

Supplementary material for “A primer on coupled state-switching models for multiple interacting time series”

Jennifer Pohle¹, Roland Langrock¹, Mihaela van der Schaar^{2,3,4},
Ruth King^{3,5}, Frants Havmand Jensen⁶

¹ Bielefeld University, Germany

² University of Cambridge, UK

³ Alan Turing Institute, London, UK

⁴ University of California, Los Angeles, USA

⁵ University of Edinburgh, UK

⁶ Woods Hole Oceanographic Institution, Massachusetts, USA

1 Model comparison in the two case studies

For both case studies presented in the main manuscript, here we additionally provide comparisons of five different specifications of the models’ state processes:

- (1) a model assuming independent (non-coupled) state processes for the different individuals (in the dolphin example) or vital signs (in the electronic health records example);
- (2) a (non-coupled) model assuming a single state sequence shared by all individuals/vital signs;
- (3) coupled state processes with contemporaneous conditional independence assumption for the individual state variables (see Section 2.2.2 in the main manuscript);
- (4) coupled state processes using mixture transition probabilities as proposed by [Saul and Jordan \(1999\)](#), see Section 2.2.3 in the main manuscript), and
- (5) coupled state processes with a Cartesian product state space and no further restrictions on the transition probabilities (see Section 2.2.1 in the main manuscript).

In each case, the state-dependent process is modelled in the same way as described in the main manuscript (see Section 4.1 for the dolphin movement data, and Section 4.2 for the electronic health record (EHR) data), using $N = 3$ states for each individual/vital sign, respectively. All models were fitted using numerical maximum likelihood with the optimisation routine `nlm` in R ([R Core Team, 2018](#)). The model comparison presented below is based on the Akaike information criterion (AIC, [Akaike, 1973](#)).

	no. parameters	log-likelihood	AIC	Δ AIC
(1) ind. HMMs	24	30857	-61667	4251
(2) multi. HMMs	18	32902	-65768	149
(3) CHMM (cond. ind.)	48	31678	-63260	2658
(4) CHMM (mixture)	38	31546	-63015	2902
(5) CHMM (Cartesian product)	84	33043	-65918	0

Table 1: AIC values obtained for the different models fitted to the dolphin tortuosity data.

	no. parameters	log-likelihood	AIC	Δ AIC
(1) ind. MSR	54	-1163008	2326124	1368
(2) multi. MSR	42	-1246712	2493508	168752
(3) CMSR (cond. ind.)	198	-1162540	2325477	721
(4) CMSR (mixture)	96	-1162819	2325830	1074
(5) CMSR (Cartesian product)	738	-1161640	2324756	0

Table 2: AIC values obtained for the different models fitted to the EHR data.

1.1 Movements of dolphin mother and calf

Table 1 displays the maximum log-likelihood, AIC and Δ AIC values for the five different HMMs — coupled and non-coupled — fitted to the dolphin tortuosity data. According to this criterion, the Cartesian product CHMM (5) performed best. The model with two independent HMMs for the two individuals (1) performed worst (Δ AIC = 4251), which highlights the importance of taking interactions into account, especially in this example where the correlation between the individuals' behaviour is very strong. Given the high synchronisation of the two dolphins, it comes as no surprise that the multivariate HMM (2) performed second best (Δ AIC = 149). However, as the behavioural modes of the two individuals are not perfectly synchronous, this model is inferior to the Cartesian product CHMM, which allows for occasional individual differences in otherwise mainly synchronous state sequences. The sparser CHMM formulations, (3) and (4), performed worse due to their inability to reflect that state switches of the two individuals tend to occur simultaneously. In this example, these model formulations also showed some numerical instabilities in the optimisation (namely boundary estimates and strong sensitivity to the choice of initial parameter values used in the optimisation).

1.2 Electronic health record data

Table 2 displays the results obtained for the five different MSR models — coupled and non-coupled — fitted to the EHR data. Despite its high number of parameters, the Cartesian product model (5) is again favoured by the AIC. As the level of synchronicity of the three vital signs is much lower than that of the individuals in the dolphin example, the multivariate MSR model (2) performed by far the worst (Δ AIC = 168752). Particularly, its unrealistic

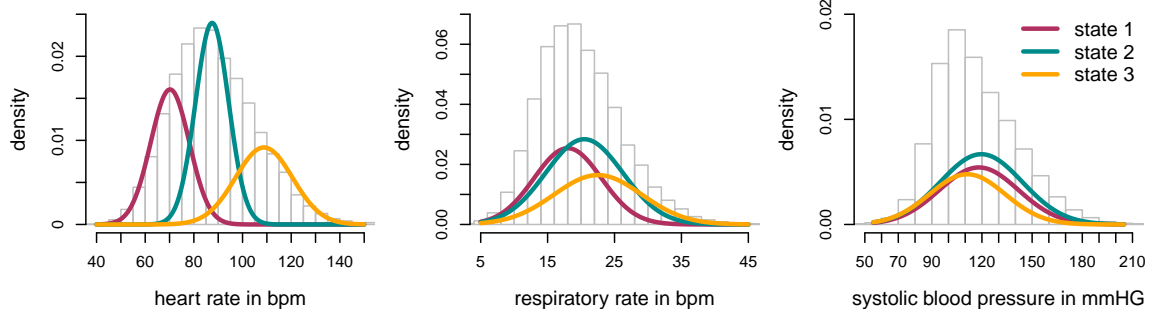


Figure 1: State-dependent distributions for heart rate, respiratory rate and systolic blood pressure, respectively, for 62-year-old males, as estimated by the multivariate MSR model.

assumption of perfect synchronicity has severe consequences in this example, as can be seen in Figure 1, which shows the state-dependent normal distributions for 62-year-old males estimated within this model formulation: The inference regarding the state dynamics is very much dominated by a single variable, heart rate, such that the model then cannot adequately distinguish different states underlying the respiratory rate and blood pressure anymore (cf. the simulation study in Section 3.1 of the main manuscript). Again due to the lower level of synchronicity, modelling the observed variables using separate independent MSR models (1) here is not quite as bad as in the dolphin example. Models (3) and (4), with sparser formulations for the coupled state processes, here also performed reasonably well — in fact these model formulations would be favoured by model selection criteria involving higher penalty terms on model complexity. However, the CMSR model with mixture transition probabilities again showed some numerical instabilities in the optimisation.

1.3 Some concluding remarks regarding model specification

Overall, the two case studies both illustrate that accounting for interaction between state processes can be important. We also found that the performance of natural candidate models, which to varying degrees can capture synchronous evolution of state sequences, very much depends on the level of synchronicity in the real processes to be modelled. Especially in the EHR example, it becomes evident that the most flexible (i.e. least restrictive) model formulation accounting for such interaction, the Cartesian product coupled model, has one major disadvantage, namely the sharp rise in model complexity when either the number of states, N , or the number of time series to be modelled, M , increases. While the AIC in the given examples still favoured this model, this is possible only due to the associated rather small penalisation of model complexity. Irrespective of the model selection criterion applied, in cases such as the EHR example, where the Cartesian product formulation yields the rather worrying number of 738 model parameters, it is desirable to somehow reduce model complexity *a priori* using expert knowledge of the system to be modelled. For example, information could be incorporated on which types of transitions are impossible

(thus fixing corresponding transition probabilities at zero) or which types of transitions can be reasonably grouped together (thus assuming blocks of shared parameters within the transition probability matrix), such that the number of free parameters is reduced accordingly (see, for example, [Sherlock *et al.*, 2013](#); [Touloupou *et al.*, 2020](#)).

2 Example of Viterbi decoding for the EHR data

To illustrate how the Cartesian product CHMM allocates the vital signs to distinct underlying states, Figure 2 displays an example time series for heart rate, respiratory rate and systolic blood pressure, respectively, colour-coded according to the Viterbi-decoded states. Overall, the Viterbi decoding produces reasonable state classifications. For example, they reflect the occasional phases with increased heart rates, which do not necessarily coincide with state switches in the other two vital signs.

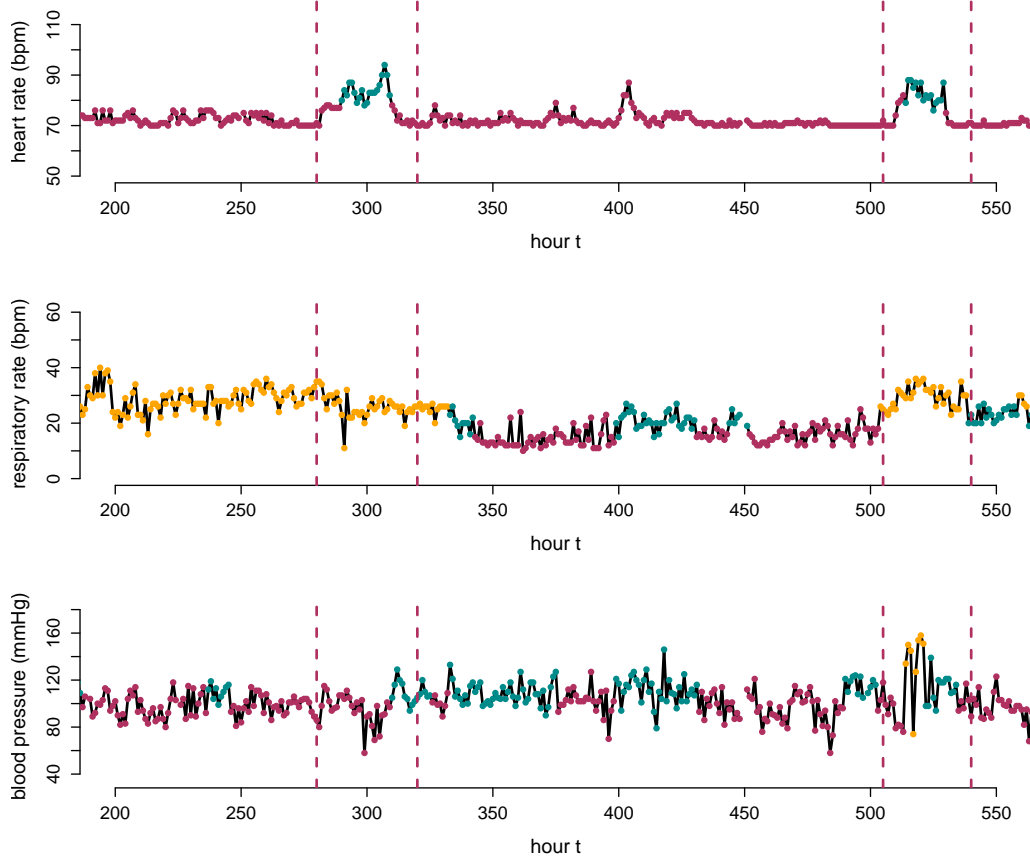


Figure 2: Example time series for heart rate, respiratory rate and systolic blood pressure, respectively, colour-coded according to the Viterbi-decoded CMSR states. The vertical red lines highlight intervals with an elevated heart rate that is not similarly reflected in the other two variables. The decoded states 1, 2 and 3 are indicated by the colours red, green and orange, respectively.

References

- Akaike, H. (1973). Information theory as an extension of the maximum likelihood theory. In Petrov, B.N. & Csaki, F., eds. *Proceedings of the Second International Symposium on Information Theory*. Akademiai Kiado, Budapest, 267–281.
- R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>.
- Saul, L.K. and Jordan, M.I. (1999). Mixed memory Markov models: Decomposing complex stochastic processes as mixtures of simpler ones. *Machine Learning*, **37**, 75–87.

- Sherlock, C., Xifara, T., Telfer, S. and Begon, M. (2013). A coupled hidden Markov model for disease interactions. *Journal of the Royal Statistical Society, Series C*, **62**, 609–627.
- Touloupou, P., Finkenstädt, B. and Spencer, S.E.F. (2020). Scalable Bayesian inference for coupled hidden Markov and semi-Markov models. *Journal of Computational and Graphical Statistics*, **29**, 238–249.