Monitoring local rainfall via hidden Markov models

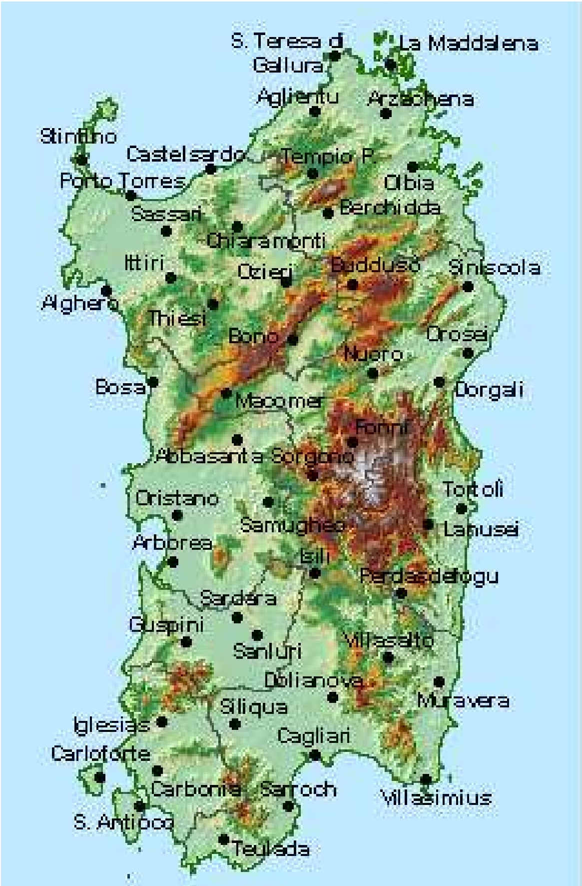
Antonella Bodini

antonella.bodini@mi.imati.cnr.it

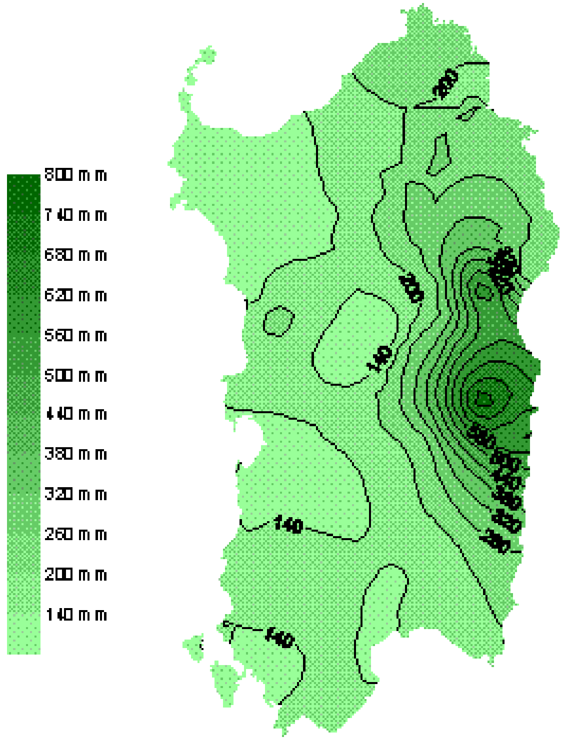
CNR-IMATI, Milano

B. Betrò	CNR-IMATI, Milano
A. Cossu	SAR-Sardegna, Sassari

The problem



DECEMBER 2004, CUMULATE



(kindly provided by SAR)

contd

1961-1990, September – January (SJ)

	Arzana	Gairo	Jerzu	Villagrande
Longitude	93144	92758	93045	93027
Latitude	395505	395126	394815	395743
altitude (m.)	674	784	550	679
Mean ...				
n. of wet days	45.9	38.3	45.9	30.0
daily rainfall	3.86	3.47	3.26	3.55
maximum over SJ	120.1	100.5	92.3	133.4
cumulate over SJ	591.0	531.1	499.2	542.5

Statistical methods

- ARMA & Co.
- Two–parts models
- resampling models
- Hidden Markov Models

correspondence between the hidden states
and the concept of discrete weather states.
Instead of explicity defining the weather states,
HMMs allow us to define them
according to the observed data.

HMM

$X_t = (X_{t1}, \ldots, X_{tq})$ r.v., q rain stations; $x_{ti} \in \mathbb{R}^+$

$C_t \in \{1, \ldots, m\}$ hidden process

$X_{1:T} := (X_1, \ldots, X_T), \; C_{1:T} := (C_1, \ldots, C_T)$

$\mathcal{L}(X) \equiv$ distribution of X

- $\mathcal{L}(X_t | X_{1:t-1}, C_{1:t}) = \mathcal{L}(X_t | C_t)$

- C_t homogeneous, first–order Markov Chain

MacDonald and Zucchini (1997)

contd

- $\mathcal{L}(X_t|C_t) = \prod_i \mathcal{L}(X_{ti}|C_t)$ and DOES NOT DEPEND ON t

Zucchini and Guttorp (1991)

- $\ast\ \mathcal{L}(X_{ti}|C_t = c) \ = w_{ic}\,\delta_0 + (1 - w_{ic})F(\,\cdot\,|\theta_{ic})$

contd

- F : (mixture of) Gamma, Exponential, ...
- BIC+cross–validation, for model selection:
  - conditional distributions
  - n. of hidden states
- MVNHMM toolbox, Kirshner (2005, up. in 2006), http://www.datalab.uci.edu/software/mvnhmm/ (EM algorithm)

Case study

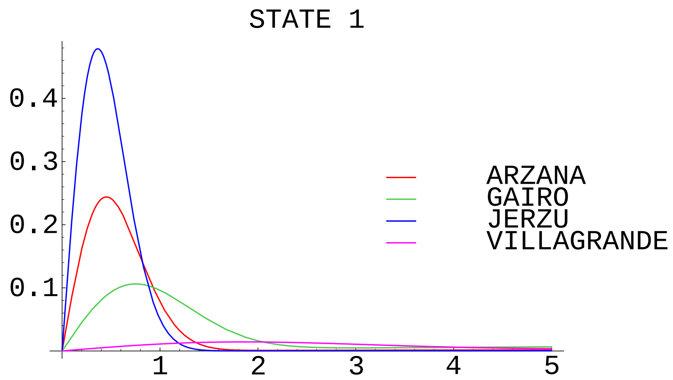
Data transformation: $x^2/100$

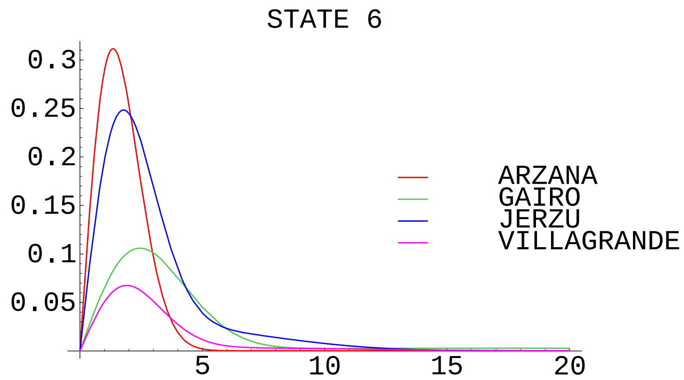
Best model for *transformed* data:

- 6 states
- $\mathcal{L}(X_{ti}|C_t = c)$ =
  mixture of δ_0 and 2 Exponential distrs.

Conditional distrs.

Dirac's weights	C=1	C=2	C=3	C=4	C=5	C=6
Arzana	0.7963	0.0446	0.0541	0.0871	0.9992	0.2945
Gairo	0.7906	0.0555	0.1738	0.3469	0.9969	0.4941
Jerzu	0.6953	0.0222	0.0453	0.1333	0.9976	0.1581
Villagrande	0.9334	0.0573	0.0978	0.3547	0.9990	0.7603

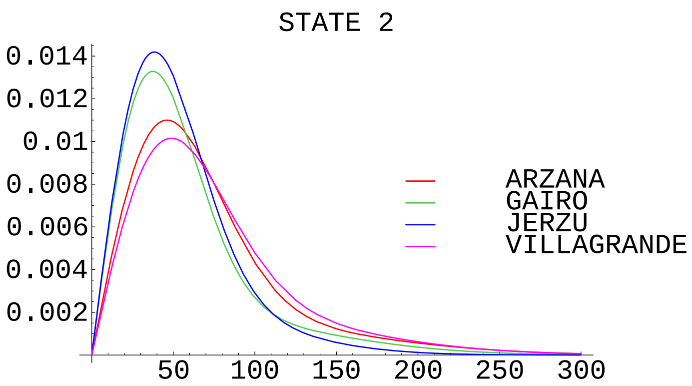


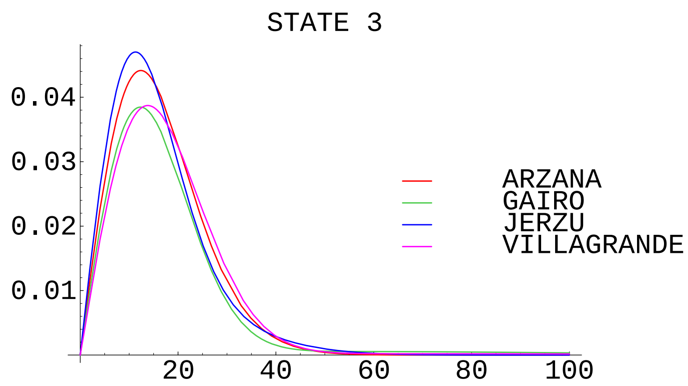


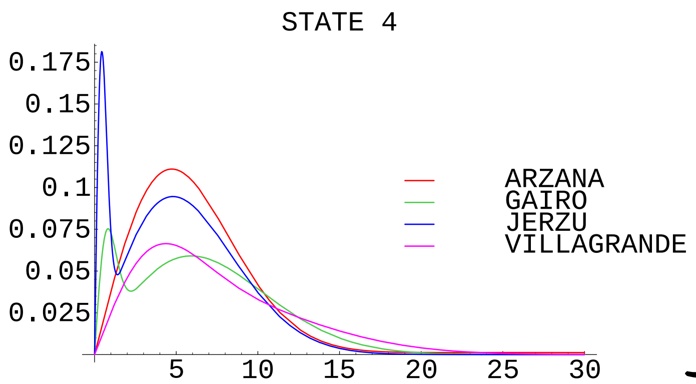
cond. distrs. for original data

contd

Dirac's weights	C=1	C=2	C=3	C=4	C=5	C=6
Arzana	0.7963	0.0446	0.0541	0.0871	0.9992	0.2945
Gairo	0.7906	0.0555	0.1738	0.3469	0.9969	0.4941
Jerzu	0.6953	0.0222	0.0453	0.1333	0.9976	0.1581
Villagrande	0.9334	0.0573	0.0978	0.3547	0.9990	0.7603



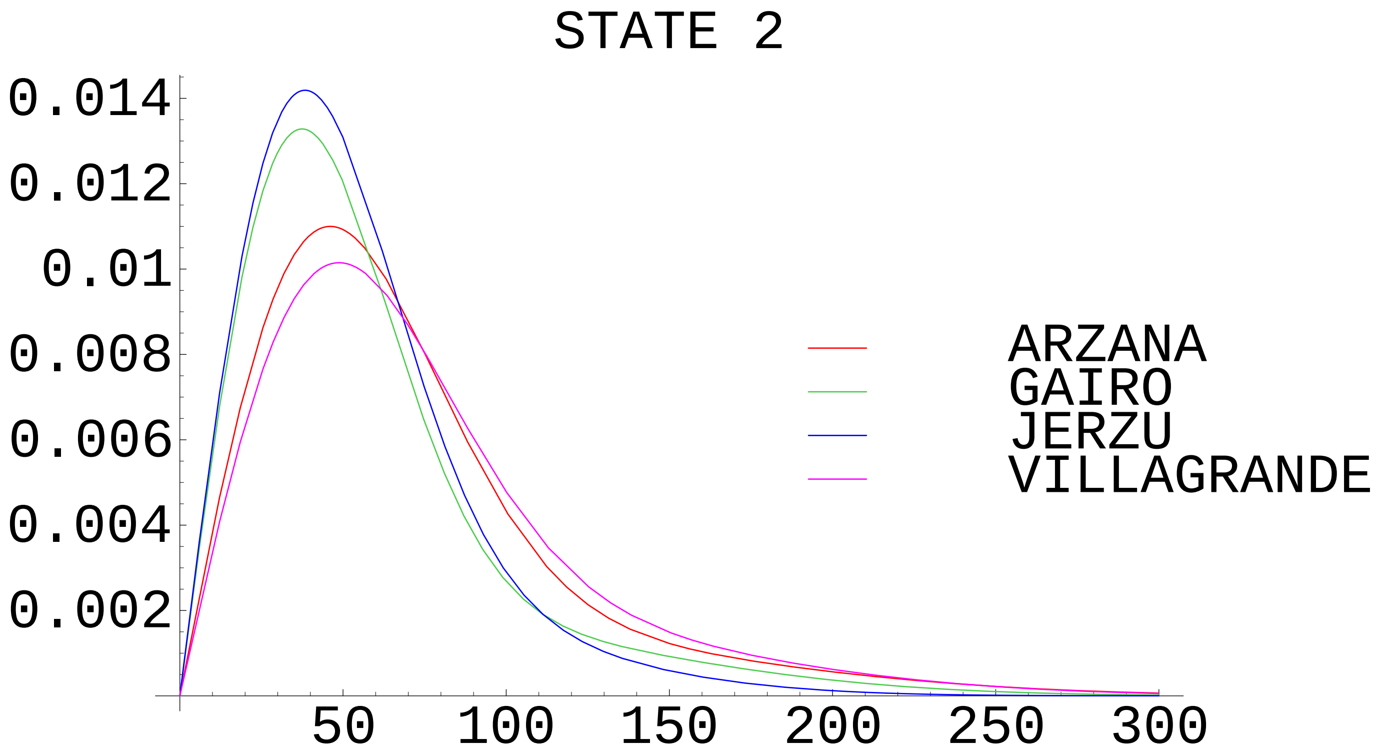




cond. distrs. for

original d.

contd


STATE 2
0.014
0.012
0.01
0.008
0.006
0.004
0.002
50
100
150
200
250
300
ARZANA
GAIRO
JERZU
VILLAGRANDE

Goodness–of–fit

- (general) good fit of frequencies

DRY DAY

model	Arzana	Gairo	Jerzu	Villagrande
empirical	0.682152	0.749334	0.65739	0.804594
estimated	0.692413	0.747431	0.656307	0.807495

MEAN DAILY RAINFALL

model	Arzana	Gairo	Jerzu	Villagrande
empirical	3.6633	3.3904	3.1552	**3.3432**
estimated	3.6211	3.3852	3.1710	**3.5337**

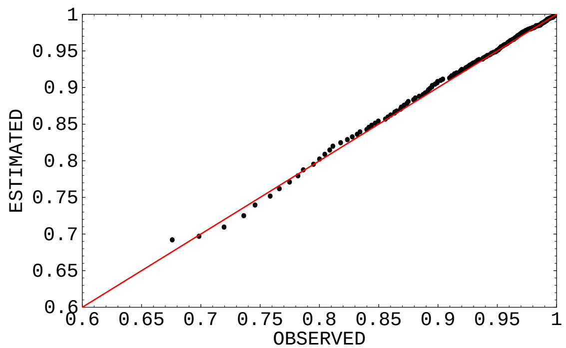
EXTREME EVENT

model	Arzana	Gairo	Jerzu	Villagrande
empirical	0.0054	0.0039	0.0021	0.0047
estimated	0.0046	0.0036	0.0021	0.0058

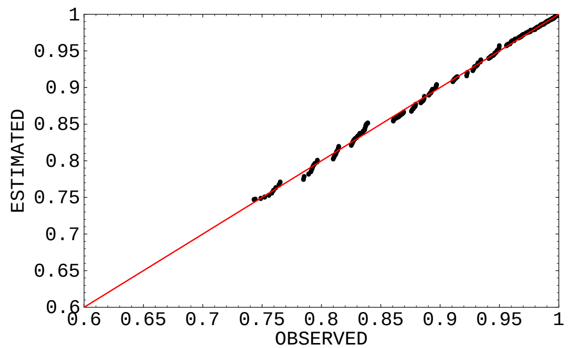
contd

- (Altman, 2004): empirical/estimated cdf

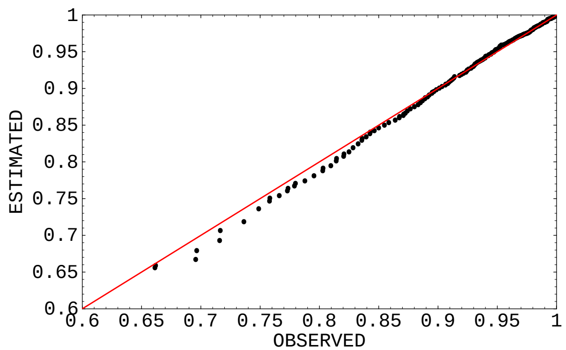
ARZANA



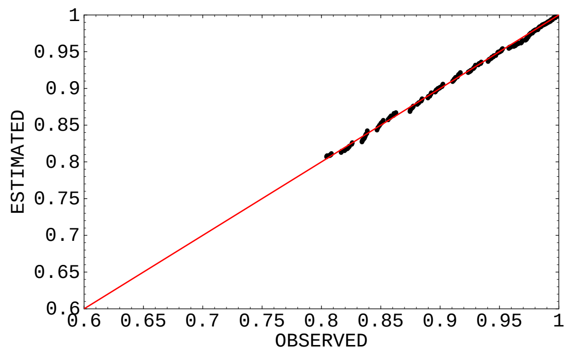
GAIRO



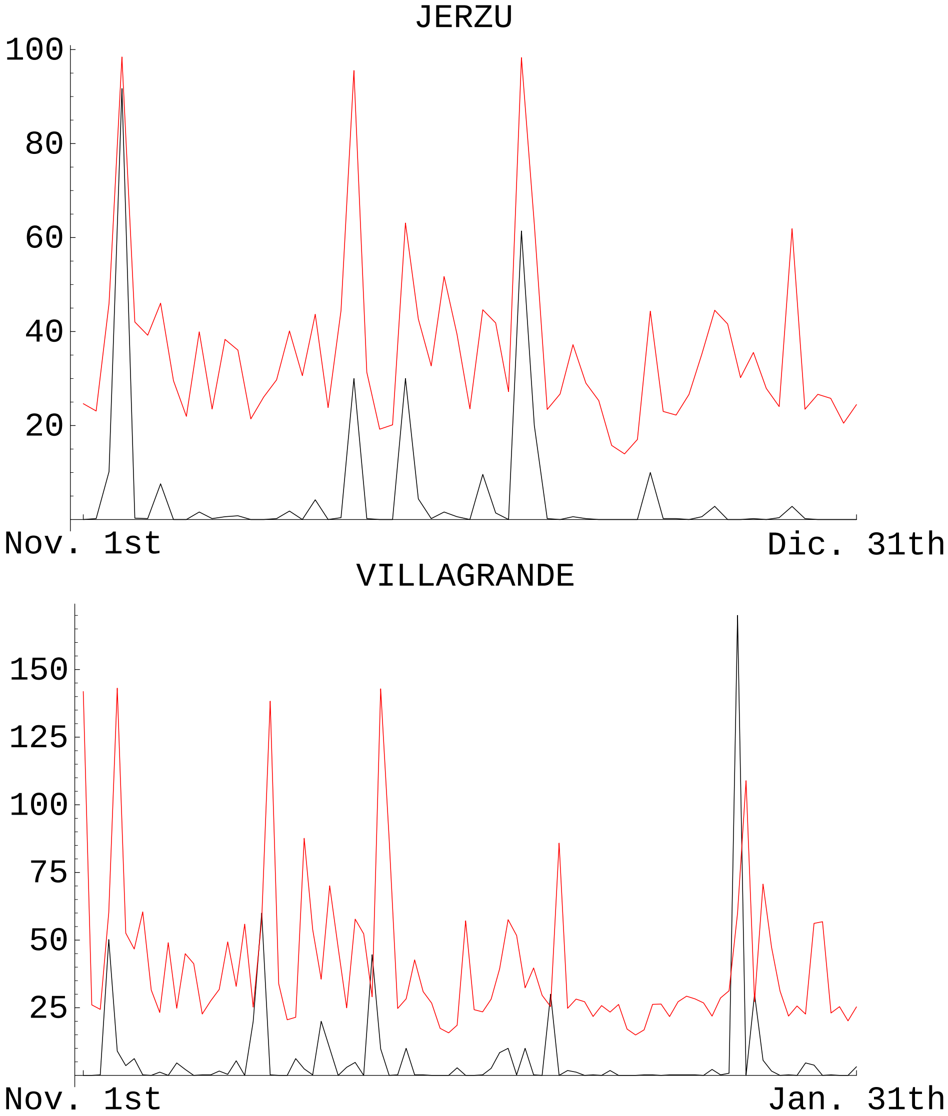
JERZU



VILLAGRANDE



Validation



data:

1997–1998

u.b 95% CI

conditional to the past

ESS

Estimated State Sequence:

the most likely seq. of states associated with data.

Median of daily rainfall for each state, according to ESS:

	C=1	C=2	C=3	C=4	C=5	C=6
Arzana	0	55.4	14.8	5.5	0	1
Gairo	0	47.5	12.	4.1	0	0.8
Jerzu	0	44.2	13.8	4.4	0	2
Villagrande	0	63.0	15.6	4.5	0	0

contd

Estimated State Sequence:

the most likely seq. of states associated with data.

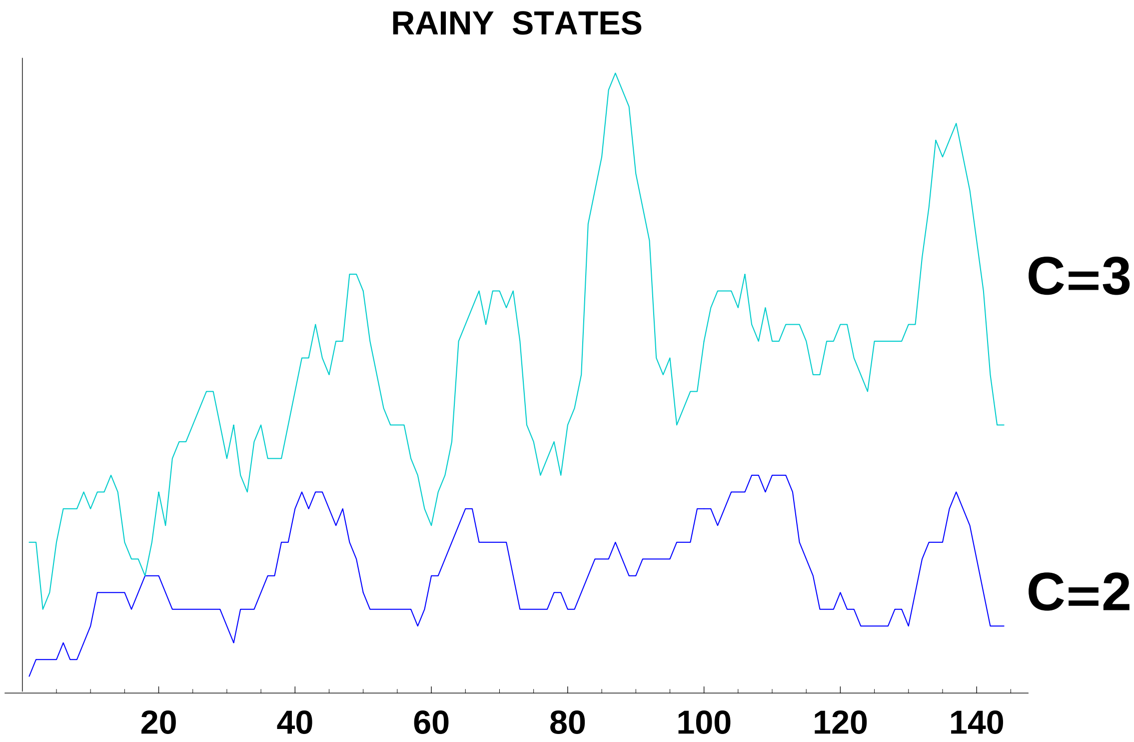
Conditional 95%–confidence intervals:

station	C=1	C=2	C=3	C=4	C=5	C=6
Arzana	(0, 1.3)	(0, 204.2)	(0, 36.7)	(0, 30.9)	(0,0)	(0, 3.8)
Gairo	(0, 10.5)	(0, 176.0)	(0, 67.3)	(0, 14.8)	(0,0)	(0, 26.3)
Jerzu	(0, 1)	(3.0, 139.9)	(0, 39.9)	(0, 12.5)	(0,0)	(0, 9.1)
Villagrande	(0, 4.2)	(0, 204.1)	(0, 43.4)	(0, 17.6)	(0,0)	(0, 5.7)

contd

Estimated State Sequence:

the most likely seq. of states associated with data.



10-day moving average of mean frequencies.

References

- **Perreault L., Fortin V., Salas J. D.** (2004) *Mixtures and Hidden Markov Models for Estimating Flood Quantiles and Risk*. American Geophysical Union, Spring Meeting 2004, abstract H54B-01.

- **MacDonald I.L., Zucchini W.** (1997) *Hidden Markov and Other Models for Discrete Time Series*. Chapmann & Hall, London.
- **Zucchini W., Guttorp P.** (1991) *A hidden Markov model for space–time precipitation.* Water Resources Research, 27, 1917–1923.
- **Altman MCK.** (2004) *Assessing the Goodness-of-Fit of Hidden Markov Models*. Biometrics, 60, 444–450.