tvst

Article

Comparative Study Between the SORS and Dynamic Strategy Visual Field Testing Methods on Glaucomatous and Healthy Subjects

Şerife Seda Kucur[1], Sebastian Häckel[2], Jan Stapelfeldt[1], Jeannine Odermatt[3], Milko E. Iliev[2], Mathias Abegg[2], Raphael Sznitman[1,#], and Rene Höhn[2,#]

[1] Artificial Intelligence in Medical Imaging Laboratory, ARTORG Center for, Biomedical Engineering Research, University of Bern, Bern, Switzerland
[2] Department of Ophthalmology, University Hospital Bern, University of Bern, Bern, Switzerland
[3] Department of Psychology, University of Bern, Bern, Switzerland

Correspondence: Şerife Seda Kucur, University of Bern, Murtenstrasse 50, Bern, 3008, Switzerland. e-mail: serife.kucur@artorg.unibe.ch

Purpose: To clinically validate the noninferiority of the sequentially optimized reconstruction strategy (SORS) when compared to the dynamic strategy (DS).

Methods: SORS is a novel perimetry testing strategy that evaluates a subset of test locations of a visual field (VF) test pattern and estimates the untested locations by linear approximation. When testing fewer locations, SORS has been shown in computer simulations to bring improvements in speed over conventional perimetry tests, while maintaining acquisition at high-quality acquisition. To validate SORS, a prospective clinical study was conducted at the Department of Ophthalmology of Bern University Hospital, over 12 months. Eighty-three subjects (32 healthy and 51 glaucoma patients with early to moderate visual field loss) of 114 participants were included in the study. The subjects underwent perimetry tests on an Octopus 900 (Haag-Streit, Köniz, Switzerland) using the G pattern with both DS and SORS. The acquired sensitivity thresholds (ST) by both tests were analyzed and compared.

Results: DS-acquired VFs were used as a reference. High correlations between individual STs ($r \geq 0.74$), as well as between mean defect values ($r \geq 0.88$) given by DS and SORS were obtained. The mean absolute error of SORS was under 3 dB with a 70% reduction in acquisition time. SORS overestimated healthy VFs while slightly underestimating glaucomatous VFs. Qualitatively, SORS acquisition yielded VF with detectable defect patterns, albeit some isolated and small defects were occasionally missed.

Conclusions: This clinical study showed that for healthy and glaucomatous patients, SORS-acquired VFs sufficiently correlated with the DS-acquired VFs with up to 70% reduction in acquisition time.

Translational Relevance: This clinical study suggests that the novel perimetry strategy SORS could be used in routine clinical practice with comparable utility to the current standard DS, whereby providing a shorter and more comfortable perimetry experience.

Introduction

Perimetry, also known as visual field testing, plays a central role in the clinical diagnosis and follow-up of glaucoma.[1] Perimetry evaluates the visual function of the examined eye by considering the central and peripheral field of view. Perimetry determines sensitivity thresholds (ST) at predefined retinotopic locations, leading to a functional map called a *visual field* (VF). Compared to an age-matched normative database of normal healthy subjects, sensitivity thresholds of a visual field are helpful for detecting abnormalities and losses in visual function due to glaucoma and other neuro-ophthalmic diseases.[2–6]

Perimetry is usually performed via a query-response procedure where the patient is presented with sequential light stimuli of different intensities at different locations of the visual field while fixing their gaze on a central fixation point. Patients are then asked to press a button each time they perceive light stimulus. The test can be lengthy in time, which can make it uncomfortable and can induce fatigue effects and concentration lapses. These negative factors correlated with higher false-positive/-negative response rates, rendering the examination results unreliable.[7–9]

To speed up visual field testing, several alternative perimetry strategies have been proposed in the literature, some of which have been adopted into routine clinical use. Early methods focused on improving testing speeds by reducing the number of presentations at one VF location through the use of various techniques that dynamically change light intensity stimuli based on past patient responses.[1,10–14] Such strategies sped up tests by up to 50% to 80% compared to conventional perimetry test strategies that take up to 15 minutes per eye. Other techniques looked for speed gains by estimating the starting light intensity at a next query location and thereby reducing the number of stimuli presented per location. Dynamic strategy (DS),[10,15] Swedish interactive threshold algorithm (SITA) standard,[16,17] SITA fast,[12,18] SITA faster,[19,20] and tendency-oriented perimetry (TOP)[21] are examples of such methods. DS uses a staircasing approach as also found in conventional techniques but with an adaptive step whose size depends on the steepness of the probability of seeing curve (POSC). With this approach, DS reduces the examination time to almost 5 minutes, with a slight compromise in the VF precision.[22] TOP brings larger reduction in examination times (e.g., less than two minutes) by exploiting the correlations within nearby VF locations.[21] Although DS can still be long for some patients to perform and TOP is used as an alternative, TOP-acquired VFs unfortunately suffer from lack of precision.[23] SITA standard and SITA fast are a family of algorithms that combine Bayesian and staircasing approaches and reduce the examination duration to approximately five to six minutes.[12,17] Although SITA strategies provide a good accuracy-speed tradeoff, an implementation of a SITA-like method was shown to perform poorly when the initial stimulus is over- or underestimated.[24] A recent version of SITA strategies, SITA faster, was designed by introducing seven modifications to SITA fast and, hence, was reported to take 2.8 minutes in average providing similar VF quality with SITA fast.[19]

More recently, methods have looked to model the relationship between VF locations in more comprehensive ways using graph-based approaches. Globally, such methods have demonstrated improvements in accuracy, time, or both.[25–28] However, these methods do not appear to bring speed improvements[25,26] or do so only on healthy patients[27] and have limited performance improvements because of a large number of parameters that must be tuned beforehand.[28]

Sequentially optimized reconstruction strategy (SORS)[29] proposes to query a subset of test locations available in the VF pattern and estimate the untested locations by linear approximation, exploiting the correlations between the tested and untested locations. In simulation, SORS was shown to improve speed compared to several strategies, including the clinically used DS and TOP without compromising precision. The purpose of this study was therefore to demonstrate the clinical reliability of SORS-acquired VFs by evaluating SORS on both normal healthy subjects and glaucomatous patients. Because DS is one of the most frequently used perimetry tests and the most used one for the often-used Octopus perimeters (Haag-Streit, Köniz, Switzerland), this study specifically aims at demonstrating the noninferiority of SORS-acquired VFs compared to DS within absolute, reduced examination time.

Methods

In this study, we performed a quantitative prospective randomized single-center study in the Department of Ophthalmology, Bern University Hospital from October 2018 to July 2019. VFs were collected and stored in a secure, web-based electronic data capture tool. The data was anonymized for further data processing and analysis. The study protocol was approved by Bernese Ethics Committee, Bern, Switzerland, and adhered to the tenets of the Declaration of Helsinki. Informed consent was obtained from the subjects about the study procedure.

Subjects

Patients were recruited in a glaucoma outpatient clinic at the Department of Ophthalmology, Bern University Hospital. The general inclusion criteria were an age between 40 and 80 years, refractive error within ±5 diopter spherical equivalent, an astigmatism less than −3 diopter, a visual acuity of more than 0.3 log Mar, a history of at least one perimetry examination and less than 20% false-positive and -negative errors for both DS and SORS examinations. Healthy subjects had mean defects (MD) of

Table 1. Age, MD, and sLV Statistics of the Patient Data Collected in the Study, Along With Mean, SD, and Max/Min Values

	No. of Patients	Age (Mean/SD [Min., Max.])	MD (Mean/SD [Min., Max.])	sLV (Mean/SD [Min., Max.])
Healthy	32	65.31/9.87 [41.41, 80.93]	1.36/1.23 [−0.70, 4.30]	2.48/0.76 [1.60, 4.80]
Glaucoma	51	67.40/8.43 [42.78, 80.93]	5.75/2.76 [0.50, 12.30]	5.67/2.12 [1.90, 11.90]
All	83	65.31/9.87 [40.41, 80.93]	4.06/3.14 [−1.70, 12.30]	4.44/2.32 [1.60, 11.90]
Test-retest	10	66.28/10.00 [41.41, 80.33]	3.32/2.85 [−0.70,9.00]	3.32/1.30 [1.60, 6.00]

less than 2 dB; glaucomatous subjects were diagnosed with either primary open-angle, pseudoexfoliation or primary angle-closure glaucoma, with early to moderate visual field loss (+2 dB < MD < +12 dB). Exclusion criteria were the inability to follow the procedure, insufficient knowledge of the project language (German or French), no history of ocular diseases other than glaucoma or cataract, or any other visual pathway conditions that might affect visual field testing (e.g., pituitary lesions, demyelinating diseases).

Overall, 114 subjects were enrolled in this study, 83 of them (32 healthy and 51 glaucoma) met the quality criteria (i.e. false positive and negative errors each less than 20% in each of the SORS and DS examinations) and were included in the study. Moreover, a randomly selected sub-group of 10 enrolled subjects (four healthy, six glaucoma) were tested twice with both DS and SORS in addition to the main study protocol to perform a test-retest analysis. Table 1 summarizes the age, MD and square-root loss variance (sLV) of the included participants, as well as the test-retest subgroup.

Visual Field Acquisition

Subjects meeting the inclusion criteria were asked to undertake an additional VF SORS examination during the same visit of their scheduled DS perimetry examination appointment. VFs of both eyes were collected using DS and one randomly determined study eye was tested using SORS. Both perimetry strategies used the G program with 59 locations within 30° field of view. For a given participant, all examinations were conducted by the same person on the same Octopus 900 Perimeter (Haag-Streit, Inc. Köniz, Switzerland). The order of DS and SORS tests was randomized to avoid fatigue or other biases. Patients were given a mandatory break of five minutes between DS and SORS examinations. Eye tracking and blinking control was turned off for all examinations.

For the collection of test-retest data, second examinations were performed less than six months following the first, with the identical experimental setup as previously described. That is, we assumed no significant change related to disease progression between measurements collected less than six months apart.

SORS: A Brief Description

SORS relies on the assumption that VF locations tested during a perimetry examination are linearly dependent to each other. Given this, a VF can be "reconstructed" by testing only a few locations and then estimating the untested locations based on the correlations between tested and untested locations.

In practice, SORS first learns what locations should be tested and what reconstruction weights should be used to produce precise VFs. Using a training dataset of VFs (see next sections) SORS learns the testing location ordering and the corresponding reconstruction weights. This is achieved by iteratively solving a least squares problem, in a greedy and sequential fashion (see Kucur and Sznitman[29] for more details). In practice, this yields for every untested location a distinct set of weights associated to tested locations. Figure 1 illustrates these weights for three different locations (shown with pink star) after 20, 36, and 59 location are tested. As more locations are tested, we observe that most of the weight mass is focused near the location to reconstruct, indicating that local neighbors provide most of the information. In cases where few locations are tested, however, we see that the algorithm establishes weight magnitudes that do not necessarily depend on the distance between the locations, but rather their correlative nature defined by the least squares optimization.

Once the learning phase is complete, a perimetry examination can be performed on new subjects by testing locations in the order learned during the training phase and reconstructing the entire VF after each new tested location. This involves computing the threshold value of the entire VF by
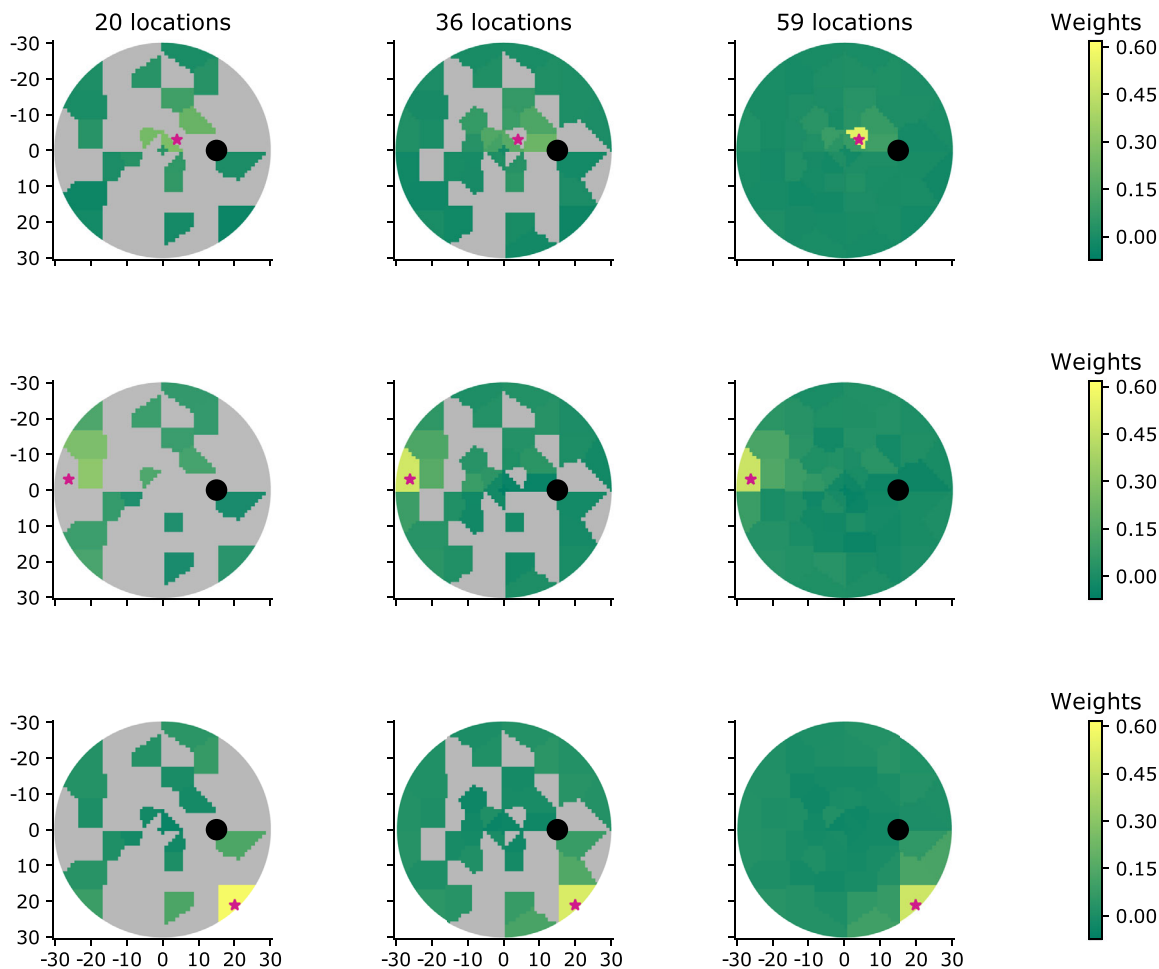
Figure 1. Color-coded weight maps corresponding to different number of tested locations. Each row corresponds to one location (*pink*) for which the reconstruction weights are shown. Columns correspond to the learned weight maps for 20, 36, and 59 tested locations. Untested locations are locations shown in *gray*.

multiplying the weight values and the measured thresholds at tested locations. Note that all VF locations, including measured locations, are thus weighted using the learned weights.

Implementation Details

SORS was implemented using the Open Perimetry Interface (OPI).[30] Both DS and SORS were run on the same Octopus 900 perimeter. We used the DS strategy provided by the manufacturer and available in the official software EyeSuite. DS measures each visual field location using adaptive staircasing and stops at the first response reversal.[10,15] It also interpolates the intermediary estimations to adjust starting STs for the next query.

To learn the SORS reconstruction weights, we used a training dataset of visual fields from the Bern University Hospital. This dataset consisted of 1168 G pattern VFs from glaucoma and normal healthy patients (potentially contained patients with other diseases), all acquired with the normal strategy.[22] The mean MD in the training dataset was of 4.88 (standard deviation [SD] = 5.87, min.-max. range [−7.82, 27.11]) and the mean sLV of 3.86 (SD = 2.33, min.-max. range [1.22, 12.61]).

As in DS, our SORS implementation used the dynamic staircasing method to measure individual locations. To account for the reduced precision of dynamic staircasing compared to the 4-2 staircasing of the normal strategy, we simulated each VF in the training dataset 10 times with dynamic staircasing using OPI.[30] This resulted in 10 noisy versions of a correct VF (we assume that the VFs acquired with normal strategy are the true VFs), leading to a noisy training dataset that was subsequently

used to learn the SORS locations and reconstruction weights.

SORS uses the same staircasing scheme and "one reversal" stopping criterion as in DS. Testing of locations are then performed in the order found during the training phase. Each time a location is completely queried, an intermediary reconstruction of the VF (i.e., estimation for the VF) is made. The starting stimulus intensity of a new location is set to the estimated value by the previous VF estimation.

In its original version, SORS queried each location until one reversal and followed the order of testing found during the training phase. However, this did not distribute the spatial attention of subjects, which has been shown to result in increased sensitivity measurements.[31–33] Instead, we adapted SORS such that a pool of four locations is randomly queried and when one location is completely queried, a next location from the testing order is included into the pool replacing the finished location. For the first four locations in the testing pool, the starting stimulus intensity was set to the normative value from the age-matched population and for the upcoming locations, they were set to the value from the previous VF reconstruction as aforementioned.

For both methods, standard background luminance of 31.4 asb (10 cd/m^2) was used, and maximum stimulus luminance was 4000 asb following Octopus 900 standards. Moreover, stimulus size was set to Goldman size III, and a white-on-white stimulus type was used. Stimulus duration was 100 ms, and the response window was 2000 ms. False-positive and -negative catch trials were randomly presented to the patient, representing up to 10% (each 5%) of the total stimuli presentations. Fixation of subjects was manually monitored, and subjects were encouraged to keep their gaze fixed at the target during the examination. In case of fixation loss, the test was paused until fixation was regained. Patients were also able to stop and take breaks during both the DS and SORS examinations.

To better assess the SORS performance at every reconstruction step, SORS measured all of the 59 test locations and the intermediary VF estimates were stored for each VF. Hence, the performance evolution with respect to the number of tested locations could be evaluated after the examination and the effective number of test locations to be queried could be properly determined.

Data Analysis/Statistics

To evaluate the correlation of DS and SORS, we computed *correlation* plots of thresholds of individual locations as well as MD values measured/estimated by DS and SORS. The correlation coefficient, *r*, which measures the goodness of the fit, is provided alongside as well. Specifically, *r* takes values between −1 and 1, where 1/−1 indicates strong positive/negative relationships, whereas values closer to 0 suggests weak or no relationship.

Estimation bias was also evaluated for both methods by means of histograms of differences between thresholds and MDs measured by DS and SORS estimates. The mean of the distributions is ideally 0 for an unbiased method.

Mean absolute error (MAE) performance of SORS in threshold estimation is computed as follows,

$$MAE = \frac{1}{N} \frac{1}{L} \sum_{n=1}^{N} \sum_{l=1}^{L} \left| VF_{n,l}^{DS} - VF_{n,l}^{SORS} \right|, \quad (1)$$

where $VF_{n,l}$ is the ST at the lth test location of the VF belonging to n^{th} patient. We show MAE performances with respect to the number of locations tested by SORS or the corresponding number of presentations.

Examination duration was compared across both methods as well as the number of tested locations by SORS. To further assess the time performances, we provide *total number of stimuli presentations* for SORS results with 20, 36, and 59 (all) compared to the DS results with 59 stimuli. *Average number of presented stimuli per location* was also computed to evaluate the gain in terms of time.

Test-retest variability analysis was performed for DS and SORS on a sub-population of subjects. This analysis has two purposes: (1) Comparison of test-retest variability of both techniques; (2) Assessment of the noninferiority of SORS compared to DS. For this purpose, histograms of differences and *Bland-Altman plots* were used to show whether the difference between DS and SORS measurements are within the range of DS test-retest variability and to inspect the SORS range of test-retest variability.

Qualitative VF examples were also provided to compare resulting VF estimates with DS measured VFs. *Error performance with respect to a gradient measure* was additionally presented for a deeper understanding of the SORS performance on estimating the isolated defects.
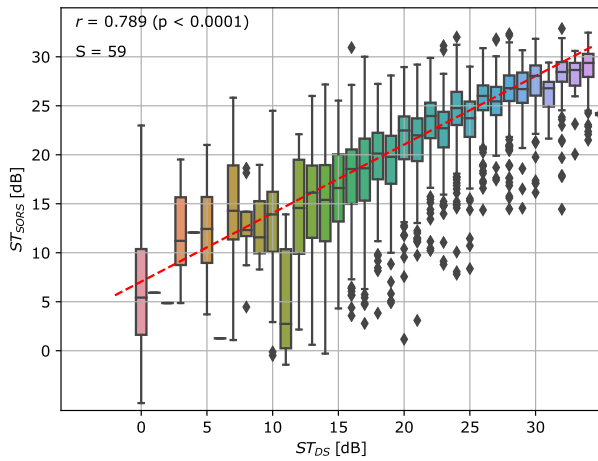
Results

Figure 2 shows the individual threshold correlations, as well as the correlation between measured MDs from DS and SORS acquisition after testing 20,
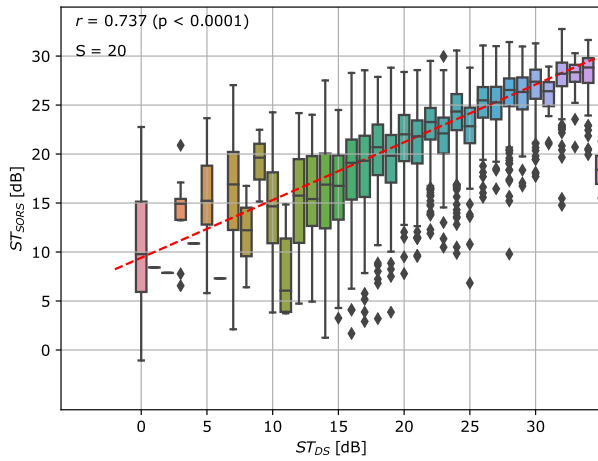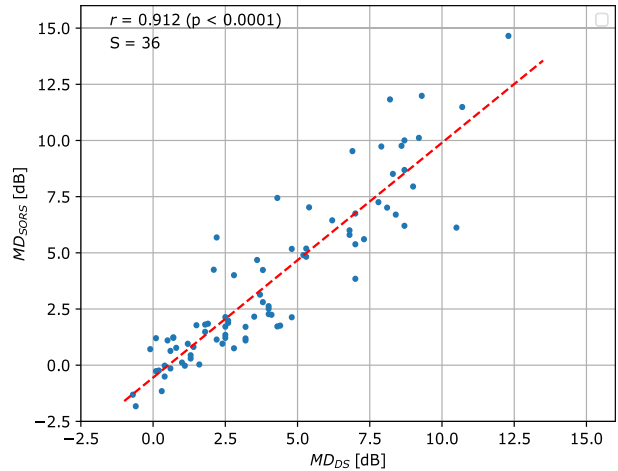
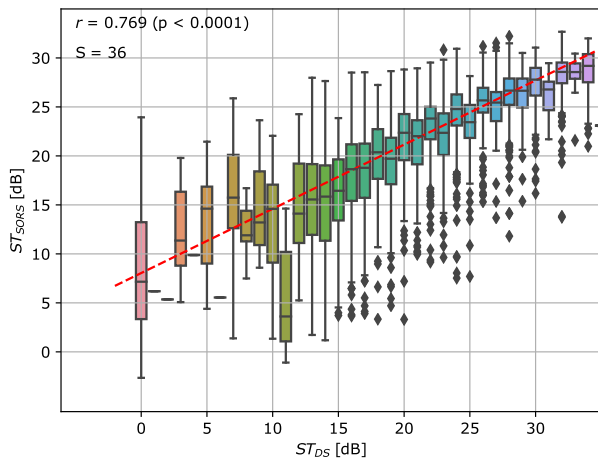(a) ST correlation (20 locations tested)
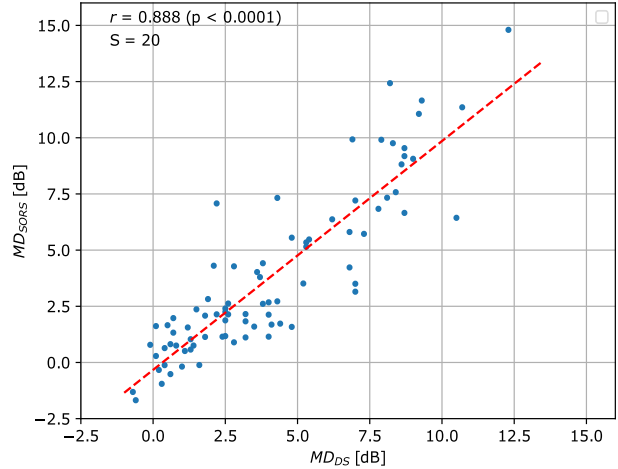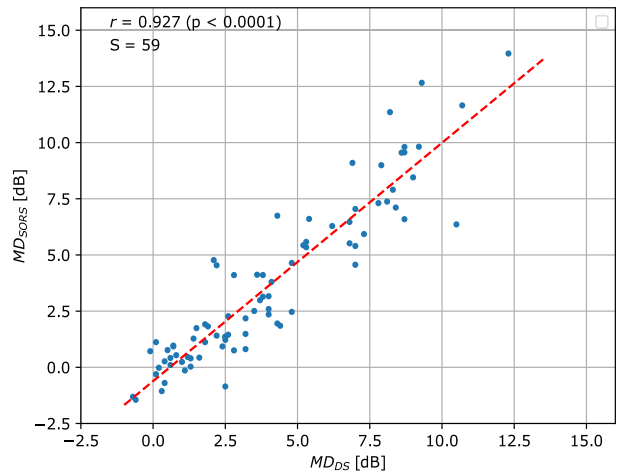
(b) MD correlation (20 locations tested)

(c) ST correlation (36 locations tested)

(d) MD correlation (36 locations tested)

(e) ST correlation (59 locations tested)

(f) MD correlation (59 locations tested)

Figure 2. Correlations between individual thresholds and MDs measured by DS and estimated by SORS testing 20 (a, b), 36 (c, d) and all (59) locations (e, f). The number of locations tested by SORS (i.e., S) and the *r* values with the corresponding *P* values are given on each plot. *Red dotted line* corresponds to the best fit line of the data points.
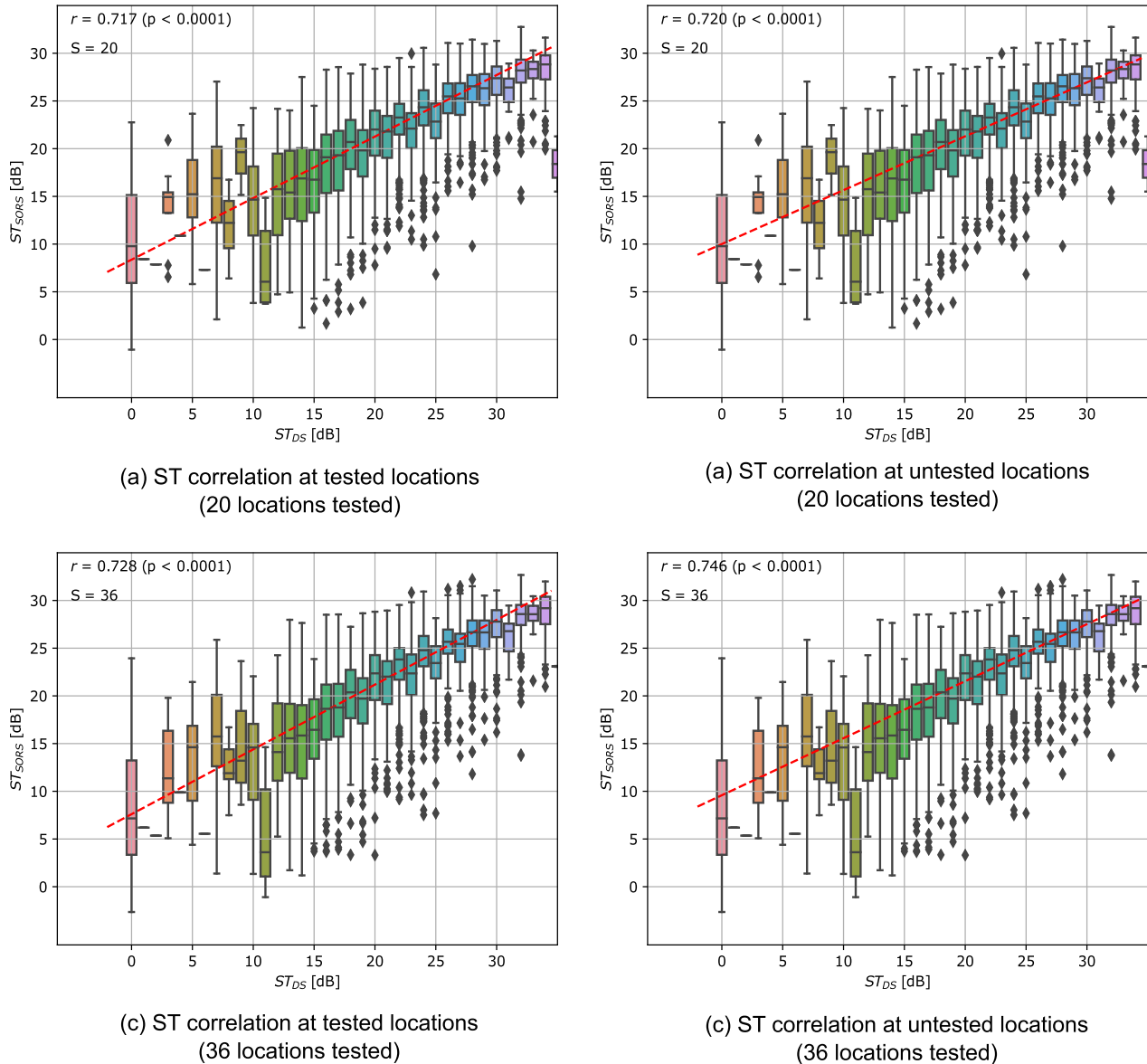
Figure 3. Correlations between individual thresholds acquired by both methods for tested (a, c) and untested locations (b, d) with 20 (a, b) and 36 (c, d) SORS tested locations. The number of locations tested by SORS (i.e., S) and the *r* values with the corresponding *P* values are given on each plot. *Red dotted line* corresponds to the best fit line of the data points.

36, and all (59) locations. *r* values for correlation in threshold measurements are 0.737, 0.746 and 0.789 for SORS testing 20, 36, and 59 locations, respectively (Wald test, $p < 0.0001$). *r* values for correlations in MD measurements are 0.888, 0.912, 0.927 for 20, 36 and 59 locations, respectively (Wald test, $p < 0.0001$). Additionally, we present the correlations for tested and untested locations separately for testing 20 and 36 locations in Figure 3.

Figure 4 shows the difference distributions in individual ST values from SORS and DS on healthy and glaucoma subgroups. In general, SORS estimated on average the healthy group thresholds slightly higher than DS when testing 20, 36 and 59 VF locations. Conversely, SORS marginally overestimated thresholds in the glaucoma group, but with mean differences below -0.1 dB across the evaluated tested location amounts. Similarly, Figure 5 shows the MD estimation bias: MDs estimated by SORS were lower than those measured by DS for both healthy and glaucoma groups with values below 1 dB across evaluated test amounts.
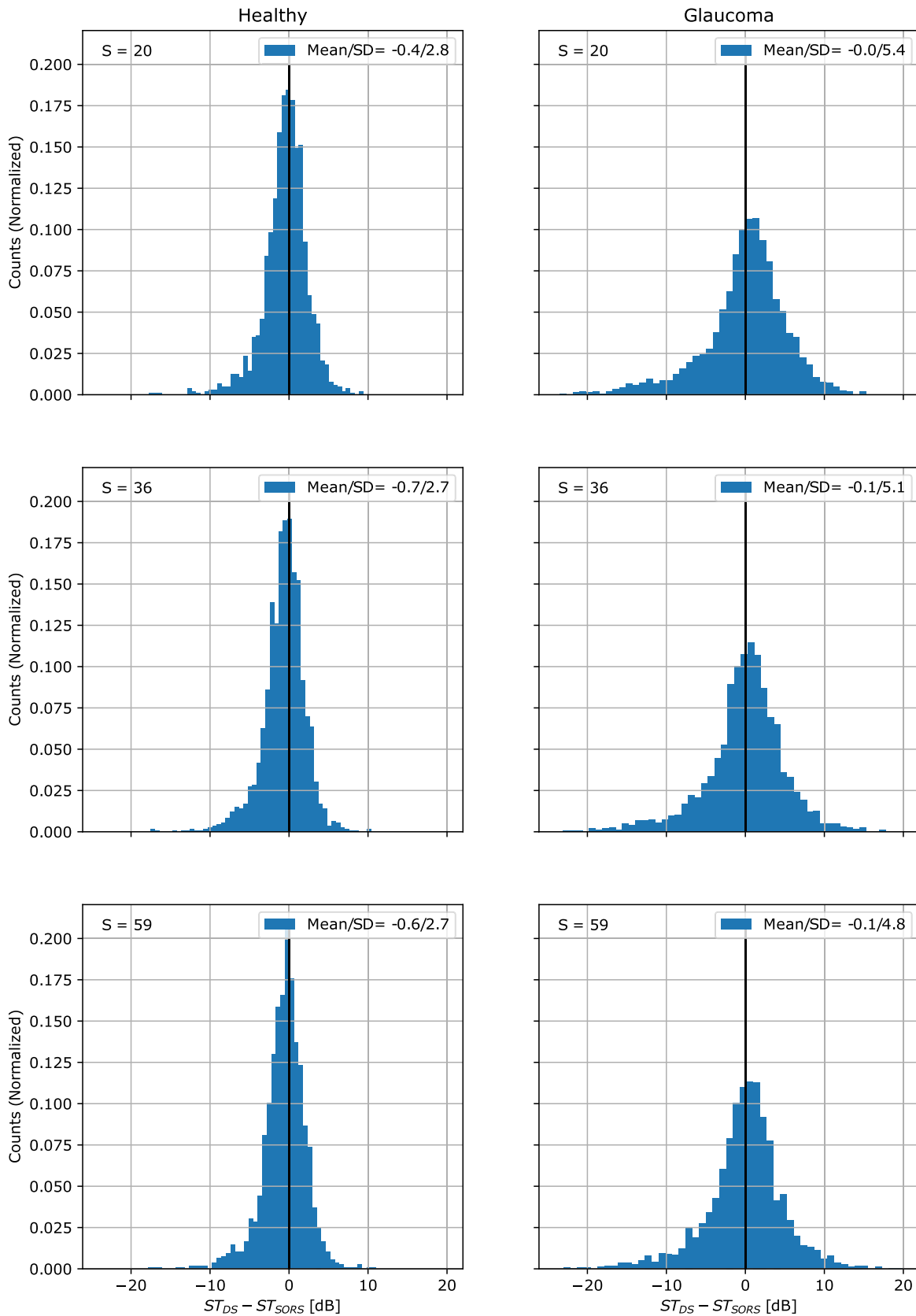
Figure 4. Estimation bias on assessing individual thresholds when SORS testing 20 (*top row*), 36 (*middle row*), and 59 (*bottom row*) for the healthy (*left column*) and for the glaucoma (*right column*) groups. Mean and SD, as well as the number of tested locations (i.e., S), are given on each plot.
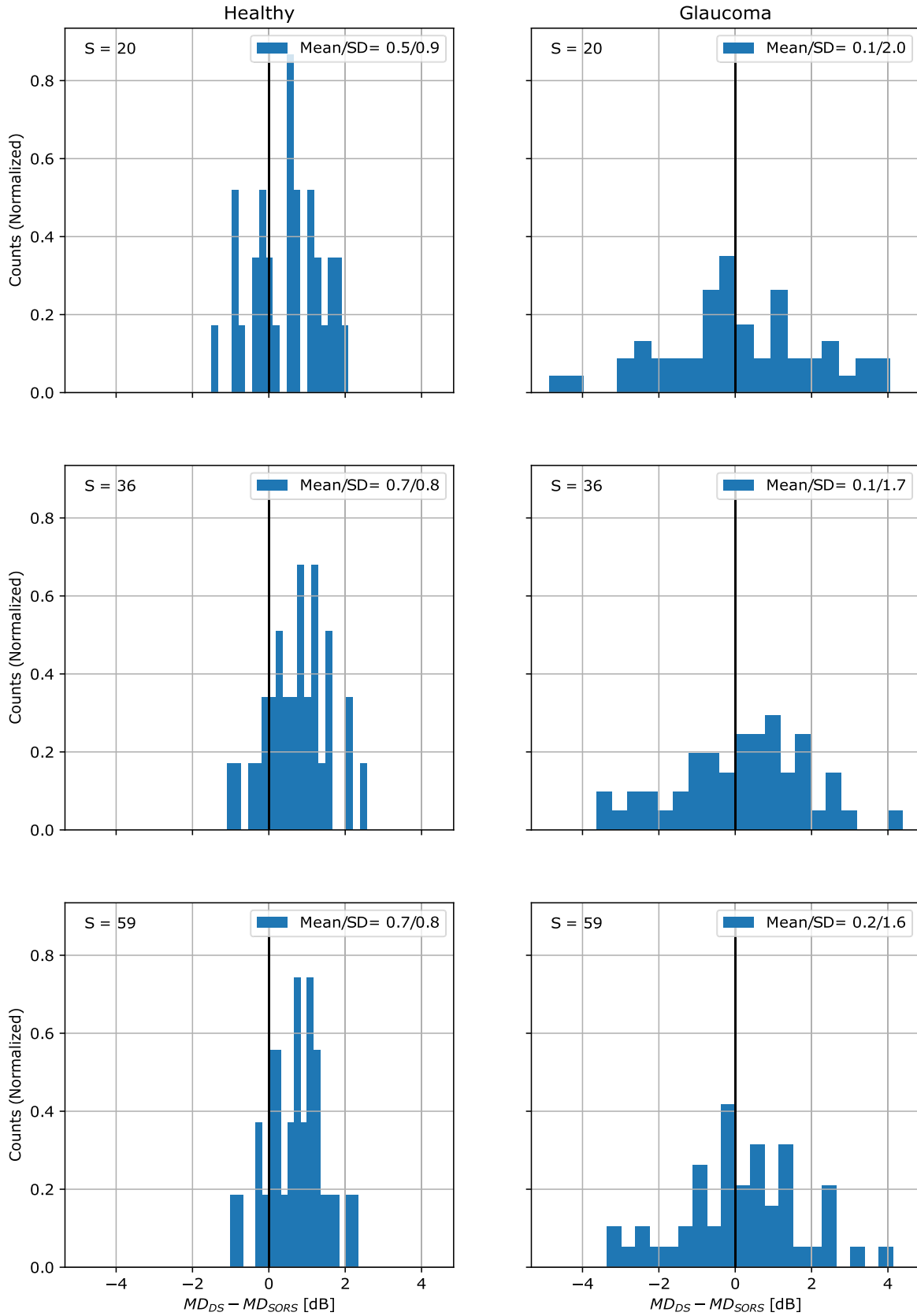
Figure 5. Estimation bias on MD measurement when SORS testing 20 (*top row*), 36 (*middle row*), 59 (*bottom row*) for the healthy (*left column*) and for the glaucoma (*right column*) groups. Mean and standard deviations (SD) as well as the number of tested locations (i.e., S) are given on each plot.

Figure 6 depicts the relation between patient age and MD estimation bias, with a general increase in result variance with age. In addition, regardless of the number of VF tested locations, SORS estimated MD values lower than DS for the 80+ age range. To better assess any systematic error for particular threshold ranges, we show absolute errors with respect to the individual threshold values in Figure 7. Accordingly, the error was found to be higher at both extreme ends of the range.

Figure 8 shows the error performance of SORS with respect to the number of presentations used, with DS thresholds taken as gold standard. MAE for healthy subjects were on average lower than for glaucoma patients, regardless of how many VF locations were tested. Over both subgroups MAEs of 3.17 (SD = 1.30), 3.00 (SD = 1.15), 2.92 (SD = 1.09) were found when testing 20, 36, and 59 locations, respectively.

With respect to the test-retest performance, Figure 9 shows that estimation bias when observing all ST values ($N = 590$, 10 subjects each with 59 locations). For each case (DS vs. DS, SORS vs. DS, and SORS vs. SORS), the mean bias was under 1 dB and the standard deviation was no larger than 3.7 dB. Figure 10 shows Bland-Altman difference plots for all comparison combinations, with the middle line corresponding to the mean difference and the upper and lower lines corresponding to 95% limits of agreements.

Median duration values with confidence intervals for observed locations are provided for healthy and glaucoma group, as well as all patients in Table 2. Medians were found to be significantly different from each other (Kruskal-Wallis test, $P < 0.0001$). A similar comparison is made with respect to the total number of presentations presented by SORS and DS, where medians were found to be significantly different to one another (Kruskal Wallis test, $P < 0.0001$). When testing both methods with 59 locations (i.e., all locations), the mean number of stimuli presentations per location by SORS was significantly lower than DS (SORS, 2.36 [SD = 0.18]; DS, 2.48 [SD = 0.20]; Mann-Whitney rank test, $P < 0.0001$).

Qualitatively, Figures 11 and 12 illustrate three VF examples from healthy and glaucoma groups, respectively. For each example, SORS VF acquisitions after testing 20, 36, and 59 locations are given and compared to the corresponding DS acquisition. For reference, MAE values are shown for each acquisition.

Since glaucoma can manifest itself as isolated defects, i.e., a small region can worsen more sharply than its neighboring locations, it is crucial for a perimetry strategy to identify such local defect regions as accurately as possible. Therefore, to evaluate SORS performance at measuring isolated defects, Figure 13
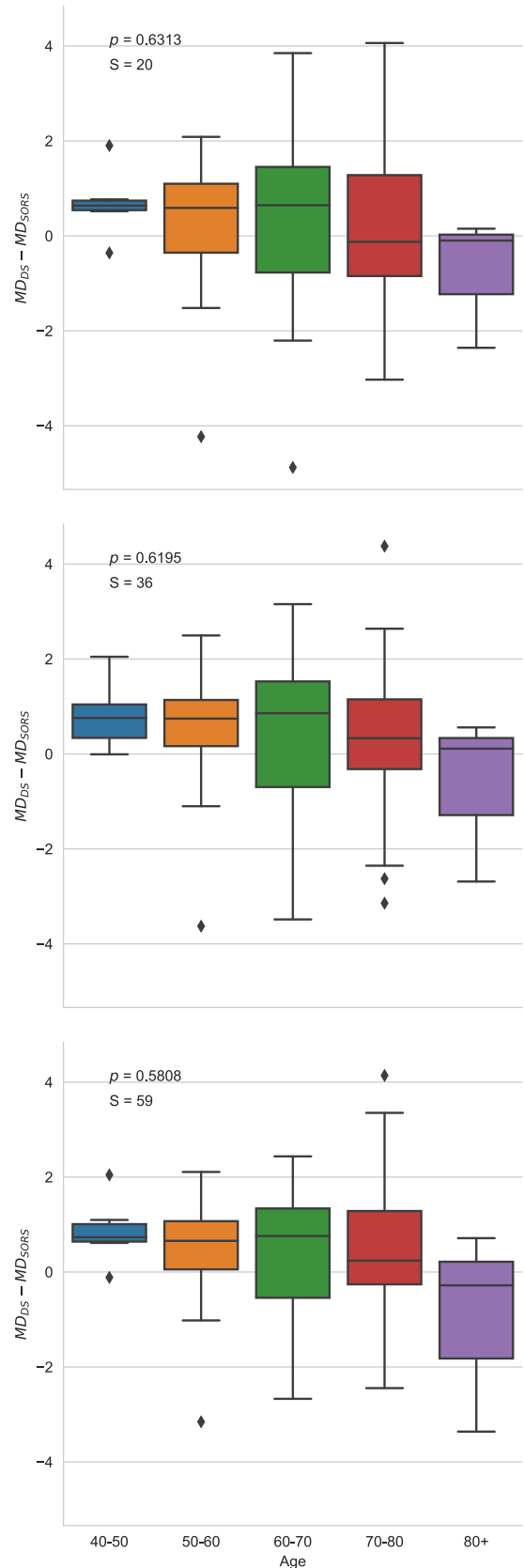

Figure 6. Estimation bias on MD estimation with respect to the age of the patient when the number of locations tested by SORS are $S = 20$ (*top*), $S = 36$ (*middle*), and $S = 59$ (*bottom*). For each subplot, Kruskal-Wallis test is performed, and *P* values are accordingly given.
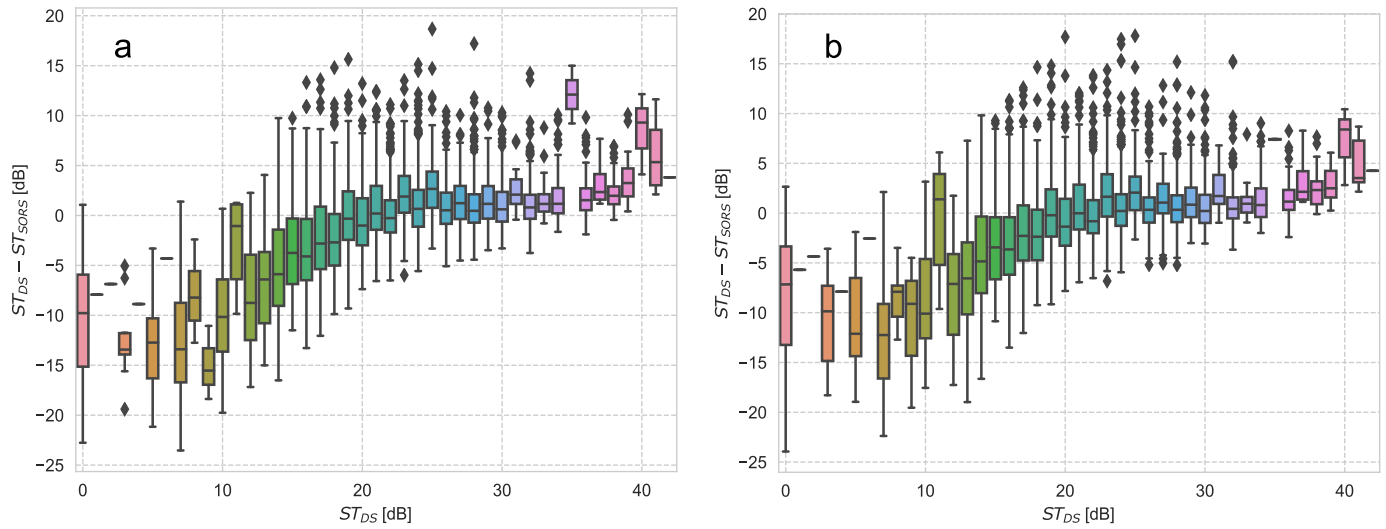
Figure 7. Differences in sensitivity thresholds as a function of DS measured sensitivity threshold: (a) 20 and (b) 36 SORS tested locations.
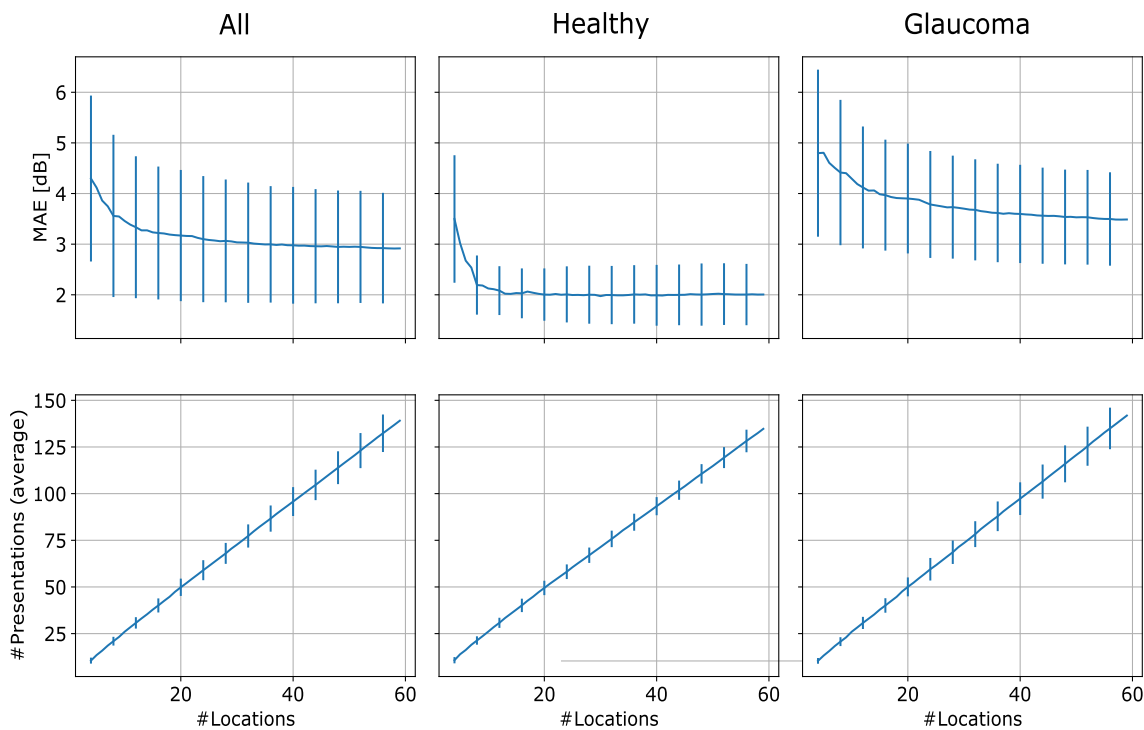

Figure 8. SORS performance in terms of accuracy and time. MAE and mean number of stimuli presentations are given with respect to the number of locations for all (*left column*), healthy (*middle column*), and glaucoma patients (*right column*). *Error bars* correspond to the SDs.

presents absolute errors with respect to a gradient measure $\Delta_l = \max\limits_{l_n \in N_l} |ST_l - ST_{l_n}|$ computed at the location l,[25,29] where N_l is the set of neighboring locations (within a radius of 9°). Δ_l corresponds to the highest difference between the ST at location l. A high Δ_l indicates the location l's ST is significantly different to its neighbor(s) (i.e., less homogeneous region) and is more challenging to predict. In Figure 13, we show the pooled errors with respect to Δ_l at all locations from
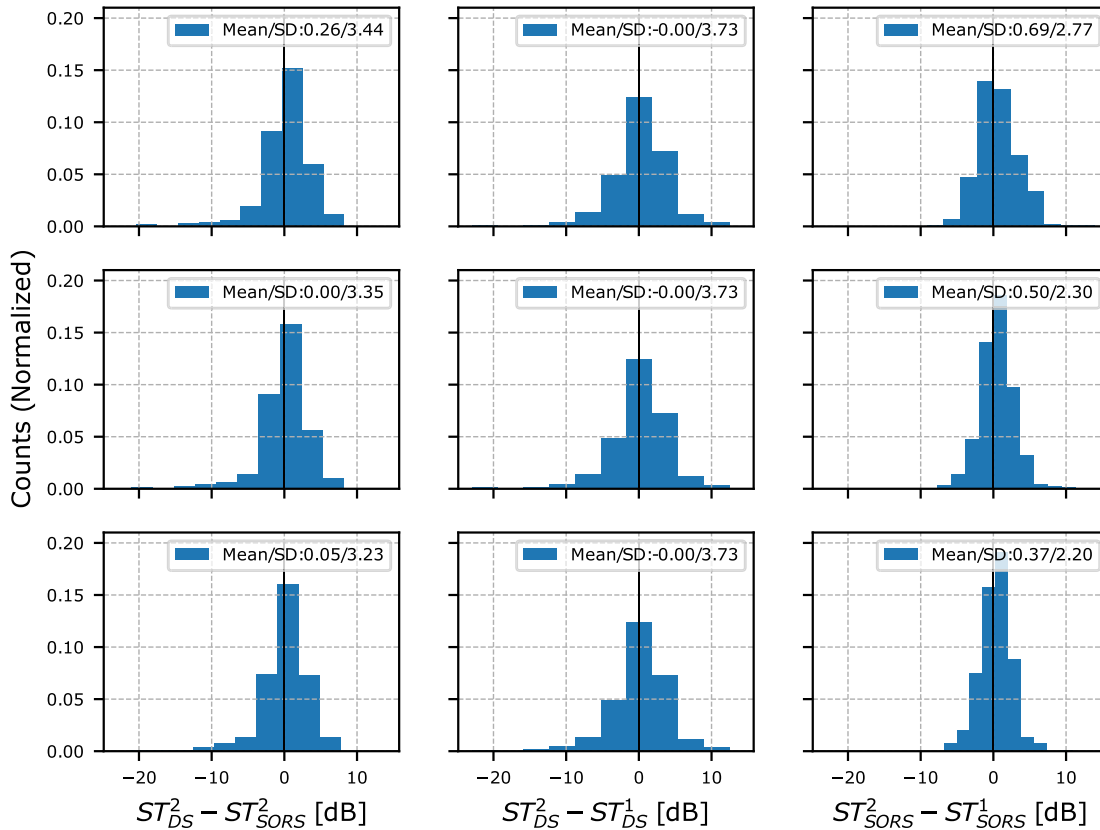
Figure 9. Histograms of threshold differences on test-retest sub-population. Columns correspond to SORS vs. DS, DS vs. DS, and SORS vs. SORS cases, respectively. Rows correspond to cases where SORS tested 20, 36, and 59 (all) locations, respectively. The counts are normalized. Mean and SDs are provided in the legends.

all healthy and glaucoma patients for the cases where 20 and 36 locations are tested by SORS.

Discussion

SORS achieved good correlations with DS in terms of individual threshold measurement even when testing only 20 locations with r values higher than 0.7. As expected, the correlation improved when increasing the number of SORS tested locations, and was consistent with the results shown in simulation.[29] Reported r values for the correlation of MD values between DS and SORS remained very high even when the number of test locations was reduced to a third (see Fig. 2). Moreover, Figure 3 shows that the correlation for tested and untested locations hardly differed from each other. This also demonstrates the strength of our proposed model that could learn to interpolate untested locations as successfully as for tested locations.

Considering the error distributions reported in Figure 4, the results indicate a slight overestimation of thresholds, and yielding lower MDs by SORS. This overestimation however is negligible with values of less than 1 dB. For glaucoma subjects though, the error distribution had larger variance which is consistent with reported VF studies.[20,34] As for the error distribution in MD estimation with respect to patient age (see Fig. 6), the SORS appears to have no significant dependency on the patient age (Kruskal-Wallis test, $P > 0.1$).

As given by Figure 7, SORS overestimated the STs less than 20 dB and slightly underestimates larger STs. The error got more remarkable for lower threshold range, especially for those less than 13 dB. One should, however, note that the bias in the low range was not significant because only patients with MD > 12 dB were included. Thus, not only were there few locations with low STs, the reliability of these is known to be low.[35] Therefore we cannot deduce a strong conclusion regarding SORS' estimation bias for the range below 15 dB. For STs greater than 15 dB, however, we conclude
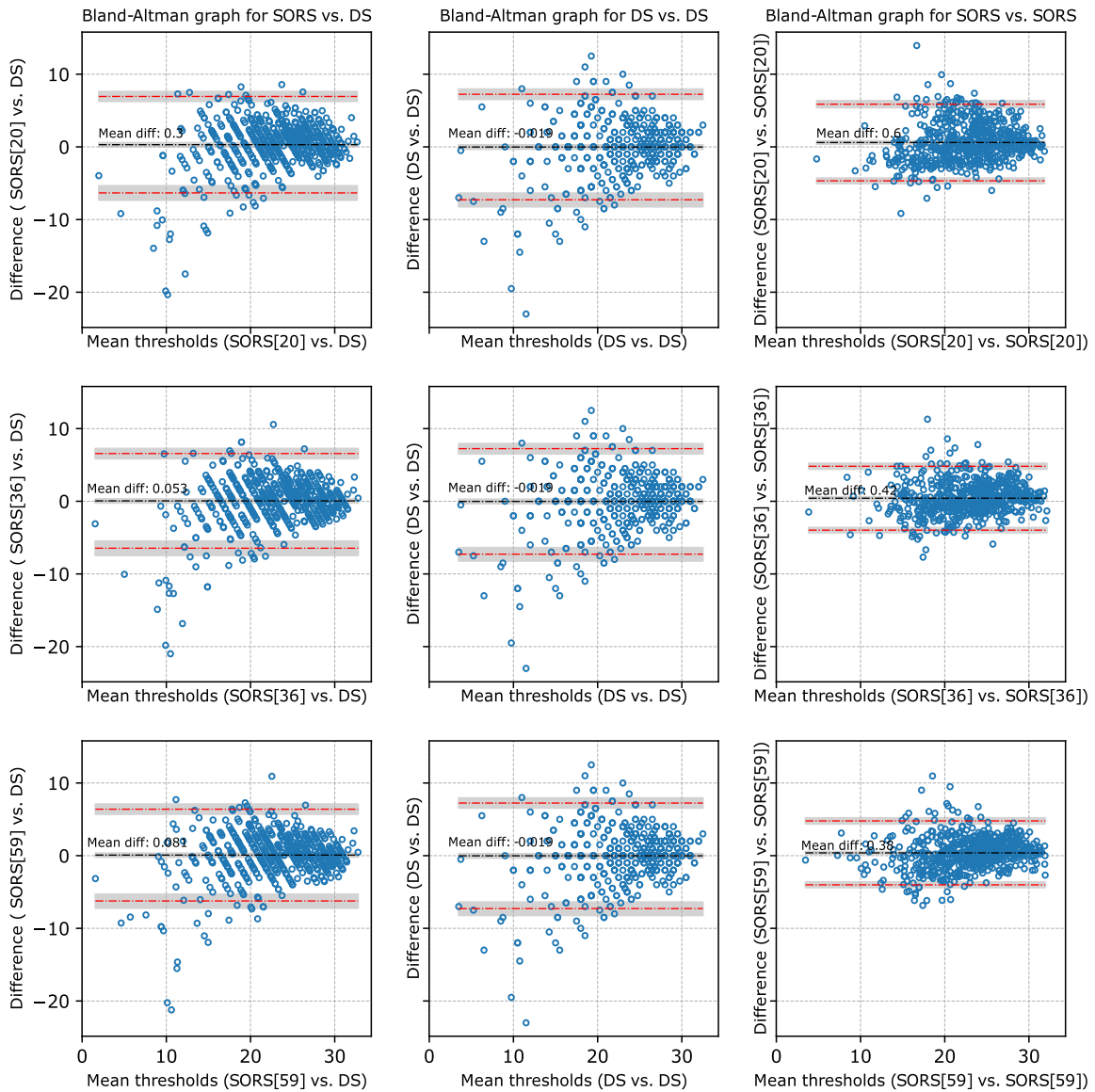
Figure 10. Bland-Altman agreement graphs for SORS vs. DS, DS vs. DS, and SORS vs. SORS for SORS testing 20, 36, and 59 (all) locations given row-wise. The *black dotted line* corresponds to the mean difference, and *red dotted lines* correspond to 95% limits of agreements (mean ± 1.96 SD). *Light-gray areas* are confidence intervals on the mean and limits of agreements.

that SORS slightly tended to overestimate above 20 dB and underestimated below, but that the error remained mainly within a [−5, 5] dB range.

Considering MAE performance for all patients, the MAE was less than 3.3 dB after 14 locations tested (MAE: 3.27 [SD = 1.38 dB], mean number of presentations: 35.37 [SD = 3.39]) and is 3.17 dB (SD = 1.30) when testing 20 locations (mean number of presentations being 52.06 [SD = 4.69]). For the healthy group, the MAE was of mean 2.02 dB (SD = 0.49 dB) with number of stimuli presentations of mean 40.10 (SD = 3.77) when observing 16 locations. This

is expected because healthy VFs are smooth, which can be easily inferred without querying many locations. The gain with additional tested locations was higher in the glaucoma group as the MAE performance continued to improve with the number of tested locations, although the improvement was still much higher for the first test locations. The MAE was under 4 dB for 20 locations (MAE = 3.90 [SD = 1.08]), with 50.04 (SD = 5.06) stimuli presentations and was less than 3.7 for 36 locations (MAE = 3.62 [SD = 0.97]), with 87.86 (SD = 7.95) stimuli presentations, which is within the acceptable range for VF accuracy.

Table 2. The Median Values [Min., Max.] for the Number of Stimuli Presentations and for the Examination Duration by SORS Testing 20, 36, and 59 (All) Locations Compared to DS Testing 59 (All) Locations

	No. of Tested Locations		
	20	36	59 (All)
No. of stimuli presentations			
All patients			
SORS	49 (48–49)	85 (84–87)	137 (135–139)
DS	—	—	145 (141–149)
Healthy			
SORS	49 (47–52)	85 (83–87)	133 (131–137)
DS	—	—	138 (134–140)
Glaucoma			
SORS	49(48–50)	86 (85–88)	138 (136–143)
DS	—	—	152 (148–155)
Examination duration (mins)			
All patients			
SORS	2.06 (2.01–2.12)	3.62 (3.56–3.67)	5.79 (5.71–5.89)
DS	—	—	6.07 (5.87–6.22)
Healthy			
SORS	2.06 (2.01–2.15)	3.57 (3.52–3.64)	5.63 (5.57–5.78)
DS	–	–	5.75 (5.57–5.83)
Glaucoma			
SORS	2.06 (2.02–2.12)	3.66 (3.56–3.75)	5.90 (5.74–6.06)
DS	—	—	6.33 (6.17–6.50)

95% confidence interval for each distribution is given in parentheses.

To further verify if the difference between DS and SORS measurements is within acceptable range, the performed test-retest experiments on a sub-group of 10 patients show some indication in this direction. By quantifying the test-retest variability, we compared the intrinsic variability of DS and SORS and evaluated whether the deviation of SORS thresholds from those of DS remained within test-retest variability range. As shown in Figure 9 and in Figure 10, test-retest variability of SORS was found to be smaller than DS as the standard deviations and limit of agreements for SORS test-retest differences were smaller than those for DS. This suggests that SORS measurements are reliable enough to decouple glaucomatous progression from measurement variation. Moreover, the variation between SORS and DS was similar (even smaller) than DS test-retest variability, as illustrated by the standard deviation and limits of agreement lines. This points to noninferiority of SORS compared to DS. These findings, however, should be taken with care because the test-retest analysis could only be performed on a subgroup of 10 patients. A larger test-retest clinical study is therefore needed for a definite conclusion.

Examination duration was two minutes with SORS testing 20 locations, which is one third of the examination duration of DS (see Table 2). This is significant for the healthy group because SORS performed very well testing only 16 locations (MAE ≈ 2 dB, see Fig. 8). This demonstrates that the examination can be reduced to almost under two minutes without noticeably affecting the VF accuracy for healthy patients. By testing 36 locations (MAE ≈ 3.7 dB, see Fig. 8), SORS led to more than 40% of time reduction compared to DS for both control and glaucoma patients. Interestingly, SORS testing all (59) locations had a 5% gain in the speed of the examination, even though they both share an adaptive staircasing scheme and stopping criterion (i.e., one reversal). This is likely due to SORS' method to select starting intensity values being closer to the real ST so that locations terminate faster with fewer stimuli per location. This was also observed with the total number of stimuli, as well as the average number of stimuli presented per location,
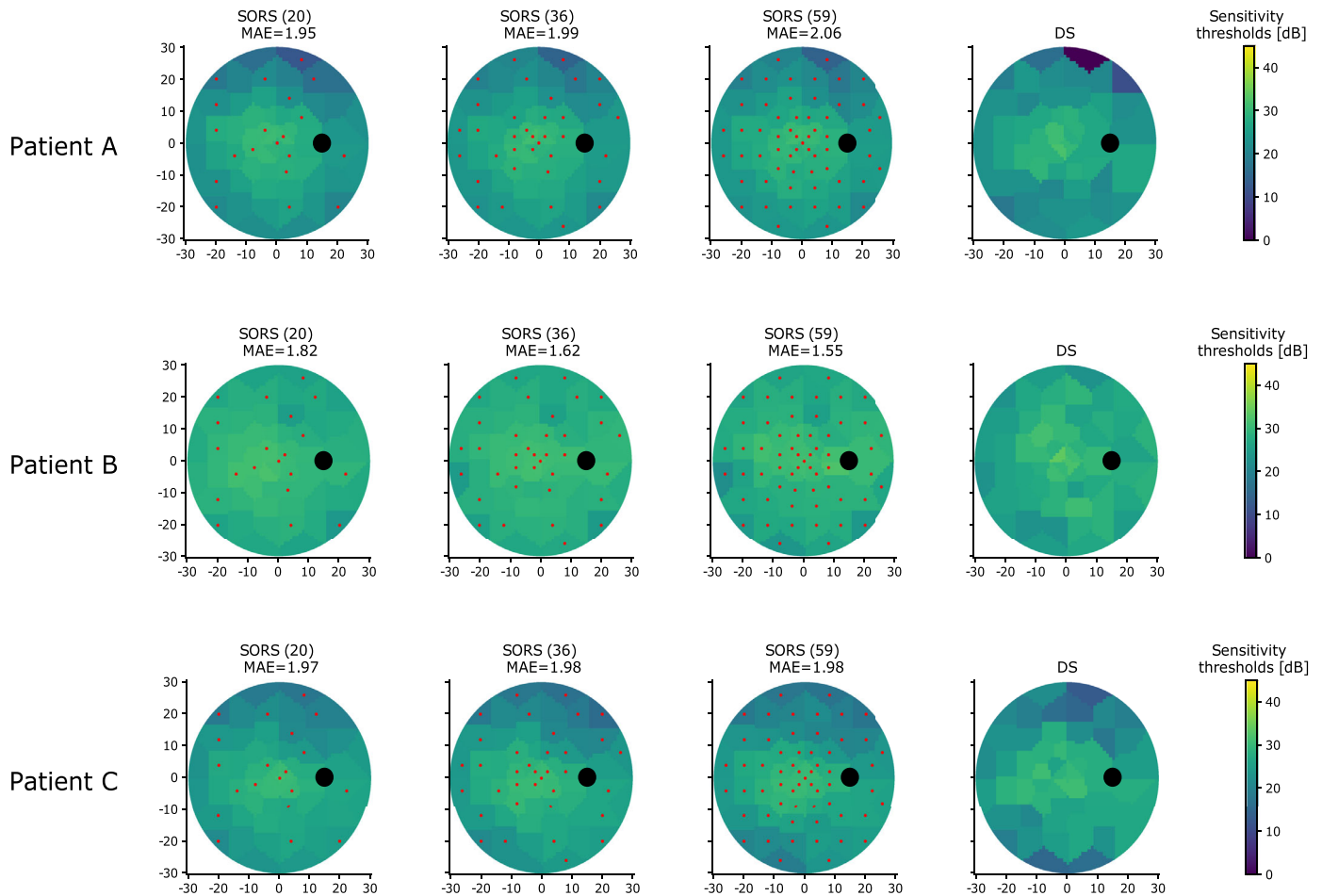
Figure 11. Three healthy visual fields examples. Each row shows acquisition output of a patient's VF for SORS with 20, 36, 59 locations tested as well as for DS. The respective method, the numbers of tested locations, and the MAEs are given above each image.

which are found to be significantly less for SORS than for DS.

In general, healthy VFs appeared smooth with relatively small acquisitions MAEs (see the qualitative examples shown in Fig. 11). Interestingly, SORS appeared to be performing well early on, when testing few locations (e.g., 20 locations) and little qualitatively and quantitatively improvements thereafter. This coincides with the stagnant error performance after testing 16 locations (see Fig. 8). In the glaucoma VF examples, the MAEs were higher because of heterogeneous defect patterns in each example in Figure 12. SORS could detect defect regions in those VFs, even with 20 locations, even though the VFs appeared smoother than those acquired with DS. This smoothness is a direct consequence of the linear model used to reconstruct missing VF locations and can lead to missing some isolated defects as in Patient D

in Figure 12. While being inevitable, the induced smoothness by SORS did not significantly affect the prediction of the isolated defects as the error of predicting highly gradient region was mostly less than 5 dB even for very high Δ_l values (e.g. 15 dB $< \Delta_l <$ 25 dB, see Fig. 13). This finding ensures that SORS, testing a fewer number of test locations, is able to capture, on average, isolated defects up to a reasonable precision, while individual counterexamples may occur.

A limitation of this study is the fact that we compared SORS to DS, which is clinically the most important but not a full threshold method that would provide more accurate ST measurements. A better option would be to compare SORS to a normal strategy,[22] alongside a test-retest protocol to estimate the variance in individual ST measurements. A follow-up study would be necessary to further clinically validate the SORS algorithm.
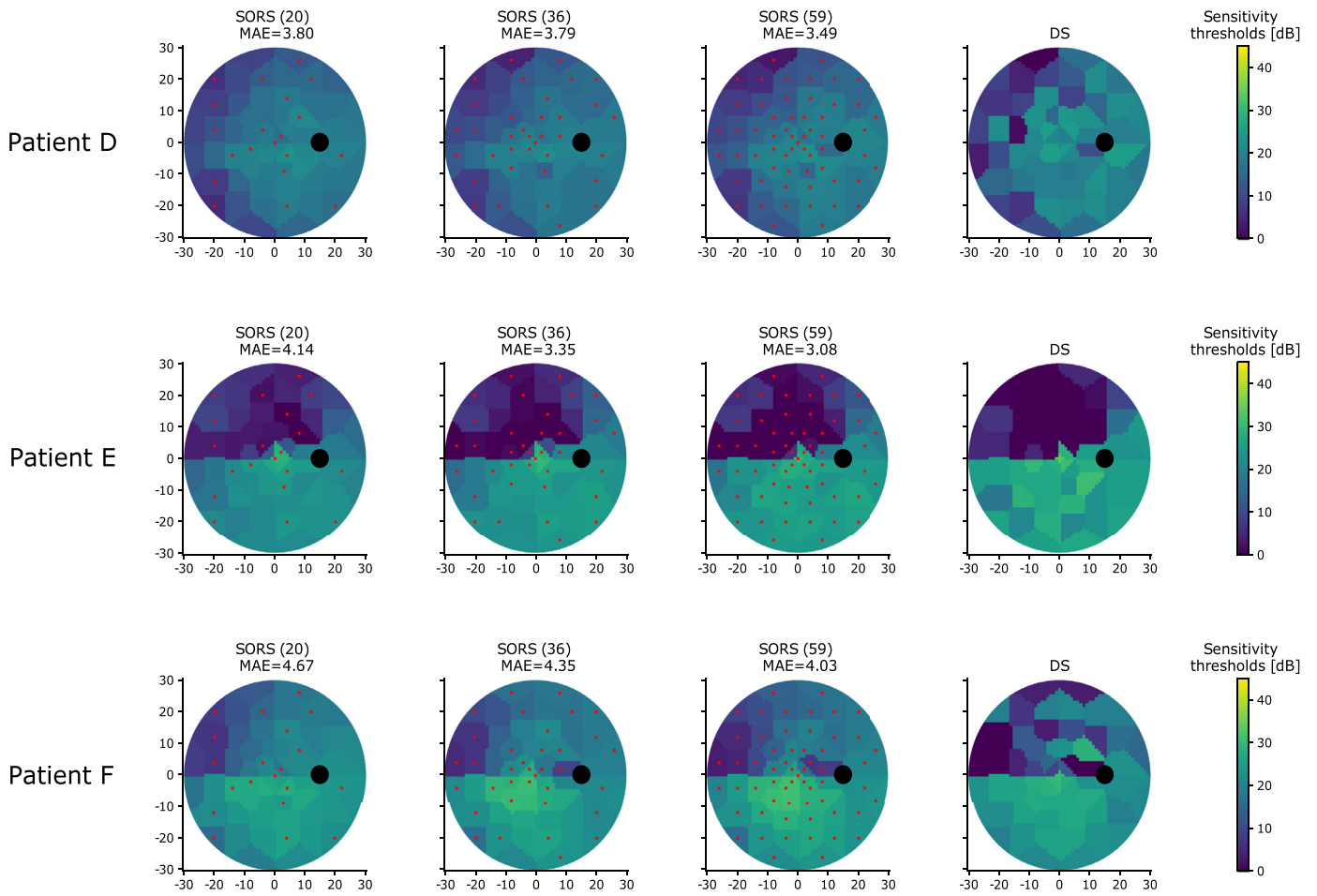
Figure 12. Three glaucomatous visual fields examples. Each row shows acquisition output of the same VF for SORS with 20, 36, 59 locations, as well as for DS. The respective method, the numbers of tested locations, and the MAEs are given above each image.
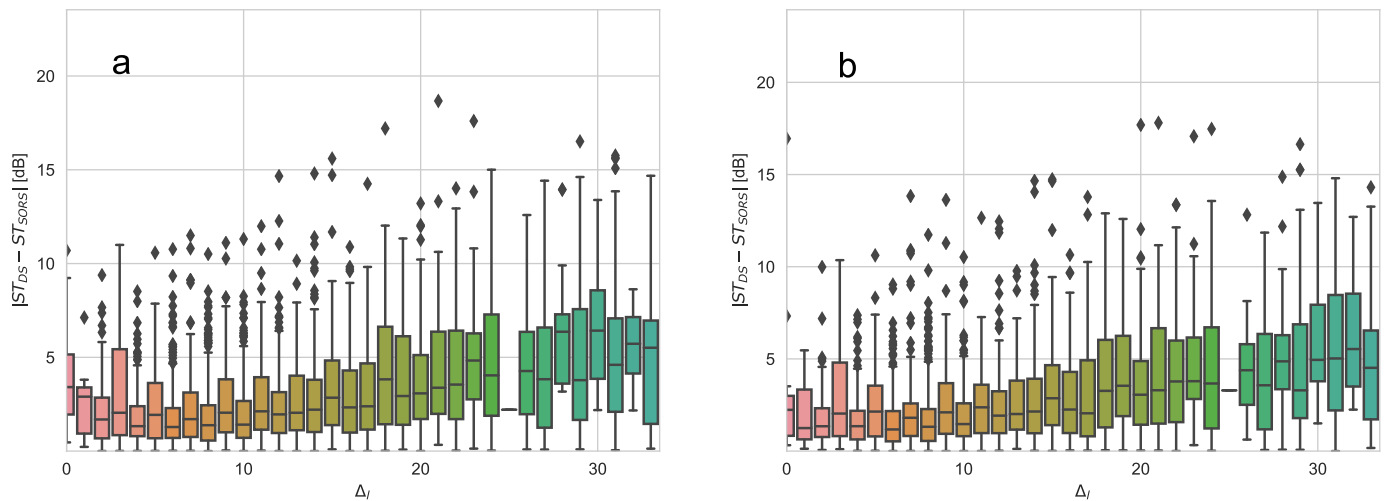

Figure 13. Absolute error with respect to gradient measure Δ_l for SORS testing 20 (a) and 36 (b) locations. Greater Δ_l value corresponds to a location having ST highly differing from its surrounding.

translational vision science & technology

Conclusion

SORS is a novel perimetry test strategy that queries fewer test locations than current conventional strategies used clinically in glaucoma care. It exploits the existing correlations between test locations and accordingly makes dynamic estimates of the VF during the examination using newly tested locations. This study has shown that SORS can achieve precise VF acquisitions, comparable to a conventional clinical technique, DS, in 40% to 70% less time. Shorter SORS examinations (i.e., testing fewer locations), seem to yield smoother VFs with potential subtle defects being missed. Nonetheless, SORS provides a good accuracy-speed tradeoff, whereby being a flexible and adaptable to any pattern and compatible with any staircasing scheme. SORS could therefore be potentially beneficial to use in routine clinical use for early to moderate glaucomatous patients.

Acknowledgments

Disclosure: **Ş.S. Kucur,** None; **S. Häckel,** None; **J. Stapelfeldt,** None; **J. Odermatt,** None; **M.E. Iliev,** None; **M. Abegg,** None; **R. Sznitman,** None; **R. Höhn,** None

# RS and RH contributed equally to this work.

References

1. Heijl A, Patella VM, Bengtsson B. *Effective Perimetry*. 4th ed. Dublin, CA: Zeiss Visual Field Primer; 2012.
2. Reitner A, Tittl M, Ergun E, Baradaran-Dilmaghani R. The efficient use of perimetry for neuro-ophthalmic diagnosis. *Br J Ophthalmol*. 1996;80:903–905.
3. Delgado MF, Nguyen NT, Cox TA, et al. Automated perimetry: A report by the american academy of ophthalmology. *Ophthalmology*. 2002;109(12):2362–2374.
4. Johnson CA, Wall M, Thompson HS. A history of perimetry and visual field testing. *Optom Vis Sci*. 2011;88(1):E8–E15.
5. Somlai J. Clinical importance of conventional and modern visual field tests in the topographical diagnostics of visual pathway disorders. In: Somlai J, Kovács T, editors. *Neuro-Ophthalmology*. Berlin: Springer; 2016:119–132.
6. Phu J, Khuu SK, Yapp M, Assaad N, Hennessy MP, Kalloniatis M. The value of visual field testing in the era of advanced imaging: clinical and psychophysical perspectives. *Clin Exp Optom*. 2017;100(4):313–332.
7. Wall M, Woodward KR, Brito CF. The effect of attention on conventional automated perimetry and luminance size threshold perimetry. *Invest Ophthalmol Vis Sci*. 2004;45:342.
8. Igarashi N, Matsuura M, Hashimoto Y, et al. Assessing visual fields in patients with retinitis pigmentosa using a novel microperimeter with eye tracking: the MP-3. *Plos One*. 2016;11:e0166666.
9. Johnson CA, Adams CW, Lewis RA. Fatigue effects in automated perimetry. *Appl Opt*. 1988;27:1030–1037.
10. Weber J, Klimaschka T. Test time and efficiency of the dynamic strategy in glaucoma perimetry. *Ger J Ophthalmol*. 1995;4:25–31.
11. King-Smith PE, Grigsby SS, Vingrys AJ, Benes SC, Supowit A. Efficient and unbiased modifications of the QUEST threshold method: theory, simulations, experimental evaluation and practical implementation. *Vis Res*. 1994;34:885–912.
12. Bengtsson B, Heijl A. Sita fast, a new rapid perimetric threshold test. description of methods and evaluation in patients with manifest and suspect glaucoma. *Acta Ophthalmol Scand*. 1998;76(4):431–437.
13. Tyrrell RA, Owens DA. A rapid technique to assess the resting states of the eyes and other threshold phenomena: the Modified Binary Search (MOBS). *Behav Res Methods Instruments Computers*. Mar 1988;20:137–141.
14. Anderson AJ, Johnson CA. Comparison of the ASA, MOBS, and ZEST threshold methods. *Vis Res*. 2006;46:2403–2411.
15. Weber J. [A new strategy for automated static perimetry]. *Fortschr Ophthalmol*. 1990;87(1):37–40.
16. Bengtsson B, Olsson J, Heijl A, Rootzn H. A new generation of algorithms for computerized threshold perimetry, SITA. *Acta Ophthalmol Scand*. 2009;75:368–375.
17. Bengtsson B, Heijl A, Olsson J. Evaluation of a new threshold visual field strategy, SITA, in normal

subjects. *Acta Ophthalmol Scand*. 1998;76:165–169.

18. King AJW, Taguri A, Wadood AC, Azuara-Blanco A. Comparison of two fast strategies, SITA Fast and TOP, for the assessment of visual fields in glaucoma patients. *Graefes Arch Clin Exp Ophthalmol*. 2002;240:481–487.

19. Heijl A, Patella VM, Chong LX, et al. A new SITA perimetric threshold testing algorithm: construction and a multicenter clinical study. *Am J Ophthalmol*. 2019;198:154–165.

20. Phu J, Khuu SK, Agar A, Kalloniatis M. Clinical evaluation of Swedish interactive thresholding algorithm–faster compared with Swedish interactive thresholding algorithm–standard in normal subjects, glaucoma suspects, and patients with glaucoma. *Am J Ophthalmol*. 2019;208:251–264.

21. Morales J, Weitzman ML, González de la Rosa M. Comparison between tendency-oriented perimetry (TOP) and octopus threshold perimetry. *Ophthalmology*. 2000;107(1):134–142.

22. Racette L, Fischer M, Bebie H, Holló G, Johnson C, Matsumoto C. Visual Field Digest. *A guide to perimetry and the Ocotpus perimeter*. 6th edition. Koniz, Switzerland: Haag-Streit AG; 2016.

23. De Tarso Ponte Pierre-Filho P, Schimiti RB, De Vasconcellos JPC, Costa VP. Sensitivity and specificity of frequency-doubling technology, tendency-oriented perimetry, SITA Standard and SITA Fast perimetry in perimetrically inexperienced individuals. *Acta Ophthalmol Scand*. 2006;84:345–350.

24. Turpin A, McKendrick AM, Johnson CA, Vingrys AJ. Properties of perimetric threshold estimates from full threshold, ZEST, and SITA-like strategies, as determined by computer simulation. *Invest Opthalmol Vis Sci*. 2003;44:4787.

25. Chong LX, McKendrick AM, Ganeshrao SB, Turpin A. Customized, automated stimulus location choice for assessment of visual field defects. *Invest Opthalmol Vis Sci*. 2014;55:3265.

26. Chong LX, Turpin A, McKendrick AM. Assessing the GOANNA visual field algorithm using artificial scotoma generation on human observers. *Transl Vis Sci Technol*. 2016;5:1.

27. Rubinstein NJ, McKendrick AM, Turpin A. Incorporating spatial models in visual field test procedures. *Transl Vis Sci Technol*. 2016;5:7.

28. Wild D, Kucur SS, Sznitman R. Spatial entropy pursuit for fast and accurate perimetry testing. *Invest Opthalmol Vis Sci*. 2017;58:3414.

29. Kucur SS, Sznitman R. Sequentially optimized reconstruction strategy: a meta-strategy for perimetry testing. *PLOS One*. 2017;12:e0185049.

30. Turpin A, Artes PH, McKendrick AM, et al. The open perimetry interface: an enabling tool for clinical visual psychophysics. *J Vision*. 2012;12:22–22.

31. Phu J, Kalloniatis M, Khuu SK. The effect of attentional cueing and spatial uncertainty in visual field testing. *Plos One*. 2016;11(3):e0150922.

32. Gould IC, Wolfgang BJ, Smith PL. Spatial uncertainty explains exogenous and endogenous attentional cuing effects in visual signal detection. *J Vis*. 2007;7(13):4–4.

33. Phu J, Kalloniatis M, Khuu SK. Reducing spatial uncertainty through attentional cueing improves contrast sensitivity in regions of the visual field with glaucomatous defects. *Transl Vis Sci Technol*. 2018;7:8–8.

34. Langerhorst CT, Carenini LL, Bakker D, Van Den Berg TJTP, De Bie-Raakman MAC. Comparison of SITA and dynamic strategies with the same examination grid. *Perimetry Update*. 1998;17–24.

35. Gardiner SK, Swanson WH, Demirel S. The effect of limiting the range of perimetric sensitivities on pointwise assessment of visual field progression in glaucoma. *Invest Ophthalmol Vis Sci*. 2016;57:288–294.