

Paralelización del modelo hidrodinámico secuencial COHERENS para sistemas multicore mediante OpenMP

Francisco López Castejón

Afiliación: Taxon Estudios Ambientales S.L

Un modelo numérico hidrodinámico es un programa computacional basado en las ecuaciones de Navier-Stokes, las cuales se tratan de un conjunto de ecuaciones no lineales en derivadas parciales que describen el movimiento de un fluido, en nuestro caso el medio marino. Reciben este nombre debido a Claude-Louis Navier y George Gabriel Stokes

Ec. Navier-Stokes

$$\left(\frac{\partial u}{\partial x}\right) + \left(\frac{\partial v}{\partial y}\right) + \left(\frac{\partial w}{\partial z}\right) = 0$$

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} + w \frac{\partial u}{\partial z} = F_{bx} - \frac{1}{\rho} \frac{\partial p}{\partial x} + \frac{\mu}{\rho} \nabla^2 u$$

$$\frac{\partial v}{\partial t} + v \frac{\partial v}{\partial x} + w \frac{\partial v}{\partial y} + w \frac{\partial v}{\partial z} = F_{by} - \frac{1}{\rho} \frac{\partial p}{\partial y} + \frac{\mu}{\rho} \nabla^2 v$$

$$\frac{\partial w}{\partial t} + v \frac{\partial w}{\partial x} + v \frac{\partial w}{\partial y} + w \frac{\partial w}{\partial z} = F_{bz} - \frac{1}{\rho} \frac{\partial p}{\partial z} + \frac{\mu}{\rho} \nabla^2 w$$

$u \Rightarrow$ Componente X de la velocidad
 $v \Rightarrow$ Componente Y de la velocidad
 $w \Rightarrow$ Componente Z de la velocidad
 $F \Rightarrow$ Sumatorio de fuerzas
 $t \Rightarrow$ Tiempo
 $P \Rightarrow$ Presión
 $\rho \Rightarrow$ densidad
 $\mu \Rightarrow$ viscosidad

$$\frac{\partial P}{\partial z} = \rho \times g$$

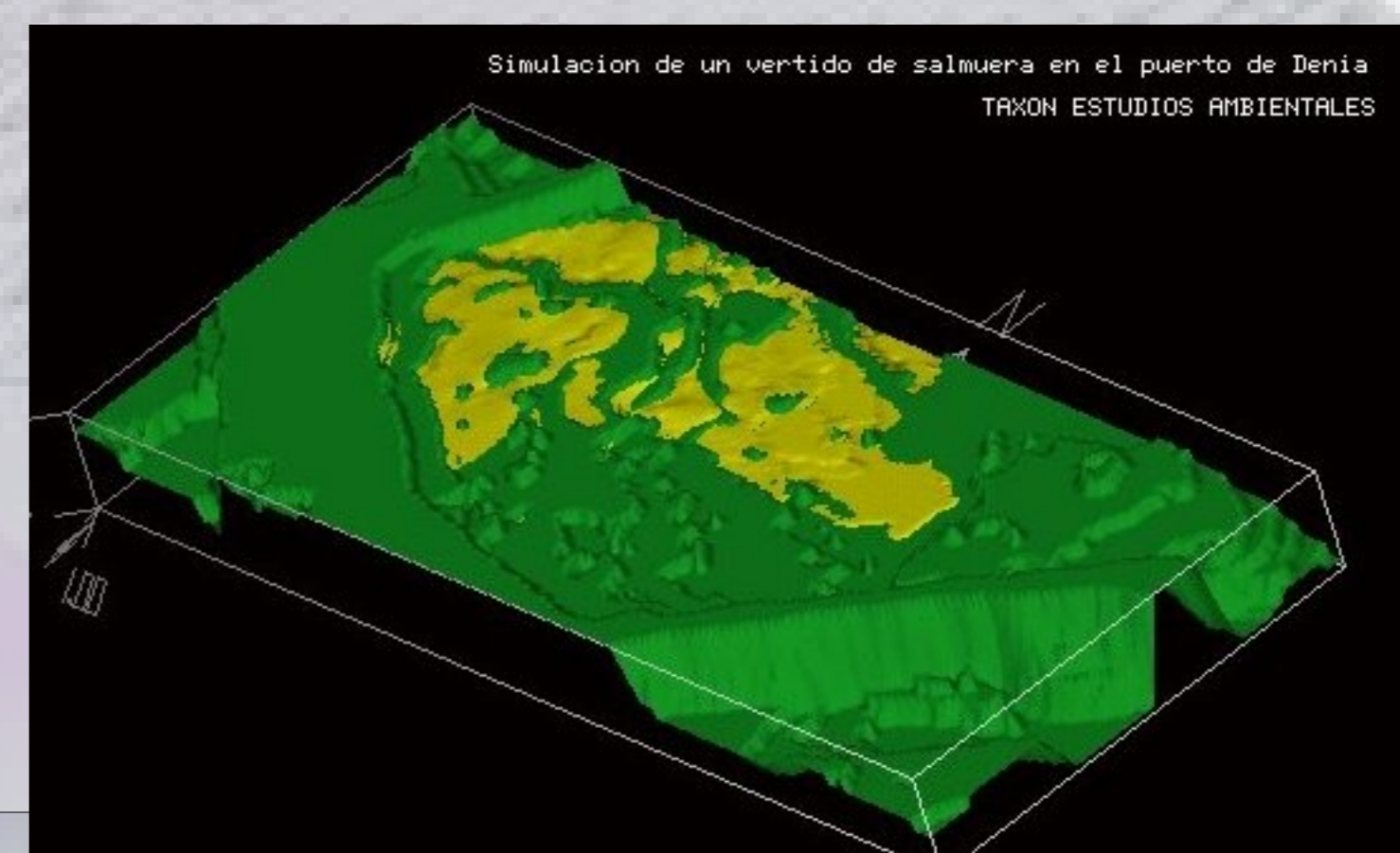
El modelo denominado COHERENS, fué desarrollado entre los años 1990 y 1998, por Management Unit of the North Sea Mathematical Models, Napier University, Proudman Oceanographic Laboratory y el British Oceanographic Data Center, dentro del proyecto europeo MAST PROFILE, NOMADS Y COHERENS. Este software ha sido usado ampliamente con fines científicos, profesionales y educativos, habiendo quedada demostrada su robustez y validez a la hora de simular procesos físicos, químicos y biológicos en el medio marino. En el presente trabajo nos centraremos en la parte hidrodinámica, ya que es esta la que constituye el esqueleto básico del programa y la que implica un mayor número de cálculos.

El software COHERENS destaca principalmente por:

- Trabajar en un espacio tridimensional.
- Ofrecer una solución dinámica al problema planteado.
- Programado en fortran 77 de forma secuencial.
- Poseer una gran cantidad de módulos acoplados, que le permite estudiar otros tipos de procesos, tales como transporte de contaminantes o variación de la calidad del agua.

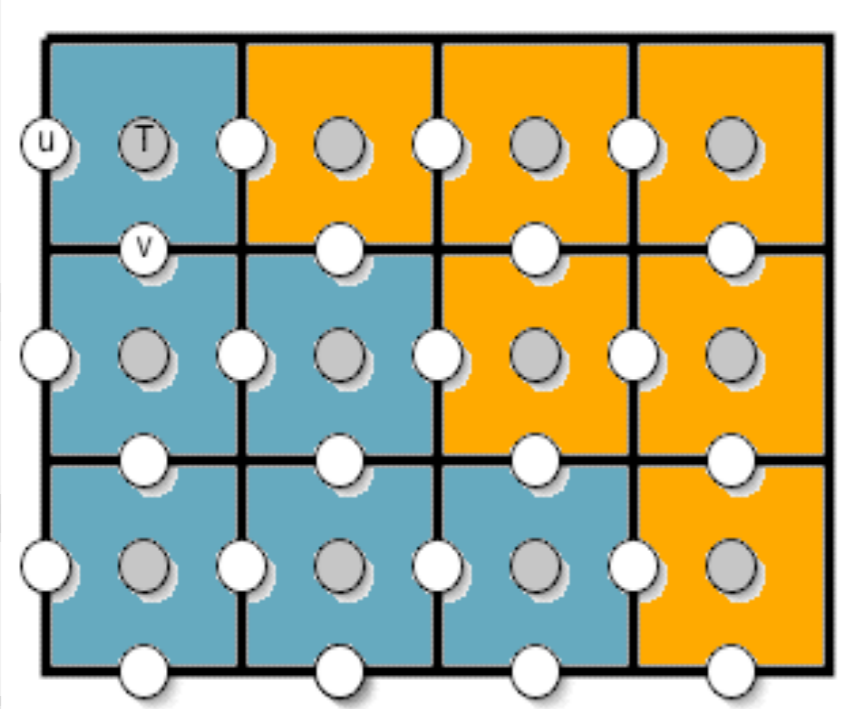
El estudio de un vertido de salmuera al mar que supone gran cantidad de cálculos a muy pequeña escala en la zona cercana al vertido, la respuesta rápida para prever la evolución accidental de un vertido en mar abierto o el estudio de patrones de circulación de mesoescala con un nivel de detalle aceptable, son algunos supuestos de trabajos que han llevado a plantear la necesidad de disminuir el tiempo de cálculo necesario para su resolución sin que esto supusiera una disminución en la calidad de los resultados. De esta manera se podrían afrontar nuevos trabajos y estudios.

Con este objetivo se decidió afrontar la paralelización del software COHERENS mediante OPENMP. La elección de este tipo de paralelismo con memoria compartida partió de la gran difusión que los sistemas multiprocesador tienen hoy en día, lo que permite acceder a un mayor número de plataformas sobre las que realizar las diferentes pruebas de rendimiento una vez paralelizado el software. Además, la empresa Taxon Estudios Ambientales dispone de equipos en los que, una vez paralelizado el código, poder usarlo. Lo cual dota de gran utilidad al objetivo de este trabajo.



El esquema del software Coherens consta de 3 bucles anidados. El más externo de ellos calcula los nuevos pasos de tiempo. En cada paso de tiempo los nuevos valores de intensidad y dirección de corriente son calculados en modo 2D. Posteriormente cada X pasos de tiempo se calculan la características de las corrientes y del resto de las variables, como por ejemplo la temperatura y salinidad, en modo 3D. La frecuencia de cálculo 3D respecto al 2D tiene valores entre 1 y 20 aproximadamente.

El esquema de cálculo utilizado por el modelo se basa en una malla de tipo Arakawa-c, en la que para cada uno de los nodos que la componen se calculan las velocidades en cada lateral, u (velocidad en el eje X) y v (velocidad en el eje Y), y otras variables como la temperatura o salinidad en el centro. El nuevo valor se obtiene a partir de los nodos adyacente y el calculado en el paso de tiempo previo. La existencia de nodos correspondientes a tierra, en los que no se realiza ningún tipo de operación, podría llevar a que algunos threads no tengan ninguna carga de trabajo mientras que otros sí, por lo que habría que detectar estos puntos para repartir la carga de trabajo de una manera equitativa entre los threads.

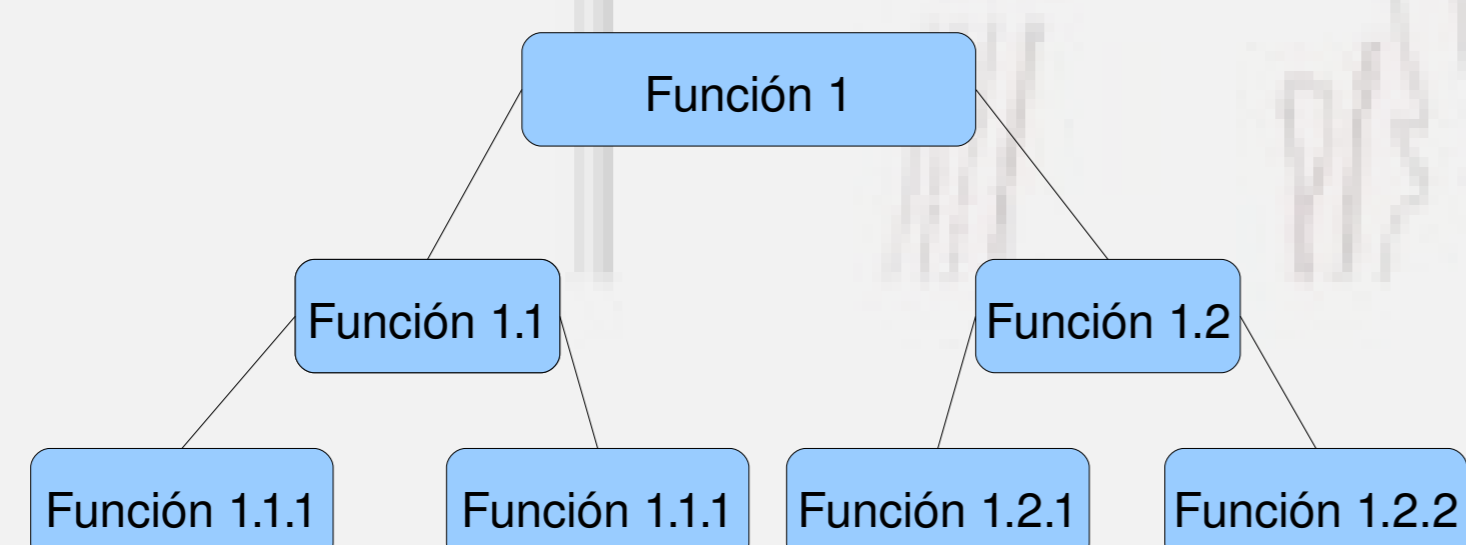


Tierra
 Mar
 U -> Velocidad en el eje X
 V -> Velocidad en el eje Y
 T -> Temperatura

La existencia de gran cantidad de variables y funciones dependientes unas de otras en el programa, podría provocar que funciones paralelizadas no pudieran iniciarse al estar dependiendo de otras no paralelizadas que no habrían terminado de completarse. Esto provocaría un aumento en el tiempo de cálculo y un peor aprovechamiento de los recursos.

Por lo que, como paso previo a la paralelización del software, es muy importante la realización de un grafo de dependencias que se pueda utilizar para asignar tareas distintas a threads distintos.

En la gráfica inferior se puede observar una esquema ilustrativo de la dependencia entre funciones y como la no paralelización de una de ellas podría influir en el resto.



Evolución temporal

Cálculo 2D

CONTNY T=0.0 %
 CRRNT2 T=1.18 %
 Tiempo total = 1.18%

Cálculo 3D

HEDDY T= 0.0 %
 DENSTY1 T= 0.0 %
 DENSTY T= 2.35 %
 VEDDY1 T=14.12 %
 VEDDY2 T=0.0 %
 CRRNT3P T= 46.47 %
 TRANSV T= 0.0 %
 CRRNT1C T= 0.0 %
 CRRNT3C T=1.76 %
 WCALC T=3.53 %
 SALT T=14.12 %
 HEAT T=16.47 %
 Tiempo total= 98.82 %

Datos de tanto por ciento de tiempo total consumido por cada uno de las funciones principales que componen el modelo. Se puede observar como el modo 3D, aunque se considerase un factor de relación con el 2D de 20, consume una gran cantidad de tiempo. Siendo la función CRRNT3P, en la que se realiza el cálculo de las corrientes para cada nodo en 3D la que consume un mayor tiempo. Por tanto, el primer paso consistirá en la paralelización de esta función.