A REPORT ON:

Seasonal Hidden Markov Models for ENSO Phase Detection

BY
Shaheen Ali
ID – 2021B4A33044H

In fulfillment of
MATH F424 – Applied Stochastic Processes
Instructor: Dr. Nirman Ganguly

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI,HYDERABAD CAMPUS

SECOND SEMESTER 2024-2025

ACKNOWLEDGMENTS

I sincerely thank BITS Pilani Hyderabad campus for providing me with this wonderful opportunity to undertake an independent study project. The university's continued commitment to facilitating student experiential learning is deeply appreciated. I am extremely grateful to my project mentor, Dr. Nirman Ganguly, for his invaluable guidance and support throughout this project.

Nirman Sir went above and beyond to ensure our success, holding weekly review meetings to provide practical feedback and help refine our approach. His suggestions were instrumental in shaping our work. Thanks to his motivation, problem-solving assistance, and the university enabling remote project work, I could participate in this cherished learning experience.

This has been an enriching learning journey under Nirman Sir’s mentorship. I feel privileged to have had this opportunity, which was enabled by BITS Pilani Hyderabad.

ABSTRACT

This project explores how a hidden state model can be used to monitor and classify changes in the El Niño–Southern Oscillation (ENSO), one of the most influential climate patterns in the world. By treating the climate phases—El Niño, La Niña, and Neutral—as hidden states that cannot be observed directly, we use sea surface temperature and subsurface warm water volume as inputs to infer these underlying conditions.

The data is organized into overlapping 3-month seasonal windows to capture how the system evolves. A separate model is trained for each season, allowing us to observe how the transitions between climate states vary throughout the year. The model outputs are analysed for accuracy and compared with actual ENSO records. This approach provides a probabilistic and seasonally sensitive method for understanding how major ocean-atmosphere patterns shift and persist over time.

INTRODUCTION

El Niño Southern Oscillation (ENSO)

The El Niño–Southern Oscillation (ENSO) is a major factor in year-to-year climate variability, having significant effects worldwide. This phenomenon, which involves the interaction between the ocean and the atmosphere, begins in the tropical Pacific Ocean and impacts global temperature, rainfall, and air circulation patterns. ENSO is composed of three distinct phases—El Niño, La Niña, and a Neutral phase—each linked to particular oceanic and atmospheric characteristics.

- **El Niño:** This phenomenon is marked by a weakening or reversal of the trade winds in the Pacific, which usually blow from east to west along the equator. Consequently, the warm surface waters that are customarily concentrated in the western Pacific move eastward, resulting in elevated sea surface temperatures in the central and eastern Pacific. This change disrupts global weather patterns, frequently leading to droughts in Australia and India, excessive rainfall in South America, and milder winters across North America.

- **La Niña:** is the opposite of El Niño and happens when trade winds intensify, moving warm surface water further to the west. This process increases the upwelling of cold, nutrient-rich water in the eastern Pacific, leading to cooler-than-normal sea surface temperatures in that area. Typically, La Niña brings higher-than-average rainfall to Southeast Asia, stronger monsoons in India, and colder winters in certain regions of North America.

- **Neutral:** In Neutral conditions, sea surface temperatures and wind patterns tend to align more closely with long-term averages, and no major anomalies related to ENSO are detected. Nonetheless, variability can still occur during these times due to the impact of other climate systems, like the Indian Ocean Dipole (IOD) or Arctic Oscillation (AO).

In India, the ENSO cycle is vital for influencing the characteristics of the Southwest Monsoon, which is responsible for about 75% of the country's yearly rainfall. El Niño years tend to be linked with weaker monsoons and drought-like conditions, which can have a serious impact on agriculture, water supplies, and economic productivity. On the other hand, La Niña phases are usually associated with more intense monsoon activity and improved crop yields. Given that the majority of India's population depends on agriculture driven by the monsoon, comprehending the relationship between ENSO and regional climate fluctuations has important socioeconomic consequences.

Even though ENSO is crucial on a global scale, effectively identifying and forecasting its phases continues to be a significant challenge. ENSO events typically evolve slowly over several months and can differ in both intensity and length. Consequently, scientists depend on oceanographic metrics, including Sea Surface Temperature (SST) anomalies and Warm Water Volume (WWV), to recognize the initiation and progression of ENSO phases. These metrics are fundamental to multiple climate forecasting models.

In this study, we aim to analyze the ENSO phenomenon using a Hidden Markov Model (HMM), a statistical approach capable of modeling time series data where the system's internal state is not directly observable. By treating ENSO phases as hidden states and SST-WWV measurements as observable variables, the HMM framework can classify, monitor, and potentially predict transitions between ENSO phases over time. This approach is particularly valuable when working with monthly data over several years, as it can uncover probabilistic patterns that might not be evident using traditional threshold-based classification.

Dataset Introduction

In order to accurately model the phases of the El Niño–Southern Oscillation (ENSO), this research employs two primary oceanic indicators: anomalies in Sea Surface Temperature (SST) and Warm Water Volume (WWV). These factors are well-established for their significance in monitoring the progression and strength of ENSO events, and they act as the measurable inputs for our Hidden Markov Model (HMM).

Sea Surface Temperature (SST) Anomalies:
SST data indicates variations from average ocean temperatures and serves as a key indicator in recognizing ENSO phases. The dataset utilized in this research specifically concentrates on the Niño 3.4 region (5°N–5°S, 170°W–120°W), which is a commonly accepted reference area for tracking ENSO. Monthly SST anomaly figures are employed to identify trends of warming or cooling in the equatorial Pacific.

- Positive anomalies (typically above +0.5°C) in this region signal the development of El Niño.

- Negative anomalies (below −0.5°C) often indicate the onset of La Niña.

- Values within this range (−0.5°C to +0.5°C) generally correspond to Neutral conditions.

Warm Water Volume (WWV): provides insights into subsurface ocean dynamics that occur before visible changes in the sea surface temperature (SST) associated with ENSO. It is characterized as the amount of water exceeding 20°C between 5°N and 5°S in the equatorial Pacific, serving as an indicator of the thermal energy stored beneath the surface.

Variations in WWV often signal upcoming changes in SST and transitions in ENSO phases. For instance, an increase in WWV generally signals the beginning of El Niño, whereas a decrease is linked to the development of La Niña. Therefore, including WWV alongside SST offers a more comprehensive understanding of the ocean-atmosphere system.

The dataset spans multiple decades, allowing for the analysis of long-term ENSO cycles and trends. It is organized into overlapping 3-month periods (e.g., DJF, JFM, FMA, etc.) to capture seasonal dynamics and smooth out short-term fluctuations. Each 3-month period is treated as a data point for HMM training and classification.

Hidden Markov Models

Hidden Markov Models (HMMs) are statistical frameworks employed to characterize systems that change over time in a probabilistic way. In an HMM, it is assumed that the system follows a Markov process with hidden (unobservable) states, where each concealed state produces observable data based on a specific probability distribution.

When predicting ENSO (El Niño–Southern Oscillation), we utilize an HMM to represent the three underlying climate states:

- El Niño,
- La Niña, and
- Neutral.

These states cannot be directly observed in the raw data, but they affect measurable oceanographic indicators such as Sea Surface Temperature (SST) anomalies and Warm Water Volume (WWV) anomalies, which we use as the observable outputs in our model. By modeling the connection between these concealed ENSO phases and the observed variables, the HMM allows us to deduce the most probable state of the climate system for each season.

Formal Definition:

A Hidden Markov Model is defined by a quintuple:

$$\lambda = (S, \mathrm{O}, \mathrm{A}, \mathrm{B}, \pi)$$

Where:

- $S = \{s_1, s_2 \dots\dots\dots\dots s_N\}$ is the set of hidden states. In our model:

$$S = \{El\ Nino, La\ Nina, Neutral\}, \qquad N = 3$$

- $\mathrm{O} = \{o_1, o_2 \dots\dots\dots\dots o_T\}$ is the sequence of observed vectors over time T. For our climate vectors, each observation vector is representated as follows:

$$o_t = \begin{bmatrix} SST_t \\ WWV_t \end{bmatrix}$$

Transition Probability Matrix (*A*)

The state transition matrix A defines the probability of transitioning from one hidden state to another:

$$A = \{a_{ij}\}, \qquad a_{ij} = P(q_{t+1} = s_j \,|q_t = s_i), \quad \sum_{j=1}^{N} a_{ij} = 1$$

Where:

- q_t represents the hidden state at time t,
- a_{ij} is the probability of transitioning from state s_i to state s_j

This reflects the probability of the ENSO phases either continuing or changing from one 3-month season to the subsequent one.

Emission Probability Distribution (B)

Since our observations are continuous and multivariate, we model emissions using a Multivariate Gaussian Distribution for each state s_i:

$$b_i(o_t) = P(o_t \mid q_t = s_i) = N(o_t; \mu_i; \Sigma_i)$$

Where:

- μ_i is the Mean vector for observations in state s_i
- Σ_i is the Covariance matrix representing variability and correlation of SST and WWV in state s_i
- o_t is the observed vector at time t

This approach enables the HMM to identify unique statistical characteristics of El Niño, La Niña, and Neutral phases derived from SST-WWV combinations.

Initial State Distribution (π)

$$\pi = \{\pi_i\}, \quad \pi_i = P(q_1 = s_i), \quad \sum_{i=1}^{N} \pi_i = 1$$

This defines the probability that the system starts in state s_i (i.e., what ENSO phase the model expects at the beginning of the time series).

Thus, the complete HMM is parametrized by:

- ***A***: *NxN* transition matrix
- ***Π***: Initial state probabilities vector of length N
- ***B***: A set of N multivariate Gaussian distributions $N(o_t; \mu_i; \Sigma_i)$

Training the HMM: Baum-Welch Algorithm

To learn the parameters (A, B, π) from the dataset, we use the Baum-Welch algorithm, a special case of the Expectation-Maximization (EM) algorithm for Hidden Markov Mocdels. The algorithm iteratively maximizes the likelihood of the observed data.

Given observations $O = \{o_1, o_2 \dots \dots \dots \dots . o_T\}$, we define:

- Forward probability:
$$\alpha_t(i) = P(o_1, o_2 \dots \dots \dots \dots . o_t, q_t = s_i | \lambda)$$

- Backward probability:
$$\beta_t(i) = P(o_{t+1}, o_{t+2} \dots \dots \dots \dots . o_T \, | q_t = s_i, \lambda)$$

Then we estimate:

- $\gamma_t(i)$: The probability of being in state s_i at time *t* given the observations
- $\xi_t(i,j)$: Probability of transitioning from s_i to s_j at time t

Using these, we re-estimate parameters:

- Updated transition probabilities:
$$a_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)}$$

- Updated means and covariances for emissions:
$$\mu_i = \frac{\sum_{t=1}^{T} \gamma_t(i) o_t}{\sum_{t=1}^{T} \gamma_t(i)}, \quad \Sigma_i = \frac{\sum_{t=1}^{T} \gamma_t(i)(o_t - \mu_i)(o_t - \mu_i)^T}{\sum_{t=1}^{T} \gamma_t(i)}$$

This iterative process continues till convergence.

Decoding the ENSO phases: The Viterbi Algorithm

After training, we use the Viterbi algorithm to determine the most probable sequence of hidden states (ENSO phases) given the observations.

The goal is to compute:
$$Q^* = \arg\max_{Q} P(Q | O, \lambda)$$

Where:

- $Q = \{q_1, q_2, \dots \dots \dots \dots q_T\}$ is a sequence of hidden states.
- $O = \{o_1, o_2 \dots \dots \dots \dots . o_T\}$ is the sequence of observations

Recurrence relations:

$$\delta_t(j) = \max_i[\delta_{t-1}(i).a_{ij}].b_j(o_t)$$

$$\psi_t(j) = \arg\max_i[\delta_{t-1}(i).a_{ij}].$$

Where $\delta_t(j)$ stores the highest probability of any path that ends in state s_j at time t, and $\psi_t(j)$ stores the corresponding backpointer.

This dynamic programming approach yields an optimal sequence $(q_1^*, q_2^*, \dots \dots q_T^*)$ which labels each time period with the most likely ENSO phase.

Note:

Although Hidden Markov Models (HMMs) aren't intended for accurate deterministic weather predictions—due to the multitude of interrelated factors and long-term effects impacting atmospheric systems—they still serve as a useful method for revealing hidden structures and probabilistic relationships in climate time series data. The core premise of HMMs, which is that the current state relies solely on the prior state (first-order Markov property), fails to fully account for the chaotic characteristics of weather systems. Nevertheless, when utilized for broader climate phenomena such as ENSO, HMMs enable us to deduce hidden regimes (for instance, El Niño, La Niña, Neutral) and examine the probabilistic changes between these states.

The main aim is not to achieve highly accurate forecasts of upcoming ENSO states, but rather to enhance our understanding of:

• The dynamics of transitions between concealed climate regimes,

• The statistical patterns identified in the observed SST and WWV,

• And the fundamental temporal coherence and persistence of ENSO phases.

This methodology provides a streamlined, interpretable model that aids in quantifying the likelihood of the climate system maintaining or shifting away from specific ENSO conditions over time.ver time.

Objectives

- Model the evolution of ENSO-related phenomena (El Niño, La Niña, Neutral) as a sequence of hidden states.
- Study the transition probability matrix to understand how likely ENSO states are to shift between phases.
- Use the Viterbi algorithm to infer the most likely sequence of ENSO phases over time and compare with known ENSO years to validate consistency.
- Apply the Baum-Welch algorithm to estimate the optimal HMM parameters from the data.
- Analyze the statistical properties (mean and variance) of the observed variables associated with each latent state, helping to characterize the influence of hidden ENSO phases.
- Train separate HMMs on overlapping 3-month windows (e.g., DJF, JFM, etc.) to analyze seasonal shifts in hidden state dynamics and transition behavior.
- Highlight the insights gained about ENSO regime shifts and climate phase behavior through state transition patterns rather than direct prediction accuracy.

Methodology

Dataset Collection and Preprocessing

- Use monthly time series data of ENSO-related indices: Niño 3.4 SST anomalies, Warm Water Volume (WWV) anomalies
- Normalize and align datasets to a common timeline.
- Segment the time series into overlapping 3-month seasonal windows (DJF, JFM, ..., NDJ).

Model Setup (HMM Initialization)

- Assume a 3-state Hidden Markov Model representing {El Niño, La Niña, Neutral}.
- Use Gaussian emission probabilities for continuous observations (SST, WWV).
- Initialize HMM parameters (transition matrix A, emission parameters $\{\mu, \Sigma\}$, and initial state distribution π) randomly or using k-means clustering.

Parameter Estimation and **State Decoding using Python Libraries**.

Analysis and Interpretation

- Analyse the learned transition matrix to uncover the stability and volatility of each ENSO phase.
- Compare emission means and covariances to characterize each hidden state's influence on SST and WWV.
- Track how transition behaviour varies across seasonal windows.

Visualization and Reporting

- Plot the observed time series overlaid with inferred states.
- Visualize transition matrices and emission distributions.
- Summarize regime dynamics across seasonal models.

Technical Implementation:

This project's analysis was carried out in Python, using a combination of data science and machine learning (TensorFlow) libraries to model and interpret ENSO (El Niño–Southern Oscillation) dynamics. The computations and visualizations were executed in an interactive notebook environment using Google Colab.

Key Libraries and their Roles:

Library	Purpose
Pandas	Data loading and processing
NumPy	Array handling, matrix algebra
Matlplotlib and Seaborn	Visualizations
Hmmlearn	Core implementation of Gaussian Hidden Markov Models used for model fitting, state prediction and probability decoding
Sklearn.metrics	Used for calculating evaluation metrics like accuracy and confusion matrices
Sklearn.cluster (KMeans)	Clustering and cluster analysis
Scipy.stats	Emission modeling

Workflow:

Data Ingestion: For every overlapping 3-month period, preprocessed seasonal ONI and WWV anomaly data (e.g., DJF, JFM, FMA) were loaded.

HMM Training: Both ONI and WWV values were used as 2D emission features to train a 3-state Gaussian HMM for each period.

State Alignment: Using a majority vote method, predicted hidden states were matched to known labels by aligning them with ENSO classes (El Niño, La Niña, and Neutral).

Model Results:
1) Initial probabilities and the transition matrix
2) Covariances and emission means for every hidden state
3) Sequences of predicted states

4) Probabilities of posterior states

Post-Model Evaluation:
1) Calculating the accuracy and confusion matrix
2) Identification of change points: Years when ENSO regimes changed
3) Year cluster analysis using state-by-state emission data

Illustrations:
1) Time series including anticipated states
2) Trends in posterior probability
3) Distributions of emissions
4) Matrix confusion

As the initial steps have already been explained in the introduction itself, I will directly proceed to the results of the time series modelling procedure, however we need a note on Classification of States and State Alignment:

Classification of States:

- The true ENSO state for each year and season was determined using the Oceanic Niño Index (ONI) value:
  - El Niño: ONI ≥ +0.5°C
  - La Niña: ONI ≤ –0.5°C
  - Neutral: –0.5°C < ONI < +0.5°C
- These thresholds align with standard climatological definitions provided by organizations such as NOAA.
- This classification created a ground truth label for each data point, enabling comparison with the HMM-inferred states.
- The labels were encoded numerically:
  - 0 → El Niño
  - 1 → La Niña
  - 2 → Neutral
- These numerical values were used to calculate:
  - Prediction accuracy of the HMM.
  - Confusion matrices to understand misclassifications
  - Aligned state labels

State Alignment:

- Since HMMs are unsupervised, the hidden states needed to be mapped to meaningful ENSO labels.
- A majority voting mechanism was used:
  - For each HMM-inferred state, the most frequent actual ENSO class (based on ONI thresholds) was identified.
  - That label (El Nino, La Nina, or Neutral) was assigned to the HMM state.
- This mapping ensured the interpretability of results and enabled consistent evaluation of model performance.
- It also made possible the computation of performance metrics.

Note on terminology:

As the data is presented in 3 month rolling means, a guide to the terminology:

DJF = December-January-February, JFM = January-February-March, FMA = February-March-April and so on.

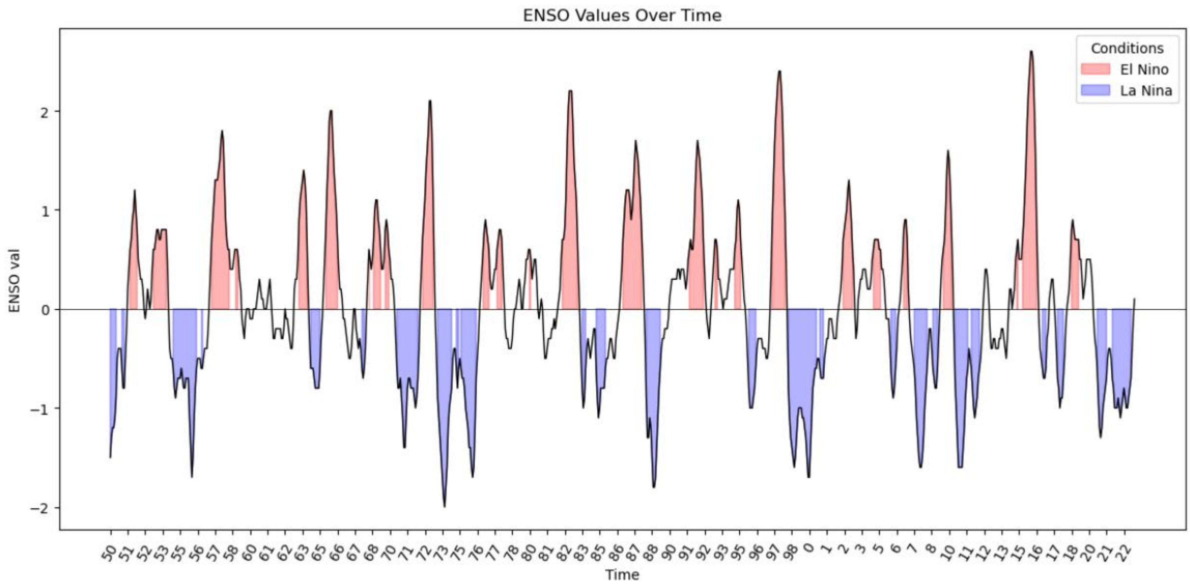


Figure: Depicting El-Nino and La-Nina events through red and blue colours respectively, y axis represents ONI values, x axis represents years from 1950 to 2023

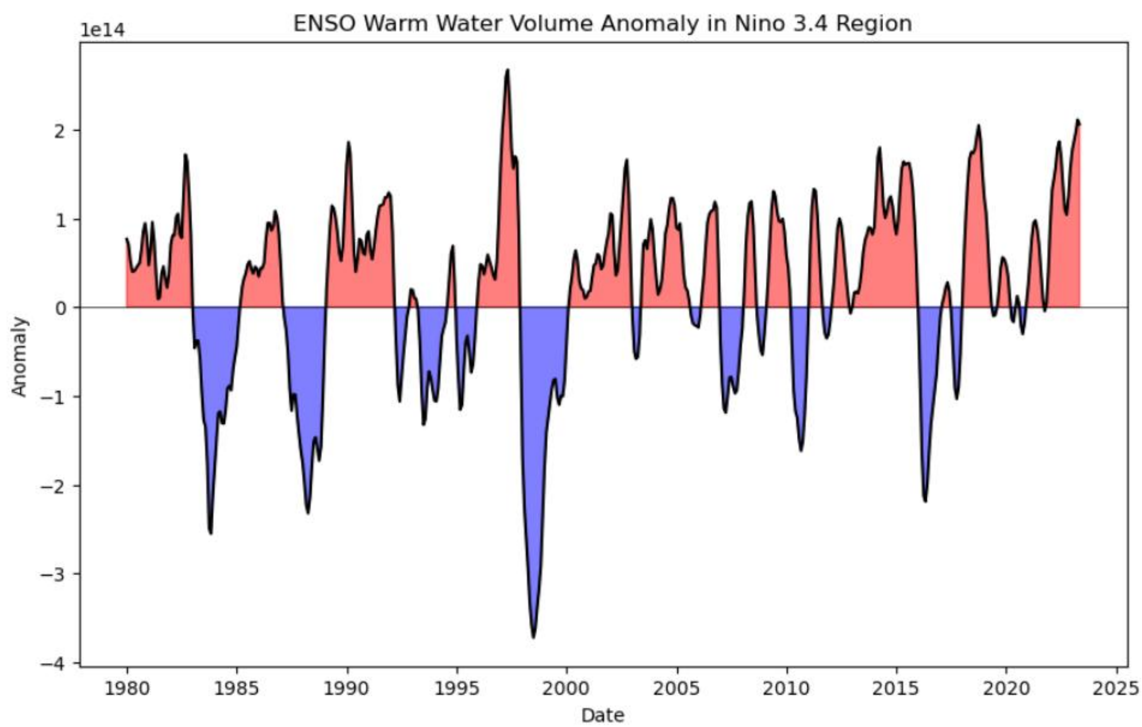


Figure: Depiction of positive and negative WWV Anomaly events using appropriate colours

Results:

Transition Matrices:

- Each 3x3 transition matrix lists probabilities of staying in or switching between El Niño (0), La Niña (1), or Neutral (2) states, with rows summing to 1.
- Diagonal values show *persistence*—the chance a state remains unchanged—while off-diagonal values show transitions to other states.

Values obtained from the model are as follows:

Month	**Transition Matrix**
DJF	[[2.26744154e-09 9.99999973e-01 2.44387261e-08] [6.82676952e-01 1.24073756e-05 3.17310641e-01] [1.58741991e-01 8.41258009e-01 1.30202873e-13]]
JFM	[[3.28218730e-01 6.71781175e-01 9.45387286e-08] [1.04512581e-01 3.64167345e-01 5.31320074e-01] [6.15147365e-01 1.10948181e-09 3.84852633e-01]]
FMA	[[1.28801913e-06 9.99998711e-01 7.48898719e-10] [8.53805698e-01 3.38763875e-05 1.46160426e-01] [3.26182984e-01 6.73816913e-01 1.02952117e-07]]
MAM	[[1.69813387e-04 9.99654945e-01 1.75241949e-04] [1.35339455e-01 7.75419252e-07 8.64659769e-01] [3.05836633e-01 2.22005828e-04 6.93941362e-01]]
AMJ	[[3.19545510e-006 8.61951009e-001 1.38045796e-001] [4.50289624e-001 5.49710376e-001 2.23130823e-032] [1.00000000e+000 1.23154782e-016 1.40508396e-122]]
MJJ	[[3.37280285e-09 9.99596915e-01 4.03081233e-04] [1.93203822e-01 1.55670104e-07 8.06796023e-01] [3.90009637e-01 5.82617577e-03 6.04164187e-01]]
JJA	[[2.77816878e-01 7.22183122e-01 9.89322968e-14] [2.78207548e-01 1.62187682e-07 7.21792289e-01] [4.41637022e-02 4.39184327e-01 5.16651971e-01]]
JAS	[[5.76093152e-04 9.99412101e-01 1.18061777e-05]

	[2.31287155e-01 2.08529458e-02 7.47859900e-01] [2.68880191e-01 7.83116873e-02 6.52808121e-01]]
ASO	[[3.15544142e-06 9.99996721e-01 1.23909016e-07] [3.55541309e-05 3.48499405e-01 6.51465041e-01] [2.20120369e-01 4.48081033e-06 7.79875150e-01]]
SON	[[2.61252723e-05 9.99973331e-01 5.43916782e-07] [1.29425133e-07 3.17556200e-01 6.82443671e-01] [2.72636365e-01 7.50415818e-05 7.27288593e-01]]
OND	[[5.52071829e-001 4.47928171e-001 2.49819755e-032] [7.51477165e-001 7.08409675e-012 2.48522835e-001] [1.00000000e+000 1.72318586e-063 9.01942876e-135]]
NDJ	[[1.50468403e-005 9.99974825e-001 1.01282959e-005] [7.79701007e-001 7.21580085e-002 1.48140985e-001] [9.99999999e-001 1.31297898e-009 3.24171610e-110]]

Key Characteristics:

- Persistence varies by season: high in winter (e.g., La Niña in DJF: 9.99999973e-01) means stable states; moderate in spring (e.g., El Niño in JFM: 3.64167345e-01) means frequent changes.
- Stability is strongest for La Niña in winter (NDJ: 9.99974825e-01), Neutral in summer (JAS: 9.99412101e-01), and El Niño in fall (SON: 7.27288593e-01), matching ENSO's seasonal cycle.

Seasonal Patterns:

- Winter (DJF, NDJ) favors La Niña staying La Niña, seen in high persistence and 33/43 La Niña clusters in DJF.
- Spring (JFM, OND) often sees La Niña shift to El Niño (e.g., 6.71781175e-01 in JFM, 7.51477165e-001 in OND), signaling El Niño onset, with 20–26 state changes.
- Summer (MJJ, JJA) sees El Niño fade to Neutral (e.g., 8.06796023e-01 in MJJ), while Neutral often turns to La Niña in late winter (FMA: 8.53805698e-01) or fall (ASO: 9.99996721e-01).

- These shifts reflect ENSO’s cycle: cold La Niña winters, warm El Niño fall peaks, and calm Neutral summers.

Core Findings:

- **Seasonal Stability:**
    - High persistence for La Niña in winter (e.g., 9.99999973e-01 in DJF), Neutral in summer (e.g., 9.99412101e-01 in JAS), and El Niño in fall (e.g., 7.27288593e-01 in SON), reflecting ENSO’s cycle of cold winters, calm summers, and warm falls.
    - Direct El Niño to La Niña transitions are rare, with Neutral often acting as a bridge, as seen in strong El Niño to Neutral transitions (e.g., 8.06796023e-01 in MJJ) and Neutral to La Niña shifts (e.g., 8.53805698e-01 in FMA).

- **Key Transitions:**
    - La Niña to El Niño shifts peak in JFM (6.71781175e-01) and OND (7.51477165e-001), signaling El Niño onset.
    - Neutral to La Niña transitions dominate in FMA (8.53805698e-01) and ASO (9.99996721e-01), indicating heat recharge.
    - El Niño to Neutral transitions are strong in MJJ (8.06796023e-01) and JJA (7.21792289e-01), marking summer dissipation.

- **Anomalies:**
    - Low persistence (e.g., 1.24073756e-05 for La Niña in DJF, 1.02952117e-07 for El Niño in FMA) suggests transient states.
    - Near-zero probabilities (e.g., 9.01942876e-135 in OND, 1.40508396e-122 in AMJ) hint at rare shifts, possibly tied to events like 1997–98 El Niño.

Emission Means and Covariances:

- Emission means give the average ONI (sea surface temperature anomaly) and WWV (ocean heat content) for each state, defining their climatic signatures.
- Covariances show variability in ONI and WWV and how they relate (correlation), revealing state consistency and dynamics.

Emission Means:

For each state $i \in \{0,1,2\}$ the emission mean is a 2-dimensional vector that represents expected values of ONI and WWV for that state.

Format:

$$\mu_i = \begin{bmatrix} \mu_{i,ONI} \\ \mu_{i,WWV} \end{bmatrix}$$

Where:

$\mu_{i,ONI}$: Mean ONI for state i

$\mu_{i,WWV}$: Mean WWV for state i

Example DJF:

$$\mu_0 = \begin{bmatrix} -0.64373172 \\ 0.24779321 \end{bmatrix}, \quad \mu_1 = \begin{bmatrix} -0.109633 \\ 0.15579874 \end{bmatrix}, \quad \mu_2 = \begin{bmatrix} 1.68340366 \\ -0.08271336 \end{bmatrix}$$

Emission Covariances:

For each state $i \in \{0,1,2\}$ the emission covariance is a 2×2 matrix that captures:

- The variances of ONI and WWV
- The covariance between ONI and WWV

Format:

$$\Sigma_i = \begin{bmatrix} \sigma_{i,ONI}^2 & \sigma_{i,ONI,WWV} \\ \sigma_{i,ONI,WWV} & \sigma_{i,WWV}^2 \end{bmatrix}$$

Where:

$\sigma_{i,ONI}^2$: Variance of ONI

$\sigma_{i,ONI,WWV}$: Covariance between ONI and WWV

$\sigma_{i,WWV}^2$: Variance of WWV

Example (DJF, State 0)

$$\Sigma_0 = \begin{bmatrix} 0.50879039 & 0.24021446 \\ 0.24021446 & 0.24813871 \end{bmatrix}$$

Month	**Emission Means**	**Emission Covariances**
DJF	[[-0.64373172 0.24779321] [-0.109633 0.15579874] [1.68340366 -0.08271336]]	[[[0.50879039 0.24021446] [0.24021446 0.24813871]] [[0.65004734 0.20579118] [0.20579118 1.12304139]] [[0.41306414 -0.19132983] [-0.19132983 0.47330649]]]
JFM	[[-0.7434777 0.25090848] [0.34801642 0.93840358] [0.29442915 -0.7986091]]	[[[0.22448681 0.01804651] [0.01804651 0.05934425]] [[0.50949129 -0.2755238] [-0.2755238 0.31296194]] [[1.05532958 -0.19529306] [-0.19529306 0.6270247]]]
FMA	[[-0.20408052 0.44216808] [-0.01372328 0.28644064] [1.16159956 -1.68103458]]	[[[0.6023601 -0.20260904] [-0.20260904 0.2491537]] [[0.3216216 0.06541301] [0.06541301 1.18378389]] [[0.23351535 -0.19893384] [-0.19893384 0.48655711]]]
MAM	[[-0.05998354 -0.19110381] [-0.19654738 -0.37894778] [0.09904913 0.50620117]]	[[[0.24824483 0.46848426] [0.46848426 2.05359865]] [[0.55299276 -0.94233537] [-0.94233537 1.72968189]] [[0.35114651 -0.22875481]

		[-0.22875481 0.34257387]]]

AMJ	[[0.03917623 0.01023852]
	 [-0.03649564 0.50615229]
	 [0.45000005 -2.69164884]]
		[[[0.34030822 0.60210957]
		  [0.60210957 1.38017864]]
		
		 [[0.33916963 -0.27610573]
		  [-0.27610573 0.46836728]]
		
		 [[0.00750007 -0.02458484]
		  [-0.02458484 0.35512995]]]

MJJ	[[-0.05737616 -0.11075926]
	 [-0.36656417 0.03107026]
	 [0.26553059 0.42595044]]
		[[[0.65819583 0.9285606]
		  [0.9285606 1.48764815]]
		
		 [[0.03194076 -0.17690559]
		  [-0.17690559 1.91067179]]
		
		 [[0.19061016 -0.23296516]
		  [-0.23296516 0.66507224]]]

JJA	[[0.24250331 -1.37210008]
	 [0.12309166 0.08883298]
	 [-0.15698725 0.74535508]]
		[[[0.4536595 0.47560306]
		  [0.47560306 1.06686103]]
		
		 [[0.74108422 0.88282432]
		  [0.88282432 1.14023236]]
		
		 [[0.11938887 0.02656588]
		  [0.02656588 0.18831336]]]

JAS	[[0.02505847 -0.4256983]
	 [-0.47876463 -0.79219552]
	 [0.19912431 0.75467621]]
		[[[0.64341234 0.29947836]
		  [0.29947836 2.6079531]]
		
		 [[0.29186807 0.17917338]
		  [0.17917338 0.27605281]]
		
		 [[0.47519754 0.26595077]
		  [0.26595077 0.18216996]]]

ASO	[[1.09135744 0.82558615]
	 [-0.47432094 -1.27077673]
	 [-0.12288214 0.38764403]]
		[[[1.2957014 0.17327382]
		  [0.17327382 1.0677738]]
		
		 [[0.33638255 0.47223849]
		  [0.47223849 1.03274006]]
		
		 [[0.40448578 0.41196652]
		  [0.41196652 0.55959173]]]

SON	[[1.16085979 0.75822943] [-0.89665283 -1.35112625]	[[[1.35356005 0.20009604] [0.20009604 0.8613416]]

	[-0.04493297 0.46865348]]	[[0.3011457 0.26558871] [0.26558871 0.73861363]] [[0.47503378 0.40639288] [0.40639288 0.4640799]]]
OND	[[-0.50384884 -0.01240295] [1.03140203 0.96358869] [-0.39997926 -2.21332461]]	[[[0.52733716 0.46305707] [0.46305707 0.59653516]] [[0.95777606 -0.02142416] [-0.02142416 0.05808325]] [[1.49002304 0.71434471] [0.71434471 0.35029442]]]
NDJ	[[-0.17207729 0.07096409] [-0.27634723 0.18014205] [2.40000792 0.23972623]]	[[[0.84295847 0.43046974] [0.43046974 1.2625349]] [[0.92328853 0.46561924] [0.46561924 0.39782929]] [[0.02999404 -0.01163461] [-0.01163461 0.24241183]]]

Key Characteristics:

- El Niño has positive ONI (e.g., 1.68340366 in DJF, 1.16085979 in SON), La Niña has negative ONI (e.g., -0.89665283 in SON, -0.7434777 in JFM), and Neutral is near zero (e.g., -0.01372328 in FMA), matching expected temperature patterns.
- WWV is positive for La Niña and Neutral (e.g., 0.46865348 in SON, 0.44216808 in FMA), showing heat buildup, and negative for El Niño (e.g., -1.68103458 in FMA, -2.21332461 in OND), indicating heat release.
- ONI varies most in El Niño (e.g., 1.05532958 in JFM, 1.49002304 in OND), while WWV varies most in Neutral/La Niña (e.g., 2.05359865 in MAM, 1.18378389 in FMA), showing different uncertainties.

Seasonal Patterns:

- Winter (DJF, JFM) has colder La Niña ONI (-0.64373172, -0.7434777) and positive WWV, signaling ocean heat recharge.

- Fall (SON, OND) shows stronger El Niño ONI (1.16085979, 1.03140203) and negative WWV, marking heat discharge.
- Summer (MJJ, JJA) has weaker ONI (e.g., -0.05737616 in MJJ), reflecting Neutral dominance.
- El Niño often shows negative ONI-WWV links (e.g., -0.433 in DJF, -0.589 in FMA), tying warmer seas to less ocean heat. La Niña links vary (e.g., positive 0.676 in DJF, negative -0.522 in FMA), while Neutral links are weak (e.g., 0.106 in FMA).
- These patterns fit the recharge-discharge theory, where heat builds in La Niña and releases in El Niño.

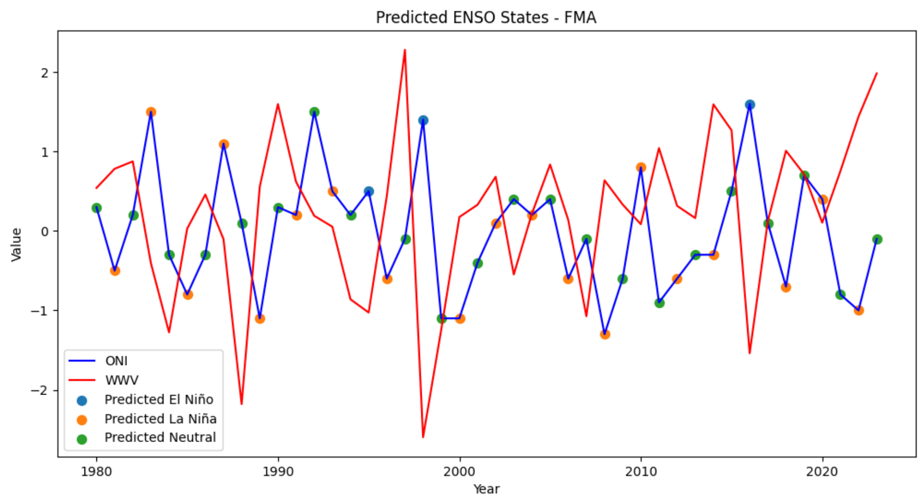
Therefore:

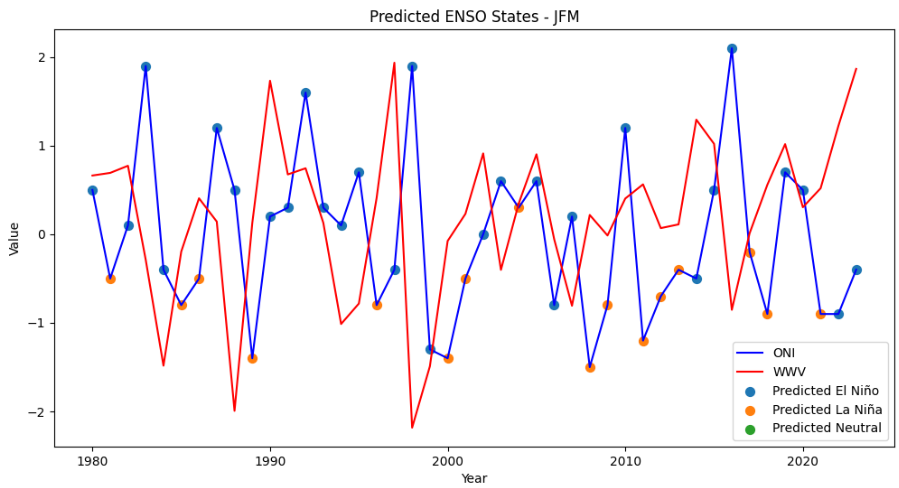
- The HMM captures ENSO's seasonal rhythm: La Niña holds strong in winter, El Niño peaks in fall, and Neutral rules summer, matching real-world patterns.
- ONI and WWV patterns support the recharge-discharge model, with heat building in La Niña/Neutral and releasing in El Niño.
- Variability highlights where ENSO states differ most, aiding prediction of diverse conditions.
- Outliers reveal rare events or model gaps, guiding further study of extreme ENSO phases.

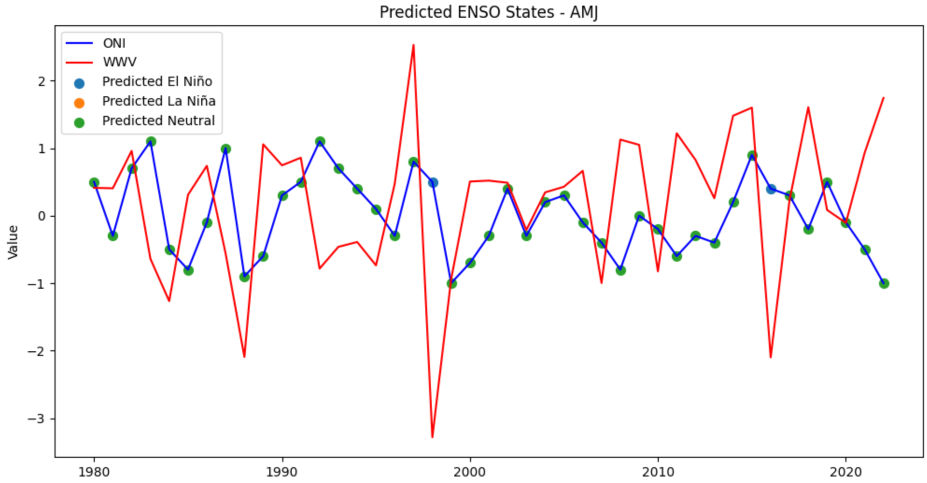
Predictions and state classification done by the Gaussian HMM:

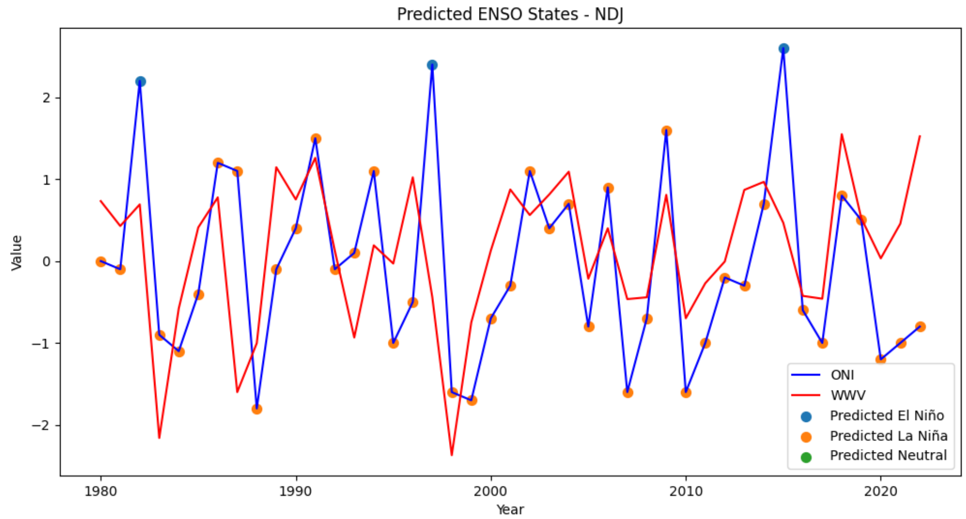
State prediction graphs:

The state prediction graphs for FMA (0.6364), JFM (0.6136), AMJ (0.5116), and NDJ (0.4884) span the full range of HMM accuracies, showcasing the model's best, intermediate, and weakest performances in classifying El Niño, La Niña, and Neutral states over time.









Accuracy scores:

Month	Prediction Accuracy
DJF	0.5581
JFM	0.6136
FMA	0.6364
MAM	0.6047
AMJ	0.5116
MJJ	0.5814
JJA	0.6279
JAS	0.5116
ASO	0.5116
SON	0.6047
OND	0.6047
NDJ	0.4884

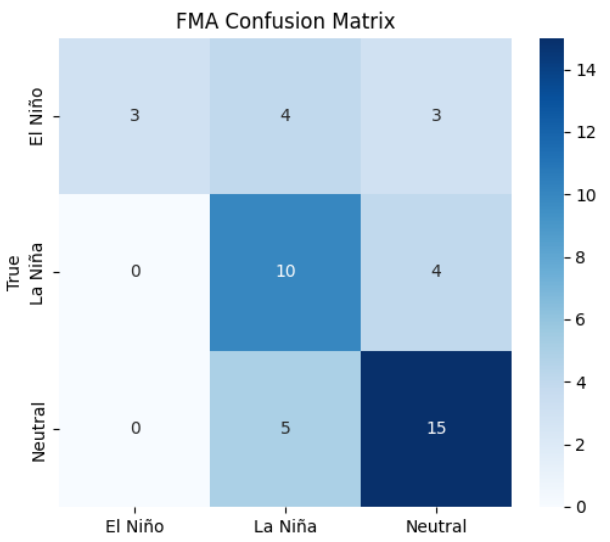
Note: Low accuracies, particularly in winter (e.g., NDJ, DJF), result from the Markovian assumption, which assumes transitions depend only on the current state, potentially oversimplifying ENSO's complex dynamics influenced by longer-term climatic memory. High accuracy was not the project's main objective; instead, the focus was on capturing ENSO's statistical properties and transition patterns to understand its dynamics, with moderate accuracies (0.4884–0.6364) deemed sufficient for this purpose.

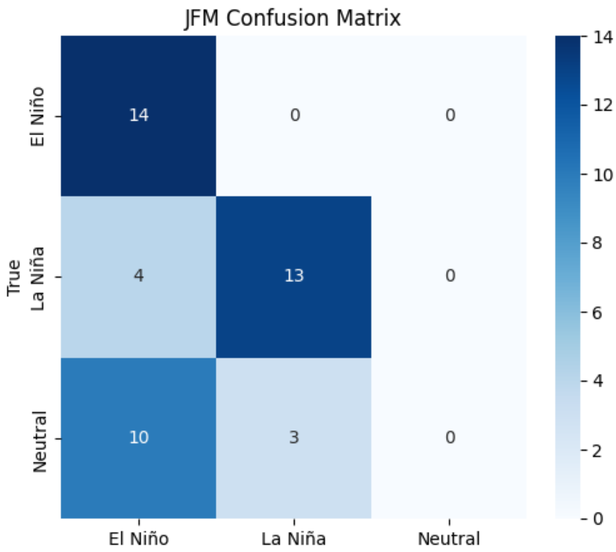
Confusion Matrices:

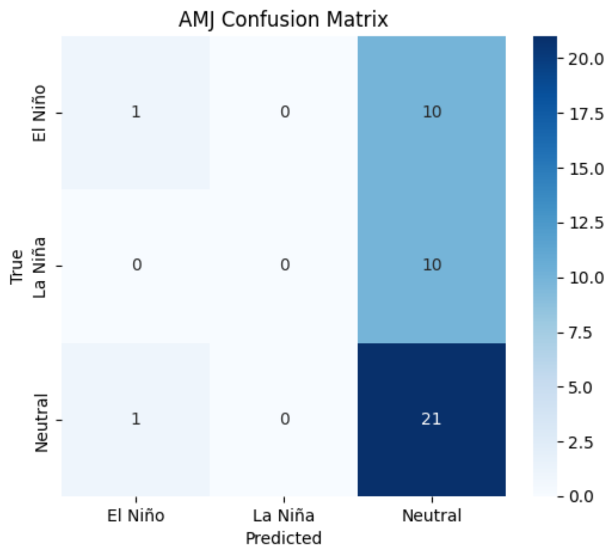
A confusion matrix C is represented as

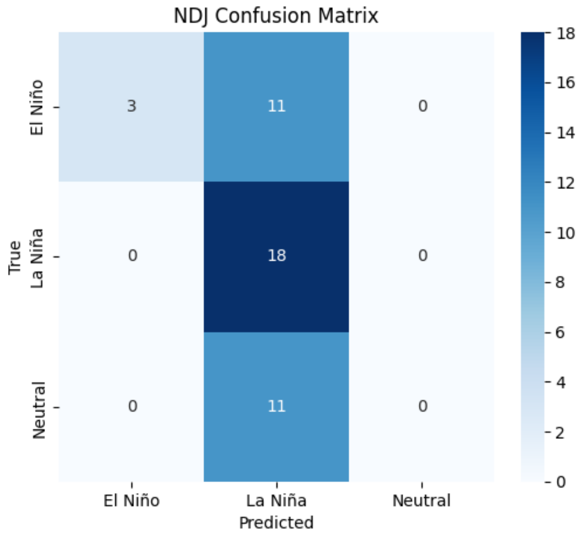
$$C = \begin{bmatrix} C_{00} & C_{01} & C_{02} \\ C_{10} & C_{11} & C_{12} \\ C_{20} & C_{21} & C_{22} \end{bmatrix}$$

Where C_{ij} = Number of samples whose true class is i and predicted class is j.









Month	Confusion Matrix
DJF	[[6 8 0] [0 18 0] [0 11 0]]
JFM	[[14 0 0] [4 13 0] [10 3 0]]
FMA	[[3 4 3] [0 10 4] [0 5 15]]
MAM	[[0 2 7] [0 4 6] [0 2 22]]
AMJ	[[1 0 10] [0 0 10] [1 0 21]]
MJJ	[[4 0 7] [4 0 4] [3 0 21]]
JJA	[[6 0 2] [4 0 5] [5 0 21]]
JAS	[[0 0 8] [0 5 8] [0 5 17]]
ASO	[[4 1 7] [1 4 8] [1 3 14]]
SON	[[5 0 8] [1 8 6] [1 1 13]]
OND	[[10 4 0] [1 16 0] [3 9 0]]
NDJ	[[3 11 0] [0 18 0] [0 11 0]]

Inferences on Confusion Matrices:

- **Classification Strengths:**
  - Strong La Niña classification in DJF (18/18 correct in State 1) and Neutral in FMA (15/20 correct in State 2) indicate robust identification of dominant states in winter and transitional seasons, reflecting clear ONI/WWV signatures.
  - High Neutral accuracy in MAM (22/24 in State 2) and JJA (21/26 in State 2) aligns with summer Neutral dominance, as per cluster summaries.

- **Common Misclassifications:**
  - Frequent El Niño misclassification as La Niña or Neutral (e.g., DJF: 11/11 State 2 misclassified, FMA: 5/5 State 2 as Neutral) suggests challenges in capturing El Niño's weaker signals, especially in winter and spring.
  - La Niña often mistaken for Neutral in JAS (8/13 State 1 as Neutral) and ASO (8/13 State 1 as Neutral), indicating overlapping ONI/WWV distributions in fall.

- **Seasonal Variability:** Confusion matrices show better El Niño classification in fall (e.g., SON: 8/13 State 0 correct, ASO: 4/12 correct) due to stronger ONI signals (e.g., 1.16085979 in SON), but winter struggles (e.g., NDJ: 0/11 State 2 correct) highlight model limitations in stable periods.

Key Insights:

- Accuracies highlight stronger performance in transitional seasons, with winter challenges due to stable states.

- Predicted state graphs visualize ENSO's cyclic patterns, emphasizing Neutral's bridging role and anomaly detection.

- Confusion matrices pinpoint reliable La Niña/Neutral classifications but El Niño struggles, particularly in winter.

Conclusion

The Seasonal Hidden Markov Models developed in this study effectively capture the probabilistic dynamics of ENSO phases, revealing distinct seasonal patterns in El Niño, La Niña, and Neutral states. The models highlight strong La Niña persistence in winter, Neutral dominance in summer, and El Niño peaks in fall, consistent with ENSO's natural cycle. Key transitions, such as La Niña shifting to El Niño in spring and El Niño fading to Neutral in summer, align with the recharge-discharge mechanism of ocean heat. Emission characteristics confirm expected patterns, with El Niño linked to warmer sea surface temperatures, La Niña to cooler temperatures, and Neutral to near-average conditions.

While achieving moderate classification accuracies (ranging from 0.4884 to 0.6364), the models perform best for La Niña and Neutral states, particularly in winter and summer, but face challenges with El Niño, especially in winter, due to the simplifying Markovian assumption. This limitation highlights ENSO's complex, long-term dependencies, which a first-order Markov model struggles to fully capture. Nevertheless, the study meets its goal of elucidating statistical properties and transition patterns rather than prioritizing high predictive accuracy.

Future work could explore higher-order Markov models, incorporate additional climate indices, or leverage advanced machine learning techniques to better capture ENSO's complexity. This study demonstrates the value of HMMs for probabilistic climate analysis, providing a robust framework for studying ENSO's behavior and its implications for climate-sensitive regions like India, where monsoon variability significantly impacts agriculture and society.