Guidelines and Techniques for Comparing System Observations with Multiple Sets of Simulation Results

Dimosthenis Anagnostopoulos Vassilis Dalakas Mara Nikolaidou

Dept of Informatics

University of Athens

Panepistimiopolis, 15771, Athens, Greece

dimosthe@di.uoa.gr, vdalakas@di.uoa.gr, mara@di.uoa.gr

Abstract--**Using simulation to support reaching conclusions for real-time processes imposes that models are thoroughly validated. This paper discusses widely adopted techniques for testing the validity of simulation results against real observations and proposes a more effective comparison technique for the case where a single system data set and multiple model data sets are available. To apply this technique, we consider the system data set as a discrete signal and use M-fold decimation, fulfilling the essential requirements imposed according to signal processing theory. Experimental results are also presented to exhibit the applicability and effectiveness of the proposed technique.**

Keywords: modelling and simulation, validation, real-time simulation, comparison techniques

I   INTRODUCTION

When simulation reaches conclusions for systems behavior in real time, it is known as real-time simulation. The term *real time*, as it relates to simulation, denotes that advancement of simulation time must occur in the real world time (i.e. not faster or slower). In faster-than-real-time simulation (FRTS), results are delivered earlier than real-time. In this case, we are capable of using system observations and model results to both test model validity and, in case of a valid model, reach predictions for the system future states. The quality of predictions can be ensured based on validation tasks involving only past and the current time points [1].

System and model state evolution in real time is depicted in figure 1. Real time points are noted as t_i. The states of the system and the model at point t_i are noted as R_i and S_i, respectively. When at time point t_x the model predicts the system state at t_n (simulation time is equal to t_n), we use the notation $Sim(t_x) = t_n$. Auditing is performed at t_{n-1}, t_n, t_{n+1} and, thus, compares states S_x and R_n at time point t_n. If model validity is consecutively ensured within a number of

consecutive auditing intervals $[t_{n-2}, t_{n-1}]$, $[t_{n-1}, t_n]$, …, it is likely that simulation predictions are also valid. Thus, plan scheduling is invoked to take advantage of predictions and experimentation resumes.
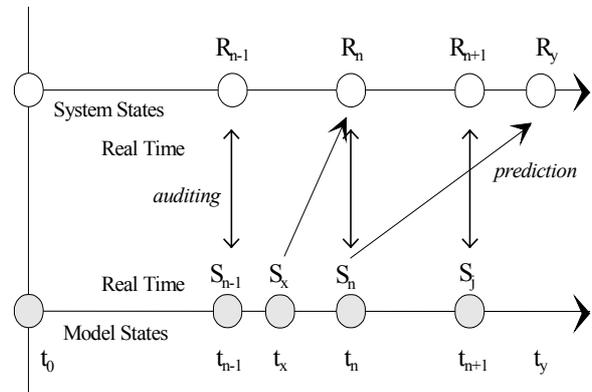
Figure 1: Experimentation in FRTS

A substantial issue is that time-dynamic system may be subjected to changes (or reformations) at any time point. Reformations involve the main features of a system, as its structure and operation parameters. Structure reformations also involve the coupling relationship between model components. We thus distinguish structure and operation parameter reformation types. To deal with them, the model must be subjected to dynamic remodeling, i.e. in real-time, without recompilation. Numerous model validation types and techniques have been discussed in the literature [2]. Both *conceptual model validation* and *computerized model verification* should be pre-assured, as only preconstructed models can be used when modifying a composite model in real time. *Operational validity* is concerned with determining that the model output behavior has the accuracy required for the intended purpose over the domain of its intended applicability [3]. This is where most of the validation testing and evaluation must take place in

FRTS. Comparing model results and real-time process observations, the following features must be provided:

1. Validation must be a real-time, computationally efficient procedure.
2. Conditions causing the model to be remodeled must be explicit. Each remodeling condition involves one or many comparisons between specific - primitive or statistical - variables, denoted as monitoring variables. The number of monitoring variables corresponding to each single condition depends on the current system configuration. For instance, comparing a single-queue, multi-teller system with its G/G/s model representation may involve the average service time of each server. As the number of servers (s) may be modified (e.g. when a teller is removed), the number of monitoring variables corresponding to this specific condition may also be modified.
3. A single comparison method is appropriate for each specific monitoring variable comparison. For single-valued variables (e.g. operation parameters), a comparison between model and system values is required. Multi-valued statistical results are compared using other approaches, such as the inspection approach, confidence intervals and hypothesis tests [4]. The acceptable deviation ranges (or comparison parameters) must also be determined for each monitoring variable comparison.

Furthermore, in FRTS, it is most probable that, due to the limited time interval within which system and model data are collected, only a single data set will be available for the system. The same does not stand for simulation, as it may run considerably faster than the system, enabling multiple (n) replications to deliver n corresponding data sets within the auditing interval.

Based on the above, we conclude that the following requirements must be met:

1. Any remodeling condition involves one or many monitoring variables, which are not used by any other condition.
2. Each monitoring variable corresponds to a single comparison method, i.e. we consider that there is a single, case-specific, optimal comparison technique for each monitoring variable, not for all variables of the same condition.
3. There is a given set of comparison methods, each being appropriate for specific comparison characteristics.
4. A single system data set and n model data sets are available for comparison.

In this paper, we discuss model validation issues in FRTS, in terms of comparison techniques between system observations and models results. As low effectiveness is indicated using well-known techniques for comparing system and model data in the context of FRTS, we introduce a technique for improving effectiveness, based on the available model and system data. This technique can be applied on a single set of IID system observations, using M-fold decimation for producing m sets of system observations, after verifying that each of the m data sets can be used to substitute the original data set. Potential applications of this technique are not restricted to FRTS, as there are many cases where a limited amount of system data is available due to the nature of the system under study.

The paper is organized as follows: in Section II, model validation is discussed, accomplished through monitoring variable comparison. In Section III, we describe the M-fold decimation-based technique for the effective comparison of a single system data set with multiple model data sets, while experimental results are presented in Section IV.

II   MODEL VALIDATION

Considering that model validation is accomplished using k monitoring variables (MV_1-MV_k), k comparisons need to be made. Monitoring variable values for the system and the model are respectively denoted as $MV_1.r$ and $MV_1.s$. Each comparison is accomplished using a single comparison technique considered as more appropriate for the specific data.

In general, we consider three comparison techniques as most appropriate for accomplishing comparisons between model and system data:

1. System - model values comparison, for single-valued variables.
2. Inspection approach, for statistical variables with one system observation data set and n model data sets [4].
3. Confidence-interval approach, for statistical variables with m system observation data sets and n model data sets) [4].

In system-model comparison, the confidence-interval approach is more effective than a hypothesis test, as the hypothesis that system and model performance measures are equal will be false in almost all cases. Moreover, confidence-interval comparison provides more information, i.e. it gives an indication of the magnitude by which the system differs from the model [4]. A reliable approximate solution in order to form a confidence interval for the difference in the two expectations (from the system and the model) is to use the approach proposed by Welch [5]. This approach is used for comparing two systems with unequal and

unknown variances, called the *Behrens-Fisher problem*, when the *Xij*'s are normally distributed [5]. This classical approach is considered as more appropriate for building a confidence interval based on a different number of independent data sets, as it enables *m* to be different than *n*, but imposes that system data sets are independent from the corresponding model data sets [4].

In the following, we discuss system-model data comparison for the three above cases (comparison techniques). In all cases, comparison parameters define the acceptable deviation range (*dr*) for each comparison, so that, when model value deviation from the corresponding system value exceeds the predetermined range, invalidity is detected. However, the meaning of this range may be different for each comparison technique. Each deviation range depends on the nature of the experiment (i.e. how close should model states be to system states) and the specific method used to compare system observations and model data.

In cases (1) and (2), deviation range determines the lower and upper endpoints of the interval $[l(MV_i.r), u(MV_i.r)]$ and the model is considered as valid when:

$$MV_i.s \in [l(MV_i.r), u(MV_i.r)],$$
$$l(MV_i.r) = MV_i.r(1-dr), \; u(MV_i.r) = MV_i.r(1+dr)$$

For single-valued variables, system and model variables ($MV_i.s$, $MV_i.r$) are directly obtained. For statistical variables with one system observation data set and *n* model data sets, a single value is derived for each data set, thus

$$MV_i.s = sum\,(MV_{i1}.s, MV_{i2}.s, \ldots, MV_{in}.s)/n,$$

where $MV_{ij}.s$ is the statistical sample obtained from replication *j* when *n* replications are made.

In the third case, where statistical variables with *m* system observation data sets and *n* model data sets are available, we build a confidence interval based on a different number of independent data sets [5]. Applying the Welch approach,

$$\overline{MV_i.r} = \frac{\sum_{j=1}^{m} MV_{ij}.r}{m}, \quad \overline{MV_i.s} = \frac{\sum_{j=1}^{n} MV_{ij}.s}{n}$$

$$S^2(MV_i.r) = \frac{\sum_{j=1}^{m}[MV_{ij}.r - \overline{MV_i.r}]^2}{m-1}$$

$$S^2(MV_i.s) = \frac{\sum_{j=1}^{n}[MV_{ij}.s - \overline{MV_i.s}]^2}{n-1}$$

The estimated degrees of freedom are computed as

$$\widehat{f} = \frac{[S^2(MV_i.r)/m + S^2(MV_i.s)/n]^2}{[S^2(MV_i.r)/m]^2/(m-1) + [S^2(MV_i.s)/n]^2/(n-1)}$$

The following interval is an approximate *100(1-a)* percent confidence interval for $MV_i.r - MV_i.s$

$$\overline{MV_i.r} - \overline{MV_i.s} \pm t_{\widehat{f},1-a/2}\sqrt{\frac{S^2(MV_i.r)}{m} + \frac{S^2(MV_i.s)}{n}}$$

Evidently, in case (3), the deviation range defines the value *a*, which means that we wish the confidence interval to cover $MV_i.r - MV_i.s$ with probability *1-a*. Suppose the upper and lower endpoints of the interval are marked as *u(a)* and *l(a),* respectively. If $0 \notin [l(a), u(a)]$, the difference between $MV_i.r$ and $MV_i.s$ is statistically significant at level *a* and we consider the model to be invalid.

III   M-FOLD DECIMATION-BASED TECHNIQUE

As previously discussed, confidence-interval-based comparison for statistical variables is more effective than other comparison techniques. The problem in FRTS is that only a single observation data set is available from the system, while *n* data sets may be produced by simulation. We thus examine how it is possible to produce *m* independent data sets from a single original data set.

Considering that $MV_i.r$ is calculated from a sequence of system values x_1, x_2, ..., x_n, where $x(n)$ is a stochastic process and *n* is an integer value. Having this single sequence $x(n)$ from the system, we consider it as a discrete signal and use well-established methods from digital signal processing to obtain a number of sequences that retain the statistical features of the initial sequence. A basic operation in multirate digital signal processing is decimation. The *M*-fold decimator takes an input sequence $x(n)$ and produces the output sequence

$$x_M(n) = x(Mn)$$

where *M* is an integer [6]. Only these samples from $x(n)$ that occur at time equal to multiples of M are retained by the decimator. Figure 2 demonstrates this idea for M=2. As is mathematically substantiated, decimation results in aliasing, unless $x(n)$ is bandlimited in a certain way. In general, it may not be possible to recover x(n) from $x_M(n)$ because of loss in information. To retain the initial statistical features of a sequence $x(n)$ using a sequence $x_M(n)$, we need to be able to reconstruct x(n) from $x_M(n)$. To achieve this, as discussed in [6], *M* must satisfy the condition

$$|\omega| < \pi/M \quad (1),$$

To calculate $|\omega|$, we have to use the Discrete Time Fourier Transformation (DTFT) on $x(n)$ [7]

$$X(\omega) = \sum_{n=-\infty}^{n=+\infty} x(n) \cdot e^{-j\omega n}$$

Condition (1) must be fulfilled due to decimation causing an expansion to the Fourier transform $X(e^{j\omega})$ of the input signal $x(n)$. Thus, when a signal is being decimated by a factor M, the stretched version of $X(e^{j\omega})$ can in general overlap with its shifted replicas [6]. If this happens, we cannot recover $x(n)$ from the decimated signal $x_M(n)$. The overlap effect is called *aliasing*. To avoid aliasing, it is widely adopted that $x(n)$ must be a bandlimited signal to the region $|\omega| < \pi/M$.
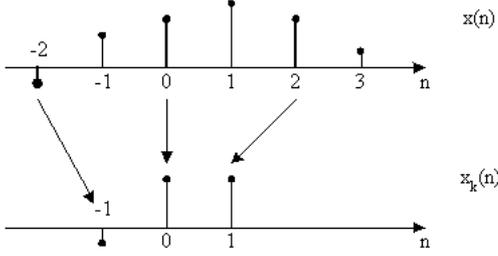
Figure 2: M-fold decimation (M=2)

In order to find the appropriate frequency ω_{max} to bandlimit our signal spectrum to the range $|\omega| < \omega_{max}$ without a significant loss of information, we estimate the power spectral density (PSD) of the signal x(n) using Welch's method [8]. With such a representation, the Fourier coefficients determine the distribution of power at the various discrete frequency components. The range of ω is $[0, \pi]$ for real $x(n)$. Using the condition $PSD(\omega) > l\% \ PSD_{max}$ we obtain the frequency range where the PSD is larger than $l\%$ of its maximum value. Depending on the choice of l, the maximum frequency in this range is $\omega_{max}(l)$, i.e. the last frequency inside the range $[0, \pi]$ having a significant contribution in $x(n)$. Thus, omitting the frequencies above ω_{max}, the spectrum $X(e^{j\omega})$ of the initial signal $x(n)$ is bandlimited to the range where $|\omega| < \omega_{max}$. However, there is an error factor depending on the choice of the coefficient l. The smaller the value of l, the smaller is the error.

Bandlimiting the signal is achieved through a digital filter, i.e. a filter that takes one sequence of numbers (the input signal) and produces a new sequence of numbers (the filtered output signal). We apply a simple low-pass filter, i.e. one that does not affects low frequencies and rejects high frequencies. The gain of the ideal low-pass filter is equal to 1 for frequencies between 0 Hz and the cutoff frequency f_c Hz, and 0 for all higher frequencies ($\omega_{max} = 2\pi f_c$) [9]. The transfer function $H(e^{j\omega})$ of this lowpass filter is depicted in figure 3. The output spectrum is produced by multiplying the input spectrum obtained from the PSD of $x(n)$ by the transfer function. The output signal is

then bandlimited to $|\omega| < \omega_{max}$ containing the amount of information included in $[0, PSD(\omega_{max})]$.
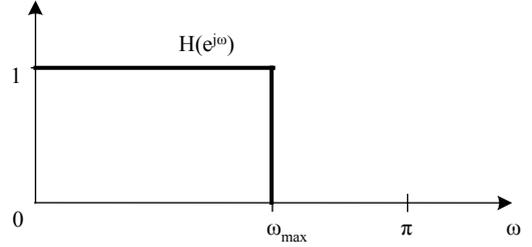
Figure 3: Digital filter

Having ensured condition (1), it is possible to obtain $x_M(n)$ that preserves the statistical features of $x(n)$, through M-fold decimation. It is evident that signals $x_k(n)$, $1 \le k \le M$, also preserve the statistical features of $x(n)$, using a lower degree of decimation. Decimation signals can be produced using a digital filter bank which splits the signal $x(n)$ into M sequences $x_k(n)$, $1 \le k \le M$, typically called subband signals [6].

The proposed technique for deriving m (M=m) system observations data sets from the original data set $x(n)$ comprises the following steps, depicted in figure 4:

1) Calculate the power spectral density (PSD) of the data set. Calculate the maximum frequency contribution ω_{max} where the PSD is greater than at least $l\%$ of PSD_{max}.
2) Calculate the maximum M using condition (1) to avoid aliasing.
3) Construct M sequences $x_k(n)$, $1 \le k \le M$, retaining the statistical features of the original sequence $x(n)$, using M-fold decimation.
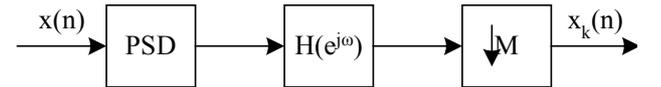
Figure 4: Producing $m \ x_k(n)$ sequences from $x(n)$

IV APPLYING THE M-FOLD DECIMATION-BASED TECHNIQUE

We present an example involving model result - system observation comparison using a simple manufacturing system as the application domain (figure 5) to illustrate the effectiveness of the proposed technique. This system assembles parts and constructs an assembled product. Time duration equal to d_i is required for this process, where d_i follow the normal distribution.

Figure 5: Simple manufacturing system

The objective of FRTS is to reach reliable conclusions for the system performance in the near future and to ensure model validity taking into consideration the current system state and performance. To achieve this, we compare the average delay (*avgdelay*) in the system and the simulation model at the end of sequential auditing intervals. Within each auditing interval, a single data set is available from the system (system observations) while *n* data sets are derived from the corresponding simulation replications. We assume that system data are IID. We thus consider a single monitoring variable MV_1 for performing comparisons between system and model data, using a 90% confidence interval (1-a=90%) as follows:

mvid	mvname	comparison type	comparison parameters
MV_1	avgd	1 system data set/ n model data sets	0.1 (a=10%)

To determine model validity using *avgdelay* as a measure of performance, it is possible to compare system and model values using a hypothesis test or, more efficiently, using an acceptable deviation range (*dr*) for the difference between estimations from the model and the system, e.g. *dr=10%*, as follows

$$MV_1.r=(d_1+ d_2+… +d_r)/r$$
$$MV_1.s= avgd_1.s +avgd_2.s+…+ avgd_n.s$$
$$avgd_i.s=(d_{i1}+ d_{i2}+… +d_{is_i})/s_i$$

assuming that *r* system observations are available and that *n* replications are made, where each replication i produces a single average delay (*avgd_i.s*) from s_i output samples. In the simple approach using an acceptable deviation range for testing model validity, the model is considered as valid when

$$MV_1.s \in[MV_1.r (1-dr), MV_1.r(1+dr)]$$

We apply the above technique for achieving a more effective comparison between system and model data, enabling *m* independent data sets to be derived from the original IID system data set. Considering the original sequence $d(r)= d_1,d_2,… ,d_r$, we derive the power spectral density (PSD) of this spectrum and then calculate the maximum frequency contribution ω_{max} where the PSD is greater than at least *l%* of PSD_{max}. The maximum *M* is accordingly calculated to avoid aliasing. Then, *m* sequences $d_k(r)$ are constructed using M-fold decimation, all sequences retaining the statistical features of the original sequence *d(r)*. Obtaining *m* system observation data sets (M=m) and *n* model data sets, we build a *100(1-a%)* confidence interval for $MV_1.r-MV_1.s$ based on a different number of independent data sets. We first calculate $\overline{MV_1.r}$, $\overline{MV_1.s}$, $S^2(MV_1.r)$, $S^2(MV_1.s)$ and and $\hat{f}$. Then, we construct the confidence interval *[l(a), u(a)]*, for *a=10%*. If $0 \notin [l(a), u(a)]$, the difference between $MV_1.r$ and $MV_1.s$ is statistically significant at level *a*.
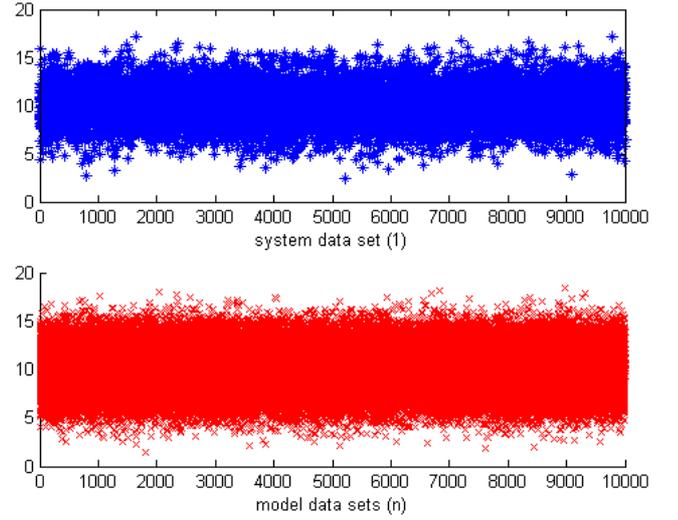
Figure 6: System-model initial data sets

We executed a number of experiments with the above sample manufacturing system. System observation data *d(r)* and 10 (n=10) model data sets $y_1(r), y_2(r), …, y_{10}(r)$, are depicted in figure 6 for μ=10 and σ²=2.0. In this figure, the model is an accurate representation of the system.
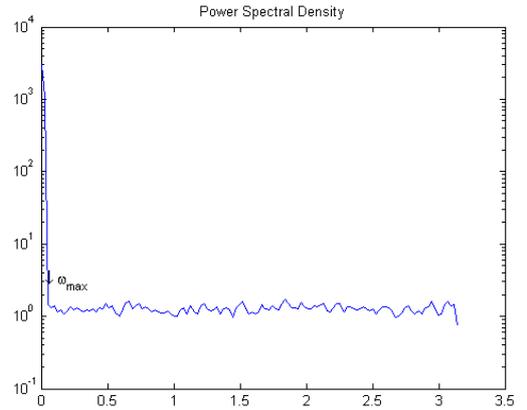
Figure 7: Power Spectral Density

Using a value of 0.001 (0.1%) for l parameter, the power spectral density of $d(r)$ and ω_{max} are depicted in Figure 7. In this case, M is equal to 128. Decimated system sequences $d_k(r)$ for various values of k are depicted in figure 8.
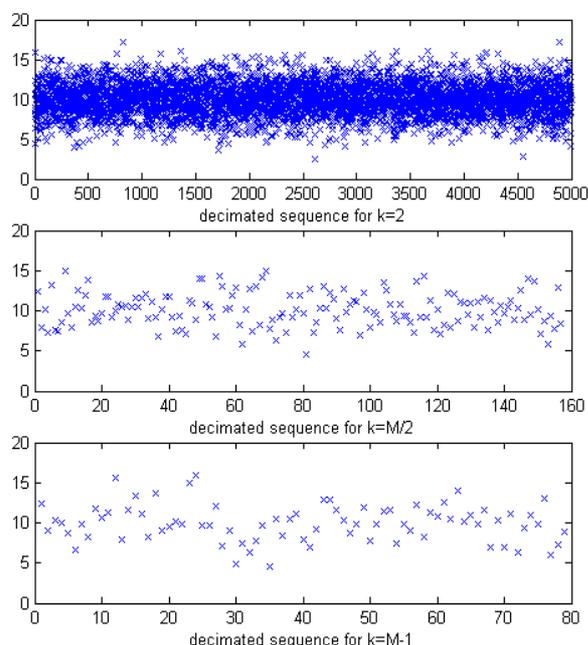
Figure 8: Decimated system data sequences

Constructing a 100(1-a)% confidence interval for $MV_1.r-MV_1.s$, results from different experiments are depicted in Table 1. To reach results, we use various system service characteristics and different model representations for the system, i.e. different parameters for the distribution of service times, as well as various values for l parameter.

Table 1: Comparison parameters and results

system	model	parameters			confidence interval	
μ/σ^2	μ/σ^2	l	1-a	M	l(a)	u(a)
10/2	10/2	10^{-3}	90%	128	0.0495	0.0961
10/0.05	10/0.005	$5*10^{-7}$	90%	42	-0.0007	0.0005
10/0.05	10/0.001	$5*10^{-7}$	90%	42	-0.0007	0.0004
1000/1	1000/1	$2*10^{-3}$	90%	128	-0.027	-0.0023
1000/1	1000/1	10^{-7}	90%	32	-0.0235	0.0007
1000/1	1000/1	$2*10^{-8}$	90%	16	-0.0221	0.0039
1000/1	1100/100	$2*10^{-3}$	90%	128	-101.60	-98.63
1000/1	1100/100	10^{-7}	90%	32	-101.61	-98.64

Experiments were conducted using a single system data set and n=10 model data sets, each consisting of 10,000 samples. Executing the steps of the above technique is an automated process, which was realized using Matlab, imposed only a low time overhead. In the case of FRTS, performance was critical. Based on the above results, the following conclusions were reached:

1) Confidence interval endpoints for identical system-model representation were extremely close, as expected, even though 0 was not always included in the confidence interval, as in the first row of Table 1. However, this effect is negligible for intervals with such limited width [4].
2) In all cases where different model representations were tested against the real system, confidence intervals clearly indicated this deviation, also providing an indication of the magnitude by which the model deviates from the system.
3) Low values for l (10^{-8} to 10^{-3}) were actually used, resulting in a practically negligible information loss in the original system data.
4) Lower values for l resulted in the calculation of lower values for M (i.e. lower decimation capabilities were provided for lower information loss).
5) Model validation was computationally efficient when applying the proposed technique.

References--
[1] Anagnostopoulos D., M. Nikolaidou, P. Georgiadis, "A Conceptual Methodology for Conducting Faster-Than-Real-Time Experiments", *SCS Transactions on Computer Simulation*, vol. 16, no 2, 1999
[2] Balci, O., "Verification, Validation and Accreditation of Simulation Models", in *Proceedings of WSC'97*, IEEE Computer Press, 1997
[3] Sargent, R., "Verification, Validation and Accreditation of Simulation Models", in *Proceedings of WSC'00*, IEEE Computer Press, 2000
[4] Law A.M., W.D. Kelton, *Simulation Modeling and Analysis,* McGraw Hill, 2000
[5] Welch B. L., "The Significance of the Difference Between Two Means when the Population Variances are Unequal", *Biometrica*, vol. 25, 1938, pp. 350-362
[6] Vaidyanathan P.P., *Multirate systems and filter banks,* Prentice Hall, Englewood Cliffs, 1993.
[7] Kaloutsides N., *Signal Processing Systems: Theory and Design*, John Wiley & Sons, 1997
[8] Welch, P.D., "The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms" *IEEE Trans. Audio Electroacoust*. Vol. AU-15 (June 1967). pp. 70-73.
[9] Proakis J., "Digital Communications", McGraw-Hill, 2001