2013 IEEE 10th International Symposium on Biomedical Imaging:
From Nano to Macro
San Francisco, CA, USA, April 7-11, 2013

SPATIOTEMPORAL MODELING OF DISTRIBUTION-VALUED DATA APPLIED TO DTI TRACT EVOLUTION IN INFANT NEURODEVELOPMENT

Anuja Sharma[1] *P. Thomas Fletcher*[1] *John H. Gilmore*[2] *Maria L. Escolar*[3]
Avantika Vardhan[1] *Aditya Gupta*[2,3] *Martin Styner*[2] *Guido Gerig*[1]

[1] School of Computing, SCI Institute, University of Utah, Salt Lake City, UT, USA
[2] Department of Psychiatry, University of North Carolina, Chapel Hill, NC, USA
[3] Department of Pediatrics, University of Pittsburgh, Pittsburgh, PA, USA

ABSTRACT

This paper proposes a novel method that extends spatiotemporal growth modeling to distribution-valued data. The method relaxes assumptions on the underlying noise models by considering the data to be represented by the complete probability distributions rather than a representative, single-valued summary statistics like the mean. When summarizing by the latter method, information on the underlying variability of data is lost early in the process and is not available at later stages of statistical analysis. The concept of 'distance' between distributions and an 'average' of distributions is employed. The framework quantifies growth trajectories for individuals and populations in terms of the complete data variability estimated along time and space. Concept is demonstrated in the context of our driving application which is modeling of age-related changes along white matter tracts in early neurodevelopment. Results are shown for a single subject with Krabbe's disease in comparison with a normative trend estimated from 15 healthy controls.

Index Terms— spatiotemporal growth trajectory, distribution-valued data, diffusion tensor imaging, early neurodevelopment, Mallow's distance

1. INTRODUCTION

The field of medical imaging has led to an improved understanding of dynamic anatomical and physiological changes. Characterizing normative growth patterns together with the natural population variability is one of the key challenges it addresses. This knowledge is crucial for clinical researchers as it enables a timely identification of deviations from expected trends enabling early diagnosis and therapy for altered trajectories. In this regard, extensive work has been done in the area of human brain development. Morphometric brain changes along time and white matter maturation have been studied using cumulative volume measurements, scalar functional data and 3D image data [1, 2, 3, 4]. However, most of these methods work with classic single-valued observations at each point in space and time for an individual subject. These methods work under the assumption that the probability distribution generating the data is well represented using scalar central tendencies like averages and quantiles. Note that this is different from encoding the population variability in an atlas model (for instance, capturing the complete data statistics in terms of a shape atlas together with individual subject variability evolving along time [1, 2]).

Data variability comes from the assumption that our data are observed random samples of an underlying probability distribution. For example, due to low image resolution or mis-registration of images, information within a voxel may not necessarily be Gaussian distributed or even unimodal. Similarly, summarizing scalar diffusion values from brain's white matter as averages within regions imposes a homogeneous and known parametric probability distribution within each region. Less samples could also result in violations of the central limit theorem's applicability. Ashburner et al. [5] explore the effect of non-Gaussian data on the accuracy of parametric statistical tests used for voxel based morphometry. HARDI imaging also highlights the presence of complex intra voxel distributions [6].

Working with data summaries and disregarding the underlying variability offers advantages in terms of simplifying the analysis. This works well if a parametric probability distribution which is consistent across all samples is either known or can be safely assumed. Otherwise, this approach poses severe limitations and can introduce errors in the framework at very early stages. This is the key issue we address in this paper. See Fig 1 for an illustration. We have two data distributions observed along time (blue:t_0, orange:t_1). A simple linear interpolation between these via their respective means (red) and medians(blue) is also shown. Since the distributions are skewed in opposite directions, the fit with mean and median are significantly different. Moreover, we have lost the data variability information at all the intermediate interpolated times.

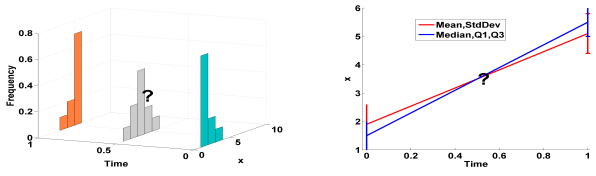

Fig. 1. Left:Two synthetically generated distributions along time (blue:t_0, orange:t_1) with translated locations and opposite skews. Right: Simple linear interpolation using classic single-valued summary statistics (means: red, medians: blue) and ignoring data variability (shown as error bars) at t_0 and t_1.

As a solution, we propose a novel nonparametric framework of spatiotemporal analysis which considers all observed data as probability distributions rather than a single scalar value. We estimate a continuous temporal growth trajectory of distribution-valued data from a discrete-time set of observations of the same. The frame-

Supported by NIH grants National Alliance for Medical Image Computing (NA-MIC) U54 EB005149, R01 HD055741, R01 MH070890, Twin R01 MH070890, Conte Center MH064065, BRP R01 NS055754, NINDS R01NS61965, DANA Foundation, NIH P30 HD03110 (UNC IDDRC), USTAR initiative at the University of Utah.

work is applicable to individual subjects as well as for delineating normative population trends. This is important because it allows designing and administering of individualized medical treatments via systematic comparisons of patient specific patterns with normative trajectories. Our driving application is the characterization of DTI- derived diffusivity changes along white matter tracts in infant neurodevelopment. Tract based changes provide an insight into the brain's maturation process and correlate with characteristic cognitive functions. This makes them better suited for assessment related to specific cognitive abnormalities [4]. Section 2 details the proposed method, section 3 includes validation using synthetic data and results of application to pediatric DTI data.

2. METHOD

2.1. Framework

As a general representation, we have spatially parametrized 3D data collected at discrete time points for a number of subjects. Let s_i be a spatial variable ($i = [0, 1, \cdots N_S]$), t_j a time variable ($j = [0, 1, \cdots N_T]$) and q_k an individual subject ($k = [0, 1, \cdots N_Q]$). N_S, N_T and N_Q are the discrete number of parametrized spatial locations, serial time points per subject and the total number of subjects respectively. Data indexed by each (s_i, t_j, q_k) vector is a probability distribution. In this paper, we follow [7, 8] and choose to represent these distributions as histograms. The reasoning lies in the fact that histograms are intuitive, computationally easy to handle and summarize the variability in the data empirically without requiring prior information of the underlying distribution. Following the notations by [7], we define a histogram random variable X such that our data is an observed sample X^{ijk} of this variable at spatial location s_i, at time t_j for subject q_k. The histogram is formally defined as

$$h^l_{X_{ijk}}(x) = (\pi^l, c^l, r^l), \tag{1}$$

where $l = [1 \cdots L]$ is the histogram bin index with L as the total number of bins and x is the variable which the histogram represents (for e.g. DTI diffusion property like FA). Each bin l is characterized by the bin frequency π^l, location of the bin's center c^l and half the bin width as bin radius r^l. r^l can vary across l allowing unequal bin widths. Histograms are normalized such that $\pi^l \geq 0$ and $\sum_{l=1}^{L} \pi^l = 1$. The corresponding cumulative distribution function (CDF) is defined as $H^l_{X_{ijk}}(x)$ and the inverse CDF or quantile function as $[H^z_{X_{ijk}}]^{-1}(u), u \in [0, 1]$ and z as the bin index for quantile functions. Assuming $h^l_{X_{ijk}}(x)$ represents the underlying probability density function within an acceptable margin of error, [7] provide a formulation for an empirical estimation of the corresponding CDF and quantile functions from it. For simplicity, we will refer to the histogram as $h_{X_{ijk}}$ from here on.

In the context of our driving application, scalar diffusion information from 3D white matter tracts is available at discrete times via Diffusion Tensor Imaging from serial scans of multiple subjects. We utilize the atlas based registration and arc length parametrization framework by [3] to achieve spatial normalization across time points and subjects (Fig 2). We extend the method by characterizing cross-sections of fiber tract bundles by distributions of diffusion values such as FA, and by representing this data as histogram variables along space and time attributed with 4D image properties. This procedure is motivated by a major limitation of the current fiber tract-based analysis which reduces local tract properties to mean values used for group statistics, thus discarding information on data variability which is important for statistical testing and inference. Moreover, taking the mean assumes normal distribution and unimodality, which is not a proper model for FA and tract locations showing mixtures of fiber bundles.

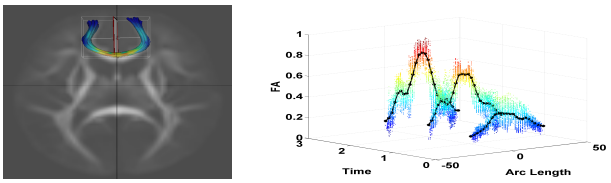

Fig. 2. Left: 3D visualization of the genu white matter tract. Right: Diffusion values along tract bundles (colored data) and cross sectional means (black) of diffusion property (FA) at discrete time points along genu tract for a single subject (based on [3]).

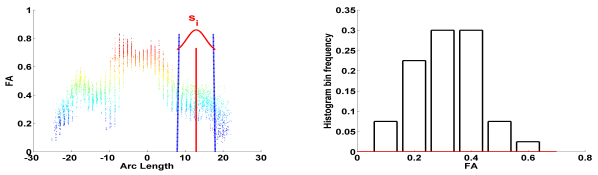

Fig. 3. Aggregating spatial information. Left: Aggregation of FA values by application of kernel weights within a kernel window to compute a histogram. Right: Computed FA histogram at location s_i.

2.2. Methodology

To obtain histogram descriptions from diffusion values along tract bundles, we use a kernel based weighting function within a moving kernel window along s (Fig 3). The weights account for the inherent functional correlation along s owing to the underlying brain anatomy and allow the distributions to represent any spatially continuous information in the absence of known priors and constraints. For an arc length parametrized tract location s_i at time t_j for subject q_k,

$$\pi^l_{X_{ijk}} = \acute{\pi}^l_{X_{ijk}} K_\sigma(s, s_i), \quad K_\sigma(s, s_i) \propto e^{\frac{-(s-s_i)^2}{2\sigma^2}}.$$

$K_\sigma(s, s_i)$ is a Gaussian kernel with standard deviation σ (which governs the amount of spatial smoothing achieved) and $\acute{\pi}^l_{X_{ijk}}$ are the unweighted histogram bin frequencies within the kernel window centered at s_i. The weights within each kernel window $K_\sigma(s, s_i)$ are normalized to 1.

We now apply a weighted average scheme based on Shepard's inverse distance weighting function applied along time where

$$w_{jk}(t) = \frac{1}{(abs(t - t_j) + \epsilon)^\alpha}, \tag{2}$$

and $\epsilon < 0.00001$ to avoid division by zero. Considering a single subject q_k, $w_{jk}(t)$ assigns weights to each time point t_j of q_k for all locations s_i. The power parameter α decides the relative influence of interpolating points located far-off with weights decreasing with increasing distance. We use $\alpha = 2$ to allow closer time points to have the most dominating effect. When considering multiple subjects for assessment of an average population trend, each subject q_k gets equal weights (assuming a homogeneous population) while each time point of q_k follows eq. (2) with total weights per subject normalized to 1. Finally, total weights across N_Q subjects are also normalized to 1. Fig 4 shows the weight allocation for 15 subjects

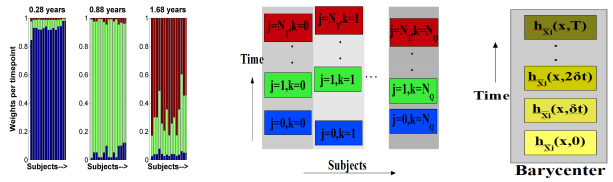

Fig. 4. Left: Weights applied to three time points (t_j approximately 1 month (blue), 1 year (green) and 2 years (maroon)) for 15 subjects. Increasing time t (0.28, 0.88, 1.68 years) shows changing relative weights with more weights given to closer time points. Center: For location s_i, N_T discrete histogram samples from each of N_Q subjects are used to find a barycentric histogram at t using weights as shown to the left (see eq. (2)). Right: Barycentric histogram $h_{\bar{\mathbf{X}}_i}(x,t)$ evaluated continuously along time at s_i.

scanned approximately at 1 month, 1 year and 2 years, calculated and displayed at $t = 0.28, 0.88, 1.68$ years. Note that t_j is not required to be the same across subjects. Using eq. (2), we compute a weighted 'average' histogram response (called a barycentric histogram $h_{\bar{\mathbf{X}}_i}(x,t)$) from $h_{X_{ijk}}$ continuously along t for each spatial location s_i ($t = [0, \delta t, 2\delta_t \cdots T]$). For a single subject this becomes an interpolation problem while for multiple subjects, the 'average' is computed in a least square sense. Fig 4 visually explains this estimation approach.

This brings us to the notion of 'distance' between two distributions. For this purpose, we use the Mallow's distance D_M^2 as our distance metric (eq. (3)). [8] decompose D_M^2 in components reflecting translation, changing width and shape of the distributions being considered. Moreover, for distributions with the same mass (e.g. normalized probability density functions), D_M^2 is conceptually the same as Earth Mover's distance which quantifies the dissimilarity between two piles of earth as the amount of work needed to transform one into another [9]. Unlike divergence based measures like KL and Hellinger, Mallow's distance is a positive L2 norm and follows desirable properties like symmetry, triangle inequality and allows proportional contribution via weights. These reasons make it an optimal choice for barycenter estimation given that a barycenter histogram minimizes the distance between itself and all other histograms [7, 8]. Mallow's distance $D_M^2(h^l_{X_{ijk}}(x), h^{l\prime}_{X_{i\prime j\prime k\prime}}(x))$ is expressed as L2 norm of the difference between u-quantiles of two distributions with $u \in [0,1]$,

$$\sqrt{\int_0^1 ([H^z_{X_{ijk}}]^{-1}(u) - [H^{z\prime}_{X_{i\prime j\prime k\prime}}]^{-1}(u))^2 du}. \quad (3)$$

For computing it, the probability range $u \in [0,1]$ is divided into $z_M = [1 \cdots Z_M]$ equi-probable intervals by combining the set of cumulative frequencies for $H^l_{X_{ijk}}(x)$ and $H^{l\prime}_{X_{i\prime j\prime k\prime}}(x)$. The quantiles corresponding to u's intervals are computed for each distribution. Assuming a uniform distribution within individual bins, each of these quantiles can be represented using their respective interval centers c^{z_M}, radii r^{z_M} and frequencies $\pi^{z_M} = u^{z_M} - u^{z_M-1}$. To summarize, all normalized density functions are mapped to a common reference frame of $u \in [0,1]$ using the CDF's and then compared with respect to the bin location and width for the corresponding quantiles. Mallow's distance $D_M^2(h^l_{X_{ijk}}(x), h^{l\prime}_{X_{i\prime j\prime k\prime}}(x))$ can then be expressed in terms of this new representation as,

$$\Sigma_{z_M=1}^{Z_M} \ \pi^{z_M}[(c^{z_M}_{X_{ijk}} - c^{z_M}_{X_{i\prime j\prime k\prime}})^2 + \frac{1}{3}(r^{z_M}_{X_{ijk}} - r^{z_M}_{X_{i\prime j\prime k\prime}})^2]. \quad (4)$$

Based on eq (2) and (4), a barycentric histogram $h_{\bar{\mathbf{X}}_i}(x,t)$ at s_i is defined as a weighted mean which minimizes

$$\Sigma_{j=0}^{N_T}\Sigma_{k=0}^{N_Q} w_{jk}(t) D_M^2(h_{\bar{\mathbf{X}}_i}(x,t), h_{X_{ijk}}(x)). \quad (5)$$

Solving this minimization problem gives $\bar{z} = [1 \cdots \bar{Z}]$ barycentric histogram bins centered at the average of $c^{\bar{z}}_{X_{ijk}}$ with bin radii as an average of $r^{\bar{z}}_{X_{ijk}}$ and bin frequencies $\pi^{\bar{z}}$. For more details, refer to [7]. Solving this for all spatial locations s_i for increasing value of $t = [0, \delta t, 2\delta_t \cdots T]$ gives a continuous time evolution of distributions in space.

3. EXPERIMENTS

3.1. Validation using synthetic data

We go back to our example from Fig 1 for validation of the proposed method. This scenario is the same as having a single subject and a single spatial location ($i = k = 0$) with distribution data observed at two discrete time points (t_0, t_1). Results are shown in Fig 5 (top row, left to right). The first plot shows the intermediate estimated distributions represented as histograms between t_0 (blue) and t_1 (orange). The weights are calculated at each t using eq.(2). Barycentric histograms $h_{\bar{\mathbf{X}}_0}(x,t)$ at $t = [0.2, 0.4, 0.6, 0.8]$ are displayed. The location, width and shape of $h_{X_{000}}$ gradually transforms over t to match $h_{X_{010}}$ at t_1. The skew slowly shifts from right to left adjusting the bin width and frequency.The CDF plots for each corresponding barycenter highlight the 'averaging' along time. This also illustrates that the barycenter computed using D_M^2 is a convex solution of the minimization problem. The third plot shows the empirically calculated mean (red) and median (blue) values using the barycenters along time. Comparing the standard deviation (red error bars) and the interquartile range (blue error bars), the skew along time becomes obvious.

3.2. Application to pediatric data

For concept demonstration, we apply our method to pediatric DTI data (FA values for the genu tract, see Fig 2) from 15 healthy infants registered to a common atlas space. Subjects are scanned around 1 month, 1 year and 2 years with scan times not in perfect correspondence. An average normative growth trajectory is estimated using all control subjects ($N_Q = 15$), time points ($N_T = 3$) and tract's locations ($N_S = 25$) to represent population trend and summarize FA's variability (see Fig. 5 (bottom row, left to right)).

The growth surface in the first plot exhibits localized temporal trends along the length of the genu tract. Overall, FA increases along time owing to the expected brain maturation in early neurodevelopment. While these agree with clinically observed patterns, interesting results show up when we zoom in on the trajectory of a single location (marked by red). The second plot draws our attention to the complete distribution associated with each point on the trajectory at this location. The quantiles immediately make it clear that the FA values do not follow a Gaussian distribution along time. The interquartile region (25th to 75th percentile) is not symmetrically placed in the center of 0th and 100th percentiles with the skew more prominent after 1 year. Without the complete variability information captured here, a simple average based growth modeling would be inaccurate. With such mixed patterns, methods to utilize distribution valued variables hold the key to more robust and reliable modeling procedures. For further illustrating the potential of this approach, we similarly estimate personalized FA evolution profile for the genu

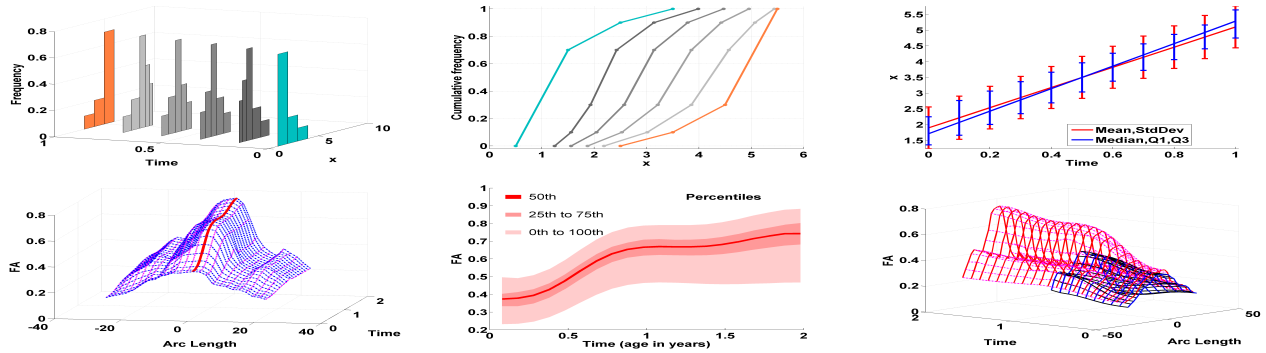

Fig. 5. TOP ROW: Evolution of synthetic distribution data continuously along time. Left: Barycenter histograms evaluated along time show translation and changing shape (increasingly lighter grayscale color from t_0(blue) to t_1(orange)). Center: Corresponding CDF plots show an 'averaging' effect between t_0 and t_1. Right: Means, standard deviations (red) and quantiles Q50, Q25 and Q75 (blue) estimated from barycentric histograms evaluated along t. BOTTOM ROW: Application to pediatric data (15 healthy controls and a single subject with Krabbe's disease) Left: Normative growth surface represented via median values estimated from barycentric histograms along s, t. Center: Quantiles (Q0, Q25, Q50, Q75, Q100) estimated for a single spatial location marked in red (left plot). Right: Comparison of spatiotemporal trajectory (median values shown) for a single subject with Krabbe's disease (blue) and a normative healthy trend (red).

tract from a subject with Krabbe's disease with only two serial scans mapped into the same atlas space. The disease is degenerative in nature affecting the myelin of the nervous system. It is often fatal without a very early diagnosis and medical intervention in infants. The third plot compares the typical growth surface obtained using controls with that of the subject with Krabbe's disease where FA values of the latter lag behind.

3.3. Discussion

The above examples provide a proof of concept for the ability of the method to model the complete data variability along time, with measurements given at only a few discrete time points. Our approach offers a two-fold advantage. First, without imposing any parametric assumptions on the underlying noise model, we estimate spatiotemporal trends utilizing the complete distribution-valued information. The Mallow's distance metric can be utilized for further data exploration like goodness-of-fit evaluation via RMSE and Pseudo R^2 and classification using clustering algorithms customized for distribution valued data [8]. Second, even if we decide to summarize data further to scalar numbers, our method facilitates a more informed decision by retaining the data variability till the last stages of statistical analysis. This increases accuracy of estimation methods and provides increased flexibility via relaxation of assumptions on noise models. Our method uses a weighted average smoothing and therefore benefits from a dense sampling of data along time. There are no constraints in terms of scan time correspondence across subjects.

4. CONCLUSION

We present a new framework for estimation of smooth temporal trends for spatially correlated distribution-valued data. Spatial aggregation employs a nonparametric kernel weighting function. A weighted average of distributions is evaluated continuously along time where weights are based on inverse distance weighting. Mallow's distance which is conceptually the same as Earth Mover's distance for equal mass distributions, is applied as the distance metric. The results illustrate the method's feasibility and its potential application to large clinical studies for comparing personalized profiles to normative data. Future work will include advanced statistical analysis schemes to utilize the complex variability information encoded in growth trajectories characterized by probability distributions. We would also like to evaluate improvements in diagnostic accuracy when compared with conventional methods based on scalar summary statistics.

5. REFERENCES

[1] K. Pohl, J. Fisher, W. Wells, et al., "Logarithm odds maps for shape representation," *MICCAI 2006*, pp. 955–963, 2006.

[2] M. Kuklisova-Murgasova, P. Aljabar, A.D. Edwards, et al., "A dynamic 4d probabilistic atlas of the developing brain," *NeuroImage*, vol. 54, no. 4, pp. 2750–2763, 2011.

[3] C.B. Goodlett, P.T. Fletcher, G. Gerig, et al., "Group analysis of dti fiber tract statistics with application to neurodevelopment," *Neuroimage*, vol. 45, no. 1, pp. S133–S142, 2009.

[4] J. Dubois, L. Hertz-Pannier, D. Le Bihan, et al., "Assessment of the early organization and maturation of infants' cerebral white matter fiber bundles: a feasibility study using quantitative diffusion tensor imaging and tractography," *Neuroimage*, vol. 30, no. 4, pp. 1121–1132, 2006.

[5] J. Ashburner and K.J. Friston, "Voxel-based morphometry-the methods," *Neuroimage*, vol. 11, no. 6, pp. 805–821, 2000.

[6] D.S. Tuch, V.J. Wedeen, et al., "High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity," *Magnetic Resonance in Medicine*, vol. 48, no. 4, pp. 577–582, 2002.

[7] J. Arroyo and C. Mate, "Forecasting histogram time series with k-nearest neighbours methods," *International Journal of Forecasting*, vol. 25, no. 1, pp. 192–207, 2009.

[8] R. Verde and A. Irpino, "Ordinary least squares for histogram data based on wasserstein distance," *COMPSTAT'2010 Book of Abstracts*, p. 381, 2008.

[9] E. Levina and P. Bickel, "The earth mover's distance is the mallows distance: some insights from statistics," in *ICCV 2001*. IEEE, 2001, vol. 2, pp. 251–256.