Applying Hidden Markov Models (HMMs) in AFRICRIX for Crisis Signal Analysis

1. Observations in AFRICRIX

*In AFRICRIX, the observations for an HMM would be **measurable crisis-related signals** captured over time from various data streams. These include:*

- *Real-time social media signals (e.g., tweet counts, trending crisis keywords, Telegram messages)*
- *News article event counts (e.g., "flood" mentions per hour)*
- *Satellite-based indices (e.g., NDVI, water extent)*
- *Citizen report counts and sensor readings (e.g., number of uploads, geolocated sensor pings)*
- *Temporal features: rolling averages, 24h signal deltas, 7-day slopes, and anomaly scores*
  Each observation at time t is a vector summarizing the system's detected signal strength and type across sources and locations.

2. Type of HMM Problem

*I **do not know the true underlying "crisis state"** (e.g., "calm", "building risk", "escalating crisis") at each timestep in real life—these are hidden. We only observe noisy, multi-source signals. Thus, **this is an unsupervised HMM (learning hidden states)**, commonly formulated as an HMM learning problem:*

- ***Objective**: Infer the sequence of hidden "crisis states" and learn their temporal dynamics purely from the observations.*

3. Training Algorithm

a. Known Values at the Start

- *The sequence of observed vectors over time ($O = o_1, o_2, ..., o$), e.g., hourly features from signals.*

- *(Optionally) Initial guesses for the number of hidden states (e.g., K = 3: calm, rising risk, crisis).*

- *(Optionally) Initial parameter estimates (can be random or from clustering, e.g., K-means).*

b. Unknown Values to Learn

- ***State transition probabilities** (A): How likely is it to move from one hidden crisis state to another (e.g., calm → rising risk)?*

- ***Emission probabilities** (B): How likely is each observable signal pattern given a particular hidden state (often modeled as multivariate Gaussian for continuous features)?*

- ***Initial state probabilities** (π): Probability of starting in each hidden crisis state.*

*The **Baum-Welch (EM) algorithm** is used to learn these parameters by maximizing the likelihood of the observed signal sequences.*

4. Parameter Updates in Training

During training, the HMM iteratively updates:

- *The **transition matrix (A)**: Probabilities of moving between hidden states, e.g., from "building risk" to "active crisis".*

- *The **emission parameters (B)**: For each hidden state, the parameters of the probability distribution generating the observed features (means/covariances for Gaussian HMM).*

- *The **initial state distribution (π)**: Probability of the system starting in each hidden state.*

*All these are updated via **expectation-maximization**: computing posteriors over state sequences (E-step) and then updating parameters to maximize data likelihood (M-step).*

In summary:
AFRICRIX uses an unsupervised HMM to model and infer the hidden crisis states underlying streams of real-time signals. The HMM is trained using the Baum-Welch algorithm, which learns the transition, emission, and initial state probabilities. The model continually updates these parameters based on incoming data, allowing the

system to detect subtle transitions from "safe" to "at-risk" to "crisis"—enabling proactive alerting before human operators can detect the pattern.