A Control Chart for a General Gaussian Process

Robert Lund
Department of Statistics
The University of Georgia
Athens, GA 30602-1952

I. V. Basawa
Department of Statistics
The University of Georgia
Athens, GA 30602-1952

W. J. Padgett
Department of Statistics
The University of South Carolina
Columbia, SC 29208

Lynne Seymour
Department of Statistics
The University of Georgia
Athens, GA 30602-1952

June 24, 1996

ABSTRACT

It has been repeatedly demonstrated that X-bar quality control charts perform poorly when the process subgroups being monitored are correlated. In this paper, we propose and investigate the performance of a control chart that accounts for subgroup correlations in a general Gaussian process. The time series innovations algorithm is used to construct the desired chart. This "innovations" chart is shown to perform as a standard X-bar chart for any Gaussian process; furthermore, the chart can be applied in the nonstationary and stationary settings.

A simulation study shows that the innovations chart performs well even when the correlation structure of the process must be estimated from a small number of subgroups. The innovations chart is then used to study two data sets. The first data set consists of motor shaft diameters and has correlated, but stationary, subgroups. The results here show that erroneous conclusions can be reached if subgroup correlations are ignored. The second data set consists of averaged monthly temperatures from Boston, MA, New York City, NY, and Washington, DC; this data consists of correlated, but decisively nonstationary, subgroups as seasonality in the first two subgroup moments are prominent. The analysis of this data helps identify historical cold snaps and heat waves and provides solid criteria for assessing their future occurrences.

KEY WORDS: Quality Control; X-bar Chart; Innovations Algorithm, ARMA Model.
1. Introduction. It is common practice in statistical process control (SPC) to utilize Shewhart $X$-bar control charts for monitoring the mean and variation of a process. These charts rely heavily on the assumption that sample statistics are independent between process subgroups; however, it has recently been demonstrated that strong autocorrelations between process subgroups frequently exist. Such autocorrelations can occur for a number of reasons. One reason is that, because of current automated measurement and recording technology, subgroup samples may be taken with a high frequency, with consecutive samples being similar in nature and hence statistically correlated. Other examples of correlated subgroups occur when items made by a worker exhibit similar characteristics due to the way a machine is handled, when a process shows seasonal patterns due to materials or weather, or when the alertness of a worker changes over time.


Most of the above authors investigate the case where one measurement is taken from the process at each sampling time. A typical approach to monitoring correlated processes when only one measurement is taken at each sampling time is to fit an ARIMA time series model to the process observations and then monitor one-step-ahead forecasts with a control chart for single measurements (see Alwan and Roberts (1988), Alwan (1992), and Montgomery and Mastrangelo (1991), among others). If the ARIMA model orders and parameters have been identified correctly, then the residuals, or differences between forecasted and observed values, should be approximately uncorrelated. A
Shewhart "individuals" chart (X chart), CUSUM chart, or other chart for single observations is then applied to the residuals to monitor the process. Willemain and Runger (1994) use a different approach based on random run lengths.

In many applications such as the ones presented in Section 5, multiple observations are taken from the process at each sampling time. It is the purpose of this paper to develop and investigate a control chart for this general situation when the subgroups are correlated. We will consider the general case where a subgroup of \( n_t \geq 1 \) measurements is sampled from a process at time \( t \) (\( n_t \) is allowed to vary with \( t \)). The methods presented below require only the knowledge or estimation of the correlation structure of the process being monitored; hence, no stationarity assumptions are required. Stationarity of subgroups is typically violated in periodic processes such as the temperature series studied in Section 5 (see Spurrier and Thombs (1990) for more on control charts for cyclic processes), and in economic applications where the nonstationary random walk is commonly used to describe process fluctuations.

The rest of this paper proceeds as follows. In Section 2, model notation and assumptions are clarified and the control chart is developed recursively from the time series innovations algorithm. Section 3 develops several examples of the methodology mathematically; Section 4 explores control chart performance via simulation; here, it is demonstrated that the proposed chart makes a vast improvement over the standard X-bar chart that neglects subgroup correlations. The chart is also shown to work well when the correlation structure of the process subgroups must be estimated. Section 5 applies the results of this paper to the motor shaft diameter data in Devor et al. (1992) and to a data set of monthly averaged temperatures from Boston, MA, New York City, NY, and Washington, DC; these two applications illustrate the need for, and flexibility of, the proposed chart. Section 6 concludes the paper with some comments.

2. A General Chart for a Gaussian Process. Suppose that at each time \( t, 1 \leq t \leq k \), a subgroup of \( n_t \) measurements is sampled from a Gaussian process. The process subgroup at time \( t \), denoted \( \{X_{ij}, 1 \leq j \leq n_t\} \), satisfies
$X_{tj} = \mu_t + \epsilon_{tj}, \quad (2.1)$

where $\{\mu_t\}$ is a Gaussian process and the subgroup "errors" $\{\epsilon_{tj}\}$ are assumed to be independent and identically distributed (iid) mean zero Gaussian random variables with $\text{Var}(\epsilon_{tj}) = \sigma^2$. One interpretation of (2.1) is that the "mean setting" of a machine at time $t$ is $\mu_t$; this mean setting can fluctuate randomly with $t$. Conditionally upon $\mu_t$ at time $t$, the subgroup sample $\{X_{tj}, 1 \leq j \leq n_t\}$ is an iid Gaussian random sample with mean $\mu_t$ and variance $\sigma^2$. For simplicity, we assume that $E[\mu_t] = 0$; adjustments to the methodology are straightforward if this is not the case (cf. Section 5). Finally, we assume that $\{\epsilon_{tj}\}$ and $\{\mu_t\}$ are independent.

The covariance structure of $\{\mu_t\}$, not necessarily stationary in form, will be denoted by

$$\text{Cov}(\mu_t, \mu_s) = \sigma(t, s). \quad (2.2)$$

For properties of the subgroups, (2.1) and $E[\mu_t] = 0$ give $E[X_{tj}] = 0$. The covariance between items in subgroups $t$ and $s$ is $\text{Cov}(X_{tj}, X_{sj}) = \sigma(t, s) + \sigma^2 e_{tj} = s \cap j = t$; hence, this model has correlated subgroups whenever $\sigma(t, s) \neq 0$ for some $t \neq s$. Notice that the subgroup $\{X_{tj}, 1 \leq j \leq n_t\}$ is only conditionally independent for any fixed $t$; in fact, the correlation between $X_{tj}$ and $X_{ti}$ when $j \neq i$ is $\text{Corr}(X_{tj}, X_{ti}) = [\sigma(t, t) + \sigma^2]^{-1} \sigma(t, t)$.

The subgroup average at time $t$ is

$$\bar{X}_t = n_t^{-1} \sum_{j=1}^{n_t} X_{tj}, \quad (2.3)$$

Setting

$$\bar{\epsilon}_t = n_t^{-1} \sum_{j=1}^{n_t} \epsilon_{tj}, \quad (2.4)$$

we obtain

$$\bar{X}_t = \mu_t + \bar{\epsilon}_t \quad (2.5)$$

from (2.1). The covariance structure of $\{\bar{X}_t\}$ is easily obtained from (2.1) – (2.5):

$$\kappa(t, s) = \text{Cov}(\bar{X}_t, \bar{X}_s) = \sigma(t, s) + n_t^{-1} \sigma^2 e_{t \cap s}. \quad (2.6)$$
Notice that \( \{ \tilde{X}_t \} \) is stationary when \( \sigma(t,s) \) depends on \( |t-s| \) only and \( n_t \) is constant in \( t \).

To develop a general control chart, we will make use of a set of one-step ahead predictions and prediction errors. Let

\[
\hat{X}_t = E[\tilde{X}_t | \tilde{X}_1, \ldots, \tilde{X}_{t-1}]
\]

be the best mean squared error predictor of \( \tilde{X}_t \) based upon \( \tilde{X}_1, \ldots, \tilde{X}_{t-1} \) and define its mean squared prediction error as

\[
v^2_{t-1} = E[(\hat{X}_t - \tilde{X}_t)^2 | \tilde{X}_1, \ldots, \tilde{X}_{t-1}].
\]

Since \( \{ \tilde{X}_t \} \) is Gaussian, conditional expectations are linear in the predicting variables and the time series innovations algorithm (see Brockwell and Davis, 1991) can be used to recursively compute \( \hat{X}_t \) and \( v^2_t \) in \( t \) provided the covariance matrix \( \kappa_n = (\kappa(i,j))_{i,j=1}^n \) is invertible for each \( n \geq 1 \). This is done through the innovations representation

\[
\hat{X}_{t+1} = \sum_{j=1}^t \theta_{tj}(\tilde{X}_{t-j+1} - \hat{X}_{t-j+1}), \ t \geq 1,
\]

with \( \hat{X}_1 = 0 \). The quantities \( \theta_{tj} \) and \( v^2_t \) can be computed recursively from \( v^2_0 = \kappa(1,1) \) and

\[
\theta_{t,t-k} = (v^2_k)^{-1}\left[ \kappa(t+1,k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j} \theta_{t,t-j} v^2_j \right], \ k = 0, \ldots, t-1,
\]

\[
v^2_t = \kappa(t+1,t+1) - \sum_{j=0}^{t-1} \theta_{t,t-j} v^2_j,
\]

for \( t \geq 1 \). Equation (2.10) can be solved in the order \( v^2_0; \theta_{11}, v^2_1; \theta_{22}, v^2_2, \ldots \).

The Projection Theorem (cf. Brockwell and Davis, 1991) shows that \( \hat{X}_t - \tilde{X}_t \) is independent of \( \tilde{X}_i \) for \( 1 \leq i \leq t-1 \); hence, since \( \{ \tilde{X}_t \} \) is Gaussian, (2.8) becomes \( v^2_{t-1} = E[\hat{X}_t - \tilde{X}_t]^2 \). We comment that a simplistic sufficient condition for \( \kappa_n \) to be invertible for each \( n \geq 1 \) follows from Proposition 5.1.1 of Brockwell and Davis (1991) when \( \{ \mu_t \} \) is stationary. Letting \( \gamma(h) = \sigma(t+h,t) \) for natural numbers \( h \), this condition is that \( \gamma(0) > 0 \) and \( \gamma(h) \to 0 \) as \( h \to \infty \).

For a control chart, we will use the standardized prediction residuals
Since $\{\mu_t\}$ and $\{\epsilon_t\}$ are Gaussian, the prediction error sequence $\{\hat{X}_t - \bar{X}_t\}$ is uncorrelated and Gaussian; hence, $\{Z_t\}$ is a sequence of iid standard normal random variables. A large $Z_t$ in absolute value indicates a large deviation between $\bar{X}_t$ and what was predicted by $\hat{X}_t$ and is associated with an out of control process. In many applications, the control limits $\pm 3$ are used and the process is called out of control at time $t$ if $|Z_t| > 3$; the probability of this happening due to chance variation when the process is actually in control is approximately 0.0027.

In practice, the parameters governing the correlation structure of $\{\mu_t\}$ and $\sigma^2_t$ are typically unknown. These parameters can be estimated from a small number of subgroup measurements under the innovations framework with the method of maximum likelihood. Letting $\mathbf{\nu}$ denote all unknown model parameters and using the notation $Z_t(\mathbf{\nu})$ to explicitly denote the dependence of the standardized residual $Z_t$ on model parameters, the model likelihood function, denoted by $L(\mathbf{\nu}; \{\bar{X}_t\})$, has the innovations form (cf. Brockwell and Davis, 1991)

$$L(\mathbf{\nu}; \{\bar{X}_t\}) = (2\pi)^{-k/2} \left( \prod_{t=1}^{k} \frac{\sigma^2_{t-1}}{v_{t-1}} \right)^{-1/2} \exp \left[ -\frac{1}{2} \sum_{t=1}^{k} Z_t^2(\mathbf{\nu}) \right] .$$

(2.12)

An estimate of $\mathbf{\nu}$, denoted by $\hat{\mathbf{\nu}}$, can be computed by minimizing the scaled negative log likelihood function $H(\mathbf{\nu}; \{\bar{X}_t\})$ defined by

$$H(\mathbf{\nu}; \{\bar{X}_t\}) = \sum_{t=1}^{k} \ln(v_{t-1}) + \sum_{t=1}^{k} Z_t^2(\mathbf{\nu})$$

(2.13)

in the argument $\mathbf{\nu}$.

Once $\hat{\mathbf{\nu}}$ is found, an estimate of the covariance function $\kappa(\cdot, \cdot)$ in (2.6) can be obtained by plugging $\hat{\mathbf{\nu}}$ in for $\mathbf{\nu}$. This estimated covariance structure can in turn be used in (2.9) and (2.10) to compute an estimated version of $\{Z_t(\mathbf{\nu})\}$ which we denote by $\{Z_t(\hat{\mathbf{\nu}})\}$. The control limits $\pm 3$ can then be applied to $\{Z_t(\hat{\mathbf{\nu}})\}$ to assess the status of the process. In general, $\{Z_t(\hat{\mathbf{\nu}})\}$ will not be an uncorrelated series; however, as $k \to \infty$, the maximum likelihood estimate $\hat{\mathbf{\nu}}$ frequently converges in
probability to \( z \). When this happens, \( Z_t(\hat{\theta}) \) converges in probability to \( Z_t(z) \) as \( k \to \infty \) for each \( t \) and \( \{Z_t(\hat{\theta})\} \) becomes an asymptotically uncorrelated series. In practice, the simulations in Section 4 will show that this estimation procedure yields a control chart that performs reasonably well for a variety of correlation structures when between 20 and 100 subgroups are used for estimation purposes.

3. Examples. In this section, we explore some specific cases of the Section 2 model that are encountered in practice.

3.1 The First Order Autoregression. Consider the AR(1) difference equation model for \( \{\mu_t\} \):

\[
\mu_t = \phi \mu_{t-1} + A_t,
\]

where \( |\phi| < 1 \) is assumed for stationarity and causality of \( \{\mu_t\} \). In (3.1), \( \{A_t\} \) is an iid mean zero Gaussian random sequence with \( \text{Var}(A_t) = \sigma_A^2 \) that is independent of \( \{\epsilon_{tj}\} \). For simplicity, we take \( n_t = n \). Using \( \gamma(h) = (1 - \phi^2)^{-1} \sigma_A^2 |\phi|^h \), we see that the correlation between items in subgroups \( t \) and \( s, t \neq s \), is

\[
\text{Corr}(X_{tj}, X_{sl}) = \frac{\sigma_A^2 |\phi|^{t-s}}{\sigma_A^2 + (1 - \phi^2) \sigma_\epsilon^2}.
\]

Notice that, in accordance with many physical processes, \( \text{Corr}(X_{tj}, X_{st}) \) exhibits geometric decay to zero as \( |t-s| \to \infty \) (of course, this geometric decay is not unique to the AR(1) model). Since \( \gamma(h) \to 0 \) as \( h \to \infty \), \( \kappa_n \) is invertible for each \( n \geq 1 \).

We note that (3.1) and (2.5) can be regarded as Kalman filter state and observation equations respectively. Using (2.5) and (3.1) in (2.7), we obtain \( \hat{X}_t = \phi \hat{\mu}_{t-1} \) where \( \hat{\mu}_{t-1} = E[\mu_{t-1} | \bar{X}_1, \ldots, \bar{X}_{t-1}] \) is the best mean squared error predictor of \( \mu_t \) from \( \bar{X}_1, \ldots, \bar{X}_{t-1} \). Further manipulations with (2.5) and (3.1) provide

\[
v_{t-1}^2 = \phi^2 v_{t-1}^2 + \sigma_A^2 + n^{-1} \sigma_\epsilon^2,
\]

where \( v_{t-1}^2 = E[(\hat{\mu}_t - \mu_t)^2 | \bar{X}_1, \ldots, \bar{X}_t] = E[(\hat{\mu}_t - \mu_t)^2] \). Hence, the control chart is easily constructed from \( \{P_t^2\} \) and \( \{\hat{\mu}_t\} \). Classical Kalman filter recursions (cf. Shumway, 1988) provide
\[ \hat{\mu}_t = \phi \hat{\mu}_{t-1} + K_t (\bar{X}_t - \phi \hat{\mu}_{t-1}), \]  

(3.3)

where \( K_t \) is the Kalman gain

\[ K_t = \frac{\phi^2 P_{t-1}^2 + \sigma_A^2}{n^{-1} \sigma^2 + \phi^2 P_{t-1}^2 + \sigma_A^2}. \]  

(3.4)

To compute \( K_t \) and \( \hat{\mu}_t \) recursively in \( t \), one uses (3.3) and (3.4) and recursively updates \( P_t^2 \) with

\[ P_t^2 = \frac{n^{-1} \sigma^2 (\phi^2 P_{t-1}^2 + \sigma_A^2)}{n^{-1} \sigma^2 + \phi^2 P_{t-1}^2 + \sigma_A^2}. \]  

(3.5)

the scheme is initiated with \( \hat{\mu}_0 = 0 \) and \( P_0^2 = \sigma_A^2 (1 - \phi^2)^{-1} \). Alternatively, the above equations are easily translated into recursions for \( \hat{X}_t \) and \( v_t^2 \). Using \( \hat{X}_t = \phi \hat{\mu}_{t-1} \) in (3.3) gives

\[ \hat{X}_{t+1} = \phi (1 - K_t) \hat{X}_t + \phi K_t \bar{X}_t \]  

(3.6)

for \( t \geq 1 \), which is similar to exponential smoothing. Using (3.2) in (3.5) gives

\[ v_t^2 = \sigma^2 + (1 + \phi^2) (n^{-1} \sigma^2) - \frac{\phi^2 [n^{-1} \sigma^2]^2}{v_{t-1}^2}. \]  

(3.7)

These recursions are initiated with \( \hat{X}_1 = 0 \) and \( v_0^2 = \sigma_A^2 (1 - \phi^2)^{-1} + n^{-1} \sigma^2 \).

3.2 The Random Walk. Consider the random walk model for \( \{\mu_t\} \):

\[ \mu_t = \mu_{t-1} + A_t, \quad t \geq 1, \]  

(3.8)

with \( \mu_0 = 0 \). Here, \( \{A_t\} \) is an iid mean zero Gaussian sequence with \( \text{Var}(A_t) = \sigma_A^2 \) that is independent of \( \{\epsilon_t\} \). The random walk is useful for modeling processes whose subgroups exhibit strong positive short-term correlations. The covariance structure of \( \{\mu_t\} \) is well known as \( \sigma(i, j) = \sigma_A^2 \text{min}(i, j) \) and (2.6) gives

\[ \kappa(i, j) = \sigma_A^2 \text{min}(i, j) + n^{-1}_i \sigma^2_{\epsilon} [i = j]. \]  

(3.9)

hence,

\[ \text{Corr}(X_{tj}, X_{sl}) = \frac{\sigma_A^2 \text{min}(t, s)}{(t \sigma_A^2 + n^{-1}_t \sigma^2_{\epsilon})^{1/2} (s \sigma_A^2 + n^{-1}_s \sigma^2_{\epsilon})^{1/2}} \]  

(3.10)

for \( t \neq s \). Note that \( \text{det}(\kappa_n) \geq (\sigma_A^2)^n \text{det}(M_n) \) where the \( (i, j) \)th entry of \( M_n \) is \( \text{min}(i, j) \). Using
$\det(M_n) = 1$ shows that $\det(\kappa_n) > 0$ for all $n \geq 1$; hence, $\kappa_n$ is invertible for all $n \geq 1$.

In the general case where $n_t$ varies with $t$, $\theta_{ij}$ and $\nu_t^2$ can be obtained recursively from (2.10). When $n_t \equiv n$, the innovation recursions can be simplified. A simplification in the computations can be made if we apply the innovations algorithm to $\{\hat{X}_1, Y_2, Y_3, \ldots\}$ instead of $\{\hat{X}_1, \hat{X}_2, \ldots\}$, where, for $t \geq 2$, (3.8) and (2.5) give

$$Y_t = \hat{X}_t - \hat{X}_{t-1} = A_t + \varepsilon_t - \overline{\varepsilon}_{t-1}. \quad (3.11)$$

Since the closed linear span of $\{\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_t\}$ equals the closed linear span of $\{\hat{X}_1, Y_2, Y_3, \ldots, Y_t\}$, one has

$$\hat{Y}_{t+1} = E[ Y_{t+1} | \hat{X}_1, Y_2, \ldots, Y_t ] = E[ \hat{X}_{t+1} - \hat{X}_t | \hat{X}_1, \ldots, \hat{X}_t ] = \hat{X}_{t+1} - \hat{X}_{t} \quad (3.12)$$

for $t \geq 2$ from (3.11). Letting $\{\eta_t^2, t \geq 0\}$ denote the one step ahead mean squared prediction errors for $\{\hat{X}_1, Y_2, Y_3, \ldots\}$, we obtain $\hat{X}_1 = 0$ and $\eta_0^2 = \nu_0^2 = \sigma^2 + n^{-1} \sigma^2$. \(\hat{Y}_2\) and $\eta_1^2$ are obtained from

$$\hat{Y}_2 = \frac{-n^{-1} \sigma^2}{\sigma^2_A + n^{-1} \sigma^2_e} \hat{X}_1, \quad \eta_1^2 = \sigma^2 + 2n^{-1} \sigma^2 - \frac{[n^{-1} \sigma^2_e]^2}{\sigma^2_A + n^{-1} \sigma^2_e}.$$ 

The crucial observation is that $\{Y_t, t \geq 2\}$ can be written as the first order moving average

$$Y_t = Z_t + \theta Z_{t-1}, \quad (3.13)$$

where $\{Z_t\}$ is Gaussian white noise with $\text{Var}(Z_t) \equiv \sigma_Z^2$; hence, $\{\hat{X}_1, Y_2, \ldots\}$ is a one dependent sequence. The parameters $\theta \in (-1, 0)$ and $\sigma_Z^2$ can be obtained by equating moments in (3.11) and (3.13); this yields $(1 + \theta^2) \sigma_Z^2 = \sigma^2_A + 2n^{-1} \sigma^2$ and $\theta \sigma_Z^2 = -n^{-1} \sigma^2_e$ which can be solved as

$$\theta = \frac{-\sigma^2_A + 2n^{-1} \sigma^2_e + \sqrt{\sigma^2_A + 2n^{-1} \sigma^2_e}}{2n^{-1} \sigma^2_e}, \quad \sigma_Z^2 = \frac{2[n^{-1} \sigma^2_e]^2}{(\sigma^2_A + 2n^{-1} \sigma^2_e) - \sqrt{\sigma^2_A + 2n^{-1} \sigma^2_e}}.$$ 

Applying (2.10) to $\{\hat{X}_1, Y_2, Y_3, \ldots\}$ yields (here, $\theta_{ij}$ refer to the prediction coefficients for $\{\hat{X}_1, Y_2, \ldots\}$)

$$\theta_{ij} = 0, \ j = 2, \ldots, t; \ \theta_{ii} = -\eta_{i-1}^{-2} [n^{-1} \sigma^2_e]; \ \text{and} \quad (3.14)$$

$$\eta_i^2 = \sigma^2_A + 2n^{-1} \sigma^2_e - \eta_{i-1}^{-2} [n^{-1} \sigma^2_e]^2.$$ 

8
for \( t \geq 2 \). Using (3.14) in (2.9) gives
\[
\hat{Y}_{t+1} = \theta_{t1}(Y_t - \hat{Y}_t),
\]
for \( t \geq 2 \) which can be combined with (3.12) to give the desired recursion
\[
\hat{X}_{t+1} = (1 + \theta_{t1})\hat{X}_t - \theta_{t1}\hat{X}_t
\]
\[ (3.15) \]
for \( t \geq 2 \). It remains to compute \( \hat{X}_2 \) and \( \nu_t^2 \) for \( t \geq 1 \). Straightforward computations give
\[
\hat{X}_2 = \frac{\sigma_A^2}{\sigma_A^2 + n^{-1}\sigma_e^2} \hat{X}_1 \quad \text{and} \quad \nu_t^2 = 2\sigma_A^2 + n^{-1}\sigma_e^2 - \frac{(\sigma_A^2)^2}{\sigma_A^2 + n^{-1}\sigma_e^2},
\]
when \( t \geq 2 \); (3.11) and (3.12) give \( \nu_t^2 = \eta_t^2 \) as required.

We again notice that (3.15) is similar to exponential smoothing; however, \( \theta_{t1} \) depends on \( t \). From (3.14), one can show that \( \theta_{t1} \to \theta \) as \( t \to \infty \); hence, asymptotically, predictions for the random walk \( \{ \mu_t \} \) obey an exponentially weighted moving average relation. We also note that (3.7) with \( \phi = 1 \) is identical to the second equation in (3.14).

3.3 Causal Linear Processes. A flexible class of stationary models for \( \{ \mu_t \} \), which includes the causal ARMA class, can be written in the form
\[
\mu_t = \sum_{k=0}^{\infty} \psi_k A_{t-k},
\]
where \( \sum_{k=0}^{\infty} |\psi_k| < \infty \) and \( \{ A_t \} \) is an iid Gaussian sequence, independent of \( \{ \epsilon_{ij} \} \), with \( \text{Var}(A_t) \equiv \sigma_A^2 \). Then \( \{ \mu_t \} \) is a mean zero covariance stationary series with covariance function
\[
\gamma(h) = \sigma(t+h,t) = \sigma_A^2 \sum_{k=0}^{\infty} \psi_k \psi_{k+h}.
\]
From \( \sum_{k=0}^{\infty} |\psi_k| < \infty \), it follows that \( \gamma(h) \to 0 \) as \( h \to \infty \); hence, \( \kappa_n \) is invertible for all \( n \geq 1 \) if \( \psi_k \neq 0 \) for some \( k \geq 0 \). When \( n_t \equiv n \), \( \{ \hat{X}_{11}, \hat{X}_{22}, \ldots \} \) will be stationary with covariance function
\[
\kappa(t,s) = \text{Cov}(\hat{X}_t, \hat{X}_s) = \gamma(t-s) + n^{-1}\sigma_e^2(t=s),
\]
hence, computations can be simplified by employing the Durbin- Levinson algorithm (cf. Brockwell and Davis, 1991) rather than the innovations algorithm.
4. Simulations. This section investigates the performance of the innovations control chart via simulation. To study a variety of correlation structures, we will consider AR(1), ARMA(1,1), and random walk models for \( \{ \mu_t \} \). The results will be compared to the case where a standard X-bar chart is applied to correlated subgroups without accounting for subgroup correlations (see Padgett et al., 1992).

To summarize control chart performance, we will use the \( \alpha_k \)-risk of Padgett et al. (1992). The \( \alpha_k \)-risk is defined as the probability of at least one out of control signal in \( k \) total subgroups when the process is actually in control:

\[
\alpha_k = P[ | Z_t(n) | > 3 \text{ for at least one } t \text{ satisfying } 1 \leq t \leq k ].
\] (4.1)

Although there are numerous other ways of summarizing control chart performance, we have selected \( \alpha_k \)-risks so that we can compare our results to those in Padgett et al. (1992). We also note that \( \alpha_k \) can be interpreted as a Type I error probability. For a Gaussian process where parameters are known, \( \alpha_k \) has a binomial form and does not depend on \( n \):

\[
\alpha_k = 1 - (1 - \alpha_1)^k,
\] (4.2)

where \( \alpha_1 = P[ | Z | > 3 ] \approx 0.0026998 \) is the probability that the absolute value of a standard normal deviate \( Z \) exceeds 3. Of course, \( \alpha_k \to 1 \) as \( k \to \infty \). Table 1 below lists several values of \( \alpha_k \).

<table>
<thead>
<tr>
<th>Table 1</th>
<th>( \alpha_k )-risks.</th>
</tr>
</thead>
<tbody>
<tr>
<td>( k )</td>
<td>( \alpha_k )</td>
</tr>
<tr>
<td>20</td>
<td>0.05263</td>
</tr>
<tr>
<td>30</td>
<td>0.07790</td>
</tr>
<tr>
<td>50</td>
<td>0.12643</td>
</tr>
<tr>
<td>100</td>
<td>0.23688</td>
</tr>
</tbody>
</table>

Our first two simulations, summarized in Tables 2 and 3 below, pertain to the stationary AR(1) model for \( \{ \mu_t \} \) in (3.1). In this and all subsequent simulations, we will examine the subgroup lengths \( k = 20, 30, 50, \) and 100, and subgroup sizes \( n_t \equiv 3, 4, 5, \) and 6. These choices of \( n_t \) and \( k \) are relatively consistent with sample sizes encountered in applications. The autoregressive parameter \( \phi \) was chosen as 0.5 and \(-0.5\) in Tables 2 and 3 respectively; this \( |\phi| \) is viewed as injecting a moderate degree of correlation in the subgroups. The values of \( \sigma_A^2 \) and \( \sigma_c^2 \) were selected to make \( V\sigma\left( \bar{X}_t \right) = 1 \)
when \( n = 1 \). In Tables 2 and 3, the upper risk is the sample \( \alpha_k \)-risk for the control chart using the estimated standardized prediction residuals \( \{ Z_t(\hat{\mu}) \} \) and the lower risk is the sample \( \alpha_k \)-risk for the control chart using the standardized prediction residuals \( \{ Z_t(\eta) \} \) where the value of \( \eta \) is assumed known. All sample risks are computed from 2000 simulations; specifically, the sample \( \alpha_k \)-risks are the proportion of simulations where a standardized prediction residual, \( Z_t(\hat{\mu}) \) or \( Z_t(\eta) \), exceeds 3 in absolute value for at least one \( t \) satisfying \( 1 \leq t \leq k \). ARMA model order selection was neglected in the likelihood computations with \( \{ Z_t(\hat{\mu}) \} \).

We first note that the sample \( \alpha_k \)-risks computed using the known value of \( \eta \) are very close to their exact values in Table 1. Hence, the chart is functioning well when model parameters are known. We contrast this to Tables 6 – 10 of Padgett et al. (1992) which show that, even when the exact \( \eta \) is used, the sample \( \alpha_k \)-risks are close to unity when a traditional \( X \)-bar chart, unmodified for subgroup correlations, is used. Hence, in this case, the innovations chart clearly improves on a traditional \( X \)-bar chart that neglects subgroup correlations.

We next note that the sample \( \alpha_k \)-risks computed from \( \{ Z_t(\hat{\mu}) \} \) are consistently smaller than the sample \( \alpha_k \)-risks computed from \( \{ Z_t(\eta) \} \). We attribute this property to a slight negative covariance in the residuals \( \{ Z_t(\hat{\mu}) \} \) due to parameter estimation with a small number of subgroups \( k \). Elaborating, if \( Z_t(\hat{\mu}) \) and \( Z_{t+1}(\hat{\mu}) \) have a negative correlation, then a value of \( Z_t(\hat{\mu}) \) around the upper control boundary \(+3\) is more likely to be followed by a \( Z_{t+1}(\hat{\mu}) \) that is significantly less than \(+3\); a similar effect holds around the lower boundary \(-3\). Hence, a negative correlation would act to reduce the number of observations exceeding control boundaries and would lower sample \( \alpha_k \)-risks.

To investigate the correlations in \( \{ Z_t(\hat{\mu}) \} \) due to parameter estimation, a separate simulation was performed with the \( n = 3 \) and \( k = 20 \) entry in Table 2. Fifty thousand sets of \( \{ Z_t(\hat{\mu}) \} \) were simulated and the sample mean and lag one sample correlations of \( \{ Z_t(\hat{\mu}) \} \) were computed in each simulation. A 95% large-sample confidence interval for the mean residual was \( 3.95 \times 10^{-4} \pm 1.96 \times 10^{-3} \); hence, the standardized prediction residuals appear to have a zero mean. In contrast, a 95% large-sample confidence interval for the lag one correlation was \(-0.04957 \pm 1.81 \times 10^{-3}\) which supports the
conjectured negative lag one correlation. Of course, if the correlation structure of \( \{Z_t(\tilde{\eta})\} \) for small \( k \) was known, this information could be used to “further standardize” \( \{Z_t(\tilde{\eta})\} \) into an iid sequence; unfortunately, we know of no such small sample results for Kalman filter residuals.

Despite the slight bias due to parameter estimation, the sample \( \alpha_k \)-risks computed from \( \{Z_t(\tilde{\eta})\} \) are much closer to the values reported in Table 1 than the values reported in Padgett et al. (1992) where subgroup correlations were ignored. Furthermore, Tables 2 and 3 show that as the series length \( k \) increases, the two sample \( \alpha_k \)-risks become closer; as discussed in Section 2, this is because as \( k \to \infty \), \( \tilde{\eta} \to \eta \) in probability and \( Z_t(\tilde{\eta}) \to Z_t(\eta) \) in probability for each fixed \( t \).

<table>
<thead>
<tr>
<th>( \text{Table 2} )</th>
<th>AR(1) ( \alpha_k )-risks</th>
<th>( \phi = 1/2, \sigma^2_{\varepsilon} = 1/2, \sigma^2_A = 3/8 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n = 3 )</td>
<td>20  30  50  100</td>
<td>( n = 3 )</td>
</tr>
<tr>
<td></td>
<td>.0160  .0485  .1060  .2250</td>
<td>( n = 3 )  .0225  .0395  .0945  .2225</td>
</tr>
<tr>
<td></td>
<td>.0435  .0830  .1255  .2365</td>
<td>( n = 4 )  .0215  .0470  .1055  .2125</td>
</tr>
<tr>
<td></td>
<td>.0520  .0730  .1410  .2470</td>
<td>( n = 5 )  .0165  .0495  .0965  .2300</td>
</tr>
<tr>
<td></td>
<td>.0170  .0505  .0880  .2535</td>
<td>( n = 6 )  .0155  .0615  .0980  .2170</td>
</tr>
<tr>
<td></td>
<td>.0575  .0855  .1210  .2650</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.0480  .0840  .1185  .2355</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( \text{Table 3} )</th>
<th>AR(1) ( \alpha_k )-risks</th>
<th>( \phi = -1/2, \sigma^2_{\varepsilon} = 1/2, \sigma^2_A = 3/8 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n = 3 )</td>
<td>20  30  50  100</td>
<td>( n = 3 )  .0225  .0395  .0945  .2225</td>
</tr>
<tr>
<td></td>
<td>.0465  .0770  .1120  .2290</td>
<td></td>
</tr>
<tr>
<td>( n = 4 )</td>
<td>.0215  .0470  .1055  .2125</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.0505  .0750  .1265  .2215</td>
<td></td>
</tr>
<tr>
<td>( n = 5 )</td>
<td>.0165  .0495  .0965  .2300</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.0605  .0755  .1335  .2410</td>
<td></td>
</tr>
<tr>
<td>( n = 6 )</td>
<td>.0155  .0615  .0980  .2170</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.0540  .0910  .1225  .2285</td>
<td></td>
</tr>
</tbody>
</table>

In a similar format to Tables 2 and 3, Tables 4 and 5 report sample \( \alpha_k \)-risks aggregated from 2000 simulations with the ARMA(1,1) model

\[
\mu_t - \phi \mu_{t-1} = A_t + \theta A_{t-1},
\]

where \( \{A_t\} \) is mean zero Gaussian white noise with \( \text{Var}(A_t) \equiv \sigma^2_A \) that is independent of \( \{\epsilon_t\} \). The white noise variances were taken as \( \sigma^2_{\varepsilon} = \sigma^2_A = 1 \) in both tables; the ARMA(1,1) parameters are chosen as \( \phi = 2/3 \) and \( \theta = 1/3 \) in Table 4 and \( \phi = -1/3 \) and \( \theta = 2/3 \) in Table 5. We note that Tables 4 and 5 have a similar structure to Tables 2 and 3 and omit further discussion.
Table 4  ARMA(1,1) $\alpha_k$-risks
\[ \phi = 2/3, \theta = 1/3, \sigma_e^2 = 1, \sigma_A^2 = 1. \]

<table>
<thead>
<tr>
<th>Number of Subgroups $k$</th>
<th>20</th>
<th>30</th>
<th>50</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 3$</td>
<td>.0210</td>
<td>.0470</td>
<td>.1040</td>
<td>.2170</td>
</tr>
<tr>
<td></td>
<td>.0600</td>
<td>.0775</td>
<td>.1175</td>
<td>.2395</td>
</tr>
<tr>
<td>$n = 4$</td>
<td>.0205</td>
<td>.0555</td>
<td>.1005</td>
<td>.2265</td>
</tr>
<tr>
<td></td>
<td>.0460</td>
<td>.0800</td>
<td>.1255</td>
<td>.2545</td>
</tr>
<tr>
<td>$n = 5$</td>
<td>.0245</td>
<td>.0510</td>
<td>.0940</td>
<td>.2015</td>
</tr>
<tr>
<td></td>
<td>.0535</td>
<td>.0760</td>
<td>.1145</td>
<td>.2390</td>
</tr>
<tr>
<td>$n = 6$</td>
<td>.0245</td>
<td>.0425</td>
<td>.1005</td>
<td>.2125</td>
</tr>
<tr>
<td></td>
<td>.0490</td>
<td>.0780</td>
<td>.1190</td>
<td>.2435</td>
</tr>
</tbody>
</table>

Table 5  ARMA(1,1) $\alpha_k$-risks
\[ \phi = -1/3, \theta = 2/3, \sigma_e^2 = 1, \sigma_A^2 = 1. \]

<table>
<thead>
<tr>
<th>Number of Subgroups $k$</th>
<th>20</th>
<th>30</th>
<th>50</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 3$</td>
<td>.0260</td>
<td>.0555</td>
<td>.1045</td>
<td>.2195</td>
</tr>
<tr>
<td></td>
<td>.0610</td>
<td>.0765</td>
<td>.1190</td>
<td>.2460</td>
</tr>
<tr>
<td>$n = 4$</td>
<td>.0170</td>
<td>.0495</td>
<td>.1095</td>
<td>.2290</td>
</tr>
<tr>
<td></td>
<td>.0440</td>
<td>.0750</td>
<td>.1300</td>
<td>.2520</td>
</tr>
<tr>
<td>$n = 5$</td>
<td>.0185</td>
<td>.0505</td>
<td>.0975</td>
<td>.2025</td>
</tr>
<tr>
<td></td>
<td>.0525</td>
<td>.0775</td>
<td>.1090</td>
<td>.2230</td>
</tr>
<tr>
<td>$n = 6$</td>
<td>.0225</td>
<td>.0435</td>
<td>.0980</td>
<td>.2115</td>
</tr>
<tr>
<td></td>
<td>.0505</td>
<td>.0840</td>
<td>.1260</td>
<td>.2405</td>
</tr>
</tbody>
</table>

Table 6 below presents an analogous simulation for the nonstationary random walk model for \{\mu_t\} in Section 3.2. The random walk has strong correlations between neighboring subgroups. The parameters for this simulation were selected as \( \sigma_e^2 = \sigma_A^2 = 1 \); for these parameter choices, (3.10) gives $Corr(X_{tj}, X_{t-1,i}) = (t - 1)/[(t + 1)t]^{1/2}$. Table 6 shows a similar structure to Tables 2-5.

**TABLE 6**  Random Walk $\alpha_k$-risks
\[ \sigma_e^2 = 1, \sigma_A^2 = 1. \]

<table>
<thead>
<tr>
<th>Number of Subgroups $k$</th>
<th>20</th>
<th>30</th>
<th>50</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 3$</td>
<td>.0190</td>
<td>.0495</td>
<td>.0995</td>
<td>.2225</td>
</tr>
<tr>
<td></td>
<td>.0515</td>
<td>.0810</td>
<td>.1295</td>
<td>.2235</td>
</tr>
<tr>
<td>$n = 4$</td>
<td>.0225</td>
<td>.0450</td>
<td>.1030</td>
<td>.2165</td>
</tr>
<tr>
<td></td>
<td>.0450</td>
<td>.0755</td>
<td>.1205</td>
<td>.2440</td>
</tr>
<tr>
<td>$n = 5$</td>
<td>.0270</td>
<td>.0495</td>
<td>.1015</td>
<td>.2120</td>
</tr>
<tr>
<td></td>
<td>.0640</td>
<td>.0820</td>
<td>.1270</td>
<td>.2295</td>
</tr>
<tr>
<td>$n = 6$</td>
<td>.0220</td>
<td>.0465</td>
<td>.0950</td>
<td>.2230</td>
</tr>
<tr>
<td></td>
<td>.0475</td>
<td>.0880</td>
<td>.1250</td>
<td>.2380</td>
</tr>
</tbody>
</table>

Overall, the above simulations show that the innovations control chart behaves as a standard $X$-bar chart with uncorrelated subgroups when the correlation structure of the subgroups is known.

When the subgroup correlation structure must be estimated from a sample of subgroups, the sample
\( \alpha_k \)-risks are slightly smaller than their true values but, nonetheless, improve greatly over values reported in Padgett et al. (1992) when subgroup correlations are ignored. Of course, the bias in the \( \alpha_k \)-risks will be asymptotically negligible when subgroup correlation parameters are estimated consistently.

5. Examples.

5.1 Motor Shaft Diameter Data. In this subsection, we apply the innovations chart to the motor shaft diameter data in Devor et al. (1992, pp. 186-188). This data has process subgroups of size \( n_s = 5 \) observed at \( k = 60 \) sampling times separated by 30 minutes each. Figure 1 plots the sample mean motor shaft diameter at each sampling time; the traditional \( X \)-bar chart upper and lower control limits, \( \bar{X} \pm A_2 \bar{R} \), computed assuming independent subgroups, are included for reference. Notice that process sample means fall outside of the control boundaries for subgroups 28 and 32; hence, a traditional \( X \)-bar chart would signal "out of control" for this process twice.

An exploratory analysis of the subgroup data suggests that a stationary model for \( \{ \mu_t \} \) is appropriate; hence, we will use an ARMA\((p,q)\) model to describe \( \{ \mu_t \} \):

\[
\mu_t - \sum_{i=1}^{p} \phi_i \mu_{t-i} = A_t + \sum_{i=1}^{q} \theta_i A_{t-i},
\]  

(5.1)

where \( \{ A_t \} \) is white noise with \( \text{Var}(A_t) \equiv \sigma_A^2 \) that is independent of \( \{ \epsilon_t \} \). Minimizing the scaled negative log likelihood in (2.13) for various \( p \) and \( q \), and using AICC to select \( p \) and \( q \), we select the AR(1) model in Section 3.1 for \( \{ \mu_t \} \). An overall sample mean of 48.793, computed as the average of all observations, was first subtracted from the measurements to obtain mean zero data. The AR(1) model estimates obtained were \( \hat{\phi}_1 = 0.741 \pm 0.106 \), and \( \hat{\sigma}_A^2 = 2.693 \). An error variance of \( \hat{\sigma}_\epsilon^2 = 39.366 \) and a negative log likelihood of 159.956 were obtained in the model fit; an AICC statistic of 326.340 was achieved. The error margin attached to \( \hat{\phi}_1 \) is one standard error and was computed by inverting the observed information matrix associated with the negative log likelihood in (2.12). The estimate of \( \phi_1 \) and its standard error strongly suggest that \( \phi_1 \) is positive; hence, we conclude that the process subgroups are positively correlated.

For confirmation of the adequacy of the AR(1) model, we will examine \( \{ Z_t(\bar{R}) \} \). Figure 2
plots $Z_t(\tilde{\mu})$ against $t$ along with $\pm 3$ control limits. If the fitted model is good, then $\{Z_t(\tilde{\mu})\}$ should have correlation properties that are similar to those of an uncorrelated sequence. Figure 3 plots the sample autocorrelation function of $\{Z_t(\tilde{\mu})\}$ over the first 20 lags; the dashed lines are 95% confidence limits for iid sequences and suggest that our fitted model is adequate. The estimated standardized residuals also passed several commonly applied tests for normality (cf. Section 9.4 of Brockwell and Davis, 1991); hence, the Gaussian AR(1) model for $\{\mu_t\}$ appears to be reasonable.

The $Z_t(\tilde{\mu})$ values in Figure 2 all lie inside the control limits $\pm 3$; hence, contrary to a traditional X-bar chart, the innovations chart suggests that the process is never out of control. Finally, we note that $Z_t(\tilde{\mu})$ follows a similar pattern in $t$ as the sample means in Figure 1, but have a smaller variability; hence, one can directly see the effects of accounting for subgroup correlations.

5.2 A Climate Example. In this section, we apply the innovations control chart to monthly average temperature data from Boston, MA, New York City, NY, and Washington, DC ($n_t \equiv 3$) for the 117 years 1871 – 1987 inclusive. The methods used here will identify excessively hot and cold months during the data record and will provide criteria for assessing their occurrences in the future.

To develop a control chart, the mean and correlation structure of the data must first be explored. Figures 4-6 plot the data from each city and clearly show a periodic cycle in mean levels at each city. Many temperature series also have periodicities in their autocovariance structure; a test to detect such periodicities based on averaged squared coherences is discussed in Lund et al. (1995) and was applied to all three series. Figures 7-9 summarize this test by plotting average squared coherences against lag along with a 99% confidence threshold after a linear trend and periodic sample means, discussed further below, were removed from each series to obtain mean zero data. A smoothing width of $M = 8$ was used to compute all coherences. The large “spike” at lag $h = 117$ in each average squared coherence plot indicates that strong periodicities exist in the correlation structure of each series.

We comment that if $\{X_{mT+n,j}\}$ is stationary, then the average squared coherence plot for city $j$ should have no large averaged squared coherences.
To model the periodicities in the first two moments of each series, and as an attempt towards obtaining mean zero stationary data, we standardize each series in a seasonal fashion by subtracting a monthly mean and dividing by a monthly standard deviation. Further allowing for a linear trend, and adopting a seasonal notation, we will apply the innovations control chart to the detrended and seasonally standardized series

\[ X^*_{mT+\nu,j} = \frac{X_{mT+\nu,j} - \mu_{\nu,j} - \beta_j(mT + \nu)}{\sigma_{\nu,j}}. \]  

(5.2)

In (5.2), the indices \( j = 1, 2, \text{ and } 3 \) refer to Boston, MA, New York City, NY, and Washington, DC respectively. The seasonal subscripting in (5.2) makes the following conventions: \( T = 12 \) is the period of the data and \( \nu \) is a monthly index that satisfies \( 1 \leq \nu \leq T \); hence, \( X_{mT+\nu,j} \) is the average monthly temperature in city \( j \) during the \( \nu \)th month of year \( m \). We note that the linear trend parameter \( \beta_j \) is allowed to change from city to city. The parameters \( \mu_{\nu,j} \) and \( \sigma_{\nu,j} \) are the detrended mean and standard deviation of the temperature at city \( j \) during month \( \nu \).

The parameters in (5.2) can be estimated directly from the subgroup data. First, we estimate \( \beta_j \) with

\[ \hat{\beta}_j = \frac{\sum_{m=0}^{d-1} \sum_{\nu=1}^{T} (X_{mT+\nu,j} - \bar{X}_j)(mT + \nu - \bar{t})}{\sum_{i=1}^{N} (i - \bar{t})^2}. \]  

(5.3)

In (5.3), \( N = 1404 \) is the length of each series, \( d = N/T = 117 \) is the number of years of data in the study, \( \bar{X}_j \) is the average of all data points for city \( j \), and \( \bar{t} = (N+1)/2 \) is the average time index.

The seasonal means are estimated with the trend adjusted seasonal sample averages

\[ \hat{\mu}_{\nu,j} = \frac{1}{d} \sum_{m=0}^{d-1} [X_{mT+\nu,j} - \hat{\beta}_j(mT + \nu)], \]  

(5.4)

and the seasonal variances estimates take the customary trend adjusted form

\[ \hat{\sigma}_{\nu,j} = \sqrt{\frac{d^{-1} \sum_{m=0}^{d-1} (X_{mT+\nu,j} - \hat{\beta}_j(mT + \nu) - \hat{\mu}_{\nu,j})^2}. \]  

(5.5)

Using estimated parameters in place of their true values in (5.2), estimated versions of \( X^*_{mT+\nu,j} \), denoted by \( \hat{X}^*_{mT+\nu,j} \), were computed and rechecked for a periodic covariance structure (it is clear that
(5.2) eliminates periodicities in the mean) with the averaged squared coherence test. Figures 10-12 summarize the results of this test, again with \( M = 8 \), and indicate that the large average squared coherence at lag \( h = 117 \) has been removed in each city; hence, \( \{ \hat{X}_{mT+\nu,j}^* \} \) appears to be a mean zero, unit variance, stationary series for each \( j \). We remark that seasonal standardizations frequently eliminate periodicities in temperature series; for further results and discussion on this point, see Lund et al. (1995).

The next step in constructing a control chart involves model development for \( \{ \hat{X}_{mT+\nu,j}^* \} \). Since \( \{ \hat{X}_{mT+\nu,j}^* \} \) is approximately stationary, we use the ARMA\((p,q)\) model in (5.1) to describe \( \{ \mu_t \} \). Because the seasonal standardization in (5.2) leaves \( \hat{X}_{mT+\nu,j} \) with an approximate unit variance, we impose the parametric restriction \( \text{Var}(\mu_t) + \sigma_e^2 = 1 \) in all likelihood computations. Optimizing (2.13) and using AICC to select \( p \) and \( q \), we select the ARMA\((1,1)\) model for \( \{ \mu_t \} \). We comment that ARMA model causality and invertibility were also enforced in the likelihood optimizations. The parameter estimates obtained were \( \hat{\beta}_1 = 0.461 \pm 0.087, \hat{\delta}_1 = -0.271 \pm 0.096, \hat{\sigma}_1 = 0.751, \) and \( \hat{\sigma}_e = 0.214 \); a negative log likelihood of \( 1857.510 \) and an AICC statistic of \( 3721.036 \) were obtained in the model fit. Attempts to add time invariant cross correlations between \( \{ \epsilon_{t1} \}, \{ \epsilon_{t2} \}, \) and \( \{ \epsilon_{t3} \} \) were pursued but proved to be insignificant.

Figure 13 plots \( \{ Z_t(\hat{\mu}) \} \) computed from \( \{ \hat{X}_{mT+\nu,j}^* \} \) in (5.2) against the control limits of \( \pm 3 \). There are three exceedances (3.791 were expected) of the \( \pm 3 \) control limits during the period of record: a hot May of 1880, a cold June of 1903, and a cold February of 1934. We comment that the \( \pm 3 \) control limits are perhaps much more stringent than the public's perception of a heat wave or cold snap; control limits of \( \pm 1.96 \) would correspond to a 5\% chance of a heat wave or cold snap occurring in any month. Although we will not list them here, there were 67 exceedances of the \( \pm 1.96 \) control limits (70.200 were expected).

For a final diagnostic check, Figure 14 plots the sample autocorrelation function of \( \{ Z_t(\hat{\mu}) \} \) and suggests no departures from white noise. The \( \{ Z_t(\hat{\mu}) \} \) sequence also passed several tests for
normality; this is not surprising as the series consists of temperatures averaged from daily values.

6. Comments. If the normality assumptions are violated, then conditional expectations may not be linear in the predicting variables and the innovations prediction theory presented in Section 2 may not yield the best mean squared error prediction of \( \bar{X}_{t+1} \) from \( \bar{X}_1, \ldots, \bar{X}_t \). We refer the interested reader to Section 13.4 of Brockwell and Davis (1991) for comments pertaining to when linear prediction is "good" in terms of mean squared error. However, we note that even if linear prediction is good, the ± 3 control limits may need to be adjusted as the distribution of \( Z_t(\theta) \) is not necessarily standard Gaussian.

Since \( \text{Corr}(X_{ij}, X_{il}) = [\sigma(t, t) + \sigma_i^2]^{-1} \sigma(t, t) \geq 0 \) when \( j \neq l \), the subgroup \( \{X_{ij}, 1 \leq j \leq n_t\} \) will contain positively correlated measurements for any fixed \( t \). This seems physically plausible as, for example, with the temperature data in Section 5.2, one expects positive pairwise correlations between \( \{X_{i1}\}, \{X_{i2}\}, \) and \( \{X_{i3}\} \) because of the close geographical proximity of the three cities in the study. Nonetheless, the model in (2.1) would be inadequate when correlations between items in a subgroup are negative (we note that negative correlations between items in different subgroups would be fine).

In practice, one may wish to use a small pilot sample of subgroups rather than the whole data set to estimate subgroup correlation parameters – we are not advocating using all subgroups in all cases. Of course, with the temperature data in Section 5.2, one step ahead predictions are frequently of interest by themselves; here, it is natural to use all available information to assess future temperatures.

Acknowledgement. Robert Lund thanks Richard Davis for useful conversations on the subject matter. I. V. Basawa's research was partially supported by grants from the ONR and the NSF; W. J. Padgett was supported in part by NSF grant DMS9503104.
BIBLIOGRAPHY


19


Figure 10: Average Squared Coherences for Boston – Seasonally Standardized Data

Figure 11: Average Squared Coherences for NYC – Seasonally Standardized Data

Figure 12: Average Squared Coherences for DC – Seasonally Standardized Data