Hidden Markov Models for Desertification Assessment in Mediterranean Karst Ecosystems

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

[1]Sapienza Università di Roma [2]Università di Bologna [3]European Space Agency [4]ZRC SAZU Karst Research Institute

1. Introduction

A large portion, as much as 40% of the Mediterranean basin, is karst, characterised by thin soils, soluble bedrock, and high hydrological connectivity. To plan for life in such areas and identify emerging threats, a systematic assessment of desertification is needed. Existing frameworks classify current condition but cannot characterise trajectories: they reveal *where* a system stands, not *where it is heading*. Such dynamics become explicit only from a transition matrix estimated on a multivariate time series.

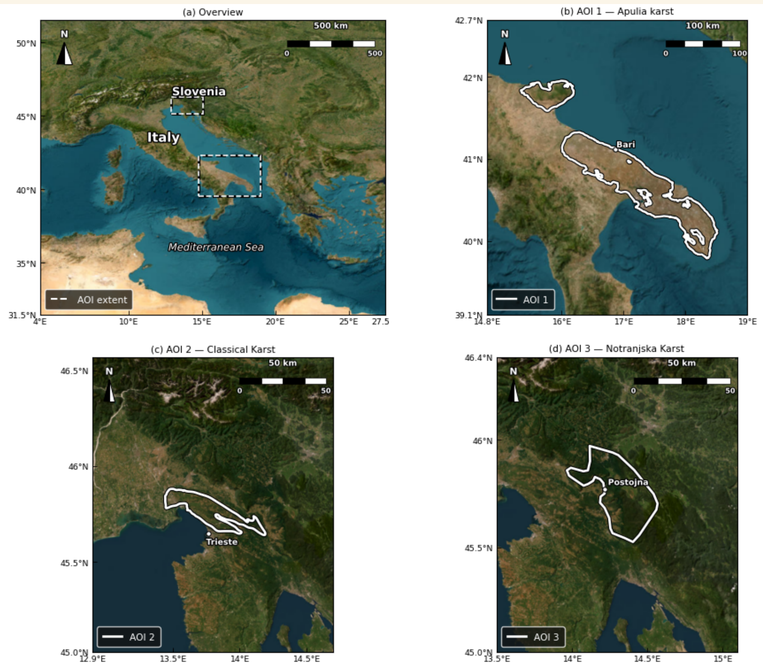

Figure 1. AOI 1 (Apulia, Italy), AOI 2 (Classical Karst), AOI 3 (Notranjska, Slovenia).

	AOI 1 Apulia	**AOI 2** Class. Karst	**AOI 3** Notranjska Karst
Area (km^2)	∼11,400	∼520	∼1,030
Elevation (m)	0–1,055	200–700	400–1,300
Mean T (°C)	∼16	∼11	∼9
Mean P (mm yr^{-1})	450–700	1,000–1,800	1,800–2,000
Soil	Terra Rossa	Rendzina, Cambisol	Deeper Cambisol
Forest (%)	∼7.5	∼73	>58
Land cover	Olive, crops	Forest, grassland	Fir-beech forest
Land use	Agric., graz.	Rural abandonment	Forestry

Table 1. Key physio-climatic characteristics of the three AOIs.

2. Data

18,290 pixels × 205 months (June 1999–June 2016). Each variable is decomposed into a 12-month *baseline* and a monthly *anomaly* (18 features/pixel-month), then standardised to *z*-scores.

Variable	Description	Unit	Type
Tasmax	Max air temperature	°C	Climate
Tasmin	Min air temperature	°C	Climate
CMI	Climatic Moisture Index	mm yr^{-1}	Climate
Pr	Precipitation	mm mo^{-1}	Climate
RSDS	Shortwave radiation	W m^{-2}	Climate
VPD	Vapour-pressure deficit	kPa	Climate
SPEI-12	12-month drought index	z-score	Climate
FAPAR	Fraction absorbed PAR	0–1	Satellite
Albedo	Surface albedo	0–1	Satellite

Table 2. Input variables (∼1 km, monthly).

3. Methods: HMM Structure

We apply a **Multivariate Gaussian HMM** (K = 10 states, trained via Expectation-Maximisation, selected by BIC) to globally available remote-sensing and reanalysis data to infer latent ecological states without field measurements. HMM is well-suited to this task: it handles multivariate time series with temporal dependencies; emission profiles and transition probabilities are directly interpretable, unlike black-box approaches.

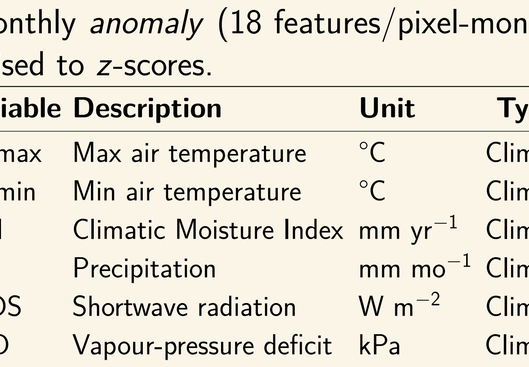

Figure 2. HMM structure: hidden states (ellipses), observations (rectangles), transitions (solid), emissions (dashed).

4. Results – Ecological States

Each point represents one pixel; clusters reflect distinct emission profiles along the ecological gradient.

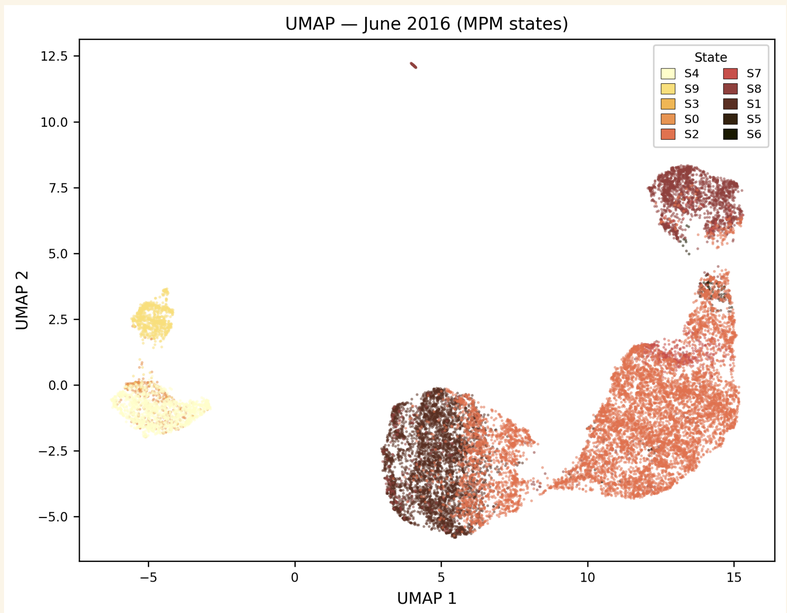

Figure 3. UMAP embedding of all pixel-months (June 2016).

Each state is characterised by a multivariate Gaussian emission distribution. Baseline emission means (SD) in physical units are reported per AOI.

(AOI 1 – Apulia)

Feature	Unit	S6	S5	S1	S8	S7	S2
Tasmax	°C	20.0±2.8	19.9±2.4	20.4±2.0	17.8±2.4	19.2±2.0	19.5±2.5
Tasmin	°C	12.2±3.5	12.1±2.4	14.6±2.6	11.8±2.9	11.8±2.5	12.3±2.9
Precip	mm mo^{-1}	40±21	48±19	57±31	55±32	60±32	63±33
VPD	kPa	0.89±0.08	0.88±0.18	0.89±0.11	0.79±0.14	0.80±0.10	0.83±0.12
SPEI-12	z-score	−0.73±0.79	0.03±0.80	0.28±0.76	0.01±0.93	0.99±0.46	−0.07±0.86
CMI	mm yr^{-1}	−662±263	−573±266	−492±379	−448±408	−423±390	−401±426
Albedo	0–1	0.175±0.017	0.175±0.010	0.167±0.013	0.157±0.016	0.173±0.014	0.173±0.018

Table 3a. Baseline emission means (SD), AOI 1.

(AOI 2 – Classical Karst & AOI 3 – Notranjska)

Feature	Unit	S0	S3	S9	S4
Tasmax	°C	13.2±3.5	12.7±3.7	15.6±2.9	12.2±2.6
Tasmin	°C	4.9±3.5	4.7±6.5	9.4±3.9	4.3±2.7
Precip	mm mo^{-1}	139±58	150±85	159±78	161±48
VPD	kPa	0.51±0.15	0.49±0.06	0.66±0.13	0.46±0.11
SPEI-12	z-score	−0.05±1.08	0.00±1.06	−0.17±1.01	0.11±0.81
CMI	mm yr^{-1}	+646±757	+777±1023	+783±963	+915±628
Albedo	0–1	0.142±0.013	0.149±0.062	0.152±0.014	0.151±0.015

Table 3b. Baseline emission means (SD), AOI 2+3.

The transition matrix shows strong diagonal dominance (high state persistence); main off-diagonal flows: S6↔S5 (∼12%), S4↔S3 (∼11–15%), S2→S7 (7.3%).

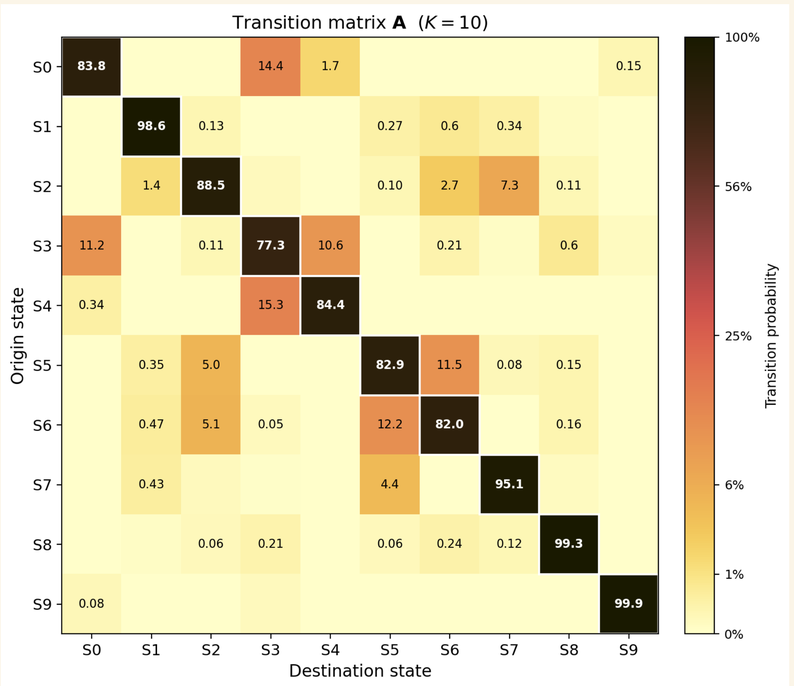

Figure 4. Transition probability matrix **A** (log scale).

States are ranked by CMI (worst to best: S6, S5, S1, S8, S7, S2, S0, S3, S9, S4).

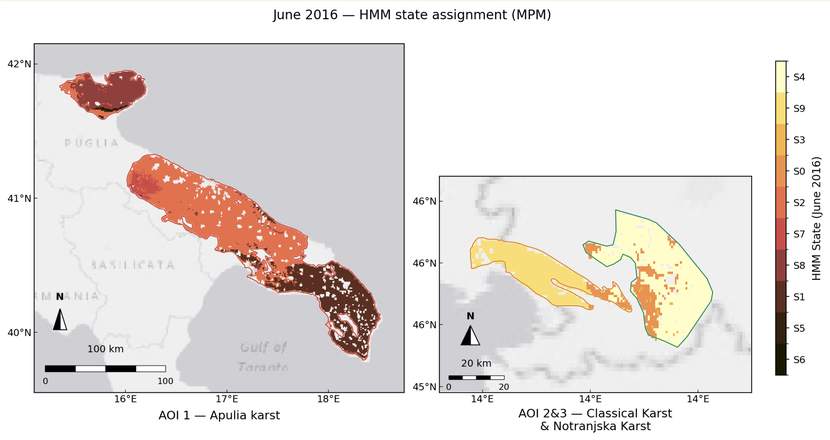

Figure 5. MPM-inferred HMM state per pixel at June 2016.

5. Results – Projection to 2025

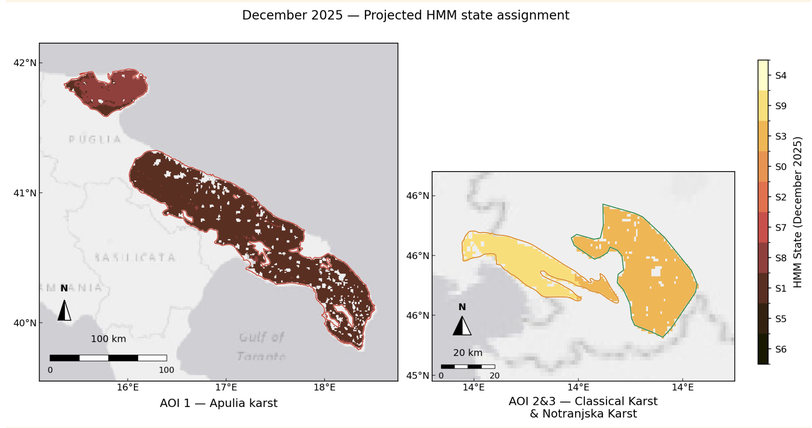

Figure 6. Projected HMM state per pixel at Dec 2025.

The stationary distribution $\boldsymbol{\pi}^*$ is the long-run equilibrium the system tends to regardless of initial conditions. The projection converges to it within ∼2 years: S2 collapses from 46% to 10.6%, while S7 rises from 3% to 18%.

6. Discussion

The HMM is **unsupervised**: states are inferred from the data alone with no ecological label. The physical emission values of each state align with independent classifications (MEDALUS, Köppen, local stations), confirming that the quality rank describes real ecological conditions. This correspondence enables objective assessment of ecological health from remote sensing alone, without field measurements.

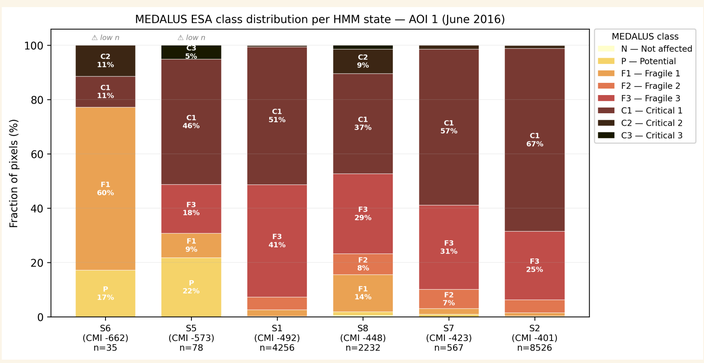

Figure 7. MEDALUS ESA class distribution per HMM state (AOI 1).

- Degraded states (S1, S2) fall in Köppen Csa in 99%/93% of pixels; resilient states (S0, S4) in Cfb/Dfb in >95%.
- S2 is out of equilibrium (46%→10.6%), signalling an ongoing shift toward drier regimes.
- States with $n < 100$ pixels (S5, S6, S3) cannot be independently validated due to insufficient spatial coverage.
- Limitations: scarce local ground data; states are correlation-based with no causal mechanism encoded.
- The 1 km timeseries covers only 1999–2016 (17 years); extending to Mediterranean scale at ∼9 km would allow substantially longer records.

7. Conclusions

- Unsupervised ML on remote-sensing data provides a scalable, field-free framework; hidden states align with Köppen, MEDALUS, and station data — operationally meaningful without ground surveys.
- Rapid convergence to $\boldsymbol{\pi}^*$ (∼2 years) indicates limited year-specific predictive power beyond long-run equilibrium.
- The HMM captures temporal dynamics but hidden states must be characterised in post-processing using soil properties and independent physical datasets.
- The method is transferable to the broader Mediterranean basin and, in principle, to karst regions globally where the same datasets are available.

References

Beck, H., McVicar, T., Vergopolan, N., Berg, A., Lutsko, N., Dufour, A., Zeng, Z., Jiang, X., van Dijk, A., and Miralles, D. (2023). High-resolution (1 km) Köppen-Geiger maps for 1901–2099 based on constrained CMIP6 projections. *Scientific Data*, 10:724.

Hasselmann, K. (1976). Stochastic climate models Part I. Theory. *Tellus*, 28(6):473–485.

Kosmas, C., Ferrara, A., Briasouli, H., and Imeson, A. (1999). Methodology for mapping Environmentally Sensitive Areas (ESAs) to desertification. In *The MEDALUS Project: Mediterranean Desertification and Land Use*, pages 31–47. European Commission. EUR 18882.

Murphy, K. (2012). *Machine Learning: A Probabilistic Perspective*. MIT Press, Cambridge, MA.