HIDDEN AND NOT-SO-HIDDEN MARKOV MODELS:
IMPLICATIONS FOR ENVIRONMENTAL DATA ANALYSIS

Rick Katz

Institute for Study of Society and Environment
National Center for Atmospheric Research
Boulder, CO USA

Email: `rwk@ucar.edu`

Web site: `www.isse.ucar.edu/HP_rick.html`

Talk: `www.isse.ucar.edu/HP_rick/pdf/katzcb06.pdf`

QUOTES

- **Jean-Baptiste Lamarck (French biologist, ~ 1800)**

  "*Tout état de choses dans l'atmosphère . . . résulte non-seulement d'une réunion de causes qui tendent à l'opérer mais encore de l'influence de l'état de choses qui existoit auparavant.*"

  "*. . . l'efficacité de toute influence que l'atmosphère reçoit, est constamment en raison de l'état préexistant des choses dans cette atmosphère . . .*"

OUTLINE

(1) Background

(2) General Modeling Framework

(3) Not-So-Hidden Markov Models

(4) Model Extensions

(5) Potential Applications

(6) Role of Hidden Markov Models

(1) Background

- **Historical Perspective**

-- "Random variables defined on Markov chain"

(or "Chain-dependent process")

-- Not clear whether state of Markov chain "hidden" or not

- **Not-So-Hidden Markov Models**

-- Suggest when hidden Markov models might be beneficial

(2) General Modeling Framework

- Underlying Markov Chain

-- Make no assumption about whether or not states of Markov chain are observed

- Definition of HMM (or not-so HMM)

-- $\{J_t : t = 1, 2, \ldots\}$ first-order Markov chain (finite state space)

-- Conditional distribution of X_t given J_t

-- X_t 's conditionally independent given J_t 's

- **Probabilistic Properties**

-- **Induced temporal dependence**

In general: $\mathrm{Cov}(X, Y) = \mathrm{E}[\mathrm{Cov}(X, Y \mid Z)] + \mathrm{Cov}[\mathrm{E}(X \mid Z), \mathrm{E}(Y \mid Z)]$

HMM: $\mathrm{Cov}(X_t, X_{t+k}) = \mathrm{Cov}[\mathrm{E}(X_t \mid J_t), \mathrm{E}(X_{t+k} \mid J_{t+k})], \; k = 1, 2, \ldots$

-- **Stationary stochastic process with weak dependence**

Central Limit Theorem

Extreme Value Theorem

(3) Not-So-Hidden Markov Models

- **Underlying Markov Chain**

-- Assume states are observed

- **Example: Chain-dependent process for precipitation**

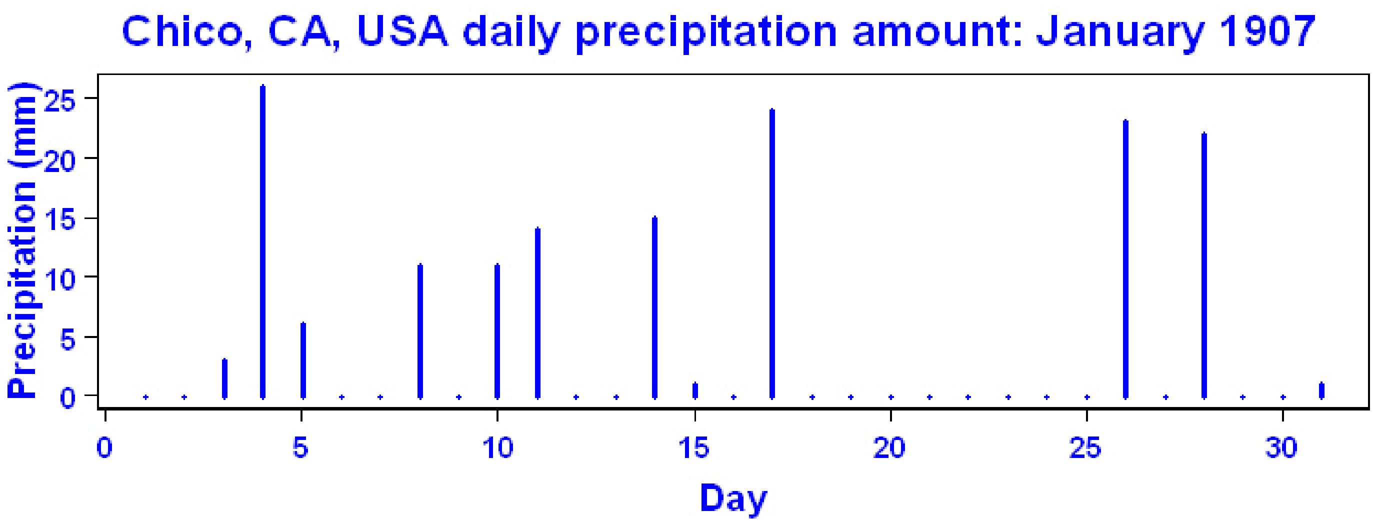
-- J_t **daily precipitation occurrence**

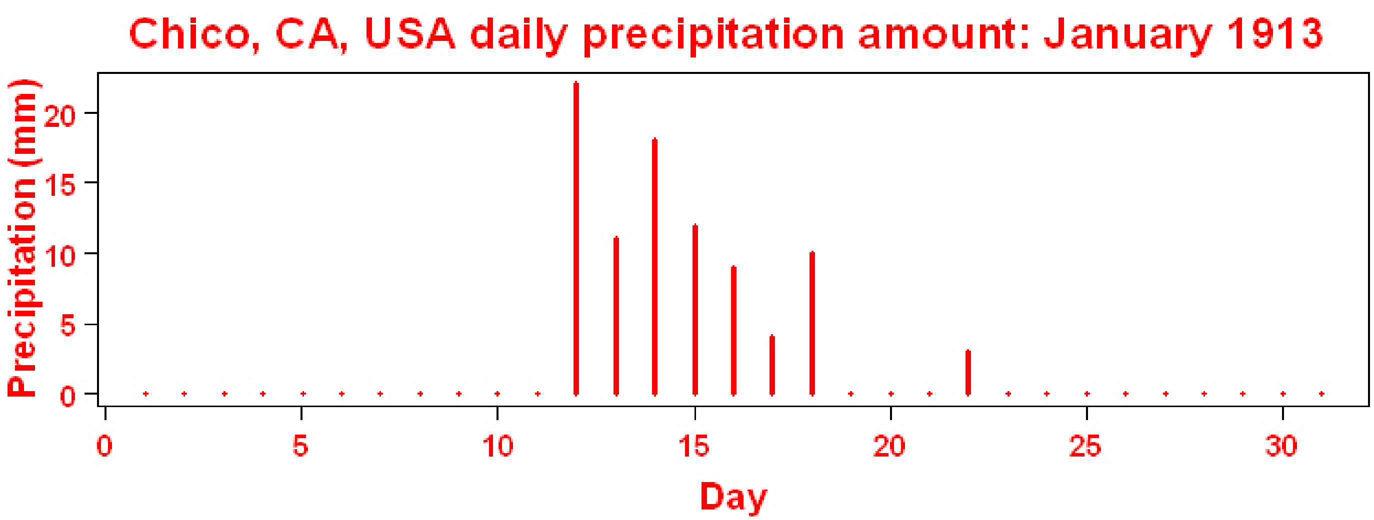
Two-state, first-order Markov chain, $\mathrm{Corr}(J_t, J_{t+1}) > 0$

$J_t = 1$ **denotes occurrence (**$J_t = 0$ **otherwise)**

-- X_t **daily precipitation amount**

$X_t = 0$ **if** $J_t = 0$ **(**$X_t > 0$ **otherwise, positively skewed distribution)**


Chico, CA, USA daily precipitation amount: January 1907
Precipitation (mm)
0
5
10
15
20
25
0
5
10
15
20
25
30
Day


Chico, CA, USA daily precipitation amount: January 1913
Precipitation (mm)
0
5
10
15
20
0
5
10
15
20
25
30
Day

- **Stochastic Weather Generator**

-- J_t daily precipitation occurrence (still first-order Markov chain)

-- X_t daily temperature variable (e.g., maximum or minimum)

-- Conditional distribution of X_t given $J_t = i$ assumed

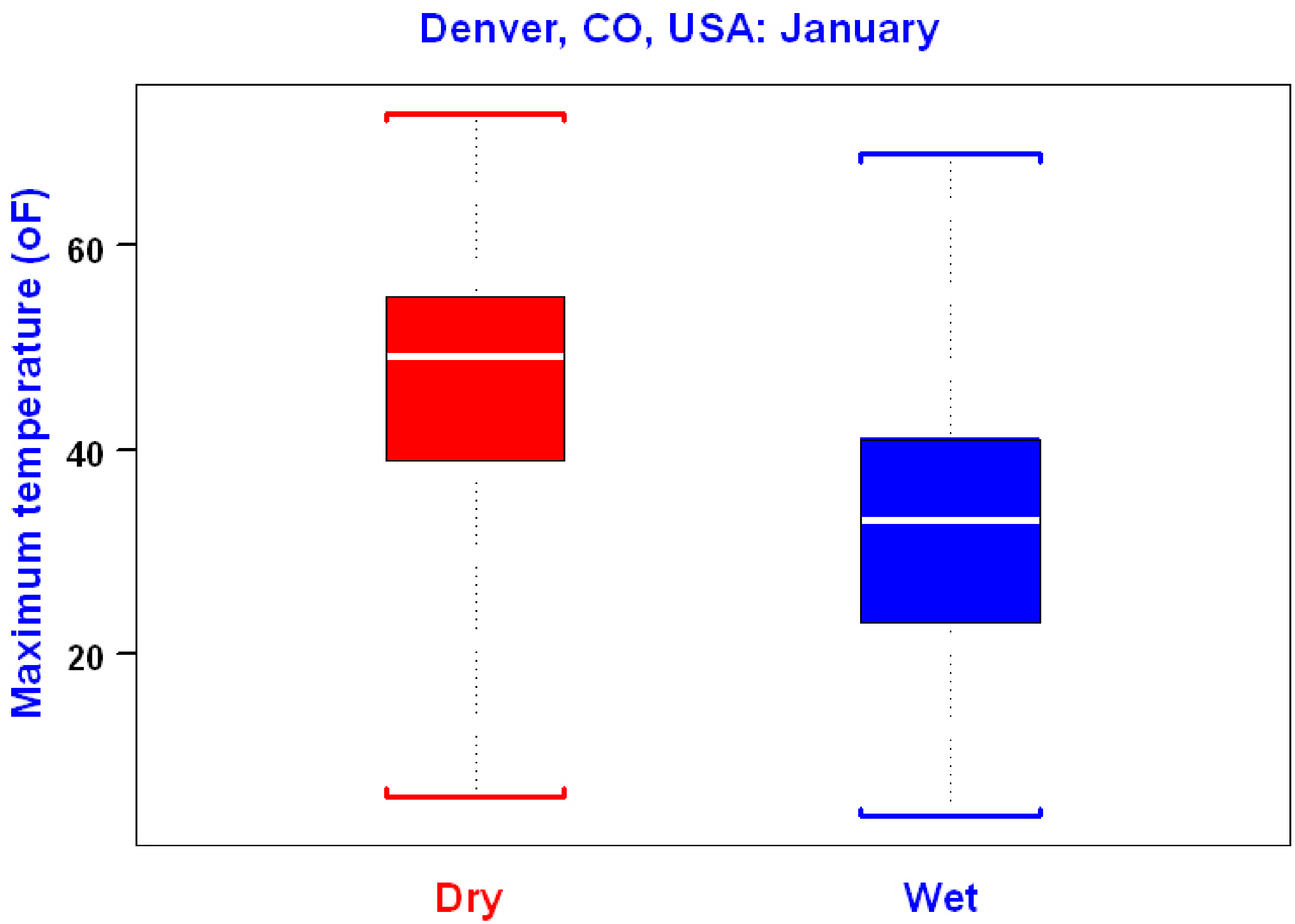
$$N(\mu_i, \sigma_i^2), \quad \mu_i = E(X_t \mid J_t = i), \quad \sigma_i^2 = \mathrm{Var}(X_t \mid J_t = i), \quad i = 1, 2$$

-- Temporal dependence of daily temperature X_t

Induced dependence (*via precipitation occurrence*) not strong enough

HMM: $\mathrm{Corr}(X_t, X_{t+1}) \leq \mathrm{Corr}(J_t, J_{t+1})$

Observe: $\mathrm{Corr}(X_t, X_{t+1}) > \mathrm{Corr}(J_t, J_{t+1})$


Denver, CO, USA: January
Maximum temperature (oF)
60
40
20
Dry
Wet

(4) Model Extensions

- **Other Types of Mixtures Involving Markov Chains**

-- Direct modeling of temporal dependence

- **Example: Stochastic Weather Generators (revisited)**

-- Random standardization

Given $J_t = i$, define $Z_t = (X_t - \mu_i) / \sigma_i$

Assume $\{Z_t, t = 1, 2, . . .\}$ is AR(1) process

$\text{Corr}(X_t, X_{t+k})$ sum of three exponentials

- **Markov Chain Conditional on Observed Covariate**

-- **Transition probabilities**

$$P_{ij} = \Pr\{J_{t+1} = j \mid J_t = i\}, \quad i, j = 0, 1$$

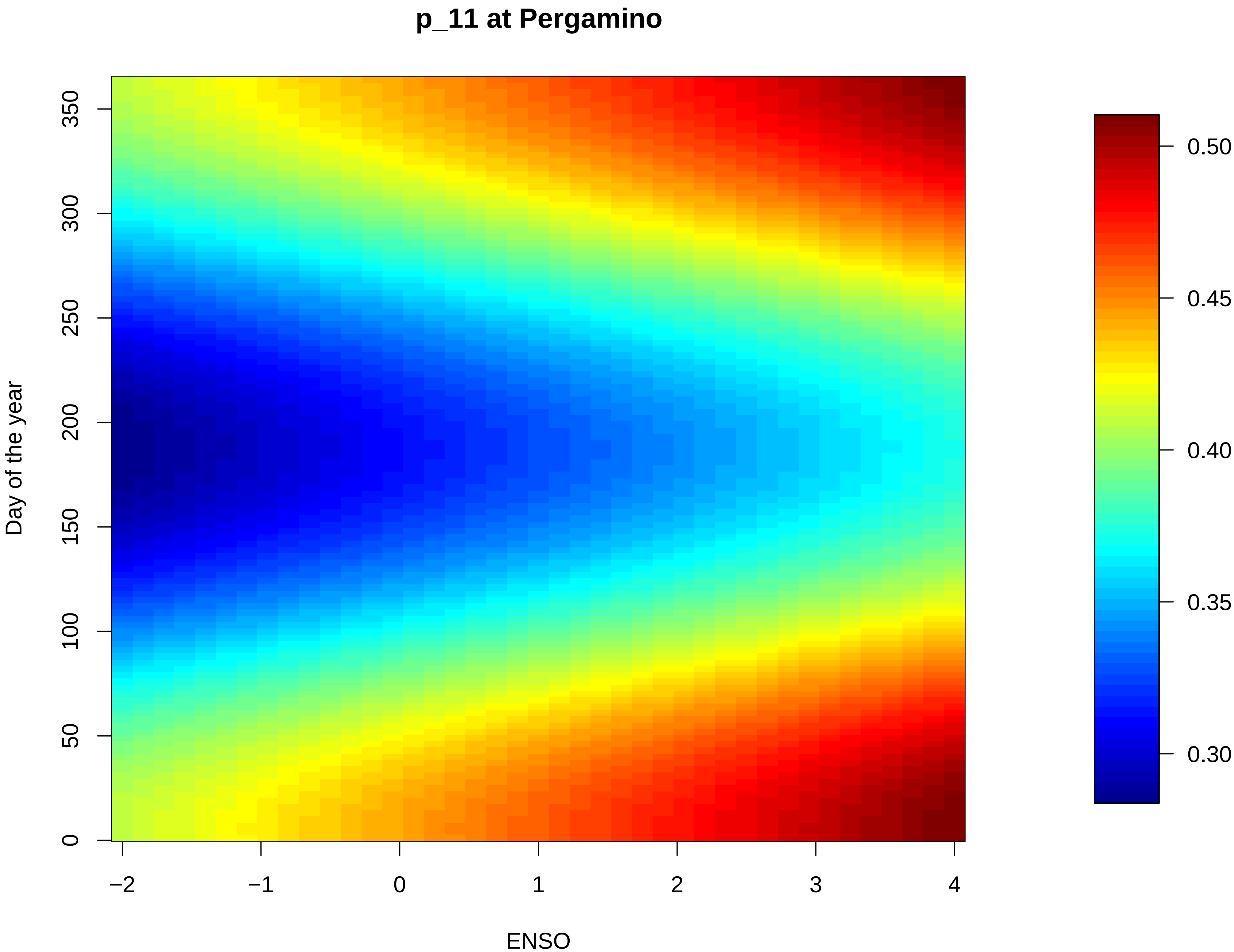
depend on covariate (e. g., via logistic transformation)

-- **Pergamino, Argentina daily precipitation occurrence example**

Annual cycle

El Niño (ENSO) index (Monthly mean)

GLM used to fit model


p_11 at Pergamino
Day of the year
0
50
100
150
200
250
300
350
−2
−1
0
1
2
3
4
ENSO
0.30
0.35
0.40
0.45
0.50

- **Chain-Dependent Process Conditional on Observed Covariate**

-- Conditional distribution of X_t given $J_t = i$ depends on covariate as well

-- Chico, CA January daily precipitation amount example

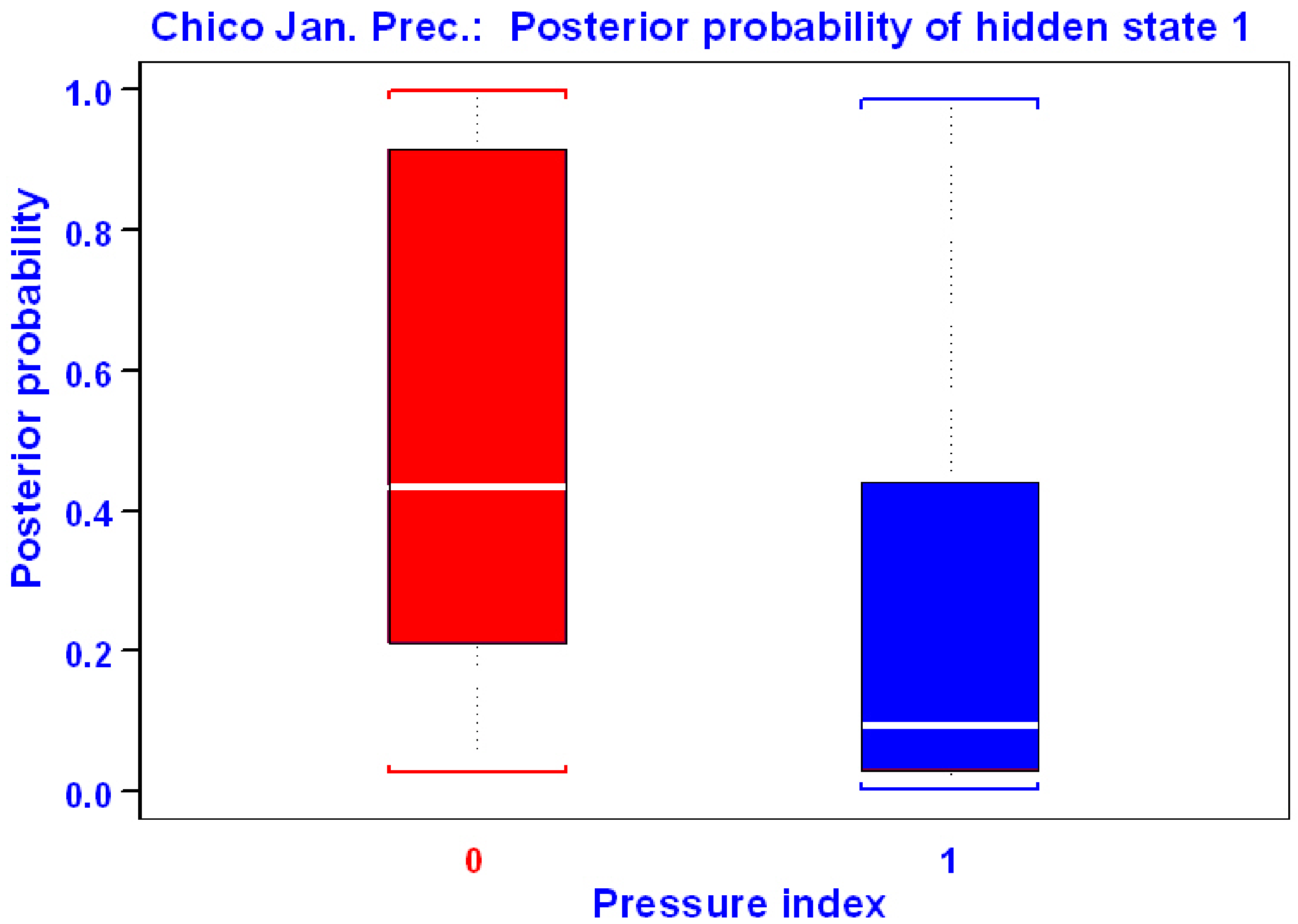
Transition probs. & precipitation "intensity" depend on:

Two-state observed pressure index (Monthly mean)

- **Chain-Dependent Process Conditional on Hidden Variable**

-- Hidden two-state index (Monthly time scale)

-- Use EM algorithm to fit model (still *not* HMM)



Chico Jan. Prec.: Posterior probability of hidden state 1
Posterior probability
1.0
0.8
0.6
0.4
0.2
0.0
0
1
Pressure index

(5) Potential Applications

- **Regime Shifts**

-- Atmospheric/Oceanic circulation patterns

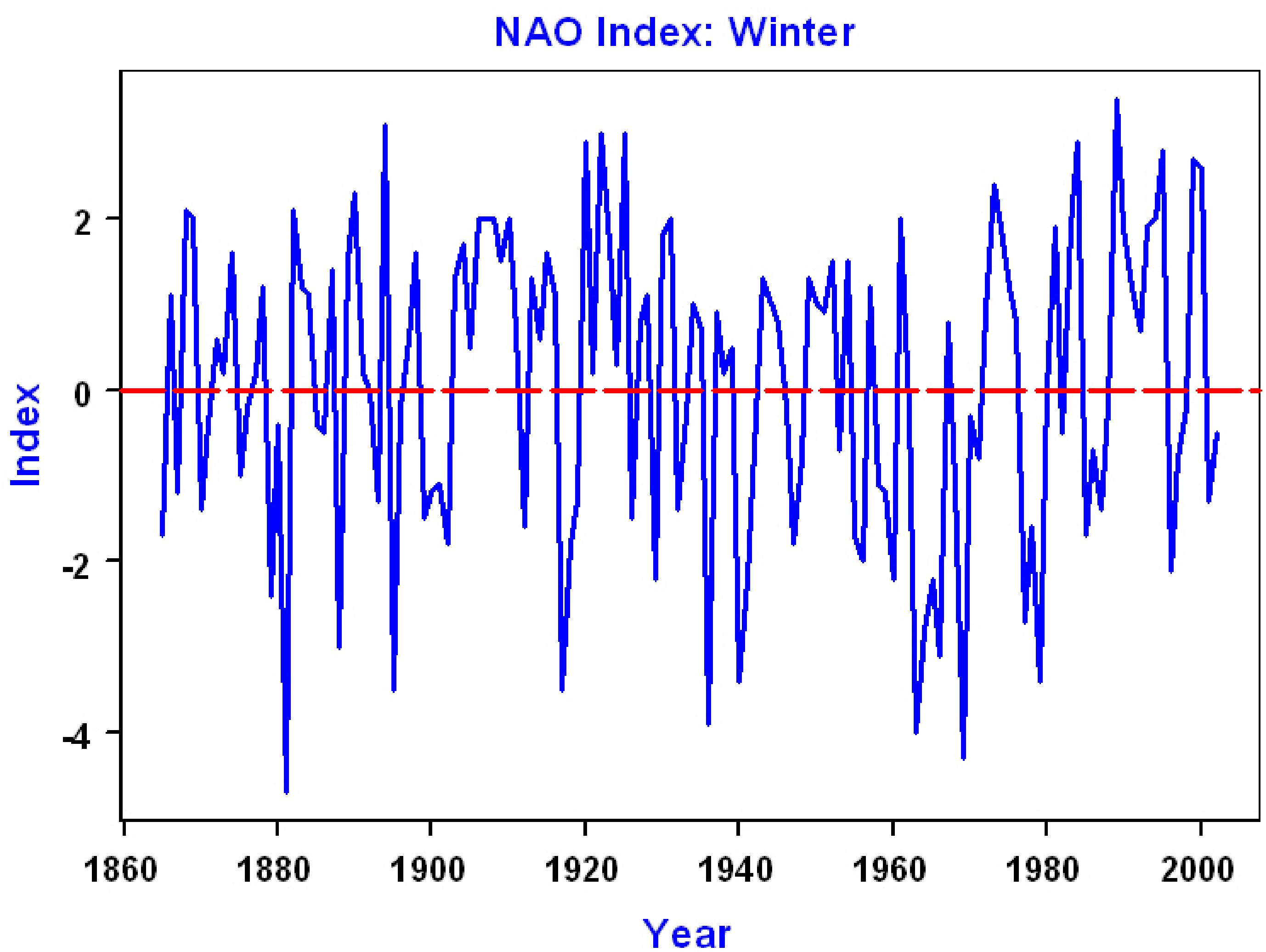
North Atlantic Oscillation (NAO)

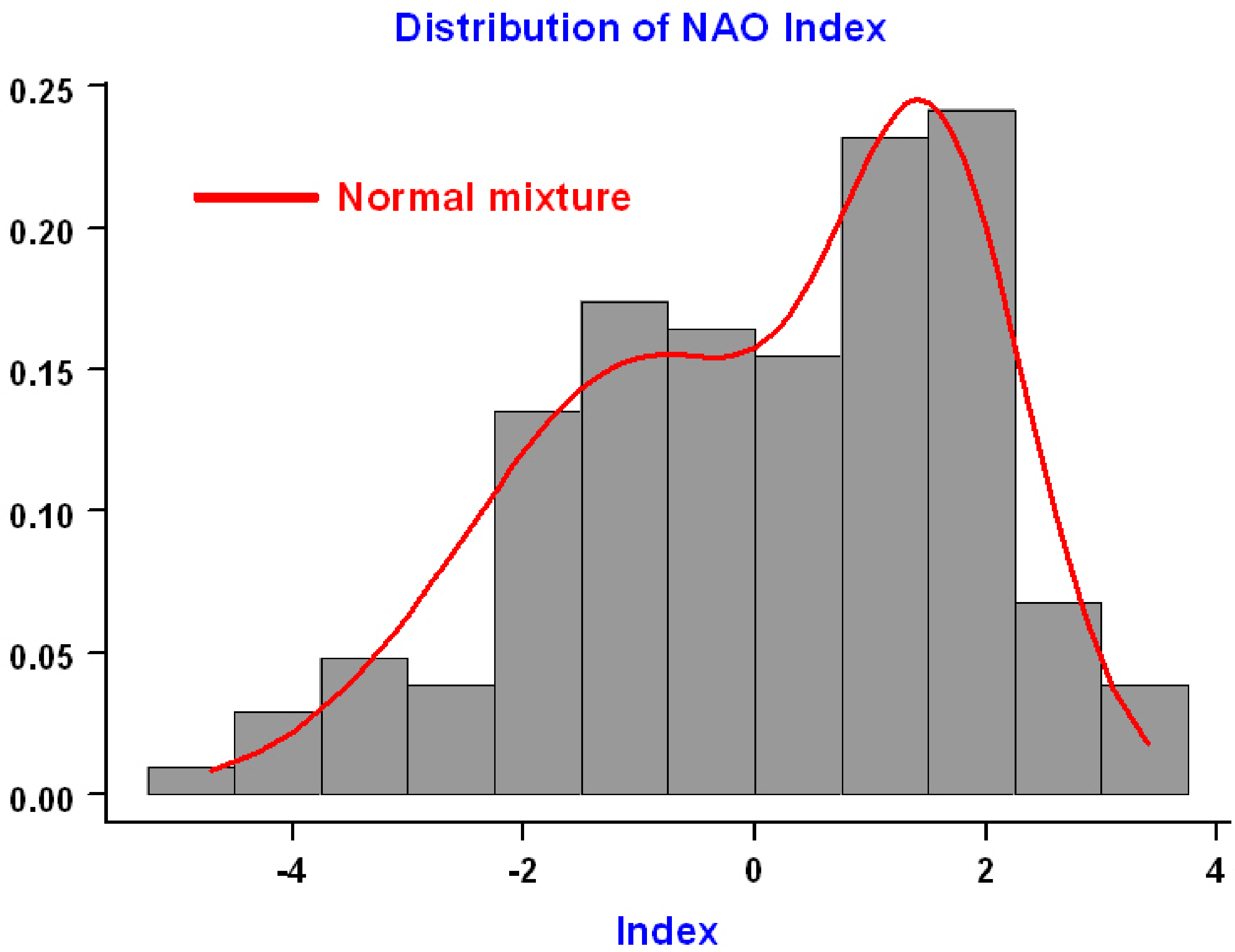
Not much temporal dependence (seasonal or annual time scale)

compared to El Niño / Southern Oscillation

Existence of “regimes”? (Conditional non-stationarity)

Mixture of two normal distributions (still *not* HMM)


NAO Index: Winter
Index
Year
2
0
-2
-4
1860
1880
1900
1920
1940
1960
1980
2000



Distribution of NAO Index
Normal mixture
0.25
0.20
0.15
0.10
0.05
0.00
-4
-2
0
2
4
Index

- **Abrupt Climate Change**

-- Palaeoclimatology

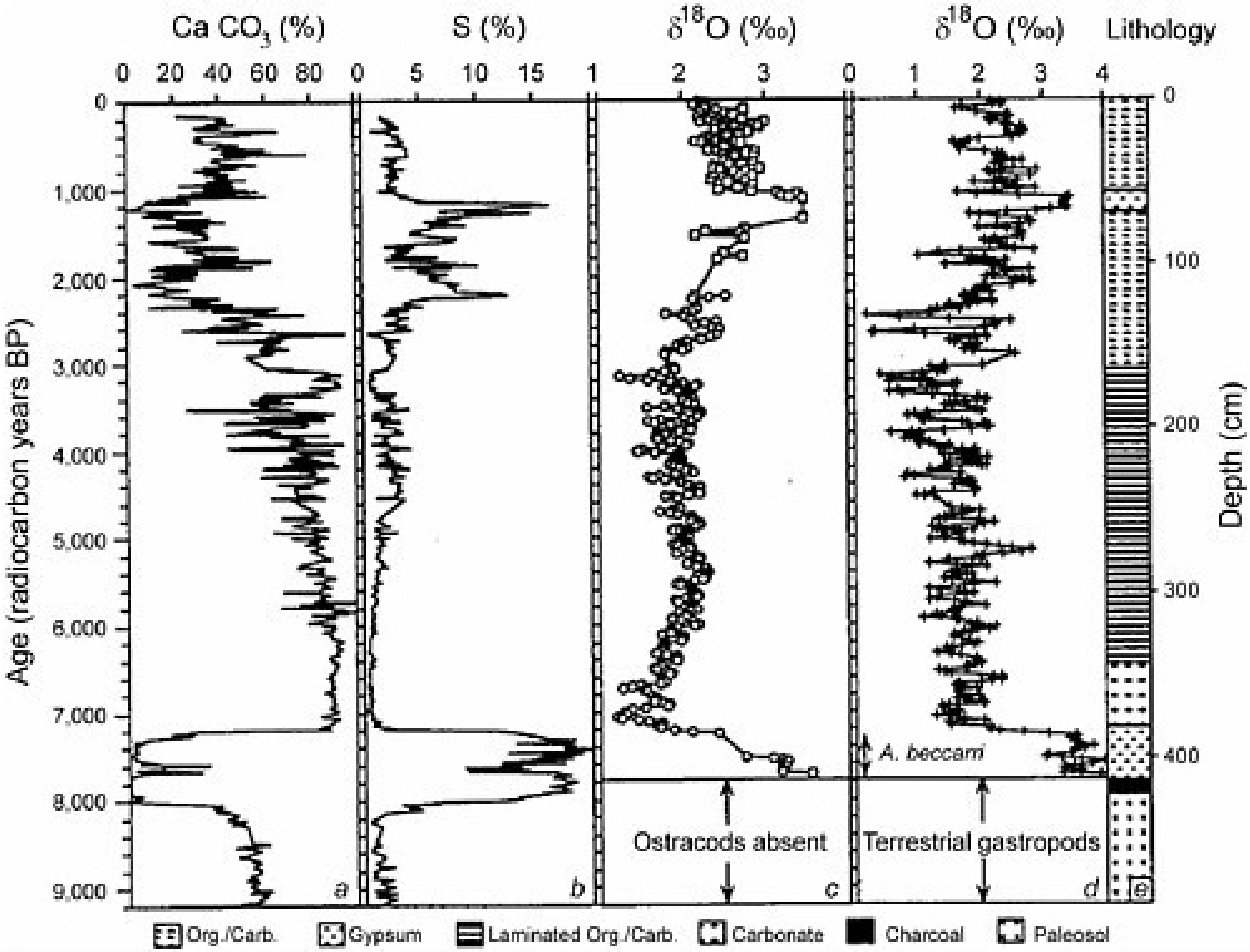
Climate reconstruction

-- Changepoint analysis

Issue of temporal dependence (Use HMM?)

-- Uncertainty analysis

Policy implications (Global climate change)



Ca CO3 (%)
S (%)
δ18O (‰)
δ18O (‰)
Lithology
Age (radiocarbon years BP)
Depth (cm)
Ostracods absent
Terrestrial gastropods
A. beccarii
a
b
c
d
e
Org./Carb.
Gypsum
Laminated Org./Carb.
Carbonate
Charcoal
Paleosol

(6) Role of Hidden Markov Models

- **Environmental/Geophysical Phenomena**

-- **What has been learned from hidden Markov models?**

(i) More flexible model

(ii) Physical interpretation

Verification of what is already known

Any new discoveries?

Domo Arigato Gozaimasu