



Filtered Trajectory Recovery: A Continuous Extension to Event-Based Model for Alzheimer's Disease Progression Modeling

Jiangchuan Du¹, Yuan Zhou²(✉),
and for the Alzheimer's Disease Neuroimaging Initiative

¹ School of Mathematical Science, Fudan University, Shanghai, China

² School of Data Science, Fudan University, Shanghai, China
yuanzhou@fudan.edu.cn

Abstract. Event-based model (EBM) is a flexible way to model the progression of Alzheimer's disease. The core of EBM is to use an ordering of events to build the trajectory of disease progression. The ordering of events is usually inferred from the data using optimization, hence may suffer from local optima and computational complexity. This paper mathematically proves that this ordering can be directly determined by the cumulative distribution function (CDF) of the data. From this standpoint, we formulate two properties—order preserving property (OPP) and distance preserving property (DPP)—that an estimated trajectory should satisfy. We show that a trajectory that satisfies these two properties is equivalent to a reparametrized version of the true trajectory. Furthermore, we show that one such reparametrized trajectory can be directly obtained from the CDF of the data by filtering. We call the algorithm filtered trajectory recovery (FTR). Extensive experiments on simulated data and real data from ADNI show that FTR can retrieve a trajectory that provides a better estimation of the event order and stages, disentangle the progression heterogeneity, without assuming a parametric form of the trajectory function. The code is released at <https://github.com/Jiangchuan-Du/FTR-master>.

Keywords: Alzheimer's Disease · Disease Progression Model · Event-based Model

1 Introduction

Alzheimer's disease is characterized by progressive loss of the neurons in the brain. The disease usually starts years before the symptoms emerge and progresses heterogeneously [1]. One way to quantify this *heterogeneity* is to group the population into subtypes: each individual belongs to a subtype and each

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-34048-2_8.

subtype corresponds to a prototypical progression *trajectory*. The progression trajectory can be measured by various biomarkers from neuroimaging scans, clinical scores, cerebrospinal fluid (CSF), etc. *Disease progression models* (DPM) aim to recover the progression trajectory from a collection of short individual time series of these biomarkers. Applying DPMs to discover these trajectories is quite significant in providing biological insight into the underlying disease mechanism. It also enables staging and subtyping the patients, hence benefits patient stratification and facilitates precision medicine in clinical trials.

A general DPM assumes that the data points follow

$$s \sim h(s), \quad \mathbf{x} \sim \mathcal{N}(\mathbf{f}(s), \Sigma(s))$$

where s is the *stage* (time from the beginning of disease progression), $h : [0, 1] \rightarrow \mathbb{R}_+$ is the density function for sampling s , whose domain is the timeline normalized to $[0, 1]$. Given sampled s , an observed data point $\mathbf{x} = [x_1, \dots, x_B]$ is sampled from a Gaussian distribution with mean $\mathbf{f}(s)$ and covariance matrix $\Sigma(s)$, where B is the number of biomarkers, $\mathbf{f} : [0, 1] \rightarrow \mathbb{R}^B : s \mapsto (f_1(s), \dots, f_B(s))$ is the entire trajectory of disease progression, $\Sigma(s)$ is the noise covariance matrix at s . Without loss of generality, we assume that the trajectory of each biomarker f_j is a *monotonically increasing* function. Given a set of observed data points, a DPM tries to estimate the underlying trajectory \mathbf{f} and possibly noise Σ .

Previous studies that try to solve the problem can be categorized into event-based models (EBM) [2–8], parametric/non-parametric models [9–11], differential equation models [12, 13], and deep learning [14, 15]. The EBMs model the trajectory as a transition from normal to abnormal using a set of predefined events [2–5, 8]. This event-based strategy is extended to model a more flexible piecewise linear function as well as multiple trajectories for quantifying the progression heterogeneity [6, 7]. Parametric/non-parametric models characterize the continuous longitudinal progress of biomarkers by a parametric (e.g. sigmoid [9]) or a non-parametric (e.g. Gaussian process [10, 11]) function of time and other variates. By assuming that toxic proteins spread along the neural network in the brain, differential equations have been introduced to model the progression of brain biomarkers [12, 13]. The last one, deep learning techniques, in particular recurrent neural networks (RNN), can also predict time series progress in dealing with longitudinal data [14, 15].

Despite a plethora of ways to model the trajectory, they more or less made some assumptions on the form of the trajectory function. We observe that a

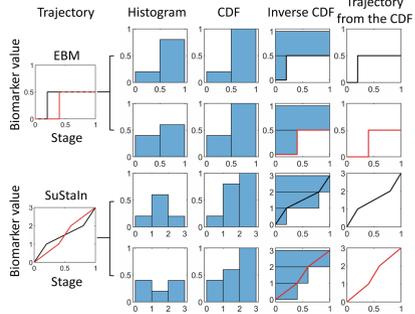


Fig. 1. Motivation of directly estimating the trajectory using the data distribution. The first column shows the trajectory modeled by the original EBM and SuStain (see details in Sect. 2). The second, third, and fourth columns show the histograms, CDF, and inverse CDF of the biomarkers, respectively, when h is a uniform distribution.

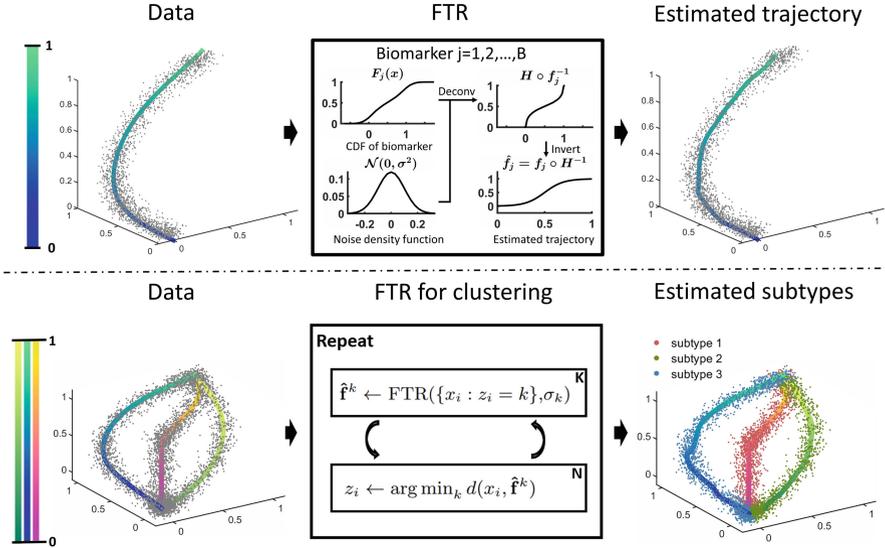


Fig. 2. Overview of the methodology. In FTR (first row), given data (gray dots) generated from a trajectory (time indicated by the colorbar on the left), the CDF of each biomarker $F_j(x)$ is deconvolved by the noise density function $\mathcal{N}(0, \sigma^2)$ to get $H \circ f_j^{-1}$, which is inverted to obtain a reparametrized trajectory $\hat{f}_j = f_j \circ H^{-1}$. Combining all such \hat{f}_j 's gives our estimated trajectory. In the second row, FTR is extended to cluster data points from 3 trajectories via a kmeans-like algorithm. (Color figure online)

trajectory is closely related to the *cumulative distribution function* (CDF) of the data in the EBM. For example, Fig. 1 shows two ways in a EBM to model the trajectory of two biomarkers (first column). Under a uniform distribution assumption on h , we see that after calculations of the histogram and CDF, the inverse CDF exactly reflects the true trajectory. This motivates a way to retrieve the trajectory directly from the data distribution.

In this paper, we developed a mathematical theory on using the data distribution for trajectory recovery. First, we show that a closed-form solution to EBM is available by using the CDF of the data (Sect. 2). Then, we extend the trajectory to an arbitrary continuous function by generalizing the idea of EBMs to an *order-preserving property* (OPP) and a *distance-preserving property* (DPP) (Sect. 3). We show that a trajectory satisfying these two properties is equivalent to a reparametrized trajectory (Proposition 2). Finally, we show that a particular reparametrized trajectory can be retrieved by deconvolving the CDF of the observed data (Theorem 3). We call this process *filtered trajectory recovery* (FTR) (see Fig. 2). Extensive experiments on both simulated datasets and a real dataset are conducted to validate the proposed FTR (Sect. 4).

2 Event-Based Model

The EBMs use an order of the events to define the entire trajectory. An *event* is the stage s when a biomarker reaches a specific target e for the first time, i.e.,

$s = \inf\{t : f_j(t) \geq e\}$. An *order* of a sequence of B events (s_1, \dots, s_B) is a bijection $o : \{1, \dots, B\} \rightarrow \{1, \dots, B\}$ that rearranges the sequence to $(s_{o_1}, \dots, s_{o_B})$ where o_i (denoting $o(i)$) is the event index at position i in the rearranged sequence. A *rank* $r : \{1, \dots, B\} \rightarrow \{1, \dots, B\}$ is the inverse function of o . Hence, r_i is the position of the i th event in the rearranged sequence.

2.1 Representative EBMs

In the original EBM, there is only 1 target e for each biomarker, leading to B events $\{s_j : j = 1, \dots, B\}$ in total [2]. The trajectory of the j th biomarker is assumed to be a step function with two outputs—normal, abnormal—with the event s_j marking the time of change from normal to abnormal. Given an order of these events, the trajectory is approximated by the corresponding ranks:

$$f_j(s) = \begin{cases} \mu_j^0 & \text{if } s < r_j \\ \mu_j^1 & \text{if } s \geq r_j \end{cases}$$

where r_j is the rank of s_j in the order, μ_j^0 is the expected value when the j th biomarker takes a normal value and μ_j^1 represents the expected abnormal value. Note that the timeline is not normalized to 0–1 in this notation. Based on the trajectory definition, a probabilistic model is introduced to estimate the order of biomarker change given a set of data points.

Considering that a step function for approximating the trajectory is far away from satisfactory, an improvement—SuStaIn—is later proposed to approximate the trajectory by piecewise linear functions [6]. In SuStaIn, multiple event targets are defined $e_1 < e_2 < \dots < e_M$ and they have corresponding events $\{s_{jl} : j = 1, \dots, B, l = 1, \dots, M\}$, i.e. $s_{jl} = \inf\{t : f_j(t) \geq e_l\}$. By fixing s_{jM} to be the endpoint on the timeline, SuStaIn uses an order of $S = (s_{11}, \dots, s_{1,M-1}, s_{21}, \dots, s_{B,M-1})$ to define the trajectory:

$$f_j(s) = \begin{cases} \frac{e_1 - e_0}{r_{j1} - r_{j0}}(s - r_{j0}) + e_0, & \text{if } r_{j0} \leq s < r_{j1} \\ \vdots \\ \frac{e_M - e_{M-1}}{r_{jM} - r_{j,M-1}}(s - r_{j,M-1}) + e_{M-1}, & \text{if } r_{j,M-1} \leq s \leq r_{jM} \end{cases}$$

where r_{jl} is the rank of s_{jl} in the order for $l \in \{1, \dots, M - 1\}$, $r_{j0} = 0$ and $r_{jM} = B(M - 1) + 1$ are the start and end points on the timeline. Similar to the original version, a probabilistic model is introduced to estimate the order of S .

2.2 Closed-Form Solution

Reviewing the above models, we see that to find the order that fits the data, we need to do optimization in a permutation space. When the dimension of the space is large, both the complexity and the robustness of finding the correct order may not be ideal. Following the intuition in Fig. 1, we show in Theorem 1 that a closed-form solution to EBMs is available under mild conditions.

Theorem 1. *Suppose the data point random variable $\mathbf{x} = [x_1, \dots, x_B]$ is generated by*

$$s \sim h(s), \quad x_j \sim \mathcal{N}(f_j(s), \sigma^2),$$

where $h : [0, 1] \rightarrow \mathbb{R}_+$ is a density function and

$$f_j(s) = \begin{cases} 0, & \text{if } 0 \leq s < a_j, \\ 1, & \text{if } a_j \leq s \leq 1. \end{cases}$$

Then, for any set of change points $\{a_j \in [0, 1]\}$, σ^2 and h , the sequence $(\hat{a}_1, \dots, \hat{a}_B)$ defined by $\hat{a}_j = F_j(\frac{1}{2})$ has the same order as (a_1, \dots, a_B) , where F_j is the CDF of the marginal distribution $p(x_j)$.

Sketch of Proof. In $F_j(x) = \int_{-\infty}^x \int_0^1 p(z; f_j(s), \sigma^2) h(s) ds dz$, the integration over s from 0 to 1 can be separated into two ranges: $[0, a_j]$ and $[a_j, 1]$. Using the definition of f_j and this separation, we can show $F_j(x) = H(a_j)\phi(x; 0, \sigma^2) + [1 - H(a_j)]\phi(x; 1, \sigma^2)$, where $\phi(x; 0, \sigma^2)$ is the CDF of $\mathcal{N}(x; 0, \sigma^2)$ and H is the CDF of h . Then we can show $F_j(\frac{1}{2}) - F_i(\frac{1}{2}) > 0$ for any $a_j > a_i$.

Theorem 1 can be extended to monotonically increasing step functions with multiple outputs, e.g. $0, 1, 2, \dots, M - 1$. A more interesting question is: Can we take the limit to make the output interval arbitrary small such that a continuous function can be estimated? In the next section, we present a way to achieve this.

3 Filtered Trajectory Recovery

Here, we assume that each f_j is strictly monotonically increasing such that the event for a biomarker reaching e can be written as $f_j^{-1}(e)$. Recalling the EBMs, we formulate two properties that an acceptable trajectory should satisfy to achieve our goals: finding the correct order of events and clustering the data points.

3.1 Theory

First, we generalize the concept of retrieving the correct order of the events to an *order-preserving property*. This property ensures that for any events, an estimated trajectory has the same order as the true one that generates the data. In addition, to make the estimated one be able to cluster the data points, we propose a *distance-preserving property*. For an estimated trajectory denoted by $\hat{\mathbf{f}} : [0, 1] \rightarrow \mathbb{R}^B : \hat{\mathbf{f}}(s) = (\hat{f}_1(s), \dots, \hat{f}_B(s))$ and the true one by $\mathbf{f} : [0, 1] \rightarrow \mathbb{R}^B : \mathbf{f}(s) = (f_1(s), \dots, f_B(s))$, these two properties are formalized as:

1. *The order-preserving property (OPP):* For any j, k, e, e' , if $f_j^{-1}(e) < f_k^{-1}(e')$, then $\hat{f}_j^{-1}(e) < \hat{f}_k^{-1}(e')$.
2. *The distance-preserving property (DPP):* For any $\mathbf{x} \in \mathbb{R}^B$, $d(\mathbf{x}, \mathbf{f}) = d(\mathbf{x}, \hat{\mathbf{f}})$, where $d(\mathbf{x}, \mathbf{f}) = \min_s \|\mathbf{x} - \mathbf{f}(s)\|$.

The order-preserving property implies that for any set of events $\{f_j^{-1}(e_l)\}$, its order is the same as that from $\{\hat{f}_j^{-1}(e_l)\}$ hence $\hat{\mathbf{f}}$ suffices to be used to retrieve the correct order of biomarker change. The distance-preserving property implies that the range of $\hat{\mathbf{f}}$ is the same as the range of \mathbf{f} , i.e. the trajectory lies on the same 1-dimensional manifold embedded in the B -dimensional space. This property enables us to use a clustering algorithm to cluster the data points. These two properties together define an admissible trajectory in estimation.

Then we ask: What kind of trajectories would satisfy these two properties? The proposition below suggests that if we reparametrized $\mathbf{f}(s)$ to $\mathbf{f} \circ H^{-1}(t)$ by using a monotonically increasing bijection H , this *reparametrized trajectory* $\mathbf{f} \circ H^{-1}$ would satisfy these two properties.

Proposition 2. *A trajectory $\hat{\mathbf{f}}$ satisfies the order-preserving property and the distance-preserving property if and only if $\hat{\mathbf{f}} = \mathbf{f} \circ H^{-1}$ where $H : [0, 1] \rightarrow [0, 1]$ is a continuous bijection that increases monotonically.*

Sketch of Proof. The sufficiency is straightforward since $f_j^{-1}(e) < f_k^{-1}(e')$ implies $H \circ f_j^{-1}(e) < H \circ f_k^{-1}(e')$. For the necessity, we can define $H_j = \hat{f}_j^{-1} \circ f_j$ for each j and show $H_j = H_k$ by contradiction (assuming $H_j(s_0) > H_k(s_0)$ for some s_0 , choose s and s' around s_0 such that $s < s'$, show that $\hat{f}_j^{-1}(e) > \hat{f}_k^{-1}(e')$ where $e = f_j(s)$ and $e' = f_k(s')$).

Now we have identified the space of trajectories that we are interested in. The next natural question is: Can we estimate such a reparametrized trajectory from the data? Note that all H 's that are continuous, bijective, and monotonically increasing would suffice for constructing $\mathbf{f} \circ H^{-1}$. For example, let H be the CDF of h , the density function of the stage, i.e. $H(s) = \int_{-\infty}^s h(\tau) d\tau$, then $\mathbf{f} \circ H^{-1}$ would satisfy the two properties. An important theoretical result we derive in this paper is that for H being the CDF of h , $\mathbf{f} \circ H^{-1}$ can be directly obtained by deconvolving the marginal CDF of each biomarker:

Theorem 3. *Suppose the data point random variable $\mathbf{x} = [x_1, \dots, x_B]$ is generated by*

$$s \sim h(s), \quad x_j \sim \mathcal{N}(f_j(s), \sigma_j^2)$$

where $f_j : [0, 1] \rightarrow [f_j(0), f_j(1)]$ is a continuous bijection that monotonically increases. Then, the marginal CDF of x_j , F_j , is a result of convolving $H \circ \tilde{f}_j^{-1}$ with the noise density function $\mathcal{N}(\cdot; 0, \sigma_j^2)$:

$$(H \circ \tilde{f}_j^{-1}) * \mathcal{N}(\cdot; 0, \sigma_j^2) = F_j$$

where H is the CDF of h , and $\tilde{f}_j^{-1} : \mathbb{R} \rightarrow [0, 1]$ is an augmented version of f_j^{-1} that extends its domain to $(-\infty, +\infty)$:

$$\tilde{f}_j^{-1}(y) = \begin{cases} 0, & \text{if } y < f_j(0), \\ f_j^{-1}(y), & \text{if } f_j(0) \leq y \leq f_j(1), \\ 1, & \text{if } y > f_j(1). \end{cases}$$

In particular, when $\sigma_j = 0$, we have $H \circ \tilde{f}_j^{-1} = F_j$.

Sketch of Proof. Let $\phi(x; \mu, \sigma^2)$ be the CDF of $\mathcal{N}(x; \mu, \sigma^2)$, we can show $F_j(x) = \int_0^1 h(s)\phi(x; f_j(s), \sigma_j^2)ds$. Let $y = f_j(s)$, $F_j(x) = \int_{-\infty}^{+\infty} \frac{dH \circ \tilde{f}_j^{-1}(y)}{dy} \phi(x; y, \sigma_j^2)dy$. Using integration by parts, $F_j(x) = \phi(x - y; 0, \sigma_j^2) \cdot H \circ \tilde{f}_j^{-1}(y) \Big|_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} H \circ \tilde{f}_j^{-1}(y)d\phi(x - y; 0, \sigma_j^2)$ where the first term vanishes and the second is the result.

Theorem 3 verifies our intuition in Fig. 1. If h is a uniform distribution on $[0, 1]$, H would be the identity function, hence f_j is directly related to the CDF F_j via convolution. Moreover, when $\sigma_j = 0$, the CDF is the inverse function of the trajectory f_j . Note that without assuming a uniform distribution on h , the reparametrized trajectory $\{f_j \circ H^{-1}\}$ still satisfies the OPP and DPP, hence suffices for ordering any events and separating the data points. This suggests a new algorithm to estimate the progression trajectory and cluster the data points.

3.2 Algorithm

According to Theorem 3, we only need to deconvolve the marginal CDF F_j using the noise density $\mathcal{N}(\cdot; 0, \sigma_j^2)$ to obtain $H \circ \tilde{f}_j^{-1}$. Then, $H \circ \tilde{f}_j^{-1}$ can be truncated to $H \circ f_j^{-1}$ according to some range $[f_j(0), f_j(1)]$ and the inverses $\{f_j \circ H^{-1} : j = 1, \dots, B\}$ can be combined to obtain the reparametrized trajectory $\mathbf{f} \circ H^{-1}$ (see Fig. 2). The details are given below.

Let $H \circ \tilde{f}_j^{-1}$, $\mathcal{N}(\cdot; 0, \sigma_j^2)$, and F_j be discretized into $\tilde{\mathbf{s}} = [\mathbf{0}_{l_0}, \mathbf{s}, \mathbf{1}_{l_1}] \in \mathbb{R}^{n+2l}$, $\mathbf{g} \in \mathbb{R}^{2l+1}$, and $\mathbf{F}_j \in \mathbb{R}^n$ respectively, where l_0 and l_1 are the lengths of the leading zeros and trailing ones. The convolution $(H \circ \tilde{f}_j^{-1}) * \mathcal{N}(\cdot; 0, \sigma_j^2) = F_j$ can be discretized into $\mathbf{K}\tilde{\mathbf{s}} = \mathbf{F}_j$, where $\mathbf{K} \in \mathbb{R}^{n \times (n+2l)}$ is a circulant matrix with $\mathbf{K}(i, i : i + 2l) = \mathbf{g}$. Let $\mathbf{K} = [\mathbf{K}_1, \mathbf{K}_2, \mathbf{K}_3]$ where \mathbf{K}_1 has l_0 columns, \mathbf{K}_3 has l_1 columns. The discretized equation becomes $\mathbf{K}_2\mathbf{s} + \mathbf{K}_3\mathbf{1}_{l_1} = \mathbf{F}_j$, or $\mathbf{A}\mathbf{s} = \mathbf{b}$ with $\mathbf{A} = \mathbf{K}_2$ and $\mathbf{b} = \mathbf{F}_j - \mathbf{K}_3\mathbf{1}_{l_1}$. Hence, \mathbf{s} can be retrieved by minimizing $\|\mathbf{A}\mathbf{s} - \mathbf{b}\|$. To keep the trajectory smooth, we add a second derivative Laplacian term $\mathbf{L}\mathbf{s}$ where $\mathbf{L}(i, i : i + 2) = [-1, 2, -1]$. The final objective function is

$$J(\mathbf{s}) = \|\mathbf{A}\mathbf{s} - \mathbf{b}\|^2 + \lambda\|\mathbf{L}\mathbf{s}\|^2 \text{ s.t. } 0 \leq s_j \leq s_{j+1} \leq 1,$$

which is a convex function with inequality constraints and can be solved by standard methods. In this work, we assume $f_j(0)$ is known to be 0 and $f_j(1)$ is unknown. Hence $l_0 = 2l$ while l_1 needs to be searched in a range to minimize $J(\mathbf{s})$. The resulting \mathbf{s} is a discretized version of $H \circ f_j^{-1}$.

This technique can be extended to cluster data points generated from multiple trajectories $\{\mathbf{f}^k : k = 1, 2, \dots, K\}$. Similar to k-means clustering, we can add a cluster label to each data point and alternately update the reparametrized trajectories, the cluster labels, and the noise variances (see Fig. 2). Note that the estimated trajectories $\hat{\mathbf{f}}^k = \mathbf{f}^k \circ (H^k)^{-1}$ may have different H^k s. To make their timelines be matched, we further reparameterize the trajectories by Euclidean arc length from the origin.

4 Results

4.1 Datasets

We use both simulated datasets and a real dataset for validation. The real dataset is from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) [16]. For each dataset, besides running the proposed FTR, we also run its variation without filtering (referred to as *FTR w.o.f.*, i.e. without the deconvolution procedure) for ablation study, and the original EBM (referred to as EBM) [2], the discriminative EBM (referred to as DEBM) [3], and SuStaIn [6] for comparison.

Simulated Datasets. For simulation, we create datasets for two experiments: (i) stage inference with a single trajectory; (ii) subtype and stage inference with multiple trajectories. In both cases, a trajectory is simulated by the following two ways:

1. *Sigmoid function.* The trajectory of the j th biomarker is assumed to follow $f_j(s) = 1/(1 + e^{-\alpha_j(s-\beta_j)})$, where α_j (controlling the slope of increase) is randomly sampled from a Gaussian distribution, $\alpha_j \sim \mathcal{N}(10, 4)$, and β_j (controlling the position of change) is sampled from a uniform distribution, $\beta_j \sim \mathcal{U}([0.1, 0.9])$. We shift each f_j vertically so that it starts at zero.
2. *Event permutation.* Assume event targets $e_l = l$ for $l = 1, 2, 3$ and $e_4 = 5$, and corresponding events $s_{jl} = f_j^{-1}(e_l)$. Randomly order $(s_{11}, \dots, s_{1,3}, \dots, s_{B,3})$ with rank r and constraint $r_{jl} < r_{j,l+1}, \forall j, l$. Then, for each j , interpolate $\{(r_{jl}, e_l) : l = 0, 1, \dots, 4\}$ ($r_{j0} = 0$ and $e_0 = 0$) linearly to obtain f_j .

Given the generated trajectory \mathbf{f} , the data points are generated by sampling stage s from a uniform distribution $\mathcal{U}([0, 1])$, and then sampling $\mathcal{N}(\mathbf{f}(s), \sigma^2 \mathbf{I})$ (covariance $(5\sigma)^2 \mathbf{I}$ for event permutation). For multiple trajectories $\{\mathbf{f}^k : k = 1, \dots, K\}$ ($K = 3$), a data point also has an associated subtype index z , which is sampled from a categorical distribution $z \sim \text{Cat}(1/K, \dots, 1/K)$ that puts equal probability to each subtype. A data point is generated by first sampling z , then sampling from trajectory \mathbf{f}^z .

The TADPOLE dataset. The Alzheimer’s Disease Prediction Of Longitudinal Evolution (TADPOLE) dataset contains preprocessed measurements from brain imaging, CSF, scores from cognitive tests, demographics, and genetics from ADNI 1, ADNI GO, and ADNI 2 [16]. There are 1737 subjects (age: 73.8 ± 7.2 , sex: 55% male) in its “D1_D2” dataset, and each subject has multiple visits with an interval of 0.5 or 1 year (also with missing visits). At each visit, a subject is also diagnosed with 3 labels: cognitively normal (CN), mild cognitive impairment (MCI), or AD.

We conducted 2 experiments. The first chooses 7 biomarkers including CSF measurements (ABETA, TAU and p-TAU), brain region volumes (Hippocampus, WholeBrain) and cognitive scores (MMSE, ADAS-Cog-13) for validating a single progression trajectory. In the second one, volume measurements of 84 brain regions from MRI scans are selected to identify subtypes with different spatiotemporal atrophy patterns. We remove the visit records with missing data,

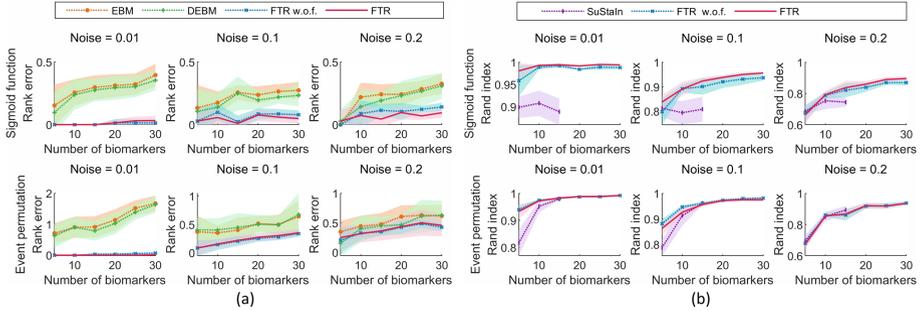


Fig. 3. Comparison of (a) rank error for a single trajectory and (b) Rand index for multiple trajectories in the simulation. Mean with standard deviation (shown in shade) is calculated over 10 randomly generated datasets for each number of biomarkers and each method (5 for SuStaIn). The first and the second rows correspond to using the sigmoid function and event permutation respectively.

leaving 693 subjects with 1812 time points in the first experiment and 1096 subjects with 3929 time points (1262 CN, 1732 MCI and 935 AD) in the second experiment. In both the experiments, we regress out age, sex and intra-cranial volume (ICV) and normalize the biomarker values into z-scores (the first experiment uses the normal group from fitted GMMs in [3] and the second one uses the CN group from the labels).

4.2 Simulation

In the simulation, we randomly generate a dataset containing 3000 data points 10 times for different numbers of biomarkers ($B = 5, 10, \dots, 30$) and different noise levels ($\sigma = 0.01, 0.1, 0.2$). For the single trajectory experiment, we also ran EBM [2] and DEBM [3] for comparison. For these two EBMs, points with stage $s < 0.1$ were set to CN and the others were set to AD. Since their objective is to estimate the correct order of biomarkers, we used the mean squared error of the ranks of the biomarkers as a metric. The results are shown in Fig. 3(a). We see that FTR has the lowest rank error in both types of trajectories, followed by FTR w.o.f., DEBM and EBM. Note that the inferior performance of EBM and DEBM in the small noise case ($\sigma = 0.01$) may be caused by their small estimated variances from the data.

For the multiple trajectory experiment, we also ran SuStaIn for comparison. For SuStaIn, we set 0.1, 0.5, 0.9, 1 (resp. 1, 2, 3, 5) for the event targets in the sigmoid function (resp. event permutation) simulation. For FTR, we ran the clustering algorithm 30 times with random initial label assignments and the best result in terms of reconstruction error was chosen. Since the objective of this task is to cluster the data points, we use Rand index as a metric. The results are shown in Fig. 3(b). We see that again FTR exceeds the baseline method in terms of clustering accuracy in most of the cases. Note that SuStaIn runs much

slower than our method hence it was only run on 5 random datasets at each number of biomarkers and also at maximum 15 biomarkers (see Table 1).

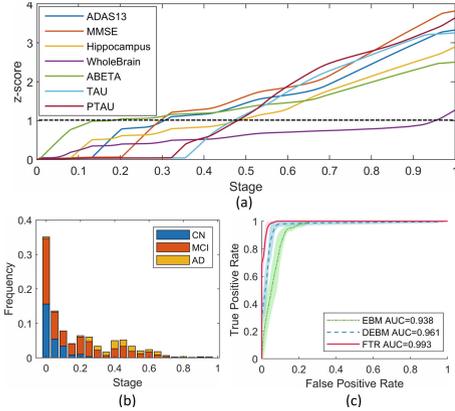


Fig. 4. Results of FTR on the 7 biomarkers from TADPOLE. (a) Estimated trajectory. (b) Distribution of estimated stages. (c) Mean ROC curves (standard deviation indicated by shade) for classifying AD from CN using the estimated stages from 5-fold cross validation.

4.3 ADNI

For the 7 biomarker experiment, the normalized z-scores of MMSE, Hippocampus, WholeBrain and ABETA decrease over time hence we add a minus sign to make them increase. The estimated trajectory from FTR is shown in Fig. 4(a). We can see that with a proper event target ($e = 1$), the order of the events—ABETA, MMSE, ADAS13, Hippocampus, TAU, p-TAU, WholeBrain—is exactly the same as that obtained from DEBM [3]. To validate the inferred stages, we split the data into a training set and a test set, trained FTR on the training set, inferred stages on the test set using the estimated trajectory, and classify AD from CN using these stages. The ROC curves from 5-fold cross validation show that FTR outperforms EBM and DEBM in this classification task, suggesting the superiority of our stage inference (see Fig. 4(c)).

The last experiment is on using FTR to identify progressive atrophy patterns in 84 brain regions measured by MRI scans. Following previous studies [1, 6], we use 3 subtypes in FTR. The trajectories are visualized by showing stages 0.2, 0.4, 0.6, 0.8, 1 in BrainPainter [17] (see Fig. 5(a)). We see that a typical progression (subtype 2) is identified along with 2 branches (subtype 1 and 3). We also plot the cognitive scores versus time from baseline for each subtype in Fig. 5(b). We can observe that the progression rate of cognitive scores calculated from linear regression is the largest for subtype 3, followed by subtype 2 and 1, which is consistent with the severity of atrophy progression from imaging.

Table 1. Running time (in seconds) of all the methods on a laptop with a Ryzen 7 4800H CPU and 16 GB memory. FTR ran much slower when $K = 3$ because the algorithm was repeated 30 times with random initializations.

Model	# of subtypes	# of biomarkers		
		5	10	15
EBM	$K = 1$	7.59	13.5	13.2
DEBM	$K = 1$	3.17	6.93	20.6
SuStaIn	$K = 3$	1299	6813	31099
FTR w.o.f.	$K = 1$	0.28	0.44	0.60
	$K = 3$	58.1	113	138
FTR	$K = 1$	6.34	12.6	19.7
	$K = 3$	760	1356	2062

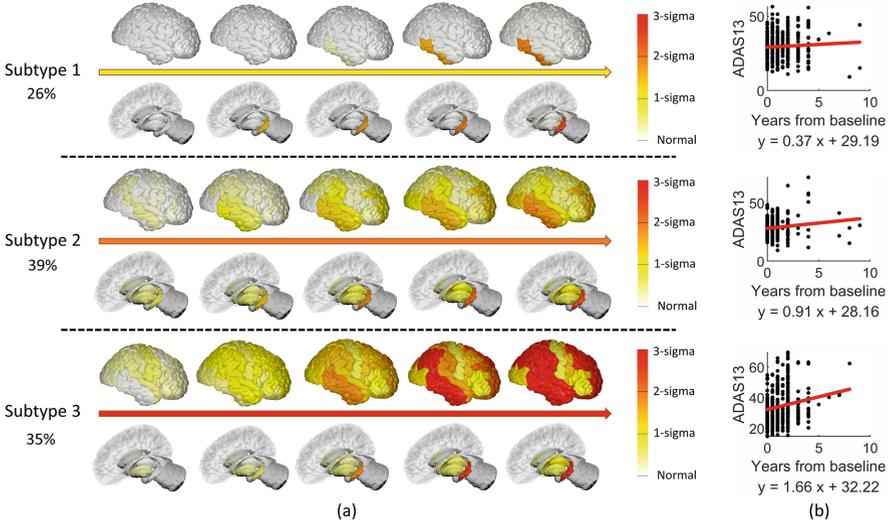


Fig. 5. Three subtypes identified by applying FTR on 84 brain regions from TAD-POLE. (a) Visualization of brain atrophy patterns at stage 0.2, 0.4, 0.6, 0.8, 1. (b) Linear regression of clinical symptom progression measured by ADAS-Cog-13.

5 Discussion and Conclusion

The results presented above show that the event order estimated from FTR is the same as that obtained from the EBM while the inferred stages are better to separate AD and CN. This is the first significant finding, which implies the superiority of our continuous trajectory estimation. The second, potentially more interesting, finding is that FTR identifies 3 fine-grained trajectories of spatiotemporal atrophy progression in 84 brain regions. Whether this subtyping scheme has any connection with the previous limbic predominant, typical, hippocampal sparing subtypes [1] will be investigated in the future.

The theory we developed on the relation between the disease progression trajectory and the CDF of the data could also have significant implications. Compared to the previous studies that make a strong assumption on the form of the trajectory function, we only assume that the function is monotonically increasing. This makes our method more flexible to retrieve a wide range of trajectories. Though the current version lacks some features like model selection, uncertainty estimation, we plan to solve these in the future.

In conclusion, this paper mathematically proves that the disease progression trajectory can be directly recovered from the CDF of the data. Based on this theory, we propose a novel FTR algorithm to estimate the trajectories and find the subtypes. FTR outperforms various state-of-the-art EBMs in our extensive simulation. It identifies a single trajectory using CSF, MRI biomarkers and cognitive scores. It also identifies 3 subtypes that differ in spatiotemporal patterns of progressive atrophy using volumes of 84 brain regions from MRI scans.

References

1. Ferreira, D., Nordberg, A., Westman, E.: Biological subtypes of Alzheimer's disease: a systematic review and meta-analysis. *Neurology* **94**(10), 436–448 (2020)
2. Fonteijn, H.M., et al.: An event-based model for disease progression and its application in familial Alzheimer's disease and Huntington's disease. *NeuroImage* **60**(3), 1880–1889 (2012)
3. Venkatraghavan, V., Bron, E.E., Niessen, W.J., Klein, S.: Disease progression timeline estimation for Alzheimer's disease using discriminative event based modeling. *NeuroImage* **186**, 518–532 (2019)
4. Venkatraghavan, V., et al.: Analyzing the effect of APOE on Alzheimer's disease progression using an event-based model for stratified populations. *Neuroimage* **227**, 117646 (2021)
5. Van Der Ende, E.L., et al.: A data-driven disease progression model of fluid biomarkers in genetic frontotemporal dementia. *Brain* **145**(5), 1805–1817 (2022)
6. Young, A.L., et al.: Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with subtype and stage inference. *Nat. Commun.* **9**(1), 1–16 (2018)
7. Vogel, J.W., et al.: Four distinct trajectories of tau deposition identified in Alzheimer's disease. *Nat. Med.* **27**(5), 871–881 (2021)
8. Firth, N.C., et al.: Sequences of cognitive decline in typical Alzheimer's disease and posterior cortical atrophy estimated using a novel event-based model of disease progression. *Alzheimer's & Dementia* **16**(7), 965–973 (2020)
9. Marinescu, R.V., et al.: DIVE: a spatiotemporal progression model of brain pathology in neurodegenerative disorders. *NeuroImage* **192**, 166–177 (2019)
10. Lorenzi, M., Filippone, M., Frisoni, G.B., Alexander, D.C., Ourselin, S., et al.: Probabilistic disease progression modeling to characterize diagnostic uncertainty: application to staging and prediction in Alzheimer's disease. *NeuroImage* **190**, 56–68 (2019)
11. Abi Nader, C., Ayache, N., Robert, P., Lorenzi, M., et al.: Monotonic Gaussian Process for spatio-temporal disease progression modeling in brain imaging data. *Neuroimage* **205**, 116266 (2020)
12. Garbarino, S., Lorenzi, M., Initiative, A.D.N., et al.: Investigating hypotheses of neurodegeneration by learning dynamical systems of protein propagation in the brain. *Neuroimage* **235**, 117980 (2021)
13. Raj, A., Kuceyeski, A., Weiner, M.: A network diffusion model of disease progression in dementia. *Neuron* **73**(6), 1204–1215 (2012)
14. Jung, W., Jun, E., Suk, H.-I., Initiative, A.D.N., et al.: Deep recurrent model for individualized prediction of Alzheimer's disease progression. *Neuroimage* **237**, 118143 (2021)
15. Nguyen, M., et al.: Predicting Alzheimer's disease progression using deep recurrent neural networks. *Neuroimage* **222**, 117203 (2020)
16. Marinescu, R.V., et al.: Tadpole challenge: prediction of longitudinal evolution in Alzheimer's disease. arXiv preprint [arXiv:1805.03909](https://arxiv.org/abs/1805.03909) (2018)
17. Marinescu, R., Eshaghi, A., Alexander, D.C., Golland, P.: BrainPainter: a software for the visualisation of brain structures, biomarkers and associated pathological processes. arXiv preprint [arXiv:1905.08627](https://arxiv.org/abs/1905.08627) (2019)