

Stochastic realization approach to the efficient simulation of phase screens

Alessandro Beghi,¹ Angelo Cenedese,² and Andrea Masiero^{1,*}

¹*Dipartimento di Ingegneria dell'Informazione, Università di Padova, via Gradenigo 6/B, 35131 Padova, Italy*

²*Dipartimento di Tecnica e Gestione dei Sistemi Industriali, Università di Padova, Stradella San Nicola 3, 36100 Vicenza, Italy*

*Corresponding author: masiero@dei.unipd.it

Received June 20, 2007; revised October 7, 2007; accepted November 5, 2007;
posted November 27, 2007 (Doc. ID 84301); published January 31, 2008

The phase screen method is a well-established approach to take into account the effects of atmospheric turbulence in astronomical seeing. This is of key importance in designing adaptive optics for new-generation telescopes, in particular in view of applications such as exoplanet detection or long-exposure spectroscopy. We present an innovative approach to simulate turbulent phase that is based on stochastic realization theory. The method shows appealing properties in terms of both accuracy in reconstructing the structure function and compactness of the representation. © 2008 Optical Society of America
OCIS codes: 010.1330, 350.5030.

1. INTRODUCTION

The introduction of computer control and in particular the application of modern control techniques to adaptive and active optics have significantly advanced the design of multiple mirror telescopes, opening the pathway to the construction of the several-meter-diameter Very Large Telescope (VLT [1]) and the next-generation telescopes such as those described in [2,3]. Adaptive optics (AO) are used to overcome the resolution limitation caused by atmospheric turbulence by compensating for factors that affect the image at fast timescales (1/100 s or even less). Such factors are not easily corrected with primary mirrors, so that AO have been developed for small corrective mirrors and recently for secondary mirrors.

As is nowadays common practice in control engineering, the design of AO control systems is performed by resorting to computer-aided control system design (CACSD) tools. In particular, simulations are required to assess the control system performance, where it is crucial to be able to reproduce the main disturbances affecting the system, such as the wavefront distortion introduced by atmospheric turbulence.

Modeling of atmospheric turbulence is not an easy task, since it is a nonlinear, chaotic process. Turbulent fluctuations in the wind velocities in the upper atmosphere mix layers of differing temperatures, densities, and water vapor content. As a consequence, the refraction index of each level of the atmosphere fluctuates and the wavefront incident on the telescope along an optical path that encounters these fluctuations has spatial and temporal variations in phase and amplitude. Across the diameter of a large telescope, the phase errors are of the order of a few micrometers and dominate the degradation of spatial resolution.

A possible way to describe turbulence in the atmo-

sphere is provided by the Kolmogorov theory [4–6], which is based on a statistical description of the refractive index, temperature, and velocity of the atmosphere. Kolmogorov introduced the concept of inner and outer scales: The outer scale is the largest-sized scale of the turbulent structure and is related to the size of the structure that initiates the turbulence. The inner scale is the smallest scale where turbulent energy starts to dissipate due to viscous friction. Wind velocity fluctuations and the motion of turbulent structures are considered to be approximately locally homogeneous and isotropic.

The spectrum of the refraction index is well modeled by Kolmogorov theory only in a limited range of frequencies (the so-called inertial range, which is the spatial range between inner and outer scales), and when there is the need to extend predictions beyond this regime, the Von Karman spectrum is preferred, which introduces a characteristic parameter called the outer scale of spatial coherence L_0 , leading to attenuation of the phase spectrum at low frequencies. This model tends to the Kolmogorov model when L_0 tends to infinity. Hereafter we use “outer scale” to indicate the outer scale of the spatial coherence L_0 .

From a computational point of view, atmospheric turbulence is often simulated by means of the so-called phase screen method. Pictorially, the phase screen is a randomly inhomogeneous thin layer placed along the path of propagation of a wave that affects the wavefront with a phase perturbation. In doing so, the phase screen introduces a planar perturbation on a horizontal plane, and along the vertical dimension the turbulence effect is modeled through the insertion of a number of screens, each contributing to the overall phase perturbation [7]. In this paper we address the problem of simulating such distorted wavefronts, in particular when the generation of atmo-

spheric phase screens for very long exposures is required. An innovative approach to simulate turbulent phases is presented, based on the stochastic realization theory, which allows us to take into account the turbulence statistics to extend an existing phase screen in time. The method is consistent with recently presented techniques [8] and shows appealing properties in terms of accuracy and compactness of the representation.

2. PROBLEM STATEMENT

The basic question is how to choose the properties of the phase screen so that it accurately models the atmosphere.

The spatial statistical characteristics of the turbulent phase ϕ are generally described by means of the structure function D_ϕ , which measures the averaged difference between the phase at two points at locations r_1 and r_2 of the wavefront, which are separated by a distance r on the aperture plane (Fig. 1),

$$D_\phi(r) = \langle |\phi(r_1) - \phi(r_2)|^2 \rangle.$$

The structure function is related to the covariance function of ϕ , $C_\phi(r) = \langle \phi(r_1), \phi(r_2) \rangle$, as

$$D_\phi(r) = 2(\sigma_\phi^2 - C_\phi(r)), \tag{1}$$

where σ_ϕ^2 is the phase variance.

According to the Von Karman theory, the phase structure function evaluated at distance r is the following [9]:

$$D_\phi(r) = \left(\frac{L_0}{r_0}\right)^{5/3} c \left[\frac{\Gamma(5/6)}{2^{1/6}} - \left(\frac{2\pi r}{L_0}\right)^{5/6} K_{5/6}\left(\frac{2\pi r}{L_0}\right) \right],$$

where $K(\cdot)$ is the MacDonald function (modified Bessel function of the third type), Γ is the gamma function, L_0 is the outer scale, r_0 is a characteristic parameter called the Fried parameter [10], and the constant c is

$$c = \frac{2^{1/6}\Gamma(11/6)}{\pi^{8/3}} \left[\frac{24}{5}\Gamma(6/5) \right]^{5/6}.$$

From relation (1) between the structure function and the covariance, the spatial covariance of the phase between two points at distance r is

$$C_\phi(r) = \left(\frac{L_0}{r_0}\right)^{5/3} \frac{c}{2} \left(\frac{2\pi r}{L_0}\right)^{5/6} K_{5/6}\left(\frac{2\pi r}{L_0}\right). \tag{2}$$

We denote with $\phi(u, v, t)$ a discrete square phase screen of size $m \times m$ pixels, being $1 \leq u, v \leq m$, as seen by the telescope pupil at time t ; and (u, v) are the Cartesian coordinates of a point on the square that inscribes the aperture plane. Without loss of generality, we assume that the

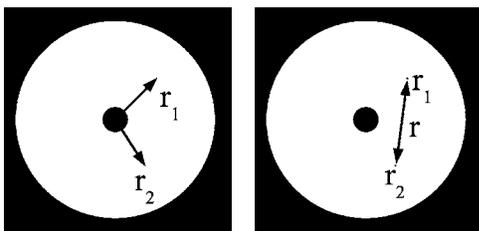


Fig. 1. Two points r_1 and r_2 at distance r on the aperture plane.

physical dimension of each pixel is $p_s \times p_s [m^2]$ (therefore the phase screen has a physical size of $D = mp_s$ in meters), although the procedure described can be easily extended to the general case of rectangular pixels.

In order to describe its temporal characteristics, the turbulence is generally modeled as the superposition of a finite number l of thin layers: The i th layer models the atmosphere from h_{i-1} to h_i meters high, where $h_l \geq \dots \geq h_i \geq h_{i-1} \geq \dots \geq h_0 = 0$. Let $\psi_i(u, v, t)$ be the value of the i th layer at point (u, v) on the telescope aperture and at time t . Then the total turbulent phase at (u, v) and at time t is

$$\phi(u, v, t) = \sum_{i=1}^l \gamma_i \psi_i(u, v, t), \tag{3}$$

where γ_i are suitable coefficients. Without loss of generality, we assume that $\sum_{i=1}^l \gamma_i^2 = 1$.

The layers are assumed to be stationary and characterized by the same spatial characteristics; i.e., all the layers are spatially described by the same structure function. The generalization to the case of layers with different spatial characteristics, e.g., different Fried parameters, is immediate. Furthermore, the layers are assumed to be zero mean and independent; hence

$$\begin{aligned} \mathbb{E}\{\psi_i(u, v, t)\psi_j(u', v', t')\} &= 0, \quad 1 \leq i \leq l, \quad 1 \leq j \leq l, \\ & \quad j \neq i, \quad 1 \leq u, v \leq m, \\ & \quad 1 \leq u', v' \leq m. \end{aligned}$$

A commonly agreed-upon assumption considers that each layer translates in front of the telescope pupil with constant velocity v_i (Taylor approximation [11]); thus

$$\psi_i(u, v, t + kT) = \psi_i(u - v_{i,u}kT, v - v_{i,v}kT, t), \quad i = 1, \dots, l, \tag{4}$$

where $v_{i,u}$ and $v_{i,v}$ are the projections of the velocity vector v_i along the direction respectively parallel and orthogonal to the wind, while kT is a delay multiple of the sampling period T .

Since all layers have the same statistical characterization, hereafter we assume $l = 1$; thus $\phi(u, v, t) = \psi_1(u, v, t)$: The generalization to the case $l > 1$ follows immediately from Eq. (3), thanks to the independence of the layers. Without loss of generality, we assume that the layer translates along the direction parallel to the wind, that is, $v_{i,u} = |v_i|$ and $v_{i,v} = 0$. Under this hypothesis the turbulent phase simulation during very long exposures is obtained by generating new columns of ϕ according to the atmospheric turbulence statistics.

In this framework, the phase screen ϕ is treated as a realization of an m -dimensional stochastic process $\Phi = \{\phi_t : t \in \mathbb{N}\}$ that we assume to be wide-sense stationary. This implies that the mean function $m_\phi(t) = m_\phi(t + \tau)$, $\forall \tau \in \mathbb{N}$ is constant ($m_\phi = 0$, without loss of generality) and that the correlation function, which with an abuse of notation here we indicate with $C_\phi(\cdot, \cdot)$, depends only on the

difference between the evaluation points $C_\phi(t_1, t_2) = C_\phi(t_1 + \tau, t_2 + \tau) = C_\phi(t_1 - t_2, 0)$, $\forall \tau \in \mathbb{N}$. Therefore, we consider the t th column of ϕ , ϕ_t (that is, $\phi_t = \phi(t, :, 0)$) as the value at time t of the stochastic process in the realization ϕ . Taking advantage of the stationarity of the process, hereafter we will write the correlation function as an univariate function $C_\phi(\cdot)$.

3. STOCHASTIC REALIZATION

A. Stochastic Realization Algorithm

The stochastic process Φ can be represented as the output y of a linear dynamic system in state space form, that is, $y_t = \phi_t$:

$$\begin{cases} x_{t+1} = Ax_t + Ke_t, \\ y_t = Cx_t + e_t \end{cases}, \quad (5)$$

where e_t is a zero-mean white-noise process with covariance matrix $\Sigma_e = \mathbb{E}\{e_t e_t^T\} = R \in \mathbb{R}^{m \times m}$. In Eq. (5), the state x and the output y vectors have dimensions of n and m , respectively, and $A \in \mathbb{R}^{n \times n}$, $K \in \mathbb{R}^{n \times m}$, and $C \in \mathbb{R}^{m \times n}$.

The problem of finding a set of parameters $\{A, C, K, R\}$ such that the covariances of the process y_t match a desired covariance matrix Σ_y is called a (partial) stochastic realization problem [12–19]. Actually, in this section we will present a particular case of the approach suggested in [14].

Moreover, in the specific phase screen case, the covariance of the stochastic process Φ is uniquely determined by the theoretical covariances given by Eq. (2).

We define Λ_i as the expected value of the product between two output samples y_{t+i} and y_t , $\Lambda_i = \mathbb{E}\{y_{t+i} y_t^T\}$, $i = 0, \dots, 2\nu - 1$, where ν is a design parameter in the procedure. From the structure of model (5), the calculation of the square matrices $\{\Lambda_i\}$ gives the following:

$$\begin{cases} \Lambda_1 = CG \\ \Lambda_2 = CAG \\ \vdots \\ \Lambda_{2\nu-1} = CA^{2\nu-2}G \end{cases}, \quad (6)$$

where $G = A\Sigma C^T + KR$ and $\Sigma = \mathbb{E}\{x_t x_t^T\}$.

Exploiting the Taylor approximation makes it possible to compute $\{\Lambda_i\}$. Letting η be the distance traveled in a sample period (proportional to the translation velocity), the values of Λ_i are simply obtained from the covariance function of Eq. (2), recalling the zero-mean assumption for ϕ_t . In other words,

$$\Lambda_i = \mathbb{E}\{y_{t+i} y_t^T\} = \mathbb{E}\{(\phi_{t+i} - m_\phi)(\phi_t - m_\phi)^T\} = C_\phi(i\eta).$$

The Λ_i are used to construct the following Hankel matrix (of size $\nu m \times \nu m$):

$$H := \begin{bmatrix} \Lambda_1 & \Lambda_2 & \cdots & \Lambda_\nu \\ \Lambda_2 & \Lambda_3 & \cdots & \Lambda_{\nu+1} \\ \vdots & \vdots & \ddots & \vdots \\ \Lambda_\nu & \Lambda_{\nu+1} & \cdots & \Lambda_{2\nu-1} \end{bmatrix} \quad (7)$$

$$= \begin{bmatrix} CG & CAG & \cdots & CA^{\nu-1}G \\ CAG & CA^2G & \cdots & CA^\nu G \\ \vdots & \vdots & \ddots & \vdots \\ CA^{\nu-1}G & CA^\nu G & \cdots & CA^{2\nu-2}G \end{bmatrix} \quad (8)$$

$$= \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{\nu-1} \end{bmatrix} [G \quad AG \quad \cdots \quad A^{\nu-1}G]. \quad (9)$$

Let T be the following Toeplitz matrix:

$$T = \begin{bmatrix} \Lambda_0 & \Lambda_1 & \Lambda_2 & \cdots & \Lambda_{\nu-1} \\ \Lambda_1^T & \Lambda_0 & \Lambda_1 & \ddots & \Lambda_{\nu-2} \\ \Lambda_2^T & \Lambda_1^T & \Lambda_0 & \ddots & \Lambda_{\nu-3} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \Lambda_{\nu-1}^T & \Lambda_{\nu-2}^T & \Lambda_{\nu-3}^T & \cdots & \Lambda_0 \end{bmatrix},$$

and let L be a Cholesky factor of T ; that is, L is a lower triangular matrix such that $T = LL^T$. Then we define the normalized Hankel matrix as follows:

$$\hat{H} := L^{-1}HL^{-T};$$

hence

$$H = L\hat{H}L^T. \quad (10)$$

Conversely, the \hat{H} matrix can be factorized according to the singular-value decomposition (SVD) algorithm:

$$\hat{H} = USV^T = US^{1/2}S^{1/2}V^T, \quad (11)$$

where U , V are unitary matrices and S is the diagonal matrix whose elements are the singular values of \hat{H} .

In a practical application of the method, most of the singular values of \hat{H} will be close to zero (Fig. 2); therefore we can use the factorization of \hat{H} even as a dimensional

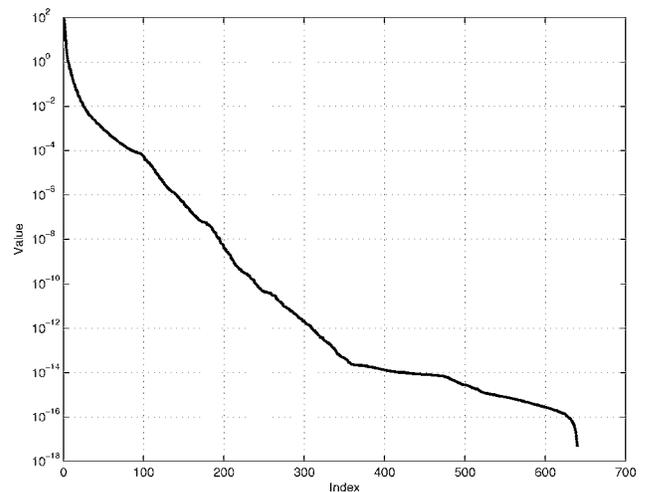


Fig. 2. Plot of the singular values of the stochastic realization model. In this case we set the parameter values to $\nu = 10$, $m = 64$; hence the size of the A matrix before the reduction step (and the number of the singular values) is $\nu m = 640$.

reduction step considering only the first \bar{n} singular values and setting the remaining ones to 0:

$$\hat{H} \approx U_{\bar{n}} S_{\bar{n}} V_{\bar{n}}^T = U_{\bar{n}} S_{\bar{n}}^{1/2} S_{\bar{n}}^{1/2} V_{\bar{n}}^T, \quad (12)$$

where

$$\begin{cases} U_{\bar{n}} = U(:, 1:\bar{n}) \\ S_{\bar{n}} = S(1:\bar{n}, 1:\bar{n}) \\ V_{\bar{n}} = V(:, 1:\bar{n}) \end{cases}.$$

In this case, the following approximate relation stands:

$$\hat{H} \approx U_{\bar{n}} S_{\bar{n}} V_{\bar{n}}^T. \quad (13)$$

From Eqs. (10) and (13) and since the factorization in Eq. (9) still holds, we can compute C and G as follows:

$$\begin{cases} C \approx \rho_1(H) L^{-T} V_{\bar{n}} S_{\bar{n}}^{-1/2} \\ G \approx (\rho_1(H^T) L^{-T} U_{\bar{n}} S_{\bar{n}}^{-1/2})^T \end{cases}, \quad (14)$$

where the $\rho_1(\cdot)$ operator selects the first m rows of a matrix.

Furthermore let $\sigma(\cdot)$ be the shift operator that, when applied to the Hankel matrix H , yields

$$\sigma(H) = \begin{bmatrix} \Lambda_2 & \Lambda_3 & \dots & \Lambda_{\nu+1} \\ \Lambda_3 & \Lambda_4 & \dots & \Lambda_{\nu+2} \\ \vdots & \vdots & \ddots & \vdots \\ \Lambda_{\nu+1} & \Lambda_{\nu+2} & \dots & \Lambda_{2\nu} \end{bmatrix}.$$

From Eqs. (10) and (13) and

$$\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{\nu-1} \end{bmatrix} A [G \quad AG \quad \dots \quad A^{\nu-1}G] = \sigma(H),$$

we can compute A in the following way:

$$A \approx S_{\bar{n}}^{-1/2} U_{\bar{n}}^T L^{-1} \sigma(H) L^{-T} V_{\bar{n}} S_{\bar{n}}^{-1/2}. \quad (15)$$

From the system equations [see Eq. (5)], it is possible to write the time evolution of $\Sigma_t = E\{x_t x_t^T\}$:

$$\Sigma_{t+1} = A \Sigma_t A^T + (G - A \Sigma_t C^T) R^{-1} (G - A \Sigma_t C^T)^T,$$

and the steady-state covariance matrix Σ is obtained by solving the following algebraic Riccati equation (ARE):

$$\Sigma = A \Sigma A^T + (G - A \Sigma C^T) (\Lambda_0 - C \Sigma C^T)^{-1} (G^T - C \Sigma A^T), \quad (16)$$

where the input noise covariance R is computed explicitly from $\Lambda_0 - C \Sigma C^T$. Let us assume that the ARE admits at least a positive semidefinite solution: The problem of the existence of such a solution will be considered in the following paragraphs. Also notice that the ARE may have multiple positive semidefinite solutions. However, there always exists two special positive semidefinite solutions Σ_- and Σ_+ such that $\Sigma_- \leq \Sigma_s \leq \Sigma_+$, where Σ_s is a generic positive semidefinite solution. Here we choose $\Sigma = \Sigma_-$, which corresponds to considering the casual factorization of the spectrum associated to the system.

Finally, the input gain K in the state equation is given by the Kalman gain: $K = (G - A \Sigma C^T) R^{-1}$.

For a generic triplet $\{A, C, G\}$, the Riccati equation, Eq. (16), may not have a solution. To explain when this may occur, let us first consider the finite covariance sequence:

$$\{\bar{\Lambda}_0, \bar{\Lambda}_1, \bar{\Lambda}_2, \dots, \bar{\Lambda}_{2\nu-1}\}, \quad (17)$$

where the matrices in the sequence are defined as follows:

$$\begin{cases} \bar{\Lambda}_0 := \Lambda_0 \\ \bar{\Lambda}_1 := CG \approx \Lambda_1 \\ \bar{\Lambda}_2 := CAG \approx \Lambda_2 \\ \vdots \\ \bar{\Lambda}_{2\nu-1} := CA^{2\nu-2}G \approx \Lambda_{2\nu-1} \end{cases}.$$

Then let us consider the infinite sequence

$$\{\bar{\Lambda}_0, \bar{\Lambda}_1, \bar{\Lambda}_2, \dots, \bar{\Lambda}_{2\nu-1}, \bar{\Lambda}_{2\nu}, \dots\} \quad (18)$$

of $m \times m$ matrices, obtained by defining

$$\bar{\Lambda}_i := CA^{i-1}G, \quad \forall i \geq 2\nu.$$

The sequence Eq. (18) is called a minimal rational extension of the finite sequence in Eq. (17) [16]. Notice that the minimal rational extension of Eq. (17) is uniquely determined by $\{A, C, G\}$. The matrices of the sequence in Eq. (18) are supposed to be the covariances of the output process in the dynamic system of Eq. (5); however, for a generic triplet $\{A, C, G\}$ satisfying $\bar{\Lambda}_i := CA^{i-1}G$, $1 \leq i \leq 2\nu-1$, Eq. (18) is not a covariance sequence. When Eq. (18) is a covariance sequence, it is called a positive sequence.

The following proposition holds:

Proposition 1. Let $\Lambda_i = C_\phi(i, \eta)$, $\forall i$ and let A, C, G be computed as in Eqs. (14) and (15). Then, there is an integer $\nu_1 \geq 2$ such that, for $\nu \geq \nu_1$ then $\{\bar{\Lambda}_0, \bar{\Lambda}_1, \bar{\Lambda}_2, \dots\}$ is a positive sequence.

The proof of Proposition 1 follows immediately from Theorem 5.3 in [14] after introducing the hypotheses that hold here.

We stress the fact that the positivity of the covariance sequence is a sufficient condition for the solvability of the Riccati equation, Eq. (16): Hence making ν sufficiently large assures the existence of a positive semidefinite solution of the ARE.

The dynamic model in Eq. (5) can be now used to synthesize new realizations of the stochastic process ϕ (or to extend in time an existing one). Indeed, given an initial state x_0 , the synthesis of new values of y is obtained by simply generating suitable samples of the input e_t and updating the state and output equations in Eq. (5). In accordance with Roddier [20], we assume that the turbulent phase has Gaussian statistics: Thus we generate e_t , for all t , taking independent samples from $\mathcal{N}(0, R)$.

Let us consider the state update equation

$$x_{t+1} = Ax_t + Ke_t. \quad (19)$$

From Theorem 13.0.1 in Meyn and Tweedie [21], whatever the initial condition x_0 is, the state probability will converge to the invariant density $\pi(\cdot)$, uniquely associated to the Markov chain described by Eq. (19). However, we can sample x_0 directly from π : In this way $p_{x_t}(\bar{x}) = \pi(\bar{x})$, $t \geq 0$, where $p_{x_t}(\cdot)$ is the state density at time t . Thus, at least theoretically, by sampling x_0 from π , we can directly sample from the dynamic system at steady state.

B. Alternative Stochastic Realization Algorithm

In Subsection 3.A we considered a general stochastic realization algorithm to compute the parameters $\{A, C, K, R\}$ of the dynamic model in Eq. (5). Taking into account our particular application, we want to reduce, as much as possible, the on-line computational complexity (off-line complexity is not a relevant issue.)

Similar to what is detailed in Subsection 3.A, we factorize \hat{H} using the SVD; however, in this case we consider the unnormalized Hankel matrix, i.e., $L=I$, $\hat{H}=H$, and

$$\hat{H} = H = USV^T = US^{1/2}S^{1/2}V^T. \quad (20)$$

Then the steps to follow for the identification of the parameters of Eq. (5) are the same as in the previous section.

For a fixed state dimension \bar{n} , this procedure does not assure the solvability of the ARE, Eq. (16). However, when the ARE is solvable, it usually allows us to achieve better performances than those of the previous section; i.e., it assures a better approximation of the theoretical covariances. Equivalently, one can obtain the same performances of the algorithm of the previous section but with a smaller \bar{n} , hence reducing the on-line computational complexity of the algorithm.

However, since in this case the Riccati equation may have no solution, it may be necessary to make a different choice for the state dimension \bar{n} and to test again the solvability of the ARE. Hence, in this case, only the off-line complexity of the algorithm is increased.

In Fig. 3 we report a comparison between the results on the replication of the theoretical structure function, obtained with the method proposed in Subsection 3.A, and those of this section. For both methods we set $\bar{n}=60$. As already claimed, when the ARE is solvable the method proposed in this section achieves better performances than those of Subsection 3.A. For this reason in Section 6 we report the results obtained with the method described in this section.

4. ASSÉMAT *et al.* METHOD

To validate the method and assess the performance of the procedure adopted, a recent work by Assémat *et al.* [8] is chosen as a reference. In [8] the problem of extending in time a phase screen of $m \times m$ pixels is considered. This, again, translates into the problem of adding new columns to the phase screen matrix. The solution proposed starts from N “old” phase values piled to form a vector z (of size

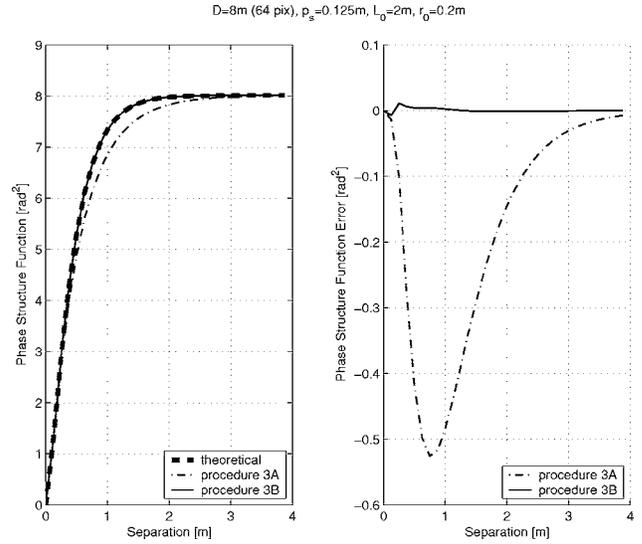


Fig. 3. Phase structure function along the wind direction. A comparison of the theoretical values (dashed curve) and those obtained with (i) the dynamic model identified with the procedure in Subsection 3.A (dashed-dotted curve) and (ii) the dynamic model identified with the procedure in Subsection 3.B (solid curve). The values of the parameters are set to $L_0=2$ m, $r_0=0.2$ m, $D=8$ m, $p_s=0.125$ m.

Nm) and a random input vector β whose components are independent Gaussian signals with zero mean and unitary covariance, which are linearly combined in a dynamic relation to form the “new” phase values y :

$$y = \tilde{A}z + \tilde{B}\beta, \quad (21)$$

where \tilde{A} and \tilde{B} are matrices of size $m \times Nm$ and $m \times m$, respectively.

To obtain the system matrices \tilde{A} and \tilde{B} , Assémat *et al.* proceed by taking the covariances

$$\Sigma_{yz} := E\{yz^T\} = \tilde{A}E\{zz^T\}, \quad (22)$$

$$\Sigma_y := E\{yy^T\} = \tilde{A}E\{zz^T\}\tilde{A}^T + \tilde{B}\tilde{B}^T. \quad (23)$$

From Eq. (22), with $\Sigma_z := E\{zz^T\}$,

$$\tilde{A} = \Sigma_{yz}\Sigma_z^{-1},$$

while from Eq. (23)

$$\tilde{B}\tilde{B}^T = \Sigma_y - \tilde{A}\Sigma_z\tilde{A}^T,$$

and hence the \tilde{B} matrix can be obtained, for example, by resorting to the SVD algorithm.

This approach can be revisited as a particular case of the stochastic realization problem. Let ν be equal to N . By assuming the notation of Section 3, ϕ_t [y in Eq. (21)] is considered as the output y_t of the following dynamic model, and the state x_t is obtained by piling the vectors $\{\phi_t, \phi_{t-1}, \dots, \phi_{t-\nu+1}\}$:

$$\begin{cases} x_{t+1} = Ax_t + Bw_t \\ y_t = Cx_t \end{cases}, \quad (24)$$

where w_t is a white-noise process with unitary covariance. Because m is the dimension of the output and $n = \nu m$ the state dimension, the process matrices $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$, and $C \in \mathbb{R}^{m \times n}$ take the form

$$A = \begin{bmatrix} \tilde{A}_1 & \tilde{A}_2 \\ I_{(\nu-1)m} & 0 \end{bmatrix} = \begin{bmatrix} \tilde{A} \\ I_{(\nu-1)m} 0 \end{bmatrix},$$

$$B = \begin{bmatrix} \tilde{B} \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

$$C = [I_m \quad 0 \quad \dots \quad 0],$$

noting that, for the sake of simplicity, the first m rows of A can be compacted in the $m \times n$ matrix \tilde{A} , and B is partitioned accordingly (with \tilde{B} of size $m \times m$).

Let the output covariances Λ_i be defined as in Eq. (6); then the state covariance matrix Σ is

$$\Sigma = \begin{bmatrix} \Lambda_0 & \Lambda_1 & \dots & \Lambda_{\nu-1} \\ \Lambda_1 & \Lambda_0 & \dots & \Lambda_{\nu-2} \\ \vdots & \vdots & \ddots & \vdots \\ \Lambda_{\nu-1} & \Lambda_{\nu-2} & \dots & \Lambda_0 \end{bmatrix}.$$

As suggested in Eq. (8), \tilde{A} can easily be computed via least squares:

$$\tilde{A} = [\Lambda_1 \quad \Lambda_2 \quad \dots \quad \Lambda_\nu] \Sigma^{-1}.$$

Moreover, since the process is assumed to be stationary, introducing matrix $Q := \tilde{B}\tilde{B}^T$ results in

$$\Sigma = A \Sigma A^T + B B^T = \begin{bmatrix} \tilde{A} \\ I_{(\nu-1)m} 0 \end{bmatrix} \Sigma \begin{bmatrix} \tilde{A}^T & I_{(\nu-1)m} \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} Q & 0 \\ 0 & 0 \end{bmatrix}.$$

Thus $Q = \Lambda_0 - \tilde{A} \Sigma \tilde{A}^T$; \tilde{B} (hence, B) can be computed from Q , for example, via SVD.

The synthesis process is substantially the same as that previously described in Section 3.

5. ZERNIKE REPRESENTATION OF TURBULENCE

In order to compare the performances of the models in Sections 3 and 4, we introduce here the Zernike representation of turbulence, which provides a low-order represen-

tation of the signal. Furthermore, the atmospheric turbulence has been statistically characterized, exploiting the Zernike representation.

One of the tests that will be used in Section 6 to compare phase screen simulation methods is the ability to reproduce the theoretical variances of the Zernike coefficients.

In this section we briefly introduce Zernike polynomials and recall some results on the statistical characteristics of their coefficients in the atmospheric turbulence framework.

Since the Zernike polynomials provide a spatial representation of the turbulence, in this section we will consider time as fixed at a constant value \bar{t} , and we will omit \bar{t} from the notation.

Zernike polynomials. Zernike polynomials are commonly used to represent signals defined inside a circle. This makes them particularly well suited to represent the turbulent phase on the aperture plane.

Let $r \in \mathbb{R}^2$ and γ be its phase, i.e., $r = |r| \exp(j\gamma)$. Then the generic Zernike polynomial Z_i , $i \geq 0$, is defined on \mathbb{R}^2 as follows:

$$Z_i(r) = \begin{cases} \sqrt{n+1} R_n^m(r) \sqrt{2} \cos(m\gamma) & \text{if } m \neq 0, i \text{ even} \\ \sqrt{n+1} R_n^m(r) \sqrt{2} \sin(m\gamma) & \text{if } m \neq 0, i \text{ odd} \\ \sqrt{n+1} R_n^m(r) & \text{if } m = 0 \end{cases},$$

where

$$R_n^m(r) = \sum_{k=0}^{(n-m)/2} \frac{(-1)^k (n-k)!}{k! \left(\frac{n+m}{2} - k\right)! \left(\frac{n-m}{2} - k\right)!} |r|^{n-2k}$$

and n, m are two integers uniquely identified by i . Notice that n, m defined in this paragraph have a different meaning from those of n, m used in the other sections. The integer n , with $n \geq 0$, is called the level of the polynomial. We use here the Noll convention [22]; however, some authors use different conventions for the relation n, m , and i . Some examples of Zernike polynomials are provided by [22].

Using the Zernike polynomials as a spatial basis, the effect of the turbulence at point r on the aperture plane can be written as follows:

$$\phi(r) = \sum_{i=0}^{+\infty} a_i Z_i \left(\frac{r}{D/2} \right), \quad |r| \leq D/2,$$

where D is the telescope aperture diameter.

Since the Zernike polynomials are orthogonal in the considered region, $a_i, i \geq 0$ can be computed from the inner product of the i th Zernike polynomial with the current turbulent phase on the aperture plane:

$$a_i = \int_{\mathbb{R}^2} \Pi \left(\frac{r}{D/2} \right) Z_i \left(\frac{r}{D/2} \right) \phi(r) dr.$$

Finally, we report the (second-order) statistical characterization of the Zernike coefficients: The turbulent phase has zero mean; hence the coefficient $a_i, i \geq 0$ has zero mean too; furthermore,

$$\mathbb{E}\{a_i a_{i'}\} = \begin{cases} = \frac{2\Gamma(11/6)}{\pi^{3/2}} \left[\frac{24}{5} \Gamma\left(\frac{6}{5}\right) \right]^{5/6} \left(\frac{D}{r_0}\right)^{5/3} \sqrt{(n+1)(n'+1)} (-1)^{(n+n'-2m)/2} \\ \times \delta_{mm'} \sum_{h=0}^{\infty} \frac{(-1)^h}{h!} \left\{ (\pi D f_0)^{2h+n+n'-5/3} \right. \\ \times \Gamma \left[h+1 + \frac{n+n'}{2}, h+2 + \frac{n+n'}{2}, h+1 + \frac{n+n'}{2}, \frac{5}{6} - h - \frac{n+n'}{2} \right] \\ \left. + (\pi D f_0)^{2h} \Gamma \left[\frac{n+n'}{2} - h - \frac{5}{6}, h + \frac{7}{3}, h + \frac{17}{6}, k + \frac{11}{6} \right] \right\} \\ \text{if } m = m', m \neq 0, m' \neq 0, i + i' \text{ even; or } m = m' = 0 \\ = 0 \text{ otherwise} \end{cases} \quad (25)$$

The above expression is derived in [23]. Other similar expressions were computed also by Takato and Yamaguchi in [24] and by Winker in [25].

6. SIMULATIONS

We report here some examples of the application of the proposed method, comparing the results with those obtained using the method of [8]. The results of the stochastic realization approach that we provide are obtained from simulations using the simplified procedure of Subsection 3.B. However the procedure of Subsection 3.A yields similar results.

We have to stress that the simulated phase screens have to reconstruct with a high level of accuracy the theoretical statistics of the turbulence in order to be of use, for instance, in the validation of the AO control procedure. Thus, we compare the methods to generate long-exposure phase screens with respect to their capability of reproducing both the structure function and the Zernike coefficient variances.

As far as the first aspect is concerned, we consider the asymptotic structure function. As explained in [21], a unique invariant density π is associated with the system in Eq. (5), characterized by its parameters $A, C, K,$ and R . Similarly, a unique invariant density is associated also with the dynamic system in Eq. (24). We assume to start simulating the turbulence at $t=t_0$. Then asymptotically for $t \rightarrow \infty$, the output density $p_{y_t}(\cdot)$ of the system in Eq. (5) converges to the invariant density π_y . Hence, we first compute the invariant density π_y , and then we use it to evaluate the corresponding structure function.

In order to provide a complete comparison between the two methods, we consider also the variances of the Zernike coefficients: In this case we compare the theoretical variances given by Eq. (25) with the sample variances estimated by sequences of 15,000 consecutive phase

screens (with wind velocity set to 4 pixels/frame). In this case the results are not asymptotic, and thus they are less accurate.

Since by hypothesis the structure function is spatially isotropic and by construction both the method of [8] and the stochastic realization approach preserve the original statistics along the direction orthogonal to the wind (see Fig. 4), most of the following examples on structure function reconstruction will show the results obtained along the direction parallel to the wind to verify the isotropic property of the structure function.

Following the guidelines for the choice of ν suggested in [8], in the examples reported we set $2 \leq \nu \leq 4$ for the method of [8]. Accordingly, the corresponding dimension

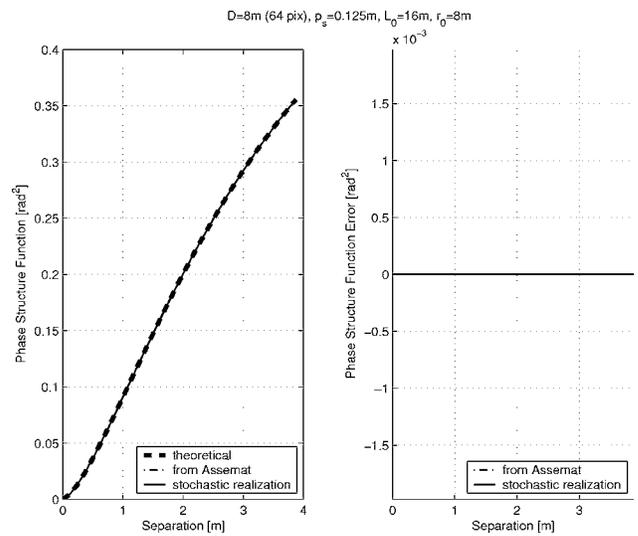


Fig. 4. Phase structure function along the direction orthogonal to the wind. A comparison of the theoretical values (dashed curve) and those obtained with (i) the dynamic model of Section 3 (solid curve) and (ii) the method of Assémat *et al.* (dashed-dotted curve). The values of the parameters are set to $L_0=16$ m, $r_0=8$ m, $D=8$ m, $p_s=0.125$ m.

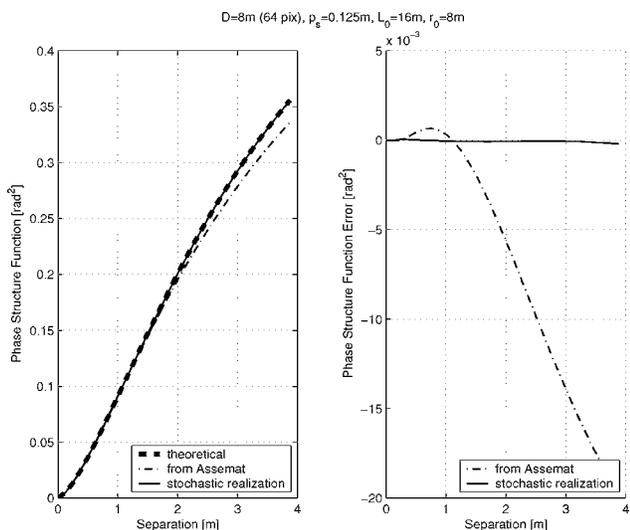


Fig. 5. Phase structure function along the wind direction. A comparison of the theoretical values (dashed curve) and those obtained with (i) the dynamic model of Section 3 (solid curve) and (ii) the method of Assémat *et al.* (dashed–dotted curve). The values of the parameters are set to $L_0=16$ m, $r_0=8$ m, $D=8$ m, $p_s=0.125$ m.

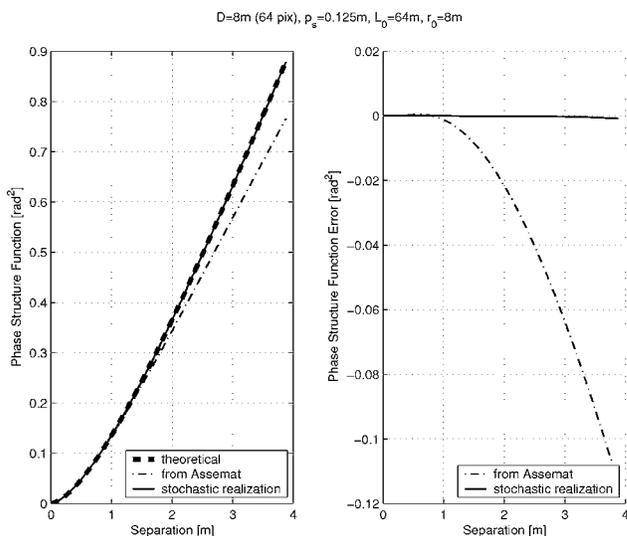


Fig. 7. Phase structure function along the wind direction. A comparison of the theoretical values (dashed curve) and those obtained with (i) the dynamic model of Section 3 (solid curve) and (ii) the method of Assémat *et al.* (dashed–dotted curve). The values of the parameters are set to $L_0=64$ m, $r_0=8$ m, $D=8$ m, $p_s=0.125$ m.

of the state is between 128 and 256. Instead, when using the procedure of Subsection 3.B, we set $\nu=10$, and the state dimension is $\tilde{n}=60$.

First, we propose three examples with parameters taken from [8]. In Figs. 4–6 we report, respectively, the structure function evaluated along the direction orthogonal to the wind, the structure function evaluated along the wind direction, and the variances of the Zernike coefficients obtained by setting the values of the parameters to $L_0=16$ m, $r_0=8$ m, $D=8$ m, $p_s=0.125$ m, and $N=2$. Then Figs. 7–10 show the structure function along the wind direction and the variances of the Zernike coefficients obtained by setting first $L_0=16$ m, $r_0=8$ m, D

$=8$ m, $p_s=0.125$ m, and $N=2$ and then $L_0=64$ m, $r_0=4$ m, $D=4$ m, $p_s=0.0625$ m, and $N=4$.

To conclude, in the last two examples we set the values of the parameters to $L_0=3.5$ m, $D=8$ m, $r_0=0.3$ m, $p_s=0.125$ m, and $N=3$ in Fig. 11, and to $L_0=1.6$ m, $D=8$ m, $r_0=0.15$ m, $p_s=0.125$ m, and $N=3$ in Fig. 12.

Notice that in the right plot of Figs. 5, 7, and 9, the error in the reconstruction of the structure function for the method of Assémat *et al.* appears to be diverging. However, this is not the case: Indeed, under the assumption of stable models that correctly represent the phase variance, the error vanishes when it is evaluated at a large distance.

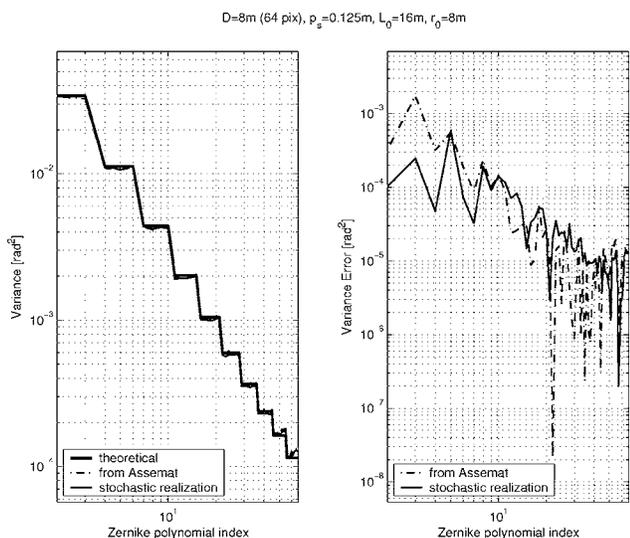


Fig. 6. Variances of the Zernike coefficients. A comparison of the theoretical values (dashed curve) and those obtained with (i) the dynamic model of Section 3 (solid curve) and (ii) the method of Assémat *et al.* (dashed–dotted curve). The values of the parameters are set to $L_0=16$ m, $r_0=8$ m, $D=8$ m, $p_s=0.125$ m.

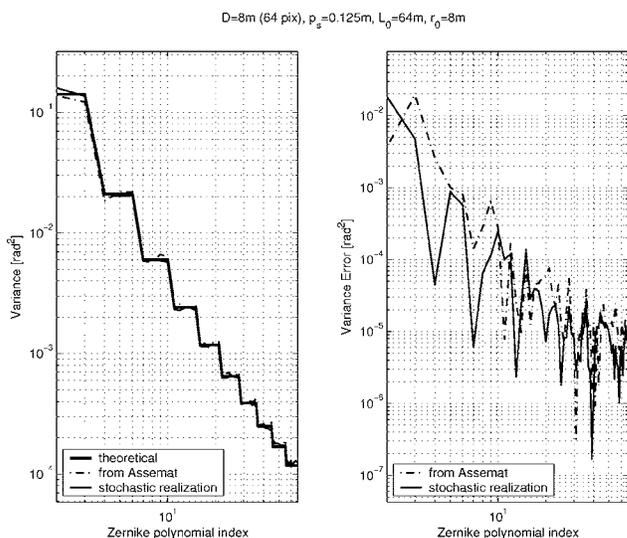


Fig. 8. Variances of the Zernike coefficients. A comparison of the theoretical values (dashed curve) and those obtained with (i) the dynamic model of Section 3 (solid curve) and (ii) the method of Assémat *et al.* (dashed–dotted curve). The values of the parameters are set to $L_0=64$ m, $r_0=8$ m, $D=8$ m, $p_s=0.125$ m.

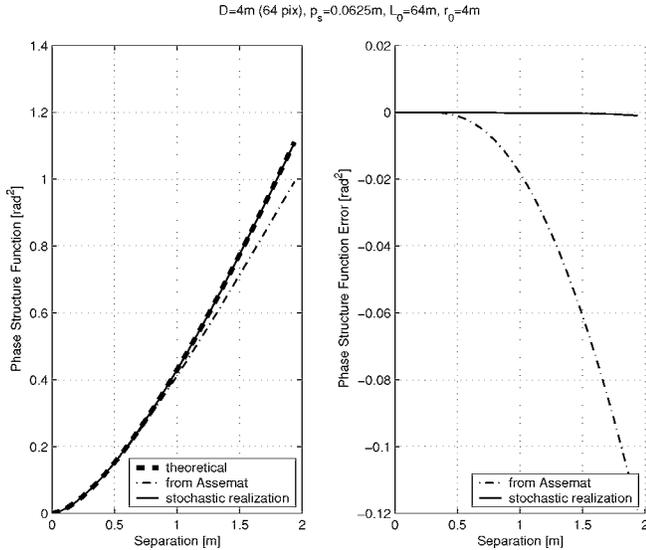


Fig. 9. Phase structure function along the wind direction. A comparison of the theoretical values (dashed curve) and those obtained with (i) the dynamic model of Section 3 (solid curve) and (ii) the method of Assémat *et al.* (dashed–dotted curve). The values of the parameters are set to $L_0=64$ m, $r_0=4$ m, $D=4$ m, $p_s=0.0625$ m.

7. DISCUSSION

To begin with, we stress the fact that the methods described in the previous sections can be successfully employed if the (wide-sense) stationarity assumption on the process ϕ stands. Furthermore, the synthesis procedure requires the A matrix in the identified model to be asymptotically stable: The procedure of Subsection 3.A ensures it, whereas this is generally not true for that in [8] (Section 4). When the stationarity assumption holds, it is simple to compute the asymptotic characteristics of both Eq. 5 and the model in Eq. (24) proposed in [8].

Two more observations are in order. First, the number of operations needed to compute a new column of the

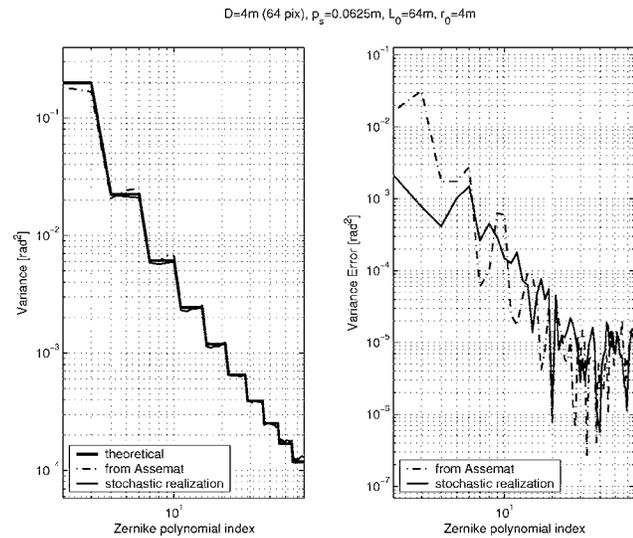


Fig. 10. Variances of the Zernike coefficients. A comparison of the theoretical values (dashed curve) and those obtained with (i) the dynamic model of Section 3 (solid curve) and (ii) the method of Assémat *et al.* (dashed–dotted curve). The values of the parameters are set to $L_0=64$ m, $r_0=4$ m, $D=4$ m, $p_s=0.0625$ m.

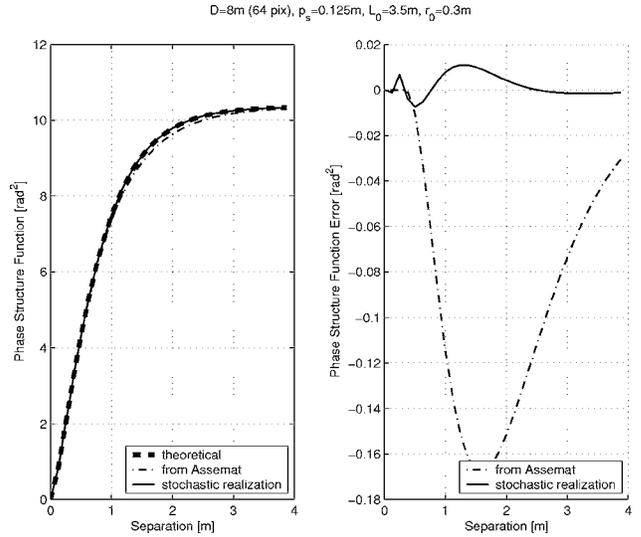


Fig. 11. Phase structure function along the wind direction. A comparison of the theoretical values (dashed curve) and those obtained with (i) the dynamic model of Section 3 (solid curve) and (ii) the method of Assémat *et al.* (dashed–dotted curve). The values of the parameters are set to $L_0=3.5$ m, $r_0=0.3$ m, $D=8$ m, $p_s=0.125$ m.

phase screen is equal to that required for sampling the new white noise e_t and for updating the state x_t and the output y_t . Since the dimensions of $\{A, C, K, R\}$, the matrices and vectors involved in the computations of x_t and e_t , depend on the size n of the state vector, it is understandable how it is critical to keep the state dimension as small as possible. To be more precise, let n_s and n_a be, respectively, the state dimensions for the procedures of Subsection 3.B (or 3.A) and Section 4, then the computational complexity is proportional to, respectively, $(m^2+n_s^2+2n_s m+n_s+2m)$ and $(m^2+m n_a+2m)$, where we have assumed that each elementary operation has the same com-

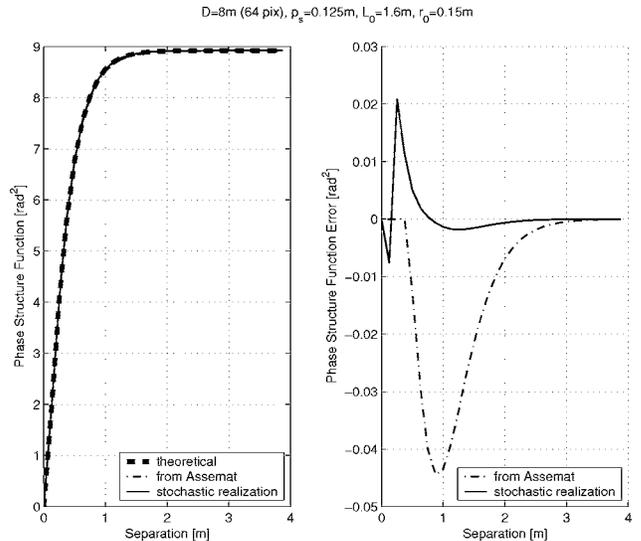


Fig. 12. Phase structure function along the wind direction. A comparison of the theoretical values (dashed curve) and those obtained with (i) the dynamic model of Section 3 (solid curve) and (ii) the method of Assémat *et al.* (dashed–dotted curve). The values of the parameters are set to $L_0=1.6$ m, $r_0=0.15$ m, $D=8$ m, $p_s=0.125$ m.

plexity (even the generation of a random number). Since $n_a = Nm$ and reasonably $n_s < m$, then the stochastic realization approach requires approximatively $4/(N+1)$ times the number of operations needed by the method of Assémat *et al.* Thus the two algorithms have similar computational complexities for small values of N . To be more precise, the method of Assémat *et al.* is computationally convenient for $N = \{1, 2\}$ (short memory system), while the stochastic realization approach becomes convenient for N larger than 3 (long memory system). Similar considerations can be made also for the memory requirement of the two algorithms.

Second, the parameter ν in both models [Eqs. (5) and (24)] corresponds to the number of covariances used in the model identification step: Large values of ν lead to better approximations of the dynamic behavior of the process. Therefore, it would be sensible to choose a large value of ν .

As far as the comparison between the stochastic realization approach (Section 3) and the original approach in [8] (Section 4) is concerned, we observe that the state vector dimension in the model of Eq. (24) is $n = \nu m$: The state dimension grows linearly with ν , and therefore there is a trade-off between the two issues mentioned before. For the state vector to show reasonable dimension, the ν parameter has to be kept small.

Conversely, one main advantage of the approach outlined in Section 3 is that we can choose n and ν separately and, thanks to the dimension reduction step in the SVD factorization of H in Eq. (11), the resulting state dimension \bar{n} will be smaller than νm .

The above considerations suggest that the method described in Section 3 provides an overall improvement over the previous method proposed in [8]. This is confirmed by the results obtained in the examples reported in Section 6. In these examples we used a much smaller state for the method of Section 3 with respect to that of [8]. On one hand, this makes the running time of the algorithm (and its memory requirements) comparable with that of [8]. On the other hand, it is evident how the output of the proposed algorithm allows us to obtain better results, especially in terms of estimation of the structure function, thanks to the larger value of ν .

Since unfortunately in practical applications the stationarity hypothesis is not typically satisfied, it is worth considering the case of nonstationary turbulence simulation. Similar to the model of Assémat *et al.*, the method described in Section 3 can also handle this case. When the nonstationarity is given by abrupt changes in r_0 , the system parameters can be easily updated. On the other hand, if the system is affected by a change in L_0 instead of in r_0 , the nonstationarity can still be handled, however the model matrices have to be recomputed following the procedure described in Section 3.

8. CONCLUSIONS

In this paper we have presented what we believe to be a new framework to develop a dynamic model used to extend the phase screen for astronomical applications.

On the one hand, we have shown how the stochastic realization approach is consistent with previous work, in

that the model by Assémat *et al.* is reinterpreted in the general framework proposed.

On the other hand, the model produced using the stochastic realization shows appealing properties of compactness, since the state dimension results are much smaller than the correspondent one in [8], and at the same time provides better results in terms of the reconstructed structure function.

ACKNOWLEDGMENTS

We acknowledge the colleagues of the ELT Project, in particular Michel Tallon at CRAL-Lyon and Enrico Fedrigo at ESO-Munich, for their precious help in supporting us with the astronomical view of the problem and for many valuable and enjoyable discussions on the subject. We also thank the reviewers for their constructive suggestions. This work forms part of the ELT Design Study and is supported by the European Commission within Framework Programme 6, contract 011863.

REFERENCES

1. ESO, "The Very Large Telescope Project," <http://www.eso.org/projects/vlt/>.
2. M. Le Louarn, N. Hubin, M. Sarazin, and A. Tokovinin, "New challenges for adaptive optics: extremely large telescopes," *Mon. Not. R. Astron. Soc.* **317**, 535–544 (2000).
3. P. Dierickx, J. L. Beckers, E. Brunetto, R. Conan, E. Fedrigo, R. Gilmozzi, N. Hubin, F. Koch, M. Le Louarn, E. Marchetti, G. Monnet, L. Noethe, M. Quattri, M. Sarazin, J. Spyromillo, and N. Yaitskova, "The eye of the beholder: designing the OWL," *Proc. SPIE* **4840**, 151–170 (2003).
4. A. N. Kolmogorov, "Dissipation of energy in the locally isotropic turbulence," *C. R. (Dokl.) Acad. Sci. URSS* **32**, 16–18 (1941).
5. A. N. Kolmogorov, "The local structure of turbulence in incompressible viscous fluid for very large Reynold's numbers," *C. R. (Dokl.) Acad. Sci. URSS* **30**, 301–305 (1941).
6. V. I. Tatarski, *Wave Propagation in a Turbulent Medium* (McGraw-Hill, 1961).
7. H. G. Booker, T. A. Ratcliffe, and D. H. Schinn, "Diffraction from an irregular screen with applications to ionospheric problems," *Philos. Trans. R. Soc. London, Ser. A* **242**, 579–609 (1950).
8. F. Assémat, R. W. Wilson, and E. Gendron, "Method for simulating infinitely long and nonstationary phase screens with optimized memory storage," *Opt. Express* **14**, 988–999 (2006).
9. A. Tokovinin, "From differential image motion to seeing," *Publ. Astron. Soc. Pac.* **114**, 1156–1166 (2002).
10. D. L. Fried, "Statistics of a geometric representation of wavefront distortion," *J. Opt. Soc. Am.* **55**, 1427–1435 (1965).
11. F. Roddier, *Adaptive Optics in Astronomy* (Cambridge U. Press, 1999).
12. M. Aoki, *State Space Modeling of Time Series*, 2nd ed. (Springer-Verlag, 1991).
13. H. P. Zeiger and A. J. McEwen, "Approximate linear realization of given dimension via Ho's algorithm," *IEEE Trans. Autom. Control* **19**, 153 (1974).
14. A. Lindquist and G. Picci, "Canonical correlation analysis, approximate covariance extension, and identification of stationary time series," *Automatica* **32**, 709–733 (1996).
15. B. L. Ho and R. E. Kalman, "Effective construction of linear state-variable models from input/output functions," *Regelungstechnik* **14**, 545–548 (1966).
16. S. Y. Kung, "A new identification and model reduction algorithm via singular value decomposition," in

- Proceedings of the 12th Asilomar Conference on Circuits, Systems and Computers* (IEEE, 1978), pp. 705–714.
17. H. Akaike, “Stochastic theory of minimal realization,” *IEEE Trans. Autom. Control* **19**, 667–674 (1974).
 18. H. Akaike, “Markovian representation of stochastic processes by canonical variables,” *SIAM J. Control* **13**, 162–173 (1975).
 19. U. B. Desai and D. Pal, “A realization approach to stochastic model reduction and balanced stochastic realizations,” in *Proceedings of IEEE Conference on Decision and Control* (IEEE, 1982), pp. 1105–1112.
 20. F. Roddier, “The effects of atmospheric turbulence in optical astronomy,” *Prog. Opt.* **19**, 281–376 (1981).
 21. S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability* (Springer-Verlag, 1993).
 22. R. J. Noll, “Zernike polynomials and atmospheric turbulence,” *J. Opt. Soc. Am.* **66**, 207–211 (1976).
 23. R. Conan, “Modélisation des effets de l’échelle externe de cohérence spatiale du front d’onde pour l’observation à Haute Résolution Angulaire en Astronomie,” Ph.D. thesis (Université de Nice-Sophia Antipolis, 2000).
 24. N. Takato and I. Yamaguchi, “Spatial correlation of Zernike phase-expansion coefficients for atmospheric turbulence with finite outer scale,” *J. Opt. Soc. Am. A* **12**, 958–963 (1995).
 25. D. M. Winker, “Effect of a finite outer scale on the Zernike decomposition of atmospheric optical turbulence,” *J. Opt. Soc. Am. A* **8**, 1568–1573 (1991).