2026 8th Global Power, Energy and Communication Conference (GPECOM) | 979-8-3315-5204-6/26/$31.00 ©2026 IEEE | DOI: 10.1109/GPECOM70462.2026.11578868

# Frequency-Constrained Unit Commitment Using SODE-Labelled Affine Surrogates

Azeez O. Olasoji
*Department of Electrical Engineering*
*University of Cape Town*
Cape Town, South Africa
olsaze001@myuct.ac.za

David T.O Oyedokun
*Department of Electrical Engineering*
*University of Cape Town*
Cape Town, South Africa
david.oyedokun@uct.ac.za

Mohammad Rajabdorri
*IIT, ETSI-ICAI*
*Universidad Pontificia Comillas*
Madrid, Spain
mrajabdorri@comillas.edu

Juan Esteban Sierra-Aguillar
*Department of Electrical Engineering*
*Universidad de Antioquia*
Medellin, Colombia
juane.sierra@udea.edu.co

Chukwuemeka E. Okafor
*Department of Electrical Engineering*
*University of Cape Town*
Cape Town, South Africa
okfchu002@myuct.ac.za

Mkhutazi Mditshwa
*Department of Electrical Engineering*
*University of Cape Town*
Cape Town, South Africa
mdtmkh001@myuct.ac.za

Komla A. Folly
*Department of Electrical Engineering*
*University of Cape Town*
Cape Town, South Africa
komla.folly@uct.ac.za

***Abstract*—Frequency nadir is one of the most informative indicators of post-contingency frequency security, but embedding it in frequency-constrained unit commitment (FCUC) remains challenging because nadir relations are nonlinear, non-convex, and are often introduced through computationally burdensome analytical approximations. This paper develops a scalable, open-source workflow for FCUC in which nadir security is enforced through calibrated affine surrogates learned from a second-order differential equation (SODE) model while preserving unit-commitment mixed-integer linear programming (MILP) structure. The study employs Adaptive Latin Hypercube Sampling (A-LHS) to generate UC-relevant operating points, thereby avoiding exhaustive grid search, and are labelled using the open-source SODE model that captures dynamic effects neglected by simplified first-order formulations. Probability-calibrated linear surrogates are then embedded directly into UC through a common affine inequality, with conservatism controlled by an operator-selected probability threshold. The study focuses on the IEEE RTS-96 system under the largest contigency event of $400MW$ and positions this benchmark as an extension of our earlier smaller-system study to a materially larger test system. Results show that the analytical first-order formulation remains systematically optimistic against the SODE model. Also the SODE-trained surrogates show a much stronger security performance with modest cost impact and substantially lower computational burden.**

***Index Terms*—frequency-constrained unit commitment, frequency nadir, low-inertia systems, second-order differential equation, machine learning surrogate, IEEE RTS-96**

## I. Introduction

The share of renewable energy sources (RES) is steadily increasing in modern power systems, driven by the need to mitigate carbon emissions from fossil-fuel-based generation. However, with increased RES penetration, power systems operate with less synchronous inertia, making frequency security after credible outages a binding operational risk [1], [2]. Consequently, power-system scheduling cannot remain purely economic, and classical N-1 security criteria are often insufficient on their own in low-inertia systems. While conventional unit commitment (UC) is usually solved as a mixed-integer linear program (MILP), embedding frequency-security constraints transforms it into frequency-constrained unit commitment (FCUC), where the most informative metric, frequency nadir, is nonlinear and non-convex.

Broadly, FCUC studies fall into analytical and data-driven strands. Analytical approaches derive closed-form nadir relations from first-order differential equations (FODE) or swing-equation assumptions. Their main strength is tractability; their main weakness is structural simplification. They rely on idealised reserve ramps and uniform frequency response to obtain closed-form nadir expressions that can later be linearised through separable programming or piecewise-linear approximations [3]–[6]. These strategies preserve tractability, but they can remain optimistic when reserve-delivery dynamics deviate materially from the assumed first-order response.

979-8-3315-5204-6/26/$31.00 ©2026 IEEE



Authorized licensed use limited to: UNIVERSIDAD PONTIFICIA DE COMILLAS. Downloaded on August 27,2026 at 10:39:23 UTC from IEEE Xplore. Restrictions apply.

A second strand moves the nonlinear burden offline. Earlier work learned linear security rules using optimal classification trees, deep neural networks, and classical linear classifiers [7]–[10]. Related large-system FCUC studies also examine stochastic low-inertia scheduling and benchmark the loss of the largest infeed as the major contingency event in the FCUC model [6], [11]. The present paper follows the methodological line developed in [12] and extends it from a smaller island system to the IEEE RTS-96 test system by combining adaptive sampling, open-source SODE labelling, probability calibration, and affine UC embedding.

The main contributions of this paper are as follows:

- an open-source SODE-labelling workflow for FCUC on the IEEE RTS-96 benchmark;
- an adaptive Latin hypercube sampling (A-LHS) strategy that replaces exhaustive grid enumeration for large-system operating-point generation;
- probability-calibrated affine surrogates that are embedded directly into UC while preserving MILP structure; and
- a comparison showing that the resulting SODE-trained FCUC schedules are less optimistic than the analytical FODE-SP benchmark while remaining computationally tractable.

The remainder of the paper presents the methodology, case study and results, and conclusions.

## II. Methodology

This section summarises the methodology adopted in this study.

### A. Classical UC Formulation

The underlying unit commitment problem minimises start-up and production costs over the scheduling horizon,

$$\min \sum_{t\in T}\sum_{i\in I} \mathrm{suc}_{t,i}(x_{t,i}) + \mathrm{gc}_{t,i}(g_{t,i}), \tag{1}$$

subject to the standard binary-logic, minimum up/down time, ramping, generation-capacity, demand-balance, and transmission constraints described in the source study. Refer to [10], [13] for a detailed breakdown.

### B. Frequency Constraint Modelling

*1) Analytical (FODE) Approach:* As seen in [12], the analytical model represents post-outage frequency dynamics through a first-order model derived from the classical swing equation:

$$P_a = P_m - P_e \tag{2}$$

$$\frac{2H}{f_0}\frac{d\Delta f(t)}{dt} + D\Delta f(t) = P_a - \sum_i \min(r_i(t), p_i) \tag{3}$$

Post-contingency system inertia:

$$H_{sys} = \sum_{i\in I_{on}} H_i M_i x_{t,i} \tag{4}$$

*a) RoCoF Constraint:*

$$H_{sys} \geq \frac{P_l f_0}{2\Delta f_{cr}^{RoCoF}} \tag{5}$$

*b) QSS Constraint:*

$$\sum_{i\in I} r_i \geq P_l - DP_{D,t}\Delta f_{cr}^{QSS} \tag{6}$$

*c) Frequency Nadir Constraint:* The reserve response during a contingency is modelled as a piecewise function:

$$R_{eff}(t) = \begin{cases} \frac{R_{tot}}{T_g} t & \text{if } t \leq T_g \\ R_{tot} & \text{if } t > T_g \end{cases} \tag{7}$$

Total reserve after outage:

$$R_{tot} = \sum_{i\in I_{on}} r_i \tag{8}$$

The nonlinear frequency nadir constraint becomes:

$$HR_{tot} - \frac{f_0 T_g (\Delta P_l^{\max})^2}{4\Delta f_{cr}^{nadir}} + \frac{DP_{D,t}T_g(\Delta P_l^{\max})f_0}{4} \geq 0 \tag{9}$$

*d) Symbol Definitions::*

- $f_0$: nominal system frequency (Hz)
- $\Delta f_{cr}^{nadir}$: critical nadir frequency deviation (Hz)
- $\Delta P$: power imbalance after contingency (MW)
- $T_g$: governor deadband delay (s)
- $P_{D,t}$: system demand at time t
- $D$: system damping coefficient (MW/Hz)
- $H$: aggregated system inertia (s)
- $R_{tot}$: total spinning reserve available (MW)

*2) Second-Order Differential Equation (SODE) Approach:* The proposed model is a second-order differential equation (SODE) benchmark that captures governor–turbine dynamics during contingency events,

$$a_{2,i}\frac{d^2 r_i(t)}{dt^2} + a_{1,i}\frac{dr_i(t)}{dt} + r_i(t) = -k_i x_i \left(\Delta f(t) + b_i \frac{d\Delta f(t)}{dt}\right), \tag{10}$$

This model reflects mechanical response characteristics and damping. For the IEEE RTS-96 benchmark, public plant-test coefficients are unavailable, so the higher-order parameters are assigned by technology class rather than by plant-specific identification.

- $r_i(t)$: frequency response from generator $i$
- $k_i = 1/R_i$: droop gain (inverse of droop constant)
- $x_i$: binary commitment status
- $a_{1,i}, a_{2,i}, b_i$: parameters of each generating unit i, can be deduced from more accurate models or field tests (In the case of La-Palma, field tests as seen in [12])

Fig. 1 shows the second-order frequency-response schematic used for SODE-based labelling in this study.


Authorized licensed use limited to: UNIVERSIDAD PONTIFICIA DE COMILLAS. Downloaded on August 27,2026 at 10:39:23 UTC from IEEE Xplore. Restrictions apply.

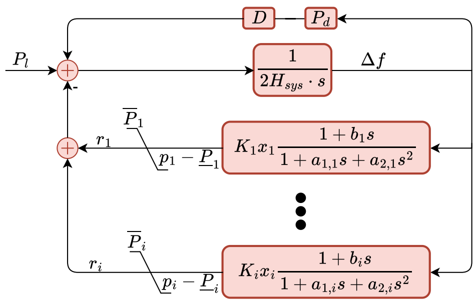

Fig. 1: Second-order frequency-response schematic.

*a) Model Interpretation:* The SODE model captures oscillatory behaviour and inertia-limited frequency support more accurately than FODE, as it is not predicated on the assumption of linear activation of reserves. However, the equation is nonlinear and not directly MILP-compatible; we simulate its response under thousands of contingency scenarios to generate labelled frequency nadir outcomes. These labels are later used to train affine ML classifiers that emulate the nadir boundary [12].

*b) ML surrogates and Surrogate tuning:* The SODE-labelled operating points are converted into safe/unsafe classes according to the adopted nadir-security threshold. Each linear surrogate then yields a decision score

$$f_\theta(x) = \Theta^\top x + \theta_0, \tag{11}$$

where $x$ collects the outage features $(p_\ell, H_{\text{after},\ell}, K_\ell, R_\ell)$. Since this raw score is not probabilistic, the features are calibrated by Platt scaling on an independent fold,

$$\hat{P}(\text{safe} \mid x) = \sigma(af_\theta(x) + b), \qquad \sigma(t) = \frac{1}{1+e^{-t}}. \tag{12}$$

An operator-selected probability floor $p^*$ is imposed by requiring $\hat{P}(\text{safe} \mid x) \geq p^*$. Defining the corresponding logit threshold $\psi^* = \log\big(p^*/(1-p^*)\big)$ gives

$$af_\theta(x) + b \geq \psi^*, \tag{13}$$

which is exactly equivalent to the deployable UC cut

$$k_0(p^*) + \sum_j k_j x_j \geq 0. \tag{14}$$

This point matters as conservatism is tuned through the probability threshold, not by changing the UC structure, adding binaries, or rebuilding the embedding each time the acceptable risk posture is adjusted.

For the IEEE RTS-96 study, the labelled library is split stratified by class into 30% external test data and 70% development data; within the development block, 80% is used for model fitting and 20% for Platt calibration. All classifier statistics reported in this paper are therefore computed on a held-out external test set after calibration has been fitted on an independent fold. Because missing unsafe points is operationally more severe than over-screening safe ones, unsafe-class recall and FN-rate remain the primary deployment metrics, while unsafe precision, unsafe F1-score, accuracy, and Brier score are reported for completeness.

TABLE I: Summary of building and labelling the IEEE-96 dataset.

| Stage | Samples | Runtime [s] |
|---|---|---|
| A–LHS generation | 42,500 | 169.0 |
| SODE labelling | 42,500 | 2394.2 |

*c) IEEE RTS-96 Dataset Generation:* In this study, we apply the SODE framework to the IEEE RTS-96 test system in order to evaluate the viability of the proposed approach presented in [12] on a materially larger benchmark. The study employs the adaptive Latin hypercube sampling (A-LHS), in contrast to the exhaustive grid-search technique used in [10], [12], to generate feasible UC operating points. Because grid search is only computationally tractable on a small test system, it is replaced with an adaptive Latin hypercube sampling (A-LHS) scheme [14], [15]. The A-LHS draws blocks of candidate dispatch vectors by stratified sampling in $(0, 1)^k$ and maps them to discrete power levels for each generator. For $k$ parameters it needs only $n$ samples to cover the space, so its complexity is $\mathcal{O}(nk)$ and scales linearly with system size.

For the IEEE-96 bus test system, the dataset is constructed around the fixed outage of unit i87 (400 MW), representing the benchmark largest-infeed N-1 contingency event, consistent with literature [6], [11]. Demand is sampled between 3891.4 MW and 8107.1 MW using 50 MW bands with a target of 500 feasible operating points per band. This sample size is a deliberate coverage–cost compromise: enlarging the library would improve boundary coverage, but at a higher offline labelling cost. The total number of demand bands is

$$B = \left\lfloor \frac{D_{\max} - D_{\min}}{\Delta D} \right\rfloor = \left\lfloor \frac{8107.1 - 3891.4}{50} \right\rfloor \approx 85, \tag{15}$$

and the corresponding target library size is therefore

$$N_{\text{samp}} = B \times N_{\text{tar}} = 85 \times 500 = 42{,}500. \tag{16}$$

Candidate dispatch vectors are generated using A-LHS and then screened through commitment realism, aggregate-demand, and post-outage reserve filters before being passed into the SODE model. The final benchmark library therefore contains 42,500 UC-feasible operating points generated in 169.0 s, while SODE labelling requires 2394.2 s. Compared with the runtime reported for the La Palma system in [12], the dataset-generation time for the IEEE 96-bus test system is reduced by a factor of approximately 16.5.

## III. Case Study and Results

### A. System Description

As stated prior the test system for evaluating this approach is the modified IEEE RTS–96 test system [16]. The test system has 96 buses, 120 lines, 96 thermal units and 19 wind plants


Authorized licensed use limited to: UNIVERSIDAD PONTIFICIA DE COMILLAS. Downloaded on August 27,2026 at 10:39:23 UTC from IEEE Xplore. Restrictions apply.

TABLE II: Technology-class coefficients used in the IEEE-96 FODE and SODE simulations.

| Technology | $a_1$ | $a_2$ | $b$ | $T_{\mathrm{FO}}$ [s] |
|---|---|---|---|---|
| OCGT | 12.0 | 3.5 | 1.6 | 10 |
| CCGT | 16.0 | 5.0 | 1.9 | 10 |
| Coal | 24.0 | 10.0 | 2.5 | 10 |
| IGCC | 20.0 | 8.0 | 2.2 | 10 |
| Nuclear | 28.0 | 12.0 | 2.8 | 10 |

TABLE III: Perturbation check for the IEEE-96 prototype SODE parameterisation over 500 representative points.

| Scenario | SODE mean [Hz] | FODE mean [Hz] | Unsafe SODE [%] | Agree. [%] |
|---|---|---|---|---|
| Class $-20\%$ | 49.499 | 49.673 | 3.6 | 98.6 |
| Class nominal | 49.474 | 49.673 | 5.0 | 100.0 |
| Class $+20\%$ | 49.454 | 49.673 | 5.4 | 99.6 |
| Common nominal | 49.459 | 49.673 | 5.4 | 99.6 |

over a 7 day horizon. All simulations on the IEEE RTS–96 bus test system focus on a **high-wind, high-demand** operating condition (HWHD), chosen to stress frequency-constrained scheduling under simultaneously tight security requirements and substantial RES penetration, in line with prior RTS–96 FCUC studies that contrast extreme wind and demand conditions [6]. Wind generation is represented by bus-located plants calibrated to 50% daily penetration. Frequency-security limits for IEEE RTS–96 are tightened to reflect large interconnected-system requirements. We use $f_0 = 50$ Hz, $\mathrm{RoCoF}_{\max} = 0.5$ Hz/s, $\Delta f_{\max} = 0.8$ Hz, and $\Delta f_{\mathrm{ss}} = 0.5$ Hz, with load damping $D_{\mathrm{load}} = 0.01$ and governor time constant $T_g = 10$ s, consistent with the RTS–96 setting adopted in [6] and in table II.

### *B. Prototype parameterisation and SODE–FODE comparison for IEEE-96*

Because higher-order plant-test coefficients are not publicly available for RTS-96, second-order coefficients are assigned by technology class rather than by plant-specific identification. Table II summarises the adopted coefficients for open-cycle gas turbine (OCGT), combined-cycle gas turbine (CCGT), coal, integrated gasification combined cycle (IGCC), and nuclear units. This design preserves dynamic heterogeneity across the IEEE-96 generating fleet without overstating plant-level realism. Unless stated otherwise, all IEEE-96 dataset statistics, feature-correlation results, and ML simulations reported in this section are based on the nominal class-based prototype parameterisation of table II. Perturbed and common-parameter cases are also introduced to evaluate the robustness of this design choice, as reported in Table III.

Across all tested scenarios, the SODE nadir remains deeper than the FODE nadir in 100% of cases and later in 98.4–98.6% of cases. Thus the qualitative discrepancy is not an artefact of one narrowly chosen parameter set.

We then evaluate the frequency trajectory of the FODE against the SODE model using the nominal class-based parameterisation in Table II. Figure 2 depicts the frequency

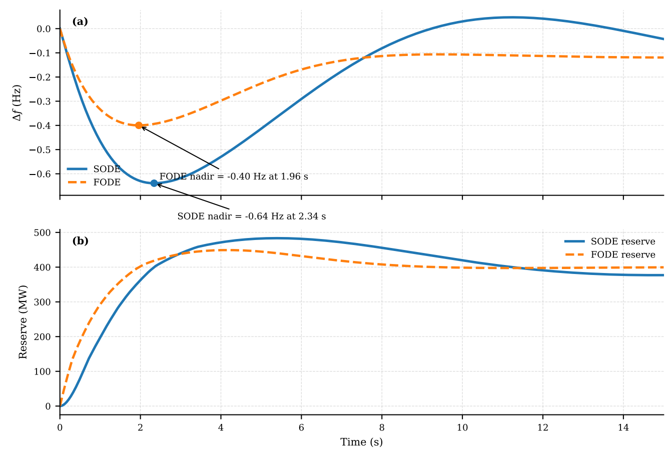


Fig. 2: Comparison of frequency response using the FODE and SODE model.

response under the benchmark contingency event. The upper panel shows the frequency deviation $\Delta f(t)$, while the lower panel reports the aggregate primary-reserve response. The FODE model reaches nadir at approximately $-0.40$ Hz after 1.96 s, whereas the SODE response reaches a deeper nadir of about $-0.64$ Hz after 2.34 s. The lower panel explains this discrepancy: in the FODE case the aggregate reserve rises rapidly and saturates with a single time constant, whereas the second-order governor-turbine dynamics in SODE produce a slower initial build-up, an overshoot, and a subsequent decay. This delayed reserve delivery translates directly into a deeper and later frequency nadir than suggested by the FODE model.

The bias is therefore structural, not merely numerical.

### *C. Dataset Summary and Feature Insights*

Using the nominal class-based parameterisation, the 42,500 simulated operating points produce a highly imbalanced SODE-labelled dataset with 873 unsafe points (2.05%) and 41,627 safe points (97.95%). Table IV reports the feature correlations with frequency nadir. Since the IEEE-96 study is built for a fixed outage only, the loss size is constant at 400 MW and its Pearson correlation is undefined by construction.

TABLE IV: Pearson correlation between SODE nadir and candidate features for the IEEE-96 dataset.

| Feature | $\mathrm{corr}(\mathrm{feature}, f_{\mathrm{nadir}})$ |
|---|---|
| Loss size $p_\ell$ | undefined (fixed at 400 MW) |
| System inertia $H_{\mathrm{after},\ell}$ | +0.803 |
| Response gain $K_\ell$ | +0.851 |
| Available reserves $R_\ell$ | +0.649 |

TABLE V: SODE nadir statistics by class for the IEEE-96 dataset under the 49.2 Hz threshold.

| Class | Mean [Hz] | Std [Hz] | Min [Hz] | Max [Hz] |
|---|---|---|---|---|
| Unsafe ($y = -1$) | $-0.930$ | 0.154 | $-2.308$ | $-0.800$ |
| Safe ($y = +1$) | $-0.478$ | 0.103 | $-0.800$ | $-0.271$ |


Authorized licensed use limited to: UNIVERSIDAD PONTIFICIA DE COMILLAS. Downloaded on August 27,2026 at 10:39:23 UTC from IEEE Xplore. Restrictions apply.

The nadir statistics in Table V clarify why threshold tuning matters. Unsafe points are relatively rare, but when they occur they are not marginal curiosities; their mean nadir deviation reaches $-0.930$ Hz and extends as low as $-2.308$ Hz. Safe points, by contrast, cluster much closer to the security boundary, with mean $-0.478$ Hz. The ML classifiers therefore see a heavily imbalanced but operationally meaningful boundary layer rather than two well-separated clouds. Any deployment strategy that rewards overall accuracy alone will be misleading.

### D. Surrogate Calibration Evaluation

The default threshold $p^* = 0.5$ proved unsuitable under the severe class imbalance of the IEEE-96 dataset. After conservative calibration to $p^* = 0.99$, unsafe-class recall improved sharply, yielding the deployment posture adopted for the final affine-surrogate evaluation in Table VI. In the ML performance discussion, the unsafe class ($y = -1$) is treated as the critical event class. Accordingly, $\mathrm{Recall}_{\mathrm{unsafe}}$ and FN-rate are used as the primary deployment metrics, since missed unsafe points are operationally more severe than conservative over-screening. $\mathrm{Precision}_{\mathrm{unsafe}}$, $\mathrm{F1}_{\mathrm{unsafe}}$, Accuracy, and Brier are reported as complementary indicators of event-detection quality, overall classification performance, and probability calibration, respectively.

TABLE VI: Final external-test performance of representative affine surrogates at $p^* = 0.99$

| Model | $\mathrm{Recall}_{\mathrm{unsafe}}$ | FN-rate | FP-rate | $\mathrm{Precision}_{\mathrm{unsafe}}$ | $\mathrm{F1}_{\mathrm{unsafe}}$ | Accuracy | Brier |
|---|---|---|---|---|---|---|---|
| SGD-Log | 1.000 | 0.000 | 0.123 | 0.123 | 0.220 | 0.879 | 0.01186 |
| LR-$C_\infty$ | 0.999 | 0.001 | 0.095 | 0.154 | 0.266 | 0.906 | 0.01083 |
| SVM-$C0.1$ | 0.999 | 0.001 | 0.102 | 0.145 | 0.253 | 0.899 | 0.01104 |
| SVM-$C1.0$ | 0.999 | 0.001 | 0.109 | 0.137 | 0.240 | 0.893 | 0.01122 |
| SVM-$C10$ | 0.997 | 0.003 | 0.113 | 0.132 | 0.234 | 0.889 | 0.01177 |
| LDA | 0.996 | 0.004 | 0.195 | 0.081 | 0.150 | 0.808 | 0.01407 |

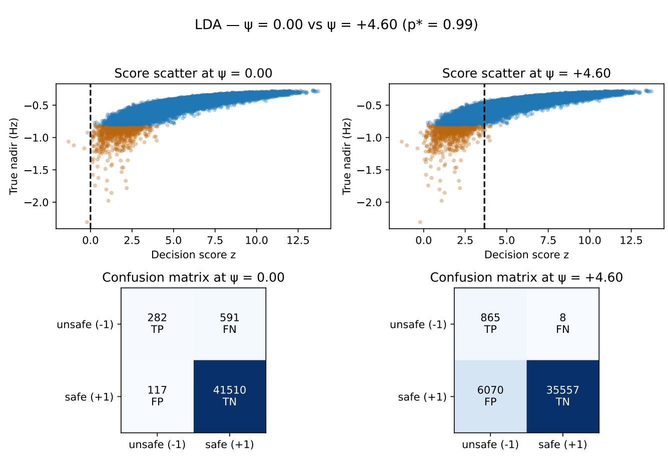


Fig. 3: IEEE RTS-96 LDA threshold-shift diagnostic: confusion matrices and the effect of conservative probability tuning.

Figure 3 further depicts the confusion matrices and shows how many unsafe points the ML surrogates miss before and after applying the calibration. Figure 3 is useful because it connects the classifier layer directly to the optimisation layer. Tightening the probability threshold does not change the algebraic form of the UC cut, but it does shift the intercept and therefore changes how conservative the feasible commitment region becomes. We next evaluate how this conservative shift impacts the FCUC. These results establish four points. The benchmark dataset can be generated and labelled at scale. The SODE model remains systematically more conservative than the FODE model. The main surrogate features are physically meaningful, with aggregate gain and inertia showing the strongest correlations to nadir severity. Finally, conservative probability calibration is not cosmetic; it is the mechanism that transforms a high-accuracy but unsafe classifier into a usable security screen in the proposed FCUC formulation.

## IV. FCUC Results on IEEE RTS-96

The learned nadir-security rules are embedded in a 7-day FCUC formulation with 50% RES penetration. Table VII summarises the average weekly UC results for the IEEE 96-bus system under the fixed outage of unit $i87$, comparing the unconstrained base case (BC), the analytical FODE-SP formulation, and the embedded SODE–ML surrogates.

TABLE VII: Average weekly UC results for the IEEE 96-bus system under a fixed N-1 outage of unit $i87$.

| Method | Obj | Unsafe (%) | Mean excess (Hz) | Time (s) | Gap (%) |
|---|---|---|---|---|---|
| BC | 6.15 | 75.60 | 3.15 | 139.54 | 0.42 |
| FODE-SP | 6.31 | 3.57 | 0.01 | 2490.04 | 0.64 |
| LDA | 6.34 | 0.00 | 0.00 | 896.06 | 0.75 |
| LR-$C_\infty$ | 6.36 | 0.00 | 0.00 | 181.62 | 0.82 |
| PA | 6.46 | 0.00 | 0.00 | 249.38 | 0.49 |
| SGD-Hinge | 6.55 | 0.00 | 0.00 | 980.83 | 0.76 |
| SGD-Log | 6.40 | 14.88 | 0.03 | 159.29 | 0.62 |
| SVM-C0.1 | 6.32 | 0.00 | 0.00 | 306.53 | 0.68 |
| SVM-C1 | 6.32 | 0.00 | 0.00 | 321.03 | 0.69 |
| SVM-C10 | 6.33 | 0.00 | 0.00 | 401.19 | 0.74 |

*Note:* Unsafe (%) denotes the percentage of hours for which the post-contingency nadir violates the adopted critical nadir-deviation limit. Mean excess denotes the average amount by which unsafe cases exceed $\Delta f^{\mathrm{crit}}_{\mathrm{nadir}} = 0.8$ Hz.

The BC yields the lowest weekly objective (6.15 M$), but it is clearly insecure, with an unsafe rate of $75.60\%$ and a mean excess of 3.15 Hz beyond the adopted nadir-deviation limit $\Delta f^{\mathrm{crit}}_{\mathrm{nadir}} = 0.8$ Hz. FODE-SP improves this substantially, reducing the unsafe rate to $3.57\%$, but residual violations remain. This confirms that the analytical FODE-based embedding is useful albeit with a bit of violations.

By contrast, most SODE-trained ML surrogates, including LDA, LR–$C_\infty$, PA, SGD-Hinge, and all SVM variants, achieve zero observed unsafe hours over the weekly horizon. SGD-Log is the only clear exception, with an unsafe rate of $14.88\%$. The operational interpretation of Table VII is therefore sequential and shows the economy–security–tractability pattern.. First, the unconstrained BC is economically attractive but unacceptable from a nadir-security perspective. Second, FODE-SP is markedly better than BC, yet it still leaves residual insecurity. Third, the secure SODE-trained surrogates recover fully secure


Authorized licensed use limited to: UNIVERSIDAD PONTIFICIA DE COMILLAS. Downloaded on August 27,2026 at 10:39:23 UTC from IEEE Xplore. Restrictions apply.

schedules without the high runtime burden of the analytical benchmark.

Among the secure surrogates, LR–$C_\infty$ provides the best overall balance between economy and tractability. Its weekly objective is $6.36$ M\$, only marginally above FODE-SP, while its runtime decreases from $2490.04$ s to $181.62$ s, corresponding to an acceleration of about $13.7\times$. The SVM surrogates are also fully secure but slower, while LDA and SGD-Hinge remain secure but computationally less attractive. Hence, the main practical value of the proposed SODE-ML FCUC framework lies not in large cost savings but in recovering fully secure weekly schedules at a fraction of the computational burden of the analytical FCUC benchmark.

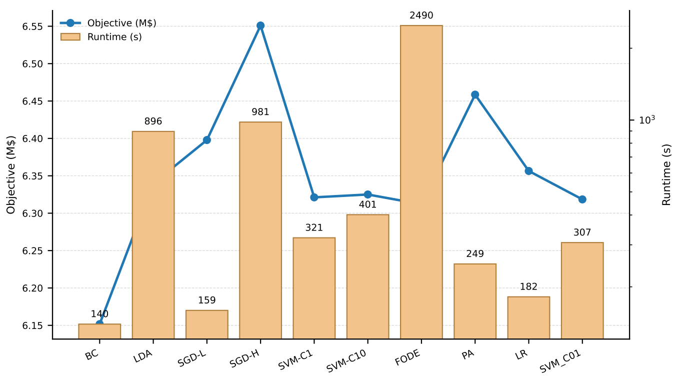


Fig. 4: Average weekly objective value and solve time across UC and FCUC methods for the IEEE RTS-96 system.

Figure 4 highlights this trade-off by jointly comparing weekly objective value and solution time across the candidate methods. The secure ML surrogates remain tightly clustered in objective value, whereas runtime differences are much larger.

## V. Discussion and Conclusion

This paper presented a data-driven FCUC framework in which frequency-nadir security is enforced through calibrated affine surrogates trained on SODE-labelled operating points and embedded directly in UC as linear constraints. Results on the IEEE RTS-96 system show that the learned SODE-based surrogates provide stronger frequency security than the analytical FODE model at substantially lower computational cost, indicating that SODE-trained affine security cuts offer a practical route toward tractable frequency-secure UC in low-inertia systems.

The scope of the present conference paper is deliberately benchmark-oriented. The IEEE RTS-96 study is built around a fixed largest-infeed outage, the second-order parameters are assigned by technology class because plant-test coefficients are not publicly available, and the operating-point library introduces a one-time offline generation and labelling burden. The present contribution is therefore methodological rather than plant-calibrated against proprietary simulators or field measurements. Instead, the purpose of the workflow is to develop a transparent open-source methodology for FCUC studies that reduces dependence on proprietary simulation environments while preserving physically meaningful dynamic screening. In addition, major topology or control-policy changes would generally require relabelling and retraining before deployment. Within those limits, the paper demonstrates that frequency-nadir security can be embedded in UC through SODE-trained affine surrogates without sacrificing MILP tractability, providing a practical alternative to analytical FCUC formulations on a materially larger test system.

## References


[1] A. Ulbig, T. S. Borsche, and G. Andersson, “Impact of low rotational inertia on power system stability and operation,” *IFAC Proceedings Volumes*, vol. 47, no. 3, pp. 7290–7297, 2014.

[2] A. O. Olasoji, D. T. O. Oyedokun, S. O. Omogoye, and C. Thron, “Review of frequency response strategies in renewable-dominated power system grids: Market adaptations and unit commitment formulation,” *Scientific African*, vol. 26, pp. 1–21, 2024.

[3] H. Ahmadi and H. Ghasemi, “Security-constrained unit commitment with linearized system frequency limit constraints,” *IEEE Transactions on Power Systems*, vol. 29, no. 4, pp. 1536–1545, 2014.

[4] L. Badesa, F. Teng, and G. Strbac, “Simultaneous scheduling of multiple frequency services in stochastic unit commitment,” *IEEE Transactions on Power Systems*, vol. 34, no. 5, pp. 3858–3868, 2019.

[5] B.Luis, F. Teng, and G. Strbac, “Optimal portfolio of distinct frequency response services in low-inertia systems,” *IEEE Transactions on Power Systems*, vol. 35, no. 6, pp. 4459–4469, Nov. 2020.

[6] C. J. Ferrandon-Cervantes, B. Kazemtabrizi, and M. C. Troffaes, “Inclusion of frequency stability constraints in unit commitment using separable programming,” *Electric Power Systems Research*, vol. 203, p. 107669, 2022.

[7] D. T. Lagos and N. D. Hatziargyriou, “Data-driven frequency dynamic unit commitment for island systems with high res penetration,” *IEEE Transactions on Power Systems*, vol. 36, no. 5, pp. 4699–4711, 2021.

[8] Y. Zhang, C. Chen, G. Liu, T. Hong, and F. Qiu, “Approximating trajectory constraints with machine learning–microgrid islanding with frequency constraints,” *IEEE Transactions on Power Systems*, vol. 36, no. 2, pp. 1239–1249, 2020.

[9] Y. Zhang, H. Cui, J. Liu, F. Qiu, T. Hong, R. Yao, and F. Li, “Encoding frequency constraints in preventive unit commitment using deep learning with region-of-interest active sampling,” *IEEE Transactions on Power Systems*, vol. 37, no. 3, pp. 1942–1955, 2021.

[10] M. Rajabdorri, B. Kazemtabrizi, M. Troffaes, L. Sigrist, and E. Lobato, “Inclusion of frequency nadir constraint in the unit commitment problem of small power systems using machine learning,” *Sustainable Energy, Grids and Networks*, vol. 36, p. 101161, 2023.

[11] M. Paturet, U. Markovic, S. Delikaraoglou, E. Vrettos, P. Aristidou, and G. Hug, “Stochastic unit commitment in low-inertia grids,” *IEEE Transactions on Power Systems*, vol. 35, no. 5, pp. 3448–3458, 2020.

[12] A. O. Olasoji, D. T. Oyedokun, M. Rajabdorri, J. E. Sierra-Aguillar, C. E. Okafor, M. Mditshwa, and K. A. Folly, “Frequency-constrained uc via second-order odes with ml surrogates,” in *2025 5th International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, 2025, pp. 1–6.

[13] A. O. Olasoji, D. Oyedokun, M. Rajabdorri, and J. E. Sierra-Aguillar, “Investigating inclusion of linear sensitivity factors in analytical frequency constrained unit commitment formulation,” in *2025 33rd Southern African Universities Power Engineering Conference (SAUPEC)*. IEEE, 2025, pp. 1–6.

[14] B. Bischl, M. Binder, M. Lang, T. Pielok, J. Richter, S. Coors, J. Thomas, T. Ullmann, M. Becker, A.-L. Boulesteix *et al.*, “Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges,” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 13, no. 2, p. e1484, 2023.

[15] P. Borisut and A. Nuchitprasittichai, “Adaptive latin hypercube sampling for a surrogate-based optimization with artificial neural network,” *Processes*, vol. 11, no. 11, p. 3232, 2023.

[16] H. Pandžić, Y. Dvorkin, T. Qiu, Y. Wang, and D. S. Kirschen, “Toward cost-efficient and reliable unit commitment under uncertainty,” *IEEE Transactions on Power Systems*, vol. 31, no. 2, pp. 970–982, 2015.




Authorized licensed use limited to: UNIVERSIDAD PONTIFICIA DE COMILLAS. Downloaded on August 27,2026 at 10:39:23 UTC from IEEE Xplore. Restrictions apply.