

# 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 suggest to segment the data and use “local” data analysis on each segment. The results of these local analyses can then be analyzed globally. We use local regression models within segments and show that analysis of optimal local regression models can be operationally identical to nonlinear dynamical modeling.

## 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 need 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 modeling 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 simple 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 [3, 8, 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 Simple Local Models.

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  $J$  segments of equal size, so that

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

where  $N = L \times J$ , and  $L$  is the size of the segment  $\mathbf{y}_j$ . Another segmentation scheme is to use a sliding window of length  $L$ , so that

$$\mathbf{y}_j = \{y_i\}_{i=j}^{j+L-1}, \quad (2)$$

where  $J = N - L + 1$ .

From a computational standpoint, the segmentation (1) is more useful because  $J \ll N$ , whereas the segmentation (2) has  $J \approx N$ .

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 segment  $\mathbf{x}_j$ . Also, define

$$\mathbf{u}_j = \mathbf{x}_j - x_{j0}.$$

We are going to study the local model

$$\mathbf{y}_j = g(\theta_j; \mathbf{u}_j) + \epsilon_j. \quad (3)$$

In this model, the first term  $g(\theta_j; \mathbf{u}_j)$  describes the “usual” component of the segment  $\mathbf{y}_j$  and  $\epsilon_i$  represents whatever remains. The remainder contains the “unusual” as well as a random component. Further,  $g(\theta_j; \mathbf{u}_j)$  is a vector valued function and  $\theta_j \in \mathbb{R}^p$  is a vector of unknown parameters. We make standard assumptions concerning the error term  $\epsilon_i$

$$E(\epsilon_j) = 0, \quad E(\epsilon_j, \epsilon_j^T) = \sigma_j^2 I,$$

where  $I$  is the identity matrix.

From a computational standpoint, the local model should be easy to fit in all segments. The parameters  $\theta_j$ ,  $\sigma_j^2$ , and possibly some other functions of the data  $\mathbf{y}_j$ , form the set of descriptors of a segment  $j$ . Generally,  $p \ll L$  so that  $\mathbf{y}_j$  is *reduced* to the set of its descriptors. A global analysis of the descriptors for all or a collection of segments reveals some interesting properties of the large data series and can identify “unusual” segments.

Although  $g(\theta_j; \mathbf{u}_j)$  can be any simple function, we use a regression model  $g(\theta_j; \mathbf{u}_j) = f(\mathbf{u}_j)\theta_j$ , where  $f(\mathbf{u}_j)$  is an  $(L \times p)$  matrix. Regression analysis is a highly developed area in statistics, from both theoretical and computational standpoints. This allows fast computation and provides some easy theoretical results. If the  $\{x_i\}_{i=1}^N$  are equally spaced, the  $\mathbf{u}_j$  are identical for all  $j$ . The result is a single regression with a matrix right

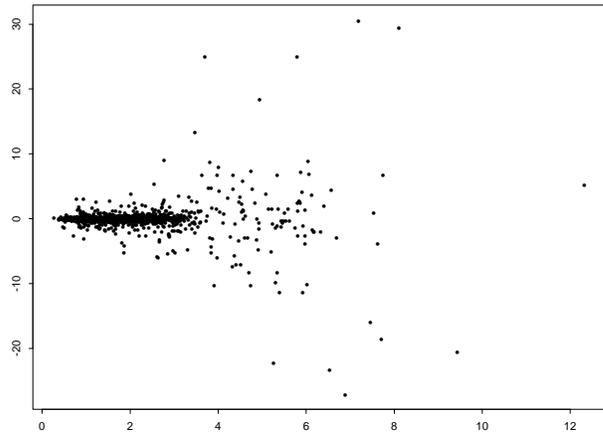


Figure 1:  $\hat{\theta}_{j1}$  (horizontal axis) vs.  $\hat{\theta}_{j2}$  for liquid water ARM data

hand side,  $[\mathbf{y}_1 | \mathbf{y}_2 | \cdots | \mathbf{y}_J]$ , producing a matrix of parameter estimates,  $[\theta_1 | \theta_2 | \cdots | \theta_J]$ . This is the case we have implemented in Splus [9] and use in all our examples.

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 [6]. 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

$$\mathbf{y}_j = \theta_{j1} + \theta_{j2} \mathbf{u}_j + \epsilon_j$$

for each segment  $j$ .

Fig. 1 shows a plot of  $\theta_{j1} \times \theta_{j2}$  for  $J = 2,000$  segments of size  $L = 100$ . Approximately 200 points are outside the thin dark elliptical region, indicating the 200 “unusual” segments that potentially contain perturbations. The scatterplot appears to be a mixture of two distributions: one tight distribution containing roughly 1800 “usual” segments without perturbations and a more diffuse distribution containing roughly 200 “unusual” segments 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 [4].

In spite of their simplicity, regression models are well suited to detecting interesting features in segmented data. In fact, as we show in Section 4, the resulting global analysis based on some optimal local regression models can be viewed as equivalent to nonlinear dynamical models. Because we are not limited to regression models, our analysis can be viewed as a generalization

of nonlinear dynamical models.

Any function of the data vector  $\mathbf{y}_j$  can be used as a descriptor of segment  $j$ , although it should be easy to compute. Two examples are the maximum absolute deviation and the sum of squared deviations. Large deviations clearly indicate the presence of something “unusual” that is not explained by the simple model.

### 3 Optimal Regression Models

Here we consider some recommendations on the selection of the matrix  $f(\mathbf{u}_j)$ .

For simplicity, 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

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

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

$$\lambda_\alpha \psi_\alpha = \mathbf{K} \psi_\alpha, \quad \lambda_1 \geq \lambda_2 \geq \dots \lambda_L \geq 0 \quad (5)$$

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

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

In fact, the optimization problem

$$\mathbf{M}^* = \arg \min_{\mathbf{M}} \|\mathbf{y}_j - \mathbf{M} \gamma\|_2, \quad (7)$$

where  $\mathbf{M}$  is an  $L \times p$  orthogonal matrix and  $\|\cdot\|_2$  is the Euclidean norm, has a solution  $\mathbf{M}^* = \Psi_p$ , the first  $p$  columns of  $\Psi$ . This is completely analogous to principal components.

Usually the  $\lambda_j$  decrease rapidly, so we can use the first  $p < L$  values of  $\theta_j$ , denoted by  $\theta_{pj}$  as descriptors of  $\mathbf{y}_j$ . That is, the regression model

$$\mathbf{y}_j = \Psi_p \theta_{pj} + \epsilon_j$$

is optimal in the sense described above.

If we now allow some of the  $\mathbf{y}_j$  to contain perturbations,

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

for some small  $0 < \delta < 1$  ( $1 - \delta$  is the corresponding confidence level for the  $\chi^2$ -distribution with  $p$  degrees of freedom) indicates the presence of a perturbation in the

$j$ -th segment. We can say that  $\delta \times J$  is approximately the number of falsely identified usual segments.

Next, if the covariance kernel  $\mathbf{K}$  is not known, we can estimate it from a “clean” subset of the data segments. The idea is that we can define the “usual” in terms of a relatively small collection of segments and use the resulting model to identify the “unusual” in the remaining segments. Let the matrix  $\mathbf{Y}$  be defined by a collection of “usual” segments as columns  $\mathbf{y}_j - \bar{\mathbf{y}}$ . Then, estimate  $\mathbf{K}$  by

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

### 4 Nonlinear Dynamical Models

Here we describe a moving window approach with an unknown covariance kernel from a nonlinear dynamical modeling perspective (often referred to as analysis of chaotic data). We use some basic terminology and results from nonlinear dynamical modeling, see for example [1], but we hope that the reader can adapt.

Let  $\mathbf{y}_i$  be a moving window segmentation

$$\mathbf{y}_i = (y_i, y_{i+k}, y_{i+2k}, \dots, y_{i+Lk})^T, \quad (8)$$

for a given pair  $(L, k)$ . Suppose we have a “clean” set of segments that is representative of a background or “usual” process from which we estimate  $\mathbf{K}$  as defined in (3). Next, we construct the representation (6) from the eigenvectors of  $\mathbf{K}$ . The sequence  $\{\mathbf{y}_i\}_{i=1}^N$  describes a trajectory in the  $\mathbf{y}$  state space,  $E_y^L$ , which is the Euclidean  $L$ -space. That is,

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

The transformation (6)

$$\Psi : E_y^L \rightarrow E_\theta^L,$$

thus produces the trajectory

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

A good choice for the parameter pair  $(L, k)$  produces eigenvalues  $\{\lambda_j\}_{j=1}^L$  that decrease rapidly, so that in general  $\Psi$  can be replaced by  $\Psi_p$ , where  $p < L$  is the number of a few selected significant eigenvalues. Normally this is the  $p$  largest eigenvalues, but often we may want to exclude some significant eigenvalues such as one representing the “level” or “trend” in the data series.

We are not trying to accurately model the background process but rather separate the unusual from the usual. Define the projection  $P_\theta^p : E_\theta^L \rightarrow E_\theta^p$  by

$$P_\theta^p(\theta) = (\theta_{m_1}, \dots, \theta_{m_p})^T \in L\{\psi_j\}_{j=m_1}^{m_p},$$

where  $L\{\psi_j\}_{j=m_1}^{m_p}$  is the linear subspace of  $E_\theta^L$  spanned by the eigenvectors  $\{\psi_j\}_{j=m_1}^{m_p}$ , and  $\{m_1, \dots, m_p\}$  index the  $p$  selected eigenvalues. Further, define

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

the projection into the  $p$ -dimensional subspace  $L\{\psi_j\}_{j=m_1}^{m_p}$  of the trajectory  $\{\theta_i\}_{i=1}^N$  in  $E_\theta^L$  constructed from the “clean” set of segments. We assume that the background process projects to a relatively small, dense region  $\tilde{B}$  for a small value of  $p$ .

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  $L\{\psi_j\}_{j=m_1}^{m_p}$  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  $\Gamma_j$  of  $l_j$  time steps,

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

is a usual segment if

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

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  $\Gamma_j$  and  $\Gamma_{j+1}$ , to be the  $j$ -th perturbed segment. The length in time steps of  $\Delta_j$  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  $\Delta_j$  is, by definition, a perturbation, or unusual segment.

Clearly, treating a sequence of local model descriptors from a moving window segmentation as a trajectory, is operationally identical to the nonlinear dynamics process described in this section. The two procedures are derived from a very different theoretical viewpoint but they operate the same. One could argue, that the local modeling approach is a generalization of the nonlinear dynamics approach, which is limited to the models

based on a principal components analysis. In an application area, where good local models are known the local modeling approach offers advantages in interpretation of what is “usual” and what is “unusual.”

## 5 Report on two Applications.

Let us begin with a data series collected under the auspices of the Atmospheric Radiation Measurement (ARM) project [6]. 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 subset without any major gaps. From a physical point of view, the background process (the usual) is a relatively clear day with dry conditions. Perturbations (the unusual) include cloud, rain, and fog events as well as many 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 [7]. 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 (8). 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 segments of  $L = 50$  observations for a total of 2,455 segments. We fit a quadratic model to each of the segments. This produces five values that characterize each segment: the three model coefficients and two measures of lack of fit ( $l_\infty$  and  $l_2$  norms of residuals). This gives a data set of

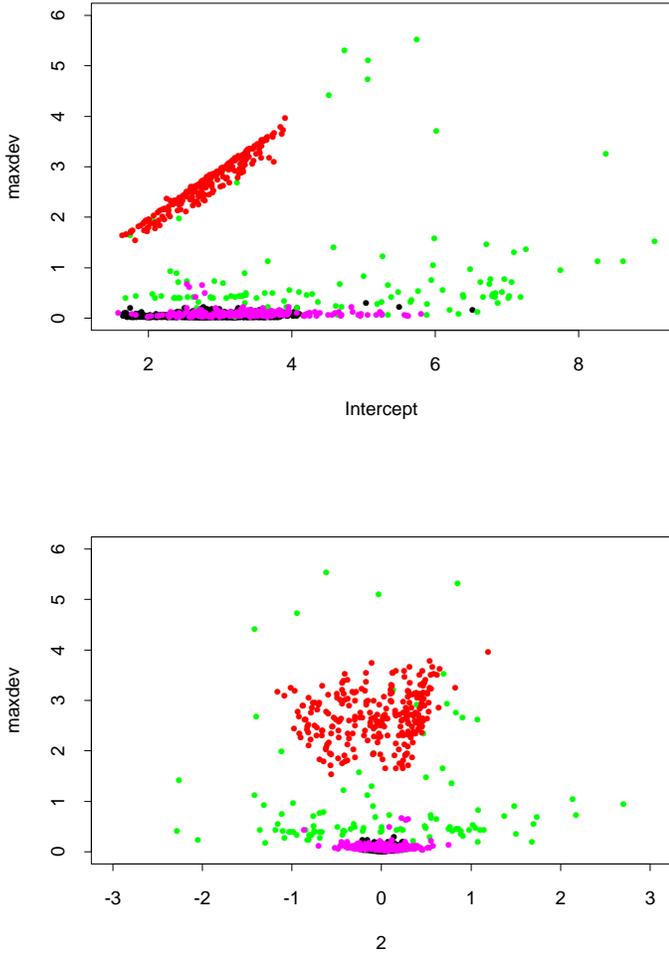


Figure 2: Clustering results for 2,455 segments of ARM data series. Usual segments are black and unusual segments are clustered in color.

2,455 observations on the five variables.

The five variables are used to select unusual (perturbation) segments. Certainly segments with a poor quadratic fit (high  $l_\infty$  and  $l_2$  norm of residuals) can be considered to contain perturbations, but also segments with unusually steep slope or a strong quadratic coefficient can be considered as parts of perturbations. In the ARM data we take the extreme (large in absolute value) 15% in any of the four variables as perturbation segments. After separating the usual from the unusual, the unusual segments are clustered with a model-based clustering algorithm [2] on all five variables. Two projections, intercept versus maximum deviation and quadratic coefficient versus maximum deviation, of the resulting clusters are in Fig 2. The black points (about

76% of data) represent the usual segments and colors indicate the three unusual clusters. The cluster colors are mapped to the segments in Fig 3 of a section of the ARM data series that contains many unusual segments (days 64 and 65). The red unusual segments are instrument anomalies and most of the other unusual segments are related to weather phenomena. 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 a cutoff can be estimated from the data. For example, given a sufficiently smooth unimodal density estimate of a variable, the quantile that determines the extremes as perturbations can be numerically estimated as the point at which the density estimate most rapidly “flattens out” into a tail (a maximum in its second derivative).

The EEG data is partitioned into 460 segments of  $L = 100$  observations. We fit a quadratic model to each segment and keep the five values (three model coefficients and two measures of fit) that characterize it. This provides 460 observations on five variables.

As in the ARM data, we take the extreme 15% in any of four of the variables as perturbation segments. A plot of a section of the data series (seconds 345 through 360) is in Fig 3 showing the usual segments in black and unusual segments in color.

**A Nonlinear Dynamical Model.** Referring to [5] for details, we find that the atmospheric(ARM) and neurophysiological(EEG) data sets can be modeled using the parameter pair values

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

Further, for both example data series we find that the dominate structure of the background process is included in the subspace of  $E_\theta^L$  spanned by the first three eigenvectors,  $E_\theta^3 = L\{\psi_j\}_{j=1}^3$ , so that  $p = 3$ . If there is a long term trend in the background process, we expect that to be reflected primarily in the first coefficient,  $\theta_1$ , of the regression model. Consequently, we use the coefficients  $(\theta_2, \theta_3)$ , 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  $E_\theta^L$  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. 4. The same sections of the data are

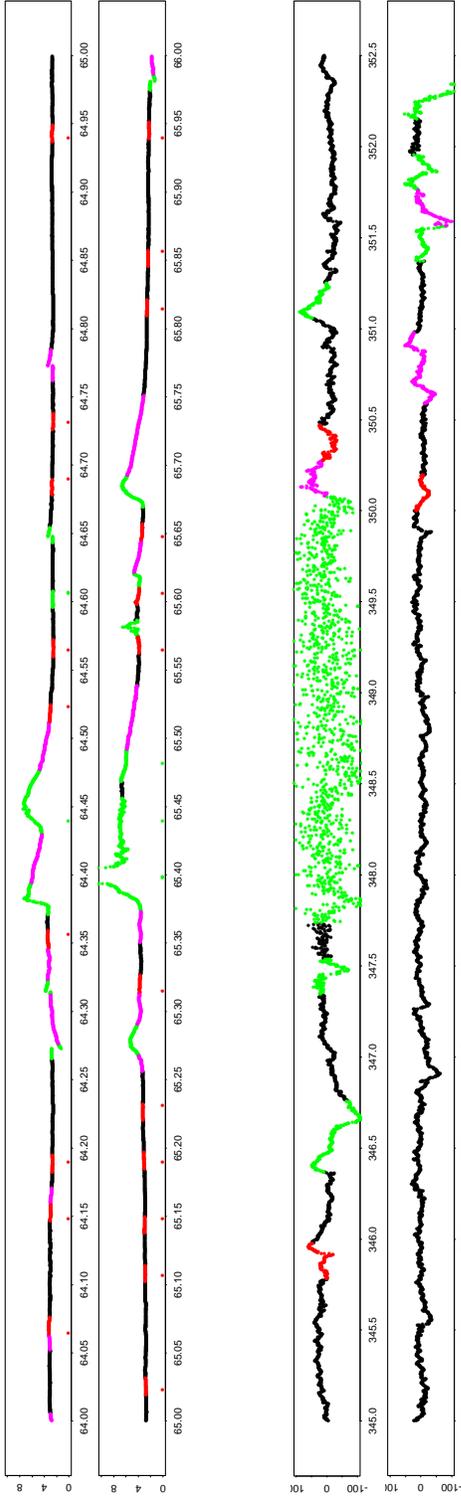


Figure 3: Local regression results for a portion of the ARM data series (days vs. cm) on the left and a portion of the EEG data series (seconds vs.  $\mu\text{v}$ ) on the right. Unusual segments are in color.

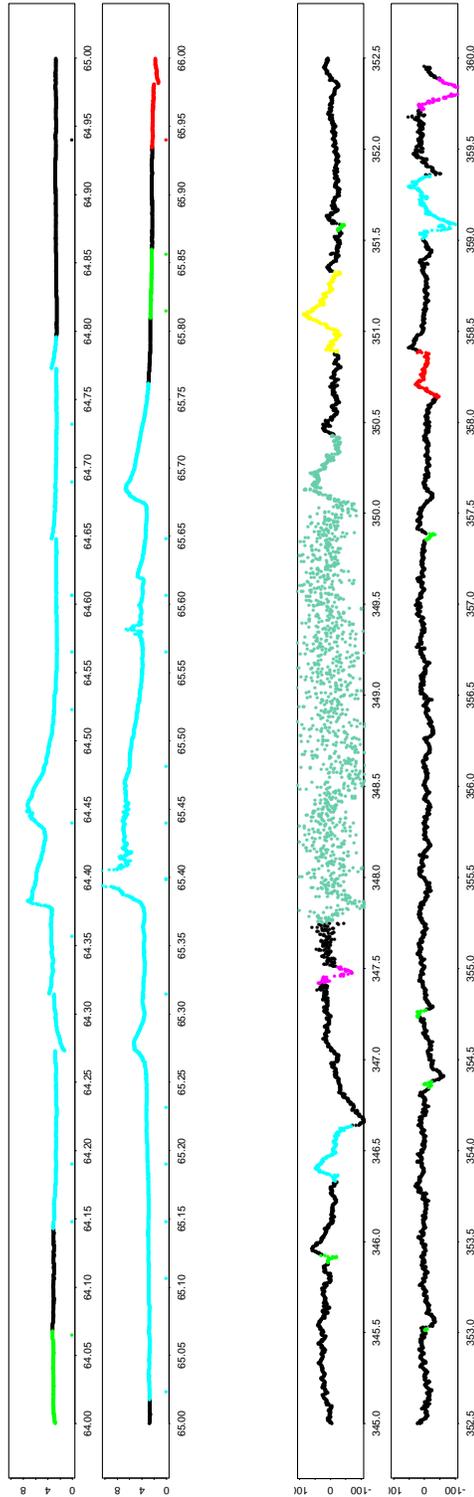


Figure 4: Nonlinear dynamical model results for the same data series as in Fig. 3. ARM data is on the left and EEG data is on the right. Perturbation segments are in color.

shown as for the local regression results in Fig. 3. The results are similar.

The ARM data has features that vary over a broad range of scales. Thus, the results illustrated in Fig. 4, 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. Color coding is according to duration of perturbation event.

Features in the EEG data are not as widely distributed across scales as the ARM data. The right portion of Fig. 4 is the perturbation coded data series. We find that the technique is very efficient at identifying artifact (eye blinks, muscle tension, etc.). Further detail of this analysis, is reported in [4].

## 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.

We show that optimal local regression models can be operationally equivalent to nonlinear dynamical model analysis.

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. Also, the techniques described in this study are extendible to multivariate form.

## References

- [1] H.D.I. Abarbanel, R. Brown, J.J. SiDorowich, and L.S. Tsimring. The analysis of observed chaotic data in physical systems. *Rev. Mod. Phys.*, 65:1331–1392, 1993.
- [2] Jeffrey D. Banfield and Adrian E. Raftery. Model-based gaussian and non-gaussian clustering. *Bio-metrics*, 49:803–821, 1993.
- [3] J. Beran. *Statistics for Long-Memory Processes*. Chapman and Hall, New York, New York, 1994.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] A. N. Kolmogorov and V. A. Uspenskii. Algorithms and randomness. *Theory Probab. Appl.*, 32:389–412, 1988.
- [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.