frontiers

Bayesian estimation of orientation and direction tuning captures parameter uncertainty

Zongting Wu[1], **and Stephen D. Van Hooser**[2,*]

[1] *Brandeis University, Department of Biochemistry, Waltham, MA, USA*
[2] *Brandeis University, Department of Biology, Waltham, MA, USA*

Correspondence*:
Stephen D. Van Hooser
vanhoosr@brandeis.edu

ABSTRACT

This study explores the efficacy of Bayesian estimation in modeling the orientation and direction selectivity of neurons in the primary visual cortex (V1). Unlike traditional methods such as least squares, Bayesian estimation adeptly handles the probabilistic nature of neuronal responses, offering robust analysis even with limited data and weak selectivity. Through the analysis of both simulated and experimental data, we demonstrate that Bayesian estimation not only accurately fits the neuronal tuning curves but also effectively captures parameter certainty or uncertainty of both strongly and weakly selective neurons. Our results affirm the complex interdependencies among response parameters and highlight the variability in neuronal behavior under varied stimulus conditions. Our findings provide guidance as to how many response samples are necessary for Bayesian parameter estimation to achieve reliable fitting, making it particularly suitable for studies with constraints on data availability.

1 INTRODUCTION

Neurons in the visual system of all examined mammals, as well as neurons in other systems such as head-direction cells and place cells in the hippocampus, often exhibit tuning for the orientation or direction of a stimulus [Hubel and Wiesel, 1959, 1962, Wilson et al., 2018, Hubel and Wiesel, 1968, Henry et al., 1974, Ohki et al., 2005] or of the movement of an animal in a coordinate system[Taube et al., 1990, McNaughton et al., 1983]. When selectivity is strong, there are many methods for quantifying the orientation and direction angle preference, the degree of tuning relative to an orthogonal angle, or the sensitivity of the tuning as the stimulus changes from the preferred angle [Carandini and Ferster, 2000, Ringach et al., 2002, Swindale, 1998, Mazurek et al., 2014].

Today, many optophysiology and multi-channel recording studies with massive electrode channel counts often characterize the response properties of dozens or hundreds of neurons simultaneously [Stringer et al., 2019, de Vries et al., 2019, Siegle et al., 2021], some of which may have high selectivity and some of which may not. Some methods of quantifying the degree of orientation and direction tuning are robust regardless of whether tuning is strong or weak. However, reliably estimating tuning parameters such as angle preference and tuning width is not possible with least squares methods when tuning is weak[Mazurek et al., 2014]. In a prior methods paper, our group recommended simply not trying to quantify parameters

other than the degree or amount (or magnitude) of tuning if a statistical test (such as Hoteling's T2 test or an ANOVA across stimuli) did not show evidence of some significant tuning [Mazurek et al., 2014].

Nevertheless, there are times when one may want to quantify properties such as preference angle or tuning width even for cells that exhibit weak tuning. For example, one may want to make a statement about how tuning has changed when following a cell over the course of development, before and after an experimental manipulation, or as another parameter (say, stimulus spatial frequency or temporal frequency) is varied [Li et al., 2008, Moore et al., 2005]. If a cell is initially poorly tuned, such that we have no certainty about its angle preference, but later exhibits strong tuning for upward motion, we may want to contrast this situation with another cell that exhibits strong tuning for downward motion but later exhibits strong tuning for upward motion. If we throw up our hands when tuning is weak, we lose the ability to make quantitative statements about these situations.

Bayesian estimation methods allow quantification of a full joint distribution of the parameters of a tuning model, so that any uncertainty or certainty of a given parameter is estimated [Bishop, 2006]. Full Bayesian estimation requires much more computation than least squares fitting or maximum likelihood estimation[Cronin et al., 2010], but advances in computing power and GPU methods now allow these methods to be applied routinely and at speeds that are similar to least squares approaches. We describe a Bayesian estimation approach for orientation and direction tuning curves that can be widely applied to other sensory tuning curves. We provide enough detail that an analyst who is unfamiliar with Bayesian estimation can apply the method to a new problem. We provide comparisons to the bootstrap technique[Efron and Tibshirani, 1993, Press and Flannery, 1992], which also allows one to describe the certainty or uncertainty of tuning parameters.

2 MATERIALS AND METHODS

Neurons or other systems can exhibit responses that vary with stimulus direction and/or orientation. These responses can be spikes or voltage deflections, modulation of activity at a stimulus frequency, or responses obtained from an indicator such as a calcium indictor. There are two major classes of methods for analyzing orientation and direction tuning curves: vector methods [Batschelet, 1981, Swindale, 1998, Mazurek et al., 2014] and fit methods [Carandini and Ferster, 2000, Ringach et al., 2002, Swindale, 1998, Mazurek et al., 2014]. Vector methods, such as the calculation of circular variance Ringach et al. [2002] are excellent for quantifying the degree or strength of orientation or direction tuning for weakly and strongly tuned cells [Mazurek et al., 2014], but in this paper we will focus on an equation model of orientation and direction tuning:

$$R(\theta; C, R_p, R_n, \theta_{pref}, \sigma) = C + R_p * e^{-\frac{(\text{angdiff}(\theta-\theta_{pref}))^2}{2\sigma^2}} + R_n * e^{-\frac{(\text{angdiff}(\theta-\theta_{null}))^2}{2\sigma^2}}. \quad (1)$$

Five parameters are included in the equation, which are C, R_p, R_n, θ_{pref} and σ. The offset, C, is the portion of the response that is constant and does not vary with orientation or direction. R_p is the response above offset for the preferred direction and R_n is the response above offset for the null direction, that is, the direction that is opposite of the preferred direction. θ_{pref} is the preferred direction. In direction space, the range of θ is $[0, 360^\circ)$, and $\theta_{null} = \theta_{pref} + 180^\circ$. The function $\text{angdiff}(\theta) = min(\theta, \theta - 360, \theta + 360)$ computes the absolute angular difference around the circle. We also sometimes refer to the response in the preferred direction and the null direction, respectively, as

$$R_{pref} = R(\theta_{pref}) \tag{2}$$

and

$$R_{null} = R(\theta_{null}). \tag{3}$$

Various different groups use different definitions of θ, with some using Cartesian coordinates where 0° refers to vertical bars moving to the right, and increasing θ implies a counter-clockwise change in angle, and other groups such as own using compass coordinates, where 0° refers to horizontal bars moving upwards and increasing θ implies a clockwise change in angle.

2.1 Bayesian formulation

Bayes' theorem is stated mathematically as the following equation:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \tag{4}$$

We constructed a Bayesian model of Equation (1). We employed a change-of-variable by letting $R_n = \alpha R_p$ with $0 \leq \alpha \leq 1$, so that we can always be sure that $R_n \leq R_p$ as we compute the likelihood of the model. Our new equation is as follows:

$$P(C, R_p, \alpha, \theta_{pref}, \sigma|D) = \frac{P(D|C, R_p, \alpha, \theta_{pref}, \sigma)P(C, R_p, \alpha, \theta_{pref}, \sigma)}{P(D)}. \tag{5}$$

In words, the equation says that the probability of the parameters $(C, R_p, \alpha, \theta_{pref}, \sigma)$ given the data D we observed is equal to the probability of seeing the data given the parameters, multiplied by the prior probability of the parameters and divided by the prior probability of the data (Bishop [2006]).

The left side of Equation 5 describes what we are seeking: how likely is it that each parameter (say, C), takes any of a set of values (say, $\{C_1, C_2, C_3, \ldots\}$). In order to calculate this probability, we need to be able to calculate the probability of the data given the parameters (called *the likelihood function*) and the prior probability of the parameters.

In our case, the data D is the set of empirical mean responses across trials to the particular angles we examined. We indicate each of i angles tested by θ_i and the mean response to that angle as $r_{\theta i}$. We need to calculate the probability, according to our model, that we observed $r_{\theta i}$ given the parameters. Equation 1 tells us that the expected response $\hat{r}_{\theta i}$ for parameters $(C, R_p, \alpha, \theta_{pref}, \sigma)$ is

$$\hat{r}_{\theta i} = R(\theta_i; C, R_p, \alpha, \theta_{pref}, \sigma). \tag{6}$$

Next, we need to take this expected response and calculate the likelihood of the actual response $r_{\theta i}$ in units of probability. Many studies of the variability of neural firing rates have found that the variance (and standard deviation) of spike responses is linearly related to the mean response rate on a log-log plot (Softky and Koch [1993], Shadlen and Newsome [1998]). We therefore assumed that the model responses would be drawn from a normal distribution with mean $\hat{r}_{\theta i}$ and a standard deviation σ_{noise} that was a function of the mean response. To estimate $\sigma_{noise}(m)$ for our data, we examined the log of the standard deviation of

each response of each cell as a function of the log of the mean response, and modeled the relationship between the set of observations $\sigma_{noise,i}$ and the corresponding mean responses m_i with a linear fit as

$$log(\sigma_{noise,i}) = a + blog(m_i),$$

obtaining the best-fit values of a and b so that we could express the expected variability as a function of the mean response as

$$\sigma_{noise}(m) = 10^a * m^b. \tag{7}$$

For each set of recordings, we pooled our observations of all $\sigma_{noise,i}$ and m_i over all responses and all cells and derived a single a and b pair that was applied to all cells. For example, when dealing with calcium imaging data from ferret visual cortex (Li et al. [2008]), we used all cells and all responses to calculate the a and b we used for the noise model for those cells. We did not derive an a and b pair for each cell individually (as the limited data for a single cell would produce a poor fit), or apply a and b across situations (for example, between recordings of spiking neurons and recordings using calcium indicators).

In our experiments, we usually record responses over several trials T rather than recording a single trial, and report the mean response averaged over trials $r_{\theta i}$. Because we average over trials, we expect the response we measure to be less noisy than the results of a single trial. According to the central limit theorem, the variability decreases by a factor of $1/\sqrt{(T)}$. Therefore, the probability that a response at a single stimulus angle was generated by a particular set of model parameters is

$$P(r_{\theta i}|C, R_p, \alpha, \theta_{pref}, \sigma) = N(r_{\theta i}; \hat{r}_{\theta i}, \sigma_{noise}(\hat{r}_{\theta i})/\sqrt{(T)}) \tag{8}$$

where $N(x; \mu, \sigma_{normal})$ is the probability density function of the normal distribution with mean μ and standard deviation σ_{normal}. We assume that the response measurements are independent, so that the probability of seeing the set of responses $\{r_{\theta 1}, r_{\theta 2}, \ldots, r_{\theta n}\}$ for each cell is

$$P(D|C, R_p, \alpha, \theta_{pref}, \sigma) = \prod_{i=1}^{i=n} N(r_{\theta i}; \hat{r}_{\theta i}, \sigma_{noise}(\hat{r}_{\theta i})/\sqrt{(T)}). \tag{9}$$

In this report, we did not make any assumptions about the *a priori* likelihood of the different values of the parameters $(C, R_p, \alpha, \theta_{pref}, \sigma)$. We estimated our distributions for discrete sets of values of each parameter (for example, C could take $\{C_1, C_2, C_3, \ldots, C_{CM}\}$, R_p could take $\{R_{p1}, R_{p2}, R_{p3}, \ldots, R_{pRPM}\}$, etc), and we assumed each value was equally likely so that

$$P(C, R_p, \alpha, \theta_{pref}, \sigma) = \frac{1}{(CM)(RPM)(\alpha_M)(\theta_M)(\sigma_M)}, \tag{10}$$

where CM, RPM, α_M, θ_M, and σ_M are the number of values of each parameter used to estimate the distribution. Perhaps in the future, by examining many populations, the expected prior distributions could be estimated differently from a uniform distribution, but we did not want to impose any biases at this time.

Finally, we did not explicitly compute $P(D)$, but we assumed that our calculation of $P(C, R_p, \alpha, \theta_{pref}, \sigma|D)$ would estimate the real probability distribution, and reported the estimated distribution normalized to 1. We end up with a matrix of likelihoods $L(c, r, \alpha_j, t, s)$ with one entry for each combination of parameters examined.

Parameter	Spiking neurons	Calcium imaging
$min(C)$	0.1	$-MX$
$max(C)$	10	$+MX$
CM	60	60
$min(R_{pref})$	0.1	0.001
$max(R_{pref})$	20	+3 $* MX$
$R_{pref}M$	60	60
$min(\alpha)$	0	0
$max(\alpha)$	1	1
αM	15	21
$min(\theta_{pref})$	0	0
$max(\theta_{pref})$	359	359
$\theta_{pref}M$	72	72
$min(\sigma)$	1	1
$max(\sigma)$	60	60
σM	60	60
	Number of calculations	Number of calculations
	$2.3 * 10^8$	$3.3 * 10^8$

Table 1. Parameter space sampled for Bayesian parameter estimation, for spiking neurons (left column) and calcium imaging (right column). Parameters were varied from a minimum value to a maximum value using the number of steps indicated for each parameter with an M. For example, CM is the number of parameters evaluated for parameter C. For calcium imaging, MX indicates the empirical maximum mean response for each cell. The total number of calculations necessary for each cell is shown at the bottom.

2.2 Practical computation and storage of Bayesian results

We calculated our Bayesian distributions using ranges of parameters that varied slightly depending upon the units of the data. In our case, we used mean spike rates and fractional change in calcium indicators from baseline. The ranges we used are shown in Table 1.

To perform Bayesian estimation for a single cell, one must perform calculations for all combinations of the 5 parameter ranges, and Table 1 indicates that billions of calculations are required for each cell. This creates burdens for computation and for storage of the results. Fortunately, modern GPUs are well-suited to performing these calculations quickly. We wrote GPU code using Matlab (MathWorks, Natick, MA) that performs these calculations in less than a minute per cell on an ordinary laptop with a GPU, so the computation is not problematic for today's computing environments.

Storing the results requires careful decisions, however. Each probability for every parameter combination is a floating point number; if stored on disk with double precision, then the results for a single cell would require on the order of 10 GB of storage. For this reason, we did not store all of the results but instead stored a much briefer summary of key statistics.

First, we stored the marginal likelihood of each parameter individually by summing over all other parameters and normalizing the distribution to 1. For example, the marginal likelihood $\hat{Lc}(c)$ of parameter C is computed as

$$Lc(c) = \sum_{r=1}^{r=RPM} \sum_{\alpha_j=1}^{\alpha_j=\alpha_M} \sum_{t=1}^{t=\theta_M} \sum_{s=1}^{s=\sigma_M} L(c, r, \alpha_j, t, s)$$

$$\hat{Lc}(c) = \frac{Lc(c)}{\sum_{c=1}^{c=CM} Lc(c)}. \tag{11}$$

While we lose information about the joint distributions of these parameters, the summary provides considerable savings, needing only $(CM) + (RPM) + (\alpha_M) + (\theta_M) + (\sigma_M)$ entries for storage instead of $(CM)(RPM)(\alpha_M)(\theta_M)(\sigma_M)$ entries (the sum instead of the product), which with our parameters means a difference between kilobytes and gigabytes per cell. The marginal distributions say a lot about the response properties of the cell, and, if we really need the joint distribution, we can recompute it.

Second, we performed estimates of the likelihoods of index values of interest while we had the full matrix of parameter estimations in memory, and stored histograms of these values. OI (Orientation Index) and DI (Direction Index) are two indexes of neuronal direction/orientation selectivity. Their magnitudes are used to characterize the sensitivity of neurons to stimuli motion in different directions/orientations in space. In this paper, we define these values as follows:

$$OI = (R_{pref} + R_{null} - (R_{orth+} + R_{orth-}))/(R_{pref} + R_{null}) \tag{12}$$

and

$$DI = (R_{pref} - R_{null})/R_{pref}. \tag{13}$$

For summary statistics, we computed histograms of the likelihood that OI and DI took values between 0 and 1 in bins of size 0.05. By using the full likelihood matrix to calculate these histograms but only storing the histogram results, we obtained a good picture of our certainty or uncertainty about the value of each index that uses the full joint information, while storing only 20 values (1/0.05) per index.

Third, we also stored the most likely value of the full joint parameters also known as the maximum likelihood estimate (MLE). These parameters are a single set of values of $C, R_p, \alpha, \theta_{pref}, \sigma$ that provided the largest joint likelihood, but by themselves do not account for our uncertainty or certainty of their values.

3 RESULTS

To illustrate the approach and to explore how the Bayesian estimation tracks a tuning curve's parameters under a ground truth situation, we generated simulated data for analysis. In 1, we show two example tuning curves: one is well-tuned for orientation and direction (1A) while the other is poorly tuned (1B). The output of Bayesian estimation is not a single value of each parameter but instead is a probability distribution over the parameters. The marginal likelihood for θ_{pref} and R_{pref} is plotted for each cell, along with histograms of the estimated orientation index (OI) and direction index (DI). When the tuning is strong, there is great certainty about the parameters and the values of the index values. When the tuning is weak, there is much more uncertainty. In particular, in the weakly-tuned case, there is great uncertainty about whether the orientation and direction index values are near 0 or might be moderate. Least-squares methods would give us a single value of all of these parameters, and our uncertainty about the orientation and direction index values would not be described.

One may ask, why bother to study the tuning properties of weakly-tuned cells? To demonstrate such a case, we imagine measuring the direction tuning of a single cell while another visual stimulus parameter (temporal frequency) is changed. For low temporal frequency, the cell exhibits high direction selectivity; for a moderate temporal frequency, the cell exhibits orientation tuning but not direction tuning; for a further temporal frequency, the cell is weakly responsive.

To evaluate the predictive accuracy and sensitivity of the Bayesian estimation model, we simulated response curves for two different types of cells and performed parameter estimation. All error bars in the text are represented using standard errors of the mean (SEM).

3.1 Bayesian estimation for simulated strong and weak spatial selectivity

As shown in **Figure 1**, multiple graphs illustrate the response characteristics of simulated neurons. The left column, labeled as "Well-tuned", features graphs that exhibit neuronal responses with higher peak firing rates and marked preferences for specific directions and orientations of stimuli. These responses are quantified in terms of frequency (Hz) and marginal likelihood across different orientations and directions, denoted by directional and orientation indexes. The right column, labeled as "Poorly-tuned", displays neurons with significantly lower response amplitudes, less distinct tuning curves, indicating a reduced selectivity for stimulus features. As shown in **Figure 1 A, B**, the black curves represent the ideal response curve of simulated data without any noise. To approximate real data more closely, we used simulated sampling points with an added 50% Gaussian noise for fitting. Each data point is based on five samples, covering a total of 36 angles. These sampling points are represented as blue dots in the figure. The noisy data, alongside results from a linear noise model, serve as inputs for Bayesian estimation, which are then interpolated based on given parameter ranges to produce the fits displayed as red curves. The parameter settings for the well-tuned simulated cells are ($C = 1$, $R_{pref} = 10$, $R_{null} = 5$, $\theta_{pref} = 90°$, and $\sigma = 30°$), while those for the poorly-tuned cells are ($C = 1$, $R_{pref} = 1$, $R_{null} = 0$, $\theta_{pref} = 90°$, and $\sigma = 30°$).

The Bayesian MLE fit for both types of simulated cells closely align with the ideal curves, particularly for the well-tuned cell in terms of the preferred angle and tuning width values. **Figure 1 C, D, E** and **F** describe the marginal probability distributions for θ_{pref} and R_p. The probability peaks accurately estimate the 'true values' of θ_{pref} and R_{pref} for the well-tuned cell. The certainty or uncertainty of the parameters is indicated by the probability distribution peaks and widths. Due to the poorly-tuned cell exhibiting weak selectivity for direction, the estimated probability for θ_{pref} is dispersed across a broader range of angles compared to the well-tuned cell. The likely values of R_{pref} are considerably smaller for the cell with poorer tuning, showing its reduced response.

Figure 1 G, H, I, J illustrates the probability distributions for the selectivity indexes DI and OI. It can be observed that the well-tuned cell displays distinct, singular probability peaks for DI and OI, accurately estimating the 'real' values ($OI = 0.86$, $DI = 0.45$). In contrast, for the poorly-tuned cell, the likelihood estimates of both are broader, although a close estimation to the 'real' value ($OI = 0.33$, $DI = 0.5$) is also observed in the probability distribution ($OI = 0.3$, $DI = 0.38$).

3.2 Bayesian estimation for simulated temporal frequency

Many studies of orientation or direction selectivity in the visual cortex employ drifting sinusoidal gratings (with a particular spatial frequency and temporal frequency) as stimuli (e.g., Movshon et al. [1978]). Some tuning properties, such as orientation tuning, are not very sensitive to spatial or temporal frequency, but others, such as direction selectivity, can change with stimulus parametersMoore et al. [2005]. For example, Moore et al. Moore et al. [2005] found that many cells in ferret visual cortex altered their direction preference angle or the amount of direction selectivity as temporal frequency was changed. Direction selectivity was often higher at low temporal frequencies than at higher temporal frequencies. If one wanted to study, say, how direction preference changed with temporal frequency, one would be out of luck with least squares approaches if selectivity became low. Here we will illustrate that the Bayesian approach allows one to study these properties quantitatively.

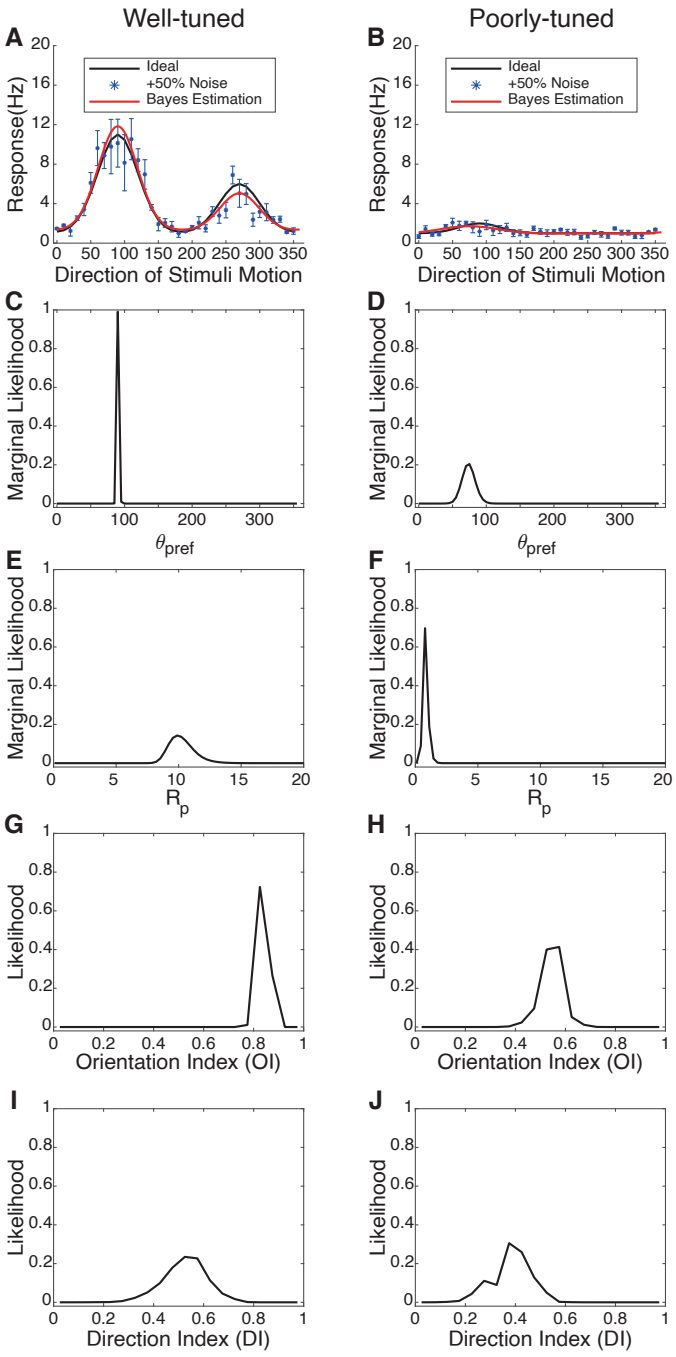

Figure 1. Bayesian estimation for two types of simulated cells with different levels of tuning. (A & B) Simulated response curve, data collection points with $+50\%$ Gaussian noise, and Bayesian estimation results. **(C & D & E & F)** The marginal likelihoods of θ_{pref} and R_{pref} in the model.**(G & H & I & J)** The probability distribution of OI and DI for spatial selectivity under Bayesian estimation.

In **Figure 2**, we simulated changes in direction selectivity of a single neuron under three different temporal frequencies: low, medium, and high. In our simulations, we hypothesized that the direction selectivity is strongest at low temporal frequency ($C = 1$, $R_{pref} = 7$, $R_{null} = 1$, $\theta_{pref} = 45°$, and $\sigma = 30°$). The maximum response amplitude is slightly smaller at the medium temporal frequency ($C = 1$, $R_{pref} = 5$, $R_{null} = 5$, $\theta_{pref} = 45°$, and $\sigma = 30°$), and the direction selectivity is very low. At the highest

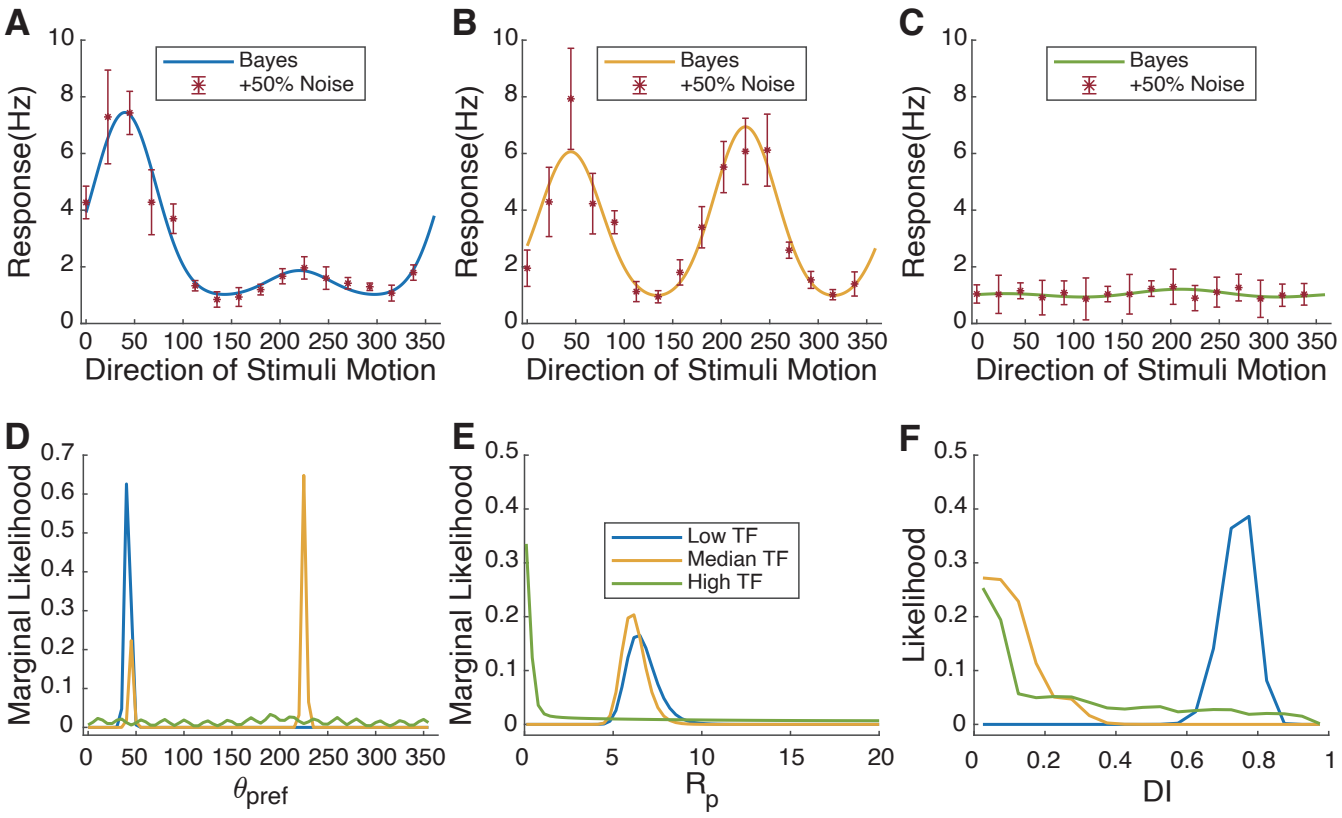

Figure 2. Neuronal responses and Bayesian estimation to stimuli at different temporal frequencies in simulated cell(A & B & C) Simulated 'real' response curves and sampling points with 50% added Gaussian noise in different temporal frequencies.**(D & E & F)** The probability distribution of θ_{pref}, R_{pref}, and DI fitted for different temporal frequencies. The changes in the likelihood of θ_{pref} and DI track the increased uncertainty of θ_{pref} with increasing temporal frequency and the decreasing DI with temporal frequency.

temporal frequency ($C = 1$, $R_{pref} = 0$, $R_{null} = 0$, $\theta_{pref} = 45°$, and $\sigma = 30°$), responses are very weak and there is an absence of direction selectivity. As shown in **Figure 2 A, B, C**, red represents the simulated sampled data points, the blue curve represents the simulated ideal curve for low temporal frequency, and yellow for medium temporal frequency, and green for high temporal frequency. To more closely replicate the actual fitting process, the number of sampling angles was reduced from 32 in **Figure 1** to 16. **Figure 2 D, E, F** shows the changes in the probability distributions of θ_{pref}, R_{pref}, and DI across the three different temporal frequencies, with the colors corresponding to the three temporal frequencies mentioned above. In **Figure 2 D**, Bayesian estimation accurately identifies the preferred angle at low temporal frequency. At the medium temporal frequency, since the response rates at R_{pref} and R_{null} are close, the model indicates two probability peaks at θ_{pref} and θ_{null}. For the high temporal frequency, the probability distribution lacks distinct peaks. This is further supported in **Figures 2 E, F**, where the probability peak of R_{pref} and DI at high temporal frequency predominantly cluster near or at zero, indicating minimal direction selectivity. For the medium temporal frequency, the obvious peak around 6 for R_{pref} and the peak near zero for DI suggest that the 'real' response curve has significant orientation selectivity with similar values of R_{pref} and R_{null}. The parameters and direction index for all three temporal frequencies successfully approximate the 'true values'.

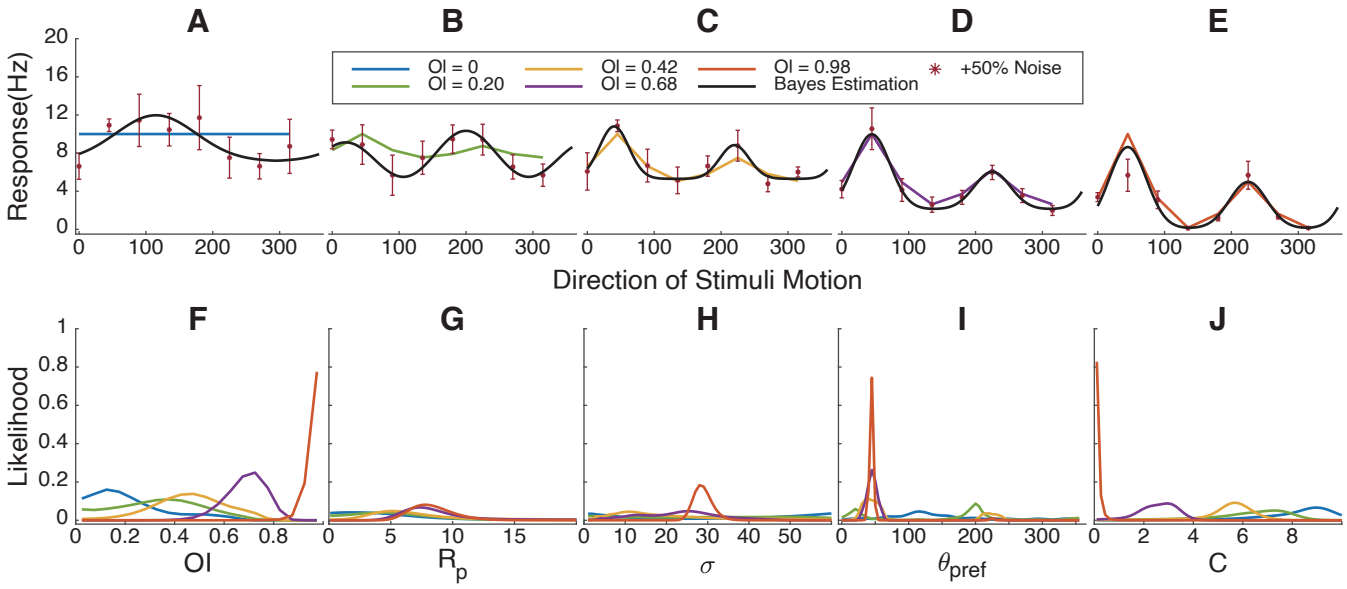

Figure 3. Bayesian estimation results for simulated cells with gradually increasing orientation index values. (A & B & C & D & E) Simulated response curve, sampling points with $+50\%$ Gaussian noise and fitting results. **(F & G & H & I & J)** The marginal likelihoods distribution of OI, R_{pref}, σ, θ_{pref} and C.

3.3 Bayesian estimation fitting for simulated data with variable orientation index and direction index values

To better test the universality and accuracy of Bayesian estimation, we simulated two groups of neurons, each consisting of five cells with gradually increasing OI or direction index DI values, ranging from 0 to 1. Additionally, we tested the sensitivity of the approach to different numbers of stimulus angles, as running additional stimulus angles requires more precious experimental time.

In **Figure 3** we simulated cells with increasing orientation selectivity index values; in **Figure 4**, we stimulated cells with increasing direction selectivity index values. In **Figure 3F**, the likelihood of the underlying OI value tracks the the true value, and θ_{pref} (**Fig. 3I**) becomes more certain as the underlying orientation index increases. Similarly, in **Figure 4F**, the likelihood of the underlying direction index tracks the true value, and again θ_{pref}

3.4 Impact of Sampling Angle Quantity on Bayesian Estimation Fitting Results

Experimental time is extremely valuable, and it is important to know how many angle steps might be necessary to estimate the parameters to a desired precision. Each color in the **Figure 5** corresponds to a different number of sampling angles, with the legend indicating the number of angles used for each Bayesian estimation fitting. The parameters of the simulated tuning curve are set as ($C = 2.5$, $R_{pref} = 7.5$, $R_{null} = 3.75$, $\theta_{pref} = 45°$, and $\sigma = 30°$).

From the probability distribution of OI in **Figure 5 F**, it can be observed that four sampling angles produce a poor parameter estimation. There is also a significant likelihood of obtaining a fit with low orientation selectivity due to the accidental collection of angles where the response frequency in the 'real' tuning curve is lower. Starting from eight sampling angles, a probability peak can be seen around 0.6, although it is somewhat broad. Similarly, in **Figures 5 G, J**, higher sampling angles concentrate the probability peaks at specific values, and the fitting curves align well with the 'real' tuning curve. This indicates that

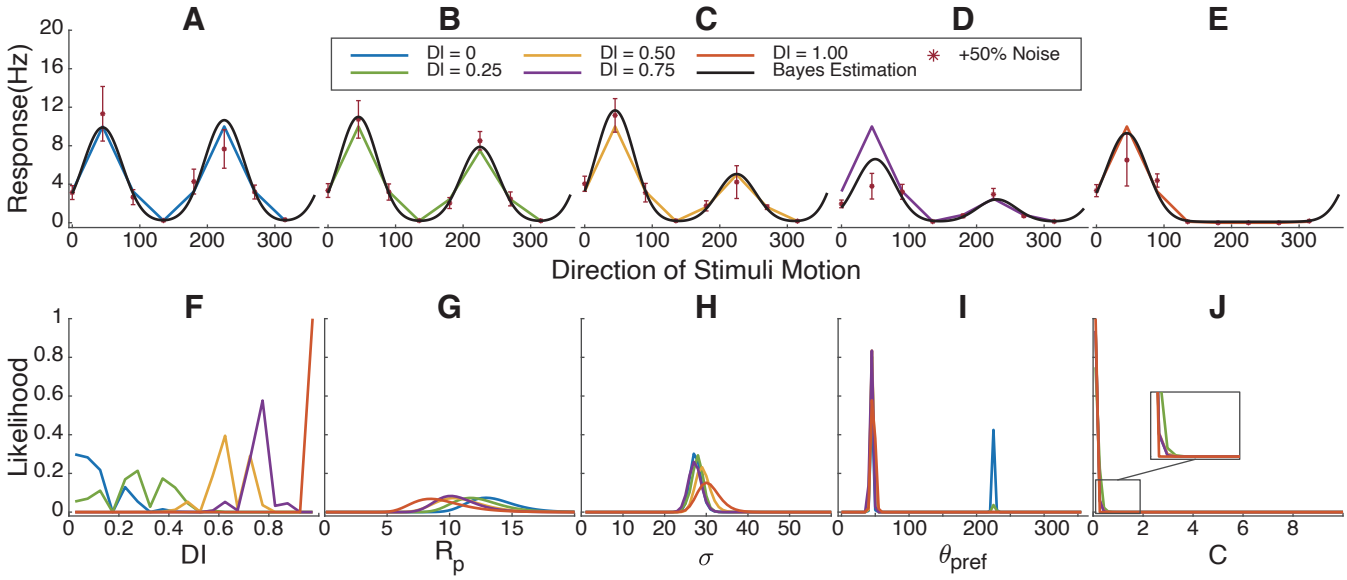

Figure 4. Bayesian estimation results for simulated cells with gradually increasing direction index values. (A & B & C & D & E) Simulated response curve, sampling points with $+50\%$ Gaussian noise and fitting results. **(F & G & H & I & J)** The marginal likelihoods distribution of DI, R_{pref}, σ, θ_{pref} and C.

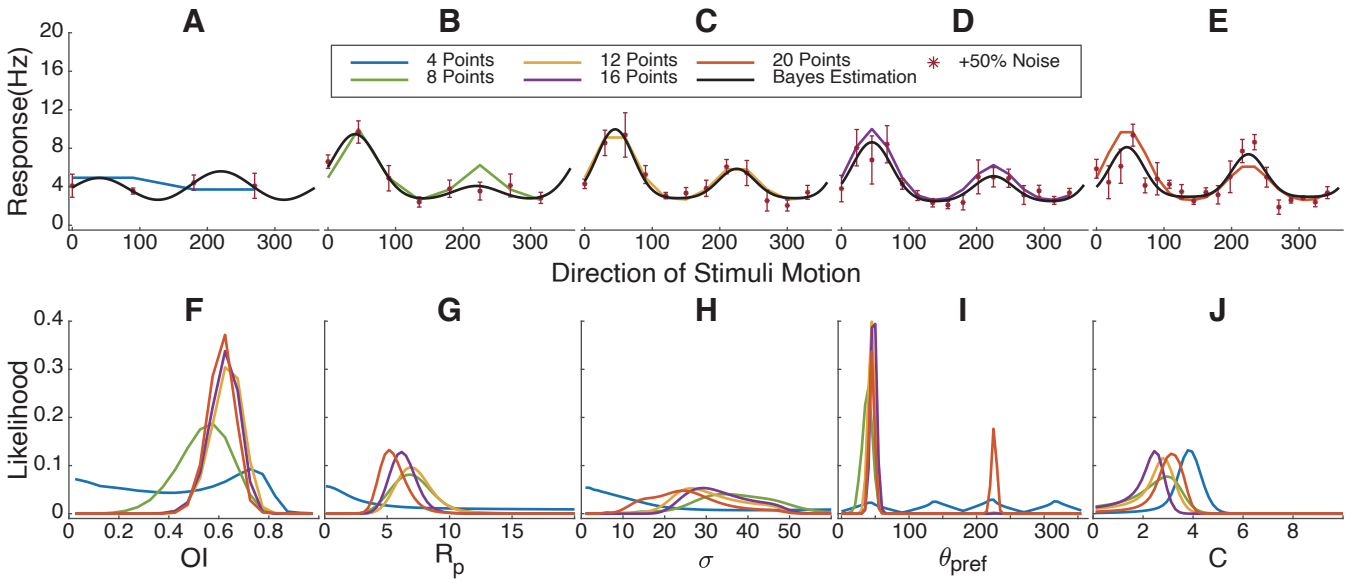

Figure 5. Impact of varying sampling angle quantities on Bayesian estimation results. (A & B & C & D & E) Simulated response curve, sampling points with $+50\%$ Gaussian noise and fitting results.**(F & G & H & I & J)** The marginal likelihoods distribution of OI, R_{pref}, σ, θ_{pref} and C.

eight or more sampling angles meet a minimum requirement for Bayesian estimation. However, under this parameter set, the probability distribution for σ remains broad, with an indistinct peak in the range from 20 to 30.

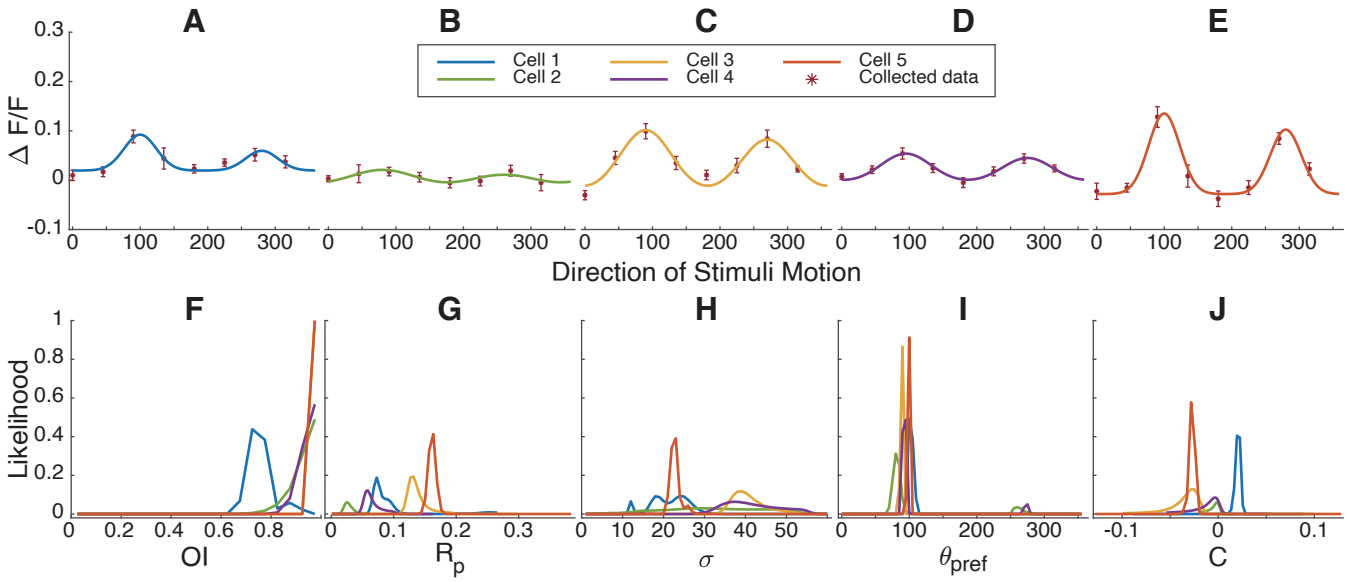

Figure 6. Impact of varying sampling angle quantities on Bayesian estimation results. (A & B & C & D & E) The response of juvenile ferret visual cortex to stimuli moving in different directions and Bayesian parameter estimation results. **(F & G & H & I & J)** The marginal likelihoods distribution of OI, R_{pref}, σ, θ_{pref} and C.

3.5 Bayesian Estimation Fitting Results for Neuronal Response Curves in the Ferret Primary Visual Cortex

Up to now, we have only examined simulated data for illustration. Next, we analyzed response tuning curves of primary visual cortex neurons from juvenile ferrets using calcium imaging from a prior study Li et al. [2008]. Five example neurons are shown in **Figure 6 A, B, C, D, E**. In **Figure 6 G**, the likelihood distribution of R_{pref} lies between 0.1 to 0.1, with the green and purple cells showing smaller R_{pref} values. Nevertheless, all cells exhibit orientation selectivity with high confidence (**Fig. 6F**), and have narrow ranges of preferred directions, with some cells showing a possibility of preferring the opposite direction (**Fig. 6I**).

3.6 Bayesian parameter estimation compared to the bootstrap

Quantifying uncertainty in model parameters when the model is not a member of a well-known class, such as a mean or a generalized linear model, has been challenging. One common method is the bootstrap (Efron and Tibshirani [1993], Press and Flannery [1992]). Under this procedure, one generates hundreds or thousands of simulated surrogate datasets for each cell by drawing from the original data with replacement. For example, to simulate a surrogate dataset for a cell that was examined with N stimulus repetitions of S stimuli, one would, for each stimulus s, draw N response values from the N actual response values with replacement, so that some trials are drawn more than once and others drawn not at all. By performing a model fit, such as a least-squares fit, of the hundreds or thousands of surrogate datasets, one can build up a probability distribution of parameter values that reflect the certainty or uncertainty of that parameter.

At first glance, using Bayesian parameter estimation or using the bootstrap for least-squares fits might seem equally appropriate, and that the chief difference might be whether one wanted to perform billions of forward calculations (where the calculations are not iterative and amendable to GPU methods) or thousands of least-squares fits (which are iterative and are less amendable to GPU methods). But we

will argue that Bayesian parameter estimation is preferred because of the practical problems with least-squares fitting, including local minima and a frequent need to place constraints on parameters to aid with convergenceMazurek et al. [2014].

We reanalyzed data from Li/Van Hooser et al. (Li et al. [2008]). In that work, the authors examined direction tuning in neurons of young ferrets before and after 3-6 hours of exposure to moving stimuli. In the naive state, cells generally exhibited relatively weak tuning for direction, whereas, after motion exposure, cells exhibited stronger tuning. One question raised in that study was whether or not individual cells were likely to have had an eventual direction preference that a) emerged from an unselective state, b) was solidified from an initial bias towards that direction preference, or c) was the result of a reversal of an initial preferred direction preference.

Figure 7 A, B, C, D, E, F displays the Bayesian likelihood distributions of θ_{pref} for three sample cells before motion exposure (BME) and after motion exposure (AME), illustrating changes in tuning properties. In **Figure 7 AB**, the preferred direction angle shifts 180 degrees, indicating a major and significant reorganization for that cell. **Figure 7 CD** shows an enhanced response strength at the original preferred angle, suggesting reinforcement of the cell's existing tuning properties. Finally, **Figure 7 EF** reveals an emergence of a single direction preference, where previously the cell exhibited similar response frequencies in two opposite directions. These findings highlight the effectiveness of Bayesian estimation in capturing nuanced changes in neuronal tuning, similar to bootstrap analysis. **Figure 7 D**.

The advantages of Bayesian parameter estimation become more apparent if we compare bootstrap and Bayesian parameter estimation across the population of 262 cells. Parameter R_p shows a monotonic relationship, indicating that both procedures do a similar job of identifying that parameter, and θ_{pref} is the same or, in a few cells that are presumably not very direction-selective, 180° apart. But, in order to converge to a meaningful fit, parameter σ was set at a floor of 22.5° because the angle step in this experiment was 45°. In many bootstrap least-squared fits (**Fig. 7J**), this floor value was the best-fit value. However, in the Bayesian approach, we find that the MLE parameter value spans a broader range. Further, the bootstrap approach often returned the constrained minimal value for R_n, which was 0, in the bootstrap approach, whereas the MLE for the Bayesian parameter estimation spanned a wider range(**Fig. 7K**).

Therefore, the Bayesian parameter estimation approach, which dutifully examines the likelihood of different combinations of parameters over many ranges, is immune from the local minima and constraint problems that can plague least-squares fitting. In Bayesian fitting, one has the burden of choosing an appropriate discretization of the parameter space and providing computational resources to carry out the computation, but with modern GPU methods these are not burdensome for relatively simple models.

4 DISCUSSION

We derived a Bayesian estimation approach for orientation and direction tuning for neurons or other processes. Least-squares fitting and vector approaches do a good job of identifying model parameters when there is some substantial selectivity, but do not readily produce a measure of certainty or uncertainty when tuning is poor Mazurek et al. [2014]. Bayesian parameter estimation explicitly calculates the likelihood of different parameter values. Until recently, the billions of calculations needed to estimate these parameter distributions were an impediment to the broad use of Bayesian methods, but the new availability of fast GPUs that can perform trillions of floating point calculations per second makes this feasible even for large populations of neurons.

4.1 Modifications of least-squares models

In order to apply Bayesian methods to tuning curves, it may be necessary to modify the original least-squares tuning model, as we did here.

4.2 Selecting what to save

The observed flattening of the probability distributions for certain parameters when there is a high response offset could be due to the interdependence among parameters. Parameters in a model, especially in a complex system like neural responses, often interact with each other. This interdependence means that changes in one parameter can affect the values and distributions of others. This lack of independence complicates the fitting process and can disturb the estimation results. When using Bayesian estimation, considering the relationships between parameters, such as their potential correlations or interactions, is crucial for accurate modeling. It may be necessary to adjust the model to account for these dependencies, possibly through a hierarchical or multi-level Bayesian approach, and to integrate it with prior probabilities [Lewicki, 1994, Cronin et al., 2010], which can better handle the complex interactions between parameters. Further analysis and modeling strategies that incorporate these interdependencies could lead to more robust and reliable estimation results.

The probability distributions of the fitting results currently depend heavily on the fitting range and quantity of each parameter. Inappropriate fitting ranges can cause an increase in the probability distributions of some parameters and selectivity indexes near zero, and also reduce the height of the probability peaks. This can misrepresent the underlying neural properties by suggesting a lack of selectivity where there might be significant neural behavior.

Comparing the bootstrap results with actual data [Li et al., 2008], the Bayesian estimation fitting results show slight shifting, even though the values are still close. This drift might be caused by the errors introduced by the Gaussian noise model used in the analysis. Further investigation is needed to explore the reasons behind this drift. Analyzing the impact of the noise model on the fitting results could help refine the estimation process and improve accuracy. It's essential to understand how different noise levels and types influence the Bayesian estimation to ensure that the model assumptions align with the actual data characteristics.

REFERENCES

E. Batschelet. *Circular Statistics in Biology*. Academic Press, New York, NY, 1981.

Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006. ISBN 0-387-31073-8.

Matteo Carandini and David Ferster. Membrane potential and firing rate in cat primary visual cortex. *Journal of Neuroscience*, 20(1):470–484, 2000. ISSN 0270-6474. doi: 10.1523/JNEUROSCI.20-01-00470.2000. URL https://www.jneurosci.org/content/20/1/470.

B. Cronin, I.H. Stevenson, M. Sur, and K.P. Körding. Hierarchical Bayesian modeling and Markov chain Monte Carlo sampling for tuning-curve analysis. *J Neurophysiol*, 103:591–602, 01 2010. ISSN 0022-3077. doi: 10.1152/jn.00379.2009.

Saskia E. J. de Vries, Jérôme A. Lecoq, Michael A. Buice, Paul A. Groblewski, Gabriel K. Ocker, Miles Oliver, David Feng, Nathan Cain, Peter Ledochowitsch, Daniel Millman, et al. A large-scale standardized physiological survey reveals functional organization of the mouse visual cortex. *Nature Neuroscience*, 22(12):1941–1950, 2019. doi: 10.1038/s41593-019-0550-9.

Bradley Efron and Robert J. Tibshirani. *An Introduction to the Bootstrap*. Number 57 in Monographs on Statistics and Applied Probability. Chapman & Hall/CRC, Boca Raton, Florida, USA, 1993.

G H Henry, B Dreher, and P O Bishop. Orientation specificity of cells in cat striate cortex. *Journal of Neurophysiology*, 37(6):1394–1409, 1974. doi: 10.1152/jn.1974.37.6.1394. URL https://doi.org/10.1152/jn.1974.37.6.1394. PMID: 4436709.

D. H. Hubel and T. N. Wiesel. Receptive fields of single neurones in the cat's striate cortex. *The Journal of Physiology*, 148(3):574–591, 1959. doi: https://doi.org/10.1113/jphysiol.1959.sp006308. URL https://physoc.onlinelibrary.wiley.com/doi/abs/10.1113/jphysiol.1959.sp006308.

D. H. Hubel and T. N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, 160(1):106–154, 1962. doi: https://doi.org/10.1113/jphysiol.1962.sp006837. URL https://physoc.onlinelibrary.wiley.com/doi/abs/10.1113/jphysiol.1962.sp006837.

D. H. Hubel and T. N. Wiesel. Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, 195(1):215–243, 1968. doi: https://doi.org/10.1113/jphysiol.1968.sp008455. URL https://physoc.onlinelibrary.wiley.com/doi/abs/10.1113/jphysiol.1968.sp008455.

Michael S. Lewicki. Bayesian Modeling and Classification of Neural Signals. *Neural Computation*, 6(5):1005–1030, 09 1994. ISSN 0899-7667. doi: 10.1162/neco.1994.6.5.1005. URL https://doi.org/10.1162/neco.1994.6.5.1005.

Ye Li, Stephen D. Van Hooser, Mark Mazurek, Leonard E. White, and David Fitzpatrick. Experience with moving visual stimuli drives the early development of cortical direction selectivity. *Nature*, 456:952–956, 2008. doi: https://doi.org/10.1038/nature07417. URL https://www.nature.com/articles/nature07417#citeas.

Mark Mazurek, Marisa Kager, and Stephen D. Van Hooser. Robust quantification of orientation selectivity and direction selectivity. *Frontiers in Neural Circuits*, 8, 2014. ISSN 1662-5110. doi: 10.3389/fncir.2014.00092. URL https://www.frontiersin.org/articles/10.3389/fncir.2014.00092.

B L McNaughton, C A Barnes, and J O'Keefe. The contributions of position, direction, and velocity to single unit activity in the hippocampus of freely-moving rats. *Exp. Brain Res.*, 52(1):41–49, 1983.

Bartlett D Moore, 4th, Henry J Alitto, and W Martin Usrey. Orientation tuning, but not direction selectivity, is invariant to temporal frequency in primary visual cortex. *J. Neurophysiol.*, 94(2):1336–1345, August 2005.

J A Movshon, I D Thompson, and D J Tolhurst. Spatial and temporal contrast sensitivity of neurones in areas 17 and 18 of the cat's visual cortex. *The Journal of Physiology*, 283(1):101–120, 1978. doi: https://doi.org/10.1113/jphysiol.1978.sp012490. URL https://physoc.onlinelibrary.wiley.com/doi/abs/10.1113/jphysiol.1978.sp012490.

Kenichi Ohki, Sooyoung Chung, Yeang H. Ch'ng, Prakash Kara, and R. Clay Reid. Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex. *Nature*, 433:597–603, 2005. doi: 10.1038/nature03274.

Teukolsky S. A. Vetterling W. T. Press, W. H. and B. P. Flannery. *Numerical Recipes in C*. Cambridge University Press, 1992.

Dario L Ringach, Robert M Shapley, and Michael J Hawken. Orientation selectivity in macaque v1: diversity and laminar dependence. *J. Neurosci.*, 22(13):5639–5651, July 2002.

Michael N. Shadlen and William T. Newsome. The variable discharge of cortical neurons: Implications for connectivity, computation, and information coding. *Journal of Neuroscience*, 18(10):3870–3896, 1998. ISSN 0270-6474. doi: 10.1523/JNEUROSCI.18-10-03870.1998. URL https://www.jneurosci.org/content/18/10/3870.

Joshua H. Siegle, Xiaoxuan Jia, S. Durand, S. Gale, C. Bennett, N. Graddis, G. Heller, T.K. Ramirez, H. Choi, J.A. Luviano, P.A. Groblewski, R. Ahmed, A. Arkhipov, A. Bernard, Y.N. Billeh, et al. Survey of spiking in the mouse visual system reveals functional hierarchy. *Nature*, 590(7847):224–229, 2021. doi: 10.1038/s41586-020-03171-x.

WR Softky and C Koch. The highly irregular firing of cortical cells is inconsistent with temporal integration of random epsps. *Journal of Neuroscience*, 13(1):334–350, 1993. ISSN 0270-6474. doi: 10.1523/JNEUROSCI.13-01-00334.1993. URL https://www.jneurosci.org/content/13/1/334.

Carsen Stringer, Marius Pachitariu, Nicholas Steinmetz, Matteo Carandini, and Kenneth D Harris. High-dimensional geometry of population responses in visual cortex. *Nature*, 571(7765):361–365, July 2019.

N V Swindale. Orientation tuning curves: empirical description and estimation of parameters. *Biol. Cybern.*, 78(1):45–56, January 1998.

JS Taube, RU Muller, and JB Ranck. Head-direction cells recorded from the postsubiculum in freely moving rats. i. description and quantitative analysis. *Journal of Neuroscience*, 10(2):420–435, 1990.

425 ISSN 0270-6474. doi: 10.1523/JNEUROSCI.10-02-00420.1990. URL https://www.jneurosci.
426 org/content/10/2/420.
427 Daniel E Wilson, Benjamin Scholl, and David Fitzpatrick. Differential tuning of excitation and inhibition
428 shapes direction selectivity in ferret visual cortex. *Nature*, 560(7716):97—101, August 2018. ISSN
429 0028-0836. doi: 10.1038/s41586-018-0354-1. URL https://europepmc.org/articles/
430 PMC6946183.

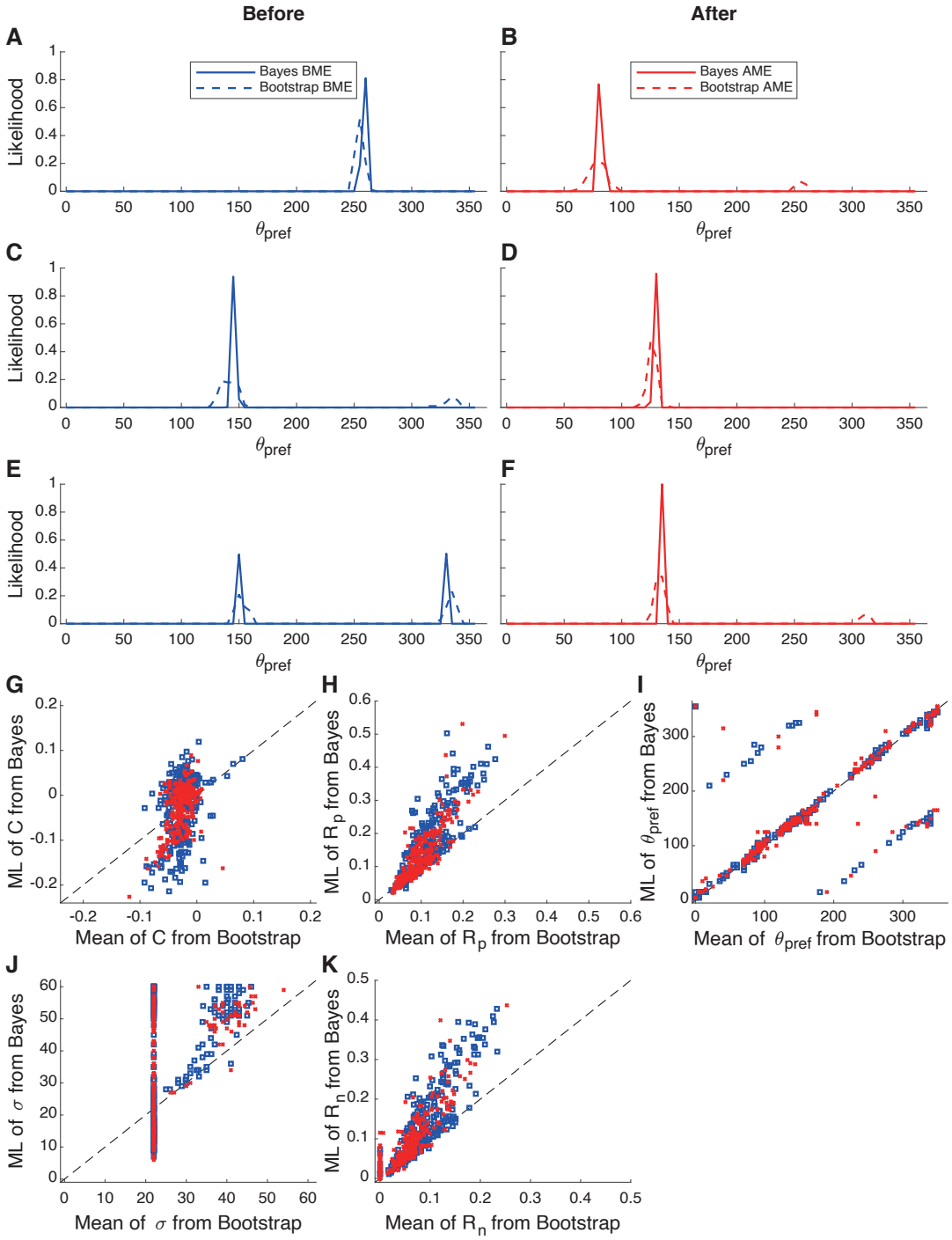

Figure 7. Bayesian Estimation and Bootstrap Fitting Results Before and After Motion Exposure. **(A & B & C & D & E & F)** Probability distribution fitting results for the parameter Thetapref in three representative cells before and after motion exposure. **(G & H & I & J & K)** Fitting results of individual parameters across 262 cells using different models. Each panel shows the relationship between the Bootstrap mean (x-axis) and the Bayesian maximum likelihood (ML) estimate (y-axis) for various model parameters, including OI, R_{pref}, σ, θ_{pref} and C.