

Simulación y Optimización como Metodología para Mejorar la Calidad de la Predicción en un Entorno de Simulación Hidrográfica.



Adriana Angélica GAUDIANI*

Directores: Emilio LUQUE FADÓN†
y Armando De GIUSTI‡

Codirector: Marcelo NAIOUF‡

Tesis presentada para obtener el grado de
Doctor en Ciencias Informáticas

Facultad de Informática
Universidad Nacional de La Plata

Julio de 2015

*Facultad de Informática, Universidad Nacional de La Plata, Argentina

†Dept. de Arquitectura de Computadores y Sistemas Operativos, Universidad Autónoma de Barcelona, España

‡Facultad de Informática, Universidad Nacional de La Plata, Argentina

*Dedicado a mis hijos
Cristian y Ezequiel*

Agradecimientos

Esta tesis representa un logro muy importante que no hubiera sido posible sin la colaboración y el acompañamiento de muchas personas que me ayudaron de miles de maneras a lograr este objetivo. Algunas conocidas gracias a este trabajo y que ya son parte de mi vida, y otras que ya formaban parte desde antes. Siento que fue un trabajo entre muchos y que me dejó un valor agregado invaluable, representado por el cariño y amistad de tantos amigos, colegas y familiares que me empujaron para finalizarlo.

Quiero agradecer a mis padres primeramente y rendir un homenaje a su memoria. Ellos me enseñaron como trabajar para alcanzar mis metas en la vida.

Mi agradecimiento eterno a la Universidad de General Sarmiento que apoyó mis investigaciones en todos estos años con financiamiento y licencias para asistir a los cursos de postgrado o viajes de investigación. También a la Fac. de Filosofía y Letras, de la UBA, por el apoyo que me dieron en estos últimos meses para terminar esta tesis.

Quiero agradecer de manera muy especial a mi Director Emilio Luque por su acompañamiento todos estos años y por confiar siempre en mí y valorar mi trabajo, especialmente en momentos difíciles. GRACIAS Emilio por tus enseñanzas y tu eterna paciencia. También les agradezco a mis directores Tito y Marcelo, por haber hecho posible el contacto con Emilio y por abrirme las puertas del III-LIDI e incluirme en su equipo de investigadores ayudándome a crecer en mi vida académica.

A mis colegas del INA, Pablo García y Mariano Re, quienes me ayudaron incondicionalmente a llevar el trabajo adelante. Sin su ayuda y el material que pusieron a mi disposición, no hubiera sido posible. Gracias Pablo por tantas horas de trabajo en conjunto y por enseñarme tanto de hidráulica.

A mis amigos y colegas de la UAB, Lola, Remo, Eduardo, Francisco, y muchos más que tuve la suerte de conocer en la UAB y aprender mucho de su trabajo. Y gracias a los responsables de la UAB que participaron en la financiación de mis viajes de investigación. Y especialmente a Loli por darme todo tu cariño en España y por hacerme sentirme como en casa, gracias por ayudarme a reflexionar ante los problemas y por darme fuerza en momentos difíciles.

A mis amigos y colegas que me ayudaron y acompañaron todo este tiempo. A todos los integrantes del III-LIDI por estar siempre que los necesité. Emmanuel, GRACIAS por tu ayuda invaluable en todo momento y por tu paciencia para ayudarme a encontrar solución a los problemas de implementación, de programación, de latex, etc.

AGRADECIMIENTOS

A mis amigos y colegas de la UNGS, por estar cuando los necesité. Todos me ayudaron de diferentes maneras, Claudio, Martín, Mariano, Santiago, y muchos más a lo largo de estos años. Florencia, gracias por valorar siempre mi trabajo y por acompañarme en este camino.

A mis amigas de tantos años, Silvina y Susana, por darme cuotas de fuerza siempre que, acertadamente, vieron que la estaba necesitando.

Y especialmente, a Cristian, Ezequiel y Ricardo, quienes son un pedazo de mi corazón y las personas más importantes de mi vida, por la ayuda que me brindaron siempre y por comprender todo el tiempo que tuve que invertir en mi trabajo. Cristian y Ezequiel, GRACIAS eternamente por acompañarme tanto.

A Elisa, mi suegra querida, quién siempre me dio alegría y esperanza.

Resumen

El modelado y la simulación computacional de eventos naturales proveen una poderosa herramienta para comprender sus mecanismos y pronosticarlos. Estas herramientas son especialmente útiles cuando estos eventos se convierten en desastres naturales que amenazan la vida humana y la economía de los pueblos afectados. En particular, las inundaciones provocadas por el desborde de ríos es uno de los desastres naturales que se benefician ampliamente con el uso de programas computacionales que ayudan a pronosticar sus efectos. Las inundaciones son fenómenos naturales que pueden ocurrir a intervalos irregulares en cualquier riachuelo o río. El asentamiento de la población en llanuras de inundación es la causa principal de los daños producidos por las inundaciones. El estudio de los procesos hidrodinámicos en las aguas superficiales, como ríos, ha desempeñado un papel pionero en el desarrollo de modelos numéricos para su uso en la ingeniería hidráulica. Esto ha llevado a la producción de gran cantidad de software relativamente sofisticado aplicado a estudiar y simular fenómenos de traslación de ondas en cauces de ríos, de manera de poder predecir eventos de desborde e inundaciones.

Aunque los modelos se construyan considerando la mayor cantidad posible de variables propias del sistema simulado existen fuentes de incertidumbre en los datos de entrada y en los parámetros propios del modelo, que aumentan el error de los datos de salida de la simulación. La mayoría de estos parámetros pueden estimarse a partir de datos de campo, por ejemplo, observaciones hidrometeorológicas, mapas de la topografía, tipos de suelo, uso del suelo, etc. Sin embargo, rara vez están disponibles datos de campo completos para apoyar plenamente la especificación de todos los parámetros. La incertidumbre debida a los valores de los parámetros de entrada, se debe a que no existe una manera de medirlos directamente y deben ser estimados o interpolados siendo muy difícil mantener actualizados sus valores durante el período de simulación. El motivo que impulsó el tema central de este trabajo fue proveer a este tipo de simulación de mayor certeza en sus resultados.

El objetivo principal de la tesis es brindar una metodología para mejorar la calidad de la simulación de un modelo computacional de cauce de ríos, aplicando la técnica de optimización vía simulación; objetivo posible de alcanzar gracias a los aportes de las técnicas y métodos del cómputo de alto rendimiento. Este método devuelve como solución un conjunto de parámetros optimizados que mejora la predicción del simulador. La mejora lograda en la calidad del simulador es medida en términos del ajuste entre los datos de salida, o simulados, y los datos reales, u observados.

El desarrollo de esta tesis requirió la selección de un modelo computacional de cauces. El modelo seleccionado fue el programa EZEIZA VI (Ezeiza), un modelo hidrodinámico desarrollado por el Instituto Nacional del Agua (INA) de la República Argentina, y el modelo del sistema físico del río Paraná cuyo comportamiento y evolución se simula con Ezeiza. El Río Paraná es de interés para este trabajo; pertenece a la Cuenca del Plata y recorre el área más poblada e industrializada de Sudamérica. Las conversaciones y reuniones con los investigadores del INA fueron decisivas para entender el funcionamiento del simulador Ezeiza y las características del sistema simulado. De esta manera fue posible elegir los parámetros críticos del modelo, como también sus posibles rangos de valores.

Optimización mediante Simulación (OvS) es un campo emergente que integra técnicas de optimización y análisis de la simulación, el cual se ha convertido en una poderosa técnica para el análisis y optimización de sistemas complejos. Mediante los aportes de la OvS, se implementa un metamodelo de simulación para encontrar el conjunto de parámetros de ajuste que optimiza la salida del simulador. El uso de un metamodelo se hace para facilitar la tarea de optimización, que implica encontrar la mejor configuración de parámetros de entrada, o variables controlables del sistema, con el fin de optimizar la función objetivo, la cual es función de las salidas del modelo de simulación, en este caso el modelo Ezeiza. El método utiliza la salida del simulador, y otros datos propios de sistema físico simulado, para evaluar su función objetivo y en base a esta medida o índice de rendimiento, propone un nuevo conjunto de valores de los parámetros.

El método de OvS aplicado a mejorar los resultados del simulador Ezeiza, es un metamodelo que utiliza una estrategia de OvS en dos etapas. Un metamodelo es un modelo empírico creado con datos generados por modelos computacionales, o sea son modelos de modelos. El método que se presenta se sitúa entre los métodos de búsquedas exhaustivas y las heurísticas propias de la optimización y es un metamodelo que utiliza una estrategia de OvS en dos etapas. La idea principal es la siguiente: En una primera etapa, mediante un proceso iterativo, se van obteniendo subregiones del espacio de búsqueda del problema

que cumplan con su función objetivo. Esta etapa se basa en la estructura de vecindad del problema para reducir la región de posibles soluciones donde se hace la búsqueda, mediante un esquema Monte Carlo combinado con un método K-Means. Se obtienen conjuntos de *regiones prometedoras*, nombradas así desde el punto de vista de la posibilidad de contener la solución óptima, o alguna subóptima, cuyo tamaño es mucho más chico. Al finalizar, se implementa una segunda etapa de búsqueda exhaustiva, pero ahora sobre la región prometedora resultante, ya mucho más pequeña.

A pesar de la reducción del espacio de búsqueda, ambas etapas implementan métodos computacionales que requieren un gran tiempo de ejecución y son posibles de llevar a la práctica gracias a los recursos de programación paralela y de cómputo de alto rendimiento. Con estos recursos se logran tiempos de ejecución aceptables, reduciendo considerablemente el tiempo computacional de cada etapa.

El método de OvS en sus dos etapas es un método fácilmente paralelizable. El algoritmo principal debe evaluar la función objetivo con cada candidato solución antes de ingresarlo al método de optimización, el cual hace un análisis de los índices obtenidos para decidir si el método debe repetirse. El simulador se ejecuta con cada uno de los elementos seleccionados al azar del espacio de búsqueda por el método Monte Carlo, el cual es un espacio de 110.592 combinaciones en el primer caso de experimentación y de 19683 combinaciones en el segundo caso. Cada evaluación es independiente de las otras, por lo tanto estamos ante un problema altamente paralelizable en un cluster mediante un modelo de programación Master-Worker. La aplicación Master-Worker del método fue escrita en C++ con MPI. Se utilizó pthreads para ejecutar el conjunto de programas que conforman el simulador Ezeiza, que se utilizan como una caja negra durante todo el proceso.

Este trabajo optimiza la predicción de Ezeiza en las estaciones de monitoreo ubicadas a lo largo del río Paraná. Para su comprobación, se implementaron dos casos de estudio del método utilizando datos reales en dos periodos de tiempo diferentes. Estos casos optimizan dos índices de rendimiento necesarios para encontrar el conjunto óptimo de parámetros. Estos índices son indicadores del rendimiento general de la simulación y dependen de los datos de salida (simulados) y de los datos observados (reales) en cada una de las 15 estaciones de monitoreo durante cada día del periodo simulado. Las funciones objetivos establecidas para la optimización están definidas por estos índices. Uno de ellos minimiza el apartamiento promedio entre ambos grupos de datos y en segundo lugar se considera una función multiobjetivo que minimiza estas diferencias y maximiza la cantidad de estaciones que mejoran.

Los escenarios de simulación que resultaron del proceso de OvS son los que contienen los valores óptimos de sus parámetros de entrada. Estos conjuntos solución, o conjunto de variables de decisión optimizadas, junto con todos los datos de entrada de una simulación completa, constituyen los escenarios de simulación que sintonizan la predicción realizada con el simulador Ezeiza.

Se logró una mejora en la predicción de Ezeiza entre el 10 % y el 25 % para el primer caso de estudio y comprobación del método y mejoras entre el 20 % y el 40 % para el segundo caso de experimentación. Estos porcentajes representan un promedio de 40cm a 60cm de altura del agua, y la importancia de estos valores radica en que pueden ser la diferencia entre la pérdida o no del ganado o de cubrir casas, al entrar el agua en las planicies de inundación afectadas. Estos escenarios optimizados mejoraron favorablemente la simulación de Ezeiza en las estaciones ubicadas en el tramo del bajo Paraná. Fue seleccionado este tramo para la optimización, pues es un tramo del río sensible para la toma de decisiones y los pronósticos. Esta zona tiene alto impacto en la región de la Cuenca del Plata, especialmente en su economía, debido a la alta cantidad de población que tiene instalada en sus planicies. No es posible pensar en un escenario de optimización de todo el sistema físico, como el que representa el río Paraná en sus 1900 km. con 76 secciones en su cauce. En cada sección se deben definir los valores de sus parámetros ajustables. El espacio de búsqueda resultante de la combinación de estos parámetros toma una dimensión tan inmensa, que optimizar todo el sistema es una misión aun imposible de llevar a la práctica.

Los resultados son alentadores y el principal aporte del trabajo fue proveer de una metodología que permitió calibrar y sintonizar un simulador, como Ezeiza, de manera automática. Esto es muy beneficioso para los ingenieros hidráulicos que deben hacer predicción con el simulador y no pueden contar con valores actualizados constantemente de los datos del río. Esta metodología es prometedora para ser utilizada en la sintonización automática de otros simuladores hidrológicos. Por otro lado, el uso del cómputo de alto rendimiento se potencia con el aporte del método de OvS, ya que se tiene una ganancia por la reducción en cantidad de simulaciones que ofrece la metodología al reducir significativamente el espacio de búsqueda de la solución, y la ganancia de la implementación en paralelo.

Índice general

Agradecimientos

Resumen I

Índice general V

Lista de figuras VII

Lista de tablas XI

Publicaciones XV

1. Introducción 1

1.1. Fundamentación 2

1.1.1. Inundaciones por Desborde del Cauce de Ríos 2

1.1.2. Modelos de Simulación de Inundaciones 3

1.2. Motivación 5

1.3. Objetivos de la Tesis y Metodología 8

1.3.1. Metodología 8

1.4. Contribución de la tesis 9

1.5. Organización de la tesis 9

2. El problema de la predicción de inundaciones 11

2.1. Ciencia Computacional e Ingeniería 12

2.2. La Simulación en Computadoras 16

2.3. Simulación de Inundaciones 18

2.3.1. Componentes del Modelo de Cauce 19

2.3.2. Del Modelo Conceptual al Modelo Computacional 20

2.3.3.	Incertidumbre de los Modelos de Inundaciones	22
2.4.	El Simulador Hidrodinámico EZEIZA	23
2.4.1.	Características del Modelo Ezeiza	24
2.4.2.	El Modelo del Río Paraná	25
2.5.	Discusión	28
3.	Mejora de la predicción mediante OvS	31
3.1.	Optimización	32
3.1.1.	La Función Objetivo y las Restricciones	33
3.1.2.	¿Qué es un Óptimo?	34
3.1.3.	Funciones Simple y Múltiple Objetivo	35
3.1.3.1.	Optimalidad de Pareto	36
3.1.4.	Una Clasificación de los Métodos de Optimización	38
3.2.	Optimización vía Simulación	41
3.2.1.	Optimización vía Simulación en Sistemas Hidrológicos	42
3.3.	Metodología propuesta para mejorar Ezeiza	44
3.4.	Un Metamodelo de Optimización vía Simulación	45
3.4.1.	Función Objetivo	46
3.4.2.	El Espacio de Búsqueda	47
3.5.	El Método de OvS en Dos Etapas	48
3.5.1.	El Método de Monte Carlo	48
3.5.2.	El Método K-Means	50
3.5.3.	Descripción de la Metodología	50
3.6.	Implementación en Paralelo del Método de OvS	51
3.7.	Discusión	52
4.	Aplicación de la metodología de OvS	55
4.1.	Las Condiciones de la Simulación	56
4.1.1.	Configuración de los Datos del Dominio	57
4.1.2.	Las Condiciones de Borde del Río	59
4.2.	Las Variables de Decisión del Modelo del Río Paraná	59
4.2.1.	Simulación Paramétrica	60
4.3.	Metricas de Evaluación de OvS	62
4.4.	Métodos de Evaluación	63
4.5.	Casos de estudio	65

4.5.1. Prueba de Concepto del Simulador	67
4.6. Caso de Estudio A	70
4.6.1. Índice de Calidad <i>IC</i>	71
4.6.2. Índice de Calidad <i>IC</i> más Índice de Mejora <i>IM - IC&IM</i>	75
4.7. Caso de Estudio B	79
4.7.1. Índice de Calidad <i>IC</i>	80
4.7.2. Índice de Calidad <i>IC</i> más Índice de Mejora <i>IM - IC&IM</i>	83
4.8. Ganancia lograda con el uso HPC	87
4.8.1. Caso de Estudio A	87
4.8.2. Caso de Estudio B	88
4.9. Discusión	89
5. Conclusiones y Líneas de Trabajo Futuras	93
5.1. Líneas futuras	96
Apéndices	98
A. Condiciones de Borde del Modelo del Paraná	101
B. Modelo Numérico de Ezeiza	105
Referencias	107

Índice de figuras

1.1. Caminos para estudiar un sistema	3
2.1. CSE como capas concéntricas.	13
2.2. Paradigmas de la Ciencia.	15
2.3. Modelado y Simulación según <i>Society for Computer Simulation</i>	17
2.4. Componentes de la escorrentía en el sistema físico	18
2.5. Componentes de la escorrentía en el modelo conceptual del fenómeno de inundaciones.	21
2.6. Ubicación del Río Paraná en sus tres tramos modelados.	26
2.7. Red de cálculo del modelo hidrodinámico.	27
3.1. Óptimo local y global.	35
3.2. Optimalidad de Pareto. Ejemplo ilustrativo.	37
3.3. Clasificación de métodos no-analíticos de optimización.	40
3.4. Coordinación entre simulación y optimización.	42
3.5. Metodología propuesta.	45
3.6. Función Objetivo.	47
3.7. Método de Optimización vía Simulación de Dos Etapas.	49
3.8. Esquema del algoritmo de la primera etapa del método de OvS.	52
4.1. Esquema del proceso de simulación paramétrica.	61
4.2. Parámetros que caracterizan una subsección del río.	62
4.3. Esquema de la sintonización de la simulación con Ezeiza para todo el espacio de búsqueda en la prueba de concepto.	64
4.4. Mejora lograda a lo largo de la serie temporal en la ciudad de Goya.	69
4.5. Mejora lograda a lo largo de la serie temporal en la ciudad de San Martín.	69
4.6. Mejora lograda a lo largo de la serie temporal en la ciudad de Rosario.	70

4.7. Mejor conjunto de configuraciones identificado por el método Monte Carlo.	72
4.8. Grupos de configuraciones identificadas por el método K-Means.	72
4.9. Comparación entre los valores de IC obtenidos en cada estación y la configuración INA, para 5 configuraciones del grupo identificado con K-Means.	73
4.10. Comparación entre los valores de IC para el Esc.OvS en cada estación y la configuración INA, en el Caso A.	75
4.11. Configuraciones que optimizan IC e IM	76
4.12. Grupos de configuraciones identificados por K-Means, considerando $IC&IM$.	76
4.13. Comparación de los valores de IC con 5 configuraciones del grupo identificado por K-Means, en cada una de las 15 estaciones. Se comparan también con el valor de IC correspondiente a la configuración INA. Caso que considera $IC&IM$ como objetivo.	77
4.14. Comparación entre los valores de IC para el Esc.OvS en cada estación y la configuración INA, en el Caso A y tomando como objetivos $IC&IM$.	78
4.15. Configuración identificada por el método Monte Carlo.	80
4.16. Grupos de configuraciones identificadas por el método K-Means.	81
4.17. Comparación entre los valores de IC obtenidos con 3 configuraciones del grupo identificado con K-Means y la configuración INA.	81
4.18. Comparación entre los valores de IC para el Esc.OvS en cada estación y la configuración INA, en el Caso B.	83
4.19. Configuraciones que optimizan IC e IM	84
4.20. Grupos de configuraciones identificados por K-Means, considerando $IC + IM$.	84
4.21. Comparación de los valores de IC con cada configuración del grupo identificado por K-Means, en cada una de las 15 estaciones. Se comparan también con el valor de IC correspondiente a la configuración INA. Caso que considera $IC&IM$ como objetivo.	85
4.22. Comparación entre los valores de IC para el Esc.OvS en cada estación y la configuración INA, en el Caso B y tomando como objetivos $IC&IM$.	86
4.23. Comparación de la ganancia lograda por la reducción del espacio de búsqueda y el uso de HPC en el Caso de Estudio A.	88
4.24. Comparación de la ganancia lograda por la reducción del espacio de búsqueda y el uso de HPC en el Caso de Estudio B.	88
4.25. Mejora lograda a lo largo de la serie temporal en la ciudad de Goya.	92

A.1. Caudales erogados por Yacyretá - 1/1/1999 al 31/12/1999	101
A.2. Alturas en Villa Constitución - 1/1/1999 al 31/12/1999	102
A.3. Alturas en Puerto Pilcomayo - 1/1/1999 al 31/12/1999	102
A.4. Caudales erogados por Yacyretá - 1/1/1999 al 31/12/1999	102
A.5. Alturas en Villa Constitución - 1/1/1999 al 31/12/1999	103
A.6. Alturas en Puerto Pilcomayo - 1/1/1999 al 31/12/1999	103

Índice de tablas

2.1. Red de Cálculo: Modelo del sistema Paraná-Paraguay	27
4.1. Datos de las estaciones de monitoreo.	58
4.2. Explosión combinatoria de escenarios en función de las secciones consideradas.	60
4.3. Valores de las variables de decisión que proveen los escenarios solución de la búsqueda exhaustiva.	68
4.4. Índices de calidad y mejora lograda en tres estaciones.	68
4.5. Variables de decisión en los escenarios determinados por el método K-Means.	74
4.6. Solución óptima lograda por la BE acotada.	74
4.7. Mejora lograda con la configuración óptima de Caso A, considerando <i>IC</i> como objetivo.	74
4.8. Solución óptima lograda con el método MC + K-Means considerando los objetivos <i>IC&IM</i>	77
4.9. Mejora lograda con la configuración óptima considerando <i>IC&IM</i> como objetivo, para el Caso A.	78
4.10. Variables de decisión en los escenarios determinados por el método K-Means.	82
4.11. Solución óptima lograda por la BE acotada.	82
4.12. Mejora lograda con la configuración óptima considerando <i>IC</i> como objetivo, para el Caso B.	83
4.13. Solución identificada con el método MC + K-Means considerando los objetivos <i>IC&IM</i>	85
4.14. Mejora lograda con la configuración óptima considerando <i>IC&IM</i> , para el Caso B.	86
4.15. Tiempo de ejecución y ganancia lograda en cada etapa del método de OvS, al aplicar HPC en la implementación.	88

4.16. Tiempo de ejecución y ganancia lograda en cada etapa del método de OvS, al aplicar HPC en la implementación.	89
---	----

Publicaciones

Revistas

- Gaudiani, A., Luque, E., García, P., Re, M, Armando De Giusti, A. Naiouf, M. *How a computational method can help to improve the quality of river flood prediction by simulation*, (2015). Advances in New Trends Springer Series 'Progress in IS' – Alemania. ISSN 978-3-8142-2317-9 - En prensa -
- Gaudiani, A., Luque, E., García, P., Re, M, Armando De Giusti, A. Naiouf, M. *Computing, a powerful tool for improving the parameters simulation quality in flood prediction*, (2014). Procedia Computer Science – Volume 29, p. 299-309, ISSN 1877-0509 - <http://dx.doi.org/10.1016/j.procs.2014.05.027>

Publicaciones relacionadas

- Argollo, E., Gaudiani, A., Rexachs, D., Luque, E. *Tuning application in a multi-cluster environment*. (2006). Springer - Lecture Notes in Computer Science, Volume 4128/2006, p. 78-88. http://dx.doi.org/10.1007/11823285_9

Proceedings en Congresos

- Gaudiani, A., Luque, E., García, P., Re, M, Armando De Giusti, A. Naiouf, M. (2014). *Computational method for prediction enhancement of a river flood simulation*, (2014). Proceedings of the 28th EnviroInfo 2014 Conference, Oldenburg, Germany. p. 325-332.
<http://dblp.uni-trier.de/rec/bibtex/conf/enviroinfo/GaudianiLGRNG14>
- Gaudiani, A., Luque, E., Armando De Giusti, A. Naiouf, M. *Un método de sintonización para mejorar la salida de un modelo computacional de cuenca de ríos*, (2013). XIX Congreso Argentino de Ciencias de la Computación. Universidad CAECE, Argentina.

<http://sedici.unlp.edu.ar/handle/10915/31096/browse?authority=4163&type=author>

Proceedings en Workshops

- Gaudiani, A., Luque, E., Armando De Giusti, A. Naiouf, M. *Simulación paramétrica paralela. Aplicación a modelos de Predicción de inundaciones*, (2008). X Workshop de Investigadores en Ciencias de la Computación. Procesamiento Paralelo y Distribuido. UNLPam, Argentina. http://weblidi.info.unlp.edu.ar/wp/?page_id=36#a2008

Proceedings en Workshops - Relacionados

- Naiouf, M., De Giusti, A, De Giusti, L., Chichizola, F., Sanz, V., Leibovich, F., Rucci, E., Gallo, S., Montes de Oca, E., Frati, E., Sanchez, M., Gaudiani, A. *Fundamentos de cómputo paralelo y distribuido de altas prestaciones*. Construcción y evaluación de aplicaciones. (2014). XVI Workshop de Investigadores en Ciencias de la Computación, WICC 2014. ISBN 978-950-34-1084-4. <http://sedici.unlp.edu.ar/handle/10915/40706/browse?authority=4163&type=author>

Capítulo 1

Introducción

En este capítulo se hará una breve introducción al problema de la simulación de cauce de ríos y los motivos que provocaron el interés en el desarrollo de esta tesis, haciendo una revisión de los trabajos relacionados con cada tema. Para describir estos motivos es necesario hacer una reseña previa de los conceptos principales del fenómeno natural de inundaciones provocadas por el desborde de ríos [1.1](#) y del modelado de los ríos como herramienta para estudiar el sistema físico y predecir su evolución. También se da un panorama del estado del arte de estos temas. Luego de la motivación [1.2](#), se expondrá el objetivo principal de este trabajo y la metodología utilizada [1.3](#). Finalmente se presentará la organización del resto de la tesis [1.5](#).

1.1. Fundamentación

Los fenómenos naturales son sucesos o transformaciones usuales que ocurren en la naturaleza sin la participación directa del hombre. Estos fenómenos pueden considerarse desastres naturales cuando provocan daños y destrucción, afectando directamente la vida de los seres humanos y la de otras especies. En áreas donde no existen intereses humanos a vulnerar, los fenómenos naturales no se consideran un peligro ni causan desastres. Las inundaciones terrestres ocurren cuando se excede la capacidad de los lechos de los ríos para conducir el agua y ésta rebalsa las riberas. El asentamiento de la población en llanuras de inundación es la causa principal de los daños producidos por las inundaciones. A continuación se brindan más detalles sobre la importancia de este evento que demanda constante estudio y desarrollo de modelos de predicción.

1.1.1. Inundaciones por Desborde del Cauce de Ríos

El fenómeno de inundaciones, provocadas por el desborde del agua que transporta el cauce de ríos, es un tema crítico en la población mundial con una relevancia en constante aumento desde hace décadas ([Jonkman y Vrijling, 2008](#)). Todos los medios que puedan ayudar a comprender y predecir estos fenómenos de la naturaleza ocupan un lugar preponderante en campos de la ciencia y la ingeniería. De estas áreas provienen principalmente, los conocimientos y la tecnología que permiten afrontar estos eventos naturales .

Durante el último siglo, dentro del contexto de un Cambio Climático, las precipitaciones han ido en sostenido aumento en el mundo, especialmente el sur de Sudamérica ha mostrado las mayores tendencias de aumento. La frecuencia de eventos de precipitación que exceden los 100 mm. en el centro y este de Argentina se ha triplicado desde la década de 1970, causando inundaciones y pérdidas de vidas y bienes ([Re y Menéndez, 2007](#)). La cuenca del Plata, ubicada en esta región, es de especial interés para el desarrollo de esta tesis. Sus características hacen que las actividades dependientes de los recursos hídricos resulten vulnerables a estos cambios, especialmente por la ocupación urbana y agropecuaria en zonas que hasta hace 40 años no tenían riesgo de inundación ([Camilloni y Barros, 2003](#)) ([Camilloni, Saurral, y Montroull, 2013](#)).

1.1.2. Modelos de Simulación de Inundaciones

El fundamento racional para usar la simulación en cualquier disciplina es la búsqueda constante del hombre para adquirir conocimientos relativos a la predicción del futuro. La simulación es una de las técnicas que permiten estudiar las respuesta de los sistemas ante diferentes estímulos y predecir su comportamiento, condicionada a supuestos previos.

Al hablar de sistema, se está haciendo referencia a cualquier proceso, organismo, fenómeno, objeto, sociedad o ente de cualquier clase, material o abstracto. Un sistema se puede definir como un conjunto de elementos que interactúan entre sí con una finalidad lógica, y la relación entre ellos establece cuál es su estructura. Cada componente del sistema es un subconjunto de variables de éste. El estado del sistema se define con el conjunto de variables necesarias para describir el sistema en un momento en particular. La necesidad de experimentar con el sistema y estudiar su comportamiento requiere la construcción de modelos cuando la experimentación con el sistema real no se puede realizar. Law y Dalton (Law y Kelton, 2000) ofrecen una sencilla y clara descripción de los medios para estudiar un sistema, el cual está graficado en la Fig.1.1

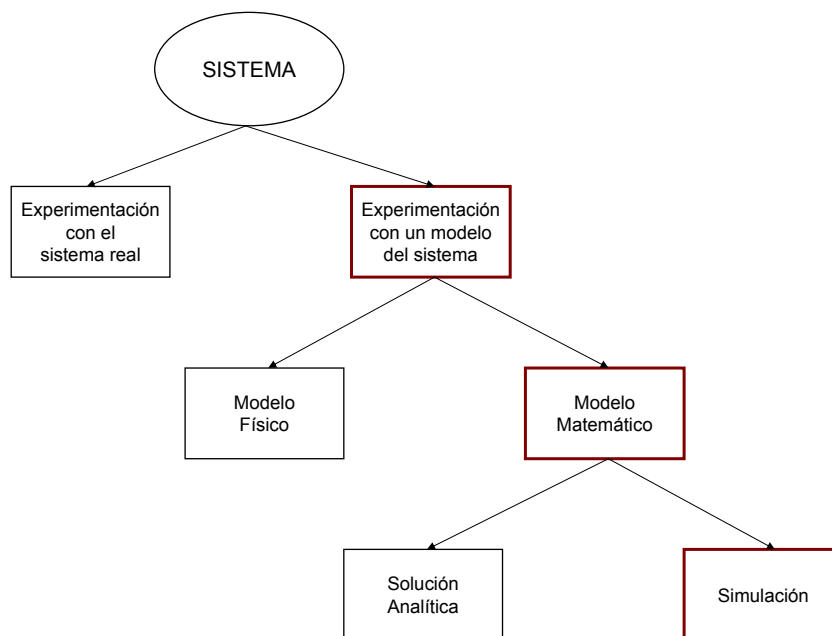


Figura 1.1: Caminos para estudiar un sistema

Es de especial interés de este trabajo los sistemas físicos o sistemas de la naturaleza y en particular, los fenómenos de inundaciones. Estos fenómenos son sistemas complejos que resultan casi imposibles de estudiar en el mundo real y deben ser modelados mediante modelos matemáticos para simular su comportamiento. Un modelo de cuenca de un río es un grupo de abstracciones matemáticas que describen las etapas más importantes del ciclo hidrológico, con el objetivo de simular la conversión de la precipitación que afecta la cuenca, en escurrimiento o caudal que no llega a infiltrarse bajo la superficie de la tierra. Esta definición y una amplia discusión del tema se puede consultar en (Ponce, 1989) (Coulthard y Van De Wiel, 2012).

La creación de modelos matemáticos de estos sistemas combinados con la gran capacidad de cálculo de las computadoras permiten que se pueda simular diversos fenómenos relacionados con los recursos hídricos. Estos modelos hidrológicos se convierten en una herramienta valiosa que aporta un mayor conocimiento del sistema, optimizan su gestión y permiten predecir su evolución futura. En el trabajo de Re et al. del 2008, sobre el modelado matemático del cauce del Río Salado, río de llanura ubicado en Argentina, los autores expresan que los usos principales de estos modelos son diagnosticar o comprender como funciona el sistema, predecir como apoyo a la gestión del sistema y efectuar análisis retrospectivos en eventos que resultaron en conflicto (Re, Badano, Lecertúa, Re, y Menéndez, 2008). Con el modelo usado en este trabajo, pudieron brindar un explicación sobre las causantes de las inundaciones extraordinarias que se sucedieron desde fines de la década del 80 hasta inicio del año 2000.

Por otro lado, la posibilidad de realizar estudios retrospectivos permite estudiar el efecto de inundaciones al modificar el medio físico que las produce, como ser si se pusieran diques o drenajes. El siguiente artículo, (Mejuto, Mayorga, y Fernández, 1997), da una visión panorámica de los modelos hidrológicos, tanto los que modelan el escurrimiento del agua en superficie como los de interés en hidrogeología. En los últimos años, la bibliografía que expresa el rol primordial de los modelos en la gestión de recursos hídricos aumentó a medida que también aumentó la necesidad de contar con ayuda en la toma de decisiones ante eventos graves de inundación. Algunos autores se enfocan en los daños económicos y pérdidas de vida (Dilley, Chen, Deichmann, Lerner-Lam, y Arnold, 2005) y en la importancia de tomar decisiones acertadas de protección ante inundaciones, tarea que se apoya en los resultados de modelos usados para pronosticar el avance del agua (Balica, Popescu, Beevers, y Wright, 2013). En este trabajo, (Hansson, Danielson, y Ekenberg, 2008), los autores combinan un modelo de inundaciones y un modelo de evaluación de estrategias

para crear un marco de trabajo que consta de una herramienta de decisión (framework) que se enfoca en el impacto social y ambiental de una inundación. El impacto de inundaciones costeras es tratada en (Re y Menéndez, 2007) mediante un modelo hidrodinámico del Río de la Plata, en Argentina, que permite cuantificar el riesgo del aumento del nivel del agua y fue utilizado para predecir la respuesta del río frente a hipotéticos escenarios futuros, con el propósito de brindar una alerta a la región.

La información que proveen los modelos de simulación hidrológica es muy útil en muchos contextos del quehacer humano. La navegación y transporte de carga, la producción hidroeléctrica, el riego y suministro de agua dulce constituyen ejemplos de actividades que se benefician con el aporte de estas predicciones (Tucci y Collischonn, 2006). En el ámbito de la ingeniería, los modelos de simulación de procesos físicos son utilizados principalmente con el propósito de pronosticar su evolución. Este es el caso de los modelos hidrodinámicos que predicen las alturas del cauce del río. Estos simuladores describen la interacción entre la topografía del cauce y el movimiento del agua, aunque lo hacen de manera simplificada a pesar que éstos son procesos muy complejos.

A continuación se citan algunos ejemplos de modelos computacionales de gran difusión en la comunidad de ingenieros e investigadores en esta área. La página oficial de US Geological Service (USGS) ofrece a la comunidad modelos de aguas superficiales, subterráneas, o de hidroquímica ¹. El modelo público HEC-RAS, de amplia difusión, es desarrollado y mantenido por el Centro de Ingeniería Hidrológica del Cuerpo de Ingenieros de la Armada de los EE.UU. ² HEC-RAS es un programa de hidráulica para modelizar inundaciones por desborde de un cauce natural o artificial. Este sitio también desarrolló HEC-HMS, el cual calcula el caudal generado por precipitaciones. La Universidad Politécnica de Valencia, España, desarrolló y puso a disposición de la comunidad científica el Modelo TETIS, el cual es un modelo de simulación hidrológica que puede resolver tanto problemas de crecidas y erosión como de recursos hídricos ³. El modelo que se utiliza en este trabajo se presentará de manera detallada en el próximo capítulo.

1.2. Motivación

Como se dijo anteriormente, la construcción de modelos para estudiar inundaciones por desborde de ríos es una de las áreas primordiales de investigación y desarrollo en todo

¹<http://water.usgs.gov/software/>

²<http://www.hec.usace.army.mil/software/hec-ras/index.html>

³<http://lluvia.dihma.upv.es/ES/software/software.html>

el mundo, como todas aquellas dedicadas a estudiar y predecir los desastres naturales. Estos modelos son herramientas que se han convertido en una ayuda indispensable en el pronóstico de crecidas. La mejora en la predicción demanda de constante investigación en materia de modelización meteorológica e hidrológica, con el propósito de crear modelos más ajustados al sistema real.

Una gran cantidad de modelos fueron diseñados para predecir las crecidas y todos, en mayor o menor medida, presentan distintas fuentes de incertidumbre. Los modelos constituyen una abstracción de la realidad y al simplificarla dejan de lado algunos aspectos del sistema real. Brugnach sostiene: “cuando un modelo es utilizado para predicción, se espera capturar las características esenciales del sistema a modelar para imitarlo de manera transparente y fiable” (Brugnach y Pahl Wostl, 2008). Recordemos que la estructura del sistema se basa en reconocer los componentes que lo forman y como interactúan entre sí y realizar esta tarea es un trabajo que no responde a un protocolo sino que depende de la experiencia que tenga el grupo humano que interviene. Warmink et al. identifican y cuantifican la incertidumbre de un modelo hidrodinámico de ríos que proviene del contexto de desarrollo del modelo y de su estructura, y tratan su impacto en el contexto de aplicación del modelo (Warmink, Van der Klis, Booij, y Hulscher, 2011). Pappenberger et al. estudian el rol de la precisión de las soluciones numéricas y su impacto en la capacidad predictiva del modelo, analizando la incertidumbre en la estructura, en los datos de entrada y en los parámetros (Pappenberger y cols., 2006). Por otro lado, los modelos hidrodinámicos se los clasifica de manera genérica como de una dimensión (1D) o de dos dimensiones (2D). Respondiendo esta clasificación a la técnica seleccionada para la discretización espacial del río. Los modelos 1D son eficientes computacionalmente, aunque a veces el modelado del río requiere un mallado más preciso debido a las características de la topografía (Finaud-Guyot, Delenne, Guinot, y Llovel, 2011). En estos casos es preciso modelar y simular con un modelo 2D y es entonces cuando se aprecia los avances en la ciencia de la computación, ya que hicieron posible el uso de estos modelos computacionales que demandan mucho más poder de cómputo que los modelos 1D. Estos últimos suelen llevar tiempos de ejecución de minutos, mientras que los 2D pueden llevar horas (Bladé y cols., 2012).

Aunque los modelos utilizados consideren la mayor cantidad posible de variables involucradas en el proceso, tratando de representar de la manera más certera el fenómeno, existen más fuentes de incertidumbre que aumentan el error de los datos de salida del modelo. Es vital considerar la incertidumbre que proviene de los errores de medición de

los datos de entrada y de los instrumentos involucrados en su captura, sin dejar de lado la producida por la discretización de dichos datos sobre una grilla. El siguiente artículo presenta una reseña de 14 métodos de análisis y valoración de fuentes de incertidumbre en las etapas del modelado de sistemas hídricos (Refsgaard, van der Sluijs, Håjberg, y Vanrolleghem, 2007). La bibliografía dedicada a analizar y tratar la incertidumbre en el modelado de ríos es muy abundante y crece año tras año, demostrando la importancia del tema y del tratamiento de un problema que sigue presente.

Los laboratorios que se dedican a la investigación y alerta diaria del estado de los ríos generalmente utilizan modelos computacionales especialmente creados y verificados para esa región en particular. Esto implica haber realizado el modelado del río a lo largo del tramo del cauce que se pretende estudiar. Esta ardua y costosa tarea requiere que personal especializado realice mediciones sobre el sistema real, apoyado en captura remota de datos y en el uso de otros recursos de medición, con el propósito que dichos datos sean utilizados por largos períodos de tiempo. De esta manera, se puede caracterizar la topografía del sistema o dominio del modelo. El río es un sistema dinámico y los valores que definen su dominio suelen tener variaciones continuamente, no siendo posible mantener actualizados estos valores. Esta tesis tomó en especial consideración la simulación del cauce del Río Paraná, el cual se caracteriza por tener grandes urbes asentadas en sus orillas. Hoy en día, se cuenta con un sistema hidrológico de la cuenca del Paraná para brindar alertas a la población que pueda verse afectada ante eventos extremos de crecida o bajante, tanto por estar asentada en su zona de influencia como por realizar actividades económicas que dependen del estado del río. Nos referiremos con más detalle sobre este tema en el próximo capítulo.

Proveer de mayor certeza este tipo de simulador fue el motivo que impulsó el tema central de este trabajo. Pero, el método de optimización del proceso de simulación, que se describe a lo largo de la tesis, consume un elevado tiempo de cómputo al requerir el lanzamiento en paralelo del simulador con gran cantidad de juegos de datos. Esto hace indispensable recurrir a *High Performance Computing* (HPC) para lograrlo. Los aportes de estas áreas del conocimiento para simular con más calidad y generar más conocimientos sobre los fenómenos naturales en estudio, resulta muy valioso. Parte de la motivación del trabajo fue poder contar con el avance suficiente de los recursos computacionales para lograr los objetivos planteados con más detalle en la sección 1.3.

1.3. Objetivos de la Tesis y Metodología

El objetivo principal de la tesis es brindar una metodología para mejorar la calidad de la simulación de un modelo computacional de cauce de ríos, aplicando la técnica de optimización vía simulación. De esta manera, se mejora un simulador que es parte de un sistema mayor de predicción del comportamiento del río. En consecuencia, se mejora una herramienta de soporte al servicio de los ingenieros hidráulicos y el personal especializado que se encargan de brindar las alertas a la población de la región afectada por posibles crecidas o bajantes del río.

1.3.1. Metodología

Para alcanzar el objetivo principal es necesario lograr lo objetivos parciales que se listan a continuación. Estos definen la metodología de la tesis.

- Conocer el comportamiento del simulador en función de cada escenario de simulación posible. Cuando se habla de escenario de simulación se hace referencia a los procesos, métodos y parámetros que serán utilizados en la sesión de simulación. La metodología implementada requiere el lanzamiento en paralelo del simulador con diferentes escenarios, en un modo básico de simulación paramétrica. Para esta etapa, se crean diferentes escenarios de simulación mediante la combinatoria de los valores de un conjunto de parámetros críticos que funcionan como variables de ajuste. Mediante búsquedas exhaustivas se llega a tener un mayor conocimiento de la respuesta del simulador a lo largo de todo su dominio mediante una exploración paso a paso de todo el espacio de búsqueda.
- Enunciar una metodología de optimización vía simulación que provea una solución para mejorar la calidad del simulador. En esta etapa se busca reducir el espacio de búsqueda mediante la implementación de un método de optimización vía simulación que inicialmente combina una heurística compuesta de un método Monte Carlo más un método de agrupamiento K-Means. Posteriormente, se busca la mejor solución posible (óptima o subóptima) mediante una búsqueda exhaustiva sobre el espacio ya reducido. O sea, de un conjunto prometedor de soluciones se busca la mejor de todas.
- Encontrar una solución al problema de optimización vía simulación en tiempos de cómputo aceptables. Encontrar una solución que optimice la simulación, es un tipo de

problemas que consume gran cantidad de cómputo. A pesar de que se presenta una técnica para reducir el espacio de búsqueda y, en consecuencia, el tiempo de cómputo, no es factible llevar el procesamiento adelante sin recurrir a las técnicas de procesamiento en paralelo y HPC.

1.4. Contribución de la tesis

Este trabajo está enfocado en la aplicación de los fundamentos y las técnicas de la matemática y la ciencia computacional para brindar mayor calidad a la simulación de un modelo hidrodinámico. De esta manera se brinda un aporte que deriva de enunciar una metodología que puede ser extendida a la simulación de otros fenómenos físicos y que necesitan tratar el problema de la incertidumbre de los datos de entrada.

Por otro lado, se mejora el pronóstico de las alturas del agua, a lo largo del cauce de un río de vital importancia en la región del Plata. Al alcanzar una mejora en las estaciones localizadas en uno de los tres tramos del cauce del Paraná, hablamos de uno a dos metros de diferencia entre la altura del agua simulada aplicando el método de optimización y la altura simulada sin aplicar la mejora. Esta diferencia lograda permite prever medidas de contingencia más acertadas, ya que el valle de inundación puede cubrir zonas habitadas si los niveles del agua aumentan en pocos metros.

Este trabajo deja establecidas las bases para avanzar en esta línea y trabajar en pos de brindar un mayor grado de certeza en los pronósticos que los ingenieros hidráulicos realizan día a día. Por otro lado, esta metodología de optimización vía simulación deja abierta una línea de investigación que pueda aplicarse a modelos hidrodinámicos de mayor grado de complejidad.

1.5. Organización de la tesis

Delineamiento de los capítulos que llevarán adelante los objetivos expuestos:

- **Capítulo 2:** En este capítulo se desarrollan los aspectos principales de la simulación computacional y el estado del arte de la simulación de inundaciones y los modelos desarrollados con este fin. Se detallan las características del simulador Ezeiza, elegido para las experimentaciones necesarias para este trabajo, y del sistema físico del río Paraná cuya simulación se implementa con Ezeiza.

- **Capítulo 3:** En este capítulo se brindan los fundamentos principales de la optimización que son necesarios para presentar la metodología de optimización vía simulación aplicada para mejorar la predicción de Ezeiza. También se presenta la implementación en paralelo del método propuesto.
- **Capítulo 4:** En este capítulo se presentan las experimentaciones utilizadas para la evaluación de la metodología propuesta en el capítulo anterior. Se detallan las condiciones de dos casos de evaluación y se discuten los resultados logrados.
- **Capítulo 5:** En este capítulo se concluye la disertación de este trabajo, se presentan las conclusiones principales y los futuros trabajos que se derivan de esta investigación.
- **Apéndices:** incluyen breve material complementario sobre el simulador Ezeiza y las condiciones de la simulación.

Capítulo 2

El Simulador de Cauce de Ríos y el Problema de la Predicción de Inundaciones

La simulación de fenómenos naturales es un problema considerado como de gran desafío para la ciencia y la ingeniería debido a la complejidad de los modelos que los representan. La simulación computacional permite tratar estos sistemas complejos, valiéndose de los conceptos y técnicas de la matemática y la ciencia de la computación para obtener un modelo computacional, y así implementar en computadora el modelo conceptual. La simulación en computadoras es usada como modelo de los sistemas reales para evaluar las respuestas obtenidas y tomar decisiones en áreas estratégicas del quehacer humano. En este capítulo se desarrollan los aspectos principales de la simulación computacional, que se fundamenta en la ciencia computacional y la ingeniería [2.1](#), para luego presentar los aspectos fundamentales de la simulación [2.2](#) y en particular de la simulación de inundaciones [2.3](#), haciendo hincapié en las distintas fuentes de errores que impactan en sus datos de salida [2.3.3](#). Se brindan las características del simulador seleccionado para esta tesis y del modelo del río Paraná sobre el que se realizan las experiencias [2.4](#).

2.1. Ciencia Computacional e Ingeniería

La ciencia computacional (*computational science*) es el área del conocimiento que posibilitó resolver problemas imposibles de tratar hace algunas décadas atrás. Por ejemplo, por ser experiencias muy costosas de llevar adelante, o en lugares inaccesibles para las personas o por ser muy peligrosas. Estos problemas complejos son generalmente aquellos problemas de sumo interés para la humanidad y poder experimentar con ellos es la única manera de estudiarlos, comprender sus límites y predecir su comportamiento. Es una ciencia que aprovecha los avances en la arquitectura de computadoras y el desarrollo de algoritmos que las controlan.

Aunque, la ciencia computacional es mucho más que el soporte de las investigaciones científicas. Sus principios y técnicas han cambiado el modo de hacer ciencia, permitiendo nuevos avances científicos a través de permitir nuevas clases de experimentaciones. Algunos investigadores la consideran como el cuarto método para hacer ciencia, junto con los métodos de observación, experimentación y teóricos. La postura de Denning y sus colegas, es considerarla como el cuarto gran dominio de las ciencias junto con las ciencias físicas, naturales y sociales, no siendo una *herramienta para la ciencia* sino siendo completamente integral a ella, permitiendo el pensamiento científico y el descubrimiento (Denning, 2010).

Para D.E. Stevenson en 1993 (Stevenson, 1993), qué era ciencia computacional estaba en debate y presentaba su definición como:

“La ciencia computacional es una disciplina emergente que se caracteriza por el uso de computadoras para proporcionar información detallada sobre el comportamiento de sistemas físicos complejos.”

Por otro lado, Gallopoulos y Sameh en 2002 hablan de ciencia computacional e ingeniería (CSE) como una ciencia en plena expansión, cuya principal ocupación está enfocada en dar entornos de resolución de problemas de la ciencia y la ingeniería (Gallopoulos y Sameh, 2002). En lugar de vistas piramidales o de intersecciones con otras ciencias, presentan un interesante esquema de capas concéntricas que “encapsulan los procesos y el flujo de la información en actividades de desarrollo e investigación en CSE”. Cada capa requiere de específicas actividades con la computadora, y los datos e información se mueven a través de ellas. Este esquema se muestra en la Fig. 2.1.

En 2005, el “2020 Science Group” presentó un detallado informe a la comunidad científica con su visión de la ciencia hacia el 2020 (Group, 2005), y el rol de la ciencia de la computación para alcanzar esta visión:

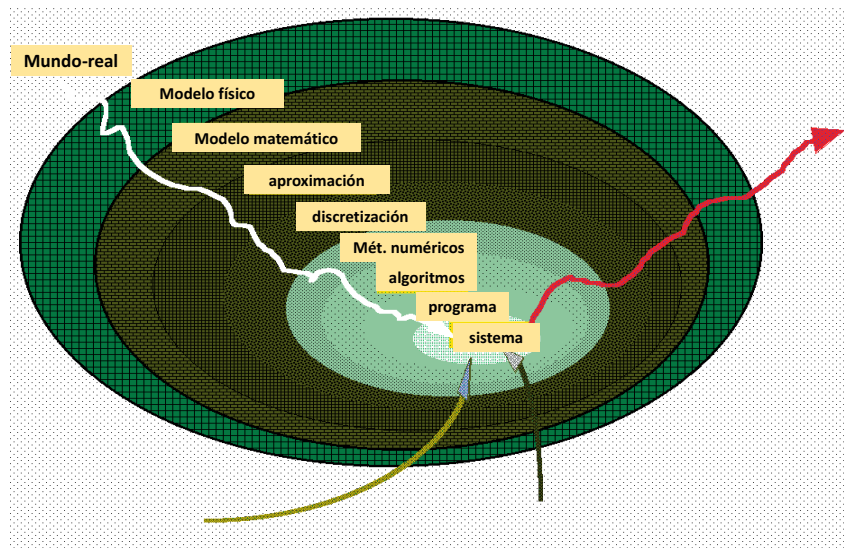


Figura 2.1: CSE como capas concéntricas.

“Creemos que los conceptos de la informática y las herramientas de la ciencia forman una tercera, y vital componente de un *triángulo de oro* a formarse con novedosas técnicas matemáticas y estadísticas en materia de ciencia junto a plataformas informáticas científicas y aplicaciones integradas dentro de la ciencia experimental y teórica.”

La ciencia computacional es una ciencia interdisciplinaria en constante crecimiento, la cual hace uso de las ciencias naturales, la matemática y la informática. Se enfoca en el desarrollo de modelos confiables de los sistemas físicos, el desarrollo de algoritmos necesarios para el proceso de modelado y su implementación en arquitecturas adecuadas para ejecutarlos de manera eficiente (Stevenson, 1993). Los aportes de los modelos y simulaciones computacionales constituyen una valiosa contribución al tratamiento de problemas en muchas áreas del conocimiento. Los científicos que se desempeñan en ciencia computacional mantienen todo su saber en informática, pero hacen énfasis en las aplicaciones científicas o ingenieriles. Como expresa Jeannette Wing, pensar como un científico de la computación es mucho más que poder hacer buenos programas de computadora, es pensar en múltiples niveles de abstracción (Wing, 2006).

La ciencia computacional no debe ser confundida con la informática. Una definición comúnmente aceptada de la informática (*computer science*) es: “La informática es la ciencia del procesamiento de algoritmos, representación, almacenamiento y transmisión de

la información”, la cual incluye aspectos de la matemática, las ciencias naturales como también de la ingeniería. A diferencia de la ciencia computacional que, a través de la informática y la matemática, estudia y trata problemas de la ingeniería y de las ciencias naturales, básicas o sociales.

Mientras que los modelos computacionales no pueden reemplazar el laboratorio, sin duda, éstos se han convertido en una parte intrínseca de la búsqueda global para el conocimiento científico, de la mano de la CSE. La CSE emergió como un área del conocimiento interdisciplinario que creció aceleradamente en las últimas décadas, con el objetivo de comprender la dinámica de sistemas complejos y predecir su comportamiento.

El diseño e implementación de modelos hidrodinámicos que se utilizan para estudiar el flujo del agua en cauces naturales, como el de los ríos, permite comprender la dinámica de esta clase de sistemas complejos. Estos modelos son una herramienta indispensable para predecir los fenómenos de inundaciones, siendo estos eventos el fenómeno físico en el que se basa el desarrollo de este trabajo. Los ingenieros hidráulicos que usan estos modelos son quienes se benefician de los aportes de la ciencia computacional.

Por otro lado, la comunidad científica nos muestra una moderna visión acerca del cuarto paradigma de la investigación científica. Haciendo una mirada hacia atrás en la evolución de la ciencia computacional y viendo su posición actual podemos encontrar las siguientes etapas:

- En el origen contábamos con la ciencia empírica, basada en la observación y la descripción de los fenómenos naturales. El discípulo de Platón, Calipo de Cycique (siglo IV a.C.), uno de muchísimos científicos que se dedicaron a la observación de los astros, con su trabajo advirtió cierta irregularidad del movimiento de los planetas.
- En los últimos siglos aparece la ciencia teórica. Dos casos de gran repercusión en la historia de la astronomía son los científicos teóricos Johannes Kepler, en 1609, e Isaac Newton, en 1687. Kepler formuló las tres leyes del movimiento planetario, basándose en las mediciones registradas durante años por el científico observacional Tycho Braher, de quién era discípulo. Newton se basa en las leyes de Kepler para enunciar su teoría de gravitación universal.
- Ya desde la segunda mitad del milenio pasado, la ciencia computacional y la simulación computacional a gran escala llevó el modo de hacer ciencia más allá del mundo físico, gracias a la aparición y vertiginosa evolución de las computadoras.

Los modelos teóricos fueron creciendo en tamaño y complejidad, siendo necesario resolverlos mediante simulación y con la “colaboración” de las computadoras.

- Actualmente, los datos que resultan de las simulaciones, junto a los que provienen de las ciencias experimentales, se generan en enormes cantidades. Esto hace que en este momento nos encontremos ante un nuevo paradigma: “data-intensive science” (Bruballa, Taboada, Cabrera, Rexachs, y Luque, 2014) (Hey, Gannon, y Pinkelman, 2012). La ciencia intensiva en datos trabaja en conjunto con la teoría, la experimentación y la simulación.

Esta visión fue graficada por James Gray en sus disertaciones sobre los paradigmas de la ciencia, como la última que dió en 2007 ^{1 2}, y que se presenta en la Fig. 2.2.

Paradigmas de la ciencia

- Mil años atrás:
La ciencia era **empírica**
- Descripción de fenómenos naturales
- Últimos siglos:
Rama **teórica**
- Utilización de modelos, generalizaciones
- Últimas décadas:
Rama **computacional**
- Simulación de fenómenos complejos
- Hoy:
Exploración de datos (eScience)
- Se unifica teoría, experimentación y simulación.
 - Los datos se capturan mediante instrumentos o se generan mediante simulador.
 - Procesados mediante software.
 - La información/conocimientos se almacenan en computadora.
 - El científico analiza la base de datos o los archivos mediante administración de datos y estadística.



$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G\rho}{3} - K\frac{c^2}{a^2}$$

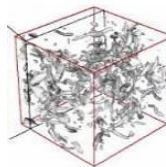


Figura 2.2: Paradigmas de la Ciencia.

Como se dijo anteriormente, en este trabajo se presenta una técnica de optimización vía simulación para sintonizar el simulador de cauce de ríos, brindando mayor calidad al modelo computacional, tema que se desarrolla en el capítulo siguiente. Esta metodología

¹Esta presentación fue la última publicada en la página web de Jim en Microsoft Research.

²<http://research.microsoft.com/en-us/um/people/gray/talks/NRCCSTBeScience.ppt>

requiere una gran cantidad de ejecuciones del simulador, ingresando los datos de escenarios completos del río y produciendo una enorme cantidad de datos resultantes de las simulaciones. En una primera etapa del trabajo, cuando se realizan experimentaciones que responden a simulaciones paramétricas, se recolecta una gran cantidad de datos de salida que deben ser procesados para comprender el comportamiento del modelo. Esta información es necesaria para implementar posteriormente el método de optimización. La sintonización del simulador requiere el procesamiento de los enormes volúmenes de datos que resultan de estas ejecuciones. La ciencia computacional y su nueva concepción como intensiva en datos, está presente en todas las etapas de esta tesis.

2.2. La Simulación en Computadoras

R. E. Shanon definió de esta manera la simulación (Shannon, 1998):

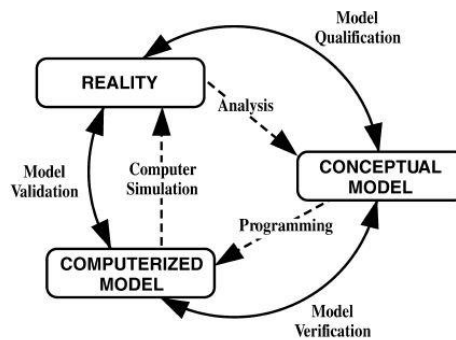
“La simulación es el proceso de diseñar un modelo de un sistema real y llevar a cabo experimentos con él, a fin de comprender el comportamiento del sistema y evaluar nuevas estrategias - dentro de los límites impuestos por un criterio o conjunto de ellos - para operar el sistema.”

Una manera muy difundida y genérica de definir la simulación es, esencialmente, viéndola como una técnica que conduce a construir modelos de una situación real aunada a la realización de experimentos con el modelo. O sea, es una técnica que simula, o imita de la manera más cercana posible, el funcionamiento de situaciones reales o sistemas del mundo real. Cuando estos procesos se desarrollan en el tiempo, su simulación lleva a pensar en temporalidad y en representar la dinámica temporal de los sucesos.

Los ingenieros en sistemas y la comunidad de científicos en investigación operativa han desarrollado y formulado muchos de los principios generales y los métodos del modelado y la simulación. En 1979, el *Technical Committee on Model Credibility of the Society for Computer Simulation*³ desarrolló un diagrama que identifica las fases del modelado y la simulación, el cual se muestra como fue creado originalmente en la Fig. 2.3.

Posteriormente, hubo consenso en la comunidad científica al enunciar los siguientes pasos para definir el proceso de simulación (Law y Kelton, 2000) (Oberkampf, DeLand, Rutherford, Diegert, y Alvin, 2002) (Sargent, 2005):

³<http://www.scs.org/>



© 1979 by Simulation Councils, Inc.

Figura 2.3: Modelado y Simulación según *Society for Computer Simulation*.

1. *Modelo conceptual del sistema físico.* Constituye una especificación del sistema físico y de su entorno, determinando qué eventos o procesos físicos serán tomados en cuenta.
2. *Modelo matemático del modelo conceptual.* Su complejidad depende de la complejidad del fenómeno considerado e incluye los conjuntos de ecuaciones, condiciones iniciales y de contorno.
3. *Discretización y selección del algoritmo para el modelo matemático.* Se enfoca en convertir una forma matemática continua en una matemática discreta, teniendo en cuenta temas como convergencia y estabilidad de los métodos numéricos, como también decisiones que derivan de las experiencias computacionales que se llevarán adelante.
4. *Programa computacional del modelo discreto.* El modelo se convierte en código de un lenguaje de computadora, aprovechando el nivel de madurez alcanzado hoy en día por la programación, luego de los grandes esfuerzos de diseño y desarrollo durante décadas. Desarrollo que alcanza la programación en paralelo para arquitecturas híbridas y en grid computing.
5. *Solución numérica del modelo computacional.* Se tiene en cuenta que sólo existen valores discretos y soluciones discretas con una precisión finita. El espacio, el tiempo y las variables sólo existen en una distribución discreta de puntos de una grilla.

6. *Representación de las soluciones numéricas.* Se refiere a la representación que permita una correcta interpretación de las soluciones logradas y de los errores que contengan. Esto incluye visualización de datos, realidad virtual y todas las técnicas de post-procesamiento de datos.

2.3. Simulación de Inundaciones

Un evento de inundación provocado por el desborde del cauce natural de ríos consiste en una onda de crecida de baja amplitud pasando por un tramo del río. Cuando se habla de ondas de alta amplitud estamos en presencia de inundación de las extensas áreas bajas adyacentes, las cuales pueden actuar como zonas de almacenamiento temporario o generar una ruta alternativa para transportar el flujo de agua. Las laderas y ríos tributarios adyacentes al cauce contribuyen a la recolección del agua de lluvia, alimentando el flujo de la inundación (Stewart, Bates, Anderson, Price, y Burt, 1999). En la Fig. 2.4 se puede ver una esquema del modelo físico correspondiente al proceso de escorrentía⁴.



Figura 2.4: Componentes de la escorrentía en el sistema físico

Las tendencias del modelamiento de procesos hidrológicos dependen del problema a tratar. En general, se describe cualquier subsistema del ciclo hidrológico como un sistema

⁴Fuente: ©The COMET Program - UCAR Community Programs

en si mismo y se representa en forma de modelo estocástico o determinístico; en ambos casos se puede realizar la simulación computacional, dependiendo del modelo propuesto. El traslado de ondas por el cauce del río se puede considerar un subsistema de un sistema más grande que contempla otros eventos que contribuyen a la intensidad del flujo.

Para describir un sistema es necesario partir de las leyes físicas que gobiernan su comportamiento, las cuales junto con la entrada (precipitación, cantidad de humedad, datos del suelo, etc.) se combinan para producir la salida esperada (escorrentía, nivel o caudal del agua en diferentes puntos del cauce).

El modelo del cauce de un río es un grupo de abstracciones matemáticas que describen fases relevantes del ciclo hidrológico, con el objetivo de simular la conversión de la precipitación en escorrentía (Ponce, 1989). Se llama escorrentía, o escurrimiento, a aquella parte de la lluvia, del agua de deshielo y/o del agua de irrigación que, en lugar de infiltrarse en el suelo, fluye hacia el cauce fluvial desplazándose sobre su superficie. Medir la escorrentía es el elemento más importante en la predicción de crecidas del río, y las características de su cuenca determinarán el área que afecta.

La predicción de la evolución del flujo en el cauce es posible mediante modelos numéricos. Estos modelos son la representación de las ecuaciones de conservación de masa y momento, mediante un conjunto de ecuaciones diferenciales parciales no-lineales, las cuales deben ser discretizadas en espacio y tiempo para ser resueltas numéricamente. La aplicación de técnicas numéricas a la resolución del sistema de ecuaciones diferenciales que se establece como modelo matemático de esta clase de flujos, se remonta a la década de 1960 (Cunge, Holly, y Verwey, 1980) (Pujol y Menéndez, 1987).

A continuación, se presentan los elementos que componen los modelos de inundaciones 2.3.1 y la implementación del modelo computacional 2.3.2, para hablar luego los errores de los modelos computacionales que impactan en los resultados y disminuyen la calidad de la simulación 2.3.3.

2.3.1. Componentes del Modelo de Cauce

Los principales componentes considerados al diseñar un modelo de cauce de ríos son:

- **Precipitación:** Constituye el dato de entrada al proceso que se modela y puede ser lluvia o nieve. Los datos que la determinan son: intensidad, duración, frecuencia, distribución espacial y temporal.

- **Abstracción hidrológica:** Son los procesos físicos que permiten medir la precipitación efectiva que afecta al modelo, como ser infiltración del suelo, evaporación, almacenamiento en superficie y transpiración. Son determinadas por las propiedades físicas de la cuenca.
- **Escurrimiento:** Es la salida del modelo, determinada por el flujo de salida. Representa el movimiento del cauce fluvial.

La precipitación es el dato de entrada al modelo, las abstracciones hidrológicas son determinadas por las propiedades de la cuenca y el escurrimiento es la salida del modelo.

2.3.2. Del Modelo Conceptual al Modelo Computacional

Los modelos de simulación se encuentran clasificados dentro de cada una de las siguientes tres dimensiones: 1) Estático vs. Dinámico, 2) Determinístico vs. Estocástico y 3) Continuo vs. Discreto.

Los modelos de procesos hidrológicos también se los suele clasificar como lineales vs. no lineales, globales vs. distribuidos, analíticos vs. numéricos o de eventos aislados vs. procesos continuos. Una detallada explicación del tema se encuentra en la bibliografía básica de hidrología (Ponce, 1989) (Ramirez, 2000).

El ingeniero que se dispone a modelar un tramo o todo el cauce de un río, puede adoptar dos enfoques diferentes, o alguna combinación de ambos. Uno es el enfoque hidrológico, en cuyo caso sólo se interesará en saber la diferencia entre el volumen de agua que entra y que sale del tramo considerado. Utilizan la ley de conservación de la masa pero realizan simplificaciones en el funcionamiento del sistema. Por otro lado, puede adoptar un enfoque hidráulico, en cuyo caso tomará en cuenta el cambio de volumen y los cambios de energía del sistema fluvial. Este enfoque requiere considerar más datos de la topografía y del flujo de agua. Frente a condiciones de flujo más complejas, es preciso recurrir a métodos hidráulicos para calcular el caudal. Los modelos hidráulicos ofrecen más precisión en los cálculos, pero el software computacional requiere más especialización para resolver ecuaciones más complejas. Debido a que se deben representar muchas condiciones físicas del flujo del agua, como también las obstrucciones y estructuras físicas, estos modelos requieren grandes cantidades de datos para describir la topografía del lecho fluvial y las secciones transversales del cauce.

El desarrollo de modelos de ríos también debe cumplir con los cinco pasos que se indicaron en la sección 2.2. El modelo matemático es estudiado mediante un modelo de

simulación computacional. Los modelos numéricos, esquemas, y algoritmos son frecuentemente utilizados como parte integral de modelos computacionales que simulan todas las fases relevantes del ciclo hidrológico. Los modelos matemáticos describen el comportamiento del sistema mediante ecuaciones matemáticas que representan las relaciones entre el estado del sistema, la entrada y la salida (Jacovkis, 1990).

En la Fig. 2.5 se muestra un esquema de los componentes principales del sistema conceptual del escurrimiento involucrados en el proceso de predicción de las inundaciones.

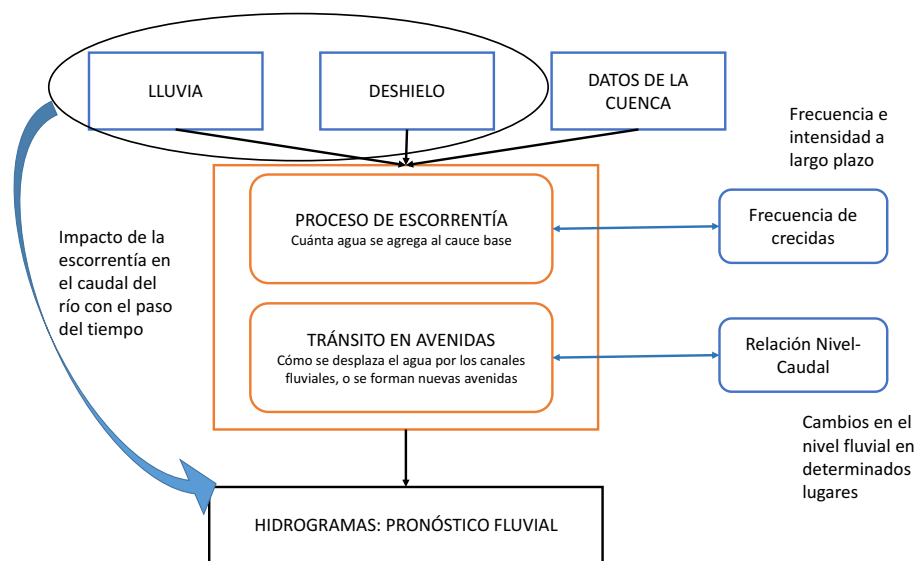


Figura 2.5: Componentes de la escorrentía en el modelo conceptual del fenómeno de inundaciones.

Una simulación numérica tiene que describir el fenómeno con suficiente precisión pero a bajo costo. La utilidad del método está condicionada por la precisión al describir el proceso físico y las leyes físicas básicas del problema que se compromete a resolver. Este compromiso debe tener limitaciones para que el modelo tenga utilidad práctica. Como expresan Pujol y Menéndez (Pujol y Menéndez, 1987):

Establecer un modelo teórico del escurrimiento significa esquematizar el flujo y sus contornos, y desechar los efectos dinámicos menos importantes, una vez identificados los esenciales.

Un ejemplo sería desechar el efecto del aire o de sedimentos en el agua, cuando no sean significativos en el modelo.

Los modelos hidráulicos utilizados para simular eventos de inundaciones y calcular la velocidad de la onda transmitida se clasifican en modelos hidrodinámicos 1D o 2D, según las dimensiones consideradas. Los modelos del escurrimiento de una dimensión, o modelos 1D, son ampliamente usados y suelen estar basados en las ecuaciones de Saint Venant en una dimensión o un esquema que se obtiene de Navier-Stokes. Se resuelven numéricamente, generalmente usando el método de diferencias finitas. Estos métodos son precisos en el cauce principal del río. Cuando se necesita una buena representación del flujo hacia la planicie de inundación, se requieren modelos de dos dimensiones, o modelos 2D. Los métodos numéricos se expresan en dos dimensiones y utilizan mapas 2D con datos de la topografía, o sea cada celda de la grilla tiene sus propios valores. La elección de las dimensiones del modelo deben tener en cuenta cuestiones de implementación del modelo computacional y los recursos de cómputo disponibles, ya que los modelos 2D demandan alto tiempo computacional. En algunas simulaciones se combinan modelos de ambas dimensiones según los requerimientos y tipo de río modelado (Timbe Castro y Willems, 2011) (Bladé y cols., 2012).

En la sección 2.4 se presenta el simulador seleccionado para esta tesis, el cual es un modelo computacional hidrodinámico 1D y se utiliza para simular el cauce del Río Paraná y Paraguay, en su tramo perteneciente a la Argentina.

2.3.3. Incertidumbre de los Modelos de Inundaciones

El buen rendimiento de todo modelo, independientemente de su dimensión, depende de la disponibilidad de los datos. Por otro lado, estos datos constituyen una de las fuentes de incertidumbre en los resultados de la simulación.

Los modelos hidrológicos necesitan al menos los siguientes datos para computar el nivel del agua y la velocidad del flujo: - Las secciones que dividen el dominio del río y los intervalos de discretización. - El índice de rugosidad del lecho del río y de las planicies adyacentes. - Las condiciones de contorno en el inicio del tramo considerado, $Q(t)$, o caudal en función del tiempo. - Las condiciones de contorno al fin del tramo, $Q(h)$, o caudal de descarga en función de la altura.

En todas estas etapas, existen fuentes de incertidumbre del modelo. La incertidumbre consiste de inseguridad e imprecisión. En el modelado de ríos, la incertidumbre más fácilmente cuantificada es la que proviene de la entrada y los parámetros del modelo, aunque se considera también la incertidumbre de la estructura interna del modelo.

La incertidumbre debida a los parámetros de entrada al modelo, se debe a que no pueden ser medidos directamente, como sucede con el coeficiente de rugosidad de la superficie del cauce del río y las planicies de inundación. Cuando esto sucede, el valor del parámetro debe ser interpolado a cada punto de discretización de todo el dominio del cauce. El valor de este parámetro depende de la clase y distribución de la vegetación dentro del área del modelo, pero ésta es clasificada en clases y las clases de vegetación son discretizadas en una grilla, por lo cual la incertidumbre que provoca depende del tamaño de las celdas de la grilla y la disponibilidad de todos los datos.

La incertidumbre en los datos de entrada que definen la geometría del cauce principal y de las planicies pueden tener un sustancial impacto en la salida del simulador. Un ejemplo son los valores de las precipitaciones que se dan en diferentes lugares del cauce. Otro parámetro que influye en la calidad de la salida es la altura de los albardones o formaciones naturales generadas por los sólidos depositados en las márgenes del cauce principal durante las crecidas. La incertidumbre debido a mediciones es causada por el instrumento y el método de medición utilizado, como también por la imposibilidad de actualizar los valores de estos datos frecuentemente.

El interés de este trabajo es la implementación de una mejora del simulador mediante un método de optimización que devuelve como solución un conjunto de parámetros optimizados. La mejora lograda en la calidad del simulador es medida en términos del ajuste entre los datos de salida, o simulados, y los datos reales, u observados. Este tema se desarrolla en el capítulo siguiente.

2.4. El Simulador Hidrodinámico EZEIZA

El desarrollo de esta tesis requirió la selección de un modelo computacional de cauces. El modelo seleccionado fue el programa EZEIZA VI (Ezeiza), un modelo hidrodinámico desarrollado por el Instituto Nacional del Agua (INA) de la República Argentina (Menéndez, 2002). Esta selección responde a los siguientes motivos.

- Ezeiza es el resultado de la investigación y desarrollo de un grupo de científicos e ingenieros de la Argentina que comenzó en 1974 en el Instituto Nacional de Ciencias y Técnicas Hídricas (INCYTH), actualmente INA, y ha continuado hasta la actualidad en constantes mejoras a cargo del Programa de Hidráulica Computacional (PHC) del Laboratorio de Hidráulica (LHA) del INA (Menéndez, 1996).

- Ezeiza se utiliza para la implementación de un modelo hidrodinámico formado por el sistema de ríos Paraná y Paraguay, en sus tramos ubicados en la Argentina. El río Paraná forma parte de la Cuenca del Plata, que se caracteriza por las grandes urbes localizadas en su zona de influencia, incluso en sus planicies de inundación.
- Ezeiza tiene como su principal aplicación ser parte del sistema utilizado para el pronóstico que actualmente lleva adelante el Servicio de Información y Alerta hidrológico de la Cuenca del Plata, correspondiente al INA. Este servicio alerta a la población costera sobre eventos de desborde o de bajantes que puedan afectar la actividad económica o poner en riesgo la vida de los habitantes. También brinda las alturas diarias del río en diferentes puntos de su cauce.

En 2011, el Ing. G. Latessa desarrolló un trabajo de mejora del modelo del río Paraná haciendo un análisis del rendimiento del modelo y mejorando la representación geométrica del Paraná para así mejorar la salida de Ezeiza. Se lograron mejoras como se detalla en su informe (Latessa, 2011), pero aun se presentan diferencias entre los datos del simulador y los observados en las estaciones a lo largo del cauce. Todo tratamiento de la simulación que mejore el servicio de alertas hidrográficas es de sumo interés de los especialistas a cargo.

Los motivos detallados anteriormente sumados a poder disponer del código fuente del programa Ezeiza, como también contar con una gran base de datos con los valores observados sobre el río Paraná a lo largo de 15 años, gracias a los investigadores del INA, fueron razones suficientes para elegir Ezeiza.

A continuación, se habla de las características del modelo hidrodinámico Ezeiza 2.4.1 y de la implementación del modelo con el sistema formado por tramos de los ríos Paraná y Paraguay 2.4.2.

2.4.1. Características del Modelo Ezeiza

Ezeiza es la representación computacional de un modelo matemático para el cálculo de la traslación de ondas en ríos y canales. Se basa en la integración numérica de las ecuaciones de Saint-Venant, para cada uno de estos canales y es capaz de describir el comportamiento hidrodinámico del sistema mediante un tratamiento unidimensional del cauce del río, debido a que la mayoría de los ríos y arroyos se caracterizan por poseer una dirección privilegiada de escurrimiento (Pujol y Menéndez, 1987) (Jaime y Menéndez, 1997). Se brindan más detalles del modelo numérico de Ezeiza en el Apéndice B.

Los datos a ingresar al modelo son los siguientes:

1. conectividad de los cauces que integran la red (cantidad y distribución geográfica de los mismos);
2. secciones transversales en ubicaciones arbitrariamente elegidas (el modelo se encarga luego de interpolar para satisfacer su condición de paso espacial);
3. características de resistencia al escurrimiento de los canales, representada a través del coeficiente de Chezy - Manning;
4. condiciones de borde de los cauces;
5. condiciones iniciales del modelo;
6. parámetros numéricos, tales como paso espacial y temporal.

El sistema computacional entrega las siguientes salidas básicas:

1. perfil de traslación la onda sobre los cauces;
2. hidrogramas de caudal y nivel en secciones previamente especificadas;
3. mapas de inundación.

2.4.2. El Modelo del Río Paraná

El Río Paraná, en la Cuenca del Plata, recorre el área más poblada e industrializada de Sudamérica, vinculando las dos mayores regiones económicas del subcontinente: el estado de Sao Paulo, al norte, y el eje fluvial-industrial Santa Fe-La Plata, al sur. El tramo del río Paraná simulado con el modelo actual, se extiende desde el embalse de Yacyretá (Corrientes) hasta la ciudad de Villa Constitución (Santa Fe), mientras que el tramo del río Paraguay va desde Puerto Pilcomayo (Formosa) hasta su confluencia con el Paraná, como puede apreciarse en la Fig. 2.6. Una amplia información puede encontrarse en (Jaime y Menéndez, 1997).

El mapa base de la zona modelada fue realizado mediante la digitalización de los croquis de los ríos Paraná y Paraguay a escala 1:50.000, publicados por el Servicio de Hidrografía Naval en 1980 y corregidos en 1984. Estos mapas sirvieron para caracterizar la geografía global de los cauces y obtener así la geometría del dominio del modelo, que fue ajustada posteriormente.

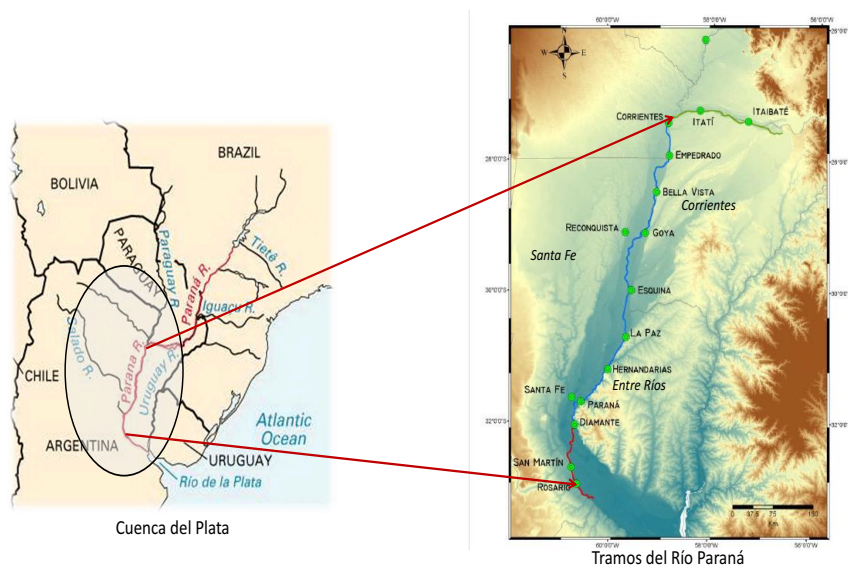


Figura 2.6: Ubicación del Río Paraná en sus tres tramos modelados.

El sistema de ríos constituye el sistema a modelar, el cual se debe caracterizar para su implementación con Ezeiza. Esta caracterización, representada en la Fig. 2.7, considera los siguientes ítems:

La extensión del dominio, el cual está delimitado por los tramos de los ríos indicados previamente.

La red de cálculo, cuyos datos se muestran en la tabla 2.1, provee al modelo de los datos de las secciones transversales en que se discretiza la geometría del sistema. La geometría del canal se provee a través de secciones ubicadas sobre el río. Pero, el paso de cálculo espacial es constante, por lo cual se generan valores intermedios a las secciones, cuyos valores se resuelven por interpolación lineal entre las dos secciones más cercanas.

Los parámetros numéricos que se utilizan para discretizar el dominio, siendo de 5 km. el paso espacial del modelo. En el caso del paso temporal se calculó sabiendo la velocidad promedio de la onda, obteniendo un $\Delta = 10$ minutos. Este paso permitió la convergencia del método numérico.

Rugosidad y albardones, son los parámetros más sensibles del modelo. Para cuantificar la resistencia hidráulica se dispone como parámetro del coeficiente de rugosidad de Manning, cuyo valor depende de la superficie que pone resistencia al escurrimiento. Es por ésto que se cuenta con un valor de *rugosidad para el cauce* y de *rugosidad para la planicie de*

inundación (Arcement y Schneider, 1989). Los albardones son formaciones naturales generadas por los sólidos depositados en las márgenes del cauce principal durante las crecidas funcionando como un umbral y son tomados en cuenta en la modelación matemática.

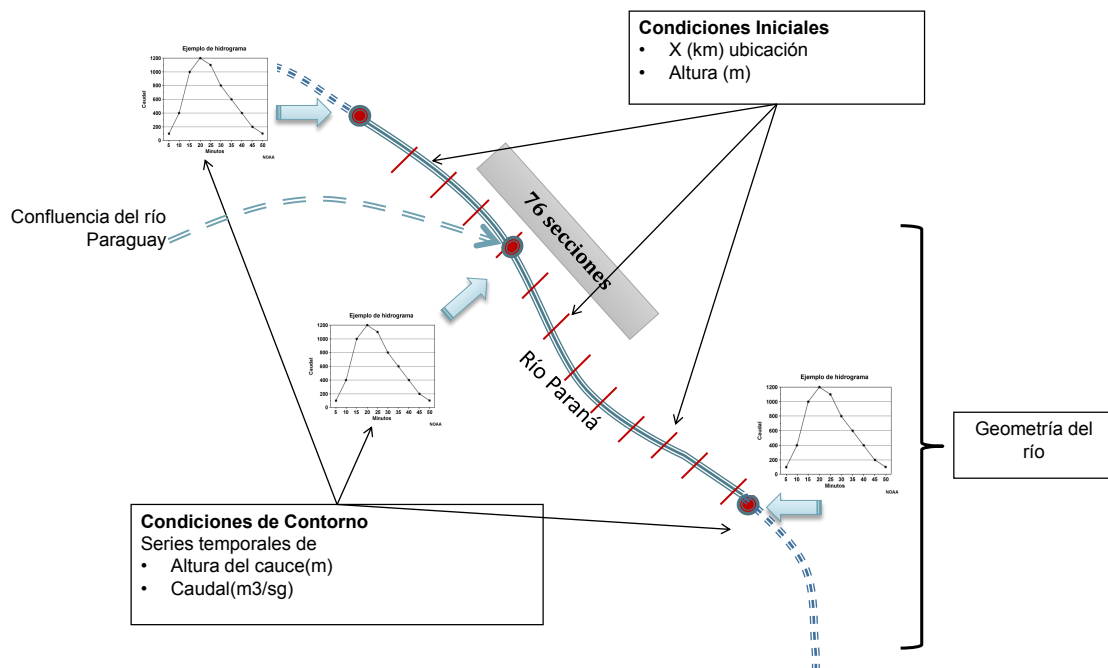


Figura 2.7: Red de cálculo del modelo hidrodinámico.

Las estaciones de seguimiento, o monitoreo, están ubicadas sobre el cauce del río Paraná y pueden verse en la Fig. 2.6. Cada una de las 15 estaciones señalan el nivel del río cada día. Sus valores son medidos diariamente por personal de Prefectura Naval Argentina y enviadas al INA. Estas series temporales de valores fueron utilizados para realizar la calibración del modelo.

Tabla 2.1: Red de Cálculo: Modelo del sistema Paraná-Paraguay

<i>Cauce</i>	<i>Curso</i>	<i>Long.(Km)</i>	<i>Secciones Ag.Arr.</i>	<i>Bordes Ag.Aba.</i>	<i>Bordes</i>
1	Paraná	1.083	76	Caudal Q	Nivel H
2	Paraguay	376	77	Nivel H	—

El Río Paraná se estudia separadamente en tramos, donde las características climáticas, geomorfológicas e hidrológicas se diferencian, estos tramos son de interés para las experiencias que se realizarán con el modelo y se detallan en el capítulo siguiente. A continuación se describe la extensión de estos tramos, que se pueden ver en la Fig. 2.6:

- **Tramo Superior:** Desde el cierre de Yacyretá (Corrientes) hasta la confluencia con el Paraguay, frente a la ciudad de Paso de la Patria (Corrientes).
- **Tramo Medio:** Desde la confluencia hasta Diamante (Entre Ríos).
- **Tramo Inferior:** Desde Diamante (Entre Ríos) hasta Villa Constitución (Santa Fe).

Mucho más detalle sobre el modelo hidrodinámico del Paraná se puede encontrar en los informes que realizó el INA en estos años de desarrollo y aplicación de Ezeiza en la predicción de inundaciones (Jaime y Menéndez, 1997) (Latessa, 2011).

2.5. Discusión

1. Se mostró inicialmente, cómo los avances en la ciencia de modelización y simulación ayudaron de manera muy beneficiosa al entendimiento y la simulación de este tipo de fenómenos. Se brindó el estado del arte de la simulación de fenómenos hidrológicos y el problema de la incertidumbre en los parámetros de entrada, el cual impacta directamente en la predicción que provee el simulador como sucede en toda simulación de procesos físicos. En las publicaciones que se hacen referencia, se pone en evidencia la necesidad de contar con herramientas para manejar esta situación.
2. La mayoría de estos parámetros pueden evaluarse a partir de datos de campo, por ejemplo, observaciones hidrometeorológicas, mapas de la topografía, tipos de suelo, uso del suelo, etc. Sin embargo, rara vez están disponibles datos de campo completos para apoyar plenamente la especificación de todos los parámetros del modelo.
3. Algunos de estos parámetros son de carácter más conceptual y no se pueden evaluar directamente. Por lo tanto, algunos de ellos deben ser determinados a través de un proceso de calibración. Esta situación es evidente al estimar el parámetro *coeficiente de Manning*, el cual no puede ser calculado y obtenerse su valor exacto, ya que depende de datos de la geometría del río los cuales no pueden conocerse con certeza.

4. Los datos de entrada del modelo están modificando sus valores continuamente con los cambios físicos que se producen en el cauce y planicies propias de la región del río considerada, o sea en el dominio de la simulación. Esta situación se manifiesta en la simulación del cauce del río Paraná, con el simulador Ezeiza.
5. Se detallaron las características principales del simulador Ezeiza y del sistema físico formado por el río Paraná y Paraguay. De esta manera, se brindaron los motivos por haberlos seleccionado para este trabajo.

Por todas las razones expuestas en este capítulo, resulta de suma utilidad brindar una metodología de calibración automática para sintonizar el simulador y así obtener más precisión en los datos de salida, siendo este objetivo el aporte principal de esta tesis. Resaltando que este aporte resulta muy beneficioso pues permite realizar este proceso tantas veces como sea necesario durante la simulación. Sus características se desarrollarán a lo largo de los capítulos que siguen.

Capítulo 3

Mejora de la Predicción Mediante Optimización vía Simulación

La búsqueda de estados óptimos fue siempre el anhelo de la humanidad, es por esto que la optimización es una antigua parte de la ciencia que está presente en la vida diaria. Por ejemplo, en economía se busca la manera de obtener el mayor rendimiento posible empleando la mínima cantidad de recursos o con el menor costo posible, mientras que en la física se plantean los problemas de maximización de la entropía y minimización de la energía. Los modelos de optimización son usados en casi todas las áreas de toma de decisiones. Sus métodos posibilitan encontrar la solución que proporciona el mejor resultado, la que logra mayores ganancias, mayor producción o satisfacción, o la que logra el menor costo o malestar. Como ya se mencionó, en esta tesis se presenta una mejora a la calidad de predicción del simulador Ezeiza, mediante un método de optimización vía simulación. En consecuencia, esta mejora también beneficia el sistema de alertas de eventos extraordinarios en la región de influencia del cauce del río Paraná, del cual Ezeiza es un componente hidrodinámico que predice las alturas del agua en las estaciones de monitoreo distribuidas en el tramo considerado. En este capítulo se brindan los fundamentos de la optimización en 3.1, para presentar la metodología de optimización vía simulación, en 3.2, y su aplicación a mejorar la predicción de Ezeiza en 3.3. En 3.4 se presenta en líneas generales el metamodelo de optimización vía simulación utilizado, para luego dar más detalles del método de dos etapas en 3.5. A continuación se dan detalles de la paralelización implementada en 3.6 para finalizar con la discusión de los temas presentados en este capítulo en 3.7.

3.1. Optimización

En la vida real se presentan problemas en los cuales el ser humano se encuentra ante la disyuntiva de elegir entre distintas alternativas. Estos problemas corresponden a áreas del conocimiento que requieren mejorar determinados procesos obteniendo las mejores soluciones posibles. Por otro lado, la complejidad de los problemas aumenta si los intereses de las personas que los plantean están dados por múltiples objetivos a lograr, por lo general contrapuestos entre sí, y demandan respuestas rápidas. Esta situación se profundiza si se toman decisiones sobre un sistema cuya información, necesaria para analizar su comportamiento, es imprecisa o proviene de un ambiente de incertidumbre. La optimización es la ciencia que proporciona las herramientas de conceptualización y análisis necesarias para buscar una solución a estos problemas de la vida diaria, con el máximo beneficio y el mínimo esfuerzo, y con el objetivo de ayudar en la toma de decisiones dentro de sus áreas de incumbencia. Como muchas otras áreas del conocimiento, esta ciencia fue beneficiada por el inmenso incremento de la capacidad de procesamiento de las computadoras en las últimas décadas, especialmente por permitirle manejar y encontrar soluciones a problemas de optimización de gran escala.

La optimización es definida como el proceso de encontrar la solución óptima, ó subóptima, de un problema bajo ciertas condiciones. Por lo general, es un proceso que dependerá del conocimiento que se tenga del problema, y un poco de sentido común al respecto. Si este problema se lo puede tratar con la suficiente abstracción seguramente se lo puede representar matemáticamente. Los métodos que buscan la solución de estos problemas se dividen, por lo general, en métodos de optimización globales y locales. Los globales buscan la mejor solución, o sea el máximo o mínimo global del problema, mientras que los locales se concentran en la vecindad de una solución encontrada, por lo cual necesitan técnicas adicionales para acercarse todo lo posible a la solución global (Chong y Zak, 2013).

El planteamiento del problema en términos matemáticos precisos debe reflejar exactamente lo que deseamos resolver. Es necesario tener en cuenta dos pasos para este planteo (Pedregal, 2006):

- Una *función objetivo* o de costo debe medir de manera fiel la idea de optimalidad, dependiente de los objetivos del problema. Si la función objetivo no refleja correctamente los criterios de optimización, seguramente la solución no será la correcta.
- El *conjunto de restricciones* debe estar explícitamente establecido y deberá cumplirse de manera que las soluciones posibles sean verdaderamente factibles para el

problema considerado. Si estas restricciones no se escriben con precisión la solución no será la esperada.

Formalmente, según los pasos anteriores, la optimización está asociada con la especificación de una función matemática f llamada función objetivo y sus restricciones, de esta forma:

$$\begin{aligned} & \max / \min f(x) \\ & \text{sujeto a } x \in S \end{aligned} \tag{3.1}$$

donde x es a variable, f es la función ($f : S \rightarrow \mathfrak{R}$), S es el conjunto de restricciones y se cumple que $\exists x_0 \in S$ tal que $f(x_0) \preceq f(x) \forall x \in S$, para minimización, y $f(x_0) \succeq f(x) \forall x \in S$ para maximización.

En el contexto de este trabajo, el problema principal es un *problema de búsqueda*, el cual es considerado como sinónimo de problema de optimización y puede expresarse así:

$$\text{encontrar } x^* \text{ que optimiza } f(x), \forall x \in S$$

donde $f : \mathfrak{R}^p \rightarrow \mathfrak{R}$ es la expresión de la función objetivo o índice de calidad. El conjunto de parámetros que se pretende ajustar es el vector x de dimensión p y x^* es un vector representativo del conjunto óptimo (o subóptimo) de parámetros. El problema o *espacio de búsqueda* está representado por el dominio $S \subseteq \mathfrak{R}^p$. El dominio podrá ser un arreglo de valores continuos o discretos o cualquier conjunto de elementos, según el problema. La especificación de este espacio proviene de un conjunto de restricciones impuestas sobre los elementos x que forman el dominio, siendo estos elementos las *soluciones candidatas*. Estas restricciones van restringiendo el tamaño de S , delimitando la *región de factibilidad*.

3.1.1. La Función Objetivo y las Restricciones

La necesidad de expresar con claridad la función objetivo permite obtener una medida certera del grado de optimalidad logrado con cada solución, pudiendo hacer una comparación cuantitativa entre cada elemento \vec{x}^* del conjunto de soluciones. Esto es posible cuando se establece una rigurosa expresión matemática del objetivo, o también llamado *índice de rendimiento*.

Los problemas de optimización representados en forma global en 3.1 pueden ser clasificados según la naturaleza de la función objetivo y la restricciones (lineal, no lineal, convexos) o por el número de variables (grande, pequeño), la forma de la función (dife-

renciable, no diferenciable). También se los clasifica en problemas con restricciones sobre las variables o sin restricciones. Estas restricciones pueden ser de distinta naturaleza, por ejemplo, igualdades, desigualdades, ecuaciones diferenciales o integrales, entre otras, y pueden estar expresadas de manera tácita o implícita. Más adelante se hace referencia a la clasificación de los métodos.

3.1.2. ¿Qué es un Óptimo?

La optimización busca la mejor solución, por lo cual es pertinente tener en claro de qué hablamos al buscar una solución óptima. La solución de problemas expresados como 3.1, requiere distinguir clases de óptimos:

- Un máximo local $\vec{x} \in \mathbb{X}$ de una función objetivo $f : \mathbb{X} \rightarrow \mathfrak{R}$ es un elemento del conjunto de entrada tal que $f(\vec{x}) \geq f(x)$ para todo x en la vecindad de \vec{x} . Si $\mathbb{X} \subseteq \mathfrak{R}$ entonces:

$$\forall \vec{x} \exists \epsilon > 0 : f(\vec{x}) \geq f(x) \forall x \in \mathbb{X}, |x - \vec{x}| < \epsilon \quad (3.2)$$

- Un mínimo local $\vec{x} \in \mathbb{X}$ de una función objetivo $f : \mathbb{X} \rightarrow \mathfrak{R}$ es un elemento del conjunto de entrada tal que $f(\vec{x}) \leq f(x)$ para todo x en la vecindad de \vec{x} . Si $\mathbb{X} \subseteq \mathfrak{R}$ entonces:

$$\forall \vec{x} \exists \epsilon > 0 : f(\vec{x}) \leq f(x) \forall x \in \mathbb{X}, |x - \vec{x}| < \epsilon \quad (3.3)$$

- Un óptimo local $\vec{x} \in \mathbb{X}$ de la función objetivo $f : \mathbb{X} \rightarrow \mathfrak{R}$ es tanto un máximo local o un mínimo local.
- Un máximo global $x^* \in \mathbb{X}$ de una función objetivo $f : \mathbb{X} \rightarrow \mathfrak{R}$ es un elemento del conjunto de entrada tal que $f(x^*) \geq f(x) \forall x \in \mathfrak{R}$.
- Un mínimo global $x^* \in \mathbb{X}$ de una función objetivo $f : \mathbb{X} \rightarrow \mathfrak{R}$ es un elemento del conjunto de entrada tal que $f(x^*) \leq f(x) \forall x \in \mathfrak{R}$.
- Un óptimo global $x^* \in \mathbb{X}$ de una función objetivo $f : \mathbb{X} \rightarrow \mathfrak{R}$ es tanto un máximo global como un mínimo global. Esta definición es válida para funciones de simple objetivo.

El caso de la función objetivo con múltiples objetivos se tratará en el próximo ítem.

En la Fig. 3.1¹ se ilustra una función f definida sobre un espacio $x \in \mathbb{R}^2$ y se ve que un óptimo global lo es para todo el dominio de \mathbb{X} , mientras que los óptimos locales lo son para un subconjunto del dominio. No se debe dejar de tener en cuenta que el verdadero sentido de hablar de una solución óptima dependerá del problema al que se refiera.

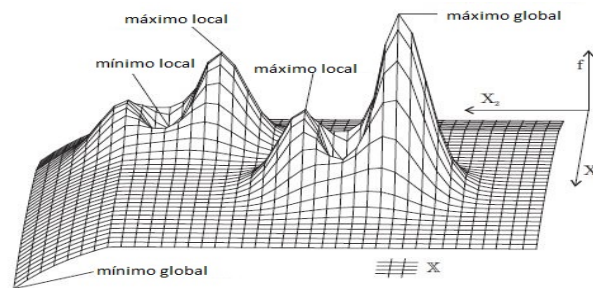


Figura 3.1: Óptimo local y global.

Al enfrentar un problema genérico de optimización, como se expresó en 3.1, por definición se está buscando un óptimo global. La realidad, tanto desde un punto de vista teórico como computacional, nos indica que en muchas circunstancias, se debería estar satisfecho al alcanzar un óptimo local. Neumaier presenta una extensa discusión y reconocimiento de problemas habituales de optimización que requieren una solución global y su estado del arte (Neumaier, 2004). Para profundizar en el tema, este artículo hace una interesante reseña de las investigaciones en optimización determinística desde 1998 a 2008 (Floudas y Gounaris, 2009).

3.1.3. Funciones Simple y Múltiple Objetivo

Cuando el problema de optimización se expresa como el planteado en 3.1, se refiere a un problema de objetivo simple. Es un problema de minimización o de maximización cuya solución puede ser un óptimo global o local.

En muchos problemas de la ingeniería, la industria, la economía y otros campos, se involucra la optimización simultánea de muchos objetivos. Los problemas donde se consideran más de un objetivo se les denomina problemas *multiobjetivo* y su propósito es optimizar todos los objetivos de forma simultánea, aunque estos objetivos, la mayoría de

¹Obtenido de Weise, Thomas. "Global Optimization Algorithms-Theory and Applications. Self Published. 2009"

las veces, se encuentran en conflicto entre ellos. Por ejemplo, no se puede mejorar uno de ellos sin deteriorar los otros.

El problema de optimización de múltiples objetivos se puede expresar en términos matemáticos como:

$$\begin{aligned} \text{optimizar } f(\vec{x}) &= |f_1(x), f_2(x), \dots, f_k(x)| \\ \text{sujeto a : } \Omega &= \{\vec{x} \in \mathfrak{R}^n \mid G(\vec{x}) \leq 0\} \end{aligned} \quad (3.4)$$

$$\begin{aligned} \text{donde } f_i : \mathfrak{R}^n &\rightarrow \mathfrak{R} \quad \forall i : 1..p \\ \text{y } \vec{x} &= [x_1, x_2, \dots, x_n]^T \end{aligned} \quad (3.5)$$

donde Ω es un conjunto no vacío de \mathfrak{R}^n y el vector $G(\vec{x})$ representa las restricciones que debe cumplir la solución, estas pueden ser mayores o menores que cero. Se busca encontrar una solución \vec{x}^* que pertenezca a la región Ω y que optimice $f(x)$ donde cada una de las funciones pueden ser maximizadas o minimizadas, dependiendo del problema.

Cuando se habla de optimalidad, se hace referencia a encontrar la mejor solución posible o una buena aproximación a ésta, dado un conjunto de limitaciones o restricciones. En la optimización de un solo objetivo, es posible determinar entre cualquier par de soluciones si uno es mejor que el otro. Esto permite comparar las posibles soluciones y, por lo general, obtener una única solución óptima (es decir, el óptimo global). Sin embargo, en optimización multiobjetivo no existe un método sencillo para determinar si una solución es mejor que otra. Una forma de hacerlo es comparar soluciones utilizando el método de Pareto por relación de dominancia que, en lugar de una única solución óptima, conduce a un conjunto de alternativas con diferentes compensaciones entre los objetivos. Estas soluciones se llaman soluciones óptimas de Pareto o soluciones no dominadas, las que se encuentran en la *frontera de Pareto* (Abbass, Sarker, y Newton, 2001) (Deb, 2014), tema que se trata a continuación.

3.1.3.1. Optimalidad de Pareto

Los fundamentos matemáticos para la optimización multiobjetivo fueron establecidos por el economista italiano Vilfredo Pareto en 1896. La noción de optimalidad de Pareto es de importancia especialmente, pero no de manera excluyente, en problemas de economía, teoría de juegos, ingeniería y ciencias sociales, ya que define una frontera de soluciones que permite manejar objetivos conflictivos de lograr de una manera satisfactoria. Definir optimalidad en el sentido de Pareto está fundamentado en el concepto de *dominancia*. Este concepto dice que la solución x_1 domina la solución x_2 si la solución x_1 es mejor que

x_2 en al menos una función objetivo y no es peor con respecto a todos los otros objetivos. Esto significa que $x^* \in \mathbb{X}$ es un óptimo Pareto si $\nexists x \in \mathbb{X}$ tal que este vector x domine al vector de decisiones x^* (Park, Datta-Gupta, y King, 2015).

En otras palabras, la definición anterior dice que un conjunto de decisiones es un óptimo de Pareto si no existe otra elección que se pueda tomar dentro del dominio del problema, que haga mejorar alguno de los objetivos sin que empeore de forma simultánea alguno de los otros.

Como ejemplo ilustrativo se muestra en la Fig. 3.2, donde dos funciones objetivo (f_1, f_2) son minimizadas con respecto a dos parámetros (x_1, x_2). En la Fig 3.2a el punto A indica los parámetros que minimizan f_1 , y el punto B los que minimizan f_2 . La solución de este problema multiobjetivo son todos los puntos que se encuentra en la línea que une A con B . Pero, como se ve en la Fig.3.2b, al moverse sobre esa línea aumenta f_1 y disminuye f_2 o viceversa al moverse en el sentido inverso. Esto muestra que cada función tiene su propia entidad y las soluciones están sobre la línea. Se encuentran en la frontera de Pareto y al no contar con información adicional no es posible decidir si alguna es buena solución o mejor que la otra. Se encuentra más información sobre el tema en (Yapo, Gupta, y Sorooshian, 1998).

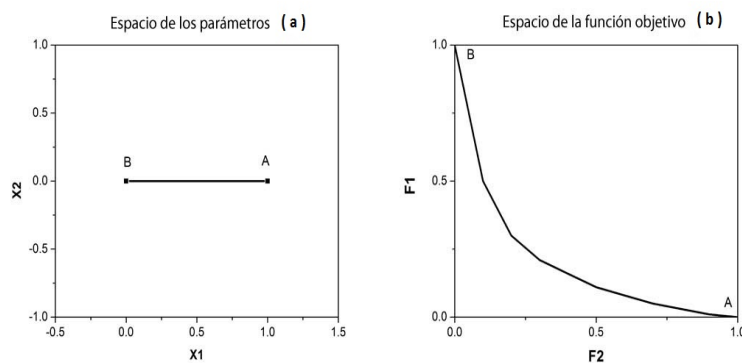


Figura 3.2: Optimalidad de Pareto. Ejemplo ilustrativo.

Hay una gran diferencia entre resolver problemas de simple y múltiples objetivos. Este último caso obliga a reflexionar los criterios a tomar en cuenta para tratar con soluciones óptimas, siendo necesario considerar la optimalidad, la precisión de la solución y la velocidad del método.

Las técnicas habituales para encontrar soluciones necesitan gran cantidad de experimentación ya que suelen proveer de a una solución por vez del conjunto óptimo de Pareto

y se necesita una gran cantidad de ensayos con distintos puntos de inicio para obtener una gran cantidad de puntos de dicho conjunto. Por otro lado, estas técnicas son altamente sensitivas a la forma de la frontera de Pareto, pueden no funcionar si esta no es convexa, por lo cual se necesitan implementar técnicas que superen estas dificultades, como son los métodos numéricos aplicados a solucionar problemas de optimización.

Es importante destacar que la solución a estos problemas de optimización multiobjetivo es utilizada por una persona, quién es la encargada de tomar decisiones sobre el entorno del problema en la vida real. La solución óptima obtenida con el método Pareto representa la "solución más adecuada", o "la preferida" por sobre las otras, por la persona que decide, quién la considera su mejor opción. Ella, o él, conoce en profundidad el problema y puede especificar la información sobre la preferencia respecto a los objetivos y a las diferentes soluciones alternativas (Miettinen, Ruiz, y Wierzbicki, 2008). Los métodos interactivos, que van recibiendo modificaciones sobre el objetivo y las restricciones de modo dinámico, hallan las soluciones convirtiendo el problema multiobjetivo en problemas de optimización de simple objetivo cuyas soluciones óptimas son soluciones del problema multiobjetivo. Métodos de estas características son discutidos y desarrollados por los autores de (Hartikainen, Miettinen, y Wiecek, 2012)

3.1.4. Una Clasificación de los Métodos de Optimización

A modo de ilustración acerca de la historia de los métodos de optimización, cuyo nacimiento se remonta a resolver temas de la ciencia desde mediados del siglo XX, Davidon describe uno de los primeros ejemplos de la aplicación de un método de optimización, en este caso uno de búsqueda directa implementado en una de las primeras computadoras digitales (Davidon, 1991).

Enrico Fermi y Nicholas Metropolis usaron en los inicios de la década de 1950, una de las primeras computadoras digitales, Los Alamos Maniac, para determinar qué valores de ciertos parámetros teóricos (phase shifts) se ajustaban mejor a datos experimentales (scattering cross sections). Ellos variaron un parámetro teórico por vez en pasos de la misma magnitud, y cuando ya ningún incremento o decremento de los parámetros ajustaba mejor a los datos experimentales, cambiaron el paso a la mitad y repitieron el proceso hasta que consideraron que los pasos eran lo suficientemente pequeños. Este simple procedimiento era lento pero seguro y fue usado varias veces en la computadora

Avidac, en Argonne National Laboratory, para ajustar seis parámetros teóricos con los datos de dispersión pion-protón que fueron recolectados usando el synchro-cyclotron de la Universidad de Chicago (Fermi y Metropolis, 1952).

Aunque este algoritmo básico, sin duda, es anterior a Fermi y Metropolis, se ha mantenido en un estándar entre la comunidad informática científica debido exactamente a la razón observada por Davidon: *es lento pero seguro* (Kolda, Lewis, y Torczon, 2003).

Desde entonces, la investigación y desarrollo sobre los métodos de optimización son evidentes actualmente por la abundante literatura al respecto generada por la comunidad que trabaja en este tema, y constantemente se intenta trabajar sobre la lentitud de los algoritmos computacionales que los resuelven, pero sin perder seguridad. En forma muy general, los métodos de optimización pueden clasificarse como métodos analíticos y no-analíticos. Los métodos analíticos se resuelven haciendo cálculos derivados de la función objetivo, como la ecuación 3.1. No siempre es posible resolver estos problemas, por ejemplo, hallando las raíces de las derivadas de las funciones que se pretende maximizar o minimizar, algo que presupone la existencia de la derivada de la función objetivo. Los problemas del mundo real, como sucede en la ciencia y la ingeniería, no se resuelven tan fácilmente pues por lo general no se caracterizan por la existencia de la derivada, o tienen una expresión compleja que no permite encontrar sus raíces analíticamente. Si existe una solución analítica entonces puede suceder que el orden de la ecuaciones sea alto o la geometría del problema sea muy compleja. Pueden ser resueltos por algún algoritmo, pero no a tiempos polinomiales de ejecución, y sería impensable una solución que no incluya un método computacional de resolución. Desde este punto de vista los métodos se pueden clasificar en analíticos o numéricos, aunque también se los clasifica según el tipo de funciones como lineales o no lineales, y según el tipo de variables como enteros/discretos o reales/continuos (Nocedal y Wright, 2006) (Fletcher, 2013).

Como vimos anteriormente, el problema deber ser caracterizado mediante la función simple o multiobjetivo y las restricciones, si existen, y una vez logrados estos pasos se debe decidir el método de solución que dependerá de la caracterización. Este método puede ser estático o dinámico, con restricciones o no, lineales o no-lineales y de una dimensión o multidimensionales.

Es importante destacar en este momento que, sin perder generalidad, la versión estática de los métodos es la forma simplificada de los dinámicos, los problemas lineales son la forma simplificada de los problemas no-lineales y los unidimensionales son la for-

ma simplificada de los multidimensionales (Cabrera, Luque, Taboada, Epelde, y Iglesias, 2012).

En la Fig. 3.3 se presenta una clasificación que considera los métodos no-analíticos. Los métodos clásicos son los métodos de búsqueda como la búsqueda exhaustiva, la cual es una de las técnicas de resolución de problemas que garantiza llegar a la solución óptima luego de testear todas y cada una de las soluciones del universo de búsqueda. El inconveniente es la lentitud del método, el cual sólo puede llevarse a la práctica con problemas pequeños o de tiempo computacional de búsqueda de orden polinomial. Si el problema no se mantiene en esos órdenes de complejidad, entonces el costo computacional se hace extremadamente alto ya que se producen explosiones combinatorias del espacio de búsqueda. En estos casos los métodos posibles de utilizar se los llama heurísticas, las cuales son técnicas de resolución de problemas que tratan de orientarse en el universo de todas las posibles soluciones en busca de la posible mejor y así bajar el tiempo computacional de búsqueda. Un inconveniente que tienen la mayoría de los problemas combinatoriales, en general, es no poder asegurar que se encontró la mejor solución (Blum y Roli, 2003).

Por otro lado, los algoritmos de optimización se clasifican según el modo en que operan, como algoritmos determinísticos y probabilísticos. Utilizar algoritmos determinísticos es posible si el espacio de búsqueda es pequeño, sino el problema se convierte en intratable computacionalmente. En estos casos los algoritmos probabilísticos son la alternativa, no proveyendo un óptimo global, sino una buena solución aproximada (Nocedal y Wright, 2006).

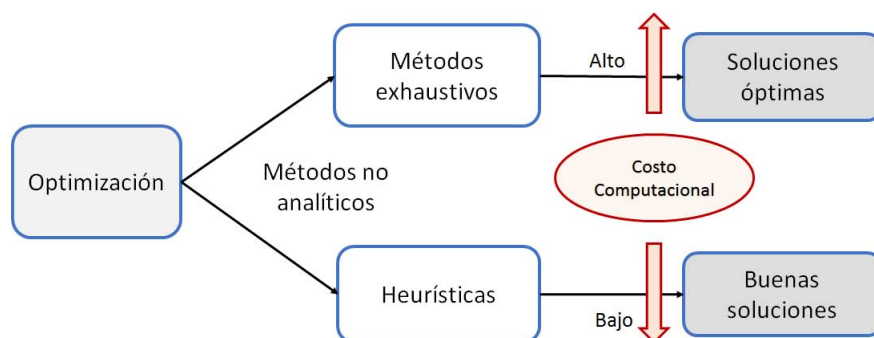


Figura 3.3: Clasificación de métodos no-analíticos de optimización.

3.2. Optimización vía Simulación

Optimización vía Simulación (OvS) es un campo emergente que integra técnicas de optimización y análisis de la simulación, el cual se ha convertido en una poderosa técnica para el análisis y optimización de sistemas complejos. El estudio de sistemas cada vez más complejos han requerido vincular esta dos áreas de investigación sin que ninguna domine, sino que se complementen aprovechando las posibilidades de la simulación de analizar los sistemas estudiados y de la optimización para encontrar soluciones que mejoren el rendimiento de los mismos. El principal objetivo de la OvS es determinar el mejor conjunto candidato de parámetros de entrada al sistema que resulten en obtener su mejor rendimiento (Fu, 2002) (Andradóttir, 2006).

Un beneficio de los experimentos de simulación sobre un modelo de simulación de un sistema, es que el usuario puede seleccionar diferentes configuraciones del sistema para mejorar el comportamiento del simulador. Esta mejora se refiere a calibrar el modelo al ir ajustando los resultados con la realidad. Cuando estas configuraciones se basan en la selección de un conjunto de valores para los parámetros de entrada del sistema, entonces el usuario podría intentar con distintos conjuntos hasta encontrar el mejor y así optimizar el funcionamiento del simulador, aunque esta búsqueda no es para nada trivial.

En principio hay una separación entre el método y el modelo. El problema de optimización está definido fuera del sistema, el cual está representado por un modelo de simulación. El modelo puede tener modificaciones sin afectar el proceso, ya que el modelo de optimización sigue siendo el mismo. En síntesis, hay una completa separación entre el modelo que representa el sistema y el procedimiento que es usado para resolver los problemas de optimización definidos con este modelo. Se produce una coordinación entre ambos procesos, como se ve en la Fig. 3.4, al recibir el método de optimización la salida del modelo de simulación que fue alimentada en éste. Este proceso continúa hasta que se cumple la condición de fin preestablecida (Glover, Kelly, y Laguna, 1996).

La OvS repite sucesivamente el análisis de la salida de un modelo de simulación con diferentes conjuntos de los parámetros hasta poder identificar el que produce la mejor simulación del sistema. Pero es importante distinguir que no estamos ante la optimización clásica determinística o la típica programación estocástica ya que no se tiene mucho conocimiento de la estructura de la función objetivo y su expresión analítica no se obtiene o quizás no existe. Debido a esto la función objetivo debe ser estimada basándose en las salidas de las corridas de la simulación. La OvS es vista como un enfoque estructurado

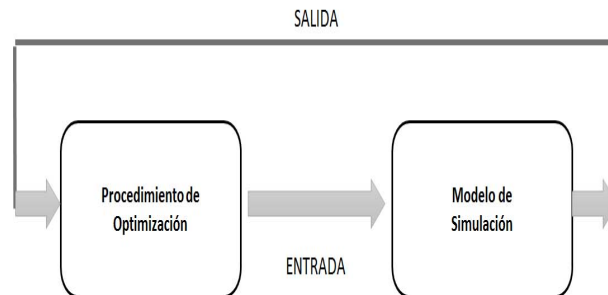


Figura 3.4: Coordinación entre simulación y optimización.

y sistemático para determinar los parámetros de entrada óptimos para lograr medir el mejor rendimiento del sistema considerado. De acuerdo con las variables de decisión, los problemas en OvS se los clasifica en los de variables continuas y los de variables discretas, los cuales se pueden subdividir según el método de solución elegido (Barton y Meckesheimer, 2006).

OvS es un campo de constantes avances en los últimos años, aunque se encuentran muchas técnicas en la literatura, aun no hay suficientes trabajos que brinden una comparación entre ellos. También existe una brecha entre el software comercial y la teoría que sustenta el tema, ya que el software se basa en metaheurísticas y no se encuentran otros basados en el métodos como, por ejemplo, de búsqueda del gradiente o técnicas de inferencia estadística. Por otro lado, los software comerciales requieren mejorar su eficiencia pues es muy complejo utilizarlos con grandes problemas de la vida real debido al alto consumo de tiempo de ejecución que los caracteriza (Wang y Shi, 2013).

3.2.1. Optimización vía Simulación en Sistemas Hidrológicos

En la bibliografía que relaciona los métodos de optimización de múltiples objetivos y los modelos hidrológicos los autores destacan que los planteos de modelos para optimizar la simulación se caracterizan por ser multiobjetivo. A continuación algunos de los más relevantes:

- En (Yapo y cols., 1998) ya se presentaba la necesidad de contar con sistemas de calibración multiobjetivo para modelos hidrológicos que simulan agua, energía, constituyentes químicos del agua, etc., para lo cual se presenta una primera etapa de un algoritmo (MOCOM-UA) para tratar un problema de optimización global múltiple objetivo, mediante un método de búsqueda random global.

- Dentro de los ejemplos de sistemas físicos, uno de ellos son la caracterización de reservorios de petróleo que son afectados por procesos sísmicos, los cuales deben tratar con la incertidumbre de los datos de entrada provenientes de datos geofísicos del sistema. Los autores de (Park y cols., 2015) presentan un método de optimización multiobjetivo basado en encontrar la frontera de Pareto, mediante algoritmos genéticos, para encontrar el conjunto de parámetros que mejore la simulación del sistema.
- Los autores de (Capocchi y Santucci, 2013) presentan un desarrollo en Python que utiliza DEVSimPy, donde integran un algoritmo de optimización con una simulación de eventos discretos, aplicando la optimización vía simulación a cauces que involucran diques, estaciones hidroeléctricas, etc.
- Los autores anteriores presentan en (Santucci y Capocchi, 2015) la integración de técnicas de optimización y simulación que llevan adelante mediante un proceso iterativo para manejar la evolución del sistema, basado en decisiones tomadas sobre los resultados previos de la simulación. En este trabajo se implementa un algoritmo de optimización vía simulación continua, usando una simulación de eventos discretos, con el objetivo de optimizar la gestión de una cuenca hídrica.
- En (Zare, Saghafian, Shamsai, y Nazif, 2012) se optimiza un modelo computacional de varios algoritmos evolucionarios multiobjetivo para obtener la solución óptima al modelado del agua por inundación en zonas urbanas. El método utiliza como funciones objetivos la calidad de la escorrentía, la cantidad de la escorrentía y los gastos (costo de mantenimiento e implementación de prácticas de control urbano del agua) y utiliza varios indicadores estadísticos para comparar el rendimiento de los algoritmos.

No se encuentran en la literatura trabajos que se ajusten al presentado en esta tesis, sobre el uso de métodos de OvS para mejorar la predicción de un sistema hidrodinámico de cauces de río, especialmente que se basen en la metodología de este trabajo, la cual se describe en lo que resta de este capítulo. Pero es muy importante destacar que este trabajo toma como punto de partida, y a su vez realimenta, los trabajos desarrollados y actualmente en curso sobre el tema de esta tesis en el grupo *High Performance Computing for Efficient Applications & Simulation de la Universidad Autónoma de Barcelona*² ,

²<http://grupsderecerca.uab.cat/hpc4eas/>

destacando los siguientes: (Cabrera, Taboada, Iglesias, Epelde, y Luque, 2012) (Taboada, Cabrera, Epelde, Iglesias, y Luque, 2013) (Bianchini, Denham, Cortés, Margalef, y Luque, 2006) (Abdalhaq, Cortés, Margalef, Bianchini, y Luque, 2006).

3.3. Metodología Propuesta para Mejorar la Predicción del Modelo de Cauce de Ríos, Aplicado a Ezeiza

En esta sección se presenta la propuesta de la metodología utilizada para mejorar la predicción del modelo de simulación Ezeiza, en conjunto con el modelo matemático hidrodinámico del río Paraná. Mediante los aportes de la OvS, se implementa un metamodelo de simulación para encontrar el conjunto de parámetros de ajuste que optimiza la salida del simulador.

Justamente, Ezeiza, por ser un simulador, permitiría experimentar con distintas configuraciones (o escenarios) y así encontrar, sin importar cuántas veces se intente, un conjunto ajustado de parámetros. Se está haciendo referencia a testear cada candidato del espacio de búsqueda hasta encontrar la mejor solución, método que garantizaría encontrar la solución óptima. Pero, también se habló en 3.1.4 sobre la complicación de intentar llevar adelante esta búsqueda exhaustiva por la explosión combinatoria que convierte el método en intratable computacionalmente. Encontrar la solución óptima de esta manera implica encontrar un vector de parámetros $x^* \in S$ que al ser ingresados al simulador junto con los datos completos de un escenario de simulación ajuste de la mejor manera posible los datos simulados y los observados. El espacio de búsqueda S del problema de optimización es enorme, por lo cual encontrar el vector x^* mediante un método computacional es imposible debido a la complejidad del algoritmo como se verá en 3.4.2. En contraste, la implementación de métodos heurísticos puede ayudar a la búsqueda, pero proveen una solución considerada “buena” y con costo computacional menor.

El método que se presenta se sitúa entre los métodos de búsquedas exhaustivas y las heurísticas propias de la optimización. El espíritu del método se grafica en 3.5. En síntesis, la idea principal del método es la siguiente: En una primera etapa, mediante un proceso iterativo, se van obteniendo subregiones del espacio de búsqueda del problema que cumplan con su función objetivo. Esta etapa se basa en la estructura de vecindad del problema para reducir la región de posibles soluciones donde se hace la búsqueda, mediante un esquema Monte Carlo combinado con un método K-Means. Se obtienen conjuntos de *regiones prometedoras*, nombradas así desde el punto de vista de la posibilidad de contener

la solución óptima, o alguna subóptima, cuyo tamaño es mucho más chico. Al finalizar, se implementa una segunda etapa de búsqueda exhaustiva, pero ahora sobre la región prometedora resultante, ya mucho más pequeña.

A pesar de la reducción del espacio de búsqueda, ambas etapas implementan métodos computacionales que son posibles de llevar a la práctica gracias a los recursos de programación paralela y HPC, que permiten lograr resultados en tiempos aceptables. Se trata este tema en 3.6.

En lo que resta de este capítulo se brindan los pormenores de esta propuesta, la cual constituye una metodología cuyo esquema de procedimientos podría ser utilizado con otros simuladores de sistemas físicos para brindarles un método automático de sintonización del simulador, mediante una calibración de los parámetros y la optimización de su predicción.

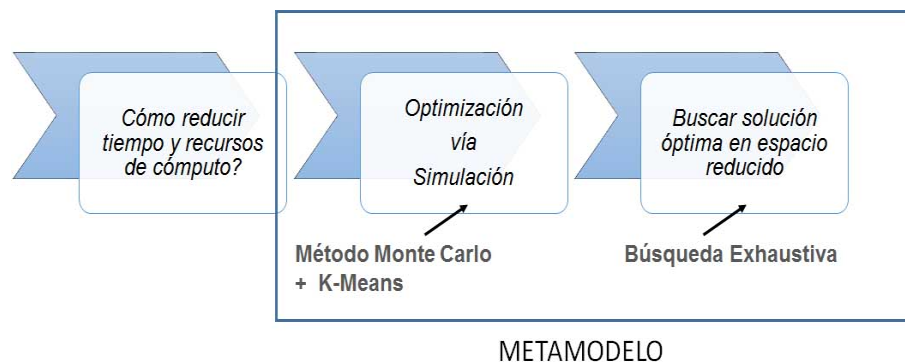


Figura 3.5: Metodología propuesta.

3.4. Un Metamodelo de Optimización vía Simulación

El método de OvS aplicado a mejorar los resultados del simulador Ezeiza, es un metamodelo que utiliza una estrategia de OvS en dos etapas. Un metamodelo es un modelo empírico creado con datos generados por modelos computacionales, o sea son modelos de modelos. El uso de un metamodelo pretende facilitar la tarea de optimización, que implica encontrar la mejor configuración de parámetros de entrada, o variables controlables del sistema, con el fin de optimizar la función objetivo, siendo ésta función de las salidas del modelo de simulación, en este caso el modelo Ezeiza. El método utiliza la salida del simulador, y otros datos propios del sistema físico simulado, para evaluar su función objetivo

y en base a esta medida o índice de rendimiento, propone un nuevo conjunto de valores de los parámetros.

Antes de ir directamente a la descripción del método, es necesario mostrar las características del problema que se pretende optimizar, como ser cuál es la función objetivo y cómo se constituye el espacio de búsqueda en el entorno del problema.

3.4.1. Función Objetivo

En este punto es importante recordar que el modelo de Ezeiza es considerado como una caja negra en el proceso de simulación, independientemente de como Ezeiza lleva adelante este proceso. Ezeiza es alimentado con cada escenario de simulación y el modelo devuelve los datos simulados (de salida) para el tiempo correspondiente a la simulación. O sea, se ingresa al simulador un vector x de parámetros de entrada y se obtiene un índice de rendimiento calculado mediante la función objetivo $f(x)$. Éste es un proceso determinístico ya que con distintos vectores de parámetros x puedo, o no, tener un mismo índice, pero cada vez que se alimenta el simulador con el mismo conjunto de parámetros y la misma configuración de entrada, entonces se obtiene siempre el mismo índice.

En el marco de los objetivos del trabajo, es necesario establecer índices que valoren los resultados obtenidos con el simulador para cada conjunto de parámetros de entrada, ya que estos parámetros son los valores de ajuste de la simulación. Estos índices se establecen en función de las diferencias entre los valores simulados y los observados, para cada escenario, dando una medida de ese ajuste. Este trabajo busca minimizar estas diferencias mediante el objetivo de minimizar los índices, o sea, cuando menores son sus valores se logra mejor ajuste y mayor calidad del simulador.

La evaluación de la función objetivo f involucra la ejecución del simulador y la estimación del índice de calidad, según el esquema de la Fig. 3.6. Los datos observados en el instante inicial t_0 son los que proveen las condiciones iniciales, los cuales junto con el vector de parámetros x definen el escenario de simulación. La función objetivo está representada por el proceso que calcula el índice de calidad I_c , recordando que en este proceso está involucrada la ejecución de Ezeiza y la selección de datos para análisis. Esta función es una rigurosa expresión matemática y permite comparaciones cuantitativas entre las distintas soluciones.

El problema que se deduce de este planteo, y que es tratado en profundidad para conseguir los resultados buscados, es la gran cantidad de posibles soluciones a evaluar. Esto se debe a la gran cantidad de escenarios de simulación que surgen de la combinación

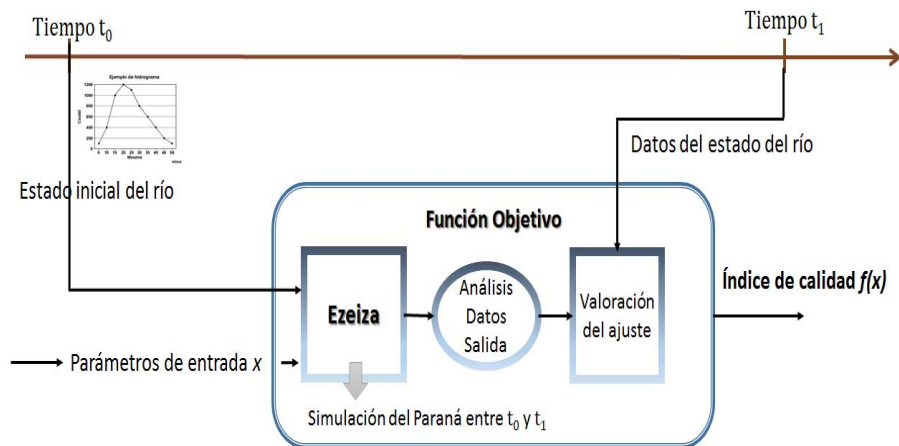


Figura 3.6: Función Objetivo.

de todos los parámetros involucrados en la simulación de cauce de ríos, como sucede con el sistema físico del río Paraná y sus afluentes utilizado en esta tesis.

Por otro lado, los índices que se tomaron en cuenta utilizan el error cuadrático medio promedio del tramo del Paraná a optimizar y el rendimiento logrado en comparación con el escenario usado actualmente por el INA, tema que se desarrollará en profundidad en el capítulo 4.

3.4.2. El Espacio de Búsqueda

En un problema de optimización el espacio de búsqueda es el conjunto de todos los elementos que pueden ser considerados en el proceso de búsqueda de la solución.

El problema de la simulación del cauce de ríos, como muchos sistemas físicos, está caracterizado por una gran cantidad de parámetros de entrada que provienen de la geometría del sistema, o sea los parámetros que resultan de la abstracción hidrológica del cauce como se explicó en el capítulo 2. Algunos de ellos son estables y no presentan variaciones significativas a lo largo del tiempo o al ocurrir algún evento de aumento o disminución significativa del caudal. Sin embargo, el coeficiente de rugosidad, o índice de Manning, como también la altura de los albardones se consideran los parámetros más sensibles del modelo, debido a su impacto en los resultados cuando sus valores se modifican levemente. Estos parámetros toman valores reales dentro de un rango, pudiendo tomar cualquier valor real en ese intervalo y por lo tanto, tendríamos infinitas soluciones posibles. Pero, para un programa computacional todo debe ser finito o discreto. Como será necesario

implementar un método computacional que logre encontrar la mejor solución posible, se deberá tener en cuenta la precisión de cómputo disponible para encontrar la manera más conveniente de discretizar los rangos de valores reales que constituyen los dominios de los parámetros.

Para ser preciso, se debe contar con los valores del parámetro Manning y altura de albardón en cada una de las 76 secciones en que se divide el dominio del río Paraná, considerando un valor de Manning para cauce y otro para planicie en cada sección. Simplificando el problema, si se tomaran en cuenta sólo 5 secciones ubicadas en el tramo inferior del río y se mantienen constantes los parámetros de las 71 secciones restantes, se estaría hablando de 15 parámetros. La cantidad estimada de combinaciones dependería de la cardinalidad de cada parámetro, que supondremos de 5 valores posibles para cada uno, por lo tanto el tamaño del espacio de búsqueda estaría dado por:

$$\left(\prod C_i\right)^5 \text{ tal que } i : 1.,15 \quad (3.6)$$

donde 15 es la dimensión del vector x y C es la cardinalidad de cada uno de sus elementos.

Bajo estas condiciones se tienen 2373×10^6 diferentes combinaciones, o sea 2×10^9 posibilidades a explorar. Con esta configuración, no es posible considerar el espacio resultante para llevar adelante la búsqueda de x^* ya que cada combinación requiere la ejecución del simulador con todo su escenario de simulación, resultando tiempos de ejecución impracticables. Por lo cual, es necesario implementar una técnica algorítmica que haga posible encontrar un conjunto de valores lo más cercanos al óptimo, pero en un espacio de búsqueda acotado.

3.5. El Método de OvS en Dos Etapas

La Fig. 3.7 grafica el método completo de OvS en dos etapas, la cual es diseñada e implementada en esta investigación, para un simulador del estilo de Ezeiza. En esta sección se describen cada uno de sus pasos, previa presentación de las características de los métodos utilizados.

3.5.1. El Método de Monte Carlo

La simulación de Monte Carlo es una técnica estocástica usada para resolver problemas matemáticos. Los métodos de Monte Carlo (MMC) generan valores aleatorios para

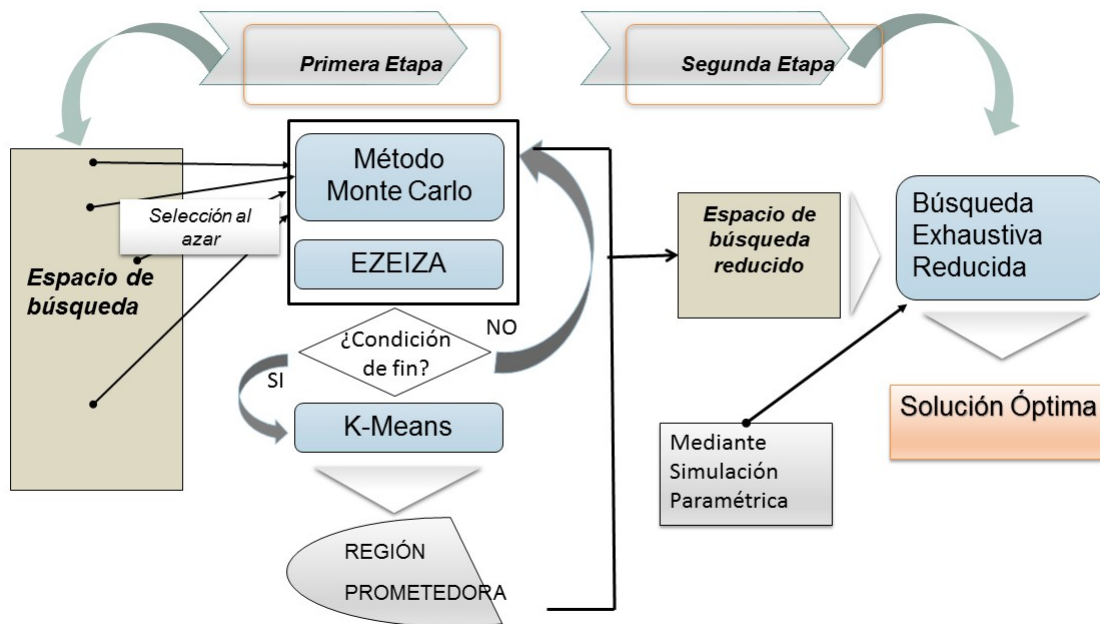


Figura 3.7: Método de Optimización vía Simulación de Dos Etapas.

crear escenarios diferentes del problema en estudio. Los MMC comprenden una variedad de técnicas que brindan la posibilidad de soluciones de problemas de índole físico o matemático, mediante pruebas aleatorias sucesivas. Dicho de otra manera estas técnicas permiten encontrar la solución de un problema numérico empleando experimentos de muestreo artificial, mediante la generación de números con una distribución de probabilidad determinada que simula los resultados de una variable física. Estas técnicas combinan conceptos estadísticos con la virtud que tienen los computadores para realizar cálculos de manera rápida. La técnica de simulación del MMC, en principio, genera valores aleatorios de parámetros de entrada estocásticos de acuerdo con sus respectivas características probabilísticas. Se genera un gran número de conjuntos de parámetros, para calcular la salida correspondiente del modelo y de esta manera poder determinar algunas características estadísticas de la salida del modelo. Es un método utilizado cuando no es posible llegar a una solución exacta usando algún algoritmo determinístico, cuando se tienen muchos grados de libertad del problema o cuando hay incertidumbre en los datos de entrada al sistema (Mackay, 1998) (Kroese, Taimre, y Botev, 2013).

3.5.2. El Método K-Means

Este método proporciona una técnica para agrupar un conjunto de datos en grupos o subconjuntos. O sea, esta técnica identifica k grupos distintos de un conjunto de n datos, siendo estos datos puntos de un espacio d -dimensional. Un algoritmo de agrupamiento K-means produce una partición de los datos o puntos $x \in A$, en k subconjunto disjuntos, $C \subseteq A$. El método trata de minimizar la distancia de los puntos de cada grupo C_i hacia el centro de dicho grupo. Cada iteración reacomoda los puntos que se distribuyen entre los grupos. Cada grupo es parametrizado por el valor de su media. La cantidad de grupos k es un dato del método y la cantidad de iteraciones dependerá de este valor. La complejidad del método depende de la cantidad de iteraciones i , de k , de la cantidad de puntos n y de la dimensión del espacio d (MacKay, 2003). En (Arthur, Manthey, y Roglin, 2009), los autores demuestran que el orden de complejidad es polinomial en k , i , d y σ , siendo este último parámetro la desviación estándar de la perturbación Gaussiana. El artículo mencionado brinda una detallada demostración al respecto.

3.5.3. Descripción de la Metodología

- a) La primera etapa, **método Monte Carlo + K-means**, devuelve un espacio de búsqueda acotado que contiene los datos de los escenarios de simulación con mejores valores promedio del índice de calidad, resultantes de medir la calidad de cada escenario. La técnica implementada es un MMC basado en el algoritmo Metrópolis (Kroese y cols., 2013) combinado con un método K-means.

El MMC selecciona, por vez, 100 escenarios al azar del espacio total de posibles escenarios de simulación que pueden ejecutarse con Ezeiza, según la zona del río Paraná (dominio del problema) seleccionada para la experimentación. El MMC delimita la región en la que se encuentra el mínimo. En cada paso, toma uno de los 100 escenarios, ejecuta la simulación y calcula su índice de calidad, para obtener finalmente un promedio de los índices. El método va guardando las configuraciones que mejoran la media ya guardada, y finaliza cuando dos iteraciones sucesivas ya no mejoran el valor medio promedio almacenado.

Una vez alcanzada la región delimitada y prometedora para encontrar el mínimo (solución óptima o subóptima) se busca agrupar los puntos obtenidos mediante el algoritmo K-Means. Esta es una estrategia "inteligente" que vuelve a reducir el espacio de soluciones prometedoras. De esta manera, de los puntos resultantes se obtienen los límites

o restricciones que se impondrán sobre los rangos de valores que pueden tomar los parámetros (espacio de búsqueda reducido). Sobre esta zona delimitada, reducida, se puede implementar la segunda etapa.

- b) La segunda etapa, **método de búsqueda exhaustiva reducida**, utiliza como espacio de búsqueda la región reducida de la primera etapa, y garantiza encontrar la solución óptima de la región de factibilidad reducida.

Este método optimización tiene un costo computacional alto, pero en este trabajo se utiliza porque es necesario garantizar la mejor solución posible. Es una solución que pudo ser llevada a la práctica debido a la implementación de la primera etapa y al uso de recursos computacionales como los que proporciona HPC.

3.6. Implementación en Paralelo del Método de OvS

El método de OvS ya detallado anteriormente en sus dos etapas es un método fácilmente paralelizable. El algoritmo principal debe evaluar la función objetivo con cada candidato solución antes de ingresarlo al método de optimización, el cual hace un análisis de los índices obtenidos para decidir si el método debe repetirse. El simulador se ejecuta con cada uno de los 100 elementos seleccionados al azar del espacio de búsqueda para el MMC. Cada evaluación es independiente de las otras, por lo tanto estamos ante un problema altamente paralelizable en un cluster mediante un modelo de programación Master-Worker. En la Fig. 3.8 se presenta un esquema de los componentes que intervienen en el algoritmo que se ejecuta en cada worker, trabajando en modo colaborativo con los demás. Se grafica el método iterativo correspondiente al MMC que ejecutado en cada worker, el cual recibe el conjunto de escenarios seleccionados al azar y enviados por el master. Al finalizar cada procesamiento, el worker se queda con la media de los índices. Si recibe una nueva carga repite el proceso mostrado en el gráfico. El master analiza los datos que va recibiendo de los workers y decide si sigue repartiendo carga o alcanzó el fin del procesamiento y obtiene un conjunto de soluciones finales. Este conjunto de soluciones serán alimentadas al módulo de K-Means para obtener un agrupamiento de regiones prometedoras.

La aplicación Master-Worker de la segunda etapa fue escrita en C++ con MPI. Se utilizó pthreads para ejecutar el conjunto de programas que conforman el simulador Ezeiza. El análisis de las series temporales que devuelve Ezeiza para el cálculo del índice, depen-

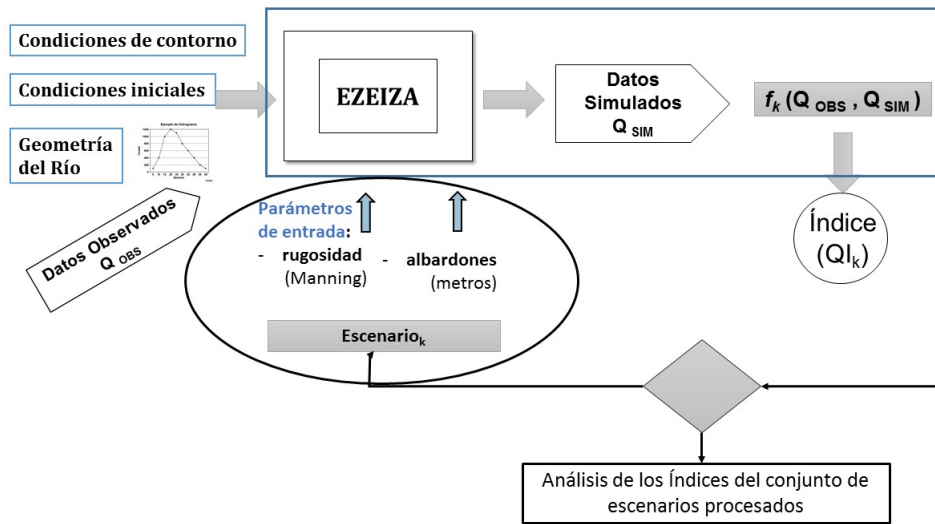


Figura 3.8: Esquema del algoritmo de la primera etapa del método de OvS.

diente de los datos simulados y observados como se muestra en la Fig. 3.8, se implementó en Java y es ejecutado en cada worker.

La etapa final de búsqueda exhaustiva es implementada, independientemente de la etapa anterior, en un modo Master-Worker mediante un programa en C++ con MPI, con reparto de carga de modo dinámico para manejar el balance de carga más pthreads para correr los programas de Ezeiza. El balance de cargas se controla en ambas etapas del método de optimización.

3.7. Discusión

1. La simulación del cauce del río Paraná requiere una gran cantidad de datos de entrada y de parámetros, los cuales también son fuente de incertidumbre ya que provienen de caracterizar un sistema físico y la dinámica de evolución del flujo de agua en el cauce. Pero, la mayoría de ellos fueron seriamente medidos y no sufren grandes modificaciones a corto plazo, o el impacto en la salida no es tan significativa. Los parámetros elegidos para calibrar y sintonizar la simulación son los que cambian cuando se producen eventos extraordinarios de aumento o disminución del caudal y son altamente sensibles en los resultados. Son los parámetros que proveen el mayor

grado de incertidumbre al modelo. Estos temas fueron largamente discutidos con los ingenieros del INA.

2. El método garantiza encontrar una buena solución. Varios factores de implementación no pueden dejarse de lado para lograr este objetivo, como son los parámetros que caracterizan la ejecución de los algoritmos. Algunos de los puntos más relevantes son la cantidad de escenarios que selecciona por vez MC, la cantidad de clusters con la que se inicia K-Means o la precisión límite para comparar los resultados. En cada problema a resolver será necesario un estudio previo del método y un ajuste sobre la marcha de estos valores, quizás de una manera artesanal. Por otro lado, una gran cantidad de iteraciones de MC son necesarias para delimitar el valor mínimo de manera confiable. Todos estos factores aumentan el tiempo de cómputo, pero la implementación del método completo en un ambiente de HPC posibilita llevarlo a la práctica en tiempos razonables de cómputo.
3. La función objetivo del problema de optimización es la función que se utiliza para valorar cada escenario de simulación. Como se dijo, los índices que se tomaron en cuenta son el error cuadrático medio promedio del tramo del Paraná a optimizar y el rendimiento logrado en comparación con el escenario usado actualmente por el INA. Pero, al buscar la solución fue necesario imponer el objetivo de no desmejorar la predicción en los tramos que no se utilizaban para la optimización. La búsqueda de una solución que cumpla la minimización de estos índices se hizo buscando las soluciones en la frontera de Pareto. Cabe destacar que el cálculo del índice podría hacerse con otros indicadores apropiados a la simulación. O sea se estaría modificando el cálculo de la función objetivo, algo que es fácilmente modificable en las aplicaciones del método, pero el método se aplicaría de igual manera.
4. Para poner en práctica esta metodología fue necesario hacer una búsqueda exploratoria reducida al inicio del trabajo de investigación, para entender más sobre las salidas obtenidas con la simulación del río Paraná y el simulador Ezeiza y tener una base para esta propuesta. Esta primera experimentación fue posible gracias a la implementación de una aplicación Master-Worker creada especialmente, y desarrollada en C++ y MPI. Estos resultados fueron luego comparados con los obtenidos en la etapa de optimización. El método de dos etapas de OvS fue también implementada considerando balance de cargas en modo Master-Worker para un cluster de computadoras.

5. Se utilizaron y combinaron métodos muy conocidos en la literatura y aplicables a muchos problemas de optimización. Pero la idea principal de crear un metamodelo de OvS para sintonizar un modelo hidrodinámico es aportar una técnica automática de calibración de parámetros y sintonización a esta clase de simulación de un fenómeno físico, y brindar también un aporte a las ya muchas aplicaciones del MMC y K-Means.

Capítulo 4

Aplicación de la Metodología de OvS propuesta, al Problema de la Predicción de la altura del Cauce del río Paraná

El método de OvS propuesto para calibrar y sintonizar el simulador de cauces de ríos Ezeiza, presentado en el capítulo 3, es puesto en evaluación mediante dos escenarios completos de simulación con datos reales. Estos casos de evaluación se basan en experimentaciones sobre el modelo hidrodinámico del río Paraná, obteniendo en cada uno de ellos los conjuntos de parámetros ajustados que mejoran la predicción de Ezeiza. Este aporte contribuye con una metodología automática de calibración y sintonización de Ezeiza, programa integrante de las herramientas computacionales que brindan las alertas diarias sobre los niveles alcanzados por el Paraná en la región. Para poner en práctica estos casos de prueba es necesario establecer condiciones sobre la ejecución de la simulación, debiendo considerarse: el modelo de simulación a utilizar, los datos de entrada que condicionan la simulación, las variables de decisión que determinan la geometría del río Paraná y las métricas utilizadas para evaluar la salida del proceso de simulación con Ezeiza. El modelo de simulación a utilizar es Ezeiza, el cual ya fue ampliamente discutido en el capítulo 2, como también el modelo del sistema físico del río Paraná. En las siguientes secciones se describen las condiciones de ejecución de cada simulación 4.1, las variables de ajuste o de decisión 4.2 y las métricas establecidas para medir la calidad de la predicción 4.3. Posteriormente, se presentan en detalle los casos de prueba con escenarios reales en 4.5, en 4.8 se presenta la ganancia lograda y finalmente se discuten los resultados logrados en 4.9.

4.1. Las Condiciones de la Simulación

El simulador Ezeiza fue presentado en 2.4. Sus características y las del modelo del Paraná fueron ampliamente conversadas con los ingenieros especialistas en hidráulica del INA, quienes utilizan diariamente este programa. De esta manera, fue posible comprender las variables hidrológicas y las propias de la geometría del cauce de un río, involucradas en la simulación de la escorrentía del río Paraná y sus afluentes.

Ezeiza es un modelo numérico que constituye una herramienta asentada y confiable para el diseño ingenieril. Permite simular sin dificultad ondas de inundación. El problema no reside en el modelo de simulación sino en la incertidumbre de los parámetros del modelo físico que simula. Es muy importante para implementar la sintonización de Ezeiza, poder ingresar los parámetros de entrada como un dato más de inicio de la simulación (en archivos de datos, por ejemplo). Se recuerda en esta sección cuáles son los datos a ingresar al programa Ezeiza en sus archivos de inicio:

- Topología de la red de flujo.
- Secciones transversales de información.
- Características de resistencia al escurrimiento de los canales, representada a través del coeficiente de Manning.
- Condiciones de bordes.
- Aportes laterales.
- Condiciones de arranque del modelo (eventualmente, éstas pueden ser generadas automáticamente).
- Parámetros numéricos (pasos espacial y temporal).

Y recordamos también las salidas de Ezeiza, en diversos archivos con resultados:

- Perfil de la onda sobre los canales del sistema.
- Hidrogramas de caudal y nivel en secciones previamente especificadas.
- Mapas de inundación.

La mayoría de sus parámetros pueden evaluarse a partir de observaciones hidrometeorológicas o mapas de la topografía. Ya se expresó en el capítulo anterior el inconveniente de no poder contar con datos de campo completos para poder especificar los parámetros del modelo, o no poder evaluarlos directamente, por lo cual necesitan siempre calibración. Otros, como propiedades hidrológicas, o algunas propiedades espaciales, pueden ser calculadas usando otras herramientas o no necesitan ajuste por calibración.

De las entradas que se detallaron anteriormente, la topología de la red de flujo (o red de cálculo en la Fig. 2.7) es la entrada de interés de este trabajo. Entre los valores que definen la topología están todos los parámetros de entrada, y especialmente los que se van a modificar para crear los escenarios de simulación. Se llama escenario a una configuración particular del conjunto de parámetros de entrada, o variables de decisión, concepto que se utilizará a lo largo de lo que resta del trabajo. Los parámetros de interés para este trabajo son el coeficiente de Manning de planicie, el coeficiente de Manning de cauce y la altura del albardón de cada sección (se toma sólo el de una de las orillas del cauce para cada sección), los cuales se pueden modificar directamente en el archivo de entrada con los datos de la geometría del río. Por otro lado, también es importante ingresar el paso temporal (días-horas) y el espacial (metros), siendo posible mantener valores por defecto. A continuación se brindan más detalles sobre la discretización del dominio y la ubicación de las secciones para ingreso de datos y las estaciones de monitoreo.

4.1.1. Configuración de los Datos del Dominio

La geometría del canal se provee a través de secciones transversales (distancia margen izquierda versus cota de fondo) arbitrariamente localizadas. El paso de cálculo espacial Δx es constante durante la simulación y los valores de los parámetros calculados en cada paso espacial se hace por interpolación, usando la información de las dos secciones más cercanas. Una vez fijado el paso espacial, el paso temporal Δt se elige de forma tal que la velocidad numérica $\Delta x/\Delta t$ sea del orden de la velocidad de propagación de la onda dominante. Se tomó un paso espacial de aproximadamente 5 km (éste se reajusta automáticamente en cada canal para determinar un paso uniforme) con lo cual se arribó a un total de 249 nodos para discretizar el dominio. El paso temporal que surge de la relación entre el paso espacial y la velocidad media de la corriente (estimada en 0,8 m/s) es de alrededor de 6000 segundos. Los cálculos se efectuaron, de todos modos, con un paso temporal sensiblemente menor; a saber, $\Delta t = 10$ minutos (600 segundos), para el cual se logra la convergencia numérica para la precisión requerida.

La estaciones de monitoreo son las 15 escalas hidrométricas ubicadas arbitrariamente a lo largo del tramo considerado sobre el río Paraná. Su ubicación se elige en función de los pronósticos que se decidan emitir en la zona. Para cada una de ellas está asociada al siguiente conjunto de datos, que se deben ingresar al simulador:

Cota Cero: El nivel desde donde se mide la altura del agua.

Progresiva: La distancia en Km. desde un cero relativo establecido al inicio del dominio, en la entrada del tramo alto del Paraná.

Río: Río al que pertenece, ya que pueden ser distintos canales.

Fuente: Organismo o institución de donde se obtienen los datos

Respecto al dato del río, cabe aclarar que el río Paraguay es un afluente del río Paraná y tiene 2 estaciones de monitoreo sobre su cauce. Sin embargo, actualmente el aporte del Paraguay al Paraná se lo toma como condición de borde. La fuente u organismos que proveen los datos son: Subsecretaría de Puertos y Vías Navegables (SSPyVN), Prefectura Naval Argentina (PNA), Cota Cero por Ministerio de Obras Públicas y Progresiva por Instituto Geográfico Militar. Estos datos se muestran en la Tabla 4.1.

Hidrómetro	Progresiva (Km)	Cota Cero (m)	Río	Fuente
Pilcomayo	0	52.36	Paraguay	SSPyVN
Formosa	169	47.16	Paraguay	SSPyVN
Ituzaingo	0	57.43	Paraná	SSPyVN
Ita Ibaté	74	51.87	Paraná	SSPyVN
Itatí	175	45.10	Paraná	SSPyVN
Corrientes	246	41.83	Paraná	SSPyVN
Empedrado	315	39.01	Paraná	SSPyVN
Bella Vista	397	33.99	Paraná	SSPyVN
Goya	482	29.12	Paraná	SSPyVN
Reconquista	506	27.13	Paraná	SSPyVN
Esquina	601	21.83	Paraná	SSPyVN
La Paz	697	15.79	Paraná	PNA y DNVN
Hernandarias	765	12.84	Paraná	SSPyVN
Santa Fe		8.95	Paraná	SSPyVN
Paraná	859	9.40	Paraná	SSPyVN
Diamante	873.69	7.319	Paraná	SSPyVN
San Martín	948.71	4.181	Paraná	SSPyVN
Rosario	983	3.599	Paraná	SSPyVN

Tabla 4.1: Datos de las estaciones de monitoreo.

Por otro lado, al hablar de ajuste se piensa en ajustar las diferencias entre los valores observados y los valores simulados. Por lo cual, todo este análisis y evaluación del método de OvS, no sería posible si no se contara con los datos observados durante los períodos de simulación que se consideraron. Esta información fue provista por el personal del

Laboratorio de Hidráulica del INA. Los datos observados se encuentran en una base de datos que cubre los años 1994-2011, conteniendo las alturas medidas del río Paraná en cada una de las 15 estaciones de monitoreo, día a día durante ese período.

4.1.2. Las Condiciones de Borde del Río

Entre los datos de inicio de la simulación se deben establecer las condiciones de borde del sistema durante el período simulado. El programa Ezeiza recibe un archivo de entrada donde se determina el tiempo de simulación, indicando el día de inicio y el día de fin del proceso. Al establecer este período, se deben generar los datos de las condiciones de borde, que son series temporales con alturas o caudales, en tres localizaciones del dominio preestablecidas. Para obtenerlos se utilizan otros programas del paquete de Ezeiza. Estos puntos están graficados en la Fig. 2.7. Los niveles del agua (H) o los caudales instantáneos (Q) en los contornos de ingreso al sistema, como en las desembocaduras, se especifican como condiciones de borde. El río Paraguay el cual aporta caudal al Paraná, es registrado en el proceso mediante una serie temporal H, siendo una de las condiciones de borde de la simulación. Los otros dos puntos considerados son Yacyretá, cuyo aporte diario Q es tomado como condición de borde aguas arriba, y Villa Constitución, cuyas aporte diario H es tomado aguas abajo.

4.2. Las Variables de Decisión del Modelo del Río Paraná

Las variables de decisión del problema son los parámetros de ajuste, los cuales fueron presentados en el ítem 2.4.2. Estos parámetros pueden modificarse a gusto, siempre dentro de las restricciones impuestas sobre el problema de optimización, y son los que permiten llevar adelante las simulaciones paramétricas del modelo. Recordamos que los parámetros de la simulación del cauce de ríos son *el coeficiente de Manning y la altura de los albardones*, los cuales son valores reales y están definidos por intervalos continuos con infinitos valores en cada intervalo. Estos dominios continuos asociados a cada parámetro deben ser discretizados, ya que no sería posible considerar sus infinitos valores en un programa computacional. Tratar estos intervalos continuos como discretos nos permite ver el problema de optimización como un problema discreto combinatorio.

En la Tabla 4.2 se muestra el excesivo crecimiento de la cantidad de escenarios en función de las secciones que se consideran, y tan solo tomando en cuenta dos parámetros: Manning de cauce, M_c , y Manning de planicie, M_p . La cantidad máxima de secciones es 76. Es evidente la explosión combinatoria y el enorme espacio de búsqueda que caracteriza el problema.

Cardinalidad		Número de Secciones Consideradas					
M_c	M_p	2	4	6	8	10	76
2	4	64	4096	262144	1.7 E+07	1.1 E+09	4.3 E+68
3	5	225	50625	1.1 E-07	2.6 E+09	5.8 E +11	2.4 E+89
4	5	400	160000	6.4 E+11	2.6 E+10	1.0 E+13	7.6 E+98
5	6	625	390625	2.4 E+08	1.5 E+11	9.5 E+13	1.8 E+106

Tabla 4.2: Explosión combinatoria de escenarios en función de las secciones consideradas.

A continuación se describe la implementación de la técnica de simulación paramétrica utilizada en dos etapas del trabajo. En un primer momento, al comenzar este trabajo, se implementó una búsqueda exhaustiva mediante esta técnica, para obtener información básica sobre el comportamiento del modelo al variar y permutar sus parámetros de entrada. El espacio de búsqueda debió ser acotado y los algoritmos de búsqueda fueron implementados mediante programación en paralelo para tener tiempos de ejecución razonables. Se brindan más detalles al comienzo de los casos de estudio en 4.5. Posteriormente, se vuelve a repetir la técnica al implementar la segunda etapa del método de OvS, presentado en el capítulo anterior. Se hace una búsqueda exhaustiva sobre el espacio acotado que devuelve la primera fase de MC y K-means para obtener la mejor solución posible al problema de optimización.

4.2.1. Simulación Paramétrica

Una simulación paramétrica consiste en cambiar los valores de un conjunto de parámetros de entrada para utilizar cada combinación resultante como entrada al simulador, lanzando tantas corridas como conjuntos de parámetros se tiene, y con un objetivo predefinido. Al referirnos al simulador Ezeiza ese objetivo sería medir el error de la salida del simulador mediante los hidrogramas de salida con las alturas del río en cada paso de tiempo (día) de la simulación. En la Fig. 4.1 se esquematiza el proceso de la simula-

ción paramétrica que irá ejecutando el simulador con cada escenario, hasta encontrar la solución de la búsqueda.

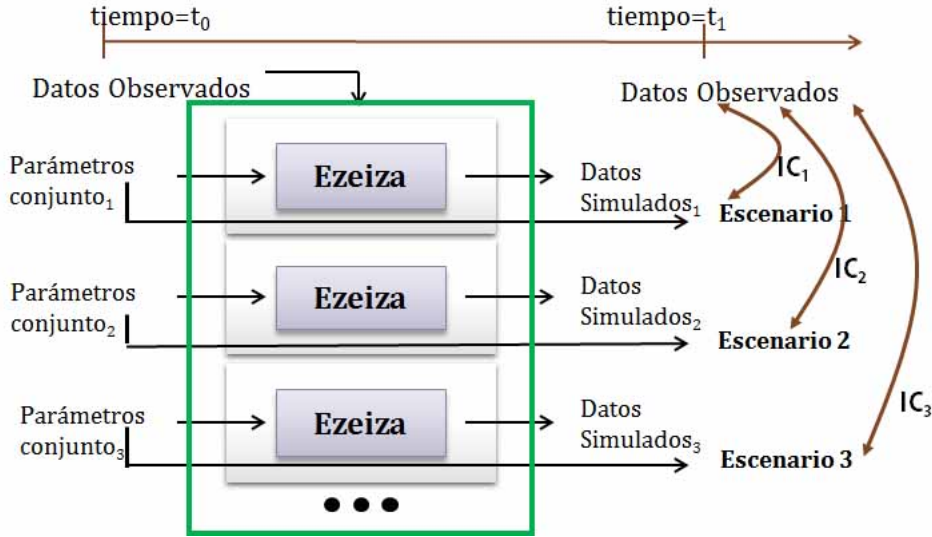


Figura 4.1: Esquema del proceso de simulación paramétrica.

Si se consideran N parámetros, la cantidad de posibles escenarios está determinada por la cardinalidad de cada uno, C_i . Cada parámetro tiene un intervalo de valores asociados y se establece para cada uno un paso de barrido sobre ese intervalo. Cada intervalo está determinado por un límite superior, un límite inferior y la medida del paso, como se muestra en la ecuación 4.1:

$$\langle [L_{inf}^i, L_{sup}^i], \Delta_i \rangle \quad (4.1)$$

donde el intervalo del parámetro i está acotado por L_{inf}^i y L_{sup}^i y es dividido en pasos de longitud Δ_i . La cantidad de escenarios posibles está dado por

$$\# Escenarios = \prod_{i=1}^N C_i \quad (4.2)$$

$$donde C_i = ((L_{sup}^i - L_{inf}^i + \Delta_i) / \Delta_i) \quad (4.3)$$

El dominio del río Paraná está dividido en 76 secciones. Según se describió en el capítulo 2.4.2, los parámetros considerados para el proceso de sintonización son el Mc, Mp y albardones. Existe un valor de Mc asociado a cada una de las subsecciones en que se

divide cada sección. La cantidad de subsecciones varía de una sección a otra pues depende de la orografía del cauce; la Fig. 4.2 muestra gráficamente esta división. Por otro lado, existe un valor de M_p y una altura de albardón para cada sección. Los valores de M_c varían muy poco en cada subsección, dentro de una misma sección. Por esta razón, se tomó la determinación de hacer una primera simplificación al dominio de los parámetros, tomando un único coeficiente de M_c por sección, independientemente de las subsecciones que tenga. Esta decisión simplifica la cantidad de parámetros a 3 por sección.

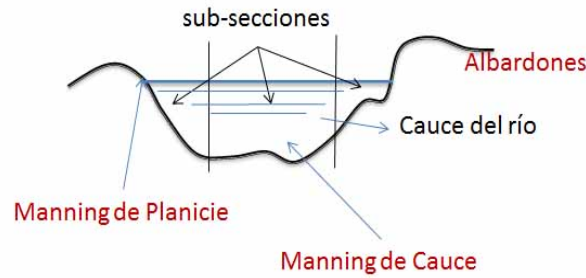


Figura 4.2: Parámetros que caracterizan una subsección del río.

4.3. Métricas de Evaluación de la Optimización via Simulación de Ezeiza

El índice utilizado como indicador del rendimiento general de la simulación, IC , depende de los datos de salida simulados, Q_{sim} , y los datos observados, Q_{obs} , obtenidos en cada una de las 15 estaciones de monitoreo, como ya se dijo al definir el modelo del río Paraná en 2.4.2. El período de simulación, n , es medido en días. La siguiente ecuación define el índice de calidad de cada escenario de simulación, en función de los índices medidos en cada estación Ie_j :

$$IC = \frac{\sum_{j=1}^{15} Ie_j}{15} \quad (4.4)$$

$$\text{donde } Ie_j = \sqrt{1/n \cdot \sum_{i=1}^n (Q_{sim}^i - Q_{obs}^i)_j^2} \quad (4.5)$$

El índice IC depende de las condiciones en que se desarrolla la simulación, o sea depende de las variables de decisión seleccionadas vistas en 4.2, del periodo de tiempo de la simulación y del tramo del río que se pretende mejorar. Si se consideran las 15 estaciones, entonces se obtiene la medida de calidad de la simulación sobre todo el dominio del río, aunque a veces se desee seleccionar el tramo a simular. En este caso, el índice corresponderá a las estaciones involucradas solamente. O sea, al hacer referencia a IC se está midiendo la calidad del escenario sobre todo el río, en cambio un índice relativo se expresa IC_{Ei}^{Ef} , donde el intervalo $[Ei .. Ef]$ estará indicando las estaciones consideradas para hacer la medición. En cambio, si se lo expresa como IC_E se está hablando de Ie_j , calculado en la estación E según la ecuación 4.5, con $j = E$.

El problema de OvS que se trata en esta tesis, cuyo principal objetivo es sintonizar el simulador Ezeiza, compara la salida del simulador, o datos simulados, con los datos observados para evaluar la función objetivo y calcular el índice de rendimiento. Pero, lo interesante de esta mejora es brindar un escenario de simulación que mejore los resultados que el INA logra actualmente con su escenario de pronóstico. Es necesario entonces definir un índice que mida esta mejora, IM , siendo su expresión:

$$IM = (\# \text{ Estaciones} : (IC_j^{Sim} < IC_j^{INA})_{j=1}^{15}) \quad (4.6)$$

IM mide la cantidad de estaciones cuyo índice de calidad, IC_j , sea menor que el logrado con el escenario INA, IC_{INA} , para las 15 estaciones de monitoreo.

Una condición necesaria para aceptar una solución es que mejore la mayor cantidad posible de estaciones sin empeorar los resultados en las restantes. Medir cuánto empeoran requiere establecer una tolerancia **tol**, tal que la diferencia entre los valores observados y los simulados no sea mayor a este parámetro, propio del proceso de optimización. De esta manera se asegura un equilibrio entre todo el dominio de simulación.

4.4. Métodos de Evaluación

En una primera etapa se implementó una prueba de concepto con el objetivo de verificar la metodología que se iba a proponer con este trabajo. Ésta consistió en implementar una búsqueda exhaustiva para encontrar un resultado al problema de optimización y así sintonizar el simulador. Se determinó una zona del dominio y se estableció una cantidad acotada de valores para las variables de decisión. Esta etapa permitió encontrar un conjunto solución, de parámetros ajustados, que mejoró la predicción de Ezeiza en alguna de las

estaciones de monitoreo del Paraná. El método se implementó mediante una simulación paramétrica en paralelo y verificó la metodología como factible de encontrar conjuntos de ajuste entre los parámetros de entrada. La Fig. 4.3 esquematiza los pasos que se tomaron en la prueba de concepto.

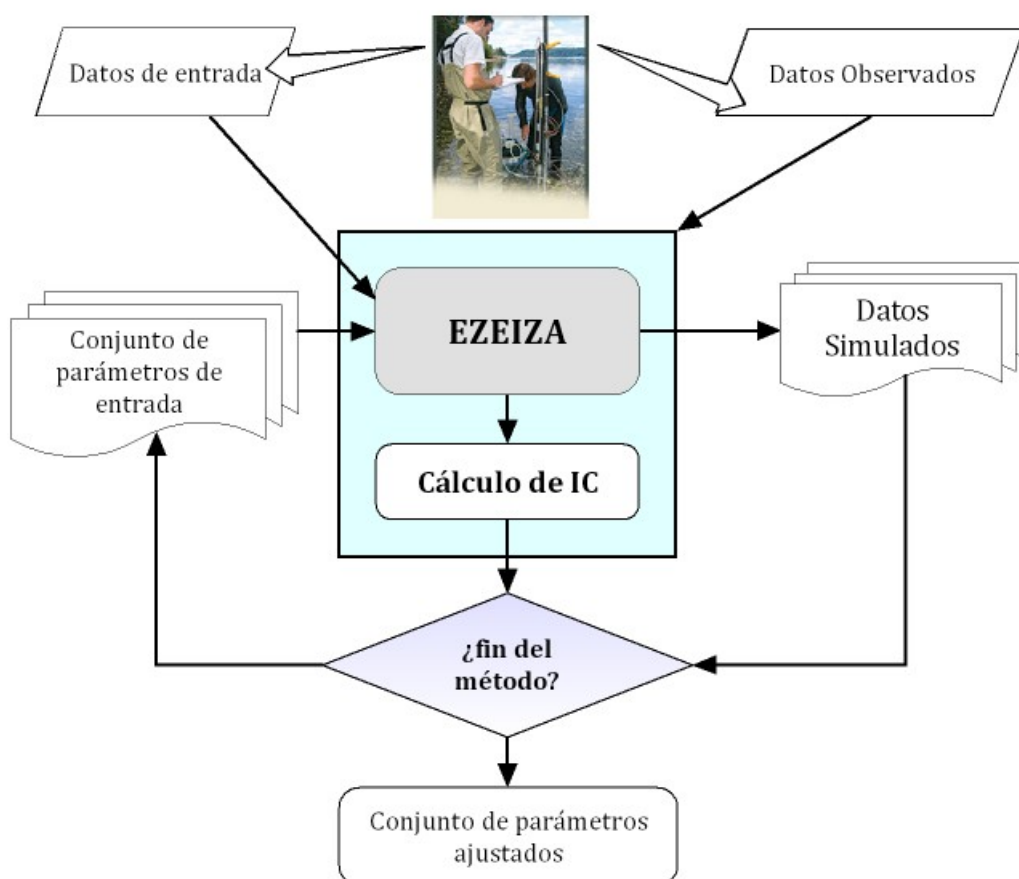


Figura 4.3: Esquema de la sintonización de la simulación con Ezeiza para todo el espacio de búsqueda en la prueba de concepto.

A fin de evaluar cuantitativamente la metodología propuesta, se establecen dos casos de estudio. Un Caso A que mejora la prueba de concepto considerando un espacio mucho mayor de escenarios y aplicando la metodología de OvS que se explicó en el capítulo anterior. Se realiza la simulación sobre el año 1999 completo. El Caso B se implementa sobre un periodo de simulación de dos meses del año 2007. Ambos periodos fueron también

considerados por el INA para hacer una calibración del modelo en el año 2011 (Latessa, 2011).

Implementación en Paralelo del Método OvS

Todas las simulaciones del método de OvS que se describen en este capítulo fueron implementados y ejecutados en equipamiento del Instituto de Investigación en Informática LIDI (III-LIDI) de la UNLP. Se utilizó un cluster de multicores HP Blade que consiste de 16 nodos, donde cada uno cuenta con 2 procesadores quad-core Intel Xeon e5405 de 2.0GHz. A su vez, cada núcleo cuenta con una caché L1 privada de 64KB (dividida en 32KB para instrucciones y otros 32KB para datos) y una caché L2 de 6MB compartida de a pares de núcleos. Con respecto a la memoria RAM disponible, 10 nodos cuentan con 10GB mientras que el resto dispone de 2GB. El sistema operativo instalado es GNU/Linux Debian 6.0 de 64 bits (versión del kernel 2.6.32).

4.5. Casos de estudio

El problema de optimización es el que se describió en el capítulo anterior y, por comodidad, se repite acá la expresión general del mismo:

$$\begin{aligned} \max / \min f(x) \\ \text{sujeto a } x \in S \end{aligned} \tag{4.7}$$

Las métricas de evaluación, que se expusieron en la sección 4.3, calculan el índice de calidad IQ según la ecuación 4.4 y el índice de mejora, según la ecuación 4.6. Las funciones objetivo que se establecen en los casos de estudio optimizan dichos índices y sus expresiones matemáticas son dependientes de los datos simulados Q_{sim} , los datos observados Q_{obs} y los obtenidos con las variables de decisión utilizadas por el INA Q_{INA} . La expresión matemática de cada objetivo es:

- Para la prueba de concepto, IC :

$$\text{Minimizar } IC : f(Q_{sim}, Q_{obs}) \tag{4.8}$$

$$\begin{aligned} \text{donde } 0,1 \leq Mp \leq 0,2 \\ 0,01 \leq Mc \leq 0,04 \end{aligned} \tag{4.9}$$

En esta prueba se optimiza el índice IC , limitando los rangos de las variables de decisión y sin tomar en cuenta los albardones. Los detalles de implementación se dan en 4.5.1.

- Para el índice IC :

$$\text{Minimizar } IC : f(Q_{sim}, Q_{obs}) \quad (4.10)$$

$$\begin{aligned} \text{donde } \quad & 0,1 \leq Mp \leq 0,4 \\ & 0,015 \leq Mc \leq 0,045 \\ & 0 \leq Al \leq 50 \end{aligned} \quad (4.11)$$

En este caso se busca el mejor conjunto de variables de decisión, para ajustar los datos observados y los simulados por Ezeiza. Las restricciones están impuestas sobre los rangos posibles de las variables de decisión, que incluyen los albardones Al . Este objetivo se aplica a los casos de estudio A y B, detallados en la próxima sección.

- Para el índice IM :

$$\begin{aligned} \text{Maximizar } IM : f(Q_{sim}, Q_{obs}, Q_{INA}) \\ \text{tal que } \frac{1}{k} \cdot \sum (IC_{\hat{e}}^{Sim} - IC_{\hat{e}}^{INA}) < tol \end{aligned} \quad (4.12)$$

donde k representa la cantidad de estaciones \hat{e} para las cuales se cumple:

$$IC_{\hat{e}}^{Sim} - IC_{\hat{e}}^{INA} < tol$$

y se mantienen las restricciones sobre las variables de decisión:

$$\begin{aligned} 0,1 \leq Mp \leq 0,4 \\ 0,015 \leq Mc \leq 0,045 \end{aligned}$$

Este objetivo maximiza la cantidad de estaciones de monitoreo que mejoran la predicción sin empeorar el resto, en no más de una tolerancia dada por tol ; siendo éste un parámetro de tolerancia propio del proceso de OvS.

- Para mejorar un tramo del río Paraná, IC_{bP} : En los casos de estudio se presenta el objetivo de minimizar IC para el bajo Paraná, pero sin empeorar la simulación en las estaciones restantes. Por lo cual, es necesario combinar los objetivos anteriores. Se

recuerda que las estaciones 12 a 15 son las que están en ese tramo, como se visualiza en la Fig. 2.6. La nueva expresión objetivo es:

$$\begin{aligned} \text{Minimizar } IC_{bP} &: f(Q_{sim}, Q_{obs}, Q_{INA}) & (4.13) \\ \text{tal que } & \left(\sum_{i=1}^{11} IC_i \right) / 11 > tol \end{aligned}$$

4.5.1. Prueba de Concepto del Simulador

Inicialmente, se consideraron 3 secciones de las 76 posibles, ubicadas en el área de Paraná Alto (Sección 76), Medio (Sección 36) y Bajo (Sección 01) y se implementó la búsqueda del mejor escenario con una búsqueda exhaustiva acotada, pero no se obtuvieron resultados satisfactorios al no poder mejorar el escenario del INA. Al ser un espacio de búsqueda tan limitado y, especialmente, al estar las secciones elegidas tan distanciadas sobre la geografía del río, no fue posible encontrar algún conjunto de soluciones mejor. Posteriormente, se trabajó con 4 secciones que estaban en el tramo inferior del río Paraná y fue posible encontrar soluciones satisfactorias. Las secciones elegidas fueron la 70, 72, 74 y 76.

En esta prueba de concepto no se consideraron los albardones como variables del problema de búsqueda. El espacio de búsqueda quedó determinado por la cardinalidad de los parámetros, los coeficientes de Mc y Mp , y por la cantidad de secciones. El rango de Mc fue $[0,01,0,04]$, tomado en pasos de 0,01, y el rango de Mp fue $[0,1,0,2]$, tomado en pasos de 0,1. Según las ecuaciones 4.2 y 4.3, resultó:

- Cardinalidad Mp : $((0.2 - 0.1) + 0.1) / 0.1 = 2$
- Cardinalidad Mc : $((0.04 - 0.01) + 0.01) / 0.01 = 4$

El tamaño del espacio de búsqueda se calculó como: $(4 \times 2)^4 = 4096$. Resultaron 4096 combinaciones posibles entre las que se buscó la solución óptima. Es importante aclarar que el algoritmo de búsqueda implementado proporcionó el óptimo de todas estas combinaciones, ya que se implementó una búsqueda exhaustiva acotada. Este óptimo no es para nada el óptimo de *todas* las posibles combinaciones que se daban con estos rangos, ya que son infinitos valores de números reales que han sido discretizados para su cómputo.

Luego de implementar el algoritmo de búsqueda mediante una simulación paramétrica, se llegó a obtener tres escenarios que mejoraron la simulación en tres estaciones ubicadas

en el bajo Paraná, sin perjudicar las restantes, las cuales mantuvieron valores simulados similares al del escenario INA. En la Tabla 4.3 se muestran los valores de los parámetros ajustados para dichos escenarios y los parámetros que utiliza el escenario INA, tomado como referencia al calcular las mejoras.

Sección	Escenario INA		Escenario 1		Escenario 2		Escenario 3	
	Planicie	Cauce	Planicie	Cauce	Planicie	Cauce	Planicie	Cauce
70	0.2	0.028	0.1	0.3	0.2	0.2	0.2	0.1
72	0.2	0.02	0.1	0.2	0.2	0.2	0.2	0.2
74	0.12	0.031	0.1	0.3	0.1	0.3	0.1	0.2
76	0.12	0.035	0.1	0.3	0.1	0.3	0.2	0.2

Tabla 4.3: Valores de las variables de decisión que proveen los escenarios solución de la búsqueda exhaustiva.

Utilizando la expresión de IC en 4.4 y de Ie en 4.5, se obtuvieron los valores que se muestran en la Tabla 4.4. El porcentaje de mejora, M_{sim} , relaciona la calidad de la simulación, obtenida con los parámetros ajustados, y la calidad de la simulación obtenida con el escenario INA. Para cada estación se calculó de la siguiente manera:

$$M_{sim} = \text{valor absoluto}(IC_{sim} - IC_{INA})/IC_{INA} \quad (4.14)$$

Las estaciones Goya, San Martín y Rosario fueron las elegidas, pues fueron las que mejor sintonizaron la simulación con Ezeiza y cuyos valores son los porcentajes de mejora. Una manera de visualizar estos resultados es mediante los hidrogramas de salida que permite crear Ezeiza, los cuales grafican la altura del agua en cada paso de tiempo que se simula, para una estación de monitoreo en particular. Los pasos de tiempo son siempre de días y las estaciones son las ciudades mejoradas. El tiempo de simulación determina la serie de datos observados a utilizar, y el mismo abarca los primeros 180 días del año 1999.

Ciudad	Escenario _i	IC _i	IC _{INA}	M _{sim}
Goya	1	0.029	0.039	25 %
San Martín	2	0.016	0.036	56 %
Rosario	3	0.025	0.031	19 %

Tabla 4.4: Índices de calidad y mejora lograda en tres estaciones.

Estos hidrogramas se muestran en las Fig. 4.4, 4.5 y 4.6. La mejora se ve en el mejor ajuste de la serie temporal de cada simulación, *Goya-Sim*, *SanMartín-Sim* y *Rosario-Sim* y la serie de los datos observados (o reales) en *Goya-Obs*, *SanMartín-Obs* y *Rosario-Obs*. Siendo durante varios períodos de tiempo más ajustada que los datos de la simulación obtenidos con el escenario INA, *Goya-INA*, *SanMartín-INA* y *Rosario-INA*.

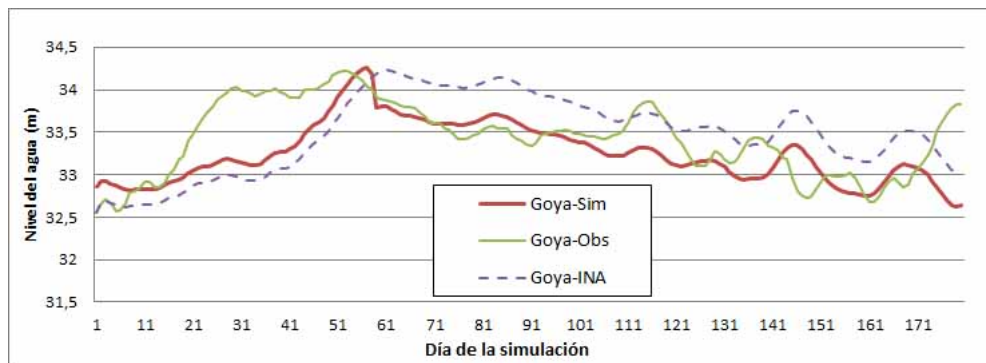


Figura 4.4: Mejora lograda a lo largo de la serie temporal en la ciudad de Goya.

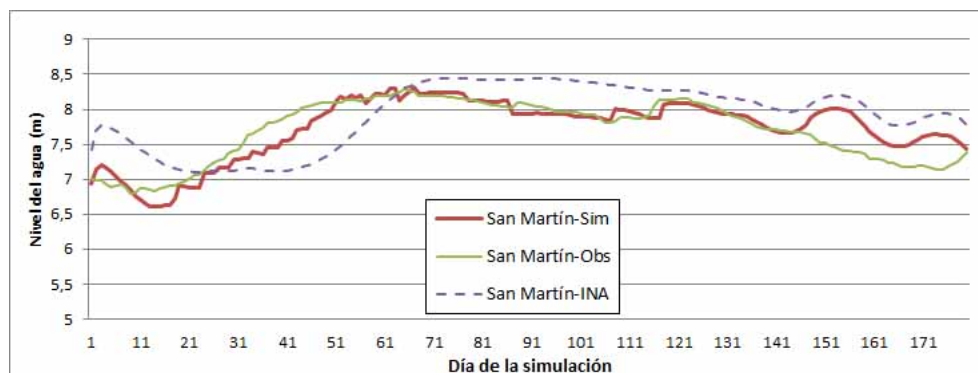


Figura 4.5: Mejora lograda a lo largo de la serie temporal en la ciudad de San Martín.

Con los datos obtenidos se promedió la distancia promedio, de los 180 días, entre las curvas de cada ciudad: datos simulados con este trabajo vs. datos observados y datos simulados con el escenario INA vs. datos observados.

1. Goya: ajuste promedio simulados: 29 cm. y promedio de ajuste con el escenario INA: 39 cm.
2. San Martín: ajuste promedio simulados: 16 cm. y promedio de ajuste con el escenario INA: 43 cm.

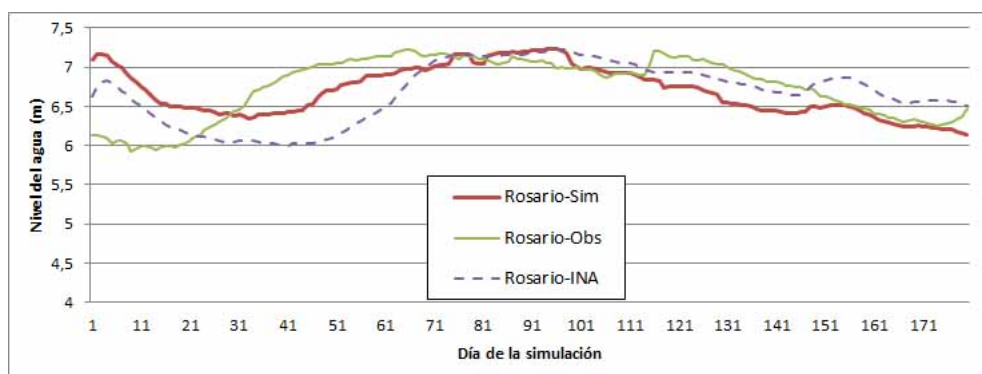


Figura 4.6: Mejora lograda a lo largo de la serie temporal en la ciudad de Rosario.

3. Rosario: ajuste promedio simulados: 24 cm. y promedio de ajuste con el escenario INA: 32 cm.

En las tres ciudades (o estaciones de monitores) se mejoró el promedio de ajuste de los datos simulados en esta prueba de concepto. Se puede ver que cuando hay una suba o una bajante importante de los niveles del agua, tanto la simulación con el escenario INA como con los escenarios seleccionados, la tendencia es apartarse de los datos observados.

4.6. Caso de Estudio A

Tiempo de simulación: 1 de enero de 1999 al 31 de diciembre de 1999 (365 días). Se elige este período por tener un período de bajante y otro de subida de los niveles del cauce del río Paraná. Se generaron las condiciones de borde y de inicio de este periodo de la simulación, de los cuales se brindan más detalles en el Apéndice A.

Restricciones respecto a los rangos de las variables de decisión, tomados según las ecuaciones 4.2 y 4.3:

- Valores de Mc en $[0,015 \dots 0,045]$, tomado en pasos de 0,01.
- Valores de Mp en $[0,2 \dots 0,4]$, tomado en pasos de 0,1.
- Valores de Al en $[0 \dots 15]$, tomado en pasos de 5.

- Cardinalidad Mp : $((0.4 - 0.2) + 0.1) / 0.1 = 3$
- Cardinalidad Mc : $((0.045 - 0.015) + 0.01) / 0.01 = 4$
- Albardones Al : $((15 - 0) + 5) / 5 = 4$

Se utilizaron 3 secciones ubicadas en el bajo Paraná, para la combinación de los parámetros. Las secciones reales seleccionadas fueron 72, 74 y 76, que se las llamará como S1, S2 y S3.

Según se muestra en el cálculo de los cardinales de las variables de decisión, el espacio total es enorme a pesar de restringir sus intervalos de posibles valores. El espacio de búsqueda, bajo estas condiciones, tiene $(4 \star 3 \star 4)^3$, o sea 110.592 configuraciones diferentes. La primera etapa del método de optimización utiliza el algoritmo de MC para configurar conjuntos de 100 escenarios en cada iteración, asignando valores aleatorios a los parámetros de cada una de las 3 secciones, dentro de los valores permitidos para cada uno. Cada configuración requiere la ejecución del simulador Ezeiza con el escenario de entrada completo, y así se van obteniendo los índices de calidad que serán analizados por el método para decidir si avanzar a la iteración siguiente, o terminar y devolver el conjunto mejor evaluado, según su *IC* promedio. Cada vez que se mide la mejora lograda en cada estación con una configuración en particular, es comparada con la lograda con la configuración del escenario INA. Si la mejora es mayor que una tolerancia *tol* del 10%, se considerará como lograda. Este límite, que al sobrepasarlo se considera una mejora, fue consensuado con los investigadores del INA.

4.6.1. Índice de Calidad *IC*

Los valores promedios del *IC* obtenidos con el método MC, con conjuntos de 100 configuraciones por vez, se distribuyeron entre 0,799 y 0,933. Luego de testear 52 conjuntos no se pudo encontrar mejor solución. El método de selección evaluó cada *IC* promedio para decidir el conjunto con las configuraciones que mantendrá almacenado como el mejor logrado. De esta manera, cuando las condiciones no son mejoradas en cinco iteraciones sucesivas el método para y devuelve el conjunto guardado. La decisión de parar en 5 iteraciones fue una decisión tomada de manera empírica ya que con más iteraciones no se obtenía ningún beneficio.

El método MC devolvió el mejor conjunto de configuraciones que encontró, el cual se muestra en la Fig. 4.7. Cada punto es el *IC* calculado para las 100 ejecuciones.

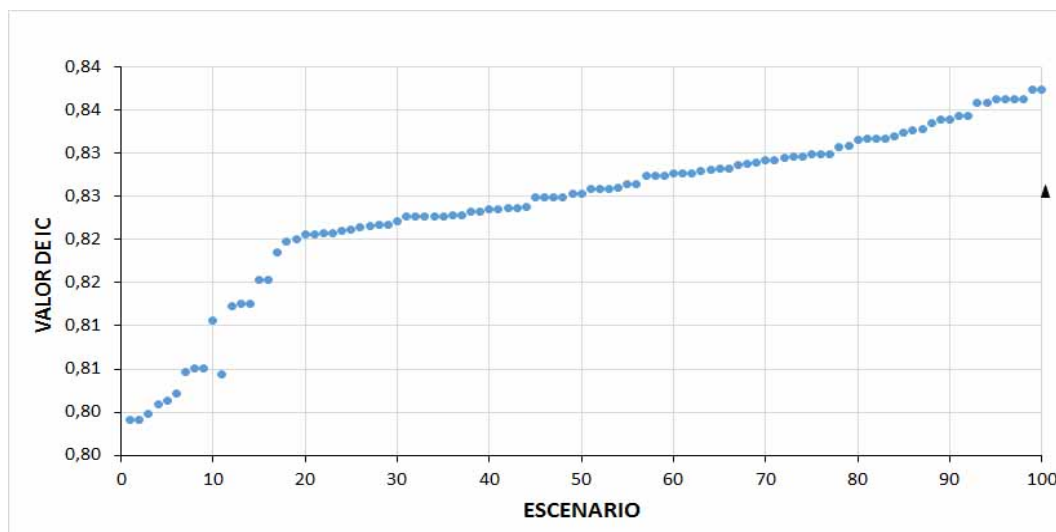


Figura 4.7: Mejor conjunto de configuraciones identificado por el método Monte Carlo.

El conjunto de escenarios que devuelve MC es agrupado mediante el algoritmo de K-Means. Los grupos identificados se muestran en la Fig. 4.8. De este resultado se observan 3 grupos de los cuales el grupo con las mejores configuraciones son los 8 puntos con valores mínimos (color azul) del *Grupo 1*.

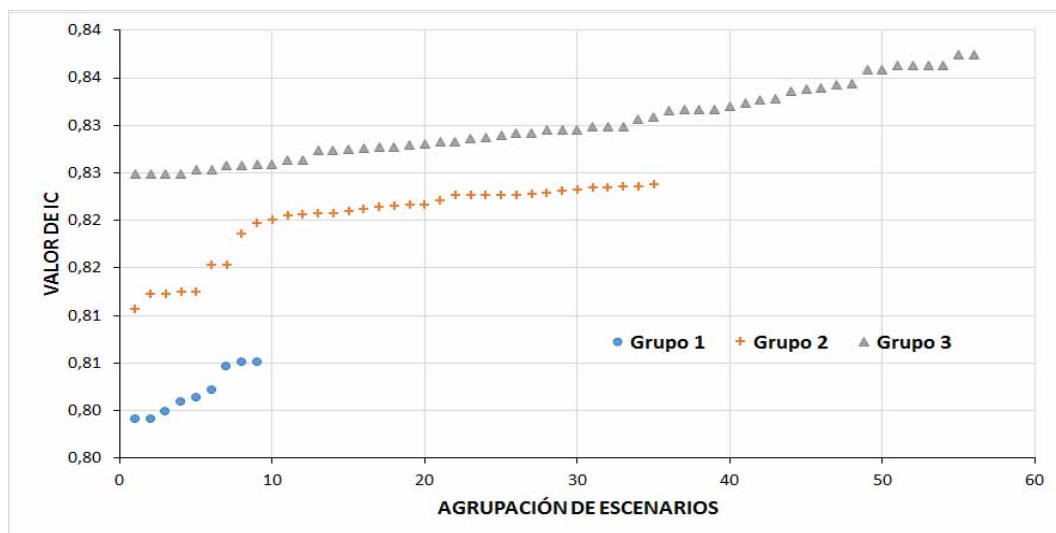


Figura 4.8: Grupos de configuraciones identificadas por el método K-Means.

Los 8 puntos del *Grupo 1* de K-Means, son los que definirán la búsqueda exhaustiva para hallar la solución. La búsqueda exhaustiva es la segunda parte del método, y es posible gracias a la reducción del espacio de búsqueda. Al finalizar esta parte se comparó

la solución con el escenario de referencia INA para evaluar la mejora obtenida. En la Fig. 4.9 se grafica la mejora lograda con 5 de los escenarios identificados por K-Means para las 15 estaciones.

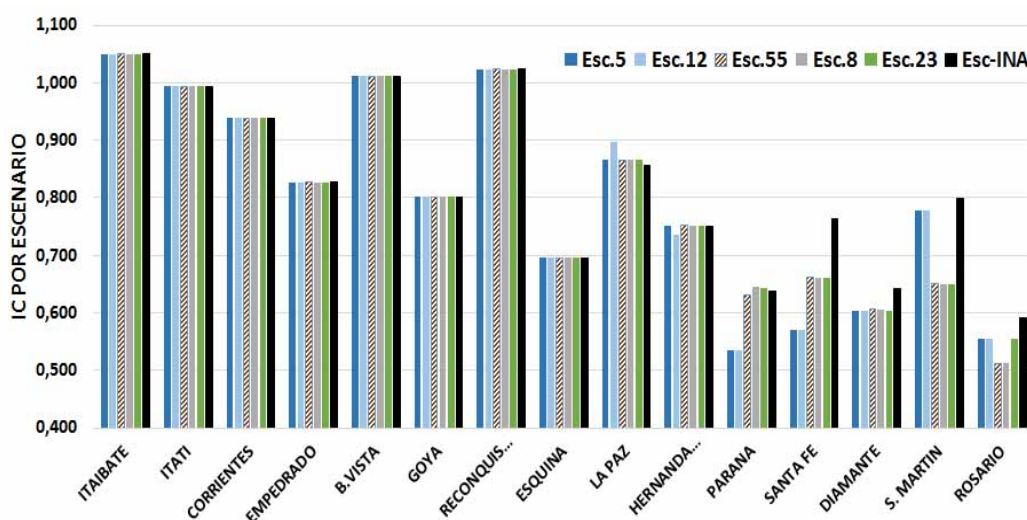


Figura 4.9: Comparación entre los valores de IC obtenidos en cada estación y la configuración INA, para 5 configuraciones del grupo identificado con K-Means.

Para facilitar la graficación, se muestran 5 escenarios solamente, pues los 3 restantes dieron resultados idénticos a los mostrados. Para todas las estaciones e se comparó su IC_e , con el índice del escenario INA, IC_{INA} . Las estaciones Paraná, Santa Fe, Diamante, San Martín y Rosario tuvieron mejoras con los escenarios elegidos. Las estaciones restantes se mantuvieron con valores estables y no perjudicaron la predicción en ninguna de estas estaciones (diferencias menores a tol).

La tabla 4.5 resume los valores de los parámetros que corresponden a los escenarios seleccionados mediante K-Means, los cuales acotaron el espacio de búsqueda de cada parámetro. La BE acotada los utilizó para crear el nuevo espacio, de mucho menor tamaño, combinando todas sus posibilidades. Resultando, 4 combinaciones para la Sección 1, 6 combinaciones para la Sección 2 y 8 combinaciones para la Sección 3.

El nuevo espacio tiene 192 combinaciones.

La mejor configuración, resultante de la BE en la segunda parte del método, es la configuración óptima mostrada en la Tabla 4.6.

Escenario	Sección 1			Sección 2			Sección 3		
	Mc	Mp	Al	Mc	Mp	Al	Mc	Mp	Al
1	0,02	0,2	0	0,02	0,1	5	0,03	0,2	0
2	0,02	0,2	0	0,02	0,2	5	0,03	0,2	0
3	0,02	0,2	5	0,02	0,1	5	0,04	0,1	0
4	0,02	0,2	0	0,02	0,2	0	0,04	0,1	0
5	0,02	0,1	5	0,02	0,1	0	0,03	0,2	5
6	0,02	0,1	5	0,02	0,1	0	0,04	0,1	5
7	0,02	0,1	0	0,02	0,2	0	0,04	0,1	0
8	0,02	0,1	0	0,02	0,2	5	0,03	0,2	0

Tabla 4.5: Variables de decisión en los escenarios determinados por el método K-Means.

Solución Óptima - IC			
Sección	Mc	Mp	Al
1	0,02	0,2	5
2	0,02	0,2	5
3	0,03	0,1	0

Tabla 4.6: Solución óptima lograda por la BE acotada.

El *IC* de esta configuración, identificada por la BE en la segunda etapa del método, fue: **0,792**. La mejora lograda se muestra en la Tabla 4.7. Se aprecia que las estaciones mejoradas fueron Santa Fe y San Martín en un buen porcentaje cada una. Otras estaciones del tramo inferior mejoraron un poco su predicción, pero por debajo de *tol*. Se llama Escenario *OvS* al que se determina con la configuración óptima que se muestra en la Tabla 4.6.

Estación	Escenario		Mejora
	OvS	INA	
Santa Fe	0,6127	0,764	20 %
San Martín	0,6051	0,799	24 %

Tabla 4.7: Mejora lograda con la configuración óptima de Caso A, considerando *IC* como objetivo.

La Fig. 4.10 compara la solución óptima del Caso A, Escenario *Esc.OvS* obtenido con el objetivo *IC*, y los *IC* que se obtienen con el escenario INA para cada estación del dominio.

En este gráfico se aprecian las mejoras significativas logradas en las estaciones ubicadas sobre el tramo inferior del Paraná.

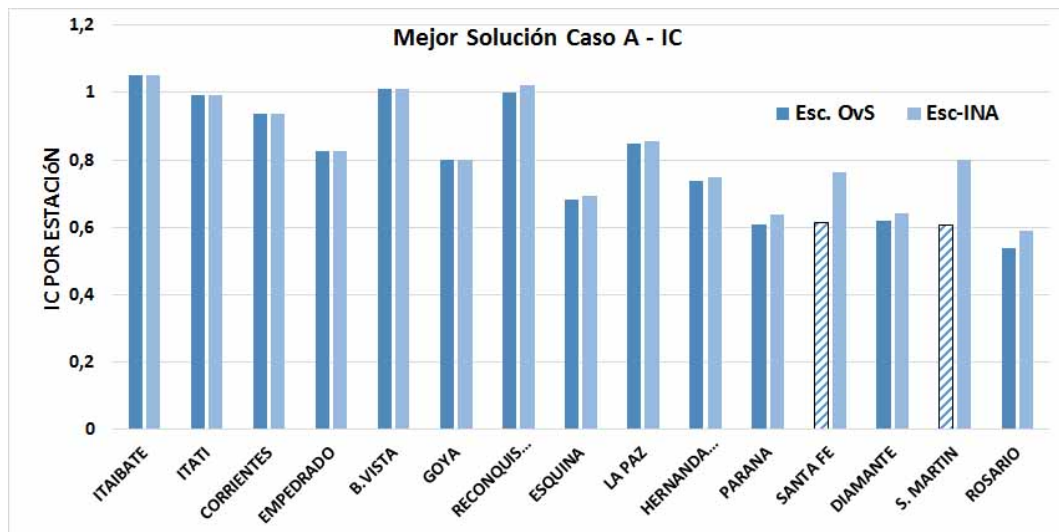


Figura 4.10: Comparación entre los valores de IC para el Esc.OvS en cada estación y la configuración INA, en el Caso A.

4.6.2. Índice de Calidad IC más Índice de Mejora IM - $IC&IM$

Al considerar también el objetivo de maximizar IM , según se expresa en 4.12, en esta experimentación se buscaron las configuraciones que a la vez pudieron minimizar IC y maximizar IM . O sea, Monte Carlo analizó cada conjunto de 100 escenarios y decidió hasta cuándo continuar el método controlando ambos objetivos. El conjunto de configuraciones obtenido fue tal que su IC promedio se encontró en el mismo rango de valores que antes, el cual se muestra en la Fig. 4.11, donde se puede ver el punto rojo correspondiente a la solución encontrada por la BE acotada y el valor del IC del INA. En este gráfico se muestran los valores de IC y de IM correspondiente a cada escenario del conjunto identificado por el método MC, donde el círculo encierra la zona en la que se encuentra la solución que cumple con ambos objetivos.

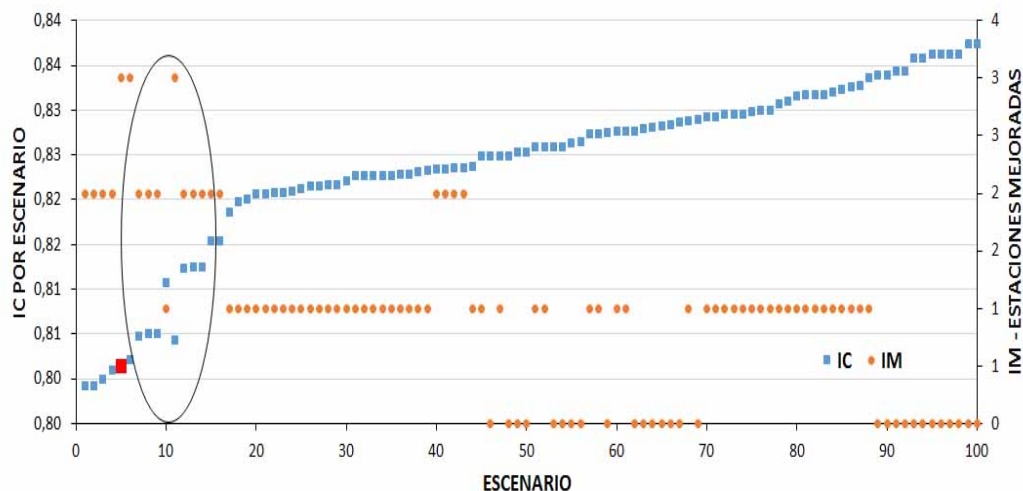


Figura 4.11: Configuraciones que optimizan IC e IM

El método K-Means identificó tres grupos de puntos, como se ve en la Fig. 4.12. El *Grupo 1*, en el círculo azul, encierra la zona factible que contiene la solución.

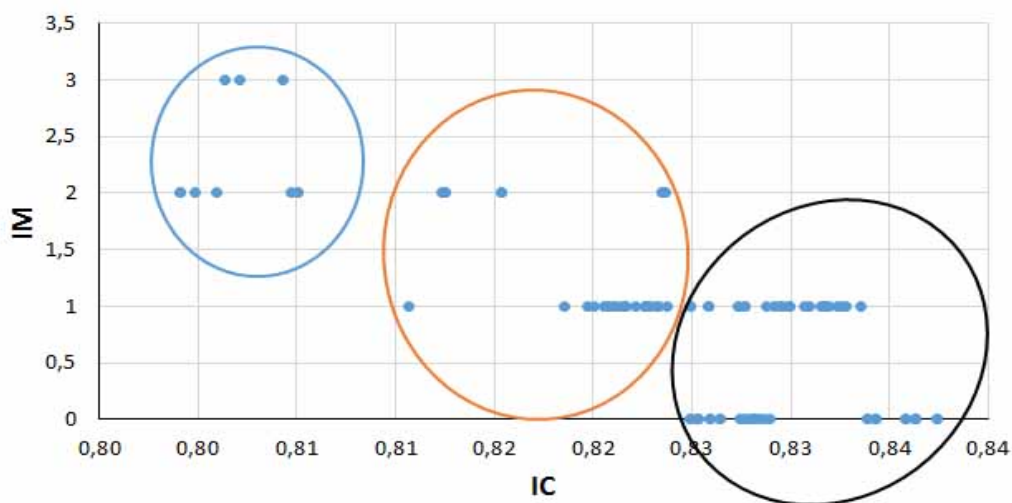


Figura 4.12: Grupos de configuraciones identificados por K-Means, considerando IC & IM .

En la Fig. 4.13 se comparan los escenarios considerados más significativos en cada una de las 15 estaciones, en función de la mejora aportada a la simulación. Estos escenarios forman parte de los puntos del *Grupo 1* encontrado por K-Means y se forman a partir de dichas configuraciones.

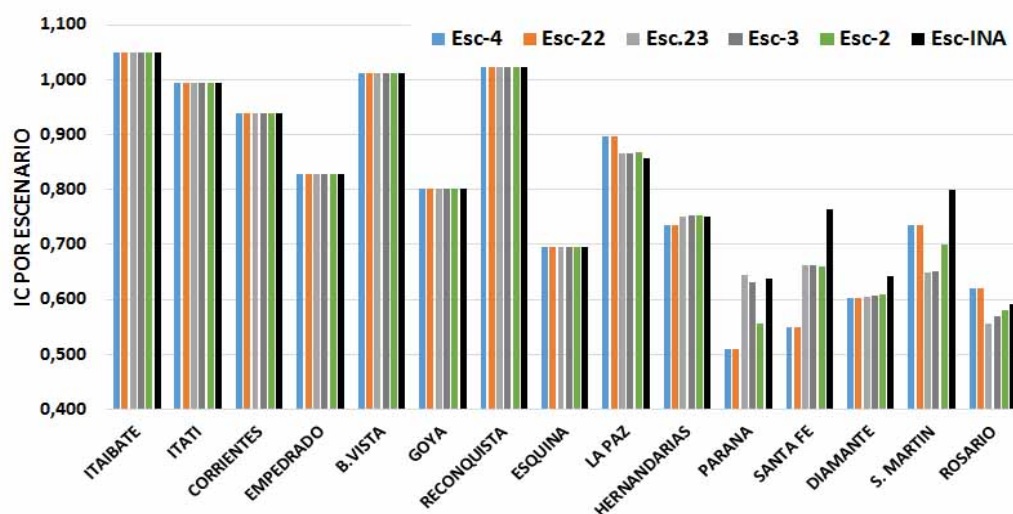


Figura 4.13: Comparación de los valores de IC con 5 configuraciones del grupo identificado por K-Means, en cada una de las 15 estaciones. Se comparan también con el valor de IC correspondiente a la configuración INA. Caso que considera $IC&IM$ como objetivo.

Utilizando este grupo de configuraciones, se acotó el espacio de búsqueda y sobre este espacio la BE acotada identificó como solución óptima la configuración que se muestra en la Tabla 4.8. Es importante aclarar que la solución óptima no corresponde al escenario que tiene el *mínimo IC del conjunto*, sino el de menor valor IC que a la vez maximiza IM .

Solución Óptima - $IC+IM$			
Sección	Mc	Mp	Al
1	0,02	0,2	5
2	0,02	0,2	5
3	0,04	0,2	0

Tabla 4.8: Solución óptima lograda con el método MC + K-Means considerando los objetivos $IC&IM$.

La solución identificada por la BE acotada dentro de la región factible, provee el conjunto de parámetros ajustados que determinan el escenario óptimo de simulación, con el cual se obtuvo la mejora mostrada en la Tabla 4.9. El índice IC óptimo del método es **0,801** y el valor de IM es: **3** estaciones.

Estación	Escenario		Mejora
	OvS	INA	
Santa Fe	0,662	0,764	13 %
San Martín	0,651	0,799	19 %
Rosario	0,512	0,59	13 %

Tabla 4.9: Mejora lograda con la configuración óptima considerando $IC&IM$ como objetivo, para el Caso A.

La Fig. 4.14 compara la solución óptima del Caso A y el objetivo dado por $IC&IM$ ($Esc.OvS$) con el escenario INA, en cada estación del dominio. Diamante y Paraná mejoran, pero no fueron consideradas en el cálculo del índice IM , lo cual se debe a que su mejora fue inferior a tol . En cambio, en las tres estaciones mejoradas se superó el límite impuesto por tol .

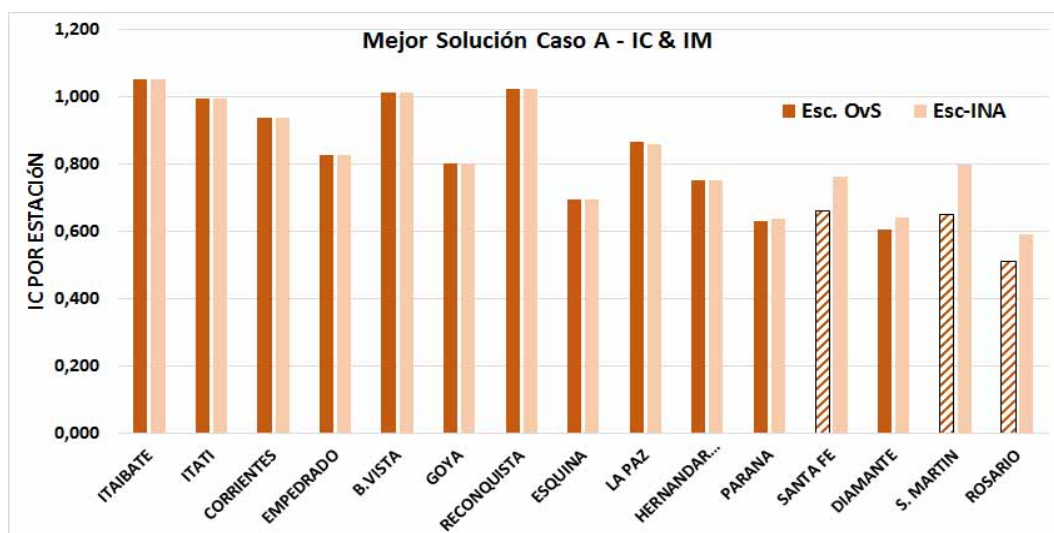


Figura 4.14: Comparación entre los valores de IC para el $Esc.OvS$ en cada estación y la configuración INA, en el Caso A y tomando como objetivos $IC&IM$.

La cantidad de estaciones que pudieron ser mejoradas a la vez alcanzó a 3, en este caso. Estas estaciones ajustaron y sintonizaron la predicción de Ezeiza con un mismo escenario. Las estaciones que mejoraron están ubicadas en la zona del bajo y medio Paraná, donde también se encuentran las secciones seleccionadas para variar sus parámetros de ajuste, como sucede con todas las experiencias ya desarrolladas. Se puede afirmar que la configuración encontrada ajusta y mejora la predicción en las estaciones Santa Fe, San Martín y Rosario, y en una menor medida en Diamante y Paraná.

El valor de IC de la configuración óptima fue de 0,801, siendo levemente mayor al obtenido con el objetivo IC solamente, que fue de 0,792. En este caso, con el multiobjetivo $IC&IM$, se pierde un poco en la calidad de la simulación con el escenario $Esc.OvS$, pero se gana en la cantidad de estaciones mejoradas simultáneamente, que en este caso son 3 estaciones.

4.7. Caso de Estudio B

Tiempo de simulación: 01/02/2007 a 01/04/2007. Se elige este período por tener una crecida importante, con un avance de onda brusco. En este nuevo intervalo de simulación fue necesario modificar las series temporales de las condiciones de borde e iniciales del simulador. En el Apéndice A se brindan más detalles.

Restricciones respecto a los rangos de las variables de decisión, tomados según las ecuaciones 4.2 y 4.3:

- Valores de Mc en $[0,02 \dots 0,04]$, tomado en pasos de 0,01.
- Valores de Mp en $[0,2 \dots 0,3]$, tomado en pasos de 0,05.
- Valores de Al en $[0 \dots 15]$, tomado en pasos de 5.

- Cardinalidad Mp : $((0.2 - 0.3) + 0.05) / 0.05 = 3$
- Cardinalidad Mc : $((0.02 - 0.04) + 0.01) / 0.01 = 3$
- Albardones Al : $((15 - 5) + 5) / 5 = 4$

Se utilizaron 3 secciones ubicadas entre el Paraná medio y la parte inicial del bajo Paraná, para la combinación de los parámetros. Las secciones reales seleccionadas fueron 64, 67 y 70, que se las llamará como S1, S2 y S3.

Según se muestra en el cálculo de los cardinales de las variables de decisión, el espacio total es enorme también en este caso de experimentación, a pesar de restringir sus intervalos de posibles valores. El espacio de búsqueda, bajo estas condiciones, tiene $(3 \star 3 \star 4)^3$, o sea 46656 configuraciones diferentes.

Este caso de estudio del método se desarrolló con la misma metodología que el caso A explicado anteriormente. Se mantuvo la misma tolerancia tol del 10%.

4.7.1. Índice de Calidad IC

Los valores promedios del IC obtenidos con el método MC, con conjuntos de 100 configuraciones por vez, se distribuyeron entre 0,658 y 0,935. Luego de testear 38 conjuntos no se pudo encontrar mejor solución. El método de selección evaluó cada IC promedio para decidir el conjunto con las configuraciones que mantendrá almacenado como el mejor logrado. En total se hicieron 3800 ejecuciones del simulador. De esta manera, cuando las condiciones no son mejoradas en cinco iteraciones sucesivas el método para y devuelve el conjunto guardado. La decisión de parar en 5 iteraciones como en el caso A se mantuvo por las mismas razones.

El método MC devolvió el mejor conjunto de configuraciones que encontró, el cual se muestra en la Fig. 4.15. Cada punto es el IC calculado para las 100 ejecuciones. En el gráfico se muestra el valor obtenido con la configuración INA.

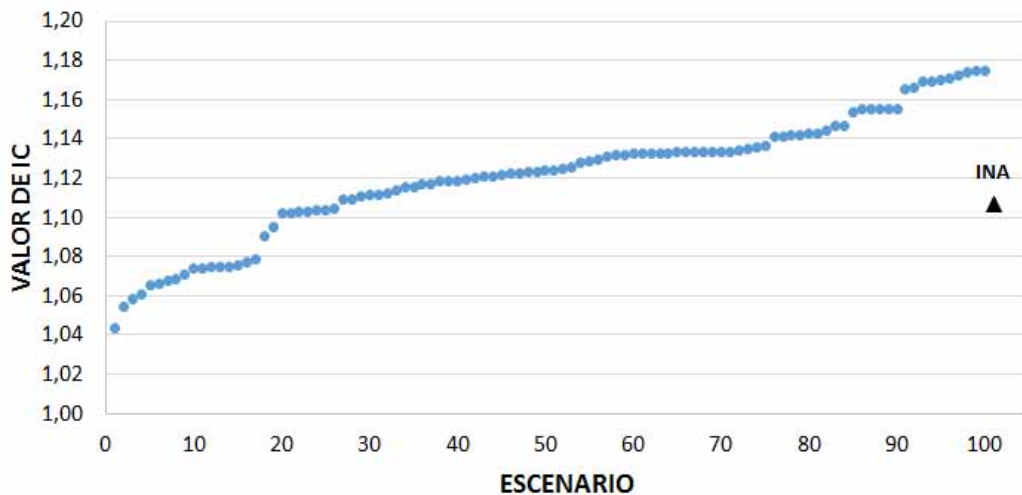


Figura 4.15: Configuración identificada por el método Monte Carlo.

Los grupos identificados mediante el método K-Means se muestran en la Fig. 4.16. De este resultado se identificaron 3 grupos de los cuales el grupo con las mejores configuraciones son los 17 puntos (color azul) del *Grupo 1*.

En la Fig. 4.17 se grafica la mejora lograda con los escenarios identificados por K-Means para las 15 estaciones. Para facilitar la graficación, se muestran los 3 escenarios más significativos respecto a la mejora lograda. Para toda estación e se comparó su IC_e , con el índice del escenario INA, IC_{INA} . Las estaciones Paraná, Santa Fe, Diamante, San Martín y Rosario tuvieron mejoras con los escenarios elegidos. Las estaciones restantes se

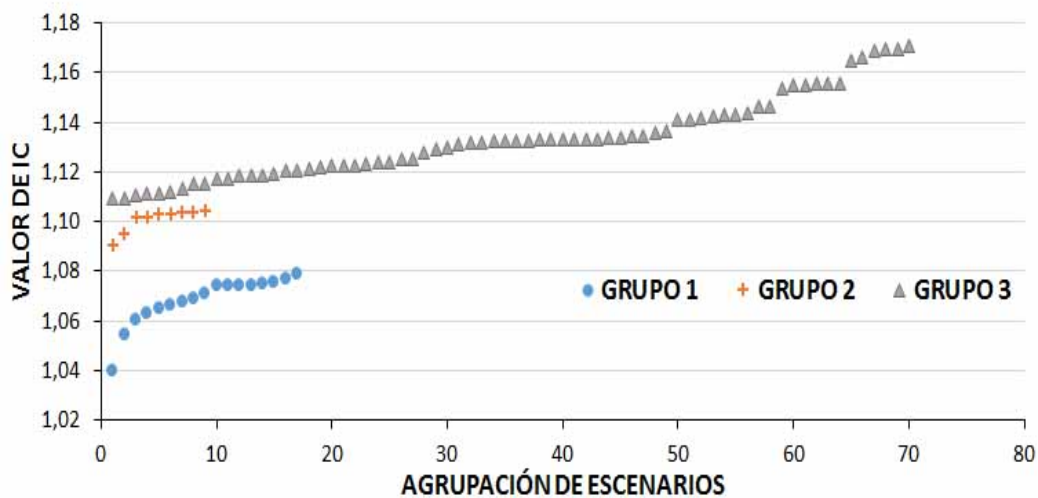


Figura 4.16: Grupos de configuraciones identificadas por el método K-Means.

mantuvieron con valores estables, excepto en el caso de Hernandarias donde el rendimiento disminuye, pero no superó la tolerancia dada por *tol*. Esta condición se debe cumplir con toda configuración seleccionada por el método

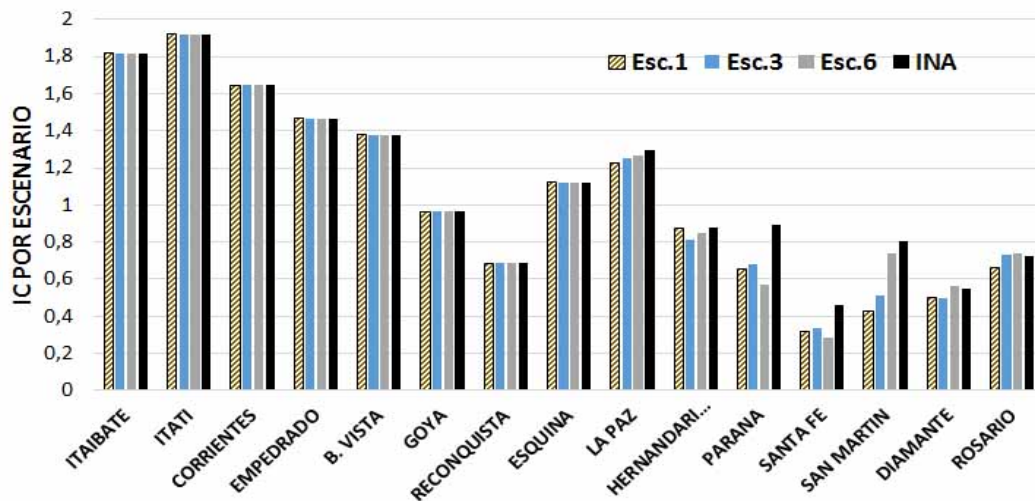


Figura 4.17: Comparación entre los valores de *IC* obtenidos con 3 configuraciones del grupo identificado con K-Means y la configuración INA.

Los 17 puntos del *Grupo 1* de K-Means definen el espacio de factibilidad sobre la que se lleva adelante la BE acotada para hallar la solución. Al finalizar esta parte se comparó la solución con el escenario de referencia INA para evaluar la mejora obtenida.

La tabla 4.10 resume los valores de los parámetros que corresponden a los escenarios seleccionados mediante K-Means, los cuales acotaron el espacio de búsqueda de cada parámetro. Resultando 12 combinaciones para la Sección 1, 12 combinaciones para la Sección 2 y 6 combinaciones para la Sección 3.

El nuevo espacio tiene ahora 864 combinaciones.

Escenario	Sección 1			Sección 2			Sección 3		
	Mc	Mp	Al	Mc	Mp	Al	Mc	Mp	Al
1	0,03	0,1	0	0,025	0,2	5	0,02	0,1	10
2	0,03	0,2	0	0,025	0,2	5	0,02	0,2	10
3	0,03	0,2	5	0,025	0,2	10	0,02	0,1	10
4	0,02	0,2	0	0,025	0,1	0	0,02	0,1	10
5	0,03	0,1	5	0,025	0,1	0	0,02	0,2	10
6	0,03	0,1	5	0,025	0,1	0	0,02	0,1	5
7	0,03	0,1	10	0,025	0,2	0	0,02	0,1	5
8	0,03	0,1	10	0,025	0,2	5	0,02	0,1	5
9	0,03	0,1	5	0,025	0,1	5	0,02	0,2	5
10	0,03	0,1	5	0,03	0,1	5	0,02	0,1	5
11	0,03	0,1	5	0,025	0,1	5	0,02	0,2	10
12	0,03	0,1	10	0,025	0,2	5	0,02	0,1	5
13	0,02	0,1	0	0,025	0,1	0	0,02	0,1	10
14	0,02	0,1	0	0,025	0,2	0	0,02	0,2	0
15	0,02	0,1	0	0,025	0,1	0	0,02	0,1	5
16	0,02	0,1	0	0,03	0,2	0	0,02	0,2	5
17	0,02	0,1	0	0,03	0,1	0	0,02	0,2	0

Tabla 4.10: Variables de decisión en los escenarios determinados por el método K-Means.

La mejor configuración, resultante de la BE en la segunda parte del método, es la configuración óptima mostrada en la Tabla 4.11.

Solución Óptima - IC			
Sección	Mc	Mp	Al
1	0,03	0,2	0
2	0,025	0,2	10
3	0,03	0,1	5

Tabla 4.11: Solución óptima lograda por la BE acotada.

El IC de esta configuración, identificada por la BE en la segunda etapa del método, fue: **1,04**. La mejora lograda se muestra en la Tabla 4.12. Se aprecia que las estaciones mejoradas fueron Paraná, Santa Fe y San Martín en un buen porcentaje cada una. Otras estaciones del tramo inferior mejoraron un poco su predicción, pero por debajo de *tol*.

Estación	Escenario		Mejora
	Mejor	INA	
Paraná	0,547	0,895	39 %
Santa Fe	0,355	0,457	22 %
San Martín	0,557	0,808	31 %

Tabla 4.12: Mejora lograda con la configuración óptima considerando IC como objetivo, para el Caso B.

La Fig. 4.18 muestra la comparación entre la solución óptima del Caso B con el objetivo dado por IC , comparada con el escenario INA.

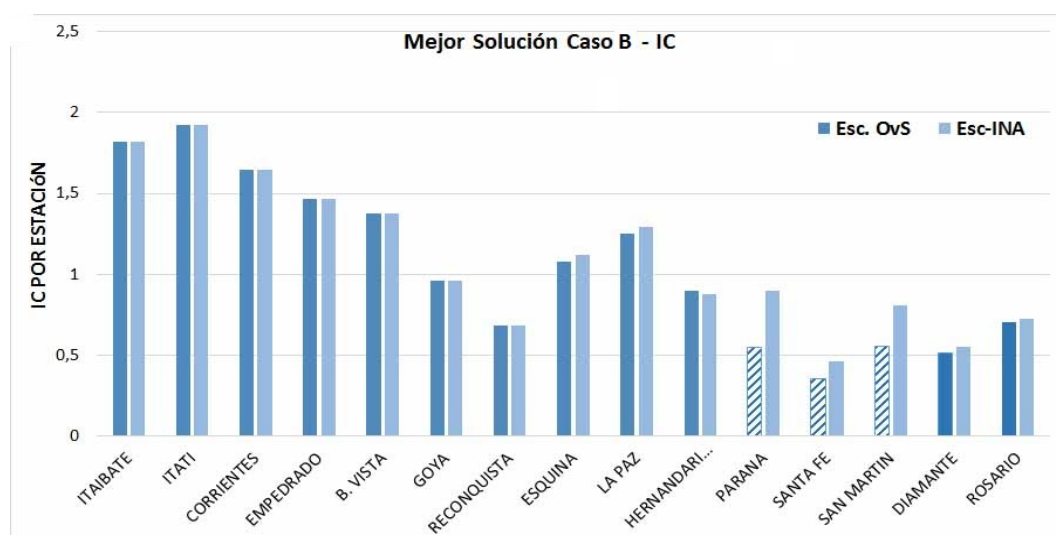


Figura 4.18: Comparación entre los valores de IC para el Esc.OvS en cada estación y la configuración INA, en el Caso B.

4.7.2. Índice de Calidad IC más Índice de Mejora IM - $IC&IM$

En esta parte se considera también el objetivo de maximizar IM , según se expresa en 4.12, en esta experimentación se buscaron las configuraciones que a la vez pudieron minimizar IC y maximizar IM . El conjunto de configuraciones obtenido por el método MC fue un conjunto cuyo IC promedio se encontró en el mismo rango de valores que antes. En la Fig. 4.19 se muestra cada escenario del grupo identificado por el método MC con su valor de IC y de IM . Se puede ver el punto rojo correspondiente a la solución de la BE acotada y el valor correspondiente al IC del INA. El círculo encierra la zona en la que se encuentra la solución que cumple con ambos objetivos.

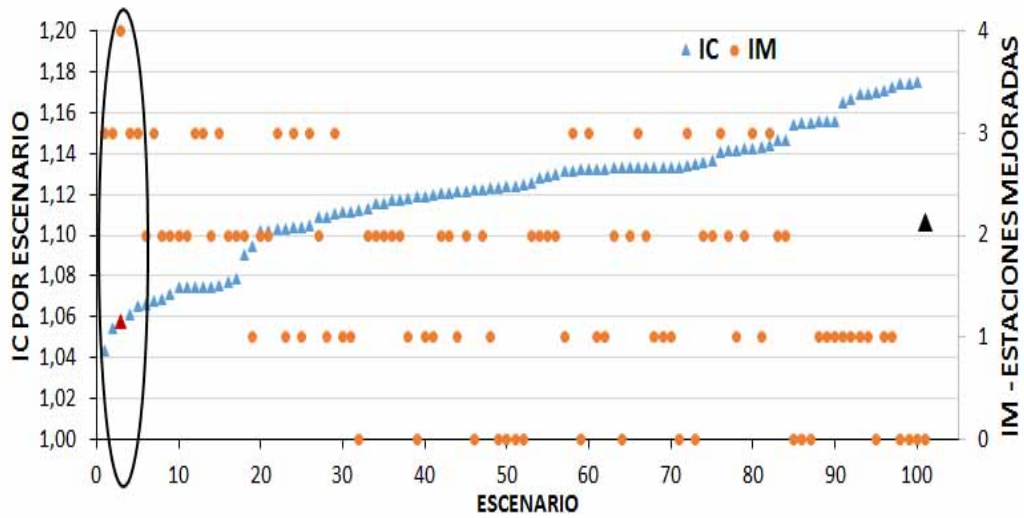


Figura 4.19: Configuraciones que optimizan IC e IM

El método K-Means identificó tres grupos de puntos, como se ve en la Fig. 4.20. El *Grupo 1*, en el círculo azul, encierra la zona factible que contiene la solución.

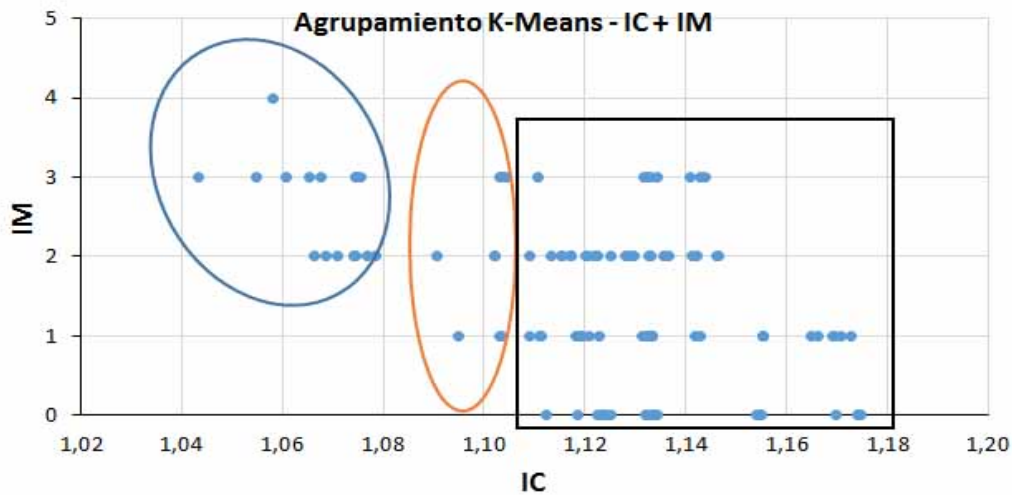


Figura 4.20: Grupos de configuraciones identificados por K-Means, considerando $IC + IM$.

En la Fig. 4.21 se comparan los escenarios considerados más significativos en cada una de las 15 estaciones, o sea en función de la mejora aportada a la simulación. Estos escenarios forman parte de los puntos del *Grupo 1* obtenido por K-Means.

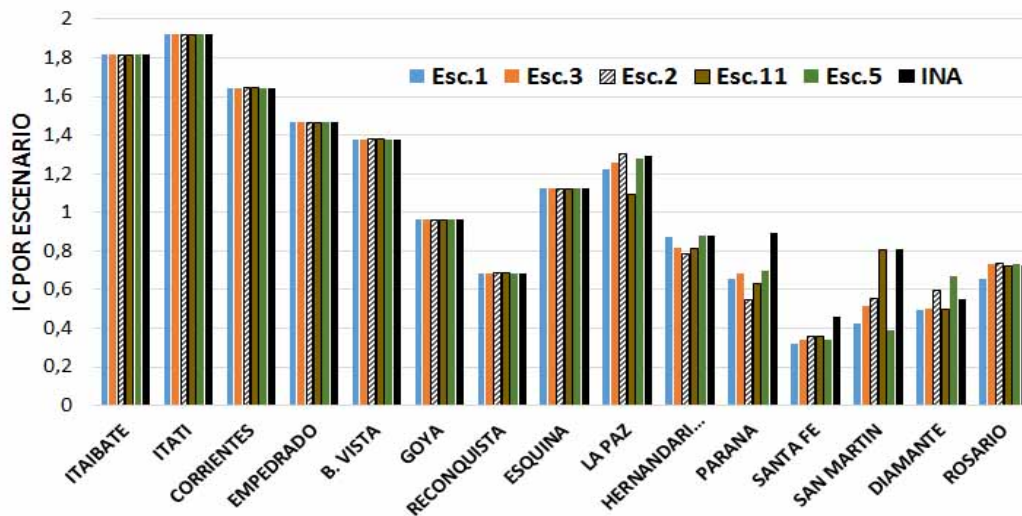


Figura 4.21: Comparación de los valores de IC con cada configuración del grupo identificado por K-Means, en cada una de las 15 estaciones. Se comparan también con el valor de IC correspondiente a la configuración INA. Caso que considera $IC&IM$ como objetivo.

Utilizando este grupo de configuraciones se obtuvo un espacio reducido de búsqueda, sobre el cual la BE acotada identificó como solución óptima la configuración que se muestra en la Tabla 4.13.

Solución Óptima - $IC + IM$			
Sección	Mp	Mc	AI
1	0,03	0,2	0
2	0,025	0,2	10
3	0,03	0,1	5

Tabla 4.13: Solución identificada con el método MC + K-Means considerando los objetivos $IC&IM$.

Las estaciones que mejoraron están ubicadas en la zona del bajo y medio Paraná. En este caso, la configuración solución no alcanzó a mejorar a Rosario. La causa de esta situación es que las secciones seleccionadas para variar sus parámetros de ajuste, están ubicadas en la zona norte del bajo Paraná y en la zona media del Paraná, pero Rosario está al sur del bajo Paraná y no fue alcanzada por la experimentación. La búsqueda de la solución que cumpla ambos objetivos está limitada por el *Grupo 1*.

La configuración óptima identificada por la BE acotada, permitió lograr las mejoras que se describen en la Tabla 4.14. El índice IC óptimo del método es **1,06** y el valor de IM es: 4 estaciones.

Estación	Escenario		Mejora
	Mejor	INA	
Hernandarias	0,785	0,878	11 %
Paraná	0,627	0,895	30 %
Santa Fe	0,360	0,457	21 %
San Martín	0,611	0,808	24 %

Tabla 4.14: Mejora lograda con la configuración óptima considerando $IC&IM$, para el Caso B.

La cantidad de estaciones que pudieron ser mejoradas a la vez alcanzó a 4 con el objetivo múltiple. Se ajustó y sintonizó la predicción de Ezeiza con un mismo escenario. Las estaciones que mejoraron están ubicadas en la zona del bajo y medio Paraná, donde también se encuentran las secciones seleccionadas para variar sus parámetros de ajuste, como sucede con todas las experiencias ya desarrolladas.

La Fig. 4.22 compara la solución óptima del Caso B y el objetivo dado por $IC&IM$, con el escenario INA, en cada estación del dominio.

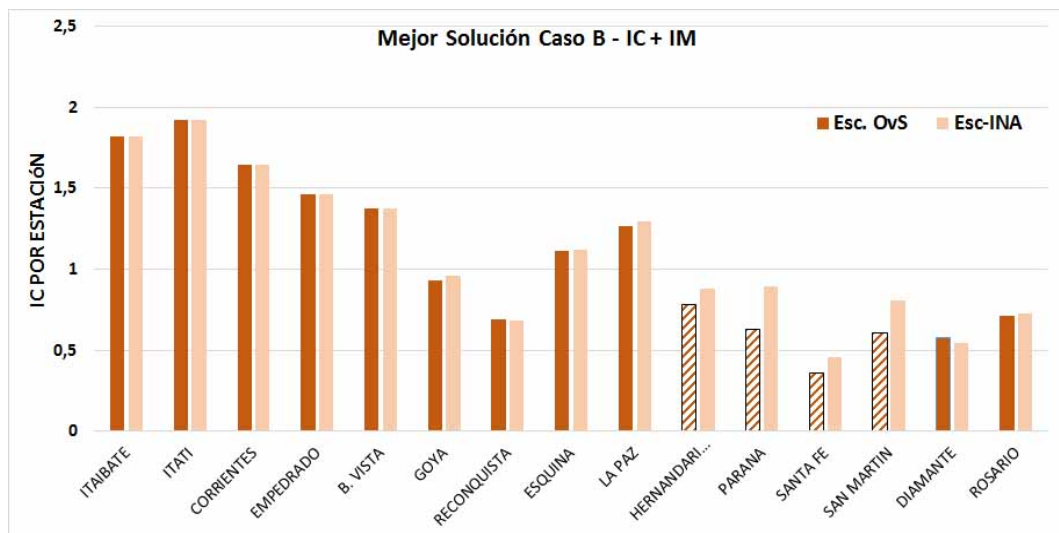


Figura 4.22: Comparación entre los valores de IC para el Esc.OvS en cada estación y la configuración INA, en el Caso B y tomando como objetivos $IC&IM$.

Como se muestra gráficamente, la solución encontrada por la BE acotada mejoró la predicción en las 4 ciudades Hernandarias, Paraná, Santa Fe y San Martín, considerando

siempre la precisión dada por tol . Como se ve, se mejoran estas estaciones y no se empeora la predicción en las restantes, excepto en Diamante donde tampoco se pasa del límite dado por tol . Esto hace que el método de OvS considere esta configuración como válida primero y la seleccione como óptima al finalizar.

El valor de índice de calidad IC de la configuración óptima fue de 1,06. En este caso, como en el Caso A fue levemente mayor al obtenido con el objetivo IC solamente, que fue de 1,04. En este caso, con el multiobjetivo $IC&IM$, se pierde un poco en la calidad de la simulación con el mejor escenario $Esc.OvS$, pero se gana en la cantidad de estaciones mejoradas simultáneamente, que en este caso son 4 estaciones.

4.8. Ganancia lograda con el uso HPC

En las secciones previas se presentó la reducción del espacio de búsqueda lograda en cada caso de estudio, al aplicar la metodología propuesta de OvS en dos etapas. Con estos resultados se muestra la mejora que el método logra en la predicción del simulador utilizado en las experiencias y que son discutidas en 4.9.

A continuación, se muestran los resultados obtenidos por la implementación computacional en paralelo del método de OvS y la ganancia en general que brindó esta propuesta.

4.8.1. Caso de Estudio A

El método propuesto en esta tesis permitió lograr una ganancia en la eficiencia de la implementación, que se puede analizar en dos niveles de su desarrollo. En un primer lugar, el método de OvS permitió reducir del espacio de búsqueda y, en consecuencia, obtener una ganancia medida en cantidad de simulaciones o número total de escenarios. En segundo lugar, la implementación del método utilizando los beneficios del HPC brindó su propia ganancia que se suma a la primera.

La ganancia del método Monte Carlo y la ganancia global están representadas en la Fig. 4.23.

La tabla 4.15 compara los tiempos de ejecución del método, en el caso de la implementación de todo el espacio de búsqueda y en los dos niveles del método de OvS. Para cada uno, se comparan los tiempos al ejecutar una versión secuencial y una en paralelo con 16 procesadores.

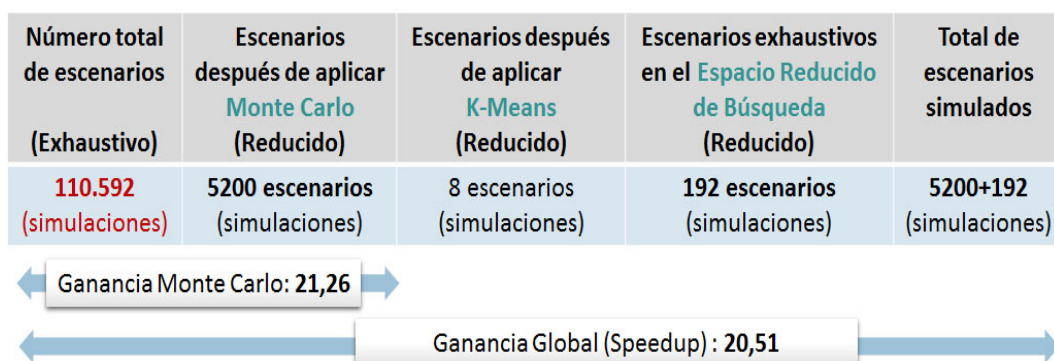


Figura 4.23: Comparación de la ganancia lograda por la reducción del espacio de búsqueda y el uso de HPC en el Caso de Estudio A.

CASO A	Secuencial		16 cores		Ganancia	1000 cores
Simulaciones	minutos	días	minutos	días	speedup	minutos
110592	243302	169	15896	11	15,306	240
5200	11440	8	756	0,5	15,132	11
8						
192	422	0,3	27	0	15,63	0,4

Tabla 4.15: Tiempo de ejecución y ganancia lograda en cada etapa del método de OvS, al aplicar HPC en la implementación.

4.8.2. Caso de Estudio B

La ganancia del método Monte Carlo y la ganancia global están representadas en la Fig. 4.24.

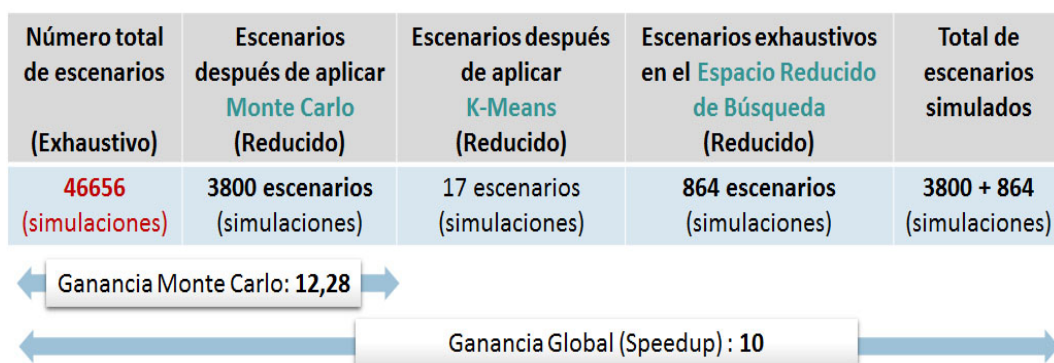


Figura 4.24: Comparación de la ganancia lograda por la reducción del espacio de búsqueda y el uso de HPC en el Caso de Estudio B.

La tabla 4.16 compara los tiempos de ejecución del método, en el caso de la implementación de todo el espacio de búsqueda y en los dos niveles del método de OvS. Como en el Caso A, se comparan los tiempos al ejecutar una versión secuencial y una en paralelo con 16 procesadores.

CASO B	Secuencial		16 cores		Ganancia	1000 cores
Simulaciones	minutos	días	minutos	días	speedup	minutos
46656	100643	70	6298	4,4	15,98	100
3800	8245	6	525	0,4	15,71	8
17						
864	1876	1,3	124	0	15,13	2

Tabla 4.16: Tiempo de ejecución y ganancia lograda en cada etapa del método de OvS, al aplicar HPC en la implementación.

4.9. Discusión

Las conclusiones más importantes de este capítulo son las siguientes:

1. En este capítulo se describió la experimentación con la que se puso a prueba el método OvS propuesto para sintonizar y mejorar la predicción con el simulador Ezeiza, identificando los puntos de fortaleza y debilidad de la metodología desarrollada como marco de experimentación y validación.
2. Las conversaciones y reuniones con los investigadores del INA ayudaron a determinar los parámetros propios de la metodología implementada, como tolerancias en el cómputo. También fueron decisivas al momento de elegir los parámetros críticos y los rangos valores que cada uno puede tomar sobre el sistema físico del cauce del río, y como suceden sus variaciones a lo largo de cada tramo y en función del comportamiento de los niveles del cauce.
3. El índice IC resulta adecuado para obtener una solución que mejore el rendimiento general de la predicción sobre todo el dominio del río. Podría suceder que se mejoren bastante algunas estaciones y que se empeore el pronóstico en otras, ya que el índice mide promedios en todo el dominio del río. Para ofrecer un objetivo que contemple esta situación se consideró un índice combinado como $IC+IM$ para obtener la mejora sobre la mayor cantidad posible de estaciones, pero desechando

aquellas configuraciones que empeoran alguna estación por encima de la tolerancia establecida a priori. La solución hallada para este objetivo múltiple representa las configuraciones que son Pareto soluciones.

4. Esta metodología fue implementada pensando en aportar una mejora al sistema de pronósticos diarios dados por el conjunto de aplicaciones a la que pertenece Ezeiza. Se logró una mejora en la predicción de Ezeiza entre el 10 % y el 25 % para el primer caso de comprobación del método y mejoras entre el 20 % y el 40 % para el segundo caso de experimentación. Estos porcentajes representan un promedio de 40cm a 60cm de altura del agua, y la importancia de estos valores radica en que pueden ser la diferencia entre la pérdida o no del ganado o de cubrir casas, al entrar el agua en las planicies de inundación afectadas.
5. En la prueba de concepto, se llegó a mejorar tres ciudades del dominio del río, usando tres diferentes conjuntos de valores de ajuste. Estos resultados fueron la base para continuar con la investigación de manera de poder mejorar más de una ciudad a la vez. En los casos A y B, se mejoran 3 y 4 ciudades respectivamente pero con el mismo escenario, el cual surgió de la configuración óptima que devolvió la segunda etapa del método OvS.
6. En todo momento se hace referencia al escenario INA. Este escenario es necesario para tener una referencia que permita decidir si esta metodología ofrece una mejora a la predicción. El INA usa el conjunto de variables de decisión que resultaron de la calibración hecha al modelo hidrodinámico del río Paraná en 2011. Los datos simulados que se obtienen en este trabajo, con Ezeiza, son contrarrestados con los datos simulados con el escenario INA bajo las mismas condiciones de simulación, para decidir si fue posible o no sintonizar Ezeiza con la solución encontrada.
7. Se puede decir que la metodología de OvS para sintonizar Ezeiza es un método acotado por el volumen de los datos y no por el cómputo. Esto significa que el método de OvS no está limitado por el tiempo de ejecución del simulador, que es bajo, sino por el tiempo que surge de la implementación de método de OvS en sus dos etapas. Esta situación fue manejada mediante los aportes de HPC. La programación en paralelo se hizo en un modo Maestro-Trabajador con balanceo de carga.
8. El método de optimización ofreció una beneficiosa ganancia general del método computacional. Gracias a la reducción del espacio de búsqueda se obtuvo una ga-

nancia del método de Monte Carlo que potencia la ganancia obtenida por el uso de HPC. Esto resulta en un importante beneficio general del método.

9. El óptimo hallado es una solución relativa a la zona que se desea mejorar. Al crear el espacio de búsqueda y decidir sobre los posibles valores de los parámetros de entrada es necesario tomar unos pocos valores de cada uno. Estas variables de decisión son variables continuas (de tipo real), y dentro de un intervalo pueden tomar infinitos valores. Se seleccionan unos pocos valores sobre el intervalo de cada una, para combinar luego una con otra y determinar así el tamaño del espacio de búsqueda.
10. Las estaciones que obtuvieron mejoras con la solución encontrada con el método de OvS, están ubicadas en la misma zona de donde fueron seleccionadas las secciones cuyos parámetros constituyen las variables de decisión de la optimización. Aunque decidir qué caso será el más conveniente para utilizar con el pronóstico brindado por Ezeiza debe ser decidido por los Ingenieros que analizan la situación del río Paraná cada día.
11. Cuando se eligieron secciones dispersas sobre todo el dominio del río Paraná no se pudo sintonizar la simulación, ya que no se encontró una configuración que mejore la predicción respecto al INA, y generalmente se empeoró mucho más. La realidad es que se utilizan 3 a 5 secciones para implementar el método de OvS, pues usar más fue impracticable por la dimensión del espacio de búsqueda. Si fuese posible combinar más secciones para cubrir todo el dominio se podría encontrar una mejor solución, algo que no se pudo llevar a la práctica con esta metodología por el elevado e inalcanzable tiempo de cómputo.
12. Se eligió mejorar las estaciones del bajo Paraná, pues están ubicadas en un tramo del río sensible para la toma de decisiones y los pronósticos. Esta zona tiene alto impacto en la región de la Cuenca del Plata, especialmente en la economía y por la alta cantidad de población que tiene.

Por último, es importante dejar en claro que el conjunto de parámetros óptimo que se obtiene con el método de OvS en sus dos etapas, es el mejor para el intervalo de simulación considerado. Por ejemplo, 6 meses en el caso de la prueba de concepto, 1 año en el caso de estudio A y 60 días en el caso de estudio B. Los errores de predicción que se producen en ese lapso están indicando que los períodos deberían acortarse y obtener

sucesivos conjuntos óptimos correspondientes a intervalos de menores de tiempo para ajustar aun más la predicción.

Esta situación se puede ver en la Fig. 4.25, ya presentada en las experimentaciones de la prueba de concepto en 4.5.1. Los errores más significativos de la predicción se dan entre el día 15 y 50 del hidrograma. Una de la soluciones a aplicar a esta situación podría ser disminuir los intervalos de tiempo en el que se desarrolla la simulación, por ejemplo de 10 días como se muestra en la gráfica de Goya.

Este es un tema propuesto en las líneas futuras de esta tesis, en el capítulo siguiente.

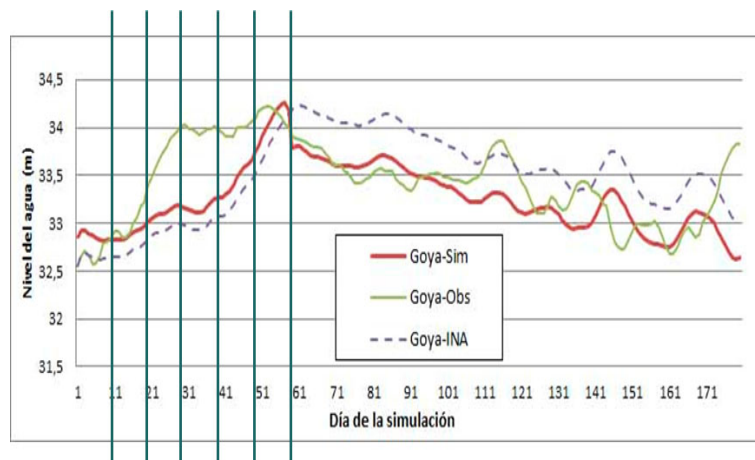


Figura 4.25: Mejora lograda a lo largo de la serie temporal en la ciudad de Goya.

Capítulo 5

Conclusiones y Líneas de Trabajo Futuras

Esta tesis planteó como objetivo principal *brindar una metodología para mejorar la calidad de la simulación de un modelo computacional de cauce de ríos, aplicando la técnica de optimización via simulación.*

Se definió un proceso de investigación guiado por objetivos específicos, a través de los cuales se transitó por diferentes niveles de profundidad del conocimiento, tanto sobre el sistema físico con el que se debería tratar como sobre el problema al que se brindaría una mejora. Esto fue posible desde los métodos y técnicas de la ciencia de la computación y el cómputo de alto rendimiento, pudiendo así dar cumplimiento al objetivo planteado.

A continuación, se enumeran los logros alcanzados con este proceso:

- Se definió un marco teórico sobre la modelación y simulación de procesos físicos para delimitar las características de los simuladores computacionales aplicados a simular el transporte de agua en el cauce de ríos de llanura, con el objetivo de conocer su dinámica en función de la cantidad de agua que recibe durante la evolución de la simulación y utilizar esta información para predecir los efectos del aumento o la disminución de los niveles del agua. En especial para predecir desbordes en los valles de inundación. Este estudio incluyó una reseña de simuladores con diferentes aplicaciones a los fenómenos hidrológicos.
- Se presentó una completa caracterización del modelo físico del cauce del río Paraná y la justificación por la que se eligió este sistema y el simulador Ezeiza, desarrollado en el INA, Argentina. Se brindaron detalles de este río como integrante principal de la

Cuenca del Plata y de su impacto en la zona de influencia cuando los niveles del agua no son los normales. Este análisis permitió comprobar los parámetros más sensibles que caracterizan a este sistema y la incertidumbre de los datos que impactan en los resultados de la simulación.

- Uno de los aportes más interesantes de esta tesis fue proveer de una metodología que permitió sintonizar un simulador, como Ezeiza, de manera automática. El interés de este aporte radica en que una calibración de todo el sistema físico, necesita de tomar datos de campo en ubicaciones específicas de río y estos datos no siempre son fáciles de obtener por una persona, incluso algunos no son posible de medir. Los valores de estos datos suelen modificarse constantemente y de manera natural, por lo cual la incertidumbre sigue siempre presente especialmente al estimar los parámetros que son claves en la predicción. Por estas razones, una calibración automática y una sintonización del simulador constituyeron aportes útiles a esta área del conocimiento, pensando que esta metodología podría aplicarse con otros modelos de simulación hidrológicos.
- La optimización, a través de la metodología de simulación propuesta, está basada en una estructura de vecindad con el objetivo de reducir la región de búsqueda. La metodología está constituida de dos etapas. La primera es una etapa de exploración global sobre todo el espacio de búsqueda con el objetivo de identificar regiones prometedoras para encontrar la solución. La reducción de la región de factibilidad se logra mediante una heurística basada en el método Monte Carlo más el método de K-means, lo cual fue posible gracias a una estructura de vecindad cuya información se fue modificando a medida que avanzó el proceso de optimización. La segunda etapa consistió en la búsqueda de la mejor solución, ya sea la solución óptima o una sub-óptima, encontrada en la frontera de Pareto mediante la realización de una *búsqueda exhaustiva reducida* en tales regiones prometedoras.
- La evaluación de la metodología propuesta se hizo mediante la implementación de dos casos de simulación correspondientes a dos periodos utilizados por el INA para hacer sus calibraciones, el año 1999 y dos meses del año 2007. En ambos casos se mejoró el pronóstico en las ciudades ubicadas principalmente en el tramo inferior del río Paraná. La mejora de la predicción en estas ciudades (estaciones) obedeció a que se utilizaron secciones ubicadas en esa misma región para obtener las variables de decisión a combinar. También se comprobó que no resultó satisfactorio combinar

los parámetros de secciones distribuidas al azar a lo largo de todo el cauce del río. Se deben restringir a la zona elegida para mejorar, aprovechando que fue posible encontrar una solución que no perjudicó la predicción en las estaciones localizadas en las zonas restantes.

- La comprobación de los objetivos propuestos, permitió llegar a porcentajes de mejora hasta el 10 % para las ciudades Santa Fe, San Martín y Rosario, en el primer caso de simulación. Se obtuvieron mejoras entre el 10 % y el 40 % para las ciudades Hernandarias, Paraná, Santa Fe y San Martín, en el segundo caso de simulación. Estos porcentajes corresponden a los ajustes logrados en los niveles del agua respecto a los datos observados y en relación a los ya obtenidos por el INA. Es importante destacar que estos valores se refieren a medidas en las alturas que van desde 40cm a 60cm de diferencia, y pueden ser la diferencia entre la pérdida o no del ganado, de cubrir o no vías terrestres de comunicación, o de tapar o no por completo una casa, ya que esa diferencia de nivel se mide en más o menos planicie de inundación afectada.
- El óptimo hallado es una solución dependiente de la zona que se eligió para mejorar. Las variables de decisión del problema son variables continuas (de tipo real), y dentro de su intervalo de validez pueden tomar infinitos valores. Por lo cual, la solución óptima es una solución propia de este espacio acotado determinado antes de poner en práctica el método de optimización por simulación. Pensar en un óptimo de todo el sistema, como el río Paraná en sus 1900 km. con 76 secciones con parámetros ajustables, aun es una misión imposible de llevar a la práctica.
- La solución óptima que devuelve el método es un conjunto de sus variables de decisión, que definen el valor de los parámetros de entrada. Como se demostró, según la función objetivo se obtienen soluciones distintas al problema de sintonización del simulador. En un caso se resigna lograr un porcentaje de ajuste menor de los datos simulados para mejorar más cantidad de estaciones a la vez. No se debe dejar de tener en cuenta que se está ante un problema real y cada solución deberá ser evaluada en todo momento por los especialistas que toman decisiones con esta información, y serán ellos quiénes decidirán cuál es la mejor y cómo usar la respuesta del método de optimización.

- Uno de los objetivos es hallar una respuesta óptima en tiempos de espera aceptables, es por eso que la puesta en práctica de esta metodología fue posible en todo momento gracias a las técnicas del procesamiento en paralelo y el aprovechamiento de los recursos del cómputo de alto rendimiento. Ambos recursos fueron una herramienta invaluable en este trabajo. La prueba de concepto, con su simulación paramétrica en paralelo fue implementada en modo Maestro-Trabajador (MW) en un cluster de multiprocesadores, utilizando pthreads para ejecutar el simulador Ezeiza. Los programas creados para llevar adelante la experimentación y el análisis de las bases de datos resultantes fueron programados en lenguaje C++, con balanceo de carga. El método MC fue implementado en C++, usando GSL-GNU. La búsqueda exhaustiva acotada sobre la región de factibilidad reducida se implementó también en modo MW.
- De las experiencias llevadas adelante se puede deducir que mientras más recursos de cómputo se tienen a disposición se podría disminuir a gusto el tiempo de una búsqueda exhaustiva de la solución óptima. Aunque, uno de los propósitos de esta tesis es aportar una metodología que mejore la predicción con el mínimo de recursos computacionales posible. Como se demostró, la optimización ofreció una beneficiosa ganancia general en dos niveles: Primero una reducción del espacio de búsqueda que brinda una ganancia en cantidad de simulaciones, o ganancia del método de Monte Carlo. Segundo la ganancia obtenida por el uso de HPC. Ambas se potencian y brindan la ganancia global del método.

5.1. Líneas futuras

Los resultados presentados muestran que la metodología da buenos resultados y que el objetivo general se pudo cumplir. Sin embargo, aun quedan puertas abiertas para la continuación de este trabajo, como también la posibilidad de abrir nuevas líneas de trabajo derivadas de esta tesis, como ser:

- Estudiar la posibilidad de definir intervalos de validez de la solución encontrada, de manera de establecer porcentajes de mejora al utilizar los parámetros ajustados que devuelve el método de OvS, para predecir el futuro del sistema. Este análisis dependerá de estar en momentos de bajante, de subida o de condiciones normales de los niveles de agua del río. Al detectar nuevos desajustes en la simulación, es posible

repetir la optimización via simulación y recalibrar el modelo para obtener una nueva solución con nuevos parámetros ajustados cada vez que sea necesario. Esta propuesta permitiría crear un entorno automático de calibración del simulador.

- Estudiar la factibilidad de aplicación del método al considerar más variables de ajuste en conjunto. Esta posibilidad requiere hacer un estudio de la complejidad del método por las características del problema de optimización y de estimar los límites de los recursos de cómputo de alto rendimiento disponibles para llevar el método a la práctica.
- Considerar nuevas de funciones de objetivos múltiples, que respondan a las necesidades de toma de decisiones con el sistema, y estudiar la posibilidad de obtener la solución de Pareto para estos nuevos objetivos.
- Agregar a la metodología de optimización un análisis de sensibilidad estadística de las variables de ajuste, considerando los errores numéricos del modelo.
- Estudiar el comportamiento del método de optimización al aplicar otras heurísticas combinadas que permitan también disminuir el espacio de búsqueda o aplicando nuevas decisiones inteligentes al elegir un nuevo candidato de dicho espacio, testarlo e ir convergiendo a la solución.
- Utilizar esta metodología con otro simulador de cauce de ríos, con otro modelo físico distinto al río Paraná, o con un simulador de dos dimensiones (2D). Utilizar este tipo de simulador computacional e interpretar sus resultados es un esfuerzo considerable para un investigador que no es del área de hidráulica computacional, por lo cual la colaboración con especialistas en el esta área del conocimiento es condición necesaria para llevar la investigación por buenos caminos. Otro tema que hace imprescindible tener esta colaboración, es la interpretación correcta de los resultados y la necesidad de acceder a los datos observados del fenómeno físico que se está estudiando.
- Estudiar la posibilidad de implementar una solución al problema de OvS, utilizando las imágenes del área de las planicies de inundación para encontrar parámetros de ajuste de la predicción, en los casos de ondas de crecidas con inundación.

Apéndices

Apéndice A

Condiciones de Borde del Modelo del Paraná

Para el río Paraná se ha utilizado como condición de borde de aguas arriba la serie de caudales medios diarios erogados en la presa de Yacyretá, mientras que aguas abajo fue utilizada como condición de borde la serie de alturas hidrométricas diarias registradas en la localidad de Villa Constitución. La información relevada y utilizada para la validación del modelo que realizó el INA en 2011, consistió en la serie de niveles hidrométricos observados en cada estación de seguimiento.

Para cada caso de experimentación que fue considerado en este trabajo fue necesario contar con las condiciones de borde del modelo para cada día de la simulación. A continuación se muestran los gráficos correspondientes al periodo del Caso A, que abarca del 1 de enero de 1999 hasta el 31 de diciembre de 1999.

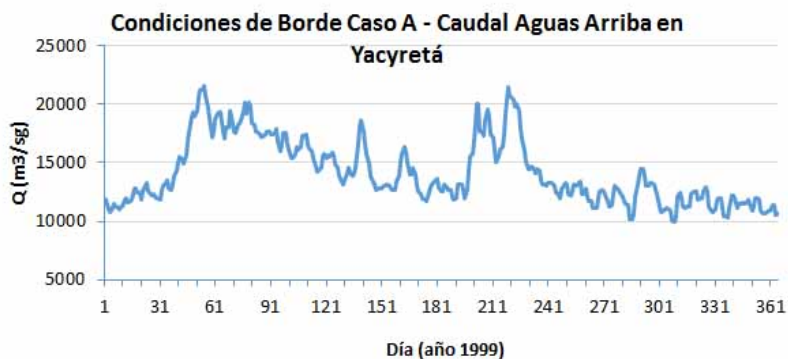


Figura A.1: Caudales erogados por Yacyretá - 1/1/1999 al 31/12/1999

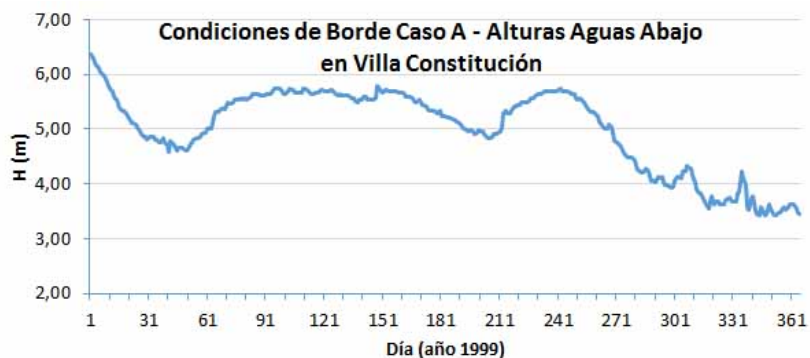


Figura A.2: Alturas en Villa Constitución - 1/1/1999 al 31/12/1999

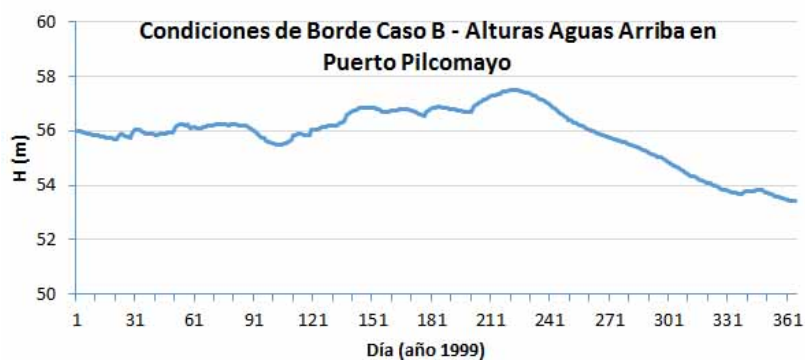


Figura A.3: Alturas en Puerto Pilcomayo - 1/1/1999 al 31/12/1999

A continuación se muestran los gráficos correspondientes al periodo del Caso B, que abarca del 1 de febrero de 2007 hasta el 30 de abril de 2007.

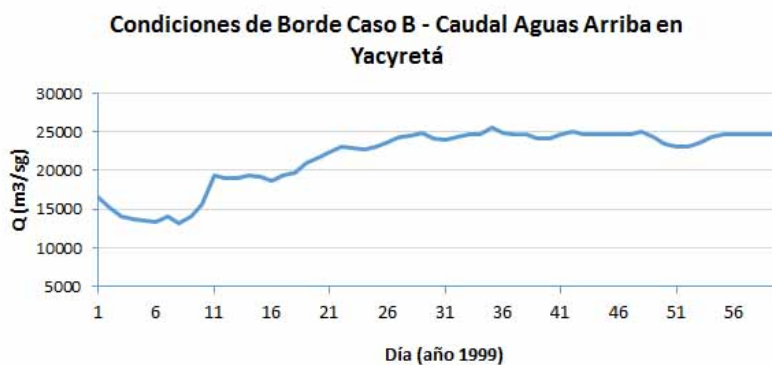


Figura A.4: Caudales erogados por Yacyretá - 1/1/1999 al 31/12/1999

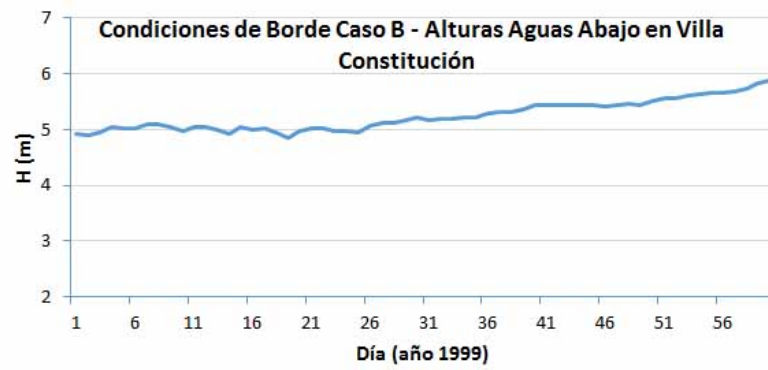


Figura A.5: Alturas en Villa Constitución - 1/1/1999 al 31/12/1999

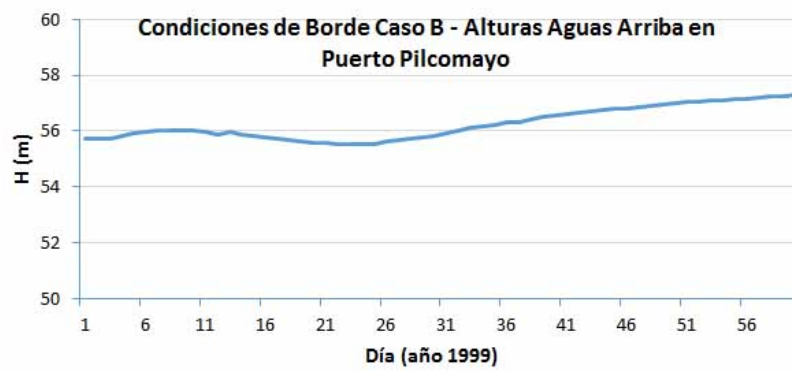


Figura A.6: Alturas en Puerto Pilcomayo - 1/1/1999 al 31/12/1999

Apéndice B

Modelo Numérico de Ezeiza

El software EZEIZA V es un sistema computacional para el cálculo de la traslación de ondas en ríos y canales, desarrollado en el INA desde la década de 1970.

EZEIZA V se basa en la resolución de las ecuaciones de Saint Venant, que constituyen la representación matemática de los principios de conservación de la Mecánica: masa (continuidad) y cantidad de movimiento (longitudinal):

$$\frac{\partial \Omega}{\partial t} + \frac{\partial Q}{\partial x} = q$$

$$\frac{\partial q}{\partial t} + \frac{\partial}{\partial x} \left(\frac{Q^2}{\Omega} + \frac{g}{2} \text{Sigma} \right) = g\Omega(I_0 - I_f) + \frac{g}{2}\gamma + q \cdot u_L$$

donde t es la coordenada temporal, x la coordenada espacial, Ω es el área de la sección transversal al escurrimiento, Q es caudal, q es la descarga lateral entrante por unidad de longitud, u_L es la componente de la velocidad de la descarga lateral entrante en la dirección del escurrimiento, g la aceleración de la gravedad, $S = 2 \cdot h_{ce} \cdot \Omega$ el momento areal (h_{CE} es la profundidad del centroide de la sección), γ el coeficiente de divergencia ($\partial/\Sigma|_{h=cte}$), h el tirante, I_0 la pendiente del fondo, y I_f la pendiente de fricción.

Las ecuaciones anteriores forman un sistema hiperbólico de ecuaciones diferenciales en derivadas parciales, que representan la evolución de ondas superficiales en aguas poco profundas. El sistema de ecuaciones diferenciales está complementado por condiciones iniciales y de contorno. En régimen subcrítico, que es el caso típico de una red de canales, se especifican como condiciones de borde los niveles o caudales instantáneos en los contornos de ingreso al sistema y los niveles o las relaciones altura-caudal en las desembocaduras.

Como condiciones iniciales, deben proveerse las distribuciones de nivel y caudal sobre todo el sistema. Las ecuaciones se resuelven numéricamente, para régimen subcrítico, por el esquema implícito en diferencias finitas de Preissmann. En la unión de dos o más canales de la red se impone un balance de masa y de energía mecánica. Las planicies de inundación se incorporan a través de la consideración de cauces de sección compuesta. El modelo es capaz de simular el flujo a través de estructuras tales como puentes, alcantarillas, vertederos, etc., considerándolas como singularidades y tratándolas separadamente.

La versión más avanzada, EZEIZA V, permite el tratamiento de una red de flujo arbitraria.

Referencias

- Abbass, H., Sarker, R., y Newton, C. (2001). Pde: a pareto-frontier differential evolution approach for multi-objective optimization problems. En *Proceedings of the 2001 congress on evolutionary computation, 2001* (Vol. 2, p. 971-978). doi: 10.1109/CEC.2001.934295
- Abdalhaq, B., Cortés, A., Margalef, T., Bianchini, G., y Luque, E. (2006). Between classical and ideal: enhancing wildland fire prediction using cluster computing. *Cluster Computing*, 9(3), 329–343.
- Andradóttir, S. (2006, octubre). Simulation optimization with countably infinite feasible regions: Efficiency and convergence. *ACM Trans. Model. Comput. Simul.*, 16(4), 357–374. Descargado de <http://doi.acm.org/10.1145/1176249.1176252>
- Arcement, G. J., y Schneider, V. R. (1989). *Guide for selecting manning's roughness coefficients for natural channels and flood plains*. US Government Printing Office Washington, DC, USA. Descargado de http://ponce.sdsu.edu/usgs_report_2339.pdf
- Arthur, D., Manthey, B., y Roglin, H. (2009, Oct). K-means has polynomial smoothed complexity. En *Foundations of computer science, 2009. focs '09. 50th annual ieee symposium on* (p. 405-414). doi: 10.1109/FOCS.2009.14
- Balica, S., Popescu, I., Beevers, L., y Wright, N. (2013). Parametric and physically based modelling techniques for flood risk and vulnerability assessment: A comparison. *Environmental Modelling & Software*, 41, 84-92. doi: <http://dx.doi.org/10.1016/j.envsoft.2012.11.002>
- Barton, R. R., y Meckesheimer, M. (2006). Metamodel-based simulation optimization. En S. G. Henderson y B. L. Nelson (Eds.), *Simulation* (Vol. 13, p. 535 - 574). Elsevier. Descargado de <http://www.sciencedirect.com/science/article/pii/S0927050706130182> doi: [http://dx.doi.org/10.1016/S0927-0507\(06\)13018-2](http://dx.doi.org/10.1016/S0927-0507(06)13018-2)

- Bianchini, G., Denham, M., Cortés, A., Margalef, T., y Luque, E. (2006). Improving forest-fire prediction by applying a statistical approach. *Forest Ecology and Management*(234), S210.
- Bladé, E., Gómez-Valentín, M., Dolz, J., Aragón-Hernández, J., Corestein, G., y Sánchez-Juny, M. (2012, junio). Integration of 1d and 2d finite volume schemes for computations of water flow in natural channels. *Advances in Water Resources*, 42, 17–29. Descargado 2015-06-05, de <http://linkinghub.elsevier.com/retrieve/pii/S0309170812000760> (00021) doi: 10.1016/j.advwatres.2012.03.021
- Blum, C., y Roli, A. (2003, septiembre). Metaheuristics in combinatorial optimization: Overview and conceptual comparison. *ACM Comput. Surv.*, 35(3), 268–308. Descargado de <http://doi.acm.org/10.1145/937503.937505> doi: 10.1145/937503.937505
- Bruballa, E., Taboada, M., Cabrera, E., Rexachs, D., y Luque, E. (2014, Aug). Simulation and big data: A way to discover unusual knowledge in emergency departments: Work-in-progress paper. En *Future internet of things and cloud (ficloud), 2014 international conference on* (p. 367-372). doi: 10.1109/FiCloud.2014.65
- Brugnach, M., y Pahl Wostl, C. (2008). A broadened view on the role for models in natural resource management: Implications for model development. En *Adaptive and integrated water management*. Springer Berlin Heidelberg.
- Cabrera, E., Luque, E., Taboada, M., Epelde, F., y Iglesias, M. (2012). Optimization of emergency departments by agent-based modeling and simulation. En *Information reuse and integration (iri), 2012 ieee 13th international conference on* (p. 423-430). doi: 10.1109/IRI.2012.6303040
- Cabrera, E., Taboada, M., Iglesias, M. L., Epelde, F., y Luque, E. (2012). Simulation optimization for healthcare emergency departments. *Procedia CS*, 9, 1464-1473.
- Camilloni, I., y Barros, V. (2003). Extreme discharge events in the paraná river and their climate forcing. *Journal of Hydrology*, 278(1-4), 94 - 106. Descargado de <http://www.sciencedirect.com/science/article/pii/S0022169403001331> doi: [http://dx.doi.org/10.1016/S0022-1694\(03\)00133-1](http://dx.doi.org/10.1016/S0022-1694(03)00133-1)
- Camilloni, I., Saurral, R., y Montroull, N. (2013). Hydrological projections of fluvial floods in the uruguay and paraná basins under different climate change scenarios. *Inter-*

national Journal of River Basin Management, 11(4), 389-399. Descargado de <http://dx.doi.org/10.1080/15715124.2013.819006> doi: 10.1080/15715124.2013.819006

Capocchi, L., y Santucci, J. F. (2013). Discrete optimization via simulation of catchment basin management within the devsimpy framework. En *Proceedings of the 2013 winter simulation conference: Simulation: Making decisions in a complex world* (pp. 205–216). Piscataway, NJ, USA: IEEE Press. Descargado de <http://dl.acm.org/citation.cfm?id=2675983.2676007>

Chong, E. K. P., y Zak, S. H. (2013). *An introduction to optimization* (Vol. 76; J. Wiley y I. Sons, Eds.). Wiley. Descargado de <https://books.google.com.ar/books?id=iD5s0iKXHP8C>

Coulthard, T. J., y Van De Wiel, M. J. (2012). Modelling river history and evolution. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 370(1966), 2123–2142. doi: 10.1098/rsta.2011.0597

Cunge, J. A., Holly, F. M., y Verwey, A. (1980). *Practical aspects of computational river hydraulics*. London: Pitman Publishing.

Davidon, W. C. (1991). Variable metric method for minimization. *SIAM Journal on Optimization*, 1(1), 1-17. doi: <http://dx.doi.org/10.1137/0801001>

Deb, K. (2014). Multi-objective optimization. En E. K. Burke y G. Kendall (Eds.), *Search methodologies* (p. 403-449). Springer US. Descargado de http://dx.doi.org/10.1007/978-1-4614-6940-7_15 doi: 10.1007/978-1-4614-6940-7_15

Denning, P. (2010). The great principles of compu. *Computing Science*, 98(5), 369-372.

Dilley, M., Chen, R., Deichmann, U., Lerner-Lam, A., y Arnold, M. (2005). *Natural disaster hotspots. a global risk analysis* (Vol. 5). ISBN 0-8213-5930-4: The World Bank and Columbia University.

Fermi, E., y Metropolis, N. (1952). *Los alamos unclassified report ls-1492* (Inf. Téc.). Los Alamos National Laboratory, Los Alamos.

Finaud-Guyot, P., Delenne, C., Guinot, V., y Llovel, C. (2011). 1d-2d coupling for river flow modeling. *Comptes Rendus Mécanique*, 339(4), 226 - 234. Descargado de <http://www.sciencedirect.com/science/article/pii/S163107211100026X> doi: <http://dx.doi.org/10.1016/j.crme.2011.02.001>

- Fletcher, R. (2013). *Practical methods of optimization*. Wiley. Descargado de https://books.google.com.ar/books?id=_WuAvIx0EE4C
- Floudas, C., y Gounaris, C. (2009). A review of recent advances in global optimization. *Journal of Global Optimization*, 45(1), 3-38. Descargado de <http://dx.doi.org/10.1007/s10898-008-9332-8> doi: 10.1007/s10898-008-9332-8
- Fu, M. C. (2002). Feature article: Optimization for simulation: Theory vs. practice. *INFORMS Journal on Computing*, 14(3), 192-215. Descargado de <http://dx.doi.org/10.1287/ijoc.14.3.192.113> doi: 10.1287/ijoc.14.3.192.113
- Gallopoulos, E., y Sameh, A. (2002). One viewpoint on computational sscomp. cse: Content and product. *Computational Science & Engineering, IEEE*, 4(2), 39-43. doi: 10.1109/99.609830
- Glover, F., Kelly, J. P., y Laguna, M. (1996). New advances and applications of combining simulation and optimization. En J. M. Charnes, D. J. IVlorrice, D. T. Brunner, y J. J. Swain (Eds.), *Proceedings of the 28th conference on winter simulation* (pp. 144–152).
- Group, T. . S. (2005). *The towards 2020 science workshop*. The Towards 2020 Science workshop. Descargado de http://research.microsoft.com/en-us/um/cambridge/projects/towards2020science/background_2020.htm
- Hansson, K., Danielson, M., y Ekenberg, L. (2008). A framework for evaluation of flood management strategies. *Journal of Environmental Management*(86), 465-480.
- Hartikainen, M., Miettinen, K., y Wiecek, M. M. (2012). Paint: Pareto front interpolation for nonlinear multiobjective optimization. *Computational optimization and applications*, 52(3), 845–867.
- Hey, T., Gannon, D., y Pinkelman, J. (2012). The future of data-intensive science. *Computer*, 45(5), 81 - 82. doi: 10.1109/MC.2012.181
- Jacovkis, P. (1990). Modelos numéricos hidrodinámicos en redes fluviales complejas. *Revista internacional de métodos numéricos para cálculo y diseño en ingeMecánica- UPC. Barcelona*, 6(4), 543-572. Descargado de <http://hdl.handle.net/2099/7594>

- Jaime, P., y Menéndez, A. (1997). *Modelo hidrodinámico del río paraná desde yacyretá hasta la ciudad de paraná* (Inf. Téc. n.º Informe LHA 01-165-97). Instituto Nacional del Agua - Secretaría de Desarrollo Sustentable - Rep. Argentina. Descargado de <http://50.30.34.57/~modelacion/es/publicaciones/informes-tecnicos>
- Jonkman, S., y Vrijling, J. (2008). Loss of life due to floods. *Journal of Flood Risk Management*, 1(1), 43–56. Descargado de <http://dx.doi.org/10.1111/j.1753-318X.2008.00006.x> doi: 10.1111/j.1753-318X.2008.00006.x
- Kolda, T. G., Lewis, R. M., y Torczon, V. (2003). Optimization by direct search: New perspectives on some classical and modern methods. *SIAM review*, 45(3), 385–482. Descargado de <http://epubs.siam.org/doi/abs/10.1137/S003614450242889>
- Kroese, D. P., Taimre, T., y Botev, Z. I. (2013). *Handbook of monte carlo methods* (Vol. 706). John Wiley & Sons.
- Latessa, G. (2011). *Modelo hidrodinámico del río paraná para pronóstico hidrológico: Evaluación del performance y una propuesta de redefinición geométrica* (Trabajo profesional Ingeniería Civil). Universidad de Buenos Aires - Facultad de Ingeniería. Descargado de <http://50.30.34.57/~modelacion/es/publicaciones/tesis>
- Law, A. M., y Kelton, W. D. (2000). *Simulation modeling and analysis*. McGraw-Hill Higher Education. Industrial Engineering Series.
- Mackay, D. (1998). Introduction to monte carlo methods. En M. Jordan (Ed.), *Learning in graphical models* (Vol. 89, p. 175-204). Springer Netherlands. Descargado de http://dx.doi.org/10.1007/978-94-011-5014-9_7 doi: 10.1007/978-94-011-5014-9_7
- MacKay, D. J. (2003). *Information theory, inference and learning algorithms*. Cambridge university press.
- Mejuto, M., Mayorga, A., y Fernández, S. (1997). La utilización de modelos en hidrología. *Ensayos: Revista de la Educación de Albacete*, 12(305-318).
- Menéndez, A. (1996). Ezeiza v. un programa computacional para redes de canales. *Mecánica computacional*, XVI, 63-71.
- Menéndez, A. (2002). Three decades of development and application of numerical simulation tools at ina hydraulics lab. *Mecánica Computacional*, XXI, 2247-2266. Descargado de <http://cimec.org.ar/ojs/index.php/mc/article/view/1035/982>

Miettinen, K., Ruiz, F., y Wierzbicki, A. P. (2008). Introduction to multiobjective optimization: Interactive approaches. En J. Branke, K. Deb, K. Miettinen, y R. Stowinski (Eds.), *Multiobjective optimization* (Vol. 5252, p. 27-57). Springer Berlin Heidelberg. Descargado de http://dx.doi.org/10.1007/978-3-540-88908-3_2 doi: 10.1007/978-3-540-88908-3_2

Neumaier, A. (2004). Complete search in continuous global optimization and constraint satisfaction. En (Vol. 13, pp. 271-369). Cambridge University Press.

Nocedal, J., y Wright, S. (2006). *Numerical optimization*. Springer. Descargado de <https://books.google.com.ar/books?id=VbHYoSyelFcC>

Oberkampf, W., DeLand, S., Rutherford, B., Diegert, K., y Alvin, K. (2002). Error and uncertainty in modeling and simulation. *Reliability Engineering & System Safety*, 75(3), 333 - 357. Descargado de <http://www.sciencedirect.com/science/article/pii/S095183200100120X> doi: [http://dx.doi.org/10.1016/S0951-8320\(01\)00120-X](http://dx.doi.org/10.1016/S0951-8320(01)00120-X)

Pappenberger, F., Matgen, P., Beven, K. J., Henry, J.-B., Pfister, L., y de, P. F. (2006). Influence of uncertain boundary conditions and model structure on flood inundation predictions. *Advances in Water Resources*, 29(10), 1430 - 1449. Descargado de <http://www.sciencedirect.com/science/article/pii/S0309170805002666> doi: <http://dx.doi.org/10.1016/j.advwatres.2005.11.012>

Park, H.-Y., Datta-Gupta, A., y King, M. J. (2015). Handling conflicting multiple objectives using pareto-based evolutionary algorithm for history matching of reservoir performance. *Journal of Petroleum Science and Engineering*, 125, 48-66. doi: [10.1016/j.petrol.2014.11.006](http://dx.doi.org/10.1016/j.petrol.2014.11.006)

Pedregal, P. (2006). *Introduction to optimization* (Vol. 46; S. S. . B. Media, Ed.). Springer New York. Descargado de <https://books.google.com.ar/books?id=94XfBwAAQBAJ>

Ponce, V. M. (1989). *Engineering hydrology: Principles and practices*. Prentice Hall.

Pujol, A., y Menéndez, A. (1987). *Análisis unidimensional de escurrimiento en canales*. Editorial Universitaria de Buenos Aires.

Ramirez, J. A. (2000). Prediction and modeling of flood hydrology and hydraulics. En E. E. Wohl (Ed.), (p. 293-333). Cambridge University Press: Cambridge, UK.

- Re, M., Badano, N., Lecertúa, E., Re, F., y Menéndez, A. (2008, Noviembre). Modelización matemática de una cuenca de llanura extensa. *Mecánica Computacional*, 27, 351-368.
- Re, M., y Menéndez, A. (2007). Impacto del cambio climático en las costas del río de la plata. *Revista Internacional de Desastres Naturales, Accidentes e Infraestructura Civil - Open Journal System*, 7(1). Descargado de <http://academic.uprm.edu/laccei/index.php/RIDNAIC/article/view/121>
- Refsgaard, J. C., van der Sluijs, J. P., HÃjberg, A. L., y Vanrolleghem, P. A. (2007). Uncertainty in the environmental modelling process: A framework and guidance. *Environmental Modelling & Software*, 22(11), 1543 - 1556. Descargado de <http://www.sciencedirect.com/science/article/pii/S1364815207000266> doi: <http://dx.doi.org/10.1016/j.envsoft.2007.02.004>
- Santucci, J. F., y Capocchi, L. (2015, enero). Optimization via simulation of catchment basin management using a discrete-event approach. *Simulation*, 91(1), 43-58. (ISSN:0037-5497) doi: 10.1177/0037549714559608
- Sargent, R. G. (2005). Verification and validation of simulation models. En *Proceedings of the 37th conference on winter simulation* (pp. 130-143). Winter Simulation Conference. Descargado de <http://dl.acm.org/citation.cfm?id=1162708.1162736>
- Shannon, R. E. (1998). Introduction to the art and science of simulation. En *Proceedings of the 30th conference on winter simulation* (pp. 7-14). Los Alamitos, CA, USA: IEEE Computer Society Press. Descargado de <http://dl.acm.org/citation.cfm?id=293172.293175>
- Stevenson, D. (1993). Science. computational science and computer science: At a crossroads. En *Proceedings of the 1993 acm conference on computer science* (p. 7-14). ACM.
- Stewart, M., Bates, P., Anderson, M., Price, D., y Burt, T. (1999). Modelling floods in hydrologically complex lowland river reaches. *Journal of Hydrology*, 223(1), 85 - 106. Descargado de <http://www.sciencedirect.com/science/article/pii/S0022169499001122> doi: [http://dx.doi.org/10.1016/S0022-1694\(99\)00112-2](http://dx.doi.org/10.1016/S0022-1694(99)00112-2)
- Taboada, M., Cabrera, E., Epelde, F., Iglesias, M. L., y Luque, E. (2013). Using an agent-based simulation for predicting the effects of patients derivation policies in emergency departments. *Procedia Computer Science*, 18, 641-650.

- Timbe Castro, L. M., y Willems, P. (2011). Desempeño de modelos hidráulicos 1d y 2d para la simulación de inundaciones. *Maskana. Revista científica de la Universidad de Cuenca*, 2(1), 91-98. Descargado de <http://dspace.ucuenca.edu.ec/handle/123456789/5401>
- Tucci, C., y Collischonn, W. (2006). Predicción de crecidas. *Boletín de la OMM*, 55(3), 179-184.
- Wang, L.-F., y Shi, L.-Y. (2013). Simulation optimization: A review on theory and applications. *Acta Automatica Sinica*, 39(11), 1957 - 1968. Descargado de <http://www.sciencedirect.com/science/article/pii/S1874102913600816> doi: [http://dx.doi.org/10.1016/S1874-1029\(13\)60081-6](http://dx.doi.org/10.1016/S1874-1029(13)60081-6)
- Warmink, J., Van der Klis, H., Booij, M., y Hulscher, S. (2011). Identification and quantification of uncertainties in a hydrodynamic river model using expert opinions. *Water Resources Management*, 25(2), 601-622. Descargado de <http://dx.doi.org/10.1007/s11269-010-9716-7> doi: 10.1007/s11269-010-9716-7
- Wing, J. M. (2006, marzo). Computational thinking. *Communications of ACM*, 49(3), 33-35. Descargado de <http://doi.acm.org/10.1145/1118178.1118215> doi: 10.1145/1118178.1118215
- Yapo, P. O., Gupta, H. V., y Sorooshian, S. (1998). Multi-objective global optimization for hydrologic models. *Journal of Hydrology*, 204(1), 83 - 97. doi: [http://dx.doi.org/10.1016/S0022-1694\(97\)00107-8](http://dx.doi.org/10.1016/S0022-1694(97)00107-8)
- Zare, S. O., Saghafian, B., Shamsai, A., y Nazif, S. (2012). Multi-objective optimization using evolutionary algorithms for qualitative and quantitative control of urban runoff. *Hydrology & Earth System Sciences Discussions*, 9(1), 777-786. Descargado de <http://connection.ebscohost.com/c/articles/71773287/multi-objective-optimization-using-evolutionary-algorithms-qualitative-quantitative-control-urban-runoff>