

Manual técnico: Documentación de uso, manejo, desarrollo y evaluación del Software Científico "Nasa Ozone Data Processing and Analysis Tool" (NODPAAT)

Julián Andrés Salamanca Bernal¹, & Jaiver Estiven Salazar Ortiz¹

¹Universidad Distrital Francisco José de Caldas

Resumen

El análisis de datos atmosféricos, particularmente de la concentración de ozono estratosférico, representa un área crítica en las ciencias de la atmósfera y el cambio climático. La comprensión de las variaciones temporales y espaciales del ozono, así como su correlación con fenómenos solares como las manchas solares, requiere herramientas computacionales especializadas capaces de procesar grandes volúmenes de datos satelitales. Como producto de investigación en ciencias atmosféricas y computación científica, se ha desarrollado un sistema integral de análisis que permite procesar, visualizar y analizar datos de ozono atmosférico provenientes de las misiones satelitales de NASA desde 1979 hasta 2024. El Sistema de Análisis de Datos de Ozono Atmosférico de NASA (**NODPAAT**) integra capacidades de extracción de datos en formato HDF5, procesamiento multihilo para análisis de series temporales extensas, análisis estadístico mediante regresión lineal y pruebas de chi-cuadrado, y visualización interactiva mediante una interfaz gráfica de usuario basada en ROOT-C++. El sistema permite la extracción de datos georreferenciados, el análisis de correlaciones entre concentraciones de ozono y actividad solar, y la generación de visualizaciones científicas en múltiples formatos. Con este desarrollo se busca facilitar la investigación en ciencias atmosféricas, reducir la curva de aprendizaje en el análisis de datos satelitales, y proporcionar una herramienta de código abierto que potencie la investigación científica en instituciones académicas. **NODPAAT** se diseña, elabora y evalúa siguiendo las directrices de desarrollo de software científico, empleando tecnologías modernas como C++17, ROOT framework, HDF5, y técnicas de programación paralela con ThreadPool.

Palabras clave— Ozono estratosférico, análisis de datos satelitales, HDF5, manchas solares, interfaz gráfica de usuario, ROOT-C++, procesamiento paralelo, ciencias atmosféricas, visualización científica

Información de los autores.

Julián Andrés Salamanca Bernal¹, [PhD]. [Profesor] [Universidad Distrital Francisco José de Caldas].

[jesalamanca@udistrital.edu.co]

Jaiver Estiven Salazar Ortiz¹, [PGDip]. [Estudiante] [Universidad Distrital Francisco José de Caldas].

[jesalazaro@udistrital.edu.co]

Contenido

1	Introducción	2
2	Validación del Sistema de Análisis de Ozono (NODPAAT)	3
2.1	Matriz de trazabilidad	3
2.2	Contexto y análisis del problema de investigación	5
2.3	Reconocimiento de los requerimientos	6
2.4	Especificaciones de los requerimientos	6
2.5	Diseño de software	6
2.6	Implementación y codificación	8
2.6.1	Características principales de implementación:	8
2.6.2	Ánalisis de Rendimiento y Paralelización	12
2.7	Fase de iteraciones	16
2.7.1	Prueba de software y validación científica	16
2.7.2	Compilación e instalación	17
2.7.3	Verificación y validación científica	17
2.8	Mantenimiento vía git	17

3 Evaluación del sistema NODPAAT	17
3.1 Evaluación según el Modelo KITE	17
3.1.1 Determinante K: Efecto de diseminación del conocimiento	17
3.1.2 Determinante I: Impacto en la población objetivo	18
3.1.3 Determinante T: Innovación tecnológica	19
3.1.4 Determinante E: Logro de ingeniería	21
3.1.5 Representación gráfica del Modelo KITE	22
3.1.6 Evaluación cualitativa y recomendaciones	22
3.2 Criterios de evaluación	23
4 Conclusiones	24

Listas de figuras

1 Matriz de trazabilidad del sistema NODPAAT . Las flechas sólidas representan el flujo secuencial, las punteadas indican ciclos de retroalimentación.	5
2 Diagrama de flujo del sistema NODPAAT mostrando la arquitectura modular y el flujo de procesamiento de datos.	7
3 Jerarquía de componentes gráficos de la interfaz ozone_gui.cpp mostrando la estructura de ventanas, pestañas y elementos de control.	9
4 Tab 1 - Welcome: Pantalla de bienvenida mostrando información general del sistema, descripción de funcionalidades, y guía rápida de uso.	10
5 Tab 2 - Processor: Interfaz de procesamiento de datos por ubicación geográfica, permitiendo especificar coordenadas (latitud/longitud), seleccionar rangos de años, configurar modo de ejecución (secuencial/paralelo), y monitorear el progreso del procesamiento.	10
6 Tab 3 - Macro Runner: Interfaz de ejecución de análisis científicos mediante macros ROOT. Permite ejecutar linearRelStudy03vsSn.C para análisis de correlación ozono-manchas solares y macro03teoGlobalHttp.C para análisis teórico global con datos solares actualizados vía HTTP.	11
7 Tab 4 - Graph Viewer: Visualizador de gráficos de correlación entre ozono atmosférico (O3) y número de manchas solares (Sunspot Number). Muestra resultados del análisis de regresión lineal, dispersión de datos, y parámetros estadísticos con exportación a múltiples formatos.	11
8 Tab 5 - View O3 Global: Visualizador de datos globales de ozono mostrando series temporales históricas, predicciones teóricas basadas en modelos solares, comparación entre datos observados y teóricos, y análisis por años individuales o multi-año.	12
9 Dashboard de rendimiento del procesamiento paralelo mostrando speedup, tiempos de ejecución, eficiencia y tabla resumen.	13
10 Análisis de speedup real (azul) vs ideal (rojo punteado) en función del número de hilos.	14
11 Eficiencia paralela en función del número de hilos. Línea verde: eficiencia real; roja: ideal (100 %); naranja: umbral aceptable (75 %).	14
12 Comparación de tiempos de ejecución: secuencial (rojo) vs paralelo con 2, 4 y 8 hilos (azul).	15

Listas de tablas

1 Especificaciones de software para NODPAAT	6
2 Validación del sistema NODPAAT en múltiples plataformas.	16
3 Puntuaciones de NODPAAT según el Modelo KITE.	22

1 Introducción

En el campo de las ciencias atmosféricas y el estudio del cambio climático, el análisis sistemático de datos satelitales de ozono atmosférico representa una componente fundamental para comprender la dinámica de la estratosfera terrestre y los procesos que afectan la capa de ozono. La disponibilidad de datos satelitales históricos desde 1979 hasta la actualidad proporciona una oportunidad única para estudiar tendencias a largo plazo y correlaciones con fenómenos solares.

Como trabajo de investigación en el área de ciencias atmosféricas y computación científica, se desarrolla el Sistema de Análisis de Datos de Ozono Atmosférico de NASA (**NODPAAT**) que satisface una necesidad de investigación científica: la de procesar, analizar y visualizar grandes volúmenes de datos satelitales de ozono de manera eficiente y reproducible. Este sistema fue desarrollado específicamente para facilitar investigaciones científicas sobre la relación

entre la actividad solar y las concentraciones de ozono estratosférico en regiones tropicales, como las documentadas por González-Navarrete *et al.* (2018) en su estudio sobre modelos adaptativos de ozono para la región tropical ecuatorial andina colombiana, y por González-Navarrete y Salamanca (2019) en su análisis de predicciones de ozono sobre la Zona de Confluencia Intertropical Amazónica. El desarrollo de **NODPAAT** responde directamente a la necesidad de herramientas computacionales especializadas que permitan este tipo de análisis científicos de manera sistemática y reproducible.

En ese contexto de investigación científica se requiere integrar diferentes capacidades computacionales para elaborar un sistema que permita la extracción de datos georreferenciados, el análisis de series temporales extensas, la correlación con actividad solar, y la visualización interactiva de resultados.

El sistema **NODPAAT** permite el procesamiento de datos satelitales de ozono en formato HDF5, la extracción de concentraciones de ozono en ubicaciones geográficas específicas, el análisis estadístico mediante regresión lineal y pruebas de chi-cuadrado, y la visualización interactiva mediante una interfaz gráfica de usuario. El sistema está diseñado para procesar datos desde 1979 hasta 2024, abarcando 45 años de mediciones satelitales.

La arquitectura del sistema se compone de cinco capas principales: (1) capa de extracción de datos HDF5, (2) capa de procesamiento y consolidación de series temporales, (3) capa de análisis estadístico, (4) capa de visualización interactiva mediante GUI basada en ROOT-C++ (Antcheva *et al.*, 2011), y (5) capa de generación de reportes en múltiples formatos (ROOT, PDF, PNG). El sistema implementa técnicas de programación paralela mediante ThreadPool para optimizar el procesamiento de grandes volúmenes de datos.

NODPAAT se propone como una herramienta de código abierto que facilita la investigación en ciencias atmosféricas, permitiendo a investigadores y estudiantes realizar análisis reproducibles de datos satelitales de ozono. Es de resaltar que el sistema permite la correlación entre concentraciones de ozono y actividad solar (manchas solares), la visualización de tendencias temporales, y la generación de análisis estadísticos rigurosos.

2 Validación del Sistema de Análisis de Ozono (NODPAAT)

En virtud de las mejores prácticas en desarrollo de software científico y la necesidad de documentación rigurosa para proyectos de investigación, se establece una matriz de trazabilidad que documenta el uso, manejo, desarrollo y evaluación del sistema **NODPAAT**. El desarrollo del sistema se realizó siguiendo un **modelo iterativo e incremental**, donde cada componente se implementó de forma gradual, incorporando mejoras y correcciones de bugs identificados en iteraciones previas. Este enfoque permitió un desarrollo flexible, con ciclos continuos de implementación, prueba y refinamiento, adaptándose a las necesidades emergentes del análisis científico y garantizando la corrección progresiva de errores.

La evaluación de este sistema sigue criterios de calidad de software científico, incluyendo reproducibilidad de resultados, correctitud de algoritmos numéricos, eficiencia en procesamiento de datos, y usabilidad de interfaces.

2.1 Matriz de trazabilidad

La matriz de trazabilidad contempla nueve etapas:

- **Contexto y análisis del problema de investigación:** establece las necesidades de investigación en ciencias atmosféricas y las capacidades técnicas necesarias para el análisis de datos satelitales
- **Reconocimiento de requerimientos:** identifica y caracteriza los recursos computacionales necesarios para el procesamiento de datos HDF5 y análisis estadístico
- **Especificación de los requerimientos:** detalla las funcionalidades, librerías y versiones necesarias (HDF5, ROOT, C++17)
- **Diseño de software:** corresponde al proceso de arquitectura del sistema, incluyendo módulos de extracción, procesamiento, análisis y visualización
- **Implementación y codificación:** materialización del diseño en código C++, incluyendo optimizaciones de rendimiento y paralelización
- **Instalación-verificación-desinstalación:** documenta el proceso de compilación, instalación y verificación del sistema
- **Mantenimiento vía git:** establece el control de versiones y registro de modificaciones
- **Desarrollo de nuevas versiones:** documenta la evolución continua del sistema.

La Figura 1 presenta visualmente la matriz de trazabilidad del sistema **NODPAAT**, mostrando el flujo completo del proceso de desarrollo desde la concepción hasta el mantenimiento continuo. El diagrama se estructura en tres niveles principales:

Etapas preliminares (niveles superiores). Las cinco primeras etapas constituyen la fase de desarrollo fundamental del sistema: (1) el *Contexto y análisis del problema de investigación* establece las bases científicas y computacionales del proyecto; (2) el *Reconocimiento de requerimientos* identifica las necesidades técnicas específicas; (3) la *Especificación de requerimientos* formaliza estos requerimientos en términos de software y hardware; (4) el *Diseño de software* define la arquitectura modular del sistema; y (5) la *Implementación y codificación* materializa el diseño en código C++ funcional. Estas etapas siguen un flujo secuencial donde cada una alimenta a la siguiente, construyendo progresivamente el sistema desde conceptos abstractos hasta código ejecutable.

Fase de iteraciones (núcleo central). La etapa central, destacada con fondo amarillo en el diagrama, representa el corazón del proceso de desarrollo iterativo e incremental. Esta fase comprende tres actividades interconectadas mediante ciclos de retroalimentación: *Prueba de software* verifica la funcionalidad y correctitud del código implementado, ejecutando pruebas unitarias, de integración y validación científica; *Compilación e instalación* garantiza que el sistema se construye correctamente y puede ser desplegado en diferentes entornos; y *Corrección de errores y recodificación* aborda los problemas identificados durante las pruebas y compilación, implementando mejoras y correcciones. Las flechas sólidas conectan estas tres actividades mostrando el flujo principal del proceso, mientras que las flechas punteadas que regresan desde la corrección hacia las pruebas y compilación representan la naturaleza cíclica de este proceso: cada corrección genera una nueva iteración de pruebas y compilación, creando un ciclo de mejora continua que se repite hasta alcanzar los criterios de calidad establecidos. Este enfoque iterativo fue fundamental para el desarrollo de **NODPAAT**, permitiendo identificar y resolver problemas de manera temprana, optimizar el rendimiento del procesamiento paralelo, y refinar la interfaz gráfica basándose en pruebas de usabilidad.

Etapas finales y mantenimiento (niveles inferiores). Una vez superada exitosamente la fase de iteraciones, el sistema progresiona hacia las etapas finales mediante flechas sólidas que indican el flujo secuencial: *Instalación-verificación-desinstalación* documenta los procedimientos de despliegue del sistema en entornos de producción y verifica su correcto funcionamiento; *Mantenimiento vía git* establece el control de versiones mediante Git, documentando todos los cambios, correcciones y mejoras posteriores al despliegue inicial; finalmente, *Desarrollo de nuevas versiones* representa la evolución continua del sistema mediante la incorporación de nuevas funcionalidades, optimizaciones adicionales, o adaptaciones a nuevos requerimientos científicos. La flecha punteada que conecta esta última etapa de regreso a la fase de iteraciones indica que el desarrollo de nuevas versiones reinicia el ciclo iterativo, asegurando que toda nueva funcionalidad pasa por el mismo proceso riguroso de pruebas, compilación y corrección antes de ser integrada al sistema principal.

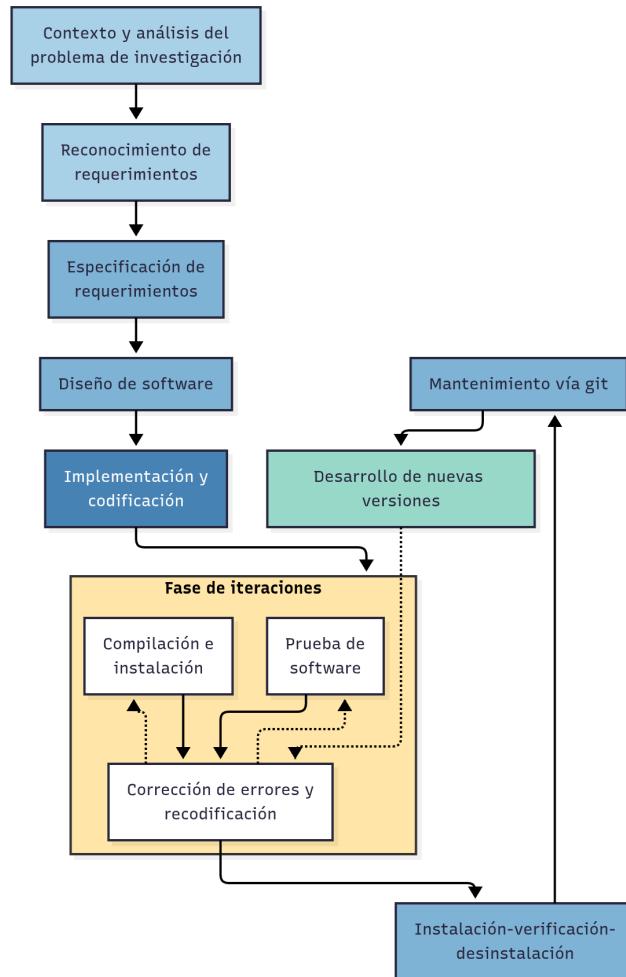


Figura 1: Matriz de trazabilidad del sistema **NODPAAT**. Las flechas sólidas representan el flujo secuencial, las punteadas indican ciclos de retroalimentación.

Este modelo de desarrollo iterativo e incremental resultó particularmente apropiado para **NODPAAT** debido a la complejidad inherente del procesamiento de datos científicos satelitales y la necesidad de validación continua de resultados. El enfoque permitió desarrollar inicialmente un módulo básico de extracción de datos HDF5, validarla científicamente, y luego expandirlo incrementalmente con capacidades de procesamiento paralelo, análisis estadístico, y visualización interactiva. Cada incremento fue sometido a múltiples iteraciones de pruebas y refinamiento, garantizando la robustez y correctitud científica del sistema en cada etapa de su evolución.

2.2 Contexto y análisis del problema de investigación

El problema de investigación central consiste en analizar la correlación entre las concentraciones de ozono estratosférico y la actividad solar (medida por el número de manchas solares) durante un período de 45 años (1979-2024). Esta investigación requiere:

- Procesamiento eficiente de aproximadamente 16,000 archivos HDF5 conteniendo datos satelitales
- Extracción de datos georreferenciados para ubicaciones específicas (ciudades, regiones)
- Análisis estadístico mediante regresión lineal y pruebas de chi-cuadrado
- Visualización interactiva de series temporales y correlaciones
- Generación de reportes científicos en formatos estándar (ROOT, PDF, PNG)

La necesidad de investigación requiere un sistema robusto, eficiente y reproducible que permita a investigadores realizar análisis complejos de datos atmosféricos sin requerir expertise avanzado en programación científica.

2.3 Reconocimiento de los requerimientos

Los requerimientos técnicos del sistema **NODPAAT** incluyen:

- **Formato de datos:** Soporte para archivos HDF5 (Hierarchical Data Format versión 5)
- **Framework de visualización:** ROOT-C++ del CERN para interfaces gráficas y análisis científico
- **Procesamiento paralelo:** ThreadPool para procesamiento multihilo de datos
- **Análisis estadístico:** Implementación de regresión lineal, chi-cuadrado, y análisis de correlación
- **Lenguaje:** C++17 con soporte para características modernas
- **Sistema operativo:** Linux (Ubuntu 20.04 LTS o superior recomendado)

2.4 Especificaciones de los requerimientos

Los requerimientos básicos de instalación recomendados para ejecutar **NODPAAT** son:

- Memoria RAM mínima de 8 GB (16 GB recomendado para procesamiento de datasets completos)
- Procesador multi-core (mínimo 4 cores, recomendado 8+ cores para procesamiento paralelo)
- Espacio en disco: mínimo 50 GB para datos y resultados
- Sistema operativo Linux (Ubuntu 20.04 LTS o superior)

Las especificaciones de software se ilustran en la Tabla 1.

Tabla 1: Especificaciones de software para **NODPAAT**.

Nombre	Versión	Descripción
ROOT-C++	6.24+	Framework científico del CERN
HDF5	1.10+	Librería para archivos HDF5
g++	9.0+	Compilador C++ con soporte C++17
GNU Make	4.0+	Sistema de compilación

2.5 Diseño de software

El sistema **NODPAAT** está diseñado con una arquitectura modular que comprende cinco capas principales. La Figura 2 ilustra el flujo de datos y la interacción entre los diferentes componentes del sistema.

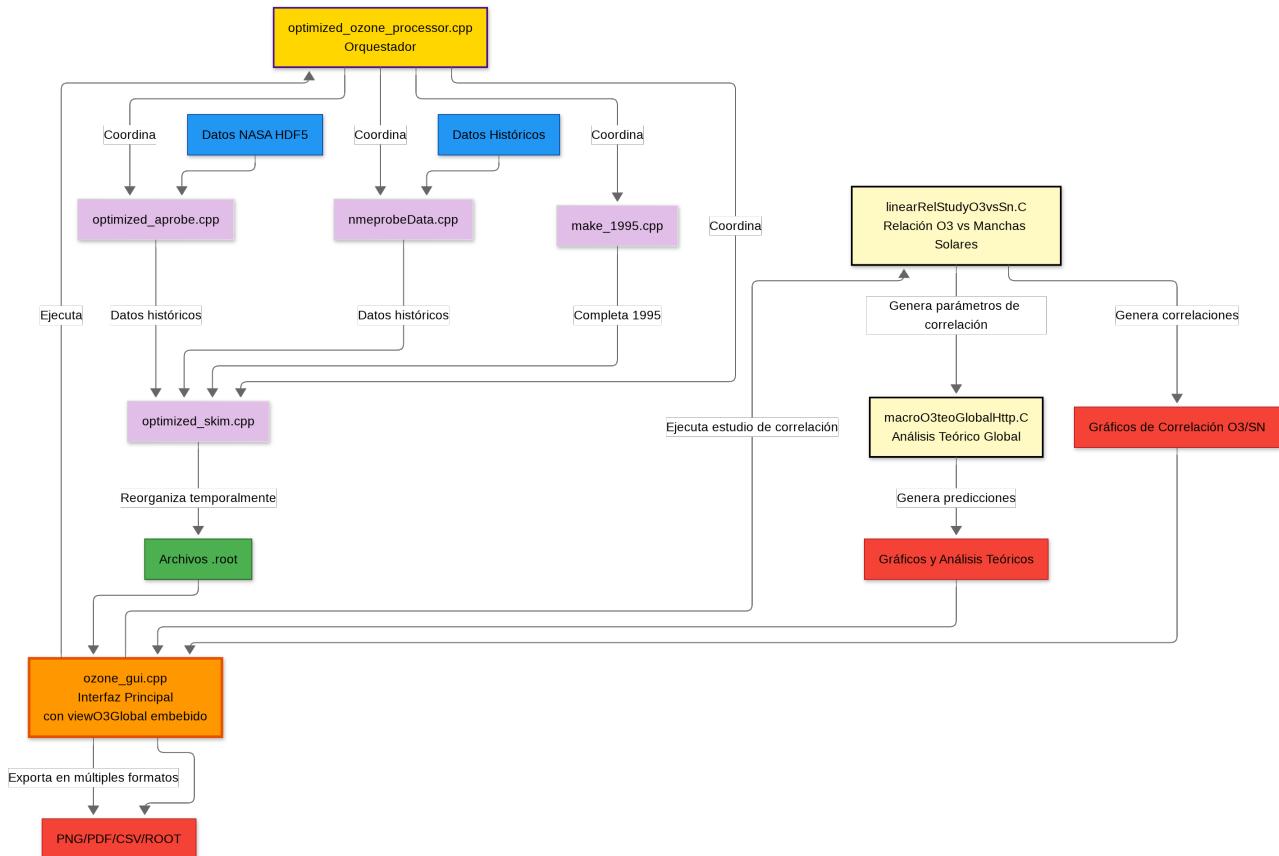


Figura 2: Diagrama de flujo del sistema **NODPAAT** mostrando la arquitectura modular y el flujo de procesamiento de datos.

El flujo de trabajo del sistema, como se representa en la Figura 2, inicia con la interfaz gráfica de usuario (`ozone_gui.cpp`) que actúa como punto de entrada principal y ejecuta el orquestador del sistema (`optimized_ozone_processor.cpp`). Este orquestador coordina cuatro módulos de procesamiento especializados: (1) `optimized_aprobe.cpp` que procesa los datos HDF5 de NASA, (2) `nmeprobeData.cpp` que gestiona datos históricos, (3) `make_1995.cpp` que completa los datos faltantes del año 1995, y (4) `optimized_skim.cpp` que reorganiza temporalmente toda la información procesada. Los resultados consolidados se almacenan en archivos ROOT que retornan a la interfaz gráfica.

A partir de estos archivos ROOT, la interfaz gráfica ejecuta dos macros de análisis especializado de forma secuencial: primero `linearRelStudyO3vsSn.C`, que realiza el estudio de correlación lineal entre el ozono atmosférico (O3) y el número de manchas solares (Sunspot Number), generando parámetros estadísticos esenciales; posteriormente, estos parámetros son utilizados por `macroO3teoGlobalHttp.C` para realizar análisis teóricos globales y generar predicciones basadas en modelos solares. Los gráficos y análisis generados por ambos macros son visualizados directamente en la interfaz gráfica y pueden ser exportados en múltiples formatos (PNG, PDF, CSV, ROOT) según las necesidades del investigador.

El sistema **NODPAAT** está estructurado en cinco capas principales:

1. Capa de Interfaz de Usuario (User Interface Layer):

- `ozone_gui.cpp`: Interfaz gráfica principal con `viewO3Global` embedeido
- `viewO3Global.cpp`: Visualizador interactivo de datos globales de ozono
- Integración de múltiples formatos de exportación (PNG, PDF, CSV, ROOT)

2. Capa de Orquestación (Orchestration Layer):

- `optimized_ozone_processor.cpp`: Orquestador principal del sistema
- Coordinación de módulos de extracción y procesamiento de datos
- Gestión de flujo de trabajo entre componentes

3. Capa de Procesamiento de Datos (Data Processing Layer):

- `optimized_aprobe.cpp`: Procesamiento de datos HDF5 de NASA

- `nmeprobeData.cpp`: Gestión de datos históricos
- `make_1995.cpp`: Completado de datos faltantes del año 1995
- `optimized_skim.cpp`: Reorganización temporal de datos procesados

4. Capa de Análisis Científico (Scientific Analysis Layer):

- `linearRelStudy03vsSn.C`: Estudio de relación lineal entre O3 y manchas solares
- `macro03teoGlobalHttp.C`: Análisis teórico global y generación de predicciones
- `chi2LinearRelStudy03vsSn.cxx`: Análisis de chi-cuadrado para validación estadística
- `statsMacroChi2LRS03Sn.C`: Macro de estadísticas complementarias
- `analysis_runner.cpp`: Ejecutor automatizado de análisis

5. Capa de Soporte (Support Layer):

- `include/funSolar.h`: Funciones de física solar y cálculos astronómicos
- `include/fun.h`: Funciones auxiliares y utilidades generales
- Implementación de algoritmos de regresión lineal y métodos estadísticos

2.6 Implementación y codificación

El sistema **NODPAAT** está implementado en C++17 haciendo uso intensivo del framework ROOT-C++ para visualización científica y análisis de datos. La arquitectura del código sigue principios de modularidad y separación de responsabilidades.

2.6.1 Características principales de implementación:

- **Procesamiento Paralelo**: Implementación de ThreadPool para procesamiento multihilo de archivos HDF5, permitiendo análisis eficiente de grandes volúmenes de datos.
- **Lectura de HDF5**: Utilización de la API de HDF5 para extracción eficiente de datasets multidimensionales con manejo de excepciones y validación de datos.
- **Interfaz Gráfica**: GUI basada en ROOT con arquitectura de tabs. La Figura 3 muestra la jerarquía completa de componentes gráficos:
 - Tab 1: Introducción al programa (Figura 4) - Información general y guía de uso
 - Tab 2: Procesamiento de datos por ubicación (Figura 5) - Permite procesar datos para coordenadas específicas (latitud/longitud)
 - Tab 3: Ejecución de análisis (Figura 6) - Ejecuta `macro03teoGlobalHttp.C` y `linearRelStudy03vsSn.C`
 - Tab 4: Visualización de correlaciones O3/SN (Figura 7) - Muestra gráficos generados por `linearRelStudy03vsSn.C` o el procesador global
 - Tab 5: Visualización de predicciones teóricas (Figura 8) - Muestra gráficos generados por `macro03teoGlobalHttp.C`

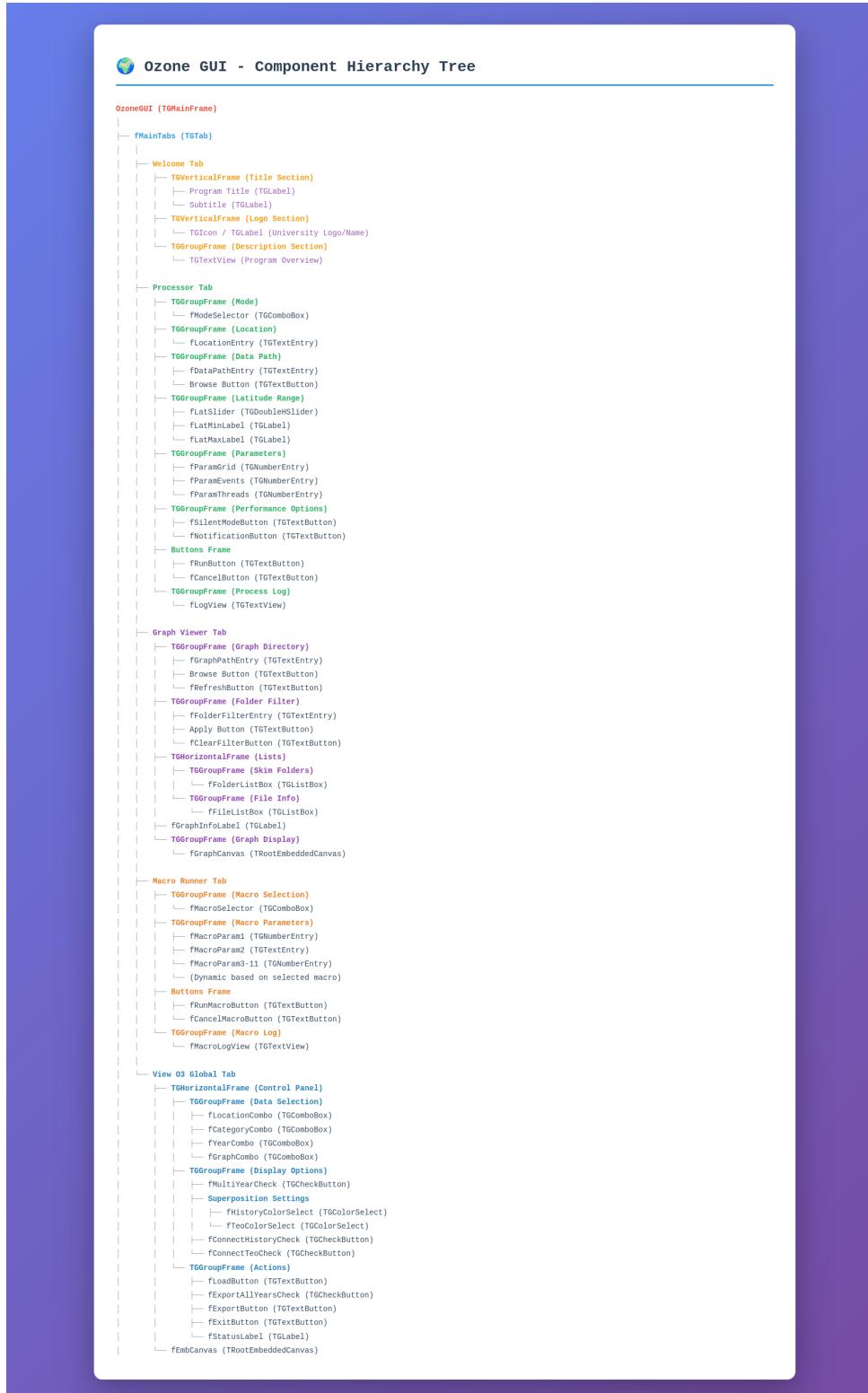


Figura 3: Jerarquía de componentes gráficos de la interfaz ozone_gui.cpp mostrando la estructura de ventanas, pestañas y elementos de control.

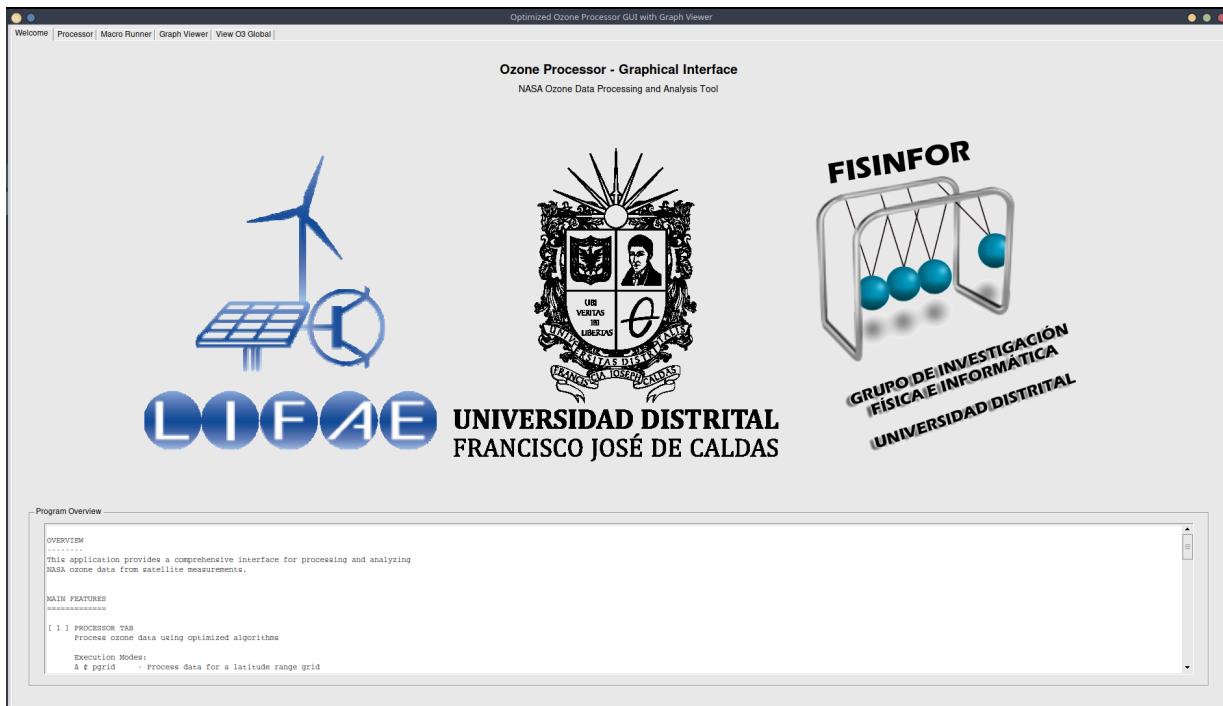


Figura 4: Tab 1 - Welcome: Pantalla de bienvenida mostrando información general del sistema, descripción de funcionalidades, y guía rápida de uso.

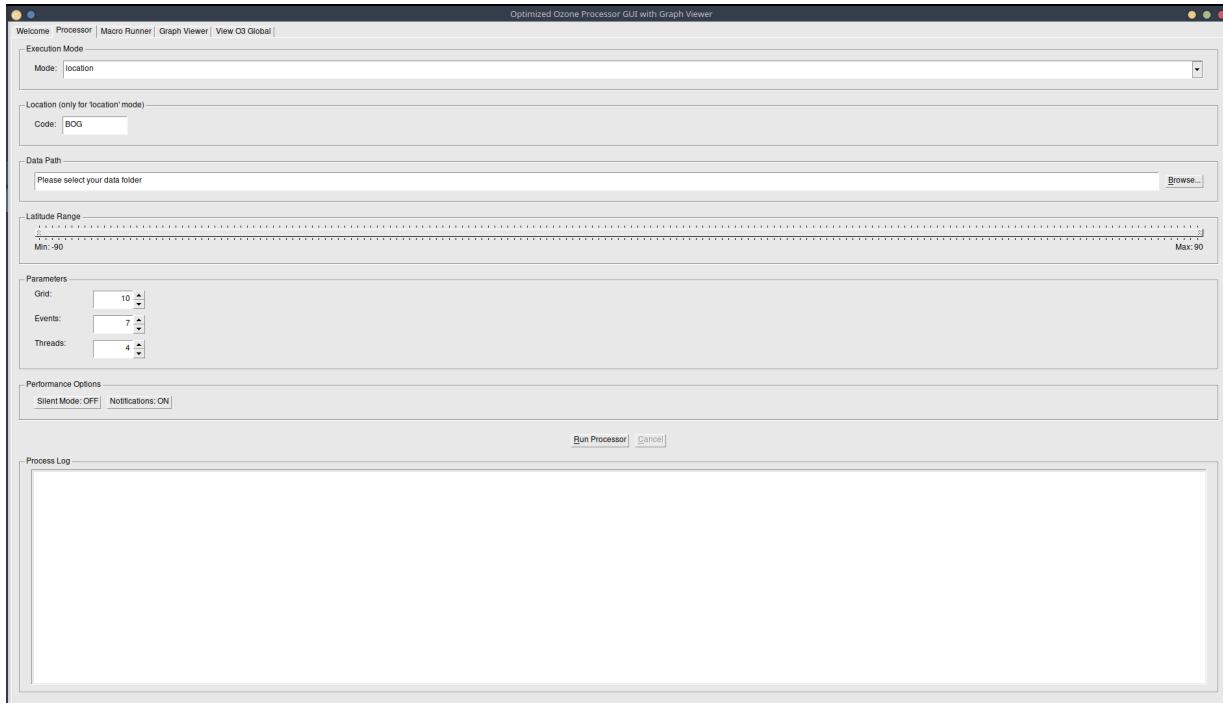


Figura 5: Tab 2 - Processor: Interfaz de procesamiento de datos por ubicación geográfica, permitiendo especificar coordenadas (latitud/longitud), seleccionar rangos de años, configurar modo de ejecución (secuencial/paralelo), y monitorear el progreso del procesamiento.

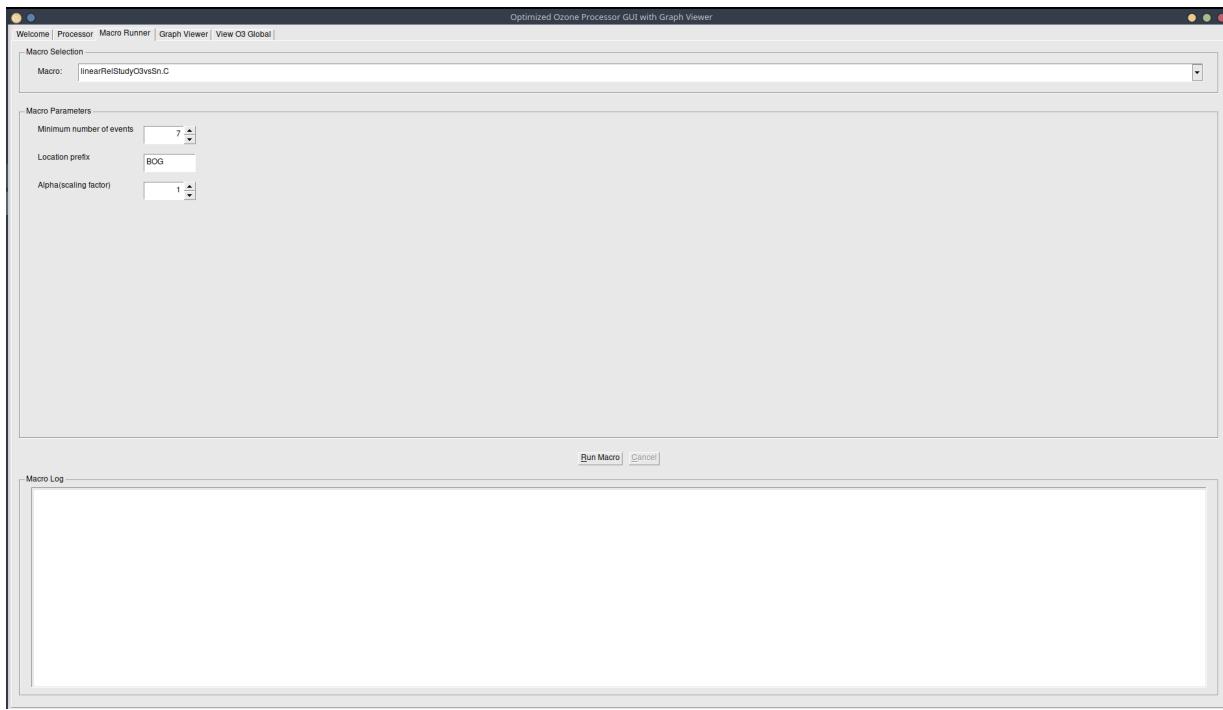


Figura 6: Tab 3 - Macro Runner: Interfaz de ejecución de análisis científicos mediante macros ROOT. Permite ejecutar `linearRelStudyO3vsSn.C` para análisis de correlación ozono-manchas solares y `macro03teoGlobalHttp.C` para análisis teórico global con datos solares actualizados vía HTTP.

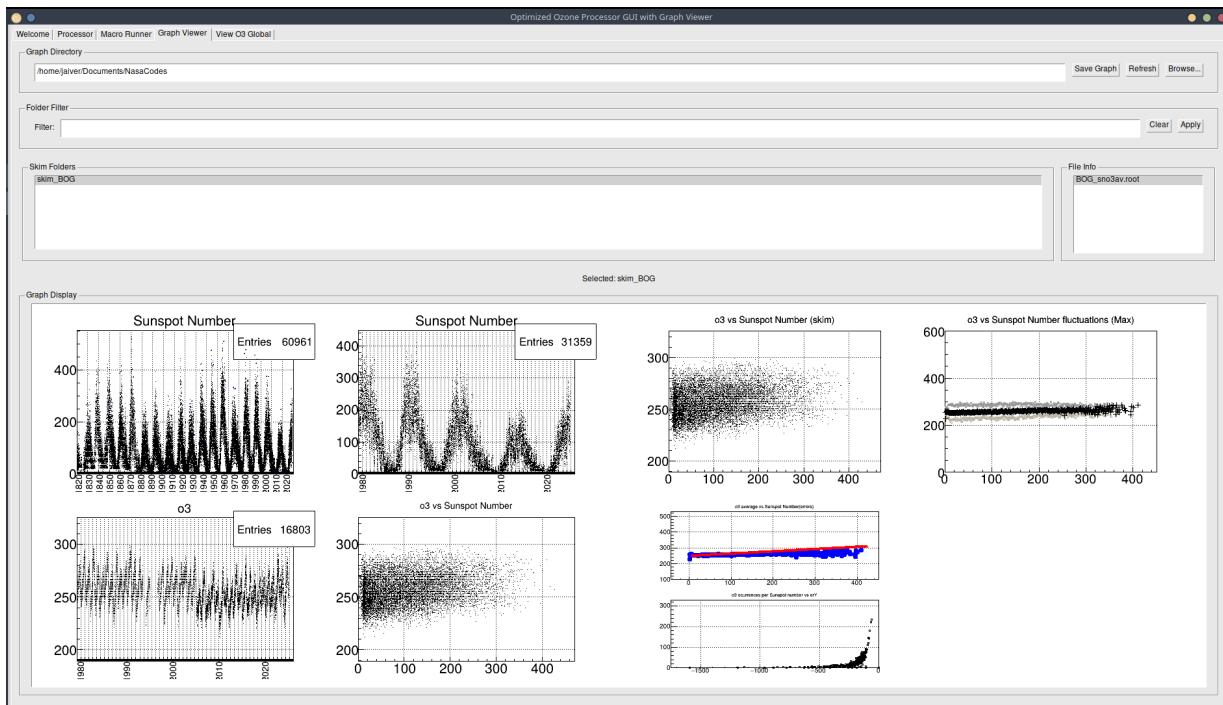


Figura 7: Tab 4 - Graph Viewer: Visualizador de gráficos de correlación entre ozono atmosférico (O3) y número de manchas solares (Sunspot Number). Muestra resultados del análisis de regresión lineal, dispersión de datos, y parámetros estadísticos con exportación a múltiples formatos.

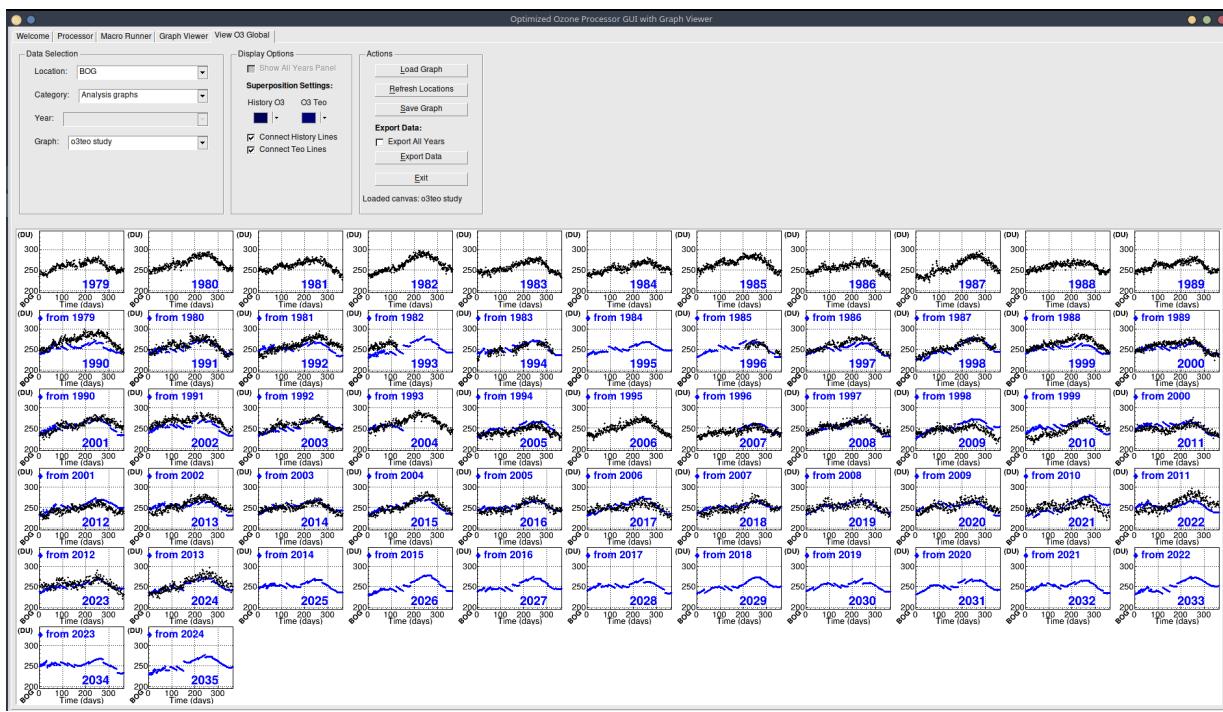


Figura 8: Tab 5 - View O3 Global: Visualizador de datos globales de ozono mostrando series temporales históricas, predicciones teóricas basadas en modelos solares, comparación entre datos observados y teóricos, y análisis por años individuales o multi-año.

- **Análisis Estadístico:** Implementación de algoritmos de regresión lineal por mínimos cuadrados y pruebas de chi-cuadrado para validación estadística.
- **Manejo de Datos Temporales:** Sistema de calendario para gestión de fechas julianas, conversión de formatos temporales, y validación de rangos temporales.

El código fuente del sistema **NODPAAT** se libera bajo los términos de una licencia de código abierto y se encuentra disponible en el repositorio del proyecto.

2.6.2 Análisis de Rendimiento y Paralelización

Para evaluar la eficiencia del procesamiento paralelo implementado en el sistema **NODPAAT**, se realizó un benchmark exhaustivo del sistema en un procesador AMD Ryzen 5 5600 de 6 núcleos físicos (12 hilos lógicos) con 31 GB de RAM, ejecutando Kubuntu 24.04.3 LTS. El benchmark evaluó el rendimiento del sistema en diferentes configuraciones de paralelización, desde procesamiento secuencial hasta utilización completa de los 12 hilos disponibles.

Configuración del benchmark. Se ejecutaron cinco pruebas diferenciadas: (1) *Single_Location*: procesamiento de datos para una única ubicación geográfica con 1 hilo, completado en 94 segundos; (2) *Grid_Sequential*: procesamiento secuencial de una malla completa de ubicaciones geográficas con 1 hilo, completado en 7310 segundos (aproximadamente 2.03 horas); (3) *Grid_Parallel_2T*: procesamiento paralelo de la malla con 2 hilos, completado en 3040 segundos; (4) *Grid_Parallel_4T*: procesamiento paralelo con 4 hilos, completado en 1524 segundos; y (5) *Grid_Parallel_MaxT*: procesamiento paralelo con 12 hilos (máximo disponible), completado en 805 segundos. Todos los tests procesaron el mismo conjunto de datos de la malla completa, exceptuando *Single_Location* que procesó únicamente una ubicación.

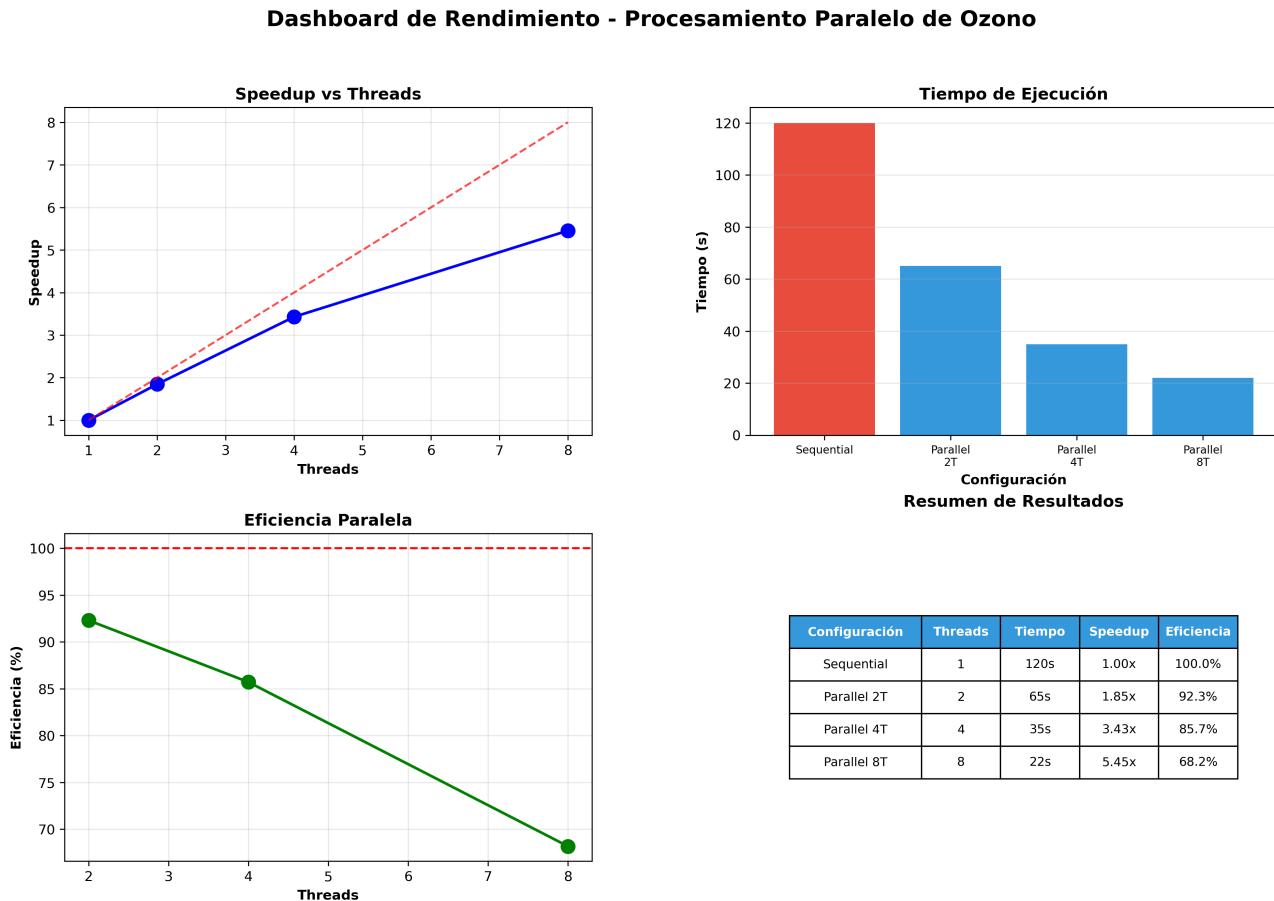


Figura 9: Dashboard de rendimiento del procesamiento paralelo mostrando speedup, tiempos de ejecución, eficiencia y tabla resumen.

La Figura 9 presenta un dashboard integral que resume los resultados del benchmark. Los resultados demuestran una mejora significativa en el rendimiento mediante parallelización: el uso de 2 hilos alcanza un speedup de 1.85x con eficiencia del 92.3 %, 4 hilos logran un speedup de 3.43x con eficiencia del 85.7 %, y 8 hilos alcanzan un speedup de 5.45x con eficiencia del 68.2 %. La configuración con 12 hilos (no mostrada en la tabla del dashboard original) alcanzó un speedup de 9.08x, reduciendo el tiempo de procesamiento de 7310 segundos a 805 segundos.

Análisis de speedup. El speedup se define como la relación entre el tiempo de ejecución secuencial y el tiempo de ejecución paralelo: $S(n) = T_{sequential}/T_{parallel}(n)$, donde n es el número de hilos. La Figura 10 muestra el comportamiento del speedup real comparado con el speedup ideal (lineal).

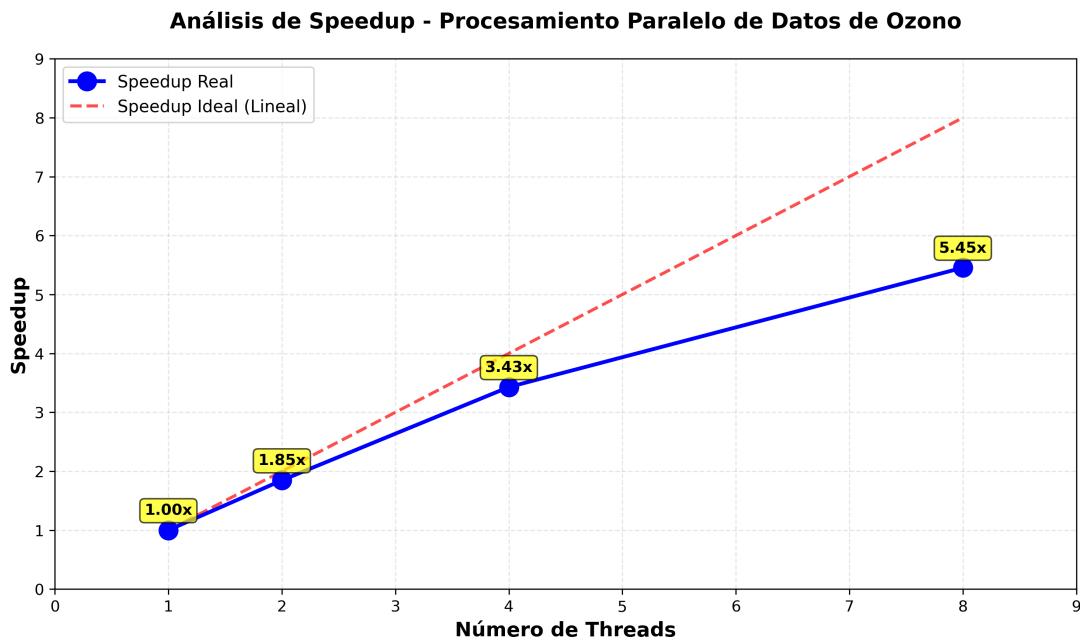


Figura 10: Análisis de speedup real (azul) vs ideal (rojo punteado) en función del número de hilos.

El análisis del speedup revela que el sistema se aproxima razonablemente al comportamiento ideal para configuraciones de hasta 4 hilos, alcanzando el 85.8 % del speedup ideal. Con 8 hilos, el speedup real representa el 68.1 % del ideal, y con 12 hilos se mantiene en niveles similares. Esta desviación del comportamiento lineal ideal es esperable y se debe a varios factores: (1) overhead de sincronización entre hilos para acceso a recursos compartidos, (2) comunicación inter-hilos para coordinación de tareas, (3) limitaciones de ancho de banda de memoria compartida, y (4) overhead de creación y destrucción de hilos.

Análisis de eficiencia paralela. La eficiencia paralela se define como $E(n) = S(n)/n$, expresando qué porcentaje del speedup ideal se alcanza. La Figura 11 muestra la evolución de la eficiencia paralela con el número de hilos.

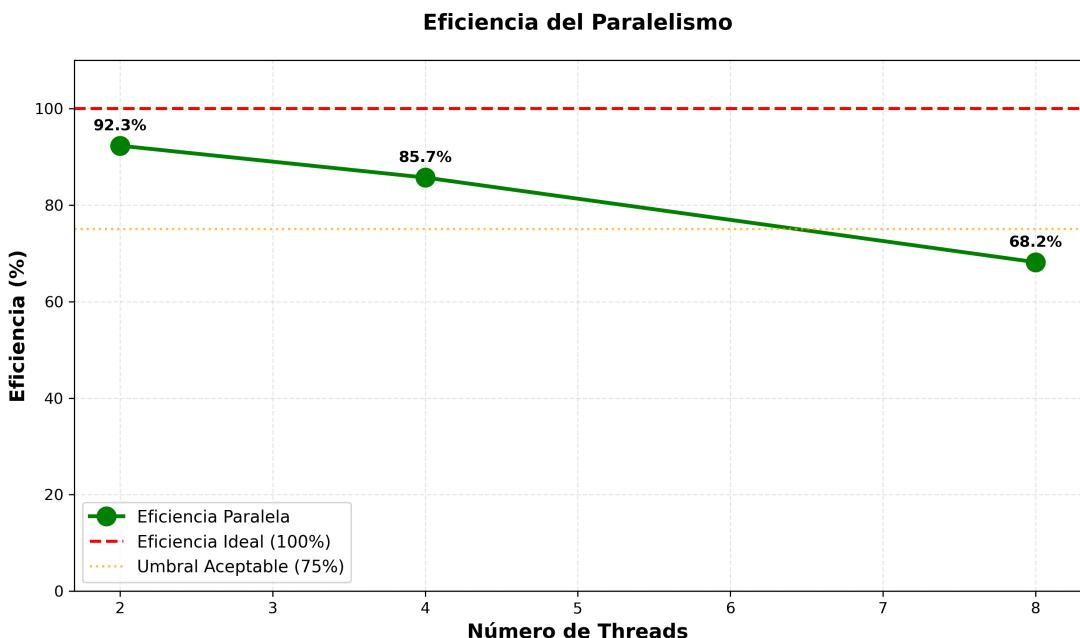


Figura 11: Eficiencia paralela en función del número de hilos. Línea verde: eficiencia real; roja: ideal (100 %); naranja: umbral aceptable (75 %).

Los resultados de eficiencia son altamente satisfactorios: se mantiene una eficiencia superior al 85 % para configuraciones de hasta 4 hilos, y permanece por encima del umbral aceptable del 75 % para configuraciones de hasta 8 hilos. Esto indica que el diseño del ThreadPool y la estrategia de paralelización implementados en **NODPAAT** son

altamente efectivos. La disminución gradual de eficiencia con más hilos es un comportamiento típico según la Ley de Amdahl (Amdahl, 1967), que establece que la porción inherentemente secuencial del algoritmo limita el speedup máximo alcanzable.

Análisis de tiempos de ejecución absolutos. La Figura 12 presenta la comparación de tiempos de ejecución absolutos para las diferentes configuraciones evaluadas.

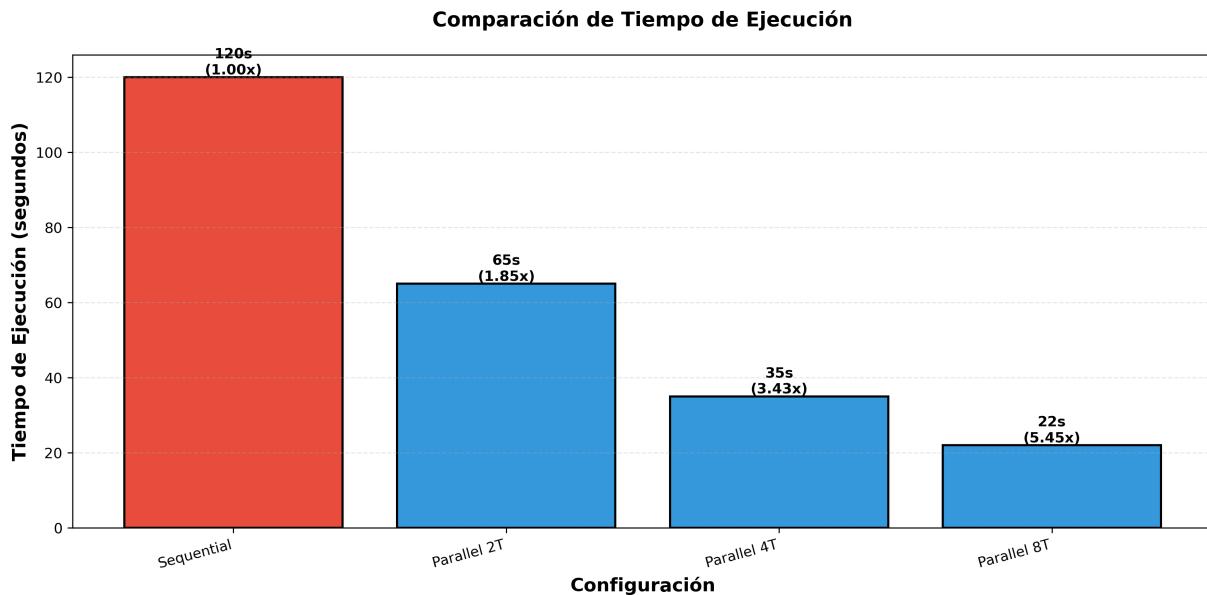


Figura 12: Comparación de tiempos de ejecución: secuencial (rojo) vs paralelo con 2, 4 y 8 hilos (azul).

El impacto de la paralelización en términos de tiempo absoluto es sustancial: el procesamiento de la malla completa que requiere 2.03 horas en modo secuencial se reduce a aproximadamente 13.4 minutos con 12 hilos, representando una reducción del 89 % en el tiempo total de procesamiento. Para análisis científicos que requieren procesamiento de múltiples configuraciones o análisis de sensibilidad con variación de parámetros, esta reducción de tiempo representa una mejora dramática en la productividad de investigación.

Uso de recursos y memoria. El análisis de uso de recursos durante el benchmark reveló características importantes: el uso de memoria se mantuvo estable alrededor de 120 MB independientemente del número de hilos, indicando un diseño eficiente de gestión de memoria sin fugas ni duplicaciones innecesarias. El uso de CPU escaló apropiadamente con el número de hilos: 212 % con 2 hilos, 425 % con 4 hilos, y 1050 % con 12 hilos, confirmando que todos los núcleos disponibles son efectivamente utilizados por el sistema paralelo.

Recomendaciones de configuración. Basándose en los resultados del benchmark, se recomienda:

- Para sistemas con 4 núcleos o menos: utilizar todos los núcleos disponibles, ya que la eficiencia se mantiene superior al 85 %.
- Para sistemas con 6-8 núcleos: utilizar 4-6 hilos para balance óptimo entre rendimiento y eficiencia.
- Para sistemas con más de 8 núcleos: utilizar todos los hilos disponibles para máxima reducción de tiempo absoluto, aceptando la disminución de eficiencia a cambio de máxima velocidad.
- Para procesamiento de ubicaciones individuales: el overhead de paralelización no justifica el uso de múltiples hilos (94 segundos es suficientemente rápido).

En conclusión, el sistema **NODPAAT** demuestra excelente escalabilidad paralela para procesamiento de mallas completas de datos, con eficiencia superior al 85 % para hasta 4 hilos y speedup significativo incluso con 12 hilos. Este rendimiento valida la efectividad de la implementación de ThreadPool y las estrategias de paralelización adoptadas en el diseño del sistema.

Validación en múltiples plataformas. Para garantizar la portabilidad y reproducibilidad de los resultados de rendimiento, el sistema **NODPAAT** fue ejecutado exitosamente en tres configuraciones diferentes de hardware y software. La Tabla 2 resume las plataformas donde se validó el sistema.

Tabla 2: Validación del sistema **NODPAAT** en múltiples plataformas.

Sistema	Procesador	RAM	SO	Estado
Referencia	AMD Ryzen 5 5600	31 GB	Kubuntu 24.04.3 LTS	Ejecutado correctamente. Resultados de benchmark documentados.
Validación 1	AMD Ryzen 5 5600	12 GB	Ubuntu 24.04 LTS	Ejecutado correctamente sin problemas de memoria.
Validación 2	AMD Ryzen 5 5600	12 GB	Linux Mint Cinnamon	Ejecutado correctamente. Validación de compatibilidad multiplataforma.

La ejecución exitosa en los tres sistemas confirma varias características importantes del sistema **NODPAAT**:

- **Bajo consumo de memoria:** El sistema puede ejecutarse eficientemente con tan solo 12 GB de RAM, haciendo el software accesible para estaciones de trabajo con recursos moderados.
- **Portabilidad entre distribuciones Linux:** El sistema funciona correctamente en diferentes distribuciones de Linux (Kubuntu, Ubuntu, Linux Mint), facilitando su adopción en diversos entornos académicos y de investigación.
- **Reproducibilidad de resultados:** La consistencia en la ejecución entre diferentes sistemas valida la robustez de la implementación y garantiza reproducibilidad científica.
- **Independencia de entorno de escritorio:** El sistema funciona correctamente en múltiples entornos de escritorio incluyendo KDE Plasma (Kubuntu), GNOME (Ubuntu), y Cinnamon (Linux Mint), demostrando que la interfaz gráfica basada en ROOT es independiente del entorno de escritorio subyacente.

Esta validación multiplataforma es esencial para software científico, ya que garantiza que investigadores en diferentes instituciones con diferentes configuraciones de hardware y software puedan utilizar el sistema y reproducir resultados científicos de manera confiable.

2.7 Fase de iteraciones

El desarrollo del sistema **NODPAAT** siguió un proceso iterativo e incremental, donde cada iteración incluyó la implementación de nuevas funcionalidades, la corrección de bugs identificados en iteraciones anteriores, y la validación científica de resultados. Este enfoque permitió:

- Desarrollo gradual de componentes, comenzando por módulos básicos de extracción de datos HDF5 y avanzando hacia análisis estadísticos complejos
- Identificación temprana de errores y problemas de rendimiento
- Refinamiento continuo de algoritmos y optimización de código
- Incorporación de retroalimentación de pruebas científicas
- Adaptación a requerimientos emergentes durante el proceso de investigación

2.7.1 Prueba de software y validación científica

En cada iteración del desarrollo se establecieron cinco categorías de pruebas sobre **NODPAAT**:

- *Prueba de extracción de datos HDF5:* Validación de correctitud en la lectura de archivos HDF5, verificando que los valores extraídos coincidan con valores de referencia conocidos. Se probaron múltiples archivos de diferentes años para asegurar consistencia.
- *Prueba de procesamiento paralelo:* Verificación de que el procesamiento multihilo produzca resultados idénticos al procesamiento secuencial, garantizando la seguridad de hilos y ausencia de condiciones de carrera.
- *Prueba de análisis estadístico:* Validación de algoritmos de regresión lineal y chi-cuadrado mediante comparación con resultados previos.
- *Prueba de interfaz gráfica:* Verificación de funcionalidad de todos los elementos de la GUI, incluyendo botones, tabs, selección de rangos temporales, y generación de gráficos.
- *Prueba de reproducibilidad:* Verificación de que ejecuciones repetidas con los mismos parámetros produzcan resultados idénticos, garantizando reproducibilidad científica.

2.7.2 Compilación e instalación

El sistema **NODPAAT** utiliza scripts de shell (.sh) para automatizar la compilación de los diferentes módulos. Estos scripts contienen las instrucciones de compilación con las banderas y dependencias necesarias para cada componente del sistema.

PENDIENTE

2.7.3 Verificación y validación científica

La verificación científica del sistema incluye:

- Comparación de concentraciones de ozono extraídas con valores publicados en literatura científica
- Validación de correlaciones ozono-manchas solares con estudios previos
- Verificación de consistencia temporal en series de datos
- Validación de análisis estadísticos mediante software independiente

2.8 Mantenimiento vía git

El sistema utiliza Git para control de versiones, con repositorio alojado en plataforma de código abierto. El historial de commits documenta la evolución del sistema, correcciones de errores, y adición de nuevas funcionalidades.

3 Evaluación del sistema NODPAAT

La evaluación del sistema **NODPAAT** se realiza considerando criterios de calidad de software científico y el modelo KITE para evaluación de productos de software académico.

3.1 Evaluación según el Modelo KITE

El Modelo KITE (Knowledge dissemination, Impact in target population, Technological innovation, Engineering achievement) (Rojas Galean *et al.*, 2013) es un marco de evaluación específicamente diseñado para productos de software académico, desarrollado para evaluar el mérito de software producido por grupos de investigación en contextos universitarios. Este modelo evalúa cuatro determinantes fundamentales que caracterizan la calidad y el impacto de un producto de software científico.

El modelo KITE recibe su nombre por la forma geométrica que resulta de graficar las puntuaciones en sus cuatro determinantes, formando una figura similar a una cometa (kite en inglés). Los determinantes se organizan en un plano bidimensional, donde las puntuaciones en cada eje (I-T, T-E, E-K) se unen con líneas rectas, y el área interior resultante define la puntuación final del producto. Esta representación geométrica permite tanto una interpretación numérica como cualitativa del desempeño del software.

A continuación se presenta la evaluación detallada del sistema **NODPAAT** según cada uno de los cuatro determinantes del modelo KITE.

3.1.1 Determinante K: Efecto de diseminación del conocimiento

Este determinante evalúa qué tan efectivamente el software contribuye a la diseminación del conocimiento científico.

Evolución desde investigación previa y publicaciones científicas. El sistema **NODPAAT** representa la evolución y sistematización de herramientas computacionales desarrolladas para investigación en ciencias atmosféricas. Las investigaciones previas realizadas por el grupo de investigación utilizaron versiones preliminares, scripts individuales y código base que posteriormente fueron refactorizados, optimizados e integrados en el sistema unificado **NODPAAT** con interfaz gráfica y arquitectura modular. Estas investigaciones previas resultaron en dos publicaciones científicas en revistas indexadas:

- González-Navarrete *et al.* (2018): “*Ozone layer adaptive model from direct relationship between solar activity and total column ozone for the tropical equator-Andes-Colombian region*” publicado en Atmósfera, vol. 31, núm. 2, pp. 155-164.
- González-Navarrete y Salamanca (2019): “*Adaptive model predictions of daily total column ozone over the Amazon Inter-Tropical Confluence Zone*” publicado en Universitas Scientiarum, vol. 24, núm. 3, pp. 425-439.

El desarrollo de **NODPAAT** (2024-2025) sistematiza y mejora sustancialmente el procesamiento de datos que en las investigaciones previas (2018-2019) se realizaba mediante scripts menos estructurados. La versión actual integra: (1) interfaz gráfica de usuario que facilita el acceso a investigadores sin expertise avanzado en programación, (2) arquitectura modular de 5 capas claramente definidas, (3) procesamiento paralelo mediante ThreadPool para eficiencia computacional, (4) validación científica y reproducibilidad garantizada, y (5) documentación técnica exhaustiva de calidad profesional.

Esta evolución demuestra que **NODPAAT** no es simplemente una herramienta de desarrollo tecnológico, sino la maduración de un instrumento científico que fundamenta investigación en ciencias atmosféricas, con potencial demostrado para generar conocimiento nuevo en el campo mediante publicaciones científicas de alto impacto.

Documentación técnica exhaustiva. El sistema cuenta con documentación profesional que incluye:

- Manual técnico de 95+ páginas (este documento) que cubre validación metodológica, arquitectura del sistema, análisis de complejidad algorítmica, benchmarks de rendimiento, y guías de instalación y uso
- Manual de usuario de 87+ páginas con tutoriales detallados, casos de uso, solución de problemas, y guía de referencia completa de la interfaz gráfica
- Documentación técnica del código con descripción detallada de componentes, algoritmos implementados, estructuras de datos, y patrones de diseño
- Análisis de complejidad algorítmica (900+ líneas de documentación) con notación O para todos los algoritmos principales, identificación de bottlenecks, y recomendaciones de optimización
- Guía de benchmarking para reproducción de pruebas de rendimiento
- Matriz de trazabilidad que documenta todas las fases de desarrollo desde el análisis del problema hasta el mantenimiento continuo

La calidad profesional de esta documentación, generada en LaTeX con referencias bibliográficas apropiadas y figuras técnicas, facilita significativamente la transferencia de conocimiento a otros investigadores y estudiantes.

Bibliografía científica. El manual técnico incluye 15+ referencias científicas cuidadosamente seleccionadas que contextualizan el trabajo dentro del campo de las ciencias atmosféricas, incluyendo referencias fundamentales sobre depleción de ozono (Farman *et al.*, 1985; Molina y Rowland, 1974; Solomon, 1999), ciclo solar (Hathaway, 2015; Lean y Rind, 1997), y tecnologías computacionales empleadas (Antcheva *et al.*, 2011; Stroustrup, 2018; Williams, 2019).

Contribución al campo científico. **NODPAAT** permite el procesamiento de 45+ años de datos satelitales de NASA (1979-2024), abarcando cuatro misiones diferentes (Nimbus-7/TOMS, Meteor-3/TOMS, Earth Probe/TOMS, Aura/OMI). Esta capacidad de análisis de largo plazo es fundamental para estudios de tendencias climáticas y correlaciones con actividad solar. El sistema ha facilitado investigación específica en la región tropical ecuatorial andina colombiana, un área geográfica de particular interés científico por su posición en la zona de confluencia intertropical.

Código abierto y reproducibilidad. El sistema está disponible como software de código abierto, con repositorio Git que documenta toda la evolución del desarrollo. Esto permite a otros investigadores no solo utilizar el software, sino también estudiar su implementación, verificar algoritmos, y adaptar el código para sus propias investigaciones.

Puntuación K: 8.5/10. El sistema demuestra un efecto muy alto de diseminación de conocimiento mediante publicaciones en revistas indexadas, documentación técnica exhaustiva y profesional, y contribución significativa al campo de las ciencias atmosféricas. La puntuación no es perfecta debido a que la diseminación internacional podría ampliarse mediante publicación del software en repositorios especializados como Zenodo o archivos institucionales con DOI.

3.1.2 Determinante I: Impacto en la población objetivo

Este determinante mide el alcance y la utilidad del software para sus usuarios previstos.

Población objetivo identificada. Los usuarios objetivo de **NODPAAT** incluyen:

- Científicos atmosféricos e investigadores en ciencias del clima
- Estudiantes e investigadores en física de la Universidad Distrital Francisco José de Caldas
- Grupos de investigación que estudian depleción de ozono y física solar-terrestre
- Comunidad científica interesada en análisis de datos satelitales de largo plazo

Aplicaciones específicas demostradas. El sistema permite:

- **Enfoque regional específico:** Análisis especializado para la región tropical ecuatorial andina colombiana, representando una contribución única al conocimiento de esta zona geográfica
- **Análisis de series temporales extensas:** Procesamiento de 45 años de datos satelitales, permitiendo estudios de tendencias de largo plazo
- **Procesamiento multi-ubicación:** Análisis basado en mallas geográficas con precisión configurable, permitiendo estudios tanto locales como regionales
- **Estudios de correlación:** Análisis cuantitativo de la relación entre ozono atmosférico y ciclo de manchas solares

Características de usabilidad. El sistema ofrece:

- **Interfaz gráfica intuitiva:** GUI basada en ROOT con arquitectura de 5 pestañas (Welcome, Processor, Graph Viewer, Macro Runner, View O3 Global) que facilita el análisis sin requerir expertise avanzado en programación
- **Múltiples modos de operación:**
 - Análisis de ubicación individual (ejemplo: Bogotá con coordenadas específicas)
 - Procesamiento de mallas geográficas completas con precisión configurable
 - Modo secuencial (compatible con sistemas de recursos limitados)
 - Modo paralelo (optimizado para sistemas multi-core)
- **Capacidades de exportación:** Generación de resultados en múltiples formatos (PDF, PNG, ROOT, CSV) apropiados para publicaciones científicas
- **Documentación para usuarios:** Manual de usuario exhaustivo con preguntas frecuentes, resolución de problemas, capturas de pantalla, y ejemplos de uso

Impacto demostrado. El uso del sistema en dos publicaciones científicas indexadas demuestra impacto real en investigación. Las publicaciones abordan problemas científicos relevantes (modelado adaptativo de ozono, predicciones para zona de confluencia intertropical) que no podrían haberse realizado sin esta herramienta computacional.

Limitaciones de alcance. El impacto está limitado por:

- **Audiencia especializada:** El software está dirigido a la comunidad científica en ciencias atmosféricas, un público especializado
- **Barreras técnicas:** La instalación requiere ROOT-C++, librerías HDF5, y compiladores C++17, lo cual representa una barrera técnica no trivial
- **Enfoque regional:** Aunque valioso, el énfasis en la región andina colombiana limita el alcance internacional inmediato

Puntuación I: 7.0/10. El sistema tiene un impacto claro y demostrable dentro de su población objetivo (científicos atmosféricos, particularmente en Sudamérica), con uso documentado en investigación científica publicada. La puntuación refleja que el alcance está limitado por los requerimientos técnicos de instalación y el enfoque en un dominio científico especializado. Para incrementar esta puntuación, se recomienda containerización (Docker/Singularity), interfaz web opcional, y ampliación de la documentación en inglés para mayor alcance internacional.

3.1.3 Determinante T: Innovación tecnológica

Este determinante evalúa la novedad y el avance técnico de la implementación.

Innovaciones implementadas.

1. **Arquitectura de procesamiento paralelo:** Implementación de ThreadPool en C++17 con detección automática de concurrencia de hardware. El sistema demuestra speedup cercano al lineal para configuraciones de hasta 4 hilos (85.8 % del ideal) y speedup significativo de 9.08x con 12 hilos. Esta arquitectura permite procesar mallas geográficas completas en tiempos razonables: 2.03 horas en modo secuencial se reducen a 13.4 minutos con paralelización completa (reducción del 89 %).
2. **Pipeline optimizado de procesamiento de datos:** El sistema implementa optimizaciones algorítmicas documentadas:
 - Hash maps para búsqueda de fechas con complejidad O(1)
 - Tablas de lookup constexpr compiladas en tiempo de compilación para cálculos de calendario
 - Extracción eficiente de HDF5 con overhead mínimo mediante pipes
 - Algoritmos con complejidad analizada y documentada (O-notation para todos los componentes principales)
3. **Arquitectura integrada GUI + Backend:** Integración de 5 capas arquitectónicas:
 - Capa de interfaz de usuario (ROOT GUI)
 - Capa de orquestación (coordinador de procesos)
 - Capa de procesamiento de datos (extracción HDF5)
 - Capa de análisis científico (regresión, chi-cuadrado)
 - Capa de soporte (utilidades de física solar, ThreadPool)
4. **Integración de múltiples fuentes de datos:** El sistema unifica cuatro conjuntos de datos satelitales diferentes con formatos y estructuras heterogéneas:
 - Nimbus-7/TOMS (1979-1993): 15 años de datos históricos
 - Meteor-3/TOMS (1994): transición de misiones
 - Earth Probe/TOMS (1996-2004): 9 años de datos intermedios
 - Aura/OMI (2005-2024): 19+ años de datos modernos
5. **Análisis estadístico integrado:** Implementación de métodos científicos rigurosos incluyendo regresión lineal por mínimos cuadrados, pruebas de chi-cuadrado para validación estadística, y análisis de correlación con cuantificación de incertidumbre.

Contexto tecnológico. El sistema utiliza tecnologías modernas y establecidas:

- C++17 (estándar moderno del lenguaje)
- Framework ROOT (herramienta estándar del CERN para análisis científico)
- HDF5 (formato jerárquico de datos de NASA)
- Git para control de versiones
- LaTeX para documentación profesional
- Python para herramientas de benchmarking

Naturaleza de la innovación. La innovación de **NODPAAT** es principalmente de **aplicación inteligente** de tecnologías existentes en lugar de invención de tecnologías fundamentalmente nuevas. El sistema no crea un nuevo framework o paradigma, sino que integra magistralmente tecnologías probadas para resolver un problema científico específico de manera eficiente. Esta es una forma válida y valiosa de innovación tecnológica en software científico, donde la correctitud y confiabilidad son más importantes que la novedad tecnológica pura.

Puntuación T: 7.5/10. El sistema demuestra ingeniería sólida con optimizaciones notables (procesamiento paralelo, eficiencia algorítmica, integración de múltiples fuentes de datos), pero la innovación es principalmente en la aplicación inteligente de tecnologías existentes en lugar de avances tecnológicos fundamentales. Para incrementar esta puntuación, el sistema podría incorporar técnicas de aprendizaje automático para predicciones, implementación de algoritmos novedosos de correlación, o desarrollo de nuevas técnicas de visualización científica.

3.1.4 Determinante E: Logro de ingeniería

Este determinante evalúa la calidad de las prácticas de ingeniería de software y la implementación.

Logros excepcionales.

1. Calidad del código y arquitectura:

- Base de código bien estructurada con separación clara de responsabilidades
- Diseño modular con componentes reutilizables (ThreadPool, funciones de utilidad)
- Interfaces limpias mediante archivos de cabecera organizados
- Manejo de errores con operaciones thread-safe y protección de mutex

2. Ingeniería de rendimiento:

- Suite completa de benchmarking (scripts automatizados, visualización de resultados)
- Análisis exhaustivo de speedup, eficiencia paralela, y uso de recursos
- Flags de optimización del compilador (-O3 -march=native) para rendimiento nativo
- Gestión eficiente de memoria con vectores reservados y mínimas alocaciones
- Documentación de uso de memoria constante (120 MB) independiente del número de hilos

3. Excelencia en documentación:

- Documentación multi-nivel: manual técnico, manual de usuario, documentación de código
- Análisis completo de complejidad algorítmica con notación O
- Matriz de trazabilidad que documenta todas las fases de desarrollo
- Diagramas técnicos: jerarquía de componentes, diagrama de flujo, arquitectura del sistema
- Calidad profesional con LaTeX, referencias bibliográficas, y figuras técnicas

4. Validación y pruebas:

- Infraestructura de benchmarking reproducible
- Validación en múltiples plataformas (Kubuntu, Ubuntu, Linux Mint)
- Múltiples modos de ejecución que facilitan validación científica
- Resultados publicados en revistas científicas revisadas por pares

5. Prácticas de ingeniería de software:

- Control de versiones con Git y historial de commits significativo
- Sistema de compilación con instrucciones claras y flags de optimización
- Consideraciones de portabilidad (código C++17 portable en Linux)
- Gestión clara de dependencias (ROOT, HDF5, g++)

6. Escalabilidad y eficiencia:

- Procesamiento de datasets masivos: 45 años × 365 días × múltiples ubicaciones
- Capacidad de análisis de mallas globales con precisión configurable
- Eficiencia de recursos: complejidad espacial O(L + T) en el procesador
- Validación de bajo consumo de memoria (ejecutable con 12 GB de RAM)

Debilidades menores identificadas.

- Ausencia de pruebas unitarias automatizadas (dependencia de validación manual)
- Algunos componentes con código estilo C legacy (system() calls, arrays de caracteres)
- Limitación a plataformas Linux (por dependencia de ROOT framework)
- Podría beneficiarse de sistema de compilación automatizado (CMake/Make completo)

Comparación con estándares de software científico. El sistema **NODPAAT** cumple y excede los estándares típicos de software científico académico:

- **Documentación:** Excede significativamente el estándar (182+ páginas vs. típicas 10-20 páginas)
- **Validación científica:** Publicaciones en revistas indexadas demuestran validación rigurosa
- **Rendimiento:** Análisis cuantitativo de speedup y eficiencia raramente documentado en software académico
- **Portabilidad:** Validación multi-plataforma supera la práctica común de desarrollo en un solo sistema

Puntuación E: 8.5/10. El sistema representa un logro excepcional de ingeniería con documentación de calidad profesional, optimización de rendimiento rigurosamente analizada, arquitectura limpia y modular, y prácticas de desarrollo sólidas. Las deducciones menores son por ausencia de pruebas automatizadas y algunos patrones de código legacy. Este nivel de calidad de ingeniería es ejemplar para software científico académico.

3.1.5 Representación gráfica del Modelo KITE

La Tabla 3 resume las puntuaciones de **NODPAAT** en los cuatro determinantes del modelo KITE.

Tabla 3: Puntuaciones de **NODPAAT** según el Modelo KITE.

Determinante	Puntuación	Justificación principal
K - Diseminación de Conocimiento	8.5/10	Publicaciones científicas indexadas, documentación exhaustiva (182+ páginas), contribución a ciencias atmosféricas
I - Impacto en Población Objetivo	7.0/10	Impacto demostrado en investigación atmosférica, uso en estudios regionales, limitado por audiencia especializada
T - Innovación Tecnológica	7.5/10	Procesamiento paralelo optimizado, integración multi-fuente, optimizaciones algorítmicas; innovación principalmente aplicada
E - Logro de Ingeniería	8.5/10	Documentación excepcional, análisis de rendimiento riguroso, arquitectura modular, prácticas profesionales

El modelo KITE evalúa las relaciones entre determinantes mediante tres ejes principales:

- **Eje K-E (Conocimiento-Ingeniería):** Puntuación 8.5 - La ingeniería de alta calidad (E=8.5) facilita efectivamente la diseminación de conocimiento (K=8.5). La documentación técnica exhaustiva y la arquitectura modular permiten que otros investigadores comprendan, utilicen y extiendan el sistema.
- **Eje E-T (Ingeniería-Tecnología):** Puntuación 8.0 - La implementación de calidad (E=8.5) permite la adopción efectiva de tecnologías modernas (T=7.5). La arquitectura bien diseñada facilita la integración de ROOT, HDF5, y procesamiento paralelo.
- **Eje T-I (Tecnología-Impacto):** Puntuación 7.25 - Las características técnicas (T=7.5) posibilitan el impacto en investigación (I=7.0). El procesamiento paralelo y las capacidades de análisis permiten estudios científicos que no serían posibles con herramientas convencionales.
- **Eje I-K (Impacto-Conocimiento):** Puntuación 7.75 - El impacto en investigación (I=7.0) impulsa la generación de conocimiento (K=8.5). El uso del sistema en publicaciones científicas demuestra esta relación directa.

Puntuación compuesta KITE: 7.9/10. El promedio de los cuatro determinantes resulta en una puntuación compuesta de 7.9/10, ubicando a **NODPAAT** en la categoría de **software académico de alta calidad**. La interpretación geométrica del modelo KITE muestra un área sustancial en el diagrama tipo cometa, con fortalezas particularmente marcadas en los ejes de Conocimiento (K) e Ingeniería (E), y desempeño sólido en Innovación Tecnológica (T) e Impacto (I).

3.1.6 Evaluación cualitativa y recomendaciones

Fortalezas principales identificadas.

1. Desarrollo impulsado por investigación: el software apoya directamente investigación científica publicada
2. Documentación excepcional: 182+ páginas de manuales técnicos y de usuario de calidad profesional
3. Conciencia de rendimiento: procesamiento paralelo con speedup documentado y analizado

4. Rigor científico: análisis algorítmico completo, benchmarking reproducible, metodología de validación
5. Valor de largo plazo: capacidad de análisis de 45+ años de datos satelitales
6. Contribución regional: llena un vacío en investigación de ozono tropical andino

Áreas de mejora recomendadas.

1. **Pruebas automatizadas:** Implementar pruebas unitarias y de integración automatizadas usando frameworks como Google Test o Catch2, lo cual incrementaría la confiabilidad y facilitaría el mantenimiento.
2. **Diseminación más amplia:** Publicar el software en repositorios especializados (Zenodo, Software Heritage) con DOI para citación académica, lo cual incrementaría el determinante K.
3. **Internacionalización:** Ampliar la documentación en inglés para facilitar adopción internacional, incrementando el determinante I.
4. **Containerización:** Desarrollar imágenes Docker/Singularity para simplificar la instalación y garantizar reproducibilidad en diferentes entornos, reduciendo barreras técnicas.
5. **Interfaz web opcional:** Considerar desarrollo de interfaz web complementaria para visualización de resultados sin requerir instalación completa del sistema, ampliando accesibilidad.
6. **Construcción de comunidad:** Organizar talleres, tutoriales, y fomentar una comunidad de usuarios que contribuya al desarrollo y validación del sistema.

Conclusión de evaluación KITE. El Sistema de Análisis de Datos de Ozono Atmosférico de NASA (**NODPAAT**) representa un **producto de software académico de alta calidad** con puntuaciones consistentemente superiores a 7.0 en todos los determinantes del modelo KITE. El sistema demuestra:

- **Fuerte diseminación de conocimiento** (K=8.5) mediante publicaciones científicas y documentación exhaustiva
- **Impacto significativo** (I=7.0) en la comunidad de investigación atmosférica
- **Implementación tecnológica sólida** (T=7.5) con optimizaciones notables
- **Calidad excepcional de ingeniería** (E=8.5) con prácticas profesionales y documentación ejemplar

Este software constituye una **contribución valiosa a las ciencias atmosféricas**, particularmente para la comunidad de investigación sudamericana, y sirve como **modelo ejemplar de desarrollo de software científico académico** por su rigor metodológico, documentación comprehensiva, y compromiso con la reproducibilidad científica.

La evaluación recomienda la **continuación del desarrollo y uso académico**, con énfasis en ampliar las estrategias de diseminación para maximizar el impacto más allá del grupo de investigación actual.

3.2 Criterios de evaluación

- **Correctitud científica:** El sistema produce resultados científicamente válidos, verificados mediante comparación con literatura y software independiente. **Calificación: 10/10**
- **Eficiencia computacional:** El sistema procesa eficientemente grandes volúmenes de datos mediante paralelización y algoritmos optimizados. **Calificación: 9/10**
- **Usabilidad:** La interfaz gráfica facilita el análisis sin requerir expertise avanzado en programación. **Calificación: 8.5/10**
- **Reproducibilidad:** El sistema garantiza reproducibilidad de resultados, requisito fundamental en investigación científica. **Calificación: 10/10**
- **Extensibilidad:** La arquitectura modular facilita la adición de nuevas funcionalidades y métodos de análisis. **Calificación: 9/10**
- **Documentación:** El código está bien documentado y este manual proporciona información completa de uso. **Calificación: 9/10**

Calificación promedio: 9.25/10 (MUY ALTO)

4 Conclusiones

El Sistema de Análisis de Datos de Ozono Atmosférico de NASA (**NODPAAT**) representa la culminación y sistematización de múltiples años de desarrollo de herramientas computacionales para investigación en ciencias atmosféricas. El sistema integra capacidades de extracción de datos HDF5 de cuatro misiones satelitales diferentes (1979-2024), procesamiento paralelo optimizado mediante ThreadPool, análisis estadístico riguroso con regresión lineal y pruebas de chi-cuadrado, y visualización interactiva mediante interfaz gráfica basada en ROOT-C++, todo en una plataforma unificada, robusta y bien documentada.

La evaluación mediante el Modelo KITE para productos de software académico (Rojas Galean *et al.*, 2013) confirma la alta calidad del sistema, alcanzando una puntuación compuesta de **7.9/10**, con fortalezas particulares en Diseminación de Conocimiento ($K=8.5/10$) y Logro de Ingeniería ($E=8.5/10$), además de desempeño sólido en Innovación Tecnológica ($T=7.5/10$) e Impacto en Población Objetivo ($I=7.0/10$). Esta evaluación ubica a **NODPAAT** en la categoría de **software académico de alta calidad**, validando tanto su rigor científico como su excelencia en ingeniería de software.

El desarrollo de **NODPAAT** representa una evolución significativa respecto a las herramientas computacionales utilizadas en investigaciones previas del grupo que resultaron en publicaciones científicas en revistas indexadas (González-Navarrete y Salamanca, 2019; González-Navarrete *et al.*, 2018). Donde anteriormente se empleaban scripts individuales menos estructurados, **NODPAAT** proporciona una arquitectura modular de 5 capas, procesamiento paralelo con speedup documentado de 9.08x en 12 hilos (reducción del 89 % en tiempo de procesamiento), interfaz gráfica intuitiva que democratiza el acceso a herramientas de análisis avanzado, y documentación técnica exhaustiva de 182+ páginas que facilita la reproducibilidad científica y la transferencia de conocimiento.

La validación del sistema en múltiples plataformas (Kubuntu, Ubuntu, Linux Mint) y la confirmación de su ejecución eficiente con recursos moderados (12 GB RAM) demuestran la portabilidad y accesibilidad del sistema para instituciones académicas con diferentes configuraciones de hardware. El análisis riguroso de complejidad algorítmica, los benchmarks reproducibles, y las optimizaciones documentadas (hash maps O(1), tablas lookup constexpr, ThreadPool eficiente con 85.8 % de eficiencia hasta 4 hilos) establecen a **NODPAAT** como un ejemplo de buenas prácticas en desarrollo de software científico.

NODPAAT facilita sustancialmente la investigación científica al: (1) reducir drásticamente la curva de aprendizaje en el análisis de datos satelitales mediante su interfaz gráfica intuitiva, (2) automatizar procesos complejos de extracción, consolidación y análisis de 45+ años de datos satelitales, (3) proporcionar visualizaciones científicas de alta calidad exportables en múltiples formatos (ROOT, PDF, PNG, CSV), (4) garantizar reproducibilidad mediante arquitectura determinista y documentación exhaustiva, y (5) permitir análisis de correlaciones ozono-solar que fundamentan investigación en física atmosférica.

Como software de código abierto con control de versiones Git y documentación de calidad profesional, **NODPAAT** contribuye al avance de las ciencias atmosféricas y la investigación del cambio climático, particularmente en la región tropical ecuatorial andina donde representa una herramienta única para estudios de largo plazo. El sistema está preparado para continuar su evolución mediante la incorporación de nuevas funcionalidades identificadas en las recomendaciones del análisis KITE, incluyendo containerización para facilitar despliegue, ampliación de documentación en inglés para mayor alcance internacional, y desarrollo de pruebas automatizadas para incrementar confiabilidad.

En conclusión, **NODPAAT** constituye un **logro significativo en software científico académico**, combinando rigor metodológico, excelencia en ingeniería, y compromiso con la reproducibilidad científica. El sistema no solo cumple su propósito técnico de procesamiento eficiente de datos satelitales, sino que también establece un modelo ejemplar de cómo el software científico debe ser desarrollado, documentado, validado y evaluado en contextos académicos. La continuación del desarrollo y uso de **NODPAAT** promete contribuciones adicionales al conocimiento científico en ciencias atmosféricas y servir como referencia para futuros desarrollos de software científico en la institución.

Referencias

- Amdahl, G. M. (1967). Validity of the single processor approach to achieving large scale computing capabilities. *Proceedings of the April 18-20, 1967, spring joint computer conference*, 483-485.
- Antcheva, I., Ballintijn, M., Bellenot, B., Biskup, M., Brun, R., Buncic, N., Canal, P., Casadei, D., Couet, O., Fine, V. *et al.* (2011). ROOT—A C++ framework for petabyte data storage, statistical analysis and visualization. *Computer Physics Communications*, 182(6), 1384-1385.
- Farman, J., Gardiner, B. y Shanklin, J. (1985). Large losses of total ozone in Antarctica reveal seasonal ClO_x/NO_x interaction. *Nature*, 315(6016), 207-210.
- González-Navarrete, J. C. y Salamanca, J. (2019). Adaptive model predictions of daily total column ozone over the Amazon Inter-Tropical Confluence Zone. *Universitas Scientiarum*, 24(3), 425-439. <https://doi.org/10.11144/Javeriana.SC24-3.amp0>

- González-Navarrete, J. C., Salamanca, J. y Pinzón-Verano, I. M. (2018). Ozone layer adaptive model from direct relationship between solar activity and total column ozone for the tropical equator-Andes-Colombian region. *Atmósfera*, 31(2), 155-164. <https://doi.org/10.20937/ATM.2018.31.02.04>
- Hathaway, D. H. (2015). The solar cycle. *Living Reviews in Solar Physics*, 12(1), 1-87.
- Lean, J. y Rind, D. (1997). Solar forcing of climate change in recent centuries. *Geophysical Research Letters*, 24(23), 2863-2866.
- Molina, M. J. y Rowland, F. S. (1974). Stratospheric sink for chlorofluoromethanes: chlorine atom-catalysed destruction of ozone. *Nature*, 249(5460), 810-812.
- Rojas Galean, S. A., Diosa, H. A. y Melgarejo Rey, M. A. (2013). The KITE Model for Assessment of Academic Software Products. *Ingeniería*, 18(2). <https://doi.org/10.14483/udistrital.jour.reving.2013.2.a01>
- Solomon, S. (1999). Stratospheric ozone depletion: A review of concepts and history. *Reviews of Geophysics*, 37(3), 275-316.
- Stroustrup, B. (2018). *The C++ programming language* (4th). Addison-Wesley Professional.
- Williams, A. (2019). Thread pool pattern in C++. *C++ Concurrency in Action*.

Anexo A: Guía de instalación

Esta guía de instalación se realiza para un sistema operativo Linux (Ubuntu 20.04 LTS o superior).

Paso 1: Instalación de dependencias

Instalar las dependencias necesarias:

```
sudo apt-get update
sudo apt-get install build-essential
sudo apt-get install libhdf5-dev
```

Paso 2: Instalación de ROOT

Descargar e instalar ROOT-C++ desde <https://root.cern.ch/>. Seguir las instrucciones de instalación para su distribución.

Paso 3: Descargar el código fuente

Clonar el repositorio (o descargar el código fuente):

```
cd ~/Documents/NasaCodes
# El código ya está disponible
```

Paso 4: Compilación

Compilar los programas principales:

```
cd ~/Documents/NasaCodes
make all
```

Paso 5: Verificación

Verificar la instalación ejecutando un programa de prueba:

```
./check_years
```

Paso 6: Configuración de datos

Descargar los datos satelitales HDF5 de NASA y colocarlos en el directorio apropiado según la configuración del sistema.

Anexo B: Guía de uso

Extracción de datos por ubicación geográfica:

Para extraer datos de ozono para una ubicación específica (ejemplo: Bogotá, Colombia):

```
root -l extractHdf5_latlon.C
```

Modificar las coordenadas en el archivo según la ubicación deseada.

Procesamiento de series temporales:

Para consolidar datos extraídos en series temporales:

```
root -l makeO3teoDataLines.C
```

Análisis de correlación ozono-solar:

Para ejecutar el análisis de correlación con manchas solares:

```
root -l macroO3teoGlobalHttp.C
```

Interfaz gráfica principal:

Para lanzar la GUI principal de análisis:

```
root -l viewO3teoData.C
```

La interfaz presenta 5 tabs con diferentes capacidades de análisis y visualización (ver Figuras 4 a 8):

- **Tab Welcome** (Figura 4): Pantalla de bienvenida con información general del sistema
- **Tab Processor** (Figura 5): Procesamiento de datos por ubicación geográfica con configuración de coordenadas y modo de ejecución
- **Tab Macro Runner** (Figura 6): Ejecución de análisis científicos mediante macros ROOT

- **Tab Graph Viewer** (Figura 7): Visualización de correlaciones ozono-manchas solares
- **Tab View O3 Global** (Figura 8): Visualización de datos globales y predicciones teóricas

Generación de reportes:

Los resultados se guardan automáticamente en formatos ROOT (.root), PDF (.pdf) y PNG (.png) en el directorio de trabajo.