A SURVEY OF TES MODELING APPLICATIONS

BENJAMIN MELAMED *
JON R. HILL †

NEC USA, Inc.
C&C Research Laboratories
4 Independence Way
Princeton, NJ 08540

Abstract

TES (Transform-Expand-Sample) is a versatile methodology for modeling general stationary time series, and particularly those that are autocorrelated. From the viewpoint of Monte Carlo simulation, TES represents a new and flexible input analysis approach. The salient feature of TES is its potential ability to simultaneously capture first-order and second-order properties of empirical time series (field measurements): Given an empirical sample, TES is designed to fit an arbitrary empirical marginal distribution (histogram) and to simultaneously approximate the leading empirical autocorrelations. Practical TES modeling is computationally intensive and can be effectively carried out only with computer support. A software modeling environment, called TEStool, has been designed and implemented to support the TES modeling methodology, through an interactive heuristic or algorithmic search approach employing extensive visualization.

The purpose of this paper is to introduce TES modeling, and to offer some illustrative examples from a range of applications. The paper first presents an overview of TES modeling and TEStool as well as the underlying TES theory. It then proceeds to survey a number of models from various application domains, including source modeling of compressed video and fault arrivals, financial modeling and texture generation. These examples demonstrate the efficacy and versatility of the TES modeling methodology, and underscore the high fidelity attainable with TES models.

Keywords and Phrases: Autocorrelated Processes, Input Analysis, Modeling and Simulation, Software Systems, Stochastic Processes, TES Processes.

*Current address: RUTCOR, Rutgers University, P.O. Box 5062, New Brunswick, NJ 08903
†Current address: Smith Barney, Inc., 1345 Avenue of the Americas, New York, NY 10105.

1 INTRODUCTION

Temporal and spatial dependencies are commonplace in a host of real-world random phenomena. Temporal dependence accounts, to a large extent, for burstiness in telecommunications traffic, especially in the context of high-speed communications networks. File transfer and compressed VBR (variable bit rate) video are typical examples from emerging ISDN (Integrated Services Digital Network) applications. Locality of reference in caches and data bases is underpinned by spatial dependence of successive memory accesses. The combined effects of temporal and spatial dependencies underlie fault cascades observed in network management. When formulated mathematically as real-valued stochastic processes, temporal dependencies are often proxied by autocorrelations within a stochastic process, and by cross-correlations among multiple processes.

Although dependencies in real-world systems abound, in practice, the natural impulse of modelers is to minimize or eliminate dependence from model descriptions in order to simplify analysis. A typical case in point is what might be termed "classical queueing". The bulk of queueing systems are comprised of one or more $GI/GI/m$ queues, where GI (General Independent) stands for a renewal process; that is, interarrival times and service times are specified as iid (independent identically distributed) random sequences. The standard argument in support of this approach is that these assumptions endow the models with analytical or numerical tractability. A related argument is motivated by the insight that analytical models may provide, especially when dealing with systems in the design stage, or with extant systems in the early stages of their life cycle. This argument is entirely legitimate, since in such scenarios one is often merely interested in qualitative understanding of design tradeoffs, and the modeling effort calls for calculation of performance measures which are rough rather than accurate.

Consider, however, performance analysis at a later stage of a system's life cycle. Suppose that the randomness in the system is driven by some stochastic process, and that field measurements from it are available to the modeler (at this point we don't care to distinguish between performance evaluation approaches, be they analytical or Monte Carlo simulation). The experienced analyst is immediately confronted with two generic and sequential questions:

1. Should one ignore dependence in empirical time series and use a simple model, say, from the class of renewal processes?

2. If dependence is to be modeled, how can it be done in an effective and systematic manner?

Experience shows that all too often, the prevailing attitude is to obviate the answer to the second question by simply electing a renewal process model. Certainly, by the Principle of Parsimony, modelers should only care to invest in more elaborate modeling to the extent that the extra effort will buy sufficiently increased accuracy in the resulting *performance measures of interest*. The point is that modelers should be aware of the fact that over-simplified renewal models can carry grave modeling risks.

To fix the ideas, assume that we have field measurements from a traffic (or workload) process, and we wish to gauge system performance measures resulting from offering this traffic to a queueing system. A little introspection on the nature of burstiness in arrival processes should convince the reader of its deleterious effect on waiting times: Multiple customers arriving in a burst will obviously suffer from increased waiting times, while the lulls separating bursts waste server utilization. Indeed, various studies [5, 24, 29] have shown that when autocorrelated traffic is offered to a queueing system, the resulting performance measures are much worse than those corresponding to renewal traffic, differing by orders of magnitude.

A growing realization of the impact of bursty traffic on queueing system performance has provided the initial motivation for devising input analysis methods that are able to capture dependence in time series; no doubt, this realization is bound to extend to other modeling domains. The TES (*Transform-Expand-Sample*) modeling methodology [25, 11, 12, 13] is just such an input analysis approach, specifically designed to address the second question above. In essence, TES is a modulo-1 reduction of a simple linear autoregressive-based scheme, followed by additional transformations. Furthermore, while the basic TES formulation is Markovian, these transformations usually result in non-Markovian processes. Thus, TES is a non-linear autoregressive-based scheme, encompassing both Markovian and non-Markovian processes.

The TES approach is tailored to a specific world view of what constitutes a good model, namely, that both first-order and second-order properties of empirical time series should be adequately captured. Specifically, it stipulates three requirements that should be satisfied by good prospective models:

Requirement 1: The marginal distribution of the model should match its empirical counterpart.

Requirement 2: The autocorrelation function of the model should approximate its empirical counterpart.

Requirement 3: Sample paths generated by a Monte Carlo simulation of the model should "resemble" the empirical data.

Note that the these requirements are arranged in decreasing stringency. The first two constitute quantitative goodness-of-fit criteria, whereas the third one is a qualitative requirement which cannot be defined with mathematical crispness. We argue that, in fact, there is no compelling need to. Obviously, Requirement 3 is a highly subjective statement, but its intuitive meaning and purpose should be clear: qualitatively "similar" sample paths can considerably enhance a practitioner's confidence in a proposed model. Furthermore, qualitative similarity is quite emphatically not a substitute for the two preceding quantitative criteria; rather, it is required *in addition*, not *instead*. A similar view is adopted in [23, 21, 32]. An excellent comprehensive survey of methods for constructing stochastic processes with prescribed marginals and autocorrelations may be found in [15], which defines a taxonomy of input analysis methods. In this taxonomy, TES methods, as well as those proposed in [23, 21, 32], would be classified as *approximate correlation distortion methods*.

TES processes and the underlying TES methodology conform neatly to this paradigm. First, the TES modeling methodology guarantees an exact fit to arbitrary marginal distributions. In particular, it can match any empirical density (histogram). Second, it affords a great deal of freedom in approximating the empirical autocorrelations, while maintaining an exact match to the empirical histogram. TES autocorrelation functions have diverse functional forms including monotone, oscillatory and others. And third, TES processes generate a wide qualitative variety of simulated sample paths, including cyclical and non-directional sequences.

In addition, TES processes enjoy two important computational advantages. The first advantage is that TES sequences are easily generated on a computer, and their periods are much longer than that of the underlying pseudo-random number stream. Their generation time complexity is typically not much higher than that of the underlying pseudo-random number generator, and their space complexity is negligible. The second advantage is that TES autocorrelations can be computed from fast and accurate numerical formulas, obviating time-consuming calculations of simulation-based statistics. Consequently, TES modeling of empirical data may be carried out interactively and visually. This observation motivated the design and implementation of a TES-based modeling package, called TEStool, which makes heavy use of visualization in order to provide a pleasant interactive environment for modeling dependent stationary time series. Two search approaches have been implemented: heuristic and algorithmic. The former affords more user control through direct interactive search, while the latter allows more elaborate algorithmic searches, thereby largely automating the search activity. The TEStool environment speeds up the modeling search procedure, cuts down on modeling errors and relieves the tedium of repetitive search. Since the search is formulated in visual terms, it is easy to carry out, and can be conducted by experts and non-experts alike. It appears that TES is the only input analysis method in the spirit of the three goodness-of-fit requirements listed above, which is currently supported by an elaborate modeling software environment.

This paper is organized as follows. Section 2 contains a brief overview of TES processes, an outline of the TES modeling methodology (including its heuristic and algorithmic variants), and a concise description of the TEStool software modeling environment. Section 3 demonstrates the efficacy of the TES modeling methodology and the TEStool modeling environment by a range of examples from various application domains. Finally, Section 4 contains the conclusion of this paper.

2 OVERVIEW OF TES MODELING

TES processes are treated in some detail in [25, 11, 12, 13]. An extensive overview of TES processes and the corresponding modeling methodology may be found in [26], and a short one in [27]. A brief overview extracted from these references now follows.

2.1 TES Processes

For any real x, let $\lfloor x \rfloor = \max\{\text{integer } n : \ n \le x\}$ be the integral part of x, and define $\langle x \rangle = x - \lfloor x \rfloor$ to be the fractional part of x. Let $\{V_n\}$ denote a sequence of iid random variables with a common, though arbitrary, density f_V. Further, let U_0 be uniform on [0,1) and independent of the sequence $\{V_n\}$. The random variables V_n are referred to as *innovations.*

There are two classes of TES processes: TES$^+$ and TES$^-$. Each TES model consists of two random sequences in lockstep, called the *background* and *foreground* sequence, respectively. A TES$^+$ background sequence $\{U_n^+\}$ is of the form

$$U_n^+ \;=\; \begin{cases} U_0, & n = 0 \\ \langle U_{n-1}^+ + V_n \rangle, & n > 0 \end{cases} \tag{2.1}$$

while a TES$^-$ background sequence $\{U_n^-\}$ is of the form

$$U_n^- \;=\; \begin{cases} U_n^+, & n \text{ even} \\ 1 - U_n^+, & n \text{ odd.} \end{cases} \tag{2.2}$$

The superscripts in Eqs. (2.1) and (2.2) are suggestive of the fact that TES processes achieve coverage of the full range of feasible lag-1 autocorrelations [11]; TES$^+$ processes cover the positive range $[0,1]$, while TES$^-$ processes cover the negative range $[-1,0]$. The superscripts will be omitted when the TES flavor is immaterial. It can be shown [11, 26] that the basic TES processes of Eqs. (2.1) and (2.2) constitute stationary Markovian sequences with uniform marginals on $[0,1)$, regardless of the innovation sequence selected; in fact, the choice of innovations determines just the second-order structure of a TES process.

TES background sequences are auxiliary, the real interest lying in TES foreground sequences, $\{X_n^+\}$ and $\{X_n^-\}$, obtained from Eqs. (2.1) or Eq. (2.2) by some transformation D from $[0,1]$ to the real line (called a *distortion*), i.e.,

$$X_n^+ \;=\; D(U_n^+), \quad X_n^- = D(U_n^-). \tag{2.3}$$

The TES modeling methodology (to be explained in Section 2.2) typically employs compound distortions composed of two successive transformations

$$X_n \;=\; F^{-1}(S_\xi(U_n)), \tag{2.4}$$

where $\{U_n\}$ is any background TES sequence. The inner transformation, S_ξ, is called a *stitching transformation*; it is selected from a family, parameterized by $0 \le \xi \le 1$, of the form

$$S_\xi(y) \;=\; \begin{cases} y/\xi, & 0 \le y < \xi \\ (1-y)/(1-\xi), & \xi \le y < 1. \end{cases} \tag{2.5}$$

It can be shown [11, 26] that each S_ξ, $0 < \xi < 1$, constitutes a "smoothing" transformation, in the sense that the corresponding stitched sequence $\{S_\xi(U_n)\}$ appears "smoother" than the underlying $\{U_n\}$. Moreover, all S_ξ, $0 \le \xi \le 1$, preserve uniformity [25], i.e., $S_\xi(U_n)$ is also distributed uniformly on $[0,1)$. The outer transformation, F^{-1}, is the inverse of some distribution function F. Notice that by the inversion method [3, 17], each X_n in (2.4) has marginal distribution F. Consequently, distortions of the form $D(x) = F^{-1}(S_\xi(x))$, $x \in [0,1]$, allow us to generate foreground sequences with any prescribed marginal distribution F. In particular, we can match any empirical distribution function H, obtained from an empirical density (histogram), h. The corresponding *histogram distortion* $D_H = H^{-1}$ [12] is explicitly given by

$$D_H(x) \;=\; \sum_{n=1}^{N} 1_{[C_{n-1},C_n)}(x)\,[l_n + (x - C_{n-1})\,\frac{w_n}{p_n}], \quad 0 \le x \le 1, \tag{2.6}$$

where 1_A is the indicator function of set A, N is the number of histogram cells of the form $[l_n, r_n)$, $w_n = r_n - l_n$ is the width of cell n, p_n is the probability of cell n, and $C_n = \sum_{j=1}^n p_n$ is the cumulative distribution of $\{p_n\}$ ($C_0 = 0$ and $C_N = 1$). Recall that a histogram h is mathematically a step-function density, i.e., a probabilistic mixture of uniform densities.

Our goal, however, is to construct a TES model, which not only can match a prescribed marginal distribution, but which can simultaneously approximate a prescribed autocorrelation function. Naturally, a Monte Carlo simulation of Eqs. (2.1) and (2.2) can always provide an estimate of the model's autocorrelation function, provided a sufficient sample size is generated. This approach, however, can be costly in terms of computation time, especially when high autocorrelations necessitate large sample sizes for adequate statistical reliability. It is, therefore, essential to develop fast and accurate computational techniques for these calculations. Reference [12] provides such formulas for the autocorrelation functions of $\{X_n^+\}$ and $\{X_n^-\}$, for a range of distortions of the form (2.4).

Intuitively, the modulo-1 arithmetic (fractional part operation), used in defining background TES$^+$ sequences in Eq. (2.1), has a simple geometric interpretation as a random walk on the *unit circle* (circumference 1), with random step size V_n [11, 26]. To fix the ideas, consider the important class of step-function innovation densities f_V of the form

$$f_V(x) \;=\; \sum_{k=1}^{K} 1_{[L_k,R_k)}(x)\,\frac{P_k}{R_k - L_k}, \tag{2.7}$$

corresponding to mixtures of uniform densities whose support is contained in [-0.5,0.5). These densities can be specified by triplet sets of the form $\{(L_k, R_k, P_k)\}_{k=1}^K$, where K is the number of triplets, $[L_k, R_k)$ is the support of step k, and $0 < P_k \le 1$ is the mixing probability of step k ($\sum_{k=1}^K P_k = 1$); for convenience, it is also required that triplets do not overlap (i.e., $R_k \le L_{k+1}$, $1 \le k \le K-1$). We point out that, in fact, any interval of length 1 can be used to support innovation densities, due to the modulo-1 arithmetic employed in the definition of background TES sequences. However, the symmetry of the interval $[-0.5, 0.5)$ about zero is intuitively compatible with the interpretation of innovation variates as modulo-1 increments (decrements) of a TES$^+$ sequence on the unit circle. Observe that step-function densities can approximate any density arbitrarily closely, while enjoying an ease of specification via the triplets (L_k, R_k, P_k).

A great deal is known about the qualitative behavior of sample paths and autocorrelation functions generated by TES processes, as a function of the stitching parameter and innovation density parameters. These are discussed in some detail in references [25, 11, 26], and briefly in the following section.

2.2 The TES Modeling Methodology

Assume that we have at our disposal some empirical sample path data representing a partial process history (e.g., interarrival times of packets on a communications link) to which we wish to fit a TES model. A complete specification of a TES process requires the modeler to specify a TES flavor (TES$^+$ or TES$^-$), a distortion D composed of a stitching transformation S_ξ and an inverse distribution F^{-1}, and an innovation density f_V. An outline of a typical heuristic TES modeling scenario follows [11, 26].

Selecting the TES Flavor: The selection of a TES flavor is based on the modeler's experience and knowledge of TES processes. Experience shows that TES$^+$ models are most prevalent; see Figures 1 and 6. TES$^-$ models are more appropriate for empirical sample paths and autocorrelation functions which have a zigzag appearance; see Figures 4 and 5.

Selecting the Inverse Distribution: When modeling empirical data, the inverse distribution function of the empirical data is the histogram inverse D_H, given in Eq. (2.6). It is determined by modeler-provided histogram parameters (number of cells and cell width) and the empirical data.

Selecting the Stitching Parameter and Innovation Density: The core activity of TES modeling is a heuristic search for a suitable stitching parameter, ξ, and innovation density, f_V. The modeler searches through stitching parameters in the range $[0, 1]$

and innovation densities in the space of step-function densities, whose support is contained in $[-0.5, 0.5)$. The search, though heuristic, is far from being a blind one. Rather, it is guided by the following qualitative guidelines: (a) the larger the support of the innovation density, the smaller the autocorrelation magnitude. In fact, a single innovation density step with a maximal support gives rise to a renewal (iid) TES process; (b) the more unsymmetric the innovation density is about the origin, the more oscillatory the sample paths and autocorrelation functions are; (c) the effect of $0 < \xi < 1$ is to "smooth" sample paths. The magnitude of the autocorrelation function increases as ξ approaches 0.5 from the left or from the right. In cyclical TES processes (those with$E[V_n] \neq 0$), ξ can be used to skew sample paths in accordance with the corresponding stitching transformation. The reader is referred to [11, 26] for more details.

We stress that the modeling scenario outlined above is essentially a manual procedure, guided by visual feedback and aided by the guidelines above. The modeler selects an initial model arbitrarily. Thereafter, the modeler modifies the model interactively and may loop back to any step based on the quantitative fit of the current model's autocorrelation function and the qualitative fit of Monte Carlo sample paths to their respective empirical counterparts.

While heuristic TES modeling searches are fully under user control, they can be lengthy and try the user's patience. To alleviate this problem, a TES modeling algorithm has been devised to supplement the heuristic mode [14]. The algorithm first carries out a brute-force search over a subspace of step-function innovation densities and various stitching parameters, and calculates the corresponding TES model autocorrelation functions. The best n combinations of pairs, (ξ, f_V), are then selected, that is, those TES models whose associated autocorrelation functions have the smallest error (sum of squared deviations) relative to the empirical autocorrelation function. The algorithm then performs non-linear optimization on each model candidate to further reduce the aforementioned error. Finally, the analyst peruses the results and selects from the n optimized candidate models the one whose Monte Carlo sample paths bear the "most resemblance" to the empirical record, in addition to having a small error. Preliminary experience has shown that algorithmic TES modeling often produces better and faster results than its heuristic counterpart [14].

Finally, we point out that TES modeling can be extended to time series which are not stationary or are composite models with stationary components. For instance, Section 3.2 contains an example of a non-stationary superposition process with stationary TES components. Another example is the Markov-Renewal-Modulated process, with modulated stationary TES components, described in [28].

2.3 The TEStool Modeling Environment

The modeling activities of both the heuristic and algorithmic TES modeling would be carried out most efficiently with visually-oriented computer support. TEStool is a visual interactive software environment designed for fitting TES models to empirical data [7, 27, 9]. It provides services to construct and modify TES models in the heuristic search mode, to generate simulated sample paths and to examine their statistics, as well as to launch algorithmic searches.

TEStool distinguishes between three types of statistics. *Empirical* statistics are those associated with the empirical data (sample paths, histograms, autocorrelation functions and spectral density). *Simulated* statistics are similarly calculated (estimated) from Monte Carlo simulations of TES models. *Analytical* statistics consist of numerical computations of autocorrelation functions and spectral densities of TES models, based on analytical formulas developed in [11, 12].

Figures 1 – 5 display each a typical TEStool screen composed of four tiled canvases. These contain, respectively, sample paths (upper-left canvas), histograms (upper-right canvas), autocorrelation functions (lower-left canvas), and a graphical specification of a TES model (lower-right canvas). The first three canvases each contain a pair of statistics: the respective TES model statistics (diamonds), superimposed for comparison on their empirical counterparts (bullets). The fourth canvas consists of a joint specification of a TES sign (flavor), a step-function innovation density, a stitching parameter and an inverse-distribution distortion; the latter is selected from a menu, including the histogram distortion (2.6) constructed from the empirical histogram. A TES model can also be specified in TEStool in text mode by populating text fields in a popup subwindow. The buttons in the top border of the display and at the bottom of each canvas support various services. These include reading and writing datasets, subdividing the screen real estate, opening a TES specification window or menu, performing various types of computations and quitting the session. The most important services are the visual specification of a TES model and the interactive computations associated with heuristic searches, and the algorithmic search facilities.

The advantage of a graphical specification stems from the fact that a visual representation is easy to grasp; furthermore, it lends itself to intuitive visual interactions, which significantly increase the heuristic modeling efficiency, while substantially decreasing modeling tedium. In the visual specification mode, the mouse is used to draw and manipulate non-overlapping rectangles representing a step-function density. For example, rectangle "stretching" is used to modify a step's height and width, while rectangle "dragging" is used to translate a step horizontally. In the visual interactive mode, any changes in model specification trigger an immediate recomputation and display update of model statistics. The modeler can then compare the new statistics to their empirical counterparts for goodness of fit, and decide whether the current fit is satisfactory, or whether another iteration is required. The heuristic search can proceed efficiently, and the visual feedback has the additional advantage of aiding the modeler in learning from experience. Finally, since the search is cast in a visual-interactive mold, it can be performed by experts and non-experts alike in the same way that a player of an arcade game can concentrate on the task at hand without having to understand the underlying software details.

The algorithmic TES modeling activity is controlled by a separate menu which is mainly used to specify the extent of the algorithmic search [14].

3 APPLICATIONS OF TES MODELS

This section surveys a range of models constructed via the TES modeling methodology from empirical data. It presents examples from a variety of application domains, including compressed-video traffic in high-speed telecommunications networks, machine fault arrivals in a manufacturing plant, financial time series, and texture synthesis for graphics applications. These illustrative examples are meant to demonstrate the efficacy and versatility of the TES modeling approach.

3.1 H.261-Compressed Video

Data compression (coding) is extensively used to reduce the transmission bandwidth requirements of telecommunications traffic. The idea is to code the data at the source (thereby compressing it to a fraction of its original size), transport the compressed data over a network, and then decode the data at its destination. Video service in emerging ISDN (Integrated Services Digital Network) is a typical application, for which the exact reproduction of the original signal is not necessary. In fact, redundant visual information (to which the human eye is relatively insensitive) may be removed without degrading the perceptual quality of the decoded image, with the additional benefit of substantially reducing the requisite transmission bandwidth. H.261 is a popular coding standard, which makes use of DCT (Discrete Cosine Transform) and other techniques to compress video spatial units (frames or subframes) [22, 30]. Since such coded units have random (but highly autocorrelated) transmission requirements (say, in bits), the corresponding coding schemes are referred to as VBR (Variable Bit Rate), the bit rate being the compressed picture (frame) size.

Figure 1 depicts the results of a heuristic TES modeling for an empirical sample path of frame bit rates, in which the coding scheme used was a variant of the H.261 compression standard [1]. The model was used, in turn, to study the performance of a coded video multiplexer [19]. The lower-right canvas contains a visual specification of a TES$^+$ model with stitching parameter $\xi = 0.5$ and an innovation density function consisting of two steps (a probabilistic mixture of two uniform densities). We remark that the two-step innovation density of Figure 1 improves on the single-step innovation density reported in the original study [19]. Indeed, the original model was created painstakingly from repeated batch runs without the benefit of TEStool's visual interactive facilities; the heuristic search for the original model took about a week, whereas the model reported here was constructed via TEStool in under an hour.

Figure 2 depicts the corresponding modeling results, obtained by an algorithmic search for a TES model, using the same empirical bit rate data. A comparison of Figures 1 and 2 is instructive. First, an examination of the lower-right canvases reveals that the algorithmic search yielded a much more elaborate innovation density in Figure 2 than the two-step density, obtained by a heuristic search, in Figure 1. Certainly, manual generation of elaborate innovation densities is limited by physical factors such as human dexterity, eye-hand coordination and sheer tedium, to which computers are not subject, though humans are still much better at pattern recognition. Second, in this case the more elaborate innovation density has evidently given rise to a better model: not only is the autocorrelation function in Figure 2 slightly more accurate than its counterpart in Figure 1 (lower-left canvases), but the the TES sample path in the former resembles the empirical data far more than in the latter (upper-left canvases). The histogram fit is identical as expected (upper-right canvases). We also mention that the elapsed time of the algorithmic search was on the order of minutes, thereby comparing favorably with that of the heuristic search.

3.2 MPEG-Compressed VBR Video

MPEG (Moving Picture Expert Group) is an emerging family of compression standards designed to encode audio-visual signals over broadband transmission channels [18]. This section focuses on MPEG-video, designed to compress a full-motion video stream into a bit

[1]The empirical data were provided by Amy Reibman, Princeton University.

rate of about 1.5 Mbits/second. The importance of MPEG derives from its planned central role in facilitating future delivery of multi-media services to customer premises.

A salient feature of MPEG, from a modeling point of view, is its use of prediction and interpolation to achieve high compression ratios, while maintaining acceptable picture quality. Coded picture sequences in MPEG are composed of cycles. A coded picture can be either an Intrapicture (I-frame), Predicted picture (P-frame) or Bidirectionally Predicted picture (B-frame). The sequence of picture (frame) types within each cycle are deterministic, though the corresponding bit rates are random. MPEG type sequences can be chosen as an MPEG parameter, depending on the application. The particular type sequence chosen in the case study described here had a length-nine cycle of the form IBBPBBPBB... (see Figure 3). Observe that the marginal distributions of I-frames (largest magnitudes), P-frames (medium magnitudes) and B-frames (smallest magnitudes) are apparently very different. This is indeed the case, as the three types were listed in increasing compression ratios. Consequently, MPEG-compressed sequences are decidedly non-stationary, which precludes the straightforward modeling approach described in Section 3.1.

The modeling approach required for a good MPEG-coded video model was more complex [2] [31]. First, each subsequence of MPEG frame types was modeled as a separate TES sequence as in Section 3.1 — I-frames and B-frames each by a TES^+ model and P-frames by a TES^- model. Second, a Monte Carlo simulation generated each of the three MPEG frame type bit rate subsequences, and the sample paths were superposed (interleaved) from the three TES bit rate models to form the correct type sequence. And third, in order to induce cross-correlations into frame bit rates comprising individual cycles, the TES background variate of each cycle-inaugurating I-frame was reused to obtain the corresponding background variates for the first P-frame and B-frame within the same cycle. Subsequent P-frames and B-frames within the same cycle were generated normally from their individual TES models. This modeling device led to more accurate statistics, as compared to superposing stochastically independent subsequences; see [31] for more details.

Figure 3 displays the modeling results for the aforementioned superposed MPEG model following the format of Section 3.1, except that the lower-right canvas displays the spectral densities corresponding to the autocorrelation functions shown in the lower-left canvas. It is important to realize that the autocorrelation functions and spectral densities in Figure 3 were formally computed from a single sample path *as if* the sequences were stationary, and therefore represent averaged estimates of different correlation coefficients. Nevertheless, they provide an indication of how well the superposed TES models capture temporal dependencies exhibited by the empirical data, since they were all computed from the corresponding sample paths in the same way.

Figure 4 displays the modeling results for the P-frame subsequence — one of the component TES models used to construct the superposed model. Two features are noteworthy here. First, the model was obtained using the algorithmic modeling facilities of TEStool; and second, the algorithm settled on a TES^- model in order to capture the zigzag nature of both the empirical data and its autocorrelation function.

Both Figure 3 and 4 show excellent agreement of all statistics. In particular, the empirical distributions are well-captured even in the superposed model, for which an exact fit of the marginal distribution is not guaranteed, it being a composite of TES models rather than a pure TES model. Furthermore, the autocorrelation functions fit very closely (they appear

[2] The empirical data were provided by Dipankar Raychaudhuri and Daniel Reininger, David Sarnoff Research Center.

hardly distinguishable to the unaided eye), and the corresponding sample paths exhibit a marked resemblance of their functional form.

We mention that the video TES-based models, in this section and the previous one, were based on relatively short empirical records, on the order of hundreds, corresponding to a single video scene. An elaborate Markov-Renewal-Modulated TES (MRMT) model for an entire movie (some 171000 bit rate observations) is detailed in [28] for the "Star Wars" movie, compressed under a variant of the JPEG coding [6]. Using clustering analysis, all scenes were grouped into four classes, each corresponding to a state in a modulating Markov chain, while scene durations were modeled by renewal processes, one per scene class. The resultant MRMT model showed excellent agreement with the empirical first-order and second-order statistics.

3.3 Machine Fault Arrivals

Many of the faults in semiconductor equipment are related to a machine processing wafers "out of specification", rather than mechanical failures. In practice, it is often difficult to get the equipment back "in specification". Consequently, several corrective actions may be related to an inaugurating "out of specification event", resulting in bursty fault arrivals. Such burstiness may manifest temporal dependencies among up times (the other cause of burstiness may be a multimodal marginal distribution of up times). Since poor equipment reliability has been identified as a major cause of uncertainty in semiconductor manufacturing [8], it is very important that the corresponding reliability be modeled correctly. Traditional reliability models, however, analyze the arrival of faults by assuming that their interarrival times (up times) are independent random variables [1, 33].

Figure 5 depicts the results of a heuristic TES modeling for machine up times, incorporating temporal dependence and based on measurements taken from a semiconductor manufacturing line. Specifically, these data represent several months of up times of a machine engaged in the photolithographic processes of a wafer fabrication facility [3]. An interesting feature of this example is that it relies solely on the more rare TES$^-$ model, so as to capture the functional form of the autocorrelation function. We remark that for $\xi = 0.5$, it can be shown that TES$^+$ and TES$^-$ are stochastically equivalent. Although the graphical specification panel shows $\xi = 0.448$, it was found that this TES$^-$ model provided a substantially better approximation for the empirical autocorrelation function than any TES$^+$ model tested.

The reader's attention is drawn to several aspects of the model fit quality. First, the simulated histogram closely approximates its empirical counterpart as advertised (recall that TES theory guarantees an exact asymptotic match as the simulated sample size increases). Second, since the sample size was very small (some 49 observations), it makes sense to attempt to match only the first 7–8 autocorrelations or so. Within this range, it would be preferable to approximate shorter-lag autocorrelations at the expense of longer-lags. Furthermore, rather than achieving an exact match to the leading autocorrelations, one primarily aims to capture their functional form (in our case their oscillatory decreasing structure). Indeed, while the model does not precisely match the empirical autocorrelation function, it does provide a good approximation which captures its functional form. And third, although sample path "resemblance" is a highly subjective judgment, the qualitative similarity of the sample paths is surprisingly close to the unaided eye, thereby increasing our confidence in the model.

[3] The empirical data were provided by John Fowler, Sematech.

3.4 Financial Time Series

Stochastic models of financial time series are extensively used by traders to guide buy/sell decisions in financial markets. There are two main uses for financial models. The first is short-term forecasting to guide day-to-day trading. The second is the generation of scenarios for use in mid-term to long-term planning (1 to 5 years). Most quantitative investment strategies generate Monte Carlo scenarios to estimate internal parameters of various financial models [10]. For example, there are many proposals for modeling interest rates in which the underlying model is a diffusion process. Their purpose is, typically, to estimate the current fair market value of a security, rather than to forecast its future price. These models share the shortcoming that they are difficult to calibrate in a formal manner.

The TES modeling methodology has a natural advantage in this regard, since it is designed to fit first-order and second-order statistics of empirical data, and the graphical displays could provide much-needed information. Having obtained a TES model based on historical data whose current (most recent) observation represents the present value (current state) of a financial empirical time series, the modeler can generate TES sample paths conditioned on the current state. In particular, one can generate multiple TES sample paths having the current state as their initial value. Furthermore, there is no need to separate the modeling and forecasting stages. By forcing a TES model to have the current state as its starting point, one can use *time reversal* [4] [16] to simulate each sample path backward in time from the present (modeling), and simultaneously, forward in time from the present (forecasting), thereby combining the two stages into one. To see that time-reversed paths can be easily generated, we recall the fact that all background TES^+ and TES^- sequences are stationary Markovian, and consequently, the corresponding time-reversed processes are readily obtained; see [11], Section 4, for details.

Figure 6 displays a TES model of a sequence of quarterly "real" returns of Treasury Bills (nominal interest rates minus inflation rate) for the period 1950-1990 [5]. Each data point is the average of all daily returns in the respective quarter. The quality of the fit is similar to that in the previous example. In particular, the model approximates the leading autocorrelations and gives rise to sample paths with considerable qualitative resemblance to the observed historical price sequence. In view of the relatively low magnitude of the empirical autocorrelations, it is doubtful that such "black box" models would serve as reliable predictors of future yields. However, they can be used to generate random scenarios to drive financial simulations.

3.5 Texture Synthesis

Simulations of two dimensional random fields are used in diverse disciplines including meteorology, picture synthesis and compression, and visual simulation [2, 4, 20]. In particular, an operation known as *texture synthesis* is commonly used to fill in the visual interior of delineated objects. Examples include water surfaces, clouds, tree bark, etc.

In this section, we demonstrate texture synthesis based on families of coupled 2-dimensional TES sequences $\{(X_n, Y_n)\}$ taking values in a rectangle of the Euclidean plane. We start with an $M \times N$ pixel array (bitmap), which is initially set to zero (black). Next, the process $\{(X_n, Y_n)\}$ is simulated. Each pair (X_n, Y_n) is translated into plane coordinates

[4] Any stochastic process can be *reversed in time*, but only select cases are *reversible*, namely, the forward and backward processes obey each the same probability law. We do not claim that TES processes are, in general, reversible.

[5] The empirical data were provided by John Mulvey, Princeton University.

and mapped to a bitmap pixel by scaling it into a pair of integer indices in the range $\{1, \ldots, M\} \times \{1, \ldots, N\}$. The corresponding pixel in the bitmap is then colored with a maximal value (white). This procedure is repeated until a desired fraction of the bitmap pixels has been colored; the overall intensity of the image corresponds to this fraction. Finally, the resulting black-and-white image is convolved with a 2-dimensional filter to obtain a digital approximation to a continuous image.

For TES-based textures, assume that both $\{X_n\}$ and $\{Y_n\}$ are coupled TES processes. More specifically, define

$$\begin{aligned} X_n &= \langle X_{n-1} + L_X + (R_X - L_X)\, Z_n \rangle, && (3.8) \\ Y_n &= \langle Y_{n-1} + L_Y + (R_Y - L_Y)\, (Z_n + \gamma_{X,Y}\, W_n) \rangle, && (3.9) \end{aligned}$$

where $-0.5 \le L_X < R_X < 0.5$, $-0.5 \le L_Y < R_Y < 0.5$, $0 \le \gamma_{X,Y} < 1$, the Z_n are iid uniform on $[0, 1)$, and the W_n are iid uniform on $[-1, 1)$ and independent of the Z_n.

Figure 7 displays some TES-based textures of the form (3.8) – (3.9), for various TES parameter settings of L_X, R_X, L_Y, R_Y, $\gamma_{X,Y}$, and different filters [6]. In particular, a band-pass filter was used to produce the upper-left image, while a low-pass Gaussian filter was used for the remainder.

Referring to Figure 7, the effect of TES parameters on the visual features of textures can be qualitatively described as follows.

1. For large values of $R_X - L_X$ and $R_Y - L_Y$, the marginal random walks, $\{X_n\}$ and $\{Y_n\}$, are approximately iid (lower-left texture).

2. A small value of $R_X - L_X$ or $R_Y - L_Y$ produces streaks; the direction of the streaks can be changed by choosing different magnitudes for the L and R parameters (lower-right texture).

3. When the L and R parameters have the same sign, one obtains characteristic oscillatory textures. When both processes are oscillatory, a quasi-periodic texture results (upper-right texture).

Finally, the influence of the $\gamma_{X,Y}$ coupling parameter is not dramatic.

[6] These textures were generated by John P. Lewis, NEC USA, Inc.

4 CONCLUSION

This paper has overviewed TES modeling and surveyed a range of TES models, fitted to empirical data from a variety of application domains. The purpose of the survey is to demonstrate the efficacy of the TES modeling approach in simultaneously capturing first-order and second-order statistics of empirical time series, and to exhibit its versatility and flexibility by addressing a wide range of applications with a mix of TES-based models.

The major applications of TES to date have been to capture the burstiness exhibited by autocorrelated traffic in high-speed telecommunications networks, mainly VBR video. Two TES-based modeling approaches to coded video have been exemplified in some detail: a simple TES source model of frame bit rates for H.261 coding and a superposition of TES models interleaved in a fixed order for MPEG coding. A source model of machine fault interarrival times in a manufacturing context has demonstrated the ability of TES models to capture bursty fault arrivals. An example from financial modeling suggests that TES models can capture the statistical signature of financial time series, and demonstrates its potential to model random phenomena which drive simulations of financial scenarios. A final example illustrates the manner in which coupled 2-dimensional TES sequences can synthesize diverse, interesting-looking graphical textures.

Based on these examples and our own cumulative experience, we propose the TES modeling methodology as a new and powerful approach to input analysis, particularly in the context of Monte Carlo simulations. Practical TES modeling currently relies on computer support to carry out heuristic or algorithmic searches for TES models which conform to the three modeling requirements of Section 1. The TEStool visual interactive software package has been presented as an effective TES modeling environment. The user interface of TEStool casts the heuristic search into an intuitive activity of successively modifying a visual parametric representation of a TES model, guided by visual feedback and some knowledge of TES processes. The TEStool modeling environment has been found to yield remarkably accurate TES models in a reasonable amount of time. It also alleviated the tedium of repetitive heuristic search, turning modeling into a pleasant activity not unlike a video arcade game. Finally, TEStool's algorithmic search facilities largely automate the search procedure, and often yield improved models.

Acknowledgments

We are grateful to John Fowler, John Mulvey, Dipankar Raychaudhuri, Amy Reibman, and Daniel Reininger for providing the empirical data and supplying us with valuable information on the respective applications. We especially thank David Jagerman for a careful reading of the manuscript, and John Lewis for programming and generating the TES textures.

References

[1] R.E. Barlow and F. Proschan. *Statistical Theory of Reliability and Life Testing: Probability Models.* Holt, Rinehart and Winston, New York, 1975.

[2] P. Bouthemy and A. Beneviste. Modeling of Atmospheric Disturbances in Meteorological Pictures. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. PAMI-6, No. 5, 587–600, 1984.

[3] P. Bratley, B.L. Fox and L.E. Schrage. *A Guide to Simulation.* Springer-Verlag, New York, NY, 1987.

[4] R. Chellappa, S. Chatterjee and R. Bagdazian. Texture Synthesis and Compression using Gaussian-Markov Random Field Models. *IEEE Trans. on Systems, Man, and Cybernetics*, Vol. 15, No. 2, 298–303, 1985.

[5] K.W. Fendick, V.R. Saksena and W. Whitt. Dependence in Packet Queues. *IEEE Trans. on Comm.*, Vol. 37, 1173–1183, 1989.

[6] M.W. Garrett. Contributions Toward Real-Time Services on Packet Networks. Ph.D. Dissertation, Columbia University, 1993.

[7] D. Geist and B. Melamed. TEStool: An Environment for Visual Interactive Modeling of Autocorrelated Traffic. *Proceedings of ICC '92*, Chicago, Illinois. 1285–1289, 1992.

[8] J.M. Harrison, C.A. Holloway, and J.M. Patell. Measuring Delivery Performance: A Case Study from the Semiconductor Industry. In *Measures for Manufacturing Excellence*, R.S. Kaplan (ed.), Harvard Business School Press, 1990.

[9] J.R. Hill and B. Melamed. TEStool: A Visual Interactive Environment for Modeling Autocorrelated Time Series. *Performance Evaluation*, 1995, to appear.

[10] J. Hull. *Options, Futures, and Other Derivative Securities.* Prentice-Hall, 1989.

[11] D. Jagerman and B. Melamed. The Transition and Autocorrelation Structure of TES Processes Part I: General Theory. *Stochastic Models*, Vol. 8, No. 2, 193–219, 1992.

[12] D. Jagerman and B. Melamed. The Transition and Autocorrelation Structure of TES Processes Part II: Special Cases. *Stochastic Models*, Vol. 8, No. 3, 499–527, 1992.

[13] D. Jagerman and B. Melamed. The Spectral Structure of TES Processes. *Stochastic Models*, Vol. 10, No. 3, 599–618, 1994.

[14] P. Jelenkovic and B. Melamed. Algorithmic Modeling of TES Processes. *IEEE Trans. on Automatic Control*, to appear.

[15] G.E. Johnson. Construction of Particular Random Processes. *Proc. of the IEEE*, Vol. 82, No. 2, 270–285, 1994.

[16] F.P. Kelly. *Reversibility and Stochastic Networks.* Wiley, 1979.

[17] A.M. Law and W.D. Kelton. *Simulation Modeling & Analysis*, (second edition). McGraw-Hill, 1991.

[18] D. Le Gall. MPEG: A Video Compression Standard for Multimedia Applications. *Communications of the ACM*, Vol. 34, 46–58, 1991.

[19] D.-S. Lee, B. Melamed, A. Reibman and B. Sengupta. Analysis of a Video Multiplexor using TES as a Modeling Methodology. *Performance Evaluation*, Vol. 16, 21–34, 1992.

[20] J. Lewis. Generalized Stochastic Subdivsion. *ACM Trans. on Graphics*, Vol. 6, No. 3, 167–190, 1987.

[21] P.A.W. Lewis and E. McKenzie. Minification Processes and Their Transformations. *J. Appl. Prob.*, Vol. 28, 45–57, 1991.

[22] M. Liou. Overview of the px64 kbit/s Video Coding Standard. *Communications of the ACM*, vol. 34, no. 4, 59–63, 1991.

[23] B. Liu and D.C. Munson. Generation of a Random Sequence Having a Jointly Specified Marginal Distribution and Autocovariance. *IEEE Transactions on Acoustics, Speech and Signal Processing* Vol. 30, No. 6, 973-983, 1982.

[24] M. Livny, B. Melamed and A.K. Tsiolis. The Impact of Autocorrelation on Queuing Systems. *Management Science*, Vol. 39, No. 3, 322–339, 1993.

[25] B. Melamed. TES: A Class of Methods for Generating Autocorrelated Uniform Variates. *ORSA J. on Computing*, Vol 3, 317–329, 1991.

[26] B. Melamed. An Overview of TES Processes and Modeling Methodology. In *Performance Evaluation of Computer and Communications Systems*, (L. Donatiello and R. Nelson, Eds.), 359–393, Lecture Notes in Computer Science, Springer-Verlag, 1993.

[27] B. Melamed, D. Goldsman and J.R. Hill. The TES Methodology: Nonparametric Modeling of Stationary Time Series. *Proceedings of WSC '92*, Arlington, Virginia, 135–144, 1992.

[28] B. Melamed and D. Pendarakis. A TES-Based Model for Compressed "Star Wars" Video. *Proceedings of GLOBECOM Communications Mini Conference*, San Francisco, California, 120–126, November, 1994.

[29] B.E. Patuwo, R.L. Disney and D.C. McNickle. The Effects of Correlated Arrivals on Queues. *IIE Transactions*, Vol. 25, No. 3, 105–110, 1993.

[30] A.R. Reibman. DCT-based Embedded Coding for Packet Video. *Signal Processing: Image Communication*, Vol. 3, 231–237, 1991.

[31] D. Reininger, B. Melamed and D. Raychaudhuri. Variable Bit Rate MPEG Video: Characterization, Modeling and Multiplexing. *Proceedings of the 14-th International Teletraffic Congress*, Antibes Juan-les-Pins, France, Vol. 1a, 295–306, June 1994.

[32] B.W. Schmeiser. Simulation Experiments. In *Handbook of Operations Research and Management Science: Vol. 2* (D. Heyman and M. Sobel, editors). North Holland, New York, 1990.

[33] S. Zacks. *Introduction to Reliability Analysis: Probability Models and Statistical Methods*. Springer-Verlag, New York, 1992.

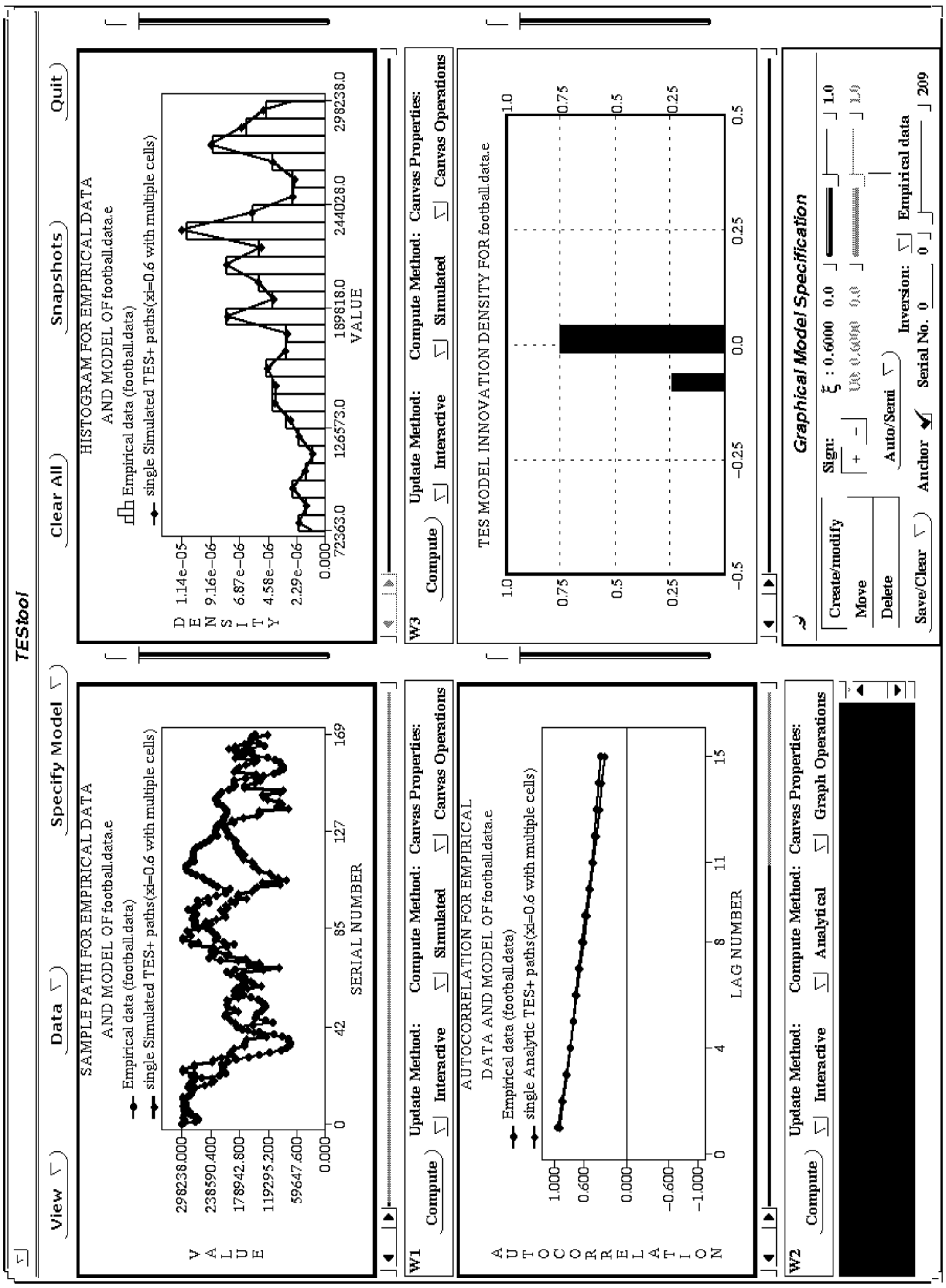

Figure 1: A TES$^+$ model of H.261-compressed VBR video, obtained by a heuristic search.

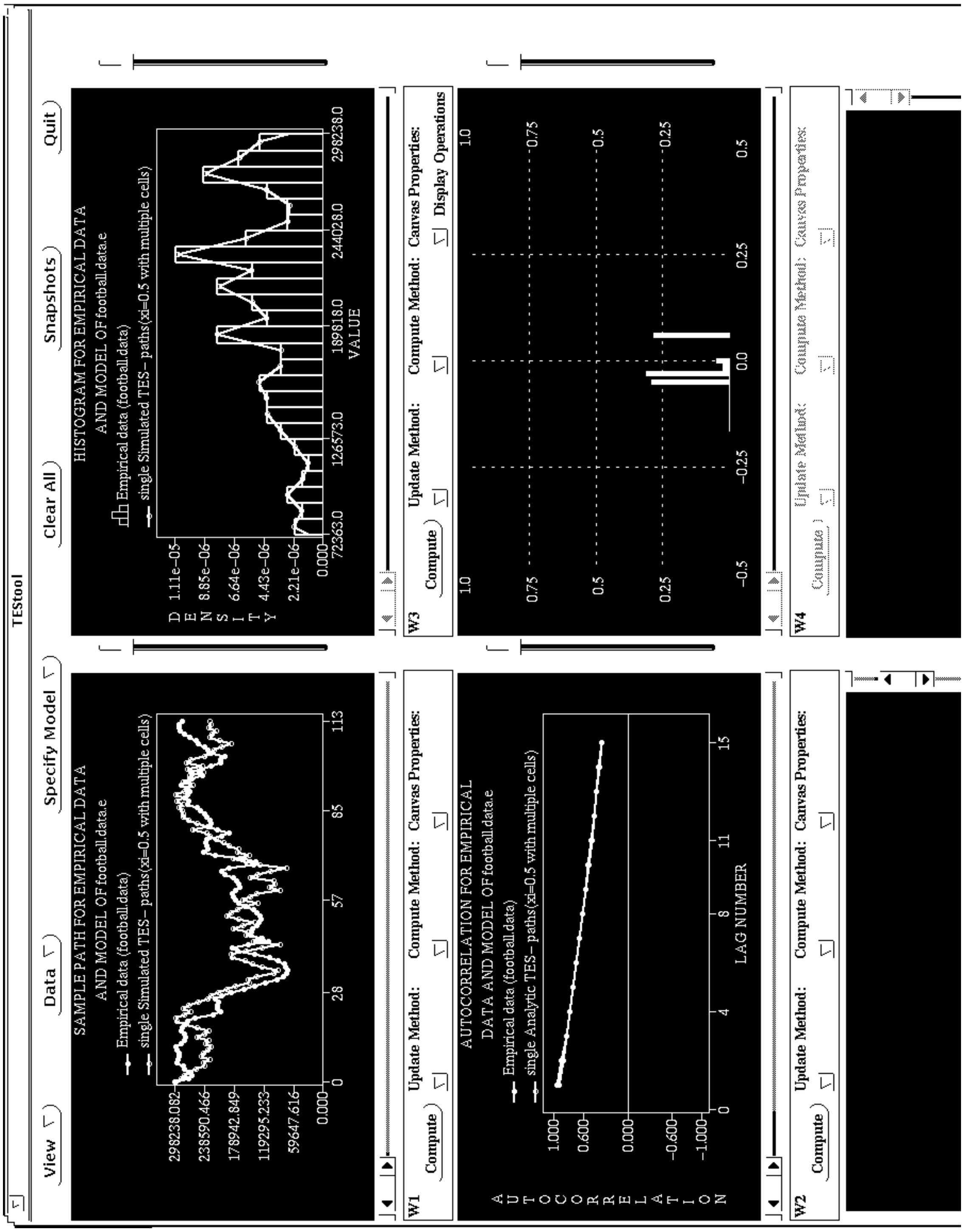

Figure 2: A TES$^+$ Model for a H.261-Compressed VBR video, obtained by an algorithmic search.

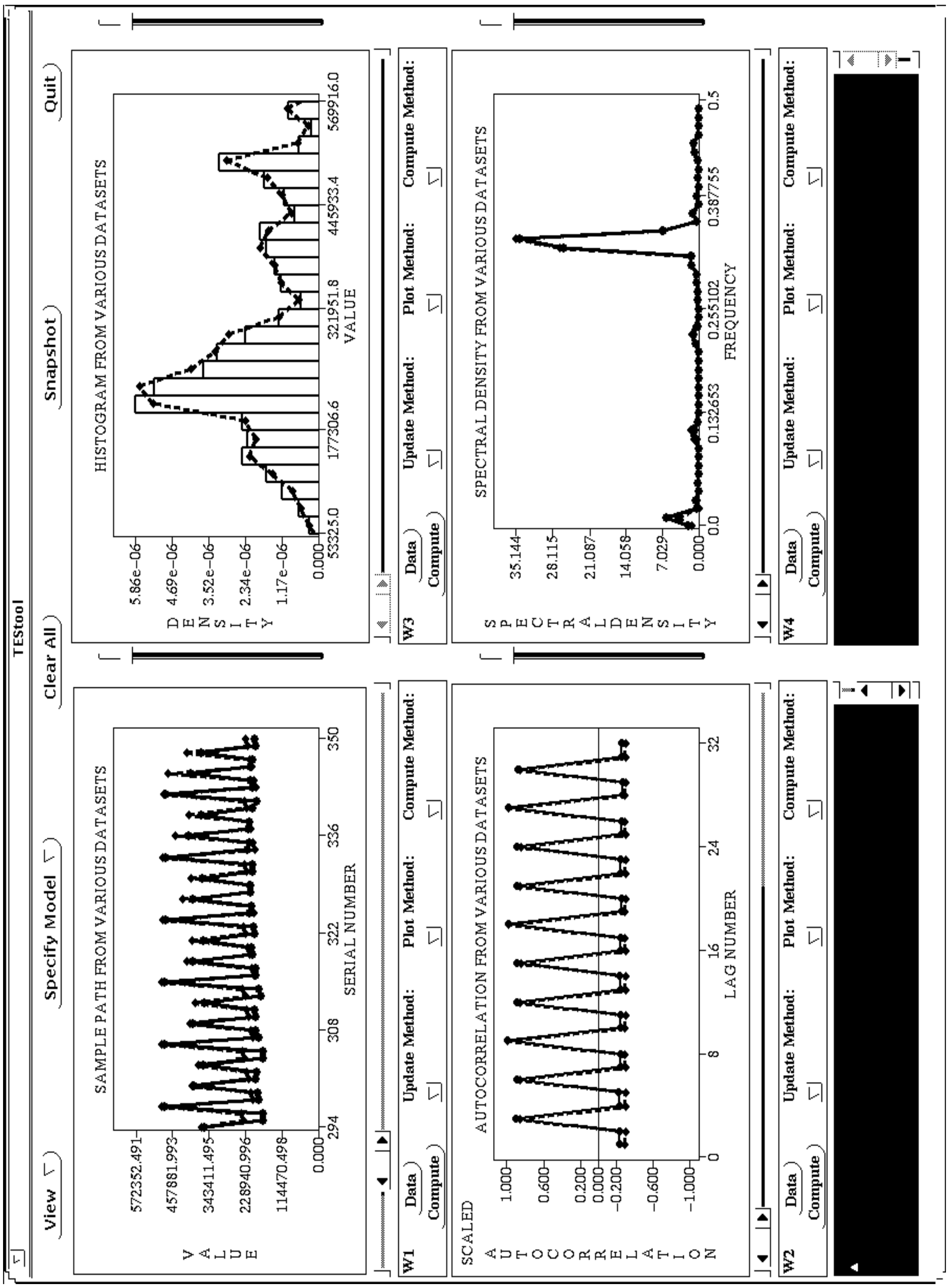

Figure 3: A composite (superposed) TES-based model of MPEG-compressed VBR video.

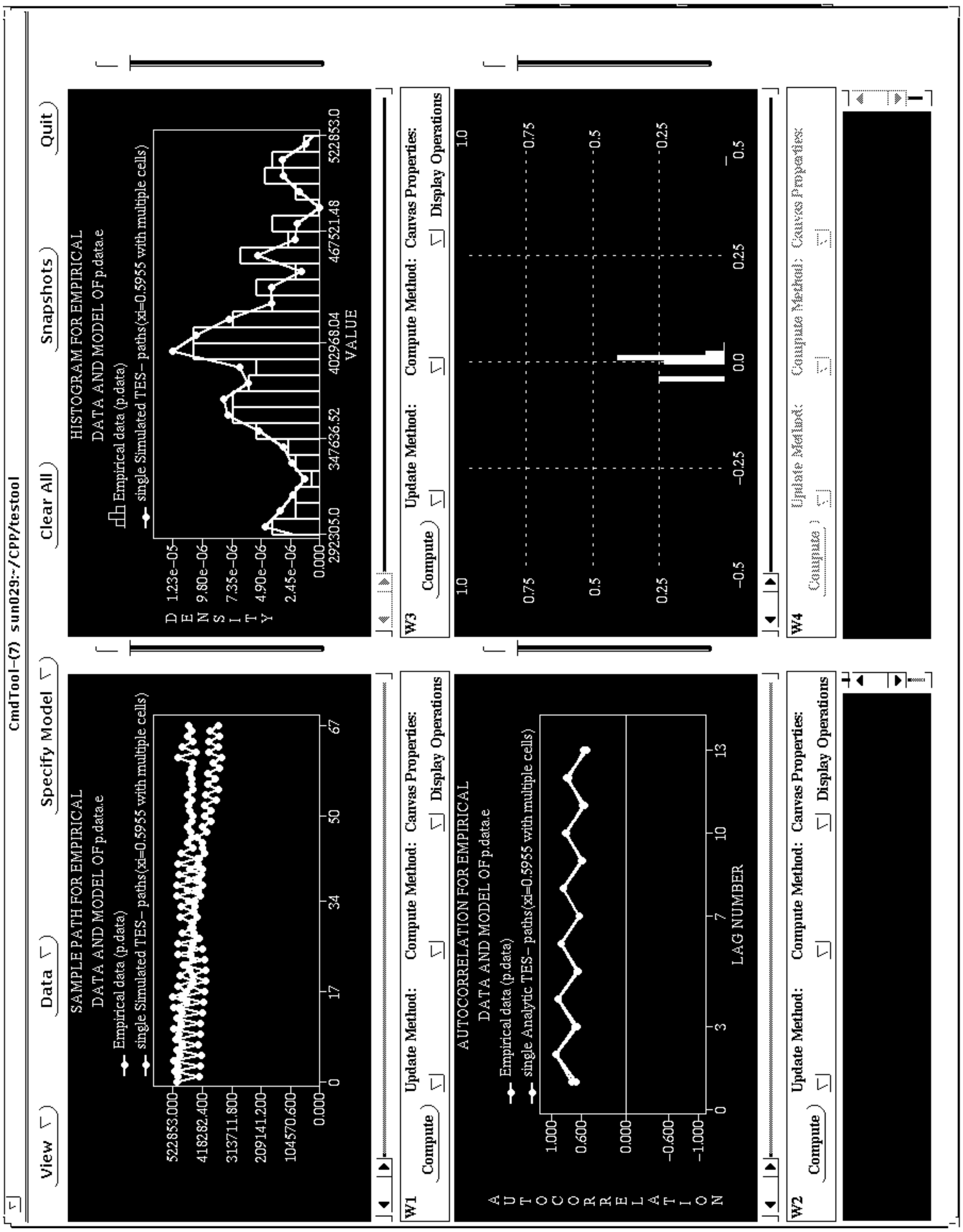

Figure 4: A TES^{-} Model for the P-Frame subsequence from an MPEG-Compressed VBR video, obtained by an algorithmic search.

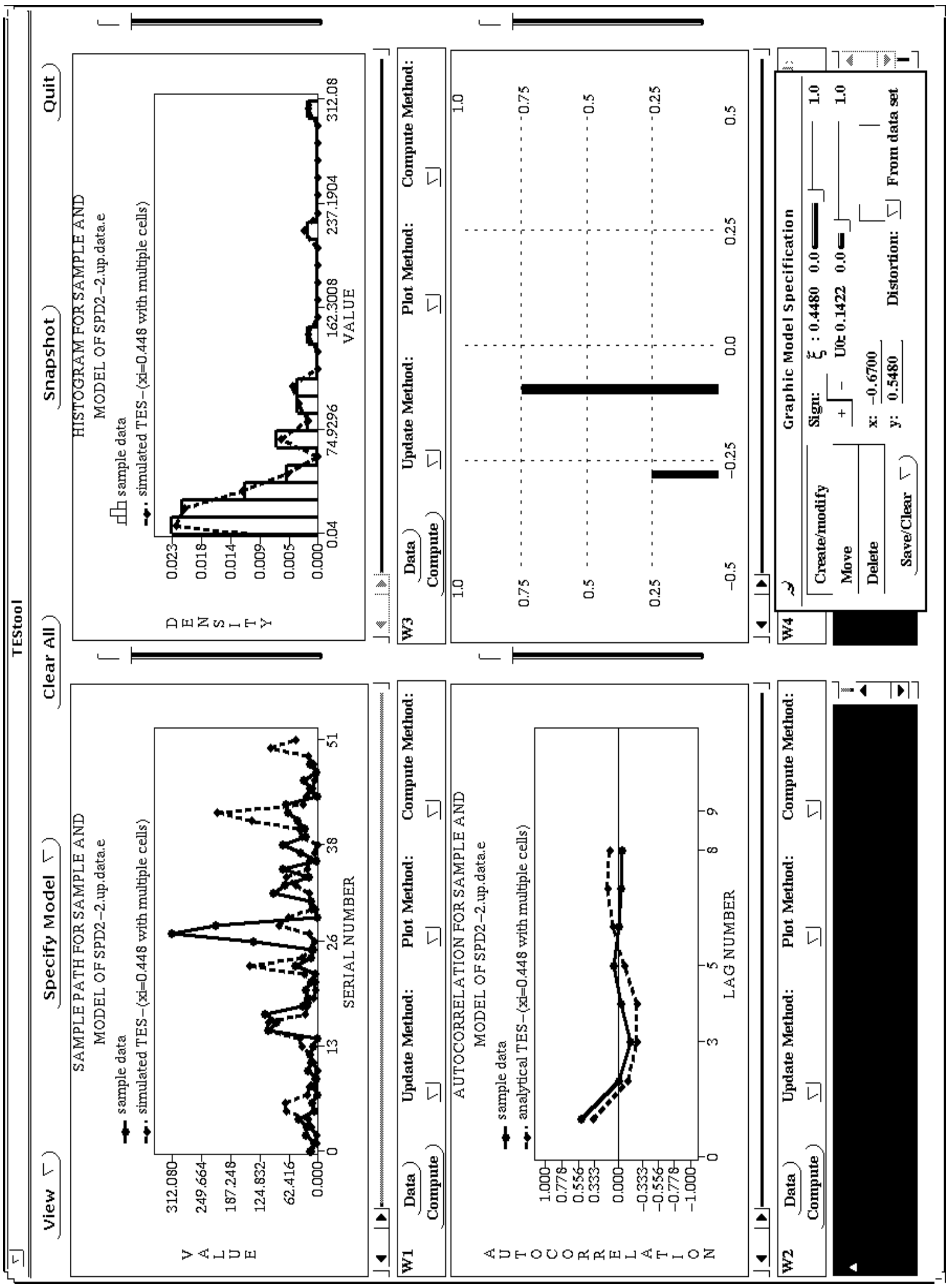

Figure 5: A TES$^-$ model of machine fault interarrival times, obtained by a heuristic search.

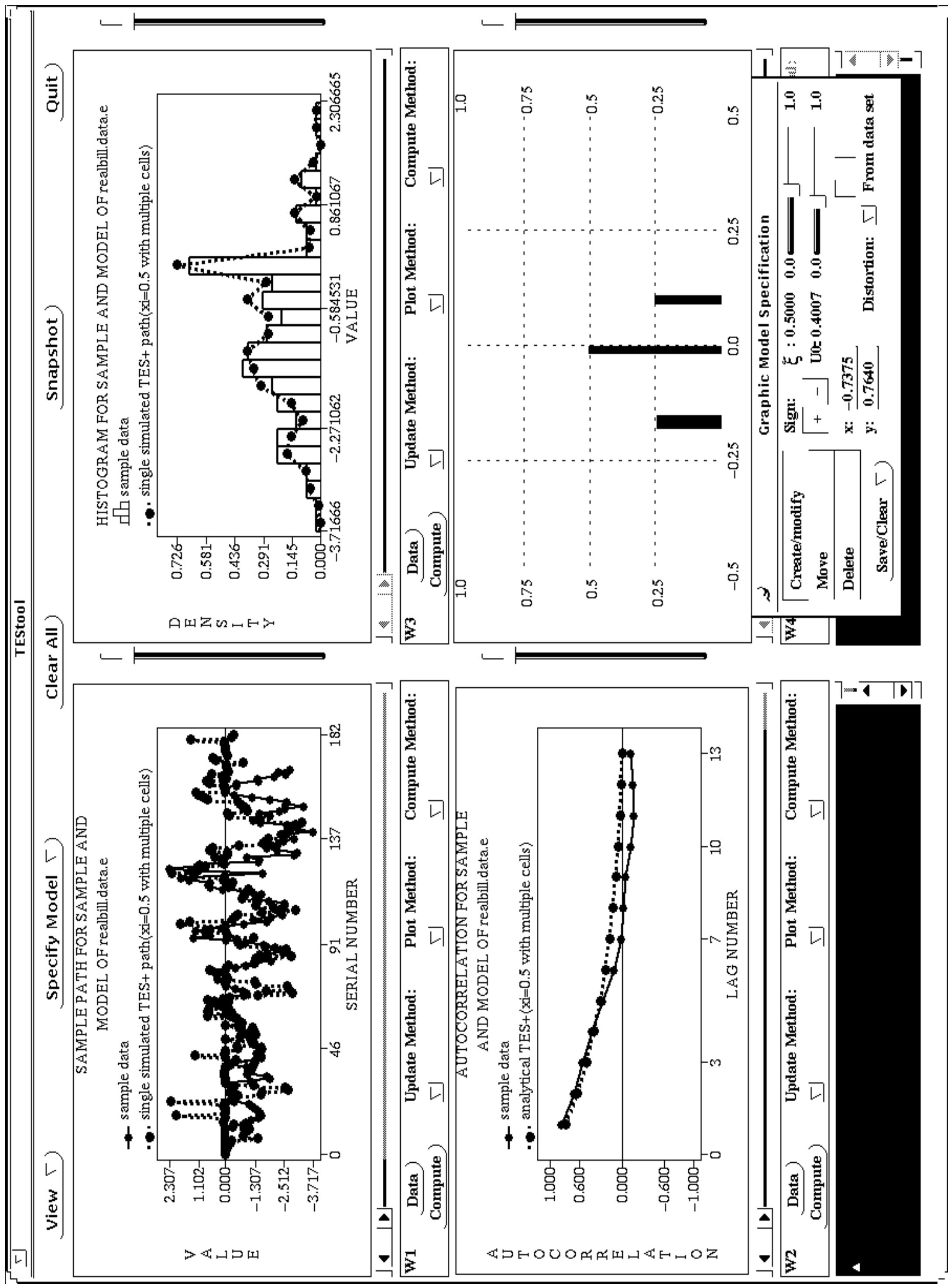

Figure 6: A TES$^+$ model of real Treasury Bill yields, obtained by a heuristic search.

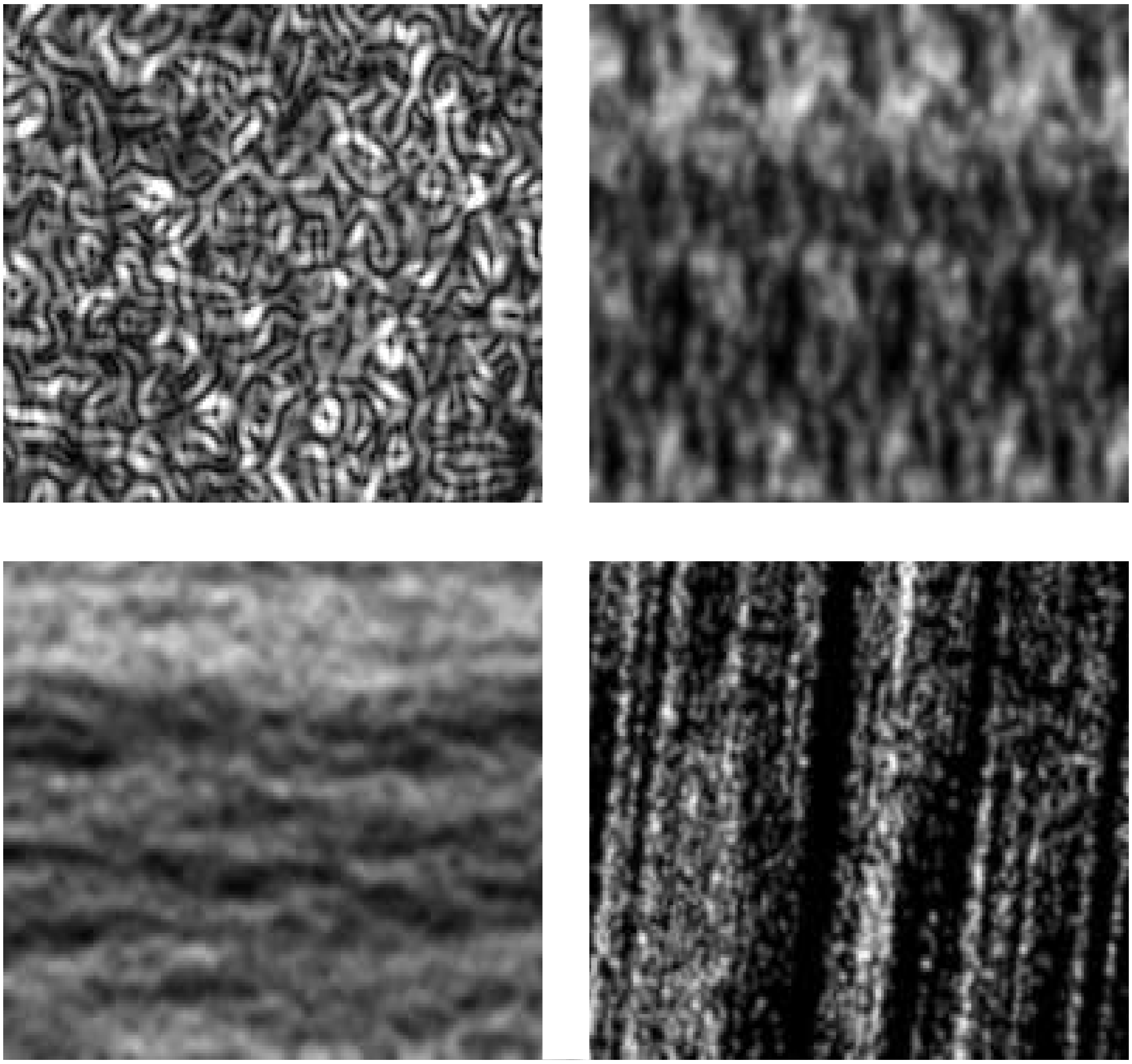

Figure 7: Graphics textures, generated by coupled 2-dimensional TES^+ processes.

AUTHOR BIOGRAPHIES

Benjamin Melamed is a Fellow at the C&C Research Laboratories, NEC USA Inc., at Princeton, New Jersey. His research interests include system modeling and analysis, simulation, stochastic processes and visual modeling environments. He authored or co-authored over 70 papers. Melamed received a B.Sc. degree in Mathematics and Statistics from Tel Aviv University in 1972, and a M.S. and Ph.D. degrees in Computer Science from the University of Michigan in 1973 and 1976, respectively. From 1977 to 1981 he taught at the department of Industrial Engineering and Management Science at Northwestern University. He joined the Performance Analysis Department at Bell Laboratories in 1981, and became an AT&T Bell Laboratories Fellow in 1988. He joined NEC in 1989 and served there as a Deputy Director and Department Head of the Performance Analysis Department. Melamed was elected IEEE Fellow in 1994.

JON HILL is with Smith Barney, Inc. in New York City. His interests include simulation output analysis, graphical interfaces to modeling tools, statistics, data analysis and graphical techniques for displaying data. After receiving a Ph.D. in Biophysics from the University of Utah in 1980, he joined AT&T Bell Laboratories at Holmdel, NJ, and participated in both the Microsystems Design Laboratory and the Operations Research Department there. In 1992 he joined the Performance Analysis Department at the C&C Research Laboratories, NEC USA, Inc., in Princeton, NJ, where he was the chief developer of the TEStool package until moving to Smith Barney in 1993.