A Novel Mixture Model for Characterizing Human Aiming Performance Data

Yanxi Li [1], Derek S. Young [1], Julien Gori [2], and Olivier Rioul [3]

[1] Dr. Bing Zhang Department of Statistics, University of Kentucky, Lexington, KY, USA

[2] ISIR, Sorbonne Université, Paris, France

[3] Télécom Paris, Institut Polytechnique de Paris, Paris, France

---

Address for correspondence: Derek S. Young, Dr. Bing Zhang Department of Statistics, University of Kentucky, Lexington, KY, USA.

E-mail: derek.young@uky.edu.

Phone: (+1) 859 218 3408.

Fax: (+1) 859 323 1973.

Abstract: Fitts' law is often employed as a predictive model for human movement, especially in the field of human-computer interaction. Models with an assumed Gaussian error structure are usually adequate when applied to data collected from controlled studies. However, observational data (often referred to as

data gathered "in the wild") typically display noticeable positive skewness relative to a mean trend as users do not routinely try to minimize their task completion time. As such, the exponentially-modified Gaussian (EMG) regression model has been applied to aimed movements data. However, it is also of interest to reasonably characterize those regions where a user likely was not trying to minimize their task completion time. In this paper, we propose a novel model with a two-component mixture structure – one Gaussian and one exponential – on the errors to identify such a region. An expectation-conditional-maximization (ECM) algorithm is developed for estimation of such a model and some properties of the algorithm are established. The efficacy of the proposed model, as well as its ability to inform model-based clustering, are addressed in this work through extensive simulations and an insightful analysis of a human aiming performance study.

Key words: block relaxation; ECM algorithm; exponentially-modified Gaussian; Fitts' law; human-computer interaction; model-based clustering

1 Introduction

An individual's reaction time and movement time are important markers about the status of their neurological system. Neurologists believe that reaction time is, perhaps, the most widely-used measure in neuroscience and psychology for noninvasively assessing processing in the brain (Wong et al., 2017). For instance,

patients with Parkinson's disease are found to have prolonged reaction time and movement time (Evarts et al., 1981), while slowed reaction time has also been regarded as an early feature of Alzheimer's disease (Gordon and Carson, 1990).

Fitts' law (Fitts, 1954) is an empirical law that describes movement time for human voluntary movement. In the Human-Computer Interaction (HCI) field, it has been adapted to model selection times in graphical user interfaces (GUIs) (MacKenzie and Buxton, 1992). Specifically, the minimum movement time t needed to select a rectangular target located at a distance d away, with width w and height h, is given by $t = a + b \log_2 \left(1 + \frac{d}{\min(h,w)}\right)$. This relationship has been demonstrated in numerous GUI contexts through controlled experiments, where participants have been asked to maximize their movement performance by going "as quickly and precisely" as possible. Usually, movement time data collected this way has relatively low variance, and the parameters of the linear model, a and b, are directly estimated using maximum likelihood estimation (or equivalently, ordinary least squares).

However, Fitts' model is often ill-fitting when the data arises from non-controlled settings, such as crowdsourced web-experiments (Goldberg et al., 2014) or field studies (Chapuis et al., 2007). It was recently argued that in non-controlled settings, Fitts' law should be interpreted as a model of minimum observed times (Gori et al., 2017, 2018). The idea is that, in these studies, one cannot control for perturbation or participant motivation, which may increase (but not decrease) the movement time needed to select the target. At the same time, one cannot naïvely

fit a lower bound to the dataset by identifying minimum movement times, since some movements may be poorly segmented. Another opportunity for lower than possible movement times is when participants accidentally click on a target (involuntary movement). It was previously shown that in this case, Fitts' law could be recovered using an exponentially-modified Gaussian (EMG) regression (Gori and Rioul, 2019).

The EMG distribution is defined as a convolution of the distributions of two independent random variables, where one follows a Gaussian distribution and the other follows an exponential distribution. A random variable X follows an EMG distribution if the density has the form

$$f(x;\mu,\sigma,\alpha) = \frac{\alpha}{2}\exp\left\{\frac{\alpha}{2}(2\mu+\alpha\sigma^2-2x)\right\}\operatorname{erfc}\left(\frac{\mu+\alpha\sigma^2-x}{\sqrt{2}\sigma}\right), \tag{1.1}$$

where $\mu \in \mathbb{R}$ and σ^2 are the variance and mean, respectively, of the Gaussian component, $\alpha > 0$ is the rate of the exponential component, and $\operatorname{erfc}(\cdot)$ is the complementary error function. We will write $X \sim EMG(\mu,\sigma,\alpha)$ to denote when a random variable follows the EMG distribution as defined above. Due to its characteristic positive skewness from the exponential component, the EMG distribution has provided insight into applied problems across a diverse cross-section of fields, such as microarray preprocessing (Silver, 2009), cell biology (Golubev, 2010), chromatography (Kalambet et al., 2011), and neuropsychology (Palmer and Horowitz, 2011). In the present study on human aiming performance, we seek a more critical examination of the data, which begins with analyzing EMG regression fits for individual subjects. We seek additional flexibility to understand from which

process the individual's performance arises: the one characterized by the Gaussian distribution or the one characterized by the exponential distribution. The EMG distribution does not allow for classifying such an individual observation, so we propose a competing regression model where the error structure is assumed to be a two-component mixture of a Gaussian and an exponential distribution. Thus, both the EMG regression model and our novel mixture model are able to characterize data with positive residuals relative to a mean trend, but the latter can also serve to perform model-based clustering.

We must also address some computational challenges of the two models in this work. For estimating the EMG regression model, we have found the existing computational routines to not be particularly robust, especially for large datasets like those analyzed in this work. We develop a block-relaxation algorithm for estimating an EMG regression model with (potentially) multiple predictors. We then develop an expectation-conditional-maximization (ECM; Meng and Rubin, 1993) algorithm for estimating our novel mixture-of-regressions model. A computational advantage of our novel mixture-of-regressions model is the global convergence of its corresponding ECM algorithm, which we lack in the block-relaxation algorithm for the EMG regression model.

The rest of this paper is organized as follows. In Section 2, we introduce the EMG regression model and our novel mixture-of-regressions model, which we refer to as a *mixture-of-regressions model with flare*, or *flare regression model*, in short. In Section 3, we detail the algorithms used for estimating both the EMG regres-

sion model and the flare regression model. We further establish some theoretical properties of the block-relaxation algorithm for estimating the EMG regression model, and prove the global convergence of the ECM algorithm for estimating the flare regression model. Estimation of standard errors for the estimated model parameters and details about a model-based clustering strategy using the flare regression model are also addressed. In Section 4, a large simulation study is conducted to assess the performance and robustness of the ECM algorithm for the flare regression model. In Section 5, we analyze human aiming performance data. We emphasize the results from the flare regression model, which are benchmarked against the EMG regression results. Other candidate models are considered in our analysis, but the metrics used demonstrate superior performance of the flare regression model in the presence of more extreme positive residuals. Finally, we conclude with a summary of the main results in Section 6.

2 The Models

For both of the models that we present, let $Y_1, \ldots, Y_n$ denote a random sample of size n, where each of these univariate random variables is measured with a vector of p-dimensional predictors, $p \in \mathbb{N}^+$, given by $\mathbf{X}_1, \ldots, \mathbf{X}_n$. We use the convention that $X_{i,1} \equiv 1$, $i = 1, \ldots, n$, to reflect an intercept in our models. We further let $(y_i, \mathbf{x}_i)$ denote the realizations of the pairs $(Y_i, \mathbf{X}_i)$. Thus, our focus will be on linear regression models of the form

$$y_i = \mathbf{x}_i \boldsymbol{\beta} + \epsilon_i, \tag{2.1}$$

but where non-traditional (i.e., non-Gaussian) distributional structures of ϵ_i will be explored.

First we consider the EMG regression setting, where the error structure for the model in (2.1) is $\epsilon_i \sim EMG(0, \sigma, \alpha)$, $i = 1, \ldots, n$. We next consider the model where $\epsilon_i \sim \lambda\mathcal{N}(0, \sigma^2) + (1 - \lambda)Exp(\alpha)$. For this two-component mixture structure on the error terms, $\lambda \in [0, 1]$ is the mixing proportion, $\mathcal{N}(0, \sigma^2)$ is the Gaussian distribution with mean 0 and variance σ^2, and $Exp(\alpha)$ is the exponential distribution with mean α^{-1}, where $\alpha > 0$ is the rate parameter. This structure gives us the model we refer to as a mixture-of-regressions model with flare, or simply a flare regression model. The etiology of the term "flare" for our purposes comes from the phenomenon that occurs in gamma-ray bursts, where flaring is an erratic emission of a huge amount of energy on a relatively short timescale (Bernardini et al., 2011). From a data perspective, this behavior manifests as an overall (piecewise) linear trend between the response and predictor(s), but a subset of the data clearly deviates more substantially from the linear trend than the rest of the data. One may also envision this as a form of one-sided contamination. Scatterplots of simulated data from an EMG regression model and a flare regression model are given in Figure 1 and Figure 2, respectively, as well as in the Supplemental Material (S.7).

3 Algorithms, Estimation, and Some Properties

3.1 EMG Regression and a Block-Relaxation Algorithm

Following (1.1), the density function for the EMG regression model is

$$f(y_i; \mathbf{x}_i, \boldsymbol{\psi}) = \frac{\alpha}{2} \exp\left\{\frac{\alpha}{2}[\alpha\sigma^2 - 2(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})]\right\} \operatorname{erfc}\left(\frac{\alpha\sigma^2 - (y_i - \mathbf{x}_i^\top \boldsymbol{\beta})}{\sqrt{2}\sigma}\right),$$

which yields the corresponding data loglikelihood

$$\ell(\boldsymbol{\psi}) = n\left(\log\frac{\alpha}{2} + \frac{\alpha^2\sigma^2}{2}\right) - \sum_{i=1}^{n}\left\{\alpha(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}) - \log\left[\operatorname{erfc}\left(\frac{\alpha\sigma^2 - (y_i - \mathbf{x}_i^\top \boldsymbol{\beta})}{\sqrt{2}\sigma}\right)\right]\right\}.$$

Here, $\boldsymbol{\psi} = (\boldsymbol{\beta}^\top, \sigma^2, \alpha)^\top$ is the parameter vector of interest. To estimate $\boldsymbol{\psi}$, we partition it into two blocks, $(\boldsymbol{\psi}_1^\top, \boldsymbol{\psi}_2^\top)^\top$, where $\boldsymbol{\psi}_1 = \boldsymbol{\beta}$ and $\boldsymbol{\psi}_2 = (\sigma^2, \alpha)^\top$. Maximum likelihood estimation is performed by setting the objective function $\mathbf{Q}(\boldsymbol{\psi}) = \ell(\boldsymbol{\psi})$. By using the partitioning we defined for $\boldsymbol{\psi}$, we can then apply the iterative block-relaxation algorithm of de Leeuw (1994) to estimate $\boldsymbol{\psi}$. See **Algorithm 1** in the Supplemental Material (S.1) for additional details.

The following theorem about concavity properties of EMG models allows us to comment about the concavity of $\mathbf{Q}(\boldsymbol{\psi})$.

Theorem 1 *Let $Y|\mathbf{X} \sim EMG(\mathbf{X}^\top\boldsymbol{\beta}, \sigma, \alpha)$ be (conditionally) an EMG random variable.*

1. *The logarithm of $f(y; \mathbf{x}, \boldsymbol{\psi})$ is strictly concave in y.*

2. *$\ell(\boldsymbol{\psi})$ is strictly concave in $\boldsymbol{\beta}$.*

3. $\ell(\boldsymbol{\psi})$ *is strictly concave in* α *if* $\alpha\sigma < 1$.

See the Supplemental Material (S.2–S.4) for a detailed proof of the above.

From Theorem 1, we conclude the objective function $\mathbf{Q}(\boldsymbol{\psi}) = \ell(\boldsymbol{\psi})$ is always strictly concave in $\boldsymbol{\beta}$ and is strictly concave in α if $\alpha\sigma < 1$. It is still challenging to derive sufficient conditions that ensure the strict concavity of $\mathbf{Q}(\boldsymbol{\psi})$ in σ. Hence, getting sufficient conditions that ensure the strict concavity of $\mathbf{Q}(\boldsymbol{\psi})$ in $\boldsymbol{\psi}$ remains an open problem. Due to the lack of concavity of $\mathbf{Q}(\boldsymbol{\psi})$ in $\boldsymbol{\psi}$, the global convergence of the block-relaxation algorithm cannot be guaranteed. However, this problem is circumvented in the flare mixture regression setting, which we show after developing the corresponding objective function for estimating its parameters.

3.2 The Flare Regression Model and an ECM Algorithm

For the flare regression model, the density function is

$$\begin{aligned} f(y_i; \mathbf{x}_i, \boldsymbol{\theta}) = \frac{\lambda}{\sqrt{2\pi\sigma^2}} \exp\Big\{ -\frac{1}{2\sigma^2}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2 \Big\} \\ + (1-\lambda)\alpha \exp\Big\{ -\alpha(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}) \Big\} I\Big\{ (y_i - \mathbf{x}_i^\top \boldsymbol{\beta}) > 0 \Big\}, \end{aligned} \tag{3.1}$$

which yields the corresponding (observed) data loglikelihood

$$\begin{aligned} \ell_o(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log \Bigg\{ \frac{\lambda}{\sqrt{2\pi\sigma^2}} \exp\Big\{ -\frac{1}{2\sigma^2}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2 \Big\} \\ + (1-\lambda)\alpha \exp\Big\{ -\alpha(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}) \Big\} I\Big\{ (y_i - \mathbf{x}_i^\top \boldsymbol{\beta}) > 0 \Big\} \Bigg\}. \end{aligned} \tag{3.2}$$

Here, $\boldsymbol{\theta} = (\lambda, \boldsymbol{\beta}^\top, \sigma^2, \alpha)^\top$ is the parameter vector of interest. Note, however, that finding $\hat{\boldsymbol{\theta}}$ by simply using (3.2) is challenging as in most finite mixture models, so

we consider the $(y_i, \mathbf{x}_i)$ as incomplete data resulting from non-observed complete data. The data is made complete by augmenting the problem with the unobserved indicators $Z_i = I\{\text{observation } i \text{ belongs to the Gaussian component}\}$. Thus, the complete-data loglikelihood is easily found to be

$$\ell_c(\boldsymbol{\theta}) = \sum_{i=1}^{n} \Bigg[Z_i \log \left(\frac{\lambda}{\sqrt{2\pi\sigma^2}} \exp \Big\{ -\frac{1}{2\sigma^2}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2 \Big\} \right) + (1 - Z_i) \log \Big((1-\lambda)\alpha \exp \Big(-\alpha(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}) I\Big\{ (y_i - \mathbf{x}_i^\top \boldsymbol{\beta}) > 0 \Big\} \Big) \Big) \Bigg]. \tag{3.3}$$

Maximum likelihood estimation for finite mixture models is typically performed via an expectation-maximization (EM) algorithm (Dempster et al., 1977). In many classic parametric mixtures, solutions of the maximization-step (M-step) exist in closed form; see McLachlan and Peel (2000). However, we cannot directly estimate $\boldsymbol{\theta}$ for the flare regression model using the above complete-data setup. In particular, we lack a closed-form solution of the regression coefficient vector $\boldsymbol{\beta}$ in the M-step. However, we mitigate this issue by implementing an iterative procedure within a conditional-maximization-step (CM-step) of an ECM algorithm.

In the first expectation-step (E-step) for iteration t, $t = 0, 1, \ldots$, we compute the expected complete-data loglikelihood as

$$\mathbf{Q}(\boldsymbol{\theta}; \boldsymbol{\theta}^{(t)}) = \sum_{i=1}^{n} \Bigg[Z_i^{(t)} \log \left(\frac{\lambda}{\sqrt{2\pi\sigma^2}} \exp \Big\{ -\frac{1}{2\sigma^2}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2 \Big\} \right) + (1 - Z_i^{(t)}) \log \Big((1-\lambda)\alpha \exp \Big(-\alpha(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}) I\Big\{ (y_i - \mathbf{x}_i^\top \boldsymbol{\beta}) > 0 \Big\} \Big) \Big) \Bigg],$$

where

$$Z_i^{(t)} = \frac{\frac{\lambda^{(t)}}{2\pi\sigma^{2(t)}} \exp\left\{ -\frac{1}{2\sigma^{2(t)}}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}^{(t)})^2 \right\}}{f(y_i; \mathbf{x}_i, \boldsymbol{\theta}^{(t)})} \tag{3.4}$$

is the posterior membership probability of observation i belonging to the Gaussian component of the flare regression model and the denominator is the flare regression density in (3.1). We then partition $\boldsymbol{\theta}$ into $(\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$, where $\boldsymbol{\theta}_1 = \boldsymbol{\beta}$ and $\boldsymbol{\theta}_2 = (\sigma^2, \alpha, \lambda)^\top$.

For the first CM-step, we calculate $\boldsymbol{\theta}_1^{(t+1)} = \arg\max_{\boldsymbol{\theta}_1} \mathbf{Q}(\boldsymbol{\theta}; \boldsymbol{\theta}^{(t)})$. In this step, we are updating $\boldsymbol{\beta}^{(t+1)}$ by maximizing the objective function $m(\boldsymbol{\beta}) = \mathbf{Q}(\boldsymbol{\beta}; \boldsymbol{\theta}^{(t)})$ with respect to $\boldsymbol{\beta}$, subject to the linear inequality constraints $(1 - Z_i^{(t)})(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}) \geq 0$, $i = 1, \ldots, n$. Not surprisingly, it is challenging to calculate the closed form for the maximum likelihood estimate (MLE) of $\boldsymbol{\beta}$. Instead, we iteratively update $\boldsymbol{\beta}$ using a gradient algorithm introduced by Lange (1995):

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} - \left[\frac{d^2 m}{d\boldsymbol{\beta}^2}\right]^{-1} \Bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}^{(t)}} \frac{dm}{d\boldsymbol{\beta}} \Bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}^{(t)}}, \tag{3.5}$$

where

$$\begin{aligned} \frac{dm}{d\boldsymbol{\beta}} &= \sum_{i=1}^{n} \left[\frac{Z_i^{(t)}}{\sigma^{2(t)}} \mathbf{x}_i (y_i - \mathbf{x}_i^\top \boldsymbol{\beta}) + \alpha^{(t)} (1 - Z_i^{(t)}) \mathbf{x}_i \right] \quad \text{and} \\ \frac{d^2 m}{d\boldsymbol{\beta}^2} &= -\sum_{i=1}^{n} \frac{Z_i^{(t)}}{\sigma^{2(t)}} \mathbf{x}_i \mathbf{x}_i^\top. \end{aligned} \tag{3.6}$$

Next, set $\boldsymbol{\theta}^{(t+1/2)} = (\boldsymbol{\theta}_1^{(t+1)\top}, \boldsymbol{\theta}_2^{(t)\top})^\top$ for the second E-step of the current iteration, and obtain the updated posterior membership probabilities as

$$Z_i^{(t+1/2)} = \frac{\frac{\lambda^{(t)}}{2\pi\sigma^{2(t)}} \exp\left\{ -\frac{1}{2\sigma^{2(t)}}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}^{(t+1)})^2 \right\}}{f(y_i; \mathbf{x}_i, \boldsymbol{\theta}^{(t+1/2)})}. \tag{3.7}$$

With $\boldsymbol{\theta}_1$ fixed at $\boldsymbol{\theta}_1^{(t+1)}$, we find $\boldsymbol{\theta}_2^{(t+1)} = \arg\max_{\boldsymbol{\theta}_2} \mathbf{Q}(\boldsymbol{\theta}; \boldsymbol{\theta}^{(t+1/2)})$ in the second CM-step, which yields the following MLEs that are weighted using the updated posterior membership probabilities $Z_i^{(t+1/2)}$, $i = 1, \ldots, n$:

$$\lambda^{(t+1)} = \frac{1}{n} \sum_{i=1}^{n} Z_i^{(t+1/2)} \tag{3.8}$$

$$\sigma^{2(t+1)} = \frac{\sum_{i=1}^{n} Z_i^{(t+1/2)} (y_i - \mathbf{x}_i^\top \boldsymbol{\beta}^{(t+1)})^2}{\sum_{i=1}^{n} Z_i^{(t+1/2)}} \quad \text{and} \tag{3.9}$$

$$\alpha^{(t+1)} = \frac{\sum_{i=1}^{n} (1 - Z_i^{(t+1/2)})}{\sum_{i=1}^{n} (1 - Z_i^{(t+1/2)})(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}^{(t+1)})}. \tag{3.10}$$

See **Algorithm 2** in the Supplemental Material (S.1) for additional details.

Letting $\boldsymbol{\theta}^{(\infty)}$ and $Z_i^{(\infty)}$, $i = 1, \ldots, n$, denote, respectively, the parameter estimates and posterior membership probabilities obtained upon convergence of **Algorithm 2**, we proceed to set $\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}^{(\infty)}$ as our estimate for $\boldsymbol{\theta}$. Moreover, the $Z_i^{(\infty)}$ and $1 - Z_i^{(\infty)}$ are the probabilities that an observation's error term came from, respectively, the Gaussian component or the exponential component. A decision rule can then be defined to determine component membership based on the $Z_i^{(\infty)}$s when compared to a pre-determined cut-off probability p^*. Specifically, the model-based clustering strategy involving our estimated flare regression model is to classify observation i as belonging to the exponential component if $1 - Z_i^{(\infty)} \geq p^*$, otherwise it is classified as belonging to the Gaussian component. The value used for p^* in our analysis will be discussed later.

While traditional EM algorithms are sensitive to the choice of initial values, thus potentially influencing final estimation outcomes (Dempster et al., 1977; Press et al., 2007; Bilmes, 1998), our ECM algorithm exhibits satisfactory precision in

estimation even when starting from initial values far from the true parameters. The default process for generating initial values (i.e., at $t = 0$) is as follows. $\boldsymbol{\beta}^{(0)}$ is set at the estimates from the simple linear regression fit to the data. We then take the residuals from that fit to help with the remaining initial values. $\sigma^{(0)}$ is generated from an exponential distribution with rate equal to the inverse of the square root of the residual sum of squares ($1/\sqrt{RSS}$); i.e., the mean equals $\sqrt{RSS}$. $\alpha^{(0)}$ is generated from a folded normal distribution with location equal to the inverse of the sum of all positive residuals and a standard deviation of 1. This reflects enough variability in terms of likely candidate values for α. Finally, $\lambda^{(0)}$ is determined by the proportions resulting from K-means clustering ($K = 2$) on the residuals. We employ this strategy throughout all of our numerical analyses.

Standard errors for mixture models like our flare regression model can be estimated in various ways. We briefly highlight two ways. First is to simply bootstrap to obtain the standard errors (see Chapter 2 of McLachlan and Peel, 2000). Second is to employ the method due to Louis (1982), which calculates the *observed-data information matrix* as the difference between the *complete-data information matrix* and the *missing-data information matrix*; i.e.,

$$\mathbf{I}(\boldsymbol{\theta}) = -\mathbb{E}_{\widehat{\boldsymbol{\theta}}}\left(\frac{\partial^2 \ell_c(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right) - \mathbb{E}_{\widehat{\boldsymbol{\theta}}}\left[\left(\frac{\partial \ell_c(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial \ell_c(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)^\top\right] + \mathbb{E}_{\widehat{\boldsymbol{\theta}}}\left(\frac{\partial \ell_c(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)\mathbb{E}_{\widehat{\boldsymbol{\theta}}}\left(\frac{\partial \ell_c(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)^\top,$$

where, again, $\ell_c(\boldsymbol{\theta})$ is the complete-data loglikelihood in (3.3). Detailed derivations of $\mathbf{I}(\boldsymbol{\theta})$ are in the Supplemental Material (S.5).

3.2.1 Convergence of the ECM Algorithm

We will now show the convergence of our ECM algorithm. For brevity, we denote the above ECM algorithm map as $A(\boldsymbol{\theta})$. Also, denote the update $\boldsymbol{\beta}^{(t+1)} \equiv m(\boldsymbol{\beta}^{(t)})$ and $A(\boldsymbol{\theta}_2; \boldsymbol{\theta}_1)$ as the iterative updating procedure of A assuming a fixed $\boldsymbol{\theta}_1 = \boldsymbol{\beta}$.

Theorem 2 *For any fixed $\boldsymbol{\theta}_2$, the gradient updating procedure m converges to a local maximum $\boldsymbol{\beta}^{(\infty)}$.*

Proof. Assuming $\dim(\boldsymbol{\beta}) = p$, $p \in \mathbb{N}^+$, the matrix $d^2m/d\boldsymbol{\beta}^2$ is negative-definite in every iteration t. Also, $m(\boldsymbol{\beta})$ is a continuous concave function in $\mathbb{R}^p$. Hence, the set $\{\boldsymbol{\beta} \in \mathbb{R}^p : m(\boldsymbol{\beta}) \geq c\}$ is compact for every constant c. The result follows as an immediate consequence of Proposition 1 in Lange (1995). □

Theorem 3 *For any fixed $\boldsymbol{\theta}_1$, the iterative updating procedure $A(\boldsymbol{\theta}_2; \boldsymbol{\theta}_1)$ converges to a local maximum and $\mathbf{Q}(\boldsymbol{\theta}_2^{(t+1)}; \boldsymbol{\theta}_1) > \mathbf{Q}(\boldsymbol{\theta}_2^{(t)}; \boldsymbol{\theta}_1)$.*

Proof. Assume a fixed $\boldsymbol{\theta}_2$. Because the updating procedure m converges to the point $\boldsymbol{\beta}^{(\infty)}$, following the result from Proposition 2 in Lange (1995), we can conclude that for all sufficiently large t, either $\boldsymbol{\beta}^{(t)} = \boldsymbol{\beta}^{(\infty)}$ or $\mathbf{Q}(\boldsymbol{\theta}_1^{(t+1)}; \boldsymbol{\theta}_2) > \mathbf{Q}(\boldsymbol{\theta}_1^{(t)}; \boldsymbol{\theta}_2)$, where recall that $\boldsymbol{\theta}_1 = \boldsymbol{\beta}$. Then, given any fixed $\boldsymbol{\beta}$, the existence of the closed-form MLEs of $\boldsymbol{\theta}_2$ is guaranteed; see Equations (3.8)–(3.10). Assuming a fixed $\boldsymbol{\theta}_1$, the iterative updating procedure $A(\boldsymbol{\theta}_2; \boldsymbol{\theta}_1)$ is, thus, a standard EM algorithm. Hence, the convergence of $A(\boldsymbol{\theta}_2; \boldsymbol{\theta}_1)$ and the monotonicity of $\mathbf{Q}$ (i.e., $\mathbf{Q}(\boldsymbol{\theta}_2^{(t+1)}; \boldsymbol{\theta}_1) > \mathbf{Q}(\boldsymbol{\theta}_2^{(t)}; \boldsymbol{\theta}_1)$) is guaranteed by Wu (1983). □

Theorem 4 *All the limit points of the ECM sequence above are stationary points of the observed-data loglikelihood* $\ell_o(\boldsymbol{\theta})$.

Proof. From Theorems 2 and 3, we conclude $\ell_c(\boldsymbol{\theta}^{(t+1)}) > \ell_c(\boldsymbol{\theta}^{(t)})$ for every t. Since the objective function $\ell_c(\boldsymbol{\theta})$ is a density function belongs to the exponential family, it is jointly continuous and concave in $\boldsymbol{\theta}$. Hence, the corresponding CM-step always converges to a stationary point. Thus, the result is an immediate consequence of Theorem 4 in Meng and Rubin (1993). □

In summary, unlike the block-relaxation method, the ECM algorithm is guaranteed to converge to a stationary point of the loglikelihood function, assuming the flare regression model. Due to the fact that the density function of the flare regression model belongs to the exponential family, the limiting point to which the ECM algorithm converges is a local maximum of the likelihood function.

4 Simulation Study

We conduct a large simulation study to assess the performance of the ECM algorithm for the flare regression model. This involves calculating and comparing root-mean-square errors (RMSEs) and mean biases, as well as analyzing its robustness under the presence of outliers and contaminated data, and classification performance. Finally, we conduct a broader model comparison study with three other candidate models, including the EMG regression model, that are also estimated using the simulated data. Bayesian information criterion (BIC; Schwarz, 1978) values are calculated to characterize the performance of the flare regression

model and its corresponding ECM algorithm relative to the estimates obtained from the other candidate models.

We consider two different conditions for the regression predictors: one with a single predictor and one with two predictors. The predictors under each condition are generated as $x_{i,j} \sim Unif[-10, 10]$, $i = 1, \ldots, n$, $j = 2, 3$, and, again, setting $x_{i,1} \equiv 1$. We further consider three different scenarios on the mixture components for the errors: well-separated components, moderately-separated components, and overlapping components. For each scenario, we randomly generated $B = 1000$ Monte Carlo samples for each of the sample sizes $n \in \{100, 500, 1000\}$. The explicit parameter settings for all 12 data-generating models are given in Table 1. Please also refer to Figures 3–6 in the Supplemental Material (S.7) for visualizations of these simulation settings.

4.1 General Estimation Performance

Tables of the RMSEs and biases are given in Tables 2–5 in the Supplemental Material (S.6). Visualizations of these tabulated results are also given in Figures 7–18 of the Supplemental Material (S.7). From these results, we can summarize some of the behavior exhibited by the RMSEs and biases across the 12 simulation models.

In 11 out of the 12 simulation settings, both the calculated RMSEs and mean biases show sufficiently low orders of magnitude in absolute value, with setting M12 being the only exception. This demonstrates satisfactory precision of the

Table 1: Parameter settings for the simulation regarding the flare regression model

Setting	**Component Structure**	$(\lambda, 1-\lambda)$	$\boldsymbol{\beta}$	σ	α
M1	Well-Separated	$(0.333, 0.667)$	$(9, 3)$	0.5	0.05
M2	Moderately-Separated	$(0.333, 0.667)$	$(9, 3)$	0.5	0.17
M3	Overlapping	$(0.333, 0.667)$	$(9, 3)$	0.5	0.5
M4	Well-Separated	$(0.9, 0.1)$	$(9, 3)$	0.5	0.05
M5	Moderately-Separated	$(0.9, 0.1)$	$(9, 3)$	0.5	0.17
M6	Overlapping	$(0.9, 0.1)$	$(9, 3)$	0.5	0.5
M7	Well-Separated	$(0.5, 0.5)$	$(-2, 1, 13)$	0.5	0.04
M8	Moderately-Separated	$(0.5, 0.5)$	$(-2, 1, 13)$	0.5	0.2
M9	Overlapping	$(0.5, 0.5)$	$(-2, 1, 13)$	0.5	0.5
M10	Well-Separated	$(0.9, 0.1)$	$(-2, 1, 13)$	0.5	0.04
M11	Moderately-Separated	$(0.9, 0.1)$	$(-2, 1, 13)$	0.5	0.2
M12	Overlapping	$(0.9, 0.1)$	$(-2, 1, 13)$	0.5	0.5

ECM algorithm. The imprecise parameter estimates from setting M12 occurs due to the fact that only a small proportion of data were generated from the exponential component ($\lambda = 0.9$) and that the exponential rate was set to be a large value ($\alpha = 0.5$). A small mixing proportion for the exponential component, along with this larger exponential rate, will obfuscate the identifiability of the mixture model. Like traditional EM algorithms, estimating with ECM algorithms suffer when faced with model identifiability problems.

In most of the simulation settings, both the calculated RMSEs and mean biases noticeably decrease when the sample size increases from $n = 100$ to $n = 1000$. As expected, this behavior shows that our ECM algorithm, like other optimization algorithms, tends to perform better as the sample size becomes larger.

Finally, in most of the simulation settings, the ECM algorithm outputs estimates with lower RMSE and mean bias values under the simulation scenario with well-separated components, and outputs estimation results with higher RMSE and mean bias values under the simulation scenarios with overlapping components. This shows that the ECM algorithm consistently produces more precise estimates when data arise from a mixture with well-separated components. Similar to the reason noted earlier about the subpar performance using data generated from setting M12, a lack of model identifiability emerges when the mixture components heavily overlap with each other.

4.2 Classification Analysis

As introduced in Section 3.2, the flare model and its corresponding ECM algorithm have a unique advantage over the EMG regression model in performing model-based clustering on data that exhibit distinct components: one component with a linear trend and the other with a positive residual. In this subsection, we assess its performance using the mean of the number of correct allocations (MCA) as the criterion. This measure quantifies the accuracy of the clustering by calculating the average number of data points correctly assigned to their respective clusters.

Table 6 in the Supplemental Material (S.6) presents the table of Misclassification

Error Rates (MCAs) and their relative percentages, each was calculated from $B = 1000$ Monte Carlo samples, featuring different settings for the cut-off probability $p^* \in \{0.5, 0.85\}$. The ECM algorithm exhibits excellent clustering performance under both p^* values. Across model settings M1–M12, the algorithm consistently achieved high accuracy: for well-separated cases, it correctly identified over 95% of all clusters; for moderately-separated cases, it correctly identified over 85% of all clusters (mostly exceeding 90%); and for overlapping cases, it correctly identified over 64% of all clusters (with a majority exceeding 80%).

4.3 Robustness Analysis

4.3.1 Outlier Test

Frequently, technical or human errors lead to outliers, posing significant challenges when attempting to achieve a good fit for the model in various empirical studies. In this investigation, we sought to assess the robustness of our flare regression model and its corresponding ECM estimation algorithm by applying them to data deliberately contaminated with outliers.

Following the simulation settings outlined at the beginning of Section 4, we adopt the main structures from models M1–M12. We then introduce outliers with extreme positive skewness. These outliers are generated from an exponential distribution with a rate parameter of 0.001. We simulate outlier percentages of 1%, 5%, and 10% of the original simulated dataset. For each percentage value, we randomly generated $B = 1000$ Monte Carlo samples for the sample size $n = 200$. Subsequently, the dataset, now containing outliers, will be fitted with the flare model.

We will then evaluate the performance of the ECM algorithm by comparing the RMSEs and mean biases with our earlier results obtained from the uncontaminated data, which did not include outliers. The specific parameter settings and results are provided in Tables 7–8 of the Supplemental Material (S.6).

Our ECM algorithm consistently demonstrated robust performance across various sample sizes, even in the presence of outliers. Across different outlier percentages, our model exhibited comparable performance to data without outliers. Specifically, model configurations with well-separated and moderately-separated structures reported similar magnitudes for RMSEs and mean biases, irrespective of the outlier percentage (including 1% simulated outliers). However, when the outlier percentages increased to 5% and 10% of the original simulated dataset, we observed slightly- to moderately-elevated values for both RMSEs and mean biases compared to the values calculated from the outlier-free data. This phenomenon can be attributed to the identifiability issue inherent in all mixture models.

4.3.2 t-Distribution Contamination Test

To further test the robustness of the ECM algorithm, we simulate a contaminated dataset following the model described in Eqn(2.1), with an error structure characterized by a mixture of a t-distribution with degrees of freedom ν (for heavier tails) and an exponential distribution with rate α. Similar to the setting outlined at the beginning of Section 4, we consider a single predictor, generated as $x_{i,2} \sim \text{Unif}[-10, 10]$, $i = 1, \ldots, n$, while setting $x_{i,1} \equiv 1$. We further explore a total of six scenarios with both well-separated and overlapping structures for the

mixture components of the errors, and varying the t-distribution's degrees of freedom $\nu \in \{5, 50, 500\}$. For each scenario, we randomly generated $B = 1000$ Monte Carlo samples for a sample size of $n = 200$. The explicit parameter settings for the six new mixture-of-regressions models are given in Table 2.

Table 2: Parameter settings for the simulation regarding the mixture of t and exponential regression model

Setting	**Component Structure**	$(\lambda, 1-\lambda)$	$\boldsymbol{\beta}$	ν	α
M13	Well-Separated	$(0.6, 0.4)$	$(-2, 6)$	5	0.05
M14	Well-Separated	$(0.6, 0.4)$	$(-2, 6)$	50	0.05
M15	Well-Separated	$(0.6, 0.4)$	$(-2, 6)$	500	0.05
M16	Overlapping	$(0.4, 0.6)$	$(6, -2)$	5	0.5
M17	Overlapping	$(0.4, 0.6)$	$(6, -2)$	50	0.5
M18	Overlapping	$(0.4, 0.6)$	$(6, -2)$	500	0.5

When fitting the flare model to the contaminated data simulated from models M13 to M18, as outlined in Table 2, we compare the estimated regression coefficients ($\hat{\boldsymbol{\beta}}$) with the theoretical regression coefficients ($\boldsymbol{\beta}$). Additionally, we assess the agreement between the estimated Gaussian variance ($\hat{\sigma}^2$) and the theoretical t variance ($\frac{\nu}{\nu-2}$), and between the estimated exponential rate ($\hat{\alpha}$) and the theoretical exponential rate (α), using metrics such as RMSEs and mean biases. With remarkably low magnitudes in both RMSEs and mean biases, our observations highlight the excellent robustness of the ECM algorithm when applied to data simulated from structurally different models. Further details, including RMSEs

and mean biases after fitting the ECM algorithm to data generated from models M13–M18, are provided in Tables 9 and 10 of the Supplemental Material (S.6).

4.4 Broader Model Comparison Study

We further examine the efficacy of the flare regression model by fitting three other models to the simulated data: the EMG regression model, the classic linear regression model, and a two-component mixture-of-linear-regressions model. The EMG regression model and its corresponding block-relaxation algorithm are as presented in Sections 2 and 3. The classic linear regression model is just the model in (2.1), but where $\epsilon_i \sim N(0, \sigma^2)$, for $i = 1, \ldots, n$. Here, the parameter vector of interest is $(\boldsymbol{\beta}^\top, \sigma^2)^\top$, which is estimated by ordinary least squares. For the two-component mixture of linear regressions, we have:

$$y_i = \begin{cases} \mathbf{x}_i\boldsymbol{\beta}_1 + \epsilon_{i1}, & \text{with probability } \lambda; \\ \mathbf{x}_i\boldsymbol{\beta}_2 + \epsilon_{i2}, & \text{with probability } 1-\lambda, \end{cases}$$

where the $\epsilon_{ij} \sim N(0, \sigma_j^2)$ are (conditionally) *iid*, $i = 1, \ldots, n$ and $j = 1, 2$. In this mixture model, the parameter vector of interest is $(\lambda, \boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_1^\top, \sigma_1^2, \sigma_1^2)^\top$, whose closed-form MLEs can easily be derived using a standard EM algorithm (De Veaux, 1989). This EM algorithm is implemented by the `regmixEM()` function in the `R` package `mixtools` (Benaglia et al., 2009).

On average, over 90% of the time the flare regression model outperforms the other candidate models in terms of having the lowest BIC values. Please see Table 1 of the Supplemental Material (S.6) for detailed percentages of the lowest BIC

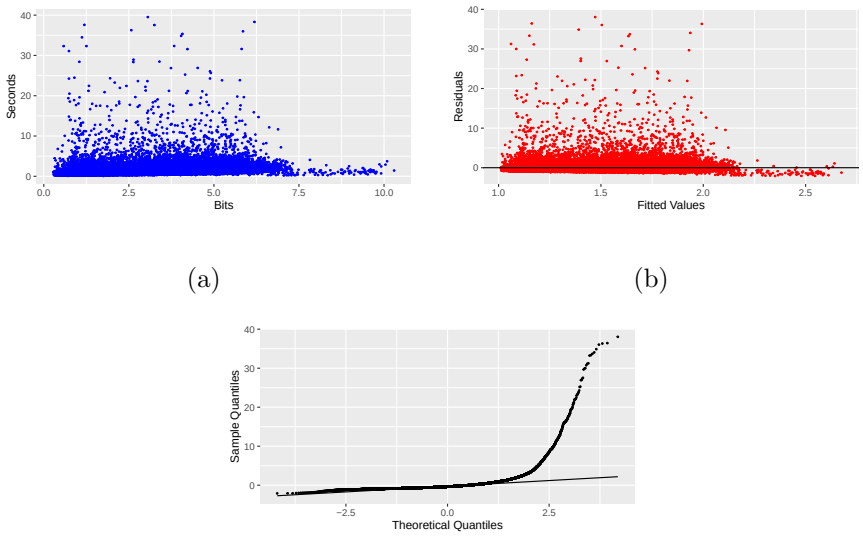
values from all the candidate models estimated using the data generated from each simulation setting. Overall, this demonstrates the strong performance of the flare regression model, especially in the context of data that one might typically consider modeling with EMG regression.

5 Application: Human Aiming Performance Data

5.1 Data Description and Model Settings

We now analyze data from the field study by Chapuis et al. (2007), which was produced by unobtrusively collecting mouse input and corresponding GUI data from 24 users over several months. The dataset consists of more than 2 million movements. Many variables of interest were collected, including time, cursor position, mouse movements, mouse and button events (click, drag, long click), type and properties of the selected target, such as size and role of the target (e.g., resizing button, edge of a window), as well as information regarding the system used (which input device, desktop/laptop, Operating System). In this work, we used only information on movement time, distance to the target, and target size, which is consistent with applying Fitts' model.

The theoretical model is given by $y = \beta_0 + \beta_1 x$, where $y = t_e/1000$ (converting milliseconds to seconds) and $x = \log_2\left(1 + \frac{dist}{\min(w_t, h_t)}\right)$. The variable x is considered a difficulty measure, whose units are in bits. Typically, in controlled studies with computer mice, $\beta_0 \in [-0.1, 0.1]$ and $\beta_1 \in [0.1, 0.2]$, where β_0 is in seconds, and β_1 is in seconds/bit. Compared to data typically collected in controlled studies, these



(a) (b)

(c)

Figure 1: (a) Scatterplot of real-world aiming performance from a user (User 1 in the Supporting Information), (b) scatterplot of the residuals versus the fitted values from a classic linear regression fit, and (c) the corresponding quantile-quantile plot

data display noticeable positive skewness because users do not routinely try to minimize their task completion time. Figure 1(a) is a scatterplot of the data from one user in our data. Notice the variability and considerable positive skewness in the task completion times. Unlike aiming data collected in controlled studies, a linear regression assuming zero-centered Gaussian noise is not an appropriate model for the present data. This is indicated by a plot of the residuals versus the fitted values (Figure 1(b)) and the corresponding quantile-quantile plot (Figure 1(c)) when fitting a simple linear regression model to the data.

In Gori and Rioul (2019), the EMG regression model was estimated with a very

small subset of the "in the wild" data. Compared to classic linear regression with Gaussian errors, the estimated EMG regression parameters fall within the typical range of those for controlled experiments, and the fitted line matches well with the idea of minimum movement time. We extend this previous work by fitting and comparing the four models used in the simulation study discussed in Section 4.4. Additionally, instead of only the small subset analyzed in Gori and Rioul (2019), we use the entire "in the wild" dataset when estimating the four candidate models for each of the 24 users.

5.2 Data Truncation and Estimation Results

Besides the characteristic positive skewness of the "in the wild" data, technical difficulties associated with trajectory segmentation frequently produce outliers. To obtain informative estimates, outliers produced by technical errors should be eliminated. However, there is no definitive indicator as to when an observation is an outlier. Thus, four different cut-off thresholds are investigated: $T = 10s$, $T = 20s$, $T = 30s$, and $T = 40s$. When we set a fixed cut-off threshold, only observations with response time y less than the threshold will be considered (i.e., $y_i \leq T$). As the cut-off threshold increases, more extreme values of long reaction times are present in the corresponding truncated data.

After fitting four candidate models discussed in Section 4.4, we find the EMG regression model and the flare regression model consistently outperform the other two candidate regression models (i.e., simple linear regression with Gaussian errors and the two-component mixture of linear regressions) by producing significantly

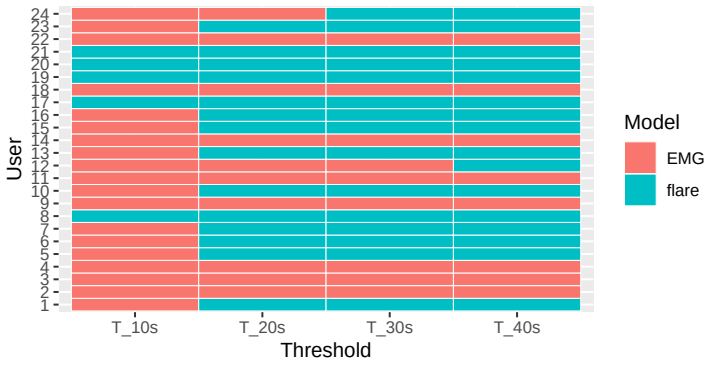


Figure 2: BIC comparisons for the EMG regression and flare regression model fits with different cut-off values. Red cells indicate lower BIC values for the EMG model; blue cells indicate lower BIC values for the flare regression model

lower BIC values under all four cut-off thresholds. As we increase the cut-off threshold, however, more extreme values are naturally present, and the flare regression model tends to perform better than the EMG regression model. Figure 2 provides a visualization of how the BIC values for each participating user change as the cut-off threshold increases. In terms of their BIC values, blue cells correspond to the EMG regression being a better fit, while the red cells correspond to the flare regression model being a better fit. As the threshold increases, more red cells appear in the figure. Thus, we see the ability of the flare regression model relative to better characterize more extreme values relative to the EMG regression model for these aiming performance data.

After balancing the need to eliminate outliers produced by possible technical errors and the necessity of preserving observations with long movement times, the cut-

off threshold $T = 40s$ is selected. When the entire dataset is truncated using $T = 40s$, the flare regression model outperforms all the other candidate models with much lower BIC values for 16 out of the 24 participating users. Moreover, we receive similar parameter estimates after fitting the four candidate models to the data from each user. Similar parameter estimates show users tend to have similar movement times while completing the aiming tasks. Exact parameter estimates obtained for the four candidate models using the truncated data with threshold $T = 40s$ and BIC comparisons are in, respectively, Tables 12–15 and Table 11 of the Supplemental Material (S.6). Besides the BIC values, linear regression yields parameter estimates outside the typical intervals for β_0 and β_1 in controlled studies. The EMG and flare regression models, on the other hand, yield parameter estimates within the typical intervals. However, the two models behave differently: the EMG regression model tends to yield more estimates for the intercept inside the typical interval, whereas the flare regression model tends to yield more estimates for the slope inside the typical interval. A visualization for this comparison is in Figure 19 of the Supplemental Material (S.7). Comparing to the intercept, researchers consider the slope to be a more informative parameter when measuring movement difficulty (Zhai, 2004; Guiard and Olafsdottir, 2011).

5.3 Classification and Interpretations

As noted in Section 1, both the EMG regression model and our flare regression model are able to effectively handle data with positive residuals relative to their underlying mean trend, which is a prominent feature of this "in the wild" data.

However, as demonstrated in Section 3.2, we can further perform model-based clustering on this "in the wild" data based on $Z_i^{(\infty)}$, the posterior membership probabilities. Those observations classified to the Gaussian component would represent the typical movement times of individuals in a controlled study, consistent with Fitts' law. Those observations classified to the exponential component would represent where a user is not trying to maximize their performance as well as any possible outliers that have not been removed due to the truncating strategy employed earlier.

For example, Figure 3 is a scatterplot for the same user in Figure 1(a) after fitting the flare regression model. In this figure, the flare regression model fit has been overlaid along with each observation color-coded according to their component membership based on their maximum posterior membership probability (i.e., the cut-off probability p^* is set to be 0.50). We have been able to effectively characterize the regions where the user has almost certainly not been performing in an optimal capacity for the aiming task. Moreover, this region could still include some outlying values associated with trajectory segmentation.

Note that we have done a hard classification based on an observation's posterior membership probabilities. However, the noticeable delineation between the Gaussian component and exponential component, as seen in Figure 3, appears in each user's fit. Further examination shows that the posterior membership probabilities unsurprisingly hover around 0.50 for the two components in this region as this is where the two components have more substantial overlap. If interested, one could

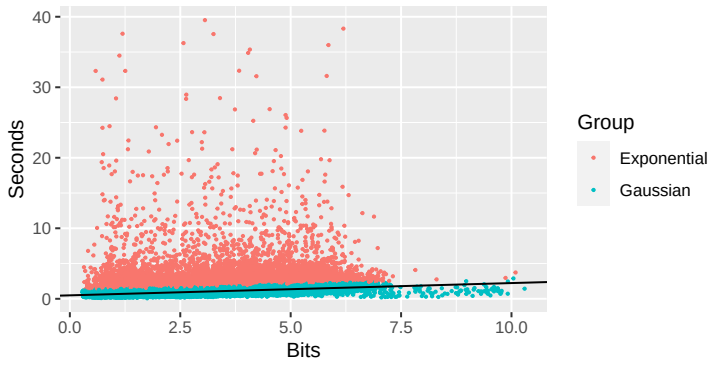


Figure 3: Scatterplot of the user's data in Figure 1(a) (User 1), but with the flare regression model fit ($\hat{\beta}_0 = 0.49$, $\hat{\beta}_1 = 0.17$) overlaid along with each observation color-coded according to their component membership based on their maximum posterior membership probabilities; refer to the first row of Table 12 in the Supplemental Material (S.6) for detailed estimation results

apply a color gradient relative to the membership probabilities to visualize the uncertainty of assignment to one component over the other, thus providing a more nuanced interpretation about the user's performance.

The "in the wild" data distinguishes itself from other data collected in controlled studies by displaying observations with extremely long task completion times. Hence, identifying outliers is essential to obtain informative results. In this study, we proceed with a conservative approach by selecting a uniform cut-off threshold ($T = 40s$), and drop all of the observations that exceed this threshold. Researchers may be interested in finding alternative methods to remove outliers. In the flare regression model, observations with exceedingly long duration times are highly

likely to be classified to the exponential component. This clustering feature allows a framework for outlier removal and for HCI researchers to focus on those observations that are more consistent with what is typically observed in controlled studies. Subsequently, even more candidate models could be investigated for characterizing the remaining observations in the data, thus allowing more nuanced comparisons between "in the wild" data and data from controlled studies for HCI research.

5.4 System Running Time

When the sample size gets (excessively) large, system running time is used in assessing the performance of algorithms. All of the algorithms used in this study were implemented in `R`. The mean sample size of the user datasets is about 19667. The mean system running time of the ECM algorithm is 8.8502 seconds,whereas, the mean system running time of the block-relaxation algorithm is 193.6914 seconds. For every user, the system elapsed time of the ECM algorithm is significantly shorter than the block-relaxation algorithm. Table 16 in the Supplemental Material (S.6) summarizes each user's sample size and system elapsed times across the four candidate models. Theoretical and technical reasons behind this empirical finding remain a potential future direction of research.

6 Concluding Remarks

The EMG distribution is a practical model applied in various fields when researchers encounter positively skewed data, especially when it involves timing studies of tasks with human subjects like the human aiming performance data that

motivated this study. This paper addressed some of the computational challenges in estimating an EMG regression model with multiple predictors by developing an iterative block-relaxation algorithm. Even though some concavity properties of the EMG regression are proved, the fact the the EMG distribution is not a member of the exponential family prevents us from guaranteeing global concavity of the loglikelihood and convergence of the block-relaxation algorithm. Alternatively, we introduced our novel flare regression model consisting of a two-component mixture structure on the errors, consisting of a Gaussian component and an exponential component. We developed an ECM algorithm for estimation, which unlike the block-relaxation algorithm, is guaranteed to converge to a local maximum of its likelihood function. After obtaining point estimates of the flare regression model, we briefly addressed the calculation of estimated standard errors for the parameter estimates.

Both the extensive simulation study and the analysis of the human aiming performance data showed significant advantages of the flare regression model over the EMG regression model and other existing regression models. Not only is the flare regression model fit typically better than the EMG regression model fit (in terms of BIC values), the former also provides us with additional insight into different performance regions in the human aiming task. Moreover, a timing comparison between the block-relaxation method for the EMG regression and the ECM algorithm for the flare regression shows superior performance for the latter. Overall, we have shown that the flare regression model is highly efficacious as a way for

characterizing the human aiming performance data analyzed in this work.

As noted in the analysis of Section 5, there is, obviously, subject-to-subject variability in terms of performance on this task. Incorporation of random effects to allow for such subject heterogeneity would likely provide an even more informative model. Thus, generalizing both the EMG and flare regression models by incorporating random effects, and then comparing the results, would be an informative direction for future research.

Acknowledgements

The authors thank the editor, the associate editor, and an anonymous referee for helpful comments on an earlier draft of this article.

References

Benaglia, T., Chauveau, D., Hunter, D. R., and Young, D. S. (2009). mixtools: An R Package for Analyzing Finite Mixture Models. *Journal of Statistical Software*, **32**(6), 1–29. URL http://www.jstatsoft.org/v32/i06/.

Bernardini, M. G., Margutti, R., Chincarini, G., Guidorzi, C., and Mao, J. (2011). Gamma-Ray Burst Long Lasting X-ray Flaring Activity. *Astronomy and Astrophysics*, **526**(A27), 1–9.

Bilmes, J. A. (1998). A Gentle Tutorial of the EM Algorithm and Its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models.

CTIT Technical Reports Series. URL https://api.semanticscholar.org/CorpusID:260604709.

Chapuis, O., Blanch, R., and Beaudouin-Lafon, M. (2007). Fitts' Law in the Wild: A Field Study of Aimed Movements. Technical Report Number 1480, Laboratoire de Recherche en Informatique. URL https://insitu.lri.fr/~bibli/Rapports-internes/2007/RR1480.pdf.

de Leeuw, J. (1994). Block-Relaxation Algorithms in Statistics. In Bock, H.-H., Lenski, W., and Richter, M. M., editors, *Information Systems and Data Analysis*, pages 308–324, Berlin, Heidelberg, 1994. Springer Berlin Heidelberg. ISBN 978-3-642-46808-7.

De Veaux, R. D. (1989). Mixtures of Linear Regressions. *Computational Statistics and Data Analysis*, **8**(3), 227–245.

Dempster, A. P., Laird, N., and Rubin, D. (1977). Maximum Likelihood Estimation From Incomplete Data via the EM Algorithm (with Discussion). *Journal of the Royal Statistical Society Series B (Methodological)*, **39**(1), 1–38.

Evarts, E., Teräväinen, H., and Calne, D. (1981). Reaction Time in Parkinson's Disease. *Brain*, **104**(Pt 1), 167–86. doi: 10.1093/brain/104.1.167.

Fitts, P. M. (1954). The Information Capacity of the Human Motor System in Controlling the Amplitude of Movement. *Journal of Experimental Psychology*, **47**(6), 381–391.

Goldberg, K., Faridani, S., and Alterovitz, R. (2014). Two Large Open-Access Datasets for Fitts' Law of Human Motion and a Succinct Derivation of the Square-Root Variant. *IEEE Transactions on Human-Machine Systems*, **45**(1), 62–73.

Golubev, A. (2010). Exponentially Modified Gaussian (EMG) Relevance to Distributions Related to Cell Proliferation and Differentiation. *Journal of Theoretical Biology*, **262**(2), 257–266.

Gordon, B. and Carson, K. (1990). The Basis for Choice Reaction Time Slowing in Alzheimer's Disease. *Brain and Cognition*, **13**(2), 148–66. doi: 10.1016/0278-2626(90)90047-r.

Gori, J. and Rioul, O. (2019). Regression to a Linear Lower Bound with Outliers: An Exponentially Modified Gaussian Noise Model. 27th European Signal Processing Conference (EUSIPCO). doi: 10.23919/EUSIPCO.2019.8902946.

Gori, J., Rioul, O., Guiard, Y., and Beaudouin-Lafon, M. (2017). One Fitts' Law, Two Metrics. 16th IFIP Conference on Human-Computer Interaction (INTERACT), pages 525–533. doi: 10.1007/978-3-319-67687-6_36.

Gori, J., Rioul, O., and Guiard, Y. (2018). Speed-accuracy Tradeoff: A Formal Information-Theoretic Transmission Scheme (Fitts). *ACM Transactions on Computer-Human Interaction (TOCHI)*, **25**(5), 1–33.

Guiard, Y. and Olafsdottir, H. B. (2011). On the Measurement of Movement Difficulty in the Standard Approach to Fitts' Law. *PLOS ONE*, **6**(10), 1–

15. doi: 10.1371/journal.pone.0024389. URL https://doi.org/10.1371/journal.pone.0024389.

Kalambet, Y., Kozmin, Y., Mikhailova, K., Nagaev, I., and Tikhonov, P. (2011). Reconstruction of Chromatographic Peaks Using the Exponentially Modified Gaussian Function. *Journal of Chemometrics*, **25**(7), 352–356.

Lange, K. L. (1995). A Gradient Algorithm Locally Equivalent to the EM Algorithm. *Journal of the Royal Statistical Society Series B (Methodological)*, **57**(2), 425–437.

Louis, T. A. (1982). Finding the Observed Information Matrix when Using the EM Algorithm. *Journal of the Royal Statistical Society Series B (Methodological)*, **44**(2), 226–233.

MacKenzie, I. S. and Buxton, W. (1992). Extending Fitts' Law to Two-dimensional Tasks. Proceedings of the SIGCHI conference on Human factors in computing systems, pages 219–226. doi: 10.1145/142750.142794.

McLachlan, G. and Peel, D. (2000). *Finite Mixture Models*. John Wiley & Sons, New York.

Meng, X. and Rubin, D. B. (1993). Maximum Likelihood Estimation via the ECM Algorithm: A General Framework. *Biometrika*, **80**(2), 267–78.

Palmer, E. and Horowitz, T. (2011). What are the Shapes of Response Time Distributions in Visual Search? *Journal of Experimental Psychology: Human Perception and Performance*, **37**(1), 58–71.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (2007). *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press, USA, 3 edition. ISBN 0521880688.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, **6**(2), 461–464.

Silver, J. (2009). Microarray Background Correction: Maximum Likelihood Estimation for the Normal–Exponential Convolution. *Biostatistics*, **10**(2), 352–363.

Wong, A. L., Goldsmith, J., Forrence, A. D., Haith, A. M., and Krakauer, J. W. (2017). Reaction Times Can Reflect Habits Rather Than Computations. *eLife.*, **6**(e28075.), 1–18. doi: 10.7554/eLife.28075.

Wu, C. J. (1983). On the Convergence Properties of the EM Algorithm. *The Annals of Statistics*, **11**(1), 95–103.

Zhai, S. (2004). Characterizing Computer Input with Fitts' Law Parameters—the Information and Non-Information Aspects of Pointing. *International Journal of Human-Computer Studies*, **61**(6), 791–809. ISSN 1071-5819. doi: https://doi.org/10.1016/j.ijhcs.2004.09.006. URL https://www.sciencedirect.com/science/article/pii/S1071581904001041.