Diagnosing the relative impact of lagged forecast sequence structure using a simple dynamic decision model

Justin McLay

Global Prediction Section
Marine Meteorology Division
NRL-Monterey

DTC Workshop
03 November 2010

Preview of the talk

The objective: Incorporate the lagged forecast sequences into a Markov decision process

1) Why is this problem of interest?

2) **Markov-chain modeling** of lagged-forecast sequences

    1) Estimation of the chain
    2) Properties of the estimated Markov chain

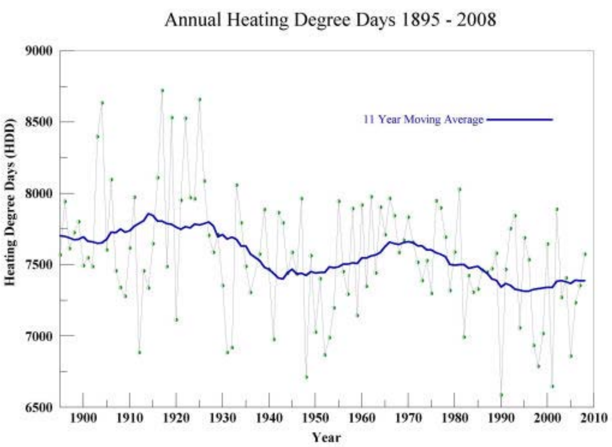
3) **Monte Carlo simulation** of the decision process: The lagged-forecast Markov chain in a dynamic decision model

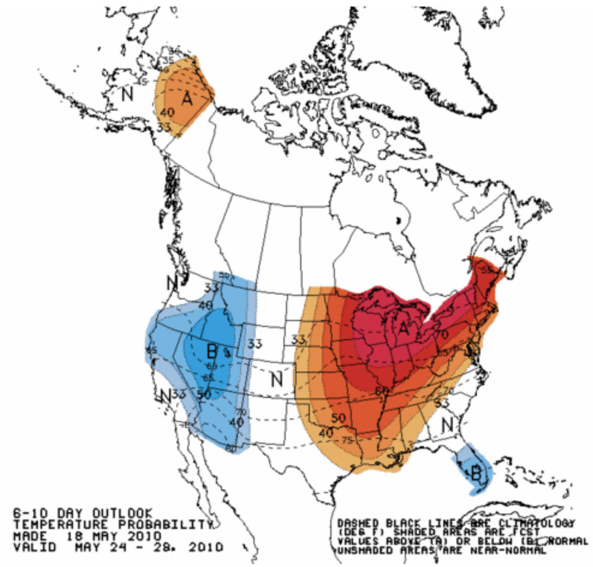
    1) Stochastic dynamic programming
    2) Calculation of conditional expected expense
    3) Results for a basket of cost functions

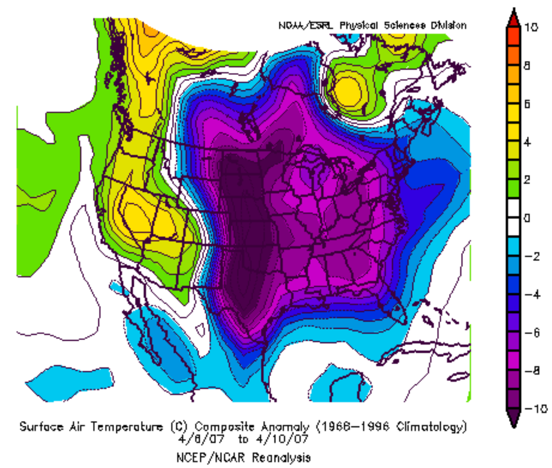
4) Future directions

Why put lagged forecasts into a decision process?

- **Aquila Energy - weather derivatives trading based upon Heating Degree Days (HDD)/Cooling Degree Days (CDD)**







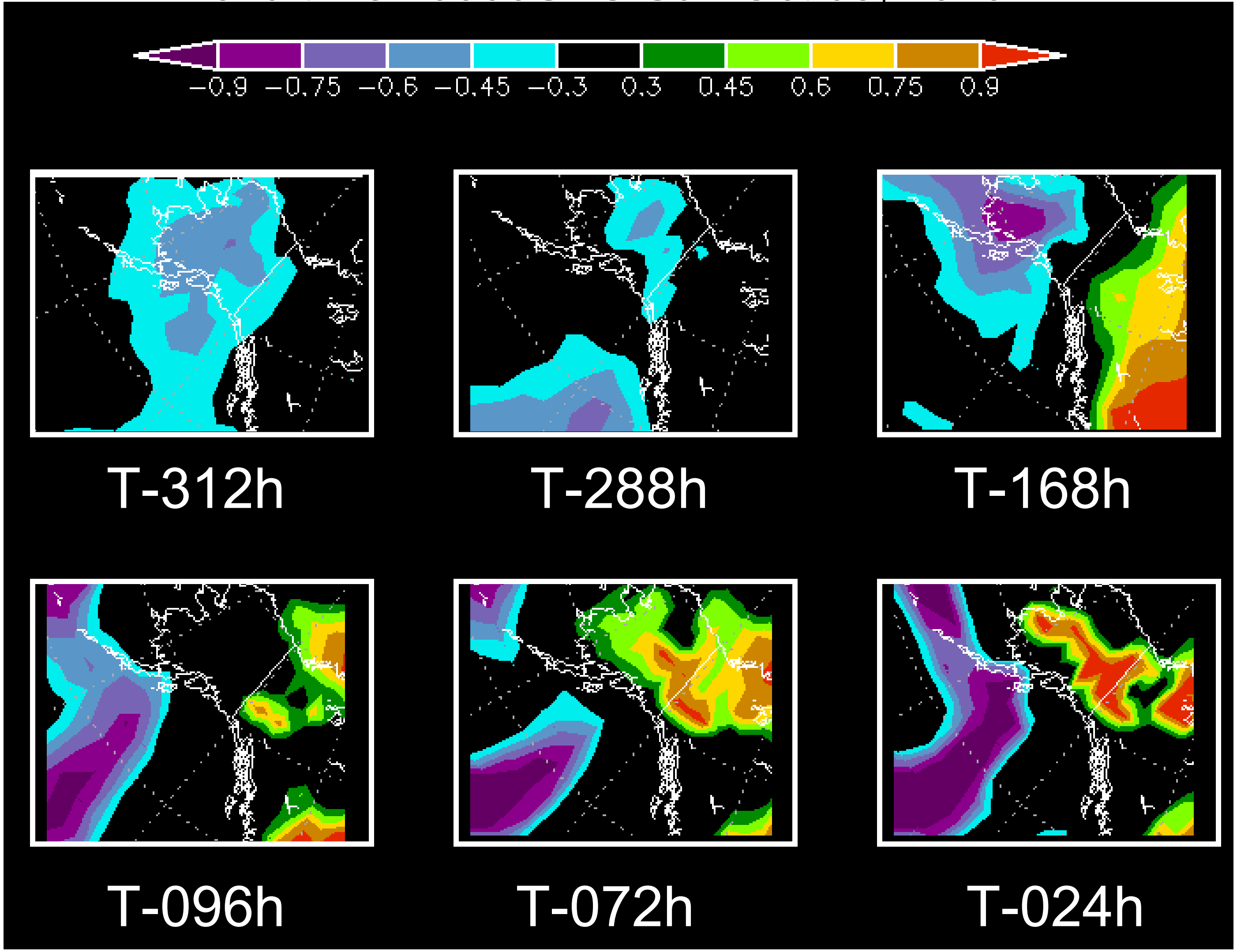
Traders claimed that lagged forecasts with “sneak” and “phantom” cold-air outbreaks and heat-waves were a notorious problem…

Sneaks: An event looks unlikely at longer lag (τ), but then abruptly becomes likely at shorter τ.

Phantoms: An event looks likely at longer τ, but then abruptly becomes unlikely at shorter τ.

Example of sneak in 850 hPa temperature anomaly probability forecast

Ens anom prob (1 sigma) – 850 hPa temp
Valid time: 0000UTC Sun Oct 03, 2010



Why put lagged forecasts into a decision process?

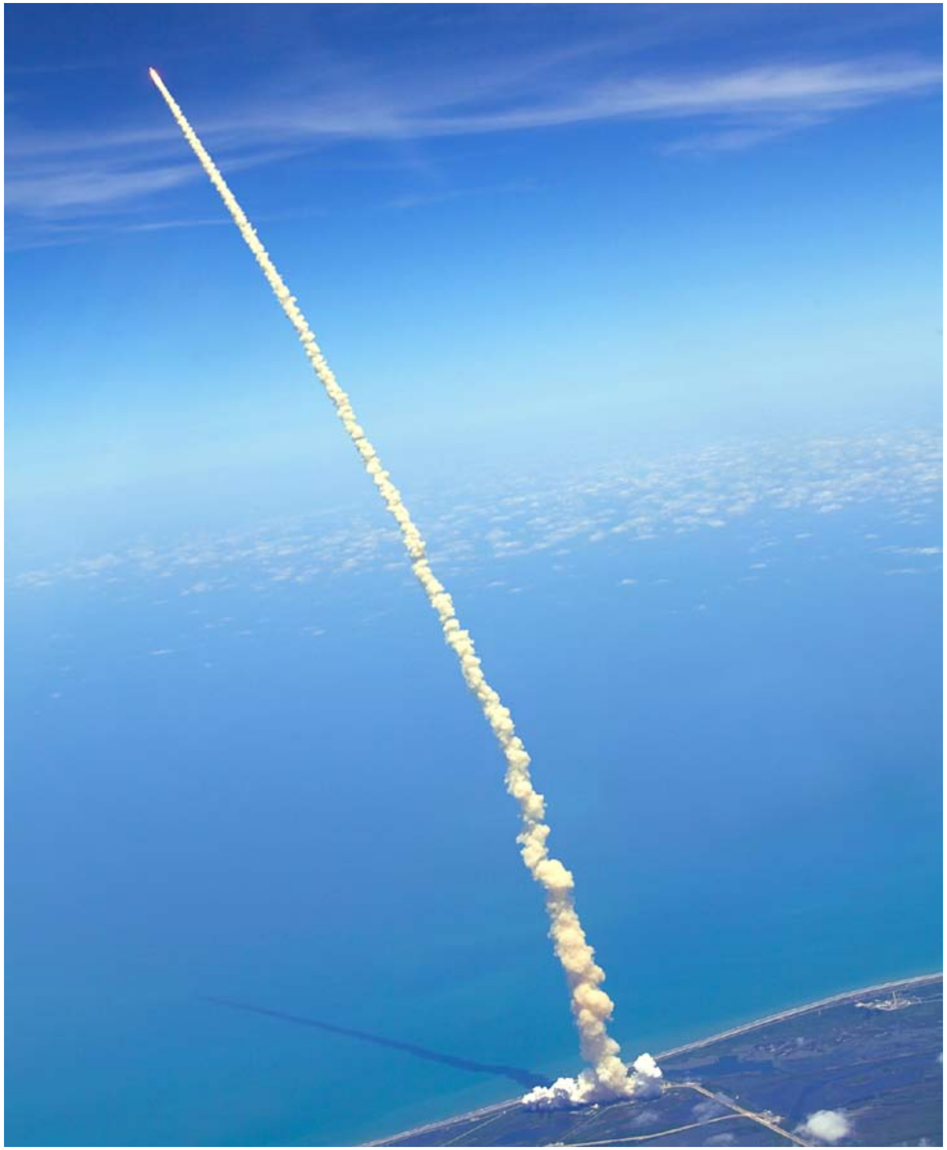
It's easy to find examples where lagged forecasts figure in decision problems…

e.g. :

03 November 2010

STS-133 Space shuttle Discovery launch

(2nd to last launch for the space shuttle)



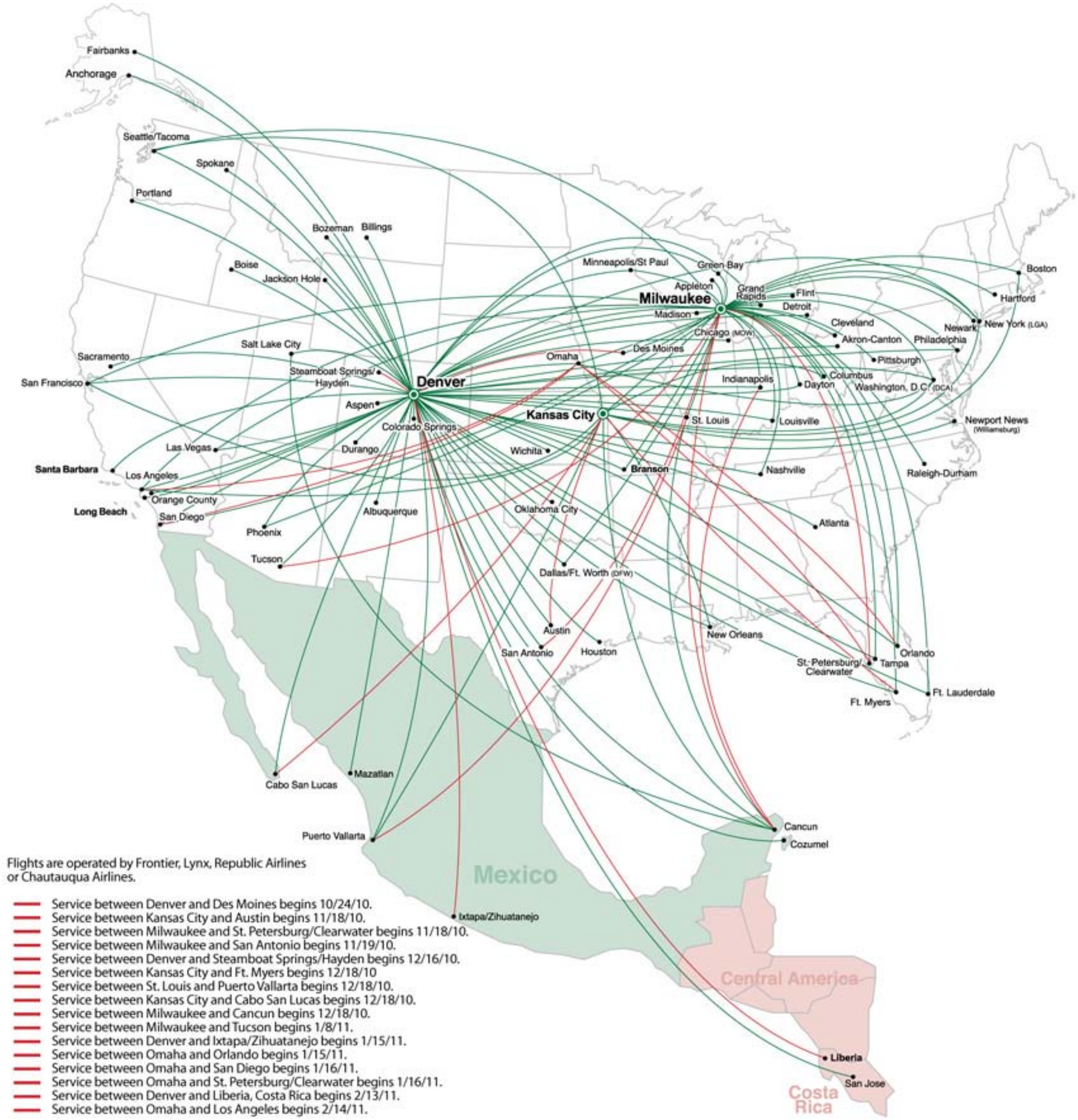
Why put lagged forecasts into a decision process?

It's easy to find examples where lagged forecasts figure in decision problems:

e. g.:

Airline hub operations on a holiday weekend

(United, Frontier hubs at DEN)



Why put lagged forecasts into a decision process?

- **The Navy (and the broader NWP community) wants to start measuring forecast ensemble performance by its impact on end users' decisions**

- **We need quantitative measures of the impact of volatility in lagged forecasts.**

Volatility in stochastic processes is an important, longstanding problem

e.g. Hurricane track prediction
"windshield wiper" effect
(Elsberry and Dobos 1990)

e.g. NWS forecast discussions

Excerpt from the San Francisco Bay Area NWS forecast discussion for 0430 UTC Tuesday 16 Dec. 2009:

Models have not come to a consensus among each other or from run to run and have been flip-flopping on their solutions for Friday with both the GFS and ECMWF showing a chance of light rain over the North Bay. Will maintain the dry forecast for now but may need to reevaluate overnight as new runs of the models come in.

e.g. (Taylor et al. 2005) GDP
(Karakatsani and Bunn 2008) electricity prices
(Kanas and Kouretas 2007) currency valuation

MARCH 1990 RUSSELL L. ELSBERRY

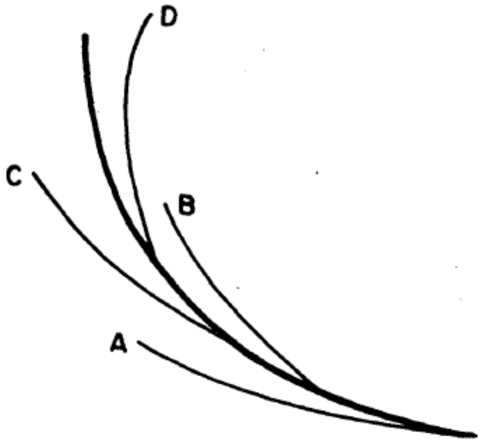


FIG. 1. Schematic of sequence of track forecasts (thin lines) labeled A through D that oscillate across the actual storm track (thick line) and illustrate the windshield wiper effect.

The Markov decision process and dynamic decision modeling

Elements of a dynamic decision model:

1) **A decision interval**

2) **A set of decision points**

3) **Meteorological states**

4) **A means to govern movement between meteorological states**

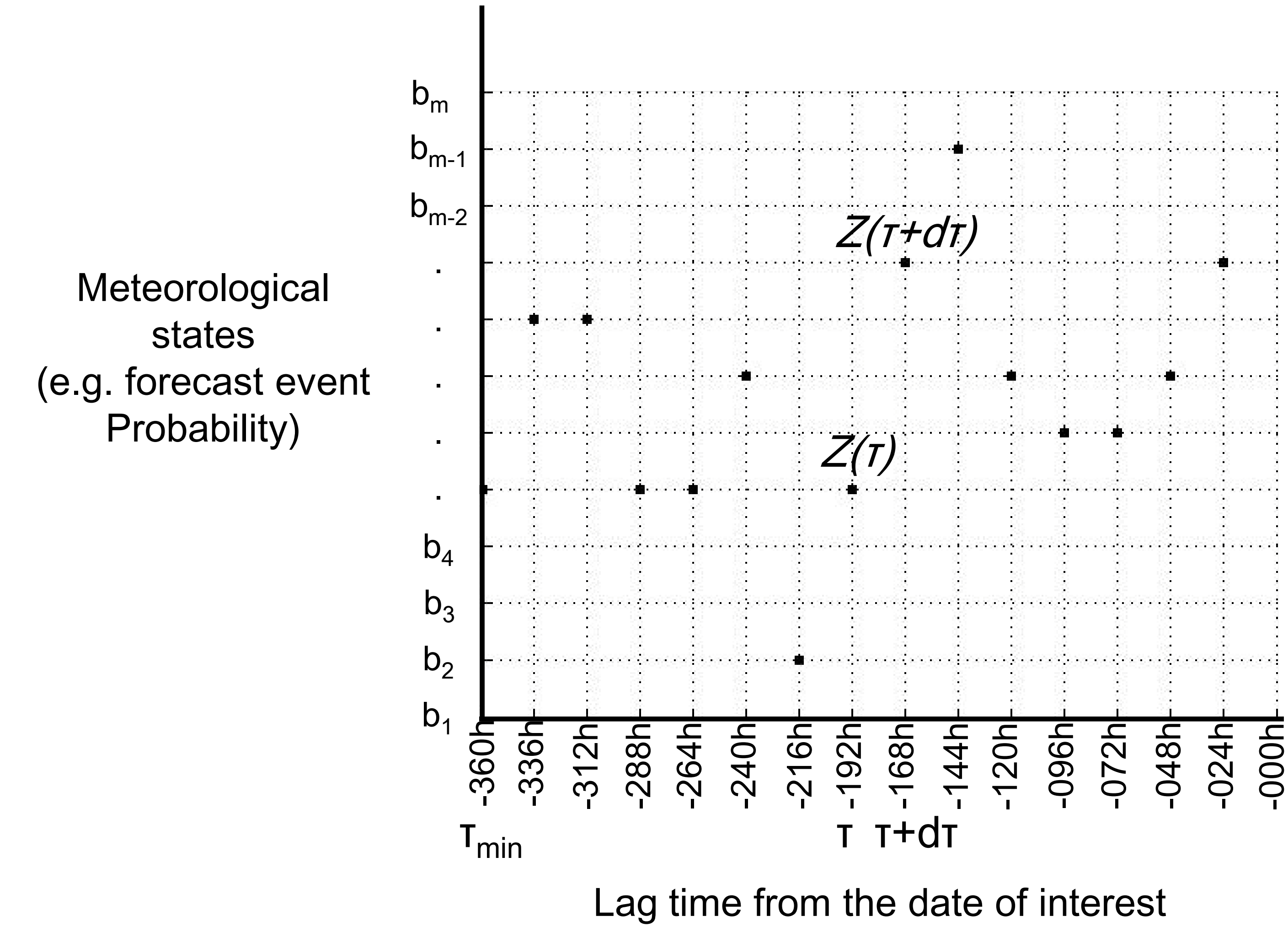
5) Non-meteorological states

6) Actions that govern movement between non-meteorological states

7) A function that describes the cost of each action

8) A loss function

The physical space of the problem:
Decision interval, points and meteorological states



The Markov chain perspective

In a Markov process **P[$Z(\tau+d\tau)=$ b$_i$]** is assumed conditional upon the r most recent realizations of Z, where r is the order of the Markov chain:

$$\mathbf{P}[Z(\tau+d\tau)=\mathbf{b}_i] = \mathbf{P}[Z(\tau+d\tau)=\mathbf{b}_i \,/\, Z(\tau)=\alpha_1,\ldots,Z(\tau-(r-1)d\tau)=\alpha_r]$$

For an order-1 chain:

$$\mathbf{P}[Z(\tau+d\tau)=\mathbf{b}_i] = \mathbf{P}[Z(\tau+d\tau)=\mathbf{b}_i \,/\, Z(\tau)=\alpha]$$

The chain is described by its "transition law" [π_o, $\mathbf{P}^{Nd\tau}(\tau)$]

$\mathbf{P}^{Nd\tau}(\tau)$: The (r+1)-dimensional N-step "transition probability matrix", with m+1 elements in each dimension.

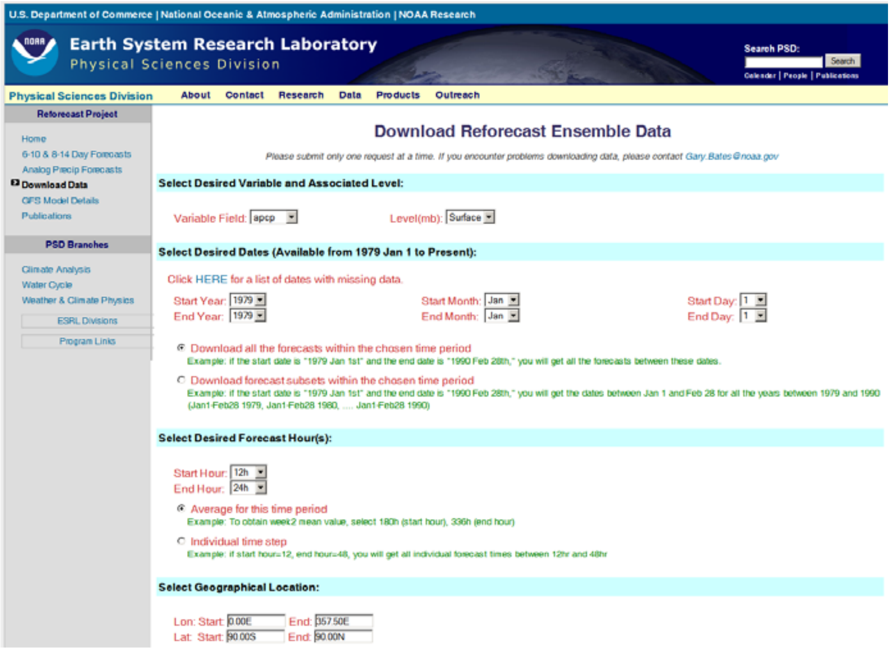
For an order-1 chain, the ij'th element of $\mathbf{P}^{Nd\tau}(\tau)$ is P[$Z(\tau+Nd\tau)=$b$_j$] | $Z(\tau)=$b$_i$, i,j∈{1,...,m}.

π_o: The "starting vector", such that π_o= P[$Z(\tau_{min})=$ b$_i$], i ∈{1,...,m}.

Maximum likelihood estimation of transition law $[\pi_o, P^{Nd\tau}(\tau)]$

Given an ensemble reforecast dataset (Hamill et al. 2006)

- 11-members
- initialized at 0000UTC
- T62L28
- 1998 version of NCEP GFS
- $2.5^o \times 2.5^o$
- 15 lead times (T_o+24h, T_o+48h, ..., T_o+360h)
- 01 January1979 to 31 May 2005



Provides 9650 realizations of lagged forecast sequences

Use maximum likelihood estimation (MLE) to obtain the transition law

For an order-1 chain:

$$\widehat{\mathbb{P}}^{dt}_{ij}(t) = K_{ij}/K_i$$

where K_{ij} is the number of transitions from state $Z(\tau) = b_i$ to state $Z(\tau+d\tau) = b_j$ observed in the dataset, and K_i is the total number of transitions from state $Z(\tau) = b_i$ observed in the dataset.

Transition law estimation – the assumptions

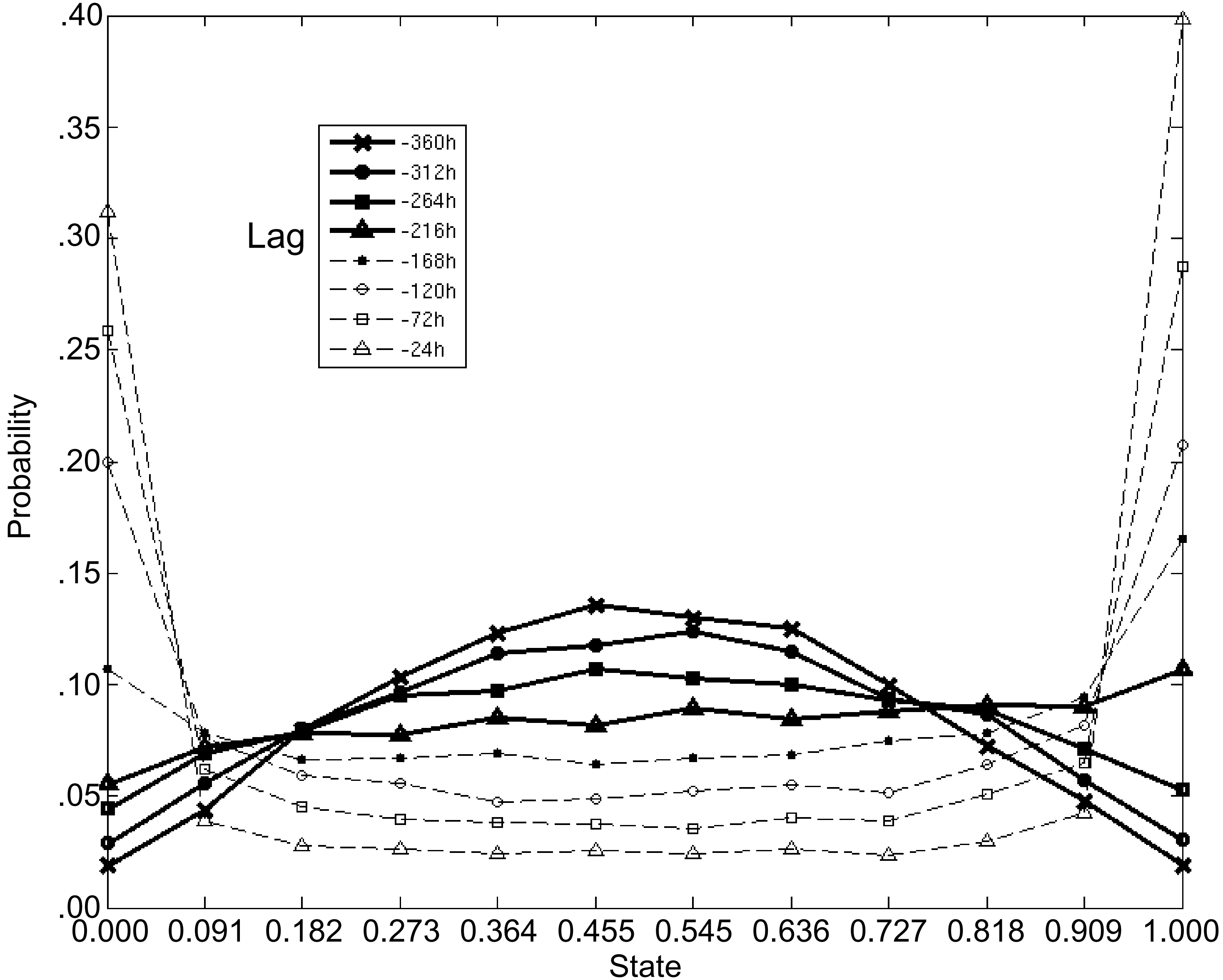
1) Neglect seasonality/regime dependence of transition law

2) Ignore the moderate positive correlation between the reforecast lagged sequences

3) Use raw ensemble event probabilities (no post-processing)

4) Choose to look at a binary event for 500 hPa geopotential height

Unconditional probability that *Z* will occupy a given state.



Some properties of the Markov process

1. Order

   ~ first-order (but could be simulated as second-order)

2. Homogeneity

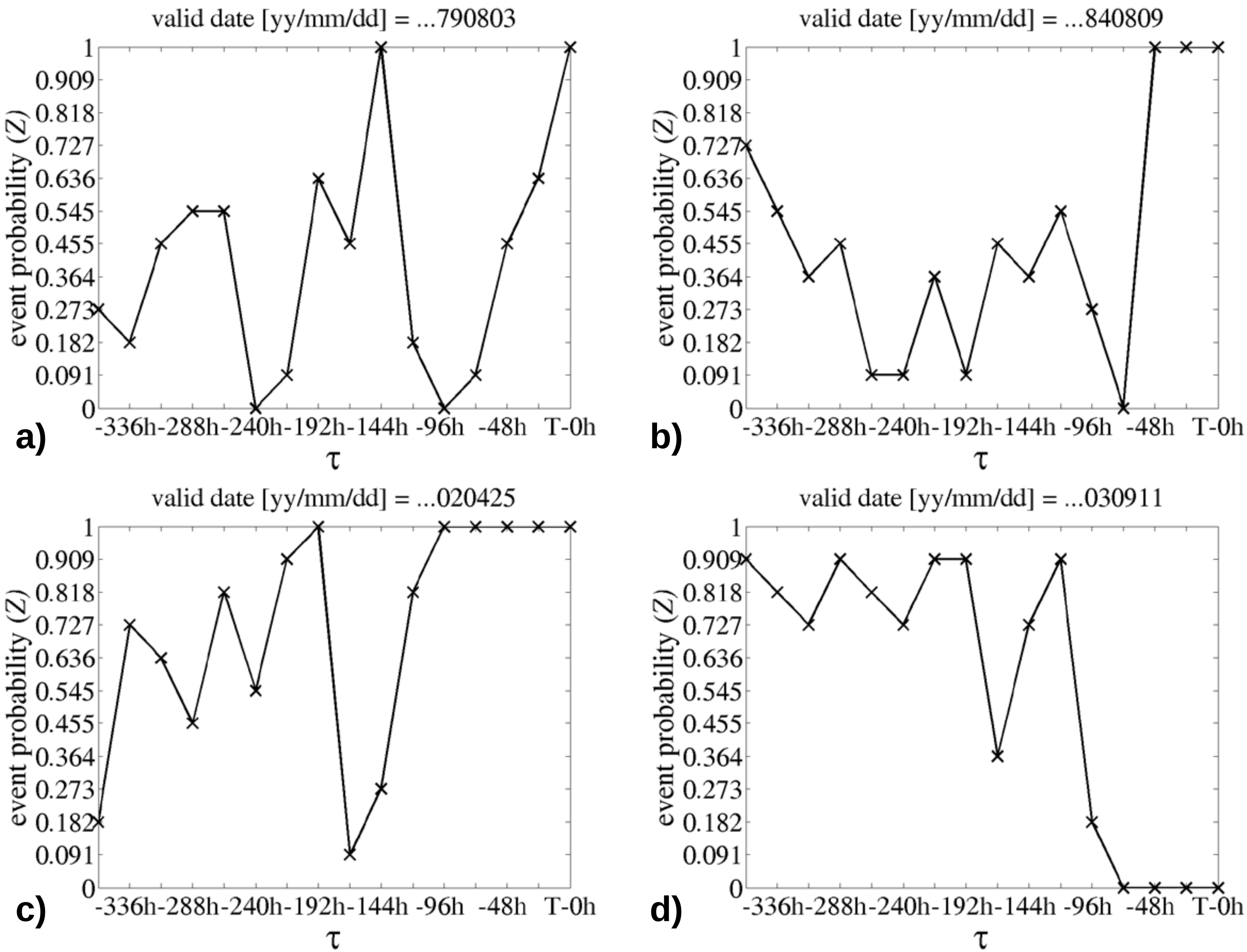
   Inhomogeneous with lag

3. Accessibility of states

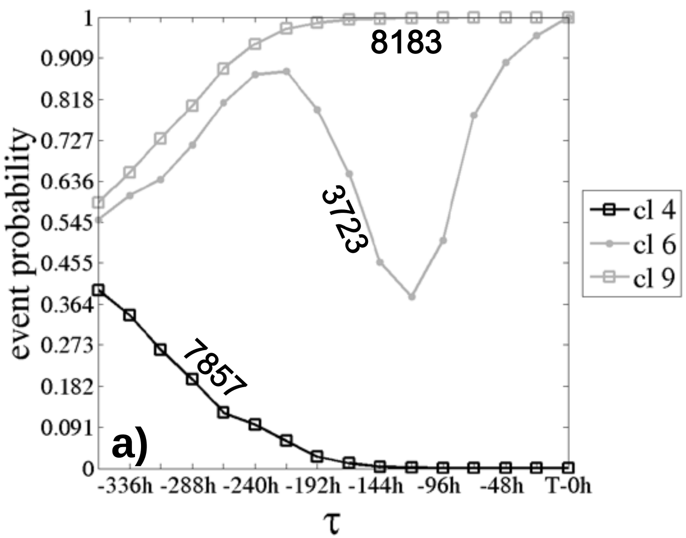
   $P_{ij}^{d\tau}(\tau) \neq 0$ and $P_{ji}^{d\tau}(\tau) \neq 0$ for all i, j, τ

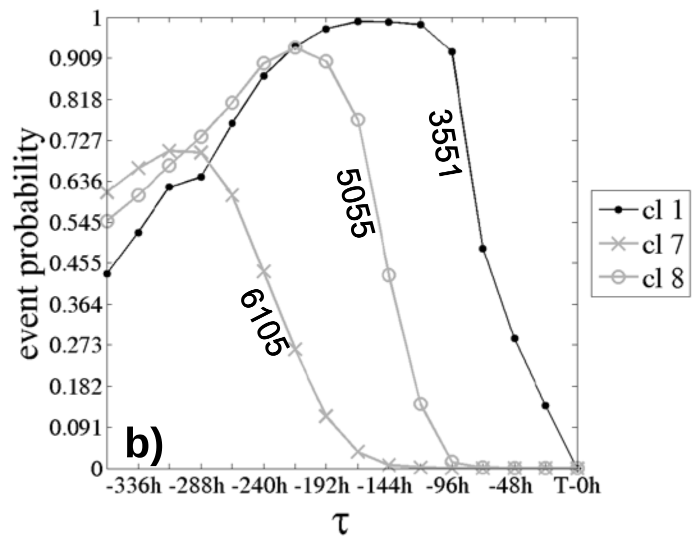
   - 12^{14} possible sequence realizations
   - enables volatility

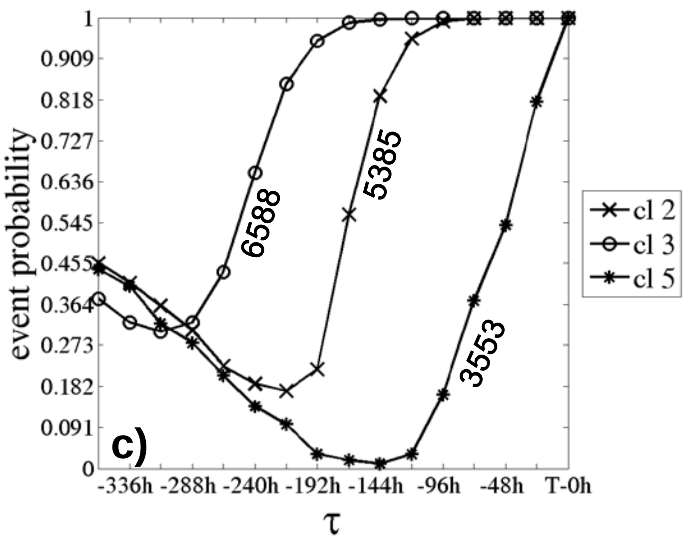
Four lagged probability forecast sequences from the reforecast dataset that illustrate the volatility allowed by the Markov process



Preferred transition paths: Objective classification of sequences with k-means clustering

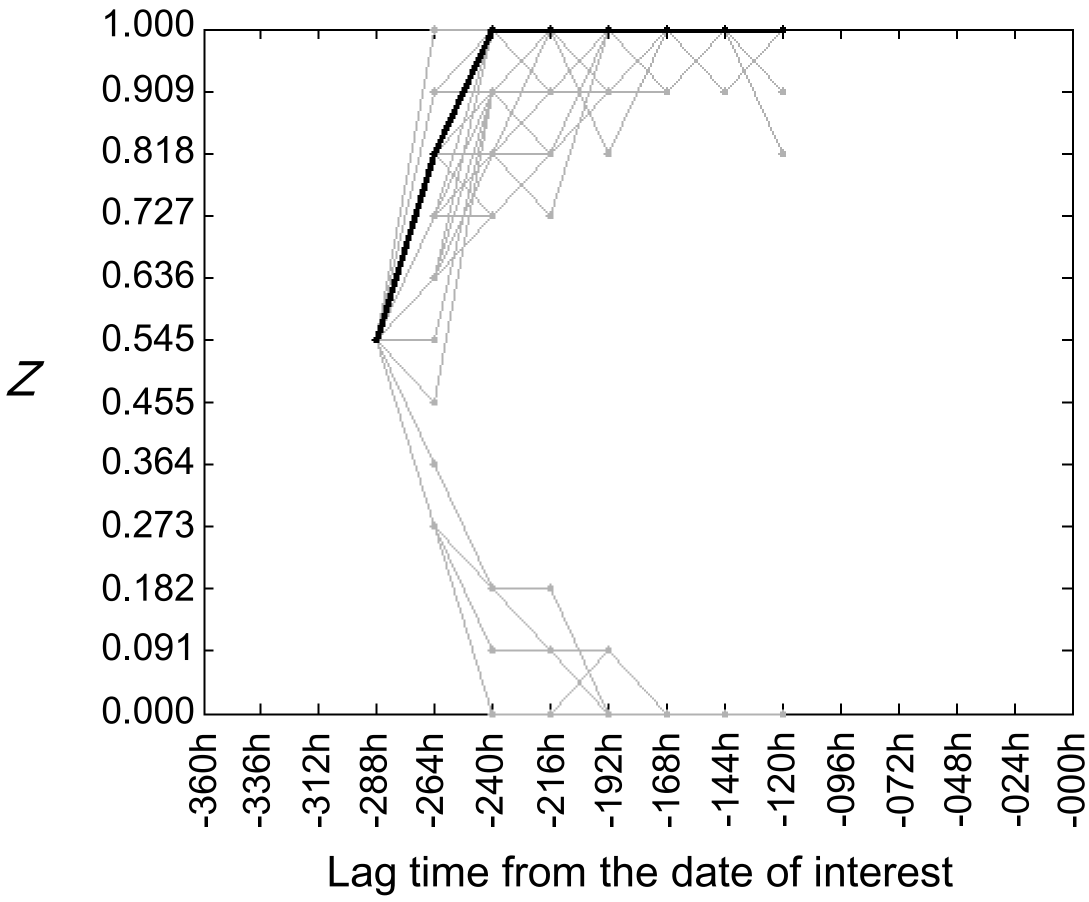






The within-cluster expected sequence for each of the nine optimal clusters. a) Expected sequences for clusters 4, 6, and 9. Clusters identified by the legend at right. The number of sequence realizations assigned to a given cluster is indicated in bold. Clusters identified by the legend on the right. b) As for a), but for clusters 1, 7, and 8. c) As for a) but for clusters 2, 3, and 5.

Preferred transition paths:
Conditional most likely path of the Markov process



The 100 most likely sequences for the interval [-288h,-120h] conditional upon *Z*(-288h) = 6/11.

The most likely sequence is indicated by the thick line.

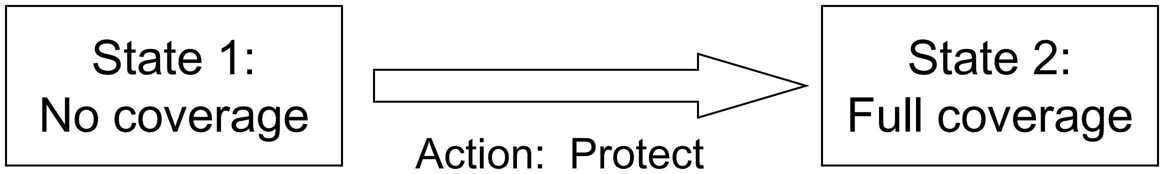
The Markov decision process and dynamic decision modeling

Elements of a dynamic decision model:

1) A decision interval
2) A set of decision points
3) Meteorological states
4) A means to govern movement between meteorological states
5) **Non-meteorological states**
6) **Actions that govern movement between non-meteorological states**
7) **A function that describes the cost of each action**
8) **A loss function**

The non-meteorological elements

A two-state, two-action decision (2x2) framework is common in the meteorological literature that deals with simple decision models (e.g. Murphy 1977, Katz and Murphy 1982, Murphy et al. 1985, Epstein and Murphy 1988, Katz 1993, Wilks 1991).



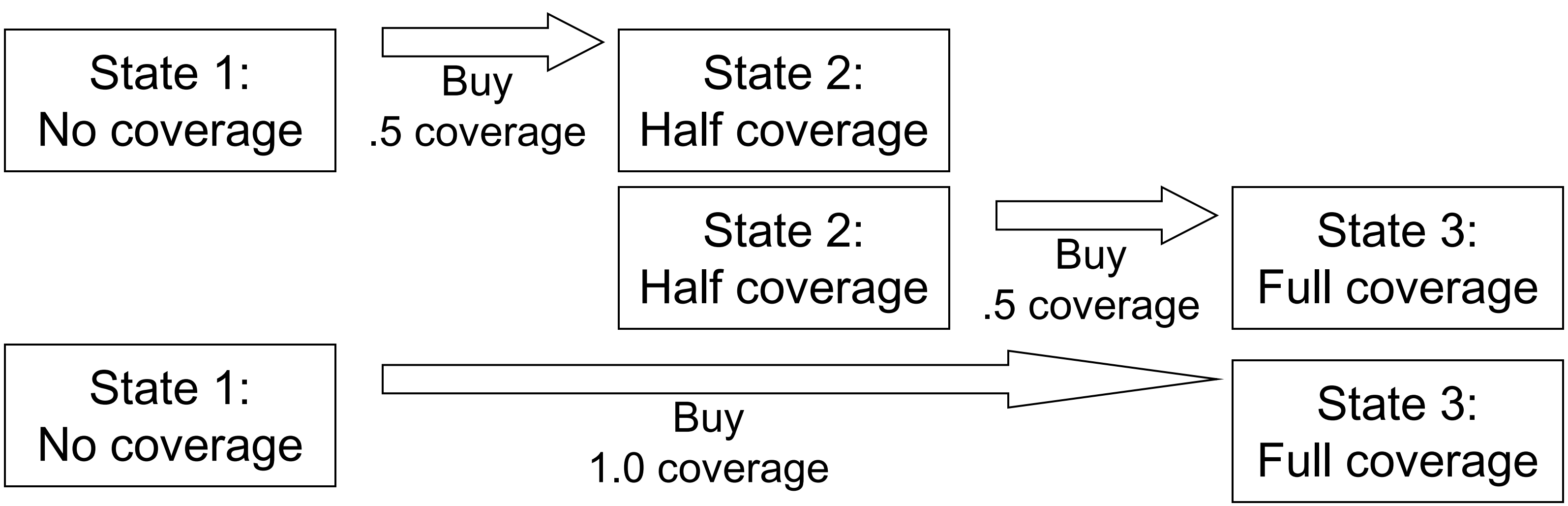
Here, a three-state, three-action (3x3) framework is employed.

Three non-meteorological states (λ=1,2,3):

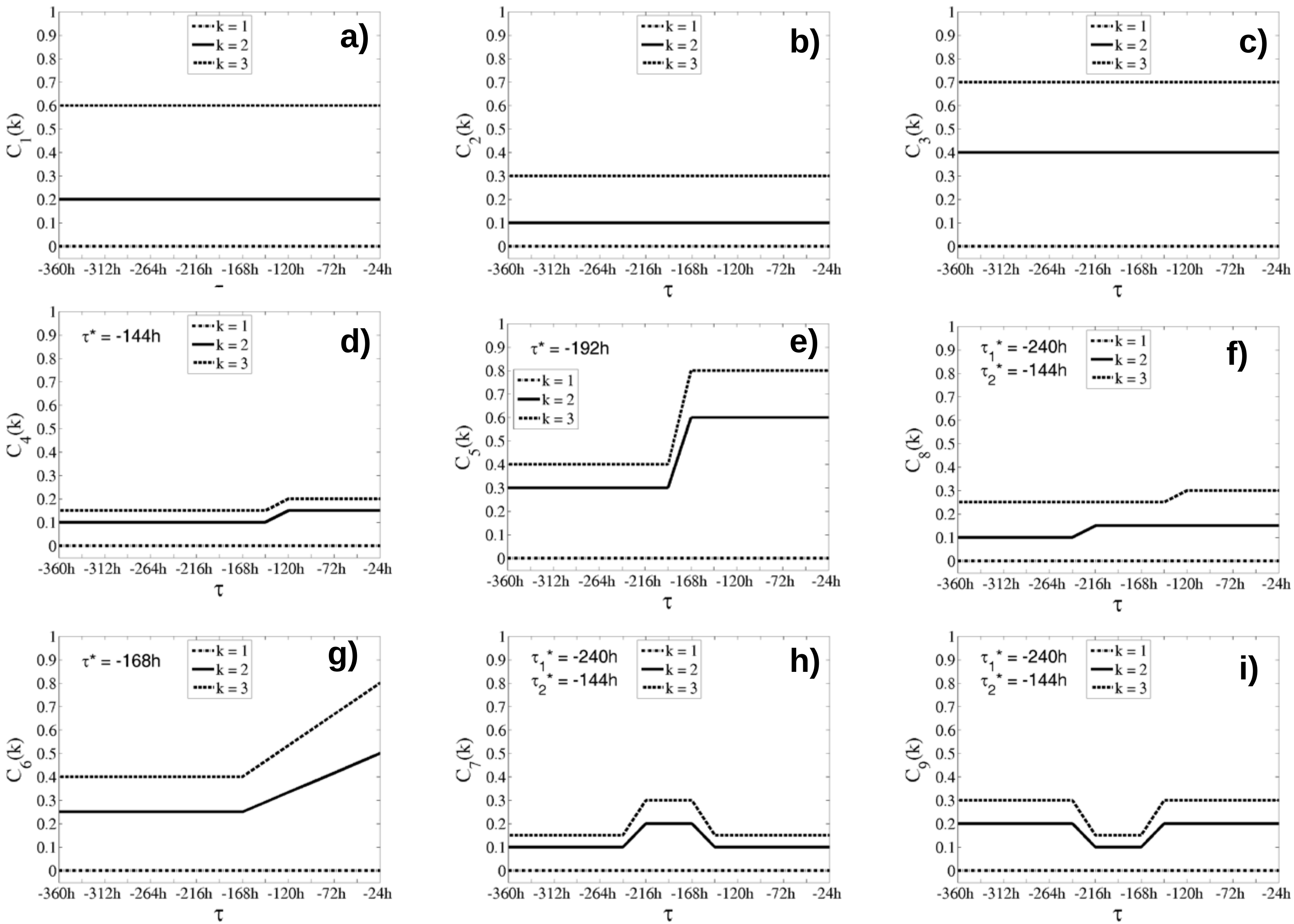
0.0 units coverage
0.5 units coverage
1.0 units coverage

Three actions A(k) (k=1,...,3):

Buy 0.0 units coverage
Buy 0.5 unit coverage
Buy 1.0 units coverage



Nine different simple cost functions for the 3x3 decision framework.



Stochastic dynamic programming

Bellman equation for stochastic dynamic programming with finite horizon:

$$\begin{aligned} V(b_i, \lambda, \tau - d\tau) &= \min_k \left[E_j \left[C(k) + V(b_j, G(\lambda, A(k)), \tau)\right]\right] \\ &= \min_k \left[\sum_{j=1}^{12} \mathbb{P}_{ij}^{d\tau}(\tau - d\tau)\left[C(k) + V(b_j, G(\lambda, A(k)), \tau)\right]\right], \end{aligned}$$

$V(b_i,\lambda,\tau\text{-}d\tau)$ represents the minimum expected expense over the remaining sub-interval $[\tau\text{-}d\tau, 0h]$ given the state (b_i,λ) of the MDP at $\tau\text{-}d\tau$

The solution of Eq. 3 yields optimal actions $A^*(b_i,\lambda,\tau)$ $\forall$ $i=1,\ldots,12$, $\lambda=1,\ldots,3$, $\tau=-360h,\ldots,-24h$.

When it comes time to apply the MDP to a given case, the optimal sequence of 15 actions over the decision interval is obtained by taking, at each successive decision point $\tau=-360h,\ldots,-24h$, the optimal action A^* that corresponds to the current state (b_i,λ) of the MDP.

Monte Carlo simulation
(The Markov chain + the dynamic decision model)

1) **Create a large number ($5x10^4$) of sequence realizations by sampling the transition law $[\pi_o, P^{Nd\tau}(\tau)]$**

2) **Apply the decision model to each of the sequence realizations**

   This incurs an expense in accordance with the combination of the optimal actions $A^*(b_i,\lambda,\tau)$, the event verification, and the loss function.

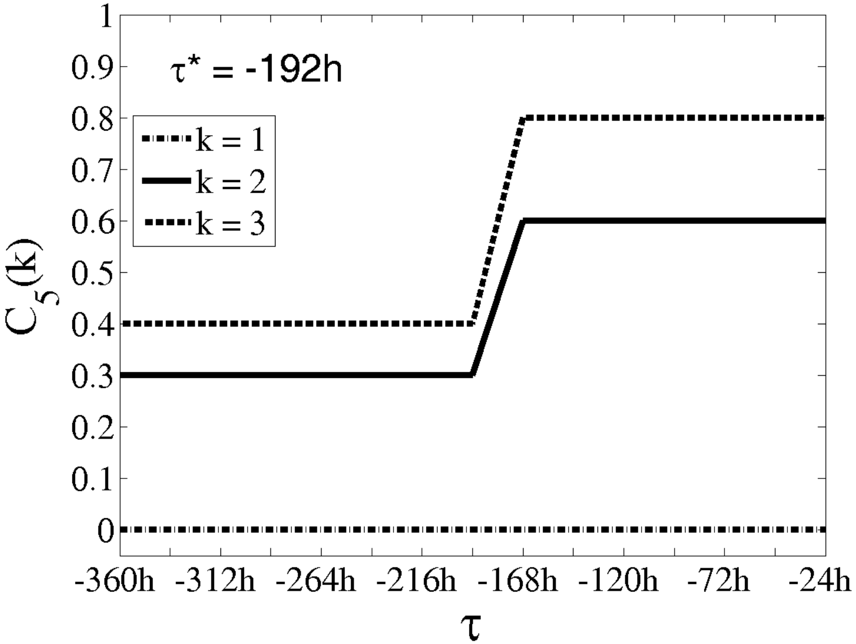
3) **Classify the sequence structure**

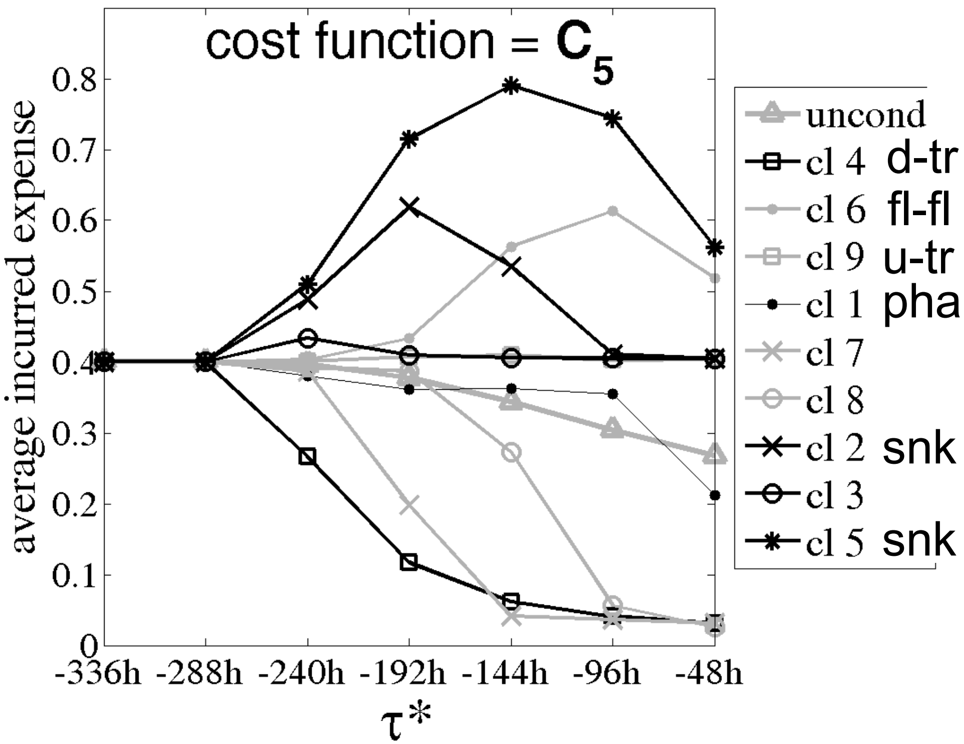
   a) cluster analysis
   b) volatility

- **Condition the decision model expected expense on each sequence structural class**

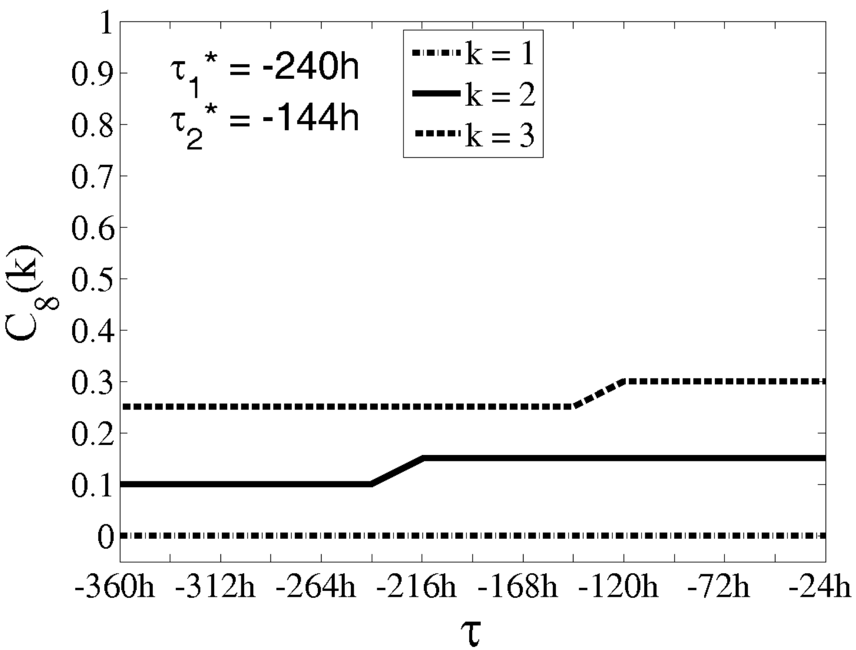
  Calculate the conditional average incurred expense for each structural class.

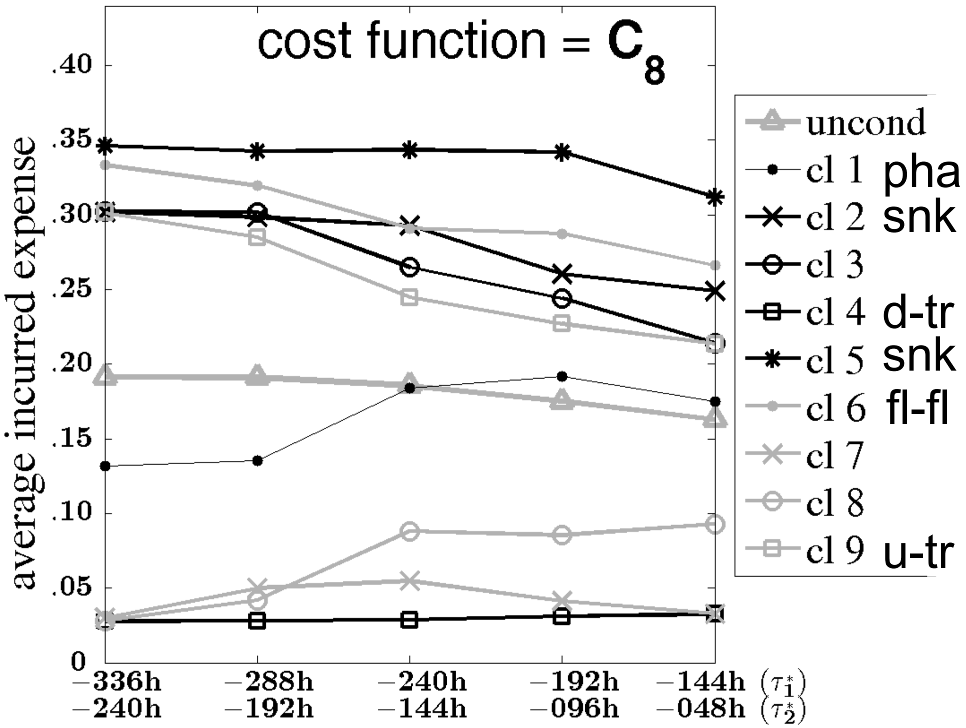
Structure classified by cluster analysis; Conditional average incurred expense w/ cost function C_5



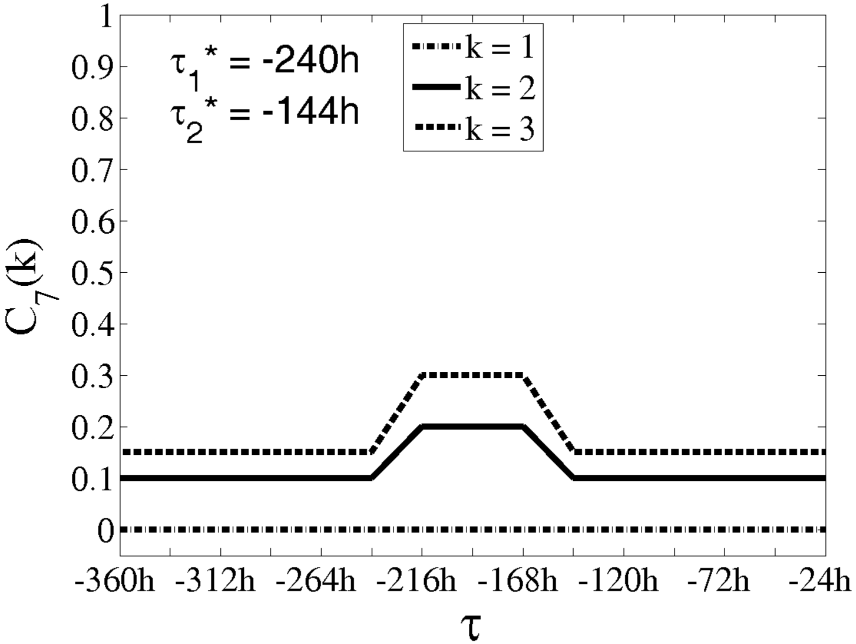


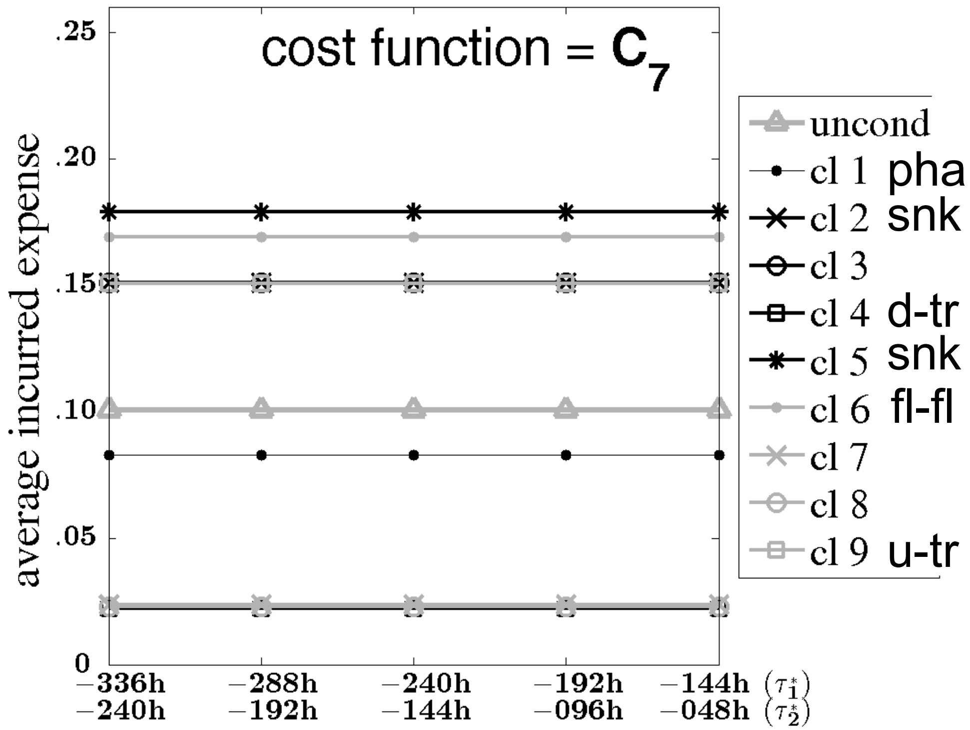
Structure classified by cluster analysis; Conditional average incurred expense w/ cost function C_8



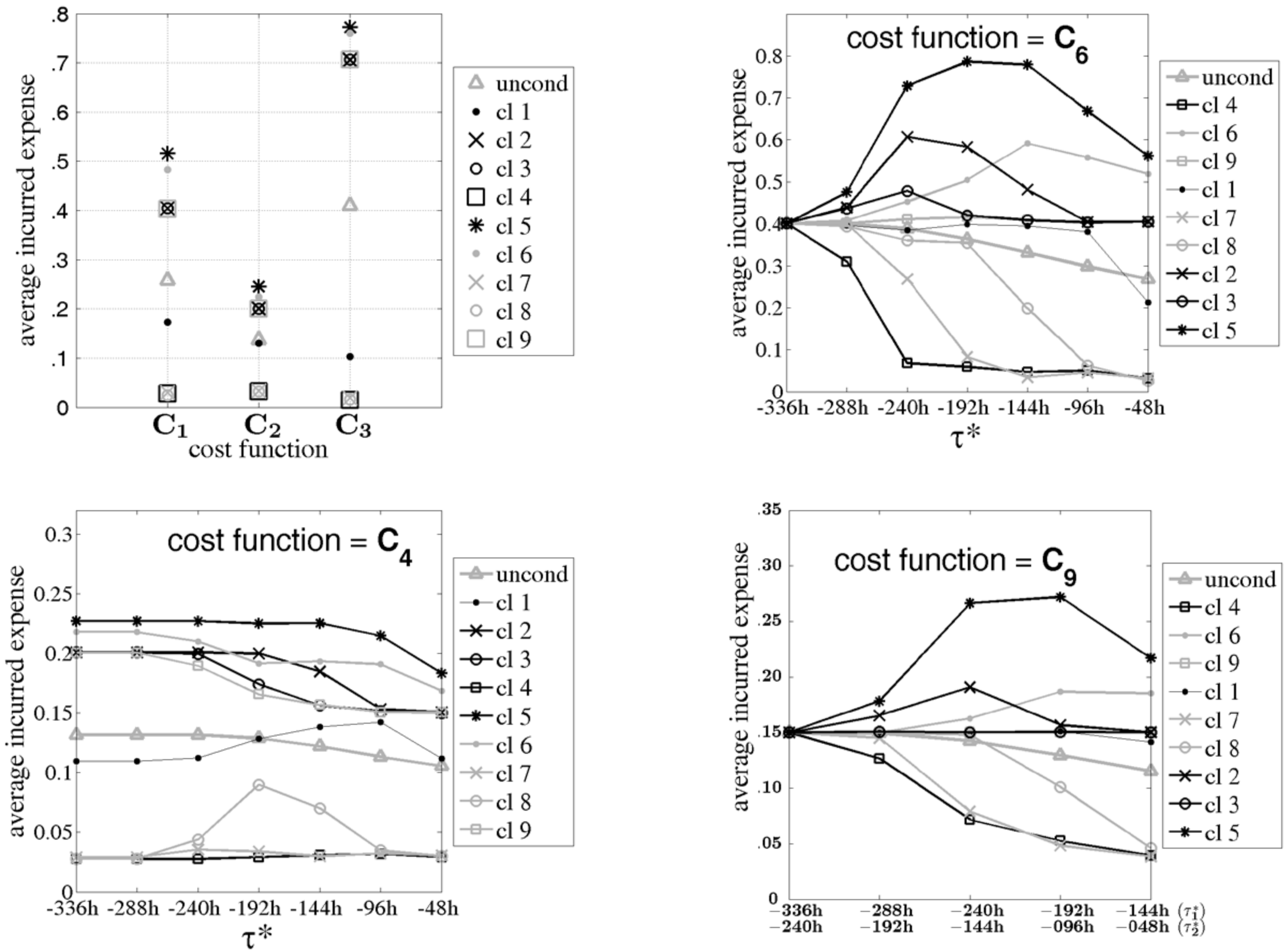


Structure classified by cluster analysis; Conditional average incurred expense w/ cost function C_7





d) Average incurred expense both unconditional and conditional upon cluster when using cost function C_6. The value of parameter τ^* is indicated by the abscissa. The cluster is indicated by the legend at right. e) As for d) but when using cost function C_7. f) As for d) but when using cost function C_8. The values of parameters τ_1^* and τ_2^* are indicated by the abscissa. g) As for f) but when using cost function C_9.



Classification of sequences by volatility

Based upon the maximum absolute change in the random variable Z from one decision point τ to the next decision point $\tau+d\tau$.

$$|\Delta Z|_{max}=\max[|Z(\tau+d\tau)-Z(\tau)|, \tau = -360h,\ldots,-24h]$$

The percentiles of $|\Delta Z|_{max}$ are calculated.

"volatile" sequences are those associated with percentiles 90 and higher

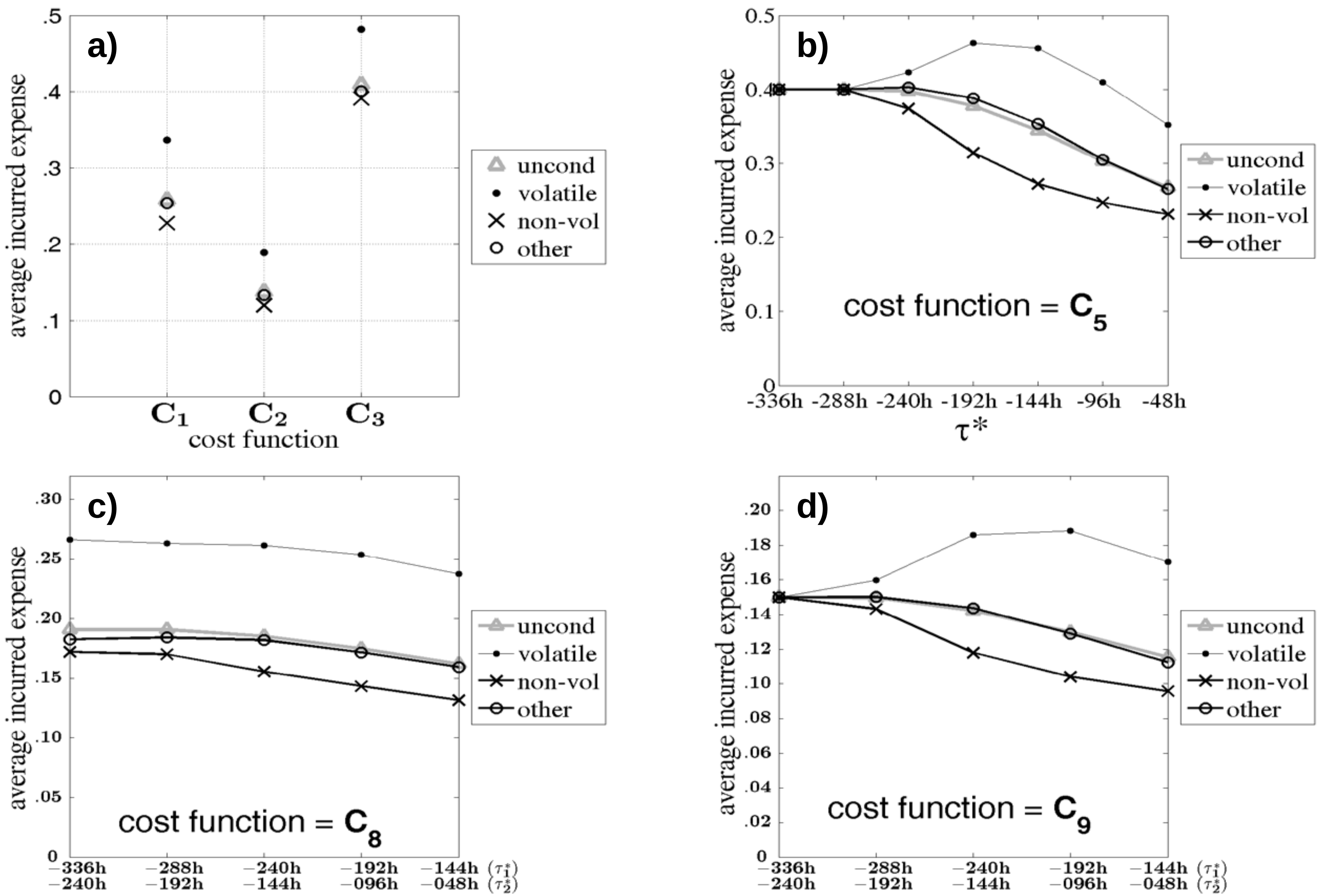
"nonvolatile" sequences are those associated with percentiles 50 and less

The remaining sequences are classified as "other" sequences.

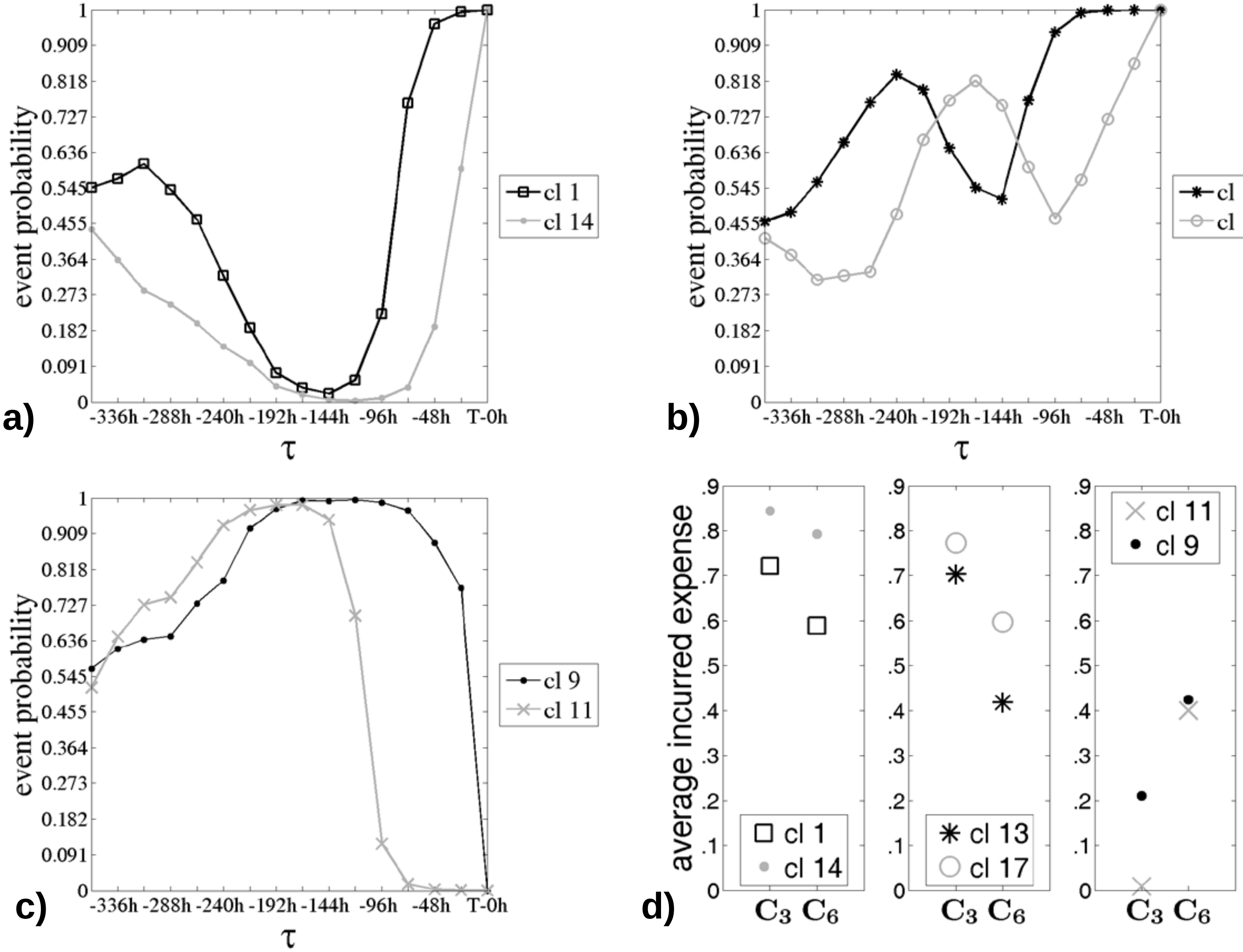
With these definitions, the non-volatile model's transition law doesn't allow transitions between states that are more than three states apart:

$$P[Z(\tau+d\tau)=b_j \,/\, Z(\tau)=b_i]=0 \text{ for all } i,j \text{ such that } |j-i|>4.$$

Structure classified by volatility; Conditional average incurred expense



Sensitivity of average incurred expense to advance warning



Options to reduce expense of sneaks, flip-flops, and volatility

1. Reduce the conditional expense of sneaks, flip-flops, and volatile sequences when they occur:

   a) Adaptive decision theory
   b) Higher-order Markov process
   c) Gaming strategies
   d) Other approaches…?

2. Reduce the frequency of occurrence of sneaks, flip-flops, and volatile sequences:

   a) Problem of data assimilation and ensemble design (more ensemble members, more cycle-to-cycle consistency of observation coverage & count,…?)

Summary

- It's fairly straightforward to used lagged forecasts in a Markov decision process.
- The main challenge is estimation of transition law. Defining the non-meteorological structure and cost function(s) may also be a challenge.
- The most problematic lagged forecast structures are those with rapid changes in event probability at short lag.
- The decision model cannot react quickly enough to some rapid changes (i.e. the extreme cases of "sneaks")
- The results are sensitive to problem specifics (e.g. cost function type, the parameters of a given cost function type)
- The expected expense is fairly sensitive to changes in advance warning. The Markov decision process can be used to quantify the benefit of improvements in advance warning.

Future Directions

- Kalman-filter estimation of transition law (π_o, $\mathbf{P}^{d\tau}(\tau)$)
- Analysis of regime-switching and discontinuous (jump) processes
- Scale-dependence of results
- Differences across single-model, multi-model, and post-processed (bias-corrected & calibrated) forecast ensembles
- Differences across deterministic versus probabilistic

References:

McLay, J. G., 2008: Markov chain modeling of sequences of lagged NWP ensemble probability forecasts: An exploration of model properties and decision support applications. *Mon. Wea. Rev.*, **136**, 3655-3670.

McLay, J. G., 2010: Diagnosing the relative impact of “sneaks”, “phantoms”, and volatility in sequences of lagged ensemble probability forecasts with a simple dynamic decision model. *Mon. Wea. Rev.*, in press, available through early online release at www.ametsoc.org.