

# Sampling geometries for ocular aberrometry: A model for evaluation of performance

Luis Diaz-Santana<sup>1</sup>, Grace Walker<sup>1</sup> and Salvador X Bará<sup>2</sup>

<sup>1</sup>*Applied Vision Research Centre, Henry Wellcome Laboratories for Visual Science  
Department of Optometry and Visual Science  
City University, Northampton Square, London EC1V 0HB, UK*

<sup>2</sup>*Area de Optica, Dept. Fisica Aplicada. Facultade de Fisica  
Universidade de Santiago de Compostela, Campus Sur, E-15782 Compostela, Galiza, Spain*

**Abstract:** The purpose of this work is to outline a simple model to assess the relative merits of different sampling grids for ocular aberrometry and illustrate it with an example. While in traditional Hartmann-Shack setups the sampling grid geometries have been somewhat restricted by the geometries of the available microlens arrays, other techniques such as laser ray tracing or spatially resolved refractometry allow for a greater freedom of choice. For all available setups, including HS, it is worth studying which of these choices perform better in terms of accuracy (closeness of the obtained results to the actual ones) and precision (uncertainty of the obtained results). Whilst the mathematical model presented in this paper is quite general and it can be applied to optimise existing or new aberrometers, the numerical results presented in the example are only valid for the particular aberration sample used and centroiding algorithms studied, and should not be generalised outside of these boundaries.

© 2005 Optical Society of America

**OCIS codes:** (010.1080) Adaptive optics; (330.5370) Physiological optics; (330.4460) Ophthalmic optics

---

## References and links

1. P.B. Liebelt, *An Introduction to Optimal Estimation*, (Addison-Wesley, Reading, MA, 1967) pg. 138.
2. S. Bará, "Measuring eye aberrations with hartmann-shack wave-front sensors: Should the irradiance distribution across the eye pupil be taken into account?," *J. Opt. Soc. Am. A* **20**, 2237–2245 (2003).
3. B. P. Medoff, *Chap. 9: Image reconstruction from limited data: theory and applications in computerized tomography*. In H. Stark, editor, *Image Recovery: Theory and Application*. (Academic, New York, 1987).
4. J. Herrmann, "Least-squares wave front errors of minimum norm," *J. Opt. Soc. Am.* **70**, 28–35 (1980).
5. W.H. Southwell, "Wave-front estimation from wave-front slope measurements," *J. Opt. Soc. Am.* **70**, 998–1006 (1980).
6. L. Diaz Santana Haro. *Wavefront sensing in the human eye with a Shack-Hartmann sensor*. (PhD thesis, Imperial College of Science Technology and Medicine, 2000)
7. Jim Schwiegerling. "Scaling Zernike expansion coefficients to different pupil sizes," *J. Opt. Soc. Am. A.* **19**, 1937–1945 (2002).

---

## 1. Introduction

We present in this paper a simple model to assess quantitatively the relative merits of different sampling grids for ocular aberrometry. There are usually different possible choices for the

microlens arrays to be used in a Hartmann-Shack (HS) wavefront sensor, as well as for the sampling patterns of sequential aberrometers such as laser ray-tracers or spatially resolved refractometers. It is then worth developing an evaluation model to ascertain which grids will perform better in terms of accuracy (closeness of the estimated aberration coefficients to the actual ones) and precision (uncertainty of the results), for a given population and measuring device.

By “sampling grid” we mean a given set of values for the following parameters: number of subpupils, spatial distribution of their centers (square, hexagonal, polar lattice), subpupil shape, size, and apodization (uniform irradiance vs gaussian apodization as in laser ray-tracing). Although it is in principle possible to compare several grids building them and checking them experimentally by measuring a given number of actual or model eyes with known aberrations this is not always practical, especially if what is sought is to analyze the performance of a relatively large set of grids and to assess the effects of modifying some of their parameters. In this paper we present a complementary approach to cope with this problem, based on the evaluation of the grid performance using equations derived from a small set of assumptions. The performance of a grid is quantified by its first and second-moment error matrices which provide, respectively, the expected and the mean squared differences between the actual and the estimated aberrations, averaged over measurements and population. The required calculations can be carried out straightforwardly, allowing in each particular case the selection of the best grid candidates out of a given range of possible choices.

The accuracy and precision of the aberrations measured using a given sampling grid depend not only on the particular parameters of that grid but also on the statistics of the aberrations of the population, the measurement noise and, last but not least, the phase estimator, *i.e.* the algorithm used to retrieve the wave aberration from the raw data provided by the sensor. All these factors are included in this model. The population statistics is described by the first and second-moment matrices whose elements are the average values of the aberration coefficients in the population and the average values of their cross-products, respectively. The noise statistics is included through the corresponding first and second-moment matrices of noise, which give the average values of the measurement noise at each microlens and the averages of their cross-products. Regarding the phase estimator, the only assumption made in this model is that it is linear, *i.e.*, that the aberration coefficients are estimated as linear combinations of the sensor measurements (which is the usual case: except for some methods based on the use of, for instance, neural networks, most estimation procedures used in aberrometry are linear. This holds for the conventional least-squares approach as well as for any method based on the integration of the measured slopes).

Given this dependence, it must be stressed that any statement about grid performance is not absolute, but relative to the population, the sensor noise and the estimator chosen to calculate the aberrations. These factors are necessary inputs of the evaluation process. In principle, then, there is no “best grid” universally valid for all possible situations and choices. The model presented here is meant as a flexible tool for grid performance evaluation, and its use requires substituting in each case the proper values for these input factors.

As an example of an application, we used the general equations presented in this paper to compute the expected performance of several sampling grids with square, hexagonal and polar geometry, for the aberration statistics of a population sample measured in the Department of Optometry and Visual Science at The City University in London (UK) and a particular model of sensor noise, assuming the use of the conventional unweighted linear least squares estimator (LSQ) which has found a widespread use in eye aberrometry. This example is meant to illustrate the application of the general equations to a particular case: the behavioral trends of performance obtained on our study do not claim any validity beyond that range.

The structure of this paper is as follows: in section 2 we present the general algebraic expressions for the first and second-order estimation error. In section 3 these results are applied to the assessment of the behaviour of several sampling grids for an example with particular values of statistics, noise and estimator. Finally, discussion and conclusions are drawn in section 4.

## 2. Matrix model of wavefront estimation

Let  $W(\mathbf{r})$  be the aberration function to be measured by a wavefront sensor using a sampling grid with  $N$  subapertures. The measurements provided by the sensor are the local wavefront slopes averaged over each subpupil, weighted by the local irradiance distribution<sup>2</sup>. They can be arranged as a  $2N$  dimensional vector,  $\mathbf{m}$ , whose elements  $m_s$  ( $s = 1, \dots, 2N$ ) are given by

$$m_s = \int_{\infty} P_s(\mathbf{r}) \nabla_s W(\mathbf{r}) d^2\mathbf{r} + v_s \quad (1)$$

where:

$P_s(\mathbf{r})$  is the normalized irradiance pupil function associated to the  $s$ -th sampling subaperture, equal to the squared modulus of the amplitude pupil function including any irradiance distribution present in the incoming wave.  $P_s(\mathbf{r}) = P_{s+N}(\mathbf{r})$ , for  $s = 1, \dots, N$ ,

$\nabla_s$  is  $d/dx$  (for  $s = 1, \dots, N$ ) and  $d/dy$  (for  $s = N + 1, \dots, 2N$ ),

$d^2\mathbf{r}$  is the surface element ( $dx dy$ ) in the pupil plane, the integration limits being extended formally to the whole plane (the spatial extent of individual subpupils is defined by the function  $P_s(\mathbf{r})$ ),

and finally  $v_s$  is the measurement noise, whose  $s = 1, \dots, 2N$  values can also be arranged as a  $2N$  vector  $\mathbf{v}$ .

For a typical HS sensor measuring a constant irradiance wave with all microlenses of the same area  $a$ ,  $P_s(\mathbf{r})$  equals  $1/a$  for points inside the  $s$ -th microlens and 0 everywhere else; if the irradiance  $I(\mathbf{r})$  of the wave is not constant,  $P_s(\mathbf{r})$  equals

$$\frac{I(\mathbf{r})}{\int_a I(\mathbf{r}) d^2\mathbf{r}} \quad (2)$$

for points inside the microlens and 0 otherwise. For a laser ray tracer with a Gaussian spot of half-width  $\sigma$  at the eye pupil,  $P_s(\mathbf{r})$  is a normalized Gaussian function of this width, centered at the corresponding sampling location.

Expressing the aberration as a sum of basis functions (not necessarily orthogonal, although this is the most common choice) we can write

$$W(\mathbf{r}) = \sum_{i=1}^{M'} a_i Z_i(\mathbf{r}) \quad (3)$$

Where the  $Z_i(\mathbf{r})$  are usually the Zernike polynomials and the  $a_i$ ,  $i = 1 \dots M'$  are the actual coefficients of the aberrated wavefront. This sum has to be extended to a  $M'$  large enough as to faithfully capture any spatial variation of the wavefront. In strict mathematical terms  $M'$  should equal infinity in order to represent exactly any aberration function  $W(\mathbf{r})$ , independently from their discontinuities, although physically it is not expected that contributions of order higher than a given  $M'$  will have any relevance in eye aberrometry. \* In what follows we will take  $M'$  as an arbitrarily large number, however finite in order to ease some matrix computations.

\*Note that the actual number of modes present in the wavefront ( $M'$ ) may be a very large number, probably greater than the number of modes which may have by themselves a direct clinical relevance. As it will be shown below, higher order modes introduce aliasing into lower order modes. This means that modes that may not be clinically relevant, may nevertheless introduce an error in the estimation of modes that are clinically relevant. Hence one should aim to reconstruct a wavefront as completely as possible.

Substituting Eq. (3) into Eq. (1) it is easy to obtain the matrix form of Eq. (1) as

$$\mathbf{m} = \mathbf{A}\mathbf{a} + \mathbf{v} \quad (4)$$

where  $\mathbf{a}$  is the modal coefficients vector (of length  $M'$ ) whose elements are the actual aberration coefficients  $a_i$  ( $i = 1, \dots, M'$ ), and  $\mathbf{A}$  is a  $(2N \times M')$  matrix whose  $(s, i)$ -th element is given by

$$\int_{\infty} P_s(\mathbf{r}) \nabla_s Z_i(\mathbf{r}) d^2\mathbf{r} \quad (5)$$

In practice the aberration  $W(\mathbf{r})$  is estimated by a truncated finite sum of terms

$$W_e(\mathbf{r}) = \sum_{i=1}^M \hat{a}_i Z_i(\mathbf{r}) \quad (6)$$

whose coefficients  $\hat{a}_i$  ( $i = 1, \dots, M$ ),  $M < M'$ , have to be estimated fulfilling some merit criteria. If the estimation procedure is linear (as it is the usual case in ocular aberrometry) then each coefficient is estimated as a linear combination of the measurements. In matrix form this can be written as

$$\hat{\mathbf{a}} = \mathbf{R}\mathbf{m} \quad (7)$$

where  $\hat{\mathbf{a}}$  is the  $M$ -dimensional vector of estimated coefficients  $\hat{a}_i$  and  $\mathbf{R}$  is a  $M \times 2N$  matrix known as the “estimator” or “estimation matrix”. The elements of  $\mathbf{R}$ ,  $R_{is}$ , (where  $i = 1, \dots, M$ ;  $s = 1, \dots, 2N$ ), are the weights used to obtain the estimated coefficients  $\hat{a}_i$  as linear combinations of the measurements  $m_s$ .

### 2.1. First order statistics: coupling matrix, aliasing and estimation bias

Equations (4) and (7) are useful to construct a fast picture of the way modal estimation works. Let us assume that the aberration to be estimated may vary according to some statistics and that the measurement noise is a zero-mean random variable, and let the brackets  $\langle \rangle$  denote averaging over measurements and population, except when indicated otherwise. Substituting Eq. (4) into Eq. (7) we have<sup>†</sup>

$$\hat{\mathbf{a}} = \mathbf{R}(\mathbf{A}\mathbf{a} + \mathbf{v}) = \mathbf{R}\mathbf{A}\mathbf{a} + \mathbf{R}\mathbf{v} \quad (8)$$

Since the measurement noise averages to zero after a set of enough measurements the expected value of  $\hat{\mathbf{a}}$  is given by

$$\langle \hat{\mathbf{a}} \rangle = \langle \mathbf{R}\mathbf{A}\mathbf{a} + \mathbf{R}\mathbf{v} \rangle = \mathbf{R}\mathbf{A} \langle \mathbf{a} \rangle \quad (9)$$

and the expected error, that is, the expected difference between the actual and the retrieved coefficients will be

$$\langle \mathbf{a} - \hat{\mathbf{a}} \rangle = (\mathbf{I} - \mathbf{R}\mathbf{A}) \langle \mathbf{a} \rangle \quad (10)$$

where  $\mathbf{I}$  is the identity matrix. If, additionally, the actual wavefront remains constant between measurements we have

$$\langle \hat{\mathbf{a}} \rangle = \mathbf{R}\mathbf{A}\mathbf{a} \quad (11)$$

<sup>†</sup>A note on dimensions:  $\hat{\mathbf{a}}$  is  $M \times 1$ ,  $\mathbf{R}$  is  $M \times 2N$ ,  $\mathbf{A}$  is  $2N \times M'$ ,  $\mathbf{a}$  is  $M' \times 1$ ,  $\mathbf{v}$  is  $2N \times 1$ .  $\mathbf{R}\mathbf{A}$  is then  $M \times M'$ . In order to write expressions such as  $\langle \mathbf{a} - \hat{\mathbf{a}} \rangle$  or  $\mathbf{I} - \mathbf{R}\mathbf{A}$  it is formally useful to extend  $\hat{\mathbf{a}}$  to have as many elements as  $\mathbf{a}$ , setting  $\hat{a}_i = 0$  for  $i = M + 1, \dots, M'$ , and similarly to extend  $\mathbf{R}$  to be a  $M' \times 2N$  matrix, completing it with  $M' - M$  identically zero rows. With this extension  $\mathbf{R}\mathbf{A}$  becomes of size  $M' \times M'$  and all sums and differences of vectors or matrices have the proper dimensions. For the sake of simplicity, however, in loose speaking we will usually refer to  $\hat{\mathbf{a}}$ ,  $\mathbf{R}$  and  $\mathbf{R}\mathbf{A}$  as being of sizes  $M \times 1$ ,  $M \times 2N$  and  $M \times M'$ , respectively, ignoring their trivial extensions.

and

$$\langle \mathbf{a} - \hat{\mathbf{a}} \rangle = (\mathbf{I} - \mathbf{RA})\mathbf{a} \quad (12)$$

Note from Eqs. (9) and (11) that if the product  $\mathbf{RA}$  is not equal to the identity the retrieved coefficients will be biased, in the sense that their average value after a set of measurements will not be equal to the expected value of the actual coefficients. The retrieved coefficients will show a systematic bias no matter how many measurements could be made to average to zero the measurement error.

An exception is the case of a population where  $\langle \mathbf{a} \rangle = 0$ , which trivially ensures always unbiased estimation in the broad sense indicated by Eqs. (9) and (10); this is indeed the case in atmospheric optics, although it seems not hold for the eye, at least for some terms like for instance spherical aberration which does not average to zero over the population. Note also that unbiased estimation over a population, in the sense of Eqs. (9) and (10) is a less strict condition than unbiased estimation over measurements for a given individual of that population, in the sense of Eqs. (11) and (12)

The matrix  $\mathbf{RA}$  gives useful information.  $\mathbf{RA}$  is a ‘‘coupling matrix’’ which determines how the actual coefficients mix up and combine among themselves to produce the values of the retrieved ones. Bearing in mind that the vector of actual coefficients,  $\mathbf{a}$ , has length  $M'$  and that the vector of retrieved ones,  $\hat{\mathbf{a}}$ , is of length  $M$ ,  $\mathbf{RA}$  is an  $M \times M'$  matrix. In an ideal case the square sub-matrix formed by the first  $M \times M$  elements of  $\mathbf{RA}$  should be the identity and the remaining elements (that is, all columns of  $\mathbf{RA}$  of index  $k$  such that  $M < k < M'$ ) should be identically zero. This would ensure that each retrieved coefficient would be exactly equal to the actual one, i.e.  $\hat{a}_i = a_i$  ( $i = 1, \dots, M$ ). However, as will be shown below, this ideal situation does not arise in practice (excepting when  $M = M'$ , and the least-squares estimator is correctly calculated).

Note that if one or more diagonal elements of the first  $M \times M$  square box of  $\mathbf{RA}$  are different from 1 they will give rise to an under or overestimation of the corresponding actual coefficient; that is, were the actual wavefront had just only one of these modes, it would be retrieved with bias (said otherwise, there would be a wrong self-modal coupling). In turn, the off-diagonal elements of this box different from zero as well as any other nonzero element in the columns  $M + 1 \dots M'$ , will give rise to aliasing (cross-modal coupling): the value of the  $i$ -th retrieved coefficient  $\hat{a}_i$  will depend not only on the corresponding actual coefficient  $a_i$  but will have contributions from other modes  $a_j$  ( $j \neq i$ ) present in the incoming wavefront.

In this work we use the word ‘bias’ to denote the first-order statistical expectation error  $\langle \mathbf{a} - \hat{\mathbf{a}} \rangle$  arising from the fact that  $\mathbf{RA}$  differs from the identity matrix  $\mathbf{I}$  (for the  $M$  retrieved modes). ‘Bias’ encompasses in a single word the effects of wrong self-modal coupling and aliasing, since these two effects play a similar formal role in the calculations: they arise from the same source, the nonzero elements of  $(\mathbf{I} - \mathbf{RA})$ , and give rise to the same consequence, a systematic error in the estimation of the aberrations. However, they may have different origins, and to give some insight of them it is necessary to choose the estimator  $\mathbf{R}$ . The common choice for  $\mathbf{R}$  in eye aberrometry is the unweighted least-squares one.

$\mathbf{A}$  is not a square matrix, hence it has no inverse in strict sense and we cannot choose  $\mathbf{R}$  as the otherwise obvious choice  $\mathbf{R} = \mathbf{A}^{-1}$  which would assure  $\mathbf{RA} = \mathbf{I}$ . However, if the  $M' \times M'$  matrix  $\mathbf{A}^T \mathbf{A}$  is nonsingular then a so-called Moore-Penrose pseudoinverse can be defined as:<sup>3,4</sup>

$$\mathbf{A}^\dagger = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \quad (13)$$

where  $T$  stands for ‘transpose’. Choosing  $\mathbf{R} = \mathbf{A}^\dagger$  we would have  $\langle \mathbf{a} \rangle = \langle \hat{\mathbf{a}} \rangle$  and hence no bias. Note however that avoiding this bias would require computing exactly the estimator

$$\mathbf{R} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \quad (14)$$

and this cannot be done satisfactorily in many practical cases, since the matrix  $\mathbf{A}$  in this equation has  $2N \times M'$  elements, that is, as many columns as modes are actually in the aberrated wavefront. This number may be exceedingly high, and to ensure that  $\mathbf{A}^T \mathbf{A}$  is nonsingular there must be enough measurement points  $N$ . In practice the finite sampling grid size is a limiting factor which excludes using the matrix  $\mathbf{A}$  with the complete number of modes to construct the estimator; the common solution in aberrometry is to compute  $\mathbf{R}$  using instead of  $\mathbf{A}$  a matrix of smaller dimensions (let it be denoted by  $\mathbf{A}_M$ ) with the same  $2N$  rows but with only the first  $M$  columns ( $M < M'$ ) corresponding to the first  $M$  modes, such that the available measurement points are enough to ensure that  $\mathbf{A}_M^T \mathbf{A}_M$  has inverse. The resulting least squares (LSQ) estimator  $\mathbf{R} = (\mathbf{A}_M^T \mathbf{A}_M)^{-1} \mathbf{A}_M^T$  is biased. This effect is graphically shown in Figure 1(a), where we show the values of the elements of the first  $M' = 35$  columns of the coupling matrix  $\mathbf{R}\mathbf{A}$  for a square lattice of 69 square microlenses. It assumes that only the first  $M=20$  Zernike modes are included in the estimator  $\mathbf{R}$ . Note that the first  $M \times M$  square box is the identity (there is no wrong self-modal coupling) but there are nonzero elements in the columns of index greater than  $M$ . These nonzero elements give rise to aliasing of high order-modes to low-order ones (The figure is displayed in a logarithmic grayscale).

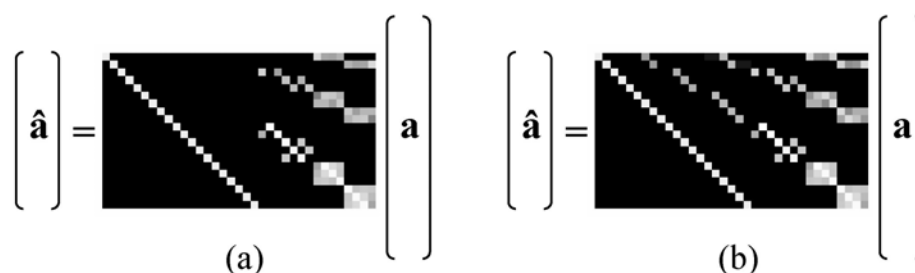


Fig. 1. First  $M' = 35$  columns of the coupling matrix  $\mathbf{R}\mathbf{A}$  for a square grid of 69 square microlenses assuming that only the first  $M=20$  Zernike modes are included in the estimator  $\mathbf{R}$ . (a)with  $\mathbf{R}$  constructed using the correct model of measurements (wavefront slopes spatially averaged over each subpupil); (b)with an incorrect measurement model (wavefront slopes evaluated at the center of each subpupil). The values are shown in a logarithmic grayscale comprising four decades (white = 1, black  $< 0.0001$ ).

In the preceding paragraph we implicitly assumed that the elements of  $\mathbf{A}_M$  used to construct  $\mathbf{R}$  are correctly calculated using Eq. (5), which in turn results from Eq. (1) which describes the actual working of the wavefront sensor. But if these elements are calculated with wrong assumptions (for instance, neglecting spatial averaging over subpupils or the actual irradiance distribution) then some additional bias will appear.<sup>2</sup> The reason is that no matter what our assumptions may be about the measurements (and hence how we compute  $\mathbf{R}$ ) the actual measurements will behave according to Eq. (1) with the correct parameters for  $\mathbf{A}$ . So, if  $\mathbf{R}$  is not computed using the correct matrix  $\mathbf{A}_M$ , the first  $M \times M$  box of  $\mathbf{R}\mathbf{A}$  will not be the identity. This effect is shown in Figure 1(b), which shows the elements of  $\mathbf{R}\mathbf{A}$  for the same grid of Fig 1(a), but in this case assuming that the elements of  $\mathbf{R}$  were incorrectly computed neglecting the spatial averaging of the wavefront slope that takes place at each sampling subaperture (that is, each element of  $\mathbf{A}_M$  was now calculated as the local derivative of the corresponding Zernike mode evaluated at the center of each microlens, instead of averaged over the subpupil). Note that this choice of the estimator would lead in this example to aliasing also from modes with

index smaller than  $M$ . Actual situations, however, may be even richer and more complex. As shown in ref.<sup>2</sup> if the uneven irradiance distribution actually present across each microlens is not correctly taken into account when constructing  $\mathbf{R}$ , there may appear modal cross-coupling of low-order modes to high-order ones, in addition to the usual "high-to-low" aliasing, even for the case  $M = M'$  (see figures 2 and 3 of that reference for details).

This example shows that, even for the same population and noise statistics, the grid performance is strongly tied to the choice of  $\mathbf{R}$ . For the usual least-squares estimator,  $\mathbf{R} = (\mathbf{A}_M^T \mathbf{A}_M)^{-1} \mathbf{A}_M^T$ , the performance depends on how close is  $M$  to  $M'$  and also (and this is a fact frequently overlooked) on how well the way of calculating each element of  $\mathbf{A}_M$  reproduces the measurements that the sensor would give were the corresponding Zernike mode, with unit amplitude, being incident on it (since this is essentially the meaning of Eq(5)).

Note that the magnitude of the systematic errors in the estimated coefficients  $\hat{\mathbf{a}}$  depends not only on  $\mathbf{R}\mathbf{A}$  but also on  $\mathbf{a}$ , the actual aberration. If the first-order aberration statistics  $\langle \mathbf{a} \rangle$  is available, the expected bias can be obtained straightforwardly using Eqs. (9) and (10), or (11) and (12) for the limiting case of a static aberration.

Different sampling grids will give rise to different estimation biases because the elements of  $\mathbf{A}$  (Eq. (5)) depend on the distribution and shape of the subpupils and any sensible choice for the estimation matrix  $\mathbf{R}$  will be a more or less sophisticated function of  $\mathbf{A}$  or of a subset of its elements.

The effects of noise are dependent on the estimation matrix  $\mathbf{R}$ , as indicated by Eq. (8). However, with regard to first-order statistics, if the noise is of zero mean it will produce no biasing effects. It will, however, be a relevant term of the overall rms error given by the second order statistics, as it will be shown in the next section.

## 2.2. Second-order statistics: Expected rms estimation error for a given population

The expected bias is just one factor of concern. Another relevant quantity is the expected rms estimation error. It is determined by the second-order statistics of the aberration coefficients and noise, as well as by the estimator and the sampling grid.

A useful parameter to assign a single figure to the estimation error is the squared difference between the estimated and actual wavefronts, spatially averaged over the eye pupil  $\Pi$ . For a given actual aberration pattern  $\mathbf{a}$  this difference is given by:

$$\sigma^2(\mathbf{a}) = \int_{\Pi} [W(\mathbf{r}; \mathbf{a}) - W_e(\mathbf{r}; \mathbf{a})]^2 d^2\mathbf{r} \quad (15)$$

where the dependence on  $\mathbf{a}$  has been explicitly indicated. Working with a given population, a single figure useful to assess the relative merits of different reconstruction schemes is  $\sigma^2$ , that is, the expected value of  $\sigma^2(\mathbf{a})$  averaged over population and measurements, *i.e.* over all possible values for  $\mathbf{a}$  and the measurement noise. Smaller values of  $\sigma^2$  will correspond to better (in the rms sense) reconstruction schemes. If we expand the wavefronts in terms of the basis functions and these are orthogonal, it is easy to see that

$$\begin{aligned} \sigma^2 &= \langle \sigma^2(\mathbf{a}) \rangle = \left\langle \int_{\Pi} [W(\mathbf{r}; \mathbf{a}) - W_e(\mathbf{r}; \mathbf{a})]^2 d^2\mathbf{r} \right\rangle \\ &= \sum_{i=1}^M \langle (\hat{a}_i - a_i)^2 \rangle + \sum_{i=M+1}^{M'} \langle (a_i)^2 \rangle \end{aligned} \quad (16)$$

where the brackets indicate averaging over population and measurements. If a non orthogonal basis would be used, double-index sums and crosscorrelation terms like  $\langle (\hat{a}_i - a_i)(\hat{a}_k - a_k) \rangle$

would appear in Eq. (16). The first sum in Eq. (16) arises from the estimation error in the  $M$  retrieved coefficients. The second sum corresponds to the error due to those aberration terms of order higher than  $M$  are not estimated and hence are not included in  $W_e$ . We can write Eq. (16) as a single sum up to  $M'$ , with the provision that  $\hat{a}_i = 0$  for  $i > M$ .

Given two column vectors  $\mathbf{f}$  and  $\mathbf{g}$ , of lengths  $N_f$  and  $N_g$ , their second-moment matrix is defined as the  $N_f \times N_g$  matrix

$$\mathbf{C}_{fg} = \langle \mathbf{fg}^T \rangle$$

It can be seen then, that  $\sigma^2$  is just the trace of the matrix  $\mathbf{C}_e$ , the second-moment matrix of the estimation errors of the modal coefficients, given by

$$\mathbf{C}_e = \langle (\hat{\mathbf{a}} - \mathbf{a})(\hat{\mathbf{a}} - \mathbf{a})^T \rangle \quad (17)$$

Computing  $\mathbf{C}_e$  as a function of useful parameters is an easy task, if we make use of Eqs (7) and (4). We get the long but straightforward expression:

$$\begin{aligned} \mathbf{C}_e &= \langle \hat{\mathbf{a}}\hat{\mathbf{a}}^T \rangle - \langle \hat{\mathbf{a}}\mathbf{a}^T \rangle - \langle \mathbf{a}\hat{\mathbf{a}}^T \rangle + \langle \mathbf{a}\mathbf{a}^T \rangle \\ &= \mathbf{R}[\mathbf{A}\mathbf{C}_a\mathbf{A}^T + \mathbf{A}\mathbf{C}_{av} + \mathbf{C}_{av}^T\mathbf{A}^T + \mathbf{C}_v]\mathbf{R}^T - \mathbf{R}[\mathbf{A}\mathbf{C}_a^T + \mathbf{C}_{av}^T] \\ &\quad - [\mathbf{C}_a\mathbf{A}^T + \mathbf{C}_{av}]\mathbf{R}^T + \mathbf{C}_a \end{aligned} \quad (18)$$

where:

- $\mathbf{C}_a = \langle \mathbf{a}\mathbf{a}^T \rangle$  is the second-moment matrix of the actual coefficients (size  $M' \times M'$ ), computed from the known statistics of the population.
- $\mathbf{C}_{av} = \langle \mathbf{a}\mathbf{v}^T \rangle$  is a second-moment matrix whose elements are the expected values of the cross-products of the actual coefficients and measurement noise (size  $M' \times 2N$ )
- $\mathbf{C}_v = \langle \mathbf{v}\mathbf{v}^T \rangle$  is the second-moment matrix of noise (size  $2N \times 2N$ ), computed from a good knowledge of the experimental setup or from a run of calibration measurements made with a static reference wavefront (In order to measure this noise term under relevant experimental conditions, artificial eyes with scattering retinas and phase plates introducing aberrations of magnitude typical of those found when measuring human eyes should be used. As pointed out by an anonymous reviewer, measuring noise with a scattering-free and nonaberrated artificial eye will give rise to an underestimation of  $\mathbf{C}_v$ )
- $\mathbf{R}$  is any estimator whose performance we want to assess, and  $\mathbf{A}$  has been defined in Eq. (5).

This formula is the main result of this model, and it allows one to compute the expected performance (in the rms sense) of any combination of grid, population, noise and estimator, through the single parameter  $\text{trace}(\mathbf{C}_e)$ , once the required second-moment matrices have been measured or calculated. Note that the geometry of the sampling grid only appears in  $\mathbf{A}$  and possibly in  $\mathbf{R}$ . The other matrices are determined by the population statistics and/or the sensor noise.

Some reasonable assumptions allow to simplify somewhat this equation. For instance, as a first approximation the actual coefficients and the measurement noise can be taken as essentially uncorrelated for usual experimental setups and conditions. Then, setting  $\mathbf{C}_{av} = 0$ , taking into account that  $\mathbf{C}_a = \mathbf{C}_a^T$ , and rearranging terms, Eq. (18) becomes:

$$\mathbf{C}_e = (\mathbf{I} - \mathbf{R}\mathbf{A})\mathbf{C}_a(\mathbf{I} - \mathbf{R}\mathbf{A})^T + \mathbf{R}\mathbf{C}_v\mathbf{R}^T \quad (19)$$



A further simplification can be made taking into account that usually the measurement noise has an equal rms value at all the subapertures, and there is no correlation of noise between different subapertures and/or along different directions. The matrix  $\mathbf{C}_v$  is then diagonal with all diagonal terms equal to the noise variance  $\sigma_v^2$  so that

$$\mathbf{C}_e = (\mathbf{I} - \mathbf{R}\mathbf{A})\mathbf{C}_a(\mathbf{I} - \mathbf{R}\mathbf{A})^T + \sigma_v^2\mathbf{R}\mathbf{R}^T \quad (20)$$

The first term in the right-hand side of Eq. (20) accounts for the expected squared wavefront error due to the biased estimation of the modal coefficients. It depends on the population statistics through  $\mathbf{C}_a$  and on the first-order bias through the matrix  $(\mathbf{I} - \mathbf{R}\mathbf{A})$ . The second one accounts for the propagation of the measurement noise, and it depends on the noise  $\sigma_v^2$  and on the estimator  $\mathbf{R}$ . Different grids will contribute differently to  $\mathbf{C}_e$ , again due to the definition of  $\mathbf{A}$  and to the  $\mathbf{A}$ -dependent choices for  $\mathbf{R}$ . Eq. (20) may be used as a good tool for grid evaluation, through the parameter  $\sigma^2 = \text{Tr}(\mathbf{C}_e)$ .

If  $\mathbf{R} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$ , with all  $M'$  modes included in  $\mathbf{A}$  and a correct computation of its elements, then the first term of Eq. (20) would identically vanish since  $\mathbf{R}\mathbf{A} = \mathbf{I}$ , and hence no bias, independently from the chosen grid. But even in this wildly optimistic case, the rms uncertainty due to noise propagation would remain. Again, different grids would give rise to different rms wavefront errors due to noise propagation. For this particular choice of  $\mathbf{R}$ , and taking into account that  $(\mathbf{A}^T\mathbf{A})^{-1}$  is symmetrical we get

$$\begin{aligned} \mathbf{C}_e &= \sigma_v^2\mathbf{R}\mathbf{R}^T = \sigma_v^2[(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T][(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T]^T \\ &= \sigma_v^2[(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T][\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}] = \sigma_v^2(\mathbf{A}^T\mathbf{A})^{-1} \end{aligned} \quad (21)$$

If, as usual,  $\mathbf{R}$  is chosen as  $\mathbf{R} = (\mathbf{A}_M^T\mathbf{A}_M)^{-1}\mathbf{A}_M^T$  with  $M < M'$ , the noise contribution to  $\mathbf{C}_e$  will be the trace of the matrix

$$\sigma_v^2(\mathbf{A}_M^T\mathbf{A}_M)^{-1} \quad (22)$$

### 3. Performance of sampling grids: a case study

In this section we evaluate the performance of different sampling geometries and densities applying the model described above to a particular choice of the input factors: population statistics, sensor noise and estimator. This section is intended as an illustration of the application of this model, and its results do not claim validity for situations where the assumptions made here about these input factors do not hold.

We have divided this analysis in two subsections each one addressing a particular term of the overall estimation error  $\sigma^2 = \text{Tr}(\mathbf{C}_e)$ . First we analyse how the measurement noise propagates in different grids assuming the same noise characteristics as that used to derive Eq. (20). Secondly we analyse the contribution of the bias introduced by different geometries.

Under the usual assumptions the variance of the wavefront slopes measurement error  $\sigma_v^2$  is a scalar and a function of the detector used, SNR, centroiding algorithm, etc. The noise contribution to the total error in the wavefront estimation is given by

$$\sigma_v^2\text{Tr}[(\mathbf{A}_M^T\mathbf{A}_M)^{-1}] \quad (23)$$

This last result is in agreement with Southwell.<sup>5</sup>

The contribution to the total error due to the bias introduced by the sampling geometry is given by

$$\text{Tr}[(\mathbf{I} - \mathbf{R}\mathbf{A})\mathbf{C}_a(\mathbf{I} - \mathbf{R}\mathbf{A})^T] \quad (24)$$

This contribution depends on the second-moment matrix  $\mathbf{C}_a$  of the particular population being tested. In this study we have used a population sample of 93 eyes taken at City University

with a commercial wavefront sensor. The information in this sample is restricted to modes up to the 4-th order (14 first non-trivial Zernike terms). Hence, no aliasing effects coming from modes higher than the 14-th may be studied, averaged over the population, using this dataset. To our knowledge, comprehensive datasets of the statistics eye aberrations including the detailed values for all cross-correlation terms  $\langle a_i a_j \rangle$  up to high orders are not yet easily available in the open literature. In spite of this limitation of our statistical sample, it is perfectly valid for illustrating an example of the way in which this model may be applied, although the numerical results of each grid performance would of course vary to some extent were the information from the statistics of modes higher than 14 be included in  $\mathbf{C}_a$ . Details of the sample's statistics are given in section 3.2.1.

### 3.1. Noise propagation

The sampling geometries explored were square, hexagonal and 3 different polar geometries, as shown in Fig. 2. The number of sampling points in each ring of the polar geometries is defined as

$$\begin{aligned} S_n &= Mn; & n >= 2 \\ S_n &= 1; & n = 1; \end{aligned} \quad (25)$$

where  $S_n$  is the number of sampling points in ring  $n$ , and  $M$  is 2, 4, and 6 for the geometries Polar2, Polar4 and Polar6 respectively. The rings are counted from the center with the central spot defined as the first ring,  $n = 1$ .

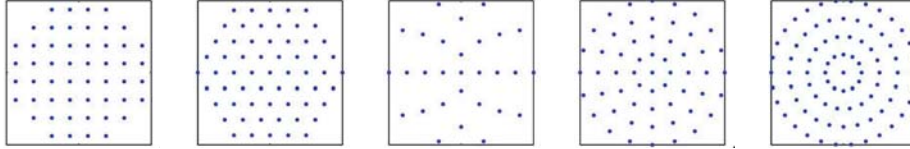


Fig. 2. Different geometries analysed. From left to right: Square, Hexagonal, Polar2, Polar4 and Polar6.

Each one of these geometries defines a particular matrix  $\mathbf{A}$ . The usual least squares estimator for  $M$  modes  $\mathbf{R} = (\mathbf{A}_M^T \mathbf{A}_M)^{-1} \mathbf{A}_M^T$  gives rise to a corresponding “noise propagator matrix”  $\mathbf{N}_\sigma = (\mathbf{A}_M^T \mathbf{A}_M)^{-1}$ , and according to Eq. (23), the diagonal of this matrix quantifies how the noise is propagated for each mode. Hence the contribution of noise to the overall estimation error will be

$$\begin{aligned} \sigma^2 &= \sigma_v^2 \text{Tr}(\mathbf{A}_M^T \mathbf{A}_M)^{-1} \\ &= \sigma_v^2 \text{Tr}(\mathbf{N}_\sigma) \\ &= \sigma_v^2 N_\sigma \end{aligned} \quad (26)$$

where the scalar  $N_\sigma = \text{Tr}(\mathbf{N}_\sigma)$  is the overall mean square noise propagator.

For each geometry in Fig. 2 different sampling densities were analysed. First, the number of sampling points were varied discretely, depending on each geometry. The square and hexagonal geometries were varied by increasing in one the number of sampling points over the horizontal diameter of the pupil while the polar geometries were varied by increasing the number of rings by one. Then, the number of reconstructed modes,  $M$ , was varied from 4th order up to 8th order (That is 14, 20, 27, 35, 44 terms respectively). Finally, for each geometry, sampling density and

number of modes a matrix  $\mathbf{A}_M$  and its corresponding mean square error propagator  $N_\sigma$  were calculated.

If the number of modes is larger than the number of sampling points the matrix  $\mathbf{A}_M^T \mathbf{A}_M$  may become singular or close to singular. These configurations were eliminated from the analysis. The chosen base was a Zernike polynomial base following the OSA normalisation.

Figure 3a shows the results for all the geometries together. While Fig. 3b shows the results for 14 and 44 modes with each geometry in a different colour. These plots show very similar behaviour for all the geometries tested. Firstly, it is evident that the noise propagator diminishes as the sampling density increases. Conversely the noise propagator increases as the number of modes to be reconstructed also increases. The changes with geometry are negligible compared with the changes observed with increased sampling density. The latter showing a power law behaviour:

$$N_\sigma = Ax^{-\alpha} \quad (27)$$

where  $x$  has replaced the previously used  $N$  as the number of sampling points –lenslets in HS for instance– (This change of notation is done to avoid confusion with  $N_\sigma$  in the remaining of the text),  $A$  defines the offset of each curve and  $\alpha$  the negative of its slope in a log-log plot. Table 1 shows the values for  $\alpha$  for the different sampling geometries under study and several number of reconstructed modes. The larger the value of  $\alpha$  is the more rapidly the magnitude of the noise propagator increases as the number of sampling points is reduced.

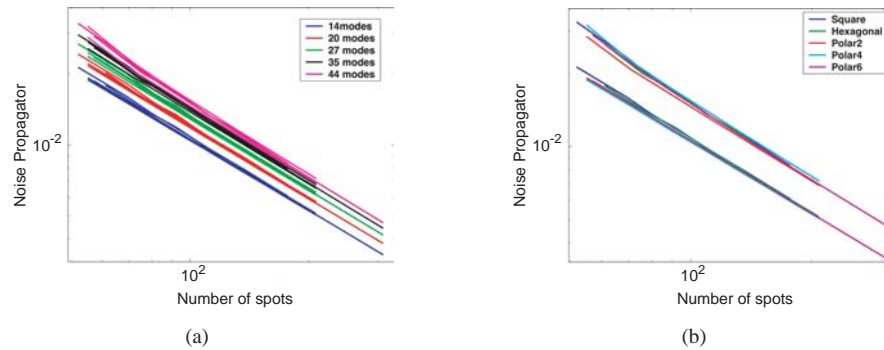


Fig. 3. noise propagators for different geometries and number of modes

Table 1. Values of  $\alpha$  for different geometries and different number of modes. As the number of modes increases so does the value of  $\alpha$ . Also note that  $\alpha$  takes maximum values for a square geometry and minimum for polar4. See text for further details

	Square	Hexagonal	Polar2	Polar4	Polar6
$\alpha$ 14 modes	1.0209	1.0103	0.9992	1.0003	1.0028
$\alpha$ 20 modes	1.0307	1.0166	1.0055	1.0071	1.0113
$\alpha$ 27 modes	1.0401	1.0216	1.0106	1.0127	1.0188
$\alpha$ 35 modes	1.0526	1.0284	1.0168	1.0197	1.0295
$\alpha$ 44 modes	1.077	1.0393	1.0347	1.028	1.0474

The implications of these results are not as straight forward as it may appear. One would be tempted to conclude that maximising the density of the sampling geometry would minimise the

noise propagation as a direct consequence of Eq. (27). However, according to Eq. (26) the total propagated rms error depends not only on the mean square noise propagator  $N_\sigma$  but also on the error in the measurement  $\sigma_v^2$ .

Note on the one hand that Eq. (27) is only valid for the geometries and sampling densities explored, and that any generalisation beyond these parameters is not justifiable. On the other hand, note that Eq. (26) is much more general and the coupling between  $\sigma_v^2$  and  $N_\sigma$  should always be studied.

The amount of light reaching an ocular wavefront sensor depends on the irradiance of the light used to illuminate the eye, the reflectivity of the subject's retina and the scattering and absorption in ocular media. This in turn, combined with instrument losses and sensor sensitivity defines the error in the measurement  $\sigma_v^2$ . Because the amount of light allowed into the eye is limited by safety considerations, the error in the measurement  $\sigma_v^2$  will depend for a particular instrument on the patient reflectivity and SNR at each sampled location. In turn, the SNR will depend in many "real life" parameters like size of the retinal spot, intraocular scattering, subjects' refraction, aberration dynamics, speckle, etc. It is through Eq. (26) that these "real life" parameters can be brought into this model. This can be achieved by using a realistic model eye to experimentally measure  $\sigma_v^2$  or a realistic numerical model to computationally estimate it.

In the case of a HS sensor, or similar devices, the SNR at each lenslet will decrease as the number of lenslets increases. The same energy is distributed over more sub-apertures. In this case we could expect an optimal number of lenslets that would minimise Eq. 26 as a function of energy per lenslet.

Sequential sensors, as a ray tracing device, collect all the light reflected by the retina at each sampling location. In such a sensor a similar situation occurs, but now in the temporal domain. As the number of sampled location increases, the time allocated to sample each one is reduced. This in turn leads to a reduction in SNR. Moreover, as the eye is a dynamic system, all measurements need to be completed in a very short time interval to ensure all of them correspond to the same temporal state of the eye. The situation is again that of the SNR being reduced as the number of samples is increased, and the need to optimise the number of measurements as a balance between noise and noise propagator described by Eq. (26) arises again. In the same fashion as with a HS sensor, the SNR and hence  $\sigma_v^2$  in this case, is dictated by "real life" parameters, and it can be estimated either experimentally or computationally. The way this is done does not affect the validity of Eq. (26).

### 3.1.1. Signal to noise ratio vs sampling density.

In this section we proceed to further explore the coupling between the measurement error  $\sigma_v^2$  and the noise propagator  $N_\sigma$ . To do this we simulated as a first approximation to a real ocular HS spot, a Gaussian spot and added photon and electronic noise to it. Several sets of 100 sampling points each were generated. Each set with the same noise statistics, and their centroids found using an iterative centroiding algorithm assisted with a Gaussian mask. The performance of the centroiding algorithm for the different noise statistics were then analysed. Full details of these simulations can be found in reference.<sup>6</sup> In this paper we only summarise those results and link them to the error propagation problem in the context of this particular example.

Figure 4(a) shows the centroiding error, which is proportional to the slope measurement error  $\sigma_v^2$ , as a function of spot intensity and for several levels of electronic noise. A power law behaviour similar to that in Eq. (27) is observed.

$$\sigma_v^2 = DI^{-\beta} \quad (28)$$

Table 2 shows the different values that  $\beta$  takes for different levels of electronic noise. It is clear that  $\beta$  increases together with the electronic noise. This means that a drop in intensity has a

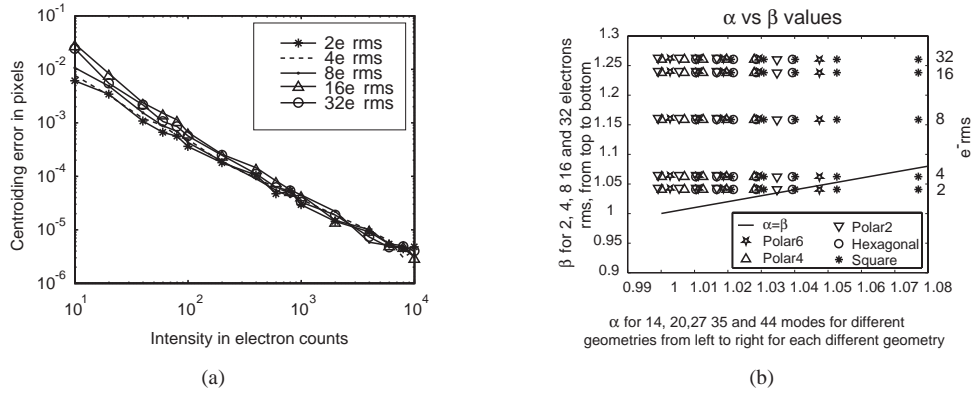


Fig. 4. (a) Centroiding error as a function of spot intensity for different levels of electronic noise. (b) Comparison between  $\alpha$  and  $\beta$ . Each row depicts a level of electronic noise, from 2 to 32 electrons rms from the top row to the bottom one. Each column depicts a different number of reconstructed modes and each symbol a different geometry. For each geometry the column furthest to the left corresponds with 14 modes (4th order) increasing to the right up to 44 modes (8th order). The diagonal line represents  $\alpha = \beta$

Table 2. Values of  $\beta$  for several different levels of electronic noise. Note that  $\beta$  increases together with the electronic noise

Read-out noise in electrons rms	2	4	8	16	32
$\beta$	1.0405	1.0622	1.1585	1.2379	1.2602

major impact on the accuracy of the centroiding for larger levels of electronic noise. This is not unexpected. Note also that  $\beta$  takes values close to one in most cases.

In order to couple these results with those obtained in the previous section we need to relate the number of sampled locations with the intensity  $I$  in Eq. (28). In the case of a HS sensor this is straight forward, as the number of sampled locations is equal to the number of lenslets utilised and the intensity of each spot produced by the lenslet array is proportional to  $1/x$ ,

$$I = C/x \quad (29)$$

where  $x$  is the number of lenslets and  $C$  is the number of electron counts that a sensor with one single lens would produce on the same CCD camera. Substituting this in Eq. (27) we obtain

$$\begin{aligned} N_{\sigma} &= A(C/I)^{-\alpha} \\ &= BI^{\alpha} \end{aligned} \quad (30)$$

where  $B = AC^{-\alpha}$ .

The total error in the reconstruction is then found by substituting Eqs. (28) and (30) into Eq. (26)

$$\begin{aligned} \sigma^2 &= \sigma_v^2 N_{\sigma} \\ &= \Gamma x^{\beta-\alpha} \end{aligned} \quad (31)$$

where  $\Gamma = ADC^{-\beta}$ .

Finally, Fig. 4 (b) compares  $\alpha$  and  $\beta$  in an attempt to explain the meaning of Eq. (31). The diagonal line is the locus of  $\alpha = \beta$ . Any point that lies on this curve represents an equilibrium state: an increase in the number of sampling points is perfectly balanced by an increase in the centroiding error. In this case  $\alpha - \beta = 0$  and Eq. (31) becomes

$$\sigma^2 = \Gamma = \text{constant} \quad (32)$$

The error is constant and depends on  $ADC^{-\beta}$ .  $A$  is fixed by the sampling geometry and the number of estimated modes,  $D$  depends on the CCD camera's electronic noise,  $C$  the total number of photons over the entire pupil and  $\beta$  which is fixed by the CCD camera noise characteristics and the centroiding algorithm used.

We can see that only an asterisk (\*) and a circle (o) intersect (or almost intersect) this line. The asterisk corresponds with the 3rd column from the right and bottom row for the square geometry (27 modes reconstruction and  $2 e^-$  readout noise rms) while the circle corresponds with the first column (again from the right) and the bottom row for the hexagonal geometry (44 modes reconstruction and  $2 e^-$  readout noise rms). In these two cases the number of HS lenslets does not affect the result, provided that there are at least 27 and 44 lenslets in each case. Increasing or decreasing the sampling density does not affect the accuracy of the wavefront estimation.

All the points below the diagonal represent configurations in which increasing the number of lenslets will increase the accuracy of the result. The limit in this case is the number of photons available over the whole pupil. In this case we can increase the sampling density until being photon limited in each lenslet to maximise the accuracy.

All the points above the diagonal, represent configurations in which minimising the sampling density is the best strategy. In this cases the minimum number of sampling areas would be limited by the need of preventing  $\mathbf{A}_M^T \mathbf{A}_M$  of becoming singular or close to singular.

These results can be summarised as follows: If the algorithm used for centroiding in this example was to be used in a HS sensor that produced similar spots to those used in the simulations, and a very low electronic noise CCD camera, in such a way that  $\beta < \alpha$  the best design would need to maximise the number of samples. On the other hand, if the CCD camera to be used had a higher electronic noise, and this would make  $\beta > \alpha$ , then it would be better to minimise the number of SH lenslets. The authors feel important to stress once again that in principle it is not justified to generalise these results beyond the boundaries of this example, and that the coupling between  $N_\sigma$  and  $\sigma_v$  needs to be addressed by instrument designers on an instrument to instrument basis.

### 3.2. Sampling rms error.

We turn now our attention to the rms error introduced in wavefront measurements by the different sampling geometries.

#### 3.2.1. Population statistics

We applied this model to a population of 93 eyes measured at City University using a commercial aberrometer, a WASCAs Wave Front Analyzer. (Asclepion Meditec AG), which provides the value of the coefficients of the first  $M' = 14$  non trivial Zernike polynomials (*i.e.* Zernike aberrations up to a radial 4-th degree).

A random sample of 50 subjects was taken from a university setting. The age range was from 18 to 52 years with an average age of 25.5 years and a standard deviation of 7.1 years. Two hundred and one aberration measurements from 50 right eyes and one hundred and twenty seven aberration measurements from 43 left eyes were taken. Each subject was measured with their

usual refractive correction of glasses or contact lenses and it was assumed that this gave each subject best-corrected vision. Thirty subjects were emmetropes, eleven subjects wore glasses and nine subjects wore contact lenses. No exclusion criteria were applied to the sample.

No mydriatic or cycloplegic drugs were administered and each eye was measured in its natural state. The data was collected in a darkened room. The subject placed his head on the chin rest. To align the centre of the pupil with the axis of the instrument six dots of corneal reflections have to be brought into focus and aligned concentrically with the pupil. The subject was asked to blink before the measurement was taken to establish a clear and uniform tear film. The subject was then asked to look through the target image of a spider web, into the distance, to reduce accommodation effects. Several image captures were taken until a satisfactory result was obtained where the pupil size was comparatively large and well aligned with the axis of the instrument. The smallest pupil size measured was 2.34mm and the largest was 7.27mm in diameter.

The measurements from each eye were averaged, the aberration coefficients reduced to a common pupil size<sup>7</sup> from the resulting coefficient sets the second-moment matrix  $\mathbf{C}_a = \langle \mathbf{a}\mathbf{a}^T \rangle$  was calculated. The  $(i, j)$  element of  $\mathbf{C}_a$  is the product of the  $i$ -th and  $j$ -th aberration coefficients averaged over the whole population,  $\langle a_i a_j \rangle$ . The result showed that in this population there is a variable degree of correlation between different Zernike terms, hence, the second-moment matrix was not diagonal. Other populations may well show a different matrix  $\mathbf{C}_a$ , depending on their particular composition; hence previous measurements of the eye aberrations of a statistically significant sample of eyes are required to assess the behaviour of the sampling grids. We have used for this study the first 14 Zernike terms, including tilts, which give a population rms aberration  $\sigma_a = [\text{Tr}(\mathbf{C}_a)]^{1/2}$  of about  $2.86 \mu\text{m}$ . The population rms aberration associated exclusively to high-order aberrations (*i.e.* excluding prismatic and second-order spherocylindrical terms) is about  $0.45 \mu\text{m}$ . These statistical parameters are characteristic of this sample of population and by no means are intended to be representative of other groups of people.

### 3.2.2. rms error introduced by different sampling geometries

In this section we discuss the effects of geometry on the rms error introduced in the estimation of the eye aberrations. The second-moment matrix  $\mathbf{C}_a$  from the sample described in section 3.2 was used in this calculations. The sample geometries explored were Square, Hexagonal and Polar2 from Fig. 2. Polar4 and Polar6 were not used in this case as the number of sampling points in these two geometries increases considerably faster than on the other three, making it difficult to make a meaningful comparison across the different geometries in this particular context.

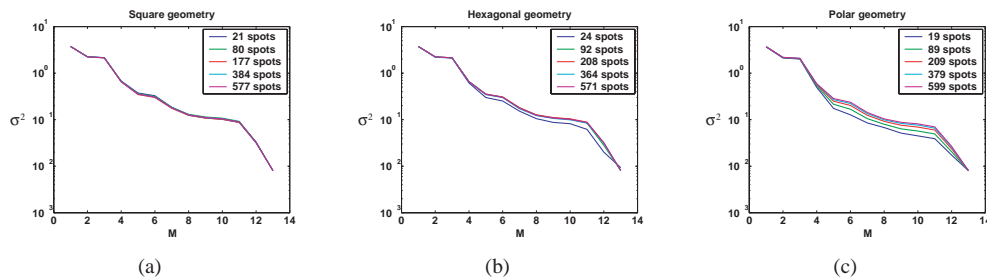


Fig. 5. Bias introduced for the different samples as a function of grid density. (a) Square grid, (b) Hexagonal grid, (c) Polar grid.

For each geometry five sampling grids were produced, each one of larger sampling density

than the previous one. For each grid  $\sigma^2 = \text{Tr}(\mathbf{C}_e) = \text{Tr}[(\mathbf{I} - \mathbf{R}\mathbf{A})\mathbf{C}_a(\mathbf{I} - \mathbf{R}\mathbf{A})^T]$  was calculated using least-squares reconstruction matrices  $\mathbf{R} = (\mathbf{A}_M^T \mathbf{A}_M)^{-1} \mathbf{A}_M^T$  with  $M$  ranging from 1 to  $M' - 1$ . The case  $M = M' = 14$  was not included, since with the input data of this example (database extending only up to the 14-th mode) it gives an identically zero rms error. The results are shown in Figs. 5 and 6

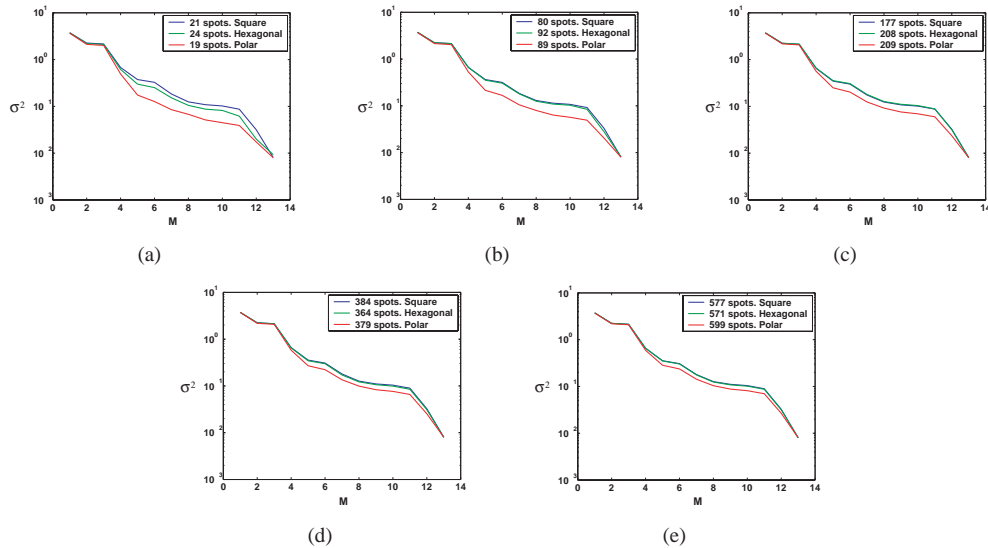


Fig. 6. Bias introduced for the different samples as a function of grid density. (a) Very low density, (b) Low density, (c) Medium density, (d) High density and (e) Very high density.

Figure 5 shows that in the square and hexagonal grids the sampling density has a very small impact on the rms error introduced, apart from very low densities in the hexagonal case. The variations with grid density in the case of the square one are almost null. Instead, in the case of the polar grid, lower sampling densities consistently produced a smaller error. These results, together with the previous ones regarding noise propagation suggest that a low density grid may actually be the best one for ocular wavefront sensing under the particular assumptions on the statistics of the population and noise used for this example.

Figure 6 shows the behaviour of grids of comparable density for different geometries. It is clear from this figure that the polar geometry produces consistently a lower rms error than the other two. The hexagonal grating performs moderately better than the square one, and the square one is the worst in all cases. Although as the sampling density increases the performance of the polar grids moves closer to that of the hexagonal and square ones. Once again, this result is valid for the assumptions made in this particular example. Polar grids may be more sensitive to aliasing from higher order modes with angular frequencies multiple of the frequency of the grid, whose statistics were not available in our clinical sample.

#### 4. Discussion

Ocular aberrometry has become a very popular technique in the fields of visual and physiological optics. The analysis presented in this paper is just a step for a better understanding of how measurements are affected both by random and systematic errors. In this section we discuss the limitations of the methodology presented and some of the problems that need to be addressed



in the future.

We presented in Section 2 of this paper a simple yet reasonably comprehensive model for evaluating the performance of sampling grids in eye aberrometry. The main results of this paper are summarized in Eqs. (10) and (18) relative to the first order estimation bias and the rms estimation error. The model uses as inputs, besides the grid parameters, the statistics of the aberrations in the population, the sensor noise and the estimator used to retrieve the aberrations from the raw data provided by the aberrometer. The range of validity of this model is wide, as far as first- and second-order error effects are of concern. The only limiting assumptions made to deduce Eqs. (10) and (18) are that the aberrometer raw-data are the displacements of centroids, that the paraxial wave equation holds (both are the basic assumptions leading to Eq. (1)) and that the modal aberration estimation is linear, that is, the aberration coefficients are estimated as linear combinations of the centroid displacements. The model accounts in an intuitive way for the sources and effects of aliasing and wrong self-modal coupling, through the analysis of the coupling matrix  $\mathbf{RA}$  and its effects in the first- and second-order estimation errors. This model is intended to serve as an evaluation tool which shall be applied in each case using the appropriate values for the input parameters, corresponding to the particular clinical or experimental settings.

In Section 3 we applied this model to a particular choice of statistics, noise and estimator. This section is intended as an example of application and the particular results obtained therein should not be taken as guidelines for aberrometer design outside the range of validity of the assumptions made to get them. Different population statistics, estimator or noise behaviour will give rise to different performance results. Two main limitations are present in this numerical example. One of them is due to the fact that the clinical database available for our study was limited to Zernike modes up to 4-th order. This allowed us to study the effects of aliasing due to the first 14 modes. This leaves out of consideration aliasing from higher order modes whose inclusion would certainly change to a bigger or lesser extent the results obtained here, depending on its statistical expectation values in the population. For instance, it is expected that taking into account these additional contributions to aliasing, a higher number  $M$  of modes should be retrieved in order to keep bias within acceptable limits; this in turn would push for a higher number of sampling subpupils than that suggested by our example. This limitation may be at least partially overcome in future studies using clinical databases with higher order modal data. It may also be overcome by using a model for the aberration power spectrum: we are currently developing this last approach.

Another limitation of the numerical results of Section 3 is that we used Eq. (20) instead of the more general Eq. (18) to calculate the rms estimation error. Eq. (20) stems from the simplifying assumption that the phase slope measurement noise is uncorrelated with the actual aberration of the wave incident on the aberrometer (i.e., we set  $\mathbf{C}_{av} = 0$ ). However, for highly aberrated wavefronts some correlation may be expected. If the aberration is strong and has relevant high-order terms, the focal spots of the microlenses will get deformed and blurred. This blurring, in itself, would not be a problem were it not for the CCD noise. The noiseless centroids of these blurred spots behave as expected, *i.e.* are proportional to the aberration slope averaged across the microlens subpupil (Eq. (1)). However, since under strong aberrations the same amount of energy is spread over a wider focal region and there is always an unavoidable amount of detector noise, the centroiding noise will increase. Centroid SNR can in principle be partially restored by increasing the irradiance of the illuminating beam, but of course this has unavoidable limits related to eye safety. Additionally, if blurring is severe, some overlapping of neighboring focal regions may appear, making more difficult their isolation to proceed to centroid computation, maybe causing some loss of relevant information. All this means that the measurement noise will be higher than that considered in our numerical example and, additionally, will be correlated

with the actual aberration coefficients. Elaborating a quantitative model for the correlation of aberrations and noise deserves further study, in order to properly account for the effects of nonzero  $C_{av}$ .

The general model introduced in Section 2 can be used to evaluate the performance of any sampling grid, for any given combination of linear estimator and statistics of aberrations and noise. Two related and interesting topics not addressed here are: first, how to use this model to compute the optimum estimator for any given combination of grid, population and noise statistics. By “optimum” we mean that linear estimator which provides the smaller minimum mean squared error (LMMSE) between the actual and the estimated aberration coefficients, averaged over measurements and individuals in the population using the given sampling grid. In general this optimum estimator will not be equal to the usual LSQ. The LMMSE estimator can be straightforwardly calculated by a direct application of the Gauss-Markov linear estimation theorem<sup>1</sup> and its study has been extensively developed for wavefront sensing by the astronomy community.

And secondly, the selection of the best grid in a more general sense, that is, the sampling grid which –in combination with its corresponding LMMSE estimator, which takes into account the population and noise statistics– will give the smallest possible error, leaving the grid parameters as optimization parameters free to vary between certain bounds. Solving this last problem requires the use of nonlinear optimization methods in a relatively high-dimensional space. Both topics will be the subject of forthcoming work.