



UGANDA CHRISTIAN UNIVERSITY

A Centre of Excellence in the Heart of Africa

SCHOOL OF RESEARCH & POSTGRADUATE STUDIES DISSERTATION CORRECTION COMPLIANCE FORM (POST VIVA FORM)

Date: 14/09/2026

Name of Candidate: LORDIN LUTALO

Reg.No: J25M19/004

Title of Dissertation: Machine Learning Forecasting for Electricity Demand Using Climate and Socioeconomic Data in Sub-Saharan Africa

S/N	COMMENTS BY EXTERNAL EXAMINER	ACTION TAKEN	INDICATOR
1	Revise the general objective so that it expresses the actual research purpose rather than the generic process of training and evaluating a machine-learning model.	The general objective was rewritten to state the purpose of the research rather than the modelling process. It now reads: "To establish how far integrated climate and socioeconomic data can improve annual electricity demand forecasting in Sub-Saharan Africa, and to make the resulting forecasts and demand drivers available to energy planners through a decision-support tool." The three specific objectives and the three research questions were rephrased in the same way, so that each states an outcome instead of a task. Specific Objective 2 now names the three-way comparison between the lag-only, exogenous-only and multivariate feature sets which is the comparison the study actually measures. The Abstract, Executive Summary, Chapter 3 introduction and Chapter 6 summary were all updated so that the wording of the objective is consistent throughout the report.	Abstract, Executive Summary, Page 3: Sections 1.3, 1.4, 3.1, 6.2 and 6.3
2	Strengthen the research framing by presenting the background from global, regional and national	The Background of the Study was rewritten and now moves through three levels in order. The global level covers the rise in world electricity use. The regional level	Sections 1.1 and 1.5

	perspectives and linking the study to relevant development priorities.	covers Sub-Saharan Africa. The national level uses Uganda as a worked example, citing the Energy Policy for Uganda. A paragraph that links the study to SDG 7, SDG 13 and African Union Agenda 2063, was added	
3	Improve the methodology chapter by clearly distinguishing the research approach from the research design.	The two were separated into stand-alone sections. Section 3.2 covers the research approach, including its quantitative, deductive, and predictive orientation, and explains why this approach was appropriate for the study while Section 3.3 covers the research design, including the panel structure, unit of analysis, comparative modelling approach, and chronological data splitting strategy.	Sections 3.2 and 3.3
4	Strengthen the literature review by explicitly connecting previous research to the study's rationale and research gap.	Each of the five gaps in the Research Gap section now names the specific studies it rests on, states what those studies could not do, and states how this study responds. The chapter summary was extended to state which part of the study design each gap shaped.	Sections 2.8 and 2.9
5	Expand the discussion of findings, particularly by explaining the significance and implications of the results rather than merely presenting them.	A new section, Significance of the Findings, was added to Chapter 5. It highlights three major findings: the importance of the lag-only baseline, the strong role of structural variables in long-term planning, and the different temperature-demand relationship in Sub-Saharan Africa. A chapter summary was added, which Chapter 5 previously didn't have and closing paragraphs on practical consequence were added to the sections on lag dominance and on population as a structural driver.	Sections 5.2, 5.3, 5.12, Sec. 5.13 and 5.15.

S/N	COMMENTS BY INTERNAL EXAMINER	ACTION TAKEN	INDICATOR
1			
2			
3			

S/N	COMMENTS BY VIVA VOCE PANNEL	ACTION TAKEN	INDICATOR
1	How did you arrive at the features used for this study? Why were only 7 features chosen out of the 27 features?	A feature selection table was added to Section 3.7, clearly explaining the feature selection process from 27 engineered features to the final 7 features used in Model C. Two raw log variables were replaced by their standardised versions, leaving 25 features. These were separated into 2 lag features and 23 exogenous features because the lag features had very strong correlations with demand ($r = 0.999$) and dominated the ranking methods. Four exogenous features were then removed because of multicollinearity above 0.75, leaving 19 features. The top 5 exogenous features were selected and combined with the 2 lag features to form the final 7 features for Model C. The section also explains that selecting five exogenous features was a pre-defined decision to limit the number of features, rather than simply choosing a cutoff from the results. The sixth and seventh features had lower scores (0.420 and 0.399) compared with 0.444 for the fifth feature, supporting the decision to stop at five.	Sec. 3.7, Table 3.3 (Feature Selection Funnel) and Sec. 4.6
2	The research objectives should be rephrased.	The general objective and all three specific objectives were rewritten to state outcomes rather than processes, and the three research questions were rephrased to match. This was addressed together with the external examiner's first comment.	Abstract, Executive Summary, Sections 1.3, 1.4, 3.1, 6.2 and 6.3
3	Why did you choose these models over others?	Section 3.8 now explains the reason for selecting the six modelling approaches and groups them into four model families: Linear Regression, panel models, Random Forest, and boosting models. It also explains why ARIMA, SARIMA, LSTM/deep learning, Prophet, and Support Vector Regression were not used, mainly due to the small annual dataset, limited ability to use climate and socioeconomic predictors, lack of interpretability, and limited added value over the selected models.	Section 3.8

4	How did you do feature selection in your pipeline?	The full two-stage procedure is set out in Section 3.7 and its results in Section 4.6. Stage 1 evaluated the two lag features against each other while Stage 2 set them aside and applied five methods (Pearson and Spearman correlation, mutual information, RFECV, Random Forest importance and XGBoost importance) to the exogenous features alone, after a multicollinearity screen at an absolute correlation of 0.75. Scores from each method were normalised to a 0 to 1 scale and averaged into a single consolidated ranking.	Sections 3.7, 3.11, 4.6
5	How did you do the data splits?	Two tables were added to show the data splits and walk-forward validation process. The modelling data covers 2002-2021, with 747 rows for training and validation and 188 rows for final testing. The section also explains why 2002 was used as the starting year, since the lag and rolling features required earlier years. The removed early rows were not imputed. A transparency note was also added to clarify that the final model was refitted on the full period for the dashboard, while all reported accuracy results were based only on the training data.	Sections 3.11, Tables 3.4 and 3.5
6	Who is the intended user of the dashboard developed in this research?	Section 3.12 now identifies one primary user: an energy planning officer in a national ministry, utility, or regulatory authority. The dashboard is designed for users with policy or engineering backgrounds, so results are shown in terawatt-hours and the model is pre-trained. Two secondary users are also identified: policy analysts for cross-country comparison and researchers for model inspection. The section also states two main limits: the dashboard does not forecast beyond 2021 and is intended for national annual planning, not operational use.	Sections 3.12 and 4.10

7	Merge the variables from the different dashboards to ensure that they are the same.	Acknowledged and noted in the report. The Overview page uses fewer pooled indicators than the Country Explorer page to keep regional trends clear and easy to read. This was a deliberate design choice, and aligning the two indicator sets is planned as a future improvement.	Sections 3.12 and 4.10
8	Clearly indicate the importance of the climate and socioeconomic variables to the forecasting model, since Model A which uses lags had similar performance.	A new section was added in Chapter 5, and it separates the predictive and structural contributions of climate and socioeconomic variables. The findings show that the five exogenous variables add almost nothing to one-year-ahead prediction when lagged demand is available, but they explain about 91% of the variation in log demand when past demand is excluded. This means the variables are structurally important but become redundant for short-term forecasting because lagged demand already captures a country's structural position. A table was added setting out six planning questions and which models can answer them: Model A answers one of the six, because it cannot be used beyond a one-year horizon, cannot explain differences between countries, and contains no policy lever. The earlier justification for selecting Model C, which rested partly on the study's original intent, was replaced with this evidence-based argument.	Sections 5.9, 5.8, Sec. 5.6 Table 5.1 (Planning Questions Addressed by the Models)

Candidate's Name: **LORDIN LUTALO**

Signature: 

Supervisor's Name

Signature: 

Dr. Terhemba Michael Ahile

Mr. Senabulya Brian

Signature: 

NB: Post Viva compliance form is designed to capture all the corrections recommended by internal examiner (supervisor), external examiner and viva panel.