

Ingeniería Inversa Aplicada a Software Numérico: Modelos Climáticos

Fernando G. Tinetti*, Pedro G. Cajaraville#, Juan C. Labraga##, Mónica A. López##, María G. Olguín#

#Departamento de Informática, Facultad de Ingeniería – (UNPSJB¹)

##Centro Nacional Patagónico (CENPAT-CONICET²)

III-LIDI, Facultad de Informática – (UNLP³)

fernando@info.unlp.edu.ar, gustavo.cajaraville@gmail.com, {labraga, monica}@cenpat.edu.ar, gabriela.olguin@gmail.com

CONTEXTO

Se presentan los avances realizados en el marco de la exposición efectuada con el mismo nombre en el WICC 2008 de General Pico, La Pampa. A su vez este trabajo forma parte del proyecto de investigación “Cómputo Optimizado y Paralelo en Clusters de PCs Aplicado a Modelos Numéricos del Clima”, que se está desarrollando en la Sede Puerto Madryn de la Universidad Nacional de la Patagonia San Juan Bosco. El objeto de estudio es el software heredado que implementa un modelo simulación de circulación general de la atmósfera.

RESUMEN

Los objetivos perseguidos fueron: recuperar el diseño del software que implementa al modelo; redocumentar las subrutinas incluidas en los niveles superiores del árbol de invocaciones y especificar los aspectos metodológicos que puedan ser utilizados para efectuar ingeniería inversa de otros modelos de simulación basados en el lenguaje Fortran. Se emplearon las técnicas de ingeniería inversa que se aplican para el análisis de software científico. Los principales resultados obtenidos fueron: a) Un conjunto de reglas para depuración de código quitando comentarios innecesarios y código anulado; b) La definición y aplicación de métricas para evaluar el volumen y la eficiencia de la depuración efectuada; c) La redocumentación del sistema mediante la generación de árboles de invocaciones partiendo de distintas subrutinas del sistema; d) La recopilación de información sobre el intercambio de datos entre subrutinas (uso de bloques common y pasaje de parámetros).

El modelo de referencia es en realidad el acoplamiento de un modelo atmosférico y un modelo oceánico. Estructuralmente el *modelo acoplado* es el *encargado* de invocar a rutinas de ambos modelos: atmosférico y oceánico, todo programado en Fortran 77. Se trata de un software heredado (legacy code) del que se contó con muy poca documentación sobre su diseño. Esta situación es bastante usual en el campo de los modelos numéricos, habitualmente implementados en programas de varios miles de líneas de código. Esto por supuesto dificulta las tareas de mantenimiento y cambios de funcionalidades, por lo que es muy conveniente la aplicación de ingeniería inversa para recuperar información del diseño del software.

Palabras Clave: ingeniería inversa, software heredado, modelado climático.

¹ Universidad Nacional de la Patagonia San Juan Bosco

² Centro Nacional Patagónico - Consejo Nacional de Investigaciones Científicas y Técnicas

³ Universidad Nacional de La Plata

* Investigador Adjunto sin Director, Comisión de Investigaciones Científicas de la Provincia de Buenos Aires

Profesor Titular, Facultad de Ingeniería, Universidad Nacional de la Patagonia San Juan Bosco

1. INTRODUCCIÓN

La comprensión del sistema climático es un problema de gran interés científico mundial. Si bien es cierto que se han realizado avances considerables en el tema, aún son muchos los factores que continúan limitando la capacidad de detectar, atribuir y comprender el cambio climático actual y proyectar los cambios climáticos que podrían ocurrir en el futuro (*Intergovernmental Panel on Climate Change* [1]). En la actualidad, las herramientas más confiables para la investigación del clima, sus fluctuaciones y variaciones, son los Modelos de Circulación General de la Atmósfera (MCGA).

Un MCGA es una representación numérica espacial y temporal aproximada de los principales procesos físicos que ocurren en la atmósfera y de las interacciones con los otros componentes del medio ambiente. Esencialmente, está constituido por un sistema de ecuaciones diferenciales en derivadas parciales que expresan los principios de conservación de la cantidad de movimiento, la energía termodinámica y la masa del sistema. Estas ecuaciones expresan las leyes físicas que gobiernan la atmósfera. Actualmente, los MCGA están integrados con modelos numéricos de funcionamiento de los océanos, la criosfera, y representaciones simplificadas de la biosfera (MCGA *acoplados*), lo que permite simular las interacciones de la atmósfera con su entorno, en distintas escalas de tiempo.

Mediante el uso de computadoras con gran capacidad de procesamiento y la aplicación de diversos métodos de cálculo numérico, pueden obtenerse soluciones numéricas aproximadas del sistema de ecuaciones de un MCGA. De este modo, se obtiene la evolución temporal y espacial del sistema climático (en forma tridimensional), en función de las condiciones iniciales y de contorno elegidas y de los valores de ciertos parámetros climáticos (ejemplo: concentración de CO₂ atmosférico). La solución numérica de un MCGA en condiciones preestablecidas se denomina *experimento climático*. Integrando el modelo atmosférico con una representación de la criósfera y una representación de los océanos, se obtienen pronósticos de cambios climáticos y anomalías en el clima. Se utiliza por lo general un método espectral, esto es, representación de capas atmosféricas.

1.1 El Modelo de Circulación General de la Atmósfera

El modelo de circulación general de la atmósfera analizado es uno de los MCGA utilizado por la comunidad científica internacional y participó junto con otros modelos del *Program for Climate Model Diagnosis and Intercomparison* (PCMDI) [7] [3] [4] [5].

En el año 1999 fue cedido, luego de un período de capacitación, a un grupo de investigación de la Unidad de Investigación de Oceanografía Física y Meteorología (UIOM) del CENPAT (Centro Nacional Patagónico)-CONICET. Las aplicaciones de este modelo en la UIOM han permitido avances en el conocimiento de temas cruciales para el desarrollo económico y social de la Argentina: la predictibilidad de anomalías climáticas estacionales, el impacto del calentamiento global sobre el ciclo hidrológico en la Argentina, el calentamiento global y los cambios climáticos regionales y el impacto del cambio climático en los recursos hídricos de la cordillera de los Andes. El sistema informático de pronóstico de anomalías climáticas basado en este modelo desarrollado por el equipo de la UIOM, publica mensualmente en forma experimental información objetiva sobre la magnitud y la probabilidad de condiciones atípicas del clima.

El código original del modelo fue desarrollado para supercomputadoras (CRAY, Silicon Graphics, etc.) con varios procesadores bajo sistema operativo UNIX. Para su utilización en la Argentina fue necesario portar el modelo en computadores de rango intermedio (Sun Ultra10, Enterprise) con un

solo procesador. En una etapa posterior se implementó también en la supercomputadora Cray Origin 2000 “Clementina II”, instalada en las dependencias de la Secretaría de Ciencia, Tecnología e Innovación Productiva (SeCyT) para uso científico. En la actualidad la ejecución de este modelo se implementa en Computadoras Personales (PC) con procesadores Intel, bajo sistema operativo Linux y consta de un programa principal y más de 250 subrutinas.

1.2 Ingeniería Inversa

La ingeniería inversa del software es el proceso consistente en analizar un programa, en un esfuerzo por crear una representación del mismo con un nivel de abstracción más elevado que el código fuente. Es un proceso de recuperación del diseño. Las herramientas a utilizar extraen información acerca de los datos, arquitectura y diseño de procedimientos de un programa ya existente. A diferencia de lo usual en otras disciplinas, donde este proceso se aplica a los productos de la competencia, en ingeniería de software con frecuencia se debe aplicar a los propios trabajos de la organización, realizados hace muchos años. En el caso de este proyecto, se aplica a un trabajo hecho por otra organización, pero que se intenta optimizar con varios objetivos: desde la optimización de rendimiento numérico hasta el cambio/mejora de funcionalidades.

Según va aumentando la abstracción crece la complejidad del trabajo, así como la necesidad de comprensión de la aplicación. La ingeniería inversa debe ser capaz de abstraer, a partir del código fuente, información significativa del procesamiento que se realiza, las estructuras de datos que se usan en el programa y el interfaz con el usuario que se utiliza.

Los sistemas heredados tienen características propias que dificultan su comprensión y mantenimiento. En caso de aplicaciones científicas, de grandes dimensiones y gran complejidad algorítmica, ante la necesidad de adaptar el software a nuevos requerimientos, se intenta mantener, dentro de lo posible, la mayoría de los elementos de software. La ingeniería inversa en estos casos es un proceso iterativo que involucra combinaciones de actividades automáticas y manuales.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Se definieron dos líneas de investigación: en forma específica, recuperar el diseño del software que implementa al modelo acoplado y mejorar su documentación, y en forma genérica, avanzar en la especificación de metodologías que puedan ser utilizadas para efectuar ingeniería inversa de código heredado basado en el lenguaje Fortran.

Para la comprensión de la estructura de los programas de cómputo secuencial que implementan el modelo acoplado océano-atmósfera, se utilizó un método de refinamiento sucesivo, acorde a lo planteado por Chikofsky en “Reverse Engineering and Design Recovery: A Taxonomy” [2], buscando identificar, en principio, bloques de subrutinas y sus relaciones funcionales, para luego analizar con mayor detalle las subrutinas más significativas, identificando entradas, salidas y procesos involucrados, procediendo luego a su redocumentación. En definitiva, se intentó comprender:

- la estructura del software y los datos que se manejan: bloques de memoria compartida, variables pasadas como parámetro, etc.;
- el procesamiento que se realiza en el código: invocaciones entre subrutinas (en particular cuando son de modelos diferentes) y la dinámica de activaciones y pasaje de datos entre modelos.

El proceso realizado alternó tareas automáticas y manuales. Incluyó una primera etapa de recopilación de información y capacitación, luego depuración del código fuente disponible, una etapa iterativa para el análisis de la comunicación entre los modelos atmosférico y oceánico, y finalmente una etapa de redocumentación, para organizar y entregar los resultados obtenidos

Las herramientas utilizadas fueron seleccionadas de entre varias encontradas en Internet para entorno Linux (preferentemente de uso libre), y orientadas a ingeniería inversa, análisis de sistemas o procesamiento de texto, como ser: Understand for Fortran: para ingeniería inversa; gprof: profiler, para análisis dinámico; KProf: para graficar la salida del gprof y realizar comparaciones entre distintas corridas del profiler; SourceNav: para recorrer el código fuente en las tareas de redocumentación; Graphviz: paquete para generar grafos jerárquicos; Perl: lenguaje de programación orientado a procesamiento de texto.

3. RESULTADOS OBTENIDOS/ESPERADOS

El objetivo principal planeado y cumplido de este proyecto fue recuperar el diseño del modelo climático acoplado océano-atmósfera en un nivel alto de abstracción, es decir:

- Diagramas de estructura de la comunicación entre los componentes de los modelos numéricos involucrados: acoplado, atmosférico y oceánico.
- Para cada modelo (acoplado, atmosférico y oceánico):
 - Árbol de invocación.
 - Listado de las rutinas utilizadas.
 - Listado de bloques *common* de Fortran 77.
 - Código re-documentado de las subrutinas.
- Implementación de reglas para automatizar el proceso de depuración del código fuente, que básicamente fueron:
 - Quitar líneas de comentario referidas a evolución histórica pero que no aportan información sobre diseño ni funcionalidad.
 - Quitar líneas de comentario que anulan código en desuso.

Si bien muchos de los resultados obtenidos en este proyecto estuvieron relacionados con el software específico en estudio, otros pueden aplicarse en forma genérica a sistemas heredados escritos en lenguaje Fortran 77:

- reglas generales aplicables a cualquier código Fortran 77 para depuración de código en desuso;
- definición de métricas para evaluar el volumen de código comentado en un sistema, y la eficiencia de su depuración;
- generación de árboles de invocaciones con representación gráfica de la estructura partiendo de distintas subrutinas del sistema;
- recopilación de información sobre el intercambio de datos entre subrutinas (uso de

bloques common y pasaje de parámetros), más la posibilidad de graficar o mostrar esos datos en una grilla.

4. FORMACIÓN DE RECURSOS HUMANOS

La consecución de estos objetivos enunciados anteriormente, permitió la consolidación del grupo de investigación como tal, avanzar en el conocimiento sobre modelos numéricos, ingeniería inversa de código científico heredado y cómputo paralelo.

5. BIBLIOGRAFÍA

[1] Intergovernmental Panel on Climate Change, Tercer Informe de Evaluación – Cambio Climático 2001 – Base Científica, disponible en <http://www.ipcc.ch/pub/un/ipccwg1s.pdf>

[2] Elliot J. Chikofsky, James H. Cross II, Reverse Engineering and Design Recovery: A Taxonomy, 1990, IEEE Software, vol. 7, no. 1, www.lore.ua.ac.be/Research/Chikofsky1990-Taxonomy.pdf.

[3] G. J. Boer, K. Arpe, M. Blackburn, M. Déqué, W. L. Gates, T. L. Hart, H. le Treut, E. Roeckner, D. A. Sheinin, I. Simmonds, R. N. B. Smith, T. Tokioka, R. T. Wetherald, D. Williamson, “Some Results From an Intercomparison of the Climates Simulated by 14 Atmospheric General Circulation Models”, J. Geophys. Res., 97(D12), 12,771–12,786.

[4] G. J. Boer et al, “Intercomparison of climates simulated by 14 atmospheric general Circulation model”, WMO/TD-No 425, CAS/JSC Working Group on Numerical Experimentation.

[5] Global and Planetary Change, An Overview of results from the Coupled Model Intercomparison Project (CMIP), Feb. 2001, UCRL-JC 140274.

[6] <http://www.supercomputo.secyt.gov.ar/>

[7] Program for Climate Model Diagnosis and Intercomparison <http://www-pcmdi.llnl.gov/>

[8] F. G. Tinetti, P. G. Cajaraville, J. C. Labraga, M. A. López, "Cómputo Paralelo Aplicado a Modelos Numéricos del Clima", IX Workshop de Investigadores en Ciencias de la Computación, Universidad Nacional de La Patagonia San Juan Bosco (UNPSJB), Trelew, Chubut, Argentina, Mayo 3-4 de 2007.

[9] World Meteorological Organization, Numerical Method Used in Atmospheric Models, GARP Publication Series 17, 1979.