CSI Simulation: Why Additive Noise Fails and How to Fix It

Aymen Bouferroum
Inria Lille-Nord Europe
Lille, France
0009-0002-9133-3152

Ildi Alla
University of Luxembourg
Luxembourg, Luxembourg
0009-0008-6290-6998

Vincent Lenders
University of Luxembourg
Luxembourg, Luxembourg
0000-0002-2289-3722

Valeria Loscri
Inria Lille-Nord Europe
Lille, France
0000-0003-2558-1801

***Abstract*—Channel State Information (CSI) has become a widely used wireless channel sensing modality for applications such as indoor localization, activity recognition, and respiration monitoring. Because collecting labeled data under every target condition is impractical, training CSI-based models often relies on simulated data produced by adding noise or perturbations to recorded channel estimates, most commonly additive white Gaussian noise (AWGN). This practice assumes that the receiver chain between the antenna and the channel estimator is *linear and gain-invariant*. We test this assumption empirically using RF jamming as a controlled perturbation on 6 commodity receivers across 2 indoor environments. The assumption does not hold. Automatic gain control compresses the channel estimate multiplicatively before digitization, producing amplitude distributions that no additive noise variance can reproduce. To close the resulting fidelity gap, we propose $\mathbf{M_{QTC}}$, a measurement-calibrated model that learns the per-subcarrier distribution transformation through quantile mapping, temporal filtering, and copula-based cross-subcarrier reordering. $\mathbf{M_{QTC}}$ reduces amplitude error 8-fold and closes 89% of the aggregate fidelity gap across four complementary dimensions. The improvement transfers directly to downstream tasks, where 5 classifiers from different families trained on $\mathbf{M_{QTC}}$-simulated data recover 93% of real-data jamming detection performance, while AWGN-trained classifiers remain near chance.**

***Index Terms*—Channel State Information; simulation validation; receiver chain; Wi-Fi sensing; sim-to-real transfer; data augmentation**

I. Introduction

Channel State Information (CSI) captures the complex-valued frequency response that an orthogonal frequency-division multiplexing (OFDM) receiver estimates on every decoded frame as part of channel equalization. Because these per-subcarrier estimates capture fine-grained amplitude and phase distortions imposed by the propagation environment, CSI has become a general-purpose sensing modality for indoor localization, human activity recognition, gesture detection, and respiration monitoring [1], [2]. CSI extraction is now supported across commodity Wi-Fi chipsets [3]–[5], enabling broad deployment for sensing research.

Many of these applications depend on simulated or synthetically augmented CSI for model training. CrossSense generates synthetic CSI from a single measurement set to enable cross-site sensing [6]. Noise-based perturbation of recorded CSI is a standard augmentation strategy in gesture recognition [7], activity recognition [8], indoor localization [9], and 5G positioning [10]. These approaches inherit, explicitly or implicitly, the textbook additive white Gaussian noise (AWGN) model [11]. Interference is added to a clean signal in the complex domain, where the model is well founded for raw baseband signals. In-phase and quadrature (I/Q)-level tools such as JamRF [12] generate interference waveforms along the same lines. The *critical assumption* is that what holds for raw baseband signals also holds for CSI after receiver processing. This assumption has not been empirically tested. Generative models bypass the additive formulation by learning CSI distributions directly [13], [14], but require large training sets and have not been validated against controlled measurements. A recent study on Wi-Fi sensing augmentation concluded that no prior work had systematically explored radio data augmentation, and that existing approaches are ad-hoc [15].

Before a received signal becomes a channel estimate, it passes through automatic gain control (AGC), analog-to-digital conversion (ADC), a fast Fourier transform (FFT) stage, and channel estimation from known training symbols [16]. Each stage introduces nonlinear distortions that an additive model does not capture. When interference raises total received power, AGC reduces the analog front-end gain before digitization, compressing the entire channel estimate by a factor that depends on the instantaneous power level [17], [18]. This compression is multiplicative and acts on signal and noise together, so the amplitude distribution reported by the receiver has a fundamentally different shape than what an additive model predicts. Baseband filtering adds further subcarrier-dependent distortions [19], and a controlled cable-coupled comparison of commercial receivers recently confirmed that AGC behavior is the dominant source of cross-device CSI variation [20]. The additive simulation model ignores all of these stages. Whether the resulting CSI distributions are realistic is therefore not a modeling detail but a testable empirical question.

RF jamming provides an ideal case study for this test. A jammer injects a controlled waveform of known power and bandwidth, producing a measurable perturbation at the CSI level that can be directly compared against simulation predictions. Because jamming raises total received power substantially, it forces a large AGC response and makes any deviation from the additive model clearly visible. Jamming also carries practical relevance. CSI-based interference detectors need realistic training data, yet collecting paired clean and jammed

recordings requires controlled environments and authorization to transmit interference [21], [22]. Recent empirical work confirmed that CSI carries exploitable jamming signatures [23], and simulation-based jamming detectors have been proposed without validating the fidelity of the underlying simulation [24]. In developing our own CSI-based jamming detection and classification pipeline [25], we observed that simulated CSI distributions diverge substantially from real receiver output, motivating the systematic investigation presented in this work. The *sim-to-real gap* in RF machine learning is a recognized problem across localization [26] and signal classification [27], but it has not been quantified at the CSI level.

The contributions of this work are as follows.

- Empirical falsification of the receiver-chain linearity assumption: the AWGN model produces a Wasserstein distance of 3.09 compared to 0.39 for the proposed model, an 8-fold gap, providing, to our knowledge, the first quantitative test at the CSI level.
- A calibrated simulation pipeline, $\mathrm{M_{QTC}}$ (Quantile-Temporal-Copula), that closes 89% of the aggregate fidelity gap across amplitude, phase, temporal, and spectral dimensions, validated on three external public datasets beyond the jamming scenario. Ablation identifies copula-based cross-subcarrier reordering as the dominant mechanism (9-fold aggregate reduction).
- Sim-to-real transfer validation: classifiers trained on $\mathrm{M_{QTC}}$-simulated data recover 93% of the detection performance achieved with real training data (AUC 0.904 vs. 0.967) across 5 classifier families and 5 receivers, while AWGN-trained classifiers operate near random decision (AUC 0.522).

To our knowledge, the paired clean and jammed CSI recordings released with this work are the first such public dataset. The framework code, dataset, and evaluation scripts are available online.[1]

II. Background and related work

A. CSI simulation practice

The simulation approaches used throughout the CSI literature fall into three categories, none of which explicitly model how the receiver chain transforms the signal before the channel estimator reports it.

The first and most common category is *additive noise injection*. Complex Gaussian noise is added directly to recorded CSI to simulate degraded conditions. Table I catalogues representative work that follows this pattern across five application domains. The underlying model is straightforward. If the true channel is H_k and independent Gaussian noise N_k is added, the simulated output is $\hat{H}_k = H_k + N_k$. This formulation is correct for raw baseband signals, where the noise enters before any receiver processing. The implicit assumption, rarely stated explicitly, is that the same formulation holds after the receiver chain has processed the signal. Serbetci et al. [9] observed that naive Gaussian injection produces worse localization performance than physics-motivated methods modeling oscillator drift and amplifier fluctuations, but did not identify the receiver chain as the underlying cause.

TABLE I. Representative work adding noise or perturbation directly to CSI. Whether the resulting distributions match real receiver output has not been tested.

Reference	Domain	Noise method
Wang et al. [7]	Gesture recog.	Additive Gaussian
Strohmayer et al. [8]	Activity recog.	Amplitude scaling
Barahimi et al. [29]	Activity recog.	$\mathcal{N}(0, 0.1)$ to CSI
Thariq Ahmed et al. [30]	Sign language	AWGN to raw CSI
Serbetci et al. [9]	Localization	$\mathcal{CN}(0, P)$ at SNR
Gao et al. [10]	5G positioning	Noise at 3 SNR levels
Zhang et al. [6]	Cross-site sensing	Synthetic generation

The second category operates at the *I/Q level* [28]. Tools such as JamRF [12] generate interference waveforms for software-defined radio (SDR) transmission, producing physically correct baseband signals that nevertheless skip the receiver entirely. The gap between I/Q-level and CSI-level simulation has not been measured. A signal that is received at the antenna may produce a very different channel estimate after passing through AGC, digitization, and the FFT.

The third category uses *generative models*. Bhat et al. [13] applied generative adversarial networks (GANs) to synthesize mmWave CSI for pose classification, and Wang et al. [14] used conditioned diffusion models for integrated sensing and communication (ISAC) networks. These approaches learn the data distribution directly but require large training sets and have not been validated against controlled measurements where ground truth is available. All three categories share a *common limitation*. The simulator output has not yet been compared, distribution by distribution, to what a real receiver produces under identical conditions.

B. Receiver-chain effects on CSI

The receiver hardware between the antenna and the channel estimator introduces systematic distortions that have been studied independently but not connected to simulation validity. Ratnam et al. [17] developed a mathematical model for the gain and phase errors that AGC, sampling clock offset, and carrier frequency offset introduce into Wi-Fi CSI, and showed that these errors must be compensated before CSI can be used reliably for sensing. Wei et al. [18] demonstrated that AGC-induced amplitude scaling explains the discrepancy between received signal strength indicator (RSSI) and CSI power measurements on commodity devices. Jiang et al. [19] showed that baseband filters introduce subcarrier-dependent amplitude and phase distortions that are pervasive across Wi-Fi network interface card (NIC) architectures and proposed the PicoScenes platform to expose these effects. Most recently, Portner et al. [20] conducted a controlled cable-coupled comparison of commercial receivers and confirmed that AGC behavior, rather than channel propagation, is the dominant source of cross-device CSI variation. Cominelli et al. [31] demonstrated that CSI sensing performance depends strongly on the deployment environment, with generalization across environments remain-

[1] https://github.com/aymenbouferroum/sim2real-csi

ing an open problem. These studies establish that the receiver chain matters for CSI applications, but none connect it to how CSI is simulated.

C. Jamming detection and the validation gap

The standard jammer taxonomy (constant, deceptive, random, reactive) was established by Xu et al. [21], who detected attacks from aggregated link metrics. Subsequent work continued to operate on link-level or spectrum-level representations. The transition to CSI-level analysis is recent. Mykytyn et al. [23] provided the first CSI-level characterization of jamming effects across static and mobile networks, confirming that jamming produces *detectable per-subcarrier signatures* in both amplitude and phase. Panitsas et al. [24] argued that detectors evaluated on simulated jamming data may not generalize to real over-the-air conditions, but did not trace the gap to the receiver chain or propose a calibrated simulation model.

The *sim-to-real gap* is increasingly recognized across wireless machine learning. Manukyan et al. [26] reported that RF localization models achieving 25 m error on synthetic data degraded to 184 m on real measurements. Clark et al. [27] showed that synthetically trained RF classifiers lose significant accuracy without channel-aware augmentation. These studies confirm that naive simulation is insufficient, but operate at different signal representations (raw I/Q, RSSI, spectrograms) rather than at the CSI level. Validating simulated channels against physical measurements is standard practice in propagation modeling [16], yet such validation has not been performed at the CSI level, and the mechanism that breaks the additive assumption remains unidentified.

III. The Linearity Assumption

CSI simulation rests on one core assumption: the receiver chain between antenna and channel estimator is *linear and gain-invariant*. Under this assumption, adding noise at the CSI output is equivalent to adding it at the antenna input. We state the assumption formally, explain why it fails in practice, and derive the baseline model that embodies it.

An OFDM receiver estimates the channel from known training symbols X_k by observing $Y_k = H_k X_k + N_k$ and computing:

$$\hat{H}_k = \frac{Y_k}{X_k}. \tag{1}$$

When wideband interference adds complex Gaussian noise J_k on every subcarrier, the jammed estimate becomes

$$\hat{H}_k = H_k + \frac{J_k}{X_k}, \tag{2}$$

and since the training symbols have approximately unit magnitude in 802.11 OFDM, J_k/X_k remains complex Gaussian with unchanged variance. If the receiver chain preserves linearity and gain, Equation (2) predicts that interference simply adds independent Gaussian samples to the channel estimate. The standard AWGN approach [11] operationalizes this prediction by adding independent complex Gaussian noise to each subcarrier of a clean recording:

$$\hat{H}_k^{(j)}[n] = H_k^{(c)}[n] + \mathcal{CN}(0,\ \hat{\sigma}^2), \tag{3}$$

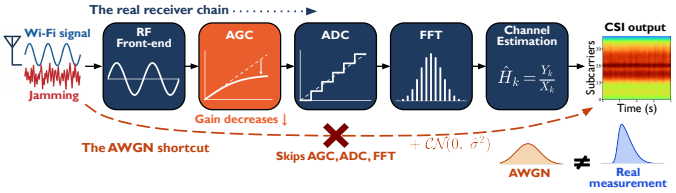

Fig. 1. The receiver chain between the antenna and the CSI output. Real interference enters at the antenna and crosses AGC, digitization, and the FFT before reaching the channel estimator. M_1 adds noise directly at the output, skipping all three stages. The two paths coincide only if the chain is linear with fixed gain.

where the noise variance $\hat{\sigma}^2$ is calibrated from the mean per-subcarrier power difference between jammed and clean recordings:

$$\hat{\sigma}^2 = \frac{1}{K}\sum_{k=1}^{K}\Big(\mathbb{E}\big[|H_k^{(j)}|^2\big] - \mathbb{E}\big[|H_k^{(c)}|^2\big]\Big). \tag{4}$$

This single-parameter model, which we refer to as $\mathbf{M_1}$ (AWGN), treats all subcarriers identically, ignores temporal structure, and assumes amplitude distributions shift additively. It is the principled baseline against which the receiver-chain-aware models are measured.

The receiver, however, is not linear (Fig. 1). Three stages between the antenna and the channel estimator violate the assumption. *First*, AGC adjusts the analog front-end gain based on total received power. When interference raises total power from P_{clean} to $P_{\text{clean}} + P_{\text{jam}}$, AGC reduces the gain by a factor that depends on the power ratio, compressing the entire channel estimate before digitization [11], [17]. This gain reduction acts on signal and interference together. The jammed CSI amplitude changes far less than the injected power would suggest under a linear model, and the distributional shape is altered in a way that no scalar noise variance can reproduce. *Second*, the analog-to-digital converter introduces quantization noise and, at high instantaneous power, clipping artifacts. At the moderate jamming power levels used in this study, quantization effects are secondary to AGC compression but become relevant at higher power levels where the AGC range is exceeded. *Third*, subcarrier-dependent baseband filtering adds amplitude and phase distortions that vary across the frequency band [19]. These three effects are present in every commodity Wi-Fi receiver, regardless of chipset [20].

M_1 predicts that jammed amplitudes grow as the square root of added noise power, following a Rice or Rayleigh distribution shifted by independent Gaussian perturbations. A real receiver instead reports amplitudes compressed by AGC, with heavier tails from occasional gain-settling transients and cross-subcarrier correlations imposed by the wideband gain reduction. The assumption may hold approximately at very low interference levels, where the total received power changes negligibly and AGC gain remains nearly constant. As interference power increases, the multiplicative compression grows and the additive model becomes increasingly inaccurate. Whether this discrepancy matters for downstream applications is the empirical question addressed in Section VI. Two

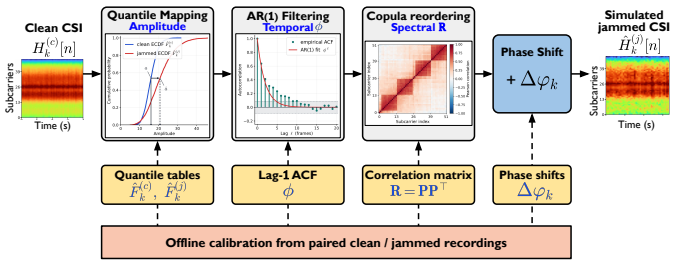

Fig. 2. M_{QTC} pipeline. Three calibrated stages: *quantile mapping* (amplitude), *AR(1) filtering* (temporal), and *Iman-Conover copula reordering* (cross-subcarrier correlation).

intermediate models that relax the assumption incrementally are introduced alongside the evaluation in Section VI.

IV. M_{QTC}: Quantile-Temporal-Copula Model

M_{QTC} addresses the three types of receiver-chain distortion identified in Section III through dedicated calibrated stages, as illustrated in Fig. 2. It follows the same interface as M_1: given paired clean and jammed recordings, calibrate parameters offline, then transform unseen clean CSI into simulated jammed CSI via $\hat{H}_k^{(j)}[n] = f_\theta\big(H_k^{(c)}[n]\big)$.

a) Stage 1: Per-subcarrier quantile mapping. *Quantile mapping* is a distribution-transfer technique widely used for bias correction in climate modeling [32]. For each subcarrier $k \in \{1, \ldots, K\}$, we estimate the empirical cumulative distribution functions (ECDFs) $\hat{F}_k^{(c)}$ and $\hat{F}_k^{(j)}$ of the clean and jammed amplitude distributions from the calibration data, evaluated at L equally spaced percentiles. During simulation, a clean amplitude value a is mapped to the jammed domain by composition of the clean ECDF with the jammed quantile function:

$$\hat{a}_k[n] = \big(\hat{F}_k^{(j)}\big)^{-1}\Big(\hat{F}_k^{(c)}\big(|H_k^{(c)}[n]|\big)\Big). \tag{5}$$

Values at the tails are linearly extrapolated from the two nearest percentiles. This guarantees that the simulated marginal distribution for each subcarrier matches the jammed reference exactly, regardless of its shape. The mapping captures AGC compression, power broadening, and any other distortion present in the per-subcarrier amplitude histograms, without requiring an explicit hardware model. The storage cost is $2KL$ floating-point values (two ECDFs per subcarrier), and the per-frame mapping cost is $O(K \log L)$ via binary search.

b) Stage 2: AR(1) temporal filtering. Quantile mapping produces correct marginals but no temporal structure beyond what the clean input carries. Real jammed CSI exhibits frame-to-frame autocorrelation that differs from the clean signal, because AGC gain adjustments operate on a time scale comparable to the frame interval. We model this with a first-order autoregressive process [33], chosen because low frame rates typical of commodity CSI extraction are insufficient to resolve higher-order dynamics. We estimate the lag-1 autocorrelation ϕ from the median autocorrelation function (ACF) of amplitude residuals across subcarriers. Rather than adding temporally correlated noise on top of the quantile-mapped amplitudes (which would destroy the marginals), we use ϕ to filter the random scores that drive the rank reordering in Stage 3:

$$s_k[n] = \phi \cdot s_k[n-1] + \sqrt{1-\phi^2} \cdot z_k[n], \tag{6}$$

where $z_k[n]$ are Cholesky-correlated normal scores produced by the copula stage below (Equation (7)) and the innovation scaling $\sqrt{1-\phi^2}$ preserves unit variance.

c) Stage 3: Iman-Conover copula reordering. After Stage 1, cross-subcarrier correlation still reflects the clean input rather than the wideband structure imposed by AGC. Since AGC reduces gain for the entire band simultaneously, independent per-subcarrier mapping destroys the correlated amplitude changes it creates. We restore the target correlation using the *Iman-Conover method* [34]: from jammed calibration amplitudes, we estimate the $K \times K$ Pearson correlation matrix $\mathbf{R}$ and compute its Cholesky factor $\mathbf{P}$ (with diagonal regularization $\epsilon\mathbf{I}$). Standard normal samples $\mathbf{W} \in \mathbb{R}^{N \times K}$ are correlated via $\mathbf{Z} = \mathbf{W}\mathbf{P}^T$ and filtered through the AR(1) process to produce rank scores $\mathbf{S}$, which reorder the quantile-mapped amplitudes:

$$\tilde{a}_k[n] = \mathrm{sort}\big(\hat{a}_k\big)\Big[\mathrm{rank}\big(s_k[n]\big)\Big]. \tag{7}$$

Because this step only permutes existing values (the sorted amplitudes are simply rearranged), it preserves the per-subcarrier marginals exactly while imposing the target cross-subcarrier dependence. The Cholesky decomposition is $O(K^3)$ but is computed once offline.

d) Phase reconstruction and output. On commodity Wi-Fi receivers, phase estimation is typically dominated by hardware noise, limiting the benefit of complex phase models. We apply a per-subcarrier circular mean phase shift $\Delta\varphi_k$ estimated from the calibration data:

$$\Delta\varphi_k = \arg\left(\frac{1}{N}\sum_{n=1}^{N}\exp\Big(j\big(\angle H_k^{(j)}[n] - \angle H_k^{(c)}[n]\big)\Big)\right). \tag{8}$$

The final simulated jammed CSI is $\hat{H}_k^{(j)}[n] = \tilde{a}_k[n] \cdot \exp\big(j(\angle H_k^{(c)}[n] + \Delta\varphi_k)\big)$. More sophisticated phase models were not pursued because the circular-mean shift already matches the observed phase variation within measurement noise. Algorithm 1 summarizes the online simulation procedure.

Algorithm 1 M_{QTC} Online Simulation

Input: Clean CSI $\{H_k^{(c)}[n]\}$; ECDFs $\hat{F}_k^{(c)}, \hat{F}_k^{(j)}$; Cholesky $\mathbf{P}$; AR coeff. ϕ; phase shifts $\Delta\varphi_k$
Output: Simulated jammed CSI $\{\hat{H}_k^{(j)}[n]\}$
for each subcarrier $k = 1, \ldots, K$ **do**
  $\hat{a}_k[n] \leftarrow (\hat{F}_k^{(j)})^{-1}(\hat{F}_k^{(c)}(|H_k^{(c)}[n]|))$ ▷ Quantile map
end for
Draw $\mathbf{W} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, size $N \times K$
$\mathbf{Z} \leftarrow \mathbf{W}\mathbf{P}^T$ ▷ Induce cross-subcarrier correlation
for each subcarrier $k = 1, \ldots, K$ **do**
  $s_k[n] \leftarrow \phi \cdot s_k[n-1] + \sqrt{1-\phi^2} \cdot z_k[n]$ ▷ AR(1) filter
  $\tilde{a}_k[n] \leftarrow \mathrm{sort}(\hat{a}_k)[\mathrm{rank}(s_k[n])]$ ▷ Copula reorder
end for
$\hat{H}_k^{(j)}[n] \leftarrow \tilde{a}_k[n] \cdot \exp\big(j(\angle H_k^{(c)}[n] + \Delta\varphi_k)\big)$ ▷ Reconstruct

e) Computational cost. Online simulation is $O(NK^2)$, dominated by the matrix multiply $\mathbf{Z} = \mathbf{W}\mathbf{P}^T$. Copula reorder-

ing adds $O(KN \log N)$ and quantile mapping $O(KN \log L)$, subdominant since K exceeds both $\log N$ and $\log L$ for practical traces. M_1 runs in $O(NK)$, so M_{QTC} adds a factor-K overhead. Offline calibration is also $O(NK^2)$ from the correlation matrix, plus $O(K^3)$ Cholesky and $O(KN \log N)$ quantile estimation. Storage is $2KL + K^2 + K + 1$ floats.

V. Experimental Setup

A. Fidelity metrics

A simulation model f_θ transforms clean CSI $H_k^{(c)}[n]$ into simulated jammed CSI $\hat{H}_k^{(j)}[n] = f_\theta(H_k^{(c)}[n])$. We measure fidelity as the mean distance across four dimensions:

$$\mathcal{F}(\hat{H}^{(j)}, H^{(j)}) = \tfrac{1}{4} \sum_{d \in \mathcal{D}} F_d, \tag{9}$$

where $\mathcal{D}$ = {amplitude, phase, temporal, spectral}. The *amplitude metric* uses the first Wasserstein distance [35] between per-subcarrier magnitude distributions, averaged across all K subcarriers:

$$F_{\text{ampl}} = \frac{1}{K} \sum_{k=1}^{K} W_1\big(|\hat{H}_k^{(j)}|,\ |H_k^{(j)}|\big). \tag{10}$$

The *phase metric* uses the absolute difference in circular variance [36], where $V(\theta) = 1 - \left|\frac{1}{N}\sum_n e^{j\theta[n]}\right|$:

$$F_{\text{phase}} = \frac{1}{K} \sum_{k=1}^{K} \big| V(\angle \hat{H}_k^{(j)}) - V(\angle H_k^{(j)}) \big|. \tag{11}$$

The *temporal metric* uses the ℓ_2 norm of the ACF difference over $L_{\max}$ lags:

$$F_{\text{temp}} = \frac{1}{K} \sum_{k=1}^{K} \Big\| \text{ACF}_{L_{\max}}\big(|\hat{H}_k^{(j)}|\big) - \text{ACF}_{L_{\max}}\big(|H_k^{(j)}|\big) \Big\|_2. \tag{12}$$

The *spectral metric* uses the Frobenius norm of the difference between the $K \times K$ cross-subcarrier Pearson correlation matrices:

$$F_{\text{spec}} = \big\| \mathbf{R}(\hat{H}^{(j)}) - \mathbf{R}(H^{(j)}) \big\|_F. \tag{13}$$

All four dimensions receive equal weight. Their natural scales differ substantially, so we report per-dimension scores alongside the aggregate throughout. Equal weighting avoids introducing arbitrary preferences across dimensions that have no common unit. The aggregate serves as a compact summary rather than the sole evaluation. The per-dimension breakdown in every table allows readers to assess each dimension independently and reweight according to their application.

B. Testbed and data collection

Both environments share the same hardware (Fig. 3). A Raspberry Pi 5 access point (802.11n HT20, channel 13), one or more ESP32-C6 receivers extracting K=52 subcarrier CSI estimates at roughly 3 fps via MQTT, and a HackRF One transmitting pre-generated jamming waveforms. The *controlled room* has a single receiver in fixed line-of-sight positions with minimal ambient Wi-Fi. Each recording comprises approximately 20 minutes of clean CSI followed by 15 minutes of jammed CSI. The *laboratory* deploys 5 receivers simultaneously at different positions and orientations in an active research environment with neighboring Wi-Fi networks and human movement. Laboratory recordings use interleaved 3-minute clean/jammed blocks to reduce drift. Constant Gaussian jamming was recorded at 5, 10, 15, and 20 dB intermediate frequency (IF) gain in the controlled room, and at 10, 15, and 20 dB in the laboratory. Within-scenario validation used 70/30 window-level train/test splits over windows of 32 consecutive frames, repeated over 5 random seeds. The controlled room isolates simulation fidelity from environmental confounds, while the lab tests whether findings generalize across devices and propagation conditions.

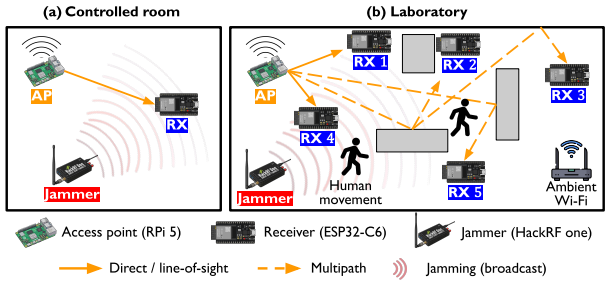

Fig. 3. Experimental testbed. (a) Controlled room: single ESP32-C6, Raspberry Pi 5 AP, and HackRF One in fixed positions. (b) Laboratory: 5 ESP32-C6 receivers in an active environment with ambient Wi-Fi and multipath.

TABLE II. Within-scenario fidelity at 20 dB, controlled room (mean±std, 5 seeds). Lower is better; phase is tied at this precision.

Model	Ampl.	Phase	Temp.	Spect.	Aggr.
M_1 (AWGN)	3.09	0.02	0.53	19.84	5.87±0.06
M_2 (Scaling)	0.82	0.02	0.37	18.10	4.83±0.09
M_3 (Scaling+Corr)	0.82	0.02	0.58	1.48	0.73±0.04
M_{QTC}	0.39	0.02	0.49	1.61	**0.63±0.07**

VI. Evaluation

We evaluate the linearity assumption (Section III) and M_{QTC} (Section IV) using the testbed and metrics of Section V.

A. Within-scenario fidelity

To isolate which receiver-chain corrections matter, we define two intermediate models alongside M_1. *M_2 (Scaling)* scales each subcarrier by its jammed-to-clean power ratio, $\alpha_k = \big(\mathbb{E}[|H_k^{(j)}|^2]/\mathbb{E}[|H_k^{(c)}|^2]\big)^{1/2}$, and adds Gaussian noise with cross-subcarrier correlation estimated from the residuals. This model captures the correct mean power per subcarrier but still assumes the amplitude distribution retains a Gaussian shape after scaling. *M_3 (Scaling+Corr)* adds Iman-Conover rank reordering on top of M_2, matching the jammed cross-subcarrier correlation while preserving M_2's per-subcarrier marginals exactly.

Table II reports the results. The amplitude column directly tests the linearity assumption. M_1 produced a Wasserstein distance of 3.09, confirming that adding Gaussian noise at the measured power difference yields an amplitude distribution far from the one observed. Fig. 4 makes this visible across the full band, where M_{QTC} tracks the real jammed amplitude profile on all 52 subcarriers while M_1 diverges around the null subcarrier region. No scalar noise variance can reproduce a

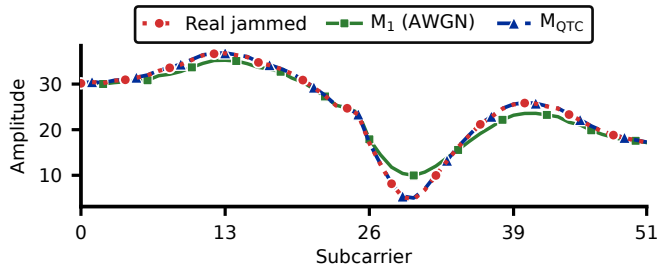

Fig. 4. Mean amplitude across all 52 subcarriers at 20 dB. M_{QTC} tracks the real jammed profile across the entire band, while M_1 diverges around the null subcarriers.

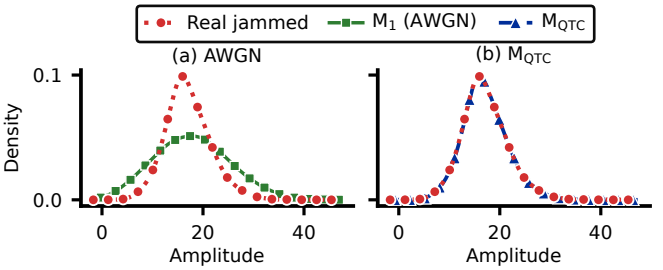

Fig. 5. Amplitude density on subcarrier 26 at 20 dB. M_1 overshoots the real jammed distribution while M_{QTC} closely tracks it.

distribution whose shape was altered by a multiplicative gain stage. This finding is consistent with Serbetci et al. [9], who showed that naive Gaussian injection on CSI produces worse localization than physics-motivated augmentation.

M_{QTC} reduced the amplitude gap to 0.39, an 8-fold improvement, because quantile mapping learns the actual per-subcarrier distribution transformation rather than assuming additivity. Fig. 5 shows this in detail on subcarrier 26, where M_1 overshoots the real jammed distribution while M_{QTC} closely tracks it. The aggregate score dropped from 5.87 to 0.63, closing 89% of the gap.

Phase fidelity was identical across all models at approximately 0.02, because phase estimation on commodity receivers is dominated by hardware noise that masks the small phase perturbation introduced by jamming. M_2 reduced the aggregate only modestly because the mean power ratio is correct but the distributional shape remains wrong. M_3 achieved a larger reduction by restoring cross-subcarrier correlation, yet retains a 2-fold amplitude residual because it corrects the dependence structure but inherits M_2's Gaussian-shaped marginals, whereas M_{QTC} replaces those marginals entirely through quantile mapping before reordering. In the laboratory (all 5 devices pooled), M_{QTC} achieved an aggregate of 1.30 compared to 7.21 for M_1, maintaining the same ordering.

B. Ablation

Table III decomposes M_{QTC} into four variants. All share the same amplitude score because quantile mapping is present in every configuration. The critical distinction is spectral. Q-only produced 21.15, *worse* than M_1 (19.84), because independent per-subcarrier mapping destroys the cross-subcarrier correlation that the clean input carried. Correcting marginals alone is not sufficient when AGC imposes wideband dependence that independent processing erases.

TABLE III. Ablation of M_{QTC} at 20 dB (mean±std, 5 seeds). Q = quantile mapping, T = AR(1), C = copula. Per-dimension scores: controlled room; aggregates: both environments.

Variant	Ampl.	Temp.	Spect.	Aggr. (ctrl)	Aggr. (lab)
Q-only	0.39	0.79	21.15	5.59±0.11	6.51±0.17
QT	0.39	0.50	20.30	5.30±0.05	7.97±0.08
QC	0.39	0.58	1.52	0.63±0.05	1.67±0.16
QTC	0.39	0.49	1.61	0.63±0.07	**1.30±0.18**

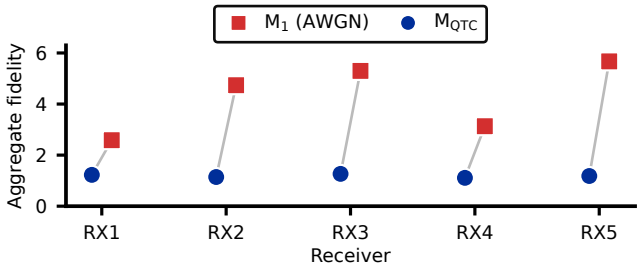

Fig. 6. Per-device aggregate fidelity in the laboratory at 20 dB (5 seeds, lower is better).

Adding copula reordering (QC) reduced the spectral score to 1.52 and the aggregate to 0.63, a 9-fold improvement over Q-only. Adding AR(1) temporal filtering (QT) improved the temporal dimension from 0.79 to 0.50 but did not address cross-subcarrier correlation. The full QTC achieved the same aggregate (0.63) as QC in this setting, because adding T improves the temporal dimension (QC 0.58 to QTC 0.49) but this gain is offset by a slight spectral shift. Copula reordering is the dominant mechanism across all conditions. The laboratory column in Table III confirms this in a second environment. QTC achieved the lowest aggregate (1.30±0.18), a 22% improvement over QC (1.67±0.16), driven by better temporal fidelity (2.83 to 1.34). The temporal component's contribution is larger in the laboratory, where richer frame-to-frame dynamics are present. Across all seven tested conditions (four power levels in the controlled room, three in the laboratory), QTC achieved the lowest or tied-lowest aggregate, confirming that the full pipeline generalizes beyond a single operating point.

C. Per-device consistency

Fig. 6 reports per-device fidelity in the laboratory. The assumption failure is consistent. Every device shows large M_1 divergence (aggregate 2.58 to 5.67) and successful M_{QTC} recovery (1.11 to 1.27), a 2.1- to 4.8-fold improvement. Per-device calibration produced lower scores than the merged-lab result (1.30), confirming that device-specific calibration captures variation that pooling averages out. In *leave-one-device-out* evaluation, M_{QTC}'s aggregate rose to 5.82, confirming that the transformation is position-specific. This is expected, as AGC compression depends on total received power, which varies with position and orientation, tying calibration to deployment geometry. The practical cost is a single 15-minute paired recording, after which unlimited realistic data can be generated from any clean recording without additional jammer hardware.

TABLE IV. Sim-to-real detection AUC at 10 dB in the laboratory (per-device calibration, mean±std across 5 receivers, 3 seeds).

Training data	LR	SVM	RF	MLP	AE	Mean
Oracle (real)	0.961	0.989	0.990	0.900	0.997	0.967±0.009
M_{QTC}	**0.920**	**0.905**	**0.929**	**0.817**	**0.949**	**0.904±0.022**
M_3	0.875	0.756	0.866	0.771	0.850	0.824±0.061
M_2	0.836	0.848	0.878	0.742	0.828	0.826±0.062
M_1 (AWGN)	0.469	0.408	0.623	0.461	0.649	0.522±0.051

D. Sim-to-real detection

The downstream consequence of simulation fidelity becomes visible at moderate power. Table IV evaluates at 10 dB IF gain, the lowest laboratory jammer power, where the detection task is most challenging. We used 5 classifiers spanning different families: logistic regression (LR), support vector machine (SVM) with RBF kernel, random forest (RF), a two-layer multilayer perceptron (MLP), and an autoencoder (AE)-based classifier (encoder-decoder with a classification head trained jointly on reconstruction and binary cross-entropy losses). Each classifier was trained on K-dimensional per-subcarrier mean amplitude features from simulated jammed data and tested on real jammed CSI (3 seeds per device, averaged across devices). An oracle trained on real jammed data served as the upper bound.

M_{QTC} achieved the highest AUC for every classifier, with a mean of 0.904 vs. 0.522 for M_1 and 0.967 for the oracle, recovering 93% of oracle performance. Three of five M_1-trained classifiers fell *below a random classifier*. The additive model overestimates the amplitude shift, so classifiers learn incorrect decision boundaries that invert when applied to real data where AGC has compressed the amplitude differences.

E. External validation

The receiver-chain distortion that breaks the additive assumption is not specific to jamming. AGC compression acts on any change in total received power, whether caused by interference, human presence, or environment change. To test whether M_{QTC}'s advantage generalizes, we compared it against four published CSI augmentation methods on three external public datasets spanning two hardware platforms. Wallhack1.8k [8] provides paired no-person and activity CSI from ESP32 receivers (52 subcarriers), SignFi [37] provides paired lab and home recordings from Intel 5300 receivers (30 subcarriers), and Widar 3.0 [38] provides paired cross-room gesture recordings from Intel 5300 receivers (30 subcarriers). Each method was given its most generous calibrated form, and fidelity was measured as the amplitude Wasserstein distance on held-out test data (5 seeds).

Table V reports the results. M_{QTC} achieved the lowest amplitude error on all three datasets, with a 2-fold improvement over the best baseline on Wallhack (0.19 vs. AirFi at 0.37), an 8.6-fold improvement on SignFi (0.96 vs. Strohmayer et al. at 8.28), and a 5.7-fold improvement on Widar (0.06 vs. Strohmayer et al. at 0.34). The advantage holds across both hardware platforms and all three types of CSI modification. All baseline methods apply additive or multiplicative perturbations that preserve the distributional shape, whereas M_{QTC}'s quantile mapping learns the true per-subcarrier distribution regardless of its form.

TABLE V. Amplitude Wasserstein distance on three external public datasets (5 seeds, lower is better). Wallhack1.8k uses ESP32 (52 subcarriers), SignFi and Widar use Intel 5300 (30 subcarriers).

Method	Wallhack1.8k (presence)	SignFi (lab→home)	Widar 3.0 (room change)
M_{QTC}	**0.19**	**0.96**	**0.06**
Gao et al. [10]	0.55	14.41	1.31
AirFi [7]	0.37	13.83	1.30
Strohmayer et al. [8]	0.51	8.28	0.34
Serbetci et al. [9]	0.55	10.23	0.60

F. Computational cost

For our configuration (K=52, L=1000 quantile percentiles), M_{QTC} storage is 0.8 MB. Simulating 1 000 frames takes approximately 14 ms on a single CPU core, roughly 6 times slower than M_1 but more than 10^4 times faster than the 3 fps collection rate. Calibration from 3 000 frames completes in under one second. The overhead relative to M_1 is modest given the 8-fold fidelity improvement, and the entire calibrate-then-simulate cycle remains fast enough to generate unlimited realistic data once a single paired recording is available.

VII. Discussion

Because AGC operates upstream of the FFT, the distortion mechanism identified in this work is present in OFDM receivers across standards, suggesting that the assumption failure extends to 802.11be and 5G NR, though validation on those platforms remains future work. Our experience developing a real-data CSI-based detection pipeline [25] highlighted the gap between simulated and measured receiver output that motivated this investigation. Calibrated simulation could complement such systems by enabling rapid data generation for new deployment sites from a single paired recording. The cost of this fidelity is position-specificity. M_{QTC}'s richer parameterization captures the local AGC operating point and must be recalibrated at each deployment, consistent with the environment dependence reported for CSI sensing [31]. M_1, by contrast, transfers with negligible change because its single parameter carries no position-specific information. The residual amplitude error shrinks further with larger calibration sets, reaching the 0.10 sampling-noise floor when calibration and evaluation use the same recording. For wider-bandwidth systems where K exceeds 1 000, the correlation matrix is naturally banded, so block-diagonal or sparse approximations can maintain tractability. All experiments in this study used the ESP32-C6 chipset, though the external validation (Section VI-E) provides indirect evidence of cross-platform applicability on Intel 5300 data. Platforms based on Intel 5300 [3] or Nexmon-modified Broadcom [4] may exhibit different distortion patterns, but Portner et al. [20] confirmed via cable-coupled measurements that AGC-induced amplitude scaling is the dominant cross-device CSI effect across Intel, Broadcom, and Espressif chipsets, so the assumption failure is structural, not chipset-specific.

VIII. Conclusion

This paper presented, to our knowledge, the first empirical validation of CSI-level simulation against real wireless measurements. Using paired clean and jammed recordings from 6 commodity Wi-Fi receivers across 2 environments, we showed that the standard AWGN model fails because the receiver chain transforms interference multiplicatively before the channel estimator reports it. The proposed M_{QTC} pipeline closes 89% of the fidelity gap through measurement-calibrated statistical transforms, with copula reordering as the dominant mechanism. Fidelity translates to utility. M_{QTC}-trained classifiers recovered 93% of oracle detection performance while AWGN-trained classifiers remained near random classifiers. The broader takeaway is that simulation models must be validated against the *hardware they target*, not just against the channel model they implement.

Acknowledgment

This work was supported by the French National Research Agency (ANR) under grant ANR-22-PEFT-0007 (NF-FITNESS project) as part of the France 2030 program, and by the Luxembourg National Research Fund under the PEARL project P24/IS/18423066/SCICT.

References

[1] F. Wang, T. Zhang *et al.*, "A survey on Wi-Fi sensing generalizability: Taxonomy, techniques, datasets, and future research prospects," *IEEE Commun. Surveys Tuts.*, vol. 28, pp. 5227–5266, 2026.

[2] I. Alla and V. Loscri, "Sec5GLoc: Securing 5G indoor localization via adversary-resilient deep learning architecture," in *Proc. IEEE Conf. Commun. Netw. Security (CNS)*, 2025, pp. 1–9.

[3] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: Gathering 802.11n traces with channel state information," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 41, no. 1, p. 53, 2011.

[4] F. Gringoli, M. Schulz, J. Link, and M. Hollick, "Free your CSI: A channel state information extraction platform for modern Wi-Fi chipsets," in *Proc. 13th Int. Workshop Wireless Netw. Testbeds Exp. Eval. Characterization (WiNTECH)*, 2019, pp. 21–28.

[5] Espressif Systems, "ESP-CSI: Wi-Fi channel state information application," 2024, GitHub repository, accessed Sep. 2026. [Online]. Available: https://github.com/espressif/esp-csi

[6] J. Zhang, Z. Tang, M. Li, D. Fang, P. Nurmi, and Z. Wang, "CrossSense: Towards cross-site and large-scale WiFi sensing," in *Proc. 24th Annu. Int. Conf. Mobile Comput. Netw. (MobiCom)*, 2018, pp. 305–320.

[7] D. Wang, J. Yang, W. Cui, L. Xie, and S. Sun, "AirFi: Empowering WiFi-based passive human gesture recognition to unseen environment via domain generalization," *IEEE Trans. Mobile Comput.*, vol. 23, no. 2, pp. 1156–1168, 2024.

[8] J. Strohmayer and M. Kampel, "Data augmentation techniques for cross-domain WiFi CSI-based human activity recognition," in *Artificial Intelligence Applications and Innovations (AIAI 2024), IFIP AICT*, vol. 711, 2024, pp. 42–56.

[9] O. G. Serbetci, J.-H. Lee, D. Burghal, and A. F. Molisch, "Simple and effective augmentation methods for CSI based indoor localization," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2023, pp. 3947–3952.

[10] K. Gao, H. Wang, H. Lv, and W. Liu, "Toward 5G NR high-precision indoor positioning via channel frequency response: A new paradigm and dataset generation method," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 7, pp. 2233–2247, 2022.

[11] A. Goldsmith, *Wireless Communications*. Cambridge Univ. Press, 2005.

[12] Technology Innovation Institute, "JamRF: Jamming RF framework," 2022, GitHub repository (archived), accessed Sep. 2026. [Online]. Available: https://github.com/tiiuae/jamrf

[13] N. N. Bhat, R. Berkvens, and J. Famaey, "CSI4Free: GAN-augmented mmWave CSI for improved pose classification," in *Proc. IEEE 4th Int. Symp. Joint Commun. Sensing (JC&S)*, 2024, pp. 1–6.

[14] J. Wang, C. Zhao *et al.*, "Generative AI enabled robust data augmentation for wireless sensing in ISAC networks," *IEEE J. Sel. Areas Commun.*, vol. 44, pp. 975–990, 2026.

[15] W. Hou and C. Wu, "RFBoost: Understanding and boosting deep WiFi sensing via physical data augmentation," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 8, no. 2, pp. 1–26, 2024.

[16] T. S. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. Prentice Hall PTR, 2001.

[17] V. V. Ratnam, H. Chen, H.-H. Chang, A. Sehgal, and J. C. Zhang, "Optimal preprocessing of WiFi CSI for sensing applications," *IEEE Trans. Wireless Commun.*, vol. 23, no. 9, pp. 10 820–10 833, 2024.

[18] B. Wei, H. Song, J. Katto, and T. Kikkawa, "RSSI-CSI measurement and variation mitigation with commodity WiFi device," *IEEE Internet Things J.*, vol. 10, no. 7, pp. 6249–6258, 2023.

[19] Z. Jiang, T. H. Luan *et al.*, "Eliminating the barriers: Demystifying Wi-Fi baseband design and introducing the PicoScenes Wi-Fi sensing platform," *IEEE Internet Things J.*, vol. 9, no. 6, pp. 4476–4496, 2022.

[20] F. Portner, F. Gringoli, M. Hollick, and A. Asadi, "Same signal, different story: Demystifying receiver effects in Wi-Fi channel state information," *IEEE Internet Things J.*, vol. 13, no. 10, pp. 20 963–20 977, 2026.

[21] W. Xu, W. Trappe, Y. Zhang, and T. Wood, "The feasibility of launching and detecting jamming attacks in wireless networks," in *Proc. 6th ACM Int. Symp. Mobile Ad Hoc Netw. Comput. (MobiHoc)*, 2005, pp. 46–57.

[22] H. Pirayesh and H. Zeng, "Jamming attacks and anti-jamming strategies in wireless networks: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 2, pp. 767–809, 2022.

[23] P. Mykytyn, R. Chitauro, Z. Dyka, and P. Langendoerfer, "Channel state information analysis for jamming attack detection in static and dynamic UAV networks – an experimental study," in *Proc. 21st Int. Conf. Distrib. Comput. Smart Syst. Internet Things (DCOSS-IoT)*, 2025, pp. 322–327.

[24] I. Panitsas, Y. Yigit, L. Tassiulas, L. A. Maglaras, and B. Canberk, "JamShield: A machine learning detection system for over-the-air jamming attacks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2025, pp. 1067–1072.

[25] A. Bouferroum, I. Alla, V. Loscri, A. Benslimane, and V. Lenders, "CITADEL: CSI-based jamming detection and open-set classification for IIoT networks," *arXiv preprint arXiv:2606.22939*, 2026.

[26] A. Manukyan, R. Mkrtchyan, A. Saribekyan, T. P. Raptis, and H. Khachatrian, "Bridging the sim-to-real gap in RF localization with large-scale synthetic pretraining," *Inf. Fusion*, vol. 130, p. 104104, 2026.

[27] W. H. Clark, S. Hauser, W. C. Headley, and A. J. Michaels, "Training data augmentation for deep learning radio frequency systems," *J. Defense Model. Simul.*, vol. 18, no. 3, pp. 217–237, 2021.

[28] I. Alla, M. Zhang, J. Ashdown, V. Loscri, and F. Restuccia, "Finding a needle in a (spectrum) haystack: Multi-band multi-device radio fingerprinting," *Comput. Netw.*, vol. 280, p. 112148, 2026.

[29] B. Barahimi, H. Tabassum, M. Omer, and O. Waqar, "Context-aware predictive coding: A representation learning framework for WiFi sensing," *IEEE Open J. Commun. Soc.*, vol. 5, pp. 6119–6134, 2024.

[30] H. F. Thariq Ahmed, H. Ahmad, K. Narasingamurthi, H. Harkat, and S. K. Phang, "DF-WiSLR: Device-free Wi-Fi-based sign language recognition," *Pervasive Mobile Comput.*, vol. 69, p. 101289, 2020.

[31] M. Cominelli, F. Gringoli, and F. Restuccia, "Exposing the CSI: A systematic investigation of CSI-based Wi-Fi sensing capabilities and limitations," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. (PerCom)*, 2023, pp. 81–90.

[32] A. J. Cannon, S. R. Sobie, and T. Q. Murdock, "Bias correction of GCM precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes?" *J. Climate*, vol. 28, no. 17, pp. 6938–6959, 2015.

[33] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*, 5th ed. Wiley, 2015.

[34] R. L. Iman and W. J. Conover, "A distribution-free approach to inducing rank correlation among input variables," *Commun. Statist.-Simul. Comput.*, vol. 11, no. 3, pp. 311–334, 1982.

[35] Y. Rubner, C. Tomasi, and L. J. Guibas, "The earth mover's distance as a metric for image retrieval," *Int. J. Comput. Vis.*, vol. 40, no. 2, pp. 99–121, 2000.

[36] K. V. Mardia and P. E. Jupp, *Directional Statistics*. Wiley, 2000.

[37] Y. Ma, G. Zhou, S. Wang, H. Zhao, and W. Jung, "SignFi: Sign language recognition using WiFi," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 2, no. 1, pp. 1–21, 2018.

[38] Y. Zheng, Y. Zhang *et al.*, "Zero-effort cross-domain gesture recognition with Wi-Fi," in *Proc. ACM MobiSys*, 2019, pp. 313–325.