

Simulación computacional, ciencia de los datos, cómputo de alto rendimiento y optimización aplicados a mejorar la predicción de modelos de simulación que representan la evolución de sistemas complejos

Mariano Trigila², Adriana Gaudiani^{1,4}, Emilio Luque³, Marcelo Naiouf¹

¹ Instituto de Investigación en Informática LIDI, Facultad de Informática, Universidad Nacional de La Plata - Centro Asociado de la CIC
mnaiouf@lidi.info.unlp.edu.ar

² Facultad de Ingeniería y Ciencias Agrarias, Pontificia Universidad Católica Argentina
marianotrigila@gmail.com

³ Depto. de Arquitectura de Computadores y Sistemas Operativos, Universidad Autónoma de Barcelona, 08193 Bellaterra (Barcelona) España
emilio.luque@uab.es

⁴ Instituto de Ciencias, Universidad Nacional de General Sarmiento.
agaudi@ungs.edu.ar

Resumen

El objetivo principal de esta propuesta es mejorar de manera automática la calidad de la simulación de un sistema dinámico complejo, proponiendo una metodología computacional de calibración.

En particular se utilizará con un modelo de translación de ondas a lo largo del cauce de un río [9]. Esta metodología se centrará en manejar la incertidumbre en los parámetros del modelo para optimizar la calidad de la predicción del simulador.

Este trabajo continúa con la línea de investigación previa donde se propuso una metodología computacional de optimización que fue validada con este simulador [6]. En esa oportunidad se elaboró una mejora computacional independiente del sistema simulado, pero en este trabajo se aprovecharán las características propias del dominio del sistema para calibrar el modelo. Se aprovecharán las propiedades de localidad en los valores de los parámetros, proponiendo una calibración en pasos sucesivos, en tramos seleccionados sobre el cauce del río.

La implementación de este trabajo requiere la utilización de técnicas de las ciencias de los datos, del cómputo de alto rendimiento y métodos del campo de la optimización [7] [10], buscando lograr el mayor ahorro en el uso de recursos computacionales.

Palabras Claves

Optimización y simulación, calibración automática, ciencia de los datos, cómputo eficiente, cómputo de alto rendimiento.

Contexto

El desarrollo de este trabajo se construye en base a la relación de colaboración establecida con los siguientes equipos de investigación.

- III-LIDI, Facultad de Informática, Universidad Nacional de La Plata.
- High Performance Computing for Efficient Applications and Simulation Research Group (HPC4EAS), Universidad Autónoma de Barcelona.
- Instituto de Ciencias de la Universidad Nacional de General Sarmiento.

- Proyecto LAPP: Análisis y exploración de visualización de datos espacio-temporales en el dominio de hidrología. Pontificia Universidad Católica Argentina.

- Instituto Nacional del Agua (INA), Laboratorio de Hidráulica Computacional (LHC).

- Group on Non-Equilibrium Processes and Transport Phenomena. Instituto de Ciencias de la Universidad Nacional de General Sarmiento.

El trabajo comenzó en 2017 y se lleva adelante en el ámbito de una tesis doctoral en Ciencias Informáticas, en la UNLP. El grupo que acompaña el proyecto viene trabajando en este tema desde 2010, y de estas investigaciones resultó una tesis doctoral en la que se basa esta propuesta, y varias publicaciones [5] [11]. Los resultados esperados son de interés para los grupos con los que se mantiene colaboración.

Introducción

Las inundaciones por desborde de ríos, cuyo impacto involucra a la sociedad, son eventos que afectan los recursos económicos en las zonas de influencia de subidas o bajantes del agua, incluso ponen en riesgo las vidas humanas. La simulación y predicción de estos eventos juegan un rol primordial como herramientas para brindar alertas. Estas son las causas principales de la elección de trabajar, en este proyecto, sobre mejorar la certeza provista en los datos simulados que proveen los modelos hidrológicos. En particular, los que se utilizan para predecir los niveles del agua del río Paraná, en Argentina, y en consecuencia, para predecir inundaciones en su zona de influencia [1] [9].

Las inundaciones de ríos de planicie es un fenómeno natural que ocurre con recurrencia, y con el cambio del clima

global se espera que acontecimientos extremos más severos y más frecuentes tengan ocurrencia en un futuro cercano [1] [8].

Por su propia naturaleza, el cauce y el lecho del río suelen modificarse con el transcurrir del tiempo y con ello se modifican los contornos y las áreas inundables, y en consecuencia se modifican los parámetros que caracterizan los modelos que lo representan, ya que se está en presencia de un sistema físico, complejo y dinámico. Este dinamismo le agrega mayor incertidumbre a la predicción que los especialistas realizan utilizando los datos simulados que se obtienen de los modelos computacionales de dicho sistema [4] [12] [13].

La Ciencia de la Computación y las tecnologías involucradas proveen los métodos y técnicas que brindarán los aportes necesarios para optimizar el proceso de simulación de estos sistemas [1]. En consecuencia, se podría ofrecer un aporte que mejore los pronósticos sobre el comportamiento de dichos sistemas físicos. Se propone aportar una mejora en la predicción de la simulación de un modelo computacional utilizado para brindar alertas a la población sobre crecidas y bajantes del cauce de un río, con gran impacto social y económico en la población costera [5] [10]. Se propone ofrecer una metodología de calibración automática del simulador, que se adapte a períodos de creciente o de bajante del río, debido a que estos eventos incrementan fuertemente el porcentaje de error en la predicción [8]. Se espera obtener amplios beneficios en las mejoras propuestas mediante los avances en la Ciencia de los Datos y en la Optimización vía Simulación, como también en el uso de técnicas más modernas del HPC [2] [3] [7]. En particular se optimizará el uso de recursos computacionales al implementar esta metodología.

Líneas de Investigación, Desarrollo e Innovación

Este trabajo está enmarcado por las líneas de investigación:

a) *computación eficiente y segura para la simulación*

b) *optimización de aplicaciones y optimización del uso de arquitecturas paralelas híbridas y cloud.*

Ambas son líneas de investigación relacionadas con los trabajos de los grupos de colaboradores nombrados en el contexto de esta presentación.

El desarrollo del mismo contempla ir cumpliendo con los siguientes objetivos:

- Estudiar el modelo de simulación del INA (Instituto Nacional del Agua)¹, utilizado para brindar alertas sobre el cauce del Río Paraná, en la Cuenca del Plata, mediante un estudio exhaustivo de la estructura de sus datos, de los datos de entrada y salida del modelo y del comportamiento del simulador [8][9].

- Estudiar el problema de optimización y calibración automática del simulador seleccionado para este trabajo, como también el problema de la búsqueda de patrones de comportamiento del río seleccionado en función de eventos extraordinarios, con el objetivo de analizar períodos de validez de las soluciones encontradas [6][12].

- Analizar y aplicar las posibles opciones que brinden una metodología de optimización vía simulación de la predicción, en modo dinámico, que se

adapte a la mayor extensión posible del dominio del río (incrementando la cantidad de parámetros de ajuste todo lo posible), con el objetivo de manejar más eficientemente los errores en la predicción que se originan por subidas o bajantes bruscas del nivel del cauce [2] [5] [13].

- Estudiar y aplicar las técnicas y metodologías que ofrece la programación en paralelo y el HPC en la implementación de la metodología propuesta.

- Analizar y utilizar las tecnologías, plataformas, frameworks y lenguajes de programación necesarios para diseñar e implementar las soluciones que sean necesarias para llevar adelante las experiencias requeridas durante el desarrollo de este trabajo.

El uso de los avances de HPC será imprescindible tanto para crear soluciones eficientes como para llevar adelante las experiencias de validación de la metodología propuesta. Se requiere una gran carga de cómputo al necesitar múltiples ejecuciones del simulador y al procesamiento de las enormes cantidades de datos que se obtendrán [1][3][7]. Se trabajará especialmente en minimizar los recursos computacionales, como tiempo de uso de CPU's, pensando en el mayor ahorro energético posible.

Resultados y Objetivos

Describir una metodología o una serie de procedimientos particulares basados en ciencias de los datos, desarrollo de heurísticas y simulación computacional, con aplicación de los métodos de la computación de alto rendimiento. Se espera proveer certeza en la capacidad de predicción del modelo, mediante un método de calibración automática; especialmente ante eventos extremos de

¹ G. Latessa, "Modelo hidrodinámico del río Paraná para pronóstico hidrológico: Evaluación del performance y una propuesta de redefinición geométrica.," INA - UBA, Buenos Aires , 2011

subida o bajante del cauce del río, siendo en estos momentos (de sequía o de lluvias) durante los que se desencadenan importantes desajustes al modelo de simulación.

Desarrollar herramientas de soporte necesarias para llevar adelante las investigaciones de este trabajo, con posibilidad de ofrecer herramientas creativas e innovadoras que puedan utilizarse en la toma de decisiones en alguna de las áreas de incumbencia de este proyecto.

Describir una serie de procedimientos generales basados en Ciencia de la Computación y Ciencias de los Datos para proveer un modelo de optimización de la predicción.

Estudiar la factibilidad de extender la metodología desarrollada para optimizar la predicción de otro modelo de simulación de fenómenos naturales.

Formación de Recursos Humanos

El desarrollo de este trabajo corresponde al ámbito de la tesis doctoral en Ciencias Informáticas de la UNLP, llevada adelante por Mariano Trigila bajo la dirección del Dr. E. Luque y la asesoría científica de la Dra. Gaudiani.

En el transcurso de este año se presentará la solicitud de becas de investigación otorgada por la UNGS a estudiantes avanzados de la carrera de Licenciatura en Sistemas para participar en temas de investigación relacionados con esta propuesta.

Referencias

[1] S. Balica, "Parametric and physically based modelling techniques for flood risk and vulnerability assessment: A comparison," *Environmental*

Modelling & Software, vol. 41, pp. 84-92, 2013.

[2] E. Cabrera, M. Taboada, M. Iglesias, F. Epelde y E. Luque, "Simulation Optimization for Healthcare Emergency Departments," *Procedia Computer Science*, vol. 9, pp. 1464-1473, 2012.

[3] A. Chi Zhou, B. He y I. Shadi, "A Taxonomy and Survey on eScience as a Service in the Cloud," *Cornell University Library*, 2014.

[4] A. Domeneghetti, A. Castellarin y A. Brath, "Assessing rating-curve uncertainty and its effects on hydraulic model calibration," *Hydrology and Earth System Sciences*, vol. 16, pp. 1191-1202, 2012.

[5] A. Gaudiani, E. Luque, P. García, M. Re, M. Naiouf y A. De Giusti, "How a Computational Method Can Help to Improve the Quality of River Flood Prediction by Simulation.," *Advances and New Trends in Environmental and Energy Informatics. Springer.*, pp. 337-351, 2016.

[6] A. Gaudiani, "Simulación y optimización como metodología para mejorar la calidad de la predicción en un entorno de simulación hidrográfica," *Tesis de doctorado - UNLP*, 2015.

[7] T. Hey, S. Tansley y K. Tolle, "The

- Fourth Paradigm: Data-Intensive Scientific Discovery,” *Proceedings of the IEEE*, vol. 99, nº 8, pp. 1324-1337, 2009.
- [8] A. Menéndez, “Three decades of development and application of numerical simulation tools at INA Hydraulics Lab,” *Mecánica Computacional*, vol. 21, pp. 2247-2266, 2002.
- [9] M. Re, N. Badano, E. Lecertúa, F. Re y A. Menéndez, “Modelización matemática de una cuenca de llanura extensa,” *Mecánica Computacional*, vol. 27, pp. 351-368, 2008.
- [10] J. F. Santucci y L. Capocchi, “Optimization via Simulation of Catchment Basin Management Using a Discrete-event Approach,” *Simulation*, vol. 91, pp. 43-58, 2015.
- [11] M. Trigila, “Cluster LAPP: Implementación de un cluster de bajo costo para procesamiento paralelo” de *World Congress on Systems Engineering and Information Technology - WCSEIT'2015*, Vigo, España, 2015.
- [12] J. Warmink, J. Janssen, M. Booij y M. Krol, “Identification and classification of uncertainties in the application of environmental models,” *Environmental Modelling & Software*, vol. 25, nº 12, pp. 1518-1527, 2010.
- [13] Q. Wu, S. Liu, Y. Cai, X. Li y Y. Jiang, “Improvement of hydrological model calibration by selecting multiple parameter ranges,” *Hydrology and Earth System Sciences.*, vol. 21, pp. 393-407, 2017.