

# HYPERSTREAMHS: UN MODELLO IDROLOGICO CONTINUO BASATO SU UN APPROCCIO MPI-DUAL-LAYER PER UNA MODELLAZIONE A GRANDE SCALA DELLE INFRASTRUTTURE IDRAULICHE

Diego Avesani<sup>1\*</sup>, Andrea Galletti<sup>1</sup>, Alberto Bellin<sup>1</sup>, Majone Bruno<sup>1</sup>

(1) Dipartimento di Ingegneria Civile, Ambientale e Meccanica, Università degli Studi di Trento (Trento)

\*email: diego.avesani@unitn.it

## ASPETTI CHIAVE

- Modellazione a grande scala
- High performance computing
- Gestione delle risorse idriche
- Hydropower

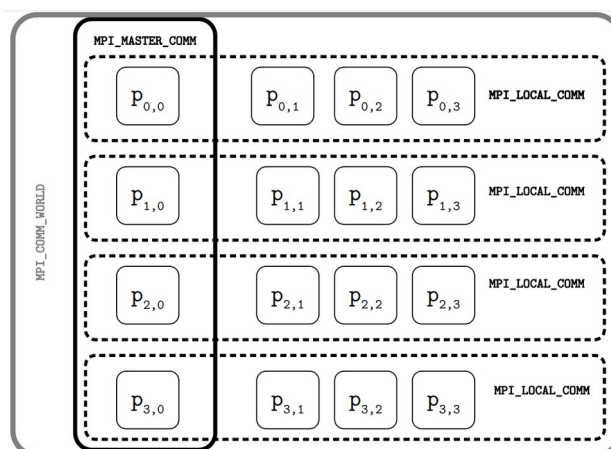
## 1 INTRODUZIONE

I modelli idrologici su larga scala hanno la necessità di un utilizzo massivo di potenza di calcolo e di grande disponibilità di memoria ad accesso casuale (RAM) (Vivoni *et al.* 2011). Questo è dovuto alla esigenza di eseguire un elevato numero di simulazioni per un'accurata valutazione dell'incertezza parametrica e di alta risoluzione in grado di rappresentare correttamente l'alterazione dei regimi idrologici dovuta alla presenza di infrastrutture idrauliche quali dighe e opere di derivazione (Liu *et al.* 2014).

In questo contesto, la programmazione parallela permette da un lato l'utilizzo di memoria distribuita, garantendo quindi elevate capacità in termini di utilizzo di dati, e dall'altro di offrire prestazioni elevate in termini di velocità di calcolo. Tuttavia, la parallelizzazione degli algoritmi utilizzati nella simulazioni idrologiche è fortemente limitata per la presenza di parti degli stessi non parallelizzabili, come ad esempio la modellazione del routing lungo la rete idrologica (Piccolroaz *et al.* 2015) e la modellazione delle infrastrutture idrauliche artificiali.

Al fine di superare queste limitazioni, è stato sviluppato il software HYPERstreamHS (Avesani *et al.* 2021). Questo adotta una strategia di parallelizzazione cosiddetta a doppio strato (dual-layer), in cui la versione parallela del kernel idrologico costituisce il primo strato, mentre il secondo strato si occupa della modellazione inversa (calibrazione). Inoltre, HYPERstreamHS è dotato di vari moduli per la modellazione esplicita delle alterazioni antropiche dei deflussi naturali, che lo rendono in grado di simulare i processi idrologici in grandi bacini fluviali con deflusso alterato (Avesani *et al.* 2021).

## 2 METODOLOGIA



**Figura 1.** Esempio di suddivisione dei  $n_p = 16$  processori disponibili in due layer. Ciascun gruppo di processori operanti sul primo layer è coordinato da un processore master che si occupa di comunicare con gli altri master (secondo layer).

## 2.1 Strategia Dual-Layer

La parallelizzazione dual-layer è ottenuta sfruttando il paradigma master-slaves presente nel protocollo MPI (Mpi Forum, 1994). A scopo illustrativo, la Figura 1 mostra un esempio con 16 processori ( $np = 16$ , denominati  $p_0, \dots, p_{15}$ ), ipoteticamente disponibili in un sistema HPC (High Performance Computing). I 16 processori vengono divisi in 4 sottoinsiemi che costituiscono ognuno il primo livello della parallelizzazione MPI. Ogni sottogruppo è quindi in grado di eseguire simulazioni idrologiche parallele ed indipendenti. Si può notare come in questo esempio specifico ogni sottoinsieme è dotato di un master e di 3 slaves. Il secondo layer è invece costituito da un ulteriore insieme dato dai master di ogni sottogruppo ed ha lo scopo di gestire l'algoritmo di calibrazione o di valutazione dell'incertezza idrologica.

## 2.2 Infrastrutture Idrauliche

Il modello prevede la modellazione esplicita delle alterazioni dei deflussi naturali legate ad attività idroelettriche, quali serbatoi, opere di derivazione, restituzione, etc. Le infrastrutture idrauliche sono inserite nel modello come nodi all'interno del reticolo fluviale. Tali nodi (definiti *artificiali*) si differenziano dai cosiddetti nodi *naturali* in quanto, mentre in questi ultimi viene effettuato ad ogni step temporale un bilancio di massa tra tutti gli afflussi prima di effettuare il routing dei deflussi a valle, nei nodi *artificiali* viene applicato uno specifico bilancio di massa vincolato alle specificità tecniche ed operative del nodo, dopo il quale i deflussi sono opportunamente partizionati nelle risultanti componenti artificiali (flussi circolanti in condotte/canali o immagazzinati in un serbatoio) e naturali.

## 3 CASO STUDIO

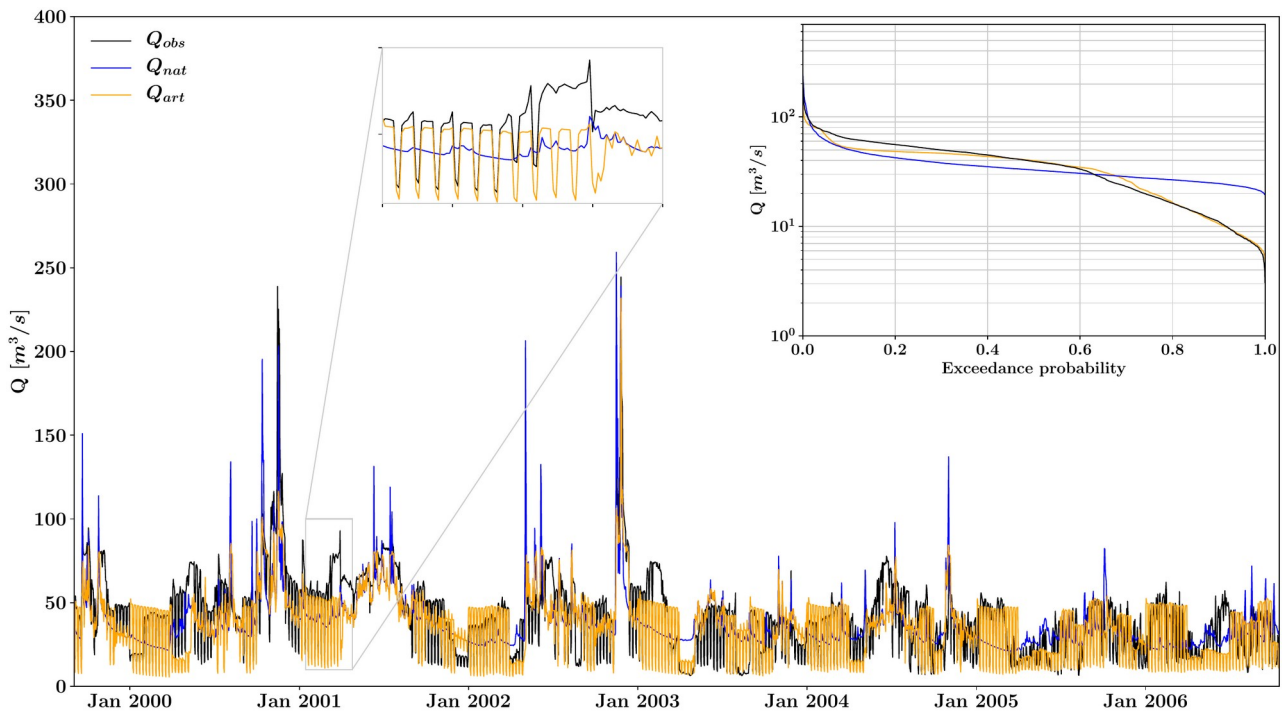
Il modello è stato testato con applicazione al bacino del fiume Adige, chiuso alla stazione di Vò Destro (circa 10600 km<sup>3</sup>). Sono state testate sia la scalabilità del modello sia la sua abilità di riprodurre i deflussi osservati in sezioni particolarmente alterate. In particolare, i test di scalabilità sono stati effettuati con riferimento all'intero bacino, includendo 30 serbatoi, 41 opere di presa e 40 grandi centrali idroelettriche (potenza installata > 3 MW) per un totale di 138 nodi includendo le sezioni naturali. La capacità del modello di riprodurre i deflussi in bacini antropizzati è stata valutata limitatamente al bacino del fiume Noce (ca. 1350 km<sup>2</sup>) chiuso alla stazione di Mezzolombardo: i deflussi in questa sezione sono fortemente influenzati dai rilasci della centrale idroelettrica di Mezzocorona, collocata 2.5 km a monte e con una portata massima turbinabile pari a 60 m<sup>3</sup>/s, a fronte del deflusso naturale medio di 20 m<sup>3</sup>/s. Tutti i test sono stati effettuati sia in regime artificiale che in regime naturale, ovvero includendo le alterazioni dovute all'utilizzo idroelettrico o meno, in modo da poterne determinare l'impatto sulle performance di scalabilità e sulla capacità di riprodurre i deflussi.

## 4 RISULTATI E CONCLUSIONI

HYPERstreamHS include due algoritmi per la calibrazione e lo studio dell'incertezza parametrica: il primo è il cosiddetto Particle Swarm Optimizer (PSO) (*Kennedy et Eberhart*, 1995) mentre il secondo è lo schema denominato stratified Latin Hypercube Sampling (LHS) (*Majone et al.* 2010, *Piccolroaz et al.* 2015). Lo scopo di questa sezione è quindi di confrontare le performance della strategia dual-layer utilizzando lo schema LHS e lo schema PSO per la calibrazione di HYPERstreamHS nel caso studio presentato in sezione 3.

### 4.1 RIPRODUZIONE DEI DEFLUSSI

I deflussi sono stati modellati nella sezione di Mezzolombardo, posta a chiusura del bacino del fiume Noce, e i cui deflussi sono fortemente alterati da attività idroelettriche (cfr. sezione 3). I parametri del modello idrologico naturale sono stati calibrati tramite PSO con riferimento alla stazione di Malè, ottenendo un indice NSE pari a 0.77. La riproduzione dei deflussi alla sezione di Mezzolombardo in condizione naturale e artificiale ha ottenuto un indice NSE pari a 0.27 e 0.43, rispettivamente. Nonostante tali valori indichino che la ricostruzione del deflusso alla sezione di Mezzolombardo non raggiunga un risultato soddisfacente, è possibile notare (Figura 2) che il realismo delle predizioni migliora drasticamente in regime artificiale, con particolare riferimento ai flussi bassi registrati durante l'hydropeaking in periodi di magra.



**Figura 2.** Flussi simulati artificiali (arancio), naturali (blu) e osservati (nero) alla stazione di Mezzolombardo. In sotto-figura è rappresentato l'andamento temporale dei deflussi negli anni 2000-2006. L'inset mostra le CDF dei deflussi simulati ed osservati.

## 4.2 SCALABILITÀ

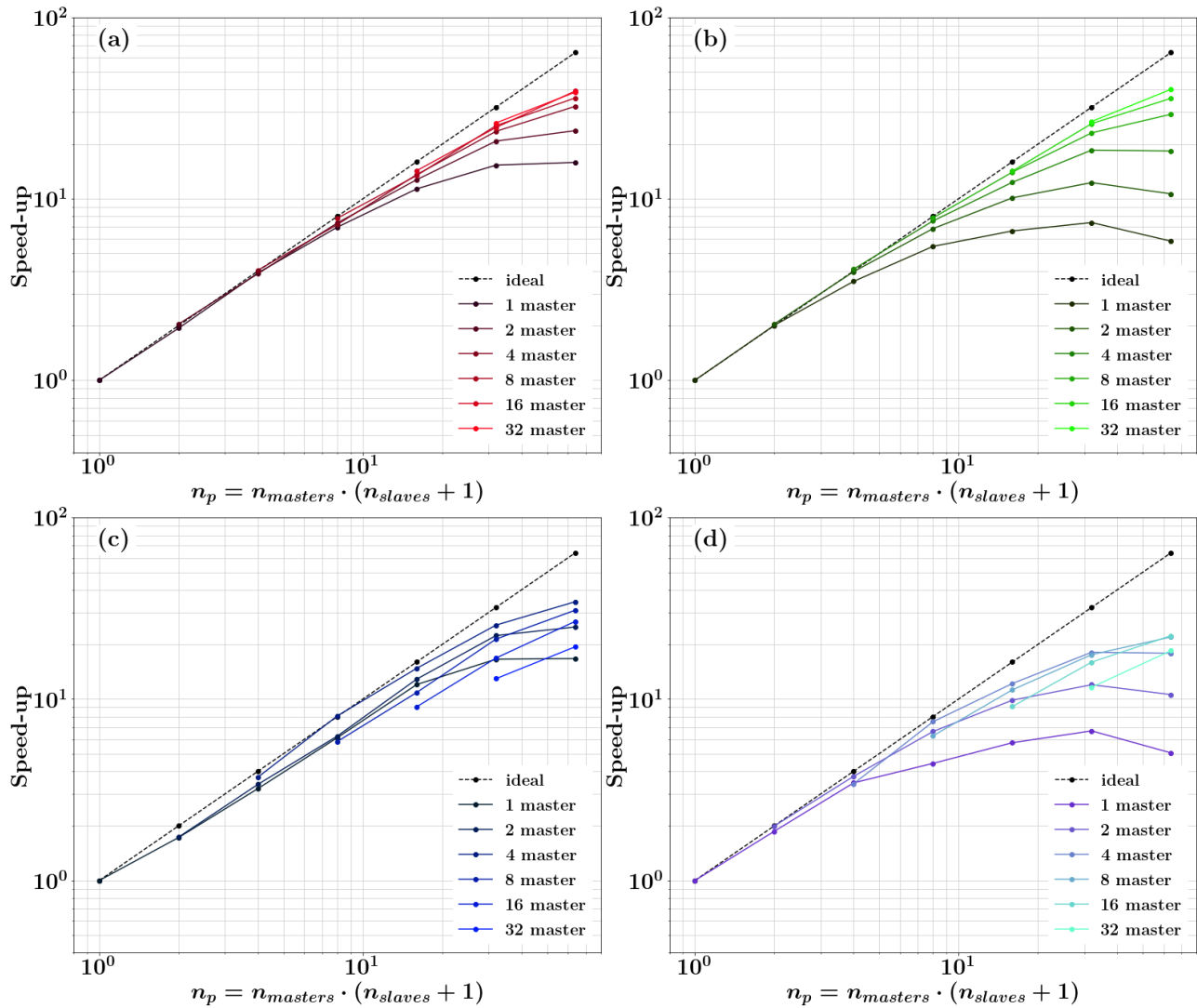
Sono state analizzate diverse combinazioni tra il numero di master e il numero di slave. In particolare le combinazioni scelte sono state identificate tra due casi estremi. Il primo estremo è rappresentato dall'utilizzo di un singolo processore che funge sia da master che da slave. Il secondo estremo invece è costituito dall'uso di 32 master ognuno dotato di un proprio slave per un totale di 64 processori. Il primo estremo rappresenta quindi il tempo di calcolo di riferimento riferendosi al caso di standard di calcolo sequenziale. Tutte le analisi sono state eseguite utilizzando una rete idrologica costituita da 5 nodi e una rete idrologica costituita da 138. La figura 3 mostra lo speed-up ottenuto grazie alla strategia dual-layer applicata agli schemi LHS e PSO, rispettivamente.

Fig. 3a e 3b mostra i risultati delle simulazioni condotte con lo schema LHS. Si può notare che nel caso caratterizzato dalla presenza di 5 nodi, quando viene utilizzato un singolo master, o in altre parole quando il secondo layer MPI è disattivato, lo speed-up ottenuto si allontana velocemente dallo speed-up teorico all'aumentare del numero dei processori. Al contrario, all'aumentare del numero del master lo speed-up ottenuto si avvicina allo speed-up massimo teorico. In altre parole, maggiore è il numero dei master minore è il tempo dedicato alla comunicazione tra i processori, aumentando di conseguenza le prestazioni dell'algoritmo parallelo. Al contrario, l'aumentare del numero dei processori appartenenti al primo layer aumenta il tempo dedicato alla comunicazione fra gli stessi riducendo il tempo dedicato al calcolo.

Nel caso invece dello schema PSO (Fig. 3c-d), all'aumentare del numero dei master non sempre corrisponde un miglioramento delle performance e quindi un avvicinamento delle curva allo speed-up teorico. Questo è dovuto alla differenza tra gli algoritmi PSO e LHS. Mentre infatti nello schema LHS non vi è sostanzialmente comunicazioni tra i masters, il PSO necessita uno scambio di informazioni tra i master ad ogni step iterativo. In quest'ultimo caso, il tempo macchina dedicato allo stesso diventa progressivamente più rilevante con il numero di masters.

In conclusione, il modello idrologico HYPERstreamHS ha presentato ottime caratteristiche di scalabilità per l'utilizzazione su infrastrutture di calcolo caratterizzate da un numero di processori intermedio, evidenziando i benefici dell'implementazione di un secondo livello di programmazione in parallelo a parità di processori impiegati nel calcolo. Il modello può essere calibrato facendo uso di vari algoritmi genetici e permette di orientare la calibrazione a numerose funzioni obiettivo, garantendo la massima flessibilità

rispetto ai fini della modellazione. Per le sue caratteristiche, il modello si presta in particolare alla soluzione di problemi di modellazione legati alla gestione e/o ottimizzazione dell'utilizzo delle risorse idriche in contesti antropizzati.



**Figura 3.** Performance di scalabilità dello schema dual-layer applicato a LHS (sotto-figure a e b) e a PSO (sotto-figure c e d), nelle configurazioni a 5 nodi (sotto-figure a e c) e 138 nodi (sotto-figure b e d).

## RIFERIMENTI BIBLIOGRAFICI

- Avesani, D., Galletti, A., Piccolroaz, S., Bellin, A., & Majone B., A Dual layer MPI continuous large-scale hydrological model including Human Systems. *Environmental Modelling & Software* (2021), 139, 105003.
- Kennedy, J., Eberhart, R., Nov, 1995. Particle swarm optimization. In: *Proceedings of ICNN'95 - International Conference on Neural Networks*, vol. 4, pp. 1942–1948.
- Liu, J., Zhu, A.-X., Liu, Y., Zhu, T., Qin, C.-Z., 2014. A layered approach to parallel computing for spatially distributed hydrological modeling. *Environ. Model. Software* 51, 221-227
- Mpi Forum, M.P., 1994. *Mpi: a message-passing interface standard*. Tech. rep. (Knoxville, TN, USA).
- Majone, B., Bertagnoli, A., Bellin, A., 2010. A non-linear runoff generation model in small alpine catchments. *J. Hydrol.* 385 (1), 300–312.
- Piccolroaz, S., Di Lazzaro, M., Zarlenga, A., Majone, B., Bellin, A., & Fiori, A. Hyperstream: a multi-scale framework for streamflow routing in large-scale hydrological model. *Hydrology & Earth System Sciences*, 2016, 20(5).
- Vivoni, E.R., Mascaro, G., Mniszewski, S., Fasel, P., Springer, E.P., Ivanov, V.Y., Bras, R. L., 2011. Real-world hydrologic assessment of a fully-distributed hydrological model in a parallel computing environment. *J. Hydrol.* 409 (1), 483–496.