A FILTER BANK APPROACH FOR MODELING AND FORECASTING SEASONAL PATTERNS

Ta-Hsin Li[1]

Department of Statistics and Applied Probability
University of California
Santa Barbara, CA 93106-3110

Melvin J. Hinich

Applied Research Laboratories
University of Texas
Austin, TX 78713-8029

April 27, 2000
Version 3

Technical Report No. 355
Department of Statistics and Applied Probability
University of California, Santa Barbara

[1]Current Address: Department of Mathematical Sciences, IBM T. J. Watson Research Center, Yorktown Heights, NY 10598-0218. E-mail: thl@watson.ibm.com

Abstract

A novel method for modeling and forecasting seasonal time series is proposed. Unlike the traditional approach that depends solely on dynamic models, the proposed method combines stochastic dynamic modeling with an analysis filter bank designed to reduce dimensionality and to extract persistent components for reliable long-term forecasting. The filter bank decomposes the time series of interest into seasonal components, and only those components that are highly coherent across the periods are selected for subsequent modeling and forecasting. Experiments show that under suitable conditions, the use of highly coherent (High-C) components not only reduces the modeling complexity and the required amount of training data but also limits the impact of noise and occasional corruption in the training data and thus provides robust forecasts with reduced variability. Fourier and wavelet filter banks are discussed in detail. Simulated and real-data examples are used to illustrate the method.

Acknowledgment. This work was supported in part by NSF grant DMS-9817552. The global irradiance data were obtained from `http://www.geo.umass.edu/climate/TILPHTML/TILPhome.html`. The authors thank the editors and the anonymous referees for their helpful comments and suggestions that have led to an improved presentation.

Key Words. Autoregressive, coherence, cyclostationary, detection, Fourier transform, multiresolution, nonstationary, periodicity, signal processing, spectral analysis, time series, wavelet.

Running Title. Modeling and Forecasting Seasonal Patterns

1 Introduction

Periodicities are inherent in many physical, business, and economical time series (e.g., Dehay and Hurd, 1994; Franses, 1996). Figure 1 shows two examples: the hourly global irradiance and the monthly number of airline passengers (see Section 7 for more information). These so-called periodic or seasonal time series are not deterministic since they do not perfectly repeat themselves from one period to another. In fact, there are considerable random variations in the waveforms over the periods. This article presents a new approach for modeling and forecasting such *changing* seasonal time series. Here, the term "season" is generic; it represents any integral periodicity (e.g., 24 hours, 12 months, 52 weeks, etc.). Since our focus is on the seasonal patterns, it is assumed that non-seasonal trends (if any), such as linear and exponential growth patterns, have been removed by some detrending techniques (e.g., differencing, regression, etc.).

Long-term forecasting of seasonal time series is critical in many applications such as strategic business planning and operation. For example, at the end of each week, an airline company may be interested in forecasting hourly passenger flows at a particular gate of an airport for the next few weeks in order to meet the demand on its services (e.g., security personnel). A manufacturing company may need at the end of each year to forecast monthly demand of its product for the next few years in order to secure sufficient parts from suppliers. Reliable forecasting up to such long-term horizons is the main concern in this article.

There are many possible ways to model changing seasonality, depending on how one assumes that the seasonal patterns would evolve over time. Most widely-used models for changing seasonality can be roughly classified into two broad categories: models with seasonal unit roots and models with season-dependent parameters. In the first category, the seasonal ARIMA (or SARIMA) model, developed by Box and Jenkins (1976), is perhaps the most popular and has been used as benchmarks in many comparative studies of forecasting (e.g., Funke, 1990; Kulendran and King, 1997; Novales and de Fruto, 1997). In this model, the seasonal pattern in one period is assumed to repeat itself in the next, only with an additive random perturbation. In other words, if $\mathbf{x}(m)$ represents the seasonal vector in year m, then the SARIMA model assumes that $\mathbf{x}(m) = \mathbf{x}(m-1) + \boldsymbol{\epsilon}(m)$, where $\boldsymbol{\epsilon}(m)$ is a zero-mean stationary random process (e.g., ARMA). This random-walk-type seasonality is also assumed in the state-space models (Harvey, 1993; Kitagawa and Gersch, 1996).

Although effective in many cases, a drawback of SARIMA is the lack of power for reliable *long-term* forecasting because of its random-walk-type model assumption. For a typical SARIMA

process, the long-term forecast function depends crucially on the most recent individual seasonal pattern in the training data, and the forecast consists of a periodic extension of this pattern and an additive modifying component which tends to zero exponentially quickly as the horizon becomes large. However, in the presence of random fluctuation, the individual seasonal patterns are not only unlikely to duplicate themselves in the future but also extremely vulnerable to corruption by occasional abnormal behavior in the training data (e.g., a large outlier). As a consequence, the SARIMA forecasts can be very unreliable in these situations, especially when the forecast horizon is not short. For reliable long-term forecasting, it is crucial to find more *persistent patterns*.

A typical example in the second category is the periodic ARMA model in which the ARMA parameters are allowed to vary with the season rather than being kept constant (e.g., Pagano, 1978; Vecchia, 1985; Bloomfield, Hurd, and Lund, 1994). This model can be reformulated as multivariate (or vector) ARMA for the seasonal vectors. The forecast of multivariate ARMA consists of the seasonal mean plus a modifying component determined by the autocorrelation properties of the time series across the periods.

The full multivariate approach is very flexible and versatile, but it has a drawback in situations where a large number of variables are contained in the seasonal vectors. For example, a 52-dimensional model would be required for weekly observations when the annual period is of interest. In situations such as this, a full multivariate ARMA model would involve vectors and matrices of huge dimension, and thus the computational and modeling complexity can become too burdensome to handle. Reliable estimation of a large number of parameters would also require a huge amount of training data which may not always be available. To efficiently utilize the resources in modeling such highly complex time series, it is desirable to develop certain *dimension reduction* techniques. One approach of dimension reduction is to impose special structures on the model parameters by using, for example, periodicity (e.g., Bloomfield, Hurd, and Lund, 1994), subset modeling (e.g., Lütkepohl, 1993), and Bayesian priors (e.g., Litterman, 1986; Lütkepohl, 1993).

In this article, we take a different approach to dimension reduction. The basic idea is to decompose the seasonal patterns into different components and distinguish the seasonal components that are important to the modeling and forecasting objectives from those that are less important and thus unworthy of the computational and data resources. By combining the pattern extraction and dimension reduction technique with parametric dynamic models, we propose a general methodology for the modeling and forecasting of seasonal time series.

The novel technique that we propose to extract persistent patterns and to reduce the dimension

in the subsequent modeling is within the framework of statistical time-frequency analysis (STFA). The STFA (also known as parametric filtering) is a general approach for time series analysis based on the combination of analysis filter banks with simple output statistics (Li, 1996; 1997; Li and Gibson, 1996). In this approach, the filter bank enhances the desirable features of the input time series and, at the same time, suppresses unwanted interference and noise. The simple statistics at the output describe important statistical properties of the filtered time series for robust detection, estimation, and modeling.

Filtering with a linear filter bank can be interpreted as projecting the input time series onto a set of basis functions represented by the filters' impulse response. If the basis functions are judiciously designed (e.g., sines and cosines), then the projections uniquely determine the input time series. When the filters are applied to each seasonal vector of a seasonal time series, the resulting projections can be regarded as waveforms that propagate from one period to another. Among these waveforms, there are some that are *highly coherent* (High-C) over the periods (Hinich, 1996). In this article, we propose to explore these High-C waveforms as a basis for parametric modeling and forecasting. Our preliminary study shows that under suitable conditions the proposed STFA method not only reduces the modeling complexity but also provides more reliable forecasts.

The parametric modeling in the proposed approach is based on the second-order summary statistics of the High-C projection coefficients. The objective is to capture the essential statistical characteristics of the projections as the period progresses. Standard vector time series models, such as vector ARMA (e.g., Lütkepohl, 1993), can be employed for the parametric modeling. Of course, the complexity is considerably reduced as compared to the full multivariate ARMA method, because the modeling is now performed only on a *subset* of the projections.

2 Decomposition by Filter Banks

The first component of the proposed approach is the decomposition of the seasonal time series by an STFA filter bank. For simplicity, we consider orthogonal filter banks only, although the orthogonality is not required as long as the transformation is invertible.

Let $\{x(t), t = 0, \pm 1, \pm 2, \cdots\}$ be the (univariate) time series of interest. Suppose that $\{x(t)\}$ has been observed over n periods of length $p > 1$ so that the available data can be expressed as

$$\mathbf{x}(m) := [x(mp-1), \cdots, x(mp-p)]^T \qquad (m = 1, \cdots, n), \tag{1}$$

where $\mathbf{x}(m)$ is the seasonal vector of the mth period. (The first element of $\mathbf{x}(m)$ is the most recent

observation in period m.) Let $\mathbf{w}_k := [w_k(1), \cdots, w_k(p)]^T$ $(k = 1, \cdots, p)$, be a set of orthonormal finite-impulse-response (FIR) filters, satisfying

$$\mathbf{w}_j^T \mathbf{w}_k = \begin{cases} 1 & \text{if } j = k, \\ 0 & \text{if } j \neq k. \end{cases}$$

With $\{x(t)\}$ as input, the output from the kth filter $\mathbf{w}_k$ can be expressed as

$$\xi_k(t) := \sum_{\tau=1}^{p} w_k(\tau)\, x(t-\tau), \qquad (k = 1, \cdots, p). \tag{2}$$

Subsampling the output at $t = mp$ (i.e., one sample per period) yields

$$\xi_k(mp) = \sum_{\tau=1}^{p} w_k(\tau)\, x(mp-\tau) = \mathbf{w}_k^T \mathbf{x}(m). \tag{3}$$

If the samples are collected across the filter bank to form a vector

$$\boldsymbol{\xi}(m) := [\xi_1(mp), \cdots, \xi_p(mp)]^T,$$

then it follows from (3) that

$$\boldsymbol{\xi}(m) = \mathbf{W}\, \mathbf{x}(m), \tag{4}$$

where $\mathbf{W}$ is an orthogonal matrix defined by

$$\mathbf{W} := [\mathbf{w}_1, \cdots, \mathbf{w}_p]^T. \tag{5}$$

Thus the filter-bank output $\boldsymbol{\xi}(m)$ is an orthogonal transform of the seasonal data vector $\mathbf{x}(m)$. Note that the $\mathbf{w}_k$, also known as the *analysis* filters, are the row vectors of $\mathbf{W}$.

Since $\mathbf{W}^T\mathbf{W} = \mathbf{I}$, one can reconstruct $\mathbf{x}(m)$ from $\boldsymbol{\xi}(m)$ according to $\mathbf{x}(m) = \mathbf{W}^T\boldsymbol{\xi}(m)$. The reconstruction can be accomplished by using a *synthesis* filter bank that comprises the *column* vectors of $\mathbf{W}$, i.e., $\mathbf{w}(\tau) := [w_1(\tau), \cdots, w_p(\tau)]^T$ $(\tau = 1, \cdots, p)$, which are also orthonormal. The reconstruction can be expressed as

$$\mathbf{x}(m) = \sum_{k=1}^{p} \mathbf{x}_k(m) = \sum_{k=1}^{p} \xi_k(mp)\, \mathbf{w}_k, \tag{6}$$

where $\mathbf{x}_k(m) := \xi_k(mp)\, \mathbf{w}_k$ is the projection of $\mathbf{x}(m)$ on $\mathbf{w}_k$, with $\xi_k(mp)$ being the projection coefficient. Equation (6) can be interpreted as a decomposition of $\mathbf{x}(m)$ into p orthogonal projections $\mathbf{x}_k(m)$, $(k = 1, \cdots, p)$. In this decomposition, the intra-period characteristics of $\{x(t)\}$ are encoded by the filters into the projection coefficients, and the inter-period variations of $\{x(t)\}$ are transformed into the dynamics of the projection coefficients.

3 Modeling With Coherent Waveforms

For each fixed k, the filter $\mathbf{w}_k$ can be interpreted as a *waveform* of unity power, with the index k representing the wavenumber. According to (6), the seasonal vector $\mathbf{x}(m)$ is a superposition of p such waveforms whose coefficients, $\xi_k(mp)$, vary randomly from one period to another, thus resulting in changing seasonality.

3.1 Definition of Coherent Waveforms

The component waveforms of $\mathbf{x}(m)$ can be classified into three categories according to the statistical characteristics of their coefficients $\xi_k(mp)$. In the following, let us assume that for each k, $\{\xi_k(mp)\}$ is a second-order stationary process with mean μ_k and variance σ_k^2 such that $n^{-1}\sum_{m=1}^{n}\xi_k(mp) \to \mu_k$ and $n^{-1}\sum_{m=1}^{n}\{\xi_k(mp)-\mu_k\}^2 \to \sigma_k^2$ as $n \to \infty$ in probability or almost surely.

Definition 1 *Assuming $\mu_k^2+\sigma_k^2 > 0$, the coherence of the kth waveform of $\mathbf{x}(m)$ is defined as*[1]

$$\gamma_k := \frac{\mu_k^2}{\mu_k^2+\sigma_k^2}. \tag{7}$$

The kth waveform of $\mathbf{x}(m)$ is said to be (a) *completely coherent if $\gamma_k = 1$,* (b) *incoherent if $\gamma_k = 0$,* (c) *partially coherent if $0 < \gamma_k < 1$. Let $\hat{\gamma}_k$ be an estimate of γ_k from a finite sample. Then, the kth waveform of $\mathbf{x}(m)$ is said to be highly coherent, or High-C, if $\hat{\gamma}_k > \theta$, where θ is a threshold whose choice is discussed in Section 3.3.*

Because $\mu_k^2+\sigma_k^2$ is the second moment of $\xi_k(mp)$, the coherence γ_k, which is always between zero and unity, can be interpreted as the fraction of the inter-period variability of $\xi_k(mp)$ around zero that can be attributed to the mean of $\xi_k(mp)$. By definition, incoherent waveforms do not have any long-term effect because their coefficients have zero mean ($\mu_k = 0$) and thus lack persistence in the long run. On the other hand, completely coherent waveforms become deterministic in the long run ($\sigma_k^2 = 0$) and are perfect for long-term forecasting.

In general, the High-C waveforms are most suitable for long-term forecasting because they possess long-lasting effects (i.e., non-zero mean) as the period progresses. This observation gives rise to a novel idea of pattern extraction and dimension reduction: Utilize only the High-C waveforms in the subsequent modeling.

[1] This definition of coherence should not be confused with the cross-correlation or the spectral coherence for multiple time series.

3.2 A Seasonal Time Series Model

Let $\mathcal{C} \subset \{1, \cdots, p\}$ be the index set of High-C waveforms. In the proposed approach, it is assumed that the time series $\{x(t)\}$ satisfies the following condition:

$$\mathbf{x}(m) = \sum_{k \in \mathcal{C}} \xi_k(mp)\, \mathbf{w}_k + \sum_{k \notin \mathcal{C}} \xi_k(mp)\, \mathbf{w}_k, \tag{8}$$

where $\boldsymbol{\xi}_C(m) := \text{vec}\{\xi_k(mp), k \in \mathcal{C}\}$ is a multivariate stationary random process to be further modeled, and $\boldsymbol{\xi}_N(m) := \text{vec}\{\xi_k(mp), k \notin \mathcal{C}\}$ is white noise with $\xi_k(mp) \sim \text{IID}(0, \sigma_k^2)$ and is independent of $\{\boldsymbol{\xi}_C(m)\}$. In this model, the projection coefficients $\xi_k(mp)$ are assumed to be stationary across the periods (i.e., with respect to m) but are allowed to be nonstationary within the periods (i.e., with respect to k). Since correlated coefficients are permitted, the model in (8) can be regarded as a generalization of the concept of harmonizable processes by including non-sinusoidal basis functions and non-independent random coefficients (increments).

Let $\mathbf{R}(j) := [r_{kk'}(j)]$ be the autocovariance function (ACF) of $\{\boldsymbol{\xi}(m)\}$, where $r_{kk'}(j) := \text{Cov}\{\xi_k((m+j)p), \xi_{k'}(mp)\}$. Then, under (8), the ACF of $\{x(t)\}$ can be expressed as

$$\begin{aligned} C(t,t') &:= \text{cov}\{x(t), x(t')\} \\ &= \mathbf{w}(\tau)^T\, \mathbf{R}(m-m')\, \mathbf{w}(\tau') \\ &= \sum_{k,k' \in \mathcal{C}} r_{kk'}(m-m')\, w_k(\tau)\, w_{k'}(\tau') + \sum_{k \notin \mathcal{C}} \sigma_k^2\, w_k(\tau)\, w_k(\tau'), \end{aligned}$$

where $t = mp - \tau$ and $t' = m'p - \tau'$. Clearly, $C(t,t')$ is a function of $m - m'$, τ, and τ', but it is not necessarily a function of $t - t' = (m-m')p - (\tau - \tau')$. Furthermore, it is easy to see that $E\{x(t+p)\} = E\{x(t)\}$ and $C(t+p, t'+p) = C(t,t')$ for any t and t'. A random process with this property is called *periodically correlated*, or *cyclostationary* (Gladyshev, 1961; Dehay and Hurd, 1994). In general, a time series $\{x(t)\}$ is periodically correlated if and only if the corresponding $\boldsymbol{\xi}(m)$ defined by (4) is stationary. Therefore, the model in (8) defines a *special* cyclostationary process: Instead of any stationary $\boldsymbol{\xi}(m)$, this model imposes a structure on the stationarity in which a subset of $\boldsymbol{\xi}(m)$ is assumed to be white noise and independent of the rest. The model can also be viewed as a generalized regression model in which the coefficients change randomly across the periods and the error term has a special *non-i.i.d.* structure.

Under the assumptions in (8), the non-High-C waveforms do not make any contribution to the statistical inference of the High-C waveforms, and vise versa, because $\{\boldsymbol{\xi}_N(m)\}$ and $\{\boldsymbol{\xi}_C(m)\}$ are mutually independent. The non-High-C waveforms are also unpredictable because their coefficients

are independent of each other, both element-wise (across the wavenumbers) and across the periods. Therefore, one can safely ignore them in the subsequent modeling. The best forecast of $\xi_k(mp)$ for $k \notin \mathcal{C}$ is simply its mean: $\mu_k = 0$ (recall that $\xi_k(mp) \sim \text{IID}(0, \sigma_k^2)$ for $k \notin \mathcal{C}$).

3.3 Detection of High-C Waveforms

To detect the High-C waveforms, consider, for each fixed k, the hypothesis testing problem:

$$H_0 : \ \mu_k = 0 \quad \text{vs.} \quad H_1 : \ \mu_k \neq 0. \tag{9}$$

Given n observations of $\xi_k(mp)$, let $\hat{\mu}_k$ and $\hat{\sigma}_k^2$ be the sample mean and the sample variance, respectively. Then, we have the following results (see Appendix for a proof).

Proposition 1 *For each k, if $\{\xi_k(mp)\}$ is a white noise process, then, under H_0 in (9), $n\,\hat{\mu}_k^2/\hat{\sigma}_k^2 \sim \chi^2(1)$ asymptotically. In general, if $\{\xi_k(mp)\}$ is a general linear process, then the assertion remains valid, provided $\hat{\sigma}_k^2$ is replaced by a consistent estimator of $f_k(0) := \sum_{j=-\infty}^{\infty} r_{kk}(j)$, which is the spectral density of $\{\xi_k(mp)\}$ at frequency zero*[2].

According to Proposition 1, the index set $\mathcal{C}$ can be constructed as follows: k is included in $\mathcal{C}$ if and only if $\chi_k^2 := n\,\hat{\mu}_k^2/\hat{\sigma}_k^2 > \chi_{1-\alpha}^2(1)$, where $\chi_z^2(1)$ is the CDF of $\chi^2(1)$ and $\alpha \in (0,1)$ is the (approximate) significance level of the test. Let $\hat{\gamma}_k := \hat{\mu}_k^2/(\hat{\mu}_k^2 + \hat{\sigma}_k^2)$ be the estimated coherence. Then, it is easy to see that $\hat{\gamma}_k = \chi_k^2/(n + \chi_k^2)$. Therefore, an equivalent procedure for constructing $\mathcal{C}$ is as follows: k is included in $\mathcal{C}$ if and only if $\hat{\gamma}_k > \theta$, where $\theta := \chi_{1-\alpha}^2(1)/(n + \chi_{1-\alpha}^2(1))$. By Definition 1, the waveforms selected in this way are High-C waveforms.

For a successful application of the proposed approach, especially for good short-term forecasting, it may be necessary to include in $\mathcal{C}$ some additional components that contribute significantly to the total variation of the time series (i.e., large σ_k^2, even though $\mu_k = 0$). Since these components may be correlated with the High-C waveforms, lumping them into the second term of (8) may give rise to inferior forecasts due to a mismatch with the assumptions in (8). Incidentally, the components with large σ_k^2 can also be called High-C, "C" for contribution. In general, the filter bank should be designed such that the resulting High-C waveforms in $\mathcal{C}$ are *efficient* in the sense that $\sum_{k\in\mathcal{C}}(\mu_k^2 + \sigma_k^2) \gg \sum_{k\notin\mathcal{C}}(\mu_k^2 + \sigma_k^2)$ and $q \ll p$, where q denotes the cardinality of $\mathcal{C}$. A measure of efficiency is the *entropy* defined by $H := -\sum_{k=1}^{p} \rho_k \log \rho_k$, where $\rho_k := (\mu_k^2 + \sigma_k^2)/\sum_{j=1}^{p}(\mu_j^2 + \sigma_j^2)$.

[2] This quantity, which can be infinite for long-memory processes, is assumed to be finite in Proposition 1.

The smaller the entropy, the fewer the High-C waveforms, and thus the more efficient the proposed approach.

4 Parametric Modeling

Once the High-C waveforms are detected, the next step is to model their inter-period dynamics in order to forecast them. Without loss of generality, let us assume that $\mathcal{C} = \{1, \cdots, q\}$ for some $0 < q < p$ so that $\boldsymbol{\xi}_C(m) = [\xi_1(mp), \cdots, \xi_q(mp)]^T$ and $\boldsymbol{\mu}_C := E\{\boldsymbol{\xi}_C(m)\} = [\mu_1, \cdots, \mu_q]^T$. In the following, we assume that $\{\boldsymbol{\xi}_C(m)\}$ is a causal vector AR (VAR) process of the form

$$\sum_{j=0}^{\kappa} \mathbf{A}_j \{\boldsymbol{\xi}_C(m-j) - \boldsymbol{\mu}_C\} = \boldsymbol{\epsilon}(m), \tag{10}$$

where $\mathbf{A}_0 = \mathbf{I}$ and $\boldsymbol{\epsilon}(m) \sim \text{IID}(\mathbf{0}, \boldsymbol{\Sigma}_\epsilon)$. Note that the VAR model is not the only choice for modeling $\{\boldsymbol{\xi}_C(m)\}$. Depending on the situation, other parametric models (such as vector ARMA or state-space models) may be more effective in representing the dynamics of $\{\boldsymbol{\xi}_C(m)\}$. We choose not to discuss the model selection issue here because there already exist many good references on this subject, a classic example being Box and Jenkins (1976). The VAR model (10) is employed here mainly because of its versatility and computational simplicity.

Given the observations $\{\boldsymbol{\xi}_C(1), \cdots, \boldsymbol{\xi}_C(n)\}$ with $n > \kappa$, one can compute the sample mean $\hat{\boldsymbol{\mu}}_C$ and the sample ACF $\hat{\mathbf{R}}_C(j)$ for $|j| \leq \kappa$. Based on these summary statistics, the estimates of $\{\mathbf{A}_1, \cdots, \mathbf{A}_\kappa\}$ and $\boldsymbol{\Sigma}_\epsilon$ can be easily obtained by the multivariate Levinson-Durbin algorithm (e.g., Brockwell and Davis, 1991, pp. 422–423). The order κ can be determined by the standard criteria such as AIC (e.g., Brockwell and Davis, 1991, p. 432).

5 Forecasting

Facilitated with the VAR model (10), one can forecast $\{x(t)\}$ for the next $h \geq 1$ periods on the basis of the past observations $\{\mathbf{x}(1), \cdots, \mathbf{x}(n)\}$.

To derive the forecasts of $\{x(t)\}$, we first need the following lemma, a proof of which can be found in Brockwell and Davis (1991, pp. 175–184).

Lemma 1 *Assume that* $\{\boldsymbol{\xi}_C(m)\}$ *satisfies* (10). *Given* $\{\boldsymbol{\xi}_C(1), \cdots, \boldsymbol{\xi}_C(n)\}$, $\boldsymbol{\mu}_C$, κ, $\{\mathbf{A}_1, \cdots, \mathbf{A}_\kappa\}$, *and* $\boldsymbol{\Sigma}_\epsilon$, *the best linear unbiased predictor (BLUP) of* $\boldsymbol{\xi}_C(n+h)$ *is*

$$\hat{\boldsymbol{\xi}}_C(n+h) := \boldsymbol{\mu}_C + \tilde{\boldsymbol{\xi}}_C(n+h),$$

where $\tilde{\boldsymbol{\xi}}_C(n+h)$ *is computed recursively according to*

$$\tilde{\boldsymbol{\xi}}_C(n+\ell) := -\sum_{j=1}^{\kappa} \mathbf{A}_j \, \tilde{\boldsymbol{\xi}}_C(n+\ell-j) \quad (\ell = 1, \cdots, h),$$

with the initial values $\tilde{\boldsymbol{\xi}}_C(n-j) := \boldsymbol{\xi}_C(n-j) - \boldsymbol{\mu}_C$ $(j = 0, 1, \cdots, m-1)$. *The mean-squared prediction error*[3] *is given by* $\boldsymbol{\Sigma}_C(h) := E\{[\boldsymbol{\xi}_C(n+h) - \hat{\boldsymbol{\xi}}_C(n+h)]\,[\boldsymbol{\xi}_C(n+h) - \hat{\boldsymbol{\xi}}_C(n+h)]^T\} = \sum_{j=0}^{h-1} \mathbf{B}_j \boldsymbol{\Sigma}_\epsilon \mathbf{B}_j^T$, *where the* $\mathbf{B}_j$ *are the coefficient matrices in the vector* $MA(\infty)$ *representation of* $\{\boldsymbol{\xi}_C(m)\}$.

Equipped with this lemma, we obtain the following proposition immediately.

Proposition 2 *Assume that* (8) *and* (10) *are satisfied. Given the observations* $\{\mathbf{x}(1), \cdots, \mathbf{x}(n)\}$ *and the model parameters* $\mathcal{C}$, $\boldsymbol{\mu}_C$, κ, $\{\mathbf{A}_1, \cdots, \mathbf{A}_\kappa\}$, *and* $\boldsymbol{\Sigma}_\epsilon$, *the best* h*-period-ahead linear unbiased predictor of* $\{x(t)\}$ *can be expressed as*

$$\hat{\mathbf{x}}(n+h) := \sum_{k \in \mathcal{C}} \hat{\xi}_k((n+h)p)\, \mathbf{w}_k. \tag{11}$$

An equivalent expression of the predictor is

$$\hat{x}(t+hp) := \sum_{k \in \mathcal{C}} \hat{\xi}_k((n+h)p)\, w_k(\tau), \tag{12}$$

where $t = np - \tau$ *and* $\tau = 1, \cdots, p$. *Without loss of generality, assume that* $\mathcal{C} = \{1, \cdots, q\}$. *Then,* $\hat{\xi}_k((n+h)p)$ *is the* k*th element of* $\hat{\boldsymbol{\xi}}_C(n+h)$ *obtained from Lemma 1. Furthermore, the mean-squared prediction error is given by* $\boldsymbol{\Sigma}_x(h) := E\{[\mathbf{x}(n+h) - \hat{\mathbf{x}}(n+h)]\,[\mathbf{x}(n+h) - \hat{\mathbf{x}}(n+h)]^T\} = \mathbf{W}^T \boldsymbol{\Sigma}_\xi(h)\, \mathbf{W}$, *where* $\boldsymbol{\Sigma}_\xi(h) := \mathrm{diag}\{\boldsymbol{\Sigma}_C(h), \boldsymbol{\Sigma}_N\}$ *and* $\boldsymbol{\Sigma}_N := \mathrm{diag}\{\sigma_{q+1}^2, \cdots, \sigma_p^2\}$.

For the VAR model, the predictor $\hat{\boldsymbol{\xi}}_C(n+h)$ tends to $\boldsymbol{\mu}_C$ as the lead period h approaches infinity. Consequently, the forecast $\hat{\mathbf{x}}(n+h)$ converges in the long run to $\boldsymbol{\eta}_C := \mathbf{W}^T[\boldsymbol{\mu}_C^T, \mathbf{0}^T]^T$ (assuming $\mathcal{C} = \{1, \cdots, q\}$). This limiting value coincides with the seasonal mean $\boldsymbol{\eta} := E\{\mathbf{x}(m)\} = \mathbf{W}^T \boldsymbol{\mu}$ under the assumptions in (8) because $\mu_k = 0$ for all $k \notin \mathcal{C}$. In general, one can regard $\boldsymbol{\eta}_C$ as a *dimension-reduced* version of the seasonal mean. It can be shown that when estimated from data, the dimension-reduced seasonal mean is able to produce smaller mean-squared error than the full sample mean of $\mathbf{x}(m)$ in situations where $|\mu_k|$ is sufficiently small for $k \notin \mathcal{C}$. Furthermore, unlike the long-term SARIMA forecasts, the dimension-reduced seasonal mean has less variability in the presence of changing seasonality, because it depends on the *average*, rather than the individual, seasonal patterns. Note that the variability of a SARIMA process grows boundlessly in the long run because of the seasonal unit roots. Unless the unbounded variability can be justified, one has to exercise certain caution when employing SARIMA in long-term forecasting.

[3] For mean-squared error with estimated parameters, see Lütkepohl (1993, pp. 85–89).

6 Examples of Filter Bank

The aim of the STFA filter banks is to extract the High-C waveforms as discussed in Section 3. In principle, any orthogonal transformation of $\mathbf{x}(m)$ is a valid filter bank in the proposed method. However, different types of waveforms in $\mathbf{x}(m)$ require different filter banks to detect effectively. In this section, we provide two examples of such filter banks, the use of which is partially motivated by their success in applications such as data representation (coding) and data compression (e.g., Vetterli and Kovačević, 1995). We emphasize the unique interpretations of these filter banks because they are important factors in selecting suitable filter banks for a given data set.

6.1 Fourier Analysis

The Fourier filter bank is of special interest because it decomposes the seasonal vectors into sines and cosines that play a fundamental role in the interpretation of many physical phenomena. Hannan (1964) was among the first to use random-coefficient sinusoids to describe changing seasonal patterns. His main focus, however, was on the case of small p in which dimension reduction may be unnecessary. In this article, we consider not only the modeling and forecasting of the random coefficients but also the detection and extraction of what is known as the "stable pattern."

With $\omega_j := 2\pi j/p$ $(j = 0, 1, \cdots)$, let $\mathbf{c}_j := (2/p)^{1/2}[\cos\omega_j, \cos 2\omega_j, \cdots, \cos p\omega_j]^T$ and $\mathbf{s}_j := (2/p)^{1/2}[\sin\omega_j, \sin 2\omega_j, \cdots, \sin p\omega_j]^T$. Then, the Fourier filter bank takes the following form: If p is even, $\mathbf{w}_1 = \frac{1}{\sqrt{2}}\mathbf{c}_0$, $\mathbf{w}_{2j} = \mathbf{c}_j, \mathbf{w}_{2j+1} = \mathbf{s}_j$ $(j = 1, \cdots, \frac{1}{2}p - 1)$ and $\mathbf{w}_p = \frac{1}{\sqrt{2}}\mathbf{c}_{p/2}$. If p is odd, $\mathbf{w}_1 = \frac{1}{\sqrt{2}}\mathbf{c}_0$, $\mathbf{w}_{2j} = \mathbf{c}_j$, and $\mathbf{w}_{2j+1} = \mathbf{s}_j$ $(j = 1, \cdots, \frac{1}{2}(p-1))$. In both cases $\boldsymbol{\xi}(m)$ is equal to the real-form discrete Fourier transform (DFT) of $\mathbf{x}(m)$.

Conventionally, the DFT of $\mathbf{x}(m)$ is defined in the complex form

$$z_j(m) := \frac{1}{\sqrt{p}} \sum_{\tau=1}^{p} x(mp - \tau)\, e^{i\omega_j(mp-\tau)} \quad (j = 0, 1, \cdots, p-1),$$

where $i := \sqrt{-1}$. Assuming that p is even (similar comments can be made when p is odd), one obtains $z_j(m) = \xi_j(mp)$ if $j = 0, \frac{1}{2}p$ and $z_j(m) = \frac{1}{\sqrt{2}}\{\xi_{2j}(mp) - i\,\xi_{2j+1}(mp)\}$ if $j = 1, \cdots, \frac{1}{2}p - 1$. Motivated by this relationship, we define the coherence of $\mathbf{x}(m)$ at frequency ω_j as

$$\gamma_j^c := \frac{|\mu_j^c|^2}{|\mu_j^c|^2 + (\sigma_j^c)^2},$$

where μ_j^c and σ_j^c are the mean and standard deviation of $z_j(m)$. Clearly, $\gamma_j^c = \gamma_j$ if $j = 0, \frac{1}{2}p$. For $j = 1, \cdots, \frac{1}{2}p - 1$, it is easy to show that $|\mu_j^c|^2 = \frac{1}{2}(\mu_{2j}^2 + \mu_{2j+1}^2)$ and $(\sigma_j^c)^2 = \frac{1}{2}(\sigma_{2j}^2 + \sigma_{2j+1}^2)$.

Therefore, it follows that

$$\gamma_j^c = \frac{\mu_{2j}^2 + \mu_{2j+1}^2}{\mu_{2j}^2 + \mu_{2j+1}^2 + \sigma_{2j}^2 + \sigma_{2j+1}^2}.$$

As can be seen, γ_j^c is a diagnostic that measures the *combined* coherence of the sine and cosine components with frequency ω_j. This diagnostic is useful when there is no need to distinguish between the sine and cosine waves with the same frequency.

Since $\mathbf{w}_{2j}$ and $\mathbf{w}_{2j+1}$ are sinusoids with the same frequency ω_j, it is convenient to combine $\mathbf{x}_{2j}(m)$ with $\mathbf{x}_{2j+1}(m)$ to form a combined waveform $\mathbf{y}_j(m) := [y_j(mp-1), \cdots, y_j(mp-p)]^T := \mathbf{x}_{2j}(m) + \mathbf{x}_{2j+1}(m)$. It is easy to show that

$$y_j(t) = \beta_j(m)\cos\{\omega_j t + \varphi_j(m)\} \quad (t = mp - \tau), \tag{13}$$

where $\beta_j(m) := (2/p)^{1/2}\{\xi_{2j}^2(mp) + \xi_{2j+1}^2(mp)\}^{1/2}$ and $\varphi_j(m) := -\arctan\{\xi_{2j+1}(mp)/\xi_{2j}(mp)\}$. As can be seen, within a period, the combined waveform $\mathbf{y}_j(m)$ behaves as a sinusoid with frequency ω_j; and across the periods, the waveform preserves the sinusoidal pattern but randomly changes its amplitude $\beta_j(m)$ and its phase $\varphi_j(m)$.

Equation (13) provides an opportunity to better understand the proposed approach. For long-term forecasting, the phase is critically important, because it changes the synchronization of the ups and downs in the waveform as the wave travels from one period to another. The waveform would be totally out of synchronization in the long run if the phase $\varphi_j(m)$ changes randomly with a uniform distribution in $(-\pi, \pi]$. This happens when $\xi_{2j}(mp)$ and $\xi_{2j+1}(mp)$ are *i.i.d.* Gaussian with zero mean, in which case $\gamma_{2j} = \gamma_{2j+1} = \gamma_j^c = 0$, so $\mathbf{x}_{2j}$ and $\mathbf{x}_{2j+1}$ are incoherent by definition. Exclusion of such waveforms in long-term forecasting is therefore justified.

Instead of separately examining the coherence of $\mathbf{w}_{2j}$ and $\mathbf{w}_{2j+1}$ as discussed in Section 3, one can combine these waveforms, for $j = 1, \cdots, \frac{1}{2}p - 1$, to determine whether the frequency ω_j is coherent in $\mathbf{x}(m)$. This can be done by considering the following hypotheses:

$$H_0 \colon\ \mu_{2j} = \mu_{2j+1} = 0 \quad \text{vs.} \quad H_1 \colon\ \mu_{2j} \neq 0 \text{ or } \mu_{2j+1} \neq 0. \tag{14}$$

The next corollary is a direct result from Proposition 1.

Corollary 1 *For fixed $j = 1, \cdots, \frac{1}{2}p - 1$, assume that $\{\xi_{2j}(mp)\}$ and $\{\xi_{2j+1}(mp)\}$ are mutually-uncorrelated general linear processes. Then, under H_0 in (14), $n\{\hat{\mu}_{2j}^2/\hat{f}_{2j}(0) + \hat{\mu}_{2j+1}^2/\hat{f}_{2j+1}(0)\} \sim \chi^2(2)$ asymptotically, where $\hat{f}_{2j}(0)$ and $\hat{f}_{2j+1}(0)$ are consistent estimators of $f_{2j}(0)$ and $f_{2j+1}(0)$. The $\chi^2(1)$ test in Proposition 1 remains valid for testing the coherence of frequency ω_0 and, when p is even, of frequency $\omega_{p/2}$.*

6.2 Wavelet Analysis

Whereas the seasonal patterns detected by Fourier analysis are persistent in terms of their frequencies, there are plenty of examples in practice where patterns are better characterized in terms of their scales. In particular, some patterns may be transient within seasonal vectors but repeat themselves coherently from one period to another (e.g., annual promotions in weekly sales data, daily rush hours in hourly air pollution data). Such transient patterns are difficult to detect by Fourier analysis, because the fixed long time-scale of Fourier analysis, which is equal to the length of the entire period, prevents the Fourier filters from effectively isolating the transient patterns in the presence of non-transient ambient noise.

Wavelet filter banks, on the other hand, allow the time scale to vary across the filters and thus provide a powerful tool for multiresolution (multiscale) analysis. Using the wavelet filters, not only can one take a 'grand view' of a seasonal vector but also zoom into smaller fractions of the seasonal vector for a 'closer look.' This zoom-in-zoom-out capability, combined with the coherence measure, makes it possible to effectively detect persistent patterns of variable sizes, especially those that are considerably shorter than the length of the period.

For compactly-supported orthogonal wavelet filter banks (e.g., Daubechies, 1992), the corresponding $\mathbf{W}$ not only is an orthogonal matrix but also has highly interpretable structures, as shown in Figure 2. In general, $\mathbf{W}$ can be partitioned into $L+1$ subbanks $\mathbf{W} = [\mathbf{W}_1^T, \cdots, \mathbf{W}_{L+1}^T]^T$, where L is defined as the largest integer such that $2^{-L}p$ is odd, or as the smallest integer such that $2^{-L}p$ is less than N (the length of the shortest wavelet filter), whichever is smaller. For $\ell = 1, \cdots, L$, the subbank $\mathbf{W}_\ell$ contains $2^{-\ell}p$ equal-length filters which are circularly translated copies of each other: the jth filter in $\mathbf{W}_\ell$ can be obtained by circularly translating the first one $(j-1)2^\ell$ samples to the right. These filters are bandpass in the frequency domain, with the effective passband $(2^{-\ell}\pi, 2^{-\ell+1}\pi)$. The length of the filters in $\mathbf{W}_\ell$ doubles (approximately) as ℓ is increased to $\ell+1$ (for $\ell < L$). The last subbank $\mathbf{W}_{L+1}$ has the same structure as $\mathbf{W}_L$ except that it consists of lowpass filters with the effective passband $(0, 2^{-L}\pi)$.

To understand the potential of wavelet filters for extracting multiscale seasonal patterns, it is necessary to briefly discuss the wavelet-based *multiresolution analysis* (Mallat, 1989). Let $\mathcal{S}_\ell$ and $\mathcal{W}_\ell$ represent the linear subspaces formed by the wavelets (row vectors) in $\{\mathbf{W}_{L+1}, \cdots, \mathbf{W}_\ell\}$ and in $\mathbf{W}_\ell$, respectively. Then, the orthogonality of $\mathbf{W}$ implies that

$$\mathcal{S}_\ell = \mathcal{S}_{\ell+1} \oplus \mathcal{W}_\ell \quad (\ell = L, L-1, \cdots, 1). \tag{15}$$

Since the wavelets in $\mathcal{W}_\ell$ are shorter than those in $\mathcal{S}_{\ell+1}$ (for $\ell < L$), the subspace $\mathcal{S}_\ell$ in (15) can be interpreted as being formed by adding to $\mathcal{S}_{\ell+1}$ the higher-resolution innovations (or details) that cannot be attributed to any wavelets in $\mathcal{S}_{\ell+1}$. The subspace $\mathcal{S}_{L+1}$ has the lowest resolution, which is true not only because $\mathcal{S}_{L+1}$ is formed by the longest filters, but also because the filters in $\mathcal{S}_{L+1}$ are lowpass (less oscillatory) whereas those in $\mathcal{W}_L$ are bandpass (more oscillatory) even though they have the same length. In summary, the subspaces $\mathcal{S}_{L+1} \subset \mathcal{S}_L \subset \cdots \subset \mathcal{S}_1$ constitute a nested chain of increasing resolution (decreasing scale).

By using the wavelet filter bank, a data vector $\mathbf{x}$ can be decomposed as a low-resolution trend and a sequence of higher-resolution innovations. More precisely, let $\boldsymbol{\xi} := [\boldsymbol{\xi}_1^T, \cdots, \boldsymbol{\xi}_{L+1}^T]^T := \mathbf{W}\mathbf{x}$, where $\boldsymbol{\xi}_\ell := \mathbf{W}_\ell \mathbf{x}$ is the (subsampled) output from the subbank $\mathbf{W}_\ell$. Then, for any $\ell = 1, \cdots, L$, the data vector $\mathbf{x} = \mathbf{W}^T\boldsymbol{\xi}$ can be expressed as

$$\mathbf{x} = \mathbf{s}_{\ell+1} + \mathbf{d}_\ell + \cdots + \mathbf{d}_1, \tag{16}$$

where $\mathbf{s}_{\ell+1} := \mathbf{W}_{L+1}^T\boldsymbol{\xi}_{L+1} + \cdots + \mathbf{W}_{\ell+1}^T\boldsymbol{\xi}_{\ell+1} \in \mathcal{S}_{\ell+1}$ and $\mathbf{d}_{\ell'} := \mathbf{W}_{\ell'}^T\boldsymbol{\xi}_{\ell'} \in \mathcal{W}_{\ell'}$ $(\ell' = 1, \cdots, \ell)$. In this expression, $\mathbf{s}_{\ell+1}$ can be interpreted as a trend in $\mathbf{x}$ whose scale is greater than or equal to $\ell + 1$, and $\{\mathbf{d}_\ell, \cdots, \mathbf{d}_1\}$ can be interpreted as a sequence of details in $\mathbf{x}$ whose scale decreases from ℓ to 1.

7 Numerical Experiments

Comparing different modeling and forecasting methods in any generality is a formidable, if not impossible, task because different methods are based on different assumptions about the seasonality that may not be entirely compatible. These assumptions, and hence the corresponding methods, should be justified only within the context of specific data (or types of data). Therefore, the objective of this section is not to show the superiority of the proposed approach over the others in handling all seasonalities. Rather, it is to demonstrate the possibility of an alternative modeling approach for seasonal time series that are capable of producing improved performance under suitable conditions. Whether or not the proposed method is successful for specific data depends on whether or not the model assumptions can be justified. More elaborate experiments are needed to draw any general conclusions, but they are beyond the scope of this article.

To measure the performance of forecasts up to a given horizon, we employ the root mean-squared error (RMSE), defined as the square-root of $\mathrm{MSE}(1,h) := h^{-1}\sum_{j=1}^{h}\mathrm{MSE}(j)$, and the R^2, defined as $R^2(1,h) := 1 - \mathrm{MSE}(1,h)/\mathrm{MST}(1,h)$, where $\mathrm{MST}(1,h) := h^{-1}\sum_{j=1}^{h}\mathrm{MST}(j)$, $\mathrm{MSE}(j) := p^{-1}\|\mathbf{x}(n+j) - \hat{\mathbf{x}}(n+j)\|^2$ and $\mathrm{MST}(j) := p^{-1}\|\mathbf{x}(n+j) - \bar{x}(n+j)\mathbf{1}\|^2$. In these expressions, $\|\cdot\|$

denotes the L_2 norm, $\bar{x}(n+j)$ is the intra-period sample mean of $\{x(t)\}$ in period $n+j$, and $\mathbf{1}$ is the vector of unity of length p.

7.1 Simulation Study.

Before testing the proposed method on real data, it is necessary to understand the method under controlled conditions. To this end, we carry out the following simulation.

In this simulation study, the seasonal time series consists of three sinusoidal components with period $p = 24$, plus a Gaussian white noise (GWN) component; the amplitudes of the sinusoids remain constant but the phases change randomly from one period to another as independent AR(1) processes. More precisely, $x(t) = s(t) + e(t)$, where $e(t) \sim \text{GWN}(0, \sigma^2)$, $s(t) := \sum_{i=1}^{3} \beta_i \cos\{2\pi f_i t + \phi_i(m)\}$ $(t = mp - p, \cdots, mp - 1)$, $\phi_i(m) := \pi\{b_i\, \psi_i(m) - c_i\}$, $\psi_i(m) = a_i\, \psi_i(m-1) + \epsilon_i(m)$, and $\epsilon_i(m) \sim \text{GWN}(0, 1 - a_i^2)$. In this experiment, $(\beta_1, \beta_2, \beta_3) = (1, 0.7, 0.7)$, $(f_1, f_2, f_3) = (1/p, 2/p, 4/p)$, $(a_1, a_2, a_3) = (0.8, 0.8, 0.8)$, $(b_1, b_2, b_3) = (0.2, 0.3, 0.4)$, and $(c_1, c_2, c_3) = (1, 0.5, 0.1)$. The value of σ^2 is chosen such that the 'signal'-to-'noise' ratio, SNR $:=$ $\text{var}\{s(t)\}/\sigma^2$, is equal to 10 dB. A realization of $x(t)$ is shown in Figure 3.

Based on the DFT of $n = 50$ seasonal vectors of this time series, the chi-square statistics are calculated according to Corollary 1 (under the white noise assumption). Plotting these statistics against the frequency gives rise to the coherence plot shown in Figure 3. As can be seen, three frequencies are statistically significant in the chi-square test. Corresponding to these frequencies are six (not three) High-C coefficient sequences, $\{\xi_{2j}(mp)\}$ and $\{\xi_{2j+1}(mp)\}$ $(j = 1, 2, 4)$. In this experiment, they are modeled as a VAR(1) process, where the optimal order $\kappa = 1$ is selected by the AIC criterion among $\kappa = 0, 1, 2$. The VAR parameters are estimated on the basis of the 50-period training data. Note that using only the High-C components gives rise to a 75% (24 to 6) dimension reduction. The reduced dimension makes it possible that the 50-period training data suffice for reliable VAR estimation. (A full VAR model would result in $d.f. = n - p\kappa - 1 = 1$ for the error estimation in the case of $\kappa = 2$.) Based on the estimated model, Figure 4 shows the forecasts of a ten-period segment that immediately follows the training data.

The same segment is forecasted by a SARIMA $(0, 0, 1) \times (0, 1, 1)_p$ model (i.e., the famous "airline model" without non-seasonal integration). The forecasts, shown in Figure 5, are obtained by the internal S-Plus functions, `arima.mle` and `arima.forecast`, which perform the maximum likelihood estimation (very time consuming because of the iterative optimization routine used) and the state-space forecasting by a Kalman filter (e.g., Harvey, 1993).

For the ten-period forecasts in Figure 4, the proposed method yields RMSE = 0.702 and R^2 = 56.4%. The SARIMA method produces much inferior forecasts, with RMSE = 0.863 and R^2 = 34.4%. The rapid degradation of long-term SARIMA forecasts, as can be seen in Figure 5, is mainly responsible for the inferior overall performance. In this particular case, SARIMA is even inferior to the sample seasonal mean (not shown) which gives rise to RMSE = 0.818 and R^2 = 40.8%. The improved performance of the proposed method over the seasonal mean can be attributed to its utilization of the inter-period correlation of the High-C components (in short-term forecasts) and to its proper reduction of dimensionality (in long-term forecasts).

For further comparison, consider Table 1 and Figure 6 where the performance indices are computed on the basis of 140 Monte Carlo runs. The chi-square test with $\alpha = 0.001$ is employed to detect the High-C waveforms. Note that the RMSE and R^2 in Table 1 are defined by using the *horizon-specific* quantities MSE(h) and MST(h) in place of the *horizon-averaged* quantities MSE($1, h$) and MST($1, h$) which are used in Figure 6. Therefore, Table 1 shows the performance for the specific forecasting horizons whereas Figure 6 shows the averaged performance over the lead periods up to the specified horizon.

As can be seen from Table 1 and Figure 6, SARIMA outperforms the proposed method (and the others) in one-period-ahead predictions, but the proposed method produces more accurate results for longer horizons. The seasonal-mean forecasts can be derived from a regression model in which the seasonal pattern, described by p dummy variables, is assumed to be deterministic rather than stochastic. This model, which may be viewed as the antithesis of SARIMA, is very good at capturing the asymptotic behavior of the seasonal patterns but performs poorly in tracking their short-term random fluctuations. In this experiment, the proposed method achieves a compromise between the two and provides satisfactory forecasts for a broader range of horizons.

The performance of the simplest periodic AR model, PAR(1), is also included in Table 1 and Figure 6. In this experiment, the PAR(1) results are nearly identical to those of the seasonal mean, although slightly better for one-period-ahead forecasting. The PAR(1) model, which generates a special periodically correlated process as does the proposed model in (8), assumes that $\tilde{x}(t) = \alpha(\tau)\,\tilde{x}(t-1) + \sigma(\tau)\,\epsilon(t)$ $(t = mp - \tau)$, where $\tilde{x}(t) := x(t) - E\{x(t)\}$ and $\epsilon(t) \sim \text{IID}(0,1)$. The parameters $\alpha(\tau)$ and $\sigma(\tau)$ are estimated by a procedure discussed in Bloomfield *et al.* (1994). One can show that PAR(1) is a special VAR(1) model for the seasonal vectors: $\mathbf{x}(m) = \boldsymbol{\eta} + \mathbf{A}\{\mathbf{x}(m-1) - \boldsymbol{\eta}\} + \mathbf{B}\,\boldsymbol{\epsilon}(m)$, where $\boldsymbol{\epsilon}(m) \sim \text{IID}(\mathbf{0}, \mathbf{D})$ and $\mathbf{D} := \text{diag}\{\sigma^2(1), \cdots, \sigma^2(p)\}$. The matrices $\mathbf{A}$ and $\mathbf{B}$ are determined by the seasonal AR coefficients $\alpha(\tau)$: the kth element in the first column of

Table 1. Comparison of Performance

Horizon	Proposed Method		Seasonal ARIMA		Periodic AR(1)		Seasonal Mean	
(period)	RMSE	R^2 (%)	RMSE	R^2 (%)	RMSE	R^2 (%)	RMSE	R^2 (%)
1	.6494	57.51	.6261	60.57	.8008	36.40	.8012	36.28
2	.6806	53.26	.6997	49.94	.8101	34.66	.8101	34.66
3	.7491	42.74	.7516	41.69	.8174	33.47	.8174	33.47
4	.7619	41.09	.7928	35.03	.8205	32.72	.8205	32.72
5	.7880	37.21	.8270	29.27	.8241	32.04	.8241	32.04
6	.7967	35.93	.8471	25.52	.8250	31.93	.8250	31.93
7	.7995	35.40	.8637	22.36	.8250	31.81	.8250	31.81
8	.8066	34.55	.8818	19.56	.8276	31.50	.8276	31.50
9	.8008	35.16	.8928	17.75	.8268	31.50	.8268	31.50
10	.8039	34.76	.8979	17.04	.8273	31.43	.8273	31.43

$\mathbf{A}$ is equal to $\prod_{\tau=1}^{p-k+1} \alpha(\tau)$ and the remaining columns are all equal to zero; the (k, k')th element of $\mathbf{B}$ is equal to $\prod_{\tau=p-k'+2}^{p-k+1} \alpha(\tau)$ if $k < k'$, equal to unity if $k = k'$, and equal to zero otherwise. Therefore, given $\boldsymbol{\eta}$ and the $\alpha(\tau)$, the PAR(1) forecasts of the seasonal vectors can be obtained from the equivalent VAR model. To compare PAR(1) with the proposed method in this experiment, one should bear in mind that PAR(1) employs $2p = 48$ estimated parameters (p seasonal means and p seasonal AR coefficients), and this number is greater than that of the proposed method, which is equal to $q^2 + q = 42$ ($q = 6$). Therefore, the inferior performance of PAR(1) is not because it is based on a simpler model. Rather, it should be attributed to the different model assumptions.

Note that this simulation study employs a parametric process to generate the data simply because the properties of the process can be easily controlled. The entire modeling and forecasting approach is better viewed as nonparametric or semiparametric, because not all information about the parametric data-generating process is fully utilized. For example, although the $\phi_i(m)$ have AR(1) structures, the coefficients $\beta_i \cos(\phi_i(m))$ and $\beta_i \sin(\phi_i(m))$ of the corresponding zero-phase sinusoids are not exactly AR(1) processes. In fact, these coefficients evolve *nonlinearly* with time. This nonlinear structure is ignored in the modeling-forecasting procedure. It is also ignored that the sum of the squared coefficients, being equal to β_i^2, remain constant across the periods. Improved performance is expected if these structural constraints are enforced in the modeling.

7.2 Global Irradiance

Now consider the global irradiance data shown in Figure 1(a) ($t = 0 \bmod 24 =$ solar noon). The data were collected in the summer of 1992 by the weather station Delta at Taconite Inlet Lake C2 in northern Ellesmere Island, Canada. More information about the data and the Taconite Inlet Project is given in Bradley *et al.*, (1996).

Because of the sinusoid-like seasonal patterns, the Fourier filter bank is chosen. As shown in Figure 7, a strong and highly coherent daily frequency is detected by Fourier analysis. For this component, a VAR(1) model is estimated. The resulting one-week-ahead forecasts are shown in Figure 8, together with the training data and the actual observations for the seven-day lead period. Figure 8 also contains the forecasts of a SARIMA $(0, 0, 1) \times (0, 1, 1)_p$ model and the forecasts of a PAR(1) model for the same data.

As can be seen, the proposed method yields the most accurate predictions in this case. The poor performance of SARIMA is primarily due to its heavy dependence on the last day's observations in the training data, which in this case happen to be rather abnormal. The forecasts by the proposed method, and by PAR(1) to a lesser extent, are more robust because they are based on more persistent patterns.

To compare the averaged performance, Figure 9 shows the horizon-specific MSE and R^2 of the forecasts, averaged over a 25-day period. In each case, the detection, modeling, and forecasting are based on the two-week training data that immediately precede the seven-day forecasting period. The High-C waveforms are determined by the chi-square test with $\alpha = 0.001$. The VAR(1) model is always used for modeling the High-C coefficients.

Figure 9 shows that the proposed method provides the best forecasts for $h \geq 2$. The PAR(1) method gives the best one-day-ahead forecasts; but as the horizon increases, the performance of PAR(1) deteriorates rapidly and becomes nearly identical to that of the seasonal mean. The SARIMA method produces satisfactory one-day-ahead forecasts but the worst longer-term forecasts that can be explained largely by the lack of robustness as illustrated in Figure 8.

7.3 The Airline Data

The next example is the famous "airline data" that contain the monthly passenger totals in international air travel from 1949–1960 (Box and Jenkins, 1976). The original data exhibit an increasing trend as well as a strong seasonality. The former is removed by a quadratic regression since the

latter is our primary concern in this experiment. The regression is preceded by a natural log transform to stabilize the increasing variability. The pre-processing leads to a monthly time series, a fraction of which is shown in Figure 1(b). The objective of this exercise is to forecast the time series of the last three years on the basis of the previous nine-year observations.

Figure 10 shows the coherence plots of Fourier and wavelet analyses. At $\alpha = 0.0001$, Fourier analysis detects five significant frequencies that correspond to ten High-C components. The ratio of dimension reduction is therefore 12:10. A higher reduction ratio of 12:6 (50%) is achieved by wavelet analysis which detects six High-C components at the same significance level. This is not entirely surprising because the seasonal patterns in the airline data, as shown in Figure 1(b), are dominated by transient features that are very suitable for wavelet analysis (e.g., a sharp peak in July-August followed abruptly by a deep valley in November; a rapid rise from February to March followed by a small valley in May). Due to the limited training data, an independence constraint is imposed on the High-C components in the VAR modeling: each component is modeled as an independent univariate AR with an optimal order selected by AIC among the choices of $\kappa = 0, 1, 2, 3$.

Figure 10 also contains the wavelet-based forecasts along with the actual observations. The RMSE and R^2 over the three-year forecasting period are 0.0409 and 92.5%, respectively. The Fourier method (not shown) yields RMSE = 0.0479 and $R^2 = 89.7\%$.

Historically, the SARIMA method in general and the $(0,1,1)\times(0,1,1)_{12}$ model in particular have been closely associated with the airline data. In this example, the lag-one difference is unnecessary since the non-seasonal trend has been removed by regression. So the model reduces to $(0,0,1) \times (0,1,1)_{12}$. The forecasts by this model give rise to RMSE = 0.0476 and $R^2 = 89.8\%$, which are comparable to the Fourier forecasts but inferior to the wavelet forecasts. The PAR(1) and the seasonal-mean forecasts are the least accurate. For PAR(1): RMSE = 0.0546, $R^2 = 86.6\%$; and for the seasonal mean: RMSE = 0.0540, $R^2 = 86.8\%$.

7.4 The Retail Sales Index

Our final example concerns the monthly growth rates of retail sales in the Netherlands from October 1960 to September 1995. The time series, a ten-year fraction of which is shown in Figure 11, is obtained by lag-one differencing of the logs of the monthly retail sales index given in Franses (1998). The objective of this exercise is to forecast the time series up to 24 months (two years) on the basis of the previous 60-month (5-year) training data. To measure the forecasting performance, the RMSE and R^2 are averaged over 28 cases, each corresponding to a shift of the 5-year training

Table 2. Forecasting the Retail Sales Data

Horizon	Proposed (FT)		Proposed (WT)		Seasonal ARIMA		Periodic AR(1)		Seasonal Mean	
(period)	RMSE	R^2 (%)	RMSE	R^2 (%)	RMSE	R^2 (%)	RMSE	R^2 (%)	RMSE	R^2 (%)
1	.0516	71.2	.0526	70.6	.0539	68.8	.0547	67.7	.0556	66.4
2	.0546	66.7	.0549	67.2	.0734	42.7	.0558	65.1	.0558	65.1
Overall	.0539	69.0	.0542	69.1	.0650	55.9	.0558	66.7	.0562	66.0

window by $p = 12$ months.

Table 2 contains the horizon-specific and the horizon-averaged (overall) RMSE and R^2 of different methods. The High-C components in the proposed method are detected at $\alpha = 0.05$ and fitted with independent AR(2) models. The tenth-order least-asymmetric Daubechies wavelets are employed in the wavelet analysis. As can be seen, the proposed method (with both Fourier and wavelet filters) has the best overall performance as well as the best performance for both one-year and two-year horizons. This is followed by the PAR(1) method and the seasonal mean. The SARIMA $(0, 0, 1) \times (0, 1, 1)_{12}$ method has the third best performance for one-year forecasting but the worst performance for the two-year horizon; the latter is responsible for its poor overall performance. A closer examination also shows that for the 28 cases the proposed method with Fourier filters outperforms SARIMA 93% of the time in terms of the horizon-averaged RMSE (and R^2); it outperforms PAR(1) and the seasonal mean 70% and 67% of the time, respectively. The corresponding numbers for the proposed method with wavelet filters are 89%, 63%, and 67%.

8 Concluding Remarks

The method presented in this article utilizes the coherent (High-C) seasonal components, extracted by an STFA filter bank, to capture long-term seasonal effects; it employs the stochastic dynamics of the High-C components to model the short-term random fluctuation of the seasonality. The intra-period relationship of the seasonal patterns is described by the filter bank, and the inter-period variation is modeled by the autocorrelation structure of the changing High-C coefficients.

Since the definition of seasonal vectors is flexible in the proposed method, it is recommended that they be defined in such away that the leading element of the last seasonal vector in the training data coincides with the most recent observation of the time series. For example, if the most recent observation is made in November, then the seasonal vectors should contain the observations in

January through November as well as the observation in December of the previous year. Although it differs from the annual cycle of January through December of the same year, this definition always ensures that the latest observation be included in the training data. It is also the most natural definition from the viewpoint of forecasting because the resulting short-term as well as long-term forecasts always contain the most recent information about the time series. The model can be retrained (perhaps with redefined seasonal vectors) once new observations become available.

The article carries two important messages: (a) Seasonal patterns may be decomposed into 'relevant' (High-C) and 'irrelevant' components, and the limited computational and data resources should be directed toward the former. (b) Different seasonal patterns may require different filter banks to extract efficiently, and there are many possible ways to design the STFA filter bank to decompose seasonal vectors.

Seeking efficient filter banks has been a classic problem in signal processing for data representation and compression. Traditional methods include the Karhunen-Loève (KL) transform and the Fourier transform. More recently, the wavelet transform has found tremendous success. If estimated from training data, the KL transform corresponds to a filter bank that consists of the so-called empirical orthogonal functions (EOFs) which have many successful applications (e.g., Kim, North, and Huang, 1996). Because it is fully nonparametric without any structural constraints, the KL transform is often considered too costly to implement in terms of the computational complexity and the large amount of training data required for reliable estimation (i.e., $n \gg p\kappa$). The Fourier and wavelet transforms are superior in both computation and estimation aspects, although not necessarily optimal in efficiency. In the proposed approach, an efficient filter bank is required to extract waveforms that are also highly coherent. In practice, one may have to experiment with various possible filter banks using, for example, the so-called "best-basis selection" techniques (e.g., Mallat and Zhong, 1993; Chen, Donoho, and Saunders, 1998). These techniques, however, must be modified to suit the objectives of forecasting rather than data compression.

Appendix: Proof of Proposition 1

The central limit theory ensures that $\sqrt{n}\,(\hat{\mu}_k - \mu_k) \to N(0, \sigma_k^2)$ in distribution and the law of large numbers guarantees that $\hat{\sigma}_k^2 \to \sigma_k^2$ in probability. The first assertion follows from Slutsky's theorem. For general linear processes, it can be shown (e.g., Brockwell and Davis, 1992) that $\sqrt{n}\,(\hat{\mu}_k - \mu_k) \to N(0, f_k(0))$ in distribution, where $f_k(0)$ is the spectral density of $\{\xi_k(mp)\}$ at zero

frequency. Slutsky's theorem is applicable if $f_k(0)$ is estimated by a consistent estimator.

References

Bloomfield, P., Hurd, H. L., and Lund, R. B. (1994), "Periodic correlation in stratospheric ozone data," *Journal of Time Series Analysis*, vol. 15, no. 2, pp. 127–150.

Bradley, R. S., Retelle, M. J., Ludlam, S. D., Hardy, D. R., Zolitschka, B., and Lamoureux, S. F. (1996), "The Taconite Inlet Lakes Project: a systems approach to paleoclimatic reconstruction," *Journal of Paleolimnology*, vol. 16, pp. 97–110.

Box, G. E. P. and Jenkins, G. M. (1976), *Time Series Analysis, Forecasting and Control*, Revised Edition, San Francisco, CA: Holden-Day.

Brockwell, P. J. and Davis, R. A. (1991), *Time Series: Theory and Methods*, 2nd Edition, New York: Springer-Verlag.

Chen, S. S., Donoho, D. L., and Saunders, M. A. (1998), "Atomic decomposition by basis pursuit," *SIAM Journal of Scientific Computing*, vol. 20, no. 1, pp. 33–61.

Daubechies, I. (1992), *Ten Lectures on Wavelets*, Philadelphia, PA: Society for Industrial and Applied Mathematics.

Dehay, D. and Hurd, H. L. (1994), "Representation and estimation for periodically and almost periodically correlated random processes," in *Cyclostationarity in Communications and Signal Processing*, edited by W. A. Gardner, Piscataway, NJ: IEEE Press.

Franses, P. H. (1996), *Periodicity and Stochastic Trends in Economic Time Series*, Oxford, UK: Oxford University Press.

Franses, P. H. (1998), *Time Series Models for Business and Economic Forecasting*, Cambridge, UK: Cambridge University Press.

Funke, M. (1990), "Assessing the forecasting accuracy of monthly vector autoregressive models," *International Journal of Forecasting*, vol. 6, pp. 363–378.

Gladyshev, E. G. (1961), "Periodically correlated random sequences," *Soviet Math.*, vol. 2, pp. 385–388.

Hannan, E. J. (1964), "The estimation of a changing seasonal pattern," *Journal of the American Statistical Association*, vol. 59, pp. 1063–1077.

Harvey, A. C. (1993), *Time Series Models*, 2nd ed., Cambridge, MA: MIT Press.

Hinich, M. J. (1996), "A statistical theory of signal coherence," preprint.

Kim, K.-Y., North, G. R., and Huang, J. (1996), "EOFs of one-dimensional cyclostationary time series: computations, examples, and stochastic modeling," *Journal of the Atmospheric Sciences*, vol. 53, no. 7, pp. 1007–1017.

Kitagawa, G. and Gersch, W. (1996), *Smoothness Priors Analysis of Time Series*, New York: Sprinter.

Kulendran, N. and King, M. L. (1997), "Forecasting international quarterly tourist flows using error-correction and time-series models," *International Journal of Forecasting*, vol. 13, pp. 319–326.

Li, T. H. (1996), "Discrimination of time series by parametric filtering," *Journal of the American Statistical Association*, vol. 91, pp. 284–293.

Li, T. H. (1997), "Time series characterization, Poisson integral, and robust divergence measures," *Technometrics*, vol. 39, no. 4, pp. 357–371.

Li, T. H. and Gibson, J. D. (1996), "Speech analysis and segmentation by parametric filtering," *IEEE Trans. Audio and Speech Processing*, vol. 4, no. 2, pp. 203–213.

Litterman, R. B. (1986), "Forecasting with Bayesian vector autoregressions – five years of experience," *Journal of Business, and Economic Statistics*, vol. 4, pp. 25–38.

Lütkepohl, H. (1993), *Introduction to Multiple Time Series Analysis*, 2nd ed., New York: Springer-Verlag.

Mallat, S. (1989), "A theory of multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 11, pp. 674–693.

Mallat, S. and Zhong, Z. (1993), "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Processing*, vol. 41, no. 12, pp. 3397–3415.

Novales, A. and de Fruto, R. F. (1997), "Forecasting with periodic models: a comparison with time invariant coefficient models," *International Journal of Forecasting*, vol. 13, pp. 393–405.

Pagano, M. (1978), "On periodic and multiple autoregression," *Annals of Statistics*, vol. 6, pp. 1310–1317.

Vecchia, A. V. (1985), "Periodic autoregressive moving-average (PARMA) modeling with applications to water resources," *Water Resour. Bull.*, vol. 21, pp. 721–730.

Vetterli, M. and Kovačević (1995), *Wavelets and Subband Coding*, Upper Saddle River, NJ: Prentice Hall.

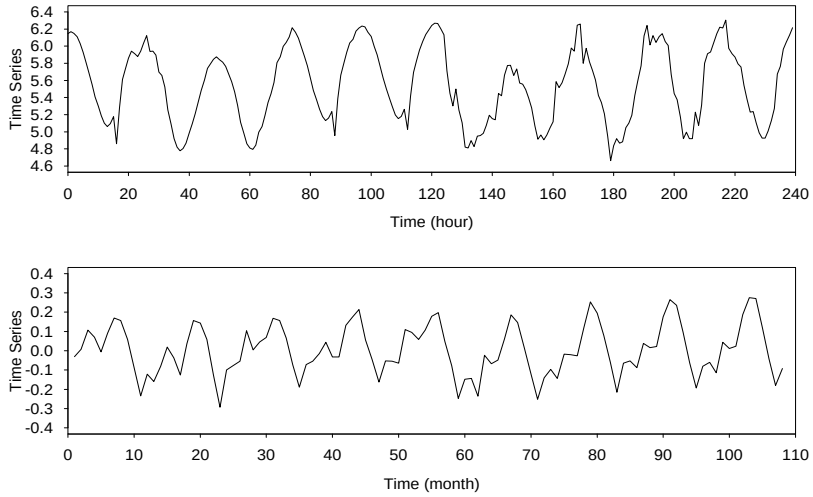

Figure 1: Examples of seasonal time series. Top, a 10-day time series of hourly mean global irradiance (in logarithmic scale with adjusted mean). Bottom, a 9-year time series of monthly passager totals in international air travel (in logarithmic scale with trend removed).

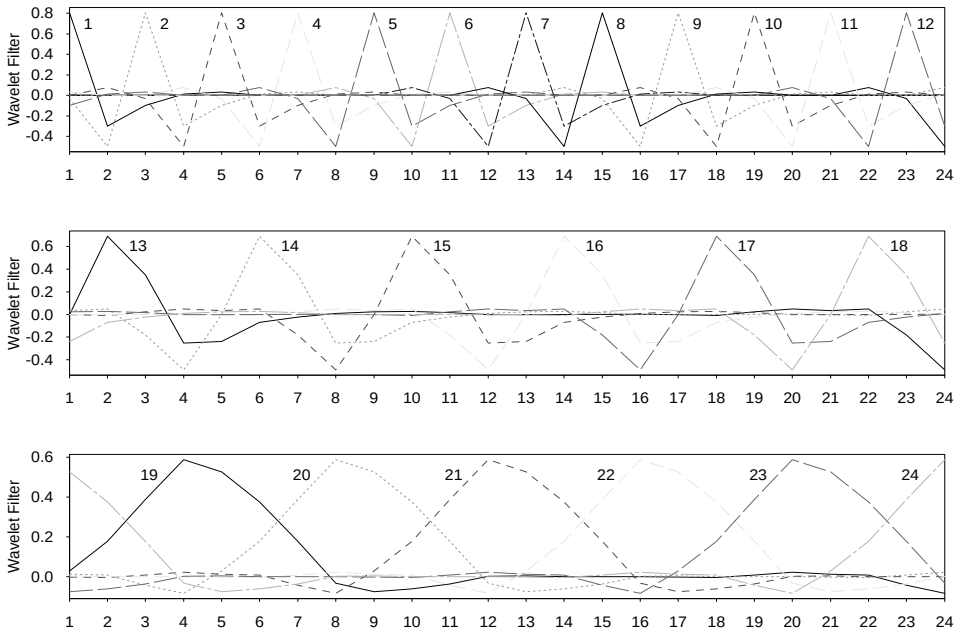

Figure 2: A wavelet filter bank of size $p = 24$, corresponding to the eighth-order ($N = 8$) least-asymmetric Daubechies wavelets. The 3 subbanks comprise: top, $2^{-1}p = 12$ highpass filters of scale 1; middle, $2^{-2}p = 6$ highpass filters of scale 2; bottom, remaining $2^{-2}p = 6$ lowpass filters of scale 2.

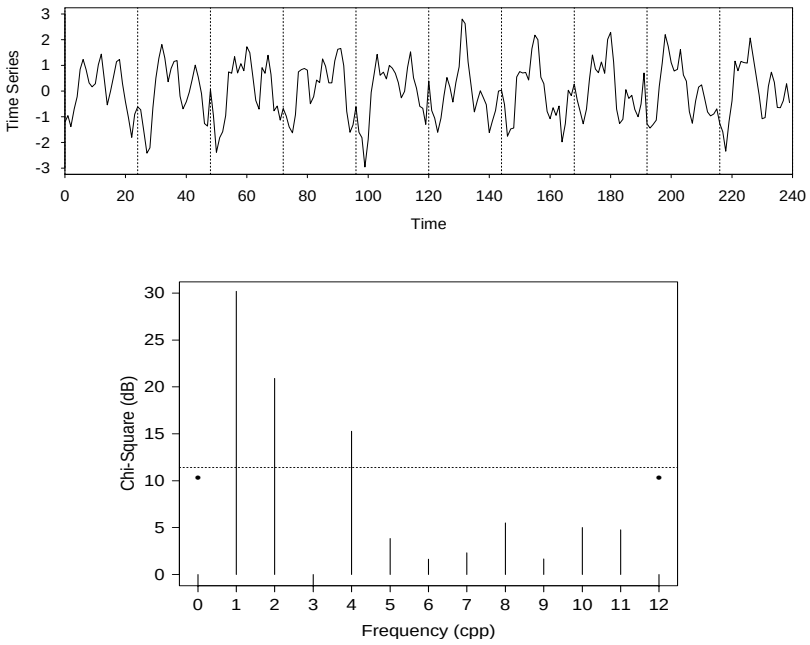

Figure 3: Top, a simulated seasonal time series with periodicity $p = 24$. Bottom, coherence plot (chi-square statistics in decibels) computed from the DFT of 50 seasonal vectors of the time series. The dotted line represents the 99th percentile of $\chi^2(2)$ and the dots are the 99th percentile of $\chi^2(1)$, both under the white noise assumption. (cpp = cycles per period.)

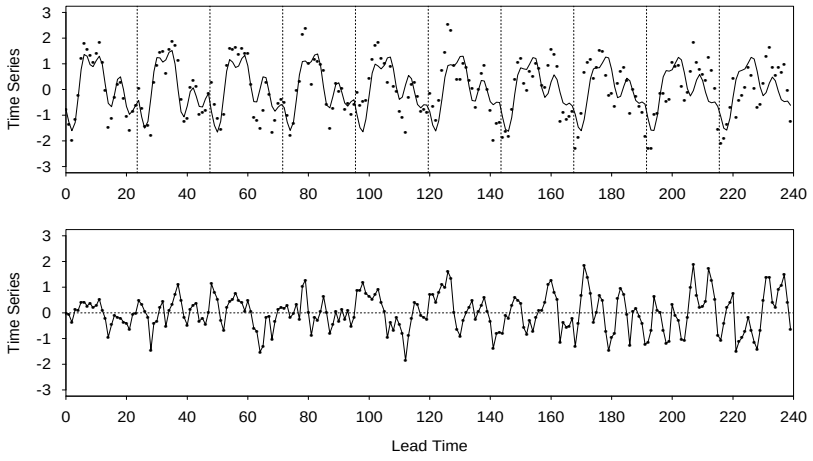

Figure 4: Forecasts of the time series shown in Figure 3 by the proposed method with Fourier filters. Top, forecasts (solid curve) and actual values (dots). Bottom, forecasting errors (RMSE = 0.702 and $R^2 = 56.4\%$).

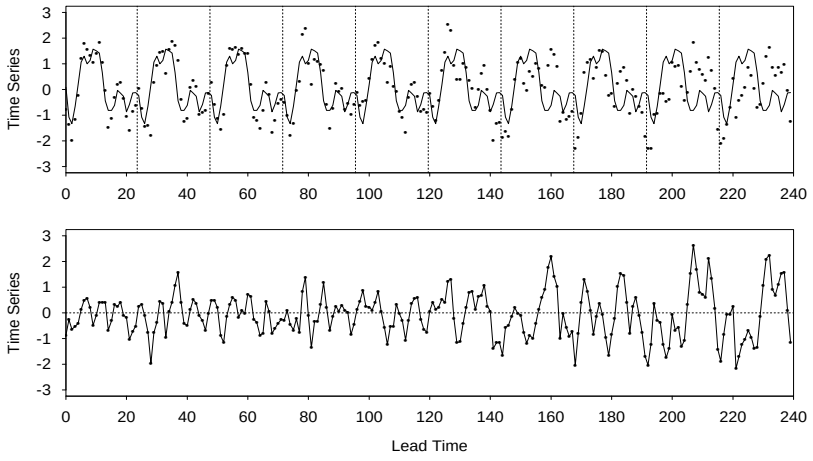

Figure 5: Forecasts of the time series shown in Figure 3 by SARIMA $(0,0,1) \times (0,1,1)_p$. Top, forecasts and actual values. Bottom, forecasting errors (RMSE = 0.863 and $R^2 = 34.1\%$). Notice the rapid growth of forecasting error as the horizon increases.

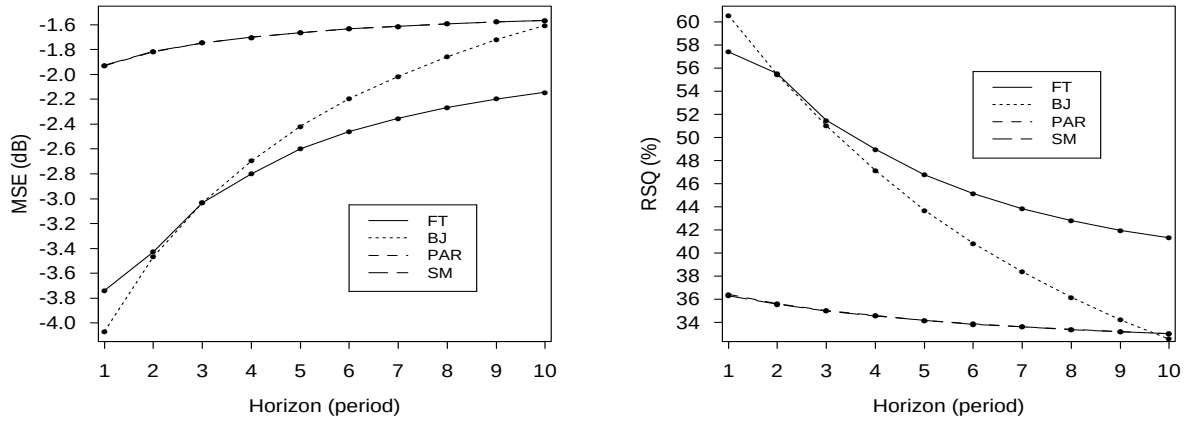

Figure 6: Averaged forecasting performance for the time series shown in Figure 3. Left, MSE in decibels. Right, R^2 in percentage. Forecasting methods: FT, the proposed method with Fourier filters; BJ, the SARIMA method; PAR, the PAR(1) method; SM, the seasonal mean. PAR and SM are indistinguishable.

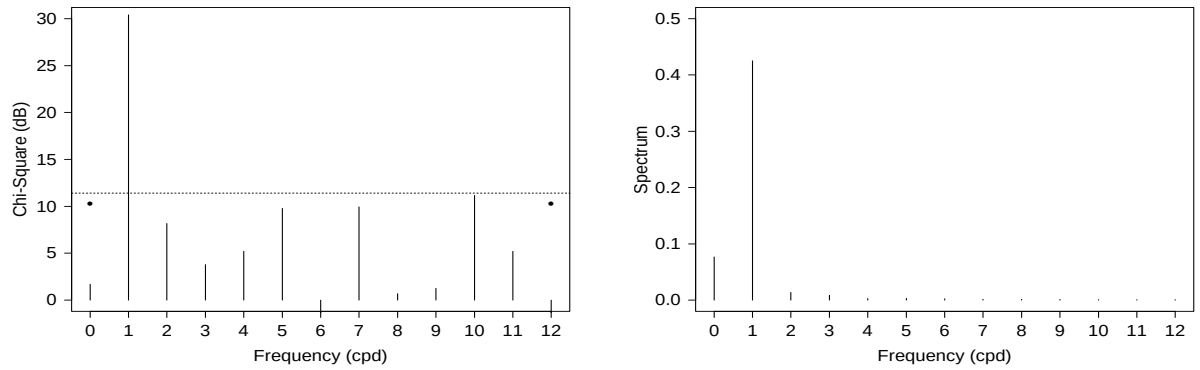

Figure 7: Fourier analysis of the hourly global irradiance, based on 14-day training data. Left, Fourier coherence plot. Right, Fourier power spectrum. (cpd = cycles per day.)

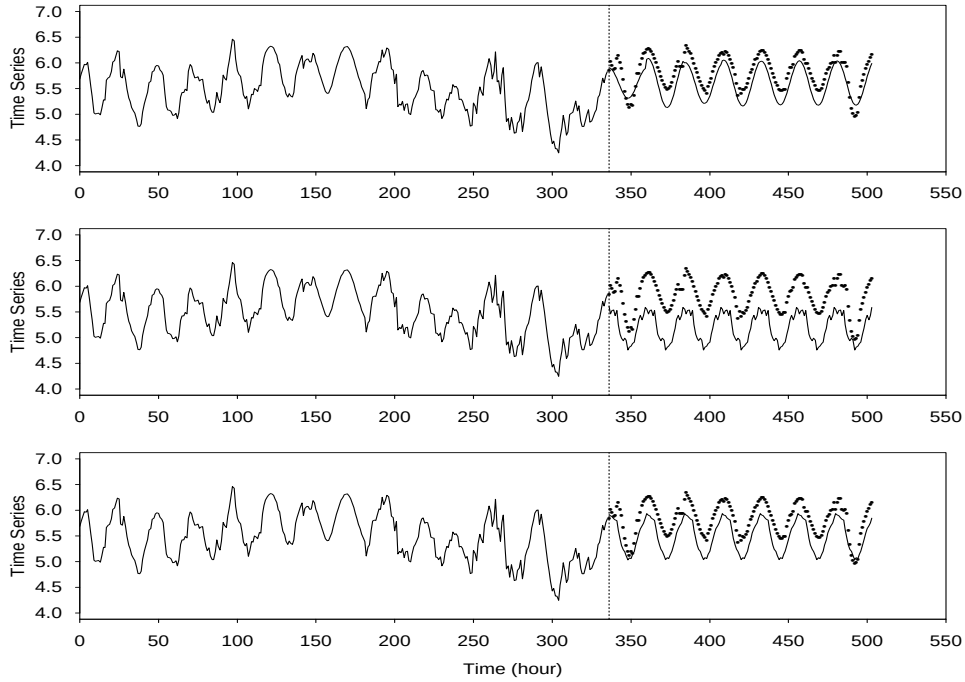

Figure 8: An example of abnormal patterns in training data of the hourly global irradiance. Top, forecasts by the proposed method with Fourier filters (RMSE = 0.260, $R^2 = 32.4\%$). Middle, forecasts by SARIMA $(0,0,1) \times (0,1,1)_p$ (RMSE = 0.644, $R^2 = -314\%$). Bottom, forecasts by PAR(1) (RMSE = 0.356, $R^2 = -26.9\%$). The vertical line separates the training data from the forecasts.

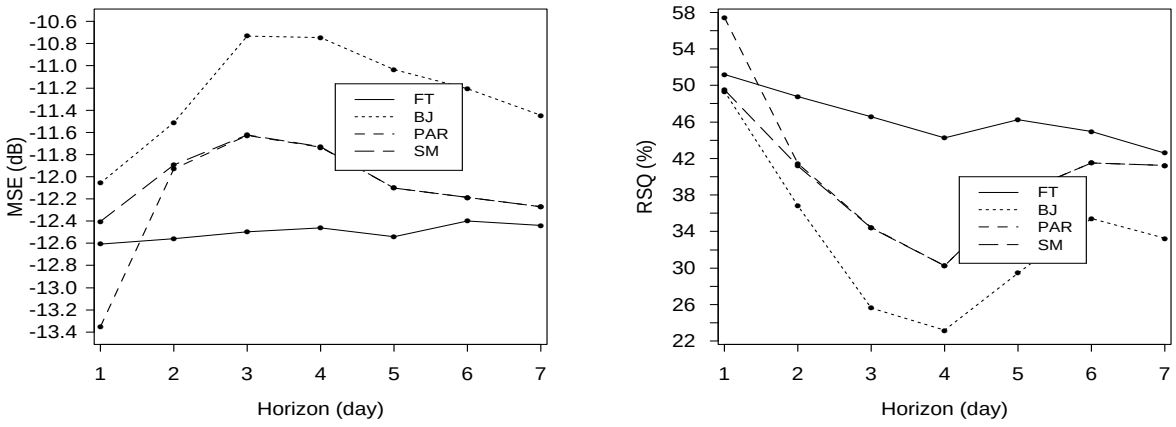

Figure 9: Averaged forecasting performance for the global irradiance with 14-day training. Left, horizon-specific MSE in decibels. Right, horizon-specific R^2 in percentage.

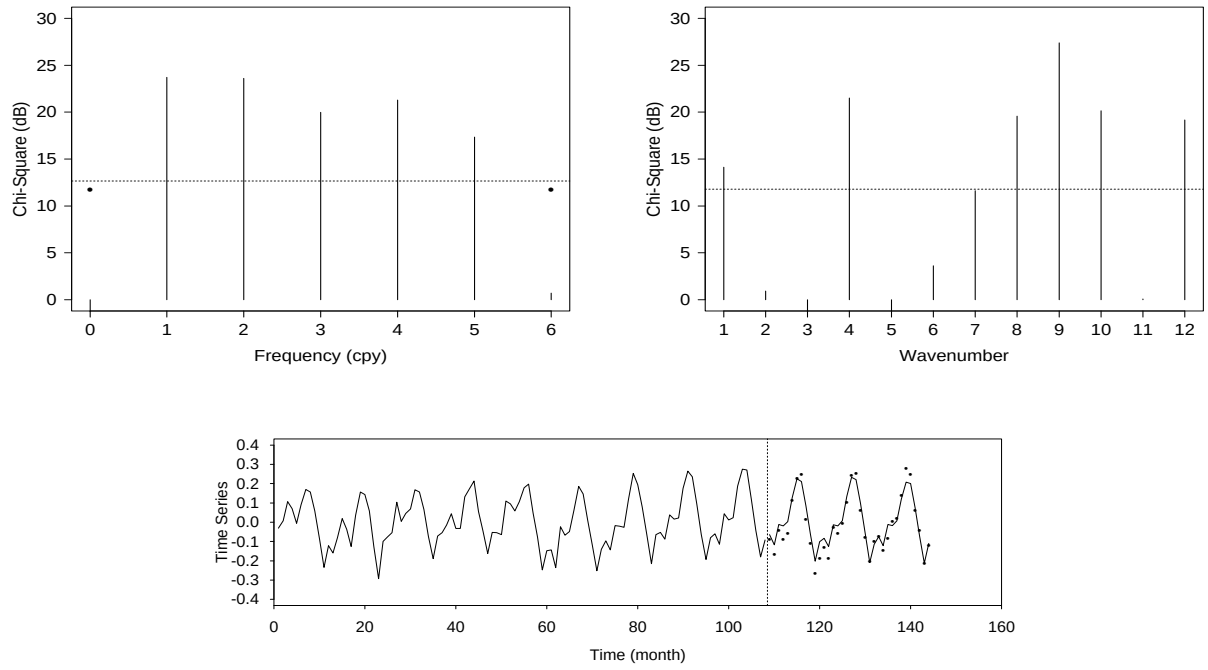

Figure 10: Top, coherence plots of the airline data ($\alpha = 0.0001$): left, Fourier analysis (cpy = cycles per year); right, wavelet analysis based on the four-order least-asymmetric Daubechies wavelets. Bottom, forecasts of the airline data by the proposed method with the wavelet filters (RMSE = 0.0409, $R^2 = 92.5\%$).

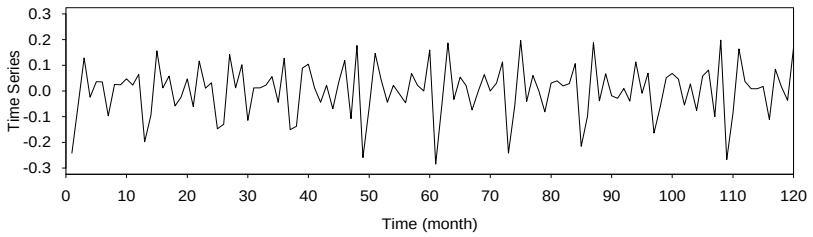

Figure 11: Monthly growth rates of retail sales in the Netherlands (October 1985 – September 1995).