Some Aspects of Geometry Driven Statistical Models

Kanti V. Mardia

Department of Statistics, University of Leeds and
Department of Statistics, University of Oxford.

1 New demands

Non-Euclidean data driven mainly by underlying geometry arise in a variety of important new applications generated by areas such as bioinformatics, meteorology, new energy sources, and finance. In these applications, variables are observed on several different manifolds: circle, sphere, cylinder, space of orthogonal matrices, Stiefel manifold, Grassmann manifold, shape spaces. Obvious cases for circular data include the 24-hour clock, general calendar measurement, and compass direction.

Models in directional statistics have been built on these spaces starting from the von Mises distribution as early in 1919, followed by the Fisher distribution in 1953. Mardia and Jupp (2000) and Jammalamadaka and SenGupta (2001) have summarised the state of knowledge of directional statistics up to the time of their publication whereas Dryden and Mardia (1998) have summarised shape statistics. Although some multivariate models have been developed recently (eg., Kent et al, 2006; Mardia et al, 2008 and the references below in Section 7), the area really needs a major boost to meet the new demands of modern applications.

2 History

Perhaps one of the landmark papers to initiate this area is of Fisher (1953). He has outlined why the actual topology has to be taken into account in general though his main focus in that paper was spherical data.

1. The theory of errors was developed by Gauss primarily in relation to the needs of astronomers and surveyors, making rather accurate angular measurements.

2. The actual topological framework of such measurements, **the surface of a sphere, is ignored in the theory as developed**, with a certain gain in simplicity.

3. It is, therefore, of some little mathematical interest to consider how the theory would have had to be developed if the observations under discussion had in fact involved errors so large that the actual topology had had to be taken into account.

4. **The question is not, however, entirely academic, for there are in nature vectors with such large natural dispersions.**

   (The bold letters and the listing are done for this paper.)

Directional Statistics has a curious history.

1. 1956-1970 Geoff Watson and Michael Stephen made several contributions.

2. 1970-1980 Its heyday was when the subject came into the limelight.

3. 1972, partly with my directional book.
   1975 more so with my discussion paper in the Journal of the Royal Statistical Society.

4. 1990-1999 almost a lull in the subject.

5. **2000 or so after a resurgence of interest led to new strides being taken, partly through the recognition by Image Analysts and Life Scientists of its importance.**

6. This momentum now continues. Figures 1 and 2 capture this citation trend of Mardia (1975) and Mardia and Jupp (2000).

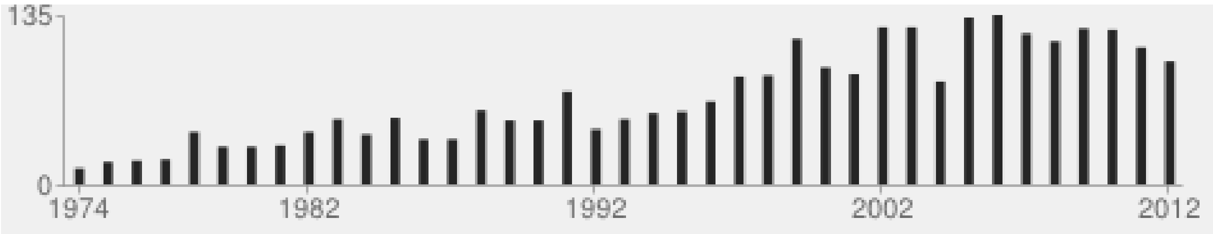

Figure 1: Mardia, 1975, Statistics of directional data, J Royal Statist. Soc. B, with Discussion, August 2012: CITATIONS 2777

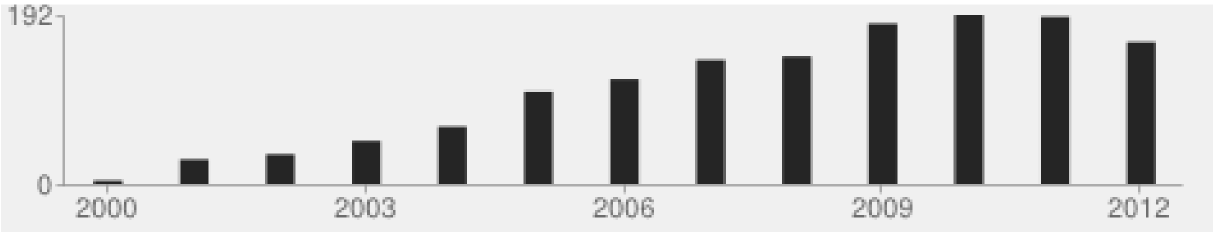

Figure 2: Mardia and Jupp, 2000, Directional Statistics, Wiley, August 2012: CITATIONS 1427

3 Strategies for constructing models

We now give important ways in which useful models have been constructed so far. We first outline how the models in multivariate analysis can be characterised in terms of increasing level of sophistication.
1. Univariate distributions.
2. Isotropic multivariate normal distribution.
3. Full multivariate normal distribution.
4. Skew multivariate distributions.
5. More general non-normal distributions.

Various developments in directional statistics have followed the same pattern...sometimes motivated by a data, at other times by extensions as a mathematical statistics or a combination of two.
1. The von Mises distribution on circle.

2. Independent von Mises distributions on torus.
3. The bivariate sine distribution (Singh et al., 2002) followed by the multivariate sine distributions on torus (Mardia et al., 2008).

Another multivariate path of development starting from the von Mises distribution on the circle has led to the following spherical distributions.
1.The Fisher distribution and the Watson distribution (isotropic).
2.The Fisher distribution followed by the Kent distribution, and the Watson distribution followed by the Bingham distribution to allow for the anisotropy.
3. In shape spaces for the planar configuration, the complex Watson distribution followed by the complex Bingham distribution (with complex symmetry) followed by the complex Bingham quartic distribution (anisotropic; Kent et al (2006)).

The area of Spherical distributions is well matured but the distributions on torus are still evolving. However, we treat the multivariate sine distribution as the front runner; the approach to multivariate normal for the concentrated case is direct; the global unimodality condition has been proven by Mardia and Voss (2013). One of the main challenges is to develop plausible skew distributions on torus as it seems the circular skew distributions have good candidates.

From the theoretical point of view, there are four basic approaches to model building in directional and shape statistics, which may be termed embedding, tangent, intrinsic and wrapping approach. Indeed, the wrapping approach has been a powerful tool for the circle where a distribution on the line is wrapped on the circle. Wrapping the multivariate normal distribution leads to a distribution on torus and is a plausible competitor to the multivariate sine distribution (eg. Jona–Lasinio et al 2012). In shape analysis, model building was initiated using an embedding approach starting from so called the Mardia-Dryden model. It seems the first three stages of multivariate analysis in directional statistics are now largely achieved and we are now entering the next phase of the skew distributions in directional statistics. This proposal aims to push forward the development of skew directional models, circular skew distributions (Section 3), skew distributions on Torus (Section 4), and spherical skew distributions (Section 5). The need for the first two areas has already been established by various real examples in recent papers. The third case here is also relevant to shape space where data involves landmarks and "edgels".

4 Circular skew distributions

There have been various approaches for constructing the univariate circular skew distribution, mainly around the von Mises distribution. The circular case is dominated by the von Mises distribution of which the probability density function (pdf) is

$$VM(\theta;\mu,\kappa) = \{2\pi I_0(\kappa)\}^{-1}\exp\{\kappa\cos(\theta-\mu)\},\ \ 0<\theta,\mu<2\pi,\kappa>0,$$

where $I_0(\kappa)$ is a Bessel function, μ is the mean direction and κ is the concentration parameter. A major consideration has been to preserve unmodality. Another consideration is to have a tractable normalizing constant. Jones and Pewsey (2012) have given an excellent review of the current state of play in this area. Let $f(x)$ be a pdf of some well-known distribution such as von Mises. There are at least two general principles: the first is Azzalini's asymmetry principle and second is Jones-Pewesey's composition principle. Azzalini(2013) and Mardia (2012) have given an overview. We repeat a few paragraphs from Mardia (2012):

Azzalini Type I Approach. In this we start from a symmetric distribution and use reweighting to introduce skewness. For the von Mises case, the pdf can be taken as (with $-1 \leq \nu \leq 1$)

$$f_\mu(\theta) = VM(\theta; \mu, \kappa)(1 + \nu \sin(\theta - \mu))$$

where the weights are from the cardioid distribution (eg, Umbach and Jammalamadaka, 2009).

Azzalini Type II Approach. In this we start with one of the conditional characterizations of Azzalini. Consider his characterization where x and y are bivariate normal with zero means, unit variances, and correlation ρ. Then the distribution of

$$x|y > 0 \quad \text{is} \quad \mathrm{SN}(\rho/\sqrt{1-\rho^2}),$$

where $\mathrm{SN}(a)$ is the standard Azzalini distribution with pdf $2\phi(x)\Phi(ax)$; $\phi(.)$ is the pdf of the standard normal and $\Phi(.)$ is the distribution function of the standard normal. Consider now the bivariate sine distribution (Singh et al 2002) with the pdf

$$\mathrm{Const}\exp\{\kappa \cos\theta + \kappa \sin\phi + \lambda \sin\theta \sin\phi\}.$$

Then it can be seen that the pdf of θ given $\theta < \phi < \pi$ is proportional to

$$\mathrm{VM}(\theta; 0, \kappa) \int_0^\pi \exp\{\kappa \cos\phi + \lambda \sin\phi \sin\theta\} \mathrm{d}\phi.$$

This integral can be expressed in terms of the distribution function of the von Mises distribution and, further for small κ and λ, it reduces to the pdf $f_\mu(\theta)$ given above.

Composition Approach. Jones and Pewsey (2012) have developed what we have called a composition approach. This is a powerful approach with various good properties including a remarkable property related to the normalising constant. But unfortunately, it can be shown that their normalising constant property does not hold for the bivariate case, eg. for the bivariate sine model.

5 Skew distributions on torus

This area seems to be wide open.

Extensions. Consider the Azzalini Type 1 Approach. Azzalini and Dalla Valle (1996) have extended SN to the multivariate case. Without any loss of generality, consider the bivariate case. In this case, the pdf of (x, y) is given by

$$2\phi(x, y; \Omega)\Phi(ax + by)$$

where $\phi(.)$ is the pdf of bivariate normal with zero means and correlation matrix Ω, and $\Phi(.)$ is the distribution function of the standard normal. We can extend the method to the bivariate circular case considering the effect of say $ax + by$ where x and y are angles when $a = -1 \; or \; = 1$ and $b = -1 \; or \; = 1$ and where the bivariate normal distribution is replaced by the bivariate sine distribution (given above). This idea extends itself to the multivariate case. There are various characterizations of Azzalini's skew normal distribution $\mathrm{SN}(\alpha)$, and their extension to the circular case could lead to different distributions as we have seen in the previous section.

6 Spherical skew distributions and edgels

As it was pointed out in Mardia (2012) that in some problems in Bioinformatics, we have data on the full polar-coordinates in 3–D giving the Cartesian coordinates (eg, Baker and Hubbard, 1984; Burdett et al, 2013)

$$x = r \sin\theta \sin\phi, \quad y = r \sin\theta \cos\phi, \quad z = r \cos\theta,$$

where $r > 0, \quad 0 < \phi < \pi, \quad 0 < \theta < 2\pi$ and another variable t as latitude $0 < t < \pi$; the longitude is missing. It turns out that the distribution of (θ, ϕ) given r becomes critical because it has a skew distribution depending on the underlying shapes.
One procedure is to use the equal-area projection known as the Schmidt projection. Namely if the spherical variables with θ colatitude and ϕ longitude then

$$x_1 = u \cos\phi, x_2 = u \sin\phi, u = 2 \sin\theta/2, 0 < u < 2.$$

Thus we can use any bivariate distribution of (x_1, x_2) on R^2 and then transform to (θ, ϕ) to construct a distribution on sphere. For example, it can be seen that if (x_1, x_2) are (truncated) bivariate normal with zero means, unit variances and correlation ρ then we get a new characterization of the Fisher distribution. Note that for the Kent distribution, we need to start with the pdf proportional to

$$e^{-\frac{1}{2}ax_1^2 - \frac{1}{2}bx_2^2 - \frac{1}{2}c(x_1^4 - x_2^4)}, x_1^2 + x_2^2 < 4,$$

with some constrains on a,b and c. Note that in transforming back, the quartic term in the exponent does not allow singularity at u=2 (south pole). For skew distribution, we have now a wide choice including Azzalini's bivariate distribution, a higher order distribution than FB_5 (used here).

7 More Directional Models, Landmarks and Edgels

In the last few years, there has been upsurge of arrival of new directional models —some extending the current ones and others introducing new ones. Oualkacha and Rivest (2012), Kume et al (2013) and Wood (2013) and Arnold et al (2013) are some examples. The models of Oualkacha and Rivest (2012), Kume et al. (2013) and Wood (2013) contain the application to walking movement; the example was initiated in Rivest et al. (2008). The paper by Kume et al. (2013) extends the cylindrical model of Mardia and Sutton (1978) to higher dimensions which is then applied to this illustrative problem. The models of Arnold et al (2013) have many applications including in crystallography. These all belong to the exponential family and as usual the normalizing constants are intricate: Kume et al. (2013) have produced a saddle-point approximation to overcome this stumbling block.

Another area where one foresees development is in generalized shape defined by landmarks with edgels. Edgel is a direction at a landmark (see, for example, Mardia et al, 2004). Mardia et al (2004) gave an example of a triangular shape with an edgel in 3D. Mardia (2012) has pointed out a example from Structural Bioinformatics where there are two landmarks each with a single edgel in 3D (see Burdett, et al, 2013). These are in the form space so that the rigid body transformation is filtered out.

Deane et al. (2013) have given another example in Structural Bioinformatics where there are two landmarks, each having two edgels. The form space is complicated but we can use Bookstein type coordinates. Consider a simple example in 3D. Let x_i, i=1,2 be two landmarks in 3D each with two edgels (see Figure in the cover; the two edgels at each landmarks are orthogonal). Without any loss of generality, we take x_1 to be the origin and the two edgels to form the coordinate frame so we are left with x_2, say x, and a rotation matrix (associated with its edgels), say A. Thus then it can be seen that there are 6 degrees of freedom under rigid transformation (3 for each landmark and 3 for rotation, noting that 6 variables are required for registration). This is a different representation than Deane et al. (2013) but note that the both systems have six degrees of freedom. It is also to be noted their registration system is knowledge–based and is more meaningful for biologists.

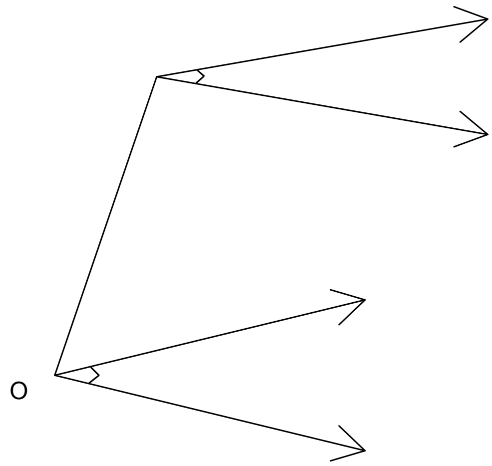

Figure 3: Two Landmarks with two edgels in 3D

Coming now to a plausible model, one can think of a Hidden Markov Model with x as a trivariate normal and A as a Fisher matrix distribution. Both distributions can be easily simulated as well as fitted. Note that Oualkacha and Rivest (2012) and Kume et al. (2013) have assumed independence of the rotation matrix and a linear variable but in practice the HMM strategy get rid of this assumption, though the HMM model is a latent one so we cannot write its distribution explicitly. The main point is that there are various new directional and shape data in Structural Bioinformatics which includes Crystallography, NMR imaging (see for example, Hamelryck et al, 2012).

8 Shape and Directional Discrimination

Kent and Mardia (2013) have given a spherical discriminant through FB_5 which needs parallel transport. Kent et al (2006) have shown that FB_5 provides a plausible model for triangular shapes so we can use this discriminant for triangular shapes. This is motivated by Linney et al (2006) on classifying facial expressions. For higher order shapes in 2D, a model called Complex Bingham Quartic Distribution (Kent et al, 2006) can be used in the same way, Another approach is to use Kents tangent projection; this approach has been used effectively by Bookstein and Kowell (2010) in Fetal Alcohol Spectrum Disorders (FASD). See also Mardia et al.(2013).

9 Circadian

Obvious cases for circular data include the 24-hour clock. Hastings (2013) has given various challenging examples. One of the key aspects is "that across all biological groups tested, circadian clocks are pivoted around a true oscillator, most frequently modelled as a limit–cycle

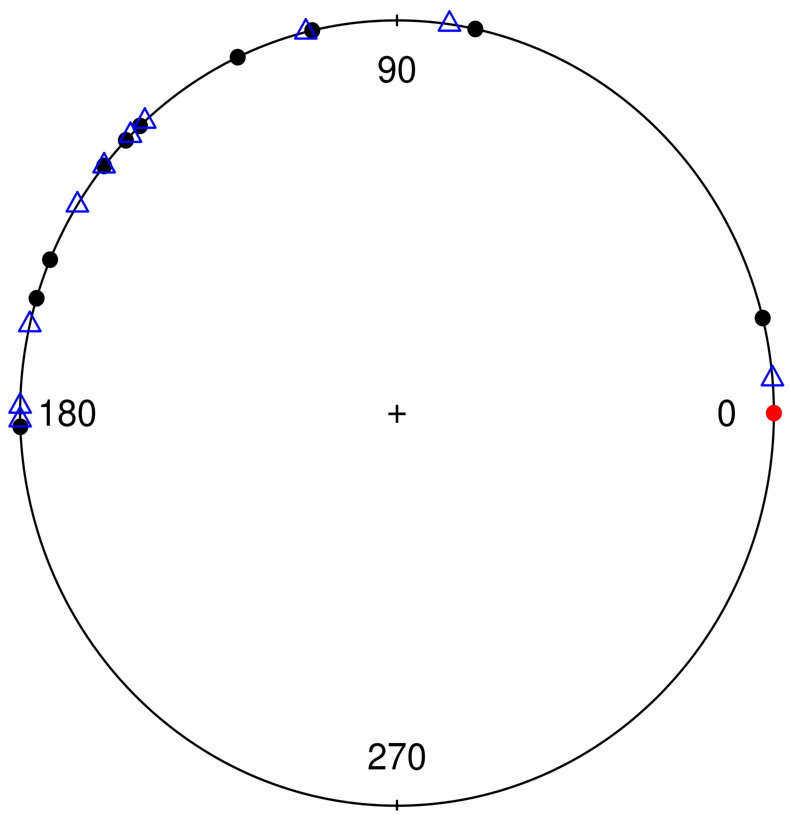

Figure 4: Systolic Blood Pressure: Peak time on two different days (Data from Downs and Mardia, 2002.)

oscillator". Limit cycle is a trajectory for which energy of the system would be constant over a cycle, that is, on an average there is no loss or gain of energy. Limit–cycle is an outcome of delicate energy balance due to the presence of nonlinear terms in the equation of motion. A.T. Winfree has pioneered this area with several contributions starting form Winfree(1970).

Another area on circular "inner" ordering sector-wise is examined by Cristina Rueda et al.(2013); the motivating example is a very important problem related to cell cycle genes.

10 Discussion

Non-Euclidean data can be analysed through Euclidean methodology by embedding, tangent approximations and so on which can bypass the need for a model. It seems this point has been emphasized in various papers including Fred Bookstein, Stephan Huckemann, Vic Patrangenaru, Stanislav Katina, Simon Byrne. John Kent will describe also related posters under this theme in his preview.

Finally, we note that we have not pointed out many other possible developments for directional and shape analysis such as Bayesian methods, graphical models, regression problems and so on, but we expect that the work will have direct impact.

11 Acknowledgments

I wish to express my thanks to Fred Bookstein, Anthony Riley, and Zhengzheng Zhang for their help.

12 References

Arnold, R., Jupp, P.E., and Schaeben, J. (2013). In this Proceedings.

Azzalini, A. (2013). In this Proceedings.

Azzalini, A. and Dalla Valle, A.(1996) The multivariate skew-normal distribution. *Biometrika*, **83**, pp. 715–726.

Baker, E.N. and Hubbard, R.E. (1984) Hydrogen bonding in globular proteins. *Prog. Biophys. Molec. Biol.*, **44**, pp.97–179.

Bookstein, F. L. and Kowell, A. P. (2010) Bringing morphometrics into the fetal alcohol report: statistical language for the forensic neurologist or psychiatrist. *Journal of Psychiatry and Law* **38**, pp. 449–473.

Burdett, P.M., Mardia, K.V. and Barber, S. (2013). In this Proceedings.

Deane, C. M., Dunbar, J., Fuchs, A., Mardia, K.V., Shi. J., Wilman, H.R., Barber, S., and Other, A.N.(2013). In this Proceedings.

Downs, T.D and Mardia, K.V (2002). Circular regression. *Biometrika*, **89**, pp 683-697.

Dryden, I.L. and Mardia, K.V. (1998). *Statistical Shape Analysis*, John Wiley.

Fisher. R. A. (1953). Dispersion on a sphere. *Proc. Roy. Soc. London* Ser A,**217**, pp.295-305.

Hamelryck, T., Mardia, K.V. and Ferkinghoff-Borg, J. (2012) *Bayesian Methods in Structural Bioinformatics*. (Editors) Springer.

Hastings, M.H.(2013). In this Proceedings.

Jammalamadaka, S.R. and SenGupta, A. (2001) *Topics in Circular Statistics.* World Scientific.

Jona–Lasinio, G.J., Gelfand, A. and Jona–Lasinio, M.J. (2012). Analyzing spatial directional data using wrapped Gaussian processes. *Annals of Applied Statistics*. **6**, pp. 1478-1498

Jones, M. C. and Pewsey, A. (2012). Inverse Batschelet distributions for circular data. *Biometrics*, **68** pp.183-193

Kent, J.T. and Mardia, K.V.(2013). In this Proceedings.

Kent, J.T., Mardia, K.V. and McDonnell, P. (2006). The complex Bingham quartic distribution and shape analysis. *Journal of Royal Statistical Society Series B*, **68**, pp.747-765.

Kume, A., Preston, S.P. and Wood, A.T.A. (2013). Saddlepoint approximations for the normalising constant of Fisher-Bingham distributions on products of spheres and Stiefel manifolds. *Biometrika*, In press.

Linney, A., McDonald, D., and Wright, A. (2006). Alter ego - live video classification of facial expression. Annual *LASR 2006 Proceedings.* Leeds University Press, pp.31-34.

Mardia, K.V. (1975) Statistics of directional data (with discussion). *J. Royal Statist. Soc. Series* B, **37**, pp. 349–393.

Mardia, K. V. (2012). Statistics of some topics in turbulence and biomolecular fields. *LASR 2012 Proceedings.* Mardia, K.V., Gusnanto, A., Riley A.D. and Voss J. (Eds.) Leeds University Press, pp.9-20

Mardia, K.V.(2013). Statistical Approaches to three key challenges in Protein Structural Bioinformatics, *Journal of the Royal Statistical Society*, Series C, *Applied Statistics*, **62**, pp.487-514.

Mardia, K.V. and Sutton, T. (1978) A Model for Cylindrical Variables with Applications. *J.Roy.Statist.Soc.* B, **40**, pp. 229-233.

Mardia, K.V., Kirkbride, J., and Bookstein, F.L. (2004). Statistics of shape, direction and cylindrical variables. *J. Applied Statist.*, **31**, pp.465–479.

Mardia, K.V., Hughes, G., Taylor, C.C. and Singh, H. (2008) A multivariate von Mises distribution with applications to bioinformatics. *Can. J. of Statist.*, **36**, pp.99–109.

Mardia, K.V., Bookstein, F.L., and J.T. Kent. (2013). Alcohol, babies and the death penalty: Saving lives by analysing the shape of the brain. *Significance* **10**, June issue, pp.12-16.

Mardia, K.V. and Voss, J. (2013) Some fundamental properties of multivariate von Mises distributions. *Department of Statistics, Research Report*, **STAT11–05**.In press

Oualkacha, K. & Rivest, L.–P. (2012). On the estimation of an average rigid body motion. *Biometrika*, **99**, pp. 585–598.

Rivest, L.-P., Baillargeon, S. & Pierrynowski, M.R. (2008). A directional model for the estimation of the rotation axes of the ankle joint. *J. Am. Statist. Assoc.* **103** ,pp.1060–1069.

Rueda, C., Fernández, M.A.,Barragán, S., and Peddada, S.D. (2013). In this Proceedings.

Singh, H., Hnizdo, V., and Demchuk, E. (2002). Probabilistic model for two dependent circular variables. Biometrika 89, 719723.

Umbach, D. and Jammalamadaka, S. R. (2009) Building asymmetry into circular distributions. *Statistics and Probability Letters*, **79**, pp. 659–663.

Winfree, A.T. (1970). Integrated view of resetting a circadian clock. *J. Theor. Biol.*, **28**, pp.327-374.

Wood, A.T.A. (2013). In this Proceedings.