

Bayesian Modelling of the Temporal Aspects of Smart Home Activity with Circular Statistics

Tom Diethe^(✉), Niall Twomey, and Peter Flach

Intelligent Systems Laboratory, University of Bristol, Bristol, UK
{tom.diethe,niall.twomey,peter.flach}@bristol.ac.uk

Abstract. Typically, when analysing patterns of activity in a smart home environment, the daily patterns of activity are either ignored completely or summarised into a high-level “hour-of-day” feature that is then combined with sensor activities. However, when summarising the temporal nature of an activity into a coarse feature such as this, not only is information lost after discretisation, but also the strength of the periodicity of the action is ignored. We propose to model the temporal nature of activities using circular statistics, and in particular by performing Bayesian inference with Wrapped Normal (\mathcal{WN}) and \mathcal{WN} Mixture (\mathcal{WNM}) models. We firstly demonstrate the accuracy of inference on toy data using both Gibbs sampling and Expectation Propagation (EP), and then show the results of the inference on publicly available smart-home data. Such models can be useful for analysis or prediction in their own right, or can be readily combined with larger models incorporating multiple modalities of sensor activity.

1 Introduction

One of the central hypotheses of a “smart home” is that a number of different sensor technologies may be combined to build accurate models of the Activities of Daily Living (ADL) of its residents. These models can then be used to make informed decisions relating to medical or health-care issues. For example, such models could help by predicting falls, detecting strokes, analysing eating behaviour, tracking whether people are taking prescribed medication, or detecting periods of depression and anxiety. Since 2007, the Centre for Advanced Studies in Adaptive Systems (CASAS) research group has been collecting data from homes with various different sensor layouts and differing numbers of residents (see *e.g.* [2]).

In most of the approaches taken to date [6, 7, 13], classifiers are learnt which put weights over individual sensors, and then take linear combinations of these weights to produce a decision function for the set of active sensors at any given time. In addition, an extra “hour-of-day” feature is often added, which in some sense attempts to capture the periodic nature of many of the activities under examination. However this can produce undesirable effects, since this is a rather coarse discretisation. This in turn can result in border effects, such as activities that are short-lived but often span an hour boundary.

We propose instead that it is more satisfactory to take a model-based approach, in which the temporally periodic nature of the activities (*i.e.* circadian or diurnal rhythms) is taken directly into account. A natural framework for this is the area of “circular” statistics [5, 9, 18], where univariate data is defined on an angular scale, typically the (unit) circle.

In addition we suggest that, rather than using frequentist methods to fit circular distributions to the data, a full Bayesian approach would be advantageous in this setting. To begin with, this allows for a principled way of incorporating prior knowledge (or results of a previous round of inference in order to perform on-line learning) if such knowledge exists. However, beyond this, inferring the full distribution over the parameters facilitates model comparison and hypothesis testing. Furthermore, if the results of inference are to be used in a decision-making context, such as for the medical application being considered here, the optimal decision is the Bayesian decision [17]. The model-based approach is also appealing as it allows us to consider building larger models, such as hierarchical models that enable us to reason about the differences between individuals and groups of people (using shared hyper-priors), and also to consider transfer learning.

In order to solve the (intractable) inference problems, we will take two approaches. Firstly, we will use Gibbs sampling [3], which is a Markov chain Monte Carlo (MCMC) algorithm for obtaining a sequence of observations which are approximated from a specified multivariate probability distribution. Gibbs sampling has the advantage of being easy to implement, and is particularly well-adapted to sampling the posterior distribution of a Bayesian network, since Bayesian networks are typically specified as a collection of conditional distributions.

We will also consider the deterministic approximation method Expectation Propagation (EP) [11], a generalisation of Belief Propagation (BP) in which the true posterior distribution is approximated with a simpler distribution, which is close in the sense of Kullback-Leibler (KL) divergence. EP approximates the belief states with expectations, such as means and variances, giving it much wider scope than would be possible with BP.

2 Related Work

Many methods and statistical techniques have been developed to analyse and understand circular data, mainly from a frequentist perspective. The popular approaches have been embedding, wrapping and intrinsic approaches (see *e.g.* [5, 9]). Here we focus on the wrapping approach, and specifically the Wrapped Normal (\mathcal{WN}) distribution [9]. A survey of Bayesian analysis of circular data using the wrapping method was given by [15], and the approaches herein build upon this work.

The use of circular statistics to model circadian or diurnal rhythms was first considered by [9], and also discussed by [16], in which various procedures for the analysis of circadian rhythms at population, organism, cellular and molecular

levels were examined, ranging from visual inspection of time plots to several mathematical methods of time series analysis.

A multivariate \mathcal{WN} Mixture (\mathcal{WNM}) model was defined used by [1] for the modelling of high-rate quantisation of phase data of speech, in which the authors used the Expectation Maximisation (EM) algorithm to learn the location and covariance parameters. Note however that a maximisation algorithm such as EM is capable of only returning a single point from the distribution, rather than a full distribution over the parameters.

Recently two non-parametric Bayesian models of circular variables based on Dirichlet Process (DP) Mixtures of normal distributions were introduced [14]: the first was a projected DP mixture of of bi-variate normals and the second was based on \mathcal{WN} s. Inference was done in this case using Gibbs slice sampling, and has the appeal that in theory it is possible to learn the number of mixture components rather than having to pre-specify or use model comparison. However, inference in this case is extremely expensive, with large numbers of iterations (40,000 were used) required, and large numbers of data points are required to fit the large number of parameters in the model.

3 Methods

Let x be a circular random variable defined on the circumference of a circle. The corresponding circular probability density function (pdf) $f(\cdot)$ is periodic with period γ : $f(x) = f(x + w\gamma), \forall w \in \mathbb{Z}, \gamma \geq 0$. Usually the distributions are defined over the unit circle, in which case $\gamma = 2\pi$, but arbitrary $\gamma \geq 0$ can be considered by a simple rescaling of x . The function $f(\cdot)$ integrates to 1 over $(0, \gamma]$. For notational simplicity, we will assume that all circular variables are constrained to their principal values, obtained by taking the modulo operation $x \leftarrow x \bmod \gamma$.

The circular distance between two points x, z for a given period γ is given by [9, eq.2.3.13]:

$$d_\gamma(x, z) = \min(x - z, \gamma - (x - z)) = \frac{\gamma}{2} - \left| \frac{\gamma}{2} - |x - z| \right|. \tag{1}$$

There exist distributions directly defined on the (unit) circle, such as the von-Mises or Circular Normal distribution (see [9, section2.2.4]), but for reasons given below we will focus on the \mathcal{WN} distribution.

3.1 The Wrapped Normal (\mathcal{WN}) Distribution

A “wrapped” distribution is one that results from wrapping the pdf of a linear random variable to the circumference a (unit) circle (infinitely many times). The corresponding distributions are called wrapped distributions, and any continuous pdf can be wrapped in this way. The Wrapped Normal (\mathcal{WN}) distribution is the circular analog of the normal distribution, achieved by wrapping in this way. In practice, the von-Mises and the \mathcal{WN} distribution are very similar [9]. However,

the wrapped Normal distribution is more convenient for Bayesian inference, as many of the technical details can be brought over from the (well studied) Normal distribution – for example, it is closed under convolution [9]. The probability density function of the wrapped normal distribution is [9]

$$f_{\mathcal{WN}}(x; \mu, \sigma, \gamma) = \frac{1}{\sigma\sqrt{2\pi}} \sum_{k=-\infty}^{\infty} \exp\left[\frac{-(x - \mu + \gamma k)^2}{2\sigma^2}\right], \quad (2)$$

with $x \in [0, 2\pi)$, location parameter $\mu \in [0, 2\pi)$, and uncertainty parameter $\sigma > 0$. We will use $\tau = \frac{1}{\sigma^2}$ to denote the precision. Because the summands of the series converge to zero, it is natural to approximate the pdf with the finite series:

$$\hat{f}_{\mathcal{WN}}(x; \mu, \sigma, \gamma) = \frac{1}{\sigma\sqrt{2\pi}} \sum_{k=-K}^K \exp\left[\frac{-(x - \mu + \gamma k)^2}{2\sigma^2}\right] \approx f_{\mathcal{WN}}(x; \mu, \sigma, \gamma), \quad (3)$$

where only $2K + 1$ summands are considered. However, one can intuitively see that for small values of K , this will only be a good approximation for small values of σ .

The \mathcal{WN} can also be expressed in terms of the Jacobi theta function (see [5, eq.(2.2.15)]), which leads to a second approximation that is more accurate for large values of σ .

$$\begin{aligned} \tilde{f}_{\mathcal{WN}}(x; \mu, \sigma, \gamma) &\approx f_{\mathcal{WN}}(x; \mu, \sigma, \gamma) \\ &= \frac{1}{\gamma} \left(1 + 2 \sum_{k=1}^K e^{-\frac{\sigma^2}{2} \left(\frac{2\pi k}{\gamma}\right)^2} \cos\left(\frac{2\pi k}{\gamma}(x - \mu)\right) \right), \end{aligned} \quad (4)$$

where only K summands are considered. Theoretical bounds are given in [8] that show that the errors of both approximations decrease exponentially with the number of summands, and show that the first representation performs well for small σ whereas the other performs well for large σ .

3.2 Bayesian Inference

The \mathcal{WN} distribution possesses the additive property [5], *i.e.* the convolution of two \mathcal{WN} distributed variables is also a \mathcal{WN} distribution. Hence for the purposes of Bayesian inference, the conjugate prior for the location parameter μ of a \mathcal{WN} distribution is another \mathcal{WN} distribution, which we denote as $\mathcal{WN}_0(\mu; \mu_0, \sigma_0, \gamma)$. The conjugate prior for the precision τ is the Gamma distribution, denoted by $\mathcal{Ga}(\tau; \alpha_0, \beta_0)$ for shape and rate parameters α_0 and β_0 respectively, as would be the case for the Normal distribution.

In [Figure 1a](#) we show the factor graph for the \mathcal{WN} model, where the shading of the x variable indicates that it is observed, and the box around x and the \mathcal{WN} factor is a plate, indicating that this part of the graph is repeated N times. Inference can be performed in this model using Gibbs sampling, where we use the approximations given in [Equation 3](#) and [Equation 4](#), and where we the former is used if $\sigma^2 < 0.15$ and the latter in the reverse case, as suggested by [8].

3.3 \mathcal{WN} Mixture (\mathcal{WNM}) Models

We define the \mathcal{WN} Mixture (\mathcal{WNM}) distribution (*i.e.* a mixture of \mathcal{WN} distributions) in the following way,

$$f_{\mathcal{WNM}}(x; \mu, \sigma, \gamma, \phi, M) = \sum_{m=1}^M w_m \mathcal{WN}(x; \mu_k, \sigma_k^2, \gamma), \quad (5)$$

where $w \in \mathbb{R}^M : \sum_{m=1}^M w_m = 1, w_m \geq 0$, *i.e.* $w \sim \text{Cat}(\phi)$ are the mixing coefficients which are drawn from a categorical (discrete) distribution of dimension $M > 0$ with a probability vector ϕ . The conjugate prior for ϕ is the Dirichlet distribution, with a concentration parameter vector $\alpha_\phi \in \mathbb{R}^M : \alpha > 0$.

In Figure 1b we show the factor graph for the \mathcal{WNM} model using gate notation for representing the mixture model [12]. ϕ is the Dirichlet distributed variable, from which the discrete variable $\Psi \in \mathbb{R}^m$ is drawn, where m is the number of mixture components, representing the gate selector is sampled.

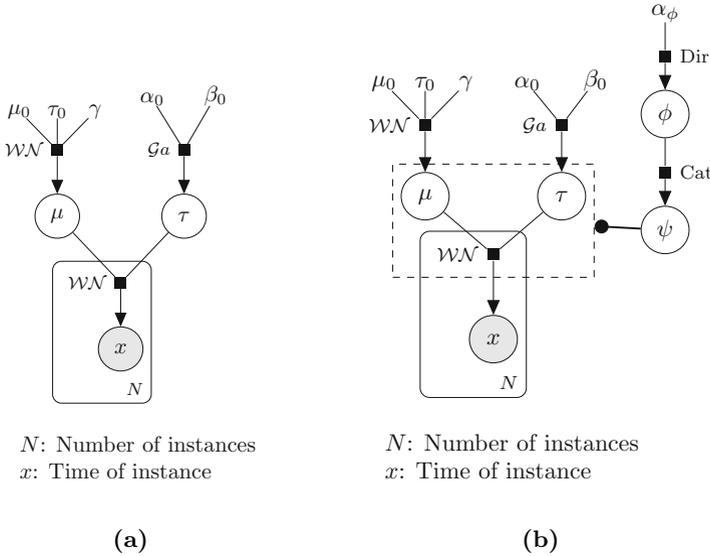


Fig. 1. (a) Wrapped Normal (\mathcal{WN}) and (b) \mathcal{WN} Mixture (\mathcal{WNM}) models.

As has been noted by [14], the standard \mathcal{WNM} model suffers from issues of identifiability, which we also found when trying to perform inference using the model. There the authors tackle the problem by “unwrapping” the distribution by conditioning on the wrapping number k_i , which results in a complex sampling procedure. Here we will take a simpler approach, that also allows us to use EP as well as Gibbs sampling.

3.4 Approximate \mathcal{WN} (\mathcal{AWN})

We approximate the \mathcal{WN} with mixture of \tilde{M} normal distributions. If we insist that \tilde{M} is odd, and define a vector of offsets, $\delta = (\delta/2)_{\delta=-\tilde{M}-1}^{\tilde{M}-1}$, the \mathcal{AWN} model is defined by

$$f_{\mathcal{AWN}}(x; \mu, \sigma^2, \tilde{M}) = \frac{1}{\tilde{M}} \sum_{\delta \in \delta} \mathcal{N}(x; \mu + \delta\gamma, \sigma^2). \quad (6)$$

This model cascades a series of \tilde{M} Gaussian distributions along the real line where adjacent distributions are a distance of γ apart, all distributions share the same variance, and the mean of the central component is constrained to be found within the periodic range, $[0, \gamma)$ (this is the only component that will fall within this range). The components whose means fall outside the periodic range contribute to modelling by mimicking the wrapped tails of the \mathcal{WN} model. Indeed, as \tilde{M} tends towards infinity the \mathcal{AWN} approximation approaches \mathcal{WN} . \mathcal{AWN} models requires specification of three parameters: μ , σ^2 and \tilde{M} , and the factor graph for this model is shown in [Figure 2a](#).

By modelling periodic distributions in this manner, we can approximate the posterior distributions of the \mathcal{WN} parameters as one would estimate Bayesian mixture model parameters. We can again use Gibbs sampling to perform inference for the \mathcal{AWN} model. However, since we have replaced the \mathcal{WN} distribution with standard normal distributions, we can also use Expectation Propagation (EP). EP has a major advantage over Gibbs sampling in this setting, which is that it is relatively easy to compute model evidence (see [Equation 7](#) in [subsection 3.6](#)) which will allow us to do model comparison. In the first set of experiments (see [subsection 4.1](#) and [5.1](#)) we will compare the two inference methods for this model.

3.5 Approximate \mathcal{WNM} (\mathcal{AWNM})

Generalisation of the \mathcal{AWN} models to an \mathcal{AWNM} is achieved by straightforward application of a standard mixture model gate over the parameter means, variances and approximation factors. The factor graph for this is given in [Figure 2b](#), where mixing factors have been introduced.

As with the \mathcal{AWN} model, we can again use either Gibbs sampling or EP to perform inference for the \mathcal{AWNM} model. The computation of evidence [Equation 7](#) plays an even greater role here, since it gives us a method to select the number of mixture components K (see [5.1](#)). In the first set of experiments (see [subsection 4.1](#) and [5.1](#)) we will compare the two inference methods for this model.

3.6 Model Comparison

We also perform Bayesian model comparison, in which we marginalise over the parameters for the type of model being used, with the remaining variable being

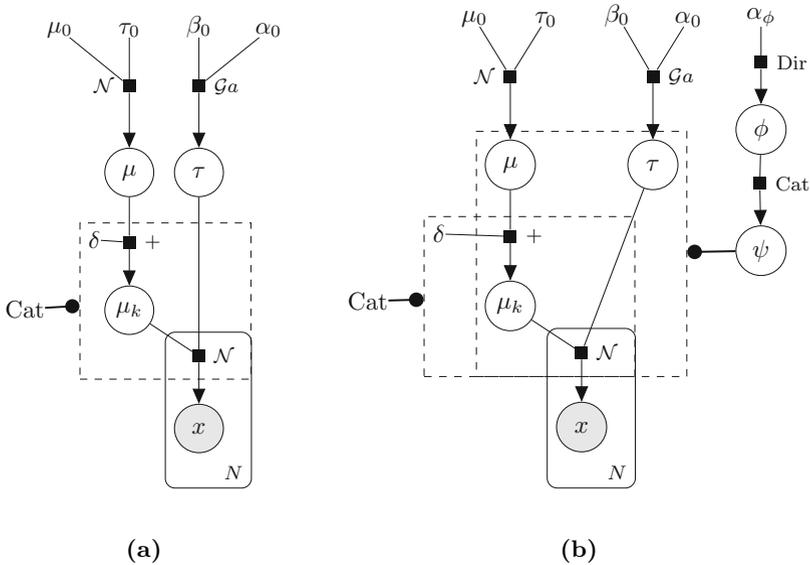


Fig. 2. (a) AWN and (b) $AWNM$ models.

the identity of the model itself. The resulting marginalised likelihood, known as the model evidence, is the probability of the data given the model type, not assuming any particular model parameters. Using D for data, θ to denote model parameters, H as the hypothesis, the marginal likelihood for the model H is

$$p(D|H) = \int p(D|\theta, H) p(\theta|H) d\theta \tag{7}$$

This quantity can then be used to compute the ‘‘Bayes factor’’ [4], which is the posterior odds ratio for a model H_1 against another model H_2 ,

$$\frac{p(H_1|D)}{p(H_2|D)} = \frac{p(H_1)p(D|H_1)}{p(H_2)p(D|H_2)}. \tag{8}$$

3.7 Rose Diagrams

A useful variant of the circular histogram is a ‘‘rose diagram’’, in which the bars of a histogram are replaced by segments. The area of each segment is proportional to the frequency of the corresponding group. As such, for groups of equal width, the radius should be proportional to the square root of the relative frequency [9]. We will use these, but with a slight abuse (since the maximum value of a pdf is arbitrary) we will plot the WN and WNM pdfs over the rose diagrams with the maximum of the pdf coinciding with the outside of the plot.

Table 1. Parameter settings for toy the data generated from (a) \mathcal{WN} distribution and (b) \mathcal{WNM} distribution of Equation 9

(a) \mathcal{WN} parameters			(b) \mathcal{WNM} parameters				
Data set	μ	σ^2	Data set	μ_1	σ_1^2	μ_2	σ_2^2
0	0.0	10.0	0	0.0	2.0	12.0	2.0
1	21.0	2.0	1	6.0	4.0	18.0	4.0
2	3.0	2.0	2	6.0	10.0	9.0	10.0
3	10.0	10.0	3	2.0	2.0	3.0	2.0

4 Experiments

All models were implemented using Infer.NET [10], a framework for running Bayesian inference in graphical models. Model specifications will be provided in the supplementary material accompanying this paper.

4.1 Toy Data

In order to evaluate the models, we first created toy datasets where we sampled from \mathcal{WN} and \mathcal{WNM} distributed data. For testing the uni-modal models, data were generated from \mathcal{WN} distribution with the settings for μ and σ^2 given in Table 1a. For testing the mixture models, data were generated from the following mixture model:

$$f(x) = 0.6 \mathcal{WN}(x; \mu_1, \sigma_1^2, \gamma) + 0.4 \mathcal{WN}(x; \mu_2, \sigma_2^2, \gamma) \tag{9}$$

where $\mu_1, \sigma_1^2, \mu_2, \sigma_2^2$ were set as in Table 1b. The first two are in some sense “easy”, since the means are well separated, with the two cases being used to ensure there were no inference pathologies. The third and fourth are harder problems as the variances are large with respect to the difference in means, where in data set 2 the variances are large and in data set 3 the variances are smaller.

We measure the mean difference (MD) for the estimated moments of the \mathcal{WN} components:

$$MD_\mu = \frac{1}{n} \sum_{i=1}^n |d_\gamma(\mu_i, \hat{\mu}_i)|, \quad MD_\sigma = \frac{1}{n} \sum_{i=1}^n |\sigma_i - \hat{\sigma}_i| \tag{10}$$

where $d_\gamma(x, z)$ is the circular distance defined in Equation 1 and n is the number of random repetitions used.

4.2 The CASAS HH101 Dataset

We next examine some real-world data collected by the CASAS research group [2]. The HH101 data set¹ contains 3 months of single-resident apartment data

¹ <http://casas.wsu.edu/datasets/hh101.zip>

with partial annotations, with 30 different activities appearing in the annotations. The house was equipped with motion sensors, door sensors, temperature sensors, and ambient light sensors, which were recorded asynchronously. We chose this data for the length of recording, and due to the fact that it was from a single resident, to avoid further complications caused by multiple residents. The layout of the house with sensor locations marked with circles can be seen in Figure 3.

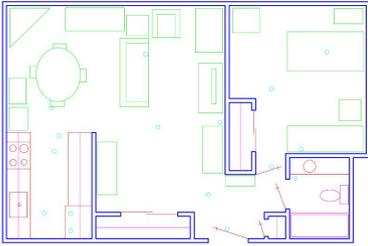


Fig. 3. Floorplan of the CASAS HH101 dataset.

Figure 4 shows the log of the total time spent performing each activity for each of the labelled activities in the CASAS HH101 dataset. It’s worth noting that this dataset is dominated by 3 activities (Sleep, Sleep_Out_Of_Bed, and Watch_TV), which is perhaps in part due to the ease of labelling these activities, and in part due the fact that the resident was an elderly person. This will clearly play an important role in the quality of inference, simply due to the number of examples available.

Despite not modelling the sensor activations themselves, our data instances are in fact dependent on the sensor activations, since the dataset only contains annotations where sensor activations exist. In order to provide samples of the times of activity occurrences to our models, we could take the start end times of the activity and then re-sample from within this range (uniformly or otherwise). Here for simplicity we assume that the sensor activations in the period between the start and end annotations themselves provide independent samples of the times of an activity.

4.3 Priors

The period γ of all \mathcal{WN} distributions in our experiments were set to 24,

We note that there are sometimes errors in the data, such as ON/OFF events not being paired up correctly. When parsing the data we take a conservative approach, finding only OFF events that follow ON events. As with the sensors, there are sometimes errors in the activity labelling. We use the same conservative method. Note also that there are sometimes activity labels that are orphaned – *i.e.* there is no BEGIN/END trigger but simply a single label next to a sensor activation – these are ignored.

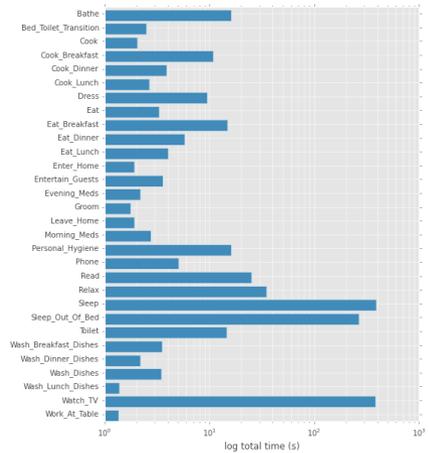


Fig. 4. Log of total time spent performing each activity for each of the labelled activities in the CASAS HH101 dataset.

representing the 24 hours in the day. In the \mathcal{WN} model we set the location parameter of the prior over the location to 0, and the precision to $(\frac{\gamma}{2})^{-2}$, meaning that two σ (wrapped standard deviations) in each direction will reach around the period, which roughly corresponds to a uniform distribution over the circle. We set the Gamma prior hyper-parameters were set to $\alpha_0 = 1$, $\beta_0 = 1$, which simply favours smaller precisions (and therefore larger variances).

In the \mathcal{AWN} model we set the location parameter of the prior over the location to $\frac{\gamma}{2}$, as this is the uninformative prior for the approximated model. All other hyper-parameters were the same as for the \mathcal{WN} model. In the $\mathcal{AWN}\mathcal{M}$ model we set the location parameter of each of the mixture components to $\frac{\gamma}{2}$. The prior precision was set to $(\frac{\gamma}{2K})^{-2}$ where K is the number of mixture components. The Gamma hyper-parameters were as with the uni-modal case.

4.4 Symmetry Breaking

In a normal mixture model, it is well known there is a symmetry in the mixture component assignments that needs to be broken by randomly initialising each data point to one of the components. In the $\mathcal{AWN}\mathcal{M}$ model, this symmetry is also present, but there is an additional symmetry caused by the approximation. Fortunately, both symmetries can be broken using a different method, where the means of the components are initialised to $\frac{m\gamma}{M}, k = 1, \dots, M$, where M is the number of mixture model components (not approximation components \tilde{M}), *i.e.* we distribute the prior means evenly around the circle. Once the means have been initialised in this way, it is no longer necessary to randomly assign the mixture components (and in fact may slow down convergence).

5 Results

We first present results for the \mathcal{WN} model and the \mathcal{AWN} model using both Gibbs sampling and Expectation Propagation (EP) on data generated from a \mathcal{WN} distribution, to show that the EP \mathcal{AWN} model is sufficiently accurate for our purposes. This validation is useful, since although EP is a deterministic algorithm, there is no guarantee of convergence if there are any loops present in the graph. We then show that this accuracy carries over to the $\mathcal{AWN}\mathcal{M}$ model on data generated from an $\mathcal{WN}\mathcal{M}$ distribution. We then show results on a smart home dataset from the Casas group.

In the following experiments we monitored the convergence of the models after each round of inference, where a round was determined to be a single full iteration of EP, or 100 iterations of Gibbs. The convergence criterion was that the means of each component had not moved by more than 30 seconds ($= \gamma/1800 \approx 0.01$) from one the previous round (other criteria are possible, but this was simple and effective).

5.1 Toy Data

Uni-modal Data. Details of the data generating process are in [subsection 4.1](#) using the parameter settings in [Table 1a](#), where we generated 100 data points and performed 5 repetitions of each data set with different random seeds. The results of learning the \mathcal{WN} model using Gibbs sampling, and \mathcal{AWN} using Gibbs sampling and EP are shown in [Figure 5](#), where performance is measured in terms of MD_μ and MD_σ as defined in [Equation 10](#). We can see that the \mathcal{AWN} model using Gibbs sampling performs almost identically to the \mathcal{WN} model in the estimation of both moments of the distribution. The \mathcal{AWN} model using EP has slightly degraded performance in terms of estimating the location μ , but is able to accurately estimate σ . The average running times were \mathcal{WN} : 0.12s, \mathcal{AWN} (Gibbs): 0.40s, and \mathcal{AWN} (EP): 0.33s.

Data set	Model					
	WN(Gibbs)		AWN(Gibbs)		AWN(EP)	
	Average of MD_μ	Average of MD_σ	Average of MD_μ	Average of MD_σ	Average of MD_μ	Average of MD_σ
0	0.15	0.29	0.14	0.29	0.30	0.30
1	0.07	0.13	0.06	0.13	0.06	0.13
2	0.06	0.13	0.06	0.13	0.07	0.13
3	0.15	0.29	0.14	0.29	0.15	0.29
Overall Average	0.11	0.21	0.10	0.21	0.15	0.21

Fig. 5. Results on data generated from a uni-modal \mathcal{WN} distribution, comparing the \mathcal{WN} model with \mathcal{AWN} model for both Gibbs and EP.

Mixture Model Data. In the following experiments we generated 100 data points in each data set, and repeated the experiments 5 times with different random seeds. The results in [Figure 6](#) indicate that for fairly small data sets, the EP version of the model is in fact more accurate in terms of MD for the estimated moments. EP and Gibbs required on average over all of the experiments ≈ 20 and 2100 iterations to converge respectively, and EP reached convergence in on average roughly one fifth of the computation time required by Gibbs sampling.

Data set	Average of $MD_{\mu 1}$		Average of $MD_{\sigma 1}$		Average of $MD_{\mu 2}$		Average of $MD_{\sigma 2}$	
	AWN(EP)	AWN(Gibbs)	AWN(EP)	AWN(Gibbs)	AWN(EP)	AWN(Gibbs)	AWN(EP)	AWN(Gibbs)
0	0.17	0.07	0.08	0.04	0.15	0.15	0.13	0.73
1	0.08	0.10	0.17	0.17	0.31	0.26	0.20	0.57
2	0.65	1.34	0.45	0.91	1.15	0.94	1.54	1.79
3	0.19	0.98	0.11	0.11	1.15	1.26	0.33	0.27
Overall Average	0.27	0.62	0.21	0.30	0.69	0.65	0.55	0.84

Fig. 6. *Small data set:* Accuracy of inference of the \mathcal{AWN} model on data generated from an \mathcal{WN} distribution (details in [subsection 4.1 Table 1b](#)).

The results in [Figure 7](#) indicate that for larger data sets, the Gibbs version of the model is more accurate in terms of MD for the estimated moments than the EP version. This is explained by the difficulty of data set “2”, which corresponds

to the “pathological” case outlined in the model comparison discussion below, and as such the errors that we see EP making here are that it estimates there is a single mode rather than two, which is not wholly unreasonable. EP and Gibbs required on average over all of the experiments roughly 28 and 1115 iterations to converge respectively, and EP reached convergence in roughly half the computation time than is required by Gibbs sampling.

Data set	Average of MD μ_1		Average of MD σ_1		Average of MD μ_2		Average of MD σ_2	
	AWNM(EP)	AWNM(Gibbs)	AWNM(EP)	AWNM(Gibbs)	AWNM(EP)	AWNM(Gibbs)	AWNM(EP)	AWNM(Gibbs)
0	0.06	0.10	0.01	0.12	0.05	0.62	0.02	1.32
1	0.06	0.14	0.04	0.23	0.07	0.59	0.06	1.11
2	0.77	1.41	0.43	0.87	2.16	1.23	0.38	1.45
3	0.35	0.24	0.20	0.18	1.02	0.46	0.31	0.29
Overall Average	0.31	0.47	0.17	0.35	0.82	0.72	0.19	1.04

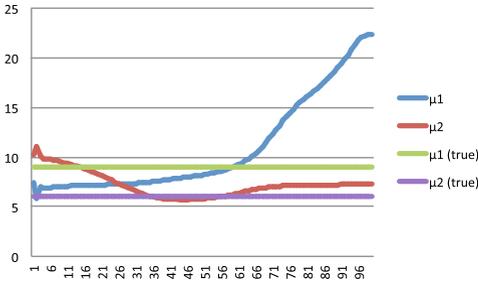
Fig. 7. *Larger data set:* Accuracy of inference of the \mathcal{AWNM} model on data generated from an \mathcal{WNM} distribution (details in subsection 4.1 Table 1b).

Model Selection. We are able to take advantage of the fact that we are using EP in the \mathcal{AWNM} model to perform model selection, since the model evidence computations are more straightforward for EP than for Gibbs sampling, and in fact have already been implemented in Infer.NET. In order to test the ability to use model evidence for model selection purposes, we ran the following experiment. We generated data from a \mathcal{WNM} distribution with $K = \{1, 2, 3, 4\}$ components. We sampled the mixture weights for the components from a symmetric Dirichlet distribution $\text{Dir}(10, 10)$ (which gives roughly equal mass to each of the components) and then sampled 200 data points from the mixture distribution according to those weights. We then computed the model evidence for the \mathcal{AWNM} model with $K = \{1, 2, 3, 4\}$ components, *i.e.* we learnt a model for each possible pair of true K and model K (16 in total). The results are shown in Figure 8. The true K values lie on the diagonal (*i.e.* where the correct K was supplied to the model). As can be seen in bold, the model gives the highest evidence to the those values of K across each row, meaning that by selecting the model with the highest evidence we would indeed choose the correct value for K .

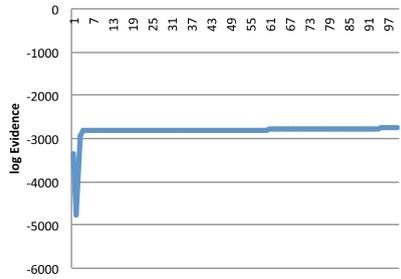
However Figure 9 shows a seemingly pathological case. In this example the true means $\mu_1 = 6, \mu_2 = 9$ are quite close together, with a large $\sigma^2 = 10$ for both components. We can see that the model at first seems to converge to the correct means, but then appears to diverge away. At the end of this inference run, the estimated weights for the components were $w_1 \approx 0.02, w_2 \approx 0.98$, showing that the model had put all of the mass on the second component, with the mean being close to the average of the true mean

True K	Model K			
	1	2	3	4
1	-710.3	-712.6	-718.9	-720.1
2	-763.9	-742.2	-746.6	-751.5
3	-819.1	-811.4	-810.7	-815.3
4	-858.5	-846.9	-860.5	-841.7

Fig. 8. Model evidence computation for the \mathcal{AWNM} model using EP. See text for details.

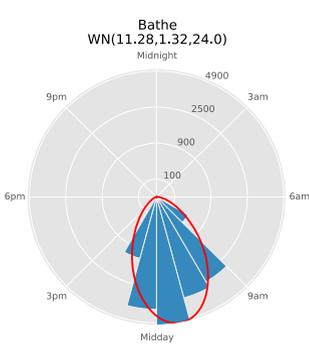


(a) Estimated and true means.

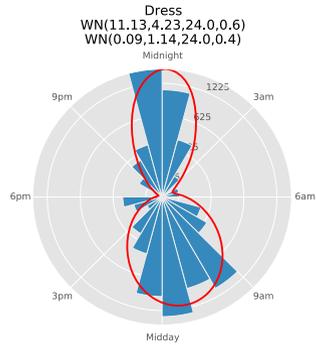


(b) Model evidence.

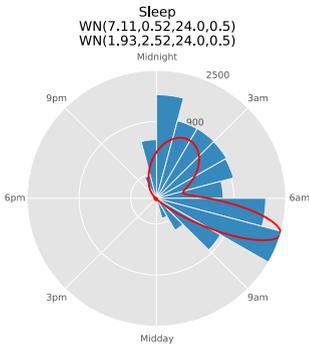
Fig. 9. A pathological case. The x -axis in both figures shows the EP iteration count. The true means $\mu_1 = 9, \mu_2 = 6$ are quite close, with a large $\sigma^2 = 10$ for both. Note that the model evidence continues to rise, despite the estimated means diverging from the truth. This is because in this case there is insufficient evidence for a bimodal model due to the high variances.



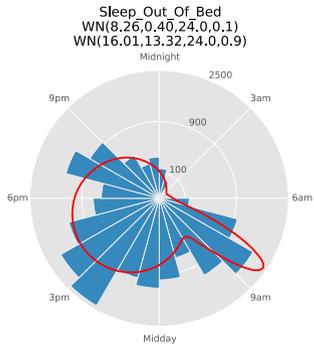
(a) Bathe



(b) Dress



(c) Sleep



(d) Sleep_Out_Of_Bed

Fig. 10. Posterior means of the \mathcal{AWNM} model fitted to activities from the CASAS HH101 dataset. Note that the model correctly captures the multi-modal nature of the activities. The \mathcal{WN} parameters are given in the subtitles of each subplot (mixture weights not shown).

Activity	K		
	1	2	3
Bathe	-2.65894389	-2.66002403	-2.661022468
Bed_Toilet_Transition	-10.19686747	-2.333433735	-2.349457831
Cook	-4.919811739	-2.632668882	-2.284385382
Cook_Breakfast	-2.25312476	-2.254319898	-2.255758157
Cook_Dinner	-4.144262708	-2.324962255	-4.364614997
Cook_Lunch	-1.750226804	-1.763010309	-1.774391753
Dress	-4.233888048	-3.645742382	-3.727730711
Eat	-2.827253886	-2.855440415	-2.880103627
Eat_Breakfast	-2.293162839	-2.297160752	-2.300803758
Eat_Dinner	-11.75947075	-2.220389972	-2.163091922
Eat_Lunch	-2.010291971	-2.031423358	-2.050072993
Enter_Home	-3.487927461	-3.40022946	-3.352487047
Evening_Meds	-3.223430398	-2.777507102	-2.766882102
Groom	-2.167755102	-1.994489796	-2.321428571
Leave_Home	-3.453540445	-3.391535756	-3.236617819
Morning_Meds	-4.422052195	-2.312746145	-2.316844603
Personal_Hygiene	-4.240142814	-4.241790439	-4.144003307
Phone	-3.553333333	-3.283061224	-3.321496599
Read	-5.354507351	-2.660965548	-2.62588984
Relax	-3.424166667	-3.185714286	-3.153554422
Sleep	-3.569722472	-3.219689469	-3.219689469
Sleep_Out_Of_Bed	-3.955734676	-3.896021116	-3.896580311
Toilet	-2.164069125	-3.981019063	-3.964657046
Wash_Breakfast_Dishes	-2.656940897	-2.661735099	-2.666066225
Wash_Dinner_Dishes	-3.874854167	-2.685444444	-2.520513889
Wash_Dishes	-4.064689332	-3.543100821	-3.334431419
Wash_Lunch_Dishes	-1.510576132	-1.534403292	-1.555349794

Fig. 11. Log evidence on the CASAS HH101 dataset for different values of mixture components M .

components. Interestingly, the model evidence continues to rise throughout, indicating that the model has favoured parsimony, which in the sense of Occam’s razor would be the sensible thing to do.

5.2 Smart Home Data

We now give results on some real-world data from the CASAS HH101 dataset, as described in subsection 4.2. We will use the \mathcal{AWNM} model for the following experiments, where we run inference with $M = 1, 2, 3, 4$ mixture components, $M = 3$, priors set as given in subsection 4.3, and symmetry breaking initialisation as given in subsection 4.4. We use model evidence Equation 7 to choose the number of mixture components K . We plot the posterior moments for four of the activities in Figure 10, where we take the posterior means $\mu_m, m = 1, \dots, M$ and the posterior mean of $\tau_m, m = 1, \dots, M$, the posterior mean of the mixture weights ϕ , and construct a \mathcal{WNM} distribution using these parameters.

Note that many of the activities are clearly multi-modal, such as the “Dress” activity in Figure 10b, and the model is able firstly to correctly identify the number of mixture components, and also to capture the multi-modal nature of the periodicity of the activities. The “Sleep_Out_Of_Bed” activity is interesting as there is a prominent “lobe” of the distribution from the narrower of the mixture components. Figure 11 shows the log model evidence scaled by the number of data points for the models learnt with $K = 1, 2, 3$ mixture components for each activity, showing the number of components chosen by model selection. Note that in some cases it is quite clear cut that one of the models should be preferred, but in other cases the choice is more borderline.

6 Discussion

The results indicate that the Approximate \mathcal{WNM} (\mathcal{AWNM}) is able to accurately estimate the moments of \mathcal{WN} Mixture (\mathcal{WNM}) distributed data, as demonstrated by the experiments on the toy dataset. Furthermore, the Expectation Propagation (EP) implementation is appealing, since it gives comparable results to the Gibbs sampling approach, but is generally faster and also enables model selection through evidence computation.

The inferred posteriors could then be used to (a) generate continuous feature(s) to be used in a classifier (probabilistic or otherwise), *e.g.* by using the log probability of activities given time-stamp. We would expect to see some improvement over a simple “hour-of-day” feature as it is a more refined representation of the distribution over time.

Potentially more interesting, however, is that since we have full distribution over the parameters, we can use these in a larger probabilistic model. For example, we can easily perform a modelling of the periodicity of the sensors activations in the same way, and then learn a mapping from sensor to activity which would in effect be a form of periodic regression.

7 Conclusions

In this paper we have shown that Bayesian inference for the \mathcal{WN} distribution (using Gibbs) is easy to implement and accurate for data generated from a the model. The \mathcal{WNM} suffers from identifiability issues, so we introduced an approximate version \mathcal{AWN} which can be easily implemented using either Gibbs or EP in the modelling framework Infer.NET [10], and that this model accurately approximates the \mathcal{WN} model. We then showed that we could extend this to the mixture modelling \mathcal{AWNM} , and demonstrated how model evidence can be used for model selection (choosing the number of mixture components). We then showed some results of performing inference using the \mathcal{AWNM} model on a real-world smart home data set.

7.1 Further Work

An appealing extension would be to construct a multivariate \mathcal{WNM} model to model all of the sensor activations (from binary sensors) in a smart home together, with the resulting covariance matrix giving a description of the periodic linkage between sensors. Following on from this, it would be interesting to combine such a multivariate model with the univariate model, either by adding an extra dimension for the activities, or by constructing a circular regression task, for example by using the circular regression approach outlined by [15].

Another appealing extension would be to consider a hierarchical model for different residents in a given home, where common hyper-priors are shared between the residents, and individual priors are then inferred for each resident. This would be a natural path to being able to transfer such models to new homes and new residents.

Acknowledgments. This work was performed under the Sensor Platform for HEalth-care in Residential Environment (SPHERE) Interdisciplinary Research Collaboration (IRC) funded by the UK Engineering and Physical Sciences Research Council, Grant EP/K031910/1.

References

1. Agiomyrgiannakis, Y., Stylianou, Y.: Wrapped Gaussian mixture models for modeling and high-rate quantization of phase data of speech. *IEEE Transactions on Audio, Speech, and Language Processing* **17**(4), 775–786 (2009)
2. Cook, D.J., Schmitter-Edgecombe, M.: Assessing the quality of activities in a smart environment. *Methods Inf. Med.* **48**(5), 480–485 (2009)
3. Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B.: *Bayesian data analysis*, vol. 2. Taylor & Francis (2014)
4. Goodman, S.N.: Toward evidence-based medical statistics. 2: The Bayes factor. *Annals of Internal Medicine* **130**(12), 1005–1013 (1999)
5. Jammalamadaka, S.R., Sengupta, A.: *Topics in circular statistics*, Series on Multivariate Analysis, vol. 5. World Scientific (2001)
6. Kim, E., Helal, S., Cook, D.: Human activity recognition and pattern discovery. *IEEE Pervasive Computing* **9**(1), 48–53 (2010)
7. Krishnan, N., Cook, D.J., Wemlinger, Z.: Learning a taxonomy of predefined and discovered activity patterns. *Journal of Ambient Intelligence and Smart Environments* **5**(6), 621–637 (2013)
8. Kurz, G., Gilitschenski, I., Hanebeck, U.D.: Efficient evaluation of the probability density function of a wrapped normal distribution. In: *Sensor Data Fusion: Trends, Solutions, Applications (SDF)*, 2014, pp. 1–5. IEEE (2014)
9. Mardia, K.V., Jupp, P.E.: *Directional statistics*. John Wiley & Sons, Chichester (2000)
10. Minka, T., Winn, J., Guiver, J., Webster, S., Zaykov, Y., Yangel, B., Spengler, A., Bronskill, J.: *Infer.NET 2.6*. Microsoft Research Cambridge (2014). <http://research.microsoft.com/infernet>
11. Minka, T.P.: Expectation propagation for approximate Bayesian inference. In: *Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence*, pp. 362–369. Morgan Kaufmann Publishers Inc. (2001)
12. Minka, T., Winn, J.: Gates. In: *Advances in Neural Information Processing Systems*, pp. 1073–1080 (2009)
13. Nazerfard, E., Das, B., Holder, L.B., Cook, D.J.: Conditional random fields for activity recognition in smart environments. In: *Proceedings of the 1st ACM International Health Informatics Symposium*, pp. 282–286. ACM (2010)
14. Nuñez-Antonio, G., Ausín, M.C., Wiper, M.P.: Bayesian nonparametric models of circular variables based on Dirichlet process mixtures of normal distributions. *Journal of Agricultural, Biological, and Environmental Statistics*, 1–18 (2014)
15. Ravindran, P., Ghosh, S.K.: Bayesian analysis of circular data using wrapped distributions. *Journal of Statistical Theory and Practice* **5**(4), 547–561 (2011)
16. Refinetti, R., Cornelissen, G., Halberg, F.: Procedures for numerical analysis of circadian rhythms. *Biological Rhythm Research* **38**(4), 275–325 (2007)
17. Robert, C.: *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Springer Science & Business Media (2007)
18. Wilks, D.S.: *Statistical methods in the atmospheric sciences*, vol. 100. Academic press (2011)