CONF-970241--2

Large Data Series: Modeling the Usual to identify the Unusual

D. J. Downing *
V. V. Fedorov *
W. F. Lawkins *
M. D. Morris *
G. Ostrouchov *

* Mathematical Sciences Section
Oak Ridge National Laboratory
Computer Science and Mathematics Division
P. O. Box 2008, Bldg. 6012
Oak Ridge, TN 37831-6367

Research sponsored by the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory, managed for the United States Department of Energy, Office of Energy Research, under contract number DE-AC05-96OR22464 with Lockheed Martin Energy Research Corporation.

DISCLAIMER

Large Data Series: Modeling the Usual to Identify the Unusual

D. J. Downing, V. Fedorov, W. F. Lawkins, M. D. Morris, and G. Ostrouchov,
Mathematical Sciences Section, Oak Ridge National Laboratory

Abstract

"Standard" approaches such as regression analysis, Fourier analysis, Box-Jenkins' procedure, et al., which handle a data series as a whole, are not useful for very large data sets for at least two reasons. First, even with computer hardware available today, including parallel processors and storage devices, there are no effective means for manipulating and analyzing gigabyte, or larger, data files. Second, in general it can not be assumed that a very large data set is "stable" by the usual measures, like homogeneity, stationarity, and ergodicity, that standard analysis techniques require. Both reasons dictate the necessity to use "local" data analysis methods whereby the data is segmented and ordered, where order leads to a sense of "neighbor," and then analyzed segment by segment. The idea of local data analysis is central to the study reported here.

1 Introduction.

Our focus is on identifying "unusual" segments of data from very long streams of data series. The adjective "unusual" is intended to convey a sense that the fraction of unusual to usual segments over the whole data stream is small. We prefer the expression "data series" to "time series" to emphasize that the data frequently does not admit a natural sense of "future" and "past." Even if the physics of the process being observed does allow a sense of time to be identified with the data series, the analysis techniques we consider are unlike traditional time series methods that focus on "forecasting " or "online" analyses. We compare a segment to a collection of its neighbors or the whole population.

The main idea is based on partitioning the data series into relatively short segments and then model each segment using a relatively simple, low-order model. Segmentation may be either static, or moving. The latter is computationally more demanding but frequently leads to a better visualization of unusual events. The parameters of the model are expected to have typical values and not display significant variation over the collection of usual segments. However, over the collection of unusual segments, the model parameters are expected to vary significantly. If the unusual segments are reflections of a limited number of distinct digressions from the process corresponding to the usual segments, then the relatively large variation in model parameter values over the whole collection of unusual segments may cluster into a few classes of relatively small variation. From a statistical standpoint, we talk about a mixture of populations of different sizes and the detection and segregation of those populations.

We present results for the univariate case. Our methods can, in principle, be generalized to large multivariate data sets that admit a meaningful segmentation.

The data segmentation problem is both critical and application dependent. We assume the scientist, engineer, biologist, or whoever is using these methods has an idea about the nature, including space and time scales, of the events or perturbations of importance for their particular application. Evidently there are applications for which a multistage, or hierarchical segmentation structure is most appropriate, especially in the case of self-similar processes (for example, see [2, 7, 11]). The techniques discussed here can be applied at every segmentation level. Developing techniques for defining hierarchical segmentation structures is an interesting topic for future studies.

Most of the theoretical material included in this article is described in terms of regression models. Hopefully a reader will be able to propagate the idea for other types of local modeling such as autoregressive models, kernel estimation, wavelet expansions, etc. We report on the analysis of two data series: an atmospheric process and a neurophysiological process.

2 Local Regression Models.

Regression analysis is one of the most developed areas in statistics, both from the theoretical and computational standpoints. In spite of their simplicity, regression models are well suited to detecting interesting features in segmented data.

Let $\{y_i\}_{i=1}^N$ be observations made under conditions $\{x_i\}_{i=1}^N$. The essence of the problem is that N is very large and it may be difficult to manipulate or even to store the set $Y = \{y_i\}_{i=1}^N$ on a relatively small computing platform. We consider methods based on partitioning Y into J subsets of equal size, so that

$$Y = \cup_{j=1}^J \mathbf{y}_j, \quad \mathbf{y}_j = \{y_i\}_{i=(j-1)L+1}^{jL}, \tag{1}$$

where $N = L \times J$, and L is the size of the subset $\mathbf{y}_j$.

A natural generalization of the segmentation scheme (1) is to use a sliding window of length L, so that

$$Y = \cup_{j=1}^J \mathbf{y}_j, \quad \mathbf{y}_j = \{y_i\}_{i=j}^{j+L-1}, \tag{2}$$

where $J = N - L + 1$.

From a computational standpoint, for the segmentation (1) we have $J \ll N$ while for the segmentation (2) we have $J \approx N$.

Let

$$x_{j0} \in \mathbf{x}_j = \{x_i\}_{i=(j-1)L+1}^{jL},$$

and suppose that x_{j0} is not close to one of the boundary points of the interval $\mathbf{x}_j$. Also, define

$$u_{ji} = x_i - x_{j_0}.$$

We are going to study the local regression model

$$y_i = \theta_j^T f(u_{ji}) + \gamma_j^T \phi(u_{ji}) + \epsilon_i, \tag{3}$$

where $(j-1)L + 1 \le i \le jL$. In this regression model, the first term $\theta_j^T f(u_{ji})$ describes the "standard" component of the data set $\mathbf{y}_j$, and the second term $\gamma_j^T \phi(u_{ji})$ describes occasional perturbations, like contaminations, outliers, thresholds, etc. Further, the expressions $f(u)$ and $\phi(u)$ are vectors of known functions, $\theta \in \mathbb{R}^p$ and $\gamma \in \mathbb{R}^q$ are unknown parameters, and ϵ_i represents whatever can not be modeled better than by a "random noise" model; i.e., ϵ_i are random variables. Concerning the random noise ϵ_i, we assume

$$E(\epsilon_i) = 0, \quad E(\epsilon_{i_1}, \epsilon_{i_2}) = \sigma_j^2 \delta_{i_1 i_2},$$

where $\delta_{i_1 i_2} = 1$, for $i_1 = i_2$ and is zero otherwise. The parameters θ_j, γ_j, σ_j^2, or some functions of them, form the set of descriptors of an interval. Generally, $p \ll L$ and $q \ll L$, meaning that the vectors θ_j, γ_j are small relative to the data $\mathbf{y}_j$.

Treating each interval $\mathbf{x}_j$ separately permits us to use very standard and simple statistical techniques. In fact, that approach applied simultaneously to all intervals $\mathbf{x}_j$ reveals some interesting properties of large data sets. Suppose there exists a physically meaningful partition $\{\mathbf{x}_j\}_{j=1}^J$. In general, the data for almost all intervals are perturbation free and can be described by the regression model (3) without the second term. Occasionally, a perturbation appears and the perturbation term $\gamma_j^T \phi(u_{ji})$ becomes significant, that is,

$$\max_{x \in \mathbf{x}_j} |\gamma_j^T \phi(x - x_{j_0})| \gg \sigma_j.$$

We are assuming that the size of a perturbation relative to the size of the interval $\mathbf{x}_j$ is small. With appropriate distributional assumptions, we can obtain the least squares estimators $\hat{\theta}_j$ and $\nu^2(\hat{\theta}_j)$, the associated sum of squared errors. These may be used as descriptors of the j-th interval. An unusual value of $\hat{\theta}_j$ and a relatively large value of $\nu^2(\hat{\theta}_j)$ indicates the presence of a perturbation.

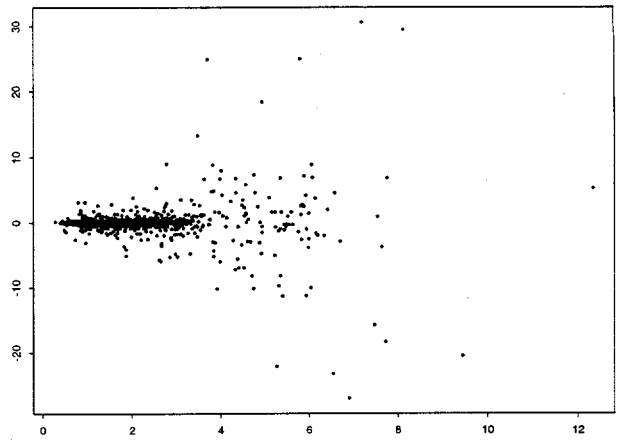

Figure 1: $\hat{\theta}_{j1}$ (horizontal axis) vs. $\hat{\theta}_{j2}$ for liquid water ARM data

Consider a segment of data from measurements of liquid water content of the atmosphere. This data was collected under the auspices of the Atmospheric Radiation Measurement (ARM) project [5]. The segment contains 200,000 observations taken at 20 second intervals, which is about 46 days of measurements. We compute a least squares fit of the simple linear model

$$y_i = \theta_{j1} + \theta_{j2} u_{ji} + \epsilon_i$$

for each segment j, and $(j-1)L + 1 \le i \le jL$.

Fig. 1 shows a plot of $\theta_{j1} \times \theta_{j2}$ for 2,000 segments of size 100. All plots in this report are produced with Splus [9]. Approximately 200 points are outside the thin dark elliptical region, indicating the 200 intervals that potentially contain perturbations. The scatterplot appears to be a mixture of two distributions: one tight distribution containing roughly 1800 intervals without

perturbations and a more diffuse distribution containing roughly 200 intervals with perturbations. Also note that almost all points lie in a cone with its vertex at the origin. There exists a simple statistical explanation of this phenomenon [3].

3 Basis Selection

In the following, we consider some recommendations on the selection of a basis function $f(u)$.

Known Covariance Kernel. For simplicity, let us assume that random errors ϵ_i may be neglected. At the same time we consider $\mathbf{y}_j$ as realizations of some random vector $\mathbf{y}$. In total we have J such realizations. Assume the covariance structure of $\mathbf{y}$ is known. Let

$$\mathbf{K} = E\left[(\mathbf{y} - E(\mathbf{y}))\,(\mathbf{y} - E(\mathbf{y}))^T\right], \tag{4}$$

be the known covariance kernel. Implicitly we also assume there are no perturbations or short-term tends. We introduce the eigenvalues, λ_α, and eigenvectors, ψ_α, of the covariance matrix $\mathbf{K}$, so that

$$\lambda_\alpha \psi_\alpha = \mathbf{K}\psi_\alpha, \quad \lambda_1 \geq \lambda_2 \geq \ldots \lambda_L \geq 0 \tag{5}$$

Define matrix $\mathbf{\Psi} = (\psi_1, \psi_2, \ldots, \psi_L)$ and vector $\theta_j = (\theta_{j1}, \theta_{j2}, \ldots, \theta_{jL})^T$, and note that $\mathbf{\Psi}^T\mathbf{\Psi} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix. Any vector $\mathbf{y}_j$ may be represented in the form

$$\mathbf{y}_j = \mathbf{\Psi}\theta_j, \quad \text{where} \quad \theta_j = \mathbf{\Psi}^T\mathbf{y}_j. \tag{6}$$

It is useful to note that θ_j may be calculated not only by (6), but also as the solution to the least squares problem

$$\theta_j = \arg\min_{\theta_j} \|\mathbf{y}_j - \mathbf{\Psi}\theta_j\|_2, \tag{7}$$

where $\|\cdot\|_2$ is the Euclidean norm.

In fact, if we consider (7) as an optimization problem with θ_j and $\mathbf{\Psi}$ unknown, and require that $\mathbf{\Psi}$ is orthonormal, the solution is exactly a set of eigenvectors of the matrix $\mathbf{K}$. This is analogous to principal components.

Referring to (6), the first $p < L$ values of θ_j are used as descriptors of $\mathbf{y}_j$. If

$$\sum_{\alpha=1}^{p} \theta_{j\alpha}^2 \lambda_\alpha^{-1} > \chi^2_{1-\delta}(p),$$

where $0 < \delta < 1$ and $1 - \delta$ is the corresponding confidence level for the χ^2-distribution with p degrees of freedom, the presence of a perturbation in the j-th interval is quite likely. We can say that $\delta \times N$ is approximately the number of falsely identified usual segments.

A reader familiar with the method of principal components may find that the results described to this point are mathematically identical to results concerning the optimal properties of principal components.

Moving Windows and Unknown Covariance Kernel. We describe a specific model that incorporates several ideas discussed earlier, including sliding, or moving windows and unknown covariance kernels.

Let $\{y_i\}_{i=1}^{N+n}$, where $n \ll N$, denote a given data series. Given a value for n, consider the vector

$$\mathbf{y}_i = (y_i, y_{i+1}, y_{i+2}, \ldots, y_{i+(n-1)})^T, \quad i = 1, 2, \ldots, N. \tag{8}$$

Let the matrix $\mathbf{Y}$ be defined by the columns $\mathbf{y}_i - \bar{y}$. Then, define the matrix

$$\hat{\mathbf{K}} = \frac{1}{N}\mathbf{Y}\mathbf{Y}^T$$

The matrix $\hat{\mathbf{K}}$ is symmetric, nonnegative. Again, we denote the eigenvalue, eigenvector pairs of $\hat{\mathbf{K}}$ by $\{(\lambda_\alpha, \psi_\alpha)\}_{\alpha=1}^{n}$, where

$$\lambda_1 \geq \lambda_2 \geq \ldots \lambda_n. \tag{9}$$

To simplify notation, we omit "hats" in the remainder of this section and we hope that the reader will distinguish these eigenvectors and eigenvalues from those introduced in the previous section. Define the matrix $\mathbf{\Psi}$ by

$$\mathbf{\Psi} = (\psi_1, \psi_2, \ldots, \psi_n),$$

meaning that the j-th column of $\mathbf{\Psi}$ is the eigenvector ψ_j. Next, define the orthonormal transformation

$$\theta = \mathbf{\Psi}(\mathbf{y}) = \mathbf{\Psi}^T\mathbf{y}. \tag{10}$$

For a nonsingular $\mathbf{\Psi}$, the vector θ_i contains all the information about the vector $\mathbf{y}_i$; that is,

$$\mathbf{y}_i = \mathbf{\Psi}\theta_i \tag{11}$$

Again, note that the optimization (7) with $\mathbf{y}_i$ replacing $\mathbf{y}_j$ can be interpreted as parameter fitting for the linear model (11).

For most practical situations, the eigenvalues λ_α decrease rapidly with increasing α. Therefore, the regression model (11) can be replaced by

$$\mathbf{y}_i = \mathbf{\Psi}_m\theta_{im} + \epsilon_i, \tag{12}$$

where m identifies the range of eigenvalues considered to be significant, and $\mathbf{\Psi}_m$ and θ_{im} are the corresponding m components of $\mathbf{\Psi}$ and θ_i. The remainder vector

ϵ_i is equivalent to "noise" in the statistical approach. Referring to comments following (7), it follows that

$$\frac{1}{N}\sum_{i=1}^{N}\min_{\theta_i}\|\mathbf{y}_i - \mathbf{\Psi}_m\theta_{im}\|_2 = \sum_{\alpha=m+1}^{n}\lambda_\alpha.$$

Note that all the preceding formulas can be generalized for the "lagged" case where the definition of the vector $\mathbf{y}_i$ (8) is replaced by

$$\mathbf{y}_i = (y_i, y_{i+k}, y_{i+2k}, \ldots, y_{i+(n-1)k})^T. \qquad (13)$$

4 Nonlinear Dynamical Model.

Here we describe a moving window approach with an unknown covariance kernel from a nonlinear dynamical model perspective. Suppose we have a "training" set of data, $\{y_i\}_{i=1}^N$ that is representative of a background process. Let $\mathbf{y}_i$ be defined by (13) for a given pair (n,k) and suppose we have gone through the procedure of constructing the representation (10). The sequence $\{\mathbf{y}_i\}_{i=1}^N$ describes a trajectory in the $\mathbf{y}$ state space, E_y^n, which is the Euclidean n-space. That is,

$$\{\mathbf{y}_i\}_{i=1}^N \subset E_y^n.$$

The transformation (10)

$$\mathbf{\Psi} : E_y^n \to E_\theta^n,$$

thus produces the trajectory

$$\{\mathbf{\Psi}(\mathbf{y}_i)\}_{i=1}^N = \{\theta_i\}_{i=1}^N \subset E_\theta^n.$$

A good choice for the parameter pair (n,k) produces eigenvalues $\{\lambda_j\}_{j=1}^n$ (9) that decrease rapidly, so that in general the model (11) can be replaced by the model (12), where $m < n$ is the range of significant eigenvalues and ϵ_i is a random noise vector used to model the insignificant eigenvalue components.

The regression model (12) is used to identify perturbations relative to the background process. We are not trying to accurately model the background process but rather separate the unusual from the usual. Define the projection $P_\theta^{m_1,m_2} : E_\theta^n \to E_\theta^{m_2-m_1+1}$ by

$$P_\theta^{m_1,m_2}(\theta) = (\theta_{m_1}, \ldots, \theta_{m_2})^T \in L\{\psi_j\}_{j=m_1}^{m_2},$$

where $L\{\psi_j\}_{j=m_1}^{m_2}$ is the linear subspace of E_θ^n spanned by the eigenvectors $\{\psi_j\}_{j=m_1}^{m_2}$. Further, define

$$\tilde{B} = P_\theta^{m_1,m_2}(\{\theta_i\}_{i=1}^N),$$

the projection into the (m_2-m_1+1)-dimensional subspace $L\{\psi_j\}_{j=m_1}^{m_2}$ of the trajectory $\{\theta_i\}_{i=1}^N$ in E_θ^n constructed from the training set $\{y_i\}_{i=1}^N$. We assume that the background process projects to a relatively small, dense region $\tilde{B}$ for a small value of $m_2 - m_1 + 1$.

The above assumptions concerning the background process versus perturbations implies that the region $\tilde{B}$ is a concentrated region in state space associated with usual data segments and that unusual segments of the data series will produce trajectory segments in $L\{\psi_j\}_{j=m_1}^{m_2}$ that move outside $\tilde{B}$. Let T_s be a characteristic time scale associated with the background process. We define a background event, or a usual segment of the observed data series as any trajectory segment that remains in $\tilde{B}$ for at least a duration T_s. More formally, a segment Γ_j of l_j time steps,

$$\Gamma_j = \{\theta_i\}_{i=i_j}^{i_j+l_j-1},$$

is a usual segment if

$$\begin{aligned} P_\theta^{m_1,m_2}(\theta_{i_j-1}) &\notin \tilde{B}, \\ P_\theta^{m_1,m_2}(\Gamma_j) &\subset \tilde{B}, \\ P_\theta^{m_1,m_2}(\theta_{i_j+l_j}) &\notin \tilde{B}, \\ l_j \times t_s &\geq T_s, \end{aligned}$$

where t_s is the sample time for the data series. In turn, we define the trajectory segment

$$\Delta_j = \{\theta_i\}_{i=i_j+l_j}^{i_{j+1}-1},$$

which separates segments Γ_j and Γ_{j+1}, to be the j-th perturbed segment. The length in time steps of Δ_j is

$$p_j = (i_{j+1} - 1) - (i_j + l_j) + 1 = i_{j+1} - (i_j + l_j).$$

Note that this definition allows a perturbed trajectory segment to pass through the region $\tilde{B}$ so long as the time it takes is less than the time scale T_s. The data series segment corresponding to Δ_j is, by definition, a perturbation, or unusual segment.

5 Report on two Applications.

Let us begin with a data series collected under the auspices of the Atmospheric Radiation Measurement (ARM) project [5]. The data series are measurements of liquid water content of the atmosphere near Oak Ridge, Tennessee, over a period of 257 days beginning in March of 1994. The observations are taken at 20 second intervals but contain gaps, some over a day in duration. Our exercise is based on a subset with 122,786 observations covering a little over 28 days. It is the longest segment without any major gaps. From a physical point of view, the background process is a relatively clear day with dry

conditions. Perturbations include cloud, rain, and fog events as well as some instrument malfunction events. An important feature of the ARM data series is that the perturbations occur on several scales.

The second data series is a subset taken from one channel of a sixteen channel electroencephalogram (EEG) record for an epileptic patient [6]. The complete EEG record is 23 minutes, recorded at the sampling rate 512 *Hz*, and it includes a seizure. A 90 second segment, which occurs well in advance of the recorded seizure, is used for the analyses presented here. EEG records typically include a great deal of "artifact," representing head movement, eye movement, muscle tension, grinding teeth, etc., in addition to unmasked neurophysiological activity. If we associate neurophysiological activity with the background process, then artifact is a perturbation relative to that background process.

The results presented for the regression model in Section 2 use non-overlapping windows of size $L = 50$, which corresponds to the segmentation 1.

The nonlinear dynamical process model in Section 4 uses a moving window paradigm of 2. However, a conclusion is reached about a collection of windows, rather than about an individual segment, as with local regression. The window length used here for both data series is approximately $L = 20$.

Local Regression Models. We begin with the ARM data. It is partitioned into intervals of $L = 50$ observations for a total of 2,455 intervals. We fit a quadratic model to each of the intervals. This produces five values that characterize each interval: the three model coefficients and two measures of lack of fit (l_∞ and l_2 norms of residuals). This gives a data set of 2,455 observations on the five variables.

We can define a perturbation interval in terms of these five variables. Certainly intervals with a poor quadratic fit (high l_∞ and l_2 norm of residuals) can be considered to contain perturbations, but also intervals with unusually steep slope or a strong quadratic coefficient can be considered as parts of perturbations. In the ARM data segment we take the extreme (large in absolute value) 15% in any of the four variables as perturbation intervals. A plot of a section of the data series that contains many perturbations (days 64 and 65) is in Fig 2 with perturbation intervals appearing bold. A more compute intensive but probably better separation of the usual from the unusual can be obtained with multivariate density estimation techniques (see [10]). The 15% cutoff quantile is arbitrary, but it can be estimated from the data. For example, given a sufficiently smooth unimodal density estimate for a given variable, the quantile that determines the extremes as perturbations can be numerically

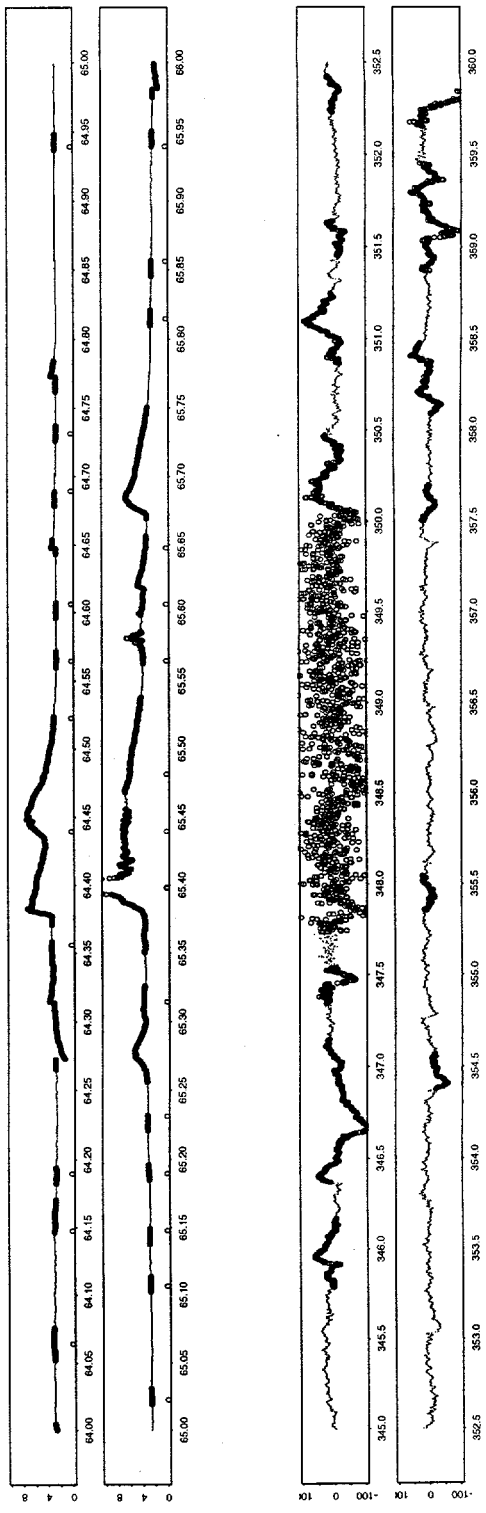

Figure 2: Local regression results for an ARM data series segment (days vs. cm) on the left and an EEG data series segment (seconds vs. μv) on the right. Perturbation intervals are in bold (plotting symbol "o").

estimated as the point at which the smooth density estimate most rapidly "flattens out" into a tail.

The EEG data segment is partitioned into intervals of $L = 100$ observations giving a total of 460 intervals. We fit a local quadratic model to each interval and keep the five values (three model coefficients and two measures of fit) that characterize the interval. This gives a data set of 460 observations on five variables.

Similarly to the ARM data, we take the extreme 20% in any of four of the variables as perturbation intervals. A plot of a section of the data series (seconds 345 through 360) is in Fig 2 showing the intervals with perturbations in bold.

We also used a model based clustering method (see [1]) to cluster the intervals containing perturbations. The clustering results are reported in [3].

A Nonlinear Dynamical Model. Referring to [4] for details, we find that the atmospheric(ARM) and neurophysiological(EEG) data sets can be modeled using the parameter pair values

$$\begin{aligned} \text{ARM} : (n,k) &= (8,3), \\ \text{EEG} : (n,k) &= (9,3). \end{aligned}$$

Further, for both example data series we find that the dominate structure of the background process is included in the subspace of E_θ^n spanned by the first three eigenvectors, $E_\theta^3 = L\{\psi_j\}_{j=1}^3$, so that $m = 3$ in the regression model (12). If there is a long term trend in the background process, we expect that to be reflected primarily in the first coefficient, θ_1, of the regression model. Consequently, we use the coefficients (θ_2, θ_3), which is equivalent to setting $(m_1, m_2) = (2,3)$. Thus $\tilde{B}$ is the projection into the two dimensional subspace $L\{\psi_j\}_{j=2}^3$ of the trajectory $\{\theta_i\}_{i=1}^N$ in E_θ^n constructed from the training set $\{y_i\}_{i=1}^N$. Finally, we find that appropriate time scales for defining usual events are

$$\begin{aligned} \text{ARM} : T_s &= 180 \times t_s = 1hr, \\ \text{EEG} : T_s &= 100 \times t_s \approx .2s. \end{aligned}$$

The analysis results for the ARM and EEG data are displayed in Fig. 3. The same sections of the data are shown as for the local regression results in Fig. 2.

The ARM data has features that vary over a broad range of scales. Thus, the results illustrated in Fig. 3, which shows perturbations coded on the given data series, reveal segments that are clearly unusual to the eye, but also mark other segments that do not appear to be unusual.

Features in the EEG data are not as widely distributed across scales as the ARM data. The right portion of Fig. 3 is the perturbation coded data series. We

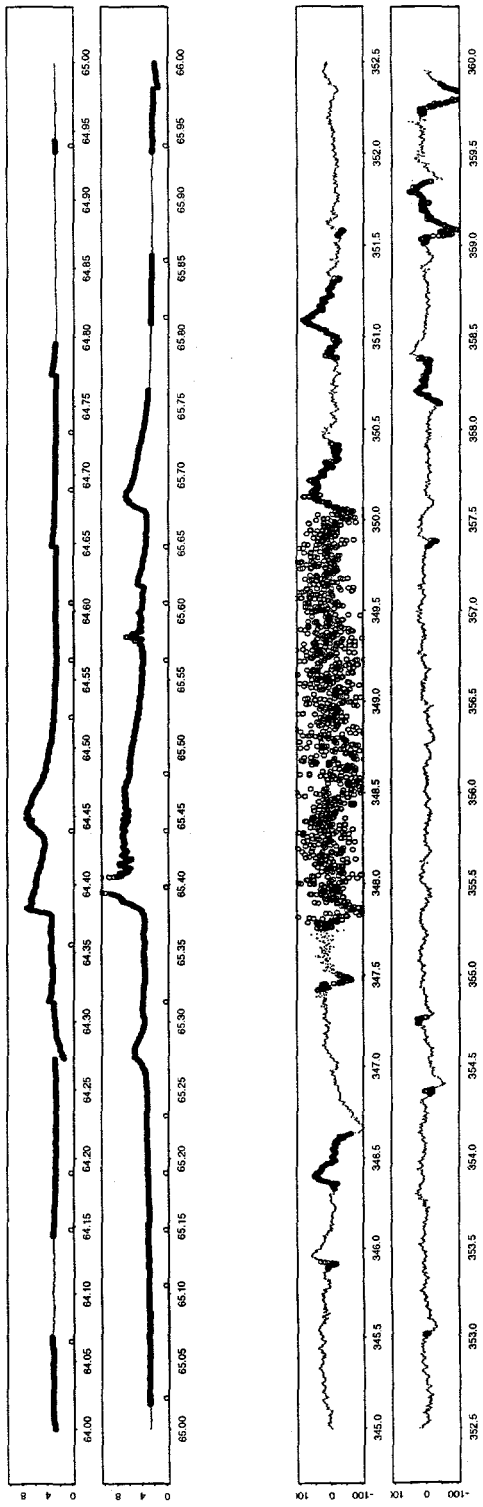

Figure 3: Nonlinear dynamical model results for the same data series as in Fig. 2. ARM data is on the left and EEG data is on the right. Perturbation intervals are in bold (plotting symbol "o").

find that the technique is very efficient at identifying artifact. Further detail of this analysis, is reported in [3].

6 Conclusions.

"Standard" approaches such as regression analysis, Fourier analysis, Box-Jenkins' procedure, et al., which handle a data set as a whole, are not admissible for very large data sets for at least two reasons. First, even with computer hardware available today, including parallel processors and storage devices, there are no effective means for manipulating and analyzing gigabyte, or larger, data files. Second, in general it can not be assumed that a very large data set is "stable" by the usual measures, like homogeneity, stationarity, and ergodicity, that standard analysis techniques require. Both reasons dictate the necessity to use "local" data analysis methods whereby the data is segmented and ordered, where order leads to a sense of "neighbor," and then analyzed segment by segment. The idea of local data analysis is central to the study reported here.

The methods described in this article are universal and may be used with virtually no a priori information about the process represented by the data. Clearly, any independent information about the process that serves to distinguish between the usual and the unusual of interest, such as time scales for example, can, and should be used in a particular application.

The segmented and ordered data structure construct taken together with the local analysis philosophy lends itself directly to parallel computational implementation. The techniques described in this study are extendible to multivariate form.

References

[1] Jeffrey D. Banfield and Adrian E. Raftery. Model-based gaussian and non-gaussian clustering. *Biometrics*, 49:803–821, 1993.

[2] J. Beran. *Statistics for Long-Memory Processes.* Chapman and Hall, New York, New York, 1994.

[3] D. J. Downing, V. Fedorov, W. F. Lawkins, M. D. Morris, and G. Ostrouchov. Large datasets: Segmentation, feature extraction, and compression. Technical Report ORNL/TM-13114, Oak Ridge National Laboratory, 1996.

[4] D. J. Downing, V. V. Fedorov, W. F. Lawkins, M. D. Morris, and G. Ostrouchov. Analysing perturbations and nonstationarity in time series using techniques based on the theory of chaotic nonlinear dynamical systems. Technical Report ORNL/TM-13115, Oak Ridge National Laboratory, Oak Ridge, TN 37831, 1995.

[5] D. O. E. Atmospheric radiation measurement program plan. Technical Report DOE/ER-0441, U. S. Department of Energy, Office of Health and Environmental Research, Atmospheric and Climate Research Division, National Technical Information Service, 5285 Port Royal Road, Springfield, Virginia 22161, 1990.

[6] L. M. Hively, N. E. Clapp, C. S. Daw, W. F. Lawkins, and M. L. Eisenstadt. Nonlinear analysis of eeg for epileptic seizures. Technical Report ORNL/TM-12961, Oak Ridge National Laboratory, Oak Ridge, TN 37831, 1995.

[7] A. N. Kolmogorov and V. A. Uspenskii. Algorithms and randomness. *Theory Probab. Appl.*, 32:389–412, 1988.

[8] C. R. Rao. *Linear Statistical Inference and Its Applications.* John Wiley and Sons, New York, New York, 1973.

[9] Statistical Sciences. *S-PLUS Guide to statistical and mathematical analysis, Version 3.2.* StatSci, a division of MathSoft Inc., Seattle, 1993.

[10] David W. Scott. *Multivariate Density Estimation: theory, practice, and visualization.* John Wiley & Sons, Inc., New York, 1992.

[11] A. Shen. Algorithmic complexity and randomness: Recent developments. *Theory Probab. Appl.*, 37:92–97, 1995.