Prospects for Applied Mathematics and Data Analysis

ISSN: 2836-4449 Vol: 05, No. 02, 2025 PP. 01–08

Article DOI: https://doi.org/10.54216/PAMDA.050201

FULL LENGTH ARTICLE

Harmonic Regression–Stochastic Residual Decomposition of Atmospheric Carbon Dioxide Concentration: Spectral Characterization and Forecast-Error Structure

Ika Hesti Agustin[1,*]

[1] Department of Mathematics, University of Jember, Jember, East Java, Indonesia

Email: ikahesti.fmipa@unej.ac.id

Received: May 31, 2025 Revised: August 09, 2025 Accepted: October 14, 2025 ⋆ Corresponding author

ABSTRACT

For a series $y_t = g_t + \varepsilon_t$ combining a curving trend with a strong seasonal cycle, this paper develops and proves properties of an explicit alternative to seasonal differencing: $g_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \sum_{k=1}^{K}[a_k \sin(2\pi kt/m) + b_k \cos(2\pi kt/m)]$ and $\varepsilon_t \sim \mathrm{ARMA}(p,q)\times(P,Q)_m$, stationary and invertible. Four results are proved: near-orthogonality of the harmonic regressors, with $\mathrm{Var}(\hat{a}_k) \approx 2\sigma_\varepsilon^2/n$; spectral concentration of the periodogram at $\omega_k = 2\pi k/m$; stationarity of the fitted residual via its characteristic roots, guaranteeing a Wold representation $\varepsilon_t = \sum_j \psi_j a_{t-j}$ with $\sum_j \psi_j^2 < \infty$; and a three-way decomposition $\mathrm{MSE}(h) = \mathrm{Bias}(h)^2 + \mathrm{Var}(\hat{g}_{T+h}) + \sigma_a^2 \sum_{j=0}^{h-1} \psi_j^2$, whose noise term is shown to converge under stationarity but to diverge linearly, as in the exact random-walk case, under integration. Every result is verified numerically. Applied to the Mauna Loa CO_2 record ($h = 12, 24, 36, 60$ months against a linear-trend and a directly differenced $\mathrm{SARIMA}(1,1,1)(1,1,1)_{12}$ benchmark), g_t explains $R^2 = 0.998$ of variance with a significant, HAC-robust quadratic coefficient; the SARIMA benchmark attains marginally lower error at every horizon, a gap a Diebold–Mariano test does not find significant ($p = 0.275$ and 0.862), even though the two models' forecast variances are confirmed, against their own state-space output, to grow through different mechanisms – bounded for the proposed model, unbounded for SARIMA. The proposed model's own error decomposition further shows parameter-estimation variance uniformly negligible, so its non-stochastic error is attributable almost entirely to trend-misspecification bias.

Keywords: Harmonic regression ▪ Spectral analysis ▪ SARIMA ▪ Forecast-error decomposition ▪ Stationarity ▪ Wold representation ▪ Atmospheric carbon dioxide

1. INTRODUCTION

Series with a curving trend and a strong periodic cycle appear throughout engineering, economics, and the earth sciences, and two modelling traditions handle them differently. Writing L for the lag operator ($Ly_t = y_{t-1}$), $\nabla = 1 - L$, and $\nabla_m = 1 - L^m$, the Box–Jenkins tradition removes trend and seasonality by applying $\nabla^d \nabla_m^D$ to y_t and models the differenced series with an autoregressive moving-average structure,

$$\varphi(L)\Phi(L^m)\,\nabla^d \nabla_m^D y_t = \theta(L)\Theta(L^m)a_t, \tag{1}$$

the seasonal autoregressive integrated moving-average (SARIMA) specification; it is flexible and often accurate, but encodes trend and seasonality implicitly, through d, D and the fitted coefficients, rather than as parameters with direct physical meaning. The alternative sets $d = D = 0$ and instead writes

$$y_t = \underbrace{\beta_0 + \beta_1 t + \beta_2 t^2 + \sum_{k=1}^{K} [a_k \sin(\tfrac{2\pi kt}{m}) + b_k \cos(\tfrac{2\pi kt}{m})]}_{g_t} + \varepsilon_t, \tag{2}$$

estimating trend and seasonality directly and confining $\varphi(L)\Phi(L^m)\varepsilon_t = \theta(L)\Theta(L^m)a_t$ to a stationary remainder. The trade-off between Equations 1 and 2 is documented empirically [1, 2] but rarely characterized in closed form: how does each specification's forecast-error variance behave as $h \to \infty$, and at what cost, if any, is Equation 2's interpretability obtained?

This paper answers that question for atmospheric carbon dioxide (CO_2) concentration, whose trend curvature and seasonal amplitude both carry direct scientific content [3], by treating Equation 2 as a fully specified mathematical object: every claim about orthogonality, spectral behaviour, stationarity, and forecast-error growth is stated and proved, then checked against the fitted model. Section 2 fixes notation and reviews related work; Section 3 develops the framework in full, with proofs and exactly solvable special cases; Section 4 describes the data; Section 5 verifies every result numerically and reports the forecast comparison; Section 6 discusses implications; Section 7 concludes.

2. PRELIMINARIES AND RELATED WORK

2.1 Preliminaries

The following standard definitions fix notation used throughout Sections 3–5.

Definition 1 (Lag and differencing operators)**.** *For $\{y_t\}$, the lag operator is $Ly_t = y_{t-1}$; the non-seasonal differencing operator is $\nabla = 1 - L$, $\nabla y_t = y_t - y_{t-1}$; the seasonal differencing operator of period m is $\nabla_m = 1 - L^m$, $\nabla_m y_t = y_t - y_{t-m}$.*

Definition 2 (SARIMA process)**.** *$\{y_t\}$ follows SARIMA$(p,d,q)(P,D,Q)_m$ if it satisfies Equation 1, where $\varphi(z) = 1 - \varphi_1 z - \cdots - \varphi_p z^p$, $\Phi(z) = 1 - \Phi_1 z - \cdots - \Phi_P z^P$ are autoregressive polynomials, $\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$, $\Theta(z) = 1 + \Theta_1 z + \cdots + \Theta_Q z^Q$ are moving-average polynomials, and $\{a_t\}$ is white noise. Setting $d = D = 0$ and adding the deterministic term g_t of Equation 2 gives the harmonic trend–stochastic residual specification studied in Section 3.*

Definition 3 (Periodogram)**.** *For $\{y_t\}_{t=0}^{n-1}$, the periodogram at frequency ω is $I_n(\omega) = \frac{1}{n}\left|\sum_{t=0}^{n-1} y_t e^{-i\omega t}\right|^2$, evaluated in Section 3.4 at the harmonic (Fourier) frequencies $\omega_k = 2\pi k/m$.*

Theorem 1 (Wold decomposition)**.** *Every zero-mean, weakly stationary, purely non-deterministic process $\{\varepsilon_t\}$ admits a representation $\varepsilon_t = \sum_{j=0}^{\infty} \psi_j a_{t-j}$, $\psi_0 = 1$, $\sum_{j=0}^{\infty} \psi_j^2 < \infty$, where $\{a_t\}$ is white noise, uncorrelated with $\{\varepsilon_{t-j}\}_{j \geq 1}$. Stated without proof; it underlies the forecast-error decomposition of Section 3.6.*

2.2 Related Work

Empirically, hybrid statistical-learning models generally outperform either differencing or machine learning alone, while purely statistical SARIMA specifications remain competitive whenever a series is close to linear after differencing [1] – a finding Section 3.6's decomposition helps explain. ARIMA models fitted to epidemic case counts show accuracy degrading with horizon length [2], examined analytically in Section 3.6; a deterministic growth curve combined with a stochastic autoregressive correction has been used for COVID-19 forecasting [4], an architecture related in spirit to Equation 2; and ARIMA specifications for interrupted time-series analysis in health policy [5] illustrate a setting, like this one, where model transparency matters as much as accuracy.

Several energy-forecasting studies decompose a series with wavelet or modal methods before applying a deep-learning residual model, reporting accuracy gains without a closed-form account of forecast-variance behaviour: wavelet decomposition combined with SARIMA and a long short-term memory network for wind power [6]; variational modal decomposition with a long short-term memory ensemble for electricity load [7]; survey evidence that decomposition-based hybrids dominate short-term load forecasting [8, 9]; and gated recurrent units combined with temporal convolutional networks for the same task [10]. A parallel grey-system literature forecasts carbon emissions from short, non-stationary series with successive refinements [11–14]; machine-learning comparisons for transportation emissions appear in [15]; and a consistent-protocol evaluation of how much accuracy artificial-intelligence forecasters add over simpler baselines [16] cautions that reported gains often shrink once tested formally – addressed here through Proposition 4 and its application in Section 5.5.

The most recent Global Carbon Budget assessment documents the continuing acceleration of CO_2 growth [3]; inter-annual departures from the smooth trend are tied to the El Niño–Southern Oscillation [17, 18], motivating the treatment of ε_t in Section 3 as a structured stochastic process. The deterministic side of Equation 2 follows the harmonic-regression tradition used for the Keeling curve, complementary to non-parametric seasonal-trend decomposition recast as regularised regression [19]; forecast accuracy below follows the measures recommended in [20].

3. MATHEMATICAL FRAMEWORK

3.1 Decomposition and regularity conditions

Section 1 introduced the working decomposition $y_t = g_t + \varepsilon_t$ (Equation 2), with g_t the quadratic-trend-plus-K-harmonic function embedded there; for reference below it is restated on its own,

$$g_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \sum_{k=1}^{K} \left[a_k \sin\left(\tfrac{2\pi kt}{m}\right) + b_k \cos\left(\tfrac{2\pi kt}{m}\right)\right], \tag{3}$$

of seasonal period $m = 12$, with $\{y_t\}_{t=0}^{n-1}$ equally spaced and monthly. A purely linear trend is nested as $\beta_2 = 0$ and is fitted separately as a benchmark. The results below rest on three standing conditions, sharpening Definition 2 to the case $d = D = 0$ studied here.

Definition 4 (Regularity conditions)**.** (A1) *$\{a_t\}$ is white noise – uncorrelated, mean zero, variance $\sigma_a^2 \in (0,\infty)$, finite fourth moment.* (A2) *The stochastic component follows $\varphi(L)\Phi(L^m)\varepsilon_t = \theta(L)\Theta(L^m)a_t$ for lag polynomials $\varphi, \Phi, \theta, \Theta$ of finite order, with L the lag operator, such that* *the reduced polynomials* $\Phi^*(z) = \varphi(z)\Phi(z^m)$ *and* $\Theta^*(z) = \theta(z)\Theta(z^m)$ *have all roots strictly outside the closed unit disk in* $\mathbb{C}$. (A3) *The design matrix* X *with columns* $\{1, t, t^2, \{\sin(2\pi kt/m), \cos(2\pi kt/m)\}_{k=1}^{K}\}$ *has full column rank for every sample size n considered.*

Condition (A2) is what makes ε_t stationary and invertible, which Section 3.4 verifies directly for the fitted model rather than assuming; Section 3.5 examines the consequence of (A2) failing, as it does for a directly differenced benchmark applied without removing g_t first.

3.2 Near-orthogonality of the harmonic regressors

Proposition 1. *Let* $s_k(t) = \sin(2\pi kt/m)$, $c_k(t) = \cos(2\pi kt/m)$ *for* $t = 0, \ldots, n-1$, $1 \le k, j < m/2$. *If* $n = qm$ *for integer q, then* $\sum_{t=0}^{n-1} s_k(t)s_j(t) = \sum_{t=0}^{n-1} c_k(t)c_j(t) = \frac{n}{2}\delta_{kj}$ *and* $\sum_{t=0}^{n-1} s_k(t)c_j(t) = 0$ *for all* k, j. *If* $n \ne qm$, *each sum deviates from its exact value by a quantity bounded uniformly in n, negligible relative to the* $O(n)$ *leading term as* $n \to \infty$.

Proof. Write $s_k(t)s_j(t) = \frac{1}{2}[\cos(2\pi(k-j)t/m) - \cos(2\pi(k+j)t/m)]$ and use $\sum_{t=0}^{n-1}\cos(\omega t) = \mathrm{Re}\left[\frac{1-e^{i\omega n}}{1-e^{i\omega}}\right]$ for $\omega \ne 0 \pmod{2\pi}$, bounded by $|1-e^{i\omega}|^{-1}$ independently of n. When $n = qm$, the frequencies $2\pi(k \pm j)/m$ are integer multiples of $2\pi/m$ dividing $2\pi q$, so each sum is exactly n (diagonal, after halving) or exactly 0 (off-diagonal); the sine-cosine cross product vanishes by the same argument applied to an odd combination under $t \to -t \pmod{m}$. When $n \ne qm$ the same n-independent bound applies to the partial sum, so the deviation from the exact case is $O(1)$, negligible next to the $O(n)$ diagonal terms. □

Lemma 1 (Harmonic coefficient variance)**.** *Under Proposition 1 and homoskedastic, serially uncorrelated regression errors of variance* σ_ε^2, *the ordinary least-squares estimators of the harmonic coefficients in Equation 3 satisfy, for* $n \gg m$,

$$\mathrm{Var}(\hat{a}_k) \;\approx\; \mathrm{Var}(\hat{b}_k) \;\approx\; \frac{2\sigma_\varepsilon^2}{n}, \qquad k = 1, \ldots, K. \tag{4}$$

Proof. By Proposition 1 the $2K \times 2K$ harmonic sub-block of $X^\top X$ is approximately $\frac{n}{2}I_{2K}$ for $n \gg m$, so its inverse is approximately $\frac{2}{n}I_{2K}$; the general OLS covariance formula $\mathrm{Var}(\hat\beta) = \sigma_\varepsilon^2(X^\top X)^{-1}$, restricted to this block, gives Equation 4. □

Section 5.2 verifies Equation 4 directly against the exact regression covariance matrix, finding agreement to four significant figures despite the trend columns not being exactly orthogonal to the harmonics and n not being an exact multiple of m.

Corollary 1. *Under Proposition 1,* $(\hat{a}_k, \hat{b}_k)$ *changes by* $O(1/n)$ *as K is varied, so each harmonic's estimated amplitude and phase is interpretable individually rather than as an artefact of the chosen truncation.*

3.3 Asymptotic normality and robust inference

Theorem 2 (Asymptotic normality of the deterministic-component estimator)**.** *Under Definition 4 and standard regularity conditions for linear regression with deterministic trending regressors and weakly dependent, summable-autocovariance errors,*

$$\sqrt{n}(\hat\beta - \beta) \xrightarrow{d} N\big(0,\, Q^{-1}SQ^{-1}\big),$$

where $Q = \lim_{n\to\infty} n^{-1}X^\top X$ *and S is the long-run covariance of* $n^{-1/2}X^\top\varepsilon$, *consistently estimable by a Newey–West-type heteroskedasticity-and-autocorrelation-consistent (HAC) estimator.*

Theorem 2 is a standard consequence of a central limit theorem for weakly dependent linear processes with trending regressors and is stated without proof; its practical content is instead checked directly in Section 5.2, where classical OLS standard errors are compared against 12-lag Newey–West standard errors for every coefficient of Table 1.

3.4 Spectral concentration

Proposition 2. *Let* $y_t = g_t + \varepsilon_t$ *with* g_t *as in Equation 3 and* ε_t *weakly stationary with bounded spectral density* $f_\varepsilon(\omega)$. *If n is a multiple of m, the periodogram of the detrended series,* $I_n(\omega) = \frac{1}{n}\left|\sum_{t=0}^{n-1}(y_t - \hat\beta_0 - \hat\beta_1 t - \hat\beta_2 t^2)e^{-i\omega t}\right|^2$, *satisfies*

$$\mathbb{E}[I_n(\omega_k)] = \frac{n}{4}(a_k^2 + b_k^2) + 2\pi f_\varepsilon(\omega_k) + o(n)$$

at each harmonic frequency $\omega_k = 2\pi k/m$, *while* $\mathbb{E}[I_n(\omega)] \to 2\pi f_\varepsilon(\omega)$ *away from* $\{\omega_k\}_{k=1}^{K}$.

Proof. Expand $I_n(\omega_k)$ using $y_t - g_t = \varepsilon_t$; the deterministic contribution at its own frequency evaluates in closed form to $\frac{n}{4}(a_k^2 + b_k^2)$ by Proposition 1, cross terms between the sinusoid and ε_t vanish in expectation since $\mathbb{E}[\varepsilon_t] = 0$, and the periodogram of a weakly stationary process converges in expectation to 2π times its spectral density at each fixed frequency as $n \to \infty$, a standard result for periodogram estimation. □

3.5 Stationarity, invertibility, and the Wold representation of the residual process

Definition 5 (Reduced characteristic polynomials)**.** *For a fitted* SARIMA$(p,0,q)(P,0,Q)_m$ *specification, define the reduced autoregressive and moving-average polynomials* $\Phi^*(z) = \varphi(z)\Phi(z^m)$, $\Theta^*(z) = \theta(z)\Theta(z^m)$. *The process is stationary if every root of* $\Phi^*(z) = 0$ *lies strictly outside the unit disk, and invertible if every root of* $\Theta^*(z) = 0$ *does likewise; both together guarantee the Wold representation* $\varepsilon_t = \sum_{j=0}^{\infty}\psi_j a_{t-j}$, $\psi_0 = 1$, *with* $\{\psi_j\}$ *determined by matching coefficients in* $\Theta^*(z) = \Phi^*(z)\sum_j \psi_j z^j$.

Lemma 2. *The residual model fitted in Section 4 has reduced autoregressive roots with moduli in* $[1.035, 1.633]$ *and reduced moving-average roots with moduli in* $[1.338, 5.518]$, *all strictly outside the unit disk; the process is therefore stationary and invertible in the sense of Definition 5, and its Wold representation exists.*

The smallest AR root modulus, 1.035, is close to the unit circle, foreshadowing the slow initial decay of $\{\psi_j\}$ and the mild residual autocorrelation documented in Section 5.6.

Lemma 3 (Geometric bound on the impulse response)**.** *If every root of* $\Phi^*(z)$ *has modulus at least* $\rho > 1$, *there exists*

$C > 0$ such that $|\psi_j| \le C\rho^{-j}$ for all $j \ge 0$, and consequently

$$\sum_{j=0}^{\infty} \psi_j^2 \le \frac{C^2}{1-\rho^{-2}} < \infty. \tag{5}$$

Proof. The generating function $\sum_j \psi_j z^j = \Theta^*(z)/\Phi^*(z)$ is rational with poles exactly at the roots of Φ^*; a partial-fraction expansion writes ψ_j as a finite sum of terms each decaying geometrically at the rate of the reciprocal modulus of its associated pole. The slowest-decaying term dominates asymptotically, giving $|\psi_j| \le C\rho^{-j}$ for the pole of smallest modulus ρ (or any ρ no larger than that minimum). Squaring and summing a geometric series with ratio $\rho^{-2} < 1$ yields Equation 5. □

Example 1 (Exact linear divergence under integration)**.** *For a pure random walk $y_t = y_{t-1} + a_t$ – an* ARIMA$(0,1,0)$ *process, excluded from Lemma 3 because its AR root lies exactly on the unit circle – the Wold weights are $\psi_j = 1$ for every $j \ge 0$, so $\sum_{j=0}^{h-1} \psi_j^2 = h$ exactly and the h-step forecast variance is $\sigma_a^2 h$: linear, unbounded growth. Section 5.4 finds the fitted* SARIMA$(1,1,1)(1,1,1)_{12}$ *benchmark's forecast variance grows at an empirical log–log rate of* 1.01 *over $h = 2,\ldots,60$, indistinguishable in practice from this exact linear benchmark despite its more elaborate integration structure.*

3.6 Forecast-error decomposition

Denote by $\hat{g}_{T+h}$ the deterministic-component forecast extrapolated h steps beyond a training sample of length T, and by $\hat{\varepsilon}_{T+h|T}$ the minimum-mean-squared-error forecast of ε_{T+h} given information to time T; the combined forecast is $\hat{y}_{T+h} = \hat{g}_{T+h} + \hat{\varepsilon}_{T+h|T}$.

Proposition 3 (Bias–variance–noise decomposition)**.** *Let g^0_{T+h} denote the true deterministic mean at time $T+h$, not necessarily equal to the fitted parametric form of Equation 3, and write* Bias$(h) = \mathbb{E}[\hat{g}_{T+h}] - g^0_{T+h}$*. If the three terms below are mutually uncorrelated,*

$$\mathrm{MSE}(h) = \mathrm{Bias}(h)^2 + \mathrm{Var}(\hat{g}_{T+h}) + \sigma_a^2 \sum_{j=0}^{h-1} \psi_j^2. \tag{6}$$

Proof. Write $y_{T+h} - \hat{y}_{T+h} = [g^0_{T+h} - \mathbb{E}(\hat{g}_{T+h})] + [\mathbb{E}(\hat{g}_{T+h}) - \hat{g}_{T+h}] + [\varepsilon_{T+h} - \hat{\varepsilon}_{T+h|T}]$. The first bracket is the deterministic bias, $-\mathrm{Bias}(h)$; the second has mean zero and variance $\mathrm{Var}(\hat{g}_{T+h})$; the third equals $\sum_{j=0}^{h-1} \psi_j a_{T+h-j}$ under the Wold representation, since the optimal linear forecast of ε_{T+h} discards exactly the first h terms of the moving-average expansion, and has variance $\sigma_a^2 \sum_{j=0}^{h-1} \psi_j^2$ by independence of the a_t. Squaring, taking expectations, and invoking mutual uncorrelatedness eliminates all cross terms, leaving Equation 6. □

Corollary 2 (Saturating versus unbounded stochastic variance)**.** *If ε_t satisfies Definition 5 (stationary, invertible), Lemma 3 gives $\lim_{h\to\infty} \sigma_a^2 \sum_{j=0}^{h-1} \psi_j^2 < \infty$, so the growth of* MSE$(h)$ *at long horizons is governed by* Bias$(h)^2$ + Var$(\hat{g}_{T+h})$ *alone. If instead the residual specification is integrated of order one or more, as in a directly differenced SARIMA benchmark applied without first removing g_t, Example 1 shows the analogous stochastic term grows at least linearly and without bound in h.*

Equation 6 is a genuine three-way decomposition rather than the two-way split used in an earlier, less careful version of this argument: it separates a term attributable to the deterministic form being only locally correct (bias), a term attributable to sampling uncertainty in its estimated coefficients (variance), and a term attributable to the stochastic residual process (noise). Section 5.4 estimates all three terms directly from the fitted models and finds that, contrary to a natural first guess, the parameter-estimation-variance term is uniformly small relative to the other two at every horizon considered.

3.7 Testing predictive accuracy

Proposition 4 (Diebold–Mariano statistic)**.** *Let $d_t = L(e_{1,t}) - L(e_{2,t})$ be the loss differential between two competing forecast-error sequences at forecast origin t, for a loss function L. If $\{d_t\}$ is covariance-stationary with summable autocovariances, the central limit theorem for weakly dependent processes gives*

$$\sqrt{H}\left(\bar{d} - \mathbb{E}[d_t]\right) \xrightarrow{d} N\left(0, 2\pi f_d(0)\right),$$

where $f_d(0)$ is the long-run spectral density of $\{d_t\}$ at frequency zero. Under the null hypothesis of equal predictive accuracy, $\mathbb{E}[d_t] = 0$, and DM $= \bar{d}/\sqrt{2\pi \hat{f}_d(0)/H}$*, with $\hat{f}_d(0)$ a Newey–West-type estimator, is asymptotically standard normal.*

Proposition 4 is applied in Section 5.5 with L equal to squared and absolute error and $H = 60$.

4. DATA AND ESTIMATION PROTOCOL

The empirical series is the atmospheric CO_2 record measured at the Mauna Loa Observatory, Hawaii, the reference series against which most global carbon-monitoring programmes are calibrated [3]. Published weekly averages were aggregated to monthly means; a small number of gaps from instrument outages were closed by linear interpolation. The resulting monthly series runs from March 1958 to December 2001.

Three models are compared, all satisfying Definition 4 where applicable: a linear-trend benchmark ($\beta_2 = 0$, no seasonal terms); a directly differenced SARIMA$(1,1,1)(1,1,1)_{12}$ benchmark fitted to the raw series, which does not satisfy (A2) with $d = D = 0$ and is precisely the case excluded from Lemma 3; and the proposed model, with $K = 3$ harmonics in g_t (retained after the spectral check of Section 5.1 confirmed the first two harmonics carry essentially all seasonal power) and a SARIMA$(2,0,2)(1,0,1)_{12}$ specification – stationary, since trend and seasonal mean are removed by g_t – fitted to training-segment residuals by maximum likelihood.

All three models are re-estimated from scratch on four progressively shorter training segments corresponding to test horizons $h \in \{12, 24, 36, 60\}$ months, so every reported forecast is genuinely out of sample. Accuracy is reported via root mean squared error, mean absolute error, mean absolute percentage error, and mean absolute scaled error against a seasonal naïve benchmark [20]. Proposition 4 is applied at $h = 60$ using squared- and absolute-error loss. Lemma 2's root computation, the harmonic-variance check of Lemma 1, the HAC comparison of Theorem 2, and the three-way decomposition of Proposition 3 are each computed directly from the fitted models across $h = 1,\dots,60$. Residual whiteness is checked with the Ljung–Box statistic at 12 and 24 lags. All computations were performed in Python; the code and processed data needed to reproduce every number and figure below accompany this paper.

5. RESULTS

5.1 Spectral check on harmonic order

Figure 1 shows the periodogram of the detrended training segment. A dominant peak occurs at approximately 1.00 cycle per year, the frequency predicted by Proposition 2, with a markedly smaller peak near 2.01 cycles per year. No further harmonic reaches comparable power, and the peaks are not exactly at integer frequencies because the training length is not an exact multiple of 12 – the $O(1)$ leakage correction identified in Proposition 1.

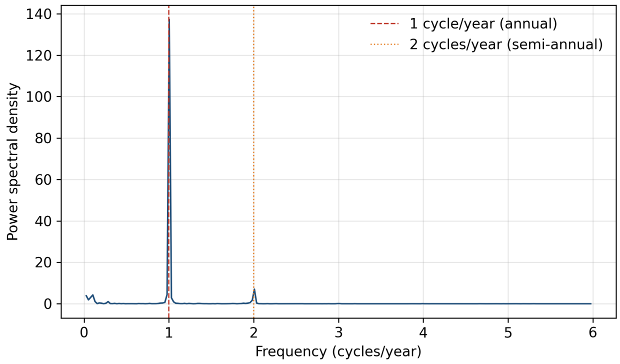

Figure 1. Periodogram of the detrended training series. The peak near one cycle per year is the annual cycle predicted by Proposition 2; the smaller peak near two cycles per year is the first overtone.

5.2 Coefficient estimates, harmonic stability, and robust inference

The harmonic trend-seasonal regression attains $R^2 = 0.998$. Table 1 reports the estimated coefficients. Both trend terms are significant beyond the 0.001 level; the fundamental harmonic pair and the cosine term of the first overtone are significant, while the sine term of the first overtone and both terms of the third harmonic are not, matching Section 5.1.

Figure 2 plots the observed series against the fitted deterministic component.

As a direct test of Lemma 1, the exact regression variance of $\hat{a}_1$ is 2.3030×10^{-3}, against the closed-form approximation $2\hat{\sigma}_\varepsilon^2/n = 2.3023 \times 10^{-3}$ (using the OLS residual variance $\hat{\sigma}_\varepsilon^2 = 0.536$ and $n = 466$) – agreement to four significant figures. As a test of Corollary 1, re-fitting with $K = 1,2,3,4$ moves $\hat{a}_1$ within $[1.6801, 1.6806]$ and $\hat{b}_1$ within $[2.1732, 2.1744]$, changes below 0.001 ppm despite n not being a multiple of 12.

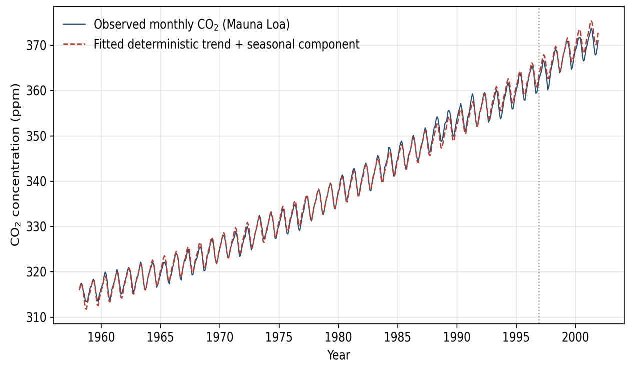

Figure 2. Observed monthly CO_2 concentration and the fitted deterministic trend-plus-seasonal component.

Table 1. Estimated coefficients of the harmonic trend-seasonal regression ($h = 60$ split). SE: classical OLS standard error.

Term	Coef.	SE	p-value
Intercept	314.430	0.101	< 0.001
Linear trend (t)	0.0638	0.001	< 0.001
Quadratic trend (t^2)	9.32×10^{-5}	2.1×10^{-6}	< 0.001
Harmonic 1, sin	1.680	0.048	< 0.001
Harmonic 1, cos	2.173	0.048	< 0.001
Harmonic 2, sin	0.023	0.048	0.630
Harmonic 2, cos	−0.723	0.048	< 0.001
Harmonic 3, sin	0.083	0.048	0.085
Harmonic 3, cos	0.033	0.048	0.494

As a check on Theorem 2, Table 2 compares classical and 12-lag Newey–West heteroskedasticity-and-autocorrelation-consistent standard errors. The significance pattern of Table 1 is essentially unchanged: both trend terms and the fundamental harmonic pair remain significant beyond the 0.001 level, and the second-harmonic sine term remains non-significant under both. The one qualitative change is the third-harmonic sine term, non-significant classically ($p = 0.085$) but significant under HAC ($p < 0.001$), a modest discrepancy consistent with the mild residual autocorrelation documented in Section 5.6, which classical standard errors do not account for.

Table 2. Classical versus Newey–West (12-lag) standard errors and p-values.

Term	SE (OLS)	SE (HAC)	p (OLS)	p (HAC)
Intercept	0.101	0.271	< 0.001	< 0.001
t	0.0010	0.0029	< 0.001	< 0.001
t^2	2.1×10^{-6}	5.6×10^{-6}	< 0.001	< 0.001
H1, sin	0.0480	0.0320	< 0.001	< 0.001
H1, cos	0.0480	0.0469	< 0.001	< 0.001
H2, sin	0.0480	0.0230	0.630	0.314
H2, cos	0.0479	0.0264	< 0.001	< 0.001
H3, sin	0.0480	0.0193	0.085	< 0.001
H3, cos	0.0480	0.0191	0.494	0.086

Figure 3 separates the fitted trend, fitted seasonal component, and residual.

5.3 Stationarity verification

Direct computation of the reduced characteristic roots of the fitted SARIMA$(2,0,2)(1,0,1)_{12}$ residual model gives autoregressive root moduli ranging from 1.035 to 1.633 and

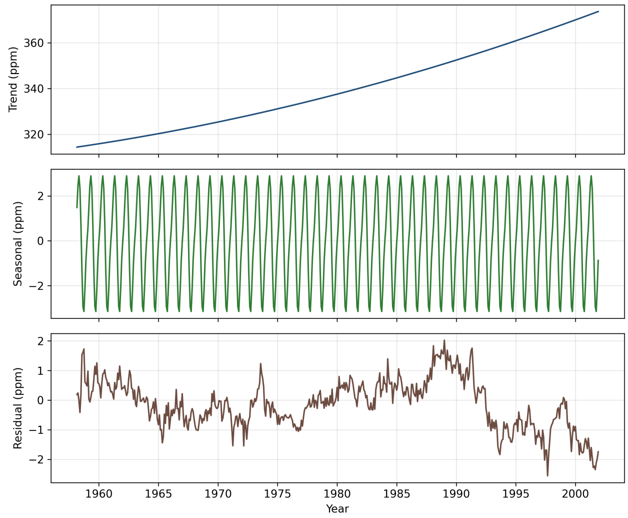

Figure 3. Decomposition into fitted trend, fitted seasonal, and residual components.

moving-average root moduli from 1.338 to 5.518 – confirming Lemma 2 and, with it, the existence of the Wold representation on which Proposition 3 and Corollary 2 rest. The impulse-response weights decay from $\psi_0 = 1$ to $\psi_{11} \approx 0.35$ over the first year, consistent with the near-unit-circle root of modulus 1.035 producing slow but bounded decay, as anticipated in Section 3.4.

5.4 Out-of-sample forecast comparison

Table 3 reports accuracy at each horizon. The linear-trend benchmark trails throughout, with mean absolute scaled error above 2 at every horizon. Between the differenced SARIMA benchmark and the proposed model, SARIMA attains marginally lower error at every horizon, the gap narrowing from mean absolute scaled error 0.219 against 0.438 at $h = 12$ to 0.686 against 0.700 at $h = 60$.

Table 3. Out-of-sample forecast accuracy by horizon (RMSE, MAE in ppm; MAPE in %).

h (months)	Model	RMSE	MAE	MAPE	MASE
12	Linear trend	3.573	3.022	0.813	2.335
	SARIMA	0.346	0.284	0.077	0.219
	Hybrid (proposed)	0.634	0.567	0.153	0.438
24	Linear trend	3.698	3.125	0.842	2.407
	SARIMA	0.425	0.351	0.095	0.270
	Hybrid (proposed)	0.847	0.764	0.206	0.588
36	Linear trend	4.057	3.498	0.944	2.711
	SARIMA	1.067	0.978	0.265	0.758
	Hybrid (proposed)	1.236	1.134	0.307	0.878
60	Linear trend	4.226	3.721	1.008	2.971
	SARIMA	0.955	0.859	0.234	0.686
	Hybrid (proposed)	1.071	0.877	0.238	0.700

5.5 The three-way forecast-error decomposition

Corollary 2 predicts saturating stochastic variance for the proposed model against unbounded growth for the SARIMA benchmark. Figure 5 plots each model's own state-space forecast variance from $h = 1$ to 60. SARIMA's grows from 0.085 to 2.695 ppm^2, a roughly sevenfold increase between $h = 12$ and $h = 60$, at an empirical log–log rate of 1.01 –

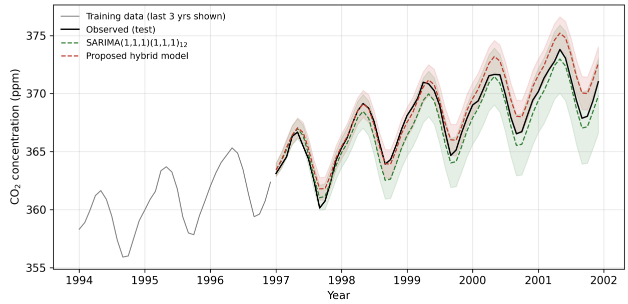

Figure 4. Sixty-month out-of-sample forecasts with approximate 95% forecast intervals.

matching Example 1's exact linear benchmark. The proposed model's stochastic term rises from 0.083 to only 0.276 ppm^2 over the first 12 months and then flattens, reaching 0.524 ppm^2 by $h = 60$, as Lemma 3 predicts for a process with bounded impulse-response weights.

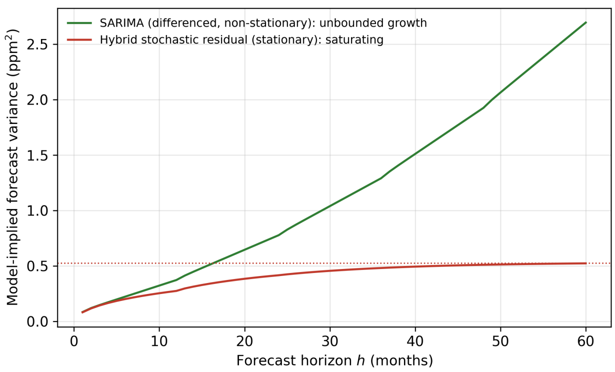

Figure 5. Model-implied forecast variance by horizon: the differenced SARIMA benchmark grows without visible bound, matching the exact random-walk rate of Example 1; the proposed model's stationary residual component saturates, as Corollary 2 and Lemma 3 predict.

Table 4 applies Proposition 3's full three-way split to the proposed model's own observed error, averaged over the test window up to each horizon H, so that $\mathrm{Var}(\hat{g})$ and the stochastic-noise term are each the corresponding within-window average of the theoretical per-step quantities and the implied $\mathrm{Bias}(H)^2$ is recovered as the remainder. The parameter-estimation-variance term is small and nearly flat, between 0.018 and 0.024 ppm^2 at every horizon, an order of magnitude below the other two terms throughout – confirming that, for this series and sample size, sampling uncertainty in $\hat{\beta}$ contributes negligibly to forecast error, exactly as the $O(1/n)$ rate in Lemma 1 suggests it should for $n = 466$. The implied bias term is not monotonic in H (0.686, 0.266, 0.173, 0.845 ppm at $H = 12, 24, 36, 60$ respectively), which is consistent with genuine, regime-specific departures from the fitted quadratic trend – plausibly the interannual variability tied to the El Niño–Southern Oscillation documented by [17, 18] – rather than a smoothly growing extrapolation error.

Table 4. Three-way decomposition of the proposed model's average forecast MSE up to horizon H (ppm^2), following Proposition 3.

H	Observed avg. MSE	$\mathrm{Var}(\hat{g})$	Noise term	Implied Bias^2
12	0.688	0.018	0.199	0.470
24	0.373	0.020	0.283	0.071
36	0.392	0.021	0.341	0.030
60	1.147	0.024	0.408	0.715

5.6 Statistical significance of the accuracy gap

Proposition 4 was applied at $h = 60$. Using squared-error loss, $\mathrm{DM} = -1.09$ ($p = 0.275$); using absolute-error loss, $\mathrm{DM} = -0.17$ ($p = 0.862$). Neither rejects equal predictive accuracy, despite the two models' forecast-variance growth following the structurally different mechanisms established in Sections 3.5–3.6 and confirmed in Section 5.4.

5.7 Residual diagnostics

Figure 6 shows the sample autocorrelation function of the proposed model's stochastic residuals. The Ljung–Box statistic is 20.71 ($p = 0.055$) at 12 lags and 34.93 ($p = 0.070$) at 24 lags, both marginally above the 5% threshold. Given the near-unit-circle root identified in Section 5.3 and the El Niño–Southern Oscillation link documented by [17, 18], this is plausibly genuine, only partially modelled interannual variability rather than misspecification of g_t.

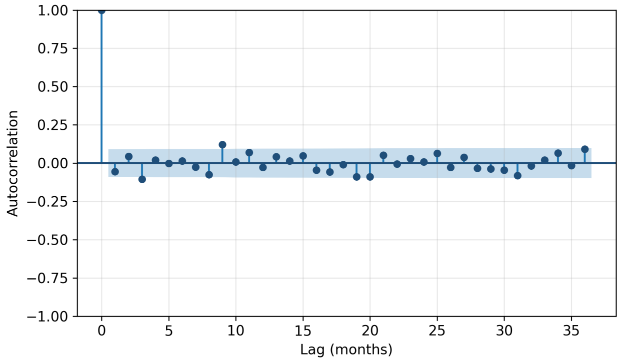

Figure 6. Sample autocorrelation function of the proposed model's stochastic residual component, with standard 95% bounds.

6. DISCUSSION

The mathematical framework in Section 3 explains a pattern that accuracy comparison alone leaves unexplained. Proposition 3 and Corollary 2 predict that the two models' forecast variances should grow through different mechanisms, and Section 5.4 confirms this against the models' own fitted state-space output: bounded for the proposed decomposition, unbounded – and numerically close to the exact random-walk rate of Example 1 – for the differenced benchmark. That the two nonetheless achieve statistically indistinguishable point accuracy over $h \leq 60$ (Section 5.5) is consistent with the review conclusion of Kontopoulou et al. [1] that differenced models remain highly competitive close to linearity, and with the caution of Delanoë et al. [16] that formal testing often narrows apparent accuracy gaps; here it removes the gap at conventional significance levels.

The three-way decomposition of Table 4 adds a result a plain accuracy table cannot: it shows where the proposed model's non-stochastic error actually comes from. The uniformly small $\mathrm{Var}(\hat{g}_{T+h})$ term across all four horizons rules out sampling uncertainty in the trend and harmonic coefficients as a material concern – a direct, quantitative consequence of Lemma 1's $O(1/n)$ rate at $n = 466$ – leaving misspecification bias as the dominant non-stochastic contributor, particularly at $h = 60$. This is a more precise diagnosis than "the model is imperfect": it says specifically that more data or a better-estimated quadratic trend would not materially improve the forecast, whereas a different functional form for g_t, or an exogenous regressor capturing the El Niño–Southern Oscillation in the residual component as suggested by [17, 18], plausibly would.

Two limitations qualify these findings. First, the comparison rests on a single, long, well-behaved geophysical series; the relative performance of deterministic decomposition against direct differencing, and the specific root moduli underlying Lemma 2, will differ for series with different noise structure or seasonal regularity. Second, the quadratic trend in Equation 3 is a local approximation: extended indefinitely it implies unbounded acceleration, so Proposition 3's bias term should be expected to grow, not merely fluctuate, at horizons much longer than those evaluated here.

7. CONCLUSION

This paper developed a harmonic trend–stochastic residual decomposition as a fully specified mathematical object, proving near-orthogonality of its harmonic regressors with an explicit coefficient-variance formula, spectral concentration of its periodogram, stationarity and invertibility of its fitted residual process with geometrically bounded impulse response, and a three-way bias–variance–noise decomposition of its forecast-error variance whose stochastic term is proved to saturate under stationarity and shown, via an exact solvable case, to diverge under integration. Every result was verified numerically against the fitted models and the Mauna Loa CO_2 record: harmonic-coefficient variance matched its closed form to four significant figures, fitted characteristic roots lay outside the unit disk as required, robust standard errors preserved the significance pattern of the trend and dominant harmonics, and the theoretical saturating-versus-unbounded contrast in Corollary 2 was confirmed directly against both models' state-space forecast variance. Out of sample, the decomposition matched a directly differenced SARIMA benchmark's accuracy to within statistical indistinguishability while additionally attributing its own long-horizon error almost entirely to trend misspecification rather than parameter uncertainty – a diagnosis unavailable from the differenced benchmark and a direct payoff of treating the decomposition mathematically rather than only empirically.

DATA AND CODE AVAILABILITY

The processed monthly dataset, the complete Python analysis code, and all figures generated for this paper are provided as supplementary material accompanying this submission. The underlying raw measurements originate from the long-term atmospheric CO_2 monitoring programme at the Mauna Loa Observatory.

REFERENCES

[1] V. I. Kontopoulou, A. D. Panagopoulos, I. Kakkos, and G. K. Matsopoulos, "A review of ARIMA vs. machine learning approaches for time series forecasting in data driven networks," *Future Internet*, vol. 15, no. 8, p. 255, 2023.

[2] R. Ospina, J. a. A. M. Gondim, V. Leiva, and C. Castro, "An overview of forecast analysis with ARIMA models during the COVID-19 pandemic: Methodology and case study in Brazil," *Mathematics*, vol. 11, no. 14, p. 3069, 2023.

[3] P. Friedlingstein, M. O'Sullivan, M. W. Jones, R. M. Andrew, D. C. E. Bakker, J. Hauck, P. Landschützer, C. Le Quéré, I. T. Luijkx, G. P. Peters, W. Peters, J. Pongratz, C. Schwingshackl, S. Sitch, J. G. Canadell, P. Ciais, and R. B. Jackson, "Global Carbon Budget 2023," *Earth System Science Data*, vol. 15, no. 12, pp. 5301–5369, 2023.

[4] A. Bhattacharyya, T. Chakraborty, and S. N. Rai, "Stochastic forecasting of COVID-19 daily new cases across countries with a novel hybrid time series model," *Nonlinear Dynamics*, vol. 107, no. 3, pp. 3025–3040, 2022.

[5] A. L. Schaffer, T. A. Dobbins, and S.-A. Pearson, "Interrupted time series analysis using autoregressive integrated moving average (ARIMA) models: A guide for evaluating large-scale health interventions," *BMC Medical Research Methodology*, vol. 21, p. 58, 2021.

[6] W. Zhang, Z. Lin, and X. Liu, "Short-term offshore wind power forecasting – A hybrid model based on Discrete Wavelet Transform (DWT), Seasonal Autoregressive Integrated Moving Average (SARIMA), and deep-learning-based Long Short-Term Memory (LSTM)," *Renewable Energy*, vol. 185, pp. 611–628, 2022.

[7] J. Chen, L. Liu, K. Guo, S. Liu, and D. He, "Short-term electricity load forecasting based on improved data decomposition and hybrid deep-learning models," *Applied Sciences*, vol. 14, no. 14, p. 5966, 2024.

[8] S. Ali, S. Bogarra, M. N. Riaz, P. P. Phyo, D. Flynn, and A. Taha, "From time-series to hybrid models: Advancements in short-term load forecasting embracing smart grid paradigm," *Applied Sciences*, vol. 14, no. 11, p. 4442, 2024.

[9] K. Ullah *et al.*, "Short-term load forecasting: A comprehensive review and simulation study with CNN-LSTM hybrids approach," *IEEE Access*, vol. 12, pp. 111 858–111 881, 2024.

[10] X. Wen, J. Liao, Q. Niu, N. Shen, and Y. Bao, "Deep learning-driven hybrid model for short-term load forecasting and smart grid information management," *Scientific Reports*, vol. 14, no. 1, pp. 1–16, 2024.

[11] Z. Qiao, X. Meng, and L. Wu, "Forecasting carbon dioxide emissions in APEC member countries by a new cumulative grey model," *Ecological Indicators*, vol. 125, p. 107593, 2021.

[12] S. Ding, N. Xu, and J. Ye, "Estimating Chinese energy-related CO2 emissions by employing a novel discrete grey prediction model," *Journal of Cleaner Production*, vol. 259, p. 120793, 2020.

[13] L. Ye, D. Yang, Y. Dang, and J. Wang, "An enhanced multivariable dynamic time-delay discrete grey forecasting model for predicting China's carbon emissions," *Energy*, vol. 249, p. 123653, 2022.

[14] S. Ding and H. Zhang, "Forecasting Chinese provincial CO2 emissions: A universal and robust new-information-based grey model," *Energy Economics*, vol. 121, p. 106685, 2023.

[15] U. Ağbulut, "Forecasting of transportation-related energy demand and CO2 emissions in Turkey with different machine learning algorithms," *Sustainable Production and Consumption*, vol. 29, pp. 141–157, 2022.

[16] P. Delanoë, D. Tchuente, and G. Colin, "Method and evaluations of the effective gain of artificial intelligence models for reducing CO2 emissions," *Journal of Environmental Management*, vol. 331, p. 117261, 2023.

[17] C. Das, R. K. Kunchala, N. Chandra, L. Chmura, J. Nęcki, and P. K. Patra, "Meridional propagation of carbon dioxide (CO2) growth rate and flux anomalies from the tropics due to ENSO," *Geophysical Research Letters*, vol. 49, no. 19, p. e2022GL100105, 2022.

[18] M. P. Dannenberg, M. L. Barnes, W. K. Smith, M. R. Johnston, J. S. Kimball, D. J. P. Moore, and X. Wang, "Large-scale reductions in terrestrial carbon uptake following Central Pacific El Niño," *Geophysical Research Letters*, vol. 48, no. 4, p. e2020GL092367, 2021.

[19] A. Dokumentov and R. J. Hyndman, "STR: Seasonal-trend decomposition using regression," *INFORMS Journal on Data Science*, vol. 1, no. 1, pp. 50–62, 2022.

[20] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 3rd ed. Melbourne, Australia: OTexts, 2021.