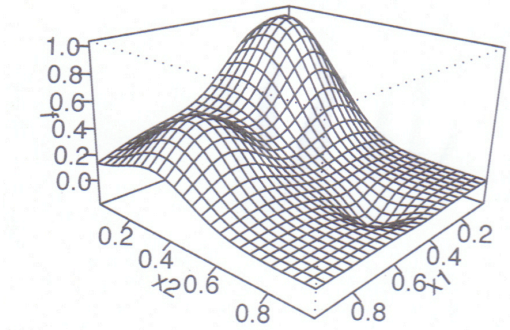
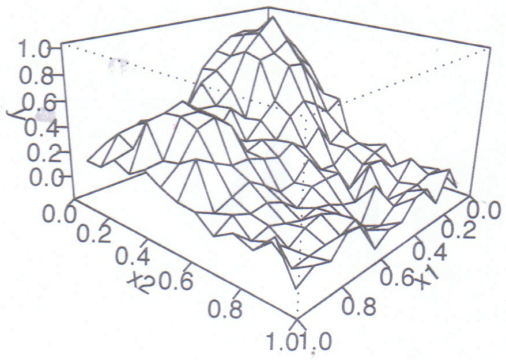
Statistics 860 Lecture 8

The next slide gives a thin plate spline demo. The 'true function' is Franke's principal test function, an old, popular example for testing two dimensional smoothing methods. This lecture will discuss the thin plate spline, and other isotropic methods for smoothing in two and sometimes more dimensions. Later we will discuss methods that are not isotropic (SS-ANOVA).

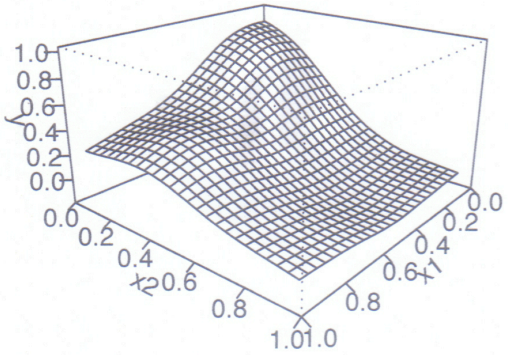
true function



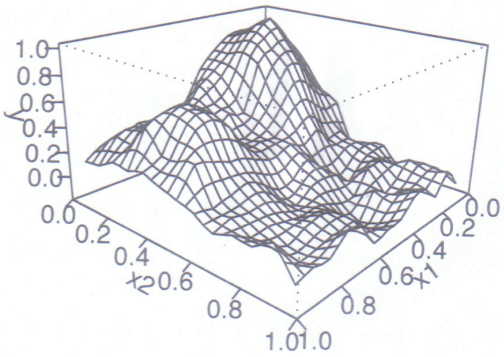
noisy data



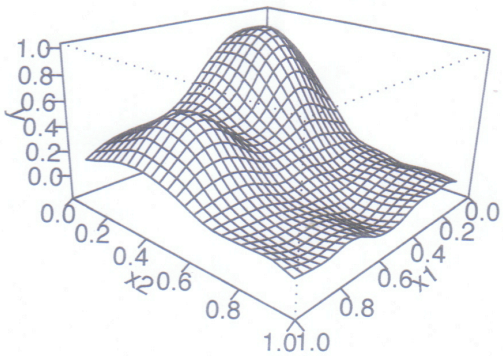
oversmoothed



undersmoothed



gcv fitting



Thin plate spline demo.

Thin plate splines: one version of splines on E^d. Let $d = 2, m = 2$,

$$< u, v >_* = \iint_{-\infty}^{\infty} u_{x_1x_1} v_{x_1x_1} + 2u_{x_1x_2} v_{x_1x_2} + u_{x_2x_2} v_{x_2x_2} dx_1 dx_2.$$

$< u, u >_* = 0$ for $u(x_1, x_2) = 1, x_1$ and x_2. (Planar functions)

$$< u, u >_* = \int \int u_{x_1x_1}^2 + 2u_{x_1x_2}^2 + u_{x_2x_2}^2 dx_1 dx_2$$

$d = 2$, general m

$$< u, v >_* = \iint \sum_{\nu=0}^{m} \binom{m}{\nu} \left(\frac{\partial^m u}{\partial x_1^\nu \partial x_2^{m-\nu}} \right) \left(\frac{\partial^m v}{\partial x_1^\nu \partial x_2^{m-\nu}} \right) dx_1 dx_2.$$

$$\| u \|_*^2 = \iint \left(\frac{\partial^m u}{\partial x_1^m} \right)^2 + \left(\begin{array}{c} m \\ 1 \end{array} \right) \left(\frac{\partial^m u}{\partial x_1^{m-1} \partial x_2^1} \right)^2 + ... + \left(\frac{\partial^m u}{\partial x_2^m} \right)^2 dx_1 dx_2.$$

Null space: $M = \begin{pmatrix} d+m-1 \\ d \end{pmatrix}$ polynomials of total degree $m-1$ or less.

For $d = 2, m = 3$, then $M = \begin{pmatrix} d+m-1 \\ d \end{pmatrix} = \begin{pmatrix} 4 \\ 2 \end{pmatrix} = 6$.

The null space is spanned by $\{\phi(x_1, x_2)\}$
$= \{1, x_1, x_2, x_1^2, x_2^2, x_1 x_2\}$.

The form for general d, m is

$$< u, v >_* = \sum_{\alpha_1 + \cdots + \alpha_d = m} \frac{m!}{\alpha_1! \cdots \alpha_d!} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}$$

$$\left(\frac{\partial^m u}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}} \right) \left(\frac{\partial^m v}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}} \right) \prod_j dx_j.$$

Need $2m - d > 0$.

Find $f = f(x_1, ..., x_d) \in \mathcal{H}$ to minimize

$$\frac{1}{n} \sum_{i=1}^{n} (L_i f - y_i)^2 + \lambda J_m^d(f)$$

where $J_m^d(f) = < f, f >_*$.

Define $E_m(s,t) = E(\|s-t\|) = E_m^d(s,t)$.

$$\begin{aligned} E(\tau) &= \theta_{m,d}|\tau|^{2m-d} \ln|\tau| && 2m-d \text{ is even} \\ &= \theta_{m,d}|\tau|^{2m-d} && \text{otherwise} \end{aligned}$$

where

$$\begin{aligned} \theta_{m,d} &= \frac{(-1)^{d/2+1+m}}{2^{2m-1}\pi^{d/2}(m-1)!(m-d/2)!} && 2m-d \text{ is even} \\ \theta_{m,d} &= \frac{\Gamma(d/2-m)}{2^{2m}\pi^{d/2}(m-1)!} && \text{otherwise} \end{aligned}$$

Generalized Divided Differences in E^d (GDD)

Let $\{\phi_\nu(s)\}_{\nu=1}^{M}$ span the $M = \begin{pmatrix} d+m-1 \\ d \end{pmatrix}$ polynomials of total degree less than m in E^d.

For example d=2, m=3.

$\phi_1(x_1, x_2) = 1, \phi_2(x_1, x_2) = x_1,$
$\phi_3(x_1, x_2) = x_2, \phi_4(x_1, x_2) = x_1^2,$
$\phi_5(x_1, x_2) = x_2^2, \phi_6(x_1, x_2) = x_1 x_2.$

$\begin{Bmatrix} t_1, ..., t_n \\ \alpha_1, ..., \alpha_n \end{Bmatrix}$, $t_i \in E^d$, α_i real numbers, are called a GDD if $\{t_1, ..., t_n\}$ is a unisolvent set and $\sum_{j=1}^{n} \alpha_j \phi_\nu(t_j) = 0, \nu = 1, ..., M$. Unisolvent means: Least squares regression on $\{\phi_\nu\}$ is unique.

Theorem Meinguet (1978)

Let $\left\{ \begin{array}{c} t_1, ..., t_n \\ \alpha_1, ..., \alpha_n \end{array} \right\}$ be a GDD, let $\left\{ \begin{array}{c} s_1, ..., s_K \\ \beta_1, ..., \beta_K \end{array} \right\}$ be a GDD, let $E_t(s) = E_m^d(\|t - s\|)$, then

$$< \sum_{j=1}^{n} \alpha_j E_{t_j}, \sum_{k=1}^{K} \beta_k E_{s_k} >_* = \sum_{j,k} \alpha_j \beta_k E_m^d(\|t_j - s_k\|).$$

"reproducing like property".

End result will be

Let $t(1), ..., t(n) \in E^d$ where $t(i) = (x_1(i), ..., x_d(i))$ and suppose the matrix $T_{n \times M}$ with $i\nu$th entry $\phi_\nu(t(i))$ is of rank M. ($\{t(i)\}$ is a unisolvet set). Let $\mathcal{X}$ be an appropriate Hilbert space of real-valued functions on E^d for which $J_m^d(f) < \infty$. Then the solution to the problem: find $f \in \mathcal{X}$ to minimize

$$\frac{1}{n} \sum_{i=1}^{n} (L_i f - y_i)^2 + \lambda J_m^d(f)$$

is unique, and ....

and has a representation

$$f_\lambda(t) = \sum_{\nu=1}^{M} d_\nu \phi_\nu(t) + \sum_{i=1}^{n} c_i E_{t(i)}(t)$$

where $d = (d_1, ..., d_M)'$ and $c = (c_1, ..., c_n)'$ satisfy

$$\begin{aligned} (K + n\lambda I)c + Td &= y \\ T'c &= 0 \end{aligned}$$

$\{K_{ij}\} = E_m^d(t(i), t(j))$.

Thus $E_m^d(s,t)$ "acts like" it is an RK even though it is not. ($E_m^d(s,s) = 0$, for example).

Remark: T of rank M and $T'c = 0 \Rightarrow \left\{ \begin{array}{c} t(1), ..., t(n) \\ c_1, ..., c_n \end{array} \right\}$ is a GDD.

Original results on thin plate splines are due to Duchon and Menguet (see references in the book) - the argument here will be different although will use their results.

Conditionally Positive Definite Functions: C. Micchelli (1986) in Constructive Approximation 2: 11-22 (in pdf1/micchelli.interpolation.86.pdf)

Let $s, t \in E^d$. $K(s,t)$ is said to be 'conditionally positive definite of order m' if for any GDD $\left\{ \begin{array}{c} t(1), ..., t(n) \\ c_1, ..., c_n \end{array} \right\}$,

$$\sum_{i,j=1} c_i c_j K(t(i), t(j)) > 0$$

Micchelli characterized the class of conditionally positive definite functions. E_m^d is conditionally p.d. of order m.

If

$$f = \sum_{\nu=1}^{M} d_\nu \phi_\nu(t) + \sum_{i=1}^{n} c_i E_{t(i)}(t)$$

with $T'c = 0$,

then $J_m^d(f) = c'Kc$ with $K_{ij} = E_m^d(\|t(i) - t(j)\|)$.

So consider

$$\begin{aligned} (K + n\lambda I)c + Td &= y \\ T'c &= 0 \qquad\qquad (*) \end{aligned}$$

$$\text{If } \; T = (Q_1 : Q_2) \begin{pmatrix} R \\ 0 \end{pmatrix}$$

then $\left\{ \begin{array}{c} t(1), ..., t(n) \\ q_{j1}, ..., q_{jn} \end{array} \right\}$ where $q_j = (q_{j1}, ..., q_{jn})'$ is a column of Q_2, is a GDD because $T'q_j = 0$.

Recall: $c = Q_2[Q_2'(K + n\lambda I)Q_2]^{-1}Q_2'y$ is the solution to $(*)$. Also K is not positive definite (only conditionally positive definite). $Q_2'KQ_2$ is positive definite. So same software can be used here as if you have a strictly positive definite RK!

D. Bates, M. Lindstrom, G. Wahba, and B. Yandell. GCVPACK-Routines for generalized cross validation. *Commun. Statist. Simul. Comput.*, 16:263–297, 1987.
[bates.lindstrom.wahba.yandell.86.pdf]

G. Wahba and J. Wendelberger. Some new mathematical methods for variational objective analysis using splines and cross-validation. *Monthly Weather Review*, 108:1122–1145, 1980.
[wahba.wendelberger.pdf]

Claim $Q_2'KQ_2$ is positive definite.

Proof: Let q be the lth column of Q_2, $\left\{ \begin{array}{c} t(1), ..., t(n) \\ q_{j1}, ..., q_{jn} \end{array} \right\}$ is a GDD because q_{li} satisfies $T'q_l = 0$. Therefore $\sum_{k=1}^{n-M} r_k q_k$ is a GDD for any $r = (r_1, ..., r_{n-M})'$.

So $r'Q_2'KQ_2r > 0$ since K is a positive definite on GDD's.

So we can write $Q_2'KQ_2 = UDU'$ as before with $d_\nu \geq 0$ and $I - A(\lambda) = Q_2U[diag(\frac{n\lambda}{n\lambda + d_\nu})]U'Q_2'$ as before. $\diamondsuit$

Radial Basis Functions (rbf):
Recently popular in machine learning. For this class we define an rbf as a basis function $K_t(\cdot)$ obtained from any symmetric positive definite function $K(s,t)$ on $E^d \times E^d$ of the form $K(s,t) = r(\| s-t \|)$ where $\| s-t \|$ is Euclidean distance.

Sometimes this definition is taken to include conditionally positive definite functions in which case E^d will be included. Also functions of the form $K(s,t) = r(\|B(s-t)\|)$ where B is some invertible $d \times d$ matrix are sometimes included.

$$K(s,t) = \int \cdots \int e^{i\omega\cdot(s-t)} H(\omega) \prod_{\nu=1}^{d} d\omega_\nu$$

with $H(\omega) \geq 0$, where $\omega = (\omega_1, ..., \omega_d)$, is always positive symmetric definite if it exists.

$$\sum_{j,k=1}^{n} a_j a_k K(t(j), t(k))$$

$$= \int \cdots \int \sum_{j,k=1}^{n} a_j a_k e^{[i\omega\cdot t(j) - i\omega\cdot t(k)]} H(\omega) \prod_{\nu=1}^{d} d\omega_\nu$$

$$= \int \cdots \int \sum_{j=1}^{n} a_j e^{i\omega\cdot t(j)} \sum_{k=1}^{n} a_k e^{-i\omega\cdot t(k)} \cdot H(\omega) \prod_{\nu=1}^{d} d\omega_\nu$$

$$= \int \cdots \int \left| \sum_{j}^{n} a_j e^{i\omega\cdot t(j)} \right|^2 H(\omega) \prod_{\nu=1}^{d} d\omega_\nu$$

≥ 0

Consider for $u \in E^d$,

$$g(u) = \int \cdots \int e^{i\omega \cdot u} H(\omega) \prod_{\nu=1}^{d} d\omega_\nu$$

Let $H(\omega) = h(\|\omega\|)$, change ω to polar coordinates.

$$H(\omega_1, ..., \omega_d) = H(\|\omega\|, 0, ..., 0) = h(\|\omega\|).$$

Let $g(u) = r(\|u\|)$ where

$$r(t) = 2^{\frac{d-2}{2}} \Gamma(\frac{d}{2}) \int_0^\infty \frac{J_{\frac{d-2}{2}}(\omega t) h(\omega)}{(\omega t)^{\frac{d-2}{2}}} d\omega,$$

where $J_\nu(x)$ is the Bessel function of first kind of ν-th order. Any $r(t)$ of this form gives rise to a family of rbf's via

$$K_t(s) = r(\|t - s\|).$$

(Skorohod and Yadrenko 1973, Th. Prob. Applic.)
*skorokhod.yadrenko.1973.pdf

Popular approximation in machine learning:

Given $y_i, t_i, i = 1, ..., n, t(i) \in E^d$;
Choose K s.p.d. on E^d

The Gaussian rbf is popular: $r(\tau) = e^{-\frac{\|\tau\|^2}{2\sigma^2}}$.
σ^2 must be chosen.
Choose "centers", $u_1, ..., u_K \in E^d$ "somehow"; usually the observation points, or a subset of the observation points if the data set is large.

Find $f(\cdot) = \sum_{k=1}^K c_k K u_k(\cdot) = \sum_{k=1}^K c_k r(\|u_k - \cdot\|)$ to minimize

$$\frac{1}{n} \sum_{i=1}^n (f(t_i) - y_i)^2 + \lambda c' J c$$

where $J_{jk} = r(\|u_j - u_k\|), j, k = 1, \cdots, K.$

Some members of the Matern class of radial basis functions (rbf's). $r(\tau)$is the rbf, $\tau = \|s - t\|$.

$r(\tau)$	
$\frac{1}{\alpha}e^{-\alpha\tau}$	$m = 0$
$\frac{1}{\alpha^3}e^{-\alpha\tau}[1 + \alpha\tau]$	$m = 1$
$\frac{1}{\alpha^5}e^{-\alpha\tau}[3 + 3\alpha\tau + \alpha^2\tau^2]$	$m = 2$
$\frac{1}{\alpha^7}e^{-\alpha\tau}[15 + 15\alpha\tau + 6\alpha^2\tau^2 + \alpha^3\tau^3]$	$m = 3$
$\cdots$	$\cdots$

Matern class of covariances are Fourier transforms of an especially simple form: $K_{m,d,\alpha}(s,t) =$

$$\int_{\infty}^{\infty} \cdots \int_{\infty}^{\infty} e^{i(s-t)\cdot\omega}(\|\omega\|^2+\alpha^2)^{-\frac{d+2m+1}{2}} d\omega_1 \cdots d\omega_d.$$

Furthermore, it can be shown that $\|f\|_K^2 =$

$$\int_{\infty}^{\infty} \cdots \int_{\infty}^{\infty} (\|\omega\|^2+\alpha^2)^{\frac{d+2m+1}{2}} |\tilde{f}(\omega_1, \cdots, \omega_d)| d\omega_1 \cdots d\omega_d$$

where d is dimension, $\tilde{f}$ is the Fourier transform of f, and m is related to the number of derivatives that are bounded linear functionals.

The next three slides are courtesy John Carew, see Carew, Dalal, Wahba and Fain, Estimating Arterial Wall Shear Stress(2003), `tr1088.pdf`

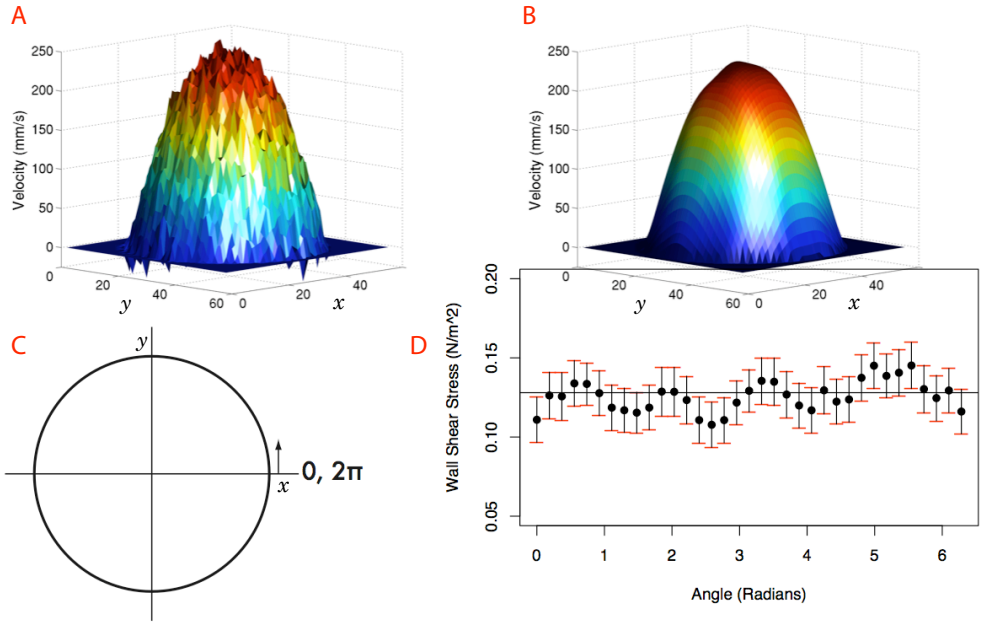


Figure 9: Fluid velocity and wall shear stress estimates in a straight glass tube phantom with a circular cross-section. The raw velocity measurements are in A. The nonparametric fit with $\nu = 4$ and $\lambda = 300$ is in B. The fitted velocity function appears parabolic, which is predicted by physical principles. In D are the WSS estimates and 95% Bayesian confidence intervals. The diagram in C shows how the measurements are ordered in D. Starting at the point with the largest value along the x-axis the order of the WSS estimates increase counter clockwise, indexed by the angle.

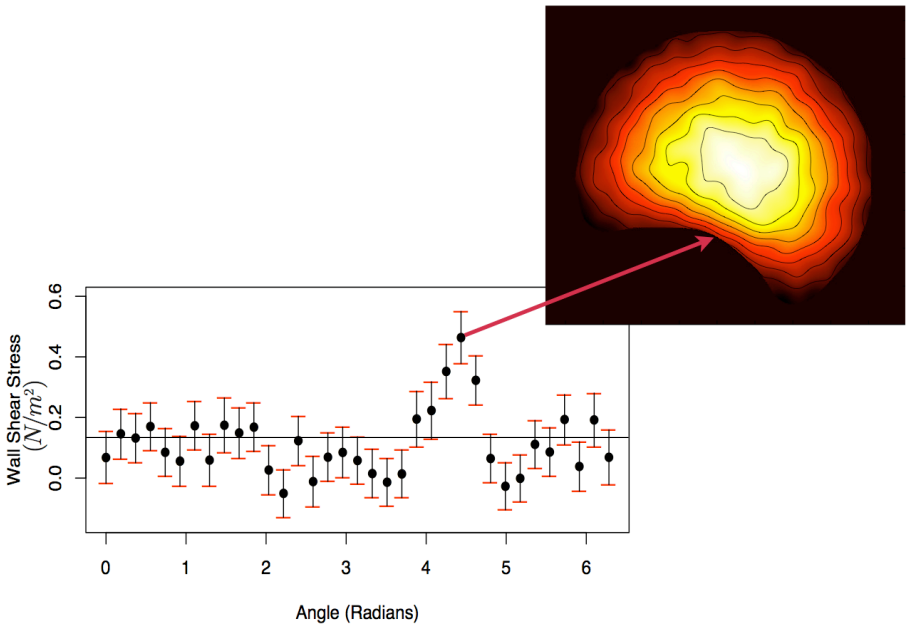


Figure 11: Fluid velocity and wall shear stress estimates in a straight glass tube phantom with non-convex cross section. The nonparametric fit uses $\nu = 4$ and $\lambda = 150$. The shear stress estimates are ordered as described in Figure 9C. The WSS is nearly constant except for the region near the indentation where the shear stress increases. The increased shear can be seen through the more closely-spaced contour lines on the velocity image near the center of the indentation. The horizontal line shows the expected value of WSS calculated for a vessel with a circular cross-section.

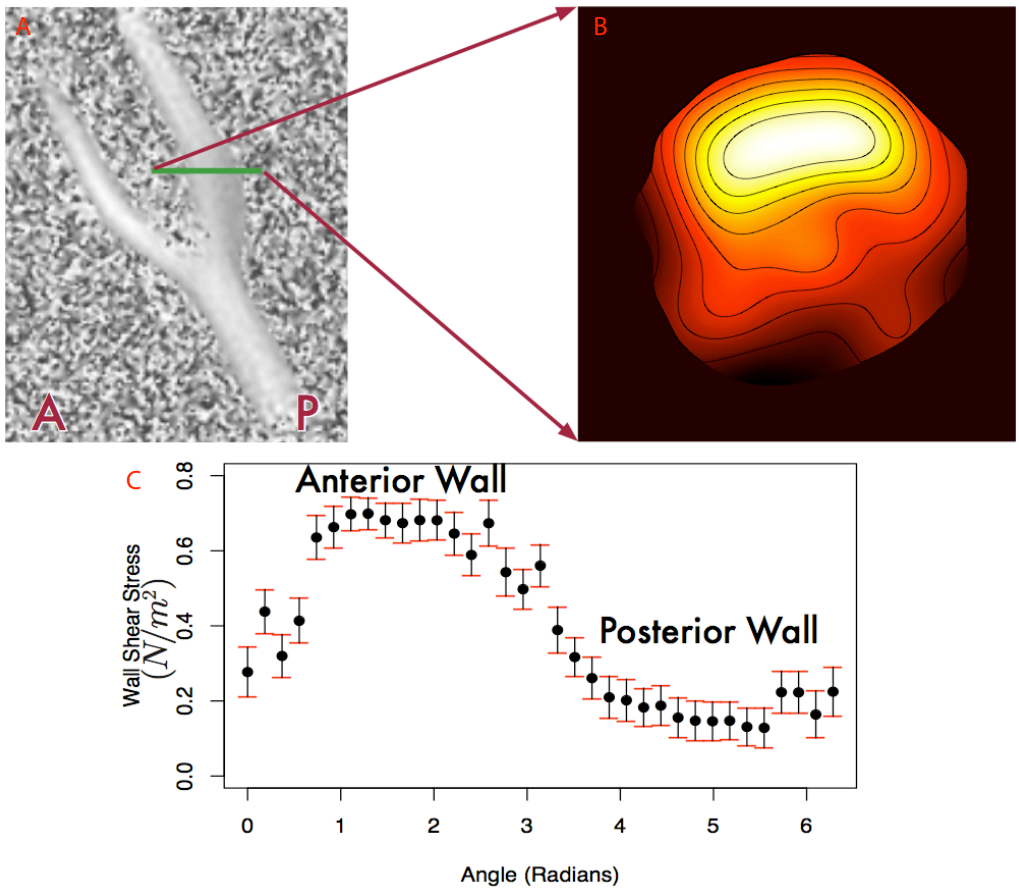


Figure 13: Velocity and WSS in a human carotid bifurcation phantom. The green line segment that bisects the internal carotid artery in A indicates the orientation of the imaging plane. The fitted velocity function with $\nu = 4$ and $\lambda = 100$ is in B. The velocity and rate of change of velocity is much higher along the anterior wall of the vessel than along the posterior wall. The effect of this flow regime is seen in the WSS estimates in C. The point estimates are ordered as in Figure 9 and the intervals are 95% Bayesian confidence intervals.