1

Some advances in constrained inference for ordered circular parameters in oscillatory systems

Cristina Rueda*, Miguel A. Fernández*, Sandra Barragán* and Shyamal D. Peddada**

*Universidad de Valladolid, Spain
**National Institute of Environmental Health Sciences, USA

1.1 Introduction

Constraints on parameters arise naturally in many applications. Statistical methods that honor the underlying constraints tend to be more powerful and result in better interpretation of the underlying scientific data. In the context of Euclidean space data, there exists over five decades of statistical literature on constrained statistical inference and at least four books on the subject (e.g. Robertson et al. (1988), Silvapulle and Sen (2005)). However, it was not until recently that these methods are being used extensively in applied research. For example, constrained statistical inference is gaining considerable interest among applied researchers in a variety of fields, such as toxicology (Peddada et al. (2007)), genomics (Hoenerhoff et al. (2012), Perdivara et al. (2011), Peddada et al. (2003)), epidemiology (Cao et al. (2011), Peddada et al. (2005)), clinical trials (Conaway et al. (2004)), cancer trials (Conde et al. (2012), Conde et al. (2013)) etc.

 While Euclidean space data are commonly encountered in applications, there are numerous instances where the underlying data and the parameters of interest reside on a unit circle. Statistical theory and methodology for analyzing such angular data has a long history (Fisher (1993), Mardia and Jupp (2000)) and, as witnessed through his publications and his highly referenced book Mardia and Jupp (2000), Professor Mardia was one of the chief architects and pioneers of this important research area. His work has wide range of applications in fields such as geosciences, spatial data, image analysis, bioinformatics etc.

In comparison to over fifty years of statistical literature on constrained inference for Euclidean space data, constrained statistical inference for circular data is almost non-existent although constraints on unit circle were encountered by applied researchers such as social psychologists and neuroscientists (cf. Schlosberg (1952), Russell (1980), Forgas (1998), Mechsner et al. (2001), Oullier et al. (2002), Posner et al. (2005)), molecular biologists (Whitfield et al. (2002), Peng et al. (2005), Hughes et al. (2009)) etc. Parameters on a unit circle are often the result of an oscillatory system.

Oscillatory systems arise naturally in many applications, such as sales of seasonal products, regulation of hormones in humans, circadian clock, periodic expression of genes participating in cell division cycle, etc. Often there are several components (or variables) involved in such oscillatory systems that act in a well-coordinated manner like an orchestra for the system to function. The system can be disrupted if one or more components go out of order. Researchers are often interested in detecting such components. For example, large scale genomic studies are routinely conducted to identify genes/proteins that have a periodic expression in a given biological system. Depending upon the underlying scientific question of interest, researchers are often interested in correlating the phases of periodic genes across different experimental conditions or species or tissues etc. Thus the statistical problem of interest is to draw inferences regarding the relative order among parameters on a unit circle.

Just as one cannot trivially extend standard methods developed for unconstrained statistical inference in the Euclidean space to circle, constrained statistical inference for Euclidean space cannot be extended to constraints on a unit circle (cf. Rueda et al. (2009)). Since constrained statistical inference on a unit circle is a relatively new topic and yet has numerous applications, the purpose of this paper is three-fold. First we describe recent theoretical and methodological advances in this field, next we shall describe some applications of the methodology in cell biology and lastly we shall present several open research problems and potential applications. More specifically, in Section 1.2 we introduce the framework and the problem of interest. In Section 1.3 we describe the problem of estimating ordered parameters on a unit circle using circular isotonic regression. Analogous to the isotonic regression estimator in the Euclidean space, circular isotonic regression estimator (CIRE) obtains ordered estimates of circular parameters under a pre-specified order among them. Using these ordered point estimators, under suitable distributional assumptions, in Section 1.4 we describe conditional tests for order among circular parameters. In Section 1.5 the problem of estimation of a global order among a set of circular objects using data from multiple experiments is described. Statistical methodology described in this paper is illustrated in Section 1.6 using data obtained from cell biology. We conclude the paper by presenting present several open research problems and potential applications in Section 1.7.

1.2   Oscillatory data and the problems of interest

Time course data are commonly obtained in many applications. However, in some applications such as in marketing research, cell biology, endocrinology, psychology etc., researchers are interested in studying various characteristics (or parameters) of the time course pattern. Although the raw data itself may reside in the Euclidean space, the underlying parameters of interest may be points on a unit circle. To illustrate this, consider data provided in the toy example described in Figure 1.1. To promote tourism to its summer resort in an

island in the pacific, suppose a travel agency runs an advertisement campaign several months before each summer. The advertisement costs in dollars over time are plotted in Figure 1 (dashed curve). The travel agency tracks the sales of airline tickets to the island over the same period (dotted curve) as well as the income revenues on the island due to tourism (solid curve). One of the parameters of interest to the travel agency is to determine the time of peak advertisement to maximize its impact on the overall sales. Thus the parameters of interest are the times that correspond to the peaks of the curve (location of the vertical lines in Figure 1.1). Since these curves are periodic they can be mapped onto a unit circle and the time to peak value of any given curve can be thought of as an angular parameter on the circle (see Figure 1.2). Thus in this example the angular parameters are ordered with the dashed value followed by the dotted value which is followed by the solid one in the anti-clockwise direction. Focus of this paper is to draw inferences regarding the relative order among these angular parameters on the unit circle. As noted in the introduction, similar examples arise in a wide range of settings and the application of interest in this paper is cell-biology which is explained in greater detail in the illustration section.
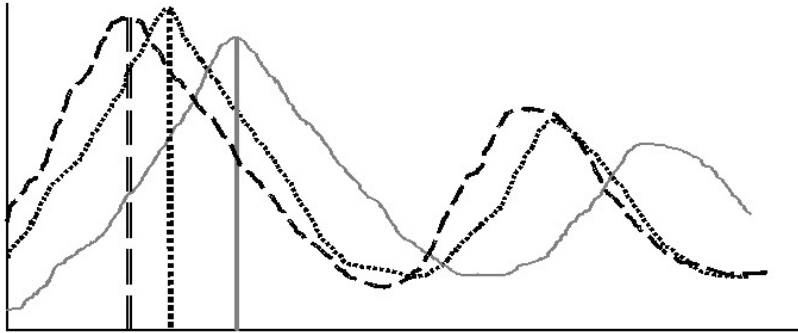

Figure 1.1   Advertisement costs in dollars over time

In Liu et al. (2004), a nonlinear model called the Random Period Model (RPM) was introduced for such time course data. Although their motivation was to describe the time course expression of cell-cycle genes, their model can be used for any such time course data. The model is given by $Y_g(t) = f(t, \eta_g) + \varepsilon_g(t)$ where t is the time, and $\varepsilon_g(t)$ is a zero mean error term with no additional distributional assumptions made. The expected response $f(t, \eta_g)$ is modeled as,

$$f(t, \eta_g) = a_g + b_g t + \frac{K_g}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \cos\left(\frac{2\pi t}{T exp(\sigma z)} + \phi_g\right) exp\left(\frac{-z^2}{2}\right) dz,$$

for all $t = 1, \ldots, n_g$ and $g = 1, \ldots, k$ and where $\eta_g = (K_g, T, \sigma, \phi_g, a_g, b_g)$ is the parameter vector. Parameters of the model are interpreted as follows. The parameters T and σ are the same for all cells and genes in the population. The parameter T governs the duration of
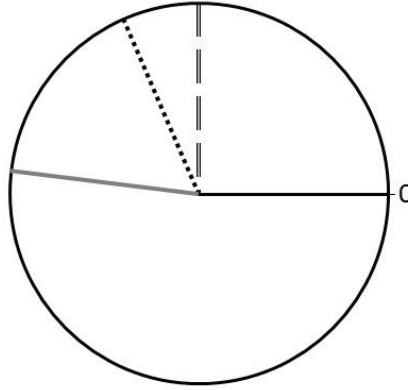
Figure 1.2   Peak costs plotted on a circle

the cell cycle, while σ measures the rate of attenuation in amplitude with each cycle (the larger σ the faster the decay in amplitude). The parameter ϕ_g is the angle of peak expression of gene g in the cell cycle with $\phi = 0$ being the point when cells are released. Parameter K_g is the amplitude of the first period and parameters a_g and b_g take into account possible drifts in the gene background expression level. The parameters T and σ are the same for all cells and genes in the population. The unconstrained estimators of all parameters of RPM, including the angular parameter ϕ_g, are obtained using nonlinear least squares methodology. Throughout this paper, we shall refer to the angular parameter ϕ_g as the phase angle due to its biological relevance.

Suppose we have k oscillatory variables (in the above tourism example we had three) and suppose for the ith variable the phase angle is denoted by ϕ_i, $i = 1, 2, \ldots, k$. Then, using the unconstrained estimators θ_i, $i = 1, 2, \ldots, k$ obtained from the RPM model, our goal is to conduct inference regarding the relative order of $\phi_1, \phi_2, \ldots, \phi_k$ around the unit circle. Suppose we travel around the circle in an anti-clockwise direction and suppose the angle ϕ_1 is followed by ϕ_2 which is followed by ϕ_3 etc. followed by ϕ_k which is finally followed by angle ϕ_1. Then we shall adopt the following notation (cf. Rueda et al. (2009), Fernández et al. (2012), 2012) to describe the relative order:

$$\phi_1 \preceq \phi_2 \preceq \ldots \preceq \phi_k \preceq \phi_1$$

It is important to note that the above order is invariant of the location of the pole of the circle. Alternatively, the above order is rotation invariant. For this reason Rueda et al. (2009) referred to the above order as an isotropic order. Focus of this paper is to discuss recent developments in the literature on the following problems: (a) To estimate $\phi_1, \phi_2, \ldots, \phi_k$ under the above order constraint using the unconstrained estimators of $\phi_i, i = 1, 2, \ldots, k$, obtained from RPM. (b) For a set of angular parameters, test the hypothesis that the above relative order is satisfied. (c) Using data from multiple experiments under different conditions test, whether the relative order among a set of phase angles is conserved across the conditions.

1.3   Estimation of angular parameters under order constraint

We begin by discussing the problem of estimating the phase angles ϕ_i, $i = 1, 2, \ldots, k$ under the order constraint $\phi_1 \preceq \phi_2 \preceq \ldots \preceq \phi_k \preceq \phi_1$ using the unconstrained estimators θ_i, $i = 1, 2, \ldots, k$ obtained from the RPM. The general idea of estimation resembles the analogous problem in the Euclidean space. Let $\mathcal{C} = \{\phi \in [0, 2\pi)^k : \phi_1 \preceq \phi_2 \preceq \ldots \preceq \phi_k \preceq \phi_1\}$. Suppose $\mathcal{C}^i = \{\phi \in [0, 2\pi)^k : 0 \le \phi_i \le \phi_{i+1} \le \ldots \le \phi_{i-1} \le 2\pi\}$, thus the pole of the unit circle is between the parameters ϕ_{i-1} and ϕ_i. Then we have $\mathcal{C} = \bigcup_{i=1}^k \mathcal{C}^i$.

For an estimator $\theta = (\theta_1, \theta_2, \ldots, \theta_k)'$ of a parameter $\phi = (\phi_1, \phi_2, \ldots, \phi_k)'$, the distance between the two is defined as the sum of circular errors (SCE) given by:

$$SCE(\theta, \phi) = d(\theta, \phi) = \sum_{i=1}^k r_i\{1 - cos(\theta_i - \phi_i)\},$$

where r_i represents a measure of concentration of θ_i about its modal direction (see Mardia and Jupp (2000) p. 17). Consequently, using the unconstrained estimator θ, the estimator of ϕ under the constraint $\phi \in \mathcal{C}$ is obtained by solving the following minimization problem:

$$min_{\phi \in \mathcal{C}} SCE(\theta, \phi) = min_{\phi \in \mathcal{C}} \sum_{i=1}^k r_i\{1 - cos(\theta_i - \phi_i)\}. \tag{1.1}$$

In the case of Euclidean space data where θ has a known diagonal covariance matrix, $\mathcal{C}$ is the simple order cone given by $\phi_1 \le \phi_2 \le \cdots \le \phi_k$ and SCE is replaced by the Euclidean distance (i.e. sum of squared errors), the corresponding minimization problem is called the isotonic regression. Typically the problem is solved using the pool adjacent violator algorithm (PAVA). The basic underlying idea of PAVA is that components of θ that violate the underlying relative order, are pooled or averaged so that the overall order is satisfied. To illustrate this, we consider the following toy example in the Euclidean space.

Example Suppose $\phi = (\phi_1, \phi_2, \phi_3) \in \mathbb{R}^3$ with $\phi_1 \le \phi_2 \le \phi_3$. Suppose the unconstrained sample means are given by $\theta_1 = 0.6$, $\theta_2 = 2.5$ and $\theta_3 = 1.5$. Since $\theta_2 > \theta_3$, therefore the order $\phi_1 \le \phi_2 \le \phi_3$ is violated. The PAVA would average the last two coordinates, yielding the $\widetilde{\phi}_1 = 0.6$, $\widetilde{\phi}_2 = \widetilde{\phi}_3 = (1.5 + 2.5)/2 = 2$ as the constrained estimates.

In the case when $\phi \in \mathcal{C}$, the unit circle, the solution to the minimization problem (1.1) is more complicated as noted in Rueda et al. (2009). Since (1.1) resembles the usual isotonic regression estimation of the Euclidean space data, Rueda et al. (2009) refer to the solution of (1.1) as circular isotonic regression estimator (CIRE). More precisely, the CIRE, denoted as $\widetilde{\phi}$, is given by:

$$\widetilde{\phi} = Argmin_{\phi \in \mathcal{C}} SCE(\theta, \phi). \tag{1.2}$$

Before we formally describe CIRE, we consider the following toy example to describe the calculation of CIRE geometrically. We remark that when considering angular data, the arithmetic means are not always appropriate for describing the average direction between a pair of angles. Instead one should use the angular mean direction (cf. Mardia and Jupp (2000), Rueda et al. (2009)).

Example Suppose $k = 3$ with $\phi_1 \preceq \phi_2 \preceq \phi_3 \preceq \phi_1$. Suppose the unconstrained estimates using the RPM are given (in radians) by $\theta_1 = 6$, $\theta_2 = 1.5$ and $\theta_3 = 2.5$ (see Figure 1.3).
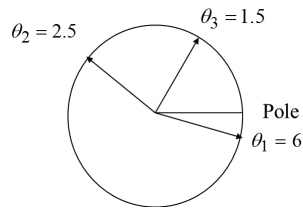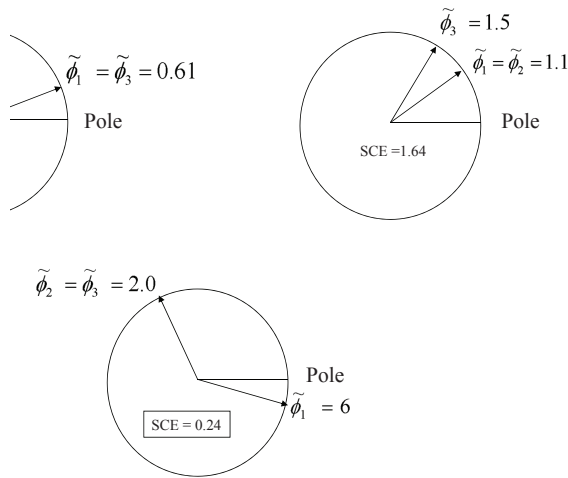
Figure 1.3   Unconstrained Estimates

Figure 1.4   Constrained Estimates

Clearly these estimates do not satisfy the desired order. In the Euclidean space example described above it was easy to identify the violator of the order and one could accordingly deal with it. However in the present case, since the data wrap around the circle, the violator may not be unique and one needs to explore all possibilities. If the violation is due to θ_1 and θ_3 then one would average these two and leave θ_2 as is. This would result in the constrained estimates given in the top left circle in Figure 1.4 with an SCE of 0.74. However, if the violation is due to θ_1 and θ_2 then one would average these two and leave θ_3 as is. This would result in the constrained estimates given in the top right circle in Figure 1.4 with an SCE of 1.64. Or the last possibility could be that θ_2 and θ_3 are in violation of the order. In which case we pool the estimates θ_2 and θ_3 and leave θ_1 as is, resulting in an SCE of 0.24. See bottom circle in Figure 1.4. Since this SCE is the smallest it is the CIRE.

An important observation to make from the above example is that in the case of circular data it is not enough to consider the adjacent violators of the order. This makes the problem computationally challenging. The main reason for this is that, unlike the arithmetic mean in the Euclidean space, the circular mean does not verify the Cauchy Mean Value property. Rueda et al. (2009) provide a general algorithm to derive CIRE and demonstrated that their algorithm is exact and computationally efficient, especially as the number of parameters increases. CIRE is implemented in the R package **isocir** (Barragán et al. (2013)). The solution to their algorithm is characterized in the following theorem.

Theorem 1.3.1. *The CIRE exists, is almost sure unique, and can be obtained from circular means of adjacent angles as,*

$$\widetilde{\phi}_g = Ave(S_{(i)}) \, for \, g = 1, ..., k, \, i = 1, ..., m,$$
$$with \, 0 < Ave(S_{(1)}) < Ave(S_{(2)}) < ... < Ave(S_{(m)}) < 2\pi,$$

where $(i)_{i=1}^m$ is a partition of $\{1, \ldots, k\}$, $Ave(S_i)$ are the circular mean directions for angles in $S_i = \{\theta_g, g \in (i)\}$, $(1), .., (m)$ are the so called level sets (cf. Robertson et al. (1988)), $n_{(i)} = \#(i)$ and $\sum_{i=1}^m n_{(i)} = n$.

In some situations, especially in cell-biology, one may be interested in partial orders of the following type:

$$\{\phi_1, \phi_2, \ldots \phi_{r_1}\} \preceq \{\phi_{r_1+1}, \ldots \phi_{r_2}\} \preceq \cdots \preceq \ldots \{\phi_{r_s+1}, \ldots \phi_k\} \preceq \{\phi_1, \phi_2, \ldots \phi_{r_1}\}. \tag{1.3}$$

In the above notation, angles within each set are not ordered but the angles one set precede the angles in the next set. Thus all angles in $\{\phi_1, \phi_2, \ldots \phi_{r_1}\}$ precede all the angles in $\{\phi_{r_1+1}, \phi_{r_1+2}, \ldots \phi_{r_2}\}$ and so on. This occurs when a biologist may hypothesize that, as a group, genes in a given set have to function before the genes in the next set function for the cell division cycle to proceed. He/she may not know the order of expression of genes within each set. Barragán et al. (2013) extended the CIRE methodology of Rueda et al. (2009) to estimate parameters under the above order constraint.

1.4   Inferences under Circular Restrictions in Von Mises Models

When dealing angular data, analogous to normal distribution on the real line, one typically uses the Von Mises distribution for performing inferences regarding the angular parameter (cf Mardia and Jupp). Accordingly, in this section we shall make a simplifying assumption that the unconstrained estimators θ_i, $i = 1, 2, \ldots, k$ are mutually independently distributed with $\theta_i \sim VM(\phi_i, \kappa)$, where VM stands for Von Mises distribution, ϕ_i denotes the angular mean direction and κ is the concentration parameter of the distribution. The probability density function (pdf) is given by:

$$g(x, \phi_i, \kappa) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(x - \phi_i)} \qquad x \in [0, 2\pi),$$

where I_0 is the modified Bessel function of first class and order zero. As noted earlier, there exists a large body of literature on statistical tests for angular data, especially under the Von Mises distribution (cf. Mardia and Jupp). However, until Fernandez et al. (2012) and

Barragan et al. (2013) there did not exist any formal literature on testing for order among angular parameters. Motivated by various applications, e.g. in social psychology, neurology, cell-biology etc., one may be interested in test the following hypotheses:

$$
\begin{aligned}
H_0 & \quad \phi_i, i = 1, ..., k \text{ follow a known order } O \\
H_1 : & \quad H_0 \text{ is not true.}
\end{aligned}
\tag{1.4}
$$

For example, O may be the circular order described earlier, i.e. $\phi_1 \preceq \phi_2 \preceq \ldots \preceq \phi_k \preceq \phi_1$. Under the above distributional assumptions, the CIRE of $(\phi_1, \phi_2, \ldots, \phi_k)'$ is the restricted maximum likelihood estimator (RMLE) of $(\phi_1, \phi_2, \ldots, \phi_k)'$ (Rueda et al. (2009)). From Theorem 1.3.1, we see that CIRE partitions the estimates into m level sets of consecutive coordinates on which $\widetilde{\phi}_i$ is constant.

Assuming κ is known, one may derive the likelihood ratio test (LRT) statistic T for hypotheses (1.4) as the angular distance between the unconstrained maximum likelihood estimator $(\theta_1, \theta_2, \ldots, \theta_k)'$ and the RMLE $(\widetilde{\phi}_1, \widetilde{\phi}_2, \ldots, \widetilde{\phi}_k)'$ which is given by:

$$
T = 2\kappa \sum_{i=1}^{k} \left(1 - \cos\left(\theta_i - \widetilde{\phi}_i \right) \right).
$$

Since in practice it is not easy to implement the LRT, Fernández et al. (2012) derived a conditional test (CT) by conditioning on the number of level sets m. Conditional tests have been well-studied in the case of order restricted inference for normal models (Robertson et al. (1988)) but unknown until Fernández et al. (2012) for Von Mises populations. The conditional test CT of Fernández et al. (2012) rejects the above null hypothesis whenever $T \geq c(m)$, where m is the number of level sets for $(\widetilde{\phi}_1, \widetilde{\phi}_2, \ldots, \widetilde{\phi}_k)'$ and $c(m)$ is chosen so that $P(\chi^2_{k-m} \geq c(m)) = \frac{\alpha}{1 - \frac{1}{(n-1)!}}$. Fernández et al. (2012) demonstrated that for large values of (κ, k), CT is an α level test (see theorem below).

Theorem 1.4.1. *Let $\phi^I = (\phi_1, ..., \phi_n)$, with $\phi_I = \pi/2, \phi_g = 3\pi/2$ for any $g \neq I$. Denote also as $(1), ..., (m)$ the level sets of $\widetilde{\Phi}$ and $R_m^n = \left\{ \theta \in [0, 2\pi)^n : \widetilde{\phi} \text{ has } m \text{ level sets} \right\}$.*

 (i) If $\phi = \phi^I$, then $P_{\phi^I} \left(T \geq c / R_m^n \right) \underset{\kappa \longrightarrow \infty}{\longrightarrow} P(\chi^2_{n-m} \geq c)$.

 (ii) For large κ, the level of the conditional test is attained at ϕ^I:
   $$P_{\phi^I} \left(T \geq c(m) \right) \underset{\kappa \longrightarrow \infty}{\longrightarrow} \alpha.$$

 (iii) If ϕ verifies the order O:
   $$P_\phi \left(T \geq c(m) \right) \underset{\kappa \longrightarrow \infty}{\longrightarrow} b \text{ with } b \leq G(q)\alpha \text{ and } G(q) \underset{q \longrightarrow \infty}{\longrightarrow} 1.$$

In practice, κ is usually unknown. In this case κ can be replaced by a consistent estimator $\widehat{\kappa}$, and accordingly T can be modified. By appealing to Mardia and Jupp (2000), pp. 87-89, $(\widetilde{\phi}_i, i = 1, 2, \ldots, k$, and $\widehat{\kappa}$ are approximately independent and furthermore

$$
\frac{\kappa}{\widehat{\kappa}} \overset{aprox.}{\rightsquigarrow} \chi^2_{q-1}.
$$

As a consequence we may approximate the distribution of CT by the central F distribution instead of the chi-squared distribution. The proof of the theorem and other theoretical details of CT are given in Fernández et al. (2012).

Barragán et al. (2013) extended the above methodology to test hypotheses regarding partial orders. More precisely, they extended the conditional test to test the following hypotheses:

$$H_0 \quad \phi_i, i = 1, ..., k \text{ follow a known partial order } O^*$$
$$H_1 : \quad H_0 \text{ is not true.} \tag{1.5}$$

where O^* may be the partial order appearing in equation (1.3).

1.5   The Estimation of a Common Circular Order from Multiple Experiments

Often data are available from multiple experiments or multiple sources and researchers are interested in estimating the common order among circular parameters. For example, using data obtained from multiple experiments on fission yeast (*S. pombe*), the yeast used in brewing alcohol, researchers are not interested in identifying periodically expressed genes but also interested in estimating their order of peak expression (see Oliva et al. (2005), Rustici et al. (2004), Peng et al. (2005)).

More precisely, our problem of interest is to determine the true relative order among k angular parameters $\phi_1, \phi_2 \ldots, \phi_k$ using the corresponding unconstrained estimators $\Theta_j = (\theta_{1j}, \; theta_{2j}, \ldots, \theta_{kj})'$, $j = 1, 2, \ldots, p$, from p independent experiments. Stacking these estimators we obtain the a $k \times p$ matrix $\Theta = (\Theta_1, ..., \Theta_p)$.

As in the Euclidean space data, combining data from multiple experiments to estimate a common parameter requires one to take into account variability between and within studies. However, since the underlying time-course data are usually based on a large number of time points, one may assume that the variability within experiments is negligible compared variability between experiments. Also it is important to recognize that in addition to estimating $\phi_1, \phi_2 \ldots, \phi_k$, we are more importantly interested in estimating their relative order.

The problem at hand resembles the classical problem of determining the "true" order or ranks among n objects using the ranks assigned by p independent "judges". For example, suppose there are k gymnasts competing in an event and there are p judges assigning ranks to each of the contestants. The goal is to estimate the true rank among the k contestants using the ranks assigned by the p judges. This is a well-studied problem in the Euclidean space (cf. Diaconis and Graham (1977), Borda (1781), Condorcet (1785), Schalekamp and Zuylen (2009)) and known to be NP-hard, see Bartholdi et al. (1989). Again, due to the underlying geometry, the Euclidean space based methods cannot be directly applied here. Barragán (2014) and Barragán et al. (2014) took the first step in addressing this problem for circular data as follows.

Let $\mathfrak{O}$ denote the set of all possible orders among k objects on a unit circle. Using data from the jth experiment, let $\widetilde{\Phi}_j^{(O)} = (\widetilde{\phi}_{1j}^{(O)}, \widetilde{\phi}_{2j}^{(O)}, \ldots, \widetilde{\phi}_{kj}^{(O)})'$ denote the CIRE under the circular order O. Then the distance between Θ_j and $\widetilde{\Phi}_j^{(O)}$ is given by:

$$d(\Theta_j, O) = SCE(\Theta_j, \widetilde{\Phi}_j^{(O)}) = \sum_{i=1}^{k} (1 - \cos(\theta_{ij} - \widetilde{\phi}_{ij}^{(O)})).$$

The average distance between Θ and the estimator of $\phi_1, \phi_2 \ldots, \phi_k$ based on the p independent experiments, called Mean Sum of Circular Errors (MSCE), is given by

$$d^*(\Theta, O) = MSCE(\Theta, \widetilde{\Phi}^{(O)}) = \sum_{j=1}^{p} \omega_j d(\Theta_j, O) \tag{1.6}$$

where ω_j is the weight associated with j^{th} experiment, which is related to the precision of the experiment j. For instance, assuming $\theta_{ij} \sim VM(\phi_{ij}, \kappa_j)$ with κ_j known, the weights may be defined as $\omega_j = \frac{\kappa_j}{\sum_{j=1}^{p} \kappa_j}$.

With this notation, Barragán (2014) and Barragán et al. (2014) restated the problem of estimating the optimum circular order $O^* \in \mathfrak{O}$ as the following minimization problem:

$$O^* = \arg\min_{O \in \mathfrak{O}} d^*(\Theta, O) = \arg\min_{O \in \mathfrak{O}} \sum_{j=1}^{p} \omega_j d(\Theta_j, O). \tag{1.7}$$

As done in the case of Euclidean space data (cf. Dwork et al. (2001a), Dwork et al. (2001b)) the methodology of Barragán (2014) and Barragán et al. (2014) consists of two steps as briefly outlined below. For more details one may refer to the above references. In the first step (Step E1) an initial approximate solution to the problem is obtained. This approximate solution is refined in the second step (Step E2) by smoothing out local "bumps" in the order.

Step E1 ($\widehat{O}^0$): In this step we cast the above optimization problem as a Traveling Salesman Problem (TSP) to obtain an approximate solution to (1.7). The TSP is well-studied in the graph theory literature (cf. Lawler et al. (1985), Reinelt (1994), Hahsler and Hornik (2011)) and is often used in numerous applications. Starting from a particular city, a salesman is required to visit each of the remaining $k - 1$ cities in his tour exactly once and then return to the city he started. The goal for the salesman is to determine the order in which he tours the cities so that total distance traveled by the salesman is shortest among all possible paths he can take. Even though this problem is considered to be computationally difficult, a large number of heuristics and exact methods are available in the literature. Some of these methods provide exact solutions when the number of cities is in tens of thousands and provide good approximations when the number of cities is in millions (Reinelt (1994)).

In our application, each experiment is represented by a graph where the objects are the cities/nodes (or estimated angles) and the length of the edges among nodes are the angular distances between the corresponding estimated angles in the experiment. There is a correspondence between tours in the graph an circular orders within the objects. For each experiment we have a distance matrix. We then aggregate (using means) the p matrices to summarize all the information in a single matrix. Finally, the heuristic algorithms implemented in R in the TSP package, Hahsler and Hornik (2011), are used to obtain the minimum length tour among nodes. The TSP solution results in an approximate circular order $\widehat{O}^0$. Not only does this strategy results in a very good approximate solution but it is also computational fast and efficient(see Barragán (2014) and Barragán et al. (2014)).

Step E2 ($\widehat{O}^*$): In this second step, Barragán (2014) and Barragán et al. (2014) fine tune the solution obtained in Step E1 by performing local smoothing to reduce the MSCE (1.6). Their solution is a modification of the *Local Kemenization* algorithm that was originally developed

by Dwork et al. (2001a) for the Euclidean data. This modification is called *Circular Local Minimization*. It consists of checking each consecutive triple (i, k, l) of adjacent cities in the order determined in $\widehat{O}^0$. The MSCE between the new order with the permutation and the data is computed. If the MSCE for the new circular order is smaller the candidate order is appropriately updated. Each time a triple is permuted, the previous ones are checked back again to ensure that no further improvement in the order is possible.

1.6   Application: Analysis of cell-cycle gene expression data

A cell division cycle in a normal eukaryotic cell consists of four phases, namely, G1, S, G2 and M phases. In the G1 phase the cell rests and grows. This is also the first check point phase where any DNA damage is detected. The G1 phase is followed by the S phase where DNA replication occurs. Following S phase, cells go through a second check point called the G2 phase to detect damage. In a normal setting, cells that cannot be repaired are not allowed to proceed to mitosis (M phase) where the cells divide. Genes involved in cell division cycle (called cell cycle genes) have a periodic expression consistent with period consistent with the duration of cell division cycle. Such genes attain peak expression just before their biological function (Jensen et al. (2006)). For a given organism, biologists are typically interested in (a) identifying cell cycle genes, (b) identifying the time to peak expression (i.e. phase angle ϕ) of a cell-cycle gene, (c) comparing the phase angles of cell cycle genes across different experimental conditions or different organisms(cf. Bähler (2005), Jensen et al. (2006), Fernández et al. (2012)). A useful database containing results from various cell cycle microarray experiments is available at www.cyclebase.org, henceforth referred as cyclebase. Cyclebase provides estimates of the peak expressions using a simple mathematical model and data from a single experiment.

To answer questions such as the above, researchers conduct long series time course gene expression studies measuring gene expressions of thousands of genes over several time points, long enough to include at least one full cell division cycle (if not more). We illustrate the methodology described in this paper using the 34 cell cycle genes *S. pombe* genes and their *S. cerevisiae* orthologs/paralogs described in Fernández et al. (2012). We used time-course data available on ten experiments conducted on *S. pombe* in three labs (five by Rustici et al. (2004), three Oliva et al. (2005) and two by Peng et al. (2005)) and six experiments conducted on *S. cerevisiae* in three labs (one experiment each by Cho et al. (1998) and de Lichtenberg et al. (2005), and two experiments each by Pramila et al. (2006) and 2 by Spellman et al. (1998)). For each gene i, $i = 1, 2, \ldots, 34$ within the jth experiment, $j = 1, 2, \ldots, 16$, we fitted the RPM to obtain the unconstrained phase angle estimates θ_{ij} for the 34 genes in the 16 experiments. Results of the estimated phase angles for the 34 genes for *S. pombe* and their *S. cerevisiae* orthologs/paralogs for the 16 experiments considered are not provided here in order to save space but can be obtained from the authors on request.

We assumed that $\theta_{ij} \sim^{independent} VM(\phi_{ij}, \kappa_j)$, where ϕ_{ij} is the true unknown phase angle for the ith gene in the jth experiment. Note that κ_j is experiment specific and not gene specific. Thus κ_j reflects the uncertainty associated with the jth experiment and phase angles of all genes within that experiment are estimated with same uncertainty. As noted earlier, since for each gene its phase angle is estimated using RPM with a reasonably large number of time points, we assume that uncertainty associated within gene is ignorable compared to the overall uncertainty associated with the experiment. The parameter κ_j is estimated using the

random effects model for circular data described in Fernandez et al. Using the methodology of Barragán (2014) and Barragán et al. (2014) described in Section 1.4 and the phase angle estimates θ_{ij} of 34 genes obtained above for Rustici et al. (2004), Oliva et al. (2005) and Peng et al. (2005) data, we obtained the common global order among the phase angles of the 34 *S. pombe* genes. Using the estimated order we obtained the constrained estimates of the phase angles using CIRE for the 34 *S. pombe genes*. These estimates along with the estimates according to cyclebase are given in Table 1.1. Similarly, using the phase angle estimates of the 34 *S. cerevisiae* orthologs/paralogs, based on the data from Cho et al. (1998), de Lichtenberg et al. (2005), Pramila et al. (2006) and Spellman et al. (1998), we estimated their global order along with their constrained estimates using CIRE (Table 1.1).

Using the conditional test CT we shall compare the global order of phase angles of the above 34 genes determined by our methodology with the order described in cyclebase for the two species of yeast. Within each species, for each experiment we tested the null hypothesis that the global order holds against the alternative that the null is not true using the CT. Thus for each experiment we obtain one p-value based on the CT. Within each species, we then combined p-values from all the experiments (i.e. $p = 10$ in the case of *S. pombe* and $p = 6$ in the case of *S. cerevisiae*) using Fisher's method to obtain $L = -\sum_{j=1}^{p} log(p - value_j)$, where $p - value_j$ is the p-value obtained for experiment j. If the $p - values$ are independently and uniformly distributed in the interval $(0, 1)$, then $2L$ is distributed as a central χ^2 random variable with p degrees of freedom. Then, if l is the observed value for L, $Fp - value = pr(\chi_p^2 > 2l)$ yields a single value to test the null hypothesis. The resulting p-values for each species and the orders considered for each species are given in Table 1.2. From the table we see that the orders estimated using the methodology proposed in Section 1.4 have a much higher p-value than those appearing in cyclebase. This happened not only for the global $Fp - value$ but for the almost all the $p - value_j$ values, suggesting that the global order provided by the cyclebase for the two species should be rejected and that the order derived by the methodology of Barragán (2014) and Barragán et al. (2014) described in Section 1.4 is plausible for the two species.

The disagreement between the order specified by the cyclebase and the order specified by the methodology of Barragán (2014) and Barragán et al. (2014) can partly be explained by noting that there are some major differences in the estimates of the phase angles between cyclebase and CIRE (identified in bold face) for some genes as seen in Table 1.1. Among them, the noticeable ones are the *S. pombe* gene *mcp1* and the *S. cerevisiae* gene *SST2*. According to cyclebase, both genes have a very high periodicity rank (i.e. have a poor periodic expression) and hence are likely to have less precise estimates of phase angles and hence not surprising that the two methods disagree in their phase angle estimates. (For this reason these genes are dropped from any further study.) Since our estimator of the global order uses information from all experiments, while taking into consideration the uncertainties associated with each experiment, we believe that our estimator of the global order is more reliable.

Since the CIRE estimators have common values for some genes (those appearing in the same level set), they also yield a partial order among the genes. The partial orders given by cyclebase and by the CIRE estimator for *S. cerevisiae* appear in Table 1.3. In that table we can see that there is no big discrepancy among the two partial orders. The most noticeable one is perhaps that of gene MOB1 which also has a high periodicity rank.

Now, we illustrate the methodology to determine a common partial order among the two

Table 1.1   Cyclebase and CIRE phase angles estimates for the two species

CEREVISIAE	CycleB	CIRE	POMBE	CycleB	CIRE
HTZ1	0.57	0.03	pht1	6.09	6.09
HHF1	6.16	0.03	htb1	6.22	6.22
HTA2	0.00	0.03	hta2	0.00	0.00
HTB2	0.00	0.03	hhf1	0.00	0.00
HHT2	6.09	0.03	hht3	0.06	0.06
HHT1	6.22	0.03	h3_3	0.06	0.06
KIP3	0.38	0.38	klp5	4.78	4.76
FKH1	0.63	0.63	fkh2	4.59	4.76
SWI5	1.57	1.57	ace2	4.71	4.76
BUD4	1.57	1.57	mid2	5.40	5.07
CDC5	1.57	1.78	plo1	4.27	4.76
CHS2	1.88	1.78	chs2	4.71	4.76
MYO1	1.88	1.78	myo3	4.65	4.76
HOF1	1.95	1.88	cdc15	4.71	4.76
MOB1	1.82	1.88	mob1	5.03	5.07
ASE1	1.88	1.88	**mcp1**	**3.83**	**4.76**
CDC20	2.26	2.26	slp1	4.65	4.76
KIN3	2.58	2.58	fin1	4.96	5.07
DBF2	2.70	2.70	sid2	4.78	4.76
CDC6	3.58	3.83	cdc18	4.90	5.07
PST1	3.77	3.83	SPAC1705_03C	4.65	4.76
DSE4	4.15	3.83	eng1	5.15	5.07
SST2	**3.14**	**5.01**	rgs1	4.78	4.76
RFA1	4.96	5.01	ssb1	5.22	4.76
MRC1	5.03	5.01	mrc1	5.09	4.76
SMC3	5.03	5.01	psm3	5.09	4.76
RNR1	5.03	5.01	cdc22	5.22	4.76
MSH6	5.03	5.01	msh6	5.09	5.07
POL1	5.03	5.01	pol1	5.09	5.07
RAD51	5.09	5.01	rhp51	4.96	4.76
MCD1	5.09	5.01	rad21	4.96	5.07
POL2	5.15	5.01	pol2	4.65	4.76
CLN2	5.15	5.01	cig2	5.09	4.76
SWE1	5.47	5.01	mik1	5.03	5.07

Table 1.2   MSCE and Fp-values for the 34 core set genes considered

Species	Order	MSCE	Fp-value
POMBE	Estimated order	0.06168913	0.8443571
POMBE	Cyclebase	0.09142278	3.616812e-07
CEREVISIAE	Estimated order	0.0281629	0.1659825
CEREVISIAE	Cyclebase	0.08753569	8.646897e-28

Table 1.3 Partial orders for *S. cerevisiae genes*

Cyclebase partial order	
Phase	Genes
G1/S	{HISTONES}$\preceq$
S/G2	{KIP3}$\preceq${FKH1}$\preceq${SWI5,**BUD4**}$\preceq${CDC5}$\preceq${**MOB1**}$\preceq$ {CHS2,**MYO1**}$\preceq${HOF1}$\preceq$
G2/M	{CDC20}$\preceq${**KIN3**}$\preceq${DBF2}$\preceq$ {**CDC6**}$\preceq${PST1}$\preceq$
M/G1	{**DSE4**}$\preceq${RFA1}$\preceq${MRC1,SMC3,RNR1,MSH6,POL1}$\preceq$ {**RAD51**,MCD1} $\preceq$ {POL2,CLN2}$\preceq${SWE1}$\preceq$
G1/S	{HISTONES}
CIRE partial order	
Phase	Genes
G1/S	{HISTONES}$\preceq$
S/G2	{KIP3}$\preceq${FKH1}$\preceq${SWI5,**BUD4**}$\preceq${CHS2,CDC5,**MYO1**}$\preceq$ {HOF1,**MOB1**}$\preceq$
G2/M	{CDC20}$\preceq${**KIN3**}$\preceq${DBF2}$\preceq$
M/G1	{**CDC6**,PST1,**DSE4**}$\preceq$ {RFA1,MRC1,SMC3,RNR1,**RAD51**,POL2,CLN2,MSH6,... ...POL1,MCD1,SWE1}$\preceq$
G1/S	{HISTONES}

species of yeasts using a subset of orthologs/paralogs by dropping genes that have either poor periodicity in at least one of the two species (cdc18 and eng1) or by dropping genes that were considered to violate the common order according to Fernández et al. (2012) (mid2, myo3, mob1, fin1, rhp51) and the corresponding *S. cerevisiae* ortholog/paralogs appearing in bold in Table 1.3). The partial orders obtained from cyclebase and the CIRE for the remaining 25 *S. pombe* genes are summarized in Table 1.4. According to cyclebase {msh6, pol1, rad21, mik1} are activated before {ssb1, cdc22}, however, based on our methodology, {ssb1, cdc22} are activated before {msh6, pol1, rad21, mik1}. It is interesting to note from Tables 1.3 and 1.4 that the partial orders derived by our methodology is satisfied by both species of yeast. Furthermore, this order is also satisfied by other previously published results (see Fernández et al. (2012)).

The methodology developed in Barragán (2014) and Barragán et al. (2014) are also useful to study phases of genes across multiple species. Thus the methodology developed in Barragán (2014) and Barragán et al. (2014) provides a general methodology to discover order among cell-cycle genes and subsequently allows biologists to explore new hypotheses regarding functional relationships and interactions among various cell cycle genes.

In general, the circular order restricted inference methods developed in Rueda et al. (2009), Fernández et al. (2012), Barragán et al. (2013), Barragán (2014) and Barragán et al. (2014) provide a general framework and tools for cell biologists to discover new biology.

Table 1.4   Partial orders for *S. pombe genes*

Cyclebase partial order
{HISTONES}$\preceq$ {plo1}$\preceq${fkh2}$\preceq${slp1,SPAC1705_03C,pol2}$\preceq$ {ace2,chs2,cdc15}$\preceq${klp5,sid2}$\preceq${rad21}$\preceq${mik1}$\preceq$ {mrc1,psm3,cig2, msh6,pol1}$\preceq${ssb1,cdc22}$\preceq$ {HISTONES}

CIRE partial order
{HISTONES}$\preceq$ {klp5,kfh2,ace2,plo1,chs2,cdc15,slp1,... ...sid2,SPAC1705_03C,ssb1,mrc1,psm3,cdc22,pol2,cig2}$\preceq$ {msh6,pol1,rad21,mik1}$\preceq$ {HISTONES}

1.7   Concluding remarks and future research

In this paper we discussed the current and ongoing research on the estimation and testing hypotheses regarding ordered parameters on a unit circle using data from multiple experiments. Although we illustrated these methods using a data from cell-biology, as described in the introduction, these methods are broadly applicable in a variety contexts including, evolutionary psychology (Russell (1980), De Quadros-Wander and Stokes (2007)), motor behavior (Baayen et al. (2012)) or circadian biology, etc.

System biologists are often interested in developing gene networks to describe inter-relationships among various genes. Many commercial software such as QIAGEN's Ingenuity®Pathway Analysis (IPA®, QIAGEN Redwood City, www.qiagen.com/ingenuity) attempt to provide such networks using curated data. However, most of those networks are based on static data. They do not take into account the temporal component in the data. However, cell division cycle is a dynamic process with each time point a collection of cell cycle genes (and others) interact and they impact on the genes that express at a later time point. In methodologies summarized in this paper describe temporal order among cell-cycle genes but it would be useful to develop dynamic networks among a collection of cell cycle genes based on the order information provided by the methods described here.

Constrained inference methods will have a natural role in other applications involving circular data, such as regression models for angular data described in Fisher and Lee (1992), Lund (1999), Downs and Mardia (2002), Kato et al. (2008) or Kato and Jones (2010). In an ongoing research project with Professor Mardia, we are exploring piecewise circular-circular regression model under constraints which may have applications in cell-biology. For instance, such models would be useful to relate phase angles of cell-cycle genes from different species or experimental groups.

All the methodology presented reviewed here are available in R language. Barragán et al. (2013) have developed a package called **isocir** (isotonic inference for circular data) which is available in the CRAN R Core Team (2014). The last version released contains CIRE and cond.test as principal functions. CIRE executes the algorithm developed in Rueda et al. (2009) to find the CIRE (1.2). The R objects called SEXP are used in C++ to improve efficiency and execution time. The function cond.test executes the conditional test described in Fernández et al. (2012) for the hypotheses (1.5). The methodology proposed to deal with the minimization problem (1.7) has also been implemented in R language as part of the new version of the **isocir** package.

References

Baayen C, Klugkist I and Mechsner F 2012 A test for the analysis of order constrained hypotheses for circular data. *Journal of Motor Behavior* **44**(5), 351–363.

Bähler J 2005 Cell-cycle control of gene expression in budding and fission yeast. *The Annual Review of Genetics* **39**, 69–94.

Barragán S 2014 *Procedimientos estadísticos para modelos circulares con restricciones de orden aplicados al anlisis de expresiones de genes*. Universidad de Valladolid. Ph.D. Dissertation.

Barragán S, Fernández M, Rueda C and Peddada S 2013 isocir: An r package for constrained inference using isotonic regression for circular data, with an application to cell biology. *Journal of Statistical Software* **54**(4), 1–17.

Barragán S, Rueda C, Fernández M and Peddada S 2014 Statistical framework for determining the temporal program in an oscillatory system. *Preprint*.

Bartholdi J, Tovey C and Trick M 1989 Voting schemes for which it can be difficult to tell who won the election.. *Social Choice Welfare* **6**, 157–165.

Borda J 1781 *Memorie sur les elections au scrutin*. Historie de l Academie.

Cao Y, Chen A, Jones R, Radcliffe J, Dietrich K, Caldwell K, Peddada S and Rogan W 2011 Efficacy of succimer chelation of mercury at background exposures in toddlers: a randomized trial. *The Journal of Pediatrics* **158**(3), 480–485.

Cho R, Campbell M, Winzeler E, Steinmetz L, Conway A, Wodicka L, Wolfsberg T, Gabrielian A, Landsman D, Lockhart D and Davis R 1998 A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol Cell.* **2**(1), 65–73.

Conaway M, Dunbar S and Peddada S 2004 Designs for single or multiple agent phase i trials. *Biometrics* **60**, 661–669.

Conde D, Fernández M, Rueda C and Salvador B 2012 Classification of samples into two or more ordered populations with application to a cancer trial. *Statistics in Medicine* **31**(28), 3773–3786.

Conde D, Salvador B, Rueda C and Fernández M 2013 Performance and estimation of the true error rate of classification rules built with additional information. an application to a cancer trial. *Statistical Applications in Genetics and Molecular Biology* **12**(5), 583–602.

Condorcet MJ 1785 *Essai sur l application de l a nalyse a la probabilite des decisions rendues a la pluralite des voix*.

de Lichtenberg U, Wernersson R, Jensen T, Nielsen H, Fausboll A, Schmidt P, Hansen F, Knudsen S and Brunak S 2005 New weakly expressed cell cycle-regulated genes in yeast. *Yeasts* **22**(5), 1191–1201.

De Quadros-Wander S and Stokes M 2007 The effect of mood on opposite-sex judgments of males' commitment and females' sexual content. *Evolutionary Psychology* **4**, 453–475.

Diaconis P and Graham R 1977 Spearmans footrule as a measure of disarray. *J. Roy. Statisti. Soc. Ser. B* **39**(2), 262–268.

Downs T and Mardia K 2002 Circular regression. *Biometrika* **89**(3), 683–697.

Dwork C, Kumar R, Naor M and Sivakumar D 2001a Rank aggregation methods for the web. *Proceedings of the 10th International World Wide Web Conference* pp. 613–622.

Dwork C, Kumar R, Naor M and Sivakumar D 2001b Rank aggregation revisited. *Manuscript*.

Fernández M, Rueda C and Peddada S 2012 Identification of a core set of signature cell cycle genes whose relative order of time to peak expression is conserved across species. *Nucl. Acids Res.* **40**(7), 2823–2832.

Fisher N 1993 *Statistical Analysis of Circular Data*. Cambridge University Press.

Fisher N and Lee A 1992 Regression models for an angular response. *Biometrics* **48**, 665–677.

Forgas J 1998 On being happy and mistaken: mood effects on the fundamental attribution error. *Journal of Personality and Social Psychology* **75**(2), 318–331.

Hahsler M and Hornik K 2011 *Traveling Salesperson Problem (TSP)*. R package version 1.0-6.

Hoenerhoff M, Pandiri A, Snyder S, Hong H, Ton T, Peddada S, Shockley K, Chan P, Rider C, Kooistra L, Nyska A and Sills R 2012 Hepatocellular carcinomas in b6c3f1 mice treated with ginkgo biloba extract for two years differ from spontaneous liver tumors in cancer gene mutations and genomic pathways. *Toxicologic Pathology*.

Hughes M, DiTacchio L, Hayes K, Vollmers C, Pulivarthy S, Baggs J, Manda S and Hogenesch J 2009 Harmonics of circadian gene transcription in mammals. *PLoS Genetics*.

Jensen J, Jensen T, Lichtenberg U, Brunak S and Bork P 2006 Co-evolution of transcriptional and post-translational cell-cycle regulation. *Nature* **443**, 594–597.

Kato S and Jones M 2010 A family of distributions on the circle with links to, and applications arising from, möbius transformation. *Journal of the American Statistical Association* **105**(489), 249–262.

Kato S, Shimizu K and Shieh G 2008 A circular-circular regression model. *Statistica Sinica* **18**, 633–645.

Lawler E, Lenstra J, Rinnooy KA and Shmoys D 1985 *The Traveling Saleman Problem*. John Wiley and Sons.

Liu D, Umbach D, Peddada S, Li L, Crockett P and Weinberg C 2004 A random periods model for expression of cell-cycle genes. *PNAS* **101**(19), 7240–7245.

Lund U 1999 Least circular distance regression for directional data. *Journal of Applied Statistics* **26**(6), 723–733.

Mardia K and Jupp P 2000 *Directional Statistics*. John Wiley & Sons.

Mechsner F, Kerzel D, Knoblich G and Prinz W 2001 Perceptual basis of bimanual coordination. *Nature* **414**, 69–73.

Oliva A, Rosebrock A, Ferrezuelo F, Pyne S, Chen H, Skiena S, Futcher B and Leatherwood J 2005 The cell-cycle-regulated genes of schizosaccharomyces pombe. *Plos. Biology* **3**, 1239–1260.

Oullier O, Bardy B, Stoffregen T and Bootsma R 2002 Postural coordination in looking and tracking tasks. *Human Movement Science* **21**, 147–167.

Peddada S, Dinse G and Kissling G 2007 Incorporating historical control data when comparing tumor incidence rates. *Journal of the American Statistical Association* **102**, 1212–1220.

Peddada S, Dunson D and Tan X 2005 Estimation of order-restricted means from correlated data. *Biometrika* **92**, 703–715.

Peddada S, Lobenhofer L, Li L, Afshari C, Weinberg C and Umbach D 2003 Gene selection and clustering for time-course and dose-response microarray experiments using order-restricted inference. *Bioinformatics* **19**, 834–841.

Peng X, Karuturi R, Miller L, Lin K, Jia Y, Kondu P, Wang L, Wong L, Liu E, Balasubramanian M and Liu J 2005 Identification of cell cycle-regulated genes in fission yeast. *The American Society for Cell Biology* **16**, 1026–1042.

Perdivara I, Peddada S, Miller F, Tomer K and Deterding L 2011 Mass spectrometric determination of igg subclass-specific glycosylation profiles in siblings discordant for myositis syndromes. *Journal of Proteome Research* **10**, 2969–2978.

Posner J, Russell J and Peterson B 2005 The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology* **17**, 715–734.

Pramila T, Wu W, Miles S, Noble W and Breeden LL 2006 The forkhead transcription factor hcm1 regulates chromosome segregation genes and fills the s-phase gap in the transcriptional circuitry of the cell cycle. *Genes Dev.* **22**(16), 2266–2278.

R Core Team 2014 *R: A Language and Environment for Statistical Computing* R Foundation for Statistical Computing Vienna, Austria. ISBN 3-900051-07-0.

Reinelt G 1994 *The Traveling Salesman. Computational Solutions for TSP Applications*. Springer-Verlag.

Robertson T, Wright F and Dykstra R 1988 *Order Restricted Statitical Inference*. John Wiley & Sons.

Rueda C, Fernández M and Peddada S 2009 Estimation of parameters subject to order restrictions on a circle with application to estimation of phase angles of cell-cycle genes. *Journal of the American Statistical Association* **104**(485), 338–347.

Russell J 1980 A circumplex model of affect. *J. of Personality and Social Psychology* **39**(6), 1161–1178.

Rustici G, Mata J, Kivinen K, Lio P, Penkett C, Burns G, Hayles J, Brazma A, Nurse P and Bahler J 2004 Periodic gene expression program of the fission yeast cell cycle. *Nature Genetics* **36**, 809–817.

Schalekamp F and Zuylen A 2009 Rank aggregation: Together we are strong. *In Proc. of 11th ALENEX* pp. 38–51.

Schlosberg H 1952 The description of facial experssions in ternos of two dimensions. *Journal of Experimental Psychology* **44**, 229–237.

Silvapulle M and Sen P 2005 *Constrained Statistical Inference: Order, Inequality, and Shape Constraints*. Wiley.

Spellman P, Sherlock G, Zhang M, Iyer V, Anders K, Eisen M, Brown P, Botstein D and Futcher B 1998 Comprehensive identification of cell cycle-regulated genes of the yeast saccharomyces cerevisiae by microarray hybridization. *Mol. Biol. Cell.* **9**(12), 3273–3297.

Whitfield M, Sherlock G, Saldanha A, Murray J, Ball C, Alexander K, Matese J, Perou, C.M.and Hurt M, Brown P and Botstein D 2002 Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol Biol Cell.* **13**, 1977–2000.