NOT ALL FREQUENCIES ARE HARMONIOUS: CONFLICT-AWARE MULTI-TASK SPECTRAL TRAINING FOR TIME-SERIES FORECASTING

Anonymous authors
Paper under double-blind review

ABSTRACT

Time-series forecasts can improve in some frequency bands while becoming less accurate in others. Aggregate error can conceal these trade-offs. Band loss values alone do not reveal how reducing one band's error affects the others during training. We therefore formulate time-series forecasting as a spectral multi-objective problem and propose Dyadic Spectral Alignment and Coordination (DySAC), a model-agnostic training framework. DySAC partitions the forecasting residual into non-overlapping dyadic bands. Each band defines a separate prediction objective for the same time series. It coordinates their gradients to seek compatible updates to the shared forecasting model. The resulting band coefficients reflect gradient directions, not just loss values or gradient norms. The aim is to distribute forecasting gains across frequencies rather than concentrate them in a few bands. The forecasting backbone and inference procedure remain unchanged. We evaluate prediction errors on the same bands to assess this balance. Across seven datasets and four backbones, DySAC yields a mean task-wise improvement of $+3.85\%$ over MSE training, with positive mean improvement in every dataset–backbone setting. The results support more balanced spectral performance, although individual tasks can still regress. Measured training-step times are 2.52–$3.30\times$ MSE in a four-setting timing benchmark, with no change to inference.

1 INTRODUCTION

Time-series forecasting is fundamental to applications such as energy management, traffic planning, and weather services (Lim & Zohren, 2021). Recent architectures, from attention-based and decomposition models (Zhou et al., 2021; Wu et al., 2021) to frequency-domain, patch-based, linear, multiscale, and variate-centric designs (Zhou et al., 2022; Nie et al., 2023; Zeng et al., 2023; Wu et al., 2023; Liu et al., 2024; Wang et al., 2024; Yi et al., 2023), have substantially improved forecasting. Despite these advances, most models are still trained with a single scalar objective, typically mean squared error (MSE) or mean absolute error (MAE).

Real-world time series contain different patterns across frequencies. Neural networks learn these components at different rates. This is known as spectral bias (Rahaman et al., 2019). A scalar loss combines the errors across frequencies. Training minimizes this aggregate error. The objective does not explicitly control trade-offs between frequency components.

Existing frequency-aware forecasting methods improve temporal representations or supervision. FEDformer (Zhou et al., 2022) and FreTS (Yi et al., 2023) model time series in the frequency domain. Structured objectives impose spectral, transformation-invariant, patch-level, transformed-label, decomposition-based, or quadratic-form multi-step constraints (Wang et al., 2025a; Lee et al., 2023; Kudrat et al., 2025; Wang et al., 2025b; Qiu et al., 2025; Wang et al., 2026). These approaches typically train the forecasting model with a scalar loss. Combining multiple losses also requires a choice of weights. Fixed weights prescribe the combination in advance. Adaptive rules based on loss values or gradient norms adjust it during training. Such statistics alone do not describe how fitting one frequency band affects another through the shared forecasting parameters.

Our forecasting results illustrate the need to examine individual bands (Section 3.3). TILDE-Q (Lee et al., 2023) improves the low-frequency bands but degrades the two highest reported bands on

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.

average. QDF (Wang et al., 2026) degrades a mid-spectrum band. Other baselines improve every band on average. Spectral trade-offs are therefore not universal, but aggregate forecast error does not explicitly control them. Gradient coordination offers a way to examine interactions between training objectives (Sener & Koltun, 2018; Yu et al., 2020; Liu et al., 2021). For time-series forecasting, we apply it to the prediction errors of different frequency bands. These objectives concern different components of the same forecast, not separate prediction tasks.

We treat each frequency band as a forecasting objective with shared model parameters. Its negative gradient indicates how to locally reduce its training error. Different bands may favor incompatible changes to the same forecasting model. Keeping their objectives separate makes these interactions explicit. Our central question is: *Can coordinating band-specific prediction objectives improve time-series forecasts across frequencies, rather than concentrate gains in a few bands?*

To this end, we propose Dyadic Spectral Alignment and Coordination (DySAC), a model-agnostic training framework for time-series forecasting. DySAC treats frequency-aware forecasting as a multi-objective optimization problem by decomposing the forecasting residual into non-overlapping dyadic frequency bands. Each band defines a distinct prediction-error objective, so training must account not only for the magnitude of each spectral error but also for how its gradient interacts with the others. Rather than assigning fixed band weights or deriving weights from scalar losses, DySAC uses the geometry of band-specific gradients to construct a coordinated update. Normalized band gradients determine a shared direction, while projections of the original gradients recover the update scale. In this way, DySAC seeks updates that are compatible across spectral objectives without prescribing a static preference over frequency bands.

We evaluate this principle using a task-aligned spectral protocol based on the same dyadic partition used during training. In addition to aggregate MSE and MAE, we report task-wise improvement, objective rankings, and average task rank to identify where forecasting gains occur across frequency bands. Across four backbones and 28 dataset–backbone settings, DySAC attains the best average task rank in 19 settings, including tied minima, and achieves the highest mean task-wise improvement, with a $+3.85\%$ gain over 119 valid task instances. It also obtains the highest mean improvement at every reported band index and the strongest worst-band mean improvement, $+2.39\%$. These results suggest that DySAC improves spectral coordination across objectives, although they do not imply dominance on every individual task or band.

Our contributions are as follows. (i) We formulate frequency-aware forecasting as a spectral multi-objective optimization problem, in which heterogeneous frequency components are preserved as explicit forecasting objectives rather than collapsed into a single scalar error. (ii) We introduce DySAC, a model-agnostic training framework that coordinates band-specific prediction errors through gradient geometry. DySAC derives band coefficients from cross-band gradient alignment instead of prescribing fixed weights from scalar statistics such as losses or gradient norms. (iii) We propose a task-aligned spectral evaluation protocol that complements standard aggregate metrics with task-wise improvement, objective ranking, and average task rank, enabling finer-grained assessment of where forecasting gains occur.

We discuss connections to forecasting architectures, spectral objectives, and gradient-based multi-objective optimization in Appendix A.

2 DySAC: Spectral Multi-Objective Training for Forecasting

In this section, we formally elaborate on the proposed DySAC training framework that modifies only the optimization step of a forecasting model. For each mini-batch, DySAC decomposes the forecasting residual into non-overlapping dyadic frequency bands and treats the error in each band as a separate objective. It then constructs a coordinated update from band-specific gradients: normalized gradients determine the shared direction, and projections of the original gradients recover the scale. The backbone and inference procedure remain unchanged.

2.1 Problem Formulation and Dyadic Spectral Objectives

Let $\mathbf{X}$ be a mini-batch of historical observations and f_θ be a differentiable forecasting model with $\theta \in \mathbb{R}^P$. For batch size B, forecast length L, and C target variables, the prediction and target satisfy

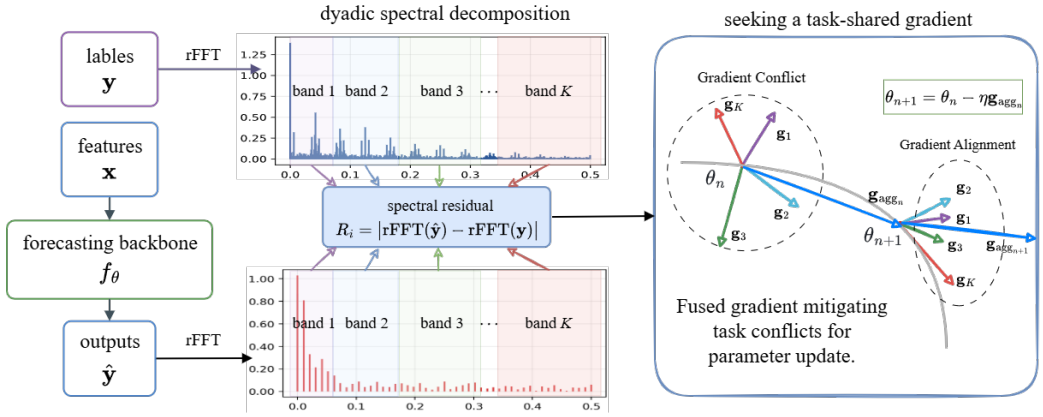


Figure 1: Overall framework of DySAC. The forecasting residual is transformed into the frequency domain and partitioned into non-overlapping dyadic bands, each defining an objective. DySAC coordinates their gradients into a shared descent direction and recovers its scale from the original gradients. The same objectives are used for task-wise evaluation.

$\hat{\mathbf{Y}} = f_\theta(\mathbf{X}) \in \mathbb{R}^{B\times L\times C}$ and $\mathbf{Y} \in \mathbb{R}^{B\times L\times C}$. Applying rFFT along the temporal dimension gives $\hat{\mathbf{Y}}^f = \text{rFFT}(\hat{\mathbf{Y}})$ and $\mathbf{Y}^f = \text{rFFT}(\mathbf{Y})$, with $M = \lfloor L/2 \rfloor + 1$ non-redundant frequency coefficients. The spectral residual is

$$\mathbf{R} = \hat{\mathbf{Y}}^f - \mathbf{Y}^f \in \mathbb{C}^{B\times M\times C}. \tag{1}$$

Inspired by dyadic multiresolution analysis (Mallat, 1989), we construct spectral tasks at multiple resolutions by partitioning the frequency axis into dyadic bands. Given a base bandwidth D_{base}, the bands are recursively defined by

$$\begin{aligned} S_1 &= 0, \qquad D_k = \min\left(D_{\text{base}} 2^{k-1},\, M - S_k\right), \\ S_{k+1} &= S_k + D_k, \qquad \mathcal{B}_k = \{S_k, \ldots, S_k + D_k - 1\}. \end{aligned} \tag{2}$$

The construction terminates when $S_{K+1} = M$, so $\{\mathcal{B}_k\}_{k=1}^K$ partitions $\Omega = \{0, \ldots, M-1\}$. It uses narrow low-frequency bands and wider high-frequency bands. This preserves low-frequency distinctions while giving weaker high-frequency components larger groups for useful update directions. Only D_{base} is required.

Using the spectral residual $\mathbf{R}$ from Eq. equation 1, we define the mean absolute spectral forecasting error for band $\mathcal{B}_k$ as

$$\mathcal{L}_k(\theta) = \frac{1}{BC|\mathcal{B}_k|} \sum_{b=1}^{B} \sum_{c=1}^{C} \sum_{m\in\mathcal{B}_k} |\mathbf{R}_{b,m,c}|\,. \tag{3}$$

Normalizing by $|\mathcal{B}_k| = D_k$ makes band losses comparable despite different widths. Treating each band loss as a separate objective gives the vector-valued problem

$$\min_\theta \mathbf{L}(\theta) = \min_\theta \left[\mathcal{L}_1(\theta), \ldots, \mathcal{L}_K(\theta)\right]^\top. \tag{4}$$

This vector objective preserves spectral identities while seeking a shared parameter solution that balances the bands.

2.2 Multi-Objective Spectral Coordination

Each band measures part of the same forecast error, and all bands share the model parameters. DySAC uses normalized gradients for direction and original gradients for scale, accounting for different error scales and learning speeds.

Band-wise gradient normalization. For each objective, we compute its gradient, normalize it, and stack the normalized gradients row-wise:

$$\mathbf{g}_k = \nabla_\theta \mathcal{L}_k(\theta), \qquad \mathbf{u}_k = \frac{\mathbf{g}_k}{\|\mathbf{g}_k\|_2 + \varepsilon}, \qquad \mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_K]^\top, \tag{5}$$

where $\varepsilon > 0$ ensures numerical stability. Normalization reduces the effect of objective-specific gradient scale, so high-energy components do not determine the direction only through larger gradients. Original magnitudes are retained for scale recovery.

Shared direction estimation. Using a regularized closed-form construction (Liu et al., 2025), we compute the coordination coefficients and shared direction:

$$\boldsymbol{\alpha} = \left(\mathbf{U}\mathbf{U}^\top + \lambda_r \mathbf{I}\right)^{-1} \mathbf{1}, \qquad \mathbf{d} = \frac{\mathbf{U}^\top \boldsymbol{\alpha}}{\|\mathbf{U}^\top \boldsymbol{\alpha}\|_2 + \varepsilon}. \tag{6}$$

Here, $\lambda_r > 0$ stabilizes the solve for correlated or nearly dependent gradients. For nonzero gradients, $[\mathbf{U}\mathbf{U}^\top]_{ij} = \rho_i \rho_j \cos \angle(\mathbf{g}_i, \mathbf{g}_j)$, with $\rho_k = \|\mathbf{g}_k\|_2 / (\|\mathbf{g}_k\|_2 + \varepsilon)$. As $\varepsilon \to 0$, the matrix records pairwise angular relations. The shared direction therefore reflects whether different bands admit compatible parameter changes.

Substituting $\mathbf{U}\mathbf{U}^\top \boldsymbol{\alpha} = \mathbf{1} - \lambda_r \boldsymbol{\alpha}$ into the definition of $\mathbf{d}$ gives $\mathbf{U}\mathbf{d} = (\mathbf{1} - \lambda_r \boldsymbol{\alpha})/(\|\mathbf{U}^\top \boldsymbol{\alpha}\|_2 + \varepsilon)$. Under ideal normalization and a consistent system $\mathbf{U}\mathbf{v} = \mathbf{1}$, the limit $\lambda_r \to 0$ gives the same positive cosine with every band gradient. Thus, each band receives the same fraction of its steepest-descent rate. Finite regularization can perturb this relation, so it does not guarantee the Adam step or final test performance. Conditions are given in Appendix B.

Coordination as solved, rather than designed, weighting. Expanding $\mathbf{d}$ in terms of the original gradients,

$$\mathbf{d} \;\propto\; \sum_{k=1}^{K} \alpha_k \mathbf{u}_k = \sum_{k=1}^{K} w_k \mathbf{g}_k, \qquad w_k = \frac{\alpha_k}{\|\mathbf{g}_k\|_2 + \varepsilon}, \tag{7}$$

so the update is a linear combination of band gradients with effective coefficients w_k. DySAC does not prescribe them from scalar statistics. Instead, $\boldsymbol{\alpha}$ depends on the full normalized gradient matrix and therefore on cross-band geometry: pairwise angles in the ideal limit, plus norm-dependent factors for finite ε. Scalar-only rules cannot distinguish configurations with identical scalar inputs but different geometry. The example in Appendix B makes this limitation explicit.

When the resulting direction satisfies $\langle \mathbf{g}_k, \mathbf{d} \rangle > 0$ for all k, the direction $-\mathbf{d}$ is a strict common first-order descent direction. A sufficiently small direct step then decreases every band objective. This is necessary but not sufficient for local Pareto optimality, and concerns the direct step rather than the realized Adam update.

Update-scale recovery. To recover the strength of the original objectives, we aggregate their projections onto $\mathbf{d}$ and perform the parameter update:

$$q = \sum_{k=1}^{K} \langle \mathbf{g}_k, \mathbf{d} \rangle, \qquad \mathbf{g}_{\text{final}} = q\mathbf{d}, \qquad \theta \leftarrow \theta - \eta \mathbf{g}_{\text{final}}, \tag{8}$$

where η is the learning rate. The projections restore update magnitude after direction normalization. The update can therefore follow a balanced direction while retaining the strength required by current forecasting errors (Liu et al., 2025).

If all band projections are positive, a sufficiently small direct step decreases every band objective. Under smooth deterministic objectives and direct updates, the projected residual admits a conditional $O(T^{-1/2})$ best-iterate bound. Conditions and proofs are given in Appendix B.2. These results do not establish Pareto stationarity, stochastic-Adam convergence, or improvement on every test band.

The backbone produces its prediction. DySAC replaces the scalar-loss gradient with $\mathbf{g}_{\text{final}}$; the architecture and inference procedure remain unchanged. The pseudocode is in Appendix D.

2.3 Task-Wise Spectral Evaluation

We evaluate every model on DySAC's dyadic partition. MSE and MAE remain the primary metrics; task-wise results show whether gains spread across bands. Details are in Appendix C.

Task-wise improvement. For objective m, seed s, and band k, $I_{k,s}^{(m)}$ is the spectral-error reduction over matched MSE (Appendix C); positive values indicate better forecasts.

For valid seeds $\mathcal{S}_k^{(m)}$, we report the mean and population standard deviation:

$$\mu_k^{(m)} = \sum_{s \in \mathcal{S}_k^{(m)}} I_{k,s}^{(m)} \Big/ \left|\mathcal{S}_k^{(m)}\right| \; ; \; \sigma_k^{(m)} = \left[\sum_{s \in \mathcal{S}_k^{(m)}} \left(I_{k,s}^{(m)} - \mu_k^{(m)}\right)^2 \Big/ \left|\mathcal{S}_k^{(m)}\right| \right]^{1/2} . \tag{9}$$

Ranking and aggregation. Missing or non-finite results are excluded rather than assigned zero. Within each setting and task, objectives are ranked by $\mu_k^{(m)}$; larger improvement gets the smaller rank, with average ranks for ties. For objective m, let $\mathcal{K}^{(m)}$ be its valid task instances and $R_k^{(m)}$ its rank on task k. We average only valid task ranks:

$$\mathrm{AvgRank}^{(m)} = \sum_{k \in \mathcal{K}^{(m)}} R_k^{(m)} \Big/ \left|\mathcal{K}^{(m)}\right| .$$

Missing tasks are excluded; lower values indicate stronger cross-frequency competitiveness, not Pareto optimality.

3 Experiments

We evaluate DySAC through four questions: **Q1,** aggregate forecasting accuracy across backbones and datasets; **Q2,** improvement on individual dyadic spectral objectives over standard MSE; **Q3,** competitiveness across heterogeneous spectral objectives; and **Q4,** whether explicit spectral multi-objective coordination outperforms direct aggregation of the same dyadic objectives.

3.1 Experimental Setup

Datasets and backbones. We evaluate seven benchmarks with four forecasting backbones spanning linear, frequency-domain, inverted-attention, and multiscale-mixing designs. This combination tests whether the training framework transfers across datasets and backbone architectures. Within each matched setting, the forecasting backbone and inference procedure remain unchanged, so comparisons focus on training objectives and their coordination. Dataset usage, backbone mappings, forecasting configurations, data splits, and preprocessing details are provided in Appendix D.

Baselines. We compare DySAC with MSE training and six structured objectives: DBLoss Qiu et al. (2025), FreDF Wang et al. (2025a), PS-Loss Kudrat et al. (2025), QDF Wang et al. (2026), TILDE-Q Lee et al. (2023), and Time-o1 Wang et al. (2025b). All methods use the same backbone, data split, configuration, and training protocol.

Metrics and implementation. We report mean squared error (MSE) and mean absolute error (MAE), with lower values preferred (Hyndman & Koehler, 2006). We also evaluate the dyadic tasks from Section 2.1. Each task is compared with a matched MSE run using the same seed. We report the mean and population standard deviation over three seeds, rank objectives within each task, and average the valid task ranks within each setting. Lower average rank indicates stronger cross-objective competitiveness. These task-wise measures describe empirical balance, not Pareto optimality. Appendix F provides a complementary evaluation with SMAPE, WAPE, per-window Pearson correlation, and an OWA-style score.

3.2 Forecasting Performance

(Q1) Table 1 reports the seven-dataset macro means for the four backbones. DySAC is best in six of eight backbone–metric columns: both MSE and MAE for DLinear, iTransformer, and TimeMixer. On FreTS, TILDE-Q has the lowest MSE (3.0931) and DBLoss the lowest MAE (0.4598); DySAC obtains 3.7328 and 0.4760. DySAC still improves over MSE training in all eight columns.

Datasets have equal weight. Because these are unnormalized means, they summarize error magnitude rather than dataset-level rank and remain sensitive to scale, especially on Fred-md. DySAC has the lowest mean column rank (2.33) over 56 columns, leads 29 columns including ties, and improves over MSE in 52. Thus the macro results do not imply universal improvement. Complete per-dataset

means, standard deviations, and non-winning cases appear in Appendix E. The aggregate and task-wise analyses use matching configurations, but the aggregate tables do not provide checkpoint or seed decompositions.

Heterogeneous error metrics. Table 2 gives equal-weight macro means over the seven datasets. DySAC is best in seven of the sixteen backbone–metric columns and improves over MSE training in all sixteen. DBLoss leads most FreTS and TimeMixer columns, while PS-Loss leads CORR for FreTS and iTransformer. These averages summarize magnitudes rather than per-dataset ranks and can hide regressions. Full results and metric definitions appear in Appendix F. Each cell averages the seven dataset-level three-seed means with equal dataset weight; test windows are not pooled. SMAPE and WAPE are percentages, CORR is averaged without a Fisher transform, and OWA is defined in Appendix F. All metrics use standardized data, and no standard deviation is reported.

Table 1: MSE/MAE averaged equally over seven datasets (lower is better). Blue bold: best; red: second-best distinct mean. Full results: Appendix E.

Objective	DLinear		FreTS		iTransformer		TimeMixer	
	MSE↓	MAE↓	MSE↓	MAE↓	MSE↓	MAE↓	MSE↓	MAE↓
Origin	2.7889	0.4573	7.9872	0.5390	3.0951	0.4080	2.4858	0.4155
+DBLoss	2.3632	0.4258	3.1058	**0.4598**	2.7682	0.3905	2.1404	0.3910
+FreDF	2.4877	0.4414	4.7588	0.4961	2.6218	0.3970	2.1972	0.4006
+PS-Loss	2.2376	0.4254	5.9104	0.4943	2.7416	0.3940	2.0537	0.3965
+QDF	2.6131	0.4475	3.6318	0.5052	2.5962	0.4005	2.2847	0.4090
+TILDE-Q	2.1340	0.4284	**3.0931**	0.4681	2.9521	0.4038	2.0722	0.4001
+Time-o1	2.8230	0.4593	5.4931	0.5231	2.6958	0.4015	2.2567	0.4167
OURS	**2.1281**	**0.4217**	3.7328	0.4760	**2.5944**	**0.3876**	**1.9129**	**0.3885**

Table 2: Dataset-macro-averaged heterogeneous metrics for four backbones. Lower is better except for CORR. Blue bold: best; red: second-best distinct mean.

Objective	DLinear				FreTS				iTransformer				TimeMixer			
	SMAPE↓	WAPE↓	CORR↑	OWA↓	SMAPE↓	WAPE↓	CORR↑	OWA↓	SMAPE↓	WAPE↓	CORR↑	OWA↓	SMAPE↓	WAPE↓	CORR↑	OWA↓
Origin	53.2366	34.5879	0.4833	0.9017	54.5083	35.1419	0.4510	0.9963	49.4133	31.2921	0.5197	0.8366	51.2040	32.6413	0.5135	0.8645
+DBLoss	**50.6524**	32.8703	0.4976	0.8532	**50.2226**	**32.1563**	0.4692	**0.8947**	47.7450	30.2265	0.5466	0.8060	**48.8011**	**31.1445**	0.5399	**0.8262**
+FreDF	52.3385	33.9121	0.5109	0.8816	52.1084	33.7068	0.4547	0.9469	48.6943	30.5606	0.5362	0.8201	50.2818	31.8548	0.5451	0.8442
+PS-Loss	50.7639	32.8536	0.5043	0.8530	51.5649	33.5974	**0.4858**	0.9384	47.9206	30.2309	**0.5616**	0.8084	49.4793	31.5840	0.5615	0.8338
+QDF	52.6706	34.1302	0.4931	0.8887	54.2933	34.2335	0.4595	0.9754	49.1171	30.9462	0.5307	0.8280	50.7463	32.4158	0.5207	0.8549
+TILDE-Q	51.8739	33.5835	0.5107	0.8678	51.5267	32.9633	0.4512	0.9232	49.0367	31.1522	0.5389	0.8272	50.3307	32.2146	0.5433	0.8511
+Time-o1	53.4229	34.6911	0.4810	0.9048	54.1029	34.4778	0.4375	0.9815	49.1163	30.9673	0.5246	0.8290	51.4278	32.9474	0.5182	0.8723
OURS	50.7697	**32.7767**	**0.5293**	**0.8499**	51.3290	32.8289	0.4749	0.9225	**47.5425**	**29.9433**	0.5589	**0.8006**	49.4541	31.3496	**0.5655**	0.8275

Training overhead. The three-seed benchmark measures complete end-to-end optimization-step times, including data transfer, forward and backward computation, objective construction, gradient clipping, and the optimizer update. Table 3 reports these costs relative to the matched MSE means for FreTS and TimeMixer on Electricity and Traffic. DySAC costs $3.30\times/3.02\times$ MSE with FreTS and $2.80\times/2.52\times$ with TimeMixer. Detailed timings and limitations are reported in Appendix G. This extra cost occurs only during training; the forecasting backbone and inference procedure remain unchanged.

Table 3: Training-step time relative to MSE (ratio of three-run means). Full timings and configuration: Table 18 in Appendix G.

Objective	FreTS		TimeMixer	
	Electricity	Traffic	Electricity	Traffic
MSE	1.00	1.00	1.00	1.00
DBLoss	1.00	1.01	1.49	2.12
FreDF	1.01	1.01	1.57	1.23
PS-Loss	0.98	1.00	1.88	1.69
QDF	0.84	0.90	1.27	2.18
TILDE-Q	0.99	1.01	1.08	1.08
Time-o1	0.78	0.84	1.04	1.52
OURS	**3.30**	**3.02**	**2.80**	**2.52**

3.3 Task-Wise Spectral Performance

(Q2, Q3) This section evaluates the trained models on exactly the dyadic objectives used during optimization. We consider 28 dataset–backbone settings across the seven datasets and four backbones (Appendix D). For each task, we compare every objective with a same-seed MSE baseline, aggregate three seeds using the population standard deviation, and rank objectives by mean improvement. The dyadic partition gives three to five tasks per setting, or 119 valid task instances per objective.

Figure 2 shows task-wise improvement for iTransformer and TimeMixer on four datasets. OURS has a positive mean in all 28 settings, and 110 of 119 seed-averaged task improvements are positive. Individual tasks can still regress. Baseline behavior also varies: QDF has a negative mean on Fred-md with FreTS, while TILDE-Q has negative dataset-level means on Electricity and Traffic despite its strong Fred-md result. Thus comparable aggregate MSE can hide different distributions of spectral gains. Panels include jittered per-seed measurements and mean markers. Values outside $[-15\%, 20\%]$ are clipped only for display.

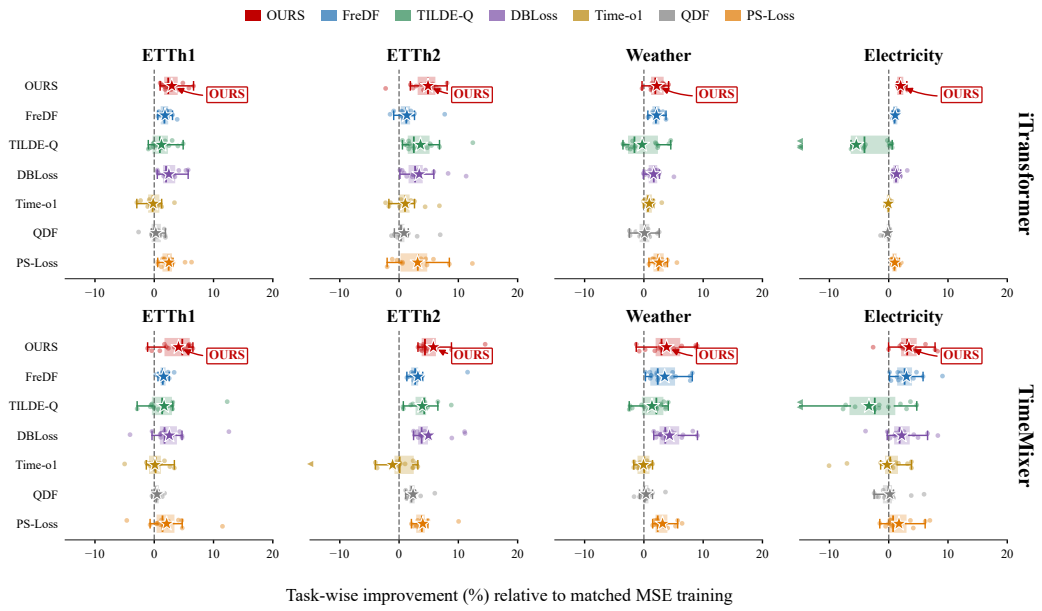


Figure 2: Task-wise spectral improvement for iTransformer (top row) and TimeMixer (bottom row) on ETTh1, ETTh2, Weather, and Electricity. Each panel shows improvement relative to MSE training; OURS is red. The complete four-backbone version is in Appendix H.

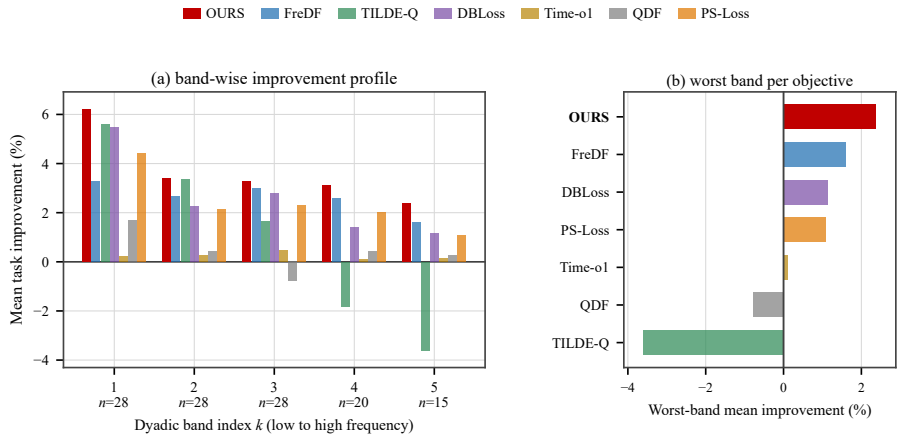


Figure 3: Band-wise improvement over matched MSE training. (a) Mean improvement by band k; (b) mean improvement of each objective's *worst* band. $k = 1$ is the lowest-frequency band.

Figure 3 resolves these results by band index. TILDE-Q improves the first two bands by $+5.59\%$ and $+3.37\%$, but its fourth- and fifth-band means are -1.81% and -3.60%. QDF improves the first band by $+1.68\%$ while degrading the third by 0.78%. These examples motivate band-level analysis. The pattern is not universal: DBLoss and PS-Loss improve every reported band, with worst-band means of $+1.15\%$ and $+1.09\%$.

DySAC's band means are $+6.22\%$, $+3.41\%$, $+3.26\%$, $+3.10\%$, and $+2.39\%$ from low to high index, the highest at every index. It has the highest worst-band mean ($+2.39\%$), followed by FreDF ($+1.60\%$); Time-o1 reaches $+0.12\%$. Its task-weighted improvement is $+3.85\%$, compared with $+2.85\%$ for DBLoss, $+2.74\%$ for FreDF, and $+2.56\%$ for PS-Loss. Because the five means use 28, 28, 28, 20, and 15 settings, respectively, they summarize band indices rather than a common physical-frequency response. They do not guarantee positive held-out improvement on every task.

Table 4: Component ablation on FreTS over Electricity and Weather. "MSE baseline" uses time-domain MSE; "Single spectral loss" applies the magnitude-L1 residual in equation 3 to the full spectrum; "Dyadic + Mean" averages gradients over dyadic bands; "Dyadic + ConFIG" coordinates the same band objectives. Cells show three-seed mean (standard deviation). Bold denotes DySAC; blue denotes the best mean per metric.

Variant	Electricity		Weather	
	MSE ↓	MAE ↓	MSE ↓	MAE ↓
MSE baseline	0.1903 (0.0006)	0.2795 (0.0004)	0.2140 (0.0029)	0.2757 (0.0044)
Single spectral loss	0.1910 (0.0006)	0.2769 (0.0010)	0.2101 (0.0028)	0.2637 (0.0070)
Dyadic + Mean	0.1879 (0.0004)	0.2743 (0.0005)	0.2100 (0.0031)	0.2608 (0.0064)
Dyadic + ConFIG	**0.1877 (0.0008)**	**0.2733 (0.0011)**	**0.2078 (0.0015)**	**0.2569 (0.0026)**

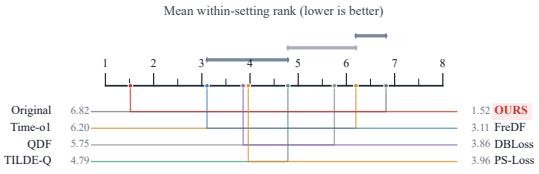


Figure 4: Critical-difference diagram of average task ranks over 28 dataset–backbone settings. Lower is better; horizontal bars connect groups that are not significantly different.

Figure 4 gives a statistical view of setting-level ranks for time-series forecasting. DySAC has the lowest mean rank, 2.01 ± 0.91, and is rank one in 19 of 28 settings, including ties. These ranks compare frequency-specific errors from the same time-series forecast and measure cross-frequency consistency. Non-significance groupings complement Table 20; they do not imply Pareto optimality.

Figure 5 relates rank to task-wise improvement in spectral forecasting. DySAC combines the best rank with the largest mean gain ($+3.85\%$); FreDF has the next-best rank (3.42), while DBLoss has the next- largest gain ($+2.85\%$) but a weaker rank (3.89). This shows that gain magnitude and cross-frequency consistency are distinct. The upper-left region therefore indicates larger spectral gains with more consistent forecasting. TILDE-Q is strong on Fred-md but less consistent elsewhere. With Table 20, plots summarize spectral balance, not universal improvement.

3.4 Ablation Study

Component ablation. **(Q4)** Table 4 separates spectral supervision, dyadic task decomposition, and gradient coordination on FreTS. A single spectral objective improves three of four metrics. Keeping the bands as separate objectives improves the results further, and ConFIG-style coordination gives the best mean in every column. The "Dyadic + Mean" row keeps the same objectives but uses uniform coefficients, $w_k = 1/K$. Therefore, the gap to the full method isolates coordination from decomposition. It does not test adaptive weighting. The full method reduces MSE by about $1.37\%/2.90\%$ and MAE by $2.22\%/6.82\%$ on Electricity/Weather; detailed results are in Appendix J.

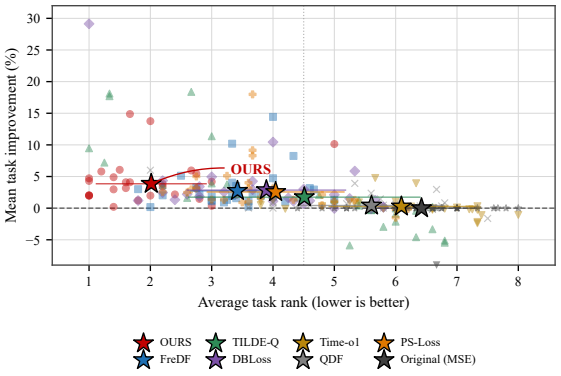


Figure 5: Cross-objective balance over 28 settings. Stars show mean rank and improvement; bars show rank standard deviation; triangles mark settings below -9%.

Gradient diagnostics. Table 22 in Appendix K provides complementary evidence on Fred-md across four backbones and three seeds. In the final epoch, conflicting gradient pairs account for 10.89%–44.00% of pairs, while no task has a negative inner product with the merged gradient in any of the twelve runs. These are different diagnostics, not a before–after change in one conflict rate. Alongside the fixed-objective ablation, they motivate using gradient geometry in the coefficient solve. They do not establish superiority over adaptive weighting or guarantee that the subsequent Adam step decreases every task loss.

Table 5: Spectral loss ablation on FreTS over Electricity and Weather. "Single spectral" uses one global objective; "Dyadic + ConFIG" is the complete framework. L1 is the magnitude-L1 residual in equation 3; L2 is squared-magnitude L2. Cells show three-seed mean (standard deviation). Bold denotes complete DySAC; blue denotes the best mean per metric.

Architecture	Spectral loss	Electricity		Weather	
		MSE ↓	MAE ↓	MSE ↓	MAE ↓
Single spectral	L1	0.1910 (0.0006)	0.2769 (0.0010)	0.2101 (0.0028)	0.2637 (0.0070)
Single spectral	L2	0.1894 (0.0002)	0.2800 (0.0014)	0.2124 (0.0028)	0.2736 (0.0045)
Dyadic + ConFIG	L1	**0.1877 (0.0008)**	**0.2733 (0.0011)**	**0.2078 (0.0015)**	**0.2569 (0.0026)**
Dyadic + ConFIG	L2	0.1897 (0.0006)	0.2789 (0.0005)	0.2095 (0.0012)	0.2703 (0.0047)

Table 6: Frequency-partition ablation on FreTS and TimeMixer over Electricity and Weather. All variants use the same magnitude-L1 spectral objective and ConFIG-style gradient coordination; only the construction of the frequency bands is changed. Cells report mean (standard deviation) over three seeds; bold marks the dyadic partition used by DySAC, and blue marks the lowest mean within each backbone and metric column.

Partition	Electricity		Weather	
	MSE ↓	MAE ↓	MSE ↓	MAE ↓
FreTS				
Dyadic	**0.1877 (0.0008)**	**0.2733 (0.0011)**	**0.2078 (0.0015)**	**0.2569 (0.0026)**
Uniform	0.1911 (0.0003)	0.2765 (0.0007)	0.2075 (0.0007)	0.2579 (0.0042)
Geometric	0.1889 (0.0007)	0.2746 (0.0005)	0.2076 (0.0019)	0.2569 (0.0016)
Random	0.1912 (0.0004)	0.2771 (0.0009)	0.2086 (0.0016)	0.2590 (0.0031)
TimeMixer				
Dyadic	**0.1698 (0.0023)**	**0.2676 (0.0025)**	**0.2103 (0.0014)**	**0.2500 (0.0008)**
Uniform	0.1752 (0.0019)	0.2714 (0.0026)	0.2146 (0.0007)	0.2532 (0.0014)
Geometric	0.1709 (0.0024)	0.2680 (0.0032)	0.2127 (0.0020)	0.2521 (0.0019)
Random	0.1795 (0.0029)	0.2758 (0.0025)	0.2139 (0.0007)	0.2520 (0.0009)

Choice of spectral loss. Table 5 compares the magnitude-L1 spectral residual of equation 3 with a squared-magnitude L2 objective on FreTS. At the single-spectrum level the two losses trade off across metrics, but within the complete Dyadic + ConFIG framework L1 attains the lower mean on all four metrics. We therefore retain magnitude-L1; the full discussion appears in Appendix J.

Choice of frequency partition. Table 6 compares dyadic, uniform, geometric, and random partitions while keeping the spectral loss and coordination fixed. Dyadic is best in six of eight comparisons and tied in one. On FreTS/Electricity, its MSE/MAE reductions over uniform and random are $1.78\%/1.16\%$ and $1.83\%/1.37\%$; on TimeMixer/Electricity they are $3.08\%/1.40\%$ and $5.40\%/2.97\%$. Gains over geometric are smaller: at most 0.64% MSE and 0.47% MAE on FreTS, and $1.13\%/0.83\%$ on TimeMixer. FreTS/Weather is not a unique dyadic win: uniform and geometric have lower MSE, while geometric ties dyadic in MAE. These small three-seed differences do not establish a unique optimal growth law. We use dyadic partitioning because it provides multiresolution bands without an additional growth-ratio hyperparameter. Further comparisons are in Appendix J.

4 Conclusion

We presented DySAC, a model-agnostic training framework for time-series forecasting. DySAC decomposes the residual into dyadic frequency bands and coordinates their gradients without changing the architecture or inference procedure. Across seven datasets and four backbones, it improves over MSE training in 52 of 56 MSE/MAE columns and achieves a mean task-wise improvement of $+3.85\%$. The mean improvement is positive in all 28 dataset–backbone settings, although bands can still regress. The cost is additional training computation because DySAC computes one gradient per band. Future work can reduce this cost, learn the frequency partition, and evaluate spectral improvements under partitions that differ from the training objectives.

ACKNOWLEDGMENTS

Use unnumbered third level headings for the acknowledgments. All acknowledgments, including those to funding agencies, go at the end of the paper.

AI USE STATEMENT

AI-based tools were used only to assist with table conversion and manuscript editing, including shortening and polishing selected descriptions to meet conference length and formatting requirements. They were not used to generate the research idea, methodology, mathematical proofs, experimental design, data analysis, results, or conclusions. The authors reviewed and verified all content and took full responsibility for the final manuscript.

REFERENCES

Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. 2018. URL https://arxiv.org/abs/1803.01271.

Cristian Challu, Kin G. Olivares, Boris N. Oreshkin, Federico Garza Ramirez, Max Mergenthaler Canseco, and Artur Dubrawski. NHITS: Neural hierarchical interpolation for time series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 6989–6997, 2023. doi: 10.1609/aaai.v37i6.25854.

Zhao Chen, Vijay Badrinarayanan, Chen-Yu Lee, and Andrew Rabinovich. Gradnorm: Gradient normalization for adaptive loss balancing in deep multitask networks. In Jennifer Dy and Andreas Krause (eds.), *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 794–803. PMLR, 10–15 Jul 2018. URL https://proceedings.mlr.press/v80/chen18a.html.

Roberto Cipolla, Yarin Gal, and Alex Kendall. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7482–7491, 2018. doi: 10.1109/CVPR.2018.00781.

Rob J. Hyndman and Anne B. Koehler. Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4):679–688, 2006. ISSN 0169-2070. doi: 10.1016/j.ijforecast.2006.03.001. URL https://www.sciencedirect.com/science/article/pii/S0169207006000239.

Dilfira Kudrat, Zongxia Xie, Yanru Sun, Tianyu Jia, and Qinghua Hu. Patch-wise structural loss for time series forecasting. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 31841–31859. PMLR, 13–19 Jul 2025. URL https://proceedings.mlr.press/v267/kudrat25a.html.

Hyunwook Lee, Chunggi Lee, Hongkyu Lim, and Sungahn Ko. TILDE-q: a transformation invariant loss function for time-series forecasting. 2023. URL https://openreview.net/forum?id=D1Sawu2y1QG.

Bryan Lim and Stefan Zohren. Time-series forecasting with deep learning: a survey. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 379 (2194):20200209, 2021. doi: 10.1098/rsta.2020.0209.

Bo Liu, Xingchao Liu, Xiaojie Jin, Peter Stone, and Qiang Liu. Conflict-averse gradient descent for multi-task learning. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 18878–18890. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/9d27fdf2477ffbff837d73ef7ae23db9-Paper.pdf.

Qiang Liu, Mengyu Chu, and Nils Thuerey. Config: Towards conflict-free training of physics informed neural networks. In Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu (eds.), *International Conference on Learning Representations*, volume 2025, pp. 59531–59566, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/file/94e85561a342de88b559b72c9b29f638-Paper-Conference.pdf.

Shikun Liu, Edward Johns, and Andrew J. Davison. End-to-end multi-task learning with attention. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1871–1880, 2019. doi: 10.1109/CVPR.2019.00197.

Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. itransformer: Inverted transformers are effective for time series forecasting. In B. Kim, Y. Yue, S. Chaudhuri, K. Fragkiadaki, M. Khan, and Y. Sun (eds.), *International Conference on Learning Representations*, volume 2024, pp. 11116–11140, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/file/2ea18fdc667e0ef2ad82b2b4d65147ad-Paper-Conference.pdf.

Spyros Makridakis, Evangelos Spiliotis, and Vassilios Assimakopoulos. The m4 competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, 36 (1):54–74, 2020. ISSN 0169-2070. doi: 10.1016/j.ijforecast.2019.04.014. URL https://www.sciencedirect.com/science/article/pii/S0169207019301128. M4 Competition.

S.G. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693, 1989. doi: 10.1109/34.192463.

Aviv Navon, Aviv Shamsian, Idan Achituve, Haggai Maron, Kenji Kawaguchi, Gal Chechik, and Ethan Fetaya. Multi-task learning as a bargaining game. 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. 16428–16446. PMLR, 17–23 Jul 2022. URL https://proceedings.mlr.press/v162/navon22a.html.

Yuqi Nie, Nam H Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=Jbdc0vTOcol.

Boris N. Oreshkin, Dmitri Carpov, Nicolas Chapados, and Yoshua Bengio. N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=r1ecqn4YwB.

Xiangfei Qiu, Xingjian Wu, Hanyin Cheng, Xvyuan Liu, Chenjuan Guo, Jilin Hu, and Bin Yang. Dbloss: Decomposition-based loss function for time series forecasting. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 27741–27768. Curran Associates, Inc., 2025. doi: 10.52202/085713-0936. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/2823c4f838727564c108c85d7e6e8507-Paper-Conference.pdf.

Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Draxler, Min Lin, Fred Hamprecht, Yoshua Bengio, and Aaron Courville. On the spectral bias of neural networks. In Kamalika Chaudhuri and Ruslan Salakhutdinov (eds.), *Proceedings of the 36th International Conference on Machine Learning*, volume 97, pp. 5301–5310. PMLR, 09–15 Jun 2019. URL https://proceedings.mlr.press/v97/rahaman19a.html.

David Salinas, Valentin Flunkert, Jan Gasthaus, and Tim Januschowski. Deepar: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, 36(3):1181–1191, 2020. doi: 10.1016/j.ijforecast.2019.07.001. URL https://www.sciencedirect.com/science/article/pii/S0169207019301888.

Ozan Sener and Vladlen Koltun. Multi-task learning as multi-objective optimization. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/file/432aca3a1e345e339f35a30c8f65edce-Paper.pdf.

Hao Wang, Lichen Pan, Yuan Shen, Zhichao Chen, Degui Yang, Yifei Yang, Sen Zhang, Xinggao Liu, Haoxuan Li, and Dacheng Tao. FreDF: Learning to forecast in the frequency domain. In Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu (eds.), *International Conference on Learning Representations*, volume 2025, pp. 6893–6922, 2025a. URL https://proceedings.iclr.cc/paper_files/paper/2025/file/1457fb1e5d72cdc4ecd88bc10f916095-Paper-Conference.pdf.

Hao Wang, Licheng Pan, Zhichao Chen, Xu Chen, Qingyang Dai, Lei Wang, Haoxuan Li, and Zhouchen Lin. Time-o1: Time-series forecasting needs transformed label alignment. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 8660–8690. Curran Associates, Inc., 2025b. doi: 10.52202/085713-0294. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/0cd62dea69635f4c5b569848267fe5a8-Paper-Conference.pdf.

Hao Wang, Licheng Pan, Yuan Lu, Zi Ciu Chan, Tianqiao Liu, Shuting He, Zhixuan Chu, Qingsong Wen, Haoxuan Li, and Zhouchen Lin. Quadratic direct forecast for training multi-step time-series forecast models. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=vpO8n9AqEG.

Shiyu Wang, Haixu Wu, Xiaoming Shi, Tengge Hu, Huakun Luo, Lintao Ma, James Zhang, and JUN ZHOU. Timemixer: Decomposable multiscale mixing for time series forecasting. In B. Kim, Y. Yue, S. Chaudhuri, K. Fragkiadaki, M. Khan, and Y. Sun (eds.), *International Conference on Learning Representations*, volume 2024, pp. 38626–38652, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/file/a7ac8a21e5a27e7ab31a5f42a0117bdb-Paper-Conference.pdf.

Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 22419–22430. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/bcc0d400288793e8bdcd7c19a8ac0c2b-Paper.pdf.

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. URL https://openreview.net/forum?id=ju_Uqw384Oq.

Kun Yi, Qi Zhang, Wei Fan, Shoujin Wang, Pengyang Wang, Hui He, Ning An, Defu Lian, Longbing Cao, and Zhendong Niu. Frequency-domain mlps are more effective learners in time series forecasting. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 76656–76679. Curran Associates, Inc., 2023. doi: 10.52202/075280-3349. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/f1d16af76939f476b5f040fd1398c0a3-Paper-Conference.pdf.

Tianhe Yu, Saurabh Kumar, Abhishek Gupta, Sergey Levine, Karol Hausman, and Chelsea Finn. Gradient surgery for multi-task learning. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 5824–5836. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/3fe78a8acf5fda99de95303940a2420c-Paper.pdf.

Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are transformers effective for time series forecasting? In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 11121–11128, 2023. doi: 10.1609/aaai.v37i9.26317.

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*, volume 35, pp. 11106–11115, 2021. doi: 10.1609/aaai.v35i12.17325.

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 2022. URL https://proceedings.mlr.press/v162/zhou22g.html.

A Related Work

A.1 Deep Models for Time-Series Forecasting

Neural forecasting models use different architectures to capture temporal dependencies. Recurrent and convolutional models learn sequential patterns (Salinas et al., 2020; Bai et al., 2018). Transformer-based models use attention, temporal patches, or variable-wise representations (Zhou et al., 2021; Wu et al., 2021; Nie et al., 2023; Liu et al., 2024). Other models use basis expansions, linear mappings, or multiscale mixing (Oreshkin et al., 2020; Challu et al., 2023; Zeng et al., 2023; Wang et al., 2024). These approaches mainly change how a model represents time series. DySAC instead changes how forecasting errors guide training. It operates on residuals and parameter gradients without changing the forecasting architecture or inference procedure.

A.2 Temporal Representations and Forecasting Objectives

Frequency-aware methods represent or supervise different components of a time series. FEDformer and FreTS incorporate frequency-domain operations into forecasting models (Zhou et al., 2022; Yi et al., 2023). FreDF supervises predictions in the frequency domain (Wang et al., 2025a). Other objectives emphasize transformation invariance, patch-level statistics, or transformed labels (Lee et al., 2023; Kudrat et al., 2025; Wang et al., 2025b). DBLoss uses decomposition-based supervision (Qiu et al., 2025), and QDF uses a quadratic-form multi-step objective (Wang et al., 2026). These methods generally combine their supervision into a scalar training loss. DySAC instead retains frequency bands as separate prediction objectives. It coordinates their training to account for interactions between different components of the same forecast.

A.3 Multi-Objective Optimization and Gradient-Based Coordination

Multi-task optimization coordinates losses that share model parameters. MGDA, PCGrad, and CAGrad use objective gradients to construct shared updates (Sener & Koltun, 2018; Yu et al., 2020; Liu et al., 2021). Nash-MTL formulates this combination as bargaining (Navon et al., 2022). Adaptive weighting instead uses uncertainty, loss descent rates, or gradient norms (Cipolla et al., 2018; Liu et al., 2019; Chen et al., 2018). These scalar statistics do not encode gradient angles. In time-series forecasting, different frequency bands may favor incompatible changes to the same model. DySAC constructs band objectives from the forecasting residual to expose these interactions. It adopts the ConFIG direction construction unchanged (Liu et al., 2025) (Section 2.2). Its contribution is the spectral forecasting formulation and evaluation on the same bands, not a new solver. This evaluation checks which components of the predicted time series improve.

B Gradient Coordination: Derivations and Examples

B.1 Equal-angle property and its qualifications

This appendix expands the equal-angle discussion of Section 2.2. Substituting $\mathbf{U}\mathbf{U}^\top\boldsymbol{\alpha} = \mathbf{1} - \lambda_r\boldsymbol{\alpha}$ into the definition of $\mathbf{d}$ in equation 6 gives

$$\mathbf{U}\mathbf{d} = \frac{\mathbf{1} - \lambda_r\boldsymbol{\alpha}}{\|\mathbf{U}^\top\boldsymbol{\alpha}\|_2 + \varepsilon}, \tag{10}$$

For the exact finite-ε construction, set $\mathbf{v} = \mathbf{U}^\top\boldsymbol{\alpha}$. Then

$$\langle \mathbf{g}_k, \mathbf{d}\rangle = \frac{(\|\mathbf{g}_k\|_2 + \varepsilon)(1 - \lambda_r\alpha_k)}{\|\mathbf{v}\|_2 + \varepsilon}. \tag{11}$$

Thus the regularized direction is a strict common descent direction if and only if $1 - \lambda_r\alpha_k > 0$ for every k. A small positive λ_r alone does not ensure this condition.

For ideal normalization ($\varepsilon = 0$ and nonzero gradients), equal positive cosines are recovered as $\lambda_r \to 0$ provided that $\mathbf{U}\mathbf{v} = \mathbf{1}$ is consistent; full row rank is a sufficient condition. In that case the regularized solution converges to the minimum-norm exact solution and $\mathbf{U}\mathbf{d} = \mathbf{1}/\|\mathbf{v}\|_2$ in the limit. If the system is inconsistent, the least-squares limit need not have equal or positive projections. With

finite ε, neither the rows of $\mathbf{U}$ nor $\mathbf{d}$ are exactly unit vectors, so their inner products must not be identified with cosines.

These identities characterize the direction handed to the optimizer. Adam's momentum and coordinate-wise preconditioning generally change the realized step direction, so a direction-level guarantee does not automatically apply to that step. Nor does a single-step property imply balanced final performance, which is evaluated empirically in Section 3.3. The configured value is $\lambda_r = 10^{-8}$ (Appendix D); the actual deviation from the limit form depends on the solved coefficients and numerical conditioning, not on λ_r alone.

B.2 Conditional descent and projected-residual convergence

We analyze the direct update of equation 8, retaining its finite λ_r and ε. This is a specialization of the smooth descent argument to the adopted ConFIG construction (Liu et al., 2025), not a new coordination solver.

Assumptions and scope. Let $\mathcal{L}_1, \ldots, \mathcal{L}_K$ be fixed deterministic objectives with ℓ_k-Lipschitz continuous gradients, $\ell_k > 0$, on a convex region containing the iterates and update segments. All gradients and directions are evaluated exactly, and the updates use neither Adam nor the mean-gradient fallback. The magnitude-L1 loss in equation 3 is not differentiable at zero residual, so this smoothness assumption is not automatic for the experiments. The single-step result applies on smooth regions satisfying the assumption. A smooth surrogate, obtained by replacing $|R|$ with $\sqrt{|R|^2 + \delta^2}$ for $\delta > 0$, can support the repeated-step analysis under suitable backbone regularity and boundedness on the region; this surrogate is not used in our reported experiments.

Proposition 1 (conditional common descent for a direct step). Let $\ell_{\max} = \max_{1 \le k \le K} \ell_k$. At an iterate θ, let $a_k = \langle \mathbf{g}_k, \mathbf{d} \rangle$, $m = \min_k a_k > 0$, and $\beta = \eta q$. For the direct update with this β, and more generally for any $\beta > 0$ whose update segment lies in the smooth region,

$$\mathcal{L}_k(\theta - \beta \mathbf{d}) \le \mathcal{L}_k(\theta) - \beta a_k + \frac{\ell_k}{2} \beta^2 \|\mathbf{d}\|_2^2 \le \mathcal{L}_k(\theta) - \beta m + \frac{\ell_{\max}}{2} \beta^2. \tag{12}$$

In particular, $0 < \beta \le m / \ell_{\max}$ ensures strict decrease of every band objective.

Proof. Fix the iterate θ and any band $k \in \{1, \ldots, K\}$. All gradients, coordination coefficients, and the direction $\mathbf{d}$ in this proof are computed at θ and held fixed along the update segment. We first establish the properties of the update, then bound the change of each objective along that segment.

Step 1: positivity of the scale and control of the direction norm. By the common descent assumption, $a_j \ge m > 0$ for every band j. The scale-recovery rule therefore gives

$$q = \sum_{j=1}^{K} a_j \ge K m > 0. \tag{13}$$

Thus $\beta = \eta q$ is positive whenever $\eta > 0$. Writing $\mathbf{v} = \mathbf{U}^\top \boldsymbol{\alpha}$, the actual finite-$\varepsilon$ normalization in equation 6 yields

$$\|\mathbf{d}\|_2 = \frac{\|\mathbf{v}\|_2}{\|\mathbf{v}\|_2 + \varepsilon} \le 1. \tag{14}$$

Moreover, $m > 0$ excludes $\mathbf{d} = \mathbf{0}$, since a zero direction would give $a_j = 0$ for all j. We use only the norm bound above, not an exact unit-norm or equal-angle assumption.

Step 2: an integral bound on the change of a band objective. Let $\theta^+ = \theta - \beta \mathbf{d}$ and parameterize the segment by $\gamma(s) = \theta - s\beta \mathbf{d}$ for $s \in [0, 1]$. By assumption, this entire segment lies in the region where $\nabla \mathcal{L}_k$ is ℓ_k-Lipschitz. The chain rule and the fundamental theorem of calculus give

$$\begin{aligned} \mathcal{L}_k(\theta^+) - \mathcal{L}_k(\theta) &= -\beta \int_0^1 \langle \nabla \mathcal{L}_k(\gamma(s)), \mathbf{d} \rangle \, ds \\ &= -\beta a_k - \beta \int_0^1 \langle \nabla \mathcal{L}_k(\gamma(s)) - \nabla \mathcal{L}_k(\theta), \mathbf{d} \rangle \, ds. \end{aligned} \tag{15}$$

For each $s \in [0,1]$, Cauchy–Schwarz and Lipschitz continuity imply

$$\begin{aligned} -\langle \nabla \mathcal{L}_k(\gamma(s)) - \nabla \mathcal{L}_k(\theta), \mathbf{d} \rangle &\le \|\nabla \mathcal{L}_k(\gamma(s)) - \nabla \mathcal{L}_k(\theta)\|_2 \|\mathbf{d}\|_2 \\ &\le \ell_k \|\gamma(s) - \theta\|_2 \|\mathbf{d}\|_2 \\ &= \ell_k s\beta \|\mathbf{d}\|_2^2. \end{aligned} \tag{16}$$

Substituting this bound into equation 15 and using $\int_0^1 s\, ds = 1/2$ yields

$$\mathcal{L}_k(\theta^+) - \mathcal{L}_k(\theta) \le -\beta a_k + \frac{\ell_k \beta^2}{2} \|\mathbf{d}\|_2^2. \tag{17}$$

This proves the first inequality in equation 12 without requiring second derivatives or convexity of the objectives.

Step 3: a uniform bound for all bands. Since $\beta > 0$, $a_k \ge m$, $\ell_k \le \ell_{\max}$, and $\|\mathbf{d}\|_2^2 \le 1$, equation equation 17 gives

$$\mathcal{L}_k(\theta^+) - \mathcal{L}_k(\theta) \le -\beta m + \frac{\ell_{\max} \beta^2}{2}. \tag{18}$$

This is the second inequality in equation 12. The right-hand side is independent of k, so the same step-size condition can be imposed simultaneously on every band.

Step 4: sufficient step size and strict decrease. If $0 < \beta \le m/\ell_{\max}$, then $\ell_{\max}\beta^2/2 \le \beta m/2$. Consequently,

$$\mathcal{L}_k(\theta^+) - \mathcal{L}_k(\theta) \le -\beta m + \frac{\beta m}{2} = -\frac{\beta m}{2} < 0. \tag{19}$$

Because k was arbitrary, this strict decrease holds for every spectral objective. Finally, substituting $\beta = \eta q$ recovers the direct-step condition. Since $q > 0$, the sufficient condition can equivalently be written as $0 < \eta \le m/(\ell_{\max} q)$, subject to the update segment remaining in the smooth region. The positivity of m is an explicit hypothesis; the regularized linear solve alone does not guarantee it. □

Proposition 2 (deterministic projected-residual bound). Let $F(\theta) = \sum_k \mathcal{L}_k(\theta)$ be bounded below by $F_{\inf}$ along the iterates, let $\ell_\Sigma = \sum_k \ell_k$, and use a constant learning rate $0 < \eta \le 1/\ell_\Sigma$. For the direct updates $\theta_{t+1} = \theta_t - \eta q_t \mathbf{d}_t$, define $\Delta_F = F(\theta_0) - F_{\inf}$. For every integer $T \ge 1$,

$$\sum_{t=0}^{T-1} q_t^2 \le \frac{2\Delta_F}{\eta}, \qquad \min_{0 \le t < T} |q_t| \le \sqrt{\frac{2\Delta_F}{\eta T}}. \tag{20}$$

Consequently, for an infinite sequence satisfying the assumptions, $q_t \to 0$ and $\|\mathbf{g}_{\mathrm{final},t}\|_2 \to 0$. This result does not require every band projection to be positive.

Proof. The gradient of F is ℓ_Σ-Lipschitz and, by the scale definition, $\langle \nabla F(\theta_t), \mathbf{d}_t \rangle = q_t$. The descent lemma therefore gives

$$\begin{aligned} F(\theta_{t+1}) &\le F(\theta_t) - \eta q_t^2 + \frac{\ell_\Sigma \eta^2}{2} q_t^2 \|\mathbf{d}_t\|_2^2 \\ &\le F(\theta_t) - \frac{\eta}{2} q_t^2. \end{aligned} \tag{21}$$

Summing over t and using $F(\theta_T) \ge F_{\inf}$ proves the sum bound. The minimum squared residual is at most its average, which proves the rate. Summability gives $q_t \to 0$, and $\|q_t \mathbf{d}_t\|_2 \le |q_t|$ proves the last claim. □

Combined interpretation. Together, the two results give complementary conditional guarantees. Proposition 1 shows that a sufficiently small direct step decreases every band when all band projections are positive. Proposition 2 shows that, under its deterministic assumptions, the aggregate projected residual has a best-iterate $O(T^{-1/2})$ bound and tends to zero. For any tolerance $\epsilon_{\mathrm{tol}} > 0$, distinct from the normalization constant ε, it guarantees an iterate with $|q_t| \le \epsilon_{\mathrm{tol}}$ whenever $T \ge 2\Delta_F/(\eta \epsilon_{\mathrm{tol}}^2)$. These results concern the prescribed direct update, not the experimental

Adam training, and do not establish parameter convergence or Pareto optimality. In particular, the Pareto-stationarity measure

$$\chi(\theta) = \max_{\|\mathbf{z}\|_2 \le 1} \min_k \langle \nabla \mathcal{L}_k(\theta), \mathbf{z} \rangle \tag{22}$$

tests all feasible directions, not only $\mathbf{d}$. Thus, a small m or $|q|$ does not establish a small $\chi(\theta)$ without an additional direction-quality bound, and we do not claim an ϵ-Pareto-stationarity rate. Mini-batch changes, Adam, clipping, and fallback steps are outside the repeated-step analysis.

B.3 Two configurations separating geometric from scalar weighting

Consider two three-band configurations with the same per-objective scalar inputs: equal band losses, equal gradient norms, and equal loss histories. Assume the scalar rule is permutation-equivariant, so identical scalar inputs receive identical weights. In configuration A, let

$$\mathbf{g}_1 = \mathbf{g}_2 = \mathbf{e}_1, \qquad \mathbf{g}_3 = \mathbf{e}_2,$$

whereas in configuration B let

$$\mathbf{g}_1 = \mathbf{e}_1, \qquad \mathbf{g}_2 = \mathbf{e}_2, \qquad \mathbf{g}_3 = \mathbf{e}_3.$$

The scalar rule receives the same inputs in both configurations and therefore cannot distinguish them. In the ideal normalization limit, however, configuration A gives

$$\mathbf{U} = \begin{bmatrix} \mathbf{u}_1^\top \\ \mathbf{u}_2^\top \\ \mathbf{u}_3^\top \end{bmatrix}, \qquad \mathbf{U}\mathbf{U}^\top = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad \boldsymbol{\alpha} \propto \left(\tfrac{1}{2}, \tfrac{1}{2}, 1\right), \tag{23}$$

while configuration B gives $\mathbf{U}\mathbf{U}^\top = \mathbf{I}_3$ and $\boldsymbol{\alpha} \propto (1, 1, 1)$. Thus the geometric solution changes when the gradient angles change, even though all scalar inputs are unchanged. In configuration A, the duplicated direction receives half the coefficient of the independent direction; in configuration B, all three directions receive the same coefficient. This proves the narrower claim used in the main text: no permutation-equivariant rule based only on the listed scalar inputs can reproduce the geometric coefficients for both configurations. The example is not a claim that every scalar weighting rule must be permutation-equivariant; it isolates the information that scalar-only inputs omit.

C Task-Wise Spectral Evaluation Protocol

The final checkpoint for each run is selected using the same validation criterion as its corresponding forecasting experiment and then evaluated on the test set. All objectives within the same dataset, backbone, sequence configuration, prediction horizon, and seed use exactly the dyadic partition defined in Section 2.1. Thus, training and evaluation share the same task identities, while the number of tasks may vary with the prediction horizon.

Task-wise relative improvement. For training objective m, random seed s, and task k, let $E_{k,s}^{(m)}$ denote the final test-set spectral error associated with $\mathcal{B}_k$, computed using the magnitude-L_1 spectral residual in equation 3. Each run is compared with the MSE model trained using the same dataset, backbone, sequence lengths, prediction horizon, and seed. The relative improvement used in Section 2.3 is

$$I_{k,s}^{(m)} = \frac{E_{k,s}^{(\mathrm{MSE})} - E_{k,s}^{(m)}}{\left| E_{k,s}^{(\mathrm{MSE})} \right| + \varepsilon},$$

where $\varepsilon > 0$ stabilizes the denominator. A positive value indicates lower task error than matched MSE training. A negative value indicates higher error, and the MSE reference is zero by construction. Matching the nominal seed does not guarantee identical random-number consumption across objectives.

Multi-seed aggregation. For each dataset, backbone, forecasting configuration, objective, and spectral task, let $\mathcal{S}_k^{(m)}$ contain the seeds with a valid task result. We report the mean and population standard deviation from Eq. equation 9. Missing or non-finite task results are excluded rather than assigned zero. The standard deviation uses the population convention (`ddof=0`).

Task-wise objective ranking. Within each fixed dataset, backbone, input configuration, prediction horizon, and task, objectives are ranked by $\mu_k^{(m)}$. Larger mean improvement receives a smaller rank, with rank 1 denoting the best objective. Exact ties receive their average rank. Only objectives with a valid result for that task participate, and ranking is applied after seed aggregation.

Average task rank. Within a fixed dataset and backbone, for objective m, let $\mathcal{K}^{(m)}$ denote the valid task instances and $R_k^{(m)}$ its rank on task k. Its average task rank is computed as described in Section 2.3. A lower average task rank indicates more consistently competitive performance. The task count $|\mathcal{K}^{(m)}|$ is the actual denominator; missing tasks enter neither the numerator nor the denominator. This rank is an empirical measure of cross-frequency balance, not a certificate of Pareto optimality.

Setting-level aggregation. For the overall spectral summary, we first compute the average task rank within each dataset–backbone setting and then average the 28 setting-level values with equal weight. This differs from pooling all 119 task ranks because settings contain different numbers of tasks. Mean task improvement instead gives equal weight to each valid task instance after seed aggregation.

Second-level ranking. Within each fixed dataset–backbone setting, objectives are ranked by their average task rank in ascending order. Exact ties receive their average rank. If multiple configurations are present, first-level rankings remain separate and their valid task ranks are subsequently averaged within the same dataset, backbone, and objective. Each configuration–task instance contributes once. The "Rank-1 settings" column of Table 20 counts settings in which an objective attains the minimum average task rank, including tied minima.

D Additional Experimental Details

D.1 DySAC Training Pseudocode

The following pseudocode summarizes one DySAC training step.

Algorithm 1 One DySAC training step.

Input: f_θ, $(\mathbf{X}, \mathbf{Y})$, D_{base}, λ_r, ε, η

1: $\hat{\mathbf{Y}} \leftarrow f_\theta(\mathbf{X})$; $\mathbf{R} \leftarrow \mathrm{rFFT}(\hat{\mathbf{Y}}) - \mathrm{rFFT}(\mathbf{Y})$; construct $\{\mathcal{B}_k\}_{k=1}^K$ by equation 2
2: **for** $k = 1, \ldots, K$ **do**
3: Compute $\mathcal{L}_k$ by equation 3; $\mathbf{g}_k \leftarrow \nabla_\theta \mathcal{L}_k$; $\mathbf{u}_k \leftarrow \mathbf{g}_k / (\|\mathbf{g}_k\|_2 + \varepsilon)$
4: **end for**
5: $\mathbf{U} \leftarrow [\mathbf{u}_1, \ldots, \mathbf{u}_K]^\top$; compute $(\boldsymbol{\alpha}, \mathbf{d})$ by equation 6
6: $q \leftarrow \sum_{k=1}^K \langle \mathbf{g}_k, \mathbf{d} \rangle$; $\theta \leftarrow \theta - \eta q \mathbf{d}$

D.2 Dataset Usage and Evaluation Protocol

The experiments use seven datasets: ETTh1, ETTh2, Electricity, Weather, Traffic, CIF-2016, and Fred-md. The main forecasting comparison and the task-wise spectral evaluation both cover all four backbones (DLinear, FreTS, iTransformer, and TimeMixer) on all seven datasets. The task-wise spectral evaluation uses DLinear, FreTS, iTransformer, and TimeMixer on ETTh1, ETTh2, Weather, Electricity, Traffic, CIF-2016, and Fred-md; each reported dataset–backbone setting uses one fixed sequence and forecasting configuration, and task-wise ranks must be computed separately per setting before any higher-level summary. The component and spectral-loss ablations use FreTS on Electricity and Weather, and the frequency-partition ablation uses FreTS and TimeMixer on the same datasets. For these ablations, the sequence–label–prediction configuration is 96–48–192. All variants within a fixed dataset–backbone setting use the same data split, forecasting configuration, optimizer settings, and random seeds. Within each backbone, all partition variants use an identical training protocol; backbone-specific training schedules follow the corresponding experimental setup, so the partition comparison is conducted within rather than across backbones. Results are aggregated over three seeds; spectral-task summaries use population standard deviation, and heterogeneous-metric summaries use sample standard deviation.

We follow the data splits and preprocessing protocols adopted in the corresponding backbone studies. Within each setting, all objectives use the same backbone configuration, data split, prediction horizon, and training protocol. Hyperparameters introduced by DySAC are selected on the validation set. Table 7 lists the sequence–label–prediction configurations used in both the task-wise spectral evaluation and aggregate forecasting comparison.

Aggregate forecasting configuration. The MSE/MAE comparison in Table 1 and Appendix E uses 192–96–96 for DLinear, iTransformer, and TimeMixer on ETTh1, ETTh2, Weather, Electricity, and Traffic, and 96–48–48 for FreTS; it uses 24–12–12 on CIF-2016 and 96–48–12 on Fred-md for all four backbones. These configurations agree with the task-wise spectral evaluation. Matching configurations alone, however, do not establish that aggregate and spectral ranks describe the same selected checkpoints. Seed-level MSE/MAE, checkpoint correspondence, and the MSE/MAE standard-deviation estimator are not established here. These comparisons therefore do not support paired significance tests or causal attribution of differences between evaluations.

Table 7: Sequence–label–prediction configurations used in every task-wise spectral evaluation setting. The same configuration is used for all objectives within a setting.

Setting	Backbone	Seq–Lab–Pred
ETTh1, ETTh2, Weather, Electricity, Traffic	DLinear, iTransformer, TimeMixer	192–96–96
ETTh1, ETTh2, Weather, Electricity, Traffic	FreTS	96–48–48
CIF-2016	DLinear, FreTS, iTransformer, TimeMixer	24–12–12
Fred-md	DLinear, FreTS, iTransformer, TimeMixer	96–48–12

ETTh1 and ETTh2 use the standard chronological split of 12, 4, and 4 months for training, validation, and testing. All other datasets are split chronologically in the ratio 7:1:2. Every dataset is standardized per variable using statistics computed on the training split only; test metrics are reported in this standardized space, which is the convention of the long-term forecasting literature and of all compared objectives.

D.3 Implementation and Reproducibility Details

All baselines are reproduced following their original configurations. Experiments are implemented in PyTorch and executed on a T-Head Zhenwu PPU cluster. Each computing node contains two Zhenwu 810E PPU accelerators, 16 CPU cores, and 32 GB of system memory. Results in the reported tables are averaged over independent runs; the ablation experiments use three random seeds.

Training protocol. The comparison runners use Adam with defaults of learning rate 5×10^{-4}, no weight decay, no learning-rate schedule, batch size 32, and a 20-epoch budget. These are script defaults rather than a verified uniform budget for every reported run: FreTS/Fred-md includes runs lasting 30 epochs. The heterogeneous TimeMixer/Fred-md evaluation uses a revised epoch setting, discussed in Appendix F. Global gradient-norm clipping at 1.0 is applied after the DySAC update $\mathbf{g}_{\text{final}}$ has been formed. Model selection uses early stopping on validation MSE with a patience of five epochs; the checkpoint with the lowest validation MSE is evaluated on the test set. Every dataset–backbone–objective combination in the spectral evaluation is evaluated over seeds $\{2026, 2027, 2028\}$; the Fred-md heterogeneous evaluation uses the same seed set. The spectral summaries use population standard deviation (`ddof=0`), whereas the heterogeneous-metric tables use sample standard deviation (`ddof=1`).

DySAC hyperparameters. The dyadic partition of equation 2 uses $D_{\text{base}} = 2$ and the `backward` rFFT normalization. The band objectives follow the magnitude-L1 form of equation 3. In the coordination step of equation 6 we set $\lambda_r = 10^{-8}$ and $\varepsilon = 10^{-12}$; both are fixed throughout training and were not tuned. The small value of λ_r alone does not guarantee a negligible equal-angle error: the perturbation also depends on the Gram matrix and the solved coefficients, as shown in equation 10. If the linear system in equation 6 fails to produce a finite direction with positive scale q, the step falls back to the plain mean of the band gradients; its activation frequency is not reported. Table 8 lists the number of dyadic bands K induced by each prediction horizon.

Table 8: Number of dyadic bands and their widths for each prediction horizon under $D_{\text{base}} = 2$. The task-wise comparison uses horizons 12, 48, and 96; horizon 192 applies to the separate ablations. The last band absorbs the remainder of the non-redundant spectrum.

Prediction length L	$M = \lfloor L/2 \rfloor + 1$	K	Band widths $D_1, \ldots, D_K$
12	7	3	2, 4, 1
48	25	4	2, 4, 8, 11
96	49	5	2, 4, 8, 16, 19
192	97	6	2, 4, 8, 16, 32, 35

E Full Aggregated Forecasting Results

The following tables report the complete results for all four backbones, with OURS in the final row of each block. Blue boldface marks the minimum per metric column and red the second-lowest distinct mean; ties at the displayed precision share the same mark. Table 1 reports the corresponding dataset-macro averages in the main text.

Table 9: Complete forecasting results for FreTS. Values are mean with standard deviation over three seeds; lower is better. Blue boldface marks the minimum mean in each metric column, and red marks the second-lowest distinct mean; ties at the displayed precision share the same mark.

Objective	ETTh1		ETTh2		Weather		Electricity		Traffic		Fred-md		CIF-2016	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Origin	0.3641 (0.0032)	0.3937 (0.0059)	0.2413 (0.0038)	0.3170 (0.0018)	0.1347 (0.0013)	0.1872 (0.0018)	0.1759 (0.0005)	0.2652 (0.0014)	0.5228 (0.0022)	0.3386 (0.0022)	51.2937 (23.6487)	1.1454 (0.2638)	3.1777 (1.0229)	1.1260 (0.2174)
+DBLoss	0.3544 (0.0013)	0.3854 (0.0012)	0.2336 (0.0049)	0.3070 (0.0060)	0.1405 (0.0007)	0.1806 (0.0032)	0.1776 (0.0008)	0.2584 (0.0007)	**0.5121 (0.0034)**	**0.3285 (0.0018)**	17.7104 (0.4547)	**0.7404 (0.0285)**	2.6119 (0.5239)	1.0184 (0.1160)
+FreDF	0.3568 (0.0072)	0.3898 (0.0084)	0.2333 (0.0072)	0.3073 (0.0057)	0.1338 (0.0009)	0.1787 (0.0012)	0.1758 (0.0003)	0.2601 (0.0009)	0.5155 (0.0018)	0.3340 (0.0021)	29.1460 (8.3139)	0.9450 (0.0458)	2.7507 (0.8642)	1.0581 (0.1899)
+PS-Loss	**0.3526 (0.0018)**	**0.3841 (0.0028)**	0.2338 (0.0025)	0.3055 (0.0020)	**0.1336 (0.0015)**	0.1755 (0.0034)	**0.1747 (0.0007)**	0.2589 (0.0005)	0.5200 (0.0015)	0.3320 (0.0015)	37.2486 (12.0543)	0.9763 (0.1271)	2.7092 (0.9930)	1.0279 (0.1902)
+QDF	0.3660 (0.0039)	0.3951 (0.0059)	0.2400 (0.0070)	0.3145 (0.0029)	0.1367 (0.0037)	0.1933 (0.0113)	0.1751 (0.0008)	0.2646 (0.0016)	0.5192 (0.0027)	0.3335 (0.0022)	20.7611 (6.6577)	0.9090 (0.0860)	3.2247 (1.1174)	1.1266 (0.2222)
+TILDE-Q	0.3673 (0.0138)	0.3908 (0.0133)	**0.2305 (0.0021)**	**0.3002 (0.0019)**	0.1348 (0.0008)	0.1761 (0.0025)	0.1876 (0.0011)	0.2730 (0.0009)	0.5294 (0.0059)	0.3405 (0.0027)	**17.6153 (0.8207)**	0.8030 (0.0178)	**2.5870 (0.8537)**	**0.9933 (0.1966)**
+Time-o1	0.3581 (0.0043)	0.3883 (0.0028)	0.2515 (0.0035)	0.3206 (0.0017)	0.1354 (0.0014)	0.1872 (0.0037)	0.1764 (0.0009)	0.2646 (0.0012)	0.5210 (0.0068)	0.3380 (0.0003)	33.7962 (14.0455)	1.0431 (0.1570)	3.2128 (0.5893)	1.1198 (0.0551)
OURS	0.3571 (0.0096)	0.3898 (0.0092)	0.2331 (0.0017)	0.3038 (0.0028)	0.1344 (0.0015)	**0.1754 (0.0034)**	0.1777 (0.0005)	**0.2583 (0.0001)**	**0.5121 (0.0025)**	0.3316 (0.0020)	21.9815 (0.6379)	0.8481 (0.0291)	2.7340 (0.7854)	1.0248 (0.1671)

Table 10: Complete forecasting results for DLinear. Values are mean with standard deviation over three seeds; lower is better. Blue boldface marks the minimum mean in each metric column, and red marks the second-lowest distinct mean; ties at the displayed precision share the same mark.

Objective	ETTh1		ETTh2		Weather		Electricity		Traffic		Fred-md		CIF-2016	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Origin	0.3870 (0.0025)	0.4074 (0.0019)	0.2907 (0.0035)	0.3512 (0.0042)	0.1850 (0.0011)	0.2485 (0.0003)	0.1522 (0.0001)	0.2473 (0.0003)	**0.4514 (0.0006)**	0.3040 (0.0015)	15.9263 (0.4846)	0.6717 (0.0073)	2.1297 (0.0497)	0.9711 (0.0146)
+DBLoss	0.3768 (0.0003)	**0.3925 (0.0007)**	0.2856 (0.0005)	0.3382 (0.0003)	0.1875 (0.0024)	0.2325 (0.0042)	0.1526 (0.0002)	**0.2460 (0.0003)**	0.4613 (0.0009)	0.2985 (0.0041)	13.1901 (0.0319)	0.5902 (0.0037)	1.8884 (0.1252)	**0.8830 (0.0436)**
+FreDF	0.3786 (0.0008)	0.3975 (0.0012)	0.2840 (0.0009)	0.3427 (0.0017)	0.1847 (0.0030)	0.2434 (0.0041)	**0.1521 (0.0001)**	0.2469 (0.0004)	0.4541 (0.0011)	0.3010 (0.0040)	13.9153 (0.3405)	0.6143 (0.0075)	2.0451 (0.0710)	0.9439 (0.0229)
+PS-Loss	**0.3757 (0.0008)**	0.3935 (0.0007)	0.2845 (0.0010)	0.3398 (0.0013)	0.1841 (0.0016)	**0.2287 (0.0037)**	0.1525 (0.0002)	0.2472 (0.0001)	0.4574 (0.0012)	**0.2980 (0.0038)**	12.2906 (0.1489)	0.5744 (0.0016)	1.9186 (0.0895)	0.8960 (0.0279)
+QDF	0.3840 (0.0018)	0.4029 (0.0013)	0.2874 (0.0019)	0.3472 (0.0018)	0.1839 (0.0015)	0.2439 (0.0035)	0.1522 (0.0001)	0.2475 (0.0009)	**0.4514 (0.0006)**	0.3037 (0.0013)	14.7672 (0.5820)	0.6368 (0.0138)	2.0654 (0.0405)	0.9503 (0.0115)
+TILDE-Q	0.3822 (0.0003)	0.3976 (0.0004)	0.2838 (0.0002)	0.3368 (0.0003)	0.1833 (0.0007)	0.2292 (0.0010)	0.1599 (0.0001)	0.2572 (0.0004)	0.4735 (0.0005)	0.3173 (0.0004)	11.5610 (0.0355)	0.5607 (0.0012)	1.8944 (0.0681)	0.9003 (0.0232)
+Time-o1	0.3843 (0.0020)	0.4035 (0.0036)	0.2880 (0.0021)	0.3469 (0.0016)	0.1839 (0.0013)	0.2437 (0.0034)	0.1523 (0.0001)	0.2477 (0.0006)	0.4517 (0.0003)	0.3049 (0.0005)	16.1072 (0.4408)	0.6767 (0.0098)	2.1939 (0.1079)	0.9915 (0.0361)
OURS	0.3759 (0.0006)	0.3944 (0.0011)	**0.2833 (0.0004)**	**0.3363 (0.0003)**	**0.1825 (0.0020)**	0.2297 (0.0047)	0.1525 (0.0001)	0.2466 (0.0003)	0.4579 (0.0002)	0.2993 (0.0027)	**11.5566 (0.3302)**	**0.5561 (0.0049)**	**1.8881 (0.0974)**	0.8893 (0.0349)

Table 11: Complete forecasting results for iTransformer. Values are mean with standard deviation over three seeds; lower is better. Blue boldface marks the minimum mean in each metric column, and red marks the second-lowest distinct mean; ties at the displayed precision share the same mark.

Objective	ETTh1		ETTh2		Weather		Electricity		Traffic		Fred-md		CIF-2016	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Origin	0.3985 (0.0054)	0.4181 (0.0027)	0.3212 (0.0053)	0.3695 (0.0052)	0.1658 (0.0023)	0.2112 (0.0022)	0.1337 (0.0002)	0.2286 (0.0007)	0.3752 (0.0009)	0.2689 (0.0017)	18.9797 (3.0220)	0.7228 (0.0640)	1.2917 (0.0705)	0.6366 (0.0135)
+DBLoss	0.3841 (0.0055)	0.4035 (0.0031)	0.3028 (0.0060)	0.3534 (0.0065)	0.1621 (0.0029)	0.2017 (0.0010)	0.1308 (0.0008)	0.2234 (0.0010)	**0.3710 (0.0001)**	**0.2589 (0.0005)**	16.8172 (1.1970)	0.6737 (0.0145)	1.2095 (0.0285)	0.6187 (0.0106)
+FreDF	0.3895 (0.0050)	0.4096 (0.0020)	0.3162 (0.0045)	0.3652 (0.0045)	0.1630 (0.0010)	0.2068 (0.0017)	0.1315 (0.0004)	0.2255 (0.0008)	0.3730 (0.0032)	0.2633 (0.0016)	15.7488 (2.0061)	0.6826 (0.0332)	1.2307 (0.0448)	0.6257 (0.0084)
+PS-Loss	0.3834 (0.0008)	0.4047 (0.0011)	0.3042 (0.0028)	0.3527 (0.0039)	**0.1590 (0.0012)**	**0.1996 (0.0012)**	0.1313 (0.0005)	0.2252 (0.0009)	0.3741 (0.0019)	0.2691 (0.0019)	16.5966 (4.0851)	0.6840 (0.0600)	1.2427 (0.0504)	0.6224 (0.0111)
+QDF	0.4001 (0.0124)	0.4165 (0.0057)	0.3148 (0.0077)	0.3650 (0.0055)	0.1657 (0.0011)	0.2104 (0.0025)	0.1344 (0.0006)	0.2293 (0.0010)	0.3758 (0.0011)	0.2682 (0.0008)	**15.5336 (2.8579)**	0.6827 (0.0406)	1.2491 (0.0115)	0.6313 (0.0031)
+TILDE-Q	0.3937 (0.0010)	0.4103 (0.0020)	**0.3020 (0.0086)**	0.3526 (0.0090)	0.1614 (0.0026)	0.2048 (0.0034)	0.1408 (0.0006)	0.2388 (0.0012)	0.3936 (0.0006)	0.2842 (0.0016)	18.0611 (2.2837)	0.7074 (0.0346)	1.2120 (0.0207)	0.6283 (0.0028)
+Time-o1	0.4010 (0.0021)	0.4185 (0.0015)	0.3193 (0.0073)	0.3669 (0.0057)	0.1646 (0.0012)	0.2077 (0.0009)	0.1338 (0.0007)	0.2291 (0.0008)	0.3767 (0.0014)	0.2694 (0.0006)	16.2320 (1.7234)	0.6908 (0.0302)	1.2433 (0.0791)	0.6279 (0.0149)
OURS	**0.3804 (0.0027)**	**0.4027 (0.0015)**	**0.3020 (0.0110)**	**0.3498 (0.0039)**	0.1652 (0.0007)	0.2077 (0.0019)	**0.1300 (0.0001)**	**0.2227 (0.0001)**	0.3744 (0.0027)	0.2616 (0.0008)	15.6489 (2.8311)	**0.6673 (0.0443)**	**1.1599 (0.0347)**	**0.6017 (0.0116)**

Table 12: Complete forecasting results for TimeMixer. Values are mean with standard deviation over three seeds; lower is better. Blue boldface marks the minimum mean in each metric column, and red marks the second-lowest distinct mean; ties at the displayed precision share the same mark.

Objective	ETTh1		ETTh2		Weather		Electricity		Traffic		Fred-md		CIF-2016	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Origin	0.4098 (0.0102)	0.4276 (0.0088)	0.3639 (0.0144)	0.3889 (0.0077)	0.1541 (0.0013)	0.2043 (0.0011)	0.1478 (0.0027)	0.2501 (0.0030)	0.4704 (0.0003)	0.3226 (0.0023)	14.6251 (2.1810)	0.6851 (0.0444)	1.2296 (0.0130)	0.6301 (0.0075)
+DBLoss	0.3980 (0.0214)	0.4157 (0.0135)	0.3132 (0.0179)	0.3550 (0.0128)	**0.1477 (0.0018)**	**0.1927 (0.0022)**	0.1423 (0.0008)	0.2415 (0.0008)	0.4557 (0.0072)	**0.3024 (0.0035)**	12.3604 (0.9917)	0.6252 (0.0124)	1.1654 (0.0155)	0.6043 (0.0073)
+FreDF	0.3998 (0.0070)	0.4216 (0.0089)	0.3372 (0.0109)	0.3704 (0.0072)	0.1528 (0.0019)	0.2018 (0.0016)	0.1413 (0.0027)	0.2407 (0.0029)	0.4585 (0.0049)	0.3170 (0.0030)	12.6909 (0.7278)	0.6353 (0.0290)	1.1996 (0.0085)	0.6172 (0.0059)
+PS-Loss	0.3998 (0.0145)	0.4192 (0.0108)	0.3370 (0.0307)	0.3696 (0.0205)	0.1506 (0.0022)	0.1970 (0.0024)	0.1440 (0.0014)	0.2438 (0.0018)	0.4695 (0.0055)	0.3123 (0.0025)	11.6950 (2.3525)	0.6217 (0.0342)	1.1798 (0.0186)	0.6122 (0.0047)
+QDF	0.4063 (0.0117)	0.4255 (0.0103)	0.3427 (0.0105)	0.3775 (0.0030)	0.1568 (0.0026)	0.2058 (0.0027)	0.1491 (0.0038)	0.2508 (0.0041)	0.4701 (0.0043)	0.3204 (0.0032)	13.2520 (1.4288)	0.6586 (0.0293)	1.2157 (0.0098)	0.6245 (0.0058)
+TILDE-Q	0.4047 (0.0086)	0.4174 (0.0072)	0.3347 (0.0286)	0.3699 (0.0190)	0.1498 (0.0020)	0.1982 (0.0022)	0.1509 (0.0022)	0.2541 (0.0036)	0.4738 (0.0015)	0.3236 (0.0052)	11.7925 (1.1081)	0.6046 (0.0159)	1.1992 (0.0116)	0.6331 (0.0036)
+Time-o1	0.4128 (0.0103)	0.4286 (0.0085)	0.4003 (0.0375)	0.4096 (0.0211)	0.1538 (0.0037)	0.2046 (0.0035)	0.1469 (0.0034)	0.2500 (0.0019)	0.4716 (0.0052)	0.3201 (0.0065)	12.9584 (1.1374)	0.6620 (0.0182)	1.2530 (0.0603)	0.6422 (0.0266)
OURS	**0.3883 (0.0079)**	**0.4128 (0.0069)**	**0.3085 (0.0040)**	**0.3538 (0.0002)**	0.1550 (0.0020)	0.2040 (0.0020)	**0.1388 (0.0019)**	**0.2377 (0.0019)**	**0.4548 (0.0021)**	0.3196 (0.0029)	**10.7985 (0.5922)**	**0.5960 (0.0078)**	**1.1462 (0.0221)**	**0.5953 (0.0080)**

E.1 Where DySAC Is Weakest in the Aggregated Results

The four forecasting tables place DySAC first in 29 of the 56 backbone–dataset–metric columns: 3/14 for FreTS, 6/14 for DLinear, 9/14 for iTransformer, and 11/14 for TimeMixer. These counts include two rounded ties, with DBLoss on FreTS/Traffic MSE and TILDE-Q on iTransformer/ETTh2 MSE. Using average ranks for ties and equal weight for the 56 columns, DySAC has the lowest mean rank (2.33), followed by DBLoss (2.87) and PS-Loss (3.25). Its per-backbone mean ranks are 3.00, 2.32, 2.07, and 1.93 for FreTS, DLinear, iTransformer, and TimeMixer, respectively; each is lowest among the eight objectives for that backbone. On FreTS, the margin over DBLoss (3.04) is small, and DySAC has fewer first places than TILDE-Q (five) and PS-Loss (four). DySAC improves over MSE in 52 columns, including all 28 MAE columns. The four exceptions are MSE regressions, not a lack of improvement on every metric of the affected setting.

Regressions relative to MSE training. The four MSE increases are DLinear/Electricity (0.1525 versus 0.1522), DLinear/Traffic (0.4579 versus 0.4514), FreTS/Electricity (0.1777 versus 0.1759), and TimeMixer/Weather (0.1550 versus 0.1541). Their respective MAEs decrease from 0.2473 to 0.2466, 0.3040 to 0.2993, 0.2652 to 0.2583, and 0.2043 to 0.2040. Thus the reported means exhibit an MSE–MAE trade-off rather than uniform degradation. MSE weights larger residuals more strongly than MAE, so the pattern is compatible with a redistribution of residual magnitudes; it does not establish that outliers caused the regressions. Paired seed-level errors and residual distributions would be needed to test that account.

Weather and ETTh1. Weather is a relative weakness: PS-Loss leads four of eight columns (DLinear MAE, FreTS MSE, and both iTransformer metrics), DBLoss leads both TimeMixer metrics, and DySAC leads DLinear MSE and FreTS MAE. TimeMixer's DySAC MSE/MAE are $0.1550/0.2040$, compared with $0.1477/0.1927$ for DBLoss; iTransformer gives $0.1652/0.2077$, compared with $0.1590/0.1996$ for PS-Loss. On ETTh1, PS-Loss leads both FreTS metrics and DLinear MSE, while DBLoss leads DLinear MAE; DySAC leads both metrics for the other two backbones. These comparisons demonstrate setting-dependent objective preferences. Attributing them to patch-wise supervision, decomposition, or particular temporal structures remains a hypothesis without matched ablations or prediction-level diagnostics.

Traffic, Electricity, and ETTh2. On Traffic, DBLoss leads five of eight columns, including the rounded FreTS MSE tie; PS-Loss leads DLinear MAE, and MSE training and QDF tie for DLinear MSE. DySAC leads TimeMixer MSE (0.4548), but DBLoss has lower MAE (0.3024 versus 0.3196). For iTransformer, DBLoss has lower MSE/MAE ($0.3710/0.2589$ versus $0.3744/0.2616$), although DySAC still improves over MSE training on both metrics. On Electricity, FreDF and DBLoss lead DLinear MSE and MAE, respectively, and PS-Loss leads FreTS MSE; DySAC leads the remaining five columns. On ETTh2, TILDE-Q leads both FreTS metrics ($0.2305/0.3002$ versus $0.2331/0.3038$) and ties DySAC on iTransformer MSE. These outcomes distinguish improvement over MSE training from improvement over the strongest competing objective; they do not support a universal Traffic-specific advantage for any one objective.

Short-horizon datasets and FreTS. On CIF-2016, TILDE-Q leads both FreTS metrics ($2.5870/0.9933$ versus $2.7340/1.0248$ for DySAC), and DBLoss leads DLinear MAE (0.8830 versus 0.8893). DySAC leads the other five columns, including both TimeMixer metrics. On Fred-md, FreTS remains a relative weakness: TILDE-Q leads MSE (17.6153 versus 21.9815), and DBLoss leads MAE (0.7404 versus 0.8481). Nevertheless, DySAC improves both metrics over the FreTS MSE baseline ($51.2937/1.1454$). DySAC leads both Fred-md metrics with DLinear and TimeMixer, while QDF leads iTransformer MSE (15.5336 versus 15.6489) and DySAC leads its MAE. Since these short-horizon settings also include several DySAC wins, limited frequency resolution alone cannot explain the FreTS shortfalls. The three-seed summaries do not establish whether optimization, checkpoint selection, or generalization accounts for them. Table 13 belongs to the separate heterogeneous evaluation and cannot diagnose these MSE/MAE runs without a checkpoint-level match.

Uncertainty and interpretation. The reported standard deviations describe variation, not uncertainty in a paired difference. For example, FreTS/CIF-2016 MSE has SD 0.7854 for DySAC and 0.8537 for TILDE-Q; FreTS/Fred-md baseline MSE has SD 23.6487. Conversely, some mean gaps are very small, such as DLinear/ETTh1 MSE (0.3759 versus 0.3757 for PS-Loss). Neither large overlapping deviations nor small absolute gaps alone settle statistical significance. These tables establish descriptive rankings and identify settings for follow-up; they do not prove a common mechanism for the 27 non-winning columns. The spectral figures summarize band-specific behavior under the matching configurations, but checkpoint correspondence is not established; the ablations use a separate configuration. Neither is used as a causal explanation for individual forecasting regressions.

F Evaluation with Heterogeneous Error Metrics

MSE and MAE need not rank objectives identically. To complement these measurements, we report an additional evaluation of the same objectives and backbones using four metrics: the symmetric mean absolute percentage error (SMAPE) and the weighted absolute percentage error (WAPE) measure relative error, the mean per-window Pearson correlation (CORR) measures shape agreement independently of amplitude, and an overall weighted average (OWA)-style score, motivated by the M4 competition (Makridakis et al., 2020), combines relative SMAPE and relative MAE against a seasonal naive reference. The reference repeats the last observed seasonal cycle of length S preceding each test window over the full horizon, with $S = 24$ for the hourly datasets (ETTh1, ETTh2, Electricity, and Traffic), $S = 144$ for the 10-minute Weather data, and $S = 12$ for the monthly Fred-md and CIF-2016 collections. With y and $\hat{y}$ denoting the target and forecast over N test windows,

H horizon steps, and C channels, and $\hat{y}^{\text{naive}}$ the seasonal naive forecast,

$$\text{SMAPE} = \frac{100}{NHC} \sum_{i,c,t} \frac{2\,|y_{i,c,t} - \hat{y}_{i,c,t}|}{|y_{i,c,t}| + |\hat{y}_{i,c,t}|}, \qquad \text{WAPE} = 100 \cdot \frac{\sum_{i,c,t} |y_{i,c,t} - \hat{y}_{i,c,t}|}{\sum_{i,c,t} |y_{i,c,t}|}, \tag{24}$$

$$\text{CORR} = \frac{1}{NC} \sum_{i,c} \frac{\sum_t (y_{i,c,t} - \bar{y}_{i,c})(\hat{y}_{i,c,t} - \bar{\hat{y}}_{i,c})}{\sqrt{\sum_t (y_{i,c,t} - \bar{y}_{i,c})^2} \sqrt{\sum_t (\hat{y}_{i,c,t} - \bar{\hat{y}}_{i,c})^2}}, \tag{25}$$

$$\text{OWA} = \frac{1}{2} \left(\frac{\text{SMAPE}}{\text{SMAPE}_{\text{naive}}} + \frac{\text{MAE}}{\text{MAE}_{\text{naive}}} \right). \tag{26}$$

All four metrics are computed in standardized space without inverse transformation. Standardization can change percentage-error values and SMAPE rankings, so these comparisons apply to this scale and should not be interpreted as raw-scale results or official M4 scores. For identical test targets and predictions, the WAPE definition above is proportional to MAE; it is not an independent ranking criterion. Windows in which the target or the forecast is constant contribute zero correlation. All four backbones (DLinear, FreTS, iTransformer, and TimeMixer) are evaluated on all seven datasets. The heterogeneous-metric runner trains models and selects its own validation checkpoints; these tables are not asserted to be a same-checkpoint re-scoring of Appendix E. Values are means and sample standard deviations (`ddof=1`) over three seeds. Rankings below use the reported means, average ranks for ties, and count all tied best means as first places. Tables 14, 15, 16, and 17 below report all four backbones. Table 2 gives the model-level summary, taking the arithmetic mean of the seven dataset-level means for each backbone and metric: $\bar{m}_{b,o} = \frac{1}{7} \sum_{d=1}^{7} m_{b,d,o}$, where o denotes the training objective. Each dataset has equal weight regardless of its number of test windows or channels. These are macro averages of reported means, not pooled test-set metrics, average ranks, or claims of uniform improvement. We do not average dataset-level standard deviations or interpret the macro means as statistically significant differences.

Across the 112 metric columns (4 backbones $\times$ 7 datasets $\times$ 4 metrics), DySAC obtains the best average rank among the eight objectives (2.04, versus 2.52 for DBLoss and 2.85 for PS-Loss), the largest number of first places (50 of 112, versus 40 for DBLoss), and improves on MSE training in 110 of the 112 columns. Per backbone, DySAC has the best average rank on DLinear (1.86; first in twelve columns) and iTransformer (1.54; first in fourteen columns), and on TimeMixer (2.29, ahead of 2.39 for DBLoss; seventeen first places). It ranks second on FreTS (2.46, behind 2.23 for DBLoss; seven first places). Aggregated per dataset, DySAC has the best average rank on ETTh2 (1.06), Fred-md (1.81), and CIF-2016 (1.69), and ties DBLoss on Electricity (2.00). DySAC takes 15, 9, 10, and 6 first places, respectively, on these datasets. DBLoss has the best average rank on ETTh1 (1.88 versus 2.06 for DySAC) and Traffic (1.75 versus 3.25); PS-Loss leads Weather (1.44 versus 2.38). DySAC improves over MSE in every column except two CORR columns discussed below. The aggregate rankings are favorable, but differences from Appendix I must not be attributed solely to the metric because checkpoint correspondence has not been established.

Where DySAC is outperformed. PS-Loss is particularly strong on Weather: it is first in nine of sixteen columns, including all four with iTransformer, three with DLinear, and CORR with FreTS and TimeMixer. DySAC wins DLinear CORR and the other three FreTS metrics, while DBLoss wins the three non-CORR TimeMixer metrics. On Traffic, DBLoss is first in thirteen of sixteen columns, including one tie, and DySAC does not lead any column. In contrast, DySAC is first in fifteen of sixteen ETTh2 columns and ten of sixteen CIF-2016 columns.

The two non-improvements over MSE training are CORR columns: Traffic with DLinear (0.8642 versus 0.8649) and Electricity with FreTS (0.8855 versus 0.8860). Both are small absolute regressions at the reported precision; neither establishes metric saturation. The Fred-md/FreTS aggregate results are reported in Table 15; Table 13 gives the seed-level diagnostic. These patterns in Tables 15, 14, 16, and 17 are compatible with metric-specific trade-offs, but do not establish their causes or account for every non-first column. Being below a competing objective is distinct from being below MSE training.

Fred-md with TimeMixer. Table 17 uses a revised training-epoch setting for all eight objectives on Fred-md, with other settings unchanged. DySAC has the best reported mean in all four metrics in this setting. The aggregate summary does not specify the revised epoch count or provide matching

per-seed logs. The TimeMixer gradient diagnostics in Appendix K retain the original 30-epoch runs, not this revised evaluation.

Fred-md with FreTS. Table 13 reports test SMAPE, best validation MSE, and training epochs for each of the three seeds. The means and sample standard deviations agree with the corresponding FreTS/Fred-md results in Table 15.

Table 13: FreTS on Fred-md (96–48–12), seeds 2026/2027/2028 (also the epoch-count order). Validation MSE is the best value per run. Seed means and SDs agree with Table 15.

Objective	Test SMAPE (%)			Best validation MSE			Epochs
	2026	2027	2028	2026	2027	2028	
Origin	27.00	25.21	23.64	0.185	0.196	0.167	30/30/30
+DBLoss	20.50	22.15	22.85	0.132	0.124	0.128	30/30/30
+FreDF	24.58	26.88	22.64	0.164	0.188	0.137	30/30/30
+PS-Loss	22.29	23.06	22.29	0.148	0.151	0.151	30/30/30
+QDF	26.49	28.45	28.50	0.183	0.186	0.188	30/30/30
+TILDE-Q	23.01	24.14	25.75	0.132	0.151	0.150	30/30/30
+Time-o1	26.46	27.27	26.02	0.166	0.177	0.162	30/30/30
OURS	24.22	24.89	23.36	0.135	0.147	0.141	30/30/30

Table 14: Heterogeneous-metric results for DLinear on all seven datasets (mean and sample standard deviation over three seeds). SMAPE and WAPE are in percent. Lower is better for SMAPE, WAPE, and OWA; higher is better for CORR. Blue boldface marks the best reported mean in each column and red the second-best distinct mean.

Objective	ETTh1				ETTh2				Weather				Electricity			
	SMAPE↓	WAPE↓	CORR↑	OWA↓	SMAPE↓	WAPE↓	CORR↑	OWA↓	SMAPE↓	WAPE↓	CORR↑	OWA↓	SMAPE↓	WAPE↓	CORR↑	OWA↓
Origin	74.7560 (0.2945)	50.7552 (0.2477)	0.5734 (0.0021)	0.9725 (0.0042)	53.5681 (0.2204)	25.9428 (0.1718)	0.4156 (0.0069)	0.9673 (0.0042)	63.3481 (0.0981)	40.9852 (0.1113)	0.3644 (0.0073)	0.9457 (0.0019)	48.9370 (0.0501)	29.8671 (0.0339)	0.9026 (0.0007)	0.8174 (0.0009)
+DBLoss	**72.3922 (0.0900)**	**49.2826 (0.1073)**	0.5780 (0.0005)	**0.9430 (0.0015)**	51.8789 (0.1136)	24.8720 (0.0556)	0.4260 (0.0020)	0.9323 (0.0020)	58.1663 (1.0914)	37.7205 (0.5363)	0.3693 (0.0077)	0.8693 (0.0144)	**48.5616 (0.0479)**	**29.6225 (0.0334)**	0.9030 (0.0001)	**0.8110 (0.0009)**
+FreDF	73.4738 (0.2063)	49.8963 (0.2124)	0.5784 (0.0010)	0.9560 (0.0033)	52.6846 (0.0770)	25.3803 (0.0769)	0.4334 (0.0089)	0.9490 (0.0019)	62.2099 (0.5707)	40.3074 (0.3778)	0.3674 (0.0063)	0.9293 (0.0086)	48.7775 (0.0531)	29.7583 (0.0494)	**0.9035 (0.0002)**	0.8146 (0.0011)
+PS-Loss	73.1331 (0.0303)	49.3828 (0.0666)	0.5781 (0.0004)	0.9489 (0.0004)	52.0847 (0.0424)	25.0530 (0.0385)	0.4237 (0.0004)	0.9375 (0.0010)	**57.6904 (0.3949)**	**37.3100 (0.2414)**	0.3762 (0.0034)	**0.8611 (0.0057)**	48.7914 (0.0337)	29.7747 (0.0267)	0.9031 (0.0001)	0.8149 (0.0006)
+QDF	74.1575 (0.7441)	50.4281 (0.4224)	0.5734 (0.0014)	0.9655 (0.0088)	53.2150 (0.2323)	25.6150 (0.1546)	0.4137 (0.0073)	0.9581 (0.0047)	62.1783 (0.7773)	40.0869 (0.5908)	0.3525 (0.0075)	0.9267 (0.0125)	48.9295 (0.0381)	29.8608 (0.0253)	0.9028 (0.0008)	0.8173 (0.0007)
+TILDE-Q	73.0866 (0.0958)	49.9146 (0.0699)	0.5697 (0.0018)	0.9535 (0.0013)	51.7394 (0.0791)	24.7560 (0.0649)	0.4356 (0.0015)	0.9289 (0.0018)	58.5248 (0.2966)	37.9386 (0.2303)	0.3750 (0.0066)	0.8745 (0.0048)	50.5602 (0.0963)	31.0005 (0.0511)	0.8971 (0.0004)	0.8464 (0.0015)
+Time-o1	74.6658 (0.4657)	50.7269 (0.4213)	0.5725 (0.0043)	0.9717 (0.0070)	53.2204 (0.1622)	25.5124 (0.1890)	0.4195 (0.0034)	0.9564 (0.0049)	63.5935 (1.8562)	41.1475 (1.2709)	0.3605 (0.0089)	0.9494 (0.0284)	48.8769 (0.0342)	29.8269 (0.0286)	0.9026 (0.0007)	0.8164 (0.0006)
OURS	72.8957 (0.1009)	49.5274 (0.1535)	**0.5806 (0.0010)**	0.9487 (0.0021)	**51.6899 (0.0588)**	**24.6593 (0.0126)**	**0.4392 (0.0023)**	**0.9267 (0.0008)**	58.0561 (0.5298)	37.5407 (0.3841)	**0.3846 (0.0014)**	0.8665 (0.0083)	48.7442 (0.1770)	29.7494 (0.1145)	0.9034 (0.0003)	0.8142 (0.0030)

Objective	Traffic				Fred-md				CIF-2016			
	SMAPE↓	WAPE↓	CORR↑	OWA↓	SMAPE↓	WAPE↓	CORR↑	OWA↓	SMAPE↓	WAPE↓	CORR↑	OWA↓
Origin	54.1782 (0.1165)	38.1450 (0.0780)	**0.8649 (0.0002)**	0.6910 (0.0014)	19.4596 (0.0586)	13.1668 (0.1422)	-0.0603 (0.0718)	0.8071 (0.0052)	58.4090 (1.0138)	43.2530 (0.6486)	0.3227 (0.0181)	1.1107 (0.0178)
+DBLoss	**52.1397 (0.8323)**	37.3435 (0.5071)	0.8629 (0.0007)	**0.6701 (0.0100)**	18.1108 (0.1056)	11.5690 (0.0723)	0.0193 (0.0296)	0.7302 (0.0044)	**53.3176 (2.3657)**	39.6817 (1.6710)	0.3245 (0.0171)	**1.0165 (0.0439)**
+FreDF	53.6189 (0.7158)	37.9599 (0.4723)	0.8644 (0.0007)	0.6856 (0.0089)	18.6458 (0.0998)	12.0423 (0.1464)	0.0927 (0.0219)	0.7558 (0.0064)	56.9587 (1.4812)	42.0401 (1.0192)	0.3368 (0.0199)	1.0812 (0.0271)
+PS-Loss	52.1924 (0.7397)	**37.2882 (0.4705)**	0.8639 (0.0007)	**0.6701 (0.0090)**	17.9354 (0.0575)	11.2604 (0.0304)	0.0526 (0.0484)	0.7171 (0.0018)	53.5199 (1.8270)	39.9063 (1.2404)	0.3326 (0.0252)	1.0213 (0.0332)
+QDF	54.0925 (0.1364)	38.1123 (0.0958)	**0.8649 (0.0001)**	0.6902 (0.0017)	18.9710 (0.1964)	12.4830 (0.2699)	0.0141 (0.1217)	0.7760 (0.0122)	57.1502 (0.8745)	42.3251 (0.5137)	0.3303 (0.0188)	1.0868 (0.0148)
+TILDE-Q	56.0454 (0.3666)	39.8704 (0.2525)	0.8570 (0.0005)	0.7181 (0.0046)	17.7619 (0.0768)	10.9904 (0.0234)	0.1093 (0.0667)	0.7052 (0.0023)	55.3987 (1.1495)	40.6140 (0.7585)	0.3311 (0.0240)	1.0480 (0.0206)
+Time-o1	54.2229 (0.1591)	38.2005 (0.0773)	**0.8649 (0.0002)**	0.6918 (0.0017)	19.5431 (0.1686)	13.2649 (0.1916)	-0.0632 (0.0558)	0.8118 (0.0091)	59.8375 (2.3200)	44.1589 (1.6066)	0.3101 (0.0111)	1.1358 (0.0426)
OURS	52.5524 (0.4694)	37.4513 (0.3338)	0.8642 (0.0003)	0.6739 (0.0060)	**17.5762 (0.0875)**	**10.9012 (0.0953)**	**0.1784 (0.0259)**	**0.6987 (0.0047)**	53.8732 (2.1488)	**39.6077 (1.5535)**	**0.3545 (0.0158)**	1.0206 (0.0403)

Table 15: Heterogeneous-metric results for FreTS on all seven datasets (mean and sample standard deviation over three seeds). SMAPE and WAPE are in percent. Lower is better for SMAPE, WAPE, and OWA; higher is better for CORR. Blue boldface marks the best reported mean in each column and red the second-best distinct mean.

Objective	ETTh1				ETTh2				Weather				Electricity			
	SMAPE↓	WAPE↓	CORR↑	OWA↓	SMAPE↓	WAPE↓	CORR↑	OWA↓	SMAPE↓	WAPE↓	CORR↑	OWA↓	SMAPE↓	WAPE↓	CORR↑	OWA↓
Origin	72.7509 (0.8253)	49.2646 (0.9743)	0.5838 (0.0069)	0.9938 (0.0151)	49.1600 (0.4852)	23.3813 (0.5234)	0.3551 (0.0153)	0.9742 (0.0143)	59.3350 (1.2674)	31.2501 (0.6613)	0.2052 (0.0148)	0.8112 (0.0171)	51.5354 (0.1790)	31.6612 (0.1256)	0.8860 (0.0008)	0.9297 (0.0034)
+DBLoss	72.1670 (0.7883)	48.1308 (0.9424)	0.5916 (0.0045)	**0.9786 (0.0146)**	47.6236 (0.1148)	22.2130 (0.1066)	0.4024 (0.0016)	0.9349 (0.0014)	48.2911 (0.5707)	29.1595 (0.4981)	0.2337 (0.0098)	0.6995 (0.0099)	**50.1837 (0.2373)**	**30.9666 (0.1545)**	0.8862 (0.0006)	**0.9072 (0.0044)**
+FreDF	72.3169 (0.8861)	48.5262 (1.0059)	0.5886 (0.0061)	0.9835 (0.0161)	47.8755 (0.1724)	22.5169 (0.0320)	0.3954 (0.0089)	0.9436 (0.0021)	49.3340 (0.5038)	29.0664 (0.2158)	0.2346 (0.0163)	0.7070 (0.0063)	50.7481 (0.1937)	31.2353 (0.1319)	0.8863 (0.0001)	0.9163 (0.0037)
+PS-Loss	72.6608 (0.4576)	**48.1200 (0.4633)**	**0.6013 (0.0024)**	0.9820 (0.0074)	47.5779 (0.0504)	22.2627 (0.2260)	**0.4201 (0.0088)**	0.9355 (0.0049)	47.9891 (0.5314)	28.6843 (0.5712)	**0.2397 (0.0138)**	0.6920 (0.0103)	50.4816 (0.1107)	31.0456 (0.0740)	**0.8879 (0.0001)**	0.9111 (0.0021)
+QDF	72.5674 (1.0406)	49.0908 (1.2282)	0.5825 (0.0049)	0.9908 (0.0192)	48.3752 (0.1401)	22.7837 (0.0617)	0.3833 (0.0014)	0.9541 (0.0016)	56.4632 (2.8562)	30.6666 (0.3101)	0.2023 (0.0163)	0.7817 (0.0253)	51.7063 (0.3547)	31.7654 (0.2266)	0.8861 (0.0005)	0.9327 (0.0065)
+TILDE-Q	**71.9298 (0.4978)**	48.3346 (0.2134)	0.5801 (0.0070)	0.9789 (0.0046)	47.5584 (0.3890)	22.1444 (0.2754)	0.4132 (0.0054)	0.9329 (0.0095)	49.0137 (0.5178)	29.2937 (0.4687)	0.2079 (0.0223)	0.7068 (0.0091)	53.2967 (0.2594)	32.9461 (0.1516)	0.8773 (0.0003)	0.9643 (0.0045)
+Time-o1	72.9788 (0.5019)	48.5529 (0.2167)	0.5813 (0.0052)	0.9884 (0.0056)	48.9651 (0.3275)	23.1959 (0.2802)	0.3598 (0.0113)	0.9685 (0.0089)	56.0958 (1.0658)	30.8148 (0.4556)	0.2041 (0.0104)	0.7803 (0.0134)	51.6734 (0.4778)	31.7737 (0.3163)	0.8857 (0.0008)	0.9325 (0.0089)
OURS	72.4474 (0.9029)	48.2185 (0.5555)	0.5931 (0.0043)	0.9814 (0.0113)	**47.3927 (0.2485)**	**21.9555 (0.1290)**	0.4157 (0.0053)	**0.9274 (0.0051)**	**47.7042 (0.6916)**	**28.5014 (0.5712)**	0.2357 (0.0100)	**0.6878 (0.0116)**	50.3705 (0.1878)	31.0474 (0.1318)	0.8855 (0.0003)	0.9101 (0.0036)

Objective	Traffic				Fred-md				CIF-2016			
	SMAPE↓	WAPE↓	CORR↑	OWA↓	SMAPE↓	WAPE↓	CORR↑	OWA↓	SMAPE↓	WAPE↓	CORR↑	OWA↓
Origin	59.6420 (0.6547)	41.5668 (0.2659)	0.8629 (0.0005)	0.8500 (0.0076)	25.2832 (1.6819)	22.4538 (5.1714)	-0.0291 (0.1437)	1.2123 (0.1922)	63.8515 (2.2177)	46.4157 (1.7160)	0.2932 (0.0433)	1.2026 (0.0430)
+DBLoss	**57.6693 (0.3518)**	**40.4586 (0.2520)**	0.8629 (0.0007)	**0.8244 (0.0050)**	**21.8349 (1.2068)**	**14.5149 (0.5594)**	0.0032 (0.0783)	**0.8977 (0.0417)**	**53.7885 (5.1633)**	**39.6504 (3.3251)**	0.3044 (0.0374)	**1.0204 (0.0913)**
+FreDF	58.4705 (0.2329)	40.8175 (0.2065)	0.8640 (0.0006)	0.8339 (0.0032)	24.7033 (2.1224)	18.5239 (0.8981)	-0.0824 (0.0527)	1.0800 (0.0476)	61.3103 (5.6335)	45.2616 (3.5840)	0.2963 (0.0210)	1.1640 (0.0981)
+PS-Loss	58.4473 (0.4021)	41.1176 (0.2446)	**0.8665 (0.0004)**	0.8365 (0.0053)	22.5484 (0.4428)	19.1384 (2.4917)	**0.0820 (0.0193)**	1.0540 (0.0853)	61.2495 (5.6905)	44.8132 (3.6074)	0.3034 (0.0401)	1.1574 (0.1001)
+QDF	58.9916 (0.4431)	41.1267 (0.1520)	0.8640 (0.0004)	0.8409 (0.0048)	27.8104 (1.1458)	17.8192 (1.6859)	0.0112 (0.0903)	1.1229 (0.0644)	64.1389 (4.7120)	46.3822 (3.1236)	0.2873 (0.0381)	1.2048 (0.0846)
+TILDE-Q	59.9699 (0.3872)	41.8881 (0.2099)	0.8574 (0.0004)	0.8555 (0.0048)	24.3004 (1.3785)	15.7404 (0.3491)	-0.0666 (0.0454)	0.9864 (0.0390)	54.6178 (5.6692)	40.3960 (3.4139)	0.2891 (0.0140)	1.0379 (0.0973)
+Time-o1	59.9132 (0.3900)	41.6724 (0.3499)	0.8618 (0.0010)	0.8531 (0.0063)	26.5855 (0.6348)	20.4477 (3.0768)	-0.1499 (0.0353)	1.1779 (0.0813)	62.5084 (7.6547)	44.8869 (4.9239)	0.3194 (0.0249)	1.1699 (0.1355)
OURS	58.2524 (0.2203)	40.6931 (0.0624)	0.8638 (0.0004)	0.8311 (0.0014)	24.1565 (0.7696)	16.6258 (0.5704)	-0.0080 (0.1186)	1.0105 (0.0331)	58.9794 (2.9855)	42.7604 (2.5871)	**0.3386 (0.0033)**	1.1093 (0.0618)

Table 16: Heterogeneous-metric results for iTransformer on all seven datasets (mean and sample standard deviation over three seeds). SMAPE and WAPE are in percent. Lower is better for SMAPE, WAPE, and OWA; higher is better for CORR. Blue boldface marks the best reported mean in each column and red the second-best distinct mean.

Objective	ETTh1				ETTh2				Weather				Electricity			
	SMAPE↓	WAPE↓	CORR↑	OWA↓	SMAPE↓	WAPE↓	CORR↑	OWA↓	SMAPE↓	WAPE↓	CORR↑	OWA↓	SMAPE↓	WAPE↓	CORR↑	OWA↓
Origin	75.5572 (1.3796)	52.3709 (0.1544)	0.5642 (0.0030)	0.9928 (0.0104)	57.2902 (0.2014)	27.0501 (0.0815)	0.3057 (0.0172)	1.0221 (0.0012)	56.0953 (0.8635)	35.7213 (0.3649)	0.3735 (0.0052)	0.8314 (0.0104)	46.1277 (0.1552)	27.7152 (0.1369)	0.9125 (0.0002)	0.7649 (0.0031)
+DBLoss	73.8776 (1.0382)	50.8310 (0.3939)	0.5800 (0.0009)	0.9673 (0.0104)	54.5955 (1.0953)	25.9894 (0.4269)	0.3979 (0.0337)	0.9778 (0.0177)	54.0841 (0.4857)	35.1686 (0.3989)	0.3878 (0.0067)	0.8093 (0.0081)	**45.0478 (0.0106)**	26.9657 (0.0268)	0.9140 (0.0001)	**0.7457 (0.0004)**
+FreDF	74.7139 (1.3335)	51.4234 (0.3389)	0.5779 (0.0008)	0.9784 (0.0121)	56.3454 (0.0326)	26.5633 (0.1276)	0.3325 (0.0031)	1.0045 (0.0022)	55.2097 (0.3339)	34.9974 (0.2081)	0.3907 (0.0050)	0.8166 (0.0047)	45.3802 (0.1486)	27.1962 (0.1095)	0.9139 (0.0000)	0.7516 (0.0027)
+PS-Loss	74.1642 (0.2251)	50.8371 (0.2530)	**0.5833 (0.0012)**	0.9693 (0.0016)	53.7028 (0.4013)	25.6464 (0.2841)	0.4193 (0.0121)	0.9633 (0.0086)	**52.6955 (0.5122)**	**33.3427 (0.1962)**	**0.4095 (0.0030)**	**0.7787 (0.0062)**	45.5228 (0.1313)	27.1898 (0.0794)	0.9141 (0.0004)	0.7528 (0.0022)
+QDF	75.8529 (1.3380)	52.5890 (0.2149)	0.5671 (0.0020)	0.9968 (0.0110)	56.5409 (0.4892)	26.8871 (0.4441)	0.3307 (0.0119)	1.0121 (0.0122)	54.7412 (0.6749)	34.6489 (0.4359)	0.3852 (0.0023)	0.8091 (0.0100)	45.9486 (0.1648)	27.5648 (0.1155)	0.9127 (0.0002)	0.7614 (0.0029)
+TILDE-Q	74.1141 (0.1116)	51.8557 (0.5075)	0.5722 (0.0015)	0.9783 (0.0045)	54.2226 (0.2356)	25.7071 (0.1255)	0.3895 (0.0205)	0.9692 (0.0044)	54.0525 (0.3662)	34.2024 (0.2730)	0.3927 (0.0015)	0.7988 (0.0056)	47.7135 (0.0845)	28.8449 (0.0500)	0.9066 (0.0001)	0.7935 (0.0014)
+Time-o1	75.1370 (1.0891)	52.4304 (0.6557)	0.5630 (0.0024)	0.9905 (0.0134)	57.6518 (0.4126)	27.2068 (0.2766)	0.3039 (0.0096)	1.0283 (0.0087)	54.5826 (0.4934)	34.7325 (0.3098)	0.3767 (0.0077)	0.8087 (0.0070)	46.0202 (0.1158)	27.6342 (0.0693)	0.9124 (0.0003)	0.7629 (0.0019)
OURS	**73.8475 (0.3281)**	**50.6682 (0.3058)**	0.5818 (0.0021)	**0.9656 (0.0030)**	**53.4859 (0.5897)**	**25.5773 (0.3162)**	**0.4274 (0.0128)**	**0.9600 (0.0111)**	53.3997 (0.6952)	34.1402 (0.3676)	0.4007 (0.0033)	0.7929 (0.0095)	45.1359 (0.2274)	**26.9540 (0.1099)**	**0.9147 (0.0001)**	0.7464 (0.0034)

Objective	Traffic				Fred-md				CIF-2016			
	SMAPE↓	WAPE↓	CORR↑	OWA↓	SMAPE↓	WAPE↓	CORR↑	OWA↓	SMAPE↓	WAPE↓	CORR↑	OWA↓
Origin	48.7821 (0.0250)	33.6910 (0.1238)	0.8913 (0.0002)	0.6169 (0.0009)	21.4460 (0.6893)	14.1685 (1.2550)	0.1201 (0.0527)	0.8790 (0.0527)	40.5945 (0.7195)	28.3280 (0.5969)	0.4708 (0.0175)	0.7488 (0.0143)
+DBLoss	**46.9440 (0.2478)**	**32.1902 (0.1035)**	**0.8928 (0.0002)**	**0.5918 (0.0026)**	**20.4404 (0.3058)**	13.2065 (0.2833)	0.1633 (0.0431)	**0.8287 (0.0063)**	39.2253 (0.2383)	27.2339 (0.4570)	0.4905 (0.0166)	0.7217 (0.0081)
+FreDF	48.0473 (0.1098)	32.9278 (0.1505)	0.8924 (0.0003)	0.6056 (0.0013)	21.4422 (0.3425)	13.3812 (0.6506)	0.1417 (0.0314)	0.8548 (0.0128)	39.7215 (0.4468)	27.4346 (0.3757)	0.5046 (0.0059)	0.7289 (0.0091)
+PS-Loss	48.5228 (0.2860)	33.6455 (0.2091)	0.8923 (0.0007)	0.6147 (0.0037)	21.1817 (1.4285)	13.4081 (1.1757)	**0.2022 (0.0143)**	0.8503 (0.0598)	39.6544 (0.3497)	27.5468 (0.2731)	0.5106 (0.0018)	0.7298 (0.0068)
+QDF	48.8592 (0.3164)	33.6146 (0.2342)	0.8912 (0.0004)	0.6169 (0.0041)	21.6425 (0.2808)	13.3831 (0.7956)	0.1388 (0.0523)	0.8591 (0.0299)	40.2344 (0.3703)	27.9356 (0.2870)	0.4894 (0.0180)	0.7403 (0.0071)
+TILDE-Q	52.0096 (0.0441)	35.6738 (0.1027)	0.8828 (0.0004)	0.6558 (0.0011)	20.5506 (0.5989)	13.8670 (0.6780)	0.1587 (0.0092)	0.8512 (0.0215)	40.5943 (0.7548)	27.9142 (0.4657)	0.4701 (0.0207)	0.7433 (0.0128)
+Time-o1	48.9086 (0.0755)	33.6111 (0.0556)	0.8910 (0.0002)	0.6172 (0.0006)	21.5263 (0.1920)	13.5411 (0.5917)	0.1367 (0.0532)	0.8615 (0.0215)	39.9876 (0.1543)	27.6150 (0.1963)	0.4884 (0.0329)	0.7338 (0.0036)
OURS	47.6269 (0.2097)	32.6301 (0.0579)	0.8927 (0.0004)	0.6002 (0.0019)	20.8849 (0.6045)	**13.0806 (0.8674)**	0.1842 (0.0247)	0.8341 (0.0365)	**38.4167 (0.2697)**	**26.5530 (0.0887)**	**0.5108 (0.0043)**	**0.7052 (0.0036)**

Table 17: Heterogeneous-metric results for TimeMixer on all seven datasets (mean and sample standard deviation over three seeds). SMAPE and WAPE are in percent. Lower is better for SMAPE, WAPE, and OWA; higher is better for CORR. Blue boldface marks the best reported mean in each column and red the second-best distinct mean.

Objective	ETTh1				ETTh2				Weather				Electricity			
	SMAPE↓	WAPE↓	CORR↑	OWA↓	SMAPE↓	WAPE↓	CORR↑	OWA↓	SMAPE↓	WAPE↓	CORR↑	OWA↓	SMAPE↓	WAPE↓	CORR↑	OWA↓
Origin	77.3420 (1.2773)	53.7174 (1.0997)	0.5493 (0.0017)	1.0173 (0.0187)	57.3142 (1.2600)	28.5453 (0.5676)	0.3364 (0.0103)	1.0491 (0.0210)	53.6846 (0.1871)	33.6509 (0.1837)	0.3814 (0.0061)	0.7900 (0.0024)	49.5208 (0.7520)	30.1352 (0.3663)	0.9050 (0.0012)	0.8260 (0.0114)
+DBLoss	**74.2684 (0.8768)**	52.2294 (1.6916)	0.5708 (0.0104)	**0.9828 (0.0206)**	54.0224 (0.9459)	26.0581 (0.9399)	0.4215 (0.0060)	0.9737 (0.0257)	**50.9789 (0.6312)**	**31.7255 (0.3632)**	0.4146 (0.0121)	**0.7478 (0.0089)**	47.9978 (0.2809)	29.1052 (0.0918)	0.9089 (0.0008)	0.7993 (0.0037)
+FreDF	77.1789 (1.6811)	52.9709 (1.1128)	0.5634 (0.0027)	1.0093 (0.0215)	55.7305 (0.2139)	27.1864 (0.5260)	0.3927 (0.0075)	1.0099 (0.0104)	53.1358 (0.2776)	33.2269 (0.2573)	0.3915 (0.0051)	0.7811 (0.0047)	48.0025 (0.4776)	29.0070 (0.3501)	0.9093 (0.0004)	0.7981 (0.0087)
+PS-Loss	74.7656 (1.2801)	52.6691 (1.3550)	0.5676 (0.0173)	0.9902 (0.0211)	55.0680 (1.3656)	27.1298 (1.5062)	0.4118 (0.0161)	1.0027 (0.0398)	52.0650 (0.3544)	32.4422 (0.4030)	**0.4171 (0.0141)**	0.7641 (0.0070)	48.4500 (0.2468)	29.3841 (0.2203)	0.9096 (0.0003)	0.8069 (0.0047)
+QDF	76.6726 (1.5312)	53.4751 (1.2716)	0.5522 (0.0029)	1.0105 (0.0220)	55.9851 (0.3431)	27.7989 (0.2405)	0.3815 (0.0128)	1.0233 (0.0074)	54.1783 (0.6425)	33.8950 (0.4565)	0.3939 (0.0117)	0.7966 (0.0099)	49.6221 (0.5062)	30.2356 (0.4981)	0.9048 (0.0011)	0.8282 (0.0108)
+TILDE-Q	74.3920 (0.9562)	52.4402 (0.9051)	0.5634 (0.0101)	0.9856 (0.0147)	54.3641 (1.0669)	27.1495 (1.3938)	0.3927 (0.0112)	0.9964 (0.0350)	52.3135 (0.2376)	32.6471 (0.3662)	0.4159 (0.0073)	0.7683 (0.0057)	50.2879 (0.5300)	30.6201 (0.4317)	0.9028 (0.0011)	0.8391 (0.0102)
+Time-o1	76.7265 (1.0959)	53.8506 (1.0624)	0.5503 (0.0069)	1.0143 (0.0158)	58.5225 (2.3550)	30.0658 (1.5453)	0.3403 (0.0105)	1.0877 (0.0496)	53.9419 (0.7675)	33.6995 (0.5810)	0.3911 (0.0180)	0.7926 (0.0121)	49.5909 (0.3106)	30.1288 (0.2306)	0.9048 (0.0006)	0.8266 (0.0057)
OURS	74.9978 (0.9175)	**51.8573 (0.8644)**	**0.5827 (0.0061)**	0.9843 (0.0126)	**54.0088 (0.2173)**	**25.9694 (0.0121)**	**0.4347 (0.0054)**	**0.9719 (0.0019)**	52.9781 (0.2157)	33.5936 (0.3370)	0.4020 (0.0073)	0.7837 (0.0050)	**47.7058 (0.2990)**	**28.6405 (0.2330)**	**0.9104 (0.0010)**	**0.7908 (0.0056)**

Objective	Traffic				Fred-md				CIF-2016			
	SMAPE↓	WAPE↓	CORR↑	OWA↓	SMAPE↓	WAPE↓	CORR↑	OWA↓	SMAPE↓	WAPE↓	CORR↑	OWA↓
Origin	58.4293 (0.5732)	40.3690 (0.2817)	0.8676 (0.0009)	0.7391 (0.0063)	22.2720 (0.6536)	14.0046 (0.2432)	0.0550 (0.0140)	0.8912 (0.0178)	39.8654 (0.8809)	28.0664 (0.3354)	0.4996 (0.0098)	0.7387 (0.0125)
+DBLoss	**54.5914 (0.8675)**	**37.8371 (0.4380)**	**0.8717 (0.0021)**	**0.6915 (0.0097)**	21.5143 (0.7224)	14.1398 (0.2749)	0.0851 (0.0878)	0.8796 (0.0234)	38.2345 (0.6730)	26.9163 (0.3235)	0.5064 (0.0063)	0.7084 (0.0105)
+FreDF	57.0966 (0.9484)	39.6683 (0.3771)	0.8700 (0.0016)	0.7240 (0.0098)	21.8009 (0.1826)	13.4271 (0.2550)	0.1870 (0.0298)	0.8637 (0.0093)	39.0277 (0.8672)	27.4973 (0.2775)	0.5021 (0.0090)	0.7234 (0.0116)
+PS-Loss	55.9535 (0.7970)	39.0767 (0.3170)	0.8708 (0.0018)	0.7111 (0.0082)	21.4021 (0.3110)	13.0540 (0.2359)	0.2309 (0.0285)	0.8440 (0.0137)	38.6510 (0.5096)	27.3323 (0.2607)	**0.5225 (0.0076)**	0.7178 (0.0080)
+QDF	57.6763 (0.6610)	40.1315 (0.5368)	0.8684 (0.0001)	0.7318 (0.0089)	21.5022 (0.8215)	13.5507 (0.0301)	0.0441 (0.0227)	0.8613 (0.0171)	39.5873 (0.8578)	27.8241 (0.2700)	0.4999 (0.0103)	0.7329 (0.0114)
+TILDE-Q	57.7132 (1.3559)	40.4708 (0.6527)	0.8650 (0.0016)	0.7348 (0.0149)	21.7954 (0.3964)	13.8788 (0.1444)	0.2090 (0.0067)	0.8774 (0.0125)	41.4485 (0.1118)	28.2955 (0.0202)	0.4544 (0.0079)	0.7562 (0.0012)
+Time-o1	57.6402 (1.0148)	40.0531 (0.8182)	0.8688 (0.0016)	0.7309 (0.0136)	22.3717 (0.7983)	14.0054 (0.4110)	0.0824 (0.0241)	0.8933 (0.0263)	41.2012 (1.6036)	28.8284 (0.8039)	0.4900 (0.0063)	0.7610 (0.0254)
OURS	57.9038 (0.7281)	39.9597 (0.3268)	0.8702 (0.0013)	0.7321 (0.0062)	**21.1626 (0.0906)**	**12.8700 (0.1734)**	**0.2537 (0.0063)**	**0.8334 (0.0070)**	**37.4217 (0.4229)**	**26.5564 (0.3162)**	0.5051 (0.0089)	**0.6962 (0.0081)**

G Training Overhead

Table 18 reports the training-step measurements for FreTS and TimeMixer on Electricity and Traffic. Every objective–backbone–dataset setting has three runs with seeds 2026, 2027, and 2028, each measuring 100 optimization steps with batch size 32 on CUDA. The benchmark uses a sequence–label–prediction configuration of 192–96–96, ten warm-up steps, the PPU-ZW810E hardware, and PyTorch 2.4.0; FreTS and TimeMixer use CUDA devices 0 and 1, respectively. The benchmark times data transfer, forward evaluation, objective construction, backward computation, gradient clipping, and the optimizer update. It synchronizes CUDA before and after each timed step. DataLoader waiting time and one-time setup are outside the timed interval. The reported SD is the sample standard deviation across the three run-level mean times, not variability across individual steps. These timings apply to this benchmark configuration and should not be generalized to every accuracy experiment; per-step hardware-load variation is not characterized.

All six structured alternatives are included in the timing benchmark, and each ratio is normalized by the matched three-run MSE mean for the same backbone–dataset setting.

DySAC takes $3.30\times/3.02\times$ the MSE step time on Electricity/Traffic with FreTS, and $2.80\times/2.52\times$ with TimeMixer. Its extra work includes per-band gradients and the coordination solve; the full step time is measured, not the solve in isolation. Across the four settings, DBLoss, FreDF, PS-Loss, QDF, TILDE-Q, and Time-o1 span $1.00\times$–$2.12\times$ MSE, $1.01\times$–$1.57\times$ MSE, $0.98\times$–$1.88\times$ MSE, $0.84\times$–$2.18\times$ MSE, $0.99\times$–$1.08\times$ MSE, and $0.78\times$–$1.52\times$ MSE, respectively. Three run-level summaries alone do not establish a reproducible speedup or identify the cause; per-step timing and hardware-load variation are not characterized. Timing variation is therefore reported rather than dismissed as noise. DySAC leaves the backbone and inference procedure unchanged; these measurements concern training only.

Table 18: Training-step time on Electricity and Traffic with FreTS and TimeMixer. Milliseconds report the mean and sample standard deviation of the three run-level mean times (seeds 2026, 2027, 2028); each run measures 100 steps with batch size 32 on CUDA. Ratios divide each objective's three-run mean by the matched MSE three-run mean, rather than averaging per-seed ratios.

Objective	FreTS / Electricity		FreTS / Traffic		TimeMixer / Electricity		TimeMixer / Traffic	
	ms (SD)	× MSE	ms (SD)	× MSE	ms (SD)	× MSE	ms (SD)	× MSE
MSE	1922.27 (14.30)	1.00	5204.26 (63.61)	1.00	32.01 (0.47)	1.00	38.54 (0.84)	1.00
DBLoss	1927.36 (4.90)	1.00	5234.42 (11.06)	1.01	47.80 (14.65)	1.49	81.81 (2.55)	2.12
FreDF	1934.66 (3.49)	1.01	5248.22 (3.20)	1.01	50.26 (14.17)	1.57	47.37 (13.69)	1.23
PS-Loss	1886.03 (92.70)	0.98	5191.59 (74.65)	1.00	60.19 (14.36)	1.88	64.98 (31.42)	1.69
QDF	1609.62 (185.16)	0.84	4684.24 (456.07)	0.90	40.80 (13.64)	1.27	84.01 (3.70)	2.18
TILDE-Q	1900.22 (19.71)	0.99	5255.97 (4.52)	1.01	34.58 (1.54)	1.08	41.61 (0.97)	1.08
Time-o1	1506.10 (259.78)	0.78	4375.43 (504.80)	0.84	33.25 (0.38)	1.04	58.41 (26.50)	1.52
DySAC	6337.46 (537.52)	3.30	15708.04 (887.90)	3.02	89.74 (1.84)	2.80	97.21 (0.38)	2.52

H Complete Task-Wise Spectral Improvement Results

Table 19 reports the mean task-wise improvement of every objective aggregated over the dyadic tasks of each dataset (three to five tasks per setting, four backbones per dataset, three seeds). Improvements are relative to matched MSE training as defined in Section 2.3; positive values indicate lower task error than MSE training. DySAC achieves the highest improvement on six of the seven datasets; on Fred-md, TILDE-Q achieves the largest mean improvement, where it also wins the DLinear and TimeMixer per-setting rankings in Table 21. DBLoss wins Fred-md with FreTS, and QDF wins Fred-md with iTransformer.

Table 19: Mean task-wise spectral improvement (%) by dataset, averaged over all valid dyadic task instances after seed aggregation, pooling the four backbones of each dataset. The best value per dataset is bolded.

Dataset	**+OURS**	+FreDF	+TILDE-Q	+DBLoss	+Time-o1	+QDF	+PS-Loss
ETTh1	**+2.89**	+1.85	+1.25	+1.97	+0.32	+0.39	+2.22
ETTh2	**+4.48**	+2.47	+3.92	+2.88	-0.15	+1.12	+3.02
Weather	**+3.69**	+3.10	+1.18	+1.94	+0.56	-0.20	+2.51
Electricity	**+1.89**	+1.40	-4.91	+1.15	-0.15	+0.06	+0.98
Traffic	**+1.40**	+1.07	-1.61	+0.92	+0.12	+0.37	+0.64
CIF-2016	**+4.77**	+2.26	+1.10	+2.60	-0.63	+0.70	+1.34
Fred-md	+10.74	+9.22	**+16.39**	+11.67	+2.19	+0.73	+9.19

Figures 6–9 report the task-wise spectral improvement for every backbone and dataset. In each panel, a box plot of the per-seed per-task improvement is combined with jittered individual measurements and a star at the mean; the red box is our method (OURS), with the same method colors used throughout. Measurements beyond $[-15\%, 20\%]$ are marked by outward triangles at the boundaries, and out-of-range means are shown as boundary stars. All statistics are computed before clipping. In the cross-objective balance figure (Figure 5), the downward triangles mark settings whose mean improvement falls below the displayed range: QDF on Fred-md–FreTS (-9.97%). The upper range includes DBLoss on Fred-md–FreTS ($+29.14\%$); this point is not clipped. DySAC on Fred-md–FreTS has a positive mean of $+10.13\%$ despite ranking fifth among objectives there.

The task identities used in this evaluation correspond one-to-one to the dyadic spectral objectives defined in Section 2.1, and every objective within a setting is evaluated on exactly the same dyadic partition.

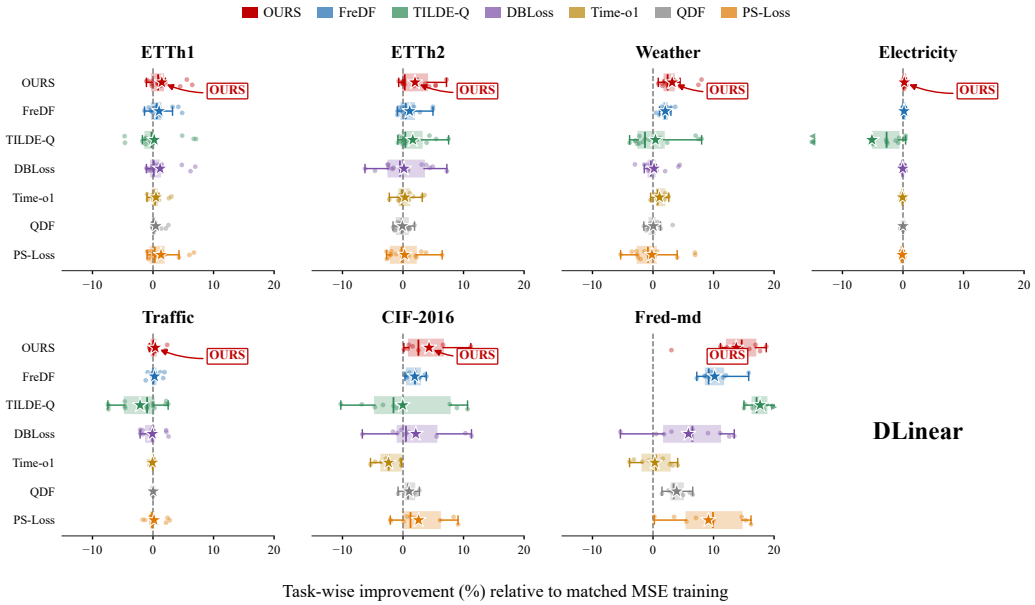


Figure 6: Task-wise spectral improvement for DLinear across seven datasets.

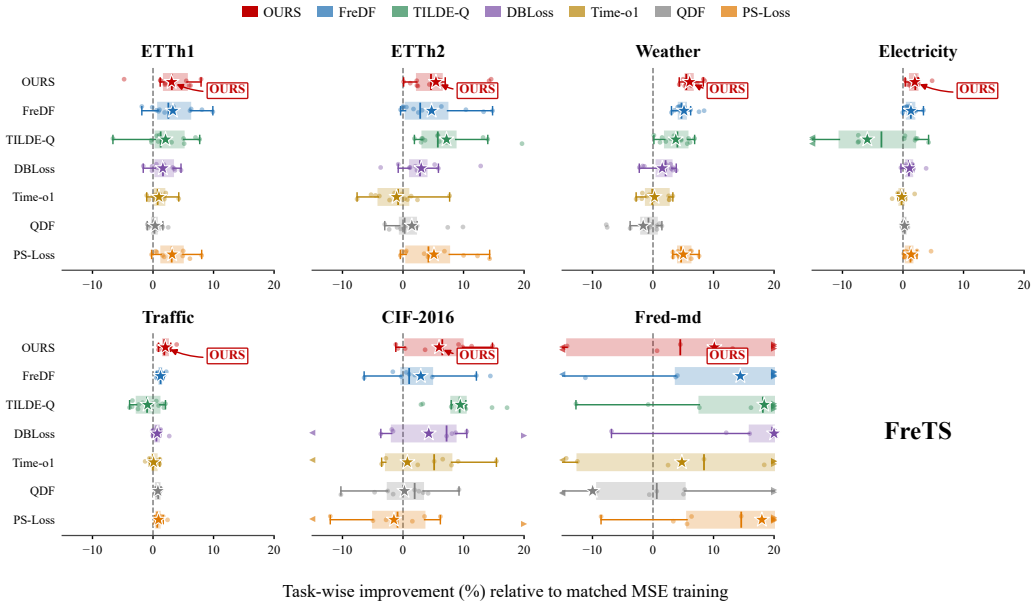


Figure 7: Task-wise spectral improvement for FreTS across seven datasets.

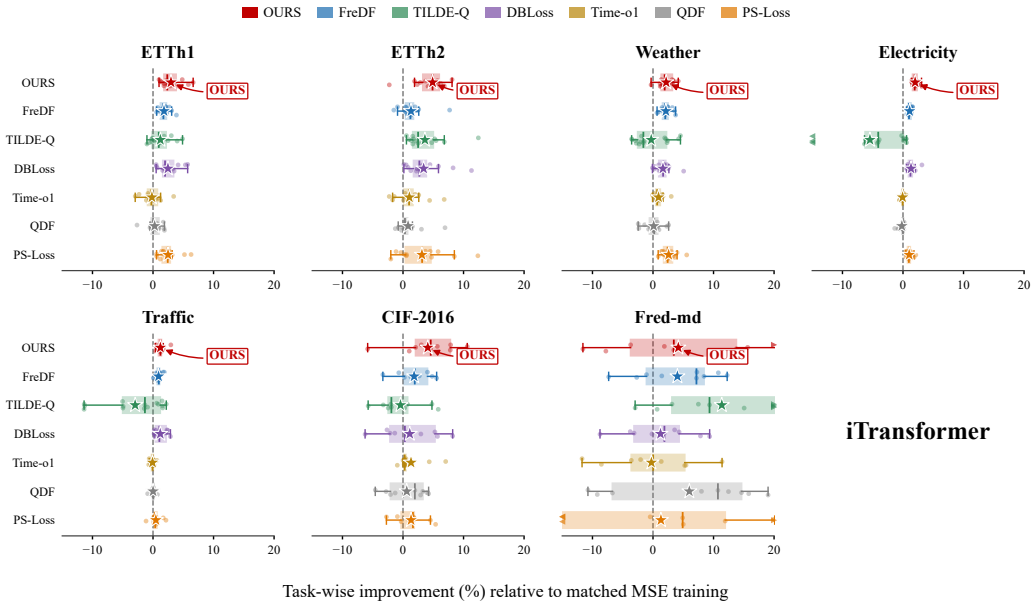


Figure 8: Task-wise spectral improvement for iTransformer across seven datasets.

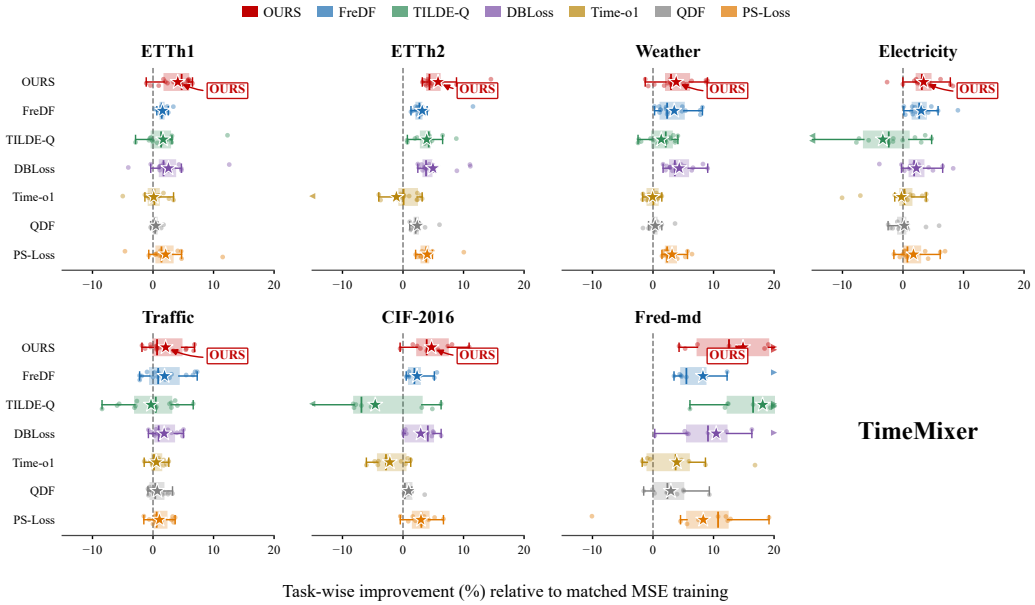


Figure 9: Task-wise spectral improvement for TimeMixer across seven datasets.

I COMPLETE TASK-WISE OBJECTIVE RANKING RESULTS

Table 21 summarizes the average-task-rank results for every dataset–backbone setting. Within each setting, objectives are ranked independently within every dyadic task by mean improvement over three seeds (exact ties receive average ranks), and the average task rank follows Section 2.3. The reported tasks correspond one-to-one to the dyadic spectral objectives defined in Section 2.1; the task count varies with the prediction horizon. DySAC achieves the best average task rank in 19 of the 28 settings: 16 uniquely and three at tied minima. The tied minima are ETTh1–FreTS (with FreDF and PS-Loss), Weather–TimeMixer (with DBLoss), and Traffic–iTransformer (with DBLoss). Under average tie ranks, the three-way minimum on ETTh1–FreTS receives second-level rank 2, but still counts as a best setting. DySAC is uniquely second in seven further settings: ETTh2–FreTS, Weather–iTransformer, Electricity–TimeMixer, Traffic–TimeMixer, CIF-2016–FreTS, Fred-md–DLinear, and Fred-md–TimeMixer. On Fred-md–iTransformer, it ties for second with second-level rank 2.5; on Fred-md–FreTS, it ranks fifth, while DBLoss ranks first.

Table 20: Cross-objective balance comparison aggregated over 28 dataset–backbone settings (119 dyadic spectral tasks per objective). Average task rank is the equal-weight mean of the setting-level values in Section 2.3; lower is better, and standard deviations are computed across settings (population definition). Rows are sorted by average task rank. "Rank-1 settings" counts the settings in which the objective achieves the best average task rank (ties are counted for all tied objectives). Mean task improvement averages all task-wise improvements relative to matched MSE training.

Objective	Avg. task rank ↓	Rank-1 settings	Mean task impr. (%)
+OURS	**2.01** ($\pm$0.91)	**19**/28	**+3.85**
+FreDF	3.42 ($\pm$0.77)	2/28	+2.74
+DBLoss	3.89 ($\pm$1.30)	4/28	+2.85
+PS-Loss	4.04 ($\pm$1.13)	2/28	+2.56
+TILDE-Q	4.51 ($\pm$1.90)	4/28	+1.74
+QDF	5.61 ($\pm$1.19)	1/28	+0.42
+Time-o1	6.09 ($\pm$1.19)	0/28	+0.27
Origin	6.42 ($\pm$0.91)	0/28	0.00

Table 21: Average task rank for every dataset–backbone setting. Lower rank is better; all objectives attaining the minimum are listed, and unique OURS wins are bolded. Config denotes sequence–label–prediction lengths.

Dataset	Backbone	Config (S-L-P)	Best objective	Best avg. rank	OURS avg. rank	Tasks
ETTh1	DLinear	192-96-96	**+OURS**	2.80	2.80	5
ETTh1	FreTS	96-48-48	+FreDF/+OURS/+PS-Loss	2.75	2.75	4
ETTh1	iTransformer	192-96-96	**+OURS**	1.40	1.40	5
ETTh1	TimeMixer	192-96-96	**+OURS**	1.60	1.60	5
ETTh2	DLinear	192-96-96	**+OURS**	2.00	2.00	5
ETTh2	FreTS	96-48-48	+TILDE-Q	1.25	2.75	4
ETTh2	iTransformer	192-96-96	**+OURS**	1.40	1.40	5
ETTh2	TimeMixer	192-96-96	**+OURS**	1.20	1.20	5
Weather	DLinear	192-96-96	**+OURS**	1.60	1.60	5
Weather	FreTS	96-48-48	**+OURS**	1.50	1.50	4
Weather	iTransformer	192-96-96	+PS-Loss	2.20	2.40	5
Weather	TimeMixer	192-96-96	+DBLoss/+OURS	2.20	2.20	5
Electricity	DLinear	192-96-96	**+OURS**	1.40	1.40	5
Electricity	FreTS	96-48-48	**+OURS**	1.00	1.00	4
Electricity	iTransformer	192-96-96	**+OURS**	1.00	1.00	5
Electricity	TimeMixer	192-96-96	+FreDF	1.80	2.20	5
Traffic	DLinear	192-96-96	**+OURS**	3.00	3.00	5
Traffic	FreTS	96-48-48	**+OURS**	1.00	1.00	4
Traffic	iTransformer	192-96-96	+DBLoss/+OURS	1.80	1.80	5
Traffic	TimeMixer	192-96-96	+DBLoss	2.80	3.40	5
CIF-2016	DLinear	24-12-12	**+OURS**	1.00	1.00	3
CIF-2016	FreTS	24-12-12	+TILDE-Q	1.00	2.67	3
CIF-2016	iTransformer	24-12-12	**+OURS**	1.67	1.67	3
CIF-2016	TimeMixer	24-12-12	**+OURS**	1.00	1.00	3
Fred-md	DLinear	96-48-12	+TILDE-Q	1.33	2.00	3
Fred-md	FreTS	96-48-12	+DBLoss	1.00	5.00	3
Fred-md	iTransformer	96-48-12	+QDF	2.00	3.00	3
Fred-md	TimeMixer	96-48-12	+TILDE-Q	1.33	1.67	3

J Ablation Study Details

Component ablation. Table 4 isolates frequency-domain supervision, dyadic task decomposition, and gradient coordination on FreTS. All relative changes in this appendix are approximate values calculated from the displayed means as $100(b - v)/b$, where b is the comparison baseline and v is the variant. The single spectral loss improves Weather MSE/MAE by 1.82%/4.35% and Electricity MAE by 0.93%, while Electricity MSE worsens by 0.37%. Dyadic + Mean improves over the single spectral loss in all four columns: 1.62%/0.94% in Electricity MSE/MAE and 0.05%/1.10% in Weather MSE/MAE. Coordination then improves over Dyadic + Mean by 0.11%/0.36% and 1.05%/1.50%, respectively. The last comparison keeps the objectives fixed and changes their gradient combination, so it measures the benefit over fixed uniform aggregation; it does not compare against adaptive weighting. The full method has the lowest mean in all four columns, improving over MSE training by 1.37%/2.22% on Electricity and 2.90%/6.82% on Weather. Three-seed mean differences alone do not establish statistical significance.

Choice of spectral loss. Equation equation 3 uses the magnitude-L1 spectral residual. Table 5 compares this with a squared-magnitude L2 objective under both the single-spectrum baseline and the complete Dyadic + ConFIG framework on FreTS. At the single-spectrum level, neither loss dominates every metric: on Electricity, L2 obtains a lower MSE (0.1894 versus 0.1910) whereas

L1 gives a lower MAE (0.2769 versus 0.2800), and on Weather L1 gives lower MSE and MAE (0.2101 versus 0.2124 and 0.2637 versus 0.2736). The comparison is more consistent for the complete method: with Dyadic + ConFIG, L1 attains the lower mean on all four metrics (Electricity MSE 0.1877 versus 0.1897, Electricity MAE 0.2733 versus 0.2789, Weather MSE 0.2078 versus 0.2095, Weather MAE 0.2569 versus 0.2703). We therefore use magnitude-L1 in the remaining DySAC experiments, not because it dominates every isolated setting, but because it provides the more consistent performance across the evaluated metrics under the complete framework.

Choice of frequency partition. Table 6 gives six unique dyadic wins and one tied win across eight comparisons at the displayed precision. On FreTS/Electricity, dyadic reduces MSE by 1.78%, 0.64%, and 1.83% relative to uniform, geometric, and random, respectively; the MAE reductions are 1.16%, 0.47%, and 1.37%. On FreTS/Weather, means span 0.2075–0.2086 in MSE and 0.2569–0.2590 in MAE. Uniform's MSE is 0.14% below dyadic, geometric's MSE is also lower, and geometric ties dyadic in MAE. Against random, dyadic improves MSE/MAE by 0.38%/0.81%; against uniform it improves MAE by 0.39%. These small differences should not be interpreted as a resolved ordering without a paired analysis.

On TimeMixer, dyadic has the lowest mean in all four columns. On Electricity, the MSE reductions over uniform, geometric, and random are 3.08%, 0.64%, and 5.40%, with MAE reductions of 1.40%, 0.15%, and 2.97%. On Weather, the corresponding MSE reductions are 2.00%, 1.13%, and 1.68%, with MAE reductions of 1.26%, 0.83%, and 0.79%. All variants retain multiple objectives and the same coordination mechanism, so these comparisons isolate partition choice within the stated setup. They are compatible with the usefulness of multiresolution organization, but do not establish a unique causal explanation or a universally optimal partition. Dyadic's positive reductions over geometric never exceed 1.13% MSE or 0.83% MAE here. At fixed base bandwidth, dyadic is a simple multiresolution choice without a separate growth-ratio parameter.

Alternative frequency partition schemes. The frequency-partition ablation compares the dyadic construction of equation 2 with three alternative ways of partitioning the non-redundant spectrum into K bands. Following Section 2.1, let $M = \lfloor L/2 \rfloor + 1$ be the number of non-redundant frequency coefficients. Every scheme specifies K strictly increasing integer boundaries

$$0 = b_0 < b_1 < \cdots < b_K = M, \tag{27}$$

and assigns the k-th band to the index set $\mathcal{B}_k = \{b_{k-1}, \ldots, b_k - 1\}$, so the bands are contiguous, non-overlapping, and together cover all frequency indices.

Uniform. The boundaries are equally spaced along the frequency index axis,

$$b_i = \mathrm{round}\left(\frac{iM}{K}\right), \qquad i = 0, \ldots, K, \tag{28}$$

so each band has width $b_k - b_{k-1} \approx M/K$, differing by at most one index when M is not divisible by K.

Geometric. Given a ratio $r > 1$ (we use $r = 1.5$), cumulative boundaries are first defined on the continuous frequency-index axis,

$$\tilde{b}_i = M \frac{r^i - 1}{r^K - 1}, \qquad i = 0, \ldots, K, \tag{29}$$

and the interior boundaries are then rounded to integers: $b_0 = 0$, $b_i = \mathrm{round}(\tilde{b}_i)$ for $1 \le i < K$, and $b_K = M$. The ideal bandwidth of the k-th band is

$$\tilde{D}_k = \tilde{b}_k - \tilde{b}_{k-1} = M \frac{r^{k-1}(r-1)}{r^K - 1}, \tag{30}$$

so adjacent ideal bandwidths satisfy $\tilde{D}_{k+1}/\tilde{D}_k = r$: the bands are narrow at low frequencies and progressively wider at high frequencies.

Random. The interior boundaries are $K-1$ distinct indices sampled uniformly without replacement from $\{1, \ldots, M-1\}$, sorted in ascending order as $c_{(1)} < \cdots < c_{(K-1)}$, with $b_0 = 0$, $b_i = c_{(i)}$ for $1 \le i < K$, and $b_K = M$. The generator seed is $s_{\mathrm{rng}} = s_{\mathrm{partition}} + M$, with a default

$s_{\text{partition}} = 2026$, so the random partition is reproducible for a fixed prediction length and number of bands and does not vary with the training seed.

If rounding produces repeated interior boundaries, the scheme falls back to the uniform construction to guarantee that every band contains at least one frequency index. With the current configuration ($L = 192$, $M = 97$, $K = 6$), the boundaries are $[0, 16, 32, 48, 65, 81, 97]$ for Uniform, $[0, 5, 12, 22, 38, 62, 97]$ for Geometric, and one fixed realization $[0, 22, 27, 48, 75, 84, 97]$ for Random.

K GRADIENT DIAGNOSTICS ON FRED-MD

We examine the twelve DySAC runs on Fred-md with four backbones, seeds 2026/2027/2028, and sequence/label/prediction lengths $96/48/12$. All twelve runs completed 30 epochs. The TimeMixer records are from the original runs, not the revised training-epoch evaluation in Table 17; matching revised gradient diagnostics are not available. The diagnostics summarize training minibatches in the last completed epoch, rather than the full training trajectory or the checkpoint selected by validation loss.

For a minibatch with $K > 1$ frequency objectives, let g_k denote the original parameter gradient of task k, and let $\tilde{g}$ denote the gradient returned by the implemented merger, including its numerical fallback if invoked. The available implementation defines

$$c_{\text{pair}} = \frac{1}{K(K-1)} \sum_{i \neq j} \mathbf{1}[\langle g_i, g_j \rangle < 0] \,,$$

$$c_{\text{merge}} = \frac{1}{K} \sum_{k=1}^{K} \mathbf{1}[\langle g_k, \tilde{g} \rangle < 0] \,.$$

The first statistic uses the sign of a stabilized pairwise cosine, which has the same sign as the gradient inner product; it measures disagreement between task gradients. The second tests each task gradient against the merged gradient, before optimizer preconditioning. These statistics have different comparison objects and denominators.

Table 22: Final-epoch gradient diagnostics for DySAC on Fred-md. Each cell reports mean (population standard deviation, $ddof = 0$) across three seeds, in percent. Each seed contributes its epoch-30 minibatch average. The two columns are distinct diagnostics, not a before–after comparison of the same quantity.

Backbone	Pairwise conflict c_{pair}	Merged conflict c_{merge}
DLinear	44.00 (2.68)	0.00 (0.00)
FreTS	28.11 (7.56)	0.00 (0.00)
iTransformer	10.89 (1.29)	0.00 (0.00)
TimeMixer	17.67 (4.03)	0.00 (0.00)

Interpretation and scope. The positive pairwise fractions show that frequency-task gradients need not agree in the recorded training regime. The zero merged fractions indicate no reported negative task–merged-gradient inner products in these final-epoch summaries. Zero inner products are not counted as conflicts, so this is not evidence of strict common descent, nor a guarantee about finite Adam updates or held-out forecasting performance. The records do not contain matched diagnostics for alternative weighting rules, a fallback activation counter, or minibatch gradient vectors. Consequently, they neither establish conflict-free behavior throughout training nor identify the causal contribution of directional geometry separately from normalization and gradient scale. Metric semantics were checked against the available implementation; the exact code revision used for these runs is not archived.