A parametric hidden semi-Markov model for toroidal time-series

Francesco Lagona[1], *, Marco Mingione[2]

[1] *Dept. of Political Sciences, University of Roma Tre, via G. Chiabrera 199 00145 Rome Italy; francesco.lagona@uniroma3.it*
[2] *Dept. of Political Sciences, University of Roma Tre, via G. Chiabrera 199 00145 Rome Italy; marco.mingione@uniroma3.it*
**Corresponding author*

Abstract. *A hidden semi-Markov model is proposed for segmenting toroidal time series according to a finite number of latent regimes, associated with torodidal densities. Data are segmented by integrating multiple sources of information: circular correlation, temporal auto-correlation, the time spent under a specific regime and the chances of regime switching. The proposal is motivated by a case study of wave and wind directions in the Adriatic sea.*

Keywords. *Hidden semi-Markov model; Toroidal data; Model-based classification; wave; wind.*

---

1 Introduction

Bivariate sequences of angles are often referred to as toroidal time series, because the pair of two angles can be represented as a point on a torus. Examples include time series of wind and wave directions [5], time series of wind mean directions and directions of the maximum gust observed each day [2] and time series of turning angles in studies of animal movement [7]. The analysis of toroidal time series is complicated by the difficulties in modeling the dependence between angular measurements over time [6]. An additional complication is given by the multimodality of the marginal distribution of the data, because environmental toroidal data are observed under time-varying heterogeneous conditions.

This paper introduces a toroidal hidden semi-Markov model (HSMM) that simultaneously accounts for dependence across circular measurements, temporal autocorrelation, multimodality and latent time-varying heterogeneity. Under this model, the distribution of toroidal data is approximated by a mixture of toroidal densities, whose parameters depend on the evolution of a latent semi-Markov process. While the toroidal density accommodates dependence between two circular variables, a mixture of toroidal densities allows for multimodality and, finally, a latent semi- Markov process accounts for temporal correlation and, simultaneously, for timevarying heterogeneity.

Our proposal extends previous approaches that are based on toroidal hidden Markov models [1, 4]. Under a toroidal hidden Markov model, the data are approximated by a mixture of toroidal densities, whose parameters depend on the evolution of a latent, first-order Markov chain with a finite number of states. The sojourn times of each state of a Markov chain are distributed according to a geometric distribution. Hence the most likely dwell time for every state of a hidden Markov model with underlying first-order Markov chain is 1. Our proposal relaxes this restrictive assumption by replacing the latent

Markov chain with a latent semi-Markov model, allowing for not necessarily geometrically distributed sojourn times that are modelled by state-specific parametric hazard functions.

2 A toroidal hidden semi-Markov model

Let $\boldsymbol{y} = (\boldsymbol{y}_t, t = 1, \ldots T)$ be a bivariate time series, where $\boldsymbol{y}_t = (y_{t1}, y_{t2})$ is a vector of two circular observations. In addition, let $\boldsymbol{u} = (\boldsymbol{u}_t, t = 1, \ldots T)$ be a sequence of latent multinomial random variables $\boldsymbol{u}_t = (u_{t1} \ldots u_{tK})$ with one trial and K classes (or states), whose binary components represent class membership at time t. Our proposal is a hierarchical model where the joint distribution of the time series is obtained by

$$f(\boldsymbol{y}) = \sum_{\boldsymbol{u}} f(\boldsymbol{y} \mid \boldsymbol{u}) p(\boldsymbol{u}).$$

The joint distribution $p(\boldsymbol{u})$ of the latent process is described by extending the notion of a Markov chain.

If $\mathbf{u}$ is a Markov chain, then $p(\boldsymbol{u})$ is fully known up to a vector of K initial probabilities $\pi_k = P(u_{1k} = 1), k = 1, \ldots, K, \sum_k \pi_k = 1$, and a $K \times K$ matrix of transition probabilities

$$\begin{pmatrix} \pi_{11} & \pi_{12} & \ldots & \pi_{1K} \\ \pi_{21} & \pi_{22} & \ldots & \pi_{2K} \\ \ldots & \ldots & \ldots & \ldots \\ \pi_{K1} & \pi_{K2} & \ldots & \pi_{KK} \end{pmatrix}$$

where $\pi_{kk'} = P(u_{tk'} = 1 | u_{t-1,k} = 1), k, k' = 1, \ldots, K, \sum_{k'} \pi_{kk'} = 1$. Formally,

$$p(\mathbf{u}) = \prod_{k=1}^{K} \pi_k^{u_{1k}} \prod_{t=2}^{T} \prod_{k=1}^{K} \prod_{k'=1}^{K} \pi_{kk'}^{u_{t-1,k} u_{tk'}}. \tag{1}$$

An alternative parametrization of the transition probabilities matrix is given by

$$\begin{pmatrix} \pi_{11} & \pi_{12} & \ldots & \pi_{1K} \\ \pi_{21} & \pi_{22} & \ldots & \pi_{2K} \\ \ldots & \ldots & \ldots & \ldots \\ \pi_{K1} & \pi_{K2} & \ldots & \pi_{KK} \end{pmatrix} = \begin{pmatrix} 1 - p_1 & p_1 \omega_{12} & \ldots & p_1 \omega_{1K} \\ p_2 \omega_{21} & 1 - p_2 & \ldots & p_2 \omega_{2K} \\ \ldots & \ldots & \ldots & \ldots \\ p_K \omega_{K1} & p_K \omega_{K2} & \ldots & 1 - p_K \end{pmatrix}$$

where $p_k = \sum_{k' \neq k} \pi_{kk'}$ is the probability of a transition from k to a different state and $\omega_{kk'}$ is the conditional probability of a transition to state $k' \neq k$, given a transition from state k. Under this setting, if the process is in state k, the time T_k up to a transition to a different state is geometric

$$P(T_k = t) = p_k (1 - p_k)^{t-1}. \tag{2}$$

More generally, let $S_k(t) = P(T_k > t) = \exp\left(-\int_0^t h(\tau) d\tau\right)$ be the survival funciton of T_k, where $h(t)$ is the associated hazard function. Then

$$p_k(t) = P(T_k \leq t + 1 \mid T_k > t) \approx 1 - \exp(-h(t + 0.5)),$$

is the conditional probability of a transition at time $t + 1$, given that the process remained in state k during a period of length t. Then

$$P(T_k = t) = p_k(t) \prod_{\tau=1}^{t-1} (1 - p_k(\tau)). \tag{3}$$

When the hazard $h_k(t)$ is time-constant, then (3) reduces to (2). Alternatively, (3) can be aproximated by

$$P(T_k = t) = p_k(m)(1-p_k(m))^{t-m} \prod_{\tau=1}^{m-1}(1-p_k(\tau)) = P(T_k = m)(1-p_k(m))^{t-m}. \tag{4}$$

Parametric hazard functions can be borrowed from the survival analysis literature and some of them are conveniently associated to a link function g that trasforms $p_k(t)$ in a linear function of time, say

$$g(p_k(t)) = \beta_{0k} + \beta_{1k}t.$$

For example, under a Gompertz hazard function, $h(t) = \alpha e^{\beta t}$, the complementary log-log transformation maps $p_k(t)$ to a linear function of time, namely

$$\log(-\log(1-p_k(t))) = \log\alpha_k + \beta_k t.$$

Under this setting, the state of the process at time t can be described by a $K \times m$ matrix $\mathbf{U}_t = (u_{tki}, k = 1\ldots K, i = 1\ldots m)$ where $u_{tki} = 1$ if and only if the process is in state k since time $t-i$. This suggests to augment the transition probabilites matrix by a $(m \times K) \times (m \times K)$ blocks matrix with $m \times m$ blocks

$$\boldsymbol{\Gamma} = (\boldsymbol{\Gamma}_{ks}, k, s = 1\ldots K),$$

where the diagonal blocks are given by

$$\Gamma_{kk}(\boldsymbol{\alpha}_k) = \begin{pmatrix} 0 & 1-p_k(1) & 0 & \ldots & 0 \\ 0 & 0 & 1-p_k(2) & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & 1-p_k(m-1) \\ 0 & 0 & 0 & \ldots & 1-p_k(m) \end{pmatrix}$$

and the off-diagonal blocks are given by

$$\Gamma_{ks}(\boldsymbol{\alpha}, \boldsymbol{\omega}_h^{\mathsf{T}}) = \begin{pmatrix} \omega_{ks}p_k(1) & 0 & 0 & \ldots & 0 \\ \omega_{ks}p_k(2) & 0 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \omega_{ks}p_k(m-1) & 0 & 0 & \ldots & 0 \\ \omega_{ks}p_k(m) & 0 & 0 & \ldots & 0 \end{pmatrix}.$$

Such augmentation allows to approximate the joint distribution of the semi-Markov process (with the desidered degree of accuracy) by

$$p(\mathbf{u}) = \prod_{k=1}^{K} \pi_k^{u_{0k}} \prod_{t=1}^{T}\prod_{h=1}^{K}\prod_{k=1}^{K}\prod_{i=1}^{m}\prod_{j=1}^{m} \gamma_{hikj}^{u_{t-1,hi}u_{tkj}}, \tag{5}$$

where γ_{hikj} is the (i,j)-th entry of the (h,k)th block of matrix Γ.

Our proposal is completed by a conditional independence assumption on the observation process. Precisely,

$$f(\boldsymbol{y} \mid \boldsymbol{u}) = \prod_{t=0}^{T}\prod_{k=1}^{K}\prod_{i=1}^{m} f(\mathbf{y}_t; \theta_k)^{u_{tki}}, \tag{6}$$

where $\theta_1, \ldots \theta_K$ is a sequence of unknown parameters. Parametric toroidal densities can be borrowed by the proposals available in the directional statistics literature. A convenient specification is for example the bivariate wrapped Cauchy distribution [3]. It is unimodal, pointwise symmetric and has a closed-form expression for the conditional distribution. A single dependence parameter controls the relationship between the two component circular variables, ranging from independence to perfect correlation. The remaining four parameters respectively indicate the two marginal means and concentrations.

3 Maximum likelihood estimation

An EM algorithm can be exploited to maximize the likelihood function of the prposed model. It is based on the following complete-data log-likelihood function

$$\begin{aligned}\log L_{\text{comp}} &= \sum_{k=1}^{K} u_{0k} \log \pi_k + \sum_{t=1}^{T}\sum_{h=1}^{K}\sum_{k=1}^{K}\sum_{i=1}^{m}\sum_{j=1}^{m} u_{t-1,hi} u_{tkj} \log \gamma_{hikj} \\ &+ \sum_{t=1}^{T}\sum_{k=1}^{K}\sum_{i=1}^{m} u_{tki} \log f(\boldsymbol{y}_t;\boldsymbol{\theta}_k). \end{aligned} \tag{7}$$

The algorithm is iterated by alternating an expectation (E) and a maximization (M) step. Given the parameter estimates, obtained at the end of the s-th iteration, the $(s+1)$-th iteration is initialized by the E-step, which evaluates the expected value of the complete data log-likelihood (7) with respect to the conditional distribution of the missing values u_{tki} given the observed data. The E step reduces to the computation of the univariate posterior probabilities of each latent state at time t, $\hat{\pi}_{tki} = P(u_{tki} = 1 \mid \boldsymbol{y}) i = 1\ldots m, k = 1\ldots K, t = 1\ldots T$, and the computation of the bivariate posterior probabilities of each pair of states in two adjacent times, say $\hat{\pi}_{t-1,t,hikj} = P(u_{t-1,hi} = 1, u_{tkj} = 1 \mid \boldsymbol{y}) \quad i,j = 1\ldots m, h,k = 1\ldots K, t = 2\ldots T$. The task of computing these posterior probabilities is generally referred to as the HMM-smoothing numerical issue and it is typically solved by specifying the posterior probabilities in terms of suitably normalized functions, which can be computed recursively, avoiding unpractical summations over the state space of latent Markov chain and numerical under- and over-flows.

The M-step of the algorithm updates the parameter estimates, by maximizing the expected value of the complete data log-likelihood (7), obtained from the previous E step. This expected value is the sum of functions that depend on independent sets of parameters and can therefore be maximized separately. Specifically, the expected log-likelihood function

$$\begin{aligned} Q &= \sum_{k=1}^{K} \hat{\pi}_{0k} \log \pi_k + \sum_{t=1}^{T}\sum_{h=1}^{K}\sum_{k=1}^{K}\sum_{i=1}^{m}\sum_{j=1}^{m} \hat{\pi}_{t-1,hi,tkj} \log \gamma_{hikj} \\ &+ \sum_{t=1}^{T}\sum_{k=1}^{K}\sum_{i=1}^{m} \hat{\pi}_{tki} \log f(\boldsymbol{y}_t;\boldsymbol{\theta}_k). \end{aligned} \tag{8}$$

Function

$$\sum_{t=1}^{T}\sum_{h=1}^{K}\sum_{k=1}^{K}\sum_{i=1}^{m}\sum_{j=1}^{m} \hat{\pi}_{t-1,hi,tkj} \log \gamma_{hikj}$$

can be split as the sum of two components:

$$Q_1 = \sum_{t=1}^{T}\sum_{h=1}^{K}\sum_{h\neq k}\sum_{i=1}^{m}\sum_{j=1}^{m} \hat{\pi}_{t-1,hi,tkj} \log \omega_{hk} \tag{9}$$

and

$$\begin{aligned} Q_2 &= \sum_{t=1}^{T}\sum_{k=1}^{K}\sum_{h\neq k}\sum_{i=1}^{m}\sum_{j=1}^{m} \hat{\pi}_{t-1,ki,thj} \log p_k(i) \\ &+ \sum_{t=1}^{T}\sum_{k=1}^{K}\sum_{i=1}^{m}\sum_{j=1}^{m} \hat{\pi}_{t-1,ki,tkj} \log(1-p_k(i)) \end{aligned} \tag{10}$$

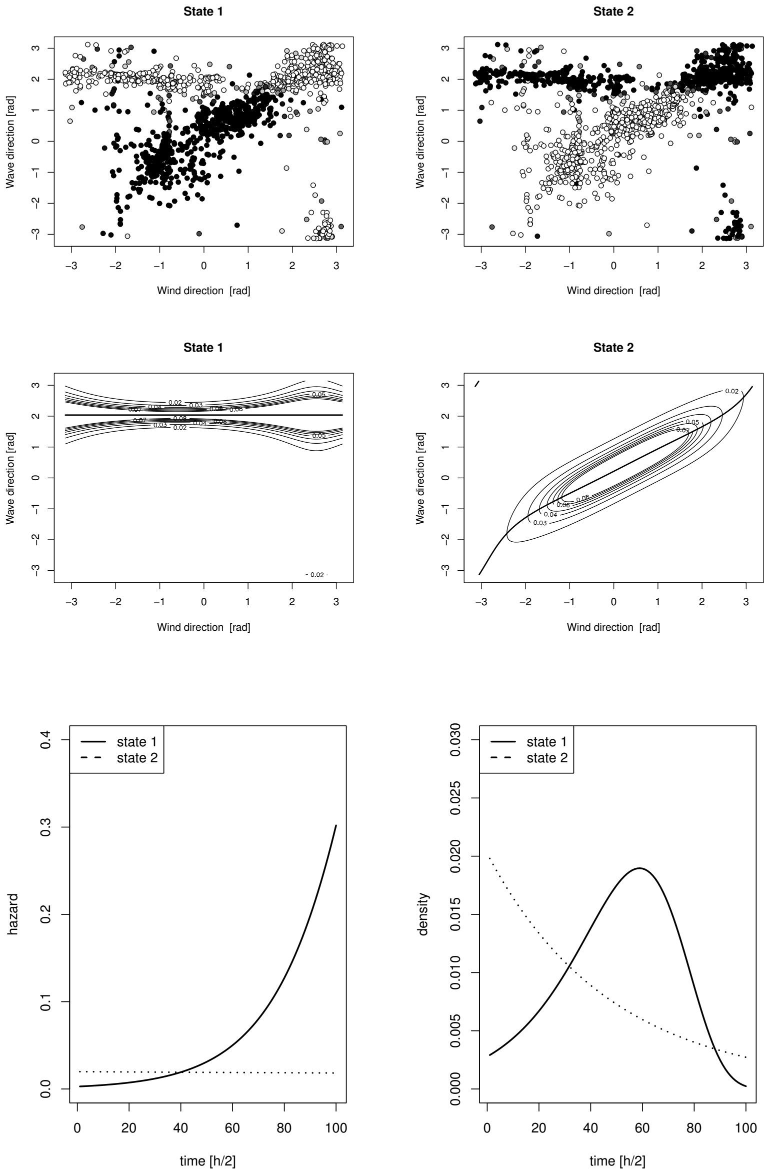

Figure 1: Top: data clustered within state 1 (left) and state 2 (right); grey levels are proportional to class membership probabilities). Middle: state-specific toroidal densities, with estimated toroidal regression lines. Bottom: state-specific hazards (left) and densities (right) of the dwell time

Function Q_1 is maximized by the estimated transition probabilities

$$\hat{\omega}_{hk} = \frac{\sum_{t=1}^{T} \sum_{i=1}^{m} \sum_{j=1}^{m} \hat{\pi}_{t-1,hi,tkj}}{\sum_{k \neq h} \sum_{t=1}^{T} \sum_{i=1}^{m} \sum_{j=1}^{m} \hat{\pi}_{t-1,hi,tkj}}.$$

Function Q_2 can be instead maximized by fitting a binomial regression model with weights and an appropriate link function.

4 Results

Figure 1 shows the results obtained on a time series of $T = 1326$ semi-hourly wind and wave directions, taken in wintertime by the buoy of Ancona, which is located in the Adriatic Sea at about 30 km from the coast. A 2-state hidden semi-Markov model with Gompertz dwell times has been used to segment the data. The model successfully segments the observations according to two clusters, and offers a clear-cut indication of the distribution of the data under each regime. Under state 1, wind and wave directions are essentially independent. Under state 2, winds appear well syncronized with waves. Interestingly, while the first state is driven by a constant hazard function, associated with an exponential dwell time distribution, the second state is driven by an exponential hazard, associates with a bell-shaped dwell time distribution.

In summary, the proposed model describes the plasticity of the wind-wave interaction in the Adriatic Sea, indicating that the joint distribution of wind and wave data changes under different environmental regimes. Regime switching changes not only the modal directions and concentrations around these modes but also, and more interestingly, the correlation structure of the data. As a result, the model indicates that the wind direction is an accurate predictor of the wave direction only under a specific, and persistent, regime (State 2). In summary, wind directions should not be used to predict wave directions, without accounting for the latent, environmental heterogeneity of the data under study.

References

[1] Bulla J, Lagona F, Maruotti A, Picone M (2012) A multivariate hidden Markov model for the identification of sea regimes from incomplete skewed and circular time series, *Journal of Agricultural, Biological, and Environmental Statistics*, **17**, 544–567

[2] Coles S (1998) Inference for circular distributions and processes. *Statistics and Computing*,**8**, 105–113.

[3] Kato S. and Pewsey A. (2015) A Mobius transformation-induced distribution on the torus. *Biometrika*, **102**, 359–370

[4] Lagona F, Picone M (2013) Maximum likelihood estimation of bivariate circular hidden Markov models from incomplete data, *Journal of Statistical Computation and Simulation*, **83**, 1223–1237.

[5] Lagona F, Picone M, Maruotti A, Cosoli S (2014) A hidden Markov approach to the analysis of space-time environmental data with linear and circular components, *Stochastic Environmental Research and Risk Assessment* **29**, 397–409

[6] Lagona F (2018) Correlated cylindrical data. In: C. Ley and T. Verdebout (Eds) *Applied Directional Statistics: Modern Methods and Case Studies*, Chapman & Hall/CRC: New York, 45–59

[7] Mastrantonio G (2018) The joint projected normal and skew-normal: A distribution for polycylindrical data, *Journal of Multivariate Analysis*, **165**, 14–26