Modelling Non-linear and Non-stationary Time Series

Chapter 7(extra): (Generalized) Hidden Markov Models

Henrik Madsen

Lecture Notes

September 2016

Hidden Markov Model

First Order Markov Property

$$p(X_t|X_{t-1}) = p(X_t|\mathcal{X}^{(t-1)}), \quad t \in \mathbb{N} \tag{1}$$

$$p(Y_t|X_t) = p(Y_t|\mathcal{X}^{(t)}, \mathcal{Y}^{(t-1)}), \quad t \in \mathbb{N} \tag{2}$$

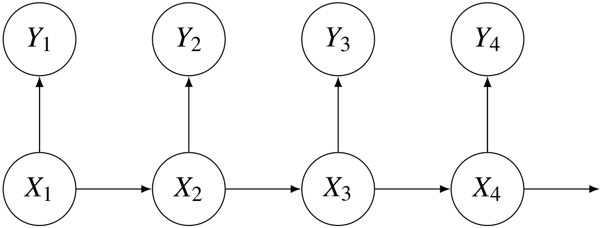


Figure : Directed graph of basic HMM. The index denotes time.

Markov Chains

Discrete state vector at time t, X_t, with m states.

Transition probability

$$p(X_t = j | X_{t-s} = i) \tag{3}$$

One-step transition probability

$$\gamma_{ij,t} = p(X_t = j | X_{t-1} = i) \tag{4}$$

One-step transition probability matrix from time $t-1$ to t

$$\mathbf{\Gamma}_t = \begin{pmatrix} \gamma_{11,t} & \cdots & \gamma_{1m,t} \\ \vdots & \ddots & \vdots \\ \gamma_{m1,t} & \cdots & \gamma_{mm,t} \end{pmatrix} \tag{5}$$

where the rows must sum to 1.

Generalized State Space Models (GSSM)

Parameter driven - model evolve independently of the past observation process.

$$p(y_t|x_t) = p\left(y_t|x_t, \mathcal{X}^{(t-1)}, \mathcal{Y}^{(t-1)}\right) \quad (6a)$$
$$p(x_{t+1}|x_t) = p\left(x_{t+1}|x_t, \mathcal{X}^{(t-1)}, \mathcal{Y}^{(t)}\right) \quad (6b)$$

Observation driven - model depends on the past observation process.

$$p(y_t|x_t) = p\left(y_t|x_t, \mathcal{X}^{(t-1)}, \mathcal{Y}^{(t-1)}\right) \quad (7a)$$
$$p\left(x_{t+1}|\mathcal{Y}^{(t)}\right) = p\left(x_{t+1}|x_t, \mathcal{X}^{(t-1)}, \mathcal{Y}^{(t)}\right) \quad (7b)$$

Forward Selection

Maximum of 5 states and model chosen by AIC

$$\text{AIC} = -2\mathcal{L} + 2p \tag{8}$$

Forecast Pseudo Residuals

$$z_t = \Phi^{-1}\left(p\left(Y_t \leq y_t | \mathcal{Y}^{(t-1)}\right)\right) \tag{9}$$

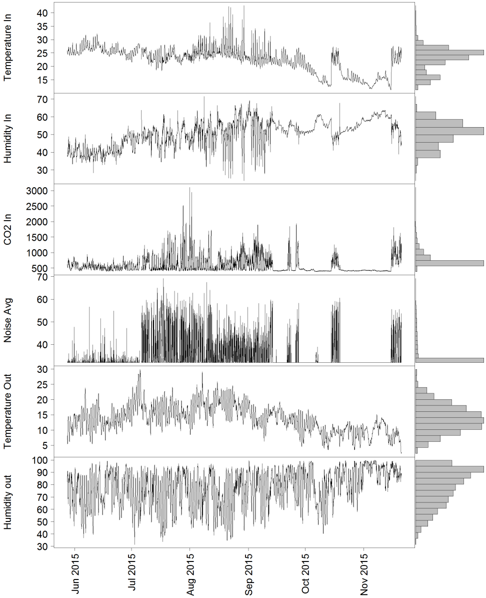
Marginal distribution

Data

IC-Meter

Table : 5-Minute Variables.

Dataset	Feature (type)	Unit	Description
Indoor-Minutes	Datetime		Date and time
Indoor-Minutes	Temperature In	$^{\circ}C$	Indoor temperature
Indoor-Minutes	Humidity In	$\%$	Indoor humidity
Indoor-Minutes	CO2 In	ppm	Indoor CO_2 content
Indoor-Minutes	Noise Avg	DB	Average noise level (5 min)
Indoor-Minutes	Noise Peak	DB	Largest average of 3 subsequent measurement (15 sec per measurement)


Temperature In
Humidity In
CO2 In
Noise Avg
Temperature Out
Humidity out
Jun 2015
Jul 2015
Aug 2015
Sep 2015
Oct 2015
Nov 2015

Homogen HMM

Original Scale

Setting

$$
\begin{aligned}
y_t &= CO_{2,t} \\
p(x_t|x_{t-1}) &\sim \mathbf{\Gamma} \\
p(y_t|x_t) &\sim \mathcal{N}\left(\mu_i, \sigma_i^2\right) \text{ for } i = 1, 2, \cdots, m
\end{aligned}
$$

Results

Table : Comparison of univariate homogen HMMs for 2 to 5 states.

	$\mathcal{L}$	p	AIC	BIC
2 states	-97903	6	195818	195864
3 states	-91239	12	182502	182595
4 states	-87492	20	175023	175178
5 states	-83968	30	167995	168227

Homogen HMM

Original Scale

Table : Fit of the HMM (CO_2) with 5 states.

	δ_i	μ_i	σ_i	γ_{i1}	γ_{i2}	γ_{i3}	γ_{i4}	γ_{i5}
State 1	0.21	410.79	6.88	0.99	0.01	0.01	0.00	0.00
State 2	0.18	432.00	7.42	0.02	0.94	0.04	0.00	0.00
State 3	0.20	477.92	24.17	0.00	0.05	0.90	0.05	0.00
State 4	0.23	618.45	68.26	0.00	0.00	0.05	0.92	0.03
State 5	0.18	1013.61	267.72	0.00	0.00	0.00	0.04	0.96

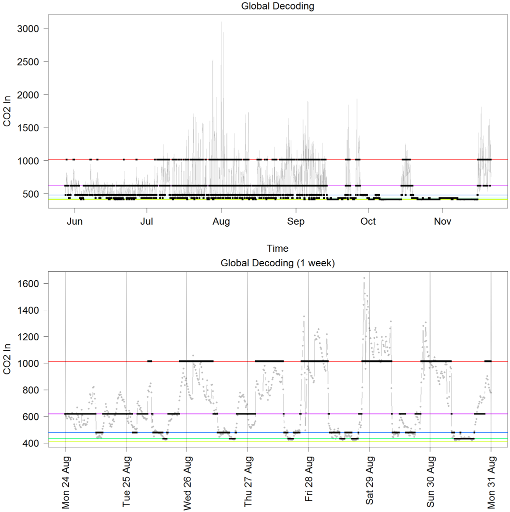


Figure : Global Decoding of the HMM (CO_2) with 5 states. Top is the entire time series. Bottom is zoomed in on one day. Vertical lines indicate 00:00. Horizontal lines indicate the state dependent mean.

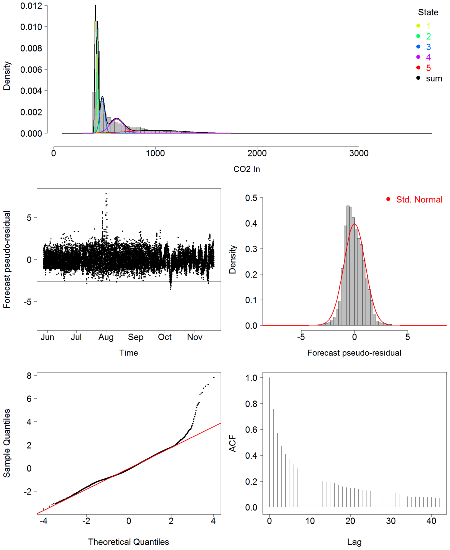


Figure : Fit of the HMM (CO_2) with 5 states.

Homogen HMM

Transformed Scale

$$h(y) = log(y - 350)$$

Setting

$$\begin{aligned} y_t &= h(CO_{2,t}) \\ p(x_t|x_{t-1}) &\sim \mathbf{\Gamma} \\ p(y_t|x_t) &\sim \mathcal{N}\left(\mu_i, \sigma_i^2\right) \text{ for } i = 1, 2, \cdots, m \end{aligned}$$

Results

Table : Comparison of univariate (log transformed CO_2) homogen HMMs for 2 to 5 states.

	$\mathcal{L}$	p	AIC	BIC
2 states	-9378	6	18768	18814
3 states	-4292	12	8609	8701
4 states	-800	20	1640	1795
5 states	2181	30	-4303	-4071

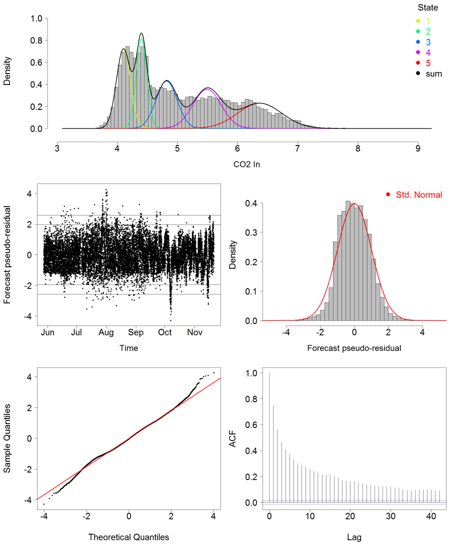


Figure : Fit of the HMM (log CO_2) with 5 states.

Generalized Form

Hierarchical Model

$$p(y_t|x_t,\boldsymbol{\theta})$$
$$p\left(x_{t+1}|x_t,\boldsymbol{\theta}\right)$$

Hierarchical model by random effects

$$\boldsymbol{Y}|(\boldsymbol{U}=\boldsymbol{u}) \sim f_{Y|u}\left(\boldsymbol{y};\boldsymbol{u},\boldsymbol{\beta}\right) \quad (10a)$$
$$\boldsymbol{U} \sim f_U\left(\boldsymbol{u};\boldsymbol{\Psi}\right) \quad (10b)$$

The likelihood is given by

$$L\left(\boldsymbol{\beta},\boldsymbol{\Psi};\boldsymbol{y},\boldsymbol{u}\right)=f\left(\boldsymbol{y};\boldsymbol{\theta}\right)=f\left(\boldsymbol{y},\boldsymbol{u};\boldsymbol{\beta},\boldsymbol{\Psi}\right)=f_{Y|u}\left(\boldsymbol{y};\boldsymbol{u},\boldsymbol{\beta}\right)f_U\left(\boldsymbol{u};\boldsymbol{\Psi}\right) \quad (11)$$

The distribution $f\left(\boldsymbol{y},\boldsymbol{u};\boldsymbol{\beta},\boldsymbol{\Psi}\right)$ is given by an exponential dispersion model with density given by Equation (**??**) with canonical link $\eta=h(\mu)$.
The random effect is given by $f_U\left(\boldsymbol{u};\boldsymbol{\Psi}\right)$ and the conditional likelihood by $f_{Y|u}\left(\boldsymbol{y};\boldsymbol{u},\boldsymbol{\beta}\right)$.

Transition Probability Matrix with Covariates

$$\boldsymbol{\Gamma}_t = \begin{pmatrix} \frac{\exp(\boldsymbol{X}_{D,t}^T\boldsymbol{\beta}_{11} + \boldsymbol{ZU}_{11})}{\exp(\boldsymbol{X}_{D,t}^T\boldsymbol{\beta}_{11} + \boldsymbol{ZU}_{11}) + \sum_{j\neq 1}\exp(\tau_{1j})} & \cdots & \frac{\exp(\tau_{1m})}{\exp(\boldsymbol{X}_{D,t}^T\boldsymbol{\beta}_{11} + \boldsymbol{ZU}_{11}) + \sum_{j\neq 1}\exp(\tau_{1j})} \\ \vdots & \ddots & \vdots \\ \frac{\exp(\tau_{m1})}{\exp(\boldsymbol{X}_{D,t}^T\boldsymbol{\beta}_{mm} + \boldsymbol{ZU}_{mm}) + \sum_{j\neq m}\exp(\tau_{mj})} & \cdots & \frac{\exp(\boldsymbol{X}_{D,t}^T\boldsymbol{\beta}_{mm} + \boldsymbol{ZU}_{mm})}{\exp(\boldsymbol{X}_{D,t}^T\boldsymbol{\beta}_{mm} + \boldsymbol{ZU}_{mm}) + \sum_{j\neq m}\exp(\tau_{mj})} \end{pmatrix} \quad (12)$$

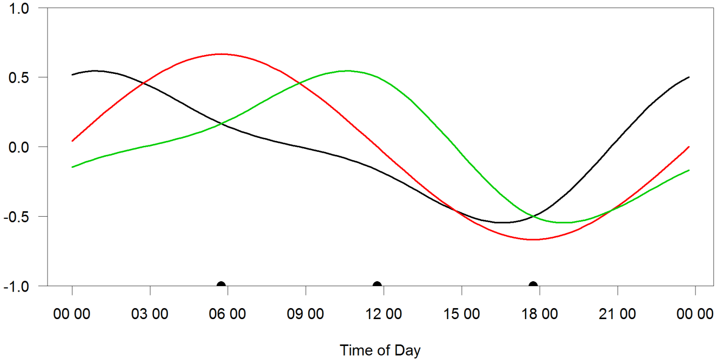


Figure : Periodic Splines. The knots are indicated on the x-axis.

Inhomogen HMM

Transformed Scale

Setting

$$
\begin{aligned}
y_t &= h(CO_{2,t}) \\
p(x_t|x_{t-1}) &\sim \boldsymbol{\Gamma}_t \\
p(y_t|x_t) &\sim \mathcal{N}\left(\mu_i, \sigma_i^2\right) \text{ for } i = 1, 2, \cdots, m
\end{aligned}
$$

Results

Table : Comparison of univariate (log transformed CO_2) inhomogen HMMs for 2 to 5 states.

	$\mathcal{L}$	p	AIC	BIC
2 states	-9319	14	18667	18775
3 states	-4229	24	8506	8692
4 states	-742	36	1556	1835
5 states	2258	50	-4417	-4030

- Same trend in residuals - not appropriate

Markov Switching AR(1)

$$p(x_{t+1}|x_t) = p\left(x_{t+1}|x_t, \mathcal{X}^{(t-1)}, \mathcal{Y}^{(t)}\right) \quad (13a)$$

$$p(y_t|x_t, y_{t-1}) = p\left(y_t|x_t, \mathcal{X}^{(t-1)}, \mathcal{Y}^{(t-1)}\right) \quad (13b)$$

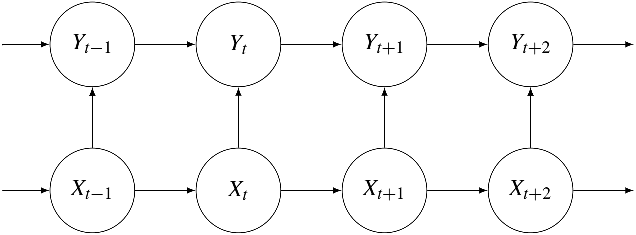


Figure : Directed graph of Markov switching AR(1).

Inhomogen Markov Switching AR(1)

Transformed Scale

Setting

$$
\begin{aligned}
y_t &= h(CO_{2,t}) \\
p(x_t|x_{t-1}) &\sim \mathbf{\Gamma}_t \\
p(y_t|x_t, y_{t-1}) &\sim \mathcal{N}\left(c_i + \phi_i y_{t-1}\ ,\ \sigma_i^2\right) \text{ for } i = 1, 2, \cdots, m
\end{aligned}
$$

Results

Table : Comparison of univariate (log transformed CO_2) inhomogen AR(1) for 2 to 5 states.

	$\mathcal{L}$	p	AIC	BIC
2 states	15845	16	-31659	-31535
3 states	16204*	27	-32354	-32145
4 states	16964	40	-33848	-33538
5 states	17336	55	-34561	-34136

- Residuals - appropriate for 4 and 5 state model!!

Final Model - 5 states

Residuals

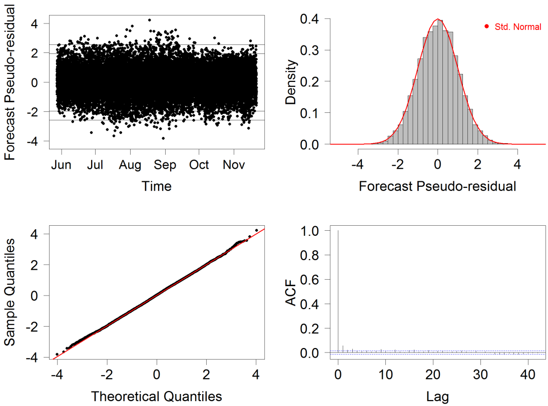


Figure : Model diagnostics of the final model.

Final Model

Marginal Distribution

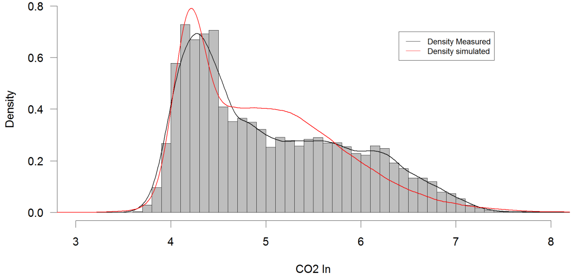


Figure : Comparison of the distribution of the measured CO_2 and the simulated distribution.

Final Model

Interpretation of the states

- State 1: Absence or sleeping
- State 2: Long term absence
- State 3: Outdoor interaction
- State 4: Presence (high activity)
- State 5: Presence (long term, low activity)

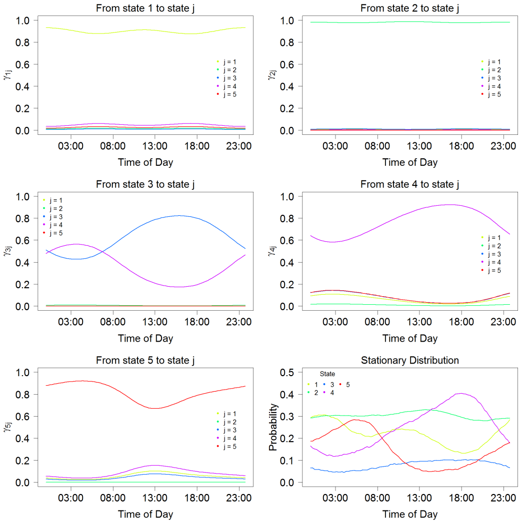


Figure : Transition probabilities over the day of the final model. The lower right plot is the stationary distribution.

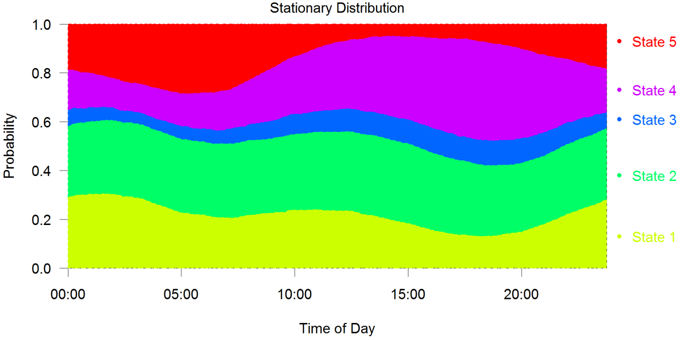


Figure : Profile of the states over the course of the day. I.e. Stacked stationary probabilities over the course of the day of the final model.

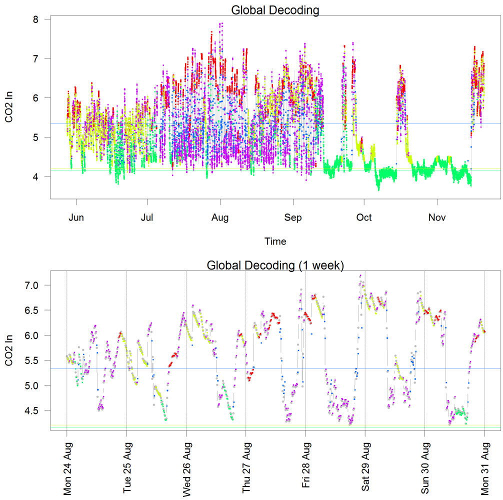


Figure : Global Decoding of the final model. Top is the entire time series. Bottom is zoomed in on one day. Vertical lines indicate 00:00. Horizontal lines indicate the state dependent mean.