

Optimization of Distributed Models Using Evolutionary Algorithms and Surrogate Modeling

Juan F. Farfán-Durán¹

Arash Heidari², Esteban Sañudo¹, Carlos Montalvo¹, Ivo Couckuyt², Tom Dhaene²,

Jerónimo Puertas¹, Luís Cea¹

¹ *Universidade da Coruna, Water and Environmental Engineering Group, Center for Technological Innovation in Construction and Civil Engineering (CITEEC)*

² *Faculty of Engineering and Architecture, Ghent University - imec, Ghent, Belgium*

This work was partially funded by the Spanish Ministry of Science and Innovation (MCIN/AEI/10.13039/501100011033) within the project “SATURNO: Early warning against pluvial flooding in urban areas” (PID2020-118368RB-I00).

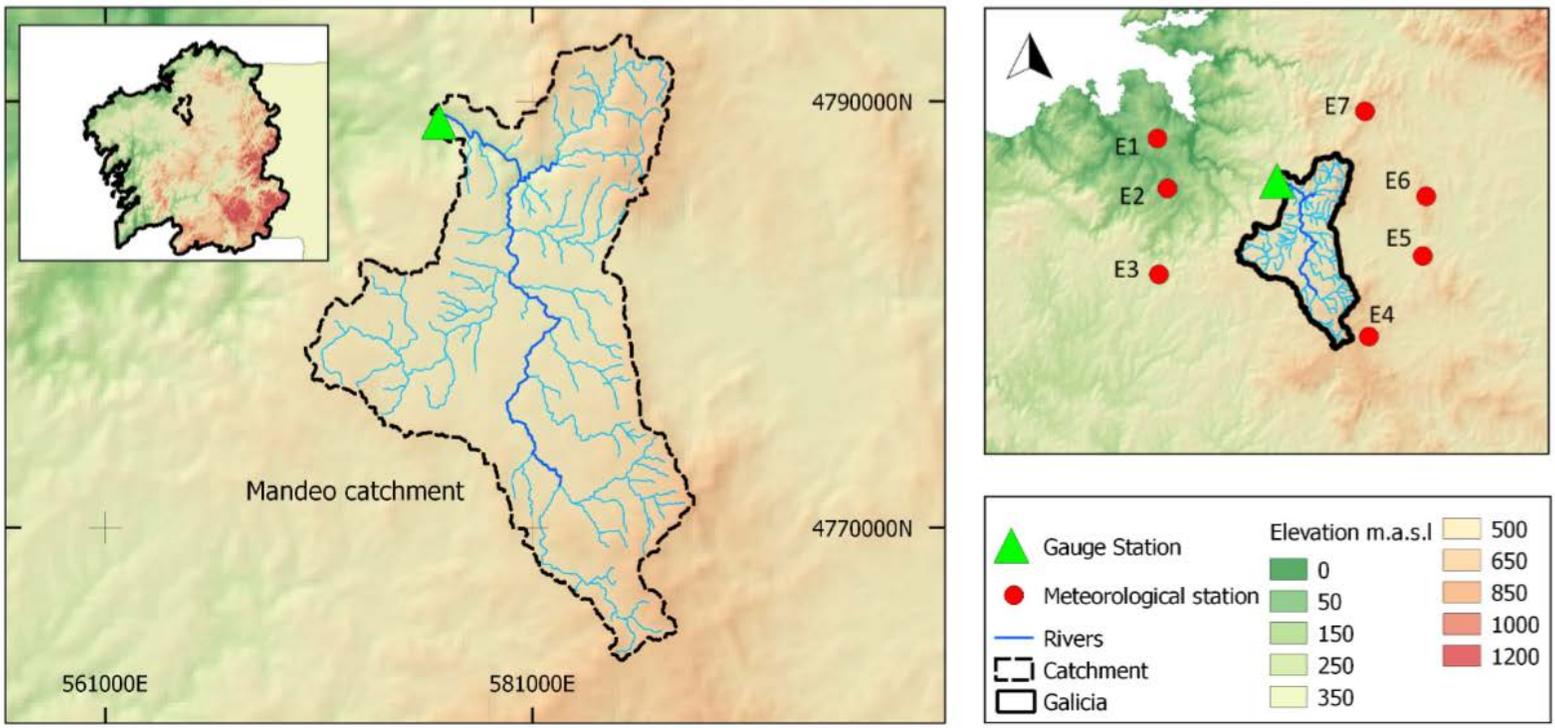
This work has been supported by the Flemish Government under the ‘Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen’ and the ‘Fonds Wetenschappelijk Onderzoek (FWO)’ programmes.

Introducción

- Los modelos hidrológicos son herramientas matemáticas esenciales para simular la respuesta de cuencas a la precipitación.
- Estos modelos pueden ser agregados, semi-distribuidos o distribuidos, cada uno con diferentes niveles de detalle espacial.
- Los modelos distribuidos son cada vez más importantes para estudiar variabilidad en respuestas hidrológicas.
- La calibración de parámetros es crucial para la precisión del modelo, pero es un proceso computacionalmente costoso.
- Los métodos de búsqueda locales y globales tienen limitaciones, incluido el alto costo computacional.
- Los modelos subrogados pueden reducir significativamente el tiempo de cálculo en la calibración de parámetros.
- Se requiere un número significativo de ejecuciones del modelo para entrenar un modelo subrogado preciso.
- Este estudio busca desarrollar un Algoritmo Evolutivo Asistido por un Subrogado (SA-EA) para la calibración eficiente de modelos hidrológicos distribuidos (Iber).

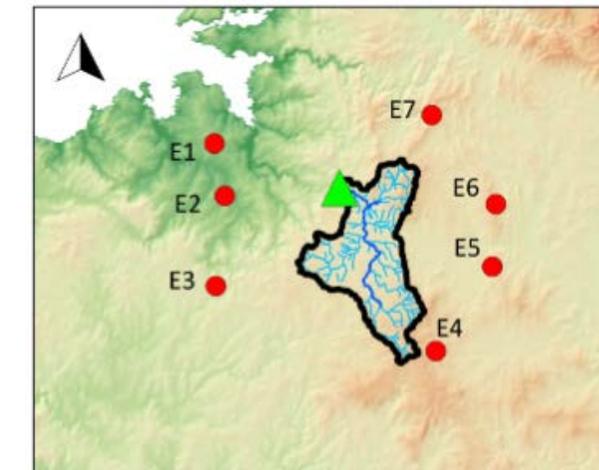
Región de Estudio

- Cuenca río del Mandeo, con un área total de 353 km² y un área modelada de 248 km²
- Cauce principal de aproximadamente 50 km hasta la estación de medición,
- Elevación en la cuenca varía entre 328 m y 810 m sobre el nivel del mar.
- Siete estaciones de lluvia usadas para interpolación en la cuenca.
- Datos históricos escala 10-minutal agregados a 1 hora.



Características de los Eventos.

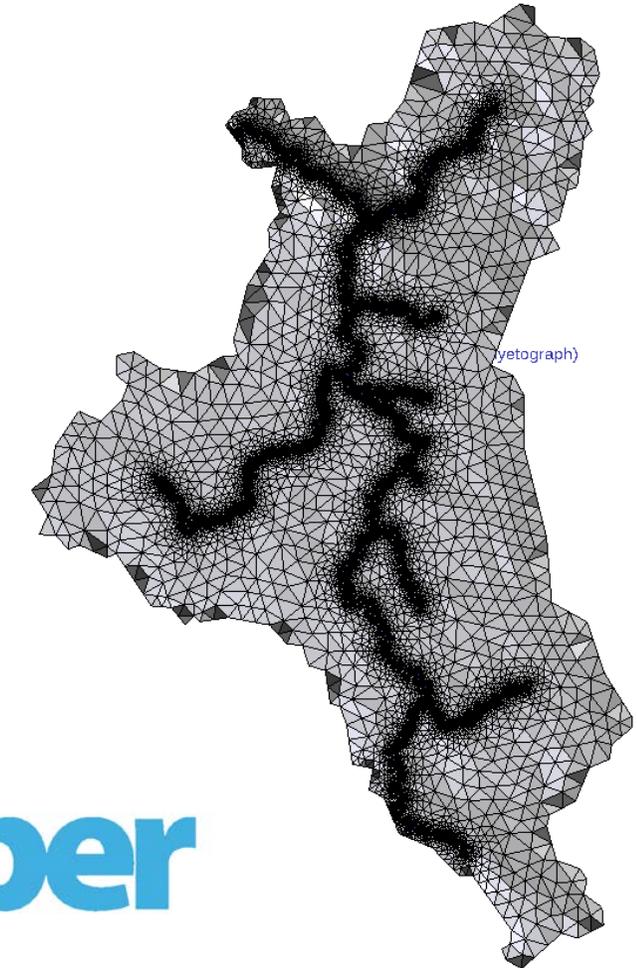
- Metodología evaluada en un evento sintético y tres reales.
- El evento sintético usó correspondiente a un evento con período de retorno de 50 años.
- Se añadieron perturbaciones aleatorias a la precipitación en cada estación.
- Los eventos reales se evaluaron con datos de tormentas de la serie observada



Event	Start	Event Duration (h)	Max 1h intensity (mm/h)	Q max (m3/s)
E ₀	-	42	11.3	201.1
E ₁	05/01/2011 12:00	38	6.1	162.5
E ₂	13/01/2016 00:00	38	10.7	171.3
E ₃	15/02/2018 15:00	50	4.1	82.5

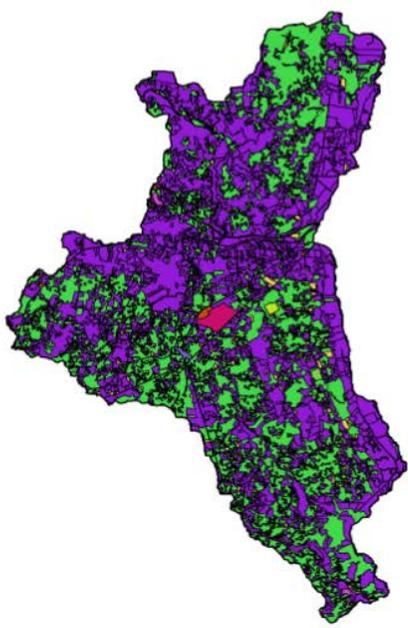
Módulo Hidrológico de Iber

- Incluye implementación mejorada con GPU pero sigue siendo computacionalmente exigente.
- Utiliza un solucionador de volumen finito no estructurado y ha sido validado en trabajos anteriores.
- La malla consta de 200,757 elementos; tamaños máximos establecidos para pendientes y ríos de 100 y 10 metros respectivamente.
- Utiliza un MDT con resolución de 25m del IGN.
- La tasa de infiltración se calcula con el modelo Green-Ampt, que tiene 6 parámetros.

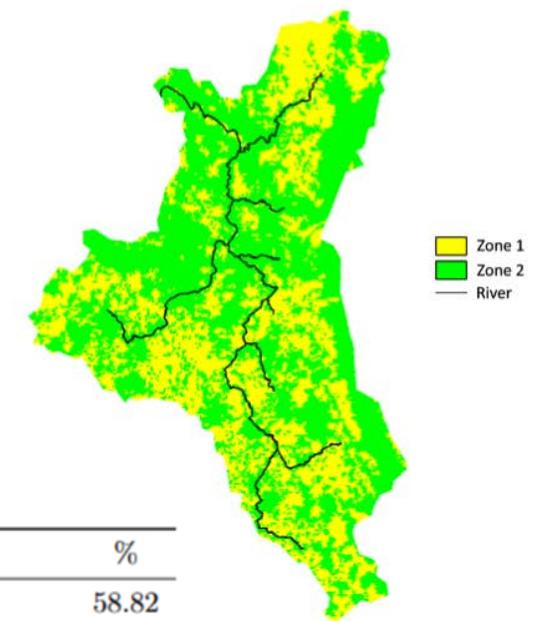
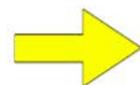


Parametrización del Modelo

- La calibración se basa en zonas distintas según tipos de uso del suelo y parámetros Green-Ampt.
- Datos de uso del suelo provienen del SIOSE español y se clasifican mediante HILUCS.
- Las áreas naturales y la agricultura dominan, ocupando el 58,82% y el 36,52% respectivamente.
- Para evitar sobreparametrización, el modelo se divide en dos zonas; el 4,66% restante se asignó a la zone 2.



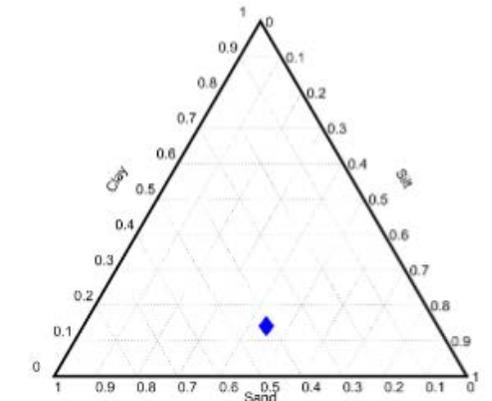
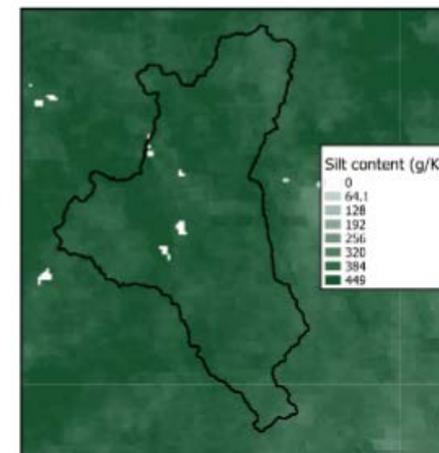
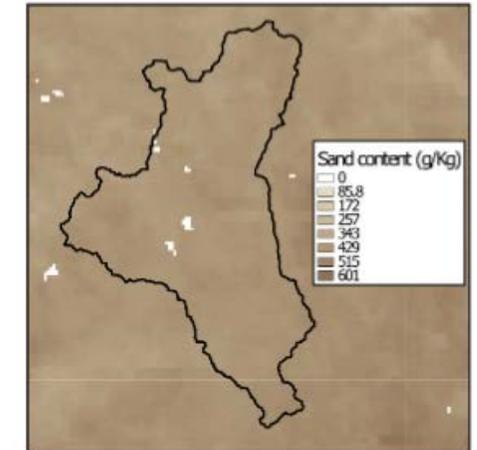
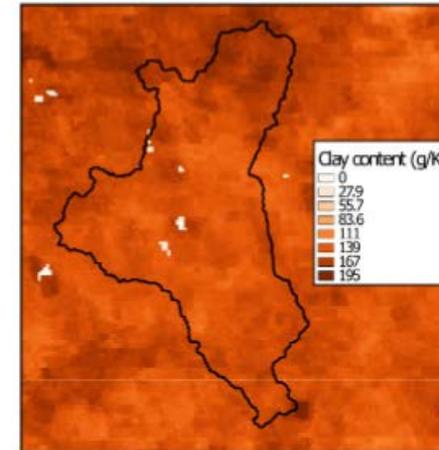
- Natural land areas
- Agriculture
- Forestry
- Residential use
- Transitional areas
- Transport networks
- Abandoned areas
- Secondary production
- Utilities
- Not known use
- Community services
- Mining and quarrying
- Water areas not in other economic use



Land use	%
Natural land areas	58.82
Agriculture	36.52
Forestry	1.92
Residential use	0.76
Transitional areas	0.59
Transport networks	0.45
Abandoned areas	0.16
Secondary production	0.15
Utilities	0.15
Not known use	0.15
Community services	0.14
Mining and quarrying	0.10
Water areas not in other economic use	0.08

Parametrización del Modelo

- Datos de tipo de suelo obtenidos de SoilGrids con resolución espacial de 200 m.
- El estudio sigue la clasificación del USDA; en promedio el suelo contiene 15% arcilla, 43% limo, 42% arena (Triángulo de texturas).
- El suelo se clasifica como marga/loam
- Con esto se puede determinar un rango de parámetros de infiltración de Green-Ampt refiriéndonos a *Handbook of Hydrology (Maidment, 1993)*.
- Proceso similar para rugosidad.
- Rugosidad del río establecida como otro parámetro a optimizar.



Parametrización del Modelo

Zone	Description	Symbol	Units	Min Value	Max Value
Zone 1	Soil suction in zone 1.	Su_1	mm	88	273
	Soil porosity in zone 1.	Por_1	-	0.3	0.5
	Initial soil saturation in zone 1.	Sat_1	-	0.05	0.85
	Saturated hydraulic conductivity of the soil in zone 1.	Ks_1	mm/h	1	4
	Initial losses in zone 1.	$loss_1$	mm	0.1	10
	Depth of soil in zone 1.	$depth_1$	m	0.25	10
	Manning roughness coefficient for the zone 1.	n_1	-	0.012	0.18
Zone 2	Soil suction in zone 2.	Su_2	mm	88	273
	Soil porosity in zone 2.	Por_2	-	0.3	0.5
	Initial soil saturation in zone 2.	Sat_2	-	0.05	0.85
	Saturated hydraulic conductivity of the soil in zone 2.	Ks_2	mm/h	1	4
	Initial losses in zone 2.	$loss_2$	mm	0.1	10
	Depth of soil in zone 2.	$depth_2$	m	0.25	10
	Manning roughness coefficient for the zone 2.	n_2	-	0.012	0.18
River	Manning roughness coefficient for the river	n_{river}	-	0.012	0.18

Funciones Objetivo

- Se utilizaron dos funciones objetivo: Coeficiente de Nash-Sutcliffe (NSE) y Coeficiente de Nash-Sutcliffe Ponderada (WNSE).
- NSE cuantifica el ajuste entre las series de caudal observado y simulado mediante una fórmula específica.
- WNSE modifica NSE ponderando las diferencias según la magnitud de los caudales observados.
- Se usó un exponente $p=1$ en WNSE para dar más peso a los caudales altos.
- Se aplicó el enfoque de *Generalized Likelihood Uncertainty Estimation (GLUE)* para evaluar la incertidumbre predictiva del modelo.
- GLUE identifica sets de parámetros “behavioral” que se ajustan bien a las observaciones.
- Se asignan pesos a cada set para tener en cuenta la incertidumbre en la modelización.
- Cada set de parámetros tiene un valor de función objetivo.
- Calibrar consiste en maximizar NSE y/o WNSE $\in [-\text{inf}, 1]$

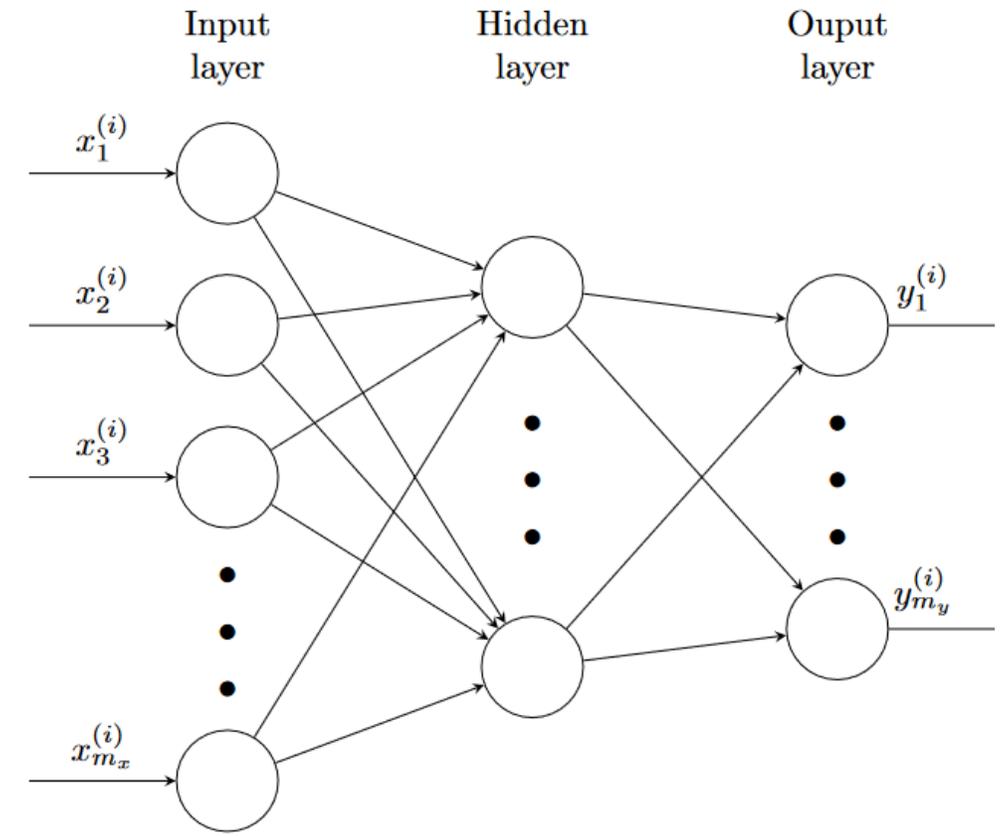
$$NSE = 1 - \frac{\sum_{i=1}^n (Q_{obs,i} - Q_{sim,i})^2}{\sum_{i=1}^n (Q_{obs,i} - \overline{Q_{obs}})^2}$$

$$WNSE = 1 - \frac{\sum_{i=1}^n w_i (Q_{obs,i} - Q_{sim,i})^2}{\sum_{i=1}^n w_i (Q_{obs,i} - \overline{Q_{obs}})^2}$$

$$w_i = \frac{(Q_{obs,i})^p}{\sum_{i=1}^n (Q_{obs,i})^p}$$

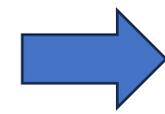
Modelo Subrogado basado en ANN

- **Entrada:** $\mathbf{X} \in \mathbb{R}^{N \times m_x}$: Individuos en la población. En nuestro caso parámetros del modelo.
- **Salida:** $\mathbf{Y} \in \mathbb{R}^{N \times m_y}$: Aptitud de los miembros de la población. Funciones objetivo (NSE, WNSE)
- 2 objective functions: NSE and WNSE
- Para un set de parámetros $\mathbf{x}^{(i)}$, la ANN intenta predecir las funciones objetivo correspondientes $\mathbf{y}^{*(i)}$
- La ANN simula el espacio de soluciones del modelo.
- Entrenamiento: minimizar Error Cuadrático Medio.
- Entrenamiento usa descenso de gradiente.

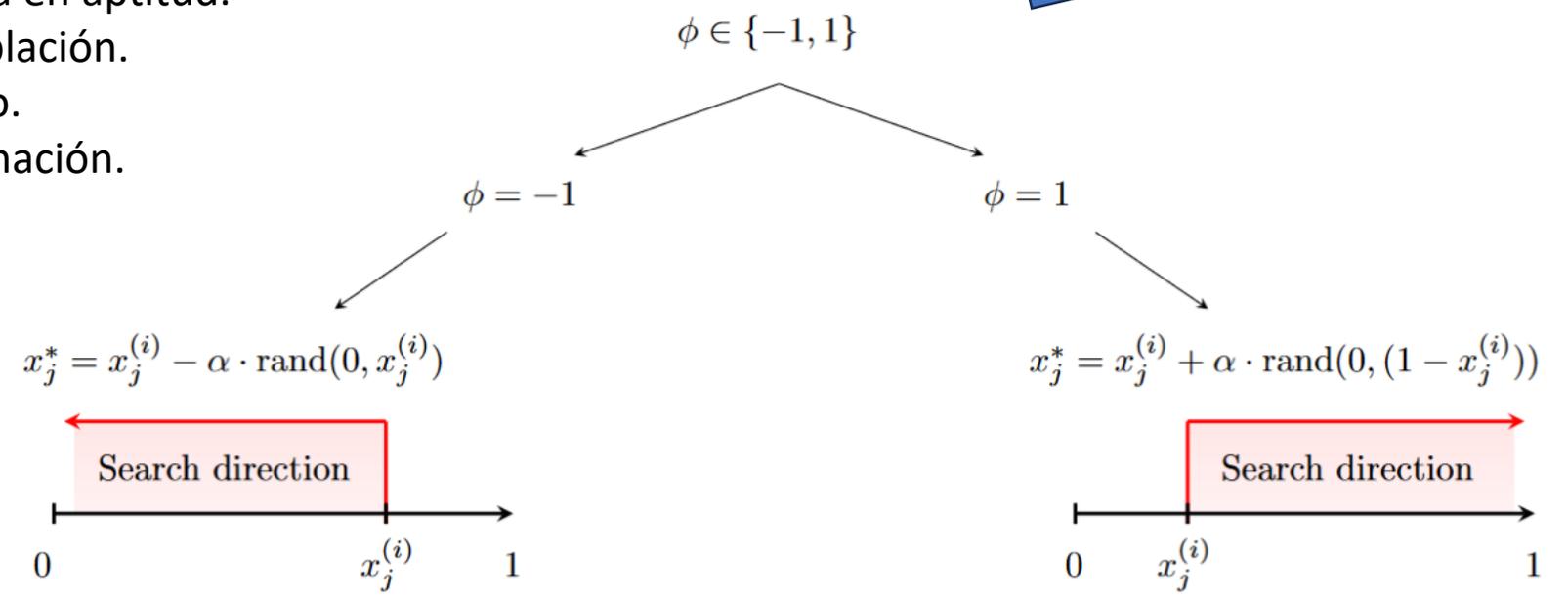


Algoritmo Evolutivo

- Aplicables en optimización (Búsqueda aleatoria dirigida).
- **Componentes clave: población, aptitud, selección, replicación, mutación.**
- Aptitud medida con funciones objetivo (i.e. NSE WNSE).
- Selección identifica mejores miembros para progresar (Sets de parámetros).
- Replicación crea nueva población basada en aptitud.
- Mutación introduce diversidad en la población.
- Tasa de mutación controlada por usuario.
- Proceso iterativo hasta criterio de terminación.

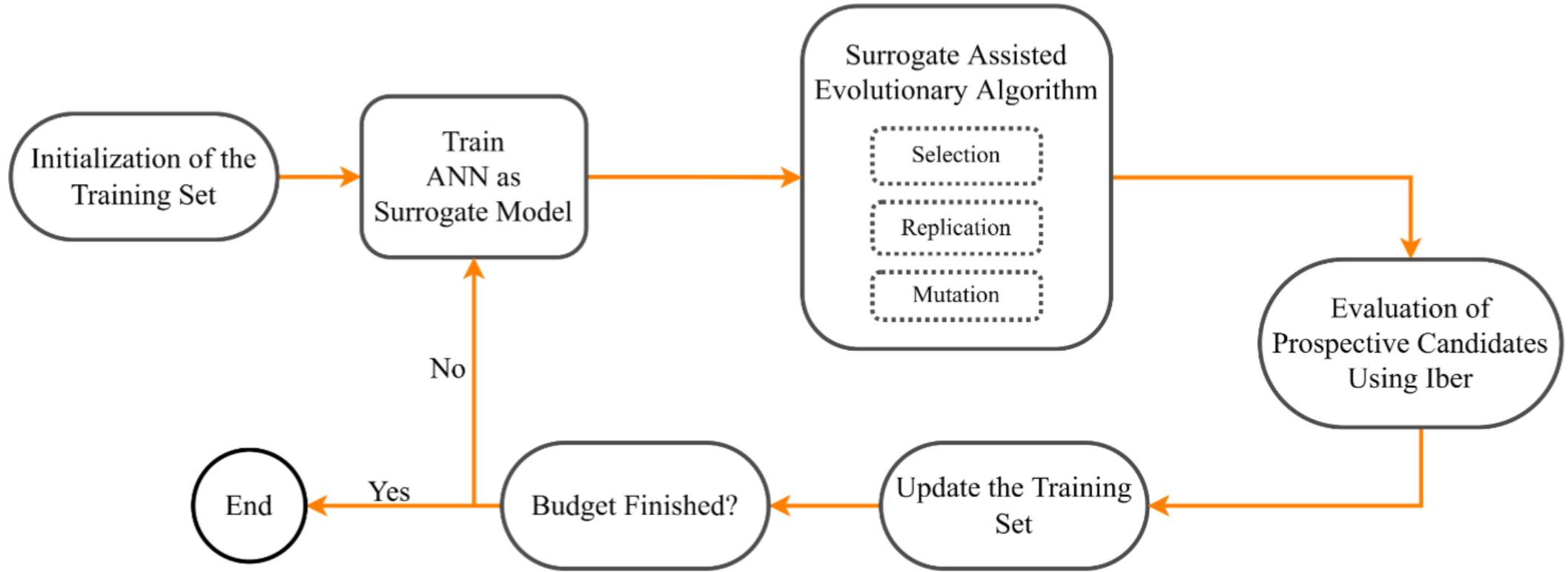


$$x_j^* = \begin{cases} x_j^* - \alpha \cdot \text{rand}(0, x_j^*) & \text{if } \phi = -1, \\ x_j^* + \alpha \cdot \text{rand}(0, (1 - x_j^*)) & \text{if } \phi = 1 \end{cases}$$



Algoritmo Evolutivo Asistido por ANN

- Algoritmos evolutivos (EAs) tradicionales requieren alta carga computacional.
- Se incorpora una ANN para reducir costos computacionales.



Algoritmo Evolutivo Asistido por ANN

- Funciones objetivo: NSE y WNSE para evaluar la aptitud.
- Entrenamiento de la ANN se realiza con validación cruzada.
- Permite replicación múltiple para aprovechar la capacidad de la ANN.
- Enfoque progresivo para actualizar el set de entrenamiento.
- Contadores de mejora y deterioro guían la evolución de los miembros.
- Proceso iterativo hasta alcanzar un criterio de parada.
- Validación de parámetros obtenidos se realiza Iber.
- Set de entrenamiento crece con cada nueva generación.
- El contador de generaciones se incrementa tras cada iteración.
- Permite un entrenamiento más preciso de la ANN.

Algorithm 2 Surrogate-Assisted Evolutionary Algorithm

```
 $\mathbf{X} = \{\mathbf{x}_i \mid \mathbf{x}_i \sim \text{Uniform}\}_{i=1}^N$  ▷ Generation of the Initial Population  
 $\mathcal{D} = \{(\mathbf{x}_i, y_i) \mid y_i = \text{fit}(\mathbf{x}_i)\}$  ▷ Fitness of each Candidate  
 $G \leftarrow 0$  ▷ Initialization of the generation counter  
while  $G < \text{NumGen}$  do  
     $\tilde{M} \leftarrow \text{TrainANN}(\mathcal{D})$  ▷ Train a surrogate model  
     $\Omega = \{\mathbf{x}_i \mid i < q \text{ in } \text{sort}(\mathcal{D} \mid \mathbf{x}_i > \mathbf{x}_j \text{ if } y_i > y_j)\}$  ▷ Selection of top  $q$  best candidates  
     $\mathbf{P} = \{p_i \mid p_i = \frac{y_i}{\sum_{i=1}^q y_i} \text{ and } \mathbf{x}_i \in \Omega\}_{i=1}^q$  ▷ Assign probability  
    for  $i = 1$  to  $GN$  do  
         $\mathbf{x}_i^* \sim \Omega, \mathbf{P}$  ▷ Replication  
         $c \leftarrow 0, d \leftarrow 0$  ▷ Initialize counters  
        while  $c < c_{\text{limit}}$  or  $d < d_{\text{limit}}$  do  
             $\mathbf{x}_i^m \leftarrow \text{Mutate}(\mathbf{x}_i^*)$  ▷ Use equation 10  
            if  $\tilde{M}(\mathbf{x}_i^m) > \tilde{M}(\mathbf{x}_i^*)$  then ▷ Estimate fit of the mutated member using  $\tilde{M}$   
                 $\mathbf{x}_i^* \leftarrow \mathbf{x}_i^m$   
                 $c \leftarrow c + 1$   
            else  
                 $d \leftarrow d + 1$   
             $\mathbf{X} \leftarrow \mathbf{X} \cup \mathbf{x}_i^*$  ▷ Update  $\mathbf{X}$   
             $\mathcal{D} \leftarrow \mathcal{D} \cup (\mathbf{x}_i^*, \text{fit}(\mathbf{x}_i^*))$  ▷ Update  $\mathcal{D}$  by evaluating  $\mathbf{x}_i^*$  using Iber  
     $G \leftarrow G + 1$ 
```

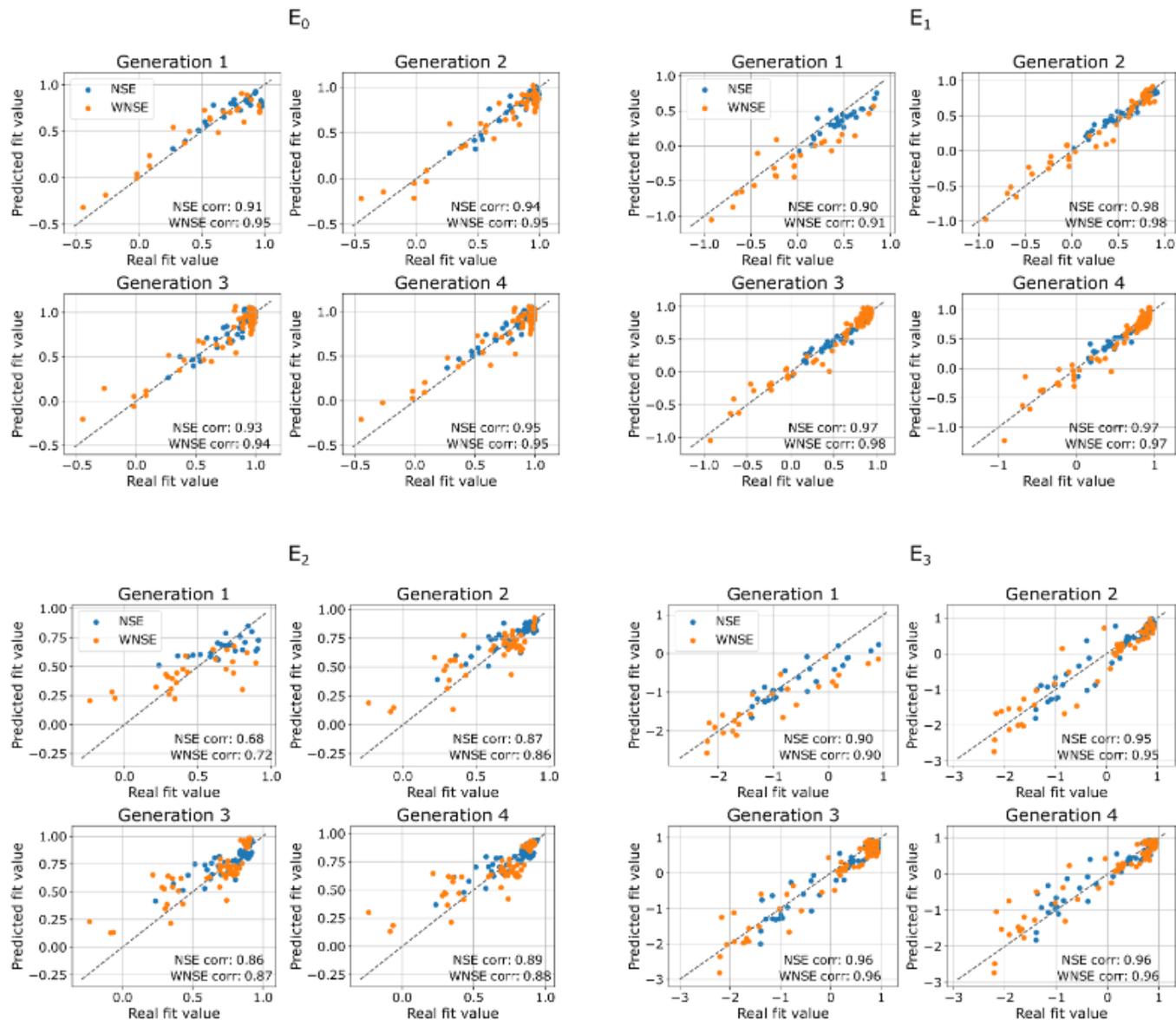
Configuración Experimental

- Número total de generaciones es NumGen=4.
- Población inicial compuesta por N=25 miembros.
- Se replican GN=25 miembros por generación.
- Se seleccionan los q=5 mejores miembros de la población.
- Contador de mejoras inicia en c=0 con límite de 25.
- Contador de deterioros inicia en d=0 con límite en 10.
- Parámetro de evolución establecido en 0.25.

Parameters for SA-EA	Symbol	Value
Generations counter (initial value)	G	1
Number of generations of SA-EA	$NumGen$	4
Number of folds for ANN training	k	5
Members of the initial population	N	25
Replicated members per generation	GN	25
Model parameters	m_x	15
Objective functions	m_y	2
Best population members	q	5
Improvement counter (initial value)	c	0
Improvement counter limit	c_{limit}	25
Deterioration counter (initial value)	d	0
Deterioration counter limit	d_{limit}	10
Evolution parameter	α	0.25

Resultados del Entrenamiento

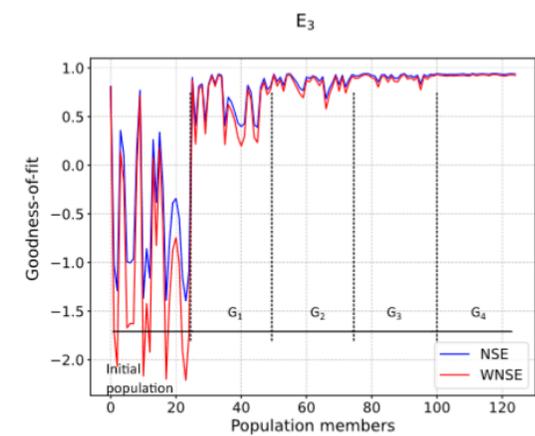
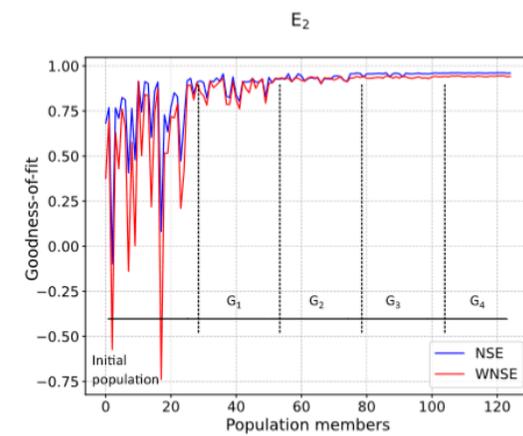
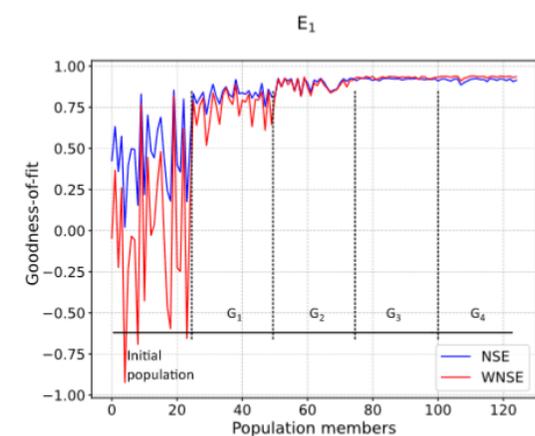
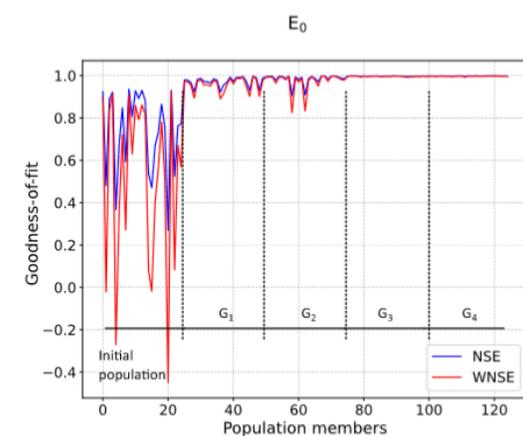
- ANN predice NSE y WNSE basados en parámetros específicos; MSE calculado para evaluar precisión.
- Disminución de MSE en cada generación indica mejora en rendimiento del modelo.
- ANN logra buen ajuste en generaciones G2-G3, dependiendo del evento.
- Generación G1 inicialmente inestable debido a datos insuficientes, pero mejora con validación y datos adicionales.
- Mejora en correlación y ajuste en cada generación; ANN más preciso en zonas con altos valores de NSE y WNSE.



Evolución de Aptitud Generaciones

- Evento sintético E_0 muestra alta adaptabilidad, con todos los miembros superando un valor de ajuste de 0.9 tras primera generación.
- SA-EA eficaz en guiar búsqueda hacia parámetros óptimos incluso con conjunto de entrenamiento pequeño.
- En eventos reales $E_1 - E_3$ mejora significativa en generaciones 3 y 4, alcanzando valores de ajuste superiores a 0.9.
- Convergencia de SA-EA generalmente lograda en generación G3, con 90-100% de miembros con alta adaptabilidad.

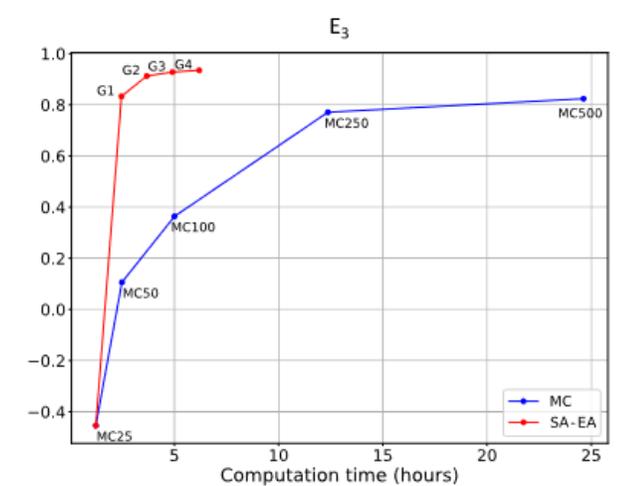
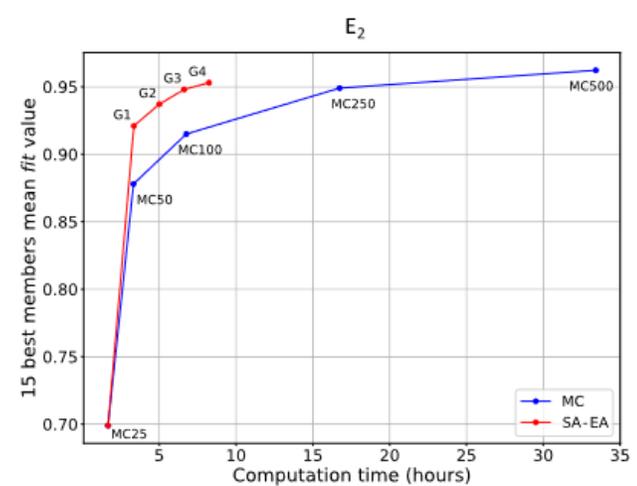
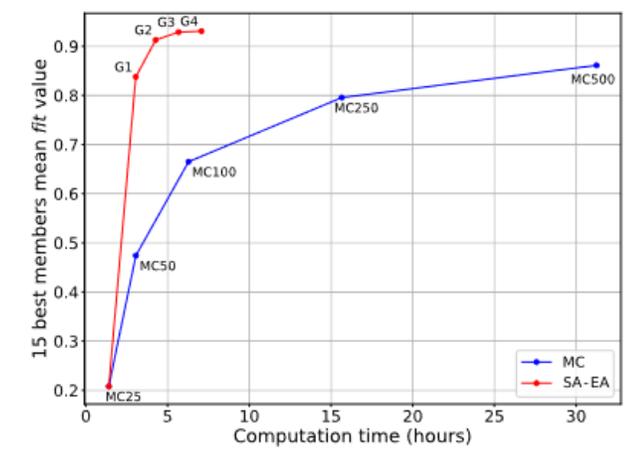
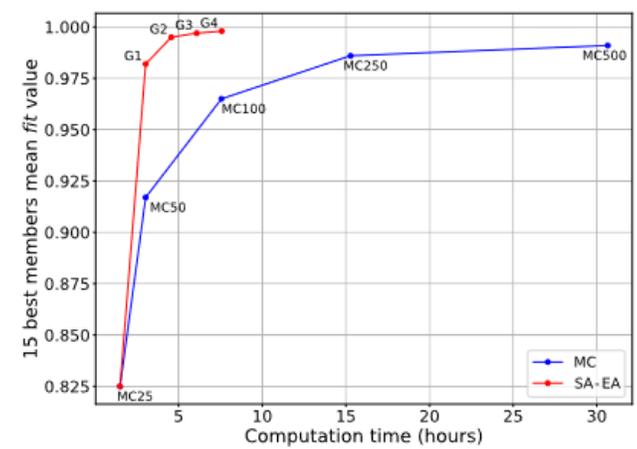
	E_0	E_1	E_2	E_3
Initial population	4/25	0/25	1/25	0/25
G1	25/25	1/25	12/25	3/25
G2	23/25	10/25	25/25	7/25
G3	25/25	25/25	25/25	20/25
G4	25/25	24/25	25/25	25/25



Tiempos de Cálculo

- ANN evaluó entre 385 y 425 mutaciones por generación, tomando solo 40-52 segundos.
- Sin ANN, el modelo hidrológico real tomaría horas para evolucionar un único miembro de la población.
- Tiempo de cómputo para validar los 25 mejores miembros con Iber varía entre 1.21 y 1.68 horas.
- SA-EA converge más rápido que el método Monte Carlo, especialmente en eventos E_0 , E_1 y E_3
- En evento E_2 Monte Carlo alcanza un ajuste ligeramente mejor y toma 5 veces más tiempo computacional.

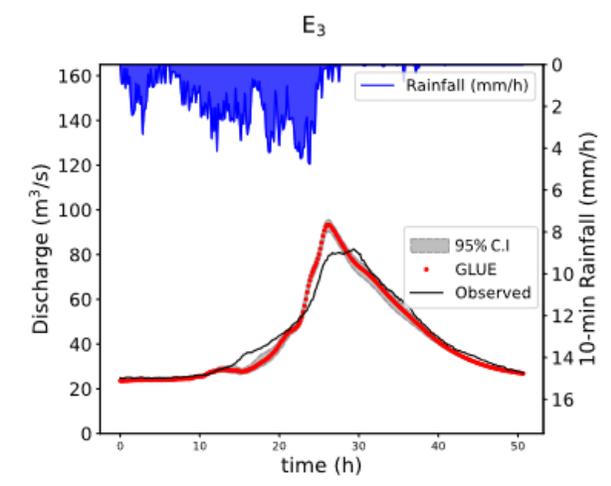
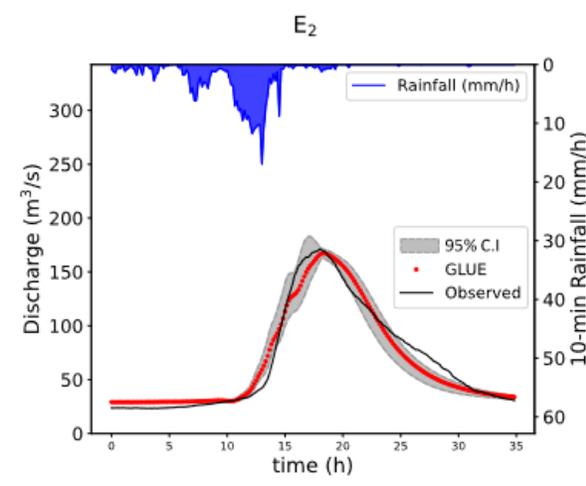
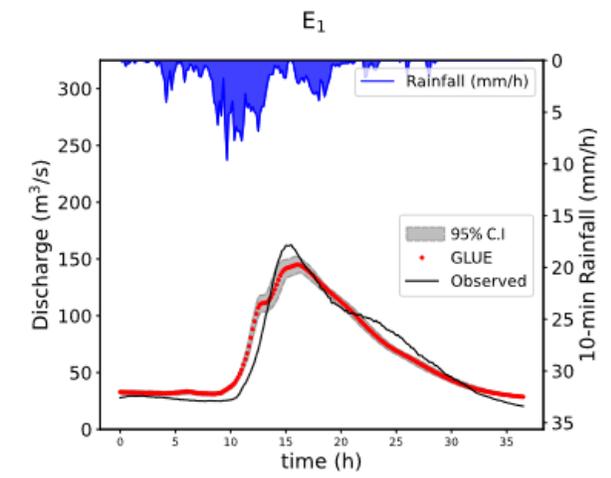
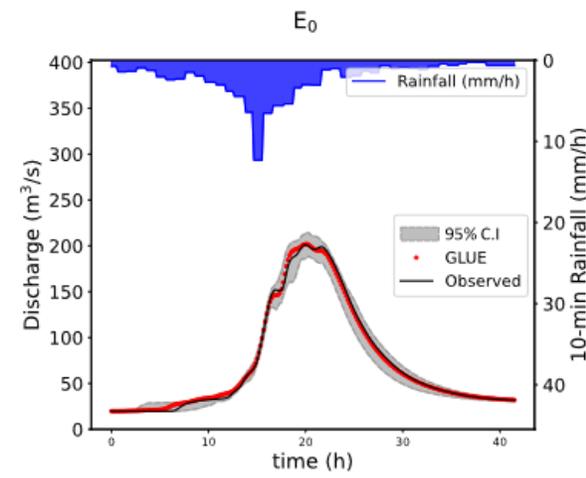
$$fit(\mathbf{x}^{(i)}) = 0.5 \cdot NSE^{(i)} + 0.5 \cdot WNSE^{(i)}$$



Resultados de la Calibración

- Enfoque GLUE utilizado para evaluar capacidad predictiva del modelo, definiendo sets de parámetros “behaviorals” con $fit > 0.90$.
- Ajuste casi perfecto en evento sintético E_0 con pequeñas desviaciones.
- Mejor rendimiento del modelo en evento sintético E_0 , con valores de NSE y WNSE de 0.99.
- Eventos reales E_1 , E_2 y E_3 muestran valores de ajuste superiores a 0.92, indicando un rendimiento del modelo muy bueno según pautas establecidas.

Event	NSE	WNSE	fit
E_0	0.99	0.99	0.99
E_1	0.92	0.93	0.93
E_2	0.94	0.93	0.93
E_3	0.93	0.92	0.93



Parámetros Obtenidos

- Valores de parámetros calibrados varían entre eventos debido a condiciones iniciales de humedad del suelo, que afectan la infiltración y descarga pico.
- Estudios previos indican que descargas fluviales en condiciones de suelo húmedo son tres veces mayores que en condiciones normales.
- Valores de rugosidad obtenidos son generalmente más altos que los comúnmente usados en hidráulica fluvial.
- Esta alta rugosidad probablemente se debe a vegetación y microtopografía no resueltas en el MDT.
- Estudios anteriores han reportado números de Manning más grandes de lo típico, influenciados por varios factores como cubierta vegetal y microtopografía.

Event	Measure	Su_1	Por_1	Sat_1	Ks_1	$loss_1$	$depth_1$	Su_2	Por_2	Sat_2	Ks_2	$loss_2$	$depth_2$	n_1	n_2	n_{river}
E_0	mean	212.8	0.34	0.49	3.2	7.5	5.6	196.1	0.36	0.76	1.7	8.6	7.4	0.09	0.09	0.05
	std	33.6	0.05	0.07	0.4	2.8	1.3	31.6	0.03	0.22	0.3	2.0	0.9	0.04	0.02	0.03
E_1	mean	136.5	0.47	0.12	1.2	2.0	4.5	222.5	0.46	0.55	1.3	6.1	6.0	0.04	0.07	0.07
	std	9.1	0.01	0.02	0.0	1.0	0.9	19.9	0.01	0.08	0.0	0.3	0.6	0.00	0.00	0.01
E_2	mean	189.2	0.40	0.67	1.8	8.5	5.2	157.9	0.37	0.63	1.8	5.4	5.0	0.13	0.07	0.09
	std	18.3	0.05	0.14	0.2	1.1	1.2	27.1	0.02	0.10	0.4	1.6	1.6	0.01	0.03	0.01
E_3	mean	119.7	0.32	0.61	2.8	6.3	9.6	148.2	0.36	0.76	1.6	8.3	0.4	0.18	0.02	0.16
	std	8.5	0.00	0.02	0.2	0.5	0.0	7.7	0.00	0.01	0.1	1.3	0.0	0.00	0.00	0.00

Conclusiones

- El estudio desarrolló un SA-EA para calibrar modelos hidrológicos distribuidos, utilizando un modelo sustituto para reducir costos computacionales.
- El SA-EA calibró eficazmente el modelo hidrológico, reduciendo tiempos de ejecución en 5-6 veces comparado con técnicas de optimización globales.
- La precisión del ANN en predecir NSE y WNSE mejoró con cada generación del EA, gracias a la incorporación de más datos validados.
- El SA-EA identificó conjuntos de parámetros con valores de ajuste superiores a 0.90, logrando tasas de éxito entre 90% y 100% en la tercera generación.
- Sin el SA-EA, la aplicación del EA sería inviable debido al alto costo computacional de resolver las ecuaciones de aguas someras en 2D múltiples veces.

Algunas referencias

- Farfán, J. F., & Cea, L. (2021). Coupling artificial neural networks with the artificial bee colony algorithm for global calibration of hydrological models. *Neural Computing and Applications*, 33, 8479-8494.
- Cea, L., & Bladé, E. (2015). A simple and efficient unstructured finite volume scheme for solving the shallow water equations in overland flow applications. *Water resources research*, 51(7), 5464-5486.
- Couckuyt, I., Gorissen, D., Rouhani, H., Laermans, E., & Dhaene, T. (2009). Evolutionary regression modeling with active learning: An application to rainfall runoff modeling. In *Adaptive and Natural Computing Algorithms: 9th International Conference, ICANNGA 2009, Kuopio, Finland, April 23-25, 2009, Revised Selected Papers 9* (pp. 548-558). Springer Berlin Heidelberg.
- Guo, Z., Leitao, J. P., Simões, N. E., & Moosavi, V. (2021). Data-driven flood emulation: Speeding up urban flood predictions by deep convolutional neural networks. *Journal of Flood Risk Management*, 14(1), e12684.

Financiamiento

This work was partially funded by the Spanish Ministry of Science and Innovation (MCIN/AEI/10.13039/501100011033) within the project “SATURNO: Early warning against pluvial flooding in urban areas” (PID2020-118368RB-I00).

This work has been supported by the Flemish Government under the ‘Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen’ and the ‘Fonds Wetenschappelijk Onderzoek (FWO)’ programmes.