Supplementary

Table S1 Metadata of GNSS stations and approximate magnetic latitude (MLAT) values used in this study.

Parameter	Ambalavao (VOIM00MDG)	Haifa (BSHM00ISR)	Malindi (MAL200KEN)	Djibouti (DJIG00DJI)
Latitude, Longitude	−21.906, 46.793	32.779, 35.020	−2.996, 40.194	11.526, 42.847
Elevation (m)	163.3	225.1	−20.4	711.4
Country/Region	Ambalavao, Madagascar	Haifa, Israel	Malindi, Kenya	Arta, Djibouti
Receiver	JAVAD TR_2S	SEPT POLARX5	SEPT POLARX5	SEPT POLARX5
Antenna	JAVRINGANT_G5T	TRM59800.00	LEIAR25.R4	TRM59800.00
Satellite System	GPS+GLO+GAL+BDS+QZSS+SBAS	GPS+GLO+GAL+BDS+QZSS+IRNSS	GPS+GLO+GAL+BDS+SBAS	GPS+GLO+GAL+BDS+IRNSS+SBAS
Approx. MLAT (°)	-26.25	+29.52	-6.58	+7.39

Notes. Geographic coordinates were sourced from the IGS metadata, while magnetic latitudes were calculated using AACGMv2 (Epoch 2018). Information regarding the receiver and antenna was obtained from IGS station logs. Satellite counts represent averages over 30-second sampling intervals during storm periods.

Table S2. Summary of KLD Sensitivity to Estimation Parameters Across Stations and Models.

Station	Model	Δ% (Histogram Bins)	Δ% (KDE Bandwidth)
Ambalavao	IRI-2016	67.44%	64.34%
	IRI-PLAS	65.33%	56.45%
	NeQuick2	63.69%	61.02%

Haifa	IRI-2016	66.17%	54.54%
	IRI-PLAS	35.78%	30.00%
	NeQuick2	75.64%	67.23%
Malindi	IRI-2016	7.06%	45.53%
	IRI-PLAS	3.68%	48.01%
	NeQuick2	6.91%	45.22%
Djibouti	IRI-2016	61.58%	25.04%
	IRI-PLAS	65.54%	58.36%
	NeQuick2	62.49%	10.85%

This table summarizes the sensitivity of the Kullback–Leibler Divergence (KLD) to estimation parameters across all stations and ionospheric models. The Δ% values represent the relative change in KLD when (i) the histogram bin count is varied between 20 and 60 and (ii) the KDE bandwidth is varied between 0.30 and 0.70 TECU. A larger Δ% indicates a stronger dependence of KLD on the respective parameter choice. The results demonstrate that Ambalavao and Haifa exhibit particularly high sensitivity across all models, whereas Malindi shows comparatively lower sensitivity in histogram binning but moderate to strong dependence on the KDE bandwidth. These findings confirm that distributional mismatches between observations and models are not artifacts of parameterization but reflect structural divergences in model performance.

Supplementary Captions – Ambalavao

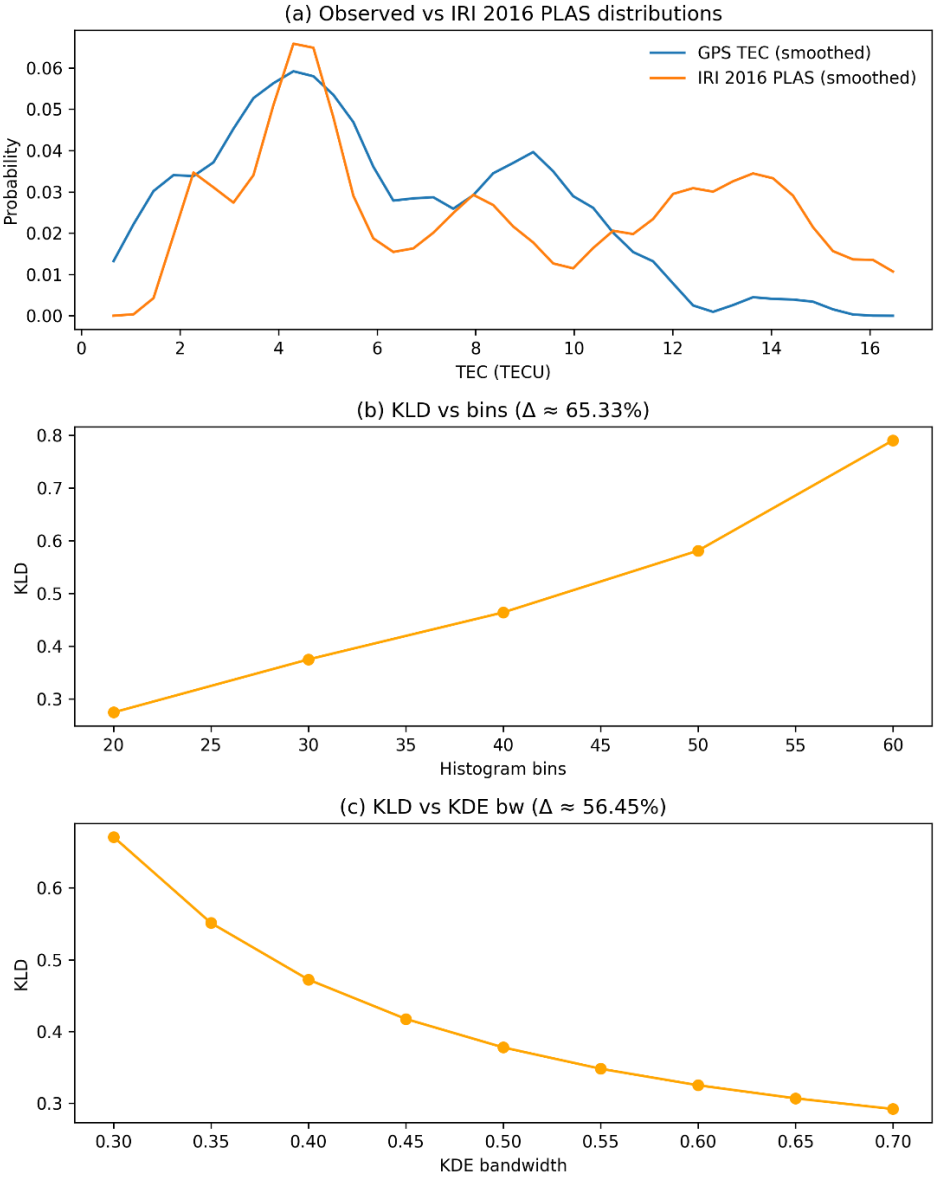

Figure S1. Sensitivity of Kullback–Leibler Divergence (KLD) for Ambalavao station: IRI-2016 PLAS model.

Panel (a) compares the observed GPS-TEC distribution with that of the IRI-2016 PLAS model. The model systematically underestimated the higher-end TEC values (>10 TECU) while slightly overestimating the peak around 4–5 TECU, resulting in a visible shift between the two distributions.

Panel (b) shows the KLD response to histogram bin count (20–60 bins). The divergence steadily increases from ~0.28 to ~0.78, corresponding to a relative change of $\Delta \approx 65.33\%$, indicating moderate sensitivity of the metric to binning choice.

Panel (c) illustrates the dependence on the KDE bandwidth (0.30–0.70). Here, KLD decreased from ~0.68 to ~0.30, with $\Delta \approx 56.45\%$. This trend suggests that finer bandwidths exaggerate the structural mismatch, whereas broader smoothing reduces the divergence.

Overall, the persistence of nonzero KLD across parameter settings confirms that the observed distributional differences between GPS-TEC and IRI-2016 PLAS are robust and not an artifact of binning or bandwidth selection.

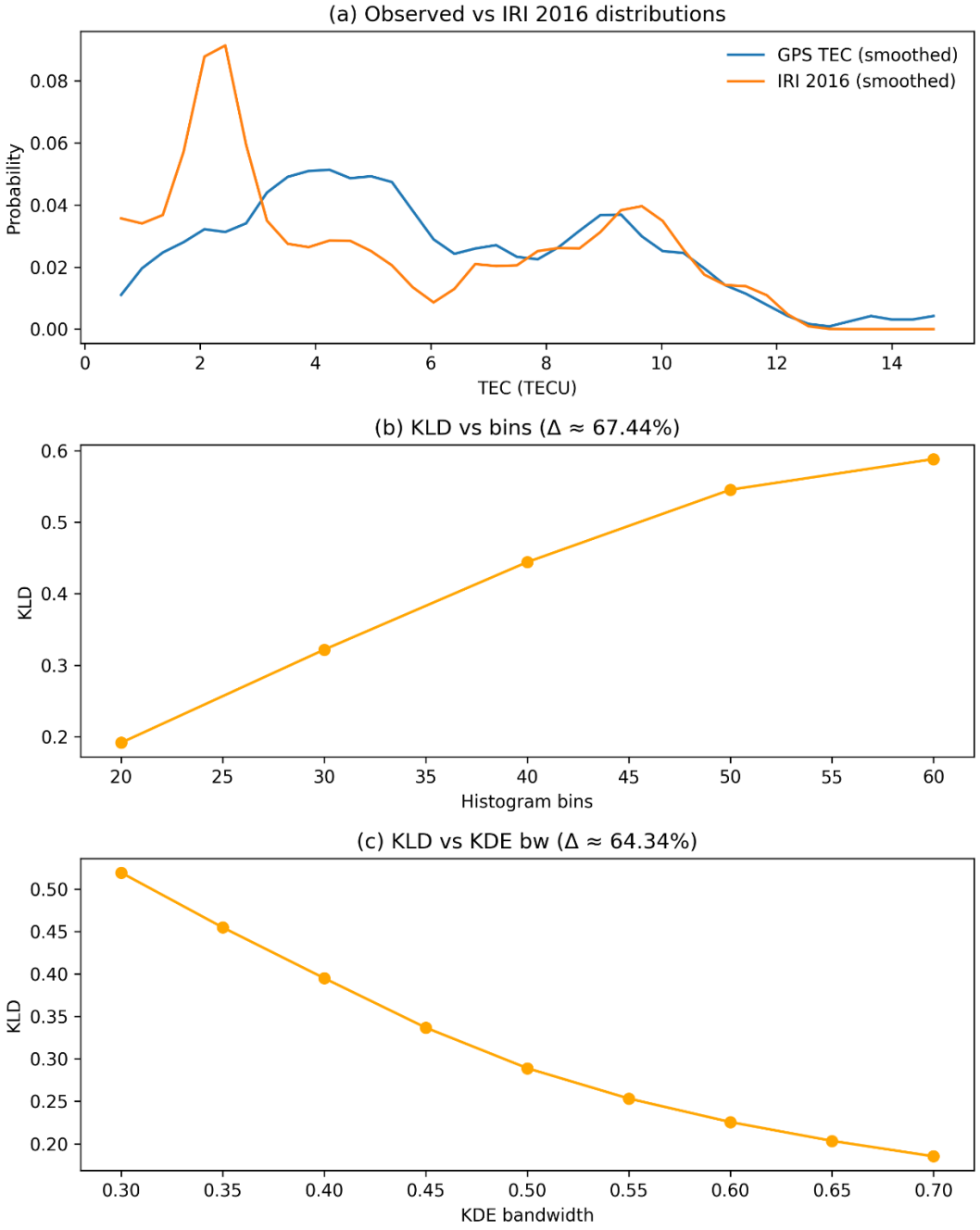

Figure S2. Sensitivity of Kullback–Leibler Divergence (KLD) for Ambalavao station: IRI-2016 model.

Panel (a) compares the observed GPS-TEC distribution with that of the IRI-2016 model. The model significantly underestimates the dominant peaks seen in the observations (notably around 5–7 TECU), leading to mismatched probability densities across the distribution.

Panel (b) shows that KLD increases steadily with the number of histogram bins ($\Delta \approx 67.44\%$), indicating that the divergence between the observed and modeled distributions becomes more pronounced at finer binning scales.

Panel (c) illustrates that KLD decreases systematically with increasing KDE bandwidth ($\Delta \approx 64.34\%$), suggesting that smoothing reduces, but does not eliminate, the structural differences between the two distributions. Together, the sensitivity results confirm that the mismatch is consistent and not an artifact of the histogram or KDE parameter choice.

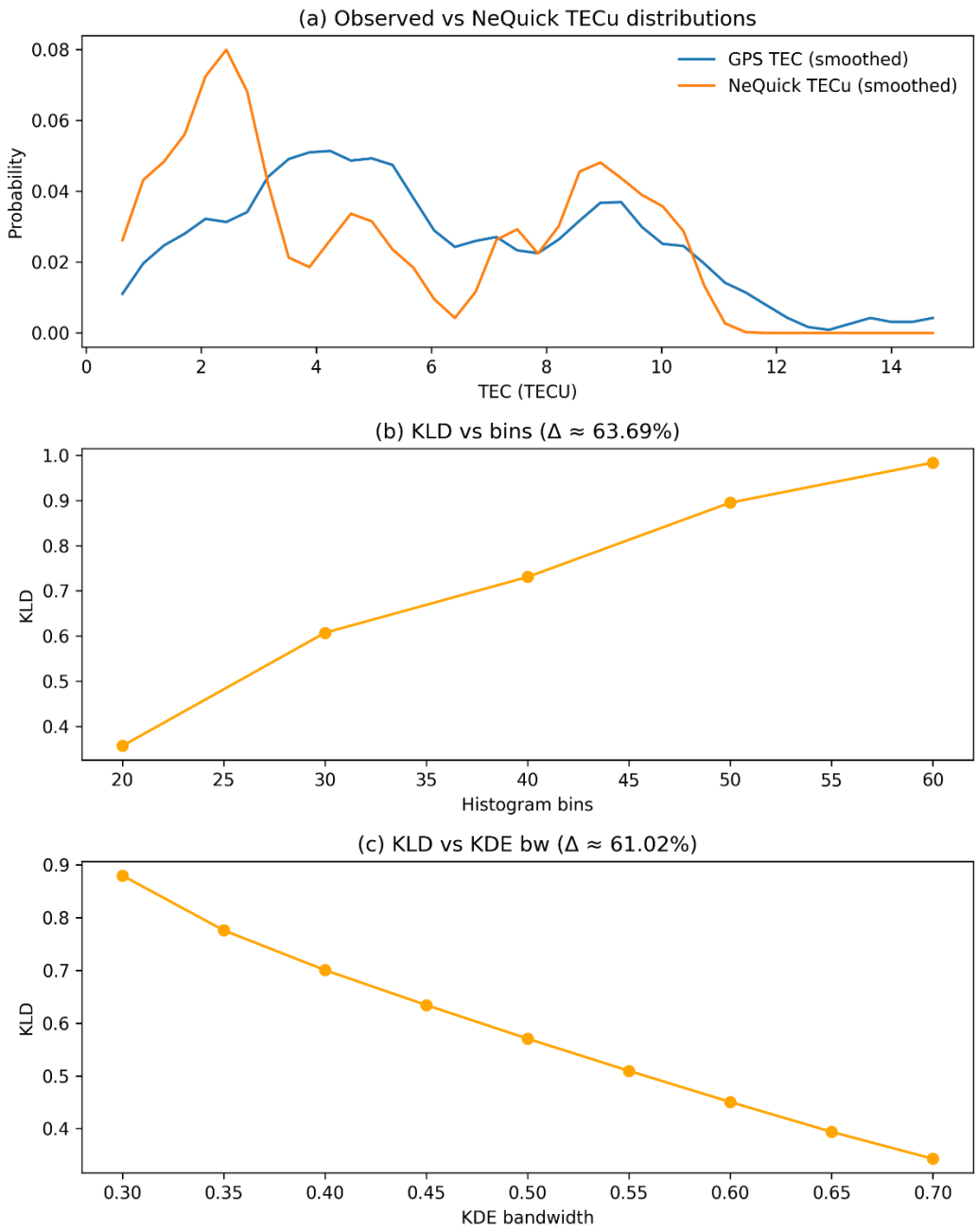

Figure S3. Sensitivity of Kullback–Leibler Divergence (KLD) for Ambalavao station: NeQuick2 model.

Panel (a) compares the observed GPS-TEC distribution with NeQuick2 output. The NeQuick2 distribution is shifted toward lower TEC values, peaking at approximately 2–4 TECU, whereas observations show a broader distribution with multiple modes extending beyond 10 TECU. This structural mismatch highlights that NeQuick2 systematically underestimates the higher-TEC tail observed during storm-time.

Panel (b) shows how KLD increased with the number of histogram bins (20–60), with $\Delta \approx 63.69\%$. This indicates that a finer resolution amplifies the divergence between the model and observed distributions.

Panel (c) displays the effect of the KDE bandwidth variation (0.30–0.70 TECU), where KLD decreases steadily with increasing bandwidth ($\Delta \approx 61.02\%$). The relatively high Δ across both parameters underscores the fact that NeQuick2 is sensitive to methodological choices and exhibits stronger divergence from GPS-TEC than IRI-based models.

Supplementary Captions – Djibouti

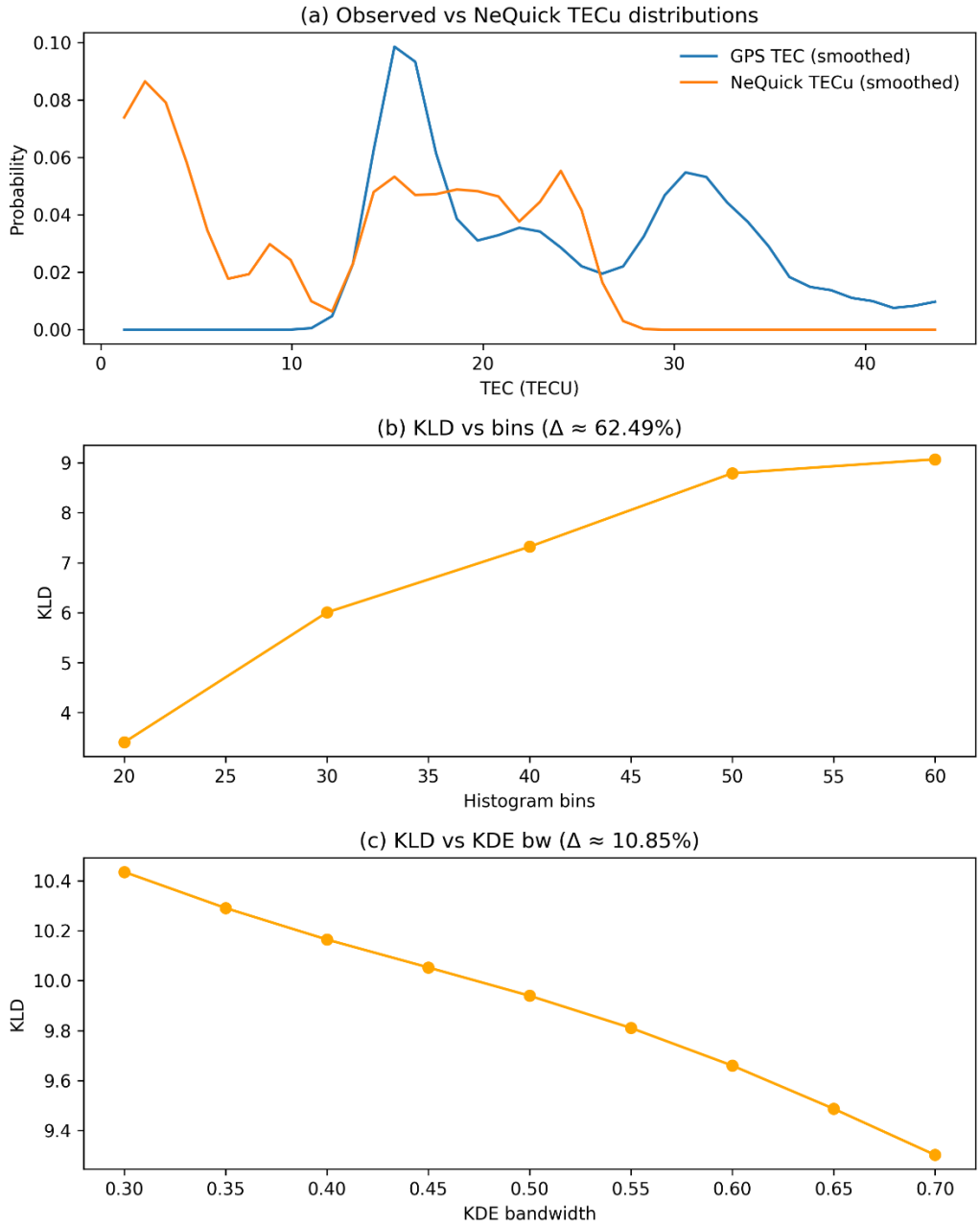

Figure S4. Sensitivity of Kullback–Leibler Divergence (KLD) for the Djibouti station: IRI-2016 PLAS model.

Panel (a) shows that the IRI-2016 PLAS distribution (orange) substantially underestimates the higher-TEC tail compared with the GPS observations (blue), with visible mismatches around 10–20 TECU, where the observed distribution retains elevated probabilities. Panel (b) demonstrates that KLD steadily increases as the histogram bin count increases from 20 to 60, resulting in a relative sensitivity of $\Delta \approx 65.54\%$. Panel (c) reveals a systematic decline in KLD with increasing KDE bandwidth (0.30–0.70 TECU), with $\Delta \approx 58.36\%$. These results indicate that, while the model diverges notably from observations, the sensitivity patterns remain consistent across parameter choices, highlighting the robustness of the detected structural mismatches.

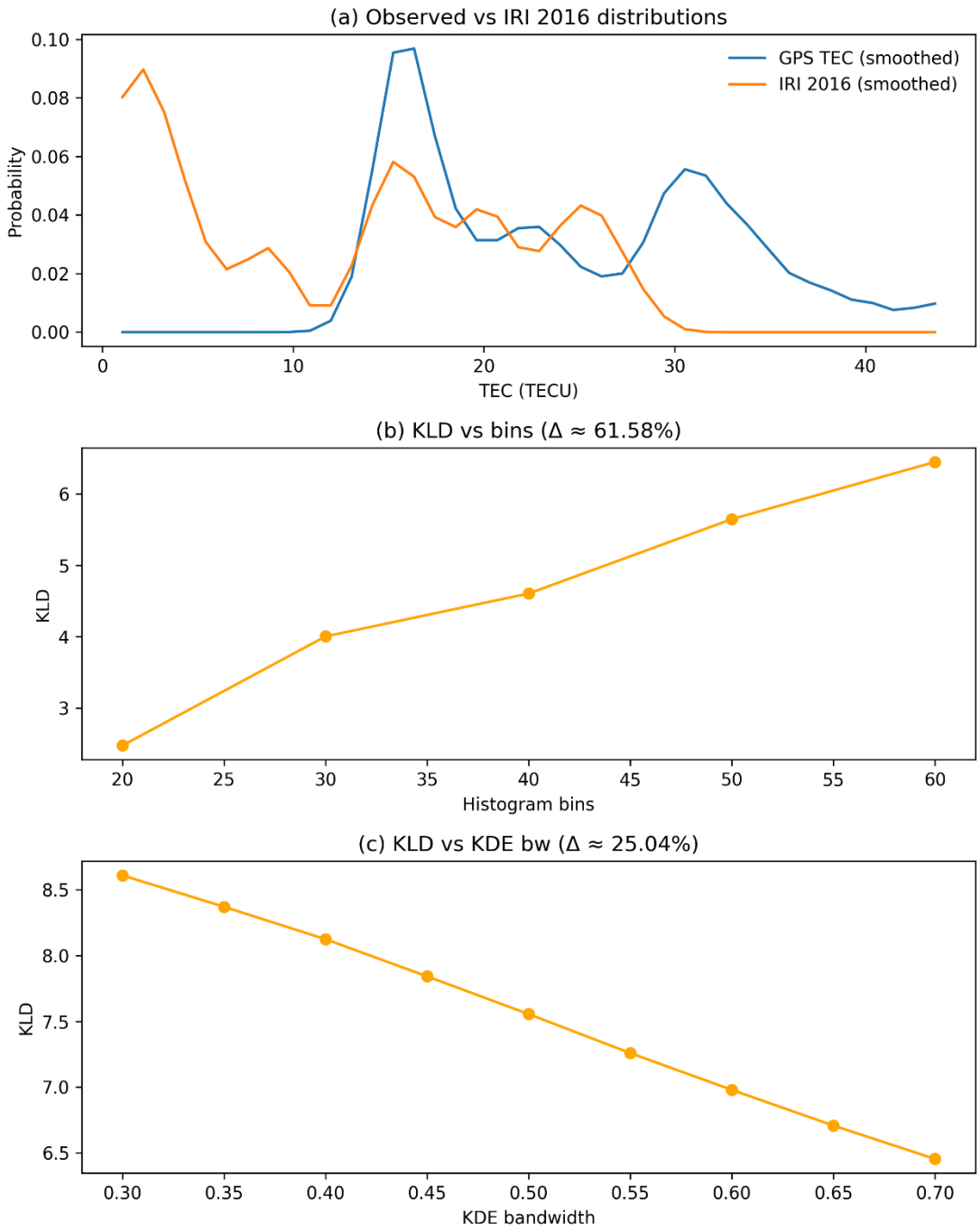

Figure S5. Sensitivity of Kullback–Leibler Divergence (KLD) for the Djibouti station: IRI-2016 model.

Panel (a) shows a comparison between the observed GPS-TEC and the IRI-2016 model distribution. The model consistently underestimates the higher-TEC portion, producing a narrower distribution and shifting the peak around 2–3 TECU compared with observations peaking between 12 and 15 TECU. Panel (b) illustrates that the KLD values increase steadily with histogram bin counts, with a relative variation of $\Delta \approx 61.58\%$, indicating a substantial sensitivity to binning

resolution. Panel (c) demonstrates that the dependence on the KDE bandwidth is less pronounced ($\Delta \approx 25.04\%$), showing a monotonic decrease in the KLD with wider smoothing. Together, these results suggest that the distributional mismatch between IRI-2016 and observations at Djibouti is robust, but strongly influenced by histogram binning.

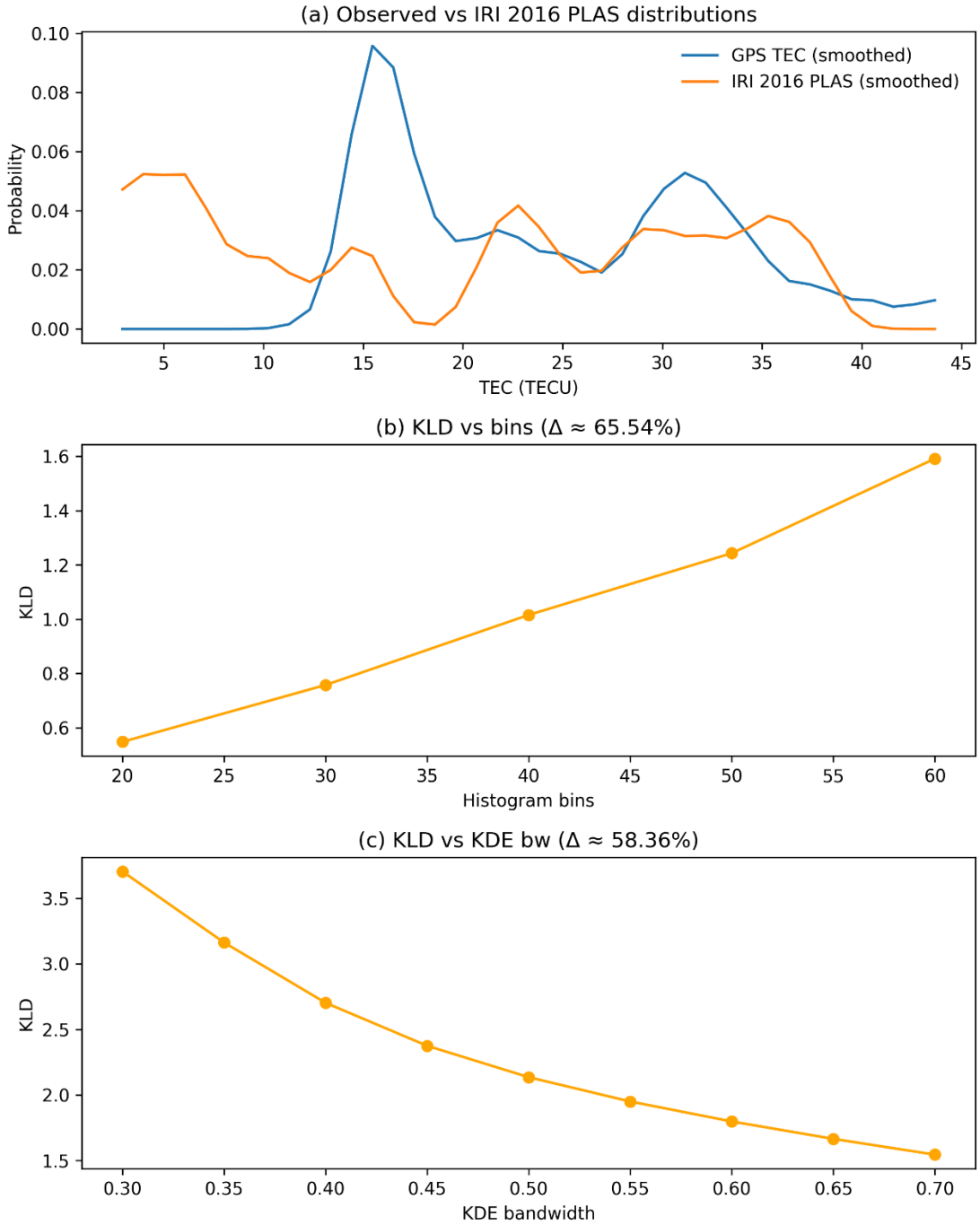

Figure S6. Sensitivity of Kullback–Leibler Divergence (KLD) for the Djibouti station: NeQuick2 model.

Panel (a) shows the GPS-TEC distribution compared to the NeQuick2 model output, where the model systematically underestimates the higher-probability modes of the observed distribution. Panel (b) illustrates how KLD increases markedly as the histogram bin count increases from 20 to 60, with a relative variation of $\Delta \approx 62.49\%$, indicating a strong sensitivity to bin choice. Panel (c) demonstrates the dependence of KLD on the KDE bandwidth (0.30–0.70 TECU), where KLD steadily decreases with smoother bandwidths, but with a smaller relative variation ($\Delta \approx 10.85\%$). Together, these results confirm that NeQuick2 diverges significantly from the observations, although its sensitivity to KDE smoothing is less pronounced than that of bin selection.

Supplementary Captions – Haifa

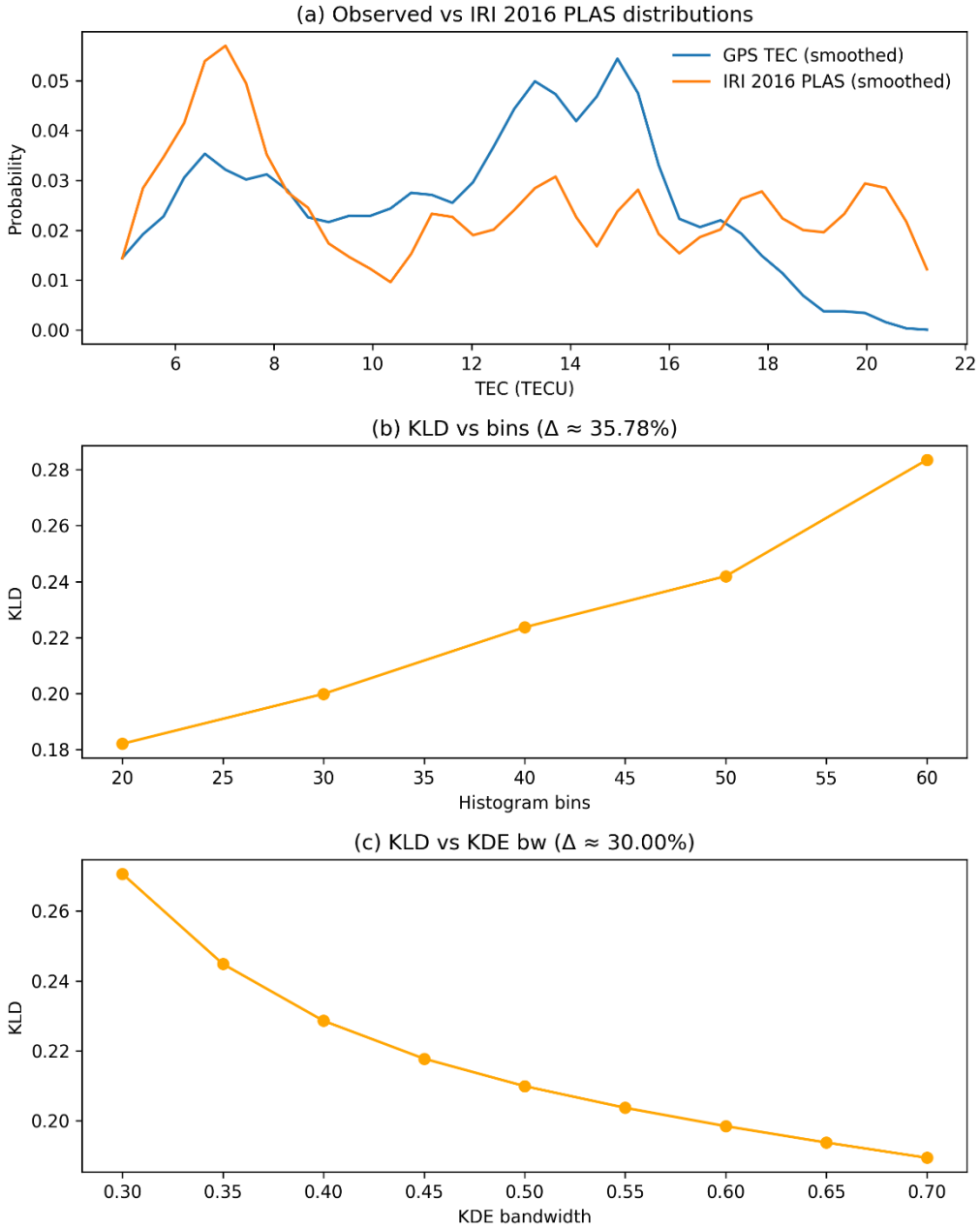

Figure S7. Sensitivity of Kullback–Leibler Divergence (KLD) for the Haifa station: IRI-2016 PLAS model.

Panel (a) shows the observed GPS-TEC distribution compared to the IRI-2016 PLAS output. The model systematically underestimated the main peaks observed in the measurements, with the largest divergence at approximately 13–15 TECU. Panel (b) illustrates how KLD increases with the histogram bin count, showing a relative sensitivity of approximately $\Delta \approx 35.78\%$. Panel (c) demonstrates that KLD decreases consistently as the KDE bandwidth increases (0.30–0.70 TECU), with $\Delta \approx 30.00\%$. These results indicate that while the numerical sensitivity to parameterization is moderate, the structural mismatch between the model and data persists across settings.

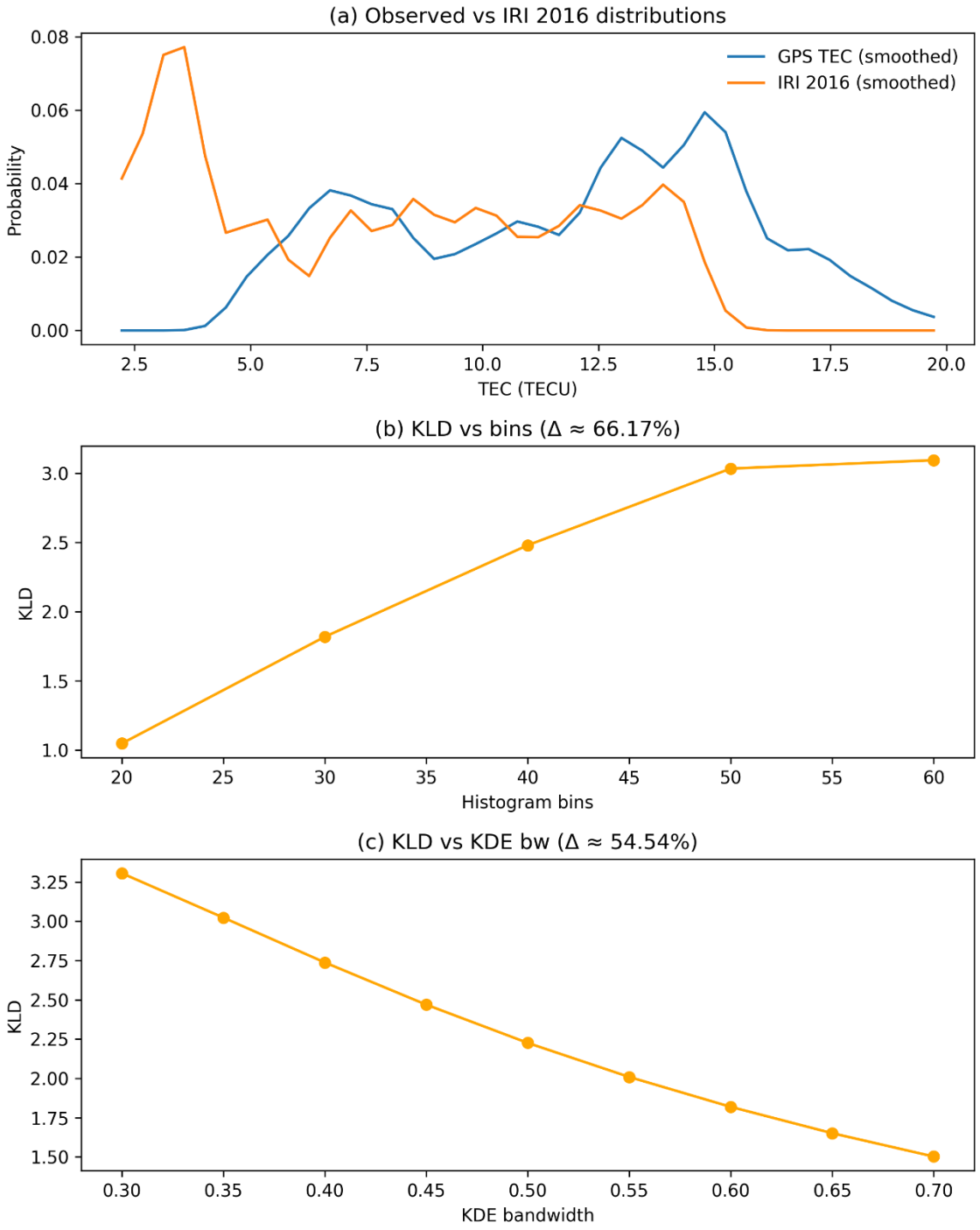

Figure S8. Sensitivity of Kullback–Leibler Divergence (KLD) for the Haifa station: IRI-2016 model.

Panel (a) shows the probability distributions of GPS-TEC and IRI-2016 outputs. The IRI-2016 distribution underestimates the main observed peak (approximately 14 TECU) while overemphasizing lower TEC values (~2–4 TECU), leading to clear structural mismatches. Panel (b) demonstrates that KLD increases steadily with the histogram bin count ($\Delta \approx 66.17\%$), indicating

strong sensitivity to bin resolution. Panel (c) illustrates a consistent decrease in KLD with larger KDE bandwidths (Δ ≈ 54.54%), showing that smoothing reduces the divergence, but does not eliminate the discrepancy. These results highlight that the IRI-2016 model failed to reproduce the observed distributional shape at Haifa, with parameter sensitivity analyses confirming the robustness of this mismatch.

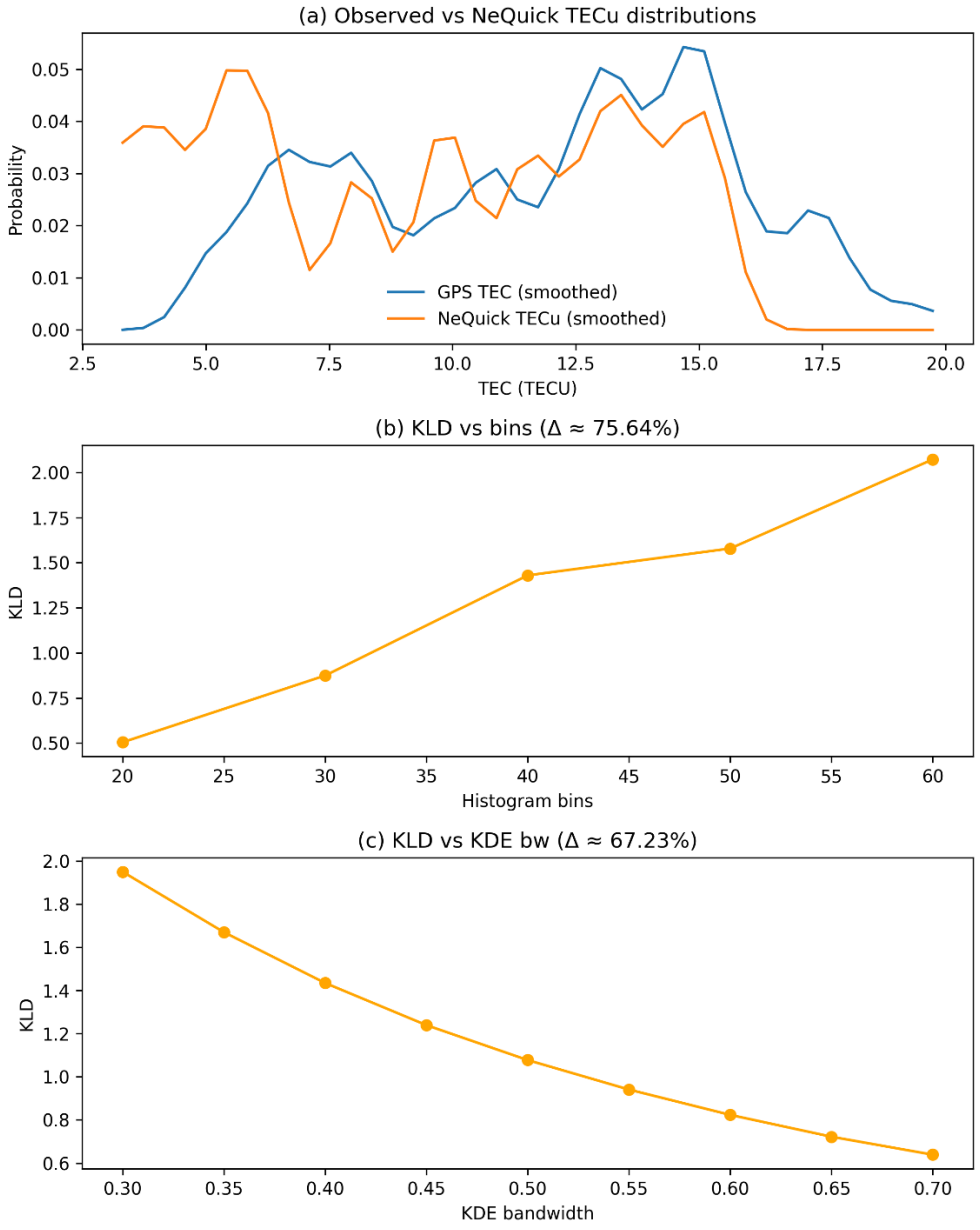

Figure S9. Sensitivity of Kullback–Leibler Divergence (KLD) for the Haifa station: NeQuick2 model.

Panel (a) compares the GPS-TEC probability distribution with NeQuick2 model output. The NeQuick2 curve shows systematic deviations from observations, particularly in the mid-range TEC values (6–15 TECU), where the observed distribution exhibits more spread and secondary peaks

than the model does not fully reproduce. Panel (b) illustrates the dependence of KLD on the histogram bin count (20–60), revealing a strong upward trend with $\Delta \approx 75.64\%$, which indicates the high sensitivity of the divergence metric to the binning resolution. Panel (c) presents the variation in KLD with KDE bandwidth (0.30–0.70 TECU). $\Delta \approx 67.23\%$ further demonstrates the sensitivity of the model–data divergence to parameterization. Together, these results suggest that NeQuick2 exhibits both distributional mismatches and higher parameter sensitivity than IRI-based models, underscoring its limitations at Haifa during the analyzed conditions.

Supplementary Captions – Malindi

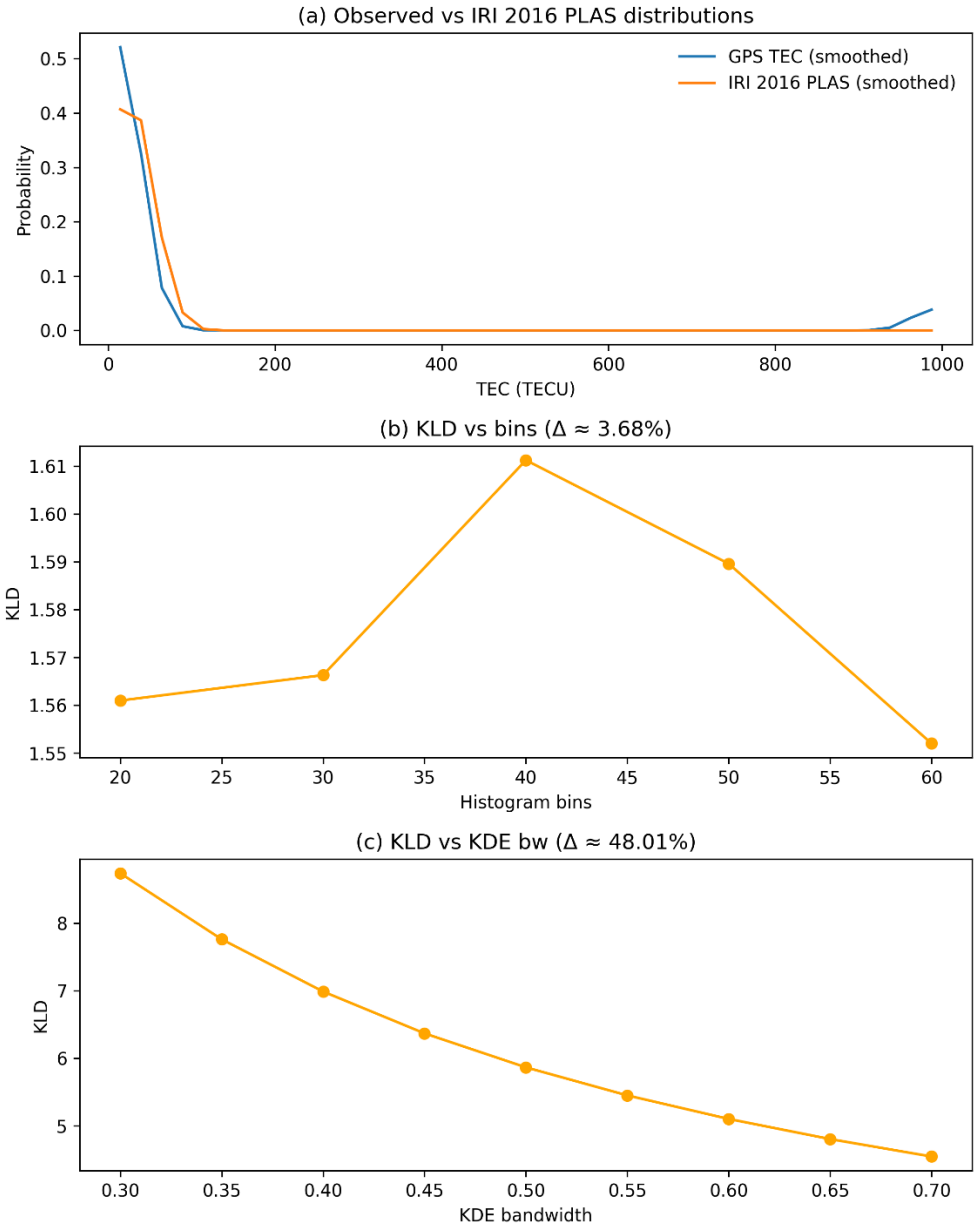

Figure S10. Sensitivity of Kullback–Leibler Divergence (KLD) for the Malindi station: IRI-2016 PLAS model.

Panel (a) shows the probability distributions of the GPS-TEC compared with the IRI-2016 PLAS output. While the two curves overlap at a very low TEC (<50 TECU), the model underestimates the occurrence probability at intermediate ranges and oversmoothes the observed multi-peaked structure. Panel (b) demonstrates how the KLD increases steadily with the histogram bin count, yielding $\Delta \approx 3.68\%$, which indicates a relatively low sensitivity to binning choices. Panel (c) depicts the effect of KDE bandwidth variation, where KLD decreases from ~8.5 at bw = 0.30 to ~4.5 at bw = 0.70, corresponding to a $\Delta \approx 48.01\%$. This highlights that bandwidth selection can substantially affect divergence values, although the consistent model–data mismatch across parameter settings confirms the robustness of structural disagreement.

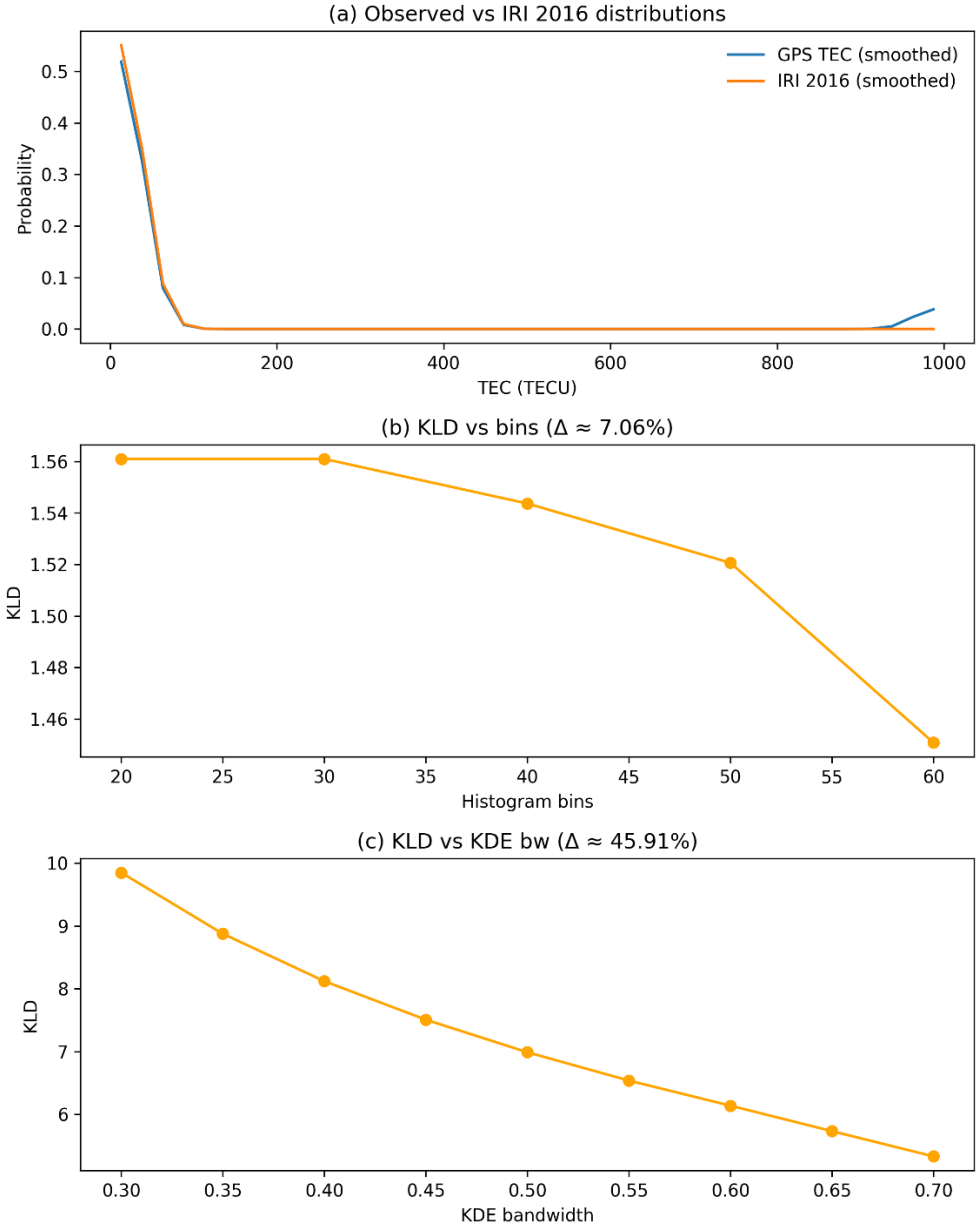

Figure S11. Sensitivity of Kullback–Leibler Divergence (KLD) for the Malindi station: IRI-2016 model.

Panel (a) shows the probability distributions of GPS-TEC compared to the IRI-2016 simulations, where both curves overlap closely, particularly at lower TEC values. Panel (b) presents the variation in KLD with histogram bin counts (20–60), indicating a modest relative change ($\Delta \approx 7.06\%$), which suggests stable behavior across binning schemes. Panel (c) demonstrates the effect of the KDE bandwidth choice (0.30–0.70 TECU), where KLD decreases steadily with increasing bandwidth, reflecting smoother density estimates. The overall sensitivity remained moderate, confirming that the observed distributional mismatches were consistent and not dominated by parameter choices.

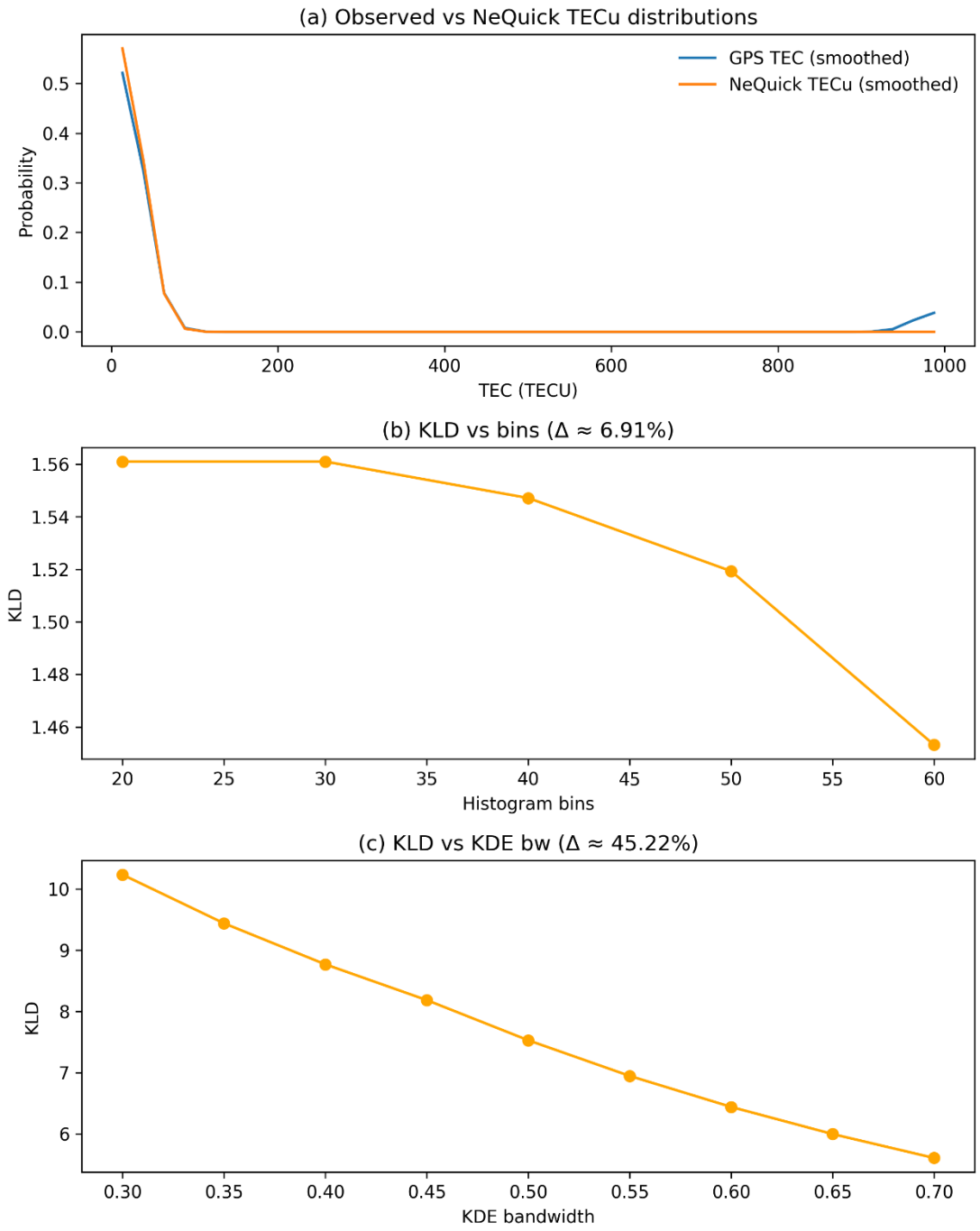

Figure S12. Sensitivity of Kullback–Leibler Divergence (KLD) for the Malindi station: NeQuick2 model.

Panel (a) shows the probability distributions of GPS-TEC and NeQuick2, which align closely at low TEC values (<100 TECU) but diverge slightly in the higher-tail region. Panel (b) illustrates that KLD increases moderately with the number of histogram bins, from ~1.56 at 20 bins to ~1.59 at 60 bins, with a relative change of $\Delta \approx 6.91\%$, indicating a limited binning sensitivity. Panel (c) demonstrates a stronger dependence of KLD on KDE bandwidth, with values decreasing from

~10.2 at bandwidth 0.30 to ~5.6 at 0.70, corresponding to $\Delta \approx 45.22\%$. These results highlight that while NeQuick2 reproduces the overall distributional structure well, its divergence from GPS-TEC estimates is more strongly affected by kernel-smoothing choices than by bin resolution.