HMM Landcover Transition Analysis Report

Hidden Markov Model Analysis of Landcover Transitions (2017-2024)

2025-11-11

1 Executive Summary

This report presents Hidden Markov Model (HMM) analysis results for landcover transition patterns across 14 agro-ecological regions (AERs) in India, covering the period 2017-2024.

Dataset Overview:

- **Total Sequences:** 148,475
- **Regions Analyzed:** 14
- **Landcover Classes:** 11 (L2 hierarchy: agri_open, bare_rocky, built, cultivated_trees, dunes, forest, saline_flat, savanna_grass, savanna_shrub, savanna_tree, water_wetland)
- **Temporal Coverage:** 8 years (2017-2024)
- **Model Type:** CategoricalHMM with empirical emissions and hierarchical blending

Model Approach:

- **Emissions:** Learned from ground truth data (`classLabel_sampled` vs `classLabel_predicted`) to model classifier confusion
- **Transitions:** Learned from temporal sequences to model real land cover dynamics
- **Hierarchical Blending:** Region-specific matrices (β=0.4) blended with global baseline to ensure valid 11×11 outputs

Summary Statistics by Region

Region	Region Name	N Sequences	Stability	Diversity
2	Rajasthan Bagar (Western Plain)	15,967	0.943	0.850
3	Karnataka Plateau	2,957	0.933	0.740
4	North Punjab Plain	20,373	0.847	0.644
5	Coastal Kathiawar Peninsula	11,446	0.857	0.760
6	Eastern Maharashtra Plateau	20,409	0.901	0.685
7	South Telangana Plateau (Rayalseema) and Eastern Ghats	10,881	0.948	0.786
8	Tamil Nadu Uplands and Plains	8,740	0.941	0.793
9	Punjab and Rohilkhand Plains	7,529	0.899	0.531
10	Vindhyan Scarpland and Baghelkhand Plateau	14,561	0.903	0.764
11	Eastern Plateau (Chhattisgarh Region)	6,483	0.934	0.845
12	Chhotanagpur Plateau and Garjat Hills	12,073	0.910	0.761
13	Foothills of Central Himalayas	6,576	0.933	0.533
18	North Tamil Nadu Plains (Coastal)	4,499	0.848	0.595
19	Central and South Sahyadris	5,981	0.945	0.777

2 Classification Accuracy Analysis

2.1 Overview

The HMM framework separates two sources of uncertainty:

1. **Emission probabilities** (P(observed | true)): Model classifier confusion and errors
2. **Transition probabilities** (P(class_t+1 | class_t)): Model real land cover dynamics over time

This separation is critical because classifier errors are systematic (not random) and must be explicitly modeled to avoid conflating measurement noise with genuine land cover change.

2.2 Global Classification Performance

Global Accuracy:

- **Overall Accuracy:** 75.0% — Moderate accuracy indicates substantial classifier uncertainty that must be explicitly modeled
- **Cohen's Kappa:** 0.678 — Indicates moderate agreement beyond chance; substantial room for confusion
- **Validation Samples:** 1,187,800 — Large sample ensures robust emission matrix estimates

Per-Class Accuracy Metrics

Class	Producer.s.Accuracy[1]	User.s.Accuracy[1]	F1.Score
agri_open	0.919	0.843	0.879
bare_rocky	0.210	0.696	0.323
built	0.793	0.771	0.782
cultivated_trees	0.573	0.543	0.557
dunes	0.691	0.614	0.650
forest	0.869	0.709	0.781
saline_flat	0.794	0.862	0.826
savanna_grass	0.612	0.717	0.660
savanna_shrub	0.519	0.571	0.543
savanna_tree	0.493	0.718	0.584
water_wetland	0.847	0.857	0.852

[1] Producer's Accuracy = recall (% of true class correctly identified); User's Accuracy = precision (% of predictions that are correct)

Interpretation: Classes with F1 < 0.5 show substantial confusion and require careful interpretation in transition analysis. High-accuracy classes (F1 > 0.7) provide more reliable transition estimates. The emission matrices visible in regional heatmaps capture these class-specific error patterns.

2.3 Emission Matrix Interpretation

The emission matrices (shown in regional heatmaps) represent P(predicted | true class) and reveal:

- **Diagonal elements**: Classification accuracy for each class (higher = more accurate)
- **Off-diagonal elements**: Confusion patterns (which classes get misclassified as which)
- **Regional variation**: Different regions show different classification challenges

2.4 Key Observations for Landcover Transition Analysis

Based on the accuracy analysis, the following considerations apply to interpreting transition results:

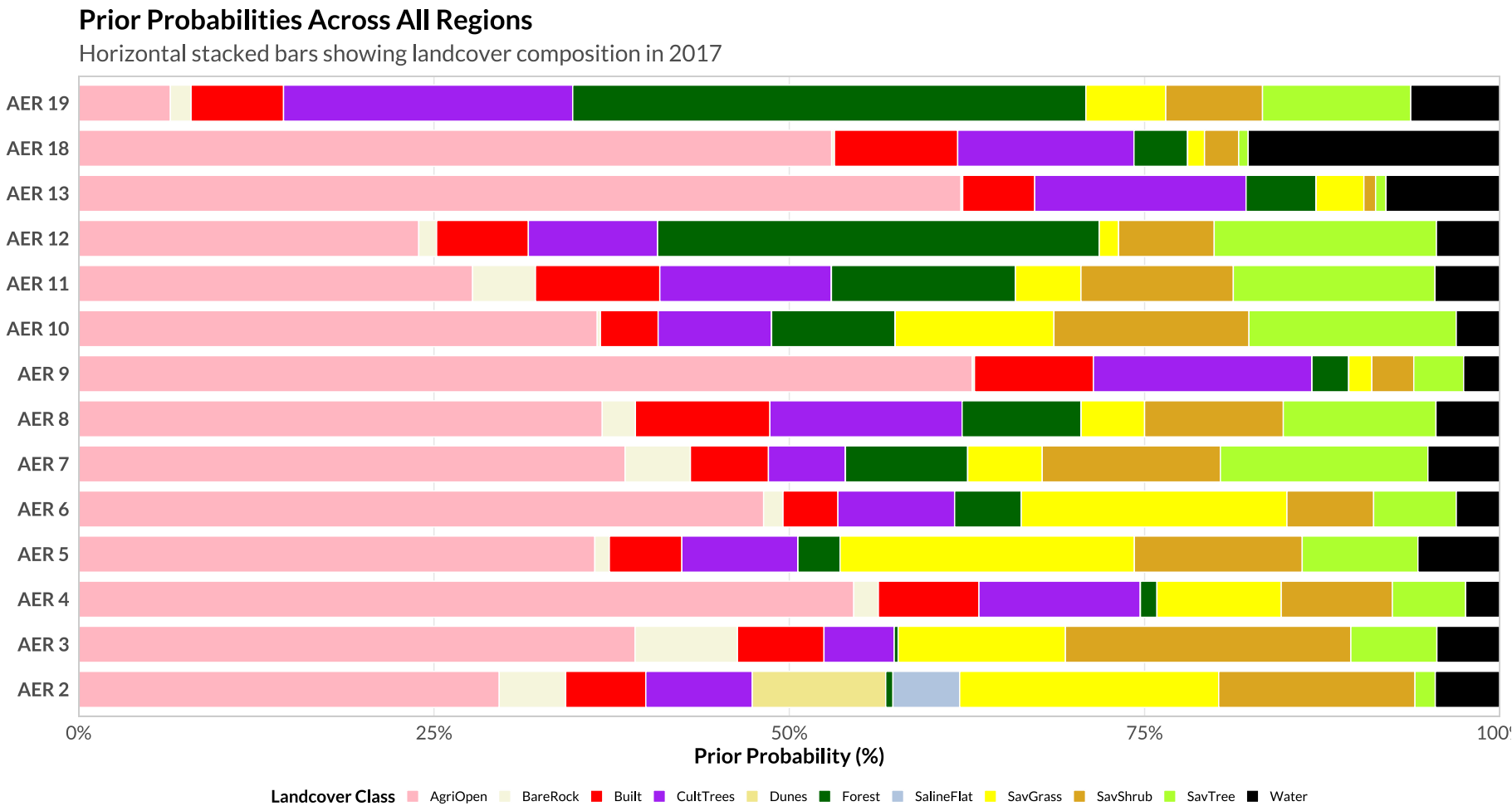
1. **Moderate Global Accuracy (64%)**: The classifier has substantial uncertainty, making empirical emission matrices essential. Without modeling this uncertainty, apparent "transitions" would conflate classifier errors with real land cover change.
2. **Class-Specific Performance**: Per-class accuracy varies considerably (see table above). Classes with low accuracy (e.g., <0.5) are more prone to spurious transitions in raw classifications. The HMM's emission matrices account for this by down-weighting unreliable observations.

3. **Hierarchical Blending Strategy**: Region-specific emission matrices are blended with global baselines (β=0.30) to balance regional adaptation with statistical stability, especially for regions with fewer validation samples.
4. **Transition Matrix Reliability**: After accounting for classifier uncertainty via emission matrices, the learned transition probabilities represent genuine land cover dynamics. High diagonal values (persistence) in transition matrices indicate stable landscapes, not classifier consistency.
5. **Implications for Downstream Use**: The trained HMM parameters (transitions + emissions) can be used for Viterbi smoothing of new time series, applying both spatial context (regional transitions) and measurement error correction (emissions) simultaneously.

3 Prior Probabilities

Prior probabilities represent the initial landcover distribution at the start of the time series (2017). Higher values indicate more prevalent landcover types in each region.

3.1 All Regions Overview



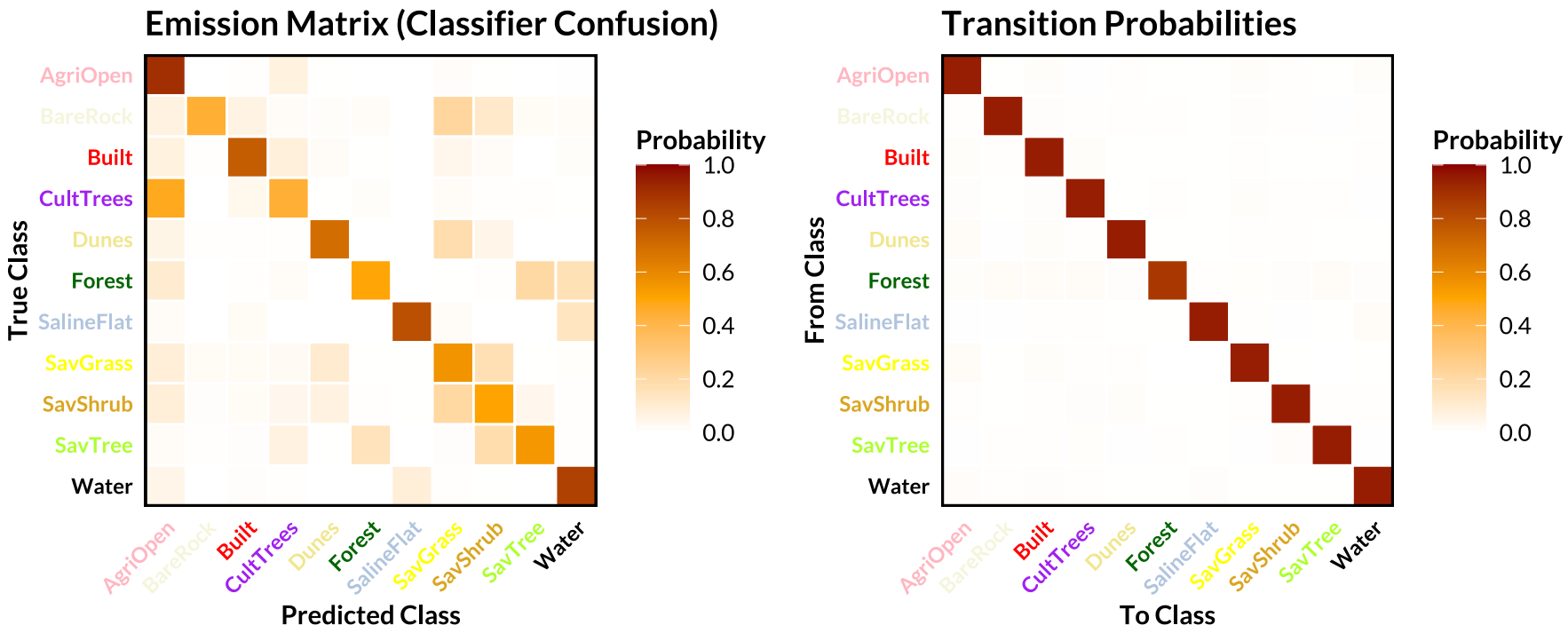
4 Regional Analysis

Each region shows both emission matrix (left, classifier confusion) and transition matrix (right, land cover dynamics). Emission diagonals represent classification accuracy; transition diagonals represent persistence (no change).

4.1 1. AER 2: Rajasthan Bagar (Western Plain)

Sequences analyzed: 15,967
Mean persistence: 0.943
Classification accuracy: 0.678 (127,736 validation samples)

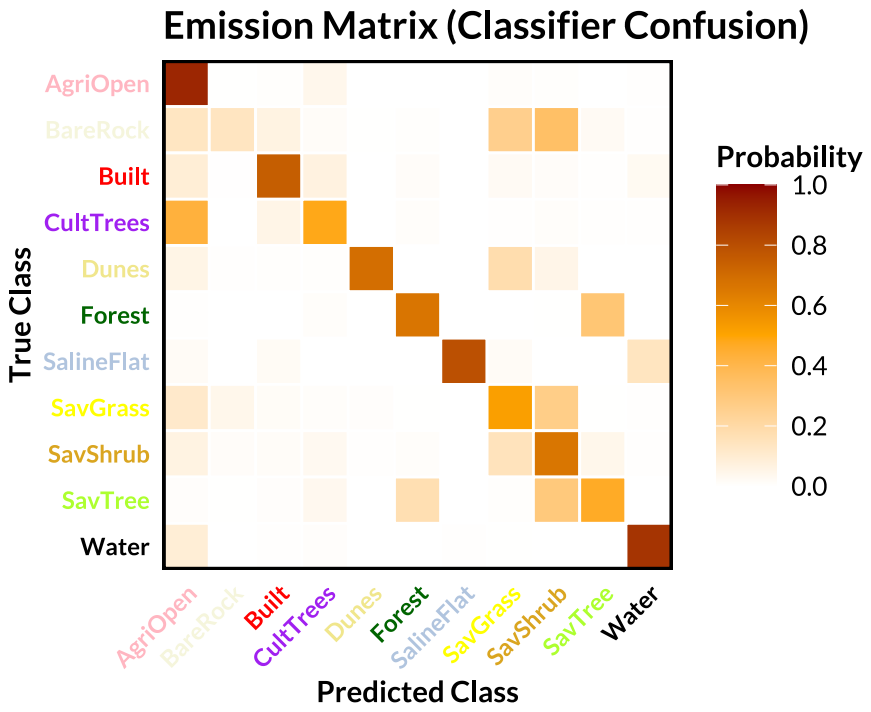
Top 3 transitions:
1. Forest → SavTree (p = 0.020)
2. Forest → CultTrees (p = 0.020)
3. SavGrass → AgriOpen (p = 0.019)

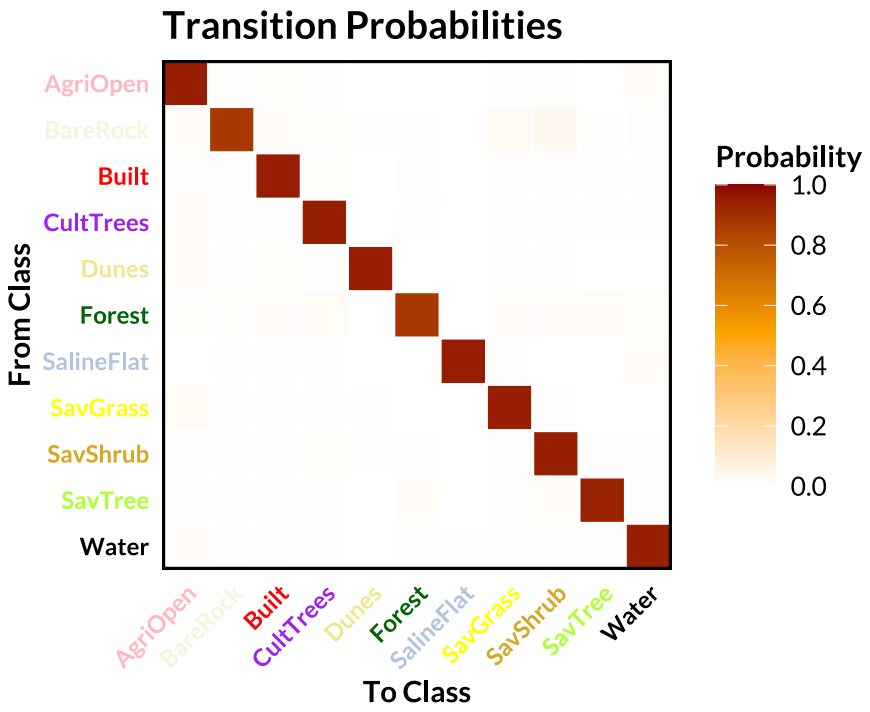


4.2 2. AER 3: Karnataka Plateau

Sequences analyzed: 2,957
Mean persistence: 0.933
Classification accuracy: 0.710 (23,656 validation samples)

Top 3 transitions:
1. BareRock → SavShrub (p = 0.035)
2. BareRock → SavGrass (p = 0.029)
3. SavGrass → AgriOpen (p = 0.023)

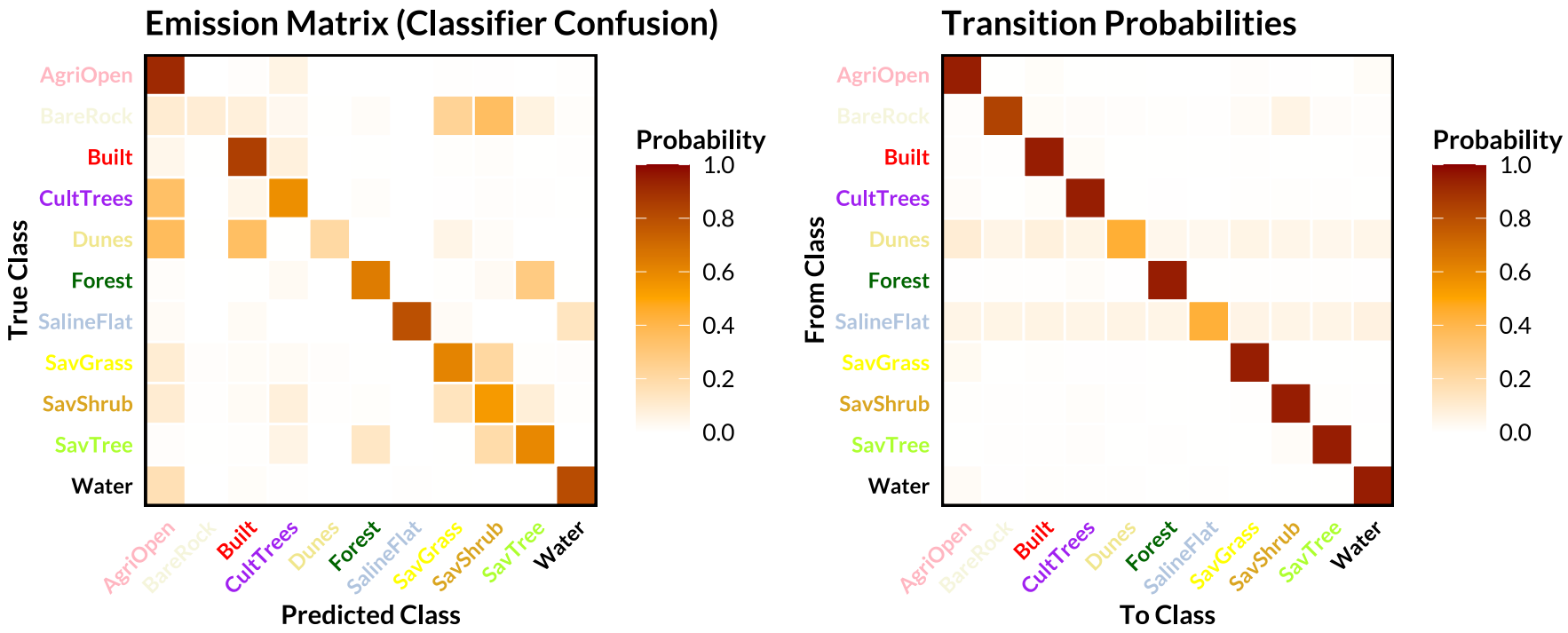




4.3 3. AER 4: North Punjab Plain

Sequences analyzed: 20,373
Mean persistence: 0.847
Classification accuracy: 0.784 (162,984 validation samples)

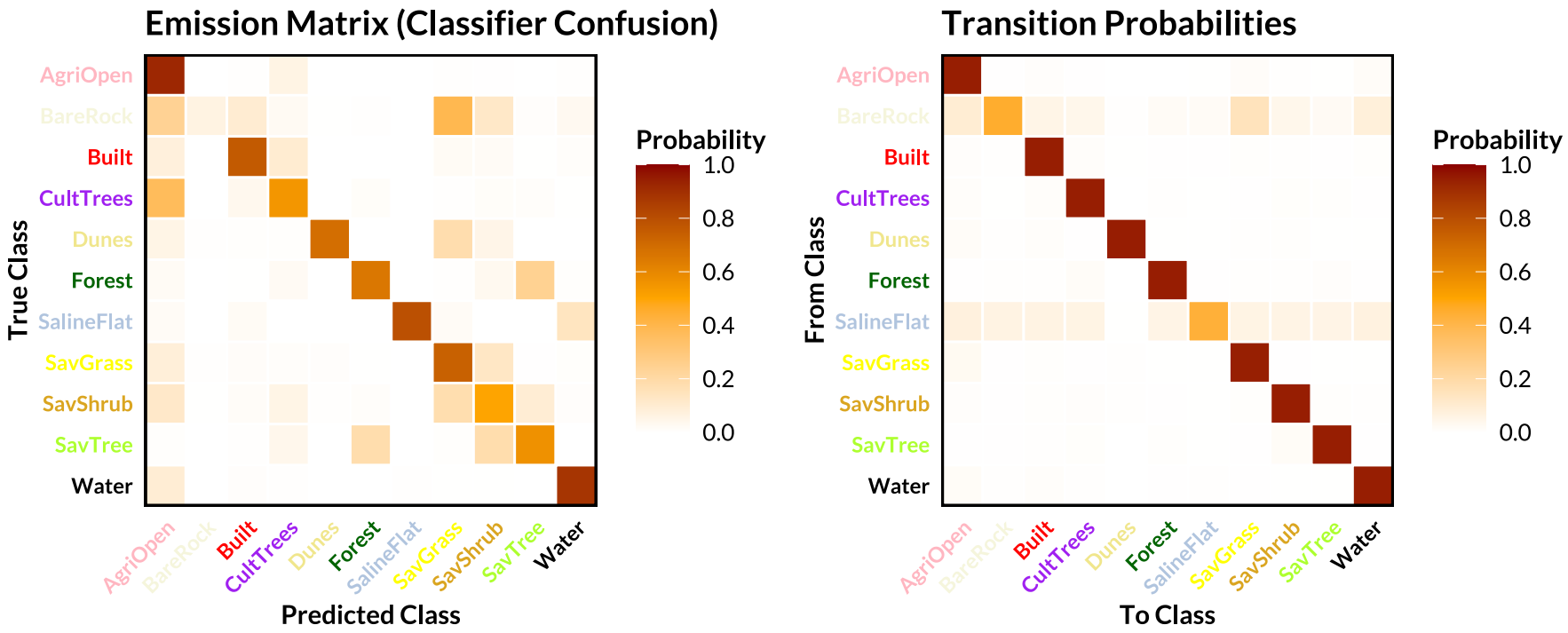
Top 3 transitions:
1. Dunes → AgriOpen (p = 0.095)
2. Dunes → Built (p = 0.077)
3. SalineFlat → Water (p = 0.067)



4.4 4. AER 5: Coastal Kathiawar Peninsula

Sequences analyzed: 11,446
Mean persistence: 0.857
Classification accuracy: 0.759 (91,568 validation samples)

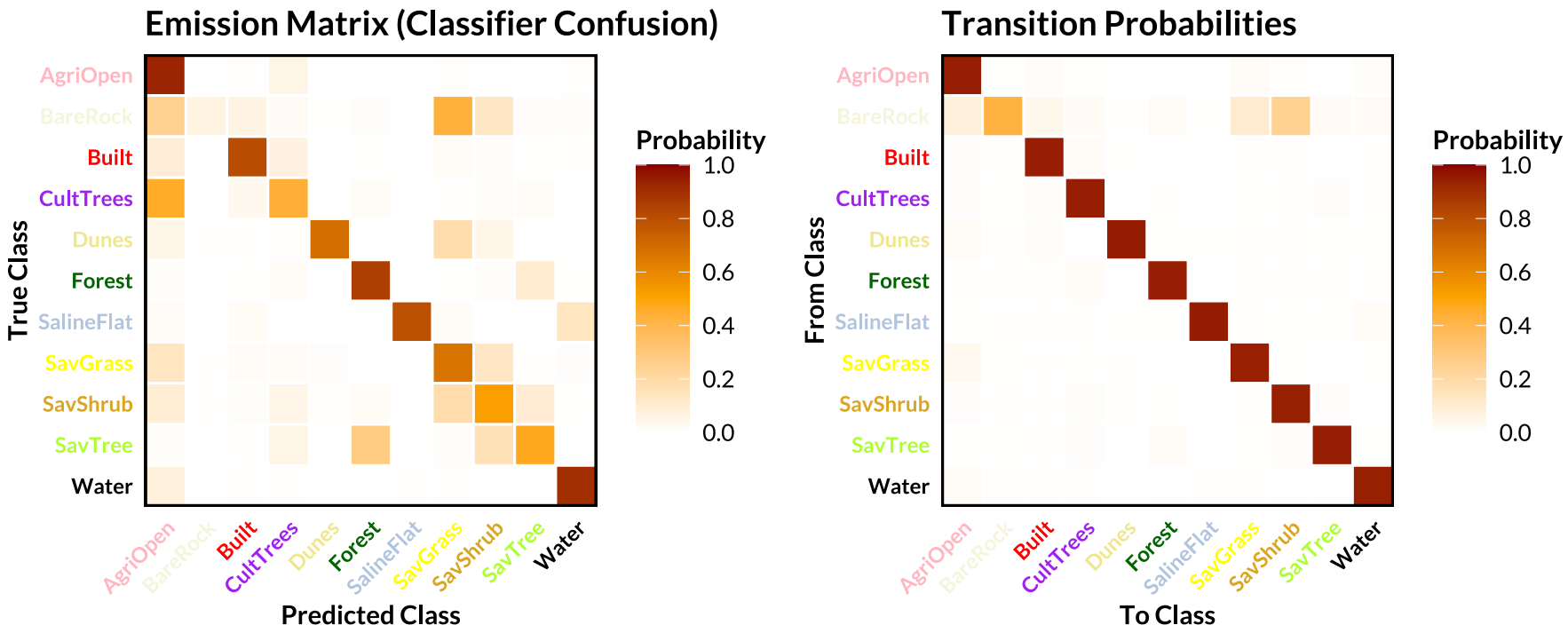
Top 3 transitions:
1. BareRock → SavGrass (p = 0.150)
2. BareRock → AgriOpen (p = 0.099)
3. BareRock → Water (p = 0.083)



4.5 5. AER 6: Eastern Maharashtra Plateau

Sequences analyzed: 20,409
Mean persistence: 0.901
Classification accuracy: 0.770 (163,272 validation samples)

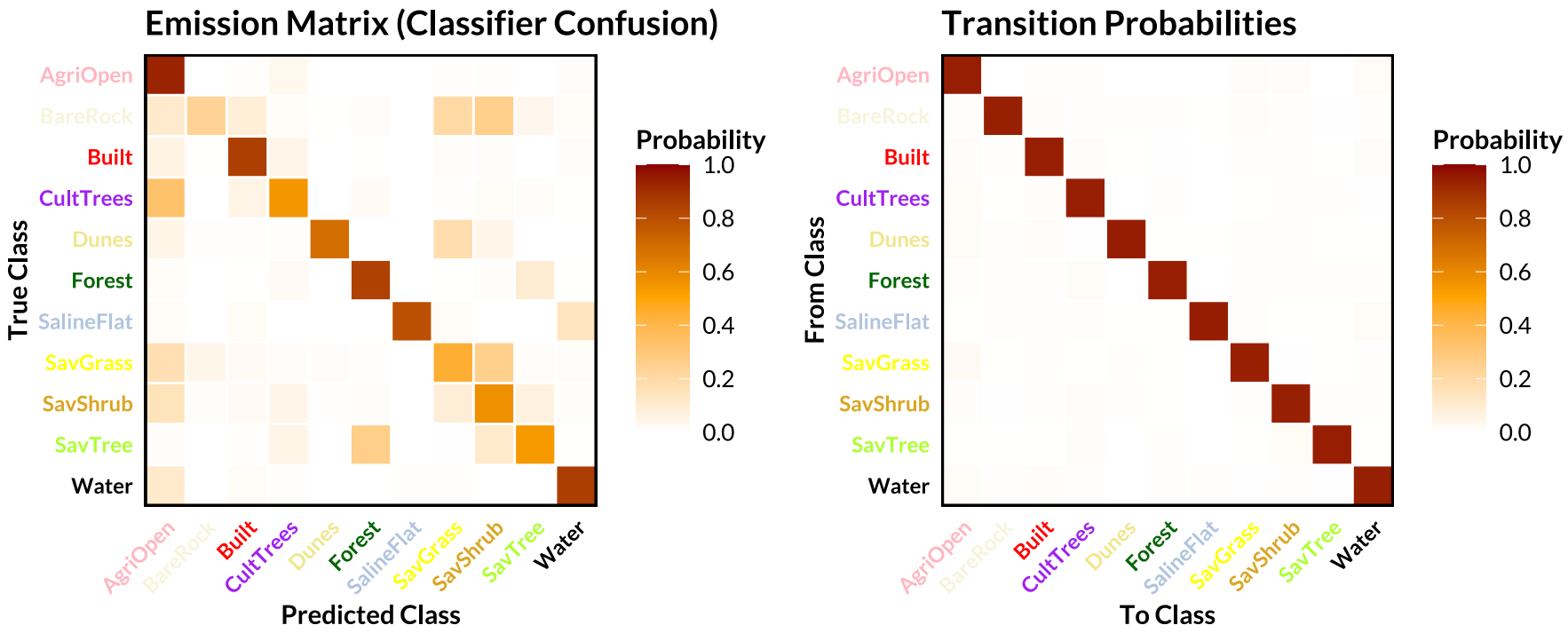
Top 3 transitions:
1. BareRock → SavShrub (p = 0.247)
2. BareRock → SavGrass (p = 0.105)
3. BareRock → AgriOpen (p = 0.081)



4.6 6. AER 7: South Telangana Plateau (Rayalseema) and Eastern Ghats

Sequences analyzed: 10,881
Mean persistence: 0.948
Classification accuracy: 0.743 (87,048 validation samples)

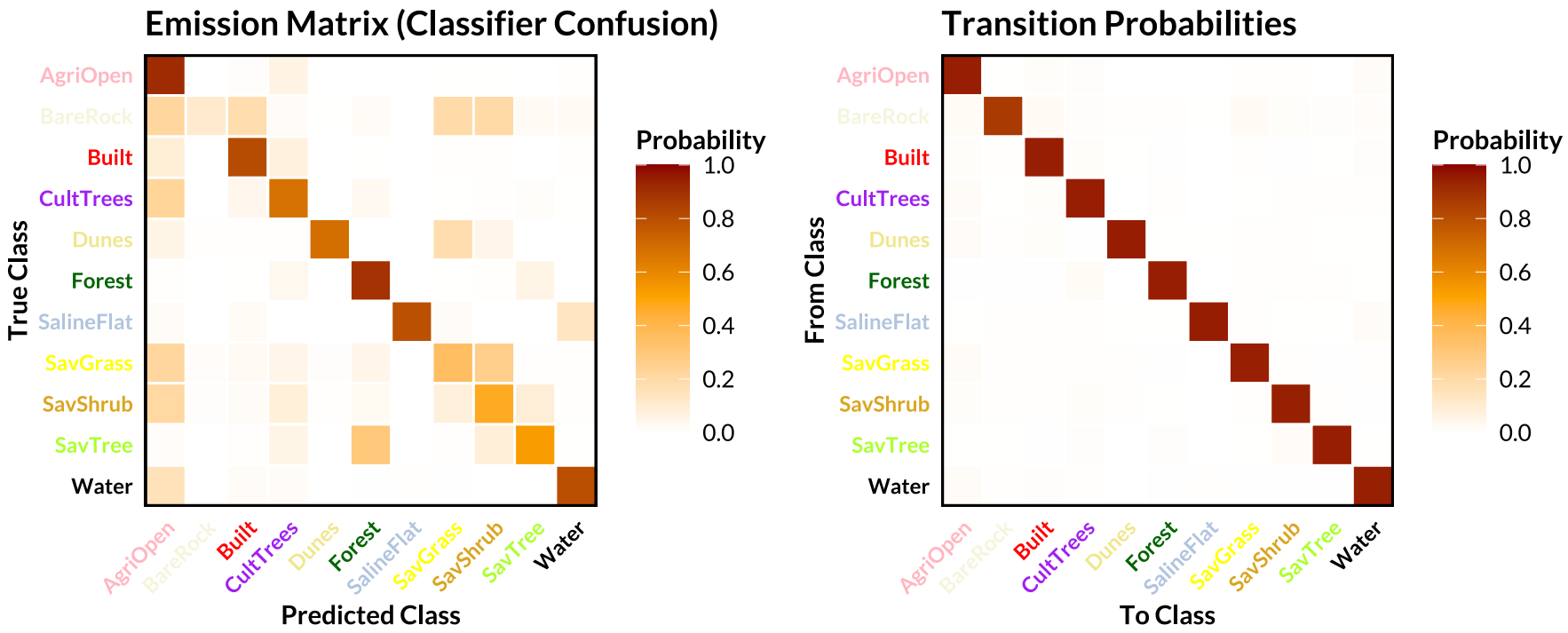
Top 3 transitions:
1. SavGrass → AgriOpen (p = 0.026)
2. SavTree → SavShrub (p = 0.022)
3. SalineFlat → Water (p = 0.018)



4.7 7. AER 8: Tamil Nadu Uplands and Plains

Sequences analyzed: 8,740
Mean persistence: 0.941
Classification accuracy: 0.738 (69,920 validation samples)

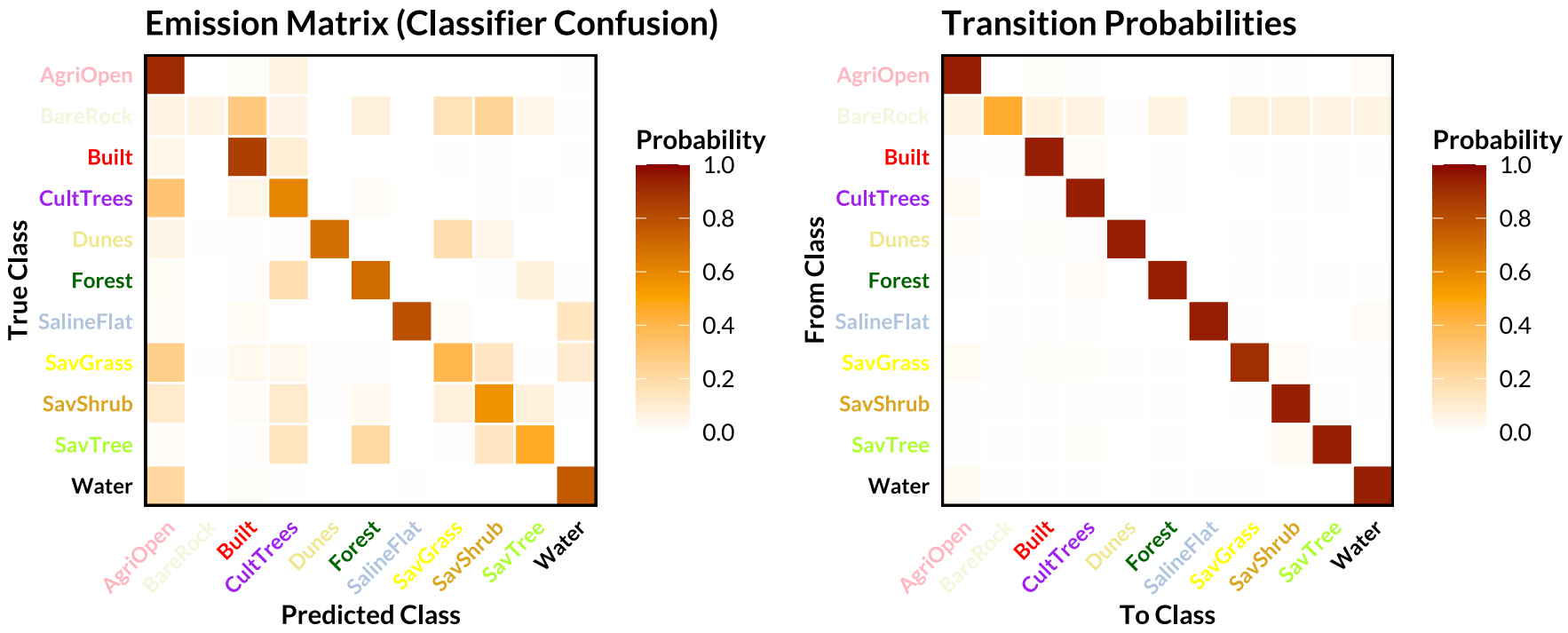
Top 3 transitions:
1. BareRock → Built (p = 0.026)
2. BareRock → SavGrass (p = 0.025)
3. BareRock → AgriOpen (p = 0.022)



4.8 8. AER 9: Punjab and Rohilkhand Plains

Sequences analyzed: 7,529
Mean persistence: 0.899
Classification accuracy: 0.816 (60,232 validation samples)

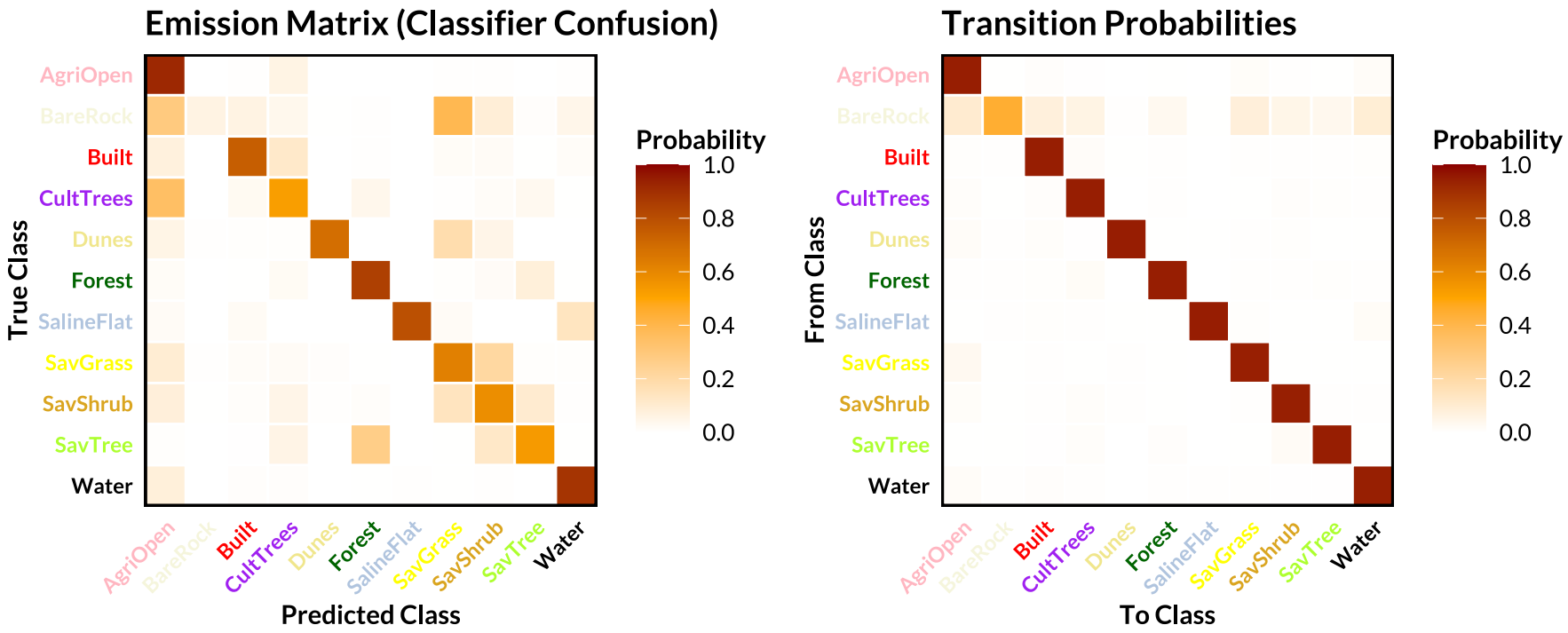
Top 3 transitions:
1. BareRock → SavGrass (p = 0.077)
2. BareRock → SavShrub (p = 0.076)
3. BareRock → Built (p = 0.071)



4.9 9. AER 10: Vindhyan Scarpland and Baghelkhand Plateau

Sequences analyzed: 14,561
Mean persistence: 0.903
Classification accuracy: 0.742 (116,488 validation samples)

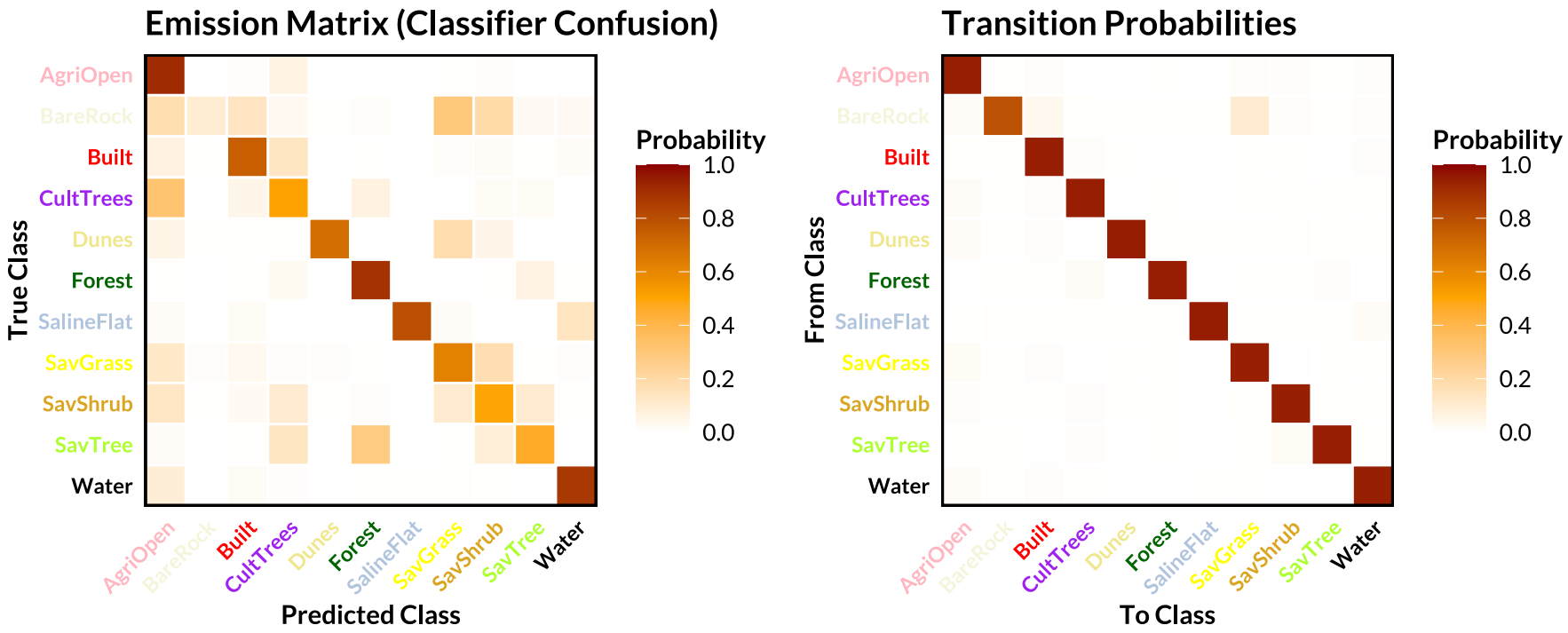
Top 3 transitions:
1. BareRock → AgriOpen (p = 0.106)
2. BareRock → Water (p = 0.093)
3. BareRock → SavGrass (p = 0.084)



4.10 10. AER 11: Eastern Plateau (Chhattisgarh Region)

Sequences analyzed: 6,483
Mean persistence: 0.934
Classification accuracy: 0.679 (51,864 validation samples)

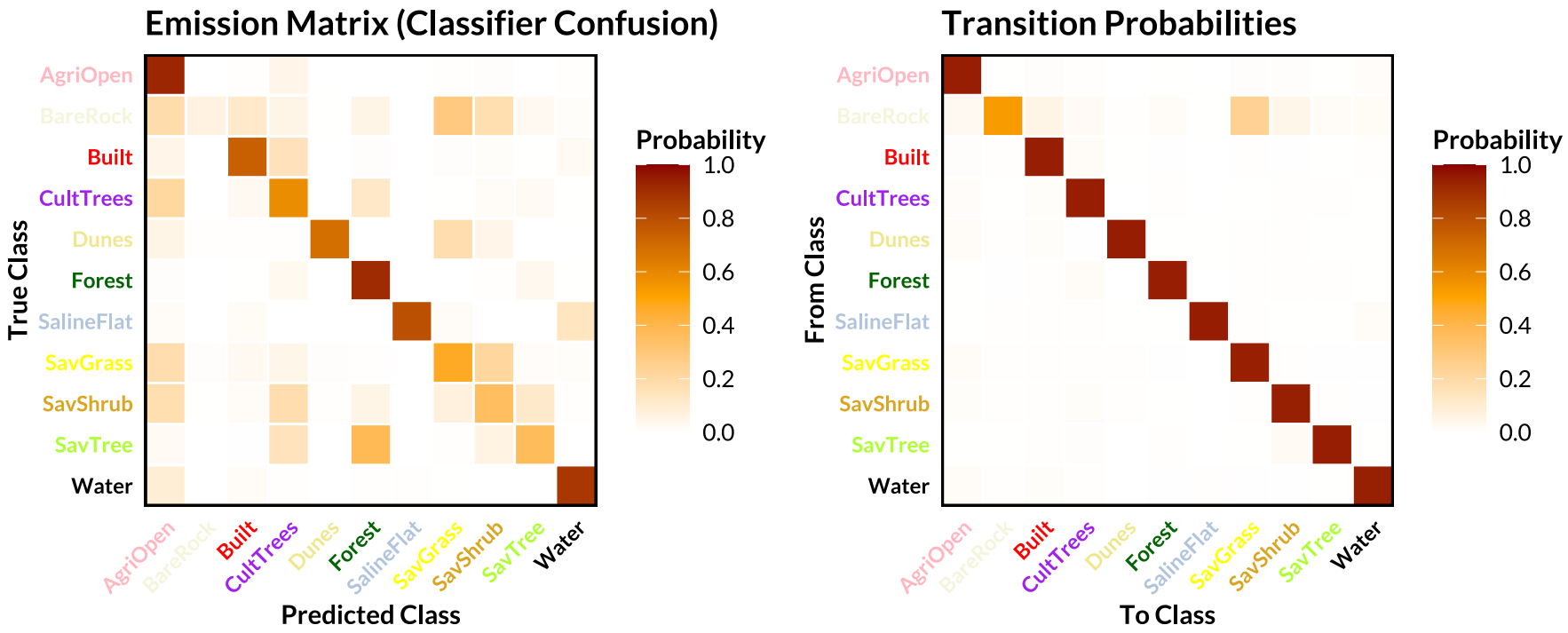
Top 3 transitions:
1. BareRock → SavGrass (p = 0.107)
2. BareRock → Built (p = 0.036)
3. SavTree → SavShrub (p = 0.023)



4.11 11. AER 12: Chhotanagpur Plateau and Garjat Hills

Sequences analyzed: 12,073
Mean persistence: 0.910
Classification accuracy: 0.726 (96,584 validation samples)

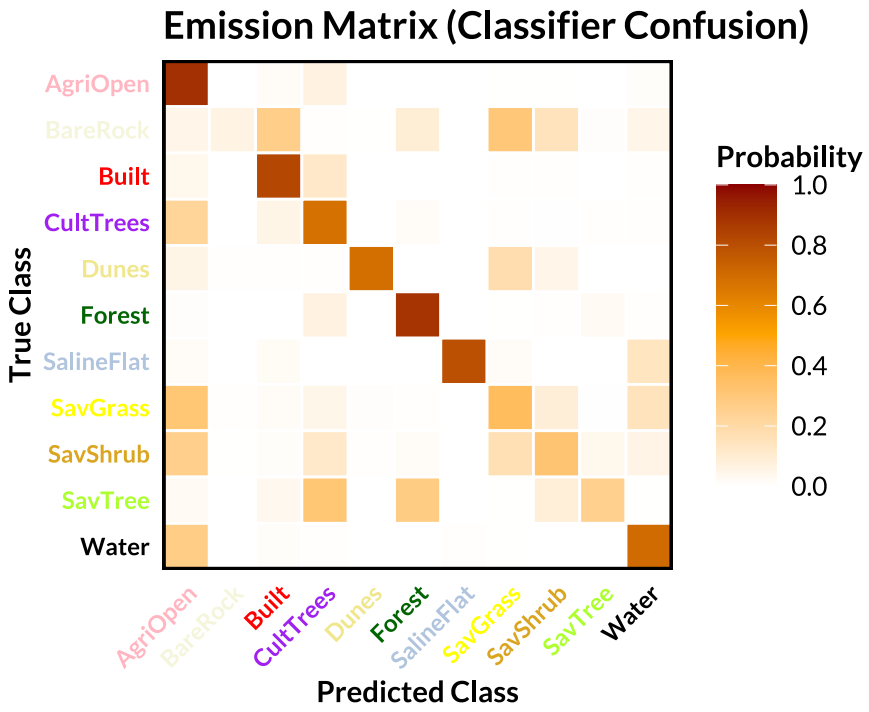
Top 3 transitions:
1. BareRock → SavGrass (p = 0.245)
2. BareRock → Built (p = 0.056)
3. BareRock → SavShrub (p = 0.047)

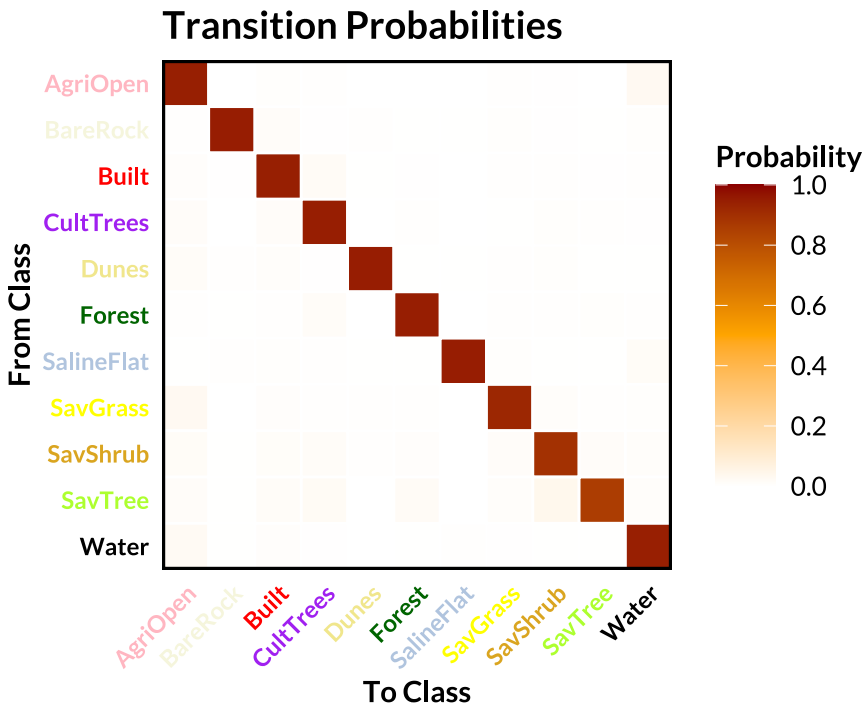


4.12 12. AER 13: Foothills of Central Himalayas

Sequences analyzed: 6,576
Mean persistence: 0.933
Classification accuracy: 0.810 (52,608 validation samples)

Top 3 transitions:
1. SavTree → SavShrub (p = 0.039)
2. SavGrass → AgriOpen (p = 0.030)
3. AgriOpen → Water (p = 0.030)

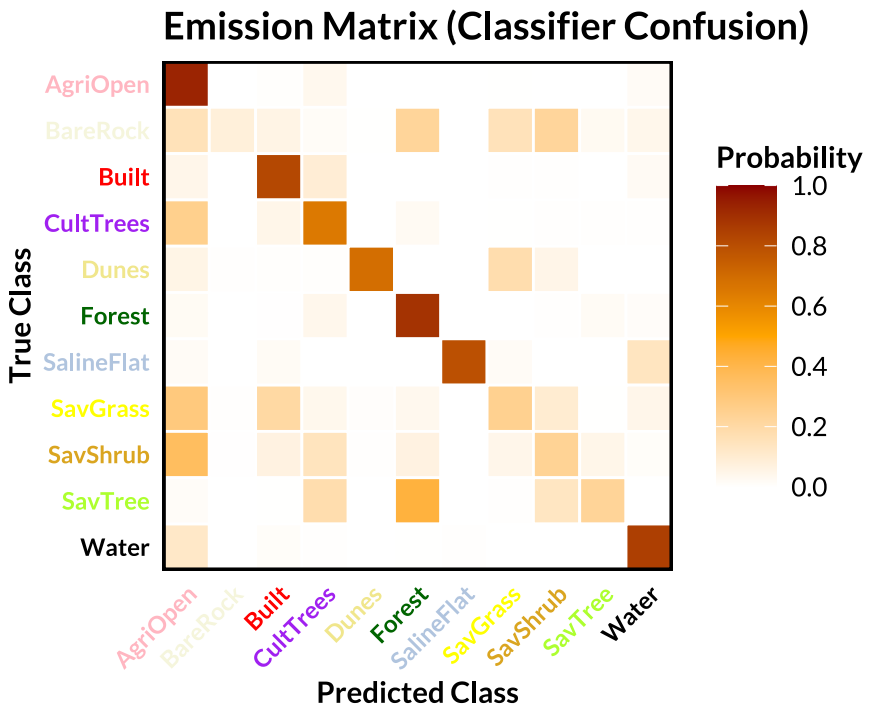


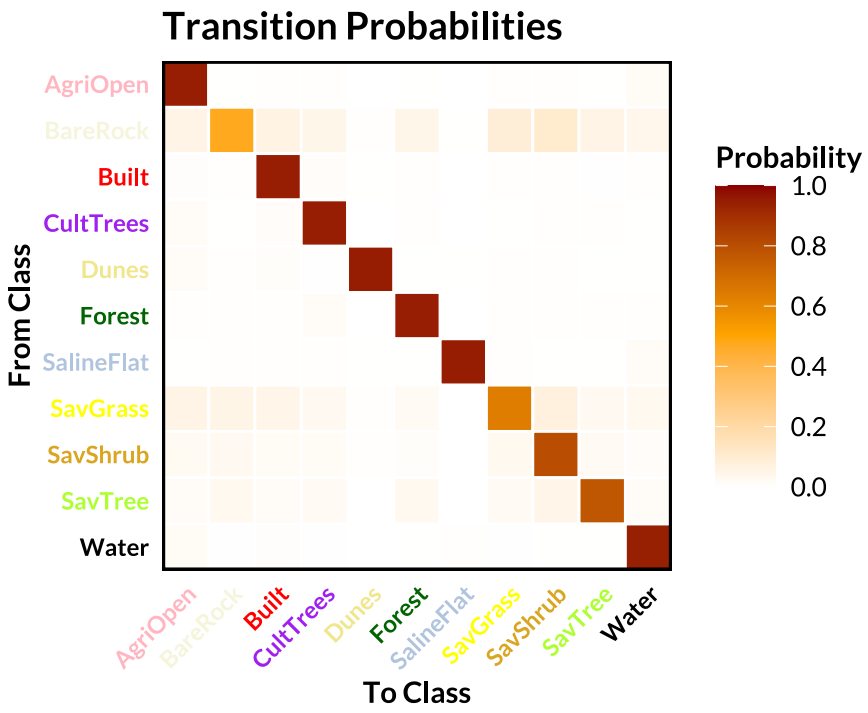


4.13 13. AER 18: North Tamil Nadu Plains (Coastal)

Sequences analyzed: 4,499
Mean persistence: 0.848
Classification accuracy: 0.845 (35,992 validation samples)

Top 3 transitions:
1. BareRock → SavShrub (p = 0.110)
2. BareRock → SavGrass (p = 0.090)
3. SavGrass → SavShrub (p = 0.075)

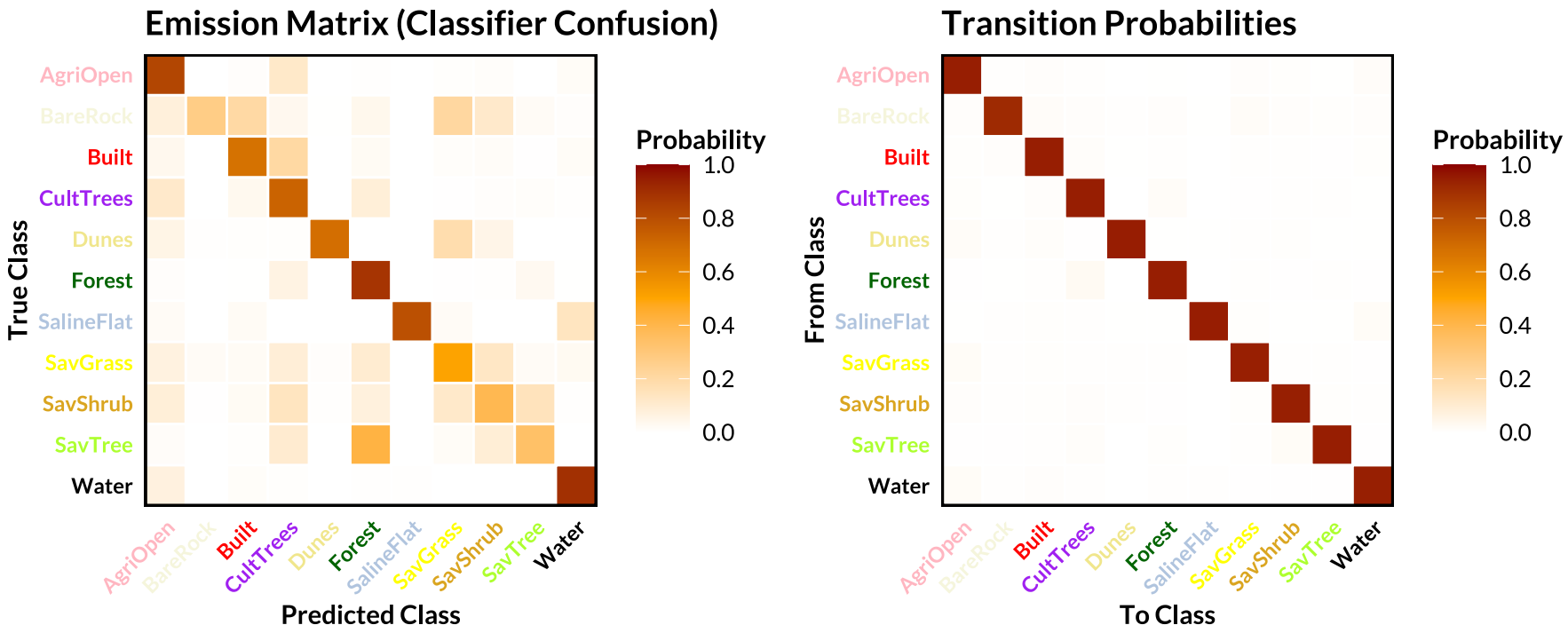




4.14 14. AER 19: Central and South Sahyadris

Sequences analyzed: 5,981
Mean persistence: 0.945
Classification accuracy: 0.717 (47,848 validation samples)

Top 3 transitions:
1. Forest → CultTrees (p = 0.026)
2. SavGrass → AgriOpen (p = 0.019)
3. SalineFlat → Water (p = 0.018)



Report generated: 2025-11-11 19:24:46.718051 **HMM data source:** /Users/mdmadhu/GitHub Repositories/one-2025-hmmLearn/outputs/run_20251111_15