Fitting Time-Series Input Processes for Simulation

Bahar Biller
Tepper School of Business, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, billerb@andrew.cmu.edu

Barry L. Nelson
Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois 60208, nelsonb@northwestern.edu

Providing accurate and automated input-modeling support is one of the challenging problems in the application of computer simulation of stochastic systems. The models incorporated in current input-modeling software packages often fall short because they assume independent and identically distributed processes, even though dependent time-series input processes occur naturally in the simulation of many real-life systems. Therefore, this paper introduces a statistical methodology for fitting stochastic models to dependent time-series input processes. Specifically, an automated and statistically valid algorithm is presented to fit autoregressive-to-anything (ARTA) processes with marginal distributions from the Johnson translation system to stationary univariate time-series data. ARTA processes are particularly well suited to driving stochastic simulations. The use of this algorithm is illustrated with examples.

Subject classifications: simulation, statistical analysis: stochastic input modeling; statistics, correlation, estimation, time series: autoregressive processes, least-squares fitting.
Area of review: Simulation.
History: Received May 2003; revision received January 2004; accepted April 2004.

1. Introduction

Dependent time-series input processes occur naturally in the simulation of many service, communications, and manufacturing systems. For example, Melamed et al. (1992) observe autocorrelation in sequences of compressed video frame bit rates, while Ware et al. (1998) report that the times between file accesses on a computer network frequently exhibit burstiness, as characterized by a sequence of short interaccess times followed by one or more long ones. Because there is no widely available, general-purpose method for fitting dependent time-series input processes, simulation practitioners typically ignore these dependencies and develop input models using software packages that are appropriate for independent and identically distributed (i.i.d.) data. Unfortunately, ignoring these dependencies while developing input models for stochastic simulation can lead to very poor estimates of performance measures. An illustration is given by Livny et al. (1993), who examine the impact of autocorrelation on the performance of single-server queues.

Much of the previous work on dependent time-series input processes is based on linear models, such as the autoregressive moving average class or those that underlie Kalman filtering and related methods (Chatfield 1999). Mallows (1967) shows that the linearity of these models imply normal marginal distributions, but there are many physical situations in which the marginals of the time-series input processes are nonnormal. Motivated by this problem, there has been considerable research on modeling time series with marginals from nonnormal families, such as exponential, gamma, geometric, or general discrete marginal distributions (see, for example, Lewis et al. 1989). However, these models often allow only limited control of the dependence structure, and a different model is required for each type of marginal distribution.

A way to overcome these limitations is to construct the desired process by a monotone transformation of a Gaussian linear process. For example, Cario and Nelson (1996, 1998) take this approach to develop models for representing and generating stationary univariate time-series processes. The central idea is to transform a Gaussian autoregressive process into the desired univariate time-series input process that they presume as having an autoregressive-to-anything (ARTA) distribution. The authors manipulate the autocorrelations of the corresponding Gaussian process so that they achieve the desired autocorrelations for the simulation input process. They assume—as is common in the simulation input-modeling literature—that the desired marginal distribution and dependence structure (specified via autocorrelations) are given. However, there is no rigorously justified method for fitting the input model when only raw data generated by an unknown process are available. To fill this gap, we attack the problem of fitting stochastic input models to stationary univariate time-series data and suggest a new method that we call the ARTAFIT method in the

remainder of the paper. We specifically fit ARTA processes with marginal distributions from the Johnson translation system. While the focus on the Johnson translation system might seem restrictive, it is less so than it first appears: In many applications, simulation output performance measures are insensitive to the specific input distribution chosen, provided that enough moments of the distribution are correct (see, for instance, Gross and Juttijudata 1997). The Johnson translation system can match any feasible finite first four moments, while the standard families incorporated in existing software packages and simulation languages often match only one or two moments. Thus, the Johnson translation system enables us to represent key features of the unknown source of the data that have significant impact on the simulation output performance measures. Although our ARTAFIT approach is particularly compatible with the Johnson translation system, our results can be extended to other continuous distribution.

The remainder of the paper is organized as follows: In §2, we provide the ARTAFIT algorithm, prove its convergence properties, and show the statistical properties of the resulting estimators as the sample size approaches infinity. We provide illustrative examples demonstrating the use of the algorithm in §3 and conclude with the expected impact of this new statistical methodology on the development of stochastic simulation inputs in §4.

2. Fitting ARTA Models

In this section, we present the framework for fitting stochastic models to dependent time-series input processes. We also provide the theory that supports the ARTAFIT framework; the proofs are available in Biller and Nelson (2004).

2.1. Notation

We let the generic univariate input random variable be denoted by X, with marginal cumulative distribution function (cdf) F_X. The cdf of the standard normal distribution is denoted by Φ and its probability density function by ϕ. The mean of a random variable is denoted by μ and its variance by σ^2.

A stationary univariate time-series input process is denoted by $\{X_t; t = 1, 2, \ldots\}$. The term "time series" means that the random variables may be dependent in sequence, such as the month-to-month orders for a product placed by a customer. We denote any realization of length n from the input process X_t by $\{x_t; t = 1, 2, \ldots, n\}$. Boldface type is used to denote column vectors, e.g., $\mathbf{x} = (x_1, x_2, \ldots, x_n)'$.

We account for dependence between random variables that are lag-h apart, say X_t and X_{t-h}, via their product-moment correlation defined as $\rho_X(h) = E[\sigma^{-2}(X_t - \mu) \cdot (X_{t-h} - \mu)]$, where X_t has mean μ and variance σ^2 for all t due to the assumption of a stationary input process. Representation of dependence by product-moment correlation is a practical compromise we make in this paper. Many other measures of dependence have been defined

(see Nelsen 1998), and they are arguably more informative than the product-moment correlation for some distributions. However, product-moment correlation is the only measure of dependence that is widely used and understood in engineering applications. We believe that making it possible for simulation users to incorporate dependence by product-moment correlation, while limited, is substantially better than ignoring dependence. Further, the ARTA model is flexible enough to incorporate dependence measures that remain unchanged under strictly increasing transformations of the random variables, such as Spearman's rank correlation and Kendall's τ, should those measures be desired.

2.2. The ARTAFIT Model

We are particularly interested in input-modeling problems in which data are plentiful and nearly automated input modeling is required. Consequently, we use a member of the Johnson translation system to characterize the marginal distribution of the input process. A robust method for fitting target distributions from the Johnson translation system to i.i.d. data is suggested by Swain et al. (1988) and implemented in software called FITTR1. They demonstrate the robustness and computational efficiency of least-squares, minimum L_1 norm, and minimum L_∞ norm techniques for estimating Johnson-type marginals. We believe that similar techniques can be effectively adapted to fitting ARTA models to dependent time-series data. We outline our approach below.

Let $\{X_t; t = 1, 2, \ldots, n\}$ denote a stationary univariate time-series input process. The goal is to approximate $\{X_t; t = 1, 2, \ldots, n\}$ by an ARTA process whose complete specification is given by

$$X_t = F_X^{-1}[\Phi(Z_t)], \quad t = 1, 2, \ldots, n, \tag{1}$$

where the base process $\{Z_t; t = 1, 2, \ldots, n\}$ is a stationary, standard Gaussian autoregressive process of order p (denoted by AR(p)) with the representation

$$Z_t = \sum_{h=1}^{p} \alpha_h Z_{t-h} + Y_t, \quad t = p+1, p+2, \ldots, n.$$

The α_h, $h = 1, 2, \ldots, p$, are fixed autoregressive coefficients that uniquely determine the autocorrelation structure of the base process, $\rho_Z(h)$, $h = 1, 2, \ldots, p$, and Y_t, $t = p+1, p+2, \ldots, n$, are i.i.d. Gaussian random variables with mean zero and variance σ_Y^2. This modeling approach works for *any* marginal distribution, although F_X^{-1} may have to be evaluated by an approximate numerical method when there is no exact closed-form expression. The inverse cdf method is an essential ingredient of the ARTA framework.

In this paper, we focus on processes having marginal distributions from the Johnson translation system defined by a cdf of the form

$$F_X(x) = \Phi\left\{\gamma + \delta f\left[\frac{x - \xi}{\lambda}\right]\right\}, \tag{2}$$

where γ and δ are shape parameters, ξ is a location parameter, λ is a scale parameter, and $f(\cdot)$ is one of the following transformations:

$$f(y) = \begin{cases} \log(y) & \text{for the } S_L \text{ (lognormal) family,} \\ \log\left(y + \sqrt{y^2 + 1}\right) \\ & \text{for the } S_U \text{ (unbounded) family,} \\ \log\left(\frac{y}{1-y}\right) & \text{for the } S_B \text{ (bounded) family,} \\ y & \text{for the } S_N \text{ (normal) family.} \end{cases}$$

This system can match any feasible finite first four moments by the manipulation of the parameters f, γ, δ, λ, and ξ. Although the first four moments of all distributions in the families S_L, S_B, S_U, and S_N are finite, the ability to match any (finite) first four moments provides a great deal of flexibility that is sufficient for many practical problems. Distributional properties of the marginals from the Johnson translation system can be found in Johnson (1949).

Clearly, fitting an ARTA process to data corresponds to the estimation of f, γ, δ, λ, ξ, p, α_h for $h = 1, 2, \ldots, p$, and σ_Y^2. However, the value of σ_Y^2 is completely determined by the autoregressive coefficients α_h, $h = 1, 2, \ldots, p$, because we select $\sigma_Y^2 = 1 - \sum_{h=1}^{p} \alpha_h \rho_Z(h)$ to force the base process Z_t to have variance 1, and each base autocorrelation $\rho_Z(h)$, $h = 1, 2, \ldots, p$, can be written as a nonlinear function of the autoregressive coefficients α_h, $h = 1, 2, \ldots, p$, by using the Yule-Walker equations (Wei 1990). For example, $\rho_Z(h) = \alpha_1^h$, $h \geqslant 1$, when the order of the underlying base process is equal to one. Therefore, we will write $\sigma_Y^2 \equiv g(p, \boldsymbol{\alpha})^2$, where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_p)'$, and no longer consider σ_Y^2 as a parameter to be estimated. For ease of presentation of our ARTAFIT algorithm, we assume that the order of the underlying base process p and the type of the Johnson transformation f are known. Clearly, these also need to be determined, and we address this issue in §3. Also, from here on, when we say "ARTA process" we mean an ARTA process with a marginal distribution from the Johnson translation system.

Let $\boldsymbol{\psi}$ correspond to the vector of ARTA parameters, i.e., $\boldsymbol{\psi} = (\lambda, \delta, \gamma, \xi, \alpha_1, \alpha_2, \ldots, \alpha_p)'$, and consider the standardized white-noise process

$$V_t(\boldsymbol{\psi}) = \frac{Y_t}{g(p, \boldsymbol{\alpha})} = \frac{Z_t - \sum_{h=1}^{p} \alpha_h Z_{t-h}}{g(p, \boldsymbol{\alpha})},$$
$$t = p+1, p+2, \ldots, n.$$

If we further write the base random variable Z_t as a function of the input random variable X_t using (1) and (2), then we get the following expression for the standardized white-noise process:

$$V_t(\boldsymbol{\psi}) = \frac{\gamma + \delta f[(X_t - \xi)/\lambda] - \sum_{h=1}^{p} \alpha_h(\gamma + \delta f[(X_{t-h} - \xi)/\lambda])}{g(p, \boldsymbol{\alpha})},$$
$$t = p+1, p+2, \ldots, n. \quad (3)$$

Now, suppose that X_t is actually an ARTA process with the parameter vector $\boldsymbol{\psi}^*$. If we have all of the parameter values correct ($\boldsymbol{\psi} = \boldsymbol{\psi}^*$), then $V_t(\boldsymbol{\psi}^*)$, $t = p+1$, $p+2, \ldots, n$, are i.i.d. standard normal random variables. Thus, our ARTAFIT procedure searches for parameters that make $V_t(\boldsymbol{\psi})$, $t = p+1, p+2, \ldots, n$, appear to be such a sample by minimizing the following objective function:

$$\frac{1}{(n-p)^2} \sum_{t=p+1}^{n} \frac{(n-p+1)^2(n-p+2)}{(t-p)(n+1-t)}$$
$$\cdot \left(\Phi\{V_{(t)}(\boldsymbol{\psi})\} - \frac{t-p}{n-p+1} \right)^2, \quad (4)$$

where $V_{(p+1)}(\boldsymbol{\psi}) \leqslant V_{(p+2)}(\boldsymbol{\psi}) \leqslant \cdots \leqslant V_{(n)}(\boldsymbol{\psi})$ denote the order statistics corresponding to the random variables $V_t(\boldsymbol{\psi})$, $t = p+1, p+2, \ldots, n$. In the remainder of the section, we explain how the objective function (4) has been derived for the estimation of the ARTA parameters.

If $\boldsymbol{\psi} = \boldsymbol{\psi}^*$, then the transformed variate $R_{(t)}(\boldsymbol{\psi}^*) = \Phi\{V_{(t)}(\boldsymbol{\psi}^*)\}$ has the distribution of the tth order statistic in a random sample of size $n - p$ from the uniform distribution on the unit interval $(0, 1)$. Because $R_{(t)}(\boldsymbol{\psi}^*)$ has mean $\rho_t = (t-p)/(n-p+1)$ (Kendall and Stuart 1979), we can write $R_{(t)}(\boldsymbol{\psi}^*) = \rho_t + \varepsilon_t(\boldsymbol{\psi}^*)$ so that the $\{\varepsilon_t(\boldsymbol{\psi}^*)$; $t = p+1, p+2, \ldots, n\}$ are translated uniform order statistics with mean zero and covariance

$$\text{Cov}\left(\varepsilon_j(\boldsymbol{\psi}^*), \varepsilon_k(\boldsymbol{\psi}^*)\right) = \frac{\rho_j(1-\rho_k)}{n-p+2}, \quad p+1 \leqslant j \leqslant k \leqslant n \quad (5)$$

(Kendall and Stuart 1979). Let $\mathbf{R_o}(\boldsymbol{\psi}) \equiv (R_{(p+1)}(\boldsymbol{\psi})$, $R_{(p+2)}(\boldsymbol{\psi}), \ldots, R_{(n)}(\boldsymbol{\psi}))'$, $\boldsymbol{\rho} \equiv (\rho_{p+1}, \rho_{p+2}, \ldots, \rho_n)'$, and $\boldsymbol{\varepsilon}(\boldsymbol{\psi}) \equiv (\varepsilon_{p+1}(\boldsymbol{\psi}), \varepsilon_{p+2}(\boldsymbol{\psi}), \ldots, \varepsilon_n(\boldsymbol{\psi}))'$, so that $\boldsymbol{\varepsilon}(\boldsymbol{\psi}^*) \equiv \mathbf{R_o}(\boldsymbol{\psi}^*) - \boldsymbol{\rho}$. Because the first and second moments of the translated uniform order statistics are known and easily computed, we exploit this fact to develop a single, distribution-free formulation of the data-fitting problem. Specifically, we minimize the distance between $\boldsymbol{\rho}$ and $\mathbf{R_o}(\boldsymbol{\psi})$ as a function of $\boldsymbol{\psi}$ with respect to some metric defined by a quadratic form in the $(n - p)$-dimensional Euclidean space. If $\mathbf{W}$ is the $(n - p) \times (n - p)$ matrix associated with this quadratic form, then the parameter estimates can be obtained via least-squares fitting given by

$$\min_{\boldsymbol{\psi}} S_{\mathbf{W}}(\boldsymbol{\psi}) \equiv (\mathbf{R_o}(\boldsymbol{\psi}) - \boldsymbol{\rho})' \mathbf{W}(\mathbf{R_o}(\boldsymbol{\psi}) - \boldsymbol{\rho})$$
$$\equiv \boldsymbol{\varepsilon}(\boldsymbol{\psi})' \mathbf{W} \boldsymbol{\varepsilon}(\boldsymbol{\psi})$$

subject to $\boldsymbol{\psi} \in \boldsymbol{\Psi}$. $\quad (6)$

We define the feasible region $\boldsymbol{\Psi}$ as follows:

$$\boldsymbol{\Psi} = \Bigg\{ (\gamma, \delta, \lambda, \xi, \alpha_1, \alpha_2, \ldots, \alpha_p)' \colon \delta \geqslant 0$$

$$\lambda \begin{cases} \geqslant 0 & \text{for } f = S_U, \\ \geqslant X_{(n)} - \xi & \text{for } f = S_B, \\ = 1 & \text{for } f = S_L \text{ and } S_N, \end{cases}$$

$$\xi \begin{cases} \leqslant X_{(1)} & \text{for } f = S_L \text{ and } S_B, \\ = 0 & \text{for } f = S_N, \end{cases}$$

$$\left| \text{RootOf}\left(1 - \sum_{h=1}^{p} \alpha_h B^h = 0, B \right) \right| \geqslant 1 \Bigg\}, \qquad (7)$$

where the function "RootOf" is a place holder for representing all the roots of the equation $1 - \sum_{h=1}^{p} \alpha_h B^h = 0$ in the variable B. The first three constraints in (7) ensure the feasibility of the Johnson parameters, depending on the family of interest, and the last constraint ensures the stationarity of the autoregressive base process, and hence the stationarity of the input process. Although an ARTA process with parameter vector $\boldsymbol{\psi}$ is well defined only if $\boldsymbol{\psi}$ falls into the interior of the region $\boldsymbol{\Psi}$, the definition of the feasible region (7) ensures the convergence of the data-fitting algorithm presented in the next section. Later in §2.3, we discuss how the statistical validity of the ARTA-fitting algorithm is affected by a solution on the boundary of the feasible region.

The least-squares fitting problem defined in (6) gives rise to different estimators depending on the form of the weight matrix $\mathbf{W}$. When the weight matrix $\mathbf{W}$ is the $(n-p) \times (n-p)$ identity matrix $\mathbf{I}$, we obtain the ordinary least-squares estimators for $\boldsymbol{\psi}^*$; and when $\mathbf{W} \neq \mathbf{I}$, we obtain weighted least-squares (WLS) parameter estimators. WLS parameter estimators are of interest because $\{\varepsilon_t(\boldsymbol{\psi}^*); t = p+1, p+2, \ldots, n\}$ are neither independent nor homoscedastic. In addition, (5) shows that the matrix $\mathbf{W} = [\text{Cov}(\varepsilon_j(\boldsymbol{\psi}^*), \varepsilon_k(\boldsymbol{\psi}^*))]_{(n-p)\times(n-p)}^{-1}$ is readily computed. For the estimation of a linear model, it is well known that such a weight matrix yields the minimum variance linear unbiased estimator of the vector of model parameters (Seber 1977); and this suggests that we should also take $\mathbf{W} = [\text{Cov}(\varepsilon_j(\boldsymbol{\psi}^*), \varepsilon_k(\boldsymbol{\psi}^*))]_{(n-p)\times(n-p)}^{-1}$ while estimating the parameters of the nonlinear model (6). However, we choose instead to use the diagonal weight matrix, $\mathbf{W} = \mathbf{D}$, defined as

$$\mathbf{D} = \text{diag}\big\{ 1/\text{Var}(\varepsilon_{p+1}(\boldsymbol{\psi}^*)), 1/\text{Var}(\varepsilon_{p+2}(\boldsymbol{\psi}^*)), \ldots,$$
$$1/\text{Var}(\varepsilon_n(\boldsymbol{\psi}^*)) \big\}, \qquad (8)$$

giving us the diagonally weighted least-squares (DWLS) parameter estimators. Our choice of the diagonal weight matrix $\mathbf{D}$ is based on the empirical comparison/analysis of Swain et al. (1988), in which DWLS is found to be superior to WLS, and the proof in Kuhl and Wilson (1999) that explains the poor performance of WLS by the cancellation of a number of residual terms from the objective function $S_{\mathbf{W}}(\boldsymbol{\psi})$.

The derivation of the objective function (4) is completed after the objective function of the DWLS estimation problem is written by using (6) and (8). Note that this formulation has two major advantages that make the ARTAFIT framework both easy to understand and easy to implement.

The first one is that the necessary first and second moments of the order statistics are known and easily computed. The second and the most important advantage is that the use of the translated uniform order statistics $\Phi\{V_{(t)}(\boldsymbol{\psi}^*)\}$, $t = p+1, p+2, \ldots, n$, for fitting permits a single formulation for not only Johnson-type distributions, but all continuous distributions.

Next, we present our ARTAFIT algorithm together with the statistical properties of the resulting estimators.

2.3. The ARTAFIT Algorithm

We can minimize the objective function (4) subject to the constraints in (7) by using a general-purpose optimization algorithm. Unfortunately, many of these algorithms are dependent upon good initial estimates of the parameters. Further, the number of model parameters we need to estimate is $p+4$, which increases linearly with the order of dependence p, making it even less likely that we can obtain robust estimates that are independent of the quality of the initial solution as p gets larger. Fortunately, there is a natural decomposition of our optimization problem, namely, the Johnson parameters $(\gamma, \delta, \lambda, \xi)$ and the base-process parameters $(\alpha_1, \alpha_2, \ldots, \alpha_p)$. We have empirically observed that solving $S_{\mathbf{D}}(\boldsymbol{\psi})$ for any fixed, feasible γ, δ, λ, and ξ provides robust estimates of $\alpha_1, \alpha_2, \ldots, \alpha_p$ in the very next iteration. Therefore, we work iteratively between improving the estimates for $(\gamma, \delta, \lambda, \xi)$ and $(\alpha_1, \alpha_2, \ldots, \alpha_p)$

Before we give a complete statement of our ARTAFIT algorithm, we introduce the notation we will use in its presentation. We first let $\mathbf{x} = (x_1, x_2, \ldots, x_n)'$ denote a vector of sample data in which ties occur with probability zero and then define $S_{\mathbf{D}}(\boldsymbol{\psi} \mid \mathbf{x})$ as the objective function dependent on the given sample. We also let $\mathbf{C}: \boldsymbol{\Psi} \to \boldsymbol{\Psi}$ and $\mathbf{D}: \boldsymbol{\Psi} \to \boldsymbol{\Psi}$ be point-to-point maps given by

$$\mathbf{C}(\boldsymbol{\psi}) \equiv \underset{\gamma, \delta, \lambda, \xi}{\arg\min} \; S_{\mathbf{D}}(\boldsymbol{\psi} \mid \mathbf{x})$$

subject to $\boldsymbol{\psi} \in \boldsymbol{\Psi}_{\mathbf{C}}$ \qquad (9)

and

$$\mathbf{D}(\boldsymbol{\psi}) \equiv \underset{\alpha_1, \alpha_2, \ldots, \alpha_p}{\arg\min} \; S_{\mathbf{D}}(\boldsymbol{\psi} \mid \mathbf{x})$$

subject to $\boldsymbol{\psi} \in \boldsymbol{\Psi}_{\mathbf{D}}$, \qquad (10)

where $\boldsymbol{\Psi}_{\mathbf{C}}$ and $\boldsymbol{\Psi}_{\mathbf{D}}$ correspond to the constraints ensuring the feasibility of the Johnson parameters and the stationarity of the underlying base process, respectively, and satisfy $\boldsymbol{\Psi} = \boldsymbol{\Psi}_{\mathbf{C}} \cup \boldsymbol{\Psi}_{\mathbf{D}}$. Finally, we define the solution set $\boldsymbol{\Omega}(\mathbf{x}) \equiv \{\overline{\boldsymbol{\psi}}: \nabla_{\boldsymbol{\psi}} S_{\mathbf{D}}(\overline{\boldsymbol{\psi}} \mid \mathbf{x}) = \mathbf{0}\}$, corresponding to the collection of parameters at which all of the entries of the gradient of the objective function $S_{\mathbf{D}}(\boldsymbol{\psi} \mid \mathbf{x})$ attain the value of zero. Next, we present the ARTAFIT algorithm:

Initialization Step. Let $k = 1$ and $\boldsymbol{\psi}_0 \in \boldsymbol{\Psi}$ be a starting parameter vector.

Main Step. 1. Let $\boldsymbol{\psi}_k \in \mathbf{C}(\boldsymbol{\psi}_{k-1})$. Replace k by $k+1$ and go to Step 2.

2. Let $\boldsymbol{\psi}_k \in \mathbf{D}(\boldsymbol{\psi}_{k-1})$. If $\boldsymbol{\psi}_k \in \boldsymbol{\Omega}(\mathbf{x})$, then stop; otherwise, replace k by $k+1$ and go to Step 1.

Starting with a parameter vector in $\boldsymbol{\Psi}$, we first solve the least-squares fitting problem for the Johnson parameters $(\gamma, \delta, \lambda, \xi)$ by keeping the base-process parameters $(\alpha_1, \alpha_2, \ldots, \alpha_p)$ fixed. We call this Step 1. Then, we solve the least-squares fitting problem for $\alpha_1, \alpha_2, \ldots, \alpha_p$ by keeping $\gamma, \delta, \lambda, \xi$ fixed, and we call this Step 2. We work iteratively between these two steps until the ARTAFIT algorithm converges.

THEOREM 1. *If $\boldsymbol{\psi}_0 \in \boldsymbol{\Psi}$ is a starting parameter vector in a convex subregion that includes any local minimum of problem* (4), *and we use the Levenberg-Marquardt algorithm to carry out Steps* 1 *and* 2, *then the ARTAFIT algorithm converges to a local minimum solution.*

PROOF. See Theorem 1 of Biller and Nelson (2004). □

Thus, using a general-purpose optimization algorithm with local convergence properties ensures that we reach a local minimum solution when we start in its convex surrounding region; however, a local minimum solution might be on the boundary of the feasible region $\boldsymbol{\Psi}$ (or, more precisely, in the limit as certain parameters approach their bounds). We address this issue here, because an ARTA process with parameter vector $\boldsymbol{\psi}$ is well defined only if $\boldsymbol{\psi}$ falls into the interior of the region $\boldsymbol{\Psi}$ defined in (7). Although we have never actually encountered such a situation in practice, we explain how we would ensure the feasibility of the ARTA distribution when the resulting estimator is on the boundary of $\boldsymbol{\Psi}$:

• By constructing a contrived example, we have been able to show that a local minimum solution can occur as $\widehat{\xi}$ approaches $x_{(1)}$ for the Johnson lognormal distribution, and as $\widehat{\xi}$ approaches $x_{(1)}$ or $\widehat{\lambda}$ approaches $x_{(n)} + \widehat{\xi}$ for the Johnson bounded distribution. We believe that little is lost by taking $\widehat{\xi} = x_{(1)}$ or $\widehat{\lambda} = x_{(n)} + \widehat{\xi}$. Although we are technically removing $x_{(1)}$ or $x_{(n)}$ from the support of the distribution, values arbitrarily close to these data points are still possible, and in the contrived example these boundary values of $\widehat{\xi}$ and $\widehat{\lambda}$ are local optima because they allow the other order statistics to be matched perfectly. The contrived example is available in Biller and Nelson (2004).

• Our ARTAFIT algorithm provides an input model with a degenerate marginal distribution, i.e., $\Pr\{X = 0\} = 1$, when $\widehat{\lambda}$ of the Johnson unbounded family is zero or the absolute value of each root of the reverse characteristic polynomial falls onto the complex unit circle. Fortunately, we can show that none of these solutions is the global minimum of the objective function $S_\mathbf{D}(\boldsymbol{\psi} \mid \mathbf{x})$; further, we can always find a value for $\widehat{\gamma}$ that results in a lower objective function value for any nonzero value of $\widehat{\lambda}$ of the Johnson unbounded distribution and $(\widehat{\alpha}_1, \widehat{\alpha}_2, \ldots, \widehat{\alpha}_p)$ of the underlying base process (Biller and Nelson 2004, Proposition 2).

We note that our approach bears some similarity to the forecasting technique of Block et al. (1990). They also consider the observed data to be a transformation, via the inverse cdf, of an underlying Gaussian process. They propose using this transformation to provide a joint distribution for the observed data, but, unlike us, they solve for the unknown parameters of the marginal distribution and the underlying Gaussian process simultaneously via maximum-likelihood estimation. However, the resulting likelihood function appears to be difficult to maximize except for the simplest models, and they provide no properties for the resulting estimators. We establish the statistical properties of our estimators in the next section.

2.4. Statistical Properties of the ARTA Estimators

If $X_1, X_2, \ldots, X_n$ are actually defined by an ARTA process, then we have the following result:

THEOREM 2. *Let $X_1, X_2, \ldots, X_n$ be identically distributed random variables with a joint ARTA distribution characterized by the parameter vector $\boldsymbol{\psi}^*$, for which the type of the Johnson transformation f and the order of the underlying base process p are known. Suppose that $\widehat{\boldsymbol{\psi}}_n$ is a solution to the diagonally weighted least-squares problem given by* (6), (7), *and* (8). *Then, as the time-series length $n \to \infty$, the following properties hold*:
  1. *$\Pr[\widehat{\lambda}_n \to \lambda^*] = 1$ and $\Pr[\widehat{\xi}_n \to \xi^*] = 1$.*
  2. *If $\widehat{\alpha}_h = \alpha_h^*$, $h = 1, 2, \ldots, p$, then $\Pr[\widehat{\boldsymbol{\psi}}_n \to \boldsymbol{\psi}^*] = 1$.*

PROOF. See Theorem 2 of Biller and Nelson (2004). □

The first result in Theorem 2 is of limited practical value. However, the second result is helpful in two ways:

1. We proposed decomposing the algorithm for solving the least-squares problem into two steps—improving the estimates of $(\gamma, \delta, \lambda, \xi)$ by keeping $(\alpha_1, \alpha_2, \ldots, \alpha_p)$ fixed and improving the estimates of $(\alpha_1, \alpha_2, \ldots, \alpha_p)$ by keeping $(\gamma, \delta, \lambda, \xi)$ fixed—because we observed that the estimates of $(\alpha_1, \alpha_2, \ldots, \alpha_p)$ are robust to poor estimates of $(\gamma, \delta, \lambda, \xi)$. Theorem 2 shows that when we get the base-process parameters right, the least-squares estimators of the remaining ARTA parameters are strongly consistent.

2. When $p = 0$, our least-squares criterion reduces to the one suggested by Swain et al. (1988) for fitting i.i.d. data. If $\{X_t; t = 1, 2, \ldots, n\}$ are actually i.i.d. Johnson-type random variables with the parameter set $\boldsymbol{\psi}^* = (\gamma^*, \delta^*, \lambda^*, \xi^*)'$, then the ARTA model with $p = 0$ is correct. Swain et al. (1988) show empirically that solving the least-squares problem provides a convenient computational method for fitting any member of the Johnson translation system when $p = 0$, but they do not provide any statistical properties of the resulting estimators. Theorem 2 indicates that their formulation gives strongly consistent estimators of the Johnson parameters.

A natural question to ask is why the DWLS ARTAFIT estimators are not consistent in general: For consistency to hold, it is necessary and sufficient that the empirical distribution of the $R_t(\boldsymbol{\psi})$, $t = p + 1, p + 2, \ldots, n$, converges to the uniform distribution on the unit interval $(0, 1)$ as $n \to \infty$. We can show that $R_t(\boldsymbol{\psi}^*)$, $t = p + 1, p + 2, \ldots, n$,

have i.i.d. uniform $(0, 1)$ marginals only if $\boldsymbol{\psi} = \boldsymbol{\psi}^*$ (Biller and Nelson 2004, Theorem 3). However, $R_t(\boldsymbol{\psi})$, $t = p + 1$, $p + 2, \ldots, n$, can be dependent, but still have uniform marginals, at parameter settings other than $\boldsymbol{\psi}^*$; for instance, at $\gamma = \gamma^*$, $\delta = \delta^*$, $\lambda = \lambda^*$, $\xi = \xi^*$, and $\alpha_h = 0$ for $h = 1, 2, \ldots, p$.

Motivated by the second result in Theorem 2, another natural question to ask is whether the following property holds: If $\widehat{\delta}_n = \delta^*$, $\widehat{\gamma}_n = \gamma^*$, $\widehat{\lambda}_n = \lambda^*$, and $\widehat{\xi}_n = \xi^*$, then $\Pr[\widehat{\boldsymbol{\psi}}_n \to \boldsymbol{\psi}^*]$ $= 1$, i.e., $\Pr[\widehat{\alpha}_h \to \alpha_h^*] = 1$ for $h = 1, 2, \ldots, p$. Although it is straightforward to prove the existence of this property when $p = 1$, it cannot be extended to higher orders of dependence. Because strongly consistent estimators of the parameters of Gaussian AR(p) processes are well known, the corresponding property would hold if we were using one of these estimators for the parameters of a Gaussian AR(p) process in the second step of the algorithm. Our motivation for not doing so, but instead using formulation (4), is to characterize the joint estimation of $(\gamma, \delta, \lambda, \xi, \alpha_1, \alpha_2, \ldots, \alpha_p)$ by a single objective that does not favor either the Johnson or the base-process parameters, leading to a direct proof of the convergence properties of the numerical algorithm. We consider alternative formulations that lead to strongly consistent estimators of the entire vector of parameters $\boldsymbol{\psi}^*$ as a subject of future research.

3. Illustrative Examples

In this section, we illustrate the use of the ARTAFIT algorithm with two different examples. The data used in the first example are generated artifically from an ARTA process, while the data in the second example come from a real process representing a pressure variable of a continuous-flow production line. In all of the experiments reported below, we use a standalone, PC-based program that implements the data-fitting algorithm as described in the previous section, but no longer assuming that the type of the Johnson transformation f and the order of the underlying base-process p are known. Instead, we fit all four members of the Johnson family and select the one that fits best, and we choose the order of the underlying base process using time-series analysis techniques; see Biller and Nelson (2004) for complete details. The key computational components of the software are written in portable C++ code and are available at http://www.andrew.cmu.edu/~billerb. We refer the interested reader to Biller and Nelson (2004) for the numerical methods used to implement the data-fitting algorithm.

3.1. Source of Data: A Known ARTA Process

In this section, we test the performance of the ARTAFIT algorithm against data that come from a known ARTA process, and measure how well the ARTAFIT algorithm recovers the parameters of the true process. The results of this section are representative of a larger study in Biller (2002), in which the method was found to do a good job of recovering the parameters of ARTA processes across a wide range of examples.

Table 1. Percent differences between the estimates and the true parameters when $p = 1$.

$\boldsymbol{\psi}$	$n = 50$	$n = 100$	$n = 500$	$n = 1{,}000$	$n = 5{,}000$
$\rho_{\mathbf{X}}(1)$	5.735	1.323	0.191	0.007	0.001
$\rho_{\mathbf{Z}}(1)$ $(=\alpha_1)$	−9.653	−6.222	−1.997	−0.819	−0.144
γ	47.345	32.950	19.236	5.293	1.278
δ	38.283	27.318	13.232	4.345	1.201
λ	19.620	15.243	9.198	3.284	0.893
ξ	20.131	18.235	7.329	2.049	0.672
R	39	36	37	34	33

Because of its capability of capturing the tail behavior of a wide variety of distributional shapes, we choose to generate data from the Johnson unbounded (S_U) marginal distribution and, in particular, we assume that $\gamma = -0.540$, $\delta = 1.540$, $\lambda = 1.140$, and $\xi = -0.510$. Marginals from other members of the Johnson family and some non-Johnson marginals were also tested in Biller (2002). As factors for the empirical analysis of this section, we select the sample size n and the order of dependence p. We let n take the values of 50, 100, 500, 1,000, and 5,000 and the autocorrelation structures be defined as $\boldsymbol{\rho}_X(1) = 0.35$ when $p = 1$; $\boldsymbol{\rho}_X(1) = 0.60$ and $\boldsymbol{\rho}_X(2) = 0.20$ when $p = 2$; and $\boldsymbol{\rho}_X(1) = -0.45$, $\boldsymbol{\rho}_X(2) = 0.20$, and $\boldsymbol{\rho}_X(3) = -0.10$ when $p = 3$.

The goal is to see whether the ARTAFIT algorithm recovers the true ARTA representation, jointly characterized by the S_U marginal distribution and the autocorrelation structures specified in the previous paragraph, as the sample size increases. We implement the algorithm by assuming that the type of the Johnson transformation f and the order of the underlying process p are not known. We identify the transformation type by fitting distributions from *all* of the families, i.e., S_L, S_B, S_U, and S_N, and comparing the goodness of the resulting fits. We determine the order of the underlying process by applying an order selection method due to Schwarz (1978): We search for the parameter p that minimizes the value of the criterion $\ln |\tilde{\sigma}_Y(p)| + ((\ln n)/n)p$, where $\tilde{\sigma}_Y(p)$ denotes the (pseudo) maximum-likelihood estimator of σ_Y for an AR(p) model. In Tables 1, 2, and 3, we report the percent differences

Table 2. Percent differences between the estimates and the true parameters when $p = 2$.

$\boldsymbol{\psi}$	$n = 50$	$n = 100$	$n = 500$	$n = 1{,}000$	$n = 5{,}000$
$\rho_{\mathbf{X}}(1)$	4.906	2.724	0.976	0.012	0.005
$\rho_{\mathbf{X}}(2)$	3.826	2.927	1.115	0.084	0.011
$\rho_{\mathbf{Z}}(1)$	2.360	0.127	0.162	0.578	0.086
$\rho_{\mathbf{Z}}(2)$	9.332	1.970	0.868	1.007	0.148
α_1	2.829	0.245	0.129	1.023	0.150
α_2	2.135	1.719	0.149	2.066	0.303
γ	39.627	25.326	16.931	5.993	2.008
δ	33.825	20.537	11.309	6.248	1.762
λ	23.911	16.208	7.427	4.396	0.879
ξ	17.243	13.157	8.502	3.685	0.713
R	51	49	45	47	43

Table 3. Percent differences between the estimates and the true parameters when $p = 3$.

ψ	$n = 50$	$n = 100$	$n = 500$	$n = 1,000$	$n = 5,000$
$\rho_X(1)$	3.986	2.071	1.237	0.283	0.003
$\rho_X(2)$	2.257	1.562	0.832	0.354	0.066
$\rho_X(3)$	5.935	2.531	0.993	0.638	0.184
$\rho_Z(1)$	14.856	3.063	0.967	1.684	0.184
$\rho_Z(2)$	−18.416	−4.780	−1.942	−1.584	−0.181
$\rho_Z(3)$	−16.542	−22.144	−21.756	−1.694	−0.075
α_1	−23.475	−3.935	−1.064	−2.573	−0.273
α_2	−76.541	−16.873	−8.718	−8.248	−0.827
α_3	25.616	10.835	8.135	2.768	0.259
γ	52.483	39.371	26.731	9.214	3.846
δ	40.319	24.946	15.247	10.004	1.962
λ	18.242	14.283	10.229	7.412	1.093
ξ	21.411	15.372	8.251	4.247	2.974
R	67	68	61	58	60

between the fitted ARTA parameters/autocorrelation structures and the true ARTA parameters/autocorrelation structures used to generate the data. We use the sign "−" to indicate that the fitted ARTA parameters underestimate the true ARTA parameters, while the lack of a sign corresponds to the overestimation of the true ARTA parameters. The numbers reported are the percent differences averaged over a number of replications determined as follows: Each experiment starts with an initial run of 30 replications, and the number of replications is increased whenever necessary to ensure an absolute error of no more than 0.1 on the scaled Kolmogorov-Smirnov test statistic. The number of replications carried out for each experimental setting is provided in the last rows of Tables 1, 2, and 3 and denoted by R. The summary test statistics reported in Tables 4, 5, and 6 later in this section are also averaged over the same number of replications.

In all of the experiments tabulated in Tables 1, 2, and 3, the ARTAFIT algorithm recovers the true ARTA processes as the sample size increases. Observe that when $n = 5,000$ the maximum percent difference between the fitted and true input autocorrelations is less than 0.184%, while the maximum percent difference between the fitted and true Johnson parameters is 3.846%. Note also that the percentage errors in the underlying AR-process parameters,

$\alpha_1, \alpha_2, \ldots, \alpha_p$, tend to be small at much lower sample sizes than the Johnson-distribution parameters. This supports our contention that the base-process parameter estimates, while not guaranteed to be consistent, are robust to errors in the Johnson parameter estimates.

Next, we measure how well our fits capture the key features of the corresponding data samples. Because an ARTA process is jointly characterized by a Johnson-type marginal distribution function F_X and an autocorrelation structure specified by $\rho_X(h)$, $h = 1, 2, \ldots, p$, neither a pure marginal fit nor a pure autocorrelation fit is sufficient for choosing a good representation for the ARTA process. Instead, we need a test that measures the goodness of the joint ARTA distribution fit. Therefore, we use the two-dimensional Kolmogorov-Smirnov (KS2) test, which is a generalization of the one-dimensional Kolmogorov-Smirnov test to bivariate distributions (Press et al. 1992). The KS2 test measures the extent to which the pairs (x_t, x_{t-h}), $t = h + 1$, $h + 2, \ldots, n$, generated by the true ARTA process are consistent with the probability distribution of the two-dimensional random variable (X_t, X_{t-h}) whose marginal distributions are identical and characterized by the Johnson parameter estimates, $\hat{f}, \hat{\gamma}, \hat{\delta}, \hat{\lambda},$ and $\hat{\xi}$, and whose bivariate input correlation is given by $\hat{\rho}_X(h)$.

Specifically, the KS2 test statistic is a measure of the largest distance between the two-dimensional empirical distribution function and the distribution function of the random variable (X_t, X_{t-h}). However, calculation of the KS2 test statistic is not as straightforward as the one-dimensional Kolmogorov-Smirnov test statistic, because the cumulative probability distribution is not well defined in more than one dimension. Based on the experience of Press et al. (1992), we choose to first calculate the fraction of data points falling in each of four natural quadrants around each

Table 4. The KS2 test statistics and observed significance levels when $h = 1$.

ρ_X	$n = 50$	$n = 100$	$n = 500$	$n = 1,000$	$n = 5,000$
0.35	0.253	0.185	0.088	0.058	0.029
	(0.291)	(0.231)	(0.211)	(0.312)	(0.246)
$(0.60, 0.20)'$	0.245	0.171	0.052	0.061	0.028
	(0.281)	(0.311)	(0.697)	(0.187)	(0.229)
$(-0.45, 0.20, -0.10)$	0.265	0.171	0.096	0.059	0.025
	(0.211)	(0.314)	(0.172)	(0.268)	(0.297)

Table 5. The KS2 test statistics and observed significance levels when $h = 2$.

ρ_X	$n = 50$	$n = 100$	$n = 500$	$n = 1,000$	$n = 5,000$
0.35	0.256	0.185	0.090	0.059	0.031
	(0.283)	(0.249)	(0.202)	(0.289)	(0.194)
$(0.60, 0.20)'$	0.261	0.182	0.061	0.067	0.030
	(0.268)	(0.266)	(0.547)	(0.153)	(0.192)
$(-0.45, 0.20, -0.10)$	0.258	0.163	0.089	0.056	0.025
	(0.242)	(0.342)	(0.217)	(0.312)	(0.315)

Table 6. The KS2 test statistics and observed significance levels when $h = 3$.

ρ_X	$n = 50$	$n = 100$	$n = 500$	$n = 1,000$	$n = 5,000$
0.35	0.272	0.189	0.093	0.059	0.032
	(0.291)	(0.247)	(0.202)	(0.284)	(0.200)
$(0.60, 0.20)'$	0.273	0.189	0.064	0.069	0.032
	(0.250)	(0.246)	(0.481)	(0.126)	(0.184)
$(-0.45, 0.20, -0.10)$	0.262	0.171	0.096	0.059	0.025
	(0.237)	(0.320)	(0.166)	(0.268)	(0.306)

point (x_t, x_{t-h}), $t = h+1, h+2, \ldots, n$, and then calculate the probabilities of $(X_t > x_t, X_{t-h} > x_{t-h})$, $(X_t < x_t, X_{t-h} > x_{t-h})$, $(X_t < x_t, X_{t-h} < x_{t-h})$, and $(X_t > x_t, X_{t-h} < x_{t-h})$ for the same set of points. By ranging both over the data points and over the quadrants, we take the maximum difference of the corresponding probabilities as the KS2 test statistic. We present the resulting KS2 test statistics with their observed significance levels in parentheses in Tables 4, 5, and 6 for $h = 1, 2$, and 3, respectively.

In our experimental setting, an observed significance level gives the extent to which a test statistic disagrees with the null hypothesis that the fitted ARTA distribution provides a good representation of the empirical data generated by the true ARTA process. We select the level of significance as 5%; thus, if an observed significance level is less than 0.05, then the test results are significant and we reject the null hypothesis, while observed significance greater than 0.05 supports goodness of the fit. Thus, Tables 4, 5, and 6 provide evidence that the fitted ARTA models provide good representations of the empirical processes in all of the experiments conducted in this section. Although we have developed the algorithm for input-modeling problems in which data are plentiful, the KS2 test statistics show that we still do a good job in fitting input models to small data samples. We attribute this to the flexibility of our ARTAFIT framework and the robustness of the least-squares fitting criterion.

3.2. Source of Data: A Real Process

It is essential that we experiment with realistic—as opposed to artificially generated—input-modeling problems to stress the proposed methodology, because real problems violate many or all of the assumptions of the simple input models in current use, including the existence of a true, underlying distribution with a simple structure. Therefore, in this section, we use our ARTAFIT algorithm to model 519 observations of a pressure variable from a continuous-flow production line that is recorded at fixed time intervals; these measurements exhibit strong serial dependence. System simulation is sometimes used to model new and existing production lines, as well as to train new operators in proper responses to process changes. Because pressure is one of the key parameters of this type of production lines, the selection of an input model for this parameter is critical in understanding its effect on the output performance measures of the corresponding production system.

Cario and Nelson (1998) suggest representing this dataset by using an ARTA process, but their approach is quite different from ours: They first solve for the unknown parameters of the marginal distribution via maximum-likelihood estimation, and then determine the input auto-correlations by using the sample autocorrelation function of the raw data. Specifically, they first fit a Weibull marginal distribution function (with shape and scale parameters of $\hat{\beta}_1 = 5.140$ and $\hat{\beta}_2 = 0.940$, respectively) to the 519 data points with the aid of the Arena Input Analyzer (Rockwell

Table 7. Comparison of KS, AD, and two-dimensional KS test statistics for pressure data.

	AF	ARF(0)	ARF(2)
KS	1.764	1.038	0.836
AD	3.295	0.758	0.759
KS2(1)	0.191	0.832	0.088
	(0.000)	(0.000)	(0.145)
KS2(2)	0.182	0.840	0.098
	(0.000)	(0.000)	(0.116)
KS2(3)	0.229	0.875	0.102
	(0.000)	(0.000)	(0.106)

Software) that assumes i.i.d. data, and then approximate the input autocorrelation structure as $\hat{\rho}_X(1) = 0.751$, $\hat{\rho}_X(2) = 0.407$, and $\hat{\rho}_X(3) = 0.121$. Finally, they employ their software ARTAFACTS to match these estimated lags 1, 2, and 3 input autocorrelations to the autocorrelations of the base process. We call this model ARTAFACT and denote it by "AF" in Table 7.

Figure 1. (a) *Q-Q* plot comparing the empirical and ARTAFACT distribution functions. (b) *Q-Q* plot comparing the empirical and ARTAFIT distribution functions.
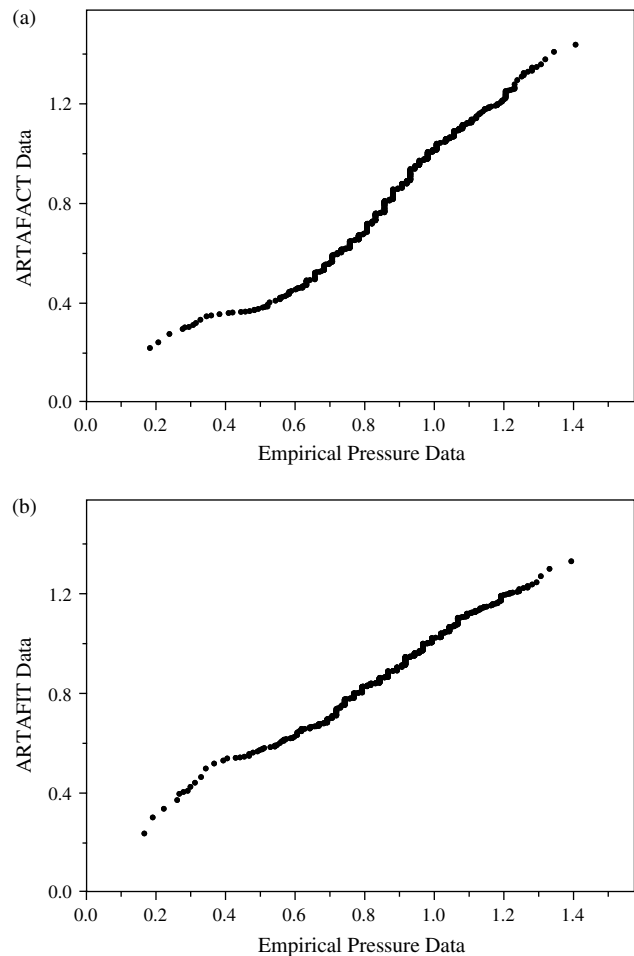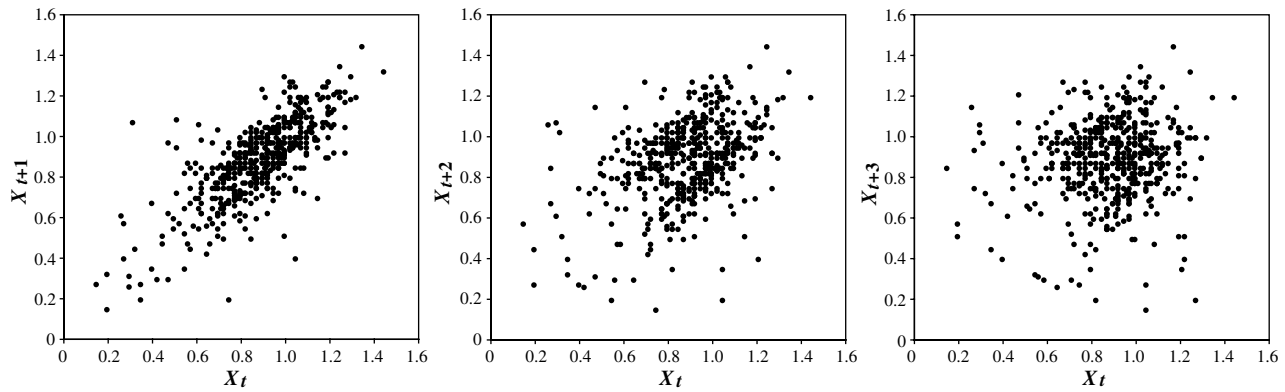

(a)

(b)

Figure 2. Scatter plots for the empirical pressure data.

Like Cario and Nelson (1998), we represent the corresponding dataset by using an ARTA process, but unlike Cario and Nelson, we solve for the unknown parameters of the marginal distribution and the input autocorrelations simultaneously via the least-squares data-fitting algorithm developed in the previous section. As in §3.1, we identify the type of the transformation f to use from the Johnson translation system by fitting distributions from *all* of the families, i.e., S_L, S_B, S_U, and S_N, and comparing the goodness of the resulting fits. Also, we relax the assumption that p is known through the application of the Schwarz criterion. Specifically, we fit a Johnson unbounded distribution and an autocorrelation structure with $p = 2$—characterized by $(\hat{\gamma}, \hat{\delta}, \hat{\lambda}, \hat{\xi}) = (2.046, 3.151, 0.457, 1.217)$ and $(\alpha_1, \alpha_2) = (1.050, -0.342)$, implying $\hat{\rho}_X(1) = 0.776$ and $\hat{\rho}_X(2) = 0.524$—to the same dataset. We denote this ARTAFIT model by "ARF(p)" in Table 7 and use p to denote its order.

Next, we provide quantile-quantile ($Q - Q$) plots comparing the ith quantile of the empirical (pressure) distribution function, $X_{(i)}$, with the ith quantile of the ARTAFACT distribution function, $\hat{\beta}_2 + (\log(1 - (i - 0.5)/519))^{\hat{\beta}_1^{-1}}$ for $i = 1, 2, \ldots, 519$ in Figure 1a and the $Q - Q$ plot comparing $X_{(i)}$ with the ith quantile of the ARTAFIT distribution

function, $\hat{\xi} + \hat{\lambda}\hat{f}^{-1}[((i - 0.5)/519 - \hat{\gamma})/\hat{\delta}]$ for $i = 1, 2, \ldots, 519$ in Figure 1b. Although the $Q - Q$ plot in Figure 1b seems more linear than the $Q - Q$ plot in Figure 1a, it is hard to conclude that there is a closer agreement between the ARTAFIT distribution and the true distribution of the physical process. Therefore, to substantiate our visual observation, we use goodness-of-fit tests to compare our results to those of Cario and Nelson (1998): We first compare the fitted Johnson cdf, $\Phi\{\hat{\gamma} + \hat{\delta}\hat{f}[(x - \hat{\xi})/\hat{\lambda}]\}$, to the empirical cdf and present the scaled Kolmogorov-Smirnov and Anderson-Darling test statistics, denoted by KS and AD, respectively, in the first two rows of Table 7. Then, we provide the KS2 test statistics together with their observed significance levels in the last three rows of Table 7 for $h = 1, 2$, and 3. The second column (AF) corresponds to the fit suggested by Cario and Nelson (1998), the third column (ARF(0)) corresponds to the Johnson fit under the assumption of independence, and the last column corresponds to the Johnson fit suggested by our data-fitting algorithm.

As noted by Moore (1982) and Gleser and Moore (1983) in the context of short-memory processes, the critical values and the corresponding nominal levels of significance of goodness-of-fit tests for i.i.d. data can be grossly incorrect when observations are dependent. Thus, we use the

Figure 3. Scatter plots for the ARTAFACT data.
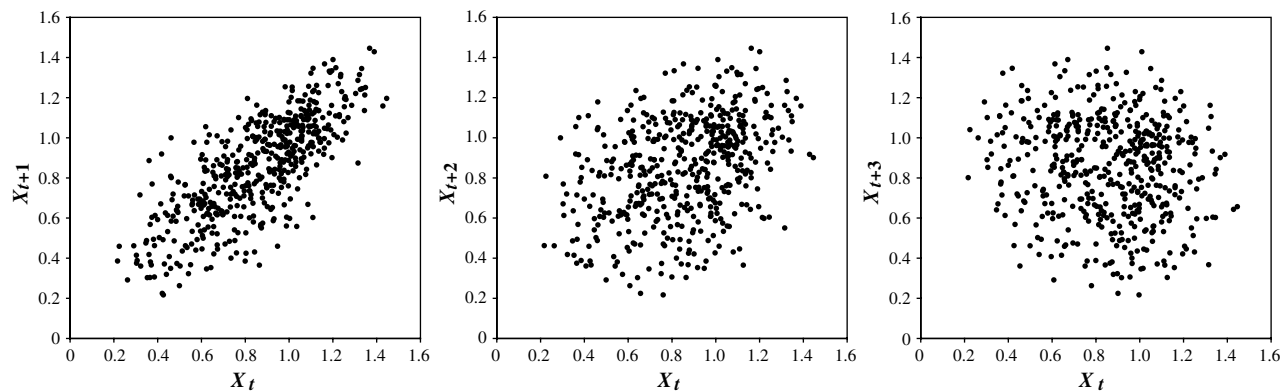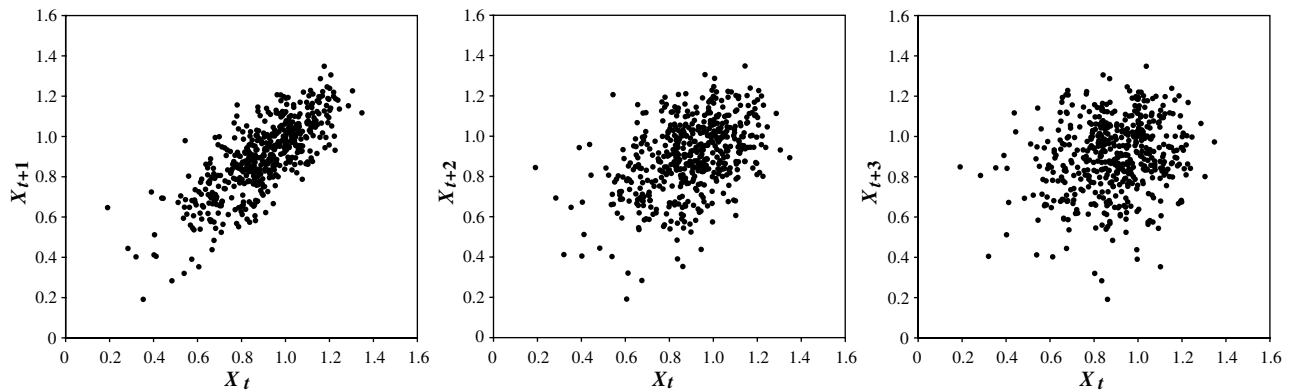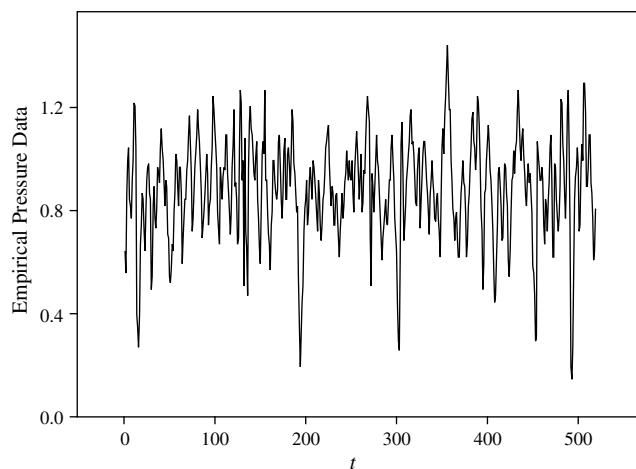
Figure 4.   Scatter plots for the ARTAFIT data.

5% critical values of 0.895 and 0.751 for the KS and AD tests, respectively, as a rough guide for judging the adequacy of the fit. A closer look at the scaled KS and AD test statistics suggests that the ARF(2) marginal is a good representation of the empirical process. Note that the AD test statistic of ARF(0) is quite comparable to the AD test statistics of ARF(2), even though the dependencies are ignored while developing the ARF(0) model. We can explain this observation by the fact that a pure marginal match is not the only thing that matters while choosing a good representation for a system with strong dependencies. The KS2 test statistics in the last three lines of Table 7 indicate that the pressure data must be drawn from a distribution significantly different from the ARF(0) distribution as well as the AF distribution. Instead, the KS2 test statistics suggest the use of our ARF(2) model.

Next, we illustrate the ARTA fit with graphical comparisons. To do this, we first generate 519 data points from each fitted model and display the scatter plots of (x_t, x_{t-1}), (x_t, x_{t-2}), (x_t, x_{t-3}) for the empirical pressure data and the data of the fitted ARTAFACT and ARTAFIT models. See Figures 2, 3, and 4. Comparison of Figures 2 and 3 shows

that the ARTAFACT data appear to be more scattered or random than the empirical data. We also provide the time-series plots in Figures 5, 6, and 7. The sample paths are qualitatively similar, but there are spikes in the empirical series that are not present in the ARTAFACT series. On the other hand, comparison of Figures 5 and 7 shows that the ARTAFIT model captures the height of the spikes reasonably well, while comparison of Figures 2 and 4 shows that our ARTAFIT data and the empirical data appear to be scattered similarly. Overall, our ARTAFIT process does a good job of capturing the key features of the underlying process.

4. Conclusion

In this paper, we propose an automated and statistically valid algorithm to fit stochastic models to dependent time-series input processes for simulation. We illustrate the algorithm using data generated by a real-world process and observe that it provides a plausible characterization of the process. Because simulation inputs form the core of every stochastic simulation model, the product of this research is expected to improve the accuracy of system representation
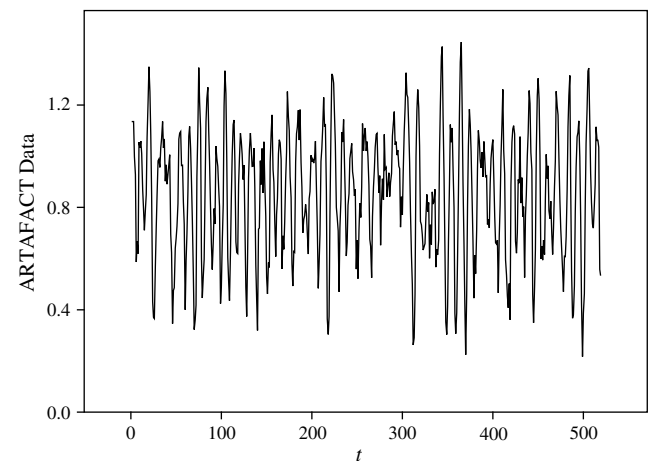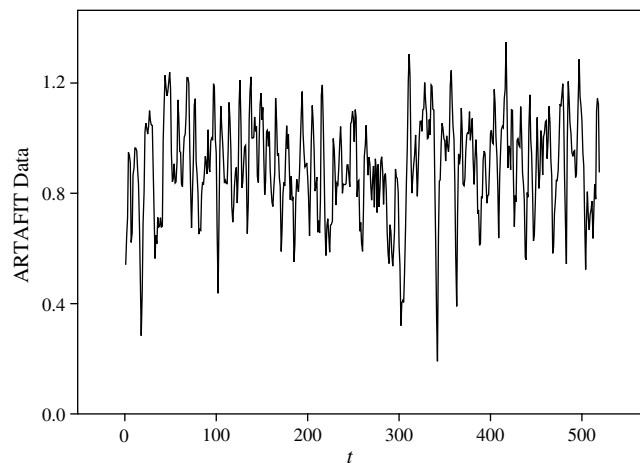
Figure 5.   Time-series plot of the empirical pressure data.

Figure 6.   Time-series plot of the ARTAFACT data.

Figure 7. Time-series plot of the ARTAFIT data.

that is necessary for the decisions the simulation model supports.

Recently, we have suggested a more comprehensive model for representing and generating stationary *multivariate* time-series input processes with arbitrary autocorrelation structures, and specifically considered the case of marginal distributions from the Johnson translation system (Biller and Nelson 2003). Our approach is very similar to the one in Cario and Nelson (1996), but we use a vector autoregressive Gaussian process that allows the modeling and generation of multivariate time-series processes. A natural extension of the work presented in this paper is to fit stochastic models to dependent, multivariate time-series input processes. This is a subject for future research.

Acknowledgments

References

Biller, B. 2002. A comprehensive input-modeling framework and software for stochastic discrete-event simulation experiments. Doctoral dissertation, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL.

Biller, B., B. L. Nelson. 2003. Modeling and generating multivariate time-series input processes using a vector autoregressive technique. *ACM Trans. Modeling Comput. Simulation* **13** 1–27.

Biller, B., B. L. Nelson. 2004. Online companion to "Fitting time-series input processes for simulation." Technical report, Graduate School of Industrial Administration, Carnegie Mellon University, Pittsburgh, PA.

Block, H. W., N. A. Langberg, D. S. Stoffer. 1990. Time series models for non-Gaussian processes. W. H. Block, A. R. Sampson, T. H. Savits, eds. *Topics in Statistical Dependence*. Institute of Mathematical Statistics, Heyward, CA, 69–83.

Cario, M. C., B. L. Nelson. 1996. Autoregressive to anything: Time-series input processes for simulation. *Oper. Res. Lett.* **19** 51–58.

Cario, M. C., B. L. Nelson. 1998. Numerical methods for fitting and simulating autoregressive-to-anything processes. *INFORMS J. Comput.* **10** 72–81.

Chatfield, C. 1999. *The Analysis of Time Series: An Introduction*. Chapman and Hall, New York.

Gleser, L. J., D. S. Moore. 1983. The effect of dependence on chi-squared test and empiric distribution tests of fit. *Ann. Statist.* **11** 1100–1108.

Gross, D., M. Juttijudata. 1997. Sensitivity of output performance measures to input distributions in queueing simulation modeling. S. Andradóttir, K. J. Healy, D. H. Withers, B. L. Nelson, eds. *Proc. 1997 Winter Simulation Conf.* Institute of Electrical and Electronics Engineers, Piscataway, NJ, 296–302.

Kendall, M. G., A. Stuart. 1979. *The Advanced Theory of Statistics*. Macmillan, New York.

Kuhl, M. E., J. R. Wilson. 1999. Least-squares estimation of non-homogeneous Poisson processes. *J. Statist. Comput. Simulation* **67** 75–108.

Johnson, N. L. 1949. Systems of frequency curves generated by methods of translation. *Biometrika* **36** 149–176.

Lewis, P. A. W., E. McKenzie, D. K. Hugus. 1989. Gamma processes. *Comm. Statist.—Stochastic Models* **5** 1–30.

Livny, M., B. Melamed, A. K. Tsiolis. 1993. The impact of autocorrelation on queueing systems. *Management Sci.* **39** 322–339.

Mallows, C. L. 1967. Linear processes are nearly Gaussian. *J. Appl. Probab.* **4** 313–329.

Melamed, B., J. R. Hill, D. Goldsman. 1992. The TES methodology: Modeling empirical stationary time series. J. J. Swain, D. Goldsman, R. C. Crain, J. R. Wilson, eds. *Proc. 1992 Winter Simulation Conf.* Institute of Electrical and Electronics Engineers, Piscataway, NJ, 135–144.

Moore, D. S. 1982. The effect of dependence on chi-squared tests of fit. *Ann. Statist.* **4** 357–369.

Nelsen, R. B. 1998. *An Introduction to Copulas*. Springer-Verlag, New York.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, B. P. Flannery. 1992. *Numerical Recipes in C. The Art of Scientific Computing*, 2nd ed. Cambridge University Press, Cambridge, UK.

Schwarz, G. 1978. Estimating the dimension of a model. *Ann. Statist.* **6** 461–464.

Seber, G. A. F. 1977. *Linear Regression Analysis*. John Wiley, New York.

Swain, J. J., S. Venkatraman, J. R. Wilson. 1988. Least-squares estimation of distribution functions in Johnson's translation system. *J. Statist. Comput. Simulation* **29** 271–297.

Ware, P. P., T. W. Page, B. L. Nelson. 1998. Automatic modeling of file system workloads using two-level arrival processes. *ACM Trans. Modeling Comput. Simulation* **8** 305–330.

Wei, W. W. S. 1990. *Time Series Analysis: Univariate and Multivariate Methods*. Addison Wesley, New York.