

# The Influence of Sampling Errors on Test–Retest Variability in Perimetry

Ted Maddess

**PURPOSE.** To determine whether visual fields measured by standard automated perimetry (SAP) can be distorted by higher-spatial-frequency image components and, in particular, whether test–retest variability of SAP fields can be explained by the combination of small scale fixational drift, small stimulus size, and coarse spatial sampling of the visual fields.

**METHODS.** Standard SAP test patterns have points 6° apart. The amplitude spectra of the perimeter's 10-2 fields (model 511 Humphrey Field Analyser [HFA]; Carl Zeiss Meditec, Inc., North Ryde, NSW, Australia) were assessed to see whether their finer grained sampling revealed spatial frequencies that could cause distortions of standard fields because of undersampling. Model visual fields were then constructed whose spectra were similar to the 10-2 fields, and test–retest variability was examined for Goldmann sizes III to VI stimuli and Gaussian fixational drift with standard deviations of 0.075° to 0.3°.

**RESULTS.** The 10-2 fields showed significant spatial frequency content up to 0.25 cyc/deg, three times the highest frequency that a 30-2 or 24-2 sample grid can resolve. As reported for SAP, test–retest variability increased with scotoma depth, and increasing the stimulus size from III to VI caused a reduction in test–retest variability, as did reduced fixation jitter.

**CONCLUSIONS.** With fixation drift half the size of that exhibited by good fixators, many of the features of SAP test–retest variability were reproduced. Reducing test–retest variability may therefore involve using large test stimuli that are blurry in appearance and that overlap somewhat when placed on the perimetric test grid. Overlap across the meridians should perhaps be avoided. (*Invest Ophthalmol Vis Sci.* 2011;52:1014–1022) DOI:10.1167/iovs.10-6014

Test–retest variability is a major problem for all forms of standard automated perimetry (SAP), including the achromatic<sup>1–3</sup> and blue/yellow programs<sup>4–5</sup> of the Humphrey Field Analyser (HFA; Carl Zeiss Meditec, North Ryde, NSW, Australia), the Octopus perimeter (Haag Streit, Köniz, Switzerland),<sup>6</sup> and the FDT (frequency doubling technology) perimeters.<sup>7,8</sup> It is often supposed that much of this variability arises from problems with the patient's vigilance and attention. Some evidence of this factor is that 5% to 10% of subjects fail quality controls, exhibiting high rates of false-positive responses and fixation losses.<sup>9–13</sup> However, some investigations indicate that

patient-related factors are not the main sources of variability in perimetry.<sup>14</sup> High test–retest variability is problematic because it interferes with our ability to track visual field progression. Thus, for example, it is recommended that new patients be given six field tests over the first 2 years to have an 80% chance of detecting a mean loss of –2 dB/y.<sup>15</sup> Is it possible that non-patient-related issues also affect the reliability of visual fields?

There is some evidence that stimulus size is a factor. The original frequency-doubling stimuli were very large, and the SD for test–retest was approximately 1 dB in patients and normal subjects.<sup>16</sup> The first FDT perimeter had 10-deg<sup>2</sup> stimuli and had lower test–retest variability than SAP<sup>7</sup>; however, halving the stimulus size for the 24-2 pattern of FDT2 (Humphrey Matrix; Carl Zeiss Meditec) increased variability.<sup>8</sup> On a similar note, Wall et al.<sup>2</sup> (*IOVS* 2009;50:E-Abstract 2239) have shown that, for achromatic SAP, variability decreases with increasing stimulus size, from Goldmann size III to size VI.

Part of the problem associated with small stimuli may be related to the very coarse grid of SAP test points, which are spaced at 6° intervals for standard HFA 24-2 and 30-2 test patterns. Thus, the standard size III stimuli test probes less than 0.41% of the test pattern area (0.146/6<sup>2</sup> · 100), which means that much can be missed. The problem may be much worse, however, if the sensitivity variations across a damaged visual field are not very smooth.

A grid of sample points spaced at 6° intervals can only accurately reconstruct spatial frequencies up to the Nyquist frequency<sup>17</sup>: ½(0.083) cyc/deg, or approximately 4 cycles across a 24-2 field. If higher spatial frequencies than 0.083 cyc/deg exist in the field and the sampling grid is orderly, then the higher frequencies will manifest themselves as lower frequencies in the band 0 to 0.083, those image components being said to be *aliased*.<sup>17,18</sup> Moire patterns are examples of the capacity of such aliased image components to distort a sampled image. Random sampling arrays produce related problems.<sup>19</sup> If aliased components distort the measured fields, then small variations in fixation would create new distortions on each retest. Microperimetry studies indicate that *good* fixators have deviations from fixation described by a distribution with a standard deviation of 0.6°.<sup>20,21</sup>

In this study, I first examined 10-2 fields, with their three times higher sampling density, to ask whether significant image components exist above the Nyquist rate for 24-2 and 30-2 fields. Such components could form distorting aliases within measured visual fields. The study then examines the effect on the measured test–retest variability of eye movements by modeling tiny deviations from the test grid between repeated field tests that mimic normal eye movements during fixation. The outcomes indicate that most of the results reported for SAP test–retest variability, including the reported changes with stimulus size, can largely be predicted by aliasing effects.

## METHODS

All data analysis and modeling was performed using commercial software (MatLab; The MathWorks, Natick, MA). The HFA 10-2 visual field

From the ARC (Australian Research Council) Centre of Excellence in Vision Science, Centre for Visual Sciences, Research School of Biology, Australian National University (ANU), Canberra, Australia.

Supported by ARC Centre of Excellence in Vision Science Grant CE0561903 and the ANU Centre for Visual Sciences.

Submitted for publication June 8, 2010; revised August 15, 2010; accepted September 26, 2010.

Disclosure: T. Maddess, Seeing Machines (F, C), P

Corresponding author: Ted Maddess, ACEVS and Centre for Visual Sciences, The Australian National University, Canberra ACT 02S00, Australia; ted.maddess@anu.edu.au.

results were de-identified data collected under the Australian National University's Human Experimentation Ethics Committee protocol 04/238 and conformed to the Declaration of Helsinki.

## Fourier Analysis

Visual field data from right eyes was flipped left to right to make them equivalent to left eye data and permit a common analysis. The threshold data from each field were then inserted into a matrix of 0s that was  $33 \times 33$  pixels in size. The amplitude spectrum of each field,  $Af_i$ , was then computed. That is to say, the 2D-Fourier transform of each of the fields was computed and the absolute value taken. The resulting spectrum characterizes the amplitude (in decibels) and the orientation of every spatial frequency in the visual field map. It is worth mentioning that any image, including a visual field map, can be exactly represented as a sum of sinusoidal gratings; hence, the spectrum identifies what spatial frequencies are found in the image from which it was derived.<sup>17</sup> As will be described in the text, these spectra were then averaged by groups,  $Af_g$ , defined by features such as the presence or absence of large nasal steps. Next amplitude spectra of flat fields with the same mean sensitivity as each group,  $Ff_g$ , were computed. The differences,  $Df_g = Af_g - Ff_g$ , were then taken to find the components of the  $Af_g$  that were not due to the windowing of the data caused by setting the points outside the field to 0. Samples, along lines through the various  $Df_g$ , so-called transects, were then taken, to examine Fourier image components that were oriented horizontally, vertically, or along the  $45^\circ$  diagonal of the fields. So, for example, the vertical meridian of the 2D spectrum describes the amplitudes (strengths) of all the spatial frequencies contained within the visual field with a vertically oriented spatial modulation.

## Model Fields

Part of the study required model visual field data that were generated on a finer spatial scale than a conventional SAP test pattern. Depending on the particular study, the spatial resolution of these model fields,  $\Delta_s$ , was  $1/10$  deg or  $1/30$  deg. The requirement was for a pattern with diffuse concentrations of damage mimicking scotoma as well as more normal parts of the field that were very smooth. A further constraint was that the amplitude spectra of these model visual fields be similar to those observed for the real 10-2 visual field data. This similarity was obtained by first initializing a matrix of resolution  $\Delta_s$  with random numbers of uniform distribution and then a large median filter, ranging from  $1.5^\circ$  to  $4.5^\circ$  on a side, was operated upon the random data. The resulting data were then scaled from 0 to 1 and the square root was taken. The resulting model visual fields,  $m_i$ , were then scaled between approximately 4 and  $-35$  dB, to imitate the range of SAP thresholds.

## Sampling

Sampling used disc shaped operators,  $o_i$ , the size of Goldmann sizes III, V, or VI stimuli. Note that, for a viewing distance of 300 mm, the radii of these stimuli in degrees are  $(0.216 \ 0.862 \ 1.724) = \text{atan}(\sqrt{2^z/\pi}, 300) \cdot 180/\pi$ , where  $z = (2 \ 6 \ 8)$ . The volume of the operator was set to 1. Sampling, then, involved locating the operator over the model fields,  $m_i$ , multiplying the points in the operator with those underlying them in  $m_i$ , and taking the sum. Since the volume of  $o_i = 1$  the final sum was the mean of the threshold data under the operator. For any given sampling of 24-2 points, the exact position of  $o_i$  was jittered by a circular Gaussian distance, with SD equal to values ranging between  $0.075^\circ$  and  $0.3^\circ$ , depending on the study. All the jitter sizes used were thus two to eight times smaller than the  $0.6^\circ$  SD reported for good fixators.<sup>20,21</sup> Strictly because the decibel values were averaged, this calculation corresponds to computing the geometric mean of the underlying linear visual field sensitivities. For completeness, studies were also performed in which the decibel thresholds were first transformed to linear sensitivities, the sample means within the stimuli were computed, and then the results were back transformed to decibel measures. Thus, in these cases, the arithmetic mean of the sensitivity,

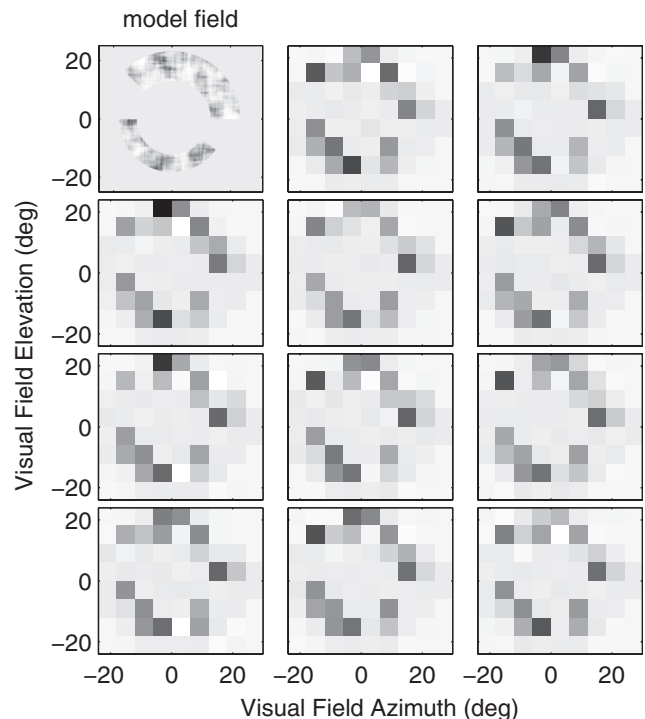
rather than the geometric mean, was computed. As expected, either method yielded very similar results.

## RESULTS

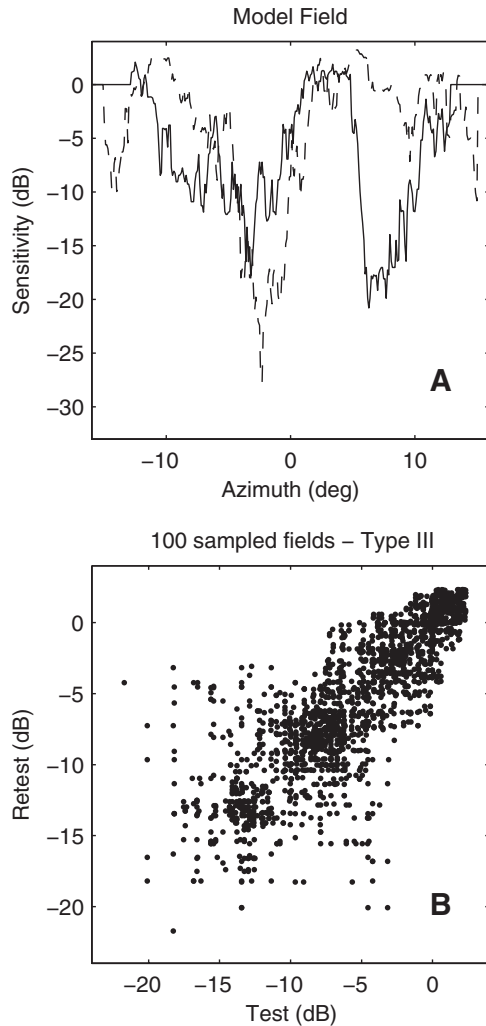
The potential problem is illustrated in Figure 1. At top left is a model visual field created as described in the Methods section. The field sensitivity data were windowed to have two arcuate scotomas. The other panels are the result of sampling the model field using a Goldmann size III stimulus at the standard test positions of a 24-2 field, but where those positions have been jittered to mimic small deviations from fixation. The jitter was described by a 2D Gaussian distribution with an SD of  $0.3^\circ$ . That is to say, the exact positions at which the field was sampled were displaced from the true 24-2 grid positions by distances described by a small circularly symmetric Gaussian probability function. Thus, 63.4% of the displacements were within  $0.3^\circ$  (1 SD) and 98.2% within  $0.6^\circ$  (2 SD) from the nominal 24-2 positions.

Figure 2A shows two horizontal transects through the upper and lower scotomas of the model field, to give an impression of the sensitivity fluctuations in the model. The test-retest variability from 100 samplings from the same model field is shown in Figure 2B. The variability increases for more damaged parts of the field, much as has been reported for repeated SAP field tests.<sup>1,2,8</sup>

The situation is perhaps worse than might be expected when considered in the context of sampling a 2D object with



**FIGURE 1.** *Top left:* a model visual field showing two arcuate scotomas containing regions of patchy damage. Horizontal transects through the upper and lower scotomas are shown in Figure 2A. The remaining images are the output of a sampling of the model field with 24-2 grid pattern plus positional errors described by a circularly symmetric Gaussian function of SD  $0.3^\circ$ . The median filter used to create the field was  $3.1^\circ$  on a side. It is noticeable that very dark and light parts of the sampled fields modulate quite markedly, in much the way that repeated visual field tests do. The gray scale for the sampled fields spans  $-20$  to  $2$  dB, for the model field  $-28$  to  $4$  dB—the difference in scale arising from the smoothing effects of the size III stimulus.



**FIGURE 2.** (A) Two horizontal transects through the model field at top left of Figure 1. *Solid trace*: the upper scotoma; *dashed trace*: the lower scotoma. (B) Test-retest measurements from 100 samples from the same field using a size III stimulus with a fixation jitter of 0.3° (SD). Each point represents a test-retest pair from some part of the field. Variability grows with scotoma depth, as is commonly reported for SAP fields with the use of a size III stimulus.

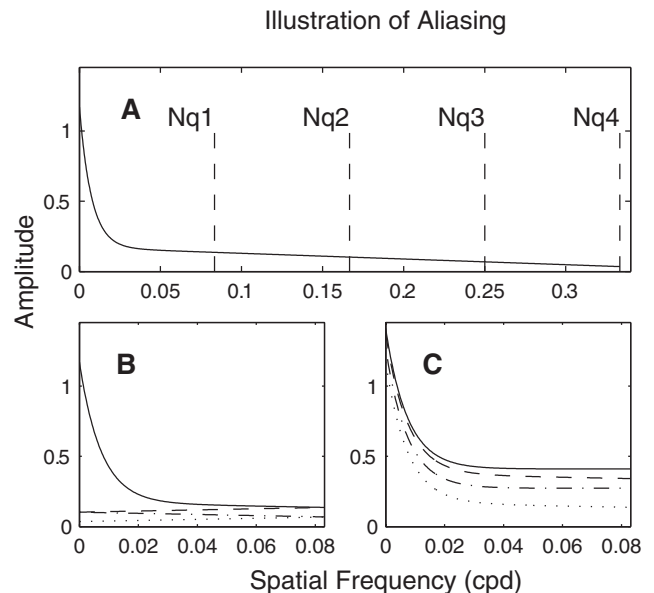
a substantially regular grid of sample points. For the 24-2 grid the separation is 6°. This defines a Nyquist sampling frequency,  $Nq_{24-2}$ , of  $1/(2 \cdot 6) = 0.083$  cyc/deg. That is to say, the highest spatial frequency that a 6° sampling array can accurately report on is a sinusoidal variation with a period of 12°, or approximately 4 cycles across a 24-2 field. For the square 24-2 sampling grid  $Nq$  is even lower along the 45° and 135° diagonal axes, being  $1/[2 \cdot \sqrt{(2 \cdot 6^2)}] = 0.059$  cyc/deg, a spatial frequency with a period of 17.0°. From the perspective of perimetry, even if the sample grid is not jittered, this poses a problem if the sampled field contains variations in sensitivity that modulate faster than 4 cycles per visual field diameter. Given the punctuate nature of glaucomatous visual field defects, more rapid variations than 0.083 cyc/deg across the field seem likely. If such image variations exist, then the problem of aliasing arises, where spatial frequencies higher than  $Nq$  masquerade as lower frequencies. The aliasing process can be understood as a folding back of the spectrum of the image about  $Nq_{24-2}$ , as illustrated in Figure 3.<sup>17,18</sup> The situation is somewhat more complex for slightly jittered sampling arrays, but the principle is the same: The sampled image will contain erroneous

modulations creating false impressions of the true visual field. These false impressions will be exacerbated if, on repeated visits, the sampling grid varies even by a small amount (Fig. 2A).

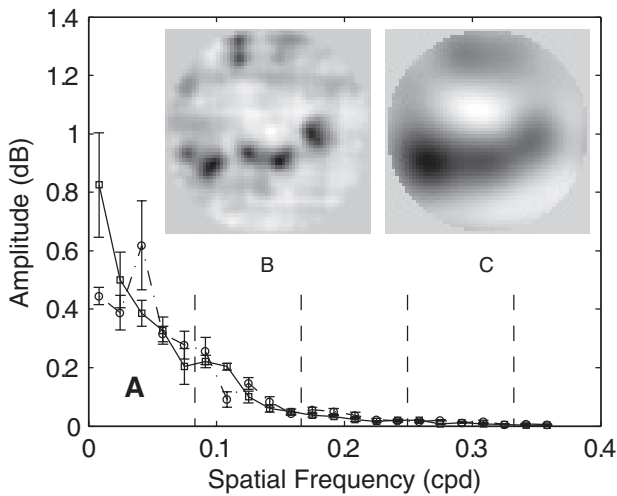
The question therefore arises: Do real visual fields contain any higher spatial frequencies than  $Nq_{24-2}$  that might present a problem, as illustrated in Figures 1 and 2A? A related issue is what a field would look like if it contained no information above  $Nq_{24-2}$ . A few papers report visual field data sampled at 1° intervals for either whole fields<sup>22,23</sup> or small parts of fields.<sup>24,25</sup> A beautiful example from Stürmer and Gloor<sup>22</sup> is shown in Figure 4, where an inferior arcuate scotoma is resolved into several islands of damage. Figure 4C shows the same data filtered to remove all content above 0.1 cyc/deg (slightly above  $Nq_{24-2}$ ), indicating what a field should look like if 6° sampling is dense enough to prevent aliasing. Figure 4B is much more in keeping with our expectations of the structure of a glaucomatous visual field than is Figure 4C.

Figure 4A shows data from the amplitude spectrum of 4B. The solid line is the mean of the six vertical transects on either side of the vertical midline of the 2D spectrum (error bars, 95% confidence limits). The dash-dot line is a similar average of the six transects bracketing the horizontal midline of the spectrum. It is clear that there is significant information up to at least 2.5 times the Nyquist rate. We also examined data from four 10-deg<sup>2</sup> fields sampled at 1° intervals from Westcott et al.<sup>24</sup> All those visual fields samples contained significant content above  $Nq_{24-2}$ .

To determine this content more quantitatively, HFA 10-2 fields were examined. For these fields,  $Nq$  is three times higher



**FIGURE 3.** (A) A putative amplitude spectrum of an image, such as a visual field sensitivity map, which is to be sampled by a 6° grid of points. The frequency axis extends to 4  $Nq_{24-2}$ , slightly more than the Nyquist frequency for a 10-2 field with a sampling period of 2° (i.e.,  $Nq_{10-2} = 0.25$  cyc/deg). Multiples of  $Nq_{24-2}$  are marked by vertical dashed lines labeled  $Nq_1$  to  $Nq_4$ . The aliasing effect seen within the band 0 to  $Nq_{24-2}$  is that the spatial frequency content of the image above  $Nq_{24-2}$  ( $Nq_1$  in A), is folded back into the range below  $Nq_{24-2}$ . (B) The folding back and forth of the spectrum. Thus, the higher frequencies masquerade as lower frequencies and are referred to as *aliased* frequencies. (C) The folded back frequencies add, the cumulative additions being shown as the different lines, with the final sum being the *solid line*. Thus, in practice a considerable part of an undersampled image can be made up of jumbled, erroneous projections of the true image.



**FIGURE 4.** (A) The means of six transects on either side of the vertical (solid) and horizontal (dash-dot) meridians of the 2D amplitude spectrum of the visual field of (B). (B) A visual field sampled by using a square grid with 1° separation. (C) The data of (B) filtered to have no content above 0.1 cyc/deg. If 6° sampling were adequate, then glaucomatous fields would look like this image. In practice, the more rapidly varying image of (B) is more in keeping with the understanding of the structure of visual fields. Vertical dashed lines: multiples of  $Nq_{24-2}$ , as in Figure 3A.

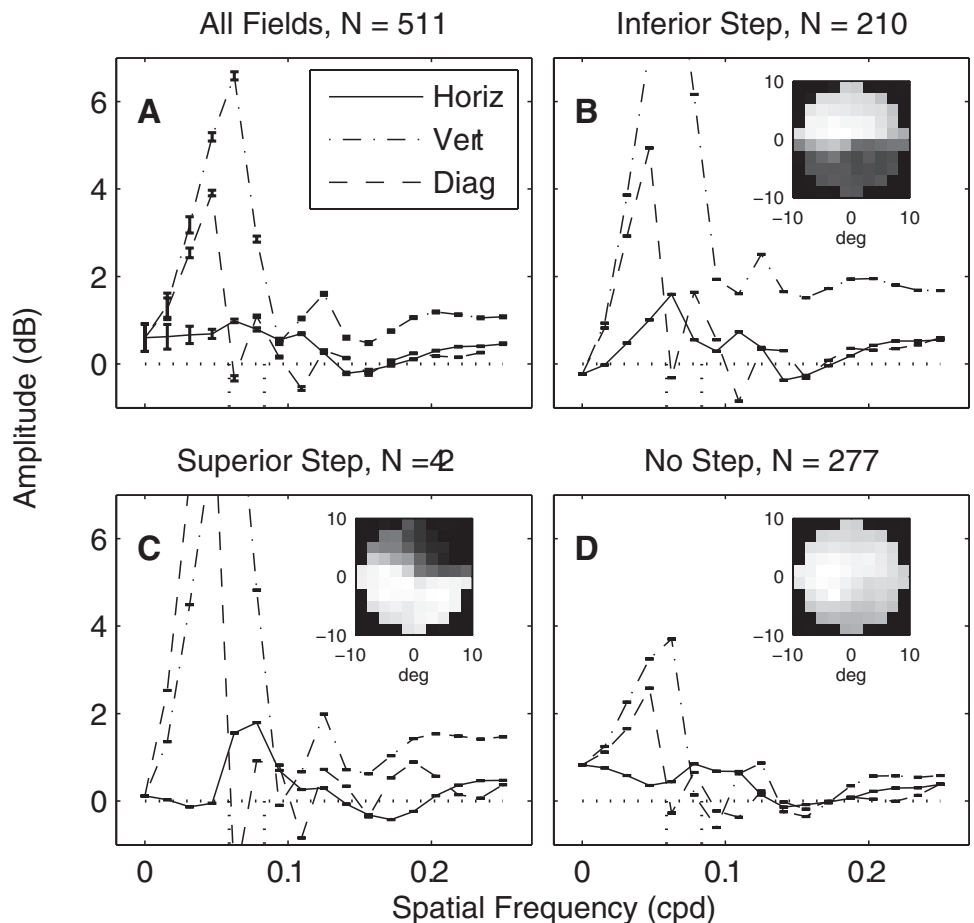
given the 2° separation of the 10-2 sample points. Figure 5 presents transects through the mean 2D spectra of different selections of the HFA fields. The three transects were the

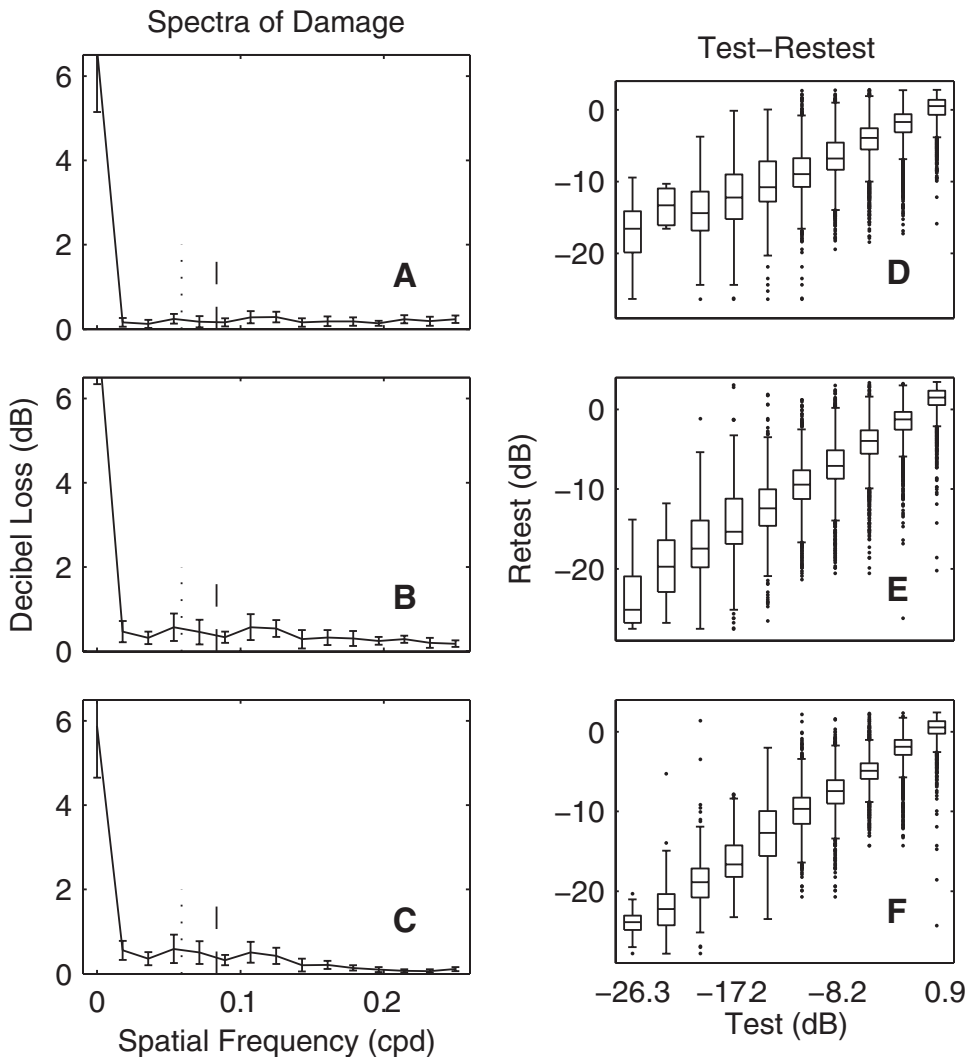
horizontal and vertical meridians of the spectra and also a transect along the 45° diagonal of the spectra (see the legend to Fig. 5A). Figure 5A shows the average of these transects and SE for all HFA 2D amplitude spectra. The spectrum of a normal, flat field was subtracted (see the Methods section), and thus the spectra in Figure 5 represent deviations from a normal flat field. The  $Nq_{24-2}$  for the horizontal, vertical, or diagonal sampling of the field by a 24-2 array (0.083 and 0.059 cyc/deg) are shown as dashed vertical lines below 0 dB. Above the  $Nq_{24-2}$  values, the amplitudes are typically less than 1 dB but are many SE away from 0 and so are highly significant. The larger values for the vertical transect are undoubtedly present, because many of these visual fields had steps.

For this reason, fields with (Figs. 5B, 5C) and without steps (Fig. 5D) also had their spectra averaged separately. While the vertical meridians of the spectra are likely to be affected by any features such as nasal steps, this is not true of the horizontal meridian transects. The transects along the diagonal of the spectra may be partially affected by steplike features, depending on the exact shapes of the defects in given fields. It is clear that most of the 12 transects have frequency components with amplitudes that are many SE away from 0 dB above  $Nq_{24-2}$ . Therefore, it seems that these image components are real and so will be aliased back into visual fields sampled with a 24-2 or similarly coarse array of test points.

The results also indicated that the spectra of the model fields used in this study should be examined. As mentioned in the Methods section, the model fields were constructed by operating a median filter on random noise and then stretching the results nonlinearly to produce somewhat tighter, deeper, model lesions. Figure 6 illustrates the effect of varying the size

**FIGURE 5.** (A) Transects through the average 2D amplitude spectra for HFA 10-2 fields (Carl Zeiss Meditec, North Ryde, NSW, Australia). The legend indicates the transects that are the horizontal and vertical meridians of the 2D spectra and also the transects taken from the 45° diagonal of the spectra. The  $Nq_{24-2}$  values for square and diagonal sampling (0.083 and 0.059 cyc/deg) are indicated as vertical dotted lines below 0 dB. Transects for separate averages are shown for fields with steps (B, C), and the 277 fields without steps (D). Insets: the mean of each type of field. For all types of field, the horizontal meridian transect is unaffected by modulations of the field in the vertical direction such as nasal steps. The diagonal transects may be somewhat affected by such steps, depending on their exact shape. The SEs are shown for every point, and in many cases are so small that they cannot be distinguished from the transect lines. It is clear that many of the spectra have amplitudes ~0.5 to 1 dB above  $Nq_{24-2}$ , which are many SE away from 0 dB and which are not due to nasal steps.





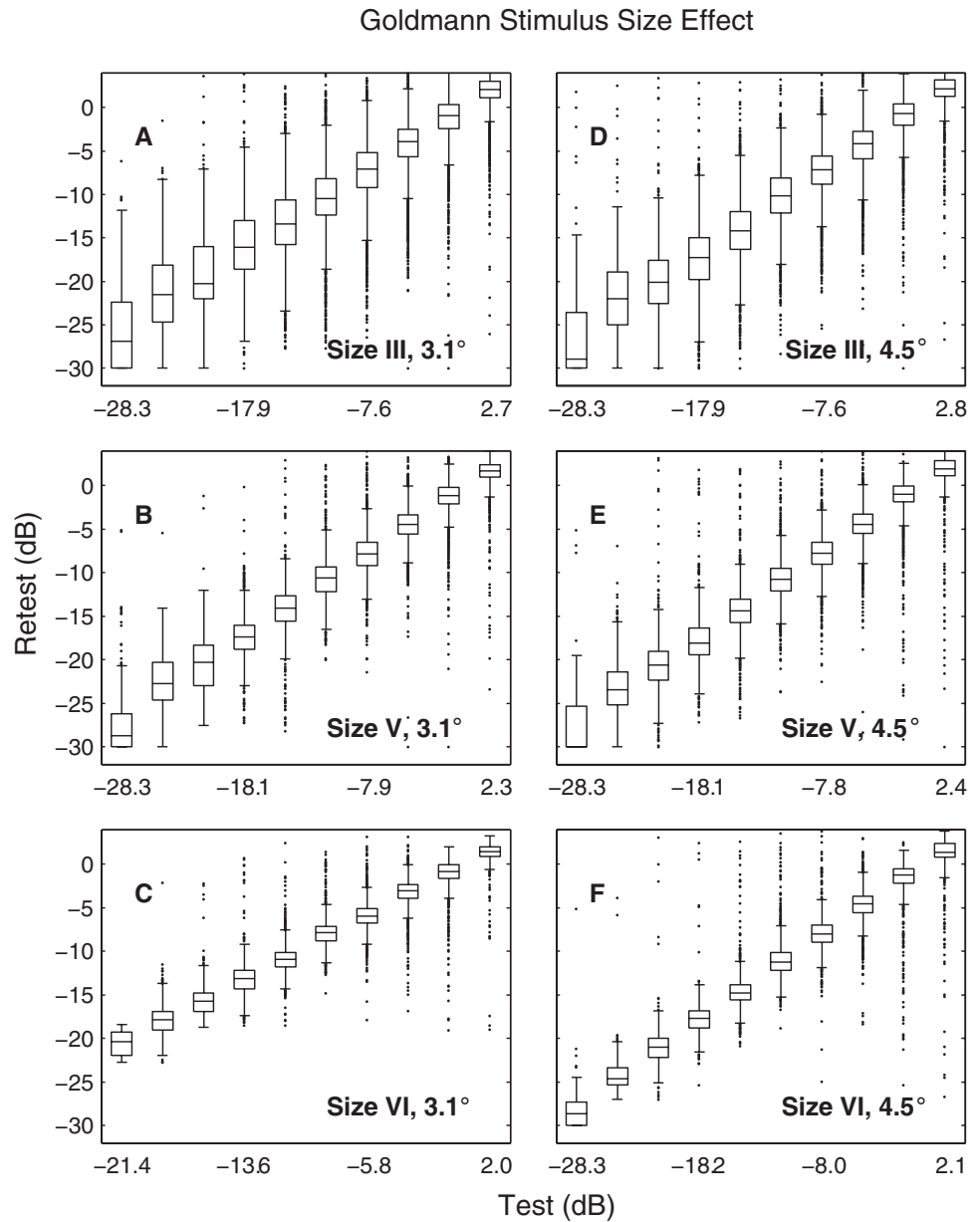
**FIGURE 6.** Six model fields were created for each row of the figure. (A–C) Horizontal transects through the 2D amplitude spectra showing their content out to 0.25 cyc/deg (cf. Figs. 3, 5). The error bars are standard deviations, and for  $n = 6$  are essentially 95% confidence limits. For each row, the model fields were created with a median filter that was (A, D) 1.5, (B, E) 3.1, and (C, F) 4.5 deg<sup>2</sup>. The test-retest plots (D–F) are each based on 1800 samplings of the model fields, and the *dots* beyond the 5th and 95th percentile whiskers are outliers. The modeled fixation jitter was 0.3° (SD). The larger median filter sizes increased the low-frequency content and reduced the high-frequency content. The square and diagonal sampling  $Nqs$  for 24-2 fields are indicated in (A), (B), and (C) by the *vertical dotted* and *dashed lines* at 0.083 and 0.059 cyc/deg.

of the median filter on the spectra (Figs. 6A–6C) and also on the corresponding test-retest variability (Figs. 6D–6F), all measured for a size III stimulus. The test-retest data are shown as box plots, where the upper and lower margins of the boxes indicate the 25th and 75th percentiles, and the whiskers the 5th and 95th percentiles. This mode of plotting was chosen for comparison with similar plots in the literature characterizing test-retest variability from 24-2 SAP fields.<sup>1,2,8</sup> From the upper to the lower rows of Figure 6, the size of the median filters ranged from 1.5° to 3.1° and then 4.5° on a side. Note that, while the median filter tends to preserve any large-scale features that emerge, it also substantially reduces the amplitude of rapid variations in the original noise image. Although the different filters altered the frequency content appreciably, the test-retest variability (Figs. 6D–6F) changed only a little—the main effect possibly being a slight reduction in variability for deeper defects (cf. Figs. 6E and 6F). That is consistent with there being less high-frequency information in Figure 6F to cause an alias for the fields created with the largest filter (cf. Figs. 6B and 6C). Overall, it appears that the assumptions of the model are met, providing there is some image content with amplitudes around 1 dB beyond the Nyquist rate for 24-2 fields.

If aliasing were occurring, then it should be ameliorated by using larger stimuli because they would effectively blur the sampled image, reducing the high-frequency content. Wall et al.<sup>2</sup> (IOVS 2009;50:E-Abstract 2239) have recently reported a reduction in test-retest variability for size V and VI Goldmann

stimuli. Figure 7 shows the effect of varying the stimulus size on the test-retest variability for model fields. As in Figure 6, two types of model fields were examined, those with median filters of 3.1° (Figs. 7A–C) and 4.5° (Figs. 7D–F) on a side. As seen in Figures 6B and 6C, the larger filter size created fields with somewhat lower frequency content, consistent with slightly larger aggregations of loss. For both filter sizes, the results are basically in accord with the observations of Wall et al. One difference is that, for size VI stimuli, the model fields created with the smaller median filter seemed to show some loss of dynamic range (Fig. 7C), whereas those for the larger median filter did not (Fig. 7F). This result may indicate that the model fields of Figures 7D to 7F are more like those of real fields, and the smaller features in the 3.1° model fields are highly blurred by the 3.45°-wide size VI stimuli. As in the data of Wall et al., a floor effect, where no measurement could be less than –30 dB, was introduced to make the comparison more equivalent.

A final question: What is the effect of varying the size of the jitter in the sampled fields? To resolve this question, the spatial grain of the model fields was reduced to 1/30 deg, while the median filter size was maintained at 3.1° square. The standard deviation of the distribution describing the sampling position error was then reduced from the standard 0.3° to 0.15° and then 0.075°. Note that the 0.3° case (Fig. 8A) is effectively a repeat of Figures 2A, 5E, and 6A, computed at higher resolu-



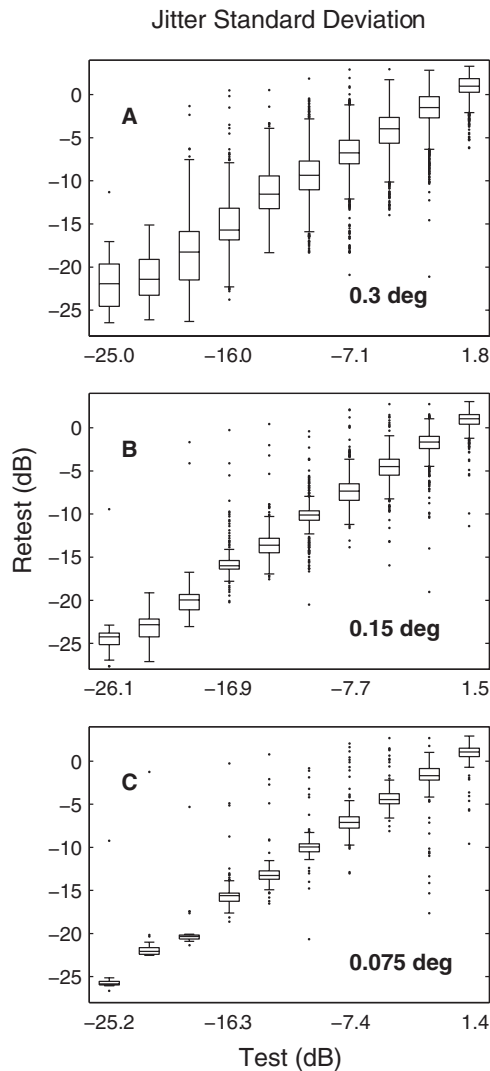
**FIGURE 7.** The effect of varying the stimulus size on the test-retest variability measured from model fields. There were nine model fields per condition and 100 samplings of each field, so that each plot is based on 900 sampled fields. As reported in the literature for SAP fields, the variability is reduced with increasing stimulus size from III to V and VI ( $0.43^\circ$ ,  $1.72^\circ$ , and  $3.45^\circ$  diameter). (A-C) The median filter used to create the model fields was  $3.1^\circ$  on a side, defining a defect size range within the model field. (D-F) The median filter size was  $4.5^\circ$  indicating larger defects than in (A-C). For size VI stimuli, there appeared to be a compression of dynamic range in (C), but not in (F), so it would appear that the large size VI stimulus greatly blurred many of the features of fields based on the smaller filter size. (F) Data are more consistent with those in Wall et al. (*IOVS* 2009; 50:E-Abstract 2239), so perhaps real fields are more like the  $4.5^\circ$  case. Box plots are as in Figure 6.

tion. Although different model fields and spatial resolutions are used in those figures, the outcomes are all very similar.

## DISCUSSION

The possibility that the test-retest variability seen in many forms of perimetry<sup>1-8</sup> is due to undersampling is suggested by the data in Figures 1 to 3. Fourier analysis of data from earlier studies in which a few glaucomatous fields were sampled at  $1^\circ$  intervals<sup>22,24</sup> supports the idea that real fields contain significant spatial frequency content that would form distorting spatial aliases if sampled with a  $6^\circ$  grid (e.g., Fig. 4A). Figures 4B and 4C illustrate the important concept that visual fields smooth enough not to have aliasing effects would look quite unlike our concept of a glaucomatous field. The result that real visual fields contain image components that could form distorting aliases was confirmed by a study of the HFA 10-2 fields (Fig. 5). Model fields were then constructed where variation between smooth, normal parts of the field and areas with concentrations of damage were created using median filters of

different sizes. The resulting amplitude spectra were similar to those of real 10-2 fields (cf. Figs. 5, 6A-C) and the test-retest variability of the model fields was similar to that reported in the literature.<sup>1,2,8</sup> The best match to published results and the 10-2 spectra was obtained for median filters that were between  $3.1^\circ$  and  $4.5^\circ$  a side, which create aggregations of damage at about those scales (Figs. 6E, 6F). Changing the stimulus from Goldmann sizes III to VI produced reductions in test-retest variability consistent with study by Wall et al.<sup>2</sup> (*IOVS* 2009; 50:E-Abstract 2239). The match was especially good for the model fields created with the larger  $4.5^\circ$  filter, in agreement with Figure 4B and other studies of the finer structure of visual fields.<sup>24</sup> The modeling studies of test-retest variability (Figs. 1, 2B, 6, 7) used a fixation error described by a circular distribution with an SD of  $0.3^\circ$ , half the known fixation jitter of good fixators.<sup>20,21</sup> Therefore, like the spatial frequency content of the model fields, the fixation error modeled was conservative. As predicted, when the fixation error SD was decreased to  $0.15^\circ$  and then  $0.075^\circ$ , test-retest variability decreased dramatically (Fig. 8).



**FIGURE 8.** The effect of reducing the SD of the fixation sampling error on the test-retest variability of model fields. Test-retest variability decreases without a change in dynamic range as jitter is reduced.

The present study could be improved in several ways. One improvement would be to study a collection of very fine grained visual fields. These tests would probably have to be repeated several times within a few weeks and the results averaged. The repeated testing would provide an accurate but slightly blurred set of fields and some data on test-retest variability. Methods such as Markov random fields, using some fine-grained fields as input, could generate a large set of model fields that might be more realistic than those used here.<sup>26</sup> That being said, the HFA 10-2 fields used in the present study already have three times higher sampling density than conventional fields and so it is not clear how much would be added. Another approach would be to track the eye movements of persons during perimetric testing and to examine the correlation between test-retest variability and the accuracy of fixation. Tracking accuracy less than  $0.1^\circ$  would be needed, and to do that while providing a  $60^\circ$  field of view would be technically difficult. Also whether anyone could achieve the required  $<0.1^\circ$  fixation accuracy is moot (Fig. 8). One could try to control the stimulus position to cancel some proportion of the fixation error, but again, the process would be technically demanding. It is worth noting that such fixation error cancellation would not remove the distorting effects of aliasing, but

would only produce more similar distortions of the field on each visit.

Much of the data on SAP test-retest variability have been collected from glaucoma patients, and so a possible conclusion is that the observed variability is a feature of glaucoma. However, investigations comparing visual field variability in glaucoma and optic neuritis by Henson et al.<sup>14</sup> indicate that visual field variability is similar whenever defects are produced by ganglion cell loss. The variability measured by those authors was the within-test variability characterized by fitting frequency of seeing (FOS) curves with a cumulative normal distribution where the SD,  $\sigma$ , is indicative of the variability.<sup>7,27,28</sup> Interestingly, Henson et al.<sup>14</sup> found that stimulus eccentricity, fixation loss rate, false-positive rate, and patient age were not significant determinants of the observed variability, consistent with the physical process of sampling, rather than patient vigilance, being the determining influence. Their results were taken to support a hypothesis put forward by other investigators<sup>7,28</sup> that the variability was in part due to the variable responses of individual ganglion cells. Aliasing would mimic that effect.

Aliasing is probably not the whole story, however. For example, in a study using very large frequency-doubling stimuli, the FOS curves are still somewhat broader in the patients than in the normal subjects,<sup>29</sup> although  $\sigma$  was only 1.3 times larger in the patients. This finding is unlike SAP studies in which  $\sigma$  was reported to increase by two to three times.<sup>7,14,27,28</sup> Other factors affecting the signal-to-noise ratios of perimetric methods are also likely to be involved.<sup>30</sup>

On the basis of results using the slightly denser sampling grid of the Competer Perimeter (Bara Elektronik AB, Lund, Sweden), Heijl<sup>31</sup> suggested that few aggregations of damage would be missed altogether by a coarse sampling grid, since they tend to be several degrees wide, suggestive of the results shown in Figure 6. Airaksinen and Heijl<sup>25</sup> have also reported arcuate scotomas  $1^\circ$  to  $2^\circ$  wide, and these would generate large amplitudes at spatial frequencies around 1 cyc/deg, approximately 10 times the Nyquist rate of a  $6^\circ$  test grid. Also, a study of combined 30-1 and 30-2 perimetry indicated that 43% of 68 eyes with measureable defects had moderate to severe disagreement between the 30-1 and 30-2 tests.<sup>32</sup> Of course, some of that disagreement would be due to test-retest variability, but the authors presented evidence that it was often because the  $6^\circ$  grid may have missed the defects. As suggested by Heijl<sup>31</sup> high variability means that retesting is probably more valuable than finer grained sampling. Recent research verifies this conclusion.<sup>15</sup>

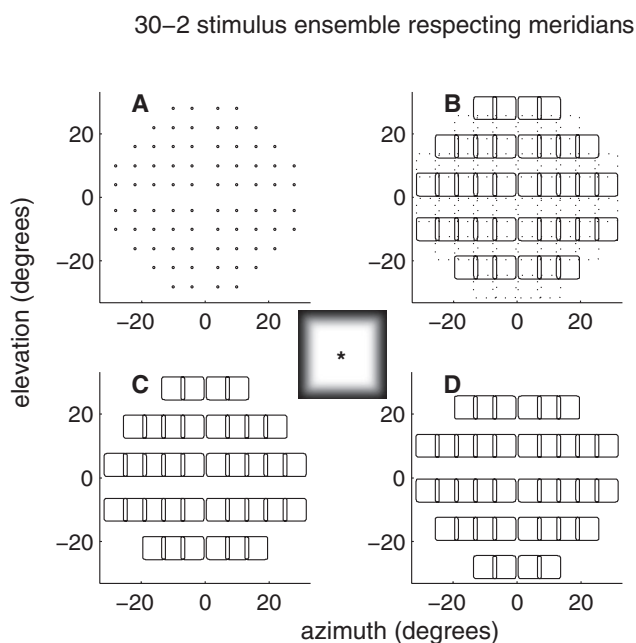
If sampling errors and the size III stimulus are the problem, then the solution to the test-retest problem is larger stimuli, not finer grained sampling or more frequent testing. The issues are how large the stimuli should be and whether any other relevant properties are necessary. Recent results from Wall et al.<sup>33</sup> suggest that size V stimuli actually increase the effective dynamic range of SAP. That being said, stimuli with sharply defined features (e.g., edges) are problematic for several reasons.

The first is that refraction must be excellent for such stimuli. By comparison, stimuli that have a blurred appearance (i.e., contain no spatial frequencies above a few cycles per degree), have limited reduction in physical contrast due to several diopters of defocus.<sup>34</sup> This concept was one of those behind the original FDT perimeter stimuli, which has been shown to be quite tolerant of misrefraction.<sup>35</sup> The main effect of misrefraction on FDT stimuli may be to blur the sharp edges of its stimuli, since physics indicates that the gratings themselves should be slightly demodulated. The smaller 24-2 stimuli of the FDT2 (Matrix; Carl Zeiss Meditec) perimeter have a greater content of edge relative to the stimulus area, perhaps biasing

patients' attention more to the edge contrast than to the grating contrast. This finding may explain the larger test-retest variability of the newer test compared with the original.<sup>7,8</sup>

Another problem with sharp-edged perimetry stimuli relates to the problems of aliasing. Stimuli with sharp features are defined by high spatial frequencies. Therefore, sampling a field with such stimuli means that the high-frequency content of the field tends to be preserved, potentially forming the basis of aliased distortion of the measured visual field map. This problem is again an argument for using large, blurry stimuli.

Fundamentally, the aliasing effect arises because the act of sampling produces copies in the spectrum of the sampled object. These copies repeat at intervals of twice the Nyquist rate.<sup>17,18,36</sup> If the sampled object is not filtered to have frequencies below the Nyquist rate, then the copies overlap in the spectrum, creating the appearance of folded back aliased frequencies (Fig. 3). The solution is to prefilter the sampled object. In perimetry terms, this effect would mean that the test stimuli would have to be composed of sufficiently low spatial frequencies (i.e., blurry), so that the resulting sampled spectra do not overlap. In practice, it means having blurry stimuli that are so large that they overlap somewhat when presented at the perimetric test grid locations. Large overlap may be unacceptable, however, particularly near the horizontal and vertical meridians of the visual field. A compromise is illustrated in Figure 9. Figure 9 shows a 30-2-like stimulus ensemble, using large stimuli that have blurred edges (Fig. 9\*). The size of the stimuli means that the blurring skirts of the stimuli overlap somewhat, except at the meridians.



**FIGURE 9.** An example of a possible stimulus ensemble that would solve the aliasing problem and provide stimuli that were quite resistant to misrefraction. (\*) Illustration of a 9-deg<sup>2</sup> blurred stimulus, suitable for tiling an SAP stimulus grid with minimal overlap. (A) The centers of the modified 30-2 grid which are offset by 1° from the horizontal and vertical meridians. (B) The ensemble of stimuli of the type shown in the \* centered on the positions shown in (A). The size of each stimulus is illustrated by its contour at half height. (C, D) Every other row of contour plots from (B) plotted separately to highlight the way the contours respect the meridians.

## CONCLUSIONS

Overall, many of the parameters of the test-retest variability of SAP visual fields seem to be predictable from sampling errors. The results of this study suggest that these errors are exacerbated by the existence of higher spatial frequencies in visual fields than can be reconstructed by the sampling array combined with the use of and the small, sharp-edged size III stimuli. One solution would be to use larger, smooth-edged stimuli that overlap somewhat except at the central meridians.

## Acknowledgments

The author thanks the reviewers for insightful comments.

## References

- Artes P, Iwase A, Ohno Y, Kitazawa Y, Chauhan B. Properties of perimetric threshold estimates from full threshold, SITA standard, and SITA fast strategies. *Invest Ophthalmol Vis Sci.* 2002;43:2654-2659.
- Wall M, Woodward KR, Doyle CK, Artes PH. Repeatability of automated perimetry: a comparison between standard automated perimetry with stimulus size III and V, matrix, and motion perimetry. *Invest Ophthalmol Vis Sci.* 2009;50:974-979.
- Heijl A, Lindgren A, Lindgren G. Test-retest variability in glaucomatous visual fields. *Am J Ophthalmol.* 1989;108:130-135.
- Blumenthal EZ, Sample PA, Zangwill L, Lee AC, Kono Y, Weinreb RN. Comparison of long-term variability for standard and short-wavelength automated perimetry in stable glaucoma patients. *Am J Ophthalmol.* 2000;129:309-313.
- Kwon YH, Park HJ, Jap A, Ugurlu S, Caprioli J. Test-retest variability of blue-on-yellow perimetry is greater than white-on-white perimetry in normal subjects. *Am J Ophthalmol.* 1998;126:29-36.
- Piltz JR, Starita RJ. Test-retest variability in glaucomatous visual fields. *Am J Ophthalmol.* 1990;109:109-111.
- Chauhan BC, Johnson CA. Test-retest variability of frequency-doubling perimetry and conventional perimetry in glaucoma patients and normal subjects. *Invest Ophthalmol Vis Sci.* 1999;40:648-656.
- Artes PH, Hutchison DM, Nicoletta MT, LeBlanc RP, Chauhan BC. Threshold and variability properties of matrix frequency-doubling technology and standard automated perimetry in glaucoma. *Invest Ophthalmol Vis Sci.* 2005;46:2451-2457.
- Bickler-Bluth M, Trick GL, Kolker AE, Cooper DG. Assessing the utility of reliability indices for automated visual fields: testing ocular hypertensives. *Ophthalmology.* 1989;96:616-619.
- Birt CM, Shin DH, Samudrala V, Hughes BA, Kim C, Lee D. Analysis of reliability indices from Humphrey visual field tests in an urban glaucoma population. *Ophthalmology.* 1997;104:1126-1130.
- Ivers RQ, Macaskill P, Cumming RG, Mitchell P. Sensitivity and specificity of tests to detect eye disease in an older population. *Ophthalmology.* 2001;108:968-975.
- Pierre-Filho Pde T, Schimmi RB, de Vasconcellos JP, 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.
- Keltner JL, Johnson CA, Cello KE, et al. Visual field quality control in the Ocular Hypertension Treatment Study (OHTS). *J Glaucoma.* 2007;16:665-669.
- Henson DB, Chaudry S, Artes PH, Faragher EB, Ansons A. Response variability in the visual field: comparison of optic neuritis, glaucoma, ocular hypertension, and normal eyes. *Invest Ophthalmol Vis Sci.* 2000;41:417-421.
- Chauhan BC, Garway-Heath DF, Goni FJ, et al. Practical recommendations for measuring rates of visual field change in glaucoma. *Br J Ophthalmol.* 2008;92:569-573.
- Maddess T, Goldberg I, Wine S, Dobinson J, Welsh AH, James AC. Testing for glaucoma with the spatial frequency doubling illusion. *Vision Res.* 1999;39:4258-4273.



17. Bracewell RN. Sampling and series. *The Fourier Transform and Its Applications*. New York: McGraw-Hill; 1986:189–218.
18. Giesler WS, Hamilton DB. Sampling-theory analysis of spatial vision. *J Opt Soc Am*. 1986;3:62–70.
19. Maddess T, Hemmi JM, James AC. Evidence for spatial aliasing effects in the Y-like cells of the magnocellular visual pathway. *Vision Res*. 1998;38:1843–1859.
20. Rohrschneider K, Becker M, Schumacher N, Fendrich T, Volcker HE. Normal values for fundus perimetry with the scanning laser ophthalmoscope. *Am J Ophthalmol*. 1998;126:52–58.
21. Springer C, Bultmann S, Volcker HE, Rohrschneider K. Fundus perimetry with the Micro Perimeter 1 in normal individuals: comparison with conventional threshold perimetry. *Ophthalmology*. 2005;112:848–854.
22. Stürmer J, Gloor B, Tobler HJ. Wie sehen glaukomgesichtsfelder wirklich aus? *Klin Monbl Augenbeilkd*. 1984;184:390–393.
23. Stürmer J. What do glaucomatous visual fields really look like in fine-grid computerized profile perimetry? *Dev Ophthalmol*. 1985;12:1–47.
24. Westcott MC, McNaught AI, Crabb DP, Fitzke FW, Hitchings RA. High spatial resolution automated perimetry in glaucoma. *Br J Ophthalmol*. 1997;81:452–459.
25. Airaksinen PJ, Heijl A. Visual field and retinal fibre layer in early glaucoma after optic disc haemorrhage. *Acta Ophthalmol*. 1983;61:186–194.
26. Zhu S, Wu Y, Mumford D. Filters, random fields and maximum entropy (FRAME): towards a unified theory for texture modeling. *Intl J Comp Vision*. 1998;27:107–126.
27. Chauhan BC, Tompkins JD, LeBlanc RP, McCormick TA. Characteristics of frequency-of-seeing curves in normal subjects, patients with suspected glaucoma, and patients with glaucoma. *Invest Ophthalmol Vis Sci*. 1993;34:3534–3540.
28. Wall M, Kutzko KE, Chauhan BC. Variability in patients with glaucomatous visual field damage is reduced using size V stimuli. *Invest Ophthalmol Vis Sci*. 1997;38:426–435.
29. Maddess T, Severt WL, Stange G. Comparison of three tests using the frequency doubling illusion to diagnose glaucoma. *Clin Exp Ophthalmol*. 2001;29:359–367.
30. Artes PH, Chauhan BC. Signal/noise analysis to compare tests for measuring visual field loss and its progression. *Invest Ophthalmol Vis Sci*. 2009;50:4700–4708.
31. Heijl A. Perimetric point density and detection of glaucomatous visual field loss. *Acta Ophthalmol*. 1993;71:445–450.
32. Weber J, Dobek K. What is the most suitable grid for computer perimetry in glaucoma patients? *Ophthalmologica*. 1986;192:88–96.
33. Wall M, Woodward KR, Doyle CK, Zamba G. The effective dynamic ranges of standard automated perimetry sizes III and V and motion and matrix perimetry. *Arch Ophthalmol*. 2010;128:570–576.
34. Campbell FW, Green DG. Optical and retinal factors affecting visual resolution. *J Physiol*. 1965;181:576–593.
35. Artes PH, Nicolela MT, McCormick TA, LeBlanc RP, Chauhan BC. Effects of blur and repeated testing on sensitivity estimates with frequency doubling perimetry. *Invest Ophthalmol Vis Sci*. 2003;44:646–652.
36. Petersen DP, Middleton D. Sampling and reconstruction of wave-number-limited functions in n-dimensional Euclidian spaces. *Inform Control*. 1962;5:279–323.