LOCAL REGRESSION FOR CIRCULAR DATA

Marco Di Marzio

DMQTE, Chieti-Pescara University

Agnese Panzera

DiSIA, Florence University

Charles C. Taylor

Department of Statistics, Leeds University

Advances in Directional Statistics, Brussels 2014

OUTLINE

- Local polynomial estimators for
  - circular-linear regression
  - linear-circular regression
  - circular-circular regression
  - spherical-linear regression
  - spherical-spherical regression
- Applications

CIRCULAR PREDICTOR: THE MODEL

Given a set $\{(\Theta_i, Y_i), i = 1, \ldots, n\}$ of independent copies of a $[0, 2\pi) \times \mathbb{R}$-valued random vector (Θ, Y), assume

$$Y_i = m(\Theta_i) + \sigma(\Theta_i)\epsilon_i, \quad i = 1, \ldots, n$$

where ϵ_is are *i.i.d.* rr. vv., independent from the Θ_is, with $E[\epsilon_i] = 0$, $\mathrm{Var}[\epsilon_i] = 1$.

A SERIES EXPANSION FOR PERIODIC FUNCTIONS

Assuming that m is smooth enough:

$$m(\phi) = \sum_{j=0}^{p} \frac{m^{(j)}(\theta)\sin^j(\phi-\theta)}{j!} + O(\sin(\phi-\theta)^{p+1}),$$

for ϕ in a neighborhood of angle θ.

Above expansion enjoys a Taylor series interpretation:

$$\sin(\phi-\theta) \simeq \phi-\theta \ \text{ for } \ \phi \to \theta.$$

CIRCULAR KERNELS I

A circular kernel K_k, with concentration parameter $k > 0$, is a real function such that

i) it admits an uniformly convergent Fourier series, i.e.

$$K_k(\theta) = \frac{1 + 2\sum_{j=1}^{\infty} \gamma_j(k)\cos(j\theta)}{2\pi},$$

ii) as k increases $\int_{-\epsilon}^{\epsilon} |K_k(\theta)| d\theta$ tends to 1 for small $\epsilon > 0$;

iii) letting $\eta_j(K_k) := \int_{-\pi}^{\pi} \sin^j(\theta) K_k(\theta) d\theta$, then

$$\eta_0(K_k) = 1\,, \quad \eta_j(K_k) = 0 \ \text{ for } \ 0 < j < r\,, \quad \text{and} \quad \eta_r(K_k) \neq 0\,.$$

CIRCULAR KERNELS II

- The concentration parameter k determines which part of the sample contributes to the estimation.

- It is *spatial* in nature, emphasizing the role of the observations which are closer to the estimation point.

Many *circular densities* are second-order circular kernels.

- **von Mises**;
- **Wrapped Normal**;
- **Wrapped Cauchy**;

LOCAL POLYNOMIALS FOR CIRCULAR PREDICTORS

A pth degree local polynomial estimator for $m(\theta)$ is the solution for β_0 of

$$\underset{\{\beta_0,\beta_1,\dots,\beta_p\}}{\text{argmin}} \sum_{i=1}^{n} \left\{ Y_i - \sum_{j=0}^{p} \frac{\sin^j(\Theta_i - \theta)\beta_j}{j!} \right\}^2 K_k(\Theta_i - \theta),$$

taking the form

$$\hat{m}(\theta; p) = \sum_{i=1}^{n} Y_i W_i(\theta),$$

where W_i is a weight function depending on $K_k(\Theta_i - \theta)$ and Θ_i. (Obviously) It is differently structured in the cases $p = 0$ and $p = 1$.

ASYMPTOTIC MEAN SQUARED ERROR

Under suitable regularity assumptions, at the estimation point θ, and for $p \in \{0, 1\}$

$$\text{AMSE}[\hat{m}(\theta; p)] = \underbrace{\frac{\eta_2^2(K_k)B^2(\theta; p)}{4}}_{\text{squared bias}} + \underbrace{\frac{R(K_k)\sigma^2(\theta)}{nf(\theta)}}_{\text{variance}}$$

where

$$B(\theta; p) := \begin{cases} m^{(2)}(\theta) + 2m^{(1)}(\theta)f^{(1)}(\theta)/f(\theta), & \text{if } p = 0, \\ m^{(2)}(\theta), & \text{if } p = 1. \end{cases}$$

OPTIMAL SMOOTHING

- The concentration parameter is not a scale factor.
- Bias and variance depend on k throughout $\gamma_j(K_k)$, $j \in \mathbb{Z}^+$. Specifically,

$$\eta_2(K_k) = \frac{1-\gamma_2(k)}{2} \quad \text{and} \quad R(K_k) = \frac{1 + 2\sum_{j=1}^{\infty} \gamma_j^2(k)}{2\pi}.$$

- Differently from the Euclidean case, the shape of the kernel and the smoothing degree come as not separated in AMSE expression. Consequently, a general structure for optimal smoothing degree is hard to obtain!

USING THE VON MISES KERNEL

For the von Mises kernel

$$\eta_2(K_k) \approx \frac{1}{k} \quad \text{and} \quad R(K_k) \approx \sqrt{\frac{k}{4\pi}}.$$

The minimizer of the resulting AMSE over k yields

$$k_{\text{AMSE}} = \left\{ \frac{2f(\theta) n \pi^{1/2} B^2(\theta; p)}{\sigma^2(\theta)} \right\}^{2/5},$$

(k_{AMSE} goes to infinity, whereas euclidean bandwidth goes to zero) which gives

$$\inf_{k>0} \text{AMSE}[\hat{m}(\theta; p)] \sim n^{-4/5}.$$

As expected, the convergence rate is the same as in the Euclidean case.

REGRESSION WITH CIRCULAR RESPONSE

When the response, say Θ, is circular and the predictor, say Δ, takes values on a generic domain, the regression function can be modeled as

$$m(\delta) := \arctan\left(\frac{E[\sin(\Theta)|\Delta = \delta]}{E[\cos(\Theta)|\Delta = \delta]}\right),$$

which minimizes the angular risk

$$E[2\{1 - \cos(\Theta - m(\Delta))\} \mid \Delta = \delta].$$

CIRCULAR RESPONSE: THE MODEL

Letting *i.i.d* rr. vv. $\{(\Delta_i, \Theta_i),\ i = 1, \cdots, n\}$, assume

$$\Theta_i = [m(\Delta_i) + \epsilon_i](\text{mod} 2\pi), \qquad i = 1, \ldots, n,$$

where the ϵ_is are *i.i.d.* random angles with zero mean direction, and finite concentration.

CIRCULAR RESPONSE: THE ESTIMATOR

Letting

$$m_1(\delta) := E[\sin(\Theta) \mid \Delta = \delta] \ \text{ and } \ m_2(\delta) := E[\cos(\Theta) \mid \Delta = \delta]$$

A local estimator for m at δ could be defined as

$$\hat{m}(\delta) := \arctan\left(\frac{\hat{m}_1(\delta)}{\hat{m}_2(\delta)}\right),$$

with

$$\hat{m}_1(\delta) := \sum \sin(\Theta_i) W(\Delta_i, \delta) \ \text{ and } \ \hat{m}_2(\delta) := \sum \cos(\Theta_i) W(\Delta_i, \delta),$$

where W is a local weight depending, as usual, on the sample observation Δ_i and the estimation point δ .

LINEAR-CIRCULAR AND CIRCULAR-CIRCULAR REGRESSION ESTIMATORS

Above estimator entails, by simple adaptations of the weight function, a *unified approach* for linear and circular predictor cases.

- When Δ is *linear*, local constant and local linear fits are obtained by using **euclidean kernel-based weights**.
- When Δ is circular, local constant and local linear fits can be obtained by using **circular kernel-based weights**.

SELECTION OF THE SMOOTHING DEGREE

Due to the circular nature of $\hat{m}(\delta)$, an accuracy measure for it can be defined as

$$L[\hat{m}(\delta)] := E[2\{1 - (\cos(\hat{m}(\delta) - m(\delta))\}],$$

Risk L is a circular version of MSE, and asymptotically (i.e. when the difference $\hat{m}(\delta) - m(\delta)$ is small) corresponds to it. In our practical experiments we have selected the smoothing degree by applying cross-validation based on an empirical version of risk L.

HYPERSPHERICAL DATA

Circular data can be also represented as unit vectors in $\mathbb{R}^2$ (*embedding approach*).

In general, a unit vector in $d \geq 2$ dimensions can be regarded as a point on the surface of the hypersphere

$$\mathbb{S}^{d-1} := \{\mathbf{x} \in \mathbb{R}^d : ||\mathbf{x}|| = 1\},$$

and defines a *hyperspherical* (also *spherical, directional*) observation.

SPHERICAL PREDICTOR: THE MODEL

For a set of independent copies $\{(\boldsymbol{X}_i, Y_i), i = 1, \dots, n\}$ of a $\mathbb{S}^{d-1} \times \mathbb{R}$-valued random vector $(\boldsymbol{X}, Y)$, assume

$$Y_i = m(\boldsymbol{X}_i) + \sigma(\boldsymbol{X}_i)\epsilon_i, \qquad i = 1, \dots, n$$

where the ϵ_is are *i.i.d.* with $\mathrm{E}[\epsilon_i] = 0$, $\mathrm{Var}[\epsilon_i] = 1$, and independent from the $\boldsymbol{X}_i$s.

TANGENT-NORMAL DECOMPOSITION

Given $\boldsymbol{a} \in \mathbb{S}^{d-1}$, the tangent-normal decomposition of a vector $\boldsymbol{b} \in \mathbb{S}^{d-1}$ is

$$\boldsymbol{b} = \boldsymbol{a}\cos(\theta) + \boldsymbol{c}\sin(\theta),$$

where $\theta \in [0, \pi]$ denotes the angle between $\boldsymbol{a}$ and $\boldsymbol{b}$, i.e. $\theta := \arccos(\boldsymbol{b}'\boldsymbol{a})$, and $\boldsymbol{c}$ is a vector orthogonal to $\boldsymbol{a}$.

A SERIES EXPANSION FOR FUNCTIONS DEFINED ON THE SPHERE

Provided that m is smooth enough, for $\boldsymbol{b}$ near $\boldsymbol{a}$, according to the tangent-normal decomposition

$$m(\boldsymbol{b}) = m(\boldsymbol{a}) + \sum_{s=1}^{p} \frac{\theta^s}{s!} \boldsymbol{c}' \mathcal{D}^s_{\bar{m}}(\boldsymbol{a}) \boldsymbol{c}^{\otimes(s-1)} + O\left(\theta^{p+1}\right)$$

where $\bar{m}(\boldsymbol{a}) := m(\boldsymbol{a}/||\boldsymbol{a}||)$, and $\mathcal{D}^s_{\bar{m}}(\boldsymbol{a})$ is the matrix of the sth order derivatives of $\bar{m}$ at $\boldsymbol{a}$.

LOCAL POLYNOMIALS FOR SPHERICAL-LINEAR REGRESSION

A pth degree local polynomial estimator of $m(\boldsymbol{x})$ can be defined as the solution for β_0 of

$$\arg\min_{\{\beta_0,\ldots,\boldsymbol{\beta}_p\}} \sum_{i=1}^{n} \left\{ Y_i - \beta_0 - \sum_{j=1}^{p} \frac{\theta_i^j}{j!} \boldsymbol{\xi}_i' \boldsymbol{\beta}_j \boldsymbol{\xi}_i^{\otimes(j-1)} \right\}^2 K_k(\cos(\theta_i)),$$

with K_k being a kernel defined on $\mathbb{S}^{d-1}$, having mean direction $\boldsymbol{X}_i$ and evaluated at the point $\boldsymbol{x}$.

SPHERICAL WEIGHTS

K_k is a unimodal density defined on $\mathbb{S}^{d-1}$ with

- rotational symmetry about its mean direction $\boldsymbol{\mu} = (0, \ldots, 0, 1)$;
- concentration parameter $k > 0$ such that, for any $W \subset \mathbb{S}^{d-1} \setminus \{\boldsymbol{\mu}\}$,

$$\lim_{k \to \infty} \int_W K_k(\boldsymbol{x}'\boldsymbol{\mu})\omega_{d-1}(d\boldsymbol{x}) = 0.$$

Example: Langevin density

$$K_k(\cos(\theta)) := \frac{\kappa^{d/2-1} e^{\kappa \cos(\theta)}}{(2\pi)^{d/2} \mathcal{I}_{d/2-1}(\kappa)}.$$

LOCAL CONSTANT AND LOCAL LINEAR FITS

A local constant fit is

$$\hat{m}(\boldsymbol{x};0) = \frac{\sum_{i=1}^{n} K_k(\cos(\theta_i))Y_i}{\sum_{i=1}^{n} K_k(\cos(\theta_i))}.$$

For $p = 1$, letting

$$\mathbb{Y} := [Y_1 \cdots Y_n]', \quad \mathbb{W} := \mathrm{diag}[K_k(\cos(\theta_1)), \ldots, K_k(\cos(\theta_n))],$$

$$\boldsymbol{\beta} := [\beta_0 \ \ \boldsymbol{\beta}_1']' \text{ and } \mathbb{X} := \begin{bmatrix} 1 & \theta_1\boldsymbol{\xi}_1' \\ \vdots & \vdots \\ 1 & \theta_n\boldsymbol{\xi}_n' \end{bmatrix},$$

the loss in our lest squares problem can be re-written as

$$||\mathbb{W}^{1/2}(\boldsymbol{Y} - \mathbb{X}\boldsymbol{\beta})||^2.$$

CONSTRAINED WEIGHTED LEAST SQUARES (1)

Minimization of the above loss over $\boldsymbol{\beta}$ admits a unique solution iff $\mathbb{X}'\mathbb{W}\mathbb{X}$ is nonsingular, and this is *not* the case for $p = 1$ because a block of matrix $\mathbb{X}$ factorizes as follows

$$\begin{bmatrix} \theta_1 \boldsymbol{\xi}_1' \\ \vdots \\ \theta_n \boldsymbol{\xi}_n' \end{bmatrix} = \begin{bmatrix} \frac{\theta_1}{\sin(\theta_1)} \boldsymbol{X}_1' \\ \vdots \\ \frac{\theta_n}{\sin(\theta_n)} \boldsymbol{X}_n' \end{bmatrix} (\boldsymbol{I} - \boldsymbol{x}\boldsymbol{x}'),$$

where $\boldsymbol{I} - \boldsymbol{x}\boldsymbol{x}'$ is singular.
Since $\boldsymbol{x}'\mathcal{D}_{\bar{m}}(\boldsymbol{x}) = 0$, we can define $\hat{m}(\boldsymbol{x}; 1)$ as the solution for β_0 of

$$\underset{\boldsymbol{\beta}}{\operatorname{argmin}} ||\mathbb{W}^{1/2}(\boldsymbol{Y} - \mathbb{X}\boldsymbol{\beta})||^2 \quad \text{subject to} \quad \boldsymbol{Q}_1'\boldsymbol{\beta} = 0,$$

with $\mathbf{Q}_1 := [0 \;\; \mathbf{x}']'$.

CONSTRAINED WEIGHTED LEAST SQUARES (2)

This yields

$$\hat{m}(\boldsymbol{x};1) = \boldsymbol{e}_1'\boldsymbol{Q}_2(\boldsymbol{Q}_2'\mathbb{X}'\mathbb{W}\mathbb{X}\boldsymbol{Q}_2)^{-1}\boldsymbol{Q}_2'\mathbb{X}'\mathbb{W}\boldsymbol{Y},$$

where $\boldsymbol{e}_1 := [1 \;\; \mathbf{0}']'$, and $\boldsymbol{Q}_2$ is a $(d+1)\times d$ matrix such that $\boldsymbol{Q}_2'\boldsymbol{Q}_1 = \mathbf{0}$, and the matrix $[\boldsymbol{Q}_1 \;\; \boldsymbol{Q}_2]$ is non-singular.

$\boldsymbol{Q}_2$ is a projector of the solution for β into the space of the vectors orthogonal to $\boldsymbol{Q}_1$: its choice does not affect the estimate.

ACCURACY AND OPTIMAL SMOOTHING

Under suitable assumptions, when the Langevin kernel is employed, for $p \in \{0, 1\}$,

$$\mathrm{E}[\hat{m}(\boldsymbol{x}; p)] - m(\boldsymbol{x}) = O\left(\frac{d-1}{k}\right), \quad \mathrm{Var}[\hat{m}(\boldsymbol{x}; p)] = O\left(n^{-1}k^{(d-1)/2}\right),$$

which yields

$$k_{\mathrm{AMSE}} \sim n^{2/(d+3)}, \quad \text{and} \quad \inf_{k>0} \mathrm{AMSE}[\hat{m}(\boldsymbol{x}; p)] \sim n^{-4/(d+3)}.$$

ROTATIONAL EQUIVARIANCE

Let $\boldsymbol{R}_\alpha$ denote the matrix performing rotations of vectors in $\mathbb{S}^{d-1}$ about the x-axis by the angle $\alpha \in (0, 2\pi)$. For a whatever location $\boldsymbol{x} \in \mathbb{S}^{d-1}$, and $p \in \{0, 1\}$,

$$\hat{m}(\boldsymbol{x}; p) = \hat{m}_R(\boldsymbol{R}_\alpha \boldsymbol{x}; p),$$

where $\hat{m}_R(\cdot; p)$ is defined as $\hat{m}(\cdot; p)$ when the sample is

$\{(\boldsymbol{R}_\alpha \boldsymbol{X}_i, \boldsymbol{R}_\alpha \boldsymbol{Y}_i), i = 1, \ldots, n\}$.

SPHERICAL-SPHERICAL REGRESSION

Let $(\boldsymbol{X}, \boldsymbol{Y})$ be a $\mathbb{S}^{d-1} \times \mathbb{S}^{q-1}$-valued random vector, $Y^{(\ell)}$ denote the ℓth cartesian coordinate of $\boldsymbol{Y}$, and set

$$m_\ell(\boldsymbol{x}) := E[Y^{(\ell)} \mid \boldsymbol{X} = \boldsymbol{x}].$$

The dependence of $\boldsymbol{Y}$ from $\boldsymbol{X}$ could be described by the minimizer of

$$E[||\boldsymbol{Y} - \boldsymbol{m}(\boldsymbol{X})||^2 \mid \boldsymbol{X}] \text{ subject to } ||\boldsymbol{m}(\boldsymbol{X})|| = 1,$$

which, at $\boldsymbol{X} = \boldsymbol{x}$, is

$$\boldsymbol{m}(\boldsymbol{x}) = ||[m_1(\boldsymbol{x}) \cdots m_q(\boldsymbol{x})]||^{-1}[m_1(\boldsymbol{x}) \cdots m_q(\boldsymbol{x})]'.$$

SPHERICAL-SPHERICAL REGRESSION MODEL

Given the random sample $\{(\boldsymbol{X}_i, \boldsymbol{Y}_i), i = 1, \ldots, n\}$, assume

$$\boldsymbol{Y}_i = \boldsymbol{m}(\boldsymbol{X}_i) + \boldsymbol{\epsilon}_i,$$

where the $\boldsymbol{\epsilon}_i$s, conditioned on the $\boldsymbol{X}_i$s, are independent random vectors with $E[\boldsymbol{\epsilon}_i \mid \boldsymbol{X}_i] = \mathbf{0}$ and $Var[\boldsymbol{\epsilon}_i \mid \boldsymbol{X}_i] = \Sigma(\boldsymbol{X}_i)$.

And the estimator is

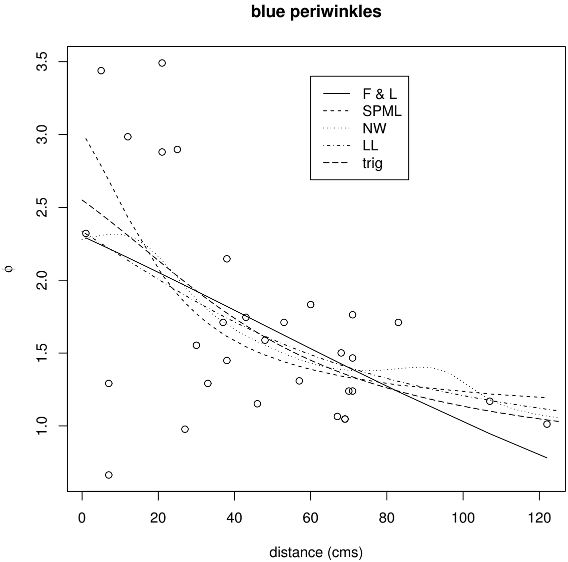
$$\hat{\boldsymbol{m}}(\boldsymbol{x}; p) = ||[\hat{m}_1(\boldsymbol{x}; p) \cdots \hat{m}_q(\boldsymbol{x}; p)]||^{-1}[\hat{m}_1(\boldsymbol{x}; p) \cdots \hat{m}_q(\boldsymbol{x} : p)]'.$$

APPLICATION TO SNAILS MOVEMENTS (1)

We consider the data collected in Fisher and Lee (1992) about distances and directions moved by small blue periwinkles after relocation.
The objective is to predict the angles given the distance moved.

We compare the local linear (LL) and local constant (NW) versions of our estimator, together with parametric fits in Fisher & Lee (1992) (f&L), Presnell et al. (1998) (SPML), and a direct trigonometric fit (trig).

APPLICATION TO SNAILS MOVEMENTS (2)

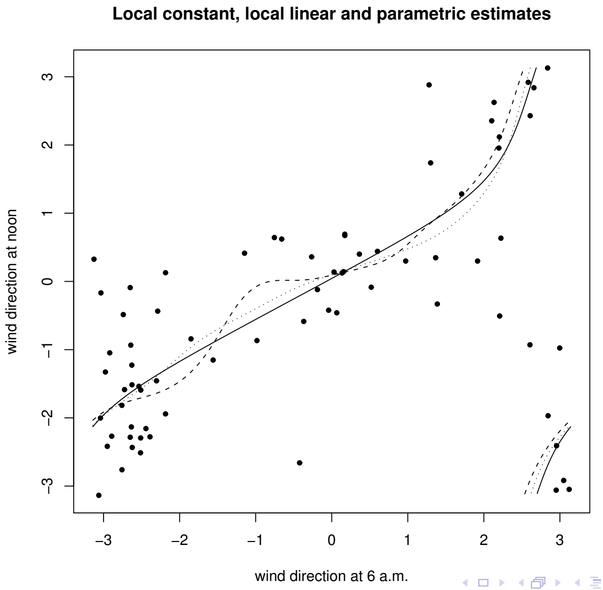


APPLICATION TO WIND DIRECTIONS (1)

We consider the data used in Kato & Jones (2010) in which the objective is to model and predict the wind direction at noon, based on the wind direction at 6 a.m. at a weather station in Texas.
They considered some parametric models based on the Mobius transformation.

Here we compare our local constant (dashed) and local linear (dotted) smoothers with a variant of their models which uses von Mises errors (continuous line).

APPLICATION TO WIND DIRECTIONS (2)



TREND ESTIMATION IN CIRCULAR TIME SERIES

Given a time series of angles $\{\Theta_t\}_{t=1}^{T}$, we assume

$$\Theta_t = [m(t/T) + \varepsilon_t](\text{mod} 2\pi)$$

where $\{\varepsilon_t\}_{t=1}^{T}$ is a stationary process with $\text{E}[\sin(\varepsilon_t)] = 0$.

Our smoothers for the circular response case could estimate the trend function at t/T.

Here, $\Delta_i = i/T$, and the kernel is an euclidean one supported on $[0, 1]$.

ORDER STATISTICS AND CIRCULAR RANK

If we treat the sample of angles $\Theta_1, \cdots, \Theta_n$ as *linear* data, and rearrange them into ascending order, we obtain

$$\Theta_{(1)} \leq \cdots \leq \Theta_{(n)}.$$

Letting r_i denote the rank of Θ_i among $\Theta_1, \cdots, \Theta_n$ $\big($i.e. Θ_i corresponds to the order statistics $\Theta_{(r_i)}\big)$, the **circular rank** of Θ_i, $i \in \{1, \cdots, n\}$, is defined as

$$\omega_i := 2\pi r_i / n.$$

CIRCULAR QUANTILES ESTIMATION

Let $\Theta_1, \cdots, \Theta_n$ be a random sample from an absolutely continuous circular distribution.
Natural local smoothing of αth quantile is

$$\hat{q}(\alpha) := \arctan(\hat{q}_1(\alpha)/\hat{q}_2(\alpha)),$$

where

$$\hat{q}_1(\alpha) := \sum_{i=1}^{n} K_k(\omega_i - 2\pi\alpha)\sin(\Theta_{(r_i)}),$$

and

$$\hat{q}_2(\alpha) := \sum_{i=1}^{n} K_k(\omega_i - 2\pi\alpha)\cos(\Theta_{(r_i)}).$$

REFERENCES

- Di Marzio, M., Panzera, A., Taylor, C.C. (2012). Nonparametric smoothing and prediction for nonlinear circular time series. *Journal of Time Series Analysis*, **33**, 620-630
- Di Marzio, M., Panzera, A., Taylor, C.C. (2012). Smooth estimation of circular distribution functions and quantiles. *Journal of Nonparametric Statistics* **24**, 935-949.
- Di Marzio, M., Panzera, A., Taylor, C.C. (2013). Nonparametric regression for spherical data. *JASA, to appear*.
- Di Marzio, M., Panzera, A., Taylor, C.C. (2013). Nonparametric regression for circular responses. *Scandinavian Journal of Statistics*, **40**, 238-255.
- Fisher, N.I., Lee, A. J. (1992). Local polynomial regression for circular predictors. *Biometrics*,**4**8, 665–677.
- Kato, S., Jones, C. (2010). A Family of Distributions on the Circle With Links to, and Applications Arising From, Möbius Transformation. *JASA*,**105**, 249–262.
- Presnell, B., Morrison, S. P., and Littell, R. C. (1998). Projected multivariate linearmodels fordirectional data. *JASA*, **93**, 1068– 1077.