# BMJ Open Paper: RP2 Leveraging Data Science and Machine Learning for Urban Climate Adaptation in Africa, the Heat Center Study Protocol

# Response to Reviewers:

## Editor comments

|  |  |
| --- | --- |
| Reviewer | Editor |
| Comment | Please revise the ‘Strengths and limitations of this study’ section of your manuscript (after the abstract). This section should contain up to five short bullet points, no longer than one sentence each, that relate specifically to the methods. The novelty, aims, results or expected impact of the study should not be summarised here. |
| Response | Acknowledged. We have revised the section to focus solely on the methods. |
| Action Taken | Revised the section to include five bullet points focusing on the methods. |

|  |  |
| --- | --- |
| Reviewer | Editor |
| Comment | Please include the planned start and end dates for the study in the methods section. |
| Response | Acknowledged. We have added the start and end dates in the methods section. |
| Action Taken | Added the start and end dates in the methods section. |

|  |  |
| --- | --- |
| Reviewer | Editor |
| Comment | We have noticed that you have used a map in figure 2. As BMJ Open publishes material under a creative commons licence, it is problematic to include copyrighted material, and we would therefore ask that if you own the copyright for this map and if not for it to be removed and/or replaced. |
| Response | Acknowledged. We have removed the map as one of the figures |
| Action Taken | Acknowledged. We have removed the map as one of the figures |

## Reviewer 1 Comments

|  |  |
| --- | --- |
| Reviewer | Reviewer 1 |
| Comment | - 2 papers discussing heat-related health risks distribution among different socio-economic groups in Johannesburg are https://www.sciencedirect.com/science/article/pii/S0169204620303947 & https://www.sciencedirect.com/science/article/pii/S2212095522002498. They could be referenced in a few instances in the paper and offer latest insights. |
| Response | Thank you for the suggestion. We have added references to these papers in the relevant sections. |
| Action Taken | Added references to the suggested papers. |

|  |  |
| --- | --- |
| Reviewer | Reviewer 1 |
| Comment | To characterize heatwaves, the authors limit themselves to temperatures. However, in general, heat stress experienced by humans involves a combination of temperature, humidity, and radiation, such as encompassed in variables as the Wet Bulb Globe Temperature or Heat Index. I would advise the authors to look at these variables and not only temperature. |
| Response | We appreciate your suggestion. Our study primarily focuses on temperature due to data availability and granularity. We acknowledge this as a limitation and will attempt to derive Heat Index values where possible. We also plan to validate different heatwave metrics, considering granularity as a priority. This limitation and our approach to addressing it will be discussed in our study. |
| Action Taken | Acknowledged the limitation and plan to derive Heat Index values where possible. |

|  |  |
| --- | --- |
| Reviewer | Reviewer 1 |
| Comment | - With respect to a heat health early warning system in a developing country, I can refer to the heat action plan in Ahmedabad (India), which offers detailed actions in case of high heat stress and the interactions with different institutional instances https://www.nrdc.org/sites/default/files/ahmedabad-heat-action-plan-2019-update.pdf. Maybe some inspiration can be obtained from their approaches. |
| Response | Thank you for the reference. We have incorporated insights from the Ahmedabad heat action plan into our study. |
| Action Taken | Added information on the Ahmedabad heat action plan and its relevance to our study in the description of the Early Warning System |

|  |  |
| --- | --- |
| Reviewer | Reviewer 1 |
| Comment | P7 L18: it must be noted that land surface temperatures are decoupled from heat stress & are probably not the best tool to measure or quantify heat stress, see e.g. https://ams.confex.com/ams/103ANNUAL/meetingapp.cgi/Paper/412571. It would be better to use a combination of temperatures and humidity. |
| Response | Thank you for your insightful comment. We acknowledge that land surface temperature (LST) may not fully capture the complexity of heat stress experienced in urban areas. In our study, while we utilize LST data from Landsat and MODIS for urban land use analysis, we will employ statistical models to estimate air temperature from remotely sensed LST and incorporate humidity data where available to provide a more comprehensive assessment of heat stress. However, we recognize the limitation that LST alone might not accurately reflect the intra-urban distribution of heat stress or the effectiveness of urban vegetation in mitigating heat stress. This limitation will be discussed in our study, emphasizing the need for more urban-scale observations and better urban-resolving models to accurately inform heat mitigation strategies within and across cities. |
| Action Taken | Modified the manuscript to acknowledge the limitation of relying solely on LST for assessing heat stress and clarified the use of statistical models to estimate air temperature and incorporate humidity data where possible. |

|  |  |
| --- | --- |
| Reviewer | Reviewer 1 |
| Comment | - P7 L23: It is not clear where the ‘geospatial climate’ data will come from yet from the paper. The data that is mentioned (CDS & ESGF) is not spatially detailed enough to distinguish different parts within the city? Recently some studies with WRF & UrbClim offer very detailed results on heat stress for cities that could potentially be used. |
| Response | We have provided additional details on the sources and spatial resolution of the geospatial climate data used in our study. |
| Action Taken | Specified the sources and resolution of the geospatial climate data in the methods section. To address the limitation of spatial resolution in climate data from the Copernicus Climate Data Store (CDS) and Earth System Grid Federation (ESGF), we will employ downscaling techniques using high-resolution climate models such as the Weather Research and Forecasting (WRF) model and the UrbClim urban climate model. These models will enhance the spatial detail of our geospatial climate data, allowing for a more precise analysis of intra-urban heat variations and improving the accuracy of our heat risk assessments for Johannesburg and Abidjan. |
| Reviewer | Reviewer 1 |
| Comment | - P9 L53: the following datasets might be of interest as input for the study: https://viewer.esa-worldcover.org/worldcover/ & https://ghsl.jrc.ec.europa.eu/datasets.php; |
| Response | Thank you for the suggestion. We have reviewed the suggested datasets and incorporated relevant data into our study. |
| Action Taken | Added information on the incorporation of the suggested datasets in the Methods section. |

|  |  |
| --- | --- |
| Reviewer | Reviewer 1 |
| Comment | - P10 L37: again, it must be noted that Landsat and MODIS are land surface temperature tools and should not directly be used for air temperature retrieval (see above) |
| Response | Acknowledged. We have addressed this concern by clarifying the use of Landsat and MODIS data in our study. |
| Action Taken | Added this text: For instance, Gudmundsson and Seneviratne (2021) demonstrated the use of such models to predict air temperature from LST data. Therefore, while Landsat and MODIS data are not direct measures of air temperature, they can be indirectly used for air temperature retrieval through the application of appropriate statistical models. |

## Reviewer 2 Comments

|  |  |
| --- | --- |
| Reviewer | Reviewer 2 |
| Comment | Improving the structure of the paper |
| Response | Acknowledged. We have improved the structure by adding numbered sections and subsections. |
| Action Taken | Improved the structure of the document with numbered sections and subsections. |

|  |  |
| --- | --- |
| Reviewer | Reviewer 2 |
| Comment | Deepening the introduction |
| Response | Acknowledged. We have expanded the introduction to provide a more comprehensive overview of past work in Africa and worldwide, and to highlight the unique contributions of our study. |
| Action Taken | In response to the reviewer's comments, the introduction has been expanded to provide a more comprehensive background on the heat-health challenges faced in African cities, particularly Abidjan and Johannesburg. The revised introduction now includes a discussion of previous studies on heat-health impacts across Africa, with specific references to research conducted in the cities of interest. It integrates the concepts of exposure, vulnerability, hazard, and adaptive capacity in the context of heat exposure studies as defined by the IPCC framework. Additionally, the introduction better articulates the study's contribution in comparison to existing literature by emphasizing the development of an Early Warning System using data science and machine learning innovations, tailored for urban African contexts. This approach is aimed at building climate-resilient cities and protecting vulnerable populations from heat hazards. The new text effectively addresses the need for a longer, more detailed introduction and sets the stage for the study's objectives and methods. |

|  |  |
| --- | --- |
| Reviewer | Reviewer 2 |
| Comment | Discussing the reliability/availability of clinical trial datasets for heat-health studies and considering administrative health data |
| Response | Acknowledged. We have provided more details on contingency plans if we cannot access enough clinical data |
| Action Taken | Added details on the contingency plans we will take if insufficient data is available for our analysis after integration of datasets |

|  |  |
| --- | --- |
| Reviewer | Reviewer 2 |
| Comment | Motivating the choice of the models, their advantages/drawbacks and how they will be validated |
| Response | Acknowledged. We have provided a more detailed explanation of the machine learning models employed in our study, including their advantages and how they will be validated. |
| Action Taken | Expanded the description of machine learning models and their validation in the methods section. |

|  |  |
| --- | --- |
| Reviewer | Reviewer 2 |
| Comment | P2 L22 : How will data during the 2020-2022 COVID-19 pandemic will be handled ? There were some major shifts/changes in mortality/morbidity during that period. Please add this information in the “Data description section”. |
| Response | We have outlined our approach to handling data during the COVID-19 pandemic, ensuring the robustness of our analysis. |
| Action Taken | Added information on handling COVID-19 pandemic data in the Methods section. In response to the shifts in mortality and morbidity during the 2020-2022 COVID-19 pandemic, we will analyze data separately for pre-pandemic, pandemic, and post-pandemic periods. Additionally, we will include COVID-19 related variables as covariates in our models to control for the pandemic's impact on health outcomes |

|  |  |
| --- | --- |
| Reviewer | Reviewer 2 |
| Comment | P4 L19 : Elaborate on the different socio-demographic groups affected by heat. |
| Response | We have expanded our discussion on the various socio-demographic groups affected by heat and their respective vulnerabilities. |
| Action Taken | Elaborated on the different socio-demographic groups affected by heat in the Introduction section. |

|  |  |
| --- | --- |
| Reviewer | Reviewer 2 |
| Comment | P4 L47 : It would be interesting to add some contexts why these two cities were selected for this study. By reading the title, we can think the study will be over the entire African continent, but it is limited to two cities. Is it because they have health data that other cities do not have ? Do they have higher population size compared to other cities ? |
| Response | We have provided a detailed rationale for selecting Johannesburg and Abidjan as the focus cities for our study. |
| Action Taken | Added context on the selection of the two cities in the Study Setting section: Abidjan and Johannesburg were selected for this study due to their unique characteristics and availability of data. Both cities have a high population density and are experiencing rapid urbanization, making them representative of the challenges faced by many African cities in the context of climate change and heat-related health impacts. Additionally, these cities have access to detailed health data from clinical trials and cohort studies, which is crucial for our analysis. The selection of these cities allows for a focused examination of heat-related health risks in urban African settings, with the potential to inform broader regional strategies for climate adaptation and public health. |

|  |  |
| --- | --- |
| Reviewer | Reviewer 2 |
| Comment | P4 L55 : The background/rationale section should be presented before the “Study setting”. Please provide numbered sections (1, 2, 3) and subsections (2.1, 2.2.) to ease understanding the structure of the document. In addition, I think “Study setting” should belong to the “Method” section. (see Major Comment #1) |
| Response | Acknowledged. We have restructured the manuscript to present the background/rationale section before the study setting. |
| Action Taken | Restructured the manuscript to present the Background/Rationale section before the Study Setting section. |

|  |  |
| --- | --- |
| Reviewer | Reviewer 2 |
| Comment | P5 L3-21: How do these definitions relate to the hazard-vulnerability-exposure framework of IPCC ? If the framework is different than IPCC, please explain why and cite the appropriate references. |
| Response | We have clarified how our definitions relate to the hazard-vulnerability-exposure framework of the IPCC. |
| Action Taken | Clarified the relationship between our definitions and the IPCC framework in the Background/Rationale section: The definitions used in this study align with the hazard-vulnerability-exposure framework of the Intergovernmental Panel on Climate Change (IPCC). According to the IPCC, 'hazard' refers to the potential occurrence of a natural or human-induced physical event that may cause harm to human health, livelihoods, assets, or ecosystems. 'Vulnerability' is defined as the propensity or predisposition to be adversely affected, which encompasses a variety of factors including sensitivity and adaptive capacity. 'Exposure' refers to the presence of people, livelihoods, species, or ecosystems in places that could be adversely affected by the hazard.  In our study, we apply these concepts to the context of urban heat stress, where the 'hazard' is the occurrence of high temperatures, 'vulnerability' includes factors such as age, health status, and access to cooling resources, and 'exposure' relates to the degree of contact individuals have with the heat hazard. Our use of these terms is consistent with the IPCC framework, which provides a comprehensive approach to understanding and addressing the impacts of climate-related hazards on various socio-demographic groups. |

|  |  |
| --- | --- |
| Reviewer | Reviewer 2 |
| Comment | P5 L41-L44 : This should be mentioned in the Introduction while adding past studies conducted in Africa. (see Major Comment #2) |
| Response | Acknowledged. We have incorporated this information into the Introduction, along with references to past studies conducted in Africa. |
| Action Taken | Incorporated information on past studies conducted in Africa into the Introduction section. |

|  |  |
| --- | --- |
| Reviewer | Reviewer 2 |
| Comment | P5 L49 : This section may be included as the final paragraph of a broader Introduction. (see Major Comment #2) |
| Response | Acknowledged. We have integrated this section into the final paragraph of the Introduction to provide a more cohesive overview. |
| Action Taken | Integrated the section into the final paragraph of the Introduction. |

|  |  |
| --- | --- |
| Reviewer | Reviewer 2 |
| Comment | P6 L38 : Using numbered sections and subsections will help the reader understand the structure of the paper. (see Major Comment #1) |
| Response | Acknowledged. We have implemented numbered sections and subsections throughout the paper to enhance readability. |
| Action Taken | Implemented numbered sections and subsections throughout the paper. |

|  |  |
| --- | --- |
| Reviewer | Reviewer 2 |
| Comment | P6 L40 : A summary table with all data sources (including health data) related to each of the three objectives would be useful here. |
| Response | Thank you for the suggestion. We have included a summary table listing all data sources related to each of the three objectives. |
| Action Taken | Included a summary table listing all data sources in the Methods section. |

|  |  |
| --- | --- |
| Reviewer | Reviewer 2 |
| Comment | P6 L50-L51 : I am curious if there is any administrative health database available (e.g., mortality, health services use). These are often the types of datasets used for heat-health studies. (See Major Comment #3) |
| Response | We have provided information on the availability and use of administrative health databases in our study. |
| Action Taken | These datasets are available but vary in quality, and therefore, we care about doing our best to use clinical trial or robust clinical trial data where possible. Contingency plans have been outlined if twe cannot gather sufficient data from the clinical trials and cohort data |

|  |  |
| --- | --- |
| Reviewer | Reviewer 2 |
| Comment | P7 L10 and L14: What about the other city? |
| Response | We have ensured that both cities, Johannesburg and Abidjan, are covered in terms of socio-economic data |
| Action Taken | We have ensured that both cities, Johannesburg and Abidjan, are covered in terms of socio-economic data |

|  |  |
| --- | --- |
| Reviewer | Reviewer 2 |
| Comment | P7 L38-39 : If health trials and cohort data are used, they need to be introduced in the Introduction and how these can be used for heat-health studies. (see Major Comment #3) |
| Response | Acknowledged. We have introduced the use of health trials and cohort data in the Introduction section. |
| Action Taken | Introduced the use of health trials and cohort data in the Introduction section. |

|  |  |
| --- | --- |
| Reviewer | Reviewer 2 |
| Comment | P8 L37-52 : I understand the point of looking for available datasets, but what if the trials and cohort data are not available/shareable for the current project ? Does this pose a risk to the achievability of the further steps ? (see Major Comment #3) |
| Response | We have addressed this concern by outlining our contingency plans and strategies for data access and availability. |
| Action Taken | Outlined contingency plans and strategies for data access in the Methods section. |

|  |  |
| --- | --- |
| Reviewer | Reviewer 2 |
| Comment | P9 L47-55 : This is somewhat of a repetition of what was introduced in the Data section of the method. A better structure of the paper will help not repeating information in different sections. (see Major Comment #1) |
| Response | Acknowledged. We have revised the manuscript to eliminate redundancy and ensure a clear presentation of information. |
| Action Taken | Revised the manuscript to eliminate redundancy in the presentation of the Data section. |

|  |  |
| --- | --- |
| Reviewer | Reviewer 2 |
| Comment | P9 L44: Please add “First objective” to this section name. |
| Response | Acknowledged. We have added "First Objective" to the section name to clearly delineate the objectives. |
| Action Taken | Added "First Objective" to the section name in the Methods section. |

|  |  |
| --- | --- |
| Reviewer | Reviewer 2 |
| Comment | P9 L58 : Please elaborate on why only PCA will be used. What are the advantages/drawbacks of this method ? (See Major Comment #4) |
| Response | We have elaborated on the rationale for using PCA and its advantages in the context of our study. |
| Action Taken | Elaborated on the use of PCA in the Methods section. |

|  |  |
| --- | --- |
| Reviewer | Reviewer 2 |
| Comment | P10 L7-10: Please add “Second objective” to this section name. |
| Response | Acknowledged. We have added "Second Objective" to the section name to clearly delineate the objectives. |
| Action Taken | Added "Second Objective" to the section name in the Methods section. |

|  |  |
| --- | --- |
| Reviewer | Reviewer 2 |
| Comment | P11 L3-4 : More details should be given related to the cross-validation and how models will be trained/calibrated/validated. (see Major Comment #4) |
| Response | We have provided additional details on the cross-validation process and the training, calibration, and validation of models. |
| Action Taken | Provided detailed information on the cross-validation process and model validation in the Methods section. |

|  |  |
| --- | --- |
| Reviewer | Reviewer 2 |
| Comment | P11 L9 : References 53 and 56 are out-of-scope for the proposed application. In addition, Ke et al. used an XGBoost model, which is not considered in the deep learning family. I recommend that the authors look at up-to-date literature for the use of deep learning applications in heat-health studies. (see Major Comment #4) |
| Response | Acknowledged. We have reviewed and updated the references to ensure they are relevant to the proposed application. |
| Action Taken | Reviewed and updated references to ensure relevance to the proposed application in the References section. |

|  |  |
| --- | --- |
| Reviewer | Reviewer 2 |
| Comment | P12 L8-L11: Add “Objective 3” to this section title. |
| Response | Acknowledged. We have added "Objective 3" to the section title to clearly delineate the objectives. |
| Action Taken | Added "Objective 3" to the section title in the Methods section. |

|  |  |
| --- | --- |
| Reviewer | Reviewer 2 |
| Comment | P12 L19-L26 : Please develop here how the models developed in objectives I and II can be used jointly for the third objective. The maps of the first objective will be at much higher resolution than the forecast model of objective II, so how these divergent resolutions be handled for objective III ? (see Major Comment #4) |
| Response | We have provided a detailed explanation of how the models developed in Objectives I and II are integrated and used jointly for Objective III. |
| Action Taken | Provided a detailed explanation of model integration for Objective III in the Methods section. The EWS integrates high-resolution heat hazard maps and a forecast model to generate alerts for areas with predicted adverse heat-health outcomes. This integration involves: 1. Downscaling the Forecast Model: Adjusting the model's resolution to match the detailed heat hazard maps. 2. Combining Data Sources: Merging the forecast model outputs with the heat hazard maps to create a comprehensive dataset. 3. Generating Alerts: Using the dataset to generate timely alerts for specific areas at risk. Additionally, the EWS incorporates heat hazard predictions to enable proactive risk management. It offers tailored guidance for at-risk individuals, suggesting strategies like adequate hydration and activity scheduling. Drawing inspiration from the Ahmedabad Heat Action Plan, our EWS will emphasize inter-agency coordination and community outreach for effective heat risk mitigation[87]. While our EWS aims to provide advanced warnings, we acknowledge the challenges of long-term forecasting. The accuracy of predictions depends on data reliability, model complexity, and weather variability. Continuous model refinement is essential for improving predictive capabilities. |

|  |  |
| --- | --- |
| Reviewer | Reviewer 2 |
| Comment | P12 L31 : Projection weeks in advance seems optimistic. Please nuance. |
| Response | We have nuanced our language to reflect the realistic expectations and limitations of projection weeks in advance. |
| Action Taken | We have nuanced our language to reflect the realistic expectations and limitations of projection weeks in advance. |