Weather is often very hard to predict even when we have a sense of the weather measures. Nevertheless, we have attempted to model weather events based on weather measures using a second order Hidden Markov Model. We have 9 hidden states representing weather events and 7 continuous real-valued weather measures as our observations. The observations have been modeled using Gaussian and Log-Normal distributions and the emissions probability is derived from a multivariate normal. We trained on 50 years of data and tested on the next five years.

* Dataset – Weather measure data for Boston over the past 55 years
* Training – first 50 years, test on the next 5 years
* TP – tracked over the entire training period – studied its stationarity
* EP – for each state – multivariate normal represents EP for the 7 observations made
* The multivariate normal constructed from underlying gaussian and log-normal distributions representing each of the 7 real-valued weather measures
* Results – overall inference accuracy of 65%
* Specifically, inferred normal days, extreme events like thunderstorms and specific events like Fog
* Model did less well on highly overlapping events like Rain, Rain-Snow, Fog-Snow etc. However, still considerable better than a naïve baseline.
* Confusion matrix to represent nature of error (not all labels are equidistant; ordering exists)
* Further improve with - a third order model, better/more parameters, better EP/TP