Neural Response Modelling and Spike Rate Estimation Techniques

Scott Monk

Department of Electrical & Computer Engineering
McGill University
Montreal, Canada

March 2014

A thesis submitted to McGill University in partial fulfillment of the requirements for the degree of Master of Engineering (M.Eng) in Electrical Engineering

2014/03/24

Abstract

Using point processes to model neural spike sequences allows the application of classical estimation techniques in their analysis. Estimation of the time varying rate at which spikes occur is often conducted to draw inference on the stimulus which triggered the response. Such estimation schemes are often founded on the assumption that spiking follows Poisson statistics, however, the observed firing rate is a product of both the stimulus and bio-physical properties of the neuron. A point process model for neural data must then incorporate dependency both on stimulus and intrinsic characteristics of the cell. To achieve this we modify the Poisson model such that it includes the refractory phenomenon observed in spiking behaviour. This results in a modified firing rate which is free from distortion caused by refractory effects. A Maximum Likelihood (ML) estimation technique for this adjusted firing rate which better represents some relation to the stimulus is presented.

We propose and justify a parametric model to represent a broad class of arbitrary firing rates. The corresponding likelihood equation for the firing rate parameters given an observed spike sequence is derived, however, several numerical methods are required to find the ML estimate. These techniques are presented in detail and include model order selection and non-convex optimization. An empirical study is conducted to determine which model selection rule, from several approaches found in the literature, is most accurate. Global maximization of the non-convex likelihood equation is carried out using a transformation method known as a filled function.

Computer simulations show that our proposed estimator can potentially lead to more accurate estimates of firing rates, as opposed to a Poisson scheme, when the data is affected by a refractory period. Results demonstrate that the error is relatively constant across data sets influenced by a range of refractory periods, indicating the estimator is robust. Rate estimates on real neural data taken from various cortices also show improved goodness of fit when contrasted with results from the Poisson estimator. A brief performance comparison with other popular estimation schemes suggests superior estimates are produced by our proposed scheme.

Sommaire

Un processus de point pour modeler des séquences de piques neuraux permet l'application des techniques d'estimation classiques dans leur analyse. L'estimation du taux variable de temps auquel les piques ont lieu est souvent faite afin de trouver l'inférence sur le stimulus qui déclenche la réaction. Ces schémas d'estimation sont souvent basés sur la supposition que la fréquence de piques élevés suit les statistiques Poisson. Cependant, le taux de piques est un produit du stimulus et des propriétés biophysiques du neurone. Un modèle de processus de point pour les données neuraux doit intégrer la dépendance du stimulus et des propriétés intrinsèques de la cellule. À cet effet, on modifie le modèle Poisson pour qu'il inclue le phénomène réfractaire observé dans le comportement piquant. Selon ce modèle ajusté, on présente la technique d'estimation Maximum de Vraisemblance (MV) pour le taux de tir qui provoque la réaction piquante.

On propose et justifie un modèle paramétrique pour représenter des taux de tir arbitraires et extensifs. L'équation de vraisemblance correspondante pour les paramètres de taux de tir se produit quand une séquence piquante est dérivée. Néanmoins, plusieurs méthodes numériques sont requises pour trouver l'estimation du MV. Ces techniques sont présentées en détail et incluent la sélection d'ordre modèle et l'optimisation non convexe. Une étude empirique, menée afin de déterminer quelle règle de sélection de modèle et inspirée de plusieurs approches trouvées dans la littérature, est la plus exacte. La maximisation globale de l'équation de vraisemblance non convexe est menée en se servant d'une méthode de transformation qui est connue comme une fonction de remplissage.

Des simulations informatiques montrent que notre estimateur proposé livre des estimations de taux de tir plus exactes qu'un schéma semblable de Poisson quand les données sont affectées par une période réfractaire. Les résultats démontrent que l'erreur est relativement constante à travers les ensembles de données influencés par plusieurs périodes réfractaires, ce qui indique un estimateur robuste. Les estimations de taux de tir sur des réelles données prises de plusieurs cortex montrent aussi une bonté de convenance (goodness of fit) lorsqu'elles sont contrastées avec les résultats de l'estimateur Poisson. Une comparaison de performance avec d'autres schémas d'estimation populaires suggère que des estimations supérieures sont produites par notre schéma proposé.

Acknowledgments

I would like to thank Prof. Harry Leib for his intellectual support, guidance, and motivation over the past several years. I also wish to thank the many other students both in and outside of my lab whom I have met during my stay at McGill. I must also give special mention to the following individuals: Kwende Kefentse, Patrick Bastedo, Shub Roy, Dan Parker and Matthew Thompson.

Finally I wish to credit those who have made their neural data available. The data analyzed in Figures 4.19 and 4.18 was collected by Pascal Fries at the National Institute for Mental Health. This data is available on the Field trip website (`fieldtrip.fcdonders.nl/`). The reach trial data used in Figures 4.21, 4.20 and 5.5 was recorded by Robert Flint at the Department of Neurology, Northwestern University and provided through the CRCNS program. The bursting data used in Figure 4.22 was collected by Ariel Rokem at the lab of Andreas Herz and was also provided through the CRCNS program (`crcns.org`).

Contents

List of Figures

List of Tables

List of Acronyms

AIC	Akaike's Information Criterion
AICC	Corrected Akaike's Information Criterion
BARS	Bayesian Adaptive Regression Splines
BIC	Bayesian Information Criterion
CDF	Cumulative Distribution Function
GLM	Generalized Linear Model
EM	Expectation Maximization
ISE	Integrated Squared Error
ISI	Inter-spike Interval
IPFM	Integral Pulse Frequency Modulation
KL	Kullback Leibler distance
MDL	Minimum Description Length
ML	Maximum Likelihood
MISE	Mean Integrated Squared Error
NMISE	Normalized Mean Integrated Squared Error
PDF	Probability Density Function
PMF	Probability Mass Function
PSTH	Post-Stimulus Time Histogram
QQ	Quantile-Quantile

Chapter 1

Introduction

Understanding the structure of neural signals is a fundamental concern of computational neuroscience [1] [2]. Since such early work as Hubel and Wiesel [3], sequences of action potentials have been studied in order to deepen our understanding of how neurons encode and transmit information about the outside world. The process of decoding these sequences, known as spike trains, is the central focus of this thesis.

Action potentials are the primary building block of any neural coding scheme, however, the pulse waveform itself is thought to give little insight into the information carried by the observed sequence [4]. Instead information seems to be somehow encrypted in the temporal occurrence of the spikes [2]. A heuristic split between two coding concepts has been made, with evidence for both existing in different areas of the central nervous system. The first, known as rate coding, conveys information in its firing frequency. Under this hypothesis the time between spikes is less important than how often they occur. Thus knowing the instantaneous firing rate is equivalent to decoding the transmitted signal. The second paradigm, termed temporal coding, dictates that the precise timing of a spike, either in relation to a stimulus feature or other spikes, conveys the message [1]. In this work we

concern ourselves with rate coding and begin with an overview of existing methods to estimate the instantaneous rate.

1.1 Literature review

The question of how to combine firing rates from single neurons to estimate a complex stimulus is context dependent, and may require some prior information as to the function of the neural circuit under measurement. More specifically, estimating the firing rate of a single cell is not likely to to yield a full complex stimulus decode. Instead, individual firing rates can be combined in some way to give an estimate of some higher dimensional complex stimulus such as a visual image, as might be done with measurements from the visual cortex, or a complex limb movement, as might done in motor neuron decodes for controlling a prosthetic. For example, one of the earliest successful schemes for predicting the intended direction of arm movement, given the firing rates of a population of motor neurons, requires some prior knowledge of the preferred direction of the neurons being recorded. This method is known as the population vector and was first introduced in [28]. In this combining scheme, preferred direction is estimated from a training set of data by noting the direction of arm movement during which the neuron in question seems to fire the most frequently. This direction is represented in three dimensional space by a vector. After the preferred direction and maximal firing rate of each neuron is known, new recordings can be used to predict the subjects arm trajectory by summing the preferred direction vectors of the population, scaled by their firing intensity. Since this model was introduced there has been some debate on whether or not the neurons are actually encoding direction or force as a function of time [29]. While such issues certainly deserve attention, what is most important from an estimation perspective is that whatever the quantity encoded may be,

the neurons communicate this through a rate coding scheme. If this assumption is accurate, then correlation between the rate estimates and experimental variables can be conducted to investigate which quantities are most likely represented by a particular neurons response.

A different type of combination is thought to take place in the visual processing pathway. A model proposed by Hubel and Weiesel, known as the complex cell model, describes how information from many neurons might be combined to allow the detection of basic shape movement, and orientation [1]. Essentially, several simple cells whose firing rates are driven by light in a certain region of the retinas field of view are fed to the input of a complex cell in the lateral geniculate nucleus (LGN). If this complex cell receives enough excitation from the simple cells then it begins to fire as well. However, the output of the complex cell correlates with more complex stimuli than simply light being incident on a certain receptor in the retina. Hubel and Weisel showed that complex cells in the LGN combine information from simple cells having only simple spatial receptive fields to detect things such as bars of light, their orientation and direction of movement. More recent research has shown that this sort of feature extraction becomes more and more specific as information is passed through the visual pathway (see for example [30]). This sort of result illustrates that the meaning of the firing rate does not remain the same and that neurons are not only relaying information, but also processing it. If the complex cell generates its output based on some function of the rates seen at its inputs, then rate coding and rate estimation can also be used to predict and model its behaviour. Even though this model is quite different from the population vector, it still relies on rate estimates in the stimulus decode, and therefore we can see that firing rate estimation is an integral part of higher level stimulus decodes, even if the rate does not directly describe the stimulus. We dedicate the remainder of this section to an examination of firing rate estimators in the literature.

A simple method for estimating the instantaneous spike rate is the post-stimulus time

histogram (PSTH) [1]. This piece-wise constant function, which approximates the rate by assuming it is constant over short intervals, is constructed as follows. Time is divided into N discrete non-overlapping bins b_i where b_i spans some interval $[\tau_i, \tau_{i+1})$, $i = 1, \ldots, N$. Each spike t_j is placed into the bin b_i for which $t_j \in [\tau_i, \tau_{i+1})$. A constant function spanning each bin is then drawn. The value of this function equals the number of spikes in the bin divided by its width $(\tau_{i+1} - \tau_i)$. While the PSTH is easy to compute, the results are highly dependent on how the bins are chosen. This dependency can be seen in two ways. First, the temporal resolution of the rate estimate decreases as bin-width increases. Second, the estimate must be an integer multiple of the inverse of the bin-width it spans. This follows from the fact that the number of spikes in any bin must take an integer value [1]. A recent method to optimize this bin-width based on minimizing the mean integrated squared error (MISE) is presented in [5]. The optimization is done assuming the spike patterns follow Poisson statistics. Traditionally, however, the bin width is chosen in a somewhat ad hoc fashion.

We illustrate this classic approach in Figure 1.1. A set of spike times is shown in frame A and two PSTHs are shown in frame B with different bin widths. We can see that the narrower bin width leads to finer temporal detail but also a more variable rate. Wider bins lack in temporal detail but provide a more slowly changing rate estimate.

One proposed improvement over the PSTH is linear filtering [1]. This approach begins by representing the observed spike train $\mathcal{T}$ as a sequence of impulses. This sequence is generated by placing a shifted impulse at each moment a spike was observed

$$\sum_{i=1}^{N(T)} \delta(t - t_i),$$

where $N(T)$ is the number of observed spikes, t_i the time of the $i'th$ spike [1] and $delta(t)$

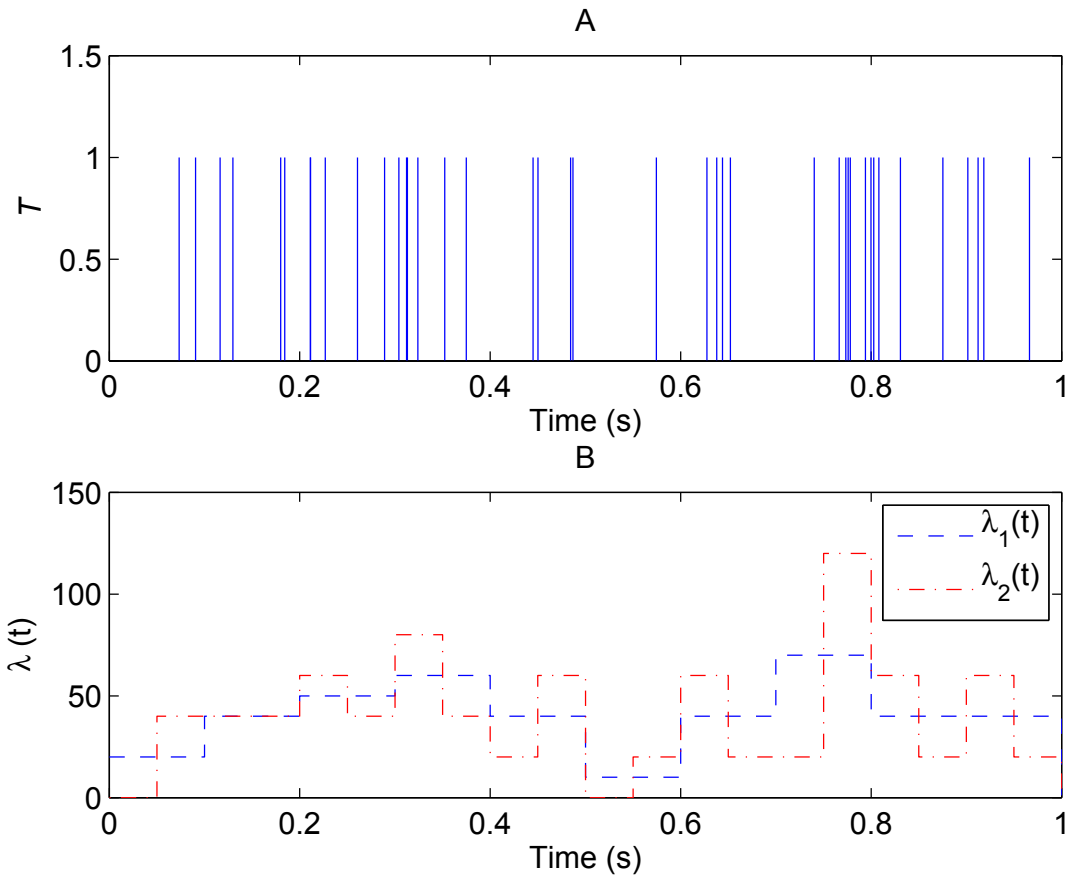

Fig. 1.1 Frame A shows a set of spike times $\mathcal{T}$ represented by impulses at times t_i. Frame B show two different rate estimates using a PSTH, having a bin widths of .1 and .05 seconds for $\lambda_1(t)$ an $\lambda_2(t)$ respectively.

is the Dirac delta function defined by

$$\delta(t) = 0, t \neq 0$$
$$\int_{-\infty}^{\infty} \delta(t)dt = 1$$

The instantaneous rate is recovered by convolving a window function $g(t)$ with the impulse train. Analogous to the bin-width in PSTH, the bandwidth of the window as well as its general form must be selected. How to optimally choose the bandwidth for $g(t)$ is the most

debated issue with regards to linear filter rate estimation schemes. A common convention is to restrict $g(t) \geq 0$ for all t, and insure $\int g(t)dt = 1$ is satisfied. The first restriction causes the resulting firing rate to always be positive, since negative rates do not have any meaning here. The second restriction enforces the correct scaling of the estimated rate, such that its units are spikes per second. These two restrictions do not put much constraint on the form of the filter, so it is natural to ask what shape for $g(t)$ is most optimal. An investigation carried out in [6] concluded that the form (ie Gaussian, boxcar, triangle) was highly irrelevant in terms of Integrated Squared Error (ISE) between the true and estimated rates. In contrast, the filter bandwidth was shown to have a very large impact on the error of the final estimate. As a result, methods which guide bandwidth selection are numerous in the literature, while the exact form receives less attention, although some authors have proposed advantages of particular functions.

A simulated data set $\mathcal{T}$ as well as two rate estimates resulting from Gaussian shaped $g(t)'s$ with different bandwidths are given in Figure 1.2. The data set is presented in frame A and the two filters used in frame B. Frame C shows the result where we can see immediately that the trade offs in filter bandwidth selection are similar to PSTH bin-width selection, however the linear filter gives a much smoother result than the PSTH regardless of the window's bandwidth. Filters that are narrower in time (wider bandwidth) give greater temporal detail, however, wider kernels tend to give smoother estimates. Once the filter is selected the rate estimate is quite easily obtained through one straightforward convolution.

A heuristic method for bandwidth selection is proposed in [6], and shown to yield close to optimal results. The procedure is summarized as follows. Start with a wide bandwidth filter and reduce it in incrementally, measuring the pairwise ISE between the current and previous estimate at each step. Plotting the pairwise ISE results should reveal a clear minimum at a particular step. The authors argue that for responses of moderate to high

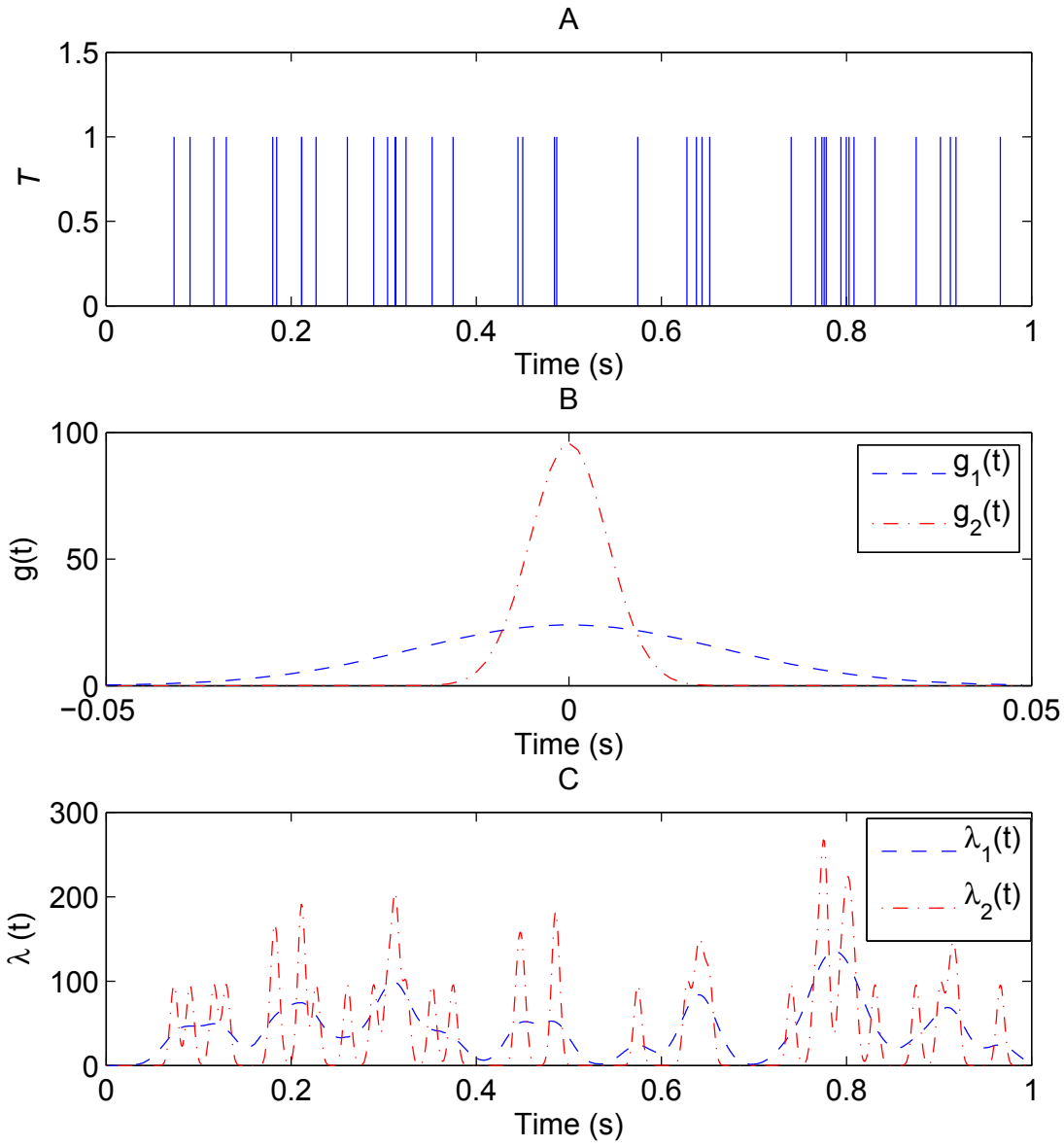

Fig. 1.2 Frame A shows a spike sequence represented by shifted impulse functions. Frame B shows two Gaussian kernels used for $g(t)$ in the firing rate estimates $\lambda(t)$ shown in frame C.

intensity (spike count), this minimum occurs close to the optimal bandwidth in terms of mean integrated squared error. Simulation results showed that this method can perform well in some cases, improving when the response contains many spikes. Further details of this heuristic optimization are given in Chapter 5, where we compare the performance of three different rate estimation schemes.

A bandwidth optimization strategy is given in [7] which suggests a Kaiser window may be the optimal filter form, and that an appropriate bandwidth can be selected by analyzing the frequency content of the spike sequence. Once the appropriate bandwidth is found, the coefficients for the Kaiser window can be calculated numerically, and the resulting linear filter can be applied to the spike sequence through convolution as discussed previously. This strategy was shown to perform quite well on both simulated and real data. The real data was measured from vestibular afferents, considered to be first order sensory neurons which employ rate coding to convey information. This implies that the intensity of the firing rate directly maps to some feature of the stimulus, in this case rotational movement of the subject. With knowledge of the experimental conditions one can then measure the accuracy of the rate estimator against the true stimulus. The simulated data was generated using an integral pulse frequency modulation scheme. Under this model a neuron acts as an integrator with the stimulus as its input. Once the integrated stimulus passes some threshold, a spike is generated and the integrator is reset to zero. It was shown that the frequency spectrum of real and simulated spike trains from this model are strikingly similar. To add some variability to the simulated trains, noise was mixed with the signal at the input to the integrator. This helped create further agreement between the generated spike sequences and those produced by irregular afferents in the vestibular system. As the input to the integrator became more randomized, the estimator gave slightly less accurate results. This is not surprising since the signal to noise ratio of the cells excitation is diminished.

This estimator is unique in that it optimizes based on a frequency domain representation of the spike sequence. The experimental design of these results also give more credibility to rate coding models, since under this paradigm estimates of the subjects rotation were accurately recovered.

Most literature regarding linear filtering reports that intense stimuli, which produce many spikes, tend to dictate narrower filters, while low intensity rates are more accurately estimated when using a wide kernel width [6]. We might hypothesize that a very dynamic firing rate which is both strong and weak over its time course could more accurately be estimated using a filter whose bandwidth in not fixed over time. This concept is investigated in [8] where, based on a Poisson spiking model, a filter that is optimal in the MISE sense is found. Essentially, the spike train is divided into sub-periods over which the filter bandwidth is fixed. The number of sub-periods is determined buy adding a smoothness or regularization term to a convex cost function derived from the MISE. Once this function is defined, simple optimization techniques can be used to find the optimal bandwidths and their periods of use. Since evaluation of the MISE requires knowledge of the true excitation to the cell, a substitute function which relies only on known variables is suggested, and then used in lieu of the MISE. Results showed that there were some gains to be had using such an approach, and the behaviour of the bandwidth over dense and sparse spiking regions was as anticipated. However, it was found that pooling of data from many trials was required before any advantages were seen over an optimal fixed bandwidth method. This result is interesting because it shows that details in complex models may be hard to estimate when only limited amounts of data is available.

While other attempts to optimize linear filters have been presented, we feel those mentioned above capture the key ideas used in the majority of the literature and move on to more innovative methods of firing rate estimation. In [9] a method known as Bayesian

Adaptive Regression Spine (BARS) is presented as a general method for the statistical smoothing of data. In one proposed application, the authors smooth a PSTH generated from spiking data. This approach fits cubic splines to a PSTH in order to achieve a smooth rate estimate. Since cubic splines are able to approximate a considerable class of time varying functions, they are therefore suitable for estimation of arbitrary firing rates. This type of fit side steps the need for a time varying linear filter, and therefore many of the issues associated with bandwidth selection. Even though the BARS method relies first on constructing a PSTH, many of the issues that arise from choosing too small a bin size are quite literally smoothed over by the fitting of cubic splines. This results in excellent temporal resolution and globally smooth estimates. The number knots (connecting points between constituent cubic polynomials) and their positions are determined using a Bayesian framework. Relevant probability densities are found by assuming the PSTH data is Poisson in nature and by using a reversible Jump Markov Chain Monte Carlo method to approximate distributions which do not yield analytic forms. A follow up report [10] elaborates on how the Poisson assumption can be altered, allowing the use of other point process models instead. The statistical nature of the model also allows inference on things such as confidence intervals, which are absent in linear filtering techniques. This richness, however, comes at the cost of added computational complexity.

A comparison between some of the previously mentioned methods, as well as others, is presented in [11]. The review focused on a particular application, arm trajectory decoding from a population of motor neurons. This type of decoding is slightly more complex than the vestibular decode previously mentioned because it requires the combination of time varying rates measured from several neurons to compose the final estimate (an arms trajectory). First the individual rate estimates from a population of neurons are acquired, then these rates are then combined in some way, based on prior information about the direction in

space each neuron seems to prefer, to estimate the trajectory of the test subject's arm over time. The first step of the comparison showed noticeable differences in the estimates produced by the various methods. Despite this, the results of the second combining step showed a low sensitivity to this rate variability. It was hypothesized that although different intensities are estimated in the first stage, the SNR of each estimate may be similar. A comparison of two different combining methods also showed that the most appropriate rate estimator changed depending on the implementation of the second step. This shows that different estimation schemes may have situation dependent strengths and weaknesses. This ties back to the result of the time varying bandwidth results in that a more complex estimator is not always going to be more accurate. It is then useful to figure out under what conditions particular schemes are beneficial.

Most statistically based methods are developed under a Poisson spiking model. This assumption is usually justified based on the limiting form of combined point processes. Loosely, as you pool many non-Poisson processes together, the statistics of the resulting process converge to those of a Poisson process [12] [13]. Methods intended for use with pooled responses from a repeated stimulus might not suffer much from modelling the data as Poisson, particularly when the number of trials is large. However, analysis of spike trains on a trial-to-trial basis shows that they are in fact not represented well by a Poisson process [13]. One striking difference is that while a Poisson process favors short inter-arrival times, real neural data seems to suggests that a refractory phenomenon is present in the spiking mechanism [14] [15]. This observation correlates with biophysical models, such as the Hodgkin-Huxley, which suggest such a period should be present based on how action potentials are generated [16] [4]. The refractory period is thought to be composed of two distinct sections. A short interval immediately following an action potential, during which time the chances of seeing another spike are extinguished, is known as the absolute

refractory period. This is followed by the relative refractory period, during which time the probability of observing a spike is less than what the stimulus would dictate otherwise. Estimators, statistically based or otherwise, intended to be used on single trial data may benefit from models which include such refractory effects. We conclude this review with a presentation of estimation schemes which include such spike models.

The main difficulty when estimating the rate from data which show history dependency is that the form of this dependency must be inferred at the same time as the cells excitation [17]. With respect to the refractory period, a common assumption [17] [18] [19], is that the present rate is a function of the input to the cell and the most recent spike. This tends to simplify the problem in a way that it can be efficiently solved. In [18] a Maximum Likelihood (ML) estimate for both the stimulus and refractory effect is presented. The form of the refractory period is constrained to a non-decreasing piece-wise constant function that scales the input excitation following a spike. The method is optimized under the assumption that the true excitation is periodic and that multiple trials may be available. While such optimization may not be broadly applicable, it is justified since the estimator is presented in the context of stimulus recovery from auditory nerve cell spike data. Once defined the likelihood is maximized using an iterative Expectation Maximization (EM) algorithm. Refractory periods estimated suggest a short absolute refractory period followed by what is at first a sharp recovery, but then slowly returns to unity over several milli-seconds. While this is one of the earlier attempts at solving this type of problem, the employment of numerical methods and non-Poisson models serves as inspiration for what follows.

A method for estimating the refractory period from the observed inter-spike interval (ISI) distribution is given in [14]. To the unfamiliar, this distribution is given more detail in Chapter 2. It's sufficient for now to note the assumptions that must be made to do so. First, the instantaneous rate only depends on the stimulus and most recent spike, and

that any refractory effects are relatively short. Second, that for the purposes of refractory estimation, the stimulus is constant. Under these conditions the form of the refractory period can be directly found from the ISI distribution. A short refractory period tends to have little impact on long inter arrivals. This means long inter-arrivals are distributed exponentially, similar to what would be seen if the data was Poisson. Hence, the true rate as determined by the stimulus can be approximated by fitting an exponential to the tail of the observed distribution. The parameter of the exponential distribution found corresponds to the rate of the constant stimulus. The refractory effect can then be teased out of the short inter-arrival times by finding the scaling function which would create such an ISI distribution given a constant stimulus. Once the refractory period has been approximated it is used to compensate the values in a PSTH style stimulus estimate. The temporary use of a constant stimulus may seem unreasonable, however, the authors show that simulated spike trains generated by non-stationary stimuli, affected by refractory periods inferred with this method, have remarkably similar statistics to real neural data. This scheme is unique in that the stimulus is estimated twice. Once as a constant function, in order to obtain the refractory period, and again as a non-stationary rate with appropriate compensation based on the refractory period. We feel the use of the ISI distribution in this way generally requires large amounts of data to be reliable, so this method is not likely to give robust results when trial-to-trial estimates are being made with no additional side information. This stems from the fitting of the exponential distribution to the tail of the ISI distribution, which is likely to be very sparsely populated on single trial data, and therefore lead to a large variance in the constant rate approximation. Another issue with this method is that it results in what is essentially a PSTH, with values compensated according to some refractory effect.

Generalized Linear Modelling (GLM) has also been applied to spike rate estimation [17]

[20] [21]. This framework has mostly been used to explore relationships between several covariates and spiking activity. For example, how the position of a mouse in its environment can modulate spike activity in its hippocampus [22], or the relationship between one neurons spiking to another [20]. GLM can also be used to estimate refractory effects, as is done in [23] and [17]. Typically the input to the neuron is modelled as a summation of a time varying stimulus and a filtered version of the neurons output. The form of the refractory filter is estimated by assuming it is a linear combination of non-linear functions, as is the excitation. Careful choice of the non-linear functions can ensure the likelihood function for their weights is convex [24]. As such, simple numerical searches can be used to estimate the rate and refractory period. Some works suggest a spline basis as the non-linear function and a linear combination of these give the rate estimate [17] [20]. In such a case the likelihood function contains one parameter per basis function per covariate. The number of basis functions used to represent the rate, refractory period, and any other covariates must be chosen in advance, but other than this the GLM framework makes few other assumptions and requires no additional side information. This framework is also given further consideration in our performance comparison in Chapter 6.

Finally, we emphasize that any estimator can be used on any point process data, regardless of the true generating mechanism. However, we might not expect a rate estimator which lacks key elements in its structure to perform as well as one which has the right enhancements. To be more precise, the problem is best approached in two steps. First we must compose a model which mimics well real spike data, and second we must have an estimator which takes into account this structure of the underlying model when estimating the cell's excitation.

1.2 Objectives, organization and contribution

The main objective of the research presented in this thesis is to develop a tractable ML approach to free firing rate estimation from neural data. As discussed in the previous section, the issue of untangling the free firing rate from other spiking factors, such as the refractory phenomenon, is a difficult one. We present a method which combines several estimation techniques and numerical methods to accurately recover a cell's excitation on a trial-to-trial basis, with no additional side information. We have made the following assumptions in our work. First, the spike sequences are generated according to a self exciting point process whose instantaneous rate is a function of the cell's excitation and most recent spike [12]. Second, the free firing rate itself can be accurately modelled by the exponential of a polynomial with an arbitrary order [25]. Third, the refractory effect is well modelled by a certain parametric function which modulates the influence of the stimulus on the instantaneous rate.

The remainder of this document is organized as follows. Chapter 2 outlines some key terminology regarding point processes. We then present a parametric model for the free firing rate and develop a spiking model which includes both relative and absolute refractory periods. The chapter concludes with a derivation of the stimulus likelihood given a spike sequence observation.

Chapter 3 details the numerical procedure we have developed in order to maximize the likelihood function derived in Chapter 2. Methods for dealing with the refractory parameters of the model are discussed. Model order selection rules from the literature are also examined and defined with respect to our model. We then look at convex and non-convex maximization techniques employed to obtain the final stimulus estimate.

A series of numerical results are given in Chapter 4. We show the fit of our free firing

rate model to different classes of time varying functions and discuss its suitability as a model for neural firing rates. We then present results which compare the performance of different model order selection rules to justify our final choice. Next we give results regarding the accuracy of the estimator using synthetic data generated by various excitations. Finally, results on real data are examined using goodness of fit metrics to validate our refractory model. Chapter 5 then compares the performance of our proposed estimator with two other common schemes. Commentary on the pros and cons of each method is provided. Concluding remarks and future directions are given in Chapter 6, followed by appendices.

The main contributions of this thesis are:

1. Develop and test a ML approach for free firing rate estimation which compensates for refractory effects.

2. Investigate the suitability of the exponential of a polynomial found in [25] as a firing rate model for neural rate estimation.

3. Present goodness of fit results on real data showing that our parametric refractory period is an accurate model. While this form, or one similar, has been suggested in certain texts [1] [26], little justification beyond heuristics is given. Our results strengthen this suggestion. We also describe conditions under which this model may not be reasonable.

4. Compare our method to other popular methods in order to benchmark its performance.

Chapter 2

Spike sequence modelling

Point processes provide a convenient framework for the analysis of spike trains. By definition a point process is a random set of localized events, each occurring at a specific moment in continuous time [12]. Thus at any moment a point process takes on one of two values: one if an event occurred at time t, and zero otherwise. Since the waveform of an action potential is not relevant for sequence decoding, the only information we require about a spike is the time at which it occurred. The connection between spike data and point processes follows immediately. Spike train measurements can be converted to a point process by constructing a signal which is a series of impulses located at the precise moments which action potentials were observed. Once the data is in point process form we can begin analysis using a probabilistic model for the spiking behaviour. This chapter first presents the notation we use throughout the thesis to describe point processes, followed by our model which defines the relationship between a cells excitation and spike response.

2.1 Point process notation

Let us denote the beginning of some observation interval as t_0, and the end of this time period as T. Then a point process realization would be a set $\mathcal{T}$ of event times $t_i, i = 1, 2, 3, ..., N(T)$ on the interval $(t_0, T]$, where $N(T)$ is an integer equal to the number of events observed over $(t_0, T]$. In general $N(t)$ is known as a counting process or counting function. Given a point process realization, $N(t)$ is equal to the number of points in $\mathcal{T}$ that satisfy $t_i < t$. From this our use of $N(T)$ to specify the total number of events in one sequence seems logical. The counting function may at times be seen taking two parameters $N(t_i, t_j)$ which is simply short hand notation for $N(t_j) - N(t_i)$. The counting process for a spike train is a random process, the statistics of which are fully specified by its associated rate function, $\lambda(t)$ defined

$$\lambda(t) = \lim_{\Delta \to 0} \frac{1}{\Delta t} Pr[N(t, t + \Delta t) = 1 | P], \tag{2.1}$$

where P represents an arbitrary event [12]. Two useful interpretations of $\lambda(t)$ are that it represents the instantaneous average event rate of the process at instant t, and that $\lambda(t)\Delta t$ approximates the probability of an arrival occurring in $(t, t + \Delta t)$ [12]. In general there is little restriction on the form of $\lambda(t)$ other than it be non-negative since negative probabilities have no meaning. If P is dependent only on t then $\lambda(t)$ defined as in (2.1) is the rate function for an inhomogeneous Poisson process. Such a rate is shown in Frame A of Figure 2.1. Frame B shows a point process realization from the rate in A. Note that spiking is more dense during periods where the rate is high, and sparse when it is low. Frame C shows the associated counting process. We can see that frame B and C convey they same information. This follows from the definition of $N(t)$, and in general $\mathcal{T}$ and $N(t)$

can be determined from each other.

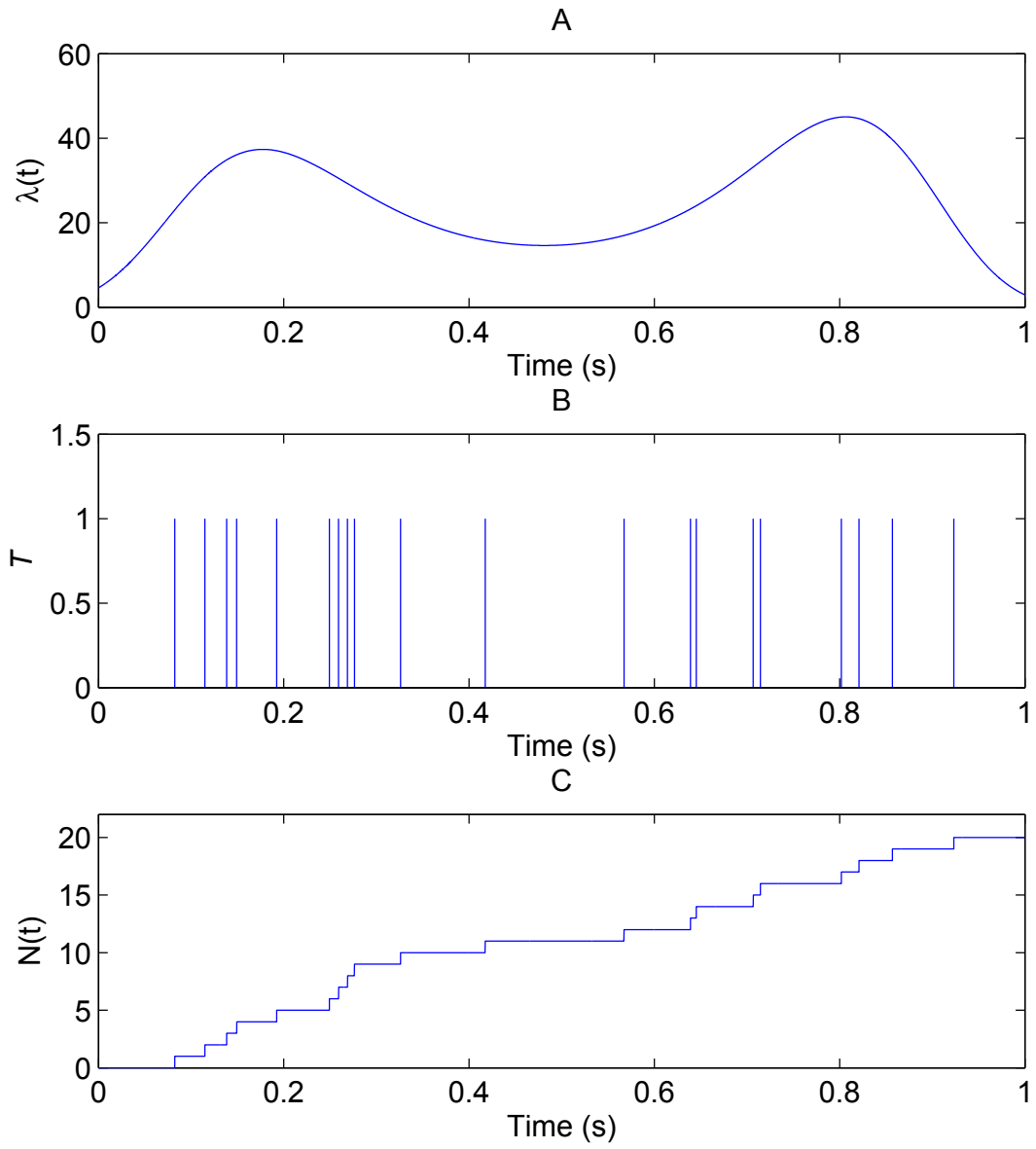

Fig. 2.1 Visualization of the relationship between $\lambda(t)$ (frame A), $\mathcal{T}$, (frame B) and $N(T)$ (frame C). Spikes are represented by impulses in the point process realization, their occurrence becoming more frequent over regions for which $\lambda(t)$ takes on larger values.

2.2 Excitation model

If spike sequences were modelled well by the Poisson process, then our goal when decoding might simply be to estimate $\lambda(t)$ the time varying firing rate of the measured cell. However, we have assumed that $\lambda(t)$ is history dependent and that the cells spiking output is influenced by the drive to its inputs, which we denote $\gamma(t)$, as well as its past spiking activity. Similarly to the models which include refractory effects mentioned in the introduction [17] [14], we propose that the instantaneous average rate $\lambda(t)$ is a product of both the input excitation (or free firing rate of preceding neurons as described in [14]) and some refractory function. The term free firing rate comes from the fact that $\gamma(t)$ is similar to $\lambda(t)$, except that it is free from refractory effects. In the remainder of this document we use the terms free firing rate and excitation synonymously.

In practice $\gamma(t)$ could take on virtually any form since its only restriction is that it must be non-negative. However, an estimator for a completely arbitrary non-negative function seems difficult to formulate. We recall from the literature review that this problem is usually solved by making further assumptions. One suggestion assumes the excitation is periodic [18], another that it is well represented by a set of basis functions [17]. In our work we propose the use of the exponential of a polynomial function defined as

$$\gamma(t) = \exp\left(\sum_{m=0}^{r} \alpha_m t^m\right), \tag{2.2}$$

to model the since this can fit any non-negative function [25] [19]. This is equivalent to representing $\log(\gamma(t))$ by a polynomial of order r. From Weierstass's polynomial approximation theorem we have that any continuous function on a closed interval can be uniformly approximated arbitrarily close by a polynomial [27]. Hence the exponential of polynomial

model is quite general, allowing the use of parametric estimation techniques resulting in relatively simple and well structured schemes.

2.3 Spike sequence model

History dependency is integrated into the estimator through modelling the spike data by a self exciting point process, as opposed to Poisson [12]. A self exciting point processes can be derived by letting the rate function (2.1) be conditioned on the spiking history up to time t. Define

$$\mu(t,0) = \lim_{\Delta\to 0} \frac{1}{\Delta t} Pr[N(t,t+\Delta t) = 1 | N(t)]$$
$$\mu(t,N(t);\mathcal{T}_{N(t)}) = \lim_{\Delta\to 0} \frac{1}{\Delta t} Pr[N(t,t+\Delta t) = 1 | N(t);\mathcal{T}_{N(t)}]$$

as in [12], where $\mathcal{T}_{N(t)}$ is a set containing the first $N(t)$ arrival times. The function $\mu(t,0)$ is free from history effects, however $\mu(t,N(t);\mathcal{T}_{N(t)})$ can depend arbitrarily on spike times occurring up to time t. The intensity of the self exciting point process modelling spike trains is defined as

$$\lambda(t) = \begin{cases} \mu(t,0), & t_0 < t \leq t_1 \\ \mu(t,N(t);\mathcal{T}_{N(t)}), & t_1 < t \leq T \end{cases} \tag{2.3}$$

where t_1 is the time of the first spike. Since $\lambda(t)$ is now conditioned on $\mathcal{T}$, it is itself a random variable.

We now define exactly how the instantaneous rate depends on past spiking history by mathematically describing the refractory effect. The refractory phenomenon is commonly classified by two distinct periods: absolute and relative. During the absolute period the cell is unable to fire due to its biophysical properties, and therefore we should never expect to see

two spikes occur over a period shorter than the absolute refractory period. Typically this period is on the order of a few milli-seconds [1]. To model this phenomenon mathematically we set $\lambda(t)$ of our self exciting process to immediately following a spike for a period of Δ_r. The variable Δ_r is the length of the absolute refractory period. Since $\lambda(t)$ is proportional to the probability of observing a spike, setting it to zero extinguishes any chance of a point being generated by the self exciting process.

The relative refractory phenomenon immediately follows the absolute period. During this time it is not impossible for the neuron to produce an spike, however, the chances are significantly diminished [1]. Over several milliseconds the refractory effects die out, and the chances of a spike occurring tend to return to those expected based on the cells excitation alone. Since we are multiplying the refractory function with $\gamma(t)$ we could represent this phenomenon mathematically by any function which takes the value zero at the end of the absolute period, and one at the end of the relative period. A suggestion similar to those found in [1] and [26] is

$$1 - \exp\left(-\beta\left(t - t_i - \Delta_r\right)\right), t \geq t_i + \Delta_r, \beta \geq 0, \tag{2.4}$$

where t_i is the time of the most recent spike. This causes the rate to smoothly transition back to normal as determined by the cell excitation $\gamma(t)$ over some time period determined by β. When β is very large, the effect of the relative period on spiking behaviour becomes negligible since under such conditions (2.4) resembles a unit step. Return to the free firing rate begins immediately after the absolute period Δ_r finishes, as such we shift the exponential scaling function by this amount. The selection of this form of the relative period is based on the fact that the theoretical inter-spike interval (ISI) distributions resemble those of real neural data [26].

To illustrate the refractory function of our model we have plotted the resulting $\lambda(t)$ immediately following a spike for a constant $\gamma(t)$ in Figure 2.2. We see that proceeding the spike $\lambda(t)$ is flat and equivalent to the free firing rate. However at $t = .005s$ the intensity drops to zero for a duration of Δ_r seconds. During this period it is impossible for the neuron to fire. The intensity then smoothly transitions back to its original level, at a rate determined by β.

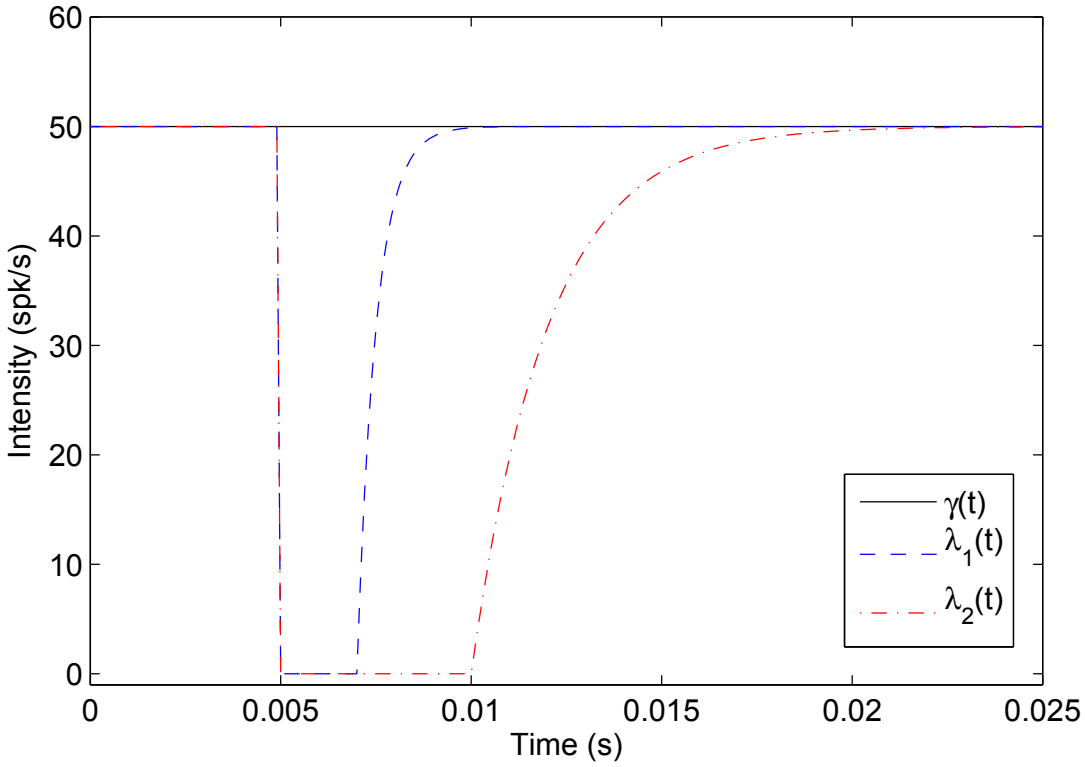

Fig. 2.2 Effect of the refractory function on the free firing rate $\gamma(t)$ assuming a spike has occurred at $t = .005$. A constant excitation is plotted with two intensities affected by different refractory periods. Intensity $\lambda_1(t)$ has a greater value of Δ_r and β than $\lambda_2(t)$. Larger values of Δ_r create longer dead times after a spike occurrence, while higher values of β cause a sharper return to the free firing rate.

Before the first measured spike we assume that the intensity process $\lambda(t)$ is free from

any history effects and is therefore fully determined by the free firing rate. Hence the dependency of the intensity (2.3) on the excitation $\gamma(t)$ including the refractory phenomenon is given by

$$\begin{aligned} \mu(t,0) &= \gamma(t) \\ \mu(t, N(t); \mathcal{T}_{N(t)}) &= \begin{cases} \gamma(t)\left[1 - \exp\left(-\beta\left(t - t_{N(t)} - \Delta_r\right)\right)\right], & \text{if } t_{N(t)} + \Delta_r \leq t < t_{N(t)+1} \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \tag{2.5}$$

Figure 2.3 further illustrates the relationship between $\gamma(t), \lambda(t)$ and $\mathcal{T}$ for our spike model. We can see that $\lambda(t)$ and $\gamma(t)$ are similar, except immediately following a spike. As prescribed by the model, $\lambda(t)$ goes to zero after a spike and then smoothly recovers to the rate dictated by $\gamma(t)$ in an exponential fashion. Although $N(t)$ is not shown in this figure, the rules for its construction are not changed by the added history dependence.

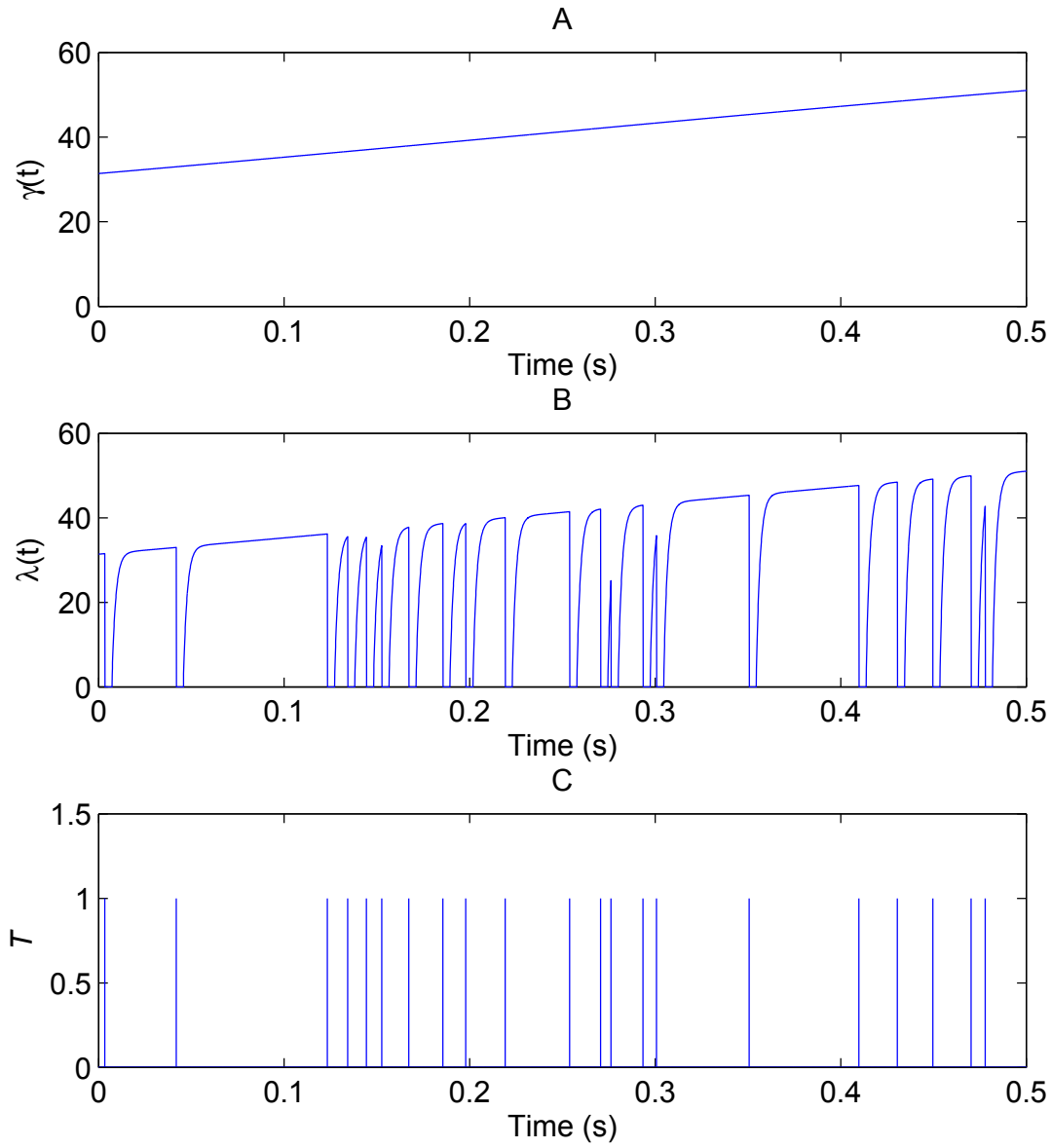

Fig. 2.3 Visualization of the relationship between $\gamma(t)$ (frame A), $\lambda(t)$ (frame B) and $\mathcal{T}$ (frame C). Following a spike $\lambda(t)$ the rate dictated by is reduced to zero, then smoothly returns to the rate prescribed by $\gamma(t)$.

The theoretical ISI distribution of a self exciting point process can be expressed in terms of $\lambda(t)$. We define $\tau_i = t_i - t_{i-1}, i = 2, 3, \ldots, N(T)$ then express the probability of the next spike taking longer than some time τ_0 to occur as the the probability that no spikes occur before τ_0. This is given by

$$\begin{aligned} Pr[\tau_i > \tau_0] &= Pr[N(t_{i-1}, t_i) = 0] \\ &= \exp\left(-\int_{t_{i-1}}^{t_{i-1}+\tau_0} \lambda(t)dt\right). \end{aligned}$$

Thus the cumulative density function (CDF) of τ_i is then

$$Pr[\tau_i < \tau_0] = 1 - \exp\left(-\int_{t_{i-1}}^{t_{i-1}+\tau_0} \lambda(t)dt\right), \tag{2.6}$$

and taking its derivative yields the probability density function (PDF)

$$p(\tau_i) = \lambda(t_{i-1}) \exp\left(-\int_{t_{i-1}}^{t_{i-1}+\tau_i} \lambda(t)dt\right). \tag{2.7}$$

If we assume that $\gamma(t)$ is constant with some value γ_0, then ISI distribution is the same for any $\tau_i, i = 2, 3, \ldots, N(T)$ and is given for our model as

$$p(\tau) = \begin{cases} \gamma_0 \left[1 - \exp\left(-\beta(\tau - \Delta_r)\right)\right] \exp\left(-\gamma_0 \int_0^\tau \left[1 - \exp\left(-\beta(\theta - \Delta_r)\right)\right] d\theta\right), & \tau \geq \Delta_r \\ 0, & \text{otherwise.} \end{cases} \tag{2.8}$$

The refractory effects on the ISI distribution is shown in Figure 2.4. Frame A shows a history independent ISI (Poisson) for three different values of γ_0. For high intensities the distribution greatly favours short inter-arrivals, while lower intensities give a more even distribution. Frame B shows the same rates as Frame A with added absolute refractory

effects. We can see this simply shifts the distribution to the right by Δ_r. This shows directly the impact of the absolute refractory period, as no inter-spike arrivals are less than Δ_r and agrees with our description of the absolute refractory period. Each distribution shows the effects of a different absolute period. Frame C shows a the distribution for the full model (2.8). Here we see the peaks of the distribution are smoothed and reduced. We see that larger values of β seem to have less affect, while smaller values greatly distort the ISI distribution. This implies that longer refractory periods do create an observable discrepancy in the statistics of spike trains when compared with Poisson data. In Chapter 5 we will see that the general shape of the distributions in frame C do in fact resemble empirical ISI distributions constructed from neural data as proposed by [26].

The refractory phenomenon makes reconstruction of the excitation somewhat complicated since $\gamma(t)$ is distorted by the refractory function. Most simply this can be seen when considering how the refractory period limits the peak firing rate of the neuron. For example a neuron with an absolute period of 1 ms could never fire at a rate more than 1000 spikes per second. As we will show in Chapter 5, the refractory effect causes a non-linear relationship between the excitation and the observed firing rate which is not present in a Poisson model. These effects can create errors if decoding is done using a Poisson model. By including the refractory period in the spiking model, the excitation can potentially be more accurately be recovered.

2.4 Likelihood equation

We now define the likelihood of a certain excitation given a certain observed spike train such that ML methods can be applied. From [12] we have that the likelihood of observing

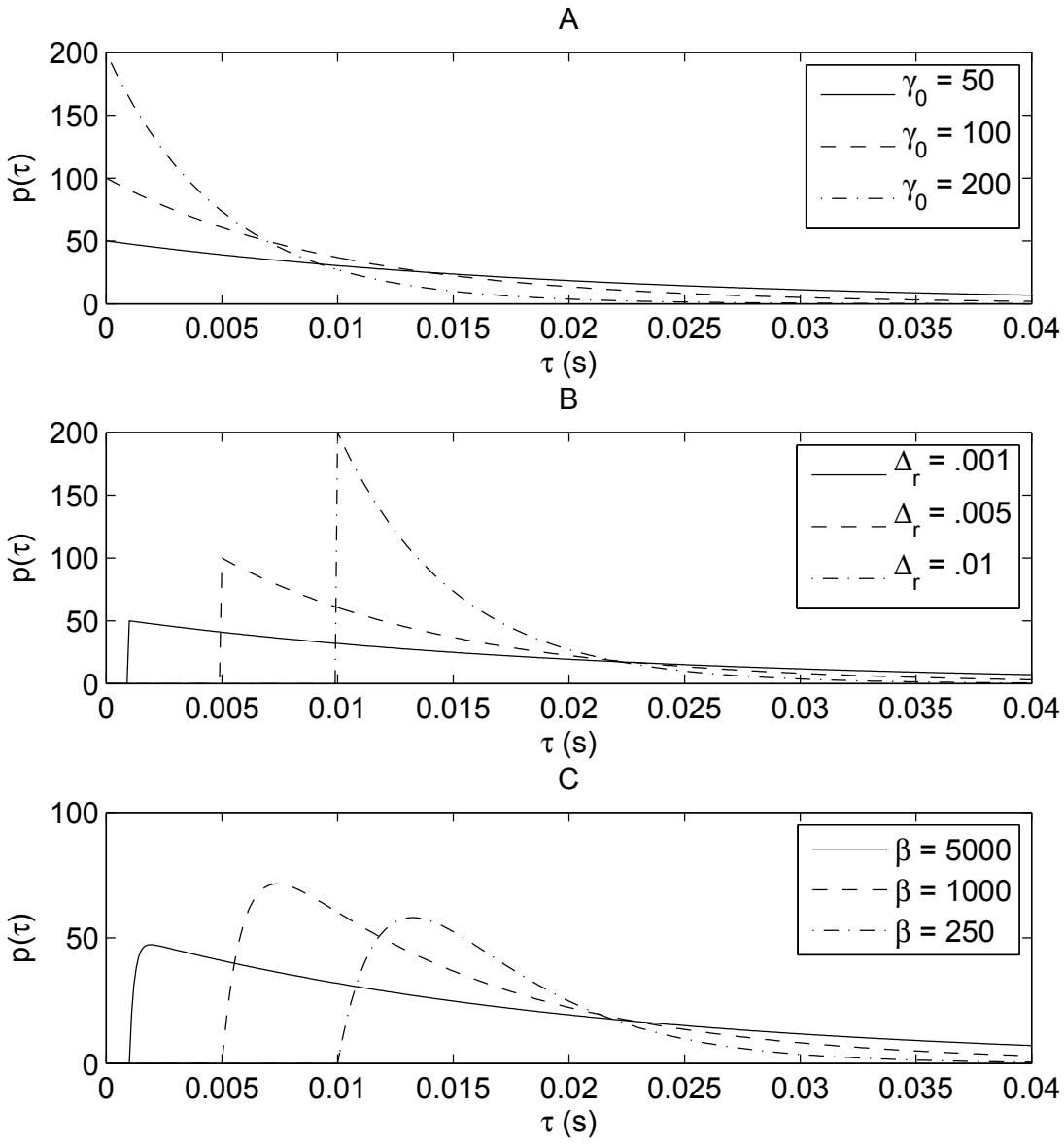

Fig. 2.4 Visualization of how the ISI distribution is affected by the addition of the refractory period. Frame A shows the ISI distribution of a Poisson process for three intensities. The frame B shows the same rates modified to include Δ_r. Frame C adds different values of β to the same distributions. Intensities and values for refractory parameters are given in the legends.

a particular spike sequence $\mathcal{T}_{N(T)}$ given $\lambda(t)$ for a self exciting point process is

$$p\left[\mathcal{T}_{N(T)}|\lambda(t)\right] = \begin{cases} \prod_{i=1}^{N(T)} \lambda(t_i) \times \exp\left(-\int_{t_0}^{T} \lambda(t)dt\right), & t_i - t_{i-1} \geq \Delta_r,\ i = 2, 3, ..., N(T) \\ 0, & \text{otherwise,} \end{cases} \tag{2.9}$$

where we have made explicit that the probability of observing a spike during the absolute refractory period is zero. The log-likelihood, when finite, is then

$$\mathcal{L} = \sum_{i=1}^{N(T)} \log \lambda(t_i) - \int_{t_0}^{T} \lambda(t)dt, t_i - t_{i-1} \geq \Delta_r, i = 2, 3, ..., N(T).$$

Using (2.2) (2.3) and (2.5) we have

$$\begin{aligned} \mathcal{L} = & \sum_{k=1}^{r} \alpha_k t_1^k + \sum_{n=2}^{N(T)} \left(\log\left[1 - \exp\left(-\beta\left(t_n - t_{n-1} - \Delta_r\right)\right)\right] + \sum_{k=0}^{r} \alpha_k t_n^k \right) \\ & - \int_0^{t_1} \exp\left(\sum_{k=0}^{r} \alpha_k t^k\right) dt - \sum_{n=1}^{N(T)-1} \int_{t_n+\Delta_r}^{t_{n+1}} \left[1 - \exp\left(-\beta\left(t - t_n - \Delta_r\right)\right)\right] \exp\left(\sum_{k=0}^{r} \alpha_k t^k\right) dt \\ & - \int_{t_{N(T)}+\Delta_r}^{T} \left[1 - \exp\left(-\beta\left(t - t_{N(T)} - \Delta_r\right)\right)\right] \exp\left(\sum_{k=0}^{r} \alpha_k t^k\right) dt \end{aligned} \tag{2.10}$$

if $t_i - t_{i-1} \geq \Delta_r, i = 2, 3...N(T)$ and $-\infty$ otherwise. We are interested in estimating $\gamma(t)$, however we can see in (2.10) that there are other parameters we must deal with, the absolute refractory period Δ_r, the relative refractory period parameter β, and the excitation polynomial order r. These issues are precisely the topic of the next chapter.

Chapter 3

Estimation techniques

Knowledge about Δ_r, β and r is required before we can proceed to maximize (2.10) with respect to the free firing rate parameters $\alpha_0, \alpha_1, ..., \alpha_r$. In this chapter we present methods for untangling the refractory and free firing rate coefficients as well as several approaches for selecting the model order of the estimate. The chapter concludes with the presentation and discussion of non-convex maximization procedures to numerically solve (2.10).

3.1 Refractory estimation

We suggest three general strategies which may be applied to find values for Δ_r and β in (2.10). The first and most simple strategy is to assume they take a plausible value. The second is to use an ensemble of single trial estimates to obtain a plausible value, and the third is simply to estimate them trial-by-trial for each spike train observation. We will see that there may be some computational advantages to the first and second but, since we are interested in accurate estimates from single realizations, the last strategy mentioned is our main focus.

3.1.1 A priori knowledge

The absolute refractory period is thought to last only a few milliseconds, while the relative refractory period might last as long as ten or twenty, but is typically also only a few milliseconds [1] [16]. This observation is supported by measured inter-spike arrival times, such as those presented in [31] [15], which show that that there are fewer inter-spike periods in the timescales consistent with the above. In light of this, it may be justifiable to simply substitute reasonable values for Δ_r and β into (2.10) and then maximize the likelihood function with respect to the free firing rate parameters. The substitution into Δ_r is direct. For example, if the assumed refractory period was 3ms, $\Delta_r = .003$. To determine β from a commonly accepted relative refractory length, we notice that from (2.5) that the effect of the relative refractory period decays after $5/\beta$ seconds, and therefore the substitution can be done by

$$\beta = \frac{5}{\tau}, \tag{3.1}$$

where τ is the accepted length of the relative refractory measured in seconds.

An immediate consequence of not needing to estimate β is that (2.10), as a function of the remaining parameters $(\alpha_0, \alpha_1..., \alpha_r)$, is concave. This property allows for a relatively quick search for the free firing rate parameters using basic optimization algorithms such as those in [32]. We prove (2.10) possesses this property by showing that it satisfies

$$\mathbf{x}^T \nabla^2 \mathcal{L} \mathbf{x} \leq 0.$$

and hence it is concave with respect to $\alpha_0, \alpha_1, ..., \alpha_r$. The entries of the gradient vector

$\nabla_i \mathcal{L}$ are

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \alpha_i} = &\sum_{n=1}^{N(T)} t_n^i - \int_{t_0}^{t_1} t^i \exp\left(\sum_{k=1}^{r} \alpha_k t^k\right) dt \\ &- \sum_{n=1}^{N(T)-1} \int_{t_n+\Delta_r}^{t_{n+1}} t^i \left[1 - \exp\left(-\beta\left(t - t_{N(T)} - \Delta_r\right)\right)\right] \exp\left(\sum_{k=0}^{r} \alpha_k t^k\right) dt \\ &- \int_{t_{N(T)}+\Delta_r}^{T} t^i \left[1 - \exp\left(-\beta\left(t - t_{N(T)} - \Delta_r\right)\right)\right] \exp\left(\sum_{k=0}^{r} \alpha_k t^k\right) dt, \end{aligned} \tag{3.2}$$

and the second order derivatives are given by

$$\begin{aligned}\frac{\partial^2 \mathcal{L}}{\partial \alpha_i \partial \alpha_j} = &- \int_{t_0}^{t_1} t^{i+j} \exp\left(\sum_{k=0}^{r} \alpha_k t^k\right) dt \\ &- \sum_{n=2}^{N(T)-1} \int_{t_n+\Delta_r}^{t_{n+1}} t^{i+j} \left[1 - \exp\left(-\beta\left(t - t_n - \Delta_r\right)\right)\right] \exp\left(\sum_{k=0}^{r} \alpha_k t^k\right) dt \\ &- \int_{t_{N(T)}+\Delta_r}^{T} t^{i+j} \left[1 - \exp\left(-\beta\left(t - t_{N(T)} - \Delta_r\right)\right)\right] \exp\left(\sum_{k=0}^{r} \alpha_k t^k\right) dt. \end{aligned} \tag{3.3}$$

Finally we have

$$
\begin{aligned}
\mathbf{x}^T \nabla^2 \mathcal{L} \mathbf{x} &= \sum_{i=0}^{r} x_i \sum_{j=0}^{r} x_j \nabla^2_{i,j} \mathcal{L} \\
&= -\int_{t_0}^{t_1} \sum_{i=0}^{r}\sum_{j=0}^{r} t^{i+j} \exp\left(\sum_{m=0}^{r} \alpha_m t^m\right) x_i x_j dt \\
&- \sum_{n=2}^{N(T)} \int_{t_{n-1}+\Delta_r}^{t_n} \sum_{i=0}^{r}\sum_{j=0}^{r} t^{i+j} \exp\left(\sum_{m=0}^{r} \alpha_m t^m\right) \left[1 - \exp\left(-\beta\left(t - t_n - \Delta_r\right)\right)\right] x_i x_j dt \\
&- \int_{t_{N(T)}+\Delta_r}^{T} \sum_{i=0}^{r}\sum_{j=0}^{r} t^{i+j} \exp\left(\sum_{m=0}^{r} \alpha_m t^m\right) \left[1 - \exp\left(-\beta\left(t - t_{N(T)} - \Delta_r\right)\right)\right] x_i x_j dt \\
&= -\int_{t_0}^{t_1} \exp\left(\sum_{m=0}^{r} \alpha_m t^m\right) \sum_{i=0}^{r} x_i t^i \sum_{j=0}^{r} x_j t^j dt \\
&- \sum_{n=2}^{N(T)} \int_{t_{n-1}+\Delta_r}^{t_n} \exp\left(\sum_{m=0}^{r} \alpha_m t^m\right) \left[1 - \exp\left(-\beta\left(t - t_n - \Delta_r\right)\right)\right] \sum_{i=0}^{r} x_i t^i \sum_{j=0}^{r} x_j t^j dt \\
&- \int_{t_{N(T)}+\Delta_r}^{T} \exp\left(\sum_{m=0}^{r} \alpha_m t^m\right) \left[1 - \exp\left(-\beta\left(t - t_{N(T)} - \Delta_r\right)\right)\right] \sum_{i=0}^{r} x_i t^i \sum_{j=0}^{r} x_j t^j dt \\
&= -\int_{t_0}^{t_1} \underbrace{\exp\left(\sum_{m=0}^{r} \alpha_m t^m\right)}_{>0} \underbrace{\left(\sum_{i=0}^{r} x_i t^i\right)^2}_{\geq 0} dt \\
&- \sum_{n=2}^{N(T)} \int_{t_{n-1}+\Delta_r}^{t_n} \underbrace{\exp\left(\sum_{m=0}^{r} \alpha_m t^m\right) \left[1 - \exp\left(-\beta\left(t - t_n - \Delta_r\right)\right)\right]}_{>0} \underbrace{\left(\sum_{i=0}^{r} x_i t^i\right)^2}_{\geq 0} dt \\
&- \int_{t_{N(T)}+\Delta_r}^{T} \underbrace{\exp\left(\sum_{m=0}^{r} \alpha_m t^m\right) \exp\left[1 - \exp\left(-\beta\left(t - t_{N(T)} - \Delta_r\right)\right)\right]}_{>0} \underbrace{\left(\sum_{i=0}^{r} x_i t^i\right)^2}_{\geq 0} dt \leq 0
\end{aligned}
$$

3.1.2 Absolute period from the data

Another option is to try and estimate the refractory parameters from the spike data without estimating the excitation. As far as Δ_r is concerned, we can profile out the refractory

parameter using an approach described in [33] known as profile likelihood. This essentially comes down to a multi-step maximum likelihood approach, the first being maximization with respect to Δ_r. The profile likelihood with respect to Δ_r is given by

$$\mathcal{L}_{\gamma,\beta} = \sup_{\Delta_r} \mathcal{L}(\gamma(t), \Delta_r, \beta). \tag{3.4}$$

Since the $\gamma(t)$ must be non-negative, we are able to find the value of Δ_r which satisfies the least upper bound in (3.4) without knowing $\alpha_0, \alpha_1, ..., \alpha_r, r$ or β. We denote this value $\hat{\Delta}_r$.

$$\begin{aligned}\hat{\Delta}_r &= \arg\max_{\Delta_r} \mathcal{L}(\gamma(t), \Delta_r, \beta) \\ &= \arg\min_{\Delta_r} \Bigg(\sum_{i=2}^{N(T)} \int_{t_{i-1}+\Delta_r}^{t_i} \underbrace{\gamma(t)\left[1 - \exp\left(-\beta(t - t_i - \Delta_r)\right)\right]}_{\geq 0} dt \\ &\quad + \int_{t_{N(T)}+\Delta_r}^{T} \underbrace{\gamma(t)\left[1 - \exp\left(-\beta(t - t_i - \Delta_r)\right)\right]}_{\geq 0} dt \Bigg).\end{aligned} \tag{3.5}$$

To minimize (3.5) we need to maximize Δ_r, while satisfying $\Delta_r \leq t_i - t_{i-1}$, $i = 2, 3, ..., N(T)$. Hence, $\hat{\Delta}_r = \min(t_i - t_{i-1})$, $i = 2, 3, ..., N(T)$. This result also turns out to be the ML estimate of Δ_r.

A similar profiling approach cannot be used for β since maximization of (2.10) with respect to β is dependant on the free firing rate parameter values. Instead it must be found in parallel with $\alpha_0, \alpha_i, ..., \alpha_r$, further details on such trial-by-trial joint maximization are given in Section 3.3.

3.1.3 Pseudo ML

ML estimators, which achieve efficiency asymptotically, may not perform well when the estimate is based on a small sample of data [34]. This implies that our trial-by-trial estimate

for the refractory parameters (and excitation) might be poor when the number of spikes observed is relatively low. In the context of neural rate estimation this situation could present itself when either the intensity of the response is low or the observation interval is short. Noting that the result of (3.5) does not depend on the the excitation, one could use a set of spike times from different trials, and even different excitations, in order to better estimate the absolute refractory period of a particular neuron.

Under the assumption that trial-to-trial results are independent, the likelihood of observing a set of K responses $\{\mathcal{T}_j, j = 1, 2, ..., K\}$ given a corresponding set of K excitations $\{\gamma_j(t), j = 1, 2, ..., K\}$ is the product of the individual trial likelihoods. It can be shown in a similar way to (3.5) that the this product is maximized with respect to Δ_r by choosing $\hat{\Delta}_r$ such that $\hat{\Delta}_r = \min(t_{i,j} - t_{i-1,j})$, $i = 2, 3, ..., N(T), j = 1, 2, ..., K$, that is the minimum inter-spike interval observed over all responses in $\{\mathcal{T}_j, j = 1, 2, ..., K\}$.

The same approach cannot be directly applied to the relative refractory period. However, since there is little evidence that the refractory phenomenon of a single neuron varies much over long timescales [15], a method which combines several single trial estimates of β could improve results even if the excitation was not the same in every trial. It is not clear what sort of scheme should be used to combine the single trial estimates. If they are sufficiently close to the true value, then one suggestion may be to take the average. However, since β affects the likelihood in a non-linear way we suspect a slightly different combination may give better results.

We have from (3.1) that the rate $\lambda(t)$ is effected for a time period inversely proportional to the value of β. In light of this we suggest first inverting the trial-to-trial estimates of β,

then averaging the inverted summation as follows

$$\beta = \frac{\left(\sum_{j=1}^{K} \beta_j^{-1}\right)^{-1}}{K}, \tag{3.6}$$

where K is the number of trials and β_j represents the single trial estimate of β from trial j. The effectiveness on any averaging scheme will depend on how the single trial estimates of β are distributed. It is not clear from the likelihood equation what sort of distribution we should expect. Conclusions on this approach are left to empirical testing, results of which are found in Chapter 4.

3.2 Model order selection

Model order selection is a common task in the field of parametric estimation. In our problem we wish to determine the value of r to use in our final free firing rate estimate. Four decision tools seem most prominent in the literature: Akaike's Information Criterion (AIC), its bias corrected version (AICC), Bayesian Information Criterion (BIC) and The Minimum Discription Length Principle (MDL) [35] [36]. All of these rules share one fundamental concept, they choose the model which minimizes some approximation to the Kullback Leibler (KL) distance between the true PDF of the data and the likelihood of the data given a particular parametric model.

3.2.1 Kullback Leibler distance

The KL discrepancy function, denoted D_{KL}, can be viewed as a measure of distance or difference between two PDFs [37]. For continuous distributions it is defined as

$$D_{KL}(f(x), g(x)) = \int_x f(x) \log \frac{f(x)}{g(x)} dx \tag{3.7}$$

where $f(x), g(x)$ are two valid PDFs. We can interpret D_{KL} as a measure of inefficiency. It represents the information lost when $g(x)$ is used in lieu of $f(x)$ [37]. If we replace $f(x)$ with the PDF of observing a certain spike sequence from the true excitation $p_0(\mathcal{T}|\gamma_0)$, and $g(x)$ with the likelihood of observing $\mathcal{T}$ from some parametric approximation to the excitation $\hat{p}(\mathcal{T}|\gamma(\theta))$ we get

$$D_{KL}(p_0(\mathcal{T}|\gamma_0), \hat{p}(\mathcal{T}|\gamma(\theta))) = \int_{\mathcal{T}} p_0(\mathcal{T}|\gamma_0) \log \frac{p_0(\mathcal{T}|\gamma_0)}{\hat{p}(\mathcal{T}|\gamma(\theta))} d\mathcal{T}, \tag{3.8}$$

where we define the integral notation with respect to $\mathcal{T}$ as

$$\int_{\mathcal{T}} f(\mathcal{T}) d\mathcal{T} = \int_{t_1} \cdots \int_{t_{N(T)}} f(\mathcal{T}) dt_1 ... dt_{N(T)}.$$

We can see from (3.8) that inefficiency is minimized when $p_0(\mathcal{T}|\gamma_0) = \hat{p}(\mathcal{T}|\gamma(\theta))$. Intuitively this should be the case, as no information is lost when the true PDF is substituted with itself. More formally

$$D_{KL}(p_0(\mathcal{T}|\gamma_0), \hat{p}(\mathcal{T}|\gamma(\theta))) \geq 0 \tag{3.9}$$

with equality if and only if $p_0(\mathcal{T}|\gamma_0) = \hat{p}(\mathcal{T}|\gamma(\theta))$. We note that (3.8) can be expressed in terms of expectation as

$$\begin{aligned} D_{KL}(p_0(\mathcal{T}|\gamma_0), \hat{p}(\mathcal{T}|\gamma(\theta))) &= E_0\left[\log\left(\frac{p_0(\mathcal{T}|\gamma_0)}{\hat{p}(\mathcal{T}|\gamma(\theta))}\right)\right] \\ &= E_0[\log p_0(\mathcal{T}|\gamma_0)] - E_0[\log \hat{p}(\mathcal{T}|\gamma(\theta))] \end{aligned} \tag{3.10}$$

where E_0 refers to expectation with respect to the true data distribution.

From (3.9) a model selection process arises naturally. Competing models could be compared by computing D_{KL} between their likelihoods and the true PDF of the data. The model which minimizes the distance is then chosen as the final estimate. Unfortunately the true distribution is not available. However, under different assumptions, several approximations to (3.9) have been suggested. Different approaches have led to different criteria. As such, short derivations based largely on those found in [35] and [38] are presented to give insight into whether or not these assumptions hold in our problem.

3.2.2 Akaike's Information Criterion

Akaike's Information Criterion was introduced in the mid 1970's by Hirotsugu Akaike as "An Information Criterion", however, it is more commonly referred to by this sections name. The criterion is formulated based on the following assumptions. First, under regularity conditions, the ML estimate is asymptotically efficient [34]. Second, the value of a parameterized likelihood function tends to be quite sensitive to small deviations in its parameters from the true value [39]. To see the second assumption consider the observations $\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_K$ as independent spike sequence realizations with PDF $p_0(\mathcal{T}|\gamma_0)$. Let the likelihood of a parameterized version of γ_0 be $\hat{p}(\mathcal{T}|\gamma(\theta)$. The average log-likelihood of this

parametric version for the K realizations is

$$\frac{1}{K}\sum_{i=1}^{K}\log\hat{p}(\mathcal{T}_i|\gamma(\theta)),$$

as K goes to infinity this sum goes with probability one to the expectation

$$\int_{\mathcal{T}} p_0(\mathcal{T}|\gamma_0)\log\hat{p}(\mathcal{T}|\gamma(\theta))d\mathcal{T},$$

which as $N(T)$ grows is asymptotically sensitive to differences between $\hat{p}(\mathcal{T}|\gamma(\theta))$ and $p_0(\mathcal{T}|\gamma_0)$ [39]. Thus both assumptions used hold at least in the asymptotic sense.

In (3.10), only the second expectation is dependant on the parametric model in question. If our goal is to compare several model likelihoods to the same true PDF, then the first expectation in (3.10) has the same value in each comparison. Therefore we can minimize D_{KL} by maximizing $E_0[\ln\hat{p}(\mathcal{T}|\gamma(\theta)]$

$$\min_{n\in N} D_{KL}\left[p_0(\mathcal{T}|\gamma_0),\hat{p}(\mathcal{T}|\gamma(\theta^n))\right] \rightarrow \max_{n\in N} E_0[\log\hat{p}(\mathcal{T}|\gamma(\theta^n))] \tag{3.11}$$

$$= \max_{n\in N}\int_{\mathcal{T}} p_0(\mathcal{T}|\gamma_0)\log\hat{p}(\mathcal{T}|\gamma(\theta^n))d\mathcal{T}, \tag{3.12}$$

where n represents the number of parameters in θ and N the set of all model sizes considered. Since neither distribution in (3.12) is available, the following procedure is suggested to approximate them. The remainder of the derivation from [39] is similar to that in [35], which we highlight key points from here.

Consider two independent spike trains: an observed set $\mathcal{T}_y$, and a fictitious set $\mathcal{T}_x$, both generated from the same excitation. Let $\hat{\theta}_y^n$ and $\hat{\theta}_x^n$ be the ML estimates of the model parameters of $\gamma(\theta^n)$ given data sets $\mathcal{T}_y$ and $\mathcal{T}_x$ respectively. An approximation of the true

data PDF using a cross validation approach might be

$$\log p_0(\mathcal{T}|\gamma_0) \approx \log \hat{p}(\mathcal{T}_y|\gamma(\hat{\theta}_x^n)) \approx E_x[\log \hat{p}(\mathcal{T}_y|\gamma(\hat{\theta}_x^n))],$$

and therefore (3.12) can be approximated by

$$E_0[\log \hat{p}(\mathcal{T}_y|\gamma(\hat{\theta}_x^n))] \approx E_y\left[E_x[\log \hat{p}(\mathcal{T}_y|\gamma(\hat{\theta}_x^n))]\right] \tag{3.13}$$

where E_x, E_y represent expectation with respect to $\mathcal{T}_x$ and $\mathcal{T}_y$. Since $\mathcal{T}_x$ was fictitious we cannot estimate $\hat{\theta}_x^n$, however, $\hat{p}(\mathcal{T}_y|\gamma(\hat{\theta}_x^n))$ in (3.13) can be approximated by taking a second order Taylor series expansion about $\hat{\theta}_y^n$.

$$\begin{aligned}
\log \hat{p}(\mathcal{T}_y|\gamma(\hat{\theta}_x^n)) &\approx \log \hat{p}(\mathcal{T}_y|\hat{\theta}_y^n) + (\hat{\theta}_x^n - \hat{\theta}_y^n)^T \left[\frac{\partial \log \hat{p}(\mathcal{T}_y|\gamma(\theta^n))}{\partial \theta^n}\bigg|_{\theta^n=\hat{\theta}_y^n}\right] \\
&+ \frac{1}{2}(\hat{\theta}_x^n - \hat{\theta}_y^n)^T \left[\frac{\partial^2 \log \hat{p}(\mathcal{T}_y|\gamma(\theta^n))}{\partial \theta^n \partial \theta^{nT}}\bigg|_{\theta^n=\hat{\theta}_y^n}\right](\hat{\theta}_x^n - \hat{\theta}_y^n) \\
&\approx \log \hat{p}(\mathcal{T}_y|\gamma(\hat{\theta}_y^n)) - \frac{1}{2}(\hat{\theta}_x^n - \hat{\theta}_y^n)^T J_y(\hat{\theta}_x^n - \hat{\theta}_y^n),
\end{aligned} \tag{3.14}$$

where J_y is the information matrix associated with the observation $\mathcal{T}_y$. The second term of the expansion evaluates to zero since the slope of the log-likelihood is zero at the ML estimate. Substituting (3.14) back into (3.13) yields

$$\begin{aligned}
&E_y\Big[E_x[\log \hat{p}(\mathcal{T}_y|\gamma(\hat{\theta}_y^n)) - \frac{1}{2}(\hat{\theta}_x^n - \hat{\theta}_y^n)^T J_y(\hat{\theta}_x^n - \hat{\theta}_y^n)]\Big] \\
&\qquad = E_y\Big[\log \hat{p}(\mathcal{T}_y|\gamma(\hat{\theta}_y^n))\Big] - \frac{1}{2}E_y\Big[E_x[(\hat{\theta}_x^n - \hat{\theta}_y^n)^T J_y(\hat{\theta}_x^n - \hat{\theta}_y^n)]\Big].
\end{aligned} \tag{3.15}$$

Since we don't know the distribution of $\mathcal{T}_y$, an asymptotically unbiased estimate of the first expectation in (3.15), owing to the properties of ML estimators, is simply $\log \hat{p}(\mathcal{T}_y|\gamma(\hat{\theta}_y^n))$.

The second expectation is shown to evaluate to $2n$ in [35]. Therefore minimizing D_{KL} is equivalent to maximizing

$$\log \hat{p}(\mathcal{T}|\gamma(\hat{\theta}^n)) - n \tag{3.16}$$

Under these asymptotic assumptions the probability of under-fitting (choosing a model order too low, when the true model is more suitably captured by some higher order model) goes to zero, and the probability of over-fitting goes to a constant as the size of the data $N(T)$ goes to infinity [35]. Previous results for a Poisson data estimator showed that the NMISE caused by under-fitting is much worse than that caused by over-fitting with our chosen free firing rate model (2.2) [19]. In Chapter 4 we present evidence this is also true for our current problem. We conclude that such a property may be desirable.

Conversely, asymptotic assumptions do not necessarily favour AIC for since $N(T)$ is far from infinite, and may at times be quite small. To remove inherent bias in AIC when $N(T)$ is finite, a slightly different version is derived in [40]. This correction to AIC known as AICC and suggests to choose the candidate model which maximizes

$$\log \hat{p}(\mathcal{T}|\gamma(\hat{\theta}^n)) - \frac{nN(T)}{N(T) - n - 1},$$

which converges to 3.16 as $N(T)$ goes to infinity.

3.2.3 Bayesian Information Criterion

The Bayesian Information Criterion attempts to estimate a prior on the parameter vector in order to evaluate (3.12). Two assumptions are made in its derivation. First, the PDF of the parameter vector $p(\theta)$ is flat around the ML estimate $\hat{\theta}$, and second that $p(\theta)$ is independent of $N(T)$ [35]. To define the prior, a fictitious realization $\mathcal{T}$ is used with an

arbitrarily large size $N(T)$. The BIC then approximates the true PDF of the data as

$$\begin{aligned}\hat{p}(\mathcal{T}|\gamma_0) &= E_\theta[\hat{p}(\mathcal{T}|\gamma(\theta^n))] \\ &= \int_\theta p(\theta^n)\hat{p}(\mathcal{T}|\gamma(\theta^n))d\theta, \end{aligned} \tag{3.17}$$

where we define the integral notation with respect to the θ vector as

$$\int_\theta f(\theta)dtheta = \int_{\theta_1} \cdots \int_{\theta_n} f(\theta)d\theta_1 ... d\theta_n.$$

We note that (3.17) is asymptotically sensitive to differences between γ_0 and $\gamma(\theta^n)$. This means that if the parametric model $\gamma(\theta)$ cannot be exactly matched to γ_0, then this approximation may be a poor choice. Hence even asymptotically BIC might only be optimal if one of the parametric models fully captures the true excitation [35].

To evaluate (3.17) we substitute $\hat{p}(\mathcal{T}|\gamma(\theta^n))$ with its second order Taylor series approximation about the ML estimate, which results in

$$\hat{p}(\mathcal{T}|\gamma(\theta^n)) \approx \hat{p}(\mathcal{T}|\gamma(\hat{\theta}^n)) \exp\left(-\frac{1}{2}(\hat{\theta}^n - \theta^n)^T \hat{J}(\hat{\theta}^n - \theta^n)\right).$$

Substituting this back into (3.17)

$$\hat{p}(\mathcal{T}|\gamma_0) \approx \int_\theta p(\theta^n)\hat{p}(\mathcal{T}|\gamma(\hat{\theta}^n)) \exp\left(-\frac{1}{2}(\hat{\theta}^n - \theta)^T \hat{J}(\hat{\theta}^n - \theta^n)\right) d\theta \tag{3.18}$$

$$\approx \hat{p}(\mathcal{T}|\gamma(\hat{\theta}^n))p(\hat{\theta}^n) \int_\theta \exp\left(-\frac{1}{2}(\hat{\theta}^n - \theta^n)^T \hat{J}(\hat{\theta}^n - \theta^n)\right) d\theta$$

$$\approx \frac{\hat{p}(\mathcal{T}|\gamma(\theta^n)p(\hat{\theta}^n)(2\pi)^{\frac{n}{2}}}{|\hat{J}|^{\frac{1}{2}}} \tag{3.19}$$

where $|\hat{J}|$ is the determinant of the information matrix evaluated at the ML parameter

estimate. We have made use of the assumption that $p(\theta)$ is flat around the ML estimate $\hat{\theta}$ such that $p(\theta) \approx p(\hat{\theta})$, and that $\hat{p}(\mathcal{T}|\gamma(\theta))$ is much larger at the ML estimate $\hat{\theta}$ than at any other θ to get from (3.18) to (3.19) [35]. Substituting (3.19) into the the model dependant term of (3.10) results in

$$\log \hat{p}(\mathcal{T}|\gamma(\hat{\theta})) + \log p(\hat{\theta}) + \frac{n}{2}\log 2\pi - \frac{1}{2}\log|\hat{J}| \approx \log \hat{p}(\mathcal{T}|\gamma(\hat{\theta}^n)) - \frac{n}{2}\log N(T) \qquad (3.20)$$

This approximation comes from the result that $|\hat{J}|$ increases approximately as $n \log N(T)$ for large $N(T)$ [35] and the previously mentioned assumption that the distribution $p(\theta)$ is independent of $N(T)$. Thus the model which maximizes the right hand side of (3.20) is the model chosen by BIC. As with AIC, BIC is derived on asymptotic principles. It is unclear form the derivation how large $N(T)$ will need to be to make BIC reliable.

3.2.4 Minimum Description Length

The philosophy behind MDL is to select the model which compresses the observed data the most [36]. More precisely, the model which explains the data in the fewest number of bits is the model chosen [38]. The first practical formulations of MDL were presented by J. Rissanen and refined over several papers in the late 1970s within which he applied the philosophy to particular problems [38] [41] [42]. Prior to this, the idea was based on minimizing Kolmologrov complexity, the shortest computer code to generate a particular data sequence [36]. Since there is no universal algorithm for doing this, MDL had to become more refined to be useful.

MDL can be derived from Gibb's Theorem. Let p_i and q_i be non-negative real numbers for $i = 1, 2, \ldots, n$ such that

$$\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} q_i$$

then we have

$$-\sum_{i=1}^{n} p_i \log p_i \leq -\sum_{i=1}^{n} p_i \log q_i \tag{3.21}$$

$$0 \leq \sum_{i=1}^{n} p_i \log p_i - \sum_{i=1}^{n} p_i \log q_i \tag{3.22}$$

which achieves equality if and only if $p_i = q_i$ for all i [38]. Let's assume that $\sum p_i = \sum q_i = 1$ such that p_i and q_i are valid probability mass distributions (PMFs) of teh random variable P and Q. Shifting both terms to the right hand side of the inequality as in (3.21) yields the discrete version of the Kullback Leibler distance. MDL treats Gibb's theorem as a size inequality, referring to the length of codes used to describe the random variables P and Q [38]. Consider a source encoder, represented by a models data distribution, which assigns binary codewords $C(a_i)$ of length l_i to source symbols $a_1, a_2, \ldots, a_N$. For an observed data string composed of concatenated codewords, the per symbol length i approaches the right hand side of (3.21) as the length of the sting goes to infinity [38]. The Kraft inequality states that a binary code is uniquely decodable if and only if

$$\sum_{i=1}^{N} 2^{-l_i} \leq 1$$

where N is the size of the set of all codewords [37]. Equality is reached when

$$l_i = \log_2 \frac{1}{q_i}, \text{ for all } i,$$

implying that the minimum code length is achieved only when $p_i = q_i$ for all i. That is, when the model and true PDF are equal.

If we assume that each parameter in θ is of the same accuracy (same number of bits),

then MDL should lead us to the model which best describes the data with the fewest parameters. Taking this approach Rissenan concluded that the following criterion should be maximized

$$\log p(\mathcal{T}|\gamma(\hat{\theta}^n)) - \frac{n}{2}\log N(T).$$

This result is identical to that found for BIC (3.20). In [35] it is suggested that this may imply some fundamental characteristic about BIC (or MDL) that has not yet been discovered. Rissanen states that despite the similarity, MDL is more principled and derived under a worst case (uniform) prior on the data, as opposed to the mathematically convenient prior chosen in BIC [38].

3.2.5 Criterion summary

All three unique selection rules are summarized in the context of our problem (3.23) (3.24) (3.25). They each combine the log-likelihood of observing the spike response given ML parameter estimates (2.10) with a penalty function based on the number of parameters in the model, and in some cases the amount of data observed,

$$\text{AIC: } \mathcal{L}_r - (r+1) \tag{3.23}$$

$$\text{AICC: } \mathcal{L}_r - \frac{N(T)(r+1)}{N(T)-r} \tag{3.24}$$

$$\text{BIC: } \mathcal{L}_r - (r+1)\log N(T) \tag{3.25}$$

where $\mathcal{L}_r$ refers to (2.10) evaluated for a particular model order r and $N(T)$ is the number of spikes observed in the response from one trial. The relationship between n in the derivations from the previous sections and r our model order is simply $r = n - 1$ since an $r'th$ order polynomial has $r+1$ coefficients. The penalty terms of each criterion are plotted in Figure

3.1. AIC and BIC are linear with r, however BIC penalizes high order models more heavily, as can be inferred from its steeper slope. Interestingly, BIC penalizes complex models more as $N(T)$ grows, while AICC shows the opposite trend. Also, at $N(T) = 100$, AICC has practically converged to AIC, thus for intense spike trains we might not expect much difference in performance between them.

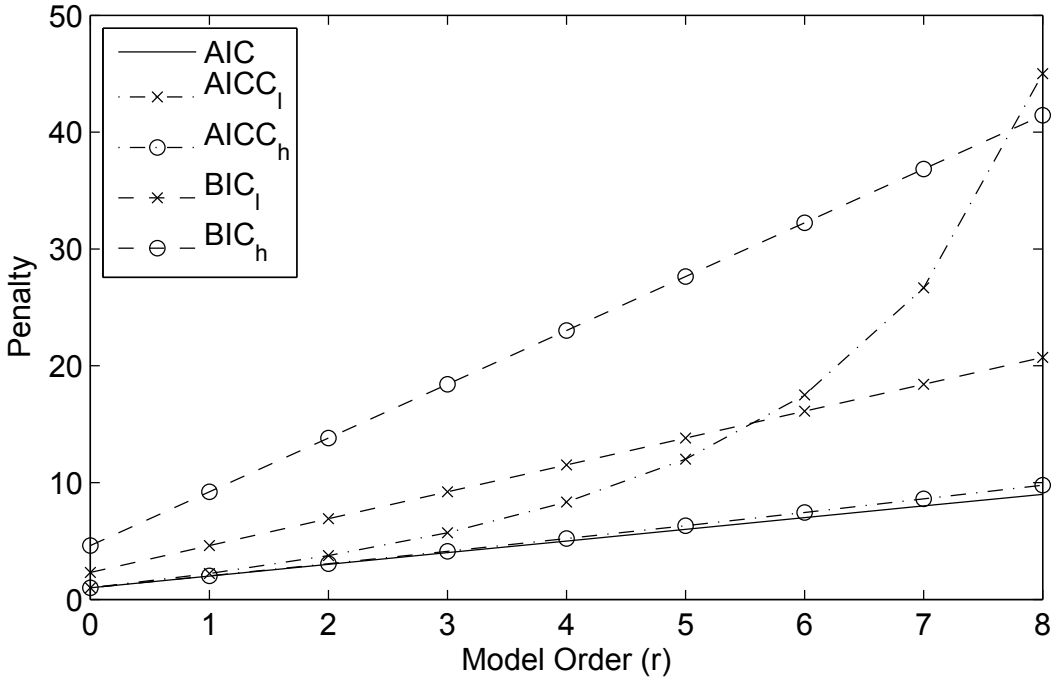

Fig. 3.1 Magnitude of the penalty term in AIC (3.23), AICC (3.24) and BIC (3.25) as a function of r. Penalty terms dependent on $N(T)$ are shown twice, subscript l shows the respective penalty when $N(T) = 10$, subscript h implies $N(T) = 100$.

Our model selection process works as follows. A set of models $\mathcal{R}$ is initially selected such that each model has a different value $r = 0, 1, ..., R$. The ML estimate is found for the parameters of each free firing rate model. The one that maximizes the respective criterion is the final model chosen by that rule. We carried out an empirical study in Chapter 4 in order to infer the most accurate criterion.

3.3 Excitation estimation

This section presents a method for finding the free firing rate coefficients, and β if it is to be inferred from the data on a trial-to-trial basis. When combined with one of the previously presented model order selection rules, the final free firing rate estimate can be found. If β is assumed to take some a priori value then only one local search need be carried out to find the ML estimate, owing to the concavity of (2.10). In this situation we suggest use of a simple Newton-Raphson search as described in [32], using the seed in section 3.3.3 as a starting point. The parameter Δ_r can be assumed known or easily estimated as per Section 3.1.2. In the more interesting situation where β is to be inferred from the data, a global maximization procedure must be used since the log-likelihood function (2.10) is no longer concave. We suggest first a robust global maximization method, and then a slightly more heuristic approach which may reduce computation time.

3.3.1 Global maximization

A search method that can cope with possible multiple local maxima and non-cancavity in the log-likelihood surface can be implemented using a filled function approach [43]. By applying a transformation to the log-likelihood we can either validate a local maximum as global, or suggest a seed which is guaranteed to lead to a better maximizer. Many filled function approaches have been proposed in the literature such as [44] [45] [43]. In our work we have employed the general filled function of [43] and adapted it for global maximization of the log-likelihood function $\mathcal{L}$ with respect to excitation coefficients and refractory parameter β. These variables constitute our parameter vector θ.

$$\theta = [\alpha_0, \alpha_1, ..., \alpha_r, \beta]^T$$

The algorithm given in [43] can be straight forwardly modified to maximize $\mathcal{L}$ instead of minimizing an arbitrary objective function. This results in the following slightly modified filled function definition. A function $P(\cdot)$ qualifies itself as a filled function if it possesses the following three properties:

1. θ^* is a strictly local maximizer of $P(\theta, \theta^*)$.

2. For any $\theta \neq \theta^*$ satisfying $\mathcal{L}(\theta) \leq \mathcal{L}(\theta^*), \theta$ is not a stationary point of $P(\theta, \theta^*)$, and satisfies $\nabla^T P(\theta, \theta^*)(\theta - \theta^*) < 0$, where ∇ is the gradient operator.

3. If θ^* is not a global maximizer of $\mathcal{L}(\theta)$ then there must exist a $\bar{\theta} \in \Omega$ such that $\bar{\theta}$ is a stationary point of $P(\theta, \theta^*)$, where Ω is the parameter space.

Any function $P(\cdot)$ satisfying the above conditions allows us to determine whether or not θ^*, a local maximum of $\mathcal{L}$, is also the global maximum. This is because a local maximum qualifies itself as global if and only if θ^* is both the only stationary point of, and a strict maximizer of $P(\theta, \theta^*)$ [43]. Put differently, if we can find a stationary point at any $\theta \neq \theta^*$ in $P(\theta, \theta^*)$ then it must satisfy $\mathcal{L}(\theta) > \mathcal{L}(\theta^*)$, and can subsequently be used as a seed for another local search in $\mathcal{L}$ to find a better maximizer. If instead we cannot find any stationary points $\theta \neq \theta^*$ in $P(\theta, \theta^*)$, then we assume θ^* is the global maximum of $\mathcal{L}$. The filled function we employ is explicitly given by

$$P\left(\theta, \theta^*, s, A\right) = \exp\left(-A\|\theta - \theta^*\|^2\right) g\left(\mathcal{L}(\theta^*) - \mathcal{L}(\theta), s\right) \tag{3.26}$$

where θ is the parameter vector, θ^* is a local maxima of $\mathcal{L}(\theta)$ the log-likelihood function (2.10), A and s are positive constants used to alleviate possible numerical issues, and $g(\cdot)$

is defined by

$$g(t,s) = \begin{cases} c, & t \geq 0 \\ h(t,c), & -s < t < 0 \\ 0, & t \leq -s \end{cases}$$

for any $s > 0, c > 0$. In [43] the function $h(t,c)$ is shown to have a general parametric form

$$h(t,c) = -\frac{qct^{2m+1}}{s^{2m+1}} - \frac{(q+1)ct^{2m}}{s^{2m}} + c$$

in our work we use $m = 1, c = 1, q = 2$. Notice that $P(\cdot)$ is twice differentiable in θ.

The algorithm requires an initialization step that sets several constants and starting parameter values. A and s in (3.26) can be adjusted to scale the filled function and avoid errors due to numerical precision. Since we want to be sure that scaling issues do not mask stationary points, a range of A and s are tested before giving up and deciding the current θ^* is the global maximum. Hence initial values A_i, s_i as well as an upper bound for A which we denote A_U and a lower bound for s, s_L must be selected. An associated scaling constant for each parameter which determines its growth at each adjustment is also used, $\hat{A}$ and $\hat{s}$. The final parameter to set is the number of directions K to perturb the seed θ^* when searching for minima in the filled function. It is suggested that K should be at least twice the number of variables in θ [43]. Since θ^* is a stationary point of $P(\cdot)$, this perturbation ensures the starting point of the minima search is in a region gradient is sufficently different from zero. After this initialization the following algorithm is run.

Global Maximization algorithm:

1. From an initial seed θ_s, find a local maximum θ^* of $\mathcal{L}(\theta)$ and proceed to the next step.

2. Set $k = 1, A = A_i, s = s_i$ and proceed to the next step.

3. If $k > K$ then go to step 5, otherwise use $\theta = \theta^* + \delta_k$ as the seed for a minima search on $P(\theta, \theta^*, s, A)$ where δ_k is a perturbation vector pointing in direction k. Denote the minimum found as θ_k and proceed to the next step. If no minimum is found, set $k = k + 1$ and repeat this step.

4. Verify that θ_k satisfies $\mathcal{L}(\theta_k) > \mathcal{L}(\theta^*)$ then again maximize $\mathcal{L}(\theta)$, using θ_k as the seed to find a new maximizer θ_k^*. Set $\theta^* = \theta_k^*$ and return to step 2.

5. Scale s by $s = s\hat{s}$. If $s \geq s_l$ then reset k to 1 and go to step 3. Else proceed to the next step.

6. Scale A by $A = A\hat{A}$. If $A \leq A_u$ then reset k to 1 and go to step 3. Else the current value of θ^* is taken as the global maximum of $\mathcal{L}$.

We propose a Newton-Raphson search to look for minima in $P(\theta, \theta^*, s, A)$ since an analytic form can be found for both its gradient and Hessian with respect to the parameter vector. A method for modifying the Newton method to be robust for non-convex surfaces

is given in Section 3.3.2. The gradient vector of $P(\cdot)$ is found to be

$$\begin{aligned}\nabla_i P(\cdot) &= \frac{\partial P(\cdot)}{\partial \theta_i} \\ &= \underbrace{-2A\overbrace{\exp\left(-A\|\theta-\theta^*\|^2\right) g\left(\mathcal{L}(\theta^*)-\mathcal{L}(\theta), s\right)}^{P(\cdot)}\left(\theta_i-\theta_i^*\right)}_{\phi_i(\cdot)} \\ &\underbrace{-\exp\left(-A\|\theta-\theta^*\|^2\right)\frac{\partial g}{\partial\left(\mathcal{L}(\theta^*)-\mathcal{L}(\theta)\right)}\left(\mathcal{L}(\theta^*)-\mathcal{L}(\theta), s\right)\nabla_i\mathcal{L}(\theta)}_{\omega_i(\cdot)}\end{aligned}$$

To find the Hessian we break the gradient entries into two equations such that

$$\nabla_{i,j} P(\cdot) = \frac{\partial \phi_i(\cdot)}{\partial \theta_j} + \frac{\partial \omega_i(\cdot)}{\partial \theta_j}$$

$$\frac{\partial \phi_i(\cdot)}{\partial \theta_j} = -2A\frac{\partial P(\cdot)}{\partial \theta_j}\left(\theta_i-\theta_i^*\right) + P(\cdot)\delta_{i,j}$$

where

$$\delta_{i,j} = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}$$

$$\begin{aligned}\frac{\partial \omega_i(\cdot)}{\partial \theta_j} &= -\exp\left(-A\|\theta-\theta^*\|^2\right)\frac{\partial g}{\partial\left(\mathcal{L}(\theta^*)-\mathcal{L}(\theta)\right)}\left(\mathcal{L}(\theta^*)-\mathcal{L}(\theta), s\right)\nabla^2_{i,j}\mathcal{L}(\theta) \\ &- \Big[-2A\exp\left(-A\|\theta-\theta^*\|^2\right)\frac{\partial g}{\partial\left(\mathcal{L}(\theta^*)-\mathcal{L}(\theta)\right)}\left(\mathcal{L}(\theta^*)-\mathcal{L}(\theta), s\right)\left(\theta_j-\theta_j^*\right) \\ &- \exp\left(-A\|\theta-\theta^*\|^2\right)\frac{\partial g^2}{\partial\left(\mathcal{L}(\theta^*)-\mathcal{L}(\theta)\right)^2}\left(\mathcal{L}(\theta^*)-\mathcal{L}(\theta), s\right)\nabla_j\mathcal{L}(\theta)\Big]\nabla_i\mathcal{L}(\theta)\end{aligned}$$

Following the definition of $g(t,s)$ and $h(t,c)$ we have

$$\frac{\partial}{\partial t}g(t,s) = \begin{cases} 0, & t \geq 0 \\ \frac{\partial}{\partial t} - \frac{6t^2}{s^3} - \frac{6t}{s^2}, & -s < t < 0 \\ 0, & t \leq -s \end{cases}$$

and

$$\frac{\partial^2}{\partial t^2}g(t,s) = \begin{cases} 0, & t \geq 0 \\ \frac{\partial^2}{\partial t^2} - \frac{12t}{s^3} - \frac{6}{s^2}, & -s < t < 0 \cdot \\ 0, & t \leq -s \end{cases} \tag{3.27}$$

That (3.27) is not continuous at $t = 0, -s$ is not an issue since the Newton-Raphson search does require entries of the Hessian to be so. What is important is that the filled function be continuous and twice differentiable [46]. With the gradient and Hessian defined, stationary points in the filled function $P(\cdot)$ can be sought.

Local maxima of $\mathcal{L}$ must also be found numerically, which is done with a similarly modified Newton-Raphson search on the surface of (2.10). In order to simultaneously maximize with respect to β, its partial derivitives should be added to our previously defined gradient vector (3.2) and Hessian matrix (3.3) of $\mathcal{L}$. These are found to be

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \beta} = & \sum_{n=2}^{N(T)} \frac{(t_n - t_{n-1} - \Delta_r)\exp\left(-\beta\left(t_n - t_{n-1} - \Delta_r\right)\right)}{1 - \exp\left(-\beta(t_n - t_{n-1} - \Delta_r)\right)} \\ & - \sum_{n=1}^{N(T)-1} \int_{t_n+\Delta_r}^{t_{n+1}} (t - t_n - \Delta_r)\exp\left(-\beta\left(t - t_n - \Delta_r\right)\right)\exp\left(\sum_{k=0}^{r}\alpha_k t^k\right)dt \\ & - \int_{t_{N(T)}+\Delta_r}^{T} \left(t - t_{N(T)} - \Delta_r\right)\exp\left(-\beta\left(t - t_{N(T)} - \Delta_r\right)\right)\exp\left(\sum_{k=0}^{r}\alpha_k t^k\right)dt \end{aligned}$$

and

$$\begin{aligned}
\frac{\partial^2 \mathcal{L}}{\partial \beta^2} &= -\sum_{n=2}^{N(T)} \frac{(t_n - t_{n-1} - \Delta_r)^2 \exp\left(-\beta(t_n - t_{n-1} - \Delta_r)\right)}{\left[1 - \exp\left(-\beta(t_n - t_{n-1} - \Delta_r)\right)\right]^2} \\
&\quad + \sum_{n=2}^{N(T)-1} \int_{t_n+\Delta_r}^{t_{n+1}} (t - t_n - \Delta_r)^2 \exp\left(-\beta\left(t - t_n - \Delta_r\right)\right) \exp\left(\sum_{k=0}^{r} \alpha_k t^k\right) dt \\
&\quad + \int_{t_{N(T)}+\Delta_r}^{T} (t - t_n - \Delta_r)^2 \exp\left(-\beta\left(t - t_n - \Delta_r\right)\right) \exp\left(\sum_{k=0}^{r} \alpha_k t^k\right) dt \\
\frac{\partial^2 \mathcal{L}}{\partial \beta \partial \alpha_i} &= -\sum_{n=2}^{N(T)-1} \int_{t_n+\Delta_r}^{t_{n+1}} t^i \left(t - t_n - \Delta_r\right) \exp\left(-\beta\left(t - t_n - \Delta_r\right)\right) \exp\left(\sum_{k=0}^{r} \alpha_k t^k\right) dt \\
&\quad - \int_{t_{N(T)}+\Delta_r}^{T} t^i \left(t - t_{N(T)} - \Delta_r\right) \exp\left(-\beta\left(t - t_{N(T)} - \Delta_r\right)\right) \exp\left(\sum_{k=0}^{r} \alpha_k t^k\right) dt.
\end{aligned}$$

To illustrate the behaviour of the filled function technique we present a simple example. In this case our objective function $\mathcal{F}$ which we seek to maximize is defined by

$$\mathcal{F} = \begin{cases} \frac{\sin(x)\sin(y)}{xy}, & x, y \neq 0 \\ \frac{\sin(x)}{x}, & x \neq 0, y = 0 \\ \frac{\sin(y)}{y}, & y \neq 0, x = 0 \\ 1, & x, y = 0. \end{cases} \tag{3.28}$$

We recognize (3.28) as a sinc function in two dimensions. This function has many local maxima, but only one global maximum at $x, y = 0$. Local searches on $\mathcal{F}$ require the

gradient vector

$$\frac{\partial \mathcal{F}}{\partial x} = \begin{cases} \frac{(x\cos(x)-\sin(x))\sin(y)}{x^2 y}, & x, y \neq 0 \\ \frac{x\cos(x)-\sin(x)}{x^2}, & x \neq 0, y = 0 \\ 0, & otherwise \end{cases}$$

$$\frac{\partial \mathcal{F}}{\partial y} = \begin{cases} \frac{(y\cos(y)-\sin(y))\sin(x)}{xy^2}, & x, y \neq 0 \\ \frac{y\cos(y)-\sin(y)}{y^2}, & y \neq 0, x = 0 \\ 0, & otherwise, \end{cases}$$

and Hessian

$$\frac{\partial^2 \mathcal{F}}{\partial x^2} = \begin{cases} -\frac{((2x\cos(x)+(x^2-2)\sin(x))\sin(y)}{x^3 y}, & x, y \neq 0 \\ -\frac{((2x\cos(x)+(x^2-2)\sin(x))}{x^3}, & x \neq 0, y = 0 \\ -\frac{\sin(y)}{3y}, & y \neq 0, x = 0 \\ -\frac{1}{3}, & x, y = 0, \end{cases}$$

$$\frac{\partial^2 \mathcal{F}}{\partial y^2} = \begin{cases} -\frac{((2y\cos(y)+(y^2-2)\sin(y))\sin(x)}{xy^3}, & x, y \neq 0 \\ -\frac{\sin(x)}{3x}, & x \neq 0, y = 0 \\ -\frac{((2y\cos(y)+(y^2-2)\sin(y))}{y^3}, & y \neq 0, x = 0 \\ -\frac{1}{3}, & x, y = 0, \end{cases}$$

$$\frac{\partial^2 \mathcal{F}}{\partial x \partial y} = \begin{cases} \frac{(x\cos(x)-\sin(x))(y\cos(y)-\sin(y))}{x^2 y^2}, & x, y \neq 0 \\ 0, & otherwise. \end{cases}$$

The gradient and Hessian for $P(\cdot)$ remain the same as previously shown, but with $\mathcal{F}$

replacing $\mathcal{L}$, and redefining the parameter vector as $\theta = [x, y]^T$.

Figure 3.2 visualizes steps taken by the filled function approach when maximizing $\mathcal{F}$. The main lower plot shows $\mathcal{F}$ in a region about the global maximum. The search is seeded at the point shown by the $\bullet$ marker with coordinates (-8,-3). A local maximization of $\mathcal{F}$ terminates at the nearby peak (-7.7, 0), marked by the $\blacklozenge$. This point is then used to construct the filled function $P(\theta, \blacklozenge, s, A)$ which we plot with $s = .2, A = .002$ in the upper left frame. The minimization starts from $\blacklozenge$ and finds a stationary point indicated by $\blacksquare$, located at approximately (-2,0). From the definition of a filled function this point must satisfy $\mathcal{F}(\blacksquare) > \mathcal{F}(\blacklozenge)$. Notice that other local maxima in $\mathcal{F}$ that are not as good as $\blacklozenge$ are suppressed and do not create cavities in the filled function. The next step is to seed another local maximization of $\mathcal{F}$ at the new found point $\blacklozenge$. This search ends at (0,0) indicated with a $\blacktriangle$. As we know from the sinc function, it is in fact the global maximum. More importantly, the algorithm is also able to determine $\blacktriangle$ is the global maximum because there are no stationary points in $P(\theta, \blacktriangle, s, A)$, plotted in the top right frame, which satisfy $\theta \neq \blacktriangle$. Subsequent searches on $P(\theta, \blacktriangle, s, A)$ then fail to find any minima, even after scaling with various values of A and s. The search gives up, and $\blacktriangle$ is returned as the global maximum of $\mathcal{F}$.

3.3.2 Local searches on non-convex surfaces

We require a modified Newton-Raphson approach to find local maxima in $\mathcal{L}$ and minima in $P(\cdot)$ since neither are convex. For the purposes of this discusion we assume a minimization problem, but note any maximization problem can be easily treated as the former by multiplying the objective function by (-1). A modification to the search is required since Newton method may not converge to a minimum if the surface is not convex [46].

In its standard for the Newton algorithm requires both the Hessian and gradient of the

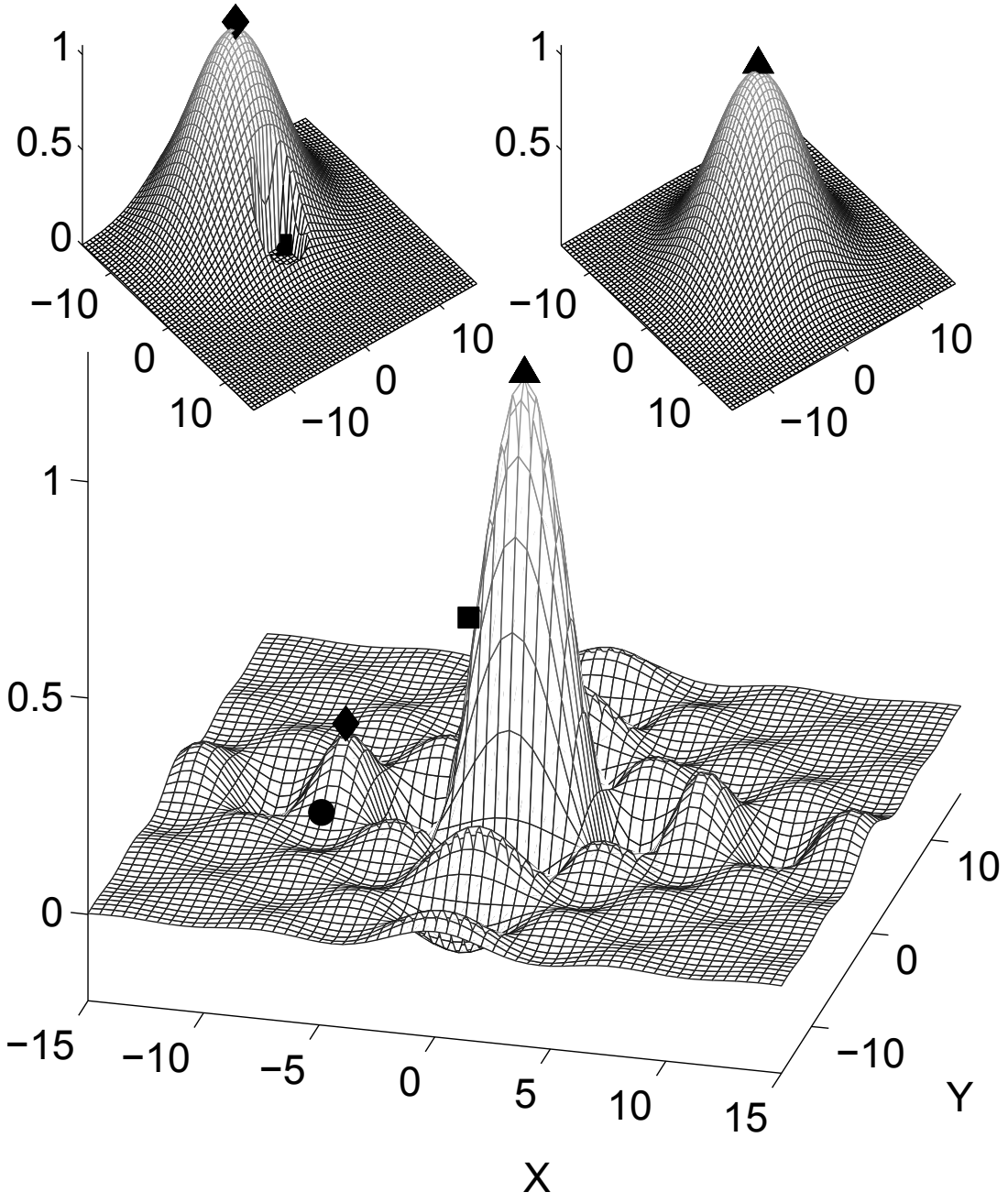

Fig. 3.2 Visualization of the filled function approach. The main lower frame shows the objective function to be maximized. Top frames show realizations of $P(\cdot)$ for different θ^*. See text for full description.

objective function $\mathcal{F}$ with respect to its parameter vector θ. An initial seed to the search θ_0 is also needed, from which point the algorithm proceeds in an iterative fashion towards the minimum by calculating a step vector v and adding it to the parameter vector θ. The step vector at iteration k is calculated as follows

$$v_k = -\nabla\mathcal{F}(\theta_k)[\nabla^2\mathcal{F}(\theta_k)]^{-1}, \tag{3.29}$$

then

$$\theta_{k+1} = \theta_k + v_k.$$

The step is in a minimizing direction if $\mathcal{F}(\theta_{k+1}) < \mathcal{F}(\theta_k)$, which implies

$$\nabla\mathcal{F}(\theta_k)^T v_k < 0.$$

This condition is guaranteed if the Hessian is positive definite since

$$\nabla\mathcal{F}(\theta_k)^T v_k = -\nabla\mathcal{F}(\theta_k)^T[\nabla^2\mathcal{F}(\theta_k)]^{-1}\nabla\mathcal{F}(\theta_k).$$

The Hessian of $P(\cdot)$ and $(-\mathcal{L})$ fail to be positive definite so (3.29) is not guaranteed to be a minimizing direction in our searches. This problem can be solved by using a substitute for $\nabla^2\mathcal{F}$ in (3.29). We denote this substitute H, defined

$$H_k = \nabla^2\mathcal{F}(\theta_k) + \delta_k I, \tag{3.30}$$

where I is the identity matrix and δ_k is a positive constant. Setting δ_k slightly larger than the most negative eigenvalue of $\nabla^2\mathcal{F}(\theta_k)$ insures H is positive definite as a consequence of

the Gerschgorin circle theorem, which in part states

$$\min_{1\leq i\leq n} \phi_i \geq \min_{1\leq i\leq n} \left(b_{ii} - \sum_{j=1, j\neq i}^{n} |b_{ij}| \right) \tag{3.31}$$

where ϕ_i are the eigenvalues of an n-by-n symmetrical matrix B whose entries are b_{ij} [46]. We note that $\nabla^2\mathcal{F}(\theta_k)$ is symmetric by definition. This theorem can be exploited in two ways. First, we can check if the Hessian evaluated at θ_k is positive definite. If the bound on the minimum eigenvalue is positive, then the Hessian must be positive definite since it follows immediately that all eigenvalues are positive. If the bound is negative, we can use its value to select δ_k in (3.30) such that H_k is Positive definite.

To see the second usage we note from (3.30) that only the diagonal terms differ between $\nabla^2\mathcal{F}(\theta_k)$ and H_k. This difference is δ_k. In (3.31) we can see directly that the minimum eigenvalue ϕ_i is bounded by a summation of one diagonal term b_{ii}, and the negative magnitude of off diagonal terms b_{ij} of the matrix B. Therefore, adding a positive value to the diagonal terms of $\nabla^2\mathcal{F}(\theta_k)$ must make the bound more positive. The value for δ_k which ensures all eigenvalues are positive is then just slightly larger than the magnitude of the most negative eigenvalue, for which we have found a bound using (3.31).

3.3.3 Initial seed

In the example shown in Figure 3.2, we knew a priori where the global maximum was and could deliberately plant a seed somewhere nearby for a quick convergence. In our actual problem, less is known about $\mathcal{L}$, but we still wish to have a reasonable first guess to avoid lengthy searches. For this we suggest a first guess which can be calculated quickly and which requires only knowledge of $N(T)$ to compute. For large β the log-likelihood equation

(2.10) of a constant excitation, (2.2) with $r = 0$, is approximated very well by

$$\begin{aligned}\mathcal{L} &= N(T)\alpha_0 - \int_0^{t_1} \exp(\alpha_0)\, dt - \sum_{n=1}^{N(T)-1} \int_{t_n+\Delta_r}^{t_{n+1}} \exp(\alpha_0)\, dt - \int_{t_{N(T)}+\Delta_r}^{T} \exp(\alpha_0)\, dt \\ &= N(T)\alpha_0 - \exp(\alpha_0)\left((T - t_0) - N(T)\Delta_r\right), \end{aligned} \tag{3.32}$$

since the effects of the relative refractory period are negligible. By taking the derivative of (3.32) with respect to α_0 and setting the result to zero, the ML estimate for a zero order model is found knowing only the observed number of spikes $N(T)$. This approximation yields

$$\begin{aligned}\frac{\partial \mathcal{L}_0}{\partial \alpha_0} &= N(T) - \exp(\alpha_0)(T - N(T)\Delta_r) = 0 \\ \hat{\alpha}_0 &= \log\left(\frac{N(T)}{T - N(T)\Delta_r}\right)\end{aligned}$$

where we have assumed $t_0 = 0$. The starting point to the global search is then $\alpha_0 = \hat{\alpha}_0$ and $\alpha_i = 0$, $i = 1, .., r$ regardless of what value of r we have assumed the free firing rate model to posses. Justification for approximating β to be large when seeding comes from the fact that since the relative refractory period is known to be short (on the order of ms), β is very unlikely to have a small value owing to (3.1).

3.3.4 Alternative global maximization

We note that one iteration of the filled function approach presented in Section 3.3.1 requires several Newton searches, and that this number grows with the size of the parameter vector and number of maxima encountered. This can be seen from the fact that a search of $P(\cdot)$ is carried out seeded from θ^* plus each perturbation vector δ_k, and that the number of

perturbation vectors is linearly dependent on the number of parameters in θ. If several local maxima are found on the way to the global maximum in a highly dimensional space, the number of Newton searches may become quite large. To possibly alleviate excessive computational effort we propose a quasi-exhaustive approach which exploits the convexity of (2.10) when β is constant.

In this approach we use some a priori knowledge to select not one, but a set of possible values of β. Although we may have no reason to think β has a specific value, it seems reasonable to expect it is within a certain range. By seeding multiple searches from the values in this range many local maxima, and likely the global maximum, can be found. The approach is implemented as follows and carried out for every $r \in R$.

Alternative Global Maximization algorithm:

1. Define the candidate vector for $\beta = \beta_1, \beta_2, ..., \beta_K$. The values of β_k should cover a reasonable range of relative refractory period lengths. Set $k = 1$. Go to next step.

2. Maximize (2.10) with β fixed at β_k using a standard Newton-Raphson search. Denote the free firing rate parameters found $\alpha_0^*, \alpha_1^*, ..., \alpha_r^*$. Go to next step.

3. Use β_k and $\alpha_0^*, \alpha_1^*, ..., \alpha_r^*$ as the seed for another Newton-Raphson search on (2.10) but with β as a variable. The method in Section 4.4.2 should be used now since the likelihood surface is no longer concave. Denote the parameter vector at the maximum found as θ_k^*. If $k \leq K$, increment k and go to step 2, else go to next step.

4. The final free firing rate and β estimates are determined by the parameters of θ_l^* such that $l = \arg\max_{k \in K} \mathcal{L}(\theta_k^*)$.

The advantage of this approach is that the number of searches required is fixed at $2K$. The disadvantage is it is unclear how large K needs to be and how to distribute the values of β_k. We suggest, however, that a spacing of such that $5/\beta_k - 5/\beta_{k+1} = 1\text{ms}$ may be adequate. We compare accuracy of the alternative method with the filled function search in Chapter 4.

Chapter 4

Simulation and data analysis

This chapter presents numerical results with respect to four key issues. We first show that the exponential of a polynomial function is a reasonable choice for $\gamma(t)$. Next we compare the performance of the model order selection rules presented in Chapter 3 to determine which is most appropriate. Finally, we conduct goodness of fit tests on real neural data to show that the refractory model proposed is suitable for use with real spike train measurements.

4.1 Metrics

Our results are primarily assessed through two metrics. The first is Normalized Mean Integrated Squared Error (NMISE) defined

$$E\left[\frac{\int_0^T \left(\gamma(t)-\hat{\gamma}(t)\right)^2 dt}{\int_0^T \gamma^2(t)dt}\right], \tag{4.1}$$

where the expectation is taken over a series of trials, $\gamma(t)$ is the true excitation, and $\hat{\gamma}(t)$ is the estimate from a particular trial. This metric is a direct comparison between the

true excitation and estimate, normalized by the true excitation power. Because the true excitation must be known in order to compute it, its use is limited in our case to simulations. We commonly express NMISE as a percentage, in which case (4.1) is multiplied by 100.

To assess estimates on real data we desire a method to infer the estimate quality without having to know the exact form of $\gamma(t)$. A result known as the time re-scaling theorem can be used to asses goodness of fit between a rate $\lambda(t)$ and a set of spike times $\mathcal{T}$. This is done by comparing the inter-spike intervals observed against those predicted by the estimated rate function [47]. Although our primary concern is to estimate $\gamma(t)$, we can easily find $\lambda(t)$ using (2.5), the estimated free firing rate, β and Δ_r.

Let $\mathcal{T}$ be a realization from a point process with intensity function $\lambda(t)$. Define

$$\Lambda(t_i) = \int_{t_i-1}^{t_i} \lambda(t)dt,$$

for $i = 2, ..., N(T)$ and assume $\Lambda(t) < \infty$ for all $t \in (t_0, T]$. The time re-scaling theorem states that the $\Lambda(t_i)$'s will follow the ISI distribution of a Poisson process with an associated constant rate $\lambda(t) = 1$, the exponential distribution [47].

Ordering the $\Lambda(t_i)$'s from smallest to largest and plotting them against the $N(T)$ quantiles of the an exponential distribution with unit parameter results in a Quantile-Quantile (QQ) plot. In our QQ plots we denote the sorted $\Lambda(t_i)$'s as z_i where z_1 is the smallest $\Lambda(t_i)$. A good fit is obtained if all the points lie close to a straight line of 45 degrees, and points away from this line indicate the estimated rate does not fit well with the data [47]. In our figures we have added dashed lines indicating a 95% confidence interval, indicating the distance from the center line each point would deviate 5% of the time if the $\Lambda(t_i)$'s were actually exponentially distributed. These confidence boundaries are calculated using the approximation described in [47].

One key point regarding QQ plots is that the goodness of fit can be used for almost any type of point process data. This means that regardless of whether $\lambda(t)$ is thought to depend on simply a cell to cell excitation or combinations of abstract higher level information, any model for $\lambda(t)$ can be compared to the measured spike times. For example in [22], a model which suggests $\lambda(t)$ depends on a two dimensional coordinates of a rat in its environment is assessed using QQ plots. Although in this example what $\lambda(t)$ is representing is different from most of our work, there is absolutely no issues applying QQ plots to both situations since $\lambda(t)$ is well defined.

We introduce $\bar{\gamma}$, the average free firing rate defined

$$\overline{\gamma} = \frac{\int_{t_0}^{T} \gamma(t)dt}{T - t_0}.$$

This is not a metric of error, but is referred to often as a characteristic of the excitation. As $\bar{\gamma}$ increases, the expected number or spikes in the response also increases up to some saturation level governed by the refractory parameters. For example, a neuron with an absolute refractory period of 5 ms could never produce more than 200 spikes per second regardless of the excitation. This is a result of our rate model definition (2.5). Since after a spike there is zero probability of seeing another during the absolute period, 200 spikes would produce 1 second of dead time. Thus while high valued excitations tend to produce more spikes, there is a fundamental limit. This concept is given some further discussion in Chapter 5.

4.2 Excitation model validation

We have chosen the parametric excitation (2.2) because the exponential of a polynomial can theoretically be fit to any non-negative function. We also desire that the number of parameters required for a good fit to be relatively small such that we can confidently set a value on the number of models we should consider in the candidate set $\mathcal{R}$. To investigate this we test the model's fit to various functions which are not of the form (2.2).

Our first example is a parametric function found in [6]. This function is suggested to model the intensity of a typical neural response and is defined

$$\gamma(t) = \theta_1 + \frac{\theta_2}{\theta_3 - \theta_4}\left[\exp\left(\frac{-t}{\theta_3}\right) - \exp\left(\frac{-t}{\theta_4}\right)\right] \tag{4.2}$$

where $\theta_i > 0, i = 1, 2, 3, 4$ are the parameters of the model and $\theta_3 > \theta_4$. This function is plotted in Figure 4.1 frame A, along with fits of various model orders. Frame B plots NMISE as a function of r of fits for on a semi-logarithmic scale. We see that a good fit, which we define by an NMISE score around 1%, is achieved by model orders above 5. The numerical values for the parameter vector θ used in Figures 4.1, 4.2, 4.3 and 4.4 are given in Table 4.1.

We also tested a sinusoidal function, defined

$$\gamma(t) = \theta_1 + \theta_2 \sin\left(\theta_3 t + \theta_4\right), \tag{4.3}$$

where we restrict $\theta_1 \geq \theta_2$ such that the free firing rate is non-negative. Results are shown in Figure 4.2. For higher orders (2.2) is almost indistinguishable from (4.3) and a good fit is achieved at $r = 7$ and above.

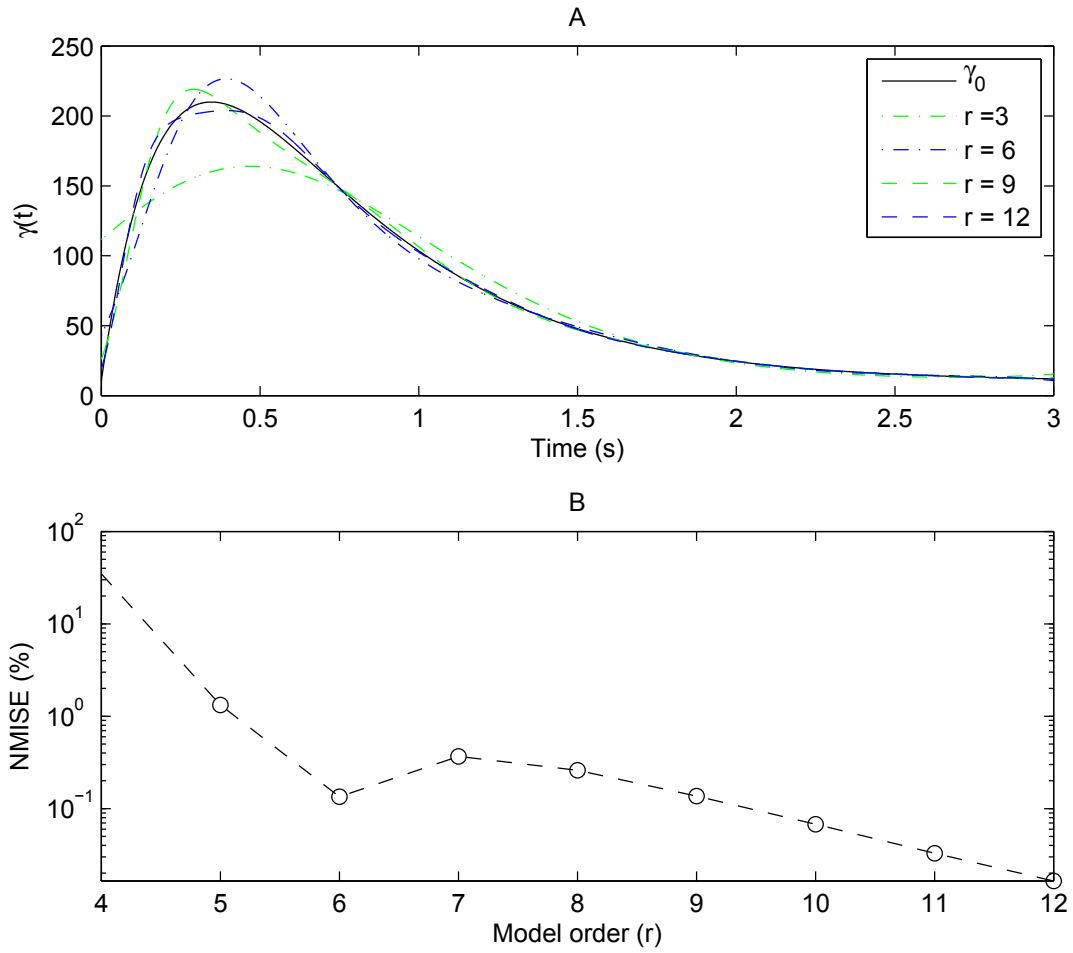

Fig. 4.1 Fit of the excitation model (2.2) to the parametric function (4.2). Frame A shows fits from various model orders denoted in the legend. Frame B shows NMISE as a function of r.

The function shown in Figure 4.3 is a boxcar function we define piece-wise as

$$\gamma(t) = \begin{cases} \theta_1, & \text{if } t < \theta_2 \\ \theta_3, & \text{if } \theta_2 \leq t < \theta_4 \\ \theta_1, & \text{if } \theta_4 \leq t \end{cases}. \tag{4.4}$$

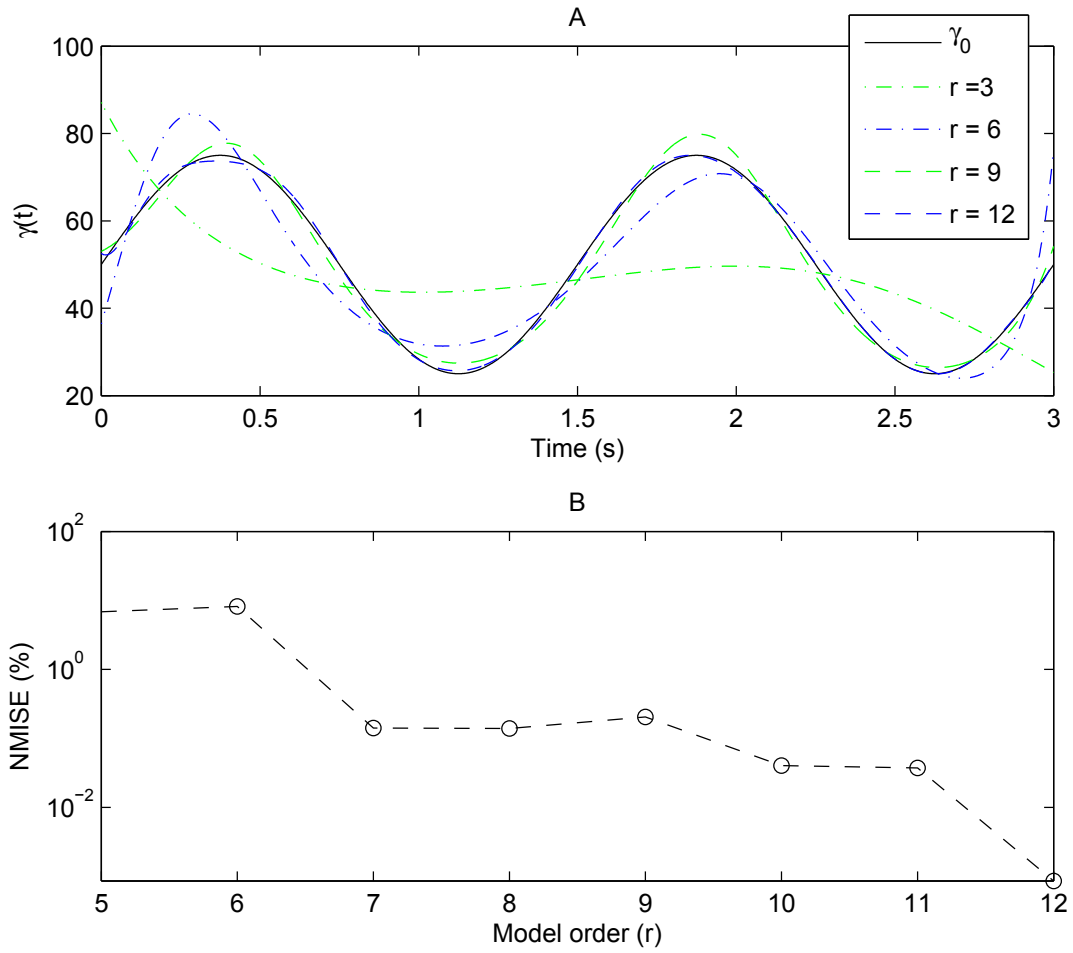

Fig. 4.2 Fit of the excitation model (2.2) to the sinusoidal function (4.3). Frame A shows fits from various model orders denoted in the legend. Frame B shows NMISE as a function of r.

We see that the model has difficulty fitting to sharp edged curves. Despite this we still feel (2.2) is an excellent model choice. In [4] a lower bound on error is derived for the case of linear filtering. The main result is that the bound depends on a ratio of the intensities bandwidth to the average firing rate. We can express this ratio as

$$\frac{T\omega_\lambda}{\int_{t_0}^{T}\lambda(t)dt} \tag{4.5}$$

where ω_λ is the intensity bandwidth. For accurate estimates to be had this ratio should be a small fraction, [4] suggesting a value as low as .01. Although our estimator is not a linear filter, we expect that a similar limit exists and that this follows from the nature of the problem, not the estimator structure.

The pulse excitation shown in Figure 4.3 has harmonics which lead to a relatively wide bandwidth. It seems reasonable to ask what typical firing rates are in measured data

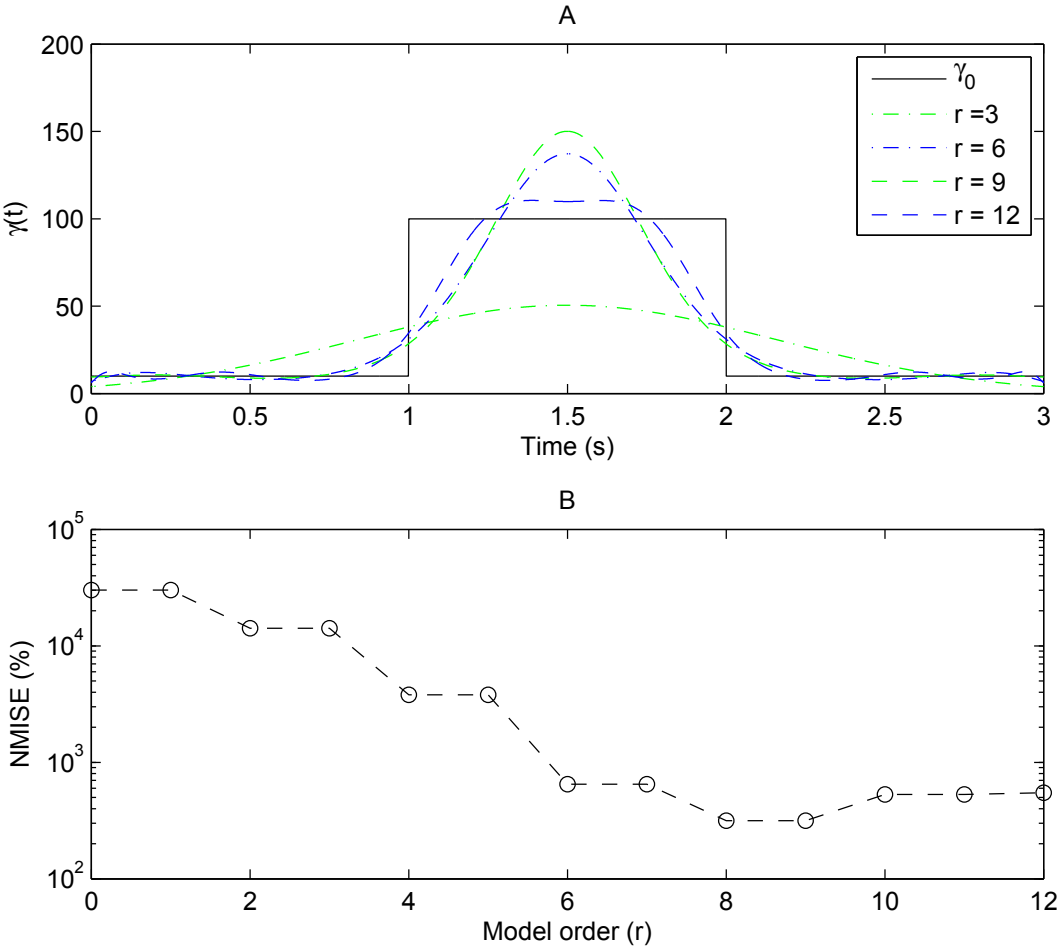

Fig. 4.3 Fit of the excitation model (2.2) to a rectangular pulse described in (4.4). Frame A shows fits from various model orders denoted in the legend. Frame B shows NMISE as a function of r.

in order to decide whether or not such a wide bandwidth excitation could be recovered with any free firing rate model. A brief survey indicates that maximal firing rates vary significantly depending on the type of cell being measured, its role in information processing, and relevance to the stimulus presented. In [48] several individual cells were recorded from the cerebellar nuclei of multiple rhesus monkeys. Cells found in the fastigial region displayed peak rates well over 300 spikes per second for certain stimuli. Trial averages over a different of stimuli ranged from 50 to 100 spikes per second.

In [49], where the function of different cell types was during smooth pursuit of visual targets, it was found that the typical firing rates of Purkinje cells is quite different than that of other brain stem cells, even when cooperating on the same task. In this experiment the range average firing rates for Purkinje cells were seen to be between 100-200 spikes per second. In the free firing rate discussion provided in [14] we see that for the retinal ganglion cells measured the peak firing rate is also around 200 spikes per second.

Some brainstem neurons that take part in saccadic eye movements can show very high firing rates during their bursts. Results such as those in [50] show these rates can peak at over 500 spikes per second, although the cell tends to be only active for brief moments during the saccade.

In contrast to these rapidly firing brainstem cell, motor neurons involved in arm trajectory execution typically have much lower firing rates, around 50 spikes per second, and tend to increase in intensity when encoding movements with greater force [51].

Such a a range of maximum rates agrees well with the notion of the refractory period only lasting a few milliseconds as discussed in Chapter 2. If we take the upper ranges found in our survey of around 500 spikes per second, then the guidelines from [4] would suggest that firing rates excited by a bandwidth of 5 Hz maximum can be well reconstructed in the single trial estimate. However, as with the majority of our analyzed data, we find often the

firing rate is much less than this max, on the order of 25-100 spikes per second. For such values the guidelines in [4] would suggest accurate estimates should only be expected on rates whose bandwidth is .25 to 1 Hz. Figure 4.4 shows a pulse filtered with a first order low-pass response at a bandwidth which results in (4.5) equal to .02, double the bound suggested in [4]. We see that the fit to the filtered pulse is considerably improved and that a good fit is achieved at $r = 5$. We conclude that the excitation model is adequate for the problem at hand.

Figure	Equation	θ_1	θ_2	θ_3	θ_4
4.1	(4.2)	10	200	.5	.25
4.2	(4.3)	50	25	$4\pi/3$	0
4.3	(4.4)	10	1	100	2
4.4	lpf (4.4)	10	1	100	2

Table 4.1 Parameter values used in excitation plots seen in Figures 4.1, 4.2, 4.3 and 4.4. Lpf refers to a first order low passed filtered version of the excitation such that (4.5) is equal to .02. For all functions the observation period is defined $t_0 = 0$, $T = 3$s.

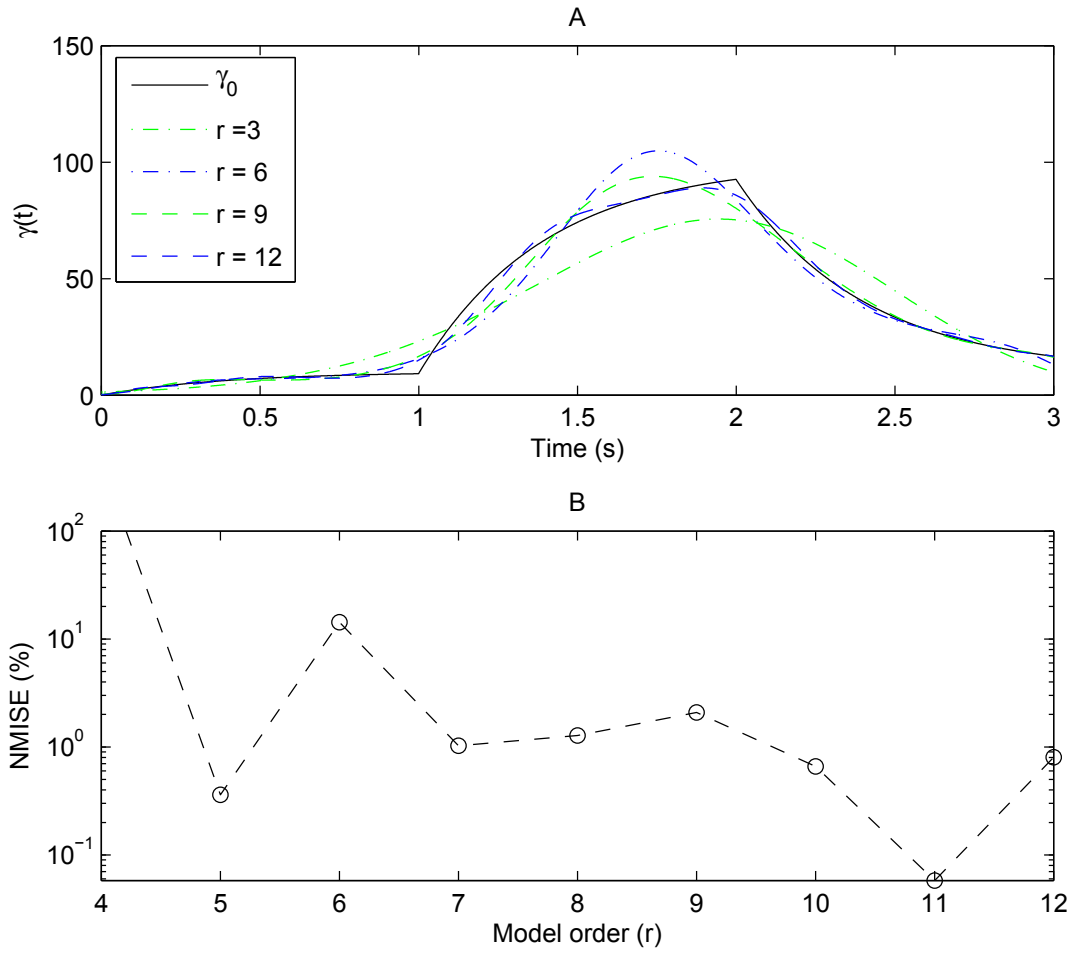

Fig. 4.4 Fit of the excitation model (2.2) to a filtered version of (4.4). Frame A shows fits from various model orders denoted in the legend. Frame B shows NMISE as a function of r.

4.3 Model order selection tool validation

In Chapter 3 we proposed that some strategy beyond a ML approach is required to determine r in (2.10). This is because the value of (2.10) evaluated at the ML estimate for $\gamma(t)$ tends to increase with r, regardless of the true model order. This is illustrated in Figure 4.5, which plots NMISE and $\mathcal{L}$ evaluated at ML free firing rate estimates for different model orders r. The true excitation was of the form (2.2) with $r = 5$. Although a clear minimum in NMISE is seen at the true model order, $\mathcal{L}$ is a non-decreasing function in r. We conclude that a ML procedure for model selection would pick a model with a much higher NMISE than the best candidate. Since it is not clear from the discussion in Section 3.3 which criterion will perform best, we performed an empirical test.

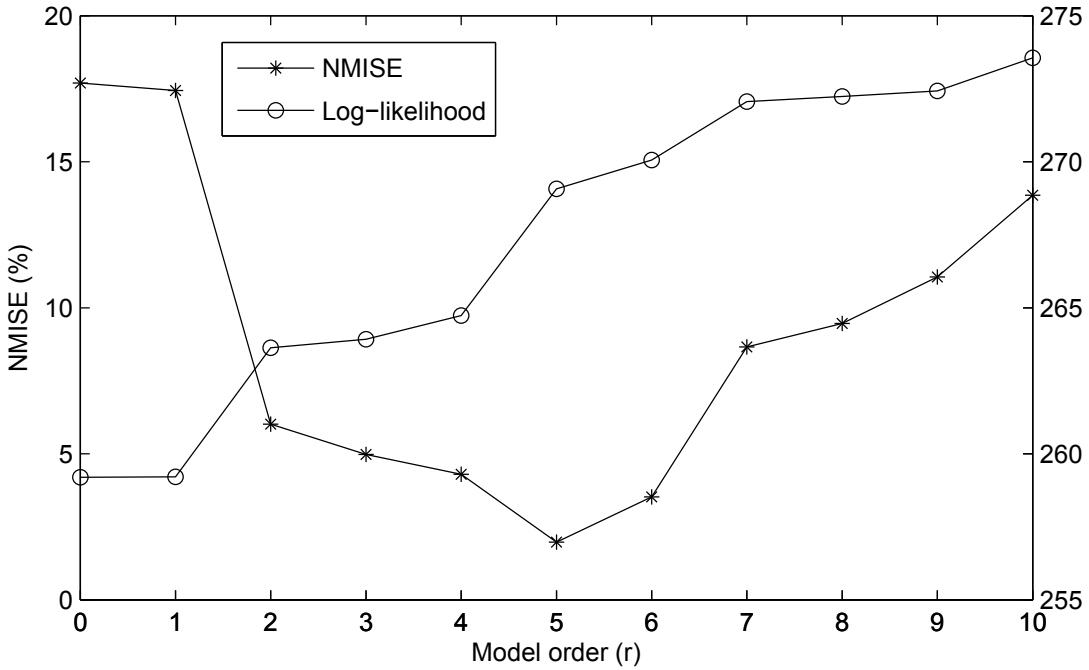

Fig. 4.5 NMISE and log-likelihood compared for ML estimates of $\gamma(t)$ as a function of model order r. A distinct minimum for the error is seen with respect to NMISE, while the log-likelihood is non-decreasing with r.

We ran a simulation in which randomly generated rate functions with the form (2.2) were used to produce spike data. Six functions with $\bar{\gamma} = 100$ for each order $r = 1, ..., 7$ were generated. For each function 500 trials worth of data were produced (3000 trials per order). We then applied the estimator to the data to find the ML estimate of $\gamma(t)$ for all model orders in $\mathcal{R}$ on a trial-to-trial basis. Three final estimates were then chosen, one by each of the model order selection rules discussed in Chapter 3 (AIC, BIC, AICC). In the test $\mathcal{R}$ contained orders $r = 0, 1, ..., 10$ and the refractory parameters were known a priori to the estimator. The full procedures for randomly generating the data and excitations is described in Appendix A. The resulting excitation coefficients are given in Appendix C, Table C.1.

Figure 4.6 shows the NMISE results as a function of true model order. We have added a fourth series to suggest a lower bound on performance of the order selection process, shown by the dashed line labelled “genie". This is the resulting NMISE when the true model order was known, not estimated by any selection rule. AICC yields the best score across all model orders considered, although as the order increases AIC and AICC tend to converge in performance. BIC performs substantially worse, except at order 1, where it is on par with AICC. These results may suggest that BIC penalizes too heavily as the model order increases, as can be seen in Figure (3.1), the slope of the penalty term is much larger as a function of r for BIC than AIC or AICC.

We also note that as the model order increases, the NMISE of any rule (including the genie) increases. Since $\bar{\gamma}$ was the same for all free firing rates, and the refractory effects constant, this is not likely a side effect of the amount of spikes available to the estimator. Instead we suggest that more complex excitations are inherently more difficult to decode. Interestingly the trend in the genie result as r increases is almost identical to those seen in AICC, with a constant offset of about 2%. Thus the error introduced by not knowing r

seems to be quite low.

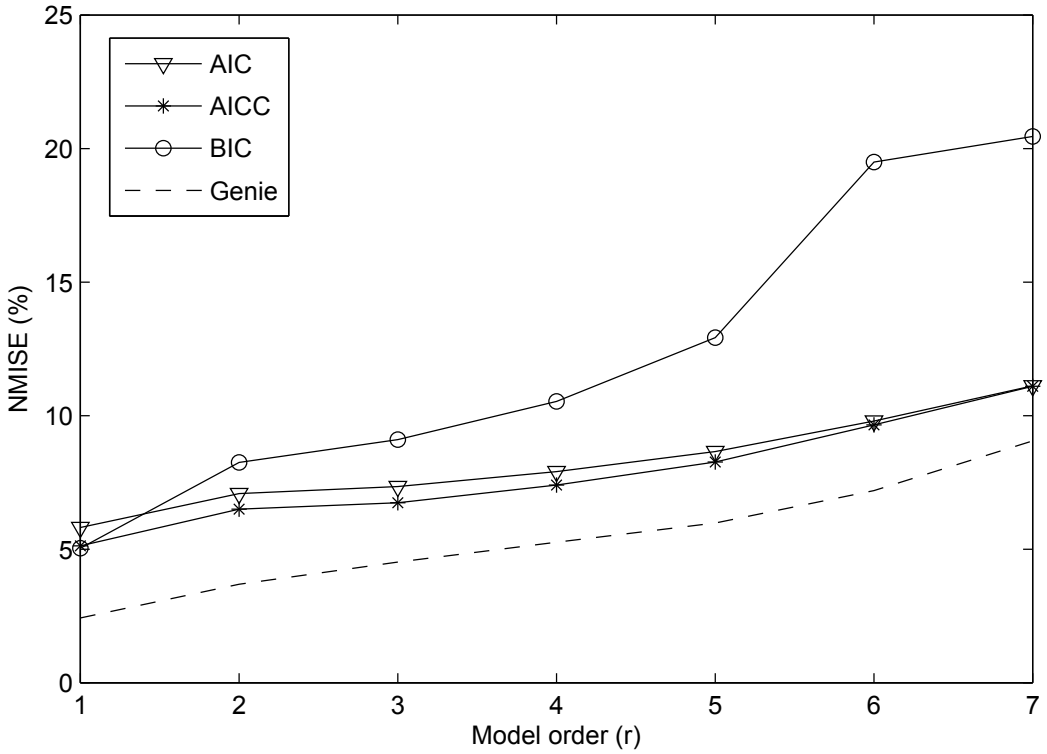

Fig. 4.6 Model selection rule comparison results. NMISE of free firing rate estimates when the true excitation had the form (2.2) and order r. AIC, AICC, BIC and a genie selection rule compared. Details of free firing rates in text.

Histograms of the trial-to-trial order estimates of the three rules were generated to investigate our suspicions regarding BIC under-fitting and to see how frequently correct orders were being selected. Figure 4.7 shows histograms for three of the seven values tested, $r = 1, 4, 7$ in frames A, B and C respectively. For $r = 1$ all rules tend to select the true order most of the time, 65.5, 69.5 and 64.3% for AIC, AICC and BIC respectively. BIC only seems to under-fit while AIC and AICC have a long tail into the upper values of r. With $r = 4$ BIC prefers to under-fit, selecting a zero order model 48.8% of the time, and the correct order in only 9.9% of the trials. AIC and AICC select the true order 45.1 and

48.2% of the time but show some tendency to under-fit. At $r = 7$ BIC continues to select order zero most of the time (49%) but has a tail that extends up to the true order, selecting the correct value in 5.3% of the trials. AIC and AICC select around the true order, being correct 31.7 and 29.7%. Despite these lower scores for AIC and AICC at high orders, we see the second most selected model was only a slight under-fit and the combined percentage for $r = 6$ and 7 were 56.5 and 55.8% for AIC and AICC respectively. One trend seen is that AICC always under-fits slightly more than AIC, and over-fits slightly less. It is not clear if this has any significant impact since, at order 7 where this is most noticeable in the histogram, the NMISE result is virtually the same for AIC and AICC as seen in Figure 4.6. Based on the results of this test we have selected AICC as the best model order selection rule for our estimator and use it exclusively in any results that follow.

4.4 Single trial free firing rate estimation

Trial-by-trial estimate accuracy of both the free firing rate and refractory effects was investigated through a computer simulation. Our first data set was generated from excitations which had the exponential of a polynomial form (2.2). We also present results for excitations of the form (4.2) and (4.3).

4.4.1 Matched free firing rate model

Simulated spike trains containing a refractory period similar to our model were generated and decoded. Spike data was produced using various $\gamma(t)$ of the form (2.2) with $r = 4$, $\Delta_r = .002$ and $\beta = 2500, 866, 500$ which corresponds to relative refractory periods of $.002, .006, .01$ seconds. Two sets of excitations were used satisfying $\overline{\gamma} = 100, 300$. Each excitation lasted a period of $T = 3$ seconds. The exact coefficients are provided in Appendix

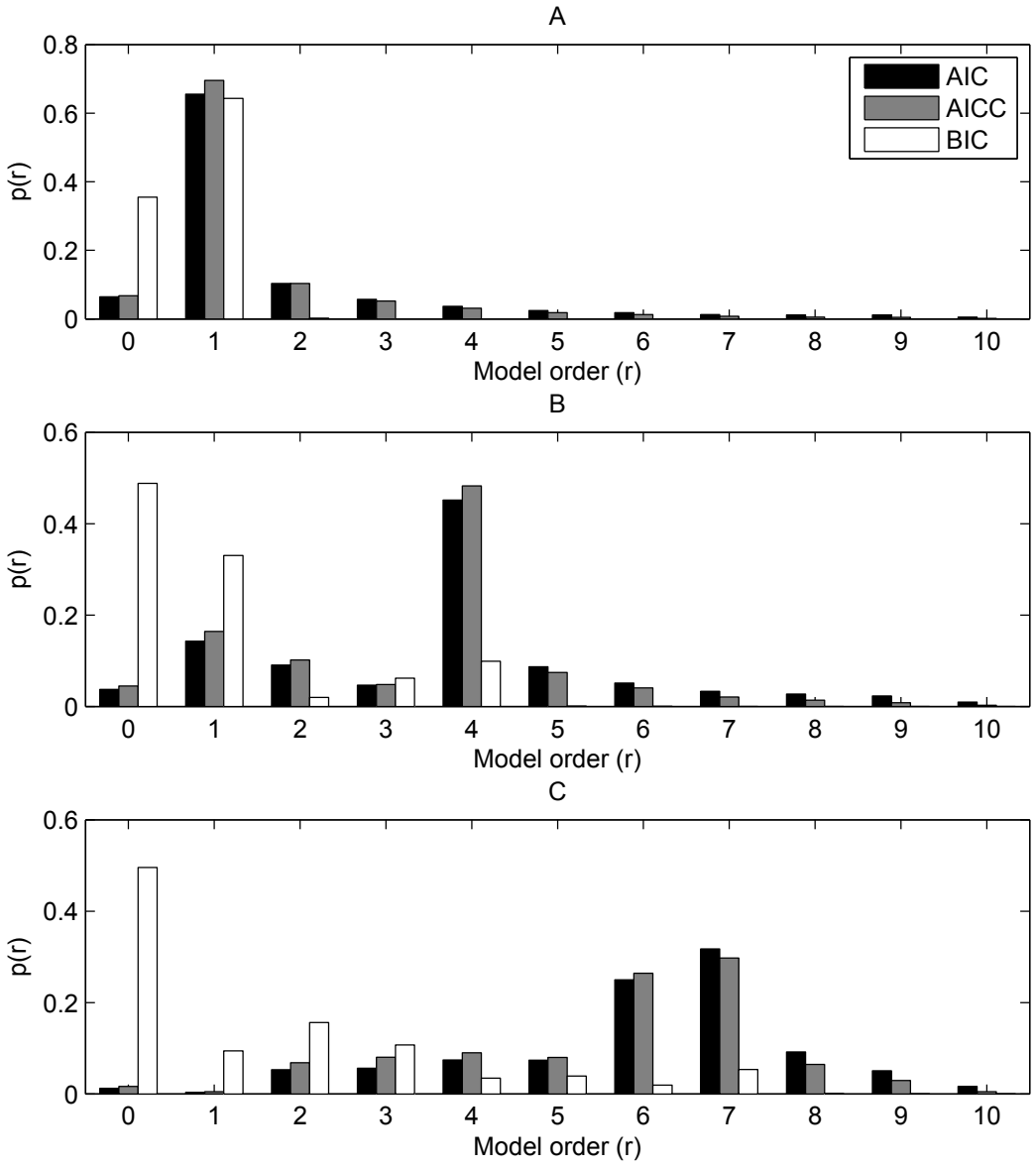

Fig. 4.7 Histograms displaying the order estimates of AIC, AICC and BIC when the true excitation had order $r = 1, 4, 7$ in frames A,B and C respectively.

C, Table C.2.

We ran our estimator in three configurations: first assuming the data was Poisson

($\Delta_r = 0, \beta = \infty$), second estimating the absolute period but assuming no relative refractory period ($\beta = \infty$) and finally estimating both absolute and relative refractory period for every trial. This allows inference on how much accuracy is lost by ignoring part or all of the refractory period.

Results are plotted in Figure 4.8 with frame A and B showing the results for $\overline{\gamma} = 100, 300$ respectively. All estimators show some degradation as the length of the relative refractory period increases. The greater performance decrease seen by the Poisson and sometimes absolute only estimator can be explained by the fact that larger relative refractory periods increase the mismatch between the data and model used in the estimator. For the matched estimator, slope in NMISE curve is minimal and may be caused by the fact that fewer spikes tend to occur over the fixed observation period as the relative refractory period lengthens. Accounting for the full refractory period yields noticeable improvement, even when it is relatively short. Overall its performance is superior by about 4-8% compared to the next best scheme. The data points in Figure 4.8 are summarized in Table 4.2.

		NMISE (%)		
Estimator	$\bar{\gamma}$	$\beta = 2500$	$\beta = 866$	$\beta = 500$
Poisson	100	32.83	45.59	57.62
Abs. Only	100	6.97	7.88	9.62
Full	100	2.84	3.58	4.99
Poisson	300	6.57	9.12	11.52
Abs. Only	300	7.73	13.79	18.96
Full	300	1.85	2.84	3.85

Table 4.2 Numerical values for results plotted in Frigure 4.8.

Figure 4.9 shows two examples of single trial estimates along with the true excitation. In frame A we can see that only the full estimator captures the general shape of the excitation correctly. Frame B shows an excitation which brings the neuron to saturation at about the .7 second mark. The full estimator does a reasonably good job at capturing the peak,

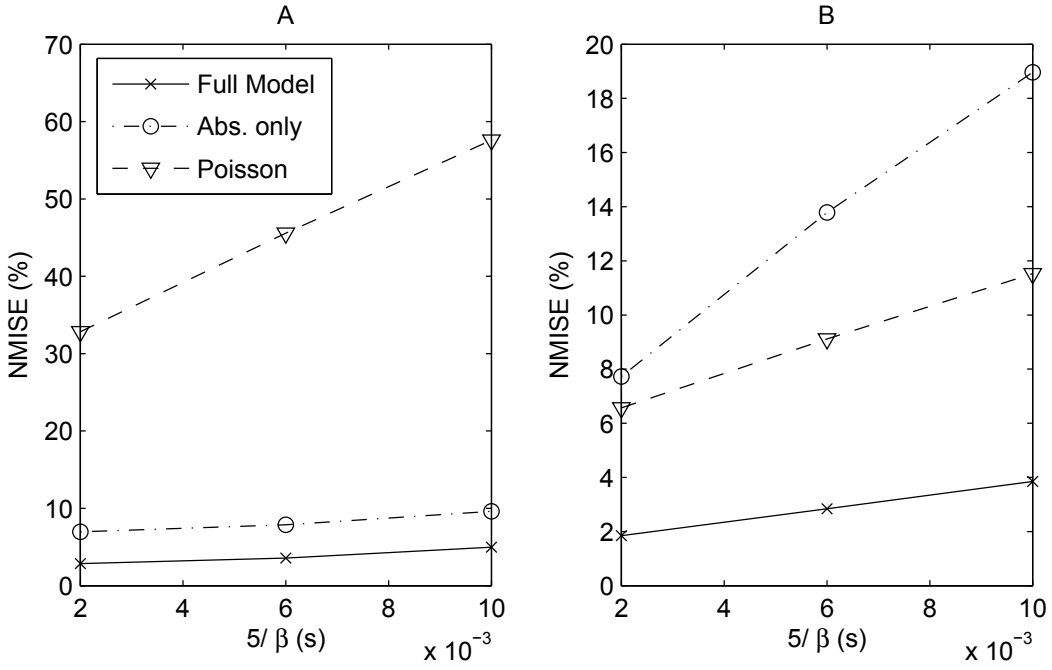

Fig. 4.8 NMISE simulation results based on synthetic data with a refractory period generated by a excitation with the form (2.2). Free firing rates were characterised by $\bar{\gamma} = 100, 300$ in frame A and B respectively.

while the other two estimators tend to ignore this feature. In both frames the Poisson estimator tends to give a lower intensity. This makes sense since it generates its estimate without compensating for any refractory effect. The absolute only model also does poorly at capturing the shape of the free firing rate in both trials although its estimate in frame A is close to the average, where the relative period was only 2ms. In frame B the relative period was 10 ms which explains the much poorer job by the absolute only estimator.

QQ plots of four representative trials are shown in Figure 4.10. The fit of the mismatched estimators worsens as the refractory effect grows, however the matched estimator fits consistently with the data. The deviation above the mid-line for the mismatched estimators indicate that the rate estimates would produce a greater number of short inter-arrival times then the data contains. This is precisely the trend we expect since these estimators

ignore some or all of the refractory effect. The mismatched estimators also give a worse fit for more intense excitations, indicating that the refractory period has a higher impact on the free firing rate estimate when the response is strong.

We also used this data set to address the pseudo-ML suggestion in Chapter 3. Figure 4.11 shows three histograms showing the estimates of Δ_r from the simulation across all excitations. Frames A,B, and C break these estimates up according the the β value used to generate the data. We see that as the relative refractory period grows, the accuracy of the Δ_r estimate slightly decreases. It was suggested that a composite estimate could be made by taking the minimum estimate Δ_r from all trials, and from the figure we can see that this would result in the correct value of Δ_r being estimated.

Figure 4.12 shows histograms of the β estimates over all trials with similar true values. We have plotted $5/\beta$ instead of simply β because it has a more intuitive meaning, namely the length of the relative refractory period. The resulting distributions are quite well centered, except in frame A where it is bound by zero. We conclude the following. As the relative refractory period grows, so does the uncertainty about it. This can be seen in the spreading of the distributions as the relative period is extended. The histograms also shows the trial-to-trial estimates are slightly biased to values of β corresponding to relative periods less than the true length. A possible explanation is that our seed starts with large β, corresponding to a very short relative period. If the likelihood surface is flat around the true maximum, the Newton search may quit early. Since we always start with large β this results in always stopping closer to a short relative period when the surface is flat about the true max. Alternatively, the ML estimate of β may simply be biased for finite data sizes. Averaging according to (3.6) gave estimates of 5832.4, 1141.8 and 660.3 when the true values were 2500, 833 and 500. These estimates correspond to refractory periods with length .85, 4.4 and 7.6ms. Averaging directly on the single trial β estimates did not yield

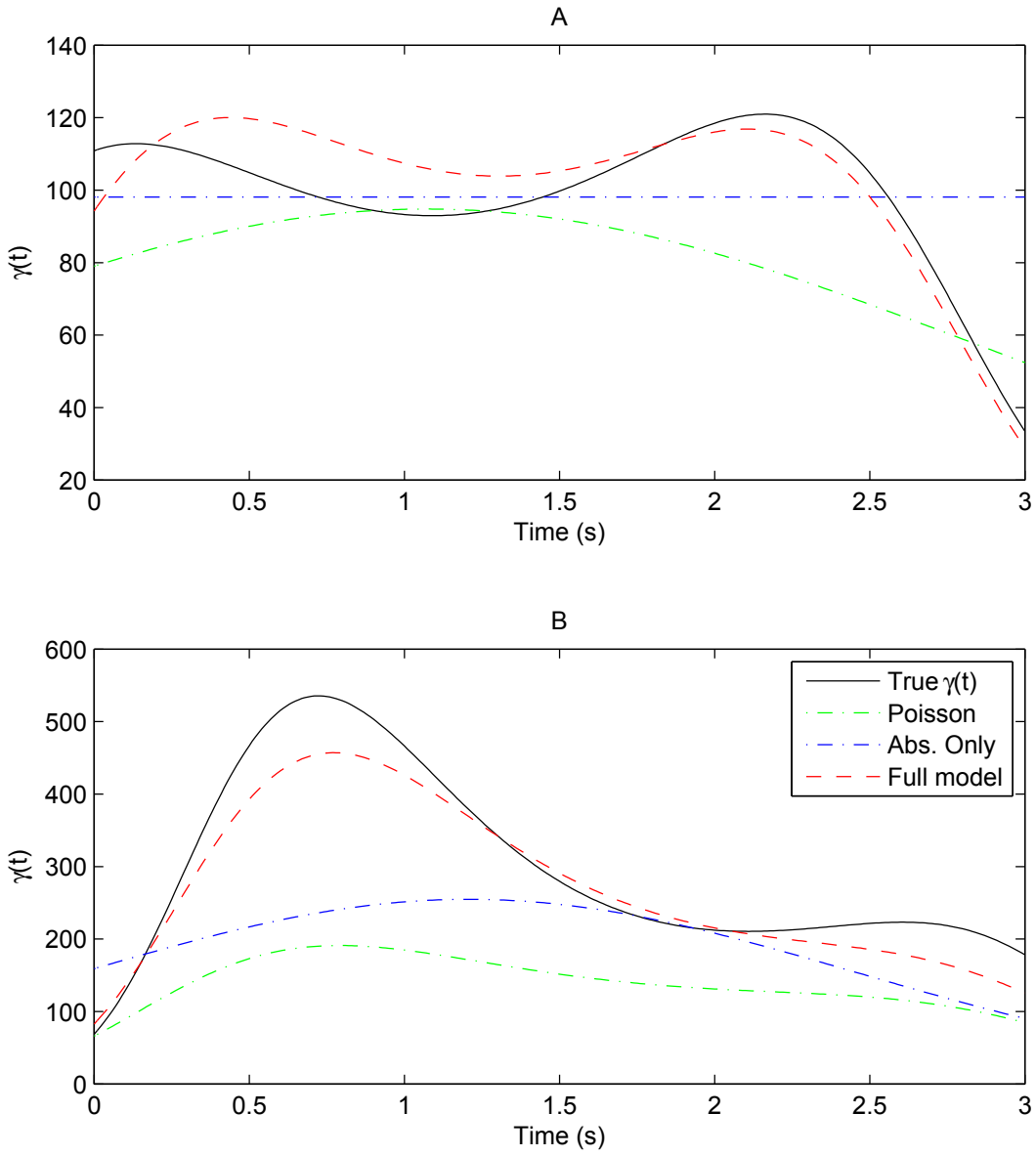

Fig. 4.9 Single trial free firing rate estimates for $\gamma(t)$ of the form (2.2). Trial data was generated using $\beta = 2500, 500$ and $\bar{\gamma} = 100, 300$ in frame A and B respectively. The absolute was the same in each frame $\Delta_r = .002$.

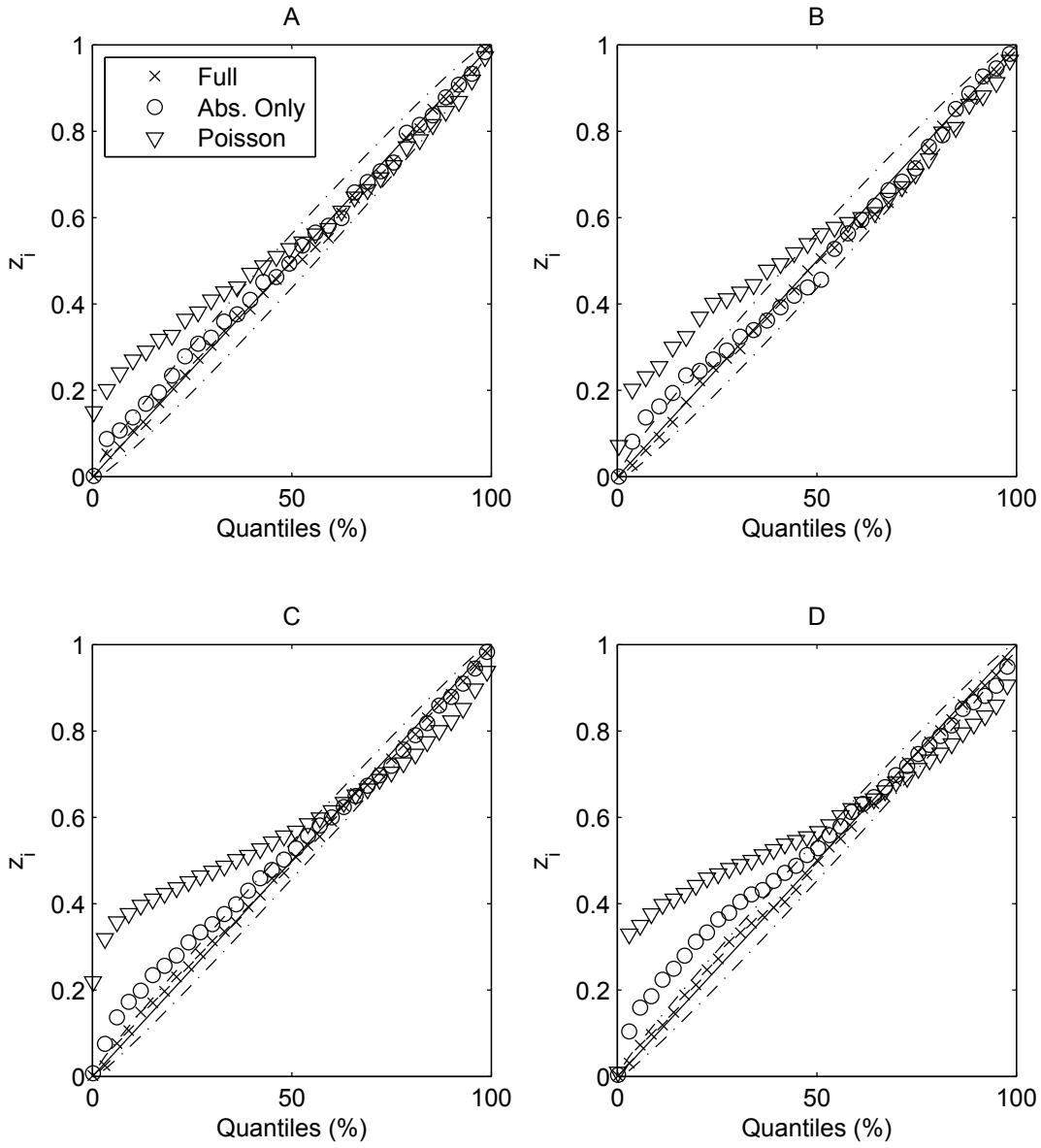

Fig. 4.10 Resulting QQ plots from free firing rates with the form (2.2). Four example trials are shown with $\beta = 2500, 866, 2500, 500$ and $\bar{\gamma} = 100, 100, 300, 300$ in frame A,B,C and D respectively.

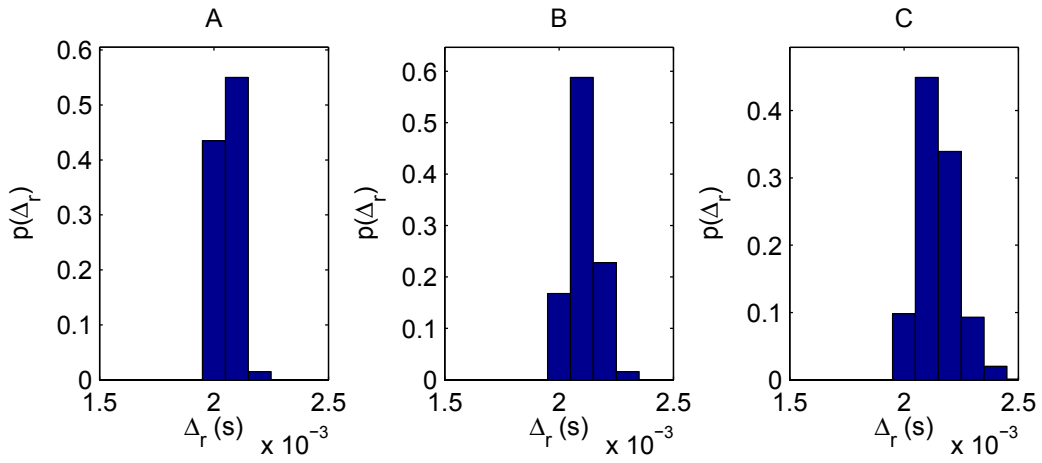

Fig. 4.11 Histograms Δ_r estimates from trials with varying β. The true absolute period $\Delta_r = .002$ while β varies from $2500, 866, 500$ in frame A, B, and C respectively.

a meaningful composite estimate. This is due to the fact that single trials for which β was estimated to be very large add too much weight to the final average value. We conclude that (3.6) is a better way to average single trial estimates of β.

Finally we investigated the accuracy of the alternative maximization approach discussed in Section 3.3.4. In this test we used a vector of β seeds corresponding to relative refractory periods of zero to twenty ms in half ms increments. The NMISE distribution for both the filled function and alternative approach is shown in Figure 4.13 for data from a randomly selected excitations from the simulation described earlier in this subsection. The error distribution for both techniques is quite similar, with their means virtually identical, 2.277 and 2.278% for the filled and alternative method respectively. Investigation into other free firing rates showed similar results. This indicates that, for searches in high parameter spaces, the alternative method may be preferred from a computational expense point of view. In this work we were not necessarily concerned with computation time and therefore no actual processing time comparison was carried out between the methods.

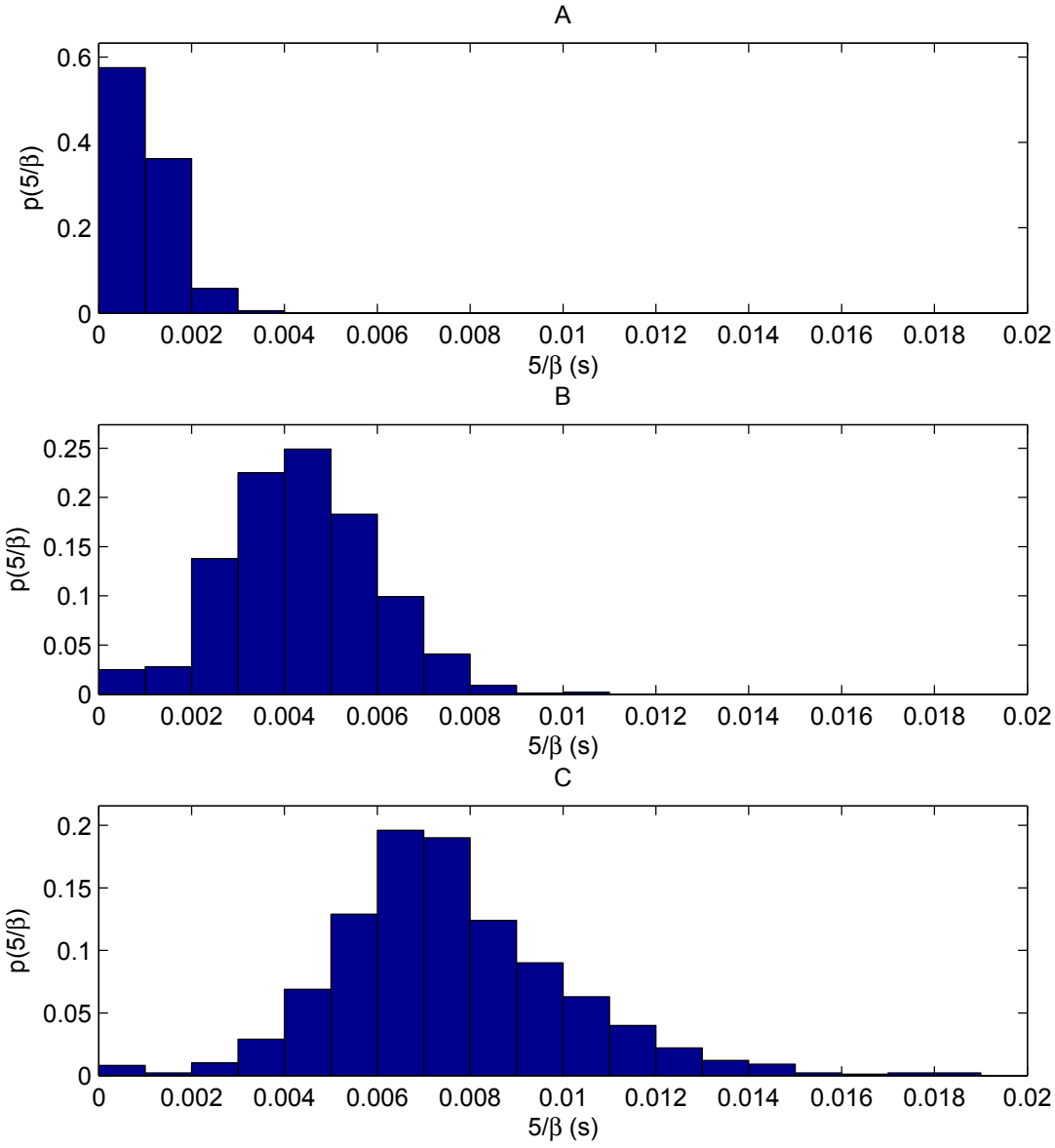

Fig. 4.12 Histograms for $5/\beta$ estimate when it was 2, 6, 10ms in frame A, B, and C respectively.

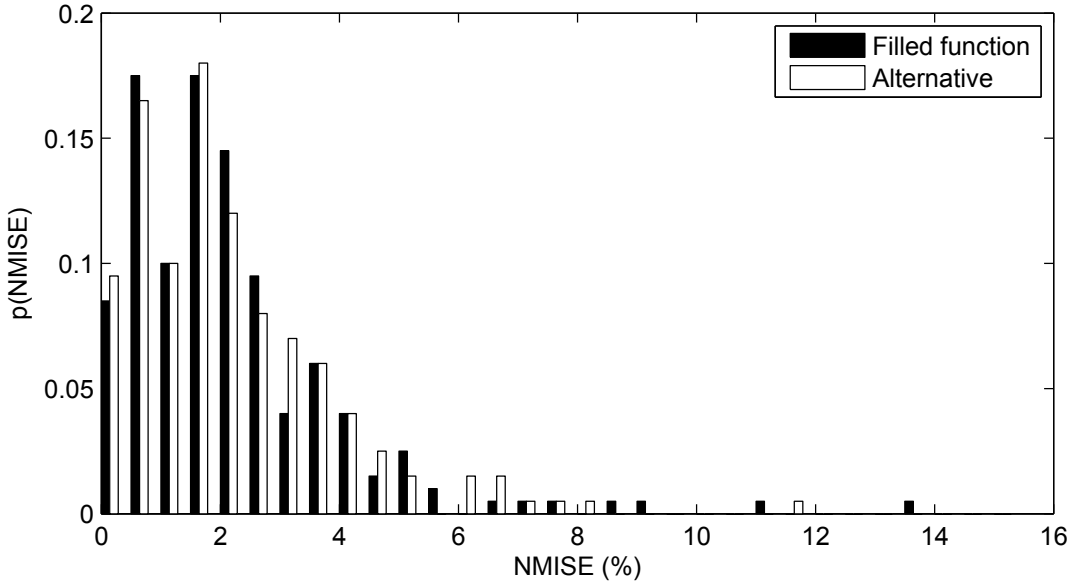

Fig. 4.13 Error distribution resulting from filled function and alternative approach as per Section 3.3.4.

4.4.2 Mismatched free firing rate model

Estimator accuracy results on excitation model (4.2) and (4.3) are given in Figures 4.14 and 4.15. In section 4.2 we saw that a good fit to these functions could be achieved with a model order around 6 or above so, as with the previous simulations, we limit $\mathcal{R}$ to contain models up to $r = 10$. Exact parameter values used for the excitations in each figure is given in Appendix C Table C.3.

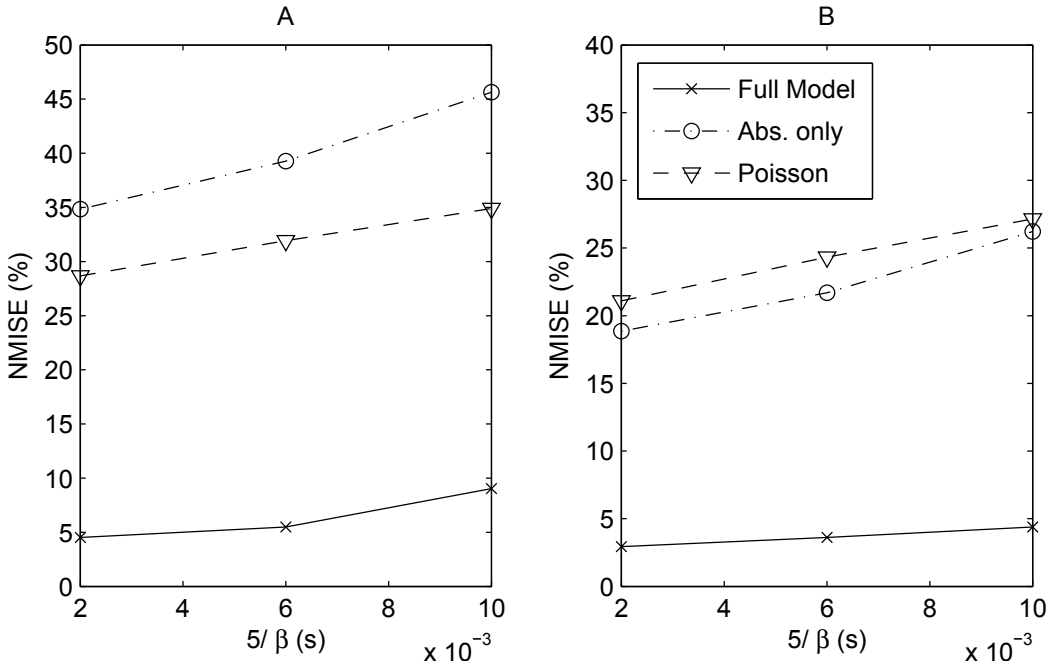

Fig. 4.14 NMISE simulation results based on synthetic data with a refractory period generated by a excitation with the form (4.2). Free firing rates used had $\bar{\gamma} = 200, 300$ in frame A and B respectively.

The NMISE scores for all three estimators are typically worse for the excutation model (4.2) than they were in the previous section. Although the full estimator performed well overall, the NMISE approached 10% when the relative period was long for the lower average intensity. At the higher intensity tested this value became less than 5%. As higher free firing rate intensities presuppose greater spike counts, we believe this may be the explanation as

opposed to model mismatch.

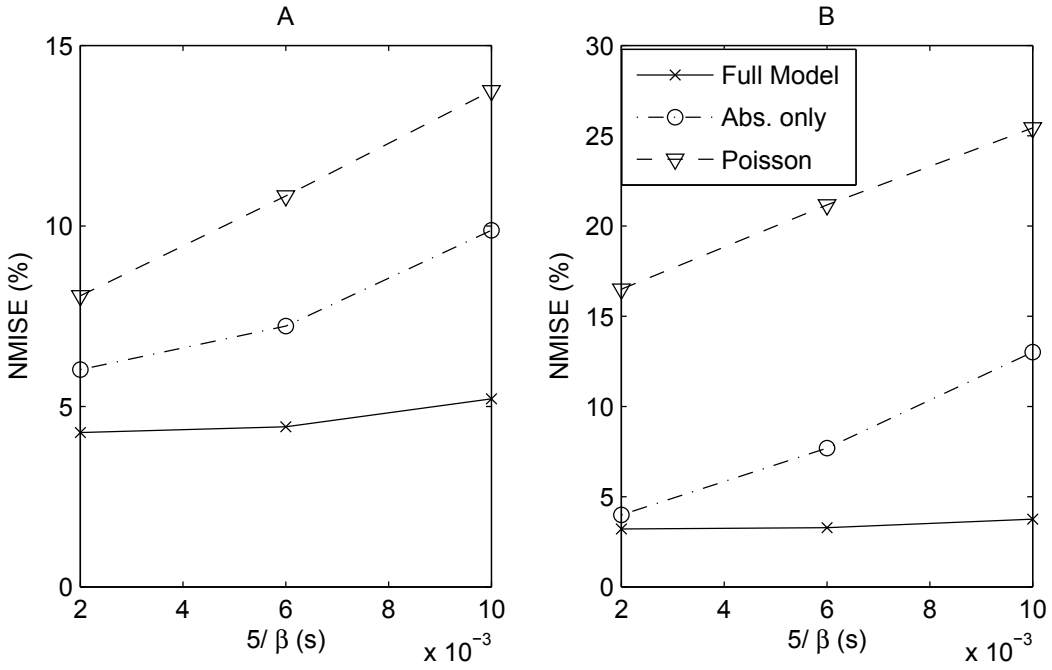

Fig. 4.15 NMISE simulation results based on synthetic data with a refractory period generated by an excitation with the form (4.3). Free firing rates used had $\bar{\gamma} = 100, 200$ in frame A and B respectively.

NMISE performance when the excitation was sinusoidal (4.3) are just slightly higher than those in the matched excitation simulation for the full estimator. We attribute this slight increase in error to the relative complexity of the excitation. If we assume the free firing rate is equivalent to a model of order 6 or 7, as was found to be a good fit, then we can see from Figure 4.6 that these NMISE values are very similar to previous results using excitations of the form (2.2) with such an order.

Two example single trial estimates are shown in Figure 4.16. In frame A we see that only the full estimator captures the peak and general shape of the excitation. In frame B it seems all estimators acceptably represent the sinusoidal nature of the excitation but that the incomplete estimators generally underestimate the magnitude of the intensity. QQ

Estimator	Excitation	$\bar{\gamma}$	NMISE (%) $\beta = 2500$	$\beta = 866$	$\beta = 500$
Poisson	(4.2)	200	28.67	31.93	34.9
Abs. Only	(4.2)	200	34.83	39.28	45.63
Full	(4.2)	200	4.52	5.48	9.04
Poisson	(4.2)	300	21.10	24.33	27.14
Abs. Only	(4.2)	300	18.86	21.70	26.21
Full	(4.2)	300	2.94	3.62	4.39
Poisson	(4.3)	100	8.06	10.83	13.74
Abs. Only	(4.3)	100	6.02	7.22	9.88
Full	(4.3)	100	4.27	4.44	5.21
Poisson	(4.3)	200	16.50	21.17	25.43
Abs. Only	(4.3)	200	3.99	7.69	13.00
Full	(4.3)	200	3.20	3.28	3.75

Table 4.3 Numerical values for results plotted in Figures 4.14, 4.15.

plots for two single trials are shown in Figure 4.17. The trends are similar to those in the first simulation, with the full estimator reporting an excellent fit to the data. We note the correlation between lowest NMISE and best fit when comparing the three estimators.

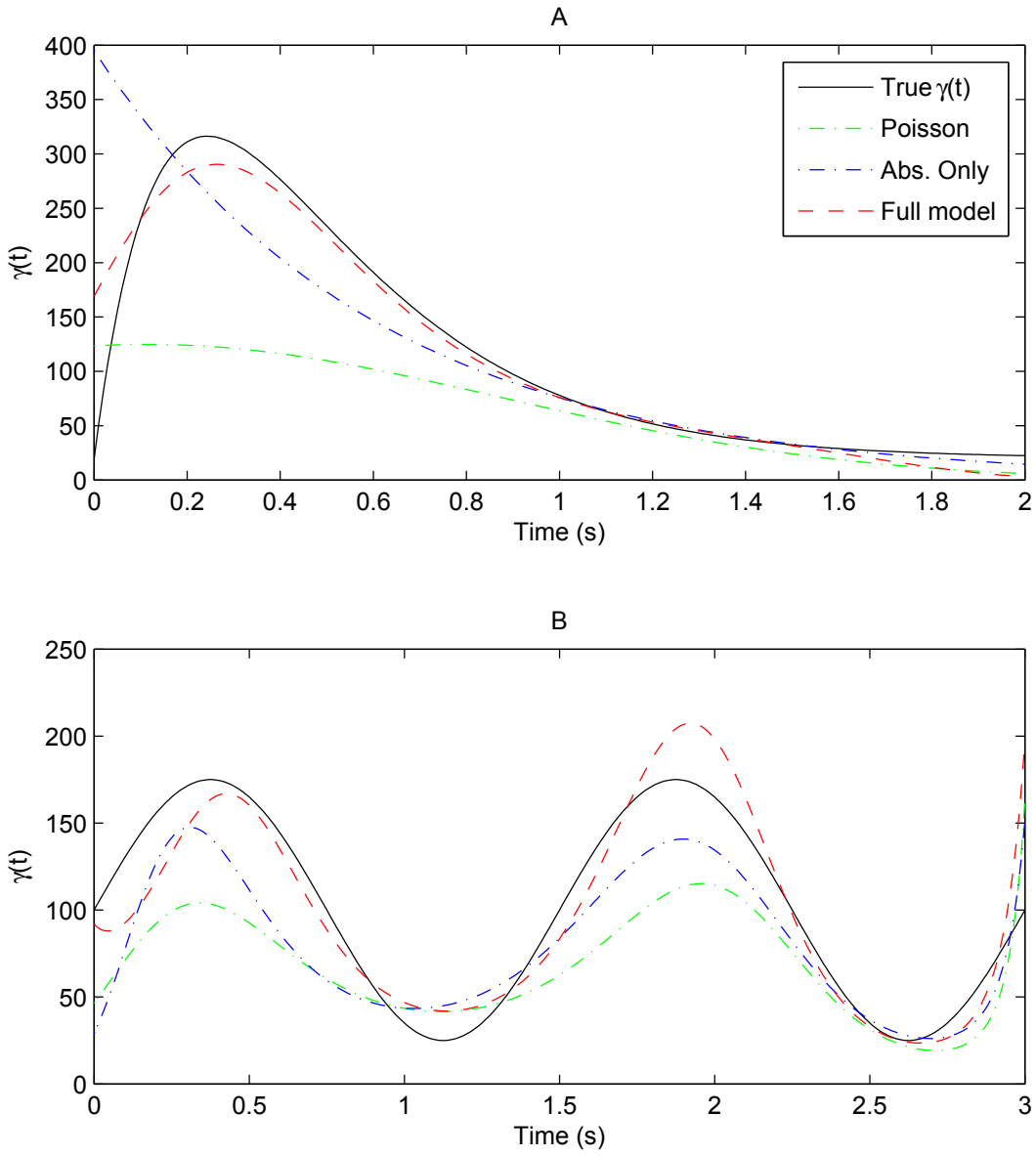

Fig. 4.16 Single trial free firing rate estimates for $\gamma(t)$ of the form (4.2) and (4.3). Trial data was generated using $\beta = 866,500$, $\bar{\gamma} = 200,100$ and $\Delta_r = .004, .002$ in frame A and B respectively.

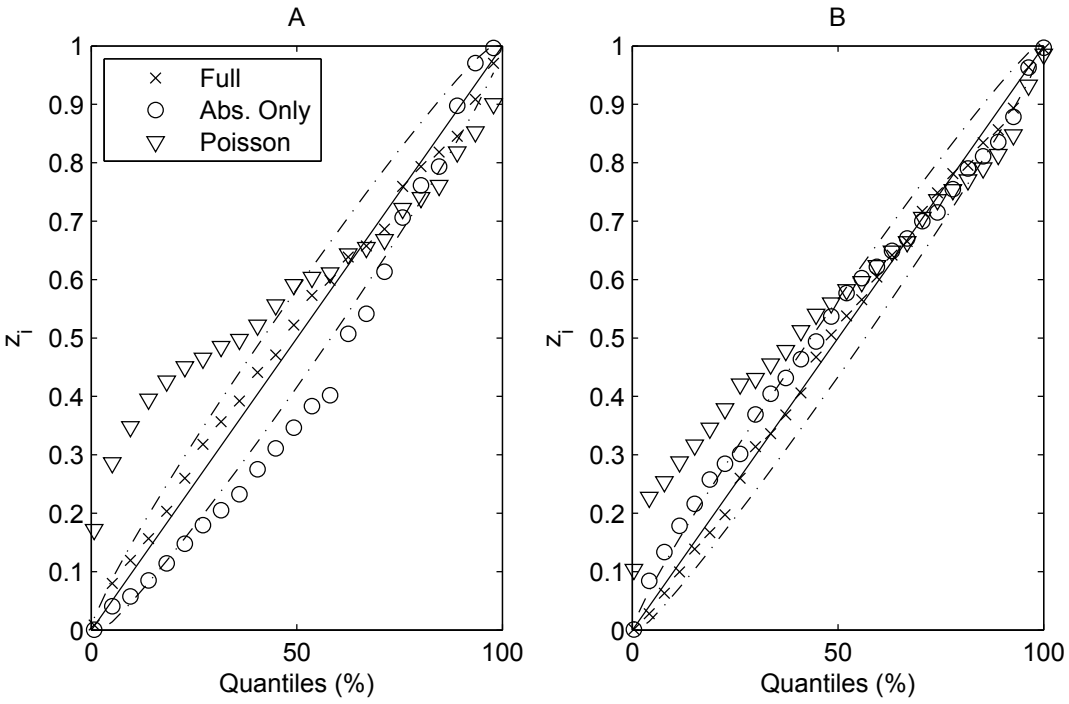

Fig. 4.17 Frame A shows a QQ plot from a trial when $\gamma(t)$ was of the form (4.2). Frame B shows a result from when $\gamma(t)$ was of the form (4.3)

4.5 Real data analysis

To validate our refractory model, goodness of fit to real data was undertaken. In this analysis the exact form of $\gamma(t)$ was not known, however, QQ plots suggest the model fits the data quite well. In combination with what we saw in the previous simulations, we expect that such a fit can only be achieved by a reasonably accurate free firing rate estimate.

Our first data set comes from an experiment described in [52], a V4 neuron in a macaque monkey monitored while it was subjected to various visual stimuli. Figure 4.18 shows the ISI distribution of this neuron over several trials. We can see that it bears a resemblance to the ISI distributions in frame C of Figure 2.4. Figure 4.19 shows two QQ plots for single trial estimates. The full model gives an excellent fit, particularly at the low quantiles which suggests our refractory model is in very good agreement with this measured data. Deviation of the simpler models in these frames is less than most of those in our simulation. We believe this is caused by the average intensity and overall spike count of the data being lower than those simulated (about $\overline{\gamma} = 45$), and the observed absolute refractory period shorter, $\Delta_r = 1.6, .8$ms in frame A and B respectively. The β estimates corresponded to relative refractory periods of only 1.6 and 2.3 ms. Even with such short refractory periods, our model is able to show an improved fit.

A second set of measurements as described in [53] was analyzed in a similar way. This set consisted of neurons from primary motor and pre-motor cortices of a rhesus monkey monitored while performing reach tasks. ISI distributions from four neurons of the set are shown in Figure 4.20. Their shapes are also reminiscent of Figure 2.4 and Figure 4.18 but with much shorter tails. This indicates that the responses are relatively intense. Corresponding QQ plots are shown in Figure 4.21. The distributions in frame A, B, C and D of Figure 4.20 are for the QQ plots of the same frame in Figure 4.21. These results

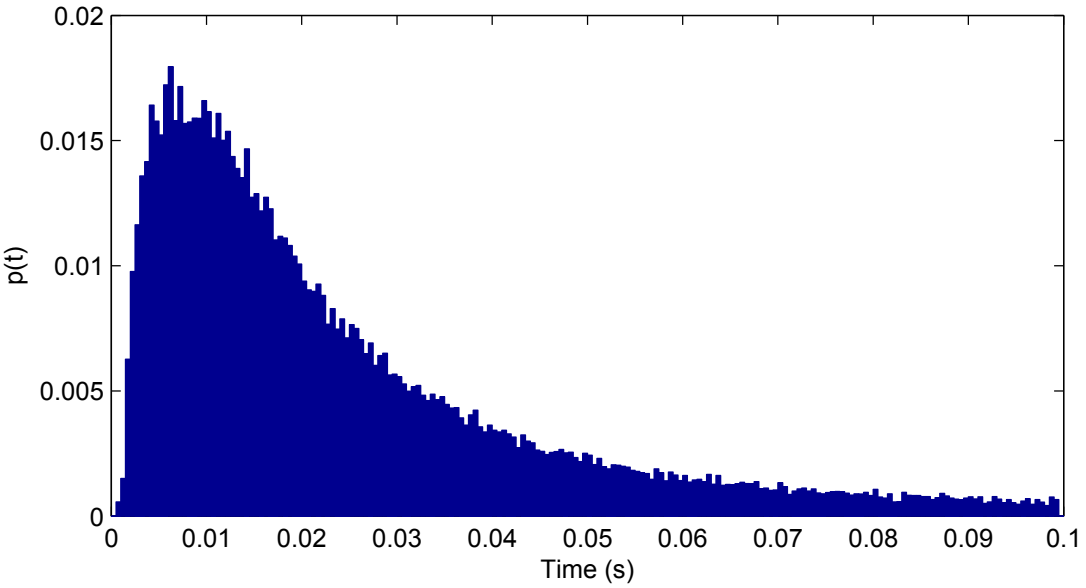

Fig. 4.18 ISI distribution for a V4 neuron from a macaque monkey observing various stimuli.

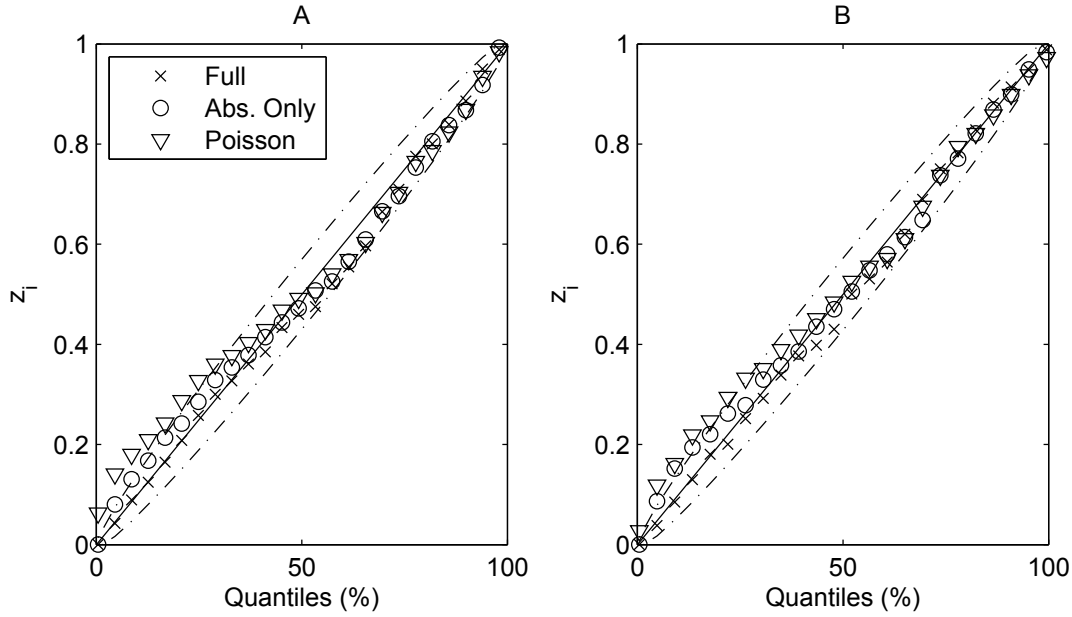

Fig. 4.19 QQ plots from two trial estimates using experimental data from a V4 neuron. Corresponding ISI distribution can be seen in Figure 4.18.

also favour our refractory model, with the low quantiles being quite close to the center-line. The difference between estimators is quite pronounced, more so than the simulation results which we find quite possible given that full estimation scheme pointed to an absolute refractory period on the order of 10ms and relative period of 70ms or more. Full details are given in Table 4.4.

Not all measured data investigated followed the gains in goodness of fit presented above. We analyze one case where this is so. This data set comes from an audio cortex neuron of a cricket and has been extensively discussed in [54]. The neural data used in this analysis is known to contain excessive bursting. Bursting is a phenomenon exhibited by certain types of neurons which tend to fire in short clusters of spikes. Since this phenomenon is not captured by our model, we wished to examine the fit of resulting estimates. The ISI distribution for bursting neurons can be characterized by its bi-modal shape. Figure 4.22 frame A shows this feature. Frame B shows a QQ plot for one trial. We at the lower quantiles the fit is quite poor and that there is no difference between the absolute only and full estimator. As reported in [54], this data appears to be a sort of temporal code and some scheme to determine which spikes are part of a burst sequence is required to fully decode the stimulus influence on the firing rate. While a bursting neuron may not always reveal itself through the ISI distribution, we can easily assess through QQ plots whether or not the model of the estimator captures the structure of the data and avoid using it on data that is inappropriate.

Figure	$\hat{\Delta}_r$	$\hat{\beta}$
4.21 A	.012	135.69
4.21 B	.009	79.58
4.21 C	.009	142.10
4.21 D	.012	106.65

Table 4.4 Refractory estimates corresponding to QQ plots in Figure 4.20

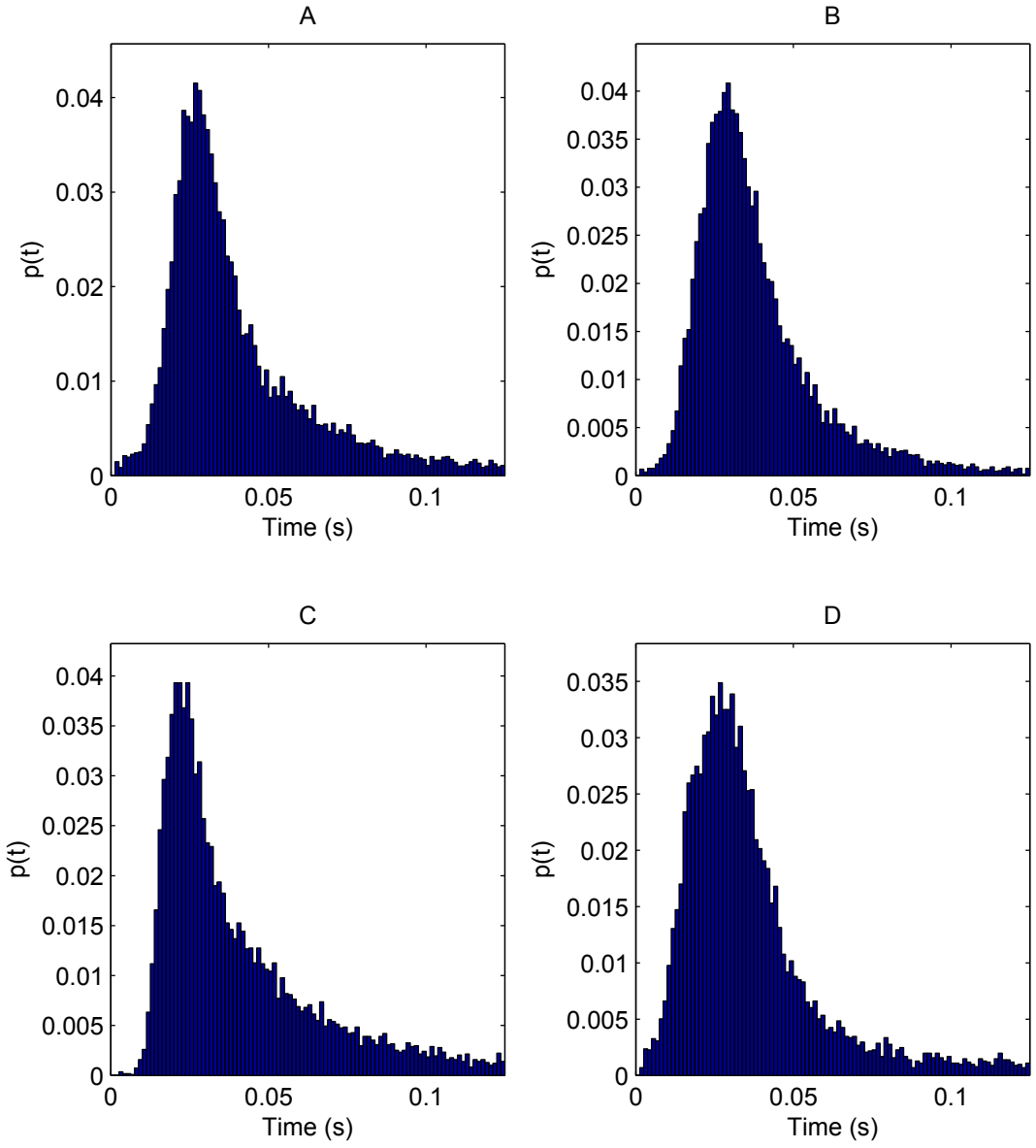

Fig. 4.20 ISI distributions for four motor cortex neurons monitored during a reach task.

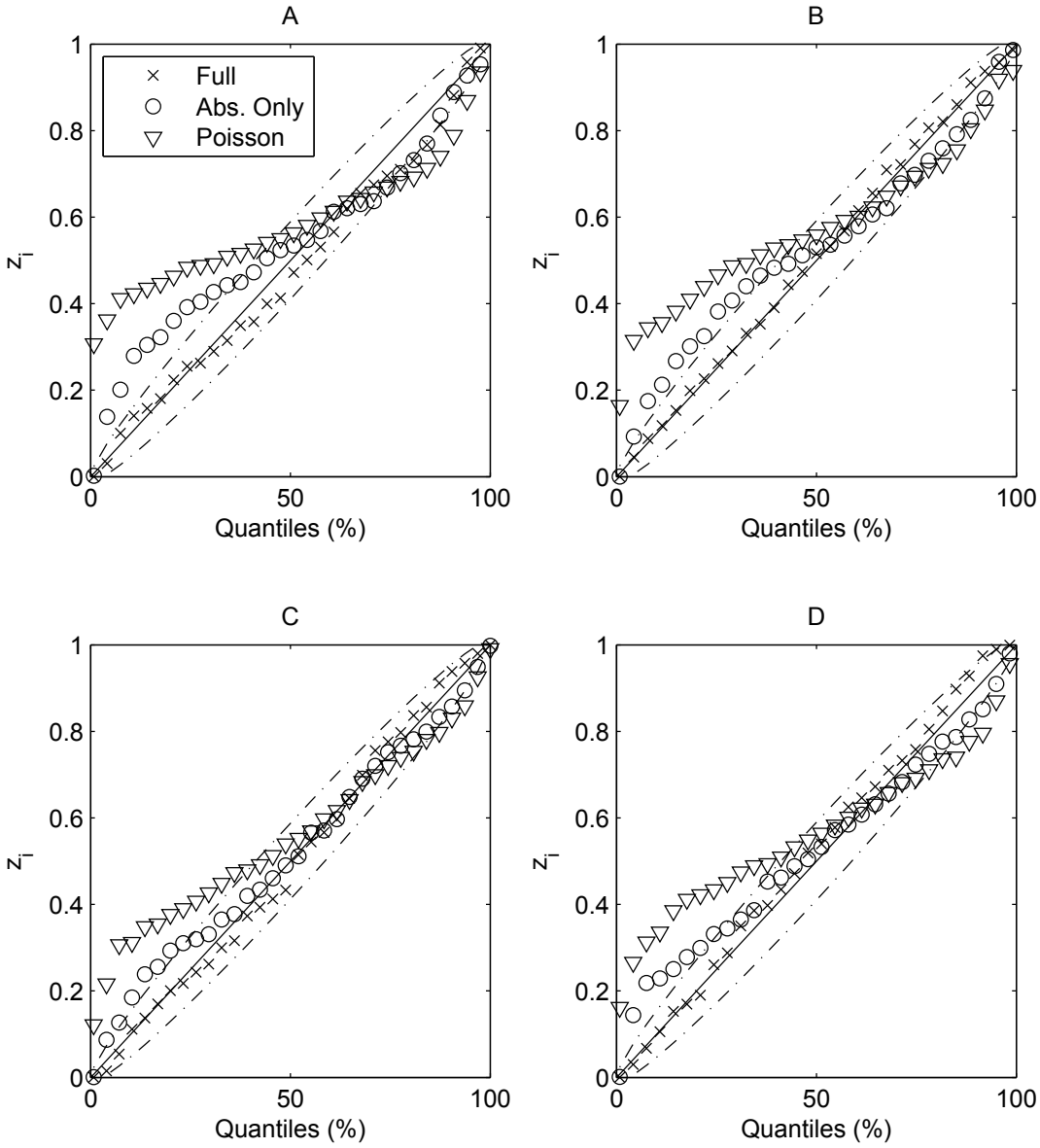

Fig. 4.21 QQ plots for four motor cortex neurons monitored during a reach task. Frames A, B, C and D are fits to the same neurons whose ISI distribution are shown in the same frame in Figure 4.20.

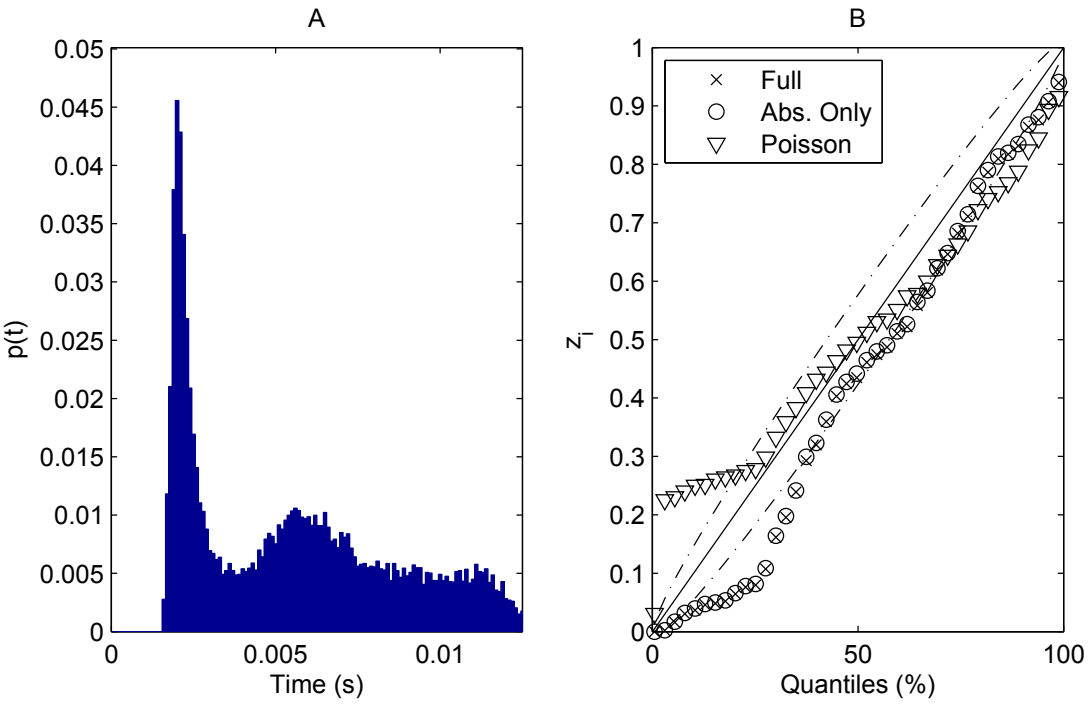

Fig. 4.22 Frame A shows the ISI distribution for a bursting neuron with the bi-modal characteristic. Frame B shows a QQ plot from one specific trial.

Chapter 5

Comparison of rate estimation techniques

Chapter 1 focused on many rate estimation schemes, the majority of which could be classified into two groups. The first would be linear filters, some of which are optimized based on simple spiking models [6] [7] [8]. The second would be those which attempt to account for refractory effects through a history dependent spiking model [18] [14] [17]. In this chapter we present a performance comparison between our proposed estimator and a typical representative from each of the two categories mentioned above. We consider our scheme to fall into the second category since it incorporates a history dependent model for the spike data.

5.1 Linear filter

The linear filter method we tested is taken from [6]. We have selected the Gaussian filter $g(t)$ parameterized by σ, and defined as

$$g(t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-t^2}{2\sigma^2}\right), -\infty < t < \infty. \tag{5.1}$$

The free firing rate estimate $\gamma(t)$ is found by convolving a sequence of impulses at the spike arrival times with (5.1). This leaves only the parameter σ to be determined. In [6] it is suggested to start with a large σ, then incrementally reduce its value. At each increment compute ϵ, the ISE between the current and previous free firing rate estimate

$$\epsilon_{i,j} = \int_{t_0}^{T} (\gamma_i(t) - \gamma_j(t))^2 dt, \tag{5.2}$$

where $\epsilon_{i,j}$ is the ISE between the two estimates $\gamma_i(t)$ and $\gamma_j(t)$. Findings in [6] show that a minimum typically becomes apparent if one plots ϵ over a range of σ, and that this minimum is close to the ideal filter width in terms of ISE between the estimate and the true excitation. We investigated this claim in Figure 5.1 which plots the ISE and ϵ as a function of σ for a single spike sequence realization. Both the minimum of ϵ and the ISE occur at the same filter width tested. We also see that the ISE curve is relatively flat about its minimum, which was typical for most trials. Hence small errors in the choice of optimal bandwidth have little impact in terms of ISE. Although this test seems heuristically derived, it performs well empirically. Since we prefer to focus on single trial performance, the optimization of σ is repeated at each trial. Our final implementation is described in the following algorithm.

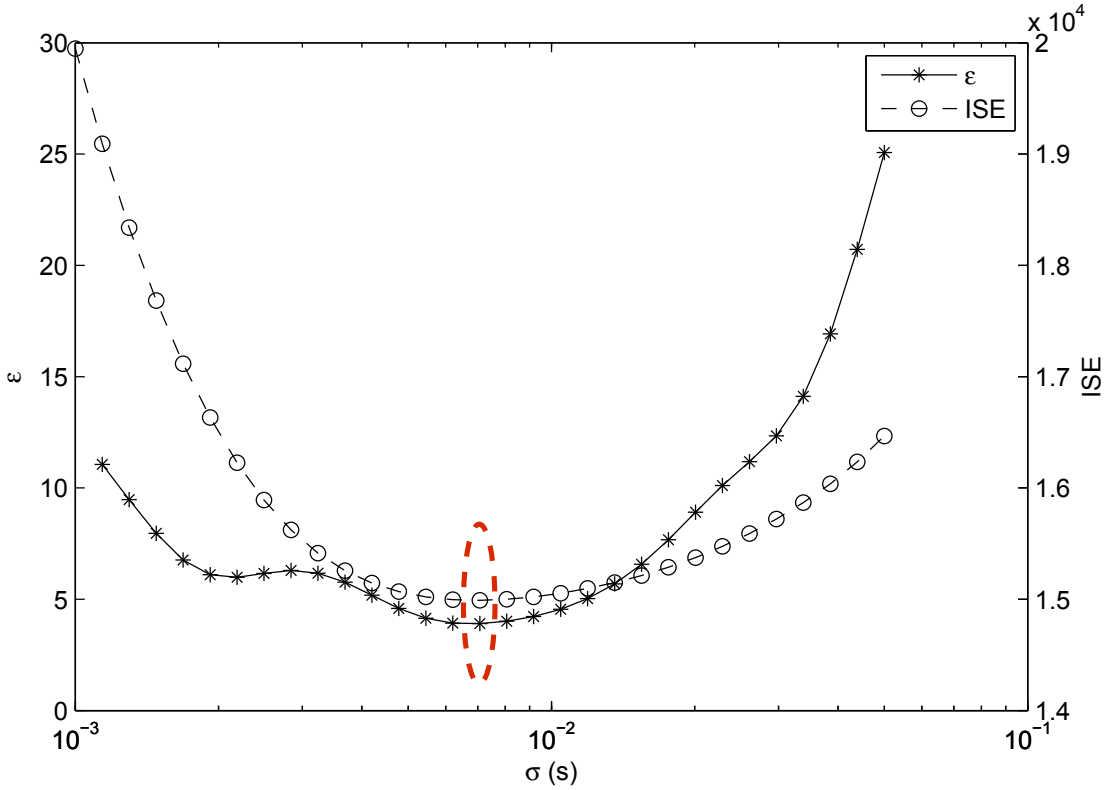

Fig. 5.1 ISE and ϵ plotted as a function of filter width σ. Line markers show actual filter widths investigated. Red ellipse highlights the fact that the minimum of each series occurs at the same point, with $\sigma \approx 7ms$.

Linear filter scheme:

1. Select a vector of candidate filter widths $\Omega = \sigma_1, \sigma_2, ..., \sigma_K$. Proceed to next step.

2. Calculate all excitation estimates $\gamma_k(t) = \sum_{i=1}^{N(T)} g_k(t - t_i), k = 1, 2, ..., K$. Proceed to next step.

3. Calculate the incremental ISE ϵ of the excitation estimates by $\epsilon_k = \int_{t_0}^{T} (\gamma_k(t) - \gamma_{k-1}(t))^2 dt, k = 2, 3, ..., K$. Proceed to next step.

4. Select $\gamma_l(t)$ as the final estimate where $l = \arg\min_{k \in K, k \neq 1} \epsilon_k$.

5.2 Generalized linear modelling

GLM extends the notion of linear regression. In this section we outline how to construct the design matrix such that the GLM method can be used in the context of rate estimation. We suggest [24] for a more in depth discussion of all the mathematical details involved. Using a Poisson regression, the rate can be estimated if it is expressed as the exponential of a linear combination of time varying functions

$$\lambda(t) = \exp\left(\sum_{i=1}^{S} \theta_i f_i(t)\right),$$

where θ_i are scaling values of $f_i(t)$, which are generally unrestricted functions of time and S the number constituent functions which form the intensity. In rate estimation it is popular to choose f_i to comprise a set of cubic Basis splines, since such splines can approximate well a large variety of continuous functions and have attractive smoothness properties [17]. A constant offset is also included, and estimated with one additional θ which scales an f_i equal to 1.

In order to incorporate history dependency, additional $f_i's$, that are functions of past spiking activity, are included. To achieve this we move the model into the discrete time domain. The discrete spike train is $\mathcal{T}$ binned such that no single bin has more than one spike. We denote this version $\mathcal{T}_n$ which equals 1 if a spike is in the $n'th$ bin, zero otherwise. Through the GLM framework a discrete version of the rate $\lambda(n)$ is then estimated where

$$\lambda(n) = \exp\left(\sum_{i=1}^{S} \theta_i f_i(n) + \sum_{j=1}^{L} \theta_j f_j(n-j)\right), \tag{5.3}$$

this allows any spikes that have occurred as many as L bins ago to influence the rate. The $f_j's$ themselves are indicator functions which equal 1 if a spike occurred at sample $n-j$,

and zero otherwise. Thus we now have a rate which depends not only on a time varying excitation, but also past spiking history.

The main issues raised are how many cubic basis functions to use, how to space them, and how much history to look back on. In our implementation we used uniform spacing for the splines, which seems logical given no prior information. A model order selection tool can then be used to determine the size of the basis, similar to the approach used in selecting r in our proposed method. The time lags were selected such that a period at least as long as the known refractory values was considered by the estimator.

We then use the GLM framework to find the ML estimate for $\lambda(n)$, which can be rewritten as

$$\lambda(n) = X\theta, \tag{5.4}$$

where X is the design matrix and θ is the parameter vector. The design matrix X is an $n-$by$-j$ matrix where n is the sample length of $\mathcal{T}_n$ and j the number of parameters to be estimated in the rate model. The $j'th$ column of the design matrix represents f_j. Those constituent functions which correspond to the spline basis are then discretized cubic splines shifted appropriately in time. The columns representing history dependency become time lagged versions of $\mathcal{T}_n$. A GLM fit will find

$$\arg\max_{\theta} \mathcal{L}(\lambda(n)|\mathcal{T}_n). \tag{5.5}$$

We note that the free firing rate $\gamma(t)$ is assumed to be the resulting spline basis estimate. The full procedure we implemented is described by the following algorithm.

GLM scheme:

1. Bin $\mathcal{T}$ such that no bin contains more than one spike, typically bins of 1 or 2ms are adequate. Proceed to next step.
2. Select a set of K spline basis, each containing $k = 1, 2, ..., K$ functions. Select a number of samples for which the spike data is history dependant. Proceed to next step.
3. Construct the design matrix X using appropriately lagged versions of the binned spike train $\mathcal{T}_n$ and spline Basis k. Proceed to next step.
4. Solve for the scaling coefficients θ using a GLM fit with exponential log link function (Poisson regression). If $k \leq K$ go to step 3.
5. Compare all k excitation estimates using AICC. The model which maximizes AICC is chosen as the final estimate.

5.3 Performance comparison

Accuracy of the algorithms discussed in the previous subsections was compared to our proposed scheme. The simulated spike data used was the same as that used to investigate excitation model mismatch in Section 4.4.2. The NMISE results for excitations of the form (4.2) and (4.3) are shown in Figure 5.2 and 5.3 respectively. Numerical values for the data points in each plot are given in Table 5.1.

All results indicate that the linear filter is inferior. This was expected since it does not account for any refractory effects. Interestingly though, we see that the NMISE results listed in Table 5.1 for the linear filter are almost identical to those of the ML Poisson estimator

in Table 4.3. This may add weight to the claim in [6] that the filter width selection scheme is somehow optimal, as its performance is similar to our Poisson based ML estimator. The GLM implementation generally does much better than the filter scheme, and performs particularly well on the excitation of form (4.3). On this excitation the NMISE was only a few percent higher than the proposed scheme of this thesis, and its accuracy as the relative refractory period increased was practically constant.

Two sample estimates from the three estimators are shown in Figure 5.4. In Frame A we see that the linear filter completely underestimates the peak of the response. This was typical for all trials. The GLM based approach does a reasonable job but still underestimates the peak somewhat. Frame B shows a similar situation. It was not typical across all trials for the GLM estimator to underestimate, but, in general it had a greater variance in trial-to-trial results. This variability may be explained by the more general refractory model which requires estimating more coefficients from the same amount of data as our proposed scheme, which uses only a two parameter refractory model.

			NMISE (%)		
Estimator	Excitation	$\bar{\gamma}$	$\beta = 2500$	$\beta = 866$	$\beta = 500$
Lin. Filter	(4.2)	200	28.19	31.69	34.43
GLM	(4.2)	200	26.41	24.99	30.52
Proposed	(4.2)	200	4.52	5.48	9.04
Lin. Filter	(4.2)	300	21.05	24.27	26.95
GLM	(4.2)	300	8.36	9.13	12.02
Proposed	(4.2)	300	2.94	3.62	4.39
Lin. Filter	(4.3)	100	9.09	11.92	14.98
GLM	(4.3)	100	6.15	6.98	6.78
Proposed	(4.3)	100	4.27	4.44	5.21
Lin. Filter	(4.3)	200	17.62	22.29	26.51
GLM	(4.3)	200	7.02	6.63	7.46
Proposed	(4.3)	200	3.20	3.28	3.75

Table 5.1 Numerical values for results plotted in Figures 5.2 and 5.3.

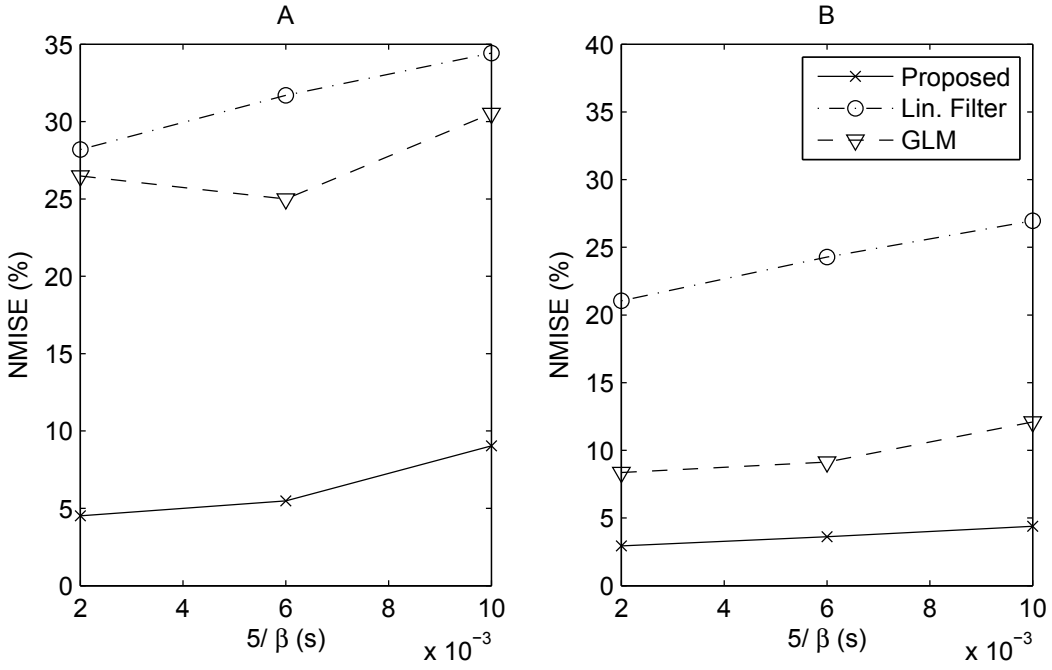

Fig. 5.2 NMISE results from excitation model (4.2). Linear filter, GLM and proposed method compared at $\beta = 2500, 833, 500$.

Figure 5.5 shows goodness of fit results on the same data used in Figure 4.21, frame A and B. We can see that the fit at short inter-arrivals of the GLM estimate is not as good as the proposed parametric scheme on this data, however, it is much better than the linear filter, Poisson, and absolute only scheme. Thus the two best fitting models to this particular data are those which attempt to compensate for the complete refractory effect.

The main advantage of the linear filter method is that it is relatively simple. Unfortunately it does not perform well in terms of NMISE or goodness of fit. Its low accuracy may be explained by the following. From [12] we have that for a Poisson process

$$\int_{t_0}^{T} \lambda(t)dt = E[N(T)], \tag{5.6}$$

the integrated intensity is equal to the expected number of spikes. Under a Poisson spiking

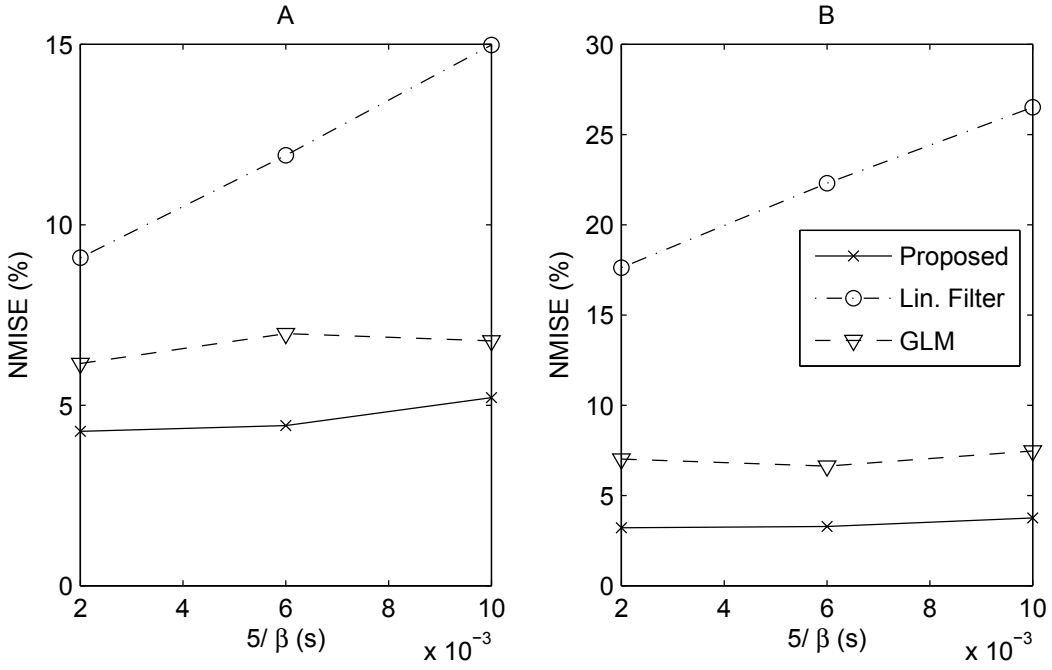

Fig. 5.3 NMISE results from excitation model (4.3). Linear filter, GLM and proposed method compared at $\beta = 2500, 833, 500$.

model $\gamma(t) = \lambda(t)$, so the integrated rate would equal the expected number of spikes. However, in our spiking model with refractory effects $\lambda(t) \leq \gamma(t)$. In Chapter 2 we established that the intensity $\lambda(t)$ is directly proportional to the probability of a spike occurring at time t. This implies that at any time t, the probability of observing a spike with our refractory model is less than or equal to probability of observing a spike from a Poisson data generator driven by the same excitation. This means we expect less spikes to be observed in the presence of a refractory period.

This cripples the performance of any linear filter which subscribes to the restriction $\int g(t) = 1$, since it results in the following relationship between the estimate $\hat{\gamma}(t)$ and the number of observed spikes

$$\int_{-\infty}^{\infty} \hat{\gamma}(t)dt = N(T). \tag{5.7}$$

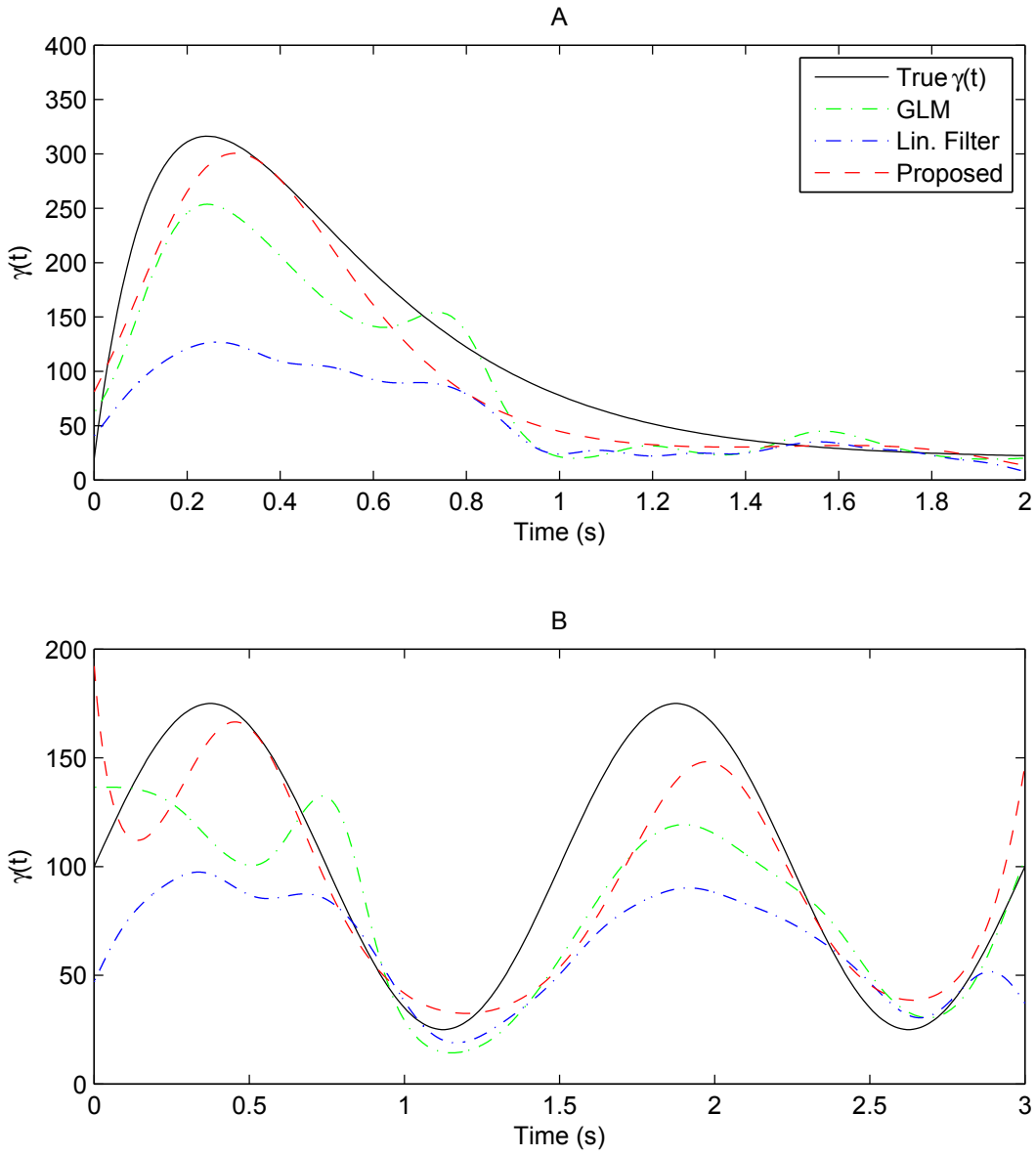

Fig. 5.4 Two sample realizations and their estimates from the linear filter, GLM, and proposed scheme.

But on account of the refractory model

$$E[N(T)] \leq \int_{t_0}^{T} \gamma(t)dt. \tag{5.8}$$

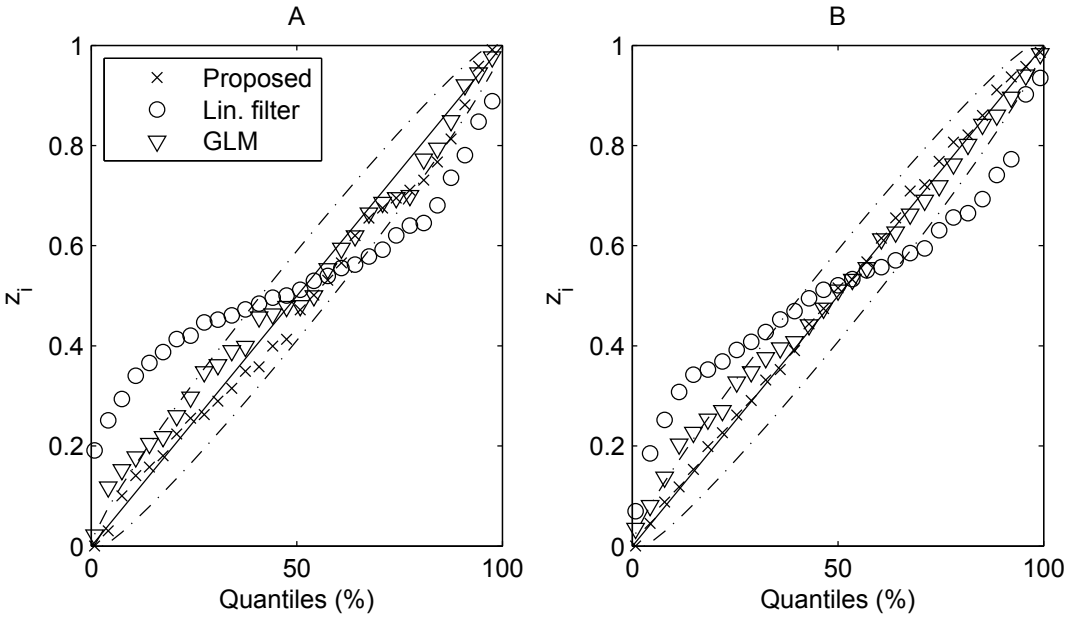

Fig. 5.5 QQ plots for two motor cortex neurons monitored during a reach task. Frames A and B are for the same trials shown in frames A and B of Figure 4.21

.

This could be interpreted as the estimate from the linear filter being guaranteed to have less energy on average than the true excitation, which leads to a biased estimate. Figure 5.6 illustrates this general phenomenon. Using a constant excitation with different $\bar{\gamma}$ and length of one second, we realized one hundred trials of data for four different spiking models. We then plotted the average $N(T)$ against $\bar{\gamma}$. For the Poisson model the relationship is linear, while for the models with refractory effects it is not. In general the disparity between the refractory and Poisson models increases with the excitation intensity, and length of refractory period. This is most likely the explanation of the error growth with relative refractory period seen in Figure 5.2 and 5.3, and the shallow estimates depicted in 5.4. Naturally one may suggest to simply scale the filter $g(t)$ in order to compensate. We note that this is not a simple problem since the scaling would require some estimate of the refractory period, and would have to happen dynamically to be accurate, since the

divergence from Poisson is not linear with intensity.

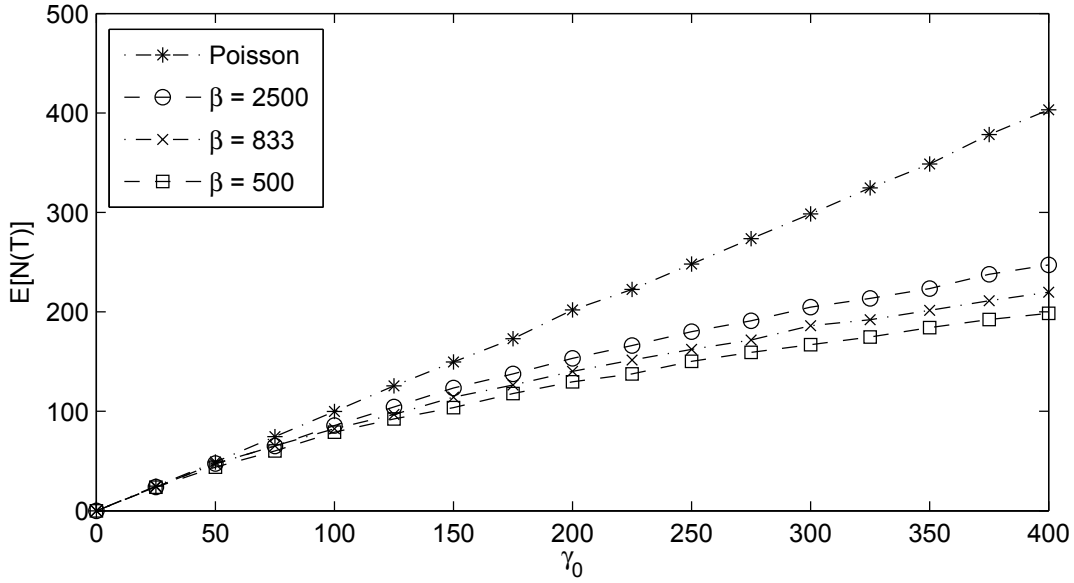

Fig. 5.6 The average number of spike $E[N(T)]$ produced by different spiking models for different excitation intensities. All excitations were constant. We compare models with values of β consistent with our simulations and $\Delta_r = .002$.

The accuracy of the GLM estimator might be improved by considering a scheme which infers more suitable knot spacing for the spline basis. For example, the performance on excitations of the form (4.2), where most of the curvature occurs early in the functions time course, was worse than the performance on excitations of the form (4.3), where the variability is more evenly distributed. This is one of the objectives of the BARS algorithm presented in [9] which attempts to put a prior on the knot positioning. In other works this is done heuristically by first eyeballing a histogram of the data, then selecting appropriate knot positions for the splines[17]. This highlights an interesting difference between the spline basis free firing rate model and the exponential of a polynomial suggestion. The polynomial approach avoids having to decide on knot positions all together, which may explain its more robust performance across the two excitation models tested. However, estimating very high

order polynomials may become computationally inefficient. One can imagine breaking the observation period into shorter intervals, and finding segmented estimates with the proposed scheme. In this case we are faced with a similar problem to knot spacing, how to optimally divide the data? Such a question has been left to future investigation.

Chapter 6

Conclusion

We have presented a parametric ML method for free firing rate estimation in the presence of refractory effects. By assuming suitable parametric forms for both the excitation and refractory period, accurate trial-to-trial estimates were achieved without aid from additional side information. The parametric excitation was tested and shown to fit well to several arbitrary functions using only a few coefficients. Use of a parametric refractory period allowed the construction of a well structured scheme to untangle the free firing rate information from biophysical effects. Our estimator also provides information about these biophysical effects in that by estimating $Delta_r$ and β we gain knowledge of the refractory period.

The log-likelihood function for our model was derived, and methods for its maximization were presented. Non-convex optimization was successfully conducted on the likelihood surface using a filled function approach in combination with a modified Newton-Raphson search. A second maximization approach, which requires only minimal a prioiri information, was also shown to give accurate results. We compared the performance of AIC, AICC and BIC for selecting the model order of our free firing rate estimate. The results clearly

indicated that AICC was most accurate in terms of NMISE. The AIC approach was not much worse, but BIC had a tendency to under-fit and gave very poor results for complex excitations.

Accuracy gains resulting from the inclusion of the full refractory effect in the estimator's model were investigated. This was accomplished by comparing free firing rate estimates between the proposed estimator, a similar scheme with Poisson assumptions, and a scheme which only accounts for an absolute refractory period. In our simulations the full refractory model consistently reduced the NMISE by 75% or more over the Poisson scheme, even when the refractory period was relatively short. When compared to the absolute only estimator, adding the relative period typically reduced NMISE at least 50%. As the refractory period lengthened, the error in the Poisson estimate grew rapidly, while the full refractory estimator was affected only slightly. We concluded that when the refractory phenomena is present in data, it should not be ignored in the excitation estimation process. This conclusion was reinforced by contrasting the performance of our estimator against other examples from the literature. A simple optimized linear filter was shown to be inadequate and insight as to why any linear filter might have trouble was presented. A scheme based on GLM which accounted for refractory affects and modelled the excitation as a sum of cubic basis functions was also investigated. While this method's performance was often far superior to the linear filter, the proposed method of this thesis still produced a lower error. We suggest that this is likely a result of the simpler parametric refractory model employed. Although in some ways less flexible than the mechanism used in the GLM approach, the parametric model shows an excellent fit to real data and is therefore justified. This reduces the degrees of freedom in the estimation process and may explain the lower variance in our proposed estimator's accuracy.

As noted, our refractory model was shown to have excellent goodness of fit to real

data. Improved fit over the Poisson model was seen in both V4 and motor cortices which displayed different refractory period lengths. In contrast, data known not to employ a rate code, and containing bursting, was shown to yield poor goodness of fit. This indicates that our estimator is not suitable for all types of data, which is not surprising as our spike sequence model doesn't account for bursting. Results indicated that the QQ plot was valid as a tool for determining the suitability of data for our proposed scheme.

Future work for this estimation scheme may include further optimization of the global maximum search. The accuracy of the scheme certainly motivates such an endeavour. Another expansion, as discussed briefly at the end of Chapter 5, could be to estimate the free firing rate piece-wise if it was determined to be very complex. An excitation such as (4.3) observed over many periods would certainly require a large value of r to be accurately fit by (2.2). A scheme to divide the data and then stitch the constituent pieces together seems at first straightforward. Owing to the polynomial form of our excitation model it is easy to smoothly interpolate between pieces using low order splines. However, what the most optimal way to split the data remains an open question.

Expanding the method to handle multi-unit recordings is another desirable extension. Similar to the approach of [23], the self exciting model could be adapted to depend not only on its own spiking history, but on the spiking history of other processes. As we used (2.4) to model the refractory effect, functions dependent on the spike times of other neurons could just as easily be combined to produce a model for such dependency.

As was discussed in Section 1.1, the free firing rate may not resemble the true stimulus at higher stages of processing. This suggests two future steps for our scheme. The first would be to obtain and test measured data from first order neurons, such as afferents in the vestibular system, since there is much evidence that these rates relate directly to rotational acceleration and position [48] [16]. Estimating rotational motion from these readings would

help indicate how much accuracy is gained by including refractory effects in the spiking model, since it is relatively straightforward to record and know the true stimulus. A second test would be to use the estimator to reconstruct a more complex stimulus by combining the free firing rates of several neurons and compare the performance with other schemes which do not account for the refractory effect. We discussed two such models in Chapter 1. A result of either investigation may give a more meaningful measurement of error caused by decoding with a model that doesn't account for the refractory phenomenon than NMISE, and allow one to conclude when it is important to account for refractory effects, and when they can be ignored.

Appendix A

Data generation

In our simulations the excitation either took a specific form as mentioned in Chapter 4 or (2.2). For data sets generated with (2.2) random excitations were created by fitting a polynomial of order r to a random vector whose entries were independent uniform random variables with prescribed variance σ^2 and zero mean. The number of entries was one thousand per second of observed time in the observation interval $(t_0, T]$. The exponential of this polynomial was then scaled to ensure a certain $\bar{\gamma}$ was achieved. Finally a check was done to make sure the maximum intensity of the excitation was reasonable. The polynomial fitting algorithm used was $polyfit$ in MATLAB which is based on a least squares fitting procedure [55].

Excitation generation algorithm:

1. Initialize $\overline{\gamma}$, σ^2, r, T, γ_u
2. Generate noise vector with σ^2 over an interval $(t_0, T]$
3. Fit a polynomial of order r to the noise, yielding coefficients $\alpha_0^*, \alpha_1^*, ..., \alpha_r^*$
4. Calculate $\overline{\gamma}^* = \int_{t_0}^{T} \exp\left(\sum_{i=0}^{r} \alpha_i^* t^i\right) dt/(T - t_0)$
5. Set $\alpha_0 = \alpha_0^* + \log\left(\overline{\gamma}/\overline{\gamma}^*\right)$ and $\alpha_i = \alpha_i^*, i = 1, 2, ..., r$. $\gamma(t) = \exp\left(\sum_{i=0}^{r} \alpha_i t^i\right)$
6. If $\arg\max_{t\in[t_0,T]} \gamma(t) > \gamma_u$ go to step 2, else deliver $\gamma(t)$.

Once satisfactory $\gamma(t)$ are found, data can be synthesized. A procedure which allows us to easily include refractory effects in the data is the thinning algorithm, noted as algorithm 6 in [56]. The algorithm is based on the following theorem. Suppose $t_1^*, t_2^*, ..., t_n^*$ are random variables representing event times from the non-homogeneous Poisson process with rate function $\lambda_u(t)$, defined over $(0, T]$. Let $\lambda(t)$ be a rate function such that $0 \leq \lambda(t) \leq \lambda_u(t)$ for all $t \in (0, T]$. If the $i'th$ event time t_i^* is independently deleted with probability $1 - \lambda(t)/\lambda_u(t), i = 1, 2, ..., n$, then the remaining event times form a non-homogeneous Poisson process with rate function $\lambda(t)$ over $(0, T]$ [57].

While this theorem refers to non-homogeneous Poisson processes, we note that any data set $\mathcal{T}$ from a self exciting process can be viewed as conditionally Poisson based on the corresponding realization of $\lambda(t)$ [12] [58]. We also note that our refractory effect only scales down $\gamma(t)$ such that $\gamma(t) \geq \lambda(t)$. One simple suggestion for the upper bounding function $\lambda_u(t)$ is then simply a constant $\lambda_u = \max \gamma(t)$. The data is then generated as follows.

Spike data generation algorithm:

1. Initialize $t = t_0$, $i = 1$.
2. Draw an independent random variable u_1 from a uniform distribution with support over $(0, 1)$.
3. Set $t^* = t - 1/\lambda_u \log u_1$. This yields the next candidate arrival time generated a process with rate λ_u. If $t^* > T$ exit algorithm and deliver $\mathcal{T} = t_1, t_2, ..., t_{i-1}$, else go to next step.
4. Draw a second independent random variable u_2 from a uniform distribution with support over $(0, 1)$. This variable helps dtermine whether or not the candidate should be included or discarded. Go to next step.
5. If $u_2 \leq \lambda(t^*)/\lambda_u$ then $t_i = t^*$ and increment i, else the candidate arrival t^* is ignored. In either case set $t = t^*$ and go back to step 2.

Note that the arrival times t_i are found sequentially in the data generating algorithm. From (2.5) we have the explicit form of $\lambda(t)$ which allows us to calculate its value based on the free firing rate $\gamma(t)$ and the most recent spike. Thus the above algorithm can easily be used to generate spike sequences for a given stimulus and the refractory model presented in Chapter 2.

Appendix B

MATLAB script details

This appendix outlines the MATLAB scripts used to generate our results. These scripts may be obtained from the author by request. In total there are six estimation schemes implemented, located in separate folders in the master *SIM* directory. Four are based on the exponential of a polynomial model with no refractory, absolute only and two schemes including full refractory effects, one solves using a filled function and the other with the alternative approach outlined in Section 3.3.4. Each estimator is composed of many files. We explain only those requiring user intervention here.

Each technique can be launched by running the *xxxMaster0.m* where *xxx* is one of the prefixes in the following table. The *xxxMaster0.m* files are contained within the *code* directories found in each estimators main folder.

The *Master0* script also chooses the data set to be inspected. These are held in the *dataSets* folder within the *fullFill* directory. Opening the *Master0* scripts will reveal a line of code *load('xxxNameArray')* where *xxx* is one of the following.

The specified name array will point the master script to the relevant data sets to load from the *dataSets* folder. The results are saved in the estimators results folder as the name

Prefix	Description
purePoi	proposed scheme with no refractory effect
justAbs	proposed scheme with only absolute refractory effect
fullFill	proposed scheme solved with filled function approach
fullVec	proposed scheme solved according to Section 3.3.4
kern	linear filter method described in Section 5.1
glm	GLM method described in Section 5.2

Table B.1 Simulation style prefixes and their meaning.

Prefix	Description
jPsim	excitation is of the exponential of a polynomial form (2.2)
nawrot	excitation is of the form (4.2)
sinusoid	excitation is of the form (4.3)
fieldTrip	real data from [52]
reach	real data from [53]
burst	real data from [54]

Table B.2 Name array prefixes and their meaning.

of the data set with the estimator prefix and *res* appended. In general the estimator outputs are coefficients specifying the ML estimates of different model orders for each trial. As such these results need to be analyzed further to find the final estimate. This is achieved by running the associated *xxxResAna.m* file with the appropriate name array variable. This creates a new file in the results folder with the data set name, estimator prefix and *ana* appended. The *ana* file contains the associated NMISE for each trial and other data such as model order if appropriate. From this data one can generate all the figures pertaining to accuracy results. Scripts for the figures themselves can be found in the *figGen* folder.

Each technique has an associated parameter script *setLocalSearchParams.mat*. As the name suggest these scripts contain variables which affect the local search methods. The filled function approach has a second script *setGlobalSearchParams.mat* which contains variables related directly to the filled function portion of the scheme. Comments are provided in the script files as to what each variable is used for.

Due to the computationally intense nature of our proposed scheme we have made use of the *Multicore* package available on MATLAB central. Additional processes for the same simulation type can be spawned simply by launching the associated *xxxSlave0.m* script. No modification of the slave script is required when the name array variable is edited in the master. To run additional slave processes the path to the *temp0* folder in the appropriate schemes folder must be accessible by the slave machines. This may require modification of the *slaveDir* variable in the appropriate *xxxMaster0.m* script. For more details on how this package works see [59].

Appendix C

Simulation excitation coefficients

This appendix contains the coefficients used for excitation of synthetic data in our simulations.

r	α_0	α_1	α_2	α_3	α_4	α_5	α_6	α_7
1	4.02	1.07						
1	5.32	-1.65						
1	4.97	-.78						
1	3.42	2.04						
1	5.49	-2.13						
1	3.92	1.24						
2	4.43	-.55	1.27					
2	3.83	4.88	-5.2					
2	5.3	-2.45	1.46					
2	4.43	2.14	-2.77					
2	5.57	-4.04	2.91					
2	3.65	-.32	2.56					
3	5.27	-6.53	15.62	-10.53				
3	2.81	6.86	-2.77	-3.58				
3	3.72	5.44	-9.96	5.83				
3	5.13	-12.25	29.92	-18.12				
3	4.55	.99	-.25	-1.56				
3	4.5	-6.57	23.48	-18.17				
4	2.91	28.93	-129.55	205.24	-105.06			
4	5.28	-7.26	40.21	-77.31	43.68			
4	2.04	44.91	-171.13	224.36	-96.88			
4	1.44	33.06	-112.24	152.44	-71			
4	3.61	12.97	-58.34	97.03	-51.75			
4	5.69	-.27	-19.94	38.45	-20.83			
5	4.24	-35.57	267.36	-679.29	730.78	-285.32		
5	1.79	33.73	-188.47	468.23	-498.76	187.43		
5	4.4	-27.03	211.36	-543.04	577.93	-220.16		
5	4.77	-13.17	114.28	-285.32	263.42	-80.15		
5	3.41	23.49	-114.19	246.07	-252.5	98.44		
5	4.74	-8.62	70.4	-229.71	316.95	-152.13		
6	-1.05	99.61	-642.04	2049.44	-3468.63	2935.91	-968.79	
6	3.09	40.51	-274.96	991.04	-2030.51	2071.03	-800.15	
6	4.27	40.12	-386.04	1442.26	-2559.1	2153.35	-692.06	
6	2.02	84.64	-804.63	3082.82	-5558.77	4734.22	-1537.5	
6	3.42	47.69	-569.73	2561.51	-5145.5	4704.57	-1601.82	
6	2	27.28	-250.07	1243.63	-2777.51	2749.93	-992.04	
7	-.28	42.1	18.21	-1424.21	6272.78	-11717.77	10110.8	-3300.31
7	4.82	39.15	-593.86	2959.53	-7053.07	8885.88	-5740.78	1503.08
7	4.66	-20.53	442.85	-3030.79	9231.52	-14090.12	10595.64	-3131.59
7	3.79	-79.22	1095.03	-5584.03	14080.68	-18757.63	12635.63	-3390.12
7	1.31	108.27	-1010.89	4352.52	-9717.3	11478.63	-6768.29	1559.65
7	-.4	215.13	-2480.11	11863.14	-28947.11	38094.96	-25724.43	6983.68

Table C.1 Parameter values for excitations used in Section 4.3, Figures 4.6 and 4.7. Free firing rate model is the exponential of a polynomial (2.2). Refractory parameters constant at $\Delta_r = .002, \beta = 1000$. Period of observation defined by $t_0 = 0, T = 1$.

r	α_0	α_1	α_2	α_3	α_4
4	3.13	7.0227	-7.867	3.2021	-0.44157
4	4.1585	3.0453	-3.4356	1.3627	-0.18945
4	4.7085	0.27269	-1.2013	0.97081	-0.21503
4	3.273	2.2027	-0.46869	-0.38116	0.1201
4	3.9945	3.9454	-6.2949	3.0756	-0.45174
4	4.4916	0.17814	0.47553	-0.29487	0.019447
4	3.6206	4.4506	-4.9737	2.0194	-0.27323
4	3.9411	1.5276	-1.2338	0.54729	-0.10164
4	4.3258	4.4274	-8.1722	4.6288	-0.81186
4	4.5087	-0.013495	-1.0834	0.84284	-0.14908

Table C.2 Parameter values for free firing rates used in Section 4.4.1, Figures 4.8, 4.9, 4.10 and 4.12 . Free firing rate model is the exponential of a polynomial (2.2). Absolute refractory parameter constant at $\Delta_r = .002$, relative varied $\beta = 2500, 866, 500$. Period of observation defined by $t_0 = 0, T = 3$.

Equation	Δ_r	θ_1	θ_2	θ_3	θ_4	t_0	T
(4.2)	.004	20	200	.3	.2	0	2
(4.2)	.004	10	300	.6	.4	0	2
(4.3)	.002	100	75	$2\pi/3$	0	0	3
(4.3)	.002	200	150	$2\pi/3$	0	0	3

Table C.3 Parameter values for excitations used in Section 4.4.2, Figures 4.14, 4.15 , 4.16 and 4.17 . Free firing rate model as specified in first column. Absolute refractory parameter constant at $\Delta_r = .002$, relative varied $\beta = 2500, 866, 500$.

References

[1] P. Dayan and L. F. Abbot, *Theoretical Neuroscience*. Wiley, 2002.

[2] *Spikes: exploring the neural code*, MIT press, July 1999.

[3] D. H. Hubel and T. N. Wiesel, "Receptive fields of single neurones in the cat's striate cortex," *Journal of Physiology*, vol. 148, pp. 574–591, 1959.

[4] D. H. Johnson, "Point process models of single-neuron discharges," *Journal of Computational Neuroscience*, vol. 3, pp. 275–299, 1996.

[5] H. Shimazaki and S. Shinomoto, "A method for selecting the bin size of a time histogram," *Neural Conputation*, vol. 19, no. 6, pp. 1503–1527, 2007.

[6] M. Nawrot, A. Aertsen, and S. Rotter, "Single-trial estimation of neuronal firing rates: From single-neuron spike trains to population activity," *Journal of Neuroscience Methods*, vol. 94, pp. 81–92, December 1999.

[7] S. Cherif, K. E. Cullen, and H. L. Galiana, "An improved method for the estimation of firing rate dynamics using an optimal digital filter," *Journal of Neuroscience Methods*, vol. 173, pp. 165–181, 2008.

[8] H. Shimazaki and S. Shinomoto, "Kernel bandwidth optimization in spike rate estimation," *Journal of Computational Neuroscience*, no. 29, pp. 171–182, 2010.

[9] I. DiMatteo, C. R. Genovese, and R. E. Kass, "Bayesian curve-fitting with free-knot splines," *Biometrika*, vol. 88, • 2001.

[10] R. E. Kass, V. Ventura, and C. Cai, "Statistical smoothing of neuronal data," *Network: Computation in Neural Systems*, pp. 5–15, January 2003.

[11] J. P. Cunningham, V. Gilja, and K. V. S. S. I. Ryu, "Methods for estimating neural firing rates, and their application to brain-machine interfaces," *Neural Networks*, no. 22, pp. 1235–1246, 2009.

[12] D. L. Snyder and M. I. Miller, *Random Point Processes in Time and Space*. Springer-Verlag, 2 ed., 1991.

[13] R. E. Kass, V. Ventura, and E. N. Brown, "Statistical issues in the analysis of neuronal data," *Journal of Neurophysiology*, no. 94, pp. 8–25, 2005.

[14] M. J. B. II and M. Meister, "Refractoriness and neural precision," *Journal of Neuroscience*, vol. 18, pp. 2200–2211, March 1998.

[15] P. R. Gray, "Conditional probablility analyses of the spike activity of single neurons," *Biophysical Journal*, vol. 7, 1967.

[16] E. R. Kandel, J. H. Schwartz, and T. M. Jessel, *Principles of neural science*. McGraw Hill, 4 ed., 2000.

[17] R. E. Kass and V. Ventura, "A spike-train probability model," *Neural Computation*, no. 13, pp. 1713–1720, 2001.

[18] K. E. Mark and M. I. Miller, "Bayesian model selection and minimum decription length estimation of auditory-nerve discharges," *Journal of the Accoustical Society of America*, vol. 91, pp. 989–1002, February 1992.

[19] S. Monk and H. Leib, "A parametric model and maximum likelihood estimation of neural rate functions," in *25th Canadian Conference on Electrical and Computer Engineering (CCECE 2012)*, 2012.

[20] J. W. Pillow, J. Shlens, L. Pananski, A. Sher, E. J. C. A. M. Litke, and E. P. Simoncelli, "Spatio-temporal correlations and visual signalling in a complete neuronal population," *Nature*, vol. 454, pp. 995–999, August 2008.

[21] P. J. Mineault, S. Barthelmé, and C. C. Pack, "Improved classification images with sparse priors in a smooth basis," *Journal of Vision*, vol. 9, no. 10, pp. 1–24, 2009.

[22] U. T. Eden, "Point process models for neural spike trains," in *Short Course 3 Syllabus*, ch. 3, 2008. Onlne: `http://www.sfn.org/index.aspx?pagename=ShortCourse3_2008` [Dec 1, 2012].

[23] W. Truccolo, U. T. Eden, M. R. Fellows, J. P. Donoghue, and E. N. Brown, "A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects," *Journal od Neurophsyiology*, vol. 93, 2005.

[24] S. N. Wood, *Generalized Additive Models*. Chapman and Hall, 2006.

[25] C. J. MacLean, "Estimation and testing of an exponential polynomial within the non-stationary Poisson process," *Biometrika*, vol. 61, no. 1, pp. 81–85, 1974.

[26] Gerstner and Kistler, *Spiking Neuron Models. Single Neurons, Populations, Plasticity*. Cambridge University Press, 2002.

[27] H. Jeffreys and B. Jeffreys, *Mathematical Physics*. Cambridge, 3 ed., 1966.

[28] A. P. Georgopoulos, A. B. Schwartz, and R. E. Kettner, "Neuronal population coding of movement direction," *Science*, vol. 233, pp. 1416–1419, September 1086.

[29] J. A. Pruszynski, I. Kurtzer, J. Y. Nashed, M. Omrani, B. Brouwer, and S. H. Scott, "Primary motor cortex underlies multi-joint integration for fast feedback control," *Nature*, vol. 478, pp. 387–390, October 2011.

[30] N. C. Rust and J. J. DiCarlo, "Selectivity and tolerance ("invariance") both increase as visual information propagates from cortical area v4 to it," *Journal of Neuroscience*, vol. 30, pp. 12978–12995, September 2010.

[31] W. Bair, C. Koch, W. Newsome, and K. Britten, "Power spectrum analysis of burtsting cells in area mt in the behaving monkey," *Jouirnal of Neuroscience*, vol. 14, pp. 2870–2892, May 1994.

[32] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[33] S. A. Murphy and A. W. van der Vaart, "On profile likelihood," *Journal of the American Statistical Association*, vol. 95, pp. 449–465, June 2000.

[34] H. V. Poor, *An Introduction to Signal Detection and Estimation*. Springer, 2 ed., 1994.

[35] P. Stoica and Y. Selen, "Model-order selection," *IEEE Signal Processing Magazine*, pp. 36–47, July 2004.

[36] P. D. Grunwald, *The Minimum Description Length Principle*. MIT Press, 2007.

[37] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. J. Wiley, 2006.

[38] J. Rissanen, "Modeling by shortest data description," tech. rep., International Federation of Automatic Control, 1978.

[39] H. Akaike, "A new look at the stasitical model identification," *IEE Transactions on Automatic Control*, vol. AC-19, pp. 716–723, Dec 1974.

[40] C. M. Hurvich and C. Tsai, "Regression and time series model selection in small samples," *Biometrika*, vol. 76, p. 297–307, 1989.

[41] J. Rissanen, "Shortest data description length and consistency of order estimates in arma-processes," tech. rep., International Federation of Automatic Control, 1978.

[42] J. Rissanen, "Estimation of structure by minimum description length," *Circuit Systems Signal Process*, vol. 1, no. 3-4, pp. 395–406, 1982.

[43] N. Yanxia and Z. Hengjun, "Generalized filled funcgion for global continuous optimization problems," in *2012 Fifth International Joint Conference on Computational Sciences and Optimiazation*, pp. 362–366, 2012.

[44] W. Wang and Y. Xu, "Simple transformation function for finding better minima," *Applied Mathematics Letters*, no. 21, pp. 502–509, 2008.

[45] C. Ng, D. Li, and L. Zhang, "Global descent method for global optimization," *Society for Industrial and Applied Mathematics Journal of Optimization*, vol. 20, no. 6, pp. 3161–3184, 2010.

[46] J. E. D. Jr. and R. B. Schnabel, *Numerical methods for unconstrained optimization and nonlinear equations*. Society for Industrial and Applied Mathematics, 1996.

[47] E. N. Brown, R. Barbieri, V. Ventura, R. E. Kass, and L. M. Frank, "The time-rescaling theorem and its application to neural spike train data analysis," *Neural Computation*, no. 14, pp. 325–346, 2001.

[48] E. P. Gardner and A. F. Fuchs, "Single-unit responses to natural vestibular stimuli and eye movements in deep cerebellar nuclei of the alert rhesus monkey," *Journal of Neurophysiology*, vol. 38, pp. 627–649, 1975.

[49] M. Joshua, J. F. Medina, and S. G. Lisberger, "Diversity of neural responses in the brainstem during smooth pursuit eye movements constrains the circuit mechanisms of neural integration," *Journal of Neuroscience*, vol. 33, pp. 6633–6647, April 2013.

[50] D. L. Sparks, "The brainstem control of saccadic eye movements," *Nature Reviews: Neuroscience*, vol. 3, pp. 952–964, December 2002.

[51] C. J. D. Luca and J. C. Kline, "Influence of proprioceptive feedback on the firing rate and recruitment of motoneurons," *Journal of Neural Engineering*, vol. 9, p. 17, 2012. DOI:10.1088/1741-2560/9/1/016007.

[52] P. Fries, J. H. Reynolds, A. E. Rorie, and R. Desimone, "Modulation of oscilatory neuronal synchronization by selective visual attention," *Science*, vol. 291, pp. 1560–1563, February 2001.

[53] R. D. Flint, E. W. Lindberg, L. R. Jordan, L. E. Miller, and M. W. Slutzky, "Accurate decoding of reaching movements from field potentials in the absence of spikes," *Journal of Neural Engineering*, vol. 9, p. 13, 2012.

[54] H. G. Eyherabide, A. Rokem, A. V. M. Herz, and I. Samengo, "Burst firing is a neural code in insect auditory system," *Frontiers in computational neuroscience*, 2008.

[55] MathWorks, "Product documentation: polyfit." Online: `http://www.mathworks.com/help/techdoc/ref/polyfit.html` [Jan. 3, 2012].

[56] R. Pasupathy, "Generating nonhomogeneous poisson processes." Online: `http://filebox.vt.edu/users/pasupath/papers/nonhompoisson_streams.pdf` [Dec. 10, 2012].

[57] P. A. W. Lewis and G. S. Shedler, "Simulation of nonhomogenous poisson processes by thinning," *Naval Research Logistics Quarterly*, vol. 26, no. 3, pp. 403–413, 1979.

[58] R. F. Serfozo, "Conditional poisson processes," *Journal of Applied Probability*, vol. 9, pp. 288–302, June 1972.

[59] M. Beuhren, "Multicore - parallel processing on multiple cores." Online: `http://www.mathworks.com/matlabcentral/fileexchange/13775-multicore-parallel-processing-on-multiple-cores` [Nov. 24, 2013], Jan 2007.