

Deconvolution of Episodic Hormone Data.

By

Finbarr O'Sullivan¹

**Department of Statistics
University of California
Berkeley, CA 94720**

and

Janet O'Sullivan

**Group in Biostatistics
University of California
Berkeley, CA 94720**

¹Research supported by
National Science Foundation under Grant No. MCS-840-3239.

**Technical Report No. 77
September 1986**

**Department of Statistics
University of California
Berkeley, California**

Deconvolution of Episodic Hormone Data.

*Finbarr O'Sullivan*¹

Department of Statistics
University of California
Berkeley, CA 94720.

Janet O'Sullivan

Group in Biostatistics
University of California
Berkeley, CA 94720.

ABSTRACT

A new approach to the analysis of episodic hormone data is described. This involves a stochastic model in which measured blood hormone concentration is represented as a convolution of individual pulses each of which is thought of as the response to a burst of neural activity. Individual pulses are not constrained to occur in a fixed regular pattern. Statistical image processing ideas are used to develop an appropriate deconvolution criterion and a repeated local adjustment deconvolution algorithm is presented. The methodology splits an observed data time series into a spike train of pulse peak times together with a set of pulse shape parameters. This decomposition motivates fresh approaches to the analysis of hormone data. A set of luteinizing hormone data on ovariectomized cows is analyzed.

Key words and phrases. endocrine data, deconvolution, local adjustment algorithm, degree of regularization, analysis of variance.

September 12, 1986

¹ Research supported by the National Science Foundation under Grant No. MCS-840-3239. Paper presented at the 13'th International Biometrics Conference, Seattle, 1986.

Deconvolution of Episodic Hormone Data.

*Finbarr O'Sullivan*¹

Department of Statistics
University of California
Berkeley. CA 94720.

Janet O'Sullivan

Group in Biostatistics
University of California
Berkeley. CA 94720.

1. Introduction

Typical experimental designs used in human and animal endocrinology involve taking repeated measurements of blood hormones concentrations over time. The resulting time series display a characteristic pattern of episodic pulses, see, for example, Evans et al.[7] and Rahe et al.[18]. There has been a great deal of interest in developing tools which can be used to obtain physiologically meaningful summaries from such data (Brinkley[3], Yates[22]). A number of procedures have been proposed, the most popular of which are the threshold based methods of Santen and Bardin[19], Merriam and Wachter[15], and Clifton and Steiner[4]. Spectral times series methods have also been applied (Rahe et al.[18] and McLeod and Craigon[14]) but typical deviations from periodicity severely limits the applicability of these methods (Clifton and Steiner[4]).

¹ Research supported by the National Science Foundation under Grant No. MCS-840-3239. Paper presented at the 13'th International Biometrics Conference, Seattle, 1986.

In threshold methods, the basic idea is to apply batteries of statistical tests to groups of consecutive data points in the time series. If the test statistics exceed some specified thresholds then the group of data points is deemed to constitute a *pulse*. An important feature of threshold methods is that there are no explicit modeling assumptions made about the data. The threshold criteria must be carefully calibrated to ensure that the number of false positives and negatives are maintained at *acceptable levels*. Apart from the work of Clifton and Steiner[4] and Ellis and Desjardins[6], there has been no attempt to systematically evaluate when threshold methods provide statistically reliable summaries and when they do not. Merriam and Wachter[15] point out that such studies would be greatly facilitated if there were a plausible collection of stochastic models for blood hormone data.

One purpose of this communication is to propose a possible model for episodic hormone release. The model specifies that hormone level in the blood is made up of a sum of elemental pulses which may occur with very flexible patterns in time. The distribution of interpulse times governs regularity in the series. Measured concentration of hormone involves a component of measurement error which is also incorporated into the model. As there are no constraints on the number or size of individual pulses it would be inappropriate to base an estimation criterion on a residual sum of squares. Statistical image processing ideas are used to develop an appropriate estimation criterion and an algorithm is provided to compute the best model. Similar kinds of algorithms have been developed by Kornylo and Mendel[11, 12] for seismic deconvolution problems. A set of data on ovariectomized cows is analyzed.

Acknowledgements

We are grateful to D. R. Brillinger, E.R. Hauser, M.M. Hinshelwood, L. Kamwanja and J. Roser for valuable comments during the course of this work.

2. Methodology

2.1. A Model for Episodic Hormone Data

The basic building block in the model is an individual pulse. Think of a pulse as the response to a burst of neural activity in the hypothalamus. We assume that pulses have a characteristic (double exponential) pattern of release described by three parameters: an amplitude α , a rate of infusion β_1 , and a rate of diffusion β_2 , see Figure 2.1. The set of pulse parameters is denoted $\theta = (\alpha, \beta_1, \beta_2)$. The time of occurrence of the pulse is identified with a pulse peak time x and the hormone concentration in the blood at time t due to the pulse is $p_\theta(t-x)$, where $p_\theta(t-x)$ is given by

$$p_\theta(t-x) = \begin{cases} \alpha e^{-|t-x|/\beta_1} & , \quad t \leq x \\ \alpha e^{-|t-x|/\beta_2} & , \quad t > x \end{cases} . \quad (2.1)$$

A graph of a pulse is given in Figure 2.1.

The total amount of hormone in the blood at any time is assumed to be a sum of contributions from several individual pulses. If there were a total number of K pulses occurring at times $x_1, x_2, \dots, x_K$; the level of hormone in the blood at time t would then be given by the sum of the contributions from the individual pulses

$$p_\theta(t-x_1) + p_\theta(t-x_2) + \dots + p_\theta(t-x_K) . \quad (2.2)$$

The measured hormone concentration at time t , denoted $y(t)$, is subject to measurement noise so our model specifies that

$$y(t) = p_\theta(t-x_1) + p_\theta(t-x_2) + \dots + p_\theta(t-x_K) + \varepsilon(t) \quad (2.3)$$

where $\varepsilon(t)$ represents the measurement error or noise at time t .

Some Examples

A range of series generated by the model are displayed in Figure 2.2. In each case, the noise is Gaussian with standard deviation chosen to be about 10% of the amplitude of an individual pulse. If the interpulse times are large by comparison to the sampling frequency then the series tends to have a very clear pattern of pulses. As the interpulse times become smaller the pattern of individual pulses becomes less clear. Finally if a series of pulses occur in rapid succession then the series simulates a surge in the hormone level. The data presented by Rahe et al.[18] demonstrates how all three of these conditions can arise in practice.

2.2. A Deconvolution (Estimation) Criterion

From a signal processing point of view the measurement in (2.3) represents a noisy convolution of pulses. Model estimation will identify the number and location of the pulse peak times as well as the individual pulse parameters, this results in a deconvolution of the observed signal. Since it is possible to obtain a perfect fit to any data set by reducing the size of the individual pulse and allowing a very large number of pulse peak times, it would be inappropriate to use the residual sum of squares as a deconvolution criterion. A more satisfactory criterion is based on statistical image processing ideas, see, for example, Titterton[20,21]. Here the residual sum of squares (RSS) criterion is replaced by a modified criterion of the form

$$RSS + \Phi(\mathbf{x}) \quad (2.4)$$

where Φ is a functional which measures the physical implausibility of the vector $\mathbf{x} = (x_1, x_2, \dots, x_K)$ of pulse peak times. For example, if apriori one assumes that the number of pulses has a Poisson distribution with mean μ then one might take

$$\Phi(\mathbf{x}) = -\log(p_K) = -\log\left\{\frac{e^{-\mu}\mu^K}{K!}\right\} \quad (2.5)$$

where p_K is the apriori probability of observing K pulses. With this the deconvolution criterion becomes

$$RSS - K \log \mu + \log(K!) \quad , \quad \mu > 0. \quad (2.6)$$

A similar kind of criterion arises in the classification and regression tree methodology of Breiman et al.[2]. There K corresponds to the number of nodes in a classification or regression tree. More sophisticated specifications for Φ might be motivated by communication theory ideas discussed in Yates[22].

The parameter μ is termed a regularization parameter; increasing μ will lead to greater numbers of estimated pulse peak times, decreasing μ will begin to restrict the number of pulse peak times. Various techniques for choosing μ have been proposed, see Titterton[20,21] for a discussion. In principle cross-validation might be used but the computational requirements tend to rule it out. When the standard deviation of the noise is known then a popular method of choosing μ is so that the residual standard deviation matches the theoretical value. The radioimmunoassays used to estimate blood hormone concentrations, see Gorbman et al.[9], typically lead to a coefficient of variation (cv) being reported for the data. Once the model has been fitted for a particular value of the regularization parameter the coefficient of variation can be estimated by the root mean square of the residuals, $\hat{r}(t)$, divided by predicted values $\hat{y}(t)$

$$\hat{cv} = \sqrt{\frac{1}{n} \sum_{i=1}^n [\hat{r}(t_i)/\hat{y}(t_i)]^2} \quad . \quad (2.7)$$

In our work we adjust μ so that $\hat{cv}$ matches the value reported in the assay and this seems to work quite well.

2.3. Repeated Local Adjustmet (RLA) Algorithm

A two stage iterative algorithm for fitting the model will now be described. In the first stage (Adjustment of Pulse Peak Times), the individual pulse parameters are fixed and the pulse peak times are updated. During the second stage (Adjustment of Pulse Parameters), the pulse peak times are fixed and the individual pulse parameters are adjusted. Each iteration is guaranteed to reduce (rather not increase) the deconvolution criterion, the procedure is terminated once stops decreasing significantly from one iteration to the next. A schematic for the RLA algorithm is as follows:

RLA Algorithm

Initialize the model parameters

$$\theta = \theta_0 \quad ; \quad x^{(0)} = (x_1^{(0)}, x_2^{(0)}, \dots, x_K^{(0)})$$

Repeat Until Convergence

- {
- (i). Adjustment of Pulse Peak Times
- (ii). Adjustment of Pulse Parameters
- }

End.

A fuller description of the two stages in the RLA algorithm follows.

(i) Adjustment of Pulse Peak Times

During this step the current information about the pulse parameters is used to update the

number and location of the pulse peak times. The updating procedure is based on a local adjustment strategy. At a typical step in the updating, consecutive pulse peak times x_{i-1} , x_i , and x_{i+1} are considered. Three possibilities are entertained: (a) the inclusion of a new pulse between x_{i-1} and x_i , (b) relocation of x_i to alternative position between x_{i-1} and x_{i+1} , and (c) removal of x_i . The condition resulting in the smallest value for the model fitting criterion is retained. Repeated passes are made through the set of pulse peak times until the local adjustment process fails to improve the fit.

(ii) Adjustment of Pulse Parameters

With the pulse peak times fixed, the adjustment of the pulse parameters is carried out by standard non-linear least squares regression techniques. A damped Gauss-Newton algorithm as described in Ortega and Rheinbold[17] is used.

Comments

The deconvolution algorithm is similar in philosophy to the SMLR (Single Most Likely Replacement) algorithm of Kornylo and Mendel[11,12] and the ICM (Iterated Conditional Modes) algorithm of Besag[1]. SMLR has been successfully used to estimate Bernoulli-Gaussian processes in seismic deconvolution problems (see also Cox and Ehrenberg[5]). ICM has been applied to various image restoration problems. Since both the number and location of the pulse peak times is being estimated the RLA algorithm is computationally intensive. No care is taken to ensure that the algorithm does not get caught in local minima and it is conceivable that things could be improved by introducing a stochastic relaxation strategy into the iterations (see Besag[1], Geman and Geman[8] and Lundy[13]). However experimentation with the algorithm (O'Sullivan[16]) has not convinced us of the immediate need for this sophistication.

3. Analysis of Luteinizing Hormone Data on Ovariectomized Cows

The collection of reproductive hormones constitute a complicated dynamical system with sophisticated feedback control mechanisms operating between the ovaries, the pituitary and the hypothalamus. Some references are Brinkley[3], Gorbman et al.[9], and Yates[22]. An important reproductive hormone is luteinizing hormone (LH) whose primary roles in many species are the formation and maintenance of the corpus luteum. LH pulsing patterns in sexually mature cows have been studied by Rahe et al.[18]. In a study of the role of season on the regulation of the onset of puberty in the bovine female, Kamwanja[10] gathered data on 46 ovariectomized cows classified by age of ovariectomy (3, 5, or 7 months) and date of birth (March or September). Ovariectomy eliminates feedback from the ovaries to the hypothalamus and the pituitary. The relationship between the LH pulsing pattern and date of birth and age of ovariectomy is studied here. For each animal blood samples were taken every 30 minutes over a 36 hour time period one month after the animals had been ovariectomized. The animals were split into a control group and a treatment group with the treatment group receiving an injection of estradiol-17 β 2.5 hours after the beginning of sampling. Only data on 22 controls are analyzed here. Blood concentrations of LH were determined by standard radioimmunoassays, see Gorbman et al.[9]. A full set of data consists of 73 blood samples (records on some animals are missing between 3 and 6 samples).

The data is plotted on the same scale in Figure 3.1. There appears to be an effect due to date of birth with March-born animals having higher levels of LH than the September-born group. There are serious departures from normality but for what its worth a repeated measures analysis of variance reveals gives that the effect of birth is statistically significant.

A range of fits of the model to data on a March born cow ovariectomized at 7 months are given in Figure 3.2. As the regularization parameter (μ) is decreased the number of pulses decreases and the individual pulses start to become bigger. The estimated coefficient of varia-

tion is plotted against $\log(\mu)$ in Figure 3.3(i). The coefficient of variation in the assay was about .15. The regularization parameter is chosen so that the estimated coefficient of variation is also .15. The estimated individual pulse and the spike train of pulse peak times are given in Figure 3.3(ii). A similar fitting strategy was used for the other animals. The individual pulses are given in Figure 3.4 Estimated spike trains are given in Figure 3.5. March born animals seem to have larger pulses and possibly more frequent pulsing than the September animals. A summary of pulse areas (total hormone output per pulse) and pulse frequencies (number of pulses per hour) are given in Table 3.1. A two-way analysis of variance was performed on pulse area; the date of birth has statistically significant ($p = .02$) effect on pulse area but date of ovariectomy does not ($p = .53$). There was no statistically significant interaction between date of birth and date of ovariectomy. An similar analysis of pulse frequency revealed no strong effects for date of birth or ovariectomy, the p -value for the date of birth effect was .12.

Table 3.1: Mean pulse areas (total hormone output per pulse) and frequencies (number of pulses per hour). The standard deviation of the means are given in parentheses. Columns correspond to dates of birth (March, September); rows to dates of ovariectomy (3, 5, 7 months).

Pulse Area			Pulse Frequency		
	Mar	Sept		Mar	Sept
3	2.18(.20)	1.56(.47)		1.00(.06)	0.74(.10)
5	2.61(.35)	1.86(.18)		0.86(.07)	0.71(.16)
7	2.49(.41)	1.80(.28)		0.89(.05)	0.91(.10)

4. Discussion

The deconvolution of episodic hormone data has close links to a range of problems arising in reflection seismology, radar, and sonar, see Kormylo and Mendel[11,12] and Cox and Ehrenberg[5]. There is much to be gained by thinking about the common structure of these problems. It is interesting that the early approaches to deconvolution in reflection seismology were based on threshold ideas similar to those of Santen and Bardin[19].

The model proposed for hormone data might be sophisticated to allow the individual pulse amplitudes to vary smoothly in time. This would be useful for modeling the effects of various treatments. It could also be used to study changes in hormone pulsing patterns during the estrous cycle. Rahe et al.[18] have presented some interesting data associated with this phenomenon.

An appropriate methodology for determining the statistical variability of the quantities recovered by the algorithm must be developed. Monte Carlo simulation can provide some answers but it may be possible to develop alternative resampling methods. A set of similar problems arise in connection with image reconstruction, see Besag[1].

References

1. Besag, J., "On the statistical analysis of dirty pictures (with discussion)," *J. R. Statist. Soc.*, vol. B 48, p. to appear, 1986.
2. Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J., *Classification and Regression Trees*, Wadsworth, Belmont, California, 1984.
3. Brinkley, H.J., "Endocrine Signaling and Female Reproduction," *Biol. of Reprod.*, vol. 24, pp. 22-43, 1981.
4. Clifton, D.K. and Steiner, R.A., "Cycle Detection: A Technique for Estimating the Frequency and Amplitude of Episodic Fluctuations in Blood Hormone and Substrate Concentrations," *Endocrinology*, vol. 112, no. 3, p. 1057, 1983.
5. Cox, D.D. and Ehrenberg, J.E., *A fast algorithm for estimation of Bernoulli-Gaussian processes*, Department of Statistics, University of Wisconsin - Madison., 1984.
6. Ellis, G.B and Desjardins, C., "Mapping Episodic Fluctuations in Plasma LH in Orchidectomized Rats," *Am. J. Physiol.*, vol. 247, p. E130, 1984.
7. Evans, J.W., Hughes, J.P., Neely, D.P., Stabenfeldt, G.H., and Winget, C.M., "Episodic LH Secretion Patterns in the Mare During the Oestrous Cycle," *J. Reprod. Fert. Suppl.*, vol. 27, p. 143, 1979.
8. Geman, S. and Geman, D., "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell*, vol. PAMI-6, pp. 721-741, 1984.
9. Gorbman, A., Dickhoff, W.W., Vigna, S.R., Clark, C.L., and Ralph,, *Comparative Endocrinology*, John Wiley & Sons, 1983.
10. Kamwanja, L.A., "Regulation of Onset of Puberty in the Bovine Female: the Role of Season," Ph.D. Dissertation, Department of Meat and Animal Science, University of

Wisconsin, Madison, 1984.

11. Kormylo, J.J. and Mendel, J.M., "Maximum likelihood detection and estimation of Bernoulli-Gaussian processes," *IEEE Trans. Inform. Theory*, vol. 28, pp. 483-488, 1982.
12. Kormylo, J.J. and Mendel, J.M., "Maximum-likelihood seismic deconvolution," *IEEE Trans. Geosci. Remote Sensing*, vol. 21, pp. 72-82, 1983.
13. Lundy, M., "Applications of the annealing algorithm to combinatorial problems in statistics," *Biometrika*, vol. 72, pp. 191-198, 1985.
14. McLeod, B.J. and Craigon, J., "Time series analysis of plasma LH and FSH concentrations as a method of assessing episodic secretion," *J. Reprod. Fert.*, vol. 74, p. 575, 1985.
15. Merriam, G.R. and Wachter, K.W., "Algorithms for the Study of Episodic Hormone Secretions," *Am. J. Physiol.*, vol. 243, 1982.
16. O'Sullivan, J., *Statistical Modelling of Episodic Hormone Release*, Unpublished, 1986. M.S. Thesis, Group in Biostatistics, UC-Berkeley
17. Ortega, J. M. and Rheinbold, W. C., *Iterative Solutions of Non-linear Equations in Several Variables*, Academic Press, 1970.
18. Rahe, C.H., Owens, R.E., Fleeger, J.L., Newton, H.J., and Harms, P.G., "Pattern of Plasma Luteinizing Hormone in the Cyclic Cow: Dependence upon the Period of the Cycle," *Endocrinology*, vol. 107, p. 498, 1980.
19. Santen, R.J. and Bardin, C.W., "Episodic Luteinizing Hormone Secretion in Man Pulse Analysis, Clinical Interpretation, Physiological Mechanisms," *J. Clin. Invest.*, vol. 52, p. 2617, 1973.
20. Titterington, D.M., "General structures of regularization procedures in image reconstruction," *Astron. Astrophys.*, vol. 144, pp. 381-387, 1985.

21. Titterington, D.M., "Common structure of smoothing techniques in statistics.," *Inter. Statist. Rev.*, vol. 53, pp. 141-170, 1985.
22. Yates, F.E., "Analysis of Endocrine Signals: The Engineering and Physics of Biochemical Communication Systems," *Biol. of Reprod.*, vol. 24, p. 73, 1981.

Figure Legends

Figure 2.1 : An individual pulse

Figure 2.2 : Examples of data generated by the model (2.3). The pulsing frequency is low in the first panel, somewhat higher in the second and the final panel shows the result of when several pulses occur very close to each other.

Figure 3.1 : LH data for control animals (March, September). All plots are on the same scale.

Figures 3.2 and 3.3 relate to the March-born cow marked by "+".

Figure 3.2 : Fits of the model corresponding to three values for the regularization parameter. These resulted in 51, 26 and 8 pulses being estimated. The solid line is the model prediction, the actual data points are marked by "*".

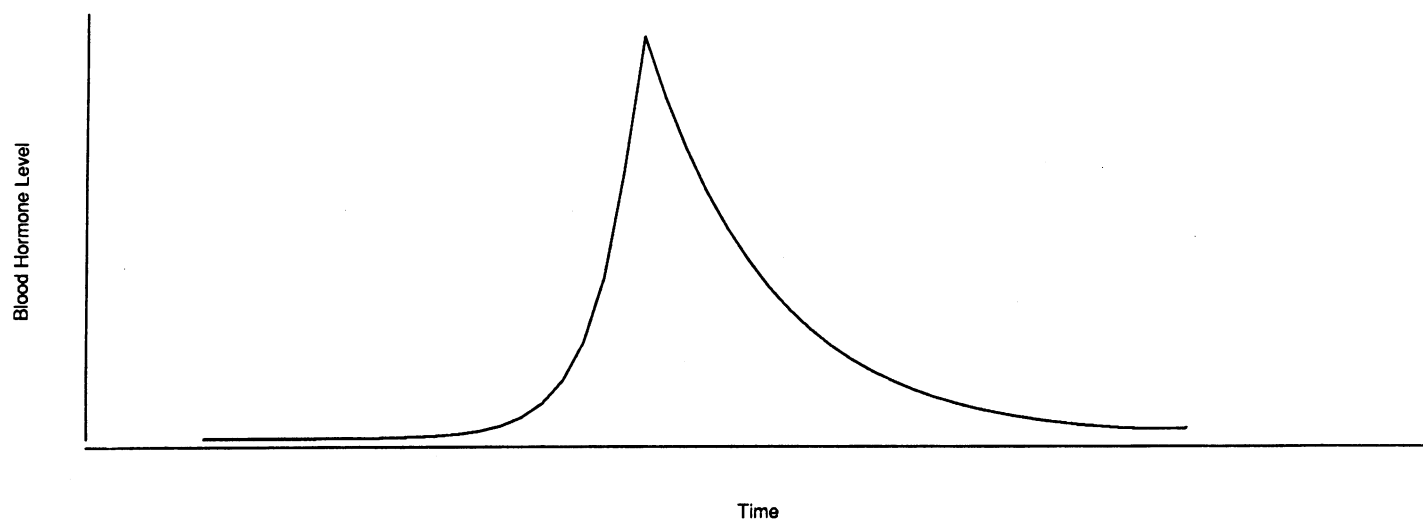
Figure 3.3(i) : Estimated coefficient of variation (cv) plotted against the logarithm of the regularization parameter.

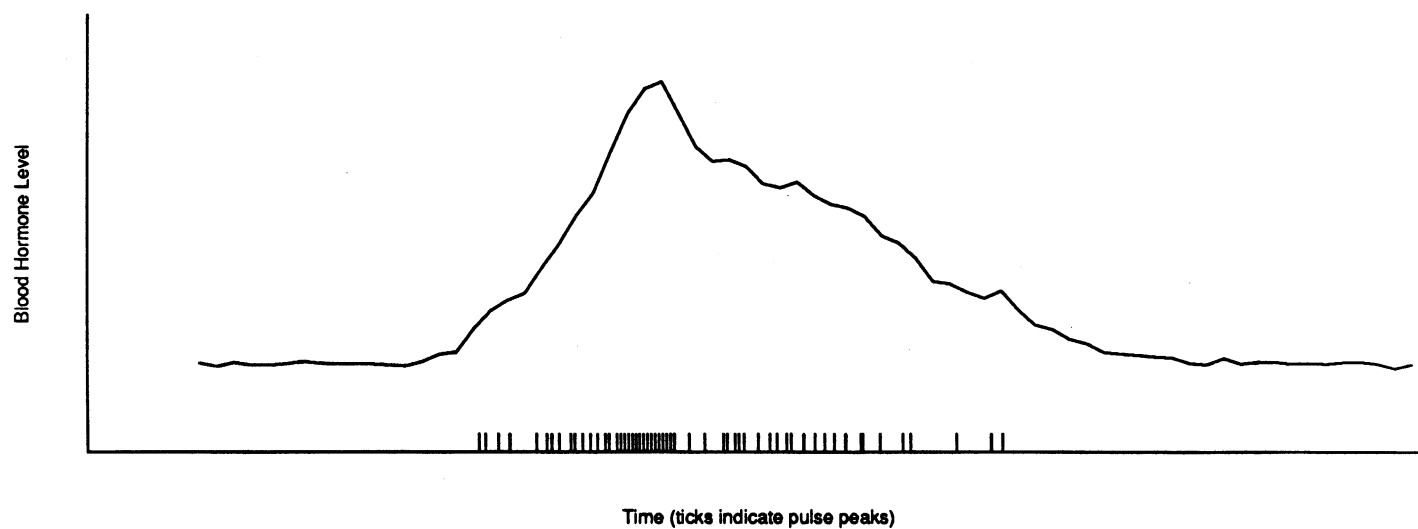
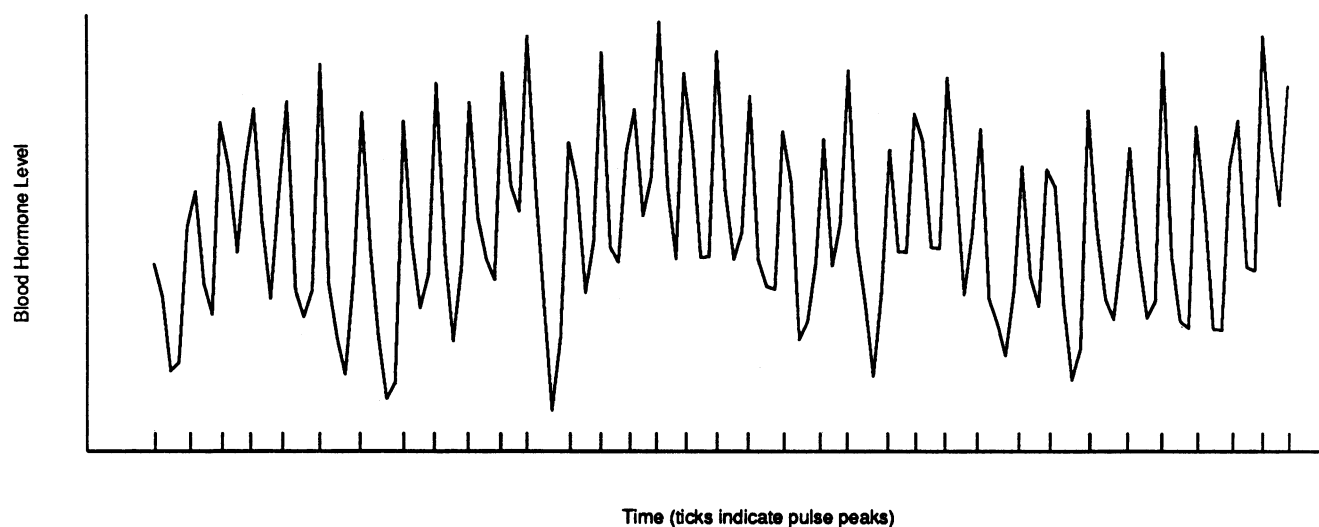
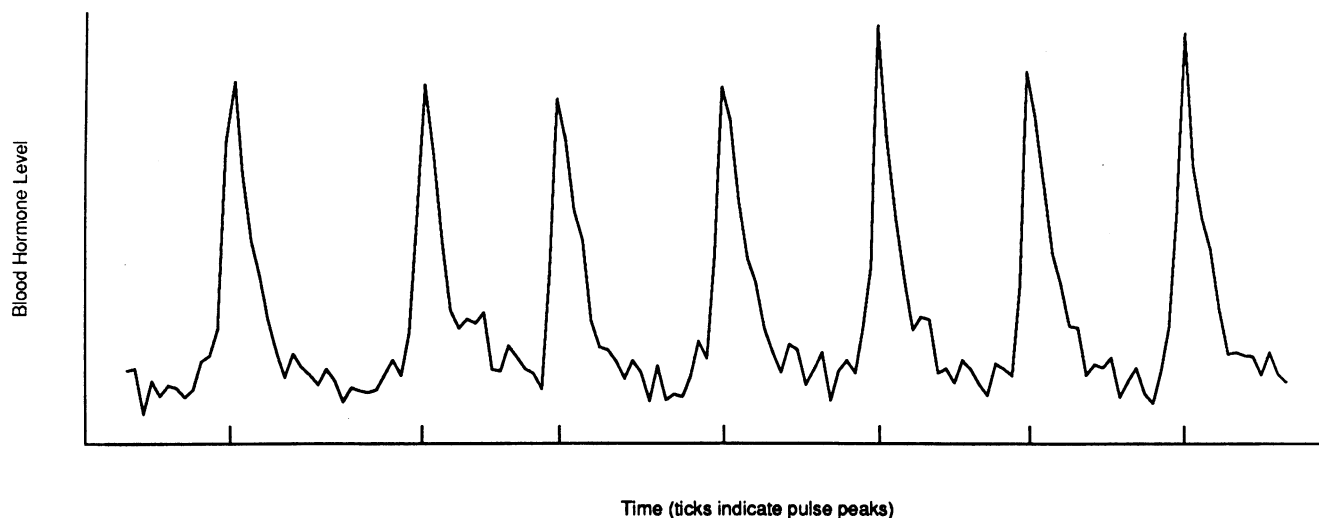
Figure 3.3(ii) : Estimated pulse and pulse peak times for the data. The regularization parameter is chosen so the estimated coefficient of variation is about .15.

Figure 3.4 : Estimated pulses (March, September) for the cow data. All plots are on the same scale. The horizontal scale is in hours and stretches from a half an hour before the beginning of sampling to half an hour past the end of the sampling period.

Figure 3.5 : Estimated pulse peak times (March, September). The ticks indicate the pulse peak times. The horizontal axes are set up as in Figure 3.3.

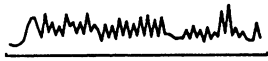
An Individual Pulse



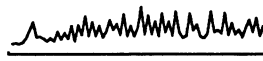


LH Data (March)

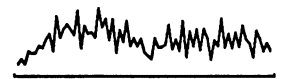
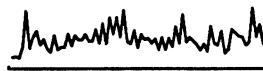
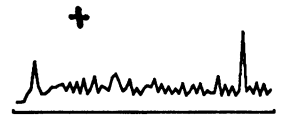
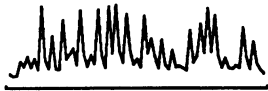
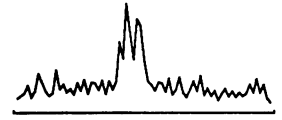
3 months



5 months

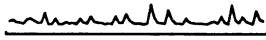


7 months



LH Data (September)

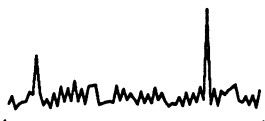
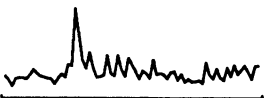
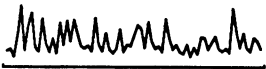
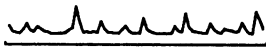
3 months



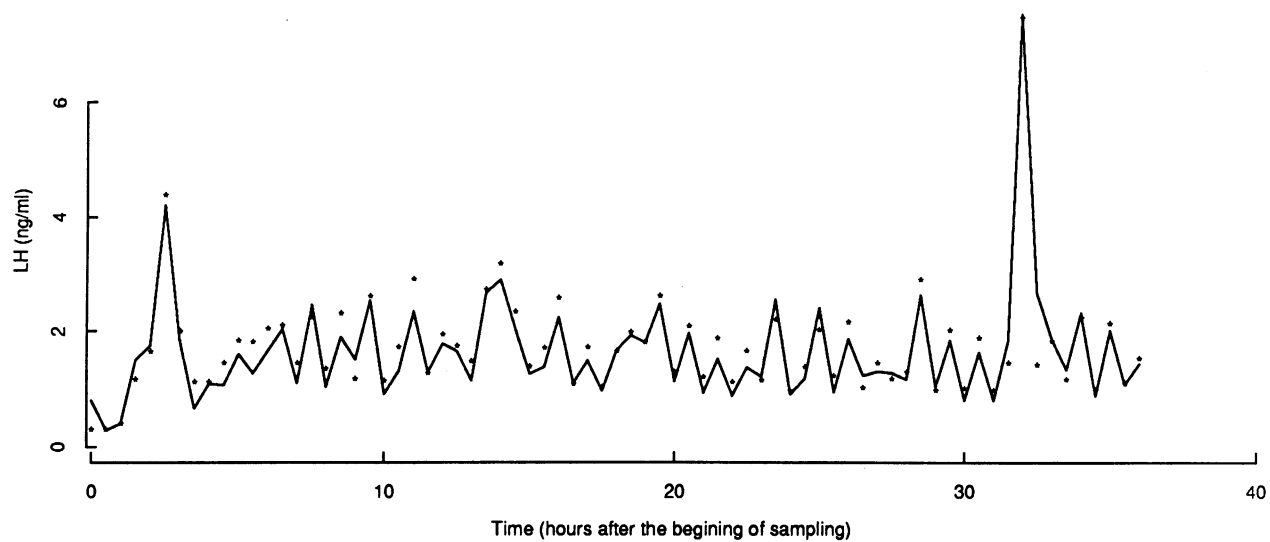
5 months



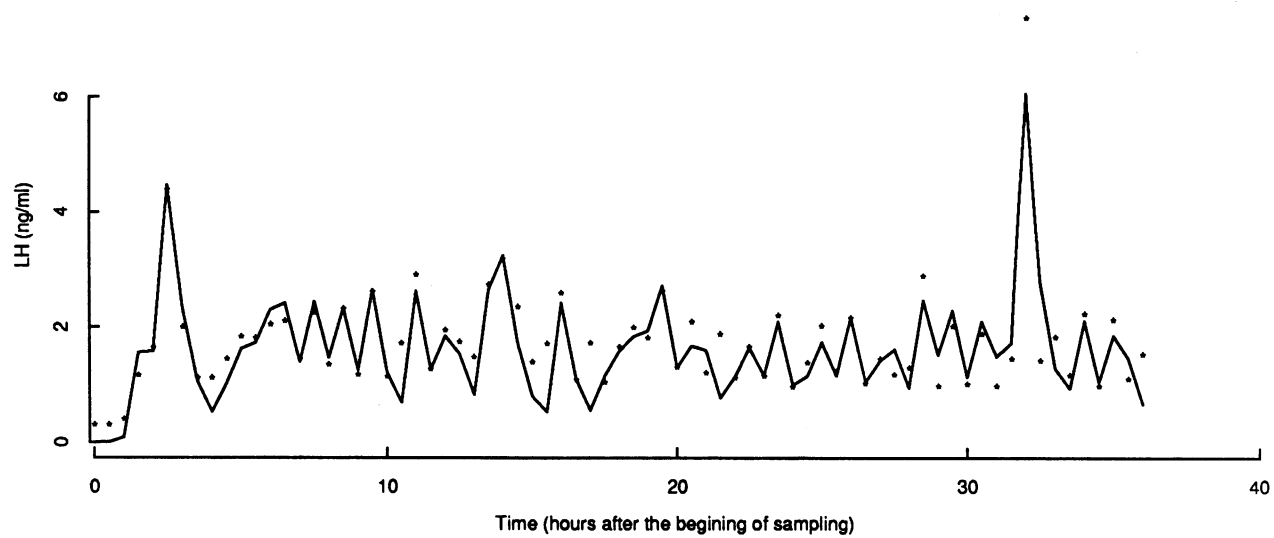
7 months



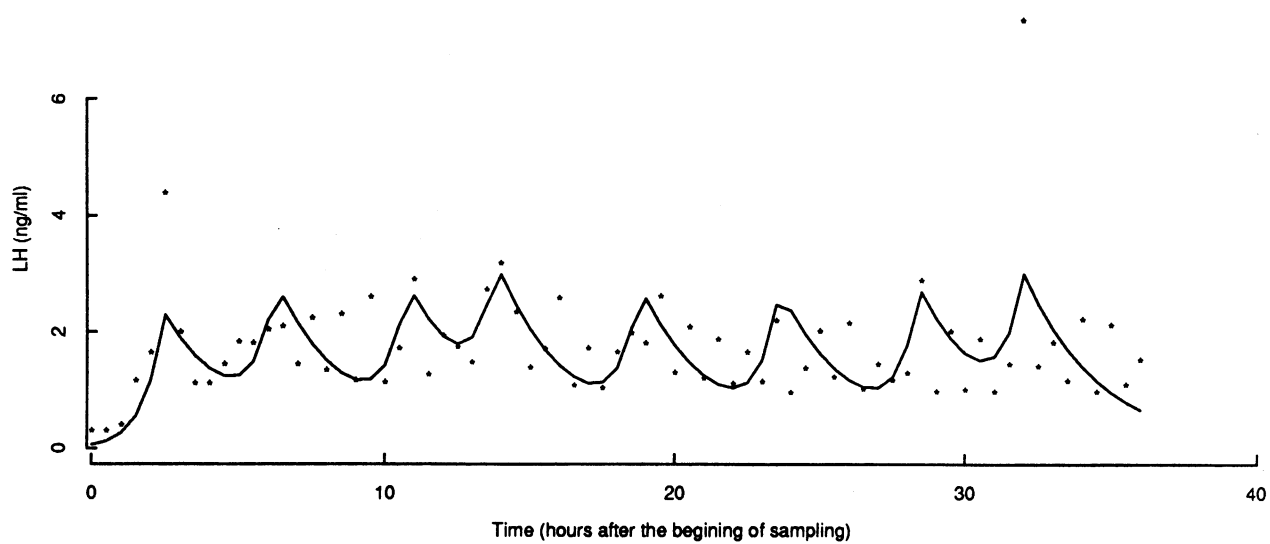
51 Pulses

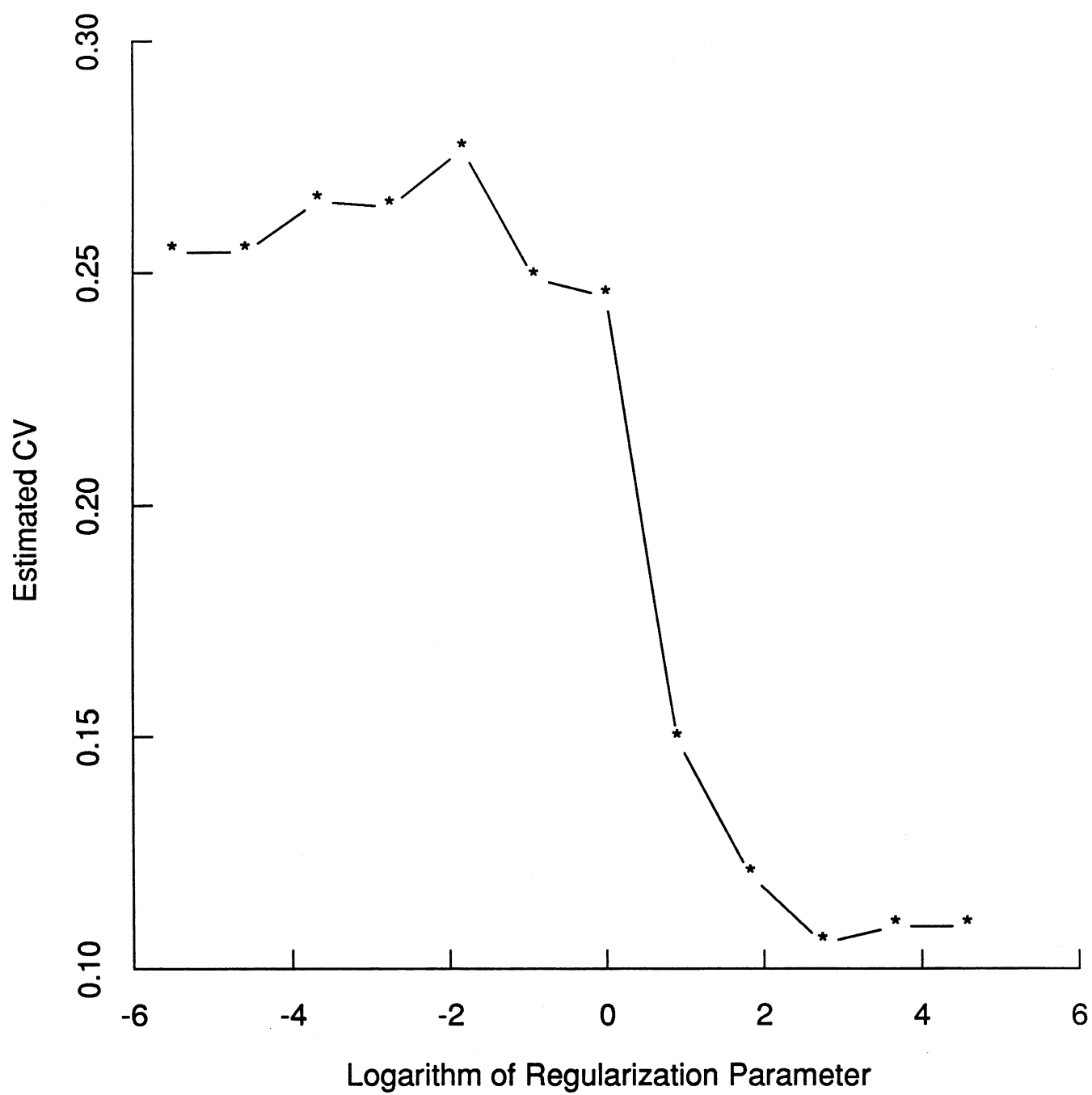


26 Pulses

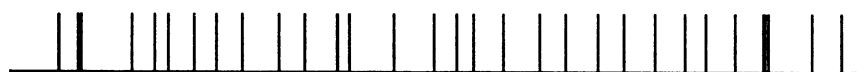
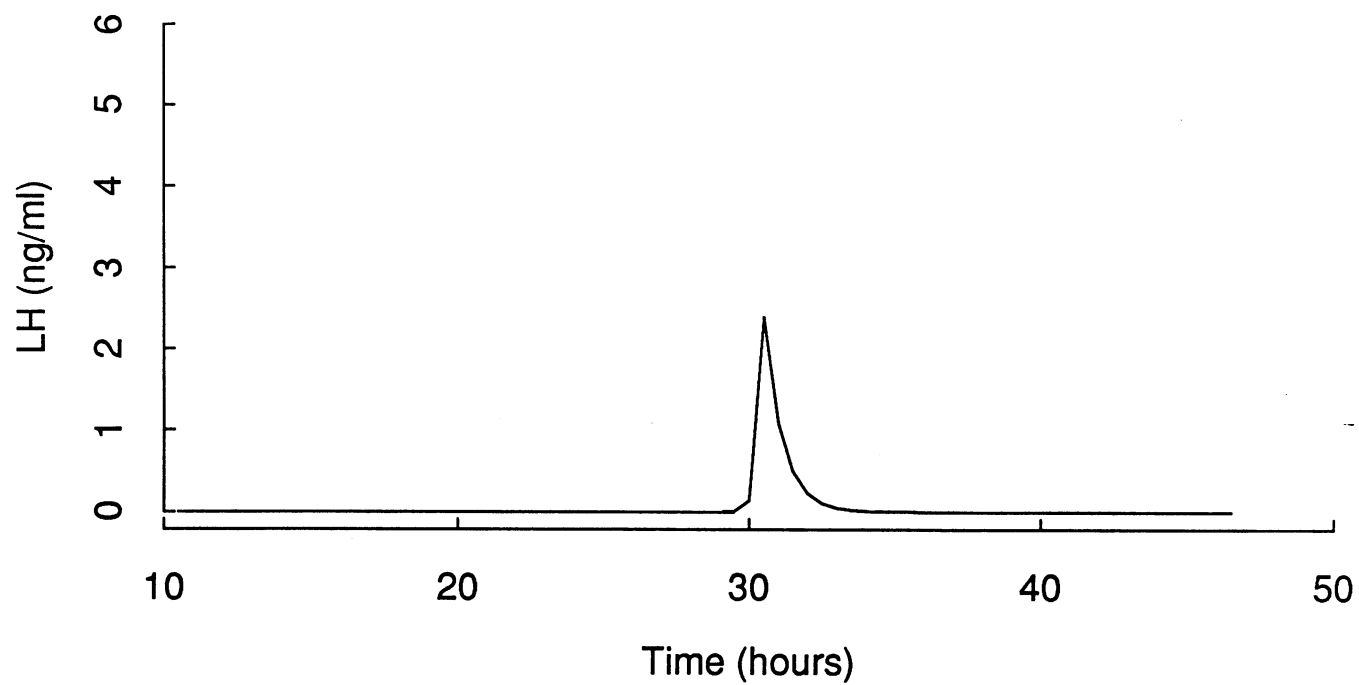


8 Pulses





Individual Pulse



Spike Train of Pulse Peak Times

Individual Pulses (March)

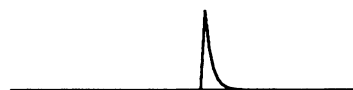
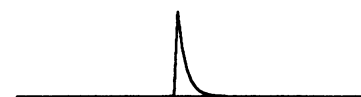
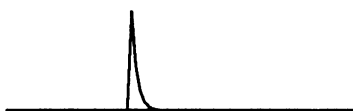
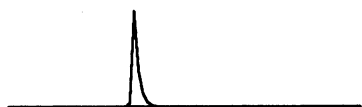
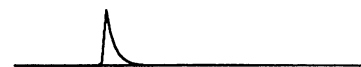
3 months



5 months



7 months



Individual Pulses (September)

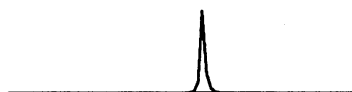
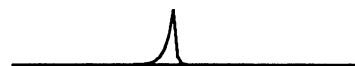
3 months



5 months



7 months

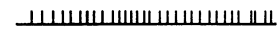
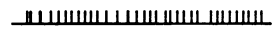
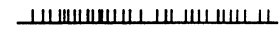
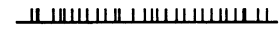
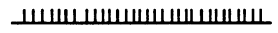
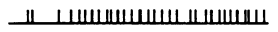
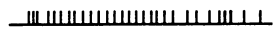


Pulse Peak Times (March)

3 months

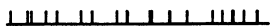
5 months

7 months



Pulse Peak Times (September)

3 months



5 months



7 months

