

Motivation: The Detector Systematic Uncertainty Challenge

Systematic uncertainty studies for LArTPC analyses require dedicated simulated samples at each detector configuration $\theta \in \Theta \subseteq \mathbb{R}^d$.

- For rare-event searches ($n-\bar{n}$ oscillation), background rejection inefficiency $\sim 10^{-4}$ \Rightarrow large statistics needed per configuration
- Producing high-fidelity (HF) samples across all θ is computationally prohibitive

Key Insight

Rather than running HF simulations everywhere, combine:

- few HF simulations: large statistics and act as ground-truth anchors
- many LF simulations: smaller statistics but broader coverage of Θ .

DetSuM = **Detector Surrogate Model**, version of RESuM¹ applied to LArTPC detector systematics. Goal is to learn $f^{\text{HF}}: \theta \rightarrow (\hat{y}(\theta), \hat{\sigma}(\theta))$ evaluated at any θ at a fraction of the full simulation cost.

¹Schuetz, Poon, & Li, ICLR 2025 [arXiv:2410.03873].

Problem Setup

Design Metric

Empirical signal efficiency per simulation:

$$y_k := \frac{1}{N_k} \sum_{i=1}^{N_k} X_{ki}$$

where $X_{ki} \in \{0, 1\}$ indicates if the i^{th} event in the k^{th} simulation passes $n-\bar{n}$ score cut.

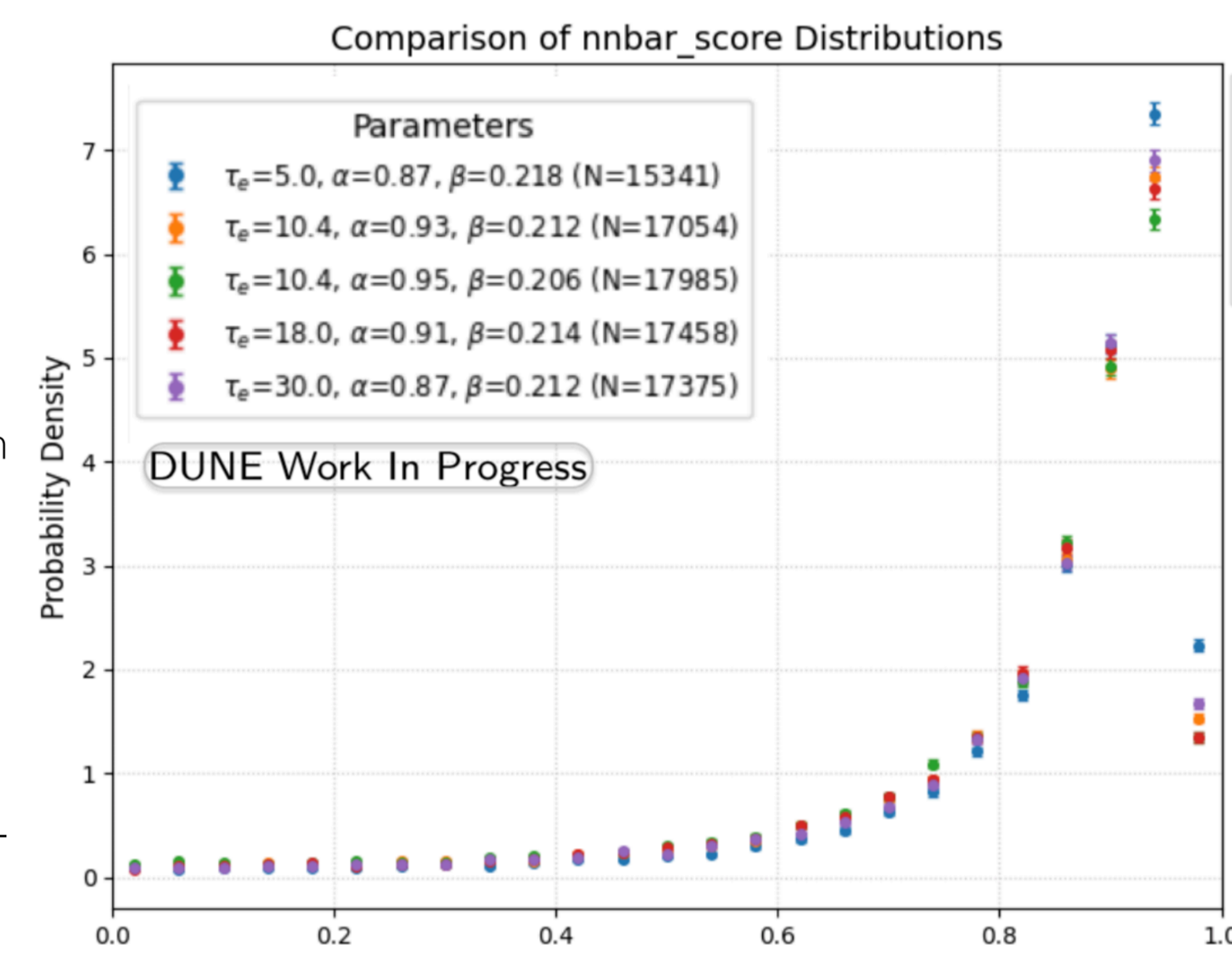


Figure: Distribution of $n-\bar{n}$ scores across five different HF simulation θ configurations.

Demonstration w/ DUNE Detector Variation Parameters ($d = 3$)

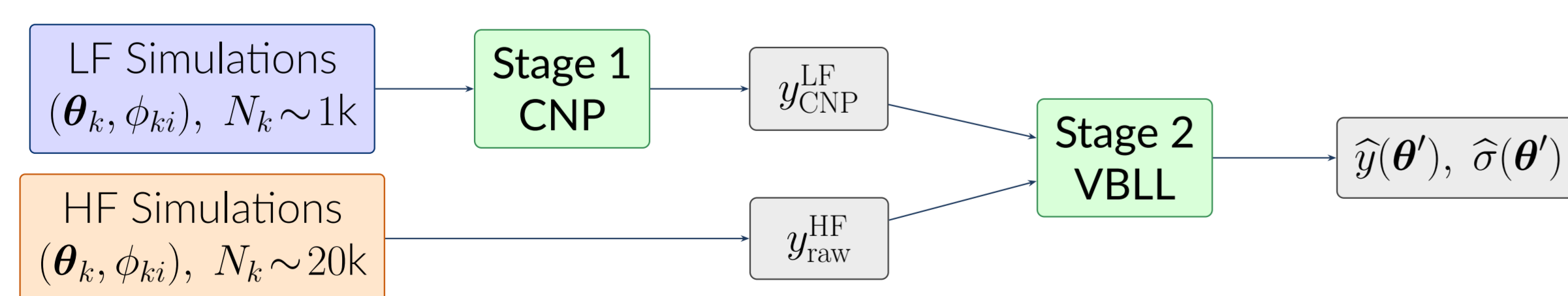
Symbol	Meaning	Min	CV	Max
τ_e	Electron lifetime (in ms)	5.0	10.4	30.0
α	Recombination	0.87	0.93	0.99
β'	Recombination	0.206	0.212	0.218

Data Split²

Sample	Configs	Evts/cfg	Total
HF training	20	$\sim 20k$	400k
LF training	100	$\sim 1k$	100k
HF validation	25	$\sim 20k$	500k

²Sample produced using 2025 implementation of classification of $n-\bar{n}$ signal from atmospheric neutrino background, not the SoTA transformerCVN update from 2026; this update will be presented by Wan & Yu at Neutrino 2026 (poster #57 on Jun. 25th).

DetSuM: Two-Stage Architecture



Stage 1 – CNP (event-level regression): Maps $(\theta_k, \phi_{ki}) \rightarrow \beta_{ki} \in [0, 1]$, a calibrated per-event signal probability. Aggregated: $y_{\text{CNP}}^{\text{LF}} = \frac{1}{N_k} \sum_i \beta_{ki}$ – a smooth, continuous surface used as prior in Stage 2.

Stage 2 – VBLL (parameter-level regression): Predictive distribution

$$p(y | \theta) = \mathcal{N}(\bar{\psi}^T \psi(\theta), \psi(\theta)^T S \psi(\theta) + \Sigma),$$

where $\psi: \mathbb{R}^3 \rightarrow \mathbb{R}^{32}$ is a *learned non-linear feature map*; $\bar{\psi}$ is the variational posterior mean; S its covariance; Σ observational noise.

Training: (1) Pre-train ψ on all LF + HF data (MSE); (2) freeze ψ , train VBLL head on HF only (ELBO). Training is light-weighted, but can be updated and scaled given newer and more extensive sample.

DUNE Parameter Rescaling

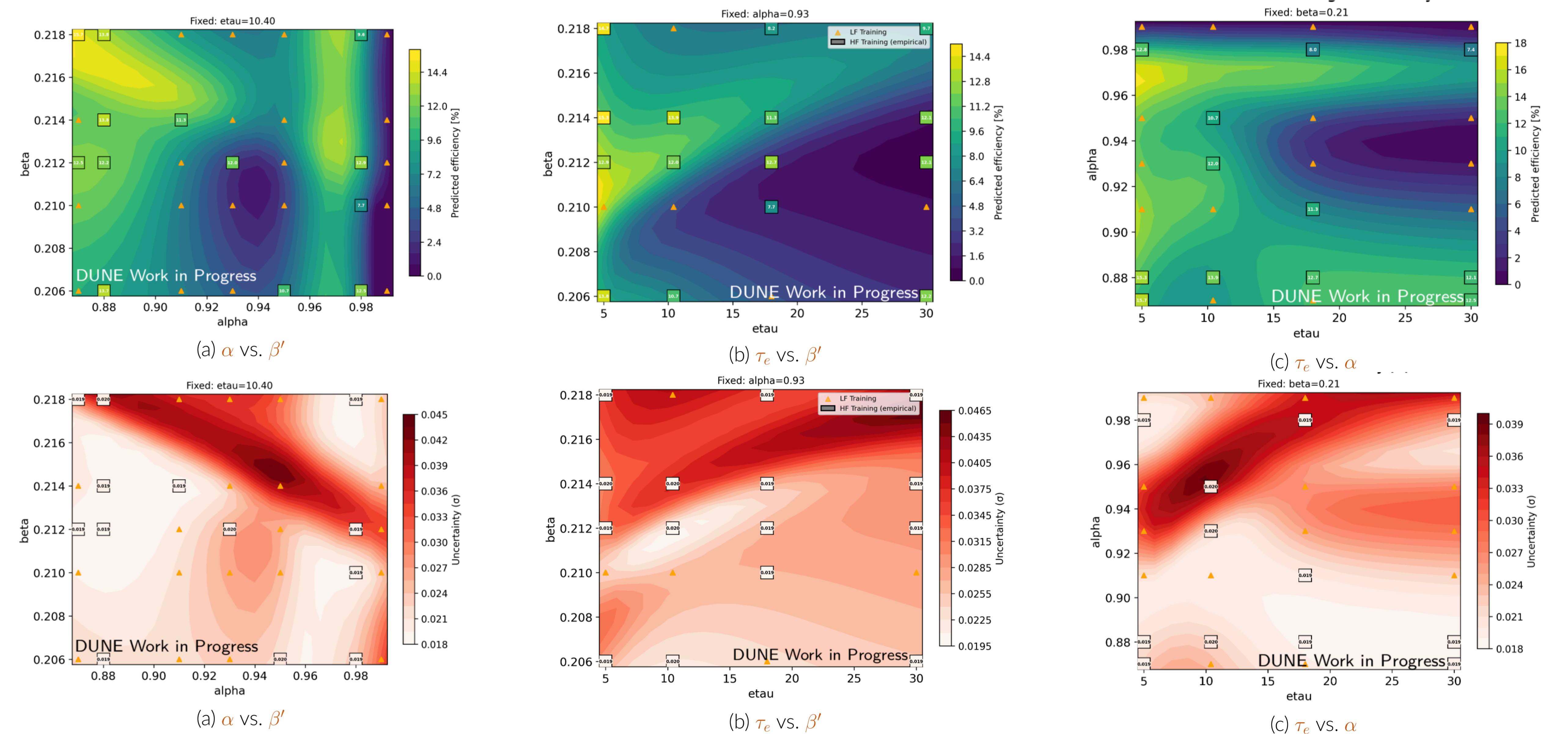
Raw ranges differ in scale and physical distribution. All parameters normalized to $[0, 1]$ before training:

- τ_e : **log-transform** then min-max (accounts for non-linear sensitivity at small values)
- α, β' : **min-max normalization** directly

Ensures comparable gradient signal across all dimensions and geometrically consistent uncertainty coverage in $\hat{\sigma}(\theta)$.

Mean Signal Efficiency and Uncertainty Predictions Across Parameter Space

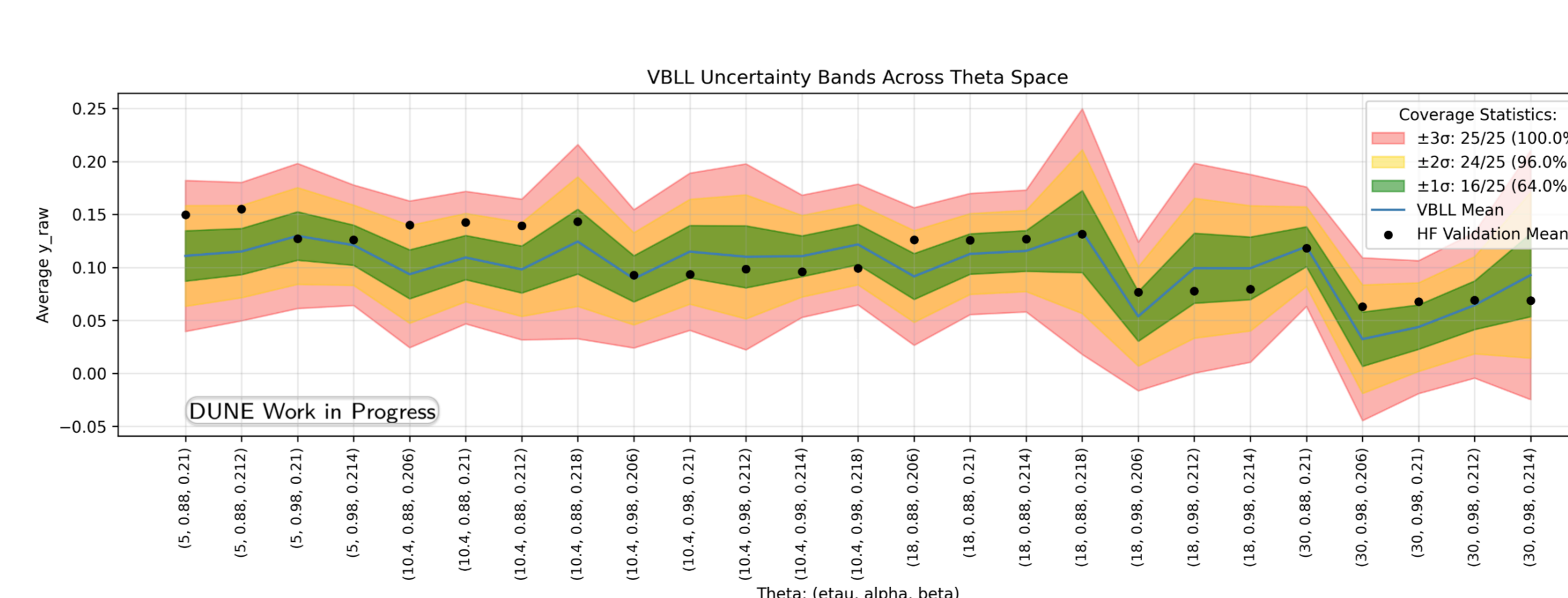
Each panel is a 2D slice through θ space with the third parameter fixed at its central value (CV). Blue/green color pattern encodes predicted signal efficiency $\hat{y}(\theta)$ (%), while reddish color pattern encodes predictive uncertainty $\hat{\sigma}(\theta)$. \blacktriangle LF training points (orange); \square HF training points (color/black).



DetSuM produces smooth, non-linear predicted efficiency contours across all three 2D subspaces. Predicted range is 0–18%, consistent with the range of y_{raw} at the HF training locations. Uncertainty is mostly consistent with the expected propagation of data sparsity given the HF samples as ground-truth anchors. Non-linear contour boundaries reflect the learned non-stationary feature map $\psi(\theta)$.

Uncertainty Quantification: Statistical Coverage on HF Validation

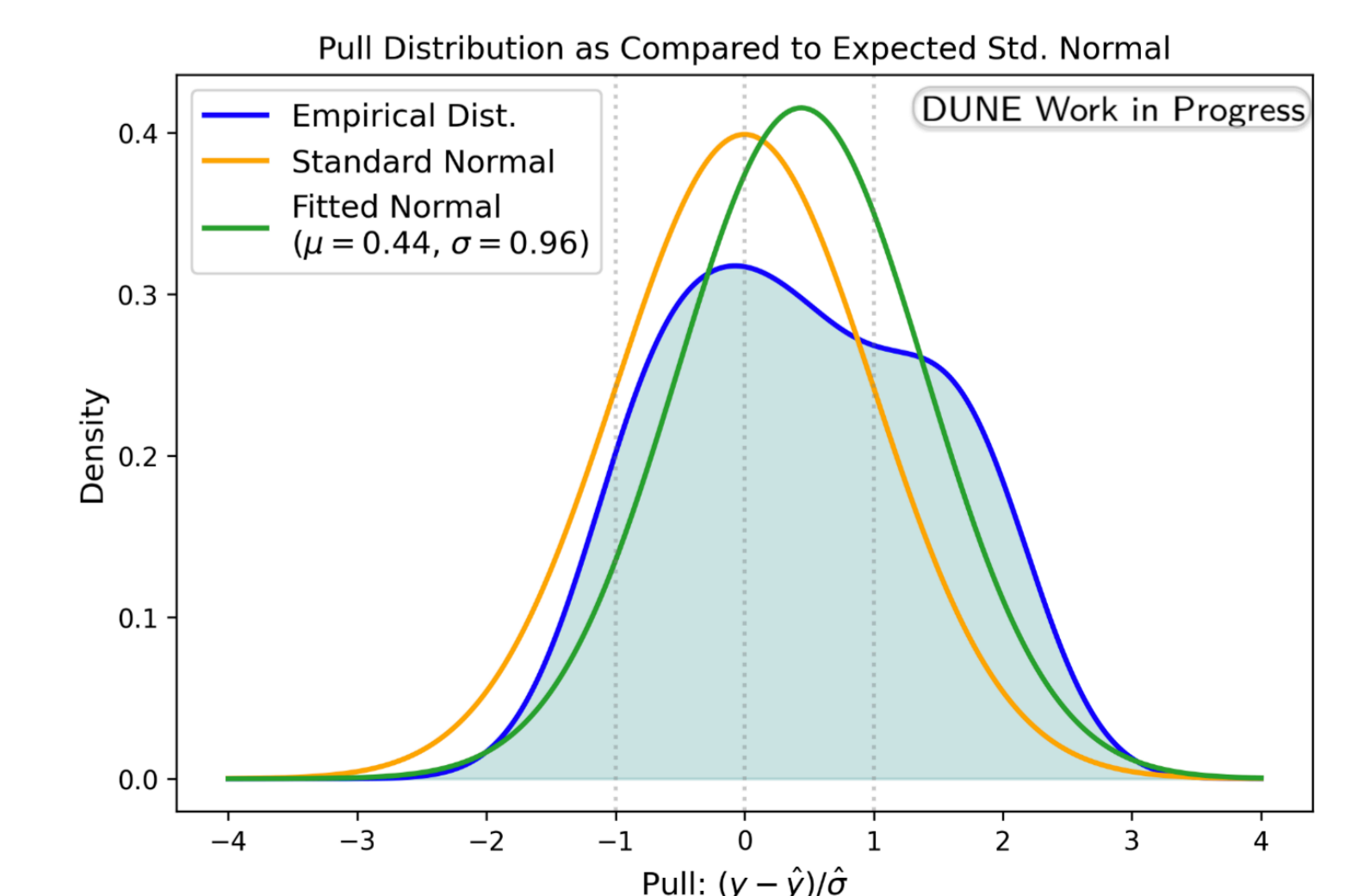
Evaluated on 25 HF validation configurations (unseen during training). VBLL $\pm 1, 2, 3\sigma$ bands (in color) compared to empirical $y_{\text{raw}}^{\text{HF}}$.



Statistical coverage of mean + uncertainty predictions across HF validation θ configurations.

Band	Exp. %	Exp. count	Obs. %	Obs. count
$\pm 1\sigma$	68.3%	17.1	64.0%	16
$\pm 2\sigma$	95.4%	23.9	96.0%	24
$\pm 3\sigma$	99.7%	24.9	100.0%	25

Table 3. Comparison of expected vs. observed statistical coverage.



Standardized residuals for HF validation θ configurations.

Tuned VBLL achieves **near-calibrated posterior uncertainty**: $\pm 2\sigma$ and $\pm 3\sigma$ statistical coverage within 1% of theoretical values, $\pm 1\sigma$ within 5% of theoretical value, and fitted uncertainty is 0.96 instead of 1.00. Positively-fitted pull distribution mean suggests predicted signal efficiency is under-reported.

Acknowledgement. This work was produced by Fermi Forward Discovery Group, LLC under Contract No. 89243024CSC000002 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics.