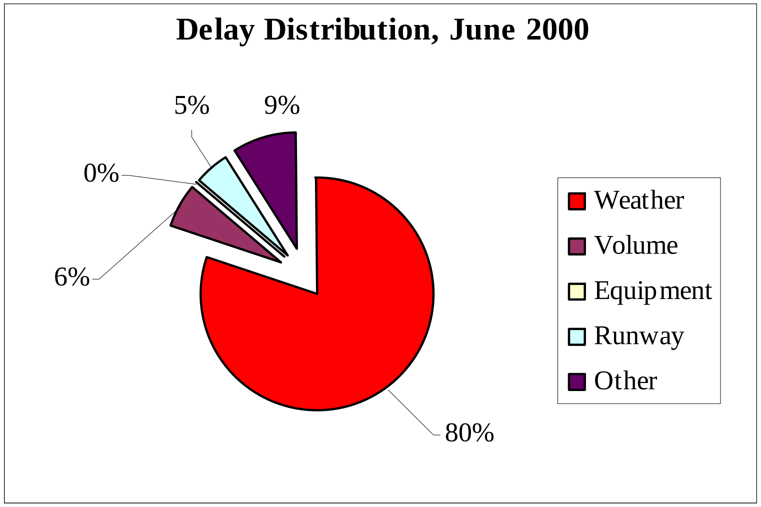
Toward Probabilistic Forecasts of Convective Storm Activity for En Route Air Traffic Management

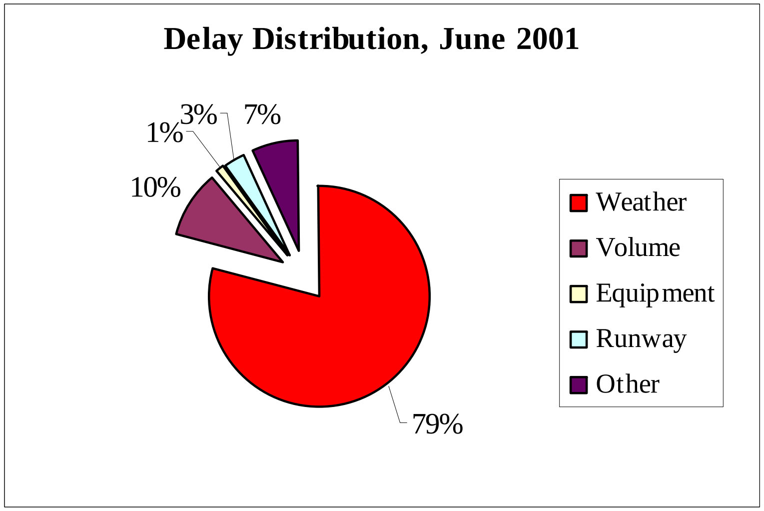
P. Barry Liu
Prof. Mark Hansen
University of California, Berkeley

September 25th, 2003

Background

- Weather is one of the primary factors in air traffic delay

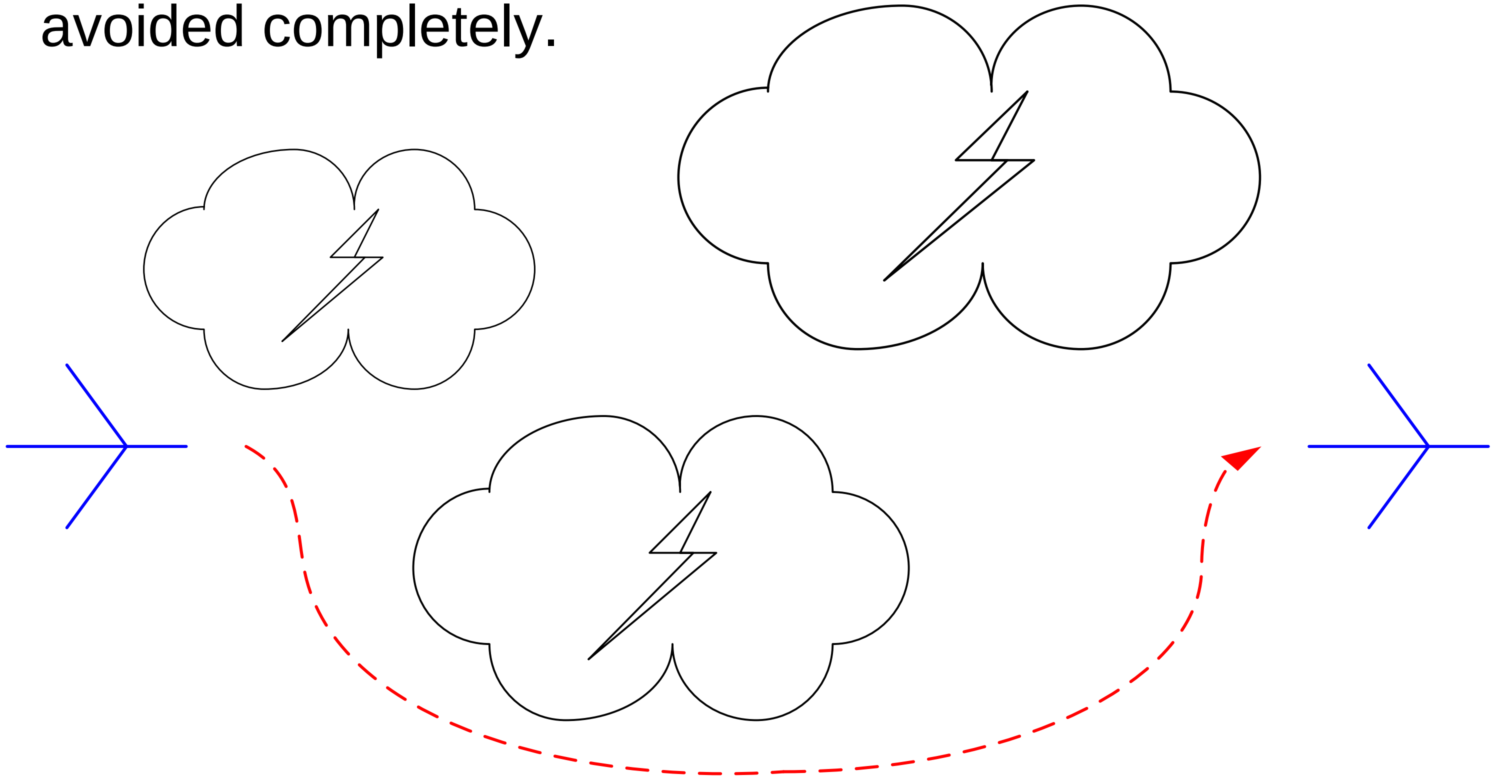




(According to FAA, from *Robust Dynamic Routing of Aircraft under Uncertainty*, Nilim et al.)

Background

- Current practice of ATM: the predicted storm zones are avoided completely.



Background

Unsatisfactory Forecast Performance

+

Complete Avoidance of the Predicted Storm Zones

||

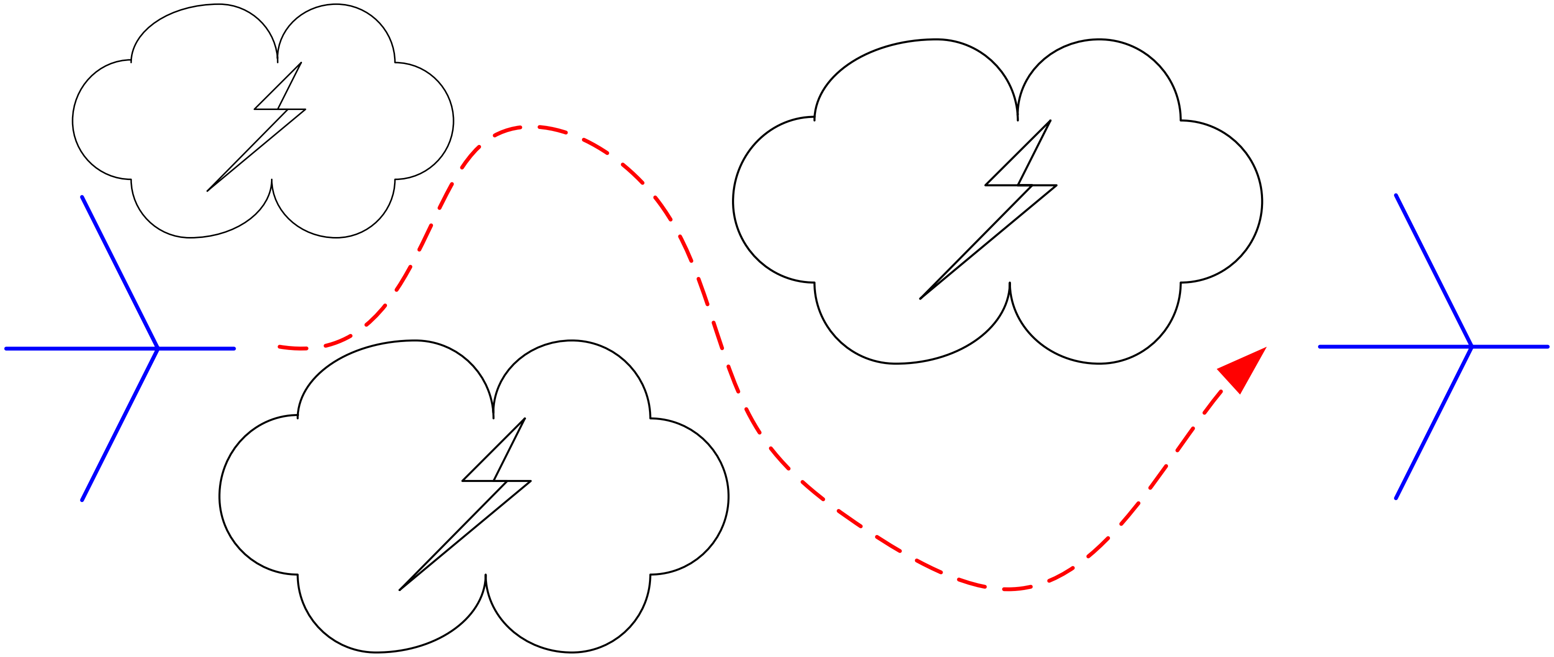
Overly Conservative Routing Decisions

+

Much More Delays than the Unavoidable

Background

- How about…take a less conservative route



Background

- Dynamic routing strategies based on this concept were developed mostly under deterministic assumptions or in a simplified probabilistic setting
- But weather is stochastic in nature...

$\Rightarrow$ Investigate ways to provide probabilistic convective weather forecasts with higher accuracy in terms of convective activity probabilities for flight links to support real-time aircraft routing decision.

Introduction

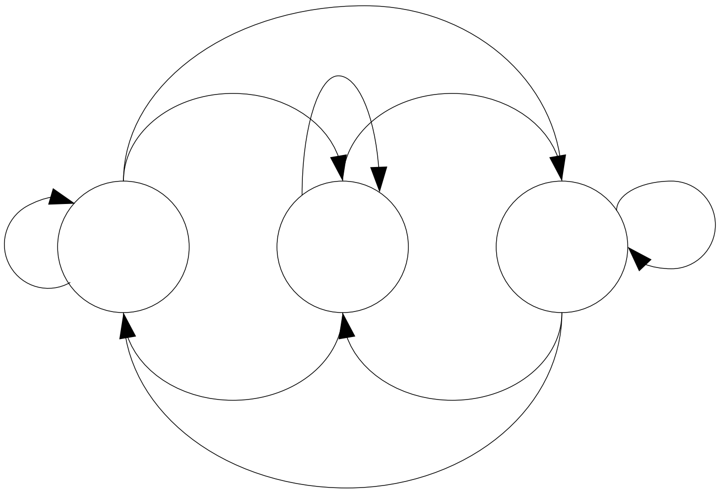
- **Goal:** provide a better prediction of convective weather in explicit probabilities defined specifically in the aircraft routing context to aid aircraft routing decision-making

- **Approach:** develop a *stochastic model* depicting the evolution of the convective weather

- **Modeling framework:** take the evolution of convective weather as a Markov process $\Rightarrow$ Future event can be predicted based on current information
  - Markov Model
  - Hidden Markov Model

Markov Model

- First order Markov chain
  - finite states (i.e. weather states)
  - future state S_{n+1}, is independent of the past states and depends only on the present state S_n
  - Transition probabilities



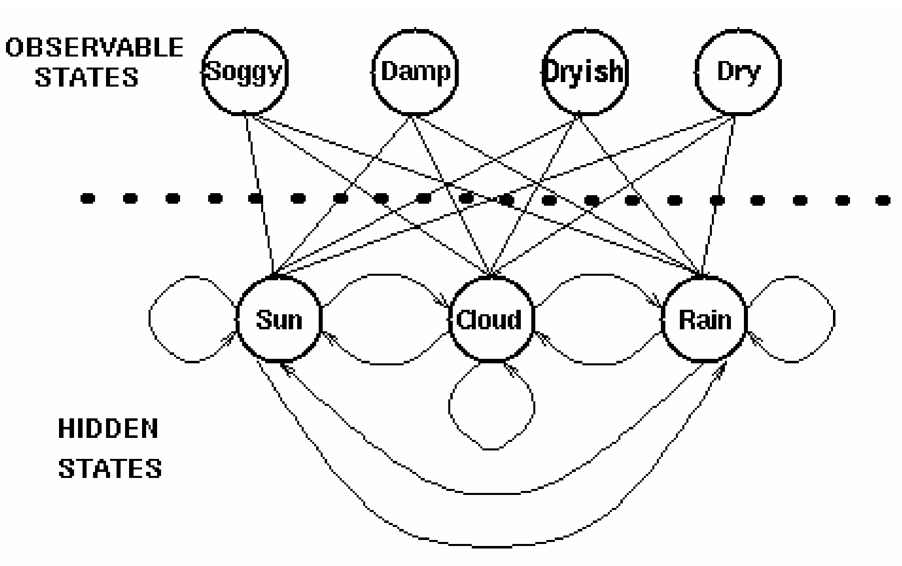
Pij

$= P\{S_{n+1} = j \mid S_n = i, S_{n-1} = i_{n-1}, \ldots, S_1 = i_1, S_0 = i_0\}$

$= P\{S_{n+1} = j \mid S_n = i\}$

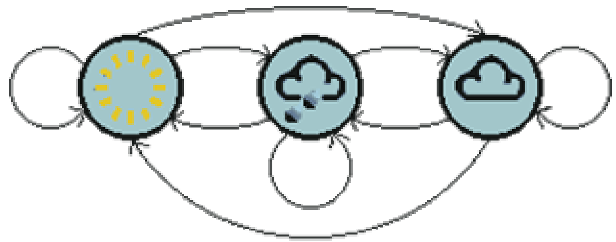
Seeing isn't believing

- What we observe does not necessarily have 1-1 mapping on the state that the system is in -- the state of the system is hidden
- Example: deduce the weather from a piece of seaweed



⇨ Hidden Markov Model!

Modeling Frameworks



- Markov Model

→ → → →

- Hidden Markov Model (HMM)

→ → → →

↓ ↓ ↓

Model Definition - Markov Model

- First-order Markov Model is fully characterized by the transition probabilities
- Example: Three states--State 1, State 2, State 3

Transition Matrix	**To**		
From	State 1	State 2	State 3
State 1	0.5	0.4	0.1
State 2	0.2	0.6	0.2
State 3	0.1	0.4	0.5

Model Development & Prediction - Markov Model

- Assume this is a discrete time Markov chain (transition occurs every t minutes)
- Model parameters: the transition probabilities
- Estimate the parameters directly from the data (the maximum likelihood estimator)

$$\hat{P}_{ij} = \frac{\#\ of\ transitions\ from\ state\ i\ to\ state\ j}{\#\ of\ visits\ to\ state\ i}$$

- Prediction: Given the current states, the prediction for n periods later could be made by applying the transition probabilities n times

Model Definition - Hidden Markov Model

- A HMM is defined by
  - number of states
  - initial state probabilities
  - state transition matrix
  - confusion matrix (emission probabilities)

→ → → →
↓ ↓ ↓

	Confusion Matrix					Transition Matrix			
	p(1 | *)	p(2 | *)	p(3 | *)	p(4 | *)	p(5 | *)	p(Low | *)	p(Med | *)	p(High | *)	p(End | *)
p(* | Low)	0.319	0.261	0.227	0.128	0.065	0.555	0.386	0.051	0.008
p(* | Med)	0.117	0.168	0.277	0.282	0.155	0.353	0.302	0.314	0.031
p(* | High)	0.115	0.133	0.291	0.263	0.198	0.12	0.417	0.432	0.031
p(* | Begin)						0.72014	0.15897	0.1209	

Model Development & Prediction - Hidden Markov Model

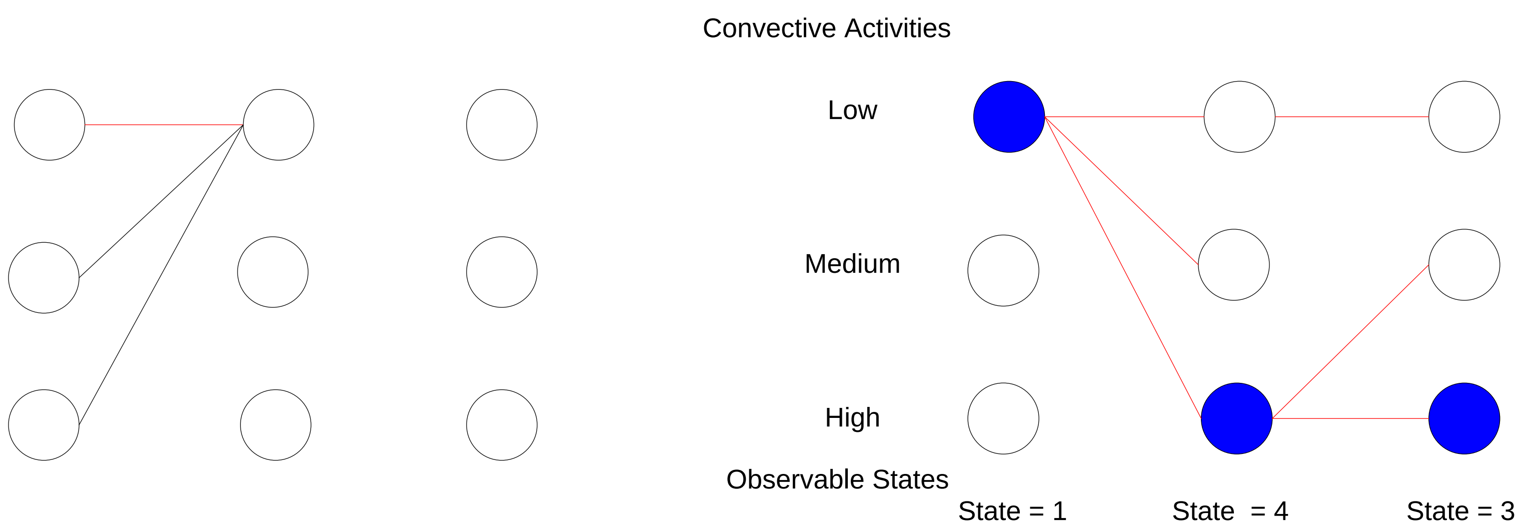
- Model parameters estimation (transition & confusion matrices)
  - Baum-Welch Algorithm
    - Posterior probabilities - Forward & Backward algorithms
    - EM algorithm - maximum likelihood with missing data
- Determine current hidden state
  - Viterbi Algorithm
    - Given the output state sequence and the model parameters
    - Determine the most probable state path
- Prediction
  - Current state + confusion matrix => probabilistic prediction of future weather

Problem Setup

- Given: An output sequence: $O_0, \ldots, O_n$
- Assume: the output sequence is generated by a HMM
- Objective function to maximize:
  - $P(O_0, \ldots, O_n)$ -- the log-likelihood of having this output sequence
- Hypothetical parameters / decision variables:
  - Transition probabilities
  - Emission probabilities
  - Initial and end hidden state probabilities
- Techniques to estimate the model parameters:
  - Calculate forward and backward probabilities recursively based on hypothetical parameters
  - Estimate the parameters (based on multiple output sequences)
  - Iterate till the objective function value reaches convergence

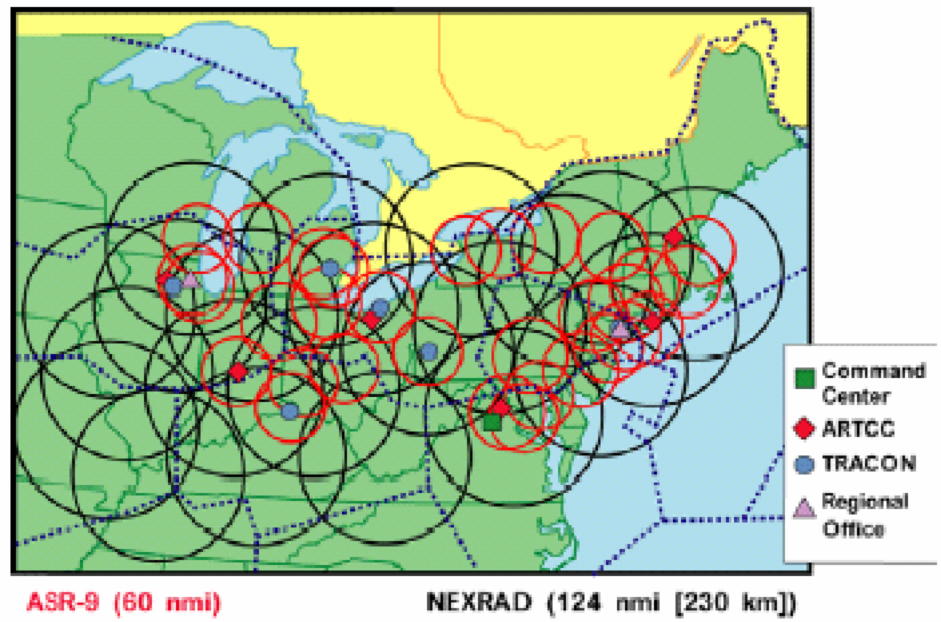
Finding the most probable hidden state path (Viterbi algorithm)

- Given:
  - An output sequence
  - HMM parameters: Transition and Confusion matrices



Data Source

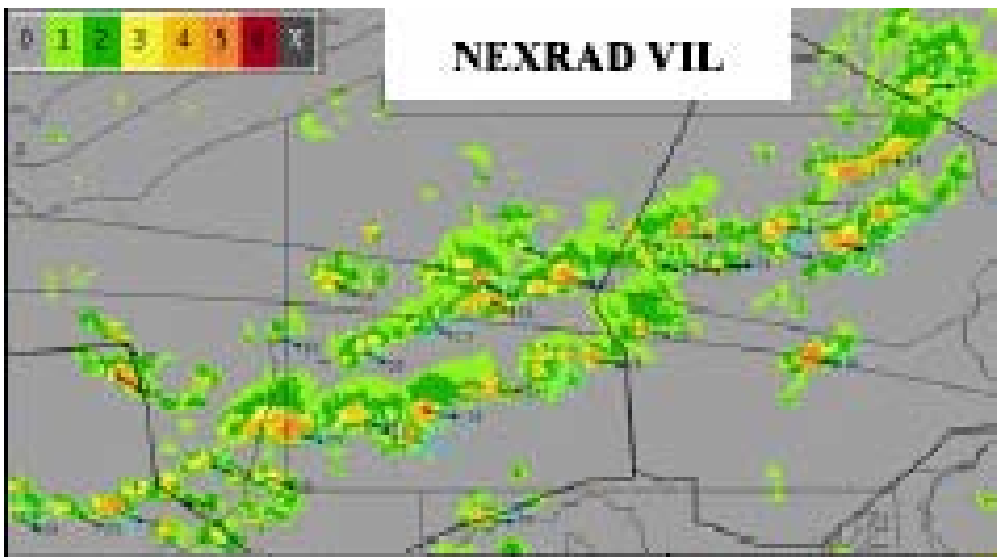
- MIT Lincoln Lab Corridor Integrated Weather System (CIWS) products
- Coverage: The Northeast Corridor in the United States ~ 4 million 1 km x 1km cells, ~700k valid cells



Coverage of sensors integrated in the 2002 CIWS demonstration.

Data Source

- Convective weather states are labeled with Video Integrator and Processor (VIP) levels from 0 to 6
- VIP level ≥ 3: not flyable
- Data include both actual and forecasts



200500	39.212509	75.008893	1
200500	39.212509	74.985601	1
200500	39.212509	74.962309	1
200500	39.212509	74.939018	1
200500	39.212509	74.915726	1
200500	39.212509	74.892435	1
200500	39.212509	74.869143	1
200500	39.212509	74.845852	1
200500	39.212509	74.822560	1
200500	39.212509	74.799269	1
200500	39.212509	74.775977	1
200500	39.212509	74.752685	1
200500	39.212509	74.729394	1
200500	39.212509	74.706102	1
200500	39.212509	74.682811	1
200500	39.212509	74.659519	1
200500	39.212509	74.636228	1
200500	39.212509	74.612936	1
200500	39.212509	74.589645	1
200500	39.212509	74.566353	1
200500	39.212509	74.543061	1

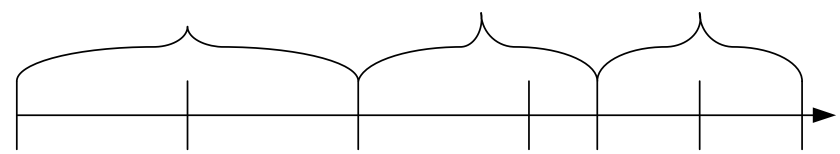
Unit Area for Evaluation - Flight Link

- Raw data: values for 1 km x 1km cells
- Assume en route flight speed ~ 500 mi
- Dimension of unit area
  - Length: distance traveled in 5 min. : ~60 km
  - Width: flight path width: ~ 12km
- Strip level = max{cell level$_1$, …, cell level$_{60}$}
- Band level = min{strip level$_1$, …, strip level$_{12}$}

A Case Study

Implementation

- Data set: Aug 24, 2002, 46 time points, 5 minutes apart



- Coded in Java
- Steps:
  - Determine the storm level for the flight links
  - Define the hidden states and output states
  - HMM parameter estimation using Baum-Welch algorithm
  - Use Viterbi algorithm to find the most probable current state
  - Use transition and confusion matrices to predict storm levels at future time periods (12 periods -- 1 hour)

Implementation Decisions

- Number of hidden states: 3
- Number of output states: 5
- Seed matrices
- Stopping rule for algorithm iterations:

  | difference of two consecutive LLs | < 0.1

<table>
<tr><th>VIP level</th><th>Output</th></tr>
<tr><td>0</td><td>1</td></tr>
<tr><td>1</td><td>2</td></tr>
<tr><td>2</td><td>3</td></tr>
<tr><td>3</td><td rowspan="2">4</td></tr>
<tr><td>4</td></tr>
<tr><td>5</td><td rowspan="2">5</td></tr>
<tr><td>6</td></tr>
</table>

	Confusion Matrix					Transition Matrix			
	p(1 | *)	p(2 | *)	p(3 | *)	p(4 | *)	p(5 | *)	p(Low | *)	p(Med | *)	p(High | *)	p(End | *)
p(* | Low)	0.3	0.3	0.2	0.1	0.1	0.5	0.4	0.05	0.05
p(* | Med)	0.2	0.2	0.2	0.2	0.2	0.3	0.3	0.3	0.1
p(* | High)	0.1	0.2	0.2	0.2	0.3	0.1	0.4	0.4	0.1
p(* | Begin)						0.5	0.3	0.2	

Estimation Experience

- With different seed transition matrices
  - Matrices with difference within certain range yield similar estimated parameters
  - Matrices with significant difference yield drastically different result parameters
  - Known issue of HMM parameter estimation: converging to local maximum
- Runtime statistics

# of locations	10	611	611
# of training periods	46	30	46
# of iterations	39	18	18
Log-Likelihood	-763.203	-16660.1	-23729.6
Estimation time (millisec)	330	3410	5060

Hardware: Pentium III processor, 128 MB RAM

Parameter estimation results

- Number of locations: 611
- Number of time periods: 20
- Number of iterations: 18
- Log-likelihood: -16660.136
- Elapsed time for parameter estimation: 3290 milliseconds
- Initial state probabilities: [0.586, 0.208, 0.205]

Transition probabilities			
from \ to	state 0	state 1	state 2
state 0	0.9365	0.0297	0.0008
state 1	0.0731	0.8478	0.0498
state 2	0.0021	0.0326	0.9278

Emission probabilities					
from \ emit	1	2	3	4	5
state 0	0.1694	0.8108	0.0199	4.76E-18	1.03E-19
state 1	1.04E-05	0.0488	0.8980	0.0533	5.34E-12
state 2	3.78E-34	3.39E-08	0.0108	0.6237	0.3655

Sample Prediction Results

Time		Band 1	Band 2	Band 3	Band 4
t=-20	Actual State	2	1	4	2
t=-15	Actual State	3	1	5	1
t=-10	Actual State	4	2	5	2
t=-5	Actual State	3	2	5	1
t=0	Actual State	3	2	4	1
t=5	P(Observable State = 1)	0.04	0.23	0.03	0.31
	P(Observable State = 2)	0.12	0.33	0.09	0.28
	P(Observable State = 3)	0.42	0.21	0.24	0.19
	P(Observable State = 4)	0.25	0.17	0.38	0.15
	P(Observable State = 5)	0.17	0.06	0.26	0.07
	Actual State	3	2	4	1
t=10	P(Observable State = 1)	0.05	0.07	0.05	0.29
	P(Observable State = 2)	0.12	0.24	0.12	0.33
	P(Observable State = 3)	0.24	0.39	0.25	0.17
	P(Observable State = 4)	0.38	0.25	0.37	0.14
	P(Observable State = 5)	0.21	0.05	0.21	0.07
	Actual State	4	3	4	2

Model Forecast Results

Hidden Markov Model

15 min ahead	Probability of Prediction	
Actual Condition	Flyable	Not Flyable
Flyable	90.85%	9.15%
Not Flyable	25.28%	74.72%

30 min ahead	Probability of Prediction	
Actual Condition	Flyable	Not Flyable
Flyable	87.60%	12.40%
Not Flyable	31.83%	68.17%

45 min ahead	Probability of Prediction	
Actual Condition	Flyable	Not Flyable
Flyable	84.25%	15.75%
Not Flyable	40.75%	59.25%

60 min ahead	Probability of Prediction	
Actual Condition	Flyable	Not Flyable
Flyable	82.18%	17.82%
Not Flyable	47.95%	52.05%

Markov Model

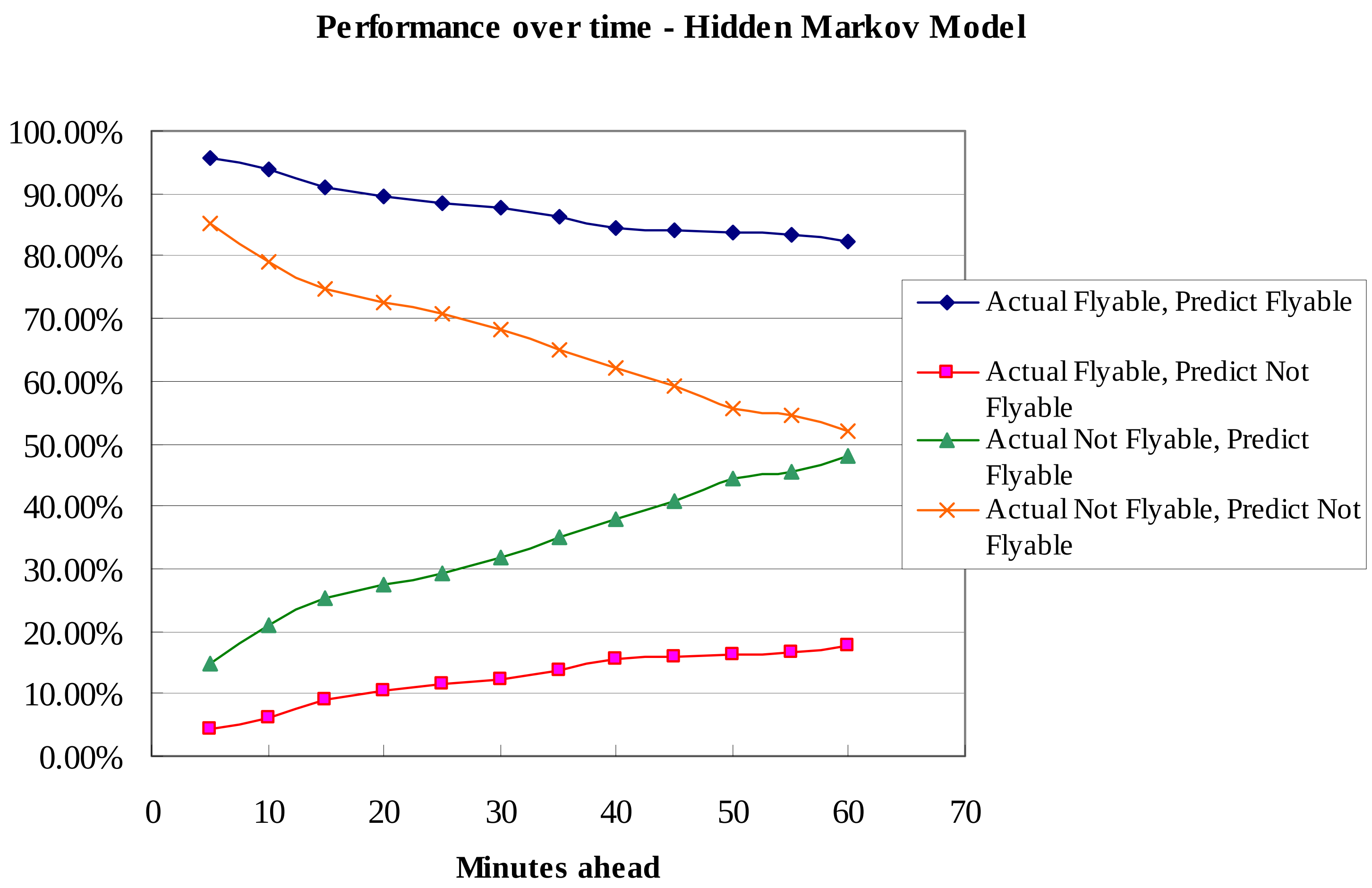
15 min ahead	Probability of Prediction	
Actual Condition	Flyable	Not Flyable
Flyable	90.08%	9.92%
Not Flyable	29.27%	70.73%

30 min ahead	Probability of Prediction	
Actual Condition	Flyable	Not Flyable
Flyable	86.23%	13.77%
Not Flyable	38.31%	61.70%

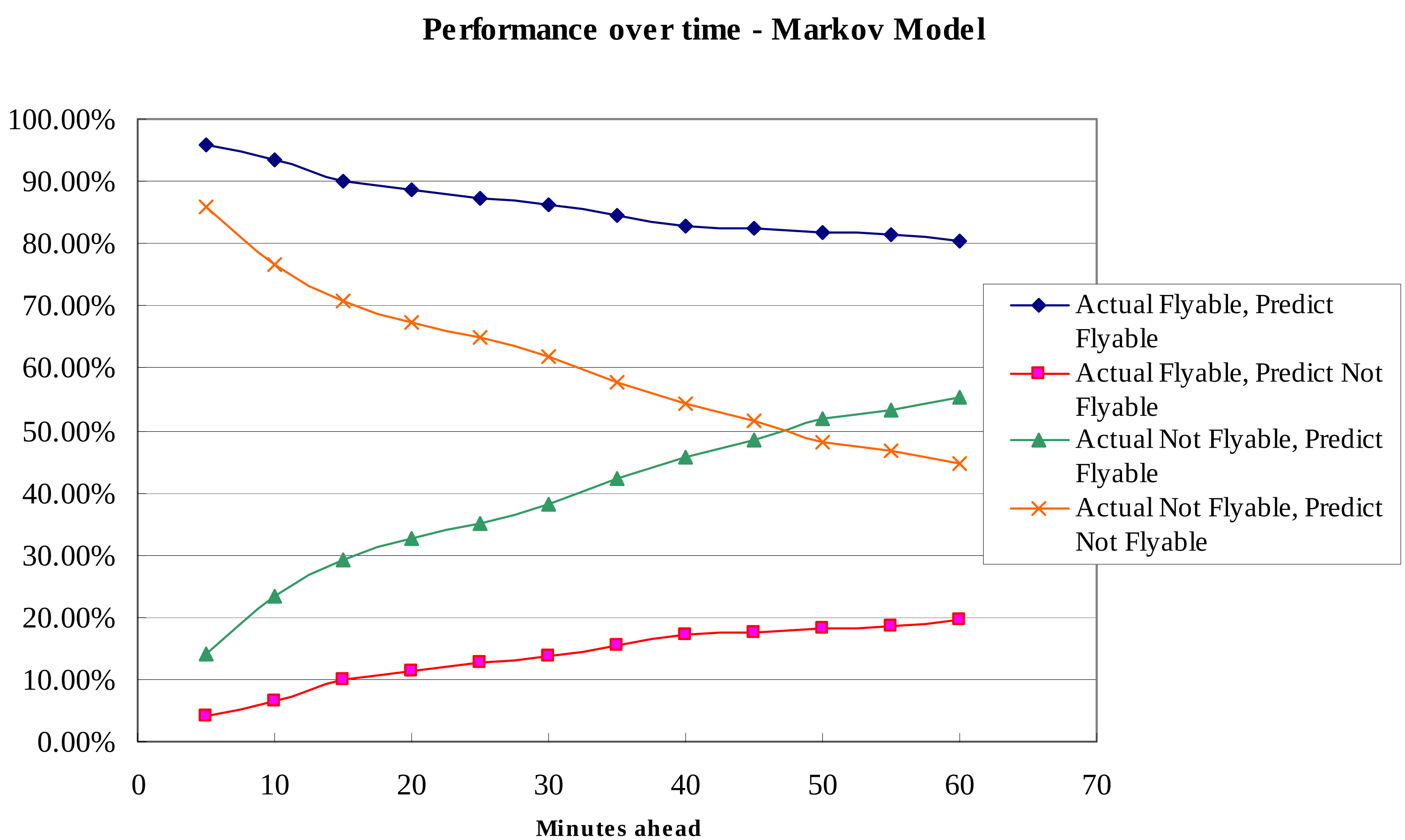
45 min ahead	Probability of Prediction	
Actual Condition	Flyable	Not Flyable
Flyable	82.42%	17.59%
Not Flyable	48.52%	51.48%

60 min ahead	Probability of Prediction	
Actual Condition	Flyable	Not Flyable
Flyable	80.28%	19.73%
Not Flyable	55.37%	44.63%

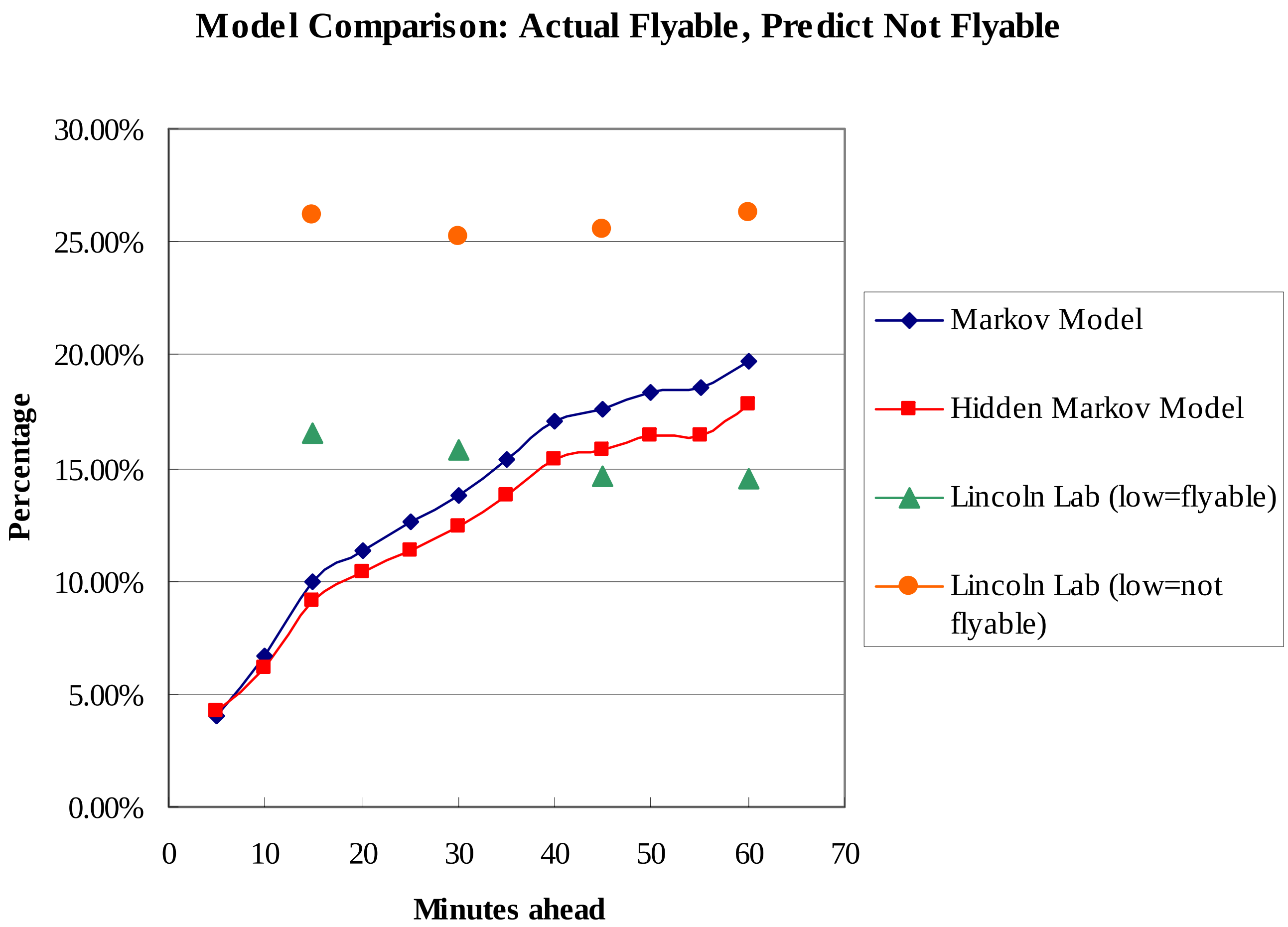
Performance of HMM



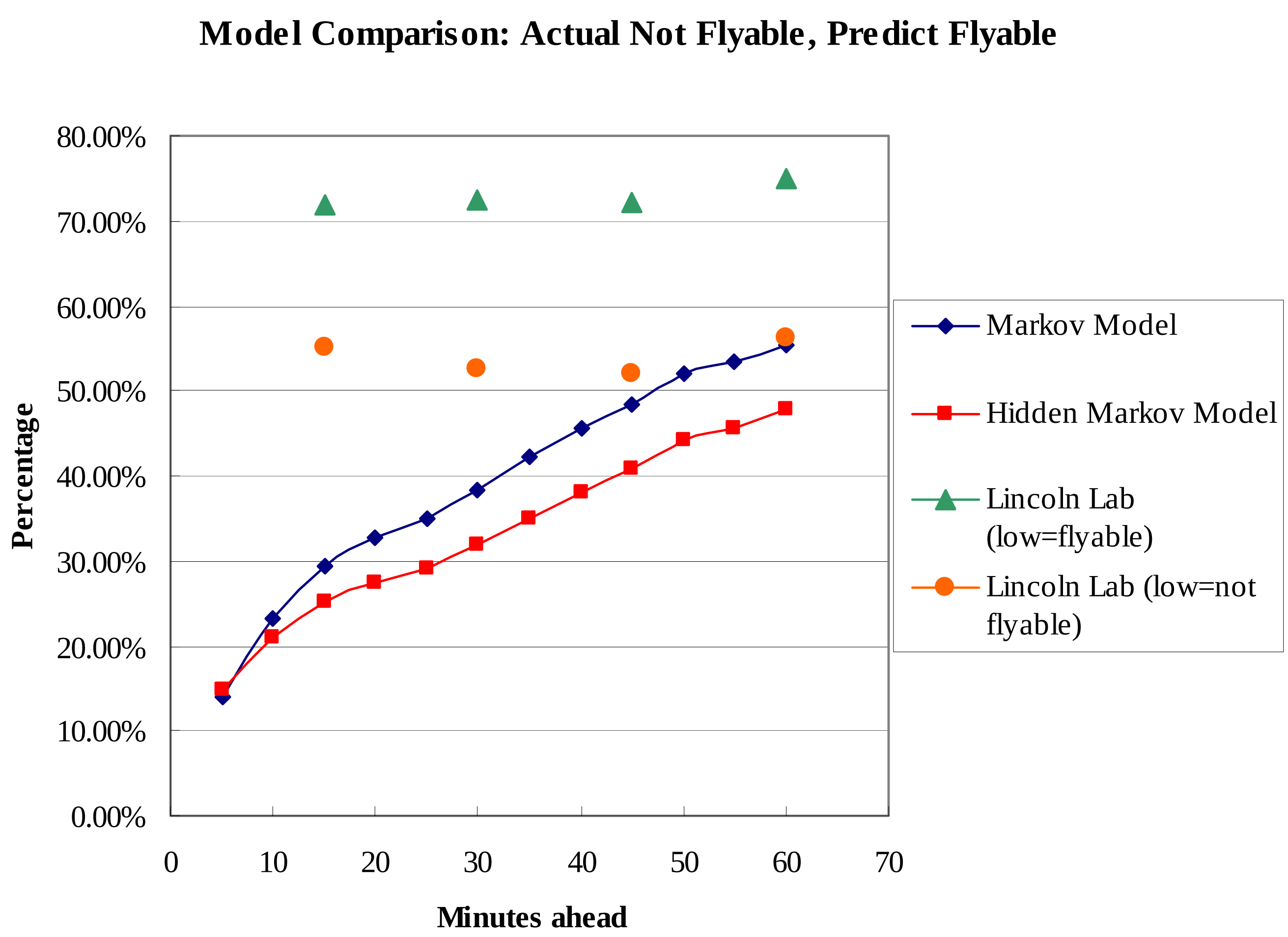
Performance of Markov Model



Model Comparison (1)



Model Comparison (2)



Conclusions

- Need for properly defined probabilistic convective weather forecasts
- Potential for Markovian models to provide such forecasts
- The states are defined in the context of airline application -- link-based states vs. cell-based states
- HMM shows promise in modeling convective weather in terms of performance and computation
- Further investigation using NHMM as modeling framework
- Future work to incorporate the convective weather forecasts in aircraft routing decision-making

Questions?

Appendix

Parameter Estimation for HMM

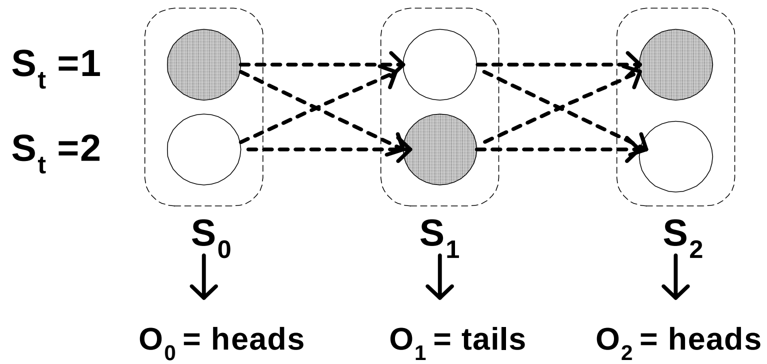
Probability of observed data

$$\begin{array}{ccccccccc} S_0 & \longrightarrow & S_1 & \longrightarrow & S_2 & \longrightarrow \cdots \longrightarrow & S_{n-1} & \longrightarrow & S_n \\ \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow \\ O_0 & & O_1 & & O_2 & & O_{n-1} & & O_n \end{array}$$

- Computing of the observed sequence involves summing over many possible hidden state sequences:

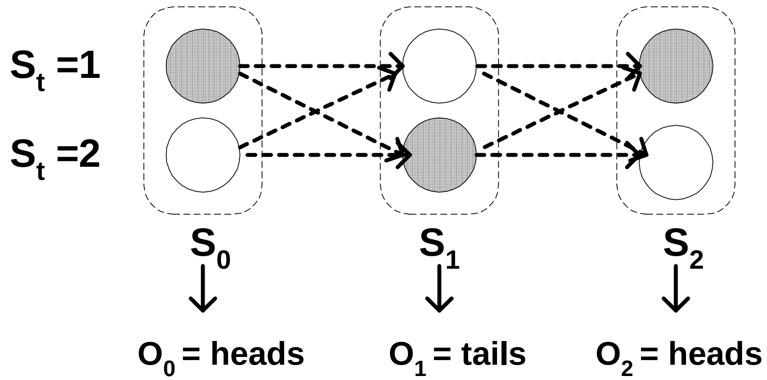
$$P(O_0, \ldots, O_n) = \sum_{S_0,\ldots,S_n} P_{init}(S_0) P_e(O_0 \mid S_0) \ldots P_{tr}(S_n \mid S_{n-1}) P_e(O_n \mid S_n)$$

Forward updates



- Forward probabilities $\alpha_t(i) = P(O_0, \ldots, O_t, S_t = i)$
  - $\alpha_0(1) = P_{init}(1)P_e(\text{heads} \mid 1)$
  - $\alpha_0(2) = P_{init}(2)P_e(\text{heads} \mid 2)$
  - $\alpha_1(1) = [\alpha_0(1)P_{tr}(1 \mid 1) + \alpha_0(2)\, P_{tr}(1 \mid 2)]\, P_e(\text{tails} \mid 1)$
  - $\alpha_1(2) = [\alpha_0(1)P_{tr}(2 \mid 1) + \alpha_0(2)\, P_{tr}(2 \mid 2)]\, P_e(\text{tails} \mid 2)$
- Generalized form:
  - $\alpha_0(i) = P_{init}(S_0 = i)P_e(O_0 \mid S_0 = i)$
  - $\alpha_t(i) = [\sum_j \alpha_{t-1}(j)P_{tr}(S_t = i \mid S_{t-1} = j)]\, P_e(O_t \mid S_t = i)$

Backward updates



- Backward probabilities $\beta_t(i) = P(O_{t+1}, \ldots, O_n \mid S_t = i)$
  - $\beta_2(1) = P_{end}(1)$
  - $\beta_2(2) = P_{end}(2)$
  - $\beta_1(1) = P_{tr}(1 \mid 1)P_e(\text{heads} \mid 1)\ \beta_2(1) + P_{tr}(2 \mid 1)P_e(\text{heads} \mid 2)\ \beta_2(2)$
  - $\beta_1(2) = P_{tr}(1 \mid 2)P_e(\text{heads} \mid 1)\ \beta_2(1) + P_{tr}(2 \mid 2)P_e(\text{heads} \mid 2)\ \beta_2(2)$
- Generalized form:
  - $\beta_n(i) = P_{end}(i)$
  - $\beta_{t-1}(i) = \sum_j P_{tr}(S_t = j \mid S_{t-1} = i)P_e(O_t \mid S_t = j)\ \beta_t(j)$

Forward-backward probabilities

$$\begin{array}{ccccccccccc} S_0 & \rightarrow \cdots \rightarrow & S_t & \rightarrow & S_{t+1} & \rightarrow & & \rightarrow & S_{n-1} & \rightarrow & S_n \\ \downarrow & & \downarrow & & \downarrow & & & & \downarrow & & \downarrow \\ O_0 & & O_t & & O_{t+1} & & & & O_{n-1} & & O_n \end{array}$$

Current estimate about S_t : $\alpha_t(i) = P(O_0, \ldots, O_t, S_t = i)$

Future evidence about S_t : $\beta_t(i) = P(O_{t+1}, \ldots, O_n \mid S_t = i)$

- The probability of generating the observations and going through state i at time t is:

  $P(O_0, \ldots, O_n, S_t = i) = \alpha_t(i)\ \beta_t(i)$
- $P(O_0, \ldots, O_n) = \sum_j \alpha_t(j)\ \beta_t(j)$ for $t = 0, 1, \ldots, n$
- The posterior probability that the HMM was in a particular state i at time t is:

  $P(S_t = i \mid O_0, \ldots, O_n) = [\alpha_t(i)\ \beta_t(i)] / [\sum_j \alpha_t(j)\ \beta_t(j)]$

  $= x_t(i)$

Forward-backward probabilities

Current estimate about S_t : $\alpha_t(i) = P(O_0, \ldots, O_t, S_t = i)$

Future evidence about S_{t+1} : $\beta_{t+1}(j) = P(O_{t+2}, \ldots, O_n \mid S_{t+1} = j)$

- The posterior probability that the HMM was in a particular state i at time t and transitioned to state j at time t+1 is:

  $P(S_t = i, S_{t+1} = j \mid O_0, \ldots, O_n)$

  $= [\alpha_t(i)\, P_{tr}(S_{t+1} = j \mid S_t = i) P_e(O_{t+1} \mid S_{t+1} = j)\, \beta_{t+1}(j)] \,/\, [\sum_j \alpha_t(j)\, \beta_t(j)\,]$

  $= y_t(i, j)$

EM algorithm for HMMs

- Assume there are M observation sequences:
  $O_0^{(m)}, \ldots O_{nm}^{(m)}$
- E-step: compute the posterior probabilities:
  - $x_t^{(m)}$ (i) for all m, i, and t (t = 0, …, n_m)
  - $y_t^{(m)}$ (i, j) for all m, i, j, and t (t = 0, …, n_{m-1})
- M-step:
  - Initial state probabilities
  - Transition probabilities
  - Emission probabilities

M-step: initial state probabilities

- Initial state probabilities
  = expected fraction of times the sequences
  started from a specific state i

$$\hat{P}_{init}(i) = \frac{1}{M}\sum_{m=1}^{M} X_0^{(m)}(i)$$

M-step: transition probabilities

- Transition probabilities:

$$\hat{P}_{tr}(j \mid i) = \frac{\#of\ \ transitions\ \ from\ i\ to\ \ j}{\#of\ \ visits\ to\ i} = \frac{\hat{N}(i, j)}{\sum_j \hat{N}(i, j)}$$

where the expected number of transitions from i to j :

$$\hat{N}(i, j) = \sum_{m=1}^{M} \sum_{t=0}^{n-1} y_t^{(m)}(i, j)$$

M-step: emission probabilities

- The emission probabilities:

$$\hat{P}_e(k \mid i) = \frac{\#of\ \ outputs\ k\ while\ in\ state\ i}{\#of\ \ visits\ to\ i} = \frac{\hat{N}_0(i,k)}{\sum_k \hat{N}_0(i,k)}$$

where the expected number of times a particular observation k was generated from a specific state i:

$$\hat{N}_0(i,k) = \sum_{m=1}^{M} \sum_{t=0}^{n_m} x_t^{(m)}(i) \delta\left(O_t^{(m)}, k\right)$$

where $\delta\left(O_t^{(m)}, k\right) = 1 \qquad if \quad O_t^{(m)} = k$

$= 0 \qquad$ otherwise

Different seed transition matrices

NumLocations:		611			
NumTimePeriods:		46			
Num of iterations:		10			
Elapsed time in milliseconds		2580			
Initial state probabilities					
0.55654	0.23009	0.21337			
End state probabilities					
0.02222	0.01828	0.0234			
Emission probabilities					
state\out	1	2	3	4	5
Low	0.16612	0.82345	0.01043	4.28E-11	1.61E-13
Med	2.45E-06	0.09386	0.87669	0.02945	3.48E-09
High	6.55E-17	9.73E-07	0.01533	0.66011	0.32456
Transition probabilities					
From\To	Low	Med	High		
Low	0.94807	0.02931	4.01E-04		
Med	0.08224	0.84123	0.05825		
High	4.68E-04	0.04093	0.9352		
Seed transition matrix					
0.4	0.35	0.2			
0.2	0.55	0.2			
0.2	0.35	0.4			

NumLocations:		611			
NumTimePeriods:		46			
Num of iterations:		10			
Elapsed time in milliseconds		2530			
Initial state probabilities					
0.57251	0.21542	0.21208			
End state probabilities					
0.02227	0.01814	0.0233			
Emission probabilities					
state\out	1	2	3	4	5
Low	0.16366	0.82105	0.01528	5.50E-11	4.65E-14
Med	1.00E-05	0.06882	0.89743	0.03374	9.10E-09
High	9.18E-18	2.02E-06	0.01307	0.66082	0.32611
Transition probabilities					
From\To	Low	Med	High		
Low	0.95002	0.02681	8.93E-04		
Med	0.07868	0.84403	0.05915		
High	0.00147	0.04039	0.93485		
Seed transition matrix					
0.5	0.3	0.19			
0.25	0.5	0.24			
0.19	0.3	0.5			

NumLocations:		611			
NumTimePeriods:		46			
Num of iterations:		10			
Elapsed time in milliseconds		2580			
Initial state probabilities					
0.17656	0.68426	0.13919			
End state probabilities					
0.02436	0.02142	0.02044			
Emission probabilities					
state\out	1	2	3	4	5
Low	0.25516	0.01706	0.03098	0.48381	0.21299
Med	0.00568	0.71909	0.26892	0.00631	6.9E-10
High	0.24117	0.0137	0.02188	0.47264	0.25062
Transition probabilities					
From\To	Low	Med	High		
Low	0.35187	0.05386	0.56991		
Med	0.01494	0.9541	0.00955		
High	0.52316	0.03105	0.42535		
Seed transition matrix					
0.2	0.35	0.4			
0.2	0.55	0.2			
0.4	0.35	0.2			

Video Integrator and Processor (VIP) Levels

VIP Level	Reflectivity (dBZ)	Precipitation Description
0	< 18	
1	[18, 30)	Light (Mist)
2	[30, 41)	Moderate
3	[41, 46)	Heavy
4	[46, 50)	Very Heavy
5	[50, 57)	Intense
6	> 57	Extreme

Some References

- Robust Dynamic Routing of Aircraft under Uncertainty, Arnab Nilim, Laurent El Ghaoui, Vu Duong
- A Hidden Markov Model for Space-Time Precipitation, Walter Zucchini, Peter Guttorp, Water Resources Research, Vol. 27, No. 8, 1991
- A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena, James P. Hughes, Peter Guttorp, Water Resources Research, Vol. 30, No. 5, 1994
- A non-homogeneous hidden Markov model for precipitation occurrence, James P. Hughes, Peter Guttorp, Stephen P. Charles, Applied Statistics, 48, Part 1, pp.15-30, 1999
- A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models, Jeff A. Bilmes, International Computer Science Institute, 1998