THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 61, No. 1, January 1982
Printed in U.S.A.

A Short-Term Forecasting Algorithm for Trunk Demand Servicing

By C. R. SZELAG

(Manuscript received December 31, 1980)

Trunk servicing is the continual process of collecting trunk group traffic measurements, monitoring network service, and augmenting the network when necessary. This study addresses the possibility of using a short-term forecast to determine the adequacy of trunk quantities planned for the imminent busy season. When seasonal patterns of demand exist, it may be possible to use observed, pre-busy-season traffic levels to predict accurately that busy-season demand will exceed the planned trunk group capacity and to determine appropriate corrective action. Toward this end, we develop a seasonal load forecasting algorithm based on Kalman filter estimation techniques and analyze the effectiveness of this approach using Bell operating company data. For trunk groups exhibiting seasonal demand, the short-term (1½ months ahead), seasonal forecast error is 50 percent less than that of the sequential projection algorithm (SPA), which linearly trends the yearly busy-season loads. Much of this improvement is attributed to the ability of the seasonal algorithm to utilize recent observations; the one-year ahead seasonal forecast error is only 20 percent less than that of SPA. We conclude that the greater generality and simplicity of SPA makes that algorithm the appropriate choice for the annual busy-season trunk forecast used in medium-range network planning. However, the seasonal algorithm demonstrated the ability to use recent data to respond quickly and accurately to various situations that result in inaccurate SPA forecasts. For this reason, the short-term forecasting algorithm developed herein is a potentially valuable tool for network administration.

I. INTRODUCTION

1.1 Motivation

Demand servicing is the process responsible for detecting and correcting overload conditions in the trunk network. Such conditions inevitably occur when unanticipated traffic levels exceed the planned capacity, which must be maintained at a reasonably low level to provide good service at low cost.

The planned capacity for each trunk group is determined primarily by the annual forecast of busy-season trunk requirements. Trunk servicing, which includes demand servicing, is the continual process of collecting trunk group traffic data, monitoring network service, and augmenting the network when needed. This paper addresses the possibility of supplementing the annual forecast and weekly monitoring process with a short-term forecast of imminent busy-season requirements. Specifically, when seasonal patterns of demand exist, it may be possible to use observed, pre-busy-season traffic levels to predict accurately that busy-season demand will exceed the planned trunk group capacity. Thus, service problems may be predicted and possibly avoided by "anticipative" demand servicing action.

When the need for demand servicing arises, the trunk servicer must decide on the locations and magnitudes of trunk group augments required to restore service. If current traffic levels already exceed those forecast, the servicer would like to know whether the peak load level has already been reached, or whether even higher levels are imminent. For each trunk group that must be augmented, the servicer should know the minimum amount of additional capacity required to both relieve the existing problem and to provide adequate service through the remainder of the busy season.

1.2 Short-term trunk forecast

Both of the trunk servicing functions described above, namely, the anticipation of imminent service problems and the determination of appropriate demand servicing augments, could be performed with the use of a short-term trunk forecasting system. Such a system would recognize within-the-year demand patterns and use this information to make accurate, short-term predictions of busy-season load. As such, the short-term forecast would serve as a "back-up" to the yearly busy-season forecast, recommending remedial action in those cases where the latter is significantly in error.

The purpose of this study is to investigate a short-term forecast to provide the trunk servicer with accurate, useful information concerning near-term traffic levels. With such a tool, service problems could be avoided by anticipative demand servicing and useful reserve capacity could be identified. This would allow the servicer to make more efficient use of existing facilities and equipment, thus, reducing the amount of reserve capacity required to maintain good service.

1.3 Overview

Section II is a discussion of the general requirements of a short-term trunk forecasting system. Motivated by these requirements, we consider in Section III the class of linear dynamic models and show how these models can be used to represent trunk group load histories exhibiting seasonal variation. In Section IV, we discuss a recursive estimation procedure, known as the Kalman filter, that is appropriate for this class of models. Using trunk group data obtained from a Bell operating company (BOC), we test the performance of a specific forecasting algorithm in Section V, and compare its performance with the year-to-year forecast produced by the recently developed sequential projection algorithm (SPA).[1,2] Section VI summarizes our findings.

II. SYSTEM REQUIREMENTS

The selection of an appropriate class of time-series models and forecasting procedures for consideration in this study depends heavily on the intended mode of operation and operating environment. The general requirements of a short-term trunk forecasting system are discussed in this section. These requirements will motivate the class of time-series forecasting algorithms considered in Section III.

2.1 Time series model

Underlying any time-series forecasting procedure is a mathematical model describing the structure of the series being forecast. For short-term trunk forecasting, the time-series model used must be sufficiently flexible to model a wide range of trunk group growth patterns accurately and to track changes in these growth patterns closely. This requirement is shown in Fig. 1. Figure 1a shows the load history of a BOC only-route group exhibiting a highly regular seasonal pattern modulating a nearly constant, linear trend. The behavior of such demand can be accurately predicted by an reasonably appropriate procedure. In particular, the annual 1-year SPA forecasts of busy-season demand are quite accurate.

In contrast, the trunk group load history shown in Fig. 1b requires a more sophisticated treatment. The distinctly nonconstant trend leads to gross errors in the SPA forecasts of busy-season loads, since the SPA algorithm design assumes approximately linear growth or decline.

Although the growth pattern shown in this figure is clearly nonconstant, it is not unpredictable. The short-term trunk forecasting algorithm should have sufficient built-in flexibility to accommodate such behavior.

2.2 Data requirements

To provide good service at low cost, the Bell System network is constantly evolving. This evolution accounts for the introduction of new and improved technology with the accompanying changes in

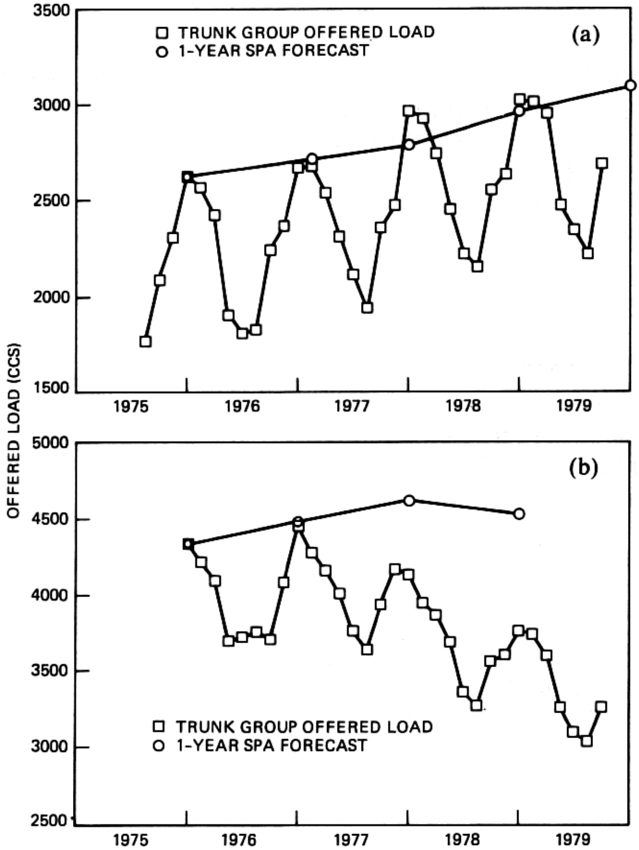

Fig. 1—(a) Trunk group AA001444—annual busy-season forecast. (b) Trunk group AA024225—annual busy-season forecast.

network configuration, and for change and growth in customer demand patterns. As the network evolves, new trunk groups, connecting new switching systems in a more economical network configuration, come into service and serve traffic previously carried by existing groups, which may be phased out of service. Also, in the past, trunk group histories were not maintained and data collection schedules were less comprehensive. For these reasons, long, uninterrupted trunk group load histories are, and will continue to be, atypical.

If it is to be useful, the short-term trunk forecasting system must be capable of operating in this type of environment. It should be capable of producing accurate forecasts based on relatively small amounts of historical data (e.g., 2 years). Also, it should be able to process and respond to information concerning semideterministic events (e.g., main station transfers and traffic reroutes) that affect the trunk group load pattern.

2.3 Computational requirements

In most BOCs, the trunk servicer relies on the Trunk Servicing System (TSS) to process recent trunk group traffic measurements and to obtain estimates of current load levels, traffic characteristics, and trunk requirements.[3] On a weekly basis, the system must process measurements and produce reports on several thousand trunk groups. Therefore, when considering candidate short-term forecasting algorithms for implementation in such a system, computational efficiency is of great importance. Equally important is the need for mechanized procedures requiring minimal intervention by the servicer, who typically must administer several hundred, and possibly thousands, of groups. Thus, computational efficiency and automation requirements rule out those procedures commonly used to perform detailed statistical analyses of individual time series. (See for example, the methods described in Ref. 4.)

2.4 Summary

In summary, the short-term trunk forecasting algorithms considered in this study should be able to track the trunk group load series accurately and adapt to dynamic changes in the demand pattern. In addition, they should perform adequately after a minimal initialization period and be computationally efficient.

In the next section, we describe a class of algorithms satisfying these requirements.

III. LINEAR DYNAMIC MODELS OF SEASONAL TIME SERIES

In this section, we consider the representation of trunk group load histories exhibiting seasonal variation by a linear, dynamic time-series model. The motivation for considering such a model for the application considered in this paper is discussed in Ref. 5. To summarize, this formulation is sufficiently general to describe a large number of time series of practical interest and is compatible with certain computationally efficient estimation techniques that make optimal use of limited amounts of data. The estimation and forecasting of these models will be considered in Section IV.

3.1 Mathematical model

A discrete time series is a sequence of observations of some quantity of interest. We think of such a series as being a realization of some stochastic process $\{\mathbf{y}_t\}$, which serves as a mathematical model explaining the observations, and which allows us to make inferences concerning future values of the series.

In the linear dynamic model, the behavior of the series is determined by an n-dimensional state-vector process $\{\mathbf{X}_t\}$ and the following two equations that describe the time evolution of $\{\mathbf{X}_t\}$ and the relation of $\mathbf{X}_t$ to the observation $\mathbf{y}_t$.

The first equation, called the system equation, is

$$\mathbf{X}_{t+1} = \boldsymbol{\phi}_t \mathbf{X}_t + \mathbf{W}_t, \tag{1}$$

where $\boldsymbol{\phi}_t$ if an $n \times n$ "transition" matrix and $\{\mathbf{W}_t\}$ is an n-dimensional, zero-mean white noise process with

$$E(\mathbf{W}_t \mathbf{W}_s') = \begin{cases} \mathbf{0} & \text{if} \quad s \neq t \\ \mathbf{Q}_t & \text{if} \quad s = t. \end{cases} \tag{2}$$

More generally, we can consider the system equation

$$\mathbf{X}_{t+1} = \boldsymbol{\phi}_t \mathbf{X}_t + \mathbf{U}_t + \mathbf{W}_t, \tag{3}$$

where $\mathbf{U}_t$ is an n-dimensional deterministic or stochastic input to the system at time t.

The matrix $\boldsymbol{\phi}_t$ in eqs. (1) and (3) describes the deterministic movement of the state variables comprising $\mathbf{X}_t$. The white noise process $\{\mathbf{W}_t\}$ explicitly allows for random variations in these state variables and, therefore, significantly enhances the flexibility of this formulation.

The second equation, called the observation equation, is

$$\mathbf{y}_t = \mathbf{H}_t \mathbf{X}_t + \boldsymbol{\epsilon}_t, \tag{4}$$

where $\mathbf{H}_t$ is a $k \times n$ "observation" matrix and $\boldsymbol{\epsilon}_t$ is a k-dimensional, zero-mean, white-noise vector sequence, independent of $\{\mathbf{X}_t\}$, with

$$E[\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_s'\} = \begin{cases} \mathbf{0} & \text{if} \quad s \neq t \\ \mathbf{R}_t & \text{if} \quad s = t. \end{cases} \tag{5}$$

Thus, the observation $\mathbf{y}_t$ is a linear function of the state variables corrupted by an additive disturbance $\boldsymbol{\epsilon}_t$.

Since we will only consider univariate time series $\{y_t\}$ in this paper, we will assume henceforth that $k = 1$.

3.2 Examples

Next, we present a few simple, but relevant, examples to demonstrate that this formulation describes many kinds of time series. Additional examples can be found in Ref. 6.

3.2.1 Linear growth model

In many applications in which a quantity is to be forecast over a relatively short time span, it is reasonable to assume that the quantity is growing approximately linearly with time. In particular, the SPA[2] design assumes that busy-season trunk group load varies from year to year along a nearly linear trend.

Linear growth can be described by the two-dimensional linear model with the following parameters:

$$\phi_t \equiv \phi = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}; \qquad \mathbf{H}_t \equiv \mathbf{H} = [1, 0];$$

$$\mathbf{X}_t = \begin{bmatrix} x_t^{(1)} \\ x_t^{(2)} \end{bmatrix}. \tag{6}$$

In this model, $x_t^{(1)}$ represents the level of the series at time t, and $x_t^{(2)}$ the instantaneous rate of growth. The inclusion of the white noise process described by $\{\mathbf{Q}_t\}$ allows the trajectory to deviate randomly from a straight line.

3.2.2 Seasonal models

A time series $\{y_t\}$ is said to exhibit seasonality if observations separated in time by some fixed interval (usually one year) exhibit similar behavior. The simplest kind of seasonality is periodicity; that is $y_t \approx y_{t+nL}$ for some positive integer L and every integer n. Two linear models of periodic behavior are described below.

A useful model of seasonal variation is available via the trigonometric representation for periodic sequences.[7] Let $\{y_t\}$ be periodic, with period L. Then $\{y_t\}$ can be represented as follows:

$$y_t = \alpha_1 + \sum_{j=1}^{\left[\frac{L-1}{2}\right]} [\alpha_{2j}\cos(j\omega t) + \alpha_{2j+1}\sin(j\omega t)] + \alpha_L(-1)^t, \tag{7}$$

where $\omega = 2\pi/L$ and the last term is omitted if L is odd. Hereafter, we will assume that L is even. The coefficients $\{\alpha_k\}$ in the trigonometric expansion are determined by

$$\alpha_1 = \frac{1}{L}\sum_{i=1}^{L} y_i$$

$$\alpha_{2j} = \frac{2}{L}\sum_{i=1}^{L} y_i\cos(ij\omega)$$

$$\alpha_{2j+1} = \frac{2}{L}\sum_{i=1}^{L} y_i\sin(ij\omega)$$

$$\alpha_L = \frac{1}{L}\sum_{i=1}^{L} y_i(-1)^i. \tag{8}$$

Note that the first coefficient α_1 represents the average level of the series $\{y_t\}$. It can be omitted from the expansion (7) if we are representing a series having zero mean.

The existence of the representation (7) follows from the fact that the set of time functions

$$\mathbf{F}_0 = [1, \cos\omega t, \sin\omega t, \cdots, (-1)^t]$$

forms an orthogonal basis for the linear space of all real, periodic sequences with period L. Also note that for every integer τ, the set of τ-translates of $\mathbf{F}_0$, defined by

$$\mathbf{F}_\tau = [1, \cos\omega(t-\tau), \sin\omega(t-\tau), \cdots, (-1)^{t-\tau}]$$

also has this property. Therefore, for each τ there exists a representation of $\{y_t\}$ as a linear combination of the elements of $\mathbf{F}_\tau$;

$$y_t = \alpha_1^{(\tau)} + \sum_j \{\alpha_{2j}^{(\tau)} \cos[j\omega(t-\tau)] + \alpha_{2j+1}^{(\tau)} \sin[j\omega(t-\tau)]\} + \alpha_L^{(\tau)}(-1)^{t-\tau} \tag{9}$$

with $\alpha_k^{(0)} = \alpha_k$ as in eq. (7).

The expansion of the series $\{y_t\}$ relative to the set of functions $\mathbf{F}_\tau$ can be viewed as a representation in which $t = \tau$ serves as the new time origin. In the linear dynamic model described below, at each new time epoch $t = \tau + 1$ we will "rearrange" the representation (9) in terms of the functions $\mathbf{F}_{\tau+1}$ via a simple, linear transformation of the coefficients $\{\alpha_k^{(\tau)}\}$.

$$\text{Let } y_t = \alpha_1^{(\tau)} + \sum_j [\alpha_{2j}^{(\tau)} \cos j\omega(t-\tau) + \alpha_{2j+1}^{(\tau)} \sin j\omega(t-\tau)] + \alpha_L^{(\tau)}(-1)^{t-\tau} \tag{10}$$

$$= \alpha_1^{(\tau+1)} + \sum_j \{\alpha_{2j}^{(\tau+1)} \cos j\omega[t-(\tau+1)] + \alpha_{2j+1}^{(\tau+1)} \sin j\omega[t-(\tau+1)]\} + \alpha_L^{(\tau+1)}(-1)^{t-(\tau+1)}. \tag{11}$$

By expanding each term in eq. (11) and equating coefficients of similar terms with the coefficients in eq. (10) yields

$$\alpha_1^{(\tau)} = \alpha_1^{(\tau+1)}$$

$$\begin{bmatrix} \alpha_{2j}^{(\tau)} \\ \alpha_{2j+1}^{(\tau)} \end{bmatrix} = \phi_j^{-1} \begin{bmatrix} \alpha_{2j}^{(\tau+1)} \\ \alpha_{2j}^{(\tau+1)} \end{bmatrix}$$

$$\alpha_L^{(\tau)} = (-1)\alpha_L^{(\tau+1)}, \tag{12}$$

where

$$\phi_j^{-1} = \begin{bmatrix} \cos j\omega & -\sin j\omega \\ \sin j\omega & \cos j\omega \end{bmatrix}. \tag{13}$$

$$\text{Let } \mathbf{X}_\tau = [\alpha_1^{(\tau)}, \cdots, \alpha_L^{(\tau)}]'. \tag{14}$$

Then

$$\mathbf{X}_\tau = \phi^{-1}\mathbf{X}_{\tau+1}, \tag{15}$$

where ϕ^{-1} is the block diagonal matrix defined by

$$\phi^{-1} = \begin{bmatrix} 1 & & & & & \mathbf{0} \\ & \phi_1^{-1} & & & & \\ & & \phi_2^{-1} & & & \\ & & & \cdot & & \\ & & & & \cdot & \\ & \mathbf{0} & & & & \cdot \\ & & & & & -1 \end{bmatrix} \tag{16}$$

Equivalently,

$$\mathbf{X}_{\tau+1} = \phi \mathbf{X}_\tau. \tag{17}$$

To summarize, the trigonometric expansion (7) of $\{y_t\}$ can be represented by a linear dynamic system with state vector $\mathbf{X}_\tau = [\alpha_1^{(\tau)}, \cdots, \alpha_L^{(\tau)}]'$ comprising the coefficients in the expansion relative to the current time τ. The transition matrix ϕ, defined by eq. (16), provides the mechanism by which the coefficients at time $\tau + 1$ are obtained from those at time τ. Finally, at time τ, we observe

$$\begin{aligned} y_\tau &= \alpha_1^{(\tau)} + \sum_j \alpha_{2j}^{(\tau)} + \alpha_L^{(\tau)} + \epsilon_\tau \\ &= \mathbf{H}\mathbf{X}_\tau + \epsilon_\tau, \end{aligned} \tag{18}$$

where the L-dimensional vector $\mathbf{H}$ is given by

$$\mathbf{H} = \mathbf{H}_\tau = [1101 \cdots 01]. \tag{19}$$

By including a nonzero white-noise input in the dynamics (16), i.e.,

$$\mathbf{X}_{\tau+1} = \phi \mathbf{X}_\tau + \mathbf{W}_\tau, \tag{20}$$

we can allow for random variations in the trigonometric coefficients, and therefore, in the seasonal pattern.

In addition to the trigonometric model described above, another commonly used seasonal representation, the seasonal index model (Ref. 6, page 217) was considered. However, empirical performance results, analogous to those to be presented in Section V, indicated that the trigonometric model gave somewhat better results. For the sake of brevity, this model will not be discussed further.

3.3 Seasonal trunk group model

In Section 3.2, we showed how the general linear model eqs. (1) to (4) can be used to represent either linear-growth or periodic time series. We now demonstrate that these models can be superimposed to form a model capable of representing a wide variety of trunk group load series.

Consider again the trunk group load histories shown in Fig. 1. These series may be characterized by an underlying linear trend with a seasonal pattern superimposed. (This observation is consistent with the assumptions used in the design of SPA.[2] See Section 3.2.1.) The former component can be represented by the linear growth model described in Section 3.2.1; the latter, by the periodic model described in Section 3.2.2. If we assume that these two effects combine in an additive fashion, we can represent the observed behavior by the linear model with state vector

$$\mathbf{X}_t = \begin{bmatrix} \mathbf{X}_t^{(l)} \\ \cdots \\ \mathbf{X}_t^{(s)} \end{bmatrix}, \tag{21}$$

where $\mathbf{X}_t^{(l)}$ is the state vector defined for the linear growth model (6) and $\mathbf{X}_t^{(s)}$ is that of the periodic model (14). (In the seasonal component $\mathbf{X}_t^{(s)}$, the "level" term $\alpha_1^{(t)}$ included in eq. (7) is omitted, since a similar term appears in the linear growth component $\mathbf{X}_t^{(l)}$.)

The dynamics of this system are described by the transition matrix

$$\boldsymbol{\phi} = \begin{bmatrix} \boldsymbol{\phi}^{(l)} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\phi}^{(s)} \end{bmatrix}, \tag{22}$$

where $\boldsymbol{\phi}^{(l)}$ and $\boldsymbol{\phi}^{(s)}$ are the transition matrices of the linear and seasonal models, respectively.

Similarly, the observation matrix, $\mathbf{H}$, is given by

$$\mathbf{H} = [\mathbf{H}^{(l)} \vdots \mathbf{H}^{(s)}], \tag{23}$$

which decomposes the observation into the sum of a trend component $\mathbf{H}^{(l)}\mathbf{X}_t^{(l)}$ and a seasonal component, $\mathbf{H}^{(s)}\mathbf{X}_t^{(s)}$.

Finally, random variation in the components of $\mathbf{X}_t$ is modeled as a white-noise input, described by the sequence $\{\mathbf{Q}_t\}$. In particular, if the disturbances to the trend and seasonal components are uncorrelated, then $\mathbf{Q}_t$ may be decomposed as

$$\mathbf{Q}_t = \begin{bmatrix} \mathbf{Q}_t^{(l)} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_t^{(s)} \end{bmatrix}.$$

3.4 Summary

In this section, we discussed the class of linear dynamic time-series models and showed how such models can be used to represent time series exhibiting both linear growth and seasonal variation. In particular, a specific model was proposed in Section 3.3 for the representation of trunk group load histories exhibiting seasonality.

In the next section, we consider the problem of estimating the parameters of such a model from the observed time series and the use of these estimates in forecasting.

IV. PARAMETER ESTIMATION AND PREDICTION

Complementing the class of linear dynamic time-series models is a sequential parameter estimation algorithm known as the Kalman filter.[8,9] With this technique, we estimate the model parameters comprising the state vector $\mathbf{X}_t$ from the observations $y_1, \cdots, y_t$ and use this estimate to make inferences concerning future values of the series (prediction).

The general procedures by which this is accomplished are discussed in this section. We begin by describing a recursive algorithm for sequentially updating the state vector estimate as new data becomes available. The recursion is initiated by a weighted least-squares estimation procedure derived in Section 4.2. The application of these procedures to the seasonal trunk group model derived in Section III yields a forecasting algorithm capable of accurately tracking and predicting the values of the seasonal time series considered in this study.

The performance of this algorithm on actual trunk group data will be examined in Section V.

4.1 The Kalman filter

Consider the linear dynamic model described by the equations

$$\mathbf{X}_{t+1} = \phi_t \mathbf{X}_t + \mathbf{W}_t \tag{24}$$

and

$$y_t = \mathbf{H}_t \mathbf{X}_t + \epsilon_t. \tag{25}$$

Assume that

$$\begin{gathered} E(\mathbf{W}_t) = \mathbf{0}, \qquad E(\epsilon_t) = 0, \\ E[\mathbf{W}_t \mathbf{W}_s'] = \begin{cases} \mathbf{0} & \text{if} \quad s \neq t \\ \mathbf{Q}_t & \text{if} \quad s = t, \end{cases} \\ E[\mathbf{W}_t \epsilon_s] = \mathbf{0}, \end{gathered} \tag{26}$$

and

$$E[\epsilon_t \epsilon_s] = \begin{cases} 0 & \text{if} \quad s \neq t \\ R_t & \text{if} \quad s = t. \end{cases}$$

Suppose that at time t, prior to observing y_t, we have available an initial estimate of the state vector $\mathbf{X}_t$. Calling this prior estimate $\hat{\mathbf{X}}_{t,t-1}$,* let us also postulate that the estimation error $\tilde{\mathbf{X}}_t = \hat{\mathbf{X}}_{t,t-1} - \mathbf{X}_t$ has zero mean and is uncorrelated with ϵ_t, the observation error in y_t. Let us also assume that the error covariance matrix of the estimate,

* In general, the notation $\hat{\mathbf{X}}_{t,s}$ will be used to denote an estimate of $\mathbf{X}_t$ based on information available at time s.

$$\mathbf{P}_{t,t-1} = E(\tilde{\mathbf{X}}_t \tilde{\mathbf{X}}_t'), \tag{27}$$

is known.

When the current observation y_t becomes available, we want to combine the prior estimate with the new observation in an optimal manner. Specifically, we seek unbiased linear estimates $\hat{\mathbf{X}}_{t+k,t}$, $k \geq 0$, of the current and future states, which minimize the quantities trace $(\mathbf{P}_{t+k,t})$, where $\mathbf{P}_{t+k,t}$ is the error covariance matrix of $\hat{\mathbf{X}}_{t+k,t}$. Thus, we are seeking minimum variance, unbiased linear state-vector estimates.

The following procedure, which yields such estimates in a recursive manner, is known as the Kalman filter.[9]

4.1.1 Filtering

The problem of determining the optimal estimate $\hat{\mathbf{X}}_{t,t}$ from data available at time t is known as the filtering problem. Its solution is given by

$$\hat{\mathbf{X}}_{t,t} = \hat{\mathbf{X}}_{t,t-1} + \mathbf{K}_t[y_t - \mathbf{H}_t\hat{\mathbf{X}}_{t,t-1}], \tag{28}$$

where the optimal gain $\mathbf{K}_t$, is determined as

$$\mathbf{K}_t = \mathbf{P}_{t,t-1}\mathbf{H}_t'[\mathbf{H}_t\mathbf{P}_{t,t-1}\mathbf{H}_t' + R_t]^{-1}. \tag{29}$$

The error covariance matrix of this estimate is given by

$$\mathbf{P}_{t,t} = (\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\mathbf{P}_{t,t-1}(\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)' + \mathbf{K}_tR_t\mathbf{K}_t' \tag{30}$$

$$= [\mathbf{I} - \mathbf{K}_t\mathbf{H}_t]\mathbf{P}_{t,t-1} \tag{31}$$

The more complex expression (30) is included since this relationship is valid for an arbitrary gain matrix $\mathbf{K}_t$, whereas (31) is valid only when $\mathbf{K}_t$ is the optimal gain (29).

4.1.2 Prediction

Having obtained the optimal linear estimate $\hat{\mathbf{X}}_{t,t}$ of $\mathbf{X}_t$ from the data available at time t, we now want to predict future values of the state vector and the series. These are obtained by extrapolation, using the linear dynamics relationship (24).

That is, the optimal linear estimates of $\mathbf{X}_{t+k}$ and y_{t+k}, based on data available at time t, are given, for $k > 0$, by

$$\hat{\mathbf{X}}_{t+k,t} = \left[\prod_{i=1}^{k} \phi_{t+i-1}\right]\hat{\mathbf{X}}_{t,t} \tag{32}$$

and

$$\hat{y}_{t+k,t} = \mathbf{H}_{t+k}\hat{\mathbf{X}}_{t+k,t}. \tag{33}$$

In particular, the optimal 1-step predictors are

$$\hat{\mathbf{X}}_{t+1,t} = \phi_t\hat{\mathbf{X}}_{t,t}, \tag{34}$$

and

$$\hat{y}_{t+1,t} = \mathbf{H}_{t+1}\hat{\mathbf{X}}_{t+1,t}.$$

The error covariance matrix of $\hat{\mathbf{X}}_{t+1,t}$ is

$$\mathbf{P}_{t+1,t} = \phi_t \mathbf{P}_{t,t} \phi'_t + \mathbf{Q}_t; \tag{35}$$

the one-step forecast error variance is

$$\text{var}(\hat{y}_{t+1,t}) = \mathbf{H}_{t+1}\mathbf{P}_{t+1,t}\mathbf{H}'_{t+1}. \tag{36}$$

4.1.3 Sequential estimation

We now show how the results of Sections 4.1.1 and 4.1.2 can be used in an efficient procedure for processing the observations $\{y_t\}$ to obtain estimates of the parameters of the underlying linear model.

First note that the 1-step prediction $\hat{\mathbf{X}}_{t+1,t}$ given by eq. (34) satisfies the requirements of the initial estimate of the state vector at time $t + 1$. Thus, we can use the filtering procedure described in Section 4.1.1 to process the next observation, y_{t+1}, when it becomes available.

In general, starting with an initial state estimate $\mathbf{X}_{t_0,t_0-1}$ at some time t_0, we can alternately apply the procedures described in Section 4.1.1 (filtering) and Section 4.1.2 (prediction) to process subsequent observations y_{t_0}, y_{t_0+1}, $\cdots$ in an efficient, recursive manner. The procedure is summarized below.

Inputs: Initial estimate $\hat{\mathbf{X}}_{t_0,t_0-1}$ with (known) error covariance matrix $\mathbf{P}_{t_0,t_0-1}$. Set $t = t_0$.

1. Compute the Kalman gain matrix $\mathbf{K}_t$, using $\mathbf{P}_{t,t-1}$ and eq. (29).
2. Use the current observation, y_t, and eq. (28) to compute the updated state-vector estimate, $\hat{\mathbf{X}}_{t,t}$. The error covariance matrix of this estimate, $\mathbf{P}_{t,t}$, is given by eq. (31).
3. Obtain $\hat{\mathbf{X}}_{t+1,t}$ and $\mathbf{P}_{t+1,t}$ using eqs. 34 and 35, respectively.
4. Set $t = t + 1$. Go to 1.

We make the following observations regarding implementation of the algorithm described above.

First, at each time t, all relevant information concerning the series is embodied in the state-vector estimate $\hat{\mathbf{X}}_{t,t}$. It is not necessary to store the history $\{y_1, \cdots, y_t\}$.

Second, the optimal gain sequence $\{\mathbf{K}_t\}$ is determined by the second-order statistics $\{\mathbf{Q}_t\}$ and $\{R_t\}$ describing the variability of the process and the measurements, respectively. Thus, if these quantities are known in advance, the gain sequence $\{\mathbf{K}_t\}$ can be precomputed.

Third, to start the recursion, an initial state estimate with known error covariance is necessary. The problem of obtaining such an estimate is considered below.

4.2 Initialization

In many Kalman filter applications, it is important to have an initial estimate of the model state vector as soon as the observations $y_1, y_2, \cdots$ begin. Since this estimate, $\hat{\mathbf{X}}_{1,0}$, is made prior to observing the series $\{y_t\}$, either a judicious guess must be made or information from an external source must be provided. In addition, the accuracy of this estimate must be quantified via the covariance matrix $\mathbf{P}_{1,0}$.

For complex models, the specification of good initialization parameters can be difficult. Also, improper specification can result in poor initial performance, because of improper weighting of the first observations. Both of these problems can be avoided completely if we can afford to wait until sufficient observations have been made that an initial state-vector estimate can be based on the data alone.

Since the short-term forecasting application considered in this paper is for use as a supplement to the yearly busy-season trunk forecast, it is neither necessary nor desirable to consider the use of the short-term forecast until sufficient data is available to make accurate predictions. For this reason, we recommend that the series $\{y_t\}$ be observed over an initialization period, say from $t = 1$ to $t = T$, so that the initial state estimate can be based on the data alone. The method by which this is accomplished is described below.

4.2.1 Linear model

Consider the linear dynamic system described by the equations

$$\mathbf{X}_{t+1} = \phi\mathbf{X}_t + \mathbf{W}_t \tag{37}$$

and

$$y_t = \mathbf{H}\mathbf{X}_t + \epsilon_t. \tag{38}$$

We assume that the matrices H and ϕ do not vary with time,* and that the latter is invertible.

We will show that each observation y_{T-k} within the initialization period can be expressed as a linear function of the state vector $\mathbf{X}_T$ at the end of the period, of future disturbances $\mathbf{W}_{T-k+j}$, and of ϵ_{T-k}, the observation error at time $t = T - k$.

We first note that, from eq. (37)

$$\mathbf{X}_{t-1} = \phi^{-1}(\mathbf{X}_t - \mathbf{W}_{t-1}), \tag{39}$$

so that

* This assumption is not necessary for the results of this section; however, it allows sufficient generality for the model considered and will simplify notation.

$$y_T = \mathbf{H}\mathbf{X}_t + \epsilon_T$$
$$y_{T-1} = \mathbf{H}[\phi^{-1}\mathbf{X}_T - \phi^{-1}\mathbf{W}_{T-1}] + \epsilon_{T-1}$$
$$y_{T-2} = \mathbf{H}[\phi^{-2}\mathbf{X}_T - \phi^{-2}\mathbf{W}_{T-1} - \phi^{-1}\mathbf{W}_{T-2}] + \epsilon_{T-2}.$$

In general, for $1 \leq k < T$,

$$y_{T-k} = \mathbf{H}\phi^{-k}\mathbf{X}_T - \mathbf{H}\left[\sum_{i=1}^{k} \phi^{-k+i-1}\mathbf{W}_{T-i}\right] + \epsilon_{T-k}. \tag{40}$$

In matrix notation,

$$\mathbf{Y}_T \equiv (y_T, \cdots, y_1)' = \mathbf{\Phi}_T\mathbf{X}_T + \mathbf{\Theta}_T, \tag{41}$$

where

$$\mathbf{\Phi}_T = \begin{bmatrix} \mathbf{H} \\ \mathbf{H}\phi^{-1} \\ \mathbf{H}\phi^{-T+1} \end{bmatrix}, \tag{42}$$

$$\mathbf{\Theta}_T = (\theta_T, \cdots, \theta_1)', \quad \text{and for} \quad 1 \leq k < T,$$

$$\theta_{T-k} = -\sum_{i=1}^{k} \mathbf{H}\phi^{-k+i-1}\mathbf{W}_{T-i} + \epsilon_{T-k}. \tag{43}$$

Equation (41) expresses the first T observations as a linear function of $\mathbf{X}_T$, plus an additive vector $\mathbf{\Theta}_T$ of zero-mean, correlated noise terms. Correlation between the components of $\mathbf{\Theta}_T$ is given by the covariance matrix

$$\mathbf{V} = (v_{jk}) = E[\mathbf{\Theta}_T\mathbf{\Theta}'_T].$$

Under the set of assumptions (26) the covariances v_{jk}, for $j \leq k$, are given by

$$\begin{aligned} v_{jk} &= \text{Cov}(\theta_{T-j+1}, \theta_{T-k+1}) \\ &= \sum_{i=1}^{j-1} E[\mathbf{H}\phi^{-j+i}\mathbf{W}_{T-i}][\mathbf{H}\phi^{-k+i}\mathbf{W}_{T-i}]' \\ &\quad + E(\epsilon_{T-j+1}\epsilon_{T-k+1}) \end{aligned} \tag{44}$$

$$= \sum_{i=1}^{j-1} \mathbf{H}\phi^{i-j}\mathbf{Q}_{T-i}(\mathbf{H}\phi^{i-k})' + \delta_{jk}R_k. \tag{45}$$

4.2.2 Minimum variance initialization estimate

Using the linear model (41) relating the first T observations to the value $\mathbf{X}_T$ of the state vector at the end of the initialization period, we can estimate $\mathbf{X}_T$ by weighted least-squares. That is, if the matrix $\mathbf{\Phi}_T$ defined in eq. (42) has rank n,* the minimum variance, unbiased linear

* For the linear growth, seasonal demand model defined in Section 3.3, the matrix (42) has rank n if $T > n$, the dimension of the state vector. That is, the matrices ϕ and $\mathbf{H}$ define an observable linear system.[9]

estimate of $\mathbf{X}_T$ is given[10] by

$$\hat{\mathbf{X}}_{T,T} = [(\mathbf{\Phi}'_T\mathbf{V}^{-1}\mathbf{\Phi}_T)^{-1}\mathbf{\Phi}'_T\mathbf{V}^{-1}]\mathbf{Y}_T. \tag{46}$$

The error covariance matrix, $\mathbf{P}_{T,T}$, is given by

$$\mathbf{P}_{T,T} = [\mathbf{\Phi}'_T\mathbf{V}^{-1}\mathbf{\Phi}_T]^{-1}. \tag{47}$$

Using eqs. (34) and (35), we can extrapolate these quantities to time $t = T + 1$. That is,

$$\hat{\mathbf{X}}_{T+1,T} = \phi\hat{\mathbf{X}}_{T,T} \tag{48}$$

and

$$\mathbf{P}_{T+1,T} = \phi\mathbf{P}_{T,T}\phi' + \mathbf{Q}_T. \tag{49}$$

Using eq. (48) as the prior state-vector estimate at time $T + 1$, we can process subsequent observations y_{T+1}, y_{T+2}, $\cdots$ sequentially, as described in Section 4.1.1.

To summarize, we have shown that an initial state-vector estimate, satisfying the conditions of Section 4.1.1, can be obtained from the linear model (41) using weighted least-squares. This estimate can then be used to start the recursive algorithm summarized in Section 4.1.3.

We will now discuss some general considerations regarding the application of these procedures to time-series forecasting.

4.3 Implementation considerations

So far in this section, we have shown how the weighted least-squares method can be used to obtain an initial state-vector estimate at time T, which is sequentially updated by the Kalman filter algorithm as new data becomes available. The application of each of these procedures is predicated on the knowledge of the second moments $\{\mathbf{Q}_t\}$ and $\{R_t\}$, which describe the variability of the state-vector process and the observations, respectively. In practice, however, these quantities are rarely known. Therefore, before we can apply these methods to an observed time series, the problem of determining appropriate values for these parameters must be considered.

4.3.1 Specification of Q_t and R_t

Various procedures have been proposed for the on-line identification of the parameters $\{\mathbf{Q}_t\}$ and $\{R_t\}$. (For example, see Ref. 11.) However, such an approach would add considerable complexity to the estimation procedures described in Sections 4.2 and 4.3 and, for the relatively short time series associated with the application considered in this paper, would probably yield little improvement in performance compared to the much simpler alternative, discussed below.

Instead of trying to estimate the parameters $\{R_t\}$ and $\{\mathbf{Q}_t\}$ describing the stochastic structure of each individual series, it may be possible to determine a single set of parameters, say $[R_t^*]$ and $[\mathbf{Q}_t^*]$†, which adequately approximates the true stochastic nature of the ensemble of time series being considered. That is, if within the domain of application, the performance of the procedures described in this section is relatively insensitive to deviations for the assumed values $[\mathbf{Q}_t^*]$ and $[R_t^*]$, then the use of a single set of parameters can be justified. Also, using this approach, a single gain sequence $\{\mathbf{K}_t\}$ and initialization matrix (46) can be precomputed and applied to all series. Thus, implementation is greatly simplified.

4.3.2 Truncated gain sequence

Another simplifying approximation is available for implementing the Kalman filter algorithm, in which the optimal gain sequence $\{\mathbf{K}_t\}$ is replaced by a simpler sequence, $\{\mathbf{K}_t^*\}$. For example, the truncated gain sequence, defined as

$$\mathbf{K}_{T+t}^* = \begin{cases} \mathbf{K}_{T+t} & \text{if} \quad t \leq \tau \\ & \\ \mathbf{K}_{T+\tau} & \text{if} \quad t \geq \tau \end{cases} \qquad (\tau \geq 1) \tag{50}$$

can often be used with good results. Two advantages of the truncated sequence over the full, optimal sequence are discussed below.

First, if τ is relatively small, only a few gain vectors $\{\mathbf{K}_{T+1}, \cdots, \mathbf{K}_{T+\tau}\}$ need to be computed and stored.

Second, and more important, is the use of the truncated gain sequence to avoid poor filter performance resulting from inaccurate specification of the parameters $\{\mathbf{Q}_t\}$. This is demonstrated in Fig. 2, which shows the theoretical performance of three seasonal trunk group algorithms based on the model described in Section 3.3.

In this example, the gain sequence has been computed under the false assumption that $\mathbf{Q}_t \equiv \mathbf{0}$‡, when in fact, $\mathbf{Q}_t > \mathbf{0}$. Under this assumption, the full gain sequence $\{\mathbf{K}_t\}$ converges to zero. Thus, new observations y_t are given insufficient weight in eq. (28) to allow the filter to track the randomly varying process.

Also shown in Fig. 2 is the theoretical performance of the truncated gain sequence ($\tau = 1$), again computed under the false assumption that $\mathbf{Q}_t = \mathbf{0}$. In this case, however, the gain is held at a constant, nonzero

† Actually, it is sufficient to determine appropriate values for $\{\mathbf{Q}_t\}$, since the gain $\mathbf{K}_t$ and initialization matrix (46) depend only on the relative magnitude of $\mathbf{Q}_t$ compared to the scalar R_t. For this reason, we will hereafter assume that $R_t \equiv 1$.

‡ The assumption that $\mathbf{Q}_t = \mathbf{0}$ implies that the process $\{\mathbf{X}_t\}$ evolves in a deterministic manner relative to a constant set of parameters, say $\mathbf{X}_0$. In this case, the minimum-variance estimates coincide with the usual (fixed-parameter) least-squares estimates. Thus, Fig. 2 illustrates the pitfalls of using such methods, when in fact, the underlying process is changing with time.

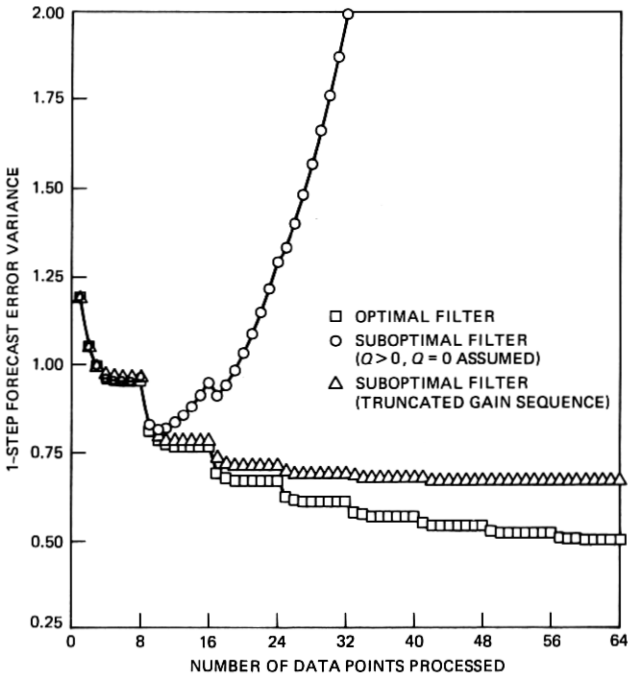

Fig. 2—Theoretical algorithm performance.

value, and the filter is able to track the series in a nearly optimal manner.

The main point being illustrated here is that, while correct specification of the parameters $\{\mathbf{Q}_t\}$ in the gain computation gives the filter sufficient responsiveness to track the process in an optimal manner, approximately the same effect can be achieved by truncating the gain sequence at some nonzero value. Thus, the truncated gain sequence is a useful implementation technique when a statistical description of the variability of the process is unavailable.

4.4 Summary

This section described the use of minimum variance, linear estimation procedures to estimate the parameters of a linear dynamic time-series model. By observing the process over a sufficiently long initialization period, an initial state-vector estimate can be obtained by the weighted least-squares method. Using the Kalman filter algorithm, this initial estimate can then be updated in an efficient, recursive manner as new observations become available.

In addition, we discussed the problem of obtaining good, suboptimal estimates when the exact statistical structure of each individual series is unknown. The application of these techniques will be illustrated in the next section, where we examine the performance of the proposed algorithm on actual trunk group data.

V. ALGORITHM PERFORMANCE-EMPIRICAL RESULTS

In Section III, we proposed a seasonal time-series model for trunk group load histories. We then described, in Section IV, a general procedure for estimating the parameters of such a model and for predicting future values of the time series. The performance of the resulting forecasting procedure on actual trunk group data is studied in this section.

We begin this section by describing the objectives that motivated us to undertake the empirical study described herein.

5.1 Study objectives

We wanted to use BOC trunk group data to determine the effectiveness of the proposed forecasting procedure in the three areas discussed below.

5.1.1 Forecast accuracy

The primary goal of the short-term trunk forecast is to obtain an accurate view of the approaching busy-season based on observed pre-busy-season traffic levels. In contrast, the yearly busy-season forecast (e.g., SPA) considers only the quantities of direct interest, namely, the time-series of yearly busy-season (peak) loads. While the short-term forecast has the advantage of operating with a shorter lead time and uses more frequent observations, the potential improvement, if any, in forecast accuracy over the busy-season forecast should be quantified.

5.1.2 Predictability

In addition to forecast accuracy, we would also like to investigate the degree with which service problems reveal themselves prior to the busy season. That is, how effective is a short-term forecast in predicting such events?

5.1.3 Error characterization

The intended application of the short-term forecast is to provide information regarding the adequacy of the planned trunk level for each trunk group. As measurements are collected and short-term forecasts are generated, the servicer may decide to

(*i*) take no action,
(*ii*) augment the group immeditately ("demand servicing"), or
(*iii*) augment the group in the near future ("anticipative demand servicing").

The removal of excess trunks is the responsibility of the yearly planned-servicing activity recommended by the yearly trunk forecast. No action should be taken unless the current or anticipated level of demand exceeds the in-service capacity.

The effective solution of the "anticipative demand servicing" decision problem requires that the uncertainty regarding the short-term forecast be understood and quantified. That is, servicing action should be taken only when, with a fair degree of certainty, action is required.

5.2 Methodology

In principal, the quantities of interest defined above can be determined (either analytically or by simulation) from the statistical structure of the time series being considered. In practice, however, the true underlying structure and dynamics are unknown. Therefore, to obtain meaningful answers to the questions posed above, we must test the performance of our forecasting procedures on real data.

5.2.1 The data

Historical trunk group data is retained in computer accessible form by most BOCs in the extended administrative history files of the Trunk Servicing System (TSS). Unfortunately, the longest histories available today consist of only about four and one-half years of data, and are available in those companies that made early use of this feature.

In the TSS system, trunk group traffic measurements are averaged and reported over "study periods" consisting of up to 20 business days. Roughly speaking, each such study period represents a rolling average of the four most recent weeks of valid data. For grade-of-service trunk groups, the busy-season corresponds to the study period (within the year) that has the highest offered load. This is the quantity that we wish to forecast.

Although we could attempt to model, track, and predict the complete time series of study period loads, this level of detail is neither necessary nor desirable. Instead, we can partition the year into some number L of "forecast periods," select the largest study period load within each forecast period, and work with the resulting time series of forecast period loads. Since the study period load series and the forecast period load series have the same maximum value in each year, it is sufficient to predict the latter.

The choice of the parameter L, the number of forecast periods per year, determines a trade-off between model complexity and response time. For large values of L, the seasonal pattern and time-series model required to accurately represent it are complex; also, data must be collected frequently. However, since the observations occur more frequently, it is possible to react more quickly to significant changes in the series $\{y_t\}$.

For the intended application, it was decided that by partitioning the year into eight short-term forecast periods, a reasonable balance between system complexity and response time would be achieved.

5.2.2 Selection of groups

To test the performance of the proposed algorithm in a controlled operating environment, it was necessary to select from the BOC data a subset of those trunk groups having reasonably "clean" data. That is, we omitted all groups whose histories exhibited one or more of the following characteristics.

(*i*) No discernible pattern: The load histories of certain trunk groups appear to have neither a within-the-year pattern nor a general growth trend. (This frequency occurs on groups serving very small volumes of traffic.) For these groups, our model is inappropriate; trending the yearly peaks (as in SPA) appears to be the most reasonable approach.

(*ii*) Very short histories: Because new groups begin and old groups leave service, the amount of available data varies among groups. We considered only those groups having at least four years of valid data.

(*iii*) Missing data: A group was not considered if, within any of the 32 forecast periods comprising the first four years, no valid study period load measurement was available.

(*iv*) Deterministic changes in the series: For certain groups, it was apparent that a major, deterministic change (e.g., a main station transfer) had drastically affected the load history. In principal, such changes can be accommodated by the model (3); however, information concerning such occurrences was not available in our data base, so these groups were not considered. (As we mentioned in Section 2.2, the ability to react appropriately to such occurrences is highly desirable for the intended application and will be considered in future work.)

A sample of approximately 300 trunk groups whose load histories satisfied conditions (*i*) to *(iv)* was selected for use in our study.

5.2.3 The forecasts

For each trunk group, the seasonal algorithm was initialized using the procedure developed in Section 4.2 and the first two years of data (16 data points). As we mentioned earlier, the application of both the minimum variance initialization and the Kalman filter algorithm requires specification of the matrix $\mathbf{Q}_t$ describing the variability of the state-vector process. This quantity was left as an adjustable parameter in the study.

The error covariance matrix $\mathbf{P}_{T+1,T}$ of the prior estimate $\hat{\mathbf{X}}_{T+1,T}$ was computed from eq. (49), using the error covariance matrix (47) of the initialization estimate $\hat{\mathbf{X}}_{T,T}$. Then, eq. (29) was used to compute the gain sequence $\{\mathbf{K}_{T+1}, \mathbf{K}_{T+2}, \cdots\}$. Since we decided to use a single value of $\{\mathbf{Q}_t\}$ for all trunk groups, the same gain sequence and initialization matrix could be computed once and applied to each time series.

Although the seasonal algorithm was designed to track the entire time series of forecast period loads, the principal quantity of interest for grade-of-service trunk groups is the yearly peak (busy season) load. To determine the accuracy with which the seasonal algorithm predicts this quantity based on data available k-periods prior to the busy season, we compared the busy-season load with the maximum value of the series predicted by our algorithm.

To compare the performance of the seasonal forecasting algorithm with SPA, both algorithms were run in parallel on the same set of data. However, since the information to be used in initializing the SPA (aggregate growth rate estimates for the offices on which the trunk group terminates[2]) was not available in our data, an alternate procedure had to be used.

To simulate the SPA initial estimates based on aggregate growth rate information, the busy-season loads were summed over all groups in each of the first two years of data. The ratio of the second-year sum to the first-year sum was used as the aggregate growth ratio R, with $R = 1.0325$ for the ensemble of trunk groups considered in the study. For each group, the busy-season load in the first year was used as its initial level estimate, $\hat{x}_0^{(1)}$, and the initial growth increment estimate was

$$\hat{x}_0^{(2)} = (R - 1)\hat{x}_0^{(1)}.$$

5.3 Examples

Before discussing the selection of algorithm parameters and the corresponding aggregate accuracy statistics, let us observe the performance of the two forecasting procedures on a few of the trunk groups considered in this study. Along with the aggregate statistics, which quantify the average improvement in accuracy provided by seasonal forecasting, these examples identify a number of general cases in which seasonal forecasting offers dramatic improvements over the yearly trending method of SPA.

Let us again consider the two example trunk group load histories discussed in Section 2.1. Figure 3 shows the load history of trunk group AA001444, except that now the results of the seasonal forecast are also shown. Recall that the first two years of data are used to obtain the initial state-vector estimate by a weighted least-squares procedure. The initialization is illustrated by the trajectory of circles, which represents an extrapolation backward in time of the state estimate at

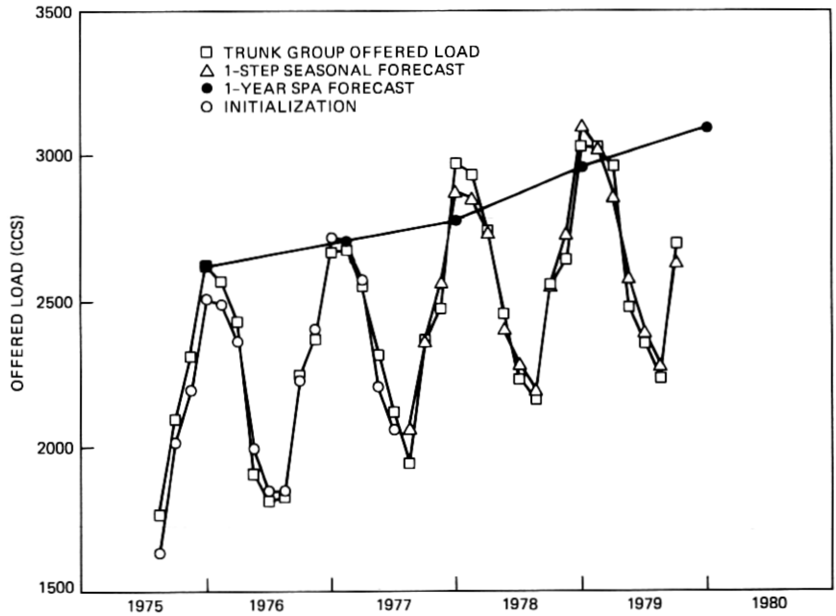

Fig. 3—Trunk group AA001444—seasonal forecast.

the end of the initialization period. Close examination reveals that the trajectory more closely fits the data at the end of the period than at the beginning; this is consistent with the fact that we are trying to estimate the current state.

The sequence of triangles beginning where the circles leave off is the sequence of one-step ahead seasonal forecasts obtained using the seasonal Kalman filter. It is evident that the seasonal algorithm is closely tracking the seasonal variation and underlying linear trend of this group and that SPA is predicting well the busy-season peaks. In fact, both procedures appear to perform equally well, on average, in predicting the yearly peak load.

A more interesting example is given in Fig. 4, which shows the performance of both algorithms on trunk group AA024225. As we mentioned earlier, the underlying growth pattern for this group exhibits a significant change in trend, which causes both algorithms to overforecast the peak load in the third year. However, by observing subsequent off-busy-season data points, the seasonal algorithm is able to detect this change, quickly home-in on the "signal," and give a reasonably good forecast of the busy-season load in the fourth year. In contrast, the yearly SPA forecast continues to be far off, since at the time the year-four forecast is made, only a single data point (the year-three busy-season load) off the previous trend line has been observed.

Fig. 4—Trunk group AA024225—seasonal forecast.

This example dramatically illustrates the potential value of the inter-busy-season information exploited by the seasonal algorithm.

A third interesting example, more in line with the intended anticipative demand servicing application discussed in Section 1.1, is given in Fig. 5. In this example, it appears that the trunk group load has undergone a moderate change in level and trend in the third year. The change occurs suddenly,* causing both algorithms to forecast significantly low in the third year. However, after processing subsequent data, the seasonal algorithm is able to forecast the next busy season with remarkable accuracy. In contrast, the SPA forecast again falls significantly low.

Finally, consider the trunk group shown in Fig. 6. In the second year, the busy-season load lies significantly above the trend for the other three years. This "outlier," which falls within the interval used by the SPA algorithm to screen for outliers, is accepted as a valid data point and processed accordingly. This causes SPA to "overshoot" the busy seasons in the next two years. In contrast, the seasonal algorithm, by processing the inter-busy-season data, recovers quickly and provides accurate forecasts in the next two years.

* Such a change, if known in advance, can be input by the forecaster, and the forecast can be changed accordingly. However, such "deterministic events" may be overlooked by the trunk forecaster, who typically has responsibility for a large number of groups.

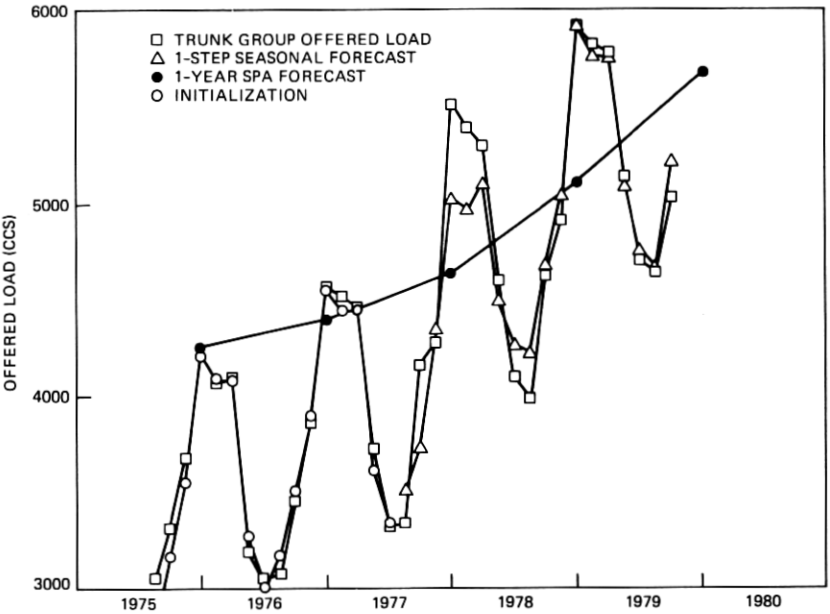

Fig. 5—Trunk group AA017330—seasonal forecast.

The significance of this fourth example lies in the additional protection against outliers provided by the seasonal forecast. If, in this example, the second SPA data point had been unusually low, it is likely that the SPA algorithm would have forecast significantly low in the next two years, possibly resulting in demand servicing activity. In such a case, the seasonal algorithm could be used to predict the service problem prior to its realization, or to estimate the additional capacity needed to relieve the problem when it occurred.

These four examples, while not constituting a meaningful statistical sample, dramatically illustrate a number of situations likely to occur in practice, where seasonal forecasting offers significant advantages over the yearly forecast of busy-season demand.

Before comparing the average accuracy results obtained for the SPA and seasonal forecasts, we will discuss the selection of the specific algorithm parameters used in the study.

5.4 Parameter selection

Recall from Section 4.3 that the ability of the algorithm to track random variations in the components of the state vector can be enhanced by either

(*i*) including a nonzero matrix $\mathbf{Q}_t$ that explicitly describes the variability of the parameters (Section 4.3.1),

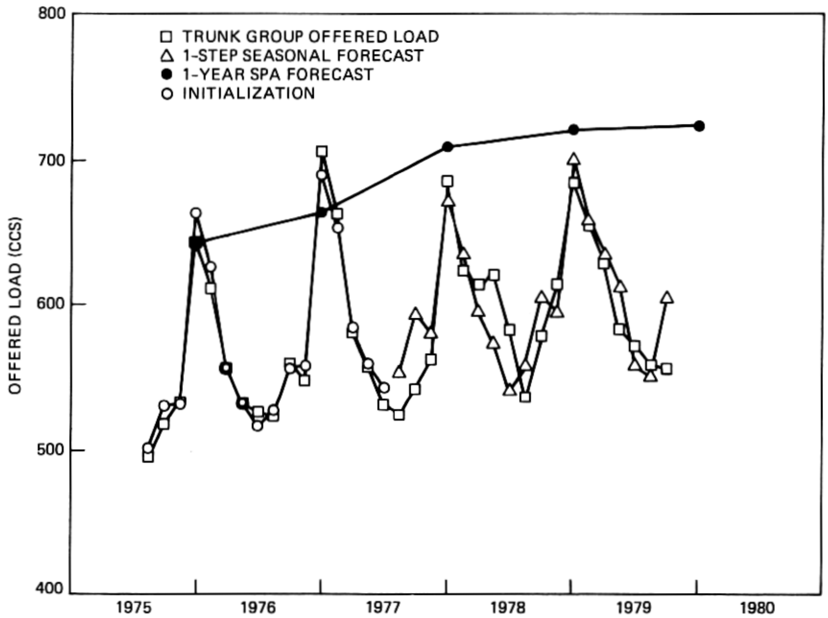

Fig. 6—Trunk group AA043161—seasonal forecast.

(*ii*) the use of certain heuristics, such as the truncated gain sequence (Section 4.3.2), or

(*iii*) some combination of (*i*) and (*ii*).

The effectiveness of these techniques was determined through experimentation. Specifically, for each version of the algorithm that we tested, the steady-state forecast error variance was estimated for each group and summed over all groups tested. With $\mathbf{Q}_t \equiv 0$, good results were obtained using the truncated gain sequence (50) with $\tau = 1$ (i.e., the first term $\mathbf{K}_{T+1}$ of the Kalman gain sequence was used to process each data point after initialization). Consistent with the hypothesis that some allowance must be made for random variation in the model parameters, this constant gain-vector sequence out-performed the full gain sequence computed under the assumption that $\mathbf{Q}_t \equiv 0$. Since the constant gain sequence also offers certain simplifications in algorithm implementation, this approach seemed very attractive.

By considering the behavior of various trunk group time series (as we discussed in Section 5.3.1), it also became clear that the seasonal forecasting algorithm should be able to follow variations in the underlying linear trend (see Fig. 2b). In addition to the responsiveness obtained through the use of the truncated gain sequence, additional responsiveness in the trend parameter is obtained by including a positive entry q_{22} in the matrix $\mathbf{Q}_t = (q_{ij})$. The optimal value of this parameter was determined empirically.

Using the average forecast error statistics, we then compared the performance of the resulting seasonal algorithm with that of SPA. The results are discussed below.

5.5 Accuracy

To quantitatively measure the performance of both the seasonal and the SPA forecasting procedures, both algorithms were used to forecast the busy-season loads on each of the trunk groups considered in the study. The results were compared using the relative error statistic, defined below.

Let a_g be the measured busy-season trunk group load in a given year, and $\hat{a}_g$ a forecast (obtained by either method) of this quantity. Then the relative forecast error is defined to be the quantity

$$e_g = \frac{\hat{a}_g - a_g}{a_g}.$$

To compare the approximately steady-state performance of the algorithms, this statistic was computed for each busy-season forecast in the fourth year. The results are shown in Fig. 7, which shows the distributions of the SPA and seasonal forecast errors, and in Table I. The latter compares the accuracy statistics of the seasonal forecasts,

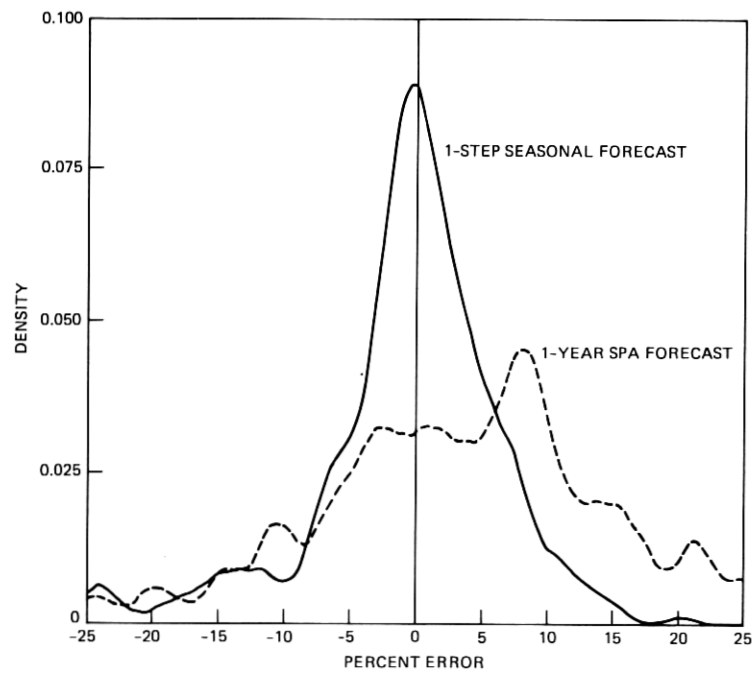

Fig. 7—Year 4 forecast error distributions.

Table I—Year 4 busy-season forecast results

Forecast	Percent		
	Bias	Average absolute Error	rms Error
1-Year SPA	3.5	10.3	13.5
1-Step seasonal	−0.7	5.1	7.5
2-Step seasonal	−0.8	6.1	8.5
3-Step seasonal	−0.5	6.5	9.0
4-Step seasonal	0.1	7.0	9.4
5-Step seasonal	0.2	7.6	10.4
6-Step seasonal	0.4	7.9	10.9
7-Step seasonal	0.5	8.0	10.8
8-Step seasonal	1.8	8.4	11.6

made one forecast period prior to the busy-season, with those of the SPA one-year-ahead forecast. The results show a significant improvement in accuracy, using either the average absolute relative error or the rms relative error. In either case, the observed error in the seasonal forecast is approximately half as large as that of the SPA forecast. Also, recall that these statistics, obtained by comparing the forecast with the measured load, also reflect load measurement error. Thus, on average, the seasonal forecast is off by less than 5 percent.

Part of this improvement must be attributed to the fact that the short-term busy-season forecast is made approximately 1½ months prior to the busy season, compared with one year for SPA. Additional improvement can be attributed to the exploitation of the additional structure of the seasonal time series. The relative importance of these two factors is illustrated in Fig. 8, which shows the relationship between short-term busy-season forecast error and lead time. Begin-

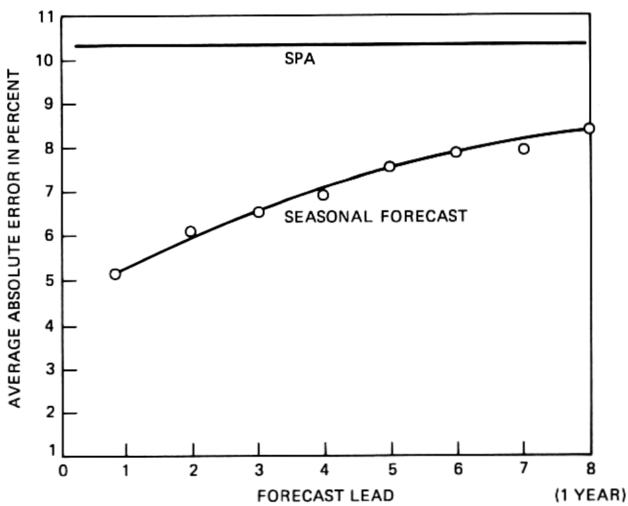

Fig. 8—Busy-season forecast error.

ning at 5.1 percent for the one-step forecast, the average absolute error statistic is a concave function of lead time, increasing to 8.4 percent for the one-year-ahead seasonal forecast. This suggests that most of the improvement in forecast accuracy can be attributed to the seasonal algorithm's ability to utilize recent data.

5.6 Summary

In this section, we described an empirical investigation of the performance of the seasonal trunk forecasting algorithm developed in Section IV. We showed that, on average, a short-term (one forecast period ahead) forecast of the approaching busy season is approximately twice as accurate as the yearly SPA forecast. This improvement in accuracy deteriorates with forecast lead time, being only 20 percent more accurate than SPA when projecting a full year ahead.

More important, we were able to identify a number of situations, likely to occur in practice, where the seasonal forecast significantly outperforms SPA. These situations tend to occur where unusual data, such as outliers or sudden, unanticipated changes in the growth pattern occur. By using the information provided by the inter-busy-season traffic levels, the seasonal algorithm is able to recover rapidly from such disturbances, and provide accurate forecasts in subsequent years. In contrast, the SPA algorithm, which processes a single data point each year, takes considerably more time to recover.

VI. SUMMARY AND CONCLUSIONS

This paper has described a comprehensive study of the use of seasonal forecasting algorithms in the trunk servicing process.

We began by discussing the basic requirements of a short-term trunk forecasting system and then described a general class of time-series models well suited for such applications. After developing a model for linear growth and seasonal demand, we considered various minimum variance procedures for estimating the parameters of such a model from a given time series. By observing the series over a two-year initialization period, an initial estimate of the model parameters is obtained by weighted least-squares. Subsequent observations are processed by an efficient recursive procedure known as the Kalman filter. Because of the optimality of these algorithms, a minimum amount of data is required before accurate forecasts can be made.

To verify the appropriateness of these procedures on actual trunk group data, an empirical investigation of the performance of the proposed algorithm was undertaken. The main conclusions of the study are discussed below:

6.1 Accuracy

On the average, the one-step (≈1½ months) ahead seasonal forecast of the busy-season load is approximately twice as accurate as the year-to-year forecast provided by the SPA. Most of this improvement can be attributed to the ability of the seasonal algorithm to utilize effectively recent observations. The one-year ahead, seasonal forecast error is only 20 percent less than that of SPA.

6.2 Anticipative demand servicing

The seasonal algorithm demonstrates the ability to respond quickly and accurately to various situations that tend to result in inaccurate SPA forecasts. Thus, the seasonal algorithm represents a potentially valuable tool for trunk network administration. In many cases, it can accurately predict that the currently planned trunk level is inadequate for the approaching busy-season demand level, and the appropriate trunk group augment. Conversely, it can also be used to identify potentially useful reserve capacity.

6.3 Relation to SPA

Since current trunk provisioning methods generally require that planned-servicing decisions be made nearly a year in advance, these results lead us to the conclusion that the greater simplicity and generality of the linear trending method of SPA makes that algorithm the appropriate choice for the annual busy-season trunk forecast. In the future, however, new technologies may make it both possible and economical to maintain the network in a near-optimal configuration through more-frequent planned-servicing adjustments. In that case, an accurate, seasonal forecast may be a better alternative for planned servicing.

REFERENCES

1. A. J. David and C. D. Pack, "The Sequential Projection Algorithm: A New and Improved Traffic Forecasting Procedure," Proc. Ninth Int'l Teletraffic Congress, Torremolinos, Spain, 1979.
2. J. P. Moreland, "A Robust Sequential Projection Algorithm for Traffic Load Forecasting," B.S.T.J., this issue.
3. R. J. Armstrong, R. Gottdenker, and R. L. Kornegay, "Servicing Trunks by Computer," Bell Laboratories Record, *54,* No. 2 (February, 1976).
4. G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day, 1976.
5. C. D. Pack and B. A. Whitaker, "Kalman Filter Models For Network Forecasting," B.S.T.J., this issue.
6. P. J. Harrison and C. F. Stevens, "Bayesian Forecasting," J. Royal Stat. Soc. (Series B), *38,* No. 3, (1976), pp. 205–47.
7. T. W. Anderson, *The Statistical Analysis of Time Series*, New York: John Wiley, 1971.
8. R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," J. Basic Eng., *82* (March 1960), pp. 34–45.
9. A. Gelb, ed., *Applied Optimal Estimation*, Cambridge, Mass.: MIT Press, 1974.
10. M. Kendall and A. Stuart, *The Advanced Theory of Statistics*, *2,* New York: MacMillan, 1979.
11. R. K. Mehra, "On the Identification of Variances and Adaptive Kalman Filtering," IEEE Trans. Automatic Control, *AC-15* (April 1970), pp. 175–84.