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

- TP – tracked over the entire training period – studied its stationarity

-

Results – overall inference accuracy of 65 percent

- 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

- Infer hidden states from 2011 to 2015 based on data from the prior 50 years

- Do the same with a third order model and show improvement in inference accuracy

○ Highlight good performance for extreme events

§ Specific examples like Hurricane Sandy (other such events), the New England winter of 2014-15

○ Talk about going from weather measures to actual weather events that impact life and society

○ Talk about inferring weather events in places/times where we can't directly experience it - weather in the past/future, weather on Mars

- Talk about limitations of this approach vis-a-vis physics based models but think about the difference between statistically predicting weather measures as opposed to statistically inferring weather events given weather measures (that are computed potentially through physics models)

- Think about improving performance using more parameters or better parameters (max/min instead of mean); better models for EP

\begin{tabular}{| c | c | c |}

\hline

\multicolumn{3}{| c |}{Hidden States} \\ \hline

Normal & Fog-Rain & Snow \\ \hline

Fog-Snow & Fog & Rain-Snow \\ \hline

Fog-Rain-Snow & Rain & Thunderstorm \\

\hline

\end{tabular}

\begin{tabular}{| c | c |}

\hline

\multicolumn{2}{| c |}{Observations} \\ \hline

Mean Temperature & Mean Dew Point \\ \hline

Mean Humidity & Mean Wind Speed \\ \hline

Precipitation & Mean Visibility \\ \hline

\multicolumn{2}{|c |}{Mean Sea Level Pressure} \\

\hline

\end{tabular}