

Universidad de Oviedo
Departamento de Explotación y Prospección de
Minas
Programa de Doctorado en Minería, Obra Civil,
Medio Ambiente y Dirección de Proyectos

**Investigación del problema inverso de
reconstrucción tomográfica en óptica
adaptativa para astronomía a través
de técnicas de minería de datos e
inteligencia artificial**

Tesis Doctoral
Marlon Gómez Victoria

Diciembre 2014

Director

Prof. Dr. Francisco Javier de Cos Juez

Prof. Dr. Fernando Sánchez Lasheras

Jurados

Prof. Dra. María Nieves Roqueñi Gutierrez

Prof. Dr. Jose Luis Calvo Rolle

Prof. Dr. Juan Albino Mendez Perez

Fecha

09.12.2014



el congreso de la sociedad internacional de óptica y fotónica SPIE (Tomographic reconstructor for multi-object adaptive optics using artificial neural networks²). En estos trabajos se ha mostrado que CARMEN es capaz de competir con otros reconstructores tales como (L&A) y (LS). De esta manera, los resultados obtenidos han abierto un nuevo horizonte en el proceso de conceptualización, desarrollo e implementación de ANN en sistemas AO multi objeto los cuales funcionan en lazo abierto llamados MOAO. Sistemas que constituyen a día de hoy un objeto de estudio para la implementación en la nueva generación de telescopios extremadamente largos denominados E-ELT.

RESUMEN (en Inglés)

In modern image reconstruction of astronomical objects, researchers develop the same concepts used for centuries, although the ideas have changed to adapt to the new age. Adaptive optics (AO) is a technique which corrects telescope disturbances for image reconstruction. Tomographic reconstruction is a mathematical method developed to measure the volume of atmospheric turbulence using artificial or natural guide stars (GS). Thanks to technological advances and the huge theoretical foundations available, researchers are able to find optimal solution for these and other problems in image reconstruction.

The opening introduction describes the historical and theoretical development of the research tools involve. These tools were developed to compensate for atmospheric turbulence related issue in order to correct the incidences in image reconstruction. The next step is to incorporate the criteria and tools used to solve technical adaptive optics problems in telescopes. To finished with the components together as an AO system. After that, this research is focus on artificial neural networks, were defined other ideas related to biological processes of information processing in artificial intelligence which are: fuzzy logic and genetic algorithms.

Next, there is a description of the design, use and conception of the reconstructor. The reconstructor is called CARMEN (Complex Atmospheric Reconstructor based in Machine lEarNing) and it was designed using R programming language helped by the AMORE library. Both the laboratory and field test provided optimistic results in the implemetation of CARMEN 's multi object adaptive optics (MOAO) system. The CARMEN results have been published in the journal Optics Express and in Monthly Notices of the Royal Astronomical Society MNRAS , a journal of Oxford University. It has also been presented at the conference for the International Society for Optics and Photonics SPIE . For more information, the reader can find further information online by typing 'Open loop tomography with artificial neural networks on CANARY: on-sky



results' or 'Tomographic reconstructor for multi-object adaptive optics using artificial neural networks'³. The papers have shown CARMEN could compete with standard reconstructors, learn and apply (L&A) and least square (LS). Those results allow the researchers to participate in the process of conceptualization, development and implementation of the ANN in multi-object AO systems which does open loop work. MOAO systems are being implemented in the new generation telescopes called Extremely Large Telescopes E-ELT.

SR. PRESIDENTE DE LA COMISIÓN ACADÉMICA DEL PROGRAMA DE DOCTORADO EN MINERÍA, OBRA CIVIL, MEDIO AMBIENTE Y DIRECCIÓN DE PROYECTOS.

3 <http://arxiv.org/abs/1405.6862>;
<http://spie.org/Publications/Proceedings/Paper/10.1117/12.925355>.

Índice general

1. Planteamiento del problema	9
2. Introducción	11
2.1. Consideraciones generales acerca de la Reconstrucción Tomográfica	11
2.2. Modelo de turbulencia atmosférica	12
2.2.1. Características	12
2.2.2. Incidencia de la turbulencia en la reconstrucción de imágenes	15
2.3. Sistemas de óptica adaptativa	18
2.3.1. Óptica Adaptativa	19
2.3.2. Espejos deformables	21
2.3.3. Sensores de frente de onda.	22
2.3.4. Sistema de procesamiento o reconstrucción	26
2.4. Sistemas de Óptica Adaptativa de Campo Amplio	28
2.4.1. Sistemas clásicos	28
2.4.2. Óptica adaptativa multi-objeto <i>MOAO</i>	30
3. Inteligencia Artificial	35
3.1. Generalidades	35
3.1.1. Lógica Difusa	36
3.1.2. Algoritmos genéticos	38
3.2. Redes neuronales Artificiales	44
3.2.1. Introducción histórica	44
3.2.2. Fundamentos	46
4. Las Redes Neuronales en nuestra investigación	53
4.1. Método de perceptrón Multicapa	53
4.2. Procesado de la información y uso de <i>CARMEN</i>	57
4.2.1. Diseño	59
4.2.2. Entrenamiento	61
5. Resultados y discusión	67
5.1. Análisis en laboratorio	67
5.1.1. Resultados de la simulación sin ruido	69
5.1.2. Resultados de la simulación con ruido	71
5.2. Validación en banco óptico	74
5.3. Validación en el cielo	76

6. Conclusiones	79
6.1. Primera fase	79
6.2. Segunda y tercera fase	81
7. Líneas futuras de investigación	83
Bibliografía	85

Agradecimientos

Se quiere dar gracias al gobierno de España, Ministerio de Economía y Competitividad por permitir llevar a cabo esta investigación mediante el Plan Nacional AYA2010-18513 de investigación predoctoral para la formación de personal investigador, al cual esta asociado el autor de esta tesis. A la Universidad Oviedo, Escuela Técnica Superior de Ingenieros de Minas Oviedo (ETSIMO) por la disposición de sus gentes para poder llevar a buen término esta investigación. Se desea mencionar explícitamente a Francisco Javier de Cos y Fernando Sánchez Lasheras por el constante soporte técnico y humano, así como también a cada uno de los miembros del tribunal de la comisión de seguimiento anual, quienes durante dos años mostraron especial interés en la investigación, por último se extiende el agradecimiento a la dirección de la ETSIMO y al personal del departamento de Explotación y Prospección de Minas por la constante ayuda técnica proporcionada. Por otro lado, se debe mencionar al profesor Cristian Dani Guzman, del departamento de Ingeniería Eléctrica, Pontificia Universidad Católica de Chile y al profesor James Osborn, del departamento de Física, Centro de Instrumentación Avanzada, Universidad de Durham, por la ayuda proporcionada en la conceptualización y comprensión de los modelos para la reconstrucción de imágenes en óptica adaptativa.

A todos ellos muchas gracias.

NOTA PARA EL LECTOR:

Se desea realizar una aclaración para ayudar a la correcta lectura de la tesis. Aunque los nombres de las herramientas aquí utilizadas están dados en Español, de mutuo acuerdo con los directores de la tesis, se ha llegado a la conclusión que lo mejor es poner todas las siglas en Inglés. Lo anterior se debe a que si se desea realizar una búsqueda en internet de los conceptos usados aquí, incluso los resultados en Español utilizan las siglas en Inglés.

Resumen

En toda investigación en reconstrucción de imágenes de objetos astronómicos mediante telescopios ubicados en la tierra, se desarrollan los mismos conceptos usados desde ya hace varios siglos. Aunque, las ideas han sido modificadas con el transcurso del tiempo adaptandolas a las herramientas existentes en nuestra época. La óptica adaptativa (*AO*) es una técnica que logra corregir las perturbaciones más importantes que sufren los telescopios al realizar la reconstrucción de imágenes astronómicas. La reconstrucción tomográfica en *AO* es un proceso matemático que permite estimar el volumen de turbulencia en la atmósfera a partir de la medición de porciones de ella mediante el uso de estrellas guías naturales o artificiales. Gracias al rápido avance tecnológico y al amplio constructo teórico al alcance de todos los investigadores interesados en esta rama del conocimiento, se han proporcionado soluciones óptimas a estos y otros problemas que se presentan en la reconstrucción de imágenes.

En este sentido, la idea fue realizar una introducción teórica e histórica de conceptos relevantes en esta investigación. Inicialmente se desarrollaron conceptos relacionados con la turbulencia atmosférica para luego trabajar la incidencia de la turbulencia en la reconstrucción de imágenes. Entre las características de la turbulencia se tienen el modelo usado en el espacio de probabilidad, el índice de refracción del modelo de turbulencia [13], algunas constantes relevantes en dicho estudio como son el parámetro de Fried [14] y la frecuencia de Greenwood [15] y por último se habló del anisoplanatismo [14] y de la razón de Strehl [16]. Una vez presentados estos conceptos, se tienen las herramientas necesarias para comprender cómo es afectado un frente de onda al entrar en la atmósfera. Seguido a este proceso, se procede a presentar los criterios y herramientas usados para solventar los problemas técnicos inducidos por el uso de la óptica adaptativa en un telescopio. Ya se tienen los elementos que componen un sistema de óptica adaptativa, como último punto y a nivel general se presentaron algunos de los sistemas de óptica adaptativa de campo amplio a disposición de los investigadores y de forma particular se describieron los sistemas *AO* multi objeto los cuales funcionan en lazo abierto llamados *MOAO*, los cuales son la base de este estudio.

Aunque esta investigación esta enfocada en el uso de redes neuronales artificiales (*ANN*) para la reconstrucción de imágenes astronómicas, desde el punto de vista general es necesario introducir ideas relacionadas con los métodos biológicos de procesamiento de la información en inteligencia artificial pues las *ANN* pertenecen a este grupo. Entre los métodos más relevantes que se tienen es este grupo se encuentran lógica difusa y los algoritmos genéticos. Para lógica difusa se definió qué es una función de pertenencia, [47] su uso y diferencia con las funciones usadas en lógica clásica. Y en los algoritmos genéticos se presentaron los conceptos y criterios necesarios para exponer el Teorema Fundamental de los algoritmos genéticos [52]. Sin embargo, al ser el objeto fundamental en este estudio las redes neuronales se desa-

rrollaron con más detalle. Se partió desde la visión histórica y de ahí en adelante se quiso presentar una idea general de los tipos de algoritmos de aprendizaje, clases de entradas en una red neuronal, clasificación según el tipo de conexión o el número de capas y las estructuras que se encuentran en las redes neuronales. También se definió el concepto de neurona y las características en el diseño de ésta.

Una vez realizado de forma general un mapa conceptual de las herramientas disponibles, espacios a considerar, métodos relevantes y en general de los criterios fundamentales al abordar una investigación con las características específicas requeridas, lo siguiente fue acotar los elementos usados en dicha investigación. Por tanto y dado que existe una amplia bibliografía relativa a las *ANN* [61, 62, 63, 64, 65, 66] entre otras, se realizó una reducción considerable en esta investigación definiendo las redes perceptrón multicapa con algoritmo de retropropagación. Esto ocurrió de esta manera pues el reconstructor con el cual se trabajó en esta investigación está enmarcado en este grupo. Lo siguiente fue describir el diseño, uso y concepción del reconstructor usado en esta investigación sintetizado así: el reconstructor se llamó *CARMEN* (Reconstructor Atmosférico Complejo basado en Aprendizaje de Máquina), fue diseñado mediante el lenguaje de programación *R* haciendo uso de la librería *AMORE*. Su topología se obtuvo después de una cantidad considerable de pruebas de laboratorio realizadas en paralelo en la Universidad de Oviedo, Universidad de Durham y en la Pontificia Universidad Católica de Chile. Se trabajó con datos representativos de simulaciones ya conocidas que contenían las entradas y salidas para la red.

Tanto las pruebas de laboratorio como las de campo proporcionaron resultados optimistas en cuanto a la implementación de *CARMEN* en un sistema de óptica adaptativa multi-objeto (*MOAO*). Los resultados obtenidos por *CARMEN* han sido objeto de publicaciones en revistas de divulgación tales como *Optics Express* y la *Revista Notas Mensuales de la Real Sociedad Astronómica (MNRAS)*, Universidad de Oxford y cuyo artículo lleva como título (Open loop tomography with artificial neural networks on CANARY: on-sky results¹). Así como también se han presentado en el congreso de la Sociedad Internacional de Óptica y Fotónica *SPIE* (Tomographic reconstructor for multi-object adaptive optics using artificial neural networks²). En estos trabajos se ha mostrado que *CARMEN* es capaz de competir con otros reconstructores tales como: aprender y aplicar (*L&A*) y mínimos cuadrados (*LS*). De esta manera los resultados obtenidos han abierto un nuevo horizonte en el proceso de conceptualización, desarrollo e implementación de *ANN* en sistemas *MOAO*. Sistemas que constituyen a día de hoy un objeto de estudio para la implementación en la nueva generación de telescopios extremadamente largos denominados *E – ELT*.

¹<http://mnras.oxfordjournals.org/content/441/3/2508.short> ó <http://arxiv.org/abs/1405.6862>

²<http://spie.org/Publications/Proceedings/Paper/10.1117/12.925355>

Para finalizar, el proceso descrito en el parrafo anterior se dividió en tres fases. Inicialmente se dejó claro cuáles fueron los resultados obtenidos en cada una de ellas, así como también se realizó una descripción de los procesos usados y las ideas nuevas introducidas en la investigación. Como se mencionó anteriormente los resultados fueron esperanzadores pues se comparó *CARMEN* con otros reconstructores y se llegó a la conclusión que aunque se trabaje en condiciones desfavorables este reconstructor responde de forma apropiada a la exigencias del sistema. No obstante, debido a las limitaciones de máquina para el entrenamiento, actualmente se trabaja en la adaptación de *CARMEN* a la programación en la unidad de procesamiento gráfico *GPU*. Se cree que esto permitirá optimizar los resultados obtenidos por *CARMEN*, pues se podrá trabajar en el entrenamiento con muestras de volúmenes más grandes. Esta última afirmación se hace pues *CARMEN* puede ser mejorada incluso mediante el uso de programación clásica. Gracias al alto grado de paralelismo de la programación en *GPU*, este método se convierte en un buen objeto de análisis para posterior aplicación en el campo.

Summary

In modern image reconstruction of astronomical objects, researchers develop the same concepts used for centuries, although the ideas have changed to adapt to the new age. Adaptive optics (*AO*) is a technique which corrects telescope disturbances for image reconstruction. Tomographic reconstruction is a mathematical method developed to measure the volume of atmospheric turbulence using artificial or natural guide stars (*GS*). Thanks to technological advances and the huge theoretical foundations available, researchers are able to find optimal solution for these and other problems in image reconstruction.

The opening introduction describes the historical and theoretical development of the research tools involve. These tools were developed to compensate for atmospheric turbulence related issue in order to correct the incidences in image reconstruction. The different tools used to eliminate atmospheric turbulence are: space of probability, refractive index [13], some relevant constants like Fried parameter [14] and Greenwood frequency [15] and concepts such as anisoplanatism [14] and Strehl ratio [16]. Once all the information from these tools is known, the effect of the form of the wavefront coming into the atmosphere can be understood. The next step is to incorporate the criteria and tools used to solve technical adaptive optics problems in telescopes. With the components together as an *AO* system, the subject of this study will be on the multi-object adaptive optics (*MOAO*) which works in an open loop.

Thanks to the innovative approach of this research *using artificial neural networks in astronomical image reconstruction*, other ideas to consider are related to biological processes of information processing in artificial intelligence which are: fuzzy logic and genetic algorithms. For fuzzy logic it was defined as the membership function of a fuzzy set [47] which is a generalization of the indicator function in classical sets. For genetic algorithms the schema theorem was proven [52]. However with artificial neural networks since they are the subject of this research they were developed in detail, beginning with a historical vision of artificial neural networks follow by the development of a general idea of the learning algorithm types, input classes in a artificial neural network, classification according to type of connection or number of layers and the artificial neural networks structures including a definition of the idea of neurons and their characteristics.

As artificial neural networks have a long bibliography [61, 62, 63, 64, 65, 66], etc., the definition of the artificial neural network used in this research was reduce to a feedforward artificial neural network such as multilayer perceptron with a backpropagation algorithm. Next, there is a description of the design, use and conception of the reconstructor. The reconstructor is called *CARMEN* (Complex Atmospheric

Reconstructor based in Machine lEarNing) and it was designed using *R* programming language helped by the *AMORE* library. The configuration of the net was obtained after hours of laboratory testing performed in conjunction with the University of Oviedo, the University of Durham and the Pontifical Catholic University of Chile.

Both the laboratory and field test provided optimistic results in the implementation of *CARMEN*'s multi object adaptive optics (*MOAO*) system. The *CARMEN* results have been published in the journal Optics Express and in Monthly Notices of the Royal Astronomical Society *MNRAS*, a journal of Oxford University. It has also been presented at the conference for the International Society for Optics and Photonics *SPIE*. For more information, the reader can find further information online by typing 'Open loop tomography with artificial neural networks on CANARY: on-sky results' or 'Tomographic reconstructor for multi-object adaptive optics using artificial neural networks'³. The papers have shown *CARMEN* could compete with standard reconstructors, learn and apply (*L&A*) and least square (*LS*). Those results allow the researchers to participate in the process of conceptualization, development and implementation of the *ANN* in multi-object *AO* systems which does open loop work. *MOAO* systems are being implemented in the new generation telescopes called Extremely Large Telescopes *E – ELT*.

The testing had several phases which were divided into three parts. In the first phase, results were gathered from trials. The second phase used laboratory results and the third phase gained results from field tests. All three phases contributed to the implementation of new ideas in this research. As previously mentioned, *CARMEN* gave promising results when it was compared with other reconstructors and one of the conclusions was that, *CARMEN* performed well even in unfavorable conditions. Due to the limitations of the computer during trials, researchers searched for new approaches like *GPU* programming language; It is thought this method could optimize *CARMEN*'s results. Even using *CPU*, *CARMEN* was an improvement over others reconstructors. The high parallelism of the *GPU* programming language makes it an optimal candidate for this type of work.

³<http://arxiv.org/abs/1405.6862>; <http://spie.org/Publications/Proceedings/Paper/10.1117/12.925355>

1 Planteamiento del problema

En este trabajo de investigación se propone analizar y validar una aproximación novedosa al problema de la reconstrucción tomográfica en óptica adaptativa, proceso que permite estimar el volumen de la turbulencia atmosférica haciendo pequeñas mediciones de ella. La hipótesis de partida es que la turbulencia atmosférica es un fenómeno tridimensional complejo y no lineal. En esencia, se plantea utilizar técnicas de minería de datos e inteligencia artificial para “aprender” las características de la atmósfera utilizando varios sensores de frente de onda repartidos en el campo de observación. Para llevar a cabo el citado análisis se simularán distintas mediciones de turbulencia tal como hubiera sido medida si se hubiesen empleado simultáneamente varias estrellas de guiado laser en el borde del campo de observación, así como con algunas estrellas guías naturales dentro del mismo. Tras el adecuado preproceso del conjunto de mediciones, las muestras obtenidas serán usadas para seleccionar y entrenar el algoritmo de inteligencia artificial capaz de reconstruir la turbulencia simulada. Se plantea validar los modelos propuestos comparando los resultados obtenidos en las simulaciones con información real del perfil de la turbulencia. Para llevar a cabo la medición real de la turbulencia se empleará un instrumento facilitado por el Centro de Instrumentación Avanzada (*CfAi*, Sedgefield, Reino Unido) y desarrollado por uno de los investigadores del equipo, Richard Wilson [1]. La principal ventaja de esta investigación es que no plantea limitar el proceso a un modelo concreto y estático tal como lo hace el estado de arte en esta disciplina, si no que, una vez escogido el algoritmo inteligente, el modelo se construirá dinámicamente a partir de las mediciones efectuadas en tiempo real. Una vez confirmada la factibilidad de las técnicas y dado que los requerimientos computacionales de esta propuesta son bastante altos, el plan de trabajo pretende centrarse en desarrollar una plataforma de entrenamiento de redes neuronales capaz de operar en pseudo tiempo real en el telescopio utilizando lógica programable. Existen otros grupos en el mundo investigando reconstrucción tomográfica basada en sistemas de óptica adaptativa multi-conjugada, en particular el sistema de laboratorio “Homer” en Francia [2], el instrumento *MAD* (Demostrador Adaptativo para un sistema Multi-conjugado) del Observatorio Europeo Austral [3] y el instrumento *GEMS/GSAOI* (cámara de Imagen de Óptica Adaptativa de Gemini Sur) del Observatorio Gemini [4], lo cual demuestra que el objetivo propuesto es de gran interés para la comunidad científica orientada a óptica adaptativa.

Dentro de los sistemas basados en óptica adaptativa multi-objeto, los errores introducidos en el proceso de reconstrucción tomográfica juegan un papel fundamental en

la resolución final del telescopio, tanto es así que en determinados casos dicho error es responsable de más del 75 % de las desviaciones totales obtenidas. La revisión del estado del arte en la temática citada pone de manifiesto que hasta la fecha sólo se han empleado modelos de reconstrucción tomográfica lineales. Por tanto, como medida para reducir los errores en los que se incurre durante el proceso de reconstrucción, este proyecto propone una alternativa que incluye la no linealidades del problema utilizando el estado del arte en herramientas no-lineales de modelacion. En síntesis, para evaluar la efectividad de la aplicación de técnicas de minería de datos e inteligencia artificial en la reconstrucción de fenómenos turbulentos complejos y realistas, en esta investigación se pretende seguir el esquema que se expone a continuación:

- Seleccionar las técnicas de modelado más eficientes en función de los datos registrados.
- Evitar las restricciones impuestas por los modelos lineales actuales.
- Evaluar las posibilidades de implementación en el campo de la óptica adaptativa de las mismas.
- Entrenar modelos de reconstrucción específicos para observatorios concretos.
- Evaluar su bondad en términos cuantitativos de mejora con respecto a técnicas existentes.

Los objetivos anteriores estan descritos en esta tesis así: En el Capítulo 4 se muestra la técnica de inteligencia artificial a usar, así como el diseño y concepción de la misma. En este capítulo también se realiza una descripción del proceso de entrenamiento de la red, las características y configuración del banco óptico, y los instrumentos usados en las pruebas de campo. En el Capítulo 5 se desarrollan los resultados obtenidos por el reconstructor en todo el proceso de entrenamiento y validación tanto en el banco de prueabs como en el cielo. Además se cuantifica la bondad del reconstructor al compararlo con reconstructores como *L&A* y *LS*. En cuanto a los capítulos restantes estan diseñados para facilitar la comprensión de los conceptos subyacentes a esta investigación y para mostrar cual seria el siguiente paso a dar en esta investigación.

2 Introducción

2.1. Consideraciones generales acerca de la Reconstrucción Tomográfica

Actualmente en toda investigación en reconstrucción de imágenes de objetos astronómicos mediante telescopios ubicados en la tierra, es necesario desarrollar los mismos conceptos usados desde ya hace varios siglos. Aunque las ideas han sido modificadas con el transcurso del tiempo adaptándolas a las herramientas existentes en nuestra época, dichos sistemas conservan la estructura desarrollada desde la antigüedad. En este sentido podemos decir que todos los grandes telescopios en el mundo cuentan o contarán con sistemas de óptica adaptativa [5]. La óptica adaptativa (*AO*) es una técnica que logra corregir las perturbaciones más importantes que sufren los telescopios al realizar la reconstrucción de imágenes astronómicas [6]. Gracias a estos sistemas es posible obtener imágenes más nítidas y de mejor resolución espacial permitiendo resoluciones cercanas al límite de difracción de estos telescopios en el infrarrojo cercano (desde 1 hasta $2,5 \mu m$). Se espera que en el futuro también se logren resultados similares dentro del espectro visible. Una de las razones para que la *AO* sea de vital importancia en el desarrollo de la siguiente generación de telescopios [7]¹ se debe a que en dichos sistemas los errores introducidos en el proceso de reconstrucción tomográfica son relevantes en la resolución final del telescopio. Para esta clase de telescopios se han tomado como referencia varias técnicas en las que se usa *AO* como son: la Óptica Adaptativa Multi-Conjugada (*MCAO*) [6], la Óptica Adaptativa de capa baja (*GLAO*) [11] y la Óptica Adaptativa Multi-Objeto (*MOAO*) [5]. Sin embargo este trabajo se enmarca dentro de sistemas en los cuales se implementa *MOAO*.

La reconstrucción tomográfica en *AO* es un proceso matemático que permite estimar el volumen de turbulencia en la atmósfera a partir de la medición de porciones de esta [8]. En algunas posiciones del campo del telescopio la estimación se hace utilizando la luz de estrellas guía naturales ó artificiales. El proceso de reconstrucción es similar a la técnica utilizada en otros ámbitos, de manera particular en radiología [9, 10] donde se reconstruyen estructuras tridimensionales a partir de diversas imágenes bidimensionales tomadas a distintos ángulos y alturas del paciente. Si bien, su uso en *AO* es fundamental debido a que ha proporcionado resultados satisfactorios

¹Denominados *Extremely Large Telescopes (ELT)*.

también trae consigo problemas inherentes a las limitaciones posicionales y espaciales [11]. El problema en esta área en particular radica en la carencia de un número ilimitado de puntos de referencia [12]. Además los diferentes ángulos que forman los puntos de referencia sólo difieren unos cuantos minutos de arco entre sí.

Uno de los problemas más relevantes que se presentan en la reconstrucción tomográfica está relacionado con la turbulencia atmosférica [8], pues dicho fenómeno es aleatorio y surge como consecuencia del calentamiento de las moléculas de gas que conforman la atmósfera. Las moléculas son calentadas a diferentes temperaturas provocando un estado de excitación entre ellas que al interactuar producen volutas de diferentes tamaños a distintas alturas. Cuando la luz entra a nuestra atmósfera se ve afectada por las diferentes capas formadas por la interacción entre moléculas ocasionando distorsión en la luz captada por el telescopio. A ese fenómeno se lo conoce como anisoplanatismo angular [11] y es la principal limitación en todas las investigaciones en reconstrucción de imágenes astronómicas pues al ser la turbulencia atmosférica [1] un fenómeno aleatorio provoca una constante variación entre el ángulo del punto de referencia y el objeto astronómico a observar.

Gracias al rápido avance tecnológico y al amplio constructo teórico al alcance de todos los investigadores interesados en esta rama del conocimiento, se han proporcionado soluciones óptimas a estos y otros problemas que se presentan en la reconstrucción de imágenes. Partiendo de ésta afirmación, uno de los objetivos es participar en el proceso de conceptualización y desarrollo de reconstructores para sistemas *AO* haciendo uso de la inteligencia artificial (*AI*). Por último se quiere hacer saber que los resultados obtenidos han proporcionado un punto de partida en la investigación, tanto es así que la primera versión del reconstructor ya está siendo implementada en un sistema *AO*.

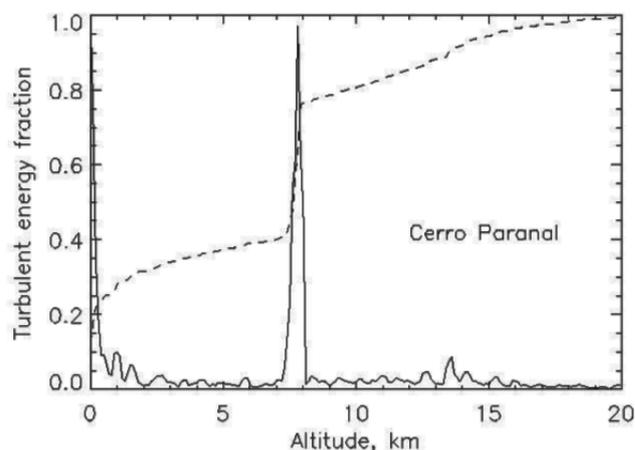
2.2. Modelo de turbulencia atmosférica

2.2.1. Características

Se sabe que este fenómeno es aleatorio y surge como consecuencia del calentamiento de las moléculas de gas que conforman la atmósfera. Los gases a diferentes temperaturas interactúan y producen volutas de diverso tamaño llamadas “*eddy*” con propiedades únicas. Por ejemplo temperatura, altura e índice de refracción entre otros. El índice de refracción de un gas respecto a la luz es función de su temperatura por lo que la turbulencia es causada por las pequeñas variaciones en la temperatura sufridas en la atmósfera de manera aleatoria producto de la interacción de las moléculas de gas. La turbulencia en observatorios astronómicos ha sido

extensamente investigada en las últimas décadas [1], fruto de estos estudios se han desarrollado modelos para analizar su comportamiento. El modelo de Kolmogorov [13] es el más utilizado para predecir el cambio en el índice de refracción de la atmósfera el cual se sabe que ocurre principalmente en capas delgadas a determinadas altitudes sobre el telescopio. Como se puede observar en la Figura 2.2.1 se muestra la energía cinética acumulada por los gases responsables de la turbulencia atmosférica medida en un observatorio astronómico². Tal y como se aprecia en la gráfica dos tercios de la energía cinética acumulada en la turbulencia pertenece a dos capas, la primera a 0 Km y la segunda a 8 Km. El tercio de energía restante se encuentra distribuida en todo el resto de la atmósfera.

Figura 2.2.1: perfil de turbulencia en Cerro Paranal (Chile). La línea sólida representa la energía de la turbulencia (en unidades relativas) y la línea punteada representa la energía acumulada. Cortesía: A. Tokovinin (Cerro Tololo) [1].



En el modelo de Kolmogorov [13], la turbulencia se trata como una sucesión de fenómenos aleatorios que ocurren en el plano a distintas alturas [1, 8]. No obstante y dado que cada eddy posee características diferentes, es necesario saber cuando un fenómeno es considerado turbulento. Para ello se usa el siguiente parámetro denominado número de Reynolds [13]

$$R_e = \frac{V_0 L_0}{\nu_0} \quad (2.2.1)$$

donde V_0 es la velocidad característica del flujo, L_0 la escala característica de la eddy a considerar y ν_0 constante llamada viscosidad cinemática del aire. De acuerdo con el modelo de Kolmogorov [13] existe un intervalo inercial donde cada L tiene una

²Concretamente en el *Observatorio Europeo Austral en Cerro Paranal* (Chile).

velocidad V garantizando una escala superior L_0 donde generalmente se da lugar a la turbulencia y una escala inferior l_0 en la cual se disipa. La existencia del intervalo inercial en el cual se pueda aplicar el modelo de Kolmogorov [13] se debe a lo siguiente:

Un espacio de probabilidad es un conjunto $\mathcal{P} = (\Omega, \mathcal{A}, P)$ que tiene tres componentes. Primero un espacio muestral finito Ω con los posibles resultados. Segundo, el conjunto de todos los sucesos aleatorios incluyendo los no elementales denotado por \mathcal{A} . Por último una medida de probabilidad P la cual asigna una probabilidad a todo suceso y cumple los axiomas de Kolmogorov.

1. Los conjuntos \emptyset (Vacío) y Ω (Omega) están en \mathcal{A} , y $P(\emptyset) = 0$, $P(\Omega) = 1$ y $0 \leq P(A) \leq 1$ para todo $A \in \mathcal{A}$.
2. Para un grupo de conjuntos disjuntos $A_1, \dots, A_n \in \mathcal{A}$, $n \in \mathbb{N}$ se tiene que, $P(A_1 \cup \dots \cup A_n) = P(A_1) + \dots + P(A_n)$

este espacio de probabilidad está dotado de otras características y propiedades pero a efectos de caracterizar el espacio usado, es suficiente con las anteriormente mencionadas.

Para describir la turbulencia atmosférica en el modelo de Kolmogorov se tiene la función estructura $D_n(\Delta x)$ definida como la diferencia entre dos valores de una variable aleatoria que depende del índice de refracción.

$$D_n(\Delta x) = \langle [n(x) - n(x')]^2 \rangle \quad (2.2.2)$$

$n(x)$ es el índice de refracción y $x' = x + r$ siendo r la distancia entre x y x' . Así la función de estructura como variación en el índice de refracción de la turbulencia en el espacio³ para el modelo de Kolmogorov [13] es:

$$D_n(\Delta x) = C_n^2(y) \Delta x^{\frac{2}{3}} \text{ entre } l_o < \Delta x < L_0 \quad (2.2.3)$$

donde $C_n(y)$ es el índice de refracción vertical del parámetro estructura que a su vez esta en función de la altura y .

³En este espacio se debe considerar al viento como el principal elemento que incide en la variación del índice de refracción de la turbulencia.

Aunque con la función estructura que depende del índice de refracción se describe la propagación de la turbulencia de la forma más adecuada, es necesario saber que la aproximación se obtiene del hecho de considerar inicialmente la función estructura como función de la velocidad. Esto se hace así pues la atmósfera posee ciertas cualidades como son: el ser localmente homogénea⁴, e isotrópica⁵ y considerarse la turbulencia como un flujo aleatorio. Dichas propiedades permiten tratar la turbulencia como un fenómeno local en un espacio.

2.2.2. Incidencia de la turbulencia en la reconstrucción de imágenes

En cuanto el frente de onda del objeto astronómico entra en la atmósfera es afectado por diferentes procesos físicos los cuales son relevantes en la construcción de telescopios [7]. Entre los problemas producidos está la deformación que sufre el frente de onda dada la fluctuación en el índice de refracción. También se debe tener en cuenta que cada capa turbulenta posee una velocidad de propagación que varía con el tiempo afectando el frente de onda en mayor medida donde la velocidad sea mayor. Para resolver estos problemas se han planteado modelos acordes con el espacio de probabilidad en el cual se trabaja, entre estos se encuentra la Longitud de Coherencia de Fried [14] y la Frecuencia de Greenwood [15]. Para describirlos se parte de la ecuación de onda.

$$\Psi = Ae^{i\beta} \tag{2.2.4}$$

donde A es la amplitud y β la fase. Al entrar en la atmósfera la luz se ve afectada debido a que su velocidad cambia de forma inversa respecto al índice de refracción y como éste varía en cada eddy, el frente de onda irá a diferentes velocidades produciendo retrasos en dicho frente de onda y ocasionando que éste ya no sea plano. En términos matemáticos, la deformación que sufre la onda al entrar en la atmósfera se describe mediante la siguiente ecuación.

$$\psi = \Delta n(y)dy \tag{2.2.5}$$

$n(y)$ representa el índice de refracción a lo largo de la línea de visión en función de la altura. De acuerdo con el comportamiento de la ecuación de onda, la distorsión producida en la atmósfera se presenta como variaciones de fase y fluctuaciones en la intensidad. En la reconstrucción de imágenes, dicha distorsión se caracteriza por

⁴La velocidad depende del vector desplazamiento entre dos posiciones

⁵La velocidad depende de la magnitud del vector desplazamiento

la variación de fase debido a que ésta incide en el proceso. Al expresar la variación de fase del frente de onda en términos de índice de refracción se tiene

$$\psi_1 = \frac{2\pi}{\lambda} \Delta n(y) dy \quad (2.2.6)$$

donde λ representa la longitud de onda. De lo que se puede concluir que existe una dependencia entre la fase y el inverso de la longitud de la onda. Por tanto, a mayor longitud de onda, menor distorsión de la imagen. A efectos de una reconstrucción óptima este es el caso a tener en cuenta. Al considerar en este espacio la fase de dos puntos muy proximos separados por una distancia Δx y teniendo presente que la atmósfera esta dividida en capas gracias al modelo de Kolmogorov [13] se tiene que la función estructura de fase en función de la estructura del índice es

$$\sigma_{\psi_1} = 2,91k^2 \sec\xi \Delta C_n^2(y) \Delta x^{\frac{5}{3}} dy \quad (2.2.7)$$

siendo k una constante y ξ el ángulo de observación cenital. Haciendo uso de aproximaciones la Ecuación 2.2.7 se reduce a

$$\sigma_{\psi_1} = 6,88 \left(\frac{\Delta x}{r_0} \right)^{\frac{5}{3}} \quad (2.2.8)$$

donde

$$r_0 = \left[0,423 \left(\frac{2\pi}{\lambda} \right)^2 (\sec\xi) \Delta C_n^2(y) dy \right]^{-\frac{3}{5}} \quad (2.2.9)$$

r_0 es el parámetro de Fried [14]. El parámetro de Fried puede ser interpretado de varias maneras, para este trabajo la más importante es la de permitir caracterizar el efecto de la turbulencia en una longitud de onda específica ayudando a optimizar la resolución de la observación R independientemente de la apertura del telescopio

$$R = \frac{1,22\lambda}{r_0} \quad (2.2.10)$$

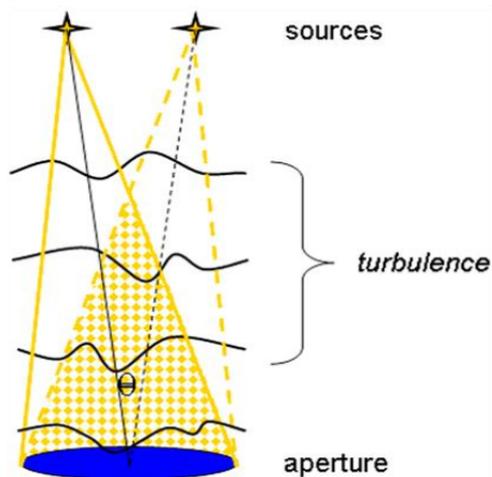
lo anterior se conoce como visión astronómica. Para solucionar el problema de las diferentes velocidades en las capas de turbulencia, solo es necesario considerar la media de las velocidades mediante series de Taylor pues una aproximacion hecha

con este modelo permite sintetizar de forma adecuada el cambio constante y rapido que se produce en la velocidades. En este sentido la Frecuencia de Greenwood [15] se define como sigue

$$\sigma_v = \tau_0 = \left[0,102 \left(\frac{2\pi}{\lambda} \right)^2 (\sec\xi) \Delta_o^\infty C_n^2(y) |v(y)|^{\frac{5}{3}} dy \right]^{-\frac{3}{5}} \quad (2.2.11)$$

lo cual representa el tiempo mínimo con que cambia la fase de la turbulencia. De esta manera, si se quiere que la función de estructura (Ecuación 2.2.8) dependa del tiempo, solo es necesario reemplazar Δx por $\bar{v}\tau_0$ siendo \bar{v} la velocidad promedio de las eddys. No obstante dichos problemas también imponen limitaciones sistemáticas. Por ejemplo, el parámetro de Fried [14] limita el número de actuadores que debe tener el espejo deformable y la Frecuencia de Greenwood [15] modifica la frecuencia con la que los actuadores modifican su estado. Esta última debe ser mayor o del orden de la frecuencia de variación de la turbulencia. Pero el principal problema en toda investigación en óptica adaptativa es el anisoplanatismo [14] Figura 2.2.2. Este fenómeno ocurre debido a las limitaciones en el campo angular del sistema. Se produce pues al ser la turbulencia un fenómeno aleatorio, la luz de objetos diferentes en el campo del telescopio, recorre distintos caminos ópticos, turbulencias diferentes y por tanto sufre aberraciones diferentes causando un rápido deterioro en la corrección del ángulo θ_0 entre el objeto astronómico y las estrellas guía.

Figura 2.2.2: *Anisoplanatismo angular.*



El anisoplanatismo depende directamente de la turbulencia atmosférica y se describe de forma analítica como sigue:

$$\sigma_{\theta} = \theta_0 = \left[2,914 \left(\frac{2\pi}{\lambda} \right)^2 (\sec\xi)^{\frac{8}{3}} \Delta_o^{\infty} C_n^2(y) y^{\frac{5}{3}} dy \right]^{-\frac{3}{5}} \quad (2.2.12)$$

usando aproximaciones, el ángulo θ queda reducido a lo siguiente

$$\theta_0 = 0,314 \cos\xi \left(\frac{r_0}{\bar{y}} \right)^{\frac{5}{3}} \quad (2.2.13)$$

donde \bar{y} representa el promedio de altura donde ocurre la turbulencia y se calcula mediante la razón de Strehl [16], herramienta que permite medir la calidad de la reconstrucción de fase. Existen varias definiciones de este concepto pero para catalogar el ángulo isoplanático respecto a la degradación de la corrección en la fase, la razón de Strehl [16] debe ser $\frac{1}{\epsilon}$ con ϵ constante. Esto simplifica el problema, puesto que el ángulo a considerar es aquel donde la fase puede ser aproximada por una función constante. En síntesis, un frente de onda está supeditado a varianzas de altura, posición y velocidad en un espacio de probabilidad, el cual depende principalmente de las características atmosféricas del lugar. Así, la varianza total al cuadrado en un frente de onda es

$$\sigma_{total}^2 = \sigma_{tipilt}^2 + \sigma_{WFS}^2 + \sigma_{retardo}^2 + \sigma_{ángulo}^2 + \sigma_{fit}^2 \quad (2.2.14)$$

donde σ_{tipilt} depende de ψ_1 . $\sigma_{retardo}$ depende de la frecuencia de Greenwood respecto al instante cuando se realiza la corrección. $\sigma_{ángulo}$ depende del ángulo isoplanático [14] respecto a un ángulo cualquiera. σ_{fit} depende del número de actuadores, el parámetro de Fried [14] y la matriz de influencia A . Y σ_{WFS} es el error asociado al sensor de frente de onda y el cual se denotará por (WFE) a lo largo de este trabajo. En la siguiente sección se desarrollarán los conceptos de actuadores, matriz de influencia y sensor de frente de onda para complementar la información acerca de la varianza ó error total en un sistema de óptica adaptativa.

2.3. Sistemas de óptica adaptativa

Existen dos clases fundamentales de telescopios desarrollados a lo largo de décadas denominados telescopios reflectantes [17] y telescopios refractantes [18]. Lo telescopios refractantes aparecieron inicialmente en 1550 gracias a Leonard Digges pero

las primeras descripciones reales datan de años posteriores, alrededor de 1600 [18]. Este telescopio capta los rayos de luz paralelos procedentes de un objeto a gran distancia y haciendo uso de lentes convergentes logra que los rayos converjan en un punto del plano focal. Son comunes en telescopios solares pero el uso de ellos trae consigo problemas relacionados con las lentes. El primero se debe a la elaboración de las lentes, si se desean usar en telescopios de gran tamaño, es difícil fabricar lentes acordes, útiles y ligeras para el objetivo, también existen inconvenientes con la calidad de la imagen debido a pequeñas burbujas de aire alojadas en algunas de las cavidades de la lente principal. Por último, para ciertas longitudes de onda los cristales de la lente son opacos perdiendo sensibilidad en partes del espectro de luz. La mayoría de los problemas que conlleva el uso de telescopios refractantes se solucionan haciendo uso de telescopios reflectantes pues éste utiliza espejos en lugar de lentes. Sus ventajas se encuentran en características tales como: solo es necesario garantizar la perfección de la superficie mientras que al trabajar con lentes este debe estar libre de imperfecciones, la aberración cromática es inexistente en un espejo y al trabajar con lentes solo pueden estar sujetos por los extremos mientras que los espejos pueden estar sujetos por toda su superficie, esta idea es fundamental si se trabajan con lentes de gran tamaño ya que la gravedad puede ocasionar distorsión en la imagen. Aunque esta clase de telescopios es más eficiente al compararse con los telescopios refractantes, no está libre de problemas ya que posee otro tipo de aberraciones ópticas que se detallan a continuación. Para finalizar se quiere aclarar que cuando se haga referencia a un telescopio se hará referencia a un telescopio de tipo reflectante [17].

2.3.1. Óptica Adaptativa

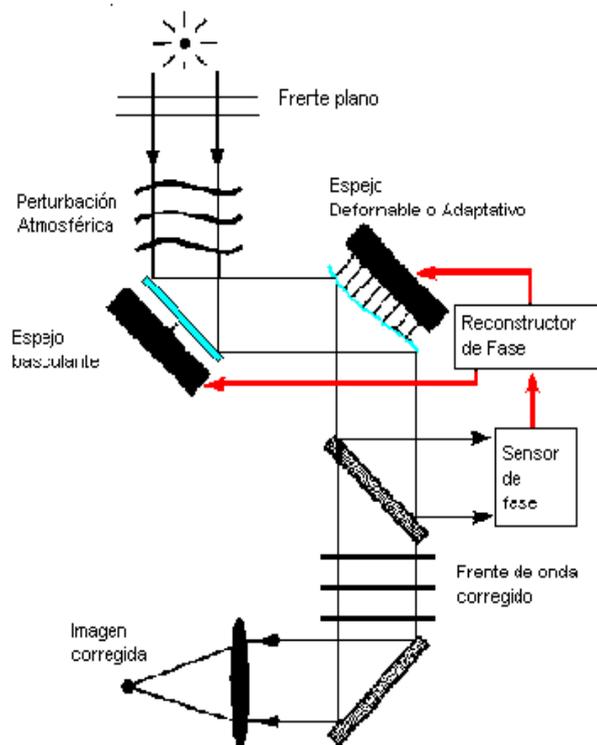
Los sistemas de óptica adaptativa en un telescopio, permiten mejorar la calidad de una imagen astronómica proporcionando las herramientas suficientes para reducir en tiempo real las perturbaciones que presenta un frente de onda al atravesar las diferentes capas atmosféricas. Los conceptos en los cuales se han estructurado dichos sistemas tienen sus fundamentos en estudios realizados desde ya hace décadas [12, 13, 15], sin embargo hasta hace poco tiempo y gracias a los avances tecnológicos que se han producido en la últimas dos décadas se han logrado obtener resultados plausibles [8, 19]. Los resultados obtenidos en los experimentos realizados han permitido que esta rama se desarrolle de forma exponencial. Gracias al interés de la comunidad científica, existen dos tipos diferentes de sistemas *AO* uno de ellos es el clásico o sistema *AO* de lazo cerrado [6, 11] y el otro es el sistema *AO* de lazo abierto [20].

Aunque existen diferencias entre los sistemas *AO* los elementos principales son: espejos deformables, sensor de frente de onda y sistema de procesamiento. La luz captada por el telescopio es reflejada por dos espejos, el primero corrige aberraciones de primer orden “*tilt mirror*” y el segundo corrige aberraciones de orden superior. Tras

cada uno de estos espejos se instala un sensor de frente de onda “*phase sensor*” cuya finalidad es medir la turbulencia residual⁶. Por último el sistema de procesamiento denominado reconstructor de fase “*phase reconstructor*” recibe la medición del sensor de frente de onda y envía las señales pertinentes para controlar ambos espejos. En la Figura 2.3.1 se puede observar la disposición de los componentes en un sistema clásico que ha sido desarrollado extensivamente en las dos últimas décadas [5, 19]. Este es un sistema en bucle cerrado por lo que el objetivo es que el sensor mida un frente de onda plano el cual es capturado por un instrumento astronómico en el plano focal “*focal plane image*”.

Cada uno de los elementos que constituyen un sistema *AO* ha sido diseñado con el fin de solventar los problemas físicos inherentes a la construcción de un telescopio moderno, es por esto que los componentes de estos sistemas están hechos de diferentes materiales y formas, haciéndolos acreedores de diferentes características.

Figura 2.3.1: *Componentes principales de un sistema de óptica adaptativa.*



⁶Mide el desplazamiento lateral de la turbulencia

2.3.2. Espejos deformables

En astronomía se usan dos tipos de espejos deformables (*DM*), los Piezo-eléctricos [20] y los Bimórficos [21]. Los espejos “piezo-eléctricos” están constituidos por una capa espejada muy fina (Deformable Membrane), unida a un conjunto de actuadores piezo-eléctricos (Actuators), cuya función es mover la capa espejada al aplicársele un voltaje determinado a cada uno de ellos. Para controlar este tipo de espejos es necesario conocer las funciones de influencia. Por definición, Las funciones de influencia están caracterizadas por la forma que adopta la superficie al aplicársele un cierto voltaje a los actuadores. En un momento en particular la forma del espejo puede obtenerse como combinación lineal de las funciones de influencia puesto que tendríamos una base vectorial tal y como se muestra en la Figura 2.3.2. Los espejos “Bimórficos” Figura 2.3.3 constan de dos capas piezo-eléctricas unidas y polarizadas de forma opuesta. Una de las capas debe estar espejada y ambas deben estar conectadas a tierra. En medio de las capas piezo-eléctricas se tiene una matriz de electrodos. Si se le aplica voltaje a un elemento de dicha matriz una de las capas se contrae y la otra se expande provocando que el espejo se curve donde se le ha aplicado el voltaje. Como la curvatura depende del voltaje al cual fue sometido, se les suele llamar espejos de curvatura. En el Cuadro 2.1 se muestran las características físicas más importantes de cada uno de los espejos deformables mencionados anteriormente.

Como se puede observar en el Cuadro 2.1 los espejos se han desarrollado teniendo en cuenta las funciones determinadas por su geometría, voltaje, frecuencia, entre otras propiedades. Sin embargo la clasificación más importante se debe a los principios físicos usados en los diferentes actuadores. Inicialmente, los primeros materiales usados en los actuadores son: la cerámica piezo-eléctrica la cual se denomina (*PZT*) por sus componentes, plomo, zirconio y titanio. Y la de material electrostrictivo o no conductor denominada (*PMN*) por sus materiales plomo y niobato de magnesio. No obstante, estos actuadores sufren de histéresis⁷ para el caso es una especie de memoria mecánica que hace depender al actuador de la posición anterior impidiendo un cálculo apropiado de la posición final de estos componentes. Dado que cada grupo de actuadores está desarrollado con diferentes materiales, se pueden encontrar diferencias entre ellos, la principal se da por el punto de Curie. El punto de Curie es la temperatura donde un material piezo-eléctrico pierde sus características y su polarización espontánea. En *PZT* el punto de Curie se encuentra en 200 y 400 grados centígrados y para *PMN* es de 0 grados centígrados.

Existen Otro tipo de actuadores en el mercado como son los Electrostáticos [22]⁸ que están hechos de una pieza de silicio en la cual se construyen los actuadores. También se encuentran los Magnéticos los cuales son actuadores de bobina de voz⁹

⁷Tendencia del material a conservar una de sus propiedades después de aplicado un voltaje.

⁸<http://www.bksv.es/Products/transducers/acoustic/accessories/electrostatic-actuators>

⁹<http://es.electrix.allitwares.com/Productlist-Los-actuadores-de-bobina-de-voz-578-1.html>

compuesta por un imán y un solenoide. Esta tecnología es relativamente nueva y tiene un plus, bajo coste y gran linealidad. Una característica fundamental de estos actuadores es la de estar libres de histéresis.

Figura 2.3.2: *Espejo deformable piezo-eléctrico. Geometría rectangular*

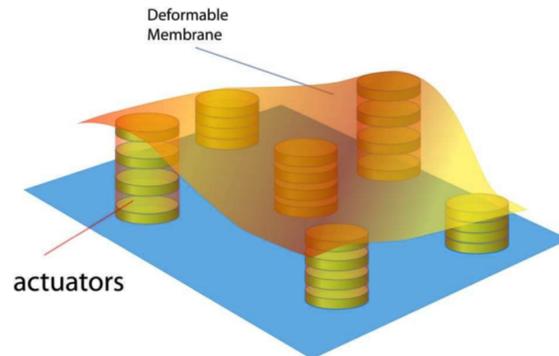
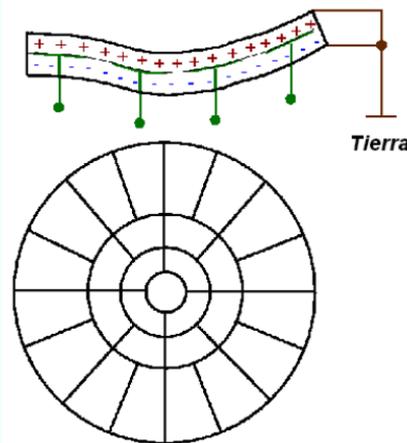


Figura 2.3.3: Espejo deformable bimórfico.



2.3.3. Sensores de frente de onda.

Un sistema *AO* incluye un componente denominado sensor de frente de onda (*WFS*) en cada uno de los espejos deformables el cual corrige las aberraciones residuales del frente de onda al pasar por los *DM*'s. Existen varios tipos de *WFS* entre los que se encuentran los sensores Shack-Hartmann (*SH*) [23], los sensores curvatura [24] y los sensores pirámide [25]. Su función es transformar las aberraciones en intensidad de la luz y luego en señales eléctricas. Esta función se realiza haciendo uso de técnicas que están supeditadas a aplicaciones de procesos indirectos. El más común es el Shack-Hartmann [23] que consta de una matriz de lentes (lenslet array) situada en el mismo plano donde salen los haces de luz del espejo primario del telescopio.

Cuadro 2.1: *Características físicas de los espejos deformables.*

	Piezo-eléctricos	Bimórficos
Número de actuadores	100-1500	13-85
Tamaño del espejo	70-200mm	30-200mm
Geometría de actuadores	Rectangular ó hexagonal	Radial
Voltaje	Cientos de voltios	Cientos de voltios
Frecuencia de resonancia	Varios Kilohercios	600Hz
coste	Muy alto	Moderado

Los haces de luz son recibidos por el sensor formando diferentes ángulos respecto de la vertical Figura 2.3.4. La matriz muestrea el frente de onda que sale de la pupila del telescopio midiendo mediante desplazamientos laterales las pendientes locales del frente de onda. Se sabe que la pendiente media sobre una apertura se describe así [23]

$$\alpha'_x = \frac{d\phi(x,y)}{dx} = \frac{\Delta x}{f} \quad (2.3.1)$$

y como cada sistema óptico necesita un factor de ampliación necesario para la reconstrucción entonces

$$\alpha_x = M\alpha'_x \quad (2.3.2)$$

donde M es el factor de ampliación. Por tanto de la Ecuación 2.3.1 y la Ecuación 2.3.2 se tiene

$$\begin{aligned} \alpha_x &= \frac{M}{f} \Delta x = \frac{\lambda}{2\pi A} \Delta_{sa} \frac{\partial\phi(x,y)}{\partial x} dx dy \\ \alpha_y &= \frac{M}{f} \Delta y = \frac{\lambda}{2\pi A} \Delta_{sa} \frac{\partial\phi(x,y)}{\partial y} dx dy \end{aligned} \quad (2.3.3)$$

en la Ecuación 2.3.3, f y A son la focal y el área respectivamente de cada subapertura y λ es la longitud de onda. La imagen es pasada por un sensor de plano focal (CCD) [26] tal y como se muestra en la Figura 2.3.4, y debe tener las siguientes características: ser rápido para obtener las muestras de la variación temporal de la atmósfera, tener resolución espacial suficiente para analizar la amplitud de las aberraciones con las cuales se trabaja y por último debe poder trabajar en condiciones de poca luminosidad. El CCD [26] mide las aberraciones mediante la localización

de la salida en cada subapertura representada con la pareja $(\Delta x, \Delta y)$, este proceso es llamado “centroiding”. Existen variedad de algoritmos pero se suele usar el algoritmo de centro de gravedad [27] pues proporciona una estructura robusta de cara a corregir los problemas del ruido y la velocidad de cálculo. Por tanto es común que la posición de cada centroide sea calculada usando la Ecuación 2.3.4

$$(\hat{x}, \hat{y}) = \left(\frac{\sum_{i,j} x_{i,j} I_{i,j}}{\sum_{i,j} I_{i,j}}, \frac{\sum_{i,j} y_{i,j} I_{i,j}}{\sum_{i,j} I_{i,j}} \right) \quad (2.3.4)$$

la pareja $(x_{i,j}, y_{i,j})$ representa las coordenadas (x, y) de cada pixel en la matriz de lentes e $I_{i,j}$ es la intensidad de dicho pixel en cada subapertura.

Como principal problema en un *WFS* se encuentra el ruido [12]. Después de muchos estudios se ha logrado caracterizar los diferentes ruidos que se producen en la reconstrucción de imágenes como ruido fotónico y ruido electrónico [28]. Cada uno de ellos caracteriza un proceso físico ocurrido en la reconstrucción. El primero se produce por la señal y el ruido de fondo del cielo. Es asociado con la probabilidad que tiene cada fotón de llegar a cada subapertura. El segundo se haya en el ruido de fondo del cielo que incluye la señal de fondo del cielo y se produce por la corriente de oscuridad y el *CCD* [26]. Es asociado con los efectos del detector y es relevante en el proceso de reconstrucción del frente de onda. Un esquema sobre la catalogación del ruido se encuentra en el Cuadro 2.2. Los efectos de estos tipos de ruido se sintetizan en las siguientes ecuaciones [29]¹⁰

$$\sigma_{\phi, N_{ph}}^2 = \frac{\pi^2 N_T^2}{\ln 4 N_{ph} N_{samp}^2} (\text{radianes}) \quad (2.3.5)$$

$$\sigma_{\phi, N_r}^2 = \frac{\pi^2 N_r^2 N_s^4}{3 N_{ph}^2 N_{samp}^2} (\text{radianes})$$

donde $\sigma_{\phi, N_{ph}}$ ruido fotónico, σ_{ϕ, N_r} ruido electrónico, N_{ph} número promedio de fotones en cada subapertura, N_r ruido de salida, N_T ancho completo de la mitad del máximo de la mancha, N_s tamaño de la subapertura en pixeles, $N_{samp} = \frac{\lambda}{d p}$ relación del tamaño del pixel angular p con la mitad de la anchura de la mancha de difracción limitada $\frac{\lambda}{d}$, siendo λ la longitud de onda y d el diámetro de la subapertura. Si queremos ver el error del sensor en el plano focal en términos de pixels tenemos la siguiente relación

$$\sigma_x = \frac{N_{samp}}{2\pi} \sigma_{\phi} \quad (2.3.6)$$

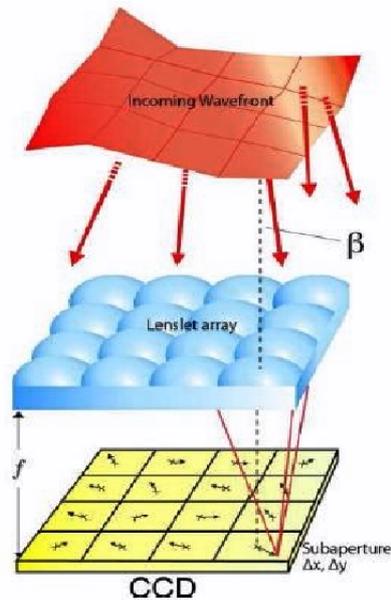
¹⁰Las ecuaciones en el sensor *SH* estan simplificadas para más detalles ver referencia.

donde σ_ϕ constituye el ruido total del sistema. Por tanto

$$\sigma_x = \sqrt{\frac{N_T^2}{8 \ln 2 N_{ph}} + \frac{N_r^2 N_s^4}{12 N_{ph}^2}} \quad (2.3.7)$$

la Ecuación 2.3.7 proporciona mediante la experimentación la configuración apropiada de parámetros como tamaño del punto y rango dinámico en el sensor de frente de onda *SH*.

Figura 2.3.4: *Sensor de frente de onda (WFS) Shack-Hartmann.*



Cuadro 2.2: *Cómo se produce el ruido en la reconstrucción de imágenes.*

$$\begin{aligned} \text{Ruido Fotónico} & \left\{ \begin{array}{l} \text{Señal} \\ \text{Ruido de fondo del cielo} \end{array} \right. \left\{ \begin{array}{l} \text{Señal de fondo del cielo} \\ \text{Ruido Electrónico} \end{array} \right. \\ \\ \text{Ruido Electrónico} & \left\{ \begin{array}{l} \text{corriente de oscuridad} \\ \text{Ruido del CCD} \end{array} \right. \end{aligned}$$

2.3.4. Sistema de procesamiento o reconstrucción

El problema a solucionar es el de reconstruir un frente de onda plano que sufre distorsiones al entrar a la atmósfera. Para ello se han desarrollado métodos para hacer uso de la información proporcionada por el *WFS* llamados reconstrucción zonal y modal. En la reconstrucción zonal se representa la fase del frente de onda como una matriz independiente de subaperturas ó zonas repartidas por la apertura del telescopio, generalmente las entradas de la matriz suelen coincidir con las posiciones de los actuadores del *DM*. En este método, el frente de onda está directamente relacionado con la cantidad de subaperturas del telescopio, por tanto cuantos más subaperturas tenga, mejor representado estará el frente de onda. El método de reconstrucción modal es considerado como el que mejor resultados proporciona si se compara con el zonal [30] y en síntesis la idea es representar el frente de onda como una familia de polinomios ortogonales a la pupila del telescopio. Como este método es el que se ha de usar por el sistema en la reconstrucción, se describirá a continuación con más de detalle. Sea la Ecuación 2.3.8 un frente de onda plano

$$\phi(x, y) = \sum_{i=1}^n a_i F_i(x, y) \quad (2.3.8)$$

donde $\{F_i(x, y)\}$ son una familia de polinomios ortonormales a la pupila del telescopio y a_i componente i -ésima del vector de coeficientes a estimar. Una matriz de lentes formada por n subaperturas en un sensor *SH* proporciona $2n$ medidas puesto que mide pendientes en los ejes x e y obtenidas al derivar la fase del frente de onda Ecuación 2.3.8. Al derivar el frente de onda se obtiene un sistema $2n \times m$ siendo m el número de modos usados en la reconstrucción.

$$\begin{aligned} \frac{\partial \phi}{\partial x} \Big|_1 &= \frac{a_1 \partial F_1(x, y)}{\partial x} \Big|_1 + \dots + \frac{a_m \partial F_m(x, y)}{\partial x} \Big|_1 \\ \frac{\partial \phi}{\partial y} \Big|_1 &= \frac{a_1 \partial F_1(x, y)}{\partial y} \Big|_1 + \dots + \frac{a_m \partial F_m(x, y)}{\partial y} \Big|_1 \\ &\vdots \\ \frac{\partial \phi}{\partial x} \Big|_n &= \frac{a_1 \partial F_1(x, y)}{\partial x} \Big|_n + \dots + \frac{a_m \partial F_m(x, y)}{\partial x} \Big|_n \\ \frac{\partial \phi}{\partial y} \Big|_n &= \frac{a_1 \partial F_1(x, y)}{\partial y} \Big|_n + \dots + \frac{a_m \partial F_m(x, y)}{\partial y} \Big|_n \end{aligned} \quad (2.3.9)$$

La Ecuación 2.3.9 se puede agrupar en un vector de pendientes

$$\vec{S} = (\alpha_{1,x}, \alpha_{1,y}, \dots, \alpha_{n,x}, \alpha_{n,y})^T \quad (2.3.10)$$

matricialmente se tiene

$$\vec{S} = A \cdot \vec{a} \quad (2.3.11)$$

donde A es la matriz de interacción y \vec{a} es el vector de coeficientes a calcular.

$$A = \begin{pmatrix} \left. \frac{\partial F_1(x,y)}{\partial x} \right|_1 & \dots & \left. \frac{\partial F_m(x,y)}{\partial x} \right|_1 \\ \left. \frac{\partial F_1(x,y)}{\partial y} \right|_1 & \dots & \left. \frac{\partial F_m(x,y)}{\partial y} \right|_1 \\ \vdots & & \\ \left. \frac{\partial F_1(x,y)}{\partial x} \right|_n & \dots & \left. \frac{\partial F_m(x,y)}{\partial x} \right|_n \\ \left. \frac{\partial F_1(x,y)}{\partial y} \right|_n & \dots & \left. \frac{\partial F_m(x,y)}{\partial y} \right|_n \end{pmatrix} \quad (2.3.12)$$

Por tanto, al reconstruir el frente onda captado por el telescopio se pretende encontrar los parámetros de \vec{a} que mejor estiman cada ϕ_i , $i = 1 \dots n$ y para que el sistema tenga solución es necesario que $m \leq 2n$. La solución se obtiene usando mínimos cuadrados ó usando SVD ¹¹. La familia de polinomios a usar para caracterizar la turbulencia atmosférica es la de polinomios de Zernike pues estos son los que mejor describen la turbulencia sobre una superficie circular [31]. El Cuadro 2.3 muestra los primeros modos de Zernikes en coordenadas polares. El mayor de ellos tiene orden seis y se escogió de esta manera pues en esta investigación las 27 salidas de la ANN representan los coeficientes de los Zernikes de hasta ese orden. De seis en adelante el sistema despreció los valores al ser considerados de orden superior. Cada polinomio es el producto de una función radial y una función polar definidos mediante las expresiones analíticas [31]

$$Z_l^k(\rho, \theta) = N_l^k(\theta) R_l^{|k|}(\rho) \quad (2.3.13)$$

$$R_l^{|k|}(\rho) = \sum_{s=0}^{\lfloor \frac{l-|k|}{2} \rfloor} \frac{(-1)^s (l-s)!}{s! \left(\frac{l+k}{2} - s\right)! \left(\frac{l-k}{2} - s\right)!} \rho^{l-2s} \quad (2.3.14)$$

$$N_l^k(\theta) = \begin{cases} \sqrt{2(l+1)} \cos(k\theta); & k \neq 0; & t = 2z \\ \sqrt{2(l+1)} \sen(k\theta); & k \neq 0; & t = 2z + 1 \\ \sqrt{2(l+1)}; & k = 0 \end{cases} \quad (2.3.15)$$

¹¹ Descomposición en valores singulares

donde t describe el número de modos para un orden y la cantidad total de estos viene dada por $t = \frac{l(l+2)+k}{2}$, $k \leq l$ y $l - k$ es par.

Al usar estos métodos se pretende resolver un problema común en matrices el cual se conoce como singularidad se considera que una matriz es singular si su determinante es igual a cero. Este es el problema principal de este trabajo de investigación debido a que al usar redes neuronales en el proceso de reconstrucción se puede disminuir considerablemente errores gracias a su alta tolerancia a fallos.

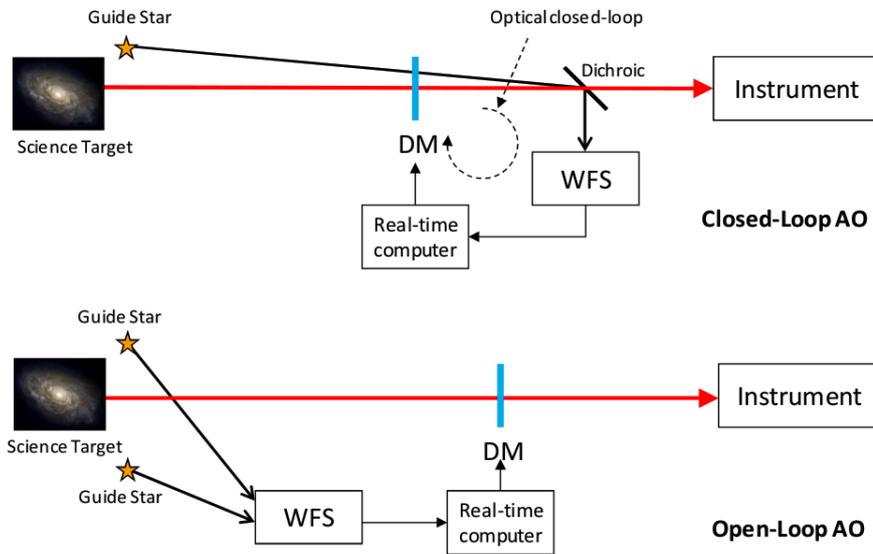
2.4. Sistemas de Óptica Adaptativa de Campo Amplio

2.4.1. Sistemas clásicos

En términos generales ya está claro cuales son los componentes de un sistema de óptica adaptativa, cómo están distribuidos y las funciones que cada uno de ellos desempeña. También se han presentado algunas de las ideas matemáticas que sustentan la teoría *AO* y todo ello pensado en la reconstrucción del frente de onda captado por el telescopio. Diferentes investigaciones acerca del tema han dejado como resultado la definición y uso de dos tipos de sistemas *AO* que suelen llamarse sistema óptica adaptativa en lazo abierto [20] y lazo cerrado [6, 11] un esquema de cada uno de estos sistemas se encuentra en la Figura 2.4.1. En esta investigación un sistema clásico o un sistema *AO* en lazo cerrado aunque útil no es conveniente debido a que uno de los objetivos es generar un sistema de redes neuronales que realice el proceso de reconstrucción tomográfica sin retroalimentación óptica. Para ello existen los sistemas *MOAO* [32] los cuales se tratarán más adelante. En ellos la reconstrucción realizada del objeto astronómico a partir de la medida del *WFS* se aplica directamente al *DM*. A diferencia de un sistema en lazo cerrado [6] el cual se realimenta continuamente deformando el *DM* iterativamente hasta que se observa que el frente de onda es lo más parecido al plano.

La limitación que presentan los sistemas clásicos de óptica adaptativa a la hora de permitir correcciones en campos visuales amplios ha sido ampliamente investigada [2] siendo los primeros sistemas concebidos para reducir dicha limitación los denominados multi-conjugados “*Multi-Conjugated Adaptive Optics*” *MCAO* Figura 2.4.2. Aunque existen dos tipos de sistemas *MCAO* [11] los cuales son: *MCAO* orientado hacia las estrellas guías (*GS*) denominado (*SO – MCAO*) y *MCAO* orientado hacia las capas de turbulencia (*LO – MCAO*), la única diferencia considerable entre ambos se debe a la configuración del *WFS*. Inicialmente la detección del frente de onda

Figura 2.4.1: *Funcionamiento sistemas lazo cerrado (arriba) y lazo abierto (abajo) en óptica adaptativa.*



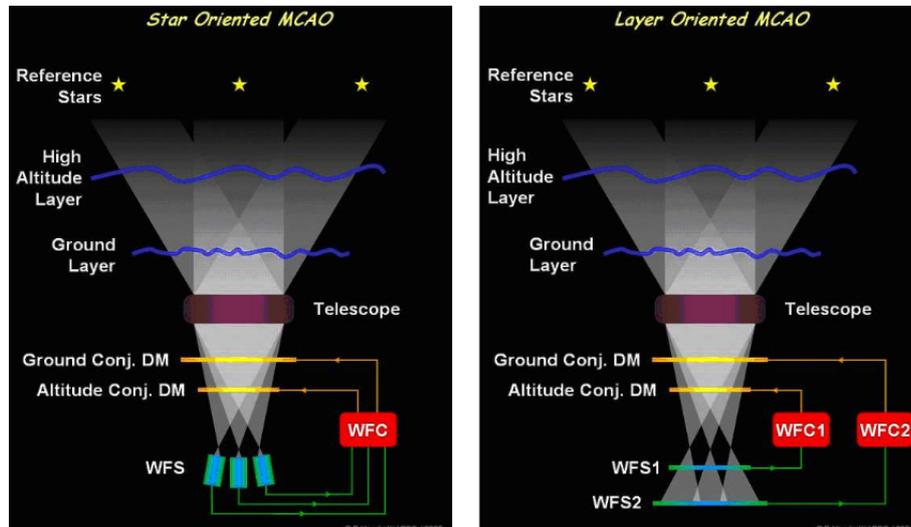
orientada hacia las *GS* surge como una generalización de los métodos existentes, los cuales usaban una sola *GS* y un sensor de frente de onda para medir la perturbación de la fase en la pupila del telescopio. *SO – MCAO* Figura 2.4.2 izquierda, permite medir una gran porción del volumen de turbulencia gracias a los diferentes *WFS* cada uno conjugado a una *GS*. Esta medida nos proporciona la fase en la pupila del telescopio a lo largo de diferentes campos visuales. Una vez se tienen las medidas del sensor, ésta es pasada por un control del frente de onda (*WFC*) que calcula los comandos para deformar los espejos. Por otro lado, en el *LO – MCAO* (Figura 2.4.2 derecha), se conjugan los *WFS* con las capas de turbulencia a considerar. Generalmente se tienen en cuenta dos capas, una de ellas a nivel del suelo y la otra con una capa alta¹², de la misma forma los *DM* están conjugados a las mismas altitudes que los *WFS*. También es necesario establecer un *WFC* para cada una de las parejas *WFS* y *DM* conjugadas a la misma altitud para que se encargue de compensar la turbulencia en cada capa.

Como segunda alternativa surgió la óptica adaptativa de capa terrestre más conocida como *óptica adaptativa de capa baja (GLAO)*. Técnica que fue inventada por Rigaut [6] como una manera sencilla de cubrir un campo amplio con un *DM* compensando solo la turbulencia situada en la zona inmediatamente superior al telescopio Figura 2.4.3. Las dos modalidades, *MCAO* y *GLAO* se basan en medir la turbulencia en distintos puntos del campo del telescopio usando estrellas guía (*naturales y artificiales*) como referencia y mediante sensores de frente de onda reducir al mínimo los problemas causados por el anisoplanatismo Figura 2.2.2. Sin embar-

¹²De acuerdo con Figura 2.2.1 esta capa se encuentra entre 6 y 8 *Km* de distancia.

go, a pesar de que los sistemas basados en *MCAO* y *GLAO* son de gran interés y existen diversos grupos de investigación que trabajan en la actualidad en la mejora de su eficiencia, ésta investigación se sustenta en óptica adaptativa multi-objetivo (*MOAO*).

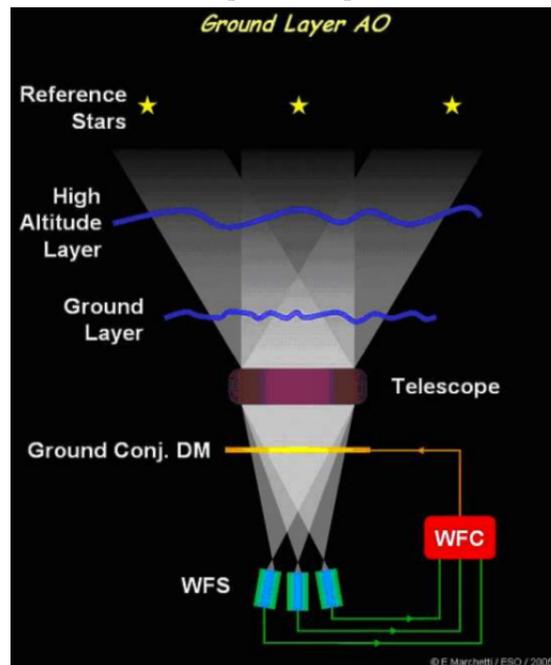
Figura 2.4.2: Sistema de óptica adaptativa de campo amplio Multi Conjugate AO. Cortesía: ESO.



2.4.2. Óptica adaptativa multi-objeto *MOAO*

La función de un sistema *MOAO* [32] es similar al de *MCAO* [11] y *GLAO* [6] pero posee una característica fundamental que lo diferencia de los otros, ésta es la de operar en lazo abierto. Lo anterior constituye un punto de inflexión en la reconstrucción de imágenes astronómicas, ya que hasta el momento los sistemas utilizados en este campo de investigación operaban en lazo cerrado. Cada *MOAO* Figura 2.4.4 contiene los siguientes subsistemas:

- Un grupo de estrellas guía láser (*LGS*). La cantidad de *LGS* del sistema depende del coeficiente de Strehl [16]. Esta medida sirve para valorar la calidad óptica del telescopio permitiendo evaluar la calidad de la imagen y el desempeño del sistema. Los *LGS* son lanzados desde el telescopio a un marco de referencia el cual contiene el objeto astronómico a observar. De esta manera se reduce el campo visual de forma considerable, para así concentrarse en zonas donde existan objetos astronómicos de interés.
- Sensores de frente de onda (*WFS*) para medir los datos de entrada sin corregir. La cantidad de *WFS* con este fin es igual a la cantidad de *LGS* utilizados. También son necesarios un grupo de *WFS* para las *estrellas guías*

Figura 2.4.3: *Sistema óptica adaptativa Ground Layer AO.*

(*GS*) naturales que puedan determinar tip-tilt. La cantidad de estos *WFS* no necesariamente deben ser igual a la cantidad *WFS* utilizados para medir los datos sin corregir.

- Un pick-off opto-mechanism para observar cada objeto. Los haces de luz son transmitidos a espejos deformables los cuales trabajan en lazo abierto respecto a los *WFS*. De esta manera se logra la corrección dinámica de los datos obtenidos.
- Una unidad de campo integral (*IFU*) utilizada para diseccionar cada uno de los componentes espaciales de los haces de luz corregidos.
- Un espectrógrafo para recuperar los espectros de cada punto de *objetos observados mediante AO*. La cantidad de espectrógrafos dependen directamente del número de objetos, ya que se puede trabajar con varios espectrógrafos funcionando en paralelo.
- Un detector de infrarrojos cercanos para obtener los espectros.
- Por último, un ordenador encargado de medir la turbulencia entrante y calcular la corrección necesaria para cada *DM*.

Todos los sistemas anteriores a *MOAO* [32], es decir, los sistemas clásicos como el de la Figura 2.3.1, así como *MCAO* [11] y *GLAO* [6] operan en lazo cerrado debido a que los sensores de frente de onda miden la turbulencia residual no corregida por el espejo. Dado que *MOAO* [32] funciona en lazo abierto, este mide la turbulencia

sin corregir con el objetivo de calcular una corrección para posiciones específicas del campo visual y así espejos deformables efectúen dicha corrección en un campo estrecho (*un espejo por objeto*) subsanando el problema que genera el anisoplanatismo [14].

En las pruebas en el cielo hechas en esta investigación se implementó un sistema *MOAO* llamado *CANARY* [33] (Simulación de un Sistema *MOAO* para el telescopio de 4,2 metros *William – Hershel* en las islas Canarias, se puede ver el sistema en la Figura 2.4.5) y que fue construido por el centro de instrumentación avanzada, departamento de Física, Universidad de Durham con la colaboración de observatorio de París y el Centro de Tecnología Astronómica del Reino Unido. *CANARY* incorpora un canal *MOAO* y se desarrolló en tres fases. En la primera, se trabajó con estrellas guías naturales (*NGS*) con 3 *WFS* en lazo abierto y un *DM* de orden bajo, 7×7 subaperturas y 8×8 actuadores. En la segunda se agregó un asterismo de 4 estrellas guías laser generadas con la dispersión de un laser de baja altura de Rayleigh que se usa en conjunción con una herramienta óptica de difracción para formar el asterismo. Los *WFS* están apuntados a *LGS*. Para encontrar el asterismo final mediante pruebas en el cielo se contó con asterismos identificados de catálogos de estrellas. Y por último se procedió a trabajar con *LGS* pero los *DM* fueron configurados sobre las condiciones de un telescopio extremadamente largo (*E – ELT*). Lo anterior cambia las características del sistema, ya que ahora los *DM* conjugados a nivel del suelo¹³ trabajan en lazo cerrado y los de orden superior, 32×32 actuadores trabajan el lazo abierto¹⁴.

¹³*DM* de orden inferior

¹⁴Los resultados de esta última parte se pueden ver en [?].

Figura 2.4.4: *Multi Object AO*.

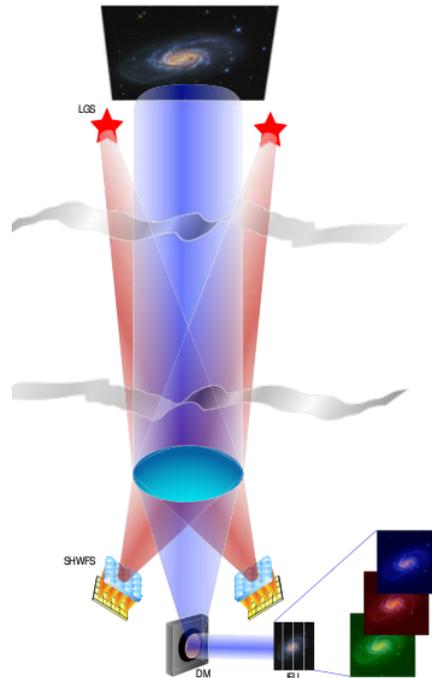
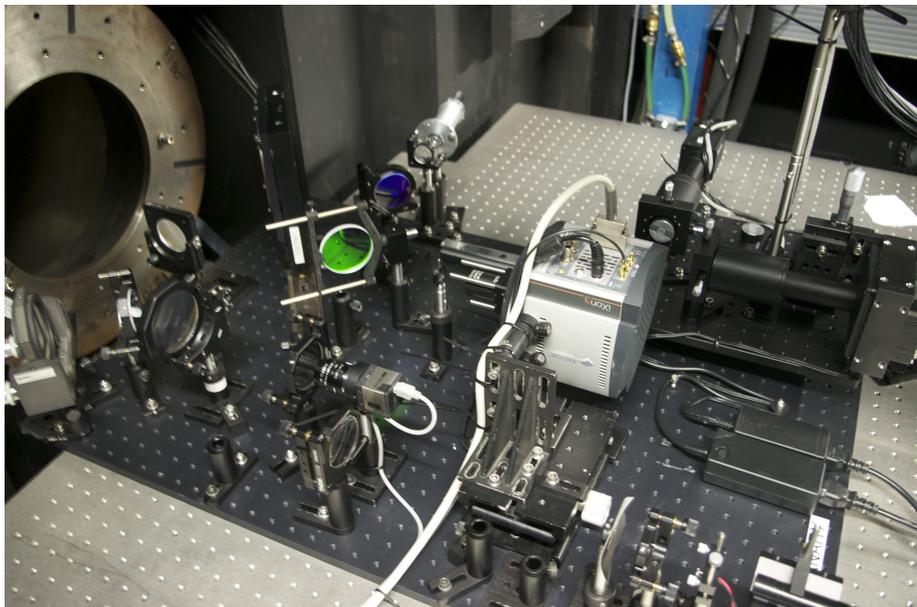


Figura 2.4.5: Experimento *CANARY*.



Cuadro 2.3: 27 primeros modos de Zernike.

índice (t)	Orden (l)	Frecuencia azimutal (k)	$Z_l^k(\rho, \theta)$
0	0	0	1
1	1	-1	$2\rho \text{sen}\theta$
2	1	1	$2\rho \text{cos}\theta$
3	2	-2	$\sqrt{6}\rho^2 \text{sen}2\theta$
4	2	0	$\sqrt{3}(2\rho^2 - 1)$
5	2	2	$\sqrt{6}\rho^2 \text{cos}2\theta$
6	3	-3	$\sqrt{8}\rho^3 \text{sen}3\theta$
7	3	-1	$\sqrt{8}(3\rho^3 - 2\rho) \text{sen}\theta$
8	3	1	$\sqrt{8}(3\rho^3 - 2\rho) \text{cos}\theta$
9	3	3	$\sqrt{8}\rho^3 \text{cos}3\theta$
10	4	-4	$\sqrt{10}\rho^4 \text{sen}4\theta$
11	4	-2	$\sqrt{10}(4\rho^4 - 3\rho^2) \text{sen}2\theta$
12	4	0	$\sqrt{5}(6\rho^4 - 6\rho^2 + 1)$
13	4	2	$\sqrt{10}(4\rho^4 - 3\rho^2) \text{cos}2\theta$
14	4	4	$\sqrt{10}\rho^4 \text{cos}4\theta$
15	5	-5	$\sqrt{12}\rho^5 \text{sen}5\theta$
16	5	-3	$\sqrt{12}(5\rho^5 - 4\rho^3) \text{sen}3\theta$
17	5	-1	$\sqrt{12}(10\rho^5 - 12\rho^3 + 3\rho) \text{sen}\theta$
18	5	1	$\sqrt{12}(10\rho^5 - 12\rho^3 + 3\rho) \text{cos}\theta$
19	5	3	$\sqrt{12}(5\rho^5 - 4\rho^3) \text{cos}3\theta$
20	5	5	$\sqrt{12}\rho^5 \text{cos}5\theta$
21	6	-6	$\sqrt{14}\rho^6 \text{sen}6\theta$
22	6	-4	$\sqrt{14}(6\rho^6 - 5\rho^4) \text{sen}4\theta$
23	6	-2	$\sqrt{14}(15\rho^6 - 20\rho^4 + 6\rho^2) \text{sen}2\theta$
24	6	0	$\sqrt{7}(20\rho^6 - 30\rho^4 + 12\rho^2 - 1)$
25	6	2	$\sqrt{14}(15\rho^6 - 20\rho^4 + 6\rho^2) \text{cos}2\theta$
26	6	4	$\sqrt{14}(6\rho^6 - 5\rho^4) \text{cos}4\theta$
27	6	6	$\sqrt{14}\rho^6 \text{cos}6\theta$

3 Inteligencia Artificial

3.1. Generalidades

La inteligencia artificial es un constructo teórico pensado para dar solución a diferentes problemas concernientes al análisis de información. Aunque el fin último es lograr que estos sistemas obtengan una mayor autonomía en el aprendizaje y toma de decisión, el objetivo todavía dista de ser cumplido. No obstante, gracias a los múltiples aportes de la comunidad científica, se puede comprobar que podría lograrse en un tiempo considerablemente corto, ya que se han desarrollado diferentes métodos y sistemas con interfaces cada vez más amigables con los investigadores. Así, la inteligencia artificial hoy día es considerada relevante en la construcción de modelos y se clasifica en dos grupos.

- **SISTEMAS BASADOS EN CONOCIMIENTO O INTELIGENCIA ARTIFICIAL CLÁSICA:** para el funcionamiento de estos sistemas se necesita definir un grupo de afirmaciones predeterminadas (Axiomas) y también un conjunto de reglas que se puedan aplicar en el modelo. Las nuevas afirmaciones del modelo son inferidas a partir de las reglas que se pueden aplicar en el modelo y los Axiomas disponibles en este. Esta teoría es llamada inteligencia artificial clásica (*GOF AI*) [34].
- **MÉTODOS BIOLÓGICOS DE PROCESADO DE LA INFORMACIÓN:** estos métodos, a diferencia de los anteriores, tienen un grado de libertad más amplio pues su misión es la de intentar reproducir actitudes inteligentes usando un esquema inductivo. Es decir, a partir de muestras concretas se desea extraer características generales de un sistema. En este grupo se encuentran herramientas tales como: la **Lógica Difusa**, los **Algoritmos genéticos** y las **Redes neuronales artificiales**.

Este capítulo se centrará en los métodos biológicos de procesamiento de la información. La evolución en este campo se produce de forma exponencial gracias a que en las investigaciones donde se hace uso de estos métodos y en las cuales no es claro qué tipo de herramienta es la apropiada para modelar una cuestión, se abordan los problemas desde diferentes perspectivas y se escoge cuál es la solución que más se adapta a las necesidades. El proceso realizado para encontrar un modelo, permite generar un bagaje conceptual rápidamente, no obstante dado que la teoría acerca

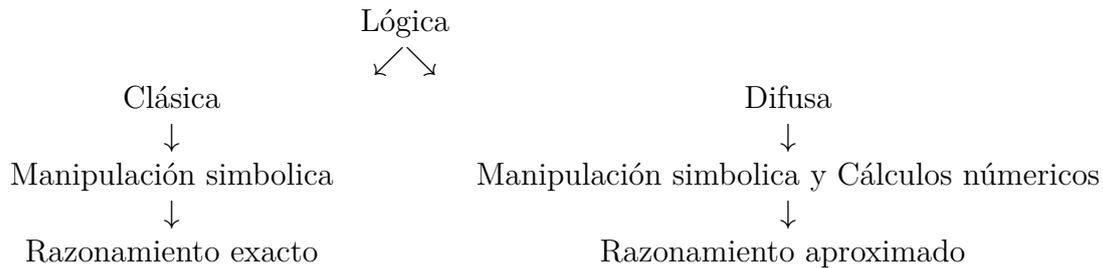
de los métodos de procesado de información se hace cada vez más extensa, si se deseara realizar una presentación detallada, se tendría que hacer una tesis dedicada solo a desarrollar sus conceptos. Es por esto que se presentará de forma escueta las ideas principales de la lógica difusa y los algoritmos genéticos para luego desarrollar las redes neuronales con un poco de más detalle. Estas últimas son la base para el estudio sobre la reconstrucción de imágenes desarrollado en este escrito.

3.1.1. Lógica Difusa

La lógica difusa [35-37] pretende acercarse al pensamiento humano dejando a un lado los postulados matemáticos de la lógica clásica dándole más importancia a la aproximación que a la precisión. Este tipo de lógica tiene sus orígenes ya desde los griegos, estos se dieron cuenta que no todas las respuestas podían obtenerse haciendo uso de la lógica clásica partiendo del criterio “nada es completamente verdadero o nada es completamente falso”. Estas ideas los llevan a darse cuenta de los diferentes grados de verdad o falsedad que pueden existir al usar el razonamiento humano. Los siguientes en detectar problemas con la lógica clásica fueron David Humme [38], George Berkeley [39] e Immanuel Kant [40] quienes mediante sus teorías pudieron observar las contradicciones que se generaban, pero fue Bertrand Russell [41] el cual popularizó esta idea ayudado por Ludwig Wittgenstein quien realizó un estudio intensivo acerca del lenguaje concluyendo que una misma palabra puede condicionar la frase a la que este subordinada [42]. La primera lógica en formalizar estos conceptos fue desarrollada en 1920 por el filósofo Jan Lukasiewicz. Para él todos los conjuntos poseían un cierto grado de pertenencia con valores oscilando entre $[0, 1]$. En 1965 Lofti Asier Zadeh relacionó los conceptos de la lógica y los conjuntos descritos por Lukasiewicz, creando así la lógica difusa [43] y en 1973 se propuso su teoría de lógica difusa [44]. Inicialmente esta teoría no fue bien recibida por los investigadores, no obstante, gracias a su estudio e implementación se pudieron desarrollar los denominados controladores difusos. El primero de este tipo de controladores fue presentado en 1975 por Assilian y E. H. Mamdani [45] y no fue hasta 1980 que F. L. Smidth & Co implementaron por primera vez un controlador basado en lógica difusa [46].

En el Cuadro 3.1 se pueden ver la diferencias principales entre lógicas. En lógica clásica un elemento pertenece o no a un conjunto. Dicha afirmación no es cierta en la teoría de conjuntos difusos pues ésta permite considerar grados de pertenencia parciales de un elemento en un conjunto con el fin de obtener reglas de inferencia que permitan procesar información que en lógica clásica no tendría mucho sentido hacerlo. De acuerdo con lo anterior un conjunto difuso se define como sigue [44].

Definición 1. sea X un conjunto universal clásico. Una función real $\mu_A : X \rightarrow [0, 1]$ se llama función de pertenencia de A y define un subconjunto A de X . A este subconjunto se le llama conjunto difuso A de X y está formado por el conjunto de parejas $(x, \mu_A(x))$ con $x \in X$.

Cuadro 3.1: *Diferencias fundamentales entre la lógica clásica y difusa.*

De manera formal la diferencia entre los dos tipos de lógica se da con la introducción del concepto función de pertenencia, así si $\mu_A(x) = 0$ $x \notin A$, si $\mu_A(x) = 1$ $x \in A$ y si $\mu_A(x) \in (0, 1)$ entonces x pertenece a A de forma parcial. Aunque son diferentes si aplicamos la 1 a un conjunto clásico no obtendríamos ninguna contradicción pues los valores de la función de pertenencia serían $\{0, 1\}$ pues $\mu_A(x) = \emptyset$ en $(0, 1)$. La ventaja al definir la función se da cuando se consideran valores $\mu_A(x)$ en $(0, 1)$, ya que se le puede asignar a una expresión ambigua en un conjunto dado una cantidad numérica la cual representa el grado de verdad ó falsedad que esta posee en el conjunto. Así los problemas quedan reducidos a la operación de cantidades sin ser relevante el modelo matemático que se está usando. El siguiente paso fue desarrollar un aparato teórico basado en estos conjuntos, sin embargo como se mencionó con anterioridad el objetivo no es presentar una estructura conjuntista ni algebraica de la lógica difusa sino presentar una introducción básica siendo suficiente el análisis presentado hasta el momento. Lo que sí se va a presentar es un ejemplo de función de pertenencia para claridad de la definición. Sea

$$\mu_A(x) = \begin{cases} 1 & \text{si } x < 0 \\ 1 - \frac{x}{a} & \text{si } x \in (0, a) \\ 0 & \text{si } x > a \end{cases} \quad (3.1.1)$$

esta función de pertenencia define un conjunto difuso en el cual, si $x < 0$ se tiene la total pertenencia al conjunto y si $x > a$ se tendría la no pertenencia. Ahora si $x \in (0, a)$ el nivel de pertenencia varía de forma decreciente. Para valores cercanos a cero se considera la pertenencia en un grado más alto que para los valores cercanos a "a". De entre las funciones de pertenencia que se pueden encontrar en la literatura se tienen: funciones trapezoidales, triangulares, sigmoidales, Gamma, Gaussiana, y Pseudo-exponenciales [47]. Como se sabe todas ellas están pensadas para dotar de estructura numérica a un grupo de predicados difusos. Por último, a día de hoy los controladores que son desarrollados con este tipo de lógica son usados en una gran variedad de electrodomésticos, sistemas de transporte, cámaras de video, aire acondicionado, etc. Por tanto, aunque estos sistemas todavía distan de ser perfectos dado su desventaja respecto a las redes neuronales, su facilidad de aplicación y

su bajo coste son los pilares con el cual se sostienen la investigaciones hechas con este método. Algunas de las investigaciones en esta materia comprenden áreas del conocimiento como Medicina e Ingeniería entre otros [48, 49, 50].

3.1.2. Algoritmos genéticos

En los algoritmos genéticos [51-53], las hipótesis propuestas por Darwin [54] en 1859 acerca de la evolución de las especies, podría decirse que constituyen el punto de partida filosófico de las ideas usadas en este método pero no es hasta 1975 que John Holland [55] plantea los principios básicos. Aunque fue Holland quien formalizó el concepto de algoritmo genético, la primera mención histórica está atribuida a Bagley [56] quien en 1967 implementó estos algoritmos en funciones de evaluación de juegos. Su idea se basó en buscar parámetros en las funciones para luego compararlos con algoritmos de correlación¹. Los algoritmos genéticos (*GA*) pertenecen a la estructura de la *computación evolutiva* que también contiene las *estrategias evolutivas*, las cuales constituyen el punto de partida de los conceptos utilizados en *GA*. Las estrategias evolutivas optimizaban parámetros reales en dispositivos específicos pero sin la idea de población o cruzamiento, solo era necesario un padre que mutara y produjera un descendiente, conservando el mejor. El científico que introdujo este concepto fue Rechenberg [57] y lo desarrolló Schwefel [58]. El siguiente desarrollo importante se denominó programación evolutiva, las ideas expuestas en él usaban métodos finitos para representar las posibles soluciones. Al igual que en las estrategias evolutivas, en esta técnica los estados de transición se realizan mediante mutación aleatoria escogiendo la mejor aproximación. Los creadores de ésta fueron Fogel, Owens y Walsh en 1966 [59]. A día de hoy existe una teoría amplia acerca de cada uno de los métodos antes mencionados pero en aquel momento, como se puede observar, todavía era necesario introducir el concepto de cruzamiento y población.

Holland [55] quería estudiar de una manera general los fenómenos de adaptación natural para luego representar las observaciones realizadas en sistemas computacionales. Mediante estas ideas logró desarrollar una técnica que le permitía encontrar una nueva población haciendo uso de una ya existente. La nueva población sería poseedora de mejores características si se compara con la anterior muestra. Esta propiedad se consigue teniendo en cuenta tres criterios simples: selección, cruce y mutación. Como ya se mencionó cada uno de los criterios en los que se basan los algoritmos genéticos tiene su razón de ser en la evolución natural de una población. Se podrían describir estos conceptos mediante su significado biológico no obstante nuestra idea es presentar los fundamentos matemáticos en el cual subyacen estos conceptos y que hacen que el modelo sea capaz de optimizar una muestra.

¹La correlación es la operación básica en los procesos de búsqueda de patrones por emparejamiento.

3.1.2.1. Esquemas

Lo primero es tener las bases suficientes para poder parametrizar un conjunto de individuos. Por esto se darán algunas definiciones concebidas para dicho fin, sin embargo es necesario saber que cada una de las herramientas a utilizar dependen directamente del problema a modelar. Por tanto y sin pérdida de generalidad todas las definiciones están basadas en el algoritmo genético simple [55] el cual se representa de la siguiente forma:

Procesos iniciales:

1. Población inicial.
2. Cálculo de la función de evaluación individual.

Ciclo nueva generación:

1. Insertar mutaciones.
2. Producir nueva generación.

Ciclo reproductivo:

1. Seleccionar dos individuos de la generación anterior.
2. Cruzar.
3. Mutar.
4. Cálculo función evaluación para cada individuo mutado.
5. Insertar mutaciones en el ciclo t nueva generación.

en el ciclo de reproducción, al escoger los individuos, es necesario que las funciones de evaluación² tengan una probabilidad proporcional. Otro concepto que debemos tener presente en la optimización es el de función objetivo. Sea la función $f : M \rightarrow \mathbb{R}$ con M un conjunto finito, si se desea optimizar f lo que se quiere encontrar es el valor máximo o mínimo de f en M . Todo depende de las condiciones de nuestro problema. No obstante, este proceso puede complicarse pues f puede tener mínimos locales o máximos locales. En este punto se hará referencia a mínimos, ya que el $\max\{f(x)\} = -\min\{-f(x)\}$.

Definición 2. sea $f : M \rightarrow \mathbb{R}$, se dice que f tiene un valor mínimo en $x \in M$ si existe un $\epsilon > 0$ tal que para todo $y \in (x - \epsilon, x + \epsilon)$ se tiene $f(x) \leq f(y)$.

²Una función de evaluación no es sino un número real que refleja la adaptabilidad del individuo al problema

cuando se haga mención a la idea de optimización ya se tiene el concepto definido y se sabrá a que se refiere. Otra aclaración pertinente, es la concerniente a la codificación del modelo. Para ello, se considerará una función $h : M \rightarrow A^l$ inyectiva, con $l \geq \log_{|A|} \|M\|$ y A un conjunto apropiado para representar los elementos de M , A se conoce como alfabeto. La inyectividad garantiza que los elementos de M sean separables dado que están relacionados con un y solo un elemento de A^l . En particular, se considerará el alfabeto $\{0, 1\}$ debido a que en la terminología comúnmente usada, al describir elementos del modelo simple, los bits representan genes los cuales a su vez conforman los cromosomas. Como se verá a continuación, las siguientes definiciones están pensadas para dotar a un alfabeto de su propia ontología y lo que pretenden es presentar una estructura que permita inferir el comportamiento de una población, objetivo del teorema fundamental de los algoritmos genéticos.

Definición 3. sea $A' = \{0, 1, *\}$ un alfabeto ampliado mediante el simbolo ”*”. Un esquema E es cualquier conjunto formado por elementos de A' . Ahora si k es la longitud de E , el número de esquemas asociados a A' es 3^k .

Definición 4. Sea r la cantidad de simbolos ”*” que tiene un esquema E , entonces a E se le asocian 2^r conjuntos. Además cada conjunto de longitud k se asocia con 2^k esquemas.

Hasta aquí se puede notar que las definiciones solo están usando conceptos conjuntistas para definir la cantidad de posibles esquemas formados a partir de un alfabeto ampliado, en lo referente a cantidad de genes. Las siguientes definiciones, aportan idea de orden y distancia en un esquema; sin embargo es necesario tener en cuenta que en un esquema estos conceptos distan de ser iguales a los trabajados comúnmente pero permiten obtener información de los operadores de mutación y cruce.

Definición 5. Al número de 0's y 1's que estén contenidos en un esquema E se le llama orden del esquema E y se denotará por $o(E)$. Por otro lado, a la distancia entre el primer y la último gen fijo de un cromosoma se le conoce como la longitud del esquema E , y se denotará por $\delta(E)$.

3.1.2.2. Propiedades de los esquemas:

1. Se puede saber el número de cromosomas en una población que son asociados a un esquema E en el tiempo t . Esta propiedad se denotará por $\xi(E, t)$ y se aplica de la siguiente manera: sea \mathcal{P}_t la población en un tiempo t y sean E_1 y E_2 dos esquemas de longitud l , entonces $\xi(E_1, t)$ y $\xi(E_2, t)$ representan la cantidad de cromosomas contenidos en la población que se pueden asociar con cada uno de los esquemas. Cabe aclarar que podría darse el caso que ninguno de los cromosomas se pueda asociar a uno de los esquemas dados.
2. También se tiene que los esquemas se adaptan con el tiempo. Esta adaptación se denotará por $val(E, t)$ y no es nada más que la media de la función objetivo de todos los cromosomas que se asocian a un esquema E . Sea

$$f : M \rightarrow A^l \rightarrow \mathbb{R} \tag{3.1.2}$$

$$x \rightarrow f(x) = g(h(x))$$

la función objetivo, entonces la media de la función objetivo de toda la población en el tiempo t se representa por $\overline{f(t)} = \frac{f(t)}{\alpha}$ donde α es el tamaño de la población inicial la cual será constante en la evolución.

lo anterior proporciona la información suficiente para enunciar el teorema fundamental de los algoritmos genéticos denominado teorema de los esquemas [52].

Teorema 6. *Teorema de los esquemas.* Sea $T : M \rightarrow \mathbb{R}$ donde

$$T_{sel\ cru\ mut}(\xi(E, t + 1)) \tag{3.1.3}$$

es el número esperado de individuos que se asocian a un esquema E y $\xi(E, t + 1)$ es el número de conjuntos que en una población en el tiempo $t + 1$ se asocian con un esquema E . Entonces, después de realizar la selección, el cruce y la mutación en un algoritmo simple se tiene

$$T_{sel\ cru\ mut}(\xi(E, t + 1)) \geq \xi(E, t) * \frac{val(E, t)}{f_1(t)} * \left(1 - \left(p_{cru} \frac{\delta(E)}{l - 1} + o(E) p_{mut} \right) \right) \tag{3.1.4}$$

No se hará de manera formal la demostración de este teorema puesto que para poder desarrollarla es necesario definir conceptos de cálculo diferencial y estadística. Sin embargo, se pretende presentar un bosquejo de la demostración³. En términos generales aunque muchas de las afirmaciones que se harán a continuación están bien definidas matemáticamente el objetivo es presentar de forma intuitiva cada una de las partes de las que consta esta demostración. En este sentido y dado

³Los cálculos de las constantes están simplificados pues algunos de ellos son extensos y como se mencionó el objetivo no es presentar una demostración formal.

que el teorema esta enunciado para un algoritmo genético simple, la demostración intuitiva se comprende de tres partes: lo primero es la selección de los individuos de una población para su posterior cruce. La probabilidad de que un individuo seleccionado se pueda asociar con un esquema E depende de su adaptación en el tiempo y de la función objetivo, esta última representa la suma de las funciones objetivos de los individuos de una población en el tiempo t .

$$Prob = \frac{val(E, t)}{f(t)} \quad (3.1.5)$$

una vez escogido los individuos se espera que la cantidad de individuos descendientes que pueden emparejarse con E después de realizar α selecciones sea

$$T_{sel}(\xi(E, t + 1)) = \alpha * \frac{val(E, t)}{f(t)} * \xi(E, t) \quad (3.1.6)$$

como la media de la función objetivo en el tiempo t es $\frac{f(t)}{\alpha}$ entonces

$$T_{sel}(\xi(E, t + 1)) = \frac{val(E, t)}{f(t)} * \xi(E, t) \quad (3.1.7)$$

ahora se supone que la evaluación del esquema se encuentra por encima de la media en una cantidad $\epsilon \overline{f(t)}$ siendo ϵ constante entonces

$$T_{sel}(\xi(E, t + 1)) = \frac{\overline{f(t)} + \epsilon \overline{f(t)}}{f(t)} * \xi(E, t) \quad (3.1.8)$$

$$T_{sel}(\xi(E, t + 1)) = (1 + \epsilon) * \xi(E, t) \quad (3.1.9)$$

y si la evaluación media de un esquema supera la media de toda la población, se puede afirmar que ésta incrementa de forma exponencial su presencia en otras generaciones. De forma contraria, si la evaluación de un esquema se encuentra por debajo de la media, ésta disminuye su presencia de forma exponencial en futuras generaciones. Partiendo de este criterio se obtiene que el operador de selección para $t = 0$ es igual a

$$T_{sel}(\xi(E, t)) = \xi(E, 0) * (1 + \epsilon)^t \quad (3.1.10)$$

por tanto, el crecimiento exponencial depende de la diferencia entre la evaluación

media de E y la global.

Para el proceso de cruce se tiene que la probabilidad de supervivencia de un esquema E al operador cruce es la siguiente

$$p_{so,cruce}(E) = 1 - p_{d,cruce}(E) \quad (3.1.11)$$

siendo $p_{d,cruce}$ la probabilidad de que el esquema E sea destruido por el operador cruce y se define como sigue

$$p_{d,cruce}(E) = \frac{\delta(E)}{l-1} * p_c \quad (3.1.12)$$

donde p_c es la probabilidad de cruce y $\delta(E)$ es una constante que mide la compatibilidad de la información contenida en una esquema E . Así la Ecuación 3.1.11 introduce una cota mínima para garantizar la supervivencia del esquema E . No obstante algunos de los esquemas sobreviven al operador $p_{d,cruce}$, pues si se consideran dos esquemas E_1 y E_2 iguales, los esquemas E_i resultantes independientemente de la restricciones impuestas por el operador mutación coinciden con E_1 y E_2 . Siendo suficiente para garantizar la supervivencia de ambos esquemas. De esta manera se puede garantizar que

$$p_{so,cruce}(E) \geq 1 - p_{d,cruce}(E) \quad (3.1.13)$$

De lo anterior se concluye que al realizar el cruce en una población seleccionada, el número esperado en dicha población es mayor que la probabilidad que tiene un grupo de individuos de ser escogidos multiplicado por la probabilidad de sobrevivir a la mutación.

$$T_{sel,cru}(\xi(E, t+1)) \geq \frac{val(E, t)}{f(t)} * \xi(E, t) * \left(1 - \frac{\delta(E)}{l-1} * p_c\right) \quad (3.1.14)$$

Por último, está el efecto producido por la mutación en un esquema. La probabilidad de que el esquema E sobreviva a la mutación se define así:

$$p_{so,mu}(E) = (1 - p_m)^{o(E)} \quad (3.1.15)$$

donde p_m es la probabilidad de mutación. Una característica que se debe tener en

cuenta para p_m es la de ser muy pequeña por lo general oscila entre 0,5% y 2% [52]. Por tanto, dado que $p_m \ll 1$, $p_{so,mu}$ se puede aproximar a lo siguiente

$$p_{so,mu}(E) \approx (1 - o(E)p_m) \quad (3.1.16)$$

ahora, al combinar el operador mutación con los operadores de cruce y selección el número esperado de individuos que se asocian a un esquema E en un tiempo t debe cumplir la siguiente ecuación

$$T_{sel,cru,mu}(\xi(E, t + 1)) \geq \frac{val(E, t)}{f(t)} * \xi(E, t) * \delta(p_c p_m) \quad (3.1.17)$$

siendo

$$\delta(p_c, p_m) = \left(1 - \frac{\delta(E)}{l-1} * p_c - o(E)p_m + \frac{\delta(E)}{l-1} o(E)p_m p_c\right) \quad (3.1.18)$$

Como ya se mencionó, las ideas desarrolladas hasta el momento están basadas en el algoritmo genético canónico presentado por Holland [55], no obstante, en el transcurso de décadas de investigación en esta disciplina ha sido necesario presentar nuevos algoritmos, los cuales se adaptan a necesidades específicas y solucionan problemas de convergencia en modelos más complejos. Lo relevante de este método es el poder predecir el comportamiento en el tiempo de una población en particular⁴. Existen trabajos que relacionan algoritmos genéticos con lógica difusa [60] pero no se pretende hablar de ello solo se pretende hacer saber que aunque cada uno de estos métodos ha sido desarrollado de forma independiente a día de hoy tanto las aplicaciones realizadas como la teoría poseen un grado de libertad mucho más amplio que el exhibido en este trabajo sobrepasando las limitaciones teóricas que se muestran en este escrito⁵.

3.2. Redes neuronales Artificiales

3.2.1. Introducción histórica

El desarrollo de las redes neuronales artificiales (*ANN*) se remonta hasta 1936 cuando Alan Turing [61] buscando dotar de ontología al mundo de la computación es-

⁴Aplicaciones algoritmos genéticos. Optimización de redes eléctricas mediante la aplicación de *GA* http://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0718-07642009000400016. Búsqueda de caminos entre dos puntos usando *GA* <http://www.tecdigital.itcr.ac.cr/revistamatematica/contribuciones-v6-n1-may2005/Geneticos/Index.htm>.

También se pueden encontrar aplicaciones en Medicina, Economía, Sociología entre otras.

⁵Algoritmos genéticos basados en lógica borrosa. Sección 4 <http://eddyalfaro.galeon.com/geneticos.html>

tudiando el cerebro sienta las bases de esta disciplina, pero no es hasta 1943 que gracias a el neurofisiólogo McCulloch y el Matemático Pitts [62] con su trabajo “introducción a la simplificación de neuronas” que se pueden ubicar los cimientos de las redes neuronales. La hipótesis que abrió un camino en esta investigación es la siguiente: el aprendizaje como un proceso inherente a la naturaleza humana no ocurría de forma estrictamente local a una porción del cerebro sino que por el contrario debía ser tratado de forma global. Inicialmente esta hipótesis lleva a supeditar la conectividad de la neuronas cerebrales al aprendizaje de cosas nuevas y la efectividad de una conexión ó sinapsis a una repetida activación entre una y otra neurona a través de la misma conexión. La gran contribución de McCulloch y Pitts [62] es la de lograr introducir características plausibles en el funcionamiento de los elementos principales (neuronas). Para ello los elementos primarios eran constituidos por el conjunto $\{0, 1\}$ los cuales tenían dos propiedades básicas:

- La existencia de un umbral por debajo del cual la neurona esta inactiva.
- Cada neurona puede recibir entradas excitadoras ó inhibitorias cuya acción es global.

en su hipótesis la inactividad de la neurona esta relacionada con la existencia de alguna de las dos particularidades anteriores. Si no existe ninguna entrada inhibidora entonces se calcula el resultado de las entradas excitadoras, si ésta es mayor al umbral, el resultado de salida es 1 y si es menor el resultado de salida es 0. Aunque este modelo distaba de ser perfecto, en su epoca constituyó un avance considerable en la comprensión de las redes neuronales dado que mediante el uso de funciones lógicas se podría caracterizar redes complejas. No obstante, esta idea no estaba lo suficientemente depurada como para solventar los problemas que se le presentaron. En 1956 en la conferencia sobre inteligencia artificial organizada en Dartmouth, Rochester, Holland, Haibt y Duda [63] siguiendo ideas conexionistas⁶ presentaron la primera simulación de una red neuronal. Esto es considerablemente significativo ya que por primera vez se podría constatar mediante simulaciones una teoría neuronal. Dado que este trabajo tenía fines puramente practicos fue necesario acotar los valores sinápticos y suponer ciertas particularidades del espacio.

En 1958 Rosenblatt [64] desarrolla el *Perceptrón* como una aproximación al problema de reconocimiento de patrones, esta red es capaz después de aprender ciertos patrones de deducir patrones similares sin haberle sido presentados anteriormente. La idea de aprendizaje era el pilar en el cual se sustentaba este concepto. En 1960 Widrow y Hoff [65] presentan *ADALINE* sistema subordinado a un algoritmo de aprendizaje que utiliza la Media de los Mínimos Cuadrados (*LMS*) y su idea era proponer un sistema adaptativo que pudiera aprender precisa y rápidamente. *ADALINE* fue aplicado en filtros adaptativos con el fin de eliminar ecos en las

⁶No tratar el aprendizaje como un proceso local

líneas telefónicas, convirtiéndola en la primera red aplicada a un problema real. El interés por esta teoría llegó a tal punto que se pensaba que las redes neuronales tenían la solución a numerosos problemas, sin embargo en 1969 se publica por parte de Minsky y Papert [66] un trabajo el cual paralizó los resultados en este campo durante un tiempo considerable. En dicho trabajo demostraron que el método de *Perceptrón* tenía limitaciones conceptuales, pues los únicos problemas que se podían abordar con este sistema eran los linealmente separables. Lo anterior causó decepción entre los investigadores interesados en las redes neuronales debido a que los problemas de ese tipo eran muy pocos. Aunque el trabajo de Minsky y Papert [66] paralizó el desarrollo en redes neuronales durante un tiempo considerablemente largo, algunos investigadores continuaron sus trabajos y no es hasta 1980 que Grossberg [67] plantea diferentes modelos muy útiles en el campo de reconocimiento de patrones llamados *ART* los cuales usan un nuevo principio de auto-organización.

En 1982 resurge el interés en las redes neuronales debido a Hopfield [68] quién gracias a su formación en Física presenta una teoría cuya base consistía en comprender la dinámica de una red recurrente con simetría en sus conexiones sinápticas haciendo uso de una función de energía. Inicialmente esta teoría estaba pensada para salidas binarias y bipolares, limitación que fue subsanada más adelante con la ampliación de la función de energía a salidas continuas. Desde este momento se puede decir que se produjo el resurgir de las redes neuronales, pues las investigaciones en este tema fueron más interesantes y variadas. Como la hecha por Kohonen [69] quien presenta su investigación acerca de mapas auto-organizativos, sustentada en la idea de que el aprendizaje en un modelo planteado no necesitaba ser supervisado. Otras como la hecha por Fukushima, Miyake e Ito [70] combinaban conocimientos en ingeniería, fisiología y teoría neuronal aplicada a reconocimiento de patrones o el aprendizaje reforzado aplicado en control. También gracias a Kirkpatrick, Gelat y Vecchi [71] se formalizan conceptos en optimización los cuales son importantes en el aprendizaje estocástico. Cada una de estas investigaciones trajo consigo beneficios, problemas y ventajas inherentes a las limitaciones en el aprendizaje. Pero no es hasta 1986 que Rumelhart, Hinton y Williams [72] presentan el algoritmo de retropropagación que junto a la investigación de Hopfield [68] representan el punto de inflexión en la investigación acerca de redes neuronales. Desde esta época en adelante, el número de investigaciones sobre redes neuronales ha alcanzado una relevancia considerable dada la cantidad de aportes realizados por los investigadores, tanto en métodos de aprendizaje como arquitecturas y aplicaciones.

3.2.2. Fundamentos

Después de esta breve introducción histórica, es posible tratar de caracterizar las redes neuronales más relevantes en las investigaciones. Las redes neuronales se clasifican teniendo en cuenta aspectos tales como: el tipo de aprendizaje, topología, valores en las entradas y funciones de activación entre otros. Es por esto que se

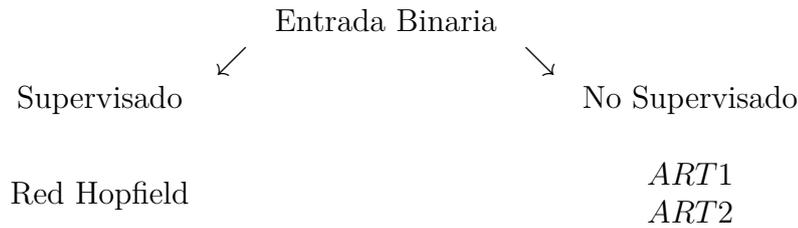
presentarán algunos esquemas y métodos los cuales exhibiran las diferentes clasificaciones encontradas en la literatura [73, 74, 75]. Inicialmente se caracterizarán la redes neuronales mediante el tipo de aprendizaje. De acuerdo con los fundamentos filosóficos el aprendizaje se ve afectado por diferentes variables que cambian con el tiempo y están relacionadas de forma directa tanto con parámetros internos como externos al individuo. Dada la naturaleza de las redes neuronales, se ha logrado deducir que estas no están exentas de dichas consideraciones, por esto y buscando catalogar las variaciones de la red para obtener las salidas deseadas se han clasificado en dos:

- **Algoritmos de aprendizaje supervisados** a este tipo de algoritmos de aprendizaje se les debe introducir las salidas que se esperan de la red. Una vez entrenada se deben comparar las salidas obtenidas con las deseadas para calcular el error cometido, el cual es usado posteriormente para modificar los pesos. De la misma manera, los algoritmos de aprendizaje supervisados se pueden dividir en dos tipos aprendizaje, por refuerzo y por corrección de error cada uno con características básicas. El aprendizaje por corrección de error consiste en ajustar los pesos en las conexiones de la red en función de la diferencia entre los valores deseados y obtenidos de salida. Esto se llama función de error cometido en la salida. El aprendizaje por refuerzo es más lento que el anterior y se basa en la idea de no disponer de la salida que se desea durante el entrenamiento ante una determinada entrada.
- **Algoritmos de aprendizaje no supervisados** en estos se conoce la salida que nos debe proporcionar la red pero a diferencia del aprendizaje supervisado, ésta organiza por si sola los diferentes valores de entradas. La asociación que se realiza en este aprendizaje depende estrictamente de las características de la señal.

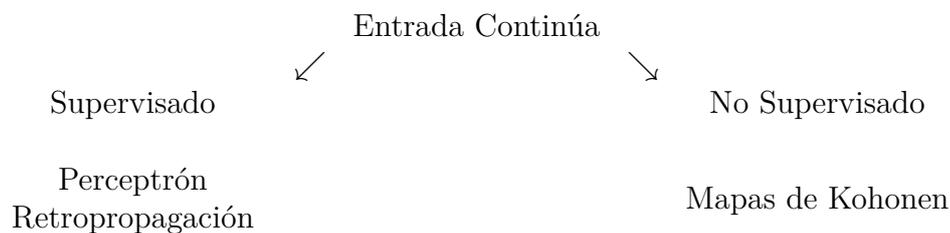
Luego se pueden catalogar las *ANN*'s dependiendo de si su entrada es binaria o continua. Como se puede observar las entradas binarias se clasifican en el Cuadro 3.2 y las entradas continuas en el Cuadro 3.3 . De acuerdo a esta clasificación y gracias a la naturaleza de esta investigación, se usarán los métodos de entrada continua.

Respecto a los modelos neuronales, como se mencionó anteriormente, el cerebro es el prototipo usado para el desarrollo de éstas, por ende el carácter ontológico de las herramientas usadas en dichos sistemas es similar al sistema biológico. Las redes neuronales artificiales son sistemas interconectados entre sí compuestos por neuronas artificiales con el objetivo de ayudarnos a comprender el funcionamiento de una red neuronal biológica o no, permitiendo interactuar con redes neuronales complejas y proporcionando aproximaciones fiables. Por ejemplo, uno de los sistemas biológicos de redes neuronales más complejo es el sistema nervioso, en éste las *ANN* intentan comprender los métodos de abstracción de información centrándose en las diferentes

Cuadro 3.2: *clasificación de las redes por entrada binaria.*



Cuadro 3.3: *clasificación de las redes por entrada continua.*



formas de procesamiento de información.

Las unidades de proceso fundamental de una red neuronal son las neuronas. Cada red neuronal goza de características específicas para su funcionamiento, sin embargo las neuronas fueron concebidas respetando las singularidades del sistema biológico. Es decir, su función es la de recoger, procesar y transmitir información. Una neurona se representa de la siguiente forma:

$$F_j(x_i, \psi_{ij}) = f_j(\sigma(x_i, \psi_{ij}), \theta_i) \tag{3.2.1}$$

donde $i, j = 1, \dots, n$ $n \in \mathbb{N}$, x_i denota el **conjunto de entradas**, ψ_{ij} los **pesos sinápticos** del grado de conexión existente entre las neuronas artificiales i y j , $\sigma(x_i, \psi_{ij})$ la **regla de propagación** de la neurona, la cual depende directamente del operador usado ó la necesidad del modelo⁷, f_j la **Función de activación** proporciona el estado de activación actual de la neurona y se define como el potencial resultante entre la regla de propagación y un estado de activación anterior de la neurona θ_i asociado comúnmente a la σ . Por último, F_i es la **Función de salida** de la neurona i , como se puede observar en la Ecuación 3.2.1 esta proporciona un valor de salida de la neurona. Las reglas de propagación se dividen en estáticas y dinámicas, no se

⁷Por lo general es representada por la suma de las entradas multiplicadas por los pesos sinápticos.

trabaja con estas últimas pero se puede decir que en este tipo de reglas el resultado esperado no solo depende de las entradas, sino también de las salidas anteriores. De entre las reglas de propagación estática más comunes se tienen las siguientes:

1. Usual o Canonica

$$\sigma(x_i, \psi_{ij}) = \sum_{i=1}^n x_i \psi_{ij} \quad (3.2.2)$$

2. Cuadrática

$$\sigma(x_i, \psi_{ij}) = \sum_{i=1}^n x_i^2 \psi_{ij} \quad (3.2.3)$$

3. Polinómica

$$\sigma(x_i, \psi_{ij}) = \sum_{i=1}^n \sum_{k=1}^n x_i x_k \psi_{ij} \quad (3.2.4)$$

4. Esférica

$$\sigma(x_i, \psi_{ij}) = \frac{1}{\rho^2} \sum_{i=1}^n (x_i - \psi_{ij})^2 \quad (3.2.5)$$

En cuanto al diseño de las funciones de activación se debe tener en consideración que el uso de estas funciones depende única y exclusivamente del modelo a desarrollar. La elección de dicho grupo se debe explícitamente a resultados obtenidos en entrenamientos de redes pero no deja a un lado el poder construir nuevas funciones que podrían ser relevantes en la investigación a desarrollar. Por tanto dichas funciones se pueden clasificar como sigue: entre $(0, 1)$ están la función McCulloch-Pitts y la función Sigmoide y entre $(-1, 1)$ se tiene la Función tangente hiperbólica, la función a trozos, y la función Gaussiana.

1. Función McCulloch-Pitts

$$f_i = \begin{cases} 1 & \sigma(x_i, \psi_{ij}) \geq 0 \\ 0 & \sigma(x_i, \psi_{ij}) \leq 0 \end{cases} \quad (3.2.6)$$

2. Función Sigmoide

$$f_i = \frac{1}{1 + e^{(-a\sigma(x_i, \psi_{ij}))}} \quad (3.2.7)$$

3. Función tangente hiperbólica

$$f_i = \frac{1 - \epsilon^{(-a\sigma(x_i, \psi_{ij}))}}{1 + \epsilon^{(-a\sigma(x_i, \psi_{ij}))}} \quad (3.2.8)$$

4. Función a trozos

$$f_i = \begin{cases} 1 & \sigma(x_i, \psi_{ij}) \geq \frac{1}{2} \\ \sigma(x_i, \psi_{ij}) + \frac{1}{2} & -\frac{1}{2} > \sigma(x_i, \psi_{ij}) > \frac{1}{2} \\ 0 & \sigma(x_i, \psi_{ij}) < -\frac{1}{2} \end{cases} \quad (3.2.9)$$

5. Función Gaussiana

$$f_i = C_1 \epsilon^{\left(\frac{\sigma(x_i, \psi_{ij}) - C_2}{C_3}\right)^2} \quad (3.2.10)$$

$C_i, i = 1, 2, 3$, constantes.

La última clasificación que se hará es la concerniente a las unidades funcionales de una red neuronal, las capas, elementos los cuales dan lugar a estructuras neuronales formadas por grupos de neuronas interactuando de formas diferentes; por ejemplo, de acuerdo al número de capas, tipo de conexión o grado de conexión. La primera interacción y la más simple de comprender es la concerniente al grado de conexión. Se dice que una red está completamente conectada si todas las neuronas de una capa se encuentran conectadas con las neuronas de otra capa o parcialmente conectada si por el contrario existe al menos una neurona de una capa que no está conectada con otra neurona perteneciente a otra capa.

Si se clasifican teniendo en cuenta el número de capas, se obtienen dos grupos

- **Redes neuronales monocapa** es la red más sencilla pues solo consta de una capa de entrada encargada de enviar la información de la red a la capa de salida cuya función es realizar los cálculos. Si nos apegamos de manera formal a la definición tendríamos dos capas en lugar de una pero por convención la capa de entrada no se cuenta dado que no efectúa cálculos.
- **Redes neuronales multicapa** constituida por una capa de salida y varias capas intermedias entre la capa de entrada y la de salida llamadas comúnmente capas ocultas, en ellas se realizan cálculos y su cantidad puede variar dependiendo de las necesidades.

Ahora según el tipo de conexiones se tienen las siguientes redes

- **Redes neuronales sin realimentación o conexión hacia adelante** (feed-forward) como su nombre indica la propagación de la información se produce en un solo sentido y no existe posibilidad de realimentación o reciclaje de la información.
- **Redes neuronales recurrentes** (feed-back) a diferencia de la anterior, en ésta sí existe la posibilidad de realimentación de la red. La conexión entre neuronas se produce de diferentes formas; conexión con neuronas de diferentes capas, la misma capa o entre la misma neurona.

En síntesis las redes neuronales tienen una arquitectura muy elaborada que consta de varios niveles: el inicial o fundamental es donde se encuentran los elementos básicos denominados apropiadamente como neurona. Teóricamente estos elementos por si solos serían los elementos más pequeños de la red pero no los menos importantes ya que gracias a ellos se pueden definir parámetros indispensables en su funcionamiento. El siguiente nivel corresponde a la configuración de la red neuronal entre los que se encuentran: tipos de conexiones en la red, configuración de los elementos básicos, cantidad de neuronas en un capa y número de capas. Hasta aquí lo visto anteriormente es suficiente para entender el funcionamiento y los usos de dichos niveles. Lo siguiente a tener en cuenta es el amplio espectro donde inciden las redes neuronales, las denominadas estructuras neuronales. Dado que se pueden usar en diferentes tipos de estructuras, las cuales dependen directamente de la aplicación que se desee. Dichas estructuras se han clasificado de la forma siguiente:

- **Estructura directa:** la red neuronal se asocia con un sistema desconocido que tendrá las mismas entradas de la red neuronal. En esta estructura el mínimo error se consigue cuando las salidas de la red neuronal son iguales a las del sistema desconocido o dicho de otra forma, la función de transferencia en ambos sistemas es la misma.
- **Estructura inversa:** en esta estructura el mínimo error se obtiene si al hacer la asociación de la *ANN* con el sistema desconocido, la salida de la red neuronal es la entrada al sistema desconocido. La función inversa tiene un papel relevante pues de ella depende la estabilidad de la estructura.
- **Estructura con retardo:** la idea es modelizar la señal de salida de la red neuronal en un instante t a partir de los valores k anteriores a ese instante. En este sentido, la salida en el instante t se obtiene realimentando la red con valores obtenidos en instantes $k \in [1, n)$ con $n < t$. El error de esta red se minimiza cuando al igual que la red directa, la salida de la red es igual a la

salida de la señal deseada.

- **Estructura con cancelador de ruido:** en esta estructura lo que se desea es minimizar el error cuadrático medio entre la señal de referencia y la salida de la red neuronal, la cual asume el papel fundamental de cancelador de ruido. La diferencia con las otras estructuras se da gracias a que la señal deseada posee un ruido, el cual se desea limpiar, por tanto el mínimo error en esta estructura se obtiene cuando la salida de la red es igual al ruido que se desea limpiar en la señal. Cabe aclarar que las entradas de la red neuronal tienen una correlación con el ruido que se encuentra en la señal a limpiar.

Una vez hecha esta breve introducción acerca de la *AI* en general y de forma particular de las *ANN* se procederá a desarrollar los conceptos relevantes en este proyecto permitiendo acotar ideas para centrarse en el análisis de resultados del reconstructor de prueba.

4 Las Redes Neuronales en nuestra investigación

Las redes neuronales han evolucionado de tal forma que en la literatura existente sobre el tema hay una variedad considerable de tipos, las cuales, como se pudo observar en el capítulo anterior, dependen de diferentes clases de funciones, tipo de conexiones entre neuronas, definición de elementos fundamentales y en general de la configuración deseada al abordar un problema en particular. Dichas características, han sido definidas a lo largo de la historia de las *ANN*'s. De esta manera y partiendo de los criterios presentados, se procederá a acotar esta investigación. Inicialmente, se trabajará con redes neuronales con conexión hacia adelante y de estructura directa. En este sentido las redes perceptrón multicapa, las cuales usan el algoritmo de retropropagación [72], fueron la primera opción dado que son las más representativas y usadas. Aunque su arquitectura es simple, no es menos eficiente pues el tipo de problema a modelar se enmarca en el procesamiento de señales y como se sabe, estas redes han logrado resultados considerables en este campo [76].

Después de acotar la investigación, lo siguiente es centrarse en las herramientas relacionadas con ésta. Se realizará una introducción acerca de la red y el tipo de algoritmo a usar, para luego discutir el diseño y entrenamiento de la red. Lo que se pretende con esto es lograr abordar de modos diferentes la investigación para crear conocimientos, los cuales servirán en esta y en futuras investigaciones.

4.1. Método de perceptrón Multicapa

Como se mostró anteriormente, históricamente el perceptrón fue la primera red neuronal y su funcionamiento se enmarca en el de estructuras directas [64]. En este tipo de red no existe la alimentación o conexión hacia adelante e inicialmente se propuso como un algoritmo monocapa cuyo objetivo era predecir la siguiente transformación

$$h : \{-1, 1\}^N \rightarrow \{-1, 1\} \quad (4.1.1)$$

en este proceso se usaba una muestra \vec{x} la cual constituía el conjunto de entradas a la red con sus salidas respectivas $h(\vec{x})$. En la definición inicial, la regla con las que asociaban los entradas y las salidas podía ser cualquiera, pero en el proceso de aprendizaje los pesos eran ajustados solamente a los valores de salida esperados

$h(\vec{x})$, esto ocurría pues no se había desarrollado de manera formal una conexión entre los elementos que conformaban la red¹. Se sabe gracias a Minsky y Papert [66] la gran variedad de restricciones que se tienen al abordar un problema con un perceptrón, pero ellos no presentaron una solución al problema del ajuste de pesos el cual fue resuelto por Rumelhart, Hinton y Williams [72] mediante la construcción de un algoritmo de aprendizaje denominado algoritmo de retropropagación. La solución consistía en realizar varios procesos que involucraban todas las capas de la red neuronal de la siguiente manera: inicialmente a la capa de entrada se le introduce una muestra, la cual se propaga capa a capa por toda la red hasta producir un conjunto de salidas de dicha muestra. El conjunto obtenido representa la respuesta en un momento específico, en este primer momento, los pesos de las conexiones se mantienen fijos y se realiza la diferencia entre la muestra obtenida y el resultado deseado; a esto se le llama señal de error. El objetivo final es minimizar la señal de error en la salida. Si es muy grande, el error obtenido es enviado hacia atrás por la red en dirección contraria a la de los pesos, permitiendo ajustarlos iterativamente hasta que la salida producida por la red sea lo más parecida a la deseada. Una vez descrito de forma práctica el funcionamiento de dicho algoritmo, sus bases teóricas subyacen en las ideas que se describirán a continuación.

En la literatura se encuentra que al algoritmo de retropropagación también se conoce como regla delta generalizada [65]. La regla delta fue desarrollada para que existiera un mecanismo que lograra adaptar los pesos hasta obtener el valor más cercano posible al deseado y se definió así: para una red neuronal monocapa, la salida de la red es la siguiente

$$f = \sum_i w_i x_i + \theta \quad (4.1.2)$$

con w_i los pesos, x_i , f las entradas y salidas respectivamente y $\theta = w_0$ peso inicial. Esta definición proporciona una relación lineal entre las entradas y las salidas. Como ya sabemos el problema es encontrar los pesos adecuados para obtener la salida deseada. Así, si x es el conjunto de entradas, s las salidas deseadas y f la salida obtenida, el error está definido por la diferencia entre $(s - f)$. Para calcular este error la regla delta usa mínimos cuadrados y se determina de la siguiente manera

$$E = \frac{1}{2} \sum (s - f)^2 \quad (4.1.3)$$

generalmente las características de cada conjunto están determinadas por la naturaleza de la muestra \vec{x} , pero es fundamental que tanto \vec{s} como \vec{f} tengan la misma dimensión. Dado que el cambio en el error se quiere proporcional a la derivada, la variación en cada peso debe ser proporcional a la siguiente relación, obtenida mediante la derivación de la función error respecto a los pesos.

$$\Delta w_i = -C \frac{\partial E}{\partial w_i} = -C \left(\frac{\partial E}{\partial f} \frac{\partial f}{\partial w_i} \right) \quad (4.1.4)$$

¹Actualmente dicha conexión esta representada por las funciones de activación.

siendo C una constante denominada constante de proporcionalidad. Al derivar la Ecuación 4.1.2 y la Ecuación 4.1.3 y reemplazar los valores en la Ecuación 4.1.4 la relación se reduce a la siguiente ecuación

$$\Delta w_i = C (s - f) x_i \quad (4.1.5)$$

como se observa, la función de la regla delta es modificar el valor de los pesos pero el gran aporte de este método es que funciona tanto para valores continuos y discontinuos ya sean de entrada o salida.

Este primer caso es importante pues ayuda a comprender la filosofía intrínseca en esta definición, la cual no se tiene presente debido a la linealidad considerada inicialmente. Para la generalización de esta regla se considera que la función de activación es una función diferenciable dada por

$$f_j = \mathcal{F}(h_j) \quad (4.1.6)$$

donde

$$h_j = \sum_i w_{ij} f_i + \theta_j \quad (4.1.7)$$

$$E = \frac{1}{2} \sum_{k=1}^n (s_k - f_k)^2 \quad (4.1.8)$$

la única diferencia notoria en esta definición se da gracias a que ahora abarca patrones más amplios, ya sea en términos de dimensión o en sus características. Pero como se verá a continuación la idea subyacente sigue el mismo proceso. Ya se sabe que s_k y f_k son la salida deseada y la obtenida, pero en este caso el conjunto de elementos varían de $1, \dots, n$. Por tanto, el error está formado por un vector o por una matriz de vectores; todo depende de las características de la muestra. De la misma forma la idea es qué el cambio en cada peso esta definido por la ecuación

$$\Delta w_{ij} = -C \frac{\partial E}{\partial w_{ij}} = -C \left(\frac{\partial E}{\partial h_j} \frac{\partial h_j}{\partial w_{ij}} \right) \quad (4.1.9)$$

en este caso al realizar la derivada de la Ecuación 4.1.7 respecto a los pesos, se obtiene

$$\frac{\partial h_j}{\partial w_{ij}} = f_i \quad (4.1.10)$$

para la segunda parte de la derivada, se derivar implícitamente de nuevo pero para elementos de dimensión mayor. Entonces

$$\frac{\partial E}{\partial h_j} = \frac{\partial E}{\partial f_j} \frac{\partial f_j}{\partial h_j} \quad (4.1.11)$$

de la Ecuación 4.1.6 se tiene

$$\frac{\partial f_j}{\partial h_j} = \mathcal{F}'(h_j) \quad (4.1.12)$$

ahora para encontrar la derivada parcial de la función error E respecto a f_j se deben de tener en cuenta dos casos; el primero se tiene al considerar que $j = k$, es decir, j es una unidad de salida de la red. Entonces la derivada parcial del error respecto a f_j se reduce a

$$\frac{\partial E}{\partial f_j} = -(s_k - f_k) \quad (4.1.13)$$

de lo anterior se deduce que los pesos se actualizarían como sigue

$$\Delta w_{ji} = C (s_k - f_k) \mathcal{F}'(h_k) f_i \quad (4.1.14)$$

para el caso que j no sea una de las salida sino una unidad perteneciente a la capa oculta no es claro el comportamiento del error de la red, sin embargo se puede representar escribiendo las salidas obtenidas en la capa oculta teniendo en cuenta las siguientes ideas: sea $j = v$ donde v pertenece a la capa oculta y

$$E = E(h_1, h_2, h_3, \dots) \quad (4.1.15)$$

mediante la regla de la cadena

$$\frac{\partial E}{\partial f_v} = \sum_{m=1}^n \frac{\partial E}{\partial h_m} \frac{\partial h_m}{\partial f_v} \quad (4.1.16)$$

$$= \sum_{m=1}^{n_1} \frac{\partial E}{\partial h_m} \frac{\partial}{\partial f_v} \sum_{i=1}^{n_2} w_{jm} f_i \quad (4.1.17)$$

de la Ecuación 4.1.6 y la Ecuación 4.1.8 se obtiene

$$= \sum_{m=1}^{n_1} \frac{\partial E}{\partial h_m} w_{vm} = - \sum_{m=1}^{n_1} (s_m - f_m) \mathcal{F}'(h_m) w_{vm} \quad (4.1.18)$$

por último, al reemplazar el resultado en la Ecuación 4.1.11, los pesos para este caso se actualizan así

$$\Delta w_{ji} = C \left(\mathcal{F}'(h_v) \sum_{m=1}^{n_1} (s_m - f_m) \mathcal{F}'(h_m) w_{vm} \right) * f_i \quad (4.1.19)$$

Las bases en las cuales se sustenta el algoritmo de retropropagación [72] están completas. Ahora es posible deducir algunas de las características relevantes subyacentes

en la redes perceptrónicas multicapa. En primera instancia queda claro la existencia de la no linealidad en la salida de cada una de las neuronas de la red, pero esta característica no ha sido considerada a la ligera, puesto que para que funcione el algoritmo de retropropagación es necesario que la función de salida sea diferenciable. De considerarse la linealidad en la relación entrada salida, ésta se convierte en una transformación lineal, principal problema de las primeras redes definidas. Otro elemento importante a tener en cuenta ocurre en las capas ocultas ya que el alto grado de aleatoriedad que poseen al ejecutar tareas de aprendizaje, les permite extraer información extra relevante y aprender de ella para reutilizarla en aras de una mejor comprensión de los datos que progresivamente están siendo ingresados en la red. Por último, el alto grado de conectividad en estas redes y las características definidas anteriormente las hacen acreedoras de una posición relevante en la solución de problemas entre los que se tienen los de clasificación y predicción. En esta investigación las redes perceptrónicas multicapa serán usadas para resolver problemas de tipo predictivo.

4.2. Procesado de la información y uso de *CARMEN*

Como se mencionó anteriormente la reconstrucción tomográfica es el pilar donde se sustentan los resultados de esta investigación, ya que se han convertido en una herramienta fundamental para los investigadores que usan sistemas de óptica adaptativa modernos [77]. Algunos de los trabajos desarrollados en óptica adaptativa [1-33] dejan ver el profundo estudio que se está realizando en esta rama del conocimiento. Para saber el funcionamiento de cada uno de estos sistemas refiérase al Capítulo 2, en él se pueden encontrar los fundamentos teóricos del comportamiento del espacio muestral y del proceso de reconstrucción, así como también diferencias, ventajas y desventajas respecto al uso de los distintos sistemas en óptica adaptativa.

En esta investigación se ha desarrollado un reconstructor basado en redes neuronales artificiales (*ANN*) llamado *CARMEN* (Complex Atmospheric Reconstructor based in Machine lEarNing) el uso de este reconstructor se enmarca en la misma línea de otros reconstructores, los cuales usan y combinan la información proporcionada por las estrellas guías. Entre los sistemas se tienen mínimos cuadrados (*LS*) [77] y el método aprender y aplicar (*L&A*) [78]. Una de las ventajas de *CARMEN* respecto de los otros reconstructores se da gracias a que para trabajar tanto con *LS* como *L&A* se necesita conocer el perfil de turbulencia y es imprescindible recalibrar los parámetros del sistema si se produce una variación en los datos, pero con *CARMEN* no es necesario realizar este proceso. En la Figura 4.2.1 se puede observar un diagrama simplificado del funcionamiento de *CARMEN* así como también el número de capas usadas, sin embargo tanto su topología como arquitectura fueron encontradas mediante experimentación, siendo el principal criterio obtener el mejor resultado. Para ello se testearon una cantidad considerable de redes con diferentes

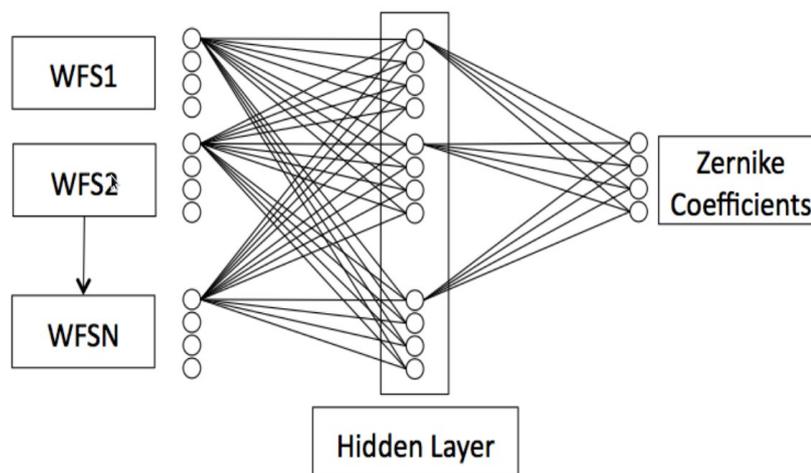
configuraciones, este proceso se discutirá más adelante.

CARMEN puede ser catalogado como un reconstructor el cual usa para su entrenamiento el algoritmo de retropropagación [72] para intentar minimizar la diferencia media cuadrada mínima sobre todos los datos de entrenamiento. Las entradas de la red son las pendientes de los objetivos medidas fuera de eje y las salidas son los coeficientes de los Zernikes [31] dentro de los ejes. Su funcionamiento se puede sintetizar como sigue: primero se entrena con un conjunto de datos los cuales simulan la turbulencia atmosférica a diferentes altitudes. Como se vio anteriormente, la turbulencia tiene un papel relevante en la distorsión del frente de onda². Los datos usados para el entrenamiento tienen la característica de estar elaborados teniendo en cuenta un gran número de casos de los cuales la salida es conocida, dicha peculiaridad es importante debido a que es necesario que la muestra considerada en esta etapa sea lo suficientemente grande como para que la red se entrene teniendo en cuenta todos los posibles escenarios. Esta es la diferencia fundamental al usar este reconstructor respecto a otros, pues una vez entrenada la red ella debe poder trabajar con cualquier perfil de turbulencia sin ser reentrenada o reoptimizada.

Como parte del proceso de aprendizaje y para efectuar la validación, se construyó un banco óptico que simuló un telescopio enfocado hacia cuatro estrellas. Una de ellas es el objeto a reconstruir y las otras tres son estrellas guías que lo rodean. En el banco las estrellas guías fueron distribuidas en forma de triángulo alrededor de la estrella a observar. La distancia desde la estrella ubicada en el centro del triángulo y cada una de las estrellas que componen las aristas de dicha figura es de 6 mm. En el proceso de simulación de las estrellas se usó fibra óptica, pues permite crear una fuente puntual de los haces de luz emitidos por las estrellas. Lo anterior se hizo con el fin de colimar los haces de luz para generar haces paralelos, no obstante en un escenario real los cuatro haces de luz pueden llegar a converger. La pantalla de fase (fase-screen) es un elemento diseñado para simular una capa de turbulencia a la altura que se desee permitiendo calcular el solapamiento que existe entre los haces del luz. Después de la pantalla de fase, se encuentra un par de lentes conjugados que vuelven a colimar y magnifican los haces de luz para que el sensor Sharck-Hartmann se pueda instalar en la posición de la pupila. En este banco solo se usó una cámara y un sensor de frente de onda para captar las cuatro estrellas, logrando que el diseño óptico sea simple además de económico. Una vez obtenidos los datos de validación estos son pasados a *CARMEN* para que pueda realizar la reconstrucción para la que fue diseñada.

²Referase al Capítulo 2.

Figura 4.2.1: Diagrama simplificado de funcionamiento y arquitectura de *CARMEN*.

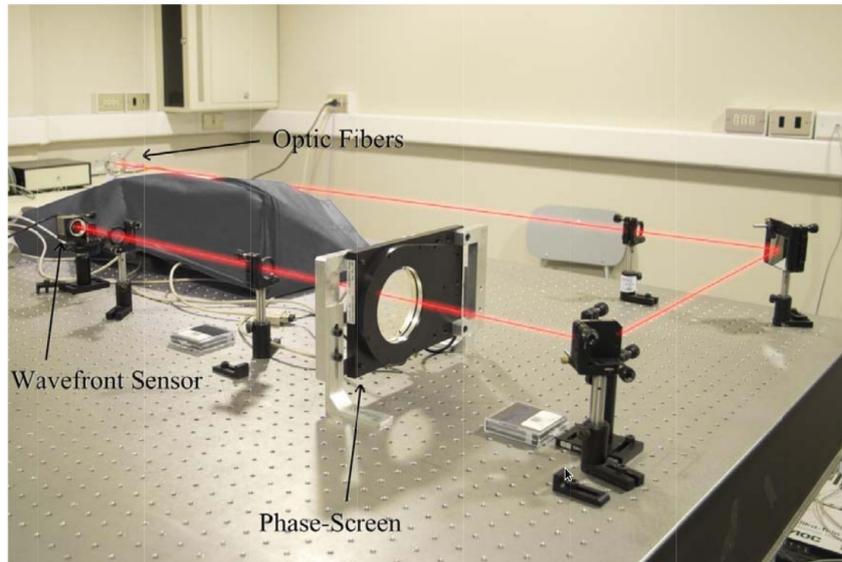


4.2.1. Diseño

Hasta ahora se ha descrito de forma general el proceso de entrenamiento de la red neuronal así como sus ventajas y desventajas. Lo siguiente será introducir las herramientas usadas en la concepción de *CARMEN*. Inicialmente, el lenguaje de programación en el cual se desarrolló el algoritmo es *R*, dada la gran variedad de librerías que posee sobre el tema. Para el proceso se usaron las librerías *AMORE* [79], *doMC* [80] y *gregmisc* [81] cada una de ellas con una función específica, por su parte *AMORE* se desarrolló como simulador de redes neuronales con la propiedad de ser flexible en cuanto al diseño de la red. Al ser flexible, permite tener acceso directo a parámetros de la red proporcionando control en los detalles de aprendizaje. Es decir, al utilizar este paquete se pueden cambiar las funciones disponibles según las necesidades del usuario. Dicha librería admite entrenar una red multicapa mediante el algoritmo de descenso por gradiente con retropropagación del error [72]. En sus versiones, tanto adaptativa (adaptative) como por lotes (batch), se pueden incrementar el conjunto de criterios en la cuantificación del error. En cuanto a la función de coste se puede elegir el trabajar entre dos opciones, utilizar funciones disponibles en el paquete o programar nuestras propias funciones en *R*. En esta investigación se trabajará con funciones predefinidas en el paquete siendo todas ellas muy comunes cuando se hace uso de redes neuronales. Los elementos dispuestos para la construcción de *CARMEN* son los siguientes:

- *newff* función diseñada para crear una red neural feed-forward teniendo los siguientes argumentos:

Figura 4.2.2: Banco óptico para obtención de datos (cortesía de la Universidad de Concepción, Chile).



- *n.neurons* vector numérico en el cual se debe introducir la cantidad de neuronas por capa y la cantidad de capas que tendrá la red, teniendo en cuenta que el primer y último elemento de este vector son las capas de entrada y salida respectivamente.
 - *learning.rate.global* y *momentum.global* velocidad de aprendizaje y momento de cada neurona.
 - *error.criterium* permite escoger el criterio con el que se va a comparar el resultado obtenido por la red con el deseado. Se pueden escoger de entre un grupo de métodos predefinidos en el paquete.
 - *hidden.layer* y *output.layer* apartado para designar las funciones de activación con las cuales se van a trabajar en las capas ocultas. Estas funciones se pueden elegir de entre una lista definida en el paquete ó en el caso de *hidden.layer* crear una función dependiendo de las necesidades del usuario.
- función *train*, como su nombre indica, entrena la red neural creada mediante la función *newff*. Su trabajo es modificar los pesos y bias con el fin lograr aproximar los datos ingresados para el entrenamiento a la salida deseada. De entre los valores más representativos tenidos en cuenta al definir esta función en *CARMEN* están: *net* la red neuronal, como ya lo mencionamos en el apartado anterior, está definida por *newff*, datos de entrenamiento tanto entrada como de salida, datos para la validación de la red tanto de entrada como de salida, *error.criterium* al igual que en *newff* se pueden elegir diferentes tipos de error

para el entrenamiento pero por cuestiones practicas se escogió el mismo que el *newff*, y cantidad de iteraciones a realizar y reportar *n.shows* y *show.step*.

- Por último una vez entrenada la red se usa la función *sim* la cual simula la respuesta que se obtendría al pasar por la red los datos que se desean validar.

a grandes rasgos estos son los argumentos relevantes usados en *CARMEN*. Si se desea ver las diferentes formas de empleo así como funciones disponibles en el paquete, argumentos de cada función y en general información detallada concerniente a éste y otros paquete ver³. Las librerías *doMC* y *gregmisc* cada una de ellas fue usada para tareas específicas, por su parte *doMC* contiene una función que paraleliza el entrenamiento de la red permitiendo entrenar a la vez redes neuronales con criterios diferentes logrando disminuir el tiempo empleado en encontrar una configuración apropiada para la red neuronal. Mientras que con el paquete *gregmisc* se puede construir un plano cartesiano en el cual se observan los resultados obtenidos en el entrenamiento y la validación.

4.2.2. Entrenamiento

4.2.2.1. Laboratorio

En el laboratorio *CARMEN* trabaja con una muestra representativa de datos con sus correspondiente entradas y salidas obtenidas de simulaciones hechas en una mesa óptica y partiendo de mediciones ya conocidas. La muestra posee la característica de unir todos los posibles escenarios debido a que si se tienen los suficientes datos independientes, los coeficientes convergen y la red debe ser capaz de trabajar con cualquier entrada similar a los datos de entrenamiento o a una combinación de ellos. Sin embargo, se tuvo especial cuidado con la muestra pues uno de los problemas que pueden ocurrir es que la red solo aprenda a crear conexiones que coincidan con el conjunto de entrenamiento. Es por esto que al usar simulaciones se quiere controlar los datos presentados a la red como guía en el proceso de aprendizaje.

Se realizaron diferentes simulaciones de escenarios para el entrenamiento, siendo el mejor escenario el generado al usar una sola capa de turbulencia donde $r_0 = 0,12$ m y $L_0 = 30$ m. La capa de turbulencia fue desplazada a 155 altitudes distintas que van desde 0 m hasta 15,500 m con 150 m de resolución y para cada altitud *CARMEN* trabajó con 1,000 pantallas de fase generadas aleatoriamente. De esta maneta *CARMEN* puede ver una cantidad considerable de posiciones de la capa para combinar las respuestas de este conjunto base y utilizarla para modelar los datos de entrada. Lo anterior permite a la red trabajar con una variedad de posibles altitudes donde ocurra una turbulencia considerable. No obstante, como se explica en el Capítulo 2 existen perfiles de turbulencia ya definidos donde se pueden observar y tener en cuenta las altitudes en las que ocurren perturbaciones importantes.

³<http://cran.r-project.org/web/packages/>.

Una característica para esta red es que incluso si ha sido entrenada considerando una sola capa de turbulencia ella está capacitada para aprender como lo haría si se le presentaran múltiples capas de turbulencia, esto debido a la combinación de datos en el entrenamiento.

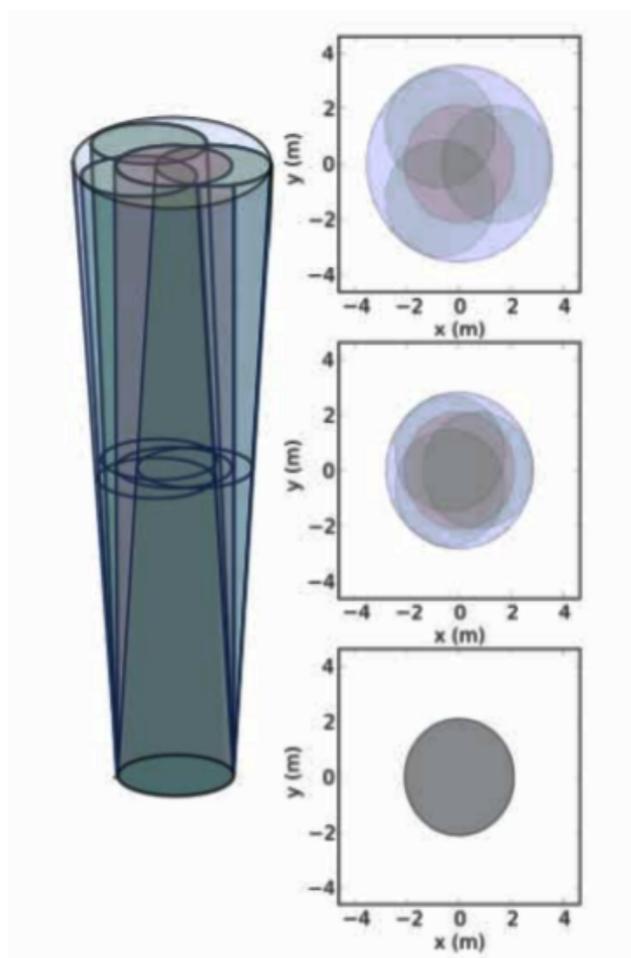
En síntesis, lo que se hace es enseñarle a la red a combinar las capas con diferentes fracciones de solapamiento de los haces de luz en el sensor *WFS* para así modelar la atmósfera con la misma resolución usada para entrenar a *CARMEN* (ver Figura 4.2.3). También al entrenar redes neuronales con datos generados mediante simulaciones simples pero que pueden considerar diferentes condiciones atmosféricas, la red puede combinar toda esta información para predecir salidas de perfiles mucho más complicados sin necesidad de cambiar los valores iniciales, incluso si las condiciones de la turbulencia cambian repentinamente en el momento de hacer las mediciones. Existen otras alternativas en la modelación como son el considerar dos capas de turbulencia, una de ellas con altura fija y la otra variable u otro modelo más realista el cual considera una serie de capas con diferentes fuerzas relativas. Sin embargo, aunque realistas, el conjunto de datos no es independiente y por tanto la red es sobre-entrenada haciendo perder la capacidad de generalizar.

En cuanto a la arquitectura de la red, de la misma forma que con las capas de turbulencia, se varió el número de neuronas en cada capa y la cantidad de capas ocultas de la red neuronal debido a que éstos definen los grados de libertad del sistema, no obstante dichas cantidades varían dependiendo del perfil de turbulencia y el ruido. Este fue uno de los inconvenientes a resolver pues no existe una arquitectura que se considere óptima, dado que ésta cambia respecto a las diferentes configuraciones de la red. Para resolver este problema, se observó qué arquitectura podría predecir los resultados esperados en una mayor cantidad de ocasiones. Después de testar las redes con diferentes arquitecturas y funciones de activación y como lo óptimo es distinto bajo diferentes condiciones, se optó por usar la aproximación más simple la cual consiste de una red perceptrón multicapa (*MLP*) con una capa oculta que contiene la misma cantidad de neuronas de entrada, un algoritmo de retropropagación con función de activación sigmooidal y cuyo valor de aprendizaje es de 0,03. La información detallada de la configuración de *CARMEN* se puede ver en el Cuadro 4.1. Las entradas de la red fueron construidas con 288 nodos que se obtuvieron de 4 sensores de frente de onda cada uno con 7×7 subaperturas reducidas a 36 subaperturas por 2 ejes coordenados. Las salidas fueron 27 que representan los coeficientes de los Zernikes.

Aunque los resultados obtenidos por redes con configuraciones más complicadas no serán estudiados de forma detallada, en el siguiente capítulo se hará un pequeña análisis para algunos de ellos. Pero en general en estos casos, todos ellos tuvieron un rendimiento inferior en diferentes circunstancias. Por ejemplo, para perfiles atmosféricos desfavorables, el error del frente de onda es ligeramente menor. De es-

ta manera y gracias al entrenamiento de redes simples, se pueden cubrir todas las posibles posiciones de las capas, para luego combinar los resultados con el fin de estimar las salidas de perfiles mucho más complicados. Esto es lo que se considera como robustez tomográfica dado que no es necesario información extra o reentrenar la red incluso si cambia drásticamente las condiciones durante la observación.

Figura 4.2.3: Diagrama topológico de los conos de luz de tres estrellas guía para un telescopio que tiene un objetivo de 4,2 m de diámetro. Las estrellas guía están distribuidas equidistantes en un anillo 30 arcsecond de radio.



4.2.2.2. En el cielo (On-sky) (CANARY)

Una vez realizada una gran cantidad de pruebas en el laboratorio el siguiente paso es probar el reconstructor en el cielo. Como se describió en el apartado anterior, *CARMEN* trabaja con una muestra representativa de datos con características básicas pero la diferencia fundamental respecto al trabajo inicial se da gracias a la

Cuadro 4.1: Topología de *CARMEN*

parámetros	valores
Tipo de entrada	Continua
Tipo de salida	Continua
Función de activación	sigmode
Conectividad de la red	Completa
Algoritmo de aprendizaje	BP con Momentum
Tasa de aprendizaje	0,3
Número de capas ocultas	1

naturaleza del cambio. La implementación en tiempo real de *CARMEN*, se hizo en el telescopio *William – Herschel* situado en el observatorio de Roque de los Muchachos en la Palma. Para generar un grupo de datos de entrenamiento se utilizó la calibración del *CANARY* [33] de la siguiente manera: se usaron dos pantallas de fase colocadas lo más cerca posible. Cada una de ellas fue rotada en sentido contrario y a diferentes velocidades angulares, a este proceso se conoce como: mediciones con realizaciones independientes de la fase. Es necesario realizar este proceso pues si no se entrena con suficiente variabilidad en la entrada de la fase, el rendimiento del reconstructor se verá comprometido. Las pantallas de fase fueron colocadas una a nivel del suelo y la otra a alturas variables tomando 10,000 iteraciones del *WFS*. Una vez definidas las velocidades angulares y estando el sistema expuesto a todas las posibles combinaciones en las pantallas de fase, lo siguiente fue pasar las capas por el espacio atmosférico simulado mediante pequeños incrementos. El conjunto de datos obtenido contiene la influencia de las capas de turbulencia para todas las posibles altitudes y está definido como sigue 504 nodos de entradas obtenidos de 3 estrellas guías naturales 4 estrellas guía laser, cada una de ellas con 36 subaperturas por 2 ejes coordenados. Las salidas generadas fueron 72 obtenidas de la cantidad de subaperturas consideradas por 2 ejes coordenados.

En general, *CARMEN* reconstruye las pendientes de los haces de luz de nuestro objeto astronómico independiente del perfil de turbulencia atmosférico. Las entradas del reconstructor serán las inclinaciones fuera del eje desde la estrella guía hasta *WFS* y las salidas son las pendientes predichas que a su vez se convierten en comandos para el *DM* mediante el sensor de matriz de control⁴. El entrenamiento se realizó para el mismo asterismo que se tendrá en el cielo. Por esta razón cada asterismo potencial requiere un reconstructor diferente. Es posible que el reconstructor usado en las estrellas guías laser sea diferente que el implementado en las estrellas guías naturales pues el asterismo de las estrellas guías laser se fija para cada telescopio, pero de la experiencia se puede concluir que un reconstructor simple siempre se puede usar.

⁴Esto proporciona un reconstructor estable incluso en condiciones atmosféricas variables.

Eso en cuanto al entrenamiento antes de realizar las respectivas mediciones en el telescopio *William – Herschel*. Ahora para la implementación de los reconstructores y posterior comparación en los resultados, se debe tener en cuenta lo siguiente: *CANARY* [33] usa *DARC* un controlador *AO* el cual provee control sobre los actuadores a tiempo real en respuesta a las entradas del *WFS* [82]. Este control es modular permitiendo usar diferentes algoritmos, *WFS* y *DMs* para ser integrados con el sistema en tiempo real. La *ANN* fue desarrollada como módulo de reconstrucción, escrito en lenguaje *C*, para *DARC* el cual aprovecha la arquitectura de canalización para disminuir la latencia o retardo temporal en la red, entre los píxeles recibidos por el *WFS* y los comandos enviados al *DM*. Gracias a la carga dinámica *DARC* permite el cambio rápido de algoritmos de control incluso cuando el bucle *AO* es iniciado. De esta manera, se puede comparar la matriz de vectores haciendo uso de los enfoques tradicionales y nuestra *ANN* con muy poco retraso. *DARC* también tiene la habilidad de usar diferentes funciones de activación para diferentes partes de cada vector intermedio pero no se utilizó esta característica aquí.

Una *ANN* puede ser representada por una secuencia de multiplicaciones entre matrices vectoriales intercaladas con la adición del bias y una función lineal así,

$$x_{i+1} = \mathcal{F}_i (M_i * x_i + b_i) \quad (4.2.1)$$

donde x_i es el estado del vector después del i – *ésimo* estado de la *ANN*, M_i es la matriz correspondiente al estado, b_i vector bias y \mathcal{F}_i es la función de activación en el estado i^5 . En el *CANARY* [33] se usaron tres etapas aunque la aplicación a tiempo real permite un número arbitrario de etapas. La primera de ellas es la aplicación de las pendientes sobre el *WFS*, conformada por 504 nodos de entrada para los siete *WFS* que tiene *CANARY* (3 estrellas guías naturales más 4 estrellas guías laser, todas con 7×7 subaperturas lo cual nos da como resultado 36 subaperturas por sensor de frente de onda). Una fase intermedia que será asignada a una capa representando las pendientes como serian vistas en el eje por el *WFS* con 72 mediciones de pendientes. Y finalmente una etapa lineal que será usada para transformar la información de las anteriores etapas a comandos para el *DM*, este proceso se realiza mediante un control de matrices de bucle cerrado y no hace parte del entrenamiento de la *ANN*. Se tiene la opción de usar más etapas intermedias entre las antes mencionadas haciendo uso de la función de activación sigmoide ó simplemente con una función de activación lineal⁶ lo que nos permitiría obtener resultados acerca del rendimiento de la *ANN*. En nuestro caso para realizar las multiplicaciones de matrices de vectores en la etapa intermedia usamos matrices densas⁷ pero para sistemas de orden superior la solución sería trivial haciendo uso

⁵Para más información referase a Sección 4.1 ó Capítulo 2.

⁶Se pueden ver las funciones más representativas en la Subsección 3.2.2

⁷Matrices con una mayoría de entradas diferentes de cero

del álgebra de matrices dispersas⁸.

Otra característica de *DARC* es la habilidad de trabajar con flujo de píxeles en lugar de imágenes. La forma como se procesan ocurre en tiempo real, así a medida que se entregan los píxeles estos van siendo procesados y no es necesario esperar a tener toda la imagen. Esto es fundamental para una baja latencia de entrega y por tanto un mejor rendimiento del *AO*. Al usar el *DARC* estandar las órdenes parciales para el *DM* pueden calcularse una vez se tengan suficientes píxeles en una subapertura determinada. Para el modulo *ANN* esto no es posible ya que para pasar de la primera capa a la siguiente, todas las mediciones de la pendiente deben ser conocidas. Sin embargo la aplicación usada aquí puede iniciar el proceso de trabajar con los píxeles tan pronto como sean recogidos por el sistema sin necesidad de esperar la llegada de toda la imagen. Por tanto en cuanto se tenga los suficientes píxeles en una subapertura dada esta es calibrada y la pendiente calculada.

Así la pendiente medida es usada para realizar una multiplicación parcial con el primer estado de la matriz de *ANN* y mientras más pendientes son medidas la primera salida de la *ANN* es construida. Finalmente, una vez las primeras salidas de la *ANN* están completas son pasadas a los siguientes estados de la *ANN* permitiendo a la primera etapa hacer uso de la corriente de píxeles y dado que se trata de una matriz más grande⁹, se pueda retener la mayor parte del flujo para ser usado en el *DARC*. Es por esto que al estar basado en una multiplicación de matrices de vectores el módulo *ANN* es un buen candidato para la implementación en la unidad de procesamiento gráfico en un sistema *E – ELT*.

⁸Matrices con una gran cantidad de entradas iguales a cero.

⁹Un factor de siete veces mas grande que la segunda matriz en el caso del *CANARY*.

5 Resultados y discusión

En este capítulo se desarrollarán de forma detallada los resultados obtenidos por *CARMEN*. Estos resultados son fruto de un proceso dispendioso y concienzudo que ha tomado 3 años y el cual se puede sintetizar mediante tres fases;

1. Simulaciones en laboratorio.
2. Pruebas en banco óptico.
3. Pruebas On-sky.

5.1. Análisis en laboratorio

Los resultados obtenidos en este ítem fueron generados mediante un simulador atmosférico propiedad de la Universidad de Durham. El simulador se configuró de la siguiente manera: tres estrellas guía fuera de eje espaciadas equitativamente y formando un anillo de 30 arcsec de radio. El objeto astronómico a observar se encontraba en el centro de este anillo. El diámetro del telescopio es de 4,2 m con 7×7 subaperturas en el *WFS*. Los parámetros de la simulación fueron escogidos similares a los de *CANARY* [33] y los resultados comparados con los reconstructores estándares *LS* [77] y *L&A* [78]. Las simulaciones usan un algoritmo de centro de gravedad estándar para el centroide. *CARMEN* se entrenó para proporcionar las seis primeras órdenes para los coeficientes de los Zernikes [31] (sin incluir pistón) en lugar de proporcionar las pendientes en la subaperturas, esto permite reducir el tiempo de entrenamiento para una investigación más eficiente. No obstante, se ha de aclarar que existe la posibilidad de entrenar *CARMEN* para devolver las pendientes.

Para una comparación apropiada, se aplicaron todos los reconstructores a un *DM* modal corrigiendo el mismo número de Zernikes. La fase de los Zernikes reconstruidos se resta a la fase de la pupila y el resultado es empleado para generar la función de dispersión (*PSF*). Los parámetros usados para evaluar estas soluciones son: error de frente de onda (*WFE* [nm]), relación *PSF* Strehl, promedio de azimutabilidad de anchura a la mitad de altura de la *PSF* de la luminosidad de una estrella (*FWHM* [arcsec]) y el diámetro de energía abarcada al 50% (*Ed50* [arcsec]) en la banda *H* (1,650 nm). El *WFE* incluye el error tomográfico y el error de instalación de las seis

órdenes radiales de los Zernikes a la fase real.

Cada uno de los reconstructores de prueba fue entrenado en tres condiciones atmosféricas diferentes, la primera favorable, la segunda intermedia y la tercera desfavorable. Lo anterior se hizo teniendo en cuenta los perfiles atmosféricos de la Palma usados en las simulaciones del *CANARY* mostrados en el Cuadro 5.1. Cada uno de los perfiles tiene cuatro capas turbulentas pero las altitudes, las fuerzas relativas de las capas y la intensidad de la turbulencia integrada en cada caso es diferente. Con el fin de comparar los reconstructores, el *LS* fue optimizado en términos de la altitud del *DM* y la densidad del actuador para experimentación. Por otro lado, la optimización de *L&A* se realizó con una atmósfera con los mismos parámetros empleados en el *LS* pero con diferentes diagramas de fase en cada caso.

Cuadro 5.1: Parámetros atmosféricos.

Parámetros	Valores			Unidades
Nombre del test	atm1	atm2	atm3	
$r_o(0,5\mu m)$	0,16	0,12	0,085	m
Capa1				
Altitud	0	0	0	<i>m</i>
Fuerza relativa	0,65	0,45	0,80	
Velocidad del viento	7,5	7,5	10	<i>m/s</i>
Dirección del viento	0	0	0	<i>C°</i>
Capa2				
Altitud	4000	2500	6500	<i>m</i>
Fuerza relativa	0,15	0,15	0,05	
Velocidad del viento	12,5	12,5	15	<i>m/s</i>
Dirección del viento	330	330	330	<i>C°</i>
Capa3				
Altitud	10000	4000	10000	<i>m</i>
Fuerza relativa	0,10	0,30	0,10	
Velocidad del viento	15	15	17,5	<i>m/s</i>
Dirección del viento	135	135	135	<i>C°</i>
Capa4				
Altitud	15500	13500	15500	<i>m</i>
Fuerza relativa	0,10	0,10	0,05	
Velocidad del viento	20	20	25	<i>m/s</i>
Dirección del viento	240	240	240	<i>C°</i>

Cuadro 5.2: Funciones de dispersión para cada reconstructor y escenario de prueba.

Nombre de la prueba	Reconstructor	Métricas			
		S-R	FWHM	Ed50	WFE
atm1	Sin corregir	0,048	0,319	0,482	644
	<i>LS</i>	0,296	0,099	0,299	293
	<i>L + A</i>	0,402	0,089	0,293	251
	<i>CARMEN</i>	0,462	0,088	0,279	231
atm2	Sin corregir	0,025	0,458	0,633	817
	<i>LS</i>	0,230	0,100	0,443	322
	<i>L + A</i>	0,300	0,091	0,436	289
	<i>CARMEN</i>	0,370	0,088	0,393	262
atm3	Sin corregir	0,012	0,684	0,912	1088
	<i>LS</i>	0,068	0,143	0,690	454
	<i>L + A</i>	0,100	0,104	0,688	409
	<i>CARMEN</i>	0,125	0,101	0,660	387

5.1.1. Resultados de la simulación sin ruido

Los resultados se encuentran sintetizados en la Tabla 5.2¹. A diferencia de los reconstructores *LS* y *L&A* no se hizo ningún cambio en *CARMEN* entre los casos de prueba. Los resultados dejan ver que *CARMEN* es capaz de adaptarse a cada escenario de prueba con éxito, ya que repercute de manera consistente en el *WFE* más bajo. De esta manera, cuanto menor sea el *WFE* mejor la aproximación, también cuanto mayor sea el coeficiente de Strehl ($S - R$) mejor la aproximación. En este sentido en el Cuadro 5.2 se observa que *CARMEN* puede funcionar tan bien como las otras técnicas puesto que los resultados son siempre mejores. Tanto *LS* como *L&A* se pueden optimizar para mejorar aún más los resultados, sin embargo, *CARMEN* no necesita ningún tipo de modificación incluso en condiciones cambiantes. Así, si el tiempo de exposición fuera lo suficientemente largo incluso para promediar los modos de orden inferior y si ejecutáramos cada uno de estos casos de prueba secuencialmente para simular un ambiente cambiante entonces la *PSF* final resulta ser simplemente la suma de las tres *PSF* en los casos de prueba. En conclusión vemos que *CARMEN* podría funcionar en un ambiente cambiante sin una reconfiguración, a diferencia de los otros reconstructores los cuales requieren ser reconfigurados para obtener resultados similares.

En el ítem *a* de la Figura 5.1.1 se muestra la *PSF* generada al usar cada uno de los reconstructores y la situación atmosférica para el caso 2². En el ítem *b* de la

¹Todas las métricas excepto *WFE* son definidas en 1650 nm, además como se puede observar los perfiles atmosféricos de prueba utilizados son similares.

²Datos de prueba caso donde la vista esta catalogada como intermedia

Figura 5.1.1: (a) Simulación de la función de dispersión. (b) Promedios de la azimutabilidad de los perfiles para cada función de dispersión.

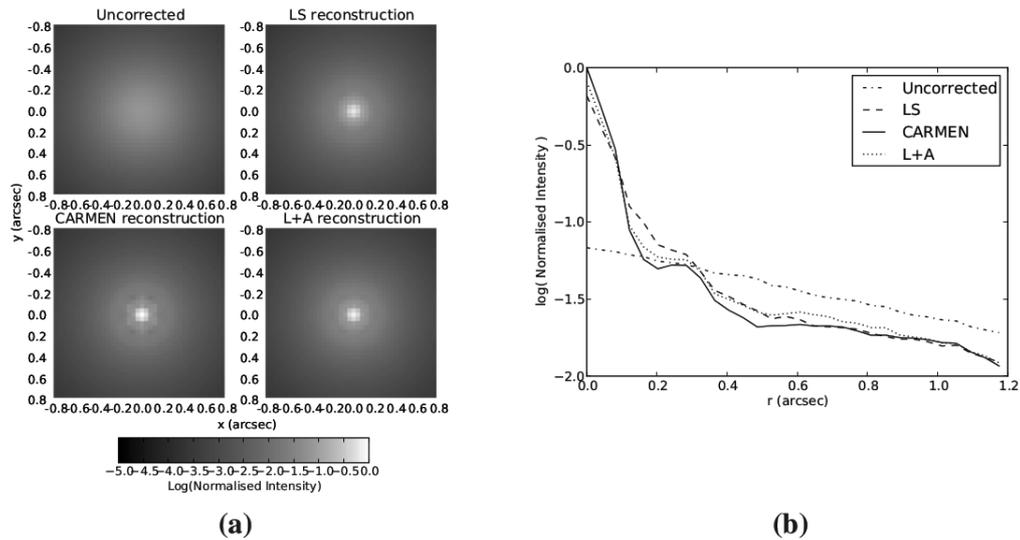
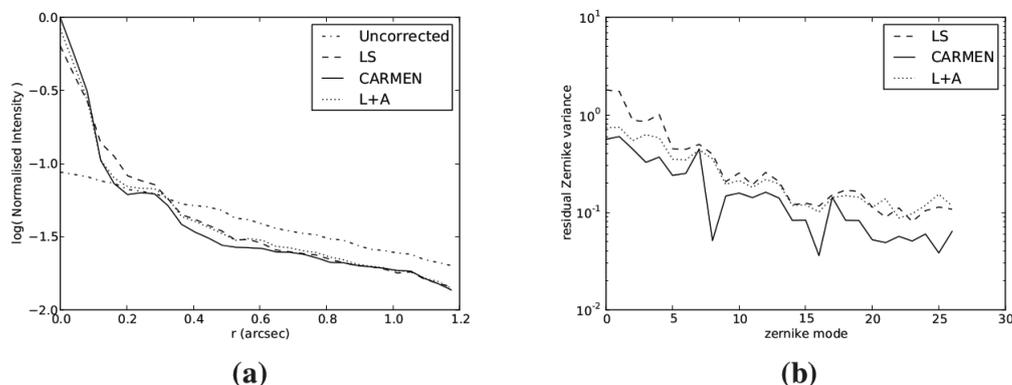


Figura 5.1.1 se muestran los promedios de azimutabilidad radial de los perfiles. El efecto de difracción no circular visto en la *PFS* se da pues se esta aproximando el frente de onda con Zernikes solo hasta el sexto orden.

La Figura 5.1.2 ítem *a* muestra los perfiles radiales azimutalmente promediados para el escenario en el que cada uno de los casos de prueba se ejecutan secuencialmente. Como se menciono anteriormente el *LS* y *L&A* se debieron configurar para cada caso de prueba atmosférica. El *WFE* para *LS*, *L&A* y *CARMEN* son 356 nm, 317 nm y 293 nm lo que corresponde a Strehl de 0,198, 0,265 y 0,319 respectivamente. El ítem *b* de la Figura 5.1.2 muestra la varianza del Zernike residual en el modo base, en ella se puede ver que la varianza residual es menor para *CARMEN* en cada modo.

También se han aplicado perfiles extremos o irreales a *CARMEN* para saber si todavía siguen siendo compensados. Para ello se introdujeron tres casos más de prueba, cada uno con dos capas de turbulencia y una fracción del 50% en la fuerza de turbulencia. Las capas fueron ubicadas una a nivel del suelo y la otra se varió a diferentes niveles. En el Cuadro 5.3 se puede ver que *CARMEN*, para cada una de las altitudes consideraras en esta prueba, obtiene el menor *WFE* y como se mencionó anteriormente, esto es un buen indicador. Lo relevante en estas pruebas radica en que tanto *LS* como *L&A* debieron ser configurados en cada una de las pruebas, a diferencia de *CARMEN* a la cual no fué necesario realizarle alteraciones. Una de las razones por la cual la corrección se reduce con *CARMEN* es debido a la reducción de la fracción de solapamiento de las pupilas en las capas altas.

Figura 5.1.2: (a) Perfiles radiales para una función de dispersión simulada haciendo uso de tres atmósferas movidas secuencialmente para simular cambios. (b) Varianza residual de los zernikes como función del número de modos.



Hasta ahora en los casos de prueba solo se tuvieron en cuenta un número pequeño de capas, lo siguiente fue experimentar con perfiles atmosféricos que contenían una cantidad mayor de capas. En este caso WFE residual para un ambiente conformado por siete capas como se muestra en la Figura 5.1.3 con *CARMEN* era de 328 nm en comparación con el WFE residual sin corregir que era de 818 nm, el valor de r_0 integrado en el sistema fue 0,12 m. Al comparar este resultado con el obtenido para los casos extremos (ver Cuadro 5.3) se puede concluir que en las funciones de red incluso con un gran número de capas de turbulencia no se produce ninguna modificación o entrada adicional.

Las líneas de la Figura 5.1.4 ítem *a* muestra el WFE de cada uno de los reconstructores dependiendo del valor de fuerza de turbulencia integrada r_0 y en el ítem *b* se puede observar el WFE dependiendo de la escala exterior L_0 . Por tanto, aunque *CARMEN* se entrenó para un único valor de fuerza integrada y también para un única escala exterior, para efectos prácticos, puede ser corregido para una amplia gama de valores realistas. Se ha variado el valor de r_0 entre [0,05, 0,25] m y L_0 entre 2 y 100 metros, en términos de diámetro se trabajó en el intervalo $[\frac{D}{2}, 25xD]$ donde D es el diámetro del telescopio y el patrón observado es consistente con las otras observaciones.

5.1.2. Resultados de la simulación con ruido

CARMEN ha sido probada con el ruido del detector simulado (ruido de disparo y el ruido de lectura) en el sensor de frente de onda. Se asumió 100 fotones por subaper-

Cuadro 5.3: Tabla de métricas para los casos extremos.

Reconstructor	Altitud de la capa alta	<i>WFE</i>	Strehl ratio
Sin corregir	5000	767	0,064
<i>LS</i>		293	0,289
<i>L + A</i>		269	0,353
<i>CARMEN</i>		211	0,520
Sin corregir	10000	818	0,025
<i>LS</i>		465	0,066
<i>L + A</i>		372	0,147
<i>CARMEN</i>		297	0,287
Sin corregir	15000	815	0,026
<i>LS</i>		574	0,043
<i>L + A</i>		466	0,069
<i>CARMEN</i>		390	0,127

tura³ de veinte por veinte píxeles por subapertura y 0,2 ruido de lectura de los electrones. Se pueden trabajar dos enfoques para entrenar *CARMEN* para el ruido. Inicialmente ejecutar las medidas del ruido en el *WFS* a través del *CARMEN* entrenada sin ruido o se puede entrenar una nueva *ANN* con pendientes que incluyan el ruido en el centroide. Después de probar las simulaciones se decidió que el último de los casos proporciona una mejor solución. El Cuadro 5.4 muestra las métricas de *PSF* resultantes generadas con los reconstructores utilizando vectores *WFS* con ruido incluido. Se observa que en presencia de ruido de disparo la diferencia entre *CARMEN* y los otros reconstructores se hace aún mayor. Este resultado no es del todo extraño pues las redes neuronales han demostrado ser buenas en los patrones de aprendizaje con datos ruidosos [28]. La red neuronal esencialmente reduce el nivel de prioridad de los modos de orden superior que ahora son indistinguibles del ruido. A diferencia de *CARMEN*, en los otros reconstructores, el ruido no fue incluido cuando se alteró la formación *L&A* y el parámetro acondicionado para maximizar el rendimiento del reconstructor *LS*.

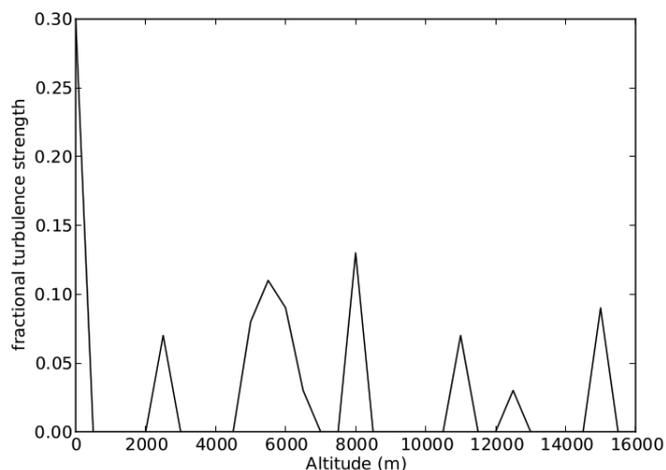
La Figura 5.1.5 (a) muestra los perfiles radiales de las *PSF* con los tres reconstructores tomográficos en el caso 2⁴. El *WFE* residual para la onda no corregida y la de los reconstructores *LS*, *L&A* y *CARMEN* para la prueba con dos atmósferas se puede observar en el Cuadro 5.4. La Figura 5.1.5 (b) muestra la varianza en los coeficientes de los Zernike residuales para cada uno de los reconstructores. En términos matemáticos la varianza se define mediante la siguiente ecuación

$$\sum \frac{(z_{recons} - z_{medido})^2}{n} \quad (5.1.1)$$

³Equivale a una estrella de magnitud 11 y el rendimiento de 50% en un telescopio de 4,2 m.

⁴Vista catalogada como intermedia.

Figura 5.1.3: Perfil turbulento de siete capas arbitrarias utilizadas para probar la habilidad de CARMEN en la reconstrucción al combinar varias capas.



donde Z_{recons} representan los coeficientes de los Zernike reconstruidos, Z_{medido} son el coeficiente de los Zernikes medidos y n es el número de iteraciones de la simulación. Se puede ver que *CARMEN* se ajusta mejor que los otros métodos en los modos de orden inferior y como la mayoría de la energía se concentra en estos modos de ahí viene la ventaja en el rendimiento de *CARMEN*. Sin embargo, con el fin de hacer esto, *CARMEN* debe ser entrenada con un conjunto de datos que contiene la misma magnitud de ruido de disparo.

Hasta ahora todo el entrenamiento se ha hecho fuera de línea en una simulación. Este enfoque tiene la ventaja de permitir seleccionar cuidadosamente los escenarios de entrenamiento para optimizar el rendimiento. Sin embargo, también podría ser beneficioso tener en línea una corrección adicional secundaria que pueda ajustar la salida de *CARMEN* para el perfil de turbulencia óptica real, los parámetros del *WFS*, la configuración óptica (por ejemplo, errores en el registro erróneo), el ruido en el centroide y cualquier otro efecto no incluido en la simulación para observar el comportamiento en un entorno real. Una opción para esta corrección secundaria sería implementar una red neuronal adicional. Al igual que con la técnica de *L&A* se requeriría un sensor real en el eje para entrenar dicha red. Una vez entrenada la red, tendrá en los vectores procedentes de los sensores de frente de onda fuera del eje, la salida inicial predicha por *CARMEN* y la salida de una mejora de la estimación de las aberraciones de fase en el eje. La desventaja es que si no se sintoniza el reconstructor tomográfico al perfil de turbulencia cambiante, entonces se perderá rendimiento al igual que con los otros reconstructores.

Figura 5.1.4: (a) Diagrama de WFE como función de la fuerza de la turbulencia integrada r_0 . (b) y como función de la escala exterior L_0 . Perfil atmosférico para el caso 3 considerado desfavorable.

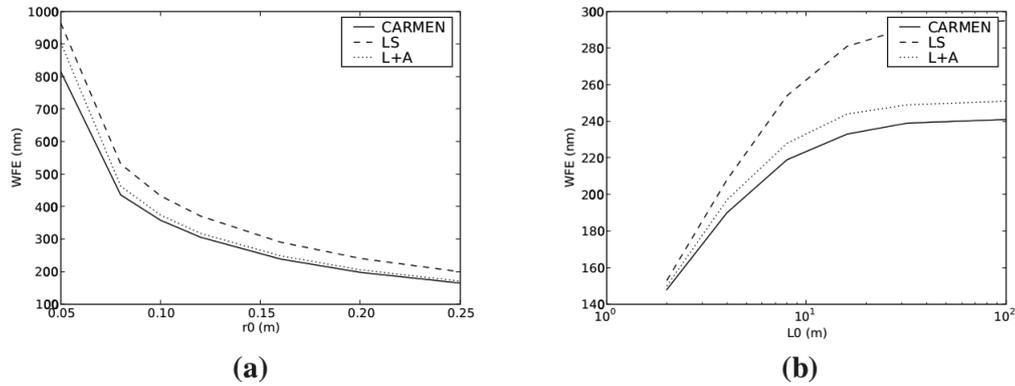
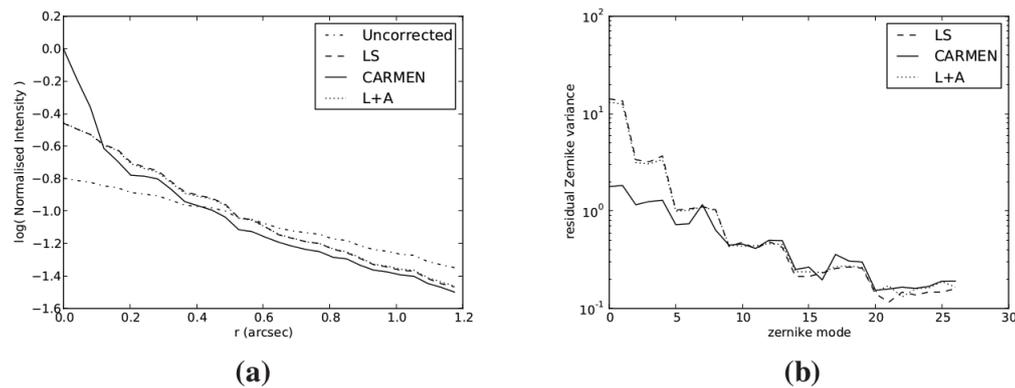


Figura 5.1.5: (a) perfiles del promedio de azimutabilidad radial para la función de dispersión de los reconstructores. (b) varianza residual de los zernikes para los tres reconstructores con ruido de disparo en el WFS .



5.2. Validación en banco óptico

De acuerdo a los resultados presentados hasta ahora, solo se han utilizado las estrellas guía naturales (GS), sin embargo se necesitarán estrellas guía láser (LGS) para aumentar la cobertura en el cielo. Esto reducirá el rendimiento de cualquier reconstructor tomográfico debido a la superposición de las metapupilas causada por el anisoplanatismo focal de los rayos [83]. Aunque no hay problemas con los LGS incluidos en el entrenamiento hay otras cuestiones prácticas que pueden complicar una aplicación en el cielo. Por ejemplo, se debe entrenar la red con el mismo perfil de densidad de la columna de sodio y los efectos fraticidas. Aunque las ANN han demostrado ser robustas, la simulación de entrenamiento debe incorporar todas estas cuestiones para optimizar el rendimiento.

Cuadro 5.4: Funciones de dispersión para cada reconstructor y escenario de prueba, incluyendo ruido.

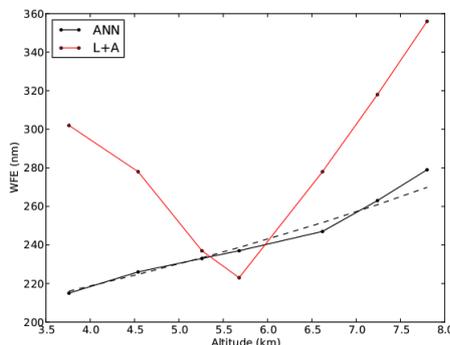
Nombre de la prueba	Reconstructor	Metricas			
		S-R	FWHM	Ed50	WFE
atm1	Sin corregir	0,048	0,319	0,482	643
	<i>LS</i>	0,106	0,187	0,378	451
	<i>L + A</i>	0,113	0,174	0,379	436
	<i>CARMEN</i>	0,274	0,095	0,359	297
atm2	Sin corregir	0,025	0,458	0,633	817
	<i>LS</i>	0,060	0,250	0,476	543
	<i>L + A</i>	0,055	0,254	0,524	547
	<i>CARMEN</i>	0,158	0,105	0,477	368
atm3	Sin corregir	0,012	0,684	0,912	1087
	<i>LS</i>	0,021	0,455	0,771	756
	<i>L + A</i>	0,020	0,455	0,773	751
	<i>CARMEN</i>	0,026	0,333	0,776	594

Para validar *CARMEN*, dentro del banco óptico, se colocaron dos pantallas de fase destinadas a simular dos capas de turbulencia que pueden ser rotadas y ubicadas a distintas alturas relativas a las pupila del telescopio. Una de ellas se colocó a nivel del suelo. Dicha pantalla es fija y tiene un altitud de cero para todos los tests, pues asumimos que la superficie de la capa turbulenta baja está siempre presente [84]. La segunda pantalla de fase fue conjugada para representar una capa de turbulencia alta y fue movida a diferentes altitudes correspondientes al rango de $[H_0 - 2000, H_0 + 2000]$ donde H_0 es la altitud para la cual el reconstructor *L&A* fue optimizado. La medida del parámetro de Fried combinado r_0 que mide la fuerza de la turbulencia atmosférica óptica integrada en ambas pantallas es aproximadamente igual a 0,16 m.

Los resultados para este experimento se ven en Figura 5.1.6. Se puede observar que el rendimiento en *L&A* fue optimizado para una altitud de 5,7 km. Como la segunda pantalla de fase fue desplazada desde esta altitud, el rendimiento con *L&A* se degradó y la curva no es simétrica respecto a la primera pantalla. Este fenomeno se produce debido a que la reducción en la superposición de la pupila proyectada induce errores adicionales al aumentar la altitud. Por otro lado, el rendimiento de *CARMEN* es lineal con el aumento de la altitud, pues como ya se sabe la reducción en el rendimiento se basa únicamente en la reducción de la superposición de la pupila proyectada en una altura. La fracción de área de los discos que simulan las capas de turbulencia, separados por una distancia x en unidades de diámetro es

$$f(x) = \arccos(x) - x(1 - x^2)^{\frac{1}{2}} \quad (5.2.1)$$

Figura 5.1.6: Residuo del WFS para los reconstructores $L\&A$ y $CARMEN$ calibrados en $CANARY$ con la posición de pantalla de fase turbulenta para una altitud dada. La línea discontinua muestra el rendimiento esperado de $CARMEN$ como función de la superposición de la pupila proyectada.



donde $x = \frac{R}{D}$. Por tanto el error residual es siempre proporcional a $\frac{1}{f(x)}$. Así, el rendimiento de $CARMEN$ sigue esta tendencia lo que demuestra que el reconstructor se ha generalizado para llegar a cabo su trabajo independientemente de la entrada de perfil de turbulencia atmosférico.

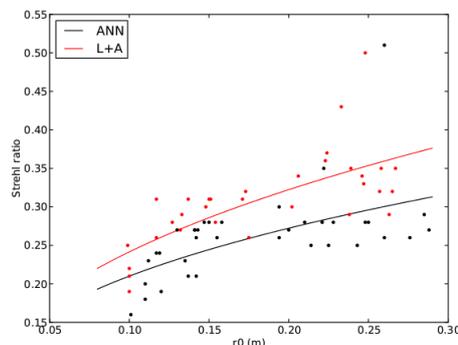
Al comparar el rendimiento de $CARMEN$ y $L\&A$ se observa que el reconstructor completamente optimizado $L\&A$ sobresale respecto a $CARMEN$, sin embargo si la capa alta es cambiada por la altitud de ~ 300 m y el algoritmo $L\&A$ no es reoptimizado entonces los dos reconstructores son iguales. Más allá de este rango de altitud $CARMEN$ supera a $L\&A$. Para un asterismo de 50 pulgadas y 7 subaperturas a través de una pupila de $4,2$ m⁵ este cambio de altitud corresponde a un desplazamiento de una décima de subapertura.

5.3. Validación en el cielo

Durante las noches del 22 y 24 de Julio del 2013 el reconstructor entrenado en banco fue implementado en el cielo. La primera de estas noches se dedicó a calibrar el reconstructor para lograr la rutina óptima en el cálculo de las aberraciones estáticas y la ganancia óptima. $CANARY$ fue operado para explosiones cortas de ~ 30 s y conmutado entre $L\&A$ y $CARMEN$. Esta metodología se utilizó para prevenir el sesgo en los resultados mediante el uso de diferentes reconstructores en distintos momentos durante las condiciones cambiantes. Por falta de tiempo solo se realizaron 36 exposiciones con cada reconstructor y la relación de Strehl [16] fue registrada por la cámara científica del $CANARY$ en la banda H . La Figura 5.1.7 muestra la relación de Strehl [16] obtenida con los dos reconstructores como función de r_0 . r_0 se estima

⁵Parámetros simulados en el banco óptico.

Figura 5.1.7: En el cielo la relación del Strehl de los reconstructores *CARMEN* (negro) y *L&A* (rojo) como función de r_0 . Los reconstructores se entrelazan temporalmente para evitar la desviación debido a las condiciones cambiantes. La línea continua indica los mínimos cuadrados a los datos con un modelo de potencia $y = \alpha x^\beta$. Se puede ver que el resultado de *L&A* es 5 por ciento mejor que el *CARMEN*.

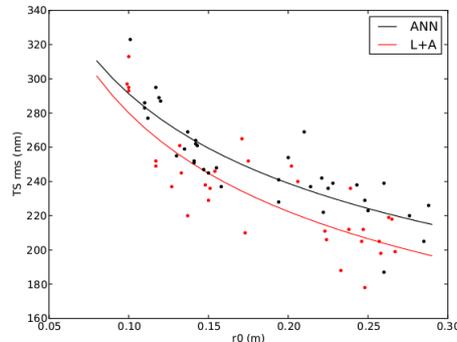


mediante el ajuste de las variaciones teóricas de una descomposición de Zernikes [31] del espectro de potencia de Kolmogorov [13] a los del frente de onda reconstruido a partir de las pendientes del *WFS*. El reconstructor *L&A* logra una relación de Strehl media de $0,31 \pm 0,06$ cuando el error dado es la desviación estandar. Por su parte *CARMEN* logra una proporción Strehl media de $0,26 \pm 0,06$ demostrando que es capaz de alcanzar casi el mismo rendimiento que *L&A* en el cielo. A partir de la desviación estandar de los datos y el examen de la Figura 5.1.7 se observa la existencia de una superposición significativa en los resultados.

Además de la relación de Strehl se pueden analizar los resultados en términos del valor eficaz residual del error del frente de onda a partir de los datos del *WFS* desde el sensor verdadero del *CANARY* (*TS rms*). La Figura 5.1.8 muestra el error residual del sensor de frente de onda de nuevo como una función de r_0 . Un ajuste de mínimos cuadrados de los datos con un modelo de ley de potencia $y = \alpha x^\beta$ indica un deficit medio de aproximadamente 15 nm en el desempeño de *CARMEN* en comparación con *L&A* para los valores de r_0 medidos entre 0,08 m y 0,29 m.

Los resultados obtenidos en el cielo por parte de *CARMEN* no coinciden con los obtenidos por *L&A* optimizado. Esta disparidad refleja las diferencias observada en las pruebas del banco y es causada por el número de pantallas de fase usadas independientemente en el proceso de entrenamiento. Si pudiéramos aumentar la variabilidad de fase en el banco, es de esperar que el rendimiento de *CARMEN* mejore. Esto se da pues si no hay suficiente variabilidad en los datos de entrenamiento el espectro de potencia medido en la fase no converge y por tanto será errónea. Así, dicho error será impreso en *CARMEN* la cual intentará forzar la salida para que coincida con

Figura 5.1.8: Relación en el cielo del TS obtenido con *CARMEN* (black) y *L&A* (rojo) como función de r_0 . Los reconstructores se entrelazan para evitar la desviación debido a las condiciones cambiantes. La línea continua indica los mínimos cuadrados de los datos. Podemos ver que la diferencia media en el error de frente de onda residual de *L&A* es aproximadamente 15 nm menor que el de *CARMEN*.



este espectro de potencia erróneo.

El uso de dos pantallas de fase induce más errores ya que hay una separación definida entre ellas. De ello se sigue que las pantallas de fase serán conjugadas a diferentes altitudes en el banco de calibración. Se estima que las dos pantallas de fase estarán separadas por aproximadamente 500 m⁶. Por tanto, el conjunto de datos formado contendrá información sobre las capas dobles de separación de 500 m. Esto llevará inevitablemente a errores en la reconstrucción de las capas individuales.

Los errores en la altitud de conjugación y en las posiciones laterales de las estrellas guías, tienen como significado que la geometría de los conos de luz en el banco será diferente a los errores en el cielo, lo cual induce errores en las pendiente reconstruidas en el eje. Esto tendrá un efecto en la superposición del haz de luz como función de la altitud y de manera efectiva significa que algunos *WFS* verán las turbulencia más lejos que otros. Ya se demostró en la simulación de [76] que con los datos de entrenamientos suficientes y con errores de alineación insignificantes, el desempeño de *CARMEN* y *L&A* son comparables. La siguiente etapa es el desarrollo de una rutina de entrenamiento que puede producir resultados comparables a *L&A*, mientras se mantiene la generalidad que se da desde el enfoque de una red neuronal.

⁶Unos 5 mm en el banco.

6 Conclusiones

Desde el momento en que se pensó y diseñó *CARMEN* transcurrió un año. En dicho año se realizaron una cantidad considerable de pruebas, algunas de ellas nos enseñaron qué era lo que no se debía hacer, pero todas contribuyeron a alcanzar el objetivo deseado. Una vez diseñado el reconstructor se realizaron diferentes procesos de validación mediante comparación con reconstructores que ya han sido trabajados en investigaciones anteriores y que son aceptados como los reconstructores a usar si se desea abordar un problema con las características particulares de esta investigación. Luego se procedió a realizar varias pruebas de banco todas ellas hechas haciendo uso de bancos ópticos ubicados en la Universidad de Durham y la Pontificia Universidad Católica de Chile. Por último y después de innumerables pruebas el siguiente paso fue realizar todo el proceso anterior en un telescopio. El telescopio disponible para esta investigación fue el *William – Herschel*, situado en el observatorio de Roque de los Muchachos en la Palma. Lo primero fue poner a punto a *CARMEN* en el banco. La idea fue entrenar lo suficiente, pero manteniendo los límites encontrados en las pruebas para no sobrentrenar la red, para que en el momento de proceder a la aplicación en el cielo está fuera capaz de considerar las posibles variaciones de la turbulencia. Acto seguido se implementó *CARMEN* en el telescopio donde se obtuvieron resultados acordes con lo esperado. El proceso de pruebas y aplicación en el cielo, una vez concebida *CARMEN* tomó dos años y se puede sintetizar en las siguientes fases.

6.1. Primera fase

Para abordar el problema en reconstrucción tomográfica se hizo uso un reconstructor basado en *ANN*, el cual se llamó *CARMEN*. Las simulaciones hechas demostrarán la versatilidad del reconstructor y gracias a los trabajos existentes sobre el tema se pudo observar lo novedoso que era. *CARMEN* fue concebido para ser entrenado con un conjunto de datos diseñados con toda la gama de posibles señales de entrada. El conjunto de datos simula una capa turbulenta colocada a diferentes alturas, de esta manera la red se entrena con tantas fracciones de solapamiento como sea posible. Después de probar varios escenarios de entrenamiento y arquitecturas en la red, se encontró que la más simple es la mejor. Para validar los resultados *CARMEN* se comparó con simulaciones realizadas con reconstructores estándar como *LS* [77] y *L&A* [78]. Esto se hizo solo para demostrar que *CARMEN* está en condiciones de

competir con otras técnicas de reconstrucción aceptadas. Aunque es posible optimizar los reconstructores *LS* y *L&A* para obtener una mejor reconstrucción, también sabemos que se puede optimizar *CARMEN* al exponerla a un proceso de aprendizaje más prolongado.

En esta fase se observó que la validez de *CARMEN* es considerable, pues se ha logrado entrenar y aplicar un reconstructor capaz de adaptarse a gran variedad de condiciones atmosféricas. En las simulaciones *CARMEN* fue probada con los casos de prueba utilizados para el proyecto *CANARY* [33], así como, también con tres ambientes más extremos. Para los ambientes extremos se realizaron pruebas con dos capas, con una fracción del 50% en la fuerza de la turbulencia fraccional y tres capas de turbulencia altas diferentes separadas a 5 km. Los resultados proporcionaron información valiosa dado que con *CARMEN* no fue necesario realizar cambios en la configuración incluso si se modificaba drásticamente el espacio. Entre las simulaciones hechas también se probó *CARMEN* con atmósferas que poseían un gran número de capas de fuerza variable fraccionada, lo cual permitió observar su capacidad de corregir exitosamente gran parte de las aberraciones de fase inducida por la turbulencia. Por otro lado, al variar la intensidad de la turbulencia total integrada y la escala exterior dentro de un rango razonable de valores¹, se comprobó que *CARMEN* sigue las mismas tendencias de *LS* y *L&A*. Así, si durante una observación las condiciones atmosféricas cambian repentinamente *CARMEN* tiene la capacidad de aportar reconstrucciones teniendo en cuenta dichas variaciones, a diferencia de *LS* y *L&A* que deben ser reoptimizados para tratar con las nuevas condiciones.

La segunda ventaja de *CARMEN* se da gracias a su capacidad de procesar ruido de disparo en las mediciones de centroides corruptos. A través de simulaciones se ha demostrado que *CARMEN* es capaz de reconstruir los coeficientes de Zernike [31] dentro del eje desde fuentes guía con ruido fuera del eje mejor que *LS* y *L&A*. En el Cuadro 5.4 se observa un ejemplo de ello, ya que al usar un caso de prueba donde la vista es catalogada como mediana, en *CANARY* los reconstructores *LS* y *L&A* tienen un *WFE* residual de 543 nm y 547 nm respectivamente mientras que *CARMEN* obtiene un *WFE* residual de 369 nm. El análisis de la variación de los residuos de Zernikes muestra que la mayoría de estas mejoras provienen de los modos de orden inferior. Por tanto, en esta primera fase se concluye que las *ANN* se pueden utilizar en la reconstrucción tomográfica en *AO* y pueden competir con otros métodos aceptados como óptimos. El siguiente paso fue probar *CARMEN* en una situación más realista. Inicialmente se realizaron pruebas en un laboratorio para después realizar pruebas en el cielo haciendo uso del *CANARY*.

¹ $0,05m < r_0 < 0,25m$ y $D/2 < L_0 < 25D$.

6.2. Segunda y tercera fase

Una vez encontrada la técnica a usar y después de entrenar y comparar *CARMEN* con datos simulados, el siguiente paso fue entrenarla haciendo uso de datos reales. Los resultados de banco mostrarán que el reconstructor funciona con datos reales a pesar de haber sido entrenado solo con datos simulados. Otro resultado positivo se da gracias a que el reconstructor fue capaz de hacer frente a los errores adicionales en el frente de onda. Errores relacionados con los datos de banco y que son distintos a los proporcionados por la turbulencia óptica, en este caso se vinculan a la vibración y el centelleo. También se constató que al usar *CARMEN*, a diferencia de otros métodos, la red necesita ser entrenada con la posición relativa de las *GS* pero no necesita reentrenamiento si el perfil de turbulencia cambia. En este sentido se deberá entrenar la *ANN* por un tiempo considerable antes de llevarse a cabo la observación, gracias a que la única variable a considerar es el perfil de turbulencia en la noche de la observación. Además, como el entrenamiento puede realizarse en una simulación por ordenador no se necesita tiempo de telescopio para dicho fin.

Se ha demostrado que *CARMEN* es de hecho insensible a los cambios en el perfil de turbulencia óptica atmosférica. Esto quedó demostrado en el banco de calibración *CANARY* pues se logró observar que *CARMEN* ofrece una reconstrucción coherente independientemente de la altitud de la turbulencia y sin ninguna información adicional. Existe una caída en el rendimiento en cuanto la altitud de la capa aumenta, lo cual es consistente con la reducción en la superposición de las pupilas proyectadas en las capas altas. También se demostró que al entrenar *CARMEN* en el banco de calibración puede alcanzar resultados comparables a *L&A*. El rendimiento fue ligeramente inferior a la de *L&A* con proporciones medias Strehl [82] de 0,31 para *L&A* y 0,26 para *CARMEN*. La razón de esto se debió a la falta de variabilidad en las pantallas de fase que utilizadas para generar el conjunto de datos de entrenamiento para *CARMEN*. Para maximizar la variabilidad de fase en la formación de datos se usaron dos pantallas de fase colocadas a contra rotación y cerca una de la otra. Esto introdujo aspectos tales como las dos pantallas de fase son realmente desplazados en altura una respecto a la otra. Por último el reconstructor se aplicó en el cielo dos noches, la primera fue usada para poner a punto *CARMEN* y la segunda para la reconstrucción.

7 Líneas futuras de investigación

Una de las características de *CARMEN* es que incluso con las condiciones actuales de hardware, si se entrena el tiempo suficiente, la red puede mejorar de manera considerable. El problema subyace en el volumen de datos a procesar y el tiempo que tomaría en realizar este proceso. Hasta ahora aunque las muestras usadas consideran una gran variedad de posibles casos, el volumen de datos con el cual se ha trabajado es muy inferior al deseado. Es por esto que se han buscado métodos que permitan a *CARMEN* trabajar con volúmenes de datos muy superiores a los trabajados hasta el momento. Uno de los candidatos para realizar esta tarea es la programación en unidades gráficas (*GPU*). Este tipo de programación se ha extendido gracias a que las compañías dedicadas a desarrollar componentes para el procesamiento de la imagen en los ordenadores, lograron construir hardware capaces de llegar a cabo tareas de paralelización mediante software específico. Dicha capacidad, permitiría la optimización de *CARMEN*, superando las limitaciones de máquina en el entrenamiento. En base a la anterior afirmación, se puede garantizar que este método lograría optimizar el rendimiento del reconstructor, en cuanto a tiempo se refiere, en más de un 50% si se hace uso de muestras obtenidas para la primera fase de entrenamiento.

De lo anterior, se puede observar que el siguiente paso es adaptar a *CARMEN* a código *GPU*, en particular el código que se va a implementar es denominado *CUDA* (Arquitectura Unificada de Dispositivos de Cómputo) y su función sería explotar las ventajas que ofrece programar en *GPU* frente a la *CPU* para entrenar redes neuronales con las características de *CARMEN*. *CUDA* es un código propiedad de la empresa *NVIDIA*¹, la cual desarrolla entre otras cosas unidades de procesamiento gráfico, y su nombre hace referencia tanto al uso de un compilador como a un grupo de herramientas para el desarrollo de aplicaciones gracias a una variación del lenguaje de programación *C*. El proceso de adaptación de *CARMEN* está pensado para en un futuro lograr su implementación en el telescopio de campo amplio *E – ELT*.

¹<http://www.nvidia.es/object/cuda-parallel-computing-es.html>.

Bibliografía

- [1] Wilson, R. W, Butterley, T, Sarazin, M, “The Durham/ESO SLODAR optical turbulence profiler”, *Monthly Notices of the Royal Astronomical Society*, 2009 vol. 399 (4) pp. 2129-213.
- [2] Costille, Anne, Petit, Cyril, Conan, Jean-Marc, Fusco, Thierry, Kulcsár, Caroline, Raynaud, Henri- François, “Optimization of MCAO performances: experimental results on ONERA laboratory MCAO bench”, *Adaptive Optics Systems*. Edited by Hubin, *Proceedings SPIE* vol 7015, pp. 113
- [3] Moretti, A, Piotto, G., Arcidiacono, C., Milone, A. P., Ragazzoni, R., Falomo, R., Farinato, J., Bedin, L. R., Anderson, J., Sarajedini, A., Baruffolo, A., Diolaiti, E., Lombini, M., Brast, R., Donaldson, R., Kolb, J., Marchetti E. and Tordo S., “MCAO near-IR photometry of the globular cluster NGC 6388: MAD observations in crowded fields”, *Astronomy and Astrophysics*, vol. 493, pp. 539, 2009
- [4] Boccas, Maxime, Rigaut, François, et al “GeMS: Gemini Mcao System: current status and commissioning plans”, *Adaptive Optics Systems*. Edited by Hubin, *Proceedings SPIE* vol 7015, pp. 22
- [5] Wizinowich P., Le Mignant, D., Bouchez, A. H., Campbell, R. D., Chin, J. C. Y., Contos, A. R., van Dam, M. A., Hartman, S. K., Johansson, E. M., Lafon, R. E., Lewis, H., Stomski, P. J., Summers, D. M., Brown, C. G., Danforth, P. M., Max, C.E., and Pennington D., “The W. M. Keck Observatory laser guide star adaptive optics system: overview,” *Publ. Astron. Soc. Pac.* 118, 297-309 (2006).
- [6] Rigaut, F, “Ground-Conjugate Wide Field Adaptive Optics for the ELTs”, *Beyond conventional adaptive optics: Proceedings of the Topical Meeting*, 2002
- [7] Stepp, L. and Strom, S. “The Thirty-Meter Telescope project design and development phase,” in *Second Backaskog Workshop on Extremely Large Telescopes*, A. L. Ardeberg and T. Andersen, eds. *Proc. SPIE* 5382, 67–75 (2004).
- [8] Sergey Bendersky, Norman S. Kopeika, and Natan Blaunstein. “Atmospheric Optical Turbulence Over Land in Middle East Coastal Environments: Prediction Modeling and Measurements” *Applied Optics*, 2004, Vol. 43, Issue 20, pp. 4070-4079.

-
- [9] Meindl T, Coppenrath E, Degenhart C, Müller-Lisse U, Reiser M. “MDCT urography: experience with bi-phasic excretory phase examination protocol”. *Eur Radiol* 2007; 17 pp 2512-2518.
- [10] Young H, Baum R, Cremerius U, K. Herholz, Hoekstra, O., Lammertsma, A. A., et al.: “Measurement of clinical and subclinical tumour response using [18F]-fluorodeoxyglucose and positron emission tomography: review and 1999 EORTC recommendations”. *European Journal of Cancer*, Vol. 35, Issue 13, 1999.
- [11] Tallon, M, Foy, R, “Adaptive Telescope with Laser Probe: isoplanatism and cone effect”, *Astronomy and Astrophysics*, 235, pp 549-557 (1990).
- [12] Fried, D L, “Statistics of a Geometric Representation of Wavefront Distortion” *Journal of the Optical Society of America*, Nov 1965.
- [13] Kolmogorov, A, “The local structure of turbulence in incompressible viscous fluids for very large Reynolds’ numbers,” in *Turbulence, Classic Papers on Statistical Theory*, S. K. Friedlander and L. Topper, eds. Wiley-Interscience, New York, 1961.
- [14] Fried, D.L. (1994) “Atmospheric Turbulence Optical Effects: Understanding the Adaptive-Optics Implications. Adaptive Optics for Astronomy.” *Proceedings of the NATO Advanced Study Institute on Adaptive Optics for Astronomy*, held in Cargese, Corse, France, June 29-July 9, 1993. Editor(s), D.M. Alloin, J.-M. Mariotti; Publisher, Kluwer Academic, Dordrecht, The Netherlands, Boston, Massachusetts, 1994. LC #: QB84.5 .N38 1993. ISBN #: 0-7923-2748-9, p.25.
- [15] Roddier, F. (1981). “The effects of atmospheric turbulence in optical astronomy”. In *Progress in optics*, North-Holland Publishing Co., Amsterdam, volume 19, pages 281–376.
- [16] Tyson, R. K. (1991). “Principles of Adaptive Optics”. Academic Press.
- [17] Smith, Warren J., “Modern Optical Engineering”, McGraw-Hill Inc., 1966, p. 400
- [18] M. Torrini (ed.), “Giovanni Battista della Porta nell’Europa del suo tempo”, Nápoles, Guida, 1990.
- [19] Rousset, Gerard, Beuzit, Jean-Luc, Hubin, Norbert, Gendron, Eric, Madec, Pierre-Yves, Boyer, Corinne, Gaffard, Jean-Paul, Richard, Jean-Claude, Vitot, M, Gigan, Pierre, Lena, Pierre J, “Performance and results of the COMEON+ adaptive optics system at the ESO 3.6-m telescope”, *Proc. SPIE Vol. 2201*, 1994 pp. 1088.
- [20] Guzmán, D., De Cos Juez, F. J., Myers, R., Guesalaga, A., Sánchez Lasheras, F., and Young, L., “Deformable mirror model for open-loop adaptive optics using multivariate adaptive regression splines”, *Optics express*, Vol. 18, No. 7, March 2010.

- [21] Kokorowski, S. A. (1979). “Analysis of adaptive optical elements made from piezoelectric bimorphs”, *J. Opt. Soc. Am.*, 69, 181-7.
- [22] Guzmán, D., De Cos Juez, F. J., Myers, R., Guesalaga, A. and Sánchez Lasheras, F., “Modeling a MEMS deformable mirror using non-parametric estimation techniques”, *Optics express*, Vol. 18, No. 20, September 2010.
- [23] De Cos Juez, F. J., Sánchez Lasheras, F., Roqueñí, N., and Osborn, J., “An ANN-Based Smart Tomographic Reconstructor in a Dynamic Environment”, *Sensors*, 12(7), 8895-8911 (2012).
- [24] Roddier F. (1988) “Curvature sensing and compensation: A new concept in adaptive optics”. *Appl Opt.* 27, 1223-5.
- [25] Ragazzoni, R.; Farinato, J. (1999). “Sensitivity of a pyramid wavefront sensor in closed loop adaptive optics”. *A&A*, 350, L23-L26.
- [26] Janesick, J. “Scientific Charged-Coupled Devices” SPIE Press, (2001).
- [27] Lardière, O., Conan, R., Clare, R., Bradley C., and Hubin. N., “Performance comparison of centroiding algorithms for laser guide star wavefront sensing with extremely large telescopes”. Escrito 14 Julio 2010 (Doc. ID 124773); publicado 5 August 2010.
- [28] Tamura, S., “An analysis of a noise reduction neural network,” in *Acoustics, Speech, and Signal Processing, 1989. ICASSP-89., 1989 International Conference on*, pp. 2001–2004 vol.3 (1989).
- [29] Thomas, S., Fusco, T., Tokovinin, A., Nicolle, M., Michau, V. and Rousset, G. “Comparison of centroid computation algorithms in a Shack-Hartmann sensor”, *MNRAS* 371, pp. 323-336 (2006)
- [30] Southwell, W. H., (1980). “Wavefront stimulation from wavefront slope measurements”, *J. Opt. Soc. Am.* 70, 998-1006.
- [31] Noll, R.J. (1976). “Zernike polynomials and atmospheric turbulence”, *J. Opt Soc. Am.* 66, 207-211.
- [32] Hammer, F., Sayède, F., Gendron, E., et al , “The FALCON concept: multi-object spectroscopy combined with MCAO in near-IR”, *ESO Workshop Proceedings: Scientific Drivers for ESO Future VLT/VLTI Instrumentation*, arXiv: astro-ph/0109289v1, (2001).
- [33] Guzman, C.D. “Technologies for Astronomical Wide-Field Adaptive Optics”. Centre for Advanced Instrumentation, Department of Physics, Durham University, United Kingdom. 2010.
- [34] Haugeland, John (1985), “Artificial Intelligence: The Very Idea”, Cambridge, Mass: MIT Press, ISBN 0-262-08153-9.
- [35] Kartalopoulos, S., *Understanding Neural Networks and Fuzzy Logic*, IEEE PRESS, NY, 1994.

-
- [36] Yager, R., Filev, D., *Essentials of Fuzzy Modeling and Control*, Wiley Interscience, NY, 1994.
- [37] Kröse, Ben, van der Smagt, Patrick, *An introduction to Neural Networks*, Eighth edition, November 1996.
- [38] Humme, David. "Tratado sobre la naturaleza humana", 1739-1740. "Investigación sobre el entendimiento humano", 1740.
- [39] Berkeley, George. "Tratado sobre los principios del conocimiento humano", 1710. "Tres diálogos entre Hylas y Philonus", 1713.
- [40] Kant, Immanuel. "Crítica a la razón pura". Primera edición 1781. Segunda edición 1787.
- [41] Russell, Bertrand, "Principia Mathematica". Stanford Encyclopedia of Philosophy or University of Michigan Historical Math Collection: Volumen I, Volumen II y Volumen III.
- [42] Wittgenstein, Ludwig, "Philosophical investigations". Prentice Hall, 1999 (ISBN 0-02-428810-1).
- [43] Zadeh, L.A., "Fuzzy sets, Information and Control" 8, 338-353, 1965.
- [44] Zadeh, L.A., "Outline of a new approach to the analysis of complex systems and decision processes", *IEEE Transactions on systems, Man and Cybernetics*, Vol 1, N. 3, 1973, 28-44.
- [45] Mamdani, E.H. y Assilan, S., "An experiment in linguistic synthesis with a fuzzy logic controller", en *Int. J. Man-Machine studies*, N. 7, 1975, 1-13.
- [46] F. L. Smith & Co, "Advanced process control for the cement industry".
- [47] Sivanandam, S.N.; Sumathi, S.; Deepa S.N. *Introduction to Fuzzy Logic using MATLAB*. Ed. Springer, Berlín, New York 2007.
- [48] Sugima, F., Mizushima, S., Kimura, M., Fukuri, Y. y Harada. Y., "A fuzzy approach to the rate control in an artificial cardiac pacemaker regulated by respiratory rate and temperature: A preliminary report", *Journal of Medical Engineering & Technology*, Vol 3, N. 15, 1991, 107-110.
- [49] Bovill, J.M., Sebel, P.S., Blackburn, C.L. and Heykants, J., "The pharmacokinetics of Alfentanil (R39209), a new opioid analgesic", *Anesthesiology*, N. 57, 1982, 439-443.
- [50] Yasunobu, S. y Miyamoto, S., "Automatic train operation by predictive fuzzy control", *Control theory and Advanced technology*, Vol 2, N. 3, 1986, 419-432.
- [51] Davis, L., (ed.) (1991). *Handbook of Genetic Algorithms*, Van Nostrand Reinhold, New York.
- [52] Goldberg D.E., (1989). "Genetic Algorithms in Search, Optimization and Machine Learning", Reading MA: Addison-Wesley.

- [53] Michalewicz Z., (1992). *Genetic Algorithms + Data Structures = Evolution Programs*, Springer-Verlag, Berlin Heidelberg.
- [54] Darwin, Charles, (1859), "On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life", 1st ed. (London: John Murray) 5 (121): 502, Bibcode:1872Natur...5..318B, doi:10.1038/005318a0, recuperado 1 Marzo 2011.
- [55] Holland, John., "Adaptation in Natural and Artificial Systems" (1975). University of Michigan Press. ISBN 0-262-58111-6.
- [56] Bagley, J. D., "The behavior of adaptive systems which employ genetic and correlation algorithms", [Tesis Doctoral] University of Michigan, 1967.
- [57] I. Rechenberg. "Evolutionstrategie: Optimierung Technischer Systeme nach Prinzipien der Biologischen Evolution". Fromman-Holzboog Verlag, Stuttgart, 1973. (German)
- [58] Schwefel, H.-P., "Numerical Optimization of Computer Models". John Wiley & Sons, New-York, 1981. 1995 – 2nd edition.
- [59] Fogel, L. J., Owens, A. J. and Walsh. M. J., "Artificial Intelligence through Simulated Evolution". New York: John Wiley, 1966.
- [60] Cordon, O., Herrera, F., Lozano. M., "A classified review on the combination fuzzy logic-genetic algorithms bibliography: 1989-1995". In E. Sanchez, T. Shibata, and L. Zadeh, editors, *Genetic Algorithms and Fuzzy Logic Systems. Soft Computing Perspectives*. World Scientific, 1999.
- [61] Teuscher, Christof, ed. (2004). "Alan Turing: Life and Legacy of a Great Thinker". Springer-Verlag. ISBN 3-540-20020-7. OCLC 53434737 62339998.
- [62] MacCulloch, W.S., Pitts, W. "A Logical Calculus of the Ideas Immanent in Nervous System". *Recopilado Neurocomputing: Foundations of Research*, pp 18-27, MIT Press, 1987.
- [63] Markram, H., Gerstner W. and Sjöström. P. J., *Frontiers Research Topics "Spike-timing-dependent-plasticity: A Comprehensive Overview"* *Frontiers in Computational neuroscience*.
- [64] Rosenblatt, F. "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain". *Neurocomputing: Foundations of Research*, pp 92-113, MIT Press, 1989.
- [65] Widrow, B., Hoff, M.E. "Adaptive Switching Circuits". *Neurocomputing: Foundations of Research*, pp 126-134, MIT Press, 1989.
- [66] Minsky, M. Papert, S. "Perceptrons". *Recopilado Neurocomputing: Foundations of Research*, pp 161-173, MIT Press, 1989.
- [67] Grossberg, S. "How Does a Brain Build a Cognitive Code". *Recopilado Neurocomputing: Foundations of Research*, pp 349-399, MIT Press, 1989.

-
- [68] Hopfield J.J. “Neurons with Graded Response Have Collective Computational Properties likes Those of Two-State Neurons”. Proc. of the National Academy of Sciences, vol 81, pp 3088-3092, Recopilado Neurocomputing: Foundations of Research, pp 579-583, MIT Press, 1989.
- [69] Kohonen, T. “Self-Organized Formation of Topologically correct Feature Maps”. Recopilado Neurocomputing: Foundations of Research, pp 511-521, MIT press, 1989.
- [70] Fukushima, K., Miyake, S., and Ito, T., “Neocognitron: a neural network model for a mechanism of visual pattern recognition”. IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-13(Nb. 3):pp. 826—834, September/October 1983.
- [71] Kirkpatrick, S., Gelat C.D., and Vecchi. M.P., “Optimization by simulated annealing”. Sciece New series, Vol. 220, N. 4598. May 13, 1983, 61-680.
- [72] Rumelhart, D.E., Hinton, G.E., Williams, R.J. “Learning Internal Representations by Error Propagation” Neurocomputing: Foundations of Reseach, pp 675-695. MIT Press, 1989.
- [73] Haykin, S. (1999) Neural Networks: A Comprehensive Foundation, Prentice Hall, ISBN 0-13-273350-1.
- [74] Wasserman, Philip (1993) Advanced Methods in Neural Computing, Van Nostrand Reinhold, ISBN 0-442-00461-3.
- [75] Gurney, K. (1997) An Introduction to Neural Networks London: Routledge. ISBN 1-85728-673-1 (hardback) or ISBN 1-85728-503-4.
- [76] Osborn, J., De Cos Juez, F. J., Guzman, D., Butterley, T., Myers, R., Gue-salaga, A., and Laine, J., “Using artificial neural networks for open-loop to-mography,” Optics Express 20, 2420 (2012).
- [77] Vidal, F., Gendron, E., and Rousset, G., "Tomography approach for multi-object adaptive optics," JOSA-A, 27, 11, pp. A253-A264 (2010).
- [78] Neichel, Benoit, Fusco, Thierry, Conan, Jean-Marc, “Tomographic recon-struction for wide-field adaptive optics systems: Fourier domain analysis and fun-damental limitations”, Journal of the Optical Society of America A, vol 26, pp 219, 2008
- [79] Castejon Limas, M., Joaquin B. Ordieres Mere, Ana Gonzalez Marcos, Marti-nez de Pison Ascacibar, F. J., Alpha V. P. E., Alba Elias F., and Perez Ramos J. M., (2014). AMORE: A MORE flexible neural network package. R package version 0.2-15. <http://CRAN.R-project.org/package=AMORE>.
- [80] Revolution Analytics (2014). doMC: Foreach parallel adaptor for the multicore package. R package version 1.3.3. <http://CRAN.R-project.org/package=doMC>.

- [81] Gregory R. Warnes. Includes code provided by William Venables and Ben Bolker. Misc Functions written/maintained by Gregory R. Warnes. Gregory R. Warnes “Gregory_R_Warnes@groton.pfizer.com. Greg’s Miscellaneous Functions, package gregmisc.
- [82] Basden, A. G. & Myers, R.M. (2012). The Durham adaptive optics real-time controller: capability and Extremely Large Telescope suitability. *Monthly Notices of the Royal Astronomical Society (MNRAS)* 424(2): 1483-1494.
- [83] Wilson R. W., Jenkins, C. R., “Adaptive Optics for astronomy: theoretical performance and limitations,” *Mon. Not. R. Astron. Soc.* 268, 39–61 (1996).
- [84] Osborn J., Wilson R. W., Butterley T., Shepherd H., Sarazin M. “Profiling the surface layer of optical turbulence with SLODAR” 2010 *MNRAS*, 406, 1405-1408.

