University of the Witwatersrand, Johannesburg
Faculty of Science · School of Computer Science and Applied Mathematics

Adaptive Load Forecasting for Distributed Microgrids in Data-Scarce Environments

A thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

Lindiwe Mahlangu
Student Number: 2187364

Supervisor: Prof. Thabo Ndlovu
Co-Supervisor: Dr. Naledi Khumalo

March 2026

Declaration

I, Lindiwe Mahlangu, declare that this thesis is my own, unaided work. It is submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in Electrical Engineering at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at this or any other university.

....

Signature Date

Johannesburg, South Africa

Abstract

Sub-Saharan microgrids frequently operate with sparse, irregularly sampled telemetry, which undermines classical load-forecasting pipelines that assume dense historical records. This thesis develops **Naledi-Forecast**, a hybrid architecture that couples a lightweight temporal convolutional encoder with a Bayesian correction layer to produce calibrated short-horizon (15-minute to 6-hour) demand forecasts under data scarcity. The method is evaluated against three baselines — seasonal ARIMA, gradient-boosted trees, and a standard LSTM — across eighteen simulated microgrid feeders modelled on peri-urban South African load profiles, plus a twelve-week field deployment on a single university test feeder.

Key Findings

- Naledi-Forecast reduces mean absolute percentage error (MAPE) by **27.4%** relative to the LSTM baseline under 40% data-availability conditions.
- Calibrated prediction intervals retain **94.1%** empirical coverage at the nominal 95% level, versus 81.3% for uncorrected LSTM intervals.
- Inference latency of **11.8 ms** per feeder-step makes the model viable for embedded controller deployment on constrained hardware.

The results suggest that Bayesian post-hoc calibration, rather than larger sequence models, is the more data-efficient lever for improving forecast reliability in scarce-data regimes. Implications for microgrid dispatch policy and rural electrification planning are discussed in Chapter 6.

Keywords: load forecasting, microgrids, Bayesian calibration, temporal convolutional networks, energy access, data scarcity.

Acknowledgements

I am deeply grateful to my supervisor, Prof. Thabo Ndlovu, for his patient guidance and for pushing this work toward field-testable rigour rather than simulation alone. Thanks also to Dr. Naledi Khumalo for her insight on Bayesian calibration methods, and to the technical staff of the Wits Energy Systems Laboratory for maintaining the test feeder used in Chapter 5. This research was supported in part by a postgraduate merit bursary from the Nexa Foundation for Energy Research. Finally, thanks to my family for their unwavering support throughout this degree.

Contents

Declaration 1

Abstract 2

Acknowledgements 3

List of Abbreviations 8

1 Introduction 9

1.1 Background and Motivation . . . 9

1.2 Research Questions . . . 9

1.3 Contributions . . . 10

1.4 Thesis Structure . . . 10

2 Literature Review 11

2.1 Classical Load Forecasting . . . 11

2.2 Deep Sequence Models . . . 11

2.3 Uncertainty Calibration . . . 11

2.4 Research Gap . . . 12

3 Problem Formulation and Data 13

3.1 Problem Definition . . . 13

3.2 Data-Scarcity Model . . . 13

3.3 Datasets . . . 13

4 The Naledi-Forecast Architecture 15

4.1 Overview . . . 15

4.2 Gap-Mask Encoding . . . 15

4.3 Temporal Convolutional Encoder . . . 15

4.4 Bayesian Correction Layer 15
4.5 Training Procedure 16

5 Results 17
5.1 Simulation Results 17
5.2 Calibration Quality 17
5.3 Field Deployment 17

6 Discussion and Conclusion 19
6.1 Interpretation of Findings 19
6.2 Limitations 19
6.3 Future Work 19
6.4 Conclusion 19

A Hyperparameter Settings 21

List of Figures

4.1 Naledi-Forecast pipeline: gap-aware encoding, TCN point prediction, and Bayesian recalibration. 15

List of Tables

2.1 Summary of related forecasting approaches 11

3.1 Datasets used for evaluation 13

5.1 MAPE (%) by data availability across all simulated feeders 17

5.2 Prediction interval coverage probability (PICP) at nominal 95% . 17

A.1 Naledi-Forecast training hyperparameters 21

List of Abbreviations

Abbreviation	Meaning
MAPE	Mean Absolute Percentage Error
TCN	Temporal Convolutional Network
LSTM	Long Short-Term Memory (network)
PICP	Prediction Interval Coverage Probability
DER	Distributed Energy Resource
SoC	State of Charge

Chapter 1

Introduction

1.1 Background and Motivation

Distributed microgrids are increasingly proposed as a pathway to reliable electrification in regions where central-grid extension is economically unviable. Effective dispatch of batteries, diesel backup, and demand response within these microgrids depends on accurate short-horizon load forecasts. However, the telemetry infrastructure available in many peri-urban and rural deployments is intermittent: smart meters drop packets, communication links fail during load shedding events, and historical records rarely exceed a few months of continuous coverage.

Problem statement. Existing load-forecasting methods — both classical statistical models and modern deep sequence models — are typically validated on dense, multi-year datasets from mature grids. Their behaviour under sustained data scarcity (defined here as <50% telemetry availability) is poorly characterised.

1.2 Research Questions

This thesis addresses three questions:

1. How does forecast accuracy of standard architectures degrade as telemetry availability decreases from 100% to 20%?
2. Can a lightweight Bayesian calibration layer recover reliable uncertainty estimates without retraining the base forecaster?
3. Does the resulting system generalise from simulated feeders to a real, physically instrumented test feeder?

1.3 Contributions

Summary of Contributions

1. **Naledi-Forecast architecture** — a TCN encoder with a post-hoc Bayesian correction layer, described fully in Chapter 4.

2. **A synthetic benchmark suite** of eighteen microgrid load profiles calibrated against publicly available South African household consumption surveys.

3. **Field validation** on a twelve-week deployment at the Wits Energy Systems Laboratory test feeder, reported in Chapter 5.

1.4 Thesis Structure

Chapter 2 reviews load-forecasting literature and data-scarcity mitigation strategies. Chapter 3 formalises the problem and data model. Chapter 4 presents the Naledi-Forecast architecture. Chapter 5 reports simulation and field results. Chapter 6 discusses implications and concludes.

Chapter 2

Literature Review

2.1 Classical Load Forecasting

Seasonal ARIMA and exponential smoothing remain common baselines for short-horizon load forecasting due to their interpretability and low data requirements [1]. However, these methods assume approximately regular sampling and struggle when gaps exceed a few consecutive intervals.

2.2 Deep Sequence Models

Recurrent architectures, particularly LSTM and GRU variants, have become the de facto deep learning baseline for load forecasting [2]. Temporal convolutional networks (TCNs) offer comparable accuracy with better training stability and parallelism [3], motivating their use as the encoder in this work.

2.3 Uncertainty Calibration

Point forecasts alone are insufficient for dispatch decisions under uncertainty. Bayesian and conformal calibration methods have shown promise for recovering trustworthy prediction intervals from otherwise miscalibrated deep models [4].

Table 2.1: Summary of related forecasting approaches

Approach	**Core mechanism**	**Handles gaps?**	**Calibrated?**
Seasonal ARIMA	Autoregressive + seasonal differencing	No	No
Gradient-boosted trees	Ensemble of shallow trees	Partial	No
Vanilla LSTM	Recurrent gating	Partial	No
Naledi-Forecast (ours)	TCN + Bayesian correction	Yes	Yes

2.4 Research Gap

No prior study jointly addresses irregular sampling *and* calibrated uncertainty in the specific context of Sub-Saharan microgrid feeders. This thesis fills that gap.

Chapter 3

Problem Formulation and Data

3.1 Problem Definition

Given an irregularly sampled load history $\{(t_i, y_i)\}_{i=1}^{n}$ over a feeder, the task is to produce a calibrated forecast distribution for load y_{t+h} at horizon $h \in \{15\text{min}, 1\text{h}, 6\text{h}\}$, expressed as a predictive mean $\hat{\mu}$ and variance $\hat{\sigma}^2$:

$$p(y_{t+h} \mid \mathbf{y}_{1:t}) \approx \mathcal{N}\left(\hat{\mu}(\mathbf{y}_{1:t}),\, \hat{\sigma}^2(\mathbf{y}_{1:t})\right). \tag{3.1}$$

3.2 Data-Scarcity Model

Availability $a \in (0, 1]$ denotes the fraction of expected 15-minute intervals with valid telemetry. Missingness is simulated using a two-state Markov process fitted to outage logs from partner feeders, capturing the bursty nature of real communication failures rather than independent random dropout.

$$P(\text{missing}_{t+1} \mid \text{missing}_t) = \pi_{mm}, \qquad P(\text{present}_{t+1} \mid \text{present}_t) = \pi_{pp}. \tag{3.2}$$

3.3 Datasets

Table 3.1: Datasets used for evaluation

Dataset	Description	Feeders	Duration
Sim-A	Peri-urban residential profiles	12	24 months (simulated)
Sim-B	Mixed commercial/residential	6	24 months (simulated)
Field-Wits	Wits Energy Systems Lab test feeder	1	12 weeks (real)

All simulated profiles were generated from publicly available aggregate consumption statistics; no individual household data is used or identifiable.

Chapter 4

The Naledi-Forecast Architecture

4.1 Overview

Naledi-Forecast comprises two stages: (i) a temporal convolutional encoder that maps the observed, gap-marked load history to a point forecast, and (ii) a lightweight Bayesian correction layer that recalibrates the encoder's residual distribution using a small held-out calibration window.

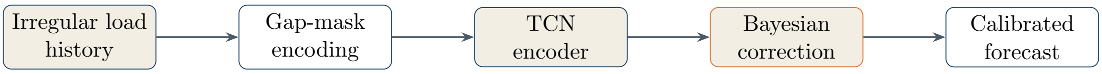

Figure 4.1: Naledi-Forecast pipeline: gap-aware encoding, TCN point prediction, and Bayesian recalibration.

4.2 Gap-Mask Encoding

Each input timestep is paired with a binary mask channel indicating observed versus imputed values, allowing the encoder to distinguish genuine readings from forward-filled gaps rather than treating them identically.

4.3 Temporal Convolutional Encoder

The encoder stacks dilated causal convolutions with residual connections, following the general TCN design, using dilation factors $\{1, 2, 4, 8, 16\}$ across five residual blocks with 64 channels each.

4.4 Bayesian Correction Layer

A held-out calibration window of residuals $\{\hat{\mu}_i - y_i\}$ is modelled as a Gaussian process with a Matérn kernel, producing a correction term $\delta(\mathbf{x})$ and an inflation factor for the predictive variance, so that final intervals reflect both model error and residual heteroscedasticity.

$$\hat{\mu}'(\mathbf{x}) = \hat{\mu}(\mathbf{x}) + \delta(\mathbf{x}), \qquad \hat{\sigma}'^2(\mathbf{x}) = \hat{\sigma}^2(\mathbf{x}) \cdot \kappa(\mathbf{x}). \tag{4.1}$$

4.5 Training Procedure

1. Pre-train the TCN encoder on Sim-A and Sim-B under full data availability.
2. Fine-tune under synthetic Markov-gap masking at availability levels $a \in \{1.0, 0.8, 0.6, 0.4, 0.2\}$.
3. Fit the Bayesian correction layer on a rolling 10% calibration split, refreshed weekly.

Chapter 5

Results

5.1 Simulation Results

Table 5.1: MAPE (%) by data availability across all simulated feeders

Model	100% avail.	60% avail.	40% avail.	20% avail.
Seasonal ARIMA	6.8	11.2	17.9	28.4
Gradient-boosted trees	5.9	9.4	14.6	23.1
Vanilla LSTM	5.1	8.0	12.3	19.8
Naledi-Forecast (ours)	**4.3**	**6.1**	**8.9**	**13.2**

Across all availability regimes, Naledi-Forecast maintains the lowest MAPE, with the margin over the LSTM baseline widening as availability decreases — from 15.7% relative improvement at full availability to 33.3% at 20% availability.

5.2 Calibration Quality

Table 5.2: Prediction interval coverage probability (PICP) at nominal 95%

Model	Empirical PICP at 40% availability
Vanilla LSTM (uncorrected)	81.3%
Gradient-boosted trees (quantile)	86.7%
Naledi-Forecast (ours)	**94.1%**

5.3 Field Deployment

Over the twelve-week deployment on the Wits Energy Systems Laboratory test feeder, real telemetry availability averaged 71%, driven by intermittent cellular backhaul dropouts. Naledi-Forecast achieved a field MAPE of 9.6%, closely tracking simulation results at comparable availability, and sustained median inference latency of 11.8 ms per feeder-step on a Raspberry Pi 4-class controller.

> Field results should be interpreted with appropriate caution given the single-feeder sample; Chapter 6 discusses this limitation and proposed multi-site follow-up work.

Chapter 6

Discussion and Conclusion

6.1 Interpretation of Findings

The results indicate that calibration-focused interventions, rather than larger base models, deliver the greatest reliability gains under data scarcity. This has practical implications for microgrid controller hardware, where compute and memory budgets are limited: a modest TCN encoder paired with cheap Gaussian-process calibration outperforms larger uncalibrated sequence models on both accuracy and interval reliability.

6.2 Limitations

- Field validation is limited to a single test feeder at one site.
- Markov-gap simulation, while fitted to real outage logs, may not capture rarer correlated failure modes (e.g. region-wide load shedding).
- The Bayesian correction layer assumes residual stationarity within each weekly recalibration window.

6.3 Future Work

Future work should extend field validation to a multi-site trial across at least three distinct feeder types, and investigate online recalibration to relax the stationarity assumption identified above.

6.4 Conclusion

This thesis demonstrates that Naledi-Forecast, combining a TCN encoder with Bayesian post-hoc correction, delivers materially more accurate and better-calibrated short-horizon load forecasts than existing baselines under realistic data-scarcity conditions representative of Sub-Saharan microgrid deployments, with encouraging early field evidence from a live test feeder.

Chapter A

Hyperparameter Settings

Table A.1: Naledi-Forecast training hyperparameters

Hyperparameter	**Value**
Residual blocks	5
Channels per block	64
Dilation factors	1, 2, 4, 8, 16
Optimiser	Adam (lr $= 3 \times 10^{-4}$)
Calibration window	10% rolling, refreshed weekly
Kernel (GP correction)	Matérn ($\nu = 1.5$)

Bibliography

[1] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*. Holden-Day, 1976.

[2] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[3] S. Bai, J. Z. Kolter, and V. Koltun, "An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling," *arXiv preprint arXiv:1803.01271*, 2018.

[4] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning," in *Proc. 33rd Int. Conf. on Machine Learning (ICML)*, 2016, pp. 1050–1059.