
Computational methods for continuous eye-tracking perimetry based on spatio-temporal integration and a deep recurrent neural network

Alessandro Grillini (a.grillini@rug.nl)

Department of Ophthalmology, University Medical Center Groningen
Hanzeplein 1 9713 GZ Groningen, The Netherlands

Alex Hernández-García

Institute of Cognitive Science, University of Osnabrück
27 Wachsbleiche, 49090 Osnabrück, Germany

Remco J. Renken

Cognitive Neuroscience Center, Department of Biomedical Sciences of Cells and Systems
University Medical Center Groningen
Hanzeplein 1 9713 GZ Groningen, The Netherlands

Giorgia Demaria

Department of Ophthalmology, University Medical Center Groningen
Hanzeplein 1 9713 GZ Groningen, The Netherlands

Frans W. Cornelissen

Department of Ophthalmology, University Medical Center Groningen
Hanzeplein 1 9713 GZ Groningen, The Netherlands

Abstract

Perimetry, the mapping of the sensitivity of different visual field locations, is an essential procedure in ophthalmology. Unfortunately, standard automated perimetry (SAP), suffers from some practical issues: it can be tedious, requires manual feedback and a high level of patient compliance. These factors limit the effectiveness of perimetry in some clinical populations. In an attempt to remove some of these limitations, alternatives to SAP have been tried based on tracking eye movements. These new approaches have attempted to mimic SAP, thus presenting stimuli on a fixed grid, and replacing manual by ocular responses. While this solves some issues of SAP, these approaches hardly exploit the high spatial and temporal resolution facilitated by eye-tracking. In this study, we present two novel computational methods that do tap into this potential: (1) an analytic method based on the spatio-temporal integration of positional deviations by means of Threshold Free Cluster Enhancement (TFCE) and (2) a method based on training a recursive deep artificial neural network (RNN). We demonstrate that it is possible to reconstruct visual field maps based on continuous gaze-tracking data acquired in a relatively short amount of time.

Preprint.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

Keywords: perimetry; visual field; eye movement; recurrent neural network

1 **1 Introduction**

2 The assessment of the quality of the visual field (also called perimetry) is a staple of ophthalmic
3 evaluation. The presence of a scotoma, a region of the visual field with reduced sensitivity, is a very
4 characteristic symptom of diseases and disorders such as macular degeneration (Tolentino et al.,
5 1994), glaucoma (Heijl & Bengtsson, 1996), hemianopia (Williams, 1997) and several forms of
6 retinopathy (Greite et al., 1981; Voipio & Karjalainen, 2009; Alexander et al., 2004).

7 The current gold standard in the diagnostic assessment of the visual field is Standard Automated
8 Perimetry (SAP) (Barton & Benatar, 2003). The main advantages of SAP are a thorough evaluation
9 of multiple visual field locations, and relatively easy-to-interpret results that are normalized with re-
10 spect to an age-matched population. However, the approach also has several limitations: the task is
11 complicated for people with limited cognitive capabilities, demands patient compliance (Szatmáry,
12 2002), and requires maintaining a stable fixation for prolonged periods of time. Furthermore, patient
13 performance is affected by learning (Schultz, 1990; Wild et al., 2006) and fatigue (Johnson et al.,
14 1988), as well as the expertise of the operator (Montolio et al., 2012). Together, these constraints
15 limit the effectiveness of SAP, particularly in clinical and rehabilitation contexts such as when deal-
16 ing with children (Walters et al., 2012), the elderly and/or cognitively impaired patients (Gangeddula
17 et al., 2017).

18 To overcome some of these issues, various groups have implemented variants of SAP in which
19 eye-movements substituted the manual responses required on each trial of SAP (Jones et al., 2019;
20 Pel et al., 2013). While this already simplifies the task, the resulting procedure is still prone to
21 issues related to patient compliance and the requirement to stably fixate. Moreover, by retaining the
22 trial- and grid-based approach of SAP, the full potential of the high spatial and temporal resolution
23 facilitated by eye tracking is not exploited. For these reasons, we recently proposed a novel eye-
24 movement-based approach that completely removes the trial-based aspect of SAP, in favor of a
25 continuous assessment of oculomotor behavior over time (Grillini et al., 2018). This technique
26 exploits the ocular following response, thus reducing the attention and compliance demanded from
27 a patient. Simultaneously, the continuous assessment reduces test time considerably. In our study we
28 showed that, based exclusively on the spatio-temporal properties of the eye-movements made during
29 a short continuous tracking task, it is possible to classify a visual field defect (VFD). A limitation
30 of our initial approach was that it could only classify a scotoma as belonging to one of the scotoma
31 shape classes on which the machine classifier had been trained, and thus not reconstruct its actual
32 location and shape. The absence of this type of information hinders a more general application of
33 this technique in clinical and rehabilitation practice.

34 To overcome this limitation, here we propose two methods for analyzing continuous gaze data ac-
35 quired during a tracking task that enables reconstructing the visual field including any VFD present.

36 Our first method is based on the intuition that, compared to a healthy participant, a patient with a
37 VFD will make larger and more prolonged tracking errors if the stimulus falls within their scotoma-
38 tous region. In essence, the method applies Threshold-Free Cluster Enhancement (TFCE) (Smith &
39 Nichols, 2009) to perform a spatio-temporal integration of a time series of eye-stimulus positional
40 deviations (the signal). This results in a weighted integration of the height and extent of the signal,
41 which in our case represent space (positional deviation) and time (duration of the deviation until it
42 is corrected), respectively. Next, we reconstruct the visual field and presence of any scotoma, by
43 back-projecting the TFCE values into visual field space. We will refer to it as the TFCE method.

44 The second method is based on training a recurrent deep artificial neural network. It constitutes a
45 data-driven approach that learns features from the time-series of gaze location data collected dur-
46 ing the tracking task. This method makes no explicit hypotheses about the underlying relationship
47 between eye movements and scotoma characteristics, and learns how the presence of a scotoma in-
48 fluences a participant's visual behavior during the tracking task. The algorithm is a 7-layer recurrent
49 neural network whose weights are optimized to minimize its time-point-wise predictions on a set of
50 labeled training data (obtained on the basis of gaze-contingent simulations of scotomas in known
51 locations). Once trained, the model is able to accurately predict whether the discrepancy between
52 eye and stimulus positions, at a given time point, is caused by the presence of a scotoma. We re-
53 construct the visual field and presence of any scotoma by back-projecting the RNN predictions into
54 visual field space. We will refer to it as the recurrent neural network (RNN) method.

55 Based on simulated gaze-contingent scotoma data previously reported on in Grillini et al. (2018), we
56 show that both methods are able to reconstruct the shape of the scotomatous regions. While further
57 improvements are still desirable, these methods constitute a stepping stone towards the design of
58 an easy and effective eye-movement-based perimetry. This technique, having fewer limitations than
59 SAP, can complement it in clinical and rehabilitation practices.

60 **2 Methods**

61 **2.1 Data acquisition**

62 Participants, apparatus, stimuli, and procedures of data acquisition have previously been described
63 in Grillini et al. (2018). To summarize, the stimulus comprised a small (0.5 deg) dot, displayed at
64 one of two possible contrast levels (5% and 50%), moving along a random walk path with or without
65 random displacements in order to induce saccades (the smooth pursuit and saccadic pursuit condi-
66 tions, respectively). The gaze of the participants was recorded with an infra-red light-based video
67 eye tracker while they tracked the stimulus with their eyes. During the experiment, the participants
68 (N=50) were subjected to different kinds of simulated gaze-contingent visual field defects: no loss,
69 central loss, peripheral loss, and hemifield loss (see Figure 1). Each trial, lasting 20 seconds, was
70 repeated 6 times for each condition (2 contrast levels \times 2 pursuit modalities) for a total of 24 trials
71 and a total test time of 480 seconds. In order to reliably assess the quality of VF reconstructed
72 using the TFCE- and RNN-based methods, we carried out a 5-fold cross-evaluation. To do this, we
73 split the data from the 50 participants into 5 sets, each containing the data from 40 participants for
74 training and 10 participants for testing.

75 In order to assess the feasibility of these methods in a clinical setting, we further assessed three
76 participants diagnosed with primary open-angle glaucoma (POAG) and three age-matched healthy
77 control participants.

78 All participants gave their written informed consent prior to participation. The study was approved
79 by the Medical Ethical Committee of the University Medical Center Groningen and the Ethics Com-
80 mittee of Psychology of the University of Groningen. The study followed the tenets of the Declara-
81 tion of Helsinki.

82 **2.2 Gaze-contingent simulated VFD**

83 Details of the procedure to simulate VFD are described in Grillini et al. (2018). A schematic repre-
84 sentation of the three types of simulated VFD used in this study is shown in Figure 1.

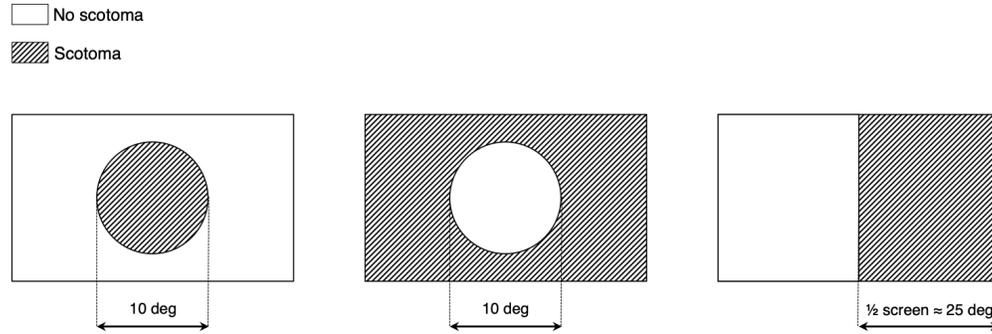


Figure 1: Examples of simulated visual field defects (VFD). From left to right: central loss, peripheral loss, hemifield loss. During the experiment, the VFD were applied in a gaze-contingent manner.

85 2.3 Method 1: Spatio-temporal integration of positional deviations by means of Threshold 86 Free Cluster Enhancement (TFCE)

87 To simultaneously factor in the magnitude of the error and its duration, we applied to our data the
88 Threshold Free Cluster Enhancement (TFCE) (Smith & Nichols, 2009). This algorithm, originally
89 developed for the analysis of the hemodynamic response in functional neuroimaging, specifically
90 helps to avoid the introduction of arbitrary thresholds when performing multiple-comparison cor-
91 rections. In our context, we applied the algorithm to the time-series of positional deviations D ,
92 where each value is the Euclidean distance between the gaze location and the stimulus position at
93 any given time point t . The positional deviations as a function of time are defined in Eq. 1:

$$D(t) = \sqrt{(p_x(t) - s_x(t))^2 + (p_y(t) - s_y(t))^2} \quad (1)$$

94 where $p(t)$ and $s(t)$ are the positions of the eye and the stimulus, respectively, divided into their
95 horizontal (x) and vertical (y) components.

96 $D(t)$ constitutes the input for the spatio-temporal integration performed with the TFCE equation,
97 described in Eq. 2:

$$D_{TFCE} = \int_{h=h_0}^{h_p} e(h)^E h^H dh \quad (2)$$

98 where e is the extent (temporal duration) and h is the height (spatial magnitude) of $D(t)$ at a given
99 point in time t (see Figure 2 for examples). This integral is implemented as a discrete sum using
100 a finite step-size dh , (in our implementation $dh = 1/2500^{th}$ of the maximum of $D(t)$); h_0 is the
101 minimum of $D(t)$ (which is always greater than or equal to 0), and E and H are the exponents.
102 The resulting $D_{TFCE}(t)$ is a time-series of positional deviations weighted for their spatio-temporal
103 integrated characteristics. Figure 2 shows some examples of using different E and H pairs: higher
104 E values (red signals) enhance clusters with longer duration and suppress shorter ones; higher H
105 values (blue signals) enhance the clusters with higher peaks and suppress the lower ones (Figure 2).
106 We set these parameters to the recommended values of $E = 2$ and $H = 0.5$ (Smith & Nichols, 2009).

107 To label each value of $D_{TFCE}(t)$ as healthy or visual loss we apply the following algorithm.

108 First, in order to obtain the frequency distribution of all possible normative values (i.e. values of
109 $D_{TFCE}(t)$ that should be considered healthy), $D_{TFCE}(t)$ is initially computed for every participant

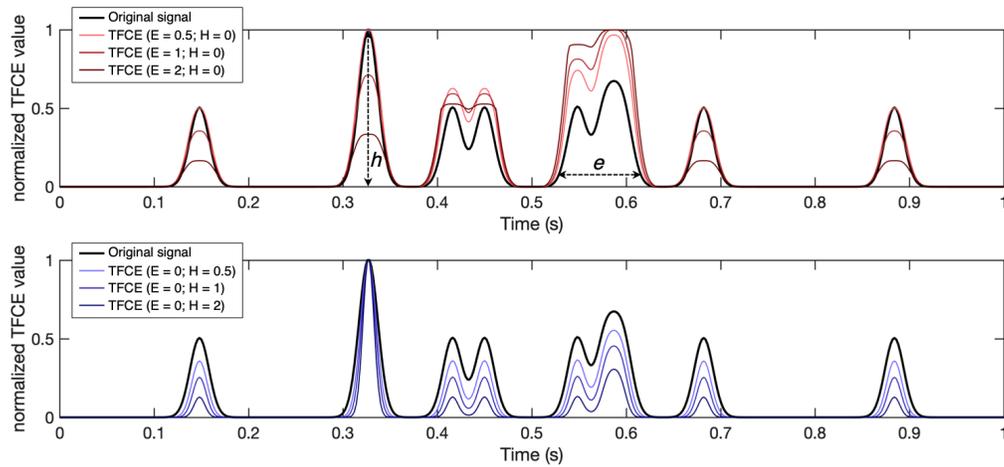


Figure 2: Examples of time series filtered using the TFCE algorithm with different combinations of parameters. A larger E weighs more the components of the original signal with longer temporal extent (e), defined as the time interval where the signal is consistently above a predefined minimum. Larger H , conversely, weighs more the highest components of the signal in terms of intensity (h). Note that the TFCE values here are normalized for visualization purposes only.

110 of the training dataset in the condition without simulated visual field loss, and the resulting values
 111 are aggregated with a histogram F .

112 Next, in order to choose the optimal boundary between “healthy” and “visual loss” values, we set
 113 a moving threshold λ_n , such that $F(\lambda_n) = n^{th}$ percentile of F . For each value of λ_n we compute
 114 $B(t; \lambda_n)$ that is the binarized form of $D_{TFCE}(t)$ such as that $B(t; \lambda_n) = 0$, if $D_{TFCE}(t) \leq F(\lambda_n)$ if
 115 “healthy” and $B(t; \lambda_n) = 1$, if $D_{TFCE}(t) > F(\lambda_n)$ if “visual loss”.

116 Finally, the visual field map is reconstructed as described in paragraph 2.5 “Time-series back-
 117 projection in visual field space”.

118 2.4 Method 2: Recursive Neural Networks (RNN)

119 In this method, we train a recursive neural network (RNN), as it is the most suitable known archi-
 120 tecture to account for the temporal properties of the data (Rumelhart et al., 1986).

121 As training input \mathbf{x}_t , we use the time series of the eye gaze positions $p(t)$ and the stimulus positions
 122 $s(t)$, as well as the luminance contrast (low contrast = 0; high contrast = 1) and type of pursuit of
 123 the stimulus (smooth pursuit = 0; saccadic pursuit = 1).

124 As training output \mathbf{y}_t , we use both the shape of the visual field defect (classified as *no loss*, *central*
 125 *loss*, *peripheral loss*, and *hemifield loss*) of the participants that generated each training sequence
 126 and, for each time time-point, whether the stimulus position lies in a location obstructed by the
 127 simulated scotoma.

128 As shown in Figure 3, the network consists of two streams that initially process the sequential data
 129 ($p(t)$ and $s(t)$) and categorical data (high/low stimulus luminance contrast and smooth/saccadic
 130 pursuit) separately. In particular, the sequential stream contains three bidirectional recurrent Gated
 131 Recurrent Unit (GRU) layers (Cho et al., 2014) to effectively process the temporal dependencies of
 132 the sequential data. The outputs of both streams are then concatenated and used to jointly train two
 133 different softmax classifiers. One is trained to classify the shape of the visual field defect (no loss,

134 central, peripheral, hemifield; analogously to what was reported by Grillini et al. (2018)), while the
 135 second one was trained to classify the visual field in a point-wise manner, i.e. “does the stimulus
 136 position in visual field space coordinates overlap with the scotoma?”.

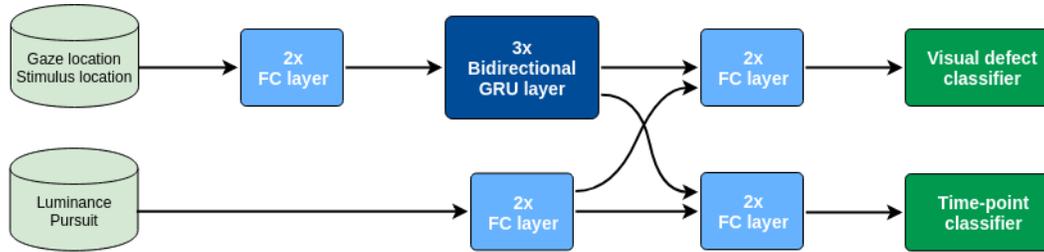


Figure 3: Schematic representation of the architecture of the deep recursive neural network.

137 We use the cross-entropy loss to define the cost function of the model, defined in Eq. 3:

$$J(\theta) = -\alpha \sum_{c=1}^{M_s} y_{s,c} \log(p_{s,c}) - \beta \sum_{c=1}^{M_d} y_{d,c} \log(p_{d,c}) \quad (3)$$

138 where M is the number of classes, y is the ground truth label and p is the predicted probability
 139 distribution, i.e. the output of each softmax classifier. Subscript s refers to the point-wise scotoma
 140 classifier and subscript d to the visual field defect shape classifier. In order to give priority to
 141 optimizing the reconstruction of the visual field, we set $\alpha = 0.75$ and $\beta = 0.25$. The parameters θ
 142 of the model are learned through mini-batch gradient descent, using RMSprop (Tieleman & Hinton,
 143 2012), for 15,000 iterations with a batch size of $B_s = 128$.

144 The training batches are formed by first selecting B_s different sequences from the set of 20-seconds
 145 trials, originally sampled at 240 Hz. Then, we randomly sampled one sub-sequence of 4.17 seconds
 146 (1,000 time steps) from each sequence and finally down-sample them at 60 Hz (250 time steps). The
 147 stimulus contrast level and pursuit modality of the corresponding sequences are also added to the
 148 training batches.

149 This deep model can be regarded as a mapping $\mathbf{y} = f(\mathbf{x}; \theta)$, where

150 $\mathbf{y} = \begin{bmatrix} pw_s \\ pw_d \end{bmatrix}$, pw_s being the point-wise scotoma prediction and pw_d the visual field defect shape pre-
 151 diction of a sub-sequence \mathbf{x} . In order to classify the visual field defect shape of one participant, we
 152 are interested here in pw_d .

153 Since the data acquisition for one participant consists of 6 repetitions of 20-seconds trials for each
 154 luminance/contrast combination, we average the predicted output probability distributions of mul-
 155 tiple sub-sequences. In particular, we average the predictions of the $6 \times 2 \times 2 = 24$ down-sampled
 156 sequences. The predicted visual field defect for a participant s is thus defined by Eq. 4:

$$\hat{y} = \arg \max_c \frac{1}{M} \sum_{i: \mathbf{x}_i \in S} f(\mathbf{x}_i; \theta) \quad (4)$$

157 where M is the number of sub-sequences in the set of trials S of participant s .

158 2.5 Time-series back-projection into visual field space

159 Both the threshold-free cluster enhancement and the recurrent neural network outputs consist of
 160 binarized time-series $B(t)$ where each entry has a value 0 if it is classified as not being obstructed

161 by a scotoma and 1 if it is. Each entry also has associated a pair of xy coordinates, where x is
162 the difference between horizontal gaze and stimulus positions at that time point and, analogously,
163 y is the difference between vertical gaze and stimulus positions. These are retinotopic coordinates,
164 meaning that they represent where the stimulus was with respect to the gaze of the participant. These
165 coordinates are then binned into an $N \times M$ grid, where each square represents 1 deg^2 of visual space
166 and N and M are the dimensions of the visual field tested. Each square contains the percentage
167 of occurrences that that specific location has been classified as being obstructed by a scotoma and
168 gets color-coded accordingly. For visualization purposes, these retinotopic coordinates can be easily
169 converted into polar coordinates (see Figure 6 and 7 for several examples).

170 A summary of the pipeline for visual field map reconstruction from gaze-tracking TFCE-filtered
171 time-series is shown in Figure 4. The RNN visual field map reconstruction is analogous, with the
172 binarized values of $B(t)$ being provided by the outcome of the time-point classifier (Figure 3) rather
173 than the threshold value $F(\lambda_n)$ applied to TFCE-filtered eye-tracking signals.

174 Note that the back-projection into visual field space has been used to reconstruct the ground truth
175 maps as well. In their case, the information about the presence or absence of the scotoma is known
176 a priori, but the reconstruction is still necessary to ensure a proper spatial comparison between the
177 ground truth and TFCE or RNN maps.

178 **2.6 Evaluation of reconstructed maps accuracy and optimization of λ**

179 In order to evaluate the accuracy of both the TFCE and the RNN methods, as well as to choose the
180 optimal λ_n parameter for the TFCE method, we compute a 2D Spearman rank correlation between
181 the reconstructed visual field maps and their respective ground truth maps obtained with the known
182 locations of the simulated scotomas.

183 The ground truth maps are obtained analogously as described in the Methods' paragraph Time-
184 series back-projection in retinotopic coordinates, using as an input the binarized time-series using
185 the known location of the visual field loss instead of the predictions made by either TFCE or RNN.

186 First, we measure the correlation between ground truth and TFCE maps of the training set recon-
187 structed using all possible values of $\lambda_n = [1, 2, 3, \dots, 100]$ (i.e. one for each percentile of the histogram
188 F). Then the average between participants is computed for each simulated visual loss condition, fol-
189 lowed by the grand average across conditions. The peak of the grand average corresponds to the
190 optimal value of λ_n that is used to reconstruct the maps of the test data. This procedure is repeated
191 for each of the 5 folds of the cross-evaluation.

192 **2.7 Clinical application of eye-tracking-based visual field reconstruction**

193 In order to provide a proof-of-concept of the viability of these methods in a clinical setting, we asked
194 3 patients and 2 healthy controls participants with various degrees of visual field loss to perform the
195 visual tracking task. We then compared the maps obtained with TFCE and RNN to those obtained
196 with Standard Automated Perimetry (SAP) by means of the Humphrey Field Analyzer (HFA), us-
197 ing the SITA-Standard algorithm. The HFA was performed monocularly on the eye affected by
198 primary open angle glaucoma (POAG) in the case of patients and in the dominant eye in the case
199 of controls. We then compared the Mean Deviation (MD) as reported by the HFA (MD_{HFA}) with
200 an MD computed on the basis of the TFCE- and RNN-reconstructed maps (MD_{TFCE} and MD_{RNN} ,
201 respectively).

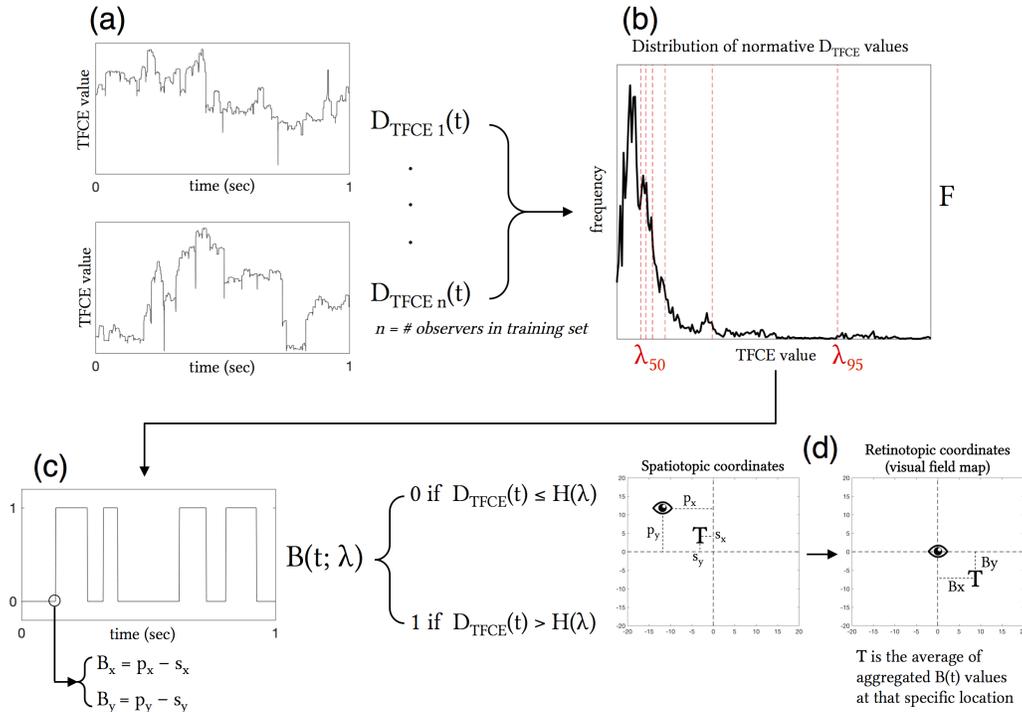


Figure 4: Schematic representation of the visual field map reconstruction algorithm pipeline. Starting from the TFCE-filtered time-series (a), a probability distribution F of all possible normative TFCE values is computed (b). For all percentiles λ_n of the distribution F , a threshold value $F(\lambda_n)$ is defined and used to binarize the TFCE-filtered signal in to be analyzed (c). Each time point of the TFCE-filtered signal has associated the spatiotopic coordinates of gaze (p_x and p_y) and stimulus positions (s_x and s_y), which are converted into retinotopic coordinates B_x and B_y . The resulting mapped retinotopic coordinates (d) are associated with a target location T which contains the average of the aggregated $B(t)$ values at that specific location, that is the expected probability that that specific location is affected by a scotoma. An analogous back-projection algorithm is implemented for the reconstruction of the visual field using the recurrent neural network, where the binarized time-series in (c) is defined by the output of the model instead of the threshold $F(\lambda_n)$.

202 Since the stimulus used for the tracking task followed a random-walk path, we could not ensure
 203 a complete coverage of the whole visual field. Therefore, the computation of the MD_{TFCE} and
 204 MD_{RNN} values comprises a correction for visual field coverage. Our MD values are computed as
 205 follows:

$$MD = \left[\frac{1}{n'} \sum_{i=1}^{n'} T(B_x, B_y) \right] \frac{n'}{n} = \frac{\sum_{i=1}^{n'} T(B_x, B_y)}{n} \quad (5)$$

206 Where n' is the number of visual field locations sampled, n is the number of all possible visual field
 207 locations and $T(B_x, B_y)$ is the expected probability that that specific location is affected by visual
 208 loss (see Figure 4).

209 **3 Results**

210 The results of the optimization of the TFCE parameter λ_n based on the maximum average accuracy
211 of each condition for each fold are shown in Figure 5, while Figure 6 shows the effect that adjusting
212 λ_n has on the visual field map reconstruction of a participant across all conditions.

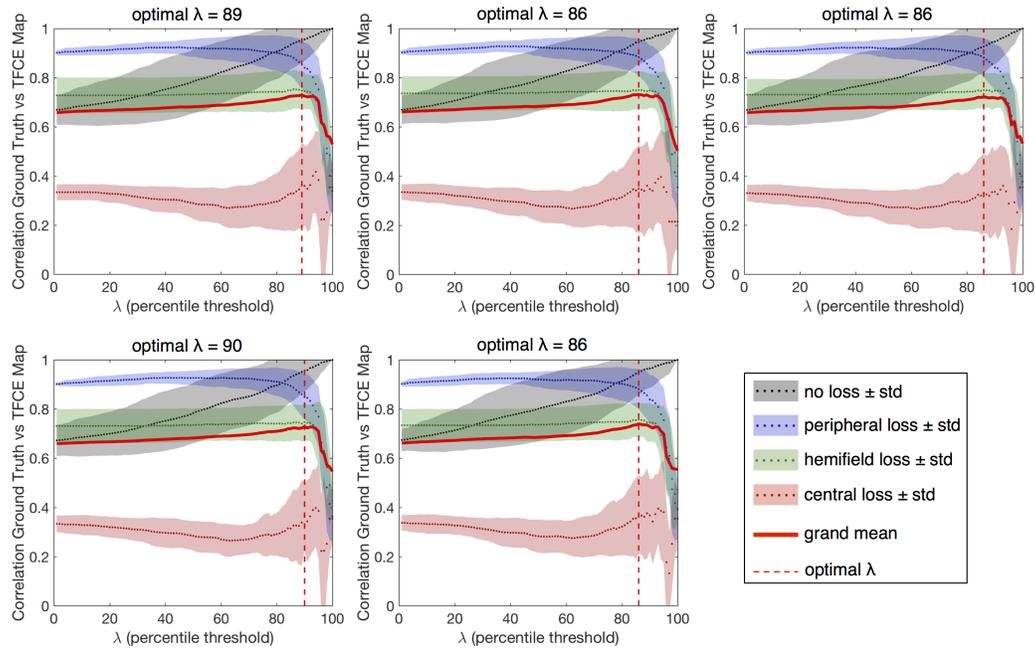


Figure 5: Results of the optimization of the λ parameter for each fold of the cross-evaluation. The optimal percentiles to be used as threshold between “healthy” and “impaired” TFCE values are in the range $86^{th} - 90^{th}$ (mean 87.4), corresponding to the peaks of the grand average between all tested conditions.

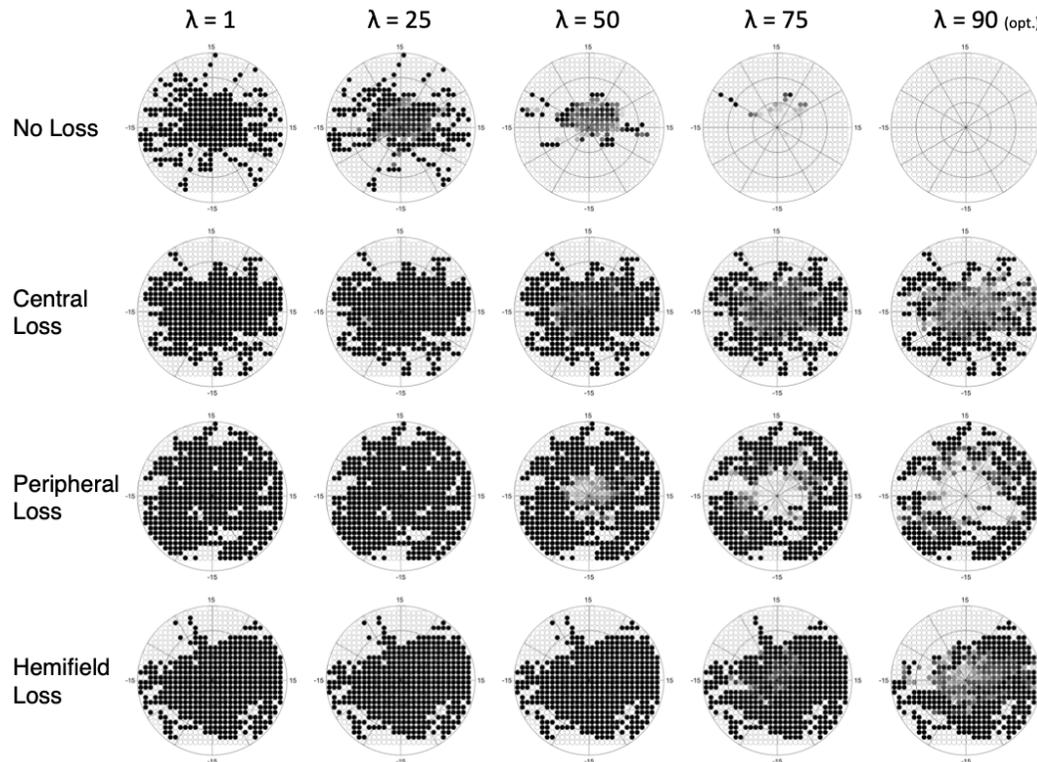


Figure 6: Example of the effect that adjusting the λ parameter has on a single participant, across all conditions. As the optimal λ is approached, the number of false positives in the “no loss” condition is minimized.

213 The results are very consistent across the 5 folds, with only the central loss condition showing
214 minimal variability. The resulting optimal λ_n is the average threshold determined for each of the
215 five folds (#1: 89; #2: 86; #3: 86; #4: 90; #5: 86).

216 Next, in the test set, we reconstruct maps using both the TFCE (with the optimized λ_n) and RNN
217 methods. For each method separately we compute the 2D Spearman rank correlation with the re-
218 spective ground truth reconstructed map. Figure 7 shows examples of reconstructed visual field
219 maps for all simulated visual loss conditions applied to one random participant of the test set.

220 We evaluated the overall performance of both the TFCE and RNN methods for all conditions tested,
221 applying the 5-fold cross-evaluation for both methods using the same 40-10 split. Each fold uses its
222 own optimised λ_n as shown in Figure 5.

223 Figure 8 shows the performance for both the TFCE and RNN methods, quantified with the spatial
224 distribution of False Positives and False Negatives (Figure 8-A), the False Positive Rate / False
225 Negative Rate (Figure 8-B) and with their accuracy, computed as the 2D Spearman rank correlation
226 between the reconstructed and ground truth maps (Figure 8-C).

227 A Kruskal-Wallis test (nonparametric one-way ANOVA) shows a statistically significant difference
228 between the two methods for all observed conditions (all p-values < 0.01), where the RNN method
229 proved to be more accurate than the TFCE in all but one condition (*No Loss*). Overall, the RNN
230 method also showed less variability between participants (see Table 1)

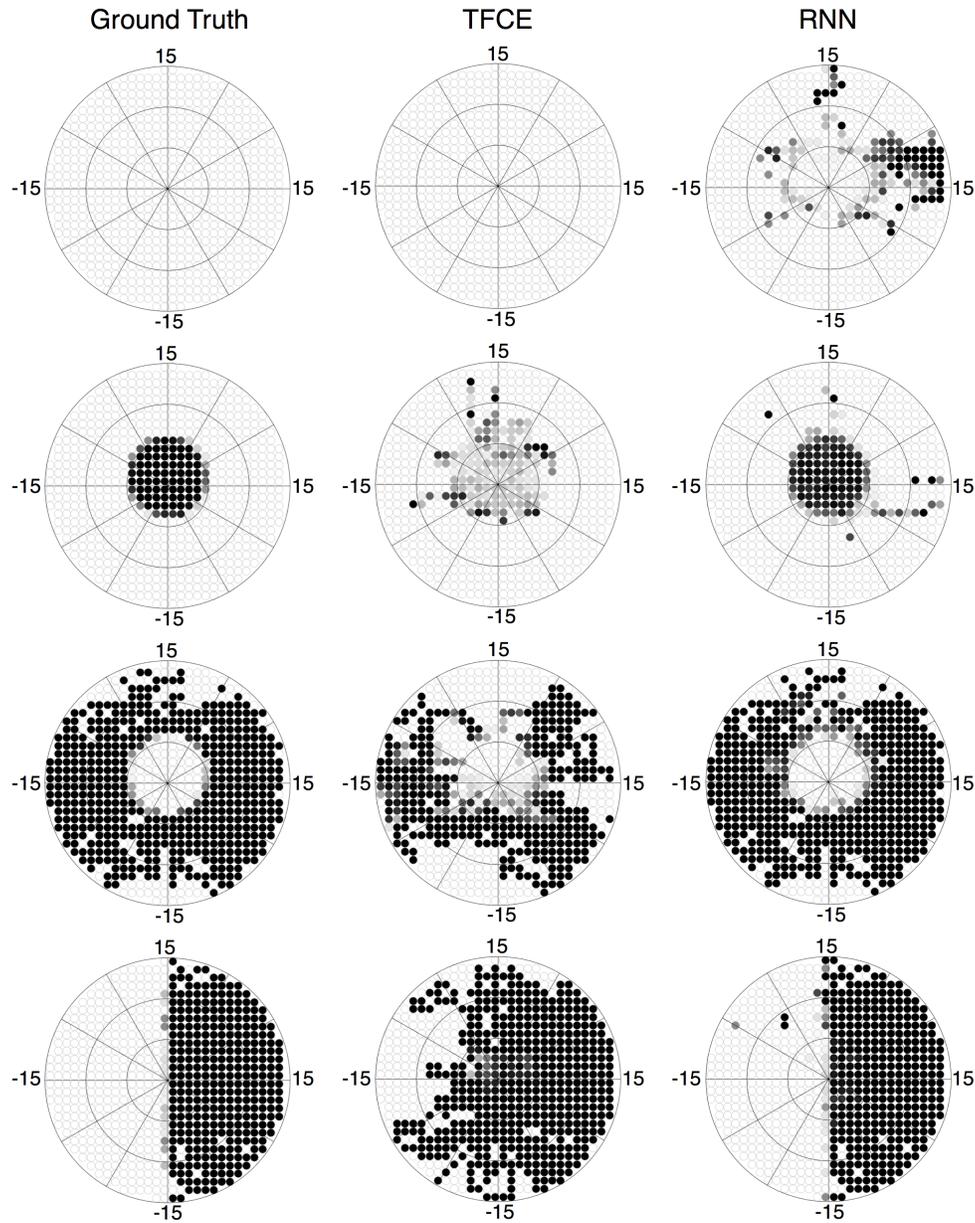


Figure 7: Reconstructed visual field maps for a single participant. Darker color indicates the higher probability of that point being affected by visual loss (white = 0%, black = 100%). The TFCE maps, although less accurate than the RNN ones, still allow to recognize the shape of the underlying scotomas. The RNN maps, on the other hand, closely resemble the ground truth maps.

231 Finally, to evaluate the generalizability of our approaches to a real clinical context, we compared the
232 TFCE and RNN methods with the perimetry maps obtained with the Standard Automated Perime-
233 ter HFA, the current gold standard in perimetry. For this comparison, we tested participants with
234 real visual field defects of different severity and compared the MD_{HFA} to the MD_{TFCE} and MD_{RNN}
235 computed as described in paragraph "Application of eye-tracking visual field reconstruction to clin-
236 ical data" in Methods. For this comparison we took as optimal TFCE λ_n the average between the
237 optimal values of each fold. This λ_n corresponds to the 87.4th percentile of the distribution of all

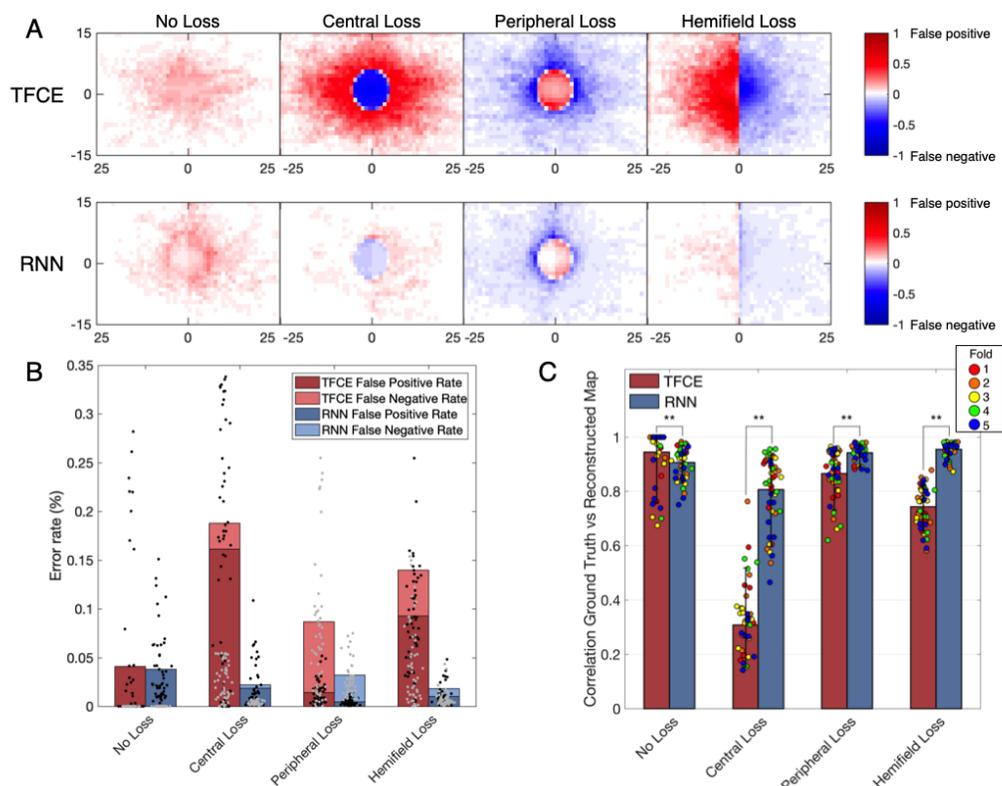


Figure 8: A. Spatial distributions of false positives and false negatives for each condition and for both TFCE and RNN methods. B. Total error rates combining false positive rate and false negative rate of TFCE and RNN. C. Comparison between TFCE and RNN visual field map reconstruction accuracies. The RNN method shows higher correlations with the ground truth as well as being more consistent than the TFCE method. Statistically significant differences ($p < 0.01$) are found across all tested conditions. Error bars show the 10th and 90th percentiles of each distribution.

	No Loss	Central Loss	Peripheral Loss	Hemifield Loss	Average
Accuracy TFCE	0.9455	0.3122	0.8664	0.7435	0.7169
Accuracy RNN	0.9070	0.8066	0.9429	0.9555	0.9030
IQR TFCE	0.0693	0.1651	0.1038	0.1138	0.1141
IQR RNN	0.0693	0.1863	0.0303	0.0268	0.0782

Table 1: Accuracy and variability of the two methods TFCE and RNN.

238 possible TFCE normative values. The clinical participants have been treated as a new independent
 239 test set, no re-training of the RNN model has been performed.

240 Contrary to the simulated scotomas, the TFCE method outperformed the RNN in this case, showing
 241 a significant correlation between its MD and the MD of the HFA. ($R^2_{TFCE} = 0.88$, $p_{TFCE} = 0.0117$;
 242 $R^2_{RNN} = 0.49$, $p_{RNN} = 0.112$).

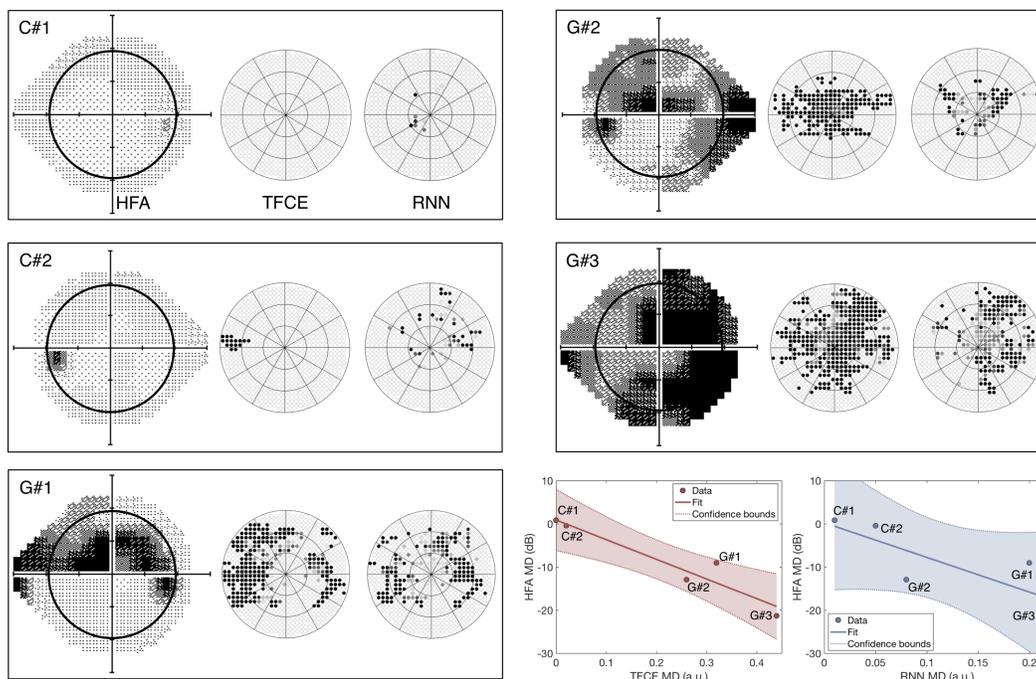


Figure 9: Perimetric maps of the five patients tested with our continuous tracking test. The TFCE maps are reconstructed using as λ_n the average of the values obtained by each individual fold (see Figure 5). C#1 and C#2 are healthy controls, while G#1, G#2, and G#3 are patients previously diagnosed with Primary Open Angle Glaucoma. The black circle within the HFA maps represents the portion of the visual field covered by the TFCE and RNN maps. The TFCE method shows a significant correlation between its Mean Deviation index and the one obtained with the HFA, whereas the RNN method does not ($R_{TFCE}^2 = 0.88$, $p_{TFCE} = 0.0117$; $R_{RNN}^2 = 0.49$, $p_{RNN} = 0.112$).

243 4 Discussion

244 Our main conclusion is that it is possible to reconstruct visual field maps, including the location
 245 of a scotoma, based on eye-tracking data acquired with a method of continuous gaze tracking. We
 246 consider this a breakthrough proof-of-principle, as it indicates a pathway towards the design of
 247 a high-resolution, patient-friendly way to perform perimetry. Below, we discuss the merits (and
 248 limitations) of the two methods for visual field map reconstruction that we presented and tested in
 249 this study: (1) spatio-temporal integration of positional deviations performed with Threshold Free
 250 Cluster Enhancement (TFCE) and (2) recursive deep artificial neural network (RNN). Moreover,
 251 we will compare our techniques to other proposed methods for eye-tracking-based perimetry and
 252 discuss possible further improvements.

253 4.1 Continuous gaze-tracking allows the reconstruction of visual field maps

254 4.1.1 TFCE

255 The spatio-temporal integration of positional deviations via Threshold Free Cluster Enhancement
 256 allows the reconstruction of visual field maps without requiring any prior knowledge about visual
 257 field defects. It is a method easy to implement, computationally inexpensive, and biologically-
 258 plausible: a loss of sensitivity in the visual field is associated with lower accuracy and higher delays
 259 of the eye-movements landing in the impaired region. This is consistent with findings from previous

260 studies involving patients with central (Van der Stigchel et al., 2013) and peripheral (Burton et al.,
261 2014) visual field defects.

262 The TFCE method performed quite well in reconstructing visual field maps with no loss or with
263 peripheral loss (accuracies of 0.95 and 0.87, respectively), while it fared less optimally in recon-
264 structing central losses and hemifield losses (accuracies of 0.31 and 0.74, respectively). A plausible
265 reason for this discrepancy is the heavy foveal bias that is inherent to continuous tracking tasks: to
266 accurately track the stimulus the observer must keep it as close as possible to the centermost part of
267 their visual field. In case of occlusion of the fovea, this is not possible, leading to prolonged errors
268 that never allow the positional deviations to return to zero (i.e. when the eye is on the target). In the
269 TFCE algorithm, this results in an erroneous definition of the baseline (the h_0 parameter in Eq. 2)
270 which in turn leads to an increased error rate towards the center of the visual field for central loss
271 and hemifield loss conditions (Figure 8-A and 8-B).

272 Furthermore, although this method fares well in detecting a scotoma in its expected location, it
273 struggles in precisely defining the edges of the defect (for examples see Figure 7, second and fourth
274 rows). This can be due to the presence of compensatory eye movement strategies in the observer
275 or the use of a different preferred retinal locus in the presence of visual field defects, whether real
276 (Coeckelbergh et al., 2002) or artificial (Cornelissen et al., 2005; McIlreavy et al., 2012).

277 Despite these limitations, the maps reconstructed with this method corresponded well to those re-
278 constructed using SAP (Figure 9) and their respective MD values significantly correlated with each
279 other, showing promising potential for the generalizability into clinical use.

280 4.1.2 RNN

281 The reconstruction of visual field maps by means of the RNN method proved to be highly accurate
282 with simulated scotomas, with an average accuracy across conditions above 0.90. This method,
283 however, did not show significant correlations with the maps of POAG patients obtained using SAP.
284 This could be due to the way the network was trained.

285 Using a limited number of predefined scotoma shapes made the RNN optimal at reconstructing sim-
286 ilar scotomas (with an average accuracy above 0.90), but not as effective in dealing with new shapes
287 not encountered before. This is also evident from the spatial distributions of false positives and false
288 negatives (Figure 8-A, lower panel), where the RNN reveals having a clear internal representation
289 of the four scotoma shapes used in the training data.

290 Another aspect that limits the implementation of the RNN in a clinical setting is that, in its current
291 form, it requires training data for which the location of the scotoma is known. This is easily achiev-
292 able using simulated gaze-contingent scotomas, but with actual patients, it is impossible to establish
293 their objective ground truth. While comparison to SAP is an obvious approach as it is the current
294 gold standard, this method can be unreliable when testing moderate to severe visual loss (Gardiner
295 et al., 2014), cognitively impaired (Diniz-Filho et al., 2017), or very young patients (< 6 yrs old)
296 (Patel et al., 2015; Tschopp et al., 1998), leading to a distorted ground truth that would not constitute
297 good training data.

298 We must note though that most of these limitations can potentially be overcome by re-designing the
299 way the training data is acquired. In the section 4.3 “*Current limitations and future improvements*”,
300 we propose possible solutions to the training issue and other problems.

301 4.2 Comparison with existing tools for eye-tracking-based perimetry

302 The rationale behind the development of a perimetric tool based on continuous gaze tracking is
303 rooted in the previous evidence that kinetic perimeters, i.e. devices such as the Goldmann Perimeter
304 and the Octopus 900 where the probing stimulus is moving, outperform SAP both in reliability
305 and ease of use in patients aged 5 to 8 (Patel et al., 2015). This suggests that the use of moving
306 stimuli in perimetry, although considered not optimal for the general population, might be relevant
307 within specific clinical contexts, and eye-tracking techniques constitute a fertile ground to explore
308 this possibility.

309 We are not the first to propose eye tracking as a means towards removing some critical aspects of
310 SAP, such as its high cognitive load and the need for manual feedback (Kim et al., 1995; Jones et al.,
311 2019; Pel et al., 2013). In fact, even the counting of fixation losses and determining blink frequency
312 can be seen as an elementary form of eye-tracking that is used to improve the reliability of SAP
313 (Ishiyama et al., 2015; Asaoka et al., 2019).

314 So far, all existing tools for eye-tracking-based perimetry employ the same working principles of
315 conventional perimetry with the primary difference being using ocular responses instead of manual
316 ones. The patient is still asked to repeatedly answer the question “do you see the stimulus?”, and
317 the answer is provided by the landing (or not) of a saccade within a Region-of-interest (ROI) around
318 the target (Kim et al., 1995; Jones et al., 2019), or by the saccadic latency to fall within the ROI of
319 a displaced target (Pel et al., 2013).

320 This approach comes with the advantage of allowing a precise sensitivity threshold estimation for
321 each tested visual field location but has two major downsides. First, since each point needs to be
322 tested individually and repeatedly, the spatial resolution is intimately interrelated to the available
323 testing time. Second, discrete eye movement perimetry is heavily reliant on optimal instrument
324 calibration: if the average calibration error exceeds the ROI radius of each target, it is very well
325 possible to have completely invalid maps.

326 Our method based on continuous tracking is designed to be less affected by these issues. First, as
327 each time sample contributes to the final map and spatial binning is applied only *a posteriori*, even
328 without any smoothing, it is possible to obtain detailed maps (see Figures 6, 7, and 9).

329 Second, although still partially affected by calibration issues, our gaze-based method of visual field
330 mapping can still allow for accurate classification of visual field defects. Our method is based on
331 the same stimuli as used in a previous study by Grillini et al. (2018), where spatio-temporal features
332 were used to classify the shape of the underlying visual field defect. The categorical classification
333 in that study was performed by training a simple decision tree with the features explicitly extracted
334 from the gaze and stimulus data. The temporal features are affected minimally by poor calibration
335 and still yield sufficient information about the type of scotoma.

336 In the present study, these features are not made explicit, but their categorical classification is still
337 performed by our neural network (see Figure 3, upper stream). As the temporal dependencies are
338 taken into account by the long short-term memory properties of the bidirectional GRUs layers (Cho
339 et al., 2014), the network should be able to perform a satisfactory visual field defect classification
340 (not reconstruction) also in the presence of rather poor calibrations. Future studies should investigate
341 this possibility.

342 4.3 Current limitations and future improvements

343 While we believe that the results presented constitute a promising proof-of-concept of the viability of
344 continuous-gaze-tracking perimetry, their implementation into a clinical setting would still require
345 a number of improvements, both in the acquisition and analysis of gaze data.

346 First, we trained the recurrent neural network using simulated visual field defects in order to es-
347 tablish a ground truth for the presence vs. absence of a scotoma. While this is not feasible with
348 real scotomas, an alternative is to train the RNN with more realistic and diverse visual loss simu-
349 lations. The ones that we used were either masking the stimulus completely or not at all, as well
350 as having stereotyped shapes lacking all the idiosyncrasies that can be present in actual visual field
351 impairments. Training the RNN with different (more organic) shapes with multiple levels of contrast
352 reduction could provide a significant improvement over our present results.

353 Second, the TFCE method requires two parameters to be defined beforehand: H and E, representing
354 the weights to be attributed to the height and to the extent of the positional deviations, respectively.
355 These parameters are typically chosen empirically, and in the present study, we adopted the values
356 recommended by the original authors Smith and Nichols, 2009 (Smith & Nichols, 2009). A future
357 implementation of this method can comprise a preliminary optimization phase in which the parame-
358 ters H and E are chosen analytically using machine learning. Possible computational approaches for
359 the optimization of the TFCE hyperparameters would be grid-search cross-validation (Abu-Mostafa
360 et al., 2012) or Bayesian optimization (Brochu et al., 2010).

361 Third, the optimization of the stimulus properties. The current stimulus is the same used to extract
362 the spatio-temporal properties of eye movements for neuro-ophthalmic screening (Grillini et al.,
363 2020), but it can be further optimized to perform visual field assessment. For instance, properties
364 such as luminance contrast, speed, and trajectory of the moving target can be adjusted with the goal
365 of maximizing visual field coverage and minimizing central bias.

366 Fourth, the simulated visual loss data used for training the RNN (healthy controls) and the clinical
367 data (glaucoma patients) used to test it were acquired from two groups with different age ranges
368 (20-30 vs 70-80). The different age ranges may have had an effect on eye movements even in the
369 absence of a clinical condition (Rottach et al., 1996).

370 Lastly, an additional improvement could be to take into account the physiology of the retina and
371 model H and E according to the “hill of vision” (Jacobs & Patterson, 1985). In this case, different
372 values of H and E could be determined for sections of the retina at different eccentricity, such as
373 the central peak (0-10 degrees), the mid-plateau (15-25 degrees) and the peripheral decay (above
374 25 degrees) and combined with the hyperparameter optimization methods mentioned above for best
375 results.

376 4.4 Conclusions

377 We developed and proposed two methods that enable the reconstruction of visual field maps by es-
378 timating retinal sensitivity using continuous gaze-tracking data: (1) spatio-temporal integration of
379 positional deviations performed with Threshold Free Cluster Enhancement (TFCE) and (2) recur-
380 sive deep artificial neural network (RNN). The two methods possess complementary qualities (and
381 downsides): the TFCE is biologically-plausible and computationally efficient while the RNN is re-
382 markably accurate when provided with proper training data. We conclude that both methods can
383 contribute to making gaze-based perimetry more viable in the future.

384 4.5 Acknowledgments

385 This project has received funding from the European Union’s Horizon 2020 research and innovation
386 programme under the Marie Skłodowska-Curie grant agreements No 641805 (“NextGenVis”) and
387 No 675033 (“EGRET”). We are thankful to Nomdo Jansonius for his valuable advice and support.

388 4.6 Disclosure

389 AG, AHG and RJR: listed as inventors on the European patent application Grillini, A., Hernández-
390 García, A., Renken, J. R. (2019). Method, system and computer program product for mapping a
391 visual field. EP19209204.7 which is partially based on the content of this manuscript.

392 AG: majority shareholder of REPERIO B.V., a private company that develops ophthalmic and neu-
393 rological tests based on eye-movements.

394 GD: none; FWC: none.

395 References

396 Abu-Mostafa, Y. S., Magdon-Ismail, M., and Lin, H.-T. *Learning from Data: A Short Course*. AML
397 Books, January 2012.

398 Alexander, K. R., Barnes, C. S., Fishman, G. A., Pokorny, J., and Smith, V. C. Contrast-Processing
399 deficits in Melanoma-Associated retinopathy. *Investigative Ophthalmology & Visual Science*, 45
400 (1):305, 2004.

401 Asaoka, R., Fujino, Y., Aoki, S., Matsuura, M., and Murata, H. Estimating the reliability of glau-
402 comatous visual field for the accurate assessment of progression using the Gaze-Tracking and
403 reliability indices. *Ophthalmology Glaucoma*, 2(2):111–119, 2019.

404 Barton, J. J. S. and Benatar, M. Automated perimetry (humphrey field analyzer). *Field of Vision*,
405 pp. 45–69, 2003.

406 Brochu, E., Cora, V. M., and de Freitas, N. A tutorial on bayesian optimization of expensive
407 cost functions, with application to active user modeling and hierarchical reinforcement learning.
408 *CoRR*, abs/1012.2599, 2010. URL <http://arxiv.org/abs/1012.2599>.

409 Burton, R., Smith, N. D., and Crabb, D. P. Eye movements and reading in glaucoma: observations
410 on patients with advanced visual field loss. *Graefes Arch. Clin. Exp. Ophthalmol.*, 252(10):1621–
411 1630, October 2014.

412 Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio,
413 Y. Learning phrase representations using RNN Encoder–Decoder for statistical machine transla-
414 tion. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*
415 (*EMNLP*), 2014.

416 Coeckelbergh, T. R. M., Cornelissen, F. W., Brouwer, W. H., and Kooijman, A. C. The effect of
417 visual field defects on eye movements and practical fitness to drive. *Vision Res.*, 42(5):669–677,
418 March 2002.

419 Cornelissen, F. W., Bruin, K. J., and Kooijman, A. C. The influence of artificial scotomas on eye
420 movements during visual search. *Optom. Vis. Sci.*, 82(1):27–35, January 2005.

- 421 Diniz-Filho, A., Delano-Wood, L., Daga, F. B., Cronemberger, S., and Medeiros, F. A. Association
422 between neurocognitive decline and visual field variability in glaucoma. *JAMA Ophthalmol.*, 135
423 (7):734–739, July 2017.
- 424 Gangeddula, V., Ranchet, M., Akinwuntan, A. E., Bollinger, K., and Devos, H. Effect of cognitive
425 demand on functional visual field performance in senior drivers with glaucoma. *Front. Aging*
426 *Neurosci.*, 9:286, August 2017.
- 427 Gardiner, S. K., Swanson, W. H., Goren, D., Mansberger, S. L., and Demirel, S. Assessment of the
428 reliability of standard automated perimetry in regions of glaucomatous damage. *Ophthalmology*,
429 121(7):1359–1369, July 2014.
- 430 Greite, J.-H., H. Greite, J., P. Zumbansen, H., and Adamczyk, R. Visual field in diabetic retinopathy
431 (DR). *Fourth International Visual Field Symposium Bristol, April 13–16, 1980*, pp. 25–32, 1981.
- 432 Grillini, A., Ombelet, D., Soans, R. S., and Cornelissen, F. W. Towards using the spatio-temporal
433 properties of eye movements to classify visual field defects. *Proceedings of the 2018 ACM Sym-*
434 *posium on Eye Tracking Research & Applications - ETRA '18*, 2018.
- 435 Grillini, A., Renken, R. J., Vrijling, A. C. L., Heutink, J. H. C., and Cornelissen, F. W. Eye move-
436 ment evaluation in multiple sclerosis and parkinson's disease using a standardized oculomotor
437 and neuro-ophthalmic disorder assessment (SONDA). *Preprint on medRxiv*, 2020.
- 438 Heijl, A. and Bengtsson, B. Early visual field defects in glaucoma: A study of eyes developing field
439 loss. *Glaucoma: Decision Making in Therapy*, pp. 75–78, 1996.
- 440 Ishiyama, Y., Murata, H., and Asaoka, R. The usefulness of gaze tracking as an index of visual field
441 reliability in glaucoma patients. *Invest. Ophthalmol. Vis. Sci.*, 56(11):6233–6236, October 2015.
- 442 Jacobs, N. A. and Patterson, I. H. Variability of the hill of vision and its significance in automated
443 perimetry. *Br. J. Ophthalmol.*, 69(11):824–826, November 1985.
- 444 Johnson, C. A., Adams, C. W., and Lewis, R. A. Fatigue effects in automated perimetry. *Applied*
445 *Optics*, 27(6):1030, 1988.
- 446 Jones, P. R., Smith, N. D., Bi, W., and Crabb, D. P. Portable perimetry using Eye-Tracking on a
447 tablet Computer—A feasibility assessment. *Translational Vision Science & Technology*, 8(1):17,
448 2019.
- 449 Kim, D. E., Eizenman, M., Trope, G. E., and Kranemann, C. Eye movement perimetry. In *Pro-*
450 *ceedings of 17th International Conference of the Engineering in Medicine and Biology Society*,
451 volume 2, pp. 1629–1630. IEEE, 1995.
- 452 McIlreavy, L., Fiser, J., and Bex, P. J. Impact of simulated central scotomas on visual search in
453 natural scenes. *Optom. Vis. Sci.*, 89(9):1385–1394, September 2012.
- 454 Montolio, F. G. J., Junoy Montolio, F. G., Wesselink, C., Gordijn, M., and Jansonius, N. M. Factors
455 that influence standard automated perimetry test results in glaucoma: Test reliability, technician
456 experience, time of day, and season. *Investigative Ophthalmology & Visual Science*, 53(11):7010,
457 2012.
- 458 Patel, D. E., Cumberland, P. M., Walters, B. C., Russell-Eggitt, I., Rahi, J. S., and OPTIC study
459 group. Study of optimal perimetric testing in children (OPTIC): Feasibility, reliability and re-
460 peatability of perimetry in children. *PLoS One*, 10(6):e0130895, June 2015.

- 461 Pel, J. J. M., van Beijsterveld, M. C. M., Thepass, G., and van der Steen, J. Validity and repeatability
462 of saccadic response times across the visual field in eye movement perimetry. *Transl. Vis. Sci.*
463 *Technol.*, 2(7):3, November 2013.
- 464 Rottach, K. G., Riley, D. E., DiScenna, A. O., Zivotofsky, A. Z., and Leigh, R. J. Dynamic properties
465 of horizontal and vertical eye movements in parkinsonian syndromes. *Ann. Neurol.*, 39(3):368–
466 377, March 1996.
- 467 Rumelhart, D. E., Hinton, G. E., and Williams, R. J. Learning representations by back-propagating
468 errors. *Nature*, 323(6088):533–536, 1986.
- 469 Schultz, J. S. The “learning effect” in automated perimetry. *Ophthalmology*, 97(7):838, 1990.
- 470 Smith, S. and Nichols, T. Threshold-free cluster enhancement: Addressing problems of smoothing,
471 threshold dependence and localisation in cluster inference. *NeuroImage*, 44(1):83–98, 2009.
- 472 Szatmáry, G. Can swedish interactive thresholding algorithm fast perimetry be used as an alternative
473 to goldmann perimetry in neuro-ophthalmic practice? *Archives of Ophthalmology*, 120(9):1162,
474 2002.
- 475 Tieleman, T. and Hinton, G. Lecture 6.5-rmsprop: Divide the gradient by a running average of its
476 recent magnitude. *COURSERA: Neural networks for machine learning*, 4(2):26–31, 2012.
- 477 Tolentino, M. J., Miller, S., Gaudio, A. R., and Sandberg, M. A. Visual field deficits in early age-
478 related macular degeneration. *Vision Res.*, 34(3):409–413, February 1994.
- 479 Tschopp, C., Safran, A. B., Viviani, P., Reicherts, M., Bullinger, A., and Mermoud, C. Automated
480 visual field examination in children aged 5-8 years. part II: Normative values. *Vision Res.*, 38(14):
481 2211–2218, July 1998.
- 482 Van der Stigchel, S., Bethlehem, R. A. I., Klein, B. P., Berendschot, T. T. J. M., Nijboer, T. C. W.,
483 and Dumoulin, S. O. Macular degeneration affects eye movement behavior during visual search.
484 *Front. Psychol.*, 4:579, September 2013.
- 485 Voipio, H. and Karjalainen, K. RETINAL AND VISUAL FIELD CHANGES IN CHLOROQUINE
486 RETINOPATHY. *Acta Ophthalmologica*, 45(2):150–158, 2009.
- 487 Walters, B. C., Rahi, J. S., and Cumberland, P. M. Perimetry in children: survey of current practices
488 in the united kingdom and ireland. *Ophthalmic Epidemiol.*, 19(6):358–363, December 2012.
- 489 Wild, J., Kim, L., Pacey, I., and Cunliffe, I. Evidence for a learning effect in Short-Wavelength
490 automated perimetry. *Ophthalmology*, 113(2):206–215, 2006.
- 491 Williams, D. Visual function in patients with homonymous hemianopia. *Neurocase*, 3(5):395f–403,
492 1997.