Conf-940806--1

PNL-SA-24222

MODELS FOR RECURRENT GAS RELEASE EVENT BEHAVIOR IN HAZARDOUS WASTE TANKS

D. N. Anderson
B. C. Arnold[a]

August 1994

Presented at the
6th International Congress of Ecology Conference
August 20, 1994
Manchester, United Kingdom

Prepared for
the U.S. Department of Energy
under Contract DE-AC06-76RLO 1830

Pacific Northwest Laboratory
Richland, Washington 99352

(a) Department of Statistics, University of California, Riverside, California

Models for Recurrent Gas Release Event Behavior in Hazardous Waste Tanks

Dale N. Anderson[1] and Barry C. Arnold[2]

Abstract

Certain radioactive waste storage tanks at the United States Department of Energy Hanford facilities continuously generate gases as a result of radiolysis and chemical reactions. The congealed sludge in these tanks traps the gases and causes the level of the waste within the tanks to rise. The waste level continues to rise until the sludge becomes buoyant and "rolls over", changing places with heavier fluid on top. During a rollover, the trapped gases are released, resulting in a sudden drop in the waste level. This is known as a gas release event (GRE). After a GRE, the waste re-congeals and gas again accumulates leading to another GRE. We present nonlinear time series models that produce simulated sample paths that closely resemble the temporal history of waste levels in these tanks. The models also imitate the random GRE behavior observed in the temporal waste level history of a storage tank. We are interested in using the structure of these models to understand the probabilistic behavior of the random variable "time between consecutive GRE's". Understanding the stochastic nature of this random variable is important because the hydrogen and nitrous oxide gases released from a GRE are flammable and the ammonia that is released is a health risk. From a safety perspective, activity around such waste tanks should be halted when a GRE is imminent. With credible GRE models, we can establish time windows in which waste tank research and maintenance activities can be safely performed.

1 Pacific Northwest Laboratory, operated for the U.S. Department of Energy by Battelle Memorial Institute under contract DE-AC06-76RLO 1830
2 Department of Statistics, University of California, Riverside, CA 92521

1.0 Introduction

The mission of the United States Department of Energy (DOE) is currently in a post cold war transition from nuclear weapons research and production to stabilizing and storing the hazardous wastes generated by these earlier activities. At the DOE Hanford facilities there are radioactive waste storage tanks that continuously generate gases as a result of radiolysis and chemical reactions. The congealed sludge in these tanks traps the gases and causes the level of the waste within the tanks to rise. The waste level continues to rise until the sludge becomes buoyant and "rolls over", changing places with heavier fluid on top. During a rollover, the trapped gases are released, resulting in a sudden drop in the waste level. This is known as a gas release event (GRE). After a GRE the waste re-congeals and gas again accumulates, which eventually results in another GRE. Figure 1.1 presents the temporal waste level history of a radioactive storage tank at the Hanford facilities. The methods presented in this paper are applied to these data. Note the evident GRE characteristics in this time series. The recurrent GRE behavior that is illustrated in Figure 1.1 is the focus of this paper.

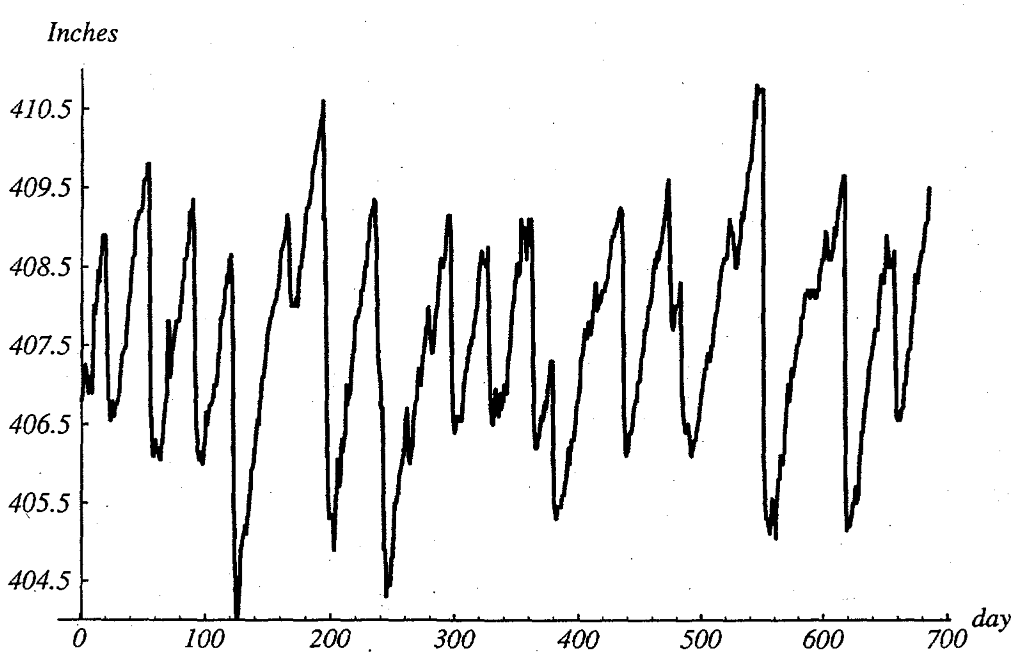

Figure 1.1. This figure presents a time series plot of the daily waste level measurements from a waste storage tank at the DOE Hanford facilities. The GRE behavior of these data is characterized by a steady increase in waste level followed by sharp level drop.

We will present nonlinear time series models that produce simulated sample paths which closely resemble the temporal waste level history of a waste tank. This modeling approach is analogous to matching a density function to a histogram. We concede that this approach is data driven, but the models presented have a simple yet sophisticated structure that may aid in describing the physical processes that create recurrent GRE behavior. The structure of each model postulates a particular description of the statistical mechanisms that produce GRE behavior. These models provide a credible foundation which we may use to address the issues associated with GRE behavior.

Ultimately we are interested in using the models in this paper to understand the probabilistic behavior of the random variable "time to a GRE". The random variable "time to a GRE" is sensibly a function of the current level of the waste in a tank. If the waste level is high, then a considerable amount of gas has accumulated and a GRE is imminent. Consequently, we denote the random variable "time to a GRE" as $T(W = w)$ where w is the current waste level. Understanding the stochastic nature of $T(W = w)$ is important because the gasses released in a GRE are flammable, and toxic. From a safety perspective, activity in and around the tank should be curtailed when a GRE is imminent. With a credible model for $T(W = w)$, we can establish time windows in which waste tank research and maintenance activities can be safely performed.

In Section 2 we present temporal waste level models based on a statistical property of the logistic probability distribution. The basic model was first introduced by Yeh, Arnold, and Robertson(1988), and enhanced by Arnold and Robertson(1989). We include in Section 2 some parameter estimation schemes and we apply these techniques to the data presented in Figure 1.1. The probability structure of the random variable $T(W = w)$ must be obtained through simulation for the model in Section 2. The simplicity of the model permits easy implementation of this simulation. The only real constraint associated with simulating the stochastic structure of $T(W = w)$ is one of computing time. Included in the application example is a simulation of the probability structure of $T(W = w)$.

In Section 3, a simple enhancement of the Tavares(1980) exponential process is presented as a temporal waste level model. The Tavares model is based on a simple characterization of the exponential probability distribution. The structure of the Tavares model is such that the probability distribution of $T(W=w)$ is mathematically tractable. We sketch the development of the probability structure of $T(W=w)$ under the Tavares model and apply the model to the temporal waste level history analyzed in Section 2. Parameter estimation schemes for the Tavares model are also presented in Section 3.

Both of the models presented in Sections 2 and 3 have the requisite structure necessary to model the GRE behavior of a waste tank. In Section 4 we comment on how to identify which of these two models provides the best fit. We also note an alternative parameter estimation scheme for the Tavares model based on the probability structure of $T(W=w)$. The models in this paper may be used to detect changes in the waste level behavior of a storage tank. The details of these ideas are also presented in Section 4. We conclude the paper with references for other potential waste level models.

2.0 Logistic Temporal Waste Level Model

A random variable W is said to have a *logistic*(μ, σ) distribution if the survival function of W is

$$P(W > w) = (1 + e^{(w-\mu)/\sigma})^{-1}. \tag{2.1}$$

The mean of a *logistic*(μ, σ) random variable is μ and the variance is $\sigma^2\pi^2/3$. The logistic distribution has a symmetric bell shape and is often used as an alternative to the normal distribution. The tails of the logistic distribution are heavy in the sense that for logistic and normal distributions with equal variances centered at zero, $P(W > w)$ for $w \gg 0$ is always largest for the logistic. The heavy tail of the logistic distribution also implies a high kurtosis (peaked nature). These properties are illustrated in Figure 2.1.

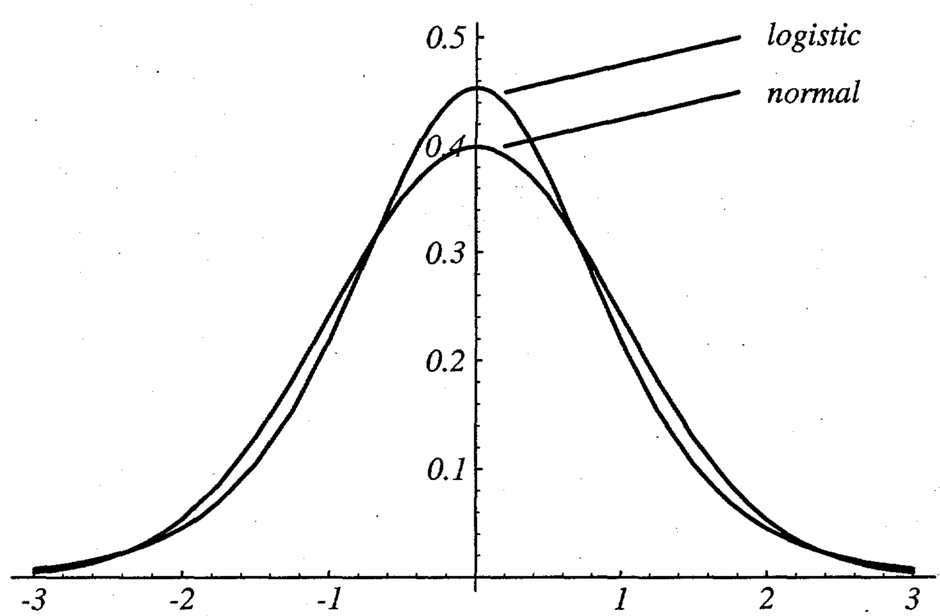

Figure 2.1. This figure illustrates the differences in tail behavior and kurtosis between the logistic and normal distributions. The graphs are of normal and logistic density functions, each with a variance of one and a mean of zero.

Let W_t be the waste level on day t and define $X_t = \gamma(W_t)$. The function $\gamma(w)$ is a monotonically increasing function that preserves the GRE behavior of the waste level history. $\gamma(w)$ has no real physical significance and is simply a tool to transform the probability structure of W_t to a shape that can be easily modeled with the models presented in Sections 2 and 3. Note that

$X_t = W_t$ when $\gamma(w) = w$. Denote the time series of observed waste levels by $\{w_t\}$, $t = 1, 2, 3, \ldots, n$. From the series $\{w_t\}$ we use the function $\gamma()$ to obtain the time series $\{x_t\}$. Again, the transformation γ is chosen to retain the GRE characteristics of the series $\{w_t\}$ and permit the marginal probability structure of $\{x_t\}$ to be modeled with the *logistic(0, 1)* distribution. Figure 2.2 is a time series plot of the daily waste level measurements (w_t in inches) of a storage tank at the Hanford facility. The 685 level measurements span the period January 1, 1982 to November 16, 1983. Included in Figure 2.2 is a histogram representing the marginal probability structure of these data.

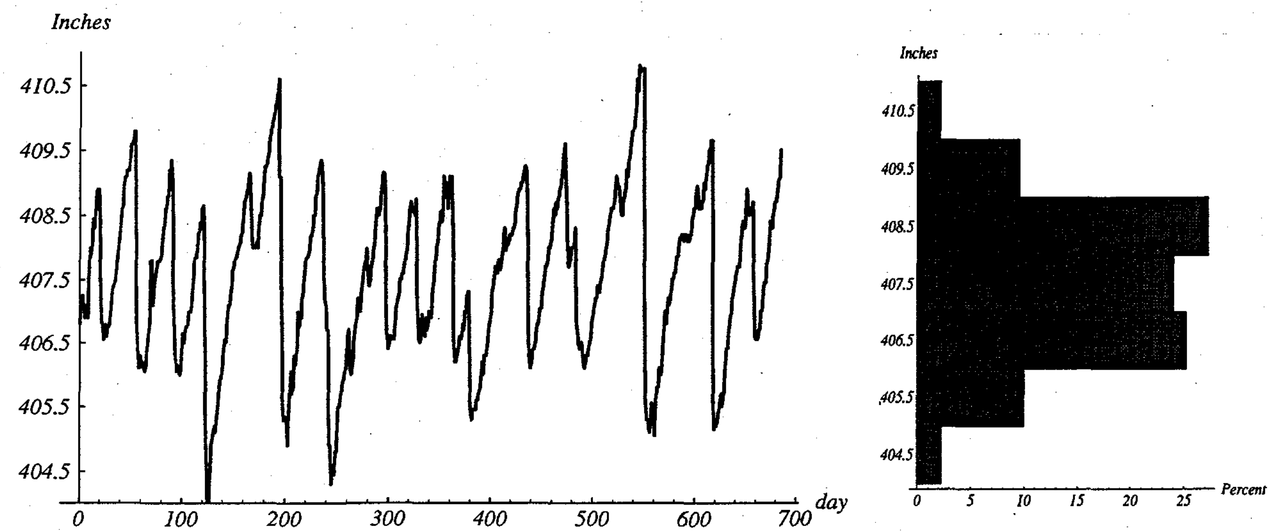

Figure 2.2. This figure presents a time series plot of the daily waste level measurements from a waste storage tank at the DOE Hanford facilities. The histogram represents the marginal (across time) probability structure of the data.

Evidently, the stationary distribution of these data might be adequately modeled with a *logistic*(μ, σ) distribution. We estimate the logistic parameters μ and σ by applying iid logistic maximum likelihood equations to the data . Thus, the level measurements w_t are transformed via

$$x_t = \gamma(w_t) = \frac{w_t - 407.562}{0.757483}. \quad (2.2)$$

To illustrate the fit of the logistic distribution to the data, we present in Figure 2.3 an overlay plot of the marginal empirical cdf and the fitted logistic cdf.

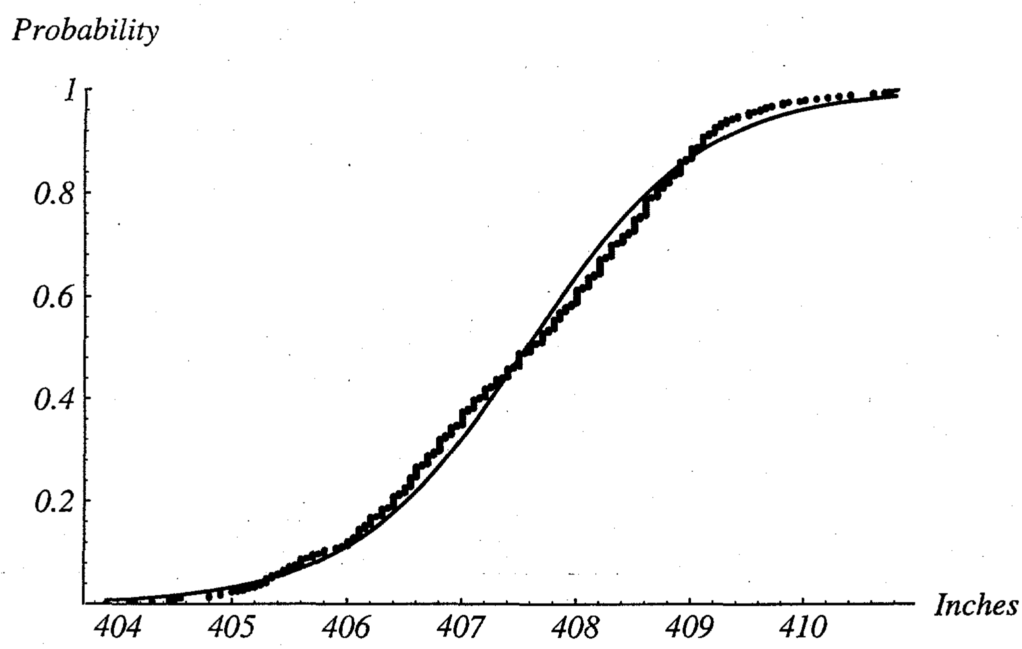

Figure 2.3. This figure is an overlay plot of the empirical cdf (•••) for the data in Figure 2.4 and the fitted logistic cdf (—). The parameters of the logistic cdf are estimated with the iid maximum likelihood equation.

Next we describe and fit a logistic model to the observed series x_t. Let Z_t be iid *logistic(0,1)* random variables and define W_t and X_t as above. Independent of the sequence $\{Z_t\}$, let B_t be iid random variables with cumulative distribution function (cdf)

$$G(\beta) = \beta^{\delta};\ \beta \in (0,1),\ \delta > 0. \tag{2.3}$$

The power-logistic temporal waste level model, introduced by Arnold and Robertson(1989), is

$$X_0 = Z_0$$

$$X_t = \begin{cases} X_{t-1} - log(B_t) & with\ probability\ B_t \\ min(X_{t-1} - log(B_t),\ Z_t) & with\ probability\ 1-B_t \end{cases},\ t = 1,\ 2,\ 3,\ldots \tag{2.4}$$

Equation (2.4) defines a stationary *logistic(0,1)* process. The autocorrelation structure for the power-logistic model is quite difficult to represent algebraically. However, we can characterize the sample path behavior of the model with a sample path fluctuation probability. Specifically it can be shown that

$$P(X_{t+1} > X_t) = \frac{2\delta+1}{2\delta+2}. \tag{2.5}$$

Figure 2.4 is an illustration of the sample path characteristics of the power-logistic process.

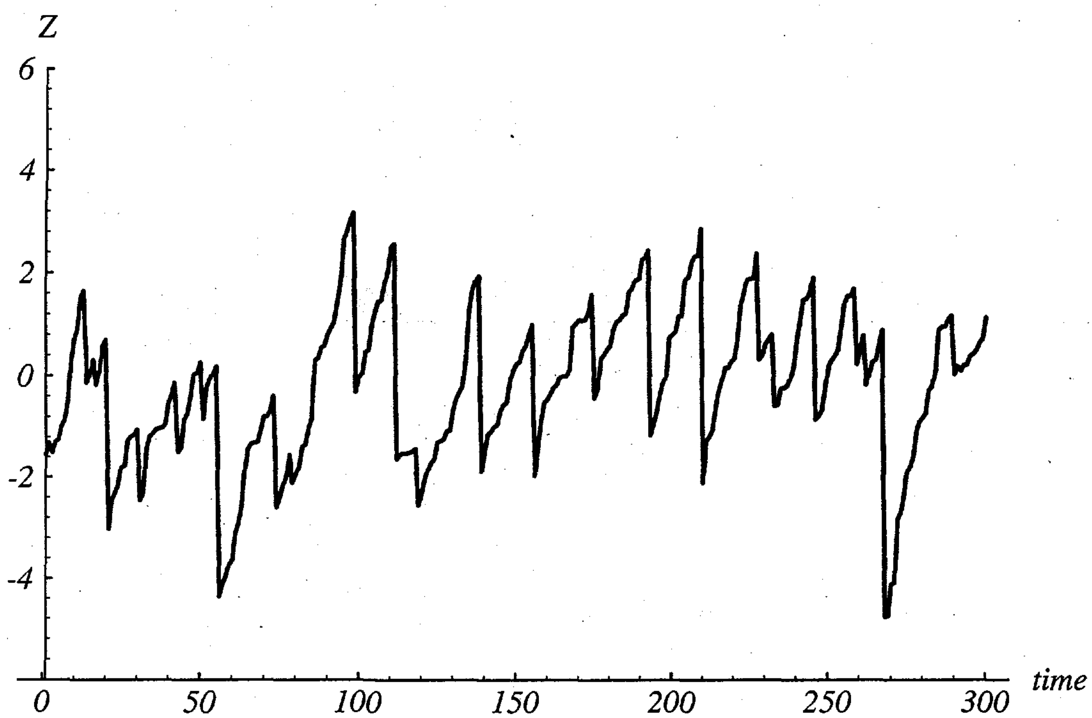

Figure 2.4. This figure illustrates through simulation the sample path of the power-logistic temporal waste level model defined by Equation (2.6). The value of δ value for the simulation is 4, which corresponds to a mean B_t value of 0.8.

We model the transformed series $\{x_t\}$ with the power-logistic process. To estimate the parameter δ we appeal to Equation (2.5) which relates δ to the sample path fluctuation probability. Based on a count of 70 drops in 685 steps we obtain a method of moments estimate $\hat{\delta}$ of 3.89. The 70 drops that yield this estimate include small drops as well as obvious GREs. The small drops may be minor GREs. Including small drops in the estimate $\hat{\delta}$ will yield a cautious or smaller activity window for a tank.

To obtain the probability structure of $T(W = w)$, the time to a GRE, we must resort to a simulation. According to the power-logistic model, a GRE occurs at time t if and only if the value of X_t is calculated from the model branch

$$min(X_{t-1} - log(B_t), Z_t) \quad (2.6)$$

and $min(X_{t-1} - log(B_t), Z_t) = Z_t$. To simulate $T(W = w)$, we simply chose or fix $W = w$, simulate the power-logistic process conditional on w, and count the number of days to a model-defined GRE. Such a simulation will yield data of the form $t_i(w)$, $i = 1, 2, 3, \ldots n$. For large n, an accurate empirical distribution for $T(W = w)$ can be obtained.

Figure 2.5 presents the average and 20th percentile of the simulated distribution of $T(W = w)$ for a suite of waste levels w. The value of w is represented on the abscissa as a percentile from the *logistic(0,1)* distribution. For this simulation, $n = 5000$ and the estimated value of $\delta = 3.89$ is used. As expected, the larger percentile (higher waste level w) yields smaller values of $T(W = w)$. If we identify only a large drop as a GRE, then we obtain a count of 22 drops in 685 steps, which yields a method of moments estimate $\hat{\delta}$ of 14.58. This parameter value will yield a larger activity window for the tank. Figure 2.6 presents the average and 20th percentile of the simulated distribution of $T(W = w)$ with $\delta = 14.58$. Again w is represented as a percentile from the *logistic(0,1)* distribution) and $n = 5000$.

To establish a safe activity window for this tank we may use the more cautious simulated distribution presented in Figure 2.5. Suppose a GRE yields a large drop (1 percentile of waste levels) in the tank waste level. From Figure 2.5, the conservative average time to the next GRE is about 20 days. We may also claim a probability of 0.8 that an activity window will be at least 12 days. This is a very short window. However, it is important to remember that the distribution presented in Figure 2.5 accounts for GREs of all sizes. We might be concerned only with large GREs. In this case, the less cautious simulated distribution presented in Figure 2.6 may be more appropriate. From Figure 2.6, the average time to the next GRE is about 70 days, while with probability 0.8, an activity window will be at least 50 days.

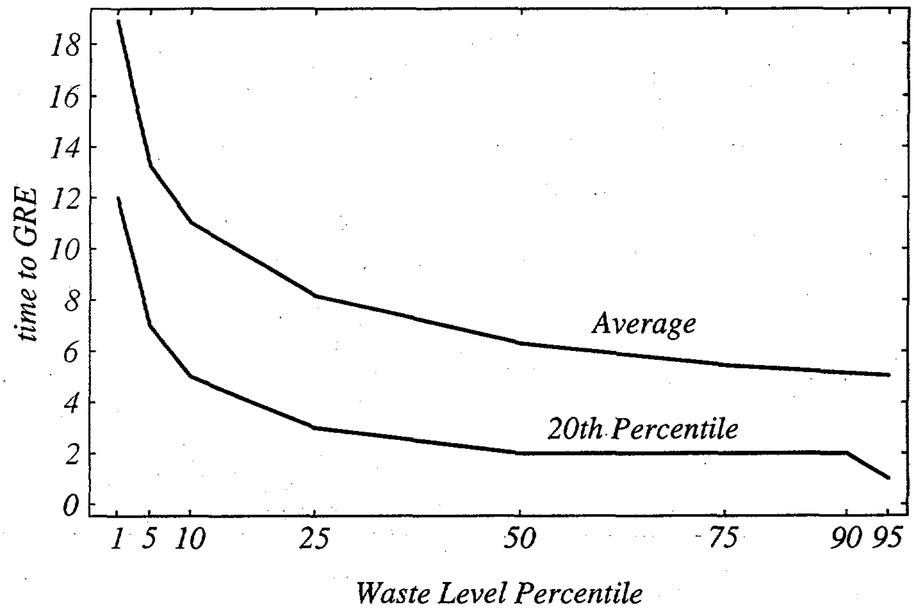

Figure 2.5. This figure presents the mean and 20th percentile of the simulated distribution of $T(W = w)$, where w is represented as a percentile from the *logistic(0,1)* distribution. For this simulation, $n = 5000$ values of T were generated at each waste level percentile. The value of $\delta = 3.89$.

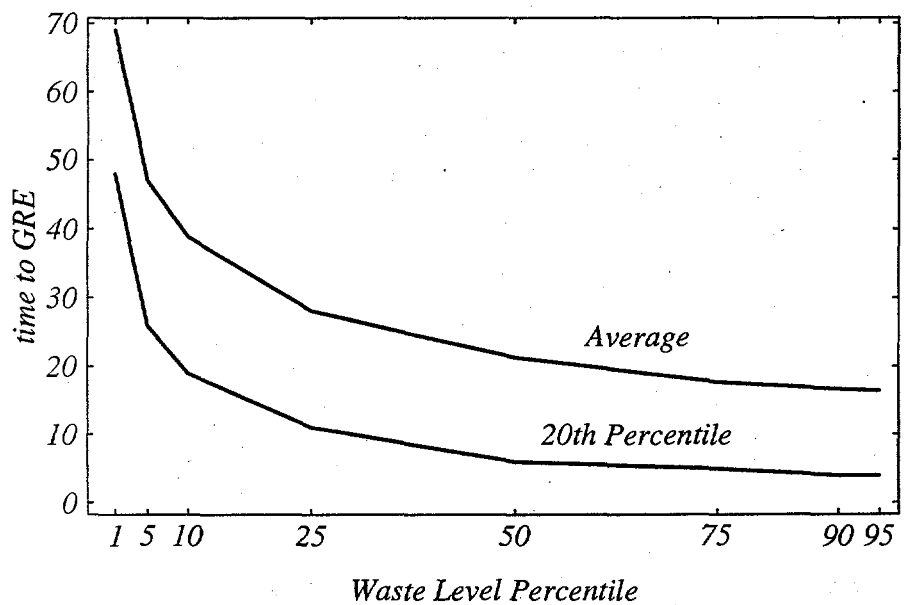

Figure 2.6. This figure presents the mean and 20th percentile of the simulated distribution of $T(W = w)$, where w is represented as a percentile from the *logistic(0,1)* distribution. For this simulation, $n = 5000$ values of T were generated at each waste level percentile. The value of $\delta = 14.58$.

An alternative estimation scheme is possible by applying conditional least squares methods introduced by Klimko and Nelson (1978). Denote the expected value of X_t conditional on X_{t-1} by $E(X_t / X_{t-1} = x)$. For the power-logistic process this quantity is complicated but tractable. If we define the hypergeometric function by

$$ {}_2F_1(a, b, c, z) = \frac{\Gamma(c)}{\Gamma(b)\, \Gamma(c-b)} \int_0^1 t^{b-1}(1-t)^{c-b-1}(1-tz)^{-a} dt, \tag{2.7} $$

we have

$$ E(X_t / X_{t-1} = x) = \frac{\delta}{(1+\delta)^2} + x + \frac{{}_2F_1(1, 1+\delta, 2+\delta, -e^{-x})}{(1+\delta)\, e^x} - \frac{\delta\, {}_2F_1(1, 1+\delta, 2+\delta, -e^{-x})}{(1+\delta)\,(2+\delta)\, e^x} + $$

$$ \log(e^{-x}) - \frac{\delta \log(e^{-x})}{1+\delta} - \frac{\log(1+e^{-x})}{1+\delta} - \frac{\delta \log(1+e^{-x})}{1+\delta} + \frac{2\delta \log(1+e^{-x})}{(1+\delta)\,(2+\delta)} + \frac{\delta^2 \log(1+e^{-x})}{(1+\delta)\,(2+\delta)} \tag{2.8} $$

The conditional least squares objective function is given by

$$ Q(\delta) = \Sigma_{i=2,n} (x_i - E(X_i / X_{i-1} = x_{i-1}))^2 \tag{2.9} $$

The value of δ that minimizes $Q(\delta)$ is a strongly consistent estimate and is known as the conditional least squares estimate of δ. For the data used in this paper, minimizing $Q(\delta)$ yields an estimate $\tilde{\delta} = 8.3$ and $Q(\tilde{\delta}) = 198.7$. Figure 2.7 gives the average and 20th percentile of the simulated distribution of $T(W = w)$ using $\tilde{\delta} = 8.3$.

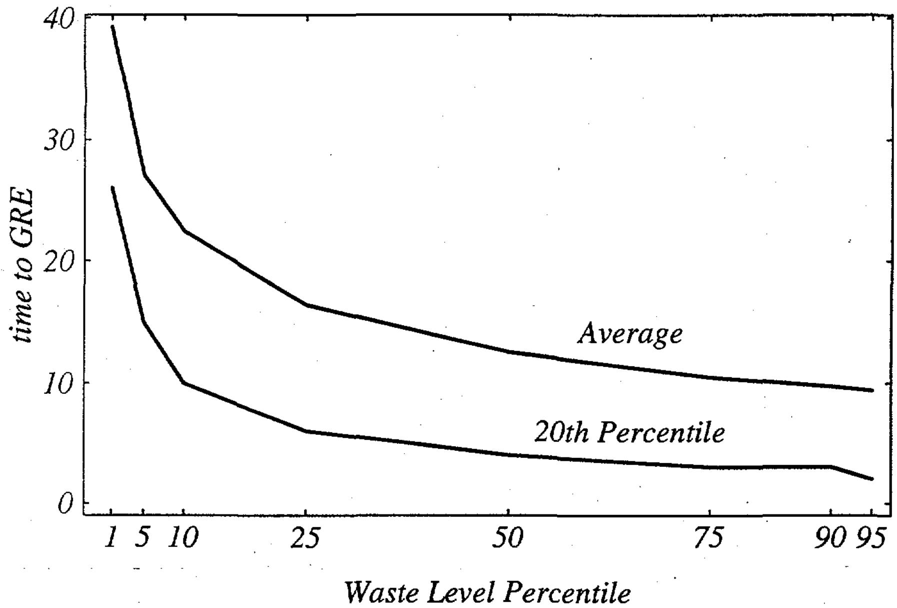

Figure 2.7. This figure presents the mean and 20th percentile of the simulated distribution of $T(W = w)$, where w is represented as a percentile from the *logistic(0,1)* distribution. For this simulation, $n = 5000$ values of T were generated at each waste level percentile. The value of $\delta = 8.3$.

3.0 Exponential Temporal Waste Level Model

A random variable W is said to have a $Weibull(\theta,\sigma,c)$ distribution if the survival function of W is

$$P(W > w) = e^{-((w-\theta)/\sigma)^c}. \tag{3.1}$$

The mean of W is $\theta+\sigma\Gamma(1+1/c)$ and the variance is $\sigma^2(\Gamma(1+2/c)-\Gamma(1+2/c)^2)$. If W is a $Weibull(\theta,\sigma,c)$ distribution then $X = ((W-\theta)/\sigma)^c$ is an exponential ($exp(\mu)$) random variable with $\mu = 1$ and

$$P(X > x) = 1-e^{-\mu x}. \tag{3.2}$$

The shape of the density functions for the $exp(\mu)$ and $Weibull(\theta,\sigma,c)$ are illustrated in Figure 3.1. The Weibull distribution has the structure and parameterization to model data that is roughly bell-shaped with a skewed tail. Since the marginal distribution of an actual waste level history may be skewed, the Weibull distribution is an appealing alternative to the logistic distribution.

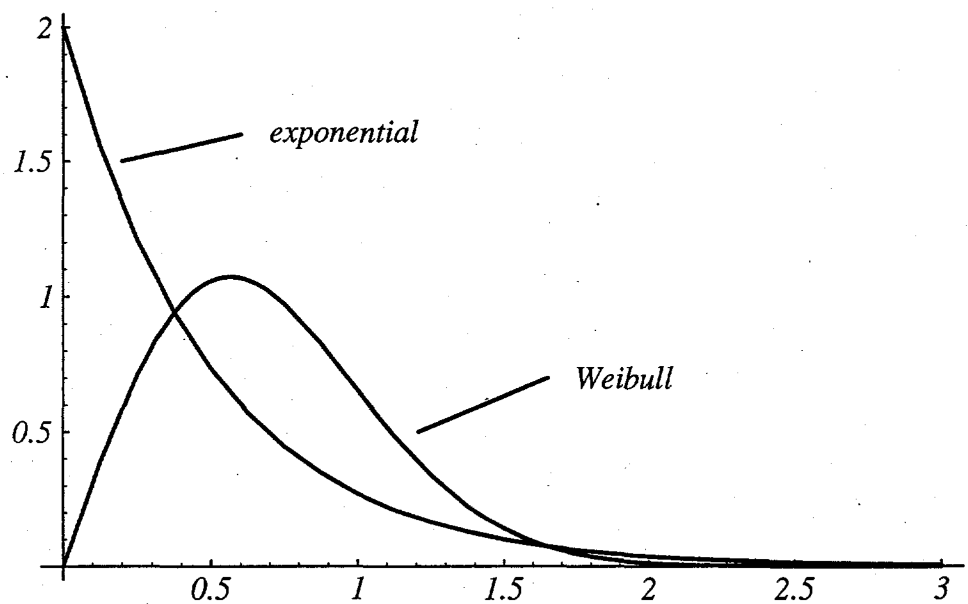

Figure 3.1. This figure illustrates the differences in shape between the exponential and Weibull density functions. For presentation purposes, both densities have support that begins at 0. The parameter values used to produce the graphs are immaterial and were chosen simply to contrast the potential differences in the shape of the two density functions.

The model that we present in this section has a stationary *exp(1)* distribution. The approach here is to model the probability structure of the actual waste levels w_t with the *Weibull*(θ,σ,c) distribution rather than the logistic as in Section 2. We then use the relationship between the Weibull and exponential distributions to transform the waste levels via

$$x_t = \gamma(w_t) = ((w_t - \theta)/\sigma)^c. \tag{3.3}$$

The x_t can then be modeled with the exponential process presented in this section. We first estimate the Weibull parameters θ, σ, and c by applying the iid Weibull maximum likelihood equations to the data w_t. Thus the level measurements w_t, presented in Figure 2.2, are transformed via

$$x_t = \gamma(w_t) = ((w_t - 403)/5)^{3.95}. \tag{3.4}$$

We illustrate the fit of the Weibull distribution in Figure 3.2 with an overlay plot of the marginal empirical cdf and the fitted Weibull cdf. Compare the fit of the Weibull distribution in Figure 3.2 with the logistic distribution fit in Figure 2.3. Evidently the Weibull distribution can better model the slight skewness in the data. In Figure 3.3 we give a time series plot of the transformed series obtained by Equation 3.4.

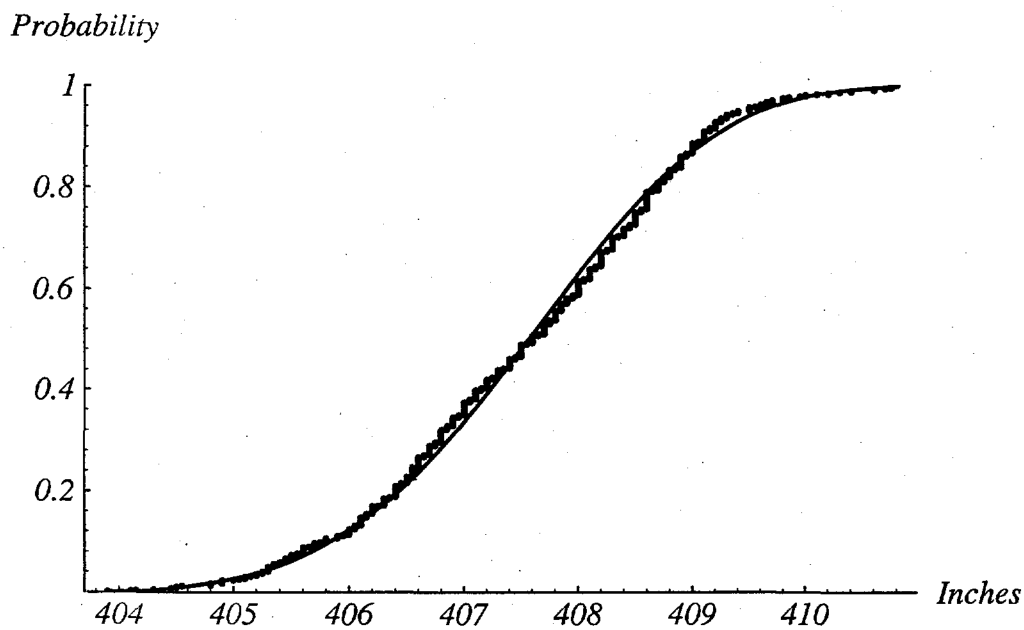

Figure 3.2. This figure is an overlay plot of the empirical cdf (•••) for the data in Figure 2.2 and the fitted Weibull cdf (—). The parameters of the Weibull cdf are estimated with the iid maximum likelihood equation.

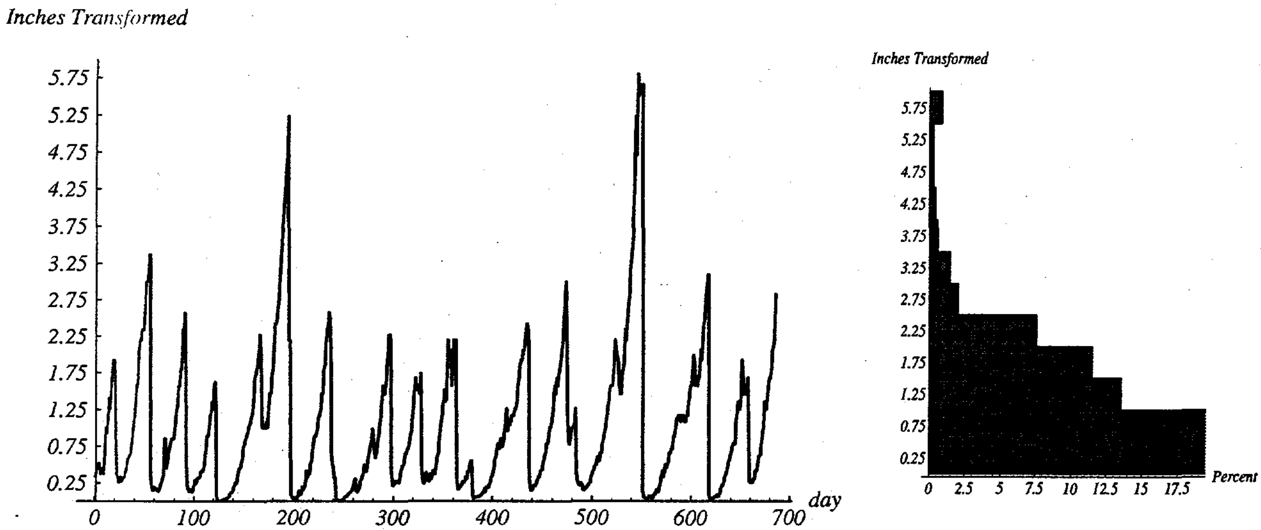

Figure 3.3. This figure presents a time series plot of the daily waste level measurements presented in Figure 2.2 and transformed by Equation 3.4.

As noted earlier, the approach in this section involves first modeling the actual waste level w_t with the *Weibull*(θ,σ,c) distribution. We may then obtain the time series x_t through the transformation (3.3). This time series will then be modeled with a marginal *exp(1)* distribution. We now present a model for the series x_t. Let Z_t be iid exponential variates with density function

$$f_{Z_t}(z) = \mu e^{-\mu z}. \tag{3.5}$$

Further, let $\{\varepsilon_t \,/ Z_t\}$ be a sequence of iid exponential variates with mean $1 / Z_t$. Thus $\{Z_t, \varepsilon_t\}$ is an iid sequence of bivariate random variables. Independent of the sequence $\{Z_t, \varepsilon_t\}$, let ε_0 be distributed *exp(1)*. A stationary generalization of the Tavares(1980) exponential process with non-singular sample paths is defined by

$$X_0 = \varepsilon_0$$

$$X_t = (1 + Z_t)\ min(X_{t-1}, \varepsilon_t),\ t = 1, 2, \ldots \tag{3.6}$$

A typical simulated sample path for the random-Tavares model (3.6) is illustrated in Figure 3.4.

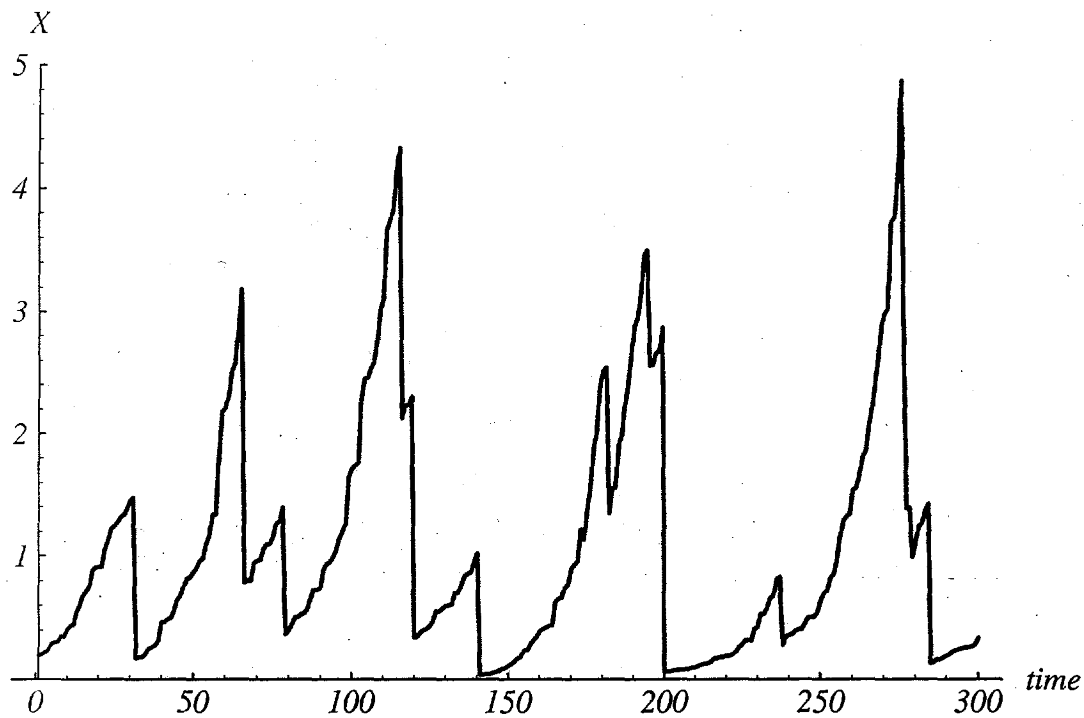

Figure 3.4. This figure illustrates through simulation the sample path of the generalized Tavares temporal waste level model. The Z_t for the simulation are iid *exp(12.4)* which corresponds to $\rho_1 = 0.93$.

The autocorrelation structure for the random-Tavares exponential model is

$$\rho_k = (\mu e^{\mu} E_1(\mu))^k, \tag{3.7}$$

where

$$E_1(x) = \int_x^{\infty} \frac{e^{-v}}{v}\, dv \tag{3.8}$$

The function E_1 is generally known as an exponential integral function. Equation (3.7) is obtained as follows. First define

$$\pi(k,t) = \Pi_{i=k,t} (1 + Z_i) \tag{3.9}$$

By recursively applying the structural definition of the model we have

$$X_t / Z_1, Z_2, \ldots, Z_t = min(\pi(1,t)\varepsilon_0, \pi(1,t)\varepsilon_1, \pi(2,t)\varepsilon_2, \ldots, \pi(t,t)\varepsilon_t) \tag{3.8}$$

We now simply reconstruct the autocorrelation development in Tavares(1980). It is well known that the minimum of iid exponential variates is again exponential. Also $X_t / Z_1, Z_2, ..., Z_t$ is distributed *exp(1)*. Consequently we may write

$$X_t / Z_1, Z_2, ..., Z_t = min(U_t, V_t) \tag{3.10}$$

where the random variables U_t and V_t are independent exponential with respective means $\pi(1,t)$ and $\frac{\pi(1,t)}{\pi(1,t) - 1}$.

By the same reasoning,

$$X_{t+k} / Z_1, Z_2, ..., Z_{t+k} = \pi(t+1,t+k)\ min(X_t, W_t) \tag{3.11}$$

where W_t is exponential with mean $\frac{1}{\pi(t+1,t+k) - 1}$.

By equations (3.10) and (3.11),

$$\boldsymbol{E}\ X_t X_{t+k} / Z_1, Z_2, ..., Z_{t+k} =$$

$$\pi(t+1,t+k)\ \boldsymbol{E}\ X_t\ min(X_t, W_t) / Z_1, Z_2, ..., Z_{t+k} = \frac{\pi(t+1,t+k) + 1}{\pi(t+1,t+k)}. \tag{3.12}$$

Thus, conditional on $Z_1, Z_2, ..., Z_{t+k}$, the lag k autocorrelation is

$$\rho_k / Z_1, Z_2, ..., Z_{t+k} = \frac{1}{\pi(t+1,t+k)}. \tag{3.13}$$

We next develop the probability structure of the random variable $T(W=w)$. For the logistic waste level model it was necessary to obtain the distribution of $T(W=w)$ through simulation. The simplicity of the model made this quite easy. The simplicity of the random-Tavares model also leads to an easy simulation of $T(W=w)$. The transformation $x = \gamma(w)$ preserves the GRE behavior of the time series w_t in the sense that $\gamma(w)$ is a monotonically increasing function. Thus, according to the random-Tavares model (3.6), a GRE will occur at time t when $min(X_{t-1}, \varepsilon_t) = \varepsilon_t$. To simulate $T(W=w)$, we first choose or fix $W=w$. This gives a value $x = \gamma(w)$. We then simulate the random-Tavares process conditional on x, and count the number of days to a model-defined GRE. Figure 3.5 presents the average and 20th percentile of the simulated distribution of $T(W = w)$ via the random-Tavares model for a suite of waste levels w. The value of w is represented on the abscissa as a percentile from the $exp(1)$ distribution. For this simulation, $n = 5000$. The sample lag one autocorrelation for the data presented in Figure 3.3 is 0.9286. The value of μ used in the simulation is obtained by setting ρ_1 in Equation (3.7) to 0.9286. This gives the estimate $\hat{\mu} = 12.13$.

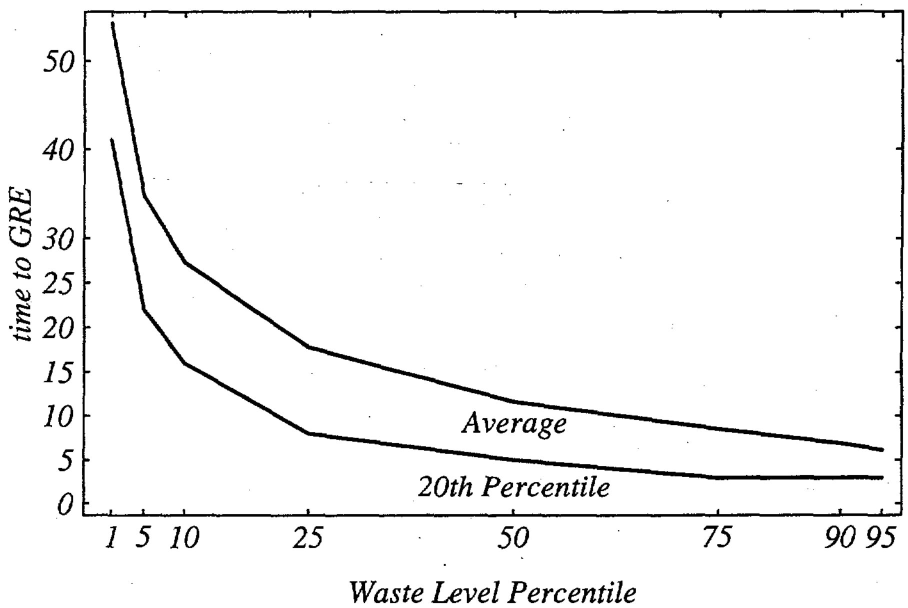

Figure 3.5. This figure presents the mean and 20th percentile of the simulated distribution of $T(W = w)$, where w is represented as a percentile from the $exp(1)$ distribution. For this simulation, $n = 5000$ values of T were generated at each waste level percentile. An estimated value of $\hat{\mu} = 12.13$ is used in the simulation.

The algebraic representation of the probability structure of $T(W)$ is nearly tractable for the random-Tavares exponential waste level model. Recall that $x = \gamma(w)$. By applying the model definition of a GRE we have

$$P(T(w) = 1) = P(x > \varepsilon_1) = \mathbf{E}_{Z_1}(1 - exp(-xZ_1)). \tag{3.17}$$

Further,

$$P(T(w) = 2) = P(x < \varepsilon_1, X_1 > \varepsilon_2) = P(x < \varepsilon_1, (1+Z_1)x > \varepsilon_2) =$$

$$E_{Z_1,Z_2}(1 - exp(-xZ_2(1+Z_1)))\ exp(-xZ_1). \tag{3.18}$$

By analogous reasoning,

$$P(T(w) = 3) = P(x < \varepsilon_1 , X_1 < \varepsilon_2 , X_2 > \varepsilon_3) =$$

$$P(x < \varepsilon_1 , (1+Z_1)x < \varepsilon_2 , (1+Z_1)(1+Z_2)x > \varepsilon_3) =$$

$$E_{Z_1,Z_2,Z_3}(1 - exp(-xZ_3(1+Z_1)(1+Z_2)))\ exp(-xZ_1)\ exp(-xZ_2(1+Z_1)). \tag{3.19}$$

In general,

$$P(T(w) = j) = E_{Z_1,Z_2,\ldots,Z_j}(1 - exp(-xZ_j\pi(1,j-1))\ \Pi_{i=1,j-1} exp(-xZ_i\pi(1,i-1)) =$$

$$E_{Z_1,Z_2,\ldots,Z_j}(1 - exp(-xZ_j\pi(1,j-1))\ exp(-x\Sigma_{i=1,j-1} Z_i\pi(1,i-1)) \tag{3.20}$$

The Tavares(1980) exponential model is constructed by constraining the $Z_j = \lambda\ (constant),\ \forall j$. In this special case

$$P(T(w) = j) = e^x(e^{-x(1+\lambda)^{j-1}} - e^{-x(1+\lambda)^j}) \tag{3.21}$$

If we equate the autocorrelation function (3.21) to the lag one sample autocorrelation, 0.9286, we can estimate $\lambda = 0.08$. The probability mass function (3.21) can then be used to construct Figure 3.6, the average and 20th percentile of the distribution of $T(W = w)$ for a suite of waste levels w. Again, the value of w is represented on the abscissa as a percentile from the *exp(1)* distribution.

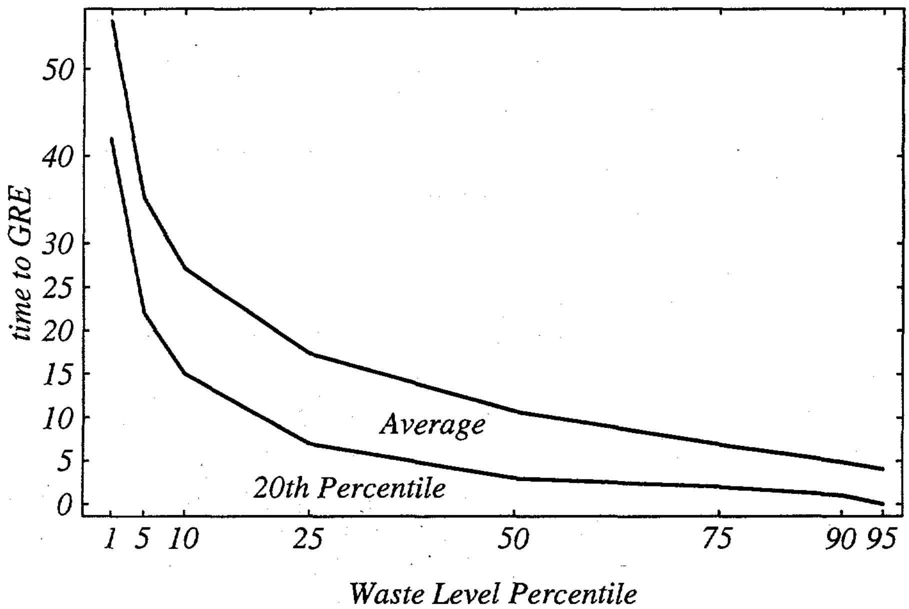

Figure 3.6. This figure presents the mean and 20th percentile of the distribution of $T(W = w)$ via the Tavares(1980) exponential waste level model. The value of w is represented as a percentile from the *exp(1)* distribution. An estimated value of $\lambda = 0.8$ was used in the probability mass function.

To apply conditional least squares estimation methods we first obtain

$$E(X_t/X_{t-1} = x) = 1 - \mu/(\mu+x) + \mu(\log(\mu+x) - \log(\mu)) \tag{3.22}$$

As in Section 2, the conditional least squares objective function is then given by

$$Q(\mu) = \Sigma_{i=2,n} (x_i - E(X_i/X_{i-1} = x_{i-1}))^2 \tag{3.23}$$

Minimizing $Q(\mu)$ yields an estimate $\tilde{\mu} = 11.5$ and $Q(\tilde{\mu}) = 89.0$. Figure 3.7 gives the average and 20th percentile of the simulated distribution of $T(W = w)$ using $\tilde{\mu} = 11.5$.

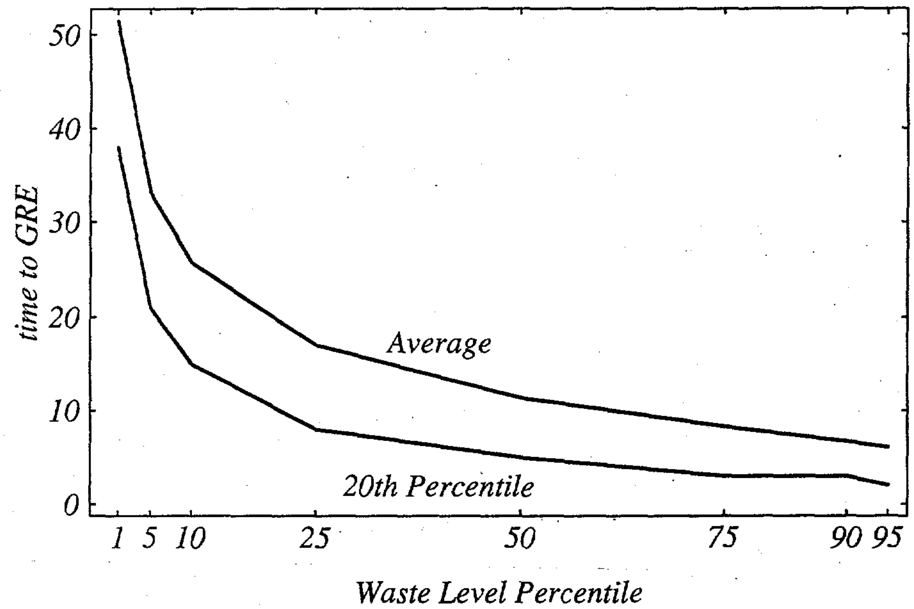

Figure 3.7. This figure presents the mean and 20th percentile of the simulated distribution of $T(W = w)$, where w is represented as a percentile from the *exp(1)* distribution. For this simulation $n = 5000$ values of T were generated at each waste level percentile. An estimated value of $\tilde{\mu} = 11.5$ is used in the simulation.

4.0 Discussion.

The decision to use the power-logistic or random-Tavares model should begin with a visual inspection of the sample path of the transformed data $x_t = \gamma(w_t)$. We can model the marginal distribution of a temporal waste level record as $logistic(\mu, \sigma)$. Under this assumption the transformation $x_t = \gamma(w_t)$ should give a sample path record that resembles the simulated sample paths of the power-logistic process. Alternatively we can model the marginal distribution of a temporal waste level record as $Weibull(\theta, \sigma, c)$. Under the Weibull assumption the transformation $x_t = \gamma(w_t)$ should give a sample path record that resembles the simulated sample paths of the random-Tavares process. The conditional least squares objective function also provides a tool to identify a model of best fit. For the waste level record analyzed in this paper we obtained $Q(\tilde{\delta}) = 198.7$ for the power-logistic model and $Q(\tilde{\mu}) = 89.0$ for the random-Tavares model. Evidently the random-Tavares model is the better of the two.

It is possible to use the probability mass function (3.21) to estimate the parameter λ. Suppose a GRE has just occurred and the waste level has dropped initiatory to another GRE.

Associated with this initial waste level is a subsequent time to the next GRE. Thus we have the bivariate data $\{w_i, t(w_i)\}$ or, through the transformation $x = \gamma(w)$, we equivalently have $\{x_i, t(x_i)\}$. Arnold and Strauss(1988) show that maximizing the psuedolikelihood function

$$PL(\lambda) = \Pi_{i=1,n} exp(x_i)\left[exp\left(-x_i(1+\lambda)^{t(x_i)-1}\right) - exp\left(-x_i(1+\lambda)^{t(x_i)}\right)\right] \qquad (3.22)$$

gives a consistent estimate for λ. The algebraic representation of the full likelihood is difficult to obtain, as it will include stochastic dependencies between consecutive GRE's. Psuedolikelihood estimation is a useful alternative when faced with the calculation of a sample autocorrelation function from a temporal waste level record with missing segments of data.

Changes in the behavior of a GRE waste tank can be subtle. To detect such changes we can use a fitted waste level model to simulate a waste level record. This waste level record can be summarized with statistics. The probability structure of these model-based statistics can also be obtained by simulating a large number of waste level records. If the observed future statistics are unusual when compared with their simulated distribution, then we can conclude a change has occurred. For example, the random-Tavares waste level model can be used to construct a simulated waste level record for say 180 days. We can then compute the lag one autocorrelation for this record. By simulating a large number of records we obtain a simulated representation of the distribution of the lag one autocorrelation. If we find that the actual 180-day lag one autocorrelation is unusual for the simulated distribution, then we would conclude that the behavior of the waste tank has changed over this 180-day period.

There are several other potential waste level models in the literature. See for example, Arnold(1988), Arnold(1989), Gaver and Lewis(1980), Lawrance and Lewis(1981) and Robertson(1989). Each of these models has the simplicity of structure exhibited in the two models described in this paper. A summary of the properties of these models can be found in Anderson(1990).

5.0 References

Anderson, D. N. (1990). *Some Time Series Models with Non-Additive Structure*, Unpublished Ph. D. Thesis, University of California, Riverside.

Arnold, B. C. (1988). "Logistic Processes involving Markovian Minimization.", *Technical Report 165, University of California, Riverside.*

Arnold, B. C. (1989). "A Logistic Process Constructed using Geometric Minimization.", *Statistics and Probability Letters*, 7, 253-257.

Arnold, B. C. and Robertson, C. A. (1989) "Autoregressive Logistic Processes.", *Journal of Applied Probability*, 26, 1989, pp. 524-531.

Arnold, B. C. and Strauss, D. J. (1988). "Pseudolikelihood Estimation.", Technical Report 164, University of California, Riverside.

Gaver, D. P. and Lewis, P. A. W. (1980) "First - Order Auto regressive Gamma Sequences and Point Processes.", *Advances in Applied Probability*, 12, 727-745.

Klimko, L. A. and Nelson, P. I. (1978) "On Conditional Least Squares Estimation for Stochastic Processes.", *Annals of Statistics*, 6, 629-642.

Lawrance, A. J. and Lewis, P. A. W. (1981) "A New Auto regressive Time Series Model in Exponential Variables (NEAR(1)).", *Advances in Applied Probability*, 13, 826-845.

Robertson, C. A. (1989). "An Exponential Process via a Switching Structure.", *Unpublished manuscript, University of California, Riverside.*

Tavares, L. V. (1980) "An Exponential Markovian Stationary Process.", *Journal of Applied Probability*, 17, 1117-1120.

Yeh, H., Arnold, B. C. and Robertson, C. A. (1988) "Pareto Processes.", *Journal of Applied Probability*, 25, 291-301.