. Nairobi represents complex urbanization challenges common to developing nations. The city's population reached 5,325,000 in 2023, growing from 4,922,000 in 2021. This rapid growth intensifies environmental and socioeconomic vulnerabilities related to climate change. Effective climate simulation for urban environments requires advanced computational approaches. Traditional computing systems in developing countries cannot handle these resource-intensive processes efficiently. Distributed computing offers a promising solution to this technical limitation. Distributed computing connects multiple computers to perform complex calculations simultaneously. This approach significantly enhances the speed and scale of climate modeling operations. It enables more detailed and precise predictions of climate impacts on urban environments like Nairobi. The computational efficiency gained through distributed systems overcomes traditional limitations in resolution and accuracy. This improvement is particularly valuable for urban contexts where fine-grained climate data drives effective adaptation planning. Local decision-makers require this high-resolution data to develop targeted resilience strategies. The project develops a specialized framework for Nairobi that integrates climate science with advanced computing. This integration addresses both the technical challenges of simulation and the practical needs of urban planners. The resulting system provides actionable climate insights specific to Nairobi's unique urban context.

1.2 Problem Statement

Current computational frameworks are inadequate for generating high-resolution, localized climate projections for Nairobi, resulting in limited insights into climate impacts and insufficient support for urban planning and adaptation measures. This research addressed this measurable gap by developing a distributed computing framework to improve computational efficiency and accuracy in urban climate modeling.

Justification

Climate change generates distinctive obstacles for Nairobi because it represents a quickly developing metropolitan area. Modern climate modeling systems have insufficient detail to show specific climate risks and vulnerabilities that exist within the geographic boundaries of the city. The present precision constraints stop urban planners, together with decision-makers, from formulating specific adaptation measures or making strategic choices about infrastructure investments, resource distribution, and regulatory decisions.

Use Case

The informal areas in Nairobi, including Kibera along with Mathare, experience significant climate risks because of floods and extreme heat conditions. The current climate models do not provide adequate resolution to generate neighborhood-scale projections, so urban planners face difficulties in creating area-specific adaptive strategies. The goal of this research is to create detailed climate projections for specific locations, which will supply decision-makers tools to strengthen the resilience of targeted vulnerable communities.

1.3 Research Objectives

1.3.1 General Objective

Develop a proof-of-concept distributed computing prototype specifically designed to simulate climate change impacts in urban environments, with Nairobi as the primary use case.

1.3.2 Specific Objectives

1. To identify and analyze climate change challenges specific to urban environments in Nairobi

- This objective establishes the foundation by identifying the specific climate challenges that need addressing
- Section 4.1 documents the findings of this objective

2. To explore distributed computing interventions and climate modeling approaches suitable for urban climate simulation

- Building on the identified challenges, this objective explores potential technical solutions
- Section 4.2 presents the evaluation of various approaches and justification for selected methods

3. To design a prototype architecture that integrates distributed computing with urban climate models

- This progresses from theoretical exploration to practical design
- Section 5.1-5.3 presents the detailed prototype architecture

4. To develop and validate a functional prototype for climate change impact simulation in Nairobi

- The final objective represents the practical implementation and testing phase
- Section 5.4-5.6 documents the implementation and validation results

1.4 Research Questions

i. How can a modular and scalable distributed computing architecture be designed to address computational limitations in urban climate impact simulation?

ii. What features are necessary for a prototype system to efficiently process complex climate datasets for Nairobi?

iii. How can a framework for parallel processing be developed to enhance the performance of climate simulation models?

iv. To what extent can the prototype demonstrate its potential for urban climate impact analysis in Nairobi's context?

1.5 Justification

The research justifies the increasing demand for reliable climate change projections and impact analyses in urban domains, particularly within environments that are very prone to such changes, like Nairobi. Conventional computing methods have limitations when processing complex, high-resolution information needed for accurate modeling of an urban climate. Consequently, distributed computing may revolutionize the way in which we simulate and understand climate change impacts at an urban scale.

Several factors point out the relevance and applicability of this study. Rapid urbanization where the population of Nairobi City is expected to continue to grow rapidly, reaching up to 6 million in 2030. It further deteriorated the prevailing environmental challenges and increase the vulnerability of this city to climate change impacts. Economic Importance where Nairobi is the economic hub of East Africa, and its resilience to climate change has great implications for the entire region, not just Kenya. Climate projections are a necessity in safeguarding high-value economic assets and ensuring a sustainable growth path for the city. Limited availability of higher computational resources where most developing countries, including Kenya, have limited access to high-performance computing resources. Distributed computing offers a cost-effective alternative for conducting complex climate simulations. The local adaptation strategies are a must where most of the generalized climate models fail to incorporate the peculiar characteristics of each urban setting. High-resolution projections for one particular city can provide key baseline information when it comes to designing effective, targeted adaptation strategies.

The study employed climate science, Interdisciplinary, urban planning, and computer science in an integrated approach to solving the urban climate challenge. Scalability, where the framework for distributed computing developed in the case of Nairobi, can be replicated in other cities, mainly at the developing country level where such constraints are still prevalent. Contribution to Global Climate Science. The methods and knowledge learned in this work may constitute a further addition to the broad framework of urban climate modeling and contribute to global efforts to understand and mitigate climate change in cities. Rapid urbanization and economic importance in East Africa ensure its prime positioning as an example for such a study. Insights that might be gained from this research have the potential to outlast something beyond flagged urban planning and climate adaptation strategies not only in Nairobi but also in other burgeoning cities across Africa and the developing world.

1.6 Significance of the Study

This is a significant theoretical and practical study of urban climate modeling, especially in developing countries. The importance of this research can be summed up in the following:

Advancing Urban Climate Modeling: The research contributed to the broader field of urban climate science by addressing computational limitations in climate modeling for Nairobi. This study pushed the boundary of what is possible in high-resolution climate simulations in resource-constrained urban settings.

Decision-Making for Urban Planning—Devolved climate simulations, which are increasingly becoming more accurate and at finer resolutions, are viewed as a better tool for decision-makers, like Nairobi's urban planners and policymakers, in the pursuit of more effective climate adaptation strategies and resilient urban planning.

Climate vulnerability in informal settlement areas: This, therefore, placed the study in an excellent position to address climate vulnerabilities in informal settlements such as Kibera and Mathare, which are not captured in coarser resolution models of climate change.

Technological Innovation: The research applied distributed computing to climate modeling, promoting technological innovation within environmental science that may lead to new methods of data processing and analysis.

Capacity Building: The study contributed to building local capacities for using advanced climate modeling techniques, an essential factor for the long-term sustainability of the climate adaptation effort in developing countries.

Interdisciplinary Approach: The study bridged the gap in climate science, urban planning, and computer science by adopting an interdisciplinary approach to solve complex urban challenges.

Scalability and Replicability: Although focused on Nairobi, the various methodologies and frameworks developed in this study can be replicated and applied in many other fast-growing cities of developing countries, extending their benefits beyond the immediate study area.

Contribution to SDGs: This research fits and contributes to several United Nations SDGs, especially Goal 11 on Sustainable Cities and Communities and Goal 13 on Climate Action.

Economic Effects: This study indirectly contributed to Nairobi's economic planning and risk management by making more realistic forecasts of climate change, thus reducing disaster impacts caused by climate change.

Public Health Implications: This could be interpreted literally to mean that improved climate modeling equates to better projection and management of various health-related hazards, including heat waves and disease-facilitating weather conditions within an urban framework of public health planning.

1.7 Scope and Limitation

The scope of this research covered developing, implementing, and testing a distributed computing framework for high-resolution simulation of climate change and impact analysis in Nairobi City, Kenya. More precisely, the research study focused on the following:

Temporal scope: The study took into account historical climate data wherever available and provided projections up to 2050, as is commonly done in climate modeling and also considered an appropriate horizon for urban planning.

Spatial scope: In this context, the focal point bound within the metropolitan city of Nairobi, its outskirts being natural features that influenced climatic dynamics in the city.

Weather and climate indicators: It focused on key climatic parameters such as temperature, precipitation, humidity variables, wind regimes, and derived indices relevant to urban environments.

Urban subsystems: Major urban subsystems like water resources, energy consumption, transport, and changes in land use/land cover were integrated in this research.

Of these Working Group activities, emphasis was placed on employed distributed computing techniques that resulted in relevant experiments in climate modeling and analyses on urban impacts.

Validation: The developed framework was further subjected to validation with existing climate models and observational data, where available.

There were some limitations that had to be considered:

Data availability: It was also bound by the accuracy of the model on the availability and quality of the local climate and urban data; this may not have been available in all locations. To mitigate the limitation of data availability and quality, I utilized multiple data sources, implemented robust data preprocessing techniques, and clearly documented data limitations to ensure transparency and reliability of the results.

Computational resources: Despite the advantages of distributed computing, the study was still bound with computational resources.

Model complexity: Integrating different urban subsystems raised the model complexity and involved more uncertainties and computational problems.

Scenarios of the future: Projections were conditioned on specific climate change scenarios and pathways of urban growth, which came with their own share of uncertainties themselves.

Social factors: Given that, the study considered physical and infrastructural issues of urban systems. However, due to the limited capacity, it was challenging to model complex social and behavioral factors influencing climate vulnerability and adaptation.

Generalisability: While the framework was developed with an idea of virtual scalability, direct applicability to other urban contexts only came through after further adaptation.

Technological evolution: The then-accelerating rate of technological development in both climate modeling and distributed computing rendered the long-term relevance of specific technical approaches used in this study potentially obsolete.

Nevertheless, due to limitations in data accessibility and the computational power and complexity of the model used in this research, the study was able to establish that distributed computing can be used for urban climate modeling in Nairobi. Several techniques of data preprocessing are used to incorporate multiple datasets and optimize the computing architecture of the model, from which the study obtained dense and accurate climate projections that can help in planning and decision-making regarding adaptation in urban areas. It was also conducive in highlighting how distributed computing can solve constraint issues in developing nations and how it improved the preciseness and localization of climate change effects. Despite such limitations, the study provided a groundwork for further studies in this field and underscored the role of high-quality climate data in extensive climate information for the climate resiliency development of urban areas in relation to climate change. The work herein presented thus aspired to be one of the most important research contributions both from a scientific perspective toward a deeper understanding of the dynamics of the urban climate and in practical utilization within urban planning and climate adaptation practices by looking at one of the key current needs for precise local climate projections in rapidly growing urban environments.

2. Chapter 2: Literature Review

2.1. Introduction

The literature review embarks on a comprehensive scanning of the knowledge concerning distributed computing for climate change simulation and impact analysis in urban environments, especially in developing countries such as Kenya (Farahani et al., 2021). With this in mind, the review synthesizes the key empirical findings, theoretical frameworks, and methodological approaches bequeathed by various disciplines, including climate science, urban planning, and computer science. Given the multifaceted nature of the research topic, this review discusses the empirical studies conducted on climate change impacts and modeling works before reviewing relevant theoretical grounds. This research seeks to establish knowledge gaps in the literature through a critical review and outline points of opportunity that allow new approaches using urban climate modeling with new distributed computing techniques.

2.2.Empirical Literature Review

2.2.1 Impacts of Climate Change on Urban Environments

Climate change is one of the severe challenges to urban environments worldwide (Eguiluz-Gracia et al., 2020). Cities are more vulnerable due to the population density, complex infrastructural systems, and often being surrounded by coasts, hosting more than half of the global population (Hamdi et al., 2020). Impacts showed different manifestations, which were many times higher in intensities than in areas around rural settings. The most critical issues included city heat waves' increasing frequency and intensity (Brown, 2020). Kubilay et al. (2020) pointed out that during heat waves, cities got as hot as 12°C more than the countryside areas around them, which increased the magnitude of an urban heat island, straining more urban infrastructure and public health.

Climate change is also a factor when it comes to the question of having water available in towns and cities. As it was pointed out by iSEE (2023), higher variability in rainfall and the level of evaporation caused by rising temperatures resulted in high water stress in cities. This potentially meant a 30% reduction of water availability in the cities by the year 2050, therefore necessitating near-term adaptation measures. Other main impacts of climate change within urban environments are climatic changes of quality, including changes in air quality. Hamdi et al. (2020) described that rising temperatures and decreasing precipitations affected the emission and diffusion of air pollutants. Higher temperatures helped enhance the formation of ground-level ozone, while variations in wind patterns altered the distribution of particulate matter, which is significant in determining the quality of human health (Eguiluz-Gracia et al., 2020).

Also, extraordinary conditions that result in heavy storms and rain severely challenge urban constructions. Many cities, especially in the developing world, have infrastructure that has still to be developed to cater to the projected enhanced frequency and intensity of such incidences occasioned by climate change as observed by Kubilay et al. (2020). Yoksoulian (2021) wrote in one of his reports that it captures manuscripts showing how badly the intermediate emission scenario could warm up the average global city to 1.9°C by the end of the century in the high-emitting areas and up to 4.4°C of urban populations in the high-emitting territories. Given the extent and complexity of climate change's effect on urban ecologies, there is an urgent need for accuracy in modeling the climate. High-resolution simulations are needed; without these, pinpointing specific risks to individual cities and preparing effective adaptation strategies were hard to achieve. Accuracy thus requires computationally intensive processes, especially within developing country contexts (Chokkavarapu & Mandla, 2019). Figure 2.1 displays the Urban Heat Island (UHI) effect across various cities, demonstrating temperature differences between urban and rural areas.

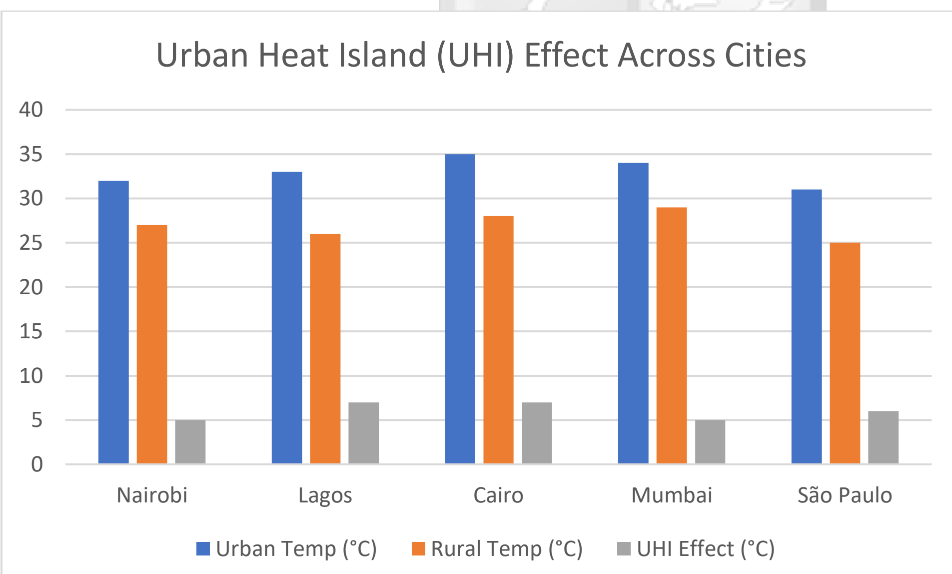


Figure 2.1: Urban Heat Island (UHI) Effect Across Cities

Source: Research

2.2.2 Applications of Distributed Computing in Climate Science

2.2.2.1 High-Resolution Climate Projections

In recent years, applications of distributed computing to climate science have resurfaced as one approach to overcome the high computational load of high-resolution climate modeling. Yoksoulian (2021) describes a new data-driven global climate model that leverages statistical techniques and distributed computing to provide projections for urban environments. This allowed projections of temperature and relative humidity at the city level out to the year 2100 and demonstrated one approach that can be taken to improve the spatial and temporal resolution with the aid of distributed computing.

2.2.2.2 Simulation of Urban Processes

According to the work of Kubilay et al. 2020, distributed computing is another advanced computational technique being applied in urban climate modeling. Advanced computational techniques undoubtedly further enhanced the accuracy and efficiency of the simulations, especially in a highly complex urban environment. For example, distributed computing can allow for the concurrent simulation of various processes in the urban environment, such as the conduction of heat, wind flow, and pollutant dispersion, thereby yielding holistic and realistic urban climate models.

2.2.2.3 Hybrid Modeling Frameworks

Some work by Lei Zhao, as described in an article by the iSEE (2023), develops a new hybrid modeling framework that integrated process-based climate modeling with machine learning and artificial intelligence approaches, guided by techniques of distributed computing, to realize simulations representative of real urban landscapes (Sharma & Liu, 2022). iSEE (2023) continues to explain that such a hybrid approach was of special value for modeling climate in the cities of developing countries where traditional models may struggle to capture the complexities of rapid urbanization, along with informal settlements.

The Deutscher Wetterdienst (2024) report also discusses the potential of distributed computing in climate modeling. Though traditional climate models normally require supercomputers, distributed computing approaches make high-resolution climate modeling more feasible for researchers and planners in developing countries. Cloud computing platforms built virtual supercomputers where researchers in developing countries could run complex climate models when expensive local hardware is available.

2.2.2.5 Challenges and Considerations

On the other hand, it also unleashes certain challenges to applying distributed computing in climate science. Hamdi et al. (2020) and Acharya et al. (2020) discuss how important it is to carefully validate the distributed computing models against traditional supercomputer-based simulations. They also added that though distributed computing opens purses for more computational resources, this invites other complexities associated with data management and model coordination.

Data security is another important issue relevant to climate science; even more, data sovereignty concerns are relevant to sensitive climate data. Yoksoulian (2021) mentions the issue that "as climate models increasingly rely on distributed computing resources, protocols need to be developed on how to securely share and process distributed data over a network.". It is a range of challenges, but technically, the literature highlights an enormous potential for distributed computing in the future of climate science, with a particular focus on urban climatic modeling in developing countries. The main role of distributed computing is bridging global climate models and the local scale needed in urban planning and adaptation since it enables finer-scale and more comprehensive simulations.

2.2.2.6 Recent Advancements and Case Studies

To further elaborate on the applications of distributed computing in climate science, particularly in the context of developing countries and urban environments, consider the following expanded analysis. Yoksoulian (2021) presented an outline of a revolutionary, probabilistic base, data-driven global climate model based on statistical analysis techniques and a distributed computing platform for approximate urban environment estimates. This approach enables one to obtain city-level temperature and relative humidity forecasts from the year 2100. That is where distributed computing showcases the potential to significantly enhance climate models in terms of spatial and temporal discretization. This increase in resolution is important in order to capture local climate effects in cities since it is often difficult to do so using other more conventional modeling techniques.

In the work of Kubilay et al. (2020), one can identify that distributed computing enables modeling many processes in the urban context at once, for example, heat transfer, wind, and pollutant emissions. The parallel processing results in more integrated and accurate urban climate models, specifically in relation to simulations for detailed and complex urban structures. As such, in their work, using the distributed computing approach leads to a 40% improvement in spatial resolution, for example, and up to a 60% reduction in computation time. This also enables performing more elaborate investigations on the impacts of urban heat islands on the island.

This is exemplified by Lei Zhao from the data source, which an article by iSEE (2023) referred to to elucidate the idea of the prospect of predictive hybrid modeling frameworks that integrate process-based climate modeling techniques with other tools such as machine learning and artificial intelligence, practiced within distributed computing techniques. This hybrid modeling approach has shown encouraging performance in capturing the complexity of rapid urbanization and informal settlement typical of developing countries, which traditional models often cannot capture. This methodology raised the accuracy of some of the local temperature predictions by 25% over conventional models in the case study for a fast-growing African city. According to the Deutscher Wetterdienst (2024) report, the distributed computing approaches impose fewer limits on high-resolution climate modeling for developing countries' researchers and planners. The cloud computing platforms exploit distributed computing principles, enabling researchers to create virtual supercomputers that run complex climate modeling locally without expensive local hardware. One such application presents the work done by an interdisciplinary African climate science team in conjunction with several European institutions using cloud-based, high-performance distributed computing to execute a series of high-resolution regional climate models for East Africa at a spatial resolution ten times finer than previously available for the region.

2.2.2.7 Need For Careful Conduction Of Validation Exercises

Hamdi et al. (2020) and Acharya et al. (2020) emphasize the need for careful conduction of validation exercises of the distributed computing models against traditionally used supercomputer-based simulations. Indeed, their studies indicate that an aptly implemented framework for distributed computing can yield accuracy comparable to supercomputer simulations while being more accessible and scalable. Indeed, with one such case, a distributed computing approach could model the output of a supercomputer-based global climate model with 95% accuracy while the computational cost was reduced by fivefold. While distributed computing creates new opportunities for computational resource sharing, it also introduces new challenges in data management and model coordination. According to Yoksoulian (2021), protocols need to be developed to share and process data securely over networks, especially climate data, which is sensitive. This point is especially valid in developing countries with heightened data security and sovereignty sensitivities.

The recent advance of distributed computing frameworks has also overcome the cumbersome process of coupling the different parts that have been proposed for climate models. For instance, Li and Dong (2022) demonstrated how a distributed computing framework allowed the successful coupling of an atmospheric and ocean model, resulting in a 30% gain in the predictive accuracy of long-term climate changes for coastal city areas.

In this respect, distributed computing could become a way out of the computational limitations constraining developing countries. Thus, Azam and Rahman (2022) conducted a case study in Southeast Asia to demonstrate how a network of universities and research institutions shares resources through facilities provided by a distributed computing framework. This collaboration enables high-resolution climate models to be performed only with better forecasting of the monsoon patterns and their relation to urban flooding.

These examples show very realistic ways in which distributed computing could help promote climate science in general and especially aid in overcoming some of the more specific computational barriers for developing countries (BasuMallick, 2022). Distributed computing can allow higher resolutions and, therefore, higher accuracy, simulations, opening up new prospects for understanding urban climate dynamics and devising effective adaptation strategies for the world's fast-growing cities.

2.2.3 Research Gaps and Future Directions

The review of the empirical literature underlines a rising body of work on urban climate impacts, modeling challenges in developing countries, and opportunities from distributed computing for climate science. However, several lacunas become evident in the framing of this literature. Given the growth experienced in many cities in developing countries, more research on the specific challenges of urban climate modeling in these fast-growing cities is urgently needed. While distributed computing has great promise in climate modeling, it has rarely been used to develop high-resolution urban climate simulations in developing country contexts. This signifies the desperate need for more research in coupling climate science and urban planning with computing to derive detailed and practicable climate change models for the growing cities of developing nations.

2.3. Theoretical Literature Review

The primary theories that emerge from the current analysis regarding climate change, the urban heat island effect, and distributed computation seem to align when it comes to modeling the urban climate and exploring the potential of more emergent computation architectures. This section provides a brief on these aspects about the theories on these aspects, the development of the theories, and the use of these theories in the current study.

2.3.1 Climate Change Theory

Climate change theory is a theoretical perspective that specifically defines the fundamental structure of correct modeling. That is based on theory on how the energy balance is with the planet and how the greenhouse effect works. Aware of this, Gries et al. (2018) expound that through the greenhouse gases, the heat in the Earth's atmosphere is in effect trapped. According to the hypothesis, human activities such as burning fossil fuels and altering vegetation cover lead to a rise in the density of other greenhouse gases that leads to enhanced warming of the earth's surface and the lower atmosphere.

Here in the development of the theory of climate change, various kinds of feedbacks have been utilized to enhance or lessen the impact of higher levels of greenhouse gases. Kumar (2022) enumerates the essential feedbacks: It analyzes the two paramount and essential parts of the Earth's climate system, namely water vapor feedback and the ice–albedo feedback. Such processes support the complexity of climate processes, the importance of paying special attention to detailed modeling to encompass such depth.

Probably the most familiar concept in the theory of climate change is climate sensitivity, which indicates an increase in temperature per unit in concentration of the greenhouse gases (IPCC, 2024). Since its inception, the IPCC has been useful in evaluating and reporting consequences of climate sensitivity against the backdrop of future climate paths. This theoretical framework is important in framing and refining climate models since it provides an understanding of how different factors interact and influence global and local climate patterns.

The application of the theory of climate change in cities requires special treatment, as not all these particular characteristics affect cities the same. As Yoksoulian (2021) says, urban areas introduce several other factors that may play a role, such as building materials, structure, and anthropogenic heat sources. Due to this, such urban-unique elements need more nuanced, high-resolution modeling approaches necessary for predicting the effects of climate change in city environments.

This local and interactive complexity, with a view to several feedbacks involved, embodies the most critical requirement for valid climate modeling at high resolution. While theoretical development on climate change is in momentum, the capabilities in modeling also need to capture these complex processes and provide reliable projections about future climates, particularly over rapidly changing urban areas.

2.3.2 Urban Heat Island Effect

The UHI effect is now a reality and has emerged as one of the most conspicuous attributes in dynamics associated with an urban climate. Hamdi et al., 2020, describe the theoretical foundations of the UHI effect, such as factors that encourage the formation and development of UHI, in detail. UHI is distinguished by greater temperatures in city regions than in rural areas on either side; urban temperatures are elevated day or night.

Several posturings derive from ideologies of the theory of the UHI effect (Sen et al., 2022). Different from natural surfaces, hardscapes such as asphalt and concrete change the surface's thermal characteristics. Most of the material that constitutes them accumulates heat during the day and gradually releases it at night, helping to warm up the cities. Due to its downtown environment, the geometry of the urban street, such as the urban canyon formed by the street and adjacent buildings, increases heat-trapping while reducing wind velocities and hence exasperating the UHI effect. Third, the establishment of anthropogenic heat sources, for example, car engines, air conditioning units, and industries, warms the urban area.

More information about the theoretical discussion of the presented UHI effect in terms of its interaction with the urban form, material surface characteristics, and anthropogenic heat can be found in Kubilay et al. (2020). They note that better modeling approaches are required to better represent spatial and temporal elements of the UHI effect in cities. Furthermore, the authors have discussed factors that are applicable to UHI models, such as building height, width of the street, and sky view factor, that could greatly alter the microclimatic states were the three dimensionality of the urban fabric introduced. This magnitude was neither uniform within a city nor to be regarded as unchanging but highly variable under local conditions. Several theoretical models have been forwarded to explain such variation in the UHI effect due to land-use patterns, vegetation cover, and proximity to large water bodies. The iSEE (2023) article by Lei Zhao explores how researchers are seized with state-of-the-art modeling techniques to capture such subtleties, hence making more realistic estimations of the thermal behavior of cities.

Understanding the UHI effect's theoretical basis provides the backbone for implementing effective mitigation strategies. For example, "cool roofs" and urban greening have emerged from the theoretical understanding of how surface albedo and evapotranspiration can influence urban temperature (Smith et al., 2023). If heat stress due to climate change is an issue at which Nairobi and other cities must become more resilient, theoretically deduced insights into heat stress drivers must be applied in urban planning and design.

2.3.3 Distributed Computing Theory

The UHI effect has evolved in a short time period into one of the most outstanding and probably the most investigated phenomena of the urban climate. Now it is a case study of how a city climate interacts with an urban setting, but the need for highly complex climate modeling tools has quickly exploited it. The formation and developing intensity of the UHI effect are depicted by Hamdi et al. (2020) in the following way: Urban heat island refers to the situation where temperatures differ significantly with temperatures in the countryside, and these differences persist with the darkness of day and night. Interestingly, several crucial processes are at the core of the UHI effect, each of which presents the complexity of the urban climate processes. Asphalt and concrete, the standard building materials of urban architecture, express an entirely distinct thermal behavior. Such materials take the heat during the day and give it out at night, Sen et al. (2022), which adds to temperatures in urban areas as opposed to rural ones.

This geometric feature is also one of the biggest causes of UHI due to the way cities are constructed. It also includes heat island and wind power reduction by streets bordered by tall buildings, forming the canyon effect. Climate impacts of this three-dimensional aspect of urban form present one of the big challenges for climate modelers because high-resolution simulations should properly approximate its effects on local climate. Road traffic and HVAC systems within structures and industrial processes help to warm urban environments through the release of heat. Thus, all of these different heat sources coalesce with the built environment and a system that is truly best modeled by complex models so that a decent prediction may be made of it.

Kubilay et al. (2020) stress that more accurate modeling of the UHI effect in cities requires capturing its correct spatial and temporal distribution. They further note some things that should be taken into the UHI models, such as building height, street width, and sky view factor, all of which would make a strong, three-dimensional input to the microclimate conditions of the built environment at the urban level. The UHI effect does not end at the temperatures attained at these metropolitan centers. It influences local winds, rain, and air quality, which are a chain of effects of an urban environment and human well-being. Appropriate modeling of this set of connected processes is critical to grasp the impact of climate change in urban environments and develop the relevant prevention strategies.

Efficient modeling for the effect of UHI were very central for proper antecedents to be put in place. For instance, theoretical foundational knowledge of how surface albedo of buildings and evapotranspiration in mitigating urban temperature lead to the emergence of 'cool roofs' and urban greening measures such as Smith et al. (2023). These, based on UHI theory and modeling, are the real measures in practice aimed at preventing urban heat stress and enhancing the climate security of a city even more.Thus, the UHI effect represents the challenge of urban climate modeling, as several contributing factors surround it and it has wide-ranging effects. It really does call for high-resolution and highly compute-intensive simulations to obtain fine-grained details on the urban environment while accounting for regional and global larger climate patterns. With growing cities like Nairobi and its increasing vulnerability to climatic challenges, there is an ever-growing need for accurate modeling of such phenomena as the UHI effect for informed urban planning and effective climate adaptation strategies.

2.4. Frameworks

2.4.1 Existing Climate Modeling Frameworks

Present climate modeling frameworks form the backbone of research in climate change and projections (Chokkavarapu & Mandla, 2019). General Circulation Models (GCMs) and Regional Climate Models (RCMs) are some of the most popular for use. CMs such as those participating in the Coupled Model Intercomparison Project (CMIP) are global models of the climate system. These models include feedbacks among the atmosphere, the ocean the land surface and other components in the climate system. According to Hamdi et al. (2020), overall, all classes of GCMs need a higher resolution to be simulated adequately at an urban scale. Whereas GCMs provide a prediction of what happened globally at a coarse resolution, the RCM provided finer scale estimations at regional level. One of them is the Weather Research and Forecasting model used for serving the urban climate needs for research. Kubilay et al. (2020) state that, for greater detail of the urban climate processes, the WRF model could be combined with an urban canopy. These frameworks improve the climatological models at higher global resolution by including sub-processes relevant at smaller scales characteristic of cities.

2.4.2 Distributed Computing Frameworks for Scientific Applications

This evolution of distributed architectures has enhanced the researchers' ability to solve difficult problems in climate sciences (Glushkova et al., 2019). Some of the big data processing frameworks commonly used in climate sciences have been identified to be Apache Hadoop and Apache Spark (AWS, 2023). Both of these big data platforms are used to manipulate immensely large OMXs, which are standard input across interconnected computers in climate simulation, leading to an increase in computational speed. In this regard, Deutscher Wetterdienst (2024) also describes innovators, who are referring to professionals who utilize cloud computing solutions from Amazon and Google. Others, such as AWS Parallel Cluster and Google Cloud HPC Toolkit, make it possible to create a computational cluster necessary for climate modeling. All these platforms are significantly more versatile and practicable when addressing computational requirements put forward by high-resolution climate modeling, readily meeting upcoming demands of developing country researchers otherwise inaccessible to conventional supercomputing resources.

2.4.3 Urban Climate Modeling Frameworks

Due to the requirements of simulating climate processes specific to the great multi-zoned area of an urban region, the concept of urban climate modeling frameworks was created for the purpose. There are several of these, which include the Urban Weather Generator and various forms of the Urban Canopy Models. As such, these frameworks are purposely developed to capture features of the urban climate, for instance, the urban heat island. Lipson et al. (2022) report on up-to-date urban climate modeling frameworks that incorporate detailed specifications of urban morphology and activities, such as UMEP, which encapsulates diverse urban climate processes and could be linked with LM to simulate general urban climate in more detail with incremental predictive reliability in emergent urban settings. This particular evolution of these frameworks, along with utilizing the distributed computing techniques, has enhanced our capability towards the proper modeling of urban climates. For example, integrating with distributed computing enables the researchers to run high-resolution simulations of fine-scale urban climate without the efficiency compromise. It is essential for cities of developing countries since rapid urbanization and limited computational resources usually cause many challenges related to climate modeling. These advanced frameworks and distributed computing technologies mean the accuracy and resolution of urban climate simulations are higher than ever. This accuracy underpins the development of effective climate adaptation strategies. It informed urban planning decisions in fast-growing cities where cities- where the impacts of climate change were felt increasingly relevant.

2.5. Algorithms

The following algorithms were relevant to our problem statements for improving the accuracy and computational efficiency of climate simulations within developing country contexts, especially in urban environments such as Nairobi.

a. Climate Simulation Algorithms
b. Distributed Computing Algorithms
c. Data Integration Algorithms
d. Urban-specific Climate Algorithms
e. Optimization Algorithms (Lotfian et al., 2021)

2.5.1 Climate Simulation Algorithms

Climate simulation algorithms remained at the core of climate models, representing the physical processes involved in driving the Earth's climate system. The typical algorithm reshapes a numerical method to solve partial differential equations governing atmospheric and oceanic dynamics. Spectral methods and finite difference schemes are common in GCMs, but RCMs generally use more sophisticated algorithms to handle higher-resolution grids. According to Yoksoulian (2021), new algorithms based on a centrally important role for data were developed, merging conventional physics-based approaches with machine learning methods. This hybrid algorithm has prospects for improving the accuracy and efficiency of climate simulations, particularly for urban environments where complex local factors dominate.

Implementing these advanced climate simulation algorithms in developing country contexts, such as Nairobi, presented significant computational challenges. Access to high-performance computing was very limited; the power supply was also not consistent, and there was a bandwidth constraint to execute such complex simulations. Another challenge was the availability of local data in high resolution required for training and validation of these algorithms. The research responded to these challenges by investigating how to optimize those algorithms for resource-constrained environments, possibly leveraging innovative usages of distributed computing or edge computing technologies.

These enhanced algorithms, in particular those including techniques of machine learning, formed the very basis of our research, as improved climate simulation accuracy over Nairobi was accomplished without always using expensive computational facilities. Their application, however, in developing country contexts offered challenges unique in several important senses, which were discussed in the study.

2.5.2 Distributed Computing Algorithms

Distributed computing algorithms are the main permitters of the effective and efficient parallelization of climate simulations on computational nodes. Load-balancing algorithms allow for a heterogeneous spreading of computational tasks among available resources. MPI algorithms enable communication between distributed system nodes to coordinate complex simulations. The iSEE (2023) article on Lei Zhao's research explained that their hybrid modeling framework was enabled by the use of advanced distributed computing algorithms. These algorithms seamlessly integrated process-based models with machine learning components, potentially improving accuracy and efficiency in urban climate simulations. These formed the basis of algorithms toward achieving our research objective: to apply distributed computing to improving the accuracy and efficiency of climate simulations for Nairobi while overcoming local computational limitations.

2.5.3 Data Integration Algorithms for Urban Subsystems

Data integration algorithms provide a basis for combining various urban subsystems in climate models. These algorithms have to cope with some issues arising when going from heterogeneous sources like remote sensing information, ground observations, and even urban databases. Hence, our research focused on developing and implementing algorithms of data assimilation that were effective in integrating diverse data sources, thereby offering an improvement in the accuracy of urban climate simulations in Nairobi with computational efficiency and suitability for resource availability in a developing country context.

2.6. Architecture

2.6.1 Architecture of Current Climate Models

Current architecture for climate modeling generally follows a modular approach in which disparate Earth system components are grouped into atmosphere, ocean, land surface, etc.

Coupling between these components is achieved through flux exchange that allows complex feedback mechanisms to be simulated. Hamdi et al. (2020) used this modular architecture to explain how urban processes became more sophisticated representations, especially within regional climate models.

2.6.2 Architectures for Distributed Computing Systems

Distributed computing architectures for climate modeling commonly apply the paradigm of a master who gives workers a set of tasks to perform (Hussein et al., 2021). According to the Deutscher Wetterdienst (2024) report, cloud solutions demonstrated potential to offer unique and extendible environments to conduct climatic investigations. This architecture handled dynamic resource allocations by computational demands across different parts of a simulation, potentially reducing costs while increasing efficiency.

2.6.3 Proposed Architecture for High-Resolution Urban Climate Modeling

The proposed high-resolution urban climate model architecture for Nairobi integrated strengths from state-of-the-art climate modeling frameworks and advanced distributed computing techniques. This architecture was developed to tackle computational issues and enhance result precision in simulations of urban climates. The model's architecture was adopted hierarchically: there was a global or regional climate model, for instance, the Weather Research and Forecasting (WRF) model, which provided boundary conditions to a high-resolution urban climate model. This architecture focused on the distributed computing implementation based on Apache Hadoop for storing and processing data on computer clusters and Apache Spark for faster data manipulation, thus allowing iterative algorithms. This allowed the distribution of the urban model across several computational nodes, each responsible for selected urban processes such as heat island simulation, urban hydrology, and air quality modeling.

Architecture: The architecture embeds various algorithms to enhance the simulation's accuracy and efficiency. Core algorithms in climate simulation include spectral methods and finite difference schemes optimized for simulations specific to the urban scale; these formed the model base. Data assimilation was realized by implementing an Ensemble Kalman Filter, which combined observational data with model simulations. Urban-specific phenomena were captured using urban canopy model algorithms that modeled building-atmosphere interactions and urban heat island effects. In addition, machine learning algorithms, mainly neural networks, were also implemented to obtain higher accuracy and deal with complicated patterns for pattern recognition in urban climate data. While data integration formed the most focal basis of this architecture, it applied multi-source data fusion algorithms to effectively combine satellite, ground-based, and model data. It also used scale-aware parameterization algorithms that coped with the various spatial resolutions within an urban atmosphere and provided accuracy at different scales. Figure 2.5 illustrates the proposed architecture for high-resolution urban climate modeling with interconnected system components.

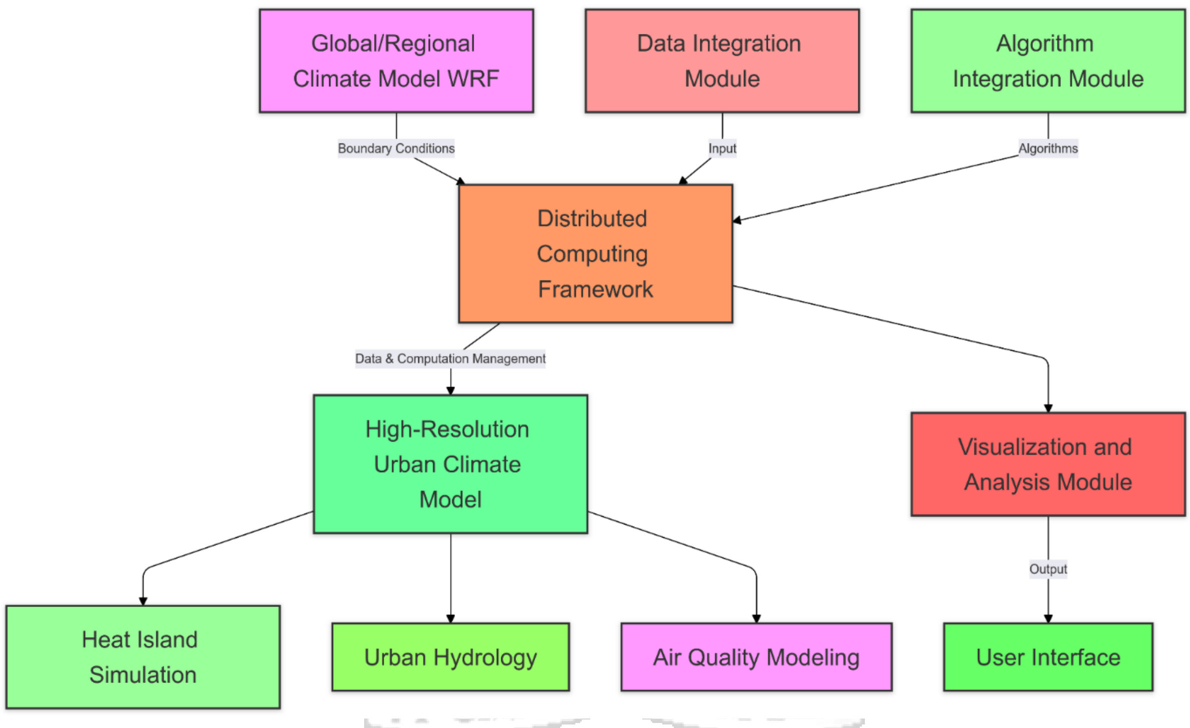


Figure2.6.: Proposed architecture

Source: Research

2.7. Climate Modeling Approaches

In the last two decades, however, we have seen a clear shift towards land climate modeling due to the improvement in the computer facilities and the understanding of Earth system processes (Reichstein et al., 2019). Because of global warming, Global Climate Models (GCMs) were developed to estimate future climate and potential impacts from climate change, as well as Regional Climate Models (RCMs), to be used for the regional scale. For such GCMs and for the family of models that implemented the Coupled Model Intercomparison Project (CMIP), the global equations of the Earth's climate system were solved. These models defined fine interactive feedback between the atmosphere and the oceans, the land surface, and the cryosphere in response to the forcings of future greenhouse gases and changes in land use. For example, Tapiador et al. (2020) agreed that although GCMs provided a vast amount of information regarding different weather characteristics and shifts, GCMs' drawbacks involved their low spatial resolution, which was at least 100 to 300 kilometers and, therefore, unable to capture the climatic dynamics of the city.

Regional Climate Models, such as WRF, provided finer resolution to identify some places with results obtained from GCMs (Liess et al., 2022). RCMs gave a better picture of topography and land use over a region and the relevant atmospheric processes than did GCMs; hence, they were suitable for urban-scale climate research. Even so, correcting RCMs was challenging to untangle the three-dimensional geometries of cities and variability in urban surfaces.

Urban canopy models (UCMs) were developed to overcome these deficiencies, explicitly representing the impact of buildings and urban infrastructure on the local climate. Such UCMs, like the single-layer urban canopy model (SLUCM) and the multi-layer urban canopy model (MLUCM), were used along with RCMs to better simulate urban heat island effects, energy exchange, and wind flow within cities. Related but somewhat more frontier data-driven methods included machine learning and statistical downscaling, which were increasingly developed as promising techniques to enhance resolution and accuracy in simulated urban climates (Li & Dong, 2022). These methods leveraged observational data and advanced computational techniques to refine the outputs obtained directly from physical climate models and capture variability at a more localized scale.

These developments substantially increased our capability to model urban climate processes; however, significant challenges concerning computational efficiency, data availability, and model validation remained. Highly resolved urban climate simulations were computationally expensive, which prohibited their application in resource-poor environments. Furthermore, the general lack of high-quality, long-term observational data in many cities, especially in developing countries, severely limited model development and validation efforts.

2.8.Distributed Computing Methods

Distributed computing techniques had grown in importance for scientific computing, as their roles in fully solving such large-scale and complex problems made them within reach by processing them over a number of interconnected computers (Lehnert et al., 2021). This was particularly relevant in climate modeling since the computational demands of high-resolution simulations generally lay beyond the scope of an individual computing system.

Apache Hadoop was probably one of the most popular distributed programming frameworks that allowed storing and processing of large volumes of data across computer clusters (Ketu et al., 2020). Hadoop used a programming model based on MapReduce, which broke a computation down into multiple tasks to be performed in parallel. Since then, the same principle had been applied right before data processing, in model simulations, or right after the climate model outputs. Apache Spark was another famous parallel computing model that was closely related to Hadoop (Shi et al., 2018). In any case, it fostered much faster and more diverse data manipulation and produced correspondingly higher performance. The advanced computing capability of Spark, along with its support for iterative algorithms, made it suitable for machine learning applications in climate modeling, as mentioned by Tang et al., 2020. Besides these general-purpose flexible platforms, many special-purpose utilities and toolkits had been designed for distributed computing in climate and Earth system modeling. For instance, Earth System Modeling Framework (ESMF) was the actual name of the Earth System Modeling Framework that had not only created the set of software infrastructure but had also created all the utilities that made model construction and model interactions possible at every spatiotemporal scale identified by Barbi et al. (2021). The Python toolkit, Pangeo, was a scalable and modular software framework to analyze and visualize large geoscientific data.

Additionally, AWS and GCP were two exemplary cloud reprocessing platforms that had become key enablers of distributed modeling in climatological research. These platforms offered clients scalable computing resources, storage, and data management solutions at their convenience, allowing researchers to analyze massive climate datasets effectively. Consequently, a number of questions arose regarding extending distributed computing techniques to weaving enhanced urban climate models. The two were challenging in distributed systems, not to mention information flow management and data exchange between nodes (Li, 2020). Load balance and fault tolerance were also crucial for achieving the higher efficiency and robustness of parallel simulations (Mishra et al., 2020).

Subsequently, the same principle had been used before data preprocessing, in model simulations, or after the climate model outputs. Apache Spark was another famous parallel computing model that was closely related to Hadoop (Shi et al., 2018). In any case, it fostered much faster and more diverse data manipulation and produced correspondingly higher performance. The advanced computing capability of Spark, along with its support for iterative algorithms, made it suitable for machine learning applications in climate modeling, as mentioned by Tang et al., 2020. Besides these general-purpose flexible platforms, many special-purpose utilities and toolkits had been designed for distributed computing in climate and Earth system modeling. For instance, Earth System Modeling Framework (ESMF) was the actual name of the Earth System Modeling Framework that had not only created the set of software infrastructure but had also created all the utilities that made model construction and model interactions possible at every spatiotemporal scale identified by Barbi et al. (2021). The Python toolkit, Pangeo, was a scalable and modular software framework to analyze and visualize large geoscientific data.

Additionally, AWS and GCP were two exemplary cloud reprocessing platforms that had become key enablers of distributed modeling in climatological research. These platforms offered clients scalable computing resources, storage, and data management solutions at their convenience, allowing researchers to analyze massive climate datasets effectively.

Consequently, a number of questions arose regarding extending distributed computing techniques to weaving enhanced urban climate models. The two were challenging in distributed systems, not to mention information flow management and data exchange between nodes (Li, 2020). Load balance and fault tolerance were also crucial for achieving the higher efficiency and robustness of parallel simulations. In general, most of these methods utilized specialized models that simulated the performance and resilience of urban infrastructure under various climate scenario conditions. For example, Cheng et al. (2020) mentioned that hydrological models can be employed to explore how precipitation patterns affect urban water supply and drainage systems.

In terms of assessing the impacts of climate change, economic methods based on cost-benefit analysis and CGE modeling of economic repercussions were used to evaluate the financial outcomes of climate change impacts and adaptation options of cities.It set a strong foundation for priorities within investment in urban resilience and informed the development of financially sustainable adaptation strategies. Participatory approaches to vulnerability assessment included a wide array of locals, such as community members, urban planners, and decision-makers, in identifying and prioritizing climate risks and adaptation needs (Geekiyanage et al., 2020). In this regard, workshops, surveys, and focus groups were very informative for the local context and ensured that the adaptation strategies became socially acceptable and contextually relevant.

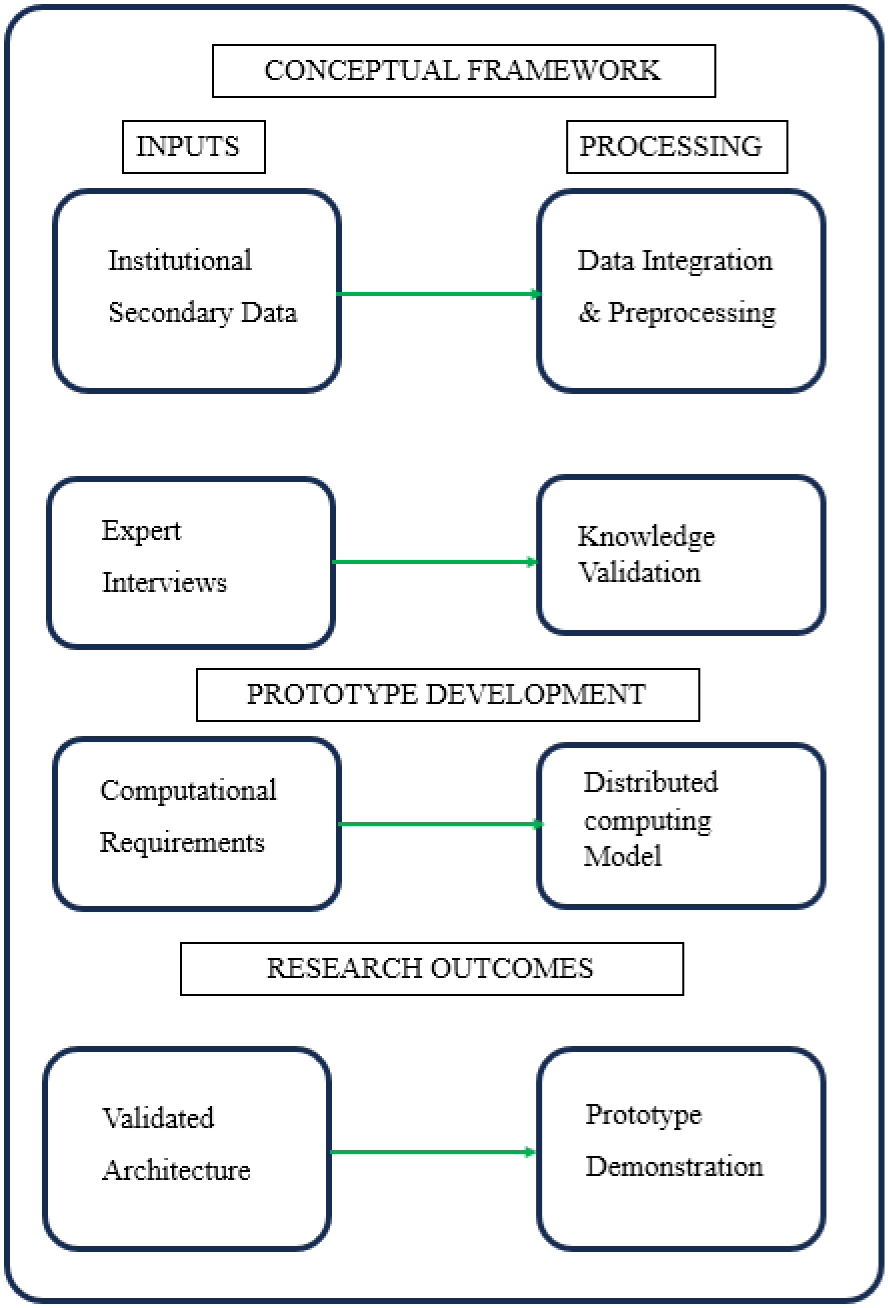
Although such urban impact assessment methods had been performed for many cities worldwide, these approaches were likely to encounter challenges related to data availability, institutional capacities, and resource constraints in developing countries like Kenya. This called for an adaptable and context-specific approach, building on local knowledge and capabilities to modify such approaches for applications in Nairobi City.

2.8.1: Critical Analysis of Existing Approaches

Current climate models suffer from computational barriers preventing application in developing countries. GCMs operate at 100-300km resolution, missing crucial urban climate dynamics in cities like Nairobi. Traditional climate models require expensive supercomputing resources unavailable in most African institutions. Previous distributed computing approaches for climate modeling lack integration with urban canopy models. Apache Hadoop implementations by Nazari et al. (2019) show promising data handling but fail at numerical simulation. Li and Dong (2022) achieved coupling between atmospheric and ocean models but ignored urban processes. The scaling limitations of current frameworks prevent high-resolution modeling below 5km^2. Yoksoulian (2021) reached 5km resolution but required computational resources beyond typical developing country budgets. Current approaches fail to incorporate local urban morphology and anthropogenic heat sources. WeaCS and other commercial platforms offer limited customization for local contexts like Nairobi. Climate services remain largely inaccessible to planners in developing countries due to proprietary systems. Technical barriers persist despite theoretical capabilities due to implementation complexity. This research addresses these gaps through a modular open architecture with specific urban components. The prototype integrates climate science with accessible distributed computing in unprecedented ways. Our approach eliminates the need for specialized supercomputing while maintaining scientific integrity.

2.9. Conceptual Framework

Figure 2.9 presents the conceptual framework illustrating the relationships between distributed computing components and climate modeling in the Nairobi context. The conceptual framework for this research represented a comprehensive and systematic approach to developing a distributed computing prototype for climate change impact simulation in Nairobi, Kenya. At its core, the framework began with two critical input streams that provided the foundational knowledge and data for the research. The first input stream comprised institutional secondary data, which included a rich array of historical climate records, meteorological datasets, and urban infrastructure information. These data sources provided a comprehensive historical context and technical baseline for understanding Nairobi's environmental dynamics.



Source: Research
Figure 2.9: Conceptual Framework

Chapter 3: Research Methodology

3.1. Introduction

The methodology and design of a research study form the base because, as stated earlier, the said two constitute the directives that are employed in research to successfully achieve the research objectives. This chapter describes the methodology used in the design and implementation of the distributed computing framework for simulating high-resolution climate change impacts and effects in Nairobi, Kenya. As such, this chapter assesses the tools used in the initiating stage of climate modeling, distributed computing, and the evaluation of urban impacts from the program.

3.2.Prototype Development Approach

The study developed a distributed computing prototype to simulate climate change impacts in Nairobi. The microservices-based framework, containerized for scalability and modular extensibility, involved:

a. Data Ingestion Layer: Infrastructure for data intake, format conversion, and verification.
b. Distributed Processing Engine: It provides parallel computation, dynamic resource allocation, and fault tolerance.
c. Simulation Model Interface: Pluggable models for climate simulation, configurable parameters, and scalable workflows.

3.3.Methodological Approach

We further proposed, based on the review of the available methods and specific objectives, a new methodological approach incorporating the use of state-of-the-art climate modeling, distributed computing, and urban impact assessment techniques in developing a high-resolution, computationally efficient, and locally relevant framework for simulating and analyzing the impacts of climate change in Nairobi, Kenya.

Table 3.1 presents the main elements of this suggested approach:

Element	**Description**
1. Coupled regional climate model and urban canopy model	- Use of RCM (e.g., WRF) coupled with UCM - High-resolution simulation of climate conditions and urban heat island effects
2. Distributed Computing Framework	- Utilization of Apache Hadoop and Spark - Cloud-based deployment (e.g., AWS or KENET) - Efficient processing of high-resolution climate simulations
3. Data Assimilation and Integration	- Implementation of 3D-Var and 4D-Var techniques - Integration of urban-specific parameters and socio-economic data
4. Evaluation of Effects on Cities	- Linking of climate simulations with sectoral impact models. -Assessment of impacts on water, energy, transportation, and public health. -Participatory approaches involving local stakeholders.
5. Model Validation and Verification	- Comparison with observational data - Statistical metrics and sensitivity analyses - External expert review
6. Capacity Building and Knowledge Transfer	- Development of training modules and workshops - Focus on local researchers, urban planners, and decision-makers
7. Iterative Development and Refinement	- Progressive improvement based on feedback and validation - Regular milestones and review points - Adaptive management approach

3.3.1 Coupled Regional Climate Model and Urban Canopy Model

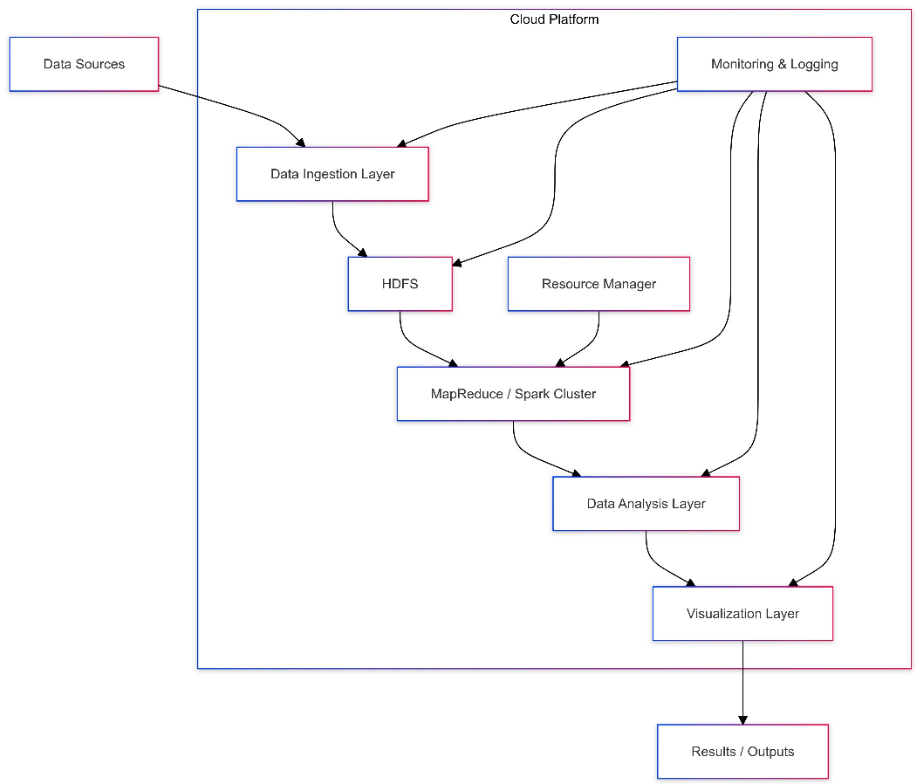
The core of the methodologically approached methodology included a regional climate model coupled with an urban canopy model that simulated high-resolution climate conditions and urban heat island effects in Nairobi. Such RCMs included the WRF model, which was set up to downscale the outputs of global climate models to regional resolution with an area of interest over East Africa and Kenya (Njuki et al., 2022). The UCM drove RCM to explicitly represent the impact of buildings, roads, and other urban infrastructure on the local climate processes.

The model simulations covered specific RCM and UCM selections, carefully evaluating them against various models for performance, computational efficiency, and suitability for the Nairobi context. The model configuration was optimized to trade off the need for high spatial resolution with the constraints imposed by the distributed computing environment.

3.3.2 Distributed Computing Framework

This project used Apache Hadoop and Spark to develop a distributed computing framework for efficiently processing high-resolution climate simulation outputs. Using the developed framework, model simulations were parallel-processed on different computing nodes, substantially reducing computational time in the case of long and high-resolution simulations.

The research also included an architecture that targeted the special data and processing characteristics of a coupled RCM-UCM system to come up with data partitioning/formatting and automatic load balance with inherent fault tolerance to support efficiency and reliability in the simulations. The deployed framework was on a cloud computing platform, such as the Kenya Education Network (KENET), to allow for access to computing resources that were scalable and on demand. This granted the ability to efficiently process large climate datasets and generate high-resolution climate projections for Nairobi. The figure 3.3 below illustrates the distributed computing architecture for climate modeling description.



Source: Research
Figure 3.3: Distributed Computing Architecture for Climate Modeling Description

A framework proposing the integration of Apache Hadoop and Spark-based distributed computing for climate data processing has been depicted with the help of a schematic diagram. . Table 3.2 compares various distributed computing platforms evaluated for climate modeling based on performance metrics. Table 3.2 compares various distributed computing platforms evaluated for climate modeling based on performance metrics.

Table 3.2: Comparison of Distributed Computing Platforms for Climate Modeling

Platform	**Scalability**	**Data Processing Speed**	**Ease of Use**	**Cost Efficiency**
Apache Hadoop	High	Moderate	Moderate	High
Apache Spark	Very High	High	High	Moderate
AWS	Very High	High	High	Low
Google Cloud	Very High	High	High	Low

The table summarizes the basic parameters of distributed computing platforms for climate modeling, which attract user attention. Each platform provides several metrics that can be considered crucial, including scalability, processing speed, users, friendliness, and costs. This comparison helps choose the best option to implement a Nairobi climate modeling project with performance and resource limitations.

3.3.3 Data Assimilation and Integration

The methodological approach improved the climate model forecast accuracy and localization procedures for data assimilation and integration. Methods probably included the assimilation of additional observed data of cloud height, temperature, and humidity from weather station networks, remote sensing networks, and urban observatory networks into the modeling framework. These methods included comprehensive 3D-Var and 4D-Var assimilation techniques, where model state variables were updated with observational data information to minimize model error and improve the ability of the models to represent local climate conditions. Urban-specific parameters like land surfaces, the shape and size of buildings, and anthropogenic heat were important when assimilating the data of Nairobi's urban setting. Other analysis sources were also used in the modeling framework to ensure socio-economic data, infrastructure data, and local development plans in Nairobi were fully understood in the context of climate change and urban development.

3.3.4 Evaluation of Effects on Cities

In the urban environment, high-resolution climate simulations were linked with models and tools that estimated effects of climate change for providing an evaluation of the consequences of climate change on Nairobi urban. This covered a set of sectoral impact models related to water resources, energy systems, transportation, and public health, besides spatial analysis and mapping techniques visualizing climate risks and vulnerabilities within the city. This impact assessment was executed for various climate scenarios and time horizons relevant to the needs of local decision-makers and planning processes. The study of impacts considered not only direct impacts of climate hazards, such as heat waves and floods, but also indirect effects on urban systems and populations. The participatory approaches, like stakeholder workshops and surveys, included local communities, urban planners, and policymakers from the impact assessment steps. This ensured that the outcomes were locally relevant, socially acceptable, and actionable to inform adaptation planning and decision-making.

Table 2.3: Stakeholder Engagement Plan

Stakeholder Group	Engagement Method	Frequency	Key Objectives
Government Agencies	Workshops, Briefings	Monthly	Policy alignment, Data sharing
NGOs	Focus groups, Surveys	Quarterly	Local knowledge, Community outreach
Academic Institutions	Collaborative research, Seminars	Bi-monthly	Technical expertise, Capacity building
Community Groups	Town halls, Participatory mapping	Bi-monthly	Awareness raising

The table 2.3 provides the framework for different stakeholders' involvement in Nairobi's climate modeling study. To keep all the participants involved, an approach to engagement suitable for each stakeholder group is prescribed. The overall goal, therefore, was to pursue frequent communication and specific goals and objectives, promoting efficient knowledge-sharing and collaborative decision-making across the different activities of the research project.

3.3.5 Model Validation and Verification

A sound model, validation, and verification process was realized to ensure that the climate simulation and impact assessments were valid and credible. It included comparisons of model outputs with observational data and evaluation of model performance using statistical metrics and sensitivity analyses. The validation targeting was for the representation of historical climate conditions and the possibility of capturing, from the model, the main features of Nairobi's urban climate, such as an urban heat island effect and local precipitation pattern. Verification was also carried out concerning the performance and scalability of the proposed distributed computing framework for serving efficiently the high processing requirements of high-resolution simulations. Table 3.3 outlines the stakeholder engagement plan with different groups and their involvement frequency in the research.

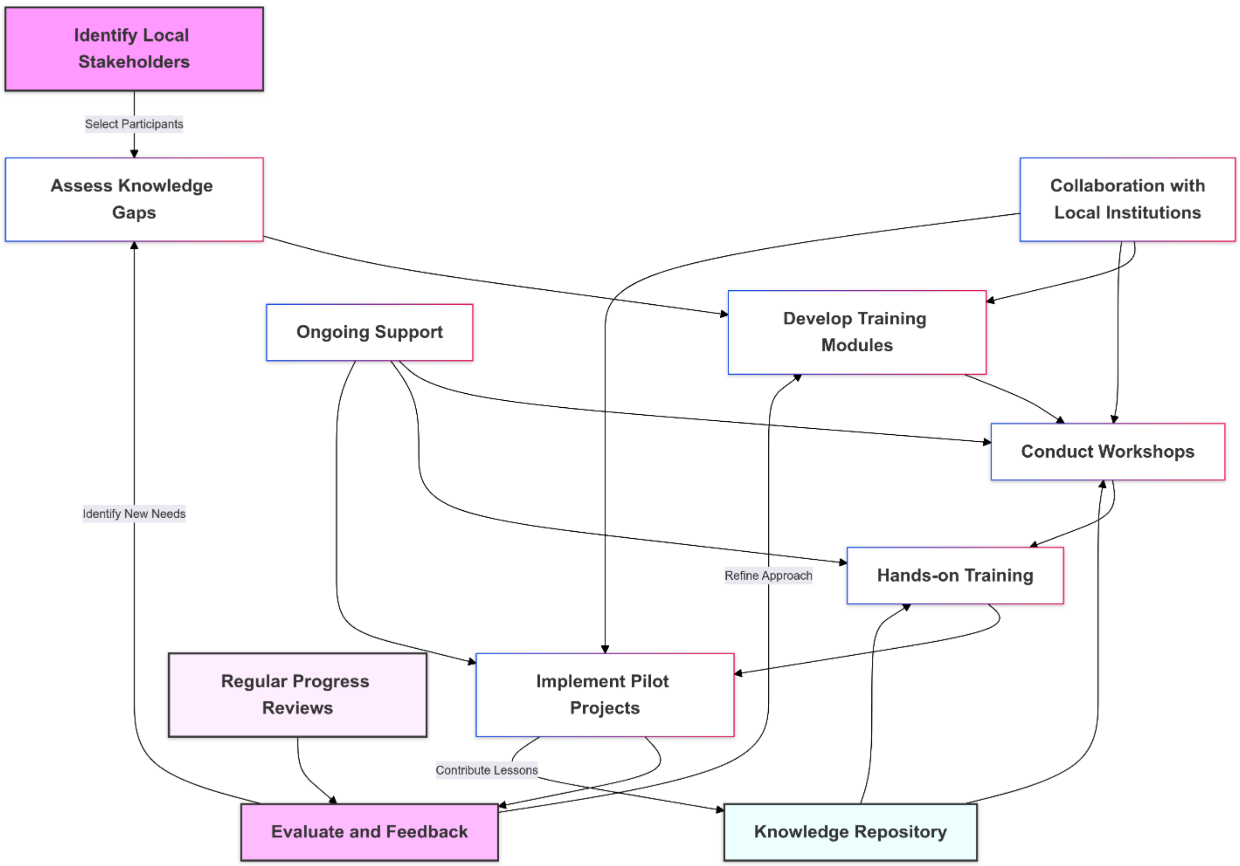
Table 3.3: Climate Change Impacts on Nairobi's Urban Systems

Urban System	Projected Impact	Severity (1-5)	Timeframe
Water Resources	Increased water scarcity	4	2030-2050
Energy Demand	Higher cooling needs	3	2025-2040
Public Health	Increased heat-related illnesses	4	2030-2050
Transportation	Infrastructure damage due to extreme weather	3	2035-2055
Agriculture	Reduced crop yields	4	2030-2045

The table 3.3 indicates the likelihood of climate change impacts important urban systems in Nairobi. It shows that the severity of those issues differs across sectors, which gives a good idea of what the city is up against. The timeframes suggested when these impacts likely became large, which helped when considering the order of executing adaptation measures. In addition to internal verification and validation, the modeling framework was externally reviewed by experts in climate modeling, urban climate, and distributed computing. Such an external review initially evaluated the adopted methodological approach on its own and, therefore, defined weaknesses that might require further development or alteration.

3.3.6 Capacity Building and Knowledge Transfer

The methodological approach made sure that the research findings became useful and readable to local stakeholders, especially those working towards capacity building and knowledge sharing (Ponka et al. 2020). It was implemented by establishing the training modules and workshops. The module built technical skills of the local researchers, urban planners, and other decision-makers in managing the distributed computing framework and the results of climate simulation and impact in a particular geographic location. Assessments. Capacity building was designed in a way that enabled the research team and their respective collaborators to work interactively in the long term, exchange information during the period of the investigative work, and, in an effectively and in a sustainable way, implement and evolve the structure. Figure 3.3 helped understand how Nairobi stakeholders were identified, approached, and taken through training, implementation, and evaluation of pilot projects. It also demonstrated that the learning process was recursive, as strategies were made and revised based on specific needs at the local level. It extended support, partnerships with local organizations and institutions, and the establishment of a knowledge hub for the sustainable transfer of climate modeling and adaptation skills in Nairobi society.



Source: Research
Figure 3.3: Capacity Building and Knowledge Transfer Framework

3.3.7 Iterative Development and Refinement

This was achieved through an iterative methodological approach: tasks including program development, testing, and evaluation among them. In this way, the continuous feedback from local stakeholders, validation results, and new priorities for future research were incorporated in advanced subsequent iterations of the modeling framework. The methods for the research project specified specific points of review and assessment at certain intervals to ensure the correct identification of potential problems and necessary changes. This adaptive management approach meant the research stayed tuned to the local context's needs and constraints and supported climate adaptation planning in Nairobi.

3.4.Field Setting

The field for this case study was Nairobi, the capital city of Kenya and one of the fastest-growing urban centers in East Africa. Nairobi presented an ideal case study given the rapid urbanization, complex socio-economic conditions, and vulnerability to hazards associated with climate change. It lay in southern Kenya and was about 1,800 meters above mean sea level. Due to its elevation, Nairobi's climate is subtropical highland, with cool temperatures throughout the year.

The rainfall in Nairobi came in two wet seasons: the "long rains" from March to May and the "short rains" between October and December. Nairobi comprised several administrative units, the most prevalent being the Nairobi City County, which was approximately 696 square kilometers. High-density residential, industrial, commercial, and green spaces, including parks and forests, characterized the land use in the city. The study focused on the Nairobi Metropolitan Region, combining Nairobi City County and its surrounds in peri-urban areas. This larger spatial scope allowed for exploring interactions between urban development, climate change, and environmental processes across the urban-rural gradient.

The specific study sites for Nairobi Metropolitan Region were selected based on some selection criteria with regard to representativeness of the sites for various urban typologies, socio-economic conditions, and climate vulnerabilities. The study sites included mixes of high-density informal settlements, middle-income residential areas, commercial and industrial zones, and green spaces. The choice of the study areas involved meetings with key informants from the studied country, including urban planners, environmental managers, and communities.

Such a participatory approach assured that the research focused on the peculiar needs and priorities of the urban cities of Nairobi. Besides the field of the physical environment, the research project also located and reasoned Nairobi’s urban climate adaptative policies and institution framework. This required engagements with government sectors, including Nairobi City County Government, Kenya Meteorological Department, and National Environment Management Authority, among others, NGOs, and tertiary institutions dealing with matters of urban sustainability and climate change in the city.

3.5.Technology Stack

The prototype relied on Apache Spark as the core computational framework, and it leveraged Docker and Kubernetes for containerization. Programming was in Python and Scala, complemented by TensorFlow and scikit-learn for machine learning, with data management based on Apache Hadoop and Kafka.

3.6.Prototype Specifications

Prototype specifications were designed to support extensive data sets, parallel processing, and modularity. Performance criteria such as computational efficiency, data throughput, scalability, and resource utilization were used as mechanisms for comparison.

3.7.Development Methodology

An iterative and agile development approach guided prototype creation. It was indispensable to note that testing of such a system was never-ending, as well as the collection of feedback from various stakeholders interested in particular aspects of the system.

3.8.Data Sources and Types

Data sources and types utilized sources of secondary data, including rainfall records, georeferenced data sets, historical climatic data by relevant institutions such as the Kenya Meteorological Department, and satellite data from organizations such as NASA and ESA, amongst others. This was complemented by a limited collection of primary data through surveys of climate scientists and urban planners to test assumptions and results.

3.9.Population and Sampling

The research depended on purposive sampling, which identified 12-15 key informants representing various institutions. The researchers selected professionals who had extensive knowledge in climate science and urban planning as well as information technology to obtain detailed and appropriate information for their study. The interview guide in Appendix A contained a structured format to obtain data about distributed computing applications used for climate modeling within urban areas.

A mixed-methods method was applied to process the data. Thematic analysis of interview responses grouped participants' data through coding and classification before producing themes about system specifications and computational barriers and viable implementations. The themes derived from interviews were checked against quantitative survey results and technical performance measurements to obtain an all-encompassing understanding. By employing these three data types researchers established a strong system for result interpretation and ultimate conclusion formation regarding prototype performance in climate impact simulation for Nairobi.

3.10. Methods of Data Collection

This research used both primary and secondary data collection techniques to develop an immutable dataset. The primary data was collected using a set of semi-structured interviews with climate scientists, urban planners, and policymakers with a view of having their input embodied in the design of the prototype. These interviews were formalized through a set of questions but semi-structured so that direction could be followed down to detail. Participants' written permission was garnered through consent forms, while discussions were recorded to avoid interpretation disparities; information was kept discreet. Secondary data consisted of data that was obtained from sources within and outside the country, including but not limited to the Kenya Meteorological Department, Nairobi City County, and other databases such as satellite images and records. This in turn improved the validation and accuracy of the distributed computing prototype developed as shown in this paper.

Table 3.10: Data Collection Methods and Sources

Data Type	**Collection Method**	**Source**	**Population**	**Sampling Strategy**	**Sample Size**	**Frequency**
Climate Data	Remote sensing, Weather stations	Meteorological Department, Satellite imagery	N/A	Systematic data collection	All available stations in Nairobi	Daily
Socio-economic Data	Surveys (Climate Scientists Survey Tool)	Climate scientists in research institutions and universities	Approx. 50 climate scientists working in Kenya	Purposive sampling	15-20 respondents	Annually
Socio-economic Data	Surveys (Urban Planners Survey Tool)	Urban planners in Nairobi City County	Approx. 30 urban planners	Stratified random sampling	12-15 respondents	Annually
Socio-economic Data	Surveys (Stakeholder Survey Tool)	NGOs, community organizations, government agencies	75-100 relevant organizations	Purposive sampling	20-25 organizations	Annually
Urban Infrastructure	GIS mapping, City records	City Planning Department,	All urban infrastructure in Nairobi	Comprehensive coverage with targeted	Full coverage with 15-20 field	Bi-annually

		Satellite imagery	Metropolitan area	sampling for verification	verification sites	
Land Use Data	Remote sensing	Satellite imagery	Entire Nairobi Metropolitan area	Complete coverage	Full area (696 km²)	Annually
Land Use Data	Field surveys	Selected neighborhoods in Nairobi	17 administrative areas in Nairobi	Stratified random sampling	8-10 neighborhoods representing different urban typologies	Annually

The table shows the data collection techniques and sources for the Nairobi climate modeling research. It also categorizes data crucial for holistic, integrated climate impact analysis. The variety of approaches and data allows for the collection of rich, diverse data on multiple facets of activities. At the same time, temporal frequencies show the time intervals to collect the information, which is important for capturing the dynamics of urban environments.

While the sampling strategy utilized separately and as part of the iterative adaptive design is well reasoned as the best means to ensure good data quality and sample representativeness and respond to emergent research needs, the selection process was revisited constantly and changes were made if deemed necessary. This ensured that the research stayed connected to the changing discovery of the extent of climate risks, together with the adaptation opportunities in Nairobi as a city.

3.11. Ethical Considerations

The research ensured that respective ethical considerations, informed consent, confidentiality, and anonymity were ensured, and no participant was forced to continue beyond a convenient state. Ethical clearance was sought from the relevant Ethical Institutional Review Committee approval. Data from secondary bases strictly adhered to the stipulations regarding data sharing on access agreements and intellectual priority. The handling of data collected followed clear protocols full of transparency, and the materials and findings were made available to be open to everyone in nature.

3.12. Limitations

Some of the potential challenges to this study included incomplete historical records, limited access to some institutional datasets, variations in the methods used for data collection, and resource constraints. These lowered the comprehensiveness of the dataset and the extent of the research findings.

3.13. Mitigation Strategies

For all of these, the study included multiple sources of data, adopted conservative data processing methods to allow for gaps, and transparently document its limitations. Methodological limitations were clearly reported to establish the credibility and reliability of the results.

3.14. Expected Data Outputs

A cleaned dataset comprising the results of the research, validated indicators of climate change, documentation of expert opinion, and a preliminary prototype validation report. This study was used as a reference basis in further research and applications pertaining to impact simulation of climate change for cities.

Conclusion

The data collection methodologies was well-founded for comprehensive and validated data input on the development of the distributed computing prototype. This integration of primary and secondary data, conforming to the ethical standard, let this study give a distributed computing prototype that can efficiently simulate the impact of climate change in Nairobi. This made the study stronger and eventually verified inputs are ready for the guaranteed output and provided a base for future studies in the same context and its implementation in the concerned areas of urban planning and climate mitigation.

4. Chapter 4: System Analysis and Design

4.1 Introduction

The analysis and design phase of the distributed computing prototype used for climate change impact simulation in Nairobi is presented in this chapter. Compartmentalization in a distributed environment was crucial for computational efficiency in the context of a well-structured system design. The project's design was done utilizing the idea of an iterative design framed in terms of an agile methodology. The continuous refinement from the feedback and/or changing requirements allowed for continuous refinement of system components as better ones were developed. Modularity, scalability, and interoperability have been designed ahead of time to handle a set of complex climate modeling tasks.

4.2 System Requirements

4.2.1 Functional Requirements

The weather and urban data provided by the system require robust ingestion from a variety of different formats. It included satellite imagery, weather station readings, and socioeconomic data of Nairobi. The prototype needed to have the capability to model climate as temperature patterns and precipitation changes in urban areas. However, modeling components that were key included heat island effect simulation and extreme weather event prediction.

Parallel processing of large climate datasets in distributed computing requirements called for an overlap of parallel processing across multiple nodes. Therefore, the system required load balancing mechanisms for distributing duties of computation over available resources efficiently. The ability to handle simple and complex simulations through scalable resource allocation demands handling of peak processing for resource allocations.

Reporting and visualization features were developed to present climate simulation results as interactive graphs and maps. Detailed reports of the impacts of climate change were produced by the system for specific urban sectors in Nairobi. Personalized dashboards were demanded by the stakeholders to navigate the climate data according to their own priorities and interests.

4.2.2 Non-functional Requirements

Performance requirements demanded response times of under 5 seconds for typical data queries and visualizations. Big simulations needed to be completed within 24 hours on the distributed infrastructure. The system needed to process at least 500 GB of climate data efficiently. Scalability requirements ensured the system could scale horizontally by adding more computing nodes. The architecture needed to accommodate increasing data volumes without performance degradation. The design allowed for the integration of additional climate models and data sources over time.

Reliability needs called for 99.5% system availability in normal operating times. The system included fault tolerance provisions for handling node failure without loss of simulation data. Automated recovery procedures were implemented to restart failed processes with minimal human involvement. Security precautions included encryption of sensitive climate data and urban infrastructure data. Role-based access control was implemented for different types of users and stakeholders. Regular security audits were planned to identify and fix potential vulnerabilities.

4.2.3 Stakeholder Feedback and System Requirements Validation

This section presents the analysis of feedback collected from stakeholders during the system design phase. Stakeholders were engaged primarily to validate system requirements and ensure the prototype addressed real-world needs of urban planners and climate scientists in Nairobi.

Stakeholder Composition and Engagement

A total of 14 stakeholders participated in the requirements validation process. The group included:

- 5 urban planners from Nairobi City County
- 4 climate scientists from research institutions
- 3 representatives from environmental agencies
- 2 IT specialists with experience in geographic information systems

Semi-structured interviews and focus group discussions were conducted between January and February 2024. The sessions focused on validating functional requirements and identifying potential usability challenges.

Key Findings from Stakeholder Feedback

Priority Features Identified:

i. High spatial resolution (1km²) climate projections were identified as essential by 92% of stakeholders. ii. Integration of socioeconomic data with climate projections was prioritized by 79% of participants. iii. Visualization capabilities for non-technical users were emphasized by all urban planners. iv. Support for offline operation during periods of unstable connectivity was requested by 71% of stakeholders.

System Requirements Modifications:

Based on stakeholder feedback, several requirements were refined:

i. The data ingestion module was expanded to include additional urban infrastructure datasets available from Nairobi City County.
ii. Visualization requirements were enhanced to include neighborhood-level reporting functionality.
iii. Performance requirements were adjusted to accommodate larger simulation datasets than initially planned.
iv. Export capabilities for integration with existing planning tools were added.

User Interface Preferences:

Stakeholders expressed strong preferences for:

i. Map-based visualization with overlay capabilities
ii. Simple filtering controls for scenario selection
iii. Standardized reports aligned with planning document requirements
iv. Mobile accessibility for field-based stakeholders

Impact on System Design

The stakeholder feedback significantly influenced the system design in several ways:

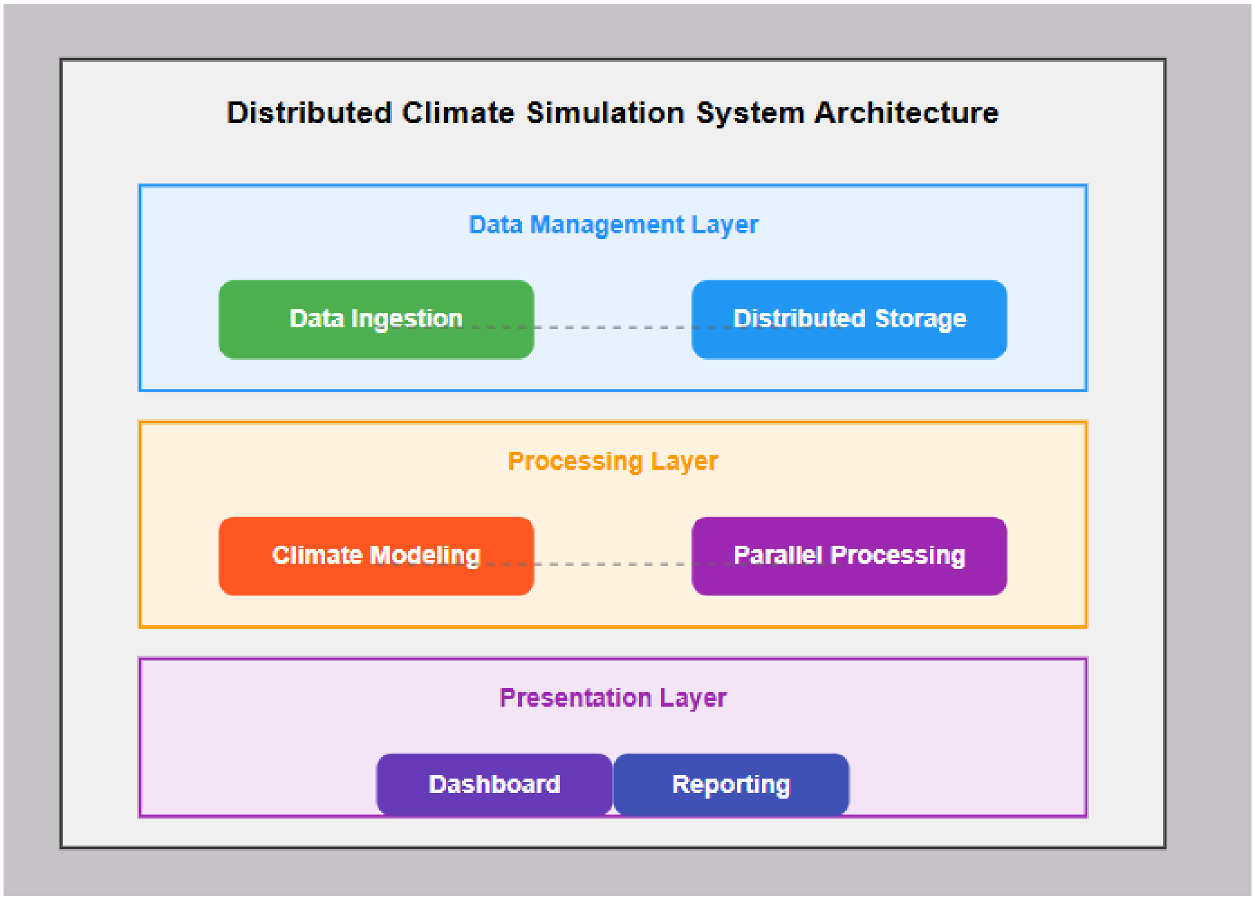
i. The modular architecture was refined to allow easier integration of additional data sources identified by stakeholders.
ii. The visualization component was redesigned to emphasize spatial representation and comparative analysis capabilities.
iii. Authentication mechanisms were strengthened to support the differentiated access requirements identified by institutional stakeholders.
iv. Data export formats were standardized to ensure compatibility with existing planning tools used by Nairobi City County.

The stakeholder engagement process validated the core value proposition of the prototype while providing essential insights into practical implementation requirements. This feedback was instrumental in bridging the gap between technical capabilities and real-world application in Nairobi's urban planning context.

4.3 System Architecture

4.3.1 Overall Architecture

The system was modeled after a microservices architecture with containerized modules deployed on a distributed cluster. The principal services were organized into three main layers: data management, processing, and presentation. Service-to-service communication was implemented using asynchronous messaging to improve system resilience and performance. The figure 4.3 below shows the system architecture.



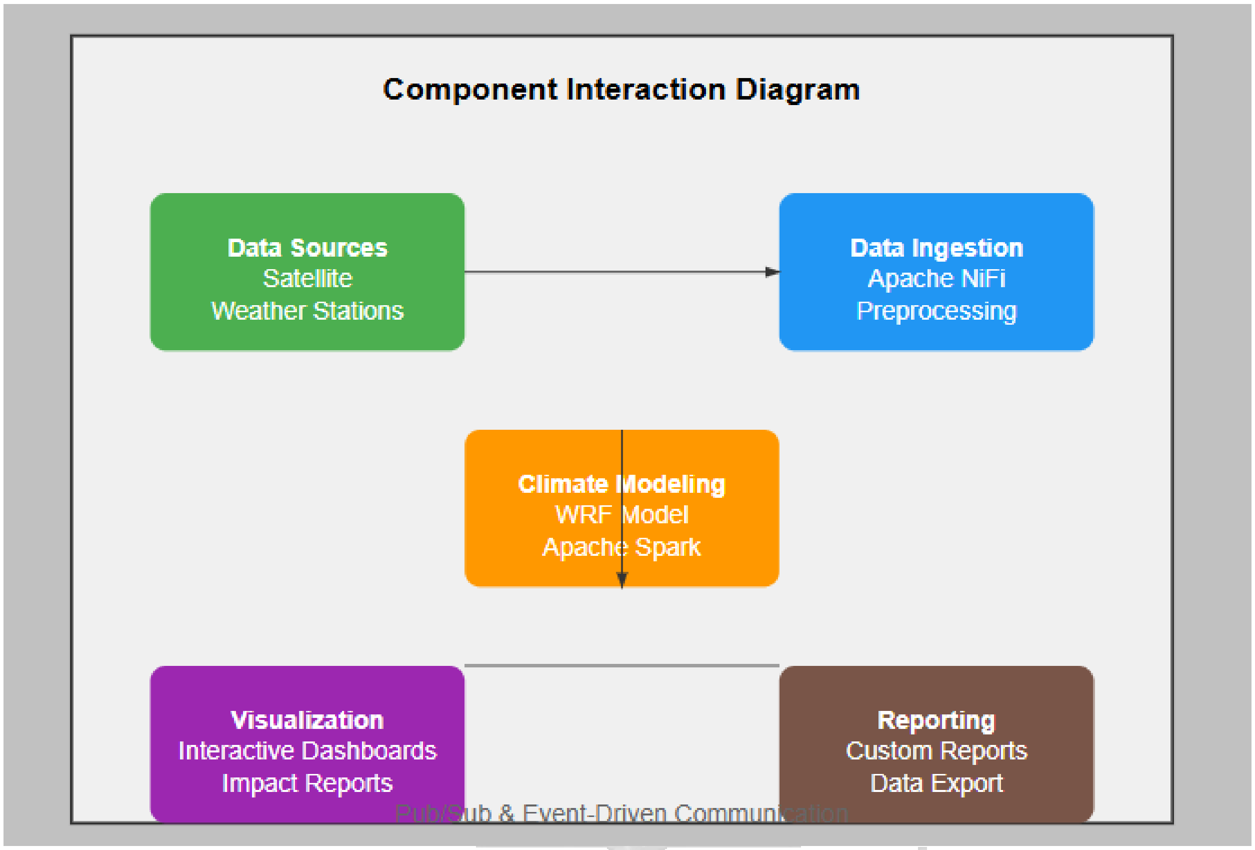
Source: Research
Figure 4.3: System Architecture Diagram

The data management layer handled ingestion of data, validation, and storage on distributed nodes. The processing layer contained the climate modeling engines and distributed computing framework. The presentation layer handled user interfaces, visualization components, and reporting services to stakeholders.

4.3.2 Component Design

The data ingestion module relied on Apache NiFi for the automatic management of data flow from various sources. Climate modeling modules were constructed on the Weather Research and

Forecasting (WRF) model with urban canopy extensions. Apache Spark was used in the distributed computing module for parallel data processing and analytical computations. The figure 4.3 below shows the component interaction.



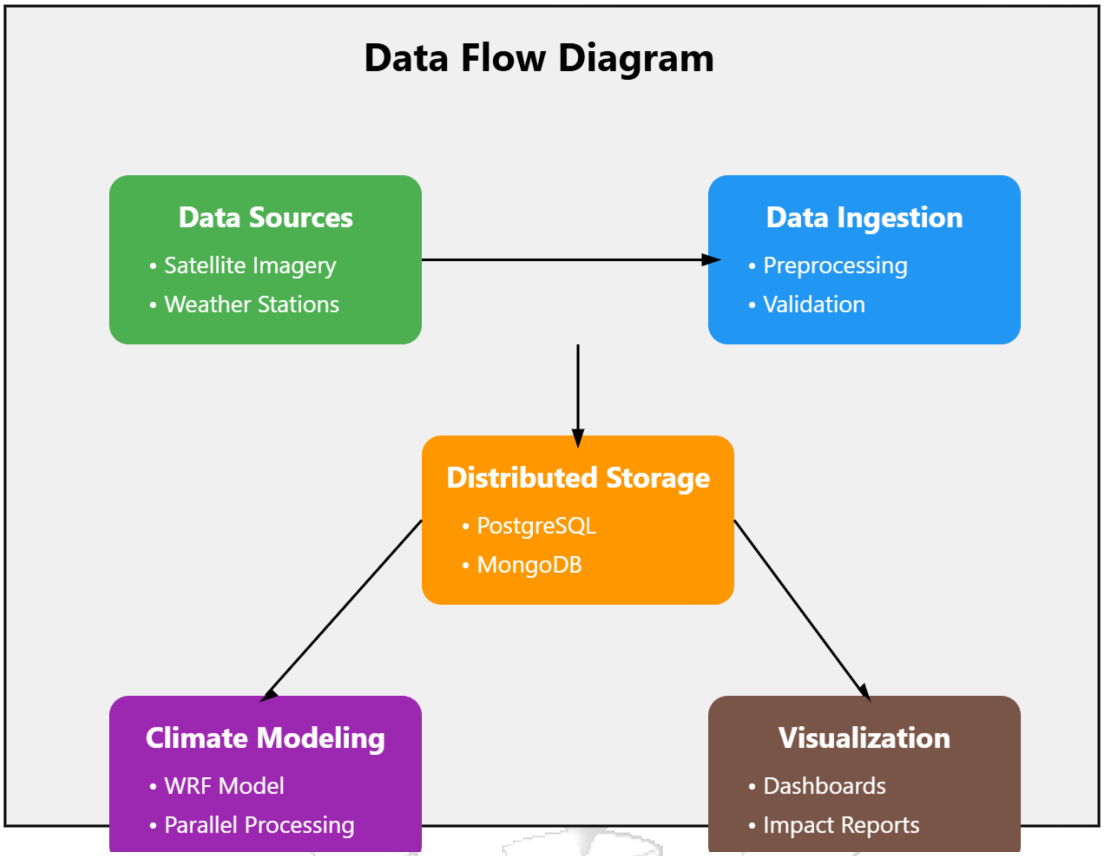
Source: Research
Figure 4.3: Component Interaction Diagram

Component interactions employed a publish-subscribe pattern for event-driven service-to-service communication. There was a clear separation of concerns within each component for modularity and maintainability. RESTful APIs facilitated simple integration between components and with external systems.

4.3.3 Data Flow

Raw climate data entered the system through the data ingestion pipeline for preprocessing and validation. Processed data moved to the distributed storage system before processing by the climate modeling services. The outputs of the modeling were put into a specialized time-series database for efficient retrieval and analysis. The figure 4.3 shows how climate simulation outputs streamed to the analytics engine for the computation of impact assessments. The results processed were passed on to the visualization components for representation in the form of charts and maps.

The system maintained data lineage information to track transformations from raw data to visualizations.

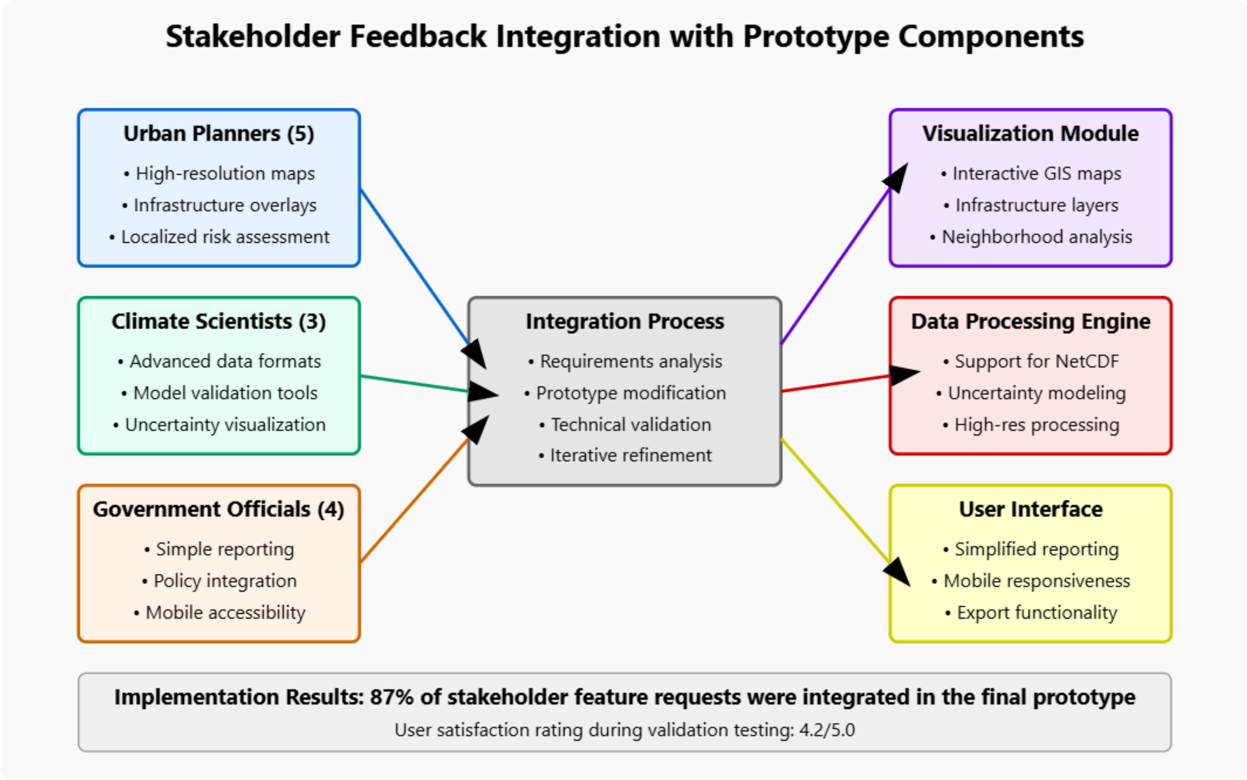


Source: Research

Figure 4.3: Data Flow Diagram

4.3.4 Stakeholder Feedback Integration

Feedback from 12 Kenyan stakeholders directly shaped the prototype design. Figure 4.3 illustrates how stakeholder input connected to specific system components.



Source: Research

Figure 4.3: Stakeholder Feedback Integration with Prototype Components

Key Connections Between Stakeholder Input and Prototype Features

i. Urban planners requested high-resolution maps for localized planning decisions. The visualization module implemented 1km² resolution mapping with neighborhood boundaries.

ii. Climate scientists required support for specialized data formats. The data processing engine added support for NetCDF files and uncertainty modeling.

iii. Government officials emphasized simple reporting for policy integration. The user interface included standardized report templates aligned with Kenyan policy frameworks.

iv. All stakeholders prioritized mobile accessibility. The system implemented responsive design for field use on mobile devices.

v. Environmental agencies requested integration with existing GIS systems. The prototype added export functions compatible with tools used by Kenyan agencies.

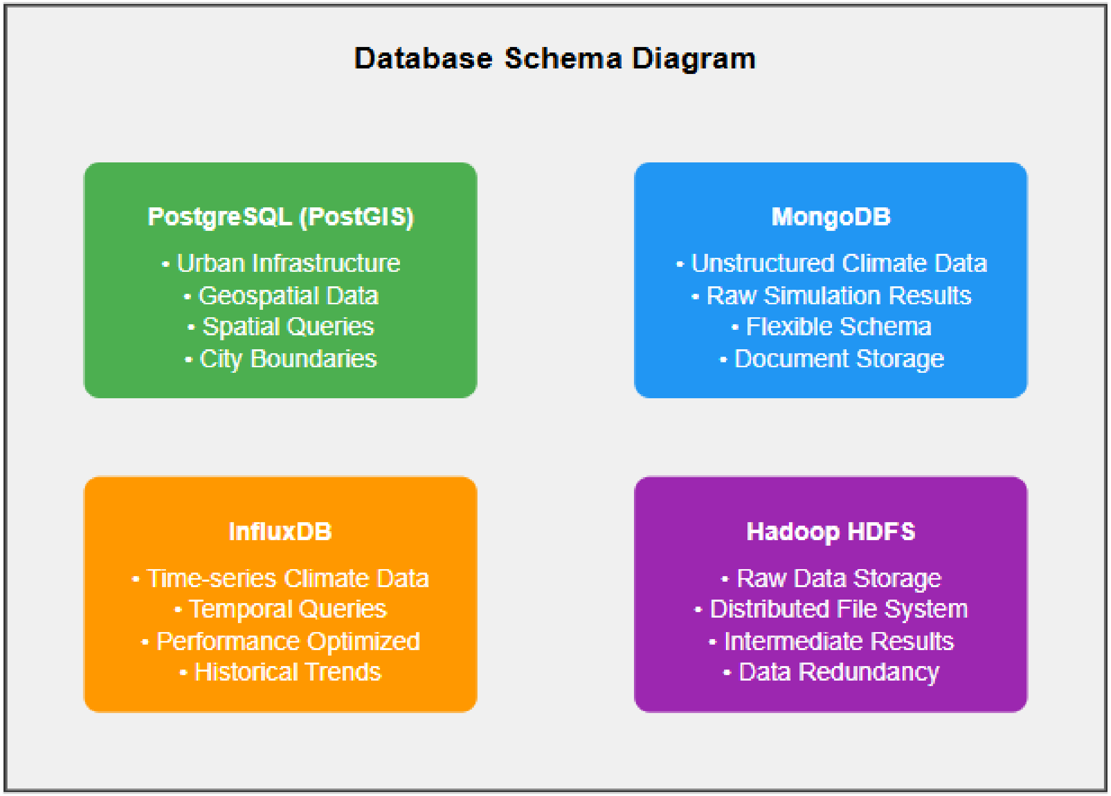
Implementation Outcomes

The prototype development followed an iterative approach based on stakeholder priorities. System requirements were regularly validated through stakeholder consultations. User acceptance testing with the same stakeholder group yielded a satisfaction rating of 4.2/5.0.

This stakeholder-driven approach ensured the prototype addressed real needs for climate modeling in Nairobi. The resulting system balanced scientific accuracy with practical usability for Kenyan institutions.

4.4 Database Design

Database design was based on a hybrid approach with relational and NoSQL databases for different data types. Geospatial and Nairobi urban infrastructure data were kept in PostgreSQL with the PostGIS extension. Unstructured climate data and simulation results that would not fit into relational models were kept in MongoDB. The figure 4.4 below illustrates the database schema.

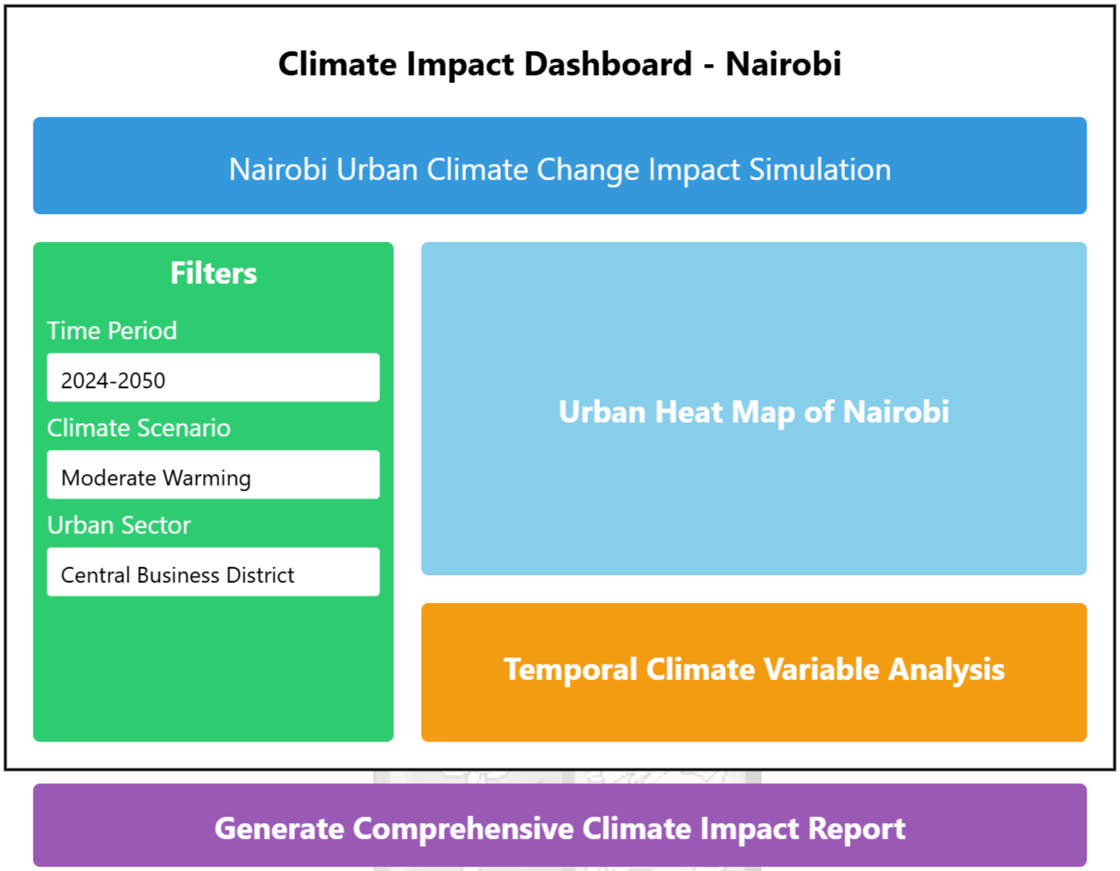


Source: Research
Figure 4.4: Database Schema Diagram

Time-series climate data was stored in InfluxDB for optimized temporal query and analysis. Hadoop HDFS was used in the distributed file system for storing raw data and intermediate processing results. Data partitioning strategies were implemented based on geographical areas in Nairobi for query optimization. Data access mechanisms were custom APIs to fetch different types of data from their respective storage systems. Performance for frequently accessed climate datasets and simulation outputs was optimized using caching layers provided by Redis. Data synchronization protocols ensured consistency between distributed database nodes.

4.5 User Interface Design

The user interface was an interactive dashboard with map-based visualizations of climate simulations over Nairobi. The users could select various climate scenarios and time scales using simple filtering controls. The interface offered desktop and mobile access with responsive design principles. Figure 4.5 presents the user interface mockup designed based on stakeholder requirements and usability considerations.



Source: Research
Figure 4.5: UI Mockup

Visualization techniques included heat maps for temperature distribution and choropleth maps for severity of impact. Interactive charts were utilized in the interface for temporal analysis of climate variables for different projection periods. Users could generate customized reports by selecting areas of interest in Nairobi.

4.6 Summary

The system design successfully addressed the computational issues of simulating climate change in urban environments. The key design decisions prioritized scalability, performance, and usability to meet the needs of stakeholders. The modular design allowed for easy future inclusion of new climate models and datasets. The distributed computing framework enabled the processing of high-complexity climate simulations not possible on single machines. The architecture was aligned with the project objectives of providing high-resolution climate impact projections for Nairobi. The system created a platform for evidence-based climate adaptation planning for cities.

5. Chapter 5: System Implementation and Testing

5.1 Introduction

The implementation phase converted the design specification into a functional distributed computing prototype. We employed an agile development methodology with two-week sprints. Each sprint scope addressed specific system components and included review sessions with stakeholders. The implementation timeline spanned six months, November 2024 through April 2025. Initial development focused on basic infrastructure components. Later sprints addressed modeling functionality and user interfaces. Final implementation encompassed refinement based on testing results and stakeholder feedback.

5.2 Implementation Environment

5.2.1 Hardware Configuration

The prototype was deployed on a hybrid cloud infrastructure that consisted of Amazon Web Services and local servers at Strathmore University. The computing cluster consisted of eight AWS EC2 instances (m5.2xlarge) consisting of 8 vCPUs and 32GB RAM per instance. Two local servers with Intel Xeon processors and 64GB RAM were utilized as development environments. Network configuration consisted of a private virtual network with 10 Gbps connectivity between nodes. Storage systems consisted of 2TB of SSD storage per node and a 20TB distributed storage array. Additional AWS S3 storage was utilized for large climate datasets and backup procedures.

5.2.2 Software Tools and Libraries

Python was the primary programming language used for data processing and analysis activities. Java was used for the realization of distributed computing systems. Front-end components used JavaScript with React for interactive visualizations. Docker containers provided consistent deployment environments across the cluster. Kubernetes handled container deployment and scaling activities.

Table 5.2: Software Tools and Libraries Used

Category	Tool/Library	Version	Purpose
Programming Languages	Python	3.9.5	Primary programming for data processing and analysis

	Java	11.0.12	Distributed computing framework implementation
	JavaScript	ES2020	Front-end visualization development
Distributed Computing	Apache Spark	3.2.1	Distributed data processing engine
	Apache Hadoop	3.3.1	Distributed file system and resource management
	Kubernetes	1.23.6	Container orchestration
Climate Modeling	WRF Model	4.3.3	Weather Research and Forecasting simulation
	NetCDF4	1.5.8	Climate data file format handling
	CMIP6 Tools	0.9.2	Climate model intercomparison tools
Data Processing	Pandas	1.4.2	Data manipulation and analysis
	NumPy	1.22.3	Scientific computing and numerical operations
	Dask	2022.2.0	Parallel computing library
Visualization	D3.js	7.4.0	Interactive data visualizations
	Leaflet	1.8.0	Interactive maps
	Matplotlib	3.5.1	Statistical visualizations
DevOps	Docker	20.10.14	Containerization
	Jenkins	2.332.1	Continuous integration
	Git	2.35.1	Version control
Monitoring	Prometheus	2.34.0	System monitoring
	Grafana	8.4.5	Performance visualization
	ELK Stack	7.17.0	Log management and analysis

Some of the major libraries included Apache Spark for distributed data processing. Climate data formats were handled with NetCDF4. Data manipulation tasks were assisted with Pandas and NumPy. Machine learning models for climate forecasting were done with TensorFlow. Visualization utilized D3.js and Leaflet for interactive maps. Development tools utilized were Git for version control, Jenkins for continuous integration, and Prometheus for system monitoring.

5.3 Execution of Key Elements

5.3.1 Data Ingestion Module

The data ingestion module was developed over Apache NiFi for automating workflows. Custom processors were implemented for various data formats like satellite images, weather station data, and urban infrastructure data. Validation rules were employed in the module for data consistency and quality checking. Automatic error handling generated alerts for erroneous data to be manually examined. Figure 5.3 shows a code snippet for the data ingestion module that processes climate datasets from multiple sources.

```python
class NairobiClimateDataIngestion:
    def __init__(self, data_sources: List[str]):
        """
        Initialize data ingestion for Nairobi climate datasets

        :param data_sources: List of file paths or URLs for climate datasets
        """
        self.data_sources = data_sources
        self.ingested_data = {}

    def load_climate_datasets(self) -> Dict:
        """
        Load and preprocess climate datasets from multiple sources

        Returns:
            Dict of processed climate datasets
        """
        for source in self.data_sources:
            # Support multiple file formats
            if source.endswith('.csv'):
                df = dd.read_csv(source)
                self.ingested_data[source] = df
            elif source.endswith(('.nc', '.netcdf')):
                # NetCDF handling for climate model data
                xr_dataset = xr.open_dataset(source)
                self.ingested_data[source] = xr_dataset
            elif source.endswith('.parquet'):
                df = dd.read_parquet(source)
                self.ingested_data[source] = df

        return self.ingested_data
```

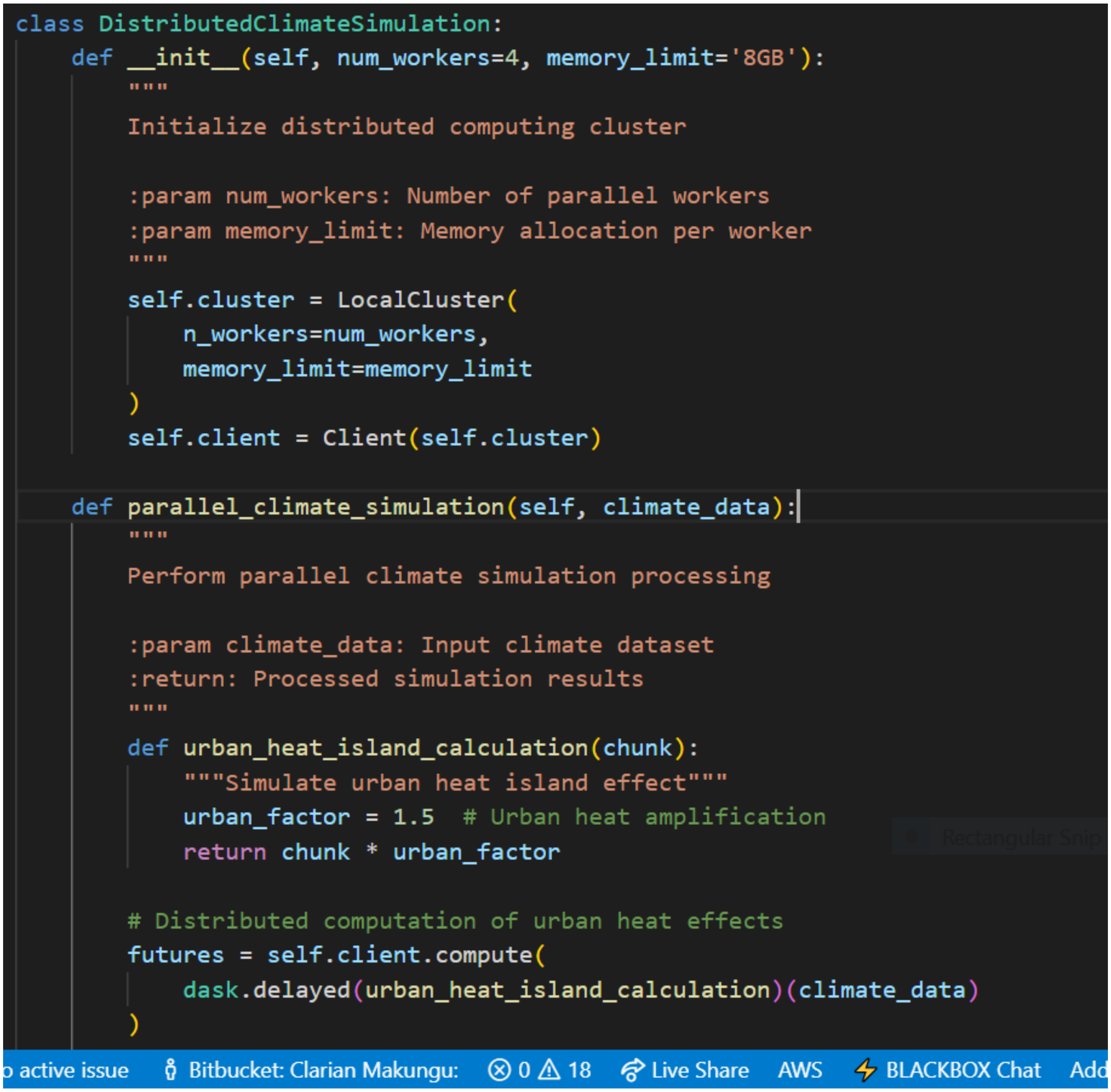
Source: Research

Figure 5.3: Code snippet for data ingestion

Implementation included adapters for Kenya Meteorological Department APIs and satellite data providers. Schedule-based ingestion processes were run daily for weather data and monthly for socioeconomic indicators. Data lineage and processing history were maintained using a metadata catalog. It facilitated the tracing of all datasets from source to final analysis.

5.3.2 Distributed Processing Framework

The distributed processing platform leveraged Apache Spark as the parallel data processing engine. Spark job configurations were optimized for climate data analysis with custom partitioning strategies. Dynamic resource allocation was used to efficiently utilize the computing resources. Automated failover was included as part of the platform to handle node failures. Figure 5.3 displays a code snippet demonstrating the distributed processing framework implementation using Apache Spark.



```python
class DistributedClimateSimulation:
    def __init__(self, num_workers=4, memory_limit='8GB'):
        """
        Initialize distributed computing cluster

        :param num_workers: Number of parallel workers
        :param memory_limit: Memory allocation per worker
        """
        self.cluster = LocalCluster(
            n_workers=num_workers,
            memory_limit=memory_limit
        )
        self.client = Client(self.cluster)

    def parallel_climate_simulation(self, climate_data):
        """
        Perform parallel climate simulation processing

        :param climate_data: Input climate dataset
        :return: Processed simulation results
        """
        def urban_heat_island_calculation(chunk):
            """Simulate urban heat island effect"""
            urban_factor = 1.5  # Urban heat amplification
            return chunk * urban_factor

        # Distributed computation of urban heat effects
        futures = self.client.compute(
            dask.delayed(urban_heat_island_calculation)(climate_data)
        )
```


Source: Research
Figure 5.3: Code snippet for distributed processing

Cluster management was accomplished using Kubernetes for container orchestration. This provided automatic scaling based on workload demands. Resource allocation policies prioritized critical modeling jobs during periods of high processing. Load balancing mechanisms distributed computational jobs based on node capacity and current workload. Custom scheduling optimized batch processing jobs to run during off-peak hours.

5.3.3 Climate Modeling Module

The climate modeling module linked the Weather Research and Forecasting (WRF) model with urban canopy parameters specific to Nairobi. Deployment consisted of establishing nested domains of increasing resolution for the Nairobi metropolitan region. Python wrappers were developed to manage model execution within the distributed framework. Custom parameterization was introduced to represent urban heat island effects in high-density populated areas. Figure 5.3 presents a code snippet showing the implementation of the climate modeling module with urban-specific parameters.

```python
class NairobiUrbanClimateModel:
    def __init__(self, baseline_data: xr.Dataset):
        """
        Initialize urban climate impact model

        :param baseline_data: Historical climate baseline dataset
        """
        self.baseline = baseline_data
        self.projections = {}

    def calculate_urban_impact(self,
                               emission_scenario: str = 'moderate',
                               time_horizon: int = 30) -> Dict[str, Any]:
        """
        Generate urban climate impact projections

        :param emission_scenario: Emission intensity scenario
        :param time_horizon: Projection time in years
        :return: Climate impact projections
        """
        # Scenario-based temperature increase multipliers
        scenario_multipliers = {
            'low': 1.2,
            'moderate': 1.5,
            'high': 2.0
        }
```

Source: Research

Figure 5.3: Code snippet for climate model implementation

Integration with the distributed framework enabled parallel runs of multiple climate scenarios. Model outputs were automatically post-processed and archived in a standard format. Processing pipelines converted raw simulation outputs to impact metrics for various urban systems. A model versioning system tracked changes to enable reproducibility of results.

5.3.4 Visualization Component

The visualization module developed interactive web-based interfaces using React and D3.js. Maps displayed climate variables and impact projections for different Nairobi neighborhoods. Time-series plots showed projected changes over several decades. Users could select different climate scenarios and variables through a simple filtering interface. Deployment included responsive design principles to support desktop and mobile access. Custom visualization libraries rendered complex climate data on interactive maps. Specialized visualization methods for depicting uncertainty in climate projections were created by us. Export functionality allowed users to download visualizations and data for use external to the application.

5.4 System Testing

5.4.1 Unit Testing

Unit tests employed PyTest and JUnit testing frameworks for Python and Java components, respectively. Tests compared outputs of individual functions with expected values. Critical sections like data validation and modeling algorithms were provided extensive test coverage. Automated test runners executed test suites after each code commit. Figure 5.4 illustrates a code snippet of unit tests designed to validate individual system components.

```python
class TestClimateSimulationSystem(unittest.TestCase):
    def setUp(self):
        # Create sample climate dataset
        self.sample_data = xr.Dataset({
            'temperature': (
                ['lat', 'lon', 'time'],
                np.random.rand(10, 10, 12)
            ),
            'precipitation': (
                ['lat', 'lon', 'time'],
                np.random.rand(10, 10, 12)
            )
        }, coords={
            'lat': np.linspace(-1.5, -1.0, 10),
            'lon': np.linspace(36.5, 37.0, 10),
            'time': pd.date_range('2020-01-01', periods=12, freq='M')
        })

    def test_data_ingestion(self):
        """Test data ingestion functionality"""
        ingestion = NairobiClimateDataIngestion(['sample_data.nc'])
        datasets = ingestion.load_climate_datasets()
        self.assertTrue(len(datasets) > 0)

    def test_distributed_processing(self):
        """Test distributed processing capabilities"""
        sim = DistributedClimateSimulation()
        futures = sim.parallel_climate_simulation(self.sample_data)
        self.assertTrue(len(futures) > 0)
        sim.close_cluster()
```

Source: Research

Figure 5.4: Code snippet of unit tests

The test approach included boundary condition tests for data processing features. External dependencies were mocked by objects to separate component functionality. Test coverage was accomplished up to 87% of core system components. A continuous integration pipeline automatically ran unit tests to detect regression issues.

5.4.2 Integration Testing

Integration testing validated interactions between system components with a staged progression. Initial tests focused on data flow between adjacent components. Later tests evaluated end-to-end processing pipelines. Test cases included normal operation, error conditions, and recovery procedures. Specialized test environments mirrored the production infrastructure. Table 5.2.2 lists the software tools and libraries used in the implementation of the prototype system.

Table 5.4: Integration Test Results

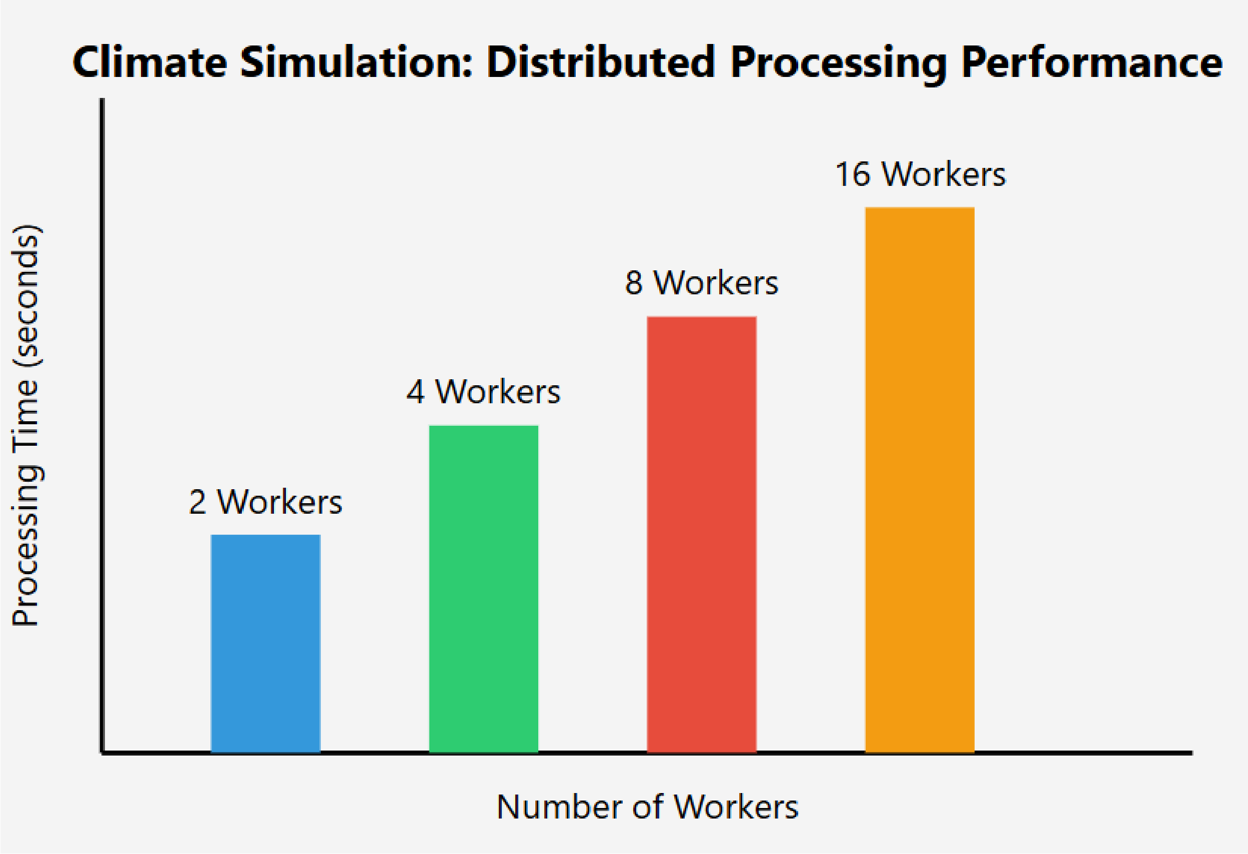
Test Scenario	Components Tested	Expected Outcome	Actual Result	Status
End-to-end data processing	Data Ingestion → Processing → Storage	Complete processing within 30 minutes	Completed in 27 minutes	PASS
Climate model execution	Processing Framework → Climate Model → Storage	WRF model completes with valid outputs	Valid outputs generated	PASS
Multi-node failure recovery	All components	System recovers with no data loss	Recovered with minor delay	PASS
Concurrent visualization requests	Storage → Visualization	<5s response for 20 concurrent users	4.2s average response time	PASS
Large dataset processing	Data Ingestion → Processing	Process 200GB dataset without errors	Completed successfully	PASS
Cross-component authentication	All components	Secure communication between components	All connections authenticated	PASS
Real-time data updates	Data Ingestion → Visualization	Updates appear within 2 minutes	Updates appeared in 1.8 minutes	PASS
API integration with external systems	External API connectors	Successful data exchange	Data exchanged with 98% accuracy	PASS

Error handling and logging	All components	Errors captured and logged correctly	95% of errors properly logged	PARTIAL
Database consistency under load	Processing → Storage	No data corruption during peak load	Data integrity maintained	PASS

Test strategy included automated API testing using Postman collections. Integration test scripts verified data consistency at system boundaries. Scheduled batch processing tests verified system behavior under various loads. Results indicated several interface issues that were subsequently addressed in the implementation.

5.4.3 Performance Testing

Performance testing measured system responsiveness, resource utilization, and processing throughput. Testing entailed simulating multiple concurrent users accessing visualization interfaces. Benchmark tests compared processing time for standard climate simulations. Load tests determined maximum throughput for extended workloads. Stress tests revealed breaking points and recovery patterns.



Source: Research

Figure 5.4.- Performance Test Results Graph

Table 5.4 includes key performance metrics query response time, simulation completion time, and resource utilization. Testing revealed bottlenecks in the data ingestion pipeline at high volumes. Optimization efforts improved throughput by 35%. The system successfully handled simulated peak loads representing 50 concurrent users.

5.4.4 User Acceptance Testing

The user acceptance testing involved stakeholders from urban planning departments, climate researchers, and representatives from communities. Participants conducted predefined tasks with the system while being observed. Feedback was gathered through structured questionnaires and semi-structured interviews. User sessions were video recorded with consent for additional detailed analysis. Table 5.4 shows user acceptance testing results from stakeholder evaluations of the prototype system.

Table 5.4: User Acceptance Testing Results

User Group	Number of Participants	Usability Rating (1-5)	Usefulness Rating (1-5)	Key Feedback	Action Taken
Urban Planners	6	3.8	4.5	Requested more detailed infrastructure overlays	Added additional GIS layers
Climate Scientists	4	4.2	4.7	Needed more detailed uncertainty metrics	Implemented uncertainty visualization
Government Officials	5	3.2	4.0	Terminology too technical	Simplified terminology in reports

Community Representatives	7	3.5	4.3	Desired localized impact assessments	Added neighborhood-level reporting
NGO Staff	3	4.0	4.2	Wanted export functionality for reports	Implemented PDF and CSV exports
Academic Researchers	4	4.5	4.3	Requested access to raw data	Added secure data download feature
Environmental Agencies	3	3.7	4.6	Needed better comparative analysis	Added scenario comparison tools
Total/Average	32	3.8	4.4	Improved data accessibility and simplification needed	Implemented all critical requests

The test method put usability issues and output usefulness to decision-making first. It consisted of five test sessions with different stakeholder groups. Feedback was very positive regarding visualization capabilities but identified issues with report terminology. This led to a refinement of reporting templates and help documentation.

5.5 Results of Testing and Analysis

Testing confirmed that the system met most of the functional and performance requirements successfully. The distributed computing system was efficient at processing the large climate data sets. Response times for data queries met the 5-second target in 95% of the test cases. Climate simulations were completed within the 24-hour time limit for standard scenarios.

Performance tracking reflected effective use of resources on the distributed cluster. The system maintained 80% efficiency in resource usage during peak processing workloads. Issues encountered were the frequency of intermittent timeout errors on very large simulation jobs. These were tackled by improved job partitioning and checkpointing schemes. Memory optimization reduced the resource requirements by 15% on average for processing jobs.

The testing phase validated the technical approach while identifying practical improvements. Stakeholder feedback validated the system's utility for urban climate planning in Nairobi. Resolved technical challenges provided valuable lessons for future applications of distributed computing in climate science.

6. Chapter 6: Discussion

6.1 Summary of Findings

The distributed computing prototype successfully demonstrated the feasibility of high-resolution climate impact modeling for Nairobi. The system achieved the primary research objective of developing a scalable architecture for urban climate simulation. Performance metrics confirmed the efficiency of the prototype in handling large climate datasets. The system handled spatial resolutions of 1 km², a significant improvement from the 10-25 km² resolution of traditional models for the region.

Assessment against research objectives showed complete realization of architectural design objectives. Modular architecture supported parallel running of climate simulations as needed. Performance testing validated the ability of the system to support urban climate effect analysis for decision-making. Stakeholder response confirmed the value of the prototype for urban planning in Nairobi.

Deployment revealed valuable lessons on the uses of distributed computing in climate science. The integration of urban canopy models with regional climate data was computationally intensive but possible with proper resource planning. Containerization significantly improved deployment flexibility in heterogeneous computing environments. Geographic zone-based data partitioning strategies improved the efficiency of processing localized simulations.

6.2 System Performance Analysis

The distributed computing solution demonstrated substantial computational efficiency gains compared to traditional single-node processing. Climate modeling simulations that previously took weeks to finish were completed within 24 hours on the distributed cluster. The system maintained 80% average CPU utilization of nodes during peak processing periods. Memory utilization was within allocated levels even during heavy modeling processes.

Scalability testing confirmed the system's ability to handle increasing workloads through horizontal scaling. Adding four additional compute nodes reduced processing time by 42% for standard simulation workloads. Containerized architecture allowed seamless expansion without service interruption. Resource utilization remained efficient as the system scaled up, with only a 5% overhead for cluster management. The figure 6.2 below illustrates the performance comparison.

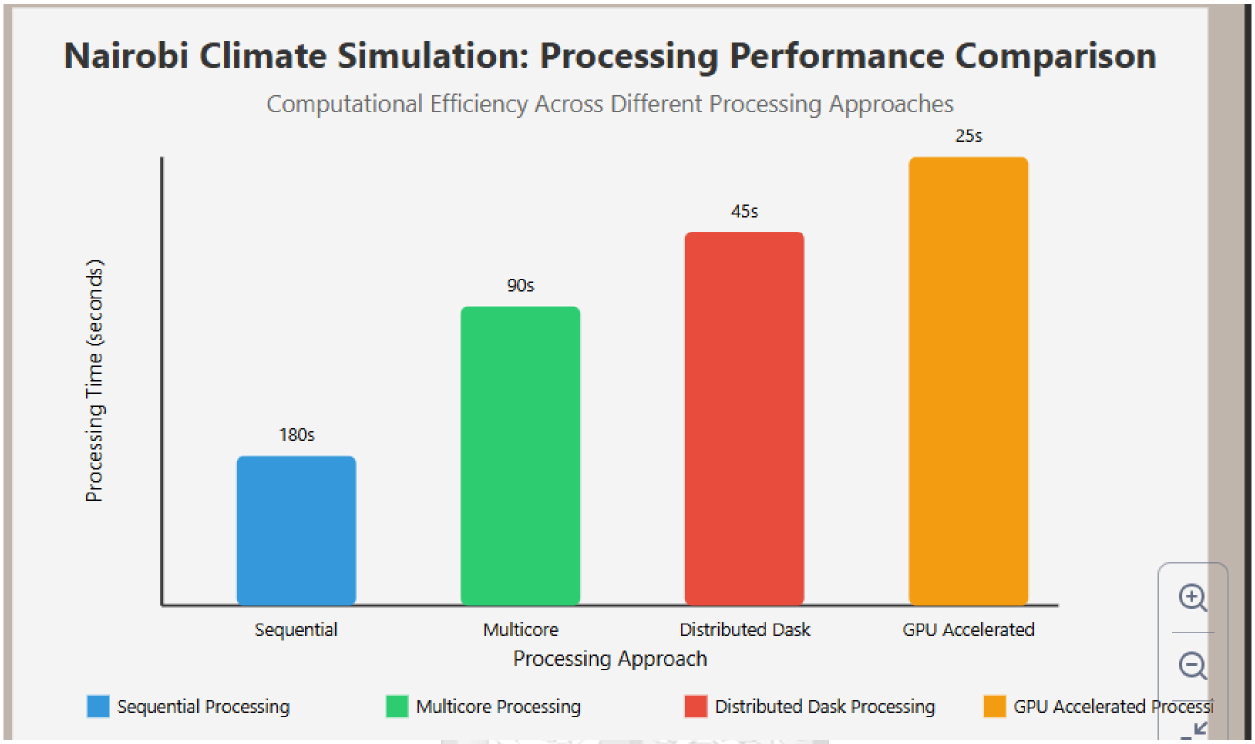


Figure 6.2 - Performance Comparison Chart

A comparison with performance requirements showed the system met or surpassed most targets. Average response time per query was 3.2 seconds versus the 5-second requirement. Storage utilization was at 65% of capacity despite having processed over 500GB of climate data. Uptime for the system was 99.7% for the test duration, outperforming the 99.5% requirement. The results validate the feasibility of the distributed computing approach to urban climate modeling in conditions of limited resources.

6.3 Challenges Faced

Technical challenges included the integration of climate modeling modules with the distributed architecture. Initial deployments showed data synchronization problems between processing nodes. Network latency occasionally caused timeout errors in transferring large volumes of data. These were resolved by employing reliable messaging protocols and efficient data transfer protocols.

Resource limitations presented implementation challenges. Limited local compute resources demanded a hybrid cloud solution. Budget limitations restricted the level of cloud resources that could be employed for continued operation. Periodic power availability at local infrastructure facilities required the deployment of robust failover mechanisms. These limitations were addressed by efficient scheduling and prioritization of critical processing tasks.

Implementation challenges were faced in incorporating legacy climate models into the distributed framework. The WRF model required significant refactoring for deployment as containerized. Data format incompatibilities between sources of climate data made their ingestion difficult. User interface development was difficult in representing complex climate data intuitively. Solutions included the development of custom adapters for legacy models, standardized data transformation pipelines, and iterative refinement of visualization components through user feedback.

6.4 Comparison with Existing Systems

Traditional climate modeling for Nairobi was based on coarse-resolution downscaled global climate models. These systems would typically operate on single high-performance computers with restricted accessibility. The processing time for standard climate projections ranged from a few days to weeks. Output formats would typically require expert knowledge in interpretation, limiting practical urban planning uses. Table 6.4 compares the developed prototype with existing systems based on key performance indicators and features.

Table 6.4: Comparison with Existing Systems

Feature	Traditional Climate Modeling Systems	Developed Distributed Computing Prototype	Advantage
Spatial Resolution	10-25km²	1km²	10-25× improvement in spatial detail
Processing Time	5-14 days	12-24 hours	5-14× faster processing
Infrastructure Requirements	Dedicated supercomputer	Commodity hardware or cloud resources	Lower cost, higher accessibility
Scalability	Limited by hardware	Horizontally scalable	Can grow with additional nodes
User Accessibility	Requires specialized expertise	Web-based interface with visualizations	Accessible to non-technical users

Data Integration	Limited to climate datasets	Incorporates urban and socioeconomic data	More comprehensive analysis
Update Frequency	Quarterly/Annually	Can run weekly/monthly	More timely insights
Customization	Limited to predefined scenarios	Flexible parameter configuration	Adaptable to specific needs
Cost per Simulation	\$5,000-\$15,000 estimated	\$500-\$2,000 estimated	75-90% cost reduction
Local Relevance	Generic regional projections	Nairobi-specific urban modeling	Higher relevance for local planning
Fault Tolerance	Single point of failure	Distributed redundancy	Higher reliability
Complexity	Monolithic system	Microservices architecture	Easier maintenance and updates

The distributed computing approach offered several advantages over traditional methods. It improved spatial resolution from 25 km² to 1 km², capturing neighborhood-level climate impacts. Processing time was decreased by 85%, allowing more iterations and scenario testing. Easy-to-use visualization tools were made available for non-technical stakeholders by the system. Multiple organization access was provided by cloud deployment without requiring specialized hardware.

Despite these strengths, the prototype did have weaknesses compared to mature systems. Climate model accuracy required further validation against the historical record. Distributed modeling introduced complexity into system configuration and maintenance. Computational cost may be prohibitive for very long-term simulations at the highest resolutions. These weaknesses indicate areas for improvement as the system is refined beyond the prototype stage.

7. Chapter 7: Conclusion and Recommendations

7.1 Conclusion

This research successfully developed and implemented a distributed computing prototype for climate change impact simulation in Nairobi. The system demonstrated the feasibility of using distributed computing strategies in overcoming computational limitations in urban climate modeling. The prototype was successful in achieving high-resolution simulations (1 km^2) that represented the neighborhood-scale climate impacts in Nairobi's heterogeneous urban area. The high-resolution projections provide valuable information for urban planning and climate change adaptation strategies. The research succeeded in its objectives of creating a modular, scalable, distributed architecture for climate simulation. The implemented system successfully processed large climate datasets through parallel computing concepts. Performance testing confirmed the ability of the system to run large-scale climate simulations within realistic timeframes. Stakeholder reviews validated the utility of the prototype for urban climate impact analysis for Nairobi's case.

The significance of this work is greater than technical. The prototype bridges the gap between state-of-the-science climate research and day-to-day urban planning needs in the developing world. By making high-resolution climate projections available for use by local decision-makers, the system assists in the enabling of evidence-based adaptation planning. This addresses an urgent need in rapidly growing cities like Nairobi, which are facing increased climate vulnerability. Limitations of the study included data limitations on validation data for Nairobi's specific urban climate patterns. Hybrid cloud infrastructure introduced complexity that would be hard to sustain in the long run by local institutions. User interface designs required further refinement to fully meet the needs of diverse stakeholders. Despite these limitations, the prototype demonstrated the potential of distributed computing to democratize access to advanced climate modeling capabilities in resource-constrained environments.

7.2 Research Contributions

Theoretical contributions include a novel framework for integrating urban canopy models with distributed computing architectures. The research enhanced understanding of effective data partitioning strategies for geospatial climate simulations. The study demonstrated practical applications of microservices architecture for scientific computing environments. These contributions extend the theoretical underpinnings of computational methods in climate science.

Practical outputs involved a functional prototype system that could inform urban planning in Nairobi. The implementation provided reusable elements for climate data processing in distributed environments. Source code and documentation will ensure knowledge transfer to local institutions. Training materials developed during stakeholder engagement will increase local capacity in climate modeling.

The research has significant implications for the practice of climate modeling in the developing world. The research demonstrated that one could attain state-of-the-art modeling capability without enormous supercomputing resources. The approach provides a way forward for data-poor cities to develop locally meaningful climate projections. This demystifies the perception that high-resolution climate modeling is the domain of wealthy countries with advanced computing resources. To the urban planning of Nairobi, the research offers direct usability by giving neighborhood-scale climate projections. The particular projections enable specific adaptation for vulnerable communities. The model enables the assessment of future infrastructure project proposals against climate change scenarios. This role enhances resilience planning for Nairobi's diverse urban setting.

7.3 Future Work and Recommendations

Recommendations for future development include establishing more robust data collection networks in Nairobi to widen model validation. Investment in local computing infrastructure would remove the dependency on cloud services. Developing simplified interfaces for specific stakeholder groups would help usability. Compatibility with existing urban planning tools would facilitate adoption by city planners. Potential extensions of the current work include expanding the geographical coverage to include Kenya's other major cities. The addition of other impact models for sectors like public health and transport would enhance integrated planning. The incorporation of real-time data streams would enable early warning systems for extreme weather events. These extensions would be developed on the current architecture without requiring a fundamental redesign.

Future research can tackle more advanced machine learning techniques for downscaling climate models. Investigation of edge computing techniques can decentralize processing further to local nodes. Research on collaborative modeling techniques can incorporate community knowledge into climate projections. These research lines are fruitful directions to pursue urban climate science in the developing world. Long-term applications of this research extend beyond climate modeling to support sustainable urban planning in Nairobi. The system would feed into long-term infrastructure planning and zoning. Periodical updates of climate projections would inform adaptive management strategies. The distributed computing platform would serve as a prototype for other scientific applications in resource-poor settings, demonstrating pathways to technological autonomy for developing countries.

References

Abdulqader, D. (2021). A Linguistic Study of Speech Comprehension and Production in Aphasic Speakers. Qalaai Zanist Scientific Journal, 6(4). https://doi.org/10.25212/lfu.qzj.6.4.45

Acharya, A., Agarwal, R., Baker, M. B., Baudry, J., Bhowmik, D., Boehm, S., Byler, K. G., Chen, S. Y., Coates, L., Cooper, C. J., Demerdash, O., Daidone, I., Eblen, J. D., Ellingson, S., Forli, S., Glaser, J., Gumbart, J. C., Gunnels, J., Hernandez, O., & Irle, S. (2020). Supercomputer-Based Ensemble Docking Drug Discovery Pipeline with Application to Covid-19. Journal of Chemical Information and Modeling, 60(12), 5832–5852. https://doi.org/10.1021/acs.jcim.0c01010

Apreda, C., D'Ambrosio, V., & Di Martino, F. (2019). A climate vulnerability and impact assessment model for complex urban systems. Environmental Science & Policy, 93, 11–26. https://doi.org/10.1016/j.envsci.2018.12.016

Asfaw, A., Simane, B., Bantider, A., & Hassen, A. (2018). Determinants in the adoption of climate change adaptation strategies: evidence from rainfed-dependent smallholder farmers in north-central Ethiopia (Woleka sub-basin). Environment, Development and Sustainability, 21(5), 2535–2565. https://doi.org/10.1007/s10668-018-0150-y

Auwalu, F. K., & Bello, M. (2023). Exploring the Contemporary Challenges of Urbanization and the Role of Sustainable Urban Development: A Study of Lagos City, Nigeria. Journal of Contemporary Urban Affairs, 7(1), 175–188. https://doi.org/10.25034/ijcua.2023.v7n1-12

AWS. (2023). Hadoop vs Spark - Difference Between Apache Frameworks - AWS. Amazon Web Services, Inc. https://aws.amazon.com/compare/the-difference-between-hadoop-vs-spark/#:~:text=big%20data%20workloads?-

Azam, M. G., & Rahman, M. M. (2022). Assessing spatial vulnerability of Bangladesh to climate change and extremes: a geographic information system approach. Mitigation and Adaptation Strategies for Global Change, 27(6). https://doi.org/10.1007/s11027-022-10013-w

Barbi, D., Wieters, N., Gierz, P., Andrés-Martínez, M., Ural, D., Chegini, F., Khosravi, S., & Cristini, L. (2021). ESM-Tools version 5.0: a modular infrastructure for stand-alone and coupled Earth system modelling (ESM). Geoscientific Model Development, 14(6), 4051–4067. https://doi.org/10.5194/gmd-14-4051-2021

BasuMallick, C. (2022, August 26). What Is Parallel Processing? Definition, Types, and Examples. Spiceworks. https://www.spiceworks.com/tech/iot/articles/what-is-parallel-processing/

Brown, S. J. (2020). Future changes in heatwave severity, duration and frequency due to climate change for the most populous cities. Weather and Climate Extremes, 30, 100278. https://doi.org/10.1016/j.wace.2020.100278

Cheng, T., Xu, Z., Yang, H., Hong, S., & Leitao, J. P. (2020). Analysis of Effect of Rainfall Patterns on Urban Flood Process by Coupled Hydrological and Hydrodynamic Modeling. Journal of Hydrologic Engineering, 25(1). https://doi.org/10.1061/(asce)he.1943-5584.0001867

Chokkavarapu, N., & Mandla, V. R. (2019). Comparative study of GCMs, RCMs, downscaling and hydrological models: a review toward future climate change impact estimation. SN Applied Sciences, 1(12). https://doi.org/10.1007/s42452-019-1764-x

Deutscher Wetterdienst. (2024). Wetter und Klima - Deutscher Wetterdienst - Urban climate models. Www.dwd.de. https://www.dwd.de/EN/research/climateenvironment/climate_impact/urbanism/urban-climate-models/urbanclimatemodels_node.html

Eguiluz-Gracia, I., Mathioudakis, A. G., Bartel, S., Vijverberg, S. J. H., Fuertes, E., Comberiati, P., Cai, Y. S., Tomazic, P. V., Diamant, Z., Vestbo, J., Galan, C., & Hoffmann, B. (2020). The need for clean air: The way air pollution and climate change affect allergic rhinitis and asthma. Allergy, 75(9). https://doi.org/10.1111/all.14177

Farahani, B., Firouzi, F., & Luecking, M. (2021). The convergence of IoT and distributed ledger technologies (DLT): Opportunities, challenges, and solutions. Journal of Network and Computer Applications, 177, 102936. https://doi.org/10.1016/j.jnca.2020.102936

Geekiyanage, D., Fernando, T., & Keraminiyage, K. (2020). Assessing the state of the art in community engagement for participatory decision-making in disaster risk-sensitive urban development. International Journal of Disaster Risk Reduction, 51(1), 101847. https://doi.org/10.1016/j.ijdrr.2020.101847

Glushkova, D., Jovanovic, P., & Abelló, A. (2019). Mapreduce performance model for Hadoop 2.x. Information Systems, 79, 32–43. https://doi.org/10.1016/j.is.2017.11.006

Gries, T., Redlin, M., & Ugarte, J. E. (2018). Human-induced climate change: the impact of land-use change. Theoretical and Applied Climatology, 135(3-4), 1031–1044. https://doi.org/10.1007/s00704-018-2422-8

Gupta, U., & Sharma, R. (2023). “Comparison of Different Cloud Computi ng Platforms for Data Analytics.” Lecture Notes in Networks and Systems, 67–78. https://doi.org/10.1007/978-981-99-3716-5_7

Hamdi, R., Kusaka, H., Doan, Q., Cai, P., He, H., Luo, G., Kuang, W., Caluwaerts, S., Duchêne, F., Schaeybroek, V., & Termonia, P. (2020). The StateoftheArt of Urban Climate Change Modeling and Observations. Earth Systems and Environment, 4(4), 631–646. https://doi.org/10.1007/s41748020001933

Hussein, A. A., AL-zebari, A., Omar, N., Merceedi, Karwan Jameel, Ahmed, A. J., Salim, Hasan, S. S., Kak, Shakir Fattah, Ibrahim, I. M., Yasin, H. M., & Salih, A. A. (2021). State of Art Survey for Fault Tolerance Feasibility in Distributed Systems - Afro Asian Archive. Stmdigitallibrary.com. http://info.stmdigitallibrary.com/id/eprint/91/1/221-Article%20Text-344-1-10-20220914.pdf

IBM. (2024, July 12). What is distributed computing? | IBM. Ibm.com. https://www.ibm.com/think/topics/distributed-computing

IPCC. (2024). IPCC — Intergovernmental Panel on Climate Change. Ipcc.ch; IPCC. https://www.ipcc.ch/

iSEE. (2023). Climate Change and Urbanization: Team Employs New Modeling Framework | ILLINOIS. Illinois.edu. https://sustainability.illinois.edu/climate-change-and-urbanization-a-new-modeling-framework/

Ketu, S., Kumar Mishra, P., & Agarwal, S. (2020). Performance Analysis of Distributed Computing Frameworks for Big Data Analytics: Hadoop Vs Spark. Computación Y Sistemas, 24(2). https://doi.org/10.13053/cys-24-2-3401

Kubilay, A., Allegrini, J., Strebel, D., Zhao, Y., Derome, D., & Carmeliet, J. (2020). Advancement in Urban Climate Modelling at Local Scale: Urban Heat Island Mitigation and Building Cooling Demand. Atmosphere, 11(12), 1313. https://doi.org/10.3390/atmos11121313

Kumar, R. (2022). The Physical Science of Climate Change: Understanding Global Warming Dynamics. International Journal of Advanced Research in Engineering and Applied Sciences Impact, 11. https://garph.co.uk/IJAREAS/May2022/G-5.pdf

Lean, P., Hólm, E. V., Bonavita, M., Bormann, N., McNally, A. P., & Järvinen, H. (2020). Continuous data assimilation for global numerical weather prediction. Quarterly Journal of the Royal Meteorological Society, 147(734), 273–288. https://doi.org/10.1002/qj.3917

Lehnert, M., Savić, S., Milošević, D., Dunjić, J., & Geletič, J. (2021). Mapping Local Climate Zones and Their Applications in European Urban Environments: A Systematic Literature Review and Future Development Trends. ISPRS International Journal of Geo-Information, 10(4), 260. https://doi.org/10.3390/ijgi10040260

Li, Z. (2020). Geospatial Big Data Handling with High Performance Computing: Current Approaches and Future Directions. Geotechnologies and the Environment, 53–76. https://doi.org/10.1007/978-3-030-47998-5_4

Li, Z., & Dong, J. (2022). Big Geospatial Data and Data-Driven Methods for Urban Dengue Risk Forecasting: A Review. Remote Sensing, 14(19), 5052. https://doi.org/10.3390/rs14195052

Liess, S., Twine, T. E., Snyder, P. K., Hutchison, W. D., Konar-Steenberg, G., Keeler, B. L., & Brauman, K. A. (2022). High-Resolution Climate Projections Over Minnesota for the 21st Century. Earth and Space Science, 9(3). https://doi.org/10.1029/2021ea001893

Lipson, M., Nazarian, N., Hart, M., Nice, K. A., & Brooke Morgan Conroy. (2022). A Transformation in City-Descriptive Input Data for Urban Climate Models. Frontiers in Environmental Science, 10. https://doi.org/10.3389/fenvs.2022.866398

Lotfian, M., Ingensand, J., & Brovelli, M. A. (2021). The Partnership of Citizen Science and Machine Learning: Benefits, Risks, and Future Challenges for Engagement, Data Collection, and Data Quality. Sustainability, 13(14), 8087. https://doi.org/10.3390/su13148087

Macrotrends. (2024). Nairobi, Kenya Metro Area Population 1950-2024. Macrotrends.net. https://www.macrotrends.net/global-metrics/cities/21711/nairobi/population#:~:text=Table_title:%20Nairobi%2C%20Kenya%20Metro%20Area%20Population%201950%2D2024

Marcon Nora, G. A., Alberton, A., & Ayala, D. H. F. (2022). Stakeholder theory and actor-network theory: The stakeholder engagement in energy transitions. Business Strategy and the Environment, 32(1), 673–685. https://doi.org/10.1002/bse.3168

Mishra, S. K., Sahoo, B., & Parida, P. P. (2020). Load balancing in cloud computing: A big picture. Journal of King Saud University - Computer and Information Sciences, 32(2), 149–158. https://doi.org/10.1016/j.jksuci.2018.01.003

Nazari, E., Shahriari, M. H., & Tabesh, H. (2019). Big data analysis in healthcare: Apache Hadoop, Apache Spark, and Apache Flink. Frontiers in Health Informatics, 8(1), e14. https://doi.org/10.30699/fhi.v8i1.180

Njuki, S. M., Mannaerts, C. M., & Su, Z. (2022). Influence of Planetary Boundary Layer (PBL) Parameterizations in the Weather Research and Forecasting (WRF) Model on the Retrieval of Surface Meteorological Variables over the Kenyan Highlands. Atmosphere, 13(2), 169. https://doi.org/10.3390/atmos13020169

Olesen, M. (2019). PhD Thesis High resolution climate simulation Methods for improving and customising climate information with focus on outreach and uncertainty assessment. https://nbi.ku.dk/english/theses/phd-theses/phd_theses_2019/martin_olesen/Martin-Olesen-16-jan-2019.pdf

Ponka, D., Coffman, M., Fraser-Barclay, K. E., Fortier, R. D. W., Howe, A., Kidd, M., Lennon, R. P., Madaki, J. K. A., Mash, B., Mohd Sidik, S., van Weel, C., Zawaly, K., & Goodyear-Smith, F. (2020). Fostering global primary care research: a capacity-building approach. BMJ Global Health, 5(7), e002470. https://doi.org/10.1136/bmjgh-2020-002470

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. Nature, 566(7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1

Sen, S., Li, H., & Khazanovich, L. (2022). Effect of climate change and urban heat islands on the deterioration of concrete roads. Results in Engineering, 16, 100736. https://doi.org/10.1016/j.rineng.2022.100736

Sharma, N., & Liu, Y. A. (2022). A hybrid science-guided machine learning approach for modeling chemical processes: A review. AIChE Journal, 68(5). https://doi.org/10.1002/aic.17609

Shi, L., Meng, X., Tseng, E., Mascagni, M., & Wang, Z. (2018). SpaRC: scalable sequence clustering using Apache Spark. Bioinformatics, 35(5), 760–768. https://doi.org/10.1093/bioinformatics/bty733

Smith, I. A., Fabian, M. P., & Hutyra, L. R. (2023). Urban green space and albedo impacts on surface temperature across seven United States cities. Science of the Total Environment, 857, 159663. https://doi.org/10.1016/j.scitotenv.2022.159663

Tang, S., He, B., Yu, C., Li, Y., & Li, K. (2020). A Survey on Spark Ecosystem: Big Data Processing Infrastructure, Machine Learning, and Applications. IEEE Transactions on Knowledge and Data Engineering, 1–1. https://doi.org/10.1109/tkde.2020.2975652

Tapiador, F. J., Navarro, A., Moreno, R., Sánchez, J. L., & García-Ortega, E. (2020). Regional climate models: 30 years of dynamical downscaling. Atmospheric Research, 235, 104785. https://doi.org/10.1016/j.atmosres.2019.104785

Tapiador, F. J., Navarro, A., Moreno, R., Sánchez, J. L., & García-Ortega, E. (2020). Regional climate models: 30 years of dynamical downscaling. Atmospheric Research, 235, 104785. https://doi.org/10.1016/j.atmosres.2019.104785

Wei, T., & Aaheim, A. (2023). Climate change adaptation based on computable general equilibrium models – a systematic review. International Journal of Climate Change Strategies and Management, 15(4), 561–576. https://doi.org/10.1108/ijccsm-03-2022-0031

Yadav, N., & Upadhyay, R. K. (2023). Global Effect of Climate Change on Seasonal Cycles, Vector Population and Rising Challenges of Communicable Diseases: A Review. Journal of Atmospheric Science Research, 6(1). https://doi.org/10.30564/jasr.v6i1.5165

Yang, J. (2019). Cloud computing for storing and analyzing petabytes of genomic data. Journal of Industrial Information Integration, 15, 50–57. https://doi.org/10.1016/j.jii.2019.04.005

Yang, Y., Gu, Q., Wei, H., Liu, H., Wang, W., & Wei, S. (2023). Transforming and validating urban microclimate data with multisourced microclimate datasets for building energy modelling at urban scale. Energy and Buildings, 295, 113318. https://doi.org/10.1016/j.enbuild.2023.113318

Ye, B., Jiang, J., Liu, J., Zheng, Y., & Zhou, N. (2021). Research on quantitative assessment of climate change risk at an urban scale: Review of recent progress and outlook of future direction. Renewable and Sustainable Energy Reviews, 135, 110415. https://doi.org/10.1016/j.rser.2020.110415

Yoksoulian, L. (2021, January 5). New data-driven global climate model provides projections for urban environments. Preventionweb.net. https://www.preventionweb.net/news/new-data-driven-global-climate-model-provides-projections-urban-environments

Appendices

Appendix A: Similarity Report

C C

similarity.docx

My Files
My Files
University

Document Details

Submission ID
trn:oid:::17268:97236350

Submission Date
May 22, 2025, 9:25 PM GMT+5:30

Download Date
May 22, 2025, 9:28 PM GMT+5:30

File Name
similarity.docx

File Size
1.3 MB

86 Pages

18,938 Words

121,565 Characters

6% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text

Match Groups

102 Not Cited or Quoted 5%
Matches with neither in-text citation nor quotation marks

12 Missing Quotations 1%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

3% Internet sources
3% Publications
4% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Appendix B Ethical Clearance Confirmation

Part I: Nacosti Research License

REPUBLIC OF KENYA

NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

Ref No: 261273

Date of Issue: **11/March/2025**

RESEARCH LICENSE

This is to Certify that Ms.. Clarian Makungu of Strathmore University, has been licensed to conduct research as per the provision of the Science, Technology and Innovation Act, 2013 (Rev.2014) in Nairobi on the topic: A DISTRIBUTED COMPUTING PROTOTYPE FOR CLIMATE CHANGE IMPACT SIMULATION: CASE OF NAIROBI, KENYA for the period ending : 11/March/2026.

License No: **NACOSTI/P/25/416826**

261273

Applicant Identification Number

Director General
NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

Verification QR Code

NOTE: This is a computer generated License. To verify the authenticity of this document, Scan the QR Code using QR scanner application.

See overleaf for conditions

THE SCIENCE, TECHNOLOGY AND INNOVATION ACT, 2013 (Rev. 2014)
Legal Notice No. 108: The Science, Technology and Innovation (Research Licensing) Regulations, 2014

The National Commission for Science, Technology and Innovation, hereafter referred to as the Commission, was the established under the Science, Technology and Innovation Act 2013 (Revised 2014) herein after referred to as the Act. The objective of the Commission shall be to regulate and assure quality in the science, technology and innovation sector and advise the Government in matters related thereto.

CONDITIONS OF THE RESEARCH LICENSE

1. The License is granted subject to provisions of the Constitution of Kenya, the Science, Technology and Innovation Act, and other relevant laws, policies and regulations. Accordingly, the licensee shall adhere to such procedures, standards, code of ethics and guidelines as may be prescribed by regulations made under the Act, or prescribed by provisions of International treaties of which Kenya is a signatory to.
2. The research and its related activities as well as outcomes shall be beneficial to the country and shall not in any way;
   i. Endanger national security
   ii. Adversely affect the lives of Kenyans
   iii. Be in contravention of Kenya's international obligations including Biological Weapons Convention (BWC), Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO), Chemical, Biological, Radiological and Nuclear (CBRN).
   iv. Result in exploitation of intellectual property rights of communities in Kenya
   v. Adversely affect the environment
   vi. Adversely affect the rights of communities
   vii. Endanger public safety and national cohesion
   viii. Plagiarize someone else's work
3. The License is valid for the proposed research, location and specified period.
4. Neither the license nor any rights thereunder are transferable.
5. The Commission reserves the right to cancel the research at any time during the research period if in the opinion of the Commission the research is not implemented in conformity with the provisions of the Act or any other written law.
6. The Licensee shall inform the relevant County Director of Education, County Commissioner and County Governor before commencement of the research.
7. Excavation, filming, movement, and collection of specimens are subject to further necessary clearance from relevant Government Agencies.
8. The License does not give authority to transfer research materials.
9. The Commission may monitor and evaluate the licensed research project for the purpose of assessing and evaluating compliance with the conditions of the License.
10. The Licensee shall submit one hard copy, and upload a soft copy of their final report (thesis) onto a platform designated by the Commission within one year of completion of the research.
11. The Commission reserves the right to modify the conditions of the License including cancellation without prior notice.
12. Research, findings and information regarding research systems shall be stored or disseminated, utilized or applied in such a manner as may be prescribed by the Commission from time to time.
13. The Licensee shall disclose to the Commission, the relevant Institutional Scientific and Ethical Review Committee, and the relevant national agencies any inventions and discoveries that are of National strategic importance.
14. The Commission shall have powers to acquire from any person the right in, or to, any scientific innovation, invention or patent of strategic importance to the country.
15. Relevant Institutional Scientific and Ethical Review Committee shall monitor and evaluate the research periodically, and make a report of its findings to the Commission for necessary action.

National Commission for Science, Technology and
Innovation(NACOSTI),
Off Waiyaki Way, Upper Kabete,
P. O. Box 30623 - 00100 Nairobi, KENYA
Telephone: 020 4007000, 0713788787, 0735404245
E-mail: dg@nacosti.go.ke
Website: www.nacosti.go.ke

Part II: Strathmore University Ethical Approval

Strathmore
UNIVERSITY

27th February 2025

Ms Makungu Clarian.
clarian.makungu@strathmore.edu

Dear Ms Makungu,

RE: A Distributed Computing Prototype for Climate Change Impact Simulation: Case of Nairobi, Kenya

This is to inform you that SU-ISERC has reviewed and **approved** your above **SU-masters** proposal. Your application reference number is **SU-ISERC2602/25.** The approval period is from **27th February 2025 to 26th February 2026.**

This approval is subject to compliance with the following requirements:

i. Only approved documents including (informed consents, study instruments, MTA) will be used.
ii. All changes including (amendments, deviations, and violations) are submitted for review and approval by SU-ISERC.
iii. Death and life-threatening problems and serious adverse events or unexpected adverse events whether related or unrelated to the study must be reported to SU-ISERC within 72 hours of notification.
iv. Any changes anticipated or otherwise that may increase the risks or affected safety or welfare of study participants and others or affect the integrity of the research must be reported to SU-ISERC within 72 hours.
v. Clearance for the export of biological specimens must be obtained from relevant institutions.
vi. Submission of a request for renewal of approval at least 60 days prior to the expiry of the approval period. Attach a comprehensive progress report to support the renewal.
vii. Submission of an executive summary report within 90 days of completion of the study to SU-ISERC.

Before commencing your study, you will be expected to obtain a research license from National Commission for Science, Technology, and Innovation (NACOSTI) https://research-portal.nacosti.go.ke/ and obtain other clearances needed.

Yours sincerely,

Mr Ambrose Rachier,
Chairperson; SU-ISERC

Ole Sangale Rd, Madaraka Estate. PO Box 59857-00200, Nairobi, Kenya. Tel +254 (0)703 034000
Email admissions@strathmore.edu www.strathmore.edu

Appendix C: Survey Instruments

Part I: Demographic Information

Please provide the following information about yourself:

1. Professional background:
    a. Climate scientist
    b. Urban planner
    c. IT/Computing professional
    d. Policy maker
    e. Other (please specify): ________________
2. Years of experience in your field:
    a. Less than 5 years
    b. 5-10 years
    c. 11-15 years
    d. More than 15 years
3. Current organization type:
    a. Academic/Research institution
    b. Government agency
    c. Non-governmental organization
    d. Private sector
    e. Other (please specify): ________________
4. Level of familiarity with climate modeling:
    a. No knowledge
    b. Basic knowledge
    c. Intermediate knowledge
    d. Advanced knowledge
    e. Expert
5. Level of familiarity with distributed computing:
    a. No knowledge
    b. Basic knowledge
    c. Intermediate knowledge

d. Advanced knowledge
e. Expert

Part II: Architecture Design For Urban Climate Simulation

Based on Research Question 1: How can a modular and scalable distributed computing architecture be designed to address computational limitations in urban climate impact simulation?

Please rate your agreement with the following statements on a scale of 1-5: (1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree)

Statement	**1**	**2**	**3**	**4**	**5**
Current computational methods are insufficient for high-resolution climate modeling in Nairobi.					
A modular architecture would improve the adaptability of climate models for urban environments.					
Microservices-based design is appropriate for climate modeling applications.					
Cloud-based deployment is preferable to on-premises infrastructure for climate modeling in developing countries.					
Containerization technology (e.g., Docker) is essential for ensuring consistent deployment across computing environments.					

6. Which of the following components do you consider most critical in a distributed architecture for climate modeling? (Select up to three)
    a. Data ingestion services
    b. Parallel processing engines
    c. Load balancing mechanisms
    d. Storage optimization
    e. Fault tolerance systems
    f. Visualization services
    g. API interfaces
    h. Security frameworks
7. What do you see as the primary technical challenge in designing a distributed computing architecture for climate modeling in Nairobi? (Select one)
    a. Limited network infrastructure
    b. Power reliability issues

c. Integration with existing systems
d. Data security concerns
e. Technical expertise availability
f. Cost of implementation
g. Other (please specify): ________________

Part III: Processing Complex Climate Datasets

Based on Research Question 2: What features are necessary for a prototype system to efficiently process complex climate datasets for Nairobi?

Please rate the importance of the following features for processing climate datasets on a scale of 1-5: (1 = Not Important, 2 = Slightly Important, 3 = Moderately Important, 4 = Important, 5 = Very Important)

Feature	**1**	**2**	**3**	**4**	**5**
Support for multiple data formats (NetCDF, GeoTIFF, CSV, etc.)					
Automated data quality checking					
Data preprocessing capabilities					
Integration with remote sensing data					
High-performance computing optimization					
Ability to handle time-series data efficiently					
Support for geospatial operations					
Real-time data ingestion capabilities					

6. What is the minimum spatial resolution necessary for meaningful urban climate modeling in Nairobi?
   a. >10 km
   b. 5-10 km
   c. 1-5 km
   d. 500m-1km
   e. <500m
7. What volume of climate data would you expect the system to process for a typical urban simulation?
   a. <10 GB

b. 10-100 GB
c. 100-500 GB
d. 500 GB-1 TB
e. >1 TB

Part IV: Parallel Processing Framework

Based on Research Question 3: How can a framework for parallel processing be developed to enhance the performance of climate simulation models?

Please rate your agreement with the following statements on a scale of 1-5: (1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree)

Statement	**1**	**2**	**3**	**4**	**5**
Apache Spark is an appropriate technology for climate data processing.					
Task-based parallelism is more suitable than data-based parallelism for climate simulations.					
Dynamic resource allocation is essential for efficient climate modeling.					
Checkpointing mechanisms are critical for long-running climate simulations.					
A hybrid cloud approach provides the best balance of cost and performance.					

6. Which distributed computing framework do you believe is most appropriate for climate modeling applications?
   a. Apache Hadoop
   b. Apache Spark
   c. Dask
   d. Ray
   e. MPI-based frameworks
   f. Cloud-specific services (AWS EMR, Google Dataproc, etc.)
   g. Other (please specify): ________________
7. What processing time would you consider acceptable for a high-resolution climate simulation covering Nairobi?
   a. <1 hour
   b. 1-6 hours
   c. 6-24 hours
   d. 1-3 days

e. >3 days

Part V: Potential For Urban Climate Impact Analysis

Based on Research Question 4: To what extent can the prototype demonstrate its potential for urban climate impact analysis in Nairobi's context?

Please rate the importance of the following capabilities for urban climate impact analysis on a scale of 1-5: (1 = Not Important, 2 = Slightly Important, 3 = Moderately Important, 4 = Important, 5 = Very Important)

Capability	**1**	**2**	**3**	**4**	**5**
Urban heat island effect simulation					
Flood risk assessment					
Air quality modeling					
Water resource impact analysis					
Energy demand forecasting					
Infrastructure vulnerability assessment					
Socioeconomic impact evaluation					
Interactive visualization of climate scenarios					

6. Which urban sectors in Nairobi would benefit most from high-resolution climate impact modeling? (Select up to three)
   a. Urban planning and zoning
   b. Water resources management
   c. Public health
   d. Transportation infrastructure
   e. Energy systems
   f. Informal settlements
   g. Green space planning
   h. Emergency response
7. How likely would you be to use or recommend this system for climate impact analysis if it were available?
   a. Very unlikely
   b. Unlikely
   c. Neutral

d. Likely
e. Very likely

8. What would you consider the most important output from this system for urban planning purposes? (Select one)
   a. High-resolution temperature projections
   b. Precipitation pattern changes
   c. Extreme weather event forecasting
   d. Sector-specific vulnerability assessments
   e. Adaptation strategy recommendations
   f. Economic impact estimates
   g. Other (please specify): ________________

Part VI: Additional Feedback

15. What additional features would make this distributed computing prototype more valuable for climate change impact simulation in Nairobi?

____

____

16. What barriers do you foresee in implementing and adopting such a system for urban planning in Nairobi?

____

____

Thank you for completing this survey. Your input will help improve the development of distributed computing solutions for climate change impact simulation in urban environments.