



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
ESCUELA DE INGENIERIA

TURBULENCE ESTIMATION IN WIDE-FIELD ADAPTIVE OPTICS SYSTEMS

NICOLAS SÉBASTIEN DUBOST ALLIGIER

Tesis presentada a la Dirección de Postgrado
como parte de los requisitos para optar al grado de
Magister en Ciencias de la Ingeniería

Profesor Supervisor:
CHRISTIAN DANI GUZMÁN CARMINE

Santiago de Chile, Julio, 2015

© MMXV, NICOLAS SÉBASTIEN DUBOST ALLIGIER



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
ESCUELA DE INGENIERIA

TURBULENCE ESTIMATION IN WIDE-FIELD ADAPTIVE OPTICS SYSTEMS

NICOLAS SÉBASTIEN DUBOST ALLIGIER

Miembros del Comité:

CHRISTIAN DANI GUZMÁN CARMINE

ANDRÉS GUESALAGA

JAIME ANGUITA

AMOKRANE BERDJA

ÁLVARO SOTO

Tesis presentada a la Dirección de Postgrado
como parte de los requisitos para optar al grado de
Magister en Ciencias de la Ingeniería

Santiago de Chile, Julio, 2015

© MMXV, NICOLAS SÉBASTIEN DUBOST ALLIGIER

to my parents

ACKNOWLEDGEMENTS

It is with great joy and relief that I can finally write these words. Such occasions provide an excuse to say thanks, not only to those that have helped me during this work, but also to my dearest: my friends and family. Overcoming the prolonged intellectual loitering of such a work is only possible through the meaning and ease of heart, brought by the people in one's life. And being as I have, one of my parents utmost priorities in theirs, I save them a special place in this acknowledgement. It is through your tireless support, meticulous guiding and unconditional love, that I have become the person I am today. Thus, I turn to our native spanish, to directly tell you *gracias totales a ustedes, mis padres*. You will always have the sincere gratitude of your loving son.

I also wish to thank all of my friends. So insistently (I am not pointing any fingers, Rodrigo and Lela), you convinced me of travelling with you to Coyhaique, only a week before defending this thesis.

To my postgraduate companions and friends, this thesis would not have come to be what it is without your contribution. My thanks to Nicole, my predecessor, for building BEAGLE, the optical test bench with which I would work. I gained much of my laboratory and practical experience from calibrating the bench and bringing it to operation. For this, I also have to thank Norman. As a sponge, I tried to absorb as much knowledge from you as I could. I am now a perfectly decent linux user and a proud member of a small but growing community of DARC users. This will prove to be even more useful now that I am headed to Durham. My thanks to Eduardo, my successor. Even though you do not work with adaptive optics, I ran every concept and idea through you. With the best disposition, you sat down to every single one of my countless and sometimes endless explanations. Always sceptic, proving anything to you was the best way of finding out how well I knew the things I thought I knew (?).

My special thanks to Amokrane. Our many and long conversations were more of a mentoring. I received much advice from you, most of which I did not understand until it was too late. But as you told me, the masters is a trial for the doctorate. I also owe you almost an entire chapter of this work, as well as much of the simulation data. But there can be but one author to a thesis (probably), so thanks.

At last, I wish to thank my advisor, Dani. Thank you for taking care of the boring stuff, so that we could take care of the nourishing stuff. Despite your tight schedule, you could always set some time aside to help us out with some piece of advice, or to write some of the recommendation letters we were all always asking. Whenever I got tangled in a problem using my distinctive, theoretical and tedious approach, you always set me in a sensible, pragmatic direction to solve it. Finally, you are one the main reasons I have been accepted for a PhD at Durham, to work with an outstanding human group. There I know, if anyone resembles you at all, I will be most welcomed. In all, thank you for the opportunities, thank you for the help and thank you for the caring.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iv
LIST OF TABLES	viii
LIST OF FIGURES	ix
RESUMEN	xii
ABSTRACT	xiii
1. INTRODUCTION	1
1.1. Addressing the problem of atmospheric turbulence in astronomical ground-based observations	1
1.2. Wide-field adaptive optics	3
2. CHARACTERIZING ATMOSPHERIC TURBULENCE	8
2.1. Definitions for wave-front & for slope	8
2.2. Slopes' Covariance & Power Spectral Density	9
2.3. Considering a Sub-aperture	10
2.4. Physical Model, von Karman's Power Spectral Density	13
3. WAVEFRONT RECONSTRUCTION & TOMOGRAPHIC RECONSTRUCTION	15
3.1. Wavefront reconstruction	16
3.1.1. Zonal representation	16
3.1.2. Modal representation	19
3.2. Tomographic reconstruction	24
3.2.1. <i>Learn & Apply</i>	26
3.2.2. Tomographic reconstructors based on artificial neural networks	29
4. CALIBRATION AND CHARACTERIZATION OF THE BEAGLE TEST BENCH	36
4.1. Calibration Routine	37

4.2.	Plate scale measurement	39
4.3.	Off-axis aberrations characterization through optimal filter signal comparison	42
4.3.1.	Description of the optimal filter signal comparison technique	42
4.3.2.	Application: Field dependent aberrations characterization	43
4.4.	Spot & Lenslet heterogeneity	47
4.5.	Phase screen characterization	49
5.	TOMOGRAPHIC RECONSTRUCTORS' RANGE OF VALIDITY	54
5.1.	Training an artificial neural network based tomographic reconstructor	55
5.2.	Correction of instrumental deviations: on-axis & off-axis gains	57
5.3.	Sensitivity analysis	61
6.	CONCLUSIONS	66
	BIBLIOGRAPHY	68

LIST OF TABLES

3.1	Indexes of Zernike polynomials. Possible pairs of indexes n, m are shown up to $n = 4$. Radial frequency is also referred to as mode	20
3.2	Noll’s sequence (sequence A176988 in OEIS). It maps every couple of indexes $(n, m) \rightarrow j$, allowing to concatenate Zernike coefficient in vectors and to operate on them.	20
3.3	Zernike polynomial with Noll’s normalization and sequence. Shortened version of Table I in (Noll, 1976)	21
3.4	Zernike polynomial derivative matrix γ_x . Shortened version of Table II in (Noll, 1976)	23
5.1	Atmospheric profiles for validation.	56
5.2	Reconstructor validation. The achieved reconstructor is compared against a reference. Here WFE-RMS[%] shows average WFE relative to average WF.	56
5.3	Instrumental deviations correction results. The ratio between WFE-RMS and WF-RMS is calculated for each random phase screen. Its average is shown in column $\mu_{WFE/WF}$ and its standard deviation in column $\sigma_{WFE/WF}$	60
5.4	WFE-RMS when introducing distortion matrices G . Simulation considers atmospheric profile ”atm1”. The ratio between WFE-RMS and WF-RMS is calculated for each random phase screen. Its average is shown in column $\mu_{WFE/WF}$ and its standard deviation in column $\sigma_{WFE/WF}$	63
5.5	WFE-RMS when rotating guide sources. Simulation considers atmospheric profile ”atm1”. The ratio between WFE-RMS and WF-RMS is calculated for each random phase screen. Its average is shown in column $\mu_{WFE/WF}$ and its standard deviation in column $\sigma_{WFE/WF}$	64

LIST OF FIGURES

1.1	Adaptive optics system’s architecture. Reproduced from (Tokovinin, 2001) . . .	3
1.2	Topological diagram of the light cylinders of a central star and three guide stars 30 arcseconds apart, for a 4.2 [m] telescope. Cut-throughs on the right are take at 0 [m], 5000 [m] and 10000 [m] of altitude. Reproduced from (Osborn et al., 2012)	4
2.1	Angle of arrival of a wave-front described by function $z(x, y)$, at a (x_0, y_0) . . .	8
2.2	Transversal view of a Shack-Hartmann wave-front sensor representation. The wave-front $z(x, y)$, propagates in the Z direction, down to the lenslet array. Each lenslet, of focal length f and surface S focuses an approximately flat segment of the wave-front. The offset of the resulting spots from the optical axis of the lenslet is used to calculate the angle of arrival α_j for sub-aperture j	11
3.1	Discrete integration of the wave-front’s slopes. Phase points z_j are located between sub-apertures of diameter d in Fried’s configuration.	17
3.2	Fried’s configuration for an $n \times n$ lenslet array. Phase points z_j are located between sub-apertures.	17
3.3	Telescope’s pupil, projected in two directions separated by an angle α , into a turbulence layer at an altitude h . As a 12×12 Shack-Hartmann is conjugated to the pupil, it is also projected in altitude. The overlap between both projections depends on the altitude and on the angle between the light sources. The star symbol represents a guide star (GS) and the spiral represents an elongated scientific target (ST).	25
3.4	Topology of a multi-layer perceptron back-propagation network. The activation pattern is ${}^{in}\vec{y}$, a neuron’s activation function in the hidden layer is f_H . The number of neurons in the input layer, in the hidden layer and in the output layer are n , n' and m respectively. In most cases $n = n'$	30

3.5	WFE for L&A and an ANN based reconstructor on the CANARY calibration bench, against altitude of the phase screen. The dashed line shows the expected performance of the ANN. Reproduced from (Osborn et al., 2014).	34
4.1	Block diagram of BEAGLE, a multi-object adaptive optics capable experiment. Here only the open-loop portion is shown, from the natural and laser guide stars to the Shack-Hartmann wavefront sensor, and going through the phase screens.	36
4.2	LED printed circuit board (PCB) containing BEAGLE's light sources. In the PCB's plane the light sources' positions are known in millimetres. This plane is the NGS plane.	40
4.3	Pupil location on the CCD camera as a function of the guide source location on the NGS plane. The CCD camera is big enough to contain all the pupils behind the SH-WFS. Pixel size is $7.4 \mu m$. $PS1$ is given by the slope of the curve. . .	41
4.4	Two light sources illuminating a moving turbulence under the frozen flow assumption. The pupil can be large and sampled by a Shack-Hartmann wavefront sensor, or small as a sub-aperture. In a test bench all altitudes and distances $h1, h2, d1, d2, d3$ and $d4$ are known. The angular separation between the light sources can therefore be determined from geometry.	42
4.5	Same turbulence slopes for two different sub-apertures . By determining the offset of maximum correlation between the signals, the common observation is found and compared. The amplitude of the signals is proportional to their standard deviation. The relative gain between these neighbouring sub-apertures is 1.3%	45
4.6	Field dependent gain, for the X and the Y direction. The phase screen moves along the X direction. Off-axis gain is mapped for each sub-aperture j	46
4.7	Slope standard deviation map per sub-aperture. The overall average for all sub-apertures depends on the turbulence profile. The relative differences depends on the aberrations.	48

4.8	Structure functions in the X and in the Y directions. Functions obtained from true data are compared with functions of a fitted model.	52
5.1	Guide stars constellation, with target star at center. X is the direction of wind propagation.	55
5.2	Star directions relative to wind direction in the X axis. Direction $i = 0$ is on-axis, i' is aligned with the wind, and i'' is the same distance from $i = 0$ as i' . Field dependent aberrations are considered azimuthally symmetric. For $\theta = 90^\circ$, $g_{x'} = g_{y''}$ and $g_{y'} = g_{x''}$	58
5.3	Demodulation and modulation. An ANN trained through simulation receives demodulated data from an instrument. The ANN produced an estimation of the target's slopes. The result is modulated at its end. The whole acts as an ANN trained with experimental data.	60
5.4	Average WFE-RMS, relative to average uncorrected WF-RMS vs. outer scale. Error bars equal 2σ , where σ is the standard deviation of WFE-RMS for all simulated cases, also relative to average uncorrected WF-RMS. At each value of L_0 , all altitudes are represented evenly.	65

RESUMEN

La hipótesis de este trabajo es que las desviaciones y aberraciones de un instrumento de óptica adaptativa pueden ser cuantificadas e incorporadas a una simulación, con el fin de entrenar un reconstructor tomográfico válido bajo condiciones experimentales. Con este propósito se caracterizó la banca óptica BEAGLE (Sáez, Basden, Guzmán, Dubost, & Berdja, 2014), un experimento de óptica adaptativa multi-objeto que emula a CANARY del William Herschel Telescope. Los parámetros adquiridos son empleados para ajustar una simulación de la banca óptica, la cual produce datos de validación.

Se propone una técnica para manejar desviaciones instrumentales, cuando las mediciones son realizadas utilizando un sensor de frente de onda Shack-Hartmann (SH). Dos desviaciones son tratadas: heterogeneidad de sub-aperturas y aberraciones dependientes del campo. La primera se debe a aberraciones estáticas en el eje y a imperfecciones en la construcción del Shack-Hartmann. La segunda es el producto de aberraciones ópticas fuera del eje. La técnica de corrección mitiga las desviaciones de las mediciones realizadas por el SH, antes de entregárselas a un reconstructor basado en una red neuronal artificial (ANN por su siglas en inglés). La intención es exponer a la ANN a datos normalizados, similares a aquellos producidos en simulación para su entrenamiento.

Una vez implementada, la corrección falla en entregar mejora alguna. Para lidiar con esto, se realiza un análisis de sensibilidad. Como es de esperar, la ANN se muestra robusta frente al ruido. Dado que la corrección propuesta puede ser vista por el reconstructor como ruido, su efecto es limitado. También se probó con otras fuentes de error, tales como el perfil atmosférico y el *pointing*. Cuando estos errores caen en órdenes de magnitud esperados, su influencia en la estimación del sistema es despreciable, bajo 1%. La variación más significativa de rendimiento se observa cuando se cambia un parámetro atmosférico: la escala externa. Trabajo futuro considera la realización de pruebas para todas las fuentes de error, en una simulación inclusiva, así como el entrenamiento de un reconstructor que considere incertidumbre en estos factores.

ABSTRACT

The hypothesis of this work is the deviations and aberrations of an adaptive optics instrument can be quantified, and incorporated into a simulation, to train a tomographic reconstructor valid in experimental conditions. To do so, the test bench BEAGLE (Sáez et al., 2014), a multi-object adaptive optics capable experiment that emulates CANARY at the William Herschel Telescope, is characterized. Parameters retrieved are employed to tune a simulation of the test bench, which produces validation data sets.

A technique is proposed to handle instrumental deviations, when measurements are made using a Shack-Hartmann (SH) wave-front sensor. Two deviations are addressed: sub-aperture heterogeneity and field-dependent aberrations. The first is due to on-axis static aberrations and imperfections in the Shack-Hartmann's construction. The second are the product of off-axis optical aberrations. The correction technique mitigates the deviations from the SH measurements before handing them over to an artificial neural network (ANN) based reconstructor. The intention is to present the ANN with normalized data, similar to that produced in simulation for its training.

When implemented, the correction fails to deliver any improvement. To address this, a sensitivity analysis is performed. As expected, the ANN is robust when facing noise. Because the proposed correction can be seen by the reconstructor as noise, its effects are limited. Other sources of error, such as atmospheric profile and pointing were tested as well. When these errors fall within expected orders of magnitude, their influence on the system's estimation is negligible, under 1%. The most significant variation in performance is observed when changing an atmospheric parameter: the outer scale. Further works remains to test all sources of error in an inclusive simulation and to train a new reconstructor that considers the incertitude in these factors.

Keywords: Turbulence, Adaptive, Optics, Tomography, Reconstructor.

1. INTRODUCTION

1.1. Addressing the problem of atmospheric turbulence in astronomical ground-based observations

Whenever looking at a star or at a planet with the naked eye, scintillation can be perceived. Contrary to popular belief, this is not related to the nature of the celestial object, but rather to the effects of atmospheric turbulence. Just as when watching over a road on a hot day, the hot air in motion deflects the light crossing through it, causing the illusion of motion. The deflection depends on the gradient of refraction index, which in turn depends on the gradient of temperature. The very same phenomenon is in action when looking at stars, but with less strength. During the night, the atmosphere is far from any source of heat. Temperature gradients are therefore very small, but they occur and cumulate over a large distance going from the ground, all the way up until the end of the atmosphere. The integrated effects accounts for a perceivable scintillation.

Although this effect seems negligible to the human eye with its little pupil and magnification, it has a destructive effect to larger instruments. Being small, a beam of light entering the pupil of the eye can be thought of being a single ray. When traversing through the atmosphere, this ray will change its phase and its direction. But for a large telescope, fitting thousands of eyes, each ray will change in its own direction and phase, resulting in a very distorted image. If a short exposure image of a star could be acquired, the result would not be the typically bell-shaped profile, but instead a swarm of speckles. The long exposure result is the integration of this moving swarm over its most probable area. The image of any star is therefore larger than that expected without turbulence. This limits the resolution of astronomical telescopes. The purpose of building ever larger telescopes, increasing the light collecting area, is not only to obtain brighter images, but also to improve the resolution by reducing the diffraction limit. But because of the effects of turbulence, the resolution of a giant telescope is no better than that of 25 [cm] pupil, amateur telescope.

There are two solutions to dealing with the problem at hand. The first and most expensive one is travelling outside of the atmosphere. This means launching space telescopes, such as the Hubble Space Telescope. This solution allows for diffraction limited images of high resolution, even during long exposures. This is why the Hubble Space Telescope was able to provide deep space observations of an unprecedented resolution. The problem is this is fairly expensive, and maintenance and upgrades require space missions. The second solution is to implement an adaptive optics (AO) system. AO systems are control systems that measure and mitigate the effects of atmospheric turbulence. The whole of the light entering the telescope's pupil is called the wave-front (WF). To measure the phase aberrations introduced to the WF by the turbulence, a wave-front sensor (WFS) is used. The most common kind is the Shack-Hartmann (SH) WFS, which consists of a square array of micro-lenses or lenslets. An optical system can be used to conjugate this array to the telescope's pupil. This is equivalent to deploying a large array of small cameras on the pupil, to image small incoming rays, thus sampling the WF.

Once the wave-front has been sampled, the information is reconstructed and handed over to a deformable mirror (DM). The DM adopts the shape of the aberrant wave-front. When reflecting upon it, the WF is corrected, becoming flat once again. This process must be repeated several times per second. Because the turbulence is ever changing, any measurement and correction ceases to be valid after a short period of time. An average AO system works at a frequency over 200 [Hz]. Finally, the DM must be placed before the WFS. This way the DM corrects and then the WFS sees the residual error, closing the control loop. Figure 1.1

This describes the basic architecture of a classical AO system, invented during the seventies for military applications. Since, the technology has found civil applications, not only in ground-based astronomy, but also in optical telecommunications and ophthalmology. With respect to its astronomical applications, AO has evolved into several variants. One of them, wide-field AO, is central to this work and is discussed in the following section.

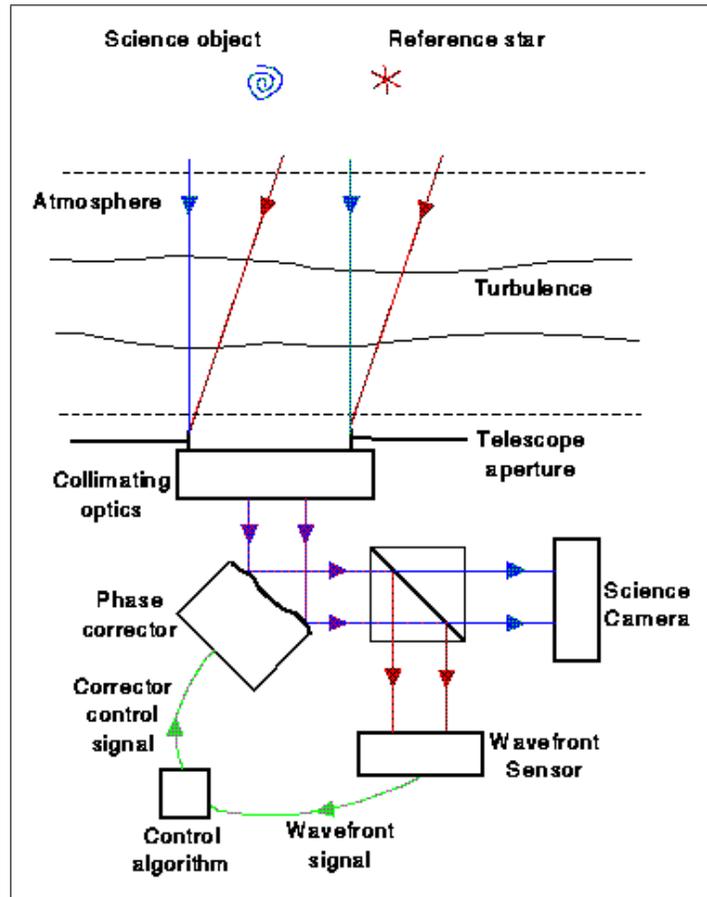


FIGURE 1.1. Adaptive optics system's architecture. Reproduced from (Tokovinin, 2001)

1.2. Wide-field adaptive optics

The light of two different stars travels across different volumes of the atmosphere. It is said that each star sees a different turbulence. Therefore, a correction that is valid for one star is not for the other. If both stars are angularly close to each other, then a correction for one remains partially valid for the other, as they see about the same turbulence. But as the stars move apart, the validity rapidly decreases to none. This is called angular anisoplanatism and is a major problem when trying to correct the aberrations seen by a target object, by measuring the turbulence on another. The phenomenon is illustrated in Figure 1.2, for four stars.

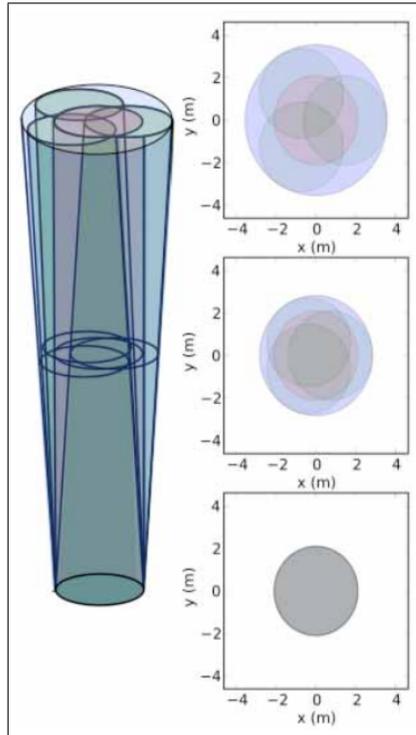


FIGURE 1.2. Topological diagram of the light cylinders of a central star and three guide stars 30 arcseconds apart, for a 4.2 [m] telescope. Cut-throughs on the right are take at 0 [m], 5000 [m] and 10000 [m] of altitude. Reproduced from (Osborn et al., 2012)

The previous case arises when observing elongated and faint targets, such as galaxies or nebulae. Because these objects are not punctual and, their shape, and therefore their incoming wave-front is not known *a priori*. On the contrary, stars are supposed to be punctual and emit flat wave-fronts. Because of this elongated targets cannot be used to sample the turbulence. In a classical AO system, if any correction is to be made upon them, it has to come from measurements of a neighbouring star. This is not always possible, as stars are required to be angularly near to the target, and of a sufficient magnitude (brightness). This calls for a new technique that is able to retrieve the aberrations of a target, from measurements of other stars.

The solution to the problem are wide-field AO (WFAO) systems. These systems make use of multiple WFS, each observing a different star, to better sample the atmosphere above the telescope. A star observed by a WFS is called a guide star, since it will be

used to sample and later correct the turbulence. When no stars are bright enough in the required field of view, artificial laser stars are launched, typically from the top or from the sides of the telescope. Both, natural guides stars (NGS) and laser guide stars (LGS) are used to illuminate volumes of the atmosphere. Each star defines a channel. When the incoming wave-front from a guide stars travels across the atmosphere, the resulting aberrant WF has integrated the effects of the volume it traversed. Therefore, only a projection of the turbulence volume can be observed at each channel. But when the channels overlap, their common volume can be retrieved from all projections by a tomographic reconstructor. In other words, the two-dimensional projection of a three-dimensional volume, allows to reconstruct the volume using tomography. There are four major techniques using tomographic reconstructors: ground layer AO (GLAO) (Tokovinin, 2004), laser tomography AO (LTAO) (Le Louarn & Hubin, 2004), multi-conjugate AO (MCAO) (Beckers, 1989) and multi-object AO (MOAO) (Assémat, Gendron, & Hammer, 2007). From the previous, MCAO and MOAO are more strictly wide-field techniques.

In the case of MOAO, the reconstructor uses the measurements to calculate the atmosphere's projection as it would be seen in a desired direction. This is equivalent to reconstructing the entire volume, and then projecting it into a direction, but with far less computations. The result can be straightforwardly fed into a controller for correction. The system is said to be multi-object because the reconstructor can estimate the projected phase aberration of multiple targets, and then correct each using individual DMs.

Two major approaches to tomography have been proposed. *Learn & Apply* (L&A) (Vidal, Gendron, & Rousset, 2010) is a linear reconstructor that uses correlation matrices calculated online with on-sky data. Each matrix is directly multiplied by a vector containing a concatenation of the phase aberrations measured for all guide sources. The resulting vector is the estimation of the phase aberration in the target direction. The alternative technique is CARMEN (Osborn et al., 2014), an artificial neural network (ANN) based reconstructor, trained using phase screens with different profiles and ranging from ground layer to an altitude of about 15500 m.

Both approaches, L&A and CARMEN, have been tested at CANARY (Morris et al., 2010; Gendron et al., 2011), the first on-sky MOAO test based at the William Herschel Telescope (WHT). Results from the implementation of L&A (Vidal et al., 2014) and CARMEN (Osborn et al., 2014) show that while achieving higher Strehl ratios when recalculated, L&A performance is highly dependent on the turbulence profile and specially on its altitude. Hence the reconstruction deteriorates over time as turbulence layers move, calling for new measurements and recalculation which take on observation time. On the other hand, because CARMEN is trained with phase screens at multiple altitudes and using various profiles, it does not show the same dependences and remains a valid reconstructor for larger periods of time. CARMEN is able to do so because an ANN can effectively include the information of multiple geometries and profiles, whereas correlation matrices strongly change with geometry.

The advantage of having a reconstructor that can be trained earlier during the week compensates for the fact that it cannot be with on-sky data previous to the observation. Artificial neural networks require large amounts of variably independent data and days to be trained. The variability will depend on the phase screens used, which are usually in short supply. Data produced via numerical simulation would not make use of facility instruments, allowing for parallel training. Through pseudo-randomness, a simulation could also guarantee a large number of independent cases, producing random phase screens at high speed. This is state of art, as both, linear and ANN based reconstructors have validated their concept and compete to be implemented in future telescopes, such as the European Extremely Large Telescope (E-ELT). Being able to train ANN based reconstructors in simulation would grant them a decisive advantage over less robust linear reconstructors.

In summary, the hypothesis is the deviations and aberrations of an instrument can be quantified, and incorporated into a simulation, to train a reconstructor valid in experimental conditions. To do so, the test bench BEAGLE (Sáez et al., 2014), a multi-object adaptive optics capable experiment that emulates the William Herschel Telescope, is characterized. Parameters retrieved are employed to tune a simulation of the test bench, which produces validation data sets.

The first chapters develop and explain the necessary tools to address the research. Chapter 2 defines and calculates the covariance and the power spectral density of the turbulence, as measured by an aperture. This relates the physical phenomenon to the instrument measuring it. Chapter 3 presents standard methods for wave-front reconstruction and for tomographic reconstruction. Their goal is to reconstruct the wave-front from wave-front sensor measurements and to project several wave-front into a target direction, respectively. Chapter 4 establishes calibration and characterization procedures for a multi-object adaptive optics capable test bench. Original work can be found from this chapter on, as new characterization procedures are proposed. They are necessary to implement the novel correction technique presented in Chapter 5.

The final chapter proposes a technique to handle instrumental deviations. Two deviations are addressed: sub-aperture heterogeneity and field-dependent aberrations. The first is due to on-axis static aberrations and imperfections in the Shack-Hartmann's construction. The second is the product of off-axis optical aberrations. The correction technique mitigates them before handing over the measurements to the reconstructor. The intention is to present the ANN with normalized data, similar to that produced in simulation for its training. When implemented, the correction fails to deliver any improvement. To address this, a sensitivity analysis is performed. As expected, the ANN is robust when facing noise. Because the proposed correction can be seen by the reconstructor as noise, its effects are limited. Other sources of error, like atmospheric profile and pointing were tested as well.

2. CHARACTERIZING ATMOSPHERIC TURBULENCE

The atmosphere, turbulent as it is, has ever changing indexes of refraction. When the collimated light of a punctual source in sky travels through it, it changes the speed of the rays, producing a change in phase. The incoming wave-front has now a phase aberration. In adaptive optics, the phase aberration must be measured so it is subtracted or corrected using a deformable mirror. The most widely used instrument to measure the wave-front is the Shack-Hartmann wave-front sensor (SH-WFS), which measured the slopes of the wave-front. It is therefore useful to statistically characterize the phase aberration in terms of its slope, so the structure or profile of the turbulence can be derived from the SH-WFS data.

2.1. Definitions for wave-front & for slope

Let X and Y be axes defining an horizontal plane, and Z be a vertical axis. A wave-front's phase can be described by a function $\varphi(x, y)$ in radians. The wave-front can also be described in metres (or nanometres) as

$$z(x, y) = \frac{\lambda}{2\pi} \varphi(x, y) \quad (2.1)$$

where λ is the light's wavelength. As shown in Figure 2.1, a measure of its slope is the angle of arrival. In general, the small-angle approximation can be applied to say $\tan(\alpha) \approx \alpha$.

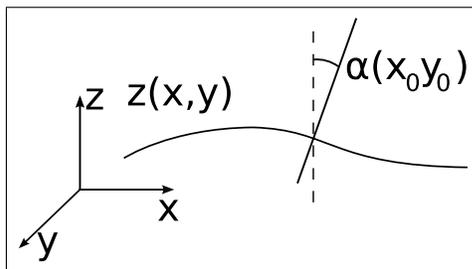


FIGURE 2.1. Angle of arrival of a wave-front described by function $z(x, y)$, at a (x_0, y_0)

From now on the angle of arrival may be referred to as slope. By definition, the angle of

arrival with respect to an X axis, also called longitudinal, at a (x_0, y_0) location is

$$\tan(\alpha_l[\lambda, x_0, y_0]) = -\frac{\partial z(\lambda, x_0, y_0)}{\partial x_0} \quad (2.2)$$

where the minus is a useful convention. Then considering the small-angle approximation

$$\alpha_l[\lambda, x_0, y_0] = -\frac{\partial z[\lambda, x_0, y_0]}{\partial x_0} \quad (2.3)$$

Finally replacing 2.1 into 2.3 gives

$$\alpha_l[\lambda, x_0, y_0] = \frac{-\lambda}{2\pi} \frac{\partial \varphi[\lambda, x_0, y_0]}{\partial x_0} \quad (2.4)$$

which relates the angle of arrival as measurable by a wave-front sensor (WFS) to the wave-front's phase.

2.2. Slopes' Covariance & Power Spectral Density

Because atmospheric turbulence is a stochastic process, it has to be characterized in the long term through statistical moments. The slopes' covariance between two points separated by $[x, y]$ is defined as

$$B_{\alpha,l}[\lambda, x, y] = \frac{\lambda^2}{4\pi^2} \left\langle \frac{\partial \varphi[\lambda, x_0, y_0]}{\partial x_0} \frac{\partial \varphi[\lambda, x_0 + x, y_0 + y]}{\partial x_0} \right\rangle_{x_0, y_0} \quad (2.5)$$

$$= \frac{\lambda^2}{4\pi^2} \lim_{\varepsilon \rightarrow 0} \left\langle \frac{\varphi[\lambda, x_0 + \varepsilon, y_0] - \varphi[\lambda, x_0, y_0]}{\varepsilon} \frac{\varphi[\lambda, x_0 + x + \varepsilon, y_0] - \varphi[\lambda, x_0 + x, y_0]}{\varepsilon} \right\rangle_{x_0, y_0} \quad (2.6)$$

$$= \frac{\lambda^2}{4\pi^2} \lim_{\varepsilon \rightarrow 0} \left(\frac{2B_\varphi[\lambda, x, y] - B_\varphi[\lambda, x + \varepsilon, y] - B_\varphi[\lambda, x - \varepsilon, y]}{\varepsilon^2} \right) \quad (2.7)$$

Then using Taylor's expansion on all terms

$$B_\varphi[\lambda, x \pm \varepsilon, y] \simeq B_\varphi[\lambda, x, y] \pm \varepsilon \frac{\partial B_\varphi[\lambda, x, y]}{\partial x} + \frac{\varepsilon^2}{2!} \frac{\partial^2 B_\varphi[\lambda, x, y]}{\partial x^2} + \dots \quad (2.8)$$

Replacing (2.8) into (2.7)

$$B_{\alpha,l}[\lambda, x, y] = \frac{\lambda^2}{4\pi^2} \lim_{\varepsilon \rightarrow 0} \left(\frac{2B_\varphi[\lambda, x, y] - 2B_\varphi[\lambda, x, y] - 2\frac{\varepsilon^2}{2!} \frac{\partial^2 B_\varphi[\lambda, x, y]}{\partial x^2}}{\varepsilon^2} \right) \quad (2.9)$$

and then simplifying, the covariance is

$$B_{\alpha,l}[\lambda, x, y] = -\frac{\lambda^2}{4\pi^2} \frac{\partial^2 B_\varphi[\lambda, x, y]}{\partial x^2} \quad (2.10)$$

From the relationship in (2.10) it is possible to find the Power Spectral Density of the slopes. To do so, for any all wide-sense stationary process X with an auto-covariance function such as

$$B_x(\tau) = \langle x(t)x^*(t - \tau) \rangle \quad (2.11)$$

Wiener-Khinchin's theorem guarantees there is a power spectral function calculated as

$$W(f) = \int_{-\infty}^{\infty} B_x(\tau) e^{-2\pi i \tau f} d\tau \quad (2.12)$$

$$\Leftrightarrow B_x(\tau) = \int_{-\infty}^{\infty} W(f) e^{2\pi i \tau f} df \quad (2.13)$$

This in terms of the turbulence's covariance becomes

$$B_\varphi[\lambda, x, y] = \iint df_x df_y W_\varphi[\lambda, f_x, f_y] e^{2\pi i(f_x x + f_y y)} \quad (2.14)$$

Replacing (2.14) into (2.10) gives

$$B_{\alpha,l}[\lambda, x, y] = -\frac{\lambda^2}{4\pi^2} \frac{\partial^2}{\partial x^2} \iint df_x df_y W_\varphi[\lambda, f_x, f_y] e^{2\pi i(f_x x + f_y y)} \quad (2.15)$$

$$= \iint df_x df_y W_\varphi[\lambda, f_x, f_y] \left(\frac{-\lambda^2}{4\pi^2} \right) \frac{\partial^2}{\partial x^2} e^{2\pi i(f_x x + f_y y)} \quad (2.16)$$

$$= \iint df_x df_y f_x^2 \lambda^2 W_\varphi[\lambda, f_x, f_y] e^{2\pi i(f_x x + f_y y)} \quad (2.17)$$

It is now possible to note that the power spectral density of the angle of arrival is

$$W_{\alpha,l}[\lambda, f_x, f_y] = \lambda^2 f_x^2 W_\varphi[\lambda, f_x, f_y] \quad (2.18)$$

2.3. Considering a Sub-aperture

As mentioned, the wave-front is sampled using a Shack-Hartmann wave-front sensor. The sensor is a lenslet array, conjugated to the pupil of the system which is in this case a

telescope. As showed in Figure 2.2, lenslets sample a small region of an incoming wave-front $z(x, y)$. They do so by focusing segments of the wave-front into spots on a CCD detector. If $\varphi(x, y)$ is limited in frequency, there is a lenslet small enough it samples an

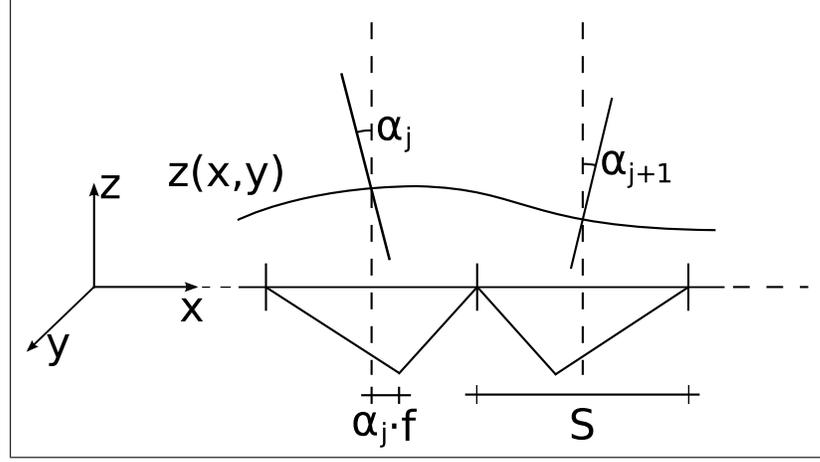


FIGURE 2.2. Transversal view of a Shack-Hartmann wave-front sensor representation. The wave-front $z(x, y)$, propagates in the Z direction, down to the lenslet array. Each lenslet, of focal length f and surface S focuses an approximately flat segment of the wave-front. The offset of the resulting spots from the optical axis of the lenslet is used to calculate the angle of arrival α_j for sub-aperture j

approximately flat segment of wave-front. In this case, it can be said the sensor measures the average slope of the sampled segment of wave-front. For sub-aperture j of surface S , the average slope is

$$\alpha_{i,p} [\lambda, x_0, y_0] = \frac{-\lambda}{2\pi S} \int_S G [x' - x_0, y' - y_0] \frac{\partial \varphi [\lambda, x', y']}{\partial x'} dx' dy' \quad (2.19)$$

where $G [x' - x_0, y' - y_0]$ is a weight function with unitary value inside sub-aperture j of coordinates $[x_0, y_0]$, and null elsewhere.

Later, the expression can be seen as a convolution between both functions. For convenience G will be centered in $[0, 0]$ so that it is pair.

$$\alpha_{i,p} [\lambda, x_0, y_0] = \frac{-\lambda}{2\pi S} \int_S G [x_0 - x', y_0 - y'] \frac{\partial \varphi [\lambda, x', y']}{\partial x'} dx' dy' \quad (2.20)$$

$$= \frac{-\lambda}{2\pi S} (G [x_0, y_0]) * \frac{\partial \varphi [\lambda, x_0, y_0]}{\partial x_0} \quad (2.21)$$

It will now be shown how the covariance of the turbulence seen through a sub-aperture can be calculated from its power spectral density and that of the sub-aperture. First lets consider the following substitution:

$$G_S [x, y] = \frac{G [x, y]}{S} \quad (2.22)$$

Then slopes' covariance through a sub-aperture is defined as

$$B_{\alpha p} [\lambda, x, y] = \left\langle \frac{-\lambda}{2\pi} \int G_S [x' - x_0, y' - y_0] \frac{\partial \varphi [\lambda, x', y']}{\partial x'} dx' dy' \right. \\ \left. \frac{-\lambda}{2\pi} \int G_S [x'' - x_0 - x, y'' - y_0 - y] \frac{\partial \varphi [\lambda, x'', y'']}{\partial x''} dx'' dy'' \right\rangle (2.23)$$

$$= \iint G_S [x' - x_0, y' - y_0] G_S [x'' - x_0 - x, y'' - y_0 - y] \\ \frac{\lambda^2}{4\pi^2} \left\langle \frac{\partial \varphi [\lambda, x', y']}{\partial x'} \frac{\partial \varphi [\lambda, x'', y'']}{\partial x''} \right\rangle dx'' dy'' dx' dy' \quad (2.24)$$

$$= \int G_S [x' - x_0, y' - y_0] \int G_S [x'' - x_0 - x, y'' - y_0 - y] \\ B_{\alpha l} [x'' - x', y'' - y'] dx'' dy'' dx' dy' \quad (2.25)$$

To continue, the translation of a function must be noted as

$$(\tau [x] f) (y) = f (y - x) \quad (2.26)$$

so that the the commutation of translations by convolution can be written as

$$\tau [x] (f * g) = (\tau [x] f) * g \quad (2.27)$$

and

$$\tau [x] (f * g) = f * (\tau [x] g) \quad (2.28)$$

Considering the notations expressed in (2.27) and in (2.28), equation (2.25) can be written as

$$B_{\alpha lp} [\lambda, x, y] = \int G_S [x' - x_0, y' - y_0] \int G_S [x_0 + x - x'', y_0 + y - y''] \tau [x', y'] B_{\alpha l} [x'' - x', y'' - y'] dx'' dy'' dx' dy' \quad (2.29)$$

$$= \int G_S [x' - x_0, y' - y_0] (\tau [-x_0, -y_0] \tau [x', y'] (G_S * B_{\alpha l}) [x, y]) dx' dy' \quad (2.30)$$

$$= \int \tau [x_0, y_0] G_S [x', y'] \tau [-x_0, -y_0] (G_S * B_{\alpha l}) [x - x', y - y'] dx' dy' \quad (2.31)$$

$$B_{\alpha lp} [\lambda, x, y] = (G_S * G_S * B_{\alpha l}) [\lambda, x, y] \quad (2.32)$$

Later, if the sub-aperture's power spectral density is defined as

$$W_{gs} [f_x, f_y] = \left(\iint_{-\infty}^{\infty} G_S [x, y] e^{-2\pi i(f_x x + f_y y)} dx dy \right)^2 \quad (2.33)$$

then it is possible to use Wiener-Khinchin's theorem in equation (2.32)

$$B_{\alpha lp} [\lambda, x, y] = \iint_{-\infty}^{\infty} W_{gs} [f_x, f_y] W_{\alpha l} [f_x, f_y] e^{2\pi i(f_x x + f_y y)} df_x df_y \quad (2.34)$$

from which the power spectral density is

$$W_{\alpha lp} [\lambda, f_x, f_y] = \lambda^2 f_x^2 W_{gs} [\lambda, f_x, f_y] W_{\varphi} [\lambda, f_x, f_y] \quad (2.35)$$

2.4. Physical Model, von Karman's Power Spectral Density

All that is needed from the previous expression is a way of calculating the turbulence's covariance or its equivalent in frequency space, the power spectral density W_{φ} . This needs a physical model and one of the first and most expanded is Kolmogorov's model (Voitskhovich, 1995). The model does not consider a finite outer scale parameter L_0 . The outer scale is the scale at which energy enters the the turbulence and then decays in an inertial regime down to the inner scale.

It is impossible to have a turbulence which correlates at infinite scale. Furthermore, a Phase Screen to simulate atmospheric turbulence in an experiment can only have a finite size. These leads to the consideration of a model which do takes into account the outer scale. This is von Karman's model (Voitsekhovich, 1995). According to it, the turbulence's power spectral density can be described as

$$W_{\varphi}[\lambda, f_x, f_y] = C(\mu)\tilde{C}_n^2\delta z\frac{1}{\lambda^2}\left[f^2 + \frac{1}{L_0^2}\right]^{-[\mu+3]/2} \quad (2.36)$$

In the equation $\tilde{C}_n^2\delta z$ is the index of refraction structure constant integrated through a turbulence layer δz , f is the frequency's norm and $C(\mu)$ is a μ depending constant of value $\mu = 2/3$ in the case of a Kolmogorov-Obukhov regime, and is calculated as follows:

$$C(\mu) = \pi\Gamma\{\mu + 2\}\sin\{\pi\mu/2\}/[2\pi]^{\mu} \quad (2.37)$$

An advantage of Equation 2.36 is that it only takes changing

$$\left[f^2 + \frac{1}{L_0^2}\right] \quad (2.38)$$

in order to change the physical model. Equation 2.38 represents the von Karman model of turbulence which considers a finite outer scale L_0 . With an infinite outer scale the model describes a Kolmogorov regime.

3. WAVEFRONT RECONSTRUCTION & TOMOGRAPHIC RECONSTRUCTION

An adaptive optics system's function is to subtract phase aberrations from incoming wave-fronts. Phase aberrations are introduced by atmospheric turbulence and constitute a wave-front that has been added to the target's wave-front. The system has to measure the aberration, then reconstruct the data into a wave-front and finally subtract it using a deformable mirror (DM). To measure a wave-front, a standard configuration is to use a Shack-Hartmann wave-front sensor (SH-WFS). The SH-WFS samples the slope of the wave-front, which is then used to reconstruct it. There are multiple methods or reconstructors. In Section 3.1 are presented two possible reconstructors. Both solve the inverse problem of calculating the wave-front's slope, here called the derivation problem. The first one does it in an all zonal representation base of the wave-front and the derivative, whereas the second goes from a zonal representation of the slopes to a modal representation of the wave-front.

In some cases a previous issue has to be addressed before the phase aberration can be reconstructed for subtraction. When a target produces an unknown wave-front, it is difficult to make the distinction between the target's and the added phase aberration. Only when the wavefront is expected to be flat or without aberrations, the wave-front is considered pure aberration due to turbulence. Unfortunately this is not always the case, as scientific targets are not always punctual light sources that produce flat wave-fronts such as stars, but elongated objects that produce unknown wave-fronts such as galaxies and nebulae. In such cases, multiple SH-WFS are deployed in order to measure nearby punctual light sources, which are called guide sources or guide stars. Because nearby stars sample the same turbulence as the target, an algorithm can be used to extrapolate the turbulence seen by the target. The algorithm is called a tomographic reconstructor. Section 3.2 presents the *Learn & and Apply* (L&A) algorithm and the application of Artificial Neural Networks (ANN) as tomographic reconstructors.

3.1. Wavefront reconstruction

Two widely used reconstructors are present in this section. The first one models the problem of calculating the slopes of a known wavefront, and then solves the inverse problem. It is said to integrate only in the sense it solves the inverse problem of the derivation. The second reconstructor goes from a zonal representation of the slopes to a description in Zernike coefficients of the wave-front. The last is called a modal representation since each coefficient weights a Zernike polynomial, which are arranged in groups or modes.

3.1.1. Zonal representation

The functions $z[x, y]$ and $\varphi[x, y]$ are representations, respectively of the wave-front and of its phase. They are called zonal representations because they produce a value for each point $[x, y]$ or zone around it. In Section 2.3, equation 2.19 describes the process through which the SH-WFS measures the slopes. With the acquired data, the intention is to reconstruct the wave-front through a process of integration.

To this purpose must be defined the places relative to the sub-apertures where the samples of the wave-front $z[x, y]$ will be reconstructed. These are called phase points. Figure 3.1 places them in the interstice between sub-apertures. The figure also shows how, through a discrete process of integration, a phase point z_{j+1} can be calculated from the previous point z_j and the slope sampled between them. This is called Fried's configuration as presented in (Southwell, 1980). For an array of square sub-apertures, each sub-aperture produces a slope described in two directions. Instead of α , slopes will now be referred by s , so the slope of in the X direction from sub-aperture j is $s_{j x}$. Fried's configuration in a two-dimensional lenslet array is as shown in Figure 3.2. Integrating as proposed in Figure 3.1 means defining a starting phase point with an arbitrary value, and using the slopes to calculate the phase points around it. As the calculation is propagated through the grid, so are the measurement errors of all the slopes used. The error will not be distributed homogeneously and will be larger the further the phase points are from the origin.

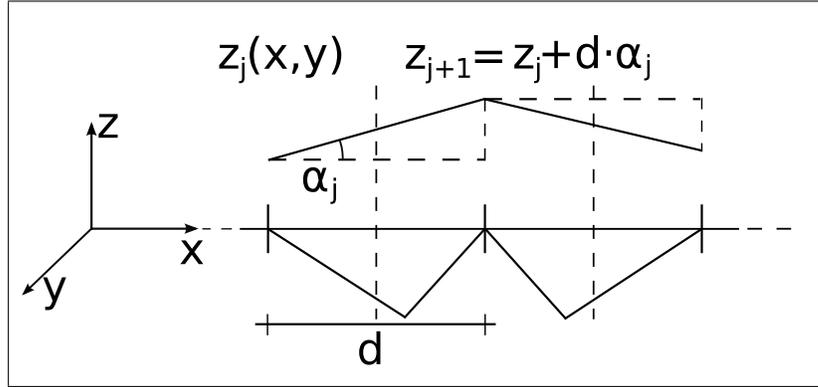


FIGURE 3.1. Discrete integration of the wave-front's slopes. Phase points z_j are located between sub-apertures of diameter d in Fried's configuration.

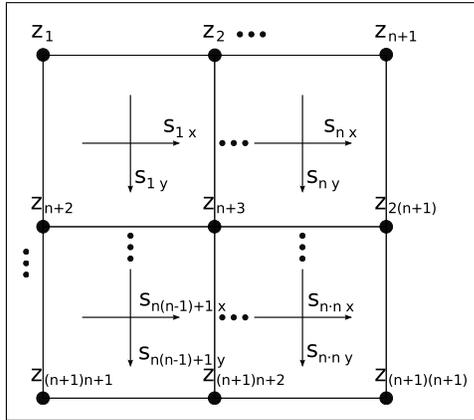


FIGURE 3.2. Fried's configuration for an $n \times n$ lenslet array. Phase points z_j are located between sub-apertures.

To solve this and to obtain a minimum error solution, the derivation problem is set and then inverted using a minimum least-squares criteria. From Figure 3.2, let all phase points be vertically concatenated in a vector

$$\vec{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_{(n+1)^2} \end{bmatrix} \quad (3.1)$$

and all slopes be vertically concatenated in a vector

$$\vec{s} = \begin{bmatrix} s_{1x} & s_{1y} & \cdots & s_{n^2x} & s_{n^2y} \end{bmatrix}^T \quad (3.2)$$

The derivation problem is to find a matrix ∇ for which

$$\nabla \vec{z} = \vec{s} \quad (3.3)$$

By examining Fried's configuration, slopes can be calculated from phase points as

$$s_{jx} = \frac{(z_{j+1} + z_{j+n+2})/2 - (z_j + z_{j+n+1})/2}{d} \quad (3.4)$$

and

$$s_{jy} = \frac{(z_{j+n+1} + z_{j+n+2})/2 - (z_j + z_{j+1})/2}{d} \quad (3.5)$$

Using these formulae, the derivation matrix becomes

$$\nabla = \frac{1}{2d} \begin{bmatrix} -1 & 1 & 0 & 0 & \cdots & -1 & 1 & 0 & \cdots \\ -1 & -1 & 0 & 0 & \cdots & 1 & 1 & 0 & \cdots \\ 0 & -1 & 1 & 0 & \cdots & 0 & -1 & 1 & \cdots \\ 0 & -1 & -1 & 0 & \cdots & 0 & 1 & 1 & \cdots \\ \vdots & \ddots & \ddots & \ddots & & & \ddots & \ddots & \ddots \end{bmatrix} \quad (3.6)$$

This matrix is finally *pseudo-inverted* so that

$$pinv(\nabla) \vec{s} \approx \vec{z} \quad (3.7)$$

Many algorithms have been implemented to perform the *pseudo-inversion*. They use Singular Value Decomposition (SVD) to minimize the square error

$$e = \langle \|pinv(\nabla) \vec{s} - \vec{z}\|^2 \rangle \quad (3.8)$$

The resulting matrix $R_\varphi = pinv(\nabla)$ is the wave-front reconstructor.

3.1.2. Modal representation

A modal representation describes a function as the weighted combination of other functions called modes. In other words, the function is described in terms of the weights necessary for its construction using a certain base. Since wave-fronts usually propagate through circular pupils, a well fitted base to describe them are Zernike polynomials (Born & Wolf, 1999). Zernike polynomials constitute an orthogonal base inside the unitary disk. If $Z_j [\rho, \theta]$ is a polynomial of index j , two polynomials are orthogonal if

$$\int d^2\rho W [\rho] Z_j [\rho] Z_{j^*} [\rho] = \delta_{jj^*} \quad (3.9)$$

where

$$W [\rho] = \begin{cases} \frac{1}{\pi}, & \rho \leq 1 \\ 0, & \rho > 1 \end{cases} \quad (3.10)$$

And because they constitute a base, then any function φ defined inside a disk of radius R can be described as

$$\varphi [R\rho, \theta] = \sum_j a_j Z_j [\rho, \theta] \quad (3.11)$$

In Zernike's definition the polynomials are defined by two indexes, the azimuthal frequency m and the radial degree n , and by their parity for they can be even or odd. When a polynomial is odd, in some cases its azimuthal frequency is considered $-m$. The definitions of the polynomials are

$$\left. \begin{aligned} Z_{n, \text{even}}^m [\rho, \theta] &= R_n^m [\rho] \cos[m\theta] \\ Z_{n, \text{odd}}^m [\rho, \theta] &= R_n^m [\rho] \sin[m\theta] \end{aligned} \right\} \begin{aligned} &0 \leq \theta < 2\pi \\ &0 \leq \rho \leq 1 \end{aligned} \quad (3.12)$$

where

$$\left\{ \begin{aligned} R_n^m [\rho] &= \sum_{k=0}^{\frac{n-m}{2}} \frac{(-1)^k (n-k)!}{k! (\frac{n+m}{2}-k)! (\frac{n-m}{2}-k)!} \rho^{n-2k} \\ |Z_n^m [\rho, \theta]| &\leq 1 \\ n &\geq |m| \geq 0 \\ n - |m| &= \text{even} \end{aligned} \right. \quad (3.13)$$

The inequations in 3.13 allow only for certain combinations of n and m , which means only certain polynomials are possible. Resulting combinations are listed in Table 3.1. When

TABLE 3.1. Indexes of Zernike polynomials. Possible pairs of indexes n, m are shown up to $n = 4$. Radial frequency is also referred to as mode

Radial degree (n)	Azimuthal frequency (m)								
	-4	-3	-2	-1	0	1	2	3	4
0					0,0				
1				1,-1		1,1			
2			2,-2		2,0		2,2		
3		3,-3		3,-1		3,1		3,3	
4	4,-4		4,-2		4,0		4,2		4,4

further using these polynomials, specially in optical applications which require using linear algebra, Zernike's notation using a pair of indexes becomes uneasy or unfitted. When representing a wave-front with Zernike coefficients, it is convenient to concatenate them in a vector so they can be operated on. This is achieved using Noll's sequence as presented in (Noll, 1976). This sequence maps every pair of indexes n, m into a single index j so that $Z_n^m \rightarrow Z_j$. The sequence, presented in Table 3.2 is constructed upon two simple rules. The first rule is n and m always start at their lowest possible values. The second rule is odd polynomials, with negative azimuthal frequencies, relate to odd values of j , and even polynomials with positive azimuthal frequencies relate to even values of j . When $m = 0$ there is no ruling, as it is not necessary. As an example, when $j = 5$ then $m = -2$, and when $j = 8$ then $m = 1$.

TABLE 3.2. Noll's sequence (sequence A176988 in OEIS). It maps every couple of indexes $(n, m) \rightarrow j$, allowing to concatenate Zernike coefficient in vectors and to operate on them.

	Index Sequence									
n,m	0,0	1,1	1,-1	2,0	2,-2	2,2	3,-1	3,1	3,-3	3,3
j	1	2	3	4	5	6	7	8	9	10

Another simplification found in (Noll, 1976) is Noll's normalization. This normalization, just as the sequence, is of standard use in adaptive optics as it is convenient for

statistical analysis. Normalizing, polynomials are now defined as

$$\left. \begin{aligned} Z_{j, \text{ even}} [\rho, \theta] &= \sqrt{n+1} R_n^m[\rho] \cos[m\theta] \\ Z_{j, \text{ odd}} [\rho, \theta] &= \sqrt{n+1} R_n^m[\rho] \sin[m\theta] \end{aligned} \right\} m \neq 0 \quad (3.14)$$

$$Z_j = \sqrt{n+1} R_n^0[\rho] \quad m = 0$$

and the resulting polynomials are shown in Table 3.3.

TABLE 3.3. Zernike polynomial with Noll's normalization and sequence. Shortened version of Table I in (Noll, 1976)

Radial degree (n)	Azimuthal frequency (m)			
	0	1	2	3
0	$Z_1 = 1$ <i>Constant</i>			
1		$Z_2 = 2\rho \cos \theta$ $Z_3 = 2\rho \sin \theta$ <i>Tilts</i>		
2	$Z_4 = \sqrt{3}(2\rho^2 - 1)$		$Z_5 = \sqrt{6}\rho^2 \sin 2\theta$ $Z_6 = \sqrt{6}\rho^2 \cos 2\theta$ <i>Astigmatism</i>	
3		$Z_7 = \sqrt{8}(3\rho^3 - 2\rho) \sin \theta$ $Z_8 = \sqrt{8}(3\rho^3 - 2\rho) \cos \theta$ <i>Coma</i>		$Z_9 = \sqrt{8}\rho^3 \sin 3\theta$ $Z_{10} = \sqrt{8}\rho^3 \cos 3\theta$

Now, as polynomials are properly defined, the final objective of reconstructing a wavefront in modal representation from measured slopes in zonal representation is addressed. This means producing a modal representation of $\frac{\partial \varphi}{\partial x}$ and of $\frac{\partial \varphi}{\partial y}$. The derivative of Equation 3.11 gives the starting point to develop the reconstructor.

$$\left\{ \begin{aligned} \frac{\partial \varphi [R\rho, \theta]}{\partial x} &= \frac{\partial}{\partial x} \left(\sum_j a_j Z_j \right) \\ \frac{\partial \varphi [R\rho, \theta]}{\partial y} &= \frac{\partial}{\partial y} \left(\sum_j a_j Z_j \right) \end{aligned} \right. \quad (3.15)$$

Let $\rho' = R \cdot \rho$, $x' = R \cdot x$ and $y' = R \cdot y$. Because coefficients a_j are constant, equations in 3.15 turn into

$$\begin{cases} R \cdot \frac{\partial \varphi[\rho', \theta]}{\partial x'} = \sum_j a_j \frac{\partial Z_j}{\partial x} \\ R \cdot \frac{\partial \varphi[\rho', \theta]}{\partial y'} = \sum_j a_j \frac{\partial Z_j}{\partial y} \end{cases} \quad (3.16)$$

Equations in 3.16 allow for calculations of the wave-front's derivative in zonal representation, when knowing its modal representation in Zernike coefficients. In other words, by knowing coefficients a_j , derivatives can be calculated if the derivatives of the polynomials are known *a priori*. This is the inverse problem of finding the coefficient by knowing the slopes. It must then be inverted by using linear algebra. The algebraic notation of the problem is

$$\begin{aligned} \frac{\partial}{\partial x} \mathbf{Z} \vec{a} &= R \cdot \vec{s}_x \\ \frac{\partial}{\partial y} \mathbf{Z} \vec{a} &= R \cdot \vec{s}_y \end{aligned} \quad (3.17)$$

In this expression

$$\vec{a} = [a_1 \ a_2 \ \dots \ a_j \ \dots \ a_m]^T \quad (3.18)$$

where sub-index m does no longer refer to a polynomial's azimuthal frequency but to the number of coefficients and therefore the finite number of Zernike polynomials used in the representation. In addition, n^2 will be the number of sub-apertures in the WFS providing the slopes. The latest are concatenated in vectors as

$$\begin{aligned} \vec{s}_x &= \frac{\partial \varphi}{\partial x} \\ &= [s_{1x} \ s_{2x} \ \dots \ s_{ix} \ \dots \ s_{n^2x}]^T \end{aligned} \quad (3.19)$$

$$\begin{aligned} \vec{s}_y &= \frac{\partial \varphi}{\partial y} \\ &= [s_{1y} \ s_{2y} \ \dots \ s_{iy} \ \dots \ s_{n^2y}]^T \end{aligned} \quad (3.20)$$

With respect to \mathbf{Z} , it is built by horizontally concatenating vectors \mathbf{Z}_j of n^2 elements so that

$$\mathbf{Z} = [Z_1 \ Z_2 \ \dots \ Z_j \ \dots \ Z_m]_{(n^2 \times m)} \quad (3.21)$$

In turn, vectors Z_j come from two-dimensional matrices, sampling Zernike polynomial modes in a Cartesian plane. Their rows have been turned into columns and vertically concatenated, just as explained in Section 3.1.1 for \vec{s} . The derivatives of these vectors have to be calculated. Equation 3.22 from (Noll, 1976) is a recursion relation between the derivative of a polynomial and polynomials of lesser order.

$$\frac{\partial}{\partial \rho} R_n^m = n (R_{n-1}^{m-1} + R_{n-1}^{m+1}) + \frac{\partial}{\partial \rho} R_{n-2}^m \quad (3.22)$$

The recursion allows to express the derivative of a polynomial in the Zernike base as

$$\nabla Z_j = \sum_{j^*} \gamma_{jj^*} Z_{j^*} \quad (3.23)$$

It also allows to derive rules to find all γ_{jj^*} coefficients and to construct a rectangular matrix γ . The rules are found in (Noll, 1976), just as are matrices γ_x and γ_y . γ_x is presented in Table 3.4.

TABLE 3.4. Zernike polynomial derivative matrix γ_x . Shortened version of Table II in (Noll, 1976)

$j \setminus j^*$	1	2	3	4	5	6
1						
2	2					
3						
4		$2\sqrt{3}$				
5			$\sqrt{6}$			
6		$\sqrt{6}$				
7					$2\sqrt{3}$	
8	$\sqrt{8}$			$2\sqrt{6}$		$2\sqrt{3}$

With matrices γ_x and γ_y the algebraic notation of Equation 3.23 is

$$\begin{aligned} \frac{\partial}{\partial x} \mathbf{Z} &= \mathbf{Z} \gamma_x^T \\ \frac{\partial}{\partial y} \mathbf{Z} &= \mathbf{Z} \gamma_y^T \end{aligned} \quad (3.24)$$

$\frac{\partial}{\partial x} \mathbf{Z}$ and $\frac{\partial}{\partial y} \mathbf{Z}$, γ_x^T and γ_y^T , and \vec{s}_x and \vec{s}_y are vertically concatenated to have

$$\begin{aligned}\nabla \mathbf{Z} &= \begin{bmatrix} \frac{\partial}{\partial x} \mathbf{Z} \\ \frac{\partial}{\partial y} \mathbf{Z} \end{bmatrix} \\ \gamma &= \begin{bmatrix} \gamma_x & \gamma_y \end{bmatrix}^T \\ \vec{s} &= \begin{bmatrix} \vec{s}_x \\ \vec{s}_y \end{bmatrix}\end{aligned}\tag{3.25}$$

so equations in 3.18 can be merged into

$$\nabla \mathbf{Z} \vec{a} = R \cdot \vec{s}\tag{3.26}$$

Here, the modal reconstructor is $R_Z = pinv(\nabla \mathbf{Z})$ so that

$$R \cdot pinv(\nabla \mathbf{Z}) \vec{s} \approx \vec{a}\tag{3.27}$$

Again, the *pseudo-inversion* uses SVG to minimize the square error

$$e = \langle \|R \cdot pinv(\nabla \mathbf{Z}) \vec{s} - \vec{a}\|^2 \rangle\tag{3.28}$$

3.2. Tomographic reconstruction

In classical adaptive optics (AO), when correcting the phase aberration of a target, its wave-front is directly measured. Using a SH-WFS, this is only possible when the target is a fairly bright punctual light source. The phase aberration of dim or elongated targets cannot be measured directly. Instead, the phase aberration of a bright enough neighbouring star is measured. Such a star is called a guide star (GS) or a natural guide star (NGS) to make the distinction with laser guide stars (LGS). As shown in Figure 3.3 the wave-front of two light sources travelling through a turbulence layer at an altitude h and into a telescope overlap, sampling the same turbulence. For $h = 0$ [m], both projections of the telescope's pupil on the layer match, sampling and mapping the same turbulence. For a small angular separation α between both sources, as altitude h increases so does proportionally the separation

between projected pupils. Therefore they have less sampled turbulence in common and in different regions of the pupil. Then, if a classical AO system was to correct the phase aberrations of one wave-front, by compensating for the aberrations seen by the other, it would only succeed for tightly overlapped pupils. The validity of the correction would quickly decline as the pupils move away from each other.

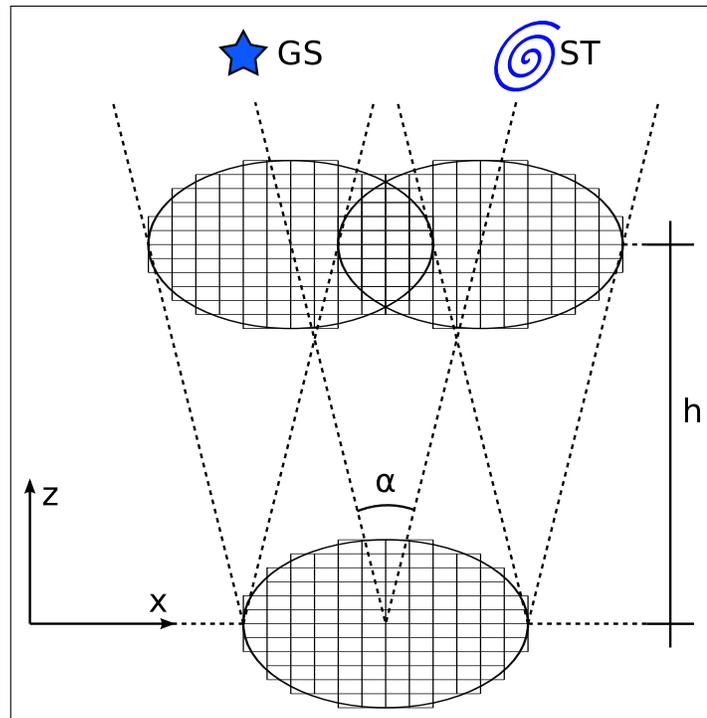


FIGURE 3.3. Telescope's pupil, projected in two directions separated by an angle α , into a turbulence layer at an altitude h . As a 12×12 Shack-Hartmann is conjugated to the pupil, it is also projected in altitude. The overlap between both projections depends on the altitude and on the angle between the light sources. The star symbol represents a guide star (GS) and the spiral represents an elongated scientific target (ST).

Instead, only the aberration of the overlapped region of the pupil projected towards the target is known from the aberration seen in the direction of the GS. A tomographic reconstructor is any algorithm relating the aberration of a pupil to that of another for later correction. The task becomes more complex as not one but multiple turbulence layers, at different and ever changing altitudes, appear above the telescope. In such a scenario, projected pupils overlap at multiple altitudes and thus in multiple configurations.

Multi-object adaptive optics (MOAO) uses multiple GSs, whether they are LGSs or NGSs, to estimate or to reconstruct the target's phase aberration. Corelation matrixes and artificial neural networks (ANN) are two of the major MOAO reconstruction paradigms. Examples of reconstructors based on them are *Learn & Apply* (Vidal et al., 2010) and CARMEN (Osborn et al., 2012) respectively. Both have been tested at the CANARY demonstrator and constitute the most promising candidates for the E-ELT MOAO capabilities. In this section, both reconstructors are presented. This is in the interest of introducing improvements to available reconstructor training techniques in Chapter 5.

3.2.1. *Learn & Apply*

An optimal reconstructor minimizes the residual phase variance between an expected and an estimated phase. The expected phase $\vec{\varphi}$, which can be directly measured when having a punctual light source, is that of the target. The estimated phase, also being that of the target, is reconstructed from a measurement vector \vec{s} , concatenating the output of WFSs pointing at GSs. When considering a linear interaction of the wave-front with all layers it is possible to describe the problem in algebraic notation as

$$\langle (\vec{\varphi} - \mathbf{R}\vec{s}) \vec{s}^T \rangle = 0 \quad (3.29)$$

where \mathbf{R} is a covariance matrix acting as the reconstructor. Solving for R gives

$$\mathbf{R} = \langle \vec{\varphi} \vec{s} \rangle \langle \vec{s} \vec{s}^T \rangle^{-1} \quad (3.30)$$

In this particular case \mathbf{R} reconstructs from a slopes base to a phase base to describe the wave-front. In the case of *Learn & Apply* (*L&A*) the reconstruction keeps the slopes base. For a concatenated vector of GS measurements \vec{s}_1 and a similar vector of target measurements \vec{s}_2 , the notation of the reconstruction is

$$\vec{s}_2(t) = \mathbf{R} \cdot \vec{s}_1(t) \quad (3.31)$$

Let \mathbf{S} be a matrix of m horizontally concatenated vectors $\vec{s}(t)$ defined as

$$\mathbf{s} = \left[\vec{s}(t_0 + \Delta t) \quad \vec{s}(t_0 + 2\Delta t) \quad \dots \quad \vec{s}(t_0 + m\Delta t) \right] \quad (3.32)$$

Then Equation 3.31 is written as

$$S_2 = \mathbf{R} \cdot S_1 \quad (3.33)$$

As S_2 is not a square matrix it cannot be inverted. Furthermore, in order to ensure that \mathbf{R} does not converge to any particular solution, the number m of concatenated vectors must be great in comparison to their dimension so all linearly independent cases are provided. Then, just as for wave-front reconstructors, *pseudo-inversion* is in order so to minimize the residual error defined as

$$e = \|\mathbf{R}S_1 - S_2\|^2 \quad (3.34)$$

which developing gives

$$e = \sum_i \sum_j \left(\sum_k r_{ik} s_{1kj} - s_{2ij} \right)^2 \quad (3.35)$$

In this case the square error is differentiated with respect to the values in \mathbf{R}

$$\frac{\partial}{\partial r_{ik}} e = \sum_i \sum_j \frac{\partial}{\partial r_{ik}} \left(\left[\sum_k r_{ik} s_{1kj} - s_{2ij} \right]^2 \right) \quad (3.36)$$

$$= \sum_i \sum_j 2s_{1kj} \left(\sum_{k'} r_{ik} s_{1k'j} - s_{2ij} \right) \quad (3.37)$$

The error is minimized by setting the derivatives to zero. Resulting equation is

$$\mathbf{R} (S_1 S_1^T) - (S_2 S_1^T) = 0 \quad (3.38)$$

Here $(S_1 S_1^T)$ is a square matrix, invertible when having linearly independent rows in S_1 .

Solving for \mathbf{R} gives

$$\mathbf{R} = (S_2 S_1^T) (S_1 S_1^T)^{-1} \quad (3.39)$$

Both terms are then divided by the number of measured cases m so that

$$\mathbf{R} = \left(\frac{1}{m} S_2 S_1^T \right) \left(\frac{1}{m} S_1 S_1^T \right)^{-1} \quad (3.40)$$

where m is also the number of columns in S_1 and S_2 . When the number of cases tends to infinity expressions $\left(\frac{1}{m} S_2 S_1^T \right)$ and $\left(\frac{1}{m} S_1 S_1^T \right)$ tend respectively to covariance matrices C_{OnOff} and C_{OffOff} . Covariance matrices are named after the convention that GSs are off-axis and the target is on-axis. Resulting expression is

$$\mathbf{R} = C_{OnOff} \cdot C_{OffOff}^{-1} \quad (3.41)$$

C_{OffOff} contains covariances of all off-axis sub-apertures with each other. C_{OnOff} contains covariances of on-axis sub-apertures with off-axis ones. The covariance between two sub-apertures i and j from WFSs p and q is $B_{\alpha p}[\lambda, x, y]$ for a single layer. $[x, y]$ is the relative position of one sub-aperture with respect to the other as projected on the turbulence layer. It is calculated as

$$[x, y] = \begin{bmatrix} x_{ip} - x_{jq} \\ y_{ip} - y_{jq} \end{bmatrix} + h \begin{bmatrix} \alpha_p - \alpha_q \\ \beta_p - \beta_q \end{bmatrix} \quad (3.42)$$

where $[\alpha_p, \beta_p]$ is the direction of source p and h is the altitude of the layer. The resulting value of the covariance is dependent on the geometry of the problem (altitude and directions) and on the profile (L_0, r_0) . It therefore yields information on the system deviations (real against expected directions) such as misalignments of the pupils and amplification of the measurements. For multiple independent layers, covariances are simply added.

The direct approach of calculating covariance matrices from on-sky data is limited for two reasons. The first is statistical convergence. In other words only a finite amount cases can be registered and the matrices not necessarily tend to covariance matrices. The second difficulty is having an on-axis WFS also called truth sensor. As the central target will usually be elongated and dim, wave-front sensing in its direction may not be possible. Both of these problems are solved by using a model to calculate the covariance matrices. Just as in Section 2.3, the model uses an *a priori* turbulence model (Kolmogorov's or Von Karman)

and includes the previously mentioned parameters on which covariance matrices depend. The model is fitted against measured covariance matrices $C_{OnOff,raw}$ (when available) and $C_{OffOff,raw}$ by minimizing the square error defined as

$$e = \|C_{OffOff,raw} - C_{OffOff}[h, r_0, L_0, [\alpha_p, \beta_p]]\|^2 + \|C_{OnOff,raw} - C_{OnOff}[h, r_0, L_0, [\alpha_p, \beta_p]]\|^2 \quad (3.43)$$

The Levenberg-Marquardt algorithm is well fitted to minimize square errors through iterations. From the optimization, the parameters of the problem are retrieved. In (Vidal et al., 2010) this part of the algorithm is called *Learn*. When on-axis data is not available, which is the second difficulty that was previously mentioned, the error to minimize is defined without C_{OffOff} . Only half of the expression remains. All parameters but the precise direction of the central target are therefore obtained. Later, the *Apply* part simply consists in calculating $C_{OffOff}[h, r_0, L_0, [\alpha_p, \beta_p]]$ and $C_{OnOff}[h, r_0, L_0, [\alpha_p, \beta_p]]$ from the model, using the parameters previously found. C_{OffOff} is relatively well estimated, even without a calibrated direction.

Since the variation time scale of atmospheric parameters is close to a few minutes, the reconstructor will only be valid during that time. Then, it has to be estimated and applied before losing validity. For longer observations, the reconstructor has to be calculated several times, with different resulting performances each time.

3.2.2. Tomographic reconstructors based on artificial neural networks

Artificial neural networks (ANN) are computational models consisting of a multitude of interconnected processing units called neurons or nodes. Inspired in biological neural networks or brains, they seek to reproduce their forecasting and pattern recognition capabilities under noisy conditions. In complex problems that remain difficult to analytically model, they provide an optimized solution for prediction. In astronomical instrumentation and AO they find a spectrum of applications ranging from modelling deformable mirrors (Guzmán, Juez, Myers, Guesalaga, & Lasheras, 2010) to open-loop tomography (Osborn et al., 2012)

Neurons receive the output of other neighbouring neurons through weighted connections called synapses. The inputs are then added and passed on to an activation function before being sent to further neurons. Neurons are arranged in ordered layers so a neuron from a given layer receive as input all the outputs from the previous layer, and send its own output as input for the following layer. A well performing topology when training ANNs to be tomographic reconstructors is the multi-layer perceptron back-propagation (MLP-BP) network. Its structure, consisting of three layers, is shown in Figure 3.4.

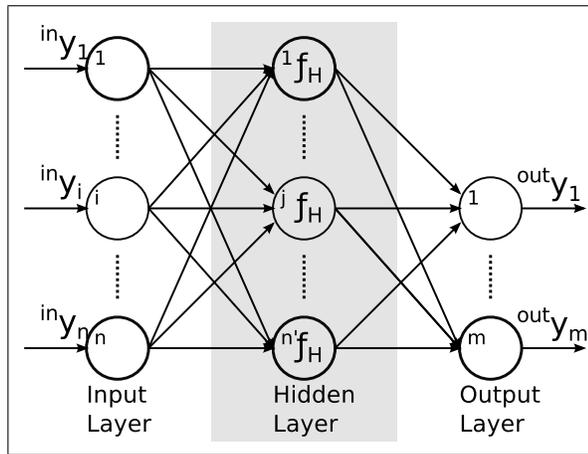


FIGURE 3.4. Topology of a multi-layer perceptron back-propagation network. The activation pattern is ${}^{in}\vec{y}$, a neuron's activation function in the hidden layer is f_H . The number of neurons in the input layer, in the hidden layer and in the output layer are n , n' and m respectively. In most cases $n = n'$.

The input layer receives an input vector ${}^{in}\vec{y}$, also called activation pattern. Each element of the vector is passed to a single neuron without being weighted. Each neuron j in the hidden layer receives the weighted sum of the input layer's outputs. The input of neuron j is calculated as

$$x_j = \sum_{i=0}^n w_{i,j} y_i \quad \begin{cases} i = 0, 1, \dots, n \\ j = 1, 2, \dots, n' \end{cases} \quad (3.44)$$

where $w_{i,j}$ is the weight of the synapse going from node i to j , and y_0 is an offset node, providing a constant value, independent of the activation pattern. For normalization purposes, the value of y_0 is typically 1. The neuron's output is produced by an activation function f_H

as

$$y_j = f_H(x_i) \quad j = 1, 2, \dots, n' \quad (3.45)$$

and weighted again when sent to the nodes in the output layer. Activation function in most cases are sigmoid functions. They act linearly within a valid range outside of which there is saturation. A common activation function is

$$f_H(x) = \frac{1}{1 + e^{-x}} \quad (3.46)$$

ANNs must be conditioned to perform a given task such as those mentioned before. It is said the ANN is trained. The training consists on adjusting the weights of all synapses arriving and leaving the hidden layer, so that the output of the ANN fits an expected output. In other words, the weights are tuned to minimize the error between the ANN's actual response to an activation pattern and the response that is expected. Training for a single scenario is irrelevant. Instead, the ANN must be able to correctly respond against a wide variety of scenarios, perhaps even predicting some not previously seen. For m nodes in the output layer and s scenarios or cases, the square error is defined as

$$e = \frac{1}{2} \sum_r^s \sum_{k=1}^m (y_{k,r} - d_{k,r})^2 \quad (3.47)$$

where $d_{k,r}$ is the expected response to the activation pattern r .

The error minimization or training algorithm for the MLP is the back-propagation (BP) algorithm (Rumelhart, Hinton, & Williams, 1988). It is a recursive algorithm that minimizes through iterations and following the inverse of the gradient. When first set, the weights are randomly assigned. Then each weight is added a step calculated as

$$\Delta w_{i,j} = -\varepsilon \cdot \frac{\partial e}{\partial w_{i,j}} \quad (3.48)$$

where ε defines the size of the step proportional to the derivative and is called the *learning rate*. Using the chain rule and replacing 3.44 into 3.48

$$\begin{aligned}
\Delta w_{i,j} &= -\varepsilon \cdot \frac{\partial e}{\partial x_j} \cdot \frac{\partial x_j}{\partial w_{i,j}} \\
&= -\varepsilon \cdot \frac{\partial e}{\partial x_j} \cdot \frac{\partial}{\partial w_{i,j}} \left(\sum_{i=0}^n w_{i,j} y_i \right) \\
&= -\varepsilon \cdot \frac{\partial e}{\partial x_j} \cdot y_i \\
&= -\varepsilon \cdot \delta_j \cdot y_i
\end{aligned} \tag{3.49}$$

To compute $\delta_j = \partial e / \partial x_j$, the ANN is evaluated at the current state (current synaptic weights) and for a given activation pattern. For a neuron k in the output layer, δ_k is retrieved directly differentiating the error in Equation 3.47 to produce

$$\delta_k = \frac{\partial}{\partial x_k} \left(\frac{1}{2} \sum_r^s \sum_{k^*=1}^m (y_{k^*,r} - d_{k^*,r})^2 \right) \tag{3.50}$$

Only hidden layer nodes have an activation function. For all other neurons $y_k = x_k$. Replacing this gives

$$\begin{aligned}
\delta_k &= \frac{\partial}{\partial x_k} \left(\frac{1}{2} \sum_r^s \sum_{k^*=1}^m (x_{k^*,r} - d_{k^*,r})^2 \right) \\
&= \sum_r^s (x_{k,r} - d_{k,r})
\end{aligned} \tag{3.51}$$

With this result it is possible to adjust all synaptic weights $w_{j,k}$ going from neurons j in the hidden layer to neurons k in the output layer. To adjust a weight $w_{i,j}$ going from a node i in the input layer to a node j in the hidden layer, Equation 3.49 says δ_j must be calculated. Again, using the chain rule gives

$$\begin{aligned}
\delta_j &= \frac{\partial e}{\partial x_j} = \sum_k \frac{\partial e}{\partial x_k} \cdot \frac{\partial x_k}{\partial y_j} \cdot \frac{\partial y_j}{\partial x_j} \\
&= f'_H(x_j) \cdot \sum_k \delta_k \cdot w_{j,k}
\end{aligned} \tag{3.52}$$

Resulting expression shows how to compute δ_j in the hidden layer using all previously calculated δ_k from a later layer. The algorithm is said to propagate backwards the gradients which gives origin to its name.

Artificial neural networks have successfully been trained into tomographic reconstructors, being CARMEN the most prominent example. First, a proof of concept for CARMEN was conducted through simulated data for training and validation (Osborn et al., 2012). The ANN receives as inputs the concatenated measurements of WFSs pointing at GSs, and is expected to produce the measurements of another WFS observing the central target. CARMEN possess a single hidden layer. It has been proved this amount is sufficient to approximate any functional relationship, provided a large enough number of neurons (Hornik, Stinchcombe, & White, 1989). It has even been shown that ANNs with two hidden layers tend to converge with less accuracy than their single-layer counterparts (De Villiers & Barnard, 1993). With respect to the number of nodes in the hidden layer, if it is too little the network fails to describe or to fit the model in enough detail. In contrast, too many neurons allow for an exceeding number of degrees of freedom and therefore for overfitting. A rule of thumb is the number has to be equal or similar to the number of nodes in the input layer. There are techniques to find an optimized number of neurons, but a trial and error procedure remains the preferred option amongst most users (Guzmán et al., 2010). After training, the network is validated with an independent set of data (Bottaci et al., 1997). If the ANN was overfitted it will not correctly extrapolate the model to fit the cases it was not presented.

Artificial neural networks are capable of modelling functions of greater complexity and dependent on more variables than matrices. Unlike linear reconstructors such as *Learn & Apply*, ANNs do not have to be optimized for turbulence layers at fixed altitudes. The best way to train an ANN based reconstructor is to present it with a large number of randomly generated phase screens, at a fixed altitude, and then repeat the process for every desired altitude. If the range of altitudes is well sampled, the ANN will be able to extrapolate when a layer is found between two altitudes it was trained for. When trained in simulation, CARMEN was presented with 1000 phase screens per altitude, with a resolution of 100

m between altitudes. And because turbulence can be considered to have a linear effect on wave-fronts, the training remains valid for any number of layers, at varying altitudes. Later, CARMEN was tested on CANARY, an adaptive optics demonstrator operated at the William Herschel Telescope (Osborn et al., 2014). The on-sky results show that when comparing CARMEN to L&A, the later outperforms the ANN by 5% in Strehl, but only when working in the configuration at which it was optimized. When layers change in altitude, L&A's performance quickly degrades, whereas the residual average wave-front error (WFE) of CARMEN increases approximately as a linear function of the altitude. In other words, the ANN remains valid for all scenarios, surpassing linear reconstructors when not re-optimized. This trend is shown in Figure 3.5, reproduced from (Osborn et al., 2014).

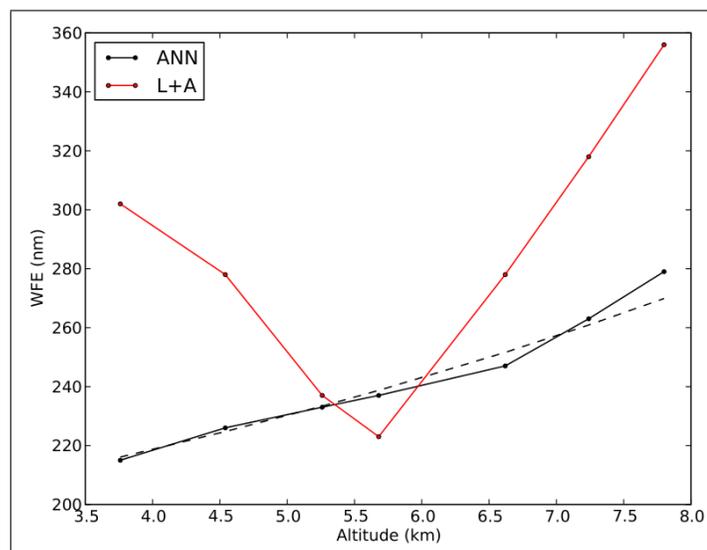


FIGURE 3.5. WFE for L&A and an ANN based reconstructor on the CANARY calibration bench, against altitude of the phase screen. The dashed line shows the expected performance of the ANN. Reproduced from (Osborn et al., 2014).

The wide range validity of ANNs is an absolute necessity. Linear tomographic reconstructors have to be recalculated when atmospheric conditions change. This process takes in the order of a few minutes and can be done in parallel while observing. ANNs instead

may take up to a week to train. Since it is not possible to accurately predict the observing conditions a few days in advance, the training has to consider all possible turbulence strengths and altitudes. This is impractical since producing experimental training data, including instrumental deviations and with enough variability and independence requires an extensive use of resources. In Chapter 5 a method will be proposed to produce training data through numerical simulations that include an instrumental characterization. This method allows to produce large data sets with pseudo-randomness and high variability, all while freeing instruments and resources.

4. CALIBRATION AND CHARACTERIZATION OF THE BEAGLE TEST BENCH

The test bench BEAGLE is a multi-object adaptive optics (MOAO) capable experiment to simulate the 4.2 m William Herschel Telescope. It is able to produce multiple natural and laser guide stars, the light from which goes through two phase screens. The first is the ground layer and the second is an altitude layer, settable at any altitude. Horizontal and vertical motions of the phase screens are controlled by step motors. Equivalent to the 4.2 m aperture, horizontal resolution is 1.5 mm and vertical resolution is 340 mm. For stars 30 arc seconds apart from the on-axis target, the vertical resolution translates into an horizontal resolution of 0.05 mm. In other words as the phase screen moves vertically in steps of 340 mm, a 30 arc seconds off-axis star perceives a virtual horizontal displacement of 0.05 mm. This will later be used in the precise measuring of the system's plate scale. A modular representation of BEAGLE is shown in Figure 4.1. Resulting wavefronts are then measured by standard 16x16 Shack-Hartmann wavefront sensors giving out slopes.

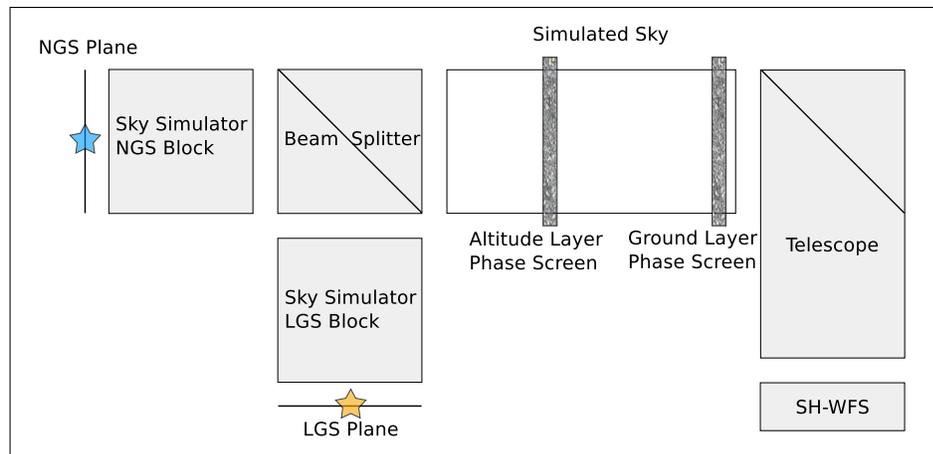


FIGURE 4.1. Block diagram of BEAGLE, a multi-object adaptive optics capable experiment. Here only the open-loop portion is shown, from the natural and laser guide stars to the Shack-Hartmann wavefront sensor, and going through the phase screens.

To produce precise and valid data, an experiment has to undergo characterizations and calibrations. Characterizations offer long term knowledge on the properties of the system such as plate scale and turbulence profile. Calibrations are performed on a routine basis to

measure and set time changing parameters such as background images, exposure time and reference centroids.

In the following sections calibration routines and characterization methods will be presented. These methods were specifically implemented for BEAGLE.

4.1. Calibration Routine

Before any acquisition is performed, dark images have to be set. A dark images or a background image is an image acquired with the camera shutter closed, blocking all light to entering the camera. If the detector senses any signal it is due to a stochastic process or noise called dark current. This process also occurs when the shutter is opened during an observation. Therefore, the dark current has to be subtracted from the image so only the observed signal remains. In a CCD camera, every pixel is characterized by its own dark current. As described in (Janesick, 2007), in a shot noise regime the dark current's noise is mostly its shot noise expressed as

$$\sigma_{D_SHOT} = (D)^{(1/2)} \quad (4.1)$$

where D is the average dark current accumulated (which is actually expressed in units of charge instead of units of charge per unit of time as would be expected from calling it "current") given as

$$D = t_I D_R \quad (4.2)$$

and where t_I is the integration time and D_R is the dark current rate (this time expressed in units of charge per unit of time). D_R has its own expression from which it can be modelled but here it will be considered as a pixel parameter, mostly dependent on the temperature. Because of this noise, the background image is determined as the average of a large number of dark images. The shot noise for a number n of averaged images is

$$\sigma_{D_SHOT} = (D)^{(1/2)} / \sqrt{n} \quad (4.3)$$

Then, as n grows the shot noise decreases, resulting in an image containing the average dark current D to be subtracted to any future image, with little noise left.

For the background image to have the same average dark current as an acquired image, they both need to have the same exposure time. An exposure time has to be determined that will produce images with a sufficient signal to noise ratio. LED stars in BEAGLE have slightly different brightness from each other. This brightness can also vary in time. Therefore a calibration routine is required to measure the intensity of each light source and then acquire a suitable background image with the same exposure time. It is important in order to be precise that a dark image is subtracted to the intensity measurement.

Because a background image can only be taken after determining the exposure time, but is necessary in the first place to do so, an iterative routine is in order. The final exposure time must be so the intensity of the light source's image is equal to a desired value. It must start with an initial guess for the exposure time and acquire a temporary background image, composed of few frames to save time. Then, by using the same exposure time, acquire an image with the target light source on and subtract the temporary background. Let the maximum value of the resulting image be I_i where sub-index i represents the iteration. Each iteration adjusts the exposure time so that I_i converges to the desired value. The exposure time for the next couple of background and light source images is calculated as

$$t_{i+1} = \frac{t_i I_{max} p}{I_i} \quad (4.4)$$

where t_i is the exposure time at iteration i , I_{max} is the saturation value of the detector, and p is a factor ranging from 0 to 1 to set the desired value as a percentage of the saturation value. For BEAGLE, I was found to be sufficiently high with $p = 0.7$.

The iteration comes to an end when the measured intensity I has an error with the desired intensity below a given threshold. Again, for BEAGLE a threshold of 5% was found to be sufficient and resulting in at most two iterations for speed. Once the definitive exposure time is found, a proper background image is acquired using a large number of frames.

Once the exposure time and background images have been set for all light sources to be used, the static aberrations of the optical system have to be accounted for. BEAGLE's Shack-Hartmann wavefront sensor (SH-WFS) intends to measure the phase aberration introduced by the system's phase screen. The phase screen simulates a layer of turbulence in the optical path above the telescope. But when propagating through the optical system, its static aberrations are added. The wavefront sensor (WFS) will therefore measure both, unless the static aberration is previously known and subtracted from the measurements. Routine calibrations can include a previous measurement of the static aberration, where the pupil's position is calibrated so the average position off all centroids issued from the Shack-Hartmann (SH) is zero and thus removing tip and tilt. The remaining offsets describe higher order aberrations. They are expressed in terms of wavefront slopes but can be transformed into any other base such a Zernikes or wavefront.

Previously measuring static aberrations has proven to be insufficient in BEAGLE. When adding the phase screen, a slight tip and tilt is introduced to the wavefront. In other words, measuring phase aberration along a phase screen produces in average tilted wavefronts when subtracting static aberration. What is done instead is to acquire the widest range possible of data and then subtract to it its own mean. For a large enough data set, the turbulence can be thought of as decorrelated so any non-zero average found is the result of static aberrations introduced by the system or by the phase screen.

4.2. Plate scale measurement

As shown in Figure 4.1, BEAGLE locates its NGS and its LGS each in separate planes. These are then collimated and joined by the Sky Simulator. After the Sky Simulator comes the portion of simulated sky where the phase screens are located. The system was designed to have the same Lagrange constant as the 4.2 m WHT. With a pupil of 18 mm, transversal magnification in simulated sky is $M_t = (4200/18)$, while longitudinal magnification is $M_l = (4200/18)^2$. Then the pupil goes through a focal reducer and into the SH-WFS formed by a lenslet array and a CCD.

In order to determine the plate scale as a unit of angle of arrival per unit of distance on the WFS image plane, two measurements must be performed. First is needed the plate scale between the NGS plane and the WFS image plane. This first plate scale $PS1$ will be expressed in units of pixels/mm, since the position of each LED source is known in millimetres as shown in Figure 4.2.

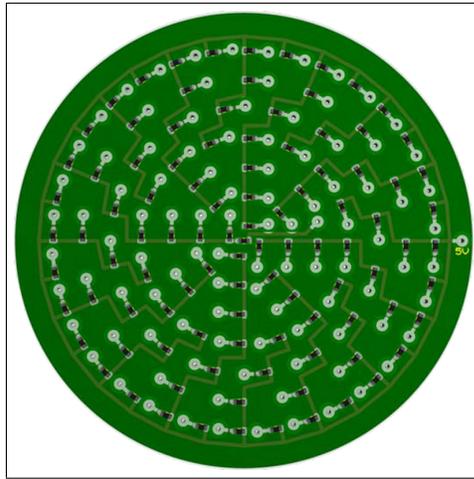


FIGURE 4.2. LED printed circuit board (PCB) containing BEAGLE's light sources. In the PCB's plane the light sources' positions are known in millimetres. This plane is the NGS plane.

To measure this plate scale the pupil's image on the WFS plane is mapped against the NGS plane. The pupil can be directly imaged with WFS's camera. Being the SH conjugated with the pupil, it can image it for all stars in the field of view (FoV) with a single fixed CCD detector. Figure 4.3 shows the X and Y positions of the pupil against the position of the light source light source.

From the slope the average plate scale is $17.2 [pixels/mm]$ in the X direction and $16.8 [pixels/mm]$ in the Y direction. It can also be seen the plate scale is constant throughout the field of view, which it would not be with off-axis aberrations. From this data set, it is interesting to calculate the error of alignment of the printed circuit board (PCB), which can be rotated relative to the camera. The coordinates of the LEDs are transformed into polar coordinates and the difference between the angle on the PCB and the one found for its image on the camera is calculated. In average, the difference in angle is $2.6 [^\circ]$.

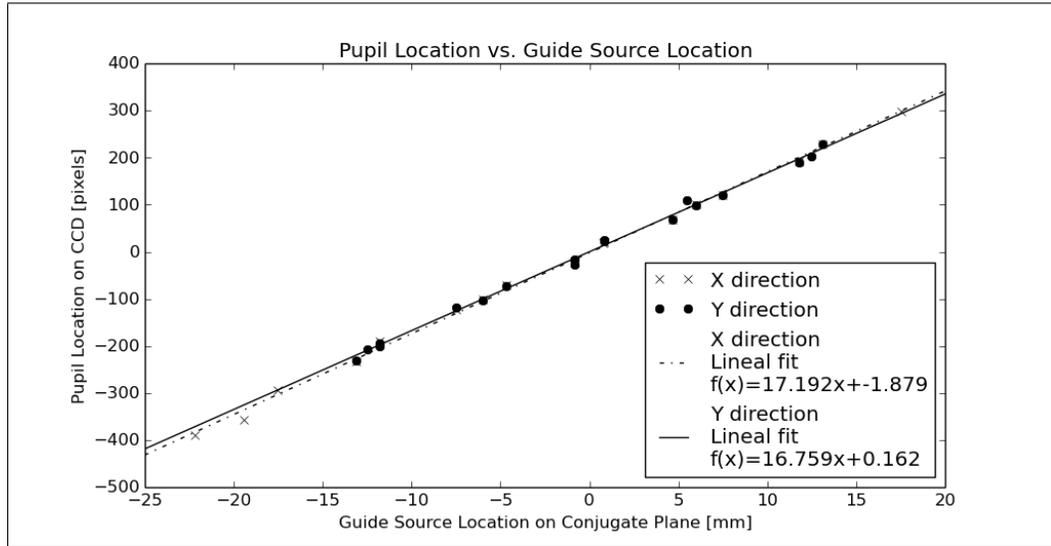


FIGURE 4.3. Pupil location on the CCD camera as a function of the guide source location on the NGS plane. The CCD camera is big enough to contain all the pupils behind the SH-WFS. Pixel size is $7.4 \mu m$. $PS1$ is given by the slope of the curve.

Now must be measured the plate scale $PS2$ between the angular position of the guide sources and their position on the NGS plane. When two stars are aligned with the direction of the wind, or in the case of a test bench, with the direction of the motion of the phase screen, they will sample the same turbulence with an offset. This is true under the frozen flow assumption, which states turbulence can be consider as frozen relative to its translation due to wind. In other words, the turbulence is seen as an invariant phase screen translating through the field of view. The angular offset can be then converted into a distance with the altitude of the layer. In an test bench this deduction is inverted. Because the phase screen's position is controlled by stepper motors and is well known, from there can be derived the angular separation of the sources. This configuration is illustrated in Figure 4.4. The ratio between the angle found and the separation of the sources on the NGS plane is $PS2$ expressed in radians/mm. The definitive plate scale PS is calculated as $PS2/PS1$ so the result is in radians/pixels. In a more convenient unit, $PS = 278 [arcsec/\mu]$. PS is only the plate scale that relates with angles of arrival on the bench. It has to be translated to angles of arrival on-sky. Using transversal magnification, finally $PS_{on-sky} = 1.19 [arcsec/\mu]$.

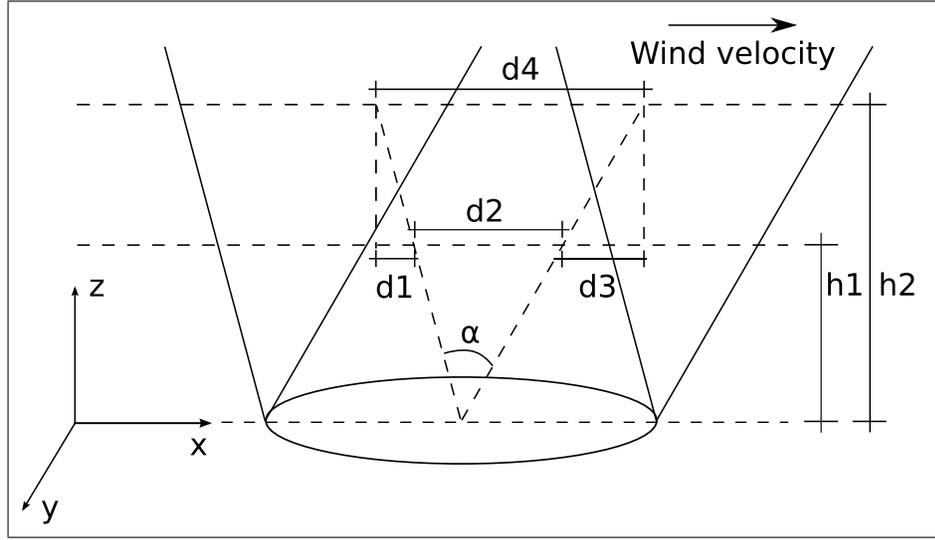


FIGURE 4.4. Two light sources illuminating a moving turbulence under the frozen flow assumption. The pupil can be large and sampled by a Shack-Hartmann wavefront sensor, or small as a sub-aperture. In a test bench all altitudes and distances h_1 , h_2 , d_1 , d_2 , d_3 and d_4 are known. The angular separation between the light sources can therefore be determined from geometry.

4.3. Off-axis aberrations characterization through optimal filter signal comparison

4.3.1. Description of the optimal filter signal comparison technique

In order to measure distances d_1 , d_2 , d_3 and d_4 as represented in Figure 4.4, an optimal filter was implemented. When an aperture observes two different light sources that are aligned with wind velocity, it samples the same turbulence with an offset. The optimal filter convolutes both signals and determines the place of maximum covariance. Let $s_1(x)$ and $s_2(x)$ be two signals observed by the same sub-aperture. The sub-index represents the star and x is the position of the phase screen in the X axis. For simplification, the stars will be aligned with this axis. Both, $s_1(x)$ and $s_2(x)$ are sampling the same turbulence but using different guide sources. Hence,

$$s_2(x) = s_1(x - \chi) \tag{4.5}$$

Now, let the covariance between both signals be

$$B(\tau) = \langle [s_1(x) - \mu_1] \cdot [s_2(x + \tau) - \mu_2] \rangle \quad (4.6)$$

$$= \langle [s_1(x) - \mu_1] \cdot [s_1(x - \chi + \tau) - \mu_2] \rangle \quad (4.7)$$

where μ_1 and μ_2 are the means of s_1 and s_2 . For any couple of signals bound by Equation 4.5

$$\operatorname{argmax}\{B(\tau)\} = \chi \quad (4.8)$$

For two comparable signals, the maximum covariance will be in the order of magnitude of the variance. The offset χ between the signals equals the offset τ between the maximum auto-covariance of one of them and the maximum covariance between the two of them.

4.3.2. Application: Field dependent aberrations characterization

There is another application for the filter described above, other than plate scale measurement. A Shack-Hartmann wavefront sensor's main component is a lenslet array. Each lenslet defines a sub-aperture that will sample a region of a wavefront by focusing it into a spot. A single sub-apertures observing the two targets can be submitted to different aberrations. These are field dependent aberrations or off-axis aberrations. They change the shape of the resulting spots, thus altering the assessment or estimation made by the centroiding algorithm. The centroiding algorithm's function is to estimate the spot's centre so the wavefront's average angle of arrival can be determined. A first degree approximation of the introduced distortion is the addition of an offset to the spot's average position and a change of its motion's amplitude. Both things can happen in the X' direction and in the Y' direction of the sub-aperture's local referential. The ratio between the amplitude of the motion with and without aberrations will be referred to as gain. Let $\vec{s}_{i,j}$ be the position of the spot's centroids without aberrations for light source i and sub-aperture j and $\vec{s}'_{i,j}$ be its position with aberrations so that

$$\vec{s}'_{i,j}(x) = \begin{bmatrix} s'_{x',i,j}(x) \\ s'_{y',i,j}(x) \end{bmatrix} \quad (4.9)$$

Let $G_{i,j}$ be a diagonal matrix with the gains for both directions of the form

$$G_{i,j} = \begin{bmatrix} g_{x' i,j} & 0 \\ 0 & g_{y' i,j} \end{bmatrix} \quad (4.10)$$

Finally, if $\vec{b}_{i,j}$ are the offsets, the aberration's effect can be modelled as

$$\vec{s}'_{i,j} = G_{i,j}\vec{s}_{i,j} + \vec{b}_{i,j} \quad (4.11)$$

Without static aberrations, atmospheric turbulence should produce null average data. In other words $\vec{b}_{i,j}$ is only due to the optical aberrations, and $\vec{s}_{i,j}$ should have null average. Then, from experimental data $\vec{b}_{i,j}$ is directly measurable as the average of $\vec{s}'_{i,j}$. For simplification, $\vec{b}_{i,j}$ will be left out and all slopes will be considered to have null average so that

$$\vec{s}'_{i,j} = G_{i,j}\vec{s}_{i,j} \quad (4.12)$$

From here, determining $G_{i,j}$ is not a direct computation. Solving for $G_{i,j}$ needs the real values of $\vec{s}_{i,j}$. These values are impossible to obtain since the turbulence is always measured through an imperfect instrument with aberrations $G_{i,j}$. Instead, if two signals were known to be equal before measuring, the ratio between their gains could be quantified. This is the case for any pair of targets i and i' that are aligned with the speed velocity of the turbulence, and are seen by sub-aperture j , so that

$$\vec{s}_{i,j}(x) = \vec{s}_{i',j}(x - \chi) \quad (4.13)$$

Using the correlation filter on measured (and therefore with aberrations) signals, offset of maximum correlation $\tau = \chi$ is determined and added so that $\vec{s}'_{i,j}(x) \simeq \vec{s}'_{i',j}(x - \tau)$. Figure 4.5 shows two matched signals seen by two different sub-apertures observing the same target. This configuration is equivalent to one with a single sub-aperture observing two different stars. Their standard deviation is an indicator of their amplitudes since signals $\vec{s}_{i,j}(x)$ and $\vec{s}_{i',j}(x - \chi)$ before measurement have the same standard deviation $\sigma[\vec{s}_{i,j}]$. It is important to note that standard deviation is measured on finite data. Therefore, two signals without aberrations can only have the same standard deviation when it is calculated over

comparable values. Continuing, since for any process X with standard deviation σ , the standard deviation of $X' = aX$ is $\sigma' = a\sigma$, the ratio between standard deviations gives out the ratio between gains.

$$\frac{\sigma [s'_{x' i',j}(x - \tau)]}{\sigma [s'_{x' i,j}(x)]} = \frac{g_{x' i',j} \cdot \sigma [s_{x' i',j}(x - \chi)]}{g_{x' i,j} \cdot \sigma [s_{x' i,j}(x)]} \quad (4.14)$$

$$= \frac{g_{x' i',j}}{g_{x' i,j}} = g_{x' (i,i'),j} \quad (4.15)$$

The same applies for direction Y' in the local sub-aperture referential. Ratios $g_{x' (i,i'),j}$ and $g_{y' (i,i'),j}$ will be called on-off gains, because index i will be fixed to refer to the on-axis target, as i' is any other off-axis target. Because these gains are relative, for simplicity they

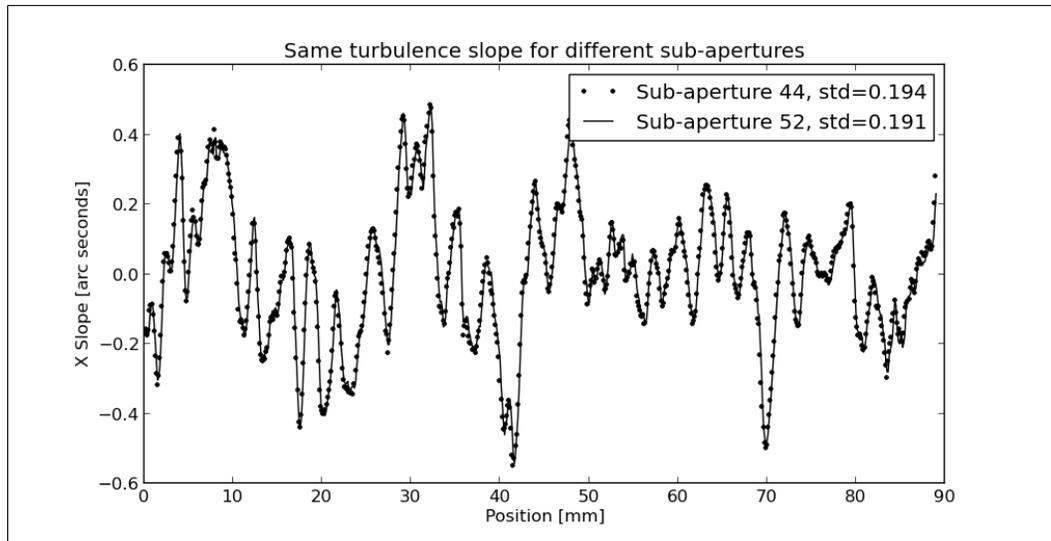


FIGURE 4.5. Same turbulence slopes for two different sub-apertures . By determining the offset of maximum correlation between the signals, the common observation is found and compared. The amplitude of the signals is proportional to their standard deviation. The relative gain between these neighbouring sub-apertures is 1.3%

will all be considered relative to the on-axis target. The relative gain of the on-axis target with itself is unitary. This way, relative gains describe the effects of off-axis aberrations over the existing on-axis aberrations. Figure 4.6 shows measurements of on-off gains in the X' and Y' directions as functions of field position. This allows to map the effects of off-axis aberrations on centroid motion throughout the field of view, producing a detailed

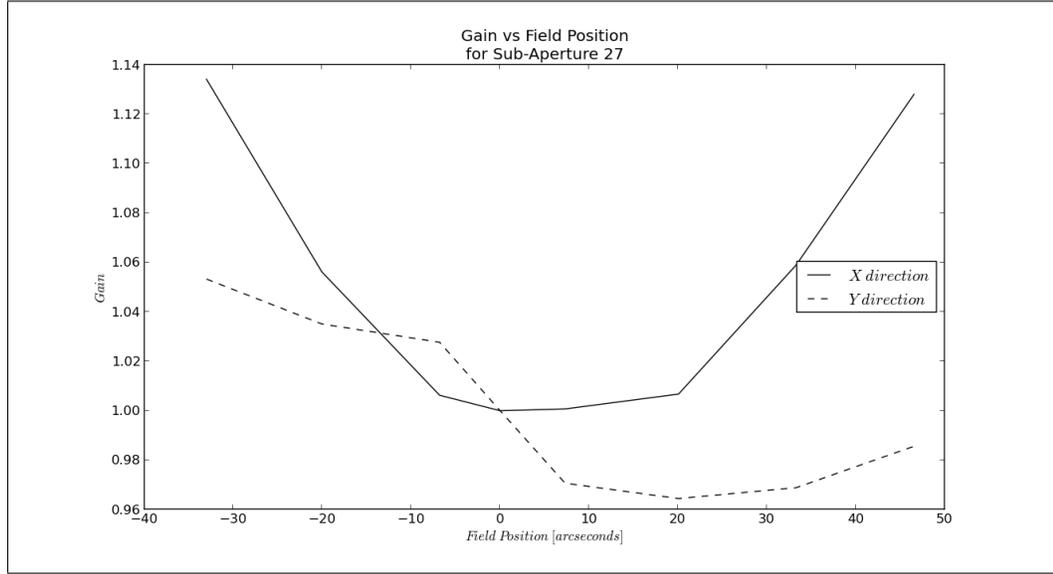


FIGURE 4.6. Field dependent gain, for the X and the Y direction. The phase screen moves along the X direction. Off-axis gain is mapped for each sub-aperture j

and comprehensive characterization of the optical system. Finally, on-ff relative gains can be arranged into a matrix

$$G_{on-off,(i,i')} = \begin{bmatrix} g_{x' (i,i'),0} & 0 & \dots & 0 \\ 0 & g_{y' (i,i'),0} & & \\ & & g_{x' (i,i'),1} & \\ \vdots & & & g_{y' (i,i'),1} & \ddots \\ 0 & & & & & g_{y' (i,i'),n} \end{bmatrix} \quad (4.16)$$

so that

$$G_{on-off,(i,i')} \cdot \vec{s}'_i(x) = \vec{s}'_{i'}(x - \tau) \quad (4.17)$$

Here, n is the number of sub-apertures. This matrix allows to take any set of measured slopes from having field-dependent aberrations in direction i , to having them in direction i' . When inverted, because it is diagonal, all values are inverted, allowing to go in the opposite sense from i' to i . This will prove useful in Section 5.2. It is important to note the off-axis aberration is only measured in the direction of the velocity of the wind. But, in polar

coordinates, if the off-axis aberrations can be consider to be only dependent on the radius and not on the angle, then a measurement in a single direction can be valid for all other angular directions.

4.4. Spot & Lenslet heterogeneity

Just as a single sub-aperture can produce two different signals observing two targets, two different sub-apertures can observe the same target, producing a signal each. Each sub-aperture, can be submitted to a different optical aberration. As explained in section 4.3.2, different aberrations will change the shape of the spots and in a first order approximation introduce a gain to its movement. Furthermore, if lenslets are not homogeneous, the difference between their focal lengths will also contribute to amplify the spots' motion. Then relative gains could be determined using the optimal filter to compare the signals' standard deviation. If both sub-apertures j and j' are aligned with the velocity of the wind, then

$$\vec{s}_{i,j}(x) = \vec{s}_{i,j'}(x - \chi) \quad (4.18)$$

Let target $i = 1$ be on-axis, as to have

$$\vec{s}_{1,j}(x) = \vec{s}_{1,j'}(x - \chi) \quad (4.19)$$

Comparing signals using the optimal filter would not be practical since, in order to compare two sub-apertures, wind velocity has to be aligned with them. Therefore, mapping all relative gains requires wind velocity to be aligned with all possible sub-aperture combinations. The solution is no to compare signals using an optimal filter, but to expose all sub-apertures to has much turbulence as possible, with a given direction of wind. If all the turbulence has the same profile, then as the data set acquired with a given sub-aperture grows, its standard deviation converges. In other words

$$\lim_{\Delta x \rightarrow \infty} \sigma [\vec{s}_{i,j}(x) |_{x_0}^{x_0 + \Delta x}] = \lim_{\Delta x \rightarrow \infty} \sigma [\vec{s}_{i,j}(x - \chi) |_{x_0}^{x_0 + \Delta x}] \quad (4.20)$$

$$= \sigma [\vec{s}_{i,j}] \quad (4.21)$$

where $\sigma [\vec{s}_{i,j}(x) |_{x_0}^{x_0+\Delta x}]$ is the standard deviation of signal $\vec{s}_{i,j}(x)$, calculated over the domain $[x_0, x_0 + \Delta x]$ for any given x_0 . The same is true for the signal with aberrations $\vec{s}'_{i,j}(x)$.

Again, just as for field dependent aberrations, but with a common target i and a couple of sub-apertures j and j' , it can be said

$$\frac{\sigma [s'_{x',i,j}]}{\sigma [s'_{x',i,j'}]} = \frac{g_{x',i,j} \sigma [s_{x',i,j}]}{g_{x',i,j'} \sigma [s_{x',i,j'}]} \quad (4.22)$$

$$= \frac{g_{x',i,j}}{g_{x',i,j'}} = g_{x',i,(j,j')} \quad (4.23)$$

These relative gains can be called on-on gains since $i = 1$ so they map on-axis aberrations. On-off gains were all relative to the on-axis direction. Here, there is no criteria to chose one sub-aperture over the others in order to make all gains relative to said sub-aperture. For now, standard deviations $\sigma [s'_{x',1,j}]$ will be consider a measure of sub-aperture homogeneity as showed in Figure 4.7.

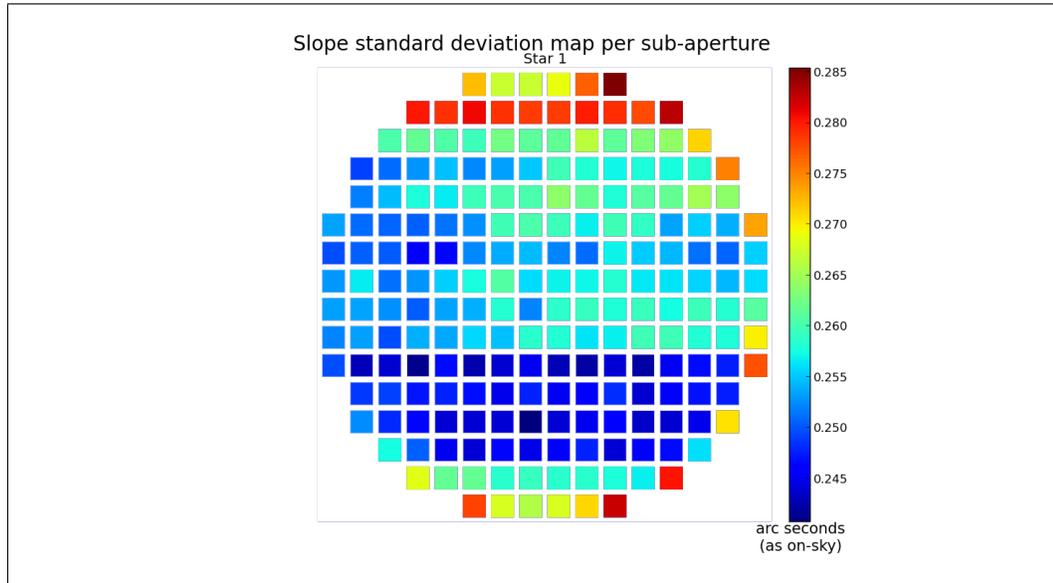


FIGURE 4.7. Slope standard deviation map per sub-aperture. The overall average for all sub-apertures depends on the turbulence profile. The relative differences depends on the aberrations.

Without having to calculate on-on gains, this map directly allows to assess the heterogeneity in how sub-apertures measure slopes. Normalizing values against their average, produces a set of gains, centred around 1. Just as with $G_{on-off,(i,i')}$, a diagonal matrix

$$G_{on-on} = \begin{bmatrix} g_{x',0} & 0 & \cdots & 0 \\ 0 & g_{y',0} & & \\ & & g_{x',1} & \\ \vdots & & & g_{y',1} & \ddots \\ 0 & & & & & g_{y',n} \end{bmatrix} \quad (4.24)$$

can be built with these normalized values so that

$$G_{on-on} \vec{s}_{i=0}(x) = \vec{s}'_{i=0}(x) \quad (4.25)$$

When inverting this matrix, a set of perfect simulated data can be tuned into having the same heterogeneity as the instrument. This technique and its purpose will be revisited in 5.2 to tune numerical simulations of wavefront sensors so they match the aberrations of true wavefront sensors, or to demodulate experimental data so it becomes homogeneous.

4.5. Phase screen characterization

To introduce phase aberrations as those caused by atmospheric turbulence, a test bench can use a phase screen. The phase screen is manufactured to match a required turbulence profile. To determine the goodness of fit of the manufactured profile with the intended one, the phase screen has to be characterized.

As seen in Chapter 2, a stochastic noise process such as turbulence is well characterized by its covariance, which translate into its power spectral density in the Fourier domain through Wiener-Khinchin's theorem. In turn, they are both defined by the strength of the turbulence $\tilde{C}_n^2 \delta z$ and the outer scale L_0 . Characterizing turbulence means finding values for a model's parameters so the error between their covariances (the model's and the measured

turbulence's) is minimized. Here, only von Karman's model will be consider as it is well suited to developed turbulence.

Both covariances have to be calculated so the model can be fitted to the data. First is calculated the covariance of the measured signal. Equation 2.13 allows to calculate a function's covariance when knowing its power spectral density. The power spectral density of a real function $s(x)$ is

$$W(f) = (\mathcal{F}\{s(x)\}) \cdot (\mathcal{F}\{s(x)\})^* \quad (4.26)$$

Replacing equation 4.26 in equation 2.13 allows for direct computation of any signal's covariance. The resulting equation,

$$B(x) = \mathcal{F}^{-1}\{(\mathcal{F}\{s(x)\}) \cdot (\mathcal{F}\{s(x)\})^*\} \quad (4.27)$$

can be defined in any number of dimensions. In Chapter 2, Equation 2.34 allows calculation of covariance $B_{\alpha lp}[\lambda, x, y]$ for two points apart in both the X and the Y directions. In the case of a single sub-aperture with wind parallel to the X axis, slopes in each local direction will constitute one-dimensional signals. From slopes $s_{x' i,j}(x)$, in the X' direction, can be calculated the longitudinal covariance $B_{\alpha lp}[\lambda, x, 0]$ and from slopes $s_{y' i,j}(x)$, in the Y' direction, can be calculated the transversal covariance $B_{\alpha tp}[\lambda, x, 0]$. A simple rotation in the reference system turns the transversal covariance into a longitudinal covariance $B_{\alpha lp}[\lambda, 0, y]$ evaluated in $x = 0$. Therefore, instead of a two-dimensional covariance matrix, a sub-aperture with one wind direction can only produce its central column and central row.

As seen previously in Section 4.4, different sub-apertures have different gains as a result of phase aberrations and heterogeneous construction of the lenslets. Since none can be defined as the reference sub-aperture, all sub-apertures have to be taken into account to measure the phase screen's covariance matrix. Resulting covariance matrices have to be averaged into a single, low noise matrix.

Later the model's covariance has to be calculated. Replacing the physical model given by Equation 2.36 into Equation 2.34 produces the model's covariance $B_{\alpha p} [\lambda, x, y]$. Sampling the theoretical covariance, with the same sample size as the WFS measurements, produces a covariance matrix from which has to be considered the central column and central row. Only they can be compared to the covariance vectors calculated from the data. Finally, the model is fitted in a least squares sense using the Levenberg-Marquardt algorithm. The error to be minimized is

$$e = \|B_{\alpha p, data} [\lambda, x, 0] - B_{\alpha p, model} [\lambda, x, 0]\|^2 + \|B_{\alpha p, data} [\lambda, 0, y] - B_{\alpha p, model} [\lambda, 0, y]\|^2 \quad (4.28)$$

It is important to consider a fair initial guess, since the algorithm can only find local minima. For a phase screen characterized with BEAGLE, and according to the manufacturer, the length coherence or Fried's parameter is $r_0 = 0.6 [mm]$. Translated into sky dimensions $r_{0 on-sky} = 14 [cm]$. This parameter has to be turned into one of the two parameters considered in the model. In (Hardy, 1998) the relationship between the strength of the turbulence and Fried's coherence length is given by

$$r_0 = \left[0.423k^2 sec \varsigma \int_{Vertical} C_n^2(z) dz \right]^{-3/5} \quad (4.29)$$

from which

$$\begin{aligned} \tilde{C}_n^2 \delta z &= \frac{r_0^{-5/3}}{0.423k^2 sec \varsigma} \\ &= 6 \cdot 10^{-9} [m^{-1/3}] \end{aligned} \quad (4.30)$$

where $k = 2\pi/\lambda$ and $\lambda = 650 \cdot 10^{-9} [m]$. This result is at a test bench scale (not the on-sky scale).

Later, the outer scale is estimated from the structure function, also provided by the manufacturer. Also in (Hardy, 1998), the structure function is defined as

$$D(\tau) = 2 \cdot (B(0) - B(\tau)) \quad (4.31)$$

so the resulting curve is visually more representative of the turbulence profile. When plotted in a log-log scale, the slope of the curve gives the r_0 and the argument at which it saturates is the outer scale.

Figure 4.8 presents structure functions for a phase screen characterized in a test bench. In the same plot are included the structure functions of the fitted model. From the data

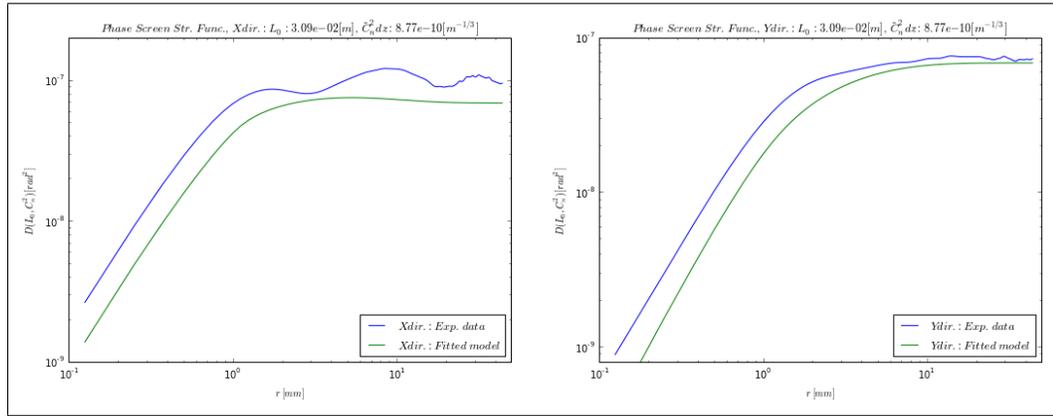


FIGURE 4.8. Structure functions in the X and in the Y directions. Functions obtained from true data are compared with functions of a fitted model.

and from the fitting, the phase screen's outer scale is $L_0 = 3.1 \cdot 10^{-2} [m]$. Translated into sky dimensions $L_{0 \text{ on-sky}} = 7.2 [m]$. This is a small outer scale for any turbulence on-sky. It explains the little variance and strength of the measured slopes, as the turbulence saturates before integrating energy from larger scales. It was also found that $\tilde{C}_n^2 \delta z = 8.77 \cdot 10^{-10} [m^{-1/3}]$. Replacing it in Equation 4.29 gives $r_0 = 1.9 [mm]$ which represents a 216% error against the manufacturer's value, and on-sky translates to a value of $r_{0 \text{ on-sky}} = 44 [cm]$.

Finally, it can also be seen the structure function oscillates for slopes parallel to wind propagation. In other words, covariance does not converge to zero, but oscillates on top of it. This is due to an imperfection in the phase screen which produces a frequency in this direction. This frequency could interfere when training a tomographic reconstructor, since it would learn it and then expect it, without being able to generalize out of it. Because of this, it is important to train reconstructors with as many different and general scenarios as possible. This way, it would not fix on any specific pattern, and therefore would learn to

generalize for all possible cases. Training through simulation as proposed in the following, Chapter 5, allows to produce training data free of fix patterns.

5. TOMOGRAPHIC RECONSTRUCTORS' RANGE OF VALIDITY

As seen in Section 3.2, tomographic reconstructors can be based on artificial neural networks (ANN). When so, their range of validity is wider than when using linear reconstructors such as *Learn & Apply*. Linear reconstructors are finely tuned to work for specific seeing conditions and altitudes. On the contrary, a sufficiently complex ANN can model any problem, and therefore, if properly trained, handle scenarios at multiple altitudes and seeing conditions. Being the most general, ANN based tomographic reconstructors may, from now on, be referred to simply as reconstructors, as opposed to linear reconstructors.

The error budget of linear reconstructors against instrumental deviations and noise is well understood (Vidal et al., 2014). This is not the case for ANN based reconstructors. These have been trained and validated in simulated conditions (Osborn et al., 2012) and also have been trained with experimental data and then validated under experimental conditions (Osborn et al., 2014). But never before has a reconstructor been trained with simulated data and then tested against real experimental conditions. In such an exercise the reconstructor's performance is expected to decrease as of the introduction of instrumental deviations and turbulence variability in long time scales.

The proposed thesis of this work is a reconstructor trained using simulated data can reach an acceptable degree of validity in experimental conditions, when considering a sufficiently precise instrumental characterization. The parameters retrieved from the characterization can be used either to adjust simulated data so it resembles experimental data, or to do the opposite. First, a network is trained in Section 5.1, using simulated data only. In Section 5.2, a method is proposed to correct instrumental on-axis and off-axis gains, using matrices G_{on-on} and G_{on-off} from Sections 4.4 and 4.3 respectively. Corrected data is passed to the reconstructor and the difference in performance is assessed. The results of this technique, although positive, are limited. Section 5.3 offers an explanation to its limitation through a sensitivity analysis. The analysis is also applied to other sources of error a reconstructor might encounter, such as mis-pointing of the guide source wave-front sensors

(WFS) and variations of the outer scale. The reconstructor’s sensitivity is discussed as it offers an understanding on how to train it so it remains valid in experimental conditions.

5.1. Training an artificial neural network based tomographic reconstructor

To test the proposed method, an artificial neural network based tomographic reconstructor is trained. The chosen constellation is as shown in Figure 5.1. Its guide stars are close to being aligned with the direction of wind propagation X and with Y . This will prove to be important in Section 5.2. To this purpose, the chosen constellation was amongst the best available in the test bench at hand, BEAGLE.

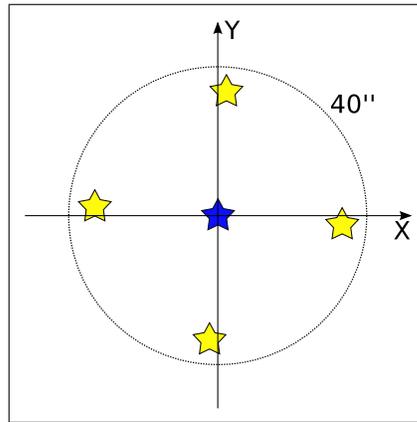


FIGURE 5.1. Guide stars constellation, with target star at center. X is the direction of wind propagation.

A numerical simulation is set using this constellation to train the network. In order to have a comparable reconstructor, the training method is similar to that used in (Osborn et al., 2012). The ANN is a multi-layer perceptron back-propagation (MLP-BP) network. It is presented with random phase screens, at individual altitudes, for which $r_0 = 25 [cm]$ and $L_0 = 30 [m]$. The tomographic volume, going from ground layer to $h_{max} = 13020 [m]$ is sampled 625 times at each individual altitude, at a resolution of 80 [m] between altitudes. The simulation represents the 4.2 [m] in diameter William-Herschel telescope, sampled by a 16x16 Shack-Hartmann WFS. The ANN is validated against three atmospheric profiles, again, similar to those in (Osborn et al., 2012), and presented in Table 5.1.

TABLE 5.1. Atmospheric profiles for validation.

		Atmosphere		
		atm1	atm2	atm3
WF-RMS [nm]		759	758	758
Layer 1	Altitude [m]	0	0	0
	Relative strength	0.65	0.45	0.8
Layer 2	Altitude [m]	4000	2500	6500
	Relative strength	0.15	0.15	0.05
Layer 3	Altitude [m]	10000	4000	10000
	Relative strength	0.10	0.30	0.10
Layer 4	Altitude [m]	13020	13020	13020
	Relative strength	0.10	0.10	0.05

The results of the best reconstructor achieved are presented in Table 5.2. Here they are compared against the results in (Osborn et al., 2012) as reference. Although having further apart guide stars, the residual wave-front error (WFE) is smaller in all cases to that of the reference. This is mainly due to having an extra guide source (four instead of three). Another contributing factor comes from using a 16x16 Shack-Hartmann (SH) instead of a 7x7 one. As the performance of the reconstructor is considered comparable to that of an accepted counterpart, it is used to test the method for correcting instrumental deviations, proposed in the following section.

TABLE 5.2. Reconstructor validation. The achieved reconstructor is compared against a reference. Here WFE-RMS[%] shows average WFE relative to average WF.

Atmosphere		μ_{WF-RMS} [nm]	$\mu_{WFE-RMS}$ [nm]	$\mu_{WFE-RMS}$ [%]
atm1	Reference	644	231	35.9
	Achieved	703	216	30.7
atm2	Reference	817	262	32.1
	Achieved	703	198	28.2
atm3	Reference	1088	387	35.6
	Achieved	694	203	29.3

5.2. Correction of instrumental deviations: on-axis & off-axis gains

When simulating an adaptive optics (AO) system, considerations on the pointing of the WFSs and on the outer scale have to be hard coded *a priori* into the simulation. If not, resulting data can not be modified *a posteriori* to match new desired values for these parameters. As seen in Sections 4.4 and 4.3, this is not true for sub-aperture heterogeneity and for field-dependent aberrations, also referred to as on-axis gain and off-axis gains respectively. If a reconstructor is trained using perfect simulated data, its performance decreases when presented experimental data, with on-axis and off-axis gains. To address this a method is devised that subtracts both types of gains from experimental data before presenting it to a reconstructor, so it resembles simulated data, and then introduced back after the reconstruction. This way the reconstructor does not operate on data as seen by a particular instrument, but is brought to a base understandable by any reconstructor.

If \vec{s}' is a vector of slopes measured by an imperfect instrument and \vec{s} are the expected slopes without aberrations, it can be said a perfect simulation produces the last. In order to add heterogeneity and field-dependent aberrations to simulated slopes, Equations 4.17 and 4.25 are used. Combining them gives

$$\begin{aligned} G_{on-off,(i=0,i')} \cdot G_{on-on} \cdot \vec{s}_{i=0}(x) &= G_{on-off,(i=0,i')} \cdot \vec{s}'_{i=0}(x) \\ &= \vec{s}'_{i'}(x - \tau) \end{aligned} \quad (5.1)$$

which can be modified into

$$G_{on-off,(i=0,i')} \cdot G_{on-on} \cdot \vec{s}_{i=0}(x + \tau) = \vec{s}'_{i'}(x) \quad (5.2)$$

Here, direction $i = 0$ is on-axis and i' is any other direction. Finally, if $\vec{s}_{i=0}(x + \tau) = \vec{s}_{i'}(x)$, then replacing into 5.2 gives

$$G_{on-off,(i=0,i')} \cdot G_{on-on} \cdot \vec{s}_{i'}(x) = \vec{s}'_{i'}(x) \quad (5.3)$$

Equation 5.3 above allows to transform slopes produced in simulation into slopes as seen by an instrument. But this is only possible when directions i and i' are aligned with

the axis X of wind velocity. Since an AO experiment only counts with a finite amount of wind propagation directions (typically one), matrix $G_{on-off,(i=0,i')}$ cannot be measured and determined for all directions i' .

To deal with this, field dependent aberrations are considered azimuthally symmetric relative to the optical axis of the system. If true, then it is possible to determine $G_{on-off,(i=0,i'')}$ for a direction i'' that is not aligned with $i = 0$ and wind velocity, from $G_{on-off,(i=0,i')}$ when i' and i are aligned. The condition is the distance between $i = 0$ and i' is the same as that of i and i'' . Figure 5.2 shows the described configuration.

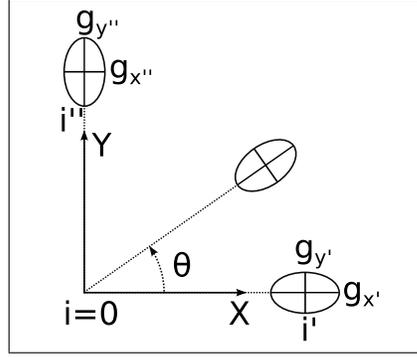


FIGURE 5.2. Star directions relative to wind direction in the X axis. Direction $i = 0$ is on-axis, i' is aligned with the wind, and i'' is the same distance from $i = 0$ as i' . Field dependent aberrations are considered azimuthally symmetric. For $\theta = 90^\circ$, $g_{x'} = g_{y''}$ and $g_{y'} = g_{x''}$.

For simplicity i'' will be considered at $X = 0$, with $\theta = 90^\circ$. As can be seen in the figure, for this particular case $g_{x'} = g_{y''}$ and $g_{y'} = g_{x''}$. This allows to calculate $G_{on-off,(i=0,i'')}$.

If field aberrations are not considered azimuthally symmetric, then gains cannot be estimated for directions i'' not aligned with X . In such gain values could be left unitary, with $g_{x''} = 1$ and $g_{y'} = 1$.

Having solved this, Equation 5.3 can be extended to a multi-object adaptive optics (MOAO) system with four guide stars $i = 1, 2, 3, 4$ and a target $i = 0$. This is done by concatenating the measurements from all WFSs and diagonally concatenating matrices

$G_{on-off,(i,i')}$ and G_{on-on} as

$$\vec{s} = \left[\vec{s}_1^T \quad \vec{s}_2^T \quad \vec{s}_3^T \quad \vec{s}_4^T \right]^T \quad (5.4)$$

$$\mathbf{G}_{on-on} = \begin{bmatrix} G_{on-on} & 0 & 0 & 0 \\ 0 & G_{on-on} & 0 & 0 \\ 0 & 0 & G_{on-on} & 0 \\ 0 & 0 & 0 & G_{on-on} \end{bmatrix} \quad (5.5)$$

$$\mathbf{G}_{on-off} = \begin{bmatrix} G_{on-off,(0,1)} & 0 & 0 & 0 \\ 0 & G_{on-off,(0,2)} & 0 & 0 \\ 0 & 0 & G_{on-off,(0,3)} & 0 \\ 0 & 0 & 0 & G_{on-off,(0,4)} \end{bmatrix} \quad (5.6)$$

Resulting equation is

$$\mathbf{G}_{on-off} \cdot \mathbf{G}_{on-on} \cdot \vec{s} = \vec{s}' \quad (5.7)$$

where $\mathbf{G} = \mathbf{G}_{on-off} \cdot \mathbf{G}_{on-on}$. If aberrations are considered azimuthally symmetric, then $\mathbf{G} = \mathbf{G}_{XY}$, if not then $\mathbf{G} = \mathbf{G}_{X1}$.

The process of subtracting gains is called demodulation, and consists in multiplying slopes by \mathbf{G}^{-1} . The process of reintroducing them is called modulation and consists in multiplying by \mathbf{G} . Figure 5.3 shows the process of feeding a reconstructor demodulated data and modulating it at its end. The entire operation is intended to emulate a reconstructor trained with experimental data.

Using the test bench, an experiment is produced with the same guide stars and target as in Figure 5.1. The turbulence layer characterized in 4.5 is swept at three different altitudes. Guide sources measurements \vec{s}' and target measurements \vec{s}'_0 are produced. The network trained in Section 5.1 was so using only simulated data. First, the ANN's performance is tested against raw measurements, without any consideration of instrumental deviations. This is considered the worst case scenario upon which the correction technique should provide with an improvement. Then the ANN is tested against perfect simulated data of turbulence at the same three altitudes. This is considered the best case scenario. Both, the

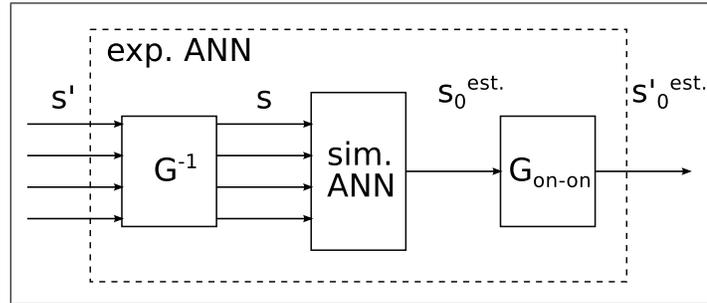


FIGURE 5.3. Demodulation and modulation. An ANN trained through simulation receives demodulated data from an instrument. The ANN produced an estimation of the target's slopes. The result is modulated at its end. The whole acts as an ANN trained with experimental data.

best and the worst case, provide a benchmark to assess the technique's results. Table 5.3 presents the results of implementing the modulation and demodulations modules. These in turn have two versions, one with G_{X1} and the other with G_{XY} .

TABLE 5.3. Instrumental deviations correction results. The ratio between WFE-RMS and WF-RMS is calculated for each random phase screen. Its average is shown in column $\mu_{WFE/WF}$ and its standard deviation in column $\sigma_{WFE/WF}$.

Altitude [m]	Origin	μ_{WF-RMS} [nm]	$\mu_{WFE/WFE}$ [%]	$\sigma_{WFE/WF}$ [%]
1442	Simulated [best case]	696	7.1	3.6
	Raw [worst case]	474	51.0	18.8
	G_{X1}	474	50.8	18.1
	G_{XY}	474	51.3	17.3
3000	Simulated [best case]	703	8.1	3.8
	Raw [worst case]	464	77.3	22.5
	G_{X1}	464	78.3	21.8
	G_{XY}	464	77.6	20.4
5000	Simulated [best case]	701	8.7	4.0
	Raw [worst case]	463	76.5	23.5
	G_{X1}	463	77.2	22.7
	G_{XY}	463	76.6	21.4

When presenting the ANN with raw instrumental data, the relative residual WFE increases in an order of magnitude. When applying the modulation technique, results do not improve. Any appreciable reduction of the error is negligible. In other words the correction fails to compensate any significant instrumental deviation. There are two non-exclusive explanations to this. The first is the characterization resulting in matrices \mathbf{G}_{on-off} and \mathbf{G}_{on-on} was not precise enough. The second is sub-aperture heterogeneity and field-dependent aberrations combined are not a significant source of error. Other deviations, not addressed by the proposed technique, would have a larger weight. If the reconstructor is insensitive to multiplying incoming data by a gain matrix \mathbf{G} , then not even the most precise characterization of on-axis and off-axis gains would result in any improvement, discarding the first explanation. The later is further explored in Section 5.3

5.3. Sensitivity analysis

When presenting an ANN based reconstructor with data from a measurement, any deviation from the training data set may lead to a decrease in the estimation's performance. As seen in the previous section, this can be critical as the error can increase up to three times from its expect value. To deal with this, a method was devised to correct field-dependent aberrations and sub-aperture heterogeneity. But there might be other differences between the experiment and the simulation that drive the reconstruction out of its space of validity. Three major sources of error are identified. The first are measurement distortions, amongst which can be found the off-axis and on-axis gains. Distortions refer, in this case, to any phenomena introducing an amplification factor or gain to the measurements. They can be modelled by a diagonal matrix \mathbf{G} multiplying the expected measurement vector. The second source of error might be the difference in outer scale. The outer scale shapes the turbulence's spatial correlation, which determines the correlation between sup-apertures. If a reconstructor learns a given correlation pattern, it might not be able to generalize when facing a new one. The last deviation is here referred to as pointing and relates to the positions of the guide sources and of the target in an experiment, against their expected ones. When preparing and adaptive optics experiment in an optical test bench, the light

sources representing the stars must be set in position to match a chosen constellation. The pointing is the process of setting the stars and then setting their corresponding WFSs. This is done with a certain precision. An error in pointing changes the direction in which the tomographic volume is seen, and therefore the correlation between sub-apertures.

The sensitivity of a reconstructor to all of these deviations is assessed. To do so, the ANN is presented with simulated data, where the simulation has been modified to include one of the errors to test. The estimation's accuracy is then compared with that obtained without the errors. Not only is this of interest in the context of the proposed hypothesis, but also when using standard reconstructors (trained with instrumental data) in changing conditions. All ANN based reconstructors are trained days before being used. When so, they all must deal with pointing errors and with a unique seeing. A sensitivity analysis is then essential to determining the range of validity of a reconstructor and to estimate its performance and the quality of its results. This is something not yet studied in detail.

The first source of error introduced are distortions. As seen, they can be represented by a diagonal gain matrix. Determining the effect these matrices have on a reconstructor's performance allows to quantify the range of improvement possible to achieve when correcting for them. To do so, random matrices \mathbf{G} are generated with a normal distribution and the same statistical parameters as used in the modulation. In Section 5.2, the elements of modulation matrix \mathbf{G} had an average of 1 by definition and a standard deviation $\sigma_{diag(\mathbf{G}_{X1})} \approx \sigma_{diag(\mathbf{G}_{XY})} \approx 0.11$. Being gains, these values are normalized and have no unit. The residual WFE-RMS of the reconstructor is tested when presented with data distorted with a large set of these random matrices. Simulation considers the first atmospheric profile described in Table 5.1. Finally, to better assess the sensitivity of the ANN to distortions, matrices are also generated using larger standard deviations. Results are presented in Table 5.4.

When $\sigma_{\mathbf{G}} = 0.11$, average WFE-RMS increases only by 0.4%. This means this is the average improvement that can be expected when correcting for distortions of that order.

TABLE 5.4. WFE-RMS when introducing distortion matrices \mathbf{G} . Simulation considers atmospheric profile "atm1". The ratio between WFE-RMS and WF-RMS is calculated for each random phase screen. Its average is shown in column $\mu_{WFE/WF}$ and its standard deviation in column $\sigma_{WFE/WF}$.

$\sigma_{\mathbf{G}}$	μ_{WF-RMS} [nm]	$\mu_{WFE/WF}$ [%]	$\sigma_{WFE/WF}$ [%]
0.00	703	35.3	15.3
0.11	703	35.7	15.3
0.22	705	37.0	15.3
0.33	707	38.8	15.3

This helps explain actual correction results in Table 5.3, where improvements are not noticeable. For larger distortions, the error grows at an increasing rate, but still at a slow one. Since a distortion with $\sigma_{\mathbf{G}} = 0.33$ is larger than could ever be expected, the ANN can be said to be very robust against them. Therefore, off-axis and on-axis gains are not a major source of error and not much can be expected when correcting for them.

The next error introduced is that of pointing. Pointing errors, specially in an optical experiments, can be the product of a misalignment of the guide source and the target, relative to the WFSs. A misalignment can change the magnification and the rotation of a constellation. The magnification affects the apparent size of an asterism, changing the estimated altitude of a turbulence layer. Since the effect of altitude changes are well understood (Osborn et al., 2014), the focus is put on rotations. Simulations are then performed where guide sources are rotated at different angles. In Section 4.2, guide sources' rotation was estimated to be 2.6 [°]. Then, the range of tested rotation angles will well include that value, leaving it at its middle. Again, simulation will consider atmospheric profile "atm1" as described in the previous section. Results are presented in Table 5.5.

When rotating an amount similar to that of the estimated rotation in the test bench, average WFE-RMS only increases 0.7%. When rotating 5.7° , which again, is a larger rotation to what can be expected in normal conditions, the error only increases 1.3%. With guide stars at about 33 [arcsec] from the optical axis, the largest rotation translates the projected pupil of 1.6 [cm] at an altitude of 1 [km], and 16 [cm] at an altitude of 10 [km]. With sub-apertures of 26.25 [cm] (16 sub-apertures across a 4.2 [m] pupil), this means

TABLE 5.5. WFE-RMS when rotating guide sources. Simulation considers atmospheric profile "atm1". The ratio between WFE-RMS and WF-RMS is calculated for each random phase screen. Its average is shown in column $\mu_{WFE/WF}$ and its standard deviation in column $\sigma_{WFE/WF}$.

Rotation [degrees]	μ_{WF-RMS} [nm]	$\mu_{WFE/WF}$ [%]	$\sigma_{WFE/WF}$ [%]
0	703	35.3	15.3
0.6	701	35.3	15.1
2.9	699	35.8	15.4
5.7	704	36.7	15.4

translations of approximately 6% and 61%, respectively. Other reconstructors such linear ones, are very susceptible to such translations. This simulation shows ANNs are fairly good at generalizing for rotation errors of pointing. Other pointing errors that are not tested here relate to each WFS's alignment. MOAO systems use pick-off mirrors to feed their WFSs. Their alignment is a difficult task, and the error of these process his difficult to quantize. Further work includes simulating individual misalignments of the WFSs, and then combining it with rotation errors.

The last difference introduced in the simulation is the change in the outer scale. Until now, ANNs are trained using a constant L_0 . It is unclear whether a reconstructor trained this way can generalize for different values of this parameter. Larger values of L_0 mean more tightly correlated turbulence. In turn, this improves the turbulence's predictability over a given distance. A good reconstructor is one that improves its estimation as the outer scale increases. A bad reconstructor is one for which the residual error increases when moving apart from the outer scale at which it was trained. For this test, multiple sets of slopes are produced through simulation. L_0 is different for each set. This time, instead of considering an atmospheric profile consisting of four layers at different altitudes, single phase screens are produced at random altitudes ranging from ground layer to 13020 [m] (maximum altitude of overlap). WFE-RMS values for each set are shown in Figure fig:outerscales.

In the figure, the error converges to a minimum as the outer scale increases. This occurs without any anomaly or noticeable minimum around $L_0 = 30 [m]$, which is the value used for training. The reconstructor can then be said to be good, as it generalizes correctly,

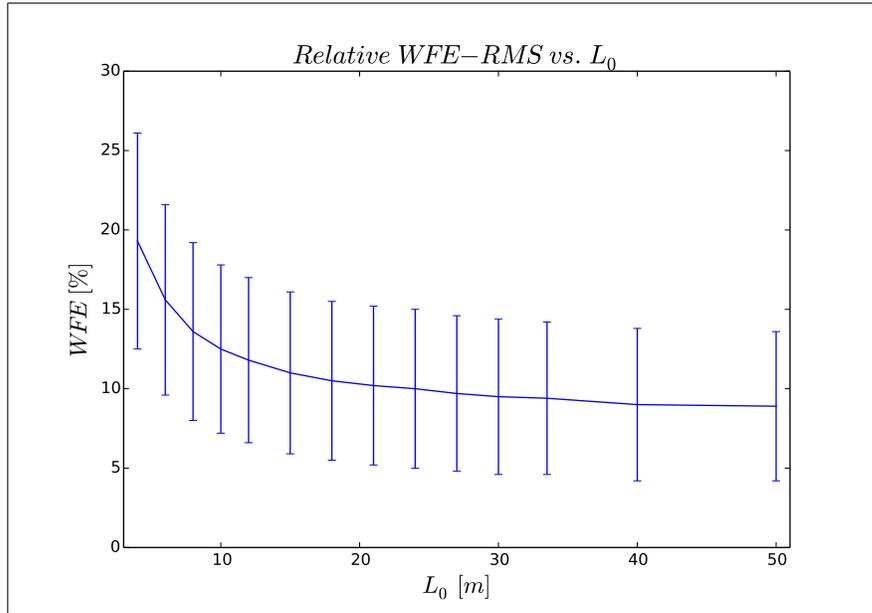


FIGURE 5.4. Average WFE-RMS, relative to average uncorrected WF-RMS vs. outer scale. Error bars equal 2σ , where σ is the standard deviation of WFE-RMS for all simulated cases, also relative to average uncorrected WF-RMS. At each value of L_0 , all altitudes are represented evenly.

without regard to the outer scale shown during training. With respect to the reconstructor’s sensitivity to this parameter, it can be said to be low when using only one layer. Considering the outer scale of a single layer in a good site like Paranal can go from about 5 [m] under bad conditions to roughly 50 [m] (Ali et al., 2010), then the average WFE-RMS would suffer at most a 15% variation. Although little, this is the largest performance variation obtained with all three sources of error introduced. Since the experimental data was measured to have an outer scale of $L_0 = 7.2$ [m], according to the results the relative WFE-RMS should have been around 15%, instead of 50% to 70%.

None of the individual errors introduced account for the difference in performance when presenting the ANN with simulated experimental data. As shown, these types of tomographic reconstructors are robust against single and expected instrumental deviations. Future work will assess scenarios with multiple sources of error, to determine whether together, their effects are magnified, or if other deviations remain to be considered.

6. CONCLUSIONS

The hypothesis of this work is the deviations and aberrations of an instrument can be quantified and incorporated into a simulation, to train a reconstructor valid in experimental conditions. With this in consideration, novel methods of characterization are presented in Chapter 4. These methods quantify sub-aperture heterogeneities and field-dependent aberrations, also called on-axis and of-axis aberrations respectively. The result is used in Chapter 5 to build correction matrices \mathbf{G}_{on-on} and \mathbf{G}_{on-off} . These matrices mitigate the previously mentioned instrumental deviations. An ANN based tomographic reconstructor would then see corrected measurements as coming from a perfect instrument or simulation. This way, data resembles that produced in simulation and presented to the ANN for training, better falling in its range of validity.

When implementing the correcting technique improvements are negligible. This calls for a sensitivity analysis. The very same error is introduced to simulated data in various degrees, to assess its progressive impact on the reconstructor's estimation. As expected, the ANN is robust to such distortions, which are handled as noise. Even when they are magnified by threefold over the estimated distortions, the overall residual WFE only increases by about 3%. The major sources of error between the simulated model and the experiment are not being addressed and correcting for them has as little impact as they do.

There may be other significant sources of error. New sensitivity analyses are conducted in order to determine their estimated impact. Two likely candidates are rotation of the field of view and variations of outer scale. From the test bench's calibration and characterization, the field's rotation and the outer scale were estimated to be $2.6 [^\circ]$ and $L_0 = 7 [m]$ respectively. Considering the simulation had no rotation and the outer scale was $L_0 = 30 [m]$, these differences between the model and the experiment could account for most of the error.

Simulations were performed adding as much as twice the estimated rotation of the field, and sweeping from $L_0 = 4 [m]$ to $L_0 = 30 [m]$. In the case of rotations, it accounts for less than a 2% increase of the WFE when rotating $5.7 [^\circ]$. Since this greatly exceeds any

expectable rotation, it can be said the ANN is robust against all rotations it might encounter. Variations of the outer scale, on the other hand, produced a larger error. Simulation shows the difference between the outer scale used for training and that of the experiment accounts for about a 5% increase of the WFE, relative to the WF-RMS. This result is most interesting since ANN based reconstructors often face seeing conditions other than those presented in training.

Individual deviations do not account for the deterioration of the ANN's estimation when presented with experimental data. Also, little is known about tomographic reconstructors' range of validity and sensitivity. Further work consists in considering other instrumental deviations for analysis, and also introducing multiple ones at a time. This could help bring new understanding on building more robust ANNs. And relative to this work's hypothesis and goal, this would help devise future correction techniques, that take into account the most relevant sources of error.

BIBLIOGRAPHY

- Ali, W. D., Ziad, A., Berdja, A., Maire, J., Borgnino, J., Sarazin, M., et al. (2010). Multi-instrument measurement campaign at paranal in 2007-characterization of the outer scale and the seeing of the surface layer. *Astronomy & Astrophysics*, 524, A73.
- Assémat, F., Gendron, E., & Hammer, F. (2007). The falcon concept: multi-object adaptive optics and atmospheric tomography for integral field spectroscopy—principles and performance on an 8-m telescope. *Monthly Notices of the Royal Astronomical Society*, 376(1), 287–312.
- Beckers, J. (1989). Detailed compensation of atmospheric seeing using multiconjugate adaptive optics. In *1989 orlando symposium* (pp. 215–219).
- Born, M., & Wolf, E. (1999). *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light*. Cambridge: Cambridge university press.
- Bottaci, L., Drew, P. J., Hartley, J. E., Hadfield, M. B., Farouk, R., Lee, P. W., et al. (1997). Artificial neural networks applied to outcome prediction for colorectal cancer patients in separate institutions. *The Lancet*, 350(9076), 469–472.
- De Villiers, J., & Barnard, E. (1993). Backpropagation neural nets with one and two hidden layers. *Neural Networks, IEEE Transactions on*, 4(1), 136–141.
- Gendron, E., Vidal, F., Brangier, M., Morris, T., Hubert, Z., Basden, A., et al. (2011). Moao first on-sky demonstration with canary. *Astronomy and Astrophysics*, 529, L2.
- Guzmán, D., Juez, F. J. d. C., Myers, R., Guesalaga, A., & Lasheras, F. S. (2010). Modeling a mems deformable mirror using non-parametric estimation techniques. *Optics express*, 18(20), 21356–21369.

Hardy, J. (1998). *Adaptive optics for astronomical telescopes* (1st ed.). Oxford, UK: Oxford University Press.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5), 359–366.

Janesick, J. R. (2007). *Photon transfer*. Orlando, FL: SPIE press San Jose.

Le Louarn, M., & Hubin, N. (2004). Wide-field adaptive optics for deep-field spectroscopy in the visible. *Monthly Notices of the Royal Astronomical Society*, 349(3), 1009–1018.

Morris, T., Hubert, Z., Myers, R., Gendron, E., Longmore, A., Rousset, G., et al. (2010). Canary: The ngs/lgs moao demonstrator for eagle. In *1st ao4elt conference-adaptive optics for extremely large telescopes* (p. 08003).

Noll, R. (1976). Zernike polynomials and atmospheric turbulence. *Journal of the Optical Society of America*, 66(3), 207-211.

Osborn, J., De Cos Juez, F. J., Guzman, D., Butterley, T., Myers, R., Guesalaga, A., et al. (2012). Using artificial neural networks for open-loop tomography. *Optics express*, 20(3), 2420–2434.

Osborn, J., Guzman, D., Cos Juez, F. de, Basden, A., Morris, T., Gendron, E., et al. (2014). Open-loop tomography with artificial neural networks on canary: on-sky results. *Monthly Notices of the Royal Astronomical Society*, 441(3), 2508–2514.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1988). Learning representations by back-propagating errors. *Cognitive modeling*, 5.

Sáez, N., Basden, A., Guzmán, D., Dubost, N., & Berdja, A. (2014). Using darc in a multi-object ao bench and in a dome seeing instrument. In *Spie astronomical telescopes+ instrumentation* (pp. 91521X–91521X). Bellingham, Washington: SPIE press.

Southwell, W. H. (1980). Wave-front estimation from wave-front slope measurements. *JOSA*, 70(8), 998–1006.

Tokovinin, A. (2001). *Adaptive optics tutorial at CTIO*. <http://www.ctio.noao.edu/~atokovinin/tutorial/intro.html>. (Accessed: 2015-06-26)

Tokovinin, A. (2004). Seeing improvement with ground-layer adaptive optics. *Publications of the Astronomical Society of the Pacific*, 116(824), 941–951.

Vidal, F., Gendron, E., & Rousset, G. (2010). Tomography approach for multi-object adaptive optics. *JOSA A*, 27(11), A253–A264.

Vidal, F., Gendron, E., Rousset, G., Morris, T., Basden, A., Myers, R., et al. (2014). Analysis of on-sky moao performance of canary using natural guide stars. *Astronomy & Astrophysics*, 569, A16.

Voitsekhovich, V. (1995). Outer scale of turbulence: comparison of different models. *JOSA A*, 12(6), 1346-1353.