

Algoritmos genéticos aplicados a la diagnosis espectroscópica en plasmas de fusión

D. Suárez*

21 de enero de 2010

*Universidad de Las Palmas de Gran Canaria. dsuarez@gmail.com

Resumen

Este trabajo expone el desarrollo de un método basado en algoritmos genéticos para realizar diagnosis espectroscópica de plasmas de interés en el área de la fusión nuclear por confinamiento inercial

Índice

1. Introducción	4
1.1. Plasmas de fusión nuclear	4
1.2. Diagnósis espectroscópica	5
1.3. Algoritmo genético	6
2. Objetivos	7
3. Métodos	9
3.1. Código de cálculo de propiedades radiativas: ABAKO	9
3.2. Función fitness	10
3.3. Algoritmo genético DIPLA	11
3.4. Aplicación de DIPLA a la diagnósis espectroscópica de plasmas de ICF	16
3.4.1. Análisis de espectros sintéticos	17
3.4.2. Análisis de espectros experimentales: experimentos de implosión en OMEGA	22
4. Resultados y discusión	22
4.0.3. Error relativo medio	22
4.1. Análisis de espectro de emisión sintético	24
4.1.1. Sin ruido	24
4.1.2. Ruido uniforme	24
4.1.3. Ruido gaussiano	24
4.2. Análisis de un espectro de emisión experimental: Implosiones en OMEGA de blancos de deuterio dopados con argón	32
4.2.1. Espectro de emisión de argón 1	32
4.2.2. Espectro de emisión de argón 2	36
5. Conclusiones	39
6. Trabajo futuro	39
6.1. Ampliación a multiobjetivo	39
6.2. Determinación de pesos w_i	39
6.3. Refinamiento	40

1. Introducción

El presente trabajo expone el trabajo realizado en relación a la aplicación de algoritmos genéticos a la diagnosis espectroscópica en el ámbito de los plasmas de fusión por confinamiento inercial (ICF, Inertial Confinement Fusion).

La investigación ha sido realizada en el contexto del trabajo tutelado de los cursos de doctorado del Instituto Universitario de Sistemas Inteligentes y Aplicaciones Numéricas en Ingeniería (SIANI) y en colaboración con el Grupo de investigación en Interacción Radiación-Materia (GIRMA) en la Universidad de Las Palmas de Gran Canaria (ULPGC). Ha sido tutelado por el Dr. Florido¹, el Dr. Winter² y la Dra. González³.

En esta introducción se describen las áreas relacionadas en la investigación: plasmas de fusión, diagnosis espectroscópica y algoritmos genéticos. En el capítulo 2 se describen los objetivos específicos del trabajo y en el capítulo 3 se exponen los métodos y herramientas que se utilizaron para conseguir llevarlos a cabo. El capítulo 4 muestra los resultados de los experimentos y los tests realizados y discute los resultados. Las conclusiones se enuncian en el capítulo 5 y a continuación se hacen unas breves reseñas a líneas de investigación abiertas y que continúan el trabajo realizado y descrito en el presente documento.

1.1. Plasmas de fusión nuclear

El plasma es el llamado cuarto estado de la materia. En la naturaleza se puede encontrar materia en estado de plasma en las estrellas (principalmente formadas por plasma de hidrógeno y helio), las capas superiores de la atmósfera terrestre, el viento solar; o, más cercano, en el fuego. Las dos características fundamentales de la materia en este estado son:

- Está compuesta de partículas neutras (átomos, neutrones) y cargados (iones, electrones).
- La evolución del plasma se ve influida por efectos colectivos debidos al gran número de interacciones de larga distancia.

En la actualidad los plasmas son de interés en diferentes áreas tecnológicas, y en particular, en la relacionada con la energética ya que una de las vías de investigación se centra en la fusión nuclear a partir de materia en estado de plasma, de forma similar a los procesos que ocurren en el interior de las estrellas. En los plasmas de fusión cabe distinguir principalmente dos tipos: fusión por confinamiento inercial (ICF) y por confinamiento magnético (MCF).

¹Departamento de Física, ULPGC

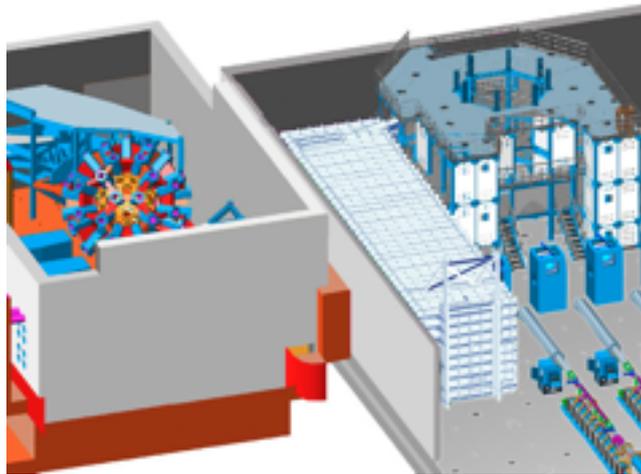
²Instituto Universitario de Sistemas Inteligentes y Aplicaciones Numéricas en Ingeniería, ULPGC

³Instituto Universitario de Sistemas Inteligentes y Aplicaciones Numéricas en Ingeniería, ULPGC

Los plasmas a los que el presente trabajo se encuentra enfocado, están enmarcados en el ámbito de la fusión nuclear por confinamiento inercial. El proceso de confinamiento inercial consigue las condiciones de fusión nuclear mediante el calentamiento y compresión del combustible, normalmente en forma de una cápsula o *microballon* que contiene una mezcla de deuterio y tritio. El método más utilizado para provocar el calentamiento y la compresión es mediante la radiación con láseres. Las condiciones de fusión así conseguidas implican una temperatura alta y una gran densidad en la cápsula de combustible.

En particular, la OMEGA Laser Facility, Rochester, Nueva York permite crear plasmas en estas condiciones radiando el núcleo de combustible nuclear mediante varios haces de luz láser.

Figura 1: OMEGA Laser Facility. Laboratory for Laser Energetics, Rochester, Nueva York.



1.2. Diagnóstico espectroscópico

Uno de los problemas a la hora de estudiar los plasmas de fusión es la dificultad de obtener información sobre magnitudes tales como la temperatura, densidad o presión.

En el caso de la fusión por confinamiento magnético, donde las dimensiones del plasma son macroscópicas, existen diferentes dispositivos para la medida de dichas propiedades. Sin embargo, debido a las altas temperaturas del plasma, los dispositivos de medida sufren y tienen grandes limitaciones.

Sin embargo, el plasma en ICF tiene dimensiones de micras y la disponibilidad de dispositivos de medida en esas condiciones es insuficiente.

Por todo ello, en ambos casos se necesita de un método de medida de dichas propiedades sin interactuar con el plasma. Esto se puede hacer mediante el análisis

de la radiación electromagnética emitida por el plasma, de la cual se puede inferir propiedades tales como la densidad y la temperatura entre otras.

Se denomina diagnóstico espectroscópico al proceso de identificar y reconocer características de un plasma a partir de su espectro de emisión electromagnético. En este caso, el interés se centra en inferir la densidad y temperatura del plasma.

Para realizar la diagnóstico se hace necesario disponer de un método teórico para producir espectros de emisión electromagnética. El método utilizado es mediante un código complejo, debido a la física que se debe considerar en el modelo (física de la estructura atómica, procesos atómicos, estadística y cinética de poblaciones de niveles atómicos, propiedades radiativas).

Su aplicación a procesos de ingeniería en la Tierra está ligada a la obtención de las características del plasma en los reactores de fusión. El conocimiento del comportamiento de los plasmas en estas circunstancias permiten ajustar y optimizar parámetros de los reactores para maximizar la producción de energía o minimizar el consumo del reactor. También, en astrofísica, y mediante la utilización de técnicas de diagnóstico se pueden averiguar características críticas de las estrellas, como su composición química y la abundancia relativa de sus elementos, su velocidad de desplazamiento y la distancia que las separa de nosotros, todo ello mediante el espectro de la luz que nos llega de ellas.

1.3. Algoritmo genético

Goldberg [7] definió los algoritmos genéticos como algoritmos de búsqueda basados en los mecanismos de selección natural y la genética natural. Combinan la supervivencia de los más adaptados con una forma estructurada, pero aleatoria, de intercambio de información. Aunque a continuación se describen las características principales de un algoritmo genético, se recomienda consultar el texto [7], donde los algoritmos genéticos son exhaustivamente explicados.

Un algoritmo genético permite realizar la optimización de funciones complejas, posiblemente no-lineales. La ventaja de los algoritmos genéticos frente a técnicas deterministas son entre otros: 1) evitar mínimos locales, 2) búsqueda de varios mínimos, 3) alta tolerancia a errores y singularidades, 4) multiobjetivo (puede haber varias funciones fitness), y 5) fácil paralelización.

El algoritmo genético optimiza una función, en general a través de su minimización, a la que se denomina función fitness. Es importante señalar que la evaluación de la función de fitness es uno de los puntos críticos en el rendimiento del algoritmo, ya que va a ser evaluada cientos o miles de veces durante el desarrollo del mismo.

Partiendo de un conjunto inicial de soluciones del espacio de búsqueda del problema, el algoritmo genético va explorando el espacio buscando la mejor solución. Para ello modifica las soluciones utilizando operadores de cambio análogos a los encontrados en la evolución natural.

El espacio de búsqueda del algoritmo es el espacio formado por todas las posibles soluciones del problema. Las soluciones que pueblan el espacio de búsqueda se codifican en una forma tratable por el algoritmo. Es muy importante seleccionar correctamente las variables que definen la solución y su forma de codificarlas en el algoritmo, así como los intervalos entre los cuales estas variables se enmarcan. Una técnica utilizada para codificar las soluciones cuando el espacio de búsqueda es muy grande (o continuo) es la discretización del espacio de búsqueda, asignando a un intervalo de valores del espacio de búsqueda un valor entero.

El algoritmo genético funciona iterativamente, donde cada iteración del algoritmo se denomina también generación, por analogía con la evolución natural. Cada generación está poblada por un conjunto de soluciones, que en la primera iteración son generadas al azar. En cada generación se evalúa la función fitness del problema para cada solución. A continuación se utilizan una serie de criterios para seleccionar aquellas soluciones que darán lugar a la siguiente generación utilizando los operadores de mutación y cruce.

El cruce consiste en la selección de dos progenitores y la mezcla de uno o más de sus valores de modo que se genere una solución que tenga características de ambas. Esta operación es análogo al mecanismo de herencia en la naturaleza.

La mutación consiste en modificar uno o más de los valores de un progenitor para generar una solución hija en la siguiente generación. Esta operación existe de forma análoga en la naturaleza.

Al final de cada iteración se comprueba si se ha alcanzado el criterio de parada, y en caso afirmativo, se da por finalizado el proceso de optimización/búsqueda.

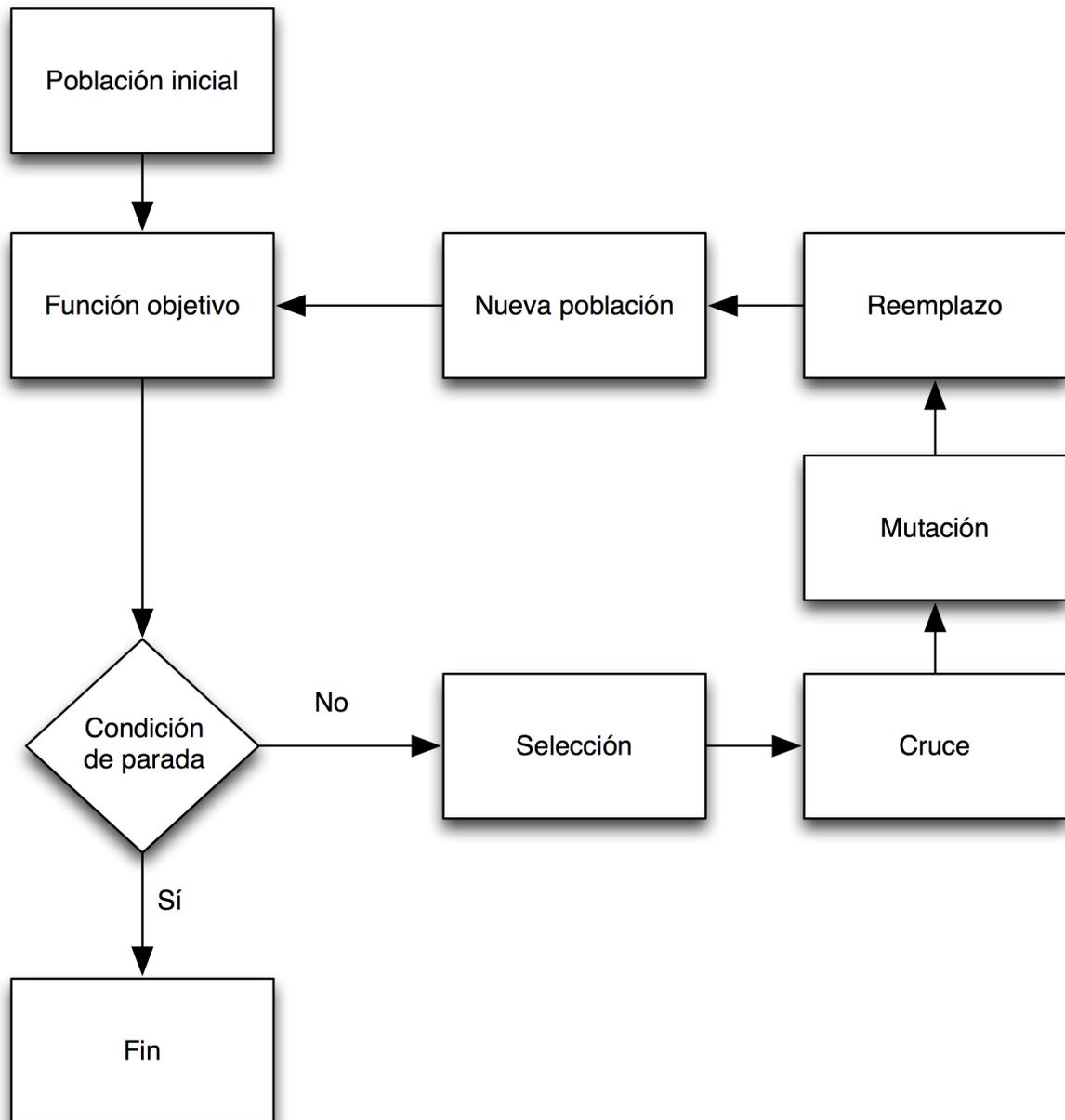
En la figura 2 se muestra el flujo de ejecución de un algoritmo genético genérico. La población inicial es evaluada mediante la función objetivo para posteriormente evaluar la condición de parada; si no se cumple la condición, se efectúan la selección de progenitores de la próxima generación, y se generan nuevos individuos mediante el cruce y la mutación. A continuación se seleccionan los nuevos individuos para formar la nueva generación. La nueva generación se evalúa mediante la función objetivo, tras lo cual vuelve a evaluarse la condición de parada..

2. Objetivos

El problema que se ha solucionado mediante el presente trabajo ha sido la diagnosis espectroscópica de plasmas de fusión. A partir de un espectro de emisión electromagnética de un plasma, supuestamente medido experimentalmente, se desean averiguar las condiciones de temperatura y densidad a las que el plasma se encuentra. El espectro de emisión del plasma a diagnosticar es nuestro espectro experimental.

Para poder lograr este objetivo se hacen necesarias una serie de herramientas

Figura 2: Flujo de ejecución del algoritmo genético



que modelen el comportamiento físico de los plasmas de fusión a partir de una temperatura y densidad dada; de modo que podamos obtener espectros de emisión teóricos.

Se pueden utilizar varias técnicas para lograr este propósito. Una búsqueda exhaustiva en el espacio de soluciones es una opción posible, una vez el espacio ha sido limitado y discretizado. Sin embargo, el tiempo de cómputo para una búsqueda exhaustiva es prohibitivo en un problema real.

En este trabajo, y mediante la aplicación de los algoritmos genéticos para dirigir la búsqueda, se busca disminuir el tiempo de cómputo de la diagnosis y mejorar su precisión mediante la utilización de los algoritmos genéticos. El algoritmo genético será empleado para realizar la búsqueda del espectro teórico que guarde el mejor acuerdo visual con el experimental, de tal modo que la función a minimizar por el algoritmo genético será la suma de las diferencias entre los espectros al cuadrado.

Con este objetivo en mente, se desarrollan un algoritmo genético que hace uso del código ABAKO de GIRMA para el cálculo de espectros de emisión. Utilizando estas herramientas se pretende buscar un conjunto adecuado de parámetros para realizar la diagnosis y garantizar que la solución es única y que puede encontrarse en un tiempo razonable.

3. Métodos

Los elementos necesarios para realizar la diagnosis fueron: 1) un código de cálculo de propiedades radiativas denominado ABAKO y desarrollado por el grupo de investigación GIRMA y 2) un algoritmo genético mono objetivo multimodal elitista. Los experimentos hechos para comprobar el funcionamiento de la diagnosis fueron de dos tipos: sintéticos y reales. Para la evaluación de la función fitness en los experimentos OMEGA se utilizó un código denominado UNIFORM, desarrollado en la Universidad de Reno, Nevada (EEUU).

Se describen a continuación tanto el código ABAKO, utilizado para realizar el cálculo de las propiedades radiativas, como el algoritmo genético desarrollado. Además se describen los experimentos utilizados para validar el algoritmo.

3.1. Código de cálculo de propiedades radiativas: ABAKO

ABAKO [5] es un código desarrollado por el grupo de investigación GIRMA de la ULPGC. ABAKO es capaz de calcular características de plasmas de fusión en un amplio rango de condiciones físicas y generar su espectro de emisión y absorción.

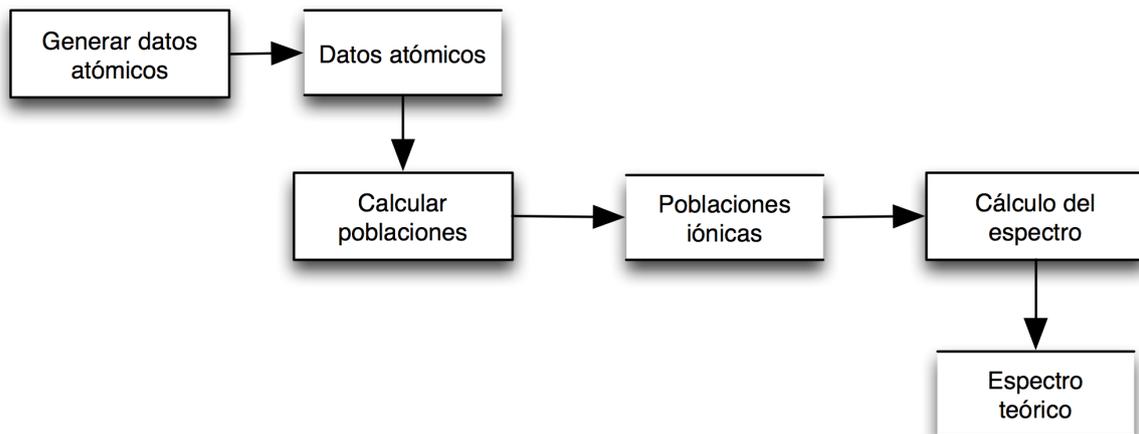
Entre sus principales características se encuentran su versatilidad, ya que es capaz de calcular propiedades de plasmas de muchos tipos y en muy diversas condi-

ciones; así como su rapidez de ejecución.

La velocidad de ejecución es de especial interés en el caso de la diagnosis, ya que será necesario calcular muchos espectros de emisión durante la ejecución del algoritmo genético. Comparado con otros códigos más complejos, ABAKO es muy rápido (varios órdenes de magnitud más rápido). A pesar de ello, los resultados proporcionados por las expresiones analíticas de ABAKO (en contraste con las expresiones mecano-cuánticas de otros códigos) no sacrifican la precisión de los resultados. Las diferencias en valores clave como ionizaciones medias y poblaciones iónicas difieren tan sólo hasta un 5% de códigos como ATOMIC y AVERROES en varios casos señalados [8, 6].

ABAKO permite generar espectros multifrecuenciales de intensidad y opacidad.

Figura 3: Flujo de ejecución de ABAKO



3.2. Función fitness

La función a minimizar por el algoritmo genético se denomina función fitness. Esta función es evaluada cientos de veces durante la ejecución del algoritmo. En el caso de la diagnosis espectroscópica, la función fitness evalúa la diferencia entre el espectro experimental y el teórico.

El valor de la función fitness es el error cuadrático absoluto, también denominada χ^2 . Un valor de la función fitness grande indica poca similitud entre los espectros, con lo que el objetivo del algoritmo genético es minimizar esta función. N es el número de puntos de la malla de energías de fotones y las I representan las intensidades absolutas del punto i -ésimo de la malla. Los superíndices indican si se trata de los valores experimentales (e) o los teóricos (t) generados por ABAKO.

$$\chi^2 = \sum_{i=1}^N w_i (I_i^t - I_i^e)^2$$

Los pesos w_i son coeficientes dependientes de la incerteza para una determinada energía. Para obtener los valores de los pesos pueden utilizarse dos sistemas: 1) Pesos fijos e iguales a 1 (con lo que se calcularía el error cuadrático); y 2) Pesos variables determinados mediante un proceso estadístico. Para calcular el peso w_i distinto a 1 se utiliza la incerteza σ_i de un punto de la malla de energías. Calcular la incerteza no es un problema trivial y por ello, el desarrollo de un método que aproveche sus posibilidades se pospone a un trabajo posterior.

$$w_i = \frac{1}{\sigma_i}$$

El código UNIFORM calcula la función fitness entre un espectro experimental y un conjunto de espectros teóricos y se utilizó para calcular el error en los experimentos de OMEGA.

Adicionalmente, y con motivo de añadir completitud, se probó a utilizar como función fitness el error relativo medio:

$$\bar{E} = \frac{1}{N} \sum_{i=1}^N \frac{|I_i^t - I_i^e|}{I_i^e}$$

cuyos resultados se discuten en 4.0.3.

3.3. Algoritmo genético DIPLA

DIPLA (Diagnosis de Plasma) es un algoritmo genético mono objetivo multimodal utilizado para realizar la diagnosis espectroscópica de plasmas de fusión y desarrollado para hacer uso de las capacidades de cálculo de propiedades radiativas de ABAKO.

El proceso de diagnosis del algoritmo necesita tener acceso al espectro experimental que se quiere diagnosticar. Son necesarios también como parámetros de entrada el elemento químico del plasma a diagnosticar y la malla de energías del espectro, cuyos puntos deben coincidir con los del espectro experimental.

Cada una de las soluciones consiste en una temperatura (T) y una densidad (N_e) que ABAKO empleará como parámetros para generar un espectro teórico que se empleará, junto con el espectro experimental, para calcular el valor de la función fitness.

Los valores de T y N_e de las soluciones se codifican en un valor entero entre 0 y un número natural positivo N_{puntos} . Este número indica el número de puntos en los que se ha discretizado el espacio de búsqueda, entre los intervalos dados de temperatura y densidad. Este número puede ser seleccionado por el usuario. Puesto que los valores de N_e son superiores a 10^7 se utilizan valores logarítmicos de la misma para discretizar el espacio.

Las expresiones utilizadas para calcular la T y N_e que ABAKO recibe como entrada son las siguientes, donde los subíndices min y max reflejan los valores mínimo y máximo del espacio de búsqueda para esa variable y x_T y x_{N_e} son los valores codificados de las variables.

$$T = T_{min} + \frac{T_{max} - T_{min}}{N_{puntos}} x_T$$

$$\log(N_e) = \log(N_{e_{min}}) + \frac{\log(N_{e_{max}}) - \log(N_{e_{min}})}{N_{puntos}} x_{N_e}$$

La función fitness χ^2 para cada solución es una medida del error entre el espectro experimental y el generado por ABAKO para esa solución. De este modo, un valor pequeño de la función de fitness significa que la T y N_e de la solución es cercana a la T y N_e que tenía el plasma que generó el espectro experimental.

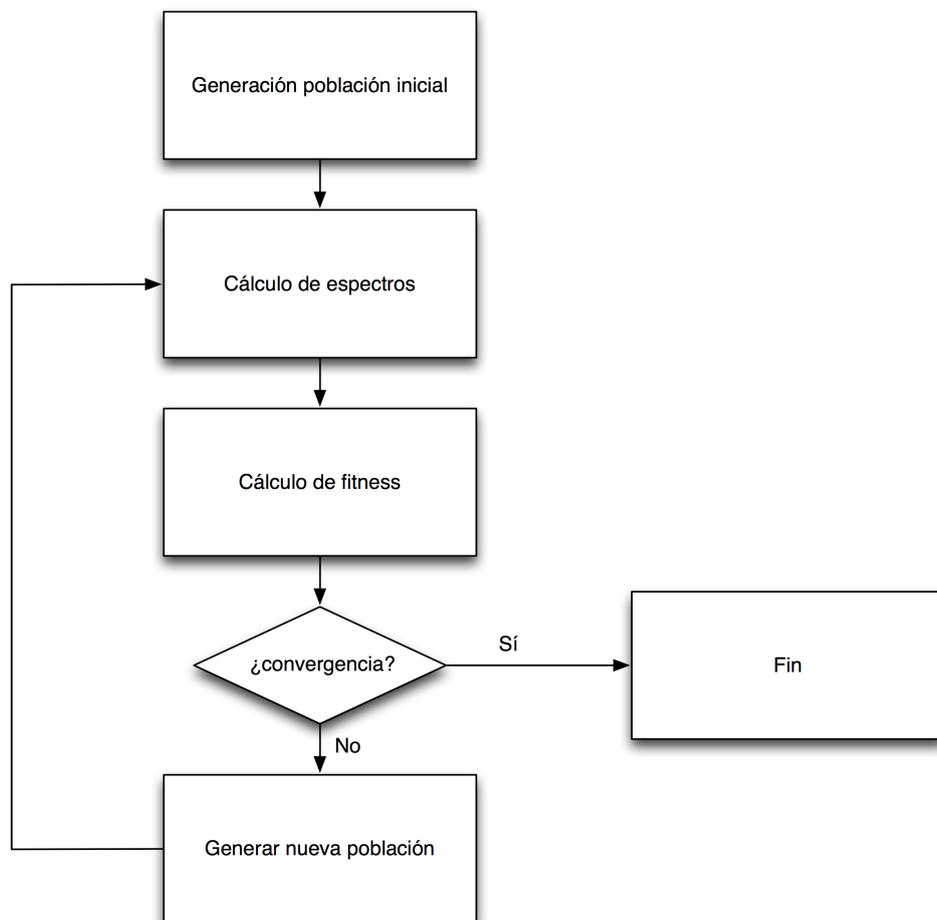
DIPLA puede utilizar elitismo para elegir uno o más elementos de una generación para que pasen a la siguiente generación sin cambios. Esta técnica se utiliza para localizar varios mínimos de la función objetivo, lo que podría significar que la solución no única. En la terminología de algoritmos genéticos se dice que DIPLA es un algoritmo multimodal.

La población de la próxima generación se obtiene mediante cruce y mutación de un conjunto de padres seleccionados de la generación actual. Se seleccionan un número de padres (un padre puede repetirse) igual al número de individuos de la generación. Para seleccionar los padres se utiliza el método del torneo: se eligen un número N_t de soluciones y se escoge como padre aquella que tenga menor valor asociado de la función fitness.

Para realizar el cruce se obtiene un número real aleatorio entre 0 y 1. Si es menor que la probabilidad de cruce, dos padres son seleccionados al azar de la lista de padres y se crea una solución hija donde una de las variables tiene el valor de un padre y la otra variable el valor de la variable del otro padre. La determinación de qué variable corresponde a cada padre se hace al azar al 50%. Una solución así generada es susceptible de ser, además, mutada.

En el caso de la mutación, se obtiene un número real aleatorio entre 0 y 1. Si es menor que la probabilidad de mutación, la solución padre es mutada (o la solución fruto de un cruce). La mutación consiste en añadir un número aleatorio entre -2^{fm} y 2^{fm} , donde fm es el valor de factor de mutación del algoritmo.

Figura 4: Flujo de ejecución de DIPLA



Cuando se hayan generado el número de soluciones para completar una generación, el algoritmo vuelve a la fase de calcular los espectros de estas nuevas soluciones mediante ABAKO y ejecutar las función fitness.

DIPLA detiene su ejecución por varios motivos, de modo que el criterio de parada utilizado es múltiple. Tres motivos pueden desencadenar la parada del algoritmo:

- Se supera el número máximo de iteraciones.
- Se ha superado un determinado umbral de error.
- Se han sucedido tres generaciones sin que el error haya disminuido.

El algoritmo que describe en alto nivel el funcionamiento del algoritmo genético implementado es el siguiente:

1. Generar población inicial aleatoria
2. Inicio de calculo de generación
3. Calcular espectros teóricos de la población mediante ABAKO
4. Calcular función de fitness para la población
5. Comprobar criterios de convergencia (si se cumplen, el algoritmo termina)
6. Crear nueva generación vacía
7. Introducir poblaciones elitistas en la nueva generación
8. Seleccionar progenitores para la siguiente generación
9. Cruzar y mutar progenitores y agregar la nueva solución a la nueva generación (hasta el número máximo de soluciones por generación)
10. Regresar a 2

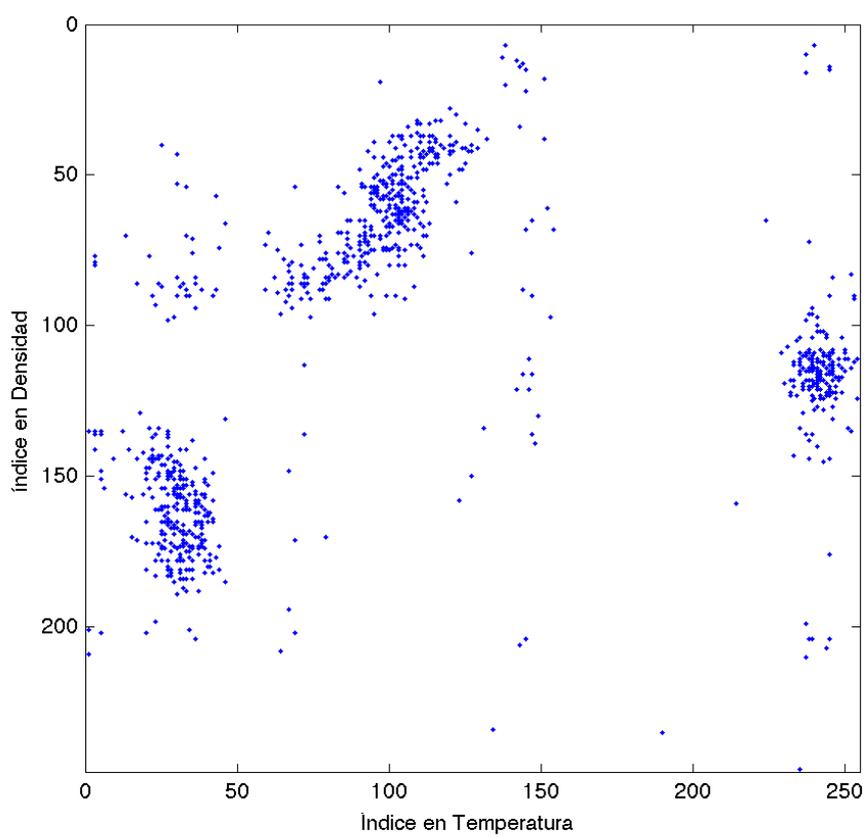
El punto crítico de DIPLA es la generación de los espectros teóricos a partir de las soluciones (T, N_e) . ABAKO es rápido pero aún así la generación de cada espectro consume un tiempo considerable. Debido a esto, se desarrolló un sistema de almacenamiento de espectros en una base de datos mediante la interfaz *shelve* que implementa persistencia de objetos en Python [1]. Una base de datos así generada sirve para un elemento químico dado y unos valores de discretización del espacio de búsqueda.

Un ejemplo de una base de datos para una serie de experimentos sintéticos puede verse en la Figura 5, donde cada punto representa una condición de T y N_e para la cual el espectro ha sido calculado y almacenado.

Esto mejora enormemente la velocidad de la diagnosis ya que, una vez que la base de datos es suficientemente grande, puede utilizarse para generar las poblaciones iniciales dando un valor de distancia euclídea mínima entre soluciones.

DIPLA es altamente parametrizable, admitiendo cualquiera de las siguientes opciones:

Figura 5: Base de datos de espectros sintéticos



- Número de soluciones por generación, es el tamaño de la población de cada iteración del algoritmo genético.
- Número máximo de generaciones, límite al número de iteraciones del algoritmo genético
- Número de torneo, tamaño del conjunto de posibles padres del cual, en una determinada iteración, dos son elegidos para generar una nueva solución hija.
- Número de procesos, cantidad de hilos de ejecución que correrá la máquina (para proceso en paralelo).
- Temperatura mínima y máxima, límites al espacio de búsqueda del algoritmo en temperatura.
- Densidad mínima y máxima, límites al espacio de búsqueda del algoritmo en densidad.
- Número de bits de temperatura y densidad, número de bits utilizados para discretizar el espacio de búsqueda, para cada una de las dos variables: temperatura y densidad.
- Elitismo, número de soluciones de una iteración que pasan sin cambios a la siguiente, y distancia mínima que debe haber entre ellas para considerarlas soluciones separadas. Esto permite al algoritmo el comportamiento multimodal.
- Probabilidad de mutación, porcentaje de que una solución generada para una nueva iteración mute.
- Probabilidad de cruce, porcentaje de que una solución sea generada mediante cruce.
- Factor de mutación, que indica que en una mutación el valor de la variable puede cambiar entre -2^{f_m} a 2^{f_m} , donde f_m es el valor del factor de mutación.
- Base de datos, fichero que contiene una base de datos de espectros previamente calculados.
- Generar primera generación de la base de datos, capacidad de obtener todas las soluciones de la primera iteración desde la base de datos, especificando una distancia mínima entre soluciones, para cubrir la mayor parte del espacio de búsqueda.

3.4. Aplicación de DIPLA a la diagnosis espectroscópica de plasmas de ICF

Los tests realizados para comprobar la robustez del algoritmo genético fueron realizados primero, contra objetivos sintéticos y a continuación, contra objetivos reales cuyos valores ya fueron diagnosticados mediante otra técnica.

Puesto que, como se dijo en la discusión de la sección 3.2 no es posible definir un valor para nuestra función objetivo que sea general para cualquier diagnosis, se

optó por tomar como valor límite en los experimentos sintéticos un error relativo medio igual a 0.5 % (que en ningún caso se alcanzó, posiblemente debido al grosor de la malla). En los experimentos de OMEGA, el criterio de parada fue únicamente una repetición del error cuadrático medio.

No se utilizó la característica multimodal, ya que en las zonas espectrales que se trataron no se pudo aseverar la existencia varias soluciones de temperatura y densidad para un mismo espectro.

3.4.1. Análisis de espectros sintéticos

Se han elegido plasmas de Carbono para realizar los experimentos sintético debido a su importancia en el contexto de plasmas de fusión. El Carbono es un elemento presente en los microballoons de combustible en los plasmas de confinamiento inercial y también como impureza común en confinamiento magnético. Su bajo número atómico permite estudiarlo en un tiempo abordable.

A continuación se muestran los espectros multifrecuenciales de emisión generados por ABAKO de los plasmas de Carbono. En general, los parámetros críticos para calcular un espectro son:

- Elemento químico
- Temperatura del plasma T
- Densidad del plasma N_e
- Rango espectral sobre el que se calcula el espectro de emisión

Cuadro 1: Tabla descriptiva de experimentos

Elemento químico	Condiciones	Intervalo de energías
Carbono	80 eV, $10^{20} cm^{-3}$	250-500 eV
Carbono	70 eV, $10^{21} cm^{-3}$	250-500 eV

Aunque son similares en cuanto a forma, es conveniente hacer notar la diferencia en órdenes de magnitud de los espectros, siendo el segundo aproximadamente 2 órdenes de magnitud más intenso que el primero.

Se definieron dos pares de valores de T y N_e experimentales, cada uno de los cuales determina un espectro sintético. El espectro a diagnosticar es el de intensidad, que en el caso de los plasmas tratados, coincide con la emisividad del plasma. Se realizaron dos diagnosis de cada experimento sintético para comprobar su robustez.

El rango de energía de la malla de fotones fue elegido entre 250 y 500 eV por un motivo especial: este rango de energías corresponde a las emisiones del Carbono

Figura 6: Carbono, espectro de emisión 80 eV, 10^{20}cm^{-3} , 250-500 eV

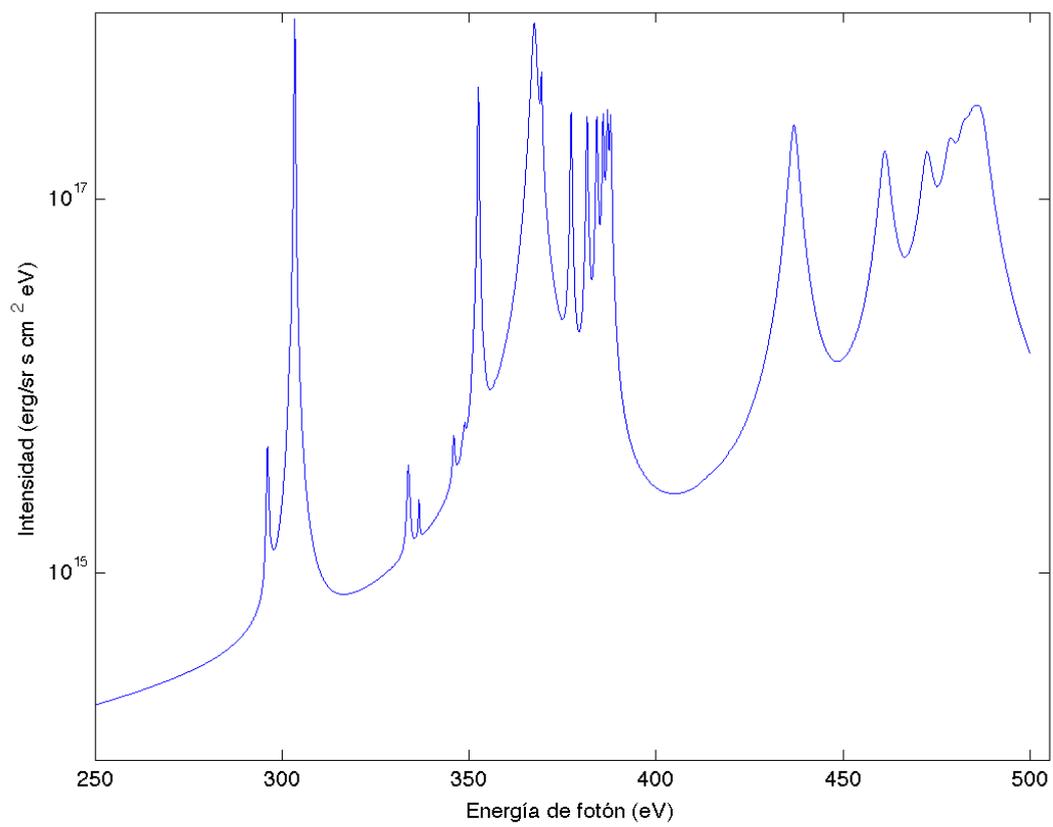
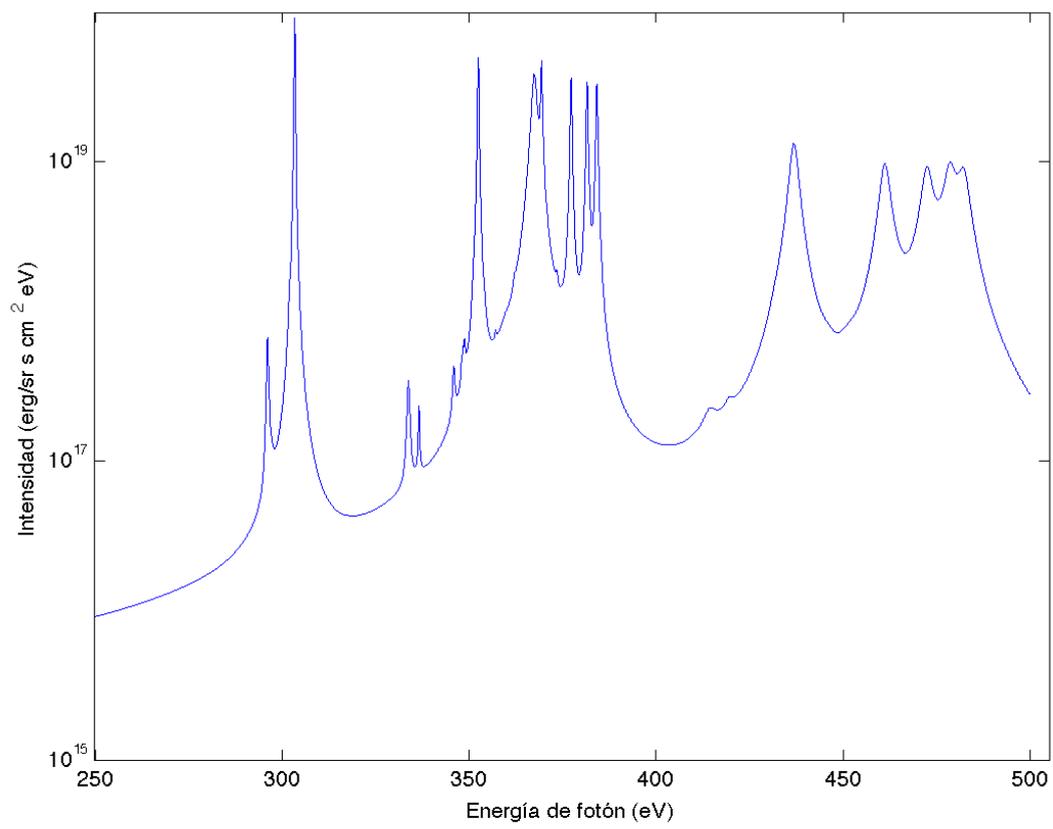


Figura 7: Carbono, espectro de emisión 70 eV, 10^{21}cm^{-3} , 250-500 eV



K-Shell. La diagnosis de las líneas espectrales en este rango de energías permite obtener las temperaturas y densidades electrónicas de los plasmas que contienen el Carbono.

Para simular la intensidad observada experimentalmente, la intensidad obtenida a partir del modelo teórico suele convolucionarse con función gaussiana de cierta anchura, que representa el posible ensanchamiento de las líneas espectrales como g . Sin embargo, tras las pruebas realizadas, esta convolución no producía un ensanchamiento significativo en el rango espectral estudiado para las resoluciones experimentales típicas entre 3 y 9 eV.

$$I^{conv}(h\nu_0) = \int g(h\nu - h\nu_0)I(h\nu)d(h\nu)$$

El espacio de búsqueda comprendía temperaturas entre 5 y 200 eV y densidades entre 10^{18} y 10^{23} electrones por centímetro cúbico. Para discretizar el espacio de búsqueda se emplearon 256 posibles combinaciones. La distancia entre los puntos de la malla en el intervalo elegido fue de 0.76 eV en temperatura y de densidad es de 0.019 en el exponente de la densidad, que indica una variación estable de aproximadamente 1 en todos los exponentes.

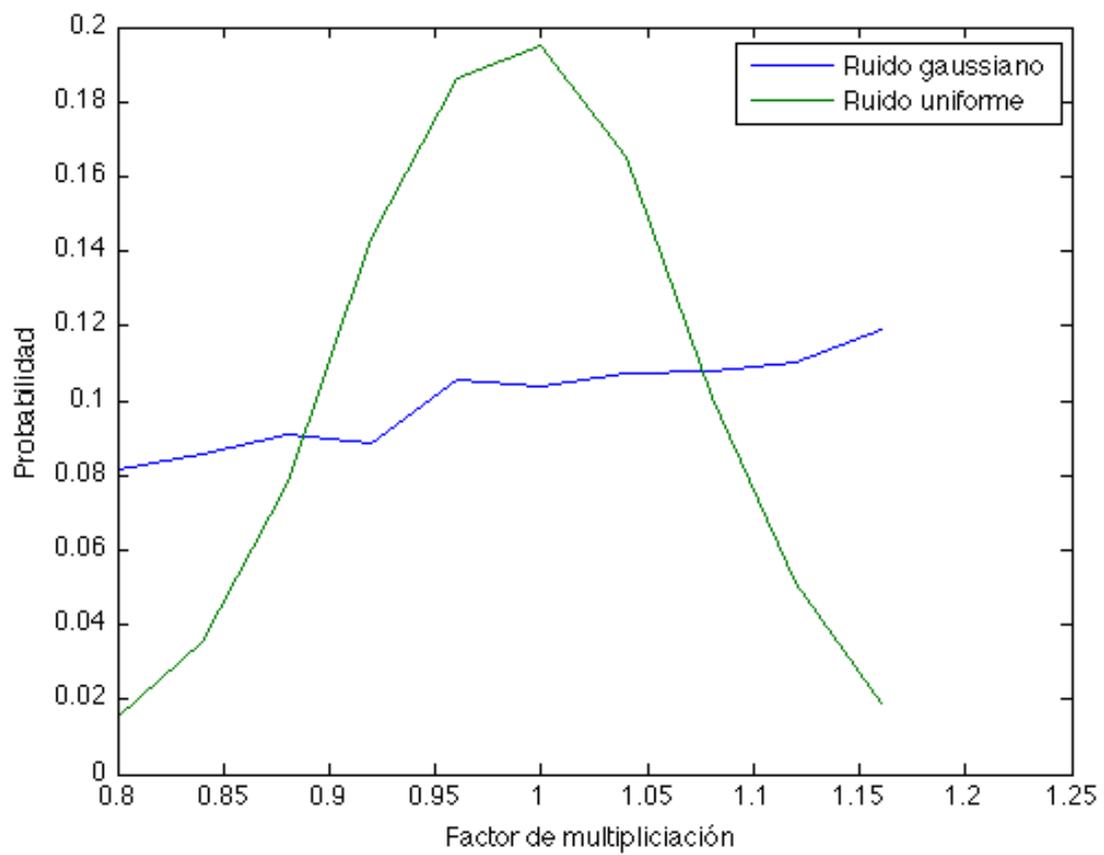
Tras una serie de pruebas, se determinó que los parámetros que daban mejor resultado fueron 0.5 para la probabilidad de mutación y cruce, un factor de mutación igual a 3, 32 individuos por generación y un máximo de 32 generaciones.

Análisis de experimentos sintéticos con ruido uniforme y gaussiano Los espectros de intensidad fueron alterados mediante un ruido aleatorio uniforme de entre el 80 % y el 120 % del valor de la intensidad de ese punto de la malla de energía de fotones. A continuación se utilizó DIPLA para diagnosticar ese espectro sintético.

Para el ruido gaussiano se procedió del mismo modo que para el caso anterior, pero la distribución de ruido aleatorio es gaussiana. Se multiplicó el valor de la intensidad para una determinada energía de fotón por el resultado de un valor aleatorio según la distribución normal gaussiana de media $\mu = 0$ y de sigma $\sigma = 0,2$. La ecuación siguiente muestra la ecuación de la gaussiana. La figura 8 muestra el resultado de simular las distribuciones aleatorias.

$$g(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Figura 8: Distribución de probabilidad aleatoria de ruido



3.4.2. Análisis de espectros experimentales: experimentos de implosión en OMEGA

Los experimentos diagnosticados fueron producidos en la OMEGA Laser Facility del Laboratory for Laser Energetics en Rochester, Nueva York [4, 3]. Los plasmas fueron generados mediante la implosión de blancos esféricos de fusión de aproximadamente 800 micras de diámetro inicial, conteniendo deuterio gaseoso y una pequeña cantidad de argón que hace las funciones de trazador espectroscópico.

Los experimentos de OMEGA vienen dados por láminas fotográficas que muestran la evolución temporal de la cantidad de fotones (intensidad de la radiación) emitidos por el plasma para cada frecuencia. La diagnosis espectroscópica se realiza sobre un instante particular del shot representado en la figura 9, extraída de [3] mediante el cuadro rojo. Se propone la diagnosis de dos espectros distintos, correspondientes a dos instantes de tiempo distintos en la gráfica de evolución del plasma.

En las diagnosis OMEGA sólo se utilizó como criterio de parada la repetición del valor de la función fitness.

Con intención de comprobar la robustez del algoritmo se realizaron 10 ejecuciones de cada test para el espectro 1. Los tests fueron hechos con distinto número de individuos por generación para comprobar la variación en la convergencia, el tiempo de ejecución y el número de generaciones.

La diagnosis del espectro 2 fue utilizada para realizar una metabúsqueda en la que se intentaron averiguar los mejores valores de probabilidad de mutación y probabilidad cruce para esta serie de diagnosis en concreto. El espectro 2 fue diagnosticado únicamente 5 veces con un número de individuos de 16. Se realizó una evaluación del cambio en el tiempo de convergencia (medido en número de iteraciones y valorando el error cuadrático medio final) para distintos valores de la variable de probabilidad de mutación y de la probabilidad de cruce, utilizando los valores: 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 y 0.9 para ambos y contemplando todas las posibles combinaciones, de modo que puede considerarse una búsqueda por fuerza bruta.

4. Resultados y discusión

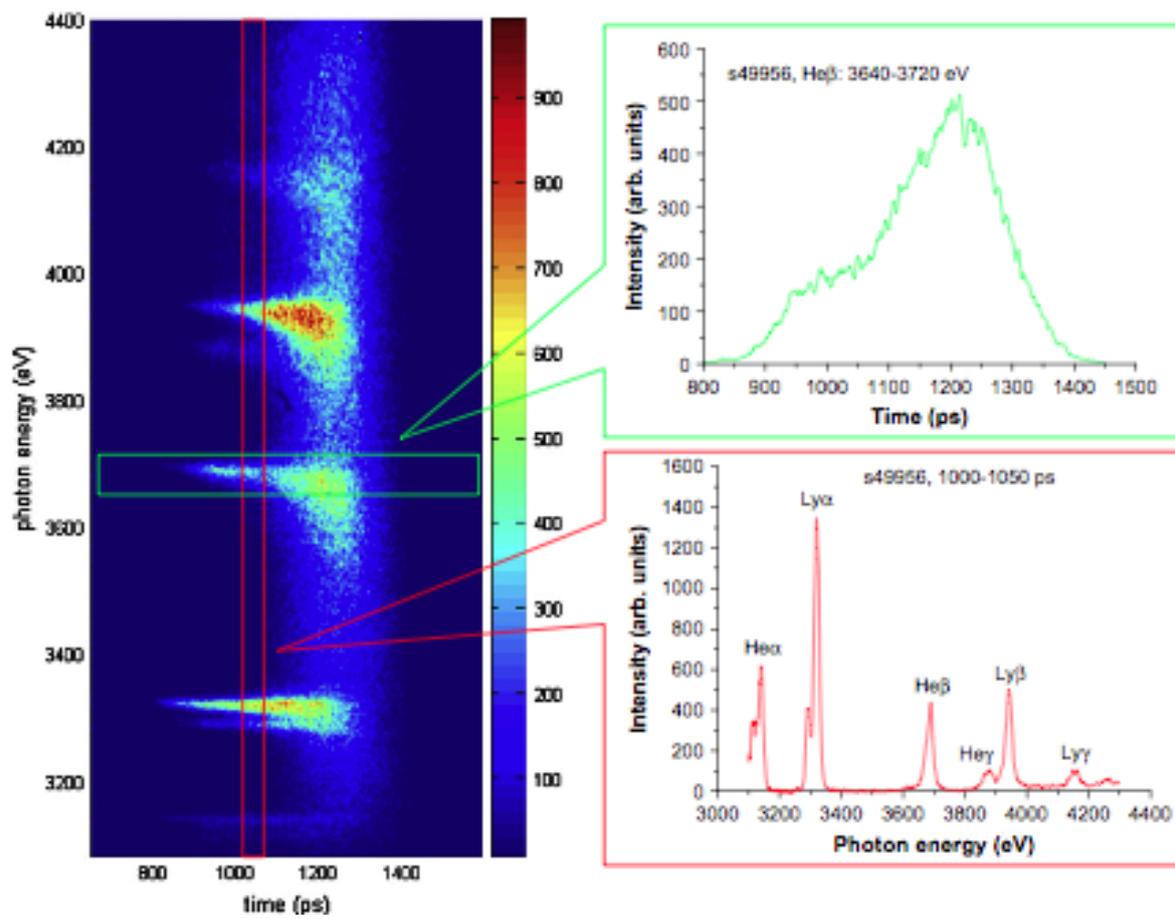
4.0.3. Error relativo medio

El error relativo medio nos proporciona una buena estimación media de la similitud de los espectros, aunque ha demostrado no ser monotónicamente descendente durante los experimentos.

De este modo, mientras que con la función fitness χ^2 se penaliza fuertemente a los espectros con valores de intensidades mayores que uno, con la función de error

Figura 9: Extracto de [3], bajo autorización del autor

R. Florido et al. / High Energy Density Physics 6 (2010) 70–75



medio relativo la clave es la diferencia relativa, sin importar el orden de magnitud absoluto de la diferencia.

Se encontró que el error relativo medio era interesante una vez la diagnosis había terminado y como una manera de obtener un valor cuantitativo de la similaridad entre los espectros, de modo que mediante esta magnitud es posible comparar dos procesos de diagnosis completamente distintos.

4.1. Análisis de espectro de emisión sintético

4.1.1. Sin ruido

La tabla 2 muestra el resumen de los resultados de los experimentos virtuales realizados. El error relativo medio se mantiene bajo, en torno al 5%, asegurando así un buen nivel de similitud entre los espectros. Las figuras 10 y 11 muestran la evolución del espectro conforme el algoritmo genético va aproximándose a la solución.

Cuadro 2: Resultados diagnosis de espectro sintético sin ruido

T y N_e	Diagnosis T y N_e	Error en T y N_e	\bar{E}
80 eV, $10^{20} cm^{-3}$	81.9 eV, $10^{20} cm^{-3}$	2.3 %, 0 %	5.7 %
70 eV, $10^{21} cm^{-3}$	69.27 eV, $1,11 \times 10^{20} cm^{-3}$	1 %, 1.1 %	4.7 %

A continuación se muestran unas figuras que muestran como evolucionó el mejor espectro a lo largo de las generaciones del algoritmo genético. Se muestra el error absoluto medio a lo largo de las generaciones y los espectros que mejor ajustan para la primera generación, la generación en la mitad del proceso y la última generación.

4.1.2. Ruido uniforme

Estos resultados ilustran de cierto modo la robustez del método de diagnosis. A pesar de haber variado el espectro mediante un ruido aleatorio, los errores relativos de temperatura y densidad se pueden considerar bajos. La evolución del espectro puede verse en las figuras 12 y 13.

4.1.3. Ruido gaussiano

Al igual que en los experimentos anteriores, los resultados en este caso proporcionan un buen acuerdo con la temperatura y densidad electrónicas. Sin embargo,

Cuadro 3: Resultados diagnosis de espectro sintético con ruido uniforme

T y N_e	Diagnosis T y N_e	Error en T y N_e	\bar{E}
80 eV, $10^{20} cm^{-3}$	81.4 eV, $1,05 \times 10^{20} cm^{-3}$	1.75 %, 5 %	8.8 %
70 eV, $10^{21} cm^{-3}$	69.56 eV, $1,07 \times 10^{20} cm^{-3}$	0.6 %, 7 %	9.2 %

el error relativo medio en la intensidad de los espectros es alta, debido a las fluctuaciones introducidas por el ruido gaussiano. Las figuras 14 y 15 muestran así, la evolución del espectro.

Cuadro 4: Resultados diagnosis de espectro sintético con ruido gaussiano

T y N_e	Diagnosis T y N_e	Error en T y N_e	\bar{E}
80 eV, $10^{20} cm^{-3}$	80.41 eV, $9,64 \times 10^{20} cm^{-3}$	0.5 %, 3.6 %	14.8 %
70 eV, $10^{21} cm^{-3}$	69.46 eV, $1,02 \times 10^{20} cm^{-3}$	0.8 %, 2 %	14.5 %

Figura 10: Diagnósis 80 eV, 10^{20} cm^{-3} , 250-500 eV. Sin ruido

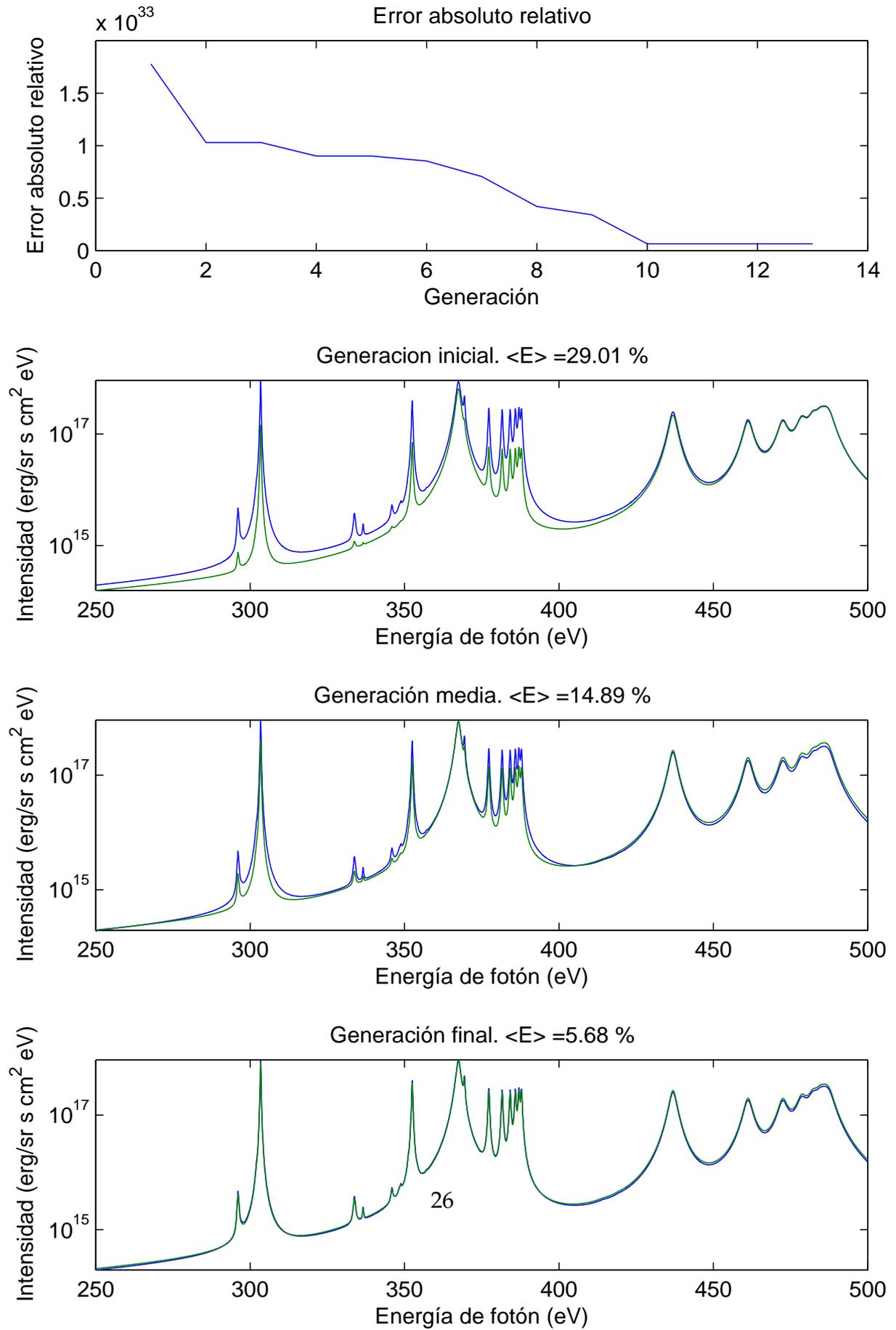


Figura 11: Diagnósis 70 eV, 10^{21} cm^{-3} , 250-500 eV. Sin ruido

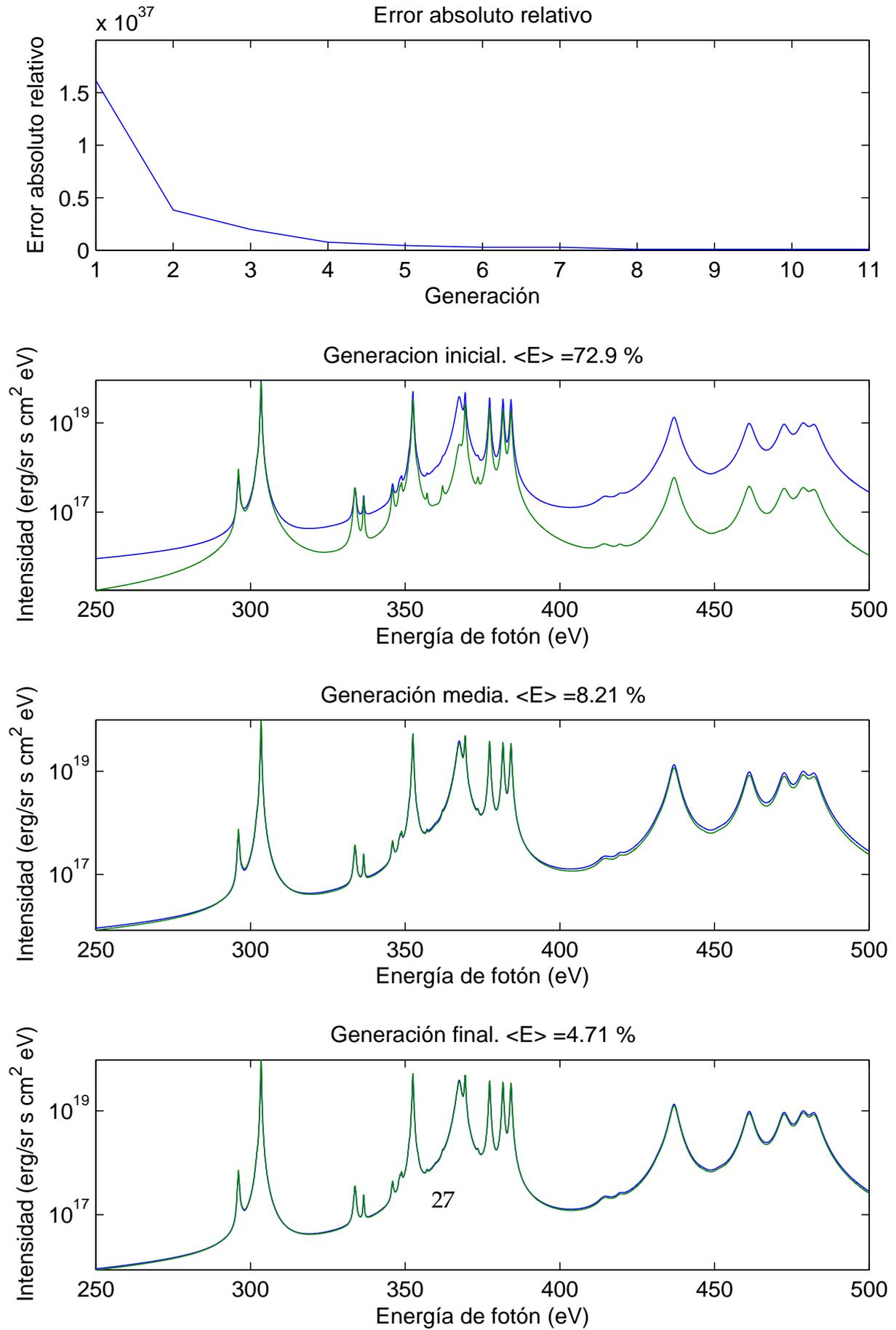


Figura 12: Diagnósis 80 eV, 10^{20} cm^{-3} , 250-500 eV. Ruido uniforme

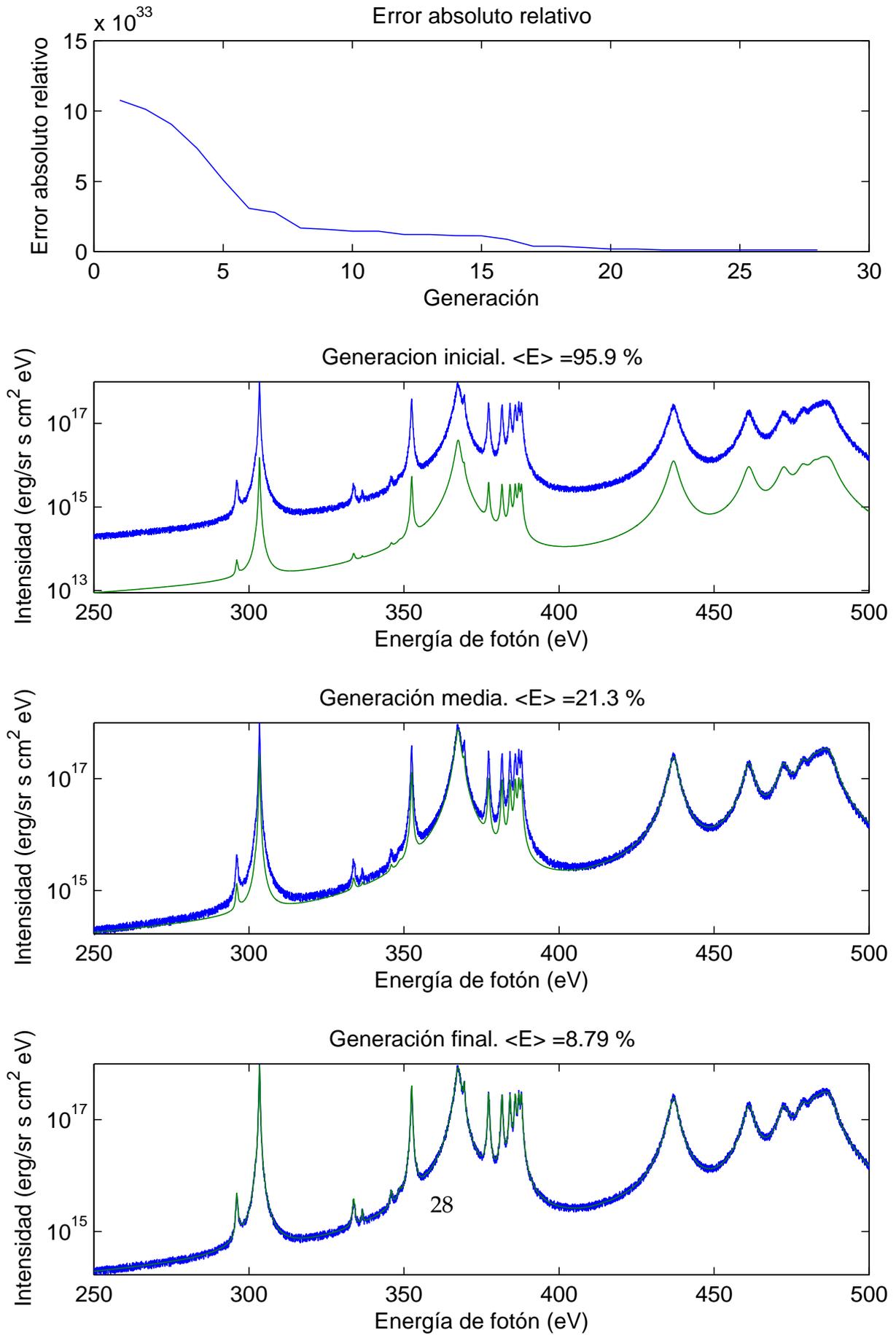


Figura 13: Diagnósis 70 eV, 10^{21} cm^{-3} , 250-500 eV. Ruido uniforme

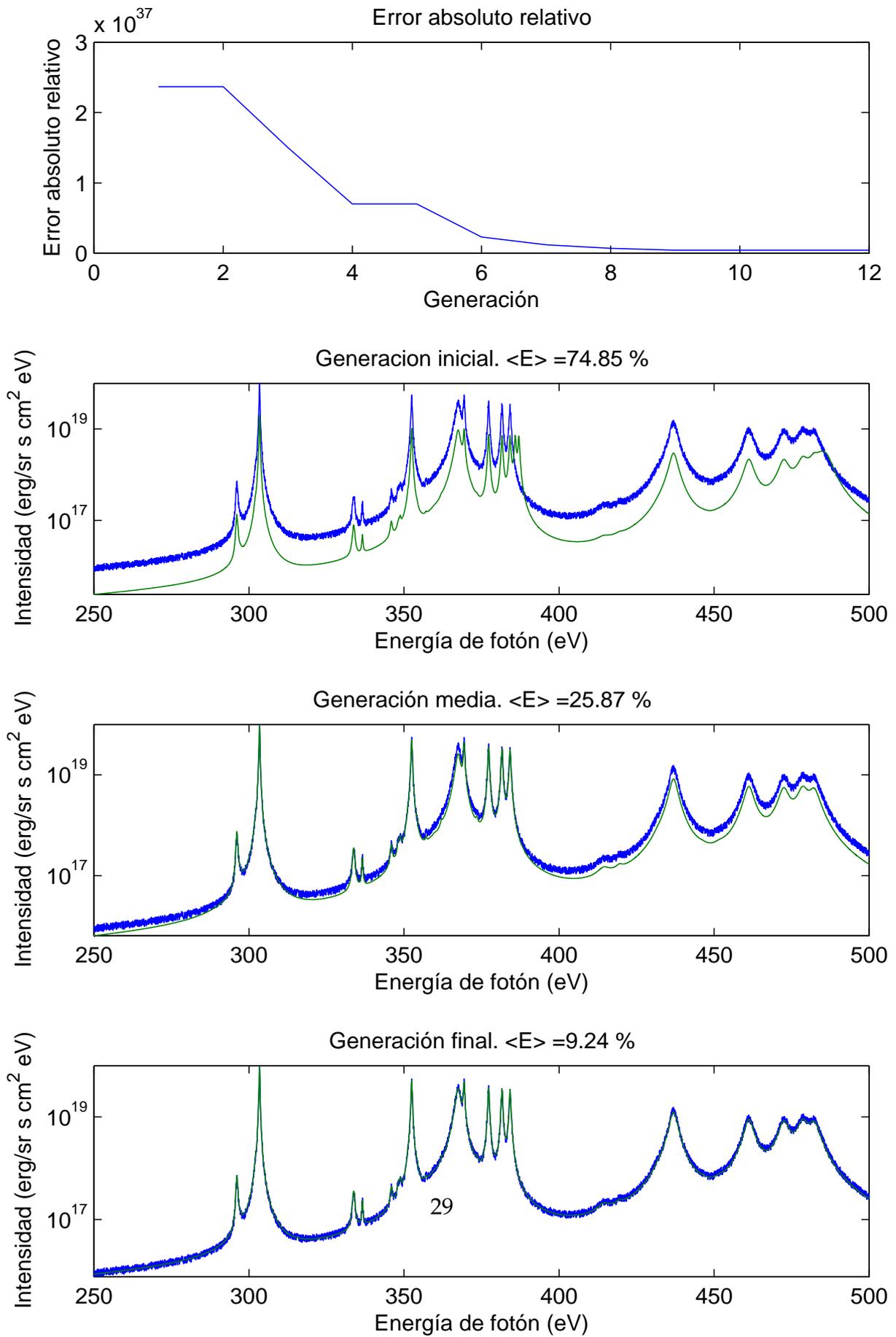


Figura 14: Diagnosis 80 eV, 10^{20} cm^{-3} , 250-500 eV. Ruido gaussiano

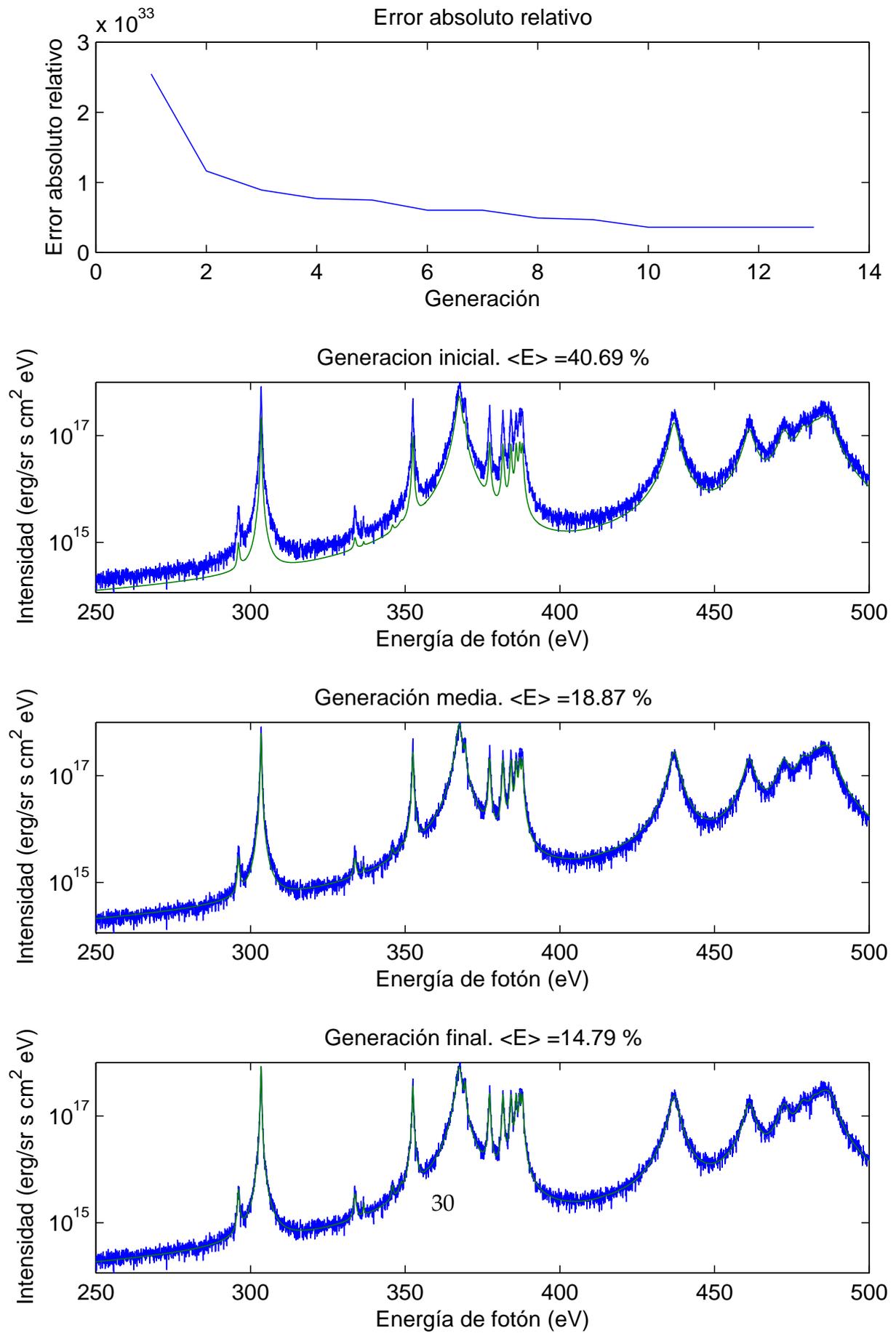
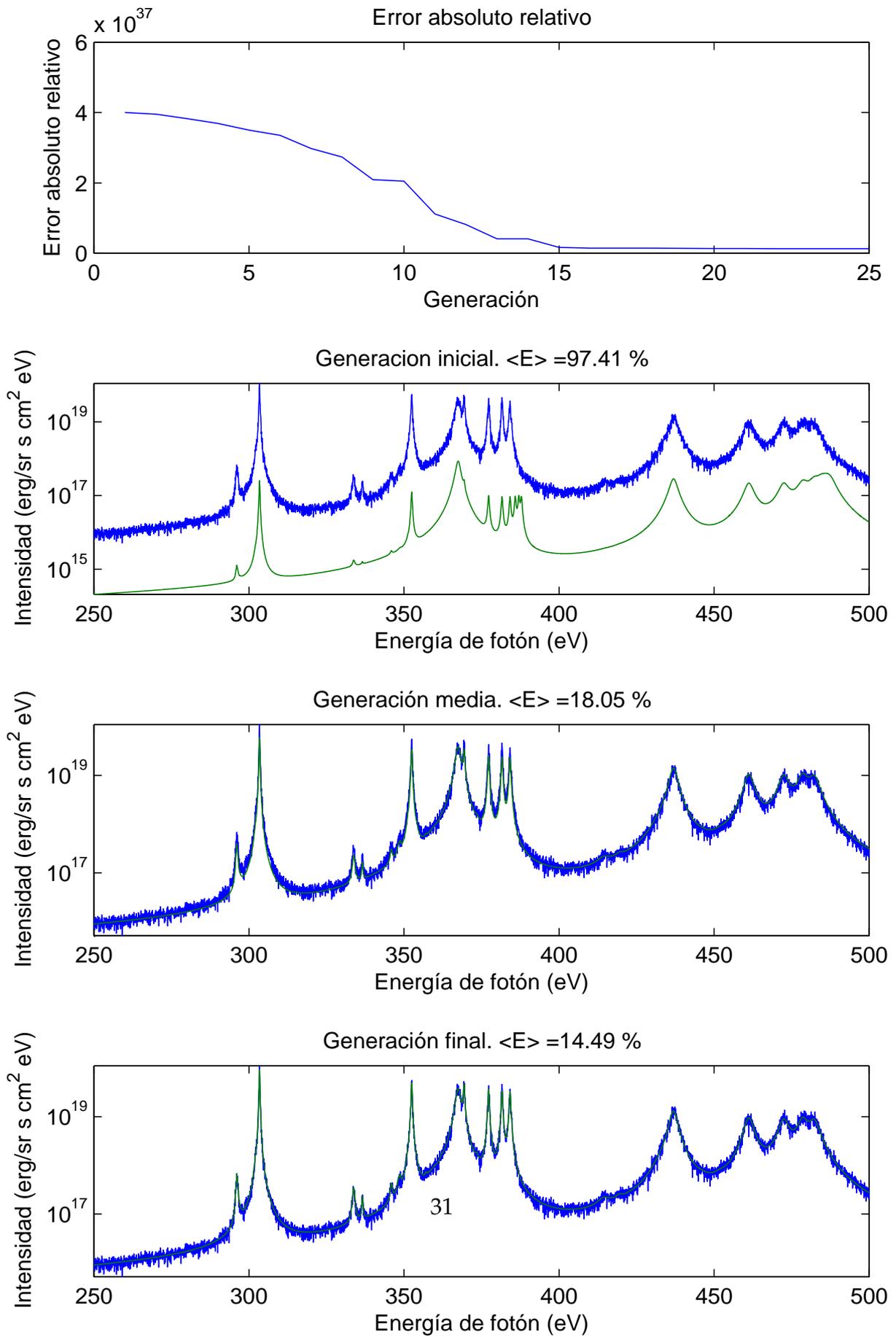


Figura 15: Diagnosis 70 eV, 10^{21} cm^{-3} , 250-500 eV. Ruido gaussiano



4.2. Análisis de un espectro de emisión experimental: Implusiones en OMEGA de blancos de deuterio dopados con argón

En este apartado se discuten los resultados de diagnosticar varios espectros de OMEGA para comprobar la robustez del algoritmo frente a la variación de varios parámetros.

Se hace notar que los tiempos de ejecución no son significativos, ya que debido a la utilización de la base de datos de espectros, los tiempos pueden variar dependiendo de si se generan soluciones ya almacenadas en la base de datos o deben calcularse. Esto explica las diferencias de tiempo entre 16, 32 y 64 soluciones por generación.

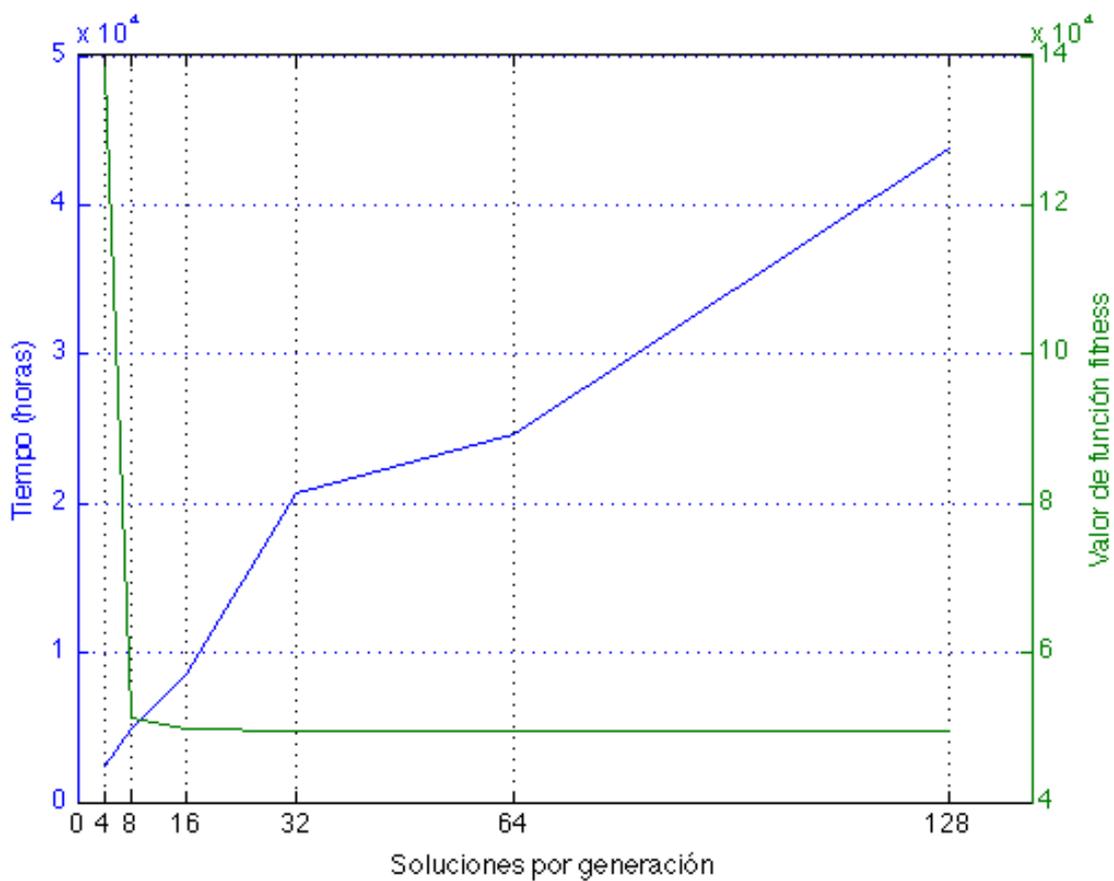
Podemos realizar una normalización para averiguar el coste en tiempo de ejecución de una solución. Para ello calculamos el número total de soluciones calculadas y dividimos entre el tiempo total. Si se realiza el cálculo para el espectro 1 obtenemos 17 centésimas de segundo por individuo. Este valor es mucho más pequeño que los 8-30 segundos que tarda un cálculo normal en ejecutarse, con lo que se deduce que la utilización de la base de datos está disminuyendo el tiempo de diagnosis en uno o dos órdenes de magnitud.

4.2.1. Espectro de emisión de argón 1

La diferencia del error no es significativa para números de soluciones por generación (SPG) superiores a 8, convergiendo bastante bien para todos los SPG superiores al tiempo que se mantiene el número medio de generaciones en 7. De este estudio se puede establecer en 16-32 el número adecuado de SPG a utilizar en diagnósticos análogos a este, dando así un margen de confianza al algoritmo.

En particular se puede observar que para un SPG de 4 el algoritmo no logra una convergencia aceptable, al no conseguir disminuir el error a valores cercanos a los otros casos. Con poblaciones superiores a 32 el tiempo de ejecución se dispara sin mejorar apreciablemente el error final.

Figura 16: Tiempo de ejecución y valor de fitness frente a las soluciones por generación



Cuadro 5: Número de generaciones medias para las diagnosis de Argón en OMEGA con varias SPG

Nº soluciones	Nº medio de gens.
4	9
8	10
16	7
32	7
64	7
128	7

Figura 17: Mejor espectro encontrado para 4 SPG

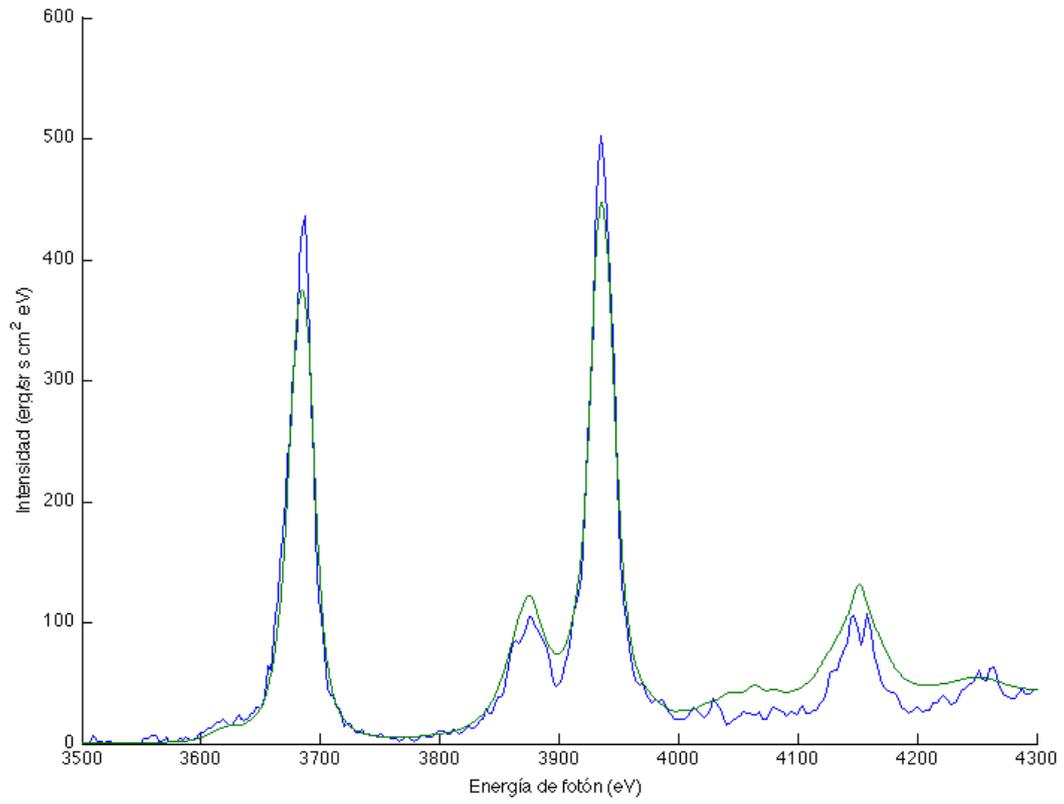
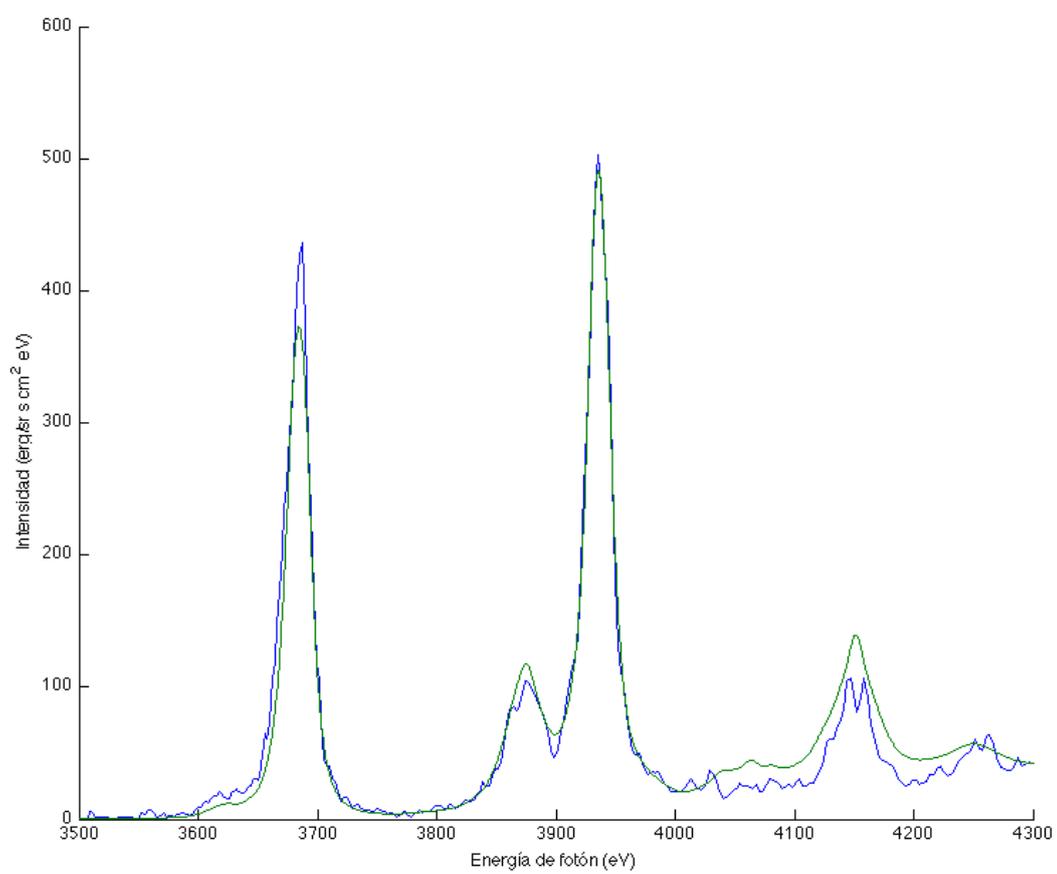


Figura 18: Mejor espectro encontrado para 32 SPG



Es valioso en este caso hacer una comparación entre los tiempos de cálculo que se obtienen con el algoritmo genético y los obtenidos utilizando fuerza bruta. En trabajos previos se han realizado diagnósticos utilizando una base de datos de 30×30 espectros (30 valores de T y 30 valores de N_e). El tiempo de cómputo estimado para generar la base de datos es de aproximadamente 5 horas. Cualquier espectro puede ahora diagnosticarse rápidamente utilizando esa base de datos en un tiempo del orden de pocos segundos. Sin embargo, el método anterior cubre un espacio de búsqueda de temperaturas y densidades más pequeño que el que puede cubrir el algoritmo genético, y el error cometido en los valores de T y N_e puede ser significativo al mallar el espacio de búsqueda en sólo 30 valores por variable.

Utilizando las características de DIPLA para usar espectros de la base de datos como generación inicial, los tiempos de cómputo pueden acelerarse mucho, que sin llegar a aproximarse a los pocos segundos del método anterior, permiten una diagnóstico más precisa en un espacio de búsqueda más amplio.

4.2.2. Espectro de emisión de argón 2

En este caso, en que pretendía buscarse los mejores valores de probabilidades de mutación y cruce para la diagnóstico, los mejores resultados se obtuvieron con valores del 0.5 en ambas variables, obteniendo con ellos un número de generaciones y un error medio más bajo.

Se trazó la frontera de Pareto en este caso entre número de iteraciones y error cuadrático medio con el fin de seleccionar las soluciones de esta metabúsqueda.

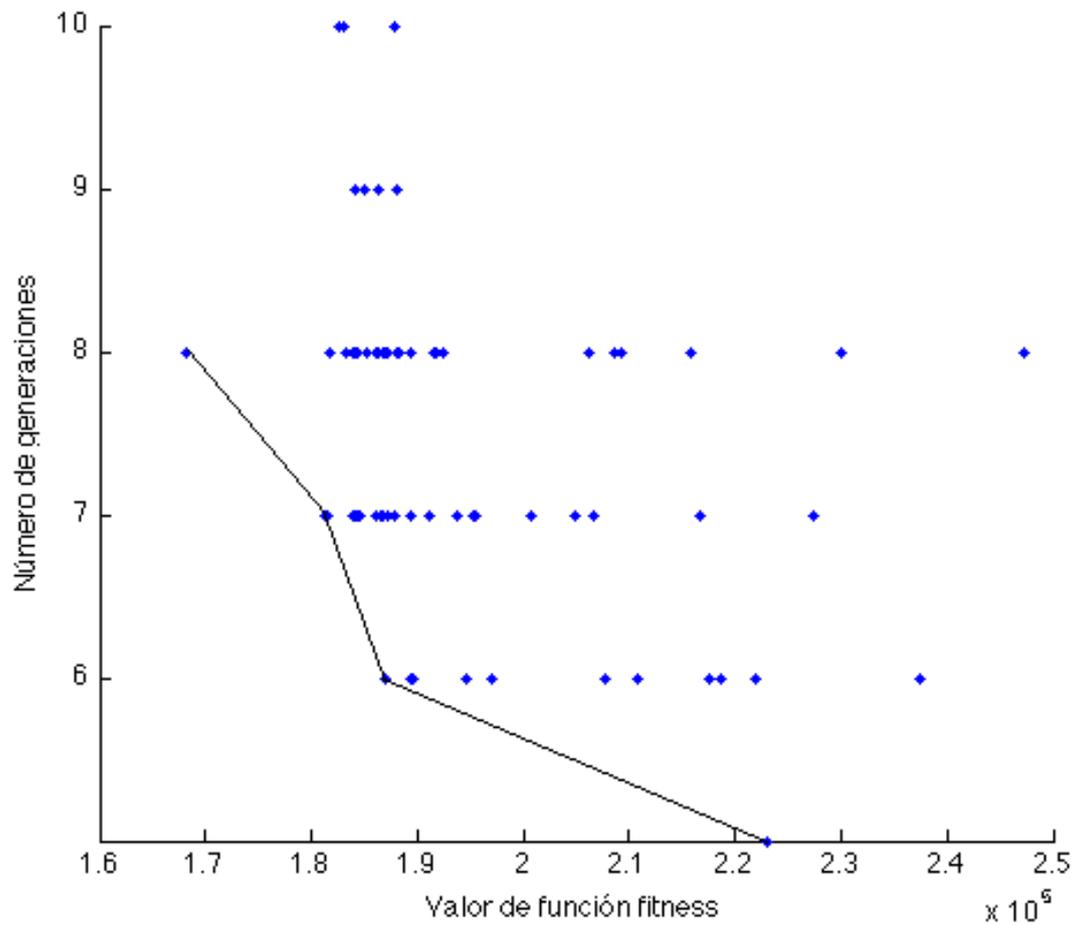
La gráfica siguiente muestra los resultados más significativos, donde se ha trazado la frontera de Pareto para las parejas de probabilidades de cruce y mutación que producen menos iteraciones y menor error absoluto medio. Los puntos de la frontera se han marcado como probabilidad de cruce y probabilidad de mutación en tanto por uno en la gráfica.

Estos resultados permiten escoger con seguridad los valores de la frontera de Pareto, que permitirán al algoritmo comportarse correctamente durante el proceso de diagnóstico.

Los resultados correspondientes a la frontera de Pareto de la búsqueda se muestran en la tabla 6. Los valores encontrados presentan el mejor acuerdo entre número de generaciones, y por tanto, tiempo computacional, y valor final de la función fitness. El compromiso de los valores 0.5 y 0.5 para probabilidad de cruce y mutación respectivamente parece muy apropiado para este tipo de diagnóstico, logrando un valor bajo de la función fitness y manteniendo el tiempo de ejecución en márgenes fácilmente controlables.

Realizando el promedio de los 5 espectros encontrados con el caso $[0.5, 0.5]$, obtenemos la gráfica de la figura 20, donde el acuerdo visual es bueno.

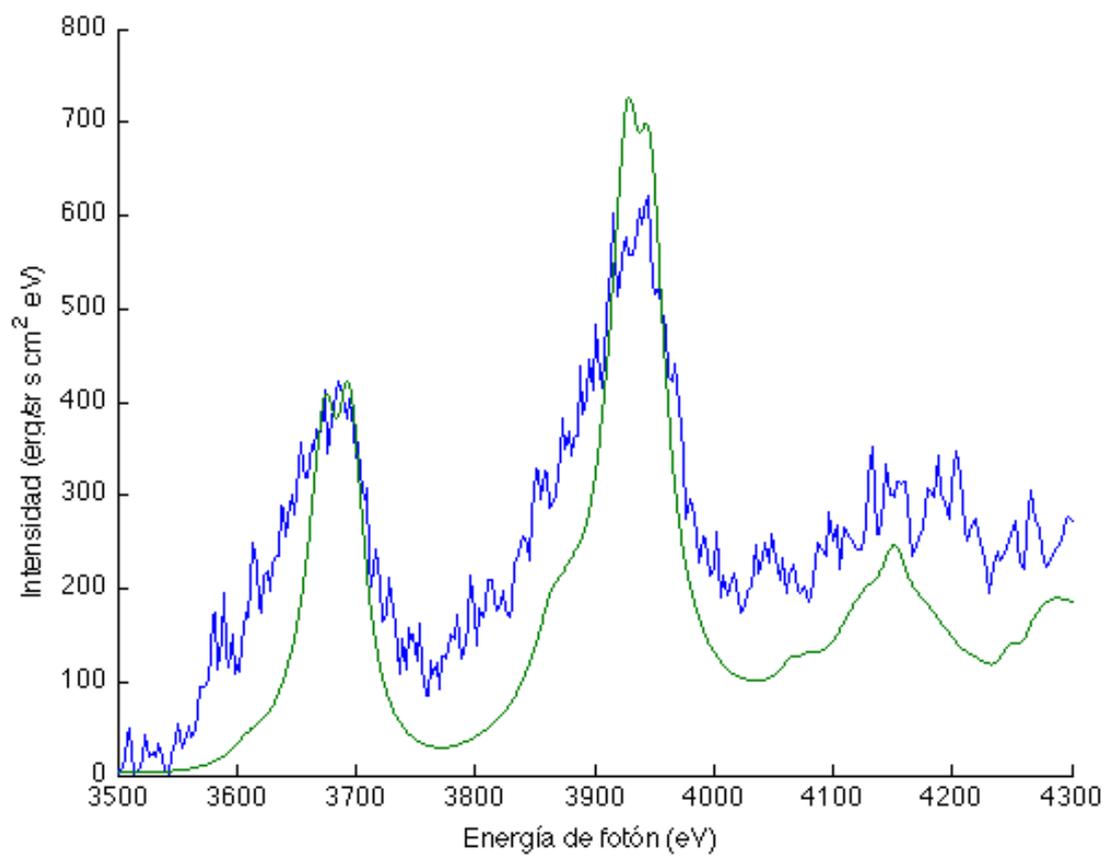
Figura 19: Generaciones y valor final de la función de fitness según probabilidad de cruce y mutación



Cuadro 6: Valores de número de generaciones y valor final de fitness de los puntos de la frontera de pareto

% de cruce	% de mutación	Valor de fitness	Nº de generaciones
0.2	0.9	223100.0	5
0.5	0.5	186890.0	6
0.7	0.9	184230.0	7
0.2	0.2	168150.0	8

Figura 20: Gráfica del mejor espectro medio para el caso: % de cruce 0.5 y % de mutación 0.5



5. Conclusiones

El presente trabajo ha demostrado la utilidad de emplear métodos inteligentes de búsqueda para realizar diagnóstico espectroscópica en plasmas de fusión en condiciones de confinamiento inercial.

Se han encontrado los parámetros que proporcionan resultados correctos para este tipo de diagnóstico, siendo estos los mostrados en la tabla 7.

Cuadro 7: Valor más eficiente de los parámetros evaluados

Parámetro	Valor
Soluciones por generación	16-32
Probabilidad de cruce	0.5
Probabilidad de mutación	0.5

Los tiempos de cómputo se mantienen dentro de los valores abordables en un ordenador personal gracias al empleo de la base de datos de espectros, lográndose en algunos casos, una diagnóstico muy precisa en un corto plazo de tiempo.

6. Trabajo futuro

6.1. Ampliación a multiobjetivo

La optimización mono objetivo produce buenos resultados de diagnóstico en muchos casos, pero no permite obtener nivel de detalle en el interior del núcleo del plasma en condiciones de ICF.

Varios experimentos realizados [10, 9, 11] utilizando algoritmos genéticos multi objetivos (en particular NSGA-II [2]) han permitido obtener temperaturas y densidades espaciales en plasmas no homogéneos, cómo los plasmas reales encontrados en condiciones ICF.

Una ampliación de DIPLA utilizando el algoritmo NSGA-II permitiría desarrollar diagnóstico punteras para el plasma de fusión.

6.2. Determinación de pesos w_i

Como se señaló en el punto 3.2, la utilización del conjunto de espectros de una población para la determinación estadística de las incertezas crea un problema de convergencia en el algoritmo genético. Se plantea la posibilidad de determinar los

pesos no mediante una población sino: 1) mediante los valores acumulados de todas las iteraciones anteriores, 2) mediante valores fijos obtenidos de la base de datos o 3) mediante un supra proceso autoconsistente que englobe el algoritmo genético.

6.3. Refinamiento

El algoritmo DIPLA puede ser mejorado de múltiples formas:

- Esquema de paralelización multiproceso (en vez del multihilo actual), lo que permitiría utilizarlo en un cluster de memoria distribuida.
- Escalado de la precisión del problema en varios aspectos: 1) Tamaño de la matriz de tasas, 2) resolución de la malla de energías del espectro variable. Con ello se acortaría el tiempo de computo de los espectros, refinándose la resolución conforme el algoritmo va acercándose a la solución.
- Utilización del algoritmo Levenberg-Marquardt cuando el algoritmo genético se encuentre relativamente cerca del mínimo. El Levenberg-Marquardt minimiza la función fitness de forma determinista a partir de una solución inicial. Es interesante aplicarlo una vez el algoritmo genético ha evadido los mínimos locales y localizado una solución dentro del entorno de un mínimo de la función fitness.

Referencias

- [1] The python standard library 2.6.2, —2009—.
- [2] K. Deb, Pratap. A, S. Agarwal, and T. Meyarivan. A fast and elitist multi-objective genetic algorithm: NSGA-II. *IEEE Transaction on Evolutionary Computation*, 6(2):181–197, —2002—.
- [3] R. Florido, R. C. Mancini, T. Nagayama, R. Tommasini, J. A. Delettrez, S. P. Regan, V. A. Smalyuk, R. Rodriguez, and J. M. Gil. Argon k-shell and bound-free emission from OMEGA direct-drive implosion cores. *High Energy Density Phys.*, 6(1):70–75, —2010—.
- [4] R. Florido, T. Nagayama, R. C. Mancini, R. Tommasini, J. A. Delettrez, S. P. Regan, V. A. Smalyuk, R. Rodriguez, and J. M. Gil. Analysis of time-resolved argon line spectra from OMEGA direct-drive implosions. *Rev. Sci. Instrum.*, 79:10E310, —2008—.
- [5] R. Florido, R. Rodriguez, J. M. Gil, J. G. Rubiano, P. Martel, E. Minguez, and R. C. Mancini. Modeling of population kinetics of plasmas that are not in local thermodynamic equilibrium, using a versatile collisional-radiative model based on analytical rates. *Physical Review E*, 80(5):056402–18, —2009—.

- [6] C. J. Fontes, J. Abdallah Jr, C. Bowen, R. W. Lee, and Yu Ralchenko. Review of the NLTE-5 kinetics workshop. *High Energy Density Phys.*, 5(1-2):15–22, —2009—.
- [7] David E. Goldberg. *Genetic algorithms in search, optimization and machine learning*. Addison-Wesley, —1989—. Book on genetic algorithms.
- [8] J. G. Rubiano, R. Florido, C. Bowen, R. W. Lee, and Yu Ralchenko. Review of the 4th NLTE code comparison workshop. *High Energy Density Phys.*, 3(1-2):225–232, —2007—.
- [9] L. A. Welser, R. C. Mancini, J. A. Koch, N. Izumi, H. Dalhed, H. Scott, T. W. Barbee, R. W. Lee, I. E. Golovkin, F. Marshall, J. Delettrez, and L. Klein. Analysis of the spatial structure of inertial confinement fusion implosion cores at omega. *J. Quant. Spectrosc. Radiat. Transf.*, 81(1-4):487–497, —2003—.
- [10] L. A. Welser, R. C. Mancini, J. A. Koch, N. Izumi, S. J. Louis, I. E. Golovkin, T. W. Barbee, S. W. Haan, J. A. Delettrez, F. J. Marshall, S. P. Regan, V. A. Smalyuk, D. A. Haynes, and R. W. Lee. Multi-objective spectroscopic analysis of core gradients: Extension from two to three objectives. *J. Quant. Spectrosc. Radiat. Transf.*, 99(1-3):649–657, —2006—.
- [11] Leslie A. Welser. *Processing and Analysis of X-ray Line Spectra and Multi-Monochromatic X-ray Images for Implosion Core Gradient Determination*. PhD thesis, —2003—.