A Unified Kernel Regression Framework on Arbitrary Surfaces and Its Application to Modeling Multiple Disconnected Subcortical Brain Structures

Chung, M.K., Schaefer, S.M., van Reekum, C.,
Peschke-Schmitz, L., Sutterer, M., Davidson, R.J.

University of Wisconsin-Madison

mkchung@wisc.edu

Abstract

We present a new local kernel regression framework on arbitrary surfaces and manifolds. Starting with a symmetric positive definite kernel as a basic building block, we formulate a new bivariate kernel regression framework that is related to heat diffusion and recently popular spectral graph wavelets. A parametric statistical inference based on the random field theory is also developed for the proposed kernel regression. The method is then applied in investigating the influence of age and gender on amygdala and hippocampus shape in the human brain. We detected a significant age effect on the posterior regions of hippocampi while there is no gender effect present in any of the structures.

1 Introduction

Biomedical Motivations. The amygdala and hippocampus are primary subcortical brain structures involved in emotion and memory (Figure 1). For normal subjects, age and gender could be major factors that affect the functions and structures of these structures, as implied by postmortem studies [1]. The findings on the atrophy of amygdalar and hippocampal structures are somewhat inconsistent. The volume reduction of amygdala and hippocampus due to aging has been found in some studies [2], while other studies did not find such association. For the effect of gender, one study reported significant differences in amygdala and hippocampus volume between the groups [3] whereas others failed to reproduce these [4]. The inconsistency between these studies are mainly due to the fact that all these studies are region of interest (ROI) based volumetry that cannot detect localized subtle anatomical changes along the surface. This provides a biological context for a need to come up with a novel surface-based approach.

Methodological Motivations. The end results of the most surface-based neuroanatomical studies are statistical parametric maps (SPM) [5]. In order to obtain stable and robust SPM, various signal smoothing and filtering methods have been proposed. Among them, diffusion equations, kernel smoothing, and wavelet-based approaches are probably most popular. Diffusion equations have been widely used in image processing as a form of noise reduction starting with Perona and Malik in 1990's [6]. Although numerous techniques have been developed for performing diffusion along surfces [7, 8], most approaches are nonparametric and requires finite element or finite difference schemes which are known to suffer various numerical instability [9]. Kernel smoothing based models have been also proposed for surface and manifolds data [10, 9]. The kernel methods basically smoothes data as weighted average of neighboring mesh vertices using mostly a Gaussian kernel and the process approximates the diffusion process. Recently, wavelets have been popularized for surface and graph data. Spherical wavelets have been used on brain surface data that has been mapped onto a sphere [11, 12]. However, spherical wavelets have an intrinsic problem that they require to establish a smooth mapping from the surface to a unit sphere, which introduces a serious

metric distortion. To remedy the limitation of spherical wavelets, spectral graph wavelet transform (SGWT) defined on a graph has been applied to arbitrary surface meshes by treating surface meshes as graphs [13, 14, 15]. Although diffusion-, kernel- and wavelet-based methods all look different from each other, it is possible to develop a unified framework that relates all of them in a coherent mathematical fashion.

Contributions. Starting with a kernel, we derive a unified kernel regression framework within the Hilbert space theory. The proposed kernel regression works for any symmetric positive definite kernel, which behaves like weights between two functional data. We show how this facilitates a coherent statistical inference for functional signals defined on an arbitrary manifold. The focus of the paper is on the theoretical development of the proposed kernel regression on manifolds.

(i) In particular, we show how the proposed bivariate kernel regression is related to diffusion-like equations. Hence, the kernel regression inherits various characteristics and properties of diffusion-like equations.

(ii) We establish the relationship between the kernel regression and recently popular spectral graph wavelets for manifolds. In fact, we show that the proposed kernel regression is equivalent to the wavelet coefficients. This mathematical equivalence levitates a need for constructing wavelets using a complicated machinery as often done in previous studies [13, 14, 15].

(iii) A unified statistical inference framework is developed for neuroimaging applications by linking the kernel regression to the random field theory [16, 17]. This levitates the need for using nonparametric procedures such as false discovery rates (FDR) or permutation tests that do not have explicate control over the scale of analysis.

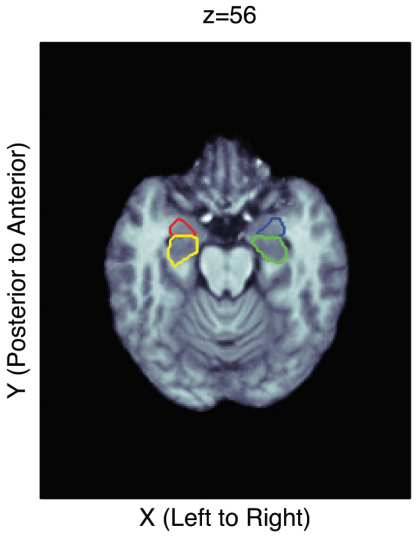

Figure 1: Left (red) and right (blue) amygdale, and left (yellow) and right (green) hippocampi are manually segmented on MRI and used to obtain subcortical surface models. We are interested in quantifying local shape changes over different gender and age within these small structures.

Subsequently, we illustrate how the kernel regression procedure can be used to localize anatomical signal within the multiple subcortical structures of the human brain.

2 Motivating Problems in Euclidean Space

In this section, we illustrate two problems in an Euclidean space that motivate the development of the proposed kernel regression.

Local Polynomial Regression. Consider measurement y_i sampled at $p_i \in \mathbb{R}^d$. The measurements are usually modeled as $y_i = h(p_i) + \epsilon_i$ with some noise ϵ_i and unknown mean function h that has to be estimated. In the well known local polynomial/kernel regression frameworks [18, 19], a univariate kernel $G(p)$ is introduced to minimize the weighted least squares of the form

$$\widehat{h}(p) = \arg \min_{\beta_0 \cdots \beta_k} \sum_{i=1}^{n} G(p - p_i) \Big| y_i - \sum_{j=0}^{k} \beta_j (p - p_i)^j \Big|^2. \tag{1}$$

Most often Gaussian kernels are used. There are many different variations and names to (1). However, in most local polynomial/kernel regression frameworks, the kernel G and polynomial basis p^j are translated by the amount of p_i in fitting the data locally. In this fashion, at each data point p_i, exactly the same shape of kernel and distance are used. However, one immediately encounters a difficulty of directly generalizing (1) to an arbitrary surface since it is unclear how to translate the kernel and basis in a coherent fashion.

Wavelets. A similar problem is also encountered in wavelets in a Euclidean space. Consider a wavelet basis $W_{t,q}(p)$ obtained from a mother wavelet W with scale and translation parameters t and q:

$$W_{t,q}(p) = \frac{1}{t} W\big(\frac{p - q}{t}\big).$$

Scaling a function on a surface is trivial. But the difficulty arises when one tries to define a mother wavelet and translate it on a surface.

To remedy these two different problems, we propose to use a bivariate kernel and bypass the problem of translating a univariate kernel in By simply changing the second argument, it has the effect of translating the kernel. By careful construction, we introduce a new kernel method that works as local kernel regression and wavelets simultaneously on an arbitrary surface.

3 Preliminary

Hilbert Space. In neuroanatomical studies, measurements are sampled densely at the resolution 1mm or less so it is more practical to model the measurements as functions. Consider a functional measurement f defined on a manifold $\mathcal{M} \subset \mathbb{R}^d$. We assume the following additive model:

$$f(p) = h(p) + \epsilon(p), \tag{2}$$

where h is the unknown signal and ϵ is a zero-mean random field, possibly Gaussian. The manifold $\mathcal{M}$ can be a single connected component or multiple disjoint components as our hippocampus/amygdala application. We further assume $f \in L^2(\mathcal{M})$, the space of square integrable functions on $\mathcal{M}$ with the inner product $\langle f, g \rangle = \int_{\mathcal{M}} f(p)g(p)\ d\mu(p)$, where μ is the Lebesgue measure. Define a self-adjoint operator $\mathcal{L}$ satisfying $\langle g_1, \mathcal{L}g_2 \rangle = \langle \mathcal{L}g_1, g_2 \rangle$ for all $g_1, g_2 \in L^2(\mathcal{M})$. Then $\mathcal{L}$ induces the eigenvalues λ_j and eigenfunctions ψ_j on $\mathcal{M}$:

$$\mathcal{L}\psi_j = \lambda_j \psi_j. \tag{3}$$

Without loss of generality, we can order the eigenvalues $0 = \lambda_0 \leq \lambda_1 \leq \lambda_2 \leq \cdots$. The eigenfunctions ψ_j form an orthonormal basis in $L^2(\mathcal{M})$. Then any symmetric positive definite kernel can be written as

$$K(p, q) = \sum_{j=0}^{\infty} \tau_j \psi_j(p) \psi_j(q) \tag{4}$$

for some τ_j. This is related to Mercer's theorem [20]. The constants τ_j are identified as follows. Consider the kernel convolution on the eigenfunction ψ_j:

$$K * \psi_j(p) = \int_{\mathcal{M}} K(p, q)\psi_j(q)\ d\mu(q). \tag{5}$$

Substituting (4) into (5), we have $K * \psi_j(p) = \tau_j \psi_j(p)$. Hence τ_j and ψ_j must be the eigenvalues and eigenfunctions of the convolution. So far the relationship between the two different sets of eigenvalues λ_j and τ_j is not given.

Fourier Analysis. In the usual Fourier analysis setting, the unknown signal h is estimated in the subspace $\mathcal{H}_k \subset L^2(\mathcal{M})$ spanned by the orthonormal basis $\{\psi_j\}$, i.e.

$$\mathcal{H}_k = \{\sum_{j=0}^{k} \beta_j \psi_j(p) : \beta_j \in \mathbb{R}\}.$$

Then the least squares estimation (LSE) of h in $\mathcal{H}_k$ is given by the shortest distance from f to $\mathcal{H}_k$:

$$\widehat{h}(p) = \arg\min_{h \in \mathcal{H}_k} \int_{\mathcal{M}} \left| f(p) - h(p) \right|^2 d\mu(p) = \sum_{j=0}^{k} f_j \psi_j(p), \tag{6}$$

where $f_j = \langle f, \psi_j \rangle$ are the Fourier coefficients. Figure 2 shows an example of LSE with $\mathcal{L}$ as the Laplace-Beltrami operator and $k = 1000$.

4 Kernel Regression on Surfaces

Instead of estimating the function h by finding the closest function in $\mathcal{H}_k$, we weight the distance with a positive definite symmetric kernel K:

$$\widehat{h}(p) = \arg\min_{h \in \mathcal{H}_k} \int_{\mathcal{M}} \int_{\mathcal{M}} K(p, q) \Big| f(q) - h(p) \Big|^2 d\mu(q)\ d\mu(p). \tag{7}$$

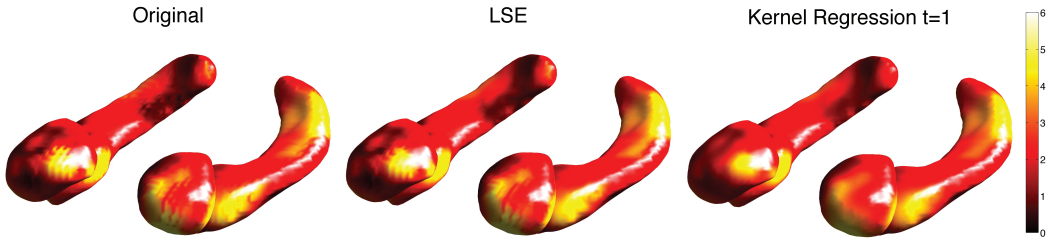

Figure 2: The displacement length, its LSE using the Laplace-Beltrami eigenfunctions and kernel regression with $t = 1$ are displayed for a subject. The strip patterns visible in the amygdale in Original and LSE are caused by image acquisition and processing and eventual discretization. They actually correspond to image slices passing through them. On the other hand kernel regression, which is equivalent to diffusion wavelet coefficients reduces such artifacts.

Without loss of generality, we will assume the kernel to be a probability distribution so that

$$\int_{\mathcal{M}} K(p,q)\, d\mu(q) = 1$$

for all $p \in \mathcal{M}$. If the kernel is a Dirac-delta function, the kernel regression simply collapses to the usual Fourier series expansion. The main contribution of the paper is the instruction of this new kernel regression framework and establishing the relationship to heat diffusions and wavelets. We can show that the solution to optimization (7) is analytically given as follows.

Theorem 1

$$\widehat{h}(p) = \arg\min_{h \in \mathcal{H}_k} \int_{\mathcal{M}} \int_{\mathcal{M}} K(p,q) \Big| f(q) - h(p) \Big|^2 \, d\mu(q)\, d\mu(p) = \sum_{j=0}^{k} \tau_j f_j \psi_j.$$

Proof. Any function $h \in \mathcal{H}_k$ can be expressed as $h(p) = \sum_{j=0}^{k} \beta_j \psi_j(p)$. By plugging the expansion into the inner integral $I(p)$, we obtain

$$I(p) = \int_{\mathcal{M}} K(p,q) \Big| f(q) - \sum_{j=0}^{k} \beta_j \psi(p) \Big|^2 \, d\mu(q).$$

Simplifying the expression, we obtain

$$I(p) = \sum_{j=0}^{k} \sum_{j'=0}^{k} \psi_j(p)\psi_{j'}(p)\beta_j\beta_{j'} - 2K * f(p) \sum_{j=0}^{k} \psi_j(p)\beta_j + K * f^2(p). \tag{8}$$

From (5), the convolution $K * f$ can bewritten as

$$K * f(p) = \sum_{j'=0}^{\infty} \tau_{j'} f_{j'} \psi_{j'}(p).$$

Since I is an unconstrained positive semidefinite quadratic program (QP) in β_j, there is no unique global minimizer of I without additional linear constraints. Integrating I further with respect to $d\mu(p)$, we collapses (8) to a positive definite QP, which yields a unique global minimizer:

$$\int_{\mathcal{M}} I(p)\, d\mu(p) = \sum_{j=0}^{k} \beta_j^2 - 2 \sum_{j=0}^{k} \tau_j f_j \beta_j + \text{ const.}$$

The minimum of the above integral is obtained when all the partial derivatives with respect to β_j vanish, i.e.

$$\int_{\mathcal{M}} \frac{\partial I}{\partial \beta_j}\, d\mu(p) = 2\beta_j - 2\tau_j f_j = 0$$

for all j. Hence $\sum_{j=0}^{k} \tau_j f_j \psi_j$ must be the unique minimizer. □

Theorem 1 implies that the kernel regression can be done by simply computing the Fourier coefficients $f_j = \langle f, \psi_j \rangle$ without doing messy numerical optimization. Note that as $k \to \infty$, the kernel regression $\widehat{h} = \sum_{j=0}^{k} \tau_j f_j \psi_j$ converges to convolution $K * f$ establishing the connection to the kernel smoothing framework [21, 9]. Hence, asymptotically kernel regression should inherit many statistical properties of kernel smoothing on manifolds.

Connection to Diffusion. The kernel regression can be shown to be related to the following diffusion-like Cauchy problem.

Theorem 2 *For an arbitrary self-adjoint differential operator $\mathcal{L}$, the unique solution of the following initial value problem*

$$\frac{\partial g(p,t)}{\partial t} + \mathcal{L} g(p,t) = 0, g(p, t=0) = f(p) \tag{9}$$

is given by

$$g(p,t) = \sum_{j=0}^{\infty} e^{-\lambda_j t} f_j \psi_j(p). \tag{10}$$

Due to the space limitation, we will not provide a proof but the readers can verify the result by plugging the solution to the equation. If we let $\tau_j = e^{-\lambda_j t}$, we can see that our proposed kernel regression $\widehat{h} = \sum_{j=0}^{k} \tau_j f_j \psi_j$ should converge to the solution of diffusion-like equation. If we let $\mathcal{L}$ be the Laplace-Beltrami operator, (9) becomes an isotropic diffusion equation as a special case and we are then dealing with heat kernel

$$H_t(p,q) = \sum_{j=0}^{\infty} e^{-\lambda_j t} \psi_j(p) \psi_j(q),$$

which is often explored mathematical objects in various fields [21, 9].

Connection to Wavelets. In order to construct wavelets on an arbitrary graph and mesh, spectral graph wavelet transform (SGWT) has been proposed recently [13, 14, 15]. However, SGWT-based so called diffusion wavelet construction has been fairly involving so far. However, it can be shown to be a special case of the proposed kernel regression. Following closely the notations in [13, 14, 15], for some scale function g that satisfies the admissibility conditions, diffusion wavelet $W_{t,p}(p)$ at position p and scale t is given by

$$W_{t,q}(p) = \sum_{j=0}^{k} g(\lambda_j t) \psi_j(p) \psi_j(q).$$

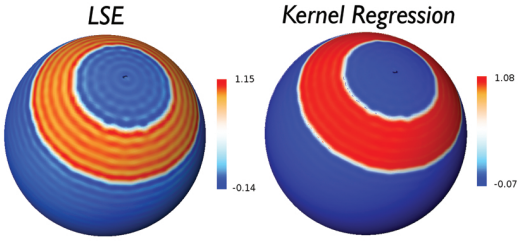

Figure 3: Gibbs phenomenon (ringing artifacts) is visible in the Fourier series expansion via LSE of the step function defined on a sphere. On the other hand, the kernel regression (wavelet coefficients) shows less visible artifacts.

If we let $\tau_j = g(\lambda_j t)$, the wavelet coefficient is given by

$$\langle W_{t,p}, f \rangle = \int_{\mathcal{M}} W_{t,q}(p) f(p) \, d\mu(p) = \sum_{j=0}^{k} \tau_j f_j \psi_j(q),$$

which is the exactly kernel regression we introduced. Hence, diffusion wavelet coefficients can be simply obtained by doing the kernel regression without an additional wavelet machinery [15]. Further, if we let $g(\lambda_j t) = e^{-\lambda_j t}$, we have

$$W_{t,p}(q) = H_t(p,q),$$

which is a heat kernel. The bandwidth t of heat kernel controls resolution while the translation is done by shifting one argument in the kernel.

Although the kernel regression is constructed using global basis functions, remarkably the kernel regression at each point p coincides with the wavelet coefficient at that point. Hence, it also inherits all the localization property of wavelets at that point. This is clearly demonstrated in a simulation given in Figure 3, where a step function of value 1 in the circular band $1/8 < \theta < 1/4$ (angle from the north pole) and of value 0 outside of the band is constructed. Then the step function is reconstructed using the Fourier series expansion via LSE and kernel regression. For the kernel regression, the heat kernel with the small bandwidth $t = 0.0001$ is used. LSE clearly shows severe ringing artifacts compared to the kernel regression (wavelet coefficients).

5 Statistical Inference for Kernel Regression

The proposed kernel regression can be naturally integrated into the random field theory based statistical inference framework [16, 17]. Consider a functional measurements $f^1, \cdots, f^n$ on a manifolds $\mathcal{M}$. In a simplest statistical setting, the measurements can be modeled as

$$f^i(p) = h(p) + \epsilon^i(p),$$

where h is an unknown group level signal and ϵ^i are zero-mean unit-variance Gaussian random fields. The model assumptions are not as restrictive as it seems since we can always normalize the data in this fashion. We further assume the random field ϵ^i is modeled as the convolution of kernel H_s on Gaussian white noise W with small bandwidth s, i.e. $\epsilon^i(p) = K_s * W(p)$. We are then interested in determining the significance of h, i.e.

$$H_0 : h(p) = 0 \text{ for all } p \in \mathcal{M} \text{ vs. } H_1 : h(p) > 0 \text{ for some } x \in \mathcal{M}. \tag{11}$$

Note that any point p_0 that gives $h(p_0) > 0$ is considered as signal. (11) is an infinite dimensional multiple comparisons problem for continuously indexed hypotheses.

Given a test statistic, often a T-field $T(p)$, which is T-statistic at each point p, we need to compute the multiple comparison corrected type-I error of rejecting the null hypothesis (there is signal) when the null hypothesis is true (there is no signal). For sufficiently high threshold z, which corresponds to the observed maximum T-statistic value, it is known that the corrected type-I error is given by

$$P\Big(\sup_{p\in\mathcal{M}} T(p) > z\Big) = \sum_{j=0}^{d} \mu_j(\mathcal{M})\rho_j(z),$$

where $\mu_d(\mathcal{M})$ is the j-th Minkowski functional or intrinsic volume of $\mathcal{M}$ and ρ_j is the j-th Euler characteristic (EC) density of T-field [16, 17].

The our application, hippocampus and amygdala surfaces are compact with no boundary, so the Minkowski functionals are simply $\mu_2(\mathcal{M}) = \text{area}(\mathcal{M})/2$, $\mu_2(\mathcal{M}) = 0$ and $\mu_0(\mathcal{M}) = \chi(\mathcal{M}) = 4 \times 2$, the Euler characteristic of $\mathcal{M}$. The EC-densities of the T-field with ν degrees of freedom [17, 16] is

$$\begin{aligned}
\rho_0(z) &= 1 - P(T_\nu \le z), \\
\rho_1(z) &= \frac{1}{\sqrt{2s^2}} \cdot \frac{1}{2\pi}\Big(1 + \frac{z^2}{\nu}\Big)^{-(\nu-1)/2}, \\
\rho_2(z) &= \frac{1}{2s^2} \cdot \frac{1}{(2\pi)^{3/2}} \frac{\Gamma(\frac{\nu+1}{2})}{(\frac{\nu}{2})^{1/2}\Gamma(\frac{\nu}{2})} z\Big(1 + \frac{z^2}{\nu}\Big)^{-(\nu-1)/2}.
\end{aligned}$$

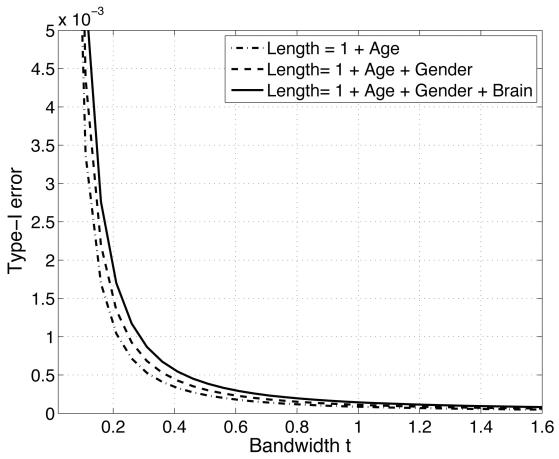

Figure 4: Type-I error plot over bandwidth t of kernel regression for three different statistical models for our amygdala/hippocampus data. As the bandwidth increases, the multiple comparisons corrected type-I error decreases. As the complexity of model increases, the error is expected to increase. That is why the most complex model has higher type-I error than the other two models. In all three models, the error flattens around the bandwidth 1 so the bandwidth $t = 1$ is chosen in the study.

Note that EC-densities has the term kernel bandwidth $2s^2$ which relates the smoothness of noise via $\epsilon^i = K_s * W$. In the usual SPM framework [16, 17], signals are usually convolved with a kernel with much larger bandwidth than s effectively masking the smoothness of noise to be t. Hence we perform kernel regression on f^i with kernel H_t and effectively model the data as

$$H_t * f^i(p) = H_t * h(p) + H_t * W(p)$$

and replace s in the EC-density with t. Figure 4 shows the type-I error plot over different bandwidth t of the kernel regression. As the bandwidth t goes to zero, the type-I error increases. When $t = 0$, the kernel regression collapse to the usual Fourier series expansion. Hence, kernel regression can be viewed as having smaller type-I error compared to the Fourier series expansion.

6 Application

Brain Imaging Data. The study consists of high resolution T1-weighted inverse recovery fast gradient echo anatomical 3D images, collected in 124 contiguous 1.2-mm axial slices (TE=1.8 ms; TR=8.9 ms; flip angle = 10°; FOV = 240 mm; 256 × 256 data acquisition matrix) of 69 middle age and elderly adults ranging between 38 to 79 years (mean age = 58.0 ± 11.3 years). There are 23 males and 46 females. The data were collected as a part of a national study for the health and well-being in the aged population. The amygdalae and hippocampi were manually segmented by a trained individual rater in the native space (Figure 1). The segmented volumes did not yield any age or gender effects at 0.05 level. This gives an additional motivation for developing a more complex surface-based shape analysis.

Brain tissues in the MRI scans were first segmented using Brain Extraction Tool (BET) [22], and then a nonlinear image registration using the diffeomorphic shape and intensity averaging technique with cross-correlation as similarity metric was performed [23]. A study-specific template was constructed. Using the deformation field of warping the individual brain to the template, we deformed the amygdala and hippocampus binary masks to the template space. The normalized masks were then averaged to produce the subcortical template. The isosurfaces of the subcortical masks are extracted using the marching cube algorithm.

The displacement from the template to an individual surface is obtained at each mesh vertex. Since the length measurement provides a much easier biological interpretation, we used the length of displacement vector as a response variable among many other possible features. Since the length on the template surface is expected to be noisy due to image acquisition, segmentation and image registration errors, it is necessary perform the proposed kernel regression and subsequently reduce the type-I error. Figure 2 shows an example of kernel regression on our data.

Numerical Implementation. The Laplace-Beltrami operator is chosen as the self-adjoint operators $\mathcal{L}$ of choice. The eigenfunctions of the Laplace-Beltrami operator on an arbitrary curved surface is analytically unknown. So it is necessary to discretize (3) using the Cotan formulation as a generalized eigenvalue problem [24, 25]:

$$\mathbf{C}\psi = \lambda \mathbf{A}\boldsymbol{\psi}, \tag{12}$$

where $\mathbf{C}$ is the stiffness matrix, $\mathbf{A}$ is the mass matrix and $\boldsymbol{\psi} = (\psi(p_1), \cdots, \psi(p_n))'$ is the eigenfunction evaluated at n mesh vertices. Once we obtained the basis functions ψ_j, the corresponding Fourier coefficients β_j are estimated as

$$\beta_j = \mathbf{f}'\mathbf{A}\boldsymbol{\psi}_j, \tag{13}$$

where $\mathbf{f} = (f(p_1), \cdots, f(p_n))'$ and $\boldsymbol{\psi}_j = (\psi_j(p_1), \cdots, \psi_j(p_n))'$ [24].

For the Laplace-Beltrami operator, the heat kernel is the corresponding kernel. We have chosen the parameters $t = 1$ and $k = 1000$ number of basis for the subsequent kernel regression. The number of eigenfunctions used is more than sufficient to guarantee relative accuracy less than 0.3% against the ground truth. Kernel regression is also performed on the surface coordinates of the template for better visualization.

Results. Since test statistics are constructed over all mesh vertices on the mandible, multiple comparisons are accounted for using the random field theory [17]. The random field theory assumes the measurements to be smooth Gaussian random field. After the displacement lengths are smoothed, we constructed the T-field testing the length difference.

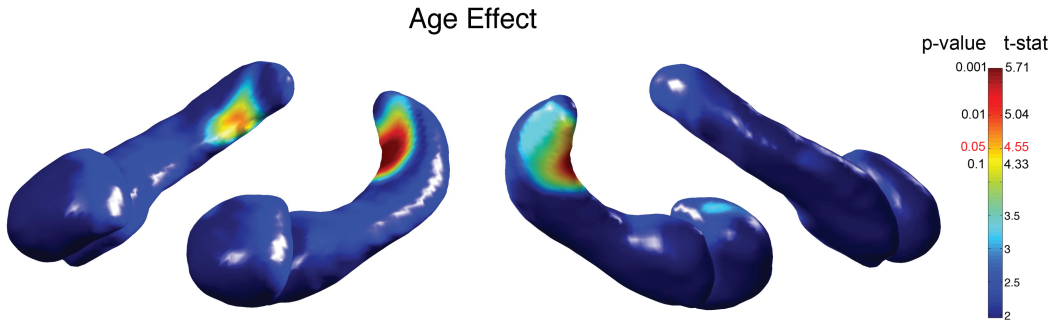

Figure 5: T-statistic and the corrected p-value maps on the amygdala/hippocampus template showing age effect. The posterior regions of the both left and right hippocampi show a significant age effect at 0.05 level. However, there is no gender effect present in any of the structures.

Subsequently $Length$ is regressed over the total brain volume and other variables:

$$Length = \beta_1 + \beta_2 \cdot Brain + \beta_3 \cdot Age + \beta_4 \cdot Gender + \epsilon, \tag{14}$$

where ϵ is zero mean Gaussian noise. The age and gender effects are determined by testing the significance of parameters β_3 and β_4 at $\alpha = 0.05$ using T-statistic and corrected for multiple comparisons. The result is displayed in Figure 5.

Age effect. We found the region of significant effect of age on the posterior part of hippocampi (left: max. T-stat = 6.25, $p = 0.00014$ (corrected); right: max. T-stat = 4.78, p = 0.024 (corrected)). Particularly, on the caudal regions of the left and right hippocampi, we found highly localized signals. It is consistent with other shape modeling studies on hippocampus [26, 5]. However, we did not find any age effects on the amygdala surface at $\alpha = 0.05$.

Gender effect. We did not detect any significant gender effects on any parts of amygdale or hippocampi at 0.05 level.

7 Conclusion

We have developed a new kernel regression that unifies bivariate kernel regression, heat diffusion and wavelets in a single coherent mathematical framework. The kernel regression is both global and local in a sense it uses global basis functions to perform regression but locally equivalent to diffusion wavelet coefficients. The proposed framework is demonstrated to reduce type-I error in modeling shape variations compared to the usual Fourier series expansion. The ability to localize subtle age-related morphological differences may provide an anatomical evidence for the functional digression in hippocampus.

References

[1] A.K.H. Miller, R.L. Alston, and J.A.N. Corsellis, "Variation with age in the volumes of grey and white matter in the cerebral hemispheres of man: measurements with an image analysis," *Neuropathology and Applied Neurobiology*, vol. 6, no. 2, pp. 119–132, 1980.

[2] E.D. Bigler, D.D. Blatter, C.V. Anderson, S.C. Johnson, S.D. Gale, R.O. Hopkins, and B. Burnett, "Hippocampal volume in normal aging and traumatic brain injury," *American Journal of Neuroradiology*, vol. 18, pp. 11, 1997.

[3] C.D. Good, I. Johnsrude, J. Ashburner, R.N.A. Henson, K.J. Friston, and R.S.J. Frackowiak, "Cerebral asymmetry and the effects of sex and handedness on brain structure: a voxel-based morphometric analysis of 465 normal adult human brains," *NeuroImage*, vol. 14, pp. 685–700, 2001.

[4] R.C. Gur, F. Gunning-Dixon, W.B. Bilker, and R.E. Gur, "Sex differences in temporo-limbic and frontal brain volumes of healthy adults," *Cerebral Cortex*, vol. 12, pp. 998–1003, 2002.

[5] Y. Xu, D.J. Valentino, A.I. Scher, I. Dinov, L.R. White, P.M. Thompson, L.J. Launer, and A.W. Toga, "Age effects on hippocampal structural changes in old men: the haas," *NeuroImage*, vol. 40, pp. 1003–1015, 2008.

[6] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 12, pp. 629–639, 1990.

[7] A. Andrade, F. Kherif, J. Mangin, K.J. Worsley, A. Paradis, O. Simon, S. Dehaene, D. Le Bihan, and J-B. Poline, "Detection of fmri activation using cortical surface mapping," *Human Brain Mapping*, vol. 12, pp. 79–93, 2001.

[8] B. Tang, G. Sapiro, and V. Caselles, "Direction diffusion," in *The Proceedings of the Seventh IEEE International Conference on Computer Vision*, 1999, pp. 2:1245–1252.

[9] M.K. Chung, S. Robbins, and A.C. Evans, "Unified statistical approach to cortical thickness analysis," *Information Processing in Medical Imaging (IPMI), Lecture Notes in Computer Science*, vol. 3565, pp. 627–638, 2005.

[10] Mikhail Belkin, Partha Niyogi, and Vikas Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *The Journal of Machine Learning Research*, vol. 7, pp. 2399–2434, 2006.

[11] D. Nain, M. Styner, M. Niethammer, J.J. Levitt, M.E. Shenton, G. Gerig, A. Bobick, and A. Tannenbaum, "Statistical shape analysis of brain structures using spherical wavelets," in *IEEE Symposium on Biomedical Imaging ISBI*, 2007.

[12] J. Bernal-Rusiel, M. Atienza, and J. Cantero, "Detection of focal changes in human cortical thickness: spherical wavelets versus gaussian smoothing," *NeuroImage*, vol. 41, pp. 1278–1292, 2008.

[13] J.-P. Antoine, D. Roşca, and P. Vandergheynst, "Wavelet transform on manifolds: old and new approaches," *Applied and Computational Harmonic Analysis*, vol. 28, pp. 189–202, 2010.

[14] D.K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Applied and Computational Harmonic Analysis*, vol. 30, pp. 129–150, 2011.

[15] W.H. Kim, D. Pachauri, C. Hatt, M.K. Chung, S. Johnson, and V. Singh, "Wavelet based multiscale shape features on arbitrary surfaces for cortical thickness discrimination," in *Advances in Neural Information Processing Systems (NIPS) 25*, 2012, pp. 1250–1258.

[16] J.E. Taylor and K.J. Worsley, "Detecting sparse signals in random fields, with an application to brain mapping," *Journal of the American Statistical Association*, vol. 102, pp. 913–928, 2007.

[17] K.J. Worsley, J.E. Taylor, F. Tomaiuolo, and J. Lerch, "Unified univariate and multivariate random field theory," *NeuroImage*, vol. 23, pp. S189–195, 2004.

[18] J. Fan and I. Gijbels, *Local Polynomial Modelling and Its Applications*, Chapman & Hall/CRC, 1996.

[19] A.C. Öztireli, G. Guennebaud, and M. Gross, "Feature preserving point set surfaces based on non-linear kernel regression," in *Computer Graphics Forum*, 2009, vol. 28, pp. 493–501.

[20] J.B. Conway, *A Course in Functional Analysis*, Springer, 1990.

[21] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," *Advances in Neural Information Processing Systems*, vol. 1, pp. 585–592, 2002.

[22] S.M. Smith, "Fast robust automated brain extraction," *Human Brain Mapping*, vol. 17, pp. 143–155, 2002.

[23] B.B. Avants, C.L. Epstein, M. Grossman, and J.C. Gee, "Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain," *Medical Image Analysis*, vol. 12, pp. 26–41, 2008.

[24] H. Zhang, O. van Kaick, and R. Dyer, "Spectral methods for mesh processing and analysis," in *EUROGRAPHICS*, 2007, pp. 1–22.

[25] A. Qiu, D. Bitouk, and M.I. Miller, "Smooth functional and structural maps on the neocortex via orthonormal bases of the laplace-beltrami operator," *IEEE Transactions on Medical Imaging*, vol. 25, pp. 1296–1396, 2006.

[26] P.M. Thompson, K.M. Hayashi, G. de Zubicaray, A.L. Janke, S.E. Rose, J. Semple, M.S. Hong, D.H. Herman, D. Gravano, D.M. Doddrell, and A.W. Toga, "Mapping hippocampal and ventricular change in alzheimer disease," *NeuroImage*, vol. 22, pp. 1754–1766, 2004.