

## SMOOTHING TIME SERIES FOR INPUT AND OUTPUT ANALYSIS IN SYSTEM SIMULATION EXPERIMENTS

Peter A. W. Lewis

Department of Operations Research  
 Naval Postgraduate School  
 Monterey, California 93940

James G. Stevens

United States Army  
 Fort Leavenworth, Kansas 66027

### ABSTRACT

Classical methods of studying the behaviour of the output of a simulation model as a function of parameters (independent variables, factors, predictor variables) can be divided into global regression and smoothing (local regression). Neither of these methods are adequate, especially when the observations are a function of a time evolution variable and are probably highly correlated. Several new smoothing methods have been proposed recently for this problem, most of them based on the use of splines. This paper concentrates on the use of multivariate adaptive regression spline (MARS) methodology for this smoothing and characterization problem, and on the use of this methodology when there is serial correlation in the data so that lagged values of the observations can be used for predictor variables. The methodology is also useful when analyzing inputs to queues.

### 1. INTRODUCTION

A common situation in simulation studies is to be confronted with the need to model or analyze a sequence of random variables  $y_i$ , where  $i$  is an ordered index, very often time or serial number, and sometimes a shorthand notation for a more general dependence on a sequence of fixed values  $x_i$ . We assume here that the  $y_i$  are continuous random variables although no assumption of normality or symmetry of their distribution is made. For example the  $y_i$  may be successive waiting times in a queue or successive service times or the number of people in a queue at successive fixed intervals. Or the  $y_i$  may be waiting times indexed by the value of the traffic intensity in the simulated queue, say  $x_i$  equals 0.05, 0.10, 0.15, ..., 0.95. Of course dependence on both types of "independent" variables may occur; the time evolution of the waiting times in a queue for different values of appropriately defined traffic intensity is clearly of interest. This will be considered later. For now we consider the case of only one "independent" variable.

Now one is interested in determining the effect of the variable  $i$  on the random variables  $y_i$ , unless the assumption is made that the sequence of random variables is independent and identically distributed, in which case, by definition, there is no effect. This might be a reasonable assumption for the service and inter-arrival processes in some congestion systems, but it is clearly not true for output processes. As a case in point, for a queue which is started empty, the effect of the serial number,  $i$ , on the expected value of the successive waiting times will be an evolution to a steady state or an explosive growth, depending on whether a "traffic intensity" is less than one or not. It is clearly an interesting problem in simulation to be able to characterize this dependence from a finite sequence of observations on the waiting times, and the fact that the behaviour changes from one region to another in the range of the dependent variable (traffic intensity) suggest that nonlinear effects are present and need to be modelled. And by nonlinearity here we mean that the whole form of the dependence changes, usually in a rather abrupt way.

To tackle this problem of the dependence of  $y_i$  on  $x_i$  one needs a model—Sir Maurice Kendall once said that "models are for thinking with"—and the usual modelling assumption is that we can characterize  $y_i$  with an additive noise model,

$$y_i = f(x_i) + \varepsilon_i,$$

where  $\varepsilon_i$  is a random variable with expected value 0, so that the expected value of  $y_i$  is  $f(x_i)$ . Of course the idea of an additive noise component is probably quite erroneous when dealing with positive random variables, but a transformation, say a log transformation, can usually be used to produce an additive noise model, or an approximation to it. We will not go into the use of transformations here, except to say that they need to be used with care, and our preference is to deal directly with the observations  $y_i$  in their original scale. (Transformation of the independent variables is another matter; often the square of an independent variable or its reciprocal may be more meaningful.)

Given the very broad model above for  $y_i$ , then we are left with the problem of determining  $f(x_i)$  from the data. There are, in general, two ways of approaching this problem, although we will mention a third, somewhat derivative method later, namely the use of splines.

The first of these approaches to determining  $f(x_i)$  is (global) regression, in which some parametric functional form is assumed for  $f(x_i)$ , over the whole range of observed  $x_i$ 's (and sometimes beyond this range) and the parameters in the functional form are estimated, for example by least squares. Unfortunately the usual linear additive forms which are assumed for  $f(x_i)$  in regression analysis are seldom adequate in queueing situations, if for no other reason than that very different functional forms would be required in, for example, the queueing case described above when the queue is an equilibrium queue or is an explosive queue. And again in simple queues the mean waiting time is proportional to  $1/(1-\rho)$ , the traffic intensity, so that if the  $x_i$ 's are values of  $\rho$ , the dependence is certainly highly nonlinear. The dependence will not be as simply expressed in more complicated queues, but the simple queue does suggest the problems one might encounter.

The second approach to determining  $f(x_i)$  is smoothing, which one might call local regression. The basic idea is simple. Assume in our model that the errors are independent and have constant variance,  $\sigma^2$ , and further that  $f(x_i)$  can be adequately approximated over the interval  $i - 1$  to  $i + 1$  by a linear (in  $i$ ) function

$$f(x_i) = \alpha(i) + \beta(i)i.$$

Note that the constants are indexed by  $i$ , implying that the linear approximation is specific to the point  $i$  and would have different coefficients elsewhere. The justification for this approximation is Taylor's theorem, and an assumption that  $f(x_i)$  is "smooth," so that we can ignore quadratic and higher terms in the variable  $i$ . Now it is easy to see that if we average the values  $y_{i-1}, y_i, y_{i+1}$ , we get a "smooth," say  $\hat{f}(x_i)$ , which has variance  $\sigma^2 / 3^{1/2}$  and which is unbiased, i.e., the smooth  $\hat{f}(x_i)$  has the same mean as  $y_i$ . To see this note that

$$\begin{aligned} E(\hat{f}(x_i)) &= E\{y_{i-1} + y_i + y_{i+1}\} / 3 \\ &= (\alpha(i) + \beta(i)(i-1) + \alpha(i) + \beta(i) + \alpha(i) + \beta(i)(i+1)) / 3 \\ &= \alpha(i) + \beta(i)i \\ &= f(x_i) \end{aligned}$$

Again, one can look at the effect of the smoothing on the relative variances of  $y_i$  and  $\hat{f}(x_i)$ . By the assumption of constant variance and independence

$$\begin{aligned} \text{Var}(\hat{f}(x_i)) &= \frac{\text{Var}(y_{k-1} + y_i + y_{i+1})}{9} \\ &= \frac{3\sigma^2}{9} \\ &= \frac{\sigma^2}{3} \end{aligned}$$

Thus, we have reduced the variance of  $y_i$  by 1/3 when we go from it to  $\hat{f}(x_i)$  as a representation of the data, or we have reduced the standard deviation by  $1/3^{1/2}$ .

Also if we do this averaging over adjacent points, we have introduced correlation between  $\hat{f}(x_{i-1})$ ,  $\hat{f}(x_i)$  and  $\hat{f}(x_{i+1})$ , which again corresponds to our notion of "smoothness" in a function.

This is easy to see since, for example,  $\hat{f}(x_{i-1})$  and  $\hat{f}(x_i)$  have in common the variables  $y_{i-1}$  and  $y_i$ . From this a formal quantification of the correlation can be made. By contrast, the additive noise for the  $y_i$ 's is assumed to be independent, and thus their sample path is anything but smooth in appearance. This is in contrast to the assumed smoothness of  $f(x_i)$ .

At this point we seem to have obtained the best of all possible worlds with this simple smoothing scheme, so why not average over  $y_i$  for indices  $(i-p), \dots, i, \dots, (i+p)$ , where  $p$  is a positive

integer and get an even better smooth  $\hat{f}(x_i)$  with an even smaller variance? The answer is that as we increase the "bandwidth" ( $p$ ) the linear approximation to  $f(x_i)$  will generally not be valid and we will introduce bias into the smooth. *This conflict between increasing bandwidth and thereby lowering variance but at the same time increasing bias is absolutely fundamental to smoothing.*

Where can we go from here? There are myriads of "better" smoothing schemes, some based on the following idea. The "smooth" described above is equivalent to fitting a linear regression function through the three values  $y_{i-1}, y_i, y_{i+1}$  and taking as the "smooth" the fitted value at  $i$ . Therefore why not fit a quadratic function to, say, five points and use as the "smooth" the new fitted value at the middle point? This is in turn equivalent to doing a moving average over five points with weights which are not all equal to 1/5 but are tapered. This is called a non-uniform smoothing window. The idea can be extended further, and an enormous amount of work has been done on the problem. For a reference see Kendall, Stuart and Ord [1987]

The drawback in the above smoothing schemes is the need to find a correct or optimum bandwidth and window, i.e., one which will minimize, locally or globally, mean square error for the "smooth." This is often done by inspection, in particular by graphing the smooth for several different bandwidths and windows. Of course there is a high degree of subjectivity in the decision to accept one smooth and not the others. Another drawback to smoothing is the fact that the same window and bandwidth may not be adequate at different points  $i$ , since the shape and form and smoothness of  $f(x_i)$  may be entirely different in different parts of the range of  $x_i$ . This has been overcome in modern smoothers, and in particular SUPERSMOOTHER, a program due to Friedman at Stanford University, by using a technique called cross-validation and some very heavy computing. Another drawback with smoothing is that the output is essentially graphics, which is fine for examining a given set of data, but is not of much use when one wants a characterization of  $f(x_i)$  with which to generate data in a simulation. Again, interpolation and prediction is difficult with smoothing.

## 2. SMOOTHING CORRELATED SEQUENCES

The criterion used for determining from the data the degree of smoothing that is necessary is based on the method of cross-validation and the generalized method of cross-validation [Craven and Whaba, 1979]. The idea is to fit at  $i$  with the value  $y_i$  removed, and

to then compare the fitted value,  $\hat{f}(x_i)$  to  $y_i$ . These quantities will be independent if the assumption of independence of the  $\epsilon_i$ 's is correct. Unfortunately this method is quite sensitive to the assumption that the errors  $\epsilon_i$  in the regression models above are independent, or at least uncorrelated. Experience has shown that this methodology, as employed in SUPERSMOOTHER, is apt to lead to serious bias problems in the smooth. These problems have been addressed by several workers; Altman (1989) has shown that standard bandwidth selection techniques such as cross-validation and generalized cross-validation favor undersmoothing when the serial correlations in the  $y_i$ 's are predominantly positive, and oversmoothing when negative. She has also shown how the selection criteria can be adjusted to correct for the effect of correlation. In another paper [Altman, (1990)] she showed how to estimate simultaneously the mean and correlation function in nonparametric regression problems. However, the approach described in the next section seems to be the most promising. Moreover the result is a functional form for  $f(x_i)$  which can be used, for example, to generate data as though it were statistically identical to the  $y_i$  sequence.

## 3. SMOOTHING WITH MARS AND ASTAR

It was suggested above that there is another method for curve estimation which is intermediate to global and local regression, and this is the method of spline smoothing. Essentially low-order polynomials are fitted for non-overlapping pieces of the independent variable—that is between knots—and conditions for smoothness are imposed at the knots. For a review of this methodology, see Silverman [1985]. The knots still have to be chosen and this is almost equivalent to choosing the bandwidth in the smoothing process. These methods are very flexible and have been extended by Friedman [1988] to a technique called multivariate (independent variable) adaptive regression splines (MARS) and the MARS methodology has been extended by Lewis and Stevens [1990] to account for serial correlation in the data. The extension of the MARS methodology to correlated data is obtained by letting one of the regressor variables be a lagged value of the data, and this leads to non-linear autoregressive threshold models for the data. Consequently the method should be applicable to very general modelling of  $y_i$  sequences when the independent variable is time or serial number.

## 4. MULTIVARIATE SITUATIONS

Of course in general one wants to look at the modelling of multivariate situations,

$$y_i = f(\mathbf{x}_i) + \epsilon_i,$$

where the boldface indicates the presence of multivariate independent variables, one of which may be time or serial number. Thus one may have data in which  $y_i$  is dependent on serial number, traffic intensity, number of servers in the queue, etc.. This kind of situation is easily accommodated in the MARS methodology and its extensions. In fact MARS is designed to work best in cases where the dimension of the independent variable  $\mathbf{x}_i$ , is large.

MARS can be conceptualized as a generalization of recursive partitioning [Morgan and Sonquist, 1963] that uses spline fitting in lieu of other simple fitting functions. Given a set of predictor variables, MARS fits a model in the form of an expansion in product spline basis functions of predictors chosen during a forward and backward recursive partitioning strategy. MARS produces continuous models for high dimensional data that can have multiple partitions and predictor variable interactions. Predictor variable contributions and interactions in a MARS model may be analyzed using an ANOVA style decomposition

## 5. AN EXAMPLE

In modelling chemical warfare the outcome—the dependent variable—is the number of deaths which have occurred at fixed time point after the initiation of the attack. Other independent variables which are controlled in the simulation experiment are the number of rounds fired, the wind direction, wind strength, the target radius, temperature, stability of the gas cloud and the breathing rate of the troops. Given the multiple levels of these independent variables which were used in the simulation experiment, there were about 230,000 data points in all for the analysis. If one uses a linear additive model for  $f(x_i)$ , where the first component of  $x_i$  is serial number, then it is easy to show that the output is serially correlated on this variable. If one models the data linearly in, say, the lagged output and the other variables, a first question which arises is whether or not to put in a constant term. Of course at time zero, or with zero rounds fired, the output should be zero, which argues against a constant term. Unfortunately if the wind direction is not “zero,” and the wind strength is not zero, then there will still be nonzero output at time zero if the regression coefficients of these independent variables are not zero. This argues for interaction terms in the equation for  $f(x_i)$ . Further analysis shows that nonlinear dependence on some of the independent variables is needed for proper modelling of the output, and this is just the kind of situation which can easily be handled with MARS.

A discussion of the application of the MARS methodology to this problem will be given.

## REFERENCES

- Altman, N.S. (1989), “Kernel Smoothing of Data with Correlated Errors,” Submitted to *Journal of the American Statistical Association*.
- Altman, N.S. (1990), *Simultaneous Estimation of the Mean and Correlation Functions in Nonparametric Regression Problems with AR(1) Errors*, submitted for publication.
- Craven, P. and G. Whaba (1979), “Smoothing Noisy Data with Spline Functions. Estimating the Correct Degree of Smoothing by the Method of Generalized Cross-Validation,” *Numerische Mathematik* 31, 317–403.
- Friedman, J.H. (1988), *Multivariate Adaptive Regression Splines*, Submitted for publication.
- Kendall, M.G., Stuart, A. and J. Keith Ord (1987), *The Advanced Theory of Statistics: Vol. 3, Design and Analysis and Time Series*. Oxford University Press: Oxford, England.
- Lewis, P.A.W., and J.G. Stevens (1990), “Nonlinear Modelling of Time Series using Multivariate Adaptive Regression Splines (MARS),” Submitted to the *Journal of the American Statistical Association*.
- Morgan, J.N. and J.A. Sonquist (1963), “Problems in the Analysis of Survey Data, and a Proposal,” *Journal of the American Statistical Association* 58, 415-434.
- Silverman, B.W. (1985), “Some Aspects of the Spline Smoothing Approach to Non-Parametric Regression Curve Fitting,” *Journal of the Royal Statistical Society, Series B*, 47, 1, 1–52.