



(RESEARCH ARTICLE)



CONDITIONAL RELIABILITY MAPS FOR LANGMUIR-PROBE INVERSION UNDER NOISE, HOT-ELECTRON TAILS, AND FORWARD-MODEL DISCREPANCY

Zainab H. Shakir *

Department of Physics, College of Science, Al-Nahrain University, Baghdad, Iraq.

* Corresponding Author

ORCID Details

Zainab H. Shakir: <https://orcid.org/0009-0009-9139-0406>

International Journal of Science and Research Archive, 2026, 20(03), 1016–1029

Publication history: Received on 14 August 2026; revised on 24 September 2026; accepted on 26 September 2026

Article DOI: <https://doi.org/10.30574/ijrsra.2026.20.3.1807>

Copyright © 2026 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution License 4.0

ABSTRACT

Langmuir probe inversions can appear precise when the assumed electron distribution or collection law is wrong. This study maps where three inversion strategies remain reliable by separating noise, voltage sampling, hot-electron structure, and forward-model discrepancy in synthetic argon plasma. Conventional semi-log analysis, robust whole-curve single-Maxwellian fitting, and multistart two-Maxwellian fitting screened by an operational BIC-shaped penalized-residual score were applied to 2,896 traces. The benchmark comprised 1,440 factorial noise and hot-fraction cases, 220 matched-family Latin-hypercube cases, 360 voltage-resolution trials, 300 out-of-family challenges, and 576 temperature-ratio cases. Under the matched Maxwellian family, whole-curve fitting reduced median absolute temperature error from 10.40% to 1.83% at 5% noise and density error from 2.58% to 1.13%. Matched-family Latin-hypercube validation produced median errors of 0.56% in temperature, 0.22% in density, and 0.014 V in plasma potential. Under the out-of-family challenge, robust-fit errors increased to 2.30%, 12.39%, and 0.543 V, revealing parameter-specific sensitivity concealed by matched-family validation. At 1% noise and a 20% hot fraction, forcing one Maxwellian produced mean-temperature biases of 5.87%, 21.69%, and 52.02% for hot/cold temperature ratios of 2, 3.5, and 5. The gate selected the two-temperature model in 0/240 sampled Maxwellian trials (exact 95% Clopper-Pearson interval, 0-1.5%), while hot-tail selection varied with noise, fraction, and temperature separation. The resulting maps are conditional reliability bounds, not universal detection limits, showing that small matched-family errors do not establish model-independent metrological accuracy.

Keywords: Langmuir Probe; Plasma Diagnostics; Inverse Problem; Bi-Maxwellian Electrons; Monte Carlo Uncertainty; Forward-Model Discrepancy; Model Selection

1 INTRODUCTION

Electrostatic probes are still one of the most practical instruments for local measurements in laboratory and space plasmas. In a swept Langmuir-probe measurement, the dependence of collected current on probe bias is used to determine the electron temperature, electron density and plasma potential. Differentiation of the characteristic can also

give information about the electron energy distribution function (EEDF). The theoretical background varies from early collector model of Mott-Smith and Langmuir [1] to later descriptions of the operating regimes, sheath formation and instrumental constraints [2-5]. An I-V sweep seems simple, but reliable parameter recovery requires a collection model consistent with probe geometry, collisionality, magnetic field, electron distribution, surface state, and circuit response [2,6-9].

In routine analysis two fundamentally different sources of errors are often confused. One comes from the measurement and its processing. Numerical differentiation amplifies noise; estimates of the plasma potential depend on the choice of smoothing and knee detection; and recovery of temperature is sensitive to the choice of electron-retarding interval. Li et al. [10] showed a strong dependence of estimates of the Maxwellian temperature on the position and width of the semi-log fit window. Hence, structured pre-processing and model-fitting workflows are becoming more and more important for wall-embedded and high-speed probe systems [11,12]. RF ripple, finite plasma resistance, contamination of the probe surface and electronic response [13-16] could contribute to further distortion.

The other source is model-form error. A single Maxwellian electron population is expected to generate an approximately linear electron-current trace on semi-logarithmic axes. Many plasmas, however, contain energetic minorities, depleted low-energy populations, drifted components, or nonextensive tails. In a bi-Maxwellian plasma, the temperature inferred from a single apparent slope may differ from the density-weighted EEDF temperature [17]. Alternative distribution models and derivative-based procedures have therefore been developed [18,19]. Experimental evidence further indicates that magnetic transport and non-Maxwellian structure can complicate classical probe interpretation [20]. Accordingly, a low residual obtained with a flexible model does not, by itself, establish the physical adequacy of the assumed distribution.

Simulation allows these effects to be separated because the generating plasma parameters and distribution are known. Resendiz Lira and Marchand [21] employed a simulation-library approach for inference from Langmuir-probe data, whereas Brooks and Paliwoda [22] quantified uncertainty in a plasma-impedance probe through Monte Carlo analysis. Recent work has also emphasized that probe noise can have a physical lower bound and need not behave as ideal white noise [23]. Together, these studies motivate a benchmark of intermediate complexity: one that is more transparent than a full particle-in-cell simulation library, yet sufficiently structured to reveal when conventional or single-population analysis fails.

This work thus finds three challenges: noise and voltage-resolution tests in the presence of a matched forward family, hot-tail distribution mismatch, and collection-law mismatch outside the inverse-model family. The same noisy traces were used to compare competing procedures within each relevant test. The analysis procedures included standard semi-log analysis, robust whole-curve single-Maxwellian fitting, and multistart two-Maxwellian fitting for the hot-tail tests. A two-temperature interpretation was kept only when a pre-defined operational penalized-score threshold was exceeded. The main outputs are two conditional reliability maps, one characterising the parameter-recovery error of the single-Maxwellian inversion, and the other reporting the operational frequency of the two-temperature selection. Separate maintenance of the three challenges avoids the confusion of matched family precision with experimental accuracy and the implication of a universal threshold for hot tail detection. Therefore, each of the bounds reported is relative to the conditions under which it was derived.

2 PHYSICAL AND NUMERICAL METHODS

2.1 Probe geometry and forward current model

A cylindrical tungsten probe was represented by its collecting area, with radius 0.125 mm and exposed length 5.0 mm. The tip area was retained, giving $A_p = 2\pi r_p L_p + \pi r_p^2 = 3.976 \times 10^{-6} \text{ m}^2$. The plasma was argon with ion mass 39.948 u. The benchmark assumes a stationary, unmagnetized, weakly collisional bulk and uses a planar thermal-flux approximation; it is therefore a diagnostic algorithm benchmark rather than a complete sheath or discharge simulation. The electron saturation current of population j was calculated from the random thermal flux [2,3,7]:

$$I_{e,s,j} = en_j A_p \sqrt{\frac{eT_j}{2\pi m_e}}. \quad (1)$$

Temperatures are expressed in electron-volts, so eT_j has units of energy. To avoid an artificial derivative discontinuity at the plasma potential, the retarding potential was regularized with a soft positive-part function

$$\Delta V_s = w \ln \left\{ 1 + \exp \left[\frac{V_p - V}{w} \right] \right\}, \quad (2)$$

where $w = 0.18$ V. The component densities were defined as $n_c = (1 - f_h) n_e$ and $n_h = f_h n_e$. For a cold population T_c and a hot fraction f_h with $T_h/T_c = 3.5$, the electron current was

$$I_e(V) = \sum_j I_{e,s,j} \exp\left(-\frac{\Delta V_{s,j}}{T_j}\right). \quad (3)$$

The density-weighted mean electron temperature used as the target quantity for reporting single-temperature bias was

$$\bar{T} = (1 - f_h)T_c + f_h T_h. \quad (4)$$

For the synthetic ion branch only, the same density-weighted arithmetic mean temperature was used as an explicit empirical scale in a mild ion-sheath expansion that prevents an unrealistically flat negative-current branch. This choice is part of the benchmark generator and is not presented as a general two-temperature Bohm criterion. With $\alpha_i = 0.10$, the ion current was

$$I_i(V) = -0.61en_e A_p \sqrt{\frac{e\bar{T}}{m_i}} \sqrt{1 + \frac{\alpha_i \Delta V_s}{\bar{T}}}. \quad (5)$$

The total clean characteristic was $I(V) = I_e(V) + I_i(V)$. Figure 1 summarizes the complete benchmark workflow, while Figure 2 shows the forward curves and the semi-log signature of a minority hot population.

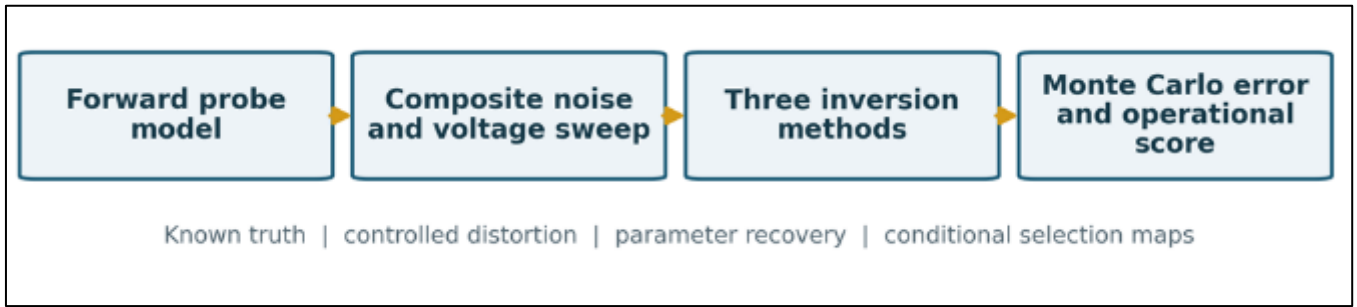


Figure 1: Benchmark workflow from a known forward model to noise distortion, three inversion strategies, and Monte Carlo reliability assessment.

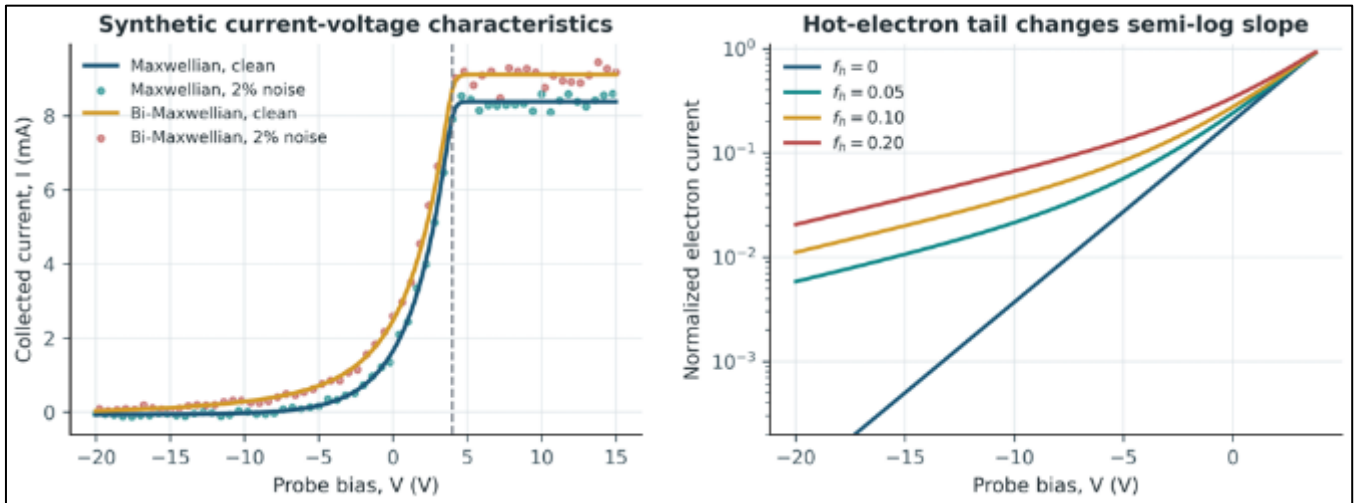


Figure 2: Representative synthetic characteristics. Left: clean and 2% noisy Maxwellian and bi-Maxwellian I-V curves. Right: normalized electron current in the retarding region for increasing hot-electron fraction.

2.2 Composite measurement noise model

The measured trace was contaminated by heteroscedastic Gaussian noise, a weak sinusoidal RF-like component, a constant offset and a linear drift. The Gaussian standard deviation at each voltage for full-scale noise fraction η and

current span ΔI was $\sigma(V) = \eta \Delta I [0.10 + 0.90\sqrt{(|I|/\Delta I)}]$. The remaining terms were randomized independently for every replicate:

$$I_{\text{meas}} = I + \varepsilon_G + A_{\text{RF}} \sin\left(\frac{2\pi V}{P} + \varphi\right) + b_0 + b_1 \frac{V - V_{\text{mean}}}{\Delta V}. \quad (6)$$

The Gaussian term satisfied $\varepsilon_G(V) \sim N[0, \sigma^2(V)]$. The RF-like amplitude was .22 $\eta \Delta I$, the period was sampled uniformly from 1.6 to 3.2 V and the phase was uniform on $[0, 2\pi]$. The standard deviations of offset and drift were $0.06\eta \Delta I$ and $0.08\eta \Delta I$, respectively. In Eq. (6), ΔV is the total sweep width (35 V) and not the voltage step size δV used in the resolution study. This composite form is not meant to reproduce a particular instrument spectrum; it avoids the benchmark degenerating to ideal independent identically distributed noise.

2.3 Inversion strategies

Conventional semi-log analysis. The current was smoothed with a third-order Savitzky-Golay filter [24]. A linear ion branch was extrapolated from the lowest 16% of the voltage samples and subtracted. The plasma-potential knee was initialized from the maximum first derivative. A sigma-clipped line was then fitted to $\ln(I_e)$ over the positive retarding-current window, normally 3.5–72% of the estimated electron saturation current. Temperature followed from the reciprocal slope, and density from Eq. (1).

Robust single-Maxwellian fit (1M). The complete trace was fitted to Eqs. (1)–(5) using bounded nonlinear least squares with a soft-L1 loss. The fitted variables were $\log_{10}(ne)$, $\log(Te)$, V_p , a bounded ion-current scale, and a small baseline offset. Heteroscedastic residual scaling prevented the electron-saturation plateau from dominating the retarding and ion branches.

Two-Maxwellian fit (2M). The full curve was fitted with ordered temperatures using seven variables: $\log_{10}(ne)$, $\log(T_c)$, Th/T_c , fh , V_p , ion scale, and baseline. Three starts, centered on hot-temperature ratios of 2.3, 3.5, and 5.0, were optimized independently; the lowest weighted residual was retained. The bounds $Th/T_c = 1.35 - 7.0$ and $fh = 0.003 - 0.35$ prevented label exchange and nonphysical population fractions.

Operational penalized model gate. For n voltage samples and k fitted variables, a BIC-shaped score was calculated using the following expression [25]:

$$\text{BIC} = n \ln\left(\frac{\text{RSS}_w}{n}\right) + k \ln(n). \quad (7)$$

RSS_w is the sum of squared heteroscedastically scaled residuals evaluated at the soft-L1 optimum; $k = 5$ for 1M and $k = 7$ for 2M. The two-temperature interpretation was selected when $\Delta \text{BIC} = \text{BIC}_{1\text{M}} - \text{BIC}_{2\text{M}} > 6$. Because the residual scale balances current branches rather than representing a separately calibrated Gaussian error variance, and because the single-population limit is a boundary or singular case of the mixture comparison, the usual likelihood interpretation of BIC is not guaranteed. Accordingly, ΔBIC is used only as an operational penalized comparison score within this benchmark, not as a Bayes factor or a universal evidence threshold. The penalty remains useful operationally because the 2M curve has two additional fitted variables and will generally reduce the raw residual.

2.4 Simulation design and reproducibility controls

Table 1 summarizes the physical baseline, factorial variables, and validation ranges.

Table 1: Physical parameters and numerical design.

Quantity	Value or range	Role
Working gas / ion mass	Ar / 39.948 u	Fixed
Probe radius; exposed length	0.125 mm; 5.0 mm	Fixed
Collecting area, A_p	$3.976 \times 10^{-6} \text{ m}^2$	Fixed
Baseline electron density	$5.0 \times 10^{16} \text{ m}^{-3}$	Fixed baseline
Cold temperature, T_c	2.5 eV	Fixed baseline
Hot/cold temperature ratio	3.5 baseline; 2.0 and 5.0 extension	Sensitivity
Plasma potential, V_p	4.0 V	Fixed baseline

Hot fraction, f_h	0, 0.02, 0.05, 0.10, 0.15, 0.20	Factorial
Full-scale noise, η	0.2%, 0.5%, 1%, 2%, 5%	Factorial
Voltage sweep / increment	-20 to 15 V / 0.20 V	Fixed baseline
Replicates	48 per factorial cell	1,440 traces
Separate Maxwellian design	n_e : 10^{16} - 10^{17} m ⁻³ ; T_e : 1.2-6.5 eV; V_p : 1.5-7.5 V; η : 0.3-3%	220 LHS cases
Voltage-resolution study	δV = 0.10, 0.20, 0.50, 1.00 V; η = 1%	360 traces
Out-of-family challenge	w : 0.08-0.35 V; γ : 0.05-0.35; α^* : 0.04-0.22; p : 0.35-0.75; ρ : 0.25-0.85	300 LHS cases
Temperature-ratio extension	T_h/T_c = 2.0 and 5.0; f_h = 0.05, 0.10, 0.20; η = 1%, 2%, 5%	576 new traces
Total unique traces	2,896	All studies
Master seed	20260801	Deterministic

The factorial design was $6 \times 5 \times 48 = 1,440$ traces. A different four-dimensional Latin-hypercube sample [26] tested the 1M inversion over the ranges of Maxwellian density, temperature, plasma potential and noise. The voltage-resolution study was conducted in 90 replicates at each increment. The master random seed was 20260801 and competing inversions had identical noisy traces. Verification included clean limit recovery, bounded solution checks, deterministic seed and 10th-90th percentile error bands. The optimisation and signal processing were done using SciPy [27].

2.5 Structural discrepancy challenge and temperature ratio extension

A second validation was designed specifically to reduce the matched-family or inverse-crime advantage. The generating model remained Maxwellian in its retarding branch but was not available to the inverse routine. Inspired by the growth of the attracting branch in nonplanar collection treatments [2,4,7,28], the electron current was modified to

$$I_e^* = I_{e,s} \exp\left(-\frac{\Delta V_s}{T_e}\right) \sqrt{1 + \frac{\gamma(V-V_p)_+}{T_e}}. \quad (8)$$

Here, $(x)_+ = \max(x, 0)$, and γ controls attractive-branch growth. The ion response was independently changed to

$$I_i^* = -I_{i,s} \left(1 + \frac{\alpha^* \Delta V_s}{T_e}\right)^p. \quad (9)$$

The nine-dimensional 300-case Latin-hypercube challenge varied density, temperature, plasma potential, noise, knee width w , attractive-branch coefficient γ , ion-sheath coefficient α^* , ion exponent p , and AR(1) noise correlation ρ over the ranges in Table 1. The inverse model retained the fixed planar knee and ion response of Eqs. (1)-(5), so its errors measure structural sensitivity rather than optimizer self-consistency. This empirical stress form is not presented as a complete orbital-motion-limited or kinetic sheath solution. Only the semi-log and 1M estimators were evaluated in this challenge; the 1M-2M operational gate was not applied to these 300 traces. The challenge therefore quantifies parameter-recovery sensitivity but does not estimate spurious hot-tail selection under collection-law discrepancy or AR(1) noise correlation.

A separate temperature-ratio extension tested $T_h/T_c = 2.0$ and 5.0 at hot fractions 0.05, 0.10, and 0.20 and noise levels 1%, 2%, and 5%, using 32 new replicates per cell. The original 48-replicate factorial cells supplied the ratio-3.5 reference without duplicating traces. The extension therefore added 576 unique cases and evaluated 1,008 trial records in the three-ratio selection matrix. Across all non-duplicated studies, 2,896 unique traces were inverted.

Model-selection entries are empirical binomial frequencies. Baseline cells contain 48 trials, whereas the added ratio-2 and ratio-5 cells contain 32 trials; individual-cell frequencies therefore have appreciable sampling uncertainty. For scale, 0/32 selections has an exact two-sided 95% Clopper-Pearson interval of 0-10.9%, and 26/48 selections, reported as 54.2%, has an interval of 39.2-68.6%. Differences between nearby cell frequencies should therefore be interpreted cautiously.

3 RESULTS

3.1 Correct model recovery under current noise

The Maxwellian subset isolates noise and processing error because the 1M forward and inverse distribution families agree. The robust whole-curve fit was consistently less sensitive than the semi-log procedure (Table 2; Figure 3). At 0.2% noise, median absolute temperature and density errors were 0.07% and 0.04% for 1M. At 5% noise, they increased to 1.83% and 1.13%, whereas the semi-log errors reached 10.40% and 2.58%. The semi-log plasma-potential estimate remained approximately 0.5 V from the true value because smoothing and the chosen derivative knee do not coincide exactly with the regularized V_p ; the 1M median absolute V_p error was 0.051 V even at 5% noise.

Table 2: Median absolute recovery errors for truly Maxwellian traces.

Noise (%)	Semi-log T error (%)	1M T error (%)	Semi-log n error (%)	1M n error (%)	1M V_p error (V)
0.2	1.28	0.07	0.33	0.04	0.002
0.5	1.14	0.27	0.30	0.10	0.007
1.0	2.49	0.44	0.58	0.15	0.010
2.0	3.94	0.77	1.16	0.33	0.014
5.0	10.40	1.83	2.58	1.13	0.051

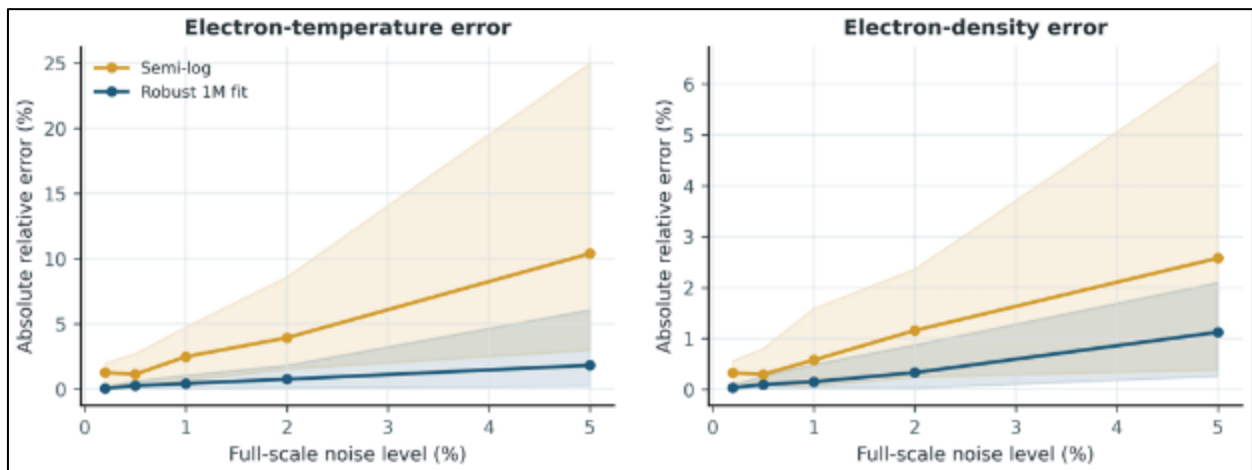


Figure 3: Median absolute relative error versus full-scale noise for Maxwellian traces. Shading denotes the 10th-90th percentile interval across 48 replicates.

3.2 Bias caused by an unresolved hot-electron minority

Table 3: Distribution-mismatch results at 1% full-scale noise (medians across 48 replicates). pp: percentage points. The 2M entries at $f_h = 0$ are reported as NA because the operational gate selected 2M in 0/48 trials; unselected component fits are not interpreted.

True f_h (%)	Semi-log mean-T bias (%)	1M mean-T bias (%)	2M mean-T bias (%)	2M error T_c (%)	2M error f_h (pp)	Observed selection (%)
0	-1.47	-0.15	NA	NA	NA	0.0
2	-1.85	0.95	0.45	1.75	1.10	95.8
5	-1.06	4.53	-0.60	1.73	1.62	100.0
10	-2.65	9.42	-0.29	2.48	2.03	100.0
15	-3.05	15.62	0.16	2.55	1.97	100.0
20	-5.87	21.69	-0.24	4.37	2.92	100.0

Noise and model mismatch produced qualitatively different behavior. At 1% noise, the 1M mean-temperature estimate was nearly unbiased for $f_h = 0.02$ but became progressively high as the hot population grew: median signed errors were +4.53%, +9.42%, +15.62%, and +21.69% for $f_h = 0.05, 0.10, 0.15$, and 0.20 , respectively (Table 3; Figure 4). The semi-log estimate was less responsive to the full mean energy and became modestly low, reaching -5.87% at $f_h = 0.20$. Agreement between the two single-temperature procedures is therefore not guaranteed, and neither recovers the two physical components.

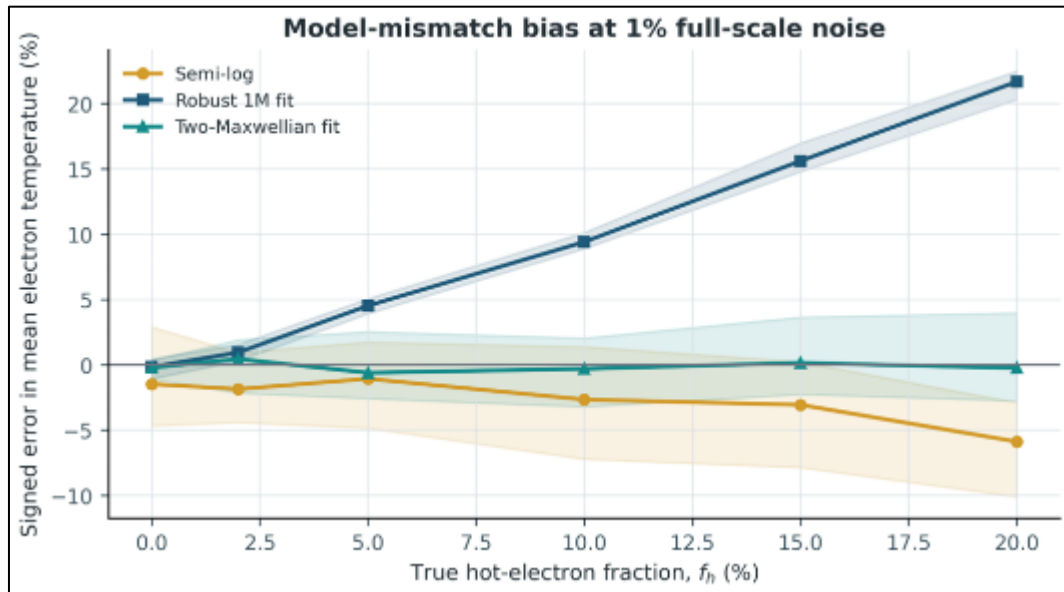


Figure 4: Signed bias in density-weighted mean electron temperature at 1% full-scale noise. Lines show medians and shading shows the 10th-90th percentile interval. The 2M point at $f_h = 0$ is displayed only as a diagnostic fit output and is not interpreted as a selected two-temperature decomposition.

3.3 Conditional two-temperature selection frequencies

The operational gate selected the 2M model in 0 of 240 truly Maxwellian trials (0%; exact two-sided 95% Clopper-Pearson interval for the sampled false-selection proportion, 0-1.5%). This finite-sample result does not imply that the underlying false-selection probability is zero. Observed selection frequencies depended sharply on hot fraction and noise (Table 4; Figure 5). A 2% hot minority was selected in 48/48 (100%), 48/48 (100%), 46/48 (95.8%), 2/48 (4.2%), and 0/48 (0%) trials at 0.2%, 0.5%, 1%, 2%, and 5% noise, respectively. At 5% noise, the corresponding frequencies were 0/48 (0%), 26/48 (54.2%), 41/48 (85.4%), and 47/48 (97.9%) for $f_h = 0.05, 0.10, 0.15$, and 0.20 . These are empirical frequencies conditional on the voltage range, 176-point base sweep, hot/cold ratio of 3.5, residual weighting, gate threshold, and specified noise model.

Table 4: Observed frequency (%) with which the operational $\Delta\text{BIC} > 6$ gate selected the two-Maxwellian model.

Noise (%)	$f_h=0\%$	$f_h=2\%$	$f_h=5\%$	$f_h=10\%$	$f_h=15\%$	$f_h=20\%$
0.2	0.0	100.0	100.0	100.0	100.0	100.0
0.5	0.0	100.0	100.0	100.0	100.0	100.0
1.0	0.0	95.8	100.0	100.0	100.0	100.0
2.0	0.0	4.2	100.0	100.0	100.0	100.0
5.0	0.0	0.0	0.0	54.2	85.4	97.9

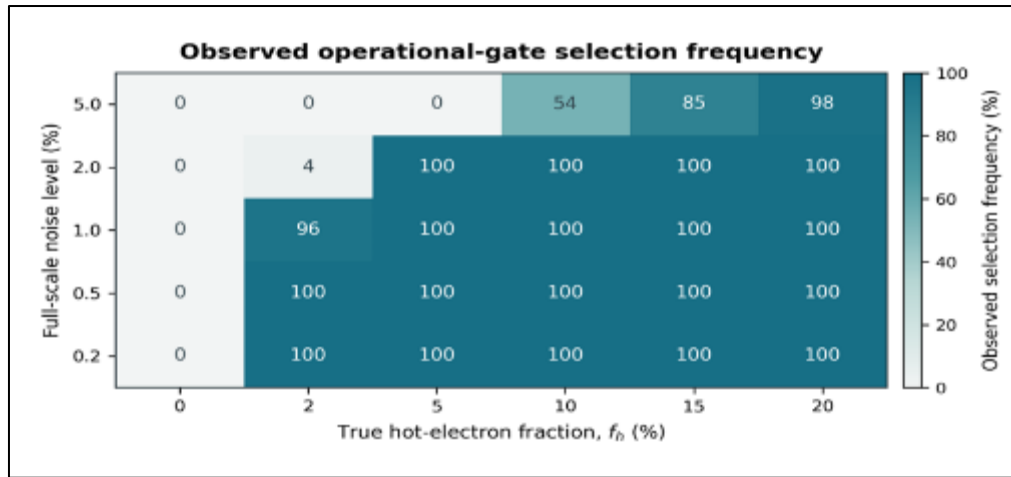


Figure 5: Observed frequency with which the operational $\Delta\text{BIC} > 6$ gate selected the two-temperature model across hot-fraction and full-scale-noise levels.

3.4 Recovery of the component parameters

Figure 6 reports diagnostic component-error summaries from the fitted 2M solutions; these summaries do not by themselves establish component identifiability. Interpretation should be restricted to cells with frequent 2M selection. In particular, at $f_h = 0.02$ and 2% noise, only 2 of 48 trials passed the gate, so the corresponding component-error summary cannot be read as reliable recovery. At 1% noise, median absolute cold-temperature error ranged from 1.73% at $f_h = 0.05$ to 4.37% at $f_h = 0.20$, and median hot-fraction error ranged from 1.10 to 2.92 percentage points. At 2% noise, these errors rose to 3.32–6.36% and 2.40–4.75 percentage points, respectively. Their variation partly reflects coupling among T_c , T_h/T_c , f_h , and the saturation-current density scale. No profile-likelihood, parameter-correlation, or boundary-hit analysis was performed; the separate component values should therefore be treated as diagnostic conditional estimates rather than unique physical decompositions.

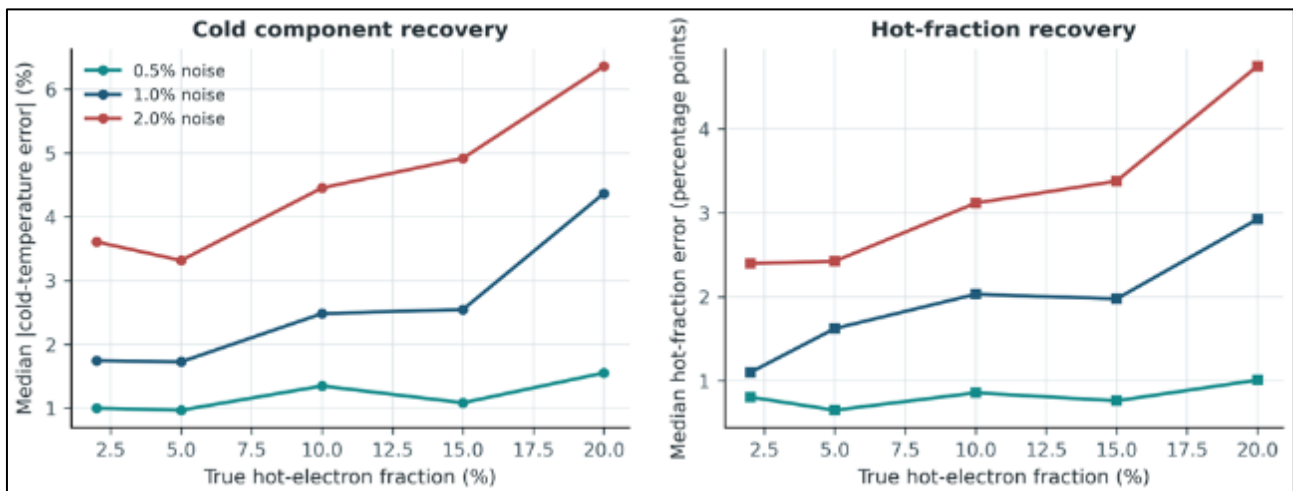


Figure 6: Diagnostic median absolute errors in cold-electron temperature and hot-electron fraction from the two-Maxwellian fits at 0.5%, 1%, and 2% noise. These summaries should be interpreted together with the selection frequencies in Figure 5.

3.5 Voltage resolution and separate Maxwellian validation

Coarsening the voltage increment from 0.10 to 1.00 V had little effect on the 1M method at 1% noise: median absolute temperature error increased from 0.33% to 0.78%, and density error from 0.11% to 0.41% (Table 5; Figure 7). The semi-log temperature error increased from 1.58% to 23.09% because only a few samples remained in the selected retarding interval. This comparison does not imply that a 1 V sweep is universally sufficient; the whole-curve fit benefits from the exact smooth forward family used in this benchmark.

Table 5: Voltage-resolution sensitivity at 1% noise (median absolute errors; 90 replicates per increment).

δV (V)	Semi-log T error (%)	1M T error (%)	1M n error (%)
0.10	1.58	0.33	0.11
0.20	1.63	0.47	0.20
0.50	6.42	0.66	0.31
1.00	23.09	0.78	0.41

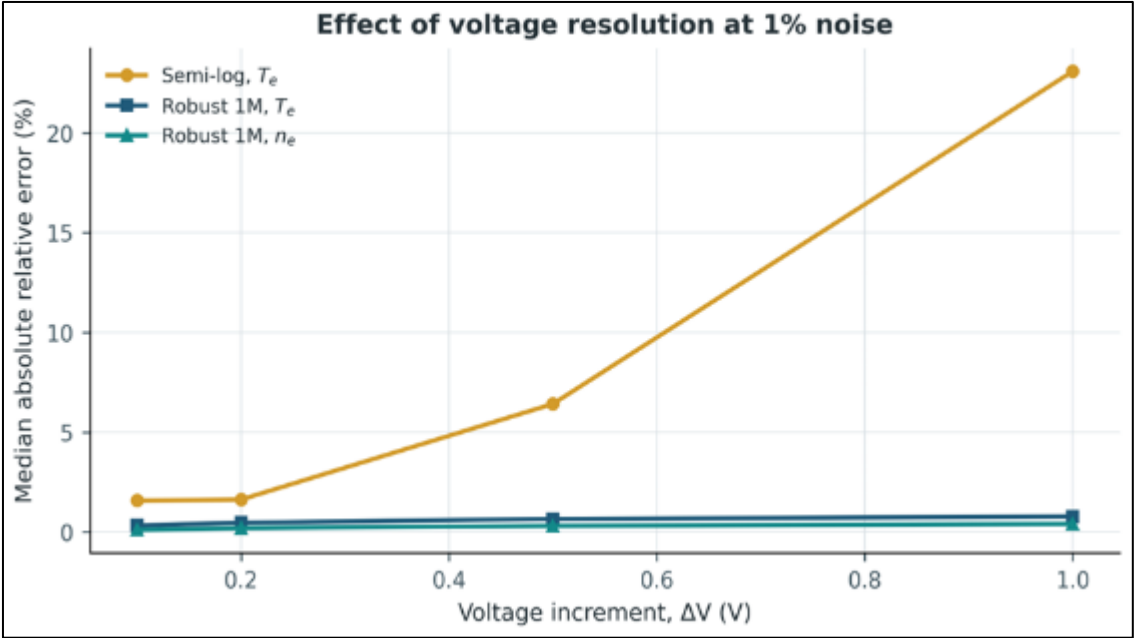


Figure 7: Median absolute recovery error versus voltage increment at 1% full-scale noise.

The separate 220-case Latin-hypercube design tested the 1M fit outside the single baseline state. Across $n_e = 10^{16}$ - 10^{17} m^{-3} , $T_e = 1.2$ - 6.5 eV, $V_p = 1.5$ - 7.5 V, and 0.3-3% noise, median absolute errors were 0.564% for temperature, 0.218% for density, and 0.014 V for plasma potential. The 90th-percentile values were 1.772%, 0.812%, and 0.055 V, respectively (Table 6; Figure 8). The near-diagonal scatter confirms numerical consistency for the specified Maxwellian forward family.

Table 6: Separate Latin-hypercube Maxwellian validation (220 cases).

Metric	Median	90th percentile	Unit
Electron temperature relative error	0.564	1.772	%
Electron density relative error	0.218	0.812	%
Plasma-potential absolute error	0.014	0.055	V

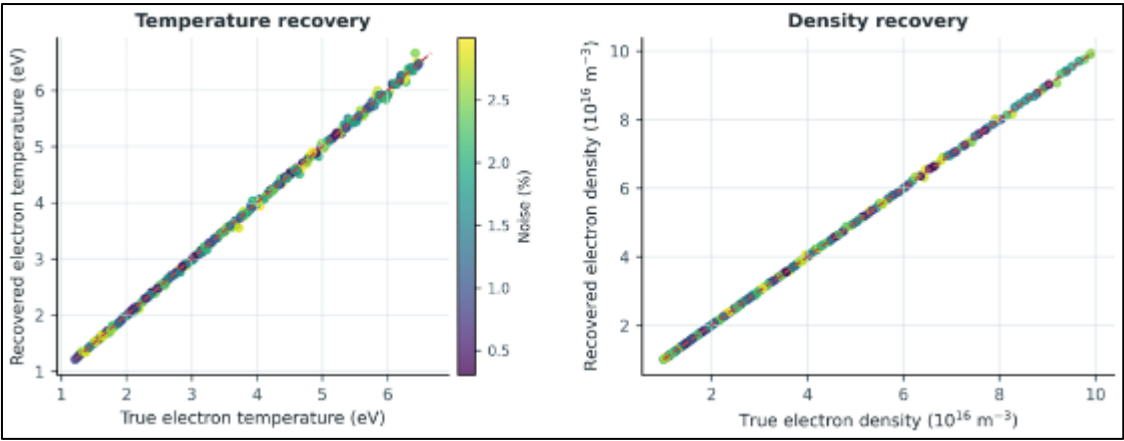


Figure 8: Recovered versus true temperature and density for the separate Maxwellian Latin-hypercube design. Color denotes full-scale noise; the dashed line is identity.

3.6 Sensitivity to forward-model discrepancy

Changing the generating collection law exposed parameter-specific fragility that was hidden by matched-family validation. Across the 300 out-of-family cases, the robust 1M fit retained median absolute temperature error of 2.30% (6.63% at the 90th percentile), but density error increased to 12.39% (22.41%) and plasma-potential error to 0.543 V (0.955 V). The semi-log method had larger temperature error, 11.76% (26.99%), and density error, 14.23% (27.21%), while its plasma-potential error was slightly lower at 0.473 V (0.859 V). Thus, whole-curve redundancy protected temperature better than density or plasma potential under this particular structural stress; it did not make all parameters model independent (Table 7; Figure 9).

Table 7: Matched-family and out-of-family recovery summaries from separate designs. The matched row contains 220 cases; each out-of-family method was evaluated on the same 300 challenged traces.

Forward method	set /	T error median / p90 (%)	n error median / p90 (%)	V _p error median / p90 (V)
Matched family	/ 1M	0.56 / 1.77	0.22 / 0.81	0.014 / 0.055
Out of family	/ 1M	2.30 / 6.63	12.39 / 22.41	0.543 / 0.955
Out of family	/ semi-log	11.76 / 26.99	14.23 / 27.21	0.473 / 0.859

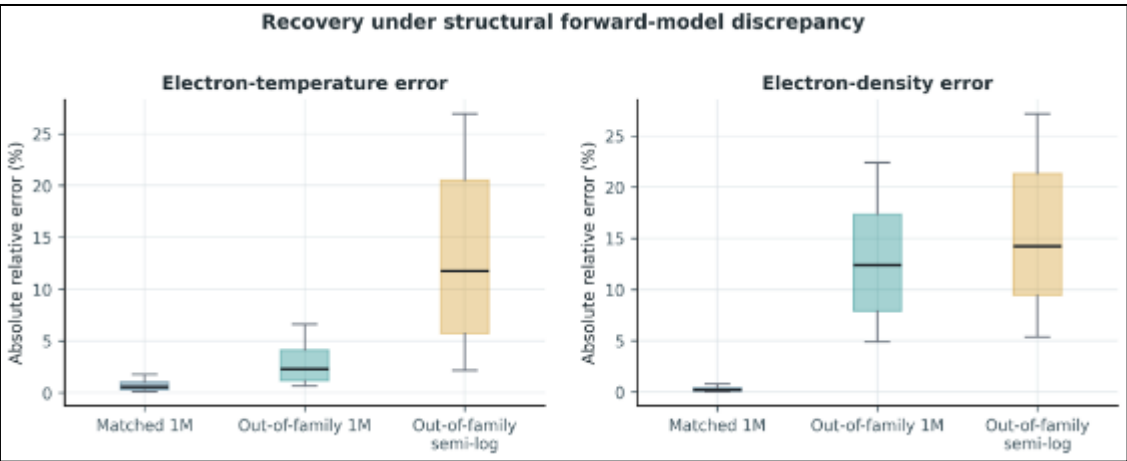


Figure 9: Recovery errors in the 300-case out-of-family challenge. Panels compare semi-log and robust single-Maxwellian inversion; boxes show the interquartile range and whiskers extend to 1.5 times that range.

3.7 Sensitivity of conditional selection to hot/cold temperature separation

Temperature separation shifted both the size and the observed selection frequency of hot-tail bias. At 1% noise and $f_h = 0.20$, the median signed 1M mean-temperature bias increased from +5.87% for $T_h/T_c = 2$ to +21.69% for 3.5 and +52.02% for 5. At 5% noise, a 5% hot fraction was selected in 0/32 (0%), 0/48 (0%), and 16/32 (50%) trials at ratios 2, 3.5, and 5, respectively; for a 10% fraction, the corresponding frequencies were 0/32 (0%), 26/48 (54.2%), and 32/32 (100%). Conversely, at ratio 2 and 5% noise, even the 20% hot population was not selected in the 32 sampled trials. The same fraction therefore cannot be assigned a universal selection threshold without specifying temperature separation and sampling uncertainty (Table 8; Figure 10).

Table 8: Observed operational-gate two-Maxwellian selection frequency (%) across temperature ratio, hot fraction, and noise. Ratio 3.5 reuses 432 factorial traces; ratios 2 and 5 comprise 576 new traces.

T_h/T_c	Noise (%)	$f_h=5\%$	$f_h=10\%$	$f_h=20\%$
2	1	15.6	100.0	100.0
2	2	0.0	18.8	78.1
2	5	0.0	0.0	0.0
3.5	1	100.0	100.0	100.0
3.5	2	100.0	100.0	100.0
3.5	5	0.0	54.2	97.9
5	1	100.0	100.0	100.0
5	2	100.0	100.0	100.0
5	5	50.0	100.0	100.0

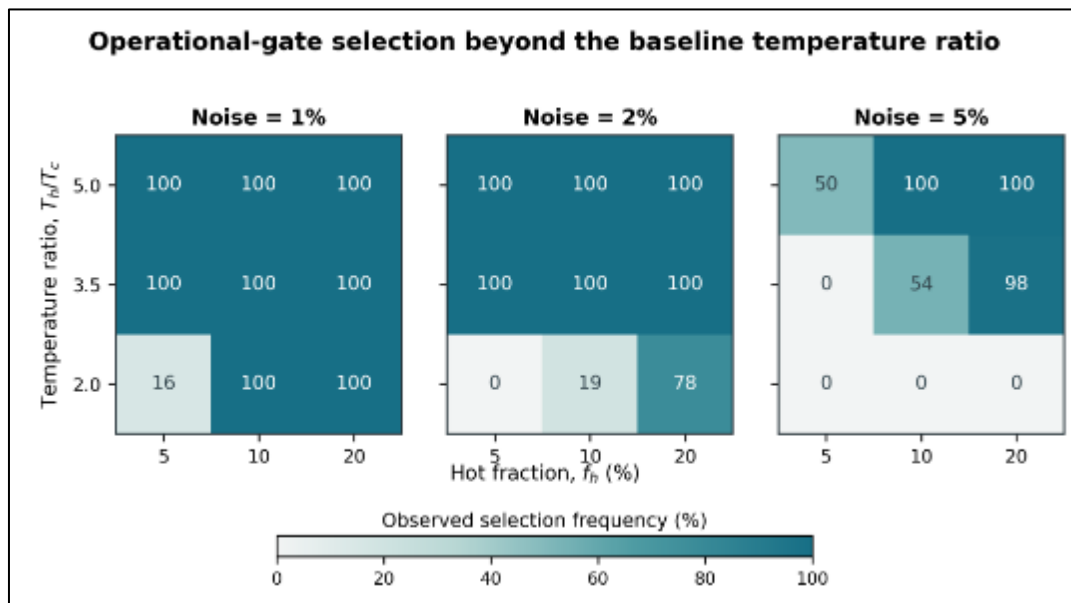


Figure 10. Observed operational-gate two-temperature selection frequency across hot/cold temperature ratio and hot fraction. Separate panels show 1%, 2%, and 5% full-scale noise; cell labels are rounded to the nearest whole percentage.

4 DISCUSSION

Under the matched Maxwellian family, whole-curve fitting used information from the ion branch, retarding region, knee, and saturation plateau, whereas the semi-log estimate depended on a local slope. This redundancy explains its lower sensitivity to noise and voltage resolution in the present benchmark. The result is consistent with evidence that fitting-window choice, ion-current subtraction, and preprocessing strongly influence classical estimates [10-12], and it supports statistical fitting and simulation-library approaches to probe inference [12,21].

Matched-family accuracy is an algorithmic lower bound, not a claim of sub-percent experimental precision. The separate matched design used the same function family for generation and inversion and therefore retained an inverse-crime advantage. Once the generating collection law was withheld from the inverse routine, median density error increased from 0.22% to 12.39% and plasma-potential error from 0.014 to 0.543 V. Density absorbed amplitude and collection-law discrepancies, while temperature remained more stable under this particular stress. A small weighted residual therefore cannot validate either the inferred density or the assumed sheath physics. The out-of-family values are conditional stress-test results, but they expose uncertainty that the matched-family test cannot reveal [12,13,20].

The hot-tail tests isolate a different failure mode. A minority energetic population bends the semi-log trace, and a forced single-Maxwellian whole-curve fit accommodates that curvature by raising an apparent temperature and shifting density. At $f_h = 0.20$ and 1% noise, the median 1M mean-temperature bias rose from +5.87% at $T_h/T_c = 2$ to +52.02% at a ratio of 5. Choi et al. [17] likewise reported differences between single-temperature derivative estimates and effective bi-Maxwellian temperatures. Nonextensive probe formulations [18,19] and recent EEDF measurements [20] further support testing the distributional assumption rather than adopting a Maxwellian by default.

4.1 Interpretation and reporting implications

The operational score is a conditional model-comparison diagnostic, not evidence that the fitted components are uniquely identifiable. The additional 2M parameters can absorb noise and drift even when they reduce the residual. The Delta BIC > 6 gate selected 2M in 0/240 sampled Maxwellian trials, but that finite benchmark result is not a universal BIC guarantee. Selection also depended on temperature separation: at 5% noise and $f_h = 0.10$, the observed frequency increased from 0% at $T_h/T_c = 2$ to 54.2% at 3.5 and 100% at 5. Reports should therefore pair the operational score with semi-log curvature, residual inspection, trial counts, sampling uncertainty, and parameter-stability or profile information. Separate component temperatures should be withheld when the decomposition is weakly constrained.

Instrument and collection effects remain part of the inference problem. RF contamination can approach the probe-signal amplitude, finite plasma resistance can shear the characteristic, and surface contamination can change collection behavior [13-16]. An inversion cannot correct an effect that its forward model omits. The present framework is therefore most useful as a pre-experimental design tool: measured noise, voltage step, probe geometry, circuit response, and a device-specific collection model can be propagated through the same workflow before an EEDF decomposition is reported.

Limitations

All traces are synthetic, and no experimental validation is claimed. The forward model is thermal, unmagnetized, and weakly collisional, with a simplified ion-sheath expansion rather than a kinetic or particle-in-cell sheath solution.

The temperature-ratio extension samples only three discrete separations (2, 3.5, and 5), and the observed selection frequencies will also change with voltage range, sample count, probe area, and the detailed noise spectrum. The two-Maxwellian family cannot represent arbitrary EEDFs, anisotropy, beams, depleted distributions, or q/κ statistics. The 300-case structural challenge varies several collection-law features empirically, but it is not a unique physical model and does not replace a kinetic, particle-in-cell, or experimentally calibrated forward calculation.

The operational Δ BIC score is not a formally calibrated BIC because the soft-L1 objective and branch-balancing residual scale do not define a standard Gaussian likelihood, and the mixture comparison contains a boundary or singular limit. Its threshold and selection maps are empirical and benchmark-specific. Selection frequencies are estimated from only 48 or 32 replicates per cell and consequently have non-negligible binomial uncertainty. The 0/240 Maxwellian result has an exact two-sided 95% interval of 0-1.5% and does not establish a zero population false-positive probability. The 2M component estimates were not evaluated using profile likelihoods, parameter-correlation matrices, boundary-hit frequencies, or convergence-frequency summaries; passing the operational gate therefore does not prove that T_c , T_h , and f_h are separately identifiable.

In the bi-Maxwellian generator, the density-weighted arithmetic mean temperature enters Eq. (5) only as an empirical ion-current scale. Multi-temperature sheath entry need not reduce to this average, so the reported hot-tail maps are conditional on this proxy. The 300 out-of-family Maxwellian traces were not passed through the 1M-2M gate; false hot-tail selection caused by structural collection-law discrepancy or correlated noise was therefore not tested.

5 CONCLUSIONS

This 2,896-trace benchmark shows that numerical precision and physical validity are separate properties of Langmuir-probe inversion. Under the matched Maxwellian family, robust whole-curve fitting outperformed the conventional semi-log procedure and retained median temperature and density errors of 1.83% and 1.13% at 5% full-scale noise. A

separate matched-family design produced median errors of 0.56% in temperature, 0.22% in density, and 0.014 V in plasma potential. Under the structural challenge, the corresponding robust-fit medians were 2.30%, 12.39%, and 0.543 V. Sub-percent matched-family recovery therefore demonstrates algorithmic consistency; it does not establish model-independent metrological accuracy.

The hot-tail reliability map establishes no universal population-fraction threshold. At 1% noise and $f_h = 0.20$, forcing one Maxwellian produced mean-temperature bias from +5.87% at $T_h/T_c = 2$ to +52.02% at a ratio of 5. The operational gate was chosen 2M in 0 of 240 sampled Maxwellian trials (exact two-sided 95% interval, 0 to 1.5%), but the observed selection rate of this gate co-varied with noise, hot fraction and temperature separation. Reliable reporting should therefore include the forward model, residual weighting, operational comparison rule, number of replicates, sampling uncertainty and evidence of component stability. The fitted components require experimental validation with measured circuit, surface, noise and sheath behaviour before they can be given a physical meaning.

REFERENCES

- [1] H. M. Mott-Smith and I. Langmuir, 'The theory of collectors in gaseous discharges,' *Physical Review* 28 (1926) 727-763. <https://doi.org/10.1103/PhysRev.28.727>.
- [2] V. I. Demidov, S. V. Ratynskaia, and K. Rypdal, 'Electric probes for plasmas: The link between theory and instrument,' *Review of Scientific Instruments* 73 (2002) 3409-3439. <https://doi.org/10.1063/1.1505099>.
- [3] R. L. Merlino, 'Understanding Langmuir probe current-voltage characteristics,' *American Journal of Physics* 75 (2007) 1078-1085. <https://doi.org/10.1119/1.2772282>.
- [4] I. H. Hutchinson, *Principles of Plasma Diagnostics*, 2nd ed., Cambridge University Press, Cambridge, 2002.
- [5] M. A. Lieberman and A. J. Lichtenberg, *Principles of Plasma Discharges and Materials Processing*, 2nd ed., Wiley, Hoboken, 2005.
- [6] V. A. Godyak and V. I. Demidov, 'Probe measurements of electron-energy distributions in plasmas: What can we measure and how can we achieve reliable results?' *Journal of Physics D: Applied Physics* 44 (2011) 233001. <https://doi.org/10.1088/0022-3727/44/23/233001>.
- [7] V. A. Godyak and B. M. Alexandrovich, 'Comparative analyses of plasma probe diagnostics techniques,' *Journal of Applied Physics* 118 (2015) 233302. <https://doi.org/10.1063/1.4937446>.
- [8] B. E. Cherrington, 'The use of Langmuir probes for plasma diagnostics: A review,' *Plasma Chemistry and Plasma Processing* 2 (1982) 113-140. <https://doi.org/10.1007/BF00633129>.
- [9] K.-U. Riemann, 'The Bohm criterion and sheath formation,' *Journal of Physics D: Applied Physics* 24 (1991) 493-518. <https://doi.org/10.1088/0022-3727/24/4/001>.
- [10] J.-Q. Li, X.-Y. Xie, Q.-H. Zhang, S.-H. Li, and W.-Q. Lu, 'Data processing of Langmuir probe I-V traces to obtain accurate electron temperature and density in Maxwellian plasmas,' *Physics of Fluids* 34 (2022) 067115. <https://doi.org/10.1063/5.0097089>.
- [11] O. Février, C. Theiler, H. De Oliveira, B. Labit, N. Fedorczak, and A. Baillod, 'Analysis of wall-embedded Langmuir probe signals in different conditions on the Tokamak à Configuration Variable,' *Review of Scientific Instruments* 89 (2018) 053502. <https://doi.org/10.1063/1.5022459>.
- [12] K. R. Trent, A. D. Gallimore, and J. E. Foster, 'Enhanced method for analyzing Langmuir probe data and characterizing the electron energy distribution function (EEDF),' *Physics of Plasmas* 26 (2019) 063513. <https://doi.org/10.1063/1.5093892>.
- [13] S. N. Andreev, A. V. Bernatskiy, N. A. Dyatko, I. V. Kochetov, V. V. Lagunov, and V. N. Ochkin, 'Accounting for plasma resistance in the interpretation of probe measurements in gas discharge plasma,' *Vacuum* 215 (2023) 112372. <https://doi.org/10.1016/j.vacuum.2023.112372>.
- [14] D. Mukhopadhyay, M. Bandyopadhyay, H. Tyagi, and K. Patel, 'Advancement of Langmuir probe-based laser photo-detachment technique for negative ion density measurement in a high-power helicon plasma source,' *Review of Scientific Instruments* 95 (2024) 043006. <https://doi.org/10.1063/5.0186429>.
- [15] K. I. Oyama, C. H. Lee, H. K. Fang, and C. Z. Cheng, 'Means to remove electrode contamination effect of Langmuir probe measurement in space,' *Review of Scientific Instruments* 83 (2012) 055113. <https://doi.org/10.1063/1.4722167>.
- [16] W. E. Amatucci, P. W. Schuck, D. N. Walker, P. M. Kintner, S. Powell, B. Holback, and D. Leonhardt, 'Contamination-free sounding rocket Langmuir probe,' *Review of Scientific Instruments* 72 (2001) 2052-2057. <https://doi.org/10.1063/1.1357234>.

- [17] I. Choi, C. Chung, and S. Y. Moon, 'Electron density and electron temperature measurement in a bi-Maxwellian electron distribution using a derivative method of Langmuir probes,' *Physics of Plasmas* 20 (2013) 083508. <https://doi.org/10.1063/1.4818609>.
- [18] F. J. Gonzalez, J. I. Gonzalez, S. Soler, C. E. Repetto, B. J. Gómez, and D. B. Berdichevsky, 'New procedure to estimate plasma parameters through the q-Weibull distribution by using a Langmuir probe in a cold plasma,' *Plasma Research Express* 4 (2022) 015003. <https://doi.org/10.1088/2516-1067/ac4f35>.
- [19] D. Xiao, X. Chen, S. Liu, H. Chen, and Y. Xiong, 'Development of the Langmuir probe under q-distribution for NCST,' *AIP Advances* 14 (2024) 015238. <https://doi.org/10.1063/5.0186389>.
- [20] E. G. Devin, J. Bak, G. Urdaneta Rincon, S. Yatom, and Y. Raitses, 'Measurements of the electron energy distribution function in partially magnetized low temperature plasmas,' *Physics of Plasmas* 33 (2026) 033507. <https://doi.org/10.1063/5.0310529>.
- [21] P. A. R. Resendiz Lira and R. Marchand, 'Simulation inference of plasma parameters from Langmuir probe measurements,' *Earth and Space Science* 8 (2021) e2020EA001344. <https://doi.org/10.1029/2020EA001344>.
- [22] J. W. Brooks and M. C. Paliwoda, 'Uncertainty analysis of the plasma impedance probe,' *Physics of Plasmas* 31 (2024) 053514. <https://doi.org/10.1063/5.0203675>.
- [23] H. L. Pécseli, 'Pitfalls in the interpretations of Langmuir probe currents,' *Physics of Plasmas* 32 (2025) 032110. <https://doi.org/10.1063/5.0251766>.
- [24] A. Savitzky and M. J. E. Golay, 'Smoothing and differentiation of data by simplified least squares procedures,' *Analytical Chemistry* 36 (1964) 1627-1639. <https://doi.org/10.1021/ac60214a047>.
- [25] G. Schwarz, 'Estimating the dimension of a model,' *Annals of Statistics* 6 (1978) 461-464. <https://doi.org/10.1214/aos/1176344136>.
- [26] M. D. McKay, R. J. Beckman, and W. J. Conover, 'A comparison of three methods for selecting values of input variables in the analysis of output from a computer code,' *Technometrics* 21 (1979) 239-245. <https://doi.org/10.1080/00401706.1979.10489755>.
- [27] P. Virtanen et al., 'SciPy 1.0: Fundamental algorithms for scientific computing in Python,' *Nature Methods* 17 (2020) 261-272. <https://doi.org/10.1038/s41592-019-0686-2>.
- [28] R. B. Lobbia and B. E. Beal, 'Recommended practice for use of Langmuir probes in electric propulsion testing,' *Journal of Propulsion and Power* 33 (2017) 566-581. <https://doi.org/10.2514/1.B35531>.