Prospects for Applied Mathematics and Data Analysis

ISSN: 2836-4449 Vol: 05, No. 02, 2025 PP. 09–14

Article DOI: https://doi.org/10.54216/PAMDA.050202

FULL LENGTH ARTICLE

Anisotropic Periodic Tikhonov Reconstruction for Structured Gaps in Hourly Urban Demand

Agnes Osagie[1,*] **Mohammad Abobala**[2]

[1] Cape Peninsula University of Technology, Faculty of Applied Science, South Africa

[2] Department of Mathematics, Faculty of Science, Tishreen University, Syria

Emails: Osagieagne2000@cput.ac.za · mohammadabobala777@gmail.com

Received: May 12, 2025 Revised: August 10, 2025 Accepted: October 07, 2025 ⋆ Corresponding author

ABSTRACT

Contiguous observation gaps are difficult to reconstruct when a count process is simultaneously periodic, weather-sensitive, and subject to abrupt operating-state changes. A single isotropic penalty treats these components as if they had the same spectral roughness; weak regularization then leaves high-frequency coefficients unstable, whereas strong regularization suppresses legitimate diurnal peaks. This paper formulates hourly reconstruction as an anisotropic periodic inverse problem. The response is represented on a square-root scale by exogenous covariates, nested daily, weekly, and annual Fourier systems, and modulated daily harmonics. A block-diagonal Tikhonov operator assigns separate regularization levels to these mechanisms and weights the kth periodic mode by k^4, the Fourier form of a squared-curvature penalty. The estimator is available in closed form. Conditions for uniqueness, a perturbation bound for reconstructed gaps, a generalized spectral-filter representation, and monotonicity of the effective dimension are derived. Numerical reconstruction on one year of hourly bicycle demand was evaluated under independent missingness, eight-hour blocks, and complete-day outages using 24 repeated masks per regime. The proposed estimator reduced RMSE relative to a periodic-profile reconstruction by 29.75%, 26.85%, and 29.24%, respectively. Against an isotropic ridge model using the same 139-dimensional basis, the anisotropic model was slightly better for random and full-day gaps, statistically indistinguishable for eight-hour blocks, and attained an effective dimension of 113.57, illustrating that the principal gain is controlled multiscale regularization rather than an increase in model size.

Keywords: Tikhonov regularization ▪ Inverse problems ▪ Fourier basis ▪ Periodic reconstruction ▪ Missing time-series data ▪ Urban mobility ▪ Structured gaps

1. OBSERVATION GAPS AS AN INVERSE PROBLEM

Hourly demand curves are not smooth in a single sense. They contain a strong within-day cycle, a weaker weekly component, slow seasonal movement, and local changes induced by temperature, precipitation, holidays, or service availability. When observations disappear in a contiguous interval, a reconstruction procedure must preserve these scales without allowing the least-observed frequencies to dominate the fit. This is different from ordinary one-step forecasting: values before and after a gap may both be observed, and the task is to recover the missing segment from an incomplete annual field.

Recent bike-sharing work has concentrated mainly on predictive architectures. Weather-based rule systems remain useful as interpretable baselines [1], while data-mining models have been used for trip-duration and demand analysis [2]. System configuration also affects observed demand and operating performance [3]; at finer spatial scales, station-level predictability and allocation around rail stations have been studied explicitly [4, 5]. Edge-computing and multiblock neural formulations were introduced for demand prediction [6, 7]; subsequent studies considered dynamic convolutions [8], deep-learning surveys [9], probabilistic forecasts [10], irregular convolutions [11], and graph adaptation across transport modes [12]. The wider forecasting literature has likewise moved toward global deep-learning and interpretable multi-horizon formulations [13, 14]. These methods are valuable when the objective is future prediction, but they do not remove the need for a small, analyzable reconstruction operator when missingness itself is the problem.

Accordingly, a linear model is adopted on a transformed response, with the main modeling choice placed in the geometry of the penalty. Modern inverse-problem treatments emphasize that regularization is not merely numerical stabilization; the operator used in the penalty determines which solution components are suppressed and which remain identifiable [15, 16]. Here that operator is designed around periodic curvature and covariate interactions.

Let $y_t \geq 0$ denote the hourly count and define $r_t = \sqrt{y_t}$. For an index set Ω of observed hours and its complement Ω^c, the target is the reconstruction map

$$\mathscr{R}_\Omega : \{r_t : t \in \Omega\} \longmapsto \{\widehat{y}_t : t \in \Omega^c\}. \tag{1}$$

The square-root transform keeps the linear inverse problem well behaved while respecting the nonnegative count scale after back-transformation.

2. PERIODIC TIKHONOV FORMULATION

2.1 A multiscale design

Write the standardized design vector at hour t as

$$\boldsymbol{x}_t^\top = \left[1,\, \boldsymbol{e}_t^\top,\, \boldsymbol{d}_t^\top,\, \boldsymbol{f}_t^\top,\, \boldsymbol{m}_t^\top,\, \boldsymbol{w}_t^\top,\, \boldsymbol{a}_t^\top\right]. \tag{2}$$

The vector $\boldsymbol{e}_t$ contains weather variables, quadratic terms for temperature, humidity, and solar radiation, operating-day and holiday indicators, season effects, and day-of-week effects. The unmodulated daily basis is

$$\boldsymbol{d}_t = \left(\sin\frac{2\pi h_t}{24}, \cos\frac{2\pi h_t}{24}, \ldots, \sin\frac{24\pi h_t}{24}, \cos\frac{24\pi h_t}{24}\right)^\top, \tag{3}$$

where $h_t \in \{0,\ldots,23\}$. Three copies of this basis are modulated by the functioning-day indicator, standardized temperature, and standardized humidity. These form $\boldsymbol{f}_t$ and $\boldsymbol{m}_t$. Weekly and annual blocks use four and six harmonic pairs, respectively. The resulting matrix $X \in \mathbb{R}^{8760\times 139}$ is modest enough for direct linear algebra yet rich enough to distinguish spectral scales.

For a generic periodic component

$$q(u) = \sum_{k=1}^{K} \{a_k \sin(2\pi k u/P) + b_k \cos(2\pi k u/P)\}, \tag{4}$$

its integrated squared curvature satisfies

$$\int_0^P \{q''(u)\}^2 du = \frac{P}{2}\left(\frac{2\pi}{P}\right)^4 \sum_{k=1}^{K} k^4 (a_k^2 + b_k^2). \tag{5}$$

Thus k^4 is not an arbitrary frequency weight: up to a fixed scale, it is the Sobolev H^2 roughness of a Fourier series.

2.2 Anisotropic penalty

Let $\boldsymbol{\theta}$ collect all regression and harmonic coefficients. Define

$$\begin{aligned} R(\boldsymbol{\lambda}) = \operatorname{diag}\{0, \lambda_e I, \lambda_d D_d, \lambda_f D_d,\\ \lambda_m D_m, \lambda_w D_w, \lambda_a D_a\}. \end{aligned} \tag{6}$$

where the first zero leaves the intercept unpenalized and each periodic matrix contains duplicated fourth-power frequencies,

$$D_K = \operatorname{diag}(1^4, 1^4, 2^4, 2^4, \ldots, K^4, K^4). \tag{7}$$

The weather-modulated daily block contains two such copies because temperature and humidity interactions are both present. Separate multipliers $\boldsymbol{\lambda} = (\lambda_e, \lambda_d, \lambda_f, \lambda_m, \lambda_w, \lambda_a)$ permit different smoothness across mechanisms.

For $n_\Omega = |\Omega|$, the estimator solves

$$\widehat{\boldsymbol{\theta}} = \operatorname*{arg\,min}_{\boldsymbol{\theta}\in\mathbb{R}^p} \left\{ \frac{1}{2n_\Omega}\|\boldsymbol{r}_\Omega - X_\Omega \boldsymbol{\theta}\|_2^2 + \frac{1}{2}\boldsymbol{\theta}^\top R(\boldsymbol{\lambda})\boldsymbol{\theta} \right\}. \tag{8}$$

The normal equations give

$$A_\Omega \widehat{\boldsymbol{\theta}} = X_\Omega^\top \boldsymbol{r}_\Omega, \qquad A_\Omega = X_\Omega^\top X_\Omega + n_\Omega R(\boldsymbol{\lambda}), \tag{9}$$

and therefore

$$\widehat{\boldsymbol{\theta}} = A_\Omega^{-1} X_\Omega^\top \boldsymbol{r}_\Omega. \tag{10}$$

Missing counts are returned to the original scale by

$$\widehat{y}_t = \left[\max\{\boldsymbol{x}_t^\top \widehat{\boldsymbol{\theta}}, 0\}\right]^2, \qquad t \in \Omega^c. \tag{11}$$

This estimator will be called anisotropic periodic Tikhonov reconstruction (APTR).

3. ALGEBRAIC PROPERTIES

3.1 Well-posedness and perturbation stability

The direct formula is useful only if the regularized system is nonsingular.

Proposition 1 (Uniqueness)**.** *The minimizer of* (8) *is unique if and only if*

$$\ker(X_\Omega) \cap \ker(R) = \{\boldsymbol{0}\}. \tag{12}$$

Proof. For any $\boldsymbol{v} \in \mathbb{R}^p$,

$$\boldsymbol{v}^\top A_\Omega \boldsymbol{v} = \|X_\Omega \boldsymbol{v}\|_2^2 + n_\Omega \boldsymbol{v}^\top R \boldsymbol{v} \geq 0. \tag{13}$$

Equality holds precisely when both terms vanish. Hence A_Ω is positive definite exactly under (12). Strict convexity then gives a unique minimizer, and (10) follows. □

Because all non-intercept directions receive positive regularization and the intercept is observed whenever $\Omega \neq \varnothing$, condition (12) holds for the reconstruction experiments considered here.

Let $X_M = X_{\Omega^c}$ and perturb the observed square-root response by $\boldsymbol{\delta}$. The corresponding change in the missing-point linear predictor is

$$\Delta\widehat{\boldsymbol{r}}_M = X_M A_\Omega^{-1} X_\Omega^\top \boldsymbol{\delta}. \tag{14}$$

Consequently,

$$\|\Delta\widehat{\boldsymbol{r}}_M\|_2 \le \|X_M A_\Omega^{-1} X_\Omega^\top\|_2 \|\boldsymbol{\delta}\|_2, \tag{15}$$

with the simpler bound

$$\|\Delta\widehat{\boldsymbol{r}}_M\|_2 \le \frac{\|X_M\|_2\|X_\Omega\|_2}{\lambda_{\min}(A_\Omega)}\|\boldsymbol{\delta}\|_2. \tag{16}$$

The denominator makes explicit why regularization matters when a gap removes information from a subset of harmonic directions.

3.2 Generalized spectral filtering

Set $G_\Omega = X_\Omega^\top X_\Omega$. Suppose first that R is positive definite on the non-intercept subspace. Let $R^{1/2}$ be its symmetric square root and diagonalize

$$R^{-1/2} G_\Omega R^{-1/2} = U\,\mathrm{diag}(\rho_1,\dots,\rho_{p-1})U^\top. \tag{17}$$

In the generalized coordinates $\boldsymbol{z} = U^\top R^{1/2}\boldsymbol{\theta}$, equation (9) acts through the factors

$$\phi_j = \frac{\rho_j}{\rho_j + n_\Omega}, \qquad 0 \le \phi_j < 1. \tag{18}$$

Directions with little data information relative to roughness ($\rho_j \ll n_\Omega$) are heavily attenuated. Because R contains k^4 weights, high-order periodic coefficients require substantially more evidence from G_Ω to survive. Equation (18) is therefore a spectral explanation of the reconstruction rule, not only a computational identity.

3.3 Gap-wise bias–variance decomposition

A second view of the regularizer follows from a local linear data model on the transformed scale,

$$\boldsymbol{r}_\Omega = X_\Omega\boldsymbol{\theta}_\star + \boldsymbol{\varepsilon}_\Omega, \qquad \mathbb{E}\boldsymbol{\varepsilon}_\Omega = \boldsymbol{0}, \qquad \mathrm{Cov}(\boldsymbol{\varepsilon}_\Omega) = \sigma^2 I. \tag{19}$$

Let $M = \Omega^c$ and temporarily omit the nonnegative back-transformation. The reconstruction is

$$\widehat{\boldsymbol{r}}_M = X_M A_\Omega^{-1} X_\Omega^\top \boldsymbol{r}_\Omega. \tag{20}$$

Since $A_\Omega = G_\Omega + n_\Omega R$, the expected coefficient error is

$$\mathbb{E}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_\star) = -n_\Omega A_\Omega^{-1} R\boldsymbol{\theta}_\star. \tag{21}$$

Regularization therefore trades variance for a directional bias determined by R, rather than shrinking every basis vector equally.

Proposition 2 (Gap-wise mean-square reconstruction error)**.** *Under (19), the expected squared error for the noiseless signal over a missing set M is*

$$\mathbb{E}\|\widehat{\boldsymbol{r}}_M - X_M\boldsymbol{\theta}_\star\|_2^2 = n_\Omega^2\|X_M A_\Omega^{-1} R\boldsymbol{\theta}_\star\|_2^2 + \sigma^2\,\mathrm{tr}\Big(X_M A_\Omega^{-1} G_\Omega A_\Omega^{-1} X_M^\top\Big). \tag{22}$$

Proof. Substitution of (19) into (10) gives the deterministic bias in (21) and random term $A_\Omega^{-1}X_\Omega^\top\boldsymbol{\varepsilon}_\Omega$. After premultiplication by X_M, the cross term has zero expectation. The trace term is the total variance because $\mathrm{Cov}(\boldsymbol{\varepsilon}_\Omega) = \sigma^2 I$. □

Equation (22) makes the geometry of the missing set explicit through X_M. Two masks with the same number of missing observations need not be equally difficult: a whole-day outage and dispersed single-hour losses generate different X_M and different information loss in G_Ω. This is the principal reason for evaluating several gap shapes rather than reporting a single random-missingness score.

3.4 Effective dimension

The observed-point smoothing matrix on the transformed scale is

$$H_\Omega = X_\Omega A_\Omega^{-1} X_\Omega^\top, \tag{23}$$

and its effective dimension is

$$\mathrm{df}_{\mathrm{eff}} = \mathrm{tr}(H_\Omega) = \mathrm{tr}\{G_\Omega A_\Omega^{-1}\}. \tag{24}$$

This quantity measures the number of freely transmitted linear degrees of freedom after regularization.

Proposition 3 (Monotone complexity under global scaling)**.** *Let $R_\eta = \eta R_0$ with $\eta > 0$ and $R_0 \succeq 0$. Then*

$$\frac{d}{d\eta}\mathrm{df}_{\mathrm{eff}}(\eta) = -n_\Omega\,\mathrm{tr}\{G_\Omega A_\eta^{-1} R_0 A_\eta^{-1}\} \le 0. \tag{25}$$

Proof. Differentiate A_η^{-1} using $dA^{-1}/d\eta = -A^{-1}(dA/d\eta)A^{-1}$ and $dA_\eta/d\eta = n_\Omega R_0$. The trace term in (25) is nonnegative because it can be written as the trace of a product of positive-semidefinite congruences. □

The proposition is useful in practice: increasing a common regularization scale cannot secretly increase model complexity, even though the design contains several interacting harmonic blocks.

4. CONTROLLED RECONSTRUCTION EXPERIMENT

4.1 Observations and basis construction

The numerical study uses the Seoul Bike Sharing Demand record distributed by the UCI Machine Learning Repository [17]. It contains 8,760 hourly observations spanning 365 consecutive days and includes hourly rental counts, meteorological variables, season, holiday status, and functioning-day status. The supplied record has no native missing entries, which allows missingness geometry to be controlled rather than confounded with an unknown acquisition mechanism. Demand ranges from 0 to 3,556 rentals per hour, with a mean of 704.60 and median of 504.5.

Figure 1 shows the annual demand surface. The strong hour-of-day organization is visible, but its amplitude changes substantially over the year. That pattern motivates separate daily,

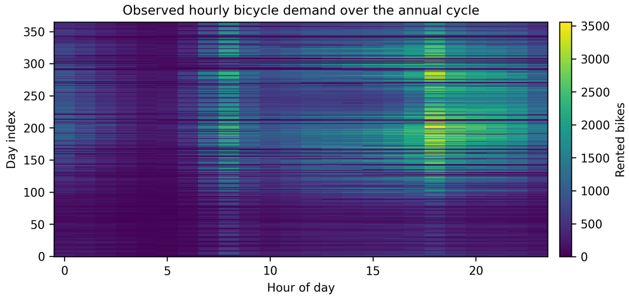

Figure 1. Observed hourly rental demand. Rows are consecutive days and columns are hours within a day.

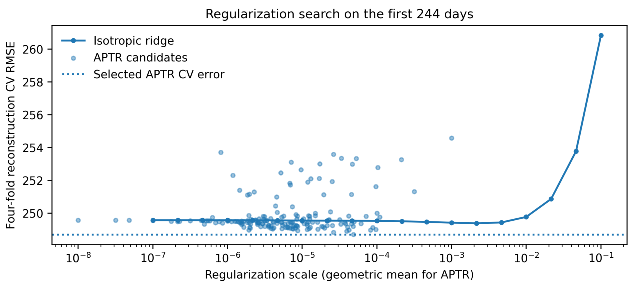

Figure 2. Regularization search restricted to the first 244 days. For APTR, the horizontal coordinate is the geometric mean of the six block coefficients and is used only for visualization.

annual, and weather-modulated terms rather than a single periodic component.

The exogenous block contains eight standardized continuous variables (temperature, humidity, wind speed, visibility, dew point, solar radiation, $\log(1+\text{rain})$, and $\log(1+\text{snow})$), three quadratic terms, functioning-day and holiday indicators, three season contrasts, and six day-of-week contrasts. Twelve daily harmonic pairs are used, together with functioning-day, temperature, and humidity modulations. Four weekly and six annual pairs complete the 139-column design.

4.2 Parameter selection without evaluation leakage

The first 244 days are used only for regularization selection. Four day-stratified reconstruction folds are formed by holding out every fourth day within that period. The isotropic comparator uses the same design but a single ridge coefficient, selected from 19 logarithmically spaced values between 10^{-7} and 10^{-1}. APTR searches 171 deterministic candidates: 11 common-scale frequency-weighted penalties plus 160 seeded anisotropic combinations. No response from the final 121 days participates in this selection.

The chosen isotropic coefficient is 2.154×10^{-3}. APTR selects

$$\begin{aligned} \lambda_e &= 4.627 \times 10^{-3}, \quad \lambda_d = 2.149 \times 10^{-7}, \\ \lambda_f &= 9.259 \times 10^{-6}, \quad \lambda_m = 8.412 \times 10^{-5}, \\ \lambda_w &= 6.898 \times 10^{-5}, \quad \lambda_a = 2.453 \times 10^{-4}. \end{aligned} \tag{26}$$

The tuning RMSE is 248.71 for APTR and 249.39 for the isotropic alternative. Figure 2 displays the search surface in a scalar projection; the ordinate, not the projected abscissa, determines selection.

4.3 Missing-data geometries

Only the final 121 days are used for reported reconstruction accuracy. Three gap mechanisms are generated independently for each of 24 replicates:

1. *Random-20*: each evaluation hour is removed with probability 0.20;
2. *Block-8h*: 45% of evaluation days receive one contiguous eight-hour gap with a uniformly selected start hour from 0 to 16;
3. *Full-day*: 18% of evaluation days are removed in complete 24-hour blocks.

All nonmasked responses remain available because the task is gap reconstruction, not forward forecasting. For comparison, a periodic-profile estimator averages observed demand within hour $\times$ functioning-state $\times$ season cells, with hourly and global fallbacks. The second comparator is isotropic ridge on the identical 139-column design. Hence the contrast between isotropic ridge and APTR isolates the penalty geometry rather than the feature set.

Accuracy is summarized by

$$\mathrm{RMSE} = \left\{ \frac{1}{m} \sum_{t \in M} (y_t - \widehat{y}_t)^2 \right\}^{1/2}, \tag{27}$$

$$\mathrm{MAE} = \frac{1}{m} \sum_{t \in M} |y_t - \widehat{y}_t|. \tag{28}$$

Paired 95% intervals for RMSE differences are obtained from 4,000 bootstrap resamples of the 24 replicate-wise differences.

5. NUMERICAL EVIDENCE

Table 1 separates two effects. First, the multiscale covariate-harmonic representation is materially better than a periodic cell mean: RMSE falls by approximately 27–30% in all three geometries. Second, anisotropic frequency regularization yields a much smaller incremental change once the same rich basis is already present. The APTR-minus-ridge RMSE difference is -0.37 for Random-20 (95% bootstrap interval $[-0.53, -0.21]$), $+0.15$ for Block-8h ($[-0.10, 0.40]$), and -0.50 for Full-day ($[-0.90, -0.11]$). Thus the blockwise penalty is beneficial for dispersed and whole-day omissions in this experiment, while the eight-hour case is practically tied.

The lack of uniform dominance is consistent with the information retained by the different masks. An eight-hour hole often leaves the same day's morning or evening segment observed, so isotropic shrinkage can exploit the dense common design nearly as well as a spectral penalty. A complete-day outage removes every within-day observation simultaneously; in that regime the explicit periodic hierarchy has a slightly clearer role.

Figure 3 shows that the profile baseline also has wider between-mask dispersion, especially for full-day outages. A single reconstructed day from the first full-day replicate is shown in Figure 4. The profile reproduces a generic daily contour but misses the day's amplitude; both regression-based reconstructions adapt to the meteorological and operating

Table 1. Reconstruction performance over 24 repeated masks. Values are mean $\pm$ standard deviation in rented bikes per hour.

Gap geometry	Method	Masked hours	RMSE	MAE	RMSE change vs. profile
Random-20	Periodic profile	593.5	398.29 ± 17.27	263.80 ± 10.36	–
	Isotropic ridge	593.5	280.17 ± 10.62	190.34 ± 6.93	-29.66%
	APTR	593.5	$\mathbf{279.81 \pm 10.59}$	$\mathbf{189.74 \pm 6.93}$	$\mathbf{-29.75\%}$
Block-8h	Periodic profile	432.0	415.69 ± 39.35	281.82 ± 26.85	–
	Isotropic ridge	432.0	$\mathbf{303.91 \pm 16.55}$	208.40 ± 12.36	-26.89%
	APTR	432.0	304.06 ± 16.58	$\mathbf{208.22 \pm 12.38}$	$\mathbf{-26.85\%}$
Full-day	Periodic profile	528.0	391.58 ± 49.82	261.64 ± 37.96	–
	Isotropic ridge	528.0	277.60 ± 15.11	188.12 ± 14.77	-29.11%
	APTR	528.0	$\mathbf{277.10 \pm 14.86}$	$\mathbf{187.40 \pm 14.50}$	$\mathbf{-29.24\%}$

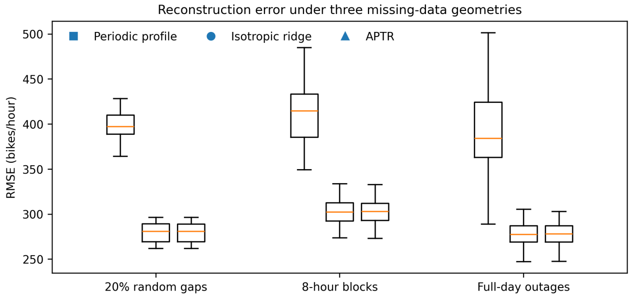

Figure 3. Distribution of replicate RMSE under each missing-data geometry.

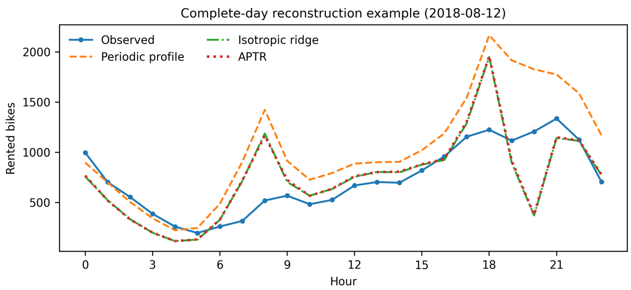

Figure 4. Example complete-day outage. All 24 target counts are withheld during fitting.

covariates.

6. WHAT THE REGULARIZATION CHANGES

The full-data APTR fit has $p = 139$ coefficients but an effective dimension of 113.57 from (24). The regularized normal matrix has condition number 1.009×10^3, and the computed quantity $\|X\|_2/\lambda_{\min}(A)$ is 5.08. These diagnostics are not predictive scores; they describe the linear map through which perturbations in the transformed observations reach the estimated coefficients.

The daily harmonic energy in Figure 5 is concentrated in the lower modes, while higher modes remain available to represent sharper peaks. Because the penalty grows as k^4, high-frequency coefficients are admitted only when the reduction in residual error compensates for a rapidly increasing curvature cost. This mechanism is qualitatively different from deleting high harmonics in advance: the basis remains complete up to $k = 12$, but its effective use is data-dependent.

The experiment has four limitations. First, gaps are imposed on a complete historical record, so the reported errors concern controlled reconstruction rather than a naturally missing process.

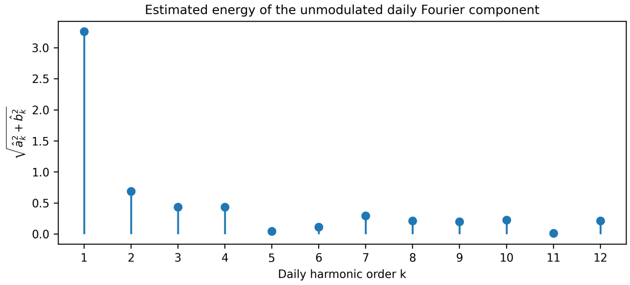

Figure 5. Norm of the fitted sine–cosine coefficient pair for each unmodulated daily harmonic.

Second, weather and calendar covariates are assumed available at missing-demand hours. Third, the square-root Gaussian least-squares criterion does not model count dispersion explicitly. Fourth, the gain over isotropic ridge is small on this dataset; the evidence supports the penalty as a transparent multiscale regularizer rather than as a universally superior predictor. More complex bike-demand models can achieve lower errors when extensive spatial history is available [18, 19, 20, 21], but their objective and information set differ from the inverse reconstruction problem considered here.

7. CONCLUSION

Structured gaps in periodic count records can be treated as a finite-dimensional inverse problem without abandoning the multiple time scales that create the observed signal. The APTR formulation combines a rich harmonic-covariate basis with an anisotropic Tikhonov operator whose fourth-power frequency weights arise directly from squared periodic curvature. The resulting estimator is closed form, uniquely defined under a simple null-space condition, spectrally interpretable, and equipped with explicit stability and complexity expressions.

The numerical experiment shows where that structure matters and where it does not. Most of the improvement over a periodic profile comes from conditioning on weather, calendar, operating state, and multiscale harmonics. Relative to isotropic ridge on the same basis, anisotropy gives small but reproducible gains for random gaps and complete-day outages and no meaningful advantage for eight-hour blocks. The comparison therefore separates the contribution of the regularizer from that of the representation. For reconstruction systems in which auditability, direct solvability, and control of high-frequency behavior are required, the formulation provides a compact alternative to opaque high-capacity models.

REFERENCES

[1] V. E. Sathishkumar and Y. Cho, "A rule-based model for seoul bike sharing demand prediction using weather data," *European Journal of Remote Sensing*, vol. 53, no. sup1, pp. 166–183, 2020.

[2] V. E. Sathishkumar, J. Park, and Y. Cho, "Seoul bike trip duration prediction using data mining techniques," *IET Intelligent Transport Systems*, vol. 14, no. 11, pp. 1465–1474, 2020.

[3] Z. Kou and H. Cai, "Comparing the performance of different types of bike share systems," *Transportation Research Part D: Transport and Environment*, vol. 94, p. 102823, 2021.

[4] Z. Yin, K. Hardaway, Y. Feng, Z. Kou, and H. Cai, "Understanding the demand predictability of bike share systems: A station-level analysis," *Frontiers of Engineering Management*, vol. 10, no. 4, pp. 551–565, 2023.

[5] L. Yu, T. Feng, T. Li, and L. Cheng, "Demand prediction and optimal allocation of shared bikes around urban rail transit stations," *Urban Rail Transit*, vol. 9, pp. 57–71, 2023.

[6] T. Xu, G. Han, X. Qi, J. Du, C. Lin, and L. Shu, "A hybrid machine learning model for demand prediction of edge-computing-based bike-sharing system using internet of things," *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 7345–7356, 2020.

[7] M. Xu, H. Liu, and H. Yang, "A deep learning based multi-block hybrid model for bike-sharing supply-demand prediction," *IEEE Access*, vol. 8, pp. 85 826–85 838, 2020.

[8] S. Qiao, N. Han, J. Huang, K. Yue, R. Mao, H. Shu, Q. He, and X. Wu, "A dynamic convolutional neural network based shared-bike demand forecasting model," *ACM Transactions on Intelligent Systems and Technology*, vol. 12, no. 6, pp. 70:1–70:24, 2021.

[9] W. Jiang, "Bike sharing usage prediction with deep learning: A survey," *Neural Computing and Applications*, vol. 34, pp. 15 369–15 385, 2022.

[10] H. Lim, K. Chung, and S. Lee, "Probabilistic forecasting for demand of a bike-sharing service using a deep-learning approach," *Sustainability*, vol. 14, no. 23, p. 15889, 2022.

[11] X. Li, Y. Xu, X. Zhang, W. Shi, Y. Yue, and Q. Li, "Improving short-term bike sharing demand forecast through an irregular convolutional neural network," *Transportation Research Part C: Emerging Technologies*, vol. 147, p. 103984, 2023.

[12] Y. Liang, G. Huang, and Z. Zhao, "Cross-mode knowledge adaptation for bike sharing demand prediction using domain-adversarial graph neural networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 5, pp. 3642–3653, 2024.

[13] B. Lim and S. Zohren, "Time-series forecasting with deep learning: A survey," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 379, no. 2194, p. 20200209, 2021.

[14] B. Lim, S. Ö. Arık, N. Loeff, and T. Pfister, "Temporal fusion transformers for interpretable multi-horizon time series forecasting," *International Journal of Forecasting*, vol. 37, no. 4, pp. 1748–1764, 2021.

[15] J. Chung and S. Gazzola, "Computational methods for large-scale inverse problems: A survey on hybrid projection methods," *SIAM Review*, vol. 66, no. 2, pp. 205–284, 2024.

[16] K. Bredies, M. Carioni, S. Fanzon, and F. Romero, "A generalized conditional gradient method for dynamic inverse problems with optimal transport regularization," *Foundations of Computational Mathematics*, vol. 23, no. 3, pp. 833–898, 2023.

[17] UCI Machine Learning Repository, "Seoul bike sharing demand," UCI Machine Learning Repository, 2020.

[18] D. Gammelli, Y. Wang, D. Prak, F. Rodrigues, S. Minner, and F. C. Pereira, "Predictive and prescriptive performance of bike-sharing demand forecasts for inventory management," *Transportation Research Part C: Emerging Technologies*, vol. 138, p. 103571, 2022.

[19] R. Harikrishnakumar and S. Nannapaneni, "Forecasting bike sharing demand using quantum bayesian network," *Expert Systems with Applications*, vol. 221, p. 119749, 2023.

[20] C. Peláez-Rodríguez, J. Pérez-Aracil, D. Fister, R. Torres-López, and S. Salcedo-Sanz, "Bike sharing and cable car demand forecasting using machine learning and deep learning multivariate time series approaches," *Expert Systems with Applications*, vol. 238, p. 122264, 2024.

[21] C. Zhou, J. Hu, X. Zhang, Z. Li, and K. Yang, "A bike-sharing demand prediction model based on spatio-temporal graph convolutional networks," *PeerJ Computer Science*, vol. 10, p. e2391, 2024.