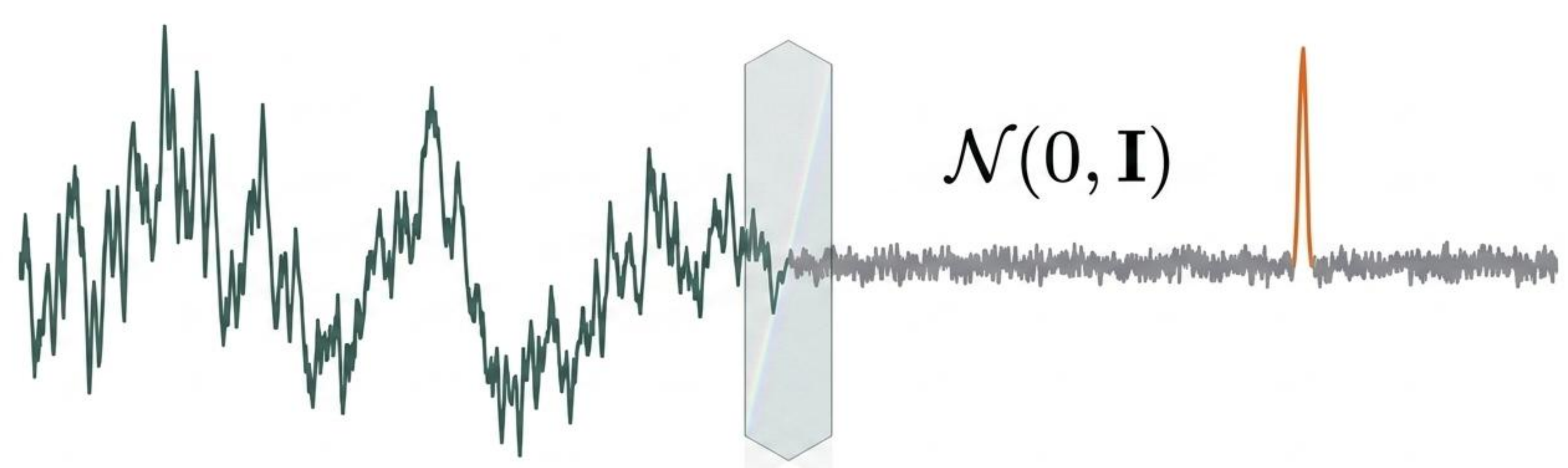
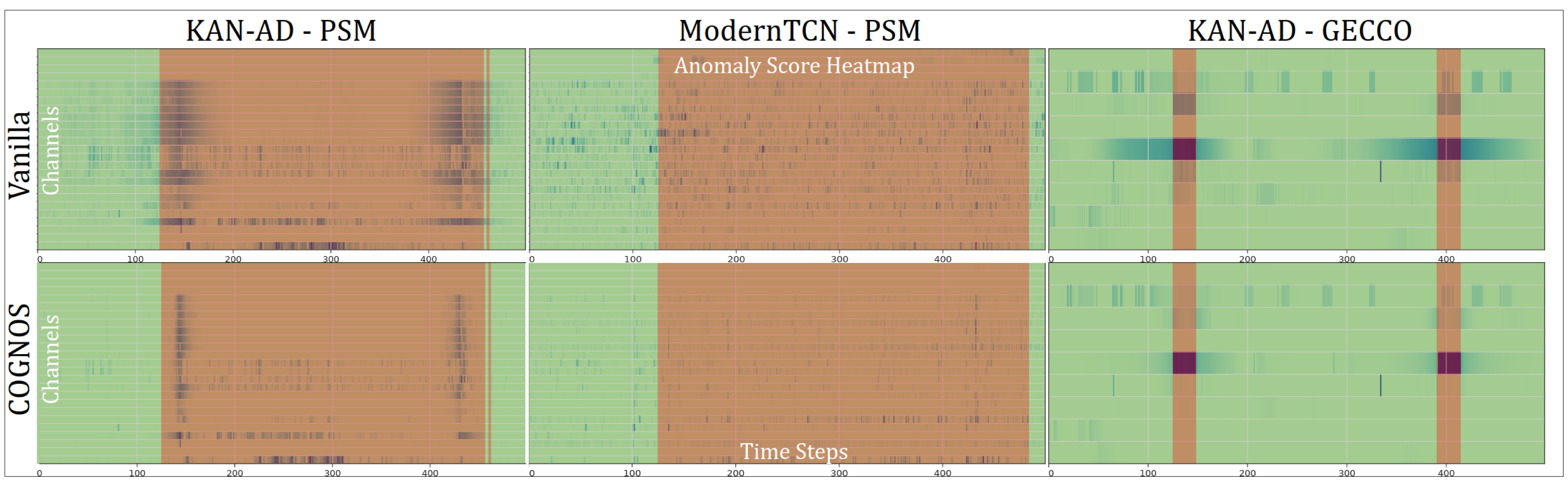
COGNOS: Universal Enhancement for Time Series Anomaly Detection via Constrained Gaussian-Noise Optimization and Smoothing

Wenlong Shang, Shihao Tian, Xutong Wan, Peng Chang
Beijing Key Laboratory of Computational Intelligence and Intelligent System,
Beijing University of Technology, China.





北京人工智能研究院
Beijing Institute of Artificial Intelligence
BEIJING UNIVERSITY OF TECHNOLOGY

ICML
International Conference On Machine Learning

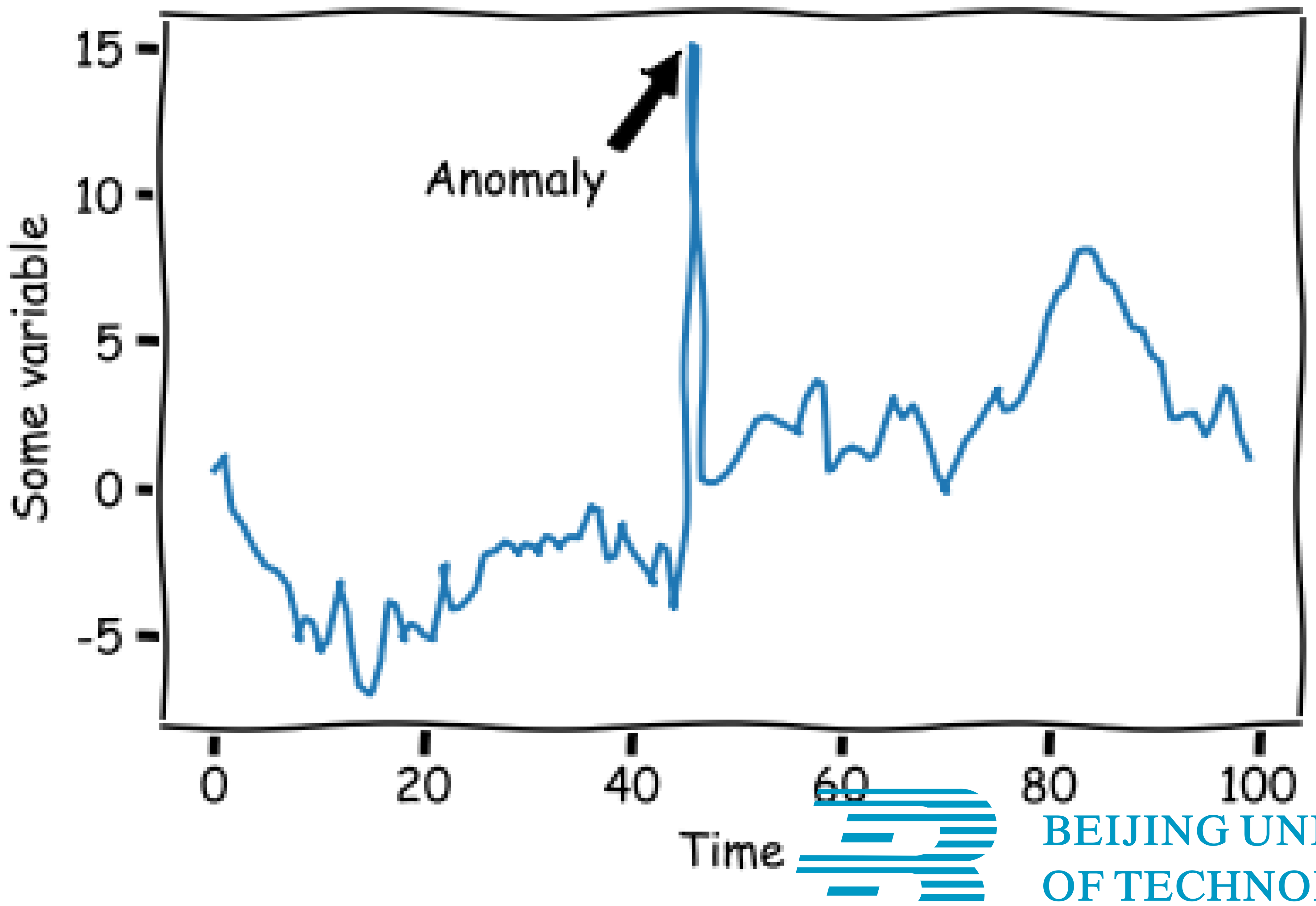
COGNOS: Universal Enhancement for Time Series Anomaly Detection via Constrained Gaussian-Noise Optimization and Smoothing

Introduction

Smart Gauge

IoT Devices

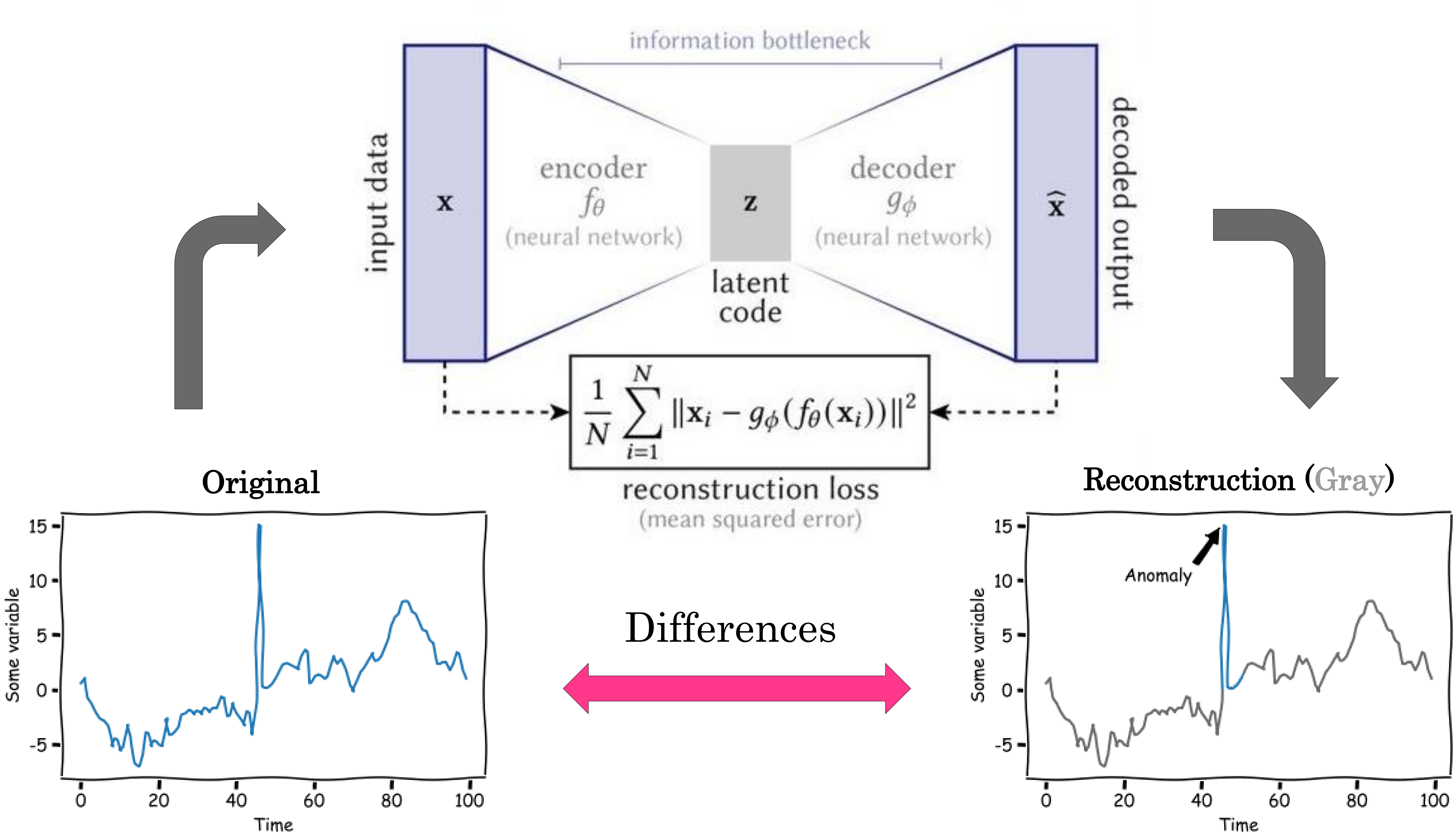
Process Industry



BEIJING UNIVERSITY
OF TECHNOLOGY

COGNOS: Universal Enhancement for Time Series Anomaly Detection via Constrained Gaussian-Noise Optimization and Smoothing

Introduction

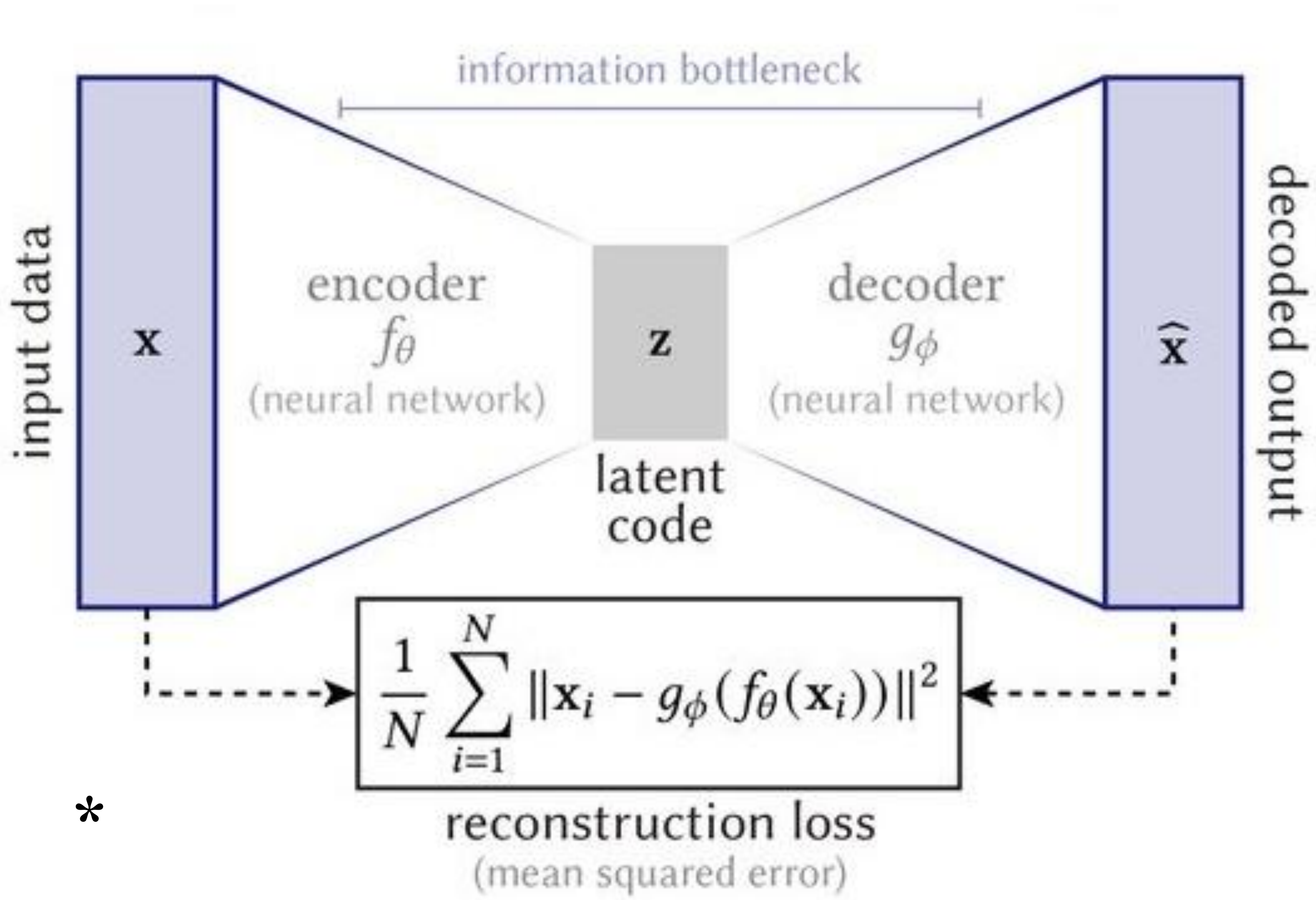


*@Keenan Crane

COGNOS: Universal Enhancement for Time Series Anomaly Detection via Constrained Gaussian-Noise Optimization and Smoothing

THE GAP: MSE Loss

The *Theory*



*

Deep models predominantly optimize Mean Squared Error (MSE).

$$\theta^*_{\mathrm{MLE}} = \underset{\theta}{\operatorname{argmax}}\, \ell(\theta) \iff \underset{\theta}{\operatorname{argmin}} \sum_{i=1}^{N} (y_i - f_\theta(\mathbf{x}_i))^2.$$

MSE acts as Maximum Likelihood Estimation (MLE) only if residuals are **i.i.d. Gaussian White Noise.**

BEIJING UNIVERSITY OF TECHNOLOGY

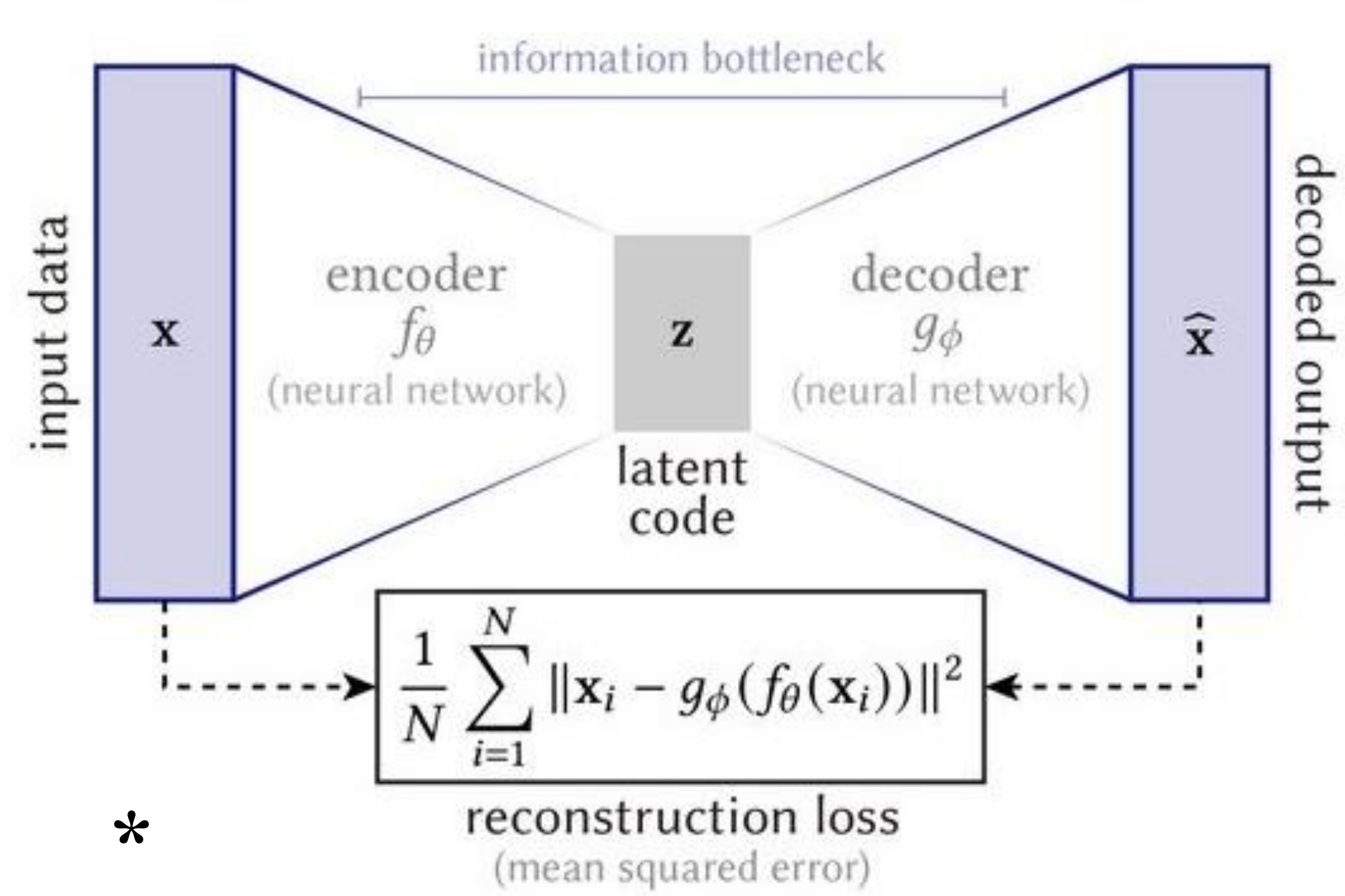
*@Keenan Crane

ICML International Conference On Machine Learning

COGNOS: Universal Enhancement for Time Series Anomaly Detection via Constrained Gaussian-Noise Optimization and Smoothing

Introduction

THE GAP: MSE Loss



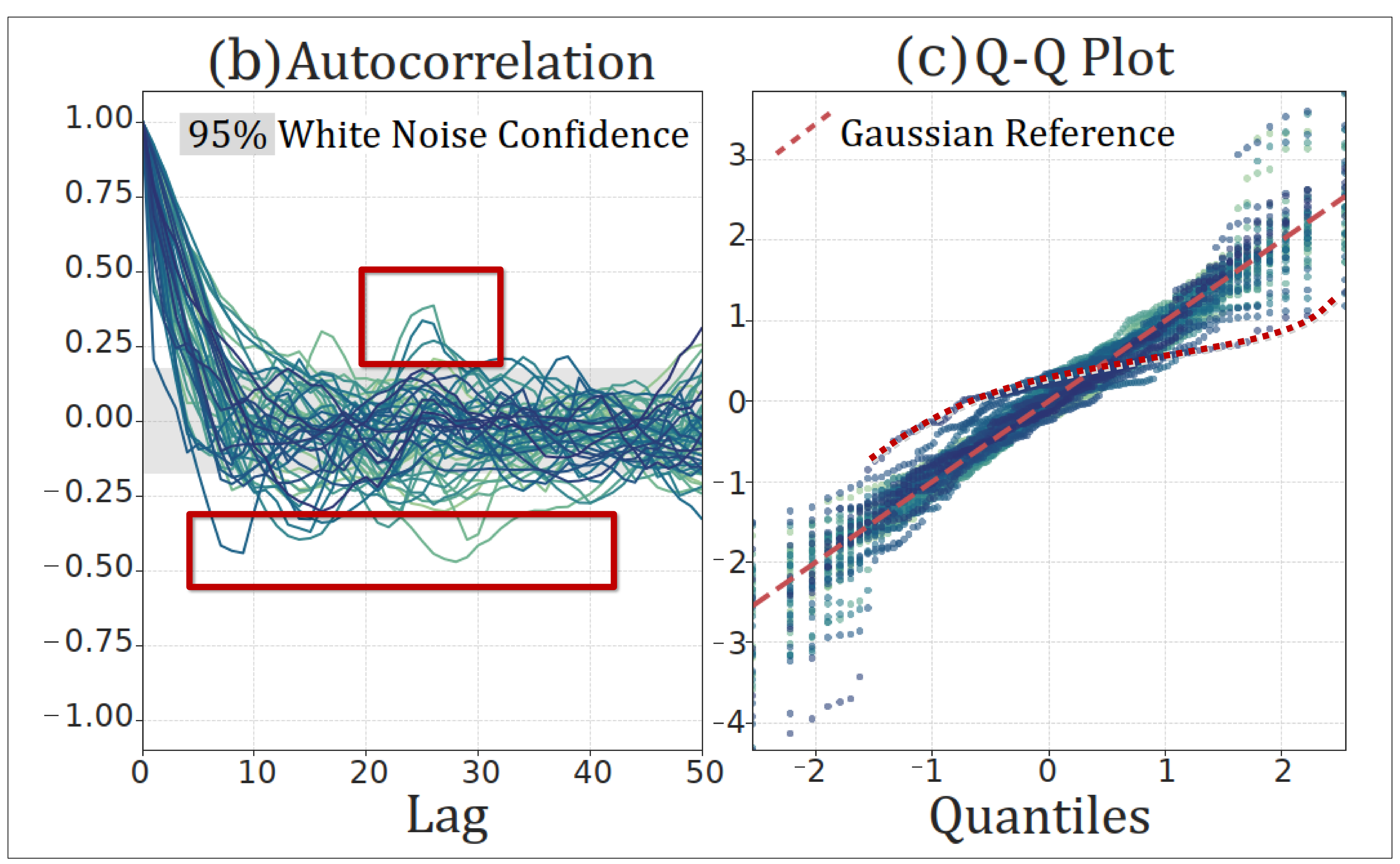
*

The *Theory*

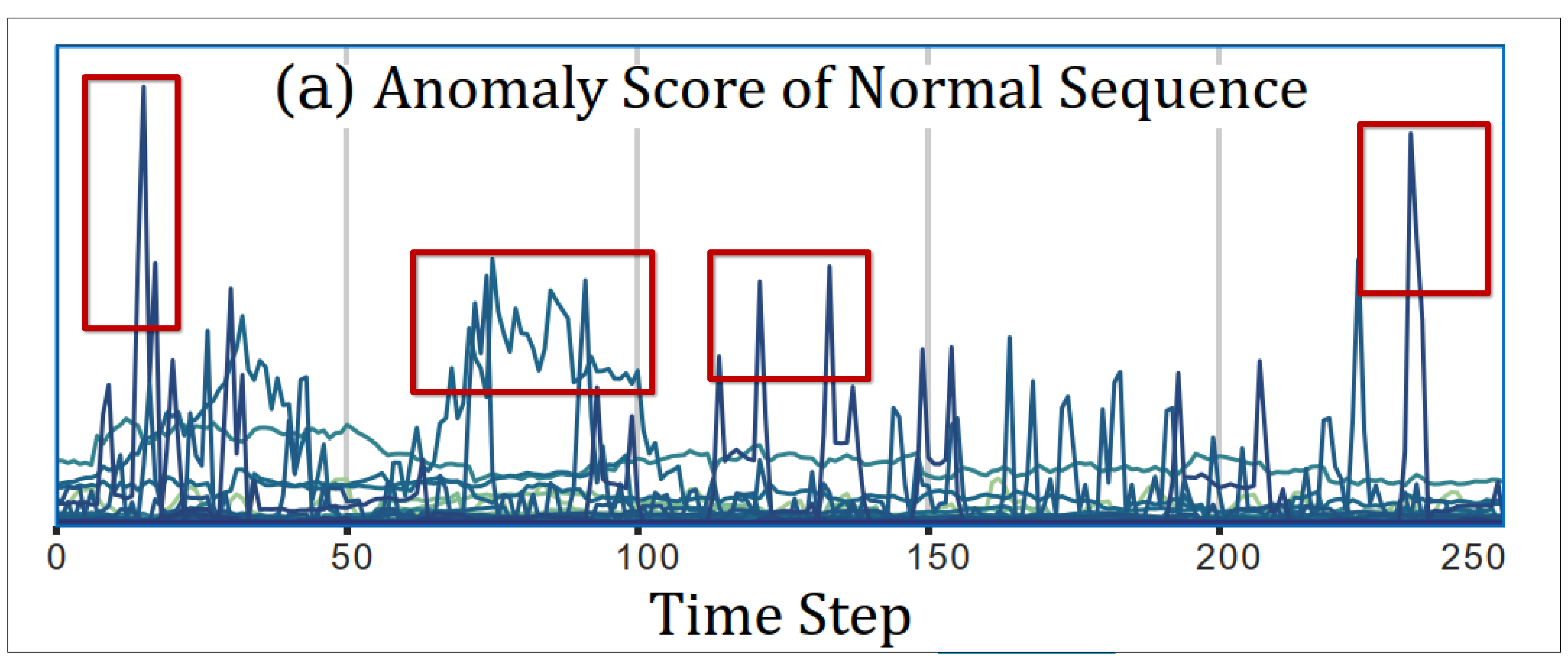
MSE acts as Maximum Likelihood Estimation (MLE) only if residuals are **i.i.d. Gaussian White Noise.**

The *Reality*

Residuals retain heavy **temporal correlation** and **non-Gaussian tails.**



A high noise floor that buries true structural breaks under **false positives**.



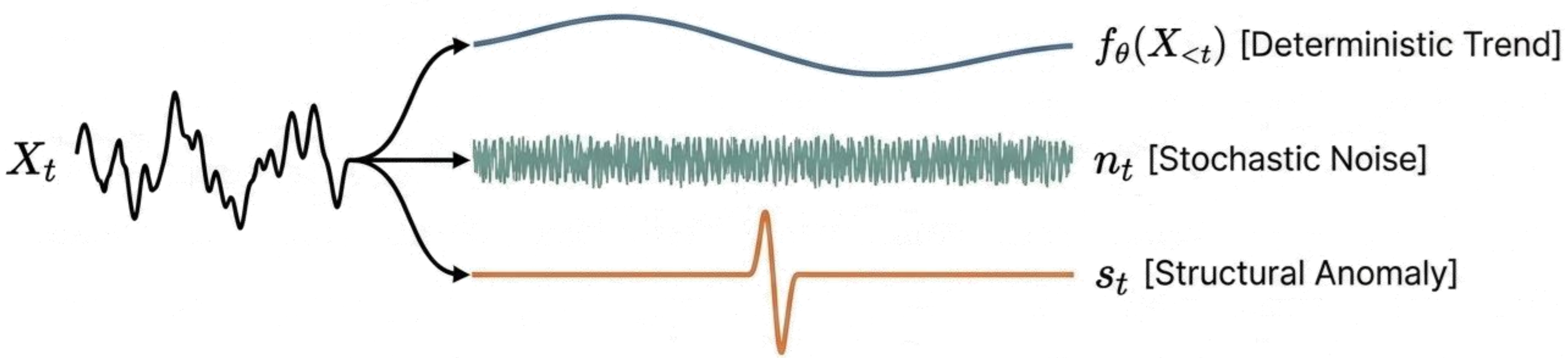
BEIJING UNIVERSITY OF TECHNOLOGY

*@Keenan Crane

Grounding in First Principles: Wold Decomposition

The Wold Decomposition:

$$\mathbf{X}_t = \underbrace{f_\theta(\mathbf{X}_{<t})}_{\text{Deterministic Trend}} + \underbrace{\mathbf{n}_t}_{\text{Stochastic Noise}} + \underbrace{\mathbf{s}_t}_{\text{Structural Anomaly}}$$



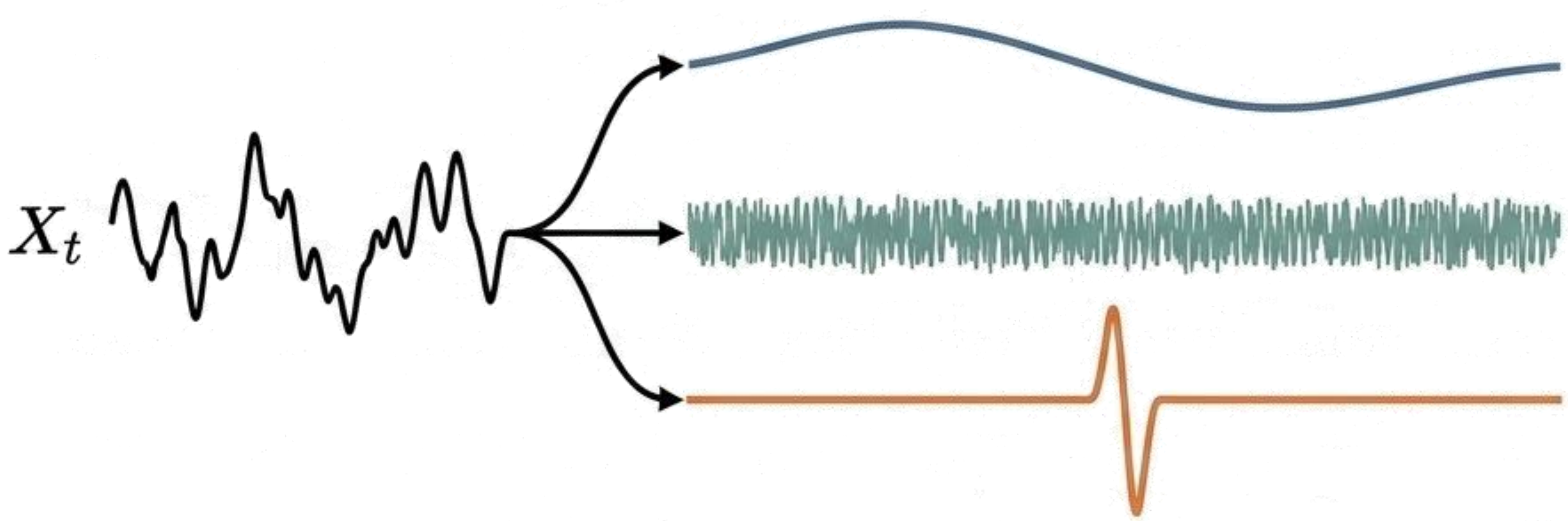
f_θ is a **deterministic function** capturing predictable dynamics.

In the reconstruction-based paradigm, a deep neural network **approximates** f_θ. Ideally, if the model is perfect, the residual should converge to the stochastic noise $\mathbf{R}_t \to \mathbf{n}_t$.

Grounding in First Principles: Wold Decomposition

The Wold Decomposition:

$$\mathbf{X}_t = \underbrace{f_\theta(\mathbf{X}_{<t})}_{\text{Deterministic Trend}} + \underbrace{\mathbf{n}_t}_{\text{Stochastic Noise}} + \underbrace{\mathbf{s}_t}_{\text{Structural Anomaly}}$$



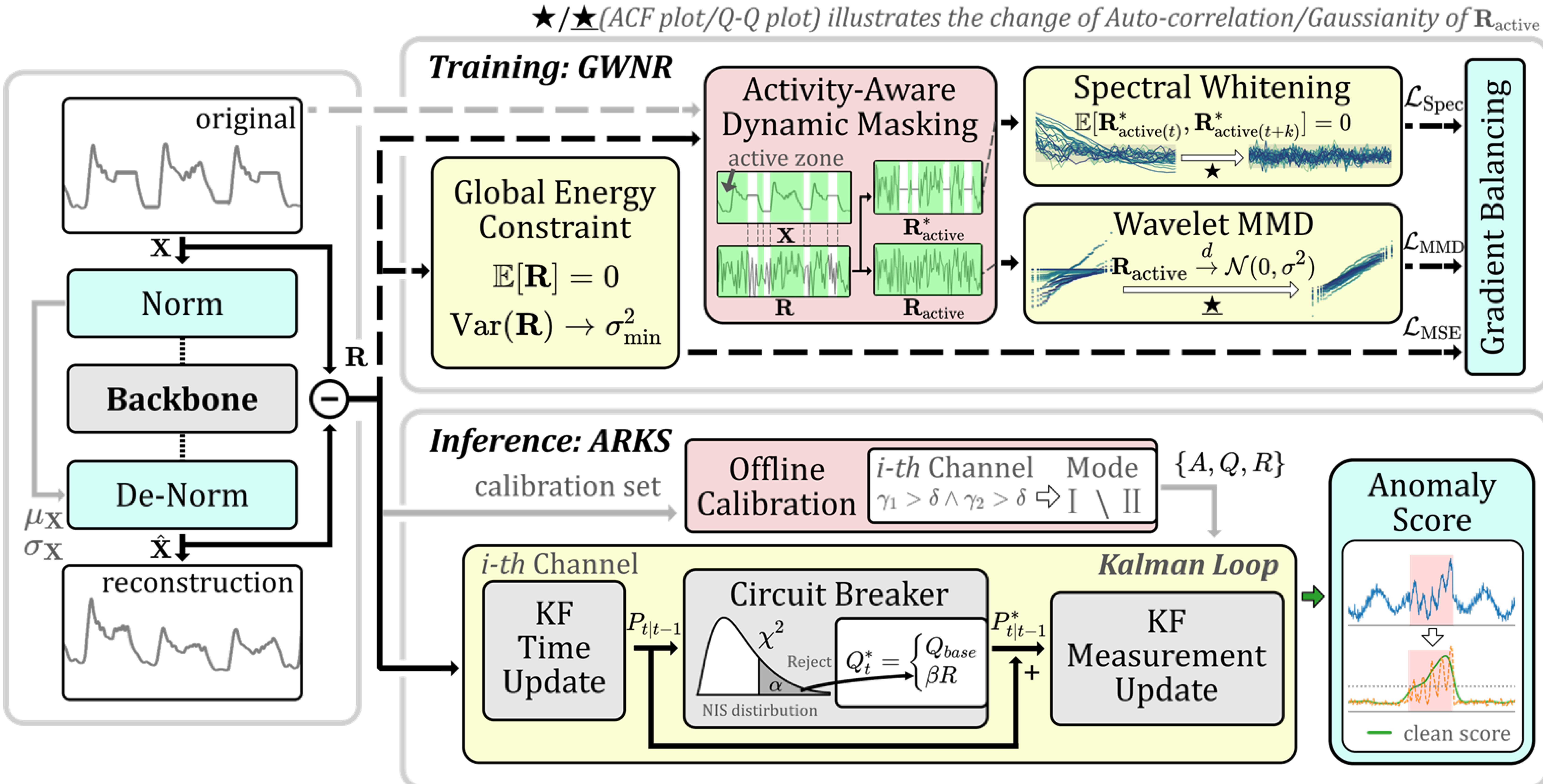
Ideally, if the model is perfect, the residual should converge to the stochastic noise $\mathbf{R}_t \to \mathbf{n}_t$.

To reliably detect sparse anomalies ($\mathbf{s}_t$), we must strictly engineer the remaining stochastic noise ($\mathbf{n}_t$) to match a valid **Null Hypothesis: Gaussian White Noise.**

COGNOS: Universal Enhancement for Time Series Anomaly Detection via Constrained Gaussian-Noise Optimization and Smoothing

Methodology

The COGNOS Architecture



Training: Enforces statistical preconditions explicitly as soft constraints.

Inference: Exploits engineered properties via optimal statistical filtering.

BEIJING UNIVERSITY OF TECHNOLOGY

COGNOS: Universal Enhancement for Time Series Anomaly Detection via Constrained Gaussian-Noise Optimization and Smoothing

Methodology

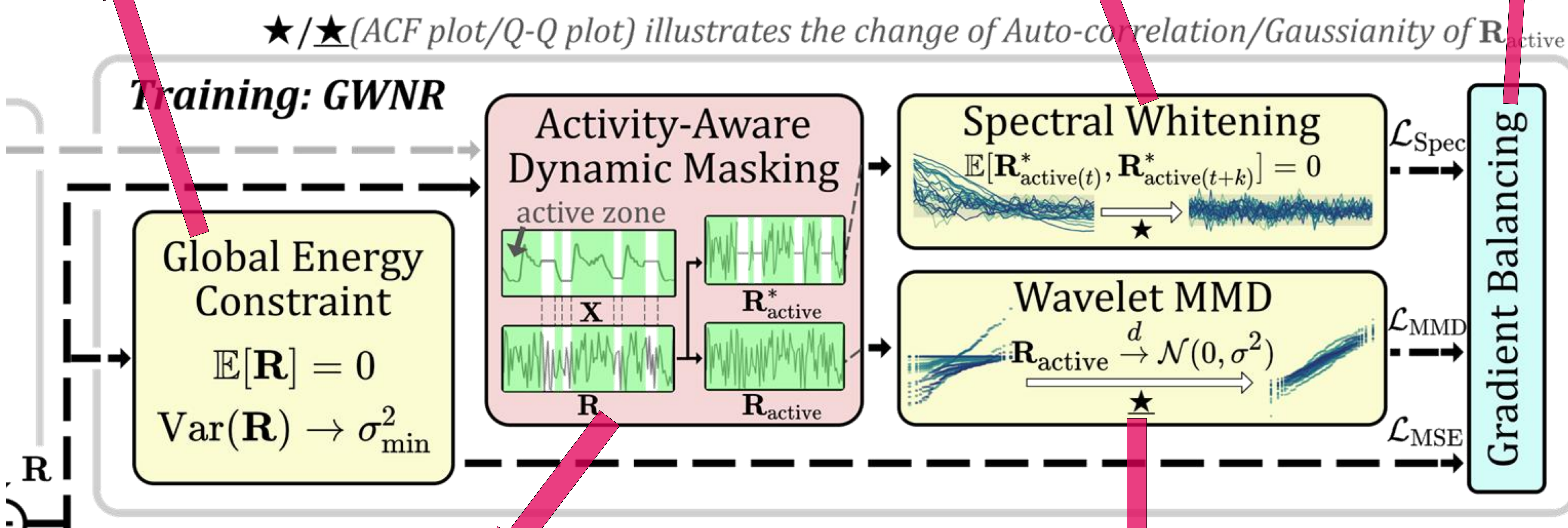
The COGNOS Architecture

MSE: Anchors the fundamental trend reconstruction fidelity.

Eliminates temporal autocorrelation.

$$\alpha_k = \mathrm{sg}\left(\frac{G_{\text{main}}}{G_k + \delta}\right) \cdot \mathbb{I}(\mathcal{L}_k > \delta),$$

$$\nabla_\theta \mathcal{L}_{\text{total}} = \nabla_\theta \mathcal{L}_{\text{MSE}} + \sum_k \alpha_k \nabla_\theta \mathcal{L}_k.$$



A mask ensures Gaussian whitening is only applied to **valid, active** signal transitions.

Satisfies optimal filter preconditions (Gaussian) **without destroying trend fitting.**

COGNOS: Universal Enhancement for Time Series Anomaly Detection via Constrained Gaussian-Noise Optimization and Smoothing

Methodology

The COGNOS Architecture

MSE: Anchors the fundamental trend reconstruction fidelity.

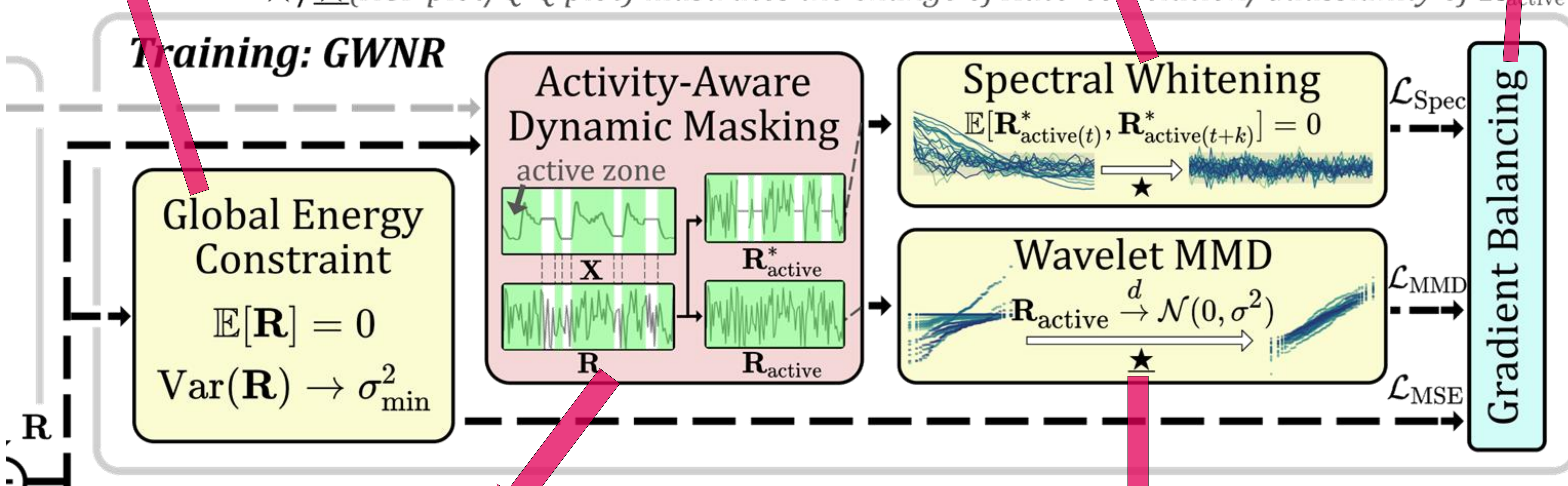
$$\mathcal{L}_{\text{Spec}} = \sum_f \tilde{S}(f) \log \tilde{S}(f) + \log K,$$

Eliminates temporal autocorrelation.

$$\alpha_k = \text{sg}\left(\frac{G_{\text{main}}}{G_k + \delta}\right) \cdot \mathbb{I}(\mathcal{L}_k > \delta),$$

$$\nabla_\theta \mathcal{L}_{\text{total}} = \nabla_\theta \mathcal{L}_{\text{MSE}} + \sum_k \alpha_k \nabla_\theta \mathcal{L}_k.$$

★/★(ACF plot/Q-Q plot) illustrates the change of Auto-correlation/Gaussianity of $\mathbf{R}_{\text{active}}$



$$\mathbf{M}_t = \max_{k \in [-w, w]} \mathbb{I}\left(\Delta_{t+k} > \delta\right).$$

$$\mathbf{R}_{\text{active}} = \mathbf{R} \odot \mathbf{M}.$$

A mask ensures Gaussian whitening is only applied to **valid, active** signal transitions.

$$(\mathbf{A}, \mathbf{D}) = \text{DWT}_{\text{Haar}}(\mathbf{R}_{\text{active}}).$$

$$\mathcal{L}_{\text{MMD}} = \mathbb{E}_{\tilde{\mathbf{D}}, \tilde{\mathbf{D}}'}[k(\tilde{\mathbf{D}}, \tilde{\mathbf{D}}')] + \mathbb{E}_{\mathcal{G}, \mathcal{G}'}[k(\mathcal{G}, \mathcal{G}')] - 2\mathbb{E}_{\tilde{\mathbf{D}}, \mathcal{G}}[k(\tilde{\mathbf{D}}, \mathcal{G})],$$

Satisfies optimal filter preconditions (Gaussian) **without destroying trend fitting.**

BEIJING UNIVERSITY OF TECHNOLOGY

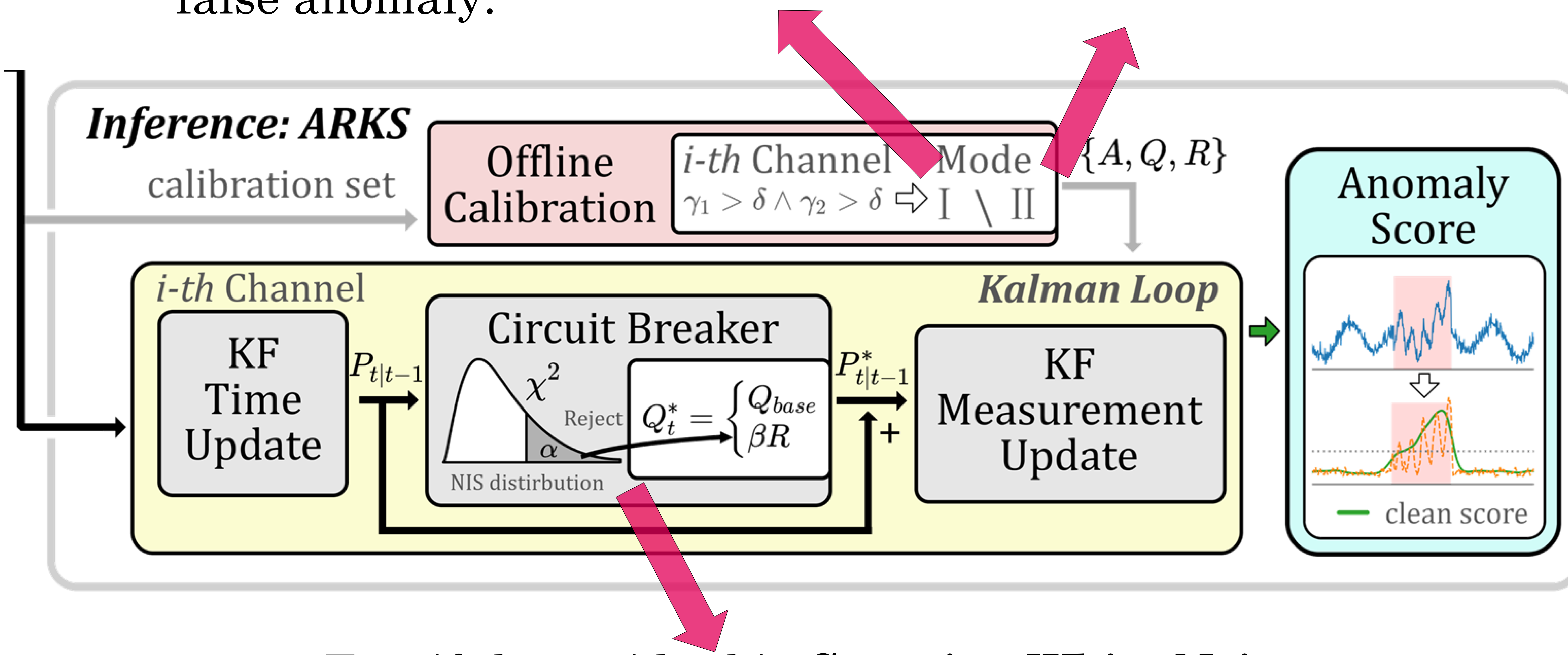
COGNOS: Universal Enhancement for Time Series Anomaly Detection via Constrained Gaussian-Noise Optimization and Smoothing

Methodology

The COGNOS Architecture

Mode I: Modeled as AR(1) process to track and absorb predictable leakage, preventing it from accumulating as a false anomaly.

Mode I I: Aggressive low-pass to suppress background noise and maximize Signal-to-Noise Ratio.



χ 2Test if the residual is **Gaussian White Noise**

Yes: suppress noise/

No: track and absorb anomaly.

ICML
International Conference On Machine Learning

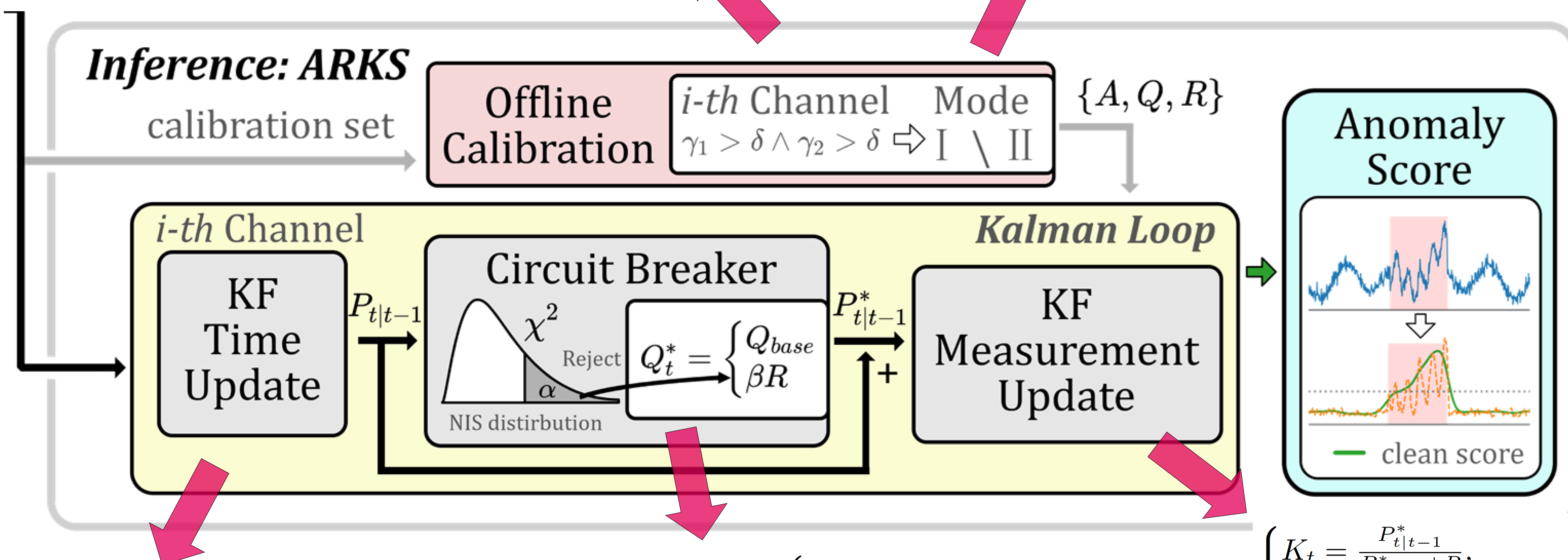
COGNOS: Universal Enhancement for Time Series Anomaly Detection via Constrained Gaussian-Noise Optimization and Smoothing

Methodology

The COGNOS Architecture

Mode I (Leakage-Tracking): If the backbone fails to capture complex periodicities, residuals exhibit autocorrelation. We model this as an AR(1) process via Yule-Walker estimation: $A = \gamma_2/\gamma_1$, $Q = \sigma_x(1 - A^2)$, and $R = \gamma_0 - \sigma_x$, where $\sigma_x = \gamma_1^2/\gamma_2$. This configuration forces the filter to track and explain the leakage, preventing it from incorrectly accumulating as an anomaly score.

Mode II (Noise-Suppression): If GWNR succeeds, residuals are white. We configure a Random Walk smoother with strong regularization: $A = 1$, $R = \gamma_0$, and $Q = \lambda R$ (with $\lambda \ll 1$). This acts as a low-pass filter ($\hat{x}_t \approx 0$), aggressively suppressing noise to enhance the Signal-to-Noise Ratio (SNR) for true anomalies.



$$\hat{x}_{t|t-1} = A\hat{x}_{t-1|t-1}.$$

$$\epsilon_t = \frac{(y_t - \hat{x}_{t|t-1})^2}{P_{t|t-1} + R}.$$

$$Q_t^* = \begin{cases} \beta R, & \text{if } \epsilon_t > \tau_\alpha \quad \text{(Anomaly)} \\ Q_{base}, & \text{otherwise} \quad \text{(Smoothing)} \end{cases},$$

$$P^*_{t|t-1} = A^2 P_{t-1|t-1} + Q_t^*,$$

$$\begin{cases} K_t = \frac{P^*_{t|t-1}}{P^*_{t|t-1} + R}, \\ \hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t(y_t - \hat{x}_{t|t-1}), \\ P_{t|t} = (1 - K_t)^2 P^*_{t|t-1} + K_t^2 R. \end{cases}$$

χ 2Test if the residual is **Gaussian White Noise**

Yes: suppress noise/ No: track and absorb anomaly.

BEIJING UNIVERSITY OF TECHNOLOGY

Experiments

Visual Proof

Mode I： Modeled as AR(1) process to track and absorb predictable leakage, preventing it from accumulating as a false anomaly.

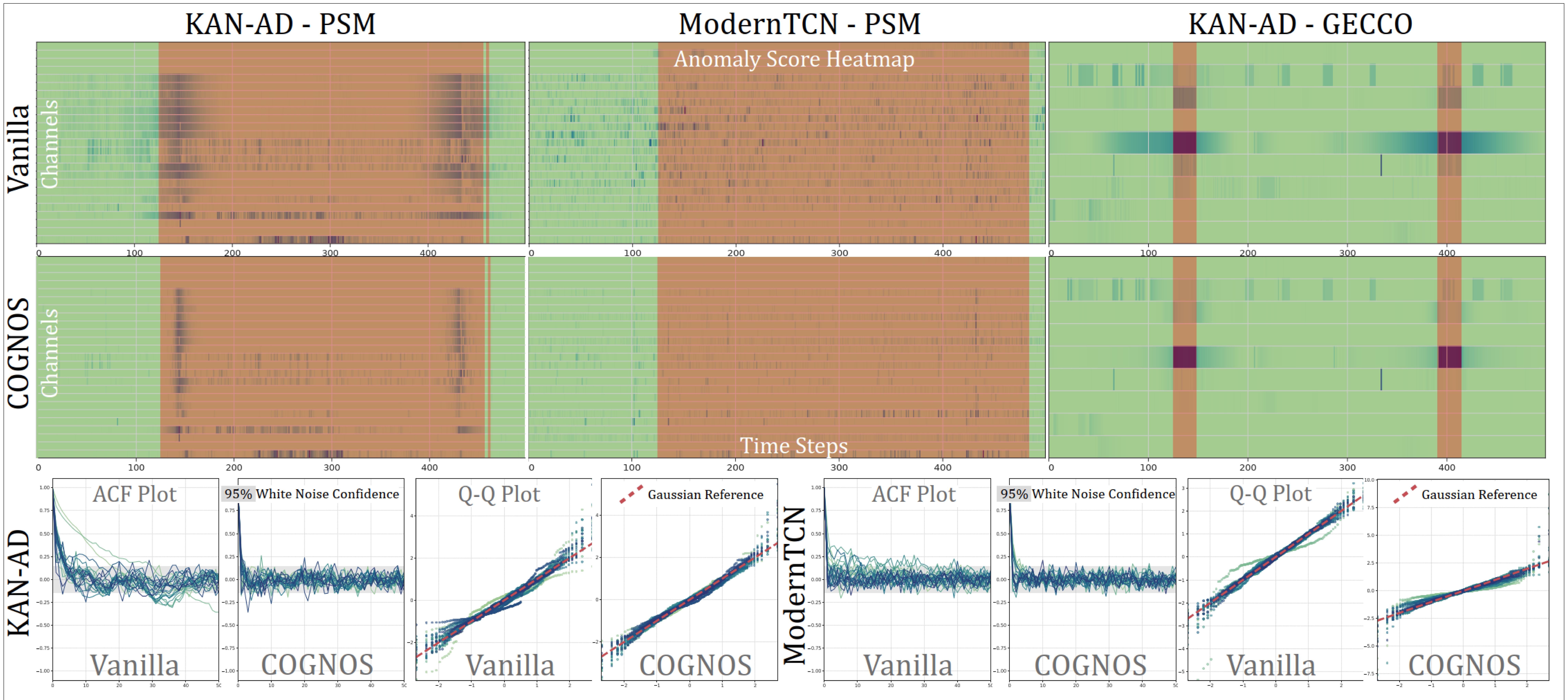


Figure 3. Qualitative Analysis on Signal Purity and Statistical Physics. **Top Row:** Anomaly score dynamics on GECCO and PSM datasets. **Bottom Row:** Statistical diagnostics of reconstruction residuals.

Experiments

Visual Proof

Mode I: Modeled as AR(1) process to track and absorb predictable leakage, preventing it from accumulating as a false anomaly.

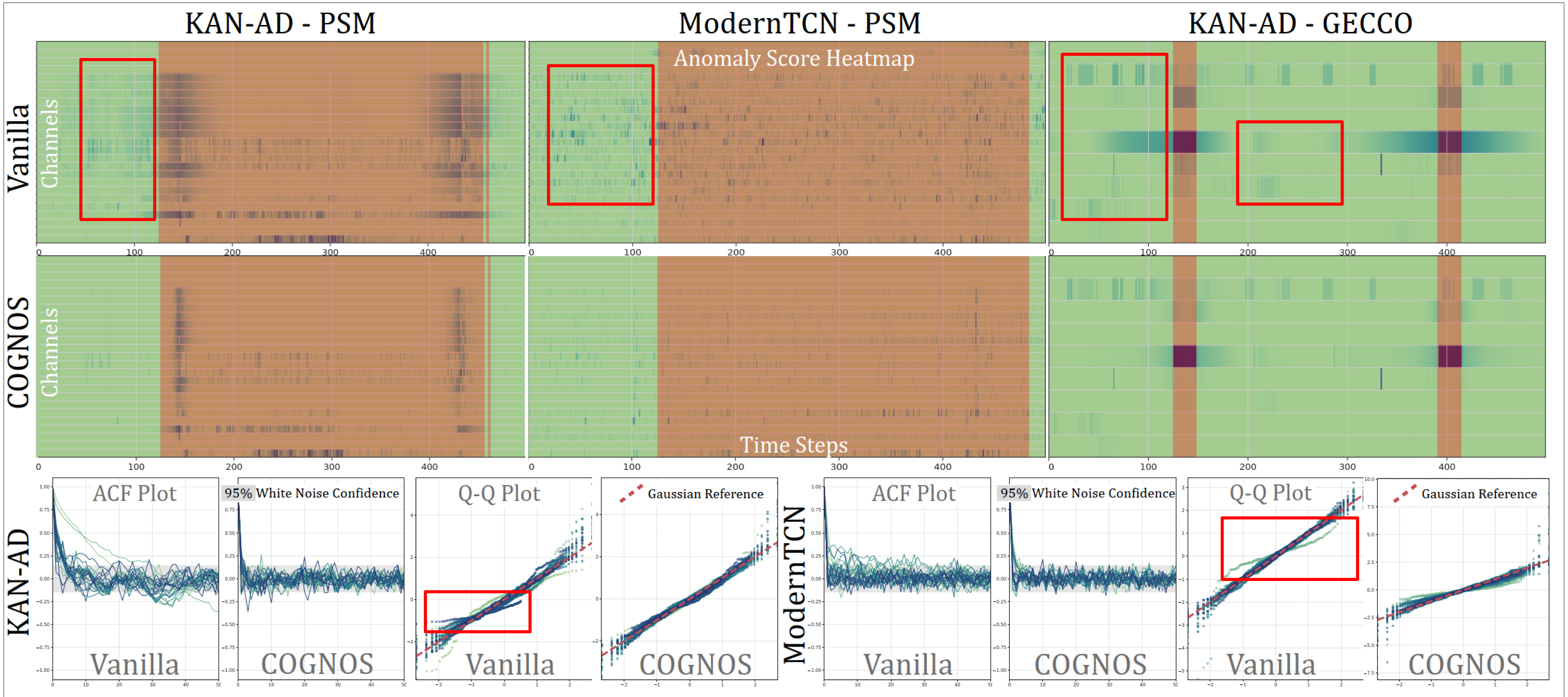


Figure 3. Qualitative Analysis on Signal Purity and Statistical Physics. **Top Row:** Anomaly score dynamics on GECCO and PSM datasets. **Bottom Row:** Statistical diagnostics of reconstruction residuals.

COGNOS: Universal Enhancement for Time Series Anomaly Detection via Constrained Gaussian-Noise Optimization and Smoothing

Experiments

Visual Proof

Table 2. Comparison of anomaly detection performance between the Vanilla method and COGNOS (Ours) across seven datasets. The table reports the **Affiliated-F1** metric. Best results are highlighted in **bold**.

Datasets	**GECCO**		**MSL**		**PSM**		**SMAP**		**SWAN**		**SWaT**		**UCR**	
Models	Vanilla	Ours	Vanilla	Ours	Vanilla	Ours	Vanilla	Ours	Vanilla	Ours	Vanilla	Ours	Vanilla	Ours
Autoformer	0.3929	**0.9426**	0.4008	**0.9388**	0.5292	**0.8743**	0.6266	**0.8521**	0.1828	**0.3604**	0.3967	**0.9571**	0.4702	**0.7427**
CrossAD	0.3115	**0.9320**	0.4135	**0.9386**	0.5904	**0.8628**	0.4871	**0.7857**	0.1829	**0.5374**	0.2973	**0.9511**	0.3383	**0.7412**
DLinear	0.4549	**0.9384**	0.5654	**0.9432**	0.7728	**0.8615**	0.6728	**0.8437**	0.2268	**0.4402**	0.4283	**0.9542**	0.4782	**0.7325**
KANAD	0.3785	**0.9465**	0.5360	**0.9157**	0.6223	**0.8628**	0.6666	**0.8491**	0.2066	**0.3936**	0.2905	**0.9543**	0.4409	**0.7279**
LSTMAE	0.3265	**0.9463**	0.5501	**0.9460**	0.6798	**0.8629**	0.6183	**0.7770**	0.2193	**0.4329**	0.4677	**0.9521**	0.4074	**0.7328**
MICN	0.9141	**0.9406**	0.4898	**0.9150**	0.7545	**0.8708**	0.6915	**0.7856**	0.2085	**0.4307**	0.6111	**0.9575**	0.6402	**0.7540**
ModernTCN	0.5646	**0.9331**	0.3545	**0.9177**	0.6903	**0.8724**	0.7321	**0.8262**	0.1850	**0.4614**	0.2495	**0.9643**	0.5369	**0.7452**
TimeMixer++	0.7936	**0.9356**	0.5179	**0.9450**	0.7018	**0.8661**	0.3764	**0.4063**	0.1889	**0.5374**	0.2946	**0.9509**	0.6232	**0.7390**
TimesNet	0.8417	**0.9342**	0.4868	**0.9171**	0.7828	**0.8672**	0.5289	**0.8301**	0.1961	**0.5249**	0.5873	**0.9601**	0.6505	**0.7619**
AVG	0.5475	**0.9388**	0.4794	**0.9308**	0.6804	**0.8668**	0.6001	**0.7729**	0.1997	**0.4577**	0.4026	**0.9557**	0.5095	**0.7419**

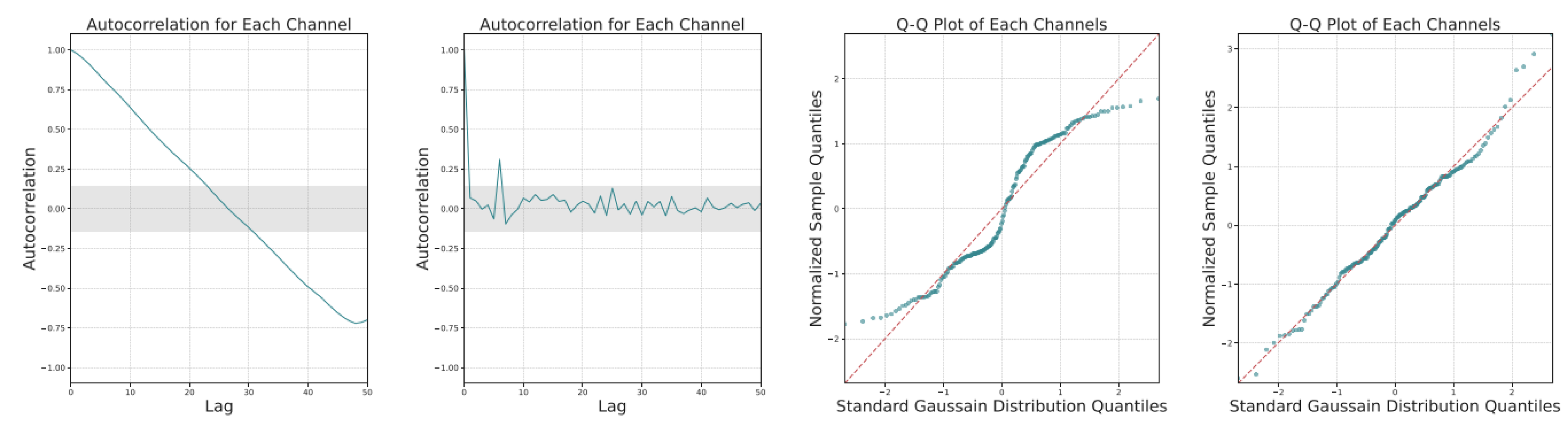


Figure 7. Residual analysis on the SMAP dataset using CrossAD. Left: vanilla method, Right: COGNOS.

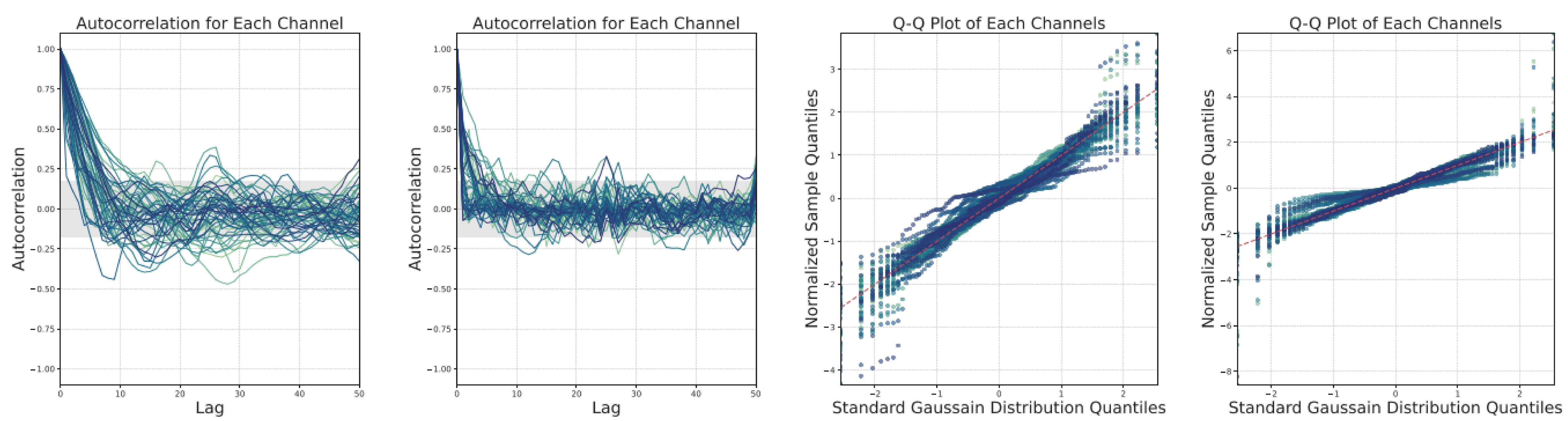


Figure 5. Residual analysis on the SWaT dataset using TimesNet. Left: vanilla method, Right: COGNOS.

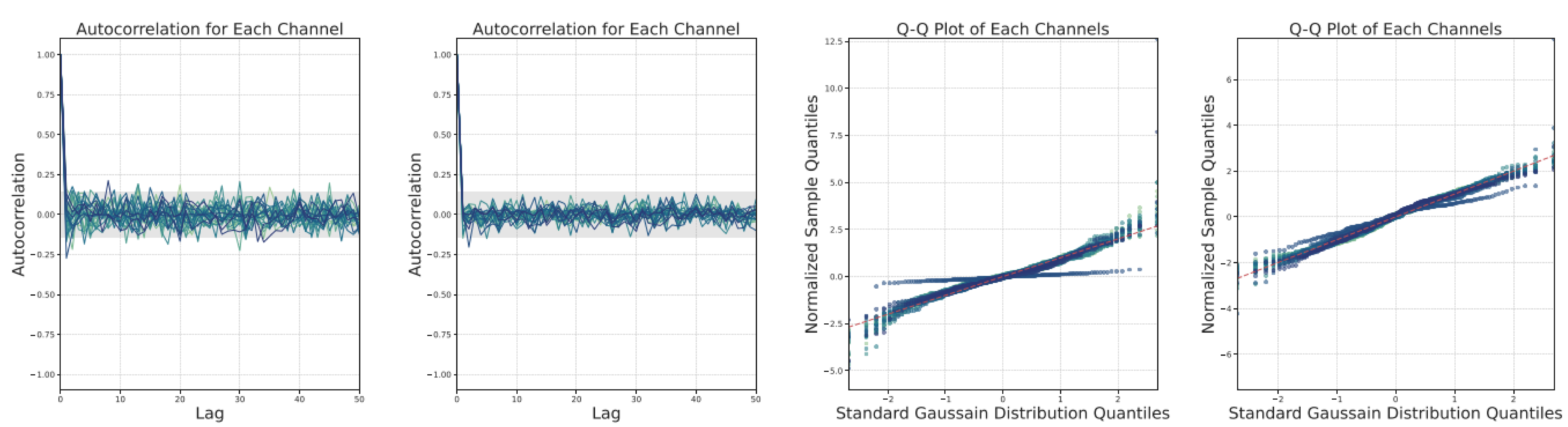


Figure 6. Residual analysis on the PSM dataset using TimeMixer++. Left: vanilla method, Right: COGNOS.

Table 3. Ablation Study of COGNOS Components and Filtering Methods. The best results are highlighted in **bold**, and the second-best are <u>underlined</u>. The full experiment results and more details can be found in the Appendix D.3.

Datasets				**GECCO**		**MSL**		**PSM**	
GWNR	ARKS	MA	LP	Std-F1	Aff-F1	Std-F1	Aff-F1	Std-F1	Aff-F1
✗				0.4738	0.3785	0.8115	0.5360	0.8946	0.6223
✗	✓			<u>0.4740</u>	<u>0.4845</u>	0.8083	0.5285	<u>0.9433</u>	<u>0.6689</u>
✓				0.3440	0.4098	<u>0.8704</u>	<u>0.5917</u>	0.9160	0.4910
✓		✓		0.3531	0.4123	0.1626	0.2564	0.7979	0.1845
✓			✓	0.3539	0.4236	0.1662	0.2584	0.8006	0.2302
✓	✓			**0.7466**	**0.9465**	**0.9165**	**0.9157**	**0.9744**	**0.8628**

BEIJING UNIVERSITY OF TECHNOLOGY

COGNOS: Universal Enhancement for Time Series Anomaly Detection via Constrained Gaussian-Noise Optimization and Smoothing

Experiments

Visual Proof

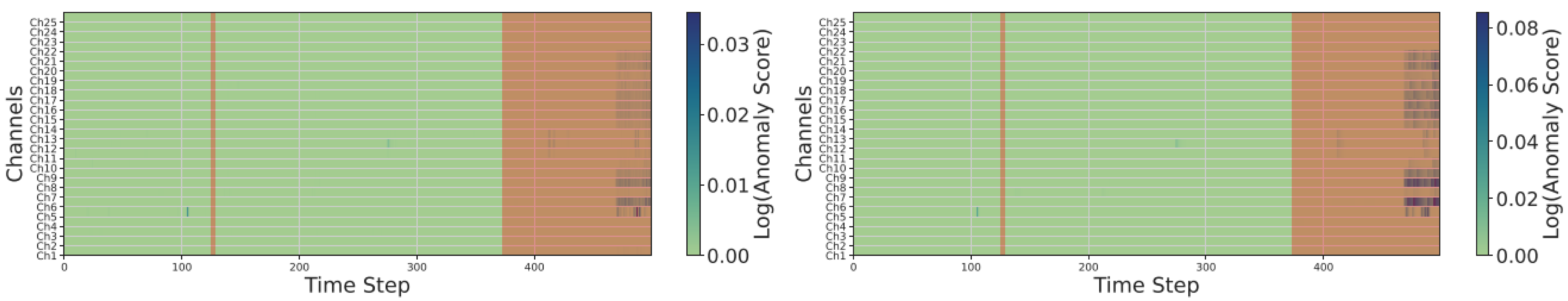


Figure 8. Qualitative comparison of anomaly scores on PSM using TimeMixer++. Left: vanilla method, Right: COGNOS.

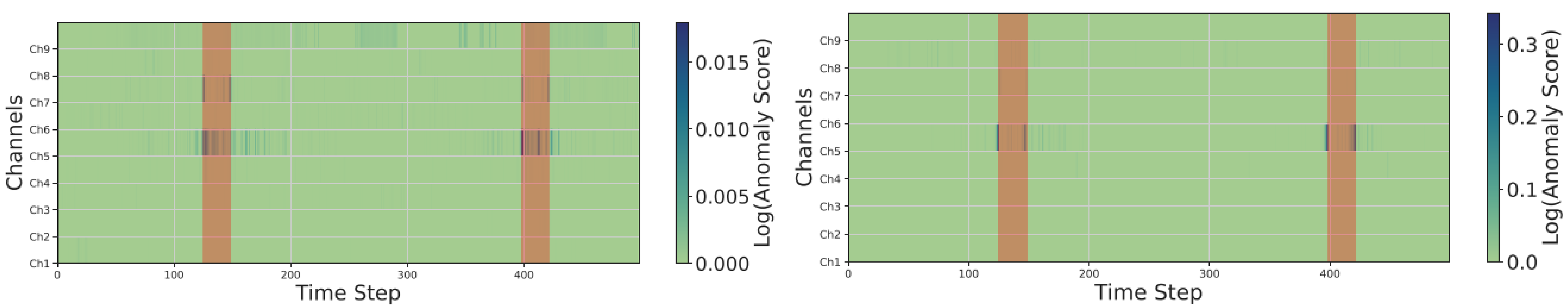


Figure 9. Qualitative comparison of anomaly scores on GECCO using ModernTCN. Left: vanilla method, Right: COGNOS.

Table 4. Computational Efficiency Analysis. We report the **average execution time (ms) per iteration** (Training) and **per sample point** (Inference), (+) denotes the additional overhead introduced by our COGNOS framework. We set batch size $= 128$ and sequence length $= 128$

Datasets	Models	Autoformer		KANAD		TimesNet	
		Vanilla	COGNOS	Vanilla	COGNOS	Vanilla	COGNOS
GECCO	Train	62.53	+41.648	39.94	+48.4	101.34	+171.49
	Inference	0.1782	+0.1274	0.0847	+0.1467	0.1694	+0.1513
PSM	Train	53.30	+82.66	97.31	+206.68	103.89	+257.55
	Inference	0.2150	+0.1031	0.2459	+0.1387	0.2054	+0.1327
SWAN	Train	87.64	+164.11	216.54	+355.01	153.57	+359.27
	Inference	0.2482	+0.0838	0.3546	+0.1268	0.2190	+0.1402

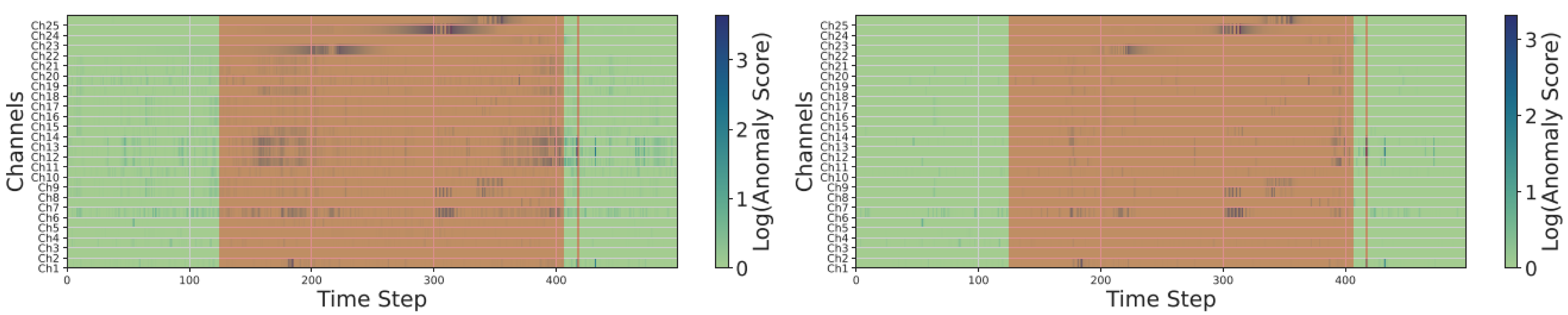


Figure 10. Qualitative comparison of anomaly scores on PSM using KAN-AD. Left: vanilla method, Right: COGNOS.

BEIJING UNIVERSITY OF TECHNOLOGY

ICML International Conference On Machine Learning

COGNOS: Universal Enhancement for Time Series Anomaly Detection via **Co**nstrained **G**aussian-**N**oise **O**ptimization and **S**moothing

Wenlong Shang, Shihao Tian, Xutong Wan, Peng Chang
Beijing Key Laboratory of Computational Intelligence and Intelligent System,
Beijing University of Technology, China.

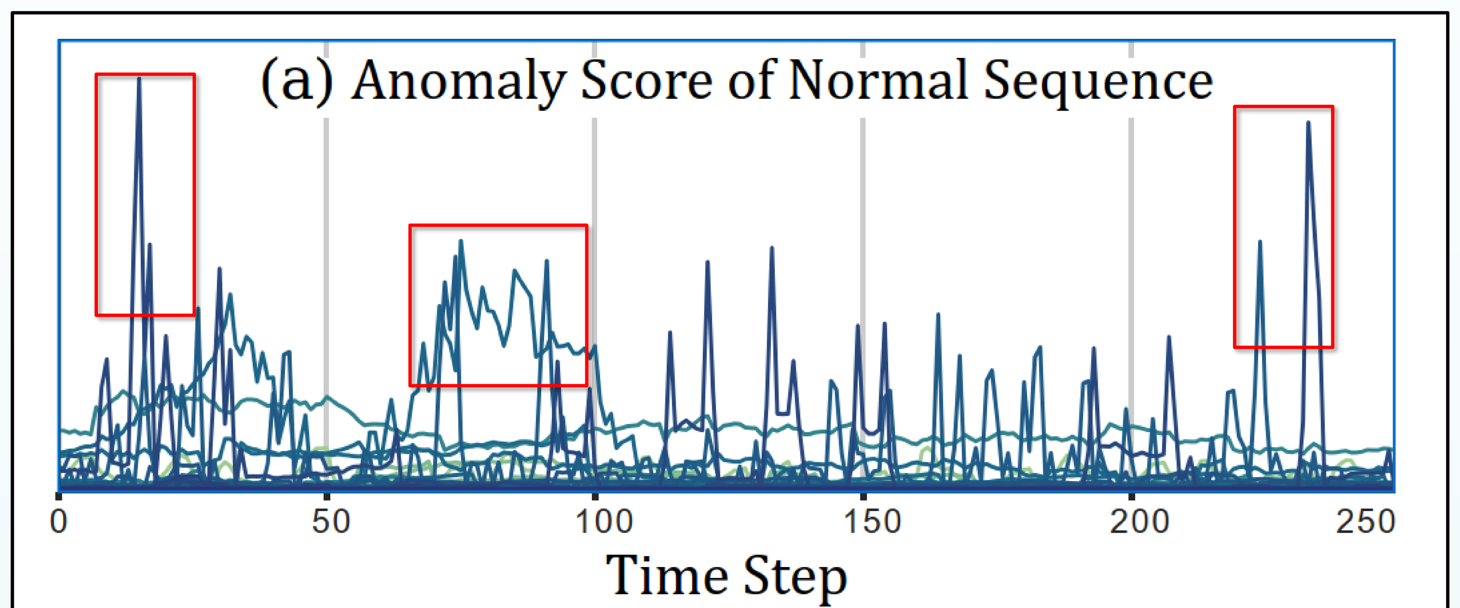
Thanks

BEIJING UNIVERSITY OF TECHNOLOGY

COGNOS: Universal Enhancement for Time Series Anomaly Detection via Constrained Gaussian-Noise Optimization and Smoothing

Wenlong Shang, Shihao Tian, Xutong Wan, Peng Chang
School of Information Science and Technology, Beijing University of Technology

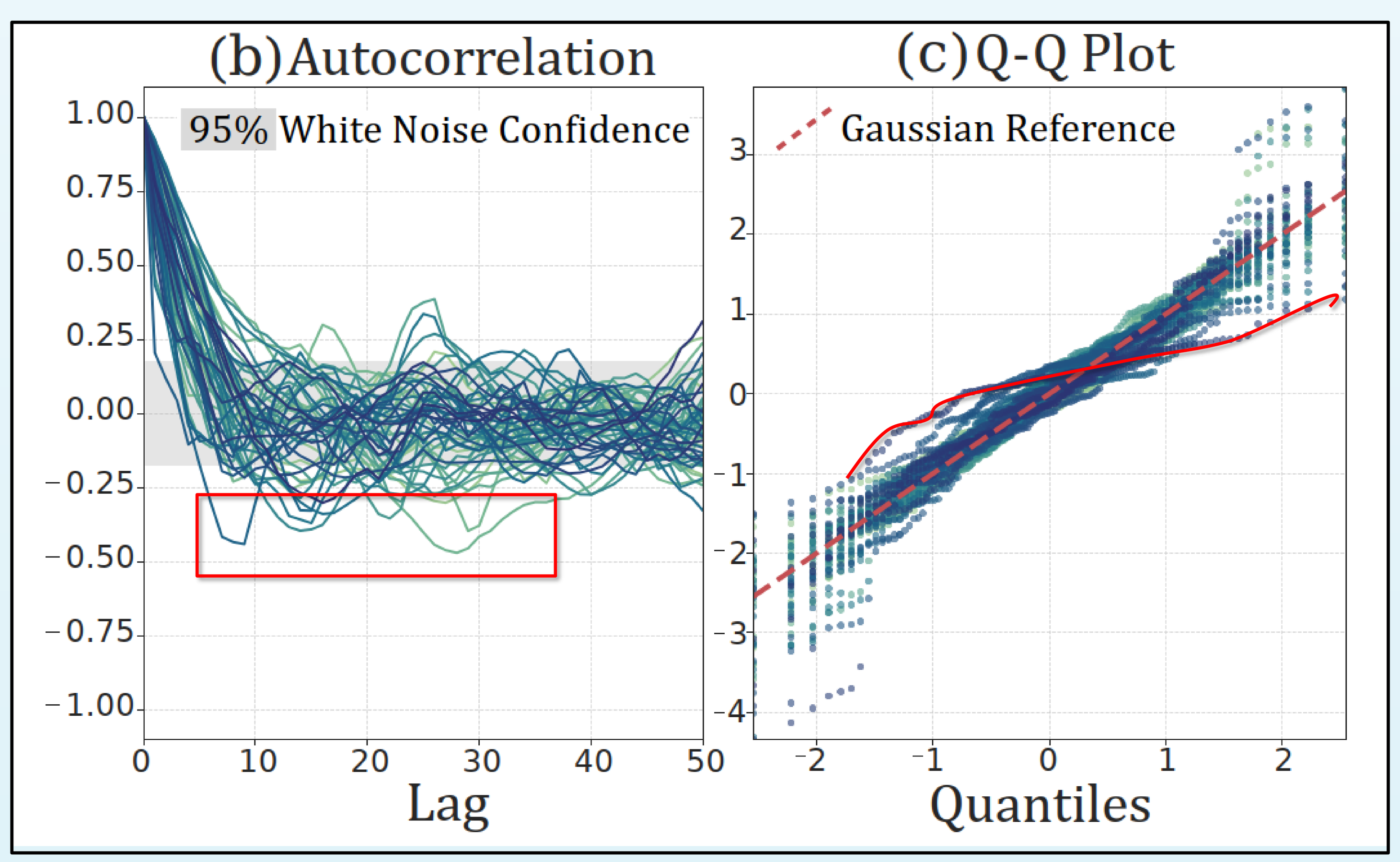
Why Reconstructive TSAD Fail



Anomaly scores are **Noisy**

Residuals are :

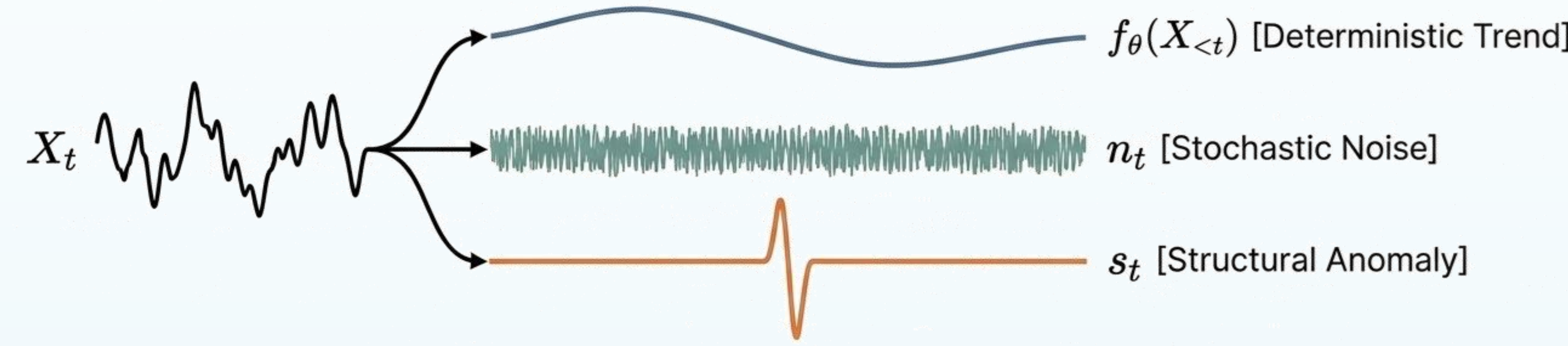
Correlated and **Non-Gaussian**



- **MSE is statistically justified only when residuals are i.i.d. Gaussian.**
- **Real TSAD residuals are correlated and non-Gaussian.**
- **Noisy residuals create unstable anomaly scores and false alarms.**

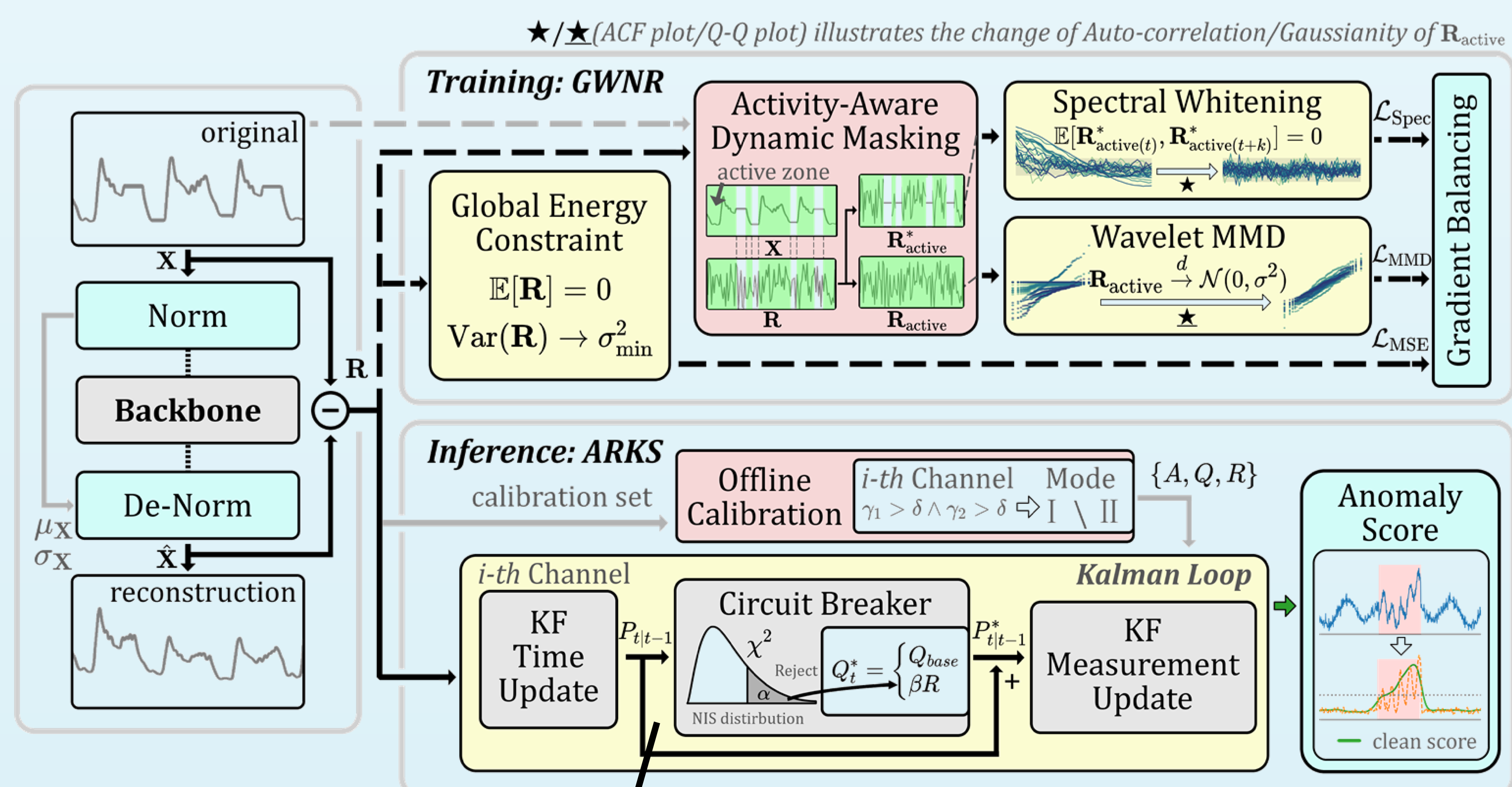
Make Residuals Testable

$$\mathbf{X}_t = \underbrace{f_\theta(\mathbf{X}_{<t})}_{\text{Deterministic Trend}} + \underbrace{\mathbf{n}_t}_{\text{Stochastic Noise}} + \underbrace{\mathbf{s}_t}_{\text{Structural Anomaly}},$$



Normal residuals should be a **statistical null hypothesis**, not just small reconstruction errors.

Methodology



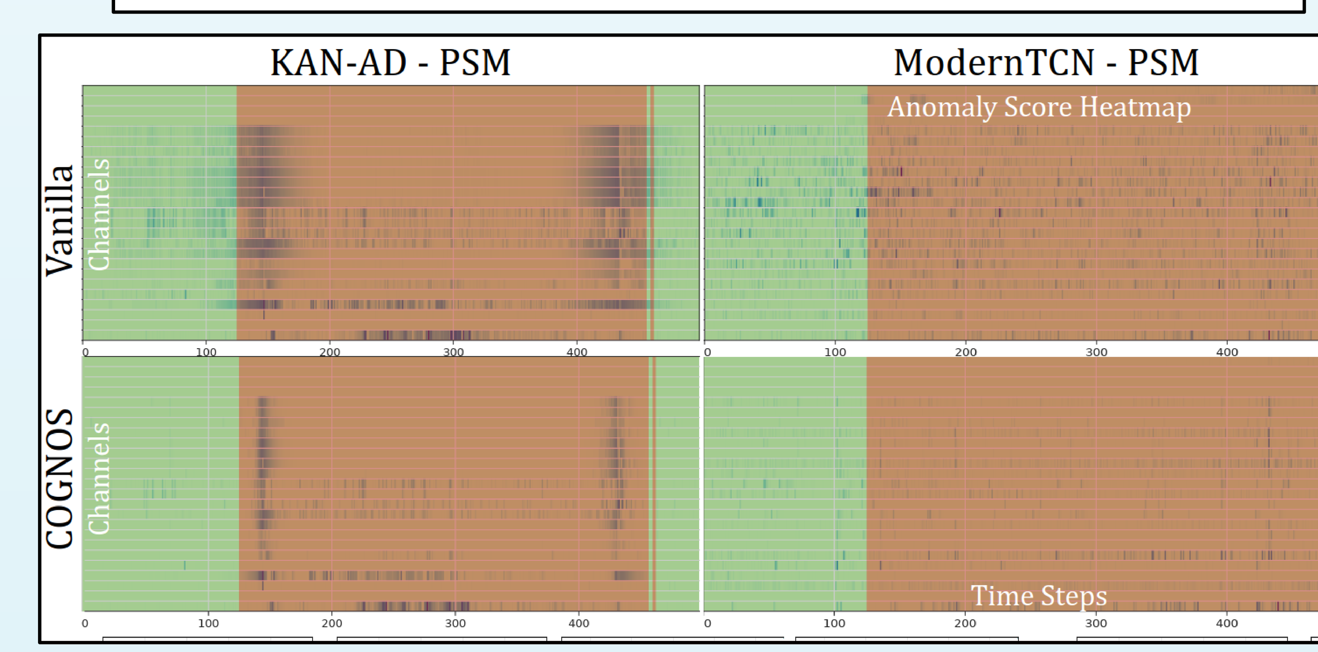
$$\epsilon_t = \frac{(y_t - \hat{x}_{t|t-1})^2}{P_{t|t-1} + R}.$$

Reject Gaussian-white null ➡ Circuit Breaker

Results

Table 1. Multi-metric evaluation of anomaly detection performance between the Vanilla method and COGNOS (Ours) across three datasets. Best results are highlighted in **bold**.

Models	Datasets	MSL		PSM		SWAN	
	Metircs	Vanilla	Ours	Vanilla	Ours	Vanilla	Ours
Autoformer	Std-F1	0.7724	**0.9321**	0.9098	**0.9759**	0.7336	**0.7950**
	Aff-F1	0.4008	**0.9388**	0.5292	**0.8743**	0.1828	**0.3604**
	R-A-R	0.6942	**0.7084**	0.6667	**0.6954**	0.9398	**0.9433**
	R-A-P	0.2366	**0.2446**	0.4851	**0.5235**	0.9362	**0.9382**
	V-R	0.6684	**0.6884**	0.6349	**0.6650**	0.9532	**0.9542**
	V-P	0.1943	**0.2039**	0.4352	**0.4740**	0.9074	**0.9125**
KANAD	Std-F1	0.8115	**0.9165**	0.8946	**0.9744**	0.7384	**0.8049**
	Aff-F1	0.5360	**0.9157**	0.6223	**0.8628**	0.2066	**0.3936**
	R-A-R	0.6646	**0.6806**	**0.7005**	0.6799	0.9227	**0.9374**
	R-A-P	0.2030	**0.2135**	**0.5092**	0.4740	0.9210	**0.9301**
	V-R	0.6450	**0.6596**	**0.6846**	0.6624	0.9414	**0.9460**
	V-P	0.1742	**0.1799**	**0.4724**	0.4432	0.8950	**0.8983**
TimesNet	Std-F1	0.7944	**0.9160**	0.9581	**0.9751**	0.7371	**0.8313**
	Aff-F1	0.4868	**0.9171**	0.7828	**0.8672**	0.1961	**0.5249**
	R-A-R	**0.6968**	0.6663	0.6742	**0.7105**	0.9387	**0.9413**
	R-A-P	**0.2323**	0.2067	0.4932	**0.5290**	0.9366	**0.9394**
	V-R	**0.6713**	0.6500	0.6408	**0.6805**	**0.9544**	0.9514
	V-P	**0.1925**	0.1760	0.4402	**0.4787**	0.9128	**0.9161**



Conclusion

- Residual engineering is complementary to architecture design: make normal errors statistically clean, then detect anomalies as principled deviations.
- **Best suited for** MSE reconstruction backbones and onset detection.