INVENTORY CONTROL PERFORMANCE OF VARIOUS FORECASTING METHODS WHEN DEMAND IS LUMPY

Adriano O. Solis[(a)], Somnath Mukhopadhyay[(b)], Rafael S. Gutierrez[(c)]

[(a)] Management Science Area, School of Administrative Studies, York University, Toronto, Ontario M3J 1P3, Canada
[(b)] Department of Information & Decision Sciences, The University of Texas at El Paso, El Paso, TX 79968-0544, USA
[(c)] Department of Industrial & Systems Engineering, The University of Texas at El Paso, El Paso, TX 79968-0521, USA

[(a)] asolis@yorku.ca, [(b)] smukhopadhyay@utep.edu, [(c)] rsgutier@utep.edu

ABSTRACT
This study evaluates a number of methods in forecasting lumpy demand – single exponential smoothing, Croston's method, the Syntetos-Boylan approximation, an optimally-weighted moving average, and neural networks (NN). The first three techniques are well-referenced in the intermittent demand forecasting literature, while the last two are not traditionally used. We applied the methods on a time series dataset of lumpy demand. We found a simple NN model to be superior overall based on several scale-free forecast accuracy measures. Various studies have observed that demand forecasting performance with respect to standard accuracy measures may not translate into inventory systems efficiency. We simulate on the same dataset a periodic review inventory control system with forecast-based order-up-to levels. We analyze resulting levels of on-hand inventory, shortages, and fill rates, and discuss our findings and insights.

Keywords: lumpy demand forecasting, neural networks, inventory control, simulation

1. INTRODUCTION
When there are intervals with no demand occurrences for an item, demand is said to be *intermittent*. Intermittent demand is also *lumpy* when there are large variations in the sizes of actual demand occurrences. Intermittent or lumpy demand has been observed in both manufacturing and service environments (Willemain, Smart, Schockor, and DeSautels 1994; Bartezzaghi, Verganti, and Zotteri 1999; Syntetos and Boylan 2001, 2005; Ghobbar and Friend 2002, 2003; Regattieri, Gamberi, Gamberini, and Manzini 2005; Teunter, Syntetos, and Babai 2010). In proposing a theoretically coherent scheme for categorizing demand into four types (smooth, erratic, intermittent, and lumpy), Syntetos, Boylan, and Croston (2005) suggest $CV^2 > 0.49$ and $ADI > 1.32$ for characterizing lumpy demand (where CV^2 represents the squared coefficient of variation of demand sizes and *ADI* is the average inter-demand interval).

We apply a number of forecasting methods to actual demand data from an electronic components distributor operating in Monterrey, Mexico, involving 24 stock keeping units (SKUs) each with 967 daily demand observations exhibiting a wide range of demand values and intervals between demand occurrences. Values of CV^2 range between 9.84 and 45.93 while values of *ADI* range between 3.38 and 5.44 (see Table 1) – all well over the cutoffs for lumpy demand as specified above.

Table 1: Basic Dataset Statistics

Series	1	2	3	4	5	6
% Nonzero Demand	30.4	32.8	32.7	34.1	35.7	36.2
Mean Demand	251.02	262.08	271.60	274.43	278.01	324.84
Std Dev	1078.80	985.19	1305.36	1221.31	1191.04	1387.20
CV^2	18.47	14.13	23.10	19.81	18.35	18.24
ADI	4.51	4.25	4.78	3.97	3.77	3.73

Series	7	8	9	10	11	12
% Nonzero Demand	32.4	33.3	34.4	33.8	35.0	35.2
Mean Demand	237.09	274.31	253.77	346.04	303.11	321.61
Std Dev	743.88	1134.55	959.19	1710.19	1229.80	1149.70
CV^2	9.84	17.11	14.29	24.43	16.46	12.78
ADI	5.21	4.73	4.03	4.83	5.14	4.83

Series	13	14	15	16	17	18
% Nonzero Demand	33.6	34.1	35.2	35.0	33.8	36.3
Mean Demand	299.15	296.07	288.78	305.81	228.74	352.32
Std Dev	1425.87	1321.28	1090.65	1257.98	889.07	1480.69
CV^2	22.72	19.92	14.26	16.92	15.11	17.66
ADI	5.44	4.68	4.39	4.41	4.30	4.09

Series	19	20	21	22	23	24
% Nonzero Demand	38.1	34.7	35.8	33.0	35.7	32.7
Mean Demand	322.98	355.48	328.70	394.84	314.33	410.00
Std Dev	1054.75	1609.05	1390.67	2675.95	1438.57	1929.56
CV^2	10.66	20.49	17.90	45.93	20.95	22.15
ADI	3.90	4.86	4.09	4.37	3.38	3.39

Seven forecasting methods were initially evaluated, namely:

- single exponential smoothing (SES)
- Croston's method
- Croston's method with two separate smoothing constants
- the Syntetos-Boylan approximation
- the Syntetos-Boylan approximation with two separate smoothing constants
- a five-period weighted moving average with optimized weights
- neural networks.

1.1 Well-Referenced Methods for Forecasting Lumpy Demand

Croston (1972) noted that SES, frequently used for forecasting in inventory control systems, has a bias that places the most weight on the most recent demand occurrence. He proposed a method of forecasting intermittent demand using exponentially weighted moving averages of nonzero demand sizes and the intervals between nonzero demand occurrences to address the bias problem. Leading application software packages for statistical forecasting incorporate Croston's method (Syntetos and Boylan 2005; Boylan and Syntetos 2007).

While Croston assumed a common smoothing constant α, Schultz (1987) suggested that separate smoothing constants, α_i and α_s, be used for updating the inter-demand intervals and the nonzero demand sizes, respectively. Eaves and Kingsman (2004) provide a clear formulation of Croston's method with 'two alpha values'. In the current study, for each demand series, we identify the combination of two alphas corresponding to the best forecast in the calibration sample. We then apply the best combination of α_i and α_s for each series to forecast the test sample.

Syntetos and Boylan (2001, 2005) reported an error in Croston's mathematical derivation of expected demand, leading to a positive bias. Syntetos and Boylan (2005) proposed what is now referred to in the literature as the Syntetos-Boylan approximation (SBA) – which involves multiplying Croston's estimator of mean demand by a factor of $(1-\alpha_i/2)$, where α_i is the exponential smoothing constant used in updating the inter-demand intervals.

We note, however, that Syntetos and Boylan (2005) used the same smoothing constant for updating demand sizes as for updating inter-demand intervals in applying SBA to monthly demand histories over a two-year period of 3000 stock-keeping units (SKUs) in the automotive industry. As we do with Croston's method in the current study, we likewise consider SBA with separate smoothing constants, α_i and α_s, for updating the inter-demand intervals and the nonzero demand sizes. Other than Schultz (1987), only Syntetos, Babai, Dallery, and Teunter (2009) and Teunter, Syntetos, and Babai (2010) have to-date reported using two separate smoothing constants on inter-demand intervals and demand sizes in empirical investigation – in the two latter studies, applied to the SBA demand estimator.

The use of low α values in the range of 0.05-0.20 has been recommended in the literature on lumpy demand (Croston 1972; Johnston and Boylan 1996). Syntetos and Boylan (2005) used the four α values of 0.05, 0.10, 0.15, and 0.20 for the SES, Croston's, and SBA methods. We use these same four values in the current study.

1.2 'Non-Traditional' Methods for Forecasting Lumpy Demand

Sani and Kingsman (1997) observed that less sophisticated (e.g., moving average) methods can prove superior to Croston's method in practice. Eaves (2002) also found that forecasting methods simpler than Croston's or SBA method can provide better forecasting results for intermittent and slow-moving demand. Regattieri, Gamberi, Gamberini, and Manzini (2005) studied monthly demand data pertaining to spare parts for Alitalia's fleet of Airbus A320 aircraft in 1998-2004. They found weighted moving average (WMA) forecasts, based on selecting the best sets of weights for three, five, and seven-month periods, to perform generally better than Croston's, SES, and other smoothing methods (SBA was not considered).

In the current study, we applied a five-day weighted moving average method with optimized weights (WMA5) – to correspond to weekly demand over a five-day work week. The method averages the last five lagged values of lumpy demand through optimized weights. The lagged value 1 means the demand during the last time period and so on. To determine the optimized weights, the method runs a standardized linear ordinary least square (OLS) regression on current period demand as target variable and the five most recent lagged period demands as predictor variables. The beta values of the lagged demands are normalized so that the values add up to 1.000. The normalized values (see Table A.1 in the Appendix) are used as the moving average weights. The method determines the weights from calibration data (as discussed in Section 2.1) only.

Researchers have used neural network (NN) models in various forecasting applications. NN models can provide reasonable approximations to many functional relationships (e.g., White 1992; Elman and Zipser 1987), with flexibility and nonlinearity cited as their two most powerful aspects. Hill, O'Connor, and Remus (1996) compared forecasts produced by NN models against forecasts generated using six time series methods from a systematic sample of 111 of the 1001 time series in a well known 'M-competition' (Makridakis, Andersen, Carbone, Fildes, Hibon, Lewandowski, Newton, Parzen, and Winkler 1982). They found NN forecast models to be significantly more accurate than those of the six traditional time series models for monthly and quarterly demand data across a number of selection criteria. Very few previous studies have used NN to forecast irregular or lumpy demand (e.g., Carmo and Rodrigues 2004; Gutierrez, Solis, and Mukhopadhyay 2008).

We used a multi-layered perceptron (MLP) trained by a back-propagation (BP) algorithm (Rumelhart, Hinton, and Williams 1988). We followed guidelines proposed by a fairly recent study on MLP architecture selection (Xiang, Ding, and Lee 2005) which suggests that one should first try a three-layered MLP. One should also start with the minimum number of hidden units required to approximate the target function.

Functions learned by a minimal net over calibration sample points work well on new samples. We used three layers of network:

- one input layer for input variables
- one hidden unit layer
- one output layer of one unit.

We chose three hidden units, which is a reasonably low number required to approximate any complex function. The network connects all hidden nodes with the input nodes representing the last time period's demand value and cumulative number of time periods with zero demand. The output node representing the current period's demand value connects to all hidden nodes. We used 0.1 for the learning rate and 0.9 for the momentum factor, as recommended by seminal research (Rumelhart, Hinton, and Williams 1988).

NN usually can approximate any function with the proper choice of parameters and a specific network structure (Lippmann 1987). Eventually, after a repeated change of network structure and parameter values, one can find a "successful" combination of calibration and validation samples which provides a false impression of model generalization. In this study, we choose a simple network structure with the same parameter values across all 24 lumpy demand series. We validate once and report the results without going back to improve upon them. If, accordingly, the NN model with this restriction outperforms other methods on the test sample, we are able to conclude the model to be superior. We do not change the parameter values of NN across all the 24 time series. On the other hand, we relax the restriction on other methods by trying out different parameter values as recommended in the literature.

2. DATA SET PARTITIONING AND FORECAST ACCURACY MEASURES

2.1 Data Set Partitioning

We initially used the first 624 observations of the 967 daily demand observations in each of the 24 time series to "train" and validate the models (the *training* sample). We then tested, at each of the four values of α, the other forecasting models under consideration on the final 343 observations (the *test* sample). This generated an approximately 65:35 (65% training data and 35% test data) partitioning. Researchers typically use an 80:20 split to validate models (Bishop 1995). To compare the forecasting methods further we have also ran the models on 50:50 and 80:20 data partitions. Due to space limitations, however, we report results only for the 65:35 data partitioning in this paper.

2.2 Forecast Accuracy Measures

Mean absolute percentage error (MAPE) is the most widely used accuracy measure for ratio-scaled data. The traditional definition of MAPE involves terms of the form $|E_t|/A_t$ (where A_t and E_t, respectively, represent actual demand and forecast error in period t). Since lumpy demand involves periods with zero demands, the traditional MAPE definition fails. We used an alternative specification of MAPE as a ratio estimate (Gilliland 2002), which guarantees a nonzero denominator:

$$\text{MAPE} = \left(\sum_{t=1}^{n} |E_t| \bigg/ \sum_{t=1}^{n} A_t \right) \times 100 \,. \tag{1}$$

Willemain, Smart, Schockor, and DeSautels (1994) conducted a study comparing performance of SES and Croston's method in intermittent demand forecasting, using (i) MAPE based on the above ratio estimate, (ii) median absolute percentage error (MdAPE), (iii) root mean squared error (RMSE), and (iv) mean absolute deviation (MAD) as forecast accuracy measures. However, they reported only MAPEs, noting that relative results were the same for all four measures. Eaves and Kingsman (2004) applied MAPE, RMSE, and MAD in comparing the performance of several methods (SES, Croston's, SBA, 12-month simple moving average, and the previous year's simple average) in forecasting demand for spare parts for in-service aircraft of the Royal Air Force (RAF) of the UK. Using demand data over a six-year period for 18750 SKUs randomly selected out of some 685000 line items, they found SBA to provide the best results overall using MAPE, but the 12-month simple moving average yielded the best MADs overall.

Armstrong and Collopy (1992) did an extensive study for making comparisons of errors across time series. For selecting the most accurate method, they recommend the median RAE (MdRAE) when few time series are available. The relative absolute error (RAE) is calculated for a given series, at a given time t, by dividing the absolute error under method m, $|F_{m,t} - A_t|$, by the corresponding absolute error for the random walk, $|F_{rw,t} - A_t|$. We compute the random walk forecast by simply adding one unit to the actual demand in the immediately preceding period. Hence,

$$\text{RAE}_t = |F_{m,t} - A_t| \big/ |(A_{t-1} + 1) - A_t| \,. \tag{2}$$

MdRAE is simply the median of all RAE_t values across the entire test sample.

Syntetos and Boylan (2005) employed two accuracy comparison measures: relative geometric root-mean square error (RGRMSE) and percentage best (PB). The first measure is as follows:

$$\text{RGRMSE} = \left(\prod_{t=1}^{n} (A_{a,t} - F_{a,t})^2 \right)^{1/2n} \bigg/ \left(\prod_{t=1}^{n} (A_{b,t} - F_{b,t})^2 \right)^{1/2n} \tag{3}$$

where the symbols $A_{m,t}$ and $F_{m,t}$ denote actual demand and forecast demand, respectively, under forecasting method m at the end of time period t. PB, another scale-free accuracy measure, is the percentage of time periods that one method outperforms all the other methods. We use absolute error as the criterion to assess alternative methods' performance under the PB approach. Gutierrez, Solis, and Mukhopadhyay (2008) used MAPE as well as RGRMSE and PB to assess performance of the SES, Croston's, SBA, and NN forecasting methods.

In the current study, we assess and compare the performance of the seven forecasting methods – specified in Section 1 – as applied to the test samples in the 24 time series in the dataset, using four scale-free error criteria: (i) MAPE, (ii) MdRAE, (iii) RGRMSE, and (iv) PB. We used SAS software release 9.1 for our empirical investigations of both forecasting performance (reported in Section 3) and inventory control performance (reported in Section 4).

3. EMPIRICAL INVESTIGATION OF FORECASTING PERFORMANCE

Like Syntetos and Boylan (2005), Gutierrez, Solis, and Mukhopadhyay (2008) applied four α values: 0.05, 0.10, 0.15, and 0.20. The latter study found the SES, Croston's and SBA methods to work best with $\alpha = 0.05$ for all 24 time series considered, which appears consistent with the lumpiness observed in the dataset.

For the Croston's and SBA methods with separate smoothing constants, we identified in the current study – for each demand series – the combination of α_i and α_s corresponding to the best forecast in the training sample, based upon a minimum MAPE criterion. We then use the best combination for each series (see Table A.2 in the Appendix) to generate forecasts on the test sample. (Eaves and Kingsman (2004), in applying the SES, Croston's and SBA methods, likewise optimized smoothing constants using MAPEs only, but cautioned that the smoothing methods may yield better results if smoothing constants were optimized using a different forecast accuracy criterion.)

Figure 1 shows the relative performance of all the seven methods with respect to MAPE under the 65:35 data partitioning. NN MAPEs are superior for 20 of the 24 time series. WMA5 is clearly the worst performer in all series. For four series (4, 22, 23, and 24), NN, Croston, SBA, and SES perform quite closely. If MAPE is the criterion to select the best method, a simple NN model is clearly the best performing method overall.

In the current study, we did not observe any substantial improvement in forecast accuracy arising from using separate smoothing constants, α_i and α_s. To execute forecasting and demand management, calibration of two-alpha combinations will add more complexity to the process. In light of practical implications, we decided to drop the two-alpha Croston and SBA methods. Moreover, because the SBA method is consistently superior to Croston's method, we proceed to investigate only four methods – SES, SBA, WMA5 and NN.

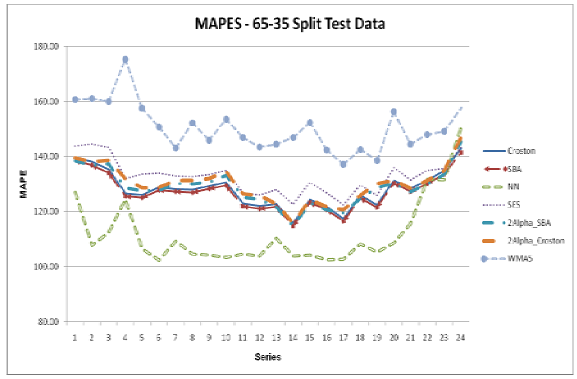

Figure 1: Comparison of MAPEs

Figure 2 shows the performance of the four remaining methods with respect to PB. NN is again the superior method overall, while WMA5 ranks second.

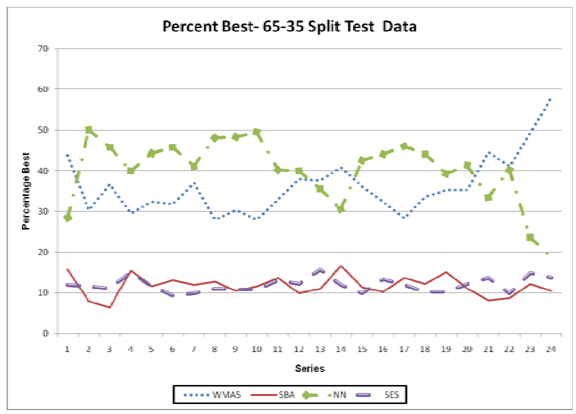

Figure 2: Comparison of Percentage Bests

Table 2 shows, for the 65:35 data partitioning, the best performing method across the 24 series for each accuracy measure. NN is the best method overall with respect to MAPE, MdRAE, and PB, while NN and WMA5 perform equally well with respect to RGRMSE. However, WMA5 performs poorly when MAPE is the criterion for selecting the best method. The other two methods, SES and SBA, which were developed and heavily researched for forecasting of intermittent/lumpy demand, did not perform as well as NN and WMA5.

4. EMPIRICAL INVESTIGATION OF INVENTORY CONTROL PERFORMANCE

Demand forecasting and inventory control have traditionally been examined independently of each other (Tiacci and Saetta 2009; Syntetos, Babai, Dallery, and Teunter 2009). In reality, demand forecasting performance with respect to standard accuracy measures may not translate into inventory systems efficiency (Syntetos, Nikolopoulos, and Boylan 2010). In an intermittent demand setting, a periodic review inventory control system has been recommended (Sani and Kingsman 1997; Syntetos, Babai, Dallery, and Teunter 2009). A number of recent studies that address both forecasting and inventory control performance for intermittent demand (e.g., Eaves and Kingsman 2004; Syntetos and Boylan 2006; Syntetos, Babai, Dallery, and Teunter 2009; Syntetos, Nikolopoulos, and Boylan 2010; Teunter, Syntetos, and Babai 2010) have employed the order-up-to (T,S) periodic review system (see, for example, Silver, Pyke, and Peterson 1998) – where T and S represent the review period and order-up-to level, respectively.

Table 2: Best Method by Forecast Accuracy Measure

	65:35 Data Partitioning			
Series	MAPE	MdRAE	RGRMSE	PB
1	NN	NN	WMA	NN
2	NN	NN	NN	NN
3	NN	NN	WMA	NN
4	NN	NN	WMA	NN
5	NN	NN	NN	NN
6	NN	NN	NN	NN
7	NN	NN	WMA	NN
8	NN	NN	NN	NN
9	NN	NN	NN	NN
10	NN	NN	NN	NN
11	NN	NN	NN	NN
12	NN	NN	NN	NN
13	NN	NN	WMA	WMA
14	NN	NN	WMA	WMA
15	NN	NN	NN	NN
16	NN	NN	NN	WMA
17	NN	NN	NN	NN
18	NN	NN	NN	NN
19	NN	NN	NN	NN
20	NN	NN	WMA	NN
21	NN	NN	WMA	WMA
22	SBA	NN	WMA	WMA
23	NN	WMA	WMA	WMA
24	SBA	WMA	WMA	WMA
Overall	**NN**	**NN**	**N/W**	**NN**

In the study by Eaves and Kingsman (2004) earlier discussed in Section 2.1, simulations of a (T,S) system were performed on actual demand data, aggregated quarterly, for the 18750 randomly selected SKUs. Forecast-based order-up-to levels S were determined as the product of the forecast demand per unit of time and the "protection interval", T+L (where L is the reorder lead time). Implied average stockholdings were calculated using a backward-looking simulation assuming a common fill rate (or percentage of total demand filled by on-hand inventory) of 100%. SBA yielded the lowest average stockholdings among the five forecasting methods evaluated.

Syntetos and Boylan (2006) used a dataset consisting of monthly demand observations over a two-year period for 3000 SKUs in the automotive industry. They modeled demand over T+L in a (T,S) system by way of a negative binomial distribution – a compound Poisson distribution whose variance is greater than its mean. Two target fill rates were considered: 90% and 95%. Using two cost policies in simulation comparisons, they demonstrated the superior inventory control performance of the SBA forecasting method relative to Croston's, SES, and 13-month simple moving average methods.

Two recent studies (Syntetos, Babai, Dallery, and Teunter 2009; Teunter, Syntetos, and Babai 2010) used a large dataset from the RAF of the UK, involving 84 monthly observations of demand for 5000 SKUs over seven years (1996-2002). The first 24 observations of each time series were used to initialize estimates of demand level and variance, and the second 24 observations were used to optimize separate smoothing constants, α_i and α_s, on inter-demand intervals and demand sizes, respectively. Simulation of inventory control performance in applying the SBA method was then performed over the final 36 observations. In the 2009 article, the authors noted that in many intermittent-demand situations the *ADI* is larger than the lead time (or the lead time plus one review period). They accordingly excluded those SKUs in the RAF dataset with *ADI* less than T+L (with T = 1 month in this case), resulting in 2455, or 49% of the original 5000 SKUs, actually being considered. In the 2010 article, lead time demand was modeled as a compound binomial process, with demands in successive periods being identically and independently distributed. Both studies introduced a new approach to determine, in a (T,S) inventory control system, order-up-to levels utilizing both inter-demand interval and demand size forecasts explicitly whenever demand occurs. Using various service-oriented and cost-oriented criteria, the two studies observed the superiority of the new approach compared to the classical approach which uses only the SBA estimate of average demand size.

Sani and Kingsman (1997) have earlier applied simulation of real data (consisting of 30 long series of daily demand data over five years for low demand items), involving a single run for each data series, as a form of empirical evaluation. In the current study, our simulations have also taken the form of a single run performed on the test sample consisting of the final 343 daily demand observations for each of the 24 series in the dataset. Simulation experiments involving multiple runs have not been attempted owing to the difficulty of mathematically modeling the degree of demand lumpiness observed in our dataset.

We assume in the current study a (T,S) periodic review inventory control system with full backordering. For initial simulation runs, T is five days (one week) and we assume a deterministic reorder lead time, L, of 10 days (two weeks). Let I_t and B_t, respectively, denote on-hand inventory and inventory shortage/ backlog at the time of review t, and F_j represent the forecast demand for period j (j = t+1, …, t+T+L). Without providing a safety stock component, the replenishment quantity based upon a forecast-based order-up-to level is

$$Q_t = \sum_{t+1}^{t+T+L} F_j - I_t + B_t . \tag{4}$$

In our simulation studies, we continue to investigate only the four methods remaining under consideration – SES, SBA, WMA5 and NN – as identified in Section 3. Figure 3 shows the mean inventory on-hand for each of the 24 SKUs throughout the test sample. We excluded WMA5 from this figure, because means for most series are well over those computed when using SBA, SES, and NN. We find that the mean inventory on-hand arising from the use of NN is lower, in most instances, than when SBA or SES is used.

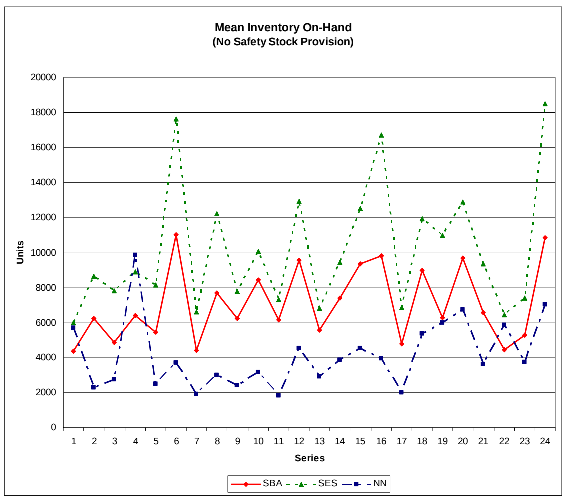

Figure 3: Mean On-Hand Inventory with No Safety Stock Provision

In Figure 4, however, we observe average backorders to be higher with NN than with SBA or SES. Mean shortages are much lower with WMA5, consistent with the much higher average on-hand inventory levels observed in Figure 3 for this method. Figure 5 shows that the percentage of time when inventory shortages occur is generally highest when NN is used. In like manner, we see in Figure 6 that the average fill rate is lowest overall when NN is used.

We reiterate that Figures 3-6 pertain to the case where there is no safety stock provided. The literature on inventory control suggests a safety stock component in order-up-to levels to compensate for uncertainty in demand during the "protection interval" $T+L$. For each demand series, we calculated the standard deviation s_{tr} of daily demand during the training sample. Initially, we set the safety stock level to be k standard deviations of daily demand during the training sample – i.e., $k \cdot s_{tr}$ – with k = 4, 6, 8, 10, and 12. We then proceeded to conduct single run simulations over the 343 observations in the test sample for each of the 24 series.

The values of k we have thus far tested give rise to safety stocks which are, more or less, comparable with the $z \cdot \sqrt{T+L} \cdot \sigma_d$ suggested when daily demand during the protection interval is assumed to be identically and independently normally distributed with standard deviation σ_d (e.g., Silver, Pyke, and Peterson 1998). With the safety stock component, the replenishment quantity to order is

$$Q_t = \sum_{t+1}^{t+T+L} F_j + k \cdot s_{tr} - I_t + B_t . \tag{5}$$

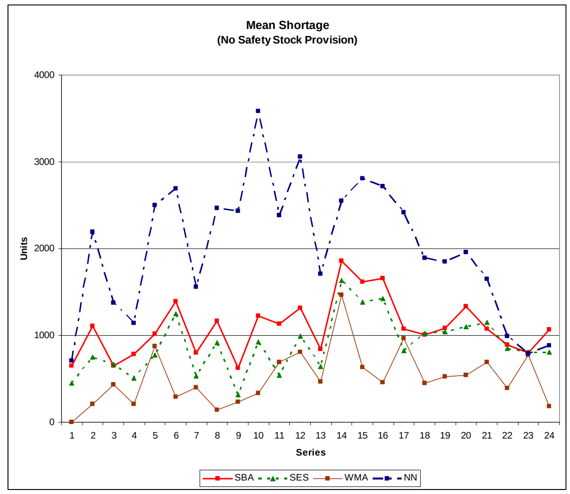

Figure 4: Mean Shortage with No Safety Stock Provision

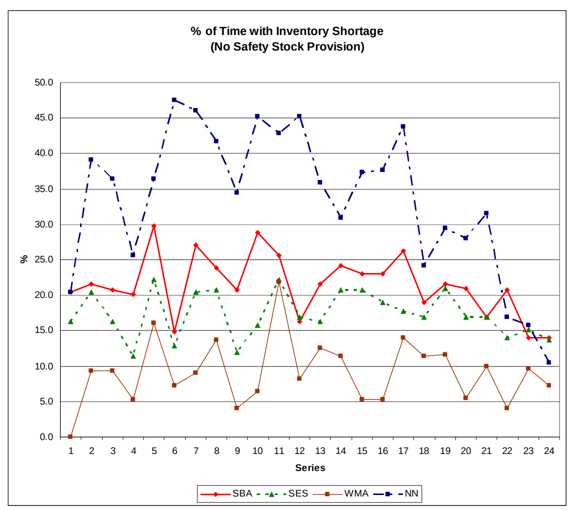

Figure 5: Percentage of Time Stocking Out with No Safety Stock Provision

When safety stock is set at $4 \cdot s_{tr}$, mean shortages as shown in Figure 7 have decreased significantly, although levels of on-hand inventory, as expected, have markedly increased. We see in Figure 8 that mean fill rates have substantially improved overall compared with those seen in Figure 6, even as mean fill rates when using NN continue to be generally lower than when WMA5, SES, and SBA are applied.

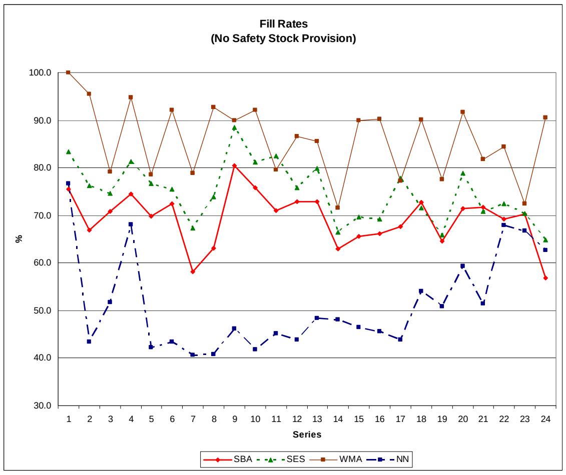

Figure 6: Average Fill Rates with No Safety Stock Provision

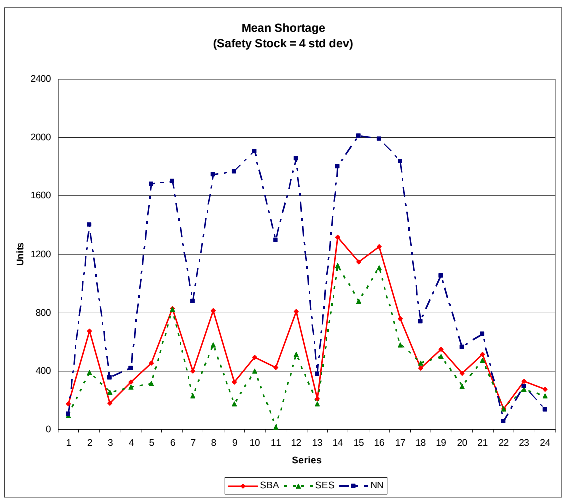

Figure 7: Mean Shortage with Safety Stock = $4 \cdot s_{tr}$

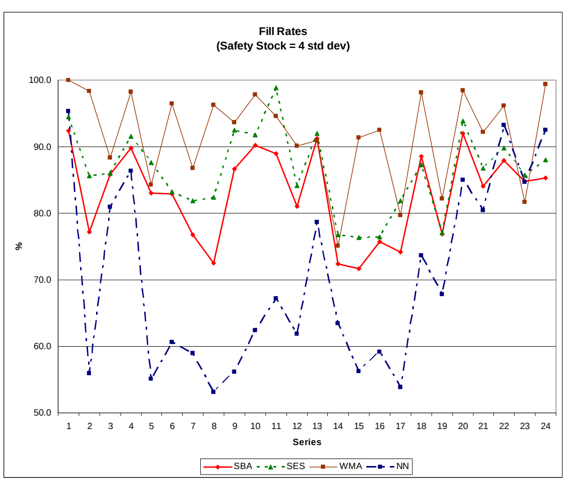

Figure 8: Average Fill Rates with Safety Stock = $4 \cdot s_{tr}$

We continue to see essentially the same mean fill rate comparisons as k is increased to 6, 8, 10, and 12. Mean fill rates when k = 8 are shown in Figure 9. We observe that all four methods under consideration lead to fill rates of 100% for series 22 and series 24.

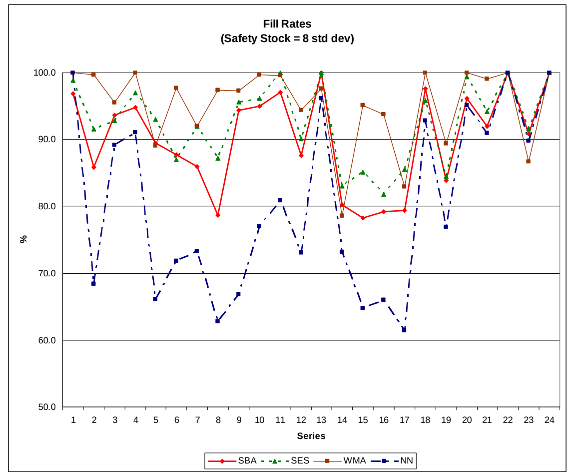

Figure 9: Average Fill Rates with Safety Stock = $8 \cdot s_{tr}$

Figure 10 shows the average on-hand inventory levels when k = 8. (Since fill rates arising from all methods are already at 100% for series 22 and series 24, these two series have been left out of Figure 10.) On the other hand, average backorder levels are shown in Figure 11. The lower mean fill rates with NN as forecasting method are clearly associated with generally lower average on-hand inventory levels but also generally greater mean shortages.

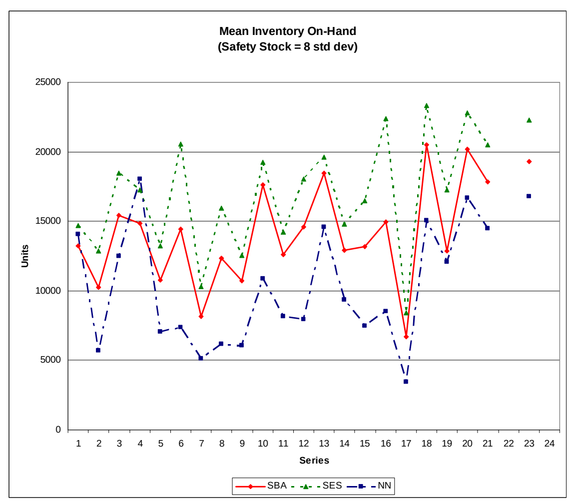

Figure 10: Mean On-Hand Inventory with Safety Stock = $8 \cdot s_{tr}$

Overall average fill rates across all 24 SKUs for each of the four methods under consideration are reported in Table 3 for the values of k tested.

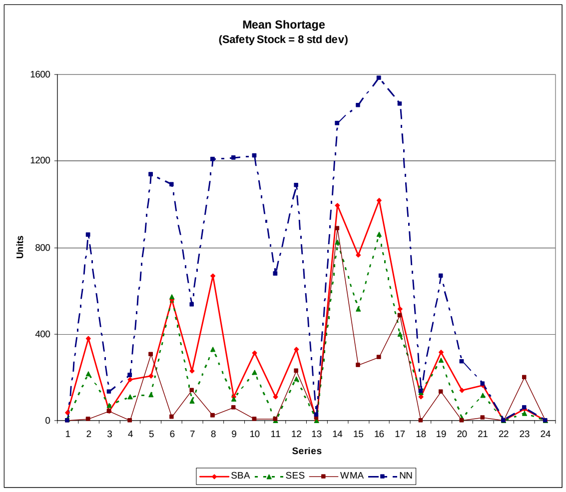

Figure 11: Mean Shortage with Safety Stock = $8 \cdot s_{tr}$

Table 3: Overall Average Fill Rates with Safety Stock = *k* Standard Deviations of Daily Demand

k	SBA	SES	WMA	NN
0	69.5	74.9	86.2	50.2
4	82.8	86.1	91.9	68.8
6	86.8	89.5	93.8	74.6
8	89.8	92.3	95.2	79.3
10	91.9	94.1	96.3	83.1
12	93.9	95.4	97.2	86.3

We have also conducted simulations with L = 3 days, for k = 0, 3, 5, 6, 7, 8, and 9. Similar comparisons of average on-hand inventory, backorders, and fill rates have arisen.

In view of the much higher levels of average on-hand inventory associated with demand forecasting using WMA5, we focus our attention on NN, SBA, and SES. At similarly specified safety stock levels, we observe much lower mean fill rates (i.e., inferior customer service levels) when NN – the "best" of the four methods based upon ratio-scaled traditional forecast accuracy measures – is applied in comparison with fill rates attained when using SES and SBA. In the same vein, NN yields relatively lower average on-hand inventory levels (i.e., lower inventory carrying costs) but higher mean shortages (i.e., higher backorder costs). While our dataset does not include specific cost information, a distributor of electronic components will be expected to pay significant attention to customer service levels and backorder costs.

Of additional interest is how SES and SBA compare in terms of stock control performance when demand is lumpy. Eaves and Kingsman (2004) and Syntetos and Boylan (2006) have found SBA to outperform several forecasting methods, SES included, when demand is intermittent though not lumpy. While SES is less sophisticated than SBA, the former yields generally higher average fill rates and lower average backorders than the latter. On the other hand, however, SES leads to somewhat higher average on-hand inventory levels than SBA.

5. CONCLUSIONS AND FURTHER WORK

In the current study, we find support for earlier assertions that demand forecasting performance with respect to standard accuracy measures may not translate into inventory systems efficiency. In particular, an NN model was found to outperform the SES and SBA methods in performance with respect to a number of scale-free traditional accuracy measures, but appears to be inferior when it comes to inventory control performance.

We intend to do further simulation work that will search, for each SKU in the dataset, for the value of k (and, hence, the safety stock component of the forecast-based order-up-to-level) that would meet a specified fill rate. Simulation studies of periodic review inventory systems generally involve searching for order-up-to-levels satisfying a target customer service level – e.g., a probability of not stocking out or a fill rate – often with a cost minimization objective (Solis and Schmidt 2009). For instance, Syntetos and Boylan (2006) evaluated performance of forecasting methods at target fill rates of 90% and 95%, while Teunter, Syntetos, and Babai (2010) considered target fill rates of 87%, 91%, 95%, and 99%. Boylan, Syntetos, and Karakostas (2008) initially set a fill rate of 95%, but later treated fill rate as a simulation parameter varying from 93% to 97%. Starting with comparable target fill rates, we will conduct simulation searches, with the resulting levels of on-hand inventory and backorders accordingly compared across the forecasting methods in terms of potential cost implications.

Based on the simulation searches outlined above, a more rational comparison between methods, especially between SES and SBA, should be possible.

APPENDIX

Table A.1: Optimized Weights for WMA5

	Optimized Weights on Lagged Demand				
Series	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5
1	0.434	0.348	0.036	0.055	0.127
2	0.113	0.086	0.174	0.344	0.282
3	0.151	0.162	0.205	0.133	0.349
4	1.017	-0.113	0.030	0.019	0.047
5	0.064	0.226	0.061	0.126	0.524
6	0.130	0.295	0.082	0.059	0.433
7	0.274	0.038	0.484	0.125	0.079
8	0.172	0.149	0.122	0.190	0.366
9	0.215	0.210	0.228	0.109	0.237
10	0.212	0.220	0.152	0.198	0.218
11	0.095	0.144	0.527	0.067	0.168
12	0.048	0.131	0.100	0.012	0.709
13	0.041	0.320	0.262	0.232	0.145
14	0.078	0.042	0.710	0.023	0.147
15	0.018	0.246	0.100	0.557	0.079
16	0.297	0.296	0.120	0.203	0.085
17	0.167	0.229	0.364	0.119	0.122
18	0.175	0.061	0.136	0.294	0.333
19	0.154	0.313	0.176	0.054	0.303
20	0.150	0.429	0.139	0.162	0.119
21	0.185	0.153	0.186	0.294	0.181
22	0.102	0.166	0.083	0.240	0.408
23	0.151	0.449	0.032	0.152	0.216
24	0.158	0.080	0.628	0.065	0.069

Table A.2: Minimum MAPEs of Two-alpha SBA and Two-alpha Croston's Methods on Training Sample

	Two-alpha SBA Method			Two-alpha Croston's Method		
Series	Minimum MAPE (%)	α_i	α_s	Minimum MAPE (%)	α_i	α_s
1	164.0	5%	5%	165.9	5%	5%
2	152.9	5%	5%	154.5	5%	5%
3	164.3	10%	5%	166.2	10%	5%
4	166.0	10%	5%	167.9	15%	5%
5	160.5	10%	5%	162.3	10%	5%
6	154.8	10%	5%	156.4	5%	5%
7	154.5	10%	5%	156.0	10%	5%
8	159.5	10%	5%	161.1	10%	5%
9	147.4	10%	5%	148.8	10%	5%
10	165.8	10%	5%	167.7	10%	5%
11	159.1	10%	5%	160.8	10%	5%
12	154.0	10%	5%	155.6	10%	5%
13	161.8	5%	5%	163.6	5%	5%
14	161.8	5%	5%	163.5	5%	5%
15	158.7	5%	5%	160.4	5%	5%
16	158.1	5%	5%	159.8	5%	5%
17	154.7	10%	5%	156.3	10%	5%
18	160.3	5%	5%	162.1	5%	5%
19	231.6	15%	20%	235.3	15%	20%
20	159.1	5%	5%	160.9	5%	5%
21	157.1	5%	5%	158.9	5%	5%
22	161.4	5%	5%	163.2	5%	5%
23	155.0	5%	5%	156.7	5%	5%
24	163.6	10%	5%	165.4	10%	5%

REFERENCES

Armstrong, J. S. and Collopy, F., 1992. Error measures for generalizing about forecasting methods: Empirical comparisons. *International Journal of Forecasting*, 8 (1), 69-80.

Bartezzaghi, E., Verganti, R., and Zotteri, G., 1999. A simulation framework for forecasting uncertain lumpy demand. *International Journal of Production Economics*, 59 (1-3), 499-510.

Bishop, C.M., 1995. *Neural Networks for Pattern Recognition*. Oxford, UK: Clarendon Press.

Boylan, J.E. and Syntetos, A.A., 2007. The accuracy of a Modified Croston procedure. *International Journal of Production Economics*, 107 (2), 511-517.

Boylan, J.E., Syntetos, A.A., and Karakostas, G.C., 2008. Classification for forecasting and stock control: a case study. *Journal of the Operational Research Society*, 59 (4), 473-481.

Carmo, J.L. and Rodrigues, A.J., 2004. Adaptive forecasting of irregular demand processes. *Engineering Applications of Artificial Intelligence*, 17 (2), 137-143.

Croston, J.D., 1972. Forecasting and stock control for intermittent demands. *Operational Research Quarterly*, 23 (3), 289-304.

Eaves, A.H.C., 2002. *Forecasting for the ordering and stock-holding of consumable spare parts*. Thesis (PhD). University of Lancaster.

Eaves, A.H.C. and Kingsman, B.G., 2004. Forecasting for the ordering and stock-holding of spare parts. *Journal of the Operational Research Society*, 55 (4), 431-437.

Elman, J.L. and Zipser, D., 1987. *Learning the hidden structure of speech*. Institute of Cognitive Science Report 8701. University of California, San Diego, USA.

Ghobbar, A.A. and Friend, C.H., 2002. Sources of intermittent demand for aircraft spare parts within airline operations. *Journal of Air Transport Management*, 8 (4), 221-231.

Ghobbar, A.A. and Friend, C.H., 2003. Evaluation of forecasting methods for intermittent parts demand in the field of aviation: a predictive model. *Computers & Operations Research*, 30 (14), 2097-2114.

Gilliland, M., 2002. Is forecasting a waste of time? *Supply Chain Management Review*, 6 (4), 16-23.

Gutierrez, R.S., Solis, A.O., and Mukhopadhyay, S., 2008. Lumpy demand forecasting using neural networks. *International Journal of Production Economics*, 111 (2), 409-420.

Hill, T., O'Connor, M., and Remus, W., 1996. Neural network models for time series forecasts. *Management Science*, 42 (7), 1082-1092.

Johnston, F.R. and Boylan, J.E., 1996. Forecasting for items with intermittent demand. *Journal of the Operational Research Society*, 47 (1), 113-121.

Lippmann, R.P., 1987. An introduction to computing with neural nets. *IEEE ASSP Magazine*, 4 (2), 4-22.

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., and Winkler, R., 1982. The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of Forecasting*, 1 (2), 111-153.

Regattieri, A., Gamberi, M., Gamberini, R., and Manzini, R., 2005. Managing lumpy demand for aircraft spare parts. *Journal of Air Transport Management*, 11 (6), 426-431.

Rumelhart, D.E., Hinton, G.E., and Williams, R.J., 1988. Learning internal representations by error propagation. In: *Parallel Distributed Processing Explorations in the Microstructure of Cognition*. Cambridge, USA: MIT Press, 328-330.

Sani, B. and Kingsman, B.G., 1997. Selecting the best periodic inventory control and demand forecasting methods for low demand items. *Journal of the Operational Research Society*, 48 (7), 700-713.

Schultz, C.R., 1987. Forecasting and inventory control for sporadic demand under periodic review. *Journal of the Operational Research Society*, 38 (5), 453-458.

Silver, E.A., Pyke, D.F., and Peterson, R., 1998. *Inventory Management and Production Planning and Scheduling*. New York, USA: John Wiley & Sons.

Solis, A.O. and Schmidt, C.P., 2009. Stochastic leadtimes in a one-warehouse, N-retailer inventory system with the warehouse actually carrying stock. *International Journal of Simulation and Process Modelling*, 5 (4), 337-347.

Syntetos, A.A. and Boylan, J.E., 2001. On the bias of intermittent demand estimates. *International Journal of Production Economics*, 71 (1-3), 457-466.

Syntetos, A.A. and Boylan, J.E., 2005. The accuracy of intermittent demand estimates, *International Journal of Forecasting*, 21 (2), 303-314.

Syntetos, A.A., Boylan, J.E., and Croston, J.D., 2005. On the categorization of demand patterns. *Journal of the Operational Research Society*, 56 (5), 495-503.

Syntetos, A.A. and Boylan, J.E., 2006. On the stock control performance of intermittent demand estimators. *International Journal of Production Economics*, 103 (1), 36-47.

Syntetos, A.A., Babai, M.Z., Dallery, Y., and Teunter, R., 2009. Periodic control of intermittent demand items: theory and empirical analysis. *Journal of the Operational Research Society*, 60 (5), 611-618.

Syntetos, A.A., Nikolopoulos, K., and Boylan, J.E., 2010. Judging the judges through accuracy-implication metrics: The case of inventory forecasting. *International Journal of Forecasting*, 26 (1), 134-143.

Teunter, R.H., Syntetos, A.A., and Babai, M.Z., 2010. Determining order-up-to levels under periodic review for compound binomial (intermittent) demand. *European Journal of Operational Research*, 203 (3), 619-624.

Tiacci, L. and Saetta, S., 2009. An approach to evaluate the impact of interaction between demand forecasting method and stock control policy on the inventory system performances. *International Journal of Production Economics*, 11 (1), 63-71.

White, H., 1992. Learning and statistics. In: H. White, ed. *Artificial Neural Networks: Approximation and Learning Theory*. Oxford, UK: Blackwell, 79.

Willemain, T.R., Smart, C.N., Shockor, J.H., and DeSautels, P.A., 1994. Forecasting intermittent demand in manufacturing: a comparative evaluation of Croston's method. *International Journal of Forecasting*, 10 (4), 529-538.

Xiang, C., Ding, S.Q., and Lee, T.H., 2005. Geometrical interpretation and architecture selection of MLP. *IEEE Transactions on Neural Networks*, 16 (1), 84-96.

AUTHORS' BIOGRAPHIES

Adriano O. Solis is an Associate Professor of Logistics Management and Management Science at York University, Canada. After receiving BS, MS and MBA degrees from the University of the Philippines, he joined the Philippine operations of Philips Electronics where he became a Vice-President and Division Manager. He later received a PhD degree in Management Science from the University of Alabama. He was previously Associate Professor of Operations and Supply Chain Management at the University of Texas at El Paso. He has published in *European Journal of Operational Research, International Journal of Simulation and Process Modelling, Information Systems Management*, *International Journal of Production Economics*, *Computers & Operations Research*, and *Journal of the Operational Research Society*, among others.

Somnath Mukhopadhyay, an Associate Professor in the Information and Decision Sciences Department at the University of Texas at El Paso, has published in numerous well-respected journals like *Decision Sciences*, *INFORMS Journal on Computing*, *Communications of the AIS*, *IEEE Transactions on Neural Networks*, *Neural Networks*, *Neural Computation*, *International Journal of Production Economics*, and *Journal of World Business*. He received MS and PhD degrees in Management Science from Arizona State University. He was a visiting research assistant in the parallel distributed processing group of Stanford University. He has over 15 years of industry experience in building and implementing mathematical models.

Rafael S. Gutierrez is an Associate Professor in the Industrial, Manufacturing and Systems Engineering Department – of which he was formerly the Department Chair – at the University of Texas at El Paso. He received MS degrees from both the Instituto Tecnologico y de Estudios Superiores de Monterrey (ITESM), in Mexico, and the Georgia Institute of Technology, and a PhD degree in Industrial Engineering from the University of Arkansas. He previously taught at the Universidad Iberoamericana, the ITESM, and the Instituto Tecnologico de Ciudad Juarez. He has been a consultant on enterprise resource planning and materials management for, among others, Emerson Electric, Lucent Technologies, United Technologies, and Chrysler's Acustar Division. He has published in *International Journal of Production Economics*, *Military Medicine*, and *International Journal of Industrial Engineering*.