Full text available at: http://dx.doi.org/10.1561/2200000054

Explicit-Duration Markov Switching Models

Silvia Chiappa
Statistical Laboratory, University of Cambridge, UK
and
Microsoft Research Cambridge, UK
chiappa.silvia@gmail.com

the essence of knowledge
Boston — Delft

Full text available at: http://dx.doi.org/10.1561/2200000054

Foundations and Trends® in Machine Learning

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

S. Chiappa. *Explicit-Duration Markov Switching Models*. Foundations and Trends® in Machine Learning, vol. 7, no. 6, pp. 803–886, 2014.

This Foundations and Trends® issue was typeset in LaTeX using a class file designed by Neal Parikh. Printed on acid-free paper.

ISBN: 978-1-60198-831-7
© 2014 S. Chiappa

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, mechanical, photocopying, recording or otherwise, without prior written permission of the publishers.

Photocopying. In the USA: This journal is registered at the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923. Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by now Publishers Inc for users registered with the Copyright Clearance Center (CCC). The 'services' for users can be found on the internet at: www.copyright.com

For those organizations that have been granted a photocopy license, a separate system of payment has been arranged. Authorization does not extend to other kinds of copying, such as that for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale. In the rest of the world: Permission to photocopy must be obtained from the copyright owner. Please apply to now Publishers Inc., PO Box 1024, Hanover, MA 02339, USA; Tel. +1 781 871 0245; www.nowpublishers.com; sales@nowpublishers.com

now Publishers Inc. has an exclusive license to publish this material worldwide. Permission to use this content must be obtained from the copyright license holder. Please apply to now Publishers, PO Box 179, 2600 AD Delft, The Netherlands, www.nowpublishers.com; e-mail: sales@nowpublishers.com

Full text available at: http://dx.doi.org/10.1561/2200000054

Foundations and Trends® in Machine Learning

Volume 7, Issue 6, 2014

Editorial Board

Editor-in-Chief

Michael Jordan
University of California, Berkeley
United States

Editors

Peter Bartlett
UC Berkeley

Yoshua Bengio
University of Montreal

Avrim Blum
CMU

Craig Boutilier
University of Toronto

Stephen Boyd
Stanford University

Carla Brodley
Tufts University

Inderjit Dhillon
UT Austin

Jerome Friedman
Stanford University

Kenji Fukumizu
ISM, Japan

Zoubin Ghahramani
University of Cambridge

David Heckerman
Microsoft Research

Tom Heskes
Radboud University

Geoffrey Hinton
University of Toronto

Aapo Hyvarinen
HIIT, Finland

Leslie Pack Kaelbling
MIT

Michael Kearns
UPenn

Daphne Koller
Stanford University

John Lafferty
CMU

Michael Littman
Brown University

Gabor Lugosi
Pompeu Fabra University

David Madigan
Columbia University

Pascal Massart
University of Paris-Sud

Andrew McCallum
UMass Amherst

Marina Meila
University of Washington

Andrew Moore
CMU

John Platt
Microsoft Research

Luc de Raedt
University of Freiburg

Christian Robert
U Paris-Dauphine

Sunita Sarawagi
IIT Bombay

Robert Schapire
Princeton University

Bernhard Schoelkopf
MPI Tübingen

Richard Sutton
University of Alberta

Larry Wasserman
CMU

Bin Yu
UC Berkeley

Full text available at: http://dx.doi.org/10.1561/2200000054

Editorial Scope

Topics

Foundations and Trends® in Machine Learning publishes survey and tutorial articles on the theory, algorithms and applications of machine learning, including the following topics:

- Adaptive control and signal processing
- Applications and case studies
- Behavioral, cognitive, and neural learning
- Bayesian learning
- Classification and prediction
- Clustering
- Data mining
- Dimensionality reduction
- Evaluation
- Game theoretic learning
- Graphical models
- Independent component analysis
- Inductive logic programming
- Kernel methods
- Markov chain Monte Carlo
- Model choice
- Nonparametric methods
- Online learning
- Optimization
- Reinforcement learning
- Relational learning
- Robustness
- Spectral methods
- Statistical learning theory
- Variational inference
- Visualization

Information for Librarians

Foundations and Trends® in Machine Learning, 2014, Volume 7, 6 issues. ISSN paper version 1935-8237. ISSN online version 1935-8245. Also available as a combined paper and online subscription.

Full text available at: http://dx.doi.org/10.1561/2200000054

Foundations and Trends® in Machine Learning
Vol. 7, No. 6 (2014) 803–886
© 2014 S. Chiappa
DOI: 10.1561/2200000054

Explicit-Duration Markov Switching Models

Silvia Chiappa[1]
Statistical Laboratory, University of Cambridge, UK
and
Microsoft Research Cambridge, UK
chiappa.silvia@gmail.com

[1]The author is currently at Google DeepMind, UK.

Full text available at: http://dx.doi.org/10.1561/2200000054

Contents

Full text available at: http://dx.doi.org/10.1561/2200000054

Full text available at: http://dx.doi.org/10.1561/2200000054

Abstract

Markov switching models (MSMs) are probabilistic models that employ multiple sets of parameters to describe different dynamic regimes that a time series may exhibit at different periods of time. The switching mechanism between regimes is controlled by unobserved random variables that form a first-order Markov chain. Explicit-duration MSMs contain additional variables that explicitly model the distribution of time spent in each regime. This allows to define duration distributions of any form, but also to impose complex dependence between the observations and to reset the dynamics to initial conditions. Models that focus on the first two properties are most commonly known as hidden semi-Markov models or segment models, whilst models that focus on the third property are most commonly known as changepoint models or reset models. In this monograph, we provide a description of explicit-duration modelling by categorizing the different approaches into three groups, which differ in encoding in the explicit-duration variables different information about regime change/reset boundaries. The approaches are described using the formalism of graphical models, which allows to graphically represent and assess statistical dependence and therefore to easily describe the structure of complex models and derive inference routines. The presentation is intended to be pedagogical, focusing on providing a characterization of the three groups in terms of model structure constraints and inference properties. The monograph is supplemented with a software package that contains most of the models and examples described[1]. The material presented should be useful to both researchers wishing to learn about these models and researchers wishing to develop them further.

S. Chiappa. *Explicit-Duration Markov Switching Models*. Foundations and Trends® in Machine Learning, vol. 7, no. 6, pp. 803–886, 2014.

DOI: 10.1561/2200000054.

[1]More information about the package is available at www.nowpublishers.com.

Full text available at: http://dx.doi.org/10.1561/2200000054

1

Introduction

Markov switching models (MSMs) are probabilistic models that employ multiple sets of parameters to describe different dynamic regimes that a time series may exhibit at different periods of time. The switching mechanism between regimes is controlled by unobserved variables that form a first-order Markov chain.

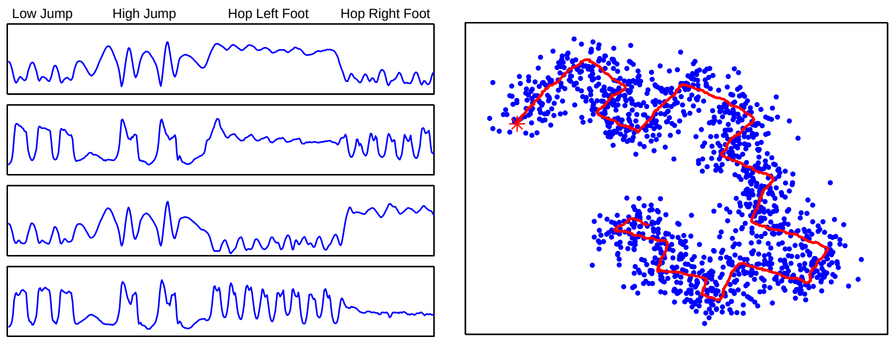
MSMs are commonly used for segmenting time series or to retrieve the hidden dynamics underlying noisy observations.

Consider, for example, the time series displayed in Figure 1.1(a), which corresponds to the measured leg positions of an individual performing repetitions of the actions low/high jumping and hopping on the left/right foot. A segmentation of the time series into the underlying actions could be obtained with a MSM in which each action forms a separate regime, *e.g.* by computing the regimes with highest posterior probabilities[1].

As another example, consider the time series displayed with dots in Figure 1.1(b), which corresponds to noisy observations of the positions of a two-wheeled robot moving in the two-dimensional space according to straight movements, left-wheel rotations and right-wheel rotations

[1]This example is discussed in detail in §3.5.3.

Full text available at: http://dx.doi.org/10.1561/2200000054



(a) (b)

Figure 1.1: (a): Body-marker recording of an individual performing repetitions of the actions low jumping up and down, high jumping up and down, hopping on the left foot and hopping on the right foot (CMU Graphics Lab Motion Capture Database). (b): Actual positions (continuous line) and measured positions (dots) of a two-wheeled robot moving in the two-dimensional space. The initial actual position is indicated with a star.

(the actual positions are displayed with a continuous line). Denoised estimates of the positions could be obtained with a MSM in which the robot movements are described with continuous unobserved variables and in which each type of movement forms a separate regime, *e.g.* by computing the posterior means of the continuous variables[2].

In standard MSMs, the regime variables implicitly define a geometric distribution on the time spent in each regime. In explicit-duration MSMs, this constraint is relaxed by using additional unobserved variables that allow to define duration distributions of any form. Explicit-duration variables also allow to impose complex dependence between the observations and to reset the dynamics to initial conditions.

Explicit-duration MSMs were first introduced in the speech community [Ferguson, 1980] and are mostly used to achieve more powerful modelling than standard MSMs through the specification of more accurate duration distributions and dependencies between the observations. In this case, the models are most commonly known with the names of

[2]This example is discussed in detail in §3.5 and in Appendix A.4.

Full text available at: http://dx.doi.org/10.1561/2200000054

hidden semi-Markov models or segment models. However, the possibility to reset the dynamics to initial conditions has recently led to the use of explicit-duration variables also for Bayesian approaches to abrupt-change detection, for identifying repetitions of patterns (such as, *e.g.*, the action repetitions underlying the time series in Figure 1.1(a)), and for performing/approximating inference[3] [Fearnhead, 2006, Fearnhead and Vasileiou, 2009, Chiappa and Peters, 2010, Bracegirdle and Barber, 2011]. In these cases, the models are most commonly known with the names of changepoint models or reset models.

Explicit-duration MSMs have been used in many application areas including speech analysis [Russell and Moore, 1985, Levinson, 1986, Rabiner, 1989, Gu et al., 1991, Gales and Young, 1993, Russell, 1993, Ostendorf et al., 1996, Moore and Savic, 2004, Liang et al., 2011], handwriting recognition [Chen et al., 1995], activity recognition [Yu and Kobayashi, 2003b, Huang et al., 2006, Oh et al., 2008, Chiappa and Peters, 2010], musical pattern recognition [Pikrakis et al., 2006], financial time series analysis [Bulla and Bulla, 2006], rainfall time series analysis [Sansom and Thomson, 2001], protein structure segmentation [Schmidler et al., 2000], gene finding [Winters-Hilt et al., 2010], DNA analysis [Barbu and Limnios, 2008, Fearnhead and Vasileiou, 2009], plant analysis [Guédon et al., 2001], MRI sequence analysis [Faisan et al., 2002], ECG segmentation [Hughes et al., 2004], and waveform modelling [Kim and Smyth, 2006]; see references in Yu [2010] for more examples.

Explicit-duration MSMs originated from the idea of explicitly modelling the duration distribution by defining a semi-Markov process on the regime variables, namely a process in which the trajectories are piecewise constant functions – with interval durations drawn from an explicitly defined duration distribution – and in which the variables at jump times form a Markov chain. The first and currently standard approach achieves that with variables indicating the interval duration, and derives inference recursions using only jump times [Rabiner, 1989, Gales and Young, 1993, Ostendorf et al., 1996, Yu, 2010]. To simplify the derivations of posterior distributions at times that are different

[3]By inference we mean the computation of posterior distributions, namely distributions of unobserved variables conditioned on the observations.

Full text available at: http://dx.doi.org/10.1561/2200000054

from jump times, Chiappa and Peters [2010] use count variables in addition to duration variables, such that the combined regime and count-duration variables form a first-order Markov chain. Other methods that explicitly model the duration distribution have been proposed with different goals and in different communities. These methods can all be viewed as different ways to define a first-order Markov chain on the combined regime and explicit-duration variables that induces a semi-Markov process on the regime variables.

In this monograph we provide a description of explicit-duration modelling that aims at elucidating the characteristics of the different approaches and at clarifying and unifying the literature. We identify three fundamentally different ways to define the first-order Markov chain on the combined regime and explicit-duration variables, which differ in encoding in the explicit-duration variables the location of (i) the preceding, (ii) the following, or (iii) both the preceding and following regime change or reset. We discuss each encoding in the context of MSMs of simple unobserved structure and of MSMs that contain extra unobserved variables related by first-order Markovian dependence. The models are described using the formalism of graphical models, which allows to graphically represent and assess statistical dependence, and therefore to easily describe the structure of complex models and derive inference routines.

The remainder of the manuscript is organized as follows. Chapter 2 contains some background material. We start with a general description of MSMs and by showing that the regime variables implicitly define a geometric duration distribution. In §2.1 we introduce the hidden Markov model, which represents the simplest MSM, and explain how to obtain a negative binomial duration distribution with regime copies. In §2.2 we introduce the framework of graphical models, and explain how to graphically assess statistical independence in a particular type of graphical models, called belief networks, that will be used for describing the models. In §2.2.1 we illustrate how belief networks can be used to easily derive the standard inference recursions of MSMs. In §2.3 we give a general explanation of the expectation maximization algorithm, which represents the most popular algorithm for parameter

Full text available at: http://dx.doi.org/10.1561/2200000054

learning in probabilistic models with unobserved variables. In Chapter 3 we describe the different approaches to explicit-duration modelling by categorizing them into three groups. The groups are introduced in §3.1, §3.2 and §3.3. In §3.4 we discuss in detail explicit-duration modelling in MSMs containing only regime variables, explicit-duration variables, and observations. In §3.5 we discuss in detail explicit-duration modelling in a popular MSM containing additional unobserved variables related by first-order Markovian dependence, namely the switching linear Gaussian state-space model, and discuss how the findings generalize to similar models with unobserved variables related by first-order Markovian dependence. The case of more complex unobserved structure is not considered. In §3.6 we describe approximation schemes to reduce the computational cost of inference. In Chapter 4 we summarize the most important points of our exposition and make some historical remarks.

Full text available at: http://dx.doi.org/10.1561/2200000054

References

R. P. Adams and D. J. C. MacKay. Bayesian online changepoint detection, 2007.

D. L. Alspach and H. W. Sorenson. Nonlinear Bayesian estimation using Gaussian sum approximations. *IEEE Transactions on Automatic Control*, 17:439–448, 1972.

D. Barber. Expectation correction for smoothing in switching linear Gaussian state space models. *Journal of Machine Learning Research*, 7:2515–2540, 2006.

D. Barber. *Bayesian Reasoning and Machine Learning*. Cambridge University Press, 2012.

D. Barber and A. T. Cemgil. Graphical models for time-series. *IEEE Signal Processing Magazine*, 27(6):18–28, 2010.

V. S. Barbu and N. Limnios. *Semi-Markov Chains and Hidden Semi-Markov Models toward Applications: Their Use in Reliability and DNA Analysis*. Springer, 2008.

C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

C. Bracegirdle. *Inference in Bayesian time-series models*. PhD thesis, University College London, London, UK, 2013.

C. Bracegirdle and D. Barber. Switch-reset models: Exact and approximate inference. In *Proceedings of The Fourteenth International Conference on Artificial Intelligence and Statistics*, volume 15, pages 190–198, 2011.

Full text available at: http://dx.doi.org/10.1561/2200000054

J. Bulla and I. Bulla. Stylized facts of financial time series and hidden semi-Markov models. *Computational Statistics and Data Analysis*, 51(4):2191–2209, 2006.

A. T. Cemgil, B. Kappen, and D. Barber. A generative model for music transcription. *IEEE Transactions on Audio, Speech Lang. Processing*, 14 (2):679–694, 2006.

M.-Y. Chen, A. Kundu, and S. N. Srihari. Variable duration hidden Markov model and morphological segmentation for handwritten word recognition. *IEEE Transactions on Image Processing*, 4(12):1675–1688, 1995.

S. Chiappa. *Analysis and Classification of EEG Signals using Probabilistic Models for Brain Computer Interfaces.* Ph.D. Thesis, EPFL, Lausanne, 2006.

S. Chiappa. A Bayesian approach to switching linear Gaussian state-space models for unsupervised time-series segmentation. In *Proceedings of Seventh International Conference on Machine Learning and Applications*, pages 3–9, 2008.

S. Chiappa and J. Peters. Movement extraction by detecting dynamics switches and repetitions. In *Advances in Neural Information Processing Systems 23*, pages 388–396, 2010.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B*, 39(1):1–38, 1977.

M. Dewar, C. Wiggins, and F. Wood. Inference in hidden Markov models with explicit state duration distributions. *IEEE Signal Processing Letters*, 19(4):235–238, 2012.

P. M. Djurić and J.-H. Chun. An MCMC sampling approach to estimation of nonstationary hidden Markov models. *IEEE Transactions on Signal Processing*, 50(5):1113–1123, 2002.

R. Durbin, S. Eddy, A. Krogh, and G. Mitchison. *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids.* Cambridge University Press, 1998.

I. A. Eckley, P. Fearnhead, and R. Killick. Analysis of changepoint models. In D. Barber, A. T. Cemgil, and S. Chiappa, editors, *Bayesian Time Series Models*, pages 205–224. Cambridge University Press, 2011.

S. Faisan, L. Thoraval, J.-P. Armspach, and F. Heitz. Hidden semi-Markov event sequence models: Application to brain functional MRI sequence analysis. In *International Conference on Image Processing*, volume 1, pages I–880–I–883, 2002.

Full text available at: http://dx.doi.org/10.1561/2200000054

P. Fearnhead. Exact and efficient Bayesian inference for multiple changepoint problems. *Statistics and Computing*, 16(2):203–213, 2006.

P. Fearnhead and Z. Liu. Online inference for multiple changepoint problems. *Journal of the Royal Statistical Society Series B*, 69(4):589–605, 2007.

P. Fearnhead and D. Vasileiou. Bayesian analysis of isochores. *Journal of the American Statistical Association*, 104(485):132–141, 2009.

J. D. Ferguson. Variable duration models for speech. In *Symposium on the Application of Hidden Markov Models to Text and Speech*, pages 143–179, 1980.

M. Gales and S. Young. The theory of segmental hidden Markov models. Technical report, Cambridge University, 1993. Technical Report CUED/F-INFENG/TR 133.

M. S. Grewal and A. P. Andrews. *Kalman Filtering: Theory and Practice*. Prentice-Hall, 1993.

H.-Y. Gu, C.-Y. Tseng, and L.-S. Lee. Isolated-utterance speech recognition using hidden Markov models with bounded state durations. *IEEE Transactions on Signal Processing*, 39(8):1743–1752, 1991.

Y. Guédon, D. Barthélémy, Y. Caraglio, and E. Costes. Pattern analysis in branching and axillary flowering sequences. *Journal of Theoretical Biology*, 212(4):481–520, 2001.

J. D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357–384, 1989.

J. D. Hamilton. Analysis of time series subject to changes in regime. *Journal of Econometrics*, 45(1-2):39–70, 1990.

J. D. Hamilton. Estimation, inference, and forecasting of time series subject to changes in regime. *Handbook of Statistics*, 11:231–260, 1993.

T. Huang, F. Li, S. Zhan, and J. Min. Variable duration motion texture for human motion modeling. In *Proceedings of the 9th Pacific Rim International Conference on Artificial Intelligence*, pages 603–612, 2006.

N. P. Hughes, S. J. Roberts, and L. Tarassenko. Semi-supervised learning of probabilistic models for ECG segmentation. In *Conference Proceedings of the IEEE Engineering in Medicine and Biology Society*, volume 1, pages 434–437, 2004.

Z. Jiang. *Hidden Markov Model with Binned Duration and Its Application*. Ph.D. Thesis, University of New Orleans, 2010.

Full text available at: http://dx.doi.org/10.1561/2200000054

S. Kim and P. Smyth. Segmental hidden Markov models with random effects for waveform modeling. *Journal of Machine Learning Research*, 7:945–969, 2006.

D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, 2009.

S. E. Levinson. Continuously variable duration hidden Markov models for speech analysis. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 11, pages 1241–1244, 1986.

Y. Liang, X. Liu, Y. Lou, and B. Shan. An improved noise-robust voice activity detector based on hidden semi-Markov models. *Pattern Recognition Letters*, 32(7):1044–1053, 2011.

J. S. Liu, R. Chen, and W. H. Wong. Rejection control and sequential importance sampling. *Journal of the American Statistical Association*, 93(443): 1022–1031, 1998.

G. J. McLachlan and T. Krishnan. *The EM Algorithm and Extensions*. John Wiley & Sons, 2008.

B. Mesot and D. Barber. Switching linear dynamical systems for noise robust speech recognition. *IEEE Transactions of Audio, Speech and Language Processing*, 15(6):1850–1858, 2007.

C. D. Mitchell, M. P. Harper, and L. H. Jamieson. On the complexity of explicit duration HMMs. *IEEE Transactions on Speech and Audio Processing*, 3(3):213–217, 1995.

M. D. Moore and M. I. Savic. Speech reconstruction using a generalized HSMM (GHSMM). *Digital Signal Processing*, 14(1):37–53, 2004.

K. P. Murphy. Hidden semi-Markov models (HSMMs), 2002. Informal Notes.

K. P. Murphy. *Machine Learning: a Probabilistic Perspective*. MIT Press, 2012.

S. M. Oh, J. M. Rehg, T. Balch, and F. Dellaert. Learning and inferring motion patterns using parametric segmental switching linear dynamic systems. *International Journal of Computer Vision*, 77:103–124, 2008.

M. Ostendorf, V. V. Digalakis, and O. A. Kimball. From HMM's to segment models: a unified view of stochastic modeling for speech recognition. *IEEE Transactions on Speech and Audio Processing*, 4(5):360–378, 1996.

V. Pavlovic, J. M. Rehg, and J. MacCormick. Learning switching linear models of human motion. In *Advances in Neural Information Processing Systems 13*, pages 981–987, 2001.

Full text available at: http://dx.doi.org/10.1561/2200000054

J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.

A. Pikrakis, S. Theodoridis, and D. Kamarotos. Classification of musical patterns using variable duration hidden Markov models. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(5):1795–1807, 2006.

J. A. Quinn, C. K.I. Williams, and N. McIntosh. Factorial switching linear dynamical systems applied to physiological condition monitoring. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(9):1537–1551, 2009.

L. R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. In *Proceedings of the IEEE*, volume 77, pages 257–286, 1989.

H. E. Rauch, F. Tung, and C. T. Striebel. Maximum likelihood estimates of linear dynamic systems. *AIAA Journal*, 3(8):1445–1450, 1965.

M. Russell. A segmental HMM for speech pattern matching. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 499–502, 1993.

M. J. Russell and R. K. Moore. Explicit modelling of state occupancy in hidden Markov models for automatic speech recognition. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 10, pages 5–8, 1985.

J. Sansom and P. Thomson. Fitting hidden semi-Markov models to breakpoint rainfall data. *Journal of Applied Probability*, 38A:142–157, 2001.

S. Särkkä. Unscented Rauch-Tung-Striebel smoother. *IEEE Transactions on Automatic Control*, 53(3):845–849, 2008.

S. C. Schmidler, J.S. Liu, and D.L. Brutlag. Bayesian segmentation of protein secondary structure. *Journal of Computational Biology*, 7(1/2):233–248, 2000.

M. Stanke and S. Waack. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics*, 19(2):ii215–ii225, 2003.

C. M. Wang. Location estimation and uncertainty analysis for mobile robots. In I. J. Cox and G. T. Wilfong, editors, *Autonomous Robot Vehicles*, pages 90–95. Springer-Verlag, 1990.

S. Winters-Hilt, Z. Jiang, and C. Baribault. Hidden Markov model with duration side information for novel HMMD derivation, with application to eukaryotic gene finding. *EURASIP Journal on Advances in Signal Processing*, 2010.

Full text available at: http://dx.doi.org/10.1561/2200000054

S.-Z. Yu. Hidden semi-Markov models. *Artificial Intelligence*, 174(2):215–243, 2010.

S.-Z. Yu and H. Kobayashi. An efficient forward-backward algorithm for an explicit-duration hidden Markov model. *IEEE Signal Processing Letters*, 10(1):11–14, 2003a.

S.-Z. Yu and H. Kobayashi. A hidden semi-Markov model with missing data and multiple observation sequences for mobility tracking. *Signal Processing*, 83(2):235–250, 2003b.

O. Zoeter. *Monitoring Non-Linear and Switching Dynamical Systems*. Ph.D. Thesis, Radboud University, Nijmegen, 2005.