

# Per-Subject Oculomotor Plant Mathematical Models and the Reliability of Their Parameters

KATERYNA MELNYK, Texas State University, USA

LEE FRIEDMAN, Texas State University, USA

DMYTRO KATRYCHUK, Texas State University, USA

OLEG KOMOGORTSEV, Texas State University, USA

The practical value of oculomotor plant mathematical models (OPMMs) has been demonstrated across various domains, including biometrics and eye movement prediction. To further enhance their utilization, new optimization approaches are commonly developed and introduced within the research community. In this study, we demonstrate a new integration of a previously developed per-subject optimization procedure for an Enderle OPMM and introduce methods to evaluate the reliability of OPMM parameters using the intraclass correlation coefficient (ICC) and Kendall's Coefficient of Concordance (KCC). We evaluated two per-subject OPMM models, Bahill and Enderle, using the 'GazeBase' eye movement dataset. The models differed in accuracy, expressed as the per-sample error in degrees of visual angle, based on the differences between the actual and predicted saccade trajectories. We found that some of the parameter estimates for both models were quite unreliable. We discuss the importance of addressing low-reliability issues and suggest methods to modify the models to enhance reliability and performance.

CCS Concepts: • **Computing methodologies** → **Modeling and simulation**; • **Human-centered computing** → **Empirical studies in HCI**.

Additional Key Words and Phrases: Eye movements, Oculomotor Plant Mathematical Model, Saccade simulation, Temporal persistence

## ACM Reference Format:

Kateryna Melnyk, Lee Friedman, Dmytro Katrychuk, and Oleg Komogortsev. 2024. Per-Subject Oculomotor Plant Mathematical Models and the Reliability of Their Parameters. In . ACM, New York, NY, USA, Article etra24-fp1021, 20 pages. <https://doi.org/10.1145/3654701>

## 1 INTRODUCTION

The oculomotor system consists of several physiological, anatomical, and neurological components. These components interact with each other to control specific eye movements. There is current literature that focuses on saccadic eye movements and attempts to model the basic elements, i.e., the oculomotor plant (OP) and the saccade generation system in the central nervous system (CNS). The OP consists of the eye globe and three pairs of extraocular muscles. The muscles are innervated by the saccade generation system in the CNS.

In 1954, Westheimer presented the first quantitative saccadic eye movement model [Westheimer 1954]. This model was able to fit the eye position data for single horizontal saccades of  $20^\circ$  but was limited to saccades of that amplitude. Further development of the eye movement models was slow due to the lack of empirical data on the exact physiological and anatomical characteristics of the OP components. In 1964, such empirical data was provided by Robinson [Robinson 1964]. During real experiments with a suction contact lens, he was able to investigate more closely the mechanics of the

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

Manuscript submitted to ACM

53 eye and test the assumptions made earlier by Westheimer. Robinson created his model which was more complex and  
54 more accurate than Westheimer's model. More importantly, it produced horizontal saccades over a range of amplitudes  
55 from  $5^\circ$  to  $40^\circ$ . Robinson's model more accurately replicated eye positional signals, but its simulated velocity trajectories,  
56 especially toward the end of each saccade, deviated more from actual velocity signals.  
57

58 Four years later, a new model for the human eye-movement mechanism was developed by Cook and Stark [Cook and  
59 Stark 1968], and soon after, it was improved by Hsu et al. [Hsu et al. 1976]. This model outperformed Robinson's model  
60 in two major ways: (1) the neural pulse generation was more realistic, and (2) the antagonist muscle, in addition to the  
61 agonist muscle, was modeled as a separate entity for the first time. In 1980 Bahill [Bahill 1980] presented an improved  
62 linear 4th order model that more accurately fit real horizontal saccade trajectories and velocity profiles. Starting in  
63 2010, Enderle and co-authors introduced a new family of models [Enderle 2010; Enderle and Zhou 2010; Ghahari and  
64 Enderle 2014, 2015]. While using Bahill's model as a baseline, Enderle et al. developed models for horizontal saccades  
65 and smooth pursuits.  
66  
67

68 At approximately the same time, Komogortsev and colleagues began to publish and use OP models. In [Komogortsev  
69 and Khan 2008], the OPMM was used for the prediction of eye-movement trajectories. A mechanical model of the  
70 OP was transformed into a Kalman filter form (OPKF) that markedly improved the prediction accuracy of saccades  
71 with various amplitudes. Komogortsev and colleagues were also interested in incorporating these OPMMs into their  
72 biometric work which focused on evaluating individual differences in OP characteristics [Komogortsev et al. 2012a,  
73 2010]. OP characteristics were used for "liveness" detection, i.e., OPMMs were used in the discrimination between  
74 living human eyes and artificial/mechanical eyes. In another study [Karpov et al. 2020], the researchers addressed the  
75 effectiveness of estimating OP characteristics and the applicability of OPMMs for real-time eye movement simulations.  
76  
77

78 OPMMs were also applied in medical settings. For example, an OPMM was employed for the correction of strabismus  
79 [Hoerantner et al. 2007; Robinson et al. 1969]. Also, OPMMs were used for the automated detection of mild traumatic  
80 brain injury (mTBI) via the application of eye movement biometrics [Komogortsev and Holland 2014]. In 2018 Wadehn  
81 and co-authors [Wadehn et al. 2018] developed a model-based approach to detect and classify different types of eye  
82 movements (fixations, saccades, and smooth pursuit). They based their analysis of saccades on Bahill's OPMM. They  
83 reported that their algorithm was robust to noise and highly accurate in detecting saccades. In 2019 Wadehn and  
84 co-workers [Wadehn et al. 2019] presented a new approach to the detection and classification of eye movements. For  
85 this, their analysis was based on Enderle's OPMM and this approach was somewhat more accurate.  
86  
87

88 All of the above studies used OPMMs with a common per-saccade optimization design where the parameters are  
89 optimized and fitted for one particular saccade. In 2022, Katrychuk and Komogortsev [Katrychuk and V. Komogortsev  
90 2022] presented a new approach that estimates one set of OPMM parameters (for a version of Bahill's model) based  
91 on all saccades made by each subject. We refer to the models presented in [Katrychuk and V. Komogortsev 2022] as  
92 per-subject models, to distinguish them from classic per-saccade models.  
93

94 In the present study, we concentrated on the per-subject baseline approach used by Katrychuk [Katrychuk and  
95 V. Komogortsev 2022] for Bahill's model. By applying per-subject optimization to Enderle's model, we aim to demonstrate  
96 the value of the current approach. By employing distinct models under the same optimization procedure, we can more  
97 precisely evaluate the performance and usefulness of the per-subject optimization approach for utilization in future  
98 studies.  
99

100 Traditionally, 'fitting accuracy' has served as the standard metric for evaluating OPMMs. In our study, we aim to  
101 address the issue of the test-retest reliability of model parameter estimates. The temporal persistence of the OPMM  
102 parameters was briefly mentioned in Friedman's et al. work [Friedman et al. 2017]. However, a detailed analysis of  
103  
104

105 physiological measures from the OPMM and the identification of the reasons for their unreliability have not been  
106 explored and discussed thoroughly. We employ two reliability-related metrics (the intraclass correlation coefficient  
107 (ICC) and Kendall’s coefficient of concordance (KCC)). We also employed a one-sample t-test to check if the optimized  
108 set of parameters for every subject was statistically significantly different from the default per-saccade OPMM values.  
109

## 111 2 METHODOLOGY

### 113 2.1 Subjects

114 For a detailed description of the subjects and the specific study characteristics, see the “GazeBase Dataset” [Griffith et al.  
115 2021]. Participants were undergraduate students at Texas State University. They were recruited through targeted emails  
116 and in-class announcements. There were 322 participants (151 self-identifying as female, 171 self-identifying as male) in  
117 what we call “Round 1”. For each new round, subjects had to have been recorded in Round 1. Round 2,  $\approx$  3 months from  
118 Round 1, had 136 subjects since not all subjects from Round 1 were available. All of our analysis was based on these 136  
119 subjects with both Round 1 and Round 2 data. Within each round, there were two sessions,  $\approx$  20 minutes apart.  
120

121 *Ethics and Privacy Statement:* All subjects provided informed consent under a protocol approved by the Institutional  
122 Research Board at Texas State University before each round of recording. As part of the consent process, subjects  
123 acknowledged that the resulting data may be disseminated in a de-identified form. No subject-identifying information  
124 is not included in the “GazeBase Dataset”.  
125

### 127 2.2 Eye-Movement Recordings

128 For a detailed description of the recording details, see the report on the “GazeBase Dataset” [Griffith et al. 2021]. Briefly,  
129 monocular (left) eye movements were recorded with an EyeLink 1000 at 1000 Hz, and scaled in degrees of visual angle  
130 (dva). Subjects performed seven tasks during each session of the recording. For the present study, only 3 tasks were of  
131 interest: (1) the Horizontal Saccade Task (HSS), (2) the Random Saccade Task (RAN), and (3) the Reading Task (TEX).  
132 For the HSS task, subjects were presented with a target that was regularly displaced between two positions: a left  
133 position ( $-15^\circ$ ) for 1 second followed by a right position ( $+15^\circ$ ) for another second. During each HSS task, there were  
134 one hundred  $30^\circ$  transitions. The RAN task was designed to elicit visually guided oblique saccades of variable amplitude  
135 through the periodic displacement of a peripheral target. The target was displaced at random locations across the  
136 monitor, ranging from  $\pm 15^\circ$  and  $\pm 9^\circ$  in the horizontal and vertical directions, respectively. During the TEX task subjects  
137 were instructed to read a poem. Overall, the dataset used for this study contains saccades from 136 subjects, performing  
138 3 tasks, each recorded for 2 rounds and 2 consecutive sessions per round.  
139

140 For eye movement classification, we employed an improved version of the previously published method [Friedman  
141 et al. 2018]. The saccade trajectories we employed were based on the saccades classified by this algorithm. To be included  
142 in our study, we only employed saccades with a horizontal amplitude  $\geq 2^\circ$  and a duration between [10..200] ms. The  
143 minimum number of saccades found for any subject, in any round, during any session, was 83. The maximum number  
144 of saccades found for any subject, in any round, during any session, was 462. The median number of saccades per  
145 subject, per round, per session, was 255.  
146

### 147 2.3 Bahill’s Model

148 The base per-saccade Bahill model consists of the 18 physiological parameters [Bahill 1980; Bahill and Stark 1977]:  
149

- 157 •  $K_{SE}^{AG,ANT}$  – series elasticity [Levin and Wyman 1927], which is an element that characterizes the change in  
158 muscle length in response to a change in force, and it is measured in  $gm_{tension}$  per degree -  $\frac{gm_{tension}}{o}$ .
- 159 •  $K_{LT}^{AG,ANT}$  – length-tension. It represents the maximum contractile force that a muscle can generate and is  
160 expressed in  $\frac{gm_{tension}}{o}$ .
- 161 •  $B_{AG}, B_{ANT}$  – force-velocity of the agonist and antagonist muscles. This is the muscular force at a speed of  
162 shortening when the muscle length is near the rest length and it is stimulated, measured in  $\frac{gm_{tension}-sec}{o}$  [Fenn  
163 and Marsh 1935].
- 164 •  $C_{AG}, C_{ANT}, C_{FIX}$  – tension slopes and tension intercept.
- 165 •  $J$  – the inertial mass of the eye globe, and is expressed in  $\frac{gm_{tension}-sec^2}{o}$ .
- 166 •  $B_p$  – passive viscosity of the tissue surrounding the eye globe, it is expressed in  $\frac{gm_{tension}-sec}{o}$ .
- 167 •  $\tau_{act}^{AG,ANT}, \tau_{deact}^{AG,ANT}$  – activation and deactivation time, they are typically measured in milliseconds (*msec*).
- 168 •  $PH_{AG,ANT}, PW_{AG,ANT}$  – neural pulse height and width [Bahill and Stark 1975]. Neural pulse height is typically  
169 expressed in  $\frac{spikes}{sec}$ . Neural pulse width is expressed in *msec*.

173 This is a 6th order model, which means we require six differential equations to describe the OP system. Each  
174 differential equation describes the particular state of the system and can be written in a convenient way using the  
175 matrix notation:  
176

$$177 \quad \dot{x} = Ax + Bu. \quad (1)$$

178 The square matrix  $A$  is a matrix of coefficients,  $u$  is an input vector and  $B$  is a matrix of the weights applied to the  
179 inputs. This model is a linear time-invariant (LTI) system. Therefore, it can be described in the form of a state-space  
180 representation. There was a vector of 6 state variables  $x = \{x_1, x_2, x_3, x_4, x_5, x_6\}$ :

- 183 •  $x_1$  – position of the eye;
- 184 •  $x_2, x_3$  – position of the agonist and antagonist node;
- 185 •  $x_4$  – eye velocity, such that  $x_4 = \dot{x}_1$ ;
- 186 •  $x_5, x_6$  – the agonist and antagonist active-state tension.

188 Following Komogortsev’s strategy [Komogortsev and Khan 2008], by using the approximation of the derivative  
189  $\dot{x}(t) = \frac{x(t+1) - x(t)}{\Delta t}$  we can present OPMM in a state transition (1) form such that:

$$191 \quad x^{t+1} = Ax^t + u. \quad (2)$$

## 193 2.4 Per-Subject Bahill’s Model

194 As mentioned earlier, Bahill’s original model is designed to simulate one saccade at a time and comprises a total of 18  
195 parameters. Subject-specific models base their fits on many saccades (generally hundreds). Because the parameters of  
196 these models account for many saccades, and because all parameters are fit for each subject, these models are more  
197 general and should be more useful in the evaluation of individual differences.

198 It is logical to propose that the neuronal control signal should be a function of saccade amplitude. This was a  
199 fundamental modification in the development of the per-subject optimization approach [Katrychuk and V. Komogortsev  
200 2022]. For the per-subject OPMM, both  $PH$  and  $PW$  for agonist muscle are considered as functions of saccade amplitude  
201 for large ( $> 5^\circ$ ) and small ( $\leq 5^\circ$ ) saccades, separately. This added 9 new parameters:

- 202 • 1 separation parameter  $SL_{TH}$  – a threshold between small and large saccades;
- 203 • 8 neurological parameters  $PH_{INT}^S, PH_{SL}^S, PW_{INT}^S, PW_{SL}^S, PH_{INT}^L, PH_{SL}^L, PW_{INT}^L, PW_{SL}^L$ .

For Bahill's per-subject model [Katrychuk and V. Komogortsev 2022], it is assumed that  $K_{SE}$  and  $K_{LT}$  have the same values for agonist and antagonist muscles as well as  $PH$  and  $PW$  which reduced the number of parameters. This approach was proposed by Komogortsev in 2007 [Komogortsev and Khan 2008]. Additionally, the two activation and deactivation timings are fixed to their defaults. This decreases the number of parameters without a significant reduction in the model performance. As a result in [Katrychuk and V. Komogortsev 2022], Bahill's per-subject model has 18 parameters as well:  $K_{SE}$ ,  $K_{LT}$ ,  $B_{AG}$ ,  $B_{ANT}$ ,  $C_{AG}$ ,  $C_{ANT}$ ,  $C_{FIX}$ ,  $J$ ,  $B_p$ ,  $PH_{SL}^{S,L}$ ,  $PH_{INT}^{S,L}$ ,  $PW_{SL}^{S,L}$ ,  $PW_{INT}^{S,L}$ ,  $SL_{TH}$ . And it is designed to simultaneously fit all the saccades from a single subject.

To illustrate per-subject Bahill's OPMM, we will use  $K$ ,  $K_2$  and  $K_3$  as defined by:

$$K = \frac{K_{SE}}{K_{SE} + K_{LT}}, \quad K_2 = \frac{K_{SE}^2}{K_{SE} + K_{LT}}, \quad K_3 = -(2 \cdot K_{SE} + K_p).$$

Using the previous notations the system state transition form is:

$$\begin{bmatrix} x_1^{t+1} \\ x_2^{t+1} \\ x_3^{t+1} \\ x_4^{t+1} \\ x_5^{t+1} \\ x_6^{t+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \Delta t & 0 & 0 \\ \frac{\Delta t K_2}{B_{AG}} & 1 - \frac{\Delta t K_{SE}}{B_{AG}} & 0 & 0 & \frac{\Delta t K}{B_{AG}} & 0 \\ \frac{\Delta t K_2}{B_{ANT}} & 0 & 1 - \frac{\Delta t K_{SE}}{B_{ANT}} & 0 & 0 & -\frac{\Delta t K}{B_{ANT}} \\ \frac{\Delta t K_3}{J} & \frac{\Delta t K_{SE}}{J} & \frac{\Delta t K_{SE}}{J} & 1 - \frac{\Delta t B_p}{J} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 - \frac{\Delta t}{\tau_{AG}} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 - \frac{\Delta t}{\tau_{ANT}} \end{bmatrix} \cdot \begin{bmatrix} x_1^t \\ x_2^t \\ x_3^t \\ x_4^t \\ x_5^t \\ x_6^t \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \frac{\Delta t N_{AG}}{\tau_{AG}} \\ \frac{\Delta t N_{ANT}}{\tau_{ANT}} \end{bmatrix},$$

where  $\Delta t$  is the OPMM internal sampling clock [Komogortsev and Khan 2008],  $N_{AG}$  and  $N_{ANT}$  are the neuronal control signals of the agonist and antagonist muscles, and  $\tau_{AG}$  and  $\tau_{ANT}$  are functions that define the low pass filtering process. These latter 2 parameters are defined by the activation and deactivation time constants.

## 2.5 Enderle's Model

In 1995 Enderle and co-authors described a key difference between the Enderle model and Bahill's OPMM:

The only structural difference between this model and the previous oculomotor muscle model is the addition of viscous element  $B_2$  and the removal of passive elasticity  $K_{pe}$ . As will be described, the viscous element  $B_2$  is vitally important to describe the nonlinear force-velocity characteristics of the muscle, and the elastic element  $K_{pe}$  is unnecessary (page 121, [Enderle 2010]).

After all changes, the order of the model was reduced from 6th to 4th order. In 2009, Zhou and Enderle [Enderle and Zhou 2010] reconsidered the previous solution and presented an updated version of OPMM where they removed one of the Voigt elements added in 2010 [Enderle 2010]. The Voigt elements pertain to the visco-elasticity of an extraocular muscle. The presence of this particular Voigt element did not affect the simulation results. After the removal of this element, the order of the model was reduced from 4th to 3rd.

While there are some parameters common to both the Bahill and Enderle per-saccade models, each model also has its unique set of parameters. The parameters for the per-saccade Enderle model that were not mentioned before are:

- $B_1, B_2$  – viscous elements in the muscle model. They are measured in  $\frac{gm \cdot tension - sec}{o}$ .
- $B_b$  – the orbital viscoelastic element connected to the eyeball sphere. It is measured in  $\frac{gm \cdot tension - sec}{o}$ .

- $K_k$  – a constant determined by steady-state analysis of the model.
- $T_1^{AG}, T_3^{ANT}$  – the initial start time of the saccade for the agonist and antagonist muscles, measured in *msec*.
- $F_{g0}, F_{t0}$  – the initial firing rate of the agonist and antagonist muscles. They are expressed in Newtons (*N*).
- $F_{gs}, F_{ts}$  – the firing rate of the agonist and antagonist muscles at the end of the saccade, expressed in *N*.

Enderle’s OPMM [Enderle and Zhou 2010] can be presented in the state transition matrix form (1) as well. To present Enderle’s OPMM more concisely, we will use the following notations. Let  $B_{12}$ ,  $K_{st}$  and  $K_k$  be equal to:

$$B_{12} = B_1 + B_2, \quad K_{st} = K_{SE} + K_{LT}, \quad K_k = \frac{K_p}{r} \cdot 5.2087 \cdot 10^3,$$

where  $B_1$  and  $B_2$  are viscous elements,  $K_p$  – is the passive elasticity for muscles,  $r$  – is the radius of the eyeball.

In addition, in the interest of simplification, we will use the following notations from Zhou and Enderle’s book [Enderle and Zhou 2010]:

$$P_0 = \frac{K_{st}K_k + 2K_{LT}K_{SE}}{JB_{12}}, \quad P_1 = \frac{2(B_1K_{SE} + B_2K_{LT}) + B_{12}K_k + K_{st}B_b}{JB_{12}},$$

$$P_2 = \frac{JK_{st} + B_{12}B_b + 2B_1B_2}{JB_{12}}, \quad \sigma = \frac{5208.7}{JB_{12}},$$

where  $B_b$  – is the voigt element.

With these substitutions, Enderle’s OPMM [Enderle and Zhou 2010] for simulation of the horizontal saccades in state transition form is:

$$\begin{bmatrix} x_1^{t+1} \\ x_2^{t+1} \\ x_3^{t+1} \end{bmatrix} = \begin{bmatrix} 1 & \Delta t & 0 \\ 0 & 1 & \Delta t \\ -\Delta t P_0 & -\Delta t P_1 & 1 - \Delta t P_2 \end{bmatrix} \begin{bmatrix} x_1^t \\ x_2^t \\ x_3^t \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \Delta t \sigma (K_{SE}(F_{AG} - F_{ANT}) + B_2(\dot{F}_{AG} - \dot{F}_{ANT})) \end{bmatrix}$$

where  $F_{AG}$  and  $F_{ANT}$  are active state tensions for the agonist and antagonist muscles.

To obtain a per-subject version of Enderle’s model, we used the same strategy as in [Katrychuk and V. Komogortsev 2022] for the simulation of the neural control signal parameters, and modeled both *PH* and *PW* as functions of saccade amplitude. It’s important to note that one of the primary differences between the per-subject versions of the Bahill and Enderle models is the way we simulate the *PH* and *PW* functions for them. In the per-subject Bahill model, neural pulse parameters were simulated separately for large and small saccades. For the per-subject Enderle model, this separation does not apply. However, in the per-subject Enderle model, neural pulse parameters were simulated separately for agonist and antagonist muscles:  $PH_{sl}^{AG,ANT}$ ,  $PH_{int}^{AG,ANT}$ ,  $PW_{sl}^{AG,ANT}$ ,  $PW_{int}^{AG,ANT}$ .

To speed up Enderle model simulation time, we simplified the neural control signal. In Enderle’s original work, he simulated the neural control signal using two peak magnitudes,  $F_{p1}$  and  $F_{p2}$ . We simplified it by using only the first peak magnitude,  $F_{p1}$ , and removing all components related to the second peak magnitude,  $F_{p2}$ . By applying all the aforementioned changes, we express the neural pulse height ( $PH^{AG}$ ) and neural pulse ( $N^{AG}$ ) for agonist muscle in the per-subject Enderle model as follows:

$$PH^{AG} = PH_{int}^{AG} + PH_{sl}^{AG} \cdot |amp|, \quad T_2 = T_1 + PW^{AG},$$

$$N^{AG} = F_{g0} \cdot (u(t) - u(t - T_1)) + PH^{AG} \cdot (u(t - T_1) - u(t - T_2)) + F_{gs} \cdot u(t - T_2),$$

where  $u(\cdot)$  is the Heaviside step function. We applied the same changes to the antagonist muscle functions as well. This significantly reduced computation time and improved simulation accuracy. As a result, this model in total has

25 parameters:  $K_{SE}, K_{LT}, K_k, B_b, B_1, B_2, J, T_1^{AG}, T_3^{ANT}, F_{g0}, F_{gs}, F_{t0}, F_{ts}, \tau_{act}^{AG,ANT}, \tau_{deact}^{AG,ANT}, PH_{sl}^{AG,ANT}, PH_{int}^{AG,ANT}, PW_{sl}^{AG,ANT}, PW_{int}^{AG,ANT}$ .

## 2.6 Comparing the Performance of Different OPMM versions

**2.6.1 Per-Saccade Models.** Per-saccade models test model performance by comparing a simulated saccade trajectory with an actual saccade trajectory. For each fitted saccade, an error is computed as the sum of the absolute value of each per-sample difference between the predicted and recorded saccade trajectory. This sum is normalized by dividing by the saccade duration, yielding an error per sample (in our case, error per msec). Different per-saccade models may perform better for some saccade amplitudes than others. Therefore, it may be important to do several comparisons for sets of saccades of particular amplitudes. This is true, in part, because different models used different saccade amplitudes during model development.

**2.6.2 Per-Subject Models.** The total number of saccades per subject was divided into a series of folds in a cross-validation manner. Every fold is consecutively chosen as a test set, while the rest represents a training set. Each per-subject model results in 5 simulation error estimates (1 per fold, 5 folds) for each subject. Error estimates were compared between each model, and within each model, training versus testing sets were also compared. We visualized the differences in the distributions of these errors by employing violin plots, which are a hybrid of a box plot and a kernel density plot. Differences in per-subject error distribution were tested with a Friedman test. A statistically significant  $\chi^2$  was followed by a multiple comparison procedure using a Tukey HSD test at  $\alpha < 0.05$ .

## 2.7 Test-Retest Reliability of our parameter estimates

We have parameter estimates for each round of data (Round 1 vs. Round 2) and for each session (sessions 1 and 2). We employed both the intraclass correlation coefficient (ICC) and the Kendall Coefficient of Concordance (KCC) to estimate the test-retest reliability. We can think of these numbers as indices of temporal persistence. We compute these statistics on a matrix with 4 columns (Round 1 Session 1; Round 1 Session 2; Round 2 Session 1 and Round 2 Session 2). According to Cicchetti [Cicchetti 1994], ICCs above 0.75 are considered excellent, between 0.6 and 0.74 are good, between 0.4 and 0.59 are fair, and below 0.4 the ICCs are poor. According to Landis and Koch [Landis JRKoch 1977], KCCs values above 0.81 are considered almost perfect, between 0.61 and 0.80 are substantial, between 0.41 and 0.60 are moderate, between 0.21 and 0.40 are fair, between 0.00 and 0.20 are slight and below 0.0 the KCCs are poor.

## 2.8 Comparing Fitted Parameters to Default Values for Saccade-Specific OPMMs

For each subject, we have two rounds, two sessions (i.e., four occasions), and five folds. This means that we have 20 parameter estimates per parameter per subject. We then statistically compared these 20 parameter estimates to a default parameter value obtained from prior saccade-specific models [Bahill 1980; Enderle 2010]. For this, we employed a one-sample t-test. We then counted the number of subjects (of a total of 136) which were significantly different from the default value ( $p < 0.05$ , 2-tailed). It is important to note that there is a 60% overlap between any two training folds, and therefore folds are not independent. This is a formal violation of the assumptions of a one-sample t-test. Nonetheless, we still thought that with 20 folds, our results would be interesting, informative, and useful to report.

## 2.9 Optimization

We used the same optimization strategy as was used for the baseline per-subject model described in [Katrychuk and V. Komogortsev 2022]. The chosen optimization method was the Nelder-Mead simplex method [Nelder and Mead 1965]. For both the Bahill and Enderle per-subject models, we divided the data from each subject into 5 non-overlapping test folds, conducting evaluations using K-Fold cross-validation, as detailed in [Katrychuk and V. Komogortsev 2022]. All calculations were conducted separately for each subject, round, and session. In the training set for each fold, the median number of saccades was 204, while in each test fold, it was 51. The average time to optimize 18 parameters for the Bahill model was 11 minutes and 40 seconds, whereas for the Enderle model with 25 parameters, it was 35 minutes and 20 seconds.

## 3 RESULTS

### 3.1 Optimized Parameter Estimates and Comparison to Default Parameters from Per-Saccade OPMMs

*3.1.1 Percent Change.* In Table 1 we present the default parameter values for the Bahill model as well as our optimized values. The range of change from default to optimized was from -30.5% for  $B_{AG}$  to 27.68% for  $C_{FIX}$ . The median percent change was  $\approx 1\%$  and the median of the absolute value change was 5.2%. In Table 2 we present the same information for the Enderle model. The range of change from default to optimized was from -21.82% for  $PW_{INT}^{ANT}$  to 17.42% for  $J$ . The median percent change was  $\approx 0.65\%$  and the median of the absolute value change was 2.5%.

*3.1.2 Number of Subjects Different from Default.* In Table 1 and Table 2 we also present counts (and percentages) of the number of subjects whose parameter estimates were significantly different ( $p < 0.05$ ) from the per-saccade default value for each model-type and parameter. For the Bahill model, subjects had parameter estimates that were significantly different from defaults for a median of 83.1% of subjects. For the Enderle model, subjects had parameter estimates that were significantly different from defaults for a median of 94.1% of subjects. For the Bahill model, there were three parameters (all related to the neural pulse height or width for small saccades) that were significantly different from the default less than 50% of the time. For the Enderle model, there were two parameters (both related to the antagonist pulse width) that were significantly different from the default less than 50% of the time.

### 3.2 Comparison of the Performance of the Enderle OPMM with Bahill OPMM

*3.2.1 Simulation Error.* To visualize the simulation accuracy achieved by each model, we present the violin and the box-and-whisker diagrams of the mean training and test errors across 5 folds. In Fig. 1 you can find the distribution plots for both Enderle and Bahill models.

The Friedman test produced at  $\chi^2 = 6,516, df = 3, p = 0.0$ . Multiple comparison testing found that, for both models, the train and test data were not statistically significantly different. This supports the claim that both models are generalizable [Katrychuk and V. Komogortsev 2022]. All other comparisons were statistically significant with p-values effectively at 0.0.

*3.2.2 Change in simulated trajectory accuracy over saccade duration.* To provide additional insight into the difference in performance between the two models, we plot the statistics of change in simulated trajectory accuracy over saccade duration. This information is computed in the following manner. First, all saccades are grouped into 4 bins according to their amplitude in degrees: (0, 5]; (5, 10]; (10, 20]; (20,  $+\infty$ ). Second, within each bin, all saccade sample timestamps are normalized to be in the range of [0..1]. Then, every saccade is resampled via one-dimensional piece-wise linear

Parameter	Number Significant	Percent	Default Value	Optimized Value	Percent Change
$K_{SE}$	117	86.03	2.5	2.726	9.04
$K_{LT}$	107	78.68	1.2	1.212	1.02
$B_p$	136	100	0.06	0.046	-23.98
$B_{AG}$	131	96.32	0.046	0.032	-30.52
$B_{ANT}$	135	99.26	0.022	0.016	-27.97
$C_{AG}$	98	72.06	0.8	0.823	2.87
$C_{ANT}$	127	93.38	0.5	0.526	5.20
$J$	132	97.06	0.000046	0.000039	-14.63
$C_{FIX}$	136	100	14	17.876	27.68
$PH_{INT}^S$	47	34.56	10	10.016	0.16
$PH_{INT}^L$	112	82.35	10	10.318	3.18
$PH_{SL}^S$	56	41.18	2	1.996	-0.20
$PH_{SL}^L$	128	94.12	2	2.213	10.65
$PW_{INT}^S$	49	36.03	10	9.949	-0.51
$PW_{INT}^L$	130	95.59	10	10.797	7.97
$PW_{SL}^S$	53	38.97	1	0.993	-0.71
$PW_{SL}^L$	136	100	1	1.132	13.22
$SL_{TH}$	79	58.09	2	2.009	0.47

Table 1. Percent of subjects with estimates statistically different from defaults for the Bahill model.

interpolation at 51 sample points equally spaced on a grid within this normalized range. Finally, the per-sample error between a simulated and a corresponding recorded saccade trajectory is computed at the resampled grid points.

In Fig. 2, you can find the resulting plots summarizing Round 1, Session 2 data for both Enderle and Bahill models. This data shows that the baseline Bahill per-subject model provided better simulation accuracy than the recent Enderle per-subject model. Although we only show the simulation error plots for the test set, the plots for the training set looked identical, as would be expected. The simulation error plot for the largest saccades (Fig. 2(D)) is distinctive. To offer a more comprehensive comparative analysis of the performance of the two per-subject OPMM models, we include in the Appendix section plots depicting original saccade trajectories alongside simulated saccades from both OPMMs.

### 3.3 Test-retest Reliability

Recall that the Enderle model has 25 parameters and the Bahill model has 18 parameters. ICC and KCC values are estimates of the test-retest reliability of each parameter. The ICCs and KCCs for the per-subject Bahill OPMM are listed in Table 3a (ICCs and KCCs were correlated at 0.998). Two of the ICCs were “excellent”, four were “good”, four were “fair” and eight were “poor”. Of the eight poor ICCs, six were related to either neural pulse height or width. The four with the lowest ICCs were related to the pulse height and width of small saccades specifically.

The ICCs and KCCs for the per-subject Enderle OPMM are listed in Table 3b (ICCs and KCCs were correlated at 0.993). Only one ICC value was “excellent”, three were “good”, four were “fair” and seventeen were “poor”. We compared the set of Bahill ICCs (median ICC = 0.44) to the set of Enderle ICCs (median = 0.36) with a Mann-Whitney U-test and

Parameter	Number Significant	Percent	Default Value	Optimized Value	Percent Change
$K_{SE}$	136	100	124.958	105.516	-15.56
$K_{LT}$	123	90.44	60.687	62.082	2.30
$K_k$	78	57.35	16.36	16.264	-0.59
$B_b$	104	76.47	0.327	0.332	1.60
$B_1$	117	86.03	5.722	5.622	-1.75
$B_2$	118	86.76	0.502	0.526	4.81
$J$	122	89.71	0.002	0.002	17.42
$T_1^{AG}$	96	70.59	0.004	0.004	10.86
$PW_{INT}^{AG}$	136	100	0.024	0.025	3.31
$PW_{SL}^{AG}$	136	100	-0.0001120	-0.0001122	0.18
$T_3^{ANT}$	136	100	0.003	0.003	-2.45
$PW_{INT}^{ANT}$	136	100	0.002	0.002	-21.82
$PW_{SL}^{ANT}$	136	100	-5.799	-5.817	0.31
$F_{g0}$	136	100	0.4	0.415	3.70
$PH_{INT}^{AG}$	136	100	1.305	1.303	-0.14
$PH_{SL}^{AG}$	136	100	-0.009	-0.009	-3.02
$F_{gs}$	136	100	0.4	0.412	2.98
$F_{t0}$	108	79.41	0.4	0.394	-1.46
$PH_{INT}^{ANT}$	114	83.82	0.391	0.395	0.97
$PH_{SL}^{ANT}$	136	100	-0.001000	-0.001014	1.36
$F_{ts}$	136	100	0.4	0.369	-7.64
$\tau_{act}^{AG}$	136	100	0.0110	0.0111	1.14
$\tau_{deact}^{AG}$	136	100	0.0050	0.0049	-1.07
$\tau_{act}^{ANT}$	136	100	0.0090	0.0094	4.20
$\tau_{deact}^{ANT}$	136	100	0.0050	0.0048	-3.19

Table 2. Percent of subjects with estimates statistically different from defaults for the Enderle model.

found that the two sets were not statistically significantly different. Of the parameters with the bottom 10 ICCs (all “poor”), six were related to either pulse height or pulse width.

#### 4 DISCUSSION

In this report, we demonstrate how to apply a per-subject optimization approach for an Enderle OPMM and assess the test-retest reliability of parameter estimates in detail. The per-subject approach, previously developed to the Bahill OPMM [Katrychuk and V. Komogortsev 2022], was applied to the Enderle model. This involved adapting the pulse-width and pulse-height functions to the model. In this report, we found that the Bahill per-subject OPMM outperforms the per-subject Enderle OPMM in terms of overall simulation error and simulation error as a function of saccade duration. This suggests that there may be room for improvement to enhance the accuracy of the per-subject Enderle model.

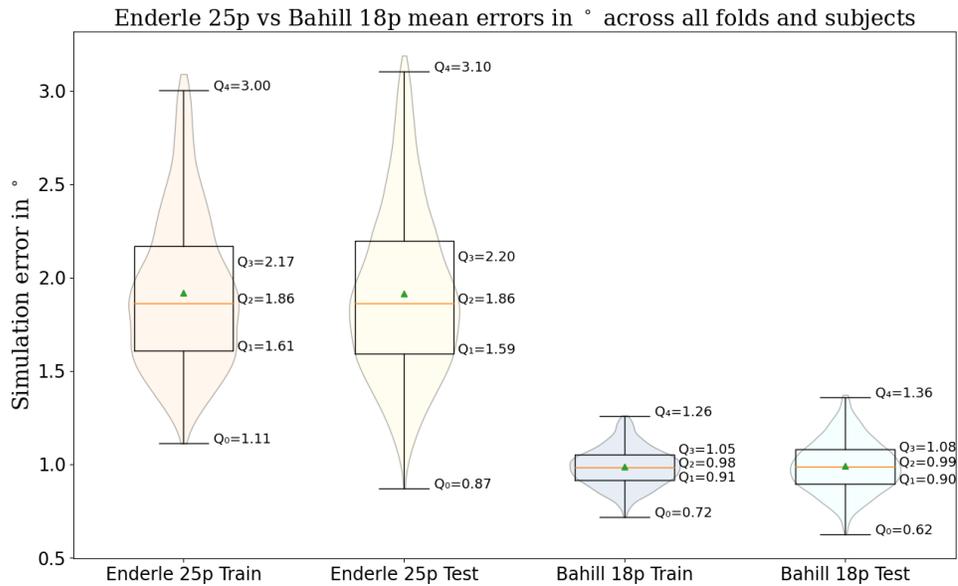


Fig. 1. Box-and-whisker plots of the mean train and test errors for both models with the outliers removed. The boxplots are presented in a standard way. The median is the red horizontal line. The mean is at the green dot. The bottom of the box is at the 25th percentile and the top of the box is at the 75th percentile. Outliers are determined using Tukey hinges. The actual drawn whisker is to the first and last point in the data that are not outliers. Outliers are not shown.

In the current study, we introduce the "test-retest" reliability as a metric for evaluating model performance. While model accuracy is a central focus in OPMM research, another crucial aspect that holds interest for many researchers is the temporal persistence (i.e., reliability) of model parameters.

#### 4.1 Reliability of OPMM Parameters

A statistical test comparing the set of ICCs for the Bahill parameters versus the Enderle parameters was not statistically significant. However, given the very small number of parameters, this test is a very weak statistical comparison. With 18 parameters for the Bahill model and 25 parameters for the Enderle model, this test has a statistical power of 0.35 to find a moderate effect size (Cohen's  $d = 0.5$ ). Typically, a 0.8 power is considered acceptable.

After analysis of the ICC values for OPMMs, we can draw a couple of interesting conclusions about both models. In Table 3a, we see that the four parameters with the lowest ICCs were related to the pulse height (PH) and width (PW) of small saccades. In the Enderle model, the PH and PW are estimated for both agonist and antagonist muscles (Table 3b). These parameters were also among the least reliable. Since the relationship between saccade amplitude and PH or PW is a central feature of both per-subject models, one possibility is that these specific per-subject models may need improvement in the estimation of PW and PH.

For the Bahill model, one solution might be to use the same neural-pulse relationship for both small and large saccades. The same suggestion can be made about the separate neural-pulse simulation for agonist and antagonist

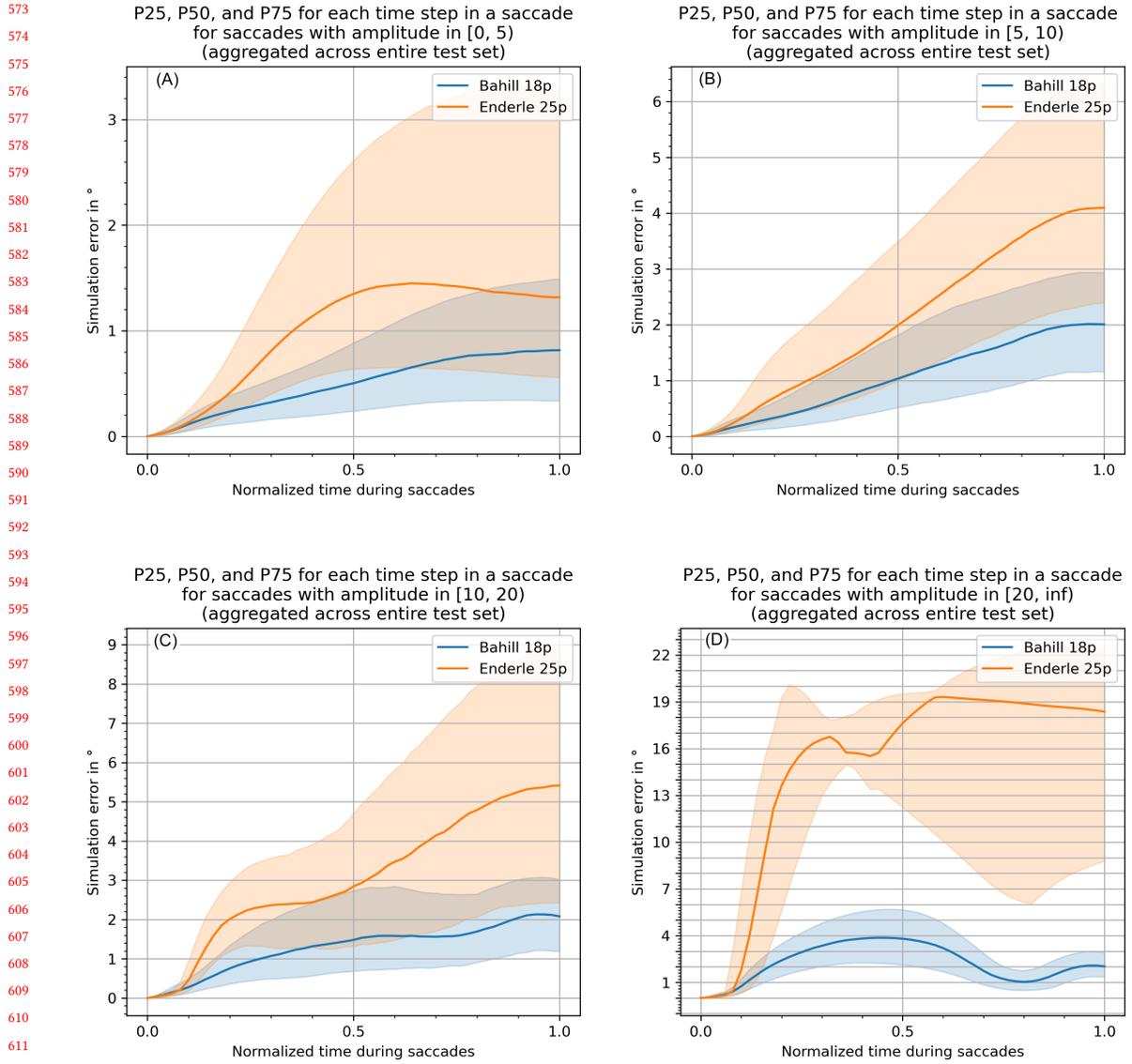


Fig. 2. Plots that summarize the change in saccade trajectory simulation accuracy over the saccade duration time (Round 1, Session 2 data) for the test dataset. The data shown in blue is the simulation error for the Bahill model, and the data shown in orange is the simulation error for the Enderle model. The lower border for each model represents the 25 percentile. The upper border for each model represents the 75 percentile. The solid line is the median. Data in a subplot (A) is for saccades with amplitude in the  $[0, 5)$  range. (B) is the same as (A) but for data in the  $[5, 10)$  range, (C) — for the data in the  $[10, 15)$  range, and (D) — for the data in the  $[20, \infty)$  range.

muscles in the per-subject Enderle model. The proposed model simplifications may enhance the accuracy and reliability of the OPMM parameters, while also reducing computation time for the modified models.

Table 3. ICC values for the per-subject OPMMs

Parameter	ICC	KCC
$B_{AG}$	0.87	0.9
$B_{ANT}$	0.822	0.858
$K_{SE}$	0.757	0.802
$B_p$	0.743	0.8
$J$	0.649	0.736
$PW_{INT}^L$	0.648	0.692
$\tilde{K}_{LT}$	0.598	0.712
$C_{FIX}$	0.522	0.64
$PH_{SL}^L$	0.47	0.616
$PH_{INT}^L$	0.419	0.551
$C_{AG}$	0.334	0.507
$PW_{SL}^L$	0.244	0.426
$SL_{TH}$	0.212	0.411
$C_{ANT}$	0.208	0.408
$PW_{INT}^S$	0.024	0.257
$PH_{SL}^S$	0.02	0.259
$PH_{INT}^S$	0.02	0.269
$PW_{SL}^S$	0.013	0.26

Parameter	ICC	KCC
$J$	0.8	0.868
$B_2$	0.668	0.766
$B_1$	0.651	0.689
$PH_{INT}^{AG}$	0.65	0.725
$F_{gs}$	0.559	0.596
$B_b$	0.45	0.6
$\tau_{act}^{ANT}$	0.441	0.546
$K_{LT}$	0.43	0.522
$\tau_{act}^{AG}$	0.399	0.534
$F_{ts}$	0.398	0.524
$PW_{INT}^{AG}$	0.384	0.514
$\tau_{deact}^{ANT}$	0.379	0.519
$K_{SE}$	0.357	0.522
$F_{t0}$	0.351	0.497
$T_3^{ANT}$	0.223	0.418
$F_{g0}$	0.175	0.379
$PH_{INT}^{ANT}$	0.132	0.352
$T_1^{AG}$	0.055	0.304
$PH_{SL}^{ANT}$	0.05	0.294
$\tau_{deact}^{AG}$	0.029	0.27
$K_K$	0.026	0.257
$PW_{SL}^{ANT}$	0.013	0.258
$PW_{SL}^{AG}$	0.007	0.25
$PH_{SL}^{AG}$	0	0.249
$PW_{INT}^{ANT}$	0	0.246

(a) ICC values for the per-subject Bahill 18 parameters model. (b) ICC values for the per-subject Enderle 25 parameters model.

It is logical and reasonable to expect that certain parameters should be more reliable than others. For example, physiological parameters related to muscle elasticity and maximum contractile force ( $K_{SE}$  and  $K_{LT}$ ) might be highly reliable as they refer to observable physical entities. This is also true of the inertia of the globe,  $J$ . On the other hand, parameters related to the neural pulse characteristics are more difficult to observe and therefore might be less reliable. However, this distinction is only partially born out by our results.

For example, for the Bahill model  $K_{SE}$  is one of the three most reliable parameters (0.757, just below the “excellent” level reliability), however for the Enderle model the reliability of the  $K_{SE}$  is very poor (0.357). For both models  $K_{LT}$  has only fair reliability.  $J$  is the most reliable parameter for the Enderle model (0.803, excellent reliability), however for the Bahill model  $J$  (“eyeball inertia”) is only within the range of good reliability (0.649).

677 Additionally, the most unreliable parts of the model may be built on incorrect or oversimplified assumptions, as they  
678 fail to explain patterns found in real data. Our least reliable parameters correspond to the neural pulse, which is indeed  
679 modeled assuming a simplistic linear relationship to saccade amplitude. Trying to fit a linear function to data exhibiting  
680 non-linear relationships may have created a loss surface with many local minima of a similar error, any of which can be  
681 a point of convergence but neither of which affords a good real data fit. The future work should be concerned with  
682 reassessing the validity of linear neural pulse models.

684 Another useful evaluation procedure we propose in this study involves examining the differences between the  
685 optimized parameters for the OPMM models and the parameter values initially presented in the original papers where  
686 these models were first introduced. As reported above, the parameter estimates for the per-subject models differed  
687 significantly from the default values of the per-saccade models in most cases. Perhaps, in such cases, it might be useful  
688 to compare the accuracy of models where some of these parameters are replaced by default values with the current  
689 models. In future research, we plan to compare various additional strategies for dealing with the issue of unreliable  
690 parameters.

692 We would also like to highlight several issues on the estimation of OPMM parameters from recorded gaze signals.  
693 The most important issue is an inability to directly assess the parameter estimation error due to the lack of appropriate  
694 ground truth. Certain OPMM parameters were designed to be simplified abstract representations of phenomena within  
695 the OP. Those are mostly related to the neural control signal, and their ground-truth values do not exist. For those  
696 parameters that do explain actual anatomical properties, such as eyeball inertia ( $J$ ), complex studies that involve surgery  
697 are required to recover their true values. Therefore, we think that the only feasible way is for the quality of the estimated  
698 parameters to be assessed indirectly, through the gaze signal fit or their reliability, as it was done in our study.

700 The next major challenge in parameter estimation is related to the choice of the optimization method. Our utilization  
701 of Nelder-Mead is carried over from the prior work on OPMMs. It is a very convenient method, as it can be applied  
702 without any assumptions on the underlying function to be optimized. However, we are not aware of any studies that  
703 evaluate other methods against it in the context of OPMM parameter optimization. Exploring other alternatives, such as  
704 Bayesian optimization, genetic algorithms, or gradient-based methods, especially in regard to their resulting parameter  
705 reliability, should be done in future work. Note that in such an experiment it is still essential to consider the quality of  
706 data fit, as it is trivial to design a method with perfect reliability that simply always returns the default parameter list.  
707 Additionally, we can circumvent the lack of ground-truth parameter values by fitting the synthetic saccade trajectories  
708 produced with the OPMM itself in a controlled experiment with known parameters. This avoids the aforementioned  
709 complexity of evaluating parameters purely on the data fit they provide and allows to benchmark different optimization  
710 methods based on the quality of their ground-truth parameters retrieval, which is unfeasible for real data.

## 716 4.2 Comparative Analysis of Bahill and Enderle Per-subject Models

717 For this, we focus on specific instances where one model may outperform the other. In the Appendix, we present  
718 12 figures, each depicting actual trajectories and simulated trajectories generated by the two OPMM models. Two  
719 generalizations emerge from an examination of these trajectories. In six of the twelve cases, the Enderle model's  
720 saccades move in the wrong direction for some period before changing direction. The peak of the "wrong" direction  
721 movement occurs between 9 and 12 milliseconds. In five of the twelve saccades, the Bahill model trajectory reaches  
722 the final endpoint sooner than the real or the Enderle trajectory. An idea for a future study would be to fit individual  
723 per-saccade-based models to these trajectories and try to determine which parameters predict these characteristics.  
724 This might help elucidate the reasons behind these differences.

### 4.3 Future Work

The primary objectives of previous OPMM studies were to improve the ability to accurately recreate horizontal saccade trajectories. However, to improve the usability of physiologically inspired models, two main issues require attention in future research. One is the enhancement of the existing level of reliability in OPMM parameters. Achieving an adequate level of reliability is crucial for OPMM models to become valuable additions to biometrics or eye-movement prediction studies. Without this improvement, unreliable parameters increase the risk of false positives or false negatives, compromising the security and effectiveness of the system they are employed in.

One approach to the low reliability would be to try models where certain constraints on individual parameters were employed or tightened. In the extreme case, rather than fitting an unreliable parameter, this parameter could be set to the default value. The analysis of trajectory errors would allow us to determine if such a procedure improves the predictive power of the model. We plan to perform such an analysis in future studies.

The other issue is simulating the 2D saccadic trajectories using OPMMs. It is not a trivial task due to the nature of saccades and how desynchronized the vertical and horizontal components can be across saccadic trajectories [Bahill and Stark 1977]. Previously, OPMMs were applied to simulate 2D movements by using two separate 1D models for vertical and horizontal channels [Komogortsev et al. 2012b, 2013]. However, we believe that it is necessary to adapt the existing horizontal saccade model to accommodate vertical saccades for more accurate results.

To sum up, we think that addressing the low reliability of current OPMM parameter estimates highlighted in our study should be a focus of future OPMM developments. Whether it can be achieved by using a simplified model with a reduced number of parameters (or tightened constraints on specific parameters), an alternative optimization method, or an extra loss function component not based on the gaze signal fit is an open research question.

## 5 CONCLUSION

In this study, we successfully applied the per-subject optimization approach to the Enderle model and established a new metric for evaluating OPMMs. We found that the current implementation of the baseline per-subject Bahill model was better than the baseline per-subject Enderle model in terms of accuracy. For each model, we have provided the test-retest reliability (ICC) for each parameter. Some of the unreliable OPMM parameters were associated with neural pulse characteristics, while others were related to different physiological properties of OP.

In summary, our study provides detailed guidelines on how to work with OPMM models, adapt the new optimization procedure to different OPMMs, and assess their performance. This assessment goes beyond mere accuracy, as it takes into account each parameter's value and its temporal persistence. The results for the current versions of per-subject models revealed significant differences between the optimized parameter estimates and the default parameter values reported for the per-saccade models. This implies that there is room for further improvements in the optimization procedure in future research.

More generally, OPMM-inspired models have been applied across various domains, including gaze-contingent techniques [Guenter et al. 2012; Patney et al. 2016], cybersecurity [Komogortsev 2017], and health assessment [Komogortsev and Holland 2014; Wadehn 2019]. However, numerous unresolved questions remain, offering scientists an opportunity to make meaningful contributions to the advancement of OPMMs that can be practically useful in eye tracking driven applications.

## REFERENCES

- 781  
782 A Terry Bahill. 1980. Development, validation and sensitivity analyses of human eye movement models. *CRC Critical Reviews in Bioengineering* 4, 4  
783 (1980), 311–355.
- 784 A Terry Bahill and Lawrence Stark. 1975. Neurological control of horizontal and vertical components of oblique saccadic eye movements. *Mathematical*  
785 *Biosciences* 27, 3-4 (12 1975), 287–298. [https://doi.org/10.1016/0025-5564\(75\)90107-8](https://doi.org/10.1016/0025-5564(75)90107-8)
- 786 A Terry Bahill and Lawrence Stark. 1977. Oblique saccadic eye movements: Independence of horizontal and vertical channels. *Archives of Ophthalmology*  
787 95, 7 (08 1977), 1258–1261.
- 788 Domenic V Cicchetti. 1994. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology.  
789 *Psychological assessment* 6, 4 (12 1994), 284. <https://doi.org/10.1037/1040-3590.6.4.284>
- 790 Gerald Cook and Lawrence Stark. 1968. The human eye-movement mechanism: experiments, modeling, and model testing. *Archives of ophthalmology* 79,  
791 4 (05 1968), 428–436.
- 792 John Enderle. 2010. *Models of Horizontal Eye Movements, Part I: Early Models of Saccades and Smooth Pursuit*. Vol. 5. Morgan & Claypool Publishers.  
793 <https://doi.org/10.2200/S00263ED1V01Y201003BME034>
- 794 John Enderle and Wei Zhou. 2010. *Models of Horizontal Eye Movements, Part II: A 3rd Order Linear Saccade Model*. Vol. 5. Morgan & Claypool Publishers.  
795 <https://doi.org/10.2200/S00264ED1V01Y201003BME035>
- 796 WO Fenn and BS Marsh. 1935. Muscular force at different speeds of shortening. *The Journal of Physiology* 85, 3 (12 1935), 277–297. <https://doi.org/10.1113/jphysiol.1935.sp003318>
- 797 Lee Friedman, Mark S Nixon, and Oleg V Komogortsev. 2017. Method to assess the temporal persistence of potential biometric features: Application to  
798 oculomotor, gait, face and brain structure databases. *PLoS one* 12, 6 (2017), e0178501.
- 799 Lee Friedman, Ioannis Rigas, Evgeny Abdulin, and Oleg V Komogortsev. 2018. A novel evaluation of two related and two independent algorithms for eye  
800 movement classification during reading. *Behavior Research Methods* 50 (05 2018), 1374–1397. <https://doi.org/10.3758/s13428-018-1050-7>
- 801 Alireza Ghahari and John Enderle. 2014. *Models of Horizontal Eye Movements: Part 3, A Neuron and Muscle Based Linear Saccade Model*. Vol. 9. Morgan &  
802 Claypool Publishers. <https://doi.org/10.2200/S00592ED1V01Y201408BME053>
- 803 Alireza Ghahari and John Enderle. 2015. *Models of Horizontal Eye Movements: Part 4, A Multiscale Neuron and Muscle Fiber-Based Linear Saccade Model*.  
804 Vol. 9. Morgan & Claypool Publishers. <https://doi.org/10.2200/S00627ED1V01Y201501BME055>
- 805 Henry Griffith, Dillon Lohr, Evgeny Abdulin, and Oleg Komogortsev. 2021. GazeBase, a large-scale, multi-stimulus, longitudinal eye movement dataset.  
806 *Scientific Data* 8, 1 (07 2021), 184. <https://doi.org/10.1038/s41597-021-00959-y>
- 807 Brian Guenter, Mark Finch, Steven Drucker, Desney Tan, and John Snyder. 2012. Foveated 3D graphics. *ACM transactions on Graphics (TOG)* 31, 6 (2012),  
808 1–10.
- 809 Robert Hoerantner, Thomas Kaltofen, Siegfried Priglinger, Christian-Michael Fock, Michael Buchberger, and Thomas Haslwanter. 2007. Model-Based  
810 Improvements in the Treatment of Patients with Strabismus and Axial High Myopia. *Investigative ophthalmology & visual science* 48, 3 (04 2007),  
811 1133–1138. <https://doi.org/10.1167/iovs.06-0769>
- 812 Frederich K Hsu, A Terry Bahill, and Lawrence Stark. 1976. Parametric sensitivity analysis of a homeomorphic model for saccadic and vergence eye  
813 movements. *Computer Programs in Biomedicine* 6, 2 (08 1976), 108–116. [https://doi.org/10.1016/0010-468X\(76\)90032-5](https://doi.org/10.1016/0010-468X(76)90032-5)
- 814 Alex Karpov, Jacob Liberman, Dillon J. Lohr, and Oleg Komogortsev. 2020. Parallel Oculomotor Plant Mathematical Model for Large Scale Eye Movement  
815 Simulation. *CoRR abs/2007.09884* (2020). [arXiv:2007.09884](https://arxiv.org/abs/2007.09884) <https://arxiv.org/abs/2007.09884>
- 816 Dmytro Katrychuk and Oleg V. Komogortsev. 2022. A study on the generalizability of Oculomotor Plant Mathematical Model. In *2022 Symposium on Eye*  
817 *Tracking Research and Applications* (Seattle, WA, USA) (ETRA '22). Association for Computing Machinery, New York, NY, USA, Article 68, 7 pages.  
818 <https://doi.org/10.1145/3517031.3532523>
- 819 Oleg Komogortsev, Cecilia Aragon, Alexey Karpov, and Larry Price. 2012a. Biometric Authentication via Oculomotor Plant Characteristics. *IEEE/IARP*  
820 *International Conference on Biometrics (ICB)* (04 2012), 413–420. <https://doi.org/10.1109/ICB.2012.6199786>
- 821 Oleg Komogortsev and Corey Holland. 2014. The Application of Eye Movement Biometrics in the Automated Detection of Mild Traumatic Brain Injury.  
822 In *CHI '14 Extended Abstracts on Human Factors in Computing Systems* (Toronto, Ontario, Canada) (CHI EA '14). Association for Computing Machinery,  
823 New York, NY, USA, 1711–1716. <https://doi.org/10.1145/2559206.2581150>
- 824 Oleg Komogortsev, Corey Holland, and Sampath Jayarathna. 2012b. Two-Dimensional Linear Homeomorphic Oculomotor Plant Mathematical Model. (07  
825 2012). <https://doi.org/10.13140/RG.2.1.2711.8486>
- 826 Oleg Komogortsev, Corey Holland, Sampath Jayarathna, and Alex Karpov. 2013. 2D Linear Oculomotor Plant Mathematical Model: Verification and  
827 Biometric Applications. *ACM Transactions on Applied Perception (TAP)* 10, 4 (10 2013), 1–18. <https://doi.org/10.1145/2536764.2536774>
- 828 Oleg Komogortsev, Sampath Jayarathna, Cecilia Aragon, and Mechehouh Mahmoud. 2010. Biometric Identification via an Oculomotor Plant Mathematical  
829 Model. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* (Austin, Texas). *Technical Reports-Computer Science*, 57–60.  
830 <https://doi.org/10.1145/1743666.1743679>
- 831 Oleg V Komogortsev. 2017. Person identification using ocular biometrics with liveness detection. US Patent 9,811,730.
- 832 Oleg V. Komogortsev and Javed I. Khan. 2008. Eye Movement Prediction by Kalman Filter with Integrated Linear Horizontal Oculomotor Plant Mechanical  
833 Model. In *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications* (Savannah, Georgia) (ETRA '08). Association for Computing  
834 Machinery, New York, NY, USA, 229–236. <https://doi.org/10.1145/1344471.1344525>

- 833 GG Landis JRKoch. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33, 1 (04 1977), 159–74. <https://doi.org/10.2307/2529310>
- 834 A Levin and J Wyman. 1927. The viscous elastic properties of muscle. *Proceedings of the Royal Society of London. Series B, containing papers of a biological*
- 835 *character* 101, 709 (1927), 218–243.
- 836 John A Nelder and Roger Mead. 1965. A simplex method for function minimization. *The computer journal* 7, 4 (1965), 308–313.
- 837 Anjul Patney, Marco Salvi, JooHwan Kim, Anton Kaplanyan, Chris Wyman, Nir Benty, David Luebke, and Aaron Lefohn. 2016. Towards foveated rendering
- 838 for gaze-tracked virtual reality. *ACM Transactions on Graphics (TOG)* 35, 6 (2016), 1–12.
- 839 DA Robinson. 1964. The mechanics of human saccadic eye movement. *The Journal of physiology* 174, 2 (12 1964), 245. <https://doi.org/10.1113/jphysiol.1964.sp007485>
- 840 D. A. Robinson, D. M. O’Meara, A. B. Scott, and C. C. Collins. 1969. Mechanical components of human eye movements. *Journal of applied physiology* 26, 5
- 841 (1969), 548–553. <https://doi.org/10.1152/jappl.1969.26.5.548> PMID: 5781605.
- 842 Federico Wadehn. 2019. *State space methods with applications in biomedical signal processing*. Vol. 31. ETH Zurich.
- 843 Federico Wadehn, David J Mack, Thilo Weber, and Hans-Andrea Loeliger. 2018. Estimation of neural inputs and detection of saccades and smooth
- 844 pursuit eye movements by sparse Bayesian learning. In 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology
- 845 Society (EMBC). *Conference proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in*
- 846 *Medicine and Biology Society. Conference 2018*, 2619–2622. <https://doi.org/10.1109/EMBC.2018.8512758>
- 847 Federico Wadehn, Thilo Weber, David J Mack, Thomas Heldt, and Hans-Andrea Loeliger. 2019. Model-based separation, detection, and classification of
- 848 eye movements. *IEEE Transactions on Biomedical Engineering* 67, 2 (05 2019), 588–600. <https://doi.org/10.1109/TBME.2019.2918986>
- 849 Gerald Westheimer. 1954. Mechanism of saccadic eye movements. *AMA Archives of Ophthalmology* 52, 5 (1954), 710–724.
- 850
- 851

## 852 A A COMPARISON OF THE SIMULATED SACCADIC TRAJECTORIES

853 We have decided to include several comparative graphs of real and predicted saccade trajectories in this Appendix.

854 These graphs will illustrate cases where one model outperforms the other in simulation accuracy, as well as cases where

855 their simulation accuracy is nearly identical. We randomly selected one subject from the available set and presented

856 instances from all 5 test folds.

857

858

859 Let’s examine examples (A) and (B). In Fig. (A), the trajectory predicted by the Enderle model begins a movement

860 away from the target, with a peak at 9 msec before changing direction towards the target. As a result, for most of the

861 trajectory, the Enderle saccade is distant from the real saccade. Overall, the Bahill trajectory is closer to the real

862 in this case. However, the Bahill saccade overshoots the target, assuming that the real saccade is accurate. In Fig. (B), the

863 tendency for the Enderle trajectory to move in the wrong direction is minimal but present, and overall the Enderle

864 saccade is more similar to the real saccade than the Bahill saccade which undershoots more the target.

865

866 Figures (C) and (D) provide examples of how OPMMs generate trajectories for saccades with amplitudes less than  $5^\circ$ .

867 In Figure (C), both predicted saccades are a poor fit for this small real saccade. The Bahill model produces a better fit

868 than the Enderle model over the interval  $[0, 19]$ . However, the landing point of the saccade from Enderle’s model was

869 closer to the original. Figure (D) demonstrates a case where the performance of both models is similar. However, at the

870 end, Enderle’s model overshoots the target.

871

872 Figures (E) and (F) illustrate examples where OPMMs generate trajectories for saccades with absolute amplitude

873 values in the range  $[5, 10]$ . Figure (E) demonstrates an instance where the Bahill model produces a saccade more

874 identical to the original one. However, the Enderle model outputs the trajectory which undershoots the real one. Figure

875 (F) illustrates a case where both models yield the same simulated trajectory with very similar simulation errors and

876 trajectory paths.

877

878

879 Figures (G) and (H) illustrate examples where OPMMs generate very similar trajectories with slightly different

880 simulation errors. Figure (G) illustrates a case where both models yield the same simulated trajectory with very similar

881 simulation errors and trajectory paths. Figure (H) presents a case where both models exhibit different behavior in

882 simulation. Once again, the Enderle saccade starts in the wrong direction but reverses at about 10 ms.

883

884

885 Figures (I) and (J) depict examples of saccades with amplitudes exceeding  $12^\circ$ . Figure (I) demonstrates a scenario  
886 where both models exhibit poor simulation scores. In Figure (J), it is evident how differently the models simulate the  
887 trajectory for the first half of the saccade. The Bahill trajectory is more similar to the real one. As the trajectories  
888 progress, both simulated trajectories undershoot the real saccade.  
889

890 Figures (K) and (L) illustrate cases where models can produce trajectories that deviate significantly from the original.  
891 In Fig. (K), both modeled trajectories are very badly formed. We see that for the first 10 ms, the Enderle saccade moves  
892 in the wrong direction, then moves in the correct direction but moves too quickly and overshoots the real saccade.  
893 The Bahill trajectory is simply incorrect. It dramatically overshoots the target 20 ms before the end and then reverses  
894 direction. Real saccades never do this. Moving on to Fig. (L), a notable difference is observed. Enderle's model generated  
895 a saccade with a significantly lower amplitude, approximately  $9^\circ$  less, which is also shifted at the beginning. In contrast,  
896 Bahill's model achieved simulation accuracy 2 times better.  
897  
898  
899  
900  
901  
902  
903  
904  
905  
906  
907  
908  
909  
910  
911  
912  
913  
914  
915  
916  
917  
918  
919  
920  
921  
922  
923  
924  
925  
926  
927  
928  
929  
930  
931  
932  
933  
934  
935  
936

937  
938  
939  
940  
941  
942  
943  
944  
945  
946  
947  
948  
949  
950  
951  
952  
953  
954  
955  
956  
957  
958  
959  
960  
961  
962  
963  
964  
965  
966  
967  
968  
969  
970  
971  
972  
973  
974  
975  
976  
977  
978  
979  
980  
981  
982  
983  
984  
985  
986  
987  
988

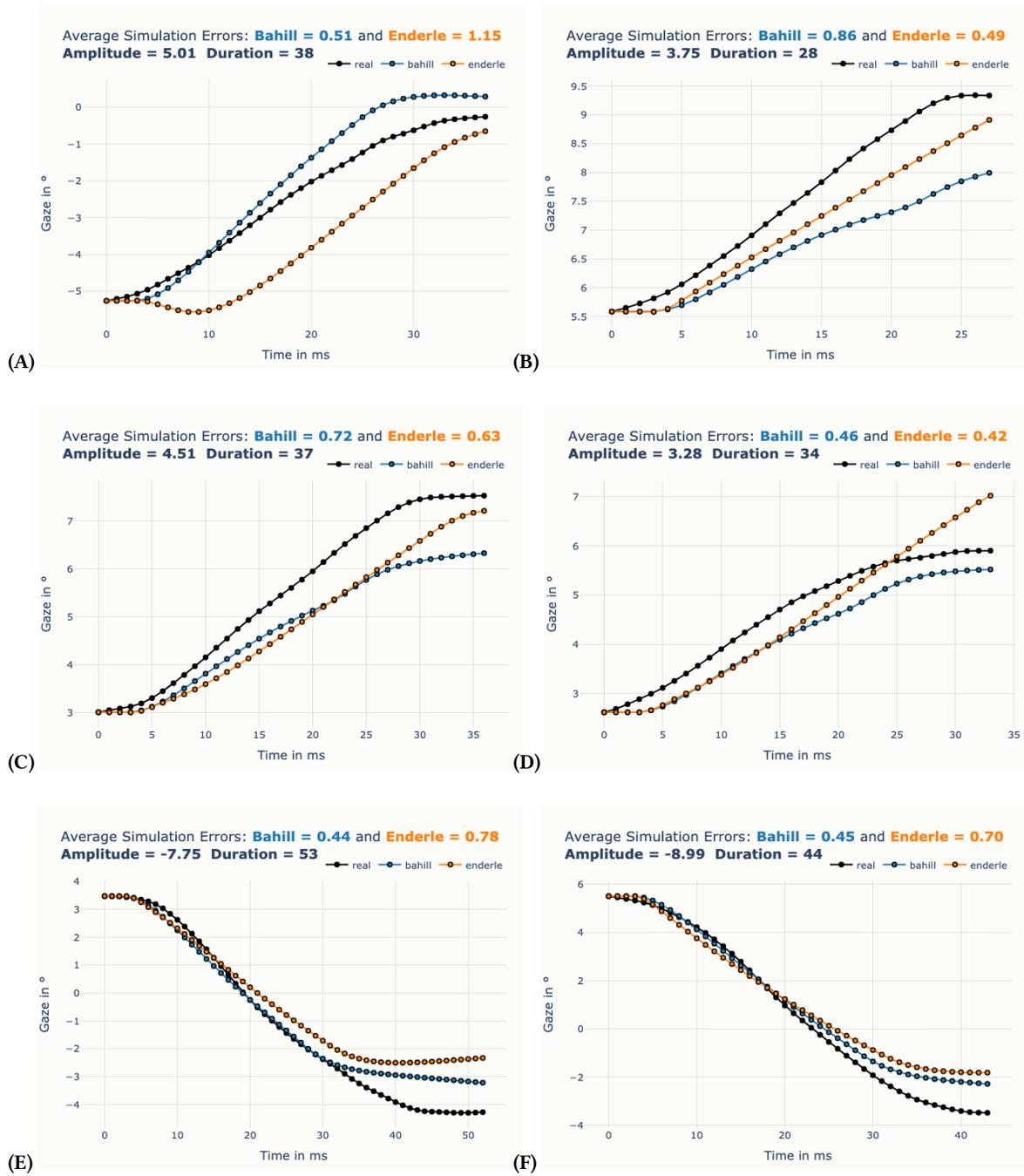


Fig. 3. Real Saccadic Trajectories Simulated by Bahill and Enderle OPMMs

989  
990  
991  
992  
993  
994  
995  
996  
997  
998  
999  
1000  
1001  
1002  
1003  
1004  
1005  
1006  
1007  
1008  
1009  
1010  
1011  
1012  
1013  
1014  
1015  
1016  
1017  
1018  
1019  
1020  
1021  
1022  
1023  
1024  
1025  
1026  
1027  
1028  
1029  
1030  
1031  
1032  
1033  
1034  
1035  
1036  
1037  
1038  
1039  
1040

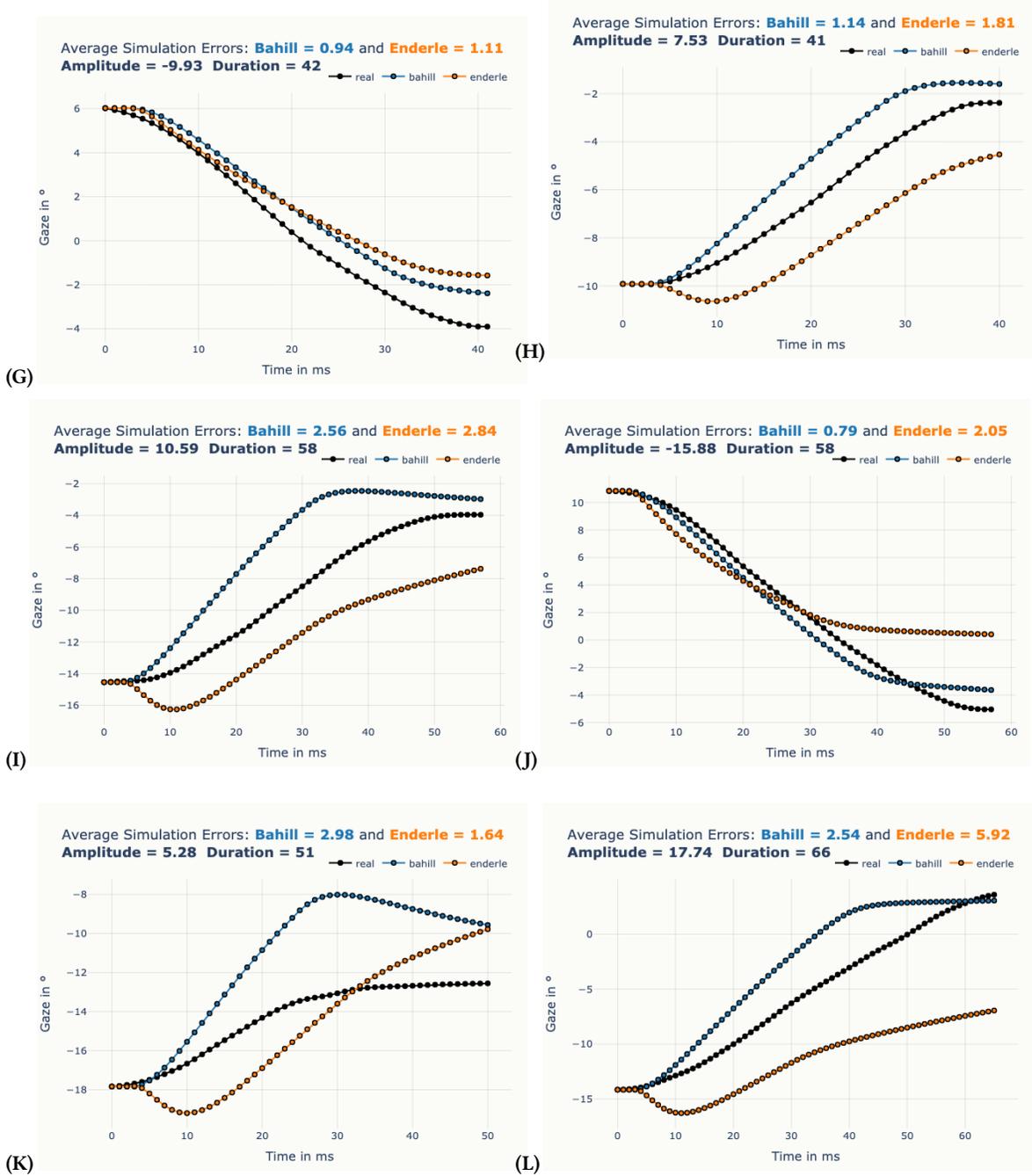


Fig. 4. Real Saccadic Trajectories Simulated by Bahill and Enderle OPMMs