

UTILIZZO DELLA TRANSFER ENTROPY PER L'IDENTIFICAZIONE DELLA SIMILARITÀ DEI PROCESSI DI TRASFORMAZIONE AFFLUSSI DEFLUSSI

Mattia Neri ^{1*}, Paulin Coulibaly ² & Elena Toth ¹

(1) Dipartimento di Ingegneria Civile, Chimica, Ambientale e dei Materiali, Università di Bologna

(2) McMaster University, Hamilton, Canada

*email: mattia.neri5@unibo.it

ASPETTI CHIAVE

- Interazione fra le serie temporali di forzanti meteorologiche e di portata osservata.
- Utilizzo della transfer entropy per la quantificazione del flusso informativo nei processi afflussi-deflussi.
- Flusso informativo come nuova signature per la caratterizzazione delle dinamiche idrologiche dominanti.
- Classificazione dei bacini basata sulle misure di transfer entropy.

1 MOTIVAZIONI E OBIETTIVI DELLO STUDIO

L'analisi della similarità fra bacini è da sempre considerata uno dei passaggi fondamentali per il trasferimento dell'informazione idrologica e ha lo scopo di individuare i processi idrologici dominanti e le loro caratteristiche. Infatti, la definizione di gruppi di bacini simili è necessaria per diversi approcci di regionalizzazione.

La maggior parte degli studi in letteratura descrive la risposta idrologica dei bacini tramite indici o "signature" idrometrici, basati su valori osservati di portata, considerando quindi la similarità fra diverse caratteristiche dell'idrogramma. D'altra parte, in caso di bacini non strumentati, sono di norma utilizzati descrittori di bacino di tipo geomorfologico o climatico (e.g. *Betterle et al.*, 2018; *Di Prinzio et al.*, 2011; *Viglione et al.*, 2007). Quando, tuttavia, il trasferimento dell'informazione riguarda informazioni sui fenomeni di trasformazione afflussi-deflussi a scala temporale fine, la similarità dovrebbe tenere conto della successione temporale dei deflussi e dei processi di generazione e propagazione degli stessi, considerando il contenuto informativo dell'intero idrogramma e delle sue relative forzanti. Per fare ciò, la letteratura si è concentrata principalmente sulla definizione della similarità della risposta idrologica tramite diverse strategie di classificazione che considerano per esempio la correlazione temporale dei deflussi (e.g., fra molti altri, *Grimaldi*, 2004; *Lombardi et al.*, 2012) o che utilizzano tecniche di data-mining per esplorare le dinamiche dell'idrogramma a diverse scale temporali (e.g. *Singh et al.*, 2016). Allo scopo di considerare la caratterizzazione sia dell'idrogramma che delle relative forzanti meteorologiche, lo studio di *Toth*, (2013) propone, per la prima volta, una classificazione dei bacini appenninici emiliano-romagnoli basata sulla variabilità e sulla struttura di correlazione a scala temporale fine di serie storiche sia di precipitazione che di deflussi, anche se tali indici sono definiti indipendentemente per lo ietogramma e per l'idrogramma. Nello studio della similarità nei processi di trasformazione afflussi-deflussi, fino ad ora la letteratura non ha infatti preso in considerazione la relazione tra la successione temporale a scala fine delle forzanti meteorologiche e quella delle portate alla sezione di chiusura. L'obiettivo di questo studio è proporre un metodo per caratterizzare ogni bacino in base ad una misura di tale *interazione*: un indice di questo tipo potrebbe risultare infatti particolarmente efficace per l'individuazione dei processi idrologici predominanti, quando si abbia come obiettivo il trasferimento di informazioni riguardanti le trasformazioni afflussi deflussi, come per esempio nella regionalizzazione dei parametri di modelli idrologici.

Data l'alta non linearità che caratterizza le relazioni che governano i processi afflussi-deflussi, tale analisi richiede l'utilizzo di una misura che guardi oltre la classica cross-correlazione fra le serie temporali. In particolare, può risultare utile l'utilizzo dei concetti della teoria dell'informazione, basati sulla nozione di entropia. Tali concetti sono infatti già ampiamente utilizzati in idrologia e nell'ottimizzazione delle risorse idriche, per applicazioni che spaziano dal design delle reti di monitoraggio fluviali o climatiche (e.g. *Keum and Coulibaly*, 2017) fino, per esempio, alla riduzione del costo computazionale per la simulazione di modelli

distribuiti (Ehret et al., 2020). Di particolare interesse è lo studio di Bennett et al. (2019) che dimostra come la *transfer entropy*, una quantità asimmetrica multivariata che analizza l'interazione tra diversi segnali, può essere utilizzata per quantificare il trasferimento attivo di informazioni fra i vari processi idrologici a diverse scale temporali, applicata in tale lavoro con lo scopo di validare il comportamento dei modelli idrologici.

Il presente studio propone di utilizzare il concetto di transfer entropy per la caratterizzazione dei bacini e per l'identificazione dei processi idrologici dominanti. In una prima fase, il flusso informativo dalle serie temporale delle tre principali forzanti meteorologiche (precipitazione, evapotraspirazione e scioglimento nivale) alla portata osservata viene quantificato in termini di transfer entropy. I valori ottenuti sono poi utilizzati come *signature* per caratterizzare le dinamiche di bacino. Infine, si procede alla classificazione dei bacini assumendo che valori simili di transfer entropy siano associati a bacini idrologicamente simili. L'analisi è applicata ad un esteso data set di oltre 200 bacini austriaci e i risultati ottenuti sono discussi confrontando i valori di transfer entropy e la classificazione con le caratteristiche dei bacini ed un set di tipiche signature.

Lo scopo del lavoro è di dimostrare il potenziale della transfer entropy per la descrizione e classificazione dei processi di trasformazione afflussi-deflussi, come strumento complementare agli indici tradizionalmente utilizzati per la caratterizzazione dei bacini.

2 BREVI CENNI SU TEORIA DELL'INFORMAZIONE E TRANSFER ENTROPY

Nella teoria dell'informazione, l'entropia quantifica l'incertezza, derivante dalla probabilità di occorrenza, associata ad una variabile. Di conseguenza, l'entropia è una misura della quantità di informazione contenuta in un generico segnale. L'entropia di Shannon $H(X)$ fornisce una formulazione matematica per definire il contenuto informativo di una variabile X (Shannon, 1948). Il contenuto informativo medio associato ad una variabile X è chiamato *marginal entropy* ed è definito come:

$$H(X) = -\sum_{i=1}^n p(x_i) \log_2 p(x_i) \quad (1)$$

dove n è il numero totale di intervalli discreti (chiamati anche *bin*) che rappresentano i diversi stati della variabile e $p(x_i)$ è la probabilità che la variabile X ricada nell'intervallo i -esimo.

In maniera simile possono essere definite versioni multivariate di entropia (qui non riportate per brevità) come la *mutual information* $I(X,Y)$ che definisce la quantità di informazione comune a più variabili e la *conditional mutual information* $I(X,Y|Z)$ che definisce l'informazione condivisa fra due variabili X e Y condizionata da quella già contenuta in una terza variabile Z . Mentre tutte le quantità citate fino ad ora sono simmetriche, la *transfer entropy* definita da (Schreiber, 2000), considera il flusso direzionale fra segnali. Essa è definita come un particolare caso di *conditional mutual information* dove le variabili sono traslate artificialmente nel tempo:

$$TE_{X \rightarrow Y}(lx, ly) = I(Y_t, X_{t-1}, \dots, X_{t-lx} | Y_{t-1}, \dots, Y_{t-ly}) \quad (2)$$

dove $TE_{X \rightarrow Y}$ è la transfer entropy da X a Y associata ai lag lx e ly . La *transfer entropy* rappresenta quindi la conoscenza aggiuntiva conseguita sullo stato di Y al tempo t , misurando X nei precedenti lx passi temporali, rispetto all'informazione nota conoscendo i precedenti ly stati di Y . Il principale vantaggio di tale grandezza è che è in grado di guardare oltre la correlazione lineare fra le variabili.

3 CARATTERIZZAZIONE DEI BACINI TRAMITE TRANSFER ENTROPY

La fase preliminare della metodologia proposta consiste nell'individuazione delle forzanti meteorologiche che governano la generazione del deflusso. Nel presente lavoro, tali variabili consistono nelle serie temporali (a scala giornaliera, così come le osservazioni di portata) di i) precipitazione P (osservata), ii) evapotraspirazione effettiva AET e iii) scioglimento nivale $melt$, che vengono stimate tramite l'applicazione di un modello afflussi deflussi semi-distribuito a simulazione giornaliera continua (TUW model, Viglione and Parajka, 2018), consolidato per l'area di studio e calibrato su un periodo di 30 anni.

La fase centrale dell'analisi consiste nella stima del flusso informativo da ciascuna delle tre forzanti alla portata osservata, per tutti i bacini. Tale flusso viene stimato calcolando i tre valori di *transfer entropy* dalle

tre componenti di generazione del deflusso. Al fine di testare la metodologia più parsimoniosa per valutare il potenziale della TE , aumentando l'affidabilità della stima, per il calcolo della *transfer entropy* si considera qui un ritardo temporale singolo ($lx=ly=1$). L'eq. (2) diventa quindi:

$$TE_{F \rightarrow Q}(1,1) = I(Q_t, F_{t-1}, |Q_{t-1}) \quad (3)$$

dove F è a turno P , AET o $melt$, ottenendo le tre quantità $TE_{P \rightarrow Q}$, $TE_{AET \rightarrow Q}$ e $TE_{melt \rightarrow Q}$. Ogni bacino risulta così caratterizzato da tre valori di *transfer entropy* che stimano l'ammontare delle informazioni trasferite dalle tre variabili meteorologiche alla portata osservata.

I risultati del calcolo delle singole TE e la loro interpretazione rispetto ai descrittori morfo-climatici di bacino e alle *signature* idrometriche classiche non vengono qui riportati per brevità. Si preferisce infatti privilegiare la presentazione dei risultati della classificazione, che permette una più facile interpretazione dei risultati accorpando l'andamento spaziale delle tre TE in un unico esito. Maggiori dettagli sia sulla metodologia che sui risultati sono descritti nel contributo in corso di pubblicazione su una rivista internazionale.

4 CLASSIFICAZIONE DEI BACINI

Nell'approccio proposto, si assume che la similarità fra bacini possa essere associata alla quantità di informazioni trasferite alla portata osservata dalle diverse componenti che generano il deflusso (ovvero bacini con simili valori delle tre TE , sono considerati anche idrologicamente simili). Perciò, l'ultima fase della metodologia (qui chiamata TE-HC, Transfer Entropy - Hierarchical Clustering) prevede l'implementazione di un algoritmo di clustering per classificare i bacini austriaci in funzione delle tre TE ottenute. Le dissimilarità fra i bacini sono calcolate tramite la distanza euclidea nello spazio tridimensionale delle tre transfer entropy e la classificazione viene effettuata con l'algoritmo di tipo gerarchico *Ward's minimum variance* (Ward, 1963).

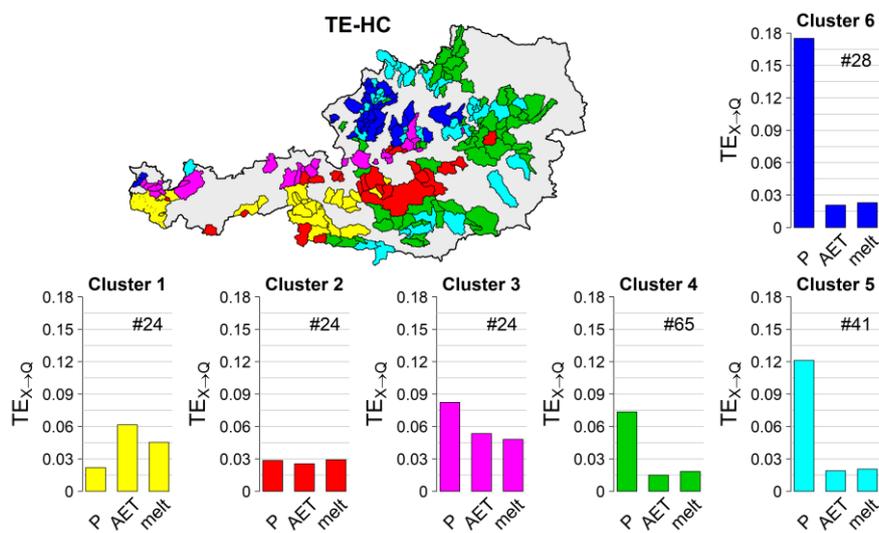


Figura 1. Risultati della TE -HC in 6 cluster. Il colore dell'area dei bacini si riferisce al cluster di appartenenza. I diagrammi a barre mostrano i valori di TE medi di ciascuna variabile nel cluster. Si riporta anche il numero di bacini che compongono ogni cluster.

La Figura 1 mostra il risultato della classificazione TE-HC. Inoltre, per favorire l'interpretazione delle classi ottenute, la Figura 2 riporta la variabilità di sei descrittori di bacino di tipo morfologico e climatico (a-f) e di cinque tipiche *signature* idrometriche (g-m), la cui descrizione è riportata in legenda, all'interno dei sei cluster. Si nota come i cluster 6 e 5 mostrino il massimo impatto di P e minimo di AET e $melt$: infatti sono bacini a bassa quota e a basso contributo nevoso, che mostrano una risposta molto veloce. Il cluster 4 mostra invece un minore impatto di P : include bacini a quote leggermente superiori e pioggia annua minima. I cluster 1 e 2 raggruppano invece i bacini col minimo impatto di P , paragonabile o inferiore a quello di AET e $melt$: in coerenza con tale risultato, tali bacini sono montani, ad alta stagionalità e mostrano la massima memoria dell'idrogramma. Infine, il cluster 3 include bacini con un effetto predominante di P , ma dove l'impatto di

AET e melt è comunque significativo. Quest'ultimo cluster risulta particolarmente interessante in quanto include i bacini più piovosi (a causa di forti piogge orografiche), ma dove l'impatto della neve sui processi afflussi-deflussi è molto significativo.

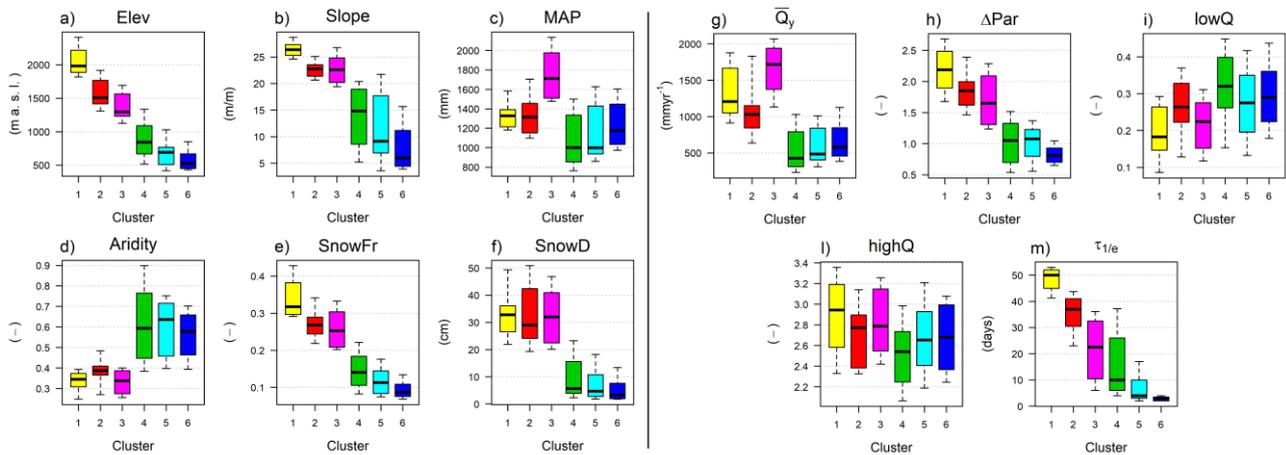


Figura 2. Variabilità all'interno dei cluster ottenuti di a) elevazione media, b) pendenza media, c) pioggia media annual, d) indice di aridità, e) frazione nevosa della precipitazione, f) accumulo nivale, g) portata media annua, h) stagionalità (coeff. Pardè), i) low flow (95° percentile), l) high flow (5° percentile) e m) memoria dell'idrogramma (stimata con integral scale).

L'approccio proposto è risultato in grado di distinguere il ruolo predominante o parziale di evapotraspirazione e scioglimento nivale sui diversi bacini della regione di studio; inoltre la $TE_{P \rightarrow Q}$ riesce a evidenziare differenze nella velocità di risposta del bacino, ma anche a catturare l'effetto di alte precipitazioni orografiche in bacini montani, con alto contributo nevoso. Tali risultati dimostrano le potenzialità della *transfer entropy* per la caratterizzazione dei bacini.

RIFERIMENTI BIBLIOGRAFICI

- Bennett, A., Nijssen, B., Ou, G., Clark, M., Nearing, G., 2019. Quantifying Process Connectivity With Transfer Entropy in Hydrologic Models. *Water Resources Research* 55, 4613–4629.
- Betterle, A., Schirmer, M., Botter, G., 2018. Cosa controlla la correlazione spaziale dei deflussi? Uso di un nuovo modello stocastico per individuare bacini idrologicamente simili.
- Di Prinzio, M., Castellarin, A., Toth, E., 2011. Data-driven catchment classification: application to the pub problem. *Hydrology and Earth System Sciences* 15, 1921–1935.
- Ehret, U., van Pruijssen, R., Bortoli, M., Loritz, R., Azmi, E., Zehe, E., 2020. Adaptive clustering: reducing the computational costs of distributed (hydrological) modelling by exploiting time-variable similarity among model elements. *Hydrol. Earth Syst. Sci.* 24, 4389–4411.
- Grimaldi, S., 2004. Linear Parametric Models Applied to Daily Hydrological Series. *Journal of Hydrologic Engineering* 9, 383–391.
- Keum, J., Coulibaly, P., 2017. Information theory-based decision support system for integrated design of multivariable hydrometric networks. *Water Resources Research* 53, 6239–6259.
- Lombardi, L., Toth, E., Castellarin, A., Montanari, A., Brath, A., 2012. Calibration of a rainfall-runoff model at regional scale by optimising river discharge statistics: Performance analysis for the average/low flow regime. *Physics and Chemistry of the Earth, Parts A/B/C* 42–44, 77–84.
- Schreiber, T., 2000. Measuring Information Transfer. *Physical Review Letters* 85, 461–464.
- Shannon, C.E., 1948. A mathematical theory of communication. *The Bell System Technical Journal* 27, 379–423.
- Singh, S.K., McMillan, H., Bárdossy, A., Fateh, C., 2016. Nonparametric catchment clustering using the data depth function. *Hydrological Sciences Journal* 61, 2649–2667.
- Toth, E., 2013. Catchment classification based on characterisation of streamflow and precipitation time series. *Hydrology and Earth System Sciences* 17, 1149–1159.
- Viglione, A., Claps, P., Laio, F., 2007. Mean annual runoff estimation in North-Western Italy. pp. 97–122.
- Viglione, A., Parajka, J., 2018. TUWmodel: Lumped Hydrological Model for Education Purposes.
- Ward, J.H., 1963. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association* 58, 236–244.