On the Learnability of Test-Time Adaptation

A Recovery Complexity Perspective

Zhi Zhou, Ming Yang, Shi-Yu Tian, Kun-Yang Yu,
Lan-Zhe Guo, Yu-Feng Li

Nanjing University

ICML 2026

The first theoretical framework for the learnability of TTA

When Can TTA Be Trusted?

Test-Time Adaptation (TTA) adapts a model to distribution shifts at test time using **only unlabeled data** — no labels, adapting on the fly as the test stream arrives.

Promise and peril

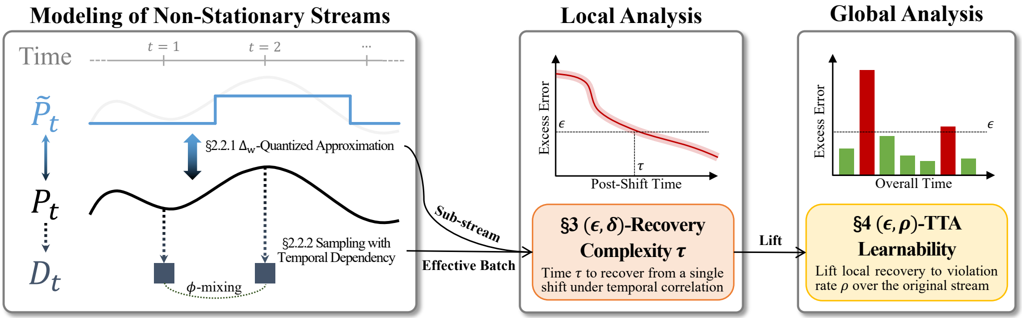
- Strong empirical success across a wide range of shifts.
- Yet it can silently fail under complex, non-stationary streams.

The gap in existing theory

Regret captures only *average* performance — not **post-shift recovery**, **per-step reliability**, or the **unlabeled** information constraint.

Open question: *when is TTA even learnable?*

Our Framework at a Glance



1. Stream model

Unify shifts (**Wasserstein quantization**) + temporal correlation (**ϕ-mixing**).

2. Local: Recovery

(ϵ, δ)-**Recovery Complexity** τ with **matching** lower & upper bounds.

3. Global: Learnability

Lift to (ϵ, ρ)-**TTA Learnability**; connect to dynamic regret.

Modeling Non-Stationary Test Streams

We model two complementary axes of non-stationarity:

Global: Δ_W-Quantized Shift

Approximate the continuous trajectory $\{\mathcal{P}_t\}$ by a **piecewise-constant** surrogate at resolution Δ_W (Wasserstein-1).

shifts bounded by path variation: $\tilde{K}_S(T) \leq \lceil 2V_T/\Delta_W \rceil$.

Local: ϕ-Mixing Dependence

Temporal correlation $\phi(i) \leq \varrho^i$ shrinks the **effective batch size**:

$$B_{\text{eff}} = \frac{B}{C_\phi} \leq B, \quad C_\phi = 1 + \frac{4\sqrt{\varrho}}{1-\sqrt{\varrho}}.$$

Correlation $\Rightarrow$ fewer informative samples.

One tractable model capturing gradual *and* abrupt shifts, i.i.d. *and* correlated streams.

Local Analysis: (ϵ, δ)-Recovery Complexity

ICML International Conference On Machine Learning

Definition

$\tau(\epsilon, \delta)$ = smallest time after a shift such that the excess risk stays $\leq \epsilon$ with probability $\geq 1 - \delta$.

Order-wise matching minimax lower & upper bounds:

$$\underbrace{\tau \;\geq\; \Omega\!\left(\frac{C_\phi}{B} \cdot \frac{1}{\alpha(\sqrt{\zeta + 2\alpha\epsilon} + \sqrt{\zeta})^2}\right)}_{\text{any algorithm (information limit)}} \qquad \underbrace{\tau \;\leq\; \mathcal{O}\!\left(\frac{C_\phi}{B\,\alpha^2 \epsilon} \log \frac{\Delta_W + \epsilon}{\epsilon}\right)}_{\text{simple TTA baseline}}$$

What governs recovery:

- alignment $\alpha \uparrow \Rightarrow \tau \sim 1/\alpha^2$ (quadratic gain)
- batch $B\uparrow$, correlation $C_\phi \downarrow \Rightarrow$ faster

An intrinsic error floor:

- misalignment $\zeta > 0$ blocks τ as $\epsilon \to 0$
- adaptivity–information trade-off

$\Rightarrow$ **Little room for improvement without stronger feedback.**

Global Analysis: (ϵ, ρ)-TTA Learnability

Definition

Problem is (ϵ, ρ)-learnable if the **fraction of time steps** violating the target ϵ is at most ρ: $\frac{1}{T}\sum_t \mathbb{P}(\ell_t(\theta_t) - R_t > \epsilon) \leq \rho$.

Lift local recovery → global learnability:

$$\rho \;\leq\; \delta + \frac{(\tilde{K}_S(T) + 1)\,\tau(\epsilon', \delta)}{T}$$

Strong transfer: fewer shifts or faster recovery $\Rightarrow$ directly smaller ρ.

Connection to dynamic regret:

$$\mathrm{Reg}(T) \leq T(\epsilon + M\rho)$$

Per-step reliability ρ directly bounds the **cumulative** regret.

- Persistent shifts ($\Delta_W = \Omega(1)$) force linear regret — unlike online learning.
- Unlabeled supervision (α, ζ) adds an intrinsic cumulative cost.
- Benign regime ($\Delta_W \to 0$) recovers classical sublinear regret.

Empirical Validation

We validate on **controlled synthetic streams** and standard **corruption benchmarks** (full setup & results in the paper).

Synthetic: scaling laws hold

Varying alignment α (at $B = 16$):

α	LB	τ	UB
0.05	59.0	322.0	311.9
0.10	15.6	77.0	78.0
0.20	4.1	19.0	19.5
0.50	0.7	4.0	3.1

$\tau \cdot \alpha^2 \approx \text{const} \Rightarrow \tau = \mathcal{O}(1/\alpha^2)$; τ stays above LB & tracks UB $\Rightarrow$ **near-tight**.

Real-world: $\tilde{\alpha}$ governs success

ImageNet-C, correlation grows as $\beta\downarrow$:

β	$\tilde{\alpha}$	ΔAcc.	#Improve
0.001	−0.028	−12.7%	0/15
0.01	+0.090	+3.6%	12/15
0.1	+0.157	+13.7%	15/15
Uniform	+0.171	+14.7%	15/15

$\tilde{\alpha} > 0$: Tent improves all 15 corruptions; as correlation rises $\tilde{\alpha} < 0$ and Tent collapses.

Alignment α is the core factor governing whether TTA succeeds.

Takeaways

Contributions

- **First learnability framework** for TTA under unlabeled, non-stationary streams.
- (ϵ, δ)-**Recovery Complexity** with order-wise **matching** lower & upper bounds.
- Lifted to (ϵ, ρ)-**TTA Learnability**; principled link to dynamic regret.

Key messages:

- TTA difficulty is set by **alignment** α, **misalignment** ζ, **batch** B, **correlation** C_ϕ.
- An **adaptivity–information trade-off** fundamentally limits TTA.

Thank you!

Questions welcome · zhouz@lamda.nju.edu.cn