When Does Decomposition Benefit Time Series Forecasting?

Konstantinos Tsikonofilos[1] (✉), Jelena Zdravkovic[1], and Panagiotis Papapetrou[1]

Department of Computer and Systems Sciences, Stockholm University, Kista 16455, Sweden {konstantinos.tsikonofilos,jelenaz,panagiotis}@dsv.su.se

Abstract. Time series forecasting often relies on decomposition-based preprocessing to improve predictive performance, yet its effectiveness varies significantly across datasets and models, and a systematic understanding of when and why it is beneficial remains lacking. In this work, we investigate the mechanisms underlying the success of decomposition using a combination of synthetic and real data. We first construct synthetic time series with known components (trend, periodicities, and noise) and show that the benefit of decomposition grows with the number of distinct, predictable periodic components and is largest when the signal is dominated by clean periodic structure rather than noise or unpredictable trends. When components must be estimated from data, a generic wavelet decomposition recovers only a fraction of the gain attained through a bandpass decomposition aligned to the signal's spectral structure. On the M4 Hourly dataset, a per-series matched bandpass decomposition improves the median series' MAE by 14.8% over a vanilla baseline, compared with 1.9% for a standard wavelet and essentially none for a naive equal-width bandpass. In contrast, on a real scale-free heart-rate signal, a wavelet decomposition is better matched to the spectrum, improving MAE by 11.6% over the vanilla baseline compared with 1.8% for an equal-width bandpass. Our findings offer a mechanistic perspective on decomposition for time series forecasting and practical guidance for matching it to a series' spectral structure.

Keywords: Time Series Forecasting · Decomposition · Synthetic Benchmarks · Spectral Structure · Wavelet Transform · Fourier Methods

1 Introduction

Decomposition has become a popular approach in time series forecasting pipelines. In its simple, univariate formulation, a time series is transformed into a set of components (e.g., trend, seasonal, and residual terms) which when combined, are able to reconstruct the original signal. This practice appears both as an explicit preprocessing step when applying a fixed decomposition before training a forecaster [11, 20], and as an architectural inductive bias in forecasting models [17]. The underlying intuition is appealing: if distinct, predictable factors of the data-generating process can be isolated and noise is removed, decomposition

may make the forecasting problem more tractable if they are simpler to model separately than in their entangled form.

Despite this intuition and empirical successes, decomposition does not yield consistent improvements. Prior empirical work has shown that gains vary substantially across datasets, decomposition methods, and forecasting models, with some combinations improving accuracy and others offering little benefit or even degrading performance [11]. This suggests that decomposition is not a universally helpful transformation, but rather, its utility depends on properties of the data, such as strength, stationarity and separability of trend or seasonal components, the amount of noise and the length of the available history, and on the model's ability to exploit the decomposed representation. A systematic understanding of *when* and *why* decomposition improves forecasting performance would be valuable by reducing the need for ad-hoc trial-and-error, guiding the choice of decomposition method for a given dataset, and clarifying which model classes are most likely to benefit from such an approach.

In this work we take a step from empirical observation towards a mechanistic understanding of when decomposition improves forecasting performance. We view decomposition primarily as a *component separation* mechanism, and hypothesize that its benefit depends on how well it isolates underlying components in the data. We therefore study performance under idealized settings with known ground-truth components and practical settings where separation is imperfect, and we connect these results to a real dataset through aligning the decomposition to its spectral structure. Our main contributions are as follows:

1. We introduce a controlled synthetic framework for studying decomposition when the ground-truth components (trend, periodic structure and noise) are known.
2. We identify factors influencing whether decomposition improves forecasting under ideal separation conditions, including the number of predictable periodic components, the distribution of signal energy across components, noise level, and trend predictability.
3. We show that under imperfect decomposition, separation quality strongly mediates performance gains obtained from decomposition.
4. We demonstrate that aligning decomposition bands with a signal's spectral structure improves the recovery of decomposition benefits in synthetic data and on the M4 Hourly dataset, where per-series matched bandpass decomposition reduces the median series MAE by 14.8% relative to the non-decomposed baseline, while for a scale-free heart-rate signal, a wavelet decomposition reduces MAE by 11.6%.
5. We translate these findings into design implications for decomposition-based forecasting pipelines, emphasizing when decomposition is likely to help and how decomposition choices should be matched to spectral structure.

To support reproducibility, all code to reproduce our experiments and figures is publicly available.[1]

[1] https://github.com/konTsik/decomp-forecasting

2 Related Work

2.1 Time series forecasting

Forecasting methods can broadly be grouped into one of two large families of models. On one end, classical statistical models including autoregressive and ARIMA [1], and exponential smoothing [2] models, typically rely on explicit parametric equations with relatively few parameters, encoding relationships between variables and timesteps. In contrast, deep learning approaches learn distributed representations over many parameters and have become increasingly prominent, particularly for large-scale and medium- to long-horizon forecasting benchmarks [14]. Across both families, using inductive biases has been shown to be beneficial for building well-performing models. Examples include autoregressive dependence, sequential processing in recurrent neural networks [4, 7], or explicit assumptions about trend and seasonal structure. Decomposition can be viewed as one such bias: instead of modeling the raw series directly, the signal is first represented through distinct components whose structure may be easier for a forecasting model to exploit [9].

2.2 Time series decomposition

Time series decomposition represents an observed series as a set of components that capture aspects of the data with different structure, such as trend, seasonality, or residual noise. In the *additive* setting considered in this work, these components sum to reconstruct the original signal.

Several families of decomposition methods have been proposed, depending on the principle that is being used to extract components. Classical forecasting-oriented methods exploit the observation that many time series contain slowly varying components, usually referred to as *trends*, together with recurrent components, usually referred to as *seasonalities*. Moving-average decompositions and STL are typical examples of this family [9]. A second family, more closely associated with signal processing, represents the signal in terms of frequency or time-frequency structure. Fourier decompositions project the signal onto sinusoidal basis functions, whereas wavelet decompositions use localized basis functions at different scales and positions [3, 15]. A third family consists of data-adaptive methods, such as empirical mode decomposition and singular spectrum analysis, which aim to extract components without relying on a fixed Fourier or wavelet basis [5, 8].

Although decomposition is often motivated as a way to reveal latent structure, the recovered components need not correspond exactly to the true data-generating factors. Imperfect separation can occur when, for example, one underlying process is split across several estimated components or multiple processes are mixed within the same component. This issue is potentially important for decomposition-based forecasting. Specifically, decomposition could be beneficial when the estimated components expose predictable structure that is easier to model than the raw series, but it may provide little benefit when the separation is poorly aligned with the signal structure.

2.3 Decomposition in time series forecasting

Several models have tried to use decomposition as a tool to exploit structure in the data and to make better forecasts. One approach is performing decomposition at the input level [11]. This views decomposition as a preprocessing step in a larger pipeline that includes forecasting each component (separately or with interactions between them) and then recombining the forecasts to construct a forecast for the original signal. This approach provides flexibility since it can be deployed in a model-agnostic function as a preprocessing step before applying any multivariate forecasting algorithm. The downside of the approach is that multiple decompositions might need to be tested to achieve good performance for each forecasting model and dataset and no method seems to clearly outperform the others [11]. Nevertheless, it has been used successfully in conjunction with a simple linear layer-based model, surpassing more sophisticated models in performance [20].

Another approach is using the decomposition as an architectural bias, with specialized blocks *learning* the decomposition from the data [19, 21, 18, 17]. While powerful and data-driven, this approach is limited in the necessity for the decomposition to be differentiable and model-specific.

2.4 Current limitations

In the light of this plethora of approaches, Kreuzer et al. [11] show that decomposition benefit depends on the dataset, forecasting model, and decomposition method. Building on it, and reusing the same linear forecasting model, we instead ask *when* and *why* decomposition improves performance. As a first step, we hypothesize that aligning the decomposition parameters with the signal's spectral structure could improve forecasting performance. A case of a decomposition not well matched to the underlying components is illustrated in Figure 1. We propose that such an imperfect separation can interact with the downstream forecaster and reduce the gains attainable from decomposition.

3 Problem Formulation

We consider a univariate time series $\mathbf{x} = (x_1, \ldots, x_T) \in \mathbb{R}^T$ and represent it as a sum of K latent components:

$$x_t = a_t + \sum_{k=1}^{K_p} p_{k,t} + n_t, \tag{1}$$

where $\mathbf{a}, \mathbf{p_k}, \mathbf{n} \in \mathbb{R}^T$ denote the trend (low-frequency), periodic/seasonal, and residual components of the time series at time point t, respectively. In real data, components must be estimated through a decomposition procedure.

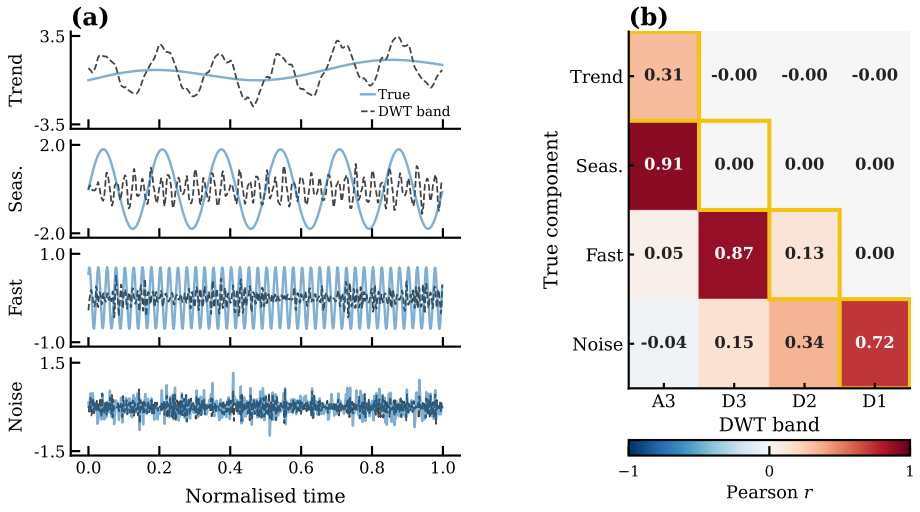

Fig. 1. Imperfect time-series decomposition via the Discrete Wavelet Transform (DWT). The observed signal is the additive sum of four known components: a slow trend, a periodic oscillation, a fast oscillation, and Gaussian noise. **(a)** Each true component (blue) overlaid with the corresponding band reconstructed from a three-level db4 DWT (dashed black). Despite applying exactly as many DWT bands as true components, none of the recovered bands closely matches its target. **(b)** Cross-correlation matrix between every true component (rows) and every DWT band (columns). A perfect decomposition would yield ones on the diagonal and zeros elsewhere. The weak diagonal values and substantial off-diagonal leakage arise because DWT bands occupy fixed dyadic (octave-width) frequency intervals that do not adapt to the spectral content of the signal.

Definition 1. *Decomposition: A decomposition method D_θ parametrized by θ, maps a time series* $\mathbf{x}$ *to a set of M estimated components*

$$D_\theta(\mathbf{x}) = \{\hat{\mathbf{c}}_1, \ldots, \hat{\mathbf{c}}_M\}, \tag{2}$$

such that

$$\mathbf{x} = \sum_{m=1}^{M} \hat{\mathbf{c}}_m \tag{3}$$

The decomposition is called perfect *if each recovered component corresponds to exactly one latent component in Equation 1.*

Definition 2. *Component Energy: Given component* $\mathbf{c}$*, we define its* energy *as follows:*

$$\mathcal{E}(\mathbf{c}) = \mathrm{Var}(\mathbf{c}). \tag{4}$$

Moreover, for a decomposition $\{\mathbf{c}_1, \ldots, \mathbf{c}_M\}$*, the* ***relative energy contribution*** *of the* m^{th} *component is*

$$r_m = \frac{Var(\mathbf{c}_m)}{\sum_{j=1}^{M} Var(c_j)}. \tag{5}$$

Based on the above definition, component energy highlights the importance of a component in terms of its contribution to the total signal variance. Experiments 2 and 3 investigate if the decomposition benefit changes as the energy distribution across trend, periodic, and noise components varies. Next, we formulate two forecasting problems, distinguishing between decomposition-based and non-decomposition-based forecasting settings.

Definition 3. *Component Predictability: Let* $\mathbf{c}$ *denote a component and let* $f^\star$ *be the optimal time series forecasting model within a model class* $\mathcal{F}$*. The* ***predictability*** *of* $\mathbf{c}$ *is defined as*

$$P(\mathbf{c}) = 1 - \frac{E(f^\star, \mathbf{c})}{\mathrm{Var}(\mathbf{c})}, \tag{6}$$

with $E(\cdot)$ *denoting the forecasting error. High values of* $P(\mathbf{c})$ *indicate that the future values of component c can be accurately inferred from its history. The normalization by* $\mathrm{Var}(\mathbf{c})$ *expresses the error relative to the component's variability, with larger values of* $P(\mathbf{c})$ *indicating higher predictability.*

Based on the above definition, the benefit from decomposition should increase when the recovered components are individually more predictable than the original data. Experiment 4 provides an assessment of this hypothesis by replacing the deterministic trend component with a random walk, while keeping the remaining signal characteristics fixed.

Problem 1. **Forecasting without decomposition:** Given an input time series segment $\mathbf{x}_{t-L+1:t}$ of length L, a vanilla baseline forecasting model $f_{vanilla} : \mathbb{R}^L \to \mathbb{R}^H$ produces a forecast $\hat{\mathbf{x}}_{t+1:t+H} = f_{vanilla}(\mathbf{x}_{t-L+1:t})$, with H being the forecasting horizon.

Problem 2. **Forecasting with decomposition:** Given a decomposition $D_\theta(\mathbf{x}) = \{\hat{\mathbf{c}}_1, \ldots, \hat{\mathbf{c}}_M\}$, a component-specific forecasting model $f^m_{dec} : \mathbb{R}^L \to \mathbb{R}^H$ is applied independently to each component $\hat{\mathbf{c}}_{m,t+1:t+H} = f^m_{dec}(\hat{\mathbf{c}}_{m,t-L+1:t})$, while the final forecast is obtained through aggregation, as follows:

$$\hat{\mathbf{x}}_{t+1:t+H} = \sum_{m=1}^{M} \hat{\mathbf{c}}_{m,t+1:t+H}. \tag{7}$$

In our experiments, $f_{vanilla}$ and f^m_{dec} are both based on linear layers, using the *G_Linear* architecture [20, 11]. The difference lies whether the model receives the raw mixture in a single layer or in independent linear layers. In the latter case the model is trained end-to-end, including all layers.

Finally, we quantify the *benefit* of a decomposition.

Definition 4. *Decomposition benefit: Let $f_{vanilla}$ denote a forecasting model trained on the original time series and f_{dec} the corresponding model trained on a decomposed representation of the time series. The* ***decomposition benefit*** *is defined as follows:*

$$\Delta(D_\theta) = 100 \cdot \frac{MAE_{vanilla} - MAE_{dec}}{MAE_{vanilla}}. \tag{8}$$

A positive value indicates that the decomposition improves forecasting performance.

The main problem studied in this paper is to characterize the conditions under which forecasting under decomposition results in lower expected error than forecasting the original time series directly. More formally, we study how the following inequality

$$\Delta(D_\theta) > 0,$$

depends on the number of latent components, their relative energy, their predictability, the amount of noise, and the alignment between D_θ and the spectral structure of the time series. More concretely, in Section 4 we consider an ideal setting with perfect access to the ground-truth components, while in Section 5 we study realistic decomposition settings, where component recovery is imperfect.

4 Perfectly Separated Synthetic Data

We study decomposition-based forecasting where the latent components are known exactly and can be recovered without estimation error. This setting allows us to isolate the factors introduced in Section 3 and assess how each factor affects the decomposition benefit $\Delta(D_\theta)$ independently of decomposition quality.

4.1 Setup

Forecasting model. We adopt the G-Linear architecture from [11, 20], which consists of one or more linear layers mapping an input window of length $L = 48$ to a forecast window of length $H = 48$. The *vanilla* baseline receives the raw time series as a single input channel, i.e., one linear layer is fitted to the observed data. The *decomposed* variant receives K separate channels (one per ground-truth component), each processed by its own linear layer. The final forecast is the sum of the per-component predictions and the model is trained end-to-end.

Training. For each configuration we generate a time series of length 5000, split it into rolling windows of input length 48 and output length 48 (stride 1), and use the first 60% for training, the next 20% for validation, and the final 20% for testing. All windows are z-normalized using the training-set statistics. We train for up to 50 epochs with the Adam optimizer (learning rate 10^{-3}) and mean squared error loss, with early stopping on the validation set (patience 10). Each experiment is repeated three times with different random seeds; we report the average MAE improvement (and for Section 5 also the root mean squared error; RMSE) over the vanilla baseline along with standard deviation.

Synthetic signal. We generate time series according to the additive model of Equation 1. Each series consists of three types of components: a low-frequency trend, a set of periodic components, and Gaussian noise. Unless stated otherwise, the trend is modeled as a sinusoid with period 200, while the periodic components are pure sinusoids with periods set as the first K primes ≥ 5. This choice ensures that the components are spectrally distinct and do not share harmonically related frequencies.

All components are normalized to zero mean and unit variance before being scaled to achieve the desired energy distribution. We define r as the fraction of total energy carried by the trend component and the remaining energy is divided equally among the periodic components. The total variance is kept constant across experiments.

Experimental design. We systematically vary four factors: the number of periodic components K, the trend energy ratio r, the noise energy fraction, and the predictability of the trend (sine vs. random walk). For each factor we hold the other variables at a default value ($K = 4$, $r = 0.3$, noise=5%, trend=sine).

4.2 Experiments

Under perfect separation, decomposition should be beneficial when the resulting components are easier to forecast than the original mixture. Thus, we formulate four hypotheses, which we test with the four experiments that follow: (H1) The decomposition benefit increases with the number of predictable latent components; (H2) The benefit depends on how signal energy is distributed across components; (H3) The benefit decreases as the proportion of unpredictable noise increases; (H4) The benefit depends on the predictability of the recovered components themselves.

Experiment 1 (H1): Number of Periodic Components We vary K from 1 to 20 while keeping $r = 0.3$ and noise energy at 5%, with a low-frequency sinusoidal trend. The vanilla model error rises steeply from 0.19 (K=1) to 0.69 (K=20), while the decomposition-aware model error stays almost constant around 0.20 (Figure 2(a)). Consequently, the MAE improvement (Figure 2(b)) grows super-linearly saturating around 72%. This suggests that the forecasting difficulty of the raw time series grows faster than linearly with the number of superimposed periodic components.

Experiment 2 (H2): Trend Energy Ratio We fix $K = 4$ and 5% noise, and sweep r from 0.05 to 0.95. At large r, the signal is almost a pure sinusoid, while at small r it is a near-even mixture of periodicities. The improvement is stable across this range, close to 10.5% (Figure 3(a)). This suggests that, under perfect separation, the forecasting difficulty depends primarily on the number and predictability of the components rather than their relative amplitudes.

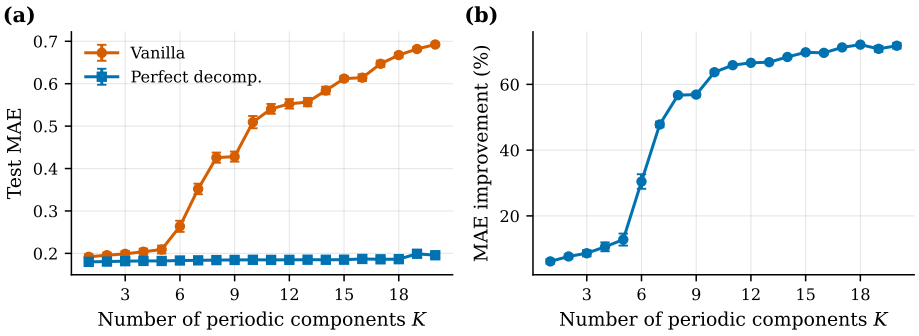

Fig. 2. Effect of the number of periodic components K. **(a)** Vanilla vs. perfect-decomposition test MAE. **(b)** The corresponding MAE improvement. Markers are means and error bars the standard deviation over three seeds, each varying both the data realization and the weight initialization.

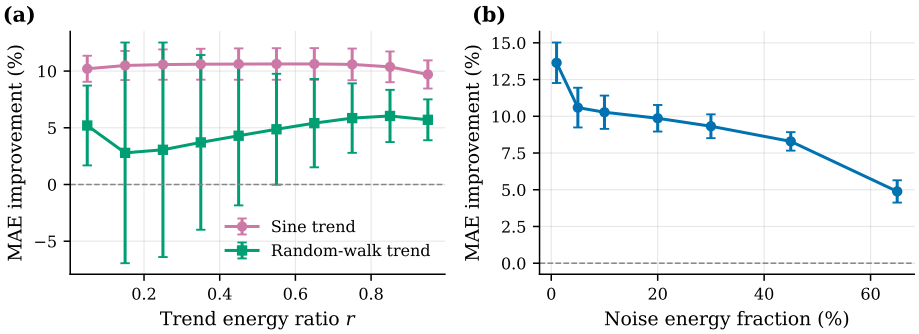

Fig. 3. Decomposition benefit under varying noise and trend predictability (K=4). **(a)** MAE improvement vs. trend energy ratio r for a sine and a random-walk trend. **(b)** MAE improvement vs. noise energy fraction (r=0.3). Means ± std over three seeds.

Experiment 3 (H3): Noise Energy With $K = 4$, $r = 0.3$, and a sine trend, we vary the noise energy from 1% to 65% of the total variance, splitting the remaining energy equally among the four periodic sines. Across this range, the benefit declines monotonically with noise (Figure 3(b)), from 13.6% at 1% noise to 10.6% at 5% and down to 4.9% at 65%.

Experiment 4 (H4): Trend Predictability We replace the low-frequency sinusoid with a random walk (normalized to unit variance, then scaled by r), keeping $K = 4$ and 5% noise. In this case, the benefit of decomposition drops to roughly 3-6% across r, well below the sine-trend curve (Figure 3(a)). This finding is consistent with Definition 3 in that the isolated random-walk trend is essentially unpredictable, and the residual benefit likely comes from the periodic components.

4.3 Summary of Findings

Across these idealized experiments, the benefit of decomposition is governed by a few consistent factors. It increases with the number of distinct periodic components (Experiment 1) and is largely insensitive to how energy is divided between trend and periodicities (Experiment 2). The advantage is largest when the signal is dominated by clean periodic structure and decreases monotonically as noise grows (Experiment 3), and the gains depend on the isolated components being predictable (Experiment 4). These results were obtained under perfect, ground-truth separation. In the next section we examine how realistic decomposition methods affect this benefit.

5 Decomposition in Practice

The perfect-separation experiments in Section 4 established the benefit of using decomposition when multiple components exist in the data. In reality, however, the true components are unknown and must be estimated from the data, and separated from noise, using actual decomposition methods. A mismatched decomposition that fails to recover the true components should yield smaller benefits than one aligned to the signal's structure.

5.1 Synthetic Data

We return to the synthetic signal with $K = 4$ periodic components (periods 5, 7, 11, 13), $r = 0.3$, and 5% white noise. We compare two single-layer-per-channel models against the vanilla baseline: a *wavelet* model receiving six components from a discrete wavelet transform (db4, level 5), and a *matched bandpass* model receiving six Fourier bands designed from the known periods: a low-frequency trend band ($f < 1/100$), one narrow (half-width equal to 40% of the distance to the nearest other periodicity, trend cutoff, or Nyquist) non-overlapping band around each known frequency $1/p$, and a residual band for the remaining spectrum. The matched bandpass represents a decomposition aligned to the signal's spectral structure as opposed to the wavelet decomposition which is not.

Figure 4d reports the MAE improvement of each model over vanilla. The wavelet model improves by 3.8% (RMSE 3.6%), whereas the matched bandpass improves by 15.4% (RMSE 14.8%). Figure 4(b,e) explains this. Measuring alignment through correlation (i.e. component *shape*, since the forecaster can rescale each channel's magnitude), wavelet decomposition spreads each periodic component across several detail bands. The matched bandpass instead places a narrow band on each known frequency, recovering a near one-to-one correspondence with the true components (Figure 4(c,f)) and confining the noise mostly to the residual band.

5.2 Real-world Demonstration

We evaluate the matched-decomposition approach on 414 hourly series from the M4 competition [13, 12], each containing 700-960 observations. We follow the

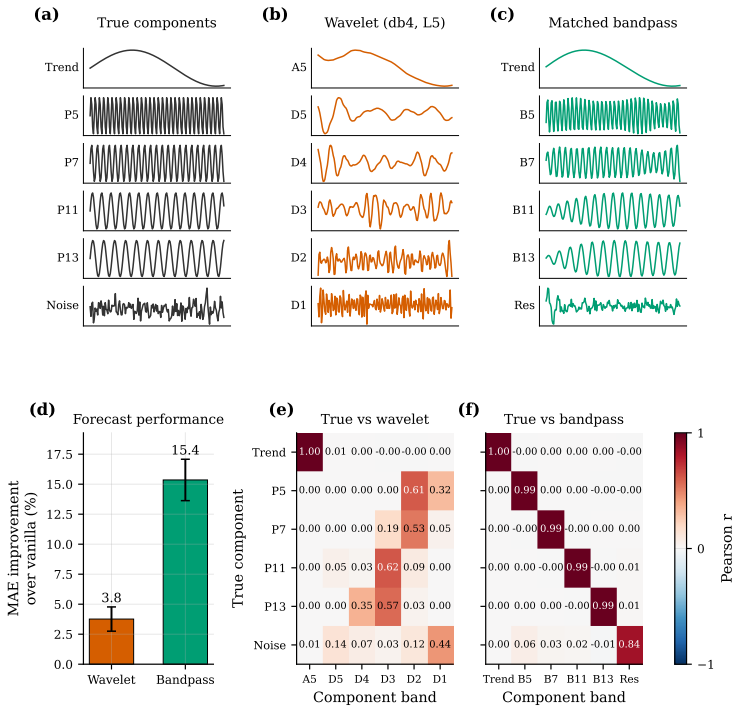

Fig. 4. Decomposition in practice (K=4, r=0.3, 5% noise). **(a-c)** True components and the components recovered by the wavelet (db4, level 5) and matched-bandpass decompositions (first 160 samples). **(d)** MAE improvement over the vanilla baseline (means ± std over three seeds). **(e, f)** Correlation between true and estimated components. The wavelet smears each periodic component across shared detail levels (P5, P7 on D2; P11, P13 on D3), whereas the matched bandpass recovers a near one-to-one correspondence and isolates the noise in its residual.

local-forecasting paradigm [16], whereby the series is treated as an independent task, and a separate model is trained using only that series' history. Each series is split into rolling windows of input length 48 and horizon length 48 (stride 1), and divided chronologically into 60% training, 20% validation, and 20% test windows. All windows are z-score normalized using the mean and standard deviation of the training windows of that series.

The vanilla model receives the raw series as a single input channel. The decomposition-based models receive multiple input channels, each processed by an independent linear layer whose outputs are summed to obtain the final forecast [11]. Because the series are forecast locally, we match the decomposition to each series individually. For every series we estimate its dominant periods from the training portion only. We compute a Hann-windowed periodogram and select prominent spectral peaks (prominence$\geq$ 5% of the maximum, periods at least 2 hours apart and up to one third of the training series length), keeping

up to $K_{\max} = 6$ peaks ordered by power. This yields K_s periods for series s. The matched bandpass then places a narrow non-overlapping band around each detected frequency, together with a low-frequency trend band (below the lowest detected period) and a high-frequency residual band, for $K_s + 2$ components in total. To attribute any difference to band *placement* rather than model capacity, the other decompositions use the same per-series channel count: the wavelet model uses a db4 transform at level $K_s + 1$ (i.e. $K_s + 2$ components), and a *naive bandpass* control uses $K_s + 2$ equal-width frequency bands, independent of the data's spectral content. All decompositions are applied globally to each series. Models are trained for up to 50 epochs with Adam optimizer [10] (learning rate 10^{-3}), MSE loss, and early stopping (patience 10).

For each model, we report the *per-series* percentage reduction in test MAE relative to the vanilla baseline. The matched bandpass achieves a median improvement of 14.8% (IQR 15.2%, median RMSE 14.9%) and reduces error on 88% of series, compared with a median of 1.9% (RMSE 2.4%) for the generic wavelet and -0.1% (RMSE 0.3%) for the naive bandpass (Figure 5a). The matched bandpass outperforms the wavelet on 86% of series (84% for RMSE), with a median per-series gap of 11.7% (Wilcoxon signed-rank $p < 10^{-36}$), and this percentage persists when restricting to the 232 series that both methods improve. The benefit of matching is largely independent of the number of detected components (Figure 5b; Spearman $\rho = 0.02, p = 0.69$), whereas the generic decompositions improve only as K grows. A potential explanation is the spectral structure of hourly data, where the daily cycle (the period detected in 99% of the cases with $K = 1$) carries a median 42% of the signal power and isolating that one dominant component with a narrow matched band attains most of the benefit. With increasing number of components, the wavelet's fixed dyadic bands and the naive equal-width bands have an increased probability of capturing that dominant component and thus their improvement could increase with K.

Because the number of channels of the matched bandpass models is tuned per series and the wavelet models are subsequently matched to control for model capacity, this absence of tuning might bias the results against the latter. To test whether this drives our result, we additionally tune the wavelet models per series by training models at every possible level with respect to the series length, select the model with the lowest *validation* MAE, and report its test MAE. The capacity-matched level was in fact optimal for only 23% of series, confirming that the main comparison under-tunes the wavelet models. Using the optimal level increases the median improvement for the wavelet models from 1.9% to 5.4%, but is nevertheless outperformed by the matched bandpass on 79% of series (median gap 8.9 percentage points; Wilcoxon signed-rank $p < 10^{-30}$), and on 81% of the 286 series that both methods improve, suggesting that spectral alignment remains the decisive factor.

5.3 Matching Without Periodic Components

The matched bandpass method improves performance by placing narrow bands on discrete spectral peaks. This suggests that when a signal has no isolated

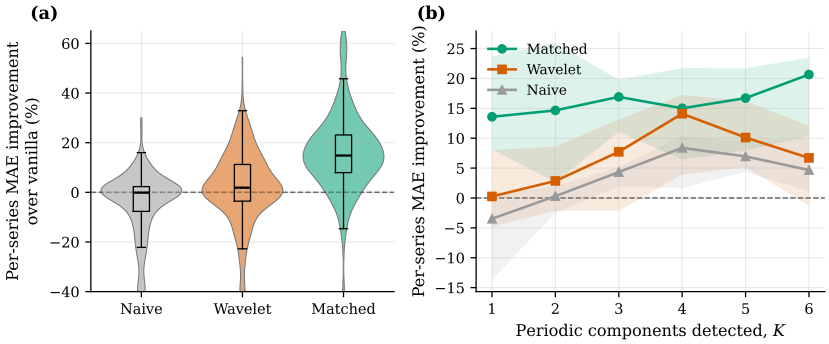

Fig. 5. Real-world demonstration on the 414 M4-Hourly series. **(a)** Per-series MAE improvement over the vanilla baseline for the naive bandpass, wavelet (db4), and per-series matched bandpass; violins show the full distribution (clipped to $[-40, 65]$% for display), boxes the median and interquartile range, dashed line the baseline. **(b)** Median improvement (shaded: IQR) versus the number of detected periodic components K; the matched bandpass benefit is large and roughly constant in K, while the generic decompositions improve only as K grows.

periodic components to match, a generic multi-scale decomposition might be preferable. We test this directly on synthetic signals with power spectral density $\propto f^{-\beta}$, generated by shaping Gaussian noise in the frequency domain ($\beta = 0$, white; $\beta = 1$, pink; $\beta = 2$, brown). Since there are no isolated periodic peaks to match with narrow bands in this setting, we compare the wavelet and naive equal-width bandpass decompositions with six components each against the vanilla baseline.

Figure 6 shows the MAE improvement of each as a function of β. For flat spectra the equal-width bandpass yields a larger improvement (15.3% vs 3.5% at $\beta = 0$), but for scale-free spectra wavelet performs better (14.5% vs 9.0% improvement at $\beta = 1$; 15.0% vs 1.1% at $\beta = 2$), suggesting that the wavelet's logarithmically-spaced bands are well matched to the self-similar structure of a 1/f spectrum, where energy is distributed across scales.

To confirm the robustness of this observation, we use an identical forecasting setup as in previous experiments, and apply the same comparison to a real scale-free signal, a publicly available univariate dataset of heart-rate (1800 samples with a sampling frequency of 2 Hz), available with the DARTS forecasting package [6]. Its power spectrum follows an approximate $f^{-1.5}$ law, and, as predicted from the synthetic experiment, a wavelet decomposition improves MAE over the vanilla baseline by 11.6% (RMSE 7.6%), as opposed to 1.8% (RMSE 1.6%) for the equal-width bandpass (Figure 6).

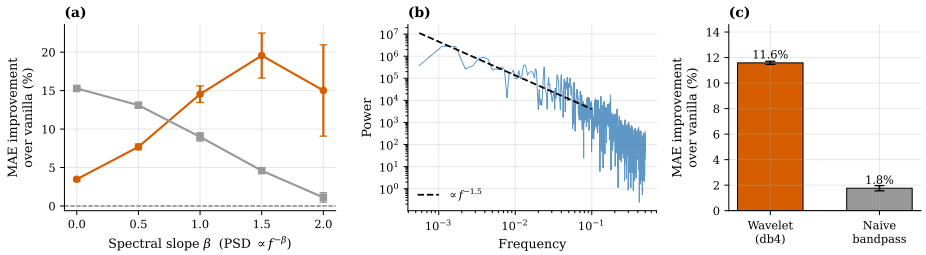

Fig. 6. Decomposition in the scale-free regime. **(a)** On synthetic $1/f^{\beta}$ signals (six components each, *means* $\pm$ *std* over three seeds). The bandpass performs better for flat spectra, while the multi-scale wavelet for scale-free ones. **(b)** A real instantaneous heart-rate recording (1800 samples) has an approximately $f^{1.5}$ spectrum (dashed), placing it in the scale-free regime. **(c)** On that signal the wavelet improves MAE by 11.6% over vanilla versus 1.8% for the equal-width bandpass, matching the synthetic prediction at comparable β. Colors are shared across panels (orange: wavelet; gray: naive bandpass).

6 Discussion

Across settings, matched bandpass gave the largest improvements on periodic data, whereas wavelet dominated the scale-free regime. Our results translate into a spectrum-based procedure for deciding whether and how to decompose time series for forecasting. As a first step, inspect the signal's power spectrum. If it shows clear discrete peaks (e.g. a strong daily or weekly cycle), place a narrow matched band at each dominant frequency. If the spectrum is scale-free, use a multi-scale wavelet decomposition instead. If the signal is instead dominated by broadband noise or by an unpredictable, stochastic trend, expect little benefit from decomposition regardless of method. In every case, bands placed blindly are not expected to provide considerable advantage over forecasting the raw series.

A principle consistent with the observed pattern of results across our experiments is that decomposition improves performance to the extent that it exposes predictable variance in a form the model can exploit. On one end, this is realized through separability, as decomposition gains increase with the number of components compared to a single linear map having access to the superimposed signal (Experiment 1). Secondly, the results from Sections 5 suggest that resolution matching could also mediate that effect, as decomposition was found to help most when its frequency bands were matched to where the predictable variance lies. Similarly, high noise levels (Experiment 3) or a stochastic trend (Experiment 4), reduced the benefit of decomposition. This also reconciles our findings on the number of components. With equal per-component variance (synthetic) the benefit grew with the number of components, whereas on M4, where the daily cycle dominates, isolating that one component captured most of the predictable variance and the benefit was stable with the number detected.

The comparison against both a wavelet and a naive bandpass decomposition strengthens our central claim that the benefit does not arise merely from pro-

viding the model with a multi-channel input, nor from using a Fourier-based decomposition specifically. Rather, it comes from aligning the frequency bands with the data's spectral structure. Furthermore, because all decomposition models in our real-data experiment share the same per-series channel count, these differences are likely attributable to band placement rather than to model capacity.

7 Conclusion

This work moves from the empirical observation that decomposition has inconsistent benefits toward a mechanistic account of when and why it does, and toward guidance practitioners can use. Using controlled synthetic signals, we isolated how decomposition benefit depends on the structure of the underlying signal. The benefit increased as more distinct predictable periodic components were superimposed, while it remained largely stable when energy was merely redistributed among already predictable components. In contrast, the benefit decreased when variance was assigned to noise or to a less predictable random-walk trend. Using estimated components, we then showed that realizing this benefit depends on aligning the decomposition with the data's spectral structure. This was achieved using a matched bandpass approach for data with clear periodicities and a wavelet approach for scale-free signals. Taken together, these results point towards decomposition providing benefits to the degree that it provides predictable variance to the forecaster, whether by separating entangled components or by matching band resolution to where that variance is concentrated in the frequency domain. Future work could relax the global-decomposition assumption by employing causal decomposition methods that avoid look-ahead bias, and extend the real-data evaluation beyond the periodic (M4-Hourly) and scale-free (heart-rate) regimes studied here to more datasets, domains, and forecasting architectures. Another promising direction is to study the interpretability benefits of decomposition-based forecasting, beyond its impact on predictive accuracy.

Disclosure of Interests. The authors have no competing interests to declare.

References

1. Box, G.E., Jenkins, G.M., Reinsel, G.C., Ljung, G.M.: Time series analysis: forecasting and control. John Wiley & Sons (2015)
2. Brown, R.G., Meyer, R.F.: The fundamental theorem of exponential smoothing. Operations Research **9**(5), 673–685 (1961)
3. Daubechies, I.: Ten lectures on wavelets. SIAM (1992)
4. Elman, J.L.: Finding structure in time. Cognitive science **14**(2), 179–211 (1990)
5. Golyandina, N., Nekrutkin, V., Zhigljavsky, A.A.: Analysis of time series structure: SSA and related techniques. CRC press (2001)

6. Herzen, J., Lässig, F., Piazzetta, S.G., Neuer, T., Tafti, L., Raille, G., Pottelbergh, T.V., Pasieka, M., Skrodzki, A., Huguenin, N., Dumonal, M., Kościsz, J., Bader, D., Gusset, F., Benheddi, M., Williamson, C., Kosinski, M., Petrik, M., Grosch, G.: Darts: User-friendly modern machine learning for time series. Journal of Machine Learning Research **23**(124), 1–6 (2022), http://jmlr.org/papers/v23/21-1177.html
7. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural computation **9**(8), 1735–1780 (1997)
8. Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.C., Tung, C.C., Liu, H.H.: The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences **454**(1971), 903–995 (1998)
9. Hyndman, R.J., Athanasopoulos, G.: Forecasting: principles and practice. OTexts (2018)
10. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
11. Kreuzer, T., Zdravkovic, J., Papapetrou, P.: Unpacking the trend: Decomposition as a catalyst to enhance time series forecasting models. Data Mining and Knowledge Discovery **39**(5), 54 (2025)
12. Makridakis, S., Spiliotis, E., Assimakopoulos, V.: The m4 competition: Results, findings, conclusion and way forward. International Journal of forecasting **34**(4), 802–808 (2018)
13. Makridakis, S., Spiliotis, E., Assimakopoulos, V.: The m4 competition: 100,000 time series and 61 forecasting methods. International Journal of Forecasting **36**(1), 54–74 (2020)
14. Makridakis, S., Spiliotis, E., Assimakopoulos, V., Semenoglou, A.A., Mulder, G., Nikolopoulos, K.: Statistical, machine learning and deep learning forecasting methods: Comparisons and ways forward. Journal of the Operational Research Society **74**(3), 840–859 (2023)
15. Mallat, S.G.: A theory for multiresolution signal decomposition: the wavelet representation. IEEE transactions on pattern analysis and machine intelligence **11**(7), 674–693 (2002)
16. Montero-Manso, P., Hyndman, R.J.: Principles and algorithms for forecasting groups of time series: Locality and globality. International Journal of Forecasting **37**(4), 1632–1653 (2021)
17. Oreshkin, B.N., Carpov, D., Chapados, N., Bengio, Y.: N-beats: Neural basis expansion analysis for interpretable time series forecasting. arXiv preprint arXiv:1905.10437 (2019)
18. Wang, S., Wu, H., Shi, X., Hu, T., Luo, H., Ma, L., Zhang, J., Zhou, J.: Timemixer: Decomposable multiscale mixing for time series forecasting. In: International conference on learning representations. vol. 2024, pp. 38626–38652 (2024)
19. Wu, H., Xu, J., Wang, J., Long, M.: Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. Advances in neural information processing systems **34**, 22419–22430 (2021)
20. Zeng, A., Chen, M., Zhang, L., Xu, Q.: Are transformers effective for time series forecasting? In: Proceedings of the AAAI conference on artificial intelligence. vol. 37, pp. 11121–11128 (2023)
21. Zhou, T., Ma, Z., Wen, Q., Wang, X., Sun, L., Jin, R.: Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting. In: International conference on machine learning. pp. 27268–27286. PMLR (2022)