Hidden Markov Model (HMM) for Flood Risk Prediction in South Sudan

Describe the Observations

The model will use multiple measurable Earth observation datasets to capture flood-related patterns:

- Rainfall data from CHIRPS (daily or weekly amounts)
- Flood proxy data from Sentinel-1 SAR (surface water extent, flooded area)
- Soil moisture data (e.g., SMAP or ESA CCI) to estimate ground saturation levels
- River discharge or flow estimates (where available, e.g., from hydrological models or local gauge data)
- Elevation data from SRTM (to group regions by topography and floodplain zones; used as static context, not as a time-varying sequence)

These combined time-series observations help the model learn patterns that signal when floods are likely to occur or persist.

Type of HMM Problem

The hidden states, actual flood risk levels over time (e.g., None, Low, Medium, High) are not known in advance. This makes it an unsupervised HMM problem: the goal is to infer the most likely hidden flood risk states and how they transition over time, using only the observed climate, soil, hydrological, and satellite data.

Training Algorithm

a. Known values:

- The historical time-series data: rainfall, flood extent, elevation
- The number of hidden states (chosen by the researcher, e.g., 3–4 risk levels)
- Initial guesses for state and transition probabilities (e.g., uniform)

b. Unknown values:

- The actual sequence of hidden flood risk states for each time step
- The transition probabilities (how likely flood risk changes over time)
- The emission probabilities (how likely certain rainfall/flood patterns appear under each risk level)

Parameter Updates

The Baum-Welch algorithm (an Expectation-Maximization method) will iteratively update:

- **Initial state probabilities** (likelihood of starting in each risk level)
- **Transition matrix (A):** how flood risk levels shift from one time step to the next

- **Emission probabilities (B):** how observed rainfall and SAR patterns match each risk level