Quantifying T2-FLAIR Mismatch using Geographically Weighted Regression and Predicting Molecular Status in Lower Grade Gliomas

Shariq Mohammed[a,b,*], Visweswaran Ravikumar[b], Elisa Warner[b], Sohil H Patel[c], Spyridon Bakas[e], Arvind Rao[a,b,f,g,h,*], Rajan Jain[i,j]

[a]Department of Biostatistics, University of Michigan, Ann Arbor, MI.
[b]Department of Computational Medicine & Bioinformatics, University of Michigan, Ann Arbor, MI.
[c]Department of Radiology & Medical Imaging, University of Virginia School of Medicine, Charlottesville, VA.
[d]Department of Radiology, University of Pennsylvania, Philadelphia, PA.
[e]Department of Pathology & Laboratory Medicine, University of Pennsylvania, Philadelphia, PA.
[f]Department of Radiation Oncology, University of Michigan, Ann Arbor, MI.
[g]Michigan Institute for Data Sciences, University of Michigan, Ann Arbor, MI.
[h]Department of Biomedical Engineering, University of Michigan, Ann Arbor, MI.
[i]Department of Radiology, New York University Grossman School of Medicine, New York, NY.
[j]Department of Neurosurgery, New York University Grossman School of Medicine, New York, NY.
[*]corresponding author. E-mail: ukarvind@umich.edu

Abstract

In this document, we provide the details about (a) geographically weighted regression, (b) computations on the space of probability density functions (PDFs), (c) permutation-based hypothesis tests for PDFs, (d) classification model, and (e) the computation time.

Keywords: geographically weighted regression, principal component analysis, probability density functions, spatial analysis

S1 Geographically Weighted Regression

Geographically weighted regression (GWR) is a spatial analysis technique to study the spatially varying relationships between the response and covariates in a regression model [1].

Model. The regression model for GWR can be fitted at each observed location of the tumor pixel from the MRI scans. For each tumor pixel $s = 1, \ldots, n$, where n is the total number of tumor pixels in the MRI, the GWR model is given as

$$y_s = \beta_{s0} + x_s \beta_s + \epsilon_s, \tag{S1}$$

where y_s and x_s are the observed response and the predictor/covariate at tumor pixel s, respectively. That is, x_s and y_s are considered as the intensities of the tumor pixel s from T2 and FLAIR MR images, respectively. Here, β_{s0} is the intercept, β_s is the regression coefficient, and ϵ_s is the random error. Let $\mathbf{x}_s = (1, x_s)$ and $\boldsymbol{\beta}_s = (\beta_{s0}, \beta_s)$ be column vectors, then the model in Equation (S1) can be simplified as $y_s = \mathbf{x}_s^\top \boldsymbol{\beta}_s + \epsilon_s$. The estimated coefficient $\hat{\boldsymbol{\beta}}_s$ at tumor pixel s is given as $\hat{\boldsymbol{\beta}}_s = [\mathbf{X}^\top \mathbf{W}_s \mathbf{X}]^{-1} \mathbf{X}^\top \mathbf{W}_s \mathbf{y}$, where $\mathbf{X} = [\mathbf{x}_1; \ldots; \mathbf{x}_n]^\top$ and $\mathbf{y} = (y_1, \ldots, y_n)$. Here, $\mathbf{W}_s$ is the local weight matrix computed at each tumor pixel $s = 1, \ldots, n$. Note that the weight matrix is different for each tumor pixel as it depends on the neighboring tumor pixels, and needs to be computed separately to compute each $\hat{\boldsymbol{\beta}}_s$.

Local Weight Matrix. To compute the local weights matrix, $\mathbf{W}_s$, we use a kernel function that assigns higher weights for other tumor pixels closer to the tumor pixel s, and these weights decrease as the distance between the tumor pixels increases. A Gaussian kernel is one of the most commonly used kernel functions and is defined as $W_{ss'} = \exp(-d_{ss'}^2/2\gamma)$, where $d_{ss'}$ is the distance between locations s and s'. Here γ is called the bandwidth parameter that controls the range and decay of the spatial correlation. There are several other choices for the kernel function (e.g. exponential and bisquare kernels) which can be chosen based on the application. However, the choice of the bandwidth parameter γ is crucial and is usually estimated based on the data. The bandwidth parameter can be estimated using different approaches such as direct assignment based on the number of neighbors of interest [8], cross-validation [1], or corrected Akaike Information Criterion [3]. We use the latter in our work.

Residuals. The GWR residuals are scaled between 0 and 1 (using the maximum and minimum residual values across all the subjects in the training data set) before constructing the PDF.

S2 Computations on the Space of PDFs

Let f be a probability density function (PDF) defined on the domain $[0,1]$ and $\mathcal{F}$ denote the collection of such PDFs defined as $\mathcal{F} = \{f : [0,1] \to \mathbb{R}_+ | \int_0^1 f(x)dx = 1\}$.

Space of Square-Root Transformations. We denote the space of SRTs by $\mathcal{H} = \{h : [0,1] \to \mathbb{R}^+ | \int_0^1 h(x)^2 dx = 1\}$. The space of the SRTs $\mathcal{H}$ represents the positive orthant of a unit Hilbert sphere [6], and is equipped with the $\mathbb{L}^2$ Riemmanian metric. If $T_h(\mathcal{H}) = \{\delta h : [0,1] \to \mathbb{R} | \int_0^1 h(x)\delta h(x)dx = 0\}$ denotes the tangent space of $\mathcal{H}$ at the SRT h, then the $\mathbb{L}^2$ Riemmanian metric on $\mathcal{H}$ can be defined as $\langle\langle \delta h_1, \delta h_2 \rangle\rangle = \int_0^1 \delta h_1(t)\delta h_2(t)dt$, where $\delta h_1, \delta h_2 \in T_h(\mathcal{H})$. The distance between any two PDFs f_1 and f_2 can now be computed using the geometry of $\mathcal{H}$ equipped with the $\mathbb{L}^2$ metric. That is, the geodesic distance between $h_1, h_2 \in \mathcal{H}$, which are the SRTs corresponding to f_1 and f_2, is given as $d(h_1, h_2) = \theta = \cos^{-1}\big(\int_0^1 h_1(x)h_2(x)dx\big)$.

Karcher Mean PDF The mean PDF can be computed using the generalized version of the mean on a metric space called the Karcher mean [4]. Let $f_1, \ldots, f_n$ be a sample of PDFs and $h_1, \ldots, h_n$ be the corresponding square-root transformations. The sample Karcher mean $\bar{h}$ on $\mathcal{H}$ is defined as the minimizer of $\rho(h) = \sum_{i=1}^n d(h_i, h)^2$, that is, $\bar{h} = \text{argmin}_{h \in \mathcal{H}}\ \rho(h)$. Note that $\bar{h}$ is the mean corresponding to the transformations $h_1, \ldots, h_n$. However, the inverse mapping of the square-root transformations to the PDFs is unique [5], and we can define $\bar{f} = \bar{h}^2$ as the average PDF for the sample $f_1, \ldots, f_n$.

We present a gradient-based approach to compute the Karcher mean on $\mathcal{H}$ [2] in Algorithm S1. Here the inverse exponential map, denoted by $\exp_{h_1}^{-1} : \mathcal{H} \mapsto T_{h_1}(\mathcal{H})$, is given by $\exp_{h_1}^{-1}(h_2) = (\theta/\sin(\theta))(h_2 - h_1\cos(\theta))$. The exponential map at a point $h_1 \in \mathcal{H}$, denoted by $\exp : T_{h_1}(\mathcal{H}) \mapsto \mathcal{H}$, is defined as $\exp_{h_1}(\delta h) = \cos(\|\delta h\|)h_1 + \sin(\|\delta h\|)(\delta h/\|\delta h\|)$, where $\|\delta h\| = \big(\int_0^1 \delta h(x)^2 dx\big)^{1/2}$.

Principal Component Analysis. Below we present an algorithm to compute the principal coefficients using the PCA.

Algorithm S1 Sample Karcher mean of densities

1: $\bar{h}_0$ (initial estimate for the Karcher mean) $\leftarrow$ any one of the densities in the sample OR the extrinsic average. Set $j \leftarrow 0$ and $\epsilon_1, \epsilon_2 > 0$ be small.
2: For $i = 1, \ldots, n$ compute $u_i = \exp_{\bar{h}_j}^{-1}(h_i)$.
3: Compute the average direction in the tangent space $\bar{u} = \frac{1}{n}\sum_{i=1}^{n} u_i$.
4: **if** $||\bar{u}||_{L^2} < \epsilon_1$ **then**
5: **return** $\bar{h}_j$ as the Karcher mean.
6: **else**
7: $\bar{h}_{j+1} = \exp_{\bar{h}_j}(\epsilon_2 \bar{u})$.
8: Set $j \leftarrow j + 1$.
9: Return to step 2.

Algorithm S2 PCA on $T_{\bar{h}}(\mathcal{H})$

1: Compute the Karcher mean of $h_1, \ldots, h_n$ as $\bar{h}$.
2: **for** $i = 1, \ldots, n$ **do**
3: Compute projections of h_i onto $T_{\bar{h}}(\mathcal{H})$, that is, $v_i = \exp_{\bar{h}}^{-1}(h_i)$.
4: Evaluate sample covariance matrix $K = \frac{1}{n-1}\sum_{i=1}^{n} v_i v_i^\top \in \mathbb{R}^{m \times m}$.
5: Compute the SVD of $K = U\Sigma U^\top$.

In Algorithm S2, U is an orthogonal matrix of principal components or principal directions of variability, and Σ is a diagonal matrix of singular values. The first r columns of U (denoted as $\tilde{U} \in \mathbb{R}^{m \times r}$) span the r-dimensional principal subspace. The value of r could be chosen to account for a specified amount (e.g. 99.99%) of cumulative variance explained by the first few principal components. Hence, the PDFs can now be expressed using coordinates in this subspace via principal coefficients computed as $X = V\tilde{U}$, where $V^\top = [v_1\ v_2\ \ldots\ v_n] \in \mathbb{R}^{m \times n}$. These principal coefficients X act as Euclidean coordinates corresponding to PDFs and are used as predictors for downstream analysis.

S3 Permutation-based Hypothesis Tests for PDFs

Let f_i denote the PDF for subject i, and $u_i \in \{0, 1\}$ denote its group indicator for $i = 1, \ldots, n$. Let h_i denote the SRT for the PDF f_i for $i = 1, \ldots, n$. Given a sample of PDFs $f_1, \ldots, f_n$, we

have a distance metric and an approach to compute the average PDF. In this section, we build a permutation-based hypothesis test to investigate any differences between two groups of PDFs. That is, we want to investigate any differences in the average PDFs of the two groups (e.g. IDH mutated vs wild-type). To do this, we propose to use a permutation-based hypothesis test which computes the average PDFs of two groups and uses the distance between the two average PDFs as the test statistic. We define $d_0 = d(\bar{h}^0, \bar{h}^1)$, where $\bar{h}^0$ and $\bar{h}^1$ are the SRTs corresponding to the average PDFs for the two groups. This value of d_0 serves as our test statistic.

We create the null distribution for the test statistic by randomly permuting the group labels u_i between the subjects. Let $(u_{\sigma(1)}, \ldots, u_{\sigma(n)})$ denote a random permutation of the group indicators $u_1, \ldots, u_n$. Using the original PDFs $f_1, \ldots, f_n$ and the permuted group indicators $u_{\sigma(1)}, \ldots, u_{\sigma(n)}$, we compute the group average PDFs $\bar{f}^0_\sigma$ and $\bar{f}^1_\sigma$, and the distance between these average PDFs as d_σ. This process is repeated m times by considering the permutations $\sigma_1, \ldots, \sigma_m$. That is, for each permutation σ_j we obtain the distance between the group average PDFs as d_{σ_j}. Here the distribution generated by $d_{\sigma_1}, \ldots, d_{\sigma_m}$ serves as the null distribution for our test statistic d_0. The p-value for this permutation-based hypothesis test can be computed as $\sum_{k=1}^{m} I(d_0 > d_{\sigma_k})/m$, where $I(d_0 > d_{\sigma_k}) = 1$ if $d_0 > d_{\sigma_k}$, and 0 otherwise.

S4 Classification

Standard classification algorithms (e.g. logistic regression, probit regression) can be employed when the predictors belong to the Euclidean space. However, in our case the data object corresponding to each subject is a PDF (i.e. the residual signature). Hence, we use the following framework that maps each PDF to a vector of values. We employ principal component analysis using the PDFs to explore the variability through their primary modes of variation. The space of PDFs can be linearized by considering the tangent space at the sample average PDF, and projecting the sample PDFs onto this tangent space [7]. A principal component analysis on this tangent space provides us with a mapping of the PDFs to a vector in the Euclidean space (details in Section 3 of Online Supplemental Data).

Hence, the PDFs are mapped to vectors, that is, the principal coefficients through principal component analysis. These principal coefficients act as Euclidean coordinates corresponding to PDFs and are used as predictors in classification models. We construct a probit regression model

which is a generalized linear model used to model a binary categorical variable using numerical and/or categorical predictors. We model $p_i = P(y_i = 1|\mathbf{x}_i)$, the probability that a subject belongs to the category 1 based on a given set of predictors $\mathbf{x}_i$. Specifically, probit regression models $p_i = \Phi(\mathbf{x}_i^\top \boldsymbol{\beta})$, where $\mathbf{x}_i$ is the i^{th} row in the predictor matrix, $\boldsymbol{\beta} \in \mathbb{R}^r$ is the coefficient vector, and $\Phi(\cdot)$ is the cumulative distribution function of a standard normal distribution. We predict the probability of class membership for a new subject as $\hat{p}_{new} = \Phi(\mathbf{x}_{new}^\top \hat{\boldsymbol{\beta}})$, where $\hat{\boldsymbol{\beta}}$ are the estimated coefficients and $\mathbf{x}_{new}$ are the predictors for the new subject. In our context, the predictors $\mathbf{x}_{new}$ are the principal coefficients corresponding to the PDF f_{new} obtained by projecting onto the tangent space generated by the training data set. Note that other prediction models (e.g. logistic regression, random forests) can also be used with principal coefficients as predictors.

S5 Computation Time

The GWR model estimation can be executed in parallel across all the subjects once the T2 and FLAIR axial slices of interest are available. We run this step in parallel for four subjects at a time and the average time taken for this estimation was about 22.5 seconds per subject with a maximum of 95.3 seconds. In Figure S1 we plot the time taken for the GWR model estimation versus the number of tumor pixels for all the subjects considered for our analysis. The computations were performed on a standard computer with an Intel(R) Core(TM) i7-7700 CPU @ 3.60GHz, 3601 Mhz, 4 cores, 8 logical processors with 32 GB RAM. Additionally, the time taken for computation of p-values based on the permutation tests increases as the number of permutations increase.

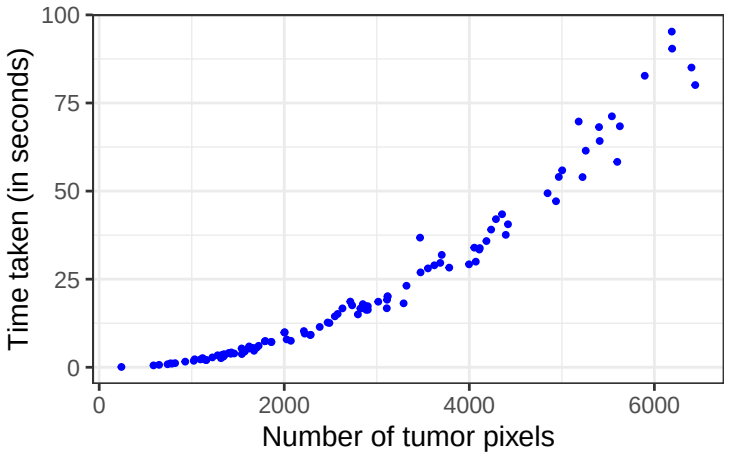
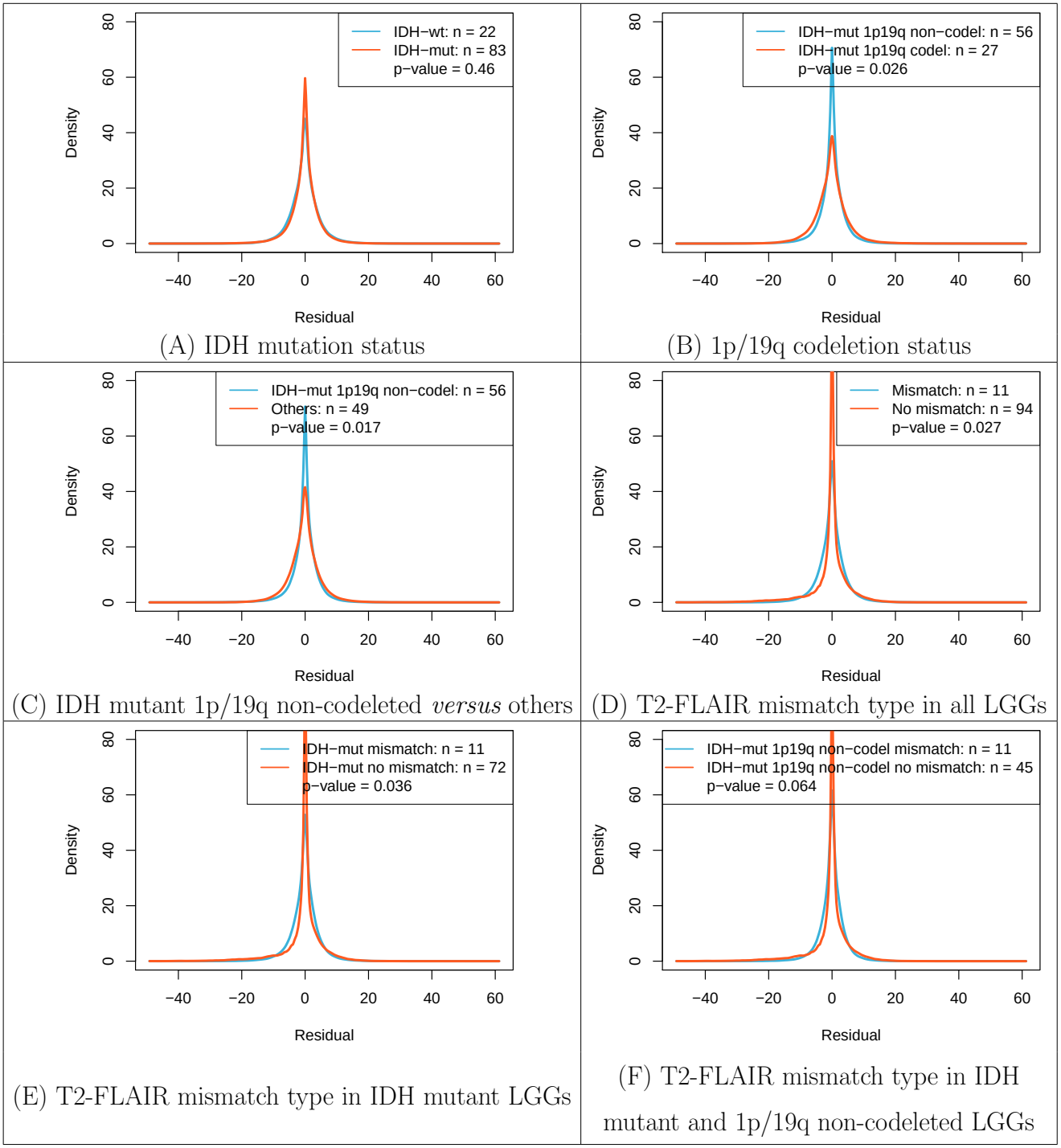


Figure S1: Time taken for the GWR model estimation versus tumor size for each subject.

S6 Supplementary Figure



(A) IDH mutation status

(B) 1p/19q codeletion status

(C) IDH mutant 1p/19q non-codeleted *versus* others

(D) T2-FLAIR mismatch type in all LGGs

(E) T2-FLAIR mismatch type in IDH mutant LGGs

(F) T2-FLAIR mismatch type in IDH mutant and 1p/19q non-codeleted LGGs

Figure S2: Average PDFs or averages of the residual signatures for each group across the six groupings (A)-(F).

References

[1] C. Brunsdon, A. S. Fotheringham, and M. E. Charlton. Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical Analysis*, 28(4):281–298, 1996.

[2] I. L. Dryden and K. V. Mardia. Statistical shape analysis, 1998.

[3] A. S. Fotheringham, C. Brunsdon, and M. Charlton. *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley & Sons, 2003.

[4] H. Karcher. Riemannian center of mass and mollifier smoothing. *Communications on Pure and Applied Mathematics*, 30(5):509–541, 1977.

[5] S. Kurtek and K. Bharath. Bayesian sensitivity analysis with the Fisher–Rao metric. *Biometrika*, 102(3):601–616, 2015.

[6] S. Lang. *Fundamentals of differential geometry*, volume 191. Springer Science & Business Media, 2012.

[7] J. Matuk, S. Mohammed, S. Kurtek, and K. Bharath. Biomedical applications of geometric functional data analysis. In *Handbook of Variational Methods for Nonlinear Geometric Data*, pages 675–701. Springer, 2020.

[8] D. P. McMillen. One hundred fifty years of land values in Chicago: A nonparametric approach. *Journal of Urban Economics*, 40(1):100–124, 1996.