

Computación y tratamiento de datos a escala extrema: el corazón del progreso en materia de predicción meteorológica y climática

por Peter Bauer¹, Michael C. Morgan² y Siham Sbill³

Las mejoras del pasado

Los progresos en la vigilancia del medioambiente y en la predicción numérica del tiempo y del clima han estado estrechamente vinculados con el avance de la supercomputación. Durante las últimas décadas, el aumento de la capacidad de cálculo nos ha permitido mejorar la calidad y el detalle de nuestros pronósticos incrementando la resolución espacial, aumentando el realismo de los modelos al incorporar representaciones más detalladas de los procesos físicos y más componentes del sistema Tierra, y utilizando técnicas de conjuntos para caracterizar la incertidumbre tanto de las condiciones iniciales como de los pronósticos (Bauer y otros, 2015).

Mejores modelos y mejores técnicas de asimilación de datos nos han permitido explotar de forma más eficiente la información disponible acerca del sistema Tierra. La asimilación de datos es computacionalmente tan costosa como la producción de pronósticos, y ese costo aumenta con la mejora de los modelos y con el aumento del volumen y la diversidad de las observaciones asimiladas. Con el perfeccionamiento de los sistemas de predicción y una mayor expectativa por parte de los usuarios de pronósticos más especializados, el volumen y diversidad de los datos de salida crecerá a ritmo similar o incluso mayor que el costo de procesamiento.

En el pasado, ese aumento del costo estuvo compensado en su mayor parte por un incremento comparable en las capacidades de cálculo y tratamiento de datos propiciado por la posibilidad de añadir más transistores a los microprocesadores (Ley de Moore⁴) y a velocidades de reloj más altas manteniendo la potencia (escala de Dennard⁵), a la vez que el precio de los procesadores se reducía. Como la densidad de transistores tiene un límite físico y las velocidades de reloj se estabilizan para limitar el consumo eléctrico, el aumento del rendimiento solo puede ser fruto de mejoras en la paralelización y en las nuevas tecnologías de procesadores que combinen esa conexión con una eficiencia energética mayor. Buena parte de esa tecnología se utiliza en la actualidad en dispositivos como los teléfonos móviles.

El reto del futuro

Se ha estimado que, en diez años, los modelos operativos por conjuntos acoplados del sistema Tierra utilizados habitualmente en predicción meteorológica y en proyecciones climáticas de alta resolución nos llevarán a necesidades de cálculo y tratamiento de datos superiores a las de hoy en un factor de 1000 o mayor (Wehner y otros, 2011). Esas necesidades ya no pueden ser satisfechas solo por la evolución de los equipos informáticos, sino que se requerirán desarrollos adicionales en métodos matemáticos numéricos y estadísticos, así como en técnicas de programación, que permitan trasladar de forma óptima las tareas de cálculo de los modelos de predicción numérica a los nuevos tipos de

1 Centro Europeo de Predicción Meteorológica a Plazo Medio, Reino Unido de Gran Bretaña e Irlanda del Norte

2 Universidad de Wisconsin-Madison, Estados Unidos de América

3 Servicio Meteorológico Nacional, Marruecos

4 www.economist.com/the-economist-explains/2015/04/19/the-end-of-moores-law

5 www.rambus.com/blogs/understanding-dennard-scaling-2/



Foto: Sala de radio de la Estación meteorológica central, Aeropuerto internacional de Santiago (Chile, 1966). Exposición organizada con motivo del Día Meteorológico Mundial de 2013

procesadores emergentes, desde CPU, GPU y FPGA hasta los altamente especializados dispositivos ASIC (circuitos integrados de aplicación específica) (Schulthess, 2015). Ese abanico puede ser incluso mayor en el futuro y el reto más importante para cualquier aplicación será explotar el potencial de los futuros equipos informáticos⁶.

Una restricción clave en los sistemas de computación de alto rendimiento es la de tener un consumo eléctrico asumible. Los actuales sistemas de escala peta (supercomputadores que permiten 10^{15} operaciones en coma flotante por segundo en tareas ejecutándose al máximo rendimiento) consumen del orden de 10^6 vatios por año, lo que implica un coste anual del orden de 10^6 dólares de los Estados Unidos de América en energía y refrigeración. En la actualidad, la mayor parte de los centros de computación de alto rendimiento se construyen con la hipótesis de que su potencia total no excederá, aproximadamente, de 20 MW, un valor muy inferior al citado incremento en un factor de 1000. Así pues, limitarse a comprar ordenadores más grandes no es una opción viable desde el punto de vista económico.

Las comunicaciones en un sistema de computación de alto rendimiento son una preocupación primordial, ya que mover datos en un chip consume unas diez veces más energía que ejecutar un cálculo, y mover datos entre chips cuesta otras diez veces más que mover datos en el mismo chip (Kogge y Shalf, 2013). Una inquietud adicional es cómo manejar a lo largo del proceso de predicción los datos observacionales de entrada y las salidas de los modelos para permitir un preproceso y postproceso eficientes que minimicen el movimiento de datos, reduzcan las necesidades de almacenamiento y aseguren al mismo tiempo la resiliencia de la producción del pronóstico. Mientras que los retos en computación y tratamiento de datos aumentan de forma drástica, los requerimientos en usabilidad y acceso rápido por parte de los usuarios también son cada vez más estrictos.

Más usuarios quieren más información, y que se suministre más rápidamente, creando tremendos retos en materia de tratamiento de datos, los cuales requerirán una mezcla de soluciones centralizadas y en la nube que permitan acercar las aplicaciones a donde se encuentren los grandes volúmenes de datos de predicción, y distribuir entre un amplio abanico de plataformas las costosas tareas de análisis y postproceso de datos en función de las necesidades de los usuarios.

6 www.newscientist.com/article/mg24031990-300-could-the-worlds-mightiest-computers-be-too-complicated-to-use/

¿Y el papel de la inteligencia artificial?

El renacimiento de los métodos de inteligencia artificial auspiciados por grandes aplicaciones comerciales puede contribuir a ese indispensable aumento de la eficiencia. Grandes compañías como IBM y Microsoft (respaldando a AccuWeather) anuncian su capacidad para suministrar a los clientes soluciones altamente especializadas tomando los productos de modelos de centros operativos nacionales e internacionales y uniéndolas a sus propios productos de predicción. Ello ha sido posible porque los procesadores de fabricación específica se ensamblan a mayor escala y los programas informáticos de aprendizaje profundo son capaces de clasificar ingentes cantidades de datos procedentes de modelos y observaciones a fin de extraer información meteorológica para fines de predicción.

La sustitución de los sistemas de predicción basados en la física por sistemas basados en el aprendizaje profundo parece improbable debido a que el número de grados de libertad del sistema Tierra, y su no linealidad, requerirían redes neuronales muy complejas que serían difíciles de entrenar y tendrían el riesgo de ser potencialmente ineficientes al ejecutarse en ordenador (Düben y Bauer, 2018). Parte del reto que las redes neuronales deben encarar al enfrentarse a pronósticos a plazos medio, estacional o climático válidos para todo el mundo será producir pronósticos físicamente coherentes, manteniendo la conservación y los balances de flujos. El manejo de sesgos y errores en el entrenamiento de los datos añaden dificultad al reto.



Sin embargo, el uso de tales técnicas en el preproceso de datos observacionales y en el postproceso de las salidas del modelo puede ayudar a distribuir mejor a lo largo del flujo de trabajo la carga que conlleva el tratamiento de datos, a extraer información útil de grandes volúmenes de datos de manera más efectiva y a reducir la carga computacional sustituyendo algunos

componentes seleccionados del modelo de predicción por redes neuronales. Esas aplicaciones ya se probaron en el pasado y son campos activos de investigación en la actualidad (Lee y otros, 2018; Hsieh y Tang, 1998).

Iniciativas actuales en materia de investigación

Los problemas expuestos plantean serios obstáculos a la mejora de las capacidades de los modelos de predicción, tanto meteorológicos como climáticos. El aumento de la percepción de ese problema ha conducido a esfuerzos en innovación e investigación a gran escala en muchos países desarrollados, que son apoyados con importantes fondos gubernamentales y privados. Los proyectos del Departamento de Energía de los Estados Unidos de América⁷ y de la Comisión Europea⁸ son ejemplos que se complementan con otros desarrollados por muchos servicios meteorológicos en los Estados Unidos, Japón, China y Europa.

Debido a la complejidad de la empresa, los investigadores deben colaborar estrechamente con la industria de la computación, y la ciencia meteorológica y climatológica necesita colaborar estrechamente con los sectores sujetos a impacto como los del agua, la energía, la agricultura y la alimentación, y la gestión de riesgos. Esas colaboraciones y la investigación científico-tecnológica de vanguardia son el foco central del proyecto ExtremeEarth⁹, propuesto como proyecto europeo de referencia que promete para esta nueva era grandes avances en la capacidad de predicción.

Esos desafíos contribuyen claramente a aumentar todavía más la diferencia de capacidades entre países más y menos desarrollados, ya que requieren un nivel singular de conocimientos especializados, colaboración entre industria e investigación y respaldo tecnológico significativo, tanto para los equipos como para los programas informáticos. Aquí es donde la colaboración internacional promovida por organizaciones como la Organización Meteorológica Mundial (OMM) será crucial para producir economías de escala sostenibles y apoyar la transferencia de conocimiento entre diferentes campos y entre los distintos países y continentes.

7 <https://e3sm.org>

8 <https://portal.enes.org>, <https://www.esiwace.eu>, <http://www.hpc-escape.eu>

9 <http://extremearth.eu>

El papel de la Organización Meteorológica Mundial

La necesidad de un esfuerzo concertado entre las ciencias del tiempo y del clima y la ciencia de la computación requiere una estrategia visible por parte de la OMM. El objetivo de ese esfuerzo sería desarrollar e intercambiar metodologías y tecnologías para la producción rentable de pronósticos y la concentración y distribución de grandes volúmenes de datos de modelos de alta resolución de complejidad creciente en todas las escalas.

A continuación se enumeran las recomendaciones detalladas para una estrategia revisada:

- Establecimiento de metodologías científicas que exploren la mejora del paralelismo y la reducción del movimiento de datos en las infraestructuras de computación de alto rendimiento a escala extrema.
- Apoyo a la normalización de estructuras de código compatible y modelos de programación que aseguren la eficiencia y legibilidad, y exploten la futura gama de procesadores y tecnologías de sistema, incluyendo criterios para la prueba de códigos, y análisis y comparativas de rendimiento.
- Diseño de entornos compatibles de tratamiento de datos para el preproceso de los datos observacionales, el postproceso de las salidas de los modelos y la distribución de productos.
- Soporte para una computación abierta, distribuida y basada en la nube, y para infraestructuras de gestión de datos para todas las etapas del proceso de producción de pronósticos, incluida la facilidad de acceso, localización de información y visualización para los usuarios finales.
- Adaptación de métodos de inteligencia artificial, como el aprendizaje profundo, para facilitar el procesamiento de datos observacionales cada vez más diversos, la extracción de información –en función del usuario– de salidas de modelos cuya complejidad va en aumento y el desarrollo de sustitutos de componentes del modelo que reduzcan el costo computacional.
- Creación de capacidad y formación entre las ciencias aplicada e informática para facilitar la captación

de las nuevas tecnologías y metodologías por parte de la comunidad.

Casi todas las áreas de aplicación de la predicción meteorológica y climática se beneficiarán de esta estrategia en la medida que las futuras capacidades en materia de computación y tratamiento de datos posibilitarán nuevos avances científicos, más rentabilidad en la manera de operar y una mayor transferencia de conocimiento desde los expertos hacia la base general de usuarios.

Referencias

- Bauer, P., A. Thorpe y G. Brunet, 2015: The quiet revolution of numerical weather prediction. *Nature*, 525, 47-55.
- Düben, P. y P. Bauer, 2018: Challenges and design choices for global weather and climate models based on machine learning. *Geoscientific Model Development*, 11, 3999-4009.
- Hsieh, W. W. y B. Tang, 1998: Applying neural network models to prediction and data analysis in meteorology and oceanography. *Bulletin of the American Meteorological Society*, 79, 1855-1870.
- Kogge, P. y J. Shalf, 2013: Exascale computing trends: Adjusting to the "New Normal" for computer architecture. *Computing in Science and Engineering*, doi: 10.1109/MCSE.2013.95.
- Lee, Y.-J., C. Bonfanti, L. Trailovic, B. J. Etherton, M. W. Govett y J. Q. Stewart, 2018: Using deep learning for targeted data selection: Improving satellite observation utilization for model initialization. 17th Conference on Artificial and Computational Intelligence and its Applications to the Environmental Sciences. <https://ams.confex.com/ams/98Annual/webprogram/Paper333024.html>.
- Schulthess, T. C., 2015: Programming revisited. *Nature Physics*, 11, 369-373.
- Wehner, M. F., L. Oliker, J. Shalf, D. Donofrio, L. A. Drummond, R. Heikes, S. Kamil, C. Kono, N. Miller, H. Miura, M. Mohiyuddin, D. Randall y W.-S. Yang, 2011: Hardware/software co-design of global cloud system resolving models. *Journal of Advances in Modeling Systems*, 3, M10003, 22 pp. <https://doi.org/10.1029/2011MS000073>.