Mappatura di stati idroclimatici latenti nel karst mediterraneo mediante l'Hidden Markov Model

Filippo Gregori[1,2], Peter Naylor[3], Nataša Ravbar[4], Michaela De Giglio[2], Marco Dubbini[2]

[1] Dipartimento di Ingegneria Civile, Edile e Ambientale, Sapienza Università di Roma, Via Eudossiana 18, 00184 Roma; filippo.gregori@uniroma1.it
[2] Dipartimento di Storia Culture Civiltà (DiSCi), sezione Geografia, Università di Bologna, via Guerrazzi 20, 40125 Bologna; filippo.gregori3@unibo.it, michaela.degiglio@unibo.it, marco.dubbini@unibo.it
[3] Agenzia Spaziale Europea, ESRIN, Via Galileo Galilei, 00044 Frascati; peter.jack.naylor@gmail.com
[4] ZRC SAZU, Istituto di Ricerca sul Carso, Titov trg 2, 6230 Postojna, Slovenia; natasa.ravbar@zrc-sazu.si

Abstract. Un Hidden Markov Model gaussiano non supervisionato, applicato a sette variabili idroclimatiche mensili (1999–2016) in tre aree carsiche mediterranee, recupera dieci stati latenti ordinati da secco a umido, senza soglie predefinite. Gli stati formano unità spazialmente coerenti e una struttura di transizione quasi indipendente per area. Portata fluviale e albedo satellitare, mai usati dal modello, confermano il gradiente mappato.

Parole chiave: Hidden Markov Model, carso, mediterraneo, data-driven, climatic reanalysis, indice di aridità

1 Introduzione

Il bacino del Mediterraneo è un hotspot riconosciuto del cambiamento climatico: temperature in aumento, precipitazioni estive in calo, siccità sempre più intense. Il carso, che copre fino al 40% del territorio mediterraneo [1,2], è tra gli ambienti più esposti: suoli sottili, praticamente irrimpiazzabili su scala umana, e una connessione idraulica diretta tra superficie e sottosuolo che trasmette lo stress climatico quasi senza filtro dell'acquifero [3,4]. Il substrato carbonatico rende inoltre l'eventuale perdita di suolo sostanzialmente irreversibile alla scala temporale umana. Le mappe di riferimento oggi disponibili, l'indice di aridità United Nations Convention to Combat Desertification (UNCCD) [5], la classificazione di Köppen-Geiger e l'indice Mediterranean Desertification and Land Use, Environmental Sensitivity Area Index (MEDALUS ESAI), offrono ciascuna una classificazione consolidata e ampiamente utilizzata, assegnando un singolo indice o una singola classe a ogni località. Resta non rappresentata una dimensione ulteriore, complementare a questa classificazione: quanto essa sia certa in un dato momento e quanto tenda a durare nel tempo. Questo lavoro

propone una rappresentazione dello stato idroclimatico come una distribuzione di probabilità su un insieme di regimi latenti, stimati senza soglie predefinite né dati di addestramento etichettati, e corredata da incertezza e persistenza esplicite.

Il metodo impiegato, un Hidden Markov Model (HMM) gaussiano, è uno strumento statistico consolidato [6], non specifico del carso: lo stesso procedimento potrebbe essere applicato, senza modifiche, a qualunque paesaggio mediterraneo semi-arido. È il contesto carsico, con la sua connessione idraulica diretta e l'irreversibilità del danno al suolo, a rendere più gravi le conseguenze pratiche del segnale che il metodo rileva, non una specificità del metodo stesso.

2 Aree di studio e dati

Lo studio copre tre aree carsiche mediterranee (AOI) lungo un gradiente da secco a umido.

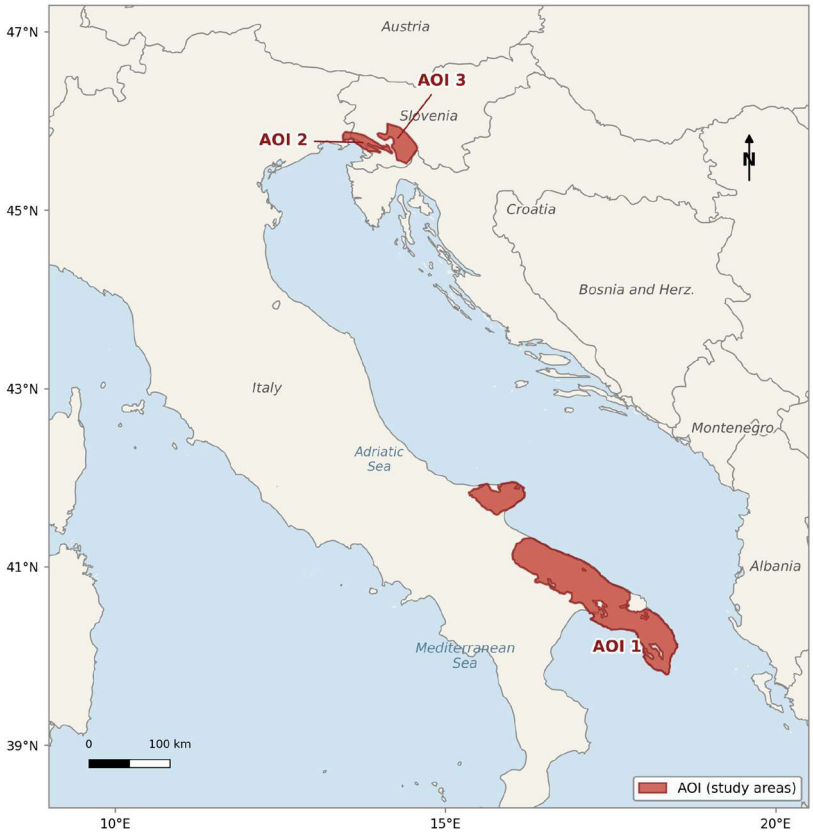

Fig. 1. Le tre aree di studio nel Mediterraneo centrale: Carso pugliese (AOI 1), Carso Classico (AOI 2) e Carso della Notranjska (AOI 3).

L'AOI 1, il Carso pugliese (Italia meridionale, circa 11.660 km²), ha un clima mediterraneo a estate calda (Köppen Csa), con un deficit idrico estivo marcato e suoli di Terra Rossa sottili e discontinui. L'AOI 2, il Carso Classico (confine Italia-Slovenia), è la località tipo da cui deriva il termine «karst» stesso, con un clima più umido (Cfb: temperato, senza stagione secca, estate calda) e una storia di spopolamento rurale e rimboschimento a pino nero. L'AOI 3, il Carso della Notranjska (Slovenia), ha un clima diviso tra le sue due classi di Köppen principali, Cfb (temperato, senza stagione secca, estate calda) e Dfb (continentale, senza stagione secca, estate calda), ed è dominato da un ampio acquifero carsico con un denso reticolo di cavità e polje (ampie depressioni carsiche a fondo pianeggiante, soggette ad allagamento stagionale), con foreste di faggio e abete gestite a copertura continua.

Le tre aree condividono un substrato carbonatico ma differiscono per clima, copertura forestale e struttura del suolo, offrendo un test naturale della generalità del metodo.

2.1 **Dataset**

Sette variabili idroclimatiche mensili (precipitazione, temperatura massima e minima, radiazione a onde corte, deficit di pressione di vapore, Standardised Precipitation-Evapotranspiration Index a 12 mesi (SPEI-12) [7], indice di umidità climatica (CMI) sono ricavate dalla rianalisi climatica Climatologies at High resolution for the Earth's Land Surface Areas (CHELSA) v2.1 [8] su griglia di ~1 km, da giugno 1999 a giugno 2016 (205 mesi). Ogni variabile è scomposta in due componenti distinte: la componente di base, una media mobile calcolata sui 12 mesi precedenti, che riflette la condizione climatica tipica e persistente del pixel; e la componente di anomalia, cioè quanto il valore di un singolo mese si discosta da questa base, per un totale di 14 feature standardizzate per pixel-mese. Albedo satellitare (Copernicus C3S, SPOT-VGT/PROBA-V) e portata fluviale osservata (stazioni idrometriche slovene sui fiumi Reka e Unica, anni idrologici 2000–2015, gli anni idrologici completi all'interno della finestra dati 1999–2016) sono assemblate come variabili indipendenti, mai fornite al modello, per una verifica a posteriori. Le aree edificate (Global Human Settlement Layer) sono escluse dalla maschera di validità, che conserva 18.290 pixel di terreno naturale o seminaturale sulle tre AOI.

3 Metodo

3.1 **Hidden Markov Model**

Un Hidden Markov Model è definito da tre componenti: un insieme di K stati latenti, non osservabili direttamente; una matrice di transizione A, dove ogni elemento dà la probabilità di passare dallo stato i allo stato j in un mese; e una distribuzione di

emissione per ciascuno stato: il profilo tipico delle 14 variabili climatiche quando il sistema si trova in quello stato, cioè i valori attesi e il modo in cui tendono a variare insieme. Qui si usa una gaussiana multivariata a covarianza «piena»: il modello include anche le correlazioni tra le variabili, non solo la loro variabilità individuale. I parametri sono stimati con l'algoritmo di Baum-Welch (un caso particolare di Expectation-Maximisation) [9] su un campione di 1.863 pixel, stratificato per AOI secondo il metodo di Cochran [10] per garantire una rappresentazione equa delle tre aree indipendentemente dalla loro estensione. Il modello stimato è poi applicato in sola inferenza a tutti i 18.290 pixel validi, tramite il decodificatore Maximum Posterior Marginal [11], che assegna ogni pixel-mese allo stato di probabilità a posteriori più alta.

3.2 **Selezione di K e caratterizzazione degli stati**

Il numero di stati è scelto su quattro criteri congiunti — riproducibilità tra semi casuali (Adjusted Rand Index), separabilità (distanza di Jeffreys-Matusita), assenza di stati di durata troppo breve, e certezza a posteriori media — anziché sul minimo di un singolo criterio statistico: Bayesian Information Criterion (BIC), Akaike Information Criterion (AIC) e Integrated Completed Likelihood (ICL) diminuiscono infatti monotonicamente da K=5 a K=14, senza un minimo interno, un comportamento già documentato in letteratura per modelli HMM costruiti per stati interpretabili piuttosto che per la sola stima di densità. Su questa base, K=10 è il valore scelto. Ogni stato è poi ordinato post-hoc lungo l'indice di aridità UNCCD (AI = P/PET, con P la precipitazione media e PET l'evapotraspirazione potenziale media), calcolato dalle sue medie di emissione fisiche, da Classe 1 (la più secca) a Classe 10 (la più umida). La Tabella 1 riassume la posizione lungo il gradiente dei dieci stati; la Fig. 5 ne mostra i valori medi di baseline per le sette variabili climatiche di base.

Tab. 1. Indice di aridità (AI = P/PET) e fascia United Nations Environment Programme (UNEP) / UNCCD [12] per i dieci stati.

Stato	**AI**	**Fascia UNEP/UNCCD**	**AOI dominante**
Classe 1	0.41	Semi-arido	Puglia (100%)
Classe 2	0.57	Sub-umido secco	Puglia (100%)
Classe 3	0.58	Sub-umido secco	Puglia (100%)
Classe 4	1.52	Umido	Carso Classico (99.8%)
Classe 5	1.82	Umido	Notranjska (96.9%)
Classe 6	1.83	Umido	Carso Classico (100%)
Classe 7	1.94	Umido	Carso Classico / Notranjska
Classe 8	2.24	Umido	Carso Classico (99.8%)
Classe 9	2.31	Umido	Notranjska (90.5%)
Classe 10	2.44	Umido	Notranjska (99.3%)

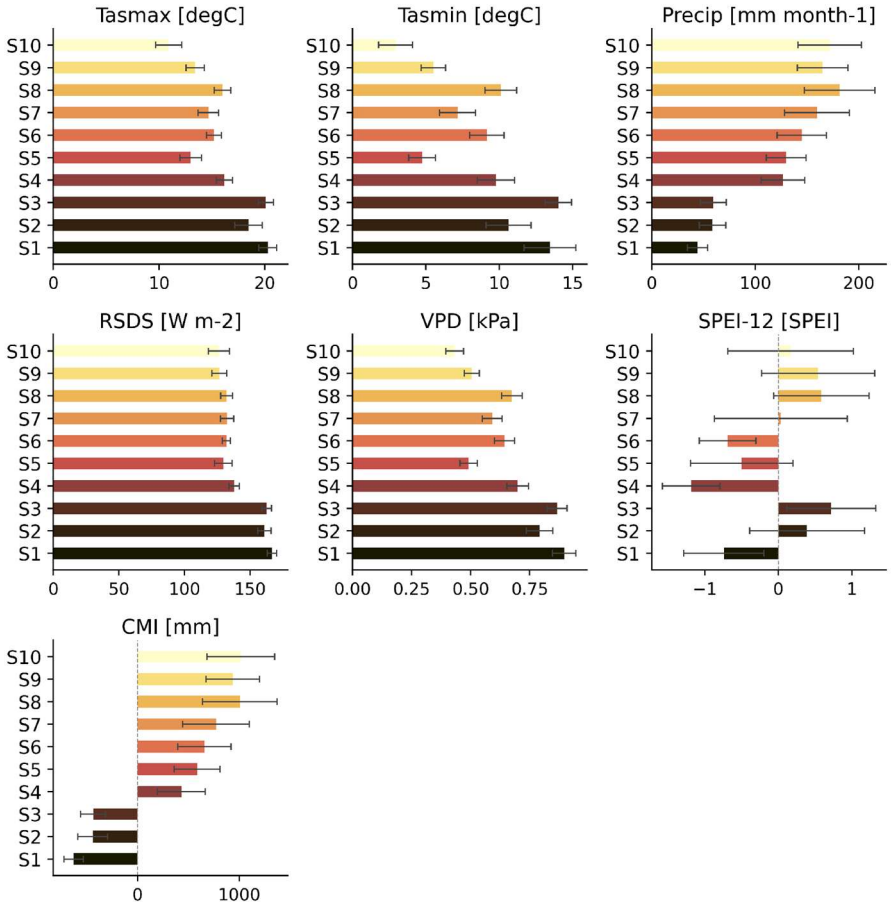

Fig. 5. Valori medi di baseline (± deviazione standard) delle sette variabili climatiche di base, per ciascuno dei dieci stati, in unità fisiche: temperatura massima e minima (Tasmax, Tasmin), precipitazione (Precip), radiazione a onde corte (RSDS), deficit di pressione di vapore (VPD), SPEI-12 e indice di umidità climatica (CMI).

4 Risultati

4.1 Distribuzione spaziale degli stati

I dieci stati recuperati si distribuiscono coerentemente con la geografia reale, pur non avendo mai ricevuto l'informazione di appartenenza ad un'area: tre stati (i più secchi, Classi 1-3) sono quasi esclusivi del Carso pugliese, gli altri sette si ripartiscono tra le due aree slovene. Gli stati disegnano regioni spazialmente coerenti sulla mappa (Fig. 2), non un mosaico casuale di pixel, un'ulteriore evidenza diretta che il modello segue un campo climatico spazialmente coerente e non rumore a livello di singolo pixel.

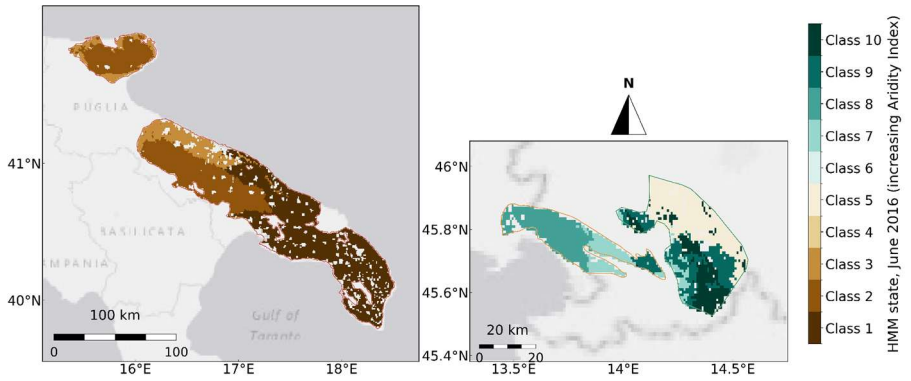

Fig. 2. Distribuzione spaziale dello stato decodificato a giugno 2016, dieci stati ordinati per indice di aridità, palette marrone per gli stati più secchi e verde-blu per quelli più umidi [13].

4.2 Struttura delle transizioni

La matrice di transizione stimata (Fig. 3) è quasi a blocchi, e i blocchi coincidono con la ripartizione spaziale tra AOI, non con un raggruppamento imposto al modello: su 3.731.160 transizioni mensili decodificate nell'intero periodo, solo 54 attraversano il confine del blocco pugliese: un isolamento dinamico quasi totale. Tra le due aree slovene i pixel passano dall'una all'altra in modo più permeabile ma comunque debole (0.03–0.25% al mese per cella). Ogni stato tende a persistere per molti mesi prima di cambiare: il tempo di permanenza teorico varia da circa 10 mesi (Classe 4, la meno stabile) a oltre 85 mesi, oltre sette anni (Classe 10, la più stabile), coerentemente con un sistema climatico che evolve con continuità, non a scatti casuali.

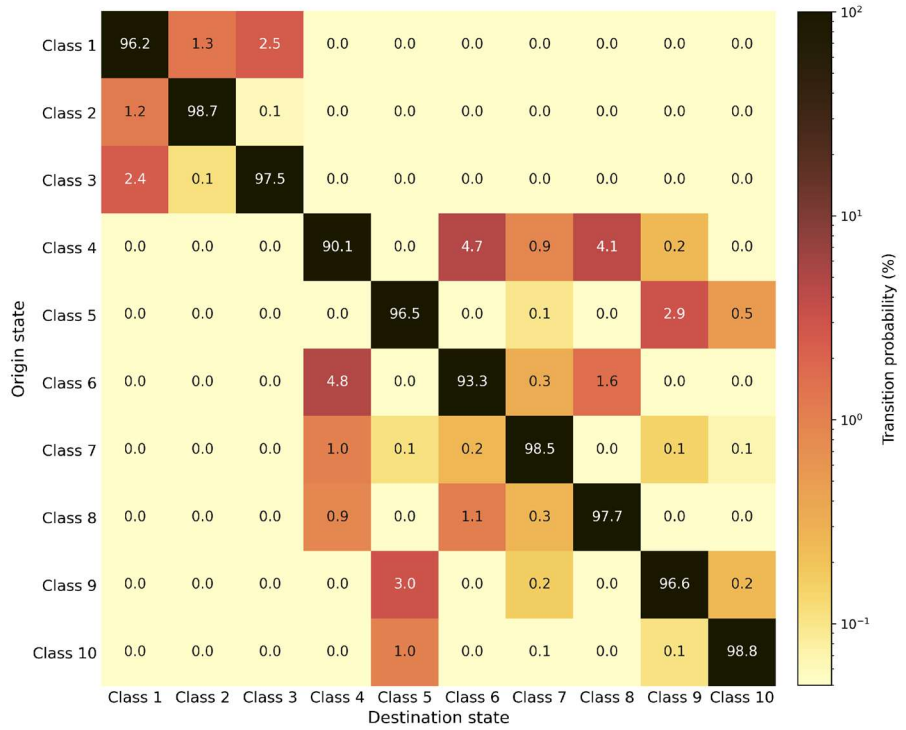

Fig. 3. Matrice di transizione (probabilità % per cella, scala di colore logaritmica), riga = stato di origine, colonna = stato di destinazione, entrambi in ordine di indice di aridità.

4.3 **Verifica indipendente**

Il gradiente mappato trova conferma nelle due variabili indipendenti introdotte nel §2.1. L'albedo satellitare correla fortemente con l'indice di aridità per stato ($r=-0.90$, ρ di Spearman$=-0.79$): gli stati più secchi riflettono sistematicamente più luce, coerentemente con una minore copertura vegetale e una maggiore esposizione del suolo nudo. La portata fluviale osservata, nelle due aree slovene dotate di idrometro, correla in modo significativo con l'indice di aridità ponderato a posteriori dal modello, aggregato per anno idrologico (ρ di Spearman$=0.81$ nel Carso Classico e 0.82 nella Notranjska, $p<0.001$ in entrambi i casi, Fig. 4): quando il modello indica una fase più umida, il fiume porta davvero più acqua.

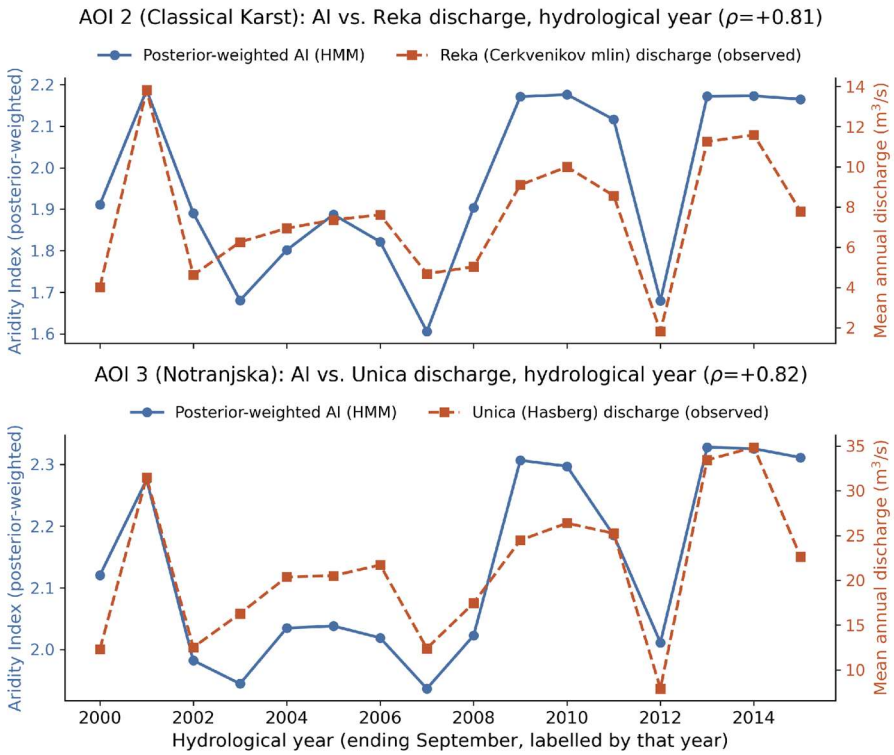

Fig. 4. Indice di aridità ponderato a posteriori (blu) contro portata fluviale osservata (arancio), per anno idrologico, fiumi Reka (AOI 2) e Unica (AOI 3).

4.4 **Controlli di robustezza**

Sei controlli indipendenti verificano la tenuta del modello. Un test di esclusione temporale (addestramento 1999–2011, verifica 2012–2016) mostra una log-verosimiglianza predittiva, quanto bene il modello si adatta a dati che non ha mai visto in fase di stima, leggermente migliore sui mesi esclusi dall'addestramento che su quelli usati per stimarlo, evidenza contro il sovradattamento. La sensibilità a K=9 e K=11 conferma la stessa struttura a blocchi. Un dubbio legittimo è che il modello aggregato stia semplicemente riconoscendo a quale AOI appartiene un pixel, senza misurare nulla di realmente climatico: per escluderlo, l'intera procedura di selezione e stima è stata ripetuta indipendentemente su ciascuna AOI, isolandola dalle altre due, senza accesso alcuno alle altre. Se il modello aggregato si limitasse effettivamente a rietichettare la posizione geografica, isolare ogni area dovrebbe sistematicamente peggiorare la distinzione tra gli stati, poiché verrebbe a mancare proprio l'informazione geografica su cui tale distinzione si baserebbe. Non è ciò che si osserva: il risultato è quantificato dal rapporto di sovrapposizione, un indice che misura quanto gli stati si sovrappongano tra loro (valori prossimi a zero indicano stati ben separati, valori prossimi a 1 indicano

stati indistinguibili). Isolando singolarmente ciascuna area, il rapporto è 0.25 in Puglia, 0.09 nel Carso Classico e 0.16 nella Notranjska, contro 0.22 nel modello aggregato: in due casi su tre la separazione migliora rispetto al modello aggregato, e nel terzo peggiora solo lievemente. Se la separazione tra stati dipendesse dall'accesso simultaneo alle tre aree, isolarle dovrebbe deteriorarla in modo sistematico; non è quanto si osserva, il che esclude che il modello si limiti a rietichettare l'appartenenza geografica. Questi controlli, insieme alla verifica indipendente del §4.3, forniscono un quadro coerente di un segnale fisico genuino, non di un artefatto statistico.

5 Conclusioni

Un Hidden Markov Model non supervisionato, applicato a sole variabili climatiche globalmente disponibili, recupera una struttura idroclimatica latente coerente con la geografia e confermata da osservazioni fisiche esterne al modello stesso. Il contributo cartografico originale è la mappa della distribuzione di probabilità sugli stati, con incertezza e persistenza esplicite, invece del singolo indice o classe delle mappe di riferimento oggi in uso.

Tre direzioni di sviluppo futuro emergono direttamente dai limiti del lavoro. La prima riguarda la scala: ri-addestrare l'HMM su circa 900.000 pixel (ERA5-Land + MODIS, dal 1950 a oggi) per produrre la prima mappa mensile dello stato idroclimatico latente per l'intero karst mediterraneo e confrontare aree con caratteristiche idroclimatiche simili. La seconda riguarda la struttura del modello: un'estensione semi-markoviana, con tempi di permanenza esplicitamente dipendenti dalla durata già trascorsa in uno stato. La terza riguarda la validazione: estendere il confronto con la portata fluviale e dati dataset da sorgenti di altri bacini carsici mediterranei, e verificare se la relazione con l'albedo osservata qui regga anche fuori dalle tre aree studiate.

Riferimenti bibliografici

1. Xanke, J., Goldscheider, N., Bakalowicz, M., Barberá, J.A., Broda, S., Chen, Z., Ghanmi, M., Günther, A., Hartmann, A., Jourde, H., Liesch, T., Mudarra, M., Petitta, M., Ravbar, N., Stevanović, Z.: Carbonate rocks and karst water resources in the Mediterranean region. Hydrogeology Journal 32, 1397–1418 (2024).
2. Goldscheider, N., Chen, Z., Auler, A.S., Bakalowicz, M., Broda, S., Drew, D., Hartmann, J., Jiang, G., Moosdorf, N., Stevanovic, Z., Veni, G.: Global distribution of carbonate rocks and karst water resources. Hydrogeology Journal 28, 1661–1677 (2020).
3. D'Ettorre, U.S., Liso, I.S., Parise, M.: Desertification in karst areas: A review. Earth-Science Reviews 253, 104786 (2024).
4. Lavee, H., Imeson, A.C., Sarah, P.: The impact of climate change on geomorphology and desertification along a Mediterranean-arid transect. Land Degradation & Development 9(5), 407–422 (1998).
5. United Nations: United Nations Convention to Combat Desertification in those countries experiencing serious drought and/or desertification, particularly in Africa. United Nations, Paris (1994).
6. McClintock, B.T., Langrock, R., Gimenez, O., et al.: Uncovering ecological state dynamics with hidden Markov models. Ecology Letters 23(12), 1878–1903 (2020).
7. Vicente-Serrano, S.M., Beguería, S., López-Moreno, J.I.: A multiscalar drought index sensitive to global warming: The Standardized Precipitation Evapotranspiration Index. Journal of Climate 23(7), 1696–1718 (2010).
8. Karger, D.N., Conrad, O., Böhner, J., et al.: Climatologies at high resolution for the Earth's land surface areas. Scientific Data 4, 170122 (2017).
9. Rabiner, L.R.: A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE 77(2), 257–286 (1989).
10. Cochran, W.G.: Sampling Techniques. 3rd edn. Wiley, New York (1977).
11. Zucchini, W., MacDonald, I.L., Langrock, R.: Hidden Markov Models for Time Series: An Introduction Using R. 2nd edn. CRC Press, Boca Raton (2016).
12. Middleton, N.J., Thomas, D.S.G. (eds.): World Atlas of Desertification. 2nd edn. UNEP/Arnold, London (1997).
13. Crameri, F., Shephard, G.E., Heron, P.J.: The misuse of colour in science communication. Nature Communications 11, 5444 (2020).