Proceedings of the 2002 Winter Simulation Conference
E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes, eds.

PARAMETER ESTIMATION FOR ARTA PROCESSES

Bahar Biller

Dept. of Manufacturing & Operations Management
Graduate School of Industrial Administration
Carnegie Mellon University
Pittsburgh, PA 15213, U.S.A.

Barry L. Nelson

Department of Industrial Engineering
& Management Sciences
Northwestern University
Evanston, IL 60208-3119, U.S.A.

ABSTRACT

Providing accurate and automated input-modeling support is one of the challenging problems in the application of computer simulation. The models incorporated in current input-modeling software packages often fall short of what is needed because they emphasize independent and identically distributed processes, while dependent time-series processes occur naturally in the simulation of many real-life systems. 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 ARTA (Autoregressive-to-Anything) processes with marginal distributions from the Johnson translation system to stationary univariate time-series data. The use of this algorithm is illustrated via a real-life example.

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 bitrates, 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. Later in this article, we model a pressure variable of a continuous-flow production line that is measured at fixed time intervals; these measurements exhibit strong series dependence and have been previously studied by Cario and Nelson (1998).

The development of models for such time-series processes is motivated by the fact that ignoring dependence can lead to performance measures that are seriously in error and a significant distortion of the simulated system. An illustration is given by Livny et al. (1993), who examine the impact of autocorrelation on queueing systems and report that when positive autocorrelation among, say, the interarrival times is ignored, then simulation results may grossly underestimate system congestion.

Much of the past work on time-series input processes is based on linear models, such as the ARMA class or those which 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 are non-normal. Motivated by this, there has been considerable research on modeling time series with marginals from specific 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 transform of a Gaussian linear process. For example, Cario and Nelson (1996) and Cario et al. (2001) take this approach to develop models for representing and generating stationary univariate time-series processes and finite-dimensional random vectors, respectively. The central idea in the former study is to transform a Gaussian autoregressive process into the desired univariate time-series input process that they presume as having an ARTA (Autoregressive-To-Anything) distribution, while the latter study transforms a multivariate normal random vector into the desired random vector that they refer to as having a NORTA (Normal-To-Anything) distribution. In both studies, the authors manipulate the correlations of the corresponding Gaussian process so that they achieve the desired correlations 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 correlations) 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. Therefore, the purpose of the present paper is to solve the problem of fitting stochastic input models to

stationary univariate time-series data and, specifically, fitting ARTA processes with marginal distributions from the Johnson translation system.

To facilitate a detailed discussion of the data-fitting problem, we first review the essential ideas involved in ARTA processes suggested by Cario and Nelson (1996). The rest of the paper is organized around the presentation of the three key levels of solving the data-fitting problem. In Section 3, we provide the iterative fitting algorithm and the statistical properties of the resulting estimators, and in Section 4, we give a brief review of the numerical methods used to implement the suggested algorithm. Section 5 provides a real-life example demonstrating the use of the algorithm and Section 6 gives concluding remarks.

2 OVERVIEW OF THE ARTA FRAMEWORK

In this section, we introduce the notation we will use and provide a brief review of the Johnson translation system; we then describe ARTA processes.

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 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 demands for a product 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 $\rho_X(h) = E\left[\sigma^{-2}(X_t - \mu)\ (X_{t-h} - \mu)\right]$, where X_t has mean μ and variance σ^2 for all t due to the assumption of stationarity of the input process.

2.2 Johnson Translation System

The Johnson translation system for a random variable X is defined by a cdf of the form

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

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:

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

There is a unique family (choice of f) for each feasible combination of finite skewness and kurtosis that determine the parameters γ and δ. Any mean and (positive) variance can be attained by any one of the families by the manipulation of the parameters λ and ξ. Within each family, a distribution is completely specified by the values of the parameters $(\gamma, \delta, \lambda, \xi)$; the range of X depends on the family of interest (Johnson 1949).

The Johnson translation system provides good representations for unimodal distributions and can represent certain bimodal shapes, but not three or more modes. Illustrations of the shapes of the Johnson-type probability density functions can be found in Johnson (1987). The first four moments of all distributions in the families S_L, S_B, S_U, and S_N are finite. Nevertheless, the ability to match any (finite) first four moments provides a great deal of flexibility that is sufficient for many practical problems.

2.3 ARTA Processes

An ARTA process is a time series with arbitrary marginal distribution and autocorrelation structure specified through finite lag p. It is based on the construction of a Gaussian standard time-series $\{Z_t;\ t = 1, 2, \ldots, n\}$ as a base process, from which we obtain a series of autocorrelated $(0, 1)$ uniform random variables $\{U_t;\ t = 1, 2, \ldots, n\}$ by using the probability-integral transformation $U_t = \Phi(Z_t)$. Then, the transformation $X_t = F_X^{-1}[U_t]$ is applied, ensuring that the input time-series process $\{X_t;\ t = 1, 2, \ldots, n\}$ has the desired marginal distribution F_X. This 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 framework described in the remainder of this section.

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

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

The α_h, $h = 1, 2, \ldots, p$, are fixed autoregressive coefficients and Y_t is white noise, representing that part of Z_t

that is not linearly dependent on past observations. The structure of Y_t is such that

$$\mathrm{E}[Y_t] = 0 \text{ and } \mathrm{E}[Y_t Y_{t-h}] = \begin{cases} \sigma_Y^2 & \text{if } h = 0, \\ 0 & \text{otherwise.} \end{cases}$$

Choosing σ_Y^2 appropriately ensures that each Z_t is marginally standard normal while the autoregressive coefficients α_h, $h = 1, 2, \ldots, p$, uniquely determine the autocorrelation structure of the base process, $\rho_Z(h)$, $h = 1, 2, \ldots, p$. Cario and Nelson (1996) have shown that the lag-h input autocorrelation $\rho_X(h)$ of the time-series input process $\{X_t; t = 1, 2, \ldots, n\}$ is a continuous, nondecreasing function of the lag-h autocorrelation $\rho_Z(h)$ of the base process. Therefore, if both the marginal distribution F_X and the desired input autocorrelations $\rho_X(h)$, $h = 1, 2, \ldots, p$, are known, then one can adjust the dependence in the Z_t process to yield the desired dependence in the X_t process by using any of the computationally feasible methods suggested by Song et al. (1996), Cario and Nelson (1998), and Chen (2001). The problem that has not been addressed is estimating the parameters of an ARTA process when only raw data produced by an unknown process are available. We will attack this problem in the remainder of the paper.

3 FITTING ARTA MODELS

In this section, we present the ARTA fitting algorithm together with the theory that supports it and the statistical properties that justify its use.

3.1 The 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 univariate 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

$$\begin{aligned} X_t &= F_X^{-1}\left[\Phi^{-1}(Z_t)\right] \\ &= \xi + \lambda f^{-1}\left[\frac{Z_t - \gamma}{\delta}\right], \end{aligned} \tag{1}$$

where

$$Z_t = \sum_{h=1}^{p} \alpha_h Z_{t-h} + Y_t, \tag{2}$$

with Y_t, $t = p+1, p+2, \ldots, n$, independent and identically distributed Gaussian random variables with mean zero and variance σ_Y^2. The value of σ_Y^2 that is required to force Z_t to have variance 1 is completely determined by $\alpha_1, \alpha_2, \ldots, \alpha_p$ (Wei 1990). Therefore, we will write $\sigma_Y \equiv g(p, \boldsymbol{\alpha})$, where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_p)$, and no longer consider σ_Y^2 as a parameter to be estimated. And from here on when we say "ARTA process," we will mean an ARTA process with Johnson-type marginals.

It is easy to see from Equations (1) and (2) that fitting an ARTA process to data corresponds to the estimation of f, γ, δ, λ, ξ, p, and α_h for $h = 1, 2, \ldots, p$. For ease of presentation of the ARTA fitting 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 in general. We address this issue in Section 4.

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

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

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

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

Suppose 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 independent and identically distributed (i.i.d.) standard normal random variables. Thus, the fitting procedure we propose searches for parameters that make $V_t(\boldsymbol{\psi})$, $t = p+1, p+2, \ldots, n$, appear to be such a sample.

Let $V_{(p+1)}(\boldsymbol{\psi}) \leq V_{(p+2)}(\boldsymbol{\psi}) \leq \cdots \leq V_{(n)}(\boldsymbol{\psi})$ denote the order statistics corresponding to the random variable $V_t(\boldsymbol{\psi})$, $t = p+1, p+2, \ldots, n$. If $\boldsymbol{\psi} = \boldsymbol{\psi}^*$, then the transformed variate $R_{(t)}(\boldsymbol{\psi}^*) = \Phi\left\{V_{(t)}(\boldsymbol{\psi}^*)\right\}$ has the distribution of the t^{th} order statistic in a random sample of size $n-p$ from the uniform distribution on the unit interval (0, 1). Since $R_{(t)}(\boldsymbol{\psi}^*)$ 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

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

Let $\mathbf{R_0}(\boldsymbol{\psi}) \equiv (R_{(p+1)}(\boldsymbol{\psi}), R_{(p+2)}(\boldsymbol{\psi}), \ldots, R_{(n)}(\boldsymbol{\psi}))'$, $\boldsymbol{\rho} \equiv (\rho_{p+1}, \rho_{p+2}, \ldots, \rho_n)'$, and $\boldsymbol{\varepsilon}(\boldsymbol{\psi}) \equiv (\varepsilon_{p+1}(\boldsymbol{\psi}), \varepsilon_{p+2}(\boldsymbol{\psi}), \ldots, \varepsilon_n(\boldsymbol{\psi}))'$, so that $\boldsymbol{\varepsilon}(\boldsymbol{\psi}^*) \equiv \mathbf{R_0}(\boldsymbol{\psi}^*) - \boldsymbol{\rho}$. Since the first and second moments of the uniformized order statistics are known and easily computed, we exploit this fact to develop a single, distribution-free formulation of the fitting problem. Specifically, we minimize the distance between $\boldsymbol{\rho}$ and $\mathbf{R_0}(\boldsymbol{\psi})$ 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

$$\begin{array}{ll} \min_{\boldsymbol{\psi}} & S_{\mathbf{W}}(\boldsymbol{\psi}) \equiv \boldsymbol{\varepsilon}(\boldsymbol{\psi})' \mathbf{W} \boldsymbol{\varepsilon}(\boldsymbol{\psi}) \\ \text{subject to} & \boldsymbol{\psi} \in \boldsymbol{\Psi}. \end{array} \tag{4}$$

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

$$\begin{aligned} \boldsymbol{\Psi} = \{(\gamma, \delta, \lambda, \xi, \alpha_1, \alpha_2, \ldots, \alpha_p)' :\ & \\ & \delta > 0, \\ & \lambda \begin{cases} > 0 & \text{for } f = \text{S}_U, \\ > X_{(n)} - \xi & \text{for } f = \text{S}_B, \\ = 1 & \text{for } f = \text{S}_L \text{ and } f = \text{S}_N. \end{cases} \\ & \xi \begin{cases} < X_{(1)} & \text{for } f = \text{S}_L \text{ and } f = \text{S}_B, \\ = 0 & \text{for } f = \text{S}_N. \end{cases} \\ & \left|\text{RootOf}\left(1 - \textstyle\sum_{h=1}^{p} \alpha_h B^h = 0, B\right)\right| > 1\}, \end{aligned} \tag{5}$$

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 of the constraints (5) 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 (Cario and Nelson 1996).

Based on the experience of Swain et al. (1988) and Kuhl and Wilson (1999), we choose to use the diagonal weight matrix, $\mathbf{W} = \mathbf{D}$, defined as

$$\mathbf{D} = \text{diag}\left\{1/\text{Var}\left(\varepsilon_{p+1}(\boldsymbol{\psi}^*)\right), \ldots, 1/\text{Var}\left(\varepsilon_n(\boldsymbol{\psi}^*)\right)\right\}, \tag{6}$$

giving us the diagonally-weighted least-squares (DWLS) parameter estimators. In expanded form, the objective function of the DWLS least-squares estimation problem, using (4) and (6), can be written as

$$\begin{aligned} S_{\mathbf{D}}(\boldsymbol{\psi}) &= \boldsymbol{\varepsilon}(\boldsymbol{\psi})' \mathbf{D} \boldsymbol{\varepsilon}(\boldsymbol{\psi}) \\ &= \sum_{t=p+1}^{n} w(n, p, t) \left(\Phi\left\{V_{(t)}(\boldsymbol{\psi})\right\} - \rho_t\right)^2, \end{aligned} \tag{7}$$

where

$$w(n, p, t) = \frac{(n-p+1)^2(n-p+2)}{(n-p)^2(t-p)(n+1-t)}$$

and $\{V_t(\boldsymbol{\psi});\ t = p+1, p+2, \ldots, n\}$ is given by (3). Notice that the use of the uniformized order statistics for fitting permits a single formulation for not only Johnson-type distributions, but all continuous distributions, because the necessary first and second moments of $\{\Phi\{V_{(t)}(\boldsymbol{\psi})\};\ t = p+1, p+2, \ldots, n\}$, are known and easily computed.

3.2 ARTA Fitting Algorithm

One can minimize the objective function (7) subject to the constraints in (5) 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 between determining the Johnson parameters $(\gamma, \delta, \lambda, \xi)$ and the base-process parameters $(\alpha_1, \alpha_2, \ldots, \alpha_p)$. Further, we have empirically observed that solving $S_{\mathbf{D}}(\boldsymbol{\psi})$ for any given fixed feasible $\gamma, \delta, \lambda, \xi$ provides pretty robust estimates of $\alpha_1, \alpha_2, \ldots, \alpha_p$. Therefore, we work iteratively between improving the estimates for $(\gamma, \delta, \lambda, \xi)$ and $(\alpha_1, \alpha_2, \ldots, \alpha_p)$.

First, we define the solution set as the collection of parameters at which all of the entries of the gradient of the objective function $S_{\mathbf{D}}(\boldsymbol{\psi})$ attain the value of zero. Then, we start the algorithm with a parameter vector in the feasible region $\boldsymbol{\Psi}$ (5). Next we 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 Stage 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 Stage 2. Until the ARTA fitting algorithm reaches a point in the solution set, we work iteratively between Stage 1 and Stage 2, converging to a stationary point.

3.3 Properties of the ARTA Estimators

Suppose that $X_1, X_2, \ldots, X_n$ are identically distributed random variables with a joint ARTA distribution characterized by the parameter vector $\boldsymbol{\psi}^*$. Even if we assume that the type of the Johnson transformation f and the order of dependence p are known, the DWLS estimators $\hat{\psi}_n$ are not necessarily consistent. In the limit, the DWLS problem requires only that the empirical distribution of $R_t(\boldsymbol{\psi}) = \Phi\{V_t(\boldsymbol{\psi})\}$ converges to $U(0,1)$, the uniform distribution on $(0,1)$. This can occur at parameter settings other than $\boldsymbol{\psi}^*$, for instance at $(\gamma = \gamma^*, \delta = \delta^*, \lambda = \lambda^*, \xi = \xi^*, \alpha_h = 0, h = 1, 2, \ldots, p)$. Of course, in finite samples the joint distribution of the order statistics does matter in minimizing (7), and, as shown in Biller and Nelson (2002b), only at $\boldsymbol{\psi} = \boldsymbol{\psi}^*$ are the $R_t(\boldsymbol{\psi})$ i.i.d. $U(0,1)$. This accounts for the performance of our algorithm in recovering $\boldsymbol{\psi}^*$ when the true process is ARTA.

Since our goal is data modeling, rather than recovery of the true distribution, consistency is not a critical property. Nevertheless, it is desirable. Although the estimator $\hat{\boldsymbol{\psi}}_n$ is not consistent in general, it does have certain limited consistency properties that are of interest:

1. $\Pr[\hat{\lambda}_n \to \lambda^*] = 1$ and $\Pr[\hat{\xi}_n \to \xi^*] = 1$.
2. If $\alpha_h = \alpha_h^*, h = 1, 2, \ldots, p$, then $\Pr[\hat{\boldsymbol{\psi}}_n \to \boldsymbol{\psi}^*] = 1$.

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

We proposed decomposing the algorithm for solving the least-squares problem into two steps—improving the estimates of $(\gamma, \delta, \lambda, \xi)$ and improving the estimates of $(\alpha_1, \alpha_2, \ldots, \alpha_p)$—because we observed that the estimates of $(\alpha_1, \alpha_2, \ldots, \alpha_p)$ are robust to poor estimates of $(\gamma, \delta, \lambda, \xi)$. The second result shows that when we get these base process parameters right, the least-squares estimators of the remaining Johnson parameters are consistent. The derivation of the corresponding result is available in Biller and Nelson (2002b).

Notice also that when $p = 0$, the first stage of the ARTA fitting algorithm reduces to the one suggested by Swain et al. (1988) and it performs least-squares fitting by treating the given sample points as independent. If $\{X_t; t = 1, 2, \ldots, n\}$ were i.i.d. Johnson-type random variables with the parameter set $\boldsymbol{\psi}^* = (\gamma^*, \delta^*, \lambda^*, \xi^*)$, then the model with $p = 0$ would be correct and the transformed random variate $R_{(t)}\left(\boldsymbol{\psi}^*\right) = \Phi\left\{\gamma^* + \delta^* f[(X_{(t)} - \xi^*)/\lambda^*]\right\}$ would have the distribution of the t^{th} uniform order statistic on the unit interval $(0,1)$. Although Swain et al. (1988) show empirically that the suggested least-squares estimators provide a convenient computational method for fitting any member of the Johnson system when $p = 0$, they do not present any statistical properties of these estimators. The second result indicates that the fitting procedure of Swain et al. gives strongly consistent estimators of the Johnson parameters.

Since strongly consistent estimators of the parameters of Gaussian AR(p) processes are well known, why not incorporate them into our algorithm? Our motivation for the formulation (7) was to characterize the joint estimation of $(\gamma, \delta, \lambda, \xi, \alpha_1, \alpha_2, \ldots, \alpha_p)$ by a single objective that did not favor either the Johnson or the base process parameters, leading to a direct proof of the convergence of the numerical algorithm. Alternative formulations that lead to strongly consistent estimators of the entire vector of parameters $\boldsymbol{\psi}^*$ are a subject of future research.

4 IMPLEMENTATION

The implementation of the ARTA fitting algorithm gives rise to a number of issues including the determination of the initial values for input process and base process, characterized by $(f, \gamma, \delta, \lambda, \xi)$ and $\left(p, \alpha_1, \alpha_2, \ldots, \alpha_p\right)$, respectively, the optimization algorithms used to solve each stage of the algorithm, the assurance of the stationarity of the input process, and the positive definiteness of the autocorrelation structure of the base process.

A common procedure for identifying the type of transformation to use from the Johnson translation system is to compute the sample skewness and kurtosis, and then pick the family associated with that point on the (skewness, kurtosis) plane. Algorithm AS 99 developed by Hill et al. (1976) does this, for instance. However, we choose to fit all of the families and compare the goodness of the fits as bias and variability in the higher sample moments causes the likelihood of identifying the wrong transformation to be very high. Using the Algorithm AS 99, we choose such starting Johnson parameters that they are feasible for both the prespecified Johnson transformation and the given collection of data points.

Using the initial estimates for the Johnson parameters, $\hat{\gamma}$, $\hat{\delta}$, $\hat{\lambda}$, $\hat{\xi}$, we transform the input data $x_1, x_2, \ldots, x_n$ via $z_t = \hat{\xi} + \hat{\lambda} f\left[\left(x_t - \hat{\gamma}\right)/\hat{\delta}\right]$, $t = 1, 2, \ldots, n$. By treating the transformed data $\hat{z}_t$ as a sample of length n from a Gaussian autoregressive process, we fit an autoregressive model by least-squares estimation and, at the same time, we determine the order of dependence $\hat{p}$ via the Schwarz criterion, an asymptotically consistent AR order-selection method that has been quite popular in recent applied work. We take the resulting estimates as the starting parameter vector for the underlying base process.

We choose to perform the DWLS estimation using a Levenberg-Marquardt (LM) optimization algorithm (Marquardt 1963). Despite its convergence properties, the convergence of the LM algorithm in practice might be slow, and often this is the case. Therefore, if the termination criterion has not been satisfied in a prespecified number of iterations, then, we resort to the Nelder-Mead algorithm (Olsson 1974).

The ARTA fitting algorithm approximates the input process by a stationary ARTA model with a positive definite autocorrelation matrix. It maintains stationarity enforcing the constraint given as the last line of the definition in (5). Further, the fitted autocorrelation matrix is always positive definite, because the autocovariance function of a covariance stationary sequence with an autoregressive representation is positive definite (Fishman 1973).

5 AN ILLUSTRATIVE EXAMPLE

In this section, we approximate an input process that has been previously studied in Cario and Nelson (1998) by an ARTA model. They use 519 data points recorded at fixed time increments on a pressure variable of a continuous-flow production line at a large chemical manufacturing plant; these measurements exhibit strong series dependence.

Cario and Nelson (1998) chose the Weibull marginal distribution function. Since their software ARTAFACTS has no capabilities for fitting marginal distributions, they determined the parameters of the Weibull distribution with the aid of the Arena Input Analyzer (Rockwell Software) that assumes i.i.d. data and uses maximum likelihood estimation. In addition, they approximated the input autocorrelation structure using the estimated autocorrelation function of the raw data and assumed an order of dependence $p = 3$. We call this model “artafacts" and denote it by “AF" in Table 1.

Our software fit a Johnson unbounded distribution and an autocorrelation structure with $p = 2$ characterized by $(\gamma, \delta, \lambda, \xi) = (2.046, 3.151, 0.457, 1.217)$ and $(\alpha_1, \alpha_2) = (1.050, -0.342)$. We call this model “artafit," denote it by “ARF(p)" in Table 1, and use p to denote its order. The histogram of the time series and the plots of the probability density functions fitted by both models are given in Figure 1.

In the first two rows of Table 1, we report the Kolmogorov-Smirnov (KS) and Anderson-Darling (AD) test statistics comparing the empirical distribution function with the fitted distributions. The second column corresponds to the fit suggested by Cario and Nelson (1998), the third column corresponds to the Johnson fit under the assumption of independence, and the other columns correspond to Johnson fits under the assumptions of orders 1, 2, and 3, respectively. The fitted Johnson-type marginal distributions when $p = 1$ and $p = 2$ are statistically superior to the one suggested by Cario and Nelson (1998), particularly in capturing the tail behavior as indicated by the AD test statistics.

In order to check the goodness of the fit of the autocorrelation structure, we choose to compare the spectral distribution functions using the Kolmogorov-Smirnov criterion (Anderson 1993). The corresponding test statistics are provided on the last row of Table 1. Although the Johnson-type fits with $p = 1$ and $p = 2$ have the same KS

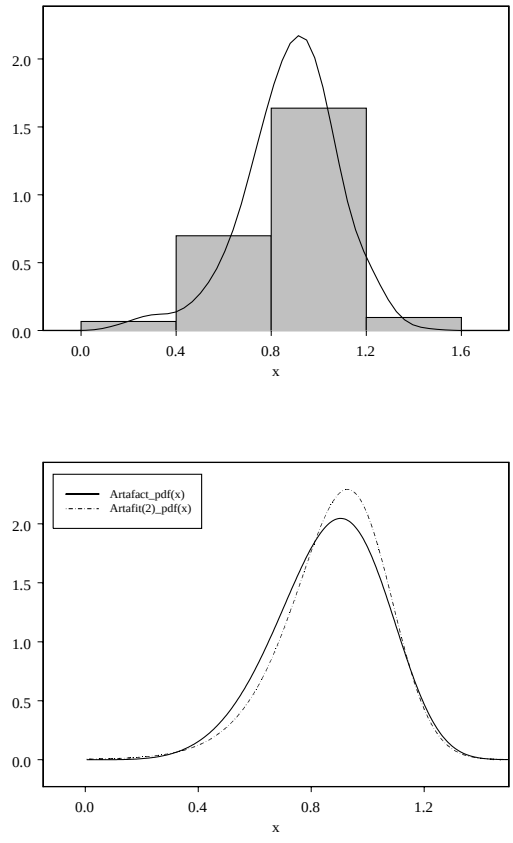

Figure 1: Histogram of the Empirical Pressure Data (with Nonparametric Density Estimate Superimposed) and the Probability Density Functions for Fitted Weibull and Johnson Unbounded Distributions

Table 1: Comparison of the Kolmogorov-Smirnov, Ander-son-Darling, and KS Spectral Test Statistics

	AF	ARF(0)	ARF(1)	ARF(2)	ARF(3)
KS_X	1.764	1.038	0.841	0.841	1.538
AD_X	3.295	0.758	0.758	0.755	2.563
KS_{ρ_X}	0.004	0.509	0.204	0.095	0.099

and AD test statistics, the one with $p = 2$ has a significantly smaller spectral test statistic, providing a better fit for the autocorrelation structure of the process. At the same time, the artafacts autocorrelations give significantly better fits for the sample autocorrelations than the autocorrelations of the artafit(2) model. However, a pure correlation match is not the only thing that matters while choosing a good representation for the underlying system. This will be clear in the the visual analysis of the time-series plots in Figures 2-4 and scatter plots in Figures 5-7.

Next, using the artafit and artafacts models, we generate 519 data points. Figures 2, 3, and 4 display the time-series plots while Figures 5, 6, and 7 provide 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 artafacts and artafit models. The sample paths are qualitatively similar, although we observe differences that can be partly attributed to the sampling error. In the empirical time series, there appear spikes that cannot be captured by the artafacts model which varies more consistently about its mean. In addition, comparison of Figure 5 and 6 shows that the artafacts data appears to be more scattered or random than the empirical data. Thus, the marginal distribution and autocorrelation structure of the artafacts process do not perform well in capturing the characteristics of the time-series process. However, the artafit(2) model appears to be more successful in representing the characteristics of the empirical data. Comparison of Figures 2 and 4 shows that the artafit(2) model captures the height of the spikes reasonably well, but the number of spikes generated by the artafit(2) process is not as large as the number of spikes generated by the empirical time series. However, the comparison of Figures 5 and 7 shows that the artafit(2) process is still reasonably successful in capturing the autocorrelation structure of the empirical time-series process. Overall, the artafit(2) process provides a plausible model for the empirical time series.

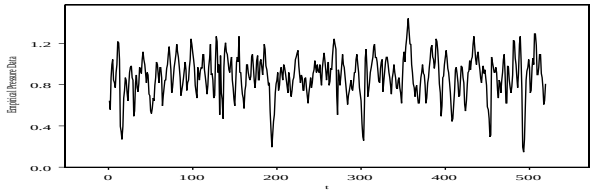

Figure 2: Time-series Plot of the Empirical Data

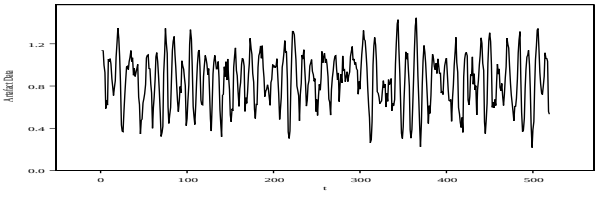

Figure 3: Time-series Plot of the Artafacts Data

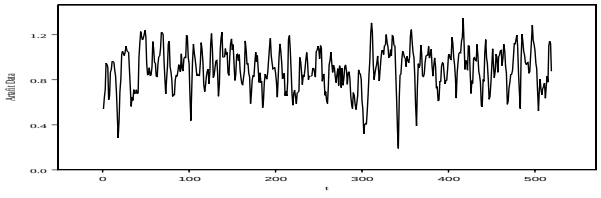

Figure 4: Time-series Plot of the Artafit Data

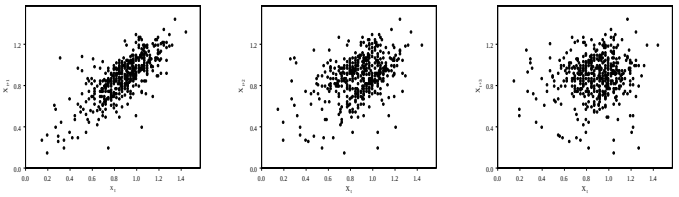

Figure 5: Scatter Plots of (x_t, x_{t+1}), (x_t, x_{t+2}), (x_t, x_{t+3}) for the Empirical Pressure Data

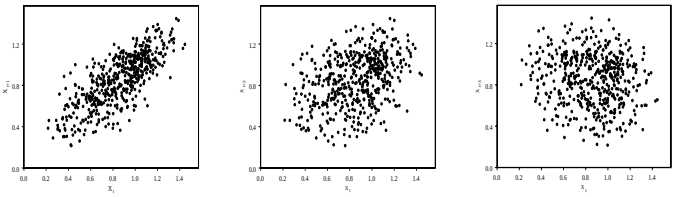

Figure 6: Scatter Plots of (x_t, x_{t+1}), (x_t, x_{t+2}), (x_t, x_{t+3}) for the Artafact Data

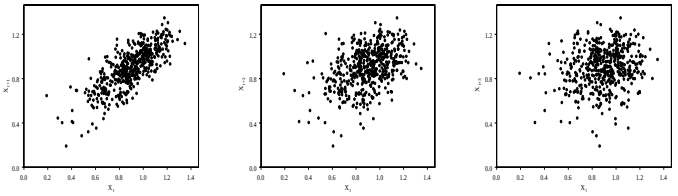

Figure 7: Scatter Plots of (x_t, x_{t+1}), (x_t, x_{t+2}), and (x_t, x_{t+3}) for the Artafit Data

6 CONCLUSION

In this paper, we propose an algorithm to fit stochastic input models to dependent univariate time-series processes. In order to demonstrate the use of the algorithm, we fit an input model to data generated by a physical process. A comprehensive empirical analysis of the ARTA fitting algorithm is reported in Biller (2002).

Recently, Biller and Nelson (2002a) 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. Their approach is very similar to the one in Cario and Nelson (1996), but they 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 of our future research.

ACKNOWLEDGMENTS

This research was partially supported by National Science Foundation Grant numbers DMI-9821011 and DMI-9900164 and Sigma Xi Scientific Research Society Grant number 142.

REFERENCES

Anderson, T. W. 1993. Goodness of fit tests for spectral distributions. *The Annals of Statistics* 21 (2): 830–847.

Biller, B. 2002. A Comprehensive Input-Modeling Framework for Stochastic Discrete-Event Simulation Experiments. Doctoral Dissertation, Department of Industrial Engineering and Management Sciences, Northwestern University, IL.

Biller, B., and B. L. Nelson. 2002a. Modeling and generating multivariate time series with arbitrary marginals using a vector autoregressive technique. *ACM TOMACS*. Forthcoming.

Biller, B., and B. L. Nelson. 2002b. Fitting ARTA input processes for simulation. Working paper, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois.

Cario, M. C., and B. L. Nelson. 1996. Autoregressive to anything: Time-series input processes for simulation. *Operations Research Letters* 19 (1): 51–58.

Cario, M. C., and B. L. Nelson. 1998. Numerical methods for fitting and simulating autoregressive-to-anything processes. *INFORMS Journal on Computing* 10 (1): 72–81.

Cario, M. C., B. L. Nelson, S. D. Roberts, and J. R. Wilson. 2001. Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix. Working Paper, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL.

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

Chen, H. 2001. Initialization for NORTA: Generation of random vectors with specified marginals and correlations. *INFORMS Journal on Computing* 13 (4): 312–331.

Fishman, G. S. 1973. *Concepts and Methods in Discrete-Event Digital Simulation*. New York: John Wiley and Sons.

Hill, I. D., R. Hill, and R. L. Holder. 1976. Fitting Johnson curves by moments. *Applied Statistics* 25 (2): 180–189.

Johnson, M. E. 1987. *Multivariate Statistical Simulation*. New York: John Wiley and Sons.

Johnson, N. L. 1949. Systems of frequency curves generated by methods of translation. *Biometrika* 36 (1): 297–304.

Kendall, M. G., and A. Stuart. 1979. *The advanced theory of statistics*. New York: Macmillan.

Kuhl, M. E., and J. R. Wilson. 1999. Least-squares estimation of non-homogeneous Poisson processes. *Journal of Statistical Computation and Simulation* 67: 75–108.

Lewis, P. A. W., E. McKenzie, and D. K. Hugus. 1989. Gamma processes. *Commun. Statist. - Stochastic Models* 5 (1): 1–30.

Livny, M., B. Melamed, and A. K. Tsiolis. 1993. The impact of autocorrelation on queueing systems. *Management Science* 39 (3): 322–339.

Mallows, C. L. 1967. Linear processes are nearly Gaussian. *Journal of Applied Probability* 4 (2): 313–329.

Marquardt, D. W. 1963. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the SIAM* 11 (2): 431–441.

Melamed, B., J. R. Hill, and D. Goldsman. 1992. The TES methodology: Modeling empirical stationary time series. In *Proceedings of the 1992 Winter Simulation Conference*, ed. J. J. Swain, D. Goldsman, R. C. Crain, and J. R. Wilson, 135–144. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Olsson, D. M. 1974. A sequential simplex program for solving minimization problems. *Journal of Quality Technology* 6 (1): 53–57.

Song, W. T., L. Hsiao, and Y. Chen 1996. Generating pseudorandom time series with specified marginal distributions. *European Journal of Operational Research* 93 (1): 1–12

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

Ware, P. P., T. W. Page, and B. L. Nelson. 1998. Automatic modeling of file system workloads using two-level arrival processes. *ACM Transactions on Modeling and Computer Simulation* 8 (3): 305–330.

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

AUTHOR BIOGRAPHIES

BAHAR BILLER is an assistant professor of Operations Management and Manufacturing at Carnegie Mellon University. She received her Ph.D. from Northwestern University. Her research interests are in computer simulation of stochastic systems and stochastic input modeling.

BARRY L. NELSON is the Krebs Professor of Industrial Engineering and Management Sciences at Northwestern University, and is Director of the Master of Engineering Management Program there. His research centers on the design and analysis of computer simulation experiments on models of stochastic systems. He has published numerous papers and two books. Nelson has served the profession as the Simulation Area Editor of *Operations Research* and President of the INFORMS (then TIMS) College on Simulation. He has held many positions for the Winter Simulation Conference, including Program Chair in 1997 and current membership on the Board of Directors. His e-mail and web addresses are <nelsonb@northwestern.edu> and <www.iems.northwestern.edu/~nelsonb/>.