DRAFTS: TIME-AWARE DECOMPOSITION WITH RESIDUAL CORRECTION FOR TIME SERIES MODELING

Anonymous authors
Paper under double-blind review

ABSTRACT

Real-world time series contain evolving underlying dynamics with irregular variations that lack stable temporal patterns and are often referred to as noise. Existing methods address this mixture by filtering frequencies or suppressing noisy observations. They either miss temporal evolution or risk suppressing useful dynamics. We propose DRAFTS, a model-agnostic framework that aims to reduce noise while preserving evolving dynamics through time-aware **D**ecomposition with **R**esidu**A**l correction **F**or **T**ime **S**eries. DRAFTS uses features derived from instantaneous amplitude and frequency to guide decomposition into a primary component intended to capture underlying dynamics. A task-specific backbone models the primary component, while a lightweight correction module uses residual information to correct the backbone output. Across four time series modeling tasks, DRAFTS improves six diverse backbones, demonstrating its effectiveness. Code is at https://anonymous.4open.science/r/DrafTS-A14F/.

1 INTRODUCTION

Deep learning is widely used for time series modeling, supporting tasks such as forecasting (Bi et al., 2023), classification (Kiyasseh et al., 2021) and anomaly detection (Xu et al., 2022). Recent recurrent (Lai et al., 2018), convolutional (Wu et al., 2023), MLP-based (Wang et al., 2024a; Zeng et al., 2023), and attention-based (Zhang & Yan, 2023; Liu et al., 2023) architectures have improved temporal dependency modeling. Yet these improvements do not necessarily enable models to distinguish underlying dynamics from noise[1]. This distinction becomes harder in real-world nonlinear and nonstationary time series, where the dynamics themselves evolve over time (Huang et al., 1998). Reducing noise while preserving these evolving dynamics therefore remains challenging.

Existing deep learning methods pursue this balance mainly via two paradigms (Figure 1 (a,b)). The first is **frequency-wise filtering**, which selects spectral components in time-series representations. FITS (Xu et al., 2024) uses a fixed frequency cutoff, while FiLM (Zhou et al., 2022a) and FilterNet (Yi et al., 2024) retain selected Fourier representations. However, such global filtering cannot track the temporal evolution of the time series. The second is **observation-wise suppression**, which uses noise estimates to regulate learning in the time domain. RobustTSF (Cheng et al., 2024) discards anomalous windows, Selective Learning (Fu et al., 2025) masks uncertainty or anomaly in the loss and DropoutTS (Zhong et al., 2026) adapts instance-level dropout. However, suppressing noisy observations as a whole may also hinder learning useful dynamics within them. Moreover, both paradigms may reduce the impact of the noise at the expense of useful dynamics.

Decomposition offers a possible way to distinguish evolving dynamics from noise. However, commonly used Fourier and discrete cosine transforms (Cooley & Tukey, 1965; Ahmed et al., 1974) apply fixed-frequency bases, which fail to track the temporal evolution of nonlinear, nonstationary series. As shown in Figure 1 (c), we propose a third paradigm: **Time-aware decomposition with correction**. Instead of frequency filtering or observation-wise suppression, we extract a primary component to capture evolving dynamics and recover missed dynamics from the residual.

[1]Throughout this paper, underlying dynamics refer to structured temporal patterns, such as trends and periodic behavior, whereas noise refers to irregular variations that do not follow stable temporal patterns.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

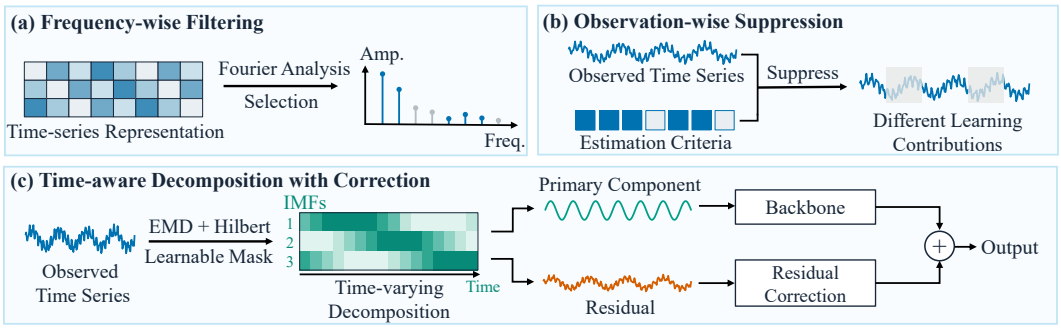


Figure 1: **Modeling paradigms for underlying dynamics and noise.** *(a) Frequency-wise filtering* overlooks temporal evolution. *(b) Observation-wise suppression* limits learning from noisy observations, suppressing useful dynamics within them. *(c) Time-Aware Decomposition with Correction* tracks evolving dynamics and exploits dynamics left in the residual.

To this end, we propose DRAFTS, a model-agnostic framework that uses time-aware **D**ecomposition to retain evolving dynamics and corrects backbone predictions with **R**esidu**A**l information **F**or **T**ime **S**eries modeling. Specifically, **in the decomposition stage**, we track time-varying oscillations using instantaneous frequency obtained through the Hilbert transform (Gabor, 1946; Boashash, 1992). Since instantaneous frequency can be difficult to interpret when fast and slow oscillations coexist, we first apply empirical mode decomposition (EMD) to decompose the input into intrinsic mode functions (IMFs) without predefined bases (Huang et al., 1998). Secondly, we apply the Hilbert transform to each IMF to construct its analytic signal, from which we derive its instantaneous amplitude and frequency. Figure 2 illustrates this process. Features derived from these quantities guide a time-aware mask that decomposes the input into a primary component and a residual. **In the modeling stage**, a task-specific backbone models the primary component, while a lightweight *Residual-Aware Correction* module exploits useful dynamics remaining in the residual to correct the backbone output. The module processes the residual sequence recurrently with fixed hidden weights, producing nonlinear representations; only the output layer is trained, so the module introduces few trainable parameters. Finally, we combine the residual correction with the backbone output to produce the task prediction. In summary, our contributions are as follows:

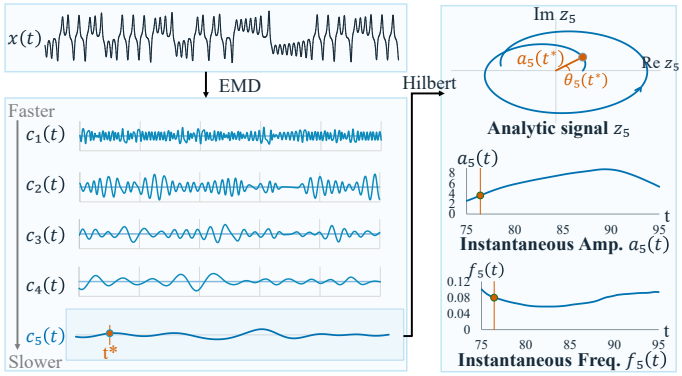


Figure 2: **EMD and Hilbert transform example.** The first five IMFs are shown. Applying the Hilbert transform to $c_5(t)$ yields $z_5(t)$ and its instantaneous amplitude and frequency. Re and Im denote real and imaginary parts; orange markers mark time t^*.

- **Novel Paradigm:** We introduce *Time-Aware Decomposition with Correction*. DRAFTS uses time-aware masks to capture evolving dynamics in a primary component and exploits useful dynamics left in the residual by imperfect decomposition to correct the backbone output.
- **Universal Compatibility:** DRAFTS augments diverse task-specific backbones with time-aware decomposition and lightweight residual correction, without modifying their internal architectures.
- **Superior Performance:** Across four time series modeling tasks, DRAFTS improves diverse backbones, demonstrating its generality and effectiveness.

2 RELATED WORK

Deep Learning for Time Series Modeling. Learning temporal dynamics that support accurate prediction is a central objective of deep learning for time series modeling. Earlier approaches use

patches (Zhou et al., 2021; Nie et al., 2023), dependency modeling (Zhang & Yan, 2023; Liu et al., 2023), linear prediction (Zeng et al., 2023), or multiscale mixing (Wang et al., 2024a). More recently, TimeFilter and L-Drive (Hu et al., 2025; Zhang et al., 2026) refine dependency and change modeling, while large time series models extend pre-training (Ansari et al., 2024; Woo et al., 2024), long contexts (Liu et al., 2025), and sparse experts (Shi et al., 2025). However, stronger modeling capacity alone does not ensure that evolving dynamics are distinguished from noise.

Modeling paradigms for noise. Prior work limits noise effects via spectral filtering or regulating observation use. **Frequency-wise filtering** includes fixed cutoffs in FITS (Xu et al., 2024) and Fourier selection in FiLM (Zhou et al., 2022a) and FilterNet (Yi et al., 2024). However, global spectral operations fail to reveal how oscillations change within a window (Wang et al., 2024c). **Observation-wise suppression** instead adjusts learning according to estimated noise, through window rejection in RobustTSF (Cheng et al., 2024), loss masking of uncertain or anomalous timesteps in Selective Learning (Fu et al., 2025), and instance-adaptive dropout in DropoutTS (Zhong et al., 2026). Yet rejecting or masking noisy observations may limit learning useful dynamics within them.

Signal Decomposition Methods. Autoformer (Wu et al., 2021) and FEDformer (Zhou et al., 2022b) use moving averages for trend–seasonal decomposition. As they do not aim to separate evolving dynamics from noise, we do not discuss them further and instead focus on methods relevant to this goal. Fourier (Cooley & Tukey, 1965) and discrete cosine transforms (DCT) (Ahmed et al., 1974) represent each window as a linear combination of fixed-frequency bases. For nonlinear and nonstationary data, global stationary bases can yield physically ambiguous frequencies. Short-time Fourier transform (STFT) (Allen & Rabiner, 1977) mitigates this issue with shorter windows, but fixed windows still limit tracking evolving oscillations. In contrast, EMD extracts time-domain oscillatory modes (Huang et al., 1998). With the Hilbert transform, it yields interpretable instantaneous amplitude and frequency, and is used for denoising across domains (Hassan, 2005; Quilfen & Chapron, 2021; Quilfen et al., 2022). However, threshold-based EMD denoising (Kopsinis & McLaughlin, 2009) may remove underlying dynamics. DRAFTS instead learns task-supervised, time-aware soft masks from analytic features and keeps the residual for correction.

3 METHODOLOGY

3.1 NOTATIONS AND PROBLEM FORMULATIONS

Let $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_T]^\top \in \mathbb{R}^{T \times N}$ be a multivariate series with T time steps and N variables. For $n = 1, \ldots, N$, let $x_{1:T}^{(n)} = (x_1^{(n)}, \ldots, x_T^{(n)})$ denote the sequence of the n-th variable. Let $\mathbf{Y}$ denote the task-specific target. For forecasting, $\mathbf{Y} \in \mathbb{R}^{S \times N}$ contains the subsequent S time steps. For classification, $\mathbf{Y} \in \{0, 1\}^C$ is the one-hot label vector, where C is the number of classes. For anomaly detection, $\mathbf{Y}$ is the next observation $\mathbf{x}_{t+1} \in \mathbb{R}^N$.

Figure 3 illustrates the overall framework of DRAFTS. Our method augments a task-specific backbone through two stages. The decomposition stage splits $\mathbf{X} = \mathbf{S}' + \mathbf{R}'$ into a primary component $\mathbf{S}'$ and the residual $\mathbf{R}'$, both in $\mathbb{R}^{T \times N}$. The modeling stage uses a task-specific backbone to model $\mathbf{S}'$ and a fixed reservoir with a trainable output layer to derive an output correction from $\mathbf{R}'$.

3.2 STAGE I: TIME-AWARE DECOMPOSITION

EMD and the Hilbert transform. We apply EMD independently to each variable $x_{1:T}^{(n)}$. We omit n for clarity. For $x_{1:T}$, EMD extracts oscillatory modes by iterative sifting (Huang et al., 1998). To extract the first IMF, set $h_t^{(0)} \leftarrow x_t$. At the q-th sifting step, EMD locates local extrema of $h_t^{(q)}$. Interpolating local maxima and minima gives the upper and lower envelopes $e_t^{+,(q)}$ and $e_t^{-,(q)}$, respectively. Their pointwise average defines the local envelope mean $\mu_t^{(q)} = \frac{e_t^{+,(q)} + e_t^{-,(q)}}{2}$. The local envelope mean is then removed by $h_t^{(q+1)} = h_t^{(q)} - \mu_t^{(q)}$. Once $h_t^{(q)}$ satisfies the IMF conditions, it is recorded as $c_{j,t}$ and removed from the current signal. Repeating this process for J modes gives

$$x_t = \sum_{j=1}^{J} c_{j,t} + \rho_t^{(J)},$$

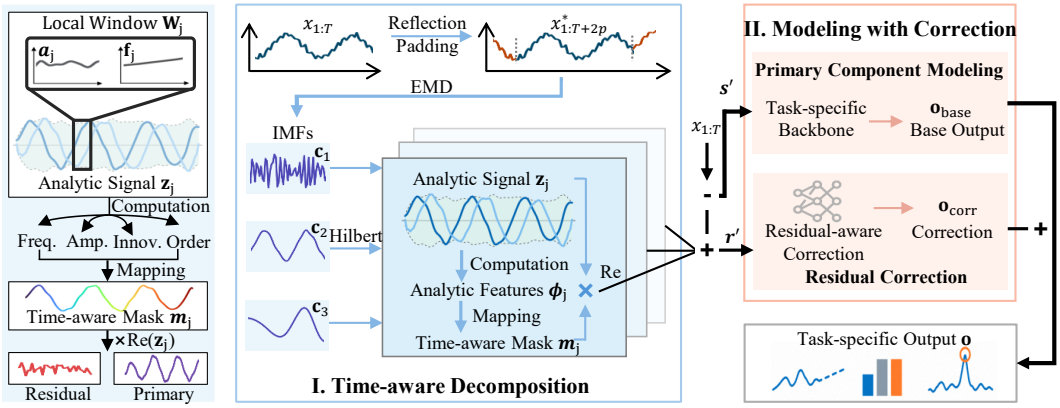


Figure 3: **Overview of DRAFTS. Right:** Time-aware decomposition splits $x_{1:T}$ into the primary component $\mathbf{s}'$ and residual $\mathbf{r}'$. The backbone and *Residual-Aware Correction* (RC) produce the base output and residual correction, respectively, which are summed to obtain $\mathbf{o}$. **Left:** Enlarged view of the blue shaded box showing how analytic features are mapped to time-aware masks for each IMF.

where $\rho_t^{(J)}$ is the component remaining after the first J IMFs have been extracted. Generally, earlier IMFs capture higher-frequency oscillations, while later IMFs capture lower-frequency ones.

For each IMF $\mathbf{c}_j \in \mathbb{R}^T$, the Hilbert transform uniquely specifies its analytic signal $z_{j,t} = c_{j,t} + \mathrm{i}\mathcal{H}[\mathbf{c}_j]_t = a_{j,t}e^{\mathrm{i}\theta_{j,t}}$, whose instantaneous amplitude, phase, and frequency are defined as

$$a_{j,t} = |z_{j,t}|, \quad \theta_{j,t} = \mathrm{unwrap}\big(\arg z_{j,t}\big), \quad f_{j,t} = \frac{\theta_{j,t} - \theta_{j,t-1}}{2\pi\Delta t}. \tag{1}$$

Here, unwrap removes 2π phase jumps along the time axis. Together, EMD and the Hilbert transform provide a physically interpretable representation of each oscillatory mode without assuming linearity or stationarity. We therefore use them as the basis for the following decomposition.

EMD for padded sequence. EMD is sensitive to boundary effects because no extrema are available beyond the observed window. This missing information can distort the interpolated envelopes and the resulting IMFs. Therefore, to obtain reliable IMFs, we apply reflection padding for the time series. Specifically, we extend both ends of the time series by reflecting the nearest p observations:

$$x^*_{1:T+2p} = [x_{p+1}, \ldots, x_2,\ x_1, \ldots, x_T,\ x_{T-1}, \ldots, x_{T-p}].$$

The reflected samples provide symmetric support beyond each boundary. Consequently, the envelopes near the original endpoints are estimated from extrema on both sides rather than through one-sided extrapolation, reducing boundary distortion in the resulting IMFs. We empirically set the padding length to one quarter of the input length, i.e., $p = 0.25T$.

We apply EMD to the padded sequence $x^*_{1:T+2p}$. Then we discard the padding and use the central T timesteps of each IMF, denoted by $c_{j,t}$, $t = 1, \ldots, T$, in subsequent processing. Previous application studies have observed noise mainly in the first two or three IMFs (Hassan, 2005; Quilfen & Chapron, 2021; Quilfen et al., 2022). Therefore, we examine the first $J \in \{2, 3\}$ IMFs for potential noise. The following local analysis determines their primary component and the residual at each timestep.

Analytic representation and feature setup. For each IMF $\mathbf{c}_j$, $j = 1, \ldots, J$, the Hilbert transform yields its analytic signal. We define the local window over which the features are computed and an amplitude-based weight that controls the contribution of each transition:

- *Local window $\mathcal{W}_{j,\tau}$:* Let $g_{j,u} = |\nabla_u \theta_{j,u}|$ denote the absolute phase gradient, where $u = 1, \ldots, T$ indexes samples and ∇_u uses unit sample spacing. We estimate the oscillation rate of $\mathbf{c}_j$ as $\tilde{f}_j = \mathrm{median}_{u:g_{j,u}>\epsilon}(g_{j,u})/(2\pi)$ in cycles/sample and set $L_j = \mathrm{round}(1.5/\tilde{f}_j)$. At timestep τ, we define the local window $\mathcal{W}_{j,\tau} = [\tau - L_j/2, \tau + L_j/2]$ and compute the subsequent statistical features within it. This design balances statistical stability and temporal locality.

- *Amplitude-based weight $w_{j,\tau}$:* Because phase is sensitive to perturbations at low amplitude, we weight the transition from $\tau-1$ to τ by adjacent amplitudes $a_{j,\tau}a_{j,\tau-1}$:

$$w_{j,\tau} = \min\left\{1, \frac{a_{j,\tau}a_{j,\tau-1}}{\left(\text{median}_{u=1,\dots,T}(a_{j,u})\right)^2}\right\}. \tag{2}$$

The squared median normalizes each IMF's weights by its amplitude scale; clipping at 1 prevents high-amplitude amplification. Low-amplitude transitions thus receive smaller weights.

Four analytic features. We compute four analytic features to guide the time-aware mask. The first three characterize local temporal variation and are illustrated in Figure 4.

1. **Frequency variation ($F_{j,t}$):** Measures the magnitude of instantaneous frequency changes:

$$F_{j,t} = \frac{\sqrt{\sum_{\tau\in\mathcal{W}_{j,t}} \bar{w}_{j,t,\tau}(\Delta f_{j,\tau})^2}}{\sum_{\tau\in\mathcal{W}_{j,t}} \bar{w}_{j,t,\tau}|f_{j,\tau}|}, \qquad \Delta f_{j,\tau} = f_{j,\tau} - f_{j,\tau-1}, \tag{3}$$

where $\bar{w}_{j,t,\tau} = \frac{w_{j,\tau}}{\sum_{u\in\mathcal{W}_{j,t}} w_{j,u}}$ is the normalized weight for $\tau \in \mathcal{W}_{j,t}$. The numerator emphasizes large frequency changes through squared increments. As a second difference of phase, $\Delta f_{j,\tau}$ is sensitive to perturbations, so we use the local weighted mean of $|f_j|$ for stable normalization.

2. **Amplitude variation ($A_{j,t}$):** Measures the magnitude of instantaneous amplitude changes:

$$A_{j,t} = \sqrt{\sum_{\tau\in\mathcal{W}_{j,t}} \bar{w}_{j,t,\tau}(\Delta a_{j,\tau})^2}, \qquad \Delta a_{j,\tau} = \frac{2|a_{j,\tau} - a_{j,\tau-1}|}{a_{j,\tau} + a_{j,\tau-1}}, \tag{4}$$

Similar to $F_{j,t}$, the numerator emphasizes large amplitude changes. Compared with $F_{j,t}$, $A_{j,t}$ directly measures the first-order change between adjacent amplitudes. Therefore, the two non-negative amplitudes provide a symmetric local reference for normalizing this change.

3. **Analytic innovation ($I_{j,t}$):** Measures the instantaneous evolution that cannot be captured by a local transition coefficient. We estimate this coefficient by minimizing the weighted one-step approximation error, whose closed-form solution is

$$\alpha_{j,t} = \frac{\sum_{\tau\in\mathcal{W}_{j,t}} \bar{w}_{j,t,\tau} z_{j,\tau} z^*_{j,\tau-1}}{\sum_{\tau\in\mathcal{W}_{j,t}} \bar{w}_{j,t,\tau}|z_{j,\tau-1}|^2}. \tag{5}$$

Writing $\alpha_{j,t} = \rho_{j,t}e^{\mathrm{i}\psi_{j,t}}$, we have $\alpha_{j,t}z_{j,\tau-1} = \rho_{j,t}a_{j,\tau-1}e^{\mathrm{i}(\theta_{j,\tau-1}+\psi_{j,t})}$. This expression estimates $z_{j,\tau}$ by scaling the previous amplitude by $\rho_{j,t}$ and advancing the previous phase by $\psi_{j,t}$. We then define the analytic innovation as the normalized error remaining after applying $\alpha_{j,t}$:

$$I_{j,t} = \frac{\sum_{\tau\in\mathcal{W}_{j,t}} \bar{w}_{j,t,\tau}|z_{j,\tau} - \alpha_{j,t}z_{j,\tau-1}|^2}{\sum_{\tau\in\mathcal{W}_{j,t}} \bar{w}_{j,t,\tau}|z_{j,\tau}|^2}. \tag{6}$$

It measures the instantaneous evolution remaining unexplained after estimating $z_{j,\tau}$ from $z_{j,\tau-1}$ using $\alpha_{j,t}$. A small $I_{j,t}$ indicates consistent evolution across adjacent timesteps.

4. **IMF order (O_j):** Encodes the normalized order index of the j-th IMF among J IMFs:

$$O_j = 1 - \frac{j-1}{\max(J-1, 1)}, \qquad j = 1, \dots, J. \tag{7}$$

O_j lets the model distinguish IMFs by their high-to-low frequency order.

Time-aware mask and decomposition. We concatenate the four features as $\boldsymbol{\phi}_{j,t} = [F_{j,t}, A_{j,t}, I_{j,t}, O_j]^\top \in \mathbb{R}^4$ and map them to a time-aware mask:

$$m_{j,t} = m_{\max}\sigma\left((\mathbf{v}^\top\boldsymbol{\phi}_{j,t} + b)\right), \tag{8}$$

where $\mathbf{v} \in \mathbb{R}^4$ and $b \in \mathbb{R}$ are learnable, σ is the sigmoid function, and $m_{\max}$ sets the maximum fraction assigned to the residual. Each candidate IMF contributes $m_{j,t}c_{j,t}$ to the residual, yielding

$$r'_t = \sum_{j=1}^{J} m_{j,t}c_{j,t}, \qquad s'_t = x_t - r'_t, \tag{9}$$

where s'_t is intended to capture underlying dynamics and r'_t is the residual. Applying this decomposition independently to all N variables yields $\mathbf{R}', \mathbf{S}' \in \mathbb{R}^{T\times N}$.

3.3 Stage II: Modeling with Correction

Residual-Aware Correction (RC) with a fixed reservoir. A recurrent reservoir with input weights $\mathbf{W}_{\text{in}}$ and recurrent weights $\mathbf{W}_{\text{res}}$ maps sequence $\mathbf{u}_{1:T}$ to output $\mathbf{y}$ via recurrent states $\mathbf{h}_t$:

$$\mathbf{h}_t = \tanh\left(\mathbf{W}_{\text{in}}\mathbf{u}_t + \mathbf{W}_{\text{res}}\mathbf{h}_{t-1}\right), \qquad \mathbf{y} = \mathbf{W}_{\text{out}}\mathbf{h}_T + \mathbf{b}_{\text{out}}. \tag{10}$$

Prior work shows that the reservoir can remain fixed (Jaeger, 2001; Maass et al., 2002). With suitably chosen heterogeneous initializations, the reservoir maps input history into diverse nonlinear states, providing an expressive temporal representation. Only the output layer $[\mathbf{W}_{\text{out}}, \mathbf{b}_{\text{out}}]$ is trained.

The residual $\mathbf{R}'$ may contain limited underlying dynamics that require expressive nonlinear temporal modeling. However, using a heavily parameterized model to capture these limited dynamics would be inefficient. We therefore build our *Residual-Aware Correction (RC)* module on the fixed recurrent reservoir with a task-specific output layer. Specifically, we use the residual as the reservoir input, $\mathbf{u}_t \leftarrow \mathbf{R}'_{t,:}$, and use the output as the task-specific correction, i.e., $\mathbf{y} \leftarrow \mathbf{o}_{\text{corr}}$. Eventually, we obtain expressive temporal representations with few trainable parameters.

Backbone modeling with residual correction. As shown in Figure 3, the task-specific backbone maps the primary component $\mathbf{S}'$ to the base output $\mathbf{o}_{\text{base}}$, while RC maps the residual $\mathbf{R}'$ to a correction $\mathbf{o}_{\text{corr}}$. The two outputs are then combined to produce the final task output:

$$\mathbf{o} = \mathbf{o}_{\text{base}} + \mathbf{o}_{\text{corr}}. \tag{11}$$

For forecasting, $\mathbf{o} \in \mathbb{R}^{S\times N}$ predicts the future sequence $\mathbf{Y}$. For classification, $\mathbf{o} \in \mathbb{R}^{C}$ gives class logits. Applying softmax yields class probabilities. For anomaly detection, $\mathbf{o} \in \mathbb{R}^{N}$ gives next-step prediction whose error defines the anomaly score.

Loss functions. The overall training objective of DRAFTS is:

$$\begin{aligned}\mathcal{L} &= \mathcal{L}_{\text{base}} + \lambda_1\mathcal{L}_{\text{corr}} + \lambda_2\mathcal{L}_{\text{fused}} \\ &= \ell_{\text{task}}(\mathbf{o}_{\text{base}}, \mathbf{Y}) + \lambda_1\ell_{\text{task}}\left(\text{detach}(\mathbf{o}_{\text{base}}) + \mathbf{o}_{\text{corr}}, \mathbf{Y}\right) + \lambda_2\ell_{\text{task}}\left(\mathbf{o}_{\text{base}} + \mathbf{o}_{\text{corr}}, \mathbf{Y}\right).\end{aligned} \tag{12}$$

Short- and long-term forecasting use sMAPE and MAE/MSE as ℓ_{task}, respectively; classification and anomaly detection use cross-entropy and one-step MSE. $\mathcal{L}_{\text{base}}$ trains the backbone and mask via $\mathbf{S}'$, $\mathcal{L}_{\text{corr}}$ trains RC on the residual with $\mathbf{o}_{\text{base}}$ detached, and $\mathcal{L}_{\text{fused}}$ supervises the fused output.

4 Experiments

To verify the effectiveness of DRAFTS, we perform extensive experiments across four mainstream modeling tasks, including short- and long-term forecasting, classification and anomaly detection.

4.1 Experimental Settings

Table 1: **Summary of the experiment datasets.** Further details are provided in Appendix A.1.

Tasks	Datasets	Metrics	Series Length
Forecasting	**Long-term:** ETTh1, ETTm1, Weather, Electricity, ILI, ECG, Laser, Noisy Lorenz63 (SNR={3,5,7})	MSE, MAE	$16 \sim 720$
	Short-term: M4 (6 subsets)	sMAPE, MASE, OWA	$6 \sim 48$
Classification	UEA (10 subsets)	Accuracy	$29 \sim 1152$
Anomaly Detection	SMD, MSL, SMAP, SWaT, PSM	Precision, Recall, F1-Score	100

Datasets. Table 1 summarizes the datasets and the settings. For forecasting, we consider both real-world benchmarks and a synthetic dataset. The synthetic dataset, Noisy Lorenz63, contains evolving underlying dynamics with noise to assess the ability of DRAFTS to model time series in which they coexist. We corrupt the input histories but use the corresponding clean Lorenz63 future as the prediction target. Specifically, we use Lorenz63 trajectories (Lorenz, 1963) as the underlying dynamics and add three noise corruptions: bursts, which represent abruptly emerging high-frequency oscillations; spikes, which represent sharp changes at isolated time points; and missing blocks to simulate periods with no recorded observations, as shown in Figure 4. We vary the signal-to-noise ratio (SNR) over $\{3, 5, 7\}$ dB to evaluate DRAFTS under different noise levels.

Table 2: **Long- and short-term forecasting results.** Results are averaged over horizons per dataset, with weighted averages over six M4 subsets. Gray cells indicate +DRAFTS, **Bold red** marks the better result within each raw/augmented pair. Full results are in Table 13 and Table 14.

Dataset	Metric	Crossformer (2023)		PatchTST (2023)		iTransformer (2023)		TimeMixer (2024a)		TimeFilter (2025)		L-Drive (2026)	
		Raw	+DRAFTS	Raw	+DRAFTS	Raw	+DRAFTS	Raw	+DRAFTS	Raw	+DRAFTS	Raw	+DRAFTS
ETTh1	MSE	0.449	**0.443** (+1.34%)	0.459	**0.457** (+0.44%)	0.448	**0.440** (+1.79%)	0.458	**0.453** (+1.09%)	0.488	**0.475** (+2.66%)	0.472	**0.465** (+1.48%)
	MAE	0.445	**0.441** (+0.90%)	0.432	**0.431** (+0.23%)	0.431	**0.431** (+0.00%)	0.429	**0.428** (+0.23%)	0.457	**0.453** (+0.88%)	0.447	**0.443** (+0.89%)
ETTm1	MSE	0.413	**0.401** (+2.91%)	0.396	**0.392** (+1.01%)	0.400	**0.394** (+1.50%)	0.393	**0.385** (+2.04%)	**0.385**	0.386 (-0.26%)	0.383	**0.376** (+1.83%)
	MAE	0.404	**0.400** (+0.99%)	0.387	**0.386** (+0.26%)	0.390	**0.389** (+0.26%)	0.385	**0.383** (+0.52%)	**0.385**	0.386 (-0.26%)	0.383	**0.382** (+0.26%)
Weather	MSE	0.241	**0.239** (+0.83%)	0.251	**0.247** (+1.59%)	0.268	**0.263** (+1.87%)	0.283	**0.241** (+14.84%)	0.238	**0.235** (+1.26%)	0.237	**0.234** (+1.27%)
	MAE	0.268	**0.267** (+0.37%)	0.270	**0.268** (+0.74%)	0.280	**0.277** (+1.07%)	0.302	**0.270** (+10.60%)	0.261	**0.260** (+0.38%)	0.258	**0.258** (+0.00%)
Electricity	MSE	0.213	**0.205** (+3.76%)	0.216	**0.213** (+1.39%)	0.216	**0.212** (+1.85%)	0.212	**0.202** (+4.72%)	0.216	**0.208** (+3.70%)	0.181	**0.179** (+1.10%)
	MAE	0.290	**0.283** (+2.41%)	0.291	**0.288** (+1.03%)	0.290	**0.287** (+1.03%)	0.282	**0.279** (+1.06%)	0.292	**0.286** (+2.05%)	0.270	**0.268** (+0.74%)
ILI	MSE	5.035	**4.509** (+10.45%)	3.322	**3.092** (+6.92%)	3.163	**3.010** (+4.84%)	3.182	**2.885** (+9.33%)	2.969	**2.499** (+15.83%)	3.110	**2.865** (+7.88%)
	MAE	1.542	**1.445** (+6.29%)	1.110	**1.094** (+1.44%)	1.069	**1.035** (+3.18%)	1.149	**1.107** (+3.66%)	1.033	**0.985** (+4.65%)	0.969	**0.945** (+2.48%)
ECG	MSE	0.771	**0.678** (+12.06%)	0.577	**0.546** (+5.37%)	0.654	**0.620** (+5.20%)	0.633	**0.620** (+2.05%)	0.579	**0.537** (+7.25%)	0.664	**0.593** (+10.69%)
	MAE	0.526	**0.506** (+3.80%)	0.438	**0.411** (+6.16%)	0.463	**0.456** (+1.51%)	0.469	**0.455** (+2.99%)	0.461	**0.416** (+9.76%)	0.514	**0.462** (+10.12%)
Laser	MSE	0.135	**0.117** (+13.33%)	0.197	**0.173** (+12.18%)	0.209	**0.182** (+12.92%)	0.232	**0.201** (+13.36%)	0.183	**0.149** (+18.58%)	0.156	**0.148** (+5.13%)
	MAE	0.166	**0.146** (+12.05%)	0.231	**0.211** (+8.66%)	0.218	**0.203** (+6.88%)	0.246	**0.230** (+6.50%)	0.198	**0.179** (+9.60%)	0.173	**0.165** (+4.62%)
Lorenz63 (SNR=3)	MSE	0.690	**0.660** (+4.35%)	0.850	**0.806** (+5.18%)	0.815	**0.791** (+2.94%)	0.837	**0.832** (+0.60%)	0.700	**0.674** (+3.71%)	0.726	**0.721** (+0.69%)
	MAE	0.497	**0.469** (+5.63%)	0.621	**0.589** (+5.15%)	0.593	**0.575** (+3.04%)	0.623	**0.620** (+0.48%)	0.506	**0.490** (+3.16%)	0.511	**0.510** (+0.20%)
Lorenz63 (SNR=5)	MSE	0.631	**0.581** (+7.92%)	0.769	**0.752** (+2.21%)	0.736	**0.720** (+2.17%)	0.772	**0.763** (+1.17%)	0.648	**0.595** (+8.18%)	0.663	**0.656** (+1.06%)
	MAE	0.491	**0.450** (+8.35%)	0.600	**0.587** (+2.17%)	0.572	**0.558** (+2.45%)	0.610	**0.605** (+0.82%)	0.513	**0.470** (+8.38%)	0.497	**0.495** (+0.40%)
Lorenz63 (SNR=7)	MSE	0.559	**0.517** (+7.51%)	0.714	**0.686** (+3.92%)	0.673	**0.655** (+2.67%)	0.717	**0.709** (+1.12%)	0.605	**0.522** (+13.72%)	0.588	**0.583** (+0.85%)
	MAE	0.462	**0.425** (+8.01%)	0.585	**0.563** (+3.76%)	0.549	**0.536** (+2.37%)	0.591	**0.588** (+0.51%)	0.509	**0.438** (+13.95%)	0.474	**0.472** (+0.42%)
M4	sMAPE	13.479	**13.231** (+1.84%)	12.402	**12.349** (+0.43%)	12.554	**12.513** (+0.33%)	13.647	**13.413** (+1.72%)	13.096	**12.845** (+1.91%)	12.893	**12.716** (+1.37%)
	MASE	1.864	**1.822** (+2.21%)	1.676	**1.662** (+0.84%)	1.680	**1.676** (+0.24%)	1.873	**1.828** (+2.41%)	1.729	**1.695** (+1.95%)	1.718	**1.682** (+2.11%)
	OWA	0.984	**0.964** (+2.02%)	0.895	**0.890** (+0.63%)	0.902	**0.900** (+0.28%)	0.993	**0.973** (+2.06%)	0.935	**0.917** (+1.93%)	0.925	**0.909** (+1.73%)

Baselines and implementation. We integrate DRAFTS into six representative SOTA backbones: (1) Transformer-based: Crossformer (Zhang & Yan, 2023), iTransformer (Liu et al., 2023), PatchTST (Nie et al., 2023); (2) MLP-based: TimeMixer (Wang et al., 2024a); (3) Graph-based: TimeFilter (Hu et al., 2025); (4) hybrid architecture: L-Drive (Zhang et al., 2026), combining CNN, GRU, and MLP modules. To ensure fair evaluation, when applying DRAFTS to baseline models, we strictly follow their original hyperparameter settings and only tune the hyperparameters introduced by DRAFTS (i.e., $\lambda_1, \lambda_2 \in [0, 1]$, $J \in \{2, 3\}$ and $m_{\max} \in (0, 1)$). Table 2 uses one fixed seed; two settings are additionally evaluated with three seeds (Appendix B.1). For each classification and anomaly detection experiment, we report means over three independent runs with different seeds. All experiments are conducted using 8 NVIDIA A800 80GB GPUs. See Appendix A for details.

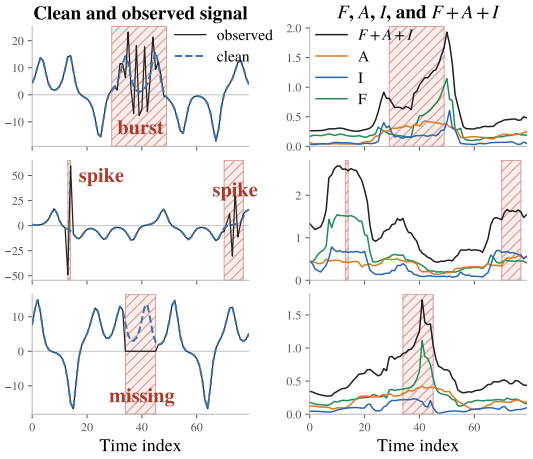


Figure 4: **Samples from Noisy Lorenz63 (SNR=3).** Clean and observed data with abrupt events are shown alongside analytic features derived from the first IMF of the observed data.

4.2 MAIN RESULTS

Long- and Short-term Forecasting Results. DRAFTS generally improves all six backbones in both forecasting settings (Table 2). For long-term forecasting, DRAFTS reduces TimeMixer's MSE by 14.84% on Weather, showing that a decomposition-based backbone also benefits from our framework. For short-term forecasting on M4, DRAFTS improves all three metrics across all six backbones, reducing Crossformer's MASE and OWA by 2.21% and 2.02%, respectively.

Classification and anomaly detection results. Figure 5 shows that DRAFTS improves all six backbones on both tasks. For classification, DRAFTS raises TimeMixer's accuracy from 45.41% to 53.12%, while iTransformer reaches 74.93%. For anomaly detection, TimeFilter achieves the largest gain in average F1: 1.12 percentage points, from 82.73% to 83.85%. These results support the broad applicability of DRAFTS across backbone architectures and time series tasks.

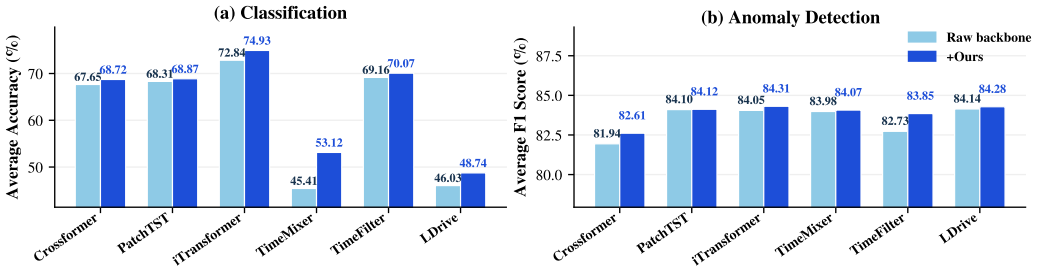


Figure 5: **Classification and anomaly detection results.** Full results see Table 15 and Table 16.

4.3 Ablations and Analysis

Effectiveness of RC. We compare four residual variants: (1) No Correction, (2) RC (ours), (3) Flatten Linear, a direct linear residual mapping, and (4) Trainable RC. For each case, we freeze the same backbone trained with the primary component and train only the residual head. Table 3 shows that RC outperforms all alternatives. Flatten Linear's weaker results suggest a potential benefit of modeling temporal dependencies in residual correction. RC does so with only 12.48K trainable parameters.

Table 3: **Residual correction on Crossformer.** Noisy Lorenz63 (SNR= 7 dB, $S = 64$); mean(std) over three head seeds. Trainable/Total: trainable/stored parameters.

Residual Module	MSE	MAE	Trainable	Total
No Correction	0.815	0.655	–	–
RC (ours)	**0.805**(0.001)	**0.653**(<0.001)	**12.48K**	**16.77K**
Flatten Linear	0.815(<0.001)	0.661(<0.001)	55.49K	55.49K
Trainable RC	0.809(0.009)	0.656(0.006)	16.77K	**16.77K**

Decomposition comparison. We isolate decomposition effects using the same TimeFilter backbone and RC, comparing time-adaptive decomposition with FFT high-frequency and low-amplitude truncation (FFT-high/low), DCT and STFT high-frequency truncation (DCT-high, STFT-high), and a sample-adaptive FFT mask (FFT-adap). Table 4 reports test forecasting errors over the shared input-energy reduction range [1.05%, 9.20%]. The strongest spectral baseline, FFT-high, performs best at only 1.28% reduction, suggesting stronger global filtering removes useful patterns. DRAFTS remains accurate at higher reductions, suggesting it better preserves forecast-relevant dynamics.

Table 4: **Decomposition Comparison.** TimeFilter test errors on noisy Lorenz63. AUC is normalized over the common energy-reduction range; parentheses give the reduction at minimum error. See Appendix A.3 for the full protocol.

Method	AUC		Minimum Error	
	MSE	MAE	MSE	MAE
Ours (DRAFTS)	**0.340**	**0.309**	**0.319** (5.49%)	**0.291** (2.26%)
FFT-high	0.352	0.321	0.322 (1.28%)	0.293 (1.28%)
FFT-low	0.403	0.358	0.347 (1.38%)	0.310 (1.38%)
DCT-high	0.362	0.331	0.344 (4.01%)	0.310 (4.01%)
STFT-high	0.363	0.331	0.333 (6.30%)	0.301 (6.30%)
FFT-adap	0.371	0.335	0.366 (3.30%)	0.331 (3.30%)

Ablation study. We compare raw TimeFilter with a matched raw-input RC control and two analytic-IMF variants, using direct attributes or ϕ, each with/without RC. Table 5 shows that neither analytic processing nor RC alone is uniformly effective: no-RC variants underperform Raw at SNR $= 3$ and SNR $= 5$, while Raw + RC helps only at SNR $= 5$ and SNR $= 7$. Yet RC improves both analytic variants relative to their no-RC counterparts at every SNR, and full DRAFTS consistently outperforms Raw + RC. Thus, the results support the complementarity of decomposition and RC and suggest that the gains do not come solely from adding RC to the raw backbone.

Table 5: **Ablation results on Noisy Lorenz63.** Each entry reports MSE / MAE averaged over $S = 16, 64$. Analyt. denotes the analytic IMFs. Full results see Appendix B.5.

Dataset	Raw	Raw + RC	Analyt. + [Amp., Freq., O]	Analyt. + ϕ	Analyt. + [Amp., Freq., O] + RC	DRAFTS (Ours)
SNR=3	0.699 / 0.505	0.713 / 0.527	0.713 / 0.522	0.716 / 0.520	0.702 / 0.511	**0.674 / 0.489**
SNR=5	0.648 / 0.513	0.636 / 0.502	0.663 / 0.522	0.667 / 0.526	0.602 / 0.474	**0.595 / 0.470**
SNR=7	0.604 / 0.509	0.561 / 0.470	0.600 / 0.500	0.548 / 0.464	0.529 / 0.445	**0.522 / 0.438**

Efficiency Analysis. Table 6 compares TimeFilter with and without DRAFTS on ILI (S=24). DRAFTS reduces MSE and MAE by 27.62% and 7.30%, respectively, while increasing trainable parameters, peak allocated GPU memory, and model-side FLOPs by only 0.23%, 0.66%, and 0.12%. Under the early-stopping protocol, it requires 22.81% fewer training iterations. These results show that DRAFTS improves forecasting performance with modest additional model-side costs.

Table 6: **Efficiency comparison.** Analytic signals are precomputed offline.

Method	Params.	Peak Alloc. (MiB)	Train Iters.	MSE	MAE	Model FLOPs/Sample
Raw	4,836,628	197.24	570	3.848	1.055	0.21179 GFLOPs
+DRAFTS	4,847,553	198.54	440	2.785	0.978	0.21204 GFLOPs
Change vs. Raw	↑ 0.23%	↑ 0.66%	↓ 22.81%	↓ 27.62%	↓ 7.30%	↑ 0.12%

Qualitative Analysis. As Figure 6 shows, under Noisy Lorenz63 (SNR=7 dB and $S = 64$), the backbone often produces overly smooth trajectories that stay near the mean or show attenuated amplitudes. In contrast, DRAFTS effectively exploits the historical context and retains the nonlinear dynamics of the underlying chaotic system. For example, DRAFTS better captures the sharp transition at the forecasting boundary and the subsequent recovery in Figure 6 (a,b). These results show that DRAFTS better captures temporal variation and oscillatory patterns under mixed noise, alleviating the excessive smoothing observed in the backbone.

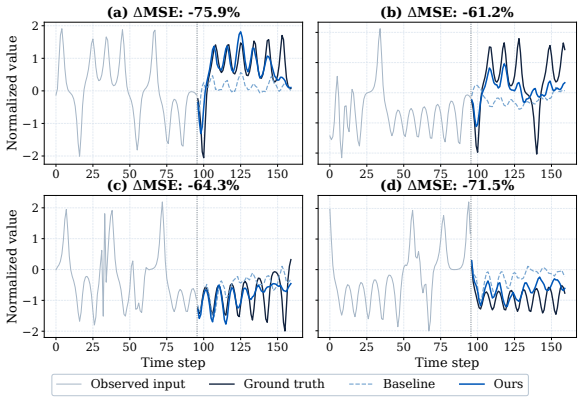


Figure 6: **Qualitative Visualization.** DRAFTS (blue) and backbone (dashed light blue) predictions, with ground truth (dark navy) and observed inputs (gray).

Decomposition Adaptivity Analysis. Using TimeFilter as the backbone, Table 7 reports results averaged over four prediction lengths for each dataset. Appendix A.2 defines residual magnitude, amplitude variance, and temporal-variation alignment. DRAFTS produces residuals on ILI that are larger, more variable, and more strongly correlated with local variation. Together, these results suggest regime-dependent decomposition profiles: residual extraction remains conservative on the regularly periodic ETTm1 series but becomes stronger and more variable on the temporally heterogeneous ILI series.

Table 7: **Decomposition adaptivity.** Temporal-variation alignment is the Pearson correlation between residual energy and local temporal variation.

Dataset	Residual Magnitude (%)	Residual Variance	Temporal-Variation Alignment
ETTm1	1.74	0.292	0.240
ILI	3.70	0.370	0.484

Hyperparameter Sensitivity. Figure 7 shows a post-hoc sensitivity analysis of Crossformer ($S = 64$), with final settings selected on validation data. We omit $J \geq 6$, as many samples lack enough valid IMFs. $J = 1$ and $J = 2$ perform similarly at SNR $= 3$, whereas $J = 2$ and $J = 3$ are optimal at SNR $= 5$ and 7, respectively. Larger J generally raises MSE, consistent with our use of $J \in \{2, 3\}$. For RC, size 64 gives the largest mean relative MSE drop across tested SNRs, though the optimum varies by SNR.

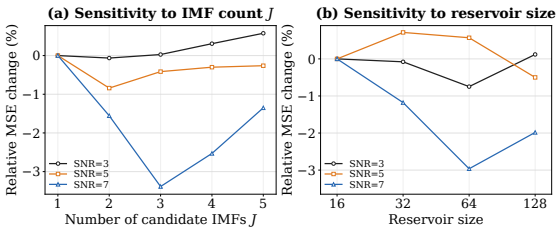


Figure 7: **Hyperparameter sensitivity.** Relative MSE changes use $J = 1$ in (a) and reservoir size 16 in (b) as references within each SNR.

5 CONCLUSION AND FUTURE WORK

This work addresses the coexistence of evolving dynamics and noise through *time-aware decomposition with correction*. Specifically, DRAFTS uses instantaneous-frequency features to extract a primary component that captures evolving dynamics, while residual correction exploits useful dynamics missed by decomposition. Improvements across four time series modeling tasks and six backbones support this approach. Future work will explore online multivariate decomposition to capture evolving cross-variable dynamics without offline preprocessing.

AI USE STATEMENT

We used generative AI tools to assist with manuscript writing and polishing, including checks of mathematical notation, dimensions, units, and formula definitions in the main text and appendix. AI tools also helped identify and organize related literature and provided feedback on experimental methodology and the interpretation of results. We manually checked the mathematical expressions and verified cited claims against the original papers. We take responsibility for the final content of this work, including all AI-assisted material.

REFERENCES

Ahmed Abdulaal, Zhuanghua Liu, and Tomer Lancewicki. Practical approach to asynchronous multivariate time series anomaly detection and localization. KDD '21, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450383325. doi: 10.1145/3447548.3467174. URL https://doi.org/10.1145/3447548.3467174.

N. Ahmed, T. Natarajan, and K.R. Rao. Discrete cosine transform. *IEEE Transactions on Computers*, C-23(1):90–93, 1974. doi: 10.1109/T-C.1974.223784.

Jont B Allen and Lawrence R Rabiner. A unified approach to short-time fourier analysis and synthesis. *Proceedings of the IEEE*, 65(11):1558–1564, 1977.

Abdul Fatir Ansari, Lorenzo Stella, Caner Turkmen, Xiyuan Zhang, Pedro Mercado, Huibin Shen, Oleksandr Shchur, Syama Syndar Rangapuram, Sebastian Pineda Arango, Shubham Kapoor, Jasper Zschiegner, Danielle C. Maddix, Michael W. Mahoney, Kari Torkkola, Andrew Gordon Wilson, Michael Bohlke-Schneider, and Yuyang Wang. Chronos: Learning the language of time series. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. URL https://openreview.net/forum?id=gerNCVqqtR.

A. Bagnall, Hoang Anh Dau, Jason Lines, Michael Flynn, James Large, Aaron George Bostrom, Paul Southam, and Eamonn J. Keogh. The uea multivariate time series classification archive, 2018. *ArXiv*, abs/1811.00075, 2018. URL https://api.semanticscholar.org/CorpusID:53166683.

Kaifeng Bi, Lingxi Xie, Hengheng Zhang, Xin Chen, Xiaotao Gu, and Qi Tian. Accurate medium-range global weather forecasting with 3d neural networks. *Nature*, 619:1–6, 07 2023. doi: 10.1038/s41586-023-06185-3.

B. Boashash. Estimating and interpreting the instantaneous frequency of a signal. i. fundamentals. *Proceedings of the IEEE*, 80(4):520–538, 1992. doi: 10.1109/5.135376.

Defu Cao, Wen Ye, Yizhou Zhang, Sam Griesemer, and Yan Liu. PINFDit: Energy-based physics-informed diffusion transformers for general-purpose time series tasks. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=EphTlUJ4XN.

Hao Cheng, Qingsong Wen, Yang Liu, and Liang Sun. Robusttsf: Towards theory and design of robust time series forecasting with anomalies. In *Proceedings of the 12th International Conference on Learning Representations*, pp. 1–28, 2024.

James W Cooley and John W Tukey. An algorithm for the machine calculation of complex fourier series. *Mathematics of computation*, 19(90):297–301, 1965.

Yisong Fu, Zezhi Shao, Chengqing Yu, Yujie Li, Zhulin An, Qi Wang, Yongjun Xu, and Fei Wang. Selective learning for deep time series forecasting. *Advances in neural information processing systems*, 2025.

Dennis Gabor. Theory of communication. part 1: The analysis of information. *Journal of the Institution of Electrical Engineers—Part III: Radio and Communication Engineering*, 93(26): 429–441, 1946.

Hassan H Hassan. Empirical mode decomposition (emd) of potential field data: airborne gravity data as an example. In *SEG International Exposition and Annual Meeting*, pp. SEG–2005. SEG, 2005.

Matthew Horak, Sowmya Chandrasekaran, and Giovanni Tobar. Nlp based anomaly detection for categorical time series. In *2022 IEEE 23rd International Conference on Information Reuse and Integration for Data Science (IRI)*, pp. 27–34, 2022. doi: 10.1109/IRI54793.2022.00019.

Yifan Hu, Guibin Zhang, Peiyuan Liu, Disen Lan, Naiqi Li, Dawei Cheng, Tao Dai, Shu-Tao Xia, and Shirui Pan. Timefilter: Patch-specific spatial-temporal graph filtration for time series forecasting. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=490VcNtjh7.

Norden E. Huang, Zheng Shen, Steven R. Long, Manli C. Wu, Hsing H. Shih, Quanan Zheng, Nai-Chyuan Yen, Chi Chao Tung, and Henry H. Liu. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 454(1971):903–995, 03 1998. ISSN 1364-5021. doi: 10.1098/rspa.1998.0193. URL https://doi.org/10.1098/rspa.1998.0193.

Kyle Hundman, Valentino Constantinou, Christopher Laporte, Ian Colwell, and Tom Soderstrom. Detecting spacecraft anomalies using lstms and nonparametric dynamic thresholding. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '18, pp. 387–395, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450355520. doi: 10.1145/3219819.3219845. URL https://doi.org/10.1145/3219819.3219845.

Herbert Jaeger. The "echo state" approach to analysing and training recurrent neural networks-with an erratum note. *Bonn, Germany: German national research center for information technology gmd technical report*, 148(34):13, 2001.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, 2015.

Dani Kiyasseh, Tingting Zhu, and David A Clifton. Clocs: Contrastive learning of cardiac signals across space, time, and patients. In *International Conference on Machine Learning*, pp. 5606–5615. PMLR, 2021.

Yannis Kopsinis and Stephen McLaughlin. Development of emd-based denoising methods inspired by wavelet thresholding. *IEEE Transactions on Signal Processing*, 57(4):1351–1362, April 2009. ISSN 1053-587X. doi: 10.1109/TSP.2009.2013885.

Guokun Lai, Wei-Cheng Chang, Yiming Yang, and Hanxiao Liu. Modeling long- and short-term temporal patterns with deep neural networks. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, SIGIR '18, pp. 95–104, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450356572. doi: 10.1145/3209978.3210006. URL https://doi.org/10.1145/3209978.3210006.

Yubo Liang, Zezhi Shao, Fei Wang, Zhao Zhang, Tao Sun, and Yongjun Xu. Basicts: An open source fair multivariate time series prediction benchmark. In *International symposium on benchmarking, measuring and optimization*, pp. 87–101. Springer, 2022.

Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. itransformer: Inverted transformers are effective for time series forecasting. *arXiv preprint arXiv:2310.06625*, 2023.

Yong Liu, Guo Qin, Xiangdong Huang, Jianmin Wang, and Mingsheng Long. Timer-XL: Long-context transformers for unified time series forecasting. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=KMCJXjlDDr.

Edward N. Lorenz. Deterministic Nonperiodic Flow. *Journal of the Atmospheric Sciences*, 20(2): 130–148, March 1963. doi: 10.1175/1520-0469(1963)020⟨0130:DNF⟩2.0.CO;2.

Tatiana S. Lugovaya. Biometric human identification based on electrocardiogram. Master's thesis, Faculty of Computing Technologies and Informatics, Saint-Petersburg Electrotechnical University "LETI", Saint Petersburg, Russian Federation, June 2005.

Wolfgang Maass, Thomas Natschläger, and Henry Markram. Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural Computation*, 14(11):2531–2560, 2002. doi: 10.1162/089976602760407955.

Aditya P. Mathur and Nils Ole Tippenhauer. Swat: a water treatment testbed for research and training on ics security. In *2016 International Workshop on Cyber-physical Systems for Smart Water Networks (CySWater)*, pp. 31–36, 2016. doi: 10.1109/CySWater.2016.7469060.

Yuqi Nie, Nam H. Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. 2023. URL https://arxiv.org/abs/2211.14730.

Tom J. Pollard, Benjamin Moody, Li wei Lehman, B. Gow, C. Fernandes, Chen Xie, Alistair E W Johnson, Roger G. Mark, and Thomas Heldt. Physionet as a global platform for biomedical research. *Nature Health*, 2026.

Yves Quilfen and Bertrand Chapron. On denoising satellite altimeter measurements for high-resolution geophysical signal analysis. *Advances in Space Research*, 68(2):875–891, 2021.

Yves Quilfen, Jean-François Piolle, and Bertrand Chapron. Towards improved analysis of short mesoscale sea level signals from satellite altimetry. *Earth System Science Data*, 14(4):1493–1512, 2022.

Xiaoming Shi, Shiyu Wang, Yuqi Nie, Dianqi Li, Zhou Ye, Qingsong Wen, and Ming Jin. Time-moe: Billion-scale time series foundation models with mixture of experts. In *International conference on learning representations*, volume 2025, pp. 34635–34667, 2025.

Ya Su, Youjian Zhao, Chenhao Niu, Rong Liu, Wei Sun, and Dan Pei. Robust anomaly detection for multivariate time series through stochastic recurrent neural network. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '19, pp. 2828–2837, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450362016. doi: 10.1145/3292500.3330672. URL https://doi.org/10.1145/3292500.3330672.

Shiyu Wang, Haixu Wu, Xiaoming Shi, Tengge Hu, Huakun Luo, Lintao Ma, James Y Zhang, and JUN ZHOU. Timemixer: Decomposable multiscale mixing for time series forecasting. In *International Conference on Learning Representations (ICLR)*, 2024a.

Yuxuan Wang, Haixu Wu, Jiaxiang Dong, Yong Liu, Mingsheng Long, and Jianmin Wang. Deep time series models: A comprehensive survey and benchmark. 2024b.

Zexin Wang, Changhua Pei, Minghua Ma, Xin Wang, Zhihan Li, Dan Pei, Saravan Rajmohan, Dongmei Zhang, Qingwei Lin, Haiming Zhang, Jianhui Li, and Gaogang Xie. Revisiting vae for unsupervised time series anomaly detection: A frequency perspective. In *Proceedings of the ACM Web Conference 2024*, WWW '24, pp. 3096–3105, New York, NY, USA, 2024c. Association for Computing Machinery. ISBN 9798400701719. doi: 10.1145/3589334.3645710. URL https://doi.org/10.1145/3589334.3645710.

A.S. Weigend and N.A. Gershenfeld. Results of the time series prediction competition at the santa fe institute. In *IEEE International Conference on Neural Networks*, pp. 1786–1793 vol.3, 1993. doi: 10.1109/ICNN.1993.298828.

Gerald Woo, Chenghao Liu, Akshat Kumar, Caiming Xiong, Silvio Savarese, and Doyen Sahoo. Unified training of universal time series forecasting transformers. In *Forty-first International Conference on Machine Learning*, 2024.

Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with Auto-Correlation for long-term series forecasting. In *Advances in Neural Information Processing Systems*, 2021.

Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. Timesnet: Temporal 2d-variation modeling for general time series analysis. In *International Conference on Learning Representations*, 2023.

Jiehui Xu, Haixu Wu, Jianmin Wang, and Mingsheng Long. Anomaly transformer: Time series anomaly detection with association discrepancy. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=LzQQ89U1qm_.

Zhijian Xu, Ailing Zeng, and Qiang Xu. Fits: Modeling time series with $10k$ parameters. In *International Conference on Learning Representations*, volume 2024, pp. 26295–26318, 2024.

Kun Yi, Jingru Fei, Qi Zhang, Hui He, Shufeng Hao, Defu Lian, and Wei Fan. Filternet: Harnessing frequency filters for time series forecasting. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 55115–55140. Curran Associates, Inc., 2024. doi: 10.52202/079017-1749. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/6323d96f79d5d49e0d3fc88835c082cd-Paper-Conference.pdf.

Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are transformers effective for time series forecasting? 2023.

Fan Zhang, Shijun Chen, and Hua Wang. L-drive: Beyond a single mapping—latent context drives time series forecasting. In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=N7hNbQ8o5Q.

Yunhao Zhang and Junchi Yan. Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting. In *International Conference on Learning Representations*, 2023.

Siru Zhong, Yiqiu Liu, Zhiqing Cui, Zezhi Shao, Fei Wang, Qingsong Wen, and Yuxuan Liang. DropoutTS: Sample-adaptive dropout for robust time series forecasting. In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=7sksHLUvhH.

Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI conference on Artificial Intelligence*, pp. 11106–11115, 2021.

Tian Zhou, Ziqing MA, xue wang, Qingsong Wen, Liang Sun, Tao Yao, Wotao Yin, and Rong Jin. Film: Frequency improved legendre memory model for long-term time series forecasting. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 12677–12690. Curran Associates, Inc., 2022a. doi: 10.52202/068431-0921. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/524ef58c2bd075775861234266e5e020-Paper-Conference.pdf.

Tian Zhou, Ziqing Ma, Qingsong Wen, Xue Wang, Liang Sun, and Rong Jin. FEDformer: Frequency enhanced decomposed transformer for long-term series forecasting. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato (eds.), *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 27268–27286. PMLR, 17–23 Jul 2022b. URL https://proceedings.mlr.press/v162/zhou22g.html.

A Experimental Details

A.1 Benchmark Details

A.1.1 Long-term Forecasting

As shown in Table 8, we evaluate the performance for long-term forecasting on seven real-world datasets, including ETT datasets (ETTh1, ETTm1), Weather, Electricity, ILI, ECG (Pollard et al., 2026; Lugovaya, 2005), Laser (Weigend & Gershenfeld, 1993) and a synthetic data Noisy Lorenz63.

Table 8: Long-term forecasting dataset descriptions. Sizes are reported as (train, validation, test).

Dataset	Dim	Input Length	Prediction Length	Dataset Size	Frequency / Information
ETTh1	7	96	$\{96, 192, 336, 720\}$	$(8640, 2880, 2880)$	Hourly / Temperature
ETTm1	7	96	$\{96, 192, 336, 720\}$	$(34560, 11520, 11520)$	15 min / Temperature
Electricity	321	96	$\{96, 192, 336, 720\}$	$(18412, 2630, 5262)$	Hourly / Electricity
Weather	21	96	$\{96, 192, 336, 720\}$	$(36887, 5269, 10540)$	10 min / Meteorology
ILI	7	24	$\{24, 36, 48, 60\}$	$(676, 96, 194)$	Weekly / Health
ECG	1	360	$\{144, 288\}$	$(7000, 1000, 2000)$	500 Hz / Health
Laser	1	100	$\{50, 100\}$	$(6055, 2018, 2020)$	Hourly / Physics
Noisy Lorenz63	3	96	$\{16, 64\}$	$(14000, 1999, 4001)$	$\Delta t = 0.075$ / Chaos

For the first five public forecasting benchmarks, we follow the standard train, validation, and test splits established in `BasicTS` (Liang et al., 2022). For the remaining datasets, we adopt dedicated forecasting protocols tailored to their signal characteristics, as described below:

- **ECG**: We use an input length of 360 samples (0.72 s) and prediction lengths of 144 and 288 samples (0.288 s and 0.576 s, respectively). This setting captures sufficient local context to capture cardiac waveform morphology while assessing short-term meaningful horizons.
- **Laser**: We use an input length of 100 to predict the next 50 and 100 observations, evaluating short- and long-horizon forecasting of nonlinear and chaotic dynamics of the laser intensity.
- **Noisy Lorenz63**: We use 96 input steps (a temporal span of 7.2) to predict 16 and 64 steps ahead (horizons of 1.2 and 4.8), evaluating both short-horizon forecasting and long-horizon forecasting under sensitive dependence on initial conditions and mixed noise.

All models are trained using Adam Kingma & Ba (2015) and a batch size of 64. Training runs for at most 100 epochs, with early stopping after 10 epochs without improvement in validation loss. Within each paired comparison, the raw and +DRAFTS variants share the same data splits, backbone configurations, optimization settings, checkpoint-selection rule, and random seed. Given the computational cost of the full benchmark, the main results use one fixed-seed run per configuration. Two selected configurations are additionally evaluated with three seeds, as reported below. Single-run comparisons do not show statistical significance or robustness across seeds; small relative gaps, especially those below 1%, warrant caution.

A.1.2 Short-term Forecasting

We use M4 for short-term forecasting, with dataset details provided in Table 9. M4 provides official training and test portions but no separate validation split. For each series, we reserve the final H observations of its official training portion for validation, where H is the forecasting horizon.

Following the M4 short-term forecasting protocol (Wu et al., 2023; Wang et al., 2024b), all models minimize sMAPE using Adam and a batch size of 32. Early stopping and checkpoint selection use validation sMAPE, with an early-stopping patience of 10. The raw and +DRAFTS variants share the same data splits, backbone configurations, optimization settings, and checkpoint-selection rule. All additional hyperparameters are selected on the validation split, and the official test set is used only for final evaluation.

Each M4 configuration is evaluated in a single run with seed 42. These matched single-run comparisons do not establish statistical significance or robustness across seeds. Small relative differences, particularly those below 1%, should therefore be interpreted cautiously.

Table 9: Short-term forecasting dataset descriptions. Sizes are reported as (train, test). M4 does not provide official validation splits; validation subsets are constructed exclusively from their official training splits as described in the experimental setup.

Dataset	Dim	Input Length	Prediction Length	Official Series (Train, Test)	Frequency / Information
M4-Yearly	1	12	6	(23,000, 23,000)	Yearly / Demographic
M4-Quarterly	1	16	8	(24,000, 24,000)	Quarterly / Finance
M4-Monthly	1	36	18	(48,000, 48,000)	Monthly / Industry
M4-Weekly	1	26	13	(359, 359)	Weekly / Macro
M4-Daily	1	28	14	(4,227, 4,227)	Daily / Micro
M4-Hourly	1	96	48	(414, 414)	Hourly / Other

A.1.3 Classification and Anomaly Detection

Classification. We evaluate classification on 10 multivariate datasets from the UEA Time Series Classification Archive (Bagnall et al., 2018), with details provided in Table 10. The archive provides official training and test splits but no separate validation split. We therefore construct a deterministic, class-stratified 80/20 training–validation split from the official training set using seed 42. This split is fixed across runs, and normalization statistics are estimated from the resulting training subset only.

PatchTST and iTransformer retain their original classification implementations. For the other four forecasting backbones without official classification implementations, we use a unified masked mean–max classification head. For each backbone, the raw and DRAFTS-augmented variants use the same head design. In the DRAFTS-augmented model, the backbone processes the primary component to produce base class logits, while RC maps the residual to a class-wise logit correction. The correction is added to the base logits before a single softmax produces the final class probabilities.

All models are trained using Adam, and a batch size of 64 for at most 100 epochs. Early stopping and checkpoint selection use validation accuracy, with an early-stopping patience of 10. All additional hyperparameters are selected on the validation split, and the official test set is used only for final evaluation after model and hyperparameter selection. Results are averaged over three training seeds: 42, 43, and 44.

Anomaly detection. We evaluate anomaly detection on five datasets: SMD (Su et al., 2019), SWaT (Mathur & Tippenhauer, 2016), PSM (Abdulaal et al., 2021), and MSL and SMAP (Hundman et al., 2018). Table 11 provides dataset details. For each backbone, the baseline and DRAFTS-augmented models share the same backbone configuration and data split. In the augmented model, the backbone processes the primary component to produce the base prediction, while RC maps the residual to a one-step correction. Their outputs are added to obtain the final prediction.

All models are trained on normal data by minimizing the mean squared forecasting error. We use Adam and a batch size of 64. Training runs for at most 100 epochs. Early stopping and checkpoint selection use validation MSE, with an early-stopping patience of 10. We repeat training with seeds 42, 43, and 44.

Table 10: UEA classification datasets. Validation sets are split from the official training sets.

Dataset	**Samples (Train, Test)**	**Variables**	**Series Length**
Handwriting	(150, 850)	3	152
LSST	(2459, 2466)	6	36
NATOPS	(180, 180)	24	51
RacketSports	(151, 152)	6	30
SelfRegulationSCP2	(200, 180)	7	1152
CharacterTrajectories	(1422, 1436)	3	60–182
SelfRegulationSCP1	(268, 293)	6	896
ArticularyWordRecognition	(275, 300)	9	144
JapaneseVowels	(270, 370)	12	7–29
UWaveGestureLibrary	(120, 320)	3	315

Table 11: Anomaly detection dataset descriptions. Sizes are reported as (train, val, test).

Dataset	Dataset sizes (train, val, test)	Variable Number	Sliding Window Length
SMD	(566724, 141681, 708420)	38	100
MSL	(44653, 11664, 73729)	55	100
SMAP	(108146, 27037, 427617)	25	100
SWaT	(396000, 99000, 449919)	51	100
PSM	(105984, 26497, 87841)	25	100

A.2 Decomposition Adaptivity Metrics

We characterize the decomposition output using residual magnitude, residual-amplitude variance, and temporal-variation alignment. Let $\mathbf{X}_{h,i}, \mathbf{R}'_{h,i} \in \mathbb{R}^{T\times C}$ denote the normalized input and its decomposition-derived residual for the i-th test window at prediction length h, respectively. Here, T and C denote the input length and number of variables, and N_h is the number of test windows. We use $\epsilon = 10^{-12}$ for numerical stability.

Residual magnitude. Residual magnitude measures the residual relative to the input:

$$M_h = \frac{100}{N_h}\sum_{i=1}^{N_h}\frac{\left\|\mathbf{R}'_{h,i}\right\|_F}{\left\|\mathbf{X}_{h,i}\right\|_F+\epsilon}. \tag{13}$$

Ignoring the numerical stabilizer, the Frobenius-norm ratio equals the RMS ratio because the input and residual have the same dimensions. This measure describes relative residual magnitude rather than an energy proportion and is therefore not restricted to $[0, 100]\%$.

Residual variance. We first calculate the residual amplitude at each timestep:

$$a_{h,i,t} = \sqrt{\frac{1}{C}\sum_{c=1}^{C}\left(R'_{h,i,t,c}\right)^2}, \qquad \bar{a}_{h,i} = \frac{1}{T}\sum_{t=1}^{T} a_{h,i,t}. \tag{14}$$

The normalized temporal variance of residual amplitude is defined as

$$V_h = \frac{1}{N_h T}\sum_{i=1}^{N_h}\sum_{t=1}^{T}\left(\frac{a_{h,i,t}-\bar{a}_{h,i}}{\bar{a}_{h,i}+\epsilon}\right)^2. \tag{15}$$

Thus, V_h measures how strongly the residual amplitude varies over time within each input window. A temporally constant residual amplitude, including an all-zero residual, contributes zero to V_h.

Temporal-variation alignment. We further measure whether residual extraction is associated with local input changes. Residual energy and local input variation are defined as

$$e_{h,i,t} = \frac{1}{C}\sum_{c=1}^{C}\left(R'_{h,i,t,c}\right)^2, \qquad d_{h,i,t} = \frac{1}{C}\sum_{c=1}^{C}\left(X_{h,i,t,c}-X_{h,i,t-1,c}\right)^2, \quad t=2,\dots,T. \tag{16}$$

Let $\mathcal{I}_h$ denote the set of test windows for which both $\{e_{h,i,t}\}_{t=2}^{T}$ and $\{d_{h,i,t}\}_{t=2}^{T}$ have nonzero temporal variance. For $|\mathcal{I}_h| > 0$, temporal-variation alignment is computed as

$$A_h = \frac{1}{|\mathcal{I}_h|}\sum_{i\in\mathcal{I}_h}\mathrm{corr}_{t=2,\dots,T}\left(e_{h,i,t}, d_{h,i,t}\right), \qquad |\mathcal{I}_h| > 0, \tag{17}$$

where corr denotes the Pearson correlation coefficient. A larger A_h indicates a more positive average temporal correlation between residual energy and local input variation. Since local input changes may contain both valid dynamics and noise, this measure describes their association rather than noise-identification accuracy. If $|\mathcal{I}_h| = 0$, A_h is treated as undefined rather than zero and is excluded from the aggregation.

Aggregation. Let $\mathcal{H}$ denote the four prediction lengths of a dataset. The reported residual magnitude and residual-amplitude variance are computed as

$$\bar{M} = \frac{1}{|\mathcal{H}|} \sum_{h \in \mathcal{H}} M_h, \qquad \bar{V} = \frac{1}{|\mathcal{H}|} \sum_{h \in \mathcal{H}} V_h. \tag{18}$$

For temporal-variation alignment, let

$$\mathcal{H}_A = \{h \in \mathcal{H} : |\mathcal{I}_h| > 0\}. \tag{19}$$

When $|\mathcal{H}_A| > 0$, its dataset-level value is

$$\bar{A} = \frac{1}{|\mathcal{H}_A|} \sum_{h \in \mathcal{H}_A} A_h. \tag{20}$$

If $\mathcal{H}_A$ is empty, $\bar{A}$ is undefined. No empty valid-window set occurred in the reported experiments.

We use $\mathcal{H} = \{96, 192, 336, 720\}$ for ETTm1 and $\mathcal{H} = \{24, 36, 48, 60\}$ for ILI. All metrics are first computed over test windows and then averaged equally across prediction lengths. All calculations are performed in the normalized model-input space using TimeFilter as the forecasting backbone.

A.3 Protocol for the Decomposition Comparison

Training and energy calibration. Table 4 evaluates TimeFilter on Noisy Lorenz63 (SNR=7) with input length 96 and prediction length 16. Each operating point is independently trained using seed 42. All variants share the data split, TimeFilter and RC architectures, loss weights, and optimization protocol. We use Adam, batch size 64, at most 100 epochs, and early stopping with patience 10. For every operating point, the checkpoint with the lowest validation MAE is retained.

For the validation windows $\{\mathbf{X}_i\}_{i=1}^{N_{\text{val}}}$ and their estimated removed components $\{\widehat{\mathbf{N}}_i\}_{i=1}^{N_{\text{val}}}$, we compute

$$r_{\text{val}} = \frac{\sum_i \|\mathbf{X}_i\|_F^2 - \sum_i \|\mathbf{X}_i - \widehat{\mathbf{N}}_i\|_F^2}{\max(\sum_i \|\mathbf{X}_i\|_F^2, 10^{-12})}. \tag{21}$$

Thus, r_{val} is a ratio of aggregated energies, rather than the mean of per-window ratios. The horizontal coordinates in Table 4 are measured on the validation set; the corresponding MSE and MAE are evaluated on the test set. No test-set energy measurement is used to construct the curves.

Common interval and nAUC. For method m, let r_{mj} denote its validation-calibrated operating points. We define the common interval as

$$a = \max\left(0.01, \max_m \min_j r_{mj}\right), \qquad b = \min\left(0.10, \min_m \max_j r_{mj}\right), \tag{22}$$

which gives $[a, b] = [1.05\%, 9.20\%]$. Therefore, every method has measured points spanning both boundaries. Points immediately outside $[a, b]$ may bracket a boundary during interpolation, but no extrapolation or endpoint extension is used.

If multiple configurations produce the same r_{val}, their test errors are averaged; the lowest test error is not selected. We then form a piecewise-linear test-error curve $\widehat{\mathcal{E}}_m(r)$. Its normalized area is computed exactly at the measured knots and interval boundaries:

$$\text{nAUC}_m = \frac{1}{b-a} \sum_{k=0}^{K-1} \frac{\widehat{\mathcal{E}}_m(s_k) + \widehat{\mathcal{E}}_m(s_{k+1})}{2} (s_{k+1} - s_k), \tag{23}$$

where $\{s_k\}$ contains a, b, and all measured points within the interval. Interpolation introduces no additional experimental observations.

The curve minimum is $\min_{j: r_{mj} \in [a,b]} \mathcal{E}_{mj}^{\text{test}}$ over measured operating points only. It is reported as an auxiliary descriptive statistic and is not used for checkpoint or hyperparameter selection. We treat nAUC as the primary comparison because curve minima remain sensitive to the number of evaluated configurations.

Table 12: Search grids used for Table 4. "Used" denotes the number of operating points within the common energy-reduction interval. Different grid sizes reflect method-specific mappings from control settings to realized energy reduction; nAUC is computed over the same common interval.

Method	Scanned settings	Runs	Used
DRAFTS	11 paired $(m_{\max}, m_{\text{init}})$ settings	11	10
FFT-high	cutoff $\{0.4, 0.6\}$; energy target 10^{-4}–0.2	10	8
FFT-low	fraction $\{0.4, 0.6, 0.8\}$; energy target 10^{-4}–0.2	11	8
DCT-high	cutoff $\{0.4, 0.6\}$; energy target 10^{-4}–0.2	10	8
STFT-high	cutoff $\{0.5, 0.6\}$; window 24; overlap 0.75	9	7
FFT-adap	mask maximum 0.1–0.212; energy target 3×10^{-5}–0.02	7	4

B COMPLETE RESULTS

Due to space constraints, the main text presents aggregated results and selected comparisons. Here we place the full results of all experiments in the following: long-term forecasting in Table 13, short-term forecasting in Table 14, classification in Table 15 and anomaly detection in Table 16.

B.1 LONG-TERM FORECASTING RESULTS

Results. As shown in Table 13, DRAFTS consistently improves six structurally different forecasting models on seven real-world datasets, achieving positive gains in 80 out of 84 average comparisons. The improvements are particularly pronounced on ILI, ECG, and Laser, demonstrating its effectiveness for non-stationary, physiological, and chaotic signals. Table 13 further shows that DRAFTS improves all 36 average comparisons on Noisy Lorenz63 across six models and three noise levels. The consistent gains at SNR $= 3$ dB demonstrate robustness to severe noise, while the substantial improvements at higher SNRs show that the benefits of DRAFTS extend to milder noise conditions.

Training-seed stability. We further evaluate training-seed variability on two representative forecasting settings: TimeFilter on ILI with a prediction horizon of 60, representing a relatively large observed improvement, and PatchTST on ETTh1 with a prediction horizon of 96, representing a typical small-gain setting. We perform full end-to-end training using seeds $\{42, 43, 44\}$. Within each seed, the Raw and +DRAFTS variants use identical data splits, backbone configurations, optimization settings, validation-MAE checkpoint-selection rules, and physical GPUs.

On TimeFilter–ILI, +DRAFTS changes MSE from 2.387 ± 0.402 to 2.000 ± 0.173 and MAE from 1.002 ± 0.074 to 0.934 ± 0.035, corresponding to mean paired reductions of $15.01\% \pm 12.73\%$ and $6.43\% \pm 6.54\%$, respectively. MSE improves in all three runs, while MAE improves in two of the three runs. On PatchTST–ETTh1, +DRAFTS changes MSE from 0.3995 ± 0.0101 to 0.3985 ± 0.0098 and MAE from 0.3957 ± 0.0045 to 0.3956 ± 0.0043, corresponding to mean paired changes of $0.26\% \pm 0.09\%$ and $0.02\% \pm 0.05\%$, respectively. Across the three seeds, DRAFTS consistently reduces MSE in both settings, with particularly substantial gains on TimeFilter–ILI. The average MAE is also improved, indicating that the benefits extend across both evaluation metrics. All deviations denote sample standard deviations across the three runs.

B.2 SHORT-TERM FORECASTING RESULTS

Setups. Performance is evaluated using sMAPE, MASE, and OWA, where lower values indicate better forecasting accuracy. OWA measures performance relative to the official Naive2 baseline. Overall M4 results are aggregated across the six frequency subsets, weighted by the number of series in each subset.

Results. As shown in Table 14, DRAFTS improves all three weighted-average metrics across all six backbones. For Crossformer, the weighted-average sMAPE, MASE, and OWA decrease by 1.84%, 2.21%, and 2.02%, respectively. On the Daily subset, DRAFTS reduces all three metrics for Crossformer by more than 10%.

Table 13: Full long-term forecasting results on seven real-world datasets and Noisy Lorenz63 at three noise levels with and without DRAFTS. Gray rows report the relative improvement (Δ) achieved by DRAFTS, where higher values are better.

Dataset	Horizon	Crossformer (2023)				PatchTST (2023)				iTransformer (2023)				TimeMixer (2024a)				TimeFilter (2025)				L-Drive (2026)			
		Raw		+DRAFTS		Raw		+DRAFTS		Raw		+DRAFTS		Raw		+DRAFTS		Raw		+DRAFTS		Raw		+DRAFTS	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	0.394	0.404	**0.383**	**0.397**	0.394	**0.392**	**0.393**	**0.392**	0.384	**0.391**	**0.380**	**0.391**	0.401	0.395	**0.392**	**0.394**	0.419	0.423	**0.416**	**0.422**	0.401	0.406	**0.388**	**0.402**
	192	0.436	0.431	**0.424**	**0.423**	**0.447**	**0.423**	**0.447**	**0.423**	0.438	0.422	**0.431**	**0.420**	0.443	**0.420**	**0.442**	**0.420**	0.482	0.449	**0.458**	**0.444**	0.449	**0.429**	**0.447**	**0.429**
	336	**0.471**	0.453	**0.471**	**0.450**	0.490	0.444	**0.488**	**0.443**	0.487	**0.446**	**0.475**	**0.446**	0.492	0.441	**0.484**	**0.440**	0.499	**0.456**	**0.494**	**0.456**	0.499	0.458	**0.489**	**0.452**
	720	0.495	**0.493**	**0.494**	**0.493**	0.506	0.470	**0.501**	**0.466**	0.481	0.466	**0.473**	**0.465**	0.496	0.460	**0.493**	**0.458**	0.551	0.500	**0.531**	**0.491**	0.538	0.493	**0.536**	**0.488**
	Avg	0.449	0.445	0.443	0.441	0.459	0.432	0.457	0.431	0.448	0.431	0.440	0.431	0.458	0.429	0.453	0.428	0.488	0.457	0.475	0.453	0.472	0.447	0.465	0.443
	Δ			1.34%	0.90%			0.44%	0.23%			1.79%	0.00%			1.09%	0.23%			2.66%	0.88%			1.48%	0.89%
ETTm1	96	0.345	0.362	**0.330**	**0.356**	0.327	**0.348**	**0.323**	**0.348**	0.330	0.351	**0.320**	**0.349**	0.318	0.342	**0.301**	**0.339**	**0.315**	0.346	**0.315**	**0.344**	0.308	0.341	**0.304**	**0.340**
	192	0.378	0.381	**0.366**	**0.374**	0.374	**0.370**	**0.371**	**0.370**	**0.379**	0.376	**0.379**	**0.375**	0.373	0.372	**0.369**	**0.371**	**0.364**	**0.371**	0.365	0.372	0.361	**0.365**	**0.356**	**0.365**
	336	0.430	0.411	**0.418**	**0.407**	0.406	0.395	**0.403**	**0.394**	0.416	0.399	**0.407**	**0.398**	0.407	0.394	**0.402**	**0.393**	**0.397**	**0.392**	0.399	0.395	0.405	0.396	**0.390**	**0.393**
	720	0.498	0.463	**0.490**	**0.461**	0.477	0.433	**0.471**	**0.431**	0.473	**0.433**	**0.471**	**0.433**	0.474	0.432	**0.466**	**0.430**	0.465	**0.432**	**0.464**	**0.432**	0.458	0.431	**0.452**	**0.430**
	Avg	0.413	0.404	0.401	0.400	0.396	0.387	0.392	0.386	0.400	0.390	0.394	0.389	0.393	0.385	0.385	0.383	0.385	0.385	0.386	0.386	0.383	0.383	0.376	0.382
	Δ			2.91%	0.99%			1.01%	0.26%			1.50%	0.26%			2.04%	0.52%			-0.26%	-0.26%			1.83%	0.26%
Weather	96	0.166	0.204	**0.161**	**0.202**	0.173	0.208	**0.168**	**0.205**	0.185	0.217	**0.177**	**0.211**	0.180	0.226	**0.162**	**0.210**	0.153	0.191	**0.150**	**0.190**	0.149	**0.186**	**0.148**	**0.186**
	192	0.208	0.244	**0.206**	**0.243**	0.215	0.248	**0.211**	**0.245**	0.235	0.258	**0.231**	**0.257**	0.246	0.282	**0.207**	**0.248**	0.200	**0.237**	**0.199**	**0.237**	0.199	0.234	**0.197**	**0.233**
	336	0.260	0.286	**0.256**	**0.284**	0.269	0.288	**0.266**	**0.286**	0.291	0.299	**0.283**	**0.294**	0.309	0.323	**0.260**	**0.286**	0.261	0.282	**0.257**	**0.281**	0.259	0.279	**0.256**	**0.278**
	720	**0.331**	0.339	**0.331**	**0.337**	0.345	0.337	**0.344**	**0.336**	0.362	0.345	**0.359**	**0.344**	0.395	0.375	**0.336**	**0.334**	0.337	**0.332**	**0.335**	0.333	0.339	**0.333**	**0.334**	**0.333**
	Avg	0.241	0.268	0.239	0.267	0.251	0.270	0.247	0.268	0.268	0.280	0.263	0.277	0.283	0.302	0.241	0.270	0.238	0.261	0.235	0.260	0.237	0.258	0.234	0.258
	Δ			0.83%	0.37%			1.59%	0.74%			1.87%	1.07%			14.84%	10.60%			1.26%	0.38%			1.27%	0.00%
Electricity	96	0.192	0.267	**0.178**	**0.257**	0.194	0.268	**0.191**	**0.266**	0.193	0.267	**0.190**	**0.263**	0.195	0.263	**0.181**	**0.259**	0.193	0.268	**0.187**	**0.263**	0.156	**0.246**	**0.152**	**0.246**
	192	0.198	**0.275**	**0.196**	**0.275**	0.199	0.276	**0.195**	**0.272**	0.198	0.275	**0.192**	**0.272**	0.193	0.267	**0.185**	**0.264**	0.199	0.276	**0.193**	**0.271**	0.170	**0.259**	**0.169**	**0.259**
	336	0.214	0.295	**0.208**	**0.288**	0.214	0.292	**0.213**	**0.290**	0.214	0.292	**0.211**	**0.289**	0.209	0.283	**0.201**	**0.280**	0.213	0.293	**0.201**	**0.287**	0.182	**0.270**	**0.181**	**0.270**
	720	0.249	0.324	**0.237**	**0.312**	0.255	0.326	**0.252**	**0.322**	0.257	0.327	**0.253**	**0.323**	0.249	0.315	**0.239**	**0.311**	0.260	0.331	**0.252**	**0.323**	0.215	0.303	**0.213**	**0.296**
	Avg	0.213	0.290	0.205	0.283	0.216	0.291	0.213	0.288	0.216	0.290	0.212	0.287	0.212	0.282	0.202	0.279	0.216	0.292	0.208	0.286	0.181	0.270	0.179	0.268
	Δ			3.76%	2.41%			1.39%	1.03%			1.85%	1.03%			4.72%	1.06%			3.70%	2.05%			1.10%	0.74%
ILI	24	4.736	1.480	**4.536**	**1.417**	3.633	1.079	**3.126**	**1.061**	3.507	1.071	**3.231**	**1.023**	3.124	1.136	**2.758**	**1.079**	3.848	1.055	**2.785**	**0.978**	3.848	0.940	**3.290**	**0.920**
	36	5.153	1.561	**4.403**	**1.423**	4.019	1.192	**3.824**	**1.159**	3.974	1.152	**3.935**	**1.146**	3.538	1.214	**3.134**	**1.144**	2.916	1.032	**2.735**	**1.007**	4.324	1.095	**4.225**	**1.070**
	48	5.244	1.576	**4.267**	**1.409**	2.939	1.099	**2.812**	**1.086**	2.513	1.005	**2.439**	**0.976**	3.055	1.130	**2.806**	**1.097**	2.744	1.037	**2.487**	**1.020**	2.241	0.926	**2.149**	**0.917**
	60	5.006	1.550	**4.831**	**1.529**	2.695	**1.071**	**2.607**	**1.071**	2.657	1.049	**2.435**	**0.994**	3.010	1.114	**2.843**	**1.108**	2.369	1.007	**1.988**	**0.935**	2.025	0.916	**1.794**	**0.871**
	Avg	5.035	1.542	4.509	1.445	3.322	1.110	3.092	1.094	3.163	1.069	3.010	1.035	3.182	1.149	2.885	1.107	2.969	1.033	2.499	0.985	3.110	0.969	2.865	0.945
	Δ			10.45%	6.29%			6.92%	1.44%			4.84%	3.18%			9.33%	3.66%			15.83%	4.65%			7.88%	2.48%
ECG	144	0.699	0.473	**0.572**	**0.466**	0.540	0.377	**0.513**	**0.365**	0.630	0.406	**0.579**	**0.397**	0.587	0.396	**0.581**	**0.383**	0.551	0.449	**0.489**	**0.387**	0.582	0.424	**0.540**	**0.404**
	288	0.843	0.579	**0.783**	**0.546**	0.613	0.498	**0.579**	**0.457**	0.677	0.519	**0.661**	**0.515**	0.679	0.542	**0.659**	**0.527**	0.607	0.472	**0.584**	**0.444**	0.746	0.604	**0.645**	**0.520**
	Avg	0.771	0.526	0.678	0.506	0.577	0.438	0.546	0.411	0.654	0.463	0.620	0.456	0.633	0.469	0.620	0.455	0.579	0.461	0.537	0.416	0.664	0.514	0.593	0.462
	Δ			12.06%	3.80%			5.37%	6.16%			5.20%	1.51%			2.05%	2.99%			7.25%	9.76%			10.69%	10.12%
Laser	50	0.106	0.147	**0.072**	**0.114**	0.140	0.182	**0.124**	**0.166**	0.150	0.166	**0.120**	**0.147**	0.170	0.194	**0.144**	**0.177**	0.126	0.152	**0.088**	**0.129**	0.106	0.132	**0.092**	**0.118**
	100	0.163	0.185	**0.162**	**0.178**	0.254	0.280	**0.221**	**0.255**	0.267	0.270	**0.244**	**0.258**	0.293	0.298	**0.257**	**0.282**	0.239	0.244	**0.210**	**0.229**	0.205	0.214	**0.204**	**0.212**
	Avg	0.135	0.166	0.117	0.146	0.197	0.231	0.173	0.211	0.209	0.218	0.182	0.203	0.232	0.246	0.201	0.230	0.183	0.198	0.149	0.179	0.156	0.173	0.148	0.165
	Δ			13.33%	12.05%			12.18%	8.66%			12.92%	6.88%			13.36%	6.50%			18.58%	9.60%			5.13%	4.62%
Lorenz63 (SNR=3)	16	0.490	0.333	**0.483**	**0.327**	0.718	0.527	**0.674**	**0.493**	0.672	0.489	**0.644**	**0.469**	0.742	0.556	**0.732**	**0.550**	0.551	0.391	**0.508**	**0.363**	0.532	0.364	**0.527**	**0.363**
	64	0.889	0.661	**0.836**	**0.610**	0.981	0.715	**0.937**	**0.685**	0.958	0.697	**0.937**	**0.680**	**0.931**	**0.689**	**0.931**	**0.689**	0.848	0.620	**0.839**	**0.616**	0.920	0.657	**0.914**	**0.656**
	Avg	0.690	0.497	0.660	0.469	0.850	0.621	0.806	0.589	0.815	0.593	0.791	0.575	0.837	0.623	0.832	0.620	0.700	0.506	0.674	0.490	0.726	0.511	0.721	0.510
	Δ			4.35%	5.63%			5.18%	5.15%			2.94%	3.04%			0.60%	0.48%			3.71%	3.16%			0.69%	0.20%
Lorenz63 (SNR=5)	16	0.410	0.320	**0.394**	**0.304**	0.626	0.500	**0.612**	**0.490**	0.574	0.459	**0.555**	**0.442**	0.654	0.531	**0.640**	**0.524**	0.508	0.415	**0.411**	**0.332**	0.444	**0.341**	**0.439**	**0.341**
	64	0.852	0.662	**0.768**	**0.595**	0.912	0.700	**0.891**	**0.683**	0.897	0.684	**0.885**	**0.674**	0.890	0.688	**0.886**	**0.685**	0.788	0.611	**0.778**	**0.608**	0.881	0.652	**0.873**	**0.649**
	Avg	0.631	0.491	0.581	0.450	0.769	0.600	0.752	0.587	0.736	0.572	0.720	0.558	0.772	0.610	0.763	0.605	0.648	0.513	0.595	0.470	0.663	0.497	0.656	0.495
	Δ			7.92%	8.35%			2.21%	2.17%			2.17%	2.45%			1.17%	0.82%			8.18%	8.38%			1.06%	0.40%
Lorenz63 (SNR=7)	16	0.303	0.268	**0.293**	**0.252**	0.560	0.481	**0.524**	**0.453**	0.496	0.429	**0.469**	**0.410**	0.581	0.504	**0.568**	**0.499**	0.414	0.377	**0.320**	**0.289**	0.362	0.312	**0.357**	**0.309**
	64	0.815	0.655	**0.741**	**0.598**	0.868	0.688	**0.847**	**0.672**	0.849	0.669	**0.841**	**0.662**	0.852	0.678	**0.850**	**0.677**	0.795	0.640	**0.724**	**0.586**	0.814	0.635	**0.808**	**0.634**
	Avg	0.559	0.462	0.517	0.425	0.714	0.585	0.686	0.563	0.673	0.549	0.655	0.536	0.717	0.591	0.709	0.588	0.605	0.509	0.522	0.438	0.588	0.474	0.583	0.472
	Δ			7.51%	8.01%			3.92%	3.76%			2.67%	2.37%			1.12%	0.51%			13.72%	13.95%			0.85%	0.42%

Table 14: Full short-term forecasting results on M4 with and without DRAFTS. Gray cells report the results obtained with DRAFTS and the corresponding relative improvement.

Subset	Metric	Crossformer (2023)		PatchTST (2023)		iTransformer (2023)		TimeMixer (2024a)		TimeFilter (2025)		L-Drive (2026)	
		Raw	+DRAFTS	Raw	+DRAFTS	Raw	+DRAFTS	Raw	+DRAFTS	Raw	+DRAFTS	Raw	+DRAFTS
Yearly	sMAPE	15.369	15.059	13.926	13.859	14.076	14.023	15.343	15.169	13.996	14.032	14.076	13.896
	MASE	3.571	3.442	3.177	3.125	3.169	3.159	3.450	3.368	3.115	3.111	3.177	3.102
	OWA	0.919	0.894	0.826	0.817	0.829	0.827	0.904	0.888	0.820	0.821	0.830	0.815
Quarterly	sMAPE	11.467	11.588	10.640	10.664	10.834	10.741	12.686	12.048	11.289	10.971	11.151	10.899
	MASE	1.302	1.369	1.257	1.257	1.260	1.251	1.552	1.460	1.326	1.264	1.371	1.306
	OWA	0.995	1.025	0.942	0.942	0.951	0.944	1.142	1.079	0.996	0.959	1.006	0.971
Monthly	sMAPE	14.388	14.011	13.329	13.241	13.493	13.478	14.200	14.127	14.397	14.014	14.043	13.883
	MASE	1.102	1.086	1.000	0.993	1.019	1.017	1.109	1.106	1.087	1.046	1.038	1.026
	OWA	1.017	0.996	0.932	0.926	0.947	0.945	1.013	1.010	1.010	0.978	0.975	0.964
Weekly	sMAPE	10.985	10.177	9.030	8.964	9.551	9.474	10.324	10.302	9.594	9.875	9.727	9.869
	MASE	3.075	3.254	2.544	2.539	2.633	2.645	3.302	3.337	2.907	2.793	2.897	2.821
	OWA	1.153	1.141	0.951	0.946	0.995	0.993	1.158	1.163	1.047	1.042	1.052	1.046
Daily	sMAPE	3.732	3.344	3.233	3.246	3.172	3.191	3.359	3.322	3.340	3.400	3.137	3.181
	MASE	4.136	3.662	3.416	3.473	3.386	3.403	3.630	3.589	3.598	3.672	3.363	3.425
	OWA	1.244	1.108	1.052	1.063	1.037	1.043	1.105	1.093	1.097	1.118	1.028	1.045
Hourly	sMAPE	21.441	20.112	18.845	18.564	17.391	17.260	18.890	17.910	19.610	19.073	17.146	16.904
	MASE	3.578	3.471	2.383	2.144	1.772	1.837	2.410	2.106	2.353	2.125	1.910	1.825
	OWA	1.330	1.272	1.010	0.952	0.843	0.853	1.017	0.927	1.025	0.962	0.865	0.841
Weighted Average	sMAPE	13.479	**13.231** (+1.84%)	12.402	**12.349** (+0.43%)	12.554	**12.513** (+0.33%)	13.647	**13.413** (+1.72%)	13.096	**12.845** (+1.91%)	12.893	**12.716** (+1.37%)
	MASE	1.864	**1.822** (+2.21%)	1.676	**1.662** (+0.84%)	1.680	**1.676** (+0.24%)	1.873	**1.828** (+2.41%)	1.729	**1.695** (+1.95%)	1.718	**1.682** (+2.11%)
	OWA	0.984	**0.964** (+2.02%)	0.895	**0.890** (+0.63%)	0.902	**0.900** (+0.28%)	0.993	**0.973** (+2.06%)	0.935	**0.917** (+1.93%)	0.925	**0.909** (+1.73%)

B.3 CLASSIFICATION RESULTS

Setups. We follow the classification protocol described in Experimental Details. Each raw/DRAFTS-augmented pair uses the same classification head design, and results are averaged over three runs with seeds 42, 43, and 44.

Table 15: Full classification results on 10 UEA datasets with and without DRAFTS averaged over three seeds. The better result in each Raw/+DRAFTS pair is highlighted in **bold red**.

Dataset	Crossformer (2023)		PatchTST (2023)		iTransformer (2023)		TimeMixer (2024a)		TimeFilter (2025)		L-Drive (2026)	
	Raw	+DRAFTS	Raw	+DRAFTS	Raw	+DRAFTS	Raw	+DRAFTS	Raw	+DRAFTS	Raw	+DRAFTS
Handwriting	18.98	**19.02**	19.84	**21.14**	11.76	**18.35**	4.59	**5.45**	9.45	**10.67**	9.22	**9.45**
LSST	51.55	**53.61**	**54.07**	53.73	57.91	**58.57**	56.27	**56.34**	60.34	**60.84**	54.41	**55.08**
NATOPS	73.33	**75.00**	74.07	**75.00**	80.19	**80.37**	31.11	**43.52**	77.96	**78.70**	29.26	**38.15**
RacketSports	71.93	**73.03**	63.38	**66.23**	70.18	**76.97**	50.44	**51.10**	61.62	**63.16**	58.55	**60.09**
SelfRegulationSCP2	55.00	**55.19**	48.15	**49.07**	51.30	**55.74**	48.89	**49.26**	**51.85**	50.56	50.56	**51.11**
CharacterTrajectories	98.31	**98.44**	**95.66**	95.59	97.63	**97.66**	93.13	**93.31**	97.03	**97.28**	90.37	**90.95**
SelfRegulationSCP1	85.21	**86.01**	**84.41**	84.30	87.94	**88.05**	77.02	**79.18**	83.96	**84.30**	**71.10**	69.85
ArticularyWordRecognition	85.11	**85.22**	**93.78**	**93.78**	96.11	**96.89**	26.11	**31.00**	96.44	**96.78**	37.67	**39.89**
JapaneseVowels	72.70	**76.49**	67.30	**67.39**	94.77	**95.59**	27.39	**62.43**	69.73	**72.25**	36.49	**49.19**
UWaveGestureLibrary	64.38	**65.21**	82.40	**82.50**	80.62	**81.15**	39.17	**59.58**	83.23	**86.15**	22.71	**23.65**
Average Accuracy	67.65	**68.72**	68.31	**68.87**	72.84	**74.93**	45.41	**53.12**	69.16	**70.07**	46.03	**48.74**

Results. As shown in Table 15, DRAFTS improves classification accuracy in 54 out of 60 backbone–dataset combinations, with one tie. All six backbones achieve higher average accuracy after incorporating DRAFTS, with gains ranging from 0.57 to 7.71 percentage points. The largest improvement is for TimeMixer, while consistent gains are also obtained for stronger classifiers such as iTransformer and TimeFilter. These results show that DRAFTS improves average classification accuracy across all six backbones, including classifiers with strong baseline performance.

B.4 ANOMALY DETECTION RESULTS

Setups. We adopt a forecasting-based unsupervised detection protocol (Hundman et al., 2018), with implementation and training details provided in Experimental Details. Given a window ending at time t, the model predicts the next observation $\mathbf{o} \in \mathbb{R}^N$. The point-wise anomaly score is

$$s_{t+1} = \frac{1}{N} \|\mathbf{x}_{t+1} - \mathbf{o}\|_2^2 . \tag{24}$$

Following prior high-quantile thresholding practices (Horak et al., 2022; Cao et al., 2026), we set the detection threshold τ to the 99.5th percentile of the normal validation scores. A test point is identified as anomalous when $s_{t+1} > \tau$. Test anomaly labels are used only for final evaluation. We report Precision, Recall, and F1-score averaged over the three runs.

Results. DRAFTS improves the average F1-score of every backbone. The largest average gain occurs on TimeFilter, from 82.73% to 83.85%. Improvements are particularly pronounced on SMD: Crossformer and TimeFilter gain 5.50 and 4.56 percentage points in F1, respectively. In both cases, DRAFTS substantially increases precision while largely preserving recall, indicating that it reduces false alarms without substantially compromising anomaly sensitivity.

Table 16: Full anomaly detection results on five datasets with and without DRAFTS, averaged over three seeds. The better result in each Raw/+DRAFTS pair is highlighted in **bold red**.

Dataset	Metric	Crossformer (2023)		PatchTST (2023)		iTransformer (2023)		TimeMixer (2024a)		TimeFilter (2025)		L-Drive (2026)	
		Raw	+DRAFTS	Raw	+DRAFTS	Raw	+DRAFTS	Raw	+DRAFTS	Raw	+DRAFTS	Raw	+DRAFTS
SMD	Precision	56.08	64.26	70.92	71.27	72.72	74.46	66.52	67.56	63.14	70.91	70.87	70.98
	Recall	88.47	87.70	89.22	88.33	87.84	88.01	89.85	89.98	89.05	87.70	88.38	88.43
	F1	68.64	**74.14**	**79.03**	78.89	79.56	**80.67**	76.44	**77.13**	73.81	**78.37**	78.65	**78.75**
MSL	Precision	82.38	82.43	80.58	79.96	79.75	79.86	81.15	80.97	80.17	81.89	81.27	80.25
	Recall	87.91	87.91	87.27	88.22	88.22	88.22	87.27	88.58	87.74	87.63	87.27	87.74
	F1	85.05	**85.08**	83.79	**83.89**	83.77	**83.83**	84.10	**84.59**	83.77	**84.62**	**84.16**	83.82
SMAP	Precision	92.98	93.48	94.20	94.13	93.34	91.80	93.78	93.83	93.70	95.04	93.99	94.39
	Recall	55.83	55.76	55.41	55.41	55.41	55.81	55.41	55.41	55.41	54.93	55.41	55.41
	F1	69.76	**69.85**	**69.78**	69.76	**69.54**	69.41	69.66	**69.68**	**69.64**	69.61	69.72	**69.83**
PSM	Precision	99.68	99.69	99.56	99.56	99.55	99.53	99.43	99.44	99.52	99.51	99.46	99.40
	Recall	84.94	84.96	86.78	86.95	85.31	85.66	90.07	88.64	85.15	85.15	87.27	88.68
	F1	91.72	**91.73**	92.72	**92.82**	91.88	**92.08**	**94.52**	93.72	**91.78**	91.77	92.96	**93.72**
SWaT	Precision	93.42	90.09	94.54	94.65	95.22	95.36	94.58	94.71	93.53	93.96	94.58	94.78
	Recall	95.64	94.81	95.81	95.81	95.81	95.81	95.81	95.81	95.81	95.81	95.81	95.81
	F1	**94.52**	92.24	95.17	**95.23**	95.52	**95.58**	95.19	**95.26**	94.66	**94.88**	95.19	**95.30**
Average F1		81.94	**82.61**	84.10	**84.12**	84.05	**84.31**	83.98	**84.07**	82.73	**83.85**	84.14	**84.28**

B.5 Ablation Study

Setups. We conduct the ablation with TimeFilter on the Noisy Lorenz63 datasets at $\mathrm{SNR} \in \{3, 5, 7\}$, using an input length of 96 and prediction horizons of 16 and 64. We compare six configurations: Raw, Raw + RC, Analyt. + [Amp., Freq., O], Analyt. + ϕ, Analyt. + [Amp., Freq., O] + RC, and the full DRAFTS. Raw + RC applies the same RC branch directly to the raw input without analyticization, controlling for the additional prediction branch. Analyt. denotes the analytic IMFs obtained through reflection-padded EMD and Hilbert analysis. The gate uses either the directly computed amplitude, frequency, and IMF order, or the four analytic features ϕ. All variants share the same data splits, TimeFilter configuration, training protocol, and seed. For each dataset–horizon pair, we reuse the corresponding preselected hyperparameters without retuning on the test set. We select checkpoints by validation MAE and report MSE and MAE from the same selected checkpoint. Table 5 averages the two horizons, while the complete horizon-level results are reported below.

Table 17: **Detailed ablation results on Noisy Lorenz63.** Each entry reports MSE / MAE for TimeFilter at seed 42. Analyt. denotes the analytic IMFs.

Dataset	Horizon	Raw	Raw + RC	Analyt. + [Amp., Freq., O]	Analyt. + ϕ	Analyt. + [Amp., Freq., O] + RC	DRAFTS (Ours)
SNR=3	16	0.551 / 0.391	0.578 / 0.426	0.582 / 0.427	0.546 / 0.389	0.524 / 0.378	**0.508 / 0.363**
SNR=3	64	0.848 / 0.620	0.849 / 0.628	0.843 / 0.617	0.886 / 0.651	0.880 / 0.645	**0.839 / 0.616**
SNR=5	16	0.508 / 0.415	0.451 / 0.369	0.499 / 0.408	0.490 / 0.399	0.412 / 0.333	**0.411 / 0.332**
SNR=5	64	0.788 / 0.611	0.821 / 0.635	0.827 / 0.635	0.843 / 0.652	0.792 / 0.615	**0.778 / 0.608**
SNR=7	16	0.414 / 0.377	0.329 / 0.298	0.420 / 0.380	0.363 / 0.334	0.324 / 0.295	**0.320 / 0.289**
SNR=7	64	0.795 / 0.640	0.794 / 0.641	0.781 / 0.621	0.733 / 0.594	0.734 / 0.596	**0.724 / 0.586**

Results. Table 17 shows that DRAFTS achieves the lowest MSE and MAE in all six SNR–horizon settings. Raw + RC is not consistently beneficial: it improves the short-horizon cases at SNR=5 and 7, but degrades both SNR=3 settings and the long-horizon SNR=5 case. The no-RC analytic variants are similarly inconsistent: neither they nor the Raw + RC control match the full model across the six evaluated settings. Compared with its no-RC counterpart, adding RC to the ϕ-based variant improves both metrics in all six settings. The full model also outperforms the Raw + RC and direct-attribute RC controls at both horizons for every SNR. These results are consistent with a complementary interaction between analytic feature encoding and residual-aware correction, with ϕ providing a stronger basis for correction than the direct attributes.