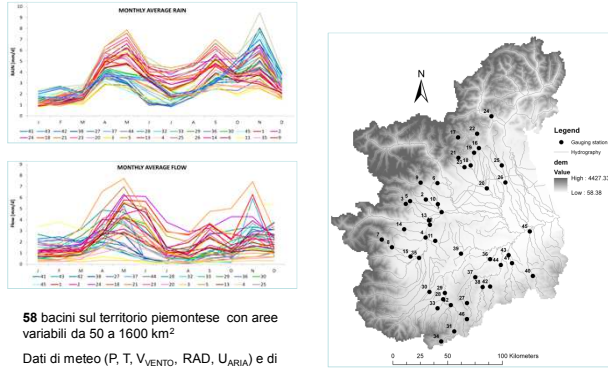


L. Boscarello¹, G. Ravazzani² e M. Mancini¹

¹Politecnico di Milano. Piazza Leonardo da Vinci, 32 Milano – e-mail: lauraanna.boscarello@polimi.it, giovanni.ravazzani@polimi.it, marco.mancini@polimi.it

LA CLASSIFICAZIONE DEI BACINI

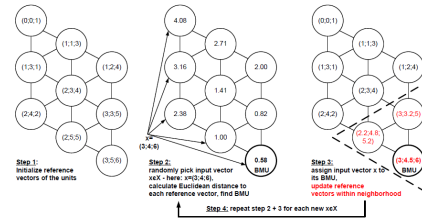
CASO STUDIO



58 bacini sul territorio piemontese con aree variabili da 50 a 1600 km²
Dati di meteo (P, T, V_{VENTO}, RAD, U_{ARIA}) e di portata dal 2000 al 2011
Disponibilità effettiva di dati di portata pari a circa 7 anni in media per bacino

Primo Livello: SELF ORGANIZING MAP

La **Self Organizing Map** o **Mapa Auto-Organizzante** è metodo di organizzazione di dati ad apprendimento non supervisionato.
Elaborato e sviluppato da Teuvo Kohonen dell'Università Tecnologica di Helsinki nel 1982. ("SOM-Toolbox for Matlab 5" [Vesanto et al., 2000; Herbst and Casper, 2008; Herbst et al., 2009].)
Essa produce una rappresentazione di dati ad elevata dimensione in una griglia a bassa dimensione, la rete neurale, costituita da neuroni a cui sono connessi tutti i vettori in ingresso.



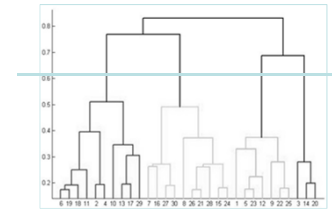
REFERENCES:
• Kohonen, T. (1995). *Self-Organizing Maps*. Berlin/Heidelberg, Germany: Springer, vol.30.
• Herbst, M., and C. Casper, (2008). *Towards model evaluation and identification using Self-Organizing Maps*. HESS, 12, 657-667.

Secondo Livello: HIERARCHICAL CLUSTERING

Il **metodo gerarchico agglomerativo** è un procedimento per l'organizzazione di dati ad apprendimento non supervisionato mediante un ricorsivo raggruppamento dei singoli dati in ingresso. È stato scelto l'approccio di Ward che non considera la distanza fra ogni cluster, ma tende a minimizzare la somma dei quadrati degli errori delle distanze all'interno dei gruppi:

$$SSE = \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$$

Il risultato è il **dendrogramma**, un diagramma ad albero, che mostra la gerarchia di partizione e i diversi livelli di similarità dei cluster. Per ottenere una specifica struttura, è necessario tagliare il dendrogramma al livello desiderato.



REFERENCES:
• Hannan, D.M., S.R. Kansakar, A.J. Gerrard, and G. Rees. (2005). *Flow regimes of Himalayan rivers of Nepal: nature and spatial patterns*. J. of Hydro., 308, 18-32.
• Köpflin, N., B. Schädel, D. Viviroli, and R. Weingartner. (2012). *Relating climate change signals and physiographic catchment properties to clustered hydrological response types*. HESS, 16, 2267-2283, doi:10.5194/hess-16-2267-2012.

INDICI IDROLOGICI

Indice	Descrizione	Formula	Unità
Q ₅₀	mediana della curva di durata normalizzata sull'area del bacino.	$\frac{Q_{50}}{A}$	[m ³ s ⁻¹ /km ²]
R _{QP}	rapporto tra la media annuale delle portate e la media annuale delle precipitazioni	$R_{QP} = \frac{Q}{P}$	-
E _{QP}	rapporto tra la variazione proporzionale annua del deflusso Q e la variazione proporzionale annua dell'altezza di precipitazione P.	$E_{QP} = \text{median} \left(\frac{dQ/P}{dP/Q} \right)$	-
BFI	rapporto tra il volume del deflusso di base e il volume delle portate totali.	$BFI = \sum_i \frac{q_b(i)}{q(i)}$	-
Q ₁₀	valore di portata che non è superato da più del 10% delle osservazioni	$\frac{Q_{10}}{Q_{50}}$	-
S _{FDC}	definita come il rapporto tra il 33° e il 66° percentile in scala semi-logaritmica	$S_{FDC} = \frac{\ln(Q_{33\%}) - \ln(Q_{66\%})}{0.66 - 0.33}$	-

REFERENCES:
• Oden, J.D., and N.L. Poff. (2003). *Redundancy and the choice of hydrologic indices for characterizing streamflow regimes*. River Res. Appl., 19, 101-121.
• Yadav, M., T. Wagener, and H. Gupta. (2007). *Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins*. Adv. in Water Resour., 30, 1756-1774.

INDICI FISICI

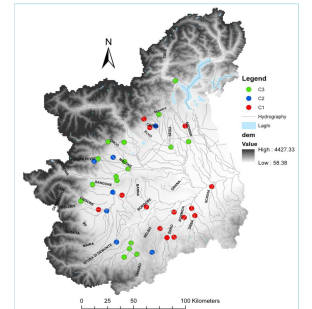
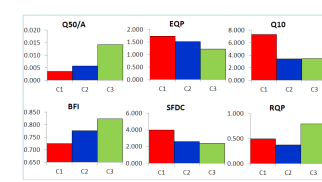
Per fornire una classificazione anche a bacini non strumentati, sono stati utilizzati indici fisici. Per la loro scelta si è proceduto con tre differenti approcci:
1. Valutazione della variabilità degli indici nelle classi idrologiche.
2. Stima del coefficiente di correlazione di Spearman tra indici fisici e idrologici.
3. Suggerimenti di autori da letteratura.

Metric	Description	Units	Min-Max
Climate			
CR	Ratio of average annual Precipitation to Evapotranspiration	-	1.103 – 3.423
RWD	Ratio of precipitation in wettest month to that of the driest	-	19.9 – 164.3
MMA	Median annual maximum 1-hour precipitation	mm/h	7.2 – 51.3
T-mo	Average monthly temperature	°C	3.5 – 12.5
MAP	Mean cumulated annual precipitation	mm	665 – 1756
NP	Number of days with precipitation	days/year	125 – 265
D.std	Standard deviation of annual precipitation	mm	227 – 680
Land use			
Agr	Percentage of arable land within the watershed	-	0.016 – 0.869
Wood	Percentage of woodland within the watershed	-	0.074 – 0.984
Physiographic			
Slope	Watershed steepness	-	0.9 – 17
H.5	5 percentile elevation	m	104 – 1017
H.95	95 percentile elevation	m	335 – 3006
Soil			
High.p	Percentage of soil with high permeability within the watershed	-	0.019 – 0.842
Low.p	Percentage of soil with low permeability within the watershed	-	0 – 0.625
FC	Watershed mean field capacity	-	0.164 – 0.76

REFERENCES:
• Yadav, M., T. Wagener, and H. Gupta. (2007). *Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins*. Adv. in Water Resour., 30, 1756-1774.
• Sanborn, S.C., and B.P. Bledsoe. (2006). *Predicting streamflow regime metrics for ungauged streams in Colorado, Washington, and Oregon*. J. of Hydro., 325, 241-261.

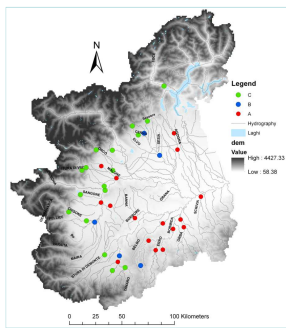
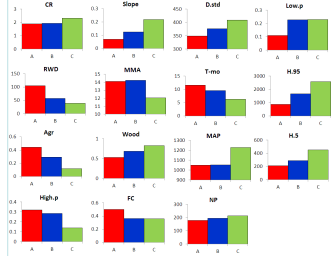
RISULTATI CLASSIFICAZIONE IDROLOGICA

Distribuzione degli indici idrologici in ogni classe



RISULTATI CLASSIFICAZIONE FISICA

Distribuzione degli indici fisici in ogni classe



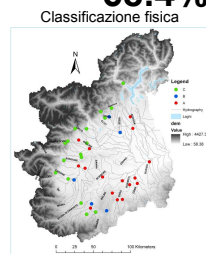
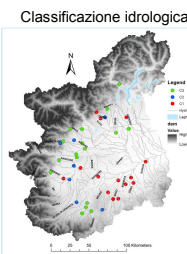
CLASSI A CONFRONTO

TABELLA DI CONTINGENZA

	A	B	C
C1	11	0	3
C2	2	3	2
C3	3	2	12

Percentuale di bacini ben classificati:

68.4%



VALIDAZIONE

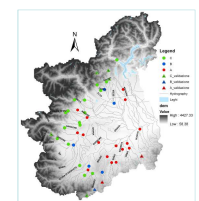
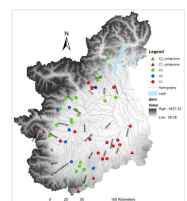
Calcolando la distanza euclidea di ciascun nuovo vettore dai neuroni delle mappe create dal SOM, si può identificare la classe di appartenenza, sia idrologica che fisica, di ogni nuovo bacino.

Classificazione idrologica

Bacino	C1	C2	C3	Classe
PERRERO	0.373	0.466	0.687	C1
BORGOGESIA	0.769	0.476	0.448	C3
CAMPERTOIGNO	0.722	0.573	0.207	C3
VARALLO	0.520	0.332	0.281	C3
GARESSIO	0.409	0.253	0.150	C3
PONTE DI NAVA	0.351	0.487	0.591	C1
ARQUATA	0.278	0.384	0.512	C1
MURIALDO	0.306	0.325	0.438	C1

Classificazione fisica

Bacino	A	B	C	Classe
PERRERO	0.86	0.70	0.59	C
BORGOGESIA	1.00	0.77	0.71	C
CAMPERTOIGNO	1.21	0.93	0.64	C
VARALLO	0.98	0.90	0.75	C
GARESSIO	0.57	0.45	0.51	C
PONTE DI NAVA	0.77	0.60	0.57	B
ARQUATA	0.54	0.86	1.21	A
MURIALDO	0.61	0.86	1.16	A



Percentuale di bacini ben classificati:
62.5%

REGIONALIZZAZIONE DELLE CURVE DI DURATA

MODELLO GLOBALE

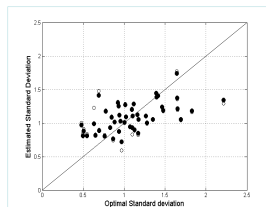
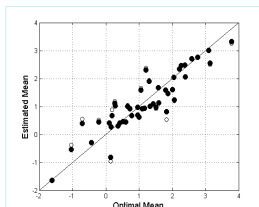
1° modello: 46 sezioni

$$\hat{\mu} = -14.878 + 1.136 \ln(MMA) + 1.026 \ln(A) + 0.944 \ln(H_{max} - H_{sta})$$

$$\hat{\sigma} = 4.561 + 0.034 \ln(Low.p) + 0.432 \ln(Wood) - 0.432 \ln(H_{max} - H_{sta})$$

R² = 0.819

R² = 0.325



Le performance nella stima dei parametri della distribuzione log-normale sono migliorabili sfruttando i risultati della classificazione?

MODELLO PER CLASSI IDROLOGICHE

2° modello:

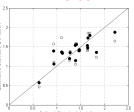
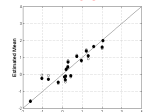
Classe Idrologica C1

$$\hat{\mu} = -29.9 + 0.679 \ln(RWD) + 2.949 \ln(MAP) + 1.175 \ln(A)$$

$$\hat{\sigma} = -7.148 - 0.695 \ln(Slope) + 0.494 \ln(RWD) - 0.552 \ln(Tmo) - 1.454 \ln(FC) + 0.674 \ln(H_{max} - H_{sta})$$

R² = 0.874

R² = 0.631



3° modello:

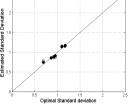
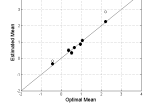
Classe Idrologica C2

$$\hat{\mu} = -15.596 + 1.537 \ln(MAP) + 0.978 \ln(A)$$

$$\hat{\sigma} = -0.383 - 0.103 \ln(RWD) + 0.240 \ln(H_{max} - H_{sta})$$

R² = 0.983

R² = 0.955



4° modello:

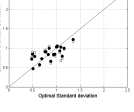
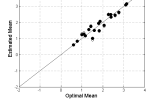
Classe Idrologica C3

$$\hat{\mu} = -3.189 + 1.076 \ln(MAP) - 1.668 \ln(NP) + 1.055 \ln(A)$$

$$\hat{\sigma} = -2.897 - 0.084 \ln(Mountain) - 0.102 \ln(High.p) + 0.866 \ln(NP) - 0.184 \ln(A)$$

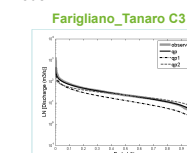
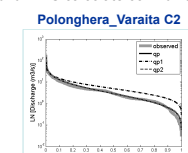
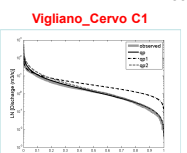
R² = 0.939

R² = 0.527



VERIFICA DEI RISULTATI

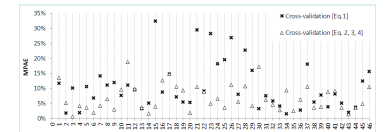
Esempi di FDC calcolate con i diversi modelli



Nel grafico: qp indica la CDP stimata con i valori ottimali di media e dev. Standard; qp1 indica la CDP stimata con il modello regionale costruito con tutti i 46 bacini, qp2 indica la CDP stimata con modello regionale dipendente dalla classe idrologica di appartenenza

Risultati Cross-Validazione

Modello	ε	σ _ε
1 - Generale	0.283	0.905
2 - Classe C1	0.156	0.470
3 - Classe C2	0.194	0.348
4 - Classe C3	0.061	0.291
Tot. (2+3+4)	0.119	0.374



CONCLUSIONS. This study tested the efficacy and the robustness that catchment classification can generate in FDCs regionalization respect to the classical regionalization procedure for a specific region.

First of all the classification framework was identified by using clustering techniques that identified three different classes in the analyzed region. This three classes highlight how in a relative small region, the differences in the hydrological regimes are quite evident; this fact is exploited to investigate possible improvements in the regionalization of FDCs.

Then, the observed FDCs were calculated from observed data and used to fit a lognormal distribution within the duration range 0.3 – 0.99 through a statistical approach. On the basis of the results of the fitting procedure, two kind of regional models were computed applying stepwise multiple linear regression. The first model was implemented considering all the 46 basins (the entire dataset), whereas the second one is composed from three models, each one for a single hydrological class obtained by cluster analysis. The performance and robustness of these models were checked through the leave-one-out cross-validation. The results show a great improvement of performance when the regionalization model is found by taking account of the three different hydrological classes, with a mean absolute percentage error that decreases from 11%, in the first case, to 7% in the second case. In conclusion, the differences in hydrological regimes are not always a consequence of geographical distance, so the spatial similarity is not always the best method to transfer information from a gauged basin to ungauged one. The subdivision in similar clusters seems to be a key point in defining a solid and good working model of regionalization. **ACKNOWLEDGEMENTS** authors acknowledge ARPA Piemonte (Regional Environment Protection Agency) for meteorological and flow data.